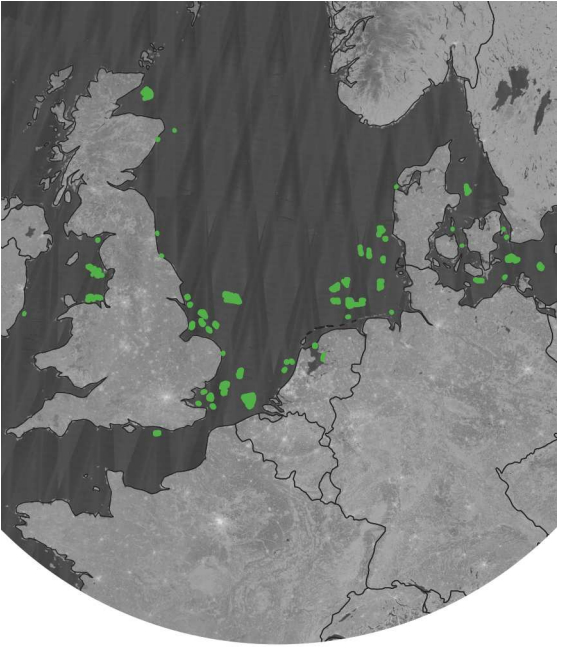
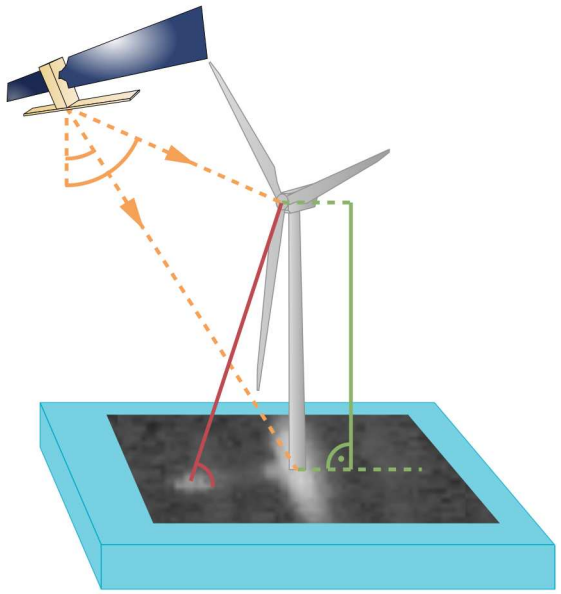
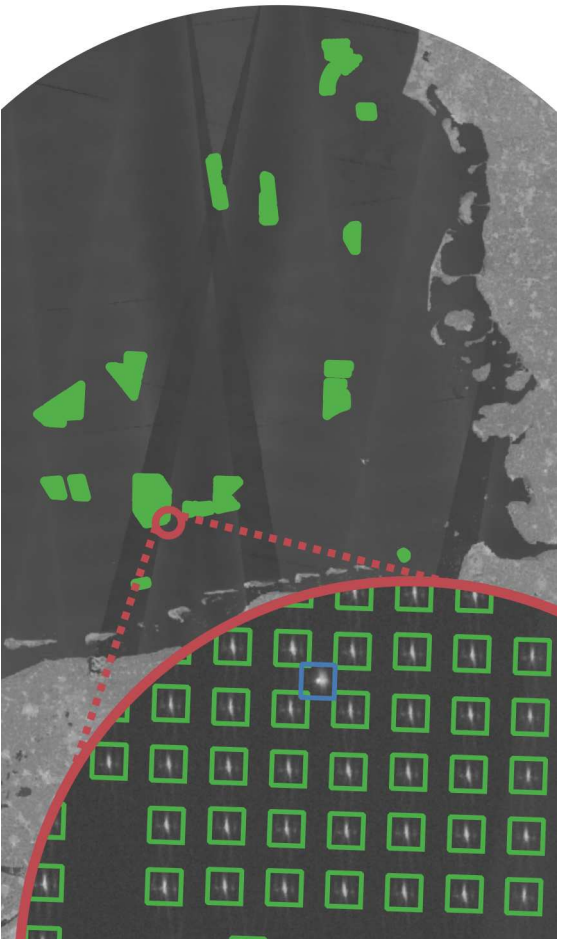


# Global Dynamics of the Offshore Wind Energy Sector Derived from Earth Observation Data



Deep Learning Based Object Detection Optimised with Synthetic Training Data for Offshore Wind Energy Infrastructure Extraction from Sentinel-1 Imagery



Dissertation zur Erlangung der Doktorwürde der Philosophischen Fakultät der Julius-Maximilians-Universität Würzburg

vorgelegt von

**Thorsten Höser**

Mai 2022

Julius-Maximilians-
UNIVERSITÄT
WÜRZBURG

Cover picture:
The upper and lower images are median composites of all ground range detected, VH polarised Sentinel-1 scenes acquired in the second quarter of 2021. The images show automatically detected offshore wind farms in the North Sea Basin and the German Bight. Furthermore, the magnification of the lower image shows detected offshore wind turbines and a single substation of the offshore wind farm Borkum Riffgrund 1. In the centre is a conceptual drawing of the Sentinel-1 sensor's acquisition geometry of an offshore wind turbine and how this geometry and the resulting radar signal can be used to calculate the turbine's hub height.

Eingereicht am: 19.05.2022

**Von:** Thorsten Höser

**Ort:** Lehrstuhl für Fernerkundung der Julius-Maximilians-Universität Würzburg, in Kooperation mit dem Deutschen Fernerkundungsdatenzentrum (DFD) des Deutschen Zentrums für Luft- und Raumfahrt (DLR)

**Erstbetreuerin:** Prof. Dr. Claudia Künzer, Universität Würzburg

**Zweitbetreuer:** Prof. Dr. Heiko Paeth, Universität Würzburg

This dissertation was prepared at the department “Land Surface Dynamics” of the German Remote Sensing Data Center (DFD), Earth Observation Center (EOC), German Aerospace Center (DLR), Oberpfaffenhofen, Germany.

Deutsches Zentrum
für Luft- und Raumfahrt
German Aerospace Center

*With enough computation and enough data, learning beats programming for complicated tasks that require the integration of many different, noisy cues.*

*Krizhevsky, Sutskever and Hinton (2017)*

# *English Summary*

The expansion of renewable energies is being driven by the gradual phaseout of fossil fuels in order to reduce greenhouse gas emissions, the steadily increasing demand for energy and, more recently, by geopolitical events. The offshore wind energy sector is on the verge of a massive expansion in Europe, the United Kingdom, China, but also in the USA, South Korea and Vietnam. Accordingly, the largest marine infrastructure projects to date will be carried out in the upcoming decades, with thousands of offshore wind turbines being installed. In order to accompany this process globally and to provide a database for research, development and monitoring, this dissertation presents a deep learning-based approach for object detection that enables the derivation of spatiotemporal developments of offshore wind energy infrastructures from satellite-based radar data of the Sentinel-1 mission.

For training the deep learning models for offshore wind energy infrastructure detection, an approach is presented that makes it possible to synthetically generate remote sensing data and the necessary annotation for the supervised deep learning process. In this synthetic data generation process, expert knowledge about image content and sensor acquisition techniques is made machine-readable. Finally, extensive and highly variable training data sets are generated from this knowledge representation, with which deep learning models can learn to detect objects in real-world satellite data.

The method for the synthetic generation of training data based on expert knowledge offers great potential for deep learning in Earth observation. Applications of deep learning based methods can be developed and tested faster with this procedure. Furthermore, the synthetically generated and thus controllable training data offer the possibility to interpret the learning process of the optimised deep learning models.

The method developed in this dissertation to create synthetic remote sensing training data was finally used to optimise deep learning models for the global detection of offshore wind energy infrastructure. For this purpose, images of the entire global coastline from ESA's Sentinel-1 radar mission were evaluated. The derived data set includes over 9,941 objects, which distinguish offshore wind turbines, transformer stations and offshore wind

energy infrastructures under construction from each other. In addition to this spatial detection, a quarterly time series from July 2016 to June 2021 was derived for all objects. This time series reveals the start of construction, the construction phase and the time of completion with subsequent operation for each object.

The derived offshore wind energy infrastructure data set provides the basis for an analysis of the development of the offshore wind energy sector from July 2016 to June 2021. For this analysis, further attributes of the detected offshore wind turbines were derived. The most important of these are the height and installed capacity of a turbine. The turbine height was calculated by a radargrammetric analysis of the previously detected Sentinel-1 signal and then used to statistically model the installed capacity. The results show that in June 2021, 8,885 offshore wind turbines with a total capacity of 40.6 GW were installed worldwide. The largest installed capacities are in the EU (15.2 GW), China (14.1 GW) and the United Kingdom (10.7 GW). From July 2016 to June 2021, China has expanded 13 GW of offshore wind energy infrastructure. The EU has installed 8 GW and the UK 5.8 GW of offshore wind energy infrastructure in the same period. This temporal analysis shows that China was the main driver of the expansion of the offshore wind energy sector in the period under investigation.

The derived data set for the description of the offshore wind energy sector was made publicly available. It is thus freely accessible to all decision-makers and stakeholders involved in the development of offshore wind energy projects. Especially in the scientific context, it serves as a database that enables a wide range of investigations. Research questions regarding offshore wind turbines themselves as well as the influence of the expansion in the coming decades can be investigated. This supports the imminent and urgently needed expansion of offshore wind energy in order to promote sustainable expansion in addition to the expansion targets that have been set.

# *Deutsche Zusammenfassung*

Der Ausbau erneuerbarer Energien wird durch den sukzessiven Verzicht auf fossile Energieträger zur Reduktion der Treibhausgasemissionen, dem stetig steigenden Energiebedarf sowie, in jüngster Zeit, von geopolitischen Ereignissen stark vorangetrieben. Der offshore Windenergiesektor steht in Europa, dem Vereinigten Königreich, China, aber auch den USA, Süd-Korea und Vietnam vor einer massiven Expansion. In den nächsten Dekaden werden die bislang größten marinen Infrastrukturprojekte mit tausenden neu installierten offshore Windturbinen realisiert. Um diesen Prozess global zu begleiten und eine Datengrundlage für die Forschung, für Entscheidungsträger und für ein kontinuierliches Monitoring bereit zu stellen, präsentiert diese Dissertation einen Deep Learning basierten Ansatz zur Detektion von offshore Windkraftanalagen aus satellitengestützten Radardaten der Sentinel-1 Mission.

Für das überwachte Training der verwendeten Deep Learning Modelle zur Objektdetektion wird ein Ansatz vorgestellt, der es ermöglicht, Fernerkundungsdaten und die notwendigen Label synthetisch zu generieren. Hierbei wird Expertenwissen über die Bildinhalte, wie offshore Windkraftanlagen aber auch ihre natürliche Umgebung, wie Küsten oder andere Infrastruktur, gemeinsam mit Informationen über den Sensor strukturiert und maschinenlesbar gemacht. Aus dieser Wissensrepräsentation werden schließlich umfangreiche und höchst variable Trainingsdaten erzeugt, womit Deep Learning Modelle die Detektion von Objekten in Satellitendaten erlernen können.

Das Verfahren zur synthetischen Erzeugung von Trainingsdaten basierend auf Expertenwissen bietet großes Potential für Deep Learning in der Erdbeobachtung. Deep Learning Ansätze können hierdurch schneller entwickelt und getestet werden. Darüber hinaus bieten die synthetisch generierten und somit kontrollierbaren Trainingsdaten die Möglichkeit, den Lernprozess der optimierten Deep Learning Modelle zu interpretieren.

Das in dieser Dissertation für Fernerkundungsdaten entwickelte Verfahren zur Erstellung synthetischer Trainingsdaten wurde schließlich zur Optimierung von Deep Learning Modellen für die globale Detektion von offshore Windenergieanlagen eingesetzt. Hierfür wurden Aufnahmen der gesamten globalen Küstenlinie der Sentinel-1 Mission der ESA

ausgewertet. Der abgeleitete Datensatz, welcher 9.941 Objekte umfasst, unterscheidet offshore Windturbinen, Trafostationen und im Bau befindliche offshore Windenergieinfrastrukturen voneinander. Zusätzlich zu dieser räumlichen Detektion wurde eine vierteljährliche Zeitreihe von Juli 2016 bis Juni 2021 für alle Objekte generiert. Diese Zeitreihe zeigt den Start des Baubeginns, die Bauphase und den Zeitpunkt der Fertigstellung mit anschließendem Betrieb für jedes Objekt.

Der gewonnene Datensatz dient weiterhin als Grundlage für eine Analyse der Entwicklung des offshore Windenergiesektors von Juli 2016 bis Juni 2021. Für diese Analyse wurden weitere Attribute der Turbinen abgeleitet. In einem radargrammetrischen Verfahren wurde die Turbinenhöhe berechnet und anschließend verwendet, um die installierte Leistung statistisch zu modellieren. Die Ergebnisse hierzu zeigen, dass im Juni 2021 weltweit 8.885 offshore Windturbinen mit insgesamt 40,6 GW Leistung installiert waren. Die größten installierten Leistungen stellen dabei die EU (15,2 GW), China (14,1 GW) und das Vereinigte Königreich (10,7 GW). Von Juli 2016 bis Juni 2021 hat China 13 GW installierte Leistung ausgebaut. Die EU hat im selben Zeitraum 8 GW und das Vereinigte Königreich 5,8 GW offshore Windenergieinfrastruktur installiert. Diese zeitliche Analyse verdeutlicht, dass China der maßgebliche Treiber in der Expansion des offshore Windenergiesektors im untersuchten Zeitraum war.

Der abgeleitete Datensatz zur Beschreibung des offshore Windenergiesektors wurde öffentlich zugänglich gemacht. Somit steht er allen Entscheidungsträgern und Stakeholdern, die am Ausbau von offshore Windenergieanlagen beteiligt sind, frei zur Verfügung. Vor allem im wissenschaftlichen Kontext dient er als Datenbasis, welche unterschiedlichste Untersuchungen ermöglicht. Hierbei können sowohl Forschungsfragen bezüglich der offshore Windenergieanlagen selbst, als auch der Einfluss des Ausbaus der kommenden Dekaden untersucht werden. Somit wird der bevorstehende und dringend notwendige Ausbau der offshore Windenergie unterstützt, um neben den gesteckten Zielen auch einen nachhaltigen Ausbau zu fördern.

# *Acknowledgement*

During this dissertation, I have received support from many sides, for which I am very grateful and which have made this special time a remarkable experience for me.

Above all, I would like to thank my supervisor at the German Remote Sensing Data Center (DFD) of the German Aerospace Center (DLR), Prof. Dr. Claudia Kuenzer. Her advice and guidance were of extraordinary help throughout the dissertation. In particular, I would like to thank her for encouraging me to achieve goals that I would not have tackled without her experience and support. The constant exchange and discussions about the developed methods and their applications were of great importance for the progress this work has made and the scientific publications that were written during this time. This dissertation under her supervision was a great experience with an outstanding exchange of ideas and fascination about Earth observation.

Furthermore, I would like to thank my mentors at the University of Wuerzburg, Prof. Dr. Heiko Paeth and PD Dr. Tobias Ullmann for their support of this dissertation and exchange, which has always led to new ideas and critical considerations.

Special thanks go to my team leaders Juliane Huth and Dr. Felix Bachofer, who welcomed me warmly to DLR and provided excellent organisation and thoughtful support, especially during the difficult time of the pandemic. I would also like to thank the entire Coast and River Basins team Dr. Marco Ottinger, Dr. Stefanie Feuerstein, Soner Üreyen, Verena Jaspersen, Dr. David Marshall, Dr. Igor Klein, Dr. Christina Eisfelder, Patrick Sogno, Zhiyuan Wang and Tuyen Ha Van for their day to day cooperation, exchange of ideas and helpful support.

I would like to thank my PhD fellows, Aiym Orynbaikyzy, Sophie Reinermann, Mariel Dirscherl, Dr. Celia Baumhoer, Dr. Ya-Lun Tsai, Jonas Koehler, Dr. Zhongyang Hu, Marius Philipp and Daniela Palacios Lopez from the Land Surface Dynamics department for the time we have spent working together on our dissertations. I would like to thank and express my admiration for my colleague Julian Zeidler for his continuous development and administration of the department's IT infrastructure and always great support when needed. You all have been an exceptionally supportive group throughout the time working on my PhD.

A heartfelt thank you goes to my family Christoph, Almut, Moritz, Carolina and Valentina, for your constant support from the beginning of my studies to this dissertation. I want to thank Janina for a time in Munich that couldn't have gone better, and Thore, for your constant reminder to get to the point and to look forward to more than only writing papers. And finally, I would like to thank an exceptional group of friends. How boring it would be without you, how extraordinary it is with you. In particular, I would like to thank Matthias, Phil and Lukas for our time on deck, which has brought me through the turbulence of recent times in good spirits.

# *Table of Contents*

# *List of Figures*

# *List of Tables*

# *Abbreviations and Acronyms*

| | |
|---|---|
| AI | Artificial Intelligence |
| ALOS-PALSAR | Advanced Land Observing Satellite-Phased Array L-Band Synthetic Aperture Radar |
| ANN | Artificial Neural Network |
| AP | Average Precision |
| API | Application Programming Interface |
| ASPP | Atrous Spatial Pyramid Pooling |
| ASTER | Advanced Spaceborne Thermal Emission and Reflection Radiometer |
| CBNet | Composite Backbone Network |
| CFAR | Constant False Alarm Rate |
| CGI | Computer Generated Imagery |
| CI | Confidence Interval |
| CNN | Convolutional Neural Network |
| COP26 | 26th United Nations Climate Change Conference of the Parties |
| CRS | Coordinate Reference System |
| DL | Deep Learning |
| DLR | German Aerospace Center |
| DoG | Difference of Gaussians |
| DOTA | Dataset for Object Detection in Aerial Images |
| ECS | East China Sea |
| EEZ | Exclusive Economic Zones |
| ERS | European Environmental Satellites |
| ESA | European Space Agency |
| EU | European Union |
| Faster R-CNN | Faster Region based-Convolutional Neural Network |
| FCN | Fully Convolutional Network |
| FP | False Ppositive |
| FPN | Feature Pyramid Network |
| FN | False Negative |
| GEE | Google Earth Engine |
| GER | Germany |
| GIS | Geographic Information System |
| GOWT | Global Offshore Wind Turbine |
| GRD | Ground Range Detected |
| GPU | Graphics Processing Unit |
| GT | Ground Truth |
| GWEC | Global Wind Energy Council |

| | |
|---|---|
| GID | Gaofen Image Dataset |
| HRSC2016 | High Resolution Ship Collection 2016 |
| ID | Identifier |
| ILSVRC | ImageNet Large Scale Visual Recognition Challenge |
| IoU | Intersection over Union |
| IS | Image Segmentation |
| IW | Interferometric Wide swath |
| ISPRS | International Society for Photogrammetry and Remote Sensing |
| IEEE | Institute of Electrical and Electronics Engineers |
| INRIA | Institut National de Recherche en Informatique et en Automatique |
| JRC | Joint Research Center |
| LCLU | Land Cover Land Use |
| LCZ42 | Local Climate Zones |
| mIoU | mean Intersection over Union |
| ML | Machine Learning |
| MLP | Multi Layer Perceptron |
| MNDWI | Modified Normalized Difference Water Index |
| MS-COCO | Microsoft-Common Objects in Context |
| MSL | Mean Sea Level |
| NAS | Neural Architecture Search |
| NOAA | National Oceanic and Atmospheric Administration |
| NSB | North Sea Basin |
| NWPU | Northwestern Polytechnical University |
| OBOE | Extensible Observation Ontology |
| OD | Object Detection |
| OSM | Open Street Map |
| OWF | Offshore Wind Farm |
| OWT | Offshore Wind Turbine |
| PANet | Path Aggregation Network |
| PASCAL-VOC | Pattern Analysis, Statistical modelling and ComputAtional Learning-Visual Object Classes |
| Pr | Precision |
| PSPNet | Pyramid Scene Parsing Network |
| Rc | Recall |
| R-CNN | Region based-Convolutional Neural Network |
| ReLU | Rectified Linear Unit |
| RFCN | Region based Fully Convolutional Network |
| RGB | Red-Green-Blue |
| RoI | Region of Interest |
| RPN | Region Proposal Network |
| RSOD | Remote Sensing Object Detection |
| RESICS | Remote Sensing Image Scene Classification |
| SAR | Synthetic Aperture Radar |
| SPPNet | Scene Pyramid Pooling Network |
| SSD | Single Shot multibox Detector |
| SegNet | Segmentation Network |
| SQLite | Strucutred Query Language data base management sytem |
| SLC | Single Look Complex |

| | |
|---|---|
| SNIPER | Scale Normalization for Image Pyramids with Efficient Resampling |
| SSDD | SAR Ship Detection Dataset |
| TP | True Ppositive |
| UAV | Unmanned Aerial Vehicle |
| UK | United Kingdom |
| US | United States |
| UTM | Universal Transverse Mercator |
| UCAS-AOD | University of Chinese Academy of Sciences-Aerial Object Detection |
| UN | United Nations |
| VEDAI | Vehicle Detection in Aerial Imagery |
| VGG | Visual Geometry Group |
| VH | Vertical transmit Horizontal receive |
| VHR | Very Heigh Resolution |
| VV | Vertical transmit Vertical receive |
| WGS84 | World Geodetic System 1984 |
| WHU | Wuhan University |
| YOLO | You Only Look Once |
| ZFNet | Zeiler and Fergus Network |

# *Chapter 1*

# *Introduction*

## 1.1 Scientific and Social Relevance

The growing global population and the associated growth in resource consumption, especially in fossil fuels, are the main drivers of steadily increasing green house gas emissions over the last decades (Steffen et al., 2020). With increasing green house gas concentration in the atmosphere, global warming is getting amplified, which widely affects the Earth system. Rising sea levels (Oppenheimer et al., 2019), more frequent extreme weather events like heavy rainfalls (Guerreiro et al., 2018; Chen et al., 2021) or droughts (Gu et al., 2020; Arneth et al., 2019), and the loss in biodiversity (Arneth et al., 2019; Chen et al., 2021) are examples how the livelihoods of humans are impacted negatively in a feedback mechanism from changes in the Earth system which are strongly related to the emissions of green house gases by humans (Steffen et al., 2020).

In response to the impacts of climate change, in 2021, at the 26th United Nations Climate Change Conference of the Parties (COP26), 65 nations agreed to a coal phaseout until 2040 to effectively reduce green house gas emissions (United Nations, 2021a). The urgency of this agreement becomes clear when looking at the green house gas emissions from fossil fuels in figure 1.1. It shows the development of $CO_2$ emissions from fossil fuel combustion since 1970 until 2020. In 2018, the $CO_2$ emissions of the three largest fossil energy sources amounted to around 34 Gt, with coal contributing 14.7 Gt (IEA, 2021b; Ritchie and Roser, 2020a). Apart from crises such as the global economic crisis in 2008/2009 or the effects of the pandemic in 2020, fossil fuel consumption and, with it, the $CO_2$ related emissions have been rising over the last decades. First statistics for 2021 report a rebound of $CO_2$ emissions by fossil fuels to pre-pandemic levels, or even higher, due to an increase in electricity generation based on coal (IEA, 2021a). Without a reduction in green house gas emissions, an increase of 2 °C of the mean average global temperature compared to pre-industrial

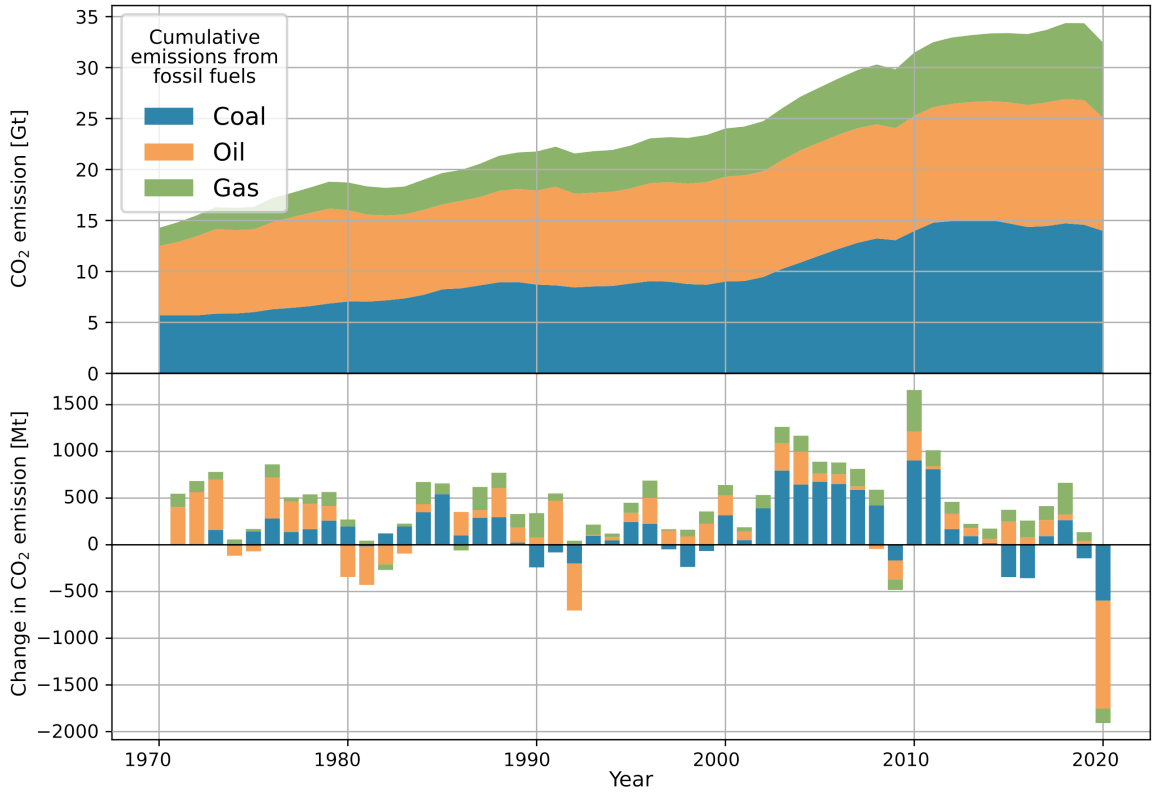



**Figure 1.1:** Global $CO_2$ emission by fossil fuels (coal, oil and gas) between 1970 and 2020 and the corresponding changes of $CO_2$ emissions from fossil fuel combustion relative to the previous year. Data source: Ritchie and Roser (2020a).

levels will take place within this century (Chen et al., 2021). The objectives of the Paris Agreement (United Nations, 2015) from 2015 would thus be missed. Decarbonisation of the energy sector as one of the largest green house gas emitters is hence a major concern in order to take action against climate change with more than just rhetoric (IEA, 2021b).

In 2020, over 60% of electrical energy was produced by burning fossil fuels, with coal at 33% is the most widely used primary energy source globally, see figure 1.2a). For the entire decarbonisation of the energy sector, large parts of existing power generating infrastructure have to be replaced. As stated earlier, coal-based electricity generation experienced a great increase after the pandemic plunge in 2020. This trend came from emerging economies with a rapidly growing demand for electrical energy. Here, new coal-based power plants are getting connected to the electrical grid, causing an increase in $CO_2$ emissions (IEA, 2021a). For decarbonisation of the energy sector in the medium term, however, it is necessary to expand carbon-neutral instead of coal-based energy production to cover the steadily increasing demand for electrical energy. Finally, about half of the $CO_2$ emissions by fossil fuels are not related to the generation of electricity or heat. One major green house gas emitter is the transportation sector (Climate Watch, 2021). Here, oil combustion is the primary

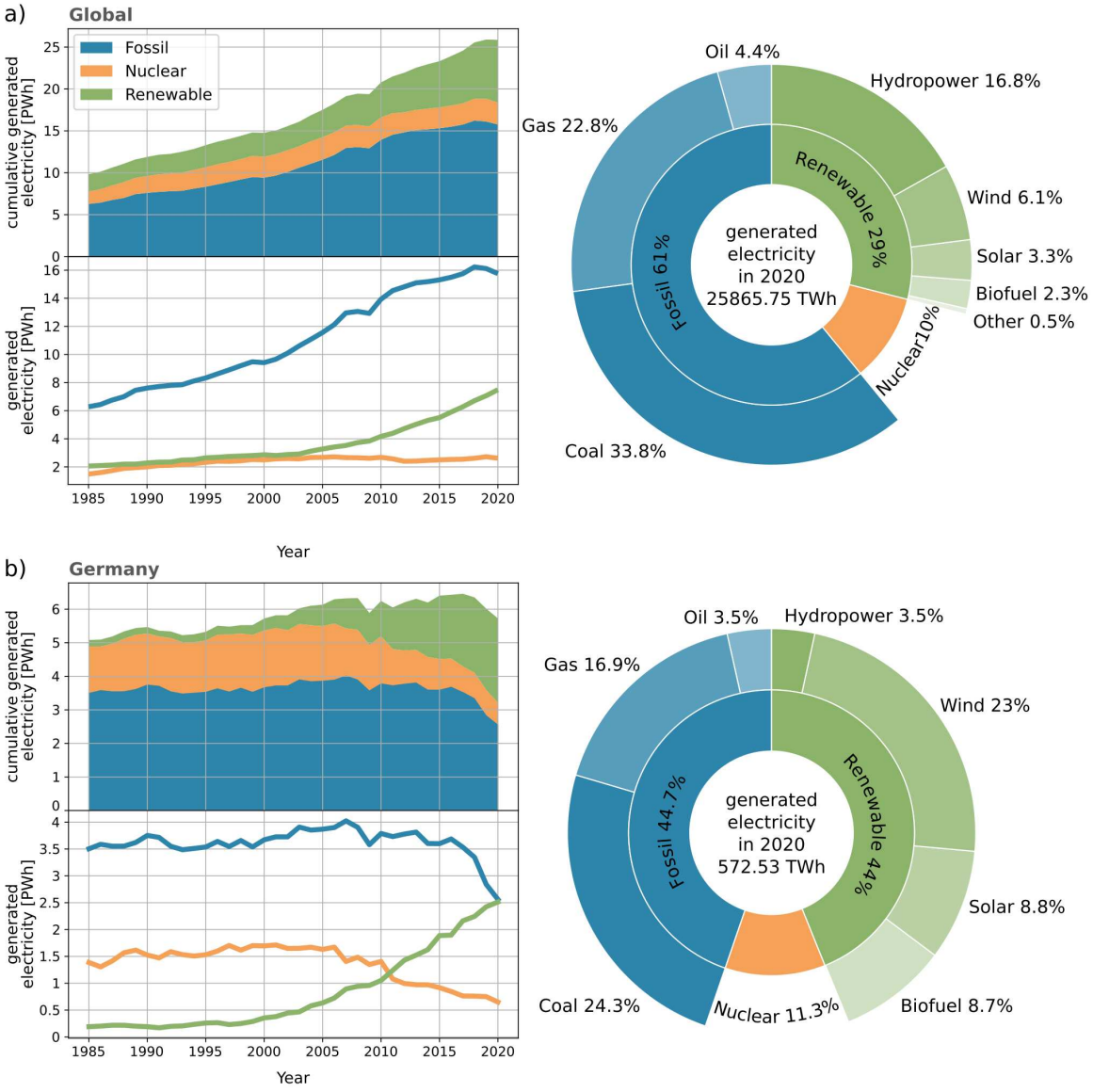



**Figure 1.2:** a) Global and b) German development of generated electricity from 1985 until 2020 for the three major primary energy sources, fossil fuels, renewables, and nuclear and the share of primary energy sources used for generating electricity in 2020. Data source: Ritchie and Roser (2020b).

energy source that directly produces $CO_2$ emissions. To further decrease $CO_2$ emissions from fossil fuels, another COP26 agreement calls for zero-emission vehicles (United Nations, 2021b). Such engines like electric motors or hydrogen fuel cells directly or indirectly need electricity, which offers the chance to use renewable energy in the transportation sector by storing electricity in batteries or producing green hydrogen. Nevertheless, as with the substitution of fossil fuel-based electricity generation, this additional electrical energy demand requires new infrastructure for generating electrical power.

Recent geopolitical events further amplify the need for alternative energy production. The Ukraine conflict is pushing nations to rapidly develop an energy economy which is independent of fossil fuel imports in order to guarantee geopolitical sovereignty, supply security, and economically viable as well as socially acceptable energy prices. From a German perspective, these geopolitical events and the accelerated coal phase-out coincide with the decision to phase out nuclear power at the end of 2022. Figure 1.2b) makes clear that despite a progressive expansion of electrical energy from renewable resources, coal and nuclear phaseout together would reduce the electrical energy production by over a third in Germany. The expected increase in electricity demand that will become necessary due to the progressive substitution of e.g. internal combustion engines has not yet been considered. Overall, future electrical energy generation is thus confronted to substantially grow its capacity in carbon-neutral energy production to cover the historical fossil fuel-based energy supply as well as to cover the steadily increasing demand for energy worldwide. The need for the development of alternative energy production to substitute existing fossil fuel-based power plants and production cycles as well as to meet increasing future energy demands is increasingly reinforced by the various environmental and geopolitical driving forces.

In order to achieve the transformation of the energy sector, an essential building block is the massive expansion of electricity generation from renewable energy sources. In 1991, Denmark installed the first offshore wind farm (OWF) Vindeby. Since then, the offshore wind energy sector has matured and, thirty years later, is in the early stage of a major growth phase. The recent expansion is driven above all by the urgent decarbonisation of the energy sector (Rodrigues et al., 2015). Offshore wind energy is an important cornerstone in a carbon-neutral energy mix due to its relatively constant availability, large untapped potential and cost-efficient energy production (Esteban et al., 2011). *The Net Zero Strategy: Build Back Greener* announced by the government of the United Kingdom (UK) in 2021 plans to increase the installed capacity of offshore wind energy infrastructure from 10.5 GW in 2020 to 40 GW by 2030 (UK Gov., 2021). Similarly, the European Union (EU) targets to increase its installed offshore wind energy capacity from 12 GW in 2020 to 60 GW by 2030 and further to 300 GW by 2050 with a total investment of EUR 800 billion, declared in *An EU Strategy to harness the potential of offshore renewable energy for a climate-neutral future* (EC, 2020). The Federal Ministry for Economic Affairs and Climate Action also announces that it will raise the offshore wind energy expansion targets for Germany to 30 GW in 2030 (BMWK, 2022).

These two strategies are exemplary for a global trend of increasing offshore wind energy capacities (Rodrigues et al., 2015). The realisation of such large scale infrastructure projects with thousands of offshore wind turbines (OWTs) and additional platforms to be in-

stalled within two to three decades represents a major technical, administrative, economic, but also ecological and social challenge. The planned offshore wind energy projects have to be realised in marine areas which are potentially used by other industries like fishery or shipping routes, are declared military exclusion zones or nature reserves or, in the case of near coast areas, raise additional conflicts with residents and the tourism industry due to a potential disruption of recreational space. Thus, the realisation of urgent carbon-neutral energy production by OWFs demands integrated spatial planning and monitoring of spatiotemporal developments to include all stakeholders and foster the exchange of information and knowledge between them (Gușatu et al., 2020, 2021). In order to support these marine infrastructure projects, which are globally widespread and unprecedented in scale, a freely accessible, global monitoring of the expansion of the offshore wind energy infrastructure is necessary. Open access to spatiotemporal data that describes the status and progression of the offshore wind energy sector allows all involved stakeholders to participate in the development of solutions to overcome the challenges related to the task of deploying thousands of offshore wind turbines in the upcoming decades.

In order to automatically derive a global data set of offshore wind energy infrastructure over a multi-year period that is independent of the accessibility of information like planning documents or operator specifications on offshore wind energy projects, spaceborne remote sensing imagery is used as a basis. With the increasing availability of spaceborne satellite images combined with modern methods of image analysis coming from the deep learning domain (LeCun et al., 2015) it is possible to derive small scale human-made object entities from Earth observation data. This makes spaceborne remote sensing imagery a strong, independent source to inform about the global developments of the offshore wind energy sector and provide freely accessible data to support integrated spatial planning. However, deep learning techniques applied to Earth observation data are in an ongoing adaption phase with open questions to be answered and domain-specific solutions to be investigated (Zhu et al., 2017; Reichstein et al., 2019; Ma et al., 2019b; Zhang et al., 2016b; Ball et al., 2017; Hoeser and Kuenzer, 2020; Hoeser et al., 2020).

Deep learning, especially convolutional neural networks (CNNs) for image analysis, received a lot of attention in 2012 when the CNN called AlexNet (Krizhevsky et al., 2012) won the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) (Russakovsky et al., 2015), an important challenge in the computer vision domain for image recognition. Figure 1.3 shows the impact of this specific CNN AlexNet after it has won the the ILSVRC in 2012. The paper of Krizhevsky et al. (2012) was cited widely in and outside the computer vision domain, which indicates that the approach presented was quickly taken up by other domains and also in practical applications (Goodfellow et al., 2016). The proportion

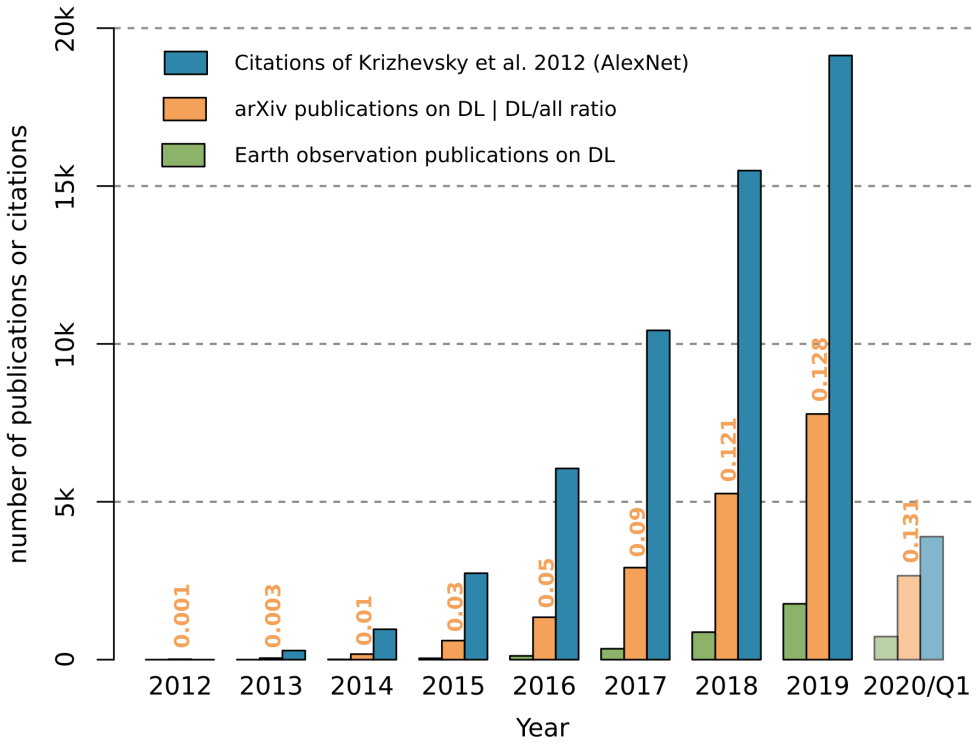



**Figure 1.3:** Development of deep learning (DL) related publications between 2012 and 2020 shown for the number of citations of Krizhevsky et al. (2012) and publications on arXiv and their share on all publications on this platform, and 16 leading Earth observation journals listed in figure 2.9. Changed after Hoeser and Kuenzer (2020, p. 2).

of publications on deep learning increased massively in the years following the publication of AlexNet, see figure 1.3. However, the onset of deep learning methods in the Earth observation domain started with a slight delay indicating domain-specific obstacles (Hoeser and Kuenzer, 2020). Over the years, deep learning became an important tool in analysing Earth observation data, and deep learning models such as the CNNs were further developed in the computer vision domain. A still ongoing challenge is to combine Earth observation data and deep learning to answer large scale research questions and implement robust workflows which are transferable in space and time (Ball et al., 2017; Hoeser and Kuenzer, 2020; Hoeser et al., 2020). Since the development of the offshore wind energy sector is happening on a global scale and already covers a period of multiple years, a workflow for the automatic extraction of its infrastructure from satellite imagery with deep learning methods faces these current challenges of deep learning applications in Earth observation.

## 1.2 Research Motivation

This work is motivated by two main issues. The first is the necessity of global and independent monitoring of the development of offshore wind farm infrastructure, which provides open access to the derived data to support the global challenge of expanding re-

newable energy production. The second issue is the application of deep learning on Earth observation data and thus a methodological focus. More specifically, deep supervised learning needs thousands of annotated training examples, no matter which topic or research domain. This prerequisite, the training data set, and its quality are of major concern regarding how well a deep learning model performs. At the same time, sufficient training data sets are very costly to compile. Therefore, the methodological emphasis of this dissertation is the development of an approach that allows for synthetically generating training data suitable for optimising deep learning models to investigate research questions and real-world problems in the Earth observation domain.

## 1.3 Research Objectives and Questions

Five research objectives were defined. The first three cover the methodological background from a deep learning perspective and the development of the synthetic data generation framework with a stronger focus on its application in the Earth observation domain. During their investigation, the methodological foundations are prepared on the basis of which the following two objectives regarding the global detection and analysis of offshore wind energy infrastructure are worked out. For each research objective, research questions were formulated in order to address specific issues and structure the research objectives into single milestones. A summary of the research objectives and how the research questions were addressed can be found in chapter 7.

***Research objective 1*** Provide an introduction to the fundamentals of supervised deep learning and convolutional neural networks, and conduct a literature review of how convolutional neural networks are applied in the Earth observation domain, in order to identify common practices and recent challenges specific for the Earth observation domain.

1. What are the key properties of the convolutional neural network that make it particularly suitable for analysing image data?
2. What is the value of training data for optimising a convolutional neural network?
3. How has the convolutional neural network been applied to Earth observation and what is a current obstacle in transferring this methodology from the field of computer vision?

***Research objective 2*** Provide an overview of how offshore wind energy infrastructure has been detected in Earth observation so far and give detailed insight into how offshore

wind energy infrastructure and its common environment appear in Sentinel-1 radar data at different spatial scales.

1. How has the detection of offshore wind energy infrastructure been researched in Earth observation so far?
2. What spatial features exist in radar images to detect and distinguish offshore wind energy infrastructure from other marine infrastructure and natural environments?

***Research objective 3*** Develop a conceptual framework to automatically generate large amounts of synthetic and task specific labelled Earth observation data by taking expert knowledge into account.

1. What structure can be used to uniformly represent expert knowledge in order to describe nested systems such as a remote sensing scene?
2. What properties must a synthetic training data set have in order to optimise a CNN so that it can reliably detect a target and at the same time distinguish it from its natural environment?

***Research objective 4*** Derive a global, multi-temporal data set of offshore wind energy infrastructure from Earth observation data by combining the previously developed framework for synthetic data generation and deep learning based object detection.

1. How can CNNs, optimised by synthetic training data, be used for object detection to globally identify individual types of offshore wind energy infrastructure in radar data?
2. How can quarterly changes in the deployment process of the offshore wind energy infrastructure be determined?
3. How many offshore wind turbines and substations had been deployed around the world by mid-2021 and how many were under construction?

***Research objective 5*** Expand the spatiotemporal information of the derived offshore wind energy infrastructure data set with important technical attributes such as the installed capacity of an offshore wind turbine as well as further spatial information about its location and conduct a global analysis about the dynamics of the offshore wind energy sector over the last five years between July 2016 and June 2021.

1. How can the installed capacity of an offshore wind turbine be determined using spaceborne radar data?

2. How far has the global expansion of offshore wind energy progressed by June 2021 measured in installed capacity and which regional differences can be identified?

3. What trends in the expansion of offshore wind energy have developed globally and regionally over the last five years?

## 1.4 Thesis Outline

***Chapter 1*** provides an overview of the scientific relevance and motivation of this dissertation. Furthermore, five research objectives along with their associated research questions are presented to determine the research focus of this dissertation.

***Chapter 2*** introduces the fundamentals of machine and deep learning with a special focus on the convolutional neural network. After this theoretical introduction, an Earth observation perspective is adopted to review the developments and applications of the convolutional neural network within the Earth observation domain.

***Chapter 3*** provides an overview of how marine and offshore wind energy infrastructure has been detected using remote sensing data to date. Furthermore, it provides an impression of how offshore wind energy infrastructure and its common natural environment appear in Sentinel-1 radar remote sensing data. Thereby a special focus is put on spatial features since they will be of importance for the later employed convolutional neural network object detector.

***Chapter 4*** focuses on the methodological development of the synthetic data generation approach. The general motivation for synthetic data is outlined, and the developed framework for generating synthetic data for Earth observation (SyntEO), is explained in depth. Theoretical explanations are underpinned with examples of the generation of synthetic images of offshore wind farms. The developed SyntEO approach is used to detect offshore wind farms on four test sites in a proof of concept study. Thereby an in-depth discussion of a sufficient data set compilation and how the convolutional neural network learns representations of the synthetic training data is carried out.

***Chapter 5*** applies the developed methods to detect offshore wind energy infrastructure globally. For this purpose, a deep learning-based workflow is implemented using the SyntEO framework. This deep learning-based workflow detects offshore wind energy infrastructure in Sentinel-1 data on a global scale automatically. Furthermore, the temporal dynamics of the detected objects are investigated and finally compiled into an offshore wind energy infrastructure data set, that is called the DeepOWT data set.

***Chapter 6*** deepens the focus on the offshore wind energy sector. The DeepOWT data set is enriched with additional attributes through spatial analysis and further examination of the radar signal, statistical modelling, and geographic information system (GIS) analysis. Finally, the spatiotemporal dynamics of the global offshore wind energy sector are derived from the data set. In the following discussion, global and regional trends are analysed, and regional differences are highlighted.

***Chapter 7*** summarises and concludes the developments and results of this work by reflecting on the research objectives and questions outlined in section 1.3. Finally, open questions and potential opportunities that have arisen during this dissertation are provided as an outlook for future research and applications.

# Chapter 2

# Convolutional Neural Networks and their Application in Earth Observation*

## 2.1 Artificial Intelligence

Figure 2.1 shows a Venn diagram that relates the term convolutional neural network (CNN) to its superordinate categories. To introduce the motivation and fundamental concepts of CNNs, these categories are going to be introduced consecutively. The main features of these categories will be identified and finally brought together in the introduction to CNNs and their application in Earth observation.

Artificial intelligence (AI) is among the most influential fields of research in the early 21st century (Zhang et al., 2021a; Liu et al., 2018). One reason for this is the vastness of the field of artificial intelligence, reaching from philosophical aspects to mathematical details and theoretical frameworks in computer science to practical implementations in software engineering. A definition of artificial intelligence is proposed by Rich and Knight (1991), which should give an idea of what is meant without being too explicit: "Artificial intelligence is the study of how to make computers do things which, at the moment, people do better" (Rich and Knight, 1991, p. 3). In order to take a closer look at artificial intelligence, two different variants can be distinguished. On the highest level, artificial intelligence can be understood as a non-natural mechanism, a so-called artificial intelligent agent that makes decisions or takes action based on inputs. Depending on the complexity of inputs and the degree of the flexibility of decisions the intelligent agent can process, the type of artificial intelligence is described as general or narrow. Thereby, artificial *general* intelligence has to be fully generalisable. That means it is task agnostic and able to generalise from past experiences to optimise actions in the future or even be somehow successful in an entirely new

*Parts of this chapter have been published in Hoeser and Kuenzer (2020) and Hoeser et al. (2020).

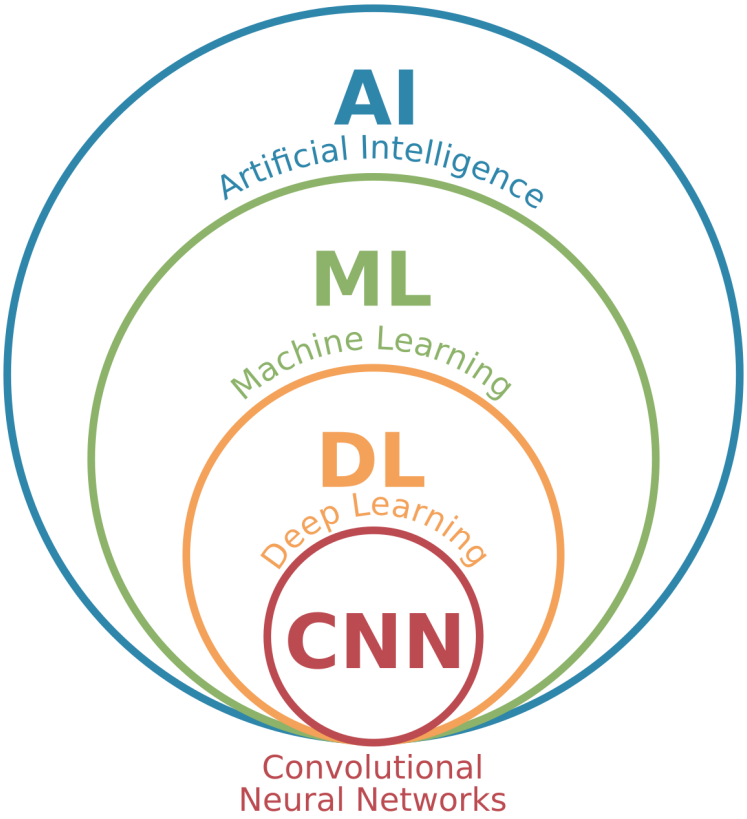



**Figure 2.1:** Relation between the terms artificial intelligence (AI), machine learning (ML), deep learning (DL), and convolutional neural network (CNN).

task. In short, to behave and learn like a self-conscious, self-adapting human being. This type of artificial intelligence is not part of this work, but the more commonly used *narrow* artificial intelligence. Here, a task of an intelligent agent is clearly outlined and the context in which the task has to be solved to some extent predefined. Rich and Knight (1991) refer to them as "*expert systems*" (p. 6). For example, such tasks can be recognising handwritten digits, text from spoken language or extracting offshore wind energy infrastructure from Earth observation data.

## 2.2 Machine Learning

This section focuses on the introduction to machine learning (ML). The primary sources used in this introduction are *Part I: Applied Math and Machine Learning Basics* in *Deep Learning* by Goodfellow et al. (2016), and *Part III: Advanced Topics, Chapter 18: Connectionist Models* in *Artificial Intelligence* by Rich and Knight (1991).

The publication of the CNN AlexNet and its outstanding performance during the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) in 2012 had a far-reaching influence on the deep learning domain, see figure 1.3. During the ILSVRC, AlexNet performed best in an image recognition task of natural red-green-blue (RGB) images, which

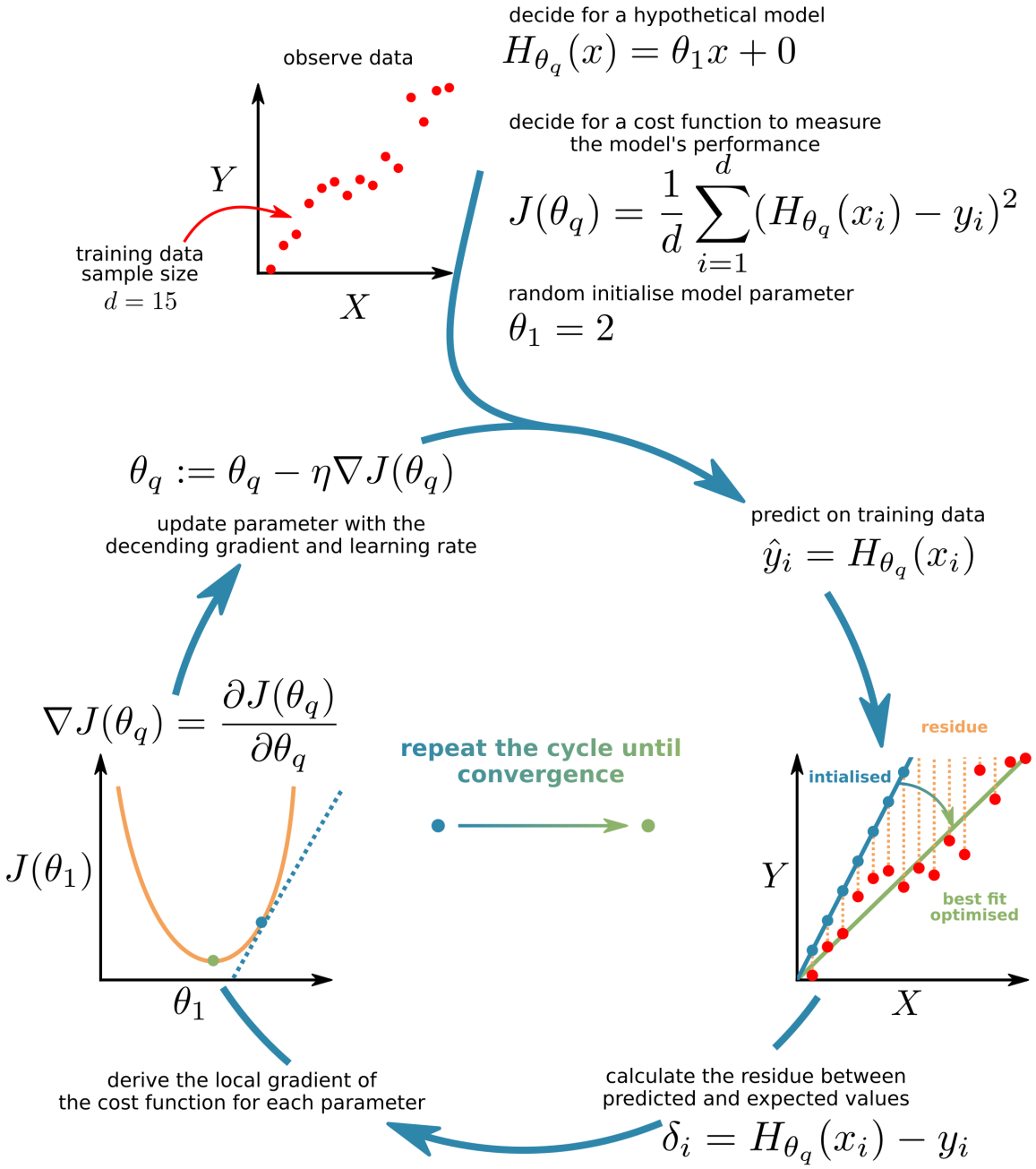



**Figure 2.2:** Schematic overview of parameter adjustment of a linear model by applying machine learning with gradient descent. The process starts with defining conditions, such as the selection of the model and cost function. The parameters are then optimised in a repetitive process in which the difference between model predictions and training data is subsequently minimised by reducing a cost function.

show 1,000 common objects. One key ability of AlexNet is that it uses self-learned features to classify a given image instead of relying on features that were programmed by a human being (Krizhevsky et al., 2012). This aspect led to the comment of Krizhevsky et al. (2017) “learning beats programming” (p. 84), a quote which will be revisited and extended multiple times during this introduction to discuss different aspects of machine and

deep learning. The example of AlexNet emphasises that a *machine* can *learn* by itself and that the learning process sufficiently adjusts a model to fulfil the rather complex task of recognising 1,000 different classes in images.

In the upcoming introduction to machine learning, the focus is put on so-called supervised learning, where training data is used to optimise a model's parameters. To provide an intuitive example of this mechanism, the complexity is reduced, and the learning process is explained on a linear regression problem. Figure 2.2 depicts this learning process as a cycle, which has an initial entry point from which the representation of the training data is learned in an iterative process by optimising the adjustable parameters of the model. To enter this machine learning cycle, first, a hypothetical model $H_{\theta_q}$ has to be defined which has the ability to solve a given task when all of its parameters $\theta_q$, with $q = 1,2,3,\ldots,Q$, are adjusted properly. The given task is to predict the output variable of $Y$ for all possible inputs of variable $X$. The observed real-world data $x_i$ and $y_i$ with $i = 1,2,3,\ldots,d$ depicted in figure 2.2 describes the relation between $X$ and $Y$ and will be used as training data in the upcoming example. They hint at the necessary complexity of a potential model and help decide on a suitable hypothetical model to initialise the machine learning cycle. In this example, the hypothetical model $H_{\theta_q}(x)$ is chosen to be a linear regression model:

$$H_{\theta_q}(x) = \theta_1 x + 0\,, \tag{1}$$

where $\theta_1$ is the slope of the linear model, and the intersection with the y-axis is set to 0 due to the observations of the real-world data and to simplify the problem in this explanation. In order to evaluate how well the model performs the residue $\delta_i$ is calculated. It is the difference between the predicted values $H_{\theta_q}(x_i) = \hat{y}_i$ and the expected results $y_i$, provided by the training data

$$\delta_i = H_{\theta_q}(x_i) - y_i\,. \tag{2}$$

The residues are plugged into a cost function $J$, which expresses in a single value how well the model performs for its current parameter setting $\theta_q$. Thus, the following requirement is to find a suitable cost function $J(\theta_q)$. In this example, it is the sum of the square residues of the estimated values $H_{\theta_q}(x_i)$ and the corresponding training examples $y_i$

$$J(\theta_q) = \frac{1}{d}\sum_{i=1}^{d}(\delta_i)^2 = \frac{1}{d}\sum_{i=1}^{d}\left(H_{\theta_q}(x_i) - y_i\right)^2\,. \tag{3}$$

Finally, all parameters of $H_{\theta_q}$ have to be initialised. In this example, the initialisation of $\theta_1$ is 2, which was randomly selected. With that done, the machine learning cycle can start to *learn* which parameter $\theta_1$ is the optimal choice to result in the best fit of the hypothetical model $H_{\theta_q}$ and the given training data. In a first iteration of the cycle, the initialised model is used to predict for all given $x_i$ values of the training data, $\hat{y}_i = H_{\theta_q}(x_i)$. The predicted values are then compared to the expected values of the training data by using the cost function $J(\theta_q)$. The higher the cost function value, the worse the model fits the training data. This clearly shows that minimising the cost function is associated with better model performance. Due to the chained dependencies of cost function and parameters, it is possible to consider changes of the cost function $J(\theta_1)$ relative to changes of $\theta_1$ by applying the chain rule

$$\frac{\partial J(\theta_q)}{\partial \theta_q} = \frac{\partial J(\theta_q)}{\partial \delta} \frac{\partial \delta}{\partial \theta_q}. \tag{4}$$

To get an intuition how the dependencies influence each other, figure 2.2 shows the cost function $J_{\theta_1}$ relative to $\theta_1$. If $\theta_1$ is further increased relative to its initialised state, this will increase $J(\theta_1)$. However, if $\theta_1$ is decreased, first $J(\theta_1)$ will decrease too, but after passing the global minimum, $J(\theta_1)$ will increase again if $\theta_1$ is decreased further. Thus, to find the global minimum, it is necessary to derive in which direction and how far to move $\theta_1$. The information in which direction to move can be obtained by partially deriving the local gradient of the cost function relative to the currently used parameter setting $\frac{\partial J(\theta_1)}{\partial \theta_1}$. By following the inverse of this gradient, $\theta_1$ always strives for the next minima. This mechanism, already described in the 19th century by Cauchy (1847), is a key feature of machine learning and is called gradient descent $\nabla J(\theta_q)$ (Lecun et al., 1998; Robbins and Monro, 1951; Bottou et al., 2018)

$$\nabla J(\theta_q) = \frac{\partial J(\theta_q)}{\partial \theta_q}. \tag{5}$$

Now that the direction in which the parameter must move is known, the question remains open as to how far the parameter must move. The amplitude itself is to some extent regularised by the gradient since when a local minimum of the cost function is approached, the amplitude of the gradient gets smaller. However, it is common practice to use an additional factor called the learning rate $\eta$. The problem is, if $\eta$ is too large, the global minimum will be missed, and $\theta_1$ ends up in another, but also not an optimal situation. However, if $\eta$ is too small, the learning process would take exceptionally long till it converges to a global or local minimum. There are many solutions how to adjust $\eta$ (Smith, 2017; Loshchilov and

Hutter, 2017), for example, by stepwise decreasing $\eta$ during the learning process. However, in this example, we consider $\eta$ to be fixed at 0.01 for simplicity. The last step in the machine learning cycle is to update the parameters according to their descending gradients and the learning rate

$$\theta_q := \theta_q - \eta \nabla J(\theta_q)\,. \tag{6}$$

With the updated parameters, a new prediction on the training data can be made, new residues calculated, and following new gradients derived to further decrease the cost function. This cycle is repeated until the cost function converges to the global or a local minimum or a predefined maximum number of iterations is reached. The optimisation of the parameter $\theta_1$ will finally end up with $\theta_1$ to be one or close to one for this example.

The key points of how a machine learns by using gradient descent are that the local gradients can be derived when comparing outputs predicted by a randomly initialised model with expected values of the training data. With the gradients, the model's parameters can be adjusted, and thus, the model subsequently learns the underlying representation of the training data. The hypothetical model in this demonstration only had one parameter, which was chosen to provide an intuitive explanation. However, the introduced mechanism allows using models with more parameters, even millions of parameters, which moves to the next section about deep learning models.

## 2.3 Deep Learning

This section focuses on the introduction to fundamental concepts in deep learning (DL). The primary sources for the following section are *Part II: Modern Practical Deep Networks* in *Deep Learning* by Goodfellow et al. (2016), *Part III: Advanced Topics, Chapter 18: Connectionist Models* in *Artificial Intelligence* by Rich and Knight (1991), and *Chapter 2: How the backpropagation algorithm works* in *Neural Networks and Deep Learning* by Nielsen (2015).

The term deep learning is closely related to a specific deep learning model type, the artificial neural network (ANN), also known as multi layer perceptron (MLP). Before the CNN is put into focus, principle ideas of deep learning are explained on the ANN. The building blocks of a deep learning model are often referred to as layers, where each layer holds adjustable parameters. The motivation of this structure is that with each additional layer, the model is able to learn more complex features. Thereby, deep learning models are exceptionally capable of learning to perform a task from data that describes highly complex

representations (LeCun et al., 2015; Goodfellow et al., 2016). The earlier cited quote of Krizhevsky et al. (2017) was only a fragment of a sentence that, when extended, becomes even more descriptive for deep learning: "learning beats programming **for complicated tasks**" (p. 84). By adding layers to a deep learning model, the model becomes *deeper*, hence the name *deep* learning, and with increasing depth, its ability to *learn* more representational features grows and thus its ability to solve complicated tasks. In the following, the structure and fundamental concepts of the ANN architecture are used to explain how a deep learning model is trained to solve a specific task with regard to the earlier introduced concepts of machine learning.

Before this introduction, a short comment to the upcoming notation to support a better understanding. At some points during the upcoming explanation it is more intuitive to refer to a single specific component within the ANN, therefore a component orientated notation is used. Thereby, the superscript $^{(l)}$ denotes the $l^{\text{th}}$ layer in an ANN with $L$ adjustable layers, where $l = 0, 1, 2, \ldots, L$. The subscript $_j$ describes the $j^{\text{th}}$ element of $m$ elements in the current layer $^{(l)}$ where $j = 1, 2, 3 \ldots, m$, and the subscript $_k$ the $k^{\text{th}}$ element of $n$ elements in the previous layer $^{(l-1)}$ where $k = 1, 2, 3, \ldots, n$. At other occasions it is helpful to refer to entire building blocks of the architecture to provide a better overview. In that case, matrix and vector notation is used. For example, a $m \times 1$ column vector with an index of $j = 1, 2, 3, \ldots, m$ that describes all elements of $a_j^{(l)}$ in the $l^{\text{th}}$ layer is denoted as $\mathbf{a}^{(l)}$. Likewise a $m \times n$ sized matrix of elements $w_{jk}$ indexed by $j = 1, 2, 3, \ldots, m$ and $k = 1, 2, 3, \ldots, n$ in the $l^{\text{th}}$ layer is denoted as $\mathbf{W}^{(l)}$, see figure 2.3 for a graphical description of an example model architecture using this notation.

Figure 2.3 shows an example of an ANN architecture and a closeup view of an artificial neuron. A single artificial neuron in an ANN combines two functionalities: It sums up all incoming values and adds a bias to generate an internal sum $z_j^{(l)}$. This sum is passed through an activation function $\sigma$ to produce a single output value $a_j^{(l)}$, the so-called activation of the neuron. In this introduction the sigmoid function with the limits 0 and 1 is used as activation function $\sigma$. Thus, the output of each neuron is a real number between 0 and 1

$$\sigma\left(z_j^{(l)}\right) = \frac{1}{1 + e^{-z_j^{(l)}}}. \tag{7}$$

The functionality of an artificial neuron allows building complex networks of many artificial neurons by connecting their out- and inputs, see figure 2.3, hence the name of this model, artificial neural network. The example ANN architecture in figure 2.3 has four layers of artificial neurons, of which three hold adjustable parameters $l = 0, 1, \ldots, L$ with $L = 3$. The layers are an input layer $\mathbf{a}^{(0)}$ with three input neurons, two hidden layers $\mathbf{a}^{(1)}$ and $\mathbf{a}^{(2)}$

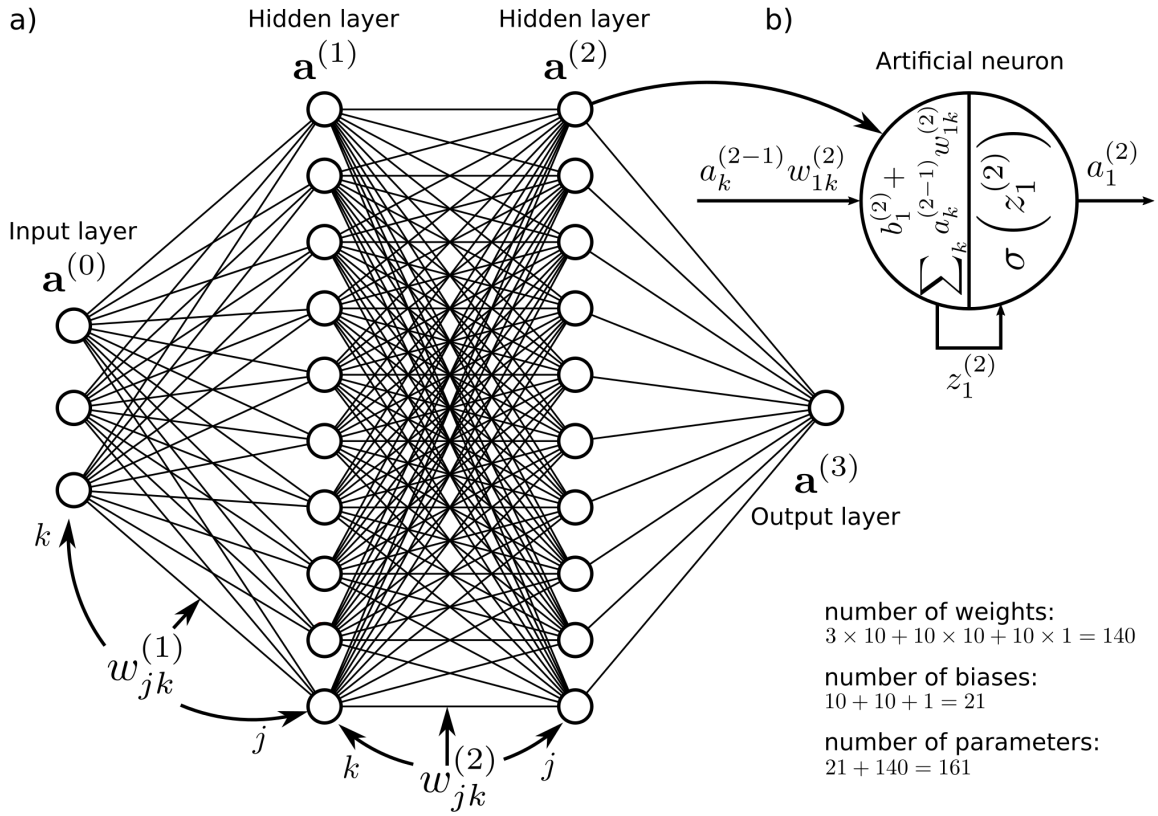



**Figure 2.3:** a) Structure of an artificial neural network (ANN) with three input neurons, two hidden layers with ten neurons each, and an output layer with a single output neuron. The lines connecting the neurons are the weights, the adjustable parameters, which multiply the output of each neuron and send it to the input of the next neuron. b) Functionality of a single neuron, all incoming values and a bias are summed up and used as input in the following activation function.

with ten neurons each, and one output layer $\mathbf{a}^{(3)}$ with a single neuron. For a better intuition, an example task is to be assumed. The example task of the ANN is to tell if the given input values represent a cat. Thereby the inputs are numerical information about the length of the tail $a_1^{(0)}$, the length of the whiskers $a_2^{(0)}$ and the bodyweight $a_3^{(0)}$. These numbers are the values of the three neurons of the input layer $\mathbf{a}^{(0)}$. The single neuron of the output layer $a_1^{(3)}$ represents the predicted probability that the input values belong to a cat. Due to the sigmoid function, the output neuron provides a number between 0 and 1 which can directly be interpreted as this probability. In order to map the input values to the output value, two so-called hidden layers with ten neurons each are used. Each neuron of a layer in an ANN $a_j^{(l)}$ is connected to each neuron of the previous layer $a_k^{(l-1)}$, see figure 2.3. Mathematically, these connections multiply the output of the neurons in the previous layer. The resulting product is used as input for the connected neuron of the current layer. That way, values can propagate from the input to the output layer. Most of the adjustable parameters of an ANN are the factors of these connections between the layers, the so-called weights $w_{jk}^{(l)}$. In addition, each neuron is assigned with a bias $b_j^{(l)}$, which is also added by the summation

function in each neuron to the incoming values. For example, the activation of the output neuron $a_1^{(3)}$, can be expressed as the summation of the incoming values from the previous layer and the assigned bias together with the subsequent activation function

$$a_1^{(3)} = \sigma\left(a_1^{(3-1)} w_{11}^{(3)} + a_2^{(3-1)} w_{12}^{(3)} \ldots + a_n^{(3-1)} w_{1n}^{(3)} + b_1^{(3)}\right). \tag{8}$$

Considering the mechanics of a single neuron, depicted in figure 2.3 and the activation function $\sigma$ in equation (7), equation (8) can be simplified to

$$a_1^{(3)} = \sigma\left(z_1^{(3)}\right). \tag{9}$$

These expressions can be generalised for every activation of each neuron in each layer to represent the forward propagation of the input values through the entire ANN

$$\begin{bmatrix} a_1^{(l)} \\ \vdots \\ a_m^{(l)} \end{bmatrix} = \sigma\left(\begin{bmatrix} a_1^{(l-1)} \\ \vdots \\ a_n^{(l-1)} \end{bmatrix} \begin{bmatrix} w_{11}^{(l)} & \ldots & w_{1n}^{(l)} \\ \vdots & \ddots & \vdots \\ w_{m1}^{(l)} & \ldots & w_{mn}^{(l)} \end{bmatrix} + \begin{bmatrix} b_1^{(l)} \\ \vdots \\ b_m^{(l)} \end{bmatrix}\right), \tag{10}$$

$$\begin{bmatrix} a_1^{(l)} \\ \vdots \\ a_m^{(l)} \end{bmatrix} = \sigma\left(\begin{bmatrix} z_1^{(l)} \\ \vdots \\ z_m^{(l)} \end{bmatrix}\right). \tag{11}$$

To further simplify these equations and to easier express upcoming matrix operations, equations (10) and (11) can be rewritten by using matrix and vector notation

$$\mathbf{a}^{(l)} = \sigma\left(\mathbf{a}^{(l-1)}\mathbf{W}^{(l)} + \mathbf{b}^{(l)}\right), \tag{12}$$

$$\mathbf{a}^{(l)} = \sigma\left(\mathbf{z}^{(l)}\right). \tag{13}$$

With this mechanic, it is possible to propagate input values forwards through the network to the final output layer. This process is also known as *feed forward* and uses all weights and biases of the network. The ANN depicted in figure 2.3 has $3 \times 10 + 10 \times 10 + 10 \times 1 = 140$ weights and $10 + 10 + 1 = 21$ biases which are used during a forward pass. That means that the described model hypothesis in figure 2.3 has 161 parameters which have to be adjusted.

The earlier mentioned example task is used for a more intuitive explanation of the learning process. The three input values of a first training example, which describes a real cat, are passed through the network. These input values propagate through the randomly initialised ANN which produces an exemplary output of 0.3 at the output neuron $a_1^{(3)}$. The interpretation is that the untrained ANN is 30% confident that this training example is a cat. Compared to the true label, which is a 100% certainty that this example describes a cat, the predicted output value indicates changing all 161 parameters in a way that the final result of 0.3 goes up to get closer to the expected value of 1. However, the combination of the 161 gradients of each parameter to change the final output to become 1 is far less intuitive than the earlier discussed linear regression problem with only one adjustable parameter. Nevertheless, the fundamentals of the machine learning cycle to receive the 161 desired gradients eventually remain the same. To follow the introduced machine learning cycle, a cost function $J$, has to be decided:

$$J = \frac{1}{2}\left(\mathbf{a}^{(L)} - \mathbf{y}\right)^2, \tag{14}$$

where $\frac{1}{2}$ is used to make it easier to calculate the derivative later.

As discussed earlier during the machine learning introduction, equation (5) shows that a change of the cost function $\partial J$ is relative to the change of the parameter $\partial\theta_q$. The same concept also applies to the parameters in the ANN. In the example network, a change of the residue provided by the cost function $J$, is relative to a change of the activation of the output neuron $a_j^{(L)}$, which is relative to the change of the sum $z_j^{(L)}$ which combines the bias $b_j^{(L)}$ and the activations of the previous layer $a_k^{(L-1)}$ multiplied by the weights $w_{jk}^{(L)}$. Following that, with the initial residue provided by the cost function, we can partially derive the gradients of the weights (15) and biases (16) in the last layer

$$\frac{\partial J}{\partial w_{jk}^{(L)}} = \frac{\partial J}{\partial a_j^{(L)}} \frac{\partial a_j^{(L)}}{\partial z_j^{(L)}} \frac{\partial z_j^{(L)}}{\partial w_{jk}^{(L)}}, \tag{15}$$

$$\frac{\partial J}{\partial b_j^{(L)}} = \frac{\partial J}{\partial a_j^{(L)}} \frac{\partial a_j^{(L)}}{\partial z_j^{(L)}} \frac{\partial z_j^{(L)}}{\partial b_j^{(L)}}. \tag{16}$$

To derive these gradients, the chained dependencies move backwards through the network and with it are able to indirectly link the error at the output neuron with the previous parameters of the same layer, see figure 2.3 to follow the described path back through the network. This mechanic can be used recursively, but only if the quantity of the error for

each neuron is known. In order to obtain these intermediate errors at each layer $\boldsymbol{\delta}^{(l)}$, the errors at the output neurons $\boldsymbol{\delta}^{(L)}$ have to be propagated backwards through the entire ANN, similar to the input values during the forward pass. This method, called backpropagation, was introduced in the context of ANNs by Rumelhart et al. in 1986 and is the fundamental tool for optimising the parameters of a deep learning model.

As already mentioned, to start the backpropagation process, $\boldsymbol{\delta}^{(L)}$ has to be derived. Therefore, the derivative of the cost function has to be applied element-wise to the derivative of the activation function to propagate the overall error of the network back to the input gate of the artificial neurons of the output layer:

$$\boldsymbol{\delta}^{(L)} = \left(\mathbf{a}^{(L)} - \mathbf{y}\right) \circ \sigma'\left(\mathbf{z}^{(L)}\right), \tag{17}$$

where $\circ$ is the element-wise multiplication, or Habermard product, of the two $1 \times m$ column vectors resulting from the residues $\mathbf{a}^{(L)} - \mathbf{y}$ and the derivative of the activation function $\sigma'\left(\mathbf{z}^{(L)}\right)$. To further derive the error for each layer $\boldsymbol{\delta}^{(l)}$, the initial error $\boldsymbol{\delta}^{(L)}$ can now be propagated back to the previous layer by using the weights of the network. Due to the backwards movement from $^{(l+1)}$ to $^{(l)}$ the transposed version of the weight matrix from equation (10) $\left(\mathbf{W}^{(l+1)}\right)^{\mathrm{T}}$ can be used, to describe the connection of the weights efficiently

$$\boldsymbol{\delta}^{(l)} = \left(\left(\mathbf{W}^{(l+1)}\right)^{\mathrm{T}} \boldsymbol{\delta}^{(l+1)}\right) \circ \sigma'\left(\mathbf{z}^{(l)}\right). \tag{18}$$

With the intermediate error at each layer obtained, it is now possible to derive the partial derivatives of all weights and biases in each layer. Thereby, the gradient of each bias can directly be described by the intermediate error at the corresponding neuron of the current layer $\delta_j^l$

$$\frac{\partial J}{\partial b_j^{(l)}} = \delta_j^l. \tag{19}$$

Furthermore, the gradient of each weight can be described by the activation of a neuron in the previous layer $a_j^{(l-1)}$ and the intermediate error at the connected neuron of the current layer $\delta_j^l$

$$\frac{\partial J}{\partial w_{jk}^{(l)}} = a_k^{(l-1)} \delta_j^l. \tag{20}$$

With this mechanism, the gradients can be derived for each parameter of the model and thus, the introduced machine learning cycle from section 2.2 can be repeated until the 161-dimensional cost function converges to a minimum. Thereby, the ANN is trained on the underlying training data set.

Especially the last expression (20) allows an intuitive discussion of what can happen during the training process. The following example is to be assumed to get a better understanding. If an activation of a neuron $a_k^{(l-1)}$ is high and the connected intermediate error at the target layer $\delta_j^l$ is high and negative, the corresponding gradient $\frac{\partial J}{\partial w_{jk}^{(l)}}$ would have a high negative value for that specific weight. When the opposite of this gradient is used to adjust the associated weight, due to the gradient descent method, the value of the weight would be increased. This increase would eventually reduce the intermediate error and contribute to minimising the entire cost function. On the other hand, if the activation of a neuron $a_k^{(l-1)}$ is small, the training effect of the associated training example would be small too.

This observation leads to the following conclusions. First, if the same example is used repeatedly in the training process, a specific path of activations and weights through the network would be used, and only the associated weights adjusted. However, other weights connected to neurons with smaller activations would be partly excluded from the training process. Thus, when using only the same training example over and over again, a certain path is burnt into the neural network. Or to put it another way, it would overfit strongly. One possibility to avoid this is a technique called dropout. Thereby, at each training step, neurons are randomly excluded from the network (Srivastava et al., 2014). Thus, even the same training example in another training step has to choose a different path through the network, resulting in different activations and gradients on different weights. This way the training effect is better distributed over all parameters in the network. Also, slightly changing the training examples by adding some kind of noise, so-called data augmentation, is helpful to change a limited number of training examples in order to use them repeatedly (Shorten and Khoshgoftaar, 2019). However, the best way to avoid only a very specifically trained or overfitted network would be to maximise the size and variance of the training data set by adding more completely new examples.

In conclusion, a deep learning model has great potential to solve complicated tasks. With many adjustable parameters, the deep learning model can learn complex representations (LeCun et al., 2015). In a supervised learning approach, training examples produce a model output that can be compared to their expected values, and the model parameters are adjusted by applying the gradient descent method to eventually minimise the difference between those two values (Goodfellow et al., 2016). In order to avoid overfitting of the network and to support a high generalisation of the model simultaneously, the employed

training data set must be large, as comprehensive as possible and composed with great care (Krizhevsky et al., 2017). Even when this last conclusion is also true for other machine learning approaches, in deep learning, due to the potentially high number of adjustable parameters, the risk of the model overfitting on an insufficient training data set is high (LeCun et al., 2015). Hence, the generation of the training data set is critical, which leads back to the next extension of the quote of Krizhevsky et al. (2017) that points out the importance of training data in deep learning: "**With [...] enough data,** learning beats programming for complicated tasks" (p. 84).

## 2.4 Convolutional Neural Networks

Finally, this last introductory section focuses on the convolutional neural network (CNN). The primary source for this introduction is *Part II: Modern Practical Deep Networks, Chapter 9: Convolutional Networks* in *Deep Learning* by Goodfellow et al. (2016).

In the introduction to deep learning, the example task was to detect if three input values belong to a cat or not. To increase the complexity, the example task is now to recognize if a RGB image shows a cat or not. This means that instead of passing three values through the network, all pixel values of an RGB image are used as input.

The CNN AlexNet, which was built for a similar image recognition task, was introduced with eight layers of adjustable weights (Krizhevsky et al., 2012). Given the assumption that the input image is an RGB image with the dimensions $256 \times 256 \times 3$, let us see what happens if an ANN with eight layers of adjustable parameters is used to perform this task. Figure 2.4 shows the structure of the ANN with the input layer $\mathbf{a}^{(0)}$ and eight layers with adjustable parameters, including the output layer. To feed images into the network, the images with a dimension of $256 \times 256 \times 3$ are flattened to an input layer with 196,608 neurons. The number of neurons of the subsequent layers starts at 1,024 and is stepwise reduced by the factor of 0.5 until 16 neurons in the 7th layer and a single output neuron in the output layer. The resulting ANN would have 202,024,521 parameters, see figure 2.4, which, considering the task of only classifying one class, is very high. For comparison, the CNN AlexNet has about 61 million parameters for classifying 1,000 classes (Krizhevsky et al., 2012). Since in a fully connected ANN each neuron has to be connected with each neuron of the next layer, 201,327,616 or 99.65% of all adjustable parameters are within the first layer of the example ANN, see figure 2.4. Considering the initial motivation of deep learning models to extract more complex features at increasing depth, a concentration of nearly all parameters in the first layer of the network is a first hint that the fully connected ANN model structure is not appropriate for image recognition.

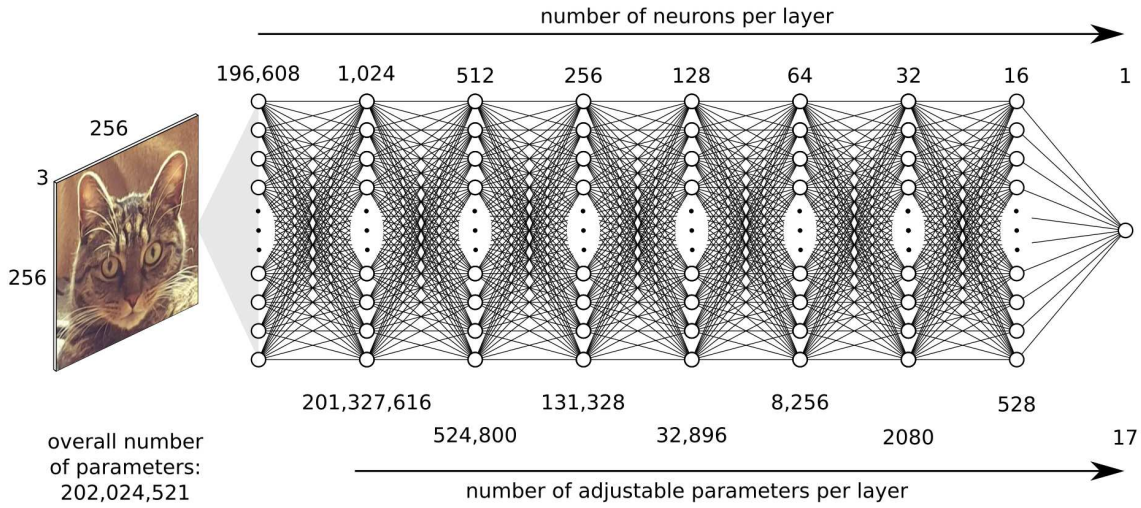



**Figure 2.4:** Structure and parameter distribution of an 8 layer fully connected artificial neural network which uses a $256 \times 256 \times 3$ input image to predict an output.

This example makes it clear that the multidimensional array structure of the input data, the image, leads to the mentioned problems. Instead of 3 rather meaningful values which describe characteristic features of a cat, like the length of the tail and whiskers and the body weight, the 196,608 pixel values are not very descriptive. Only in a specific structure, the pixel values show something that a human recognises as a cat. By flattening the pixel values in order to use them as inputs for the ANN, this fundamental structure is getting lost even before underlying features can be extracted from them.

A single channel of an image is a 2D discrete function of pixel values. The internal gridded structure is an inherent characteristic of this function (Goodfellow et al., 2016). Due to the spatial proximity of pixels and their gradients, shapes like edges become visible, which together form high-level features like ears, eyes and whiskers, which eventually represent a cat. Recognising simple features and combining them into more semantically meaningful features was found to be typical for the visual cortex of mammals (Hubel and Wiesel, 1962; Felleman and Van Essen, 1991). However, if the characteristic 2D structure is lost because the image was flattened to ingest it into the input layer of an ANN, most of the information, especially the spatial features an image can represent, is lost. Thus an image stores not only pixel values but also their local connectivity by preserving a specific structure. It follows that a model which has the task of analysing images should be able to ingest images in their original structure and investigate the spatial arrangement of their values in order to mimic the behaviour of the visual cortex (Cadieu et al., 2014; Fukushima and Miyake, 1982).

Edge filters are good examples of focal operations that can extract strong features from images by exploring pixel values and their spatial context. In order to detect vertical or horizontal edges in a single band greyscale image $\mathbf{I}$ a filter $\mathbf{W}$ (the kernel), has to be applied

pixel-wise on the image. By moving the kernel over the entire input image, pixels in regions with a strong vertical or horizontal difference return higher values in the output image **A**. The described process is called a two-dimensional discrete convolution (Goodfellow et al., 2016). The convolution, denoted with $*$, of a $s \times t$ sized padded input image **I**, indexed with $u = 1,\ldots,s$ and $v = 1,\ldots,t$, and a $m \times n$ sized kernel **W**, with a centred index $j = -\lfloor \frac{m}{2} \rfloor,\ldots,\lfloor \frac{m}{2} \rfloor$ and $k = -\lfloor \frac{n}{2} \rfloor,\ldots,\lfloor \frac{n}{2} \rfloor$, can be expressed as

$$\mathbf{A}(u,v) = (\mathbf{I} * \mathbf{W})(u,v) = \sum_j \sum_k \mathbf{I}(u-j, v-k)\mathbf{W}(j,k)\,. \tag{21}$$

In the case of CNNs, the convolution in most deep learning frameworks is usually implemented as cross-correlation, and the kernel function is not flipped (Goodfellow et al., 2016)

$$\mathbf{A}(u,v) = (\mathbf{I} * \mathbf{W})(u,v) = \sum_j \sum_k \mathbf{I}(u+j, v+k)\mathbf{W}(j,k)\,. \tag{22}$$

In the following, the cross-correlation variant as introduced in equation (22) is used and simplified as follows

$$\mathbf{A} = \mathbf{W} * \mathbf{I}\,. \tag{23}$$

Figure 2.5 demonstrates the convolution of a vertical and horizontal filter. Where the vertical filter is able to extract the eyes and ears of the cat, the horizontal filter is able to extract the whiskers. Even when these filtered images are the first promising results to extract meaningful features from the input image, it would be necessary to further increase the semantic meaning of those extracted edges by combining these features in subsequent steps. That would mean recursively applying additional filters that operate on intermediate results to find even stronger features until a set of features is programmed to recognise a cat securely. The stack of kernel functions required for this is far less intuitive compared to the kernel function in figure 2.5 to extract horizontal or vertical features. Adding another layer of functions to intermediate results is probably beyond the capabilities of most humans, let alone several successive combinations of filters. This is where the core idea of machine and deep learning comes into action, and the next extension to the quote of Krizhevsky et al. (2017) applies to this problem: “learning beats programming for complicated tasks **that require the integration of many different, noisy cues**” (p. 84).

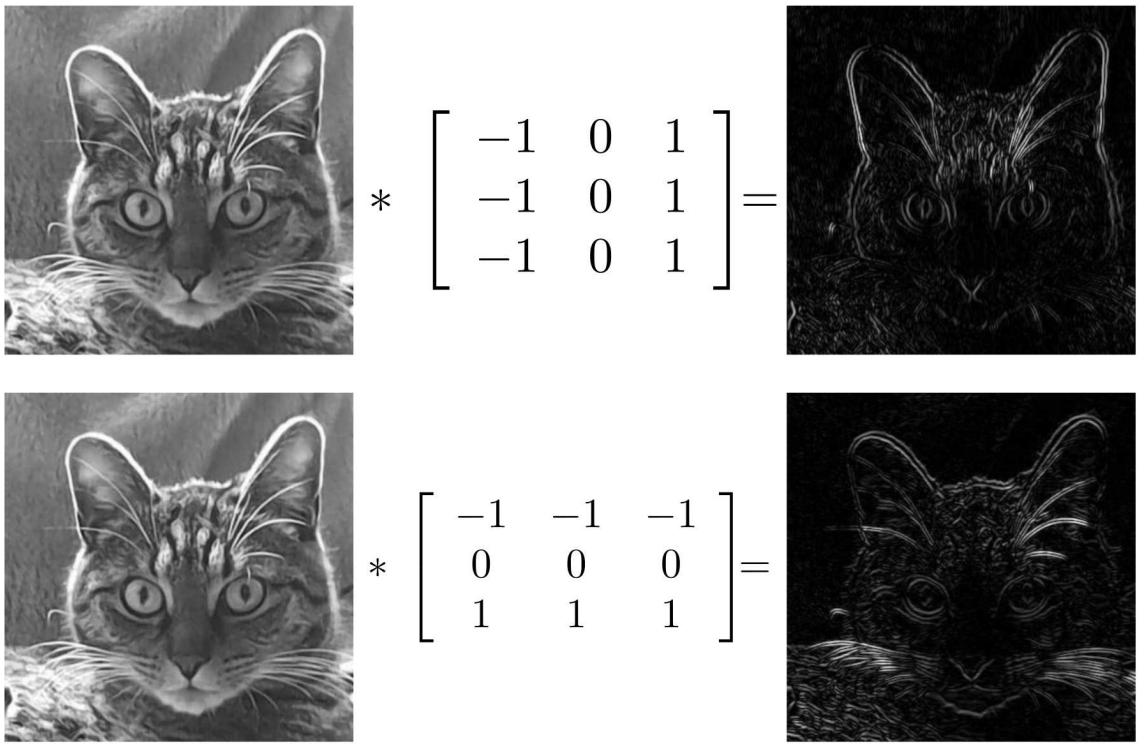



**Figure 2.5:** Convolution of a vertical and horizontal $3 \times 3$ kernel on an input image. The resulting activation map highlights vertical and horizontal features in the input image of the cat.

From a deep learning perspective, the kernel function $\mathbf{W}$ can be interpreted as weights or the adjustable parameters, and the resulting output image $\mathbf{A}$ is something related to the activation of a neuron. Since the output is not a single neuron but an entire matrix of activations, it can be understood as an activation map or feature map (Goodfellow et al., 2016). With that idea in mind, it makes sense to think about the kernel function, not as something predefined like an edge filter, but something that can be learned in the same way as the weights in an ANN. And this is the core principle of a convolutional neural network. Instead of using predefined kernel functions to extract specific features from input data, self-learned kernel functions are used in a CNN to subsequently extract highly complex features which build on each other (LeCun et al., 2015).

The forward propagation of an input image or activation map of the previous layer $\mathbf{A}^{(l-1)}$ to the current layer $\mathbf{A}^{(l)}$ in a CNN is descirbed by equation (24) which follows the overall idea of the earlier introduced forward propagation within an ANN, see equation (13)

$$\mathbf{A}^{(l)} = \sigma\left(\mathbf{W}^{(l)} * \mathbf{A}^{(l-1)} + \mathbf{b}^{(l)}\right). \tag{24}$$

This expression shows that likewise to an ANN, in a CNN a bias $\mathbf{b}^{(l)}$ is added to the results of the convolution operation, and an activation function $\sigma$ is applied to each feature map.

Figure 2.6 shows a typical CNN structure for image recognition. An input image is fed into the so-called convolutional backbone, in which features are getting extracted. While going deeper into the network, the extracted features become more representative (Krizhevsky et al., 2012). For a better intuition, one can imagine that features in the early convolutional layers represent high-level information like edges. In contrast, features in the last layer are much more semantically meaningful. Here an activation in a feature map can represent an entire ear of the cat. This is possible due to the feature maps' increasing depth and the decreasing size. The increasing depth allows for storing subsequently more features. Nevertheless, only increasing the depth would lead to extremely large feature map tensors, which might be too large to fit in the processing hardware. Thus, max-pooling layers are connected after convolutional layers. Max pooling layers are kernels with a dimension of $2 \times 2$, which reduce the corresponding values to their maximum value (Ranzato et al., 2007). They commonly move with a stride of 2 over the image, thus reducing the width and height dimension by the factor of 0.5, which has two effects, see figure 2.6a).

The first advantage of the dimension reduction by max-pooling is that the processed activation maps are half the size in dimension and thus only occupy a quarter of memory compared to their original size. This reduction in width and height allows increasing the number of feature maps in depth without exceeding the memory of the underlying hardware. Due to max pooling, less important activations are dropped to add more layers with different but more important and representational information. The second advantage is that due to the subsequent reduction of width and height of the feature maps, the spatial location of activations becomes less important. In the early layers, activations correspond directly with the original structure of the input image, whereas in the last feature maps, the spatial context of activations is less important. Instead, a specific activation in the last layers can indicate that *somewhere* in the image are whiskers. This location invariance allows highly complex features to appear in the last feature map without the necessity that they are at the same location in the image (Scherer et al., 2010; LeCun et al., 2015; Goodfellow et al., 2016).

After the convolutional layers have extracted meaningful features which describe important characteristics of the input image, such features are fed into an ANN to perform the final classification. This fully connected ANN building block, after the CNN backbone, is the so-called classifier head, see figure 2.6c). In other words, the convolutional backbone has the task to extract features from the raw input data in order to present those meaningful but selected features to an ANN. Thus the ANN is able to predict more meaningful input values instead of many less meaningful raw pixel values (Krizhevsky et al., 2012). The translation invariance (LeCun et al., 2015) in the last layers of the convolutional backbone also contributes to the fact that when those features are passed to the ANN, the loss

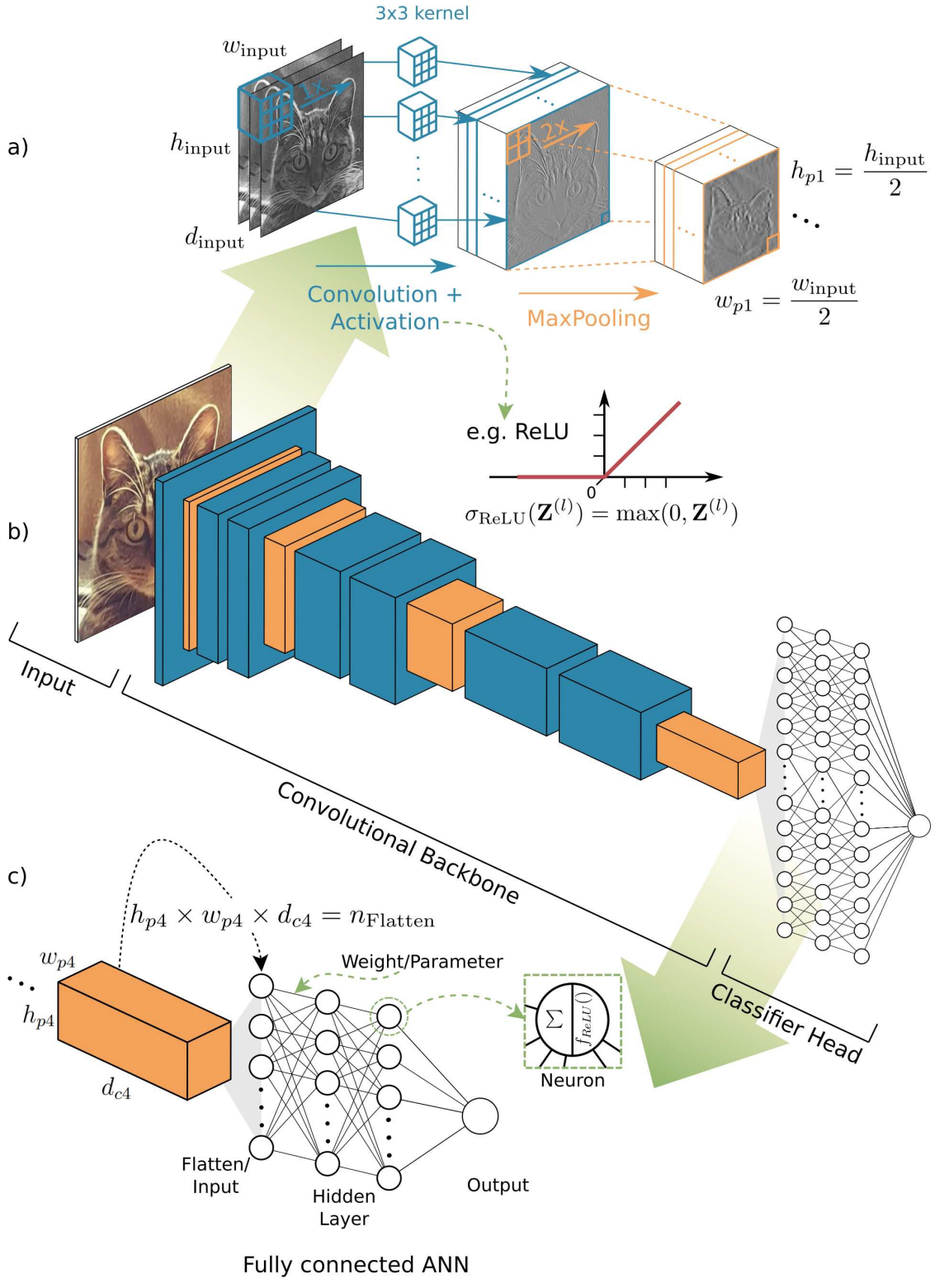



**Figure 2.6:** Schematic overview of a convolutional neural network (CNN) for image recognition. a) Convolution with $3 \times 3$ sized kernels, a rectified linear unit (ReLU) activation function, and subsequent max pooling with a $2 \times 2$ sized kernel. b) Main building blocks of a CNN, the input, the convolutional backbone, which extracts meaningful features from the input, and the classifier head, here an artificial neural network (ANN) which uses the extracted features for prediction. c) Transition from the convolutional backbone to the classifier head. Changed after Hoeser and Kuenzer (2020, p. 6).

of the internal structure by flattening the feature tensors is not that relevant any more. In section 2.3 three *predefined* characteristic input values were used in an ANN, the length of the tail, whiskers and bodyweight of a cat. Now, in a CNN, *self-learned*, characteristic features extracted from an image of a cat are passed to a connected ANN to perform the classification.

The importance of the convolutional backbone becomes even clearer when looking at the number and distribution of parameters, which was found to be a problem when ingesting a raw image directly into an ANN at the beginning of this section. A CNN, depicted in figure 2.7 with eight trainable layers, five convolutional layers with max-pooling in the convolutional backbone and three fully connected layers in the ANN classifier head is to be considered. Each kernel of each convolutional layer has the dimension $3 \times 3$, and the input image has the dimension of $256 \times 256 \times 3$. The first convolutional layer, including max-pooling, results in a tensor of feature maps with the dimension $128 \times 128 \times 16$. The corresponding number of parameters in this layer would be kernelheight $\times$ kernelwidth $\times$ inputdepth $\times$ outputdepth $+$ bias which is $3 \times 3 \times 3 \times 16 + 16 = 448$. This number of parameters is much lower compared to the ANN example with over 200 million parameters in the first layer. It becomes clear that the width and height dimensions of the input image do not influence the parameter count in a CNN. Instead, the number of parameters is dependent on the kernel dimension and depth of the input and output feature maps. The latter is promising since, with increasing depth of the feature maps, the majority of parameters are located in the last layers of the convolutional backbone, where the most representational features are extracted, see figure 2.7.

By doubling the depth of each subsequent feature map, the last convolutional layer would have $3 \times 3 \times 128 \times 256 + 256 = 295,168$ adjustable parameters. This tensor with a dimension of $8 \times 8 \times 256$ has to be flattened in order to feed it into the subsequent ANN. Assuming a first trainable layer with 1,024 neurons in the ANN, the first layer in the classifier head would have $8 \times 8 \times 256 \times 1,024 + 1,024 = 16,778,240$ which corresponds to 94.81% of all parameters of the network. Even when there is a similar accumulation of parameters in the first layer of the ANN-classifier head as in the earlier discussed ANN example, an important difference between the two models exists. The difference is that in the CNN, the majority of parameters process highly representational features instead of raw pixel values since they are located deep in the network architecture, see the comparison made in figure 2.7. Overall the CNN has about 17.7 million parameters compared to 202 million parameters of the ANN example. This means that using a CNN for image analysis is beneficial in two major ways. First, the underlying grid-like structure of the input data is preserved and taken into account by the model architecture in order to extract more

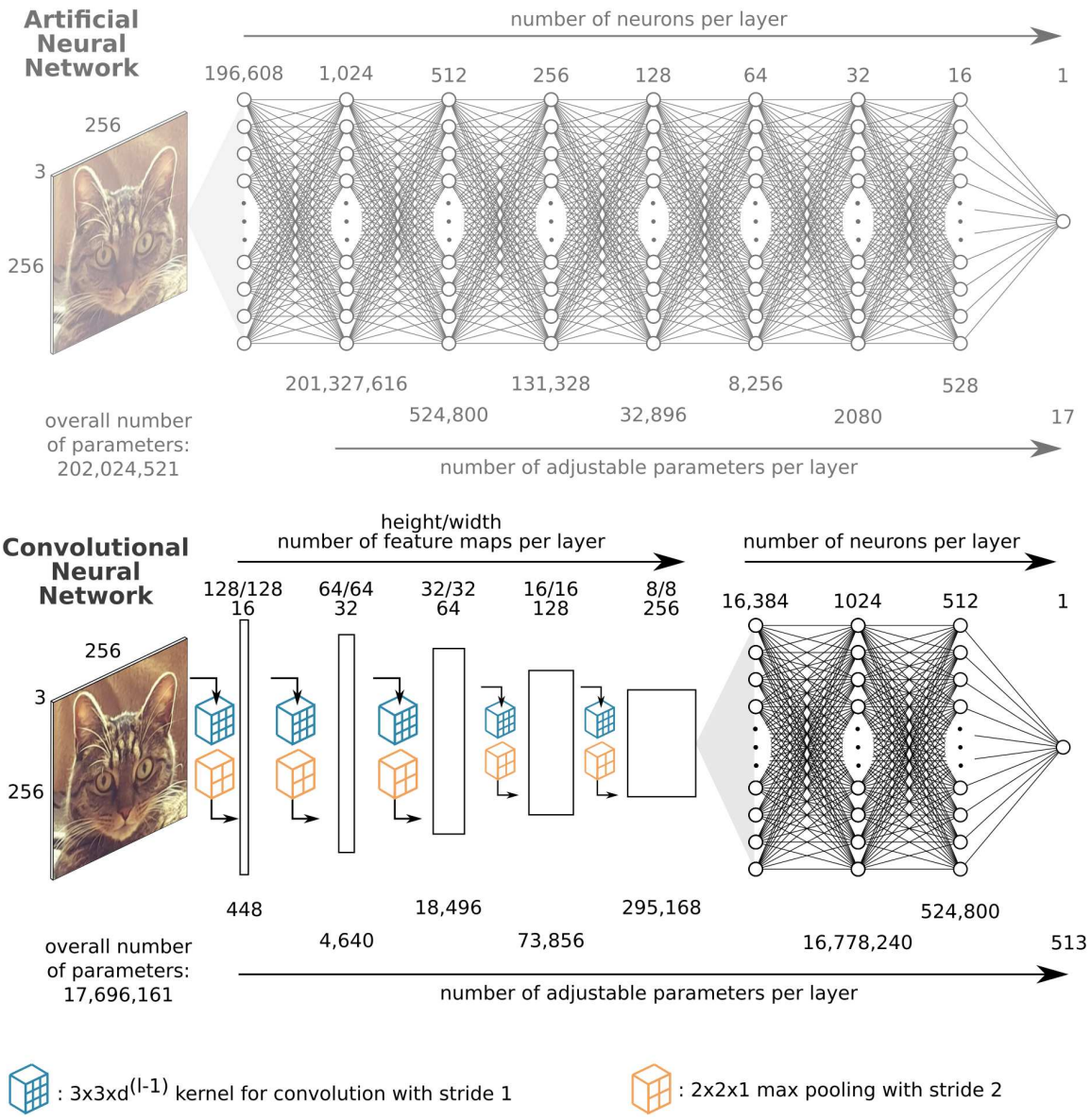



**Figure 2.7:** Comparison of the structure and parameter distribution of an eight layer depth fully connected artificial neural network (ANN) and a convolutional neural network (CNN). Both use a $256 \times 256 \times 3$ input image to predict an output, the CNN has far fewer parameters, and most of the parameters are in the deep layers of the convolutional backbone.

meaningful features. Second, the amount of parameters is reduced dramatically compared to a standard fully connected ANN. Following that, a CNN is computationally the much more efficient deep learning model type for the given example of image recognition.

The insights into CNNs demonstrate that for solving more complex tasks, like image recognition, the number of adjustable parameters in deep learning models can go up to millions or even hundreds of millions. This increases the importance of comprehensive training sets to avoid overfitting and to ensure successful training of the models, as already stated earlier. It also highlights another aspect. In order to optimise such models, extensive computational capacities are necessary (Krizhevsky et al., 2017). With the increasing capabilities

of graphics processing units (GPUs) in the early 21st century, a hardware component became popular, which significantly accelerated the training of neural networks due to its specialisation in the calculation of matrix operations (Voulodimos et al., 2018; Shrestha and Mahmood, 2019). As described in the previous introduction to deep learning, most of the expressions of deep learning models and their optimisation use matrix operations. Since the theory of backpropagation for optimising ANNs was already introduced in 1986, one crucial factor that deep learning has not become that popular then was that there was not enough processing power to handle deep architectures with many parameters technically, and with it the large matrices which express the model. Krizhevsky et al. (2017) describe this very clearly in their publication of AlexNet: "Twenty years later [respective to 1980s], we know what went wrong: for deep neural networks to shine, they needed far more labelled data and hugely more computation" (p. 84). With the recent advances in GPU processing capabilities and the availability of data in the era of big data, today, the theory behind deep learning is technically feasible. Finally, the complete quote of Krizhevsky et al. (2017), which has been extended multiple times during this introduction, summarises the fundamental ideas of deep learning and CNNs: "**With enough computation** and enough data, learning beats programming for complicated tasks that require the integration of many different, noisy cues" (p. 84).

To conclude the introduction to CNNs, it is important to emphasise that CNNs are modular architectures consisting of building blocks as shown in figure 2.6. The most important and characteristic building block is the convolutional backbone or feature extractor. However, the so-called head of the architecture can be replaced entirely, depending on the task to be fulfilled. Figure, 2.8 shows different examples of tasks that are commonly performed by using a CNN. Besides image recognition, in image segmentation, a CNN would classify every single pixel of the input image, which in Earth observation is commonly known as a classification map. In object detection, single entities of objects are detected and described by a bounding box. Whereas in instance segmentation, object detection is applied first with subsequent image segmentation to provide a mask for the detected instance of an object. The capability of CNNs to perform different tasks is important to mention since it makes clear that CNNs are highly adaptable models, and they can be applied to problems that commonly occur in the Earth observation domain. This adaptability of deep learning models explains why they became so popular across many research domains and in practice (Bengio, 2013).

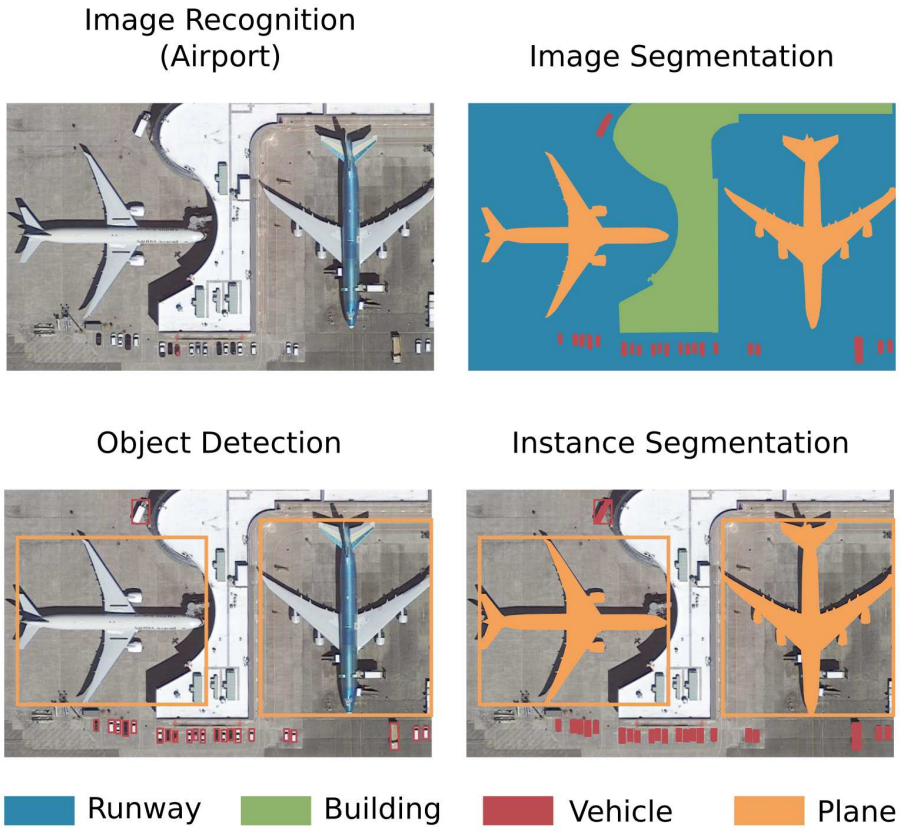



**Figure 2.8:** Examples of four typical tasks performed by convolutional neural networks (CNNs), image recognition, image segmentation, object detection and instance segmentation. Changed after Hoeser and Kuenzer (2020, p. 8).

## 2.5 CNNs in Earth Observation

The previously introduced concepts were mainly developed in machine learning, computer science, and computer vision, which are the main drivers of the latest artificial intelligence and deep learning revival. In this section, the focus is put on how the Earth observation domain adapted to the methodological developments of these domains and how deep learning subsequently became a vital tool for remote sensing image analysis (Zhang et al., 2016b; Zhu et al., 2017; Hoeser and Kuenzer, 2020). As deep learning started to emerge for image analysis and open questions about how deep learning can be used in Earth observation were identified and discussed, the number of use cases within the Earth observation community grew (Ball et al., 2017; Hoeser and Kuenzer, 2020; Hoeser et al., 2020). Eventually, the potential of deep learning was strongly promoted to help understand recent trends in data-driven Earth system science (Reichstein et al., 2019). Due to the earlier discussed characteristics, the CNN became the most widely used deep learning model type in remote sensing image analysis (Ma et al., 2019b). In order to get an overview of what happened since the introduction of AlexNet in 2012 in the Earth observation domain regarding deep

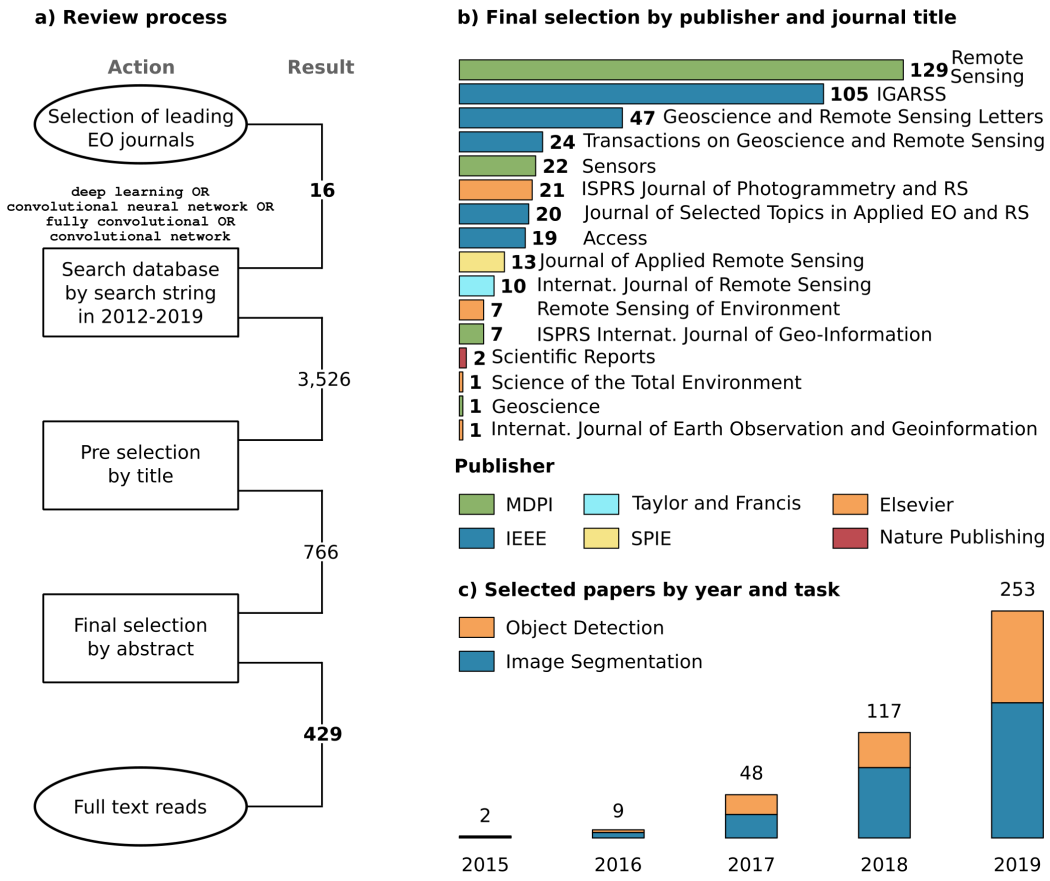



**Figure 2.9:** a) Selection process of the 429 out of 3,526 conducted publications from 16 Earth observation journals. b) Overview of the 16 Journals and numbers of publications. c) Number of publications per year and grouped by object detection and image segmentation. Changed after Hoeser et al. (2020, p. 4).

learning with CNNs, in the first year of this dissertation project, a review that includes publications until the end of 2019 was carried out, see Hoeser and Kuenzer (2020) and Hoeser et al. (2020) for the entire survey.

Figure 2.9 describes the review process. A total of 429 publications from 16 journals were reviewed. Potential publications were selected with an initial query of the publishers' databases. The following query pattern was used: `deep learning OR convolutional neural network OR fully convolutional OR convolutional network`. The final selection of publications to review was made under the condition that the investigated data is remote sensing data collected by sensors attached to drones, aircraft or satellites. Furthermore, the application domain is related to observations of the land surface, and a CNN is used as a methodological approach, either for object detection or image segmentation.

### 2.5.1 Application Domains

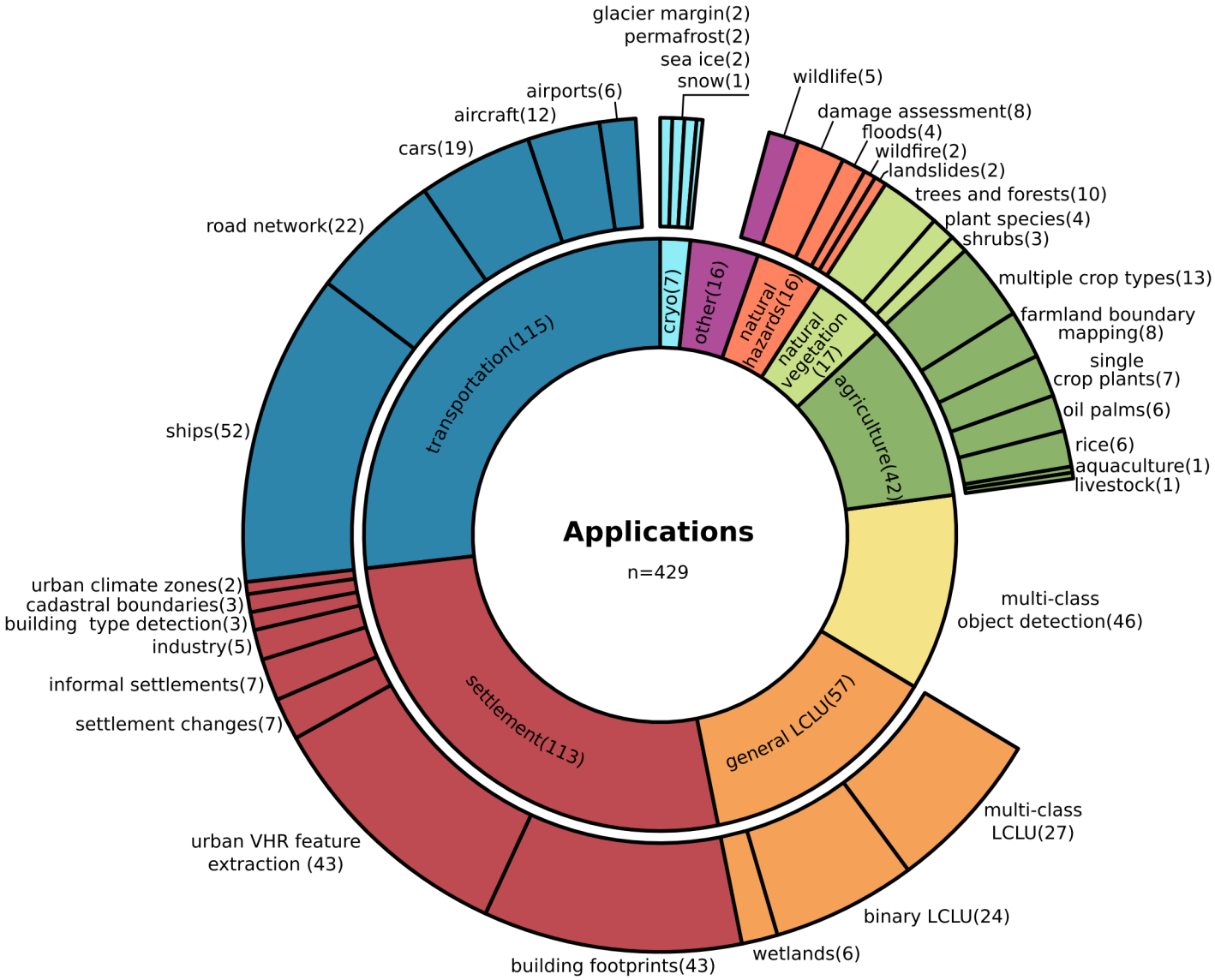



**Figure 2.10:** Grouped application categories of the 429 reviewed publications. Especially the subcategories of transportation and settlement as well as the entire multi-class object detection category, point to the extraction of small scale features which demand high-resolution imagery. Changed after Hoeser et al. (2020, p. 11).

The application domains of all 429 publications were grouped into eight categories with an additional category for miscellaneous applications as depicted in figure 2.10. Typical applications in the Earth observation domain were identified, like land cover land use (LCLU) classification (Henry et al., 2019; Ma et al., 2019a; Zhang et al., 2019a; Hu et al., 2018; Yang et al., 2019; Stoian et al., 2019), studies with a focus on agriculture like classifying crop types (Zhong et al., 2019; Ji et al., 2018; Ienco et al., 2019; Rußwurm and Körner, 2018; Teimouri et al., 2019; Li et al., 2019), investigations of large ecosystems like forests

(Sylvain et al., 2019; Hamdi et al., 2019; Safonova et al., 2019), and such focussing on glacier margins and ice shelf (Mohajerani et al., 2019; Baumhoer et al., 2019). However, when looking more closely into the categories, for example, into the investigations of urban areas, the focus is often put on extracting small scale information like single building footprints (Yang et al., 2018; Wen et al., 2019; Bittner et al., 2018; Liu et al., 2019a; Ye et al., 2019; Shrestha and Vanneschi, 2018; Ji et al., 2019) or specific classes of buildings like industrial complexes or power plants (Zhang et al., 2018; Zhang and Deng, 2019). Other examples of studies that investigate single entities on a spatially small scale are such that study numerous trees instead of large scale forest cover (Fromm et al., 2019; Weinstein et al., 2019; Dong et al., 2019; Santos et al., 2019). The focus on small scale entities becomes even more evident in the largest application domain, transportation. Here, the detection of single vehicles, like cars (Tang et al., 2017; Li et al., 2019; Deng et al., 2017; Koga et al., 2018), ships (You et al., 2019; You et al., 2019; Zhang et al., 2019c; Fan et al., 2019; He et al., 2018; Gao et al., 2019; Li et al., 2018b; Zhang et al., 2019; Voinov et al., 2019) and aircraft (Chen et al., 2018b; Wang et al., 2019c; Zhao et al., 2019) is the major focus in the transportation category. The category of multi-class object detection further increases the number of studies focusing on extracting small scale information from Earth observation data. The target classes in this domain are also vehicles and buildings but also more specific classes, like harbours, airports, roundabouts, industrial facilities like gas storages, or sports facilities like swimming pools, baseball diamonds, tennis courts or soccer pitches (Cheng et al., 2014; Xia et al., 2018; Cheng et al., 2016; Wu et al., 2018; Li et al., 2018a; Wang et al., 2019a; Su et al., 2019). This summary of the application domains shows that in Earth observation CNNs are used for established tasks like LCLU classifications but more importantly, CNNs seem to enable a sophisticated yet flexible object detection and image or instance segmentation, especially on high-resolution remote sensing data.

### 2.5.2 Remote Sensing Data and Training Data Sets

The detection of single cars, ships, specific industrial or sporting facilities, as well as the accurate segmentation of building footprints, need image data that supplies rich spatial features of the targets (Wang et al., 2019a; Xia et al., 2018). Figure 2.11 clearly shows that very high spatial resolution RGB imagery is the most commonly investigated under the 429 reviewed publications. Even among spaceborne missions, such which provide a spatial resolution of $\leq 1$ m are used more often than the established products of the Sentinel-2 or Landsat missions. In addition, RGB imagery acquired from Google Earth is often investigated since it provides high-resolution images from various spaceborne sensors. The strong focus on high-resolution RGB images in Earth observation during this early phase of implementing deep learning techniques is inherited by the computer vision domain. In

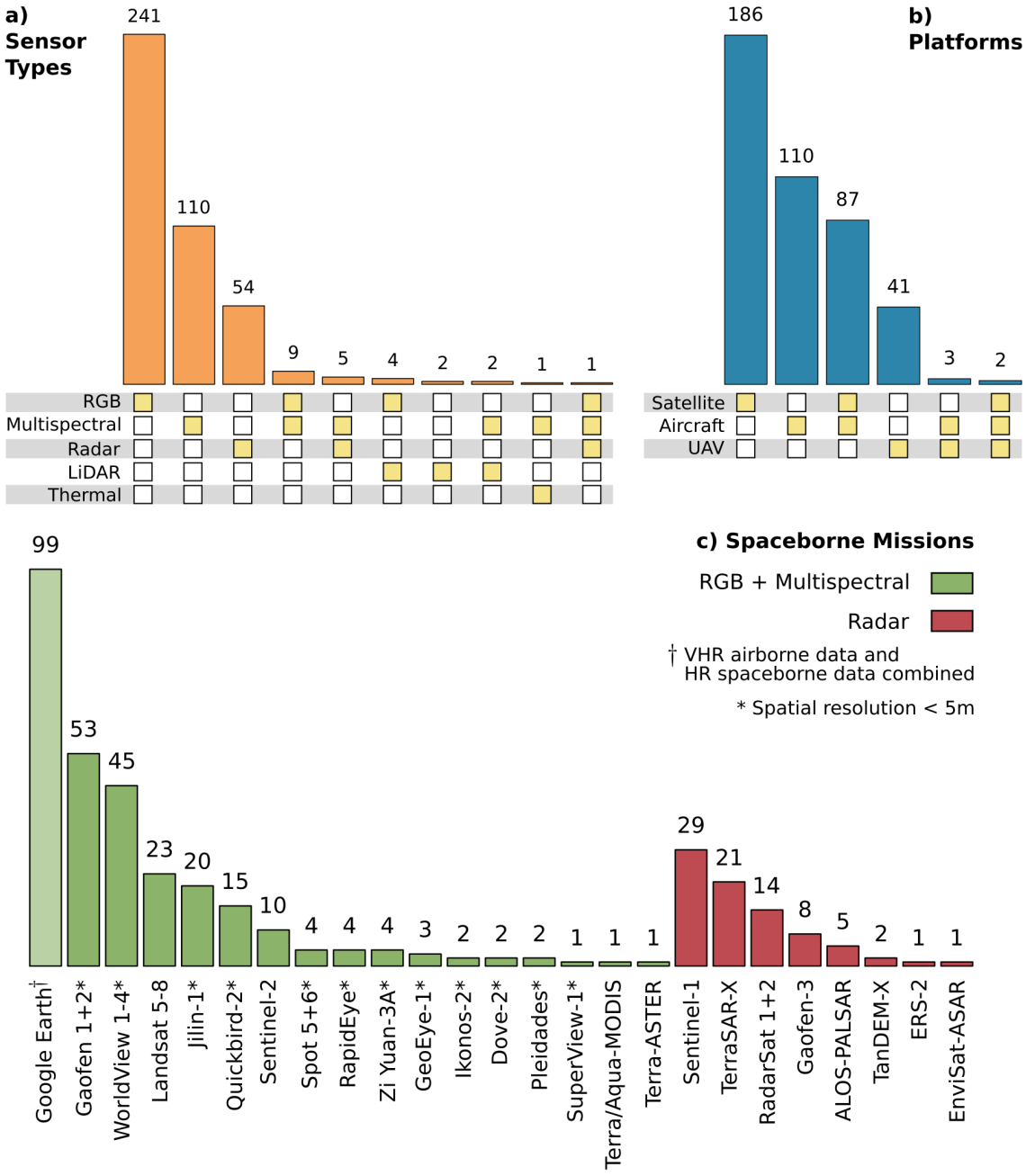



**Figure 2.11:** a) Number of studies grouped by sensor type, b) and platforms. c) Number of studies grouped by spaceborne missions, differentiated in optical and radar sensors. Google Earth, even when not a single specific mission, was also added since it is widely used among the reviewed publications as a source for high-resolution RGB images. Changed after Hoeser et al. (2020, pp. 7-8).

computer vision and the development of CNNs, influential competitions and related data sets have a focus on RGB images with a rich representation of spatial features. For example, the earlier cited ILSVRC challenge is based on the RGB ImageNet data set with more than 14 million images, which is the primary driver in the development of convolutional backbones and image recognition (Russakovsky et al., 2015). Other important data sets which are used for development in the computer vision domain are the Cityscape (Cordts

et al., 2015, 2016) and Pattern Analysis, Statistical modelling and ComputAtional Learning-Visual Object Classes (PASCAL-VOC) data set (Everingham et al., 2010, 2015) for image segmentation, and the Microsoft-common objects in context (MS-COCO) data set (Lin et al., 2014) for object detection. All of them are based on RGB data with a rich spatial feature representation. Thus, the methodological transition from the computer vision to the Earth observation domain started with the investigation of high-resolution RGB images which also provide spatial features for clearly outlined objects and surfaces since CNNs coming from the computer vision domain could directly be employed in this early stage of implementing CNNs in Earth observation.

The observations above partly explain why in Earth observation, high-resolution RGB imagery is widely used. However, it does not explain why certain application categories, such as transportation or settlement, are more represented than others. Therefore, another aspect is important to understand: Data alone is only half the job in deep learning. In order to train a neural network, it needs training examples, which means accurately labelled data sets. The necessity of training data has already been stressed during the introduction to basic concepts of deep learning and CNNs. Large annotated data sets are the fuel that drives the deep learning engine. Without them, supervised deep learning can not be applied or even investigated (Krizhevsky et al., 2017; Long et al., 2021; Xia et al., 2018). However, the annotation of such large data sets is a labour expensive task (Long et al., 2021). In Earth observation, even in times of big data, the data availability, especially of very high-resolution data can also become a major problem. The established open access archives do not supply an extensive coverage of this kind of imagery. Thus, studies of deep learning with CNNs have to rely on existing data sets, or they need to build their own data sets by spending many resources. Table 2.1 summarises data sets that have been used more than once in the reviewed articles. Within the 429 reviewed publications, these 25 data sets were used 233 times. When looking at the application domains of these data sets, it becomes clear why settlement, transportation and multi-class object detection are among the most investigated application domains: For these application domains, the most training data sets are available, see table 2.1. From this, it can be deduced that in the first years in which CNNs were used in Earth observation, the focus was primarily on a methodological transfer, and therefore, existing training data sets were reused. Only gradually did studies focus on non-technical research questions with a stronger geoscientific focus which demanded their own training data sets (Hoeser et al., 2020).

**Table 2.1:** Overview of training data sets used two times or more in the 429 reviewed publications. Most data sets are from the settlement, transportation and multi-class object detection categories. The column *Task* describes if the data set can be employed for image segmentation (IS) or object detection (OD). Data sets marked with a † are from radar sensors, all other data sets contain images which are captured by optical sensors. Changed after Hoeser et al. (2020, p. 10).

| Data set | Year | Task | Application | Times used |
|---|---|---|---|---|
| ISPRS Vaihingen (ISPRS, 2016) | 2016 | IS | settlement | 47 |
| ISPRS Potsdam (ISPRS, 2016) | 2016 | IS | settlement | 35 |
| NWPU VHR 10 (Cheng et al., 2014) | 2014 | OD | multi-class OD | 33 |
| DOTA (Data set for OD in Aerial Images) (Xia et al., 2018) | 2018 | OD | multi-class OD | 17 |
| Massachusetts Building (Mnih, 2013) | 2013 | IS | settlement | 11 |
| Munich 3K (Liu and Mattyus, 2015) | 2016 | OD | transportation | 9 |
| Massachusetts Roads (Mnih, 2013) | 2013 | IS | transportation | 9 |
| †SSDD (SAR Ship Detection Dataset) (Li et al., 2017b) | 2017 | OD | transportation | 9 |
| VEDAI (Vehicle Detection in Aerial Imagery) (Razakarivony and Jurie, 2016) | 2016 | OD | transportation | 7 |
| †AIRSAR UAVSAR (NASA/JPL, 2020) | 2016 | IS | agriculture/ transportation | 7 |
| WHU Building Aerial (Ji et al., 2019) | 2018 | IS | settlement | 6 |
| Cheng roads (Cheng et al., 2017b) | 2017 | IS | transportation | 5 |
| RSOD (Remote Sensing OD) (Long et al., 2017) | 2017 | OD | multi-class OD | 5 |
| IEEE Zeebruges (IEEE GRSS, 2015) | 2015 | IS | settlement | 4 |
| HRSC2016 (High Resolution Ship Collections) (Liu et al., 2017) | 2016 | OD | transportation | 4 |
| GID (Gaofen Image Dataset) (Tong et al., 2018) | 2018 | IS | general LCLU | 4 |
| Zhang Aircrafts (Zhang et al., 2016a) | 2016 | OD | transportation | 3 |
| SpaceNet Building (SpaceNet, 2017) | 2017 | IS | settlement | 3 |
| UCAS-AOD (Zhu et al., 2015) | 2015 | OD | transportation | 3 |
| LCZ42 (Local Climate Zone 42) (Zhu et al., 2020) | 2020 | IS | settlement | 2 |
| Busy parking lot (Mou and Zhu, 2018) | 2018 | OD | transportation | 2 |
| DeepGlobe Roads (Demir et al., 2018) | 2018 | IS | transportation | 2 |
| NWPU RESICS 45 (Remote Sensing Image Scene Classification) (Cheng et al., 2017a) | 2017 | OD | multi-class OD/settlement | 2 |
| INRIA (Institut national de recherche en informatique et en automatique) (Maggiori et al., 2017) | 2017 | IS | settlement | 2 |
| †Open SAR Ship Dataset (Huang et al., 2018; Li et al., 2017a) | 2017 | OD | transportation | 2 |

### 2.5.3 CNN Architectures

The development of CNNs since the introduction of AlexNet in 2012 (Krizhevsky et al., 2012) has been diverse and fast-paced. In order to get an overview of the methodological developments and implementations, the evolution of CNNs in computer vision is highlighted, and the most prominent adaptations of the Earth observation domain are pointed out. Furthermore, for the most frequently used model types in Earth observation, insights are provided that explain which characteristics of CNNs are important for processing remote sensing data.

#### 2.5.3.1 CNNs for Image Recognition and Convolutional Backbones

The developments of CNNs, especially from the computer vision domain, are described by many abbreviations and acronyms to communicate the models and approaches. These abbreviations can help navigate the ramified developments when their names, chronological order, and relations are understood. To introduce the most important terminology, the tables 2.2, 2.3, and 2.4 shortly comment on frequently used abbreviations and provide the primary sources for a better understanding of the upcoming overview about the evolution of CNN architectures and their adaptation in Earth observation.

The ILSVRC challenge (Russakovsky et al., 2015) on the ImageNet data set (Deng et al., 2009) has driven the development of more sophisticated convolutional backbones for feature extraction (Hoeser and Kuenzer, 2020). Table 2.2 provides an overview of the milestone architectures and their typical abbreviations developed for this task. Figure 2.12

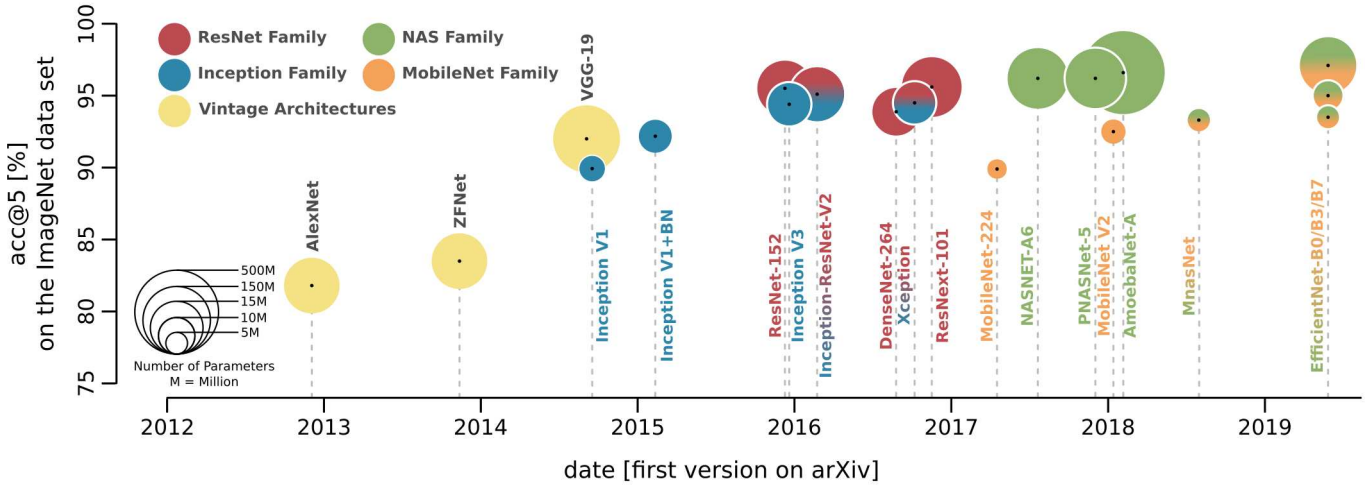



**Figure 2.12:** Evolution of convolutional neural network (CNN) architectures for image recognition with their size in parameters and performance on the ImageNet data set measured with its specific acc@5 metric (accuracy for the top 5 predicted classes of an image) (Russakovsky et al., 2015). Changed after Hoeser and Kuenzer (2020, p. 10).

shows the performance of milestone architectures on the ImageNet data set. It becomes clear that about three years after the introduction of AlexNet, an accuracy level of over 95% on the ImageNet data set was reached by evolving CNN architectures. The developments of the Inception (Szegedy et al., 2015; Ioffe and Szegedy, 2015; Szegedy et al., 2016; Chollet, 2017) and ResNet (He et al., 2016; Xie et al., 2017) architectures were particularly responsible for the initial increase. Further developments concentrated on minimising the model's parameters by keeping up the performance or even increasing it. The MobileNet models successfully proposed lightweight architectures (Howard et al., 2017). In combination with neural architecture search (NAS) (Zoph and Le, 2016; Tan et al., 2018), an approach to

**Table 2.2:** Overview of the terminology related to the developments in image recognition and convolutional backbones. The table is sorted to describe the developments of convolutional neural network (CNN) architectures with increasing complexity with proximity to content in order to group closely related CNN architectures. Changed after Hoeser and Kuenzer (2020, p. 4).

| Abbreviation | Reference | Short Comment |
|---|---|---|
| ImageNet | Deng et al. (2009) | Data set for image recognition |
| AlexNet | Krizhevsky et al. (2012) | CNN by Alex Krizhevsky et al. (2012) |
| ZFNet | Zeiler and Fergus (2014) | CNN by Zeiler and Fergus (2014) |
| VGG-16/19 | Simonyan and Zisserman (2014) | CNN by members of the Visual Geometry Group |
| Inception V1-3 | Szegedy et al. (2015); Ioffe and Szegedy (2015); Szegedy et al. (2016) | CNN architectures with Inception modules (build for extracting more complex features and intermediate reduction of feature-map depth by the so called bottle-neck design) |
| ResNet | He et al. (2016) | CNN architecture with residual connections (enabling very deep architectures which are still trainable) |
| ResNeXt | Xie et al. (2017) | Advanced CNN architecture, based on ResNet |
| Xception | Chollet (2017) | CNN architecture which combines ResNet and Inception modules |
| DenseNet | Huang et al. (2017) | Very deep CNN based on ResNet |
| NAS | Zoph and Le (2016) | Neural Architecture Search (the specific CNN architecture is *searched* by another neural network and no longer hand crafted by a human programmer) |
| NASNet | Zoph et al. (2018) | CNN architecture drafted with NAS |
| MobileNet | Howard et al. (2017) | Efficient CNN architecture with a focus on prameter reduction designed applications running on mobile platforms |
| MnasNet | Tan et al. (2018) | MobileNASNet, a parameter efficient CNN architecture drafted with NAS |
| EfficientNet | Tan and Le (2019) | A parameter efficient CNN architecture drafted with NAS and specific scaling approach to match the needed complexity |

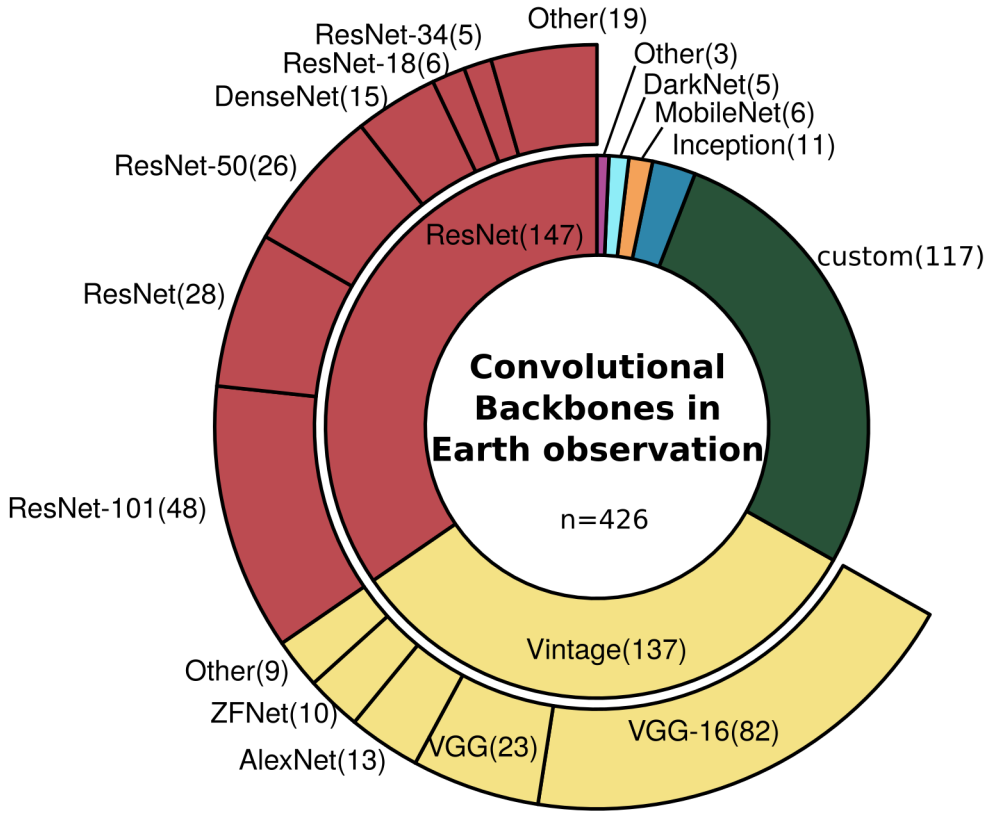



**Figure 2.13:** Number of adapted convolutional backbone architectures in the Earth observation domain. ResNet architectures are the most widely used due to their good performance and less complex structure compared to Inception architectures. They are closely followed by early architectures grouped in the Vintage category. Recent developments like the MobileNet are at the beginning of their adaptation phase. Changed after Hoeser et al. (2020, p. 17).

use another neural network to design the CNN architecture, they eventually became the EfficientNet models (Tan and Le, 2019). The EfficientNet architectures inherited the most important modules of their milestone-predecessors (Tan and Le, 2019; Hoeser and Kuenzer, 2020) which are still widely used in and outside the computer vision domain.

After an initial period where the early architectures like AlexNet (Krizhevsky et al., 2012), ZFNet (Zeiler and Fergus, 2014), and VGG (Simonyan and Zisserman, 2014), grouped as vintage architectures, were used, the Earth observation domain mainly focused on the adaptation of the ResNet (He et al., 2016) convolutional backbones, as depicted in figure 2.13. A lower complexity characterises the ResNet architecture compared to Inception and EfficientNet models, as well as a good performance with moderate use of parameters (Hoeser and Kuenzer, 2020). A crucial factor for its widespread application is that ResNet is commonly implemented in popular deep learning frameworks (Abadi et al., 2015; Chollet et al., 2015; Paszke et al., 2019) or that it can be quickly rebuilt with the functionalities of the frameworks due to its relatively low complexity. This combination makes

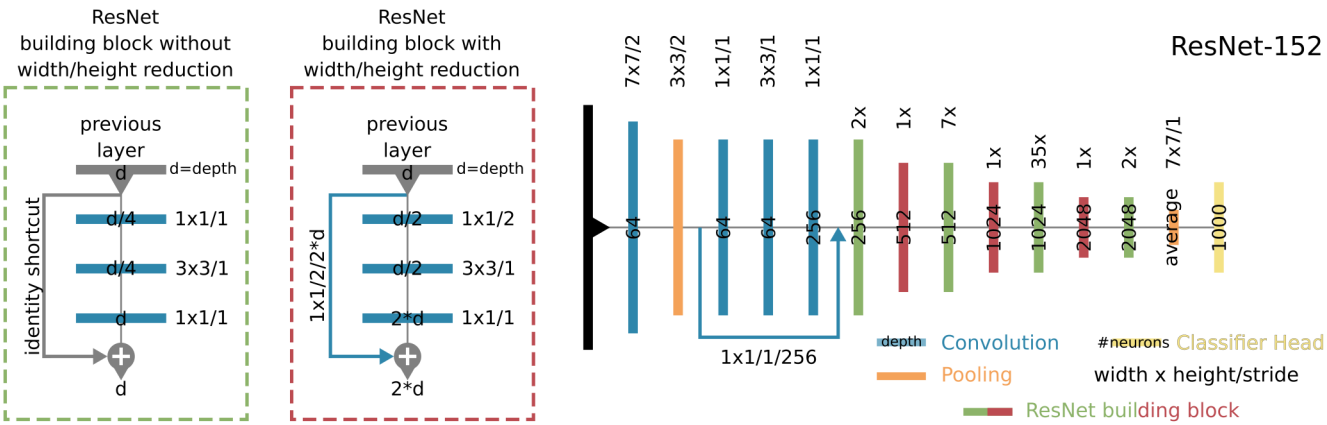



**Figure 2.14:** Schematic overview of the ResNet-152 (152 layers deep) architecture, proposed by (He et al., 2016) and the ResNet specific building block with the so-called residual connection. Changed after Hoeser and Kuenzer (2020, p. 13).

the ResNet architecture particularly interesting for researchers from the Earth observation domain. Looking closer at the figures 2.12 and 2.13 makes clear that in Earth observation, less deep variants like the 101 or 50 layers deep ResNet architectures are preferred over the 152-layer deep architecture, which shows the best performance in the computer vision domain. The reason is that the deep variants in computer vision were used for the 1,000 class problem of the ImageNet data set. Whereas in Earth observation, fewer classes are investigated in a single remote sensing scene. Thus fewer parameters and with it, shallower models are better suited (Hoeser and Kuenzer, 2020; Hoeser et al., 2020).

The core idea of the ResNet architecture is the so-called residual connection, see He et al. (2016). Figure 2.14 gives an impression of this feature. It can be understood as a bypass for feature maps around the next stack of convolutional layers. At first glance, this seems counterintuitive since less representational information is transported in deeper layers of the network. Nevertheless, this intuition changes when seen from a backpropagation perspective. When only considering a small stack of convolutional layers as the model $H(x)$, without the bypass connection, the function of the convolutional layers to learn would be $F(x)$, thus $H(x) = F(x)$. With the residual connection, the input $x$ additionally bypasses the model and is added to the model's output. This changes the function to $H(x) = F(x) + x$ which can be expressed as $H(x) - x = F(x)$. Following that, the function $F(x)$ to learn is no longer the direct mapping of $H(x)$ but the residual $H(x) - x$. Rethinking that in a neural network context means that when additionally adding the input layers to the results of the next convolutional layer, the model has to perform at least as good as in the layer before. It is impossible for it to perform worse since the layers are now nested functions, where the model's knowledge until the current layer is available through the bypass connection. That way, the current layers have only to learn the additional complexity they add to the entire model instead of learning an entirely new function. Another important effect is, that

due to the residual connection, continuously decreasing gradients during backpropagation are prevented. That way, early layers receive an optimisation signal which otherwise would vanish in deep networks without residual connections (He et al., 2016).

He et al. (2016) demonstrated that with residual connections, very deep networks with hundreds and even thousands of layers could be built and continuously become better in solving the given task. Due to the good performance of ResNet on the ImageNet and other leading data sets in the computer vision domain, and its straightforward technical implementation, the ResNet building blocks are widely accepted and used in CNN model architectures in the computer vision as well as the Earth observation domain.

### 2.5.3.2 CNNs for Image Segmentation

Image segmentation, or pixel-wise classification from an Earth observation perspective, is driven by the PASCAL-VOC (Everingham et al., 2010, 2015) and the Cityscape data sets (Cordts et al., 2015, 2016) in the computer vision domain. Table 2.3 provides an overview of important abbreviations used during this introduction. Furthermore, the temporal evolution of milestone models and how they perform on the PASCAL-VOC benchmark data set is depicted in figure 2.15. Two major model types have been proposed, the encoder-decoder networks and the naïve-decoder networks. Both start with a convolutional backbone, which extracts the representational features. In a naïve-decoder model, relatively simple bilinear

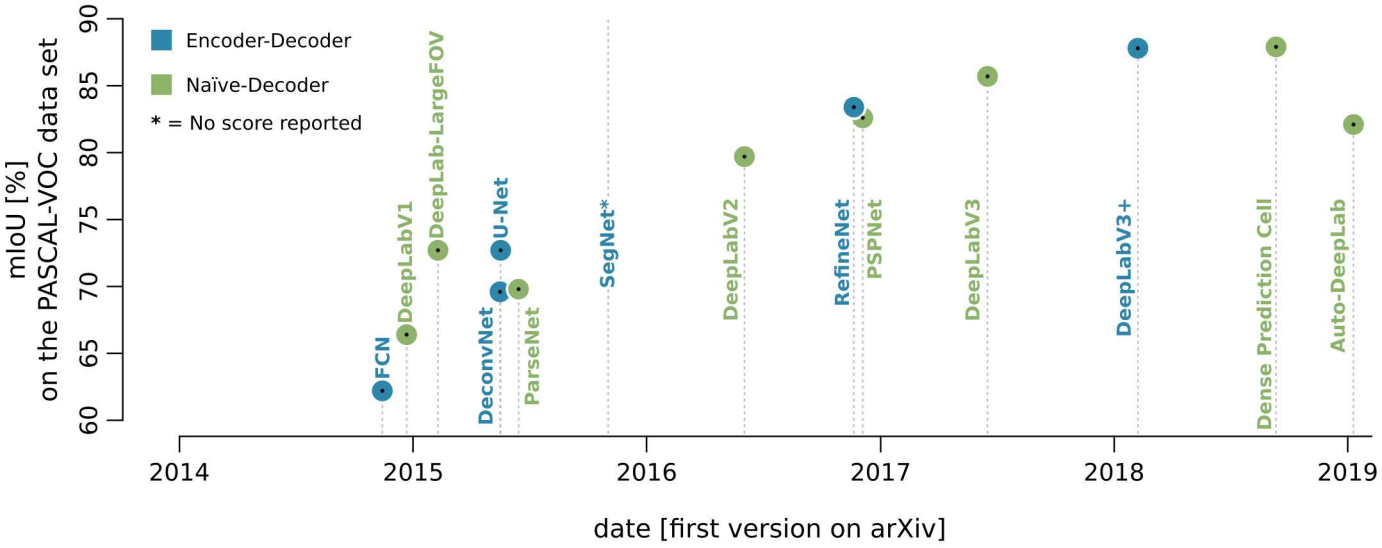



**Figure 2.15:** Evolution of convolutional neural network (CNN) architectures for image segmentation grouped by the two major architecture concepts encoder-decoder and naïve-decoder models. They are compared by their performance on the Pattern Analysis, Statistical modelling and ComputAtional Learning-Visual Object Classes (PASCAL-VOC) data set measured by the mean intersection over union (mIoU) (Everingham et al., 2010, 2015). Changed after Hoeser and Kuenzer (2020, p. 17).

**Table 2.3:** Overview of the terminology related to the developments in image segmentation with convolutional neural network (CNN). The table is sorted to describe the developments of CNN architectures with increasing complexity and proximity to content in order to group closely related CNN architectures. Changed after Hoeser and Kuenzer (2020, p. 4).

| Abbreviation | Reference | Short Comment |
|---|---|---|
| PASCAL-VOC | Everingham et al. (2010, 2015) | Pattern Analysis, Statistical modelling and Computational Learning - Visual Object Classes (data set for image segmentation and object detection) |
| FCN | Shelhamer et al. (2014) | Fully Convolutional Network |
| DeepLabV1-V3+ | Chen et al. (2014); Papandreou et al. (2015); Chen et al. (2016, 2017, 2018a) | CNN architectures developed for image segmentation with a specific focus on investigating multi-scale spatial features by introducing the ASPP module |
| ASPP | Chen et al. (2016, 2017) | Atrous Spatial Pyramid Pooling, a module which applies convolutions with different spatial scales |
| DeconvNet | Noh et al. (2015) | Deconvolutional Network CNN which inverts convolutional operations to restore input image size for refined segmentation masks) |
| U-Net | Ronneberger et al. (2015) | U-shaped encoder-decoder architecture, to combine deep features with a low spatial information and earlier extracted features with a higher spatial information depth during deconvolution |
| ParseNet | Liu et al. (2015) | Parsing global image context to pixel wise classification to include large scale image context |
| PSPNet | Zhao et al. (2016) | Pyramid Scene Parsing Network for extracting features on multiple scales |
| AutoDeepLab | Liu et al. (2019) | DeepLab related architecture drafted with NAS for image segmentation |

interpolation is used to restore the input image dimension from the deepest feature maps of the convolutional backbone and finally predict the pixel-wise segmentation masks for each class. Especially CNNs from the DeepLab family use this technique (Chen et al., 2014; Papandreou et al., 2015; Chen et al., 2016, 2017). One exception of this group is the DeepLab-V3+ variant (Chen et al., 2018a), which is an encoder-decoder model.

The general idea of encoder-decoder models, as depicted in figure 2.16, is to use deconvolutional layers in the decoder path, which can be imagined as an upside-down convolutional backbone. Due to this mirrored appearance of the model architecture, the most widely used U-Net model (Ronneberger et al., 2015) has gotten its name, where the left side of the U is the encoder, the right side the decoder and the bottom of the U can be understood as the connection between them. In addition to the encoder-decoder structure,

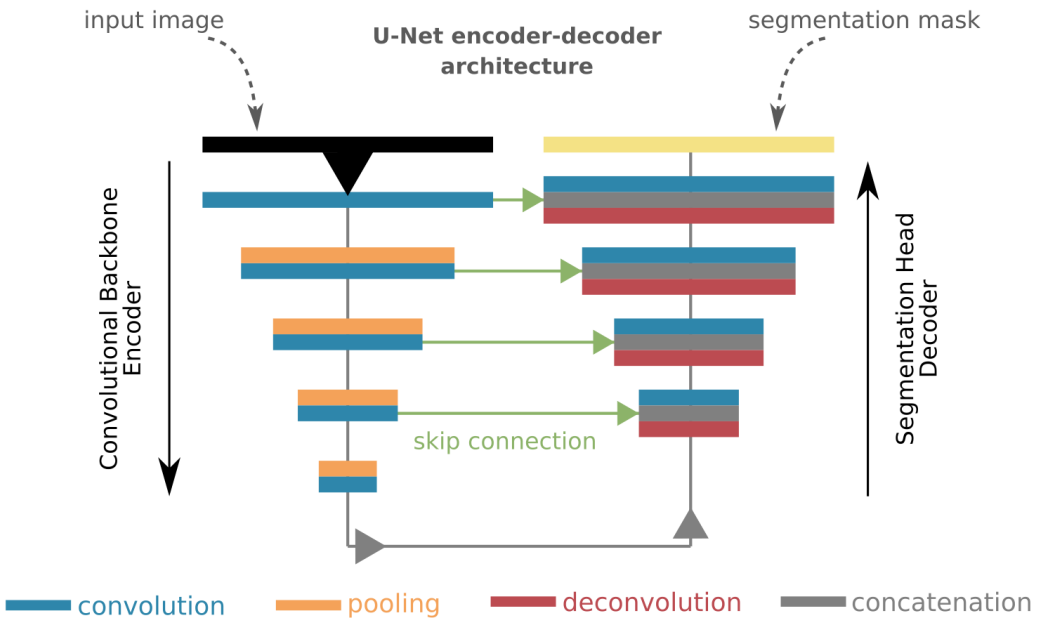



**Figure 2.16:** Schematic overview of the U-Net encoder-decoder architecture (Ronneberger et al., 2015). A convolutional backbone is the encoder, where the decoder restores the image resolution to input dimensions by using deconvolution operations. During deconvolution in the decoder, skip connections enrich the spatial accuracy of the feature maps with accurate location information from the decoder. Changed after Hoeser and Kuenzer (2020, p. 17).

U-Net and other encoder-decoder models are characterised by skip connections. They connect layers with the same dimension from the encoder with the decoder directly, which enables feature sharing between the encoder and decoder, see figure 2.16. That way, information with an accurate spatial localisation but less representational features from the encoder are combined with highly representational feature maps but a relatively low spatial accuracy in the decoder. That way, the extracted representational features are subsequently refined with accurate spatial information during the deconvolution process. Since in Earth observation, targets are often relatively small clusters of pixels within a larger scene, and small scale spatial details are of importance, encoder-decoder models, which support these details to appear correctly in the segmentation masks, are the models of choice and the reason why naïve-decoder architectures are not that widely used due to their bilinear upsampling approach, as reported in figure 2.17. Furthermore, in Earth observation, features can appear on multiple spatial scales. For instance, a lake has a homogeneous appearance at a small scale in its centre but a highly heterogeneous appearance at a large scale when its shoreline is taken into account. Thus different features on different scales can contribute to a better classification of all pixels which belong to a multi-scale object (Miao et al., 2018). A module that is capable of this is the atrous spatial pyramid pooling (ASPP) module, originating from the DeepLab family (Chen et al., 2016, 2017). It consists of multiple parallel

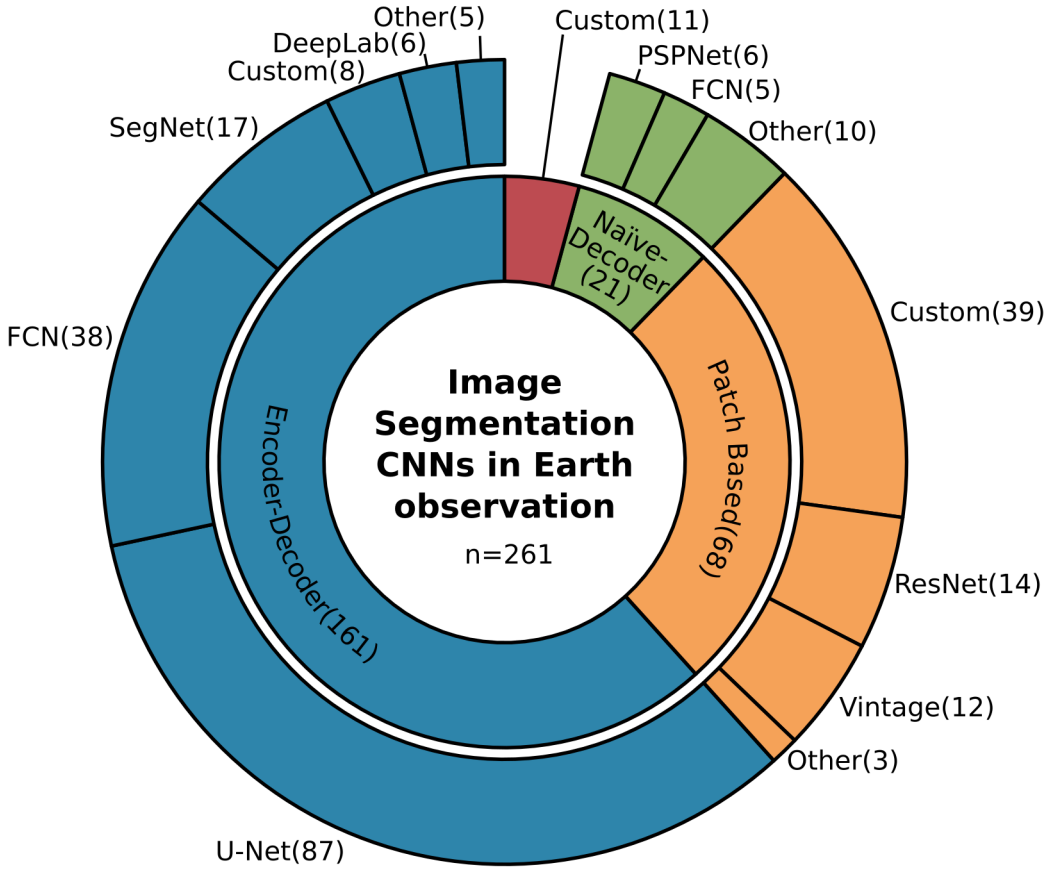



**Figure 2.17:** Number of adaptations of image segmentation architectures in the Earth observation domain. Encoder-decoder architectures are the most widely used due to their spatial refinement of features during the deconvolution process, which was found to be highly beneficial for image segmentation of detailed Earth observation data. Changed after Hoeser et al. (2020, p. 19).

convolutional layers, where each kernel has a different spatial distance between the pixels used during the convolution. These kernels can be imagined as kernels with holes, hence the name *à trous* from French for *including holes*. In the Earth observation domain, it was found that plugging in the ASPP module at the bottom of the U-Net model can enrich the model's capability of extracting multi-scale spatial features (Dirscherl et al., 2021; Miao et al., 2018; Fu et al., 2019; Li, 2019; He et al., 2019; Hoeser et al., 2020).

#### 2.5.3.3 CNNs for Object Detection

Finally, in object detection, the Microsoft-common objects in context (MS-COCO) data set (Lin et al., 2014) is commonly used as a benchmark in the computer vision domain. Microsoft compiled RGB images with bounding boxes which annotate 80 classes of common objects in context, hence the name MS-COCO. See table 2.4 for further explanation of the most important abbreviations used in this section. Two major model types can be differentiated in object detection: one-stage and two-stage detectors. The one-stage detector

**Table 2.4:** Overview of the terminology related to the developments in object detection with convolutional neural network (CNN). The table is sorted to describe the developments of CNN architectures with increasing complexity and proximity to content in order to group closely related CNN architectures. Changed after Hoeser and Kuenzer (2020, p. 4).

| Abbreviation | Reference | Short Comment |
|---|---|---|
| MS-COCO | Lin et al. (2014) | Microsoft-Common Object in Context (data set for object detection and instance segmentation) |
| R-CNN | Girshick et al. (2013) | Region based CNN, potential regions are derived in an image segmentation based preprocessing step, each derived region is then passed through multiple CNNs where each CNN detects a specific class |
| Fast R-CNN | Girshick (2015) | R-CNN+RoI pooling based CNN, region proposals are derived similar to R-CNN, but the classification process is performed by a single CNN |
| RPN | Ren et al. (2015) | Region Proposal Network, CNN based submodule which proposes regions for class agnostic objects |
| Faster R-CNN | Ren et al. (2015) | R-CNN+RPN+RoI pooling based CNN, the end-to-end trainable variant of R-CNN which can detect regions of interest and classify them in an end-to-end trainable CNN model. |
| FPN | Lin et al. (2016) | Feature Pyramid Network, multicale feature refinement of the convolutional backbone |
| Cascade R-CNN | Cai and Vasconcelos (2018) | Faster R-CNN based cascading detector for less noisy detections |
| Mask R-CNN | He et al. (2017) | Combination of Faster R-CNN and FCN (image segmentation) for instance segmentation |
| CBNet | Liu et al. (2019b) | Composite Backbone Network, Faster R-CNN based object detection with multiple backbones which subsequently share feature maps for multiscale-multidepth feature refinement |
| YOLO-V1-3 | Redmon et al. (2015); Redmon and Farhadi (2016, 2018) | You Only Look Once, fast and parameter efficient object detectors, developed for mobile platforms |
| DarkNet | Redmon and Farhadi (2016, 2018) | CNN backbone for YOLO-V2+3 |
| EfficientDet | Tan et al. (2020) | Parameter efficient CNN for object detection based on the EfficientNet convolutional backbone |

models combine object detection and classification in a single unified process. Thereby, the model variants by Redmon et al. (2015) named You Only Look Once (YOLO) (Redmon et al., 2015; Redmon and Farhadi, 2016, 2018) are together with the single shot multibox detector (SSD) (Liu et al., 2016a) the most widely used one-stage detectors. The advantage of these models is their small number of parameters, which makes them particularly interesting for applications on smartphones and other mobile platforms like unmanned aerial vehicles (UAVs). However, what can be seen in figure 2.18, one-stage detectors until late 2019 are lagging behind in performance compared to two-stage detectors. Typically,

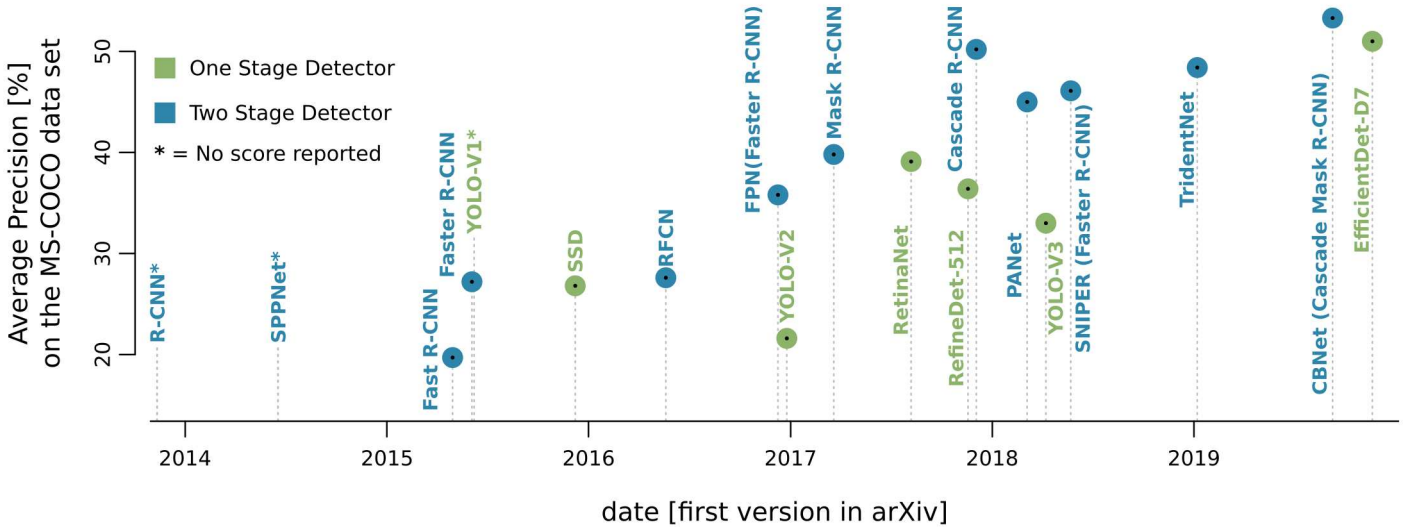



**Figure 2.18:** Evolution of convolutional neural network (CNN) architectures for object detection grouped by the two major architecture concepts one stage detector and two-stage detector. They are compared by their performance on the Microsoft-common objects in context (MS-COCO) data set measured by the average precision (AP) (Lin et al., 2014). Changed after Hoeser and Kuenzer (2020, p. 22).

two-stage detectors are heavier in parameters and need more powerful hardware and time to process an image, but generally perform better than one-stage detectors. Since both conditions, hardware capabilities and processing time, are normally not that crucial in analysing Earth observation imagery compared to mobile applications, two-stage detectors are by far the most widely used object detector CNNs in Earth observation, see figure 2.19. The region based-convolutional neural network (R-CNN) (Girshick et al., 2013) and Fast R-CNN (Girshick, 2015) models are the origins of two-stage detectors. Their successors, the Faster R-CNN (Ren et al., 2015) and Mask R-CNN (He et al., 2017) are among the most widely used in Earth observation, see figure 2.19. The overall idea of the Faster R-CNN architecture (Ren et al., 2015) is depicted in figure 2.20. An object detection head uses shared features from the convolutional backbone to perform bounding box regression and object classification. In order to do so, first, the so-called region proposal network (RPN), a submodule of the Faster R-CNN, looks for class agnostic object proposals by using predefined anchor boxes with width, heights, and scale factors that are equally distributed over the last feature map of the convolutional backbone. The bounding box regression then uses these relatively rough region proposals to refine the object location. In a parallel task, the features cropped by the region proposals are used in a classification process similar to image recognition to predict the class of the object (Ren et al., 2015).

Due to feature extraction and region proposals in the first step and subsequent location refinement and classification in the second step, they are called two-stage detectors. It is important to understand that the Faster R-CNN model uses shared features from the

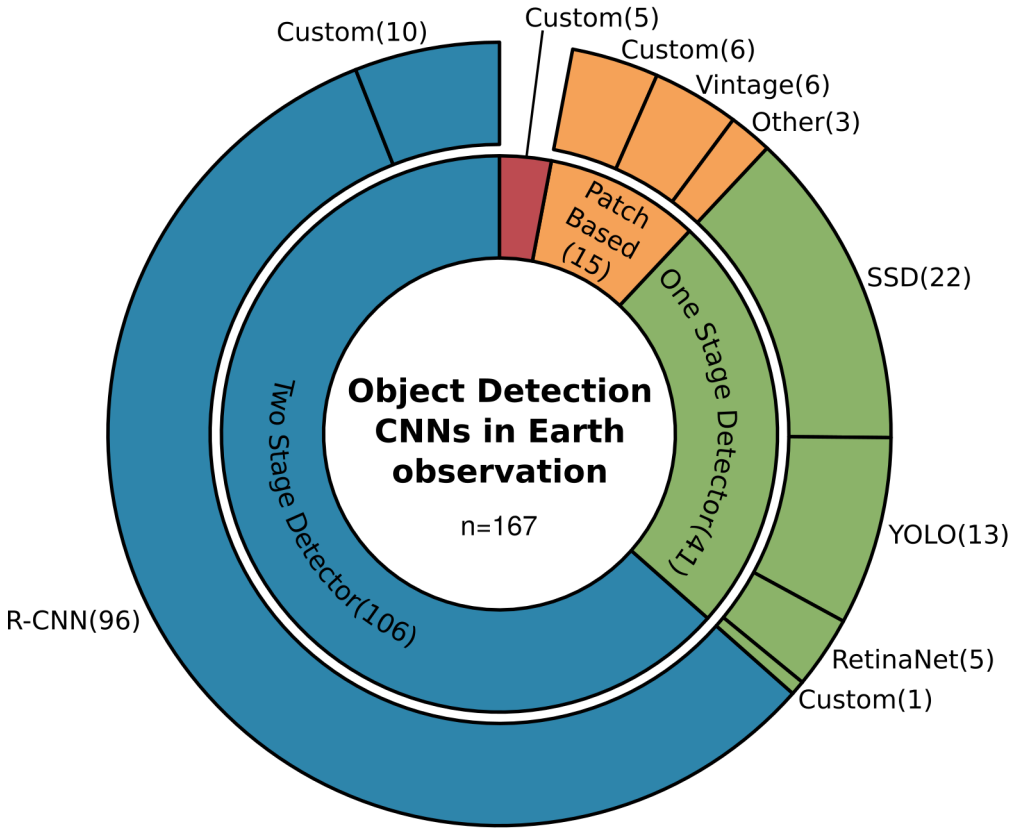



**Figure 2.19:** Number of adaptations of object detection architectures in the Earth observation domain. Two-stage detector architectures are the most widely used due to their better performance yet higher amount of parameters compared to one-stage detectors. With their modular design, two-stage detectors allow for plug in modifications which were found to be beneficial in Earth observation by adding multi-scale feature extraction (Lin et al., 2016). Changed after Hoeser et al. (2020, p. 21).

convolutional backbone to perform the location regression and the classification task. The optimisation of the convolutional backbone in order to provide features that can be used in two different tasks is possible due to the multi-task loss, see figure 2.20. The entire cost function $J_{\text{multi}}$ to optimise the network is a sum of the cost function of the bounding box regression $J_{\text{reg}}$ and the object classification $J_{\text{cls}}$. By reducing $J_{\text{multi}}$, the parameters of the network are optimised to extract features that can be used by both tasks simultaneously (Ren et al., 2015).

Further developments of the Faster R-CNN architecture are manifold (Lin et al., 2016; Cai and Vasconcelos, 2018; Liu et al., 2019b; Ghiasi et al., 2019). One important variant is the feature pyramid network (FPN) (Lin et al., 2016) which was proposed in late 2017, about one and a half years after the Faster R-CNN was introduced, see figure 2.18. Instead of using a pyramid of anchor boxes with different scales in the RPN, it combines the last layers of the convolutional backbone similar to the U-Net, with a stepwise increasing dimension of the feature maps and skip-connections between those of the origin backbone.

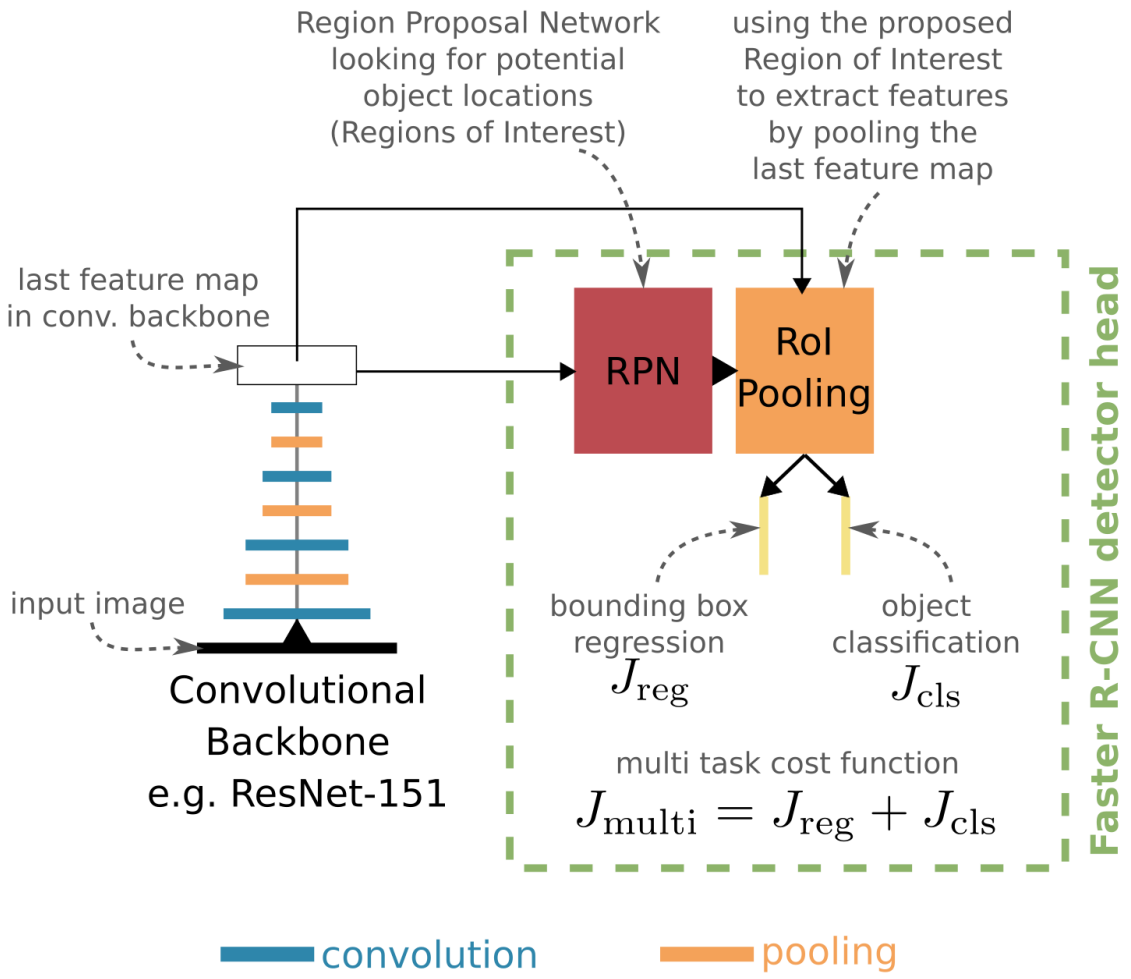



**Figure 2.20:** Schematic overview of the Faster region based-convolutional neural network (Faster R-CNN) two-stage detector architecture (Ren et al., 2015). A convolutional backbone is connected to the region proposal network (RPN) which looks for class agnostic object proposals. Their spatial location is then refined by bounding box regression, and their class is predicted by a classifier. The cost functions of both tasks are combined in a joint cost function $J_{\text{multi}}$. That way, the entire convolutional neural network can be optimised to share features for both tasks. Changed after Hoeser and Kuenzer (2020, p. 24).

That way, representational features of the last layers can appear in earlier layers with a higher spatial resolution and larger scale. The result is a feature pyramid that can be passed as inputs with multiple scales to the RPN (Lin et al., 2016). This is important from an Earth observation perspective since targets of interest are often very small compared to the input image dimension and might vanish in the last layers of a deep convolutional backbone. Due to the feature pyramid, this can be prevented, and they appear in recovered earlier layers but in combination with highly representational features from deeper layers (Hoeser and Kuenzer, 2020; Hoeser et al., 2020).

Regardless of the task, image recognition, image segmentation or object detection, one important property of the presented architectures is their ability to recognise spatial features on multiple scales. The most common approaches combine semantically deep features but with a spatially lower resolution with semantically less deep features but spatially higher resolution. With the developments in CNN architectures since 2012, these models are capable of extracting information from imagery data which are highly relevant for applications in the Earth observation domain.

## 2.6 Discussion

In this introduction, details from the fields of machine and deep learning were discussed with particular reference to CNNs and their application in Earth observation. In addition to the details discussed, it is important to consider the basic motivation of the individual areas in order to derive potential challenges from them. Therefore, three major aspects can be summarised from the previous sections:

- With machine learning methods, it is possible to automatically optimise the parameters of deep learning models and learn representations of complex relationships in order to obtain algorithms that fulfil predefined tasks.
- With state of the art developments of CNN architectures and modules coming from the computer vision domain, it is possible to investigate remote sensing data with a special focus on spatial features and context in which targets appear on the land surface.
- In supervised deep learning, a large size and comprehensive variability of the training data sets are crucial to enable the full potential of deep learning models, consisting of millions of trainable parameters. Without such training data sets, these models are exceptionally prone to overfitting.

The first aspect indicates that complex models can be optimised in an automated way, provided that technical implementations exist. With modern deep learning frameworks accessible, the focus can be shifted to the aspect of defining the task. This is crucial since, from an artificial intelligence perspective, the goal is to obtain a narrow artificial intelligence which implies that a task is well defined. The definition of the task can start with a simple semantic description, like, *the extraction of offshore wind energy infrastructure from Earth observation data*. However, providing more details narrows the task and supports a more specialised artificial agent, which finally will increase its performance or decreases the amount of complexity the model has to learn.

From an Earth observation perspective, this is important to be considered carefully since the data from this domain often is a highly complex representation of Earth systems which are often not fully understood. Thus, a definition of what is a *target* and what is a *non-target* is important. In the next step, additional information related to both classes helps refine the task description further. Do specific spatiotemporal characteristics of the target exist? Are there non-targets that have a stronger relation to the target than others? The final task is narrowed down by answering such questions, and finally, more technical decisions can be made. For instance, which Earth observation data type is suitable, which deep learning model can handle the data structure, and in the case of CNNs, which task is to perform: image recognition, object detection or image segmentation. Here we come full circle to the technical implementations. Now the building blocks contained in the machine learning frameworks can be configured in such a way that a technically explicit description of the previously made definition of the task and thus of the artificial agent can be modelled. In this way, knowledge about the conditions of the system to be investigated flows into the conception of artificial intelligence through a definition of the task to be fulfilled.

The second aspect emphasises the capability of specific CNN architectures to investigate spatial features. The strength of remote sensing data lies in the depiction of spatial relationships. Thus, a targeted analysis of these relationships in order to be able to make statements about individual components in a remote sensing scene meets the core of basic geographical principles. Tobler's first law of geography states that "everything is related to everything else, but near things are more related than distant things" (Tobler, 1970, p. 236). With that in mind, spatially aware CNNs seem to be a good choice for a hypothetical model when learning representations from Earth observation data. As discussed in detail in the introduction, there are certain architectures of CNNs that strongly enable taking spatial features into account. Of particular interest are those that can depict spatial context. With detailed knowledge about such architectures, it is possible to create an intelligent agent, which is not only an expert due to a well-defined task but also a professional in analysing spatial data by learning its underlying representations. This is also crucial from a methodological point of view. Instead of building an extremely complex model with theoretically infinite numbers of parameters, a less complex model with narrower expertise can be designed by taking specific characteristics of the input data and the task definition into account. Thus, even when using heavy parameter approaches like deep learning, it is necessary to keep in mind that fewer parameters are beneficial (Tobler, 1970) from a computational perspective but also an perspective about how well the model can be explained. A CNN which consists of specific building blocks that are designed to fulfil specific tasks, is much more approachable than a fully connected ANN with far more parameters and a less intuitive structure. Overall, from an Earth observation and spatial analysis perspective CNNs, when designed

carefully, are self-learned approximations of abstract spatial functions which contribute to the scope of established, hand-crafted spatial analysis procedures like Kriging (Krige and Magri, 1982), Morran's I (Moran, 1950; Anselin, 1995) or Wavelet analysis (Kumar and Foufoula-Georgiou, 1997).

In supervised deep learning, a well-designed CNN architecture is nevertheless only a model hypothesis with great but as yet unexploited potential. In order to unfold this potential, large, variable training data sets are required from which the model can learn a generalised representation of features to solve a given task. This point was stressed on multiple occasions during the introduction and shall be taken into account during the conceptualisation of each deep learning-based investigation. Two major questions have to be answered in supervised deep learning at an early stage: Do I have access to enough potential training data? Do I have access to enough resources to annotate the data in order to generate an extensive and precise training data set? In any case, the generation of such large and precise data sets with a wide variety of training examples by hand is a labour-intensive task. In Earth observation, the problem is even more prominent since the data can be very specific, and a high level of experience is necessary to annotate the data correctly. Thus the data set generation process can not be outsourced. The review of 429 publications in Earth observation clearly shows that, likewise to the computer vision domain, available training data sets have a massive impact on the application domains of deep learning. This indicates that at the reviewed stage, in Earth observation, the focus was mainly on transferring the methods from the computer vision to the Earth observation domain by using existing training data sets. Only recently, studies that use own data sets generated for a specific problem are becoming more frequent. The availability and variation of training data sets is thus an essential driver for the Earth observation domain to expand the fields of applications. By increasing the availability of training data sets and methods to generate them efficiently, the focus of studies can be increased from investigations of technical research questions **about** deep learning in Earth observation towards studies that investigate geoscientific research questions **with** deep learning.

# Chapter 3

# *Offshore Wind Energy Infrastructure in Earth Observation**

## 3.1 Related Research

The detection of persistent marine infrastructure from Earth observation data is of multiple interests. From an engineering perspective, automated detection supports getting an overview of technical infrastructure's spatial distribution. Precise location information supports the planning of supply routes, safety and security measures for the infrastructure itself and people working on it, and appropriate decommissioning at the end-of-life of a facility and searching for new sites by taking already occupied areas into account. From an ecological perspective, it is important to track both negative and positive impacts during the entire life cycle of marine infrastructure on marine ecosystems. Its spatial distribution and temporal development are crucial and an essential link between engineering and an ecological perspective. With spaceborne remote sensing, it is possible to map marine space frequently, which is otherwise hard to oversee. If no further differentiation is made, persistent infrastructure in marine space is referred to as rigs, platforms, maritime or marine infrastructure in the remote sensing literature.

From an Earth observation perspective, the detection of marine infrastructure above the water surface relies on the distinct contrast of objects in an otherwise homogeneous area. This property has been investigated for marine infrastructure in general and, in some cases, to extract offshore wind turbines using handcrafted approaches. Liu et al. (2016b) extracted persistent marine infrastructure from the optical sensor system of the Landsat-8 mission by weeding object candidates by their temporal and positional invariance and due to morphological characteristics. By combining images from the Landsat and Sentinel-2 sensors,

*Parts of this chapter have been published in Hoeser et al. (2022)

Xu et al. (2020) used a combination of order statistic filtering (Bovik et al., 1983) and thresholding to propose marine infrastructure object candidates. In a second step, existing geometries which describe the boundaries of OWFs in the North Sea Basin were used to extract those object candidates which are within OWF areas in order to classify them as OWTs. For areas outside the North Sea Basin, where no OWF boundaries were available in the study of Xu et al. (2020), the extracted marine infrastructure objects were visually examined to identify gridded clusters. All previously extracted object candidates within these clusters were then grouped as OWTs, without any further differentiation between other platform types which occur in OWFs. Both studies by Liu et al. (2016b) and Xu et al. (2020) demonstrate that the general detection of persistent marine infrastructure is possible by investigating multispectral radar imagery. Moreover, the study of Xu et al. (2020) shows that OWFs have specific large scale spatial features, the grid-like patterns, that can be recognized by humans. However, automatic internal differentiation of marine infrastructure, especially between different offshore wind energy infrastructure types, is more challenging and needs additional data or a human interpreter who is looking for a larger spatial context or small scale details.

In radar remote sensing, Zhang et al. (2019b) and Wang et al. (2019b) used the constant false alarm rate (CFAR) approach which earlier has been applied successfully for vessel detection in SAR imagery (Wackerman et al., 2001) to derive persistent marine infrastructure. the CFAR approach has the advantage that no static threshold needs to be set in order to identify a target object. Instead, a 2D kernel defines target and background areas that are sensitive to contrast instead of absolute values. Wong et al. (2019) further developed this method. To increase the signal to noise ratio of median composites of multiple radar images, they used the difference of gaussians (DoG) method before applying a CFAR based thresholding by taking the mean or median background values of a 250 m radius around object candidates into account. With further postprocessing of the morphology of the object candidates and complex thresholding, they were able to derive marine infrastructure from radar images. Besides detecting oil rigs in the Gulf of Mexico, their algorithm has also been applied in the exclusive economic zones (EEZ) of the UK and the East China Sea, where they were able to detect offshore wind energy infrastructure in predefined OWF test sites.

In 2021, Zhang et al. (2021b) proposed a global OWT data set derived from Sentinel-1 data by applying multiple thresholds over predefined morphological parameters. This data set demonstrates that it is possible to investigate morphological features present in radar images to detect offshore wind turbines specifically. However, the approach tends to wrongly classify offshore platforms within wind farms as turbines, although they are transformer stations, meteorological masts or foundations of unfinished OWTs. Furthermore, in

their proposed data set global offshore wind turbine (GOWT) v1.3, large accumulations of OWTs, like the largest UK offshore wind farm project named Horn Sea with hundreds of platforms, are missing. In addition to the location of an OWT, Zhang et al. (2021b) and Xu et al. (2020) both derived the first appearance of an OWT. In both approaches, the construction phase, which can last from days to months and even a year, is not considered. What is recognised by both approaches is the first appearance of a platform at a wind turbine location, but no differentiation is made between a platform under construction and a platform with a readily deployed wind turbine. This emphasises the difficulty but necessity of differentiating marine infrastructure, especially when investigating offshore wind energy infrastructure.

The following insights about remote sensing data and the task to be defined can be derived from these studies. In the context of marine infrastructure, radar data have the advantage of being independent of weather conditions, which were found to be problematic when detecting OWTs in optical remote sensing data (Xu et al., 2020; Zhang et al., 2021b). Furthermore, median composites of multi-temporal acquisitions make it possible to remove moving or only briefly lingering objects from the data (Wong et al., 2019). This crucial reduction of potential false positives before applying the detection algorithm is a strong argument for using radar imagery for detecting OWTs. A precise differentiation between other persistent marine infrastructure types is crucial to explicitly extract offshore wind energy infrastructure. Where the superordinate category marine infrastructure has a very distinct spatiotemporal signal in remote sensing images, its subcategories appear to be very close and challenging to separate. Hand-designed thresholds and morphological parameters seem to have reached their limits in the process.

A definition of the task to perform for the later designed intelligent agent can be formulated. The task is to automatically detect offshore wind energy infrastructure in non-presorted radar imagery of marine areas worldwide. It has to differentiate offshore wind energy infrastructure from other marine infrastructures. Furthermore, an internal differentiation is necessary to separate OWTs from other offshore platforms, commonly situated in offshore wind farms like transformer stations, meteorological masts, or foundations of offshore wind turbines under construction. Besides the location information, the temporal deployment stages of offshore wind energy infrastructure have to be differentiated by closely investigating changes in the derived object type to avoid an overestimation of readily deployed OWTs.

## 3.2 Offshore Wind Energy Infrastructure in Sentinel-1 Radar Imagery

In the previous section, it was discussed that specific grid-like, regular patterns were used by humans to visually examine derived marine infrastructure objects to classify them as OWFs (Xu et al., 2020). Furthermore, in the study of Zhang et al. (2021b) morphological parameters were investigated to determine OWTs in radar imagery specifically. From a CNN perspective, these are promising insights since they demonstrate that multi-scale spatial features can be exploited to extract the target of interest, a typical use case for a CNN as discussed in chapter 2.5.1. However, it is important to fully understand the signal and its spatial features, which will be investigated in order to define a suitable CNN architecture and methodological study design. The following section provides an in-depth discussion of the investigated Sentinel-1 data to provide information on how offshore wind energy infrastructure and their typical environment appear in radar imagery.

The Sentinel-1 mission belongs to the Copernicus Earth observation programme of the EU (Aschbacher, 2017), operated by the European Space Agency (ESA). ESA launched the two satellites Sentinel-1A and Sentinel-1B in April 2014 and 2016, respectively. Both satellites share the same orbit path at a flight height of about 700 km. Each platform has a revisit rate of 12 days. Thus together, the Sentinel-1 mission maps the same spot every six days. Mounted on the spaceborne platforms are active, dual-polarisation C-band synthetic aperture radar (SAR) instruments that map the Earth continuously. With a wavelength of 5.6 cm, the radar signal is independent of cloud cover and, since actively emitted, the instrument is able to acquire data day and night. The sensor geometry is a right looking geometry. This means that when acquiring data in both orbit directions, the sensor looks East in ascending orbit direction and West in descending orbit direction, see figure 3.1 (Torres et al., 2012).

Besides the raw data, ESA provides two level 1 data products, the Single Look Complex (SLC) product, which combines amplitude and phase information, and the Ground Range Detected (GRD) product. For the GRD product, the SAR signal is multi-looked, an averaging process of neighbouring pixels, which results in square pixels with reduced speckle. The pixel values describe the backscatter magnitude in decibel and, compared to the SLC product, the phase value is lost during the multi-look process. The pixel spacing of the GRD product depends on the acquisition mode of the sensor. In this work, the GRD product with the acquisition mode Interferometric Wide swath (IW) with a swath width of 250 km and incident angles between 31° and 46° is used. The final pixel spacing for this sensor specification is 10 m $\times$10 m (Torres et al., 2012).

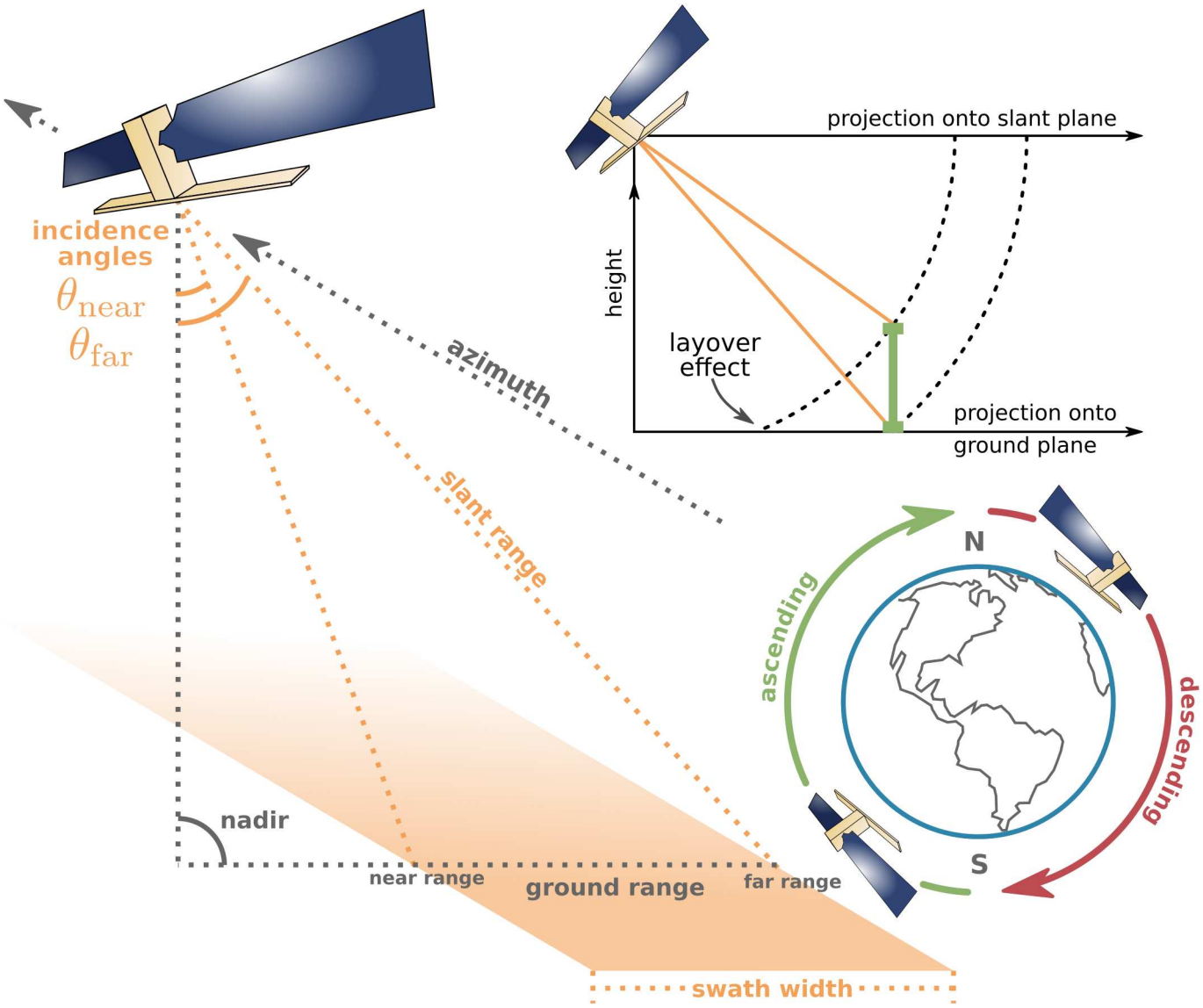



**Figure 3.1:** Conceptual overview of the right looking sensor geometry of the Sentinel-1 mission, its two possible orbit directions, and an example of the layover effect caused by the interaction of a vertical structure with the radar signal.

The selected polarisation of the radar signal is vertical transmit horizontal receive (VH). It was chosen over the vertical transmit vertical receive (VV) polarisation since only a small part of the signal returns with a changed polarisation from open water. The VH signal is less sensitive to waves, which stronger interact with the VV signal, resulting in higher backscatter amplitudes in VV bands for open water (Meyer, 2019; Tsyganskaya et al., 2018). Since the targets are offshore wind energy infrastructures within open water in windy places, the selection of VH polarisation is a first measure to increasing the contrast between the target and its natural environment. A median filter of multiple acquisitions is used to further filter and accentuate the offshore wind energy infrastructure signal. Therefore, all acquisitions within three months of both orbit directions are stacked and reduced to a median composite. Figure 3.2 provides an intuition of the acquisitions within three months and the

corresponding global median composite. Figure 3.2a) depicts the number of all acquisitions made in this period. It becomes clear that Europe has the best coverage, with both platforms acquiring data in all orbit directions, whereas other parts of the world are less frequently covered. However, due to the reduction to a median composite, all acquisitions at each pixel location are reduced to a single value, see figure 3.2b). From this, it becomes clear that the represented information in the median composite differs depending on the location and its corresponding acquisition frequency. For example, outside Europe, the occurrence of non-permanent objects due to fewer acquisitions is more probable, which has to be taken into account when processing the data later on.

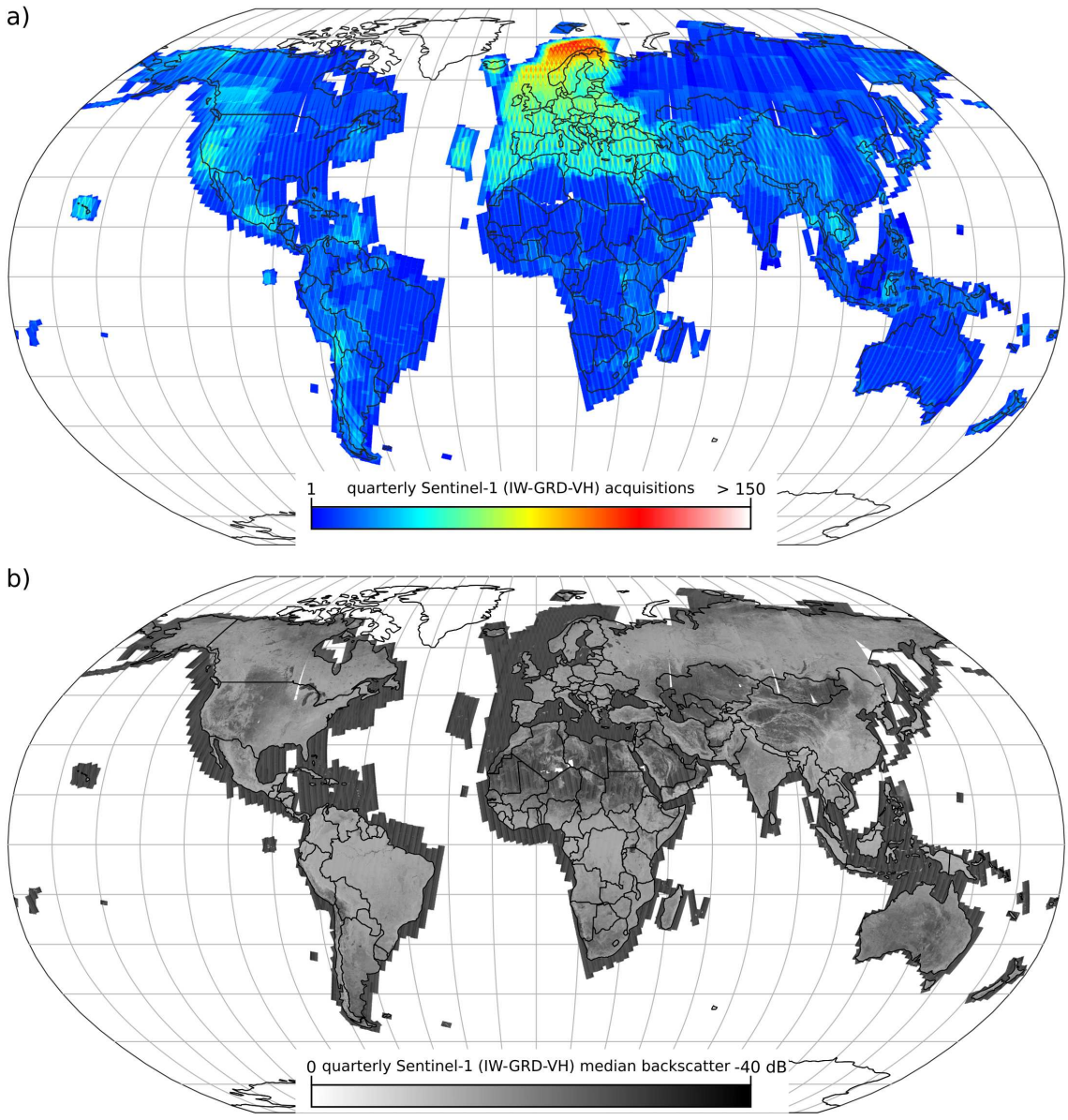



**Figure 3.2:** a) overview of the number of acquisitions of the Sentinel-1 mission for all acquisitions of both satellites and all orbit directions of the second quarter of 2021. b) the corresponding median composite for the vertical transmit horizontal receive (VH) band. Changed after Hoeser et al. (2022, p. 4255).

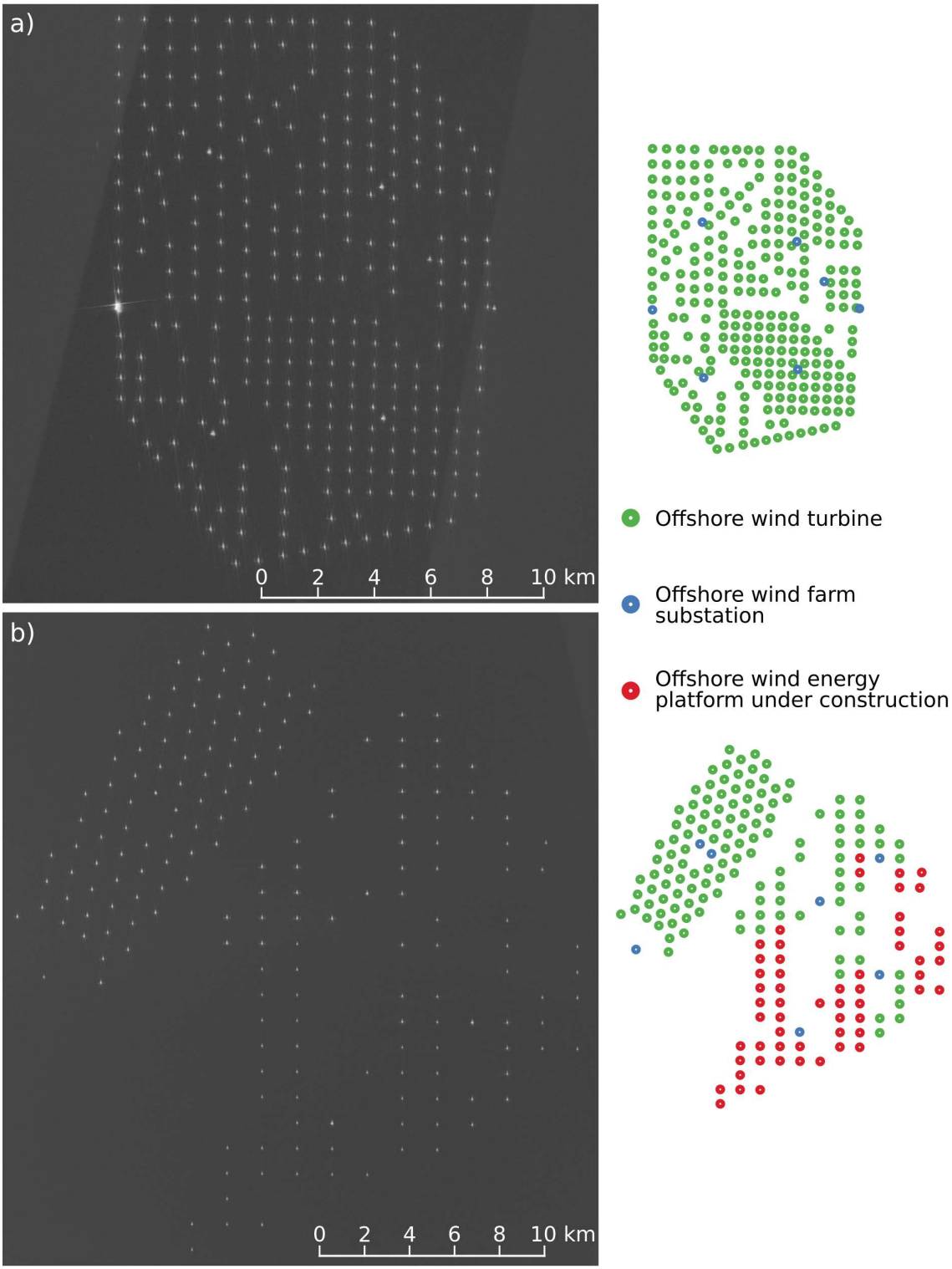
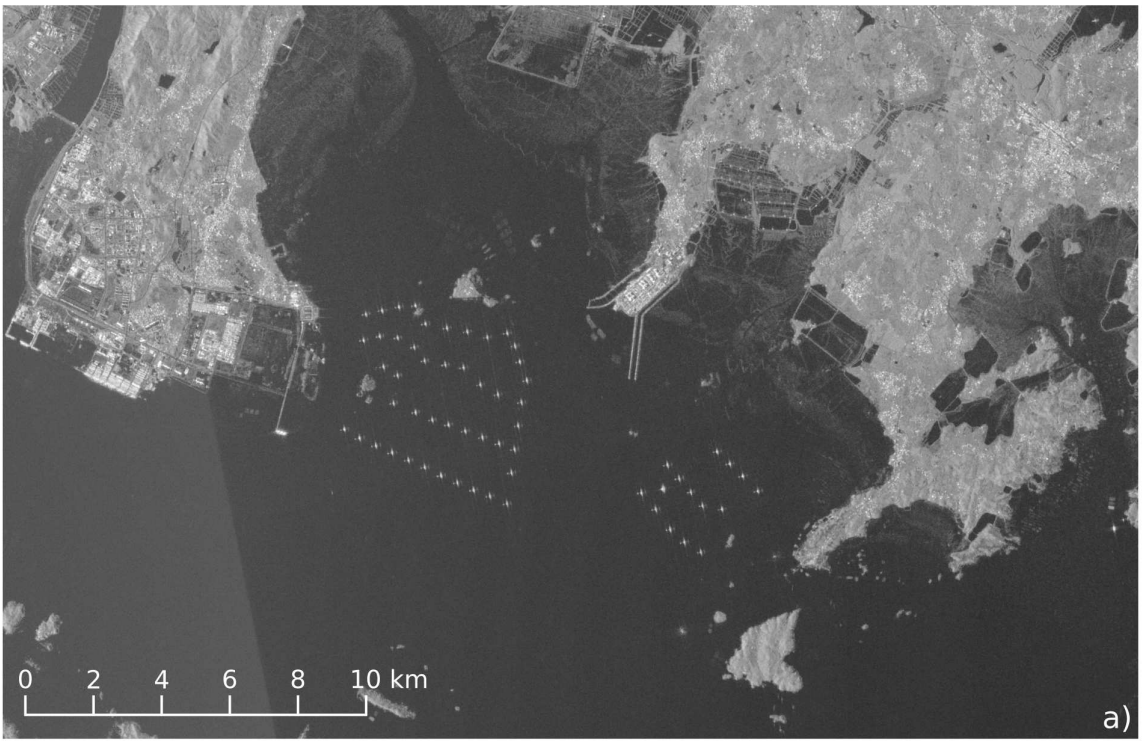



**Figure 3.3:** Examples for the large scale spatial features of offshore wind farms (OWFs) at a far off coast site. The grid-like structure of readily deployed OWF is clearly visible. a) shows an OWF site in the German Bight in the north of the island Borkum. b) shows an OWF cluster in the Moray Firth at the north-east coast of the Scottish Highlands.



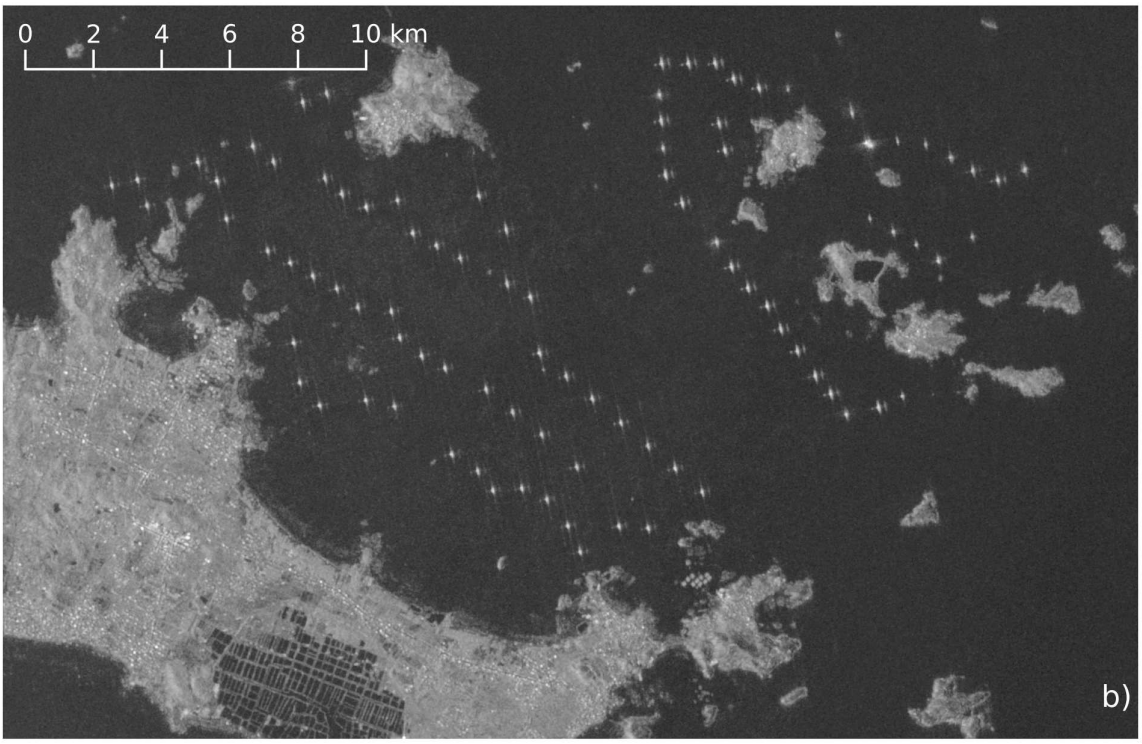




**Figure 3.4:** Examples for the large scale spatial features of offshore wind farms (OWFs) at near coast sites. The grid-like structure of OWF is still an important feature. Unstructured spatial features of natural environments and structured spatial features of human-made marine infrastructure appear at a close distance to the target OWF. a) and b) show Chinese OWF sites in the Taiwan Strait, a northern part of the South China Sea.

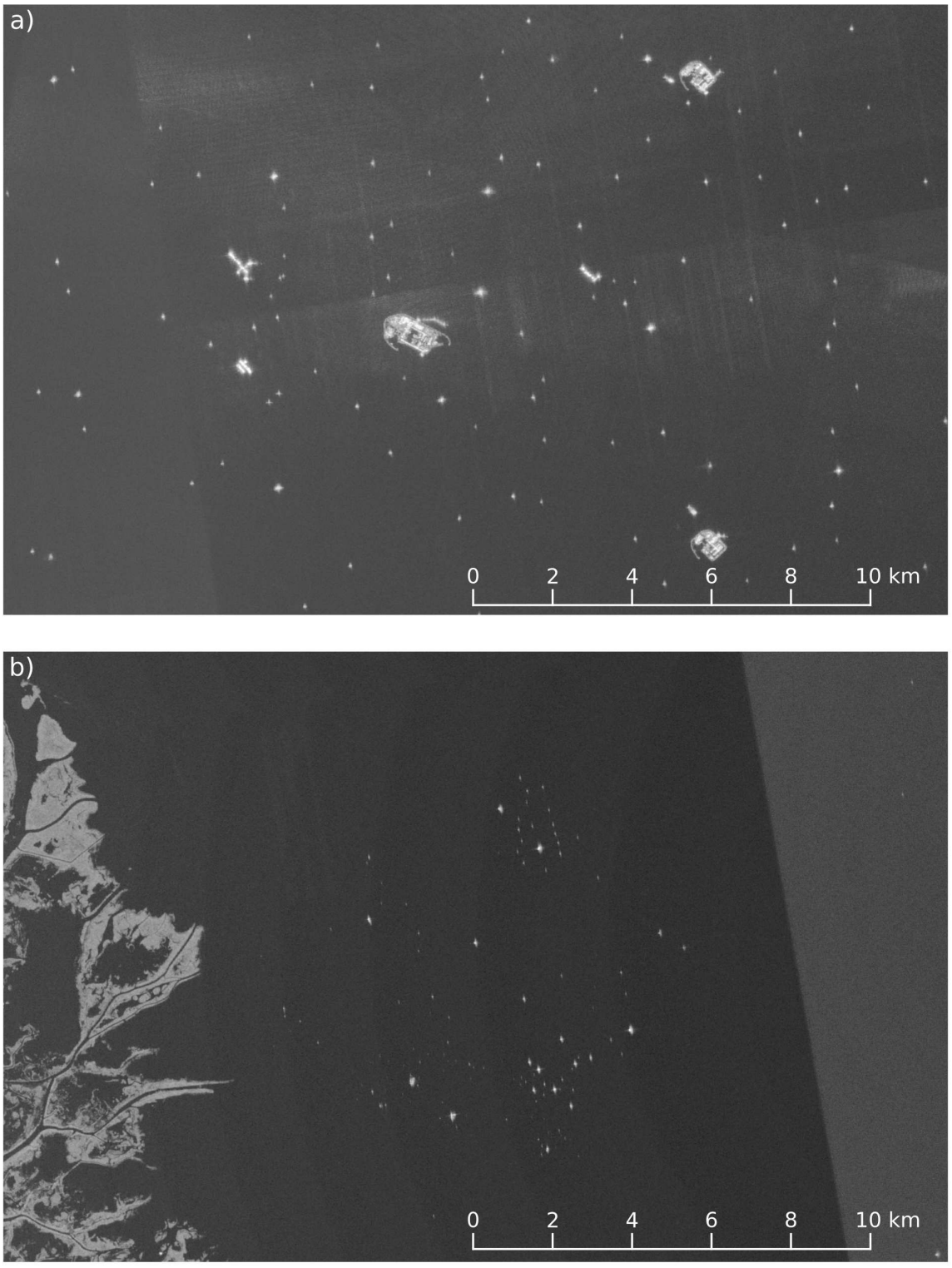



**Figure 3.5:** Non-target examples of marine infrastructure which on a small spatial scale have a similar appearance to offshore wind turbine but on a large spatial scale are unstructured and thus different compared to the typically grid-like structured offshore wind farms. a) shows oil rigs and industrial facilities in the Persian Gulf, b) shows oil rigs in the Gulf of Mexico, close to the Mississippi River delta which.

To get a first impression of how OWFs look like in an Sentinel-1 median composite, figure 3.3 shows two large OWF clusters in the North Sea Basin. Figure 3.3a), an OWF site in the German Bight in the north of the island Borkum, shows readily deployed OWTs as a well structured, grid-like pattern. Larger internal gaps distinguish different sections in this cluster, and some substations can be seen due to their more distinct backscatter signature and slightly off-grid position. Figure 3.3b) shows an OWF cluster in the Moray Firth at the north-east coast of the Scottish Highlands. Here the northern cluster is readily deployed, whereas the southern cluster has both OWTs under construction and such readily deployed. These two examples give the impression that detecting marine infrastructure is a rather straightforward task that only needs the detection of bright clusters in front of a darker background. However, the internal differentiation of different offshore wind energy infrastructure elements seems to be problematic. Furthermore, both examples show far offshore conditions with simple detection conditions, whereas near-shore conditions have a much more complex and challenging environment.

Coastal areas are compared to far offshore areas, highly complex, natural environments with river deltas, archipelagos and rugged coastlines. In addition, the natural morphology in these areas is heavily influenced by humans. Harbour infrastructure, industrial facilities, fishing and aquaculture facilities, bridges, buoys and lighthouses are only some of many features which can be found next to offshore wind energy infrastructure, which is also situated here. Figure 3.4 shows two examples of highly heterogeneous Chinese coastal areas with OWFs in the Taiwan Strait, a northern part of the South China Sea. It becomes clear that the task of detecting a single OWT under these conditions is much more complicated than at far offshore OWF sites due to the vast amount of potential false positive detections. Still, from a visual examination, the OWF components are distinguishable due to their structured patterns on a larger scale.

Finally, figure 3.5 provides further impressions of potential false positive detections. Figure 3.5a) shows oil rigs and refineries in the Persian Gulf and figure 3.5b) oil rigs in the Gulf of Mexico close to the Mississippi River delta. The small scale resemblance to offshore wind energy infrastructure is clearly visible, however on a larger scale, oil rigs appear less structured. This concludes the first impression of offshore wind energy infrastructure in Sentinel-1 radar imagery. A closer look will further provide insights into small-scale features specifically characteristic of offshore wind energy infrastructure.

Figure 3.6 provides an impression of different types of OWTs differentiated by common types of foundation construction. In general, gravity and monopile foundations are more often used in shallow water depths, whereas jacket and tripod foundations are used in deeper water depths (Wu et al., 2019). However, since most of the foundation is below

the water surface, their differences above the water surface are only slightly visible in radar imagery. More important is the size of the installed OWT. In figure 3.6 it can be seen that with increasing MW capacity, the turbine signature becomes larger, due to larger foundation constructions and transition platforms between foundation and wind turbine, but also the turbine itself. Additionally, the last example of the Alpha Ventus OWF in figure 3.6 provides a closer look at other platform types like transformer stations or meteorological masts. For them, it can be seen that their spatial signature differs slightly. These are promising observations since this difference in spatial features can later be utilised to internally differentiate these offshore wind energy infrastructure types.

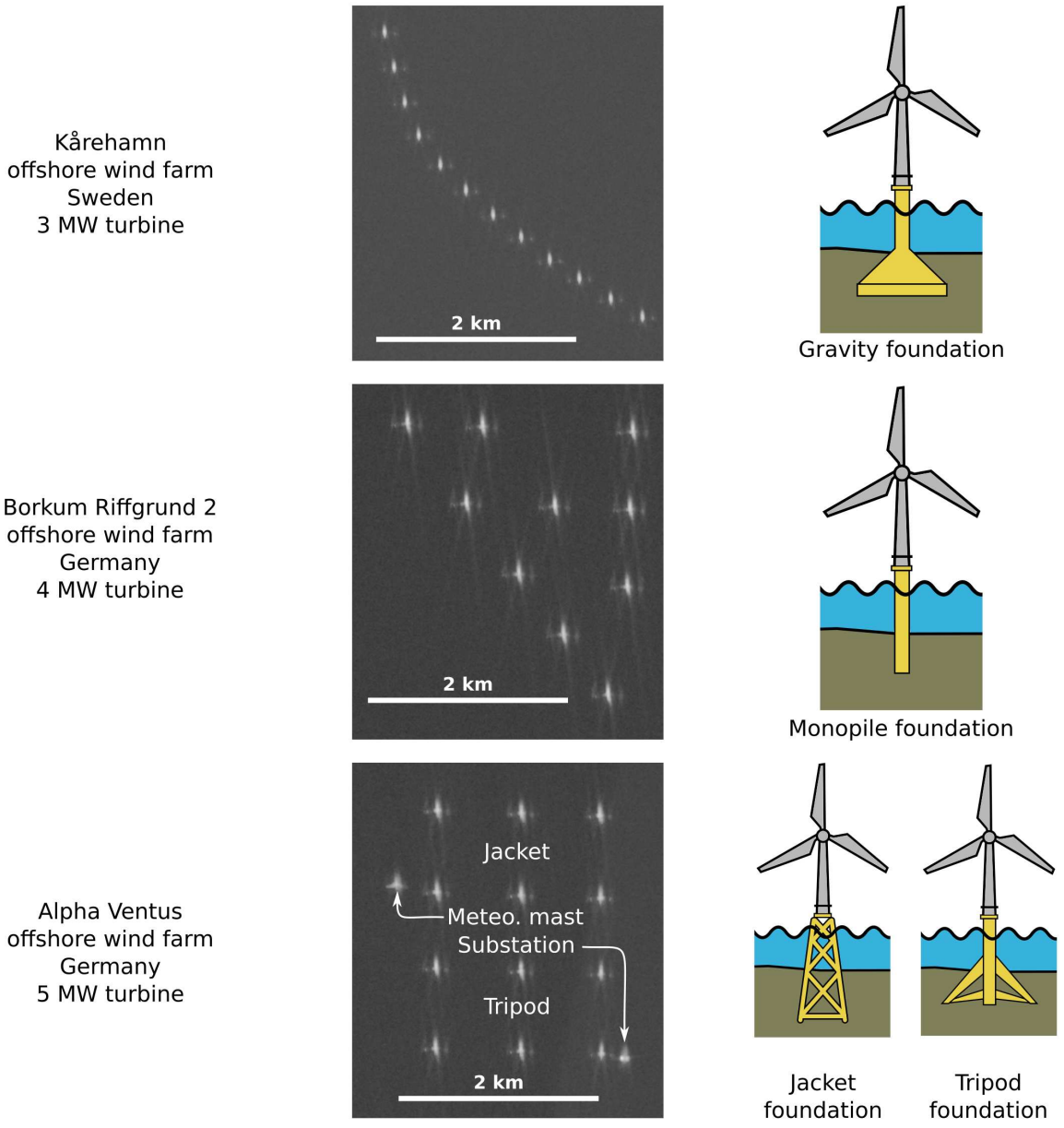



**Figure 3.6:** Different types of offshore wind turbine foundations which are directly built on the sea bed and their appearance on radar median composites. The Alpha Ventus example provides the signatures of a substation, and a meteorological mast.

Most OWTs by 2021 are built directly on the seabed. However, recent developments of platform types propose floating platforms that are only connected to the seabed by caissons, see figure 3.7. Two major advantages come from floating OWTs. First, the construction interferes less with the seabed and thus the marine ecosystem. Second, this installation type can be used in much deeper water depths, as the common foundation types discussed before (Stewart and Muskulus, 2016; Henderson and Witcher, 2010). Figure 3.7 clearly shows, that the radar signature of the floating spar variant is very similar to OWTs with conventional foundations. This is different from the semisubmersible type, which is less surprising since much of its platform is, as the name suggests, above the water surface. The metal floating body strongly interacts with the radar signal. This makes the radar signature appear wider than with other platform types.

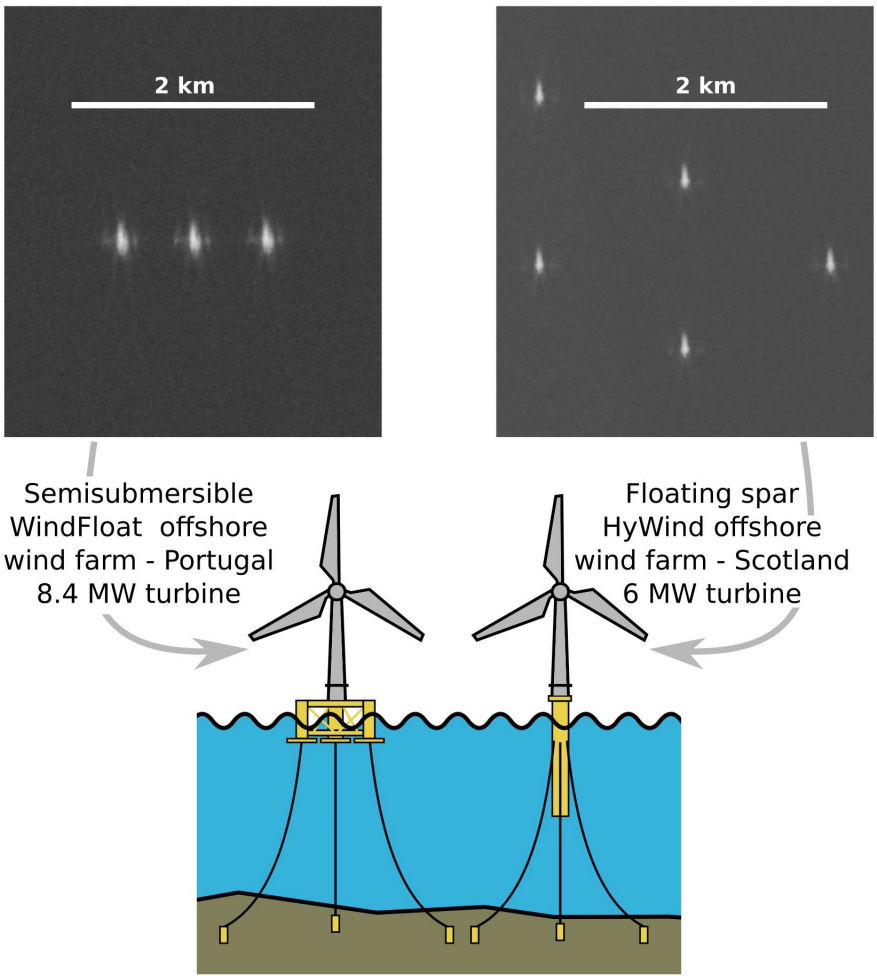



**Figure 3.7:** Different types of floating offshore wind turbine (OWT) and their appearance on radar median composites. Due to a larger floating body of the semisubmersible type, these OWTs appear broader on the radar median composites compared to the floating spar type.

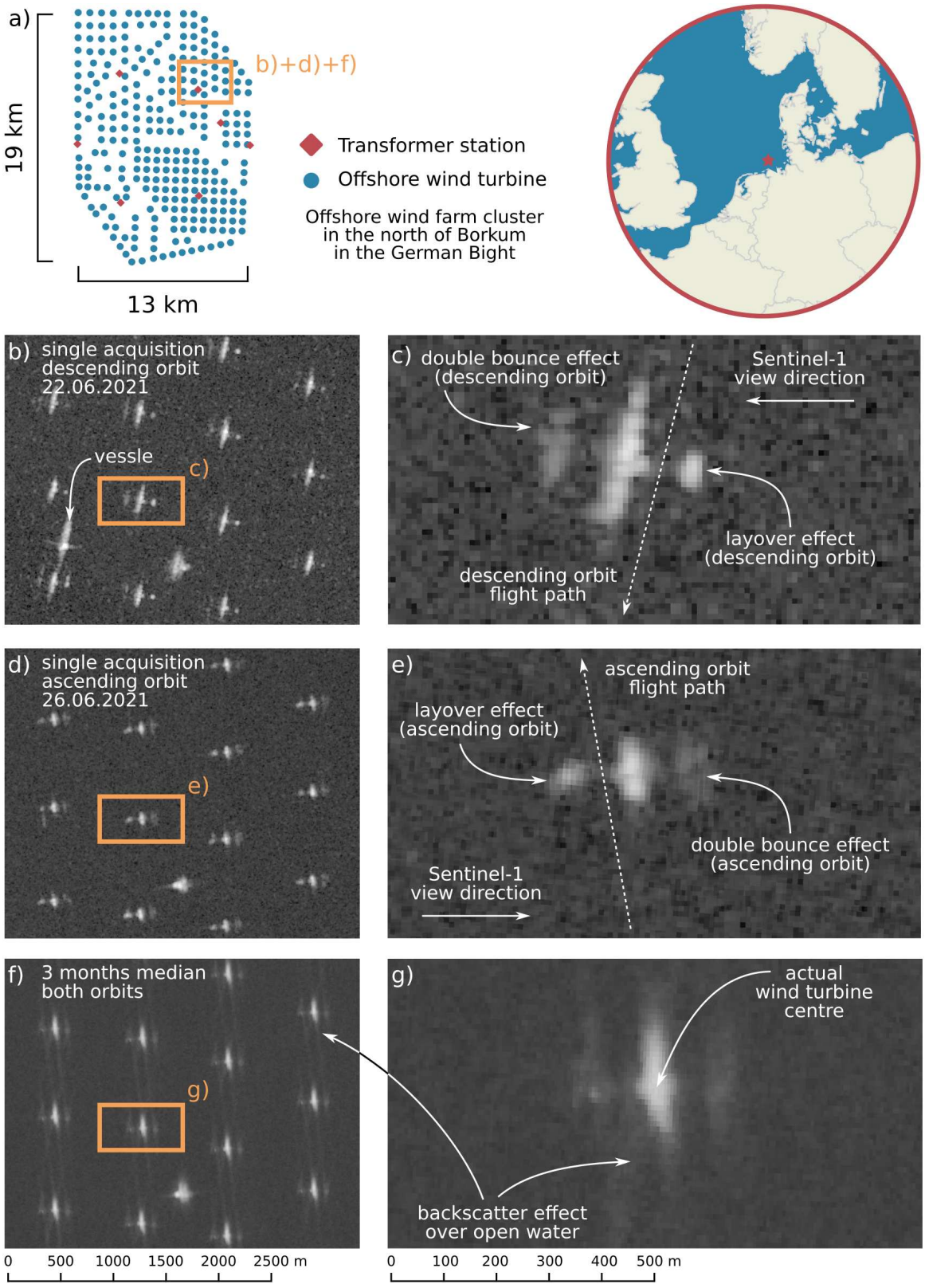



**Figure 3.8:** Examples of small scale spatial patterns of offshore wind turbine in radar imagery. b-d) show single acquisitions in ascending and descending orbit direction with the corresponding radar effects. f+g) is a three-month median composite (April to June 2021) of all orbit directions where the radar effects overlap. b) and f) also demonstrate how temporal objects like a vessel are filtered out in a median composite.

To understand the smallest spatial details of the radar signature of the object to be investigated, figure 3.8 provides an in-depth description for a single OWT and how it appears in radar imagery with different specifications. In figure 3.8, water surfaces have low backscatter values since the radar signal is specular reflected off the smooth water surface with an exit angle equal to the incident angle. Thus most of the signal is reflected away from the right looking sensor (Meyer, 2019; Ulaby and Long, 1991). In contrast to the water surface, offshore wind energy infrastructures appear as distinct backscatter clusters with high amplitudes. Here the radar signal hits the angular constructions of the different platform types and is directly reflected back to the sensor. Over open water, these strong scatterers have a distinct spatial pattern due to the imaging properties of the SAR instrument. In single acquisitions like in figure 3.8b-e), it looks as if the objects are rotated and stretched out in azimuth direction, due to the satellite's inclined orbit. In the median composites, depicted in figure 3.8f-g), this effect becomes visible on a larger scale, looking similar to the wakes of a moving ship. The fact that it is visible on median images proves that this effect is related to the sensor mechanics and a result of the interaction with this specific target, persistent offshore wind energy infrastructure, and its natural environment, the open sea.

One specific characteristic of the OWT's radar signature may already have been noticed when looking at the provided illustrations in figure 3.8. Two smaller clusters appear in the radar signature to the left and right of the bright cluster of a wind turbine. In the single acquisitions, shown in figure 3.8b-e), a bright cluster is visible, which lies in front of the turbine location, relative to the sensor view direction. This distinct cluster is the layover effect from the radar signal reflected at the turbine nacelle. Hereby the radar signal hits a steeply inclined object before its theoretical target at a smooth Earth ellipsoid is reached. When the radar signal is directly reflected from such an object, the signal's travel time is reduced. With it, the slant range is shorter than expected for the theoretical location on a smooth Earth's surface. When the signal is projected onto the ground range, a high backscatter amplitude appears in front of the real target position. In other words, it leans towards the sensor, see figure 3.1 for a theoretical ray path that explains the layover effect visually (Meric et al., 2009). The second less distinct cluster lies behind the turbine location relative to the sensor view direction. The double bounce effect causes this cluster. Here, the radar signal hits the water surface in front of the turbine and is reflected away as described above for a specular reflection. However, for a certain location, the signal hits the turbine as it goes on. The signal bounces off a second time, hence the term double bounce, reflecting towards the sensor. As the ray path is now longer, the resulting backscatter signal projected onto ground range appears behind the turbine. The same effect is also possible by a ray path where the signal first bounces off the turbine towards the water surface, where it bounces off a second time towards the sensor.

In the median image, both effects are visible but less distinct. When both orbit directions are used to generate the median composite, the stronger layover and weaker double bounce cluster are at both sides, cancelling some of the information of the opposing orbit direction. However, with a symmetrical appearance of the effects combined, a highly characteristic spatial pattern is available that describes a single OWT, see figure 3.8f-g).

## 3.3 Discussion

This in-depth investigation of the appearance of offshore wind energy infrastructure in radar remote sensing imagery provides the necessary information to set up the methodological study design and to decide on a specific CNN architecture. Furthermore, it supports compiling a precise training data set that includes the possible variability of different situations in which offshore wind energy infrastructure appears. That said, one crucial point has not been discussed yet: The availability of enough raw data to start the compilation of such a training data set. At this point, one encounters a massive problem when examining OWFs in remote sensing data. The number of OWF clusters in 2021 is about several hundred, depending on the way they are spatially differentiated. Even when each OWFs is used for training, the number of training examples would be too small to adjust the parameters of a CNN for object detection. Not even the repeated use of temporally staggered data solves this problem, as OWFs are persistent structures, so the variability in the training data would be small and overfitting very likely. Also, a training data set that already contains all the OWFs would make any further detection unnecessary. Overall, this seems like an unsolvable problem because the potential ability of a CNN optimised by supervised machine learning can not be unfolded for detecting offshore wind energy infrastructure.

In this section, characteristics of the target, offshore wind energy infrastructure, have been described on multiple spatial scales. In addition, the characteristics of non-targets, like oil rigs or heterogeneous coastal areas, have been discussed, too. Altogether, these entities express a specific spatial environment where offshore wind energy infrastructure appears. With expert knowledge about coastal environments, offshore wind energy infrastructure, and the radar signal of spaceborne sensors, the semantic description of such an environment is possible. Thus instead of a real image of an OWF this knowledge should be used to generate a virtual image. If this is possible, an infinite number of images becomes available. However, such a semantic description only exists in the expert's mind. Hence, it is necessary to bring it to a format a machine can understand. That way, features used in a human mind can be transferred into training data from which a machine can learn the underlying representation. If this knowledge transfer is possible, a machine should be able to detect offshore wind energy infrastructure within real remote sensing images.

# Chapter 4

## SyntEO Framework - Synthetic Data Generation for Earth Observation*

Synthetic data describes things that have not been recorded in the real world. They can either be entirely artificial generated information or data altered to a significant degree to have a new meaning. Synthetic training data sets are sets of synthetic data which have been automatically labelled during the generation process. Synthetic training data is strongly promoted by research coming from the computer vision and artificial intelligence domain and fields of engineering like robotics or autonomous driving (Nikolenko, 2021). Brodeur et al. (2017) and Martinez-Gonzalez et al. (2020) generated synthetic in-door settings with realistic physics from which an intelligent agent can learn how to interact with objects. Tremblay et al. (2018a,b); Hinterstoisser et al. (2019); Georgakis et al. (2017); Rajpura et al. (2017); Josifovski et al. (2018) and Nowruzi et al. (2019) generated realistic scenes with different objects and their specific context environment to train object detection algorithms in order to detect the same targets in real-world imagery. The studies of Ros et al. (2016); Richter et al. (2016, 2017) and Khan et al. (2019) used 3D engines to generate traffic scenes in urban environments, partly with changing weather conditions, to automatically generate large pixel-wise annotated data sets for image segmentation for training autonomous driving assistants. Shah et al. (2018) proposed AirSim, another 3D engine based virtual environment generator, which can be used to produce synthetic training data sets for autonomous driving but also for the navigation of UAVs with different sensor modalities. Further studies proposed how virtual environments can be used to train UAVs to navigate in different scenarios (Bonatti et al., 2020, 2019). These examples demonstrate a wide variety of possible applications and technical feasibility of the generation process itself and that a deep learning model can learn from synthetic data to perform a task in real-world settings.

*Parts of this chapter have been published in Hoeser and Kuenzer (2022b)

Despite an increasing interest in the Earth observation domain in deep learning and especially the CNN model for image analysis, a combination of synthetic training data and deep learning is rare. Studies from the Earth observation domain, which employ synthetic training data, are investigating very high-resolution RGB data and applications in the transportation and settlement categories of Earth observation. Similar to the earlier mentioned studies for generating data to train autonomous driving, Kong et al. (2020) generated 3D urban scenes to generate training data sets of overhead images for extracting building footprints. He et al. (2021) modelled scenes of ships in natural environments using the Unity game engine for the task of ship detection in very high-resolution RGB images. Furthermore, Han et al. (2017); Berkson et al. (2019); Shermeyer et al. (2021); Weber et al. (2021) proposed image generation processes to synthesize scenes that contain vehicles like cars or aircraft. All of these studies focus on very high-resolution RGB images and human-made object classes, which are rich in spatial features at the investigated spatial resolution. In contrast to these studies, Isikdogan et al. (2018) employed a synthetic approach to generate single band greyscale images of river networks in order to train an image segmentation encoder-decoder CNN to extract real river networks from Landsat 8 modified normalized difference water index (MNDWI) rasters.

All of these studies have in common that the synthetic training data is eventually used to optimise a CNN deep learning model to process real-world imagery. Thus, spatial features and context are essential to appear in the training data in order to optimise the CNN model type, which is able to learn such representations. Before synthesising remote sensing data, it is important to understand the basic properties of remote sensing imagery that allow spatial features to be represented. In order for spatial features to be recognised in image data, the spatial resolution of the capturing sensor must be **higher** than the size of the spatial feature itself. Without this capability, a single pixel value would mix the spatial expression of information at a **low** spatial resolution, and the spatial feature would not be available for image analysis. These properties are discussed by Strahler et al. (1986) as the H-resolution and L-resolution remote sensing scene model. In the following, the focus is put on the H-resolution remote sensing scene model, where targets of interest are described by multiple pixels to describe their internal spatial characteristics. Furthermore, in their fundamental work, Strahler et al. (1986) gave an important definition for a remote sensing scene: A remote sensing scene is “the spatial and temporal distribution of matter and energy fluxes [...]” (p. 122) which is “[...] not chaotic or random, but manifest spatial and temporal order” (p. 123). This is a crucial statement since it describes spatial context as a fundamental characteristic of a remote sensing scene. In the following, entities of a remote sensing scene are not divided into targets and background but into targets and non-targets. This subdivision is intended to express more strongly that non-targets in a remote

sensing scene are not arbitrary backgrounds but in a specific spatial context with the target and each other. Only with this in mind can synthetic remote sensing scenes be created that virtualise a realistic complexity of the spatial characteristics of a real remote sensing scene. This premise ultimately leads back to Tobler's first law of geography, that "everything is related to everything else, but near things are more related than distant things" (Tobler, 1970, p. 236). This rule has been discussed earlier from a CNN perspective, which are models that, when appropriately trained, are able to take the spatial context and with it this important condition into account. To train a CNN in such a way, these properties must be represented in the training data. Hence they must be considered when generating synthetic training data.

## 4.1 Theoretical Background of the SyntEO Framework

In section 3.2 the appearances of offshore wind energy infrastructure, other marine infrastructure and coastal environments in Sentinel-1 median composites have been discussed with a particular focus on their spatial features. Expert knowledge about the sensor, coastal systems, and infrastructure was combined to describe their properties semantically. The core idea of the SyntEO framework is to represent this knowledge in a structured way in order to make it accessible for an artificial data generator that produces synthetic remote sensing scenes along with task-specific annotation. In this section, the theoretical background of the framework is outlined. In order to make this explanation more intuitive, the offshore wind energy infrastructure example is used for clarification. Nevertheless, the framework is designed to be a general approach to generating synthetic remote sensing scenes upon structured expert knowledge.

Figure 4.1 compares a typical workflow of generating a training data set with the here proposed SyntEO workflow. For the typical training data set generation, depicted on the left side of figure 4.1, a limited set of remote sensing scenes is manually annotated by a domain expert. In case the complexity of the data allows it, this task can also be done by a group of domain amateurs who follow an annotation guideline formulated by the domain expert. However, both parties label the image data with task-specific labels as a bounding box around the target for object detection, resulting in a limited training data set with fixed variability. The annotation accuracy can thereby be assumed to be higher with a smaller variability if the annotation task is performed by a single domain expert and relatively lower with a higher variability when a group of domain amateurs does the annotation. In contrast to the typical training data set generation process, the SyntEO approach is depicted on the right side of figure 4.1. The SyntEO framework is based on a knowledge representation, the ontology, formulated by the domain expert. It is based on expert knowledge

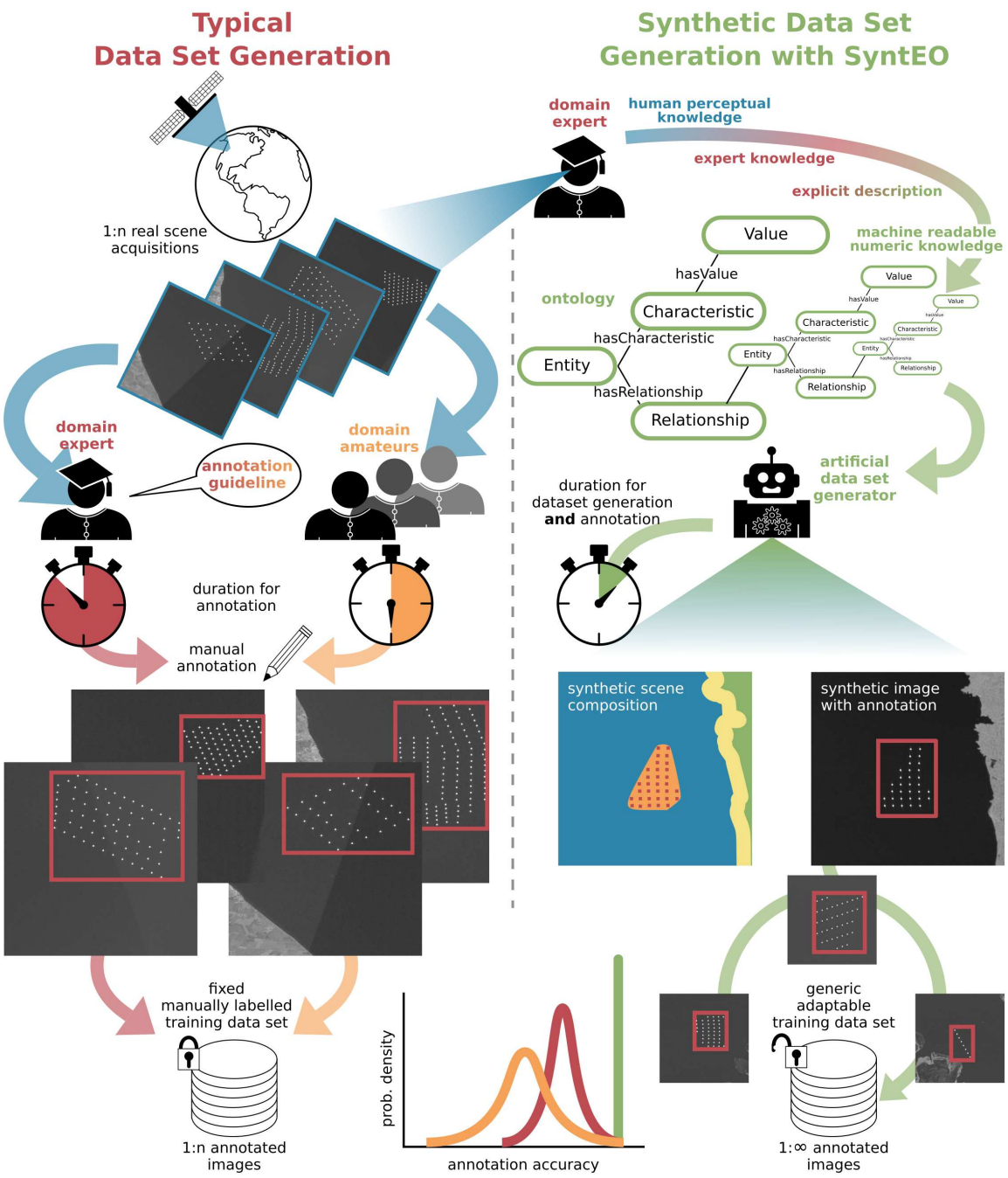



**Figure 4.1:** Comparison of a typical deep learning training data set generation process, and a training data set generation process with the SyntEO framework. Where in the typical process, one or multiple human annotators build a limited training data set, in the SyntEO process, an unlimited training data set is automatically generated by a machine upon expert knowledge made accessible by a human expert via an ontology. Changed after Hoeser and Kuenzer (2022b, p. 164).

and perceptual knowledge, coming from the observation of real-world imagery. Further on, an artificial data generator uses this structured knowledge representation to generate a remote sensing image synthetically. In a first step, this generator composes a discrete two or three-dimensional synthetic scene composition of target and non-target scene elements. The

sensor-specific texture is added to the scene composition to generate the synthetic image in a second step. The corresponding annotation for this synthetic image is derived from the scene elements of the discrete scene composition. Thereby the annotation type depends on the given task. By repeating the synthetic data generation process, a synthetic training data set with theoretically infinite size and variability is generated. Once the domain expert has formulated the ontology, the generation process of fully annotated training examples is fast, and label accuracy is stable and of maximal accuracy.

The main pillars of the SyntEO framework are ontology formulation and the synthetic data generator. Additionally, two technical building blocks are part of the entire workflow of the SyntEO framework, as pictured in figure 4.2. After the ontology formulation, technical preconditions for the synthetic data set have to be defined. These preconditions are the spatial resolution of the sensor to be simulated, and by taking the available GPU hardware and later trained CNN model specifications into account, the extents of the synthetic scene and the synthetic image. The dependencies of these parameters will be explained in detail in the upcoming introduction to the SyntEO framework. With this information provided, the synthetic data generator, which is a processing backend to produce images, starts synthesising training examples upon the knowledge represented in the ontology. In order to use the

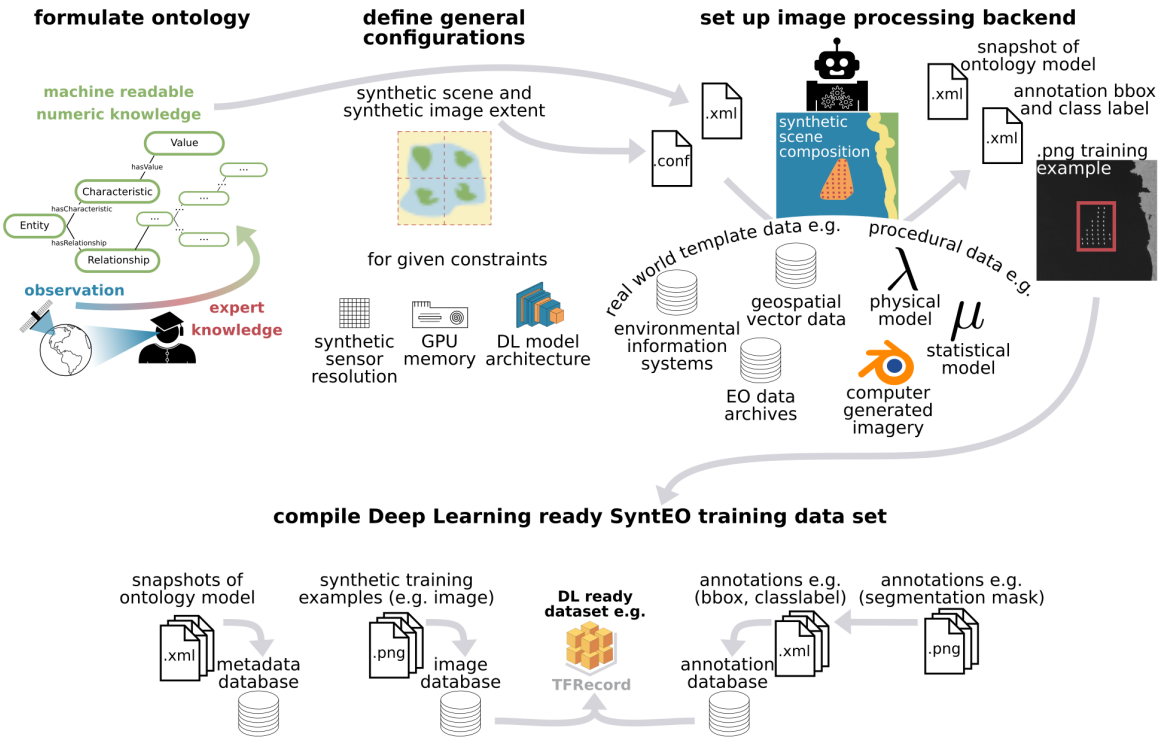



**Figure 4.2:** Overview of the SyntEO workflow with its two main components, the ontology formulation and image processing backend implementation and two subordinate steps that determine the technical conditions of the synthetic data and the composition of the training data set. Source: Hoeser and Kuenzer (2022b, p. 167).

generated images in a deep learning framework as training data, the last additional step is the generation of a deep learning ready data set. In this step, the generated image-annotation pairs are compiled to a balanced training data set by taking the metadata of each generated image into account. That way, the specific representation of the final training data set can be adjusted more precisely. Optionally, the image and annotation data files can be converted to deep learning framework-specific formats like the `.TFrecord` format optimised for the TensorFlow framework. The final result is a deep learning ready synthetic training data set for Earth observation.

### 4.1.1 The SyntEO Ontology

In artificial intelligence, knowledge engineering is to formally represent knowledge that enables an intelligent agent to make decisions or perform a task (Rich and Knight, 1991). The term ontology can be used to refer to such a formalised knowledge representation. Rich and Knight (1991) define the term ontology in two ways: “Ontology is the philosophical study of what exists. In the AI context, ontology is concerned with which categories we can usefully quantify over and how those categories relate to each other” (p. 292). Thus, in contrast to the global, philosophical perspective, the term ontology is used more narrowly to structure and formalise specific knowledge in knowledge engineering. With this, it becomes possible to design a narrow artificial intelligence using a formal ontology. More specifically, in SyntEO the term ontology is used as defined by Gruber (1995) who states that “an ontology is an explicit specification of [...] an abstract, simplified view of the world that we wish to represent for some purpose” (p. 908). The purpose of SyntEO is to generate a synthetic remote sensing scene. Thus the structure of the SyntEO ontology has to represent the necessary knowledge for image generation for specific remote sensing scenes.

By adapting the earlier cited definition by Strahler et al. (1986) of a real remote sensing scene, in SyntEO a synthetic remote sensing scene is “the spatial and temporal non-chaotic composition of scene elements which are an abstract representation of matter and energy fluxes. Hereby, scene elements can be two or three-dimensional targets and non-targets that describe the spatial location, size, distribution and shape of, e.g. land cover, land use classes, atmospheric conditions, natural or artificial landscape and objects” (Hoeser and Kuenzer, 2022b, p. 167). In order to guarantee a non-chaotic scene composition, the representation of knowledge about scene elements has to consider the spatial relationships among them. Furthermore, to support an information structure with increasing depth in details of complex elements to foster the appearance of multi-scale spatial features, the ontology should be organised as a nested, hierarchical representation of single entities which are related due to their common context (Wu, 1999).

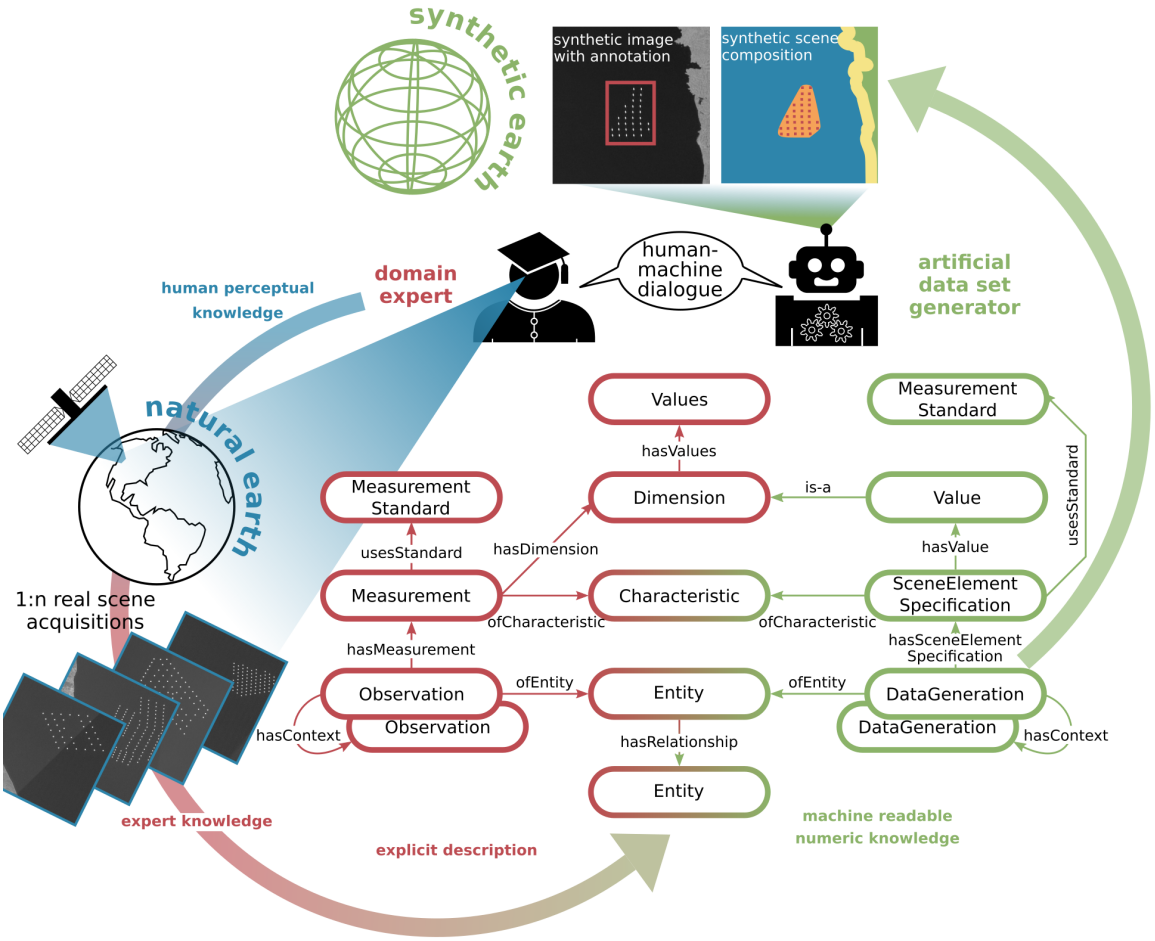



**Figure 4.3:** Conceptual overview of the SyntEO ontology building block. The red part has to be set up by a human expert, the green part is used by a machine to extract parameters to generate a synthetic scene composition and the corresponding synthetic image and annotation. The building block can be used to build interrelated nested knowledge representations about complex systems or short an ontology. Source: Hoeser and Kuenzer (2022b, p. 168).

In ecology, comparable challenges exist to capture ecosystems built of nested entities and shared context. Therefore, to scientifically capture an ecosystem, an ontology has been developed that provides such a description of nested entities as a basic structure. The Extensible Observation Ontology (OBOE) for representing scientific observations by Madin et al. (2007) is used as starting point in SyntEO and further developed for the task of representing knowledge to eventually synthesise Earth observation data. Figure 4.3 shows the building block of the adapted SyntEO ontology. The SyntEO ontology can be accessed from the perspective of a domain expert and of the artificial data generator. Hereby, it is a knowledge transfer platform between humans and machines. Observations by humans, the perceptual knowledge, combined with expert knowledge, becomes explicit by formalising semantic descriptions to explicit numeric knowledge represented in the ontology structure. A machine can interpret this structured knowledge in order to generate a synthetic scene.

To describe a scene element like an OWF, the domain expert makes an *Observation* of an *Entity*. This *Observation* has a *Measurement* of a specific *Characteristic*. This *Characteristic* can be the size of an OWF which is measured as the largest horizontal or vertical extent in kilometres. Since the final task is to generate a large data set with high variance, a single observation value is not sufficient. Instead, a *Measurement* has a *Dimension* of *Values*. This *Dimension* can be a range between two values, a statistical distribution, a list of values, a formula that describes a physical process under certain conditions, or a database of template data to provide some examples for a better intuition. For example, the *Dimension* of the *Characteristic* size of an OWF is a list of scaling factors where the list has three entries, small:5 km, medium:10 km and large:17 km. In order to describe other *Characteristics*, *Observations* can be connected with the *Context* attribution. For example, in addition to the size, an OWF is defined by its turbine density. By connecting both *Observations* with the *Context* attribution, the size can influence the *Dimension* of the turbine density values to express, that a large OWF has a low turbine density and a small-sized OWF has a high turbine density. Hereby, multi-scale spatial features are represented in a meaningful context. Finally, to guarantee a non-chaotic scene with spatial order, *Entities* can be connected via the *Relationship* attribution. For example, the *Entity* OWF has a *Relationship* of type topology with the *Entity* open sea. The associated topological rule states, that an OWF *must be inside* open sea. To further extent this example, the *Entity* open sea in turn has a *Relationship* of type topology with the *Entity* land. This topological rule states that the open sea *must not overlap* with the mainland. This nested structure allows relating OWF with land, even when there is no direct *Relation* attribution. The transitive relation between both entities states that an OWF *must not overlap* with the mainland.

This example gives an impression of the ontology from the perspective of the domain expert, who is embedding the knowledge. The opposing perspective of the data generator is similar, see figure 4.3. It starts with the *DataGeneration* of an *Entity*, like an OWF. Therefore it has to define a single *Value* for a *SceneElement Specification* of the *Characteristic* size. To obtain this *Value* it is randomly selected from the *Dimension* under consideration of any limitations coming from the *Context* attribution. That way, the declarative data generator randomly selects from the *Dimension* to define a single *Value*. Whereas the descriptive domain expert defines a variety of *Values* in this *Dimension*. By doing this, for each data generation step, the expressed characteristics have a very high variance in the final data set since many random processes are necessary to describe all scene elements in the synthetic scene composition.

### 4.1.2 Technical Configurations for Spatial Feature Representation

The ontology enables a formal representation of knowledge about multi-scale spatial features. In order to support their appearance in the final synthetic image to foster their influence on the training process of the CNN, some technical configurations have to be set. Figure 4.4 provides an impression of the configuration parameters and their potential conflicts that need to be balanced. Ideally, the generated multi-scale spatial features would appear in the final synthetic image as undistorted as possible and only influenced by the chosen sensor resolution. This condition, represented in the first row of figure 4.4, is given if there is no conflict about the chosen scene extent in metric units, the resulting image extent in pixel when sensor resolution is applied to the scene, and the available GPU memory in combination with the decided CNN architecture is large enough to train on the generated image. To make this more clear, the following example is to be assumed. The largest spatial feature of an OWF is the outer boundary which at maximal vertical and horizontal extent is roughly 20 km. At a pixel resolution of 10 m, this would lead to a synthetic image with the dimension of $2,000 \times 2,000$ pixels. Since $2^n$ is a typical value for input matrices in CNNs the image dimension is adjusted to $2,048 \times 2,048$ pixels. Following that, the synthetic scene has an extent of $20,480$ m $\times$ $20,480$ m. The CNN, which will be trained with the synthetic images, has an architecture design that ingests images with a dimension of $2,048 \times 2,048$ pixels. This architecture has a specific number of parameters and feature maps that together demand a sufficiently sized GPU memory during training. Suppose the CNN architecture becomes too large in the number of parameters and intermediate feature maps to store. In that case, there is the possibility that a small-sized GPU can not technically solve the optimisation process without running out of memory. However, if the GPU is large enough to ingest the synthetic images as derived from the synthetic scenes at full sensor resolution, there is no conflict, and no trade-off has to be balanced.

In a situation where the technical parameters change, like the GPU memory, limiting conditions are reached in which trade-offs have to be made. Thereby, the goal to keep in mind is to maximise spatial feature representation in the synthetic images in order to enable the CNN to learn the most important representation. To reach this goal, two options can be differentiated. One option is that the large scale features are the most important of the target to be visible in the final training example. However, due to the smaller GPU memory, now, the image can not be ingested in its full size of $2,048 \times 2,048$ pixels but have to be scaled down. In the example depicted in the second row of figure 4.4, a CNN architecture with an input size of $1,024 \times 1,024$ is chosen, which leads to smaller feature maps and with it, a GPU with a smaller amount of memory can be used. Since large scale spatial features are the most important in this example, the entire synthetic scene with an extent of

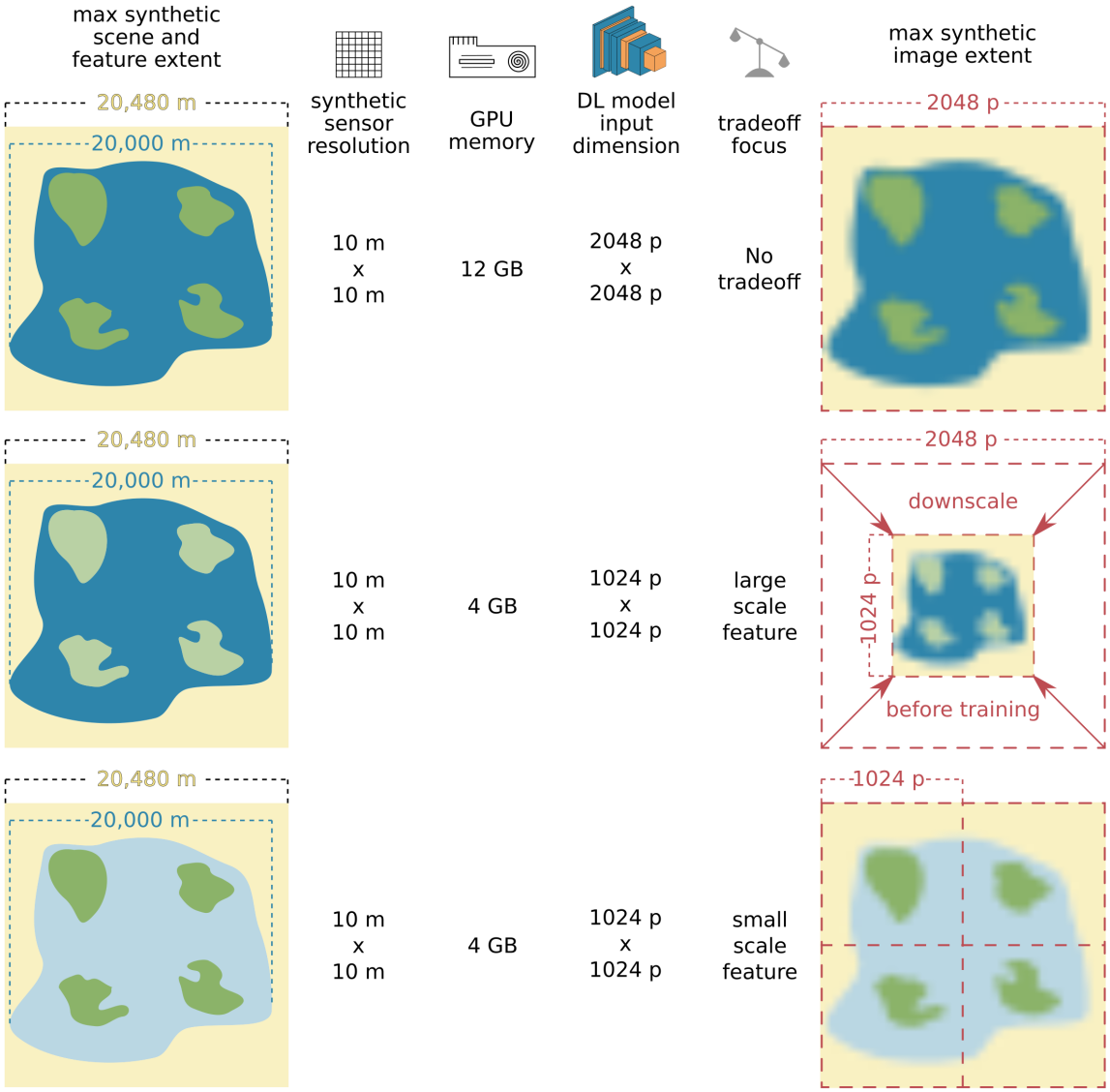



**Figure 4.4:** Three examples of how trade-offs between feature representation, sensor resolution, graphics processing unit (GPU) memory, and deep learning model architecture can be balanced. The first example has no trade-offs, and all features appear to their full extent in the synthetic image. The second example has to scale the synthetic image, to preserve large scale features at the cost of small scale features. The third example has to tile the synthetic image to show small scale features with all of their details at the cost of large scale features. Source: Hoeser and Kuenzer (2022b, p. 170).

$20,480\ \text{m} \times 20,480\ \text{m}$ is used, and an image of size $2,048 \times 2,048$ pixels is generated. This image has to be scaled down before training to the CNN ingestion size of $1,024 \times 1,024$ pixels. This leads to a virtual decrease of the spatial resolution to $20\ \text{m} \times 20\ \text{m}$. That way, large scale spatial features are preserved in the final training example. However, small-scale features, whose representation is more affected by downscaling, lose information content. This is the trade-off that has to be decided for.

Another option is to focus on small scale features in case they are the most important for the given task, as depicted in the third row in figure 4.4. Instead of downscaling the image, which leads to decreasing representation of small scale details, four images of size $1,024 \times 1,024$ are derived from the entire synthetic scene. That way, the virtual sensor resolution of 10 m $\times$ 10 m is preserved in the final image. However, large-scale features no longer appear to their full extent in the training examples. Nevertheless, the spatial context between small and large scale features is still meaningful since the synthetic scene was designed to represent both at full scale, and the images were taken from this composition. This means the ontology does not have to change since nothing changes in the spatial and contextual relationships among the scene elements. Only the technical configurations for the image generator change. This allows using the same ontology to generate suitable training data sets under changing determining technical conditions.

### 4.1.3 Artificial Data Generator

The artificial data generator is the processing backend, which generates the synthetic scene composition and image. Therefore, it ingests the ontology to build scene elements and compose them into a meaningful synthetic scene and, finally synthetic image. The actual implementation of this processing backend depends on what the domain expert decides to use or develops to generate the necessary shapes and textures technically. In the upcoming sections, some examples of image processing backends are proposed. They can be as simple as binary arrays or as complex as a three-dimensional virtual environment. Independent of the processing backend, the machine perspective of the ontology has to provide combinations of values so that the processing backend can use them. Thus, the machine-readable translation of the ontology can be seen as a large parameter file. A suitable data format for this parameter file is the `.xml` or `.json` format since both can represent the hierarchic and nested structure of the ontology and are still readable by a human interpreter.

Two main types of synthetic data can be differentiated in the SyntEO framework for the data generation process. Procedural data are entirely artificially generated, whereas template data are data that already exist but are remixed so that they appear in a new context. For procedural data, practical implementations and engines from the computer generated imagery (CGI), gaming industry, and virtual reality domains can be adapted (He et al., 2021). Especially for representing ecosystems with less structured patterns, fundamental works on generating gradient noise of Perlin (1985) and Perlin (2001) are excellent examples of how to technically implement representations of organic or natural patterns described in the ontology. Opposing this artistic approach for a processing backend is the direct use of the laws of physics, statistical distributions or libraries of scientific observations to generate

data. One example of how radar signatures with a very high resolution can be generated is the RaySAR tool (Auer et al., 2016). Another example of a database of scientific observations to generate multispectral texture is the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) spectral library by Baldridge et al. (2009).

In contrast to procedural data, template data is already existing information that can be used as scene element geometries or textures. For example, the Open Street Map (OSM) project contains many already structured geometries which can be remixed to generate an entirely new remote sensing scene composition. By adding remote sensing data to these shapes, the sensor-specific texture of radar or multispectral instruments can be extracted from freely accessible Earth observation archives in order to synthesise natural or human-made environments.

An example processing backend should bring some light to this rather abstract description of what such a backend could look like. For the most illustrative example possible, the example of offshore wind energy infrastructure in radar data will be briefly put on hold, and the generation of high-resolution RGB data for the detection of individual trees will be considered. As stated earlier, modern 3D engines provide very complex processing backends to generate synthetic images. An open-source 3D engine is Blender (Blender Online Community, 2018), which comes along with a python application programming interface (API) that allows passing parameter information in order to generate complex scenes automatically. Figure 4.5 shows an example of a scene and the derived synthetic image produced with Blender. In a first step, Perlin noise (Perlin, 1985) was used to generate a 3D terrain model to initialise the synthetic scene. The generated noise values are interpreted as height information to build the terrain. From this synthetic terrain, properties like the slope, aspect, or terrain pointiness can be derived, which are used as control parameters for further scene and texture generation. To add a basic texture to the terrain, the high-resolution optical IKONOS archive was searched for homogeneous areas from which subsets of unvegetated and vegetated areas are extracted. Together with the terrain pointiness, the two IKONOS images are passed into a shader, producing an entirely new texture. Since tree detection is the task, trees have to be placed onto the generated terrain. Therefore 3D tree models from open and proprietary libraries can be acquired or modelled by hand within Blender. With at least a single tree model, a particle system can be used for its placement. The particle system allows placing objects across a given scene randomly. It also controls randomisations like the type, rotation, scale or inclination of the objects to be placed. Furthermore, the particle system can be combined with the terrain properties via a weight map. By thresholding the terrain properties like the height and slope, a weight map can be generated, which controls the density of objects to occur at a specific location. Altogether, the 3D tree models are then

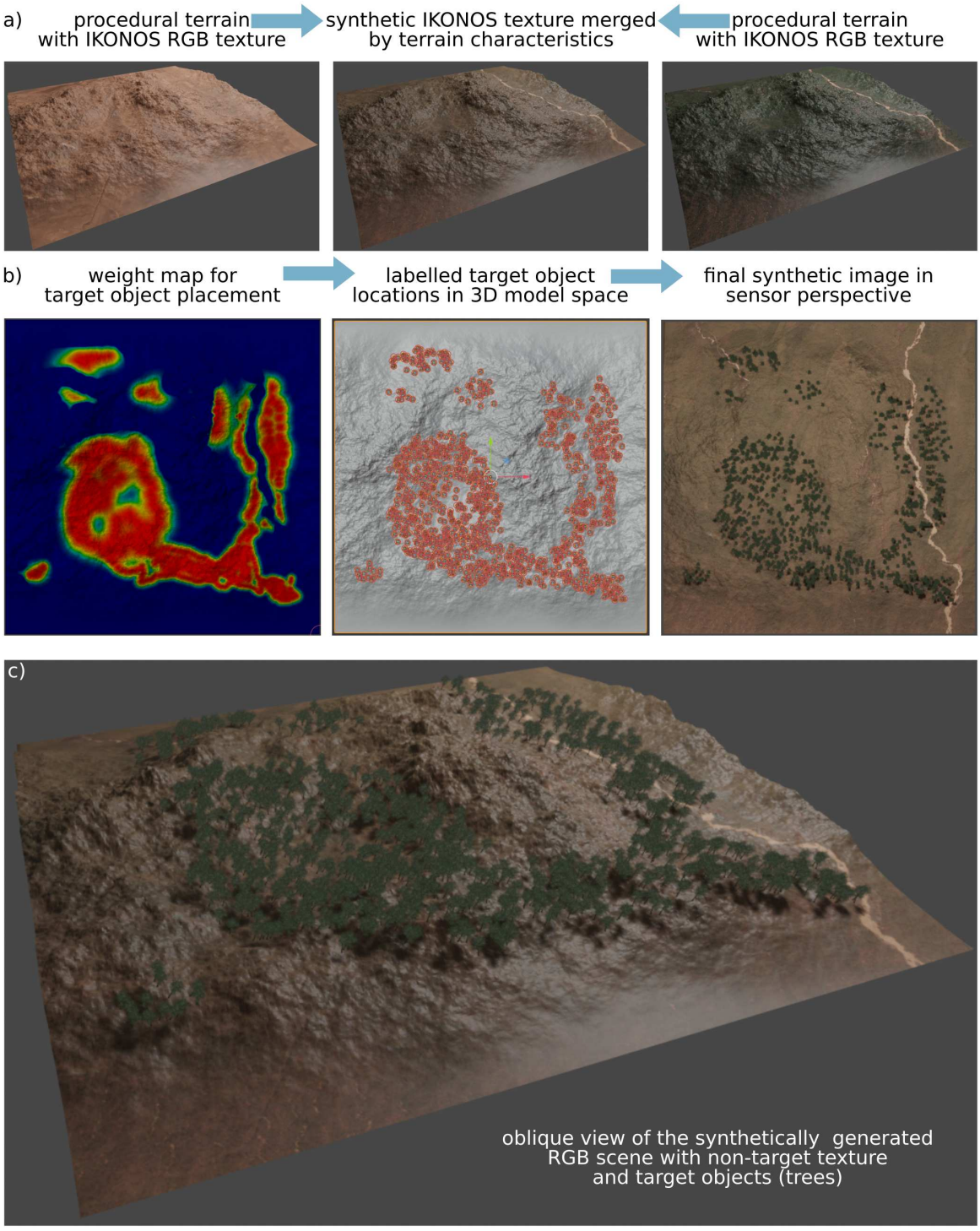



**Figure 4.5:** Generation of a synthetic RGB image with the 3D engine Blender. a) shows how non-target texture is generated by fusing IKONOS template data by taking synthetic terrain properties into account. b) shows how tree targets are getting distributed over the synthetic terrain and the final synthetic image from an overhead sensor perspective and less specular lightning. c) shows the generated synthetic environment from an oblique view for a better impression of the 3D model space. Source: Hoeser and Kuenzer (2022b, p. 171).

randomly placed over the scene. Since everything happens in a 3D model space, the location of each tree is known, and labels can be derived for each object placed by the particle system. Further on, the 3D model space allows changing the viewing angle, the camera position, and lighting conditions. By selecting an overhead camera position and less specular light as for the oblique views in the 3D model space, a synthetic remote sensing image can be taken from the modelled scene.

Since all employed modules of the Blender software can be controlled by the python API, an automated generation of these images is possible. A processing backend can be implemented with a suitable parameter file coming from an ontology that describes these environments, which supports highly complex scene generation in a modular way.

### 4.1.4 Training Data Set Compilation

One major motivation of SyntEO is to generate large training data sets, fast and with a high variance of training examples. With the previously described knowledge representation and a flexible data generation backend, a large amount of training data can be produced. However, in the end, only a large amount of data, even if annotated, is a start but not optimal for deep learning. An optimal deep learning data set must be balanced and ready to use in modern deep learning frameworks.

A proper data set balancing allows the underlying machine learning mechanism to learn representations equally. For example, if the data generator produced ten thousand images of small wind farms close to the coast but only one hundred images of large wind farms far off the coast, an unbalanced data set has been generated. Due to the unbalanced representation of examples, a trained object detector CNN on this data set would learn that OWFs usually appear near coasts and are relatively small. In a worst-case scenario, the amount of large off coast OWFs would be so small that the optimisation signal from this group of training examples vanishes during the machine learning process and large off coast OWFs are getting wrongly rejected by the object detector. Thus representation in the ontology is not enough. The final balancing of different training examples in the training data set has to be representative, too.

The introduced components of the SyntEO framework already offer everything necessary to balance the data set. In each image processing step, the data generator approaches the ontology and defines specific values in order to produce a synthetic remote sensing scene. Each entity in the ontology provides metadata that can be exported along with each image-annotation pair. That way, the large amount of data can be structured by using this metadata to set up a data database from which a balanced data set can be built. Due to this

approach, each data set gets a metadata structure, which helps build stable and fully controllable experiment environments. With stable experiment environments, it becomes possible to observe how different types of data sets affect the CNN performance. For instance, to investigate what happens when the same CNN is trained with a data set containing only small OWFs instead of all sizes of OWFs.

Finally, a last optional step can be made when the data set is balanced. Especially in deep learning, specific data formats became popular since they allow very sophisticated handling of modern deep learning frameworks and their sub APIs. One good example is the `.TFrecord` format of the TensorFlow deep learning framework for its data API. Some benefits of this API are that it supports sophisticated data augmentation on the fly during training and cares about an optimised data pipeline from storage to GPU to avoid idling hardware and thus speed up training. The `.TFrecord` format is also commonly used in other APIs, like the TensorFlow object detection API. Thus, converting image and annotation files to this specific format enables higher usability for a specific deep learning framework.

## 4.2 Proof of Concept - SyntEO for Offshore Wind Farm Detection

After a theoretical introduction with brief examples to clarify the otherwise only technical terms, a first end-to-end application of the SyntEO framework for OWF detection is performed. This detection is still part of the development process of the SyntEO framework and not the final global offshore wind energy infrastructure detection, which will be presented in the next chapter 5. The upcoming investigation is a proof of concept study which tests the practicability of the proposed SyntEO framework and looks more deeply into the single steps of the introduced workflow. Therefore, four synthetic data sets for the task of OWF detection will be produced. With each of the four data sets, a CNN object detector is trained and tested on four test sites.

### 4.2.1 Test Sites and Data

The four test sites in this proof of concept study are two hot spots of offshore wind energy production and two areas without any OWFs but that show potential false positives, see figure 4.6a). The two wind energy production hotspots are the North Sea Basin and the East China Sea. Figures 4.6b) and c) show readily deployed OWFs at far off coast locations in the North Sea Basin and a near-shore OWF cluster in the East China Sea with wind turbines under construction. Unlike the North Sea Basin, in the East China Sea, OWFs

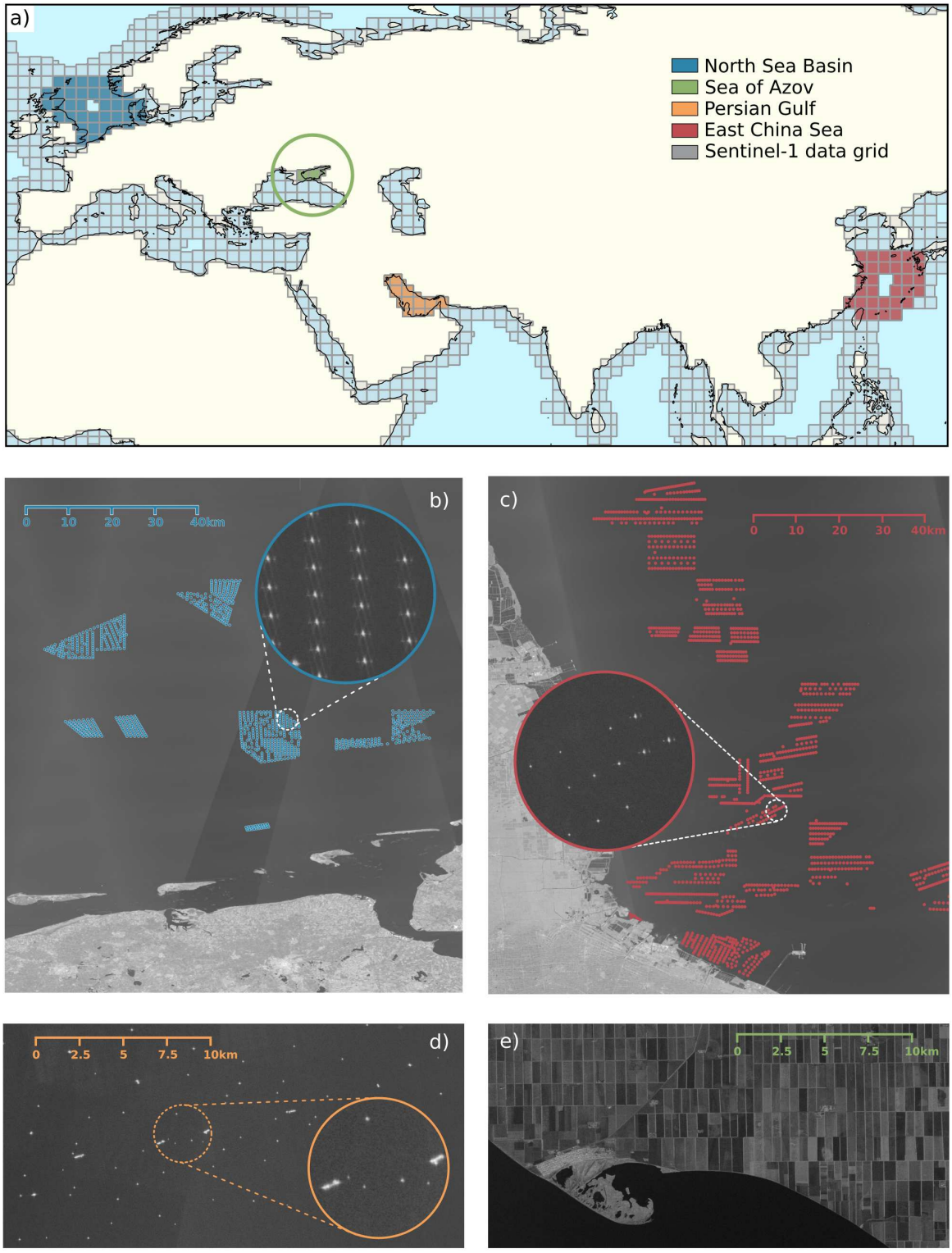



**Figure 4.6:** a) Locations of four test sites and the 1.8° data grid of the global coastline for which Sentinel-1 median composites are generated. b) example of the North Sea Basin test site in the German Bight. c) near coast offshore wind farm in the East China Sea, Jiangsu, China. d) oil rigs in the Persian Gulf near Khaji, Saudi Arabia, and c) agriculture landscape and road network at a coastal site of the Sea of Azov, near Yeysk, Russia. Source: Hoeser and Kuenzer (2022b, p. 174).

are built more closely to the shoreline and in areas with a strong tidal influence. The near coast locations of OWFs in the East China Sea are the more challenging environments for offshore wind energy infrastructure detection since OWFs are built closely to other infrastructures like harbours, buoys and aquaculture, and other natural non-targets with similar spatial signatures like small islands. Furthermore, due to OWFs on tidal flats, the typical radar signature in median images is less pronounced due to acquisitions in periods of low tides with no or less water. Due to the acquisition schedule of the Sentinel-1 mission, the North Sea Basin has more acquisitions of both orbit directions, whereas the East China Sea has fewer acquisitions in the same period with mostly one orbit direction, see figure 3.2. Thus in the East China Sea, the typical spatial features of the radar signature are further dampened by a lower signal to noise ratio due to the smaller number of acquisitions used in the stack to produce the median composite. Also, the offshore infrastructure itself differs since in the test period, the third quarter in 2020, the East China Sea has much more OWFs under construction compared to the North Sea Basin. For these turbines, only the foundations return a backscatter signal. The typical clusters to the left and right of the turbine, originating from the layover and the double bounce effect of the turbine's nacelle, are missing. The two magnifications in figures 4.6b) and c) demonstrate the differences between the signatures in the North Sea Basin and the East China Sea. Overall, these make the East China Sea a more difficult test site than the North Sea Basin.

The other two test sites are the Persian Gulf and the Sea of Azov, see figures 4.6d) and e) respectively. They only show non-targets that are potential false positive detections due to their spatial characteristics. The Persian Gulf is characterised by oil rigs and offshore refineries, which present potential false positives due to their similar backscatter signature on a small spatial scale. The onshore coastal environment in the Sea of Azov is characterised by large rectangular agricultural fields which are separated by a grid-like road network. This grid-like pattern on a larger spatial scale is this test site's potential false positive feature. Due to their characteristics, both sites are used as real-world examples to test the abstraction ability of differently trained CNNs to investigate not only the accuracy of the models but also their precision.

The ground truth data set of real-world OWFs were generated by visually examining Sentinel-1 median composites of the third quarter in 2020 (2020Q3). In addition to the Sentinel-1 median composites, high-resolution RGB imagery in Google Earth and Sentinel-2, as well as public planning documents and reports about OWF projects, were used to validate OWT locations. All readily deployed OWTs within the four test sites were annotated and aggregated to OWF clusters by their spatial proximity. The bounding boxes of each OWF cluster are the ground truth geometries for the North Sea Basin and the East China

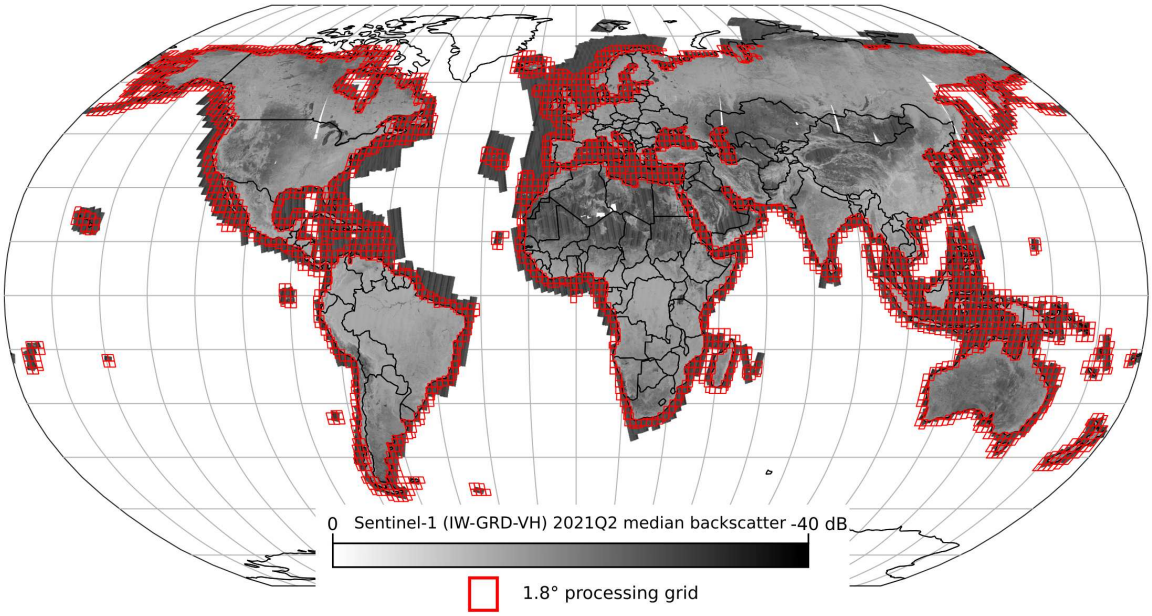



**Figure 4.7:** Overview of the entire global coastal 1.8° data grid for a 200 km are from the coastline towards the sea. The background layer is a three month Sentinel-1 median composite, which is processed and downloaded for the data grid boundary. Changed after Hoeser et al. (2022, p. 4255).

Sea. Since the two other sites, the Persian Gulf and the Sea of Azov, contain no OWFs, their ground truth information is specifically empty to assess false positive detections. That way, four real-world test sites with different characteristics are defined to assess the performance of an object detector that is optimised by synthetic training data.

Sentinel-1 median composites were acquired not only for the four test sites but also for the entire global coastline. Data that did not belong to the four test sites were used as template texture during the image generation process of the synthetic training data. For efficient data acquisition and processing, a processing grid with a cell width of 1.8° was set up, which is also the technical boundary of the four test sites, see figure 4.6a). The processing grid is initialised to cover the entire globe. Since only such grid cells are of interest which intersect an area reaching 200 km from the shoreline into the open sea, a polygon covering this area was generated to make a spatial selection of the grid cells. The selected global processing grid by this polygon is depicted in figure 4.7.

The acquired Sentinel-1 median composites contain all VH-polarised Sentinel-1 IW GRD acquisitions of the third quarter of 2020. The GRD products of the Sentinel-1 archive are freely accessible via the Google Earth Engine (GEE) (Gorelick et al., 2017). GEE has the major advantage that it provides the GRD products with necessary preprocessing already completed. The GRD products on the GEE were adjusted to their precise orbit information, border noise and thermal noise have been removed, and radiometric and terrain correction have been applied. Furthermore, the GEE allows sophisticated spatiotemporal

querying of the data archive and provides an option to build custom processing graphs, which can all be run on the GEE processing environment. That way, the quarterly median composites can be generated directly on the GEE in a highly efficient way. However, the median composite generation and the export of the results is still a processing heavy task for larger regions, and GEE limitations prevent a direct download of a single raster which would be necessary to describe the global coastline at a spatial resolution of 10 m $\times$ 10 m. To solve this problem, the previously generated processing grid is used to enable parallel data acquisition of adequately sized image tiles, as depicted in figure 4.8.

To further organise the processing grid, each grid cell is initialised with an identifier (ID), in the following referred to as grid ID. Based on this grid ID, a folder structure is set up which will hold all files which are associated with the corresponding grid ID. To start the acqusition process, a python script uses an automatic authentication to get access to the used external services, the GEE and also the later used Google Cloud Platform. By using the GEE-python API, a single grid ID geometry, the start and end date of the period of interest, the Sentinel-1 data product specification and the processing graph for the median composites are passed to the GEE, to generate a single median composite tile for the corresponding grid cell. This median composite is further reduced from 16 bit floating-point, which provides the backscatter amplitude in dB, to 8 bit integers. Thereby, the range of -40 to 0 dB is rescaled to 0 to 255. The loss of decimal places during this rescaling is acceptable since the CNN, which is used later on for object detection, is more sensitive for spatial features than for small changes of single pixel values. The resulting image is then exported to a storage bucket on the Google Cloud Platform. This is necessary, due to the design and quotas of the GEE. While the task is executed by the GEE, the python script, which started the process, continuously queries the processing stages and checks if the exported data from the GEE is available on the Google Cloud Platform storage. That way, exported data is downloaded from the Google Cloud Platform storage to a local processing machine immediately after the export from the GEE was successful, and non of the processed median composites remains on Google services.

In order to download the data for the entire globe, this process has to be repeated thousands of times, where one process can take up to 20 minutes, depending on the GEE performance. To reduce the data acquisition time, the python script to process a single grid ID is called ten times in parallel. Thus, the acquisition of a single band median composite with a pixel size of 10 m $\times$ 10 m for the entire global coastline is possible within three to five days. To organise the parallel download process for all grid cells, a SQLite database is set up with the grid IDs that manages the feedback information from both Google services, which is queried along the workflow. That way, the database distributes open grid cells to

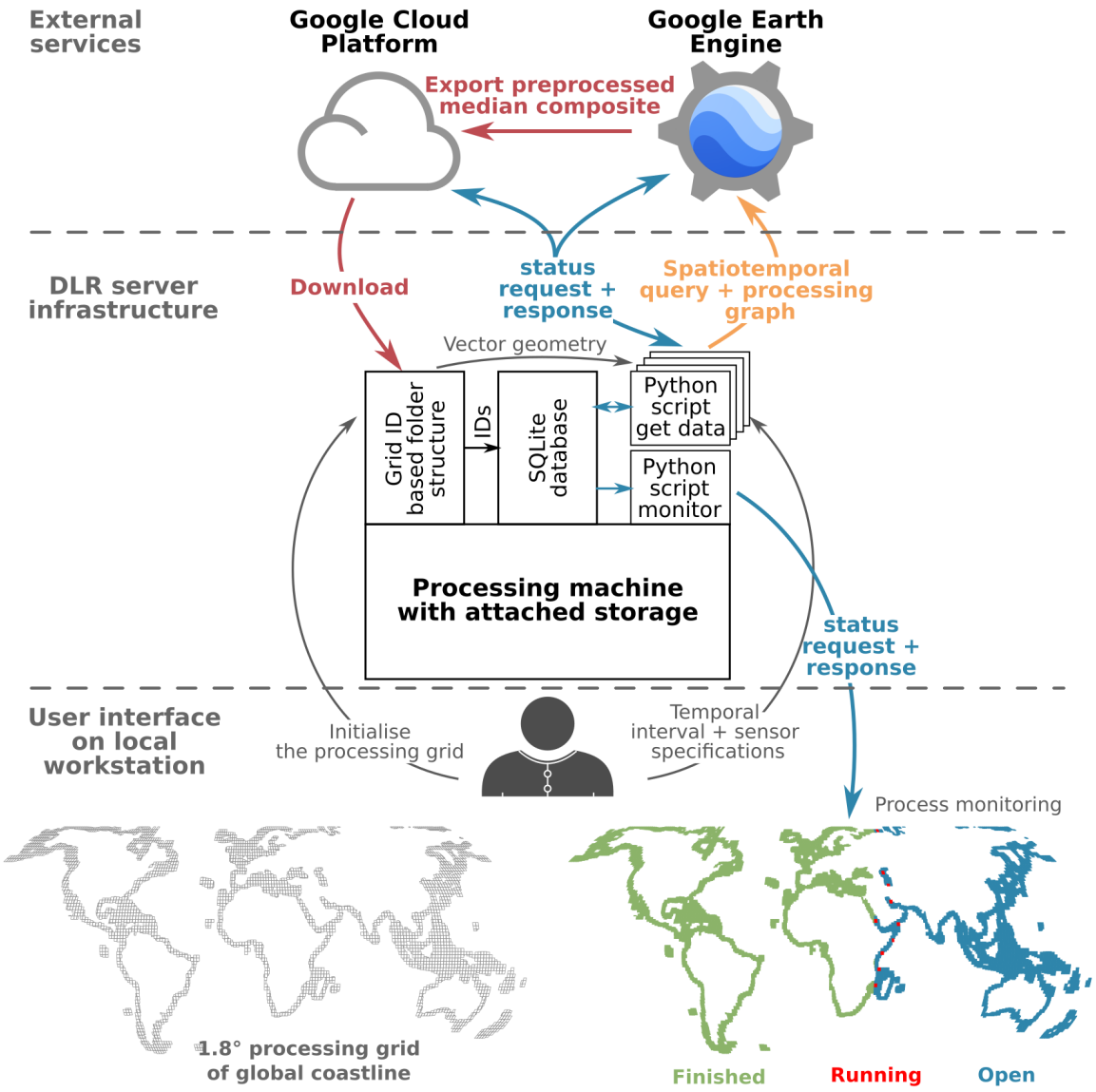



**Figure 4.8:** Conceptual depiction of the data acquisition pipeline of the global Sentinel-1 median composites. A user triggers a global download process of the data to a local processing machine. The process uses the Google Earth Engine (GEE) and Google Cloud Platform to query, process and download the median composites for each grid cell in parallel. By tracking feedback information of the Google services, monitoring of the download process is visualised to provide user feedback without interfering with the running processes.

the python scripts until all grid IDs have been processed. In addition, the feedback information stored in that database is used by another python script which outputs a summary of the entire data acquisition pipeline to allow the user quick monitoring of the process without interfering with the running tasks, see figure 4.8. In the following, whenever Sentinel-1 imagery is acquired in this dissertation, this data acquisition pipeline has been used. With the Sentinel-1 data acquisition pipeline set up, the proof of concept study and further global investigations can start to efficiently work on a large amount of data.

### 4.2.2 Ontology Example to Represent Offshore Wind Farms

In order to formalise perceptual and expert knowledge to machine-readable knowledge in an ontology, it helps to start with a semantic description of the details which have to occur in the representation. For OWFs, such a description is given by the following sentences:

"Offshore wind farms are located in the sea but can appear in coastal areas on tidal flats. Smaller wind farms are closer to the coast than larger wind farms. The wind turbine density decreases by increasing OWF size. Wind turbines are organised in a regular grid-like pattern with individual but consistent, systematic changes to the grid structure for each wind farm. The typical outer shape of the entire wind farm is a polygon with four to five sides" (Hoeser and Kuenzer, 2022b, p. 173).

Figure 4.9 provides a visual impression of how the SyntEO ontology represents the two-dimensional spatial characteristics of an OWF and how this single scene element is embedded in a meaningful synthetic scene composition with other scene elements. The given semantic description is hereby incorporated into this graphical representation in order to clarify the knowledge transfer process.

To generate a single OWF, first the characteristic *size* is defined. Therefore, the data generator randomly selects from a list of values that describe scaling factors as numerical information and keys as semantic descriptions. For example, by selecting the value *small:5*, see figure 4.9a), other OWF characteristics are scaled by the factor 5 later on to translate them into a metric system based on kilometres. With the semantic key *small* further contextual characteristics can be selected, for instance, the wind turbine density within the small-sized wind farm. In figure 4.9b) the key *small* predefines two discrete uniform distributions with ranges. From these distributions, two values are selected, which describe the dimensions of a regular turbine grid. Thus together, they express the turbine density. When the context attribution changes to another wind farm size, like medium or large, the ranges of the discrete uniform distribution also change to adjust the turbine density according to the OWF size.

Since not all OWFs have a perfectly rectangular grid layout, a deformation function to the x- and y-axis of the regular grid is applied, see figure 4.9c). This deformation function is randomly selected from a list of functions. Similar to the *size* characteristic, the deformation functions are mathematical expressions paired with semantics. The semantic notation is not necessary for the data generator but highly supports the readability of the ontology for a human interpreter. *Size*, turbine *density* and *grid deformation* are connected via context attributions to finally generate the internal *gridded wind farm layout* of the OWF scene element.

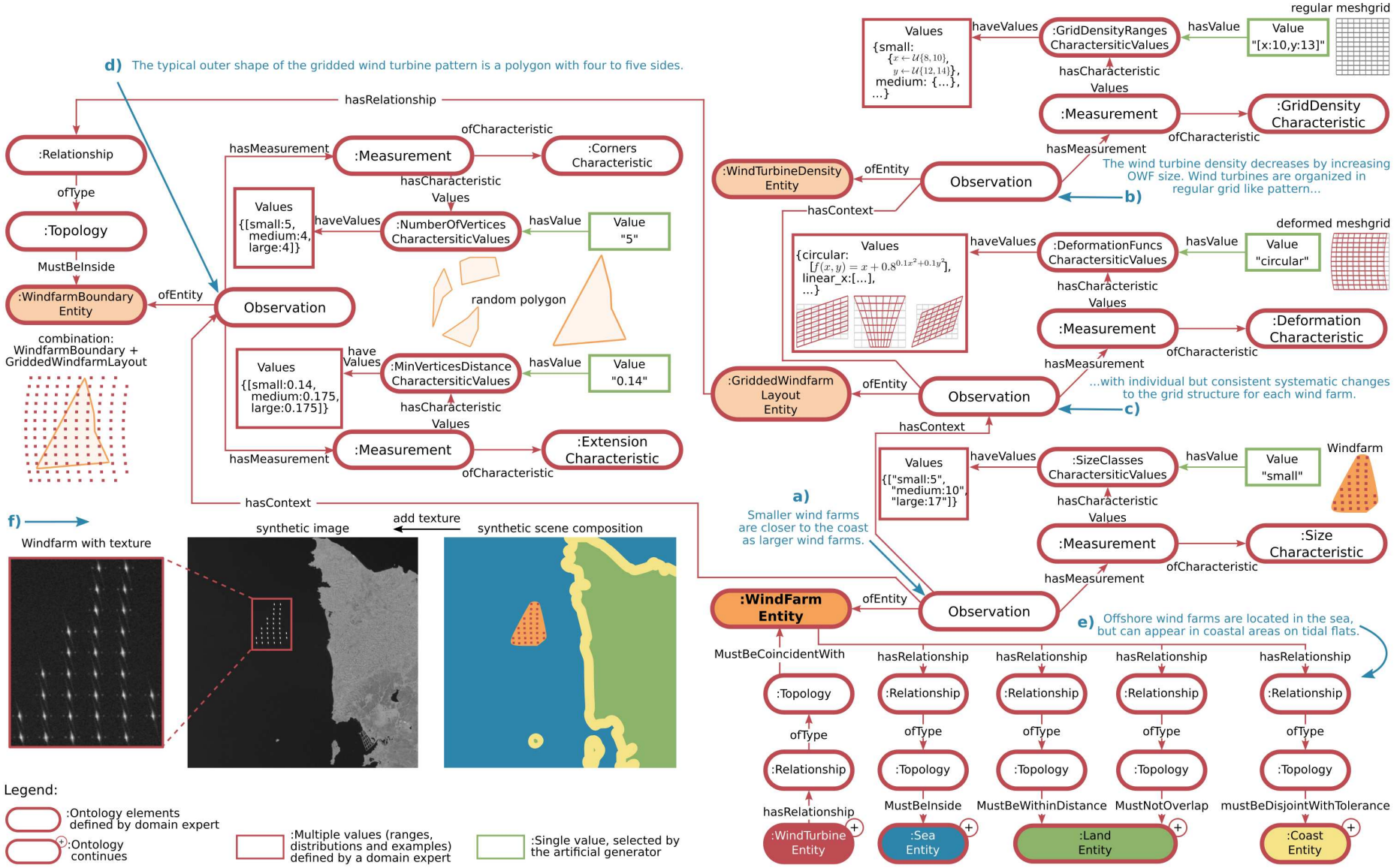



**Figure 4.9:** Excerpt of the SyntEO ontology for generating a synthetic remote sensing scene for offshore wind farms. a-d) describes the nested representation of an offshore wind farm. e) brings the wind farm entity into a spatial context with other entities. f) shows the resulting scene composition and how the synthetic remote sensing image is created by adding texture. Source: Hoeser and Kuenzer (2022b, p. 175).

Besides this internal structure of an OWF, the outer boundary has to be defined additionally. Therefore, a polygon is generated with a randomly selected *number of vertices* and a *minimum distance* between the vertices, see figure 4.9d). This *wind farm boundary* has to be combined with the internal *gridded wind farm layout*. In the ontology, this combination of the two loosely generated information is represented by a *Relationship* attribution of type *Topology* between the *gridded wind farm layout* and the *wind farm boundary*. The topological rule states that the *gridded wind farm layout must be inside* the *wind farm boundary*. The resulting subset can now be scaled with the initial *size* value to transform them into a spatially meaningful coordinate system in kilometres.

The described OWF is only one scene element of many others. The other scene elements describe non-targets, and most of their shapes are taken from template data. Therefore random crops of the OSM coastline are extracted and optionally transformed like flipped over the vertical or horizontal axis or rotated. With a coastline provided, coastal area, land, and open sea can be derived as two-dimensional geometries. Hereby, the spatial context of non-target scene elements in the scene composition is initialized. To embed the fully procedural generated OWF geometry, the topological relationships to the non-target scene elements are defined in the ontology, see figure 4.9e). By following these topological rules, a position of the generated OWF geometry is selected, and the final synthetic scene is composed. In figure 4.9f) texture is added to the scene composition. For the non-target scene elements, this texture comes from the downloaded Sentinel-1 median composites corresponding to the OSM coastline geometry. Thus, like the geometries for the non-target objects, the texture also comes from template data. However, the texture which describes the OWTs within the OWF is generated in a procedural manner, which will be discussed in detail in the next section as an example for the data generator processing backend.

With the presented excerpt of the ontology for the practical example of the representation of the geometry of an OWF, the structure of an OWF and its spatial relationships to other scene elements are clarified. The theoretical foundations from the previous section were herewith transferred into a practical example by repeatedly applying the building block of the SyntEO ontology.

### 4.2.3 Artificial Image Generator Backend

Figure 4.9 gives a visual impression of how the SyntEO ontology is structured. At the same time, it already provides hints about the processing backend, which uses the description in the ontology as input to generate the synthetic image. The transition from ontology to processing backend is, by design, smooth since the task of the ontology is to represent human knowledge in a way a machine can read and use it to generate a synthetic remote

sensing scene. Thus, some of the information within the ontology has to be close to the technical processing backend. For the OWF example, the processing back end is based on representing all information as NumPy arrays (Harris et al., 2020). Thus each distribution, geometry or pixel value is represented as a one- or multidimensional array. The processing backend for the generation of a synthetic training data set that describes offshore wind energy infrastructure was developed for this specific task and is based on the structures provided by the NumPy library. To get an impression of how this can look like, the generation of OWT texture and an entire oil rig non-target scene with texture are discussed in this section.

In figure 4.9f), texture is added to the scene composition. Thereby, the non-target texture comes from the Sentinel-1 template data. Technically, the processing backend uses the geometrical description of a coastline segment, defines areas of land, coast and open sea from it and uses these areas to clip the corresponding Sentinel-1 image to obtain texture for these areas. However, the generate OWT locations also need the texture to appear in the synthetic image. This texture does not come from template data. Instead, the texture for an OWT is generated procedurally. The technical implementation of generating pixel values that visually represent an OWT is mainly based on two-dimensional kernels. Figure 4.10 gives an overview of how the processing backend stepwise produces and merges these kernels to finally generate a texture array for an OWT location. In figure 4.10a-g), the backscatter clusters of the turbine centre and the layover and double bounce effects are generated. For the turbine centre, depicted in figure 4.10a-d), two Laplacian kernels $\mathcal{L}$, see equation (25), are generated. Each kernel is described by $i = 1, 2, \ldots, 40$ values, equally distributed in an interval from -1 to 1, $x = [-1, 1]$. Both Laplacian kernel functions have a location parameter $\mu_{\mathcal{L}}$ of 0, and a scale parameter $\sigma_{\mathcal{L}}$ of which the kernel assigned to the horizontal axes has a smaller scale parameter $\sigma_{\mathcal{L}h} \in \mathbb{Q} : 0.25 \leq \sigma_{\mathcal{L}h} \leq 0.5$ as the kernel assigned to the vertical axis $\sigma_{\mathcal{L}v} \in \mathbb{Q} : 0.75 \leq \sigma_{\mathcal{L}v} \leq 1$

$$\mathcal{L}(x_i) = \frac{1}{2\sigma_{\mathcal{L}}} e^{-\frac{|x_i - \mu_{\mathcal{L}}|}{\sigma_{\mathcal{L}}}} . \tag{25}$$

By taking the outer product of the two arrays with a dimension of $1 \times 40$, a two-dimensional kernel with a width and height of $40 \times 40$ depicted in figure 4.10a) is generated. To make the smooth texture more realistic, an array of the same dimension with random noise is added to the two dimensional Laplacian kernel. After that, the resulting raster values are stretched to typical backscatter values of OWT centres, characterised in the ontology. To manipulate the shape of the texture, a binary mask is generated, see figure 4.10c). The mask is initialised in the centre of the $40 \times 40$ shaped array and grows vertically and

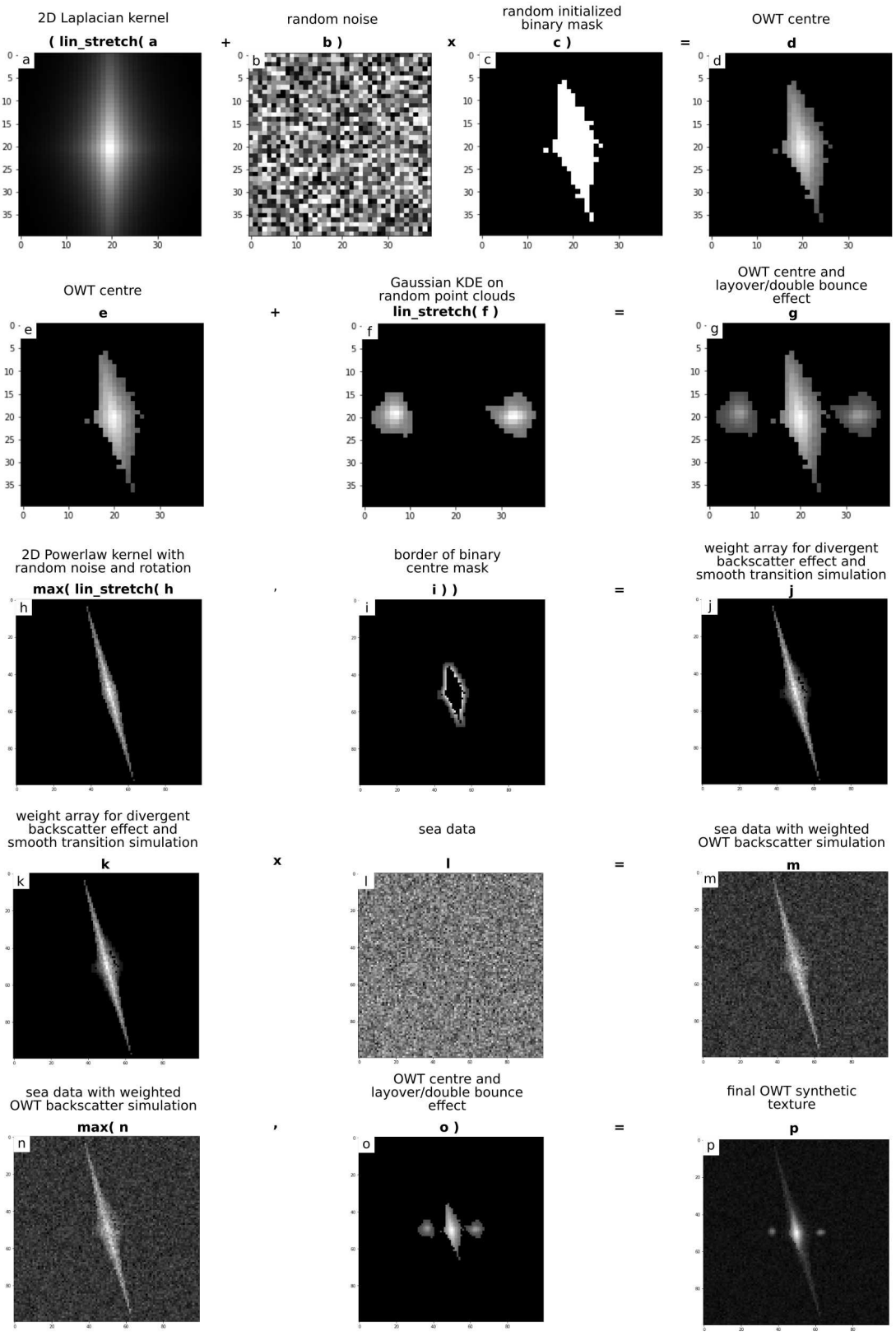



**Figure 4.10:** Example of a stepwise construction of an offshore wind turbine (OWT) texture from two-dimensional kernel functions, morphological operations and noise. Each individual turbine texture created in this way has a unique signature due to the random selection of parameters of each step.

horizontally, where the maximum horizontal extent decreases by increasing vertical distance to the centre. The ontology provides both values, which define the maximum extent. A subsequent erosion operation randomises the shape of the otherwise symmetrical mask. By applying this result to the two dimensional Laplacian kernel, the texture of the centre of an OWT is provided, see figure 4.10d).

Similar to this, the two smaller clusters to the left and right of the OWT location are generated. Therefore two point clouds are generated in a unit square with location parameters described by two normal distributions. The location parameter of these normal distributions defines the centres' vertical and horizontal locations. A two-dimensional Gaussian kernel density estimation is applied to them, and the resulting two-dimensional empirical kernels are transformed into an array with the dimension $40 \times 40$, see figure 4.10f). The approach of a kernel density estimation in a 2D plane of points which have coordinates from a Gaussian normal distribution was chosen to achieve a greater heterogeneity of the texture instead of directly using a two-dimensional Gaussian kernel which would be symmetric. Similar to the turbine centre, a linear stretch is applied to adjust the pixel values to typical backscatter amplitudes. By adding the array of the centre texture and the array of the layover and backscatter signatures, the main part of the turbine texture is finished, see figure 4.10e-g).

The next feature describes the radar signature coming from the backscatter effect of strong scatterers over open water. For simplicity, only a single orbit direction is considered in this example. As a starting point, two kernels of the power-law function $\mathcal{P}$ are generated, see equation (26), where $\sigma_{\mathcal{P}} \in \mathbb{Q} : 0.8 \leq \sigma_{\mathcal{P}} \leq 0.88$ is the shape parameter

$$\mathcal{P}(x_i) = \sigma_{\mathcal{P}} x_i^{\sigma_{\mathcal{P}}-1}. \tag{26}$$

Similar to the two dimensional Laplacian kernel, two one dimensional kernels with different parametrisation for the horizontal and the vertical axis are generated and combined by calculating their outer product to a two-dimensional kernel. This time the shape of the array is larger, $100 \times 100$, due to the larger spatial extent of the feature to generate. In addition, the array is rotated, to simulate the orbit flight path inclination, and random noise is added to it, as depicted in figure 4.10h). The earlier generated centre mask, see figure 4.10c), is used to generate a border with decreasing intensity around the mask to simulate a smooth transition between the generated turbine centre and the backscatter over open water effect texture, see figure 4.10i). Both the power-law kernel and centre border are stretched to represent weights in a range between 1 and 1.4. Both arrays are then combined by stacking them vertically and reducing the stack by selecting the maximum values, see figure 4.10h-j). Since the effect in real radar images is a result of the interaction between a strong scatterer and

open water, water has to be taken into account when producing this texture. Therefore, the water pixels which are at the location of the OWT in the scene compositions are extracted from the raster, provided by the template Sentinel-1 data. The resulting array is then multiplied by the weight array produced before. Thus the spatial pattern of the strong scatterer over open water is combined with real backscatter values that are coming from the water surface itself. By using the produced kernels as weights instead of direct backscatter values, much higher randomisation is realised in the texture, see figure 4.10k-m).

Finally, the generated arrays of the radar signatures can be combined. Therefore, the first result of the turbine centre and the two layover and double bounce clusters are padded with zero values to reach the same dimension as the second array with a larger spatial extent, compare figure 4.10g+m). Then, both arrays are vertically stacked and reduced by selecting the maximum values. This generates the texture for a single OWT location, see figure 4.10n-p). The entire procedure of procedural texture generation is repeated for each OWT location in the scene composition. To maximise variability, the parameters of the kernel functions, ranges of backscatter values to be simulated, and the overall structure of the OWT texture design are randomised. For example, an OWT which was acquired at both orbit directions would have an additional power-law kernel. Larger OWTs would have larger centre clusters and spacing between the layover and double bounce clusters. This parameter selection is described in the ontology and controlled via context attributions, e.g. an OWF of size large triggers larger OWT textures. Due to the randomisation of the texture generation, each OWT texture can be assumed to be unique.

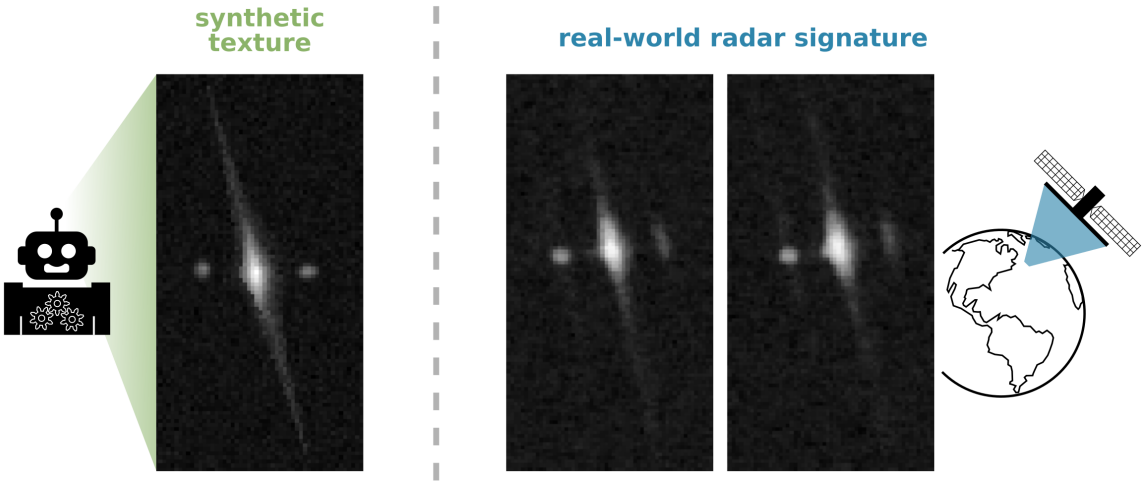



**Figure 4.11:** Visual comparison of a synthetically generated texture and two real-world textures of offshore wind turbines (OWTs) in the East China Sea. Distinct spatial features of the radar signature are approximated by the synthetic texture to communicate them during the training process of a convolutional neural network.

Figure 4.11 compares the generated example to two real-world OWF signatures which were acquired at a single orbit direction in the East China Sea. It becomes clear that when learning a generalized version of the spatial characteristics of the synthetic texture, the spatial features of a real-world radar signature of an OWT can be approximated.

In combination with the procedural shape and texture generation of OWFs and OWTs and the template data based non-target scene element generation, it is possible to synthesise thousands of individual remote sensing scenes. However, a CNN will not necessarily learn all of the features embedded in these images, but those which are the strongest and clearest, or loosely speaking, the easiest to learn. At this stage, natural environments with small islands are provided along with the OWFs since the scene elements coastline and land appear in the scene compositions. However, human-made offshore infrastructure is underrepresented in the data set. Thus, most probably, a CNN will learn to look for strong backscatter clusters within the open sea. It does not need to look for specific structures of these clusters since, in the training data set, nothing else exists which looks similar to an OWF. A brief discussion of a more intuitive problem explains this issue. If a CNN is trained to recognise dogs and the training data set does only contain images of dogs, it will not learn that there are other animals, and thus it is prone to false positives when applied in the real world. For example, if the trained CNN has to predict an image of a giraffe, it would be certain that this is a dog. Since it is a narrow artificial intelligence, it would not even think that this dog has a strangely long neck, but it would just predict that this image looks doggish, hence it has to be a dog in it. Thus, showing the artificial agent more of the world about which it makes decisions during training allows it to learn a more realistic complexity.

This example demonstrates the importance to consider potential false positives. A proper training data set has to contain training examples that show closely related classes in order to teach the CNN the detailed differences it has to look for. In the OWF study, a false positive training example would be an image of other marine infrastructure without these objects being labelled. Thus when the CNN predicts a bounding box for such an object during training, this detection would result in a very high error. The associated gradients derived from it would force the parameters of the CNN to finally find the difference between the actual targets and such non-targets, which look very similar. It can be imagined as a sort of fine-tuning of a more sophisticated combination of kernel functions within the CNN. In this way, the CNN has to find other more detailed features to learn differences, and at this moment, the CNN starts to learn the very characteristics of the gridded OWF structure, or the small scale spatial features of a single OWT. However, these images that show false positives have to be generated additionally to provide them in sufficient quantity.

The processing backend to generate these images is based on a methodology that has its origins in the production of visual effects with CGI, a very important field for developing a capable image processing backend. The large scale difference between OWF and sites with other marine infrastructure like oil rigs is that an OWF appears as a structured grid-like pattern, whereas oil rigs are accumulated in random clusters which look more organic, as discussed in section 3.2. To simulate these large scale patterns, gradient noise in its variant of OpenSimplex noise was used, see figure 4.12. OpenSimplex noise is a descendant of Perlin noise, which was originally developed by Perlin (1985) for the movie Tron in 1982. It allows generating noise that is not entirely random but follows a local gradient. Perlin noise and its variants are still commonly used, especially in video game design and computer graphics, to efficiently generate animations or surface geometries (Perlin, 2001). For the SyntEO framework and its processing backends, they are especially important since they allow for the synthesis of natural environments which are otherwise complicated to describe. Thus these techniques are promising tools for a processing backend to generate Earth observation data.

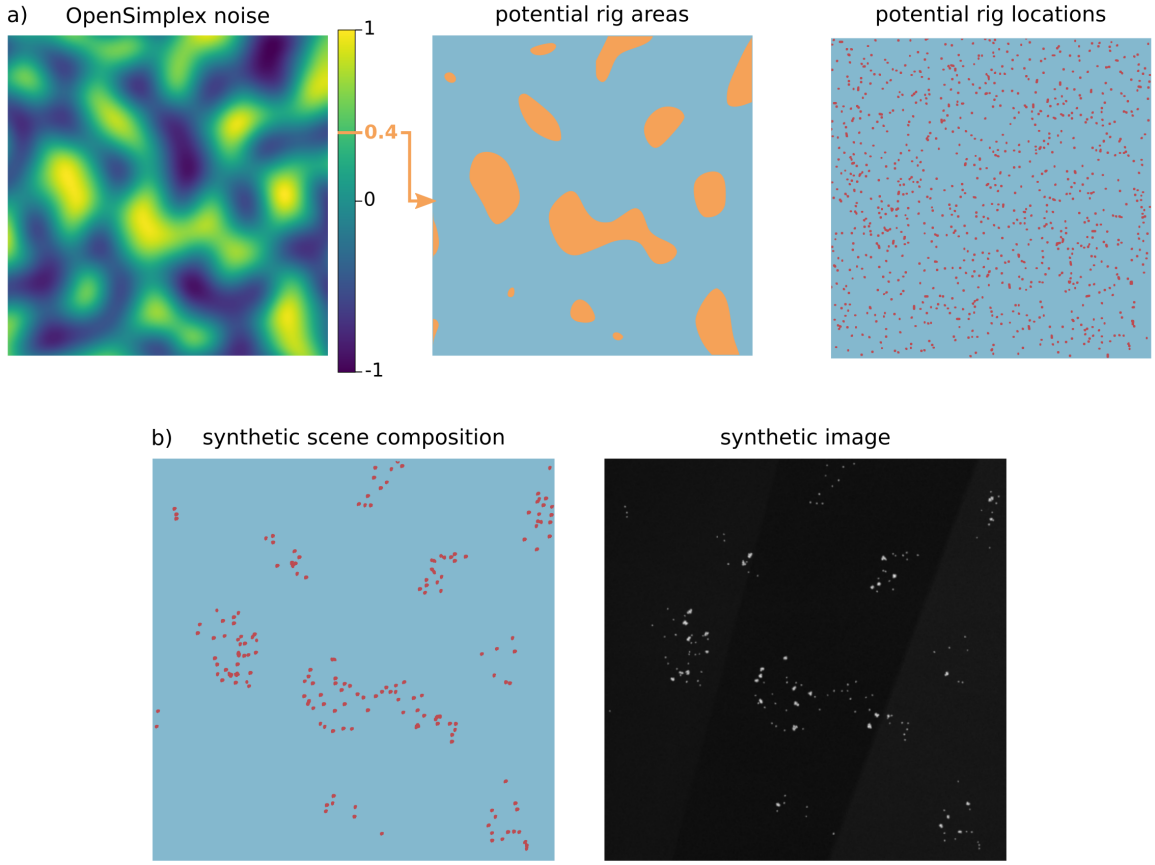



**Figure 4.12:** The generation of organic shapes by using OpenSimplex noise to finally generate a synthetic radar image showing oil rig clusters. Source: Hoeser and Kuenzer (2022b, p. 176).

In this specific example for oil rig generation, OpenSimplex noise is generated and a threshold applied to the values to design large scale spatial patterns which look unstructured with an organic shape, see figure 4.12. The next step generates platform locations by randomly drawing coordinates from a uniform distribution. By applying the topological rule *must be inside*, only such platform locations are kept which are within the generated OpenSimplex noise geometries. After this step, the final scene composition is set up. For the open sea area in this scene composition, the texture is added by selecting Sentinel-1 template data from gird cells which only show the open sea. The earlier described method of generating empirical two-dimensional kernels by applying a Gaussian kernel density estimation over point clouds is reused to generate the texture of platform locations. This time the spatial scaling factor of the point cloud coordinates and kernel width is larger to represent an entire oil rig instead of only the small clusters belonging to layover and double bounce effects of the turbine's nacelle. Finally, the image is generated with a corresponding annotation file without any entries of a target object, which makes it clear that each detection made on this image is a false positive.

The described technical insights in the implementation of the processing backend provide a better intuition about this important but task-specific component of the SyntEO framework. What applies to the ontology also applies to the processing backend: The more frequently the SyntEO framework is used to describe and generate data, the greater the volume of already prepared building blocks that can be reused. Thus, implementing a processing backend will become easier over time. The more shapes and textures have been built in the past, the more processing backend modules exist and can be reused and rearranged to implement a new processing backend.

### 4.2.4 Data Set Variants and Composition

Section 4.2.3 already emphasised the importance of large and balanced training data sets. This example demonstrated that the balancing is not only about target-feature representation but also non-target examples. Only providing randomly generated images-annotation pairs will not be sufficient to exploit the potential of a CNN. Four consecutive data sets that stepwise expand in complexity, variability, and represented targets and non-targets have been generated to prove this statement. Figure 4.13 shows the main components and changes of the four data sets.

The first data set only contains OWFs with a small size and no coastline but only land and open sea as non-target scene elements. Furthermore, the open sea is not Sentinel-1 template data but a constant value drawn from a uniform distribution, see figure 4.13ae). Data set-1 also does not include any non-target training examples like oil rigs. Overall,

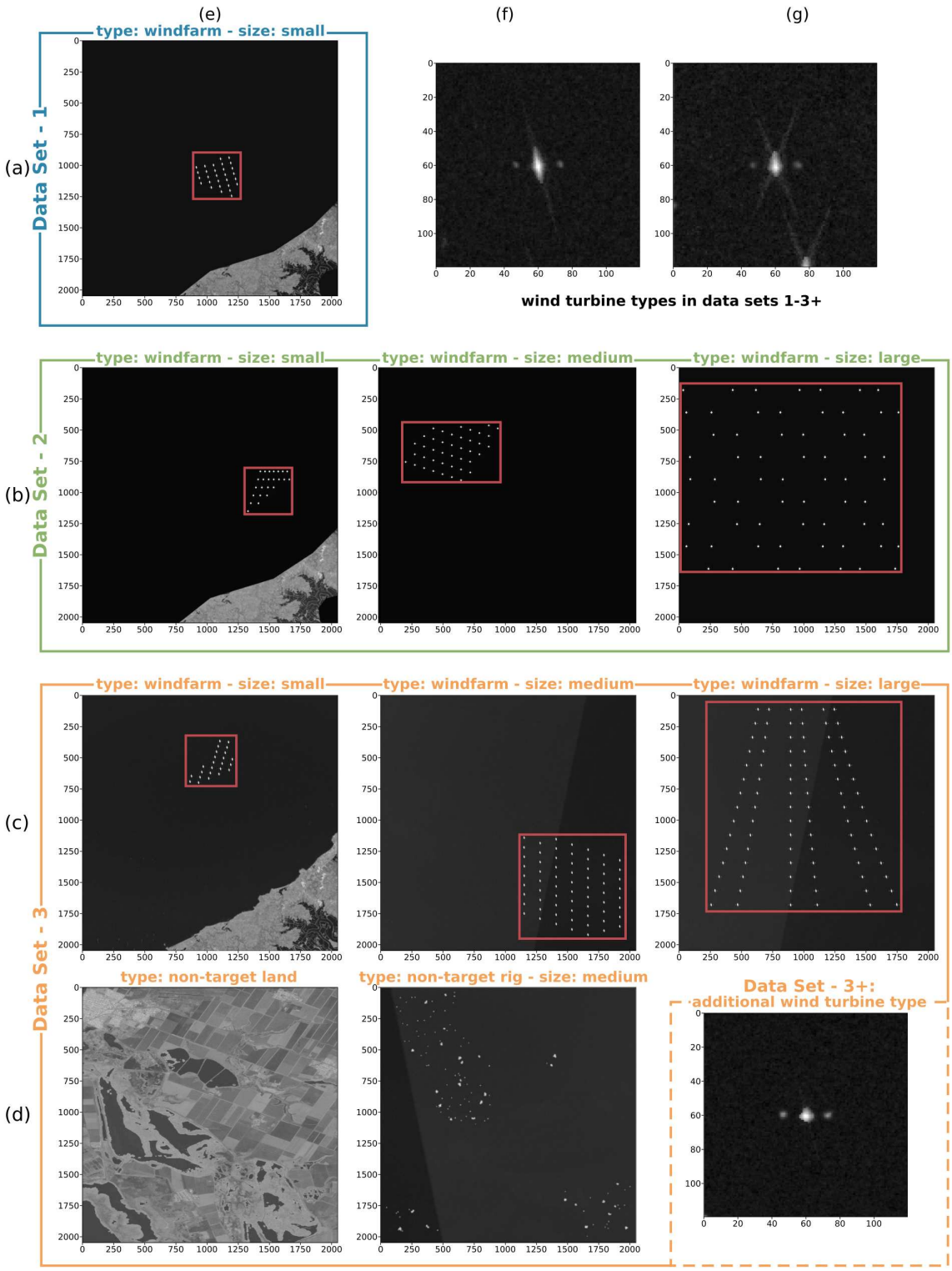



**Figure 4.13:** The evolution of synthetic training data sets for offshore wind farm detection with increasing complexity. Data set-1 shows only small scale offshore wind farms (OWFs) and mainland. Data set-2 adds two more OWF scales, medium and large. Data set-3 adds the coastline to the mainland and two non-target classes of pure land and oil rigs. Data set-3+ finally adds an additional offshore wind turbine (OWT) texture type, which can be found on tidal flats. Source: Hoeser and Kuenzer (2022b, p. 177).

data set-1 contains 45,000 images that show small OWFs, a simplified open sea, and land without a coastline. For the second data set, all OWF sizes and all associated contextual properties have been enabled in the ontology. Thus OWFs do appear at the three sizes small, medium and large in data set-2, see figure 4.13b). However, the setting for the non-target classes was kept as for data set-1. The data set-2 has 90,000 images, 1/4 of each small and large-sized OWFs, and 2/4 of medium-sized OWFs. The third data set additionally contains non-target examples, see figure 4.13c+d). The coastline area was embedded in images with a transition between land and open sea, and the open sea areas are no longer constant values but Sentinel-1 template data. More importantly, non-target examples that show unstructured marine infrastructure clusters and a second type that shows only land are added to the data set composition. The overall size of data set-3 is 90,000 training examples with small and large OWFs and non-target oil rigs and land, each contributing 1/6 of the 90,000 images. The remaining 2/6 are provided by training examples containing medium-sized OWFs.

Finally, data set-3+ was composed. As the name suggests, it has the same size and composition as data set-3, but it has an additional OWT type, see figure 4.13dg). In all other data sets, the two major variants in figure 4.13ae+ag) have been used. Thereby, ae) is a OWT signature of a turbine acquired at an ascending orbit direction and figure 4.13ag) is a signature of an OWT acquired at both orbit directions. Furthermore, both variants show turbines surrounded by the open sea. The new OWT type figure 4.13dg) shows a signature that can typically be found on tidal flats. Due to the lack of water in some images of the median composite, the typical radar signature of an OWT is less pronounced.

The data set generation and final composition process of each of the four data sets have the same workflow. The necessary parts of the ontology were enabled, corresponding to each data set variant. Thus the image generator produced data sets with increasing complexity by reusing the same ontology with additional features enabled for each more complex data set variant. Since some parts of the data generator are random processes, the total number of generated examples was chosen to be higher than the number of examples in the finally composed training data sets. This way, an unstructured *pile* of image and annotation files along with the corresponding metadata was generated. The images are stored as `.png` fiels, the annotations and metadata information as `.xml` files of which the annotation `.xmls` are following the PASCAL-VOC annotation convention (Everingham et al., 2010). The metadata files were combined into a database, and by choosing a balanced number for each group of image types, the corresponding image-annotations pairs were selected from the unsorted pile of data. The training data set compiled in this way was then converted to `.TFrecord` binary files, ready for deep learning with the TensorFlow object detection API.

### 4.2.5 CNN Model and Training

The chosen CNN model for object detection is the Faster R-CNN two-stage detector (Ren et al., 2015). The ResNet-50 convolutional backbone (He et al., 2016) was chosen as preceding feature extractor. It ingests the input image and extracts the necessary features before the Faster R-CNN uses them for object localisation, bounding box regression and class prediction. The ResNet-50 Faster R-CNN model structure was technically implemented by employing the TensorFlow deep learning framework (Abadi et al., 2015). The input size of an image to the CNN is $1,024 \times 1,024$ pixels. Thus on the fly, downscaling of the $2,048 \times 2,048$ pixel-sized synthetic training examples is performed. Since the target objects are entire OWFs, large scale features have to be preserved. Thus the downscaling before training approach of the entire synthetic image was chosen, following the concept presented in section 4.1.2. As introduced in section 2.5.3.3, the Faster R-CNN contains the region proposal network (RPN) submodule. This submodule is needed to find region proposals of potential objects. It relies on so-called anchor boxes, which can be imagined as boxes equally distributed over the last feature map of the convolutional backbone. The initial scale and aspect ratios of these boxes have to be defined manually. Aspect ratios of the anchors $\mathcal{A}_{\text{aspect}}$ were set to [0.5, 1, 2] to obtain one squared and two rectangular shapes, whereas the scale ratios $\mathcal{A}_{\text{scale}}$ where set to [0.25, 0.5, 1, 2, 3.5]. The scale ratios were calculated by taking typical OWF sizes and their size downscaled by the convolutional feature extractor in the last feature map into account. A scale one anchor box of the RPN would have the pixel dimensions $\mathcal{A}_h = 16, \mathcal{A}_w = 16$, with width $_w$ and height $_h$, in the last feature map. Typical target sizes $\mathcal{T}$ are [128, 256, 512, 1024, 1792] times 10 m in the input image. The downscaling factor of the convolutional backbone, coming from the max pooling operations, is called the stride $S$ of the network. For the employed ResNet-50 architecture, the stride is 16. Hereby an anchor scale $\mathcal{A}_{\text{scale}}$ can be calculated by taking an additional downscaling of the input image $\mathcal{I}$ due to the model's input size $\mathcal{M}$ of the CNN into account

$$\mathcal{A}_{\text{scale}} = \sqrt{\mathcal{T}_h \mathcal{T}_w \frac{\mathcal{M}_h \mathcal{M}_w}{\mathcal{I}_h \mathcal{I}_w}} \times \frac{1}{S\sqrt{\mathcal{A}_h \mathcal{A}_w}}. \tag{27}$$

In the case of the same width $_w$ and height $_h$ for each variable, this can be simplified to

$$\mathcal{A}_{\text{scale}} = \frac{\mathcal{T} \times \mathcal{M}}{\mathcal{I} \times S \times \mathcal{A}}. \tag{28}$$

This calculation of the anchor box scale factors is possible due to the available information on the size parameters of the target entities in the ontology. Normally, the proper anchor scale definition needs an in-depth investigation of the bounding boxes of the training data set (Redmon and Farhadi, 2016).

The four synthetic data sets were split into a 95% training and 5% validation set in order to use a fraction of the synthetic images to track the training process. The test set, which will later be used for evaluation, is entirely independent of the synthetic data set and only contains real-world images. The synthetic data sets-2 until 3+ with each holding 90,000 examples are split into 85,500 examples for the train set and 4,500 examples for the validation set. For the smaller data set-1 with 45,000 examples, the split is 42,750 for the train set and 2,250 for the validation set. The training set sizes can be considered as large enough for the training of the object detector CNN with a single target class OWF. In fact, the sizes are so large that no further data augmentation or iterations, so-called epochs, of the data set is necessary during training. This demonstrates the advantages of the SyntEO approach. Instead of reusing a relatively small training data set in multiple epochs to decrease the cost function to a minimum, the scaleable size of the synthetic data set allows to only use each training example once in the entire training process. That way, overfitting to specific training examples is impossible.

The optimisation approach is stochastic gradient descent (Robbins and Monro, 1951; Bottou et al., 2018), similar to the machine learning cycle introduced in section 2.2. With a batch size of 4, 21,375 training steps for the larger data sets and 10,687 training steps for the smaller data set are possible without repeating a training example to optimise the CNN. The learning rate was scheduled with the cosine decay approach (Loshchilov and Hutter, 2017). After a warm-up phase of 1,000 training steps, the base learning rate of 0.01 is reached and decreased to 0 by following the cosine function for all remaining training steps. In addition, a momentum of 0.9 was chosen. Simply speaking, the momentum describes the share of the last gradient update, which is added to the gradients of the current training step. This way, the learning process is less prone to stuck in local minima of the cost function and enables the learning process to find a better local minimum or even the global minimum faster (Polyak, 1964; Sutskever et al., 2013). The hardware used for training the models were four parallel Nvidia RTX 2080-Ti GPUs with 11 GB memory each. Finally, all four models, which are named model-1 until model-3+ corresponding to the employed training data sets, were exported and ready for predicting OWFs on real-world data.

### 4.2.6 Offshore Wind Farm Detection and Accuracy Assessment

By using the four trained CNN object detectors, OWFs were detected in the Sentinel-1 median composites of the third quarter of 2020 within the four test sites. Therefore, all Sentinel-1 median composite tiles of the test sites were sliced into image chips with a dimension of $2,048 \times 2,048$ pixels and overlap of 50%. These images chips were downscaled to $1,024 \times 1,024$ before entering the trained CNNs. All detected bounding boxes with a detection score higher than 0.8 were considered valid detections. The provided pixel coordinates of the detected bounding boxes of each OWF were transformed to geographic coordinates of the coordinate reference system (CRS) World Geodetic System 1984 (WGS84) and merged into a single file. Multiple detections of the same OWF caused by the 50% overlap or the unification of large OWF clusters which are within multiple image chips, were dissolved by a cascading procedure. Therefore, each bounding box $\mathcal{B}$, starting with the highest score, selects all intersecting bounding boxes. This way, a list of bounding boxes with $i = 1,2,3,\ldots,B$ is created. The detection score sorts this list in descending order, where the box with the highest score is called the leading box, and all other boxes in the list are member boxes.

$$\text{IoU} = \frac{\mathcal{B}_i \cap \mathcal{B}_{i+1}}{\mathcal{B}_i \cup \mathcal{B}_{i+1}} \tag{29}$$

Now each member box in this list is compared with the leading box. If the member box is entirely within the leading box or their intersection over union (IoU), see equation (29), is higher than 0.33 the geometries of both boxes are unified, and the resulting geometry updates the geometry of the leading box. If no condition of the above is valid, the member box geometry is put aside to enter another box list or is an OWF on its own. Equation (30) provides a summary of this process

$$\mathcal{B}_1 := \bigcup_{i=1}^{B-1} \begin{cases} \mathcal{B}_1 \cup \mathcal{B}_{i+1}, & \text{if} \quad \mathcal{B}_{i+1} \subseteq \mathcal{B}_1 \\ \mathcal{B}_1 \cup \mathcal{B}_{i+1}, & \text{if} \quad \frac{\mathcal{B}_1 \cap \mathcal{B}_{i+1}}{\mathcal{B}_1 \cup \mathcal{B}_{i+1}} > 0.33 \\ \mathcal{B}_1, & \text{otherwise} \end{cases} . \tag{30}$$

This procedure results in a very close boundary around a detected wind farm since it can consist of several bounding boxes which are optimised to describe the outer boundary of the detected target. Especially for large OWF clusters these boundaries are closer to the ground truth bounding boxes of an OWF, see figure 4.16 for some examples.

To assess the performance of the models, the predicted OWF locations are compared to the hand labelled ground truth bounding boxes. Thereby a true positive prediction is defined by an IoU of 0.33 or higher of a predicted OWF boundary with a ground truth bounding box. With this definition of a true positive, the precision Pr and recall Rc were calculated

$$\mathrm{Pr} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}}, \tag{31}$$

$$\mathrm{Rc} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}. \tag{32}$$

Precision and recall were combined by calculating their harmonic mean, the F1 score

$$\mathrm{F1} = 2 \times \frac{\mathrm{Pr} \times \mathrm{Rc}}{\mathrm{Pr}+\mathrm{Rc}}. \tag{33}$$

Furthermore, the commonly used average precision AP for object detection and its corresponding precision-recall curve were calculated to better visualise the differences in the model performances. Therefore, all detections were sorted in descending order by their prediction score. From this sorted list with $i = 1, 2, 3, \ldots, D$, where $D$ is the number of detections, the all point interpolated precision-recall curve $\mathrm{Pr}_{\mathrm{interp}}$ is derived, see Padilla et al. (2021)

$$\mathrm{Pr}_{\mathrm{interp}} = \max_{\tilde{\mathrm{Rc}}:\, \tilde{\mathrm{Rc}} \geq \mathrm{Rc}} \mathrm{Pr}_{\mathrm{interp}}(\tilde{\mathrm{Rc}}). \tag{34}$$

The average precision AP describes the all point interpolated precision-recall curve in a single value by calculating its area under the curve (Padilla et al., 2021):

$$\mathrm{AP} = \sum_{i=1}^{D} (\mathrm{Rc}(i) - \mathrm{Rc}(i-1)) \times \mathrm{Pr}_{\mathrm{interp}}(\mathrm{Rc}(i)), \tag{35}$$

where $\mathrm{Rc}(0) = 0$.

### 4.2.7 Detection Results

Figure 4.14 shows the detection performance of all four models. As expected, the model's performance increases with increasing training data set complexity. The highest difference can be seen between model-1 and model-2. For data set-2, the different size parameters for an OWF have been enabled in the ontology, which seems to be a highly

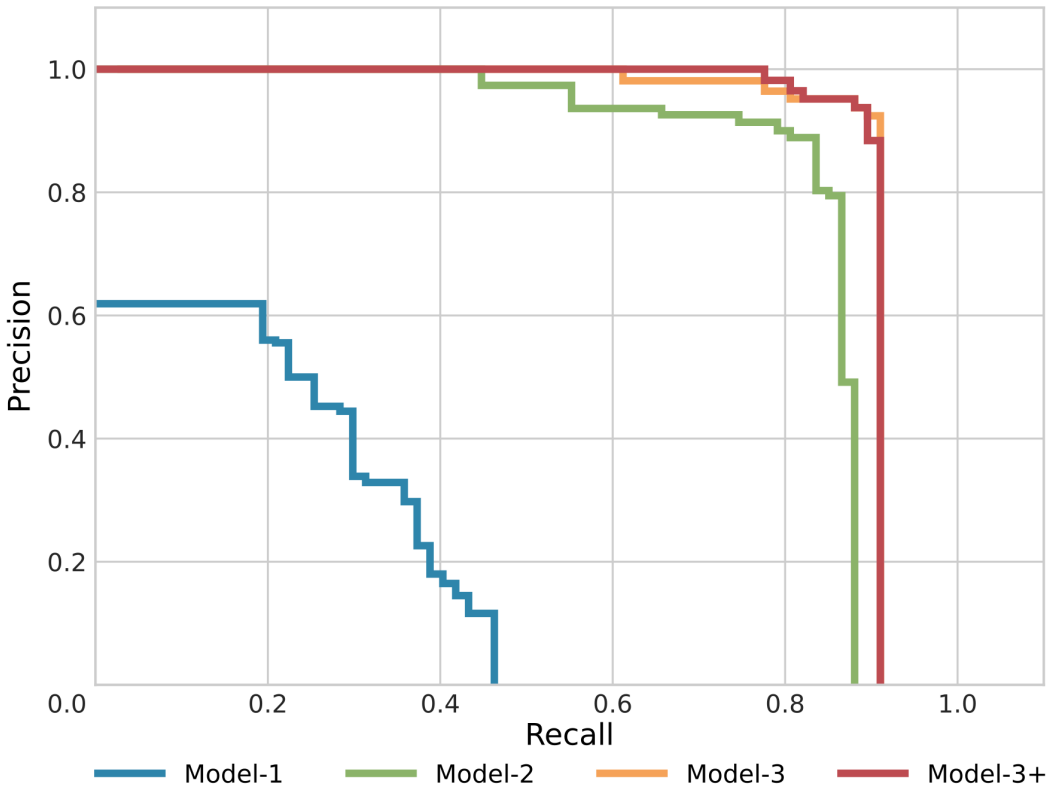



**Figure 4.14:** The four interpolated precision-recall curves for each of the trained models and the associated synthetic training data sets. The average precision scores for each curve are as follows Model-1 (0.21), Model-2 (0.842), Model-3 (0.901), Model-3+ (0.904). Source: Hoeser and Kuenzer (2022b, p. 179).

important feature. The embedding of non-target information has exactly the effect as intended. The precision values increase clearly, and even recall is slightly higher for models 3 and 3+ compared to model-2. Overall, model-3 and model-3+ are the best performing models. Table 4.1 confirms this interpretation. The recall values increase drastically from model-1 to model-2 and only slightly from thereon. For data set-3 and 3+ the precision values have their strongest increase indicating a better differentiation between true positives and false positives. The influence of data set-3 and 3+ on a better representation of false positive FP examples becomes especially clear when looking at the non-OWF test sites. Here, the performance increase for model-3 and 3+ is obvious.

These results show that by including different sizes of OWFs in the training data, this important feature is learned by the CNN and the detection of OWFs increases drastically. Furthermore, by adding non-target information in the data sets 3 and 3+, the corresponding CNN models 3 and 3+ are forced to further differentiate between coastal environments, other marine infrastructure and the real OWF targets, and with it the false detection rate decreases.

**Table 4.1:** All calculated metrics (true ppositive (TP), false ppositive (FP), false negative (FN), recall (Rc), precision (Pr), average precision (AP)) for the four models performed on the ground truth (GT) data. The metrics are presented for each site separately and **combined** by micro averaging. For the **combined** metrics, the best scores are underlined. Changed after Hoeser and Kuenzer (2022b, p. 179).

| Studysite | GT | TP | FP | FN | Rc | Pr | F1 | AP |
|---|---|---|---|---|---|---|---|---|
| **Model - 1** | | | | | | | | |
| **Combined** | **67** | **31** | **252** | **36** | **0.463** | **0.11** | **0.177** | **0.21** |
| North Sea Basin | 42 | 19 | 85 | 23 | 0.452 | 0.183 | 0.260 | 0.217 |
| East China Sea | 25 | 12 | 67 | 13 | 0.48 | 0.152 | 0.231 | 0.282 |
| Persian Gulf | 0 | | 100 | | | | | |
| Sea of Azov | 0 | | 0 | | | | | |
| **Model - 2** | | | | | | | | |
| **Combined** | **67** | **59** | **80** | **8** | **0.881** | **0.424** | **0.573** | **0.842** |
| North Sea Basin | 42 | 40 | 3 | 2 | 0.952 | 0.93 | 0.941 | 0.952 |
| East China Sea | 25 | 19 | 22 | 6 | 0.76 | 0.46 | 0.576 | 0.712 |
| Persian Gulf | 0 | | 31 | | | | | |
| Sea of Azov | 0 | | 24 | | | | | |
| **Model - 3** | | | | | | | | |
| **Combined** | **67** | **61** | **11** | **6** | **0.91** | **0.847** | **0.878** | **0.901** |
| North Sea Basin | 42 | 40 | 1 | 2 | 0.952 | 0.976 | 0.964 | 0.952 |
| East China Sea | 25 | 21 | 5 | 4 | 0.84 | 0.808 | 0.824 | 0.817 |
| Persian Gulf | 0 | | 5 | | | | | |
| Sea of Azov | 0 | | 0 | | | | | |
| **Model - 3+** | | | | | | | | |
| **Combined** | **67** | **61** | **14** | **6** | **0.91** | **0.813** | **0.86** | **0.904** |
| North Sea Basin | 42 | 40 | 0 | 2 | 0.952 | 1 | 0.976 | 0.952 |
| East China Sea | 25 | 21 | 4 | 4 | 0.84 | 0.84 | 0.84 | 0.831 |
| Persian Gulf | 0 | | 10 | | | | | |
| Sea of Azov | 0 | | 0 | | | | | |

A closer observation of the results of the four study sites provides even deeper insights and better intuition about the effects of the different synthetic training data sets. Figure 4.15 shows two subsets of the North Sea Basin test site and the different model performances. The increase in accuracy from model-1 to model-2 is striking. Where model-1 only recognises small scale OWFs and fails to capture larger OWFs as one coherent unit, model-2 learned to detect OWFs with multiple sizes. The false positive detection until model-3 is the Triton Knoll OWF which in 2020 was under construction. Magnification of figure 4.15b) shows the regular grid-like pattern, which is already a strong large scale spatial feature that leads to the detection of this OWF site, which due to its deployment stage, is not part of the ground truth data set.

Figure 4.16 shows two subsets of the East China Sea test site. Similar to the North Sea Basin, the increasing ability of multi-scale OWF detection is clearly visible by comparing model-1 and model-2. Typically for the East China Sea test site are the near cost locations of OWF which bring them close to harbour infrastructure. This infrastructure is still included

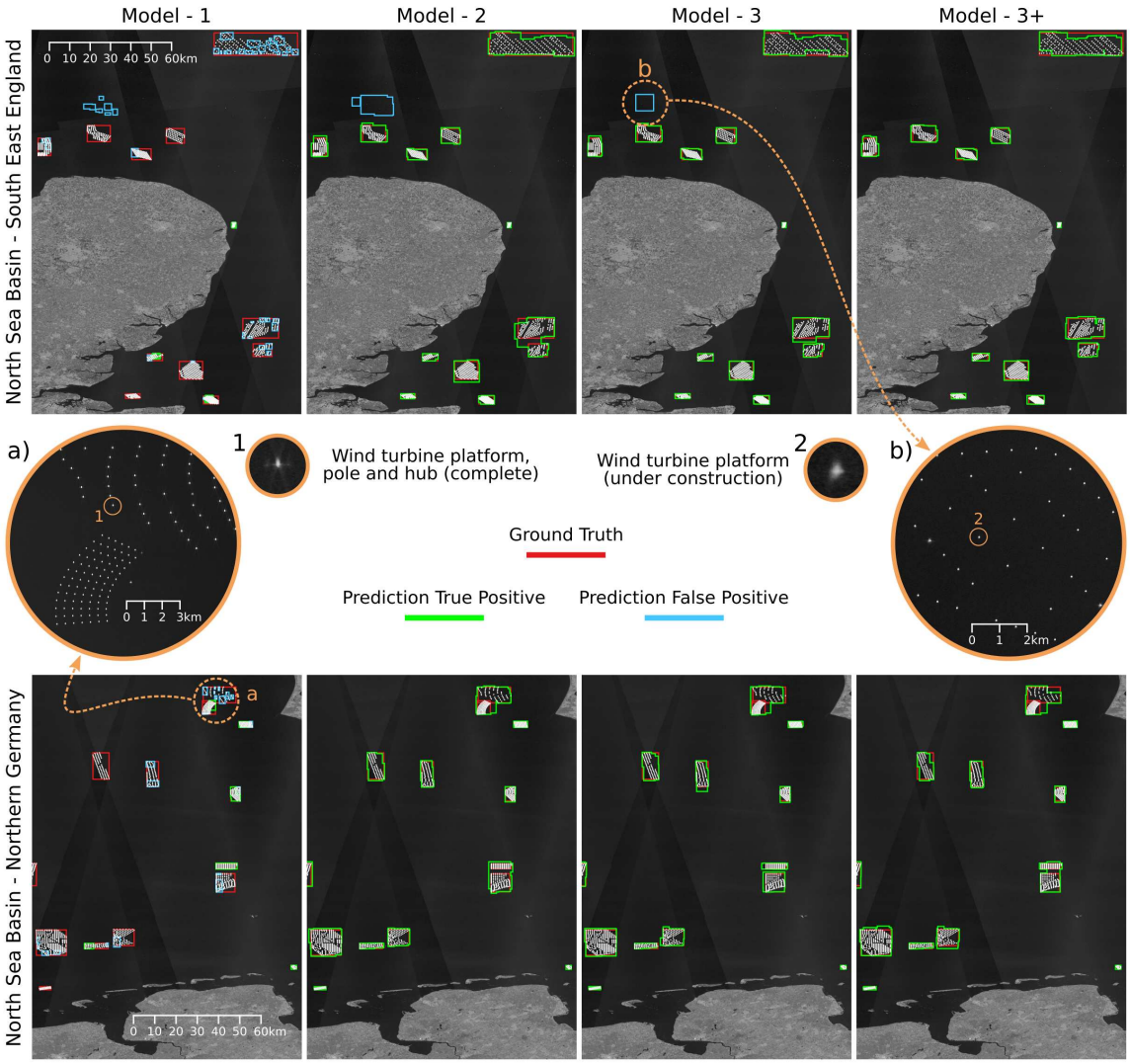



**Figure 4.15:** Progressive detection results of the four trained models for two examples from the North Sea Basin. Magnification a) shows readily deployed offshore wind farm (OWF), and b) OWF under construction. Changed after Hoeser and Kuenzer (2022b, p. 180).

as false positive detections in model-2, see the example of the Hangzhou Bay area in figure 4.16. This behaviour of model-2 is relatable since the non-target scene element coast is made available in data set-3. In model-3 and 3+ the harbour structures which were falsely detected in the Hangzhou Bay area are no longer detected. This example demonstrates the earlier discussed decrease of false positive detections and the corresponding increase in precision by including non-target information in the synthetic training data set. Furthermore, the Jiangsu example in figure 4.16 shows the impact of the additional OWT model type for locations on tidal flats. The OWF cluster built at the shoreline is situated on tidal flats. Magnification 4.16b) and b4) show the typical less distinct radar signature of an OWT in such an environment. Model-3, which is not aware of this type of OWT signature but is trained to reject other marine platforms like oil rigs, decides to reject these parts of the OWF cluster. In model-2, these OWTs were included since the non-target type oil rigs were not part of

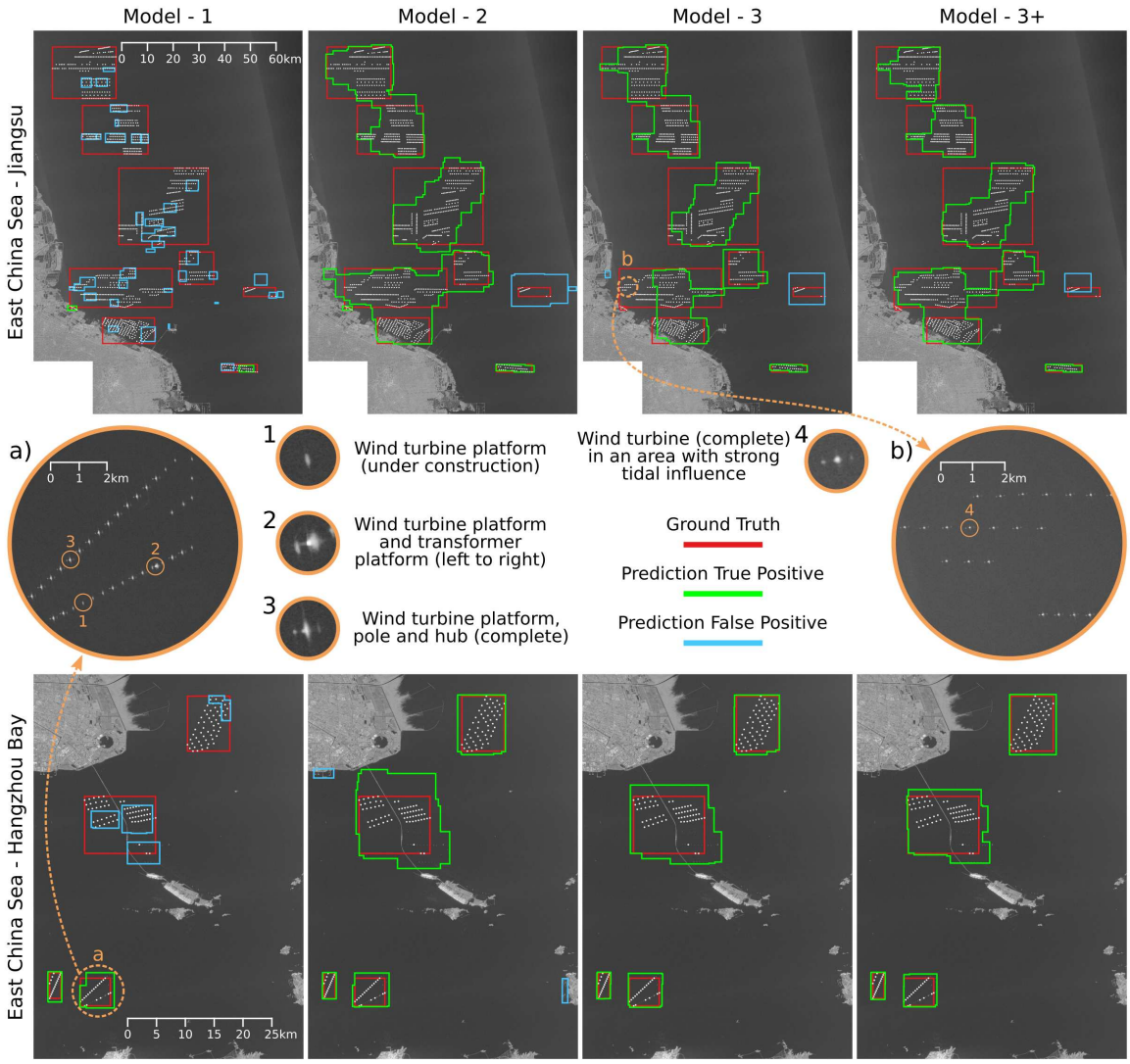



**Figure 4.16:** Progressive detection results of the four trained models for two examples from the East China Sea. Magnification a) shows an offshore wind farm (OWF) partly readily deployed and under construction, as well as a transformer substation; b) shows an OWF on a tidal flat. Changed after Hoeser and Kuenzer (2022b, p. 181).

the data set-2. However, in model-3+ after introducing the new OWT type in data set-3+, the CNN is forced to learn the detailed differences between oil rigs and this special type of OWT radar signature. Only with this addition of the new OWT type, model-3+ is able to enclose these turbines again fully.

In figure 4.17 the effects of non-target training examples become even more evident. Until model-2, accumulations of oil rigs in the Persian Gulf are getting detected. As soon as they are introduced as non-target training examples in data set-3, model-3 no longer detects them. However, in model-3+ some seem to appear again. This is due to the new OWT type, which on a smaller spatial scale looks less different compared with an oil rig than the other more distinct OWT types. Nevertheless, the effect of changing the synthetic training data is clearly visible in the model performance. Since the Sentinel-1 data is ingested unfiltered

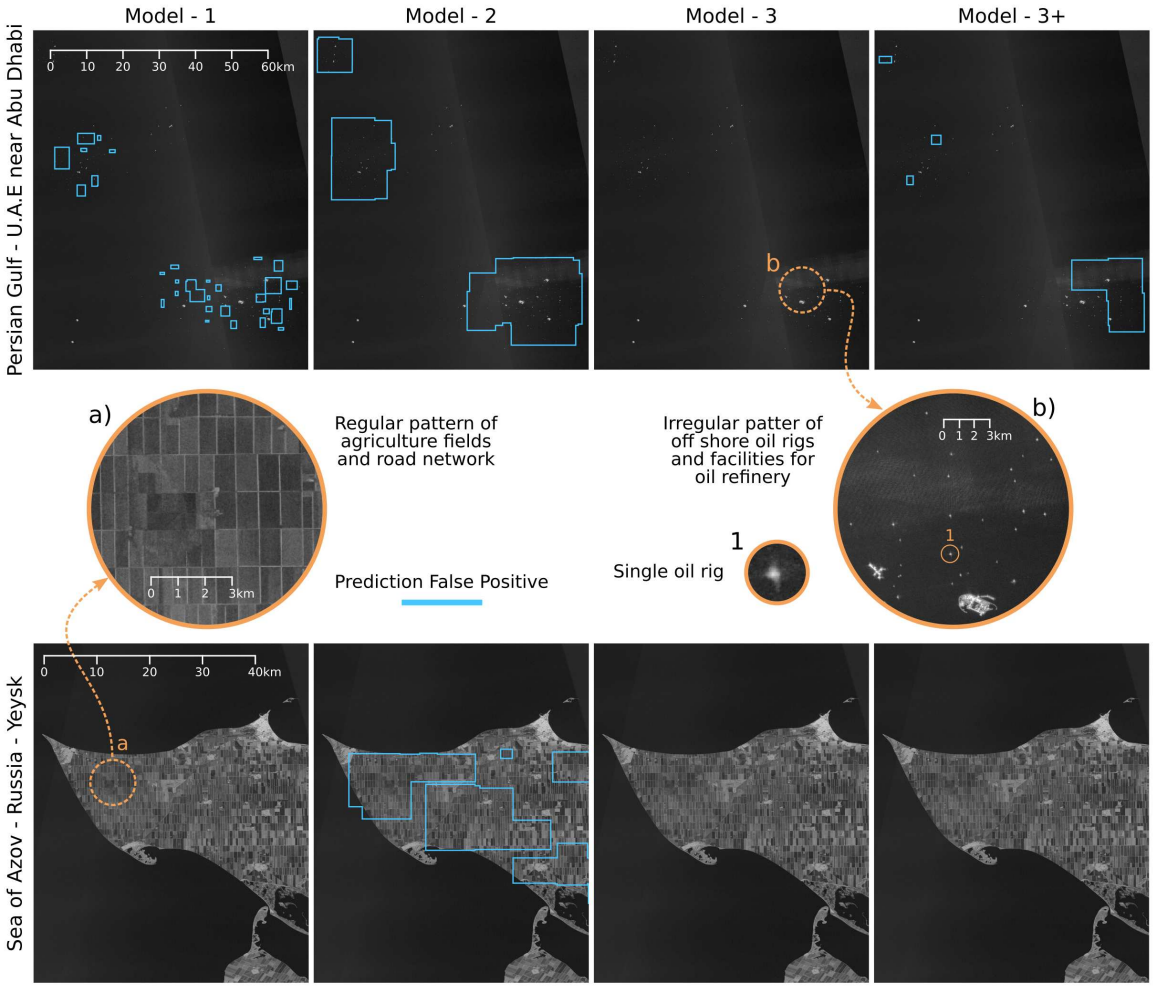



**Figure 4.17:** Progressive detection results of the four trained models for two examples from the Persian Gulf and the Sea of Azov. Magnification a) shows rectangular fields and a grid-like road network, which is a potential false positive feature; b) shows oil rigs and industrial facilities as potential false positive examples. Changed after Hoeser and Kuenzer (2022b, p. 182).

into the CNN models, not only offshore areas but also parts of onshore coastal areas are getting processed. In onshore environments, specific patterns can occur, which can be misleading for a CNN model. To understand what happens in this less intuitive example, the stepwise changing representations in the data sets help to understand which features each model has learned and is looking for in the data. Model-2 has learned to not only look for grid-like spatial patterns on a small scale but also on a large scale since large sized OWF types are included in data set-2. The spatial patterns of large gridded structures, which are typical for the turbine distribution of an OWF do also occur in the road network. Something that seems like an easily avoidable mistake is a very promising insight. It proves that the intentionally embedded grid-like pattern is an important feature for the model to detect OWF. Otherwise, the CNN would not have learned it. However, this information is misleading when the trained model is deployed for a real-world application. Only by adding land as

a non-target example in the data sets 3 and 3+, the model becomes aware of the fact that it has to look for grid-like structures only in a marine environment. This last example is very important to demonstrate that the intended knowledge transfer from a human being to a machine was successful since similar patterns that a human would use are important for the machine to fulfil its task.

## 4.3 Discussion

The two main objectives of the SyntEO framework are to enable researchers to generate large annotated remote sensing data sets and to transfer knowledge from a human Earth observation expert to a machine in order to build a narrow artificial intelligence to solve a specific task.

The introduced theory of the SyntEO framework and the proof of concept study demonstrated that by using SyntEO it is possible to start a knowledge transfer from a human domain expert to a machine. Once an ontology is set up and a processing backend connected, a data set can be produced on demand. Four consecutive data sets were generated in the proof of concept study, the largest consisting of 90,000 image-annotation pairs. After the ontology is formulated and the processing backend implemented, the technical generation of such a data set lasted 2.6 hours on a machine with 4 Intel Xeon Platinum 8260 CPUs and 2.40GHz running 190 parallel threads. This nearly instantaneous generation of training data allows for complex investigations of the target object. Furthermore, it enables a researcher to allocate more resources to answering geoscientific research questions instead of labelling ten thousand images with millions of target objects.

An argument against the SyntEO approach can be that instead of a rather complex workflow of generating synthetic images and training a deep learning model, the embedded knowledge in the ontology can directly be used as a rule-based model to fulfil the task at hand. A deep learning model would therefore be an unnecessary layer of complexity that can be avoided. The counterargument is that the ontology holds much information about individual features, but the highest density of information is only achieved through the composition of these features into a synthetic scene and image. Following that, most of the spatial information is still implicitly embedded in the ontology. However, by generating synthetic images upon the ontology, a model that uses these images indirectly learns the implicit knowledge of the ontology that otherwise would not be accessible. This becomes particularly clear with non-target scene objects. The provided examples and proof of concept study clearly demonstrate that a well-described target alone is not enough to train a generalised model. A definition of what not to detect is also necessary to minimise false

detections. In the proof of concept study, some non-targets were only loosely described, and much of their representation was taken from template data. This template data is extremely important and just as much a part of a synthetic scene as procedural information. However, the inherent characteristics of the template data are not explicitly described in the ontology and thus only accessible for a model which learns from the ontology but not for a model which is the ontology. The challenge is to express knowledge as explicitly as necessary to provide information to a machine capable of extracting implicit information from the knowledge representation. This knowledge transfer was proven to be possible for the OWF example with the proposed SyntEO framework.

Nevertheless, the question remains open as to how far SyntEO is transferable for other tasks besides the investigated OWF detection. The demonstrated examples show that parts of the ontology and the processing backend can be reused. The more the framework is used to generate synthetic remote sensing data, the more entities have been described or generated, which means that transferability can enter a self-reinforcing cycle. An advantage of the SyntEO ontology is that the same building block is reused over and over again to formalise the expert knowledge. Thus exchanging descriptions is supported by design since the same base structure is used. The more critical part is the processing backend. Here it is up to the developers of the respective processing backend to design it in a modular way that individual components function independently and prepare interfaces for adding new components. The stepwise changed data sets generated in the proof of concept study demonstrate that such a modular design and development is technically feasible. Be that as it may, only the further use and development of the framework can finally answer the question of how transferable the framework is.

What is important to mention as a potential downside of the framework is that the SyntEO ontology itself is not self-regulating. Rich and Knight (1991) describe the following dilemma:

1. “An AI system must contain a lot of knowledge if it is to handle anything but trivial toy problems.
2. But as the amount of knowledge grows, it becomes harder to access the appropriate things when needed, so more knowledge must be added to help. But now there is even more knowledge to manage, so more must be added, and so forth” (p. 27).

From the perspective of the SyntEO framework, this means that a domain expert runs the risk of getting lost in the formulation of the ontology. This has to be prevented at an early stage. Parts of this problem can be mitigated by the development of a comprehensive processing backend. This allows knowledge to be expressed more compactly in the ontology,

and the technical implementation is outsourced entirely to the processing backend. For example, instead of describing single components of how a polygon has to be generated in the ontology, the ontology can express that a polygon with four to six sides is necessary, and the processing backend takes over the rest. This supports the domain expert in not losing sight of the actual goal, the representation of crucial knowledge.

Furthermore, before starting to formulate the ontology, the task of the narrow intelligence should be defined, and the associated input data which is used to solve this task has to be well understood. After that, semantic descriptions of what should be in the ontology were found to be very helpful to set up a running thread while diving into the ontology formulation. Finally, continuously revising the ontology while drafting it is necessary to check if the currently formulated entity really adds value to the ontology, which the model can later learn. Thus, the knowledge representation process already starts before the ontology is formulated, with the description of the problem which the intelligent agent has to solve. Everything that is included in the ontology has the task to contribute to the overarching goal of finally building a narrow artificial intelligence, and as the name suggests, this goal can not be reached by a global but also a narrow ontology.

# Chapter 5

## DeepOWT - A global Offshore Wind Turbine Data Set*

The theoretical background and methodological development presented in the previous chapters are now brought together to solve the original problem: Deriving offshore wind energy infrastructure from remote sensing data. The final product will be a data set of single locations of offshore wind energy infrastructure, primarily offshore wind turbines. In addition to the spatial information, temporal information is provided for each object's location. It describes the temporal development of the deployment stages in three stages. These stages describe the first appearance of an object at a site, whether that object is under construction or whether it has been completed. Besides these model predictions, two comprehensive spatial and temporal ground truth data sets are used to evaluate the spatial and temporal information of the data set.

Figure 5.1 describes the entire workflow. By using the data acquisition pipeline introduced in section 4.2.1 Sentinel-1 median composites are acquired. Upon this data and other optical remote sensing data and publicly available information, the ground truth data sets of single offshore wind energy infrastructure objects and their temporal development are mapped by visual examination. Two synthetic training data sets are generated using the developed SyntEO framework. One synthetic training data set is used to optimise a CNN, which detects potential OWF areas globally. The other synthetic training data set optimises a second CNN, which searches for individual offshore wind energy infrastructure objects within the potential OWF areas detected before. After single object locations have been detected, their temporal development of the radar signal's spatial pattern is investigated to classify the five-year deployment stages from July 2016 until June 2021. In the last step, a sophisticated location refinement is performed to refine the bounding box detections to

*Parts of this chapter have been published in Hoeser et al. (2022)

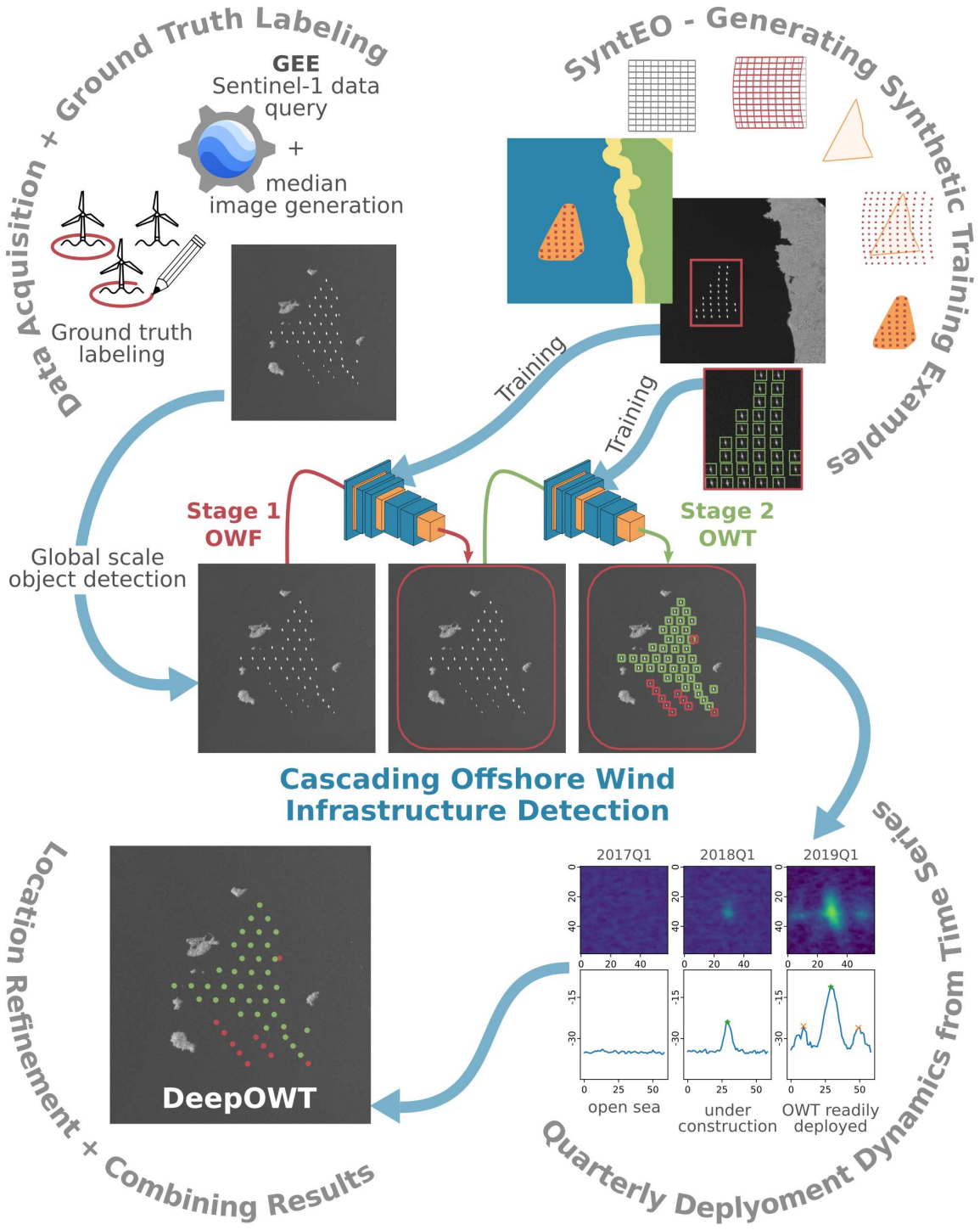



**Figure 5.1:** Overview of the workflow to derive the global multi-temporal DeepOWT data set of offshore wind energy infrastructure. A global stack of Sentinel-1, three month median composites are used as input for a cascade of two convolutional neural networks (CNNs) which are optimised with synthetic training data generated with the SyntEO framework. The first CNN detects entire offshore wind farm areas, the second single offshore wind energy infrastructure objects. A subsequent time series analysis derives the quarterly deployment stages for each detected object and combines spatiotemporal information to the final DeepOWT data set. Changed after Hoeser et al. (2022, p. 4256).

point locations. The temporal deployment stage information is added to these point locations to constitute the DeepOWT data set. Finally, the DeepOWT data set is evaluated and compared to two other sources which provide similar information about offshore wind energy infrastructure.

## 5.1 Data

The data to be investigated has a global spatial extent and a five-year temporal period. The precise spatial extent is the global coastline with a 200 km buffer towards the sea from the shoreline. The corresponding geometry of this area has already been prepared and optimised for processing by creating the processing grid in section 4.2.1. This processing grid and its associated data acquisition pipeline are reused for the global detection of OWFs. Similar to the proof of concept study, a three-month median composite of Sentinel-1 IW GRD products with a VH polarisation, both orbit directions and a pixel spacing of 10 m is generated. Before downloading the median composite of the global coastline, the pixel values were downscaled to 8-bit integers and subsequently downloaded as described earlier in section 4.2.1. For offshore wind energy infrastructure detection, a single global input data set of a median composite combining images from 01.04.2021 until 30.06.2021 (2021Q2) is acquired. For the temporal analysis of the deployment dynamics, quarterly median composites beginning in July 2016 are necessary to cover the five-year temporal period. To avoid downloading the median composites for the entire global coastline for all of these earlier periods, the median composites are only generated and downloaded in areas of OWFs which were detected in 2021Q2. For these areas, nineteen three-month interval median composites were generated and downloaded, which cover the period from 01.07.2016 until 31.03.2021. The same acquisition pipeline is used for these areas as for the global acquisition. Thus, twenty input data sets are generated. One data set from 2021Q2 with median composites that together cover the entire global coastline, and nineteen data sets of each quarter beginning with 2016Q3 with median composites of each OWF area, detected in the 2021Q2 imagery.

Besides the input data, ground truth data is necessary for the data set evaluation and as a benchmark for comparison of the DeepOWT data set with other data sources which provide offshore wind turbine information. Therefore extensive ground truth data sets were generated manually. For all ground truth data sets, the earlier introduced two test sites, the North Sea Basin and the East China Sea, are examined, see figure 5.2. As discussed in section 3.2, the East China Sea test site is the more challenging, with many OWFs under construction and more frequent near coast locations which bring them close to other marine infrastructures and small islands. Additionally, the acquisition modalities of the Sentinel-1

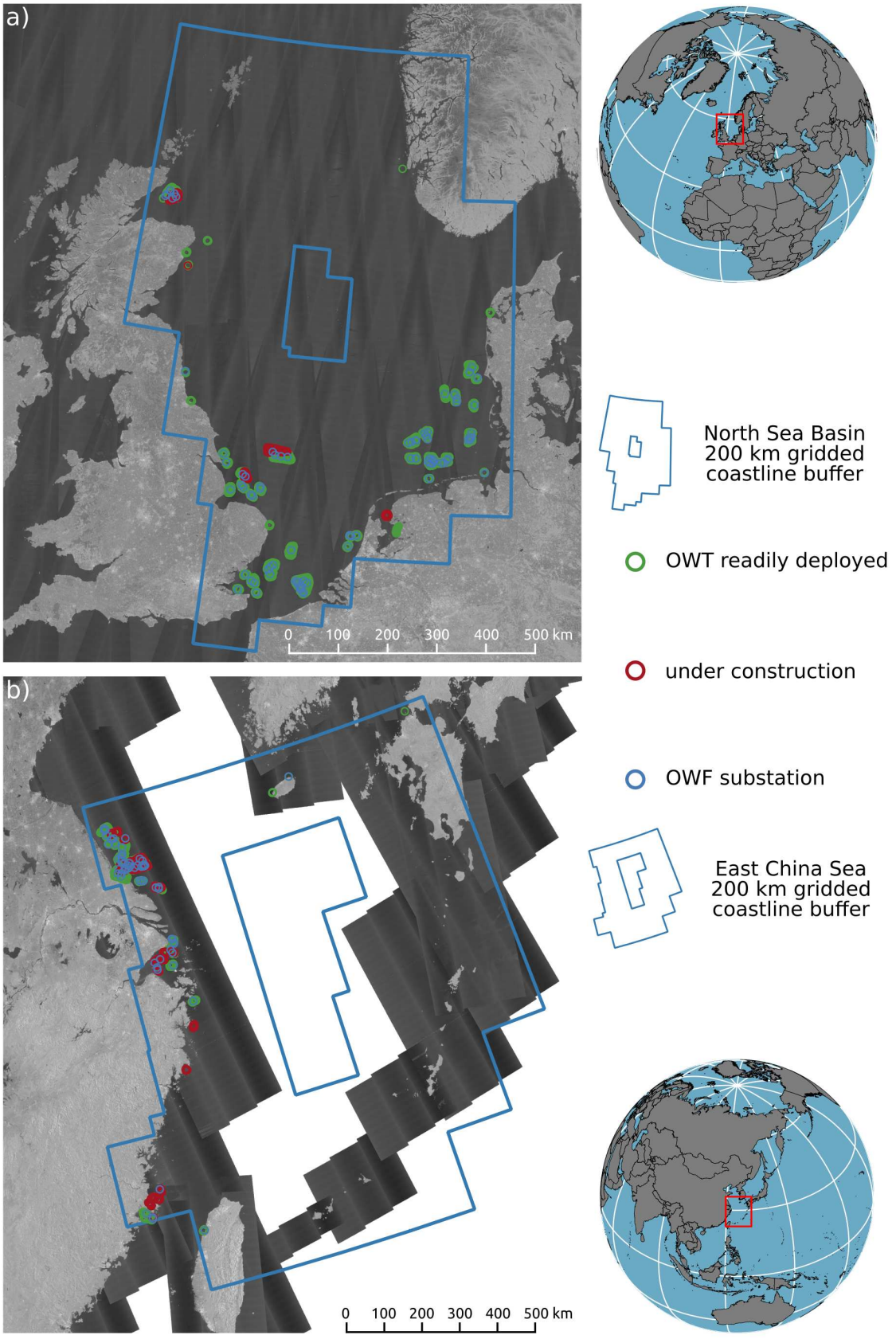



**Figure 5.2:** Overview of the two ground truth test sites (North Sea Basin and East China Sea) and the single ground truth labels for each offshore wind energy infrastructure object (offshore wind turbine (OWT), offshore wind farm (OWF)). Source: Hoeser et al. (2022, p. 4262).

mission are different for both sites. At the North Sea Basin, data is acquired for both orbit directions, ascending and descending. For the East China Sea, mostly only the ascending orbit is used, resulting in fewer acquisitions in a three-month interval which can cause less clear median composites. These differences make them suitable test sites for checking the detection and classification performance at different levels of complexity.

The Sentinel-1 median composites were used as a primary source to map ground truth offshore wind energy infrastructure locations manually. The annotation task was supported by investigating additional RGB imagery coming from the Sentinel-2 mission and Google Earth. Differences in the platform, whether the OWF is under construction or readily deployed, or it is an OWF substation, were also examined. For complicated cases, the precise classification of the platform type was derived from official planning documents or information extracted from news portals or descriptions of the OWF project by the operator.

The first two ground truth data sets were mapped for 2021Q2 to assess the object detection performance of offshore wind energy infrastructure, see figure 5.2. All offshore wind energy infrastructure locations in the North Sea Basin and the East China Sea are described in a separate ground truth data set with one of the three classes: OWT, platform under construction, and OWF substation. Thereby, the location of an object's centre is mapped with a point coordinate. These ground truth data sets contain 7,198 hand-annotated object locations. Similar ground truth data sets were mapped for the fourth quarter of 2019. They contain 5,290 hand-annotated object locations and will later be used to compare the performance of the data set derived in this work and other sources of global information about offshore wind energy infrastructure. These additional ground truth data sets were generated since all data sets that were compared with each other share the period 2019Q4. As before, for these ground truth data sets, all offshore wind energy infrastructure locations in the North Sea Basin and the East China Sea were annotated.

To assess the classification of the platform type over time, about 15% of the locations of the 2021Q2 ground truth data sets were randomly selected. All nineteen quarterly periods between July 2016 and March 2021 were investigated for these locations, and the corresponding deployment stage was labelled. The four possible labels are: OWT, platform under construction, OWF substation, and open sea. This ground truth data set contains 20,520 hand-annotated labels for the North Sea Basin and East China Sea test sites. Table 5.1 provides a detailed overview of all ground truth data sets and their characteristics as described in this section.

**Table 5.1:** Summary of the content of all ground truth data sets, the periods they cover and the number of objects for each class (offshore wind turbine (OWT)). The abbreviation of the *Site* column are North Sea Basin (NSB), and East China Sea (ECS)). * denotes the numbers for the start period 2016Q3 and the final period in 20201Q1 of the ground truth time series. † is the overall number of hand labelled temporal objects of all 19 periods in the ground truth time series. Source: Hoeser et al. (2022, p. 4259).

| Site | Time stamp | OWT | Construction | Substation | Open sea | ∑ label |
|---|---|---|---|---|---|---|
| NSB | 2021Q2 | 4,016 | 253 | 85 | - | 4,354 |
| ECS | 2021Q2 | 2,208 | 574 | 62 | - | 2,844 |
| NSB | 2019Q4 | 3,571 | 172 | 78 | - | 3,821 |
| ECS | 2019Q4 | 1,208 | 214 | 47 | - | 1,469 |
| NSB | 2016Q3-2021Q1 | 352-583* | 59-47* | 12-19* | 227-1* | 12,350† |
| ECS | 2016Q3-2021Q1 | 40-311* | 12-87* | 3-11* | 375-21* | 8,170† |

## 5.2 Global Offshore Wind Energy Infrastructure Detection with Deep Learning

### 5.2.1 Generation of Synthetic Training Data Sets

The overall design to detect offshore wind energy infrastructure on a global scale is a cascading approach of two subsequent CNN object detectors. The first stage of the cascade detects larger areas that potentially contain OWFs. The second stage closely investigates these areas in order to extract single offshore wind energy platforms. The idea is familiar with how the visual cortex of mammals recognises objects (Hubel and Wiesel, 1962; Felleman and Van Essen, 1991). First large scale features are recognised, like the agglomeration of grid-like patterns in the open sea. Then, at a closer look by a second CNN, single objects or small scale features are getting identified. In this case, the different types of offshore wind energy platforms. Two training data sets are necessary to optimise this CNN cascade. The first data set has a focus on large scale features of OWFs. The second data set has to represent small scale features of single platform types. Also, their annotation is different. Where the first data set describes the outer boundary of an entire OWF with a single bounding box, the second data set describes each location of each platform with an individual bounding box and additionally with a corresponding class label for different platform types.

The introduced SyntEO framework was used to generate these two data sets. Figure 5.3 shows synthetic training examples of the first data set. The generation of this data set is based on the developments presented in the proof of concept study in the previous chapter 4. However, in addition to OWFs with an outer boundary described by a polygon, linear OWFs were generated, see figure 5.3a+b) for some examples. Besides that, the target and non-target types and balancing of the data set-3+ as introduced in section 4.2.4 have been

reused. The data set for optimising the stage-1 CNN, for potential OWF area detection, contains 90,000 image-annotation pairs. Also, the handling of the image size during training is the same as in the proof of concept data set-3+. The images of size $2,048 \times 2,048$ pixels are downscaled before training to the input size of the stage-1 CNN, $1,024 \times 1,024$ pixels.

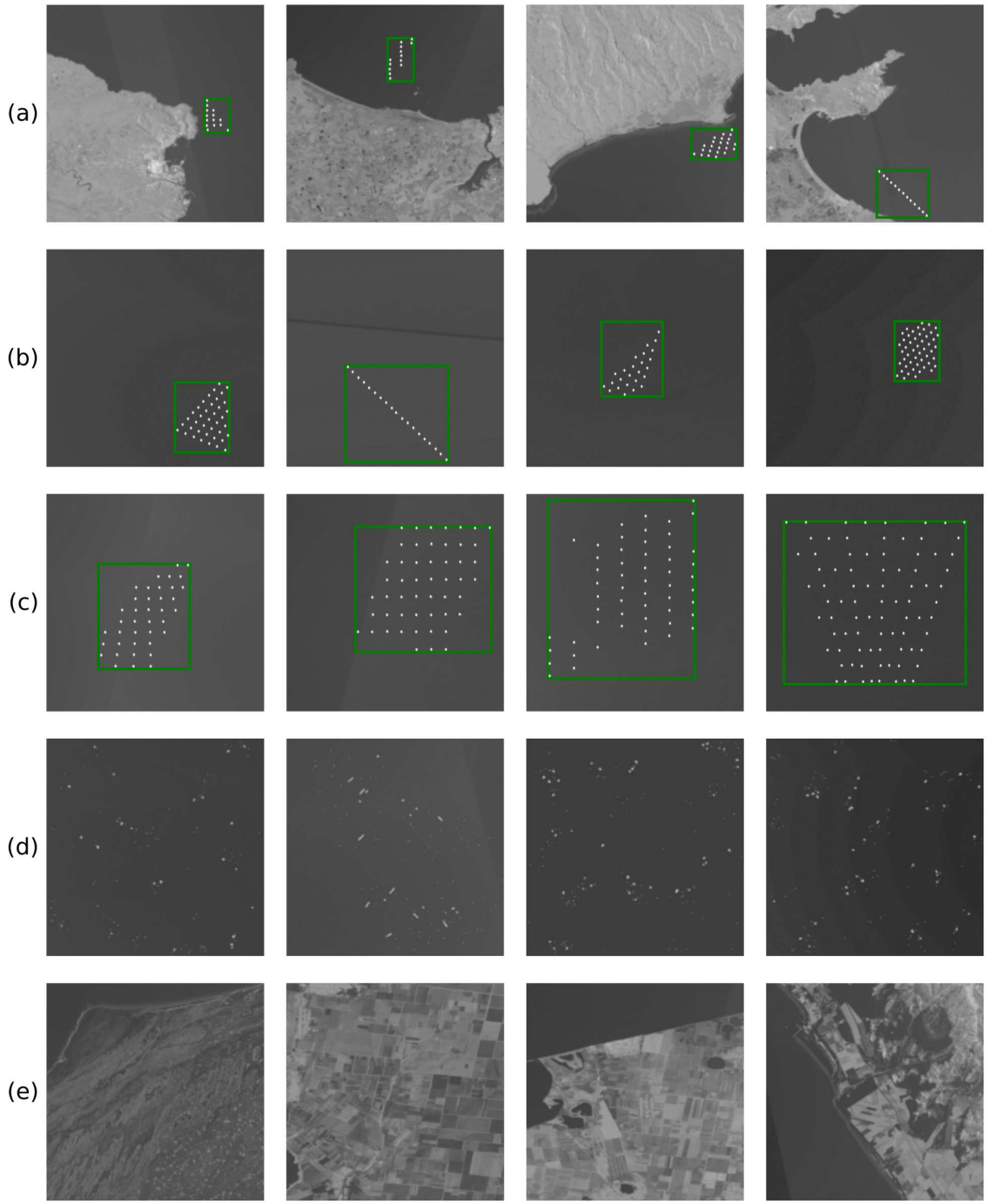



**Figure 5.3:** Examples of synthetically generated imagery for offshore wind farm (OWF) detection with their corresponding annotation. a-c) show OWF of scales small, medium and large, respectively. d) shows non-target training examples of oil rigs. e) shows non-target training examples of the mainland.

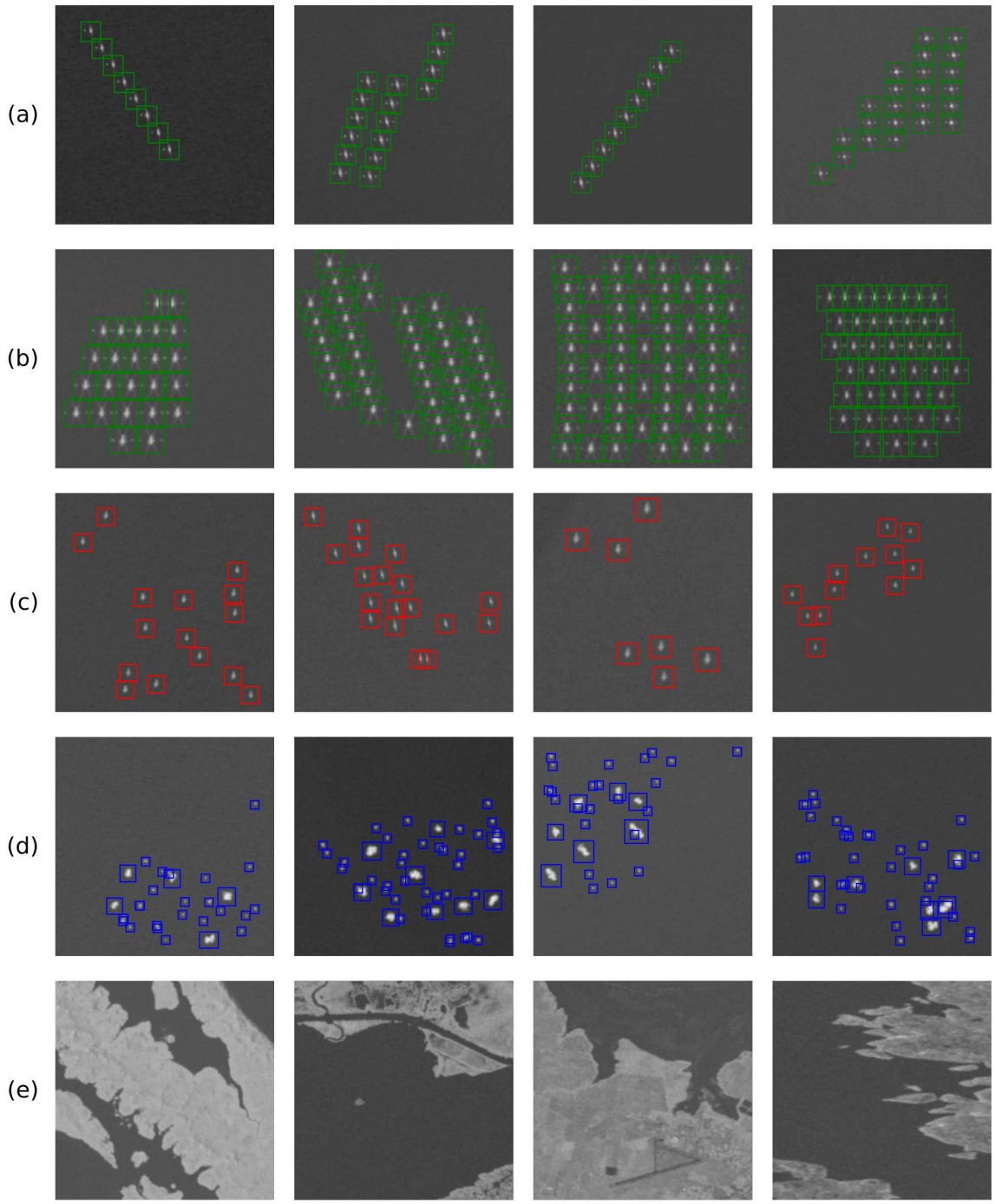



**Figure 5.4:** Examples of synthetically generated imagery for offshore wind energy infrastructure detection with their corresponding annotation. a+b) show readily deployed offshore wind turbine. c) shows offshore wind energy infrastructure under construction. d) shows rigs such as transformer platforms. e) shows non-target examples with mainland and coastal areas.

The most important change in the second data set for optimising the stage-2 CNN is the annotation of the target objects. Figure 5.4 shows, that instead of entire OWF areas, each single OWTs is annotated. Furthermore, single OWF substations and platforms that are not readily installed are two new additional classes besides the readily deployed OWT class, see

figure 5.4. Thus, a specific class label is provided for each object class in the second data set, along with the coordinates of the bounding box annotation. The second data set describes three classes, OWT, OWF substations, and platforms under construction. Furthermore, it contains a specific non-target image type that shows coastal areas for introducing potential false positives to the model during training. Thus, there are four image types in the training data set, each represented by equal shares. For this more comprehensible data set compared to the first one, a larger size has been chosen. Overall, the second data set contains 275,000 image-annotation pairs. Since the spatial features of single offshore infrastructure platforms appear on a small scale, the image size is $512 \times 512$ pixels. The stage-2 CNN for single platform detection has an input size of $512 \times 512$ pixels. That way, no downscaling before training is necessary, and the full spatial resolution of 10 m $\times$ 10 m can be used to detect the small scale objects, following the approach introduced in section 4.1.2.

The entire training of the two cascading CNNs relies on these two synthetic data sets. The training data sets were generated by employing the developed SyntEO framework. No auxiliary training data is used to optimise the two cascading CNNs to detect offshore wind energy infrastructure on a global scale.

### 5.2.2 CNN Architectures and Training

The ResNet-50 (He et al., 2016) convolutional backbone with a subsequent Faster R-CNN (Ren et al., 2015) is used as deep learning model architecture for object detection. For the first stage CNN, which is looking for potential OWF areas, the anchor ratio and scale factors of the RPN are the same as in the proof of concept study [0.5, 1, 1.5] and [0.25, 0.5, 1, 2, 3.5] respectively. The second stage CNN however, has to look for much smaller objects. Furthermore, the synthetic image and CNN input size changed both to $512 \times 512$ pixels to focus on small scale spatial features. The earlier introduced equation (27) was used to adjust the anchor scale factors of the second stage RPN to be [0.25, 0.5] with an anchor aspect ratio of [1]. The scale ratios relate to search areas with a real-world dimension of 640 m $\times$ 640 m to $1,280$ m $\times$ $1,280$ m.

The CNN architectures of stage-1 and -2 were built and optimised by using the TensorFlow deep learning framework (Abadi et al., 2015). The training scheme which was used in the proof of concept study in section 4.2.5 was found to be sufficient to optimise a similar model with similar training data. Hence the scheme was reused in both model trainings. The general set-up is a 95% train and 5% validation split of the synthetic training data sets. Both synthetic data set sizes are chosen to be large enough that each training example is only used once during the entire training. Thus only a single but long-lasting epoch is used and no data augmentation is applied to the training examples before entering the network.

The stochastic gradient descent method (Robbins and Monro, 1951; Bottou et al., 2018) with a momentum of 0.9 and a batch size of 4 is used for optimising both models. The learning rate is scheduled by using the cosine decay learning rate approach with a base learning rate of 0.01 after a warm-up phase of 1,000 steps (Loshchilov and Hutter, 2017).

The training process on four parallel Nvidia RTX 2080Ti GPUs took 6.6 hours for the stage-1 CNN and 4.8 hours for the stage-2 CNN. The finally optimised models were exported to make them available for building the upcoming object detection pipeline of offshore wind energy infrastructure for the entire Earth.

### 5.2.3 Object Detection and Postprocessing

As stated earlier, the object detection process is split into two stages. In the first stage, all Sentinel-1 median composites of the entire global coastline are used as input to the stage-1 object detector CNN. In this first stage, potential OWF areas are getting extracted. Therefore a relatively small detection score of 0.5 was chosen. Thus many potential OWF regions are getting detected by the first stage. The second stage looks closer into these potential OWF areas. The stage-2 object detector CNN detects three kinds of objects: OWT readily deployed, offshore wind energy platforms under construction and larger offshore platforms like OWF substations but also oil rigs or other kinds of persistent marine infrastructure. For these object classes to be recognised as valid detections in the second stage, a detection score of 0.75 or higher was chosen.

The core motivation of this cascade is that the designed artificial intelligence should approach the detection problem similar to a human remote sensing expert. To find offshore wind energy infrastructure, such an expert would search for obvious clusters of scatterers in spaceborne SAR median composites. If the cluster is obviously unstructured, the expert would drop it due to the knowledge that some kind of structured pattern is distinct for offshore wind energy infrastructure. However, some less clear cases need a closer look not to miss any clusters. This first decision is represented by the low detection score of the first stage. This means that more detections are made than necessary in this first stage, but it increases the certainty that no OWFs are overlooked. The human expert would then look into each preselected site more closely. If there are only non-offshore wind energy infrastructure platforms, the entire clusters can be dropped. In case most of the objects are OWTs, this cluster is truly an OWF. The technical implementation of this interpretation uses the detection results of the second stage to revise the first stage. In case that 90% or more of all platforms in a potential OWF area are non-OWT platforms, this potential OWF area is belatedly dropped. In case the site contains 15% or more OWTs, the entire OWF is considered a real OWF. By using this cascade of detection modules, a self-checking property

was implemented into the model cascade. This self-checking property maximises the number of true positive detections since many potential OWF areas are getting recognised. At the same time, it minimises the number of false positives since the second stage revises wrongly included areas due to small scale observations.

Furthermore, this approach supports a straightforward contextual decision for the object type OWF substations. Whenever a rig-like platform is detected in a group of OWT, it is classified as a OWF substation. This classification would not be that clear when otherwise, only the single rig object is detected, which can have the same radar signature as an oil rig in Sentinel-1 median composites. Only with the contextual consideration, OWF substations can be differentiated from other rig-like platforms. Overall, the cascading design of two interacting CNN object detectors was found to be highly effective in detecting small objects in large Earth observation archives. Basic concepts like Tobler's first law of geography reappear during the implementation. They demonstrate the consideration of geographic principles that have been taken into account during the design of artificial intelligence.

For the object detection of stage-2, the Sentinel-1 median composite tiles were tiled into images chips with a dimension of $512 \times 512$ pixels and an overlap of 50%. This results in multiple detections since the same object can appear in multiple images. To aggregate the detections, the cascading unification of bounding boxes as introduced in section 4.2.6

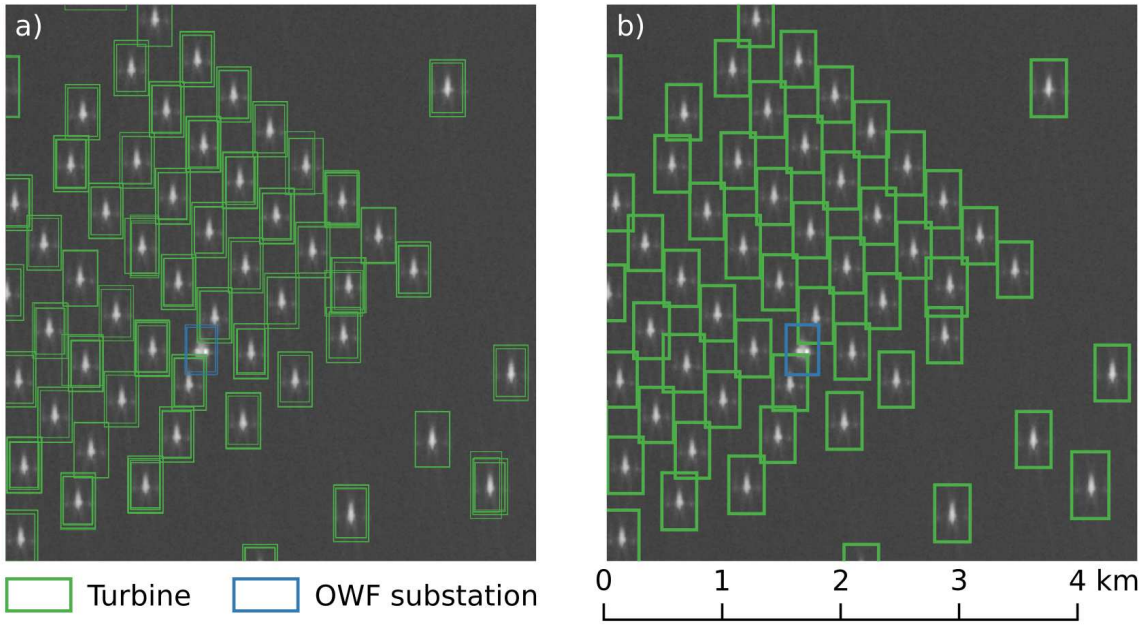



**Figure 5.5:** Unification of raw detection results in single bounding boxes for each object. a) raw, overlapping bounding box detection results of the second stage convolutional neural network. b) unified bounding boxes by taking their intersection over union (IoU) into account to preserve the detected classes and to keep individual objects separate from each other.

and equation (30) was applied to unify multiple bounding boxes of the same object to a single geometry with coordinates in the CRS WGS84. To further simplify the shape of the unified geometry, instead of the exact outer boundary of all detected boxes of one object, the bounding box of the unified geometry is the final location description. This forces all summarised shapes of a single object detection to be a rectangular box around offshore wind energy infrastructure location, see figure 5.5b). For further processing and data management, each bounding box derived that way is assigned an ID, in the following referred to as location ID.

To further increase the object localisation, the bounding boxes were refined to point locations, see figure 5.6a+b). Since not all bounding boxes are exactly centred around the object, the centroid of a bounding box would be too inaccurate. Furthermore, in some very dense OWT situations, a bounding box centred around the target OWT can also include a neighbouring OWT, island, or other marine infrastructure. Thus simply searching for the maximum backscatter amplitude in the entire bounding box can also lead to a wrong localisation. To find the precise centre of an offshore infrastructure object, the bounding box is used to crop the pixel values from the Sentinel-1 composite tile. Each column of such

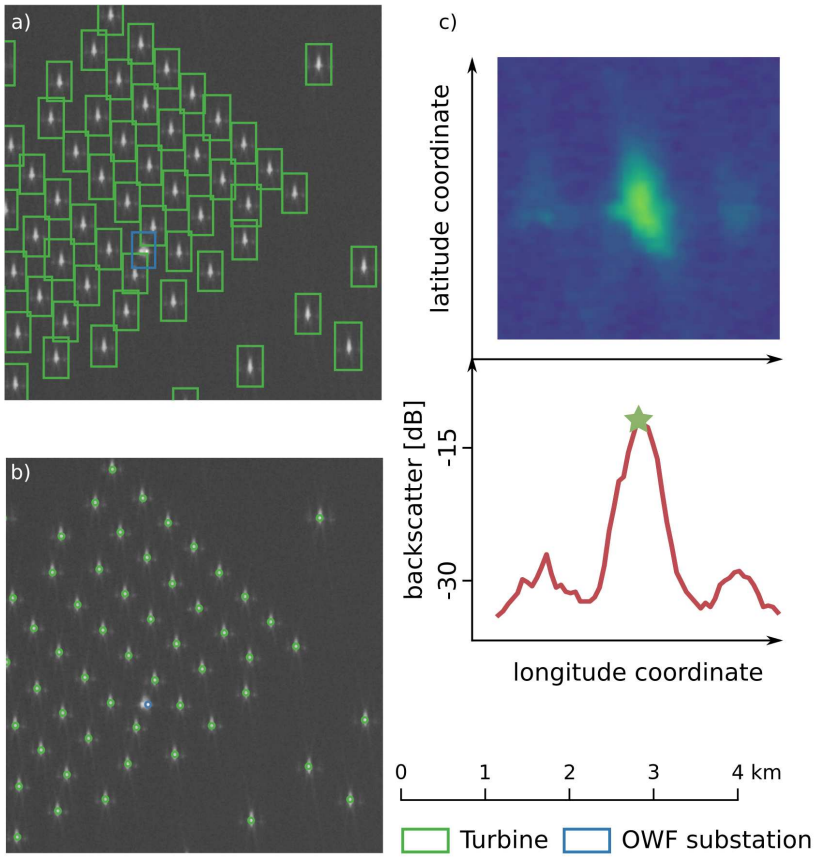



**Figure 5.6:** Location refinement from bounding boxes a) to a single point location b) by searching for a centred maximum amplitude in the radar signature of each detected object c).

an image crop is reduced to its maximum value. That way, a swath profile along the horizontal axis is generated, see figure 5.6c). A peak finder algorithm with a high prominence (Virtanen et al., 2020) is applied to this swath profile to find the longitude coordinate for the maximum value of this swath profile. In case multiple peaks are detected this way, the peak which is closest to the centre of the swath profile is chosen. The latitude coordinate is finally derived by searching for the peak value in the corresponding column of the image crop. Again, if several values in the column corresponding to this maximum value, the one that is closer to the centre of the image is selected. Thus, the refined point location of each offshore wind turbine corresponds to the pixel location with the maximum backscatter value closest to the image centre, see figure 5.6.

## 5.3 Deriving Temporal Deployment Dynamics

Besides detecting offshore wind energy infrastructure, their temporal deployment stages are of interest. This additional information provides the opportunity to analyse the temporal development of the offshore wind energy sector. In other studies, the first date when an object at an offshore wind energy infrastructure site was detected is chosen to be the date this object is considered as an OWT (Xu et al., 2020; Zhang et al., 2021b). This approach can lead to temporal overestimating of readily deployed infrastructures since construction periods, which can take up to one year, are not considered. Therefore it is necessary to differentiate if an object is under construction or readily deployed. The stage-2 CNN can already differentiate between these stages during object detection. However, it is only applied to the last investigated period 2021Q2. All other nineteen quarterly periods beginning with 2016Q3 are not investigated yet. Since the bounding boxes provided by the CNN cascade reduce the information content to a very distinct, small scale backscatter signal, it is possible to investigate the changes of this local signal to precisely derive the deployment stage.

The process relies on the previously introduced maximum swath profiles along the longitude axis for each object location, see figure 5.6c). These profiles were generated for all derived bounding boxes over all twenty quarterly periods. Figure 5.7 gives an impression of such a time series for an OWT. Visually, it can be seen that construction started in 2017Q3 and lasted until 2018Q2. Since 2018Q3 the OWT has been readily deployed and is operational. The changes in the backscatter cluster come from the initial construction of the foundation. Then in 2018Q3 pole, nacelle and blades are getting installed to complete the OWT. When the nacelle is installed, the two characteristic clusters from the layover and double bounce effect appear next to the main cluster of the OWT foundation in the radar signature. In the swath profile, the corresponding two peaks are clearly visible. To extract

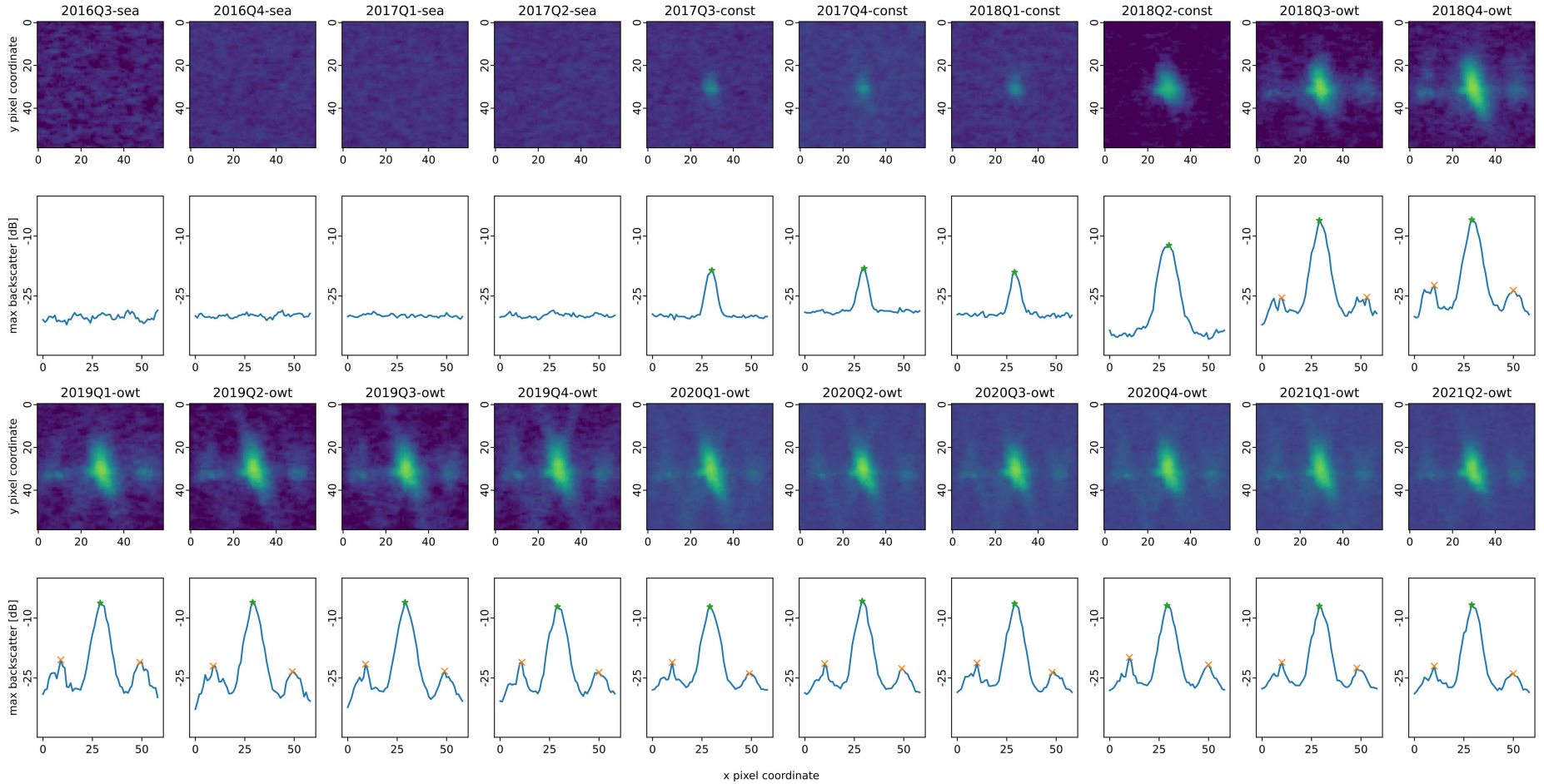



**Figure 5.7:** Time series of the radar signature for a deployment process of an offshore wind turbine (OWT) from June 2016 until July 2021 with a quarterly period. The deployment stages are classified by tracking changes in the peaks of the maximum swath profile of the location described by the bounding box coming from the convolutional neural network. The two lower peaks next to the main centre peak describe an OWT as readily deployed since they are coming from the layover and double bounce effect, which only appears when the pole and nacelle of a turbine are installed. Source: Hoeser et al. (2022, p. 4261).

all peaks automatically, the same peak finder algorithm as used before in section 5.2.3 for location refinement is applied twice. The first run has a higher prominence only to find the centre peak, and the second run has a lower prominence to find the neighbouring peaks. Of all peaks found by the second run, only the first peak to the left and the right of the centre peak are kept. All other peaks are dropped. With this procedure, peaks coming from closely built OWTs do not interfere with the target signature.

With this approach, the number of peaks, the peak amplitude, and also the peak width can be derived. Since the class of the last period, 2021Q2, is already provided by the CNN, the resulting temporal signal of the attributes can be investigated with prior knowledge. Thus the time series are analysed in a backwards-looking direction. For an OWT time series, if the number of peaks is reduced along with a decrease in centre peak amplitude, the earlier periods have to be under construction. If the main centre peak is entirely lost, all earlier periods only show the open sea. For an OWF substation, if both amplitude and width of the centre peak drop, the earlier periods are under construction. Similar to an OWT, if the centre peak disappears, the earlier periods have to be open sea. Thus, for a time series where the final class is under construction, if the centre peak disappears, construction has not started, and all of the previous periods have to be open sea. This straightforward classification of the time series is only possible due to the precise extraction and classification of object locations for the 2021Q2 period with the proposed deep learning approach.

The temporal classification process was applied to all object locations resulting in 19 additional classes besides the latest prediction of 2021Q2 made by the CNN. The total period of the data set is five years or twenty quarters, from July 2016 to June 2021. This time series information is connected via the location ID with the previously derived location information.

## 5.4 DeepOWT Data Set Summary

The combined spatiotemporal information about offshore wind energy infrastructure derived from the Sentinel-1 archive with deep learning constitutes the DeepOWT data set. It describes 9,941 globally distributed offshore wind farm infrastructure locations, most of them offshore wind turbines. In June 2021 the data set contains 8,885 OWTs, 204 OWF substations, and 852 offshore wind energy infrastructures under construction. Figure 5.8 gives a first impression of the spatial extent of the data set. It shows the outer boundaries of OWFs derived by the first stages of the cascading object detection pipeline and the refined point locations of single OWTs and OWF substations of two OWF clusters.

**Figure 5.8:** Global distribution of detected offshore wind farm clusters and two hot spot regions, the North Sea Basin and the East China Sea. The two offshore wind farm examples show an EU site, which is readily deployed, and a Chinese site, which in June 2021, is partly readily deployed and under construction. Changed after: Hoeser and Kuenzer (2022a, p. 5).

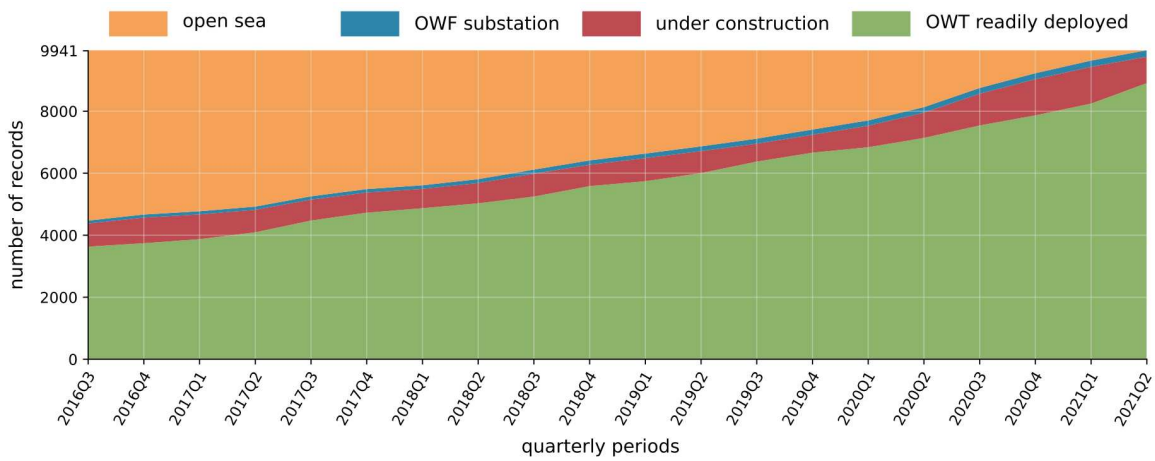



**Figure 5.9:** Temporal development of the 9,941 globally detected offshore wind energy infrastructure locations (offshore wind farm (OWF), offshore wind turbine (OWT)) from July 2016 until June 2021. Source: Hoeser et al. (2022, p. 4263).

Figure 5.9 gives a first impression of the temporal development of the number of different offshore wind energy infrastructure object types. The data set's temporal frequency is three months for each quarter of a year. The overall time span of the data set is five years, with twenty quarterly periods from July 2016 until June 2021. An in-depth spatiotemporal analysis of the DeepOWT data set is provided in the next chapter 6.

The data set is available as a single `.geojson` file with a small size of 4.1 MB, which allows flexible handling in GIS environments, even on mobile platforms and web applications. An impression of the entire data set is provided by the Coastal Explorer, a web mapping service of the German Aerospace Center (DLR)[1]. The data set file, along with all ground truth data sets, is publicly available on Zenodo[2].

## 5.5 Data Set Evaluation and Comparison

An evaluation was carried out to assess the performance of the methods used to derive the DeepOWT data set. During the assessment, a detection is considered a true positive TP if its location is within a 100 m radius of a ground truth object and the predicted class is the same as the class of this ground truth location. If a detection is made outside a ground truth area or the predicted class is wrong, this detection is a false positive FP. A ground truth location that is entirely undetected is considered a false negative FN. With this definition, the evaluation metrics can be calculated. As in section 4.2.6, precision Pr (31) and recall Rc (32) were used to calculate the F1 score (33). Furthermore, the average precision AP (35) was derived from the all point interpolated precision-recall curve $Pr_{interp}$ (34) (Padilla et al.,

[1] https://coastalx.eoc.dlr.de
[2] https://zenodo.org/record/5933967

2021) for the two test sites, the North Sea Basin and the East China Sea. In this assessment, the number of ground truth locations in the North Sea Basin is much higher than for the East China Sea, see table 5.2. In order not to distort a combined metric for both test sites by the over-represented data of the North Sea Basin, which is considered easier for offshore wind energy infrastructure detection, combined metrics are calculated using the macro averaging method after the scores have been calculated separately for each test site (Opitz and Burst, 2021)

$$\mathrm{Pr}_{\mathrm{avg}} = \frac{1}{n}\sum_{i=1}^{n}\mathrm{Pr}_i\,, \tag{36}$$

$$\mathrm{Rc}_{\mathrm{avg}} = \frac{1}{n}\sum_{i=1}^{n}\mathrm{Rc}_i\,, \tag{37}$$

$$\mathrm{F1}_{\mathrm{avg}} = \frac{1}{n}\sum_{i=1}^{n}\mathrm{F1}_i\,, \tag{38}$$

where $n$ is the number of test sites.

The first evaluation was carried out to assess the object detection performance with the CNN cascade on the 2021Q2 Sentinel-1 median composites. The corresponding ground truth data sets are of the North Sea Basin and the East China Sea for the period 2021Q2, see table 5.1. Table 5.2 and figure 5.10 provide the results of the assessment. The combined performance on OWT detection is an $\mathrm{F1}_{\mathrm{avg}}$ of 0.992, which expresses nearly perfect detection of offshore wind turbines. The precision-recall curves in figure 5.10 and corresponding

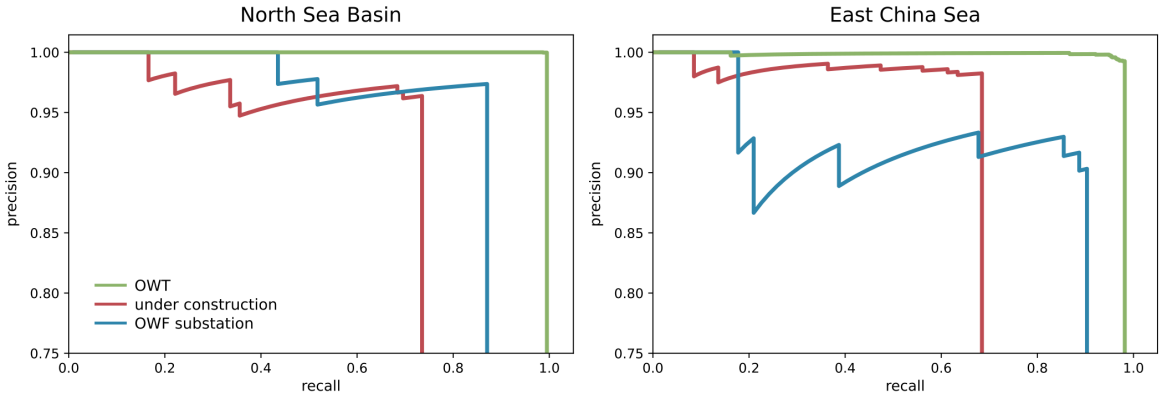



**Figure 5.10:** Precision-recall curves for the classes offshore wind turbine (OWT), offshore wind farm (OWF) substations, and offshore wind energy infrastructure under construction of the DeepOWT data set for the two ground truth sites the North Sea Basin and the East China Sea. See table 5.2 for the corresponding AP values. Source: Hoeser et al. (2022, p. 4264).

**Table 5.2:** Overview of all calculated metrics (ground truth (GT), true ppositive (TP), false ppositive (FP), false negative (FN), recall (Rc), precision (Pr), average precision (AP)) for the convolutional neural network cascade detections on the 2021Q2 global Sentinel-1 median image for each class separately. The detections were evaluated with the 2021Q2_nsb (North Sea Basin), and 2021Q2_ecs (East China Sea) ground truth data sets. Source: Hoeser et al. (2022, p. 4263).

| **Site** | **Class** | GT | TP | FP | FN | Pr | Rc | F1 | AP |
|---|---|---|---|---|---|---|---|---|---|
| North Sea Basin | OWT | 4,016 | 3,996 | 1 | 20 | 1.0 | 0.995 | 0.997 | 0.995 |
| North Sea Basin | under construction | 253 | 186 | 7 | 67 | 0.964 | 0.735 | 0.834 | 0.72 |
| North Sea Basin | OWF substation | 85 | 74 | 2 | 11 | 0.974 | 0.871 | 0.919 | 0.859 |
| East China Sea | OWT | 2,208 | 2,168 | 16 | 40 | 0.993 | 0.982 | 0.987 | 0.981 |
| East China Sea | under construction | 574 | 393 | 7 | 181 | 0.982 | 0.685 | 0.807 | 0.678 |
| East China Sea | OWF substation | 62 | 56 | 6 | 6 | 0.903 | 0.903 | 0.903 | 0.853 |
| | | | | | | $Pr_{avg}$ | $Rc_{avg}$ | $F1_{avg}$ | |
| Combined | OWT | 6,224 | 6,164 | 17 | 60 | 0.996 | 0.988 | 0.992 | |
| Combined | under construction | 827 | 579 | 14 | 248 | 0.973 | 0.71 | 0.821 | |
| Combined | OWF substation | 147 | 130 | 8 | 17 | 0.938 | 0.887 | 0.911 | |

AP values suggest a slightly better performance for the North Sea Basin (AP = 0.995, F1 = 0.997), confirming the impression that the East China Sea (AP = 0.981, F1 = 0.987) is the more complicated environment. However, the differences in the scores are minimal, which demonstrates that object detection works equally well on both test sites.

The detection of OWF substations reaches the second highest $F1_{avg}$ score of 0.911. From table 5.2 it becomes clear that for this class, false negatives FN are the weak spot, even though the performance is still very high. Compared to OWTs, the FN-rate is higher and, with it, recall Rc is lower. One explanation for this is that in the East China Sea test site, many OWFs are under construction, and often the substations are the first which are getting constructed. At an early stage of an OWF project, they can thus be spatially separated from the grid-like turbine foundations, which do not yet cover the entire area of the OWF. Thus the estimated OWF boundary of the stage-1 object detector does not include them. Following that, they are not detected by the stage-2 object detector. However, they will be included when the OWF reaches a more mature construction stage.

The most challenging object type is the platform under construction class. Here the object detection reaches an $F1_{avg}$ score of 0.821. Table 5.2 and figure 5.10 clearly show, that if a detection of a platform under construction is made by the cascading CNN detector, this detection is mostly correct, resulting in high precision Pr scores. However, not all platforms which are under construction are getting detected, hence recall Rc is comparatively low for this class. The higher FN-rate of the under construction class can be explained by the less distinct radar signature of a new build platform foundation which can easily be

confused with other marine infrastructure since it lacks most of its typical construction elements. Another reason is the chosen quarterly interval of the median composites. Suppose a foundation was built at the end of a quarter. In that case, its signature is only visible in a few images of a quarterly stack and thus not included in the generated median composite. Instead this median composite will show open sea for that period even when a foundation is already installed for a couple of days or weeks. This effect is only for a short time, and the right prediction is made for the next quarter. This makes this class more challenging to detect and easy to miss. These are important insights into the behaviour of the cascading CNN detector and essential for further developments and for the interpretation of the derived data set. Nevertheless, even for the least well performing class, the overall $F1_{avg}$ score is above 0.8.

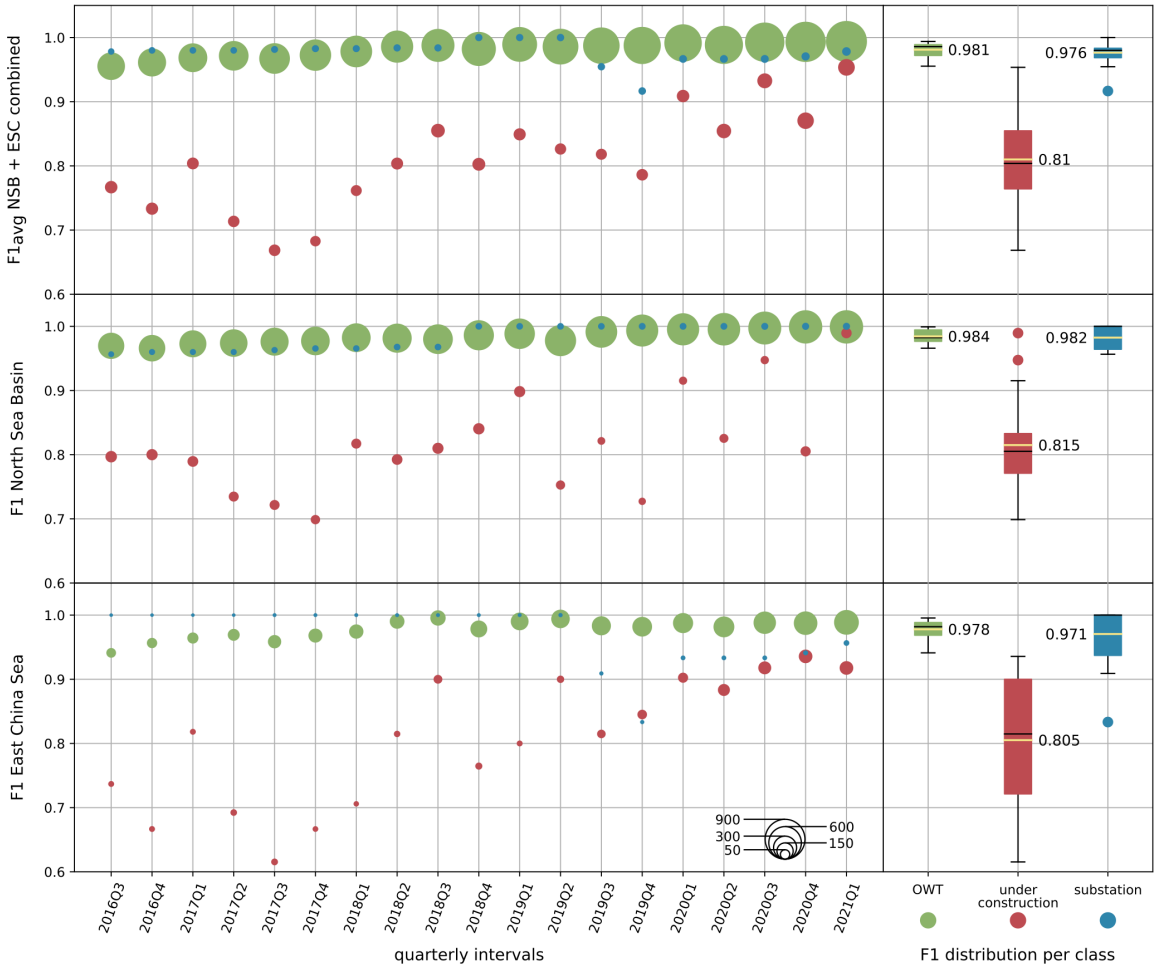



**Figure 5.11:** Evaluation of the classification of the deployment time series. For each period and class, the F1 score has been calculated separately. The point size represents the number of samples for each class and period. In addition, the F1 scores are macro averaged for each class over the entire time series, and their distributions are presented to give a more condensed overview of the classification performance. See the boxplot and average values to the right. Source: Hoeser et al. (2022, p. 4264).

What is highly important is the similar detection performances of the proposed approach on both test sites despite their different complexity. This demonstrates that the CNN object detection pipeline is able to learn a highly generalised version of the synthetic training data, which allows it to perform with top metrics in different environments. This proves the transferability of the detection approach from entirely synthetic environments to real-world settings and the sophisticated variability of the SyntEO training data set.

To evaluate the class predictions of the nineteen quarterly periods, the temporal ground truth data sets of the North Sea Basin and the East China Sea were used as the baseline, see table 5.1. Figure 5.11 provides an overview of single F1 scores for each quarter and class. In order to describe the performance for all nineteen periods, the scores were aggregated for each class by using the earlier mentioned macro averaging approach, see equation (38). OWTs have the highest $F1_{avg}$ value of 0.981, closely followed by the OWF substation class with 0.976. Again, platforms under construction is the most challenging class, which reaches an $F1_{avg}$ score of 0.81. Also, the periodical scores for this class have the largest variance. Since the most important factor for this specific class to be problematic, the quarterly median images, has not changed, the same problems as discussed earlier also apply to this classification approach. However, the general performance of the time series classification reaches top scores and is in line with the CNN classification scores. Also similar to the CNN classification is the separate performance on each test site. The predictions in the North Sea Basin are slightly higher than for the East China Sea. However, the differences are negligible, see figure 5.11.

Besides the DeepOWT data set, two openly accessible sources provide OWT locations on a global scale. They are the Open Street Map (OSM) project and the earlier discussed global offshore wind turbine (GOWT) data set v1.3 by Zhang et al. (2021b), see section 3.1. Since the GOWT v1.3 data set describes OWT locations until 2019Q4, the ground truth data set variant with all offshore wind energy infrastructure in the North Sea Basin and the East China Sea for 2019Q4 is used as baseline in this comparison, see table 5.2.

In the first comparison, all readily deployed OWT locations until 2019Q4 of the derived DeepOWT data set are compared to OWT locations in the OSM database. Figure 5.12 clearly shows the better performance of the proposed DeepOWT data set. The OSM information is based on entries coming from an open-for-all mapping community. For regions with a more active community or accessible data, the OSM coverage is better than for regions that have less active users or public data. Thus the data availability is not homogeneous but changes from region to region. This aspect becomes visible in this comparison. It clearly shows the advantages of using remote sensing data, which in the case of the Sentinel-1 mission, is a globally homogeneous and entirely independent source.

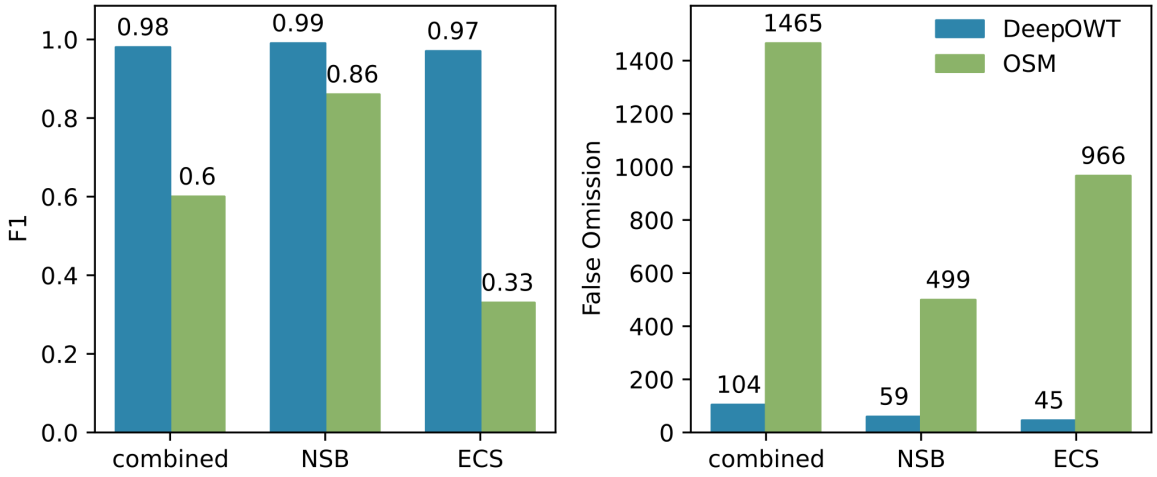



**Figure 5.12:** Comparison of the detection performance for readily deployed offshore wind turbines (OWTs) of the DeepOWT data set and entries in the Open Street Map (OSM) database on the test sites North Sea Basin (NSB) and East China Sea (ECS). Source: Hoeser et al. (2022, p. 4265).

Like the DeepOWT data set, the GOWT v1.3 data set is based on Sentinel-1 data. However, GOWT v1.3 does not differentiate between OWT under construction or OWT, which are readily deployed. It maps a OWT location as soon as it appears in the Sentinel-1 time series as readily deployed (Zhang et al., 2021b). Thus, OWT and OWT under construction from the DeepOWT data set are grouped in one class and compared to the detections of the GOWT v1.3 data set see figure 5.13. The proposed deep learning-based object detection for deriving the DeepOWT data set performs better at both test sites. The GOWT v1.3 data set is based on the investigation of handcrafted morphological features and heavy thresholding. Thus the earlier cited quote of Krizhevsky et al. (2017) also applies in the case of

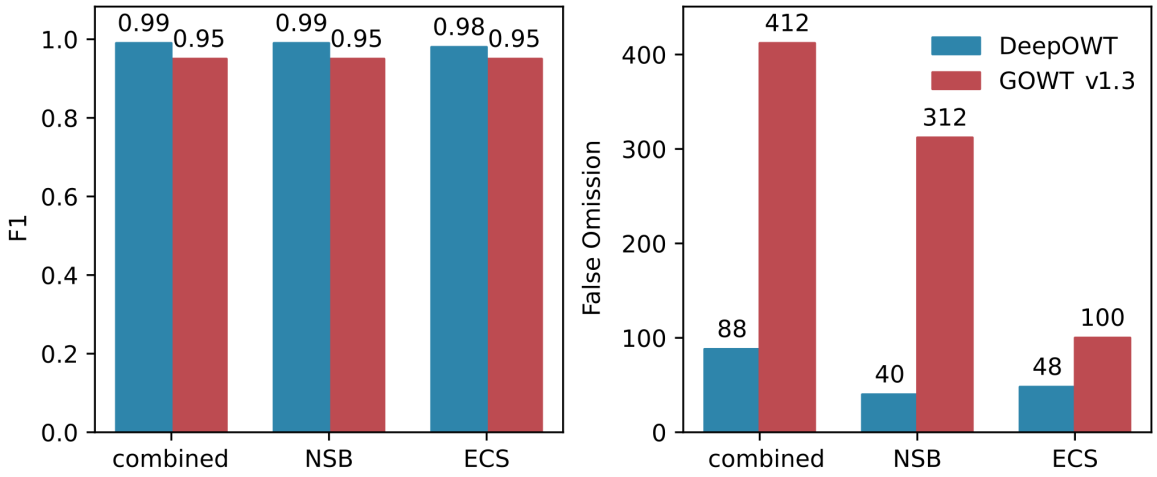



**Figure 5.13:** Comparison of the detection performance for readily deployed offshore wind turbines (OWTs) and OWTs under construction of the DeepOWT data set and the global offshore wind turbine (GOWT) data set by Zhang et al. (2021b) on the test sites North Sea Basin (NSB) and East China Sea (ECS). Source: Hoeser et al. (2022, p. 4265).

offshore wind energy infrastructure detection from Earth observation imagery: “learning beats programming” (p. 84). It is even more true if considered that the DeepOWT data set also differentiates between the different construction phases and additionally includes OWF substations as a third class, which are both not represented in GOWT v1.3.

The data set evaluation and final comparison to other data sets show that the derived information of the DeepOWT data set is of the highest quality. It makes global information on offshore wind energy infrastructure freely available on a global scale, with a unique depth of information with high accuracy and precision of the information provided.

## 5.6 Discussion

In this chapter, the previously developed deep learning-based object detection optimised with synthetic training data was used to develop an automated processing chain. This process detects all offshore wind energy infrastructures on a global scale and derives a five-year time series on their deployment status. this process can be broken down into several narrow artificial intelligences: The synthetic data generation process based on knowledge engineering, the two stage, deep learning-based object detection, and the time series analysis for the detected objects. Each component has a rather narrow task. However, as soon as these individual intelligent agents enter into an exchange with each other, the solution to a complex task, such as the global spatiotemporal mapping of the offshore wind energy infrastructure, becomes possible. Thereby a closer look at each individual component offers to disentangle the otherwise complex process and discuss individual characteristics.

The results of the CNN cascade for global offshore wind energy infrastructure detection confirm the hypothesis that CNNs are potentially able to detect the desired targets in spatially complex environments if sufficient training data is available. The SyntEO framework made it possible to generate such large, highly variable, and task-specific training data sets. It is crucial to emphasise that the CNNs were trained solely with this synthetic data. No further auxiliary data were used during the model optimisation or detection process. Nevertheless, the CNNs are able to extract the desired objects from real data. What is particularly important here is that they do this almost equally well in two environments with different complexity. This proves that the SyntEO approach in combination with an appropriate deep learning model can be used to develop a precise yet abstract model that outperforms detection methods designed by humans.

The breakdown of the entire processing chain into individual modules allows the entire workflow to be designed in a comprehensible but still understandable way. This is particularly helpful in problem analysis and allows the processing chain to be optimised for

future developments. Besides the CNN cascade for object detection, the other component of the entire detection and classification pipeline is the time series classification to derive the deployment dynamics for each quarter. By describing the object location precisely with the derived bounding box of the stage-2 CNN, the complexity of the data and, with it, the problem to solve was highly reduced. Due to this substantial reduction of complexity of the input data, a deep learning model is no longer necessary to perform the classification. However, without the preceding classification and precise location regression of the CNN the time series classification would not receive the suitable input data. This demonstrates the narrowness of the single modules, especially of the time series classification. However, when all narrow modules are combined, from synthetic data set generation and model optimisation, over CNN based global offshore wind energy infrastructure detection, to multi-temporal deployment stage classification, a complex task can be solved efficiently.

As has been shown during the performance evaluation, room for improvement is possible by investigating a higher temporal resolution than the three-month intervals. Especially the class under construction suffers from false negatives due to the quarterly median composites. On the other hand, the quarterly median composites are necessary to create the target objects' specific spatial patterns and filter temporal objects that otherwise would cause a high number of false positives. A potential solution for future applications could be to use rolling quarterly median composites every two weeks and refine the results on a temporal scale to receive even more accurate predictions when a class change happens.

The last point to discuss is the potential of the DeepOWT data set. The proposed DeepOWT data set builds a starting point for many applications in research and practice. In practice, such a global data set is a vital information source for decision-makers to flexible aggregate spatiotemporal developments in order to communicate potentials, results and challenges related to the upcoming decades of offshore wind energy expansions. Open access to the data set helps to foster the exchange of knowledge and interests between stakeholders, which was found to be beneficial for integrated spatial planning (Wever et al., 2015; Fox et al., 2006; Henderson et al., 2003). Moreover, the open accessibility of the data offers the chance to raise public awareness of the urgent need to expand renewable energy production. Early participation during the planning process of all stakeholders is closely related to integrated spatial planning of marine space, which is in the interface of practical application and research (Gușatu et al., 2020, 2021). In the DeepOWT data set, instead of complex remote sensing imagery, the information is condensed to a handy data set that can be used in GIS environments to inform stakeholders via web applications or to set up sophisticated geo processes in order to analyse competing claims for the use of marine space (Gușatu et al., 2020; Virtanen et al., 2022).

The data set is an essential layer for research to combine environmental data like sea surface conditions (Christiansen et al., 2022), underwater noise (Tougaard et al., 2008; Madsen et al., 2006; Wahlberg and Westerberg, 2005), wind resources (Frandsen et al., 2006; Cavazzi and Dutton, 2016; Badger et al., 2015), migration routes and habitats of birds, fish and mammals (Lloret et al., 2022; Drewitt and Langston, 2006; Wilson and Elliott, 2009; Bailey et al., 2014; Bergström et al., 2014; Slavik et al., 2019), shipping routes, and marine traffic (Yu et al., 2020; Ladan and Hänninen, 2012). Especially the focus on the differentiation of the deployment stages under construction and readily deployed adds important information since different impacts of noise and vessel traffic are expected in the two states. Expanding renewable energy is urgent to reduce greenhouse gas emissions and support independent energy production. However, to make this expansion also sustainable, small and large scale effects have to be investigated as early as possible. The DeepOWT data set supports these studies on a global scale.

# *Chapter 6*

## *Global Dynamics of the Offshore Wind Energy Sector 2016-2021**

The proposed DeepOWT data set, as introduced in the last chapter, provides object locations and temporal deployment dynamics. However, when speaking about the expansion of the offshore wind energy sector, one central metric is of major interest, the installed capacity. The two offshore wind energy strategies of the UK and EU use this number to set their goals. Hence, only the number of single OWTs does not sufficiently communicate the dynamics of the offshore wind energy sector. In this chapter, the DeepOWT data set is used as a comprehensive starting point to further derive commonly reported attributes of OWT and to analyse their spatiotemporal dynamics with particular focus on the installed capacity.

## 6.1 Estimation of Installed Capacity Based on Sentinel-1 Data

The attribute installed capacity can not be derived directly from Earth observation data. However, a proxy for the capacity of an OWT is the rotor diameter which again is related to the turbine height. Thus for estimating the installed capacity, first, the height of an OWT is calculated and used as an input variable in a statistical model to predict the installed capacity.

### 6.1.1 Data

The data used for height calculation comes from the Sentinel-1 mission. This is a major advantage since OWT detection, deployment stage classification, height calculation, and with it, the capacity estimation can all be derived from a single data source. Similar to the

*Parts of this chapter have been published in Hoeser and Kuenzer (2022a).

object detection input data, a median composite for the second quarter in 2021 of all available IW GRD VH polarised images is generated. As specified later in more detail, this time, only a single orbit direction is used to generate this median composite. For median composite data generation and download, the data acquisition pipeline introduced in section 4.2.1 was used. To validate the height calculation, turbine hub heights of 50 OWT clusters were collected from publicly accessible sources like regulatory planning documents or OWF operator descriptions. For the same OWT clusters, the information about installed capacity was gathered, which will later be split into a train and test set in order to model and validate the relation between hub height and installed capacity.

### 6.1.2 Radargrammetric OWT Hub Height Calculation

As already discussed in section 3.2, due to the sensor properties of the Sentinel-1 mission, the spatial signature of the radar signal produced by an OWT shows specific characteristics. These characteristics provide the necessary information to calculate the hub height of an OWT. Figure 6.1 provides a conceptual overview of the height calculation approach. Figure 6.1a) shows the ray path and resulting median composite of a single OWT. The most significant part of the radar signal is related to the centre location of the turbine. However, a smaller cluster to the left of the centre is also clearly visible and caused by the layover effect as depicted in figure 6.1b). Due to the right looking sensor geometry, this layover cluster is always to the left of the centre for images taken in ascending orbit direction and to the right for images in descending orbit direction. To simplify the following explanation, only the ascending orbit, as shown in figure 6.1, is considered. However, for height calculation, both orbits have been used, primary the ascending orbit, and in case no acquisitions were available, images from a descending orbit were processed.

The layover effect is caused by the nacelle of the turbine. The radar signal hits the elevated nacelle in the open sea in an otherwise nearly perfectly smooth and non elevated environment and is reflected directly to the radar antenna. This specific part of the signal would normally hit a point behind the turbine in range direction. Due to the now shortened ray path, the signal, when projected onto the ground range, appears closer to the sensor as the foundation at sea surface level, which is exactly at the same location only with a vertical offset , see figure 6.1b). This vertical offset is the hub height, see figure 6.1a). The centre coordinate of the foundation is already given by the DeepOWT data set, where it was precisely refined to describe the exact centre of the backscatter cluster which the turbine foundation and platform produce, see figure 5.6. The coordinate of the layover cluster is unknown. To find this coordinate, first, a 400 m $\times$ 400 m box is generated for each OWT location in the DeepOWT data set. Following that, the peak detection approach of the horizontal maximum

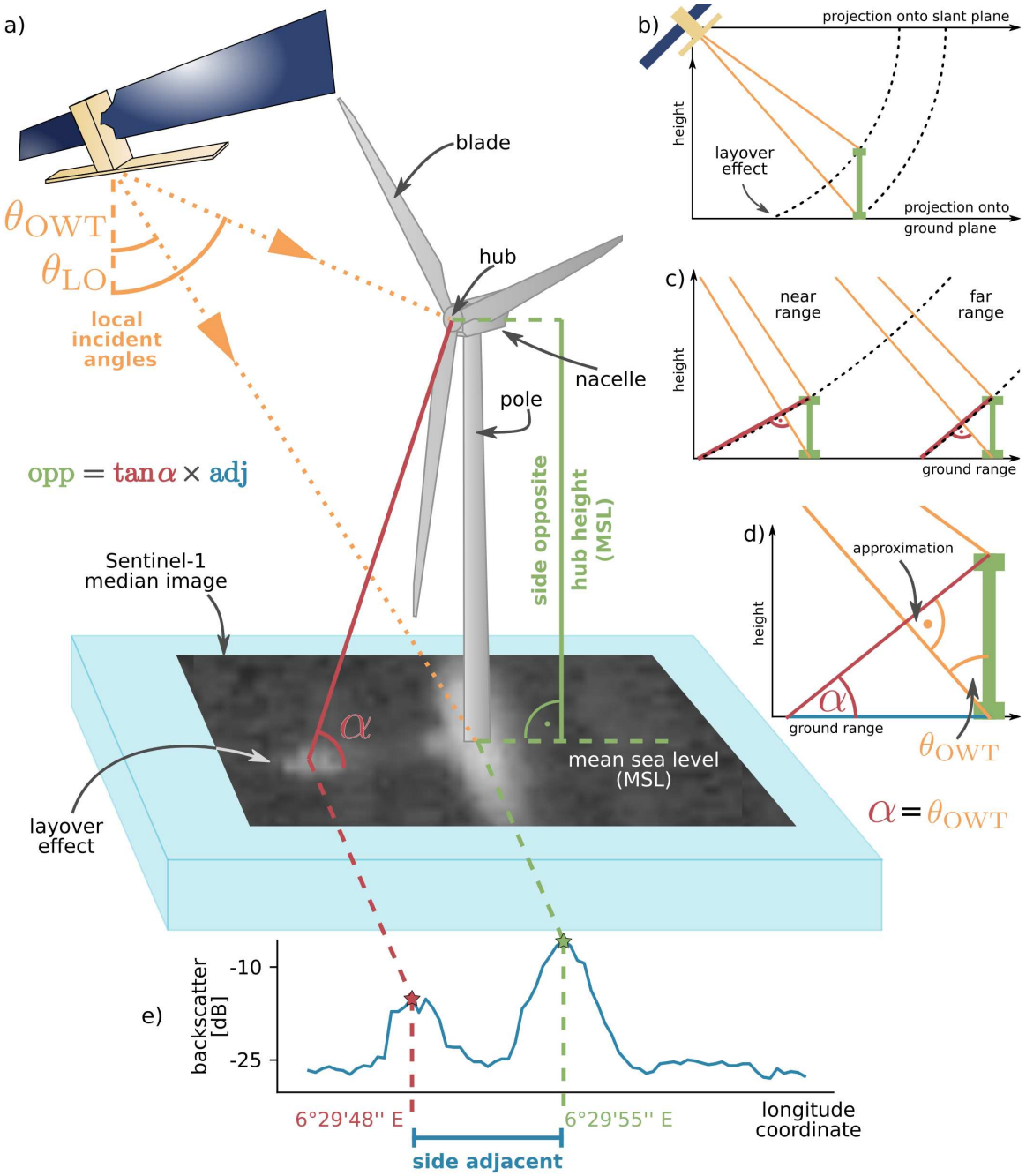



**Figure 6.1:** Conceptual overview of the radargrammetric hub height calculation from Sentinel-1 images. a) shows the investigated radar image and its typical backscatter signature of an offshore wind turbine (OWT). b-d) describe how the layover (LO) effect and imaging geometry can be used for hub height above mean sea level (MSL) calculation in both near and far range conditions. e) shows how the 2D radar image is reduced to a maximum swath profile to derive the exact location of the layover cluster to use its distance to the turbine centre for height calculation. Source: Hoeser and Kuenzer (2022a, p. 7).

swath profile of the local radar signature introduced for location refinement in section 5.2.3 is reused. This time, instead of detecting the centre peak of the turbine foundation, the smaller peak of the layover effect, left to the centre peak, was detected, see figure 6.1e). Thus, the exact coordinate of the cluster of the layover effect is available. By transforming both

coordinates to the corresponding Universal Transverse Mercator (UTM) zone, their absolute horizontal difference can be calculated in metres. With this distance and the hub height, a right-angled triangle becomes obvious, where the calculated distance is the side adjacent and the hub height the side opposite, see figure 6.1a). To derive $\alpha$ the local incident angle of the centre location is used, see figure 6.1c+d). The approximate hypotenuse between turbine house and layover coordinate is orthogonal to the incident ray path of the radar signal hitting the turbine platform and fundament at the sea surface. The resulting triangles describe the angle $\alpha$ as being equal to the local incident angle $\theta_{\mathrm{OWT}}$ at the coordinate of the turbine centre, see figure 6.1d). This property is maintained by increasing incident angle from far to near range since a turbine with the same height has a larger angle $\alpha$ in a far range situation as in a near range situation, see figure 6.1c). Since the local incident angle is provided along with the backscatter information for a Sentinel-1 acquisition, $\alpha$ can be queried for each OWT coordinate. Due to the possibility that images from neighbouring orbits can contribute to the median composite, the local incident angle at the OWT location is not derived once from a random image of the median composite stack but for each contributing image in the median composite stack. The most frequent value within this list of local incident angles is defined as $\alpha$ for the height calculation of a turbine. With $\alpha$ and the side adjacent obtained, the hub height or side opposite can be calculated, see figure 6.1a).

To reduce outliers of OWT heights within an OWT cluster, the height calculation is performed cluster wise. Therefore, in a preceding step, clusters of OWTs are defined due to the spatial proximity and deployment date of single OWTs, provided by the DeepOWT data set. Then, for each OWT, the height is calculated separately as proposed above. The resulting list of height values in one OWT cluster is subsequently reduced by selecting the median. The final hub height for an OWT is the median value of its corresponding OWT cluster. This procedure was applied to the entire DeepOWT data set. Figure 6.2 provides the results for the 50 ground truth OWT clusters along with an in-depth error discussion. A Levene test (Brown and Forsythe, 1974) whose result confirms equal variance across the ground truth height and the calculated height $\alpha_{\mathrm{crit}} = 0.05 < p = 0.272$, a Pearson's correlation coefficient of $\rho = 0.877$ ($p \ll 0.001$), and the corresponding $R^2$ of 0.77 were calculated. Thus, 77% of the variance of the calculated height can explain 77% of the ground truth height. Together with a slope of a linear regression of 1.06, the calculated results show sufficient statistical robustness to be accepted as a good approximation of the ground truth data. Furthermore, the error distributions in figure 6.2b+c) indicate that there is no systematic error, leading to errors only occurring in one direction. Rather, the error distribution follows a normal distribution with a mean of 1.03 m and a standard deviation of 8.25 m, tested after D'Agostino and Pearson (1973) $\alpha_{\mathrm{crit}} = 0.05 < p = 0.274$. The absolute error distribution in figure 6.2d) shows a right-skewed distribution with a mean absolute error of 6.45 m.

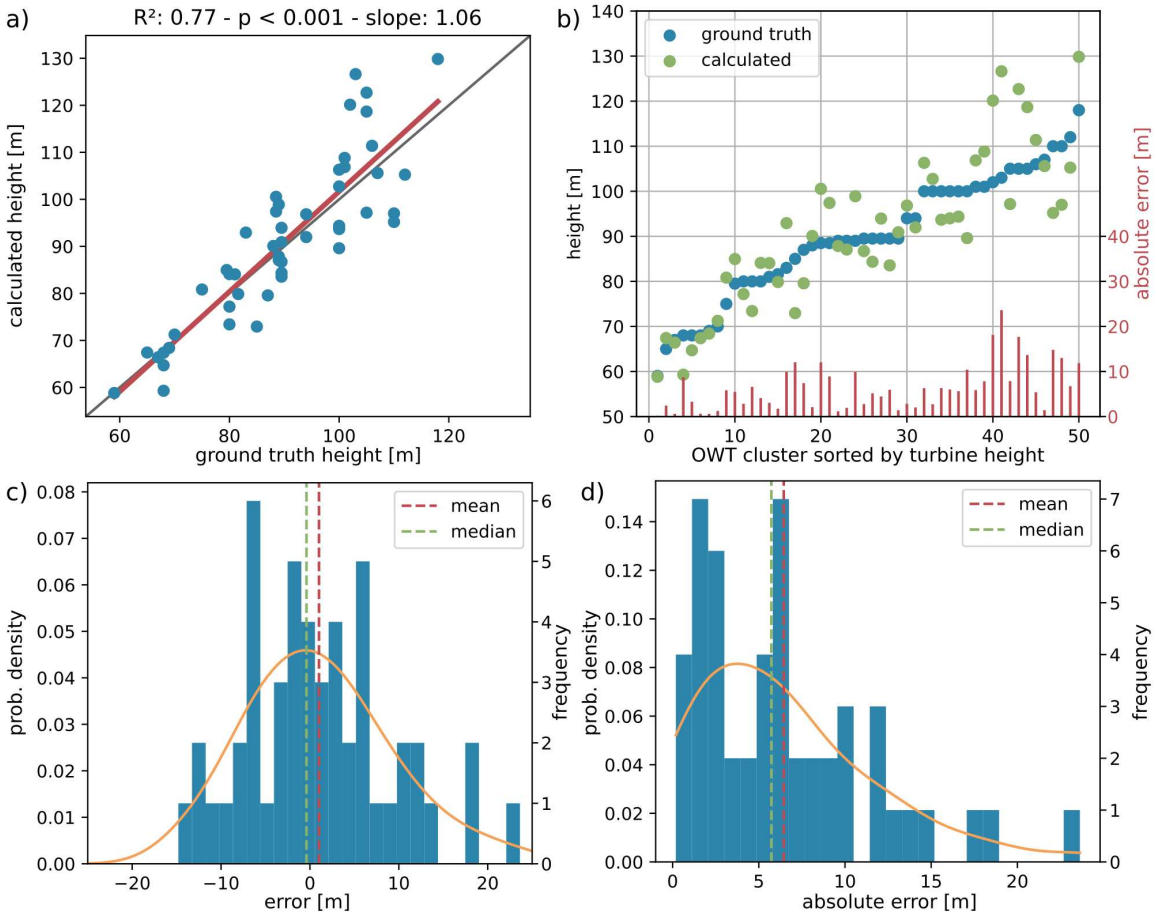



**Figure 6.2:** Graphical error discussion of the hub height calculation. a) Correlation between calculated and ground truth hub height. b) Residues of the calculated hub heights, which are sorted ascending by the ground truth hub height. c) Error distribution, which follows a normal distribution with $\mu = 1.03$ m. d) Distribution of the absolute error with a mean of 6.45 m. Source: Hoeser and Kuenzer (2022a, p. 6).

By taking potential error sources during the height calculation into account, the mean absolute error is within the technical feasibility of the Sentinel-1 mission. Figure 6.3 shows the error contribution, which is related to the 10 m pixel spacing of the investigated IW GRD product and the associated lack of precision in object localisation within the resulting pixels. One part of the overall height error $\varepsilon_h$ is related to the precision of the two coordinates, which are used to calculate the side adjacent. With a pixel spacing of 10 m, the distance between the two coordinates of OWT cluster and layover cluster is imprecise within the bounds of $\pm 10$ m. This is the theoretical error of the side adjacent $\varepsilon_{\text{adj}}$ given by the Sentinel-1 product specification. Depending on a near range (31°) or far range (46°) incident angle, the maximum of the related height error contribution $\varepsilon_{h^{\varepsilon_{\text{adj}}}}$ ranges between $\pm 6$ m and $\pm 10.36$ m respectively, see figure 6.3. Compared to the mean absolute error of 6.45 m, the general applicability of the proposed hub height calculation is pointed out. The calculated hub height values are sufficiently accurate to be used as an input in the upcoming installed capacity estimation without risking systematic or excessive error propagation.

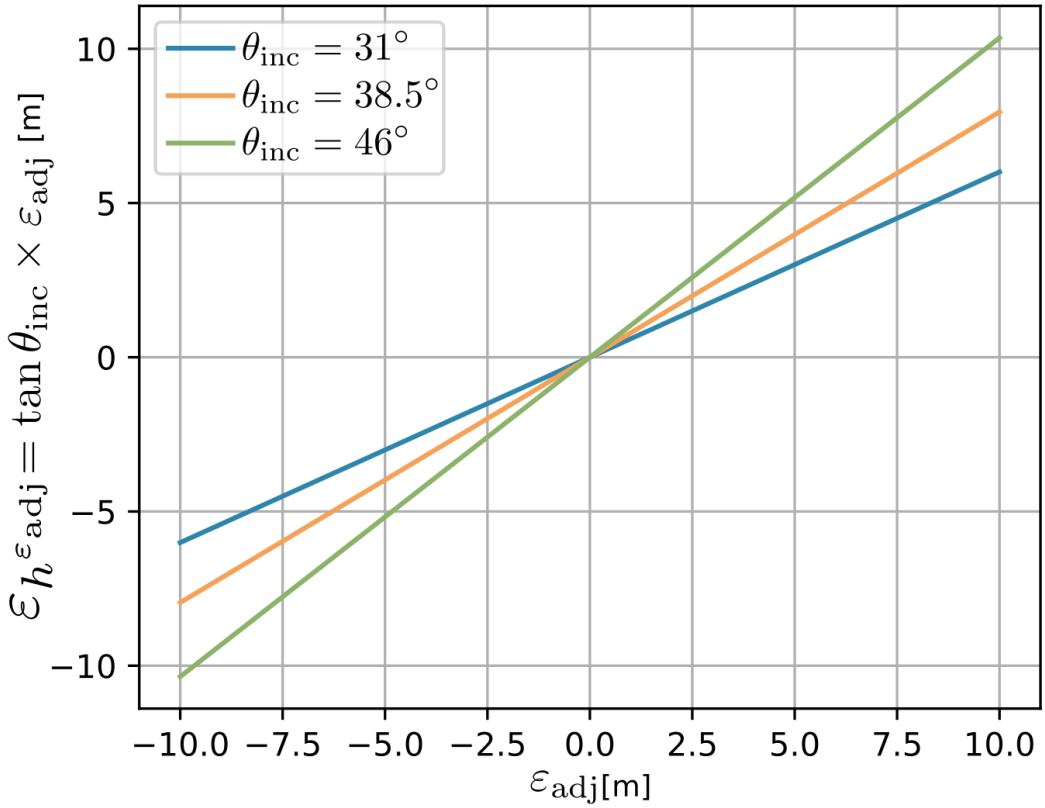



**Figure 6.3:** Theoretical bounds of the height error contribution $\varepsilon_{h}{}^{\varepsilon_{\text{adj}}}$ coming from the 10 m pixel spacing of the Sentinel-1 Interferometric Wide swath (IW) Ground Range Detected (GRD) products which contributes to the adjacent side error $\varepsilon_{\text{adj}}$. The three incident angles $\theta_{\text{inc}}$ describe a near (31°) to far (46°) range situation, thus the corresponding maximum error is ±6 m and ±10.36 m. Source: Hoeser and Kuenzer (2022a, p. 7).

### 6.1.3 Regression Model for Capacity Estimation

With the calculated hub height for each OWT available, the last step is to describe the relationship between hub height and installed capacity to finally estimate the installed capacity. The empirical correlation is obtained by observing the manually collected hub height-installed capacity pairs from publicly available sources. The 50 data samples are split randomly into a train set with 36 samples and a test set with 14 samples which will not be considered further until the evaluation. Figure 6.4 shows the train data as scatter plot. The point cloud suggests an exponential correlation for lower parts of the hub height between 60 m and around 95 m. However, the exponential trend for hub heights above 95 m is less distinct, with increasing variance and a tendency to level off. The uncertainty about a proper model fit caused by a lower point density and higher variance for higher hub heights means that there is a risk of reacting more sensitively to small changes in hub height if a purely exponential curve is assumed, which leads to greater fluctuations in the installed capacity. Instead, a suitable model hypothesis should be able to represent the clearly exponential part at the beginning without overestimating the second part of the point cloud in a

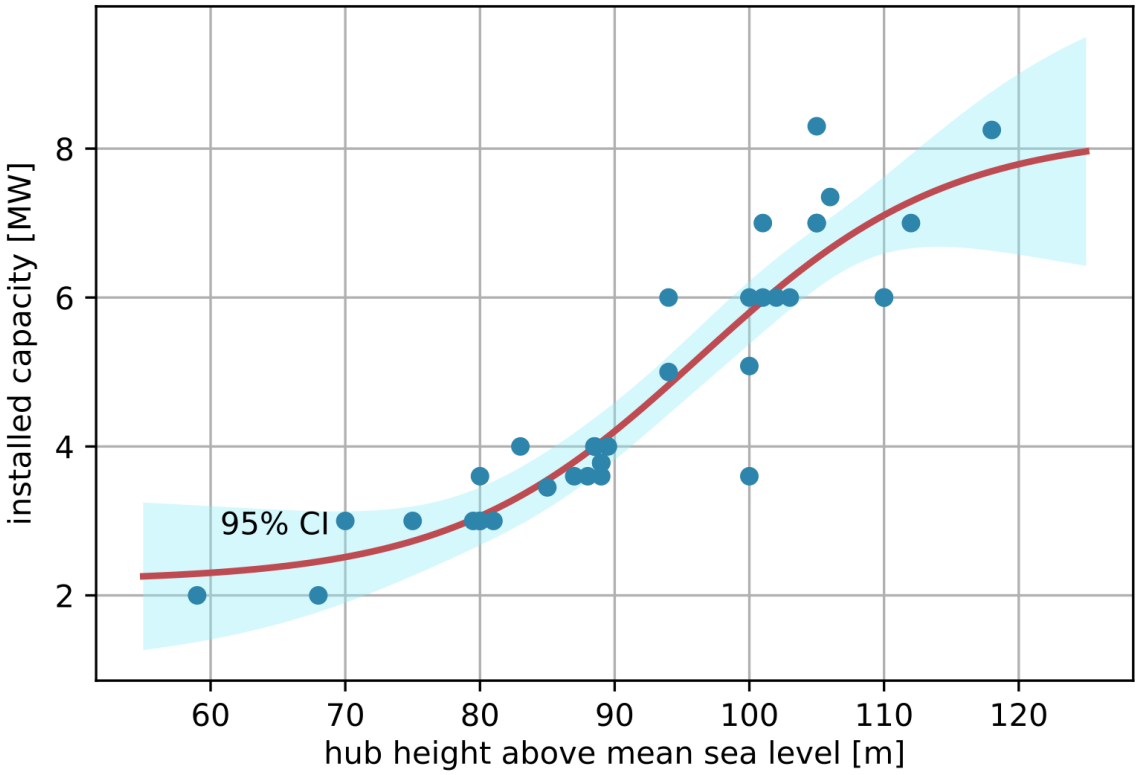



**Figure 6.4:** Relation between the turbine's hub height and installed capacity of 36 training samples and the corresponding fitted sigmoid model with its 95% confidence interval (CI). Source: Hoeser and Kuenzer (2022a, p. 8).

further exponential progression. Therefore, a model hypothesis was chosen that conservatively describes the relationship in the second part of the point cloud but can communicate the recognisable exponential progression for the first part. This decision also contributes to a smaller impact of the earlier discussed height error contribution caused by the 10 m pixel spacing of the underlying data for height calculation. In a purely exponential model, the height error contributions would also impact the estimated capacity exponentially.

$$C(h_{\mathrm{OWT}}) = \frac{\Lambda}{1 + e^{-\omega(h_{\mathrm{OWT}} - h_0)}} + \xi \tag{39}$$

By taking all these considerations into account, the hypothetical model was chosen to be a sigmoid regression as shown in figure 6.4 and equation (39). The installed capacity $C$ is estimated by the OWT height $h_{\mathrm{OWT}}$. $\Lambda$ is the sigmoid function's maximum parameter, and $\omega$ denotes the logistic growth rate. $h_0$ and $\xi$ are offset parameters of the infliction point of the function. $h_0$ controls the position along the axis of the OWT hub height, and $\xi$ controls the position along the axis of the installed capacity. These adjustable parameters of the sigmoid function were optimised according to the machine learning cycle introduced in section 2.2. The least-squares cost function was minimised to fit the sigmoid function's parameters. Figure 6.4 shows the fitted model along with its 95% confidence interval (CI).

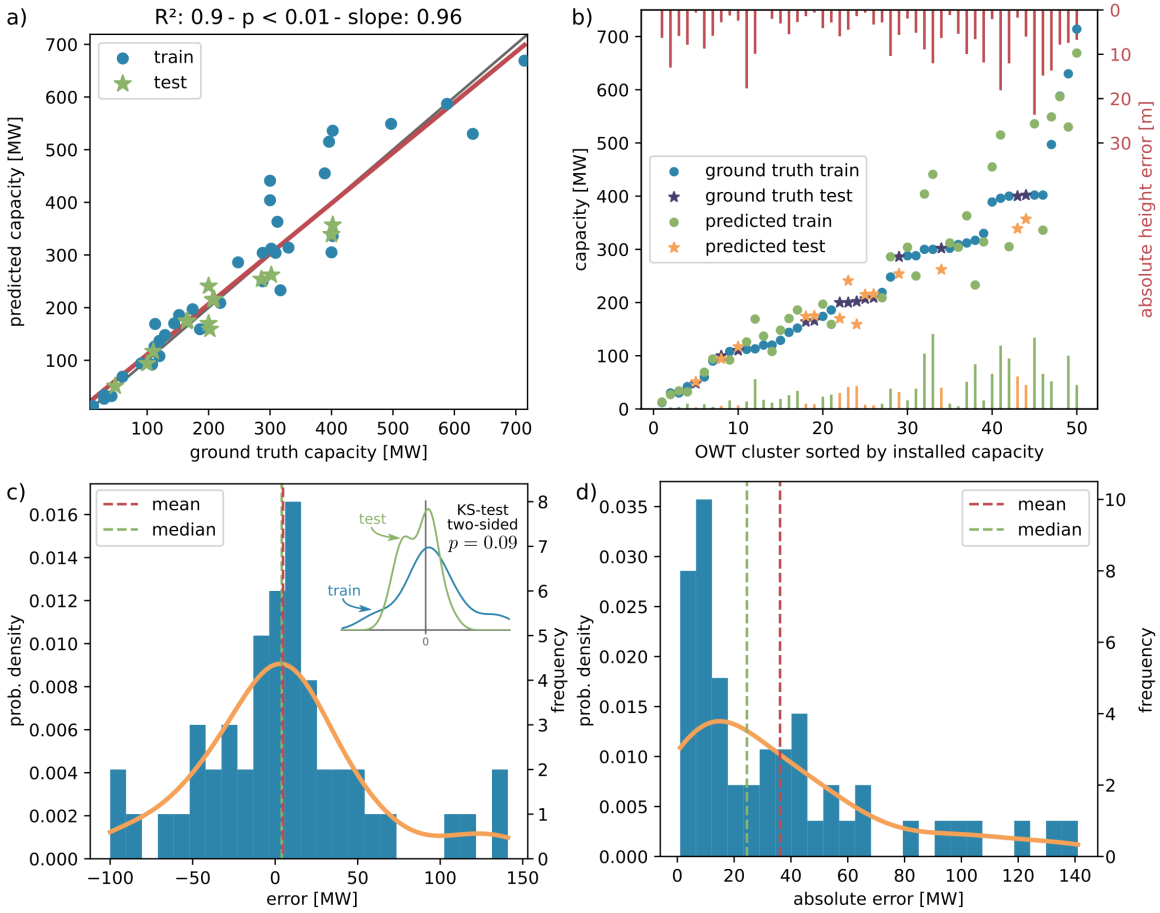



**Figure 6.5:** Graphical error discussion of the estimated installed capacity. a) shows the correlation between estimated and ground truth installed capacity. b) shows the residues of the installed capacity of the ground truth offshore wind turbine clusters, which are sorted ascending by their installed capacity. Furthermore, the absolute height error is provided to relate both residues. c) shows the error distribution of the estimated installed capacity, which follows a normal distribution with $\mu = 4.76$ MW. Furthermore, the Kolmogorov-Smirnov test in c) shows that the distributions of the errors predicted on the train and test split are indifferent and thus discussed together in c) and d). d) shows the distribution of the absolute error of the installed capacity with a mean of 36.16 MW. Source: Hoeser and Kuenzer (2022a, p. 8).

For an error discussion, the installed capacity was predicted for all 14 OWT clusters of the test but also for the 36 of the train set, separately. Therefore, the model was applied to each calculated hub height of an OWT coming from the DeepOWT data set, which lies within the boundaries of one of the 50 ground truth OWT clusters. The predicted installed capacities for each OWT in a cluster were then summed up and compared to the corresponding installed capacity of the OWT cluster of the ground truth data. That way, it is possible to investigate not only the model performance on a single OWT but also the combination of the fitted model and the underlying DeepOWT data set. Thus the upcoming error discussion is representative of estimating installed capacity with the OWT locations and hub heights coming from the DeepOWT data set.

As stated above, predictions were made for OWTs which are within the clusters described by the test and train sets. The resulting error distributions for both sets were compared by a Kolmogorov-Smirnov test (Massey, 1951). The test result $\alpha_{\text{crit}} = 0.05 < p = 0.09$ confirms the null hypothesis that both distributions are identical, see figure 6.5c). Due to this result, the upcoming error discussion combines the error values of the test and train sets. The error discussion follows the same pattern as for the height calculation. Figures 6.5a+b) provide insights into the distribution of the residues. The Levene test confirms equal variance for predicted capacity and ground truth capacity with $\alpha_{\text{crit}} = 0.05 < p = 0.977$. The correlation between predicted and ground truth installed capacity is high with a slope of 0.96 of a linear regression, a Pearson coefficient of $\rho = 0.949$ ($p < 0.01$), and an $R^2 = 0.9$. Thus, 90% of the variance of the predicted installed capacity can explain 90% of the variance of the ground truth installed capacity, see figure 6.5a). Likewise to the calculated height error values, the estimated installed capacity error values were tested for normality. With the test result $\alpha_{\text{crit}}0.05 < p = 0.067$ the error values follow a normal distribution with a mean of 4.76 MW and a standard deviation of 51.15 MW for entire OWT clusters, indicating that no tendency of the model for systematic over- or underestimation occurs. Furthermore, the mean absolute error is 36.16 MW, which is less than 15% of the mean installed capacity of 244.98 MW of the 50 ground truth OWT clusters.

Overall, the results show a sufficient performance to estimate the installed capacity with the introduced approach of calculating hub height and using the sigmoid regression to link this variable with the installed capacity. Finally, the fitted sigmoid regression model was applied on a global scale to all calculated hub heights to estimate the installed capacity of each offshore wind turbine in the DeepOWT data set. For a further comparison of the estimated installed capacity with regional reports of the installed capacity of the offshore wind energy sector, see figure 6.13 and the associated discussion in section 6.3.

## 6.2 Deriving Offshore Wind Turbine Site Specifications

### 6.2.1 National Affiliation and Water Depth

Having derived the installed capacity as a key characteristic of OWTs, other characteristics describing the location of an OWT will now be compiled. For this purpose, the spatial information of the DeepOWT data set was processed in GIS analysis. To derive the national affiliation of each OWT, the point locations in the DeepOWT data set were spatially queried by polygons of exclusive economic zoness (EEZs) globally. EEZs are areas on the use of marine resources set by the United Nations (UN). The data used is provided by Flanders Marine Institute (2020).

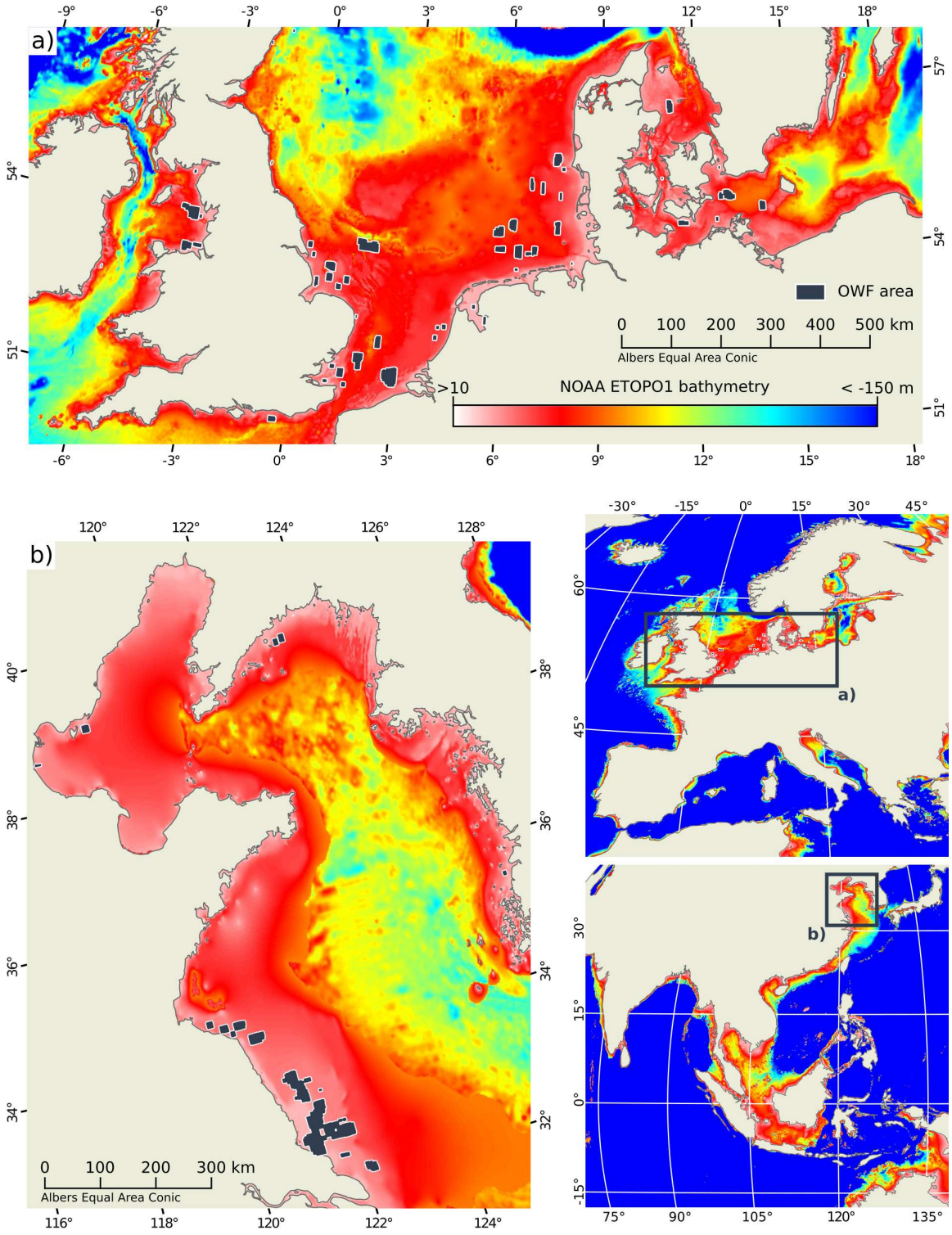



**Figure 6.6:** The derived offshore wind farm (OWF) areas of the DeepOWT data set in context with bathymetry data coming from NOOA's ETOPO1 data set. a) shows the situation of large shallow areas in the North Sea Basin. b) shows shallow water depths for coastal sites in the northern East China Sea and the Yellow Sea. Source: Hoeser and Kuenzer (2022a, p. 10).

In another overlay operation, the water depth for each OWT was queried from the ETOPO1 data set. ETOPO1 is a 1 arc-minute global raster of topography and bathymetry data provided by Amante and Eakins (2009). The ETOPO1 data set is included in the data archive of the GEE. To derive the OWT water depth, the bathymetry data was queried for the locations provided in the DeepOWT data set. Figure 6.6 provides an impression of the water depth and the derived locations of OWFs in North Europe, North East China and the West Coast of North and South Korea. The national affiliation and water depth information was appended to the DeepOWT data set attribute table for later analysis.

### 6.2.2 Distance to Coast and Neighbouring Platforms

Figure 6.7 shows OWFs in the North Sea Basin at the eastern coast of England with isopleths showing the distance to the shoreline. To derive the exact minimum distance between each OWT and the shoreline, a search radius of 200 km was defined for each OWT in the DeepOWT data set. These search areas were used to clip a global mainland polygon

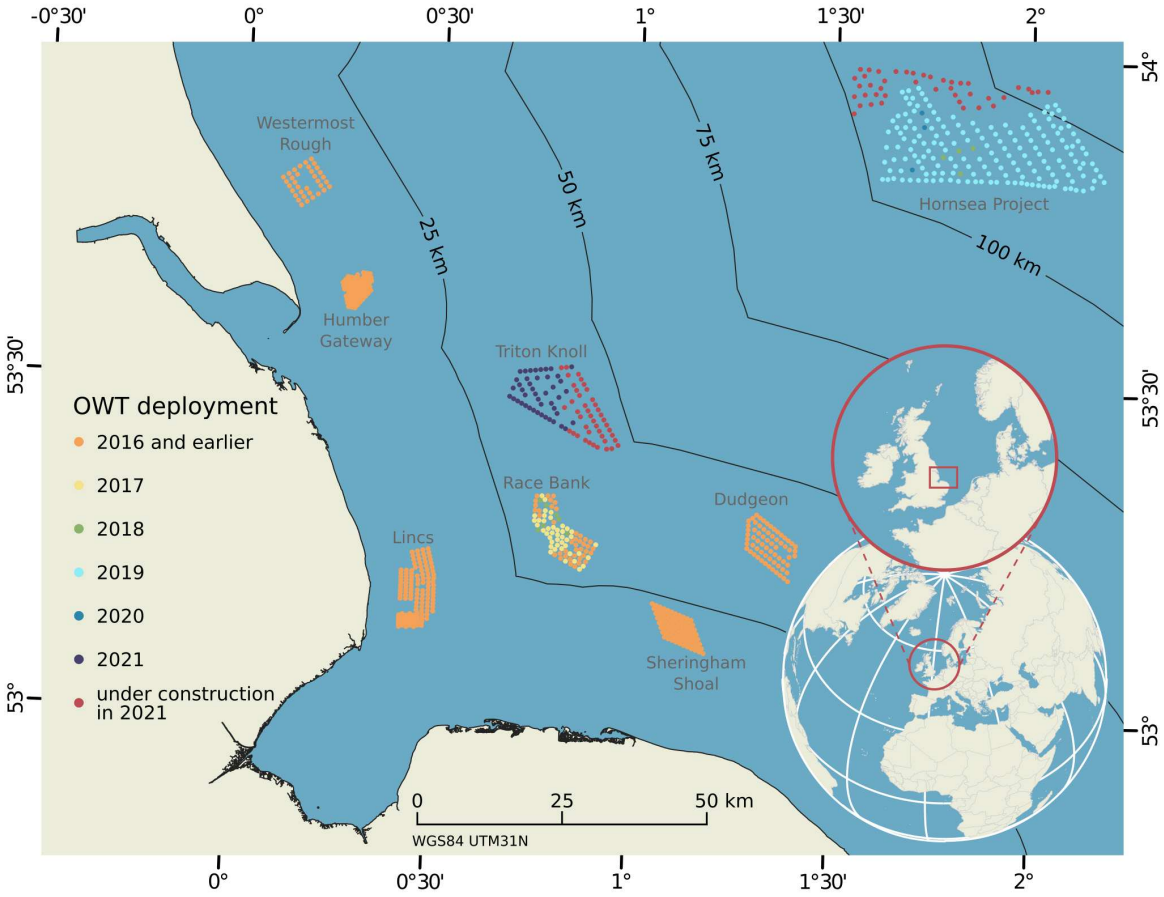



**Figure 6.7:** offshore wind turbine (OWT) locations from the DeepOWT data set grouped by their deployment date in context with their distance to the shoreline indicated by 25 km distance isopleths for the eastern coast of England. It becomes obvious that over time, the distance to the coastline increases. Source: Hoeser and Kuenzer (2022a, p. 9).

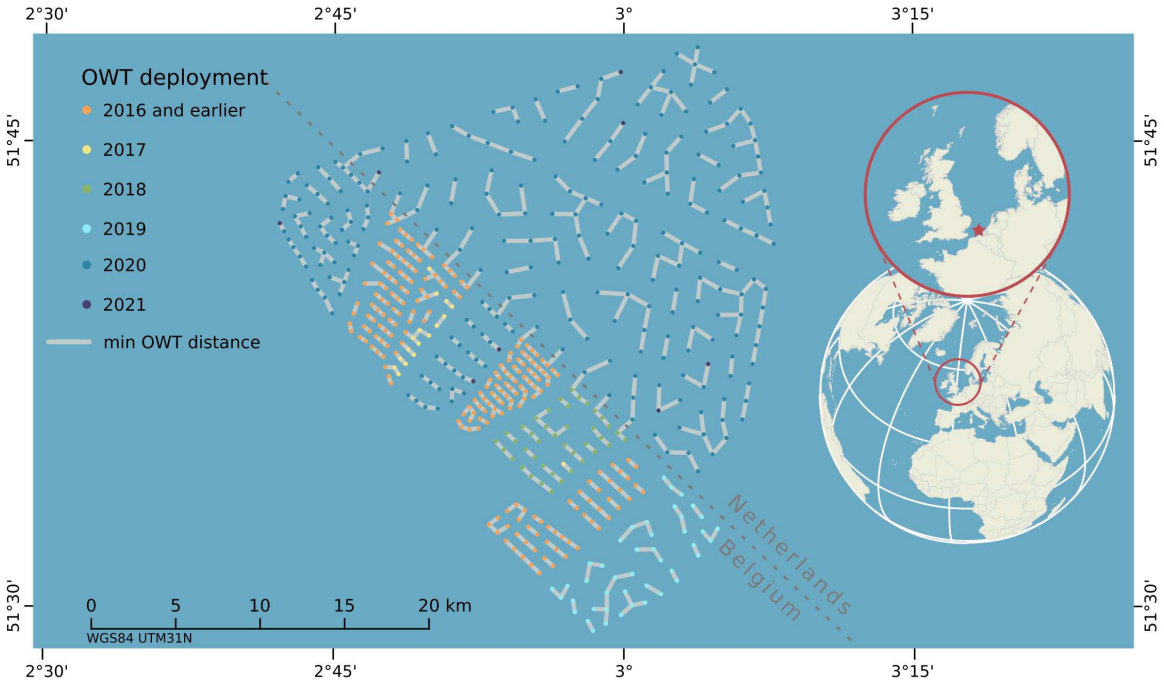



**Figure 6.8:** offshore wind turbine (OWT) locations from the DeepOWT data set grouped by their deployment date in context with their distance to the closest neighbour in an offshore wind farm cluster of the Netherlands and Belgium. It becomes obvious that over time, the distance to the closest OWT increases, which is related to larger rotor diameters of modern OWTs. Changed after Hoeser and Kuenzer (2022a, p. 9).

provided by Natural Earth (2022). The clipped areas are reprojected to the corresponding UTM zone of the OWT under investigation. With that done, the minimum distance to all polygons in the 200 km large search radius can be derived for all OWT locations.

With the same search radius of 200 km for each OWT location, all other OWT locations of the DeepOWT data set were selected. All distances between these selected OWTs and the OWT under investigation were calculated. The minimum of all of these distances was selected along with the corresponding location ID of the OWT. Figure 6.8 gives an impression of the resulting information for an OWF in the North Sea Basin in the EEZs of the Netherlands and Belgium.

## 6.3 Global Analysis of Spatiotemporal Offshore Wind Turbine Dynamics

This section brings together the methods developed and results obtained in this work. It provides a comprehensive presentation of the derived spatiotemporal dynamics of the offshore wind energy sector from Earth observation data. The most significant features were extracted from a single Earth observation archive containing data of the Sentinel-1

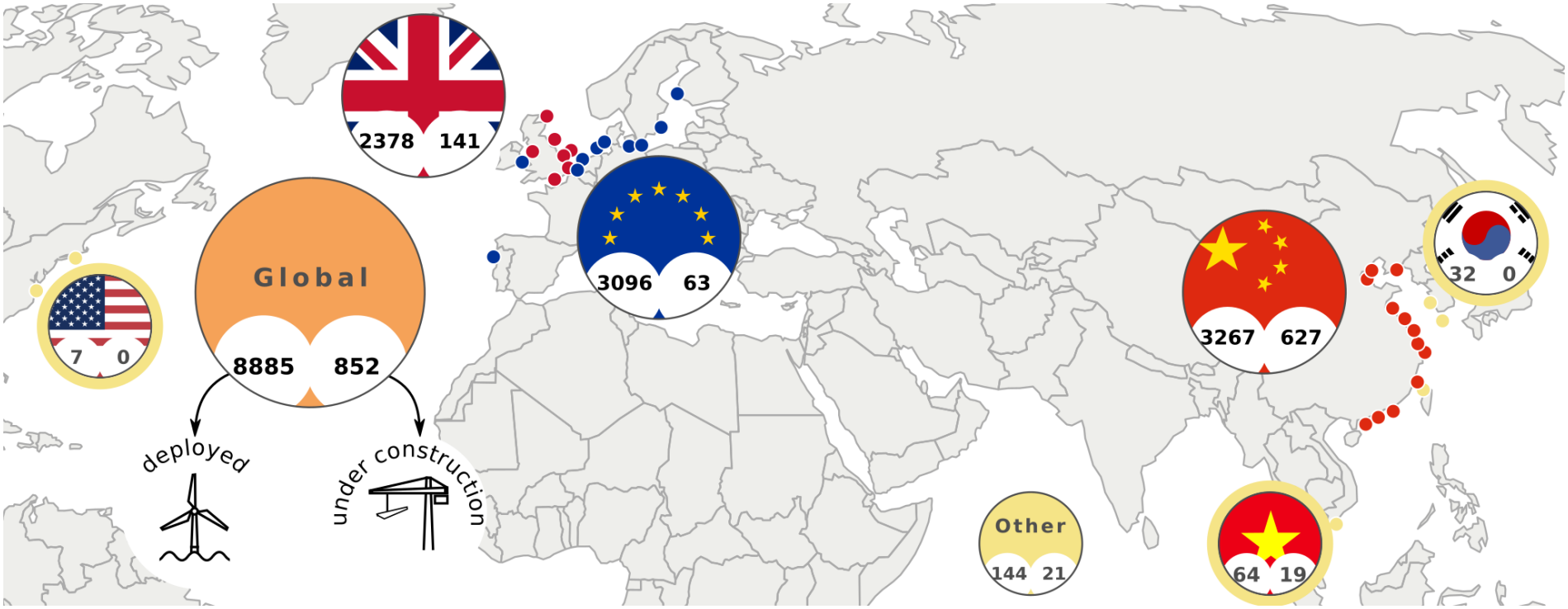



**Figure 6.9:** Global overview of the derived number of offshore wind turbines readily deployed and under construction as of June 2021. The information is solely based on Earth observation data coming from the Sentinel-1 mission. The three major offshore wind energy producers are The EU, China and the UK, with China having the most projects under construction. The points along the coastline indicate major offshore wind farm clusters. Source: Hoeser and Kuenzer (2022a, p. 11).

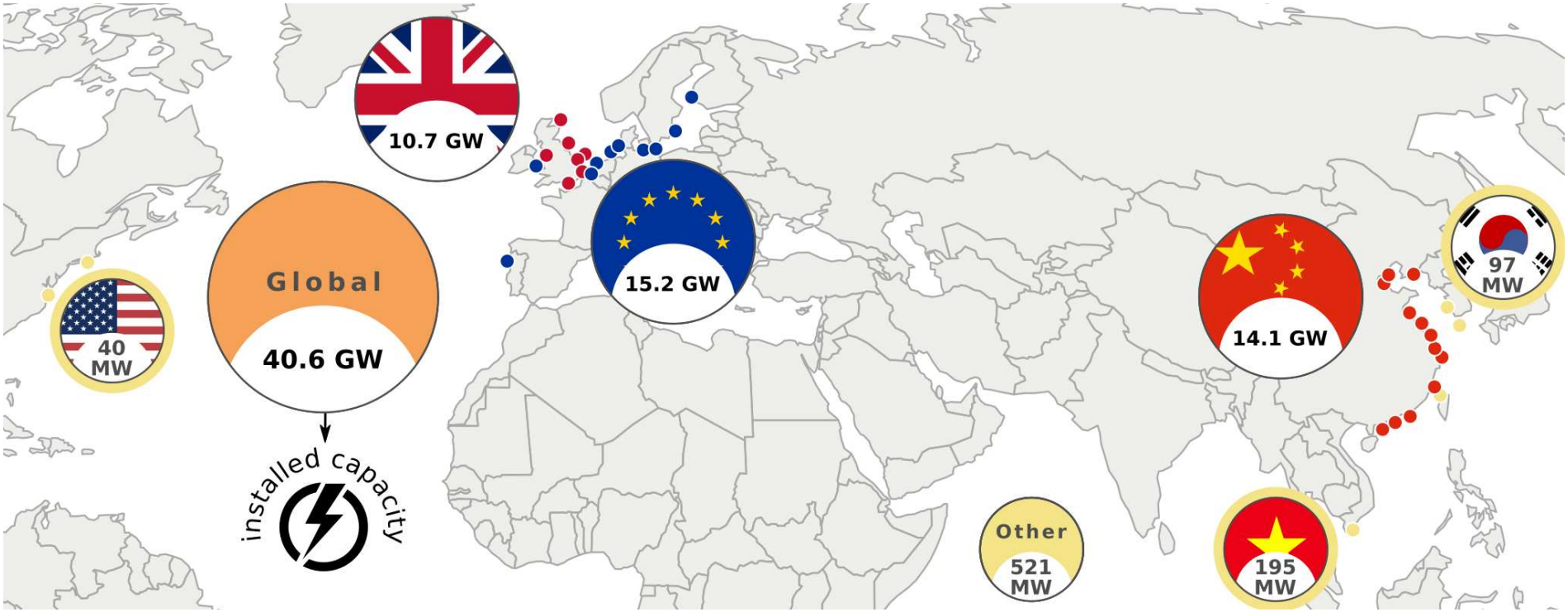



**Figure 6.10:** Global overview of the derived installed capacity for offshore wind energy production as of June 2021. The information is solely based on Earth observation data coming from the Sentinel-1 mission. The three major offshore wind energy producers are The EU, China and the UK. The points along the coastline indicate major offshore wind farm clusters. Source: Hoeser and Kuenzer (2022a, p. 11).

mission, which demonstrates the versatility of the information captured by the Sentinel-1 satellites. At the same time, it also demonstrates the versatility of the proposed methods and implemented procedures used to extract the spatiotemporal information of the offshore wind energy sector from this data source. The extracted and analysed information about the global dynamics of the offshore wind energy sector, hidden in spaceborne radar images, is hereby made accessible to a large audience.

Figure 6.9 shows the deployment situation of OWT, and figure 6.10 the installed capacity in June 2021 on a global scale as derived by applying the methodology introduced in this dissertation. Globally, 40.6 GW of installed offshore wind energy capacity was realised in June 2021 by 8,885 readily deployed OWTs. The current massive expansion of the offshore wind energy sector can be seen in the fact that 852 more OWTs are under construction in June 2021, which corresponds to about 10% of all turbines installed by then. The three major wind energy production contributors are clearly discernable in figure 6.10. With 15.2 GW the EU has the most installed offshore wind energy capacity worldwide, closely followed by China with 14.1 GW, and the UK with 10.7 GW. Despite being among the world's largest onshore wind energy producers, the United States (US) has only 40 MW of installed capacity, which relates to 7 OWT in 2 pilot projects on the Atlantic coast. However, the currently developed Vineyard Wind project will be the first commercial-scale offshore wind farm with 62 OWTs and an overall installed capacity of about 800 MW (Vineyard Wind, 2022). It is planned to be readily deployed in 2023 and hereby the starting point for the US to enter the offshore wind energy sector besides other regions which have already deployed first OWFs.

More insights into the temporal developments of the offshore wind energy sector are provided by figure 6.11. This overview broadens the perspective on the state of the global offshore wind energy sector, especially with regard to the three largest participants. It clearly shows that the most significant expansion of offshore wind energy infrastructure within the last five years took place in China. From July 2016 to June 2021, 5,268 OWTs with a combined installed capacity of 27.2 GW were deployed worldwide. Of these, 2,960 OWTs with a combined capacity of about 13 GW were deployed in China. Considering that China in July 2016 only had about 1.1 GW installed capacity, this is an increase of 1,182% of China's installed capacity, and a contribution of 48% to the global increase in installed capacity. China is thus the most important driver behind the expanding trend in offshore wind energy over the last five years. Figure 6.12 summarises the quarterly changes in a comparison of overall growth rates for a better overview. In contrast to the Chinese growth rates, the EU deployed 1,313 OWTs with 8 GW, which is a contribution of 29% to the global expansion. The UK deployed 916 OWTs with 5.8 GW, which is a contribution of

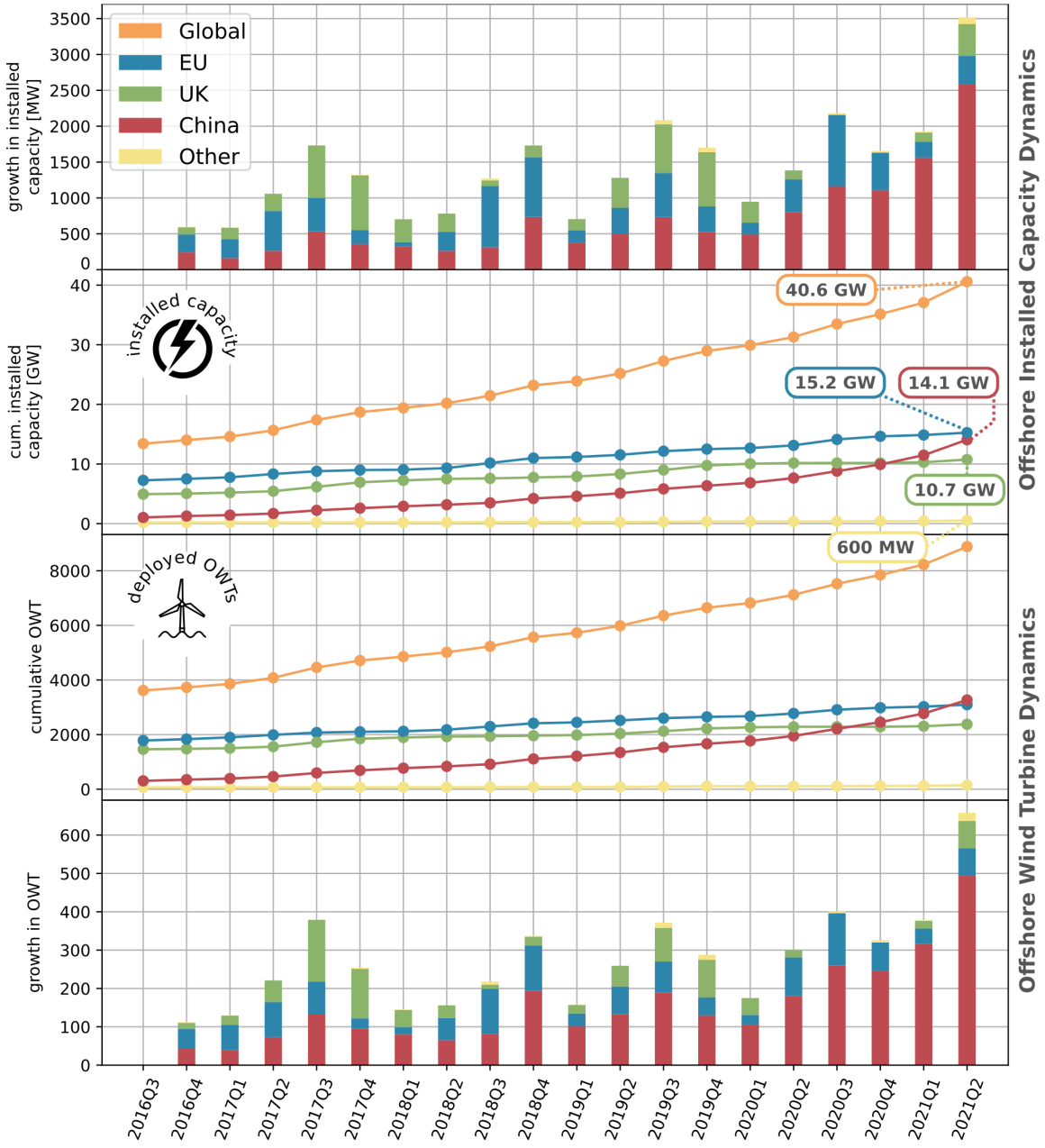



**Figure 6.11:** Derived global and regional temporal dynamics of the installed capacity for offshore wind energy generation and the number of readily deployed offshore wind turbine (OWT). Cumulative values and growth rates are represented separately. From this overview, it becomes clear that in the investigated five years, from July 2016 until June 2021, China is the most important driver of the expansion of the offshore wind energy sector. Source: Hoeser and Kuenzer (2022a, p. 12).

21% to the global offshore wind energy capacity growth between July 2016 and June 2021. Interestingly, China took over EU in the number of installed OWT in June 2021. However, the EU still has a larger amount of installed capacity, see figure 6.11, which indicates the installation of more powerful OWTs in the EU compared to China. Another detail is a very large increase in installed capacity and number of OWTs for China in the last period of the second quarter of 2021. This large increase is strongly related to the final deployment of

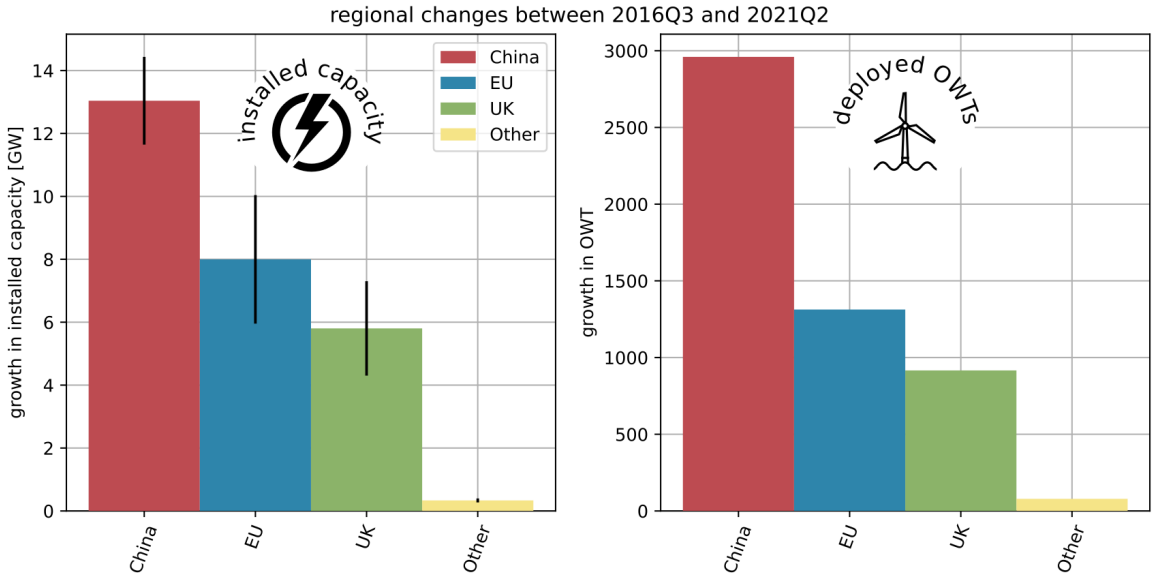



**Figure 6.12:** Derived five-year absolute growth rates of the installed capacity for wind energy generation and readily deployed offshore wind turbine (OWT) of the three major wind energy producers European Union (EU), China and United Kingdom (UK), and other participants in the wind energy sector. Source: Hoeser and Kuenzer (2022a, p. 11).

the 1.1 GW Rudong OWF cluster in the East China Sea (Xing, 2021). The offshore wind energy sector's early history and new developments can be deducted from this observation. With the Danish pilot project, Vindeby in 1991, the first OWF was developed and installed in Europe. Since then, the offshore wind energy market has matured over decades, driven mainly by developments in the EU and the UK, which found direct access to good conditions due to the shallow water depths and high and stable wind speeds in the North Sea Basin. These are the reasons why the EU and the UK have a head start in the development of wind energy in 2016 compared to other regions in the world and have partly maintained it until today. China entered the wind energy sector at a time when technology and deployment processes were at a mature stage of development. This factor, in combination with OWF sites relatively close to the shoreline, is contributing to the rapid growth in the Chinese EEZ over the last five years.

The provided spatial and temporal overview of where and when OWTs are deployed on a global scale demonstrates the flexibility of the compiled DeepOWT data set to aggregate information in time and space. Throughout the compilation of the information for the DeepOWT data set, great attention was paid to the accuracy of the methods used in order to minimise deviations in the number of completed OWTs and their installed capacity. The provided data set evaluation in section 5.5 and error discussions in section 6.1 already provided evidence for high precision and accurate performance. Figure 6.13 further undermines the flexible spatiotemporal information aggregation and consistent temporal continuity. Furthermore, it compares the calculated installed capacities with official reports of the

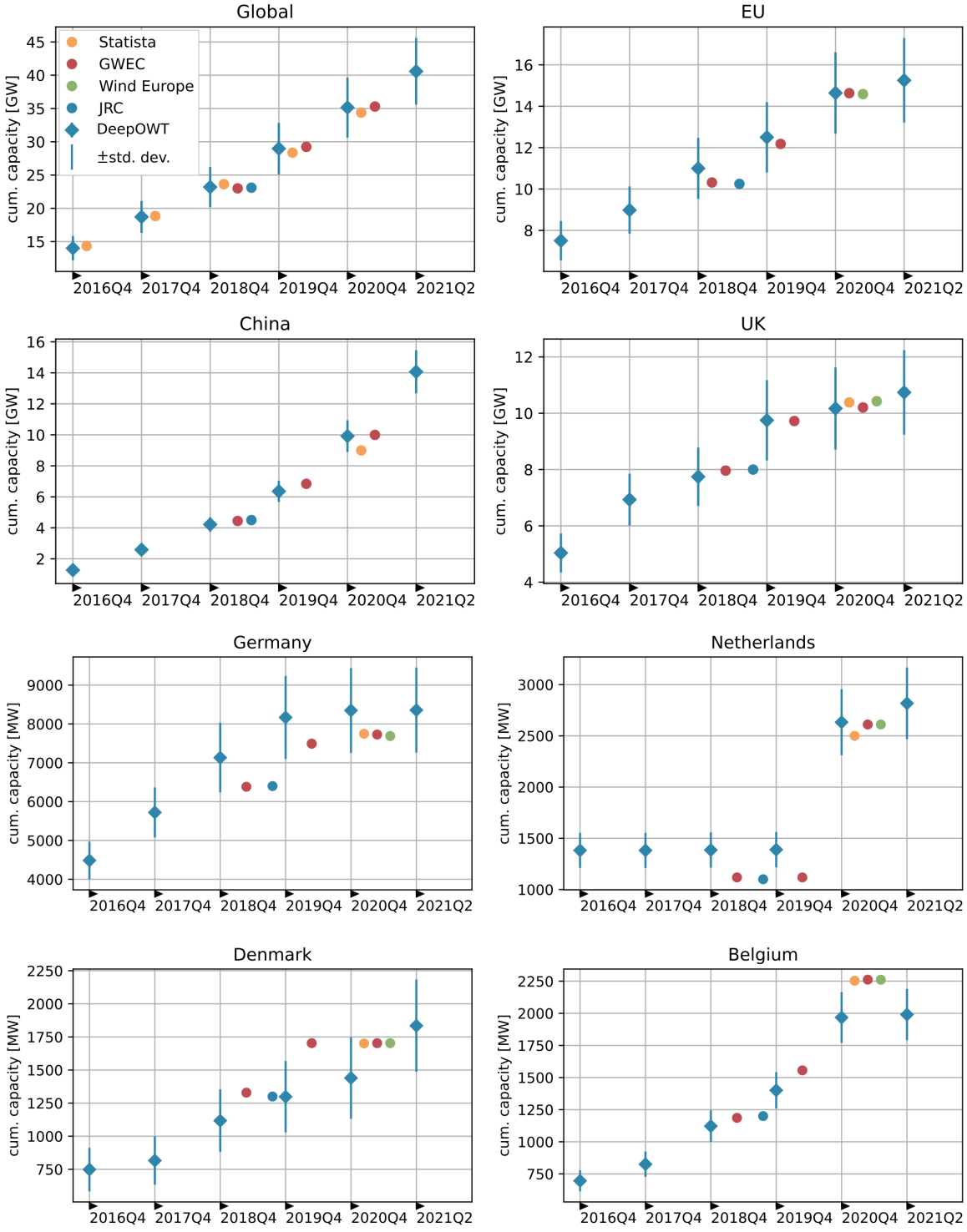



**Figure 6.13:** Comparison of data on global and regional installed capacity of offshore wind energy between the data of this work (DeepOWT) and public reports on this subject. The institutions used for comparison and their reports are Statista (Statista, 2021b,a), the Global Wind Energy Council (GWEC) (Lee and Zhao, 2021, 2020), Wind Europe (Ramírez et al., 2021), and Joint Research Center (JRC) (European Commission et al., 2019). Changed after Hoeser and Kuenzer (2022a, p. 13).

wind energy sector. These comparisons show that the temporary developments of installed capacities based on the DeepOWT data set are in line with the results of official reports by Statista (Statista, 2021b,a), the Global Wind Energy Council (GWEC) (Lee and Zhao, 2021, 2020), Wind Europe (Ramírez et al., 2021), and the Joint Research Center (JRC) (European Commission et al., 2019). It is to emphasise that the proposed methodology to derive the DeepOWT data set allows producing this information continuously. This enables the generation of seamless time series at all scale levels, which is not possible by consulting the aggregated reports used for comparison. Figure 6.13 makes clear that these continuous time series are important when comparing regional trends. Only in the five-year time series of the DeepOWT data set is the trend of strong expansion in China and stagnating construction in the EU and the UK discernible.

Besides global and regional trends of the combined installed capacity, the derived information also provides insights into trends and distributions describing OWT characteristics and their site specifications. Figure 6.14 summarises these trends on a global scale by showing distributions of the site specifications of all OWTs which have been deployed in the corresponding quarter. To better describe major trends but also new developments, quantile regressions of 50% percentile and the upper and lower 5% were performed. The regression takes all quarterly periods into account, besides the first period in 2016Q3, since it includes all OWT ever build until then. For another aggregated overview, the distribution of each site specification is also shown for different regional scales, see figure 6.14.

Figure 6.14a+b) show that on a global scale, the hub height and installed capacity have been increasing since 2016. In particular, smaller turbines with an installed capacity of 4 MW or less were no longer installed. Instead, larger variants with capacities of more than 4 MW were deployed. The technical progress of offshore wind turbine technology is also reflected in regional differences. The EU and UK, which have participated in the market since the beginning of offshore wind energy development, still have a large share of turbines with lower nominal capacity from this period. At the same time, the further development of the technology for the high wind speeds of the North Sea Basin is also taking place here, which means that very large turbines with high output can also be found in the EU and UK. In comparison, China, in particular, and other regions that entered the offshore wind energy market later have benefited from the already researched technology and immediately deployed larger turbines. Especially in China, no turbines of the smallest scale, which can be found in the EU and UK were installed, see figure 6.14a+b). Chinese offshore wind turbine development has been comparatively more focused on similar turbine types with a nominal capacity between 4 and 5 MW in the last five years, which is also the reason for the global trend focussing on this size.

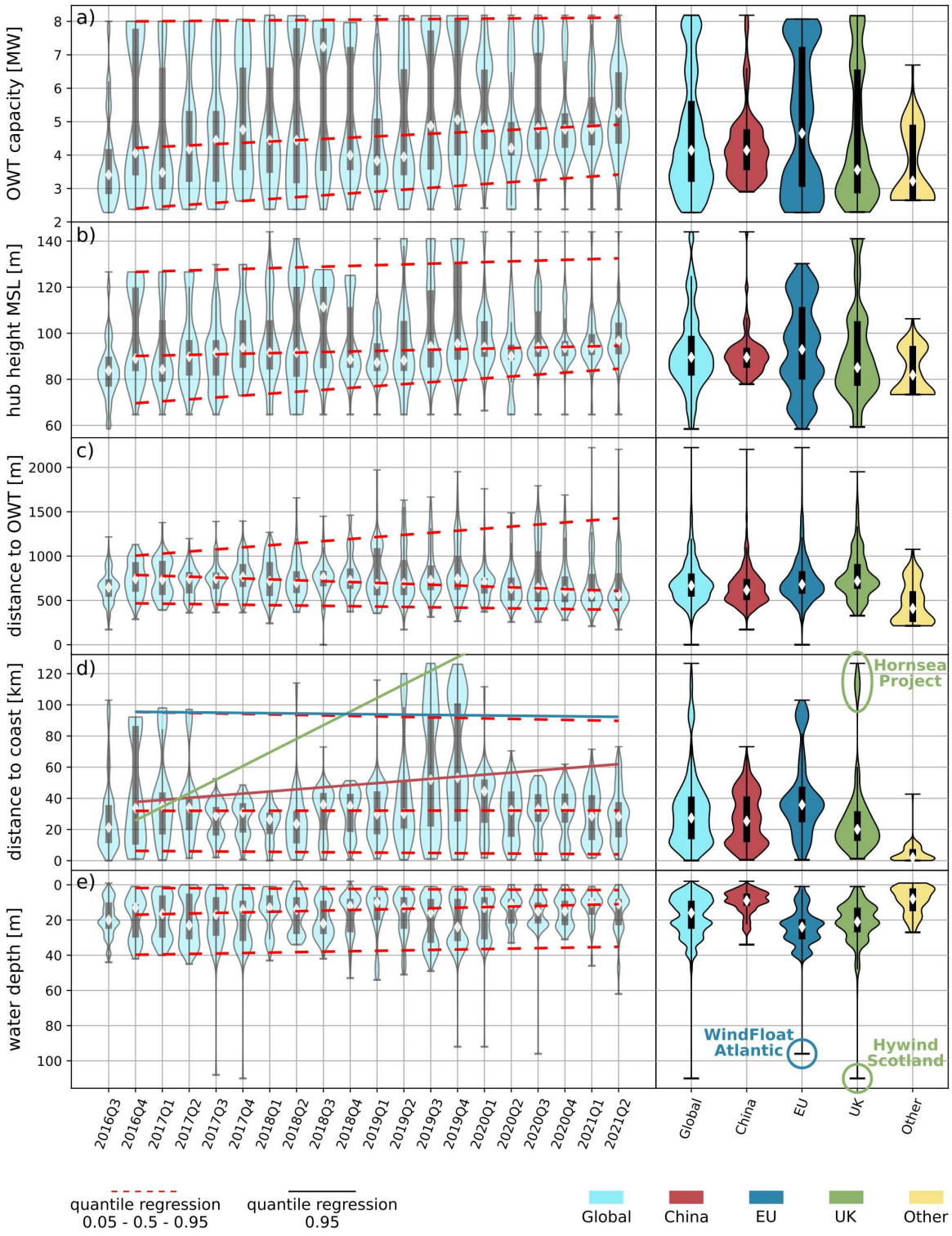



**Figure 6.14:** Temporal dynamics of multiple offshore wind turbine (OWT) characteristics on a global scale between July 2016 and June 2021 and aggregated distributions on regional scales for the European Union (EU), China, the United Kingdom (UK), and other participants of the offshore wind energy sector. The dashed red lines represent 5%, 50% and 95% quantile regressions of global trends, the solid lines in d) are the 95% quantile regressions of the EU, China, and the UK. Changed after Hoeser and Kuenzer (2022a, p. 14).

Related to the trends of hub height and installed capacity, wind turbine density, or the minimum distance between OWTs, shows an increasing trend for the upper 5% percentile, see figure 6.14c). However, the primary trend of minimum distances between OWTs in the last five years is slightly decreasing. This is due to the larger amount of medium-sized OWTs with a medium rotor size and thus a reduced need for widely spaced installations, mainly deployed by China but also other nations which joined the offshore wind energy sector lately. Since more medium spaced OWT with a smaller minimum OWT distance are necessary to realise the same installed capacity, they dominate the 50% percentile trend of decreasing minimum distance, even when new developments start in an opposite direction as indicated by the upper 5% percentile. Following that, the trend will most probably turn in the upcoming years due to developments of considerably larger OWT models, which demand a larger minimum distance to the neighbouring OWT. It may be of interest to follow this development closely since it affects the installed capacity per area ratio of OWF and thus has a direct influence on the sometimes competing demands of different stakeholders for utilisation in already heavily exploited marine areas.

Figure 6.14d) shows the development of the OWT distances to the coast. The global trend is less pronounced, which is not surprising since this specific measure depends particularly on the respective natural conditions but also on the stages of development of the regional offshore wind market and is therefore not as significant in a large-scale aggregated view. Therefore three regional trends of the upper 5% percentile are provided. For the UK the trend of the distance to the coast is strongly increasing, and when looking at the regional distribution to the right of the temporal trend in figure 6.14d) it becomes clear that after a period of nearshore sites, far offshore sites are getting developed. This indicates an advanced stage of the regional offshore wind energy sector in the UK, as for cost reasons, the coastal areas are exhausted first before more distant areas have to be developed for offshore wind energy production. The main driver for this trend to appear for the UK market, is the Hornsea project, which in June 2021 has a maximum distance to the coast of over 120 km. This trend will also continue in the future, as the UK's ambitious expansion targets require large areas that can no longer be developed close to shore (Higgins and Foley, 2014). An example of this are the planned OWF projects on the Dogger Bank with distances of over 200 km from the coast. However, at the Dogger Bank are favourable conditions, such as shallow water depths and consistently high wind speeds (Cotterill et al., 2012).

The selection of site locations of offshore wind energy projects in the Chinese EEZ are different from what can be seen recently in the UK. Most of the deployed OWTs of the last five years were constructed close to the shoreline to utilise these still untapped and easier to develop areas first. In fact, one of the largest Chinese OWF projects, the Jiangsu Rudong

Offshore Intertidal Demonstration Wind Farm, has been constructed, as the name suggests, directly on tidal flats, which are natural transition zones between the open sea and mainland. China developed most of its offshore wind energy projects close to the shoreline and only slowly moved away from the coast. This strategy allowed China to deploy their OWT fast and efficiently, which is one explanation for the massive Chinese expansion in the last five years. However, when looking at the water depth at the Chinese coast in figure 6.6 it becomes clear that China cannot easily build conventional offshore wind farms with foundations directly connected to the seabed hundreds of kilometres from the coast because the water depths here are too deep and the conditions are therefore different from those in the North Sea Basin. From a perspective of maximising installed offshore wind energy capacity, China needs to make greater use of coastal areas, as confirmed by the expansion activities of the last five years.

Similar to the global trend of distances to the coast, the trend of the EU is mixed and with it ambiguous. Too many separate national offshore wind strategies with different stages and EEZ areas in which the OWTs are getting constructed, mix different distances to the coast and obscure a trend on an EU scale. However, the overall distribution of distances to the coast for the entire EU highlights some other aspects which go along with observations of the water depth in figure 6.14e). Similar to China in the East China Sea, the EU has large areas in the North Sea Basin with tidal flats, which in China are partly used for large scale OWF projects. However, the distribution of distances to the coast shows that in the EU coastal areas are mostly kept free of offshore wind farms and constructions starting further offshore. The difference between the usage of coastal areas and shallow waters becomes even more evident when comparing the water depth of OWTs in context with the distance to the coast, see figure 6.14d-e). Despite shallow water depths in the North Sea Basin, near the coast and on tidal flats, OWT mostly appear in deeper water depth. Which is another indicator that in the EU, coastal areas and tidal flats are intentionally kept free of OWF. An important reason for this are exclusion zones, especially natural conservation areas for coastal ecosystems and tidal flats, which balance area requirements of the offshore wind energy industry and concerns for the preservation of ecological diversity and recreational space. In contrast, in China, nearly all OWTs within the last five years are set up in shallow waters, see figure 6.14e) and closer to the coast, which eases the deployment process and is more efficient from an offshore wind energy infrastructure deployment perspective.

Despite the rather constant trends, the water depth characteristics in figure 6.14e) reveal one of the most recent and highly important developments in the offshore wind energy sector. The global temporal trend and the corresponding regional summaries of the EU and UK show two groups of outliers. They belong to two pilot projects of floating OWFs,

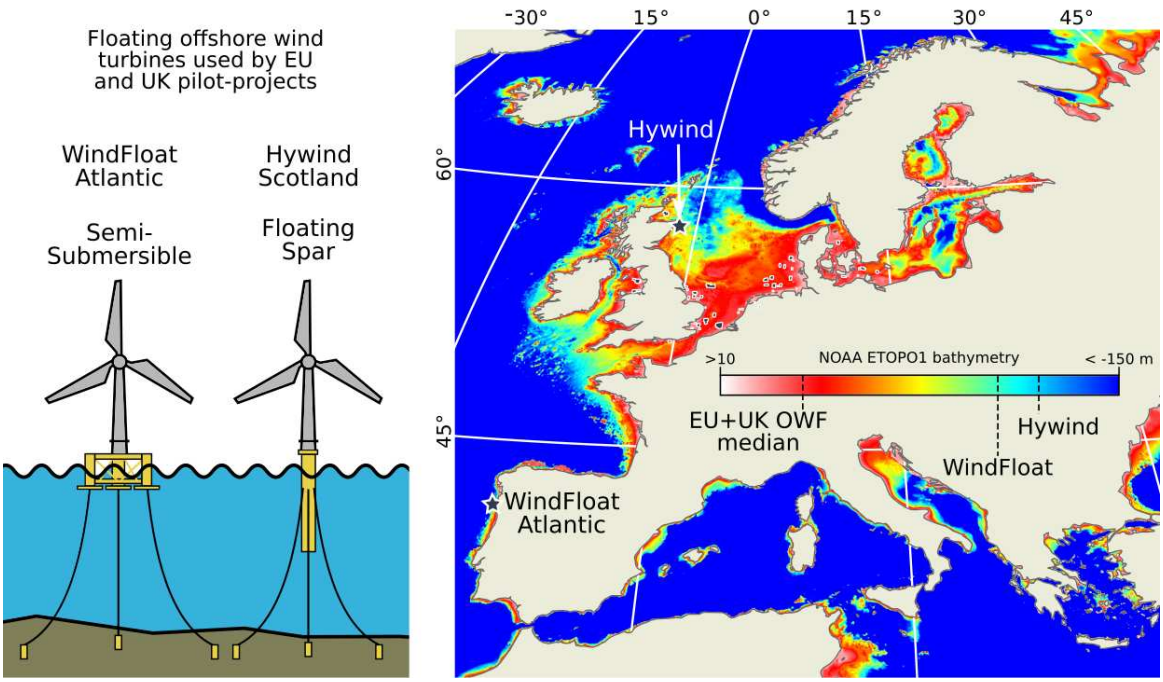



**Figure 6.15:** Two European Union (EU) and United Kingdom (UK) floating offshore wind farm projects with different platform types and their location in context with water depths for the EU and UK. The two water depths of the floating offshore wind turbine demonstrate which possibilities this technology offers when their water depths are compared to the median installation water depth of offshore wind farms in the EU so far. Source: Hoeser and Kuenzer (2022a, p. 15).

the WindFloat project in the Atlantic at the coast of Portugal with a water depth of about 95 m (Energias de Portugal, 2022), and the Hywind project in the North Sea Basin at the Scottish coast with a maximum water depth of 120 m (Equinor, 2022), see figure 6.15. For comparison, the median water depth of OWT sites in the EU is 22 m for common OWT foundation types based on the seabed. With the technology of floating OWTs, offshore wind energy production at water depths considerably deeper than 50 m is possible, and with it, the potential area in which offshore wind energy can be generated becomes significantly larger (Henderson and Witcher, 2010). This is particularly true for the EU's riparian states on the Atlantic coast and in the Mediterranean region, see figure 6.15. The 300 GW target for the EU by 2050 requires densification and further expansion in the North Sea Basin and Baltic Sea, but with new technologies that are about to be deployed in deeper water depths, such as floating wind turbines, it is also possible to integrate areas that have not yet participated in the European offshore wind energy market. This means that the realisation of the expansion targets is spread over a larger area and includes more EU members. With the open DeepOWT data set, there is a basis to spatially locate knowledge and experience from already implemented wind energy projects at an early stage and exchange it more easily, especially with regions new to the offshore wind market.

## 6.4 Discussion

The radargrammetric interpretation of the typical radar signature of an offshore wind turbine is the starting point for the calculation of the installed capacity. Any improvement in the height calculation leads to a direct improvement in the estimate of the installed capacity. The method presented here combines information from individual images and median composites of a single orbit direction. Future refinement of this method should investigate the calculation of hub heights based solely on the evaluation of individual images and from both orbit directions, hence it could further increase the accuracy of the hub heights information. However, this requires a much greater effort in the preprocessing of individual acquisitions. In order to advance this method for an even more precise calculation of the height information, additional ground truth data, in particular from OWT currently under construction, are also necessary.

More data about turbines which are currently under construction, is also essential for future modelling of the installed capacity. The proposed model for mapping hub height to installed capacity only considers OWTs with a maximum installed capacity of about 9 MW. However, recent advances in turbine technology have led to the deployment of OWTs with an installed capacity up to 14 MW. A number that will continue to rise in the future. Due to the shape of the sigmoid regression model, an update of hub heights and installed capacity values is necessary to keep track of recent developments. Furthermore, it should be checked to what extent the other derived parameters, in particular the time of installation of the turbine, can be included in the model. Together with an even more precise hub height calculation, the model's confidence interval can be kept narrow, and the model continues to provide an accurate estimate for the installed capacity as it did for the period investigated in this work.

The attributes derived for each turbine location are an important addition to the spatiotemporal information of the DeepOWT data set. Their integration into the data set increases the application potential of the data significantly. In addition to the locations, specific turbine heights and capacities can now be taken into account. They can directly be employed to model energy generation with meteorological data or to investigate the large and small-scale effects of wind turbines on wind speeds and surface water. In order for this data set to reach its audience, it is important that the data is regularly updated. For this purpose, the methods and technical processes in this chapter have to be consolidated. In view of the sensor failure of the Sentinel-1B instrument in December 2021 (ESA, 2022), the introduced methodologies should be reviewed and, if necessary, adapted to make them robust against such events.

The analysis in this chapter shows how the most important developments of the entire global wind energy sector can be observed with the help of remote sensing data. The presented analysis of the five-year period between July 2016 and June 2021 shows the flexibility with which the derived information can be evaluated in space and time with a high degree of accuracy. Global trends could be analysed as well as regional differences and specific developments in individual regions.

# *Chapter 7*

# *Synthesis and Outlook*

In order to review the methods developed and results obtained in this dissertation, they are brought together in this concluding chapter. By working on the five research objectives, three major deliverables have been produced in this dissertation: The development of the SyntEO framework for the generation of synthetic remote sensing data ready for deep learning applications, the deep learning derived global offshore wind energy infrastructure data set DeepOWT, and finally its progressive analysis and description of recent developments in the global offshore wind energy sector.

## 7.1 Summary and Conclusive Findings

Driven by an alternative to coal-fired power and other fossil fuels with the aim of reducing greenhouse gases, or to achieve greater energy autonomy for states, the expansion of wind energy will increase strongly in the upcoming decades and offshore wind energy is at the beginning of a massive expansion worldwide. On a global level, this is one of the largest marine infrastructure projects of all time. In addition to the challenge of implementing this expansion technically, it is also important to design it sustainably at an early stage. Therefore, integrated spatial planning that takes into account the concerns of all stakeholders involved is of great importance. The deployment of infrastructure to this extent also requires multidisciplinary scientific research and continuous monitoring, of the expansion itself but also of its ecological, economic and social impacts. A freely available, precise, spatially and temporally homogeneous and nevertheless global database of offshore wind energy infrastrucutre that can be continued in the decades to come is a fundamental and particularly important element in this context.

With satellite-based Earth observation data, especially from ESA's Sentinel-1 mission, a data basis is available that provides both the spatial and temporal requirements for such a database. Modern image analysis procedures, such as deep learning-based object detection,

have proven their potential especially in the computer vision domains over the last decade. Their use to derive the desired database from satellite imagery requires the development of specific methods adapted for Earth observation.

Given the challenges of the global offshore wind energy expansion and possibilities to monitor this expansion by combining Earth observation data with deep learning methods, the goal of this dissertation is as follows: Create and provide a scientifically and socially necessary, freely accessible database, which globally maps the progressive expansion of offshore wind energy infrastructures. Derive this database by investigating novel methodological developments to successfully apply modern deep learning based image processing in the Earth observation domain. To achieve this goal, the following research objectives are addressed in this dissertation.

***Research objective 1*** Provide an introduction to the fundamentals of supervised deep learning and convolutional neural networks, and conduct a literature review of how convolutional neural networks are applied in the Earth observation domain, in order to identify common practices and recent challenges specific for the Earth observation domain.

1. What are the key properties of the convolutional neural network that make it particularly suitable for analysing image data?
2. What is the value of training data for optimising a convolutional neural network?
3. How has the convolutional neural network been applied to Earth observation and what is a current obstacle in transferring this methodology from the field of computer vision?

The introduction to convolutional neural networks (CNNs) presented in chapter 2 embeds this model type in its superordinate categories of deep learning, machine learning and, starting at the highest level, artificial intelligence. During a detailed introduction to the fields of machine and deep learning, the characteristics of CNNs that make them particularly suitable for analysing image data were highlighted. This includes, above all, learning 2D kernel functions that build on each other and enable complex spatial features to be captured in images. In addition to the detailed discussion of individual aspects of the models and procedures from the fields of machine learning, it was worked out that the extent and variability of training data sets are of particular importance for supervised deep learning with CNNs. Without sufficient data, these complex models cannot be adequately optimised or are overfitted to the training data so that no generalisation of the model can be achieved. Following this general introduction to CNNs, a comprehensive review of 429 articles on

the use of CNNs in Earth observation was conducted. Overall, the analysis of these articles describes the application of CNNs in an advanced phase of the transfer of the methodology from the field of computer vision to Earth observation. This can be seen in the model and data types used in Earth observation, which are largely similar to those from the computer vision domain and are only just beginning to be developed specifically for Earth observation. At the time of the review, training data sets were being reused to investigate certain model optimisation approaches or make methodological progress. However, the number of studies that are creating their own training data sets to answer new geoscientific questions has only recently started to increase. As a result, CNNs have often been used in Earth observation for those applications for which there are existing training data. From this, the need for research can be derived to develop methods that allow researchers from the Earth observation domain to have resource efficient, large, variable training data sets created to further expand the range of applications for CNNs in Earth observation.

***Research objective 2*** Provide an overview of how offshore wind energy infrastructure has been detected in Earth observation so far and give detailed insight into how offshore wind energy infrastructure and its common environment appear in Sentinel-1 radar data at different spatial scales.

1. How has the detection of offshore wind energy infrastructure been researched in Earth observation so far?
2. What spatial features exist in radar images to detect and distinguish offshore wind energy infrastructure from other marine infrastructure and natural environments?

Chapter 3 starts with a review of methods to detect marine infrastructure in Earth observation data. Persistent marine infrastructure, and offshore wind energy infrastructure in particular, has been detected primarily using the established constant false alarm rate method. Other morphological examinations of the objects, coupled with manually determined thresholds, also make use of the high contrasts that arise between marine infrastructures and the surrounding water. These contrasts are clearly represented in both optical and radar remote sensing data. However, radar data is particularly useful as it is weather-independent and thus provides a higher temporal coverage in coastal and offshore areas. The so far published procedures indicate that offshore wind energy infrastructure is detectable, but its internal distinction cannot be performed automatically by the hand-programmed features. Thus, for example, transformer stations or platforms under construction are misclassified as wind turbines on a small spatial scale. Moreover, on a larger spatial scale, the detection of entire offshore wind farms in heterogeneous environments has to be supported

by human interpretation in order to distinguish them from other marine infrastructures and marine environments. However, a close look at the radar signal provided by the Sentinel-1 mission globally shows spatial patterns that provide an opportunity to exploit these features with complex analysis in order to not only detect offshore wind energy infrastructures, but also to distinguish them from each other. These spatial patterns can be explained by the imaging geometry and typical effects of radar instruments. In addition to these small-scale spatial features, a detailed discussion of large-scale patterns of offshore wind energy infrastructures was also pointed out. Their typical geometric arrangement of individual objects clearly stands out from other marine infrastructures and natural environments. This has highlighted unique features of the targets to be detected at different scales, which will form the basis for training CNNs to incorporate these features into their decisions. However, to teach these features to a CNN, large variable training data sets are needed, as highlighted in research objective 1 and chapter 2. In addition to the resource expensive effort of creating large training data sets by hand, the use case of offshore wind energy infrastructure is complicated by the fact that there are too few real-world examples of offshore wind farms worldwide to compile such a data set. This requires the development of a different approach to build such a training data set.

***Research objective 3*** Develop a conceptual framework to automatically generate large amounts of synthetic and task specific labelled Earth observation data by taking expert knowledge into account.

1. What structure can be used to uniformly represent expert knowledge in order to describe nested systems such as a remote sensing scene?
2. What properties must a synthetic training data set have in order to optimise a CNN so that it can reliably detect a target and at the same time distinguish it from its natural environment?

Based on the successive findings and need for research from research objective 1 and 2, in chapter 4, the SyntEO framework was developed for the creation of synthetic training data in Earth observation. In order to be able to map expert knowledge about remote sensing sensors, natural spaces and human influences as well as the nested multi-scale structure of spatial data, a special ontology was developed in the SyntEO framework. By formulating this ontology human expert knowledge becomes machine readable and can be used by an image processing backend to automatically generate complex synthetic remote sensing scenes, from which a synthetic remote sensing image is taken along with its task specific annotation. During the development of the SyntEO framework, a proof of concept study for the detection of entire offshore wind farms was carried out. It became clear that, in

addition to the object to be detected, potential false positives must also be contained in a synthetic training data set in order to ensure sufficient generalisation of the features to be learned. Finally, SyntEO is a flexible framework that enables the embedding of expert knowledge in the training data and can thus generate very large and highly variable data sets. Furthermore, SyntEO enables the structured creation of balanced training data sets, as metadata is available for each synthetically generated image, which can be taken into account when compiling the final training examples. In this way, it was possible to create precisely controllable training environments in the proof of concept study, which also made it possible to draw conclusions about the learning process of the deep learning model and thus brought human and machine closer together throughout the entire training process.

***Research objective 4*** Derive a global, multi-temporal data set of offshore wind energy infrastructure from Earth observation data by combining the previously developed framework for synthetic data generation and deep learning based object detection.

1. How can CNNs, optimised by synthetic training data, be used for object detection to globally identify individual types of offshore wind energy infrastructure in radar data?
2. How can quarterly changes in the deployment process of the offshore wind energy infrastructure be determined?
3. How many offshore wind turbines and substations had been deployed around the world by mid-2021 and how many were under construction?

In chapter 5, the methods developed in the previous research objective 3 were eventually employed for an automatic detection of individual objects such as offshore wind turbines, platforms under construction and offshore wind farm substations on the entire global coastline. Therefore, a cascade of two object detector CNNs was used, where the first stage detects large scale offshore wind farms, and the second stage detects and distinguishes single offshore wind infrastructure platform types. Based on the spatially precise localisation by the cascading CNN object detectors, the temporal signal for each object was investigated in a further step. Here, changes in the small scale spatial radar signature of a single object where tracked in a temporal stack over five years to derive the deployment stages of each infrastructure object. The combination of the spatial detection and temporal interpretation of the radar signal resulted in the DeepOWT data set. The data set covers the entire world for a five-year period divided into three-month intervals from July 2016 to June 2021. For the latest entry in the data set in June 2021, 8,885 offshore wind turbines (OWTs), 852 offshore wind energy infrastructure objects under construction, and 204 offshore wind farm substations were detected. The evaluation of the data set demonstrates the accuracy of the

applied methods, especially in the detection of offshore wind turbines, where an F1 scores of over 98% were achieved. Furthermore, the approaches used to create the data set are robust to different conditions, as shown by the detailed evaluation and comparison of results in two differing test sites, the North Sea Basin and East China Sea. Since the approach used to generate the DeepOWT data set is based exclusively on synthetic training data generated with the SyntEO framework, the data set clearly demonstrates the applicability of the developed framework. In order to promote the widest possible use of the DeepOWT data set, the entire data set was converted into a lightweight data format that can be read directly into GIS software and made freely accessible[1]. Furthermore, the data set was integrated into the Coastal Explorer[2], a openly accessible online mapping service of the German Aerospace Center, in order to provide a direct overview of the derived information.

***Research objective 5*** Expand the spatiotemporal information of the derived offshore wind energy infrastructure data set with important technical attributes such as the installed capacity of an offshore wind turbine as well as further spatial information about its location and conduct a global analysis about the dynamics of the offshore wind energy sector over the last five years between July 2016 and June 2021.

1. How can the installed capacity of an offshore wind turbine be determined using spaceborne radar data?
2. How far has the global expansion of offshore wind energy progressed by June 2021 measured in installed capacity and which regional differences can be identified?
3. What trends in the expansion of offshore wind energy have developed globally and regionally over the last five years?

A key parameter for describing offshore wind energy projects and strategies is the installed capacity. With the DeepOWT data set so far, the number and spatial distribution of offshore wind turbines can be described, but not the corresponding installed capacity. In order to determine this information for each turbine in the DeepOWT data set, chapter 6 presents a radargrammetric calculation of the hub height for each of the detected offshore wind turbines. In a subsequent step, the installed capacity was estimated by a statistical model that links the calculated hub height and the installed capacity. With the addition of further site parameters, such as nationality, water depth, and distances to the coast and nearest turbine, together with the five-year deployment time series of the DeepOWT data set, a global analysis of the development of the offshore wind energy sector was carried out.

[1] https://zenodo.org/record/5933967
[2] https://coastalx.eoc.dlr.de

In the second quarter of 2021, 40.6 GW of offshore wind energy installed capacity was available worldwide coming from 8,885 offshore wind turbines. Of this, 15.2 GW was in the EU, 14.1 GW in China, and 10.7 GW in the UK. In the last five years, a total of 5,268 offshore wind turbines with a combined capacity of 27.2 GW have been deployed globally. The most important driver of this growth is China with 13 GW. Europe has deployed 8 GW installed capacity in the same time, the UK 5.8 GW. That makes clear, that China has deployed almost its entire offshore wind energy infrastructure in the last five years. The EU and UK have approximately doubled their capacity in the same period. Their participation in the offshore wind energy sector reaches back until 1991. The massive expansion activities of the last five years in China have mainly taken place near the coast and in shallow waters by exploiting the untapped areas in the young Chinese offshore wind energy sector. This contrasts with the offshore wind energy sector in the UK, which has had to develop increasingly offshore areas over the last five years to accommodate the needs of very large offshore wind energy projects, partly because nearshore areas are being utilised by already established offshore wind farms. An evaluation of the spatial distribution of the derived turbine locations and the time of their construction has made this analysis possible. Furthermore by taking the water the depths of the derived turbine locations into account, a newly emerging trend can be seen. In the EU and the UK, pilot projects of floating offshore wind turbines are being tested in significantly deeper water depths. In future, they will complement the established foundation types, which are built directly on the seabed. This will allow future planning to take locations into account where no offshore wind energy has been developed so far. In the EU this is mainly the Mediterranean and the Atlantic coast, which offers more EU members to contribute to the necessary energy production transformation with the development of offshore wind energy projects.

## 7.2 Future Challenges and Opportunities

The summarised research objectives and their corresponding results demonstrate the potential of Earth observation data to gather detailed insights on a global scale by applying the methods developed in this dissertation. Both the developed methodological approaches and the derived database on the global development of the offshore wind energy sector offer great potential for future use and development.

Figure 7.1 revisits the EU, UK and German offshore wind energy strategies mentioned in the introduction, see chapter 1, and presents their goals in relation to the already realised installed capacity of offshore wind energy infrastructure as identified in this work. The gap to be closed in this illustration does not only exist in the expansion of offshore wind energy infrastructure but also in a continuous report of the corresponding data. The

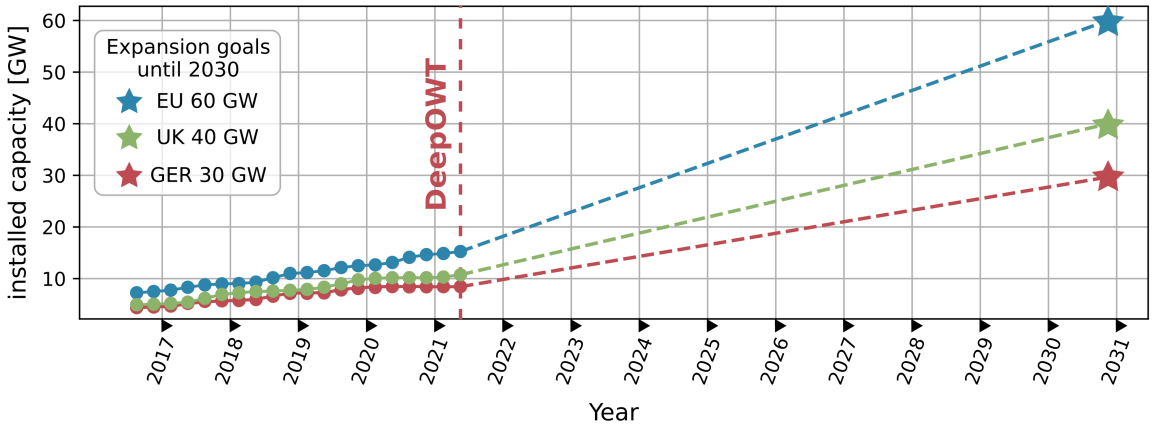



**Figure 7.1:** Development of the offshore wind energy installed capacity for the European Union (EU), United Kingdom (UK) and Germany (GER) reported by the DeepOWT data set and the expansion goals set by the offshore wind energy strategies for these regions.

global detection of offshore wind energy infrastructure must now be consolidated with the developed procedures in order to accompany the massive expansion in a timely manner and to guarantee the necessary data basis for associated research and decision-makers. An urgent expansion of offshore wind energy requires a freely accessible database of the spatiotemporal dynamics of the offshore wind energy sector to foster a timely reporting and the timely involvement of all stakeholders in order to achieve the ambitious targets of recent offshore wind energy strategies. Remote sensing data can, as this work has shown, contribute to such a database efficiently for the entire world if its analysis continues in the future. In addition to a further continuation of the DeepOWT data set, the introduced methods should be refined in the future in order to be able to describe, for example, an even higher temporal resolution of the changes in the deployment process of the wind turbines, or to be able to infer the installed capacity for new turbine types as well as is currently possible.

The SyntEO framework developed to generate synthetic training examples provides a basis for a variety of future scientific investigations and similar remote sensing products such as the DeepOWT data set. The use of modern processing backends such as 3D engines for the generation of synthetic high-resolution RGB images will open up new possibilities to make the expert knowledge that is represented in a SyntEO ontology available for deep learning analysis. As the number of environments, described by the SyntEO ontology increases, a library of representations grows that will collectively make it easier to generate complex remote sensing scenes synthetically. This will make it possible to map increasingly complex descriptions of expert knowledge and use them as a direct training signal in the optimisation of artificial intelligences. This will not only enable a wide range of potential applications but also increasingly close the gap between human knowledge and artificial intelligence. With the rapid generation of large and variable training data sets, SyntEO strengthens the use of modern deep learning methods in Earth observation.

# *Bibliography*


Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., Zheng, X., 2015. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. Last visited: 07.01.2022. URL: https://www.tensorflow.org/.

Amante, C., Eakins, B.W., 2009. ETOPO1 1 Arc-Minute Global Relief Model: Procedures, Data Sources and Analysis. NOAA Technical Memorandum NESDIS NGDC-24v. Last visited: 10.02.2022. URL: https://www.ngdc.noaa.gov/mgg/global/global.html, doi: https://doi.org/10.7289/V5C8276M.

Anselin, L., 1995. Local Indicators of Spatial Association—LISA. Geographical Analysis 27, 93–115. doi: https://doi.org/10.1111/j.1538-4632.1995.tb00338.x.

Arneth, A., Denton, F., Agus, F., Elbehri, A., Erb, K., Osman Elasha, B., Rahimi, M., Rounsevell, M., Spence, A., Valentini, R., 2019. Framing and Context, in: Shukla, P.R., Skea, J., Calvo Buendi, E., Masson-Delmotte, V., Pörtner, H.O., Roberts, D.C., Zhai, P., Slade, R., Connors, S., Diemen, R.v., Ferrat, M., Haughey, E., Luz, S., Neogi, S., Pathak, M., Petzold, J., Portugal Pereira, J., Vyas, P., Huntley, E., Kissick, K., Belkacemi, M., Malley, J. (Eds.), Climate Change and Land: an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems, pp. 77–129.

Aschbacher, J., 2017. ESA's earth observation strategy and Copernicus, in: Satellite earth observations and their impact on society and policy. Springer, Singapore, pp. 81–86. doi: https://doi.org/10.1007/978-981-10-3713-9_5.

Auer, S., Bamler, R., Reinartz, P., 2016. RaySAR - 3D SAR simulator: Now open source, in: 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), pp. 6730–6733. doi: https://doi.org/10.1109/IGARSS.2016.7730757.

Badger, J., Hahmann, A., Larsén, X., Badger, M., Kelly, M., Olsen, B., Mortensen, N., 2015. The Global Wind Atlas: An EUDP project carried out by DTU Wind Energy. DTU Wind Energy. URL: https://backend.orbit.dtu.dk/ws/portalfiles/portal/238494910/GWA_64011_0347_FinalReport.pdf.

Bailey, H., Brookes, K.L., Thompson, P.M., 2014. Assessing environmental impacts of offshore wind farms: lessons learned and recommendations for the future. Aquatic Biosystems 10. doi: https://doi.org/10.1186/2046-9063-10-8.

Baldridge, A., Hook, S., Grove, C., Rivera, G., 2009. The ASTER spectral library version 2.0. Remote Sensing of Environment 113, 711–715. doi: https://doi.org/10.1016/j.rse.2008.11.007.

Ball, J.E., Anderson, D.T., Sr., C.S.C., 2017. Comprehensive survey of deep learning in remote sensing: theories, tools, and challenges for the community. Journal of Applied Remote Sensing 11, 1 – 54. doi: https://doi.org/10.1117/1.JRS.11.042609.

Baumhoer, C.A., Dietz, A.J., Kneisel, C., Kuenzer, C., 2019. Automated Extraction of Antarctic Glacier and Ice Shelf Fronts from Sentinel-1 Imagery Using Deep Learning. Remote Sensing 11. doi: https://doi.org/10.3390/rs11212529.

Bengio, Y., 2013. Deep Learning of Representations: Looking Forward, in: Dediu, A.H., Martin-Vide, C., Mitkov, R., Truthe, B. (Eds.), Statistical Language and Speech Processing, Springer Berlin Heidelberg, Berlin, Heidelberg. pp. 1–37. doi: https://doi.org/10.1007/978-3-642-39593-2_1.

Bergström, L., Kautsky, L., Malm, T., Rosenberg, R., Wahlberg, M., Capetillo, N.Å., Wilhelmsson, D., 2014. Effects of offshore wind farms on marine wildlife - a generalized impact assessment. Environmental Research Letters 9, 034012. doi: https://doi.org/10.1088/1748-9326/9/3/034012.

Berkson, E.E., VanCor, J.D., Esposito, S., Chern, G., Pritt, M., 2019. Synthetic Data Generation to Mitigate the Low/No-Shot Problem in Machine Learning, in: 2019 IEEE Applied Imagery Pattern Recognition Workshop (AIPR), pp. 1–7. doi: https://doi.org/10.1109/AIPR47015.2019.9174596.

Bittner, K., Adam, F., Cui, S., Körner, M., Reinartz, P., 2018. Building Footprint Extraction From VHR Remote Sensing Images Combined With Normalized DSMs Using Fused Fully Convolutional Networks. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 11, 2615–2629. doi: https://doi.org/10.1109/JSTARS.2018.2849363.

Blender Online Community, 2018. Blender - a 3D modelling and rendering package. Blender Foundation. Stichting Blender Foundation, Amsterdam. URL: http://www.blender.org.

BMWK, 2022. Bundesministerium für Wirtschaft und Klimaschutz: Überblickspapier Osterpaket. Last visited: 06.04.2022. URL: https://www.bmwk.de/Redaktion/DE/Downloads/Energie/0406_ueberblickspapier_osterpaket.pdf?__blob=publicationFile&v=12.

Bonatti, R., Madaan, R., Vineet, V., Scherer, S., Kapoor, A., 2019. Learning Visuomotor Policies for Aerial Navigation Using Cross-Modal Representations. arXiv, Computer Vision and Pattern Recognition cs.CV. doi: https://doi.org/10.48550/ARXIV.1909.06993.

Bonatti, R., Wang, W., Ho, C., Ahuja, A., Gschwindt, M., Camci, E., Kayacan, E., Choudhury, S., Scherer, S.A., 2020. Autonomous aerial cinematography in unstructured environments with learned artistic decision-making. Journal of Field Robotics 37, 606 – 641. doi: https://doi.org/10.1002/rob.21931.

Bottou, L., Curtis, F.E., Nocedal, J., 2018. Optimization Methods for Large-Scale Machine Learning. SIAM Review 60, 223–311. doi: https://doi.org/10.1137/16M1080173.

Bovik, A., Huang, T., Munson, D., 1983. A generalization of median filtering using linear combinations of order statistics. IEEE Transactions on Acoustics, Speech, and Signal Processing 31, 1342–1350. doi: https://doi.org/10.1109/TASSP.1983.1164247.

Brodeur, S., Perez, E., Anand, A., Golemo, F., Celotti, L., Strub, F., Rouat, J., Larochelle, H., Courville, A., 2017. HoME: a Household Multimodal Environment. arXiv, Artificial Intelligence cs.AI. doi: https://doi.org/10.48550/ARXIV.1711.11017.

Brown, M.B., Forsythe, A.B., 1974. Robust Tests for the Equality of Variances. Journal of the American Statistical Association 69, 364–367. doi: https://doi.org/10.2307/2285659.

Cadieu, C.F., Hong, H., Yamins, D.L.K., Pinto, N., Ardila, D., Solomon, E.A., Majaj, N.J., DiCarlo, J.J., 2014. Deep Neural Networks Rival the Representation of Primate IT Cortex for Core Visual Object Recognition. PLOS Computational Biology 10, 1–18. doi: https://doi.org/10.1371/journal.pcbi.1003963.

Cai, Z., Vasconcelos, N., 2018. Cascade R-CNN: Delving Into High Quality Object Detection. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition , 6154–6162. doi: https://doi.org/10.1109/CVPR.2018.00644.

Cauchy, A.L., 1847. Méthode générale pour la résolution des systèmes d'équations simultanées. Comptes rendus hebdomadaires des séances de l'Académie des sciences - Analyse Mathématique , 536–538.

Cavazzi, S., Dutton, A., 2016. An Offshore Wind Energy Geographic Information System (OWE-GIS) for assessment of the UK's offshore wind energy potential. Renewable Energy 87, 212–228. doi: https://doi.org/10.1016/j.renene.2015.09.021.

Chen, D., Rojas, M., Samset, B.H., Cobb, K., Diongue Niang, A., Edwards, P., Emori, S., Faria, S.H., Hawkins, E., Hope, P., Huybrechts, P., Meinshausen, M., Mustafa, S.K., Plattner, G.K., Tréguier, A.M., 2021. Framing, Context, and Methods, in: Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S.L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M.I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J., Maycock, T.K., Waterfield, T., Yelekçi, O., Yu, R., Zhou, B. (Eds.), Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, p. 215.

Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L., 2014. Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs. arXiv, Computer Vision cs.CV. doi: https://doi.org/10.48550/ARXIV.1412.7062.

Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L., 2016. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. IEEE Transactions on Pattern Analysis and Machine Intelligence 40, 834–848. doi: https://doi.org/10.1109/TPAMI.2017.2699184.

Chen, L.C., Papandreou, G., Schroff, F., Adam, H., 2017. Rethinking Atrous Convolution for Semantic Image Segmentation. arXiv, Computer Vision cs.CV. doi: https://doi.org/10.48550/ARXIV.1706.05587.

Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018a. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation, in: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (Eds.), Computer Vision – ECCV 2018, Springer International Publishing, Cham. pp. 833–851. doi: https://doi.org/10.1007/978-3-030-01234-2_49.

Chen, Z., Zhang, T., Ouyang, C., 2018b. End-to-End Airplane Detection Using Transfer Learning in Remote Sensing Images. Remote Sensing 10. doi: https://doi.org/10.3390/rs10010139.

Cheng, G., Han, J., Lu, X., 2017a. Remote Sensing Image Scene Classification: Benchmark and State of the Art. Proceedings of the IEEE 105, 1865–1883. doi: https://doi.org/10.1109/JPROC.2017.2675998.

Cheng, G., Han, J., Zhou, P., Guo, L., 2014. Multi-class geospatial object detection and geographic image classification based on collection of part detectors. ISPRS Journal of Photogrammetry and Remote Sensing 98, 119 – 132. doi: https://doi.org/10.1016/j.isprsjprs.2014.10.002.

Cheng, G., Wang, Y., Xu, S., Wang, H., Xiang, S., Pan, C., 2017b. Automatic Road Detection and Centerline Extraction via Cascaded End-to-End Convolutional Neural Network. IEEE Transactions on Geoscience and Remote Sensing 55, 3322–3337. doi: https://doi.org/10.1109/TGRS.2017.2669341.

Cheng, G., Zhou, P., Han, J., 2016. Learning Rotation-Invariant Convolutional Neural Networks for Object Detection in VHR Optical Remote Sensing Images. IEEE Transactions on Geoscience and Remote Sensing 54, 7405–7415. doi: https://doi.org/10.1109/TGRS.2016.2601622.

Chollet, F., 2017. Xception: Deep Learning with Depthwise Separable Convolutions, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1800–1807. doi: https://doi.org/10.1109/CVPR.2017.195.

Chollet, F., et al., 2015. Keras. Last visited: 07.01.2022. URL: https://keras.io.

Christiansen, N., Daewel, U., Djath, B., Schrum, C., 2022. Emergence of Large-Scale Hydrodynamic Structures Due to Atmospheric Offshore Wind Farm Wakes. Frontiers in Marine Science 9. doi: https://doi.org/10.3389/fmars.2022.818501.

Climate Watch, 2021. Climate Watch Historical GHG Emissions. Last visited: 07.01.2022. URL: https://www.climatewatchdata.org/ghg-emissions.

Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., Schiele, B., 2016. The cityscapes dataset for semantic urban scene understanding, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3213–3223. doi: https://doi.org/10.1109/CVPR.2016.350.

Cordts, M., Omran, M., Ramos, S., Scharwächter, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., Schiele, B., 2015. The Cityscapes Dataset, in: CVPR Workshop on the Future of Datasets in Vision.

Cotterill, C., Dove, D., Long, D., James, L., Duffy, C., Muller, S., Forsberg, C., Tjelta, T., 2012. Dogger Bank: A geo challenge, in: Offshore Site Investigation and Geotechnics - Integrated Technologies - Present and Future. URL: http://nora.nerc.ac.uk/id/eprint/530797/.

D'Agostino, R., Pearson, E.S., 1973. Tests for Departure from Normality. Empirical Results for the Distributions of $b_2$ and $\sqrt{b_1}$. Biometrika 60, 613–622. doi: https://doi.org/10.2307/2335012.

Demir, I., Koperski, K., Lindenbaum, D., Pang, G., Huang, J., Basu, S., Hughes, F., Tuia, D., Raskar, R., 2018. DeepGlobe 2018: A Challenge to Parse the Earth Through Satellite Images, in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops.

Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L., 2009. Imagenet: A large-scale hierarchical image database, in: 2009 IEEE Conference on Computer Vision and Pattern Recognition, IEEE. pp. 248–255. doi: https://doi.org/10.1109/CVPR.2009.5206848.

Deng, Z., Sun, H., Zhou, S., Zhao, J., Zou, H., 2017. Toward Fast and Accurate Vehicle Detection in Aerial Images Using Coupled Region-Based Convolutional Neural Networks. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 10, 3652–3664. doi: https://doi.org/10.1109/JSTARS.2017.2694890.

Dirscherl, M., Dietz, A.J., Kneisel, C., Kuenzer, C., 2021. A Novel Method for Automated Supraglacial Lake Mapping in Antarctica Using Sentinel-1 SAR Imagery and Deep Learning. Remote Sensing 13. doi: https://doi.org/10.3390/rs13020197.

Dong, T., Shen, Y., Zhang, J., Ye, Y., Fan, J., 2019. Progressive Cascaded Convolutional Neural Networks for Single Tree Detection with Google Earth Imagery. Remote Sensing 11. doi: https://doi.org/10.3390/rs11151786.

Drewitt, A.L., Langston, R.H.W., 2006. Assessing the impacts of wind farms on birds. Ibis 148, 29–42. doi: https://doi.org/10.1111/j.1474-919X.2006.00516.x.

EC, 2020. European Commission: An EU Strategy to harness the potential of offshore renewable energy for a climate neutral future. Last visited: 07.01.2022. URL: https://ec.europa.eu/energy/sites/ener/files/offshore_renewable_energy_strategy.pdf.

Energias de Portugal, 2022. WindFloat Atlantic - Floating offshore wind-power generating platform. Last visited: 15.04.2022. URL: https://www.edp.com/en/innovation/windfloat.

Equinor, 2022. Industrialising floating offshore wind. Last visited: 15.04.2022. URL: https://www.equinor.com/en/what-we-do/floating-wind.html.

ESA, 2022. News - Copernicus Sentinel-1B anomaly (5th update) - 25 February 2022. Last visited: 15.04.2022. URL: https://sentinels.copernicus.eu/web/sentinel/-/copernicus-sentinel-1b-anomaly-5th-update/.

Esteban, M.D., Diez, J.J., López, J.S., Negro, V., 2011. Why offshore wind energy? Renewable Energy 36, 444–450. doi: https://doi.org/10.1016/j.renene.2010.07.009.

European Commission, Joint Research Centre, Vazquez Hernandez, C., Telsnig, T., 2019. Wind energy : technology market report. Publications Office. doi: https://doi.org/10.2760/260914.

Everingham, M., Eslami, S.M., Gool, L., Williams, C.K., Winn, J., Zisserman, A., 2015. The Pascal Visual Object Classes Challenge: A Retrospective. Int. J. Comput. Vision 111, 98–136. doi: https://doi.org/10.1007/s11263-014-0733-5.

Everingham, M., Van Gool, L., Williams, C.K.I., Winn, J., Zisserman, A., 2010. The Pascal Visual Object Classes (VOC) Challenge. International Journal of Computer Vision 88, 303–338. doi: https://doi.org/10.1007/s11263-009-0275-4.

Fan, W., Zhou, F., Bai, X., Tao, M., Tian, T., 2019. Ship Detection Using Deep Convolutional Neural Networks for PolSAR Images. Remote Sensing 11. doi: https://doi.org/10.3390/rs11232862.

Felleman, D.J., Van Essen, D.C., 1991. Distributed hierarchical processing in the primate cerebral cortex . Cerebral Cortex 1, 1–47. doi: https://doi.org/10.1093/cercor/1.1.1-a.

Flanders Marine Institute, 2020. Union of the ESRI Country shapefile and the Exclusive Economic Zones (version 3). Last visited: 10.02.2022. URL: https://www.marineregions.org/, doi: https://doi.org/10.14284/403.

Fox, A., Desholm, M., Kahlert, J., Christensen, T.K., Krag Petersen, I., 2006. Information needs to support environmental impact assessment of the effects of European marine offshore wind farms on birds. Ibis 148, 129–144. doi: https://doi.org/10.1111/j.1474-919X.2006.00510.x.

Frandsen, S., Barthelmie, R., Pryor, S., Rathmann, O., Larsen, S., Højstrup, J., Thøgersen, M., 2006. Analytical modelling of wind speed deficit in large offshore wind farms. Wind Energy 9, 39–53. doi: https://doi.org/10.1002/we.189.

Fromm, M., Schubert, M., Castilla, G., Linke, J., McDermid, G., 2019. Automated Detection of Conifer Seedlings in Drone Imagery Using Convolutional Neural Networks. Remote Sensing 11. doi: https://doi.org/10.3390/rs11212585.

Fu, Y., Ye, Z., Deng, J., Zheng, X., Huang, Y., Yang, W., Wang, Y., Wang, K., 2019. Finer Resolution Mapping of Marine Aquaculture Areas Using WorldView-2 Imagery and a Hierarchical Cascade Convolutional Neural Network. Remote Sensing 11. doi: https://doi.org/10.3390/rs11141678.

Fukushima, K., Miyake, S., 1982. Neocognitron: A new algorithm for pattern recognition tolerant of deformations and shifts in position. Pattern Recognit. 15, 455–469. doi: https://doi.org/10.1016/0031-3203(82)90024-3.

Gao, L., He, Y., Sun, X., Jia, X., Zhang, B., 2019. Incorporating Negative Sample Training for Ship Detection Based on Deep Learning. Sensors 19. doi: https://doi.org/10.3390/s19030684.

Georgakis, G., Mousavian, A., Berg, A.C., Kosecka, J., 2017. Synthesizing Training Data for Object Detection in Indoor Scenes. arXiv, Computer Vision and Pattern Recognition cs.CV. doi: https://doi.org/10.48550/ARXIV.1702.07836.

Ghiasi, G., Lin, T., Le, Q.V., 2019. NAS-FPN: Learning Scalable Feature Pyramid Architecture for Object Detection, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 7029–7038. doi: https://doi.org/10.1109/CVPR.2019.00720.

Girshick, R.B., 2015. Fast R-CNN. 2015 IEEE International Conference on Computer Vision (ICCV), 1440–1448. doi: https://doi.org/10.1109/ICCV.2015.169.

Girshick, R.B., Donahue, J., Darrell, T., Malik, J., 2013. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. 2014 IEEE Conference on Computer Vision and Pattern Recognition, 580–587. doi: https://doi.org/10.1109/CVPR.2014.81.

Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press, Cambridge, Massachusetts. URL: http://www.deeplearningbook.org.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sensing of Environment 202, 18–27. doi: https://doi.org/10.1016/j.rse.2017.06.031. big Remotely Sensed Data: tools, applications and experiences.

Gruber, T.R., 1995. Toward principles for the design of ontologies used for knowledge sharing? International Journal of Human-Computer Studies 43, 907–928. doi: https://doi.org/10.1006/ijhc.1995.1081.

Gu, L., Chen, J., Yin, J., Sullivan, S.C., Wang, H.M., Guo, S., Zhang, L., Kim, J.S., 2020. Projected increases in magnitude and socioeconomic exposure of global droughts in 1.5 and 2 °C warmer climates. Hydrology and Earth System Sciences 24, 451–472. doi: https://doi.org/10.5194/hess-24-451-2020.

Guerreiro, S.B., Fowler, H.J., Barbero, R., Westra, S., Lenderink, G., Blenkinsop, S., Lewis, E., Li, X.F., 2018. Detection of continental-scale intensification of hourly rainfall extremes. Nature Climate Change 8, 803–807. doi: https://doi.org/10.1038/s41558-018-0245-3.

Gușatu, L., Menegon, S., Depellegrin, D., Zuidema, C., Faaij, A., Yamu, C., 2021. Spatial and temporal analysis of cumulative environmental effects of offshore wind farms in the North Sea basin. Scientific Reports 11. doi: https://doi.org/10.1038/s41598-021-89537-1.

Gușatu, L.F., Yamu, C., Zuidema, C., Faaij, A., 2020. A spatial analysis of the potentials for offshore wind farm locations in the North Sea region: Challenges and opportunities. ISPRS International Journal of Geo-Information 9, 96. doi: https://doi.org/10.3390/ijgi9020096.

Hamdi, Z.M., Brandmeier, M., Straub, C., 2019. Forest Damage Assessment Using Deep Learning on High Resolution Remote Sensing Data. Remote Sensing 11. doi: https://doi.org/10.3390/rs11171976.

Han, S., Fafard, A., Kerekes, J., Gartley, M., Ientilucci, E., Savakis, A., Law, C., Parhan, J., Turek, M., Fieldhouse, K., Rovito, T., 2017. Efficient generation of image chips for training deep learning algorithms, in: Sadjadi, F.A., Mahalanobis, A. (Eds.), Automatic Target Recognition XXVII, International Society for Optics and Photonics. SPIE. pp. 15 – 23. doi: https://doi.org/10.1117/12.2261702.

Harris, C.R., Millman, K.J., van der Walt, S.J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N.J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M.H., Brett, M., Haldane, A., del Río, J.F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., Oliphant, T.E., 2020. Array programming with NumPy. Nature 585, 357–362. URL: 10.1038/s41586-020-2649-2, doi: https://doi.org/10.1038/s41586-020-2649-2.

He, B., Li, X., Huang, B., Gu, E., Guo, W., Wu, L., 2021. UnityShip: A Large-Scale Synthetic Dataset for Ship Recognition in Aerial Images. Remote Sensing 13. doi: https://doi.org/10.3390/rs13244999.

He, H., Yang, D., Wang, S., Wang, S., Li, Y., 2019. Road Extraction by Using Atrous Spatial Pyramid Pooling Integrated Encoder-Decoder Network and Structural Similarity Loss. Remote Sensing 11. doi: https://doi.org/10.3390/rs11091015.

He, K., Gkioxari, G., Dollár, P., Girshick, R.B., 2017. Mask R-CNN. 2017 IEEE International Conference on Computer Vision (ICCV), 2980–2988. doi: https://doi.org/10.1109/ICCV.2017.322.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep Residual Learning for Image Recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778. doi: https://doi.org/10.1109/CVPR.2016.90.

He, Y., Sun, X., Gao, L., Zhang, B., 2018. Ship Detection Without Sea-Land Segmentation for Large-Scale High-Resolution Optical Satellite Images, in: IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium, pp. 717–720.

Henderson, A.R., Morgan, C., Smith, B., Sørensen, H.C., Barthelmie, R.J., Boesmans, B., 2003. Offshore Wind Energy in Europe— A Review of the State-of-the-Art. Wind Energy 6, 35–52. doi: https://doi.org/10.1002/we.82.

Henderson, A.R., Witcher, D., 2010. Floating Offshore Wind Energy — A Review of the Current Status and an Assessment of the Prospects. Wind Engineering 34, 1–16. doi: https://doi.org/10.1260/0309-524X.34.1.1.

Henry, C.J., Storie, C.D., Palaniappan, M., Alhassan, V., Swamy, M., Aleshinloye, D., Curtis, A., Kim, D., 2019. Automated LULC map production using deep neural networks. International Journal of Remote Sensing 40, 4416–4440. doi: https://doi.org/10.1080/01431161.2018.1563840.

Higgins, P., Foley, A., 2014. The evolution of offshore wind power in the United Kingdom. Renewable and Sustainable Energy Reviews 37, 599–612. doi: https://doi.org/10.1016/j.rser.2014.05.058.

Hinterstoisser, S., Pauly, O., Heibel, H., Martina, M., Bokeloh, M., 2019. An Annotation Saved is an Annotation Earned: Using Fully Synthetic Training for Object Detection, in: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops. doi: https://doi.org/10.1109/ICCVW.2019.00340.

Hoeser, T., Bachofer, F., Kuenzer, C., 2020. Object Detection and Image Segmentation with Deep Learning on Earth Observation Data: A Review-Part II: Applications. Remote Sensing 12. doi: https://doi.org/10.3390/rs12183053.

Hoeser, T., Feuerstein, S., Kuenzer, C., 2022. Deepowt: a global offshore wind turbine data set derived with deep learning from sentinel-1 data. Earth System Science Data 14, 4251–4270. doi: https://doi.org/10.5194/essd-14-4251-2022.

Hoeser, T., Kuenzer, C., 2020. Object Detection and Image Segmentation with Deep Learning on Earth Observation Data: A Review-Part I: Evolution and Recent Trends. Remote Sensing 12. doi: https://doi.org/10.3390/rs12101667.

Hoeser, T., Kuenzer, C., 2022a. Global dynamics of the offshore wind energy sector monitored with sentinel-1: Turbine count, installed capacity and site specifications. International Journal of Applied Earth Observation and Geoinformation 112, 102957. doi: https://doi.org/10.1016/j.jag.2022.102957.

Hoeser, T., Kuenzer, C., 2022b. Synteo: Synthetic dataset generation for earth observation and deep learning – demonstrated for offshore wind farm detection. ISPRS Journal of Photogrammetry and Remote Sensing 189, 163–184. doi: https://doi.org/10.1016/j.isprsjprs.2022.04.029.

Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., Adam, H., 2017. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. arXiv, Computer Vision, cs.CV. doi: https://doi.org/10.48550/ARXIV.1704.04861.

Hu, Y., Zhang, Q., Zhang, Y., Yan, H., 2018. A Deep Convolution Neural Network Method for Land Cover Mapping: A Case Study of Qinhuangdao, China. Remote Sensing 10. doi: https://doi.org/10.3390/rs10122053.

Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q., 2017. Densely Connected Convolutional Networks, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2261–2269. doi: https://doi.org/10.1109/CVPR.2017.243.

Huang, L., Liu, B., Li, B., Guo, W., Yu, W., Zhang, Z., Yu, W., 2018. OpenSARShip: A Dataset Dedicated to Sentinel-1 Ship Interpretation. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 11, 195–208. doi: https://doi.org/10.1109/JSTARS.2017.2755672.

Hubel, D.H., Wiesel, T.N., 1962. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. The Journal of Physiology 160, 106–154. doi: https://doi.org/10.1113/jphysiol.1962.sp006837.

IEA, 2021a. Global Energy Review 2021. Last visited: 07.01.2022. URL: https://www.iea.org/reports/global-energy-review-2021.

IEA, 2021b. Net Zero by 2050. Last visited: 07.01.2022. URL: https://www.iea.org/reports/net-zero-by-2050.

IEEE GRSS, 2015. GRSS Data Fusion Contest. Last visited: 07.01.2022. URL: http://www.grss-ieee.org/community/technical-committees/data-fusion./2015-ieee-grss-data-fusion-contest/.

Ienco, D., Interdonato, R., Gaetano, R., Minh, D.H.T., 2019. Combining Sentinel-1 and Sentinel-2 Satellite Image Time Series for land cover mapping via a multi-source deep learning architecture. ISPRS Journal of Photogrammetry and Remote Sensing 158, 11–22. doi: https://doi.org/10.1016/j.isprsjprs.2019.09.016.

Ioffe, S., Szegedy, C., 2015. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift, in: Bach, F., Blei, D. (Eds.), Proceedings of the 32nd International Conference on Machine Learning, PMLR, Lille, France. pp. 448–456. doi: https://doi.org/10.5555/3045118.3045167.

Isikdogan, F., Bovik, A., Passalacqua, P., 2018. Learning a River Network Extractor Using an Adaptive Loss Function. IEEE Geoscience and Remote Sensing Letters 15, 813–817. doi: https://doi.org/10.1109/LGRS.2018.2811754.

ISPRS, 2016. 2D Semantic Labeling Challenge. Last visited: 07.01.2022. URL: http://www2.isprs.org/commissions/comm3/wg4/semantic-labeling.html.

Ji, S., Wei, S., Lu, M., 2019. Fully Convolutional Networks for Multisource Building Extraction From an Open Aerial and Satellite Imagery Data Set. IEEE Transactions on Geoscience and Remote Sensing 57, 574–586. doi: https://doi.org/10.1109/TGRS.2018.2858817.

Ji, S., Wei, S., Lu, M., 2019. A scale robust convolutional neural network for automatic building extraction from aerial and satellite imagery. International Journal of Remote Sensing 40, 3308–3322. doi: https://doi.org/10.1080/01431161.2018.1528024.

Ji, S., Zhang, C., Xu, A., Shi, Y., Duan, Y., 2018. 3D Convolutional Neural Networks for Crop Classification with Multi-Temporal Remote Sensing Images. Remote Sensing 10. doi: https://doi.org/10.3390/rs10010075.

Josifovski, J., Kerzel, M., Pregizer, C., Posniak, L., Wermter, S., 2018. Object Detection and Pose Estimation Based on Convolutional Neural Networks Trained with Synthetic Data, in: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 6269–6276. doi: https://doi.org/10.1109/IROS.2018.8594379.

Khan, S., Phan, B., Salay, R., Czarnecki, K., 2019. ProcSy: Procedural Synthetic Dataset Generation Towards Influence Factor Studies Of Semantic Segmentation Networks, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, pp. 88–96.

Koga, Y., Miyazaki, H., Shibasaki, R., 2018. A CNN-Based Method of Vehicle Detection from Aerial Images Using Hard Example Mining. Remote Sensing 10. doi: https://doi.org/10.3390/rs10010124.

Kong, F., Huang, B., Bradbury, K., Malof, J.M., 2020. The Synthinel-1 dataset: a collection of high resolution synthetic overhead imagery for building segmentation, in: 2020 IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 1803–1812. doi: https://doi.org/10.1109/WACV45572.2020.9093339.

Krige, D.G., Magri, E.J., 1982. Studies of the effects of outliers and data transformation on variogram estimates for a base metal and a gold ore body. Journal of the International Association for Mathematical Geology 14, 557–564. doi: https://doi.org/10.1007/BF01033879.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. ImageNet Classification with Deep Convolutional Neural Networks, in: Pereira, F., Burges, C.J.C., Bottou, L., Weinberger, K.Q. (Eds.), Advances in Neural Information Processing Systems, Curran Associates, Inc.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2017. ImageNet Classification with Deep Convolutional Neural Networks. Communications of the ACM 60, 84–90. doi: https://doi.org/10.1145/3065386.

Kumar, P., Foufoula-Georgiou, E., 1997. Wavelet analysis for geophysical applications. Reviews of Geophysics 35, 385–412. doi: https://doi.org/10.1029/97RG00427.

Ladan, M., Hänninen, M., 2012. Data sources for quantitative marine traffic accident modeling. Technical Report. Aalto University, School of Engineering.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep Learning. Nature 521, 436–444. doi: https://doi.org/10.1038/nature14539.

Lecun, Y., Bottou, L., Bengio, Y., Haffner, P., 1998. Gradient-based learning applied to document recognition. Proceedings of the IEEE 86, 2278–2324. doi: https://doi.org/10.1109/5.726791.

Lee, J., Zhao, F., 2020. GWEC - Global Wind Report 2019. Global Wind Energy Council. URL: https://gwec.net/global-wind-report-2019/.

Lee, J., Zhao, F., 2021. GWEC - Global Wind Report 2021. Global Wind Energy Council. URL: https://gwec.net/global-wind-report-2021/.

Li, B., Liu, B., Huang, L., Guo, W., Zhang, Z., Yu, W., 2017a. OpenSARShip 2.0: A large-volume dataset for deeper interpretation of ship targets in Sentinel-1 imagery, in: 2017 SAR in Big Data Era: Models, Methods and Applications (BIGSARDATA), pp. 1–5. doi: https://doi.org/10.1109/BIGSARDATA.2017.8124929.

Li, J., Qu, C., Shao, J., 2017b. Ship detection in SAR images based on an improved faster R-CNN, in: 2017 SAR in Big Data Era: Models, Methods and Applications (BIGSARDATA), pp. 1–6. doi: https://doi.org/10.1109/BIGSARDATA.2017.8124934.

Li, K., Cheng, G., Bu, S., You, X., 2018a. Rotation-Insensitive and Context-Augmented Object Detection in Remote Sensing Images. IEEE Transactions on Geoscience and Remote Sensing 56, 2337–2348. doi: https://doi.org/10.1109/TGRS.2017.2778300.

Li, L., 2019. Deep Residual Autoencoder with Multiscaling for Semantic Segmentation of Land-Use Images. Remote Sensing 11. doi: https://doi.org/10.3390/rs11182142.

Li, M., Guo, W., Zhang, Z., Yu, W., Zhang, T., 2018b. Rotated Region Based Fully Convolutional Network for Ship Detection, in: IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium, pp. 673–676. doi: https://doi.org/10.1109/IGARSS.2018.8519094.

Li, Q., Mou, L., Xu, Q., Zhang, Y., Zhu, X.X., 2019. R3-Net: A Deep Network for Multioriented Vehicle Detection in Aerial Images and Videos. IEEE Transactions on Geoscience and Remote Sensing 57, 5028–5042. doi: https://doi.org/10.1109/tgrs.2019.2895362.

Li, Z., Chen, G., Zhang, T., 2019. Temporal Attention Networks for Multitemporal Multisensor Crop Classification. IEEE Access 7, 134677–134690. doi: https://doi.org/10.1109/ACCESS.2019.2939152.

Lin, T.Y., Dollár, P., Girshick, R.B., He, K., Hariharan, B., Belongie, S.J., 2016. Feature Pyramid Networks for Object Detection. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 936–944. doi: https://doi.org/10.1109/CVPR.2017.106.

Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L., 2014. Microsoft COCO: Common Objects in Context, in: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (Eds.), Computer Vision – ECCV 2014, Springer International Publishing, Cham. pp. 740–755.

Liu, C., Chen, L., Schroff, F., Adam, H., Hua, W., Yuille, A.L., Fei-Fei, L., 2019. Auto-DeepLab: Hierarchical Neural Architecture Search for Semantic Image Segmentation, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 82–92. doi: https://doi.org/10.1109/CVPR.2019.00017.

Liu, J., Kong, X., Xia, F., Bai, X., Wang, L., Qing, Q., Lee, I., 2018. Artificial Intelligence in the 21st Century. IEEE Access 6, 34403–34421. doi: https://doi.org/10.1109/ACCESS.2018.2819688.

Liu, K., Mattyus, G., 2015. Fast Multiclass Vehicle Detection on Aerial Images. IEEE Geoscience and Remote Sensing Letters 12, 1938–1942. doi: https://doi.org/10.1109/LGRS.2015.2439517.

Liu, P., Liu, X., Liu, M., Shi, Q., Yang, J., Xu, X., Zhang, Y., 2019a. Building Footprint Extraction from High-Resolution Images via Spatial Residual Inception Convolutional Neural Network. Remote Sensing 11. doi: https://doi.org/10.3390/rs11070830.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.Y., Berg, A.C., 2016a. Ssd: Single shot multibox detector, in: Leibe, B., Matas, J., Sebe, N., Welling, M. (Eds.), Computer Vision – ECCV 2016, Springer International Publishing, Cham. pp. 21–37. doi: https://doi.org/10.1007/978-3-319-46448-0_2.

Liu, W., Rabinovich, A., Berg, A.C., 2015. ParseNet: Looking Wider to See Better. arXiv, Coputer Vision cs.CV. doi: https://doi.org/10.48550/ARXIV.1506.04579.

Liu, Y., Sun, C., Yang, Y., Zhou, M., Zhan, W., Cheng, W., 2016b. Automatic extraction of offshore platforms using time-series Landsat-8 Operational Land Imager data. Remote Sensing of Environment 175, 73–91. doi: https://doi.org/10.1016/j.rse.2015.12.047.

Liu, Y., Wang, Y., Wang, S., Liang, T., Zhao, Q., Tang, Z., Ling, H., 2019b. CBNet: A Novel Composite Backbone Network Architecture for Object Detection. arXiv, Computer Vision cs.CV. doi: https://doi.org/10.48550/ARXIV.1909.03625.

Liu, Z., Yuan, L., Weng, L., Yang., Y., 2017. A High Resolution Optical Satellite Image Dataset for Ship Recognition and Some New Baselines, in: Proceedings of the 6th International Conference on Pattern Recognition Applications and Methods - Volume 1: ICPRAM, INSTICC. SciTePress. pp. 324–331. doi: https://doi.org/10.5220/0006120603240331.

Lloret, J., Turiel, A., Solé, J., Berdalet, E., Sabatés, A., Olivares, A., Gili, J.M., Vila-Subirós, J., Sardá, R., 2022. Unravelling the ecological impacts of large-scale offshore wind farms in the Mediterranean Sea. Science of The Total Environment 824, 153803. doi: https://doi.org/10.1016/j.scitotenv.2022.153803.

Long, Y., Gong, Y., Xiao, Z., Liu, Q., 2017. Accurate Object Localization in Remote Sensing Images Based on Convolutional Neural Networks. IEEE Transactions on Geoscience and Remote Sensing 55, 2486–2498. doi: https://doi.org/10.1109/TGRS.2016.2645610.

Long, Y., Xia, G.S., Li, S., Yang, W., Yang, M.Y., Zhu, X.X., Zhang, L., Li, D., 2021. On Creating Benchmark Dataset for Aerial Image Interpretation: Reviews, Guidances, and Million-AID. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 14, 4205–4230. doi: https://doi.org/10.1109/JSTARS.2021.3070368.

Loshchilov, I., Hutter, F., 2017. SGDR: Stochastic Gradient Descent with Warm Restarts. arXiv, Machine Learning cs.LG. doi: https://doi.org/10.48550/ARXIV.1608.03983.

Ma, F., Gao, F., Sun, J., Zhou, H., Hussain, A., 2019a. Attention Graph Convolution Network for Image Segmentation in Big SAR Imagery Data. Remote Sensing 11. doi: https://doi.org/10.3390/rs11212586.

Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., Johnson, B.A., 2019b. Deep learning in remote sensing applications: A meta-analysis and review. ISPRS Journal of Photogrammetry and Remote Sensing 152, 166 – 177. doi: https://doi.org/10.1016/j.isprsjprs.2019.04.015.

Madin, J., Bowers, S., Schildhauer, M., Krivov, S., Pennington, D., Villa, F., 2007. An ontology for describing and synthesizing ecological observation data. Ecological Informatics 2, 279–296. doi: https://doi.org/10.1016/j.ecoinf.2007.05.004. meta-information systems and ontologies. A Special Feature from the 5th International Conference on Ecological Informatics ISEI5, Santa Barbara, CA, Dec. 4–7, 2006.

Madsen, P.T., Wahlberg, M., Tougaard, J., Lucke, K., P, T., 2006. Wind turbine underwater noise and marine mammals: Implications of current knowledge and data needs. Marine ecology progress series 309, 279–295. doi: https://doi.org/10.3354/meps309279.

Maggiori, E., Tarabalka, Y., Charpiat, G., Alliez, P., 2017. Can Semantic Labeling Methods Generalize to Any City? The Inria Aerial Image Labeling Benchmark, in: IEEE International Geoscience and Remote Sensing Symposium (IGARSS), IEEE. doi: https://doi.org/10.1109/IGARSS.2017.8127684.

Martinez-Gonzalez, P., Oprea, S., Garcia-Garcia, A., Jover-Alvarez, A., Orts-Escolano, S., Garcia-Rodriguez, J., 2020. UnrealROX: an extremely photorealistic virtual reality environment for robotics simulations and synthetic data generation. Virtual Reality 24, 271–288. doi: https://doi.org/10.1007/s10055-019-00399-5.

Massey, F.J., 1951. The Kolmogorov-Smirnov Test for Goodness of Fit. Journal of the American Statistical Association 46, 68–78. doi: https://doi.org/10.1080/01621459.1951.10500769.

Meric, S., Fayard, F., Pottier, E., 2009. Radargrammetric SAR Image Processing, in: Ho, P.G.P. (Ed.), Geoscience and Remote Sensing. IntechOpen, Rijeka. chapter 20, pp. 421–454. doi: https://doi.org/10.5772/8300.

Meyer, F., 2019. Spaceborne Synthetic Aperture Radar: Principles, Data Access, and Basic Processing Techniques, in: Flores-Anderson, A.I., Herndon, K.E., Thapa, R.B., Cherrington, E. (Eds.), The SAR Handbook. Comprehensive Methodologies for Forest Monitoring and Biomass Estimation. SERVIR Global, Huntsville. chapter 2, pp. 21–44.

Miao, Z., Fu, K., Sun, H., Sun, X., Yan, M., 2018. Automatic Water-Body Segmentation From High-Resolution Satellite Images via Deep Networks. IEEE Geoscience and Remote Sensing Letters 15, 602–606. doi: https://doi.org/10.1109/LGRS.2018.2794545.

Mnih, V., 2013. Machine Learning for Aerial Image Labeling. Ph.D. thesis. University of Toronto.

Mohajerani, Y., Wood, M., Velicogna, I., Rignot, E., 2019. Detection of Glacier Calving Margins with Convolutional Neural Networks: A Case Study. Remote Sensing 11. doi: https://doi.org/10.3390/rs11010074.

Moran, P.A.P., 1950. Notes on Continuous Stochastic Phenomena. Biometrika 37, 17–23. doi: https://doi.org/10.2307/2332142.

Mou, L., Zhu, X.X., 2018. Vehicle Instance Segmentation From Aerial Image and Video Using a Multitask Learning Residual Fully Convolutional Network. IEEE Transactions on Geoscience and Remote Sensing 56, 6699–6711. doi: https://doi.org/10.1109/TGRS.2018.2841808.

NASA/JPL, 2020. Airborne Synthetic Aperture Radar (AIRSAR). Last visited: 07.01.2022. URL: https://airsar.jpl.nasa.gov/index_detail.html.

Natural Earth, 2022. 1:10m Physical Vectors. Last visited: 07.01.2022. URL: https://www.naturalearthdata.com/downloads/10m-physical-vectors/.

Nielsen, M.A., 2015. Neural Networks and Deep Learning. Determination Press. URL: http://neuralnetworksanddeeplearning.com.

Nikolenko, S.I., 2021. Synthetic Data for Deep Learning. Springer International Publishing, Cham. doi: https://doi.org/10.1007/978-3-030-75178-4_1.

Noh, H., Hong, S., Han, B., 2015. Learning Deconvolution Network for Semantic Segmentation, in: 2015 IEEE International Conference on Computer Vision (ICCV), pp. 1520–1528. doi: https://doi.org/10.1109/ICCV.2015.178.

Nowruzi, F.E., Kapoor, P., Kolhatkar, D., Hassanat, F.A., Laganière, R., Rebut, J., 2019. How much real data do we actually need: Analyzing object detection performance using synthetic and real data. arXiv, Computer Vision and Pattern Recognition cs.CV. doi: https://doi.org/10.48550/ARXIV.1907.07061.

Opitz, J., Burst, S., 2021. Macro F1 and Macro F1. arXiv, Machine Learning cs.LG. doi: https://doi.org/10.48550/ARXIV.1911.03347.

Oppenheimer, M., Glavovic, B., Hinkel, J., van de Wal, R., Magnan, A., Abd-Elgawad, A., Cai, R., Cifuentes-Jara, M., DeConto, R., Ghosh, T., Hay, J., Isla, F., Marzeion, B., Meyssignac, B., Sebesvari, Z., 2019. Sea Level Rise and Implications for Low-Lying Islands, Coasts and Communities, in: Pörtner, H.O., Roberts, D., Masson-Delmotte, V., Zhai, P., Tignor, M., Poloczanska, E., Mintenbeck, K., Alegría, A., Nicolai, M., Okem, A., Petzold, J., Rama, B., N.M., W. (Eds.), Climate Change and Land: an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems. Cambridge University Press, Cambridge, UK and New York, NY, USA, pp. 321–445. doi: https://doi.org/10.1017/9781009157964.006.

Padilla, R., Passos, W.L., Dias, T.L.B., Netto, S.L., da Silva, E.A.B., 2021. A Comparative Analysis of Object Detection Metrics with a Companion Open-Source Toolkit. Electronics 10. doi: https://doi.org/10.3390/electronics10030279.

Papandreou, G., Chen, L.C., Murphy, K., Yuille, A.L., 2015. Weakly- and Semi-Supervised Learning of a DCNN for Semantic Image Segmentation. arXiv, Computer Vision cs.CV. doi: https://doi.org/10.48550/ARXIV.1502.0273.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S., 2019. PyTorch: An Imperative Style, High-Performance Deep Learning Library, in: Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., Garnett, R. (Eds.), Advances in Neural Information Processing Systems 32. Curran Associates, Inc., pp. 8024–8035.

Perlin, K., 1985. An Image Synthesizer. SIGGRAPH Comput. Graph. 19, 287–296. doi: https://doi.org/10.1145/325165.325247.

Perlin, K., 2001. Chapter 2, Noise Hardware, in: SIGGRAPH 2001, Association for Computing Machinery, New York, NY, USA. pp. 2–1 – 2–24. URL: https://www.csee.umbc.edu/~olano/s2002c36/ch02.pdf.

Polyak, B., 1964. Some methods of speeding up the convergence of iteration methods. USSR Computational Mathematics and Mathematical Physics 4, 1–17. doi: https://doi.org/10.1016/0041-5553(64)90137-5.

Rajpura, P.S., Hegde, R.S., Bojinov, H., 2017. Object detection using deep cnns trained on synthetic images. arXiv, Computer Vision and Pattern Recognition cs.CV. doi: https://doi.org/10.48550/ARXIV.1706.06782.

Ramírez, L., Fraile, D., Brindley, G., O'Sullivan, R., Miró, L., de Velde, L.V., 2021. Offshore Wind in Europe - Key trends and statistics 2020. Wind Europe. URL: https://windeurope.org/intelligence-platform/product/offshore-wind-in-europe-key-trends-and-statistics-2020/.

Ranzato, M., Huang, F.J., Boureau, Y.L., LeCun, Y., 2007. Unsupervised Learning of Invariant Feature Hierarchies with Applications to Object Recognition, in: 2007 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8. doi: https://doi.org/10.1109/CVPR.2007.383157.

Razakarivony, S., Jurie, F., 2016. Vehicle detection in aerial imagery: A small target detection benchmark. Journal of Visual Communication and Image Representation 34, 187–203. doi: https://doi.org/10.1016/j.jvcir.2015.11.002.

Redmon, J., Divvala, S.K., Girshick, R.B., Farhadi, A., 2015. You Only Look Once: Unified, Real-Time Object Detection. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 779–788. doi: https://doi.org/10.1109/CVPR.2016.91.

Redmon, J., Farhadi, A., 2016. YOLO9000: Better, Faster, Stronger. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 6517–6525. doi: https://doi.org/10.1109/CVPR.2017.690.

Redmon, J., Farhadi, A., 2018. YOLOv3: An Incremental Improvement. arXiv, Computer Vision cs.CV. doi: https://doi.org/10.48550/ARXIV.1804.027.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., Prabhat, 2019. Deep learning and process understanding for data-driven Earth system science. Nature 566, 195–204. doi: https://doi.org/10.1038/s41586-019-0912-1.

Ren, S., He, K., Girshick, R.B., Sun, J., 2015. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence 39, 1137–1149. doi: https://doi.org/10.1109/TPAMI.2016.2577031.

Rich, E., Knight, K., 1991. Artificial Intelligence. 2 ed., McGraw-Hill, New York.

Richter, S.R., Hayder, Z., Koltun, V., 2017. Playing for benchmarks, in: IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29, 2017, pp. 2232–2241. doi: https://doi.org/10.1109/ICCV.2017.243.

Richter, S.R., Vineet, V., Roth, S., Koltun, V., 2016. Playing for Data: Ground Truth from Computer Games, in: Leibe, B., Matas, J., Sebe, N., Welling, M. (Eds.), European Conference on Computer Vision (ECCV), Springer International Publishing. pp. 102–118. doi: https://doi.org/10.1007/978-3-319-46475-6_7.

Ritchie, H., Roser, M., 2020a. Our World in Data. CO2 and Greenhouse Gas Emissions. Last visited: 07.01.2022. URL: https://ourworldindata.org/co2-and-other-greenhouse-gas-emissions.

Ritchie, H., Roser, M., 2020b. Our World in Data. Energy. Based on BP Statistical Review of WOrld Energy & Ember Last visited: 07.01.2022. URL: https://ourworldindata.org/energy.

Robbins, H.E., Monro, S.A., 1951. A Stochastic Approximation Method. Annals of Mathematical Statistics 22, 400–407. doi: https://doi.org/10.1214/aoms/1177729586.

Rodrigues, S., Restrepo, C., Kontos, E., Teixeira Pinto, R., Bauer, P., 2015. Trends of offshore wind projects. Renewable and Sustainable Energy Reviews 49, 1114–1135. doi: https://doi.org/10.1016/j.rser.2015.04.092.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation, in: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, Springer International Publishing, Cham. pp. 234–241. doi: https://doi.org/10.1007/978-3-319-24574-4_28.

Ros, G., Sellart, L., Materzynska, J., Vazquez, D., Lopez, A.M., 2016. The SYNTHIA Dataset: A Large Collection of Synthetic Images for Semantic Segmentation of Urban Scenes, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3234–3243. doi: https://doi.org/10.1109/CVPR.2016.352.

Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. Learning representations by back-propagating errors. Nature 323, 533–536. doi: https://doi.org/10.1038/323533a0.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C., Fei-Fei, L., 2015. ImageNet Large Scale Visual Recognition Challenge. International Journal of Computer Vision (IJCV) 115, 211–252. doi: https://doi.org/10.1007/s11263-015-0816-y.

Rußwurm, M., Körner, M., 2018. Convolutional LSTMs for Cloud-Robust Segmentation of Remote Sensing Imagery. arXiv, Computer Vision cs.CV. doi: https://doi.org/10.48550/ARXIV.1811.02471.

Safonova, A., Tabik, S., Alcaraz-Segura, D., Rubtsov, A., Maglinets, Y., Herrera, F., 2019. Detection of Fir Trees (Abies sibirica) Damaged by the Bark Beetle in Unmanned Aerial Vehicle Images with Deep Learning. Remote Sensing 11. doi: https://doi.org/10.3390/rs11060643.

Santos, A.A.d., Marcato Junior, J., Araújo, M.S., Di Martini, D.R., Tetila, E.C., Siqueira, H.L., Aoki, C., Eltner, A., Matsubara, E.T., Pistori, H., Feitosa, R.Q., Liesenberg, V., Gonçalves, W.N., 2019. Assessment of CNN-Based Methods for Individual Tree Detection on Images Captured by RGB Cameras Attached to UAVs. Sensors 19. doi: https://doi.org/10.3390/s19163595.

Scherer, D., Müller, A., Behnke, S., 2010. Evaluation of Pooling Operations in Convolutional Architectures for Object Recognition, in: Diamantaras, K., Duch, W., Iliadis, L.S. (Eds.), Artificial Neural Networks – ICANN 2010, Springer Berlin Heidelberg, Berlin, Heidelberg. pp. 92–101. doi: https://doi.org/10.1007/978-3-642-15825-4_10.

Shah, S., Dey, D., Lovett, C., Kapoor, A., 2018. AirSim: High-Fidelity Visual and Physical Simulation for Autonomous Vehicles, in: Hutter, M., Siegwart, R. (Eds.), Field and Service Robotics, Springer International Publishing, Cham. pp. 621–635. doi: https://doi.org/10.1007/978-3-319-67361-5_40.

Shelhamer, E., Long, J., Darrell, T., 2014. Fully Convolutional Networks for Semantic Segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence 39, 640–651. doi: https://doi.org/10.1109/CVPR.2015.7298965.

Shermeyer, J., Hossler, T., Van Etten, A., Hogan, D., Lewis, R., Kim, D., 2021. RarePlanes: Synthetic Data Takes Flight, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), pp. 207–217. doi: https://doi.org/10.1109/WACV48630.2021.00025.

Shorten, C., Khoshgoftaar, T.M., 2019. A survey on Image Data Augmentation for Deep Learning. Journal of Big Data 6, 60. doi: https://doi.org/10.1186/s40537-019-0197-0.

Shrestha, A., Mahmood, A., 2019. Review of Deep Learning Algorithms and Architectures. IEEE Access , 53040–53065. doi: https://doi.org/10.1109/ACCESS.2019.2912200.

Shrestha, S., Vanneschi, L., 2018. Improved Fully Convolutional Network with Conditional Random Fields for Building Extraction. Remote Sensing 10. doi: https://doi.org/10.3390/rs10071135.

Simonyan, K., Zisserman, A., 2014. Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv, Computer Vision cs.CV. doi: https://doi.org/10.48550/ARXIV.1409.1556.

Slavik, K., Lemmen, C., Zhang, W., Kerimoglu, O., Klingbeil, K., Wirtz, K.W., 2019. The large-scale impact of offshore wind farm structures on pelagic primary productivity in the southern North Sea. Hydrobiologia 845, 35–53. doi: https://doi.org/10.1007/s10750-018-3653-5.

Smith, L.N., 2017. Cyclical Learning Rates for Training Neural Networks, in: 2017 IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 464–472. doi: https://doi.org/10.1109/WACV.2017.58.

SpaceNet, 2017. SpaceNet 2: Building Detection v2. Last visited: 07.01.2022. URL: https://github.com/SpaceNetChallenge/BuildingDetectors_Round2.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. Journal of Machine Learning Research 15, 1929–1958. doi: https://doi.org/10.5555/2627435.2670313.

Statista, 2021a. Global offshore wind energy capacity from 2009 to 2020. Last visited: 10.02.2022. URL: https://www.statista.com/statistics/476327/global-capacity-of-offshore-wind-energy/.

Statista, 2021b. Wichtigste Länder weltweit nach installierter Offshore-Windenergieleistung im Jahr 2020. Last visited: 10.02.2022. URL: https://de.statista.com/statistik/daten/studie/158501/umfrage/kapazitaeten-der-offshore-windkraft-nach-laendern-im-jahr-2009/.

Steffen, W., Richardson, K., Rockström, J., Schellnhuber, H.J., Dube, O.P., Dutreuil, S., Lenton, T.M., Lubchenc, J., 2020. The emergence and evolution of Earth System Science. Nature Reviews Earth and Environment 1, 54–63. doi: https://doi.org/10.1038/s43017-019-0005-6.

Stewart, G., Muskulus, M., 2016. A Review and Comparison of Floating Offshore Wind Turbine Model Experiments. Energy Procedia 94, 227–231. doi: https://doi.org/10.1016/j.egypro.2016.09.228.

Stoian, A., Poulain, V., Inglada, J., Poughon, V., Derksen, D., 2019. Land Cover Maps Production with High Resolution Satellite Image Time Series and Convolutional Neural Networks: Adaptations and Limits for Operational Systems. Remote Sensing 11. doi: https://doi.org/10.3390/rs11171986.

Strahler, A.H., Woodcock, C.E., Smith, J.A., 1986. On the nature of models in remote sensing. Remote Sensing of Environment 20, 121–139. doi: https://doi.org/10.1016/0034-4257(86)90018-0.

Su, H., Wei, S., Yan, M., Wang, C., Shi, J., Zhang, X., 2019. Object Detection and Instance Segmentation in Remote Sensing Imagery Based on Precise Mask R-CNN, in: IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium, pp. 1454–1457. doi: https://doi.org/10.1109/IGARSS.2019.8898573.

Sutskever, I., Martens, J., Dahl, G., Hinton, G., 2013. On the importance of initialization and momentum in deep learning, in: Dasgupta, S., McAllester, D. (Eds.), Proceedings of the 30th International Conference on Machine Learning, PMLR, Atlanta, Georgia, USA. pp. 1139–1147. doi: https://doi.org/10.5555/3042817.3043064.

Sylvain, J.D., Drolet, G., Brown, N., 2019. Mapping dead forest cover using a deep convolutional neural network and digital aerial photography. ISPRS Journal of Photogrammetry and Remote Sensing 156, 14–26. doi: https://doi.org/10.1016/j.isprsjprs.2019.07.010.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z., 2016. Rethinking the Inception Architecture for Computer Vision, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2818–2826. doi: https://doi.org/10.1109/CVPR.2016.308.

Szegedy, C., Wei Liu, Yangqing Jia, Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015. Going deeper with convolutions, in: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1–9. doi: https://doi.org/10.1109/CVPR.2015.7298594.

Tan, M., Chen, B., Pang, R., Vasudevan, V., Le, Q.V., 2018. MnasNet: Platform-Aware Neural Architecture Search for Mobile. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2815–2823. doi: https://doi.org/10.1109/CVPR.2019.00293.

Tan, M., Le, Q.V., 2019. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks, in: Chaudhuri, K., Salakhutdinov, R. (Eds.), Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA, PMLR. pp. 6105–6114.

Tan, M., Pang, R., Le, Q.V., 2020. EfficientDet: Scalable and Efficient Object Detection. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 10778–10787. doi: https://doi.org/10.1109/CVPR42600.2020.01079.

Tang, T., Zhou, S., Deng, Z., Lei, L., Zou, H., 2017. Arbitrary-Oriented Vehicle Detection in Aerial Imagery with Single Convolutional Neural Networks. Remote Sensing 9. doi: https://doi.org/10.3390/rs9111170.

Teimouri, N., Dyrmann, M., Jørgensen, R.N., 2019. A Novel Spatio-Temporal FCN-LSTM Network for Recognizing Various Crop Types Using Multi-Temporal Radar Images. Remote Sensing 11. doi: https://doi.org/10.3390/rs11080990.

Tobler, W.R., 1970. A Computer Movie Simulating Urban Growth in the Detroit Region. Economic Geography 46, 234–240. doi: https://doi.org/10.2307/143141.

Tong, X., Lu, Q., Xia, G., Zhang, L., 2018. Large-Scale Land Cover Classification in Gaofen-2 Satellite Imagery, in: IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium, pp. 3599–3602. doi: https://doi.org/10.1109/IGARSS.2018.8518389.

Torres, R., Snoeij, P., Geudtner, D., Bibby, D., Davidson, M., Attema, E., Potin, P., Rommen, B., Floury, N., Brown, M., Traver, I.N., Deghaye, P., Duesmann, B., Rosich, B., Miranda, N., Bruno, C., L'Abbate, M., Croci, R., Pietropaolo, A., Huchler, M., Rostan, F., 2012. GMES Sentinel-1 mission. Remote Sensing of Environment 120, 9–24. doi: https://doi.org/10.1016/j.rse.2011.05.028. the Sentinel Missions - New Opportunities for Science.

Tougaard, J., Madsen, P.T., Wahlberg, M., 2008. Underwater noise from construction and operation of offshore wind farms. Bioacoustics 17, 143–146. doi: https://doi.org/10.1080/09524622.2008.9753795.

Tremblay, J., Prakash, A., Acuna, D., Brophy, M., Jampani, V., Anil, C., To, T., Cameracci, E., Boochoon, S., Birchfield, S., 2018a. Training Deep Networks with Synthetic Data: Bridging the Reality Gap by Domain Randomization, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 1082–10828. doi: https://doi.org/10.1109/CVPRW.2018.00143.

Tremblay, J., To, T., Birchfield, S., 2018b. Falling Things: A Synthetic Dataset for 3D Object Detection and Pose Estimation, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 2119–21193. doi: https://doi.org/10.1109/CVPRW.2018.00275.

Tsyganskaya, V., Martinis, S., Marzahn, P., Ludwig, R., 2018. SAR-based detection of flooded vegetation – a review of characteristics and approaches. International Journal of Remote Sensing 39, 2255–2293. doi: https://doi.org/10.1080/01431161.2017.1420938.

UK Gov., 2021. Net Zero Strategy: Build Back Greener. HM Government, London. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1033990/net-zero-strategy-beis.pdf.

Ulaby, F.T., Long, D., 1991. Microwave Radar and Radiometric Remote Sensing. Artech House, Boston, London.

United Nations, 2015. The Paris Agreement, COP21. Last visited: 07.01.2022. URL: https://unfccc.int/process-and-meetings/the-paris-agreement/the-paris-agreement.

United Nations, 2021a. Global coal to clean power transition statement, COP26. Last visited: 07.01.2022. URL: https://ukcop26.org/global-coal-to-clean-power-transition-statement/.

United Nations, 2021b. Zero emission vehicles transition council: 2022 action plan, COP26. Last visited: 07.01.2022. URL: https://ukcop26.org/zero-emission-vehicles-transition-council-2022-action-plan/.

Vineyard Wind, 2022. Vineyard Wind 1. Last visited: 15.04.2022. URL: https://www.vineyardwind.com/vineyardwind-1.

Virtanen, E., Lappalainen, J., Nurmi, M., Viitasalo, M., Tikanmäki, M., Heinonen, J., Atlaskin, E., Kallasvuo, M., Tikkanen, H., Moilanen, A., 2022. Balancing profitability of energy production, societal impacts and biodiversity in offshore wind farm design. Renewable and Sustainable Energy Reviews 158, 112087. doi: https://doi.org/10.1016/j.rser.2022.112087.

Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S.J., Brett, M., Wilson, J., Millman, K.J., Mayorov, N., Nelson, A.R.J., Jones, E., Kern, R., Larson, E., Carey, C.J., Polat, İ., Feng, Y., Moore, E.W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E.A., Harris, C.R., Archibald, A.M., Ribeiro, A.H., Pedregosa, F., van Mulbregt, P., SciPy 1.0 Contributors, 2020. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. Nature Methods 17, 261–272. doi: https://doi.org/10.1038/s41592-019-0686-2.

Voinov, S., Heymann, F., Bill, R., Schwarz, E., 2019. Multiclass Vessel Detection From High Resolution Optical Satellite Images Based On Deep Neural Networks, in: IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium, pp. 166–169. doi: https://doi.org/10.1109/IGARSS.2019.8900506.

Voulodimos, A., Doulamis, N., Doulamis, A., Protopapadakis, E., 2018. Deep learning for computer vision: A brief review. Computational Intelligence and Neuroscience 2018. doi: https://doi.org/10.1155/2018/7068349.

Wackerman, C., Friedman, K., Pichel, W., Clemente-Colón, P., Li, X., 2001. Automatic Detection of Ships in RADARSAT-1 SAR Imagery. Canadian Journal of Remote Sensing 27, 568–577. doi: https://doi.org/10.1080/07038992.2001.10854896.

Wahlberg, M., Westerberg, H., 2005. Hearing in fish and their reactions to sounds from offshore wind farms. Marine Ecology Progress Series 288, 295–309. doi: https://doi.org/10.3354/meps288295.

Wang, J., Ding, J., Guo, H., Cheng, W., Pan, T., Yang, W., 2019a. Mask OBB: A Semantic Attention-Based Mask Oriented Bounding Box Representation for Multi-Category Object Detection in Aerial Images. Remote Sensing 11. doi: https://doi.org/10.3390/rs11242930.

Wang, Q., Zhang, J., Su, F., 2019b. Offshore Platform Extraction Using RadarSat-2 SAR Imagery: A Two-Parameter CFAR Method Based on Maximum Entropy. Entropy 21. doi: https://doi.org/10.3390/e21060556.

Wang, Y., Li, H., Jia, P., Zhang, G., Wang, T., Hao, X., 2019c. Multi-Scale DenseNets-Based Aircraft Detection from Remote Sensing Images. Sensors 19. doi: https://doi.org/10.3390/s19235270.

Weber, I., Bongartz, J., Roscher, R., 2021. Artificial and beneficial – Exploiting artificial images for aerial vehicle detection. ISPRS Journal of Photogrammetry and Remote Sensing 175, 158–170. doi: https://doi.org/10.1016/j.isprsjprs.2021.02.015.

Weinstein, B.G., Marconi, S., Bohlman, S., Zare, A., White, E., 2019. Individual Tree-Crown Detection in RGB Imagery Using Semi-Supervised Deep Learning Neural Networks. Remote Sensing 11. doi: https://doi.org/10.3390/rs11111309.

Wen, Q., Jiang, K., Wang, W., Liu, Q., Guo, Q., Li, L., Wang, P., 2019. Automatic Building Extraction from Google Earth Images under Complex Backgrounds Based on Deep Instance Segmentation Network. Sensors 19. doi: https://doi.org/10.3390/s19020333.

Wever, L., Krause, G., Buck, B.H., 2015. Lessons from stakeholder dialogues on marine aquaculture in offshore wind farms: Perceived potentials, constraints and research gaps. Marine Policy 51, 251–259. doi: https://doi.org/10.1016/j.marpol.2014.08.015.

Wilson, J.C., Elliott, M., 2009. The habitat-creation potential of offshore wind farms. Wind Energy 12, 203–212. doi: https://doi.org/10.1002/we.324.

Wong, B.A., Thomas, C., Halpin, P., 2019. Automating offshore infrastructure extractions using synthetic aperture radar & Google Earth Engine. Remote Sensing of Environment 233, 111412. doi: https://doi.org/10.1016/j.rse.2019.111412.

Wu, J., 1999. Hierarchy and Scaling: Extrapolating Information along a Scaling Ladder. Canadian Journal of Remote Sensing 25, 367–380. doi: https://doi.org/10.1080/07038992.1999.10874736.

Wu, X., Hong, D., Ghamisi, P., Li, W., Tao, R., 2018. MsRi-CCF: Multi-Scale and Rotation-Insensitive Convolutional Channel Features for Geospatial Object Detection. Remote Sensing 10. doi: https://doi.org/10.3390/rs10121990.

Wu, X., Hu, Y., Li, Y., Yang, J., Duan, L., Wang, T., Adcock, T., Jiang, Z., Gao, Z., Lin, Z., Borthwick, A., Liao, S., 2019. Foundations of offshore wind turbines: A review. Renewable and Sustainable Energy Reviews 104, 379–393. doi: https://doi.org/10.1016/j.rser.2019.01.012.

Xia, G.S., Bai, X., Ding, J., Zhu, Z., Belongie, S., Luo, J., Datcu, M., Pelillo, M., Zhang, L., 2018. DOTA: A Large-Scale Dataset for Object Detection in Aerial Images, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 3974–3983. doi: https://doi.org/10.1109/CVPR.2018.00418.

Xie, S., Girshick, R., Dollár, P., Tu, Z., He, K., 2017. Aggregated Residual Transformations for Deep Neural Networks, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5987–5995. doi: https://doi.org/10.1109/CVPR.2017.634.

Xing, W., 2021. China's Three Gorges connects 3.1 GW of installed wind power capacity per day. Last visited: 15.04.2022. URL: https://www.seetao.com/details/131215.html.

Xu, W., Liu, Y., Wu, W., Dong, Y., Lu, W., Liu, Y., Zhao, B., Li, H., Yang, R., 2020. Proliferation of offshore wind farms in the North Sea and surrounding waters revealed by satellite image time series. Renewable and Sustainable Energy Reviews 133, 110167. doi: https://doi.org/10.1016/j.rser.2020.110167.

Yang, H., Wu, P., Yao, X., Wu, Y., Wang, B., Xu, Y., 2018. Building Extraction in Very High Resolution Imagery by Dense-Attention Networks. Remote Sensing 10. doi: https://doi.org/10.3390/rs10111768.

Yang, X., Chen, Z., Li, B., Peng, D., Chen, P., Zhang, B., 2019. A Fast and Precise Method for Large-Scale Land-Use Mapping Based on Deep Learning, in: IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium, pp. 5913–5916. doi: https://doi.org/10.1109/IGARSS.2019.8898705.

Ye, Z., Fu, Y., Gan, M., Deng, J., Comber, A., Wang, K., 2019. Building Extraction from Very High Resolution Aerial Imagery Using Joint Attention Deep Neural Network. Remote Sensing 11. doi: https://doi.org/10.3390/rs11242970.

You, Y., Cao, J., Zhang, Y., Liu, F., Zhou, W., 2019. Nearshore Ship Detection on High-Resolution Remote Sensing Image via Scene-Mask R-CNN. IEEE Access 7, 128431–128444. doi: https://doi.org/10.1109/ACCESS.2019.2940102.

You, Y., Li, Z., Ran, B., Cao, J., Lv, S., Liu, F., 2019. Broad Area Target Search System for Ship Detection via Deep Convolutional Neural Network. Remote Sensing 11. doi: https://doi.org/10.3390/rs11171965.

Yu, Q., Liu, K., Teixeira, A., Soares, C.G., 2020. Assessment of the Influence of Offshore Wind Farms on Ship Traffic Flow Based on AIS Data. Journal of Navigation 73, 131–148. doi: https://doi.org/10.1017/S0373463319000444.

Zeiler, M.D., Fergus, R., 2014. Visualizing and understanding convolutional networks, in: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (Eds.), Computer Vision – ECCV 2014, Springer International Publishing, Cham. pp. 818–833. doi: https://doi.org/10.1007/978-3-319-10590-1_53.

Zhang, C., Sargent, I., Pan, X., Li, H., Gardiner, A., Hare, J., Atkinson, P.M., 2019a. Joint Deep Learning for land cover and land use classification. Remote Sensing of Environment 221, 173–187. doi: https://doi.org/10.1016/j.rse.2018.11.014.

Zhang, D., Mishra, S., Brynjolfsson, E., Etchemendy, J., Ganguli, D., Grosz, B., Lyons, T., Manyika, J., Niebles, J.C., Sellitto, M., Shoham, Y., Clark, J., Perrault, R., 2021a. The AI Index 2021 Annual Report. AI Index Steering Committee, Human-Centered AI Institute, Stanford University, Stanford, CA.

Zhang, F., Du, B., Zhang, L., Xu, M., 2016a. Weakly Supervised Learning Based on Coupled Convolutional Neural Networks for Aircraft Detection. IEEE Transactions on Geoscience and Remote Sensing 54, 5553–5563. doi: https://doi.org/10.1109/TGRS.2016.2569141.

Zhang, H., Deng, Q., 2019. Deep Learning Based Fossil-Fuel Power Plant Monitoring in High Resolution Remote Sensing Images: A Comparative Study. Remote Sensing 11. doi: https://doi.org/10.3390/rs11091117.

Zhang, J., Wang, Q., Su, F., 2019b. Automatic Extraction of Offshore Platforms in Single SAR Images Based on a Dual-Step-Modified Model. Sensors 19. doi: https://doi.org/10.3390/s19020231.

Zhang, L., Zhang, L., Du, B., 2016b. Deep Learning for Remote Sensing Data: A Technical Tutorial on the State of the Art. IEEE Geoscience and Remote Sensing Magazine 4, 22–40. doi: https://doi.org/10.1109/MGRS.2016.2540798.

Zhang, N., Liu, Y., Zou, L., Zhao, H., Dong, W., Zhou, H., Zhou, H., Huang, M., 2018. Automatic Recognition of Oil Industry Facilities Based on Deep Learning, in: IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium, pp. 2519–2522. doi: https://doi.org/10.1109/IGARSS.2018.8518054.

Zhang, S., Wu, R., Xu, K., Wang, J., Sun, W., 2019c. R-CNN-Based Ship Detection from High Resolution Remote Sensing Imagery. Remote Sensing 11. doi: https://doi.org/10.3390/rs11060631.

Zhang, T., Tian, B., Sengupta, D., Zhang, L., Si, Y., 2021b. Global offshore wind turbine dataset. Scientific Data 8. doi: https://doi.org/10.1038/s41597-021-00982-z.

Zhang, Y., Zhang, Y., Shi, Z., Zhang, J., Wei, M., 2019. Rotationally Unconstrained Region Proposals for Ship Target Segmentation in Optical Remote Sensing. IEEE Access 7, 87049–87058. doi: https://doi.org/10.1109/ACCESS.2019.2926227.

Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J., 2016. Pyramid Scene Parsing Network. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 6230–6239. doi: https://doi.org/10.1109/CVPR.2017.660.

Zhao, P., Gao, H., Zhang, Y., Li, H., Yang, R., 2019. An Aircraft Detection Method Based on Improved Mask R-CNN in Remotely Sensed Imagery, in: IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium, pp. 1370–1373. doi: https://doi.org/10.1109/IGARSS.2019.8900528.

Zhong, L., Hu, L., Zhou, H., 2019. Deep learning based multi-temporal crop classification. Remote Sensing of Environment 221, 430–443. doi: https://doi.org/10.1016/j.rse.2018.11.032.

Zhu, H., Chen, X., Dai, W., Fu, K., Ye, Q., Jiao, J., 2015. Orientation robust object detection in aerial images using deep convolutional neural network, in: 2015 IEEE International Conference on Image Processing (ICIP), pp. 3735–3739. doi: https://doi.org/10.1109/ICIP.2015.7351502.

Zhu, X.X., Hu, J., Qiu, C., Shi, Y., Kang, J., Mou, L., Bagheri, H., Haberle, M., Hua, Y., Huang, R., Hughes, L., Li, H., Sun, Y., Zhang, G., Han, S., Schmitt, M., Wang, Y., 2020. So2Sat LCZ42: A Benchmark Data Set for the Classification of Global Local Climate Zones. IEEE Geoscience and Remote Sensing Magazine 8, 76–89. doi: https://doi.org/10.1109/MGRS.2020.2964708.

Zhu, X.X., Tuia, D., Mou, L., Xia, G., Zhang, L., Xu, F., Fraundorfer, F., 2017. Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources. IEEE Geoscience and Remote Sensing Magazine 5, 8–36. doi: https://doi.org/10.1109/MGRS.2017.2762307.

Zoph, B., Le, Q.V., 2016. Neural architecture search with reinforcement learning. arXiv, Computer Science cs.LG. doi: https://doi.org/10.48550/ARXIV.1611.01578.

Zoph, B., Vasudevan, V., Shlens, J., Le, Q.V., 2018. Learning Transferable Architectures for Scalable Image Recognition, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 8697–8710. doi: https://doi.org/10.1109/CVPR.2018.00907.

# *Eidesstattliche Erklärung*

Versicherung an Eides statt

Ich, Thorsten Höser, versichere an Eides statt durch meine Unterschrift, dass ich die Dissertation “Global Dynamics of the Offshore Wind Energy Sector Derived from Earth Observation Data - Deep Learning Based Object Detection Optimised with Synthetic Training Data for Offshore Wind Energy Infrastructure Extraction from Sentinel-1 Imagery” selbständig und ohne fremde Hilfe angefertigt, alle Stellen, die ich wörtlich oder dem Sinne nach aus Veröffentlichungen entnommen habe, als solche kenntlich gemacht und ich auch keine anderen als die von mir angegebenen Quellen und Hilfsmittel benutzt habe. Ich versichere an Eides statt durch meine Unterschrift, dass ich die Regeln der Universität Würzburg über gute wissenschaftliche Praxis eingehalten habe, insbesondere, dass ich die Gelegenheit zum Promotionsvorhaben nicht kommerziell vermittelt bekommen und insbesondere nicht eine Person oder Organisation eingeschaltet habe, die gegen Entgelt Betreuer bzw. Betreuerinnen für die Anfertigung von Dissertationen sucht.

Ich versichere an Eides statt, dass ich die vorgenannten Angaben nach bestem Wissen und Gewissen gemacht habe und dass die Angaben der Wahrheit entsprechen und ich nichts verschwiegen habe. Die Strafbarkeit einer falschen eidesstattlichen Versicherung ist mir bekannt, namentlich die Strafandrohung gemäß § 156 StGB bis zu drei Jahren Freiheitsstrafe oder Geldstrafe bei vorsätzlicher Begehung der Tat bzw. gemäß § 161 Abs.1 StGB bis zu einem Jahr Freiheitsstrafe oder Geldstrafe bei fahrlässiger Begehung.

München, 19.05.2022

Unterschrift, Thorsten Höser