Review

Object Detection and Image Segmentation with Deep Learning on Earth Observation Data: A Review—Part II: Applications

Thorsten Hoeser [1,*], **Felix Bachofer [1]** and **Claudia Kuenzer [1,2]**

[1] German Remote Sensing Data Center (DFD), German Aerospace Center (DLR), Münchner Straße 20, D-82234 Wessling, Germany; felix.bachofer@dlr.de (F.B.); claudia.kuenzer@dlr.de (C.K.)

[2] Department of Remote Sensing, Institute of Geography and Geology, University Würzburg, Am Huband, D-97074 Wuerzburg, Germany

[*] Correspondence: Thorsten.Hoeser@dlr.de; Tel.: +49-8153-28-3966

check for updates

Abstract: In Earth observation (EO), large-scale land-surface dynamics are traditionally analyzed by investigating aggregated classes. The increase in data with a very high spatial resolution enables investigations on a fine-grained feature level which can help us to better understand the dynamics of land surfaces by taking object dynamics into account. To extract fine-grained features and objects, the most popular deep-learning model for image analysis is commonly used: the convolutional neural network (CNN). In this review, we provide a comprehensive overview of the impact of deep learning on EO applications by reviewing 429 studies on image segmentation and object detection with CNNs. We extensively examine the spatial distribution of study sites, employed sensors, used datasets and CNN architectures, and give a thorough overview of applications in EO which used CNNs. Our main finding is that CNNs are in an advanced transition phase from computer vision to EO. Upon this, we argue that in the near future, investigations which analyze object dynamics with CNNs will have a significant impact on EO research. With a focus on EO applications in this *Part II*, we complete the methodological review provided in *Part I*.

Keywords: artificial intelligence; AI; machine learning; deep learning; neural networks; convolutional neural networks; CNN; image segmentation; object detection; earth observation

1. Introduction

The availability of spatio-temporal Earth observation data has increased dramatically over recent decades. This data provides the needed information to understand and monitor land-surface dynamics on a large scale, for example: urban growth and the distribution of settlements [1], vegetation cover and its temporal dynamics [2], and water availability [3]. Traditionally such studies analyze aggregated classes: buildings and impervious surfaces are summarized as built-up areas, trees and grassland become regions with different vegetation intensities, and open water and shorelines are mapped as binary water masks. However, to investigate the intrinsic characteristics of land cover and land use classes, the spatio-temporal dynamics of the single entities which compose those classes must be taken into account. Doing so will allow us to better understand the livelihoods on our planet: is urban growth characterized by expansion or redensification? Which kind of buildings are newly built and how are existing buildings modified? Which specific species and plant communities are present in areas labeled as vegetated? How does human interaction with ecosystems specifically affect single maybe endangered species? Is a specific ship the cause for a specific oil spill?

The spatio-temporal patterns of single entities like vehicles for transportation, the accumulation of humans and goods or artificial infrastructures enables the description of space and how it is

used precisely. Therewith, answering urgent geoscientific research questions becomes possible and Earth observation as a tool can be more intensively used in everyday applications. Overall, in Earth observation, land surface dynamics could be better described as *specific things* within *general classes*.

One crucial requirement to identify objects in remote sensing images is the availability of spatio-temporal Earth observation data with a high to very high resolution. With the ongoing development of spaceborne missions and open archives, the trend in increasing data availability will continue in the near future. Nevertheless, due to the vast amount of data, processing techniques are required that are adaptive, account for spatio-temporal characteristics, fast in processing and can be automated while still competing with human performance or even outperforming them [4,5].

With the emergence of deep learning, a technique has been introduced which is able to meet these requirements [4]. During the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) in 2012, Krizhevsky et al. [6] proved the potential of convolutional neural networks (CNNs) by outperforming established methods [7]. The proposed CNN predicts only on learned features from training data instead of using handcrafted features designed by humans. Since 2012, the diversity of deep-learning models has been steadily increasing. Today, they are state-of-the-art techniques for image segmentation and object detection which can extract fine-grained information and objects from imagery data. Furthermore, the capabilities of artificial intelligence and deep learning were found to be exceptionally relevant to current ecological challenges as well as their interactions with economy and society, as formulated in the sustainable development goals [8].

These properties make deep learning the most promising tool for handling complex, large-scale environmental information, with CNNs especially capable of analyzing imagery data. Considering the leading role of deep learning in image recognition, image segmentation and object detection, these methods can automatically archive and classify Earth observation data on a scene, instance and pixel level. Hence, deep learning is ideally suited for Earth observation research in order to create inventories of objects towards a digital twin earth [9] as illustrated in Figure 1, in order to better analyze object dynamics and their impact on our planet.

With these features of deep learning in mind, the question arises how, and for which data types and applications, deep learning has been used in Earth observation research. Previous reviews about deep learning on Earth observation discussed the potential of deep learning [4,10–14] or the usage of different model types [10–14] in a broad perspective. More recent reviews also started discussing deep learning from a data type perspective, like for hyperspectral [15–17] or radar [18] data as well as from an application perspective as in [19] for change detection. In most of these reviews, the CNN deep-learning model type received the most attention and can today be seen as an established deep-learning model in Earth observation applications. Due to the importance of the CNN, not just in the field of Earth observation but also in computer vision [20], this review focuses specifically on CNNs for object detection and image segmentation applied to Earth observation data. Furthermore, to provide a detailed review of the most common data and architectures, the review focuses on the most used sensors types, like optical, multispectral and radar. To our knowledge, this review is the first in Earth observation with a detailed perspective on a specific deep-learning model type, allowing a comprehensive discussions which will help to understand recent developments and current applications.

The predecessor of this review series, *Part I: Evolution and Recent Trends* [21] hereafter referred to as "*Part I*", is dedicated to thoroughly reviewing the history of CNNs in the field of computer vision since its emergence in 2012 until late 2019. This review, *Part II: Applications*, describes the transition from CNNs developed in the field of computer vision to their application in the field of Earth observation research in depth by reviewing 429 papers published in Earth observation journals. In detail, the scope of this review is to give an overview of the:

- spatial distribution of research teams and study sites,
- most employed platforms and sensor types,
- frequently used annotated deep-learning datasets featuring Earth observation data,

- application domains and specific applications, where deep learning for object detection and image segmentation was used,
- most employed CNN architectures and their adaptations to remote sensing data applied for object detection and image segmentation,
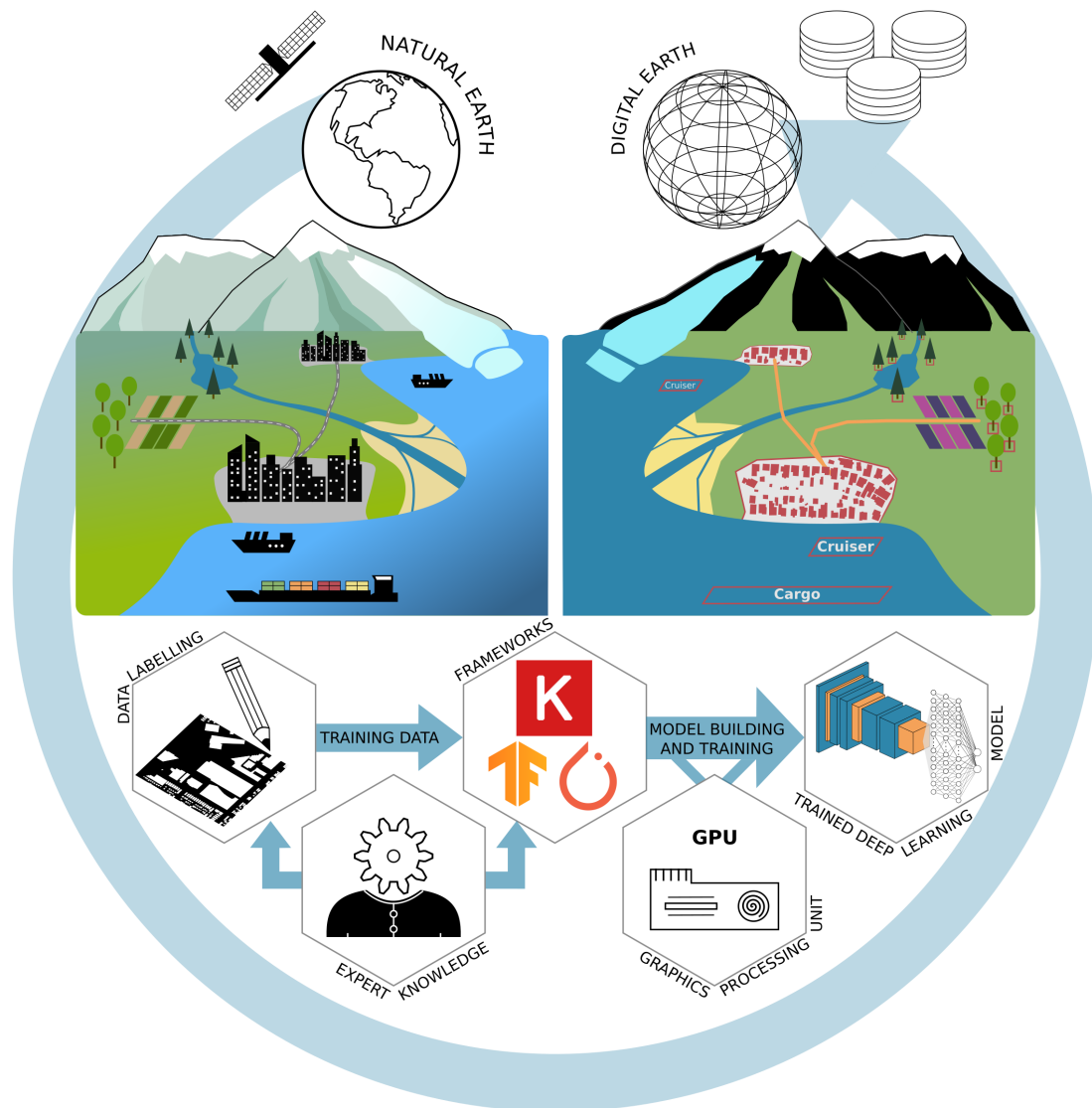- deep-learning frameworks which are commonly used by Earth observation researchers.

Figure 1. Conceptual overview of the derivation of digital entities from remotely sensed data using deep-learning techniques. The upper left part describes the natural earth with different land surfaces, objects and their interactions. The upper right part shows the digitized version of the natural earth with a focus on extracting specific entities. The lower part describes a deep-learning workflow: the creation of training data; model building and training; and application of the trained model to produce large-scale inventories of objects and their dynamics on Earth's surface.

Continuing from *Part I* [21], where the theoretical background on CNNs was provided, this review further lowers the entry barriers for researchers who want to apply deep learning in Earth observation. This is done by presenting the above-mentioned attributes of CNNs in Earth observation research and pointing to the established methods in combination with data types and applications. This will help to find the best combination of data, models and frameworks to successfully implement

a deep-learning workflow as pictured in Figure 1. Furthermore, we show new perspectives of deep-learning applications to Earth observation data by discussing recent developments.

2. Review Methodology

Figure 2a describes the review process. The majority of reviewed sources come from 14 peer-reviewed journals with a focus on Earth observation by analyzing imagery data, see Figure 2b. In addition, the IGARSS proceedings were included due to their combination of method development and application of deep learning from an Earth observation perspective. Finally, to picture the reach of Earth observation in a multidisciplinary context, Scientific Reports by Nature publishing was added, resulting in 16 journals, see Figure 2b.
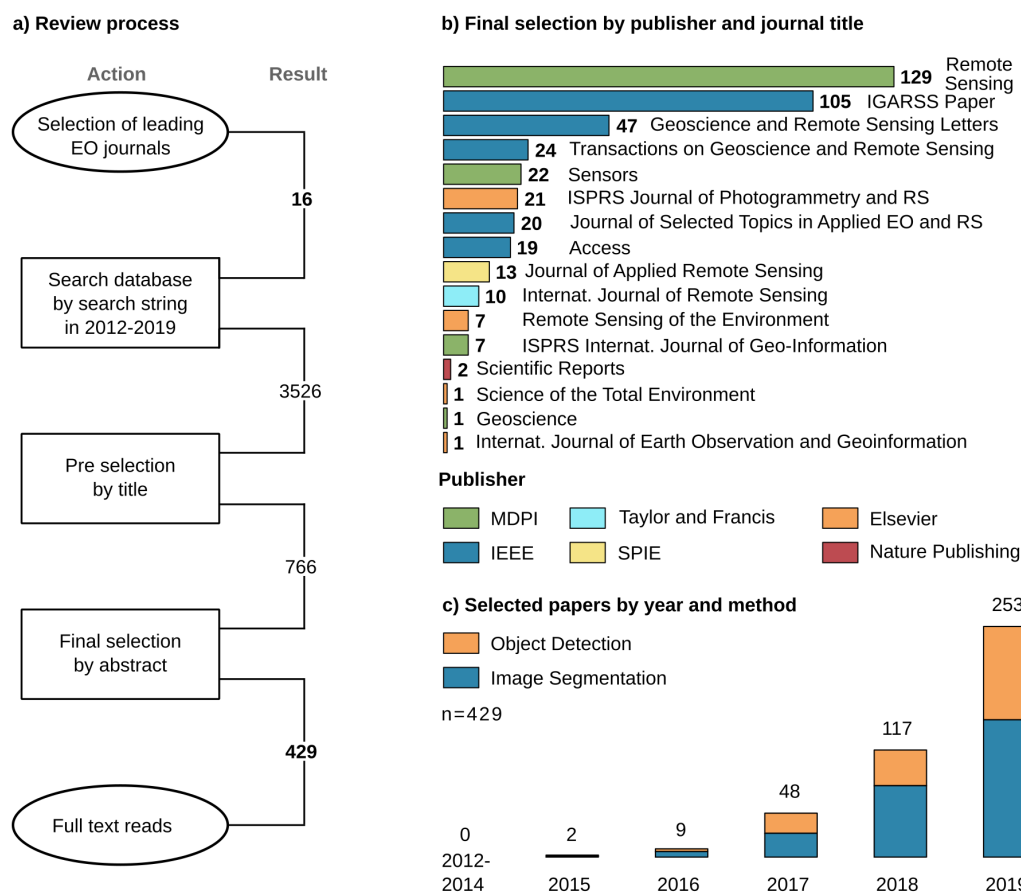

Figure 2. (**a**) Summary of the review process of 16 journals and 429 included papers. The search period for potential papers starts in 2012, due to the initial introduction of CNNs by Krizhevsky et al. [6] for image processing. (**b**) Overview of the included journals, publishers and number of papers per journal. (**c**) Temporal distribution of the 429 papers separated by task: object detection (168) and image segmentation (261).

A search string queried the journal databases: ``deep learning'' OR ``convolutional neural network'' OR ``fully convolutional'' OR ``convolutional network''. With respect to *Part I* [21], we defined the work of Krizhevsky et al. [6] in 2012 as starting point for the emergence of CNNs in image analysis and included all articles until 2019 to maintain a whole year frequency. The resulting 3526 potential papers were then filtered by title, giving 766 candidates, and further screened by abstract. To focus on publications which mainly extract spatial features from imagery data, we included articles which used optical, LiDAR, multispectral, thermal or radar data in combination with a CNN as the deep-learning model. To specifically concentrate on image segmentation and object detection, we excluded all publications which investigate image recognition tasks, in EO often referred to as scene

recognition, where a single label is predicted for the entire image. Furthermore, to focus thematically, we concentrate on land-surface processes and therewith excluded publications which investigate atmosphere or atmosphere-land interaction applications. These criteria for in- and exclusion lead to a selection of studies which can be compared and discussed together since they follow similar approaches in CNN architectures and at the same time represent a wide thematical variety.

Finally, with 429 papers, the review represents the development of deep learning in Earth observation since 2012. Despite the searched period from 2012 until 2019, the onset of CNNs used for image segmentation and object detection in an Earth observation context took place in 2015 and increased to 253 papers in 2019, see Figure 2c. Since 2016 the group of papers concerning image segmentation is larger than the group using CNNs for object detection. The review was conducted based on the 429 selected papers by full-text reads while looking for attributes like the spatial distribution of author's affiliations and study sites, researched applications as well as employed datasets, remote sensing sensors, CNN architectures and deep-learning frameworks. Therefore, the reported CNN architecture for each paper is the best performing architecture investigated, especially in ablation studies which compare multiple architectures. The complete table of all 429 papers reviewed with all determined attributes is provided in Appendix A in .csv format.

3. Results of the Review

3.1. Spatial Distribution of Studies

The spatial distribution of first author's affiliations depicted in Figure 3 clearly shows the largest share have Asian affiliations (68%), with 62% coming from China alone. Furthermore, European and American affiliations contribute with about 19% and 10% respectively. The high number of publications from Chinese affiliations can be explained by a high contribution of papers with a strong methodological background. Such papers have a focus on developing deep-learning algorithms for Earth observation data and mostly use well-established datasets for ablation studies.
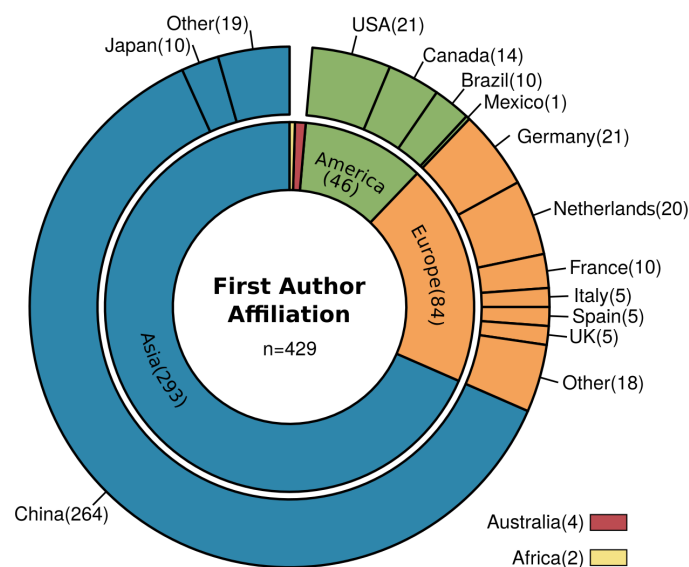

Figure 3. Overview of first author affiliation grouped by countries and continents. The largest shares among the continents are: Asia (68%), Europe (20%) and America (11%); and the largest shares among the countries are: China (62%), USA, Germany and Netherlands (each 5%).

Further details become clear when looking at the distribution of study sites in Figure 4. In Figure 4a which shows study sites of all publications, the class *Multilocal* has the biggest share and together with studies of class *Not specified*, they represent about 38% of all studies. It is characteristic that most of these studies took data from multiple locations or not-further defined places with a minor

focus on a geoscientific research question or a distinct study site. Rather, they focus on the technical implementation of CNNs on Earth observation data and proof-of-concept studies, which are an essential driver of the increasing usage of CNNs in Earth observation research. In this subgroup of studies without a specific located study site, the contribution from authors with a Chinese affiliation (84%) is by far the largest. The next two larger contributors are Germany and France with 2% each.
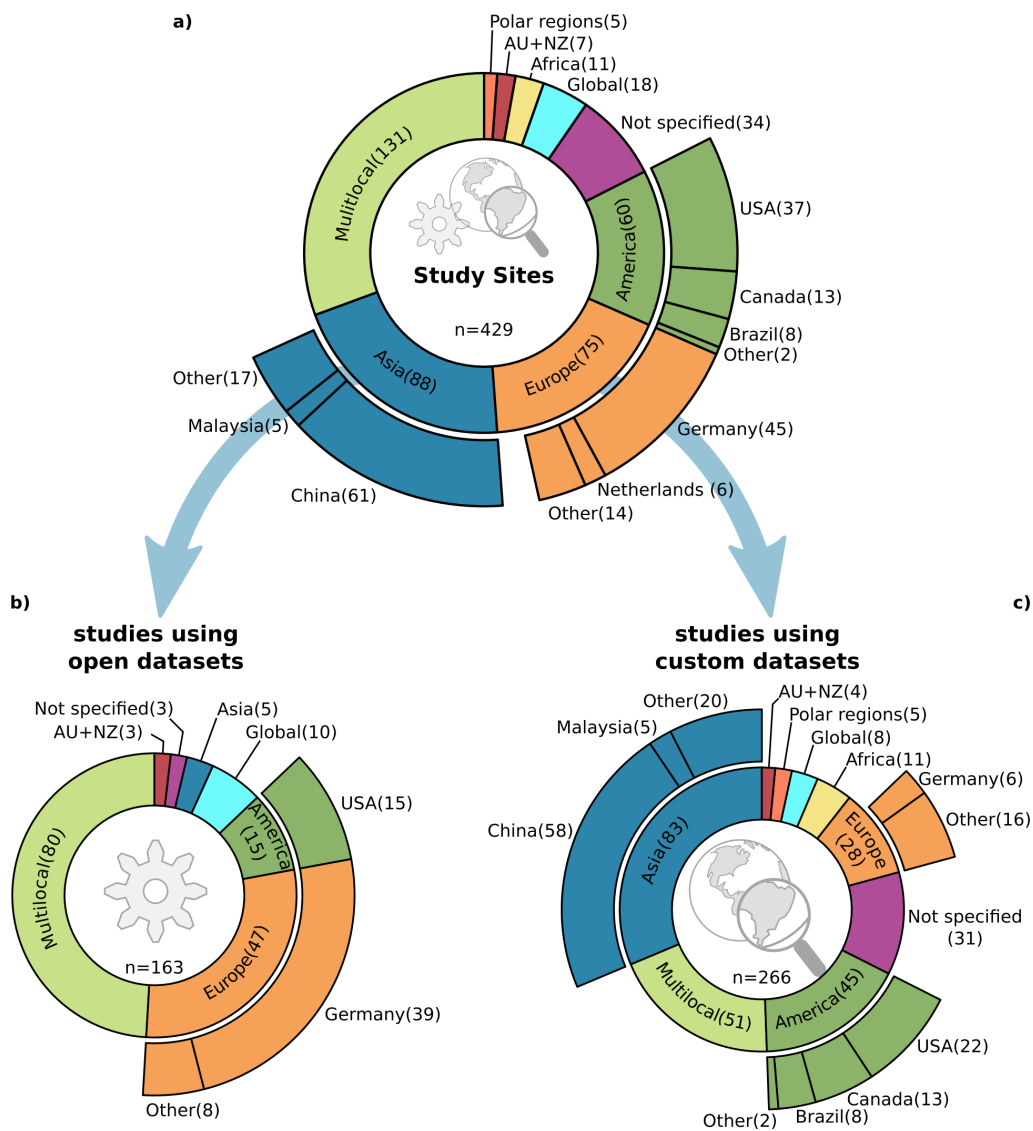

Figure 4. Overview of study site locations, where *Multilocal* refers to locations scattered over multiple continents commonly without a distinct spatial focus. (**a**) The study site locations for all 429 reviewed studies. (**b**) Study site locations of studies which used open datasets (38%), commonly focusing on method development. (**c**) Study site locations of studies which used custom datasets (62%) which investigate method development, and especially proof-of-concept studies as well as large-scale geoscientific research questions.

Furthermore, when splitting the study sites into two groups, as presented in Figure 4b,c, the substantial contribution of Chinese studies especially in publications with a methodological focus, becomes evident. The used datasets separate the two groups: the first group in (b) shows the studies, which are using established datasets. With this, the used datasets define the study sites, and the focus is on examining an algorithm for a specific task. The second group (c) shows studies which use a customized dataset and therefore the selected location is more meaningful for the underlying research question. Contributions from Chinese affiliations in the first group (b) are about 78%, higher than their

overall contribution (62%). As follows, contributions of Chinese affiliations in group (c) with custom study sites are still high (52%) but lower compared to the overall contribution.

The number of studies located in Germany is also noticeable when separating the study sites in the two groups mentioned above. In group (b) there are 39 studies, whereas the share drops to only 6 studies in group (c). This imbalance results from two well-established datasets, the ISPRS Potsdam and Vaihingen [22] datasets, commonly used for ablation studies of image segmentation algorithms. Interestingly, 54% of the 39 studies with a German location in group (b) are published by authors who are with Chinese affiliations, which further proves the Chinese contribution to method development in deep learning for Earth observation.

Furthermore, a closer look at Figure 4c shows the still small number (3%) of studies with both a global perspective and customized datasets, or in other words deep learning applied to geoscientific research questions on a global scale. However, the study site distribution of customized datasets shows a more balanced picture, except for the African continent.

3.2. Platforms and Sensors

Traditionally, deep learning with CNNs was developed in computer vision for RGB image analysis where rich feature information is present due to a sufficient image resolution such as in the ImageNet [7], PASCAL VOC [23,24], MSCOCO [25] or Cityscape [26,27] datasets. Therefore, widely used CNN models are designed for three-channel input data and are often the starting point for investigations in other domains like Earth observation. This origin can be related to the most commonly investigated data types. Optical RGB images with a very high (<100 cm) and high (100 < 500 cm) spatial resolution are the most widely used imagery data in Earth observation with 56%, as depicted in Figure 5a.
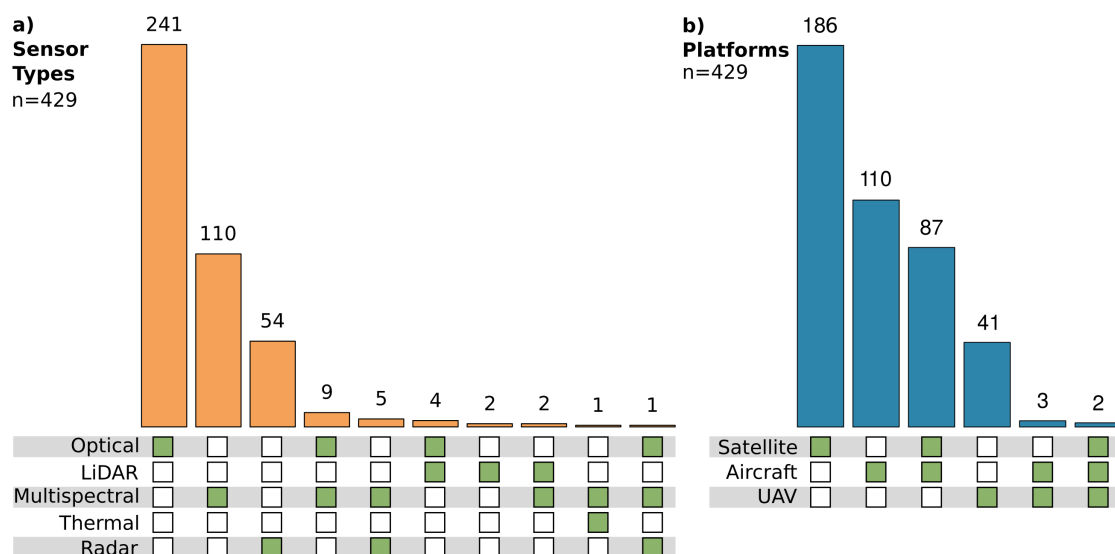

Figure 5. (**a**) Distribution of employed sensor types and their combinations with the largest shares for optical (56%), multispectral (26%) and radar (13%) sensor types. (**b**) Distribution of employed platforms and their combination with the largest shares for satellites (43%), aircraft (26%) and their combination (20%).

Studies which apply multispectral or radar data have shares of 26% and 13% respectively. However, especially in radar remote sensing, deep learning is gaining in popularity [18]. The number of publications which exploit radar data increased from 4% in 2017 to 15% in 2018 and 2019. This increase indicates the subsequent percolation of deep-learning methods in Earth observation. Sensors which provide a high spatial resolution are by far the most widely employed across all sensor types. Data with a spatial resolution <500 cm (VHR and high) are used in 79% of the publications and even when

increasing the spatial resolution to <100 cm (VHR) they still represent 43% of all studies and are therewith the largest group.

Figure 5b shows that 43% of the studies used spaceborne platforms, whereas 36% solely rely on airborne platforms. The use of airborne platforms supports the above-given statement about the usage of data with a very high to high spatial resolution. Furthermore, it also supports the findings in Section 3.1 that most publications focus on small-scale, proof-of-concept and method development studies rather than large-scale applications. However, the 43% of publications which use spaceborne platforms should have the potential to investigate larger scales, beyond regional and national data acquisition projects. Still, the majority investigate local, single scenes or smaller clippings, scattered over multiple scenes, also focusing on proof of concept or method development. Therefore, the spatial resolution of the imagery data is mainly very high or high when looking at the spaceborne missions employed by the studies, see Figure 6.
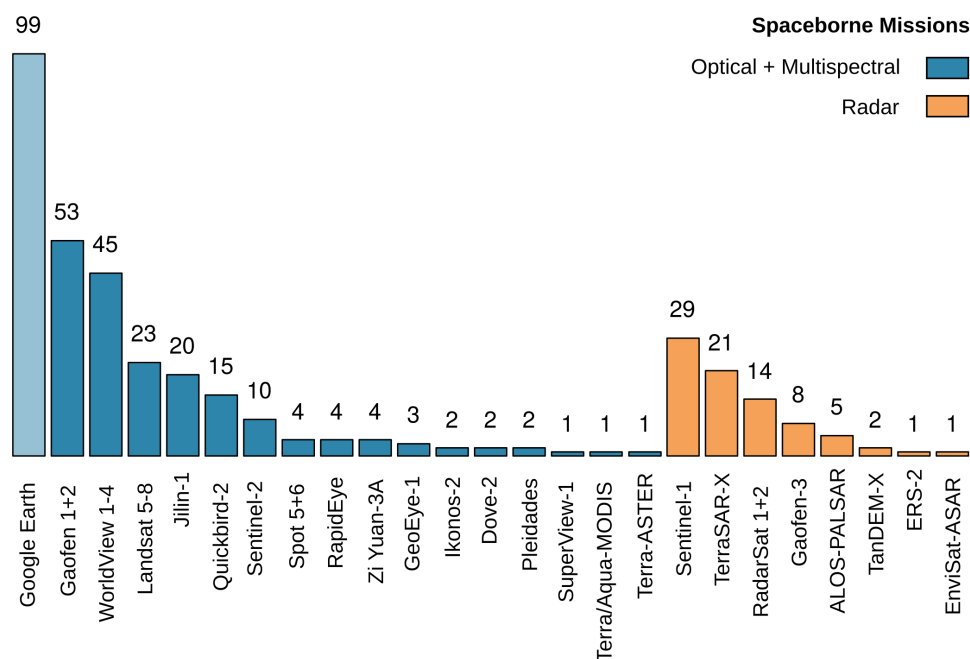

Figure 6. Overview of the most employed spaceborne missions in the 429 studies. Google Earth is also included due to its importance for receiving optical images with a high and very high spatial resolution, even when the underlying mission is not reconstructable. Overall, the overview shows the relevance of missions which focus on a high to very high spatial resolution and optical sensors.

A prominent number of studies use Google Earth as a data source. Even when the zoom level in Google Earth determines the platform and mission, the reported spatial resolution in such studies was between 50 and 200 cm, which indicates that a spaceborne optical sensor sensed the image. Hence, Google Earth appears among spaceborne optical missions, even when conveying every specific mission was not possible. Besides Google Earth, the reported spaceborne missions Gaofen 1 + 2 and WorldView 1–4 are the most established data sources, both known for their high spatial resolution. Similar to the study contributions, with the Gaofen missions, Chinese remote sensing missions are also prominent in the field of data acquisition. Frequently employed spaceborne radar missions such as TerraSAR-X, RadarSat 1 + 2 and Gaofen-3 also have a relatively high spatial resolution among radar sensors. The large number of 29 studies which use Sentinel-1 data can be related to 17 investigations that work on ship detection, which will be further discussed in Section 3.4.1.

3.3. Datasets Used

Freely available datasets and their often-associated challenges play an essential role in the development of deep-learning methods. The impact of datasets like ImageNet [7], PASCAL VOC [23,24], MSCOCO [25] or Cityscape [26,27] are striking on the evolution of CNNs as discussed in *Part I* [21]. A similar impact can be seen in the field of Earth observation. Datasets which were present since early 2013 have widely affected the development of CNN architectures as well as the fields of application. Table 1 summarizes the most frequently used datasets within the 429 studies in this review.

When looking at the datasets for image segmentation, the strong relation to urban VHR feature extraction, building footprints and road extraction becomes evident. The already mentioned ISPRS Potsdam and Vaihingen datasets [22] are the most widely used. Together with the IEEE Zeebruges dataset [28], they are applicable in both urban VHR feature extraction and less frequently used for extracting building footprints. The Massachusetts buildings and roads datasets published by Mnih [29] are also commonly used in the settlement domain. Building footprint extraction is further enhanced by several SpaceNet challenges featuring building footprints [30–33] as well as by the WHU Building dataset [34]. Road segmentation was also a topic of several datasets and challenges like the DeepGlobe Roads challenge in 2018 [35] or two SpaceNet challenges [36,37] as well as the published dataset by Cheng et al. [38].

The most popular datasets for object detection are multi-class object detection datasets like NWPU VHR-10 [39], DOTA (Dataset for OD in Aerial Images) [40] or RSOD (Remote Sensing Object Detection) [41]. They provide bounding boxes for 10 to 15 classes, and in case of DOTA the bounding boxes are also available as rotated bounding boxes. Like ISPRS Potsdam and Vaihingen [22], those datasets are mainly used for method development and ablation studies. Nevertheless, since single classes can be extracted easily from such datasets, they are also frequently used for ship, car or aircraft detection, which are all classes that are commonly present in multi-class datasets. Other datasets focus solely on a single class, often related to the transportation sector. SSDD (SAR Ship Detection Dataset) [42] and OpenSARShip [43,44] both focus on ship detection from SAR imagery, whereas Munich 3K [45], VEDAI (Vehicle Detection in Aerial Imagery) [46], and busy parking lot are used for car detection. Busy parking lot is special in the way that it is an annotated high-resolution video [47].

Overall, open datasets and challenges have a significant influence on the topics investigated by researchers, which will be further discussed in the next Section 3.4. However, 265 studies (62%) used custom datasets or combined custom datasets with the datasets mentioned above. In order to create custom datasets, labeling by hand is the most common and accurate but also labor-intensive way, whereas creating datasets from synthetic data is fast, generic and offers an inexpensive alternative as shown by Isikdogan et al. [48] and Kong et al. [49]. Since synthetic data is hard to create for multispectral and radar remote sensing, weakly supervised approaches [50–52] and studies that leverage OSM (Open Street Map) data [53–68] offer insights in how to use fuzzy data sources [69]. Thus, researchers are encouraged to use such approaches or small-scale hand-labeled datasets for proof-of-concept studies in order to build custom, large-scale, deep-learning datasets in the next step. When building deep-learning datasets, several properties should be considered, Long et al. [5] give a thorough discussion about the creation of benchmark datasets, to which we refer for further reading.

Table 1. Overview of used open datasets which were mentioned two times or more in the reviewed publications. The column, *Task* describes the main usage of the datasets from a methodological perspective divided into the two groups: object detection (OD) and image segmentation (IS). The application domains settlement, transportation and multi-class object detection are the most prominent. Of specific applications the detection of ships, cars and multiple object classes as well as the extraction of building footprints, urban VHR features and road network extraction are the most investigated topics.

Dataset	Year	Task	Domain	Application	Sensor	Count
ISPRS Vaihingen [22]	2016	IS	settlement	urban VHR feature extraction	multispectral	47
ISPRS Potsdam [22]	2016	IS	settlement	urban VHR feature extraction	multispectral	35
NWPU VHR 10 [39]	2014	OD	multi-class OD			33
DOTA (Dataset for OD in Aerial Images) [40]	2018	OD	multi-class OD		optical	17
Massachusetts Building [29]	2013	IS	settlement	building footprint	optical	11
Munich 3K [45]	2016	OD	transportation	cars	optical	9
Massachusetts Roads [29]	2013	IS	transportation	road network	optical	9
SSDD (SAR Ship Detection Dataset) [42]	2017	OD	transportation	ships	SAR	9
VEDAI (Vehicle Detection in Aerial Imagery) [46]	2016	OD	transportation	cars	optical	7
AIRSAR UAVSAR [70]	2016	IS	agriculture/ transportation	multi crop type/ship	SAR	7
WHU Building Aerial [34]	2018	IS	settlement	building footprint	optical	6
Cheng roads [38]	2017	IS	transportation	road network	optical	5
RSOD (Remote Sensing OD) [41,71]	2015	OD	multi-class OD		optical	5
IEEE Zeebruges [28]	2015	IS	settlement	urban VHR feature extraction	optical + LiDAR	4
HRSC2016 (High-Resolution Ship Collections) [72]	2016	OD	transportation	ships	optical	4
GID (Gaofen Image Dataset) [73]	2018	IS	general LCLU	multi-class LCLU	multispectral	4
Zhang Aircraft [50]	2016	OD	transportation	aircraft	optical	3
SpaceNet Building [31]	2017	IS	settlement	building footprint	multispectral	3
UCAS-AOD [74]	2015	OD	transportation	cars/aircraft	optical	3
LCZ42 (Local Climate Zone 42) [75]	2020	IS	settlement	local climate zones	multispectral + SAR	2
Busy parking lot [47]	2018	OD	transportation	cars	optical	2
DeepGlobe Roads [35]	2018	IS	transportation	road network	multispectral	2
NWPU RESICS 45 (Remote Sensing Image Scene Classification) [76]	2017	OD	multi-class OD/settlement	industry	optical	2
INRIA (Institut national de recherche en informatique et en automatique) [77]	2017	IS	settlement	building footprint	optical	2
Open SAR Ship Dataset [43,44]	2017	OD	transportation	ships	SAR	2

3.4. Research Domains and Applications

In Figure 7, nine application domains are depicted to provide an overview of the diversity of applications of deep learning in Earth observation. Three categories, settlement, transportation and multi-class object detection, are directly connected to the previously mentioned datasets. Within these categories, the classes ships, road network and cars as well as building footprints and urban VHR feature extraction and the entire group multi-class object detection, represent 53% of all publications. Through this, the influence of open datasets become distinct and show the strong data-driven paradigm, which is necessary for method development and establishing CNNs as a common tool for the Earth observation community.
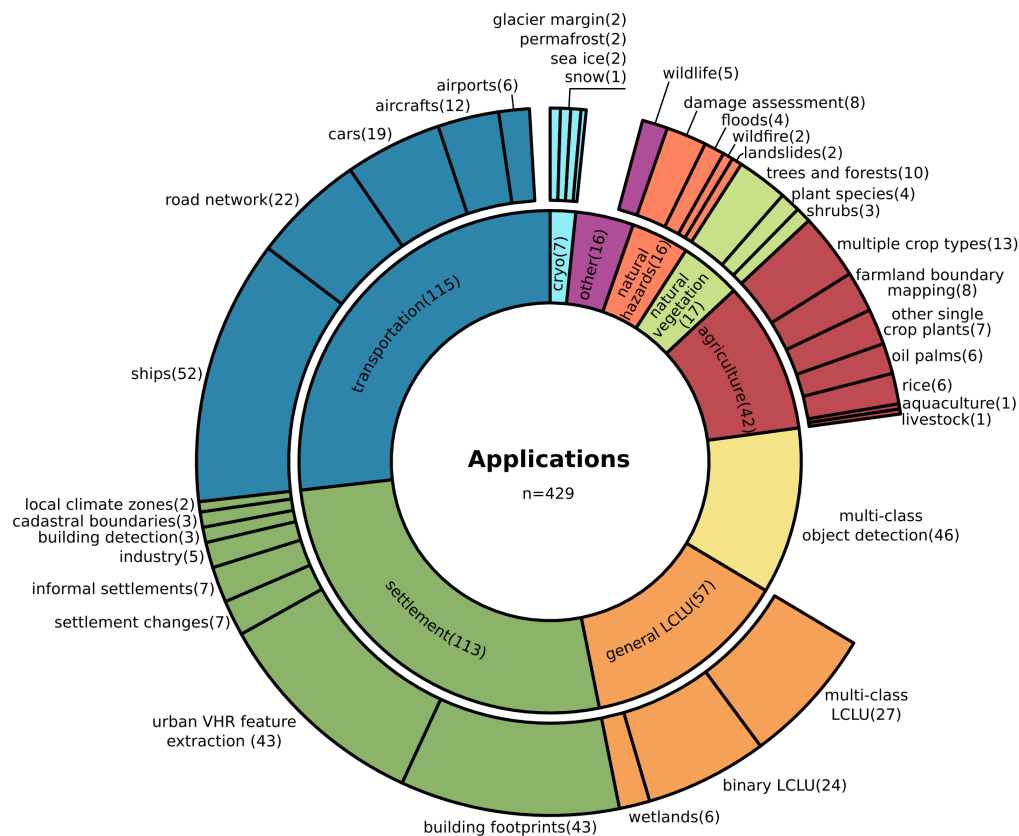

Figure 7. Detailed summary of all 429 investigated applications and their nine greater categories, called application domains. Transportation (27%), settlement(26%) and general land cover land use (LCLU) (13%) have the largest shares among the domains. Ship detection (12%), urban VHR feature extraction and building footprints (each 10%), as well as the entire group of multi-class object detection (11%), are the most investigated specific applications.

In this section, all nine application domains will be investigated thoroughly by mentioning both research topics for which CNN-based deep-learning models are already established and novel applications. In the following, the domains and single applications will be discussed by decreasing order of the number of publications, following Figure 7.

3.4.1. Transportation

Of all reviewed publications, 27% investigate targets related to the transportation sector, such as the detection of ships, cars, aircraft or entire airports as well as the segmentation of road networks. Ship detection is one of the most studied Earth observation object detection problems, with many best practice examples on how to transfer deep-learning algorithms to the specific needs of remotely sensed data [57,78–110]. Hereby, the relatively more challenging inshore ship detection, compared to offshore

situations, draws the most interest. For better detection, studies are using land-sea masks derived by elevation models or coastlines from OSM to suppress false detection of ship-like structures on land [57]. Other studies process the land-sea mask jointly on the same image the ships are detected from, within one deep-learning model. This is done by combining an early land-sea binary segmentation before the ship detection step [86–89]. Another approach is to train on negative land samples, in order to teach the model that ships do not appear in such an environment [91,92]. In addition, Chen et al. [90] proposed a method, which produces specific ship aware attention masks before detection, which are then used to support more precise ship localization during the detection process.

Images with a very high resolution were especially used to detect ships in situations where they are lying very close together, like in harbors or near harbor berths. In order to separate multiple ships or disentangle ships and harbor structures, rotated bounding boxes were found to be very useful [79,93–105]. In contrast to horizontal bounding boxes, they allow a higher signal to noise ratio and hence, better detection performance [95]. Also, the overlap of nearby bounding boxes decreases, which prevents falsely suppressed bounding boxes of dense conglomerations of ships. Other approaches optimize the detection performance for densely packed ships and near land anchoring situations by incorporating nearby spatial context into the detection model [78,97,106].

By combining object detection with further classification or segmentation modules, studies were also able to determine a specific ship class [57,103,107,108], extract instance masks for ships [95,106] or even predict their direction [103,109]. Since ships are relatively small targets against a large non-target background, the sea, weighted cost functions were used to improve training and performance [110] significantly. Overall, ship detection is an application which uses both radar and optical data. 44% use radar data, often coming from the SSDD [42] or OpenSARShip datasets [43,44], whereas 56% use optical data. However, considering the increasing performance of ship detection, there is a lack of studies which after detection, analyze the information in a spatio-temporal context.

For modern traffic managing, car detection form remotely sensed data is an important tool. Problems and also solutions are similar to ship detection: the targets are relatively small and densely packed in parking lots or traffic jams; also, the orientation is important to predict the travel direction. Hence, car detection requires rotated bounding boxes [111,112], prediction of instance attributes like the orientation after detection [113] or segmentation masks [47,114–116] likewise to ship detection. Furthermore, cars are relatively tiny objects in an even more complex environment than ships. Therefore, Koga et al. [117] and Gao et al. [118] gave special consideration to deal with complex environments and hard to train examples by employing hard example mining during the training of their deep-learning models.

The studies of Li et al. [112] and Mou and Zhu [47] investigated detection and segmentation of car instances from remotely sensed video footage of a parking lot. Therewith, they proved that deep-learning algorithms are capable of processing up to 10 frames per second [112] of very high-resolution optical data to monitor dense traffic situations.

In the aviation sector, both airports [119–121] and aircraft were detected [50,122–127]. Hereof, Zhao et al. [125] successfully extracted instance masks for each aircraft and Wang et al. [126] jointly detected and classified the type of aircraft in a military context. Hou et al. [127] tracked aircraft in video footage sensed by the spaceborne Jilin-1 VHR optical sensor, pushing the border from detection to tracking which is highly important in a transportation context.

Applications with a stronger focus on segmentation within the transportation group are mainly extracting road networks [38,62,128–137]. The early work of Mnih [29] in 2013 provided the open Massachusetts roads dataset, followed by SpaceNet [36,37] and DeepGlobe [35] challenges, as well as further datasets [38]. Since roads appear in a complex environment which contains very similar surfaces and at the same time binary road segmentation has an imbalanced target–background ratio, studies focused on handling of these problems. Lu et al. [129] counterbalanced the target–background imbalance by weighted cost functions and gained stable results on multiple scales and spatially transferable models. Wei et al. [130] incorporated expert knowledge into the cost function by describing

typical geometric structures for a better focus on roads. More recently, the combination of multiple CNN models [131,132] and heavy CNN architecture adaptation [128,133] led to further improvements.

Beside the road surface area, centerlines are of interest and models were designed to derive them jointly [38,129,134,135]. In other approaches, Wu et al. [62] and Yang et al. [136] showed how OSM centerlines could be used as the only label to segment road networks successfully. Going into even more detail, Azimi et al. [137] proposed Aerial LaneNet to segment road markings from very high-resolution imagery.

3.4.2. Settlement

Similar to the transportation sector, 26% of the publications investigated studies in a settlement context. The aforementioned ISPRS Potsdam and Vaihingen datasets [22] are responsible for the large group on urban VHR feature extraction. Most of those studies have a methodological focus, disentangling the complex structures in an urban environment [138–154]. Hence, further details are discussed in Section 3.5.2 with a focus on the progress which was made in image segmentation architectures for remote sensing.

Closely related to general urban feature extraction, are the specific building footprint applications. Few early studies focused on detecting buildings, but soon not just bounding boxes or point locations, but the footprint of the buildings was derived using image segmentation [34,53,55,56,61,63,64,66,155–167]. Again, Mnih [29] and the Wuhan university [34] provided important datasets. Furthermore, SpaceNet [30–33] and DeepGlobe [35] challenges with their own datasets increased the attention on this application. The INRIA dataset [77] is a benchmark dataset which enabled studies to take global differences in the building structure into account [155,156]. Several studies used OSM labels and obtained reasonable results [53,55,56,66], helping to build models which are invariant to large-scale differences in building structures. Still, OSM data accuracy is highly heterogeneous [53]. For example, in east African rural areas Vargas-Munoz et al. [63] amended OSM data by adding unlabeled buildings using a deep-learning approach which derived building footprints, compared them with OSM layers and added missing entries; and Maggiori et al. [55] used raw and inaccurate OSM labels as a starting point and manually labeled data to refine those during the training of the model to finally derive building footprints in France.

Since accurately derived edges of buildings and noisy results are the major issues, much attention was given to remedy this weakness by exploiting edge signals [64], edge-sensitive cost functions [157–159], sophisticated model designs [56,61,155,156,160–164], the usage of conditional random fields (CRFs) for refinement of the derived footprints [165] and multiscale approaches [166]. Even though those insights were reported for building footprint extraction, they are also important for remote sensing image segmentation in general.

Using those findings, Yang et al. [157] were able to apply a deep-learning-based approach to extract building footprints in a nationwide survey for the United States. Another large-scale application was conducted by Wen et al. [167] in Fujian Province, China. They employed an instance segmentation approach, which can derive footprints of buildings with complex structures in urban as well as rural environments.

The detection of areas of informal settlements or footprints of buildings within such areas were studied in [168–174]. A multi-temporal perspective on how those areas change over four time steps was conducted by Liu et al. [174]. In general urban settings, settlement changes like new build areas or changes in building structures were investigated by employing change detection approaches which use both multispectral and optical [175–177] as well as radar data [178–180]. Fewer studies investigated specific industrial applications like the detection of power plants and chimneys [181–183], derived cadastral boundaries [184,185] or recently differentiated urban areas due to their local climate zones [186,187].

3.4.3. General Land Cover and Land Use

LCLU mapping has a long tradition in Earth observation, and with 13% of all publications, LCLU is the third largest group of applications of the reviewed papers. The majority of multi-class LCLU publications conducted proof-of-concept studies on a local scale by demonstrating how deep-learning models can be applied in complex scenarios with many classes by reaching high spatial accuracies [188–196]. However, large-scale applications were also investigated, which often classify more aggregated classes due to the lower spatial resolution of the input data [58,73,197–201].

Further studies with a specific context focused on coastal [202] and alpine [203] environments, as well as wetlands [204–207], by deriving multi-class LCLU maps. Other LCLU applications applied a binary classification to focus on the presence of a single class. The majority of those studies looked for built-up areas [54,208–211], waterbodies [212–216], shorelines [217,218] or river networks [48,219].

3.4.4. Multi-Class Object Detection

With CNNs it became possible to detect multiple objects of different classes in images, based on their intrinsic features. In remote sensing, two datasets established themselves as baselines and pushed the capabilities of deep-learning-based object detection models. 11% of the reviewed publications investigated multi-class object detection from which 66% used the older NWPU VHR-10 dataset with 10 object classes [39] and 26% the more recent and challenging DOTA dataset with 15 object classes and rotated bounding box annotations [40].

Since objects in remotely sensed imagery appear to be relatively small, not centered in the middle of the image, densely cluttered, partly occluded and with an arbitrary orientation angle [12,21,220,221], object detection on remote sensing images is a challenging task. In order to cope with these challenges, studies trained models to be capable of handling multiple scales [222–227] or focus especially on small objects [228]. Also, the spatial context was taken into account for better detection and classification [229–232] following the idea that not only the target itself contains a distinct signal, but also the typical surrounding in which it appears. Occlusion is particularly vital in overhead images, since near ground objects, which are often targets in multi-class object detection, can be occluded by several layers of obstruction. Clouds, high buildings, steep topography and above-ground vegetation like tree crowns can all together occlude objects. Ren et al. [233] took such examples into account and trained an object detector which is aware of partly occluded objects and through this still able to detect them.

The findings in object detection studies from the transportation sector for ships, cars and aircraft are highly related to multi-class object detection. With the DOTA dataset, orientated bounding boxes became easily available and models were designed to predict the rotation offset jointly with the xy-offset to provide rotated bounding boxes [223,229,234]. These studies are performing better than the former approaches which solely trained on rotated images, to make the model rotation invariant. Finally, instance segmentation was applied in [235] and also rotated bounding boxes for object detection with subsequent instance mask segmentation, resulting in state-of-the-art performance [236].

3.4.5. Agriculture

The agricultural sector was investigated in 10% of the publications. Typically, the classification of multiple crop types has the biggest share with 29%. What is particularly important for crop type classification is the phenology, hence temporal signals must be considered to be well as spatial boundaries. Spatially aware CNNs were mainly discussed until now; however, temporal exploitation is possible with CNNs. Several studies used CNNs in a multi-temporal context to investigate crop type classification [237–245]. Zhou et al. [240] combined two time steps and classified each pixel due to the one-dimensional concatenation of the spectral signals, hence the spatial context was dismissed. Similarly using a spatial neighborhood of 8×8 pixel and a temporal receptive field of three time steps, Ji et al. [239] classified crop types by extracting three-dimensional features. Pelletier et al. [238]

proposed a spectral-temporal guided CNN and Zhong et al. [237] a 1D CNN that classifies EVI time series. In both studies, the employed CNNs were able to outperform baseline models based on random forest and recurrent neural networks.

In general, CNNs and RNNs [246] or specifically LSTMs [247] are two different types of deep-learning models, of which the latter is originally designed to exploit sequential data. Hence, RNNs and LSTMs were also used for crop type mapping or LCLU investigations [248–252]. Nevertheless, CNNs and RNNs or LSTMs do not directly compete. Several studies [253–258] proved that combinations of both can reach higher accuracies than singular models, hence CNNs complement the temporal or sequential perspective of RNNs and LSTMs with a spatial perspective. Since this review focuses on CNNs, such models are not discussed further here. However, we want to highlight their recently increasing importance, especially in applications for crop type mapping and LCLU.

More studies in the agriculture sector applied CNNs to distinguish and outline farmland and smallholder parcels [259–263]. On a much closer perspective, specific crop plants were monitored mainly by UAV-based imagery like corn [264], sorghum [265], strawberries [266], figs [267] or opium poppy [268]. Together with a study that detected livestock [269], they demonstrated new applications of very high-resolution remote sensing data. Single oil palm trees in monocultures were counted [270–272] and rice fields mapped [273–275] on a large scale using data from spaceborne sensors. Jiang et al. [275] used EVI time series to exploit a temporal rather than a spatial signal, in order to distinguish rice fields in a complex landscape. Another single proof-of-concept study segmented marine aquaculture in very high-resolution spaceborne, multispectral imagery at one location of 110 km^2 in the East Chinese Sea [276].

3.4.6. Natural Vegetation

In 4% of the reviewed publications, natural vegetation like forests and near-surface plant societies were studied. Mazza [277] used TanDEM-X SAR data to classify forested areas in a proof-of-concept study. Refs. [278–280] investigated the health of forests and trees, where Safonova et al. [280] looked for damages caused by bark beetle and Hamdi et al. [279] for areas affected by storms. On an individual level, trees and tree crowns were detected in optical imagery [281–284].

Near-surface vegetation like shrubs and weeds were studied in order to monitor the distribution of weeds which lead to an increase in fire severity [285] or to generate maps of specific endangered species [286]. Other studies focused on disentangling species from a complex environment to generate species segmentation maps, giving close detailed information about their spatial distribution and plant communities [287–290].

3.4.7. Natural Hazards

Natural hazards were also investigated by 4% of the publications. A major focus is on damage assessment after natural hazards like earthquakes [291–293], tsunamis [294,295], their combination [296] or wildfires [297]. In four studies flooded areas were derived, two by spaceborne sensors on a larger scale [298,299] and two by UAVs for fast response mapping on a local scale [300,301]. Also, on a local scale with UAVs, slope failures were investigated by Ghorbanzadeh et al. [302]. Due to the needs of fast response analysis on a large-scale, spaceborne data acquisition under all weather conditions is necessary, hence the use of radar data is crucial. The studies of Bai et al. [294] and Li et al. [299] used TerraSAR-X data for post-earthquake-tsunami and immediate flood event investigations respectively, and Zhang et al. [297] used Sentinel-1 data to locate burned areas after wildfires.

3.4.8. Cryosphere

With 2% of all publications, studies which discuss topics of the cryosphere are among the smallest groups. Applications are monitoring calving glacier margins [303], the Antarctic coastline [304] and sea ice [305,306] or detecting permafrost induced structures [307] and segmenting permafrost

degradation [308] as well as deriving ice and snow coverage [309]. Hereby, the publication of Baumhoer et al. [304] is to highlight. This proof-of-concept study develops a deep-learning-based method, which can monitor the Antarctic coastline, by segmentation of Sentinel-1 data in different locations. This CNN-based deep-learning model was recently incorporated into a set of methods to answer geoscientific research questions on a large scale [310]. Such a large-scale application of radar data with deep learning was found to be missing until now in Earth observation [18]. Therewith, these two consecutive studies describe the evolution of deep learning in Earth observation characteristically: from method development to a stronger implementation as a tool for answering large-scale research questions.

3.4.9. Wildlife

Applications using spaceborne, airborne and UAV platforms to acquire very high-resolution data, exploit the capabilities of deep-learning models to detect tiny but for ecosystems highly important features like wildlife. Using data recorded from spaceborne platforms, Guirado et al. [311] were able to detect whales and Bowler et al. [312] albatrosses. Imagery acquired by UAV overflights were used to detect mammals [313] and data sensed by airborne sensors attached to planes to detect seals [314].

3.5. Employed CNN Architectures

The decision on the appropriate architecture for a given research question and resources is of particular importance when optimizing the outcome. This became a challenging task due to the vast amount of deep-learning models, their variations and the fast-moving field. In this section, we provide an overview of the most used and established CNN architectures for image segmentation and object detection as well as feature extractors in Earth observation research. This overview can be used as a starting point in a decision process to build a deep-learning environment.

Since the introduction of AlexNet for image recognition in 2012, the diversity of applications and CNN architectures has greatly increased. A thorough introduction to the evolution and recent trends of CNN architectures is provided in *Part I* [21]. However, in order to give some intuition on the tasks and architectures of CNNs discussed in this section, the main characteristics are briefly summarized at the beginning of each subsection. Also, in the next sections, the single architectures and variations are summarized in groups, in order to provide a better overview. In case that an architecture could not be assigned to a group because its architecture design is too unique and no apparent relation to the major CNN architecture families is discernable, it was assigned to the group *Custom*. Furthermore, the group Other is used to summarize items of architecture designs which appear less than five times.

3.5.1. Convolutional Backbones

A CNN architecture mainly consists of an input layer; the feature extractor also called convolutional backbone; and the head, which performs a task-specific action on the extracted features, like image segmentation or object detection. This section focuses on the convolutional backbone, which subsequently extracts semantically high-level features from the input data and therewith has a strong influence on the performance of a CNN model. During model design, a backbone should be chosen that balances depth, number of parameters, processing power consumption and feature representation. For a detailed overview of convolutional backbones, we refer to Part I, Section 3.1 [21].

As pictured in Figure 8, architectures from the ResNet family and Vintage architectures are the most commonly used feature extractors (67%). The ResNet designs are known for their so-called residual connection, which made training deep networks possible [315]. ResNet models show a good balance in the number of parameters and accuracy, and they are made to be less complex compared to Inception designs [316], and therefore easy to adapt [21]. As the name suggests, the group *Vintage* contains relatively old architectures like the VGG-Net, introduced between 2012 and 2014 [6,317,318]. They are widely used as feature extractors since their early occurrence but compared to ResNet, they

show a lower accuracy while using more parameters [21,156]. Nevertheless, they appear to be among the most used models.

The third largest group of custom models is often similar to *Vintage* designs in their combination of stacked convolution and pooling operations. Still, single designs are especially made to take remote sensing data characteristics into account [38,61,192,241]. It is important to note the small group of six items which use MobileNets [145,268,319–322], of which five were published in 2019. They describe an onset of interest in parameter efficient models with high accuracy and they prove that such models can compete in Earth observation studies.
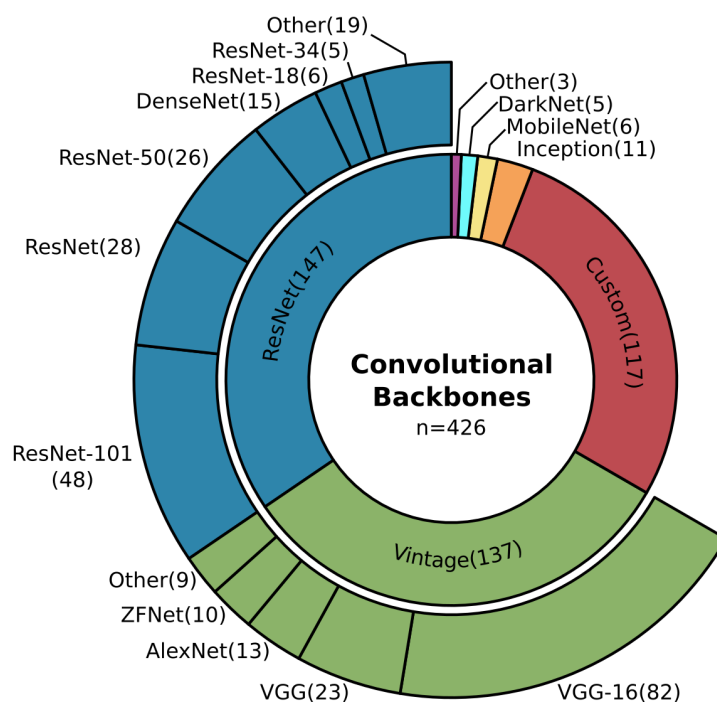

Figure 8. Overview of the commonly employed convolutional backbones or feature extractors in image segmentation and object detection. ResNet (35%) as well as Vintage designs (32%), like the VGG-Net, are the most widely used feature extractors, whereas recently developed parameter efficient architectures like the MobileNets (1%) belong to a minority in Earth observation studies.

A closer look at the specific architectures used within the groups ResNet and Vintage reveals that the depth of the architectures used in Earth observation differs from the depth of the best performing models on benchmark datasets in the field of computer vision, see Part I Section 3.1 [21]. For instance, in computer vision ResNet-152 and VGG-19 perform better on the ImageNet benchmark dataset than their shallower counterparts ResNet-100 and VGG-16 [315,318]. However, the shallower models are more frequently used in Earth observation, see Figure 8. The arguments for shallower models in Earth observation as in computer vision and especially ImageNet are manifold but reveal essential insights about remotely sensed data analyzed by deep learning.

The deeper a model, the more parameters it has. This makes it capable of classifying 1000 classes precisely, as necessary for ImageNet, but that many parameters are not needed for a 2–20 class remote sensing application. It is even more likely that such a number of free parameters could easily overfit the model. In case the Earth observation dataset is too small, many such parameters cannot be sufficiently trained, hence shallower networks which mostly have fewer parameters are in favor [38]. Also, with fewer parameters, the processable tile size of a remote sensing scene increases which minimizes border effects, allows more context and contributes to less noisy results [216]. However, since spatially tiny targets characterize remote sensing data, a deep feature extractor can easily oversee those spatially small details. With the depth of a model, the so-called receptive field grows, and the final

features extracted subsequently are no longer aware of tiny objects and fine-grained borders. Hence to preserve their information, a shallower convolutional backbone is one solution [323,324]. Finally, many studies decide on a shallower model, even when a deeper model gains a higher accuracy since the shallower model has a much better accuracy-parameter ratio and is considered computationally superior [38,118,156,325].

This reasoning reflects the overall usage of shallower architectures in the reviewed studies, where the ResNet models and the VGG models of the Vintage architectures can be considered established convolutional backbones for Earth observation applications.

When discussing convolutional backbones as well as the similarities and differences between Earth observation and computer vision, transfer learning is an important issue to highlight. In transfer learning, the convolutional backbone of a deep-learning model is pre-trained on a dataset which often comes from another domain and then fine-tuned on a smaller dataset of the target domain. The idea is to use a large dataset which initially teaches the model to understand basic feature extraction so that it starts with a better intuition when it first sees data of the target domain. Of the 429 reviewed papers, 38% used a transfer learning approach, of which 63% used the pre-trained weights of the ImageNet dataset. This makes weights pre-trained on ImageNet the most widely used for transfer learning approaches in Earth observation. Even though many examples exist where transfer learned models are performing better than models trained from scratch [157,281,326], this general assumption cannot be made in every case. One example is that models pre-trained on ImageNet are optimized for RGB images. Analyzing more than three channels is not directly possible with weights that were pre-trained on RGB data [304,327]. However, expanding the pre-trained input channels is possible [157], for instance, by doubling the first filter banks of the first convolution, as demonstrated by Bishke et al. [328]. A more critical issue is highlighted in studies, which investigated radar imagery. The speckle and geometrical properties in radar images make them appear very different to natural images. Hence, models pre-trained on ImageNet and fine-tuned on radar imagery perform even worse than the same models trained solely on the radar data [107,205,329]. In conclusion, transfer learning of pre-trained models is an opportunity for better results [114], primarily when optical data is used. In cases where the target dataset strongly differs from the pre-trained dataset, the positive effects can be smaller or even worse than a training approach from scratch [85,107,204,205,329,330].

3.5.2. Image Segmentation

Image segmentation describes the task of pixel wise classification. In order to achieve this, CNNs were first used in the so-called patch-based classification approach, where a convolutional backbone moves with a small input size like 9×9 pixel over the image and classifies the center pixel or the entire patch upon the extracted features. Later in 2014, FCNs (Fully Convolutional Networks) [331] were introduced, which use the entire image as input and after feature extraction, in this context called the encoder, restores input resolution in the decoder, and classifies each pixel afterwards. The result is an input resolution segmentation mask where each pixel is assigned to a single class. Restoration of the input resolution can be achieved using operations like bilinear interpolation which characterizes the group of naïve-decoder models. Another way to do this is upsampling by trainable deconvolutional layers and merging information of the encoder path with the decoder path, called encoder-decoder models. For further explanation and examples, we refer to Part I Section 3.2 [21].

Figure 9 illustrates that of the 261 studies which used CNNs for image segmentation, 62% are encoder-decoder models and of these, 54% can be related to the U-Net design. The dominance of encoder-decoder models in Earth observation can be related to an already mentioned property of remote sensing data: the occurrence of tiny and fine-grained targets. Due to the usage of information from early stages in the encoder during upsampling to restore input image resolution, tiny and spatially accurate features have a much stronger influence on the segmentation as in naïve-decoder models [133]. Since such details are often of significant interest in Earth observation applications, encoder-decoder designs became the most favored models.

The U-Net model, initially designed for biomedical image segmentation [332], gained much attention due to its good performance at the time it was published and its clear, structured design. This made it an intensively researched and modified model [21]. In Earth observation applications, image segmentation must deal with blobby results which are contrary to the intent to segment details and fine-grained class boundaries. In order to overcome this contradiction, atrous convolutions and the effective atrous spatial pyramid pooling module (ASPP) from the DeepLab family [333–336] were integrated into the U-Net in multiple studies [133,169,196,215,221,276,285,309,337–342]. Atrous convolution maintains image resolution during feature extraction, which supports the attention to detail [151,163,202,343,344], where the ASPP module also takes spatial context into account which results in less blobby segmentation masks [145,146,345]. In order to gain better results, the final segmentation masks were further refined using CRFs (Conditional Random Fields) [60,165,212,218,346], as well as multiscale segmentation approaches that fuse features of multiple scales before predicting the per pixel class [146,152,189].
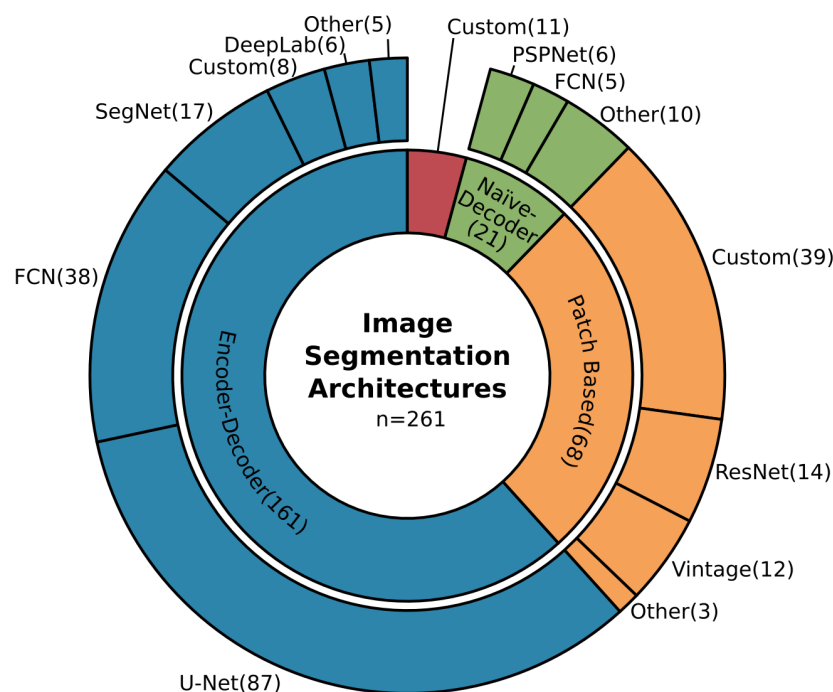

Figure 9. Overview of the commonly employed architectures for image segmentation. With 62% encoder-decoder models are the most frequently used especially the U-Net design (33%). Followed by patch-based approaches (26%) and naïve-decoder models (8%).

Other modifications in architecture design which recently received more attention are the so-called attention modules. They originate from the SE (Squeeze and Excitation) module, which multiplies the output of a convolutional operation by trainable weights [347]. Therewith, some results gain more attention than others. In Earth observation, the modules are often called channel and spatial attention modules: where channel attention modules weight the channels that hold the extracted features globally and spatial attention modules weight areas of those channels spatially [128,143,164,197,219,276,342,348,349]. This technique supports the idea that not all features which are extracted by a neural network affect the results equally.

Similarly, there are gated connections, which incorporate feature selection during the decoding process. Instead of passing raw features from the encoder to the decoder to enrich spatial accuracy, these features are refined by gated skip connections [64,148,350]. Another model modification is used in studies which exploit multi-modal data. Data fusion with CNNs offers many possibilities; in so-called early fusion, the input data is fused before or during the first convolutional operation.

On the other hand, late fusion first extracts features in parallel and fuses them deep in the network structure. Between those options, intermediate fusion and even shuffling can increase feature representation [56,58,140,144,146,152,162,188,195,203,346,351–354]. However, all these modifications in architecture design demonstrate that established CNNs for image segmentation need to be adjusted or even explicitly designed to gain state-of-the-art performance in Earth observation [161]. Nevertheless, their modular designs encourage extensive experiments and optimization.

Finally, in Figure 9 the percentage of patch-based approaches is still high (33%), despite the fact that this technique was one of the first approaches used to segment images and since the emergence of FCN in 2014, was found to be inferior. This is true for high spatial resolution images with rich feature information. Nevertheless, medium to low spatial resolution imagery, which is used in large-scale Earth observation applications like Landsat data, does not necessarily show the kind of spatial feature richness that modern CNN architectures are designed for [186,187,195,199,202,206,211,239,274,355,356]. Hence, patch-based approaches are still a model of choice for this kind of data. This statement is supported by the share of image segmentation studies with a lower spatial resolution above five meters, which is more than twice as high for patch-based models (38%) as in all other approaches (17%), where higher resolution dominates.

3.5.3. Object Detection

In object detection, instances of objects are detected in an image, commonly using a bounding box to describe their position and extent. Two major approaches exist which apply CNNs: two-stage and one-stage detectors. Two-stage detectors first predict regions of interest which contain most probable, class agnostic object candidates and perform adjacent bounding box regression and object classification. One-stage detectors perform object localization, classification and bounding box regression in a single shot. The latter commonly perform faster but are less accurate. A detailed introduction to state-of-the-art CNN-based object detection algorithms is provided in Part I, Section 3.3 [21].

Figure 10 shows that in Earth observation applications on object detection, the two-stage approach dominates with 63%. Thereof, the well-established R-CNN models [357] are the most used and of those, 73% belong to the Faster R-CNN design [358].

Like the U-Net architecture in image segmentation, R-CNN models are known for their modularity, which makes the model a good starting point for customization. Another argument for using two-stage approaches is better accuracy, even when their computational consumption is relatively high. One-stage detectors were also used and modified to better fit the needs of Earth observation applications, especially in studies where processing speed is critical or hardware limited [82,292,321,359,360]. The SSD [361] and YOLO [362–364] models are among the most widely used, which can be related to their good performance and extensive documentation, experiments and advancements in computer vision.

Similar to the challenges faced in image segmentation, tiny objects, which can be densely cluttered and with an arbitrary rotation, are the main challenges in object detection. In Section 3.4, it was already reported that predicting rotated bounding boxes contribute to better results for specific applications. 73% of the models capable of predicting the angle offset belong to two-stage detectors and 27% to one-stage detectors, which demonstrates an advantage of the modular two-stage architectures for model modifications. However, the architecture modification to predict rotated bounding boxes is one among many to tackle the challenging characteristics of remotely sensed data.

The feature pyramid network (FPN) [365] module is commonly known to enrich extracted features and pass them to the detector on multiple scales. FPN and resembling techniques were leveraged in several studies, which also proposed further modifications of this particular structure [65,80,95,100,124,125,225,227,232,235,236,325,366–375]. Other positive effects on detecting objects in remotely sensed images were found in modifications which compensate for the spatial information loss that happens during feature extraction [228,376]. Similar to image segmentation, a widely used modification is the use of atrous convolutions to maintain high

resolution during feature extraction [106,227,232,326,369,377–379]. In order to shift the model architecture to concentrate on important features, attention modules were incorporated which makes them a widely used architecture modification in both object detection and image segmentation [81,94,96,125,225,230,360,367,375,380,381].
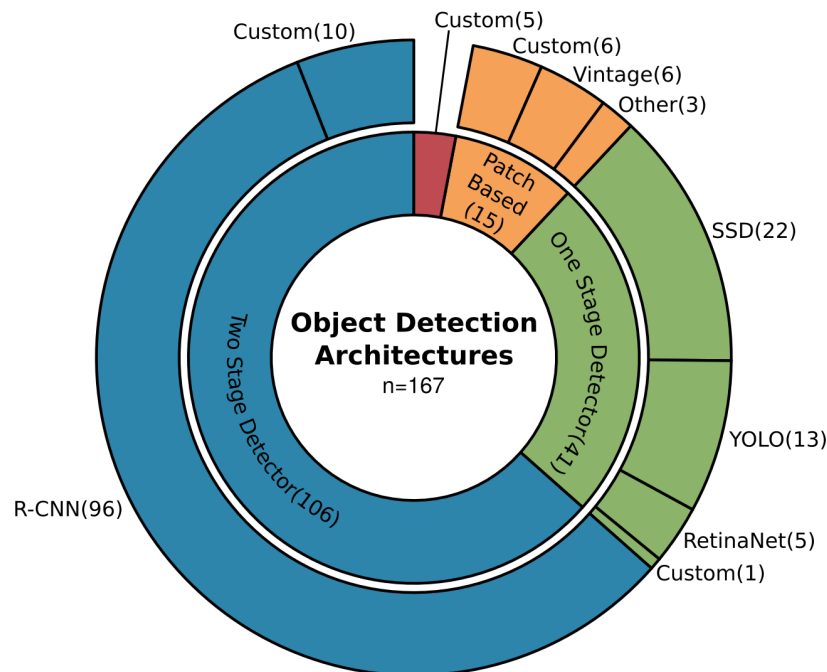

Figure 10. Overview of the commonly employed architectures for object detection. With 63% two-stage detector models are the most frequently used especially designs from the R-CNN family (57%). Followed by one-stage detector models (25%) and patch-based approaches (9%).

Further object detection specific modifications are the applications of deformable convolutions to better extract object related features [232,233,372,377,382,383], as well as to specifically take the spatial context of objects into account [78,106,125,230–232,236,366,376]. In order to better exploit complicated training examples, hard example mining is used to focus on challenging situations [117,119,326,384]. As in computer vision, the cascading design of the Faster R-CNN model [385] proved that adaptive intersection over union (IoU see Part I, Section 3.2 [21]) thresholds are an efficient way for better performance, especially on tiny objects [232,377,386]. Overall, modifications of architectures highly depend on the size, distribution and characteristics of the target defined by the application as well as the used data. For two-stage detectors, Faster R-CNN [358] and its variants of cascading models [385] and Mask R-CNN [387] are the most established and promising designs in Earth observation.

3.6. Deep-Learning Frameworks

As shown in the previous sections, the transfer of deep-learning methods from computer vision to Earth observation is covered by a large share of studies. The majority have a strong focus on method development in which they modify deep-learning architectures, cost functions and optimize model training. However, recently the share of studies which apply deep-learning methods to answer specific geoscientific research questions is getting bigger. Both groups have special requirements on deep-learning frameworks: studies with a focus on method development need to dive deep into model structures and demand flexibility and freedom in customization, whereas studies focusing on application tend to use stable, established and streamlined model building tools.

Different frameworks are known to offer more freedom or more accessible functionality. Pytorch [388] and native TensorFlow [389] offer deep and complex customization, whereas the Keras

API [390] is known for its intuitive handling. Interestingly, with recent studies focusing on applications, Keras is receiving much more attention, as pictured in Figure 11. This indicates a user community which prefers established functionalities to use deep learning as a tool, but still wants to have the opportunity to customize. Still, the share of Pytorch and TensorFlow remains high, which points to users who need more freedom to customize their models or optimize their data pipeline.
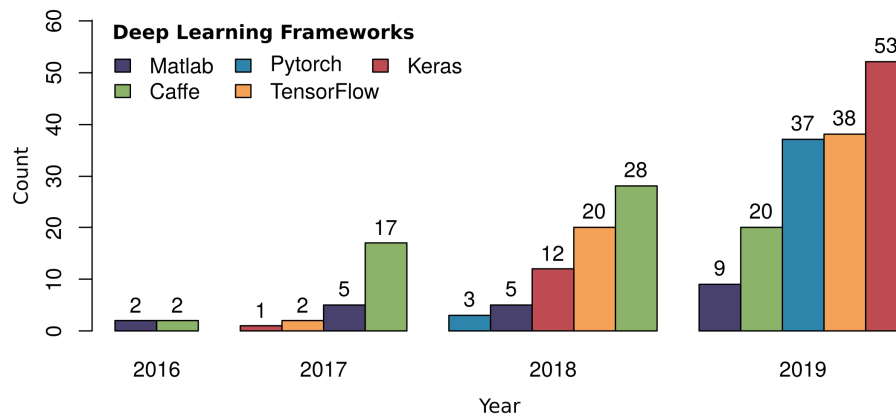

Figure 11. Temporal development of employed deep-learning frameworks by numbers of studies which reported the chosen framework. Of the 66 Keras-based studies, 47 reported the underlying backend, of which 44 used TensorFlow. This number is not included in the TensorFlow count, presented in the figure.

The usage of the early Caffe framework [391], which at some point forked and one branch eventually became Pytorch, decreases. In contrast, the number of studies which use Pytorch increases, indicating that parts of the Caffe user community turned to Pytorch. All the frameworks pictured in Figure 11 are fully open source besides the Matlab-based MatConvNet [392], which shows a comparably smaller share than the open access frameworks. Overall, the increase in deep learning as a tool to answer geoscientific research questions can be seen in the increase of users who select open access, well documented and subsequently more stable frameworks to apply deep learning.

4. Discussion and Future Prospects

4.1. Discussion of the Review Results

The results presented in the previous sections and summarized in Figure 12 demonstrate that deep learning with CNNs in Earth observation is in an advanced transition phase from computer vision. The main study focus of the reviewed publications was on method development and proof-of-concept studies, which are characterized by using existing datasets or datasets on a smaller spatial scale. These studies are essential to establish and prove the capabilities and limitations of CNNs in Earth observation research. However, recent studies have begun to investigate large-scale applications with CNNs and incorporate CNNs as a major tool in large-scale geoscientific research.

In the following, the reviewed aspects of employed sensor types, datasets, architectures and applications are discussed. The most used sensor types are those which provide optical data with a high to very high spatial resolution. The dominance of RGB data with rich spatial features can be related to the origin of deep learning with CNNs in computer vision, where also rich feature RGB input data drives the development of CNN research. The increasing use of more Earth observation specific data from multispectral and radar sensors shows the progress made in adapting CNNs for remote sensing data. However, for all sensor types, a high spatial resolution is proven to be crucial for extracting objects and fine-grained class boundaries, which can better be detected when strong representational features can be extracted from the input data. When following the motivation of CNNs

in Earth observation towards an inventory of things in order to analyze object dynamics, high spatial resolution data is and will be the main focus of employed sensor types in Earth observation.
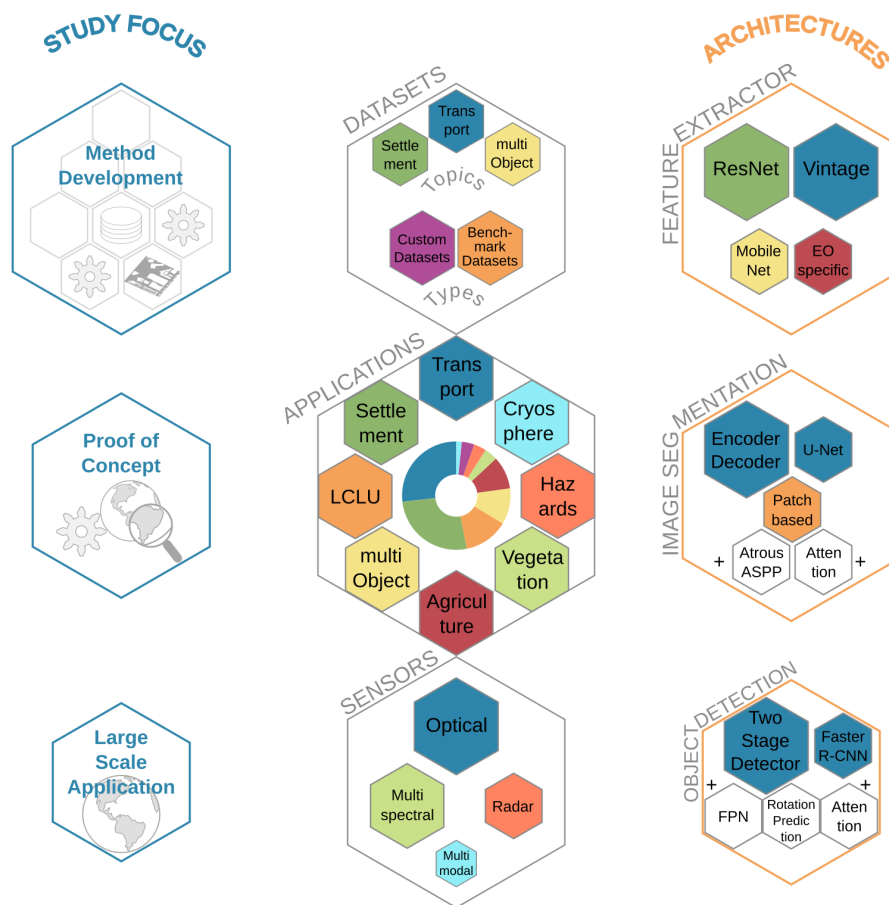

Figure 12. Visual summary of the major findings of this review, separated into study focus, datasets, applications, sensors and used architectures, from left to right. The size of the hexagons represents the importance of the labeled topic, except for applications, where the pie chart depicts the shares of the topics.

From the open datasets in Earth observation, the most widely used cover a small range of application topics: building footprints, car, ship and multi-object detection, as well as road network extraction. Even when datasets of further topics exist, and their variation is increasing recently, the majority of the reviewed publications investigate datasets which contain targets of the above-mentioned applications. Hereby, they made progress in extracting and answering research questions of the specific application targets. However, primarily they used such datasets to find solutions on how to optimize CNNs coming from the computer vision domain for analyzing Earth observation data. This strong data-driven method development in the early phase of the transition of CNNs from computer vision to Earth observation is subsequently complemented by studies which created custom datasets. Those studies indicate the recent trend of dataset creation which is driven by research questions of land-surface dynamics. Accompanying this observation in datasets is the shift of the study focus from method development to proof of concept and finally, large-scale studies, which we characterize as the advanced transition phase.

The methodological insights which were made during this transition phase are prominent in the employed architectures and their modifications. For feature extraction, architectures form the ResNet family and Vintage designs like VGG are the most commonly used. Therefore, in Earth observation shallower model variants are chosen, compared to computer vision. The most frequent arguments for

shallower models were to avoid overfitting or vanishing of tiny features through large receptive fields as well as to reach a better accuracy-parameter ratio. In image segmentation, the encoder-decoder design and the U-Net is the most widely used. The usage of this specific model is justified by better spatial refinement during the decoding process of fine-grained details by sharing entire feature maps from the encoder with the decoder. The gain in spatial accuracy due to this approach was found to be beneficial for Earth observation applications where tiny and fine-grained structures are important. However, where such fine-grained information is not prominent in the data due to a lower spatial resolution, patch-based approaches are also applied. For object detection, the highly modular R-CNN and especially the Faster R-CNN design of the two-stage detector approach dominates. The modular design allows for intensive modifications, and follow up architectures like Cascade R-CNN and Mask R-CNN show promising results in more precise predictions for deriving instance segmentation, which will be necessary when analyzing spatio-temporal object dynamics. Overall, architectures were intensively modified to better fit the needs of Earth observation studies. Here, attention modules are often found to be effective in both image segmentation and object detection. The methodological basis, which was developed during the phase of transition from computer vision to Earth observation, offers the opportunity for fast and successful architecture modification in future studies.

When finally looking at the applications where CNNs were used, the picture is similar to the most commonly used datasets: transportation, settlement and multi-object detection are among the four largest, together accounting for 66% of all studies. Building footprints and road extraction as well as car, ship and multi-object detection belong to the most intensively researched topics. With general LCLU mapping, a traditional Earth observation research topic is the third largest group with 13% of all studies. The similarity between the topics of the most used open datasets, as discussed above, and the investigated applications of all reviewed studies indicate the data-driven paradigm of the transition phase. The optimization of specific problems on available datasets is used to push method development and establish CNNs in Earth observation before answering geoscientific research questions. Furthermore, the three large application groups (transportation, settlement and multi-class object detection) have a strong focus on relatively small but numerous artificial objects in one image. This large number of studies is also attributable to the strength of CNNs to identify the characteristic features of entities in imagery data and localize them precisely. Hence, CNNs are exceptionally well suited to answer questions related to these application sectors. On the other hand, applications in LCLU, where CNNs also perform successfully, typically cope with more indifferent spatial boundaries. Furthermore, spectral and temporal signals are also necessary, hence more complex model modifications to CNNs, or combination with other model types, and expert knowledge are needed. Also, few datasets exist in this domain which are suited to deep learning. Therefore, there are fewer studies in this area compared to applications where deep learning with CNNs is traditionally more established.

However, with the subsequently proven capabilities of CNNs in Earth observation research, studies are recently starting to use CNNs on applications which generally rely on other methods like random forest. In addition to that new applications are being investigated, most with a focus on analyzing entities with an object detection approach, which widen the field in which Earth observation can be applied.

4.2. Future Prospects

With the results of the review summarized and discussed in the last section, with regards to study focus, sensor types, datasets, architectures and applications, this section provides future prospects to these aspects. Figure 13 pictures a graphical summary of upcoming trends of CNNs in Earth observation and their potential drivers.

The ongoing transition from computer vision to Earth observation will continue and change the foci of studies. This will lead to a decrease in proof-of-concept studies and at the same time, to an increase in studies which answer large-scale geoscientific research questions using CNNs.

Especially large-scale datasets which provide a global variance of training examples will lead to a better spatial transferability and thus push large-scale studies forward. Method development will remain a challenging and acutely investigated topic, due to the fast-moving developments coming from computer vision and the possibilities to modify CNN architectures in general. Also, method development will play a significant role in making CNNs spatially transferable and less likely to overfit when more global datasets become publically available.
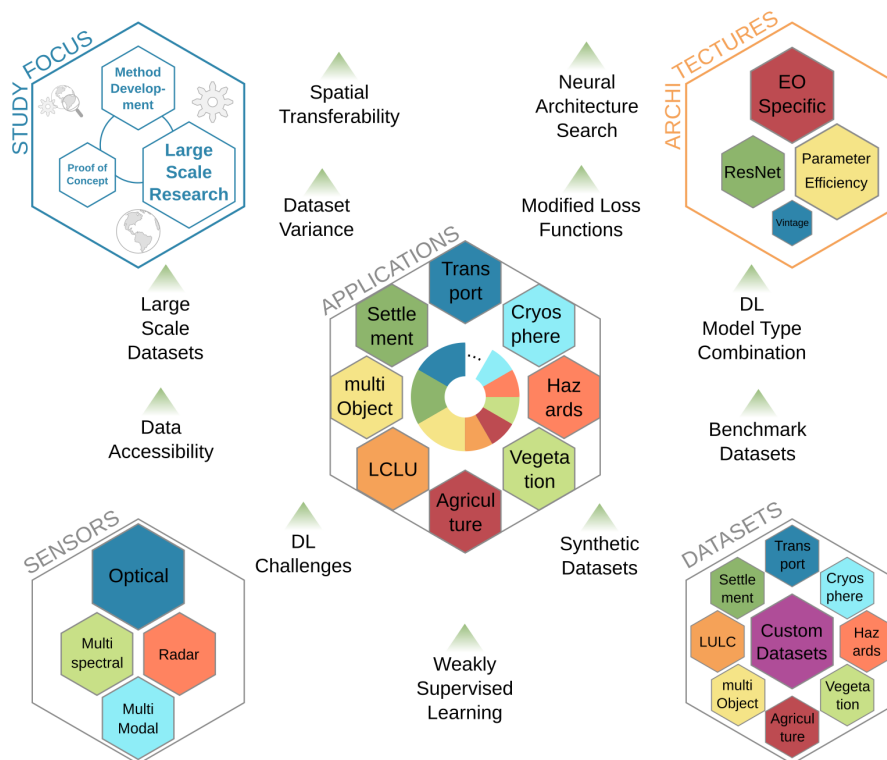

Figure 13. Visual summary of future prospects, separated into study focus, datasets, applications, sensors and used architectures. The size of the hexagons represents the importance of the labeled topic, except for applications, where the pie chart depicts the relevance of a topic. In addition, the drivers of future prospects are depicted with a green triangle. They are arranged in such a way that their proximity to the thematic fields roughly reflects their influence on the five aspects.

Optical imagery sensors with a high to very high spatial resolution will remain the most employed data sources for Earth observation with CNNs. Such data provides a rich feature depth for multiple objects at the same time and is therefore an excellent choice to detect and segment entities of multiple classes jointly in one image. Furthermore, data accessibility will increase due to the opening of archives containing previously sensed optical images with a very high spatial resolution, as well as future spaceborne missions. The latter will increase not only spatial resolution but also temporal resolution, and also provide high-resolution video footage sensed from space. However, multispectral and radar sensors, as well as their combination, will be more employed based on recent findings. CNN models became more adapted or were specially designed to deal with such input data, and insights on their behavior in deep-learning approaches were made, as reported in this review.

Another driver for increasing the attention for specific sensors are deep-learning challenges, where teams can compete by solving specific topics. The recently ended SpaceNet 6 challenge on building footprints extracted from multispectral and SAR data [33] is an example for the potential of such events and how to use them to push methods which analyze multi-sensor data. The multi-sensor focus of SpaceNet 6 as well as the recently published dataset SEN12MS [393] are characteristic for

pushing deep-learning research and are an indicator for research interest in focusing on sensors other than optical systems.

Datasets are one of the main drivers of deep-learning development, which was widely reported and discussed in this review and in *Part I* [21]. Hence more and larger datasets will be created which can be used for benchmarking. Nevertheless, the creation of datasets and particularly large and benchmark datasets is labor-intensive. In order to encourage researchers to create custom datasets, alternatives should be provided besides creating datasets just by labeling images by hand. Weakly supervised learning has already been mentioned as an opportunity to leverage fuzzy labeled data [10–12] and activate data sources unused until now [69]. Another way to create datasets, which was only used once in the reviewed papers, is to synthesize data. The authors Isikdogan et al. [48] and recently Kong et al. [49] demonstrated that using fully synthetically created training data in Earth observation is possible and could further increase the variety of datasets. Overall, with an increasing variety of datasets, researchers will be able to analyze multiple parameters of land-surface dynamics better and combine them into deep-learning workflows which help to answer complex research questions.

The modification of architectures which come from computer vision is necessary to analyze data in Earth observation successfully. Furthermore, custom model designs, which are tailored to the needs of Earth observation data, are becoming more popular. However, parameter efficiency is important in both cases to save computing power and avoid overfitting. By choosing shallower model variants of designs from the *ResNet Family* or *Vintage designs*, the number of parameters has already been taken into account. However, with the emergence of MobileNets [394,395] and neural architecture search [396–398] new possibilities are emerging to create parameter efficient designs [399] and architectures particularly built for Earth observation problems. Efficient designs will be of greater interest to increase the effectiveness of feature extractors. The application of neural architecture search was already part of the discussion of *Part I* [21]. In the meantime, Wang et al. [400] proposed RSNet, a network designed for remote sensing data by leveraging neural architecture search. This study demonstrates that neural architecture search, which increased the accuracy-parameter ratio in computer vision significantly, can also be used to create specific architectures for Earth observation.

Even when CNNs are the most widely used deep-learning model types in Earth observation, applications with a multi-temporal perspective are particularly sparse, as reported in this review. Hence, the combination with other model types, like LSTMs which support the representational capacity for sequential signals, see Section 3.4.5, will lead to an increase in applications. This will be of increasing interest not just to better derive information with a temporal signature, but also to analyze object dynamics and make future predictions.

Another driver aims more towards the training of neural networks and is not necessarily limited to CNNs. Loss functions, which assess the model's performance during training and provide the signal on which the model is optimized, offer the possibility to incorporate expert knowledge into the network. An example included in this review is the work of Wei et al. [130] who embedded geometrical structures into the loss function for better road extraction. By adding already existing knowledge from an expert's perspective to predict complex systems upon remotely sensed data, new research applications can be exploited.

All drivers will finally impact the applications of CNNs in Earth observation. More applications will be investigated, and new applications added which today have never been in focus from an Earth observation perspective since, with deep learning, more detailed information is accessible. However, it is to be expected that specific applications will further dominate, due to their interest from research and practical applications. Those applications will be in sectors where humans interact and operate economically, like transportation and settlements. When analyzing object dynamics, the livelihoods of our planet can be closely monitored and quantified. This will not just provide valid data for research and political decision-makers, but also to everyday economic managers who decide about short term strategies on a regional and global scale. Therefore, a broad inventory of things is essential to compare current situations with the past, and to predict and manage in near-real time.

5. Conclusions

In this review, we provide an extensive overview of the convolutional neural network (CNN) for image segmentation and object detection in Earth observation research by analyzing 429 publications. The main findings of this review highlight the fact that CNNs are in an advanced transition from computer vision to Earth observation. This transition is characterized by data-driven method development, which provides adapted models and small-scale proof-of-concept studies. Together they build a basis for large-scale geoscientific research questions which use CNNs. This trend will continue and finally provide a global, digital inventory of things derived by CNNs from Earth observation data, to analyze object dynamics and their impact on land surfaces.

In detail, the 429 publications were analyzed with regard to five major aspects: (1) the spatial distribution of the studies globally, (2) the employed sensor types, (3) the most commonly used open datasets, (4) the specific applications in which CNNs are used in Earth observation research, and (5) CNN architectures and modifications to fit the needs of remotely sensed data. The results are:

1. The study site locations are mainly from three continents: Asia (21%), Europe (17%), and America (14%), whereas, studies with a global perspective (4%) or a focus on polar regions (3%) and Africa (1%) have the smallest shares. The largest shares of national study sites are from China (14%), Germany (10%), and the US (9%). Studies with multiple locations (31%) or without any specification (8%) are mainly investigating method development or proof of concepts. Also 83% of the German study sites are in Potsdam and Vaihingen, from which datasets are available for method development and ablation studies.

2. The most employed sensor systems are optical sensors (56%) which provide a high to very high spatial resolution. They are followed by multispectral (26%) and radar (13%) sensors. Only 4% of the studies employ multi-sensor data. Imagery data of optical sensors with a high spatial resolution are often acquired via Google Earth. More directly, the most commonly investigated spaceborne missions with optical and multispectral sensors are Gaofen 1 + 2 and WorldView 1–4; and Sentinel-1 and TerraSAR-X for radar systems. This shows the importance of high to very high spatial resolution data, and in case of Sentinel-1 of freely available data archives.

3. Datasets are highly important for the development of deep-learning algorithms and are strong drivers for specific applications when they are publicly available. Custom datasets, which were used solely or combined with existing open datasets, were investigated in 62% of the studies. Publicly available, open datasets are prominent in the settlement and transportation sectors. Here, a specific focus is on building footprints, road extraction or car and ship detection. They are by far the most frequently used open datasets. Important datasets for method development in image segmentation are the well-known ISPRS Potsdam and Vaihingen datasets; and for object detection the NWPU VHR-10 and DOTA dataset.

4. Applications in Earth observation in which CNNs are widely used are: transportation (27%) and settlement (26%), as well as the strong method development related multi-class object detection (11%). Here, ship detection (12%), as well as building footprint and urban VHR feature extraction with both 10%, are among the most deeply studied specific applications and therewith demonstrate a focus on detecting entities in remote sensing data. Classical Earth observation domains like LCLU (13%), agriculture (10%) and natural vegetation (4%) are less frequently studied. Still, proof-of-concept studies show how research questions of these domains can be answered by analyzing many single entities and their impact of the wider land cover class they belong to. With a focus on the extraction and detection of fine-grained boundaries and entities, Earth observation with CNNs will be able to quantify object dynamics on a large scale. This will increase the interest from everyday applications for short term decision making and management in economy and practice. Hence, application domains which are characterized by artificial objects will continue to be the most investigated group.

5. In both image segmentation and object detection, CNN architectures for feature extraction are dominated by designs related to the ResNet family (35%) and the older VGG (32%) architecture. Lately, efficient designs like the MobileNets (1%) were also successfully employed. In image segmentation, encoder-decoder designs (62%) are used, especially the U-Net model (33%). They are followed by patch-based approaches (33%) for data with a lower spatial resolution. In object detection, the two-stage detector approach (63%) is the most widely used, and of these approaches, the R-CNN family with the Faster R-CNN model (57%) is the most prevalent. Commonly made modifications to adapt CNNs to Earth observation tackle tiny objects and fine-grained boundary class problems by using attention modules, atrous convolution and rotated bounding boxes.

With this *Part II* of the review, we continued to provide thorough insights into deep learning with CNNs in Earth observation, with a special focus on researchers who want to use this method. Deep learning models and CNNs in particular are an established tool in Earth observation, whose relevance will further continue to increase. Driven by a strong method development and numerous proof-of-concept studies, large-scale geoscientific research studies will provide novel answers. They will investigate the interconnections between small and numerous objects and large-scale land covers to understand land-surface dynamics on a new level. Due to the opening of archives which contain data with a very high spatial resolution, we propose a surge of studies with a focus on object dynamics which will extensively employ the CNN deep-learning model.

Appendix A. Table of Reviewed Publications

The digital version of the table `Appendix_review_PII_DLinEO.csv` can be found here: https://github.com/thho/DLinEO_review.

References

1. Marconcini, M.; Metz-Marconcini, A.; Üreyen, S.; Palacios-Lopez, D.; Hanke, W.; Bachofer, F.; Zeidler, J.; Esch, T.; Gorelick, N.; Kakarla, A.; et al. Outlining where humans live, the World Settlement Footprint 2015. *Sci. Data* **2020**, *7*, 1–14. [CrossRef] [PubMed]
2. Zhu, Z.; Bi, J.; Pan, Y.; Ganguly, S.; Anav, A.; Xu, L.; Samanta, A.; Piao, S.; Nemani, R.R.; Myneni, R.B. Global Data Sets of Vegetation Leaf Area Index (LAI)3g and Fraction of Photosynthetically Active Radiation (FPAR)3g Derived from Global Inventory Modeling and Mapping Studies (GIMMS) Normalized Difference Vegetation Index (NDVI3g) for the Period 1981 to 2011. *Remote Sens.* **2013**, *5*, 927–948. [CrossRef]
3. Klein, I.; Gessner, U.; Dietz, A.J.; Kuenzer, C. Global WaterPack–A 250 m resolution dataset revealing the daily dynamics of global inland water bodies. *Remote Sens. Environ.* **2017**, *198*, 345–362. [CrossRef]
4. Reichstein, M.; Camps-Valls, G.; Stevens, B.; Jung, M.; Denzler, J.; Carvalhais, N.; Prabhat. Deep learning and process understanding for data-driven Earth system science. *Nature* **2019**, *566*, 195–204. [CrossRef]
5. Long, Y.; Xia, G.S.; Li, S.; Yang, W.; Yang, M.Y.; Zhu, X.X.; Zhang, L.; Li, D. DiRS: On Creating Benchmark Datasets for Remote Sensing Image Interpretation. *arXiv* **2020**, arXiv:cs.CV/2006.12485.
6. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. In *Advances in Neural Information Processing Systems 25*; Pereira, F., Burges, C.J.C., Bottou, L., Weinberger, K.Q., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2012; pp. 1097–1105.

7. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. ImageNet Large Scale Visual Recognition Challenge. *Int. J. Comput. Vis. (Ijcv)* **2015**, *115*, 211–252. [CrossRef]

8. Vinuesa, R.; Azizpour, H.; Leite, I.; Balaam, M.; Dignum, V.; Domisch, S.; Felländer, A.; Langhans, S.D.; Tegmark, M.; Nerini, F.F. The role of artificial intelligence in achieving the Sustainable Development Goals. *Nat. Commun.* **2020**, *11*, 1–10. [CrossRef]

9. ESA. Copernicus Masters. ESA Digital Twin Earth Challenge. Available online: https://copernicus-masters.com/prize/esa-challenge/ (accessed on 27 July 2020).

10. Ball, J.E.; Anderson, D.T.; Chan, C.S. Comprehensive survey of deep learning in remote sensing: Theories, tools, and challenges for the community. *J. Appl. Remote Sens.* **2017**, *11*, 1–54. [CrossRef]

11. Zhu, X.X.; Tuia, D.; Mou, L.; Xia, G.; Zhang, L.; Xu, F.; Fraundorfer, F. Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources. *IEEE Geosci. Remote Sens. Mag.* **2017**, *5*, 8–36. [CrossRef]

12. Zhang, L.; Zhang, L.; Du, B. Deep Learning for Remote Sensing Data: A Technical Tutorial on the State of the Art. *IEEE Geosci. Remote Sens. Mag.* **2016**, *4*, 22–40. [CrossRef]

13. Ma, L.; Liu, Y.; Zhang, X.; Ye, Y.; Yin, G.; Johnson, B.A. Deep learning in remote sensing applications: A meta-analysis and review. *Isprs J. Photogramm. Remote Sens.* **2019**, *152*, 166–177. [CrossRef]

14. Tsagkatakis, G.; Aidini, A.; Fotiadou, K.; Giannopoulos, M.; Pentari, A.; Tsakalides, P. Survey of Deep-Learning Approaches for Remote Sensing Observation Enhancement. *Sensors* **2019**, *19*, 3929. [CrossRef] [PubMed]

15. Petersson, H.; Gustafsson, D.; Bergstrom, D. Hyperspectral image analysis using deep learning—A review. In Proceedings of the 2016 Sixth International Conference on Image Processing Theory, Tools and Applications (IPTA), Oulu, Finland, 12–15 December 2016; pp. 1–6.

16. Audebert, N.; Le Saux, B.; Lefevre, S. Deep Learning for Classification of Hyperspectral Data: A Comparative Review. *IEEE Geosci. Remote Sens. Mag.* **2019**, *7*, 159–173. [CrossRef]

17. Paoletti, M.; Haut, J.; Plaza, J.; Plaza, A. Deep learning classifiers for hyperspectral imaging: A review. *Isprs J. Photogramm. Remote Sens.* **2019**, *158*, 279–317. [CrossRef]

18. Zhu, X.X.; Montazeri, S.; Ali, M.; Hua, Y.; Wang, Y.; Mou, L.; Shi, Y.; Xu, F.; Bamler, R. Deep Learning Meets SAR. *arXiv* **2020**, arXiv:eess.IV/2006.10027.

19. Khelifi, L.; Mignotte, M. Deep Learning for Change Detection in Remote Sensing Images: Comprehensive Review and Meta-Analysis. *arXiv* **2020**, arXiv:cs.CV/2006.05612.

20. LeCun, Y.; Bengio, Y.; Hinton, G. Deep Learning. *Nature* **2015**, *521*, 436–444. [CrossRef]

21. Hoeser, T.; Kuenzer, C. Object Detection and Image Segmentation with Deep Learning on Earth Observation Data: A Review-Part I: Evolution and Recent Trends. *Remote Sens.* **2020**, *12*, 1667. [CrossRef]

22. ISPRS. 2D Semantic Labeling Challenge. Available online: http://www2.isprs.org/commissions/comm3/wg4/semantic-labeling.html (accessed on 28 July 2020).

23. Everingham, M.; Van Gool, L.; Williams, C.K.I.; Winn, J.; Zisserman, A. The Pascal Visual Object Classes (VOC) Challenge. *Int. J. Comput. Vis.* **2010**, *88*, 303–338. [CrossRef]

24. Everingham, M.; Eslami, S.M.; Gool, L.; Williams, C.K.; Winn, J.; Zisserman, A. The Pascal Visual Object Classes Challenge: A Retrospective. *Int. J. Comput. Vis.* **2015**, *111*, 98–136. [CrossRef]

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

26. Cordts, M.; Omran, M.; Ramos, S.; Scharwächter, T.; Enzweiler, M.; Benenson, R.; Franke, U.; Roth, S.; Schiele, B. The Cityscapes Dataset. In Proceedings of the CVPR Workshop on the Future of Datasets in Vision, Boston, MA, USA, 7–12 June 2015; Volume 2.

27. Cordts, M.; Omran, M.; Ramos, S.; Rehfeld, T.; Enzweiler, M.; Benenson, R.; Franke, U.; Roth, S.; Schiele, B. The Cityscapes Dataset for Semantic Urban Scene Understanding. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 3213–3223.

28. IEEE GRSS. GRSS Data Fusion Contest. Available online: http://www.grss-ieee.org/community/technical-committees/data-fusion./2015-ieee-grss-data-fusion-contest/ (accessed on 28 July 2020).

29. Mnih, V. Machine Learning for Aerial Image Labeling. Ph.D. Thesis, University of Toronto, Toronto, ON, Canada, 2013.

30. SpaceNet. SpaceNet 1: Building Detection v1. Available online: https://github.com/SpaceNetChallenge/BuildingDetectors (accessed on 1 April 2020).

31. SpaceNet. SpaceNet 2: Building Detection v2. Available online: https://github.com/SpaceNetChallenge/BuildingDetectors_Round2 (accessed on 1 April 2020).

32. SpaceNet. SpaceNet 4: Off-Nadir Buildings. Available online: https://github.com/SpaceNetChallenge/SpaceNet_Optimized_Routing_Solutions (accessed on 1 April 2020).

33. Shermeyer, J.; Hogan, D.; Brown, J.; Etten, A.V.; Weir, N.; Pacifici, F.; Haensch, R.; Bastidas, A.; Soenen, S.; Bacastow, T.; Lewis, R. SpaceNet 6: Multi-Sensor All Weather Mapping Dataset. *arXiv* **2020**, arXiv:eess.IV/2004.06500.

34. Ji, S.; Wei, S.; Lu, M. Fully Convolutional Networks for Multisource Building Extraction From an Open Aerial and Satellite Imagery Data Set. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 574–586. [CrossRef]

35. Demir, I.; Koperski, K.; Lindenbaum, D.; Pang, G.; Huang, J.; Basu, S.; Hughes, F.; Tuia, D.; Raskar, R. DeepGlobe 2018: A Challenge to Parse the Earth Through Satellite Images. In Proceedings of the The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Salt Lake City, UT, USA, 18–23 June 2018; pp. 172–181.

36. SpaceNet. SpaceNet 3: Road Network Detection. Available online: https://github.com/SpaceNetChallenge/RoadDetector (accessed on 1 April 2020).

37. Etten, A.V. City-Scale Road Extraction from Satellite Imagery v2: Road Speeds and Travel Times. In Proceedings of the The IEEE Winter Conference on Applications of Computer Vision (WACV), Snowmass Village, CO, USA, 1–5 March 2020; pp. 1786–1795.

38. Cheng, G.; Wang, Y.; Xu, S.; Wang, H.; Xiang, S.; Pan, C. Automatic Road Detection and Centerline Extraction via Cascaded End-to-End Convolutional Neural Network. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 3322–3337. [CrossRef]

39. Cheng, G.; Han, J.; Zhou, P.; Guo, L. Multi-class geospatial object detection and geographic image classification based on collection of part detectors. *Isprs J. Photogramm. Remote Sens.* **2014**, *98*, 119–132. [CrossRef]

40. Xia, G.S.; Bai, X.; Ding, J.; Zhu, Z.; Belongie, S.; Luo, J.; Datcu, M.; Pelillo, M.; Zhang, L. DOTA: A Large-Scale Dataset for Object Detection in Aerial Images. In Proceedings of the The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 3974–3983.

41. Long, Y.; Gong, Y.; Xiao, Z.; Liu, Q. Accurate Object Localization in Remote Sensing Images Based on Convolutional Neural Networks. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 2486–2498. [CrossRef]

42. Li, J.; Qu, C.; Shao, J. Ship detection in SAR images based on an improved faster R-CNN. In Proceedings of the 2017 SAR in Big Data Era: Models, Methods and Applications (BIGSARDATA), Beijing, China, 13–14 November 2017; pp. 1–6.

43. Huang, L.; Liu, B.; Li, B.; Guo, W.; Yu, W.; Zhang, Z.; Yu, W. OpenSARShip: A Dataset Dedicated to Sentinel-1 Ship Interpretation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 195–208. [CrossRef]

44. Li, B.; Liu, B.; Huang, L.; Guo, W.; Zhang, Z.; Yu, W. OpenSARShip 2.0: A large-volume dataset for deeper interpretation of ship targets in Sentinel-1 imagery. In Proceedings of the 2017 SAR in Big Data Era: Models, Methods and Applications (BIGSARDATA), Beijing, China, 13–14 November 2017; pp. 1–5.

45. Liu, K.; Mattyus, G. Fast Multiclass Vehicle Detection on Aerial Images. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 1938–1942.

46. Razakarivony, S.; Jurie, F. Vehicle detection in aerial imagery: A small target detection benchmark. *J. Vis. Commun. Image Represent.* **2016**, *34*, 187–203. [CrossRef]

47. Mou, L.; Zhu, X.X. Vehicle Instance Segmentation From Aerial Image and Video Using a Multitask Learning Residual Fully Convolutional Network. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 6699–6711. [CrossRef]

48. Isikdogan, F.; Bovik, A.; Passalacqua, P. Learning a River Network Extractor Using an Adaptive Loss Function. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 813–817. [CrossRef]

49. Kong, F.; Huang, B.; Bradbury, K.; Malof, J.M. The Synthinel-1 dataset: A collection of high resolution synthetic overhead imagery for building segmentation. In Proceedings of the 2020 IEEE Winter Conference on Applications of Computer Vision (WACV), Aspen, CO, USA, 1–5 March 2020; pp. 1803–1812.

50. Zhang, F.; Du, B.; Zhang, L.; Xu, M. Weakly Supervised Learning Based on Coupled Convolutional Neural Networks for Aircraft Detection. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 5553–5563. [CrossRef]

51. Ji, J.; Zhang, T.; Yang, Z.; Jiang, L.; Zhong, W.; Xiong, H. Aircraft Detection from Remote Sensing Image Based on A Weakly Supervised Attention Model. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 322–325.

52. Wu, X.; Hong, D.; Tian, J.; Kiefl, R.; Tao, R. A Weakly-Supervised Deep Network for DSM-Aided Vehicle Detection. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019, pp. 1318–1321.

53. Kaiser, P.; Wegner, J.D.; Lucchi, A.; Jaggi, M.; Hofmann, T.; Schindler, K. Learning Aerial Image Segmentation From Online Maps. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 6054–6068. [CrossRef]

54. Krylov, V.A.; de Martino, M.; Moser, G.; Serpico, S.B. Large urban zone classification on SPOT-5 imagery with convolutional neural networks. In Proceedings of the 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, China, 10–15 July 2016; pp. 1796–1799.

55. Maggiori, E.; Tarabalka, Y.; Charpiat, G.; Alliez, P. Convolutional Neural Networks for Large-Scale Remote-Sensing Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 645–657. [CrossRef]

56. Bittner, K.; Adam, F.; Cui, S.; Körner, M.; Reinartz, P. Building Footprint Extraction From VHR Remote Sensing Images Combined With Normalized DSMs Using Fused Fully Convolutional Networks. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 2615–2629. [CrossRef]

57. Voinov, S.; Krause, D.; Schwarz, E. Towards Automated Vessel Detection and Type Recognition from VHR Optical Satellite Images. In Proceedings of the IGARSS 2018—2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 4823–4826.

58. Piramanayagam, S.; Saber, E.; Schwartzkopf, W.; Koehler, F.W. Supervised Classification of Multisensor Remotely Sensed Images Using a Deep Learning Framework. *Remote Sens.* **2018**, *10*, 1429. [CrossRef]

59. Kim, J.H.; Lee, H.; Hong, S.J.; Kim, S.; Park, J.; Hwang, J.Y.; Choi, J.P. Objects Segmentation From High-Resolution Aerial Images Using U-Net With Pyramid Pooling Layers. *IEEE Geosci. Remote Sens. Lett.* **2019**, *16*, 115–119. [CrossRef]

60. Shahzad, M.; Maurer, M.; Fraundorfer, F.; Wang, Y.; Zhu, X.X. Buildings Detection in VHR SAR Images Using Fully Convolution Neural Networks. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 1100–1116. [CrossRef]

61. Shi, Y.; Li, Q.; Zhu, X. Building Footprint Extraction with Graph Convolutional Network. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 5136–5139.

62. Wu, S.; Du, C.; Chen, H.; Xu, Y.; Guo, N.; Jing, N. Road Extraction from Very High Resolution Images Using Weakly labeled OpenStreetMap Centerline. *Isprs Int. J. -Geo* **2019**, *8*, 478. [CrossRef]

63. Vargas-Muñoz, J.E.; Lobry, S.; Falcão, A.X.; Tuia, D. Correcting rural building annotations in OpenStreetMap using convolutional neural networks. *Isprs J. Photogramm. Remote Sens.* **2019**, *147*, 283–293. [CrossRef]

64. Huang, J.; Zhang, X.; Xin, Q.; Sun, Y.; Zhang, P. Automatic building extraction from high-resolution aerial images and LiDAR data using gated residual refinement network. *Isprs J. Photogramm. Remote Sens.* **2019**, *151*, 91–105. [CrossRef]

65. Griffiths, D.; Boehm, J. Improving public data for building segmentation from Convolutional Neural Networks (CNNs) for fused airborne lidar and image data using active contours. *Isprs J. Photogramm. Remote Sens.* **2019**, *154*, 70–83. [CrossRef]

66. Li, W.; He, C.; Fang, J.; Zheng, J.; Fu, H.; Yu, L. Semantic Segmentation-Based Building Footprint Extraction Using Very High-Resolution Satellite Images and Multi-Source GIS Data. *Remote Sens.* **2019**, *11*, 403. [CrossRef]

67. Manandhar, P.; Marpu, P.R.; Aung, Z.; Melgani, F. Towards Automatic Extraction and Updating of VGI-Based Road Networks Using Deep Learning. *Remote Sens.* **2019**, *11*, 1012. [CrossRef]

68. Zeng, F.; Cheng, L.; Li, N.; Xia, N.; Ma, L.; Zhou, X.; Li, M. A Hierarchical Airport Detection Method Using Spatial Analysis and Deep Learning. *Remote Sens.* **2019**, *11*, 2204. [CrossRef]

69. Schmitt, M.; Prexl, J.; Ebel, P.; Liebel, L.; Zhu, X.X. Weakly Supervised Semantic Segmentation of Satellite Images for Land Cover Mapping – Challenges and Opportunities. *arXiv* **2020**, arXiv:cs.CV/2002.08254.

70. NASA/JPL. Airborne Synthetic Aperture Radar (AIRSAR). Available online: https://airsar.jpl.nasa.gov/index_detail.html (accessed on 27 July 2020).

71. Xiao, Z.; Liu, Q.; Tang, G.; Zhai, X. Elliptic Fourier transformation-based histograms of oriented gradients for rotationally invariant object detection in remote-sensing images. *Int. J. Remote Sens.* **2015**, *36*, 618–644. [CrossRef]

72.	Liu, Z.; Yuan, L.; Weng, L.; Yang., Y. 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, Porto, Portugal 24–26 February 2017; pp. 324–331. [CrossRef]

73.	Tong, X.; Lu, Q.; Xia, G.; Zhang, L. Large-Scale Land Cover Classification in Gaofen-2 Satellite Imagery. In Proceedings of the IGARSS 2018—2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 3599–3602.

74.	Zhu, H.; Chen, X.; Dai, W.; Fu, K.; Ye, Q.; Jiao, J. Orientation robust object detection in aerial images using deep convolutional neural network. In Proceedings of the 2015 IEEE International Conference on Image Processing (ICIP), Quebec City, QC, Canada, 27–30 September 2015; pp. 3735–3739.

75.	Zhu, X.; Hu, J.; Qiu, C.; Shi, Y.; Kang, J.; Mou, L.; Bagheri, H.; Haberle, M.; Hua, Y.; Huang, R.; et al. So2Sat LCZ42: A Benchmark Dataset for Global Local Climate Zones Classification. *IEEE Geosci. Remote Sens. Mag.* **2020**, Early Access. [CrossRef]

76.	Cheng, G.; Han, J.; Lu, X. Remote Sensing Image Scene Classification: Benchmark and State of the Art. *Proc. IEEE* **2017**, *105*, 1865–1883. [CrossRef]

77.	Maggiori, E.; Tarabalka, Y.; Charpiat, G.; Alliez, P. Can Semantic Labeling Methods Generalize to Any City? The Inria Aerial Image Labeling Benchmark. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS). IEEE, Fort Worth, TX, USA, 23–28 July 2017.

78.	Kang, M.; Ji, K.; Leng, X.; Lin, Z. Contextual Region-Based Convolutional Neural Network with Multilayer Fusion for SAR Ship Detection. *Remote Sens.* **2017**, *9*, 860. [CrossRef]

79.	Liu, W.; Ma, L.; Chen, H. Arbitrary-Oriented Ship Detection Framework in Optical Remote-Sensing Images. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 937–941. [CrossRef]

80.	Li, Q.; Mou, L.; Liu, Q.; Wang, Y.; Zhu, X.X. HSF-Net: Multiscale Deep Feature Embedding for Ship Detection in Optical Remote Sensing Imagery. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 7147–7161. [CrossRef]

81.	Zhang, X.; Wang, H.; Xu, C.; Lv, Y.; Fu, C.; Xiao, H.; He, Y. A Lightweight Feature Optimizing Network for Ship Detection in SAR Image. *IEEE Access* **2019**, *7*, 141662–141678. [CrossRef]

82.	Zhang, T.; Zhang, X.; Shi, J.; Wei, S. Depthwise Separable Convolution Neural Network for High-Speed SAR Ship Detection. *Remote Sens.* **2019**, *11*, 2483. [CrossRef]

83.	Jiao, J.; Zhang, Y.; Sun, H.; Yang, X.; Gao, X.; Hong, W.; Fu, K.; Sun, X. A Densely Connected End-to-End Neural Network for Multiscale and Multiscene SAR Ship Detection. *IEEE Access* **2018**, *6*, 20881–20892. [CrossRef]

84.	Wu, F.; Zhou, Z.; Wang, B.; Ma, J. Inshore Ship Detection Based on Convolutional Neural Network in Optical Satellite Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 4005–4015. [CrossRef]

85.	Deng, Z.; Sun, H.; Zhou, S.; Zhao, J. Learning Deep Ship Detector in SAR Images From Scratch. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 4021–4039. [CrossRef]

86.	You, Y.; Cao, J.; Zhang, Y.; Liu, F.; Zhou, W. Nearshore Ship Detection on High-Resolution Remote Sensing Image via Scene-Mask R-CNN. *IEEE Access* **2019**, *7*, 128431–128444. [CrossRef]

87.	You, Y.; Li, Z.; Ran, B.; Cao, J.; Lv, S.; Liu, F. Broad Area Target Search System for Ship Detection via Deep Convolutional Neural Network. *Remote Sens.* **2019**, *11*, 1965. [CrossRef]

88.	Zhang, S.; Wu, R.; Xu, K.; Wang, J.; Sun, W. R-CNN-Based Ship Detection from High Resolution Remote Sensing Imagery. *Remote Sens.* **2019**, *11*, 631. [CrossRef]

89.	Fan, W.; Zhou, F.; Bai, X.; Tao, M.; Tian, T. Ship Detection Using Deep Convolutional Neural Networks for PolSAR Images. *Remote Sens.* **2019**, *11*, 2862. [CrossRef]

90.	Chen, C.; He, C.; Hu, C.; Pei, H.; Jiao, L. A Deep Neural Network Based on an Attention Mechanism for SAR Ship Detection in Multiscale and Complex Scenarios. *IEEE Access* **2019**, *7*, 104848–104863. [CrossRef]

91.	He, Y.; Sun, X.; Gao, L.; Zhang, B. Ship Detection Without Sea-Land Segmentation for Large-Scale High-Resolution Optical Satellite Images. In Proceedings of the IGARSS 2018—2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 717–720.

92.	Gao, L.; He, Y.; Sun, X.; Jia, X.; Zhang, B. Incorporating Negative Sample Training for Ship Detection Based on Deep Learning. *Sensors* **2019**, *19*, 684. [CrossRef] [PubMed]

93.	Zhang, Z.; Guo, W.; Zhu, S.; Yu, W. Toward Arbitrary-Oriented Ship Detection With Rotated Region Proposal and Discrimination Networks. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 1745–1749. [CrossRef]

94. Wang, J.; Lu, C.; Jiang, W. Simultaneous Ship Detection and Orientation Estimation in SAR Images Based on Attention Module and Angle Regression. *Sensors* **2018**, *18*, 2851. [CrossRef] [PubMed]

95. Zhang, Y.; Zhang, Y.; Shi, Z.; Zhang, J.; Wei, M. Rotationally Unconstrained Region Proposals for Ship Target Segmentation in Optical Remote Sensing. *IEEE Access* **2019**, *7*, 87049–87058. [CrossRef]

96. Chen, J.; Xie, F.; Lu, Y.; Jiang, Z. Finding Arbitrary-Oriented Ships From Remote Sensing Images Using Corner Detection. *IEEE Geosci. Remote Sens. Lett.* **2019**, 1–5. Early Access. [CrossRef]

97. Xiao, X.; Zhou, Z.; Wang, B.; Li, L.; Miao, L. Ship Detection under Complex Backgrounds Based on Accurate Rotated Anchor Boxes from Paired Semantic Segmentation. *Remote Sens.* **2019**, *11*, 2506. [CrossRef]

98. Li, M.; Guo, W.; Zhang, Z.; Yu, W.; Zhang, T. Rotated Region Based Fully Convolutional Network for Ship Detection. In Proceedings of the IGARSS 2018—2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 673–676.

99. Wang, T.; Gu, Y. Cnn Based Renormalization Method for Ship Detection in Vhr Remote Sensing Images. In Proceedings of the IGARSS 2018—2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 1252–1255.

100. Fu, K.; Li, Y.; Sun, H.; Yang, X.; Xu, G.; Li, Y.; Sun, X. A Ship Rotation Detection Model in Remote Sensing Images Based on Feature Fusion Pyramid Network and Deep Reinforcement Learning. *Remote Sens.* **2018**, *10*, 1922. [CrossRef]

101. Li, S.; Zhang, Z.; Li, B.; Li, C. Multiscale Rotated Bounding Box-Based Deep Learning Method for Detecting Ship Targets in Remote Sensing Images. *Sensors* **2018**, *18*, 2702. [CrossRef]

102. Sun, J.; Zou, H.; Deng, Z.; Cao, X.; Li, M.; Ma, Q. Multiclass Oriented Ship Localization and Recognition In High Resolution Remote Sensing Images. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 1288–1291.

103. Voinov, S.; Heymann, F.; Bill, R.; Schwarz, E. Multiclass Vessel Detection From High Resolution Optical Satellite Images Based On Deep Neural Networks. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 166–169.

104. Ma, J.; Zhou, Z.; Wang, B.; Zong, H.; Wu, F. Ship Detection in Optical Satellite Images via Directional Bounding Boxes Based on Ship Center and Orientation Prediction. *Remote Sens.* **2019**, *11*, 2173. [CrossRef]

105. Bi, F.; Hou, J.; Chen, L.; Yang, Z.; Wang, Y. Ship Detection for Optical Remote Sensing Images Based on Visual Attention Enhanced Network. *Sensors* **2019**, *19*, 2271. [CrossRef] [PubMed]

106. Feng, Y.; Diao, W.; Zhang, Y.; Li, H.; Chang, Z.; Yan, M.; Sun, X.; Gao, X. Ship Instance Segmentation from Remote Sensing Images Using Sequence Local Context Module. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 1025–1028.

107. Dechesne, C.; Lefèvre, S.; Vadaine, R.; Hajduch, G.; Fablet, R. Ship Identification and Characterization in Sentinel-1 SAR Images with Multi-Task Deep Learning. *Remote Sens.* **2019**, *11*, 2997. [CrossRef]

108. Ma, M.; Chen, J.; Liu, W.; Yang, W. Ship Classification and Detection Based on CNN Using GF-3 SAR Images. *Remote Sens.* **2018**, *10*, 2043. [CrossRef]

109. Lin, H.; Shi, Z.; Zou, Z. Fully Convolutional Network With Task Partitioning for Inshore Ship Detection in Optical Remote Sensing Images. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 1665–1669. [CrossRef]

110. Sun, S.; Lu, Z.; Liu, W.; Hu, W.; Li, R. Shipnet for Semantic Segmentation on VHR Maritime Imagery. In Proceedings of the IGARSS 2018—2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 6911–6914.

111. Tang, T.; Zhou, S.; Deng, Z.; Lei, L.; Zou, H. Arbitrary-Oriented Vehicle Detection in Aerial Imagery with Single Convolutional Neural Networks. *Remote Sens.* **2017**, *9*, 1170. [CrossRef]

112. Li, Q.; Mou, L.; Xu, Q.; Zhang, Y.; Zhu, X.X. R3-Net: A Deep Network for Multioriented Vehicle Detection in Aerial Images and Videos. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 5028–5042. [CrossRef]

113. Deng, Z.; Sun, H.; Zhou, S.; Zhao, J.; Zou, H. Toward Fast and Accurate Vehicle Detection in Aerial Images Using Coupled Region-Based Convolutional Neural Networks. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 3652–3664. [CrossRef]

114. Schilling, H.; Bulatov, D.; Niessner, R.; Middelmann, W.; Soergel, U. Detection of Vehicles in Multisensor Data via Multibranch Convolutional Neural Networks. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 4299–4316. [CrossRef]

115. Audebert, N.; Le Saux, B.; Lefèvre, S. Segment-before-Detect: Vehicle Detection and Classification through Semantic Segmentation of Aerial Images. *Remote Sens.* **2017**, *9*, 368. [CrossRef]
116. Merkle, N.; Azimi, S.M.; Pless, S.; Kurz, F. Semantic Vehicle Segmentation in Very High Resolution Multispectral Aerial Images Using Deep Neural Networks. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 5045–5048.
117. Koga, Y.; Miyazaki, H.; Shibasaki, R. A CNN-Based Method of Vehicle Detection from Aerial Images Using Hard Example Mining. *Remote Sens.* **2018**, *10*, 124. [CrossRef]
118. Gao, Z.; Ji, H.; Mei, T.; Ramesh, B.; Liu, X. EOVNet: Earth-Observation Image-Based Vehicle Detection Network. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 3552–3561. [CrossRef]
119. Li, S.; Xu, Y.; Zhu, M.; Ma, S.; Tang, H. Remote Sensing Airport Detection Based on End-to-End Deep Transferable Convolutional Neural Networks. *IEEE Geosci. Remote Sens. Lett.* **2019**, *16*, 1640–1644. [CrossRef]
120. Zhang, P.; Niu, X.; Dou, Y.; Xia, F. Airport Detection on Optical Satellite Images Using Deep Convolutional Neural Networks. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 1183–1187. [CrossRef]
121. Chen, F.; Ren, R.; Van de Voorde, T.; Xu, W.; Zhou, G.; Zhou, Y. Fast Automatic Airport Detection in Remote Sensing Images Using Convolutional Neural Networks. *Remote Sens.* **2018**, *10*, 443. [CrossRef]
122. Cai, B.; Jiang, Z.; Zhang, H.; Yao, Y.; Nie, S. Online Exemplar-Based Fully Convolutional Network for Aircraft Detection in Remote Sensing Images. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 1095–1099. [CrossRef]
123. Chen, Z.; Zhang, T.; Ouyang, C. End-to-End Airplane Detection Using Transfer Learning in Remote Sensing Images. *Remote Sens.* **2018**, *10*, 139. [CrossRef]
124. Wang, Y.; Li, H.; Jia, P.; Zhang, G.; Wang, T.; Hao, X. Multi-Scale DenseNets-Based Aircraft Detection from Remote Sensing Images. *Sensors* **2019**, *19*, 5270. [CrossRef]
125. Zhao, P.; Gao, H.; Zhang, Y.; Li, H.; Yang, R. An Aircraft Detection Method Based on Improved Mask R-CNN in Remotely Sensed Imagery. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 1370–1373.
126. Wang, H.; Gong, Y.; Wang, Y.; Wang, L.; Pan, C. DeepPlane: A unified deep model for aircraft detection and recognition in remote sensing images. *J. Appl. Remote Sens.* **2017**, *11*, 1–10. [CrossRef]
127. Hou, B.; Li, J.; Zhang, X.; Wang, S.; Jiao, L. Object Detection and Trcacking Based on Convolutional Neural Networks for High-Resolution Optical Remote Sensing Video. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 5433–5436.
128. Xu, Y.; Wu, L.; Xie, Z.; Chen, Z. Building Extraction in Very High Resolution Remote Sensing Imagery Using Deep Learning and Guided Filters. *Remote Sens.* **2018**, *10*, 144. [CrossRef]
129. Lu, X.; Zhong, Y.; Zheng, Z.; Liu, Y.; Zhao, J.; Ma, A.; Yang, J. Multi-Scale and Multi-Task Deep Learning Framework for Automatic Road Extraction. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 9362–9377. [CrossRef]
130. Wei, Y.; Wang, Z.; Xu, M. Road Structure Refined CNN for Road Extraction in Aerial Image. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 709–713. [CrossRef]
131. Li, Y.; Guo, L.; Rao, J.; Xu, L.; Jin, S. Road Segmentation Based on Hybrid Convolutional Network for High-Resolution Visible Remote Sensing Image. *IEEE Geosci. Remote Sens. Lett.* **2019**, *16*, 613–617. [CrossRef]
132. Zhang, X.; Ma, W.; Li, C.; Wu, J.; Tang, X.; Jiao, L. Fully Convolutional Network-Based Ensemble Method for Road Extraction From Aerial Images. *IEEE Geosci. Remote Sens. Lett.* **2019**, 1–5. Early Access. [CrossRef]
133. He, H.; Yang, D.; Wang, S.; Wang, S.; Li, Y. Road Extraction by Using Atrous Spatial Pyramid Pooling Integrated Encoder-Decoder Network and Structural Similarity Loss. *Remote Sens.* **2019**, *11*, 1015. [CrossRef]
134. Liu, Y.; Yao, J.; Lu, X.; Xia, M.; Wang, X.; Liu, Y. RoadNet: Learning to Comprehensively Analyze Road Networks in Complex Urban Scenes From High-Resolution Remotely Sensed Images. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 2043–2056. [CrossRef]
135. Hong, Z.; Ming, D.; Zhou, K.; Guo, Y.; Lu, T. Road Extraction From a High Spatial Resolution Remote Sensing Image Based on Richer Convolutional Features. *IEEE Access* **2018**, *6*, 46988–47000. [CrossRef]
136. Yang, X.; Li, X.; Ye, Y.; Lau, R.Y.K.; Zhang, X.; Huang, X. Road Detection and Centerline Extraction Via Deep Recurrent Convolutional Neural Network U-Net. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 7209–7220. [CrossRef]

137. Azimi, S.M.; Fischer, P.; Körner, M.; Reinartz, P. Aerial LaneNet: Lane-Marking Semantic Segmentation in Aerial Imagery Using Wavelet-Enhanced Cost-Sensitive Symmetric Fully Convolutional Neural Networks. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 2920–2938. [CrossRef]

138. Zhao, W.; Du, S.; Emery, W.J. Object-Based Convolutional Neural Network for High-Resolution Imagery Classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 3386–3396. [CrossRef]

139. Volpi, M.; Tuia, D. Dense Semantic Labeling of Subdecimeter Resolution Images With Convolutional Neural Networks. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 881–893. [CrossRef]

140. Audebert, N.; Le Saux, B.; Lefèvre, S. Beyond RGB: Very high resolution urban remote sensing with multimodal deep networks. *ISPRS J. Photogramm. Remote Sens.* **2018**, *140*, 20–32. [CrossRef]

141. Wang, Y.; Liang, B.; Ding, M.; Li, J. Dense Semantic Labeling with Atrous Spatial Pyramid Pooling and Decoder for High-Resolution Remote Sensing Imagery. *Remote Sens.* **2019**, *11*, 20. [CrossRef]

142. Liu, S.; Ding, W.; Liu, C.; Liu, Y.; Wang, Y.; Li, H. ERN: Edge Loss Reinforced Semantic Segmentation Network for Remote Sensing Images. *Remote Sens.* **2018**, *10*, 1339. [CrossRef]

143. Luo, H.; Chen, C.; Fang, L.; Zhu, X.; Lu, L. High-Resolution Aerial Images Semantic Segmentation Using Deep Fully Convolutional Network With Channel Attention Mechanism. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 3492–3507. [CrossRef]

144. Chen, K.; Fu, K.; Gao, X.; Yan, M.; Zhang, W.; Zhang, Y.; Sun, X. Effective Fusion of Multi-Modal Data with Group Convolutions for Semantic Segmentation of Aerial Imagery. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 3911–3914.

145. Zhang, G.; Lei, T.; Cui, Y.; Jiang, P. A Dual-Path and Lightweight Convolutional Neural Network for High-Resolution Aerial Image Segmentation. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 582. [CrossRef]

146. Cao, Z.; Diao, W.; Zhang, Y.; Yan, M.; Yu, H.; Sun, X.; Fu, K. Semantic Labeling for High-Resolution Aerial Images Based on the DMFFNet. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 1021–1024.

147. Jia, Y.; Ge, Y.; Chen, Y.; Li, S.; Heuvelink, G.B.; Ling, F. Super-Resolution Land Cover Mapping Based on the Convolutional Neural Network. *Remote Sens.* **2019**, *11*, 1815. [CrossRef]

148. Guo, S.; Jin, Q.; Wang, H.; Wang, X.; Wang, Y.; Xiang, S. Learnable Gated Convolutional Neural Network for Semantic Segmentation in Remote-Sensing Images. *Remote Sens.* **2019**, *11*, 1922. [CrossRef]

149. Basaeed, E.; Bhaskar, H.; Hill, P.; Al-Mualla, M.; Bull, D. A supervised hierarchical segmentation of remote-sensing images using a committee of multi-scale convolutional neural networks. *Int. J. Remote Sens.* **2016**, *37*, 1671–1691. [CrossRef]

150. Mou, L.; Hua, Y.; Zhu, X.X. Spatial Relational Reasoning in Networks for Improving Semantic Segmentation of Aerial Images. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 5232–5235.

151. Nogueira, K.; Dalla Mura, M.; Chanussot, J.; Schwartz, W.R.; dos Santos, J.A. Dynamic Multicontext Segmentation of Remote Sensing Images Based on Convolutional Networks. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 7503–7520. [CrossRef]

152. Sun, Y.; Zhang, X.; Xin, Q.; Huang, J. Developing a multi-filter convolutional neural network for semantic segmentation using high-resolution aerial imagery and LiDAR data. *ISPRS J. Photogramm. Remote Sens.* **2018**, *143*, 3–14. [CrossRef]

153. Papadomanolaki, M.; Vakalopoulou, M.; Karantzalos, K. A Novel Object-Based Deep Learning Framework for Semantic Segmentation of Very High-Resolution Remote Sensing Data: Comparison with Convolutional and Fully Convolutional Networks. *Remote Sens.* **2019**, *11*, 684. [CrossRef]

154. Yue, K.; Yang, L.; Li, R.; Hu, W.; Zhang, F.; Li, W. TreeUNet: Adaptive Tree convolutional neural networks for subdecimeter aerial image segmentation. *ISPRS J. Photogramm. Remote Sens.* **2019**, *156*, 1–13. [CrossRef]

155. Liu, Y.; Gross, L.; Li, Z.; Li, X.; Fan, X.; Qi, W. Automatic Building Extraction on High-Resolution Remote Sensing Imagery Using Deep Convolutional Encoder-Decoder With Spatial Pyramid Pooling. *IEEE Access* **2019**, *7*, 128774–128786. [CrossRef]

156. Zhang, Y.; Gong, W.; Sun, J.; Li, W. Web-Net: A Novel Nest Networks with Ultra-Hierarchical Sampling for Building Extraction from Aerial Imageries. *Remote Sens.* **2019**, *11*, 1897. [CrossRef]

157. Yang, H.L.; Yuan, J.; Lunga, D.; Laverdiere, M.; Rose, A.; Bhaduri, B. Building Extraction at Scale Using Convolutional Neural Network: Mapping of the United States. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 2600–2614. [CrossRef]

158. Hui, J.; Du, M.; Ye, X.; Qin, Q.; Sui, J. Effective Building Extraction From High-Resolution Remote Sensing Images With Multitask Driven Deep Neural Network. *IEEE Geosci. Remote Sens. Lett.* **2019**, *16*, 786–790. [CrossRef]

159. Wang, S.; Zhou, L.; He, P.; Quan, D.; Zhao, Q.; Liang, X.; Hou, B. An Improved Fully Convolutional Network for Learning Rich Building Features. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 6444–6447.

160. Guo, Z.; Wu, G.; Song, X.; Yuan, W.; Chen, Q.; Zhang, H.; Shi, X.; Xu, M.; Xu, Y.; Shibasaki, R.; et al. Super-Resolution Integrated Building Semantic Segmentation for Multi-Source Remote Sensing Imagery. *IEEE Access* **2019**, *7*, 99381–99397. [CrossRef]

161. Lin, J.; Jing, W.; Song, H.; Chen, G. ESFNet: Efficient Network for Building Extraction From High-Resolution Aerial Images. *IEEE Access* **2019**, *7*, 54285–54294. [CrossRef]

162. Schuegraf, P.; Bittner, K. Automatic Building Footprint Extraction from Multi-Resolution Remote Sensing Images Using a Hybrid FCN. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 191. [CrossRef]

163. Liu, P.; Liu, X.; Liu, M.; Shi, Q.; Yang, J.; Xu, X.; Zhang, Y. Building Footprint Extraction from High-Resolution Images via Spatial Residual Inception Convolutional Neural Network. *Remote Sens.* **2019**, *11*, 830. [CrossRef]

164. Ye, Z.; Fu, Y.; Gan, M.; Deng, J.; Comber, A.; Wang, K. Building Extraction from Very High Resolution Aerial Imagery Using Joint Attention Deep Neural Network. *Remote Sens.* **2019**, *11*, 2970. [CrossRef]

165. Shrestha, S.; Vanneschi, L. Improved Fully Convolutional Network with Conditional Random Fields for Building Extraction. *Remote Sens.* **2018**, *10*, 1135. [CrossRef]

166. Ji, S.; Wei, S.; Lu, M. A scale robust convolutional neural network for automatic building extraction from aerial and satellite imagery. *Int. J. Remote Sens.* **2019**, *40*, 3308–3322. [CrossRef]

167. Wen, Q.; Jiang, K.; Wang, W.; Liu, Q.; Guo, Q.; Li, L.; Wang, P. Automatic Building Extraction from Google Earth Images under Complex Backgrounds Based on Deep Instance Segmentation Network. *Sensors* **2019**, *19*, 333. [CrossRef]

168. Shi, Q.; Liu, M.; Liu, X.; Liu, P.; Zhang, P.; Yang, J.; Li, X. Domain Adaption for Fine-Grained Urban Village Extraction From Satellite Images. *IEEE Geosci. Remote Sens. Lett.* **2020**, *17*, 1430–1434. [CrossRef]

169. Wang, J.; Kuffer, M.; Roy, D.; Pfeffer, K. Deprivation pockets through the lens of convolutional neural networks. *Remote Sens. Environ.* **2019**, *234*, 111448. [CrossRef]

170. Persello, C.; Stein, A. Deep Fully Convolutional Networks for the Detection of Informal Settlements in VHR Images. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 2325–2329. [CrossRef]

171. Mboga, N.; Persello, C.; Bergado, J.R.; Stein, A. Detection of Informal Settlements from VHR Images Using Convolutional Neural Networks. *Remote Sens.* **2017**, *9*, 1106. [CrossRef]

172. Wurm, M.; Stark, T.; Zhu, X.X.; Weigand, M.; Taubenböck, H. Semantic segmentation of slums in satellite images using transfer learning on fully convolutional neural networks. *ISPRS J. Photogramm. Remote Sens.* **2019**, *150*, 59–69. [CrossRef]

173. Al-Najjar, H.A.H.; Kalantar, B.; Pradhan, B.; Saeidi, V.; Halin, A.A.; Ueda, N.; Mansor, S. Land Cover Classification from fused DSM and UAV Images Using Convolutional Neural Networks. *Remote Sens.* **2019**, *11*, 1461. [CrossRef]

174. Liu, R.; Kuffer, M.; Persello, C. The Temporal Dynamics of Slums Employing a CNN-Based Change Detection Approach. *Remote Sens.* **2019**, *11*, 2844. [CrossRef]

175. Arabi, M.E.A.; Karoui, M.S.; Djerriri, K. Optical Remote Sensing Change Detection Through Deep Siamese Network. In Proceedings of the IGARSS 2018—2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 5041–5044.

176. Daudt, R.C.; Le Saux, B.; Boulch, A.; Gousseau, Y. Urban Change Detection for Multispectral Earth Observation Using Convolutional Neural Networks. In Proceedings of the IGARSS 2018—2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 2115–2118.

177. Amin Larabi, M.E.; Chaib, S.; Bakhti, K.; Karoui, M.S. Transfer Learning for Changes Detection in Optical Remote Sensing Imagery. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 1582–1585.

178. Jaturapitpornchai, R.; Matsuoka, M.; Kanemoto, N.; Kuzuoka, S.; Ito, R.; Nakamura, R. Sar-Image Based Urban Change Detection in Bangkok, Thailand Using Deep Learning. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 7403–7406.

179. Jaturapitpornchai, R.; Matsuoka, M.; Kanemoto, N.; Kuzuoka, S.; Ito, R.; Nakamura, R. Newly Built Construction Detection in SAR Images Using Deep Learning. *Remote Sens.* **2019**, *11*, 1444. [CrossRef]

180. Li, L.; Wang, C.; Zhang, H.; Zhang, B. Residual Unet for Urban Building Change Detection with Sentinel-1 SAR Data. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 1498–1501.

181. Yao, Y.; Jiang, Z.; Zhang, H.; Cai, B.; Meng, G.; Zuo, D. Chimney and condensing tower detection based on faster R-CNN in high resolution remote sensing images. In Proceedings of the 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Fort Worth, TX, USA, 23–28 July 2017; pp. 3329–3332.

182. Zhang, N.; Liu, Y.; Zou, L.; Zhao, H.; Dong, W.; Zhou, H.; Zhou, H.; Huang, M. Automatic Recognition of Oil Industry Facilities Based on Deep Learning. In Proceedings of the IGARSS 2018—2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 2519–2522.

183. Zhang, H.; Deng, Q. Deep Learning Based Fossil-Fuel Power Plant Monitoring in High Resolution Remote Sensing Images: A Comparative Study. *Remote Sens.* **2019**, *11*, 1117. [CrossRef]

184. Xia, X.; Persello, C.; Koeva, M. Deep Fully Convolutional Networks for Cadastral Boundary Detection from UAV Images. *Remote Sens.* **2019**, *11*, 1725. [CrossRef]

185. Crommelinck, S.; Koeva, M.; Yang, M.Y.; Vosselman, G. Application of Deep Learning for Delineation of Visible Cadastral Boundaries from Remote Sensing Imagery. *Remote Sens.* **2019**, *11*, 2505. [CrossRef]

186. Qiu, C.; Schmitt, M.; Ghamisi, P.; Mou, L.; Zhu, X.X. Feature Importance Analysis of Sentinel-2 Imagery for Large-Scale Urban Local Climate Zone Classification. In Proceedings of the IGARSS 2018—2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 4681–4684.

187. Yang, R.; Zhang, Y.; Zhao, P.; Ji, Z.; Deng, W. MSPPF-Nets: A Deep Learning Architecture for Remote Sensing Image Classification. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 3045–3048.

188. Arief, H.A.; Strand, G.H.; Tveite, H.; Indahl, U.G. Land Cover Segmentation of Airborne LiDAR Data Using Stochastic Atrous Network. *Remote Sens.* **2018**, *10*, 973. [CrossRef]

189. Zhang, X.; Xiao, Z.; Li, D.; Fan, M.; Zhao, L. Semantic Segmentation of Remote Sensing Images Using Multiscale Decoding Network. *IEEE Geosci. Remote Sens. Lett.* **2019**, *16*, 1492–1496. [CrossRef]

190. Henry, C.J.; Storie, C.D.; Palaniappan, M.; Alhassan, V.; Swamy, M.; Aleshinloye, D.; Curtis, A.; Kim, D. Automated LULC map production using deep neural networks. *Int. J. Remote Sens.* **2019**, *40*, 4416–4440. [CrossRef]

191. Ahishali, M.; Kiranyaz, S.; Ince, T.; Gabbouj, M. Dual and Single Polarized SAR Image Classification Using Compact Convolutional Neural Networks. *Remote Sens.* **2019**, *11*, 1340. [CrossRef]

192. Ma, F.; Gao, F.; Sun, J.; Zhou, H.; Hussain, A. Attention Graph Convolution Network for Image Segmentation in Big SAR Imagery Data. *Remote Sens.* **2019**, *11*, 2586. [CrossRef]

193. Zhang, C.; Sargent, I.; Pan, X.; Li, H.; Gardiner, A.; Hare, J.; Atkinson, P.M. Joint Deep Learning for land cover and land use classification. *Remote Sens. Environ.* **2019**, *221*, 173–187. [CrossRef]

194. Pan, X.; Zhao, J. A central-point-enhanced convolutional neural network for high-resolution remote-sensing image classification. *Int. J. Remote Sens.* **2017**, *38*, 6554–6581. [CrossRef]

195. Gaetano, R.; Ienco, D.; Ose, K.; Cresson, R. A Two-Branch CNN Architecture for Land Cover Classification of PAN and MS Imagery. *Remote Sens.* **2018**, *10*, 1746. [CrossRef]

196. Li, L. Deep Residual Autoencoder with Multiscaling for Semantic Segmentation of Land-Use Images. *Remote Sens.* **2019**, *11*, 2142. [CrossRef]

197. Panboonyuen, T.; Jitkajornwanich, K.; Lawawirojwong, S.; Srestasathiern, P.; Vateekul, P. Semantic Segmentation on Remotely Sensed Images Using an Enhanced Global Convolutional Network with Channel Attention and Domain Specific Transfer Learning. *Remote Sens.* **2019**, *11*, 83. [CrossRef]

198. Guo, Y.; Chen, E.; Li, Z.; Zhao, L.; Xu, K. Convolutional Highway Unit Network for Large-Scale Classification with GF-3 Dual-Pol Sar Data. In Proceedings of the IGARSS 2018—2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 2424–2427.

199. Hu, Y.; Zhang, Q.; Zhang, Y.; Yan, H. A Deep Convolution Neural Network Method for Land Cover Mapping: A Case Study of Qinhuangdao, China. *Remote Sens.* **2018**, *10*, 2053. [CrossRef]

200. Yang, X.; Chen, Z.; Li, B.; Peng, D.; Chen, P.; Zhang, B. A Fast and Precise Method for Large-Scale Land-Use Mapping Based on Deep Learning. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 5913–5916.

201. Stoian, A.; Poulain, V.; Inglada, J.; Poughon, V.; Derksen, D. Land Cover Maps Production with High Resolution Satellite Image Time Series and Convolutional Neural Networks: Adaptations and Limits for Operational Systems. *Remote Sens.* **2019**, *11*, 1986. [CrossRef]

202. Feng, Q.; Yang, J.; Zhu, D.; Liu, J.; Guo, H.; Bayartungalag, B.; Li, B. Integrating Multitemporal Sentinel-1/2 Data for Coastal Land Cover Classification Using a Multibranch Convolutional Neural Network: A Case of the Yellow River Delta. *Remote Sens.* **2019**, *11*, 1006. [CrossRef]

203. Gao, L.; Luo, J.; Xia, L.; Wu, T.; Sun, Y.; Liu, H. Topographic constrained land cover classification in mountain areas using fully convolutional network. *Int. J. Remote Sens.* **2019**, *40*, 7127–7152. [CrossRef]

204. Mahdianpari, M.; Salehi, B.; Rezaee, M.; Mohammadimanesh, F.; Zhang, Y. Very Deep Convolutional Neural Networks for Complex Land Cover Mapping Using Multispectral Remote Sensing Imagery. *Remote Sens.* **2018**, *10*, 1119. [CrossRef]

205. Mohammadimanesh, F.; Salehi, B.; Mahdianpari, M.; Gill, E.; Molinier, M. A new fully convolutional neural network for semantic segmentation of polarimetric SAR imagery in complex land cover ecosystem. *ISPRS J. Photogramm. Remote Sens.* **2019**, *151*, 223–236. [CrossRef]

206. Pouliot, D.; Latifovic, R.; Pasher, J.; Duffe, J. Assessment of Convolution Neural Networks for Wetland Mapping with Landsat in the Central Canadian Boreal Forest Region. *Remote Sens.* **2019**, *11*, 772. [CrossRef]

207. DeLancey, E.R.; Simms, J.F.; Mahdianpari, M.; Brisco, B.; Mahoney, C.; Kariyeva, J. Comparing Deep Learning and Shallow Learning for Large-Scale Wetland Classification in Alberta, Canada. *Remote Sens.* **2020**, *12*, 2. [CrossRef]

208. Li, J.; Zhang, R.; Li, Y. Multiscale convolutional neural network for the detection of built-up areas in high-resolution SAR images. In Proceedings of the 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, China, 10–15 July2016; pp. 910–913.

209. Tan, Y.; Xiong, S.; Li, Y. Automatic Extraction of Built-Up Areas From Panchromatic and Multispectral Remote Sensing Images Using Double-Stream Deep Convolutional Neural Networks. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 3988–4004. [CrossRef]

210. Huang, B.; Zhao, B.; Song, Y. Urban land-use mapping using a deep convolutional neural network with high spatial resolution multispectral remote sensing imagery. *Remote Sens. Environ.* **2018**, *214*, 73–86. [CrossRef]

211. Guo, J.; Ren, H.; Zheng, Y.; Nie, J.; Chen, S.; Sun, Y.; Qin, Q. Identify Urban Area From Remote Sensing Image Using Deep Learning Method. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 7407–7410.

212. Feng, W.; Sui, H.; Huang, W.; Xu, C.; An, K. Water Body Extraction From Very High-Resolution Remote Sensing Imagery Using Deep U-Net and a Superpixel-Based Conditional Random Field Model. *IEEE Geosci. Remote Sens. Lett.* **2019**, *16*, 618–622. [CrossRef]

213. Pu, F.; Ding, C.; Chao, Z.; Yu, Y.; Xu, X. Water-Quality Classification of Inland Lakes Using Landsat8 Images by Convolutional Neural Networks. *Remote Sens.* **2019**, *11*, 1674. [CrossRef]

214. Isikdogan, F.; Bovik, A.C.; Passalacqua, P. Surface Water Mapping by Deep Learning. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 4909–4918. [CrossRef]

215. Miao, Z.; Fu, K.; Sun, H.; Sun, X.; Yan, M. Automatic Water-Body Segmentation From High-Resolution Satellite Images via Deep Networks. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 602–606. [CrossRef]

216. Li, Z.; Wang, R.; Zhang, W.; Hu, F.; Meng, L. Multiscale Features Supported DeepLabV3+ Optimization Scheme for Accurate Water Semantic Segmentation. *IEEE Access* **2019**, *7*, 155787–155804. [CrossRef]

217. Shamsolmoali, P.; Zareapoor, M.; Wang, R.; Zhou, H.; Yang, J. A Novel Deep Structure U-Net for Sea-Land Segmentation in Remote Sensing Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 3219–3232. [CrossRef]

218. Chu, Z.; Tian, T.; Feng, R.; Wang, L. Sea-Land Segmentation With Res-UNet And Fully Connected CRF. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 3840–3843.

219. Xia, M.; Qian, J.; Zhang, X.; Liu, J.; Xu, Y. River segmentation based on separable attention residual network. *J. Appl. Remote Sens.* **2019**, *14*, 1–15. [CrossRef]

220. Jiao, L.; Zhang, F.; Liu, F.; Yang, S.; Li, L.; Feng, Z.; Qu, R. A Survey of Deep Learning-Based Object Detection. *IEEE Access* **2019**, *7*, 128837–128868. [CrossRef]

221. Zhang, C.; Wei, S.; Ji, S.; Lu, M. Detecting Large-Scale Urban Land Cover Changes from Very High Resolution Remote Sensing Images Using CNN-Based Classification. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 189. [CrossRef]

222. Deng, Z.; Sun, H.; Zhou, S.; Zhao, J.; Lei, L.; Zou, H. Multi-scale object detection in remote sensing imagery with convolutional neural networks. *ISPRS J. Photogramm. Remote Sens.* **2018**, *145*, 3–22. [CrossRef]

223. Wu, X.; Hong, D.; Ghamisi, P.; Li, W.; Tao, R. MsRi-CCF: Multi-Scale and Rotation-Insensitive Convolutional Channel Features for Geospatial Object Detection. *Remote Sens.* **2018**, *10*, 1190. [CrossRef]

224. Wang, Y.; Dong, Z.; Zhu, Y. Multiscale Block Fusion Object Detection Method for Large-Scale High-Resolution Remote Sensing Imagery. *IEEE Access* **2019**, *7*, 99530–99539. [CrossRef]

225. Ying, X.; Wang, Q.; Li, X.; Yu, M.; Jiang, H.; Gao, J.; Liu, Z.; Yu, R. Multi-Attention Object Detection Model in Remote Sensing Images Based on Multi-Scale. *IEEE Access* **2019**, *7*, 94508–94519. [CrossRef]

226. Zhang, S.; He, G.; Chen, H.; Jing, N.; Wang, Q. Scale Adaptive Proposal Network for Object Detection in Remote Sensing Images. *IEEE Geosci. Remote Sens. Lett.* **2019**, *16*, 864–868. [CrossRef]

227. Wang, P.; Sun, X.; Diao, W.; Fu, K. FMSSD: Feature-Merged Single-Shot Detection for Multiscale Objects in Large-Scale Remote Sensing Imagery. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 3377–3390. [CrossRef]

228. Zhang, H.; Wu, J.; Liu, Y.; Yu, J. VaryBlock: A Novel Approach for Object Detection in Remote Sensed Images. *Sensors* **2019**, *19*, 5284. [CrossRef] [PubMed]

229. Li, K.; Cheng, G.; Bu, S.; You, X. Rotation-Insensitive and Context-Augmented Object Detection in Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 2337–2348. [CrossRef]

230. Zhang, G.; Lu, S.; Zhang, W. CAD-Net: A Context-Aware Detection Network for Objects in Remote Sensing Imagery. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 10015–10024. [CrossRef]

231. Mo, N.; Yan, L.; Zhu, R.; Xie, H. Class-Specific Anchor Based and Context-Guided Multi-Class Object Detection in High Resolution Remote Sensing Imagery with a Convolutional Neural Network. *Remote Sens.* **2019**, *11*, 272. [CrossRef]

232. Chen, C.; Gong, W.; Chen, Y.; Li, W. Object Detection in Remote Sensing Images Based on a Scene-Contextual Feature Pyramid Network. *Remote Sens.* **2019**, *11*, 339. [CrossRef]

233. Ren, Y.; Zhu, C.; Xiao, S. Deformable Faster R-CNN with Aggregating Multi-Layer Features for Partially Occluded Object Detection in Optical Remote Sensing Images. *Remote Sens.* **2018**, *10*, 1470. [CrossRef]

234. Cheng, G.; Zhou, P.; Han, J. Learning Rotation-Invariant Convolutional Neural Networks for Object Detection in VHR Optical Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 7405–7415. [CrossRef]

235. Su, H.; Wei, S.; Yan, M.; Wang, C.; Shi, J.; Zhang, X. Object Detection and Instance Segmentation in Remote Sensing Imagery Based on Precise Mask R-CNN. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 1454–1457.

236. Wang, J.; Ding, J.; Guo, H.; Cheng, W.; Pan, T.; Yang, W. Mask OBB: A Semantic Attention-Based Mask Oriented Bounding Box Representation for Multi-Category Object Detection in Aerial Images. *Remote Sens.* **2019**, *11*, 2930. [CrossRef]

237. Zhong, L.; Hu, L.; Zhou, H. Deep learning based multi-temporal crop classification. *Remote Sens. Environ.* **2019**, *221*, 430–443. [CrossRef]

238. Pelletier, C.; Webb, G.I.; Petitjean, F. Temporal Convolutional Neural Network for the Classification of Satellite Image Time Series. *Remote Sens.* **2019**, *11*, 523. [CrossRef]

239. Ji, S.; Zhang, C.; Xu, A.; Shi, Y.; Duan, Y. 3D Convolutional Neural Networks for Crop Classification with Multi-Temporal Remote Sensing Images. *Remote Sens.* **2018**, *10*, 75. [CrossRef]

240. Zhou, Z.; Li, S.; Shao, Y. Crops Classification from Sentinel-2A Multi-spectral Remote Sensing Images Based on Convolutional Neural Networks. In Proceedings of the IGARSS 2018—2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 5300–5303.

241. Mullissa, A.G.; Persello, C.; Stein, A. PolSARNet: A Deep Fully Convolutional Network for Polarimetric SAR Image Classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 5300–5309. [CrossRef]

242. Zhou, Y.; Wang, H.; Xu, F.; Jin, Y. Polarimetric SAR Image Classification Using Deep Convolutional Neural Networks. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 1935–1939. [CrossRef]

243. Chen, S.; Tao, C. PolSAR Image Classification Using Polarimetric-Feature-Driven Deep Convolutional Neural Network. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 627–631. [CrossRef]

244. Cué La Rosa, L.E.; Happ, P.N.; Feitosa, R.Q. Dense Fully Convolutional Networks for Crop Recognition from Multitemporal SAR Image Sequences. In Proceedings of the IGARSS 2018—2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 7460–7463.

245. Zhang, D.; Zhang, J.; Pan, Y.; Duan, Y. Fully Convolutional Neural Networks for Large Scale Cropland Mapping with Historical Label Dataset. In Proceedings of the IGARSS 2018—2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July2018; pp. 4659–4662.

246. Rumelhart, D.E.; Hinton, G.E.; Williams, R.J. Learning Internal Representations by Error Propagation. In *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Volume 1: Foundations*; Rumelhart, D.E., Mcclelland, J.L., Eds.; MIT Press: Cambridge, MA, USA, 1986; pp. 318–362.

247. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef]

248. Sun, Z.; Di, L.; Fang, H. Using long short-term memory recurrent neural network in land cover classification on Landsat and Cropland data layer time series. *Int. J. Remote Sens.* **2019**, *40*, 593–614. [CrossRef]

249. Rußwurm, M.; Körner, M. Multi-Temporal Land Cover Classification with Sequential Recurrent Encoders. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 129. [CrossRef]

250. Russwurm, M.; Korner, M. Temporal Vegetation Modelling Using Long Short-Term Memory Networks for Crop Identification From Medium-Resolution Multi-Spectral Satellite Images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Honolulu, HI, USA, 21–26 July 2017.

251. Zhao, H.; Chen, Z.; Jiang, H.; Jing, W.; Sun, L.; Feng, M. Evaluation of Three Deep Learning Models for Early Crop Classification Using Sentinel-1A Imagery Time Series—A Case Study in Zhanjiang, China. *Remote Sens.* **2019**, *11*, 2673. [CrossRef]

252. Wang, H.; Zhao, X.; Zhang, X.; Wu, D.; Du, X. Long Time Series Land Cover Classification in China from 1982 to 2015 Based on Bi-LSTM Deep Learning. *Remote Sens.* **2019**, *11*, 1639. [CrossRef]

253. Interdonato, R.; Ienco, D.; Gaetano, R.; Ose, K. DuPLO: A DUal view Point deep Learning architecture for time series classificatiOn. *ISPRS J. Photogramm. Remote Sens.* **2019**, *149*, 91–104. [CrossRef]

254. Ienco, D.; Interdonato, R.; Gaetano, R.; Minh, D.H.T. Combining Sentinel-1 and Sentinel-2 Satellite Image Time Series for land cover mapping via a multi-source deep learning architecture. *ISPRS J. Photogramm. Remote Sens.* **2019**, *158*, 11–22. [CrossRef]

255. Rußwurm, M.; Körner, M. Convolutional LSTMs for Cloud-Robust Segmentation of Remote Sensing Imagery.*arXiv* **2018**, arXiv:cs.CV/1811.02471.

256. Chang, T.; Rasmussen, B.P.; Dickson, B.G.; Zachmann, L.J. Chimera: A Multi-Task Recurrent Convolutional Neural Network for Forest Classification and Structural Estimation. *Remote Sens.* **2019**, *11*, 768. [CrossRef]

257. Teimouri, N.; Dyrmann, M.; Jørgensen, R.N. A Novel Spatio-Temporal FCN-LSTM Network for Recognizing Various Crop Types Using Multi-Temporal Radar Images. *Remote Sens.* **2019**, *11*, 990. [CrossRef]

258. Li, Z.; Chen, G.; Zhang, T. Temporal Attention Networks for Multitemporal Multisensor Crop Classification. *IEEE Access* **2019**, *7*, 134677–134690. [CrossRef]

259. García-Pedrero, A.; Lillo-Saavedra, M.; Rodríguez-Esparragón, D.; Gonzalo-Martín, C. Deep Learning for Automatic Outlining Agricultural Parcels: Exploiting the Land Parcel Identification System. *IEEE Access* **2019**, *7*, 158223–158236. [CrossRef]

260. Du, Z.; Yang, J.; Ou, C.; Zhang, T. Smallholder Crop Area Mapped with a Semantic Segmentation Deep Learning Method. *Remote Sens.* **2019**, *11*, 888. [CrossRef]

261. Parente, L.; Taquary, E.; Silva, A.P.; Souza, C.; Ferreira, L. Next Generation Mapping: Combining Deep Learning, Cloud Computing, and Big Remote Sensing Data. *Remote Sens.* **2019**, *11*, 2881. [CrossRef]

262. Masoud, K.M.; Persello, C.; Tolpekin, V.A. Delineation of Agricultural Field Boundaries from Sentinel-2 Images Using a Novel Super-Resolution Contour Detector Based on Fully Convolutional Networks. *Remote Sens.* **2020**, *12*, 59. [CrossRef]

263. Persello, C.; Tolpekin, V.; Bergado, J.; de By, R. Delineation of agricultural fields in smallholder farms from satellite images using fully convolutional networks and combinatorial grouping. *Remote Sens. Environ.* **2019**, *231*, 111253. [CrossRef]

264. Kitano, B.T.; Mendes, C.C.T.; Geus, A.R.; Oliveira, H.C.; Souza, J.R. Corn Plant Counting Using Deep Learning and UAV Images. *IEEE Geosci. Remote Sens. Lett.* **2019**, 1–5. Early Access. [CrossRef]

265. Malambo, L.; Popescu, S.; Ku, N.W.; Rooney, W.; Zhou, T.; Moore, S. A Deep Learning Semantic Segmentation-Based Approach for Field-Level Sorghum Panicle Counting. *Remote Sens.* **2019**, *11*, 2939. [CrossRef]

266. Chen, Y.; Lee, W.S.; Gan, H.; Peres, N.; Fraisse, C.; Zhang, Y.; He, Y. Strawberry Yield Prediction Based on a Deep Neural Network Using High-Resolution Aerial Orthoimages. *Remote Sens.* **2019**, *11*, 1584. [CrossRef]

267. Fuentes-Pacheco, J.; Torres-Olivares, J.; Roman-Rangel, E.; Cervantes, S.; Juarez-Lopez, P.; Hermosillo-Valadez, J.; Rendón-Mancha, J.M. Fig Plant Segmentation from Aerial Images Using a Deep Convolutional Encoder-Decoder Network. *Remote Sens.* **2019**, *11*, 1157. [CrossRef]

268. Zhou, J.; Tian, Y.; Yuan, C.; Yin, K.; Yang, G.; Wen, M. Improved UAV Opium Poppy Detection Using an Updated YOLOv3 Model. *Sensors* **2019**, *19*, 4851. [CrossRef]

269. Rahnemoonfar, M.; Dobbs, D.; Yari, M.; Starek, M.J. DisCountNet: Discriminating and Counting Network for Real-Time Counting and Localization of Sparse Objects in High-Resolution UAV Imagery. *Remote Sens.* **2019**, *11*, 1128. [CrossRef]

270. Li, W.; Dong, R.; Fu, H.; Yu, L. Large-Scale Oil Palm Tree Detection from High-Resolution Satellite Images Using Two-Stage Convolutional Neural Networks. *Remote Sens.* **2019**, *11*, 11. [CrossRef]

271. Zheng, J.; Li, W.; Xia, M.; Dong, R.; Fu, H.; Yuan, S. Large-Scale Oil Palm Tree Detection from High-Resolution Remote Sensing Images Using Faster-RCNN. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 1422–1425.

272. Freudenberg, M.; Nölke, N.; Agostini, A.; Urban, K.; Wörgötter, F.; Kleinn, C. Large Scale Palm Tree Detection in High Resolution Satellite Images Using U-Net. *Remote Sens.* **2019**, *11*, 312. [CrossRef]

273. Duan, Y.; Zhong, J.; Shuai, G.; Zhu, S.; Gu, X. Time-Scale Transferring Deep Convolutional Neural Network for Mapping Early Rice. In Proceedings of the IGARSS 2018—2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 1136–1139.

274. Zhang, M.; Lin, H.; Wang, G.; Sun, H.; Fu, J. Mapping Paddy Rice Using a Convolutional Neural Network (CNN) with Landsat 8 Datasets in the Dongting Lake Area, China. *Remote Sens.* **2018**, *10*, 1840. [CrossRef]

275. Jiang, T.; Liu, X.; Wu, L. Method for Mapping Rice Fields in Complex Landscape Areas Based on Pre-Trained Convolutional Neural Network from HJ-1 A/B Data. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 418. [CrossRef]

276. Fu, Y.; Ye, Z.; Deng, J.; Zheng, X.; Huang, Y.; Yang, W.; Wang, Y.; Wang, K. Finer Resolution Mapping of Marine Aquaculture Areas Using WorldView-2 Imagery and a Hierarchical Cascade Convolutional Neural Network. *Remote Sens.* **2019**, *11*, 1678. [CrossRef]

277. Mazza, A.; Sica, F.; Rizzoli, P.; Scarpa, G. TanDEM-X Forest Mapping Using Convolutional Neural Networks. *Remote Sens.* **2019**, *11*, 2980. [CrossRef]

278. Sylvain, J.D.; Drolet, G.; Brown, N. Mapping dead forest cover using a deep convolutional neural network and digital aerial photography. *ISPRS J. Photogramm. Remote Sens.* **2019**, *156*, 14–26. [CrossRef]

279. Hamdi, Z.M.; Brandmeier, M.; Straub, C. Forest Damage Assessment Using Deep Learning on High Resolution Remote Sensing Data. *Remote Sens.* **2019**, *11*, 1976. [CrossRef]

280. Safonova, A.; Tabik, S.; Alcaraz-Segura, D.; Rubtsov, A.; Maglinets, Y.; Herrera, F. Detection of Fir Trees (Abies sibirica) Damaged by the Bark Beetle in Unmanned Aerial Vehicle Images with Deep Learning. *Remote Sens.* **2019**, *11*, 643. [CrossRef]

281. Fromm, M.; Schubert, M.; Castilla, G.; Linke, J.; McDermid, G. Automated Detection of Conifer Seedlings in Drone Imagery Using Convolutional Neural Networks. *Remote Sens.* **2019**, *11*, 2585. [CrossRef]

282. Weinstein, B.G.; Marconi, S.; Bohlman, S.; Zare, A.; White, E. Individual Tree-Crown Detection in RGB Imagery Using Semi-Supervised Deep Learning Neural Networks. *Remote Sens.* **2019**, *11*, 1309. [CrossRef]

283. Dong, T.; Shen, Y.; Zhang, J.; Ye, Y.; Fan, J. Progressive Cascaded Convolutional Neural Networks for Single Tree Detection with Google Earth Imagery. *Remote Sens.* **2019**, *11*, 1786. [CrossRef]

284. 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.; et al. Assessment of CNN-Based Methods for Individual Tree Detection on Images Captured by RGB Cameras Attached to UAVs. *Sensors* **2019**, *19*, 3595. [CrossRef]

285. Rist, Y.; Shendryk, I.; Diakogiannis, F.; Levick, S. Weed Mapping Using Very High Resolution Satellite Imagery and Fully Convolutional Neural Network. In Proceedings of the IGARSS 2019—2019 IEEE

International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 9784–9787.

286. Guirado, E.; Tabik, S.; Alcaraz-Segura, D.; Cabello, J.; Herrera, F. Deep-learning Versus OBIA for Scattered Shrub Detection with Google Earth Imagery: Ziziphus lotus as Case Study. *Remote Sens.* **2017**, *9*, 1220. [CrossRef]

287. Langford, Z.L.; Kumar, J.; Hoffman, F.M.; Breen, A.L.; Iversen, C.M. Arctic Vegetation Mapping Using Unsupervised Training Datasets and Convolutional Neural Networks. *Remote Sens.* **2019**, *11*, 69. [CrossRef]

288. Kattenborn, T.; Eichel, J.; Fassnacht, F.E. Convolutional Neural Networks enable efficient, accurate and fine-grained segmentation of plant species and communities from high-resolution UAV imagery. *Sci. Rep.* **2019**, *9*, 1–9. [CrossRef] [PubMed]

289. Nogueira, K.; dos Santos, J.A.; Menini, N.; Silva, T.S.F.; Morellato, L.P.C.; Torres, R.D.S. Spatio-Temporal Vegetation Pixel Classification by Using Convolutional Networks. *IEEE Geosci. Remote Sens. Lett.* **2019**, *16*, 1665–1669. [CrossRef]

290. Nogueira, K.; dos Santos, J.A.; Cancian, L.; Borges, B.D.; Silva, T.S.F.; Morellato, L.P.; Torres, R.D.S. Semantic segmentation of vegetation images acquired by unmanned aerial vehicles using an ensemble of ConvNets. In Proceedings of the 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Fort Worth, TX, USA, 23–28 July 2017; pp. 3787–3790.

291. Chen, F.; Yu, B. Earthquake-Induced Building Damage Mapping Based on Multi-Task Deep Learning Framework. *IEEE Access* **2019**, *7*, 181396–181404. [CrossRef]

292. Ma, H.; Liu, Y.; Ren, Y.; Yu, J. Detection of Collapsed Buildings in Post-Earthquake Remote Sensing Images Based on the Improved YOLOv3. *Remote Sens.* **2020**, *12*, 44. [CrossRef]

293. Nex, F.; Duarte, D.; Tonolo, F.G.; Kerle, N. Structural Building Damage Detection with Deep Learning: Assessment of a State-of-the-Art CNN in Operational Conditions. *Remote Sens.* **2019**, *11*, 2765. [CrossRef]

294. Bai, Y.; Gao, C.; Singh, S.; Koch, M.; Adriano, B.; Mas, E.; Koshimura, S. A Framework of Rapid Regional Tsunami Damage Recognition From Post-event TerraSAR-X Imagery Using Deep Neural Networks. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 43–47. [CrossRef]

295. Sublime, J.; Kalinicheva, E. Automatic Post-Disaster Damage Mapping Using Deep-Learning Techniques for Change Detection: Case Study of the Tohoku Tsunami. *Remote Sens.* **2019**, *11*, 1123. [CrossRef]

296. Bai, Y.; Mas, E.; Koshimura, S. Towards Operational Satellite-Based Damage-Mapping Using U-Net Convolutional Network: A Case Study of 2011 Tohoku Earthquake-Tsunami. *Remote Sens.* **2018**, *10*, 1626. [CrossRef]

297. Zhang, P.; Nascetti, A.; Ban, Y.; Gong, M. An implicit radar convolutional burn index for burnt area mapping with Sentinel-1 C-band SAR data. *ISPRS J. Photogramm. Remote Sens.* **2019**, *158*, 50–62. [CrossRef]

298. Potnis, A.V.; Shinde, R.C.; Durbha, S.S.; Kurte, K.R. Multi-Class Segmentation of Urban Floods from Multispectral Imagery Using Deep Learning. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 9741–9744.

299. Li, Y.; Martinis, S.; Wieland, M. Urban flood mapping with an active self-learning convolutional neural network based on TerraSAR-X intensity and interferometric coherence. *ISPRS J. Photogramm. Remote Sens.* **2019**, *152*, 178–191. [CrossRef]

300. Ichim, L.; Popescu, D. Flooded Areas Evaluation from Aerial Images Based on Convolutional Neural Network. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 9756–9759.

301. Gebrehiwot, A.; Hashemi-Beni, L.; Thompson, G.; Kordjamshidi, P.; Langan, T.E. Deep Convolutional Neural Network for Flood Extent Mapping Using Unmanned Aerial Vehicles Data. *Sensors* **2019**, *19*, 1486. [CrossRef] [PubMed]

302. Ghorbanzadeh, O.; Meena, S.R.; Blaschke, T.; Aryal, J. UAV-Based Slope Failure Detection Using Deep-Learning Convolutional Neural Networks. *Remote Sens.* **2019**, *11*, 2046. [CrossRef]

303. Mohajerani, Y.; Wood, M.; Velicogna, I.; Rignot, E. Detection of Glacier Calving Margins with Convolutional Neural Networks: A Case Study. *Remote Sens.* **2019**, *11*, 74. [CrossRef]

304. Baumhoer, C.A.; Dietz, A.J.; Kneisel, C.; Kuenzer, C. Automated Extraction of Antarctic Glacier and Ice Shelf Fronts from Sentinel-1 Imagery Using Deep Learning. *Remote Sens.* **2019**, *11*, 2529. [CrossRef]

305. Wang, C.; Zhang, H.; Wang, Y.; Zhang, B. Sea Ice Classification with Convolutional Neural Networks Using Sentinel-L Scansar Images. In Proceedings of the IGARSS 2018—2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 7125–7128.

306. Gao, Y.; Gao, F.; Dong, J.; Wang, S. Transferred Deep Learning for Sea Ice Change Detection From Synthetic-Aperture Radar Images. *IEEE Geosci. Remote Sens. Lett.* **2019**, *16*, 1655–1659. [CrossRef]

307. Zhang, W.; Witharana, C.; Liljedahl, A.K.; Kanevskiy, M. Deep Convolutional Neural Networks for Automated Characterization of Arctic Ice-Wedge Polygons in Very High Spatial Resolution Aerial Imagery. *Remote Sens.* **2018**, *10*, 1487. [CrossRef]

308. Huang, L.; Liu, L.; Jiang, L.; Zhang, T. Automatic Mapping of Thermokarst Landforms from Remote Sensing Images Using Deep Learning: A Case Study in the Northeastern Tibetan Plateau. *Remote Sens.* **2018**, *10*, 2067. [CrossRef]

309. Poliyapram, V.; Imamoglu, N.; Nakamura, R. Deep Learning Model for Water/Ice/Land Classification Using Large-Scale Medium Resolution Satellite Images. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 3884–3887.

310. Baumhoher, C.; Dietz, A.; Kniesel, C.; Paeth, H.; Kuenzer, C. Driving Forces of Circum-Antarctic Glacier and Ice Shelf Front Retreat over the Last Two Decades. *The Cryosphere Discussions* **2020**, Submitted.

311. Guirado, E.; Tabik, S.; Rivas, M.L.; Alcaraz-Segura, D.; Herrera, F. Whale counting in satellite and aerial images with deep learning. *Sci. Rep.* **2019**, *9*, 1–12. [CrossRef]

312. Bowler, E.; Fretwell, P.T.; French, G.; Mackiewicz, M. Using Deep Learning To Count Albatrosses From Space. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 10099–10102.

313. Kellenberger, B.; Marcos, D.; Tuia, D. Detecting mammals in UAV images: Best practices to address a substantially imbalanced dataset with deep learning. *Remote Sens. Environ.* **2018**, *216*, 139–153. [CrossRef]

314. Salberg, A. Detection of seals in remote sensing images using features extracted from deep convolutional neural networks. In Proceedings of the 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Milan, Italy, 26–31 July 2015; pp. 1893–1896.

315. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 26 June–1 July 2016; pp. 770–778. [CrossRef]

316. Szegedy, C.; Wei Liu.; Yangqing Jia.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 1–9. [CrossRef]

317. Zeiler, M.D.; Fergus, R. Visualizing and Understanding Convolutional Networks. In *Computer Vision—ECCV 2014*; Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T., Eds.; Springer International Publishing: Cham, Switzerland, 2014; pp. 818–833.

318. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2014**, arXiv:cs.CV/1409.1556.

319. Buscombe, D.; Ritchie, A.C. Landscape Classification with Deep Neural Networks. *Geosciences* **2018**, *8*, 244. [CrossRef]

320. Allauddin, M.S.; Kiran, G.S.; Kiran, G.R.; Srinivas, G.; Mouli, G.U.R.; Prasad, P.V. Development of a Surveillance System for Forest Fire Detection and Monitoring using Drones. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 9361–9363.

321. Li, L.; Zhang, S.; Wu, J. Efficient Object Detection Framework and Hardware Architecture for Remote Sensing Images. *Remote Sens.* **2019**, *11*, 2376. [CrossRef]

322. Zhang, T.; Zhang, X. High-Speed Ship Detection in SAR Images Based on a Grid Convolutional Neural Network. *Remote Sens.* **2019**, *11*, 1206. [CrossRef]

323. Liu, W.; Cheng, D.; Yin, P.; Yang, M.; Li, E.; Xie, M.; Zhang, L. Small Manhole Cover Detection in Remote Sensing Imagery with Deep Convolutional Neural Networks. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 49. [CrossRef]

324. He, H.; Yang, D.; Wang, S.; Zheng, Y.; Wang, S. Light encoder–decoder network for road extraction of remote sensing images. *J. Appl. Remote Sens.* **2019**, *13*, 1–11. [CrossRef]

325. Ji, H.; Gao, Z.; Mei, T.; Li, Y. Improved Faster R-CNN With Multiscale Feature Fusion and Homography Augmentation for Vehicle Detection in Remote Sensing Images. *IEEE Geosci. Remote Sens. Lett.* **2019**, *16*, 1761–1765. [CrossRef]

326. Ding, P.; Zhang, Y.; Deng, W.J.; Jia, P.; Kuijper, A. A light and faster regional convolutional neural network for object detection in optical remote sensing images. *ISPRS J. Photogramm. Remote Sens.* **2018**, *141*, 208–218. [CrossRef]

327. Zhang, C.; Sargent, I.; Pan, X.; Li, H.; Gardiner, A.; Hare, J.; Atkinson, P.M. An object-based convolutional neural network (OCNN) for urban land use classification. *Remote Sens. Environ.* **2018**, *216*, 57–70. [CrossRef]

328. Bischke, B.; Helber, P.; Borth, D.; Dengel, A. Segmentation of Imbalanced Classes in Satellite Imagery using Adaptive Uncertainty Weighted Class Loss. In Proceedings of the IGARSS 2018—2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 6191–6194.

329. Henry, C.; Azimi, S.M.; Merkle, N. Road Segmentation in SAR Satellite Images With Deep Fully Convolutional Neural Networks. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 1867–1871. [CrossRef]

330. Malof, J.M.; Collins, L.M.; Bradbury, K. A deep convolutional neural network, with pre-training, for solar photovoltaic array detection in aerial imagery. In Proceedings of the 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Fort Worth, TX, USA, 23–28 July 2017; pp. 874–877.

331. Shelhamer, E.; Long, J.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2014**, *39*, 640–651. [CrossRef] [PubMed]

332. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*; Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F., Eds.; Springer International Publishing: Cham, Switzerland, 2015; pp. 234–241.

333. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs. *arXiv* **2014**, arXiv:cs.CV/1412.7062.

334. Papandreou, G.; Chen, L.C.; Murphy, K.; Yuille, A.L. Weakly- and Semi-Supervised Learning of a DCNN for Semantic Image Segmentation. *arXiv* **2015**, arXiv:cs.CV/1502.02734.

335. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *40*, 834–848. [CrossRef]

336. Chen, L.C.; Papandreou, G.; Schroff, F.; Adam, H. Rethinking Atrous Convolution for Semantic Image Segmentation. *arXiv* **2017**, arXiv:cs.CV/1706.05587.

337. Zhou, L.; Zhang, C.; Wu, M. D-LinkNet: LinkNet With Pretrained Encoder and Dilated Convolution for High Resolution Satellite Imagery Road Extraction. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 192–1924.

338. Zhang, P.; Ke, Y.; Zhang, Z.; Wang, M.; Li, P.; Zhang, S. Urban Land Use and Land Cover Classification Using Novel Deep Learning Models Based on High Spatial Resolution Satellite Imagery. *Sensors* **2018**, *18*, 3717. [CrossRef]

339. Zhu, Q.; Zheng, Y.; Jiang, Y.; Yang, J. Efficient Multi-Class Semantic Segmentation of High Resolution Aerial Imagery with Dilated LinkNet. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 1065–1068.

340. Peng, B.; Li, Y.; Fan, K.; Yuan, L.; Tong, L.; He, L. New Network Based on D-Linknet and Densenet for High Resolution Satellite Imagery Road Extraction. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 3939–3942.

341. Liu, Z.; Feng, R.; Wang, L.; Zhong, Y.; Cao, L. D-Resunet: Resunet and Dilated Convolution for High Resolution Satellite Imagery Road Extraction. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 3927–3930.

342. Dong, S.; Zhuang, Y.; Yang, Z.; Pang, L.; Chen, H.; Long, T. Land Cover Classification From VHR Optical Remote Sensing Images by Feature Ensemble Deep Learning Network. *IEEE Geosci. Remote Sens. Lett.* **2020**, *17*, 1396–1400. [CrossRef]

343. Tao, Y.; Xu, M.; Lu, Z.; Zhong, Y. DenseNet-Based Depth-Width Double Reinforced Deep Learning Neural Network for High-Resolution Remote Sensing Image Per-Pixel Classification. *Remote Sens.* **2018**, *10*, 779. [CrossRef]

344. Fu, G.; Liu, C.; Zhou, R.; Sun, T.; Zhang, Q. Classification for High Resolution Remote Sensing Imagery Using a Fully Convolutional Network. *Remote Sens.* **2017**, *9*, 498. [CrossRef]

345. Guo, R.; Liu, J.; Li, N.; Liu, S.; Chen, F.; Cheng, B.; Duan, J.; Li, X.; Ma, C. Pixel-Wise Classification Method for High Resolution Remote Sensing Imagery Using Deep Neural Networks. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 110. [CrossRef]

346. Liu, Y.; Piramanayagam, S.; Monteiro, S.T.; Saber, E. Semantic segmentation of multisensor remote sensing imagery with deep ConvNets and higher-order conditional random fields. *J. Appl. Remote Sens.* **2019**, *13*, 1–23. [CrossRef]

347. Hu, J.; Shen, L.; Sun, G. Squeeze-and-Excitation Networks. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA 18–23 June; pp. 7132–7141. [CrossRef]

348. Wu, Y.; Zhang, R.; Zhan, Y. Attention-Based Convolutional Neural Network for the Detection of Built-Up Areas in High-Resolution SAR Images. In Proceedings of the IGARSS 2018—2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 4495–4498.

349. Yang, H.; Wu, P.; Yao, X.; Wu, Y.; Wang, B.; Xu, Y. Building Extraction in Very High Resolution Imagery by Dense-Attention Networks. *Remote Sens.* **2018**, *10*, 1768. [CrossRef]

350. Wang, H.; Wang, Y.; Zhang, Q.; Xiang, S.; Pan, C. Gated Convolutional Neural Network for Semantic Segmentation in High-Resolution Images. *Remote Sens.* **2017**, *9*, 446. [CrossRef]

351. Sun, W.; Wang, R. Fully Convolutional Networks for Semantic Segmentation of Very High Resolution Remotely Sensed Images Combined With DSM. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 474–478. [CrossRef]

352. Zhang, Y.; Xia, G.; Wang, J.; Lha, D. A Multiple Feature Fully Convolutional Network for Road Extraction From High-Resolution Remote Sensing Image Over Mountainous Areas. *IEEE Geosci. Remote Sens. Lett.* **2019**, *16*, 1600–1604. [CrossRef]

353. Bergado, J.R.; Persello, C.; Stein, A. Fusenet: End- to-End Multispectral Vhr Image Fusion and Classification. In Proceedings of the IGARSS 2018—2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 2091–2094.

354. Chen, Y.; Li, C.; Ghamisi, P.; Jia, X.; Gu, Y. Deep Fusion of Remote Sensing Data for Accurate Classification. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 1253–1257. [CrossRef]

355. Rezaee, M.; Mahdianpari, M.; Zhang, Y.; Salehi, B. Deep Convolutional Neural Network for Complex Wetland Classification Using Optical Remote Sensing Imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 3030–3039. [CrossRef]

356. Kroupi, E.; Kesa, M.; Navarro-Sánchez, V.D.; Saeed, S.; Pelloquin, C.; Alhaddad, B.; Moreno, L.; Soria-Frisch, A.; Ruffini, G. Deep convolutional neural networks for land-cover classification with Sentinel-2 images. *J. Appl. Remote Sens.* **2019**, *13*, 1–22. [CrossRef]

357. Girshick, R.B.; Donahue, J.; Darrell, T.; Malik, J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 24–27 June 2014; pp. 580–587.

358. Ren, S.; He, K.; Girshick, R.B.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *39*, 1137–1149. [CrossRef]

359. Zhao, Y.; Zhao, J.; Zhao, C.; Xiong, W.; Li, Q.; Yang, J. Robust Real-Time Object Detection Based on Deep Learning for Very High Resolution Remote Sensing Images. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 1314–1317.

360. Zhang, Z.; Liu, Y.; Liu, T.; Lin, Z.; Wang, S. DAGN: A Real-Time UAV Remote Sensing Image Vehicle Detection Framework. *IEEE Geosci. Remote Sens. Lett.* **2019**, 1–5. Early Access. [CrossRef]

361. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.E.; Fu, C.Y.; Berg, A.C. SSD: Single Shot MultiBox Detector. In *Computer Vision – ECCV 2016*; Leibe, B., Matas, J., Sebe, N., Welling, M., Eds.; Springer: Cham, Switzerland, 2016; pp. 21–37.

362. Redmon, J.; Divvala, S.K.; Girshick, R.B.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 8–10 June 2015; pp. 779–788.

363. Redmon, J.; Farhadi, A. YOLO9000: Better, Faster, Stronger. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2016; pp. 6517–6525.

364. Redmon, J.; Farhadi, A. YOLOv3: An Incremental Improvement. *arXiv* **2018**, arXiv:cs.CV/1804.02767.

365. Lin, T.Y.; Dollár, P.; Girshick, R.B.; He, K.; Hariharan, B.; Belongie, S.J. Feature Pyramid Networks for Object Detection. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2016; pp. 936–944.

366. Liu, J.; Yang, S.; Tian, L.; Guo, W.; Zhou, B.; Jia, J.; Ling, H. Multi-Component Fusion Network for Small Object Detection in Remote Sensing Images. *IEEE Access* **2019**, *7*, 128339–128352. [CrossRef]

367. Gao, F.; Shi, W.; Wang, J.; Yang, E.; Zhou, H. Enhanced Feature Extraction for Ship Detection from Multi-Resolution and Multi-Scene Synthetic Aperture Radar (SAR) Images. *Remote Sens.* **2019**, *11*, 2694. [CrossRef]

368. Tayara, H.; Chong, K.T. Object Detection in Very High-Resolution Aerial Images Using One-Stage Densely Connected Feature Pyramid Network. *Sensors* **2018**, *18*, 3341. [CrossRef]

369. Bao, S.; Zhong, X.; Zhu, R.; Zhang, X.; Li, Z.; Li, M. Single Shot Anchor Refinement Network for Oriented Object Detection in Optical Remote Sensing Imagery. *IEEE Access* **2019**, *7*, 87150–87161. [CrossRef]

370. Wang, P.; Sun, X.; Diao, W.; Fu, K. Mergenet: Feature-Merged Network for Multi-Scale Object Detection in Remote Sensing Images. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 238–241.

371. Zhang, W.; Jiao, L.; Liu, X.; Liu, J. Multi-Scale Feature Fusion Network for Object Detection in VHR Optical Remote Sensing Images. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 330–333.

372. Huang, H.; Huo, C.; Wei, F.; Pan, C. Rotation and Scale-Invariant Object Detector for High Resolution Optical Remote Sensing Images. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 1386–1389.

373. Yao, Q.; Hu, X.; Lei, H. Geospatial Object Detection In Remote Sensing Images Based On Multi-Scale Convolutional Neural Networks. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 1450–1453.

374. Liu, N.; Cui, Z.; Cao, Z.; Pi, Y.; Lan, H. Scale-Transferrable Pyramid Network for Multi-Scale Ship Detection in Sar Images. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 1–4.

375. AL-Alimi, D.; Shao, Y.; Feng, R.; Al-qaness, M.A.A.; Elaziz, M.A.; Kim, S. Multi-Scale Geospatial Object Detection Based on Shallow-Deep Feature Extraction. *Remote Sens.* **2019**, *11*, 2525. [CrossRef]

376. Pang, J.; Li, C.; Shi, J.; Xu, Z.; Feng, H. $\mathcal{R}^2$-CNN: Fast Tiny Object Detection in Large-Scale Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 5512–5524. [CrossRef]

377. Yan, J.; Wang, H.; Yan, M.; Diao, W.; Sun, X.; Li, H. IoU-Adaptive Deformable R-CNN: Make Full Use of IoU for Multi-Class Object Detection in Remote Sensing Imagery. *Remote Sens.* **2019**, *11*, 286. [CrossRef]

378. Chen, S.; Zhan, R.; Zhang, J. Geospatial Object Detection in Remote Sensing Imagery Based on Multiscale Single-Shot Detector with Activated Semantics. *Remote Sens.* **2018**, *10*, 820. [CrossRef]

379. Wang, G.; Zhuang, Y.; Wang, Z.; Chen, H.; Shi, H.; Chen, L. Spatial Enhanced-SSD For Multiclass Object Detection in Remote Sensing Images. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 318–321.

380. Lin, Z.; Ji, K.; Leng, X.; Kuang, G. Squeeze and Excitation Rank Faster R-CNN for Ship Detection in SAR Images. *IEEE Geosci. Remote Sens. Lett.* **2019**, *16*, 751–755. [CrossRef]

381. Wang, C.; Bai, X.; Wang, S.; Zhou, J.; Ren, P. Multiscale Visual Attention Networks for Object Detection in VHR Remote Sensing Images. *IEEE Geosci. Remote Sens. Lett.* **2019**, *16*, 310–314. [CrossRef]

382. Cao, Z.; Li, X.; Zhao, L. Object Detection in VHR Image Using Transfer Learning with Deformable Convolution. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 326–329.

383. Xu, Z.; Xu, X.; Wang, L.; Yang, R.; Pu, F. Deformable ConvNet with Aspect Ratio Constrained NMS for Object Detection in Remote Sensing Imagery. *Remote Sens.* **2017**, *9*, 1312. [CrossRef]

384. Zhou, C.; Ye, H.; Hu, J.; Shi, X.; Hua, S.; Yue, J.; Xu, Z.; Yang, G. Automated Counting of Rice Panicle by Applying Deep Learning Model to Images from Unmanned Aerial Vehicle Platform. *Sensors* **2019**, *19*, 3106. [CrossRef]

385. Cai, Z.; Vasconcelos, N. Cascade R-CNN: Delving Into High Quality Object Detection. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 6154–6162.

386. Tian, Z.; Wang, W.; Zhan, R.; He, Z.; Zhang, J.; Zhuang, Z. Cascaded Detection Framework Based on a Novel Backbone Network and Feature Fusion. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 3480–3491. [CrossRef]

387. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R.B. Mask R-CNN. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2980–2988.

388. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *Advances in Neural Information Processing Systems 32*; Wallach, H., Larochelle, H., Beygelzimer, A., Alché-Buc, F., Fox, E., Garnett, R., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2019; Volume 32, pp. 8024–8035.

389. Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G.S.; Davis, A.; Dean, J.; Devin, M.; et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. Available online: https://www.tensorflow.org/ (accessed on 27 July 2020).

390. Chollet, F. Keras. Available online: https://keras.io (accessed on 27 July 2020).

391. Jia, Y.; Shelhamer, E.; Donahue, J.; Karayev, S.; Long, J.; Girshick, R.; Guadarrama, S.; Darrell, T. Caffe: Convolutional architecture for fast feature embedding. In Proceedings of the 22nd ACM International Conference on Multimedia,Orlando, FL, USA, 3–7 November 2014; pp. 675–678.

392. Vedaldi, A.; Lenc, K. MatConvNet – Convolutional Neural Networks for MATLAB. In Proceeding of the 23rd ACM Int. Conf. on Multimedia, Brisbane, Australia, 26–30 October 2015; pp. 689–692.

393. Schmitt, M.; Hughes, L.H.; Qiu, C.; Zhu, X.X. SEN12MS—A Curated Dataset of Georeferenced Multi-Spectral Sentinel-1/2 Imagery for Deep Learning and Data Fusion. *arXiv* **2019**, arXiv:cs.CV/1906.07789.

394. Sandler, M. MobileNet V2 ImageNet Checkpoints. Available online: https://github.com/tensorflow/models/blob/master/research/slim/nets/mobilenet/README.md (accessed on 1 April 2020).

395. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L. MobileNetV2: Inverted Residuals and Linear Bottlenecks. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 4510–4520. [CrossRef]

396. Tan, M.; Chen, B.; Pang, R.; Vasudevan, V.; Le, Q.V. MnasNet: Platform-Aware Neural Architecture Search for Mobile. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 16–20 June 2019; pp. 2815–2823.

397. Zoph, B.; Le, Q.V. Neural Architecture Search with Reinforcement Learning. *arXiv* **2016**, arXiv:cs.LG/1611.01578.

398. Zoph, B.; Vasudevan, V.; Shlens, J.; Le, Q.V. Learning Transferable Architectures for Scalable Image Recognition. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 8697–8710. [CrossRef]

399. Tan, M.; Le, Q.V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In Proceedings of the 36th International Conference on Machine Learning, ICML 2019, Long Beach, CA, USA, 9–15 June 2019; pp. 6105–6114.

400. Wang, J.; Zhong, Y.; Zheng, Z.; Ma, A.; Zhang, L. RSNet: The Search for Remote Sensing Deep Neural Networks in Recognition Tasks. *IEEE Trans. Geosci. Remote Sens.* **2020**, 1–15. Early Access. [CrossRef]