Hochschule für Technik Stuttgart

University of Applied Sciences

Master of Science Programme
Photogrammetry and Geoinformatics
Master Thesis
Winter Term 2021/2022

Evaluation of the utility of radar data to provide model parameters for energy system analysis

by

Fatjon Dine

Supervisors:

Prof. Dr. Michael Hahn (HFT)
Dr. Thomas Esch (DLR)
Dr. Annekatrin Metz-Marconcini (DLR)
Dr. Karl-Kiên Cao (DLR)

Evaluation of the utility of radar data to provide model parameters for energy system analysis.

by

Fatjon Dine

A dissertation presented in partial fulfillment of the requirements for the degree of Master of Science in the Department of Geomatics, Computer Science and Mathematics, Stuttgart University of Applied Sciences

Declaration

The following Master thesis was prepared in my own words without any additional help. All used sources of literature are listed at the end of the thesis.

I hereby grant to Stuttgart University of Applied Sciences permission to reproduce and to distribute publicly paper and electronic copies of this document in whole and in part.

Stuttgart, 30.03.2022

Dine Fatjon

Approved by:

Prof. Dr. Michael Hahn

To my beloved family.

For their endless love, support and encouragement.

Evaluation of the utility of radar data to provide model parameters for energy system analysis.

Abstract

Based on recent advances, deep learning (DL) has proven to be able to achieve an outstanding performance in remote sensing. Nevertheless, has shown to be limited to the evaluation of the optical data. Beside many previous researches to introduce DL in Synthetic Aperture Radar (SAR), analysing the energy systems by means of above-mentioned techniques offers a huge potential to be explored. From the satellite data, energy systems parameters can be derived which will lead to an improvement in quality and completeness for the existing databases. In this context, this research contributes to the generation of a new worldwide open database, by means of a uniform methodology which can be updated continuously. Hence, in this master thesis a novel methodology is developed for automatic extraction, provision of information on the type of the plant and the exact geo-location for two main energy generators respectively *Wind Turbine* and *Coal-fired Power Plants*. Simultaneously paying special heed on being competitive in accuracy and computational time. For this purpose, *Faster R-CNN* a state-of-the-art DL framework for object detection composed with *Res Net 50 + FPN* as a backbone are utilized to model the architecture. The dataset annotation is designed to be automatic and utilize manually created geolocations plus open databases e.g Global Power Plant Database (GPPD) to compass this complex and time-consuming process faster. Suitable areas with extensive geographical scope are selected for implementing the developed approach. Moreover, analysis and evaluation of the achieved results graphically shows the algorithm capability to work in different geographical scales. Finally, the conducted accuracy assessment reveals the capability of the developed methodology in this master thesis to ensure an accuracy of **87.71%** in large scale applications for individual Wind Turbine detection, and **92.41%** in large scale applications for Coal-fired Power Plants detection.

Keywords: Synthetic Aperture Radar, Convolutional Neural Networks, Target Detection, Energy System Analysis.

Table of Content

List of Figures

List of Tables

Abbreviations

AI	Artificial Intelligence
AOI	Area of interest
AP	Average Precision
AR	Average Recall
ASTER	Advanced Spaceborne Thermal Emission and Reflection Radiometer
CNN	Convolutional Neural Network
CO2	Carbon dioxide
CPU	Central processing unit
CSV	Comma Separated Value
DEM	Digital Elevation Model
DL	Deep Learning
DLR	German Aerospace Center
EO	Earth Observation
ESA	European Space Agency
FN	False Negative
FP	False Positive
GDDR	Graphics Double Data Rate
GEE	Google Earth Engine
GIS	Geographic Information System
GNSS	Global Navigation Satellite System
GPPD	Global Power Plant Database
GPU	Graphics processing unit
GRD	Ground Range Detection

H	Horizontal
IoU	Intersection-over-Union
ISE	Institute for Solar Energy Systems
it	Iteration
km	Kilometers
LUTs	Look Up Tables
m	Meters
ML	Machine Learning
MS COCO	Microsoft Common Objects in Context
MSI	Multi-Spectral Instrument
NASA	National Aeronautics and Space Administration
NN	Neural Network
POD	Precise Orbit Determination
px	Pixels
RAM	Random Access Memory
R-CNN	Region-based Convolutional Neural Network
ReLu	Rectified Linear Unit
ResNet	Residual Neural Network
ROI	Region of Interest
s	Seconds
SAR	Synthetic Aperture Radar
SGD	Stochastic gradient descent
SLC	Single Look Complex
SRTM	Shuttle Radar Topography Mission
TIFF	Tag Image File Format

TN	True Negative
TP	True Positive
TUM	Technical University of Munich
TWh	Terawatt-hours
USWTDB	United States Wind Turbine Database
V	Vertical
VGG	Visual Geometry Group
VOC	Visual Object Classes
w.r.t	with respect to
WRI	World Research Institute

1 Introduction

The introduction chapter outlines in a detailed way, divided into sections, the aim, objectives and the research importance of this master thesis. The Section 1.1 presents a definition concerning the energy processes, energy transition, its importance at present and for the future, followed by the challenges which are confronting the energy transition, afterwards to persist with the advantages and limitations of existing databases. Further, the Section 1.2 concentrate on the problem statement and research questions that arise on this master thesis. Furthermore, Section 1.3 introduces the possible solutions and new ideas the master thesis tends to bring which will facilitate the needs utilizing of state-of-the-art technologies.

1.1 Energy Transition and Energy System Analysis

Energy is fundamental ingredient which makes human life viable. Nowadays its importance is exponentially growing in every aspect as people get richer and populations increase, there is no industrial, agricultural, health, domestic, or any other sort of process that doesn't require a degree of external energy. As every sort of process in the daily life need energy supply, it therefore causes a rapid increase on the demand for energy worldwide.

Energy is divided as two basic categories therefore is either Potential or Kinetic. Potential energy is a form of the stored energy, and Kinetic energy explained as the energy in the motion of objects, waves etc. These two basic categories are at the base of many different forms of energy such as light energy, heat energy, electrical energy, chemical energy, nuclear energy or atomic energy [1]. Afterwards those forms of energy are at the base of creating electricity. An instance of the potential energy is the form of energy that has stored potential and can be used afterwards, e.g hydroelectricity generators or solar systems. On the other hand, an instance of the kinetic energy is the electricity production created by motion of air e.g wind turbines.

The electricity generation is done using infinite or finite sources, designated also as renewable and non-renewable energy. Renewable energy come from natural infinite sources such as wind, water or sun that are constantly replenished, non-renewable energies come from the finite resources of the nature such as oil, gas and coal which takes long time to replenish. The energy consumption has increased every year and from the statistics published at the journal Our World in Data by the authors Hannah Ritchie and Max Roser (see **Figure 1**) shows that how fossil fuels (coal, oil, gas) play a dominant role by offering a base load in global energy systems [2].

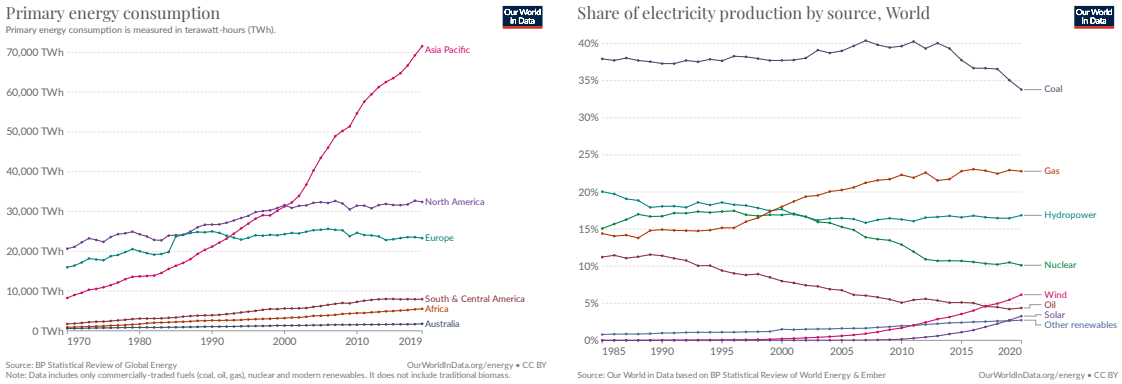


Figure 1. *Left*: Primary electricity consumption in TWh (1970-2019)
Right: Share of electricity in % World (1985-2020)
(Source: Our World in Data[1])

The big process began with the Industrial Revolution [2] in the 18th to 20th century which changed from the handicraft economy replacing it with industry and machine manufacturing, hence increasing the demand for energy and non-renewable energies still up to the present time play a key role as a fundamental supplier for all the technological, social, economic and development progress.

For two important reasons (coal, gas, oil) can not continuously be a fundamental supplier for a very long time, because first it is a finite resource and second because of the several negative impacts, such as: the CO2 emission making them the largest driver of global climate change, and as a major contributor in the air pollution which causes millions of deaths every year [2]. As a solution to avoid of these negative impacts can be a switch from the fossil-based system to the renewable energy sources through the energy transition process.

Energy transition is a process which consist on a structural transformation of the energy system, thus the path of the global energy has to shift the transform of the energy supply which is actually based on the non-renewables (fossil-based), instead to renewables (zero-carbon). As considered worldwide, energy transition along with expansion of the electricity generation are the most important challenges of the 21st century. On the United Nations Climate Change Conference which was held on November-December 2015 in Paris was reached an agreement for the reduction of climate change until 2050, and part of the climate change was considered energy transition as well [3].

[1] Our World in Data https://ourworldindata.org/grapher/primary-energy-cons Accessed 19 11 2021

[2] Britannica, The Editors of Encyclopaedia. "Industrial Revolution". Encyclopedia Britannica, 21 Jul. 2021, https://www.britannica.com/event/Industrial-Revolution. Accessed 19 11 2021.

In this agreement climate change was recognized as an urgent and potential irreversible threat to human societies and the planet. To meet the requirements for energy supply in a fast-growing world, requires a flexible design in energy sectors to be able to produce, transport and redistribute the energy between. This takes a significant range of assessment criteria and scientific analyses of energy infrastructures into account, in order to achieve it sustainably and in harmony with social, economic and environmental implications.

In the publication concerning the energy transition the Federal Ministry for Economic Affairs and Energy in Germany states "*The big challenge now is to organise the wholesale electricity market so that the electricity supply will remain secure and reliable in future*." [4, p. 17]

Energy infrastructure with a key impact on the growing economy is interpreted in different contexts and has no standard definition. As the European Commission [5] describes, it includes in particular, transmission, distribution and storage infrastructure for electricity generators e.g renewable sources (wind, solar, hydropower) and non-renewable sources (gas, coal, oil, etc.)

Energy systems require analysis in order to support the change on supply and demand structure of the energy sector. In addition, this will contribute on the management and operation of power plants, energy networks and storage systems. Through energy system analysis researches efficiently improve the methodology for finding the most cost-optimal paths to reduce CO2 emission in the energy systems. This is done by means of energy models that consist on components which allow them make analysis of the energy systems based on three different geographical scales as described by Fraunhofer Institute for Solar Energy Systems (Fraunhofer ISE) [6]:

- National and international level of energy systems
- Regional, city, community level of energy systems
- Industrial and municipal level of energy systems

Thus, as those change effects cannot be tested in real power systems, analysis are carried by using energy models which in other words can be considered as laboratory scale simulations. The conclusion achieved out of energy models that repress significant analysis, assist on the provision of the needed information for the purpose of decision-making. According to Chair of Renewable and Sustainable Energy Systems at Technical University of Munich (TUM) [7] some of them are marked as:

- Authorities or policy makers to develop their masterplans at different scales
- Utility companies and commercialized industrial technology
- Scarcity of resources
- Climate change target agreements

1.1.1 The importance of data for energy system modelling

In order to produce significantly accurate predictions and scenario analyses using energy system models for purpose to contribute on the improvement of policy advice and facilitation of the controversial debates, provision of precise data on energy infrastructures constitute an important role [6]. Hereby, different parameters for instance: exact geo-location, type of plant and power generation capacity of the energy infrastructure are essential representation for creation of a high-quality electricity feed-in forecasts, have a sustainability on the energy transition process and as well a direct impact in energy industry.

Depending on the scale and the precision needed on the research, this input data constitutes a key component for the energy transition. But usually, they are not free or as World Research Institute (WRI) states in the technical note “A Global database of Power Plants” [8] available in a very poor quality. Basically, the quality of a database can be attributed to its purpose of use. Thus, mainly quality measures can be considered the geographical coverage, completeness of information or accuracy geo-location (more extensive information is shown on Subsubsection 1.1.1.2). Nowadays many attempts are done to provide open databases e.g Global Power Plant Database (GPPD) [8] which can contain this required information and can be used as a reference database, but until now they have not succeeded to be reliable due to lack of standards or continuous updates.

On one hand, Subsubsection 1.1.1.1 describes the advantages of having an open database for the power plants which can be updated continuously. Further, proceeding with the impact an open database has for helping on a more sustainable energy transition process. On the other hand, Subsubsection 1.1.1.2 introduces the limitations of the existing open and commercial databases.

1.1.1.1 Advantages

Based on the importance of energy system databases as framed by WRI [8] is concluded that being able to possess a high-accuracy global power plant database will benefit in many aspects which will increase the efficiency on building the energy models and thereafter use them on energy transition process. Some advantages to be mentioned are:

- Independent database with continuously updated information.
- Customized extracted information depending on need.
- Comprehensive and easy navigation between the individual energy systems.
- Comparison and draw insights between analysis.
- Comprehend world’s power systems as well their impacts on development and environmental challenges.

1.1.1.2 Limitations

Although high-accuracy global power plant database has many advantages and plays a vital role for many analyses on energy systems, limitations for establishing and global database to be truly comprehensive or fully open still remains locked. Mainly existing databases e.g GPPD [8] fetch information for the power plants from other 3rd party or national sources, the reliability of their database strongly depends on the reliability of the data sources they identify and aggregate. Furthermore, including reporting data source delay and the conversion from country specific data format into adapting them to a common standard for all the sources. **Table 1** shows an overview of some selected available databases for and the limitations which exist on those databases.

Table 1. Overview of selected open power plant databases

Data Source	URL	Last Update	Limitations
Global Energy Observatory	http://globalenergyobservatory.org/	2018	No regular updates, not all countries are covered.
Enipedia	https://old.datahub.io/dataset/enipedia	2016	Low accurate geolocation and no regular updates.
Global Power Plant Database	https://datasets.wri.org/dataset/globalpowerplantdatabase	2021	Different ranks for data quality, depending on source.
Carbon Monitoring for Action	http://carma.org/	2012	Low accurate geolocation and regular updates, small set of indicators are covered.
Open Power System Data	https://data.open-power-system-data.org/	2020	Data fetched from 3rd party, unclear update frequency.
ENTSO-E Transparency Platform	https://transparency.entsoe.eu/	2015	Data are from 2015 and onwards, no geolocation for all the sources.
World Electric Power Plants Database	https://doi.org/10.7910/DVN/OKEZ8A	2017	Limited access to certain files, unclear update frequency

As some of the limitations from open databases can be mentioned [8]:

- Data availability (of small or renewable power plants)
- No data available or not recently updated information for every country
- Lack of reporting standards on open databases (different countries report in different formats)

As described in the technical note named "A Global database of Power Plants" [8] written at WRI by a group of researchers concerning the process of the database matching and design, they identify this process as the most complex one, because as the information for the power plants comes from different sources an adaptation into a common format is vital for this process. But the lack of several information e.g geo-location can require another additional work and this process becomes time consuming and complex. In **Figure 2** it is shown a donut chart of the geolocated capacity by source used by WRI for the creation of the GPPD. Some of existing databases have a lack on public access and some others in terms of spatial coverage, completeness, transnational consistency or location information.

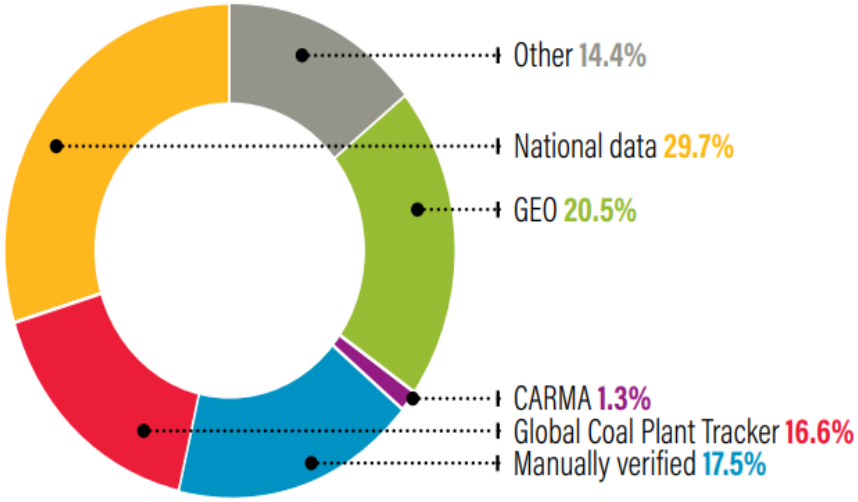


Figure 2. Percent of Geolocated Capacity by source, as of August 2020
(Source: WRI, March 2021 [8, p. 7])

1.2 Problem Statement

On one hand, based on the conducted research the limitations (see Subsubsection 1.1.1.2) which appears on the existing open databases are shown and on the other hand, as well the advantages (see Subsubsection 1.1.1.1) an open database could offer. Based on those indicators plus the importance it has on the energy transition process and not only, allows on considering the provision of a large-scale database with parameters on energy systems as a very important research topic.

Despite the tentative for generating a large-scale database mainly from 3rd party sources of the open databases mentioned in **Table 1**, this process is still complex and not sustainable.

Hence, a novel solution is required in order to answer two important research questions:

- Can nowadays the state-of-the-art technologies such as SAR imagery and DL techniques provide a valuable resource for the creation of a database which contains onshore energy system parameters?
- Can the chosen approach ensure high-detection accuracy also for large scale applications?

An answer to those research questions will strongly be a candidate novel solution for assisting on filling the existing gap and thus continuously generate an updated database. Further this solution will efficiently contribute on the energy transition process.

Important to note is that the research tends to make use of the available resources which are open access, hence the methodology will be capable to utilize the proposed resources without restrictions.

1.3 Possible solutions

This issue of creating an open database which information will be always up to date and have no dependencies or limitations as stated in Subsubsection 1.1.1.2, nowadays with the state-of-the-art technologies seems to have a solution.

As a starting point towards the solution are the satellite images, which nowadays are easily accessible and continuously available in very high spatial and temporal resolution. Spatial resolution refers to how detailed an object is acquired by the sensor during an observation. Temporal resolution refers to the amount of time a sensor requires in order to acquire the same object. Usually in remote sensing this is defined as sensor (satellite) revisit time [9]. This valuable resource for the data needed is made available with the European Space Agency (ESA) remote sensing satellite program called SENTINEL and can be downloaded for free. On the other hand, the state-of-the art on Deep Learning (DL) and Convolutional Neural Networks (CNNs) (see Section 2.2) and their applications are presenting the last advances in artificial neural networks across a wide range, and accomplishing astonishing results in different fields of applications.

A combination of those two resources at first was more focused on using Multi Spectral Imagery (MSI) [10] from Sentinel-2 [11] which has a mission to observe the Earth and continuously provide MSI at different resolution from 10-60m depending on the band. This valuable information

helped researchers to produce a more accurate on their daily tasks of land cover and land use classification and not only.

Although this method combination can produce high-accuracy results in general for some certain tasks the imagery can not produce enough accurate information. An instance of those tasks can be mentioned as most challenging the areas where shadow is present e.g next to building, trees (forest), bridges, etc. The acquisition of information in those areas is relatively challenging where spatial resolution mentioned in the above paragraphs constitute an important factor for a stable product.

Hence, scientists introduced Synthetic Aperture Radar (SAR) imagery which is retrieved from Sentinel-1 radar satellite constellation with DL technologies. Ever since the first applications the results achieved seem promising e.g [12] [13], therefore this huge potential in the field which is undiscovered remains unlocked.

Many different applications mainly for military purpose were done at first e.g. ship or vessel detection, or other military observatory applications, but now SAR imaging applications are being advanced by the use of neural networks and a very important tool due the unique characteristics of: weather independence, 24/7 information acquisition, long distance working.

Similar to the research questions formulated on this master thesis (see Section 1.2) are conducted for offshore wind turbine detection researches, but there until now there is no methodology which is capable to produce a large-scale onshore database for different energy systems types and their parameters.

This master thesis aims at identifying energy system parameters, in particular: type of the power plant, exact geo-location by means of DL techniques and SAR Imagery. The provision of SAR images utilized in this master thesis is done by German Aerospace Center (DLR).

This research focus lies on developing the methodology for a fully automated methodology on the extraction of the onshore energy generators. As a result, this contributes to the generation of a new worldwide open database, created by a uniform methodology which can be updated continuously. The energy generators in which this master thesis focusses to detect, include in the database are Wind Turbines and Coal Power Plants. However, it provides the bases for opening new perspectives of applying the methodology on the other remaining energy infrastructures.

2 Theoretical Background

This chapter aim to briefly explain the main components that assist to reach the goal of this master thesis. By dividing them into sections an introduction each of those components is made in a way that the reader will be able to understand the contribution of those components and their importance on the process. In the Section 2.1 a brief introduction concerning the definition of SAR imagery and Sentinel-1, how they are acquired by satellites and provided as open data for community use. Furthermore, in the Section 2.2 the state-of-the-art technologies in CNNs, an overview in DL techniques associated with the advantages and most common tasks those techniques are applied. Additionally, on Section 2.3 the state-of-the-art DL architectures used for object detection tasks, their supported pre-trained networks are presented and the accuracy assessment.

2.1 Synthetic Aperture Radar Imagery

2.1.1 Definition

In remote sensing exist two types of sensors which are passive and active sensors [14]. Passive sensors defined as passive because of the capability to detect only when the natural energy i.e sunlight is available. Active sensors are the type of sensors which are independent from the natural illumination and have capability to generate their own energy source and use it for illumination. Thus, a signal is transmitted from the sensor which afterwards is reflected on Earth surface and received back to sensor. **Figure 3 (a)** shows an illustration of passive sensors and **Figure 3 (b)** illustrates how active sensors are independent to natural illumination.

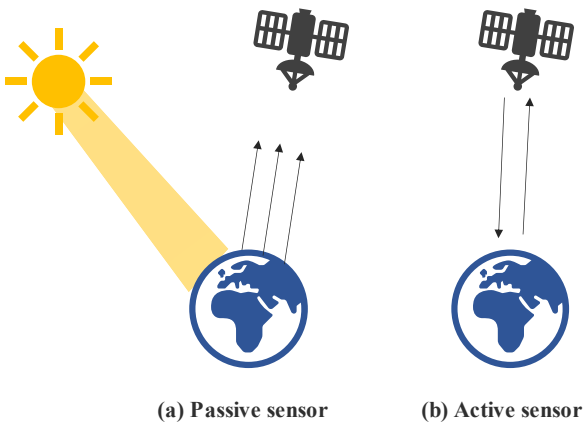


Figure 3. An illustration of passive and active sensors

The typical EO sensors are using optical devices and are launched in the sun-synchronous orbit, which means that they follow the illumination of the sun and pass over the target at the same time

[11, p. 18]. Being passive sensors optical systems are limited to overcome issues such as e.g penetrate clouds or acquire information during the night. Hence, those systems use the morning time to pass over the target, this allows them to prevent the above-mentioned issues.

The utilization of radio waves categorizes **RA**dio **D**etection **A**nd **R**anging (RADAR) sensors as active sensors (see **Figure 3**). For this reason, those type of sensors are characterized by the capability to detect objects at far range without being impended by lack of illumination. According to the online encyclopedia Encyclopaedia Britannica [3] the origin of radar principle has its beginnings on the late 1880s where the German physicist Heinrich Hertz conducted the experiments on electromagnetic radiation [15]. The conducted experiments at that time as Encyclopaedia Britannica states verified an earlier theory wrote by Scottish physicist James Clerk Maxwell [15]. Further Encyclopaedia Britannica states that: "*The first real introduction radar had it during the World War II with surveillance and detection tasks, where many countries developed radar sensors for military purposes*" [15].

According to Encyclopaedia Britannica contributors [15] in the early 1950s a new active sensor was introduced named SAR, utilized in airborne as well for space borne systems. Furthermore, they frame SAR systems were capable only to reach the high-state of development only during the progress in digital processing and advances in technology in early 1980s. This progress in technology made SAR systems to get broader usage for military, civil, or Earth Observation applications [15].

SAR as an active remote sensing system transmits radio pulses towards Earth surface. Afterwards, those radio pulses are reflected at the Earth surface and received back again from the sensor (see **Figure 3**), thus detecting Earth's surface physical characteristics. As described in the article about SAR systems by Moreira et al [16] the utilized transmitted microwave pulses of SAR systems are modulated in the frequency domain and they are designated as chirp signals. This means that the frequency of signal increases or decreases with the time. In addition, the received backscattered signals are stored along the synthetic aperture [16]. The characteristics of having a synthetic aperture on transmitting radio pulses obviously allows the SAR system to have a proportional resolution. Which means SAR is able to distinguish between two or more targets on the same bearing but at different ranges [16].

Spaceborne SAR systems instead of using optical devices they have antennas and high-frequency electronics which makes them able to overcome these issues and ensure high performance [16, pp. 6-11]. Their ability to have day and night data acquisition, penetrate clouds, being independent

[3] Encyclopaedia Britannica, https://www.britannica.com/ Accessed on 25 03 2022

from natural illumination make them to be very competitive on state-of-the art technologies for providing high-resolution image on-demand. The utilization of long wavelengths at range from centimetres to meters, which is then translated in different bandwidth gives radar sensors many special properties such the ability to penetrate the clouds and sub surfaces [17]. **Figure 4** shows an illustration of different spectral domains including the operational bands of SAR systems.

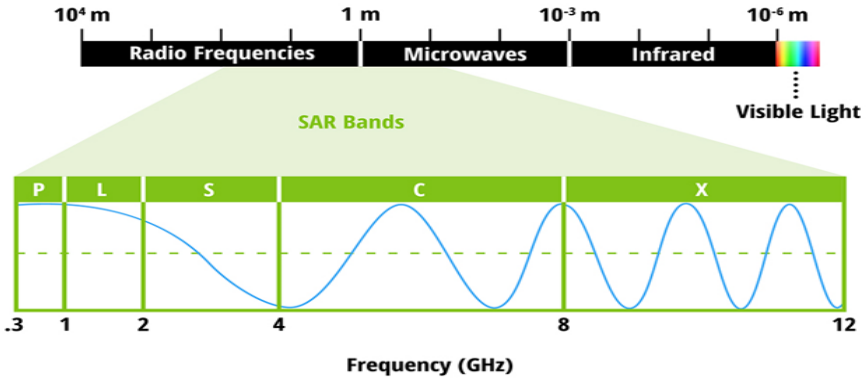


Figure 4. Electromagnetic spectrum of SAR systems
(Source: NASA SAR Handbook [17])

As SAR operates in different bandwidth, when designing a SAR system, it is apparent that the selection of the frequency band is done by considering many factors such as: the application intention of the sensor, its limited power available for spaceborne SAR and the availability of small components and atmospheric attenuation [18, pp. 25-28]. Atmospheric attenuation increases with the increase of the frequency of the radio waves; thus, radio waves might be attenuated by absorption in oxygen and water, and attenuated by scattering in liquid or solid water particles [18, pp. 25-28]. **Table 2** shows SAR imagery operational band, frequencies, wavelength and the relevant field where is applicable.

Table 2. Frequency bands used in the context of imaging radars
(Source: NASA SAR Handbook [17, p. 29])

Band	Frequency [GHz]	Wavelength [cm]	Application
Ka	27 – 40	1.1 – 0.8	Airport surveillance
Ku	12 – 18	2.4 – 1.7	Satellite altimetry
X	8 – 12	3.8 – 2.4	Short range, urban mapping, marine radar
C	4 – 8	7.5 – 3.8	Long range, global mapping, airborne weather
S	2 – 4	15 – 7.5	Moderate range, terminal traffic control, monitor agriculture
L	1 -2	30 – 15	Long range surveil lance, high penetration.
P	0.3 - 1	100 - 30	Biomass, vegetation mapping and assessment

SAR imaging systems as illustrated in **Figure 5** are mounted on a platform which moves with the constant velocity v in the azimuth direction at a certain height H. In the meantime, the platform is having a linear displacement along track the antenna of SAR system transmits sequential microwave pulses at different locations, which afterwards are received as backscattered echoes. These raw backscattered echoes received from the ground are stored on-board form a two-dimensional matrix of complex samples where the signal is sorted by the arrival time in both range and azimuth direction. The generated two-dimensional matrix is compressed thus producing an image of the surface. [17]

In order to prevent possible distortions that might occur almost all optical instruments are usually built nadir looking, SAR system antennas are mounted side-looking [17], as shown in **Figure 5** thus, an incidence angle denoted with θ is introduced. Considering that radar measures the travel time of the pulses, a nadir looking sensor could not be able to distinguish between two equidistant targets from the sensor, because the backscattered signal of those targets will have a simultaneous arrival time. Hence, an ambiguity in range on equidistant targets will appear, in order to remove this ambiguity, the SAR sensor is restricted in a side looking view [17]. Therefore, the radiated beam reach reaches two targets at different time and as result have a different backscatter time for each of the echoes, thus, being able to distinguish between targets. From **Figure 5** can be seen an illustration of the acquisition geometry of SAR imaging systems.

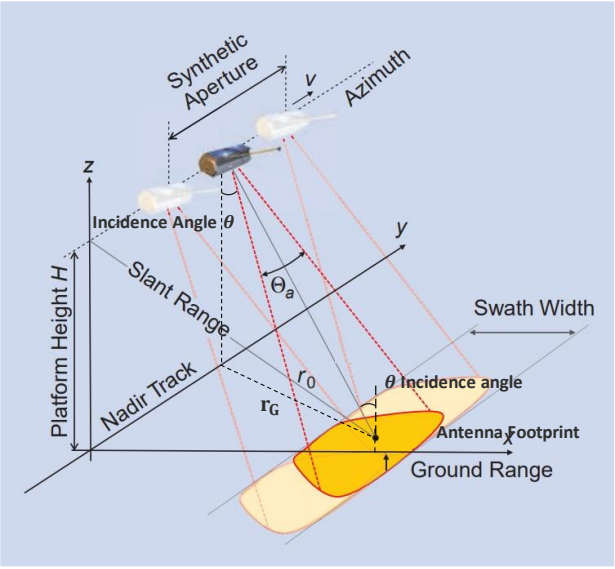


Figure 5. Illustration of the SAR imaging systems geometry
(Source: modified based on DLR [16, p. 10])

As outlined above the image in SAR systems is built up from the backscattered echoes in range and azimuth direction, hence the resolution of SAR images consists on two main components: range resolution and azimuth resolution.

2.1.1.1 SAR resolutions

The range resolution is defined by the ability of the sensor to distinguish two targets of different ranges but same bearing w.r.t reference point. On one hand, slant range is defined as the distance between the antenna and ground pixel (see **Figure 5**). Slant range r_0 is proportional to the pulse duration τ, the shorter the pulse duration the better the resolution. But in principle the slant range resolution is independent from the sensor to scene distance. In order to achieve high range resolution a large signal bandwidth is required. On the other hand, ground range r_G is defined as the distance between ground track and ground pixel. The projection of slant range into ground range is proportional to the incidence angle θ. The calculation of slant range and ground range are defined by equations (1) [16]:

$$r_0 = \frac{c \cdot \tau}{2} \qquad r_G = \frac{c \cdot \tau}{2 \cdot \sin(\theta)} \tag{1}$$

Azimuth resolution is defined by the ability of the sensor to distinguish two targets which consist of similar ranges but different bearings from the reference point. In theory this resolution is approximated as half of the antenna length [16]. The usual length of spaceborne antennas is ~ 15m, but SAR with its characteristics aims to create a simulation of a longer antenna which makes the aperture synthetic during the platform motion which yields a finer azimuth resolution. The width of angular aperture is defined by the equation (2) , where the wavelength λ of the frequency band, slant range r_0 and the antenna aperture A are considered [16].

$$\delta x = \frac{\lambda \cdot r_0}{A} \tag{2}$$

2.1.1.2 SAR Polarimetry

Polarimetry is a property with a very broad usage which refers on defining the travel direction of electric field vector. When the direction of radiation electric field is not defined then is called unpolarized. Polarization is very useful tool in remote sensing surface measurements for providing reliable polarization information of surface or volume scattering process and object geometry. In remote sensing usually only linear polarization is used, so the propagation direction of radiation electric field is either vertical or horizontal (see **Figure 6**) [19].

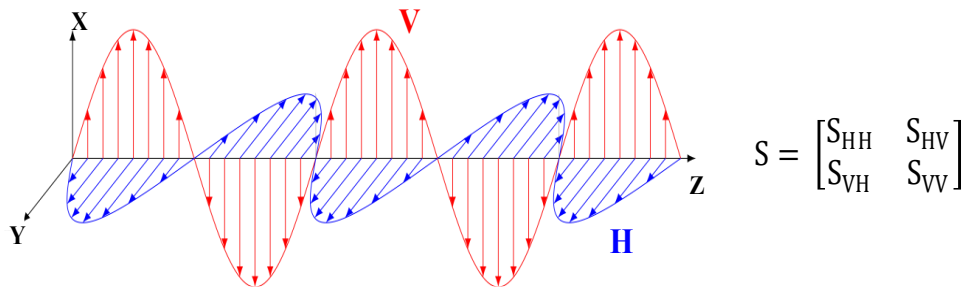


Figure 6. Illustration of Horizontal and Vertical polarization
(Source: modified based on Wikipedia [4])

Polarimetry in SAR systems consist on four different possible combinations:

- HH which consist on horizontal transmission and horizontal receive
- HV: has a horizontal transmission and vertical receive
- VV: which consist on vertical transmission and vertical receive
- VH: the radiation is vertical transmitted and horizontal received

Hence, four single image layers and fully polarized scattering matrix S (see **Figure 6**) are obtained [19]:

- Co-polarized a combination of HH, VV polarisations
- Cross-polarized is obtained out of VH, HV polarizations
- Dual-polarized systems provide two-layer combinations
- Quad-polarized systems provide four-layer combinations

2.1.1.3 SAR distortions and calibrations

Geocoding also known in SAR systems as geometric calibration is the process of the conversion of SAR images either ground range and slant range into map coordinate systems. This process is performed using two general procedures [17, pp. 37-38]:

- Ellipsoidal geocoding is defined as the method of the conversion into map coordinate system without using of DEM as reference data.
- Terrain Geocoding approach consist on the conversion into map coordinate system where reference DEM data are considered during the conversion process.

[4] Wikipedia https://en.wikipedia.org/wiki/Polarization_(waves)#/media/File:Electromagnetic_wave2.svg Accessed on 26 03 2022

In the data acquired by SAR systems it is very typical that due to the slanted geometry and topographic relief, terrain inclined areas appear in the SAR images as geometrically distorted targets. The reason behind these artefacts is that on SAR images the projection of a three-dimensional scene is measured in slant range and azimuth which causes an improper positioning w.r.t to their true position [17, p. 24]. The artefacts that can appear are foreshortening, shadow or layover (see **Figure 7**).

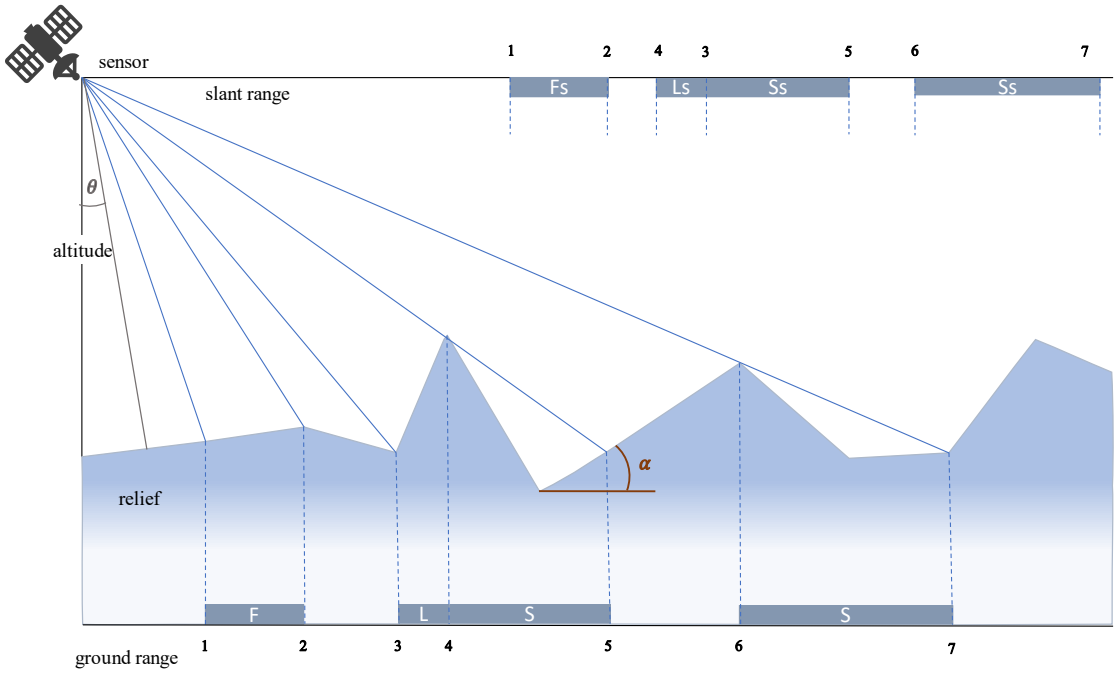


Figure 7. Most relevant geometric distortions on SAR systems

- Foreshortening $\theta > \alpha$ is an artefact which occurs on the slopes facing the sensor, those slopes appear as foreshortened and tilted towards the sensor.
- Shadow occurs in the areas with vertical structures, steeper than incidence angle θ, where there is no signal behind obstacles. These areas appear as black areas.
- Layover $\alpha \geq \theta$ is the opposite of foreshortening, in the areas where the incidence angle θ is smaller than the slope angle α. In layover the top of the mountain will appear ahead their base, thus causing a change in the target order and mixture in the signal

Beside geometric distortions, the utilization of side-looking geometry SAR systems on irregular ground surface causes sort another distortion in the output images, the so-called radiometric distortion. Therefore, in order to correct the radiometric distortions prior knowledge about the geometry of the observed scene is required. Some of the radiometric artefacts that disturb SAR images are speckle noise, brightness equalization, backscatter coefficient compensation, topography induced radiometric distortions [17, pp. 24-25].

SAR systems raw data acquired by the sensor are challenging to interpret, hence calibration of the existing distortions is very important in order to adjust the acquired data for the purpose of use. Radiometric calibration is important to successfully convert raw digital image data from satellite or aerial sensors to spectral radiance, reflectance, or brightness temperatures [17, pp. 45-48]. As one of processing steps on the conversion of raw SAR data explained in section 2 of Appendix A on SAR Handbook [17, pp. 45-48] this type of calibration is pivotal for reliable quantitative measurements of the SAR imagery.

On SAR calibration objective is to provide imagery in which the pixel values can be directly related to the radar backscatter of the scene. The SAR system sends a chirp signal towards the target and this signal is received back at the antenna of the sensor. The received signal also denoted as backscatter which defines the signal reflectance strength in a form of intensity value. [17, pp. 25-29] Depending on the application and selection case multiple type of calibration can be applied in order to transform radar reflectivity [17, pp. 25-29]. A more detailed explanation concerning the different type of radiometric calibration are shown in Appendix B, persisting with the mathematical background of those calibrations.

Table 3 shows an overview of the existing SAR sensors launched by different countries which operate on different bands and polarization.

Table 3. Overview on some of the actual spaceborne SAR sensors
(Source: derived from [16, p. 9])

Sensor	Operation	Frequency band (Polarization)	Country
Radarsat	2019-today	C (quad)	Canada
TerraSAR -X	2007-today	X (quad)	Germany
TanDEM -X	2010-today	X (quad)	Germany
RISAT -1	2012-today	C (quad)	India
ALOS -2	2014-today	L (quad)	Japan
SENTINEL -1a/1b	2014,2016-today	C (dual)	ESA
Kompsat-5	2013-today	X (dual)	South Korea
HJ-1C	2012-today	S (VV)	China

2.1.2 Sentinel – 1

Sentinel-1 is a radar satellite mission implemented under the so-called Copernicus program as a result of a collaborative work between European Commission (EC) and ESA. As ESA and EC states in the official website this mission is composed of a constellation of two satellites, Sentinel-1A and Sentinel-1B launched respectively on 3 April 2014 and 25 April 2016 [20]. Further, ESA and EC frames that the two satellite constellations share the same orbital plane phasing 180 degrees apart from each other.

On the objectives of Sentinel-1 ESA and EC states that SAR satellites are designed to ensure reliability on operational services and applications that require long time series; high and medium-resolution world coverage; short revisit time. Moreover, those advantages will provide users an enhanced advances on different EO tasks like: land monitoring, security, climate change or maritime [20].

As referred in [9] Sentinel-1 Technical guides Sentinel-1 has an altitude of about ~700 km and is launched in the near-polar sun-synchronous orbit with an inclination angle of 98.18 degrees. A single Sentinel satellite requires 12 days in order to complete a single cycle, which means that the satellite pass through the same target every 12 days. Hence, in order to shorter the revisit time to 6 days ESA has launched a composition of two Sentinel-1 satellites. Both satellites Sentinel-1A and Sentinel-1B have an orbital phase of 180 degrees apart from each other [9].

Figure 8 shows an illustration of Sentinel-1A and Sentinel-1B satellites and their positioning in the near-polar orbit.

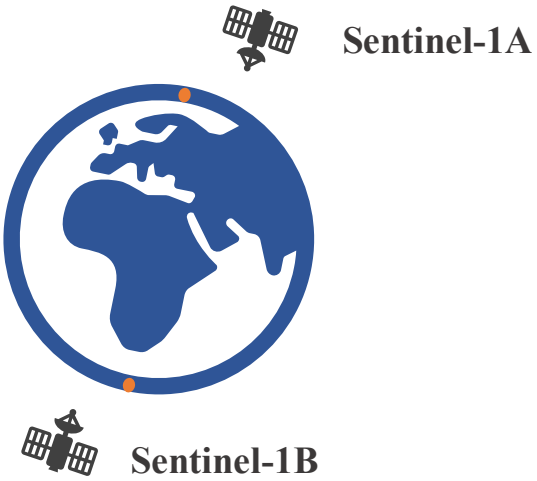


Figure 8. Sentinel-1 constellation

SAR Sentinel-1 operates in C band frequency (see **Table 2**) and supports a dual-polarization (HH+HV, VV+VH) (see **Figure 6**) which increases their efficiency on applications like land cover classification and sea-ice applications. The SAR instrument is composed of two subsystems [9]:

- Antenna subsystem is the sensor which controls the selection of polarization, beam shaping and steering over the transmitted as well as received beams.
- Electronic subsystem also considered as the core of the SAR instrument controls the architecture where transmitted chirps are generated, afterwards received as backscatters, generation of the frequency, modulation, filtering. In addition, the most important function is to store, compress and digitalize the data, and the provision of datalinks with the ground station.

Similar to all SAR systems, they are side looking systems with the antenna located on the right w.r.t to the flight direction, with the capability to operate in four modes for the acquisition of the images (see **Figure 9**) [21]:

- Interferometric Wide Swath (IW) mode provides the combination three sub-swaths of spatial resolutions 5m x 20 m which ensure the acquisition data of a large swath width of 250 km at single look. In order to create a product in IW single look mode one image per polarization channel is generated thus, a total of three images (single polarization) and six images (dual polarization).
- Wave (WV) mode consist on acquiring data in scenarios of size 20km x 20km at incidence angle 23 degrees (near range) and 36.5 degree (far range), every 100 km forming an undulated shape along the orbit. The acquisition of the images is done in single polarization (HH, VV) with spatial resolution 5m x 5m.
- Strip Map (SM) is a mode where the SAR instrument is capable to provide constant coverage with geometrical resolution of 5m x 5m with swath width of 80km. By configuring the antenna on fixed azimuth and elevation pointing, the incidence angle is changed to ensure the provision of six overlapping sub-swaths with total coverage range of 375 km.
- Extra Wide Swath (EW) mode operates very similar to IW for recording the data except the used technique which allows EW mode to acquiring a swath width of range 400km consisting of six sub-swaths of a lower resolution 20m x 40m.

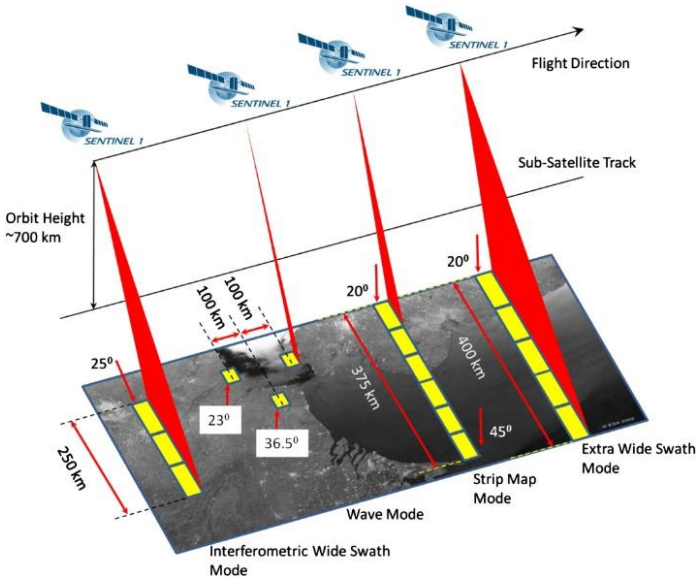


Figure 9. Sentinel -1 operational modes (Source: ESA [21])

In order to provide an accurate information concerning a process named Precise Orbit Determination (POD) [9] which is vital for Sentinel-1 quality control because it contains history records of the manoeuvre, the mass of satellite, the evolution of its gravitational center and the satellite's attitude, thus a precise orbit can be determined. Two RUAG [5] Global Navigation Satellite System (GNSS) spaceborne receivers of dual frequency are responsible for managing the provision of POD information which become available as soon as the images are generated [9].

According to [9] Sentinel-1 features provide three different levels for the products, which are generated and systematically processed as below:

- Level 0 products
- Level 1 products
- Level 2 products

A more comprehensive introduction concerning three levels for the products generated by Sentinel-1 is shown in Appendix C. What makes Sentinel-1 satellite special is not only the open access provided by ESA in different spatial and temporal resolution, that another important feature is the free software named SNAP [6] and Sentinel-1 toolboxes which allow users to efficiently filter or create images for their study purpose.

[5] RUAG - company which offers GNSS production

[6] Sentinel Application Platform (SNAP) - https://step.esa.int/main/toolboxes/snap/ Accessed 19 01 2022

2.1.3 Advantages of Sentinel – 1

Based on the latest developments on SAR systems and the capability to achieve outstanding performance, SAR systems are ranked among state-of-the-art technologies for the resolving EO tasks. Different SAR systems are proving to be able achieving an excellent performance for many applications compared to optical systems. Among them SAR systems Sentinel -1 has already raised a big interest on the usage because of the open access for users, high reliability and world-wide geographical coverage [20]. Based on ESA Sentinel-1 satellite description, the advantages the satellite offers can be derived [20]:

- Dual polarization capability
- Very short revisit time
- Rapid product delivery, open access to all users
- Operating at wavelengths not impeded by cloud cover
- Active system, not impended by lack of illumination
- Can acquire data over a site during day or night time
- Weather independent sensor
- Depending on acquisition mode high and medium spatial resolutions are available

Those great advantages provide a valuable resource by ensuring EO data at high spatial and temporal resolution which are generated and delivered continuously, hence, opening new perspectives on improving the working efficiency on different EO tasks.

2.2 Deep Learning

2.2.1 Overview on deep learning

Deep learning is a computational method inspired by the functionality of neural networks of the human brain, which performs processing of multiple layers and compass them to learn from a vast of data at different representation levels. [22] Nowadays, DL has been extremely important to data scientists because of the capability to efficiently achieve outstanding results for different complex tasks, thus outperforming other traditional methods. Although, DL for many might sound as a new trend technology, its beginnings dates back in 1940s and after many rebranding DL exponentially increased the popularity only in the last decade and researchers started revealing its fascinating computational power [22, pp. 1-26].

DL approach corresponds of artificial Neural Networks (NNs) [22, pp. 79-85] which might sound complex but in fact it applies a very simple mathematics in the background. NN consists of a

network constructed by nodes or units which take a value $\boldsymbol{x}$ as an input associated with a value $\boldsymbol{w}$ as the weight assigned based on the relative importance to other inputs and calculates the output $\boldsymbol{y}$ by applying an activation function $\boldsymbol{f}$. A comprehensive schematic illustration of NN is shown in Appendix A.

Depending on the usage purpose some NN techniques can be classified at different types such as:

- **CNNs** [22, pp. 326-336] consist on a NNs composed of several filter layers which apply convolutional operations. CNNs are very used for tasks like image processing and object detection.
- **Recurrent Neural Networks** have the form of a repeating modules of same structure which use a certain number of loops through the same network and the information persist from one step to the next. They have a wide usage for tasks like time-series analysis or natural-language processing [23].
- **Long Short-Term Memory Networks** are a type of recurrent neural networks with the only difference that on the repeated module the structure changes. They are used for time-series predictions, speech synthesis or music composition [23].
- **Generative Adversarial Networks** consist of DL technique which is composed of a model which generate fake data and map them beside real data as inputs on the discriminator model which then distinguish whether the data are true or false and afterwards perform an update of both models. This model is used for realistic images and cartoon characters, recreation of human faces and rendering 3D objects [24].
- **Radial Basis Function Networks** are type of networks with a very simple structure constructed by input layer, a hidden layer and the output layer. The hidden layer of those algorithms applies radial basis functions such as Gaussian function on the input variables. They have a wide usage for classification, regression tasks and time-series prediction [25].
- **Multilayer Perceptron** or feedforward neural network are DL models that consist of fully connected input and output layers with the multi hidden layers in between. The weights are calculated between input and hidden layers and each hidden layer contain its activation function. They are very used models on most speech recognition, image classification or object detection tasks [22, pp. 326-336].
- **Self Organizing Maps** are NN technique mainly used for unsupervised classification which are able to achieve very promising results in complex datasets. The name of the algorithm is somehow self explanatory of the network structure, where weights are assigned randomly to the input layer, thus the best match wins [26].

- **Autoencoders** are a type of DL models used for more efficient unsupervised classification. They apply an encoding of the input image afterwards compress it into a meaningful representation to end with the decoding step and producing the output. Autoencoders are very used methods for different image processing tasks [27].

2.2.2 State-of-the-art deep learning applications

Nowadays, DL methods with the capability to achieve excellent performance are considered state-of-the-art in many domains where can be mentioned [28]:

- Speech recognition
- Image Classification
- Image Localization
- Image Segmentation
- Target Detection
- Instance Segmentation, etc.

In this master thesis the main focus is on image interpretation tasks, where research is done to find the most suitable approach in order to achieve the objectives, thus object detection method and its frameworks are considered for further analysis. In **Figure 10** are illustrated some of the DL tasks mentioned above.

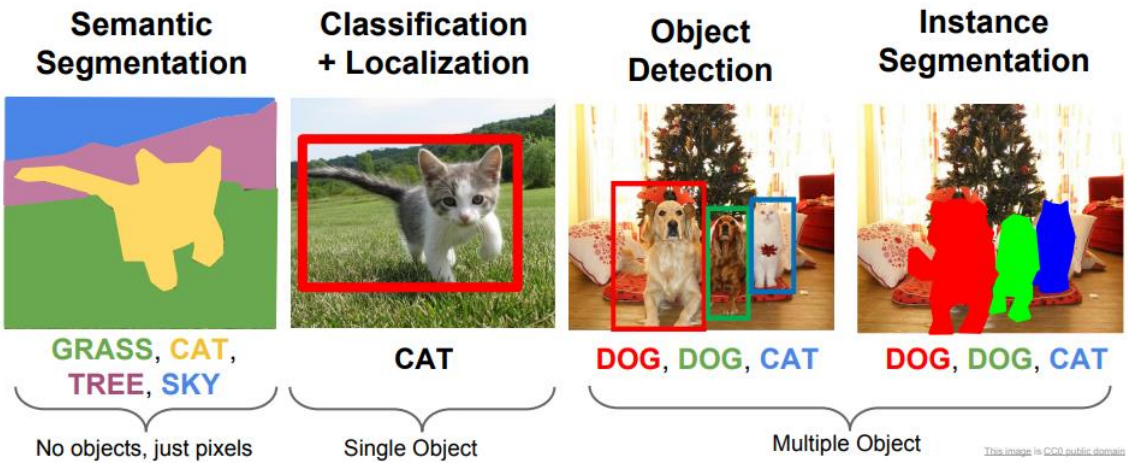


Figure 10. Comprehensive illustration of some DL tasks
(Source: Li, Johnson and Yeung, 2017 [29, p. 17])

Nowadays almost every research field or industry has been introduced to DL, while considering this approach to efficiently solve complex tasks which in the near past were considered as impossible. Some of the most popular application and industries which have a close relation with DL

technologies are: autonomous driving, virtual assistant, entertainment, security, visual recognition, healthcare, image restauration, demographic predictions, music industry, pharmaceutical, advertising, robotics, etc.

2.3 Deep Learning architectures

The exponential increase in usage and advances of DL methods is also reflected on the increase of the conducted research which utilize DL architectures, obviously also on the competition between for achieving the best performance in accuracy and speed. DL architectures are constructed based on the task they are implemented for (see Subsection 2.2.2), in addition employing the networks shown in Subsection 2.2.1. Considering the fact that in this master thesis is implemented object detection task, thus the focus is to concentrate more on giving an overview only for the architectures of object detection.

Mainly, the architectures utilized in object detectors incorporate CNNs. While, some of DL tasks such as i.e image classification, image segmentation aims to solve a single problem respectively *classification problem*, in comparison object detection which aims to solve a multi-task problem. As it can be seen in **Figure 10**, first the task of object detection is to predict the object class, this task is known as classification task. Second, for each object location to predict the bounding box coordinates, this task is denoted as *regression task*.

In the review research paper conducted by Hoeser et al. [30] the evolution of DL architectures and state-of-the-art object detectors are comprehensively reported. For instance, some of the detectors they compare are: *R-CNN, SPP Net, Fast R-CNN, Faster R-CNN, YOLO, SSD, Mask R-CNN, Retina Net, Cascade R-CNN, RefineDet512, EfficientDet-D7* [30]. According to their results, it can be concluded that Faster R-CNN shows a great advantage to be the best fit for the purpose of this master thesis. As a result, in this master thesis the focus is only Faster R-CNN detector, hence other detectors are not discussed in details.

2.3.1 Faster R-CNN

Faster R-CNN is a state-of-the-art object detection architecture used in DL techniques which incorporate deep CNNs, this approach is capable to outperform other competitors when it comes to localization and detection accuracy. In 2016 Ren et al. [31] proposed Faster R-CNN approach a new, unified network with re-modelled architecture in comparison to the successor of Fast R-CNN [32]. The architecture of the new approach consists of a two-stage detector by introducing in the model Region Proposal Network (RPN) (see Subsubsection 2.3.1.1). In **Figure 11** is shown an illustration of the components of Faster R-CNN technique.

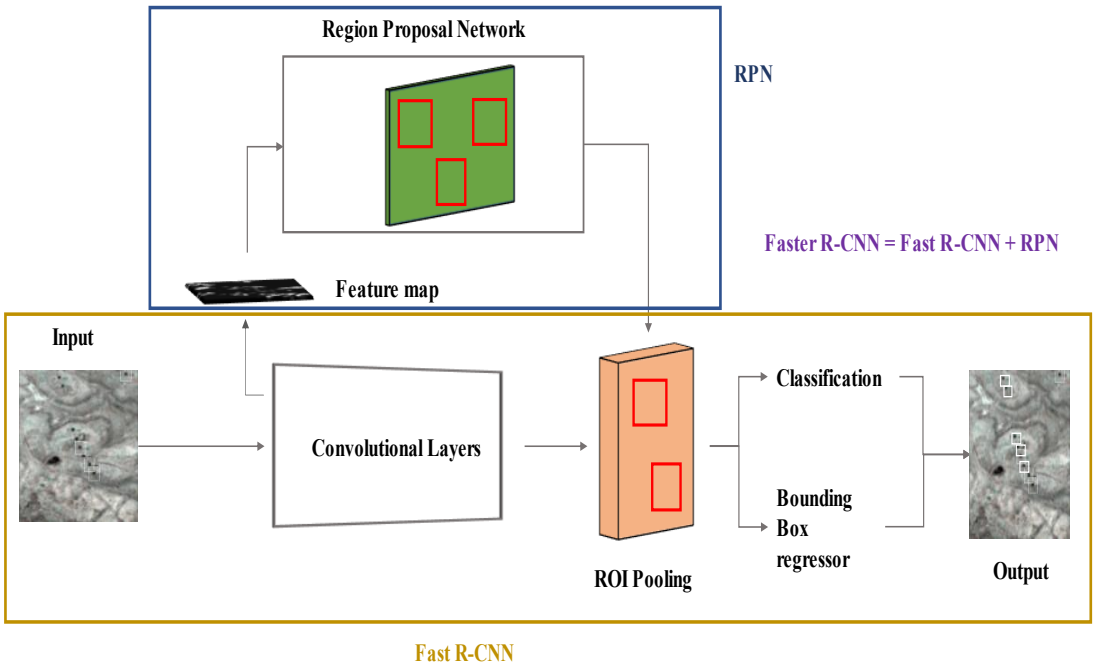


Figure 11. Faster R-CNN schematic illustration

2.3.1.1 Region Proposal Network

The introduction of the RPN module upgrades the Faster R-CNN approach by adding a mechanism in order to generate high-accuracy region proposals. This mechanism improves the classification efficiency; hence the model is trained to find the regions of interest. As stated by Ren et al. [31] RPN's are usually prevalent methods which serve as region proposal generators based on pyramid of images or filters at different representation levels.

These methods because of utilization of enumerating images are capable to ensure very good results but require more processing time. Therefore, in order to improve the processing speed in the architecture of Faster R-CNN, Ren et al. [31] proposed the generation of multi-anchors in the form of pyramid where the total number of anchors is denoted with k, over the convolutional feature map using a sliding window instead of processing the multiple features.

Anchor boxes in object detectors are a set of bounding boxes with predefined width and height. Usually, the size of an anchor box is defined as equal in height and width. In addition, an anchor box is associated with an aspect ratio. Aspect ratio refers to the proportion relationship between the anchor height and width. The left side of **Figure 12** illustrates anchor boxes and aspect rations generated on a feature map. Moreover, a feature map in which the anchors are constructed refers to the output obtained using the filtering layer on the input utilizing the convolutional layers mentioned in Subsubsection 2.3.1.2.

Hence, by using a k number of anchors located at the center of each sliding window and afterwards simultaneously multiple region proposals are predicted on the feature map. The anchors generated on the RPN module are associated with anchor size and aspect ratio, for instance if 3 different anchor sizes (e.g 8, 16, 32) and 3 different aspect ratios (e.g 0.5, 1, 2) are used, the total number of generated anchors is $k = anchor\ size \cdot aspect\ ratio = 9\ anchor\ boxes$ at each spatial location. The number of anchors used in the default configuration by Faster R-CNN is 9 but it can be customized.

Simultaneously two sibling convolution layers are applied on the boxes respectively a classification layer (*cls* layer) and a regression layer (*reg* layer). The classification layer for each anchor box proposal k produces 2 outputs, which is the estimated probability score for being an object or not. The regression layer for each anchor box proposal k generates 4 coordinates. **Figure 12** shows an illustration of the RPN mechanism and how anchors are generated on the sliding window.

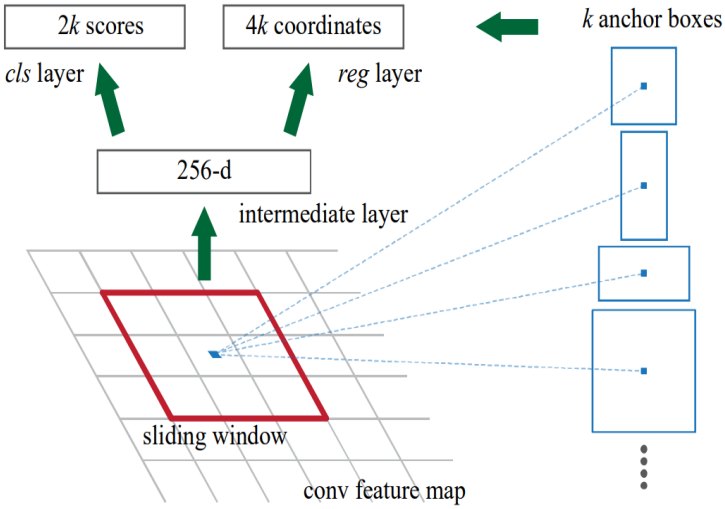


Figure 12. Illustration of RPN mechanism (Source: Ren et al. [31])

The capability of being translation-invariant approach allows Faster R-CNN to process more efficiently and increase the accuracy. Ren et al. [31] considers this property of Faster R-CNN as very important which allows the approach to reduce the model size. Further, Ren et al. states that this property guarantees on Faster R-CNN method that in the case an object is translated in an image, the same proposal ensures a translation of the object, as well the same function predict the proposal at location.

2.3.1.2 Convolutional Layers

Faster R-CNN [31] is known as an approach which employs mainly convolutional layers in order to achieve detection tasks, but at certain parts of the model Fully Convolutional, ReLu, Batch Normalization, Max Pooling, ROI Pooling layers are also utilized.

Convolutional layer consists of simple filtering by an activation function of an input and as result produce an output at different size [33]. Different from Convolutional layers, in **Fully Convolutional** layer in every node is applied an activation function and therefore the output size remains the same [33].

Rectified Linear Unit (ReLU) layer applies an activation function to the input where it will direct output if the input is bigger then 0, otherwise if smaller it will output 0 [33].

Batch Normalization layer calculates the mean and standard deviation w.r.t the mini-batch at the time normalization is applied. Input output size remains the same [33].

Max Pooling layer applies a kernel as a sliding window, which extracts the maximum value of the pool [33].

Region of interest (ROI) Pooling layer applies the same strategy as Max Pooling layer, but with the difference that is applied on ROI's in order to select the most appropriate ROI's [33].

2.3.1.3 Feature Pyramid Network

For the extraction of the feature maps optionally can be utilized Feature Pyramid Network (FPN), as Faster R-CNN approach do not use feature pyramid networks under the default configuration. FPN role on Faster R-CNN consists of generating the feature maps which are inserted into RPN module (see Subsubsection 2.3.1.1), those feature maps are build in the form of image pyramids and consist vast amount of data at different representation levels also denoted as multi-scale feature representation.

Image pyramids are considered as necessary for achieving high-accuracy especially on object detection tasks. According to the research conducted by Lin et al. [34] he states that by introducing FPN to Faster R-CNN ResNet backbone (more detailed information shown in Subsubsection 2.3.1.5), the model significantly improves the object detection either in precision and recall (see Subsection 2.3.3) by surpassing other state-of-the-art methods. FPN allows on generating scale-invariant pyramid of features (see **Figure 13**), with the ability to feed the model by semantic rich features, thus facilitate a higher recognition rate.

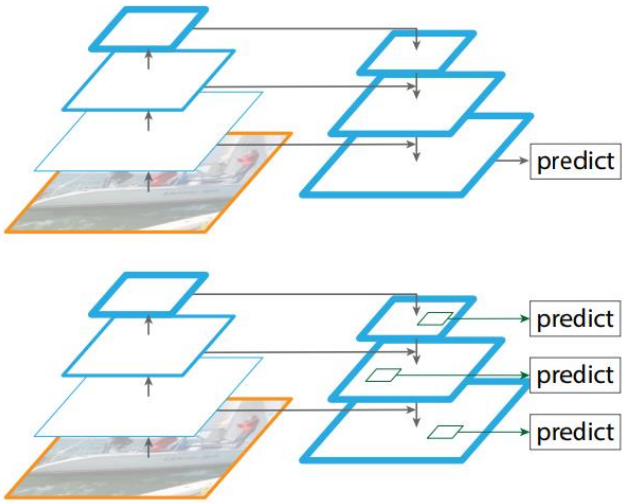


Figure 13. FPN architecture (Source: Lin et al. [34])

2.3.1.4 Loss functions

Faster R-CNN approach calculate two losses for the object detection, the classification loss and regression loss which are combined on a multi-task loss. They are calculated in the RPN module (see Subsubsection 2.3.1.1) and in the final classification. For this each anchor box generated in the RPN module is classified for being an object or not. This classification is performed based on the evaluation metric IoU (see Subsubsection 2.3.3), if the proposed anchor has an IoU overlap with the ground truth box bigger then 0.7 the anchor is classified as an object, if the proposed anchor has an IoU overlap with the ground truth smaller then 0.3 the anchor is classified as **not** an object. The anchor that are neither positive nor negative are not considered to the training.

Hence, the mathematical expression behind classification and regression loss is defined in equation (3) [31]:

$$\boldsymbol{L}(\{p_i\},\{t_i\}) = \frac{1}{N_{cls}}\sum_i \boldsymbol{L_{cls}}(\boldsymbol{p_i}, \boldsymbol{p_i^*}) + \boldsymbol{\lambda}\cdot\frac{1}{N_{reg}}\sum_i \boldsymbol{p_i^*}\cdot\boldsymbol{L_{reg}}(\boldsymbol{t_i}, \boldsymbol{t_i^*}) \qquad (3)$$

The multi-task loss is calculated for each anchor of the mini-batch which is denoted in the equation (3) with i. The classification and the regression terms are normalized using N_{cls} and N_{reg} which are determined by the mini-batch size for classification and by the number of anchor locations for regression. The λ parameter between two terms according to Faster R-CNN paper [31] is considered usually $\lambda = 10$ with the functionality of controlling the balance between two losses.

The classification loss L_{cls} equation (4) [31] applied in Faster R-CNN is known also as log loss, who make use of inputs p_i defined as the probability of an anchor i being an object and p_i^* which

is the ground truth label, this label is 1 in case the anchor is positive, otherwise 0 if the anchor is negative.

$$\boldsymbol{L_{cls}(p_i, p_i^*)} = -\log\left(p_{p_i^*}\right) \tag{4}$$

The regression loss L_{reg} equation (5) [31] is a robust L1 loss calculated over t_i as the vector representing the anchor coordinates and t_i^* which are the ground truth coordinates associated with a positive anchor.

$$\boldsymbol{L_{reg}(t_i, t_i^*)} = \sum_{i \in \{x,y.w,h\}} smooth_{L_1}(t_i - t_i^*) \tag{5}$$

in which:

$$smooth_{L_1}(t_i - t_i^*) = \begin{cases} 0.5 \cdot (t_i - t_i^*)^2 & if\ |t_i - t_i^*| < 1 \\ |t_i - t_i^*| - 0.5 & otherwise \end{cases}$$

Note that the regression function on the multi-task loss equation is only activated for positive anchors according to Faster R-CNN paper [31] and equation (3).

The bounding box regression Faster R-CNN approach according to Ren et al. [31] is following the typical bounding box regressor techniques, which are trained to regress the fixed anchor boxes to the nearby ground truth of a predefined object class.

2.3.1.5 Pre-trained backbones

Backbones in DL techniques refer to one of the most important processes performed during object detection. The backbone has an important functionality which consist on the architecture of the feature extractor network. Some of the backbones which can be used for object detection tasks can be mentioned: Visual Geometry Group (VGG) with models VGG-16 or VGG-19 developed on 2014 by Simonyan and Zisserman at University of Oxford [35] which consist of backbones with 16 and 19 layers deep respectively. Residual Network (Res Net) with models ResNet-50, ResNet-101 developed on 2015 by He et al. at Microsoft Research [36], with the architecture that combines a network of 50 and 101 layers deep.

These backbones are available as pre-trained networks which are already trained on large datasets like e.g ImageNet or Common Objects in Context (COCO) for different objects, hence obviously efficiently improves the model to detect object easier and learn faster. On the other hand, based on the research conducted there are no publicly available pretrained backbone for the SAR images for any size of input channels. As result the use of a pretrained backbone is not possible for helping the model during training.

An illustration of Res Net architecture with 50 layers deep is shown in **Figure 14**

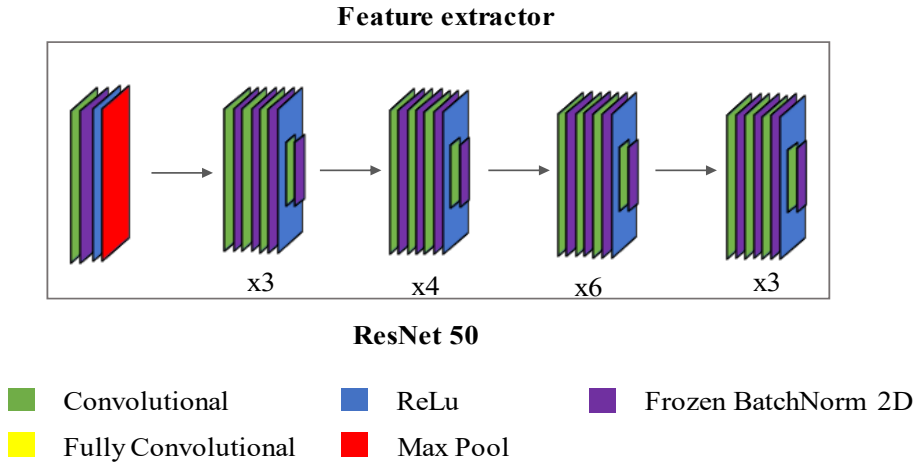


Figure 14. Illustration of ResNet-50 backbone architecture

Furthermore, in **Figure 15** is shown an illustration of VGG architecture of 16 layers deep.

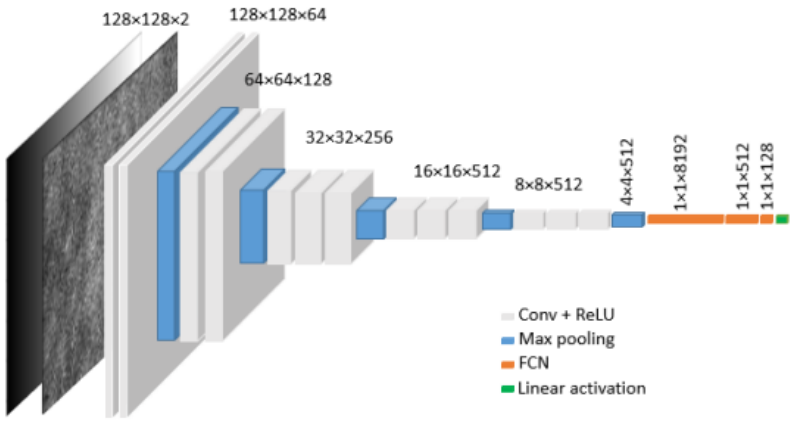


Figure 15. Illustration of VGG-16 backbone architecture (Source: Hashemi et al. [37])

2.3.2 Hyperparameter configuration

Moreover, after building the Faster R-CNN model, the hyperparameter tuning to neural networks is required in order to build the architecture before compiling into the model. As most of the time hyperparameter tuning sets depends on the datasets, which makes it a very challenging task to find the best possible sets for a specific dataset, so in a way to be capable prevent overfitting and underfitting issues.

When training a DL model is aimed that with many iterations the overall cost to become smaller. This is achieved by looping the model learning epochs for long time, hence decrease the overall cost and positively increase the reliability of the model. However, looping the model learning for long time might result on overfitting or underfitting of the model. Overfitting is considered the scenario when the model learns too much from the training dataset, which means that the overall cost of the model is relatively small but the generalization of the model is unreliable. Underfitting consist on the case when the model has a high bias in terms flexibility in finding the dominant trend due to too many patterns. Hence, this means that the model has not learned enough from the training data and as a result is not capable to even capture the well generalized trend. [38]

Some of the most important hyperparameters to be tuned before model training are:

Number of epochs [28, pp. 22-25] which defines the number of times the entire dataset will pass forward and backward through the neural network and as a result to update its weights. Number of epochs has an important role because the more optimization rounds are done on the training data, the less error will appear on the data, but on the other hand this parameter need to be monitored frequently and an early stopping is applied due to underfitting or overfitting scenarios on the model. **Figure 16** shows an example of how the early stopping is applied after a "patience" epoch as soon as the validation or training loss starts to increase.

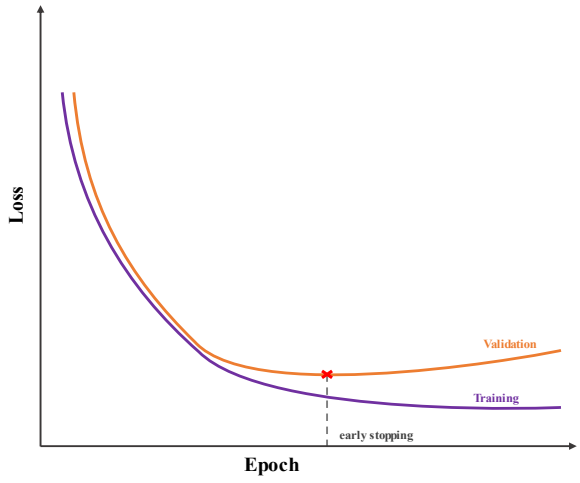


Figure 16. An example of early stopping applied during model training

Batch size [28, pp. 22-25] denotes the number of samples that will be propagated through the network. As the updating of the weights on the neural network can not be done on the whole training set due to computational complexity or noise issues, the update is done in batches of samples. Batch size importance is very high in the optimization process during training the data, as it controls the accuracy of the estimate of the gradient error.

This parameter needs to be chosen in the correct way as a too large batch size can have a negative effect on the stochasticity of the gradient descent, on the other hand a too small batch size can affect negatively by producing a noisy gradient descent. **Figure 17** shows an illustration of the above-mentioned effects of batch size in gradient descent.

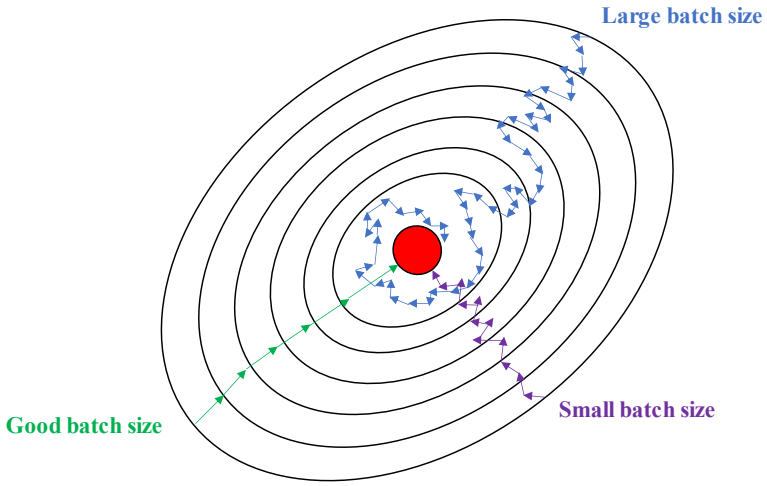


Figure 17. Illustration of batch size effects in gradient descent fluctuation

Optimizers [28, pp. 22-25] are algorithms in the neural networks designed with the functionality to minimize the cost of production or efficiently maximize the production. Optimization is an iterative procedure with the capability to find the parameters that minimizes the error by comparing various solutions. Optimization widely can affect the accuracy and the speed of the training model, therefore is executed until a finest solution is found. After each epoch during the training the weights of the model are updated and the loss function is calculated, in order to efficiently update those weights and obtain a small loss after each epoch optimizers are implemented with the functionality to modify the weights and learning rate attributes of the training model, which afterwards will help on reducing the loss and improving the accuracy.

Adam [39] is an algorithm first published in 2014 by authors Diederik P. Kingma and Jimmy Lei Ba which is used for training deep neural networks which aims to preform gradient-based optimization on loss functions. The algorithm is designed as an adaptive learning rate method, where individual learning rates of parameters are calculated and afterwards from the estimations adapt the learning rate for each weight w_t of the neural network using the first and second moment of gradient. A more extensive information about the mathematical background of the algorithm is shown in Appendix D.

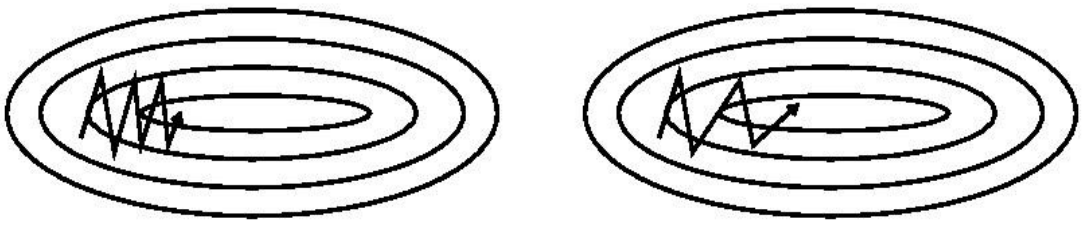
Stochastic Gradient Descent (SGD) [40] is a simple and very efficient approach used in Machine Learning (ML) algorithms and deep neural networks for the computation of the

gradient descent and afterwards perform an optimization of the objective parameters. In comparison to other gradient descent optimization algorithms which calculate and update the parameters over the whole batch, SGD algorithm tends to randomly pick a single or few examples from the training set at each iteration to reduce the gradient computations enormously. Hence, SGD was designed with the motivation behind to overcome the high cost of calculating the update parameters over the full training examples. This technique helps on the improvement of both training speed and accuracy.

Momentum [41] is popular approach with a very wide usage in deep neural networks that is known for the ability to accelerate the learning and as result improve the speed of convergence. In other words, momentum is a temporal element used to update parameters, is denoted with γ and ranges between [0 1] which is initialized with typical value about 0.9. The inclusion of the momentum hyperparameter thus transforms the formulation of the classic SGD equation (20) into equation (6):

$$w = \gamma \cdot w_{t-1} - \alpha \cdot g_t \tag{6}$$

Calculation of gradients and moving towards optimum process can be resembled with a ball pushed down a hilly terrain. The terrain slopes are called ravine, which are surface curves being more steeply in one direction then the other [42]. While searching for the optimum in the classic SGD algorithm as it can be seen in **Figure 18 (a)** it has big oscillation on the y-axis, so across the ravines which causes a slow progress towards convergence. Therefore, momentum is introduced to SGD algorithms which helps have less oscillations in the y-axis and thus accelerate the progress on x-axis towards the optimum as it can be seen in **Figure 18 (b)**.



(a) SGD without momentum **(b) SGD with momentum**

Figure 18. Illustration of momentum effect in SGD
(Source: Genevieve B. Orr [43])

Learning rate usually denoted as α can be a fixed or adaptively changed hyperparameter applied in the training of neural networks which indicates at which pace the weights should be updated. The importance of learning rate hyperparameter is described by Goodfellow et al. on his book

concerning Deep Learning as the most important where he states: "*The learning rate is perhaps the most important hyperparameter. If you have time to tune only one hyperparameter, tune the learning rate."* [22, p. 429].

The learning rate is in charge of controlling how much the model has to be changed as response to the estimated error each time a weight update is performed. Learning rate is strongly related to the loss function and it can be understood by having a look on the loss curve plot, an outline for some of the cases that occur during training are shown in the graphical way in **Figure 19**.

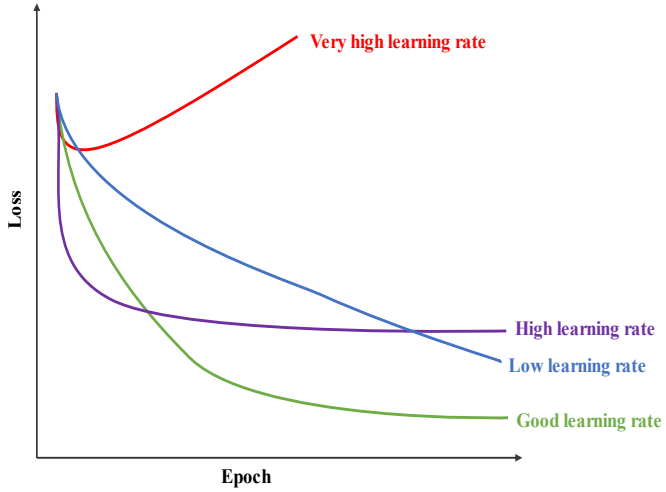


Figure 19. Example of learning rates effects that might occur during training

Learning rate can be also applied as a constant hyperparameter, but as well in order to prevent the errors during the training and validation most of the time a reduction of the learning rate, often after certain epoch is applied. The reduction of learning rate is applied when a metric has stopped improving using a patience number of epochs which can be manually configured. The learning rate reducer algorithms which are used on this master thesis are two:

- **Reducing Learning Rate on Plateau** [44] is an algorithm which reduces the learning rate based on the monitoring of evaluation metrics and if no improvement is done for a given patience epochs the default learning rate is dropped by an order of magnitude set on the factor argument (usually 0.1). Depending on the input evaluation metrics the learning rate reducer operates in two modes: minimum mode which the monitored quantity as stopped decreasing and in maximum mode when the monitored quantity as stopped increasing. Equation (7) shows the reduction of the learning rate based on Plateau algorithm.

$$lr_{epoch} = lr_{epoch_{-1}} * factor \tag{7}$$

- **Step Learning Rate** [44] is an algorithm which decays the learning rate parameter based on gamma at as step size. Gamma is a multiplicative factor which defines the decay of the learning rate. Equation (8) shows the reduction of the learning rate based on Step learning rate algorithm.

$$lr_{epoch} = Gamma * lr_{epoch_{-1}} \quad \%every\ (n)\ step\_size \tag{8}$$

In **Figure 20** it is shown graphically an example of the learning rate decay using the above-mentioned algorithms.

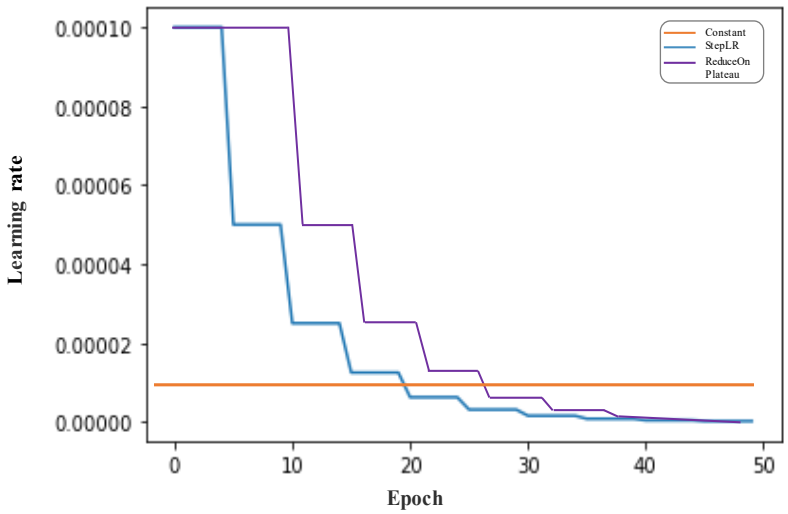


Figure 20. Example of learning rate decay with different algorithms

2.3.3 Accuracy assessment

Typically, in the context of object detection tasks in order to test the detection performance of the developed models, a set of evaluation metrics are considered. This accuracy evaluation metrics utilize the anchors predicted by the algorithm and the ground truth boxes obtained on the annotation process. Thus, for a more efficient accuracy assessment the input ground truth boxes are required and they have an important role, which otherwise with a non-accurate input ground truth box might cause inconsistencies during the automatic evaluation.

The comparison of the predicted anchor box with the relevant ground truth assist to correctly classify each of the predicted box and obtain a significant evaluation of the success rate of the network. Each of the prediction is inserted into the confusion matrix shown in **Table 4** which illustrate the relationship of four measures: True Positive, False Positive, False Negative, True Negative.

Table 4. Confusion Matrix illustration

Confusion Matrix		**True Class**	
		Positive	**Negative**
Predicted Class	**Positive**	True Positive	False Positive
	Negative	False Negative	True Negative

The determination of TP, FP, FN, TN samples in Faster R-CNN is done using the calculation of Intersection-over-Union method, which is the most popular method used in object detectors to quantify detection performance for a particular dataset. This simple calculation is expressed by the equation (9):

$$IoU = \frac{area\ of\ overlap}{area\ of\ union} \tag{9}$$

IoU is calculated utilizing the ground truth measures and the predicted anchors of the model, thus generate an output which ranges from 0-1, based on this calculation a class is assigned to the measure. On one hand the value 0 of IoU is considered as no overlap, on the other hand a value 1 of IoU is considered as the perfect match. **Figure 21** visually illustrates an example how ground truths and the predicted anchors based on their coordinates are utilized which rewards the IoU match value.

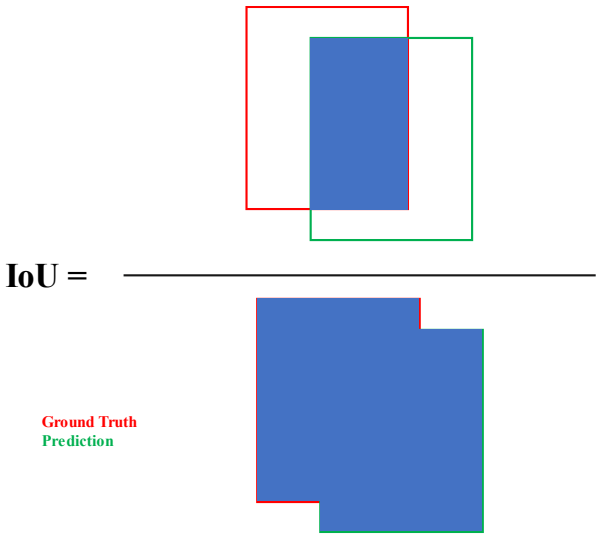


Figure 21. Intersection-over-Union calculation from ground truth boxes and predicted anchors

Furthermore, in order to categorize each of the sample from the obtained value of IoU thresholds are applied:

- **True Positive** samples are obtained when IoU between the ground truth box and the predicted anchor is above 0.5, this for the reason since the prediction anchors have different anchor size and aspect ratio might occur that the best prediction has the centroid of the box closer to the ground truth but the total box area is not sufficient to reach high IoU. Thus, when evaluating the model using COCO evaluator IoU 0.5 is also considered.
- **False Positive** samples are determined when the IoU between the ground truth and predicted anchor is below 0.5 and above 0. In addition, when applying automatic evaluation of the model FP samples are also considered new anchor predictions where ground truth is not available
- **True Negative** samples in most of the cases of object detection tasks are not considered due to lack of information for reasonable ground truth boxes.
- **False Negative** samples are considered measures where IoU is equal to 0, which means that might be either an instance where for a ground truth box a predicted anchor is available but they do not overlap each other, or an instance where for an input ground truth a prediction is not available.

As a result, the effectiveness of the object detector model can be evaluated using the number of categorized samples in order to calculate the accuracy defined as (10):

$$Acuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{10}$$

The Precision also known as Sensitivity is a quality measure of the model evaluation also defined as the proportion of which the actual positives are correct is calculated using equation (11):

$$Precision = \frac{TP}{TP + FP} \tag{11}$$

The Recall also known as Specificity is a quantity measure of model evaluation which determines the proportion the actual positives are identified correctly, by means of the following equation:

$$Recall = \frac{TP}{TP + FN} \tag{12}$$

From Precision and Recall measures the weighted mean of the two measures can be defined using the following equation (13), this metrics determine the weighted accuracy of the model.

$$F1\ score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \tag{13}$$

3 Related Research

In this chapter the intention is to review the different literatures which are linked with the influence of an up-to-date dataset of energy infrastructure in energy system analysis, different researches which utilized them, review target detection methods, DL frameworks in similar applications. Section 3.1 a brief explanation on the review for target detection tasks, the detectors which are used on those applications. Further, on Section 3.2 aim to review the energy system analysis previous researches and create an idea was already achieved similar applications. Furthermore, Section 3.3 shows conclusions from the conducted related research and how the novel methodology of this thesis purpose to overcome limitations that occurred in previous researches.

3.1 Review on target detection architectures

Earlier researches and works on SAR object detection has shown to be very successful in different applications. "*Even in small-sized networks are easily able to achieve more than 90% test accuracy on most tasks of military vehicle detection using Moving and Stationary Target Acquisition and Recognition (MSTAR) dataset or ship detection on custom datasets*" stated Xiang Zhu et al. [13] in the object detection section of the review paper written concerning deep learning and SAR.

On the review research papers concerning object detection and image segmentation commonly employed architectures Hoeser et al. [30], [45] investigated the capabilities, designs and detector models. "*In Earth observation applications on object detection, the two-stage approach dominates with 63%. Thereof, the well-established R-CNN models are the most used and of those, 73% belong to the Faster R-CNN design*" stated Hoerser et al. [45]. On one hand one-stage detectors, are designed to have only a single pass through the neural network and therefore predict all the bounding boxes in one go. On the other hand, two-stage detectors are divided by a RPN module (see Subsubsection 2.3.1.1), therefore, for each spatial location classifying and applying bounding box regression.

In the research regarding moving target detection respectively ship detection from SAR imagery Chang et al. [12] conclude from the experimental results that "*You Look Only Once v2 architecture (YOLOv2) test results showed an increase in accuracy of ship detection as well as a noticeable reduction in computational time compared to Faster R-CNN*", as well he proposed a reduced version of YOLOv2 architecture with a similar detection performance as YOLOv2, but with less computational time on a NVIDIA TITAN X GPU.

Based on the survey conducted by Jiao et al. [46] where he affirms that "*Two-stage detectors have high localization and object recognition accuracy, whereas the one-stage detectors achieve high*

inference speed" and the review research papers conducted by Hoerser et al. [30], [45], this master thesis considers two-stage detector architectures as the most suitable in order to achieve the objective, for the reason that the capability to ensure high-localization accuracy is primary considered in comparison with processing speed of the detectors. Hence, although Retina Net proposed by Alexander et al. [47] as the detector which is capable to achieve the best accuracy using high-resolution aerial images, as well Retina Net approach cited by Haopeng Zhang and Qin Deng [48] for the good balance between high accuracy and low running time. For performing object detection tasks, due to highest localization accuracy the combination of Faster R-CNN architecture associated with Res Net 50 + FPN are considered as the most suitable for modelling the methodology.

3.2 Energy Systems parameter extraction

Although the combination of DL techniques with remote sensing data nowadays are capable to provide astonishing results for segmentation, object detection or classification tasks used in different industries, still remains a huge potential to be discovered and improved. Therefore, this section is focused to review some of previous related research w.r.t detection and identification of energy infrastructure for mapping purposes using DL techniques and EO data.

In the research conducted by Manso-Callejo et al. [49] for recognition of individual wind turbines by means of DL techniques respectively hybrid architectures of semantic-segmentation and high-resolution aerial images in order to automatically generate and update a georeferenced wind turbine dataset. For the purpose of completing this task Manso-Callejo et al. utilized a dataset composed 5140 tiles of high-resolution aerial images of Spain with ~ 0.6 m pixel resolution. For each wind turbine Manso-Callejo et al. states that indenting to provide a very precise input semantic data, circles of radius 3-5 m were adopted at the bases for the generation of segmented masks.

Further, Manso-Callejo et al. concludes that the wind turbine identification of is accompanied with difficulties such as orientation of the generated shadow which depends on the image acquisition time, similarity in structure with transmission towers, as well orthorectification process errors. Last but not least, in the research [49] is proposed a strategy for automatic resolving inconsistencies of inaccurate extractions of wind turbines near the boundaries of the images. After filtering positional errors on predicted data, the obtained accuracy is 98.56 % for the tested region.

Guishen Tian et al. [50] in the study with regard to Electric Tower Target Identification proposed two-stage target detection algorithm of YOLO v2 and VGG cascade connection by virtue of transfer learning. In addition, frames that "*The proposed algorithm can be used to enhance the*

accuracy of only using YOLO v2 for target identification, thus reducing the false alarm rate of model effectively and the obtained result shows that the electric tower recall rate can reach 73.8%.".

In another research concerning grid infrastructure detection Alexander et al. [47] investigated three different object detection architectures and affirm that one-stage detector architecture of Retina Net was capable to achieve the best overall performance in both precision and recall in comparison with Faster R-CNN and YOLO v2. In the research conducted by Alexander et al. very high-resolution images were utilized (0.15m to 0.3m resolution) from four states in the US, in addition, according to Alexander et al. [47] important to notice for Retina Net model is: "*The obtained results indicate that the model detected the majority of transmission and distribution towers, while not producing an excessive number of false positives*.". Furthermore, he states that the model performance declined as image resolution decreased.

In the comparative research for different object detection architectures (one-stage and two-stage) conducted by Haopeng Zhang and Qin Deng [48] with regard to monitoring of fossil-fuel power plants by means of high-resolution satellite images and DL, they analyse the advantages and disadvantages on the experimented models. Further, they frame that is relatively challenging to suggest an architecture which can provide both the high-localization and speed at the same time.

However, Haopeng Zhang and Qin Deng [48] mention Faster R-CNN and FPN as the models capable to achieve the highest detection accuracy with Average Precision of 82%, but accompanied this with long running time. In addition, in the research paper is stated that Retina Net architecture performance was relatively balanced on both high accuracy and lowest running time. Important to notice in this research is that the object detection architectures were tested utilizing EO data which consist of 1-m spatial resolution acquired in China.

3.3 Conclusions

Based on the conducted previous research in related approaches this master thesis proposes a methodology which utilize open EO data and is capable to overcome most common issues, ensure high detection accuracy for small, medium, large-scale applications (different terrain conditions). Further, the proposed approach reveals the capability to be configured and thus, perform high accuracy either for objects small objects or large-scale objects.

In comparison to the conducted methodology by Manso-Callejo et al. [49] the proposed method in this master thesis instead of very high-resolution aerial images (which for a multi-country scale are not freely available) introduce SAR imagery. SAR images beside the advantage of offering

free data continuously updated at high-resolution, allows to overcome the shadow. In addition, also other inconsistencies of optical systems concluded in his research Manso-Callejo et al. [49]. This is illustrated in **Figure 22 (a)** where can be seen that shadow in wind turbine detection can easily decrease the accuracy, especially in areas where wind turbines are constructed very close to each other.

Furthermore, the inconsistency becomes relatively challenging when using images of lower pixel size in occasions like **Figure 22 (a)** which might result on bad precision. Hence, the proposed model in this master thesis overcome this issue by being shadow-free image as visualized in **Figure 22 (b)** where it is capable to achieve a sustainable precision on lower pixel size even for wind turbines constructed near each other. Finally, for overcoming the artefacts on detecting the object close to image borders, is proposed the methodology of utilizing overlapping images, therefore obtaining redundant predictions for a single wind turbine location which can be easily filtered afterwards in order to obtain a precise detection.

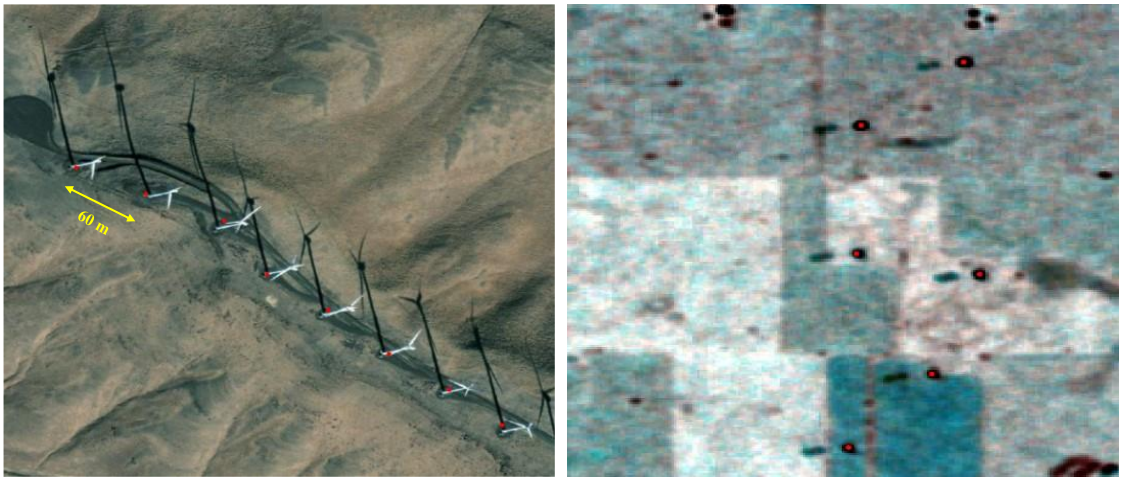


Figure 22. Wind Turbine instance artefacts inspection (*geo-location marked in red*) **(a)** Individual wind turbines visualized in high-resolution aerial images **(b)** Individual wind turbines visualized in SAR imagery utilized in this master thesis

4 Data and computational tools

The data and computational tools chapter aim to briefly introduce what data are used to accomplish the goal and by means of what hardware plus software the goal is achieved. In subsection 4.1.1 are presented the properties of the recorded images from Sentinel-1, pursuing in subsection 4.1.2 with the existing open global databases for the power plants. Furthermore, on Section 4.2 an introduction of the hardware utilized associated with the software used to support the process are presented.

4.1 Datasets

Details on the implementation of the main components on which the dataset is build construct the fundamentals and possess an essential contribution on the performance and accuracy of the result. These components which undergo through a chain of processes linked to each other build a dataset which can assure sufficient data for the model to be trained, validated and tested. Moreover, as a significant process the creation of the dataset can be complex as well as very challenging at the same time and on the other hand it requires good knowledge on understanding the structures, as it might end up of different issues during training or validation process. The dataset is automatically generated and annotated through a developed algorithm in order to compass the process faster.

4.1.1 Sentinel – 1

The all-weather radar satellite images from Sentinel-1 can be downloaded using either Copernicus Open Access Hub website *"https://scihub.copernicus.eu/"* where the data are made available from ESA as Level 1 data under two product types: Ground Range Detected (GRD) and Single Look Complex (SLC) (see Appendix C) and provided as raw data without any pre-processing applied, or can be downloaded through Google Earth Engine (GEE) *"https://developers.google.com/earth-engine/datasets/catalog/sentinel"* where the data are calibrated, orthorectified, pre-processed and can be optionally customized before download.

Data collected by GEE are of Level-1 GRD which comprises complex imagery data with amplitude and phase. The pre-processing steps of the data is implemented on each of the scene on GEE following same route as ESA provides on Sentinel-1 Toolbox, to derive the backscatter coefficient (see Subsection 2.1.1) in each pixel [51]:

- Apply orbit file
- GRD border noise removal

- Thermal noise removal
- Radiometric calibration (see Subsubsection 2.1.1.3)
- Terrain correction using Shuttle Radar Topography Mission (SRTM) 30 or ASTER Digital Elevation Model (DEM) for areas greater than 60 degrees latitude, where SRTM is not available.

Orbit metadata information contained on SAR products are generally not accurate, hence orbit metadata file is updated using the precise orbital products and auxiliary data generated by Copernicus Precise Orbit Determination (POD) Service (see Subsection 2.1.2), but those data are made available several days after the generation of the product. The data generation is associated with the availability of GPS file coverage of several orbits with significant overlaps, and result on distinct products of different timeliness and accuracy requirements. Sentinel-1 Toolbox of the Sentinel application platform allows an automatic download and update of the orbit metadata, except when the orbit metadata is not available, the update is done on a precise orbit file. [52]

GRD border noise removal is an algorithm designed to remove the radiometric artefacts that might appear on the scenes, such as low intensity noise and invalid data on scene edges. Radiometric artefacts at the image borders are caused by the azimuth and range compensation [52], these artefacts make the information on the edges not readable and therefore this information is not considered.

Intensity of Sentinel-1 images is disturbed by additive **thermal noise**, particularly this noise is shown on the cross-polarization channel [53]. Thermal noise removal is a processing step accomplished by subtracting the reconstructed noise field from the original image, in particular it normalizes the backscatter signal within the entire Sentinel-1 scene, which result on reduced discontinuities between sub-swaths for scenes in multi-swath acquisition modes [52].

The data obtained from ground range geometry are data which does not take the terrain into account, which result to some distortions on the side looking geometry (see Subsubsection 2.1.1.3). These geometric distortions such as foreshortening (F), layover (L) or shadow (S) appear in the SAR images due to terrain elevations. Terrain correction also known as orthorectification is a process which intend to compensate these distortions, hence the objects and matters that appear in SAR images will be represented in the images as close as possible to how they are in the real world. Using a reference digital elevation model to derive precise geo-location information a terrain correction is applied on SAR images to remove all the distortions caused by topography. Reference DEM used by Earth Engine to perform terrain correction are SRTM 30 and ASTER DEM when the latitude range is greater than 60° or less than -60° [51]. A more detailed explanation on the DEM used by GEE is shown in Appendix E.

GEE provides the capability to download for free Sentinel-1 images which are available from 03/10/2015 and continuously daily updated [51]. The data are available on the GRD mode (see Appendix C) and can be downloaded using Earth Engine Snippet as image collection. The collections scenes are processed using the Sentinel-1 Toolbox to generate a calibrated, ortho-corrected product, afterwards within two days they became available on the platform. Sentinel-1 GRD image collections of GEE offer a global coverage at: [54], [51]

- Three different resolutions which consist of 10, 25 or 40 meters.
- Three acquisition modes (see Subsection 2.1.2) which ranges at high and medium resolution of Level 1 products:
    - Interferometric Wide Swath,
    - Extra Wide Swath
    - Strip Map
- Polarizations (see Subsubsection 2.1.1.2) at single band VV or HH, and polarizations at dual band VV+VH and HH+HV

The data which are utilized for the master thesis are captured during the period November 2020 to October 2021, pre-processed and filtered in the temporal domain. For this period temporal statistics are available as product of intensity based multi-temporal filtering. The resolution of a pixel in the SAR images is 10 meters.

The images are cropped at the center location of 10000 individual locations for wind turbines and 2330 individual locations of coal power plants contain 12 different bands respectively 4 bands with statistics (mean, median, maximum, minimum intensity values) on VV polarization, 4 bands with statistics on (mean, median, maximum, minimum intensity values) for VH polarization and 4 bands on statistics (mean, median, maximum, minimum) out of combination VV and VH polarizations. This process of pre-processing, filtering in the temporal domain and calculating the statistics of SAR images was completed by DLR. Afterwards, at the geo-locations provided by the author the image chips were cropped and utilized in this master thesis.

Further, how polarizations and their combinations work and which polarizations are supported by Sentinel-1 SAR satellite, are shown in Subsection 2.1.2

4.1.2 Open datasets

Dataset creation and annotation of labels is a complex and the most time-consuming process because in almost every case it requires manual annotation and labelling of the classes that appear in the training, validation or testing images. Many open-source and commercial software are already developed in which the labelling process can be done, as an alternative solution many CNNs

are trained, validated or tested using some of the image databases containing thousand of class types which are available for free for researchers or non-commercial use, for example: ImageNet [55], MS COCO [56], Pascal VOC [57] (see more detailed information for those datasets on Appendix F)

Some existing datasets (e.g MSTAR [7]) are available for free also for SAR images, but these datasets are mainly used for military activities or surveillance applications such as ship detection or moving targets. Even though these datasets are able to provide a valuable resource for many general object detection tasks, when it comes to more specific tasks these datasets are not able to assist the process.

For the detection of mainly onshore energy systems such as wind turbine or coal power plants there is no existing open-database available where the images could be integrated on the process, hence the task was to establish a new database from scratch. And the establishment process of the database in which the model is trained, validated and tested involve external existing open datasets in which the geo-location of individual energy systems such as wind turbine or coal power plant is known. This information gathered from different open-sources databases is filtered, selected and distributed globally.

In order to reduce an existing drawback between the new created shapefile which locations will serve as known geo-locations for classes aimed to be detected in this master thesis. In addition, as well the issue if the targeted energy infrastructure such as wind turbine or coal power plant still exist and are not demolished or even not build yet. Such synchronization between the shapefile and SAR imagery is considered, due to the influence it may cause on mislocated energy systems. This synchronization process is performed in timeframe where the points extracted from open datasets and the recorded images from SAR consist during the period of November 2020 to October 2021.

Hence, the probability that the labelled images can contain information on the energy systems which are not on operational status is almost zero.

The use of existing open datasets with available information as geo-location of the classes which are aimed as target detection on this master thesis have an essential contribution. To automatically generate and label an own image dataset on this master thesis is a powerful resource which will open new perspective to have a up to date dataset with high accuracy and no mission information.

[7] Moving and stationary target acquisition and recognition (MSTAR) database, https://www.sdms.afrl.af.mil/index.php?collection=mstar, Accessed on 23 03 2022

Three different existing open-source power plant databases were chosen to filter power plant locations globally and crop the 10000 images, that are split in training, validation and testing as shown in **Table 5**, which consist on the individual wind turbine as the center point of the relevant image. Three of the existing open-source datasets integrated on this master thesis are:

- Global Power Plant Database
- U.S Wind Turbine Database
- ESRI Windkraftanlagen Brandenburg

Open GPPD [58] developed at WRI is a comprehensive, open-source database which has a global coverage of power plants at any size. Approximately 30,000 power plants including non-renewable (e.g. coal, gas, oil, nuclear, biomass, waste, geothermal) and renewables (e.g. hydro, wind, solar) from 164 different countries are covered on the database making it one of the best global databases for power plants with its last update on 2021. The available information out of the dataset is: geolocation and entries containing information on plant capacity, generation, ownership, and fuel type.

The advantage of this database different from other local databases is that has a global coverage and is up to date, on the other hand with its only limitation that for renewable energies such as wind turbine it includes wind farms instead of the geo-location of each individual wind turbine, thus a manual job on creation of 4000 different points inside of a shapefile distributed all over the world (see **Figure 23**) was considered as the most efficient solution to achieve a high accuracy result out of target detection model [8]. For Coal Power Plant case only GPPD database is used which contain the geo-location of 2330 plants distributed worldwide as shown in **Figure 23**.

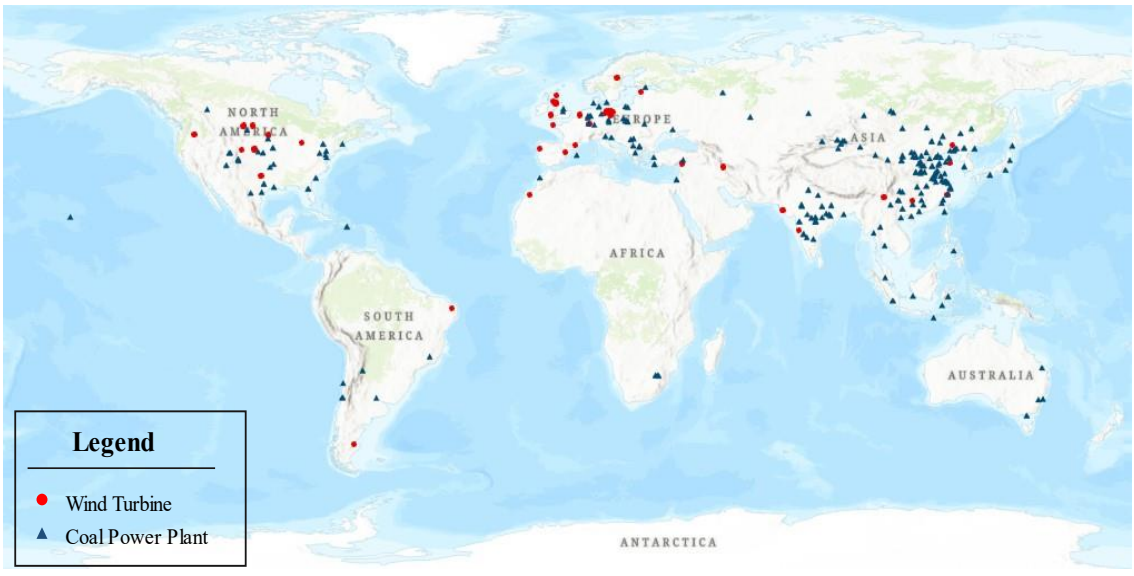


Figure 23. GPPD database Wind Turbines and Coal Power plants distribution

United States Wind Turbine Database USWTDB [59] a database developed in 2016 as a joint project named U.S Wind Turbine Database, provided by the U.S. Geological Survey, American Clean Power Association, and Lawrence Berkeley National Laboratory. With a goal to create a comprehensive and accurate database for individual wind turbines located in the U.S from different publicly available data sources and from data privately held by American Clean Power Association and Lawrence Berkeley National Laboratory the database was released to the public in April 2018. The latest release, version 4 published on November 2021 of USWT database consist of 70,142 wind turbines covering 44 states and the most recent update was done on the mostly on the second quarter of 2021 and with a few from the third quarter of 2021 [59]. From this dataset 4000 wind turbines distributed in the U.S as shown in **Figure 24** were selected to be integrated in this master thesis. The extracted wind turbines are split in training, validation and testing images (see **Table 5**).

Figure 24. USWTDB dataset wind turbine distribution

ESRI Windkraftanlagen Brandenburg [60] is an open dataset made publicly available from ESRI Germany. In 2019 ESRI published a dataset which contains information on wind turbines in the state of Brandenburg, this dataset is maintained quarterly with its last update on December 2021. A comprehensive visualization on the database is possible in operational wind turbines, approved and not yet commissioned wind turbines as well as planned wind turbines that are still in the approval process. The database contains 4725 individual wind turbine geo-locations with information on type of the plant, construction date, owner, capacity and the operational status [60]. Out of this database 2000 individual wind turbine are selected and randomly distributed

through the area (see **Figure 25**), afterwards split for training, validation and training as shown in **Table 5**.

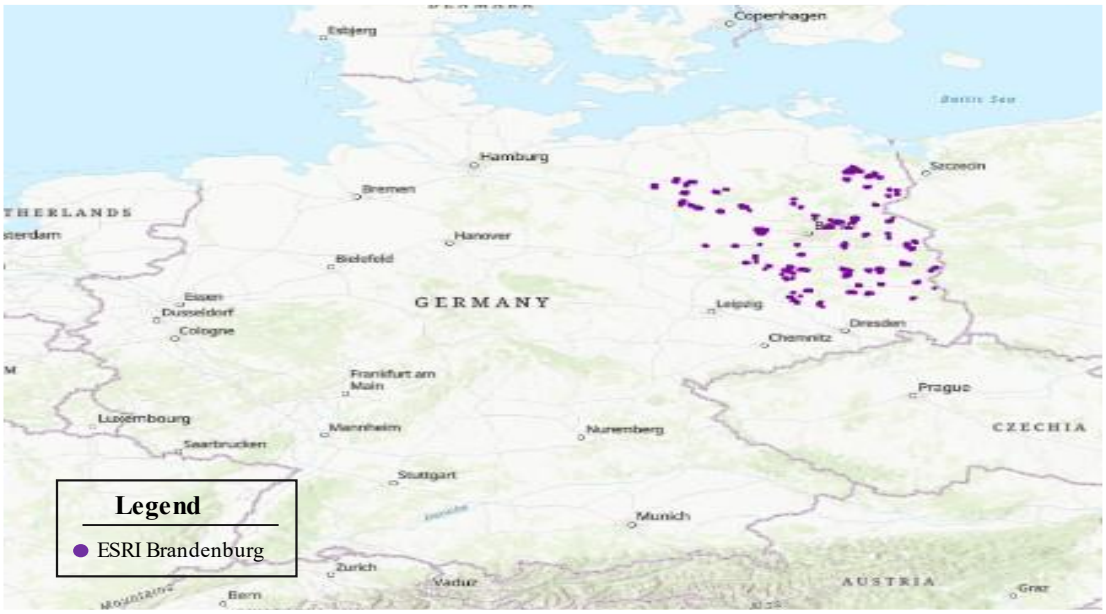


Figure 25. ESRI Windkraftanlagen Brandenburg dataset wind turbines distribution

After the download of the images at each geo-location center and completing the annotations with the help of the shapefiles, now the image coordinates of the wind turbines and coal power plants that appear on each of scenes are known. Before starting to build the model, a visualization of a batch of images in order to ensure that the conversion of coordinates is done properly. This visualization as shown in **Figure 26**, created using python programming language where only the annotated boxes of wind turbines and coal power plants extracted from relevant shapefiles are plotted.

Further, as the distance between individual wind turbines usually ranges to ~ 100 m, on the other hand images are cropped at each center geo-location, this results the same wind turbine to appear in multiple images. Basically, in order to have a more sustainable training of the model is recommended that the single wind turbine to be classified only in one of the processes i.e training, validation or testing and not in multiple processes. For a better split between processes the automatic way appeared to be quite complex, therefore a manual split using shapefiles in ArcGIS software thus classifying points was performed.

Important to note is that this operation construct the root of a very important process which has to be done in a very high accuracy in order to expect high performance from the model, otherwise inconsistencies will appear at the training or model evaluation.

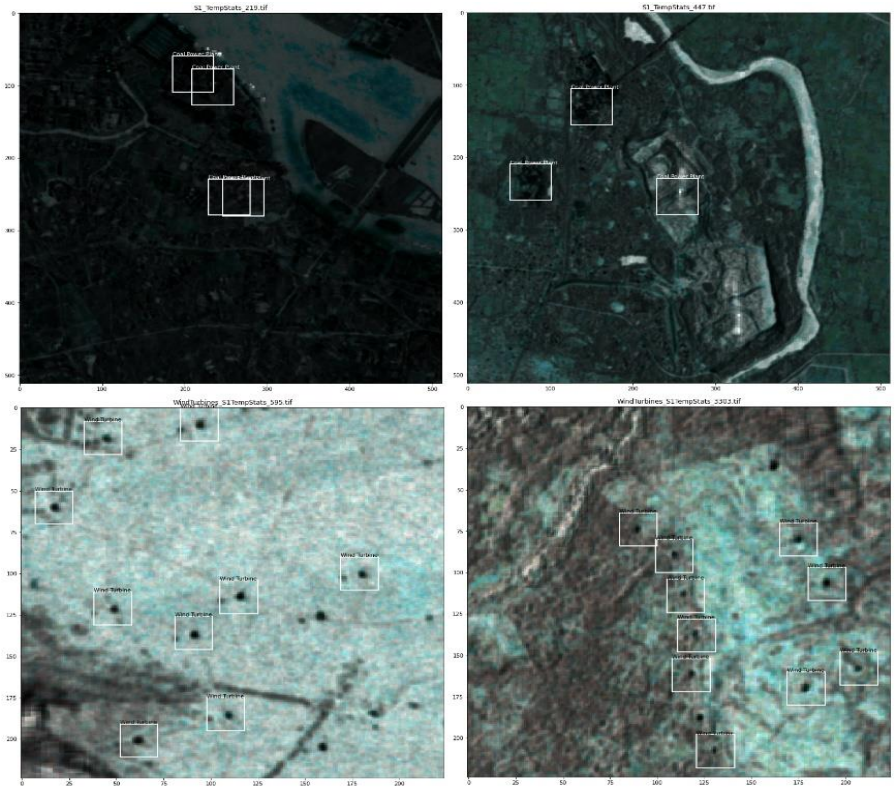


Figure 26. A batch of wind turbines (*lower*) and coal power plants (*upper*) images and their annotations

4.2 Hardware and Software

The model is built in such a way that in order to obtain a more comprehensive and user-friendly visualization of the final output is recommended the use of a Geographic Information System (GIS) software. Meanwhile during the process different parameters are plotted as graphs using programming language Python to track the process and make sure that the performance is as expected. Further, an overview of the hardware and software used for the development of the master thesis are briefly outlined.

ArcGIS Pro and QGIS is the software in which the important task of the creation of shapefiles with the wind turbine locations from all the sources and as well the separation of them into training, validation and testing. Those shapefiles are vital for the development as the information they contain compass the complex manual process and time consuming for creation of the dataset bounding boxes for object detection, to became an automatic process through a developed algorithm and extract the bounding boxes in the desired format. As well a comprehensive visualization of the final output with the support of the background maps and satellite images is possible to achieve using ArcGIS Pro.

ArcGIS Pro is a commercial professional geographical information system software developed by ESRI[8] which with the help and support of the toolboxes assist to explore, visualize, and analyse any amount of data in a professional way. In this master thesis ArcGIS Pro 2.8.3 was used with academic license provided by Hochschule Für Technik Stuttgart.

Python is the main programming language in which comprise the base for developing the algorithms. Python is a high-level programming language used widely for scientific and numeric computations. The support of many packages for mathematics, science, engineering, data analysis and modelling libraries etc. makes python one of the most efficient and used programming languages. Many different open-source libraries support the process with different computations, some of the libraries which are used are:

- Numpy
- Pandas
- GeoPandas,
- Rasterio
- Torch
- Torchvision
- Matplotlib
- CV2

Furthermore, many DL frameworks are developed in order to support the DL tasks (see **Figure 10**). Based on a conducted review research for the frameworks which have the widest usage based on their abilities, as result is finalized to chose PyTorch DL framework for constructing the model.

PyTorch [61] is a python based open-source DL framework that enables fast, flexible computations of machine learning models and allow developers to rapidly deploy models to production. PyTorch was primarily developed by Facebook's AI Research lab and initially released on September 2016. It provides rich libraries including python packages, ecosystem of tools and models, which makes the DL framework strong at GPU acceleration and capable to handle, incredibly efficient at computation of complex features like tensors. Some of the abilities that distinguish PyTorch to be mentioned are [61]:

- Ease-of-use and flexibility
- Distributed training
- Robust system
- Cloud support
- Rich ecosystem

Machine learning frameworks makes use of tensors (see **Figure 27**) which are data structure used by machine learning systems. ML frameworks tend to transform all the input features into tensors,

[8] ESRI https://pro.arcgis.com/en/pro-app/latest/get-started/get-started.htm Accessed on 02 01 2022

and eventually apply a vast of computations. Tensors are arrays of multi-dimensional distinguished by their uniform type with the ability transform inputs in order to make the numerical computation smooth, efficient on GPU and CPU. Everything transformed into tensors, divided into batches before trained and validated into the DL constructed network.

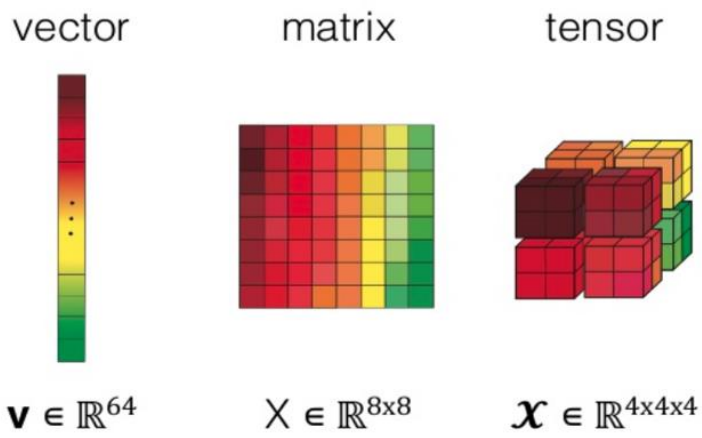


Figure 27. Comparison between vector, matrix, tensor
(Source: Berton Earnshaw [9])

The hardware which supports the build of the model and where the computations are done are the last generation hardware with a high-level performance. The process on the virtual servers is configured using Docker platform which run in a Linux operating system. The server uses a NVIDIA GeForce graphic card of type GTX TITAN X. GTX TITAN X is build in 2015 and is based on the graphics processor GM200. On GTX TITAN X the RAM memory of 12 GB GDDR5 memory has been paired which grant to graphic processing units a more efficient synchronous dynamic random-access memory. [62]

[9] A brief survey of tensors, Berton Earnshaw, 2017 https://www.slideshare.net/BertonEarnshaw/a-brief-survey-of-tensors Accessed on 25 03 2022

5 Methodology and results

In this chapter is presented in a brief way divided into sections the proposed methodology, workflow, experiments conducted in the master thesis and the results achieved. The first part of the chapter explains the methodology and workflow which is followed in order to achieve the objectives, further, the second part Section 5.2 shows the conducted experiments in which the methodology is tested by applying different hyperparameter configuration on network training. Further, Section 5.3 briefly explains the empirical results and the accuracy assessment obtained out of testing the Wind Turbine and Coal Power Plant models separately. Furthermore, Section 5.4 the author shows critical discussion and recommendation from personal experience concerning the applications outcome for the Wind Turbine and Coal Power Plant models.

5.1 Workflow

The methodology of this master thesis follows a 5-phase approach workflow (see **Figure 28**), the implementation aims to automate a complex process in which by means of state-of-the-art DL techniques and EO data respectively Sentinel-1, the algorithm is trained by a certain amount of input data and thus have the capability to perform target detection tasks. As an output the model generates a CSV file which contain geo-location and other properties of the detected features that can be easily externally loaded thus, create a comprehensive visualization using any GIS software.

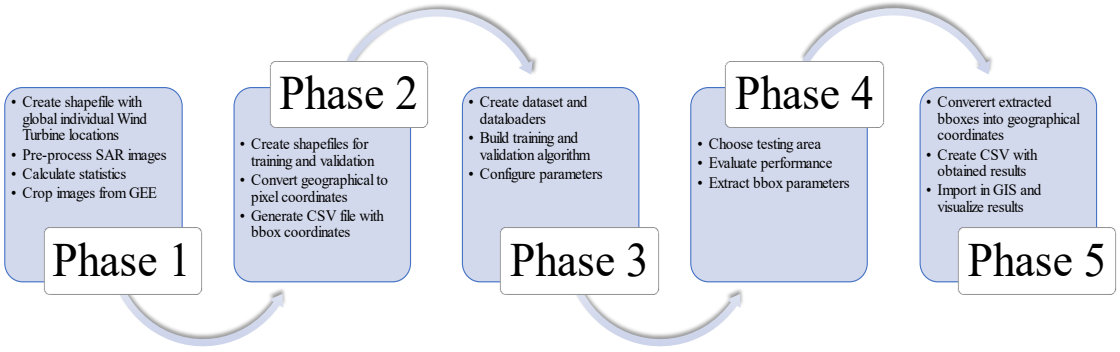


Figure 28. General workflow for energy systems parameter extraction

For each of the described phases in **Figure 28** an algorithm is developed. The starting point toward the solution consist with the creation of a shapefile which contains global individual geolocations of wind turbines, since the existing open databases are able to provide information only on the windfarms (with a certain number of wind turbines), thus no global database with single wind turbine location dataset can be found. The provision of this dataset required a manual work using ArcGIS Pro software and with the help of the open dataset of GPPD (see Subsection 4.1.2)

and orthoimages in background, the creation a new shapefile with single geo-locations of wind turbines distributed all over the world as shown in **Figure 29** (labelled in red color) is completed.

Additionally, in **Figure 29** with yellow color are labelled individual wind turbine geo-locations obtained from USWTDB (see Subsection 4.1.2), and last but not least with the purple color are labelled single wind turbines filtered from ESRI Windkraftanlagen Brandenburg database (see Subsection 4.1.2). Those inputs merged together in a single shapefile compose the individual geo-locations of 10 000 wind turbines distributed globally (see **Figure 29**). For the coal power plants model the input shapefile contains 2330 geo-locations of coal power plants distributed worldwide (see **Figure 23**), extracted based on GPPD open database. Those shapefiles together with EO data specifically Sentinel-1 data are essential components which constitute a key role for the process.

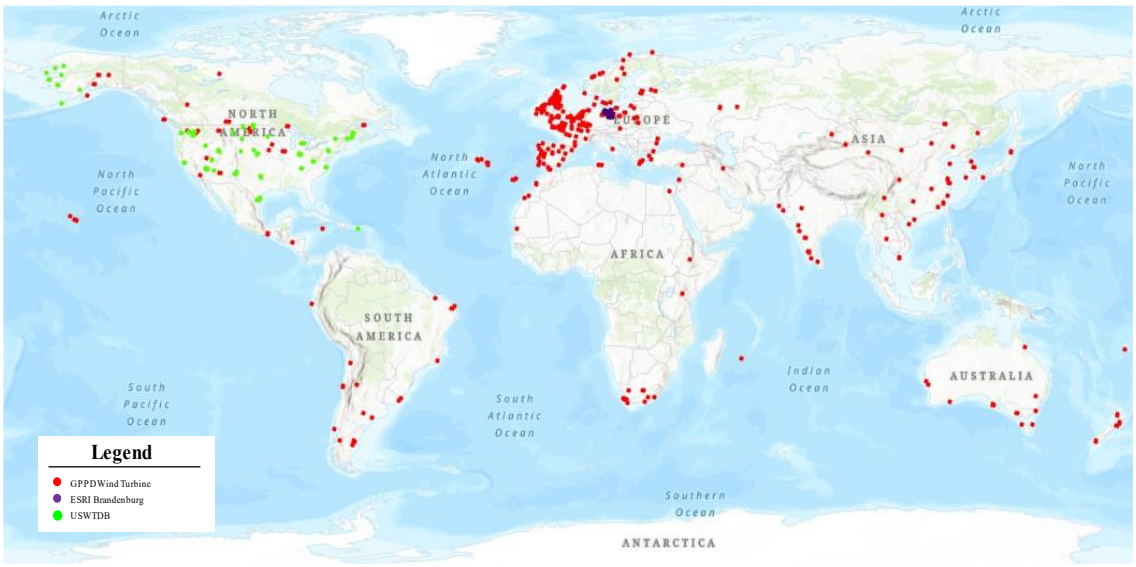


Figure 29. Distribution of manually collected wind turbines

5.1.1 Pre-processing

On **Phase 1** of the developed methodology pursues the goal that by means of GEE, the created shapefile with single locations of wind turbines and the created shapefile of coal power plants, to crop, download the pre-processed SAR images at a size which can be configured. The download of pre-processed images using GEE have an advantage in comparison to raw data from Copernicus Open Access Hub of Sentinel-1 for the reason that the above-mentioned processing (see Subsection 4.1.1) done by GEE is very complex and requires deep knowledge on SAR systems, but this processing is compulsory in order to correctly use the Sentinel-1 images. In addition, as GEE allows customization and filtering of SAR imageries on-site, the data are filtered in the temporal domain (pre-processing completed at DLR) making available statistics for each of the band during the period November 2020 to October 2021. The 12 different bands as explained in Subsection

4.1.1 are download at 10 meters pixel resolution and a single frame is cropped at size of 224px *x* 224px for the wind turbines model and 512px *x* 512px for the coal power plants model.

5.1.2 Dataset creation

Further, on **Phase 2** the developed python algorithm holds the functions to:

- Find how many points exist inside the frame of a single image when overlapping with the generated shapefile.
- Convert each existing point on the single image frame from geographical coordinates into pixel coordinates and produce bounding boxes at size 200m x 200m for wind turbines.
- For coal power plants since the size depends on the capacity of the plant, hence two different bounding boxes are created of size 500m x 500m and size 1000m x 1000m.
- Create a CSV file and store inside the information of image name, bounding boxes coordinates (stored in Pascal VOC format) and class name (wind turbine or coal power plant).

Pascal VOC [10] format (see **Figure 30**) is a bounding box annotation format which encode the bounding box coordinates with four values in pixels as:

$$[x_{min}, y_{min}, x_{max}, y_{max}]$$

where: x_{min} and y_{min} are coordinates of the top-left corner of the bounding box and x_m and y_{max} are coordinates of bottom-right corner of the bounding box.

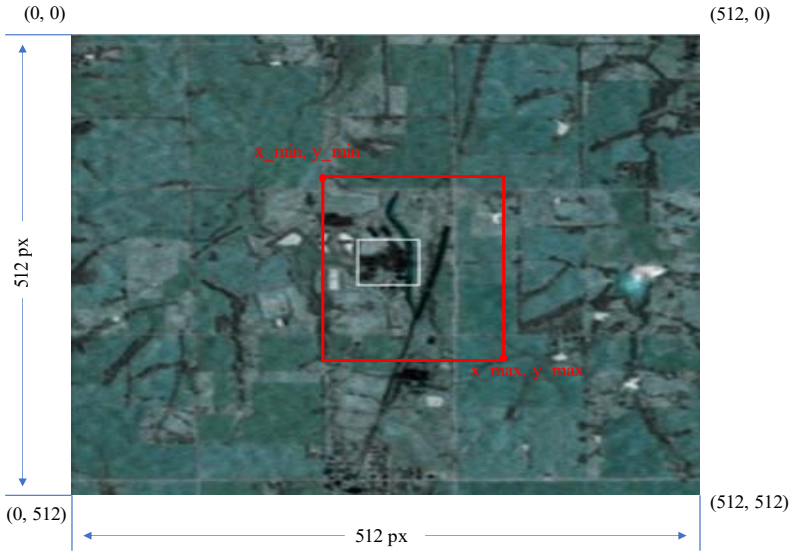


Figure 30. An example image with a bounding box annotation in Pascal VOC format

[10] Pascal VOC - http://host.robots.ox.ac.uk/pascal/VOC/ Accessed on 13 01 2022

In addition to **Phase 2**, mainly for wind turbines case, as a single wind turbine can appear in multiple images, the purpose is to include the same wind turbine and images that overlap each other only in one process training, validation or testing, so the same image will not appear in multiple processes. By preparing a random process to include scenes from different countries the distribution of training, validation and testing is at proportion as shown in **Table 5** with the number of images for each process for wind turbines. **Figure 31** shows a comprehensive visualization on the distribution of training, validation and testing wind turbines dataset supported by a background map generated using ArcGIS Pro.

Table 5. Training, validation and testing images distribution for Wind Turbines

Wind Turbine Dataset	Training images	Validation images	Testing images	Total images	Training %	Validation %	Testing %
GDDP	2767	822	411	4000	69,18	20,55	10,28
Windkraftanlagen Brandenburg	1285	495	220	2000	64,25	24,75	11
USWTDB	2753	803	444	4000	68,82	20,08	11,1
TOTAL	**6805**	**2120**	**1075**	**10000**	**67,42**	**21,79**	**10,79**

On **Table 6** the distribution of training, validation and testing images generated from GPPD geolocation database for the coal power plant model is shown.

Table 6. Training, validation and testing images distribution for Coal Power Plants

Coal Power Plant Dataset	Training images	Validation images	Testing images	Total images	Training %	Validation %	Testing %
GDDP	1667	420	233	2330	72	18	10
TOTAL	**1667**	**420**	**233**	**2330**	**72**	**18**	**10**

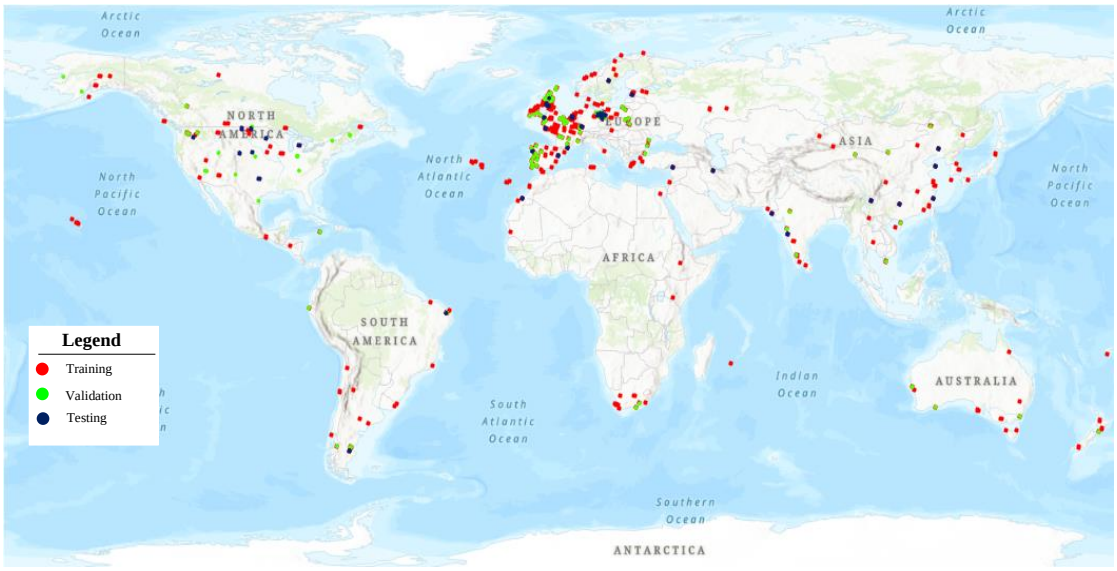


Figure 31. Training, validation and testing wind turbines images distribution

5.1.3 Model builder

Moreover, **Phase 3** also designated as the model builder aim to prepare the environment and build the model based on the proposed machine learning framework PyTorch (see Section 4.2). An algorithm is built which prepare the dataset, split the images in training, validation and testing as well perform data augmentation on images. As the DL architectures utilize tensors as inputs all the parameters are converted into tensors (see Section 4.2) before trained using Faster R-CNN object detector and simultaneously after each epoch the performance is validated using the set of validation images. In **Figure 32** a schematic workflow of model builder is shown, with all the tasks involved within the phase 3.

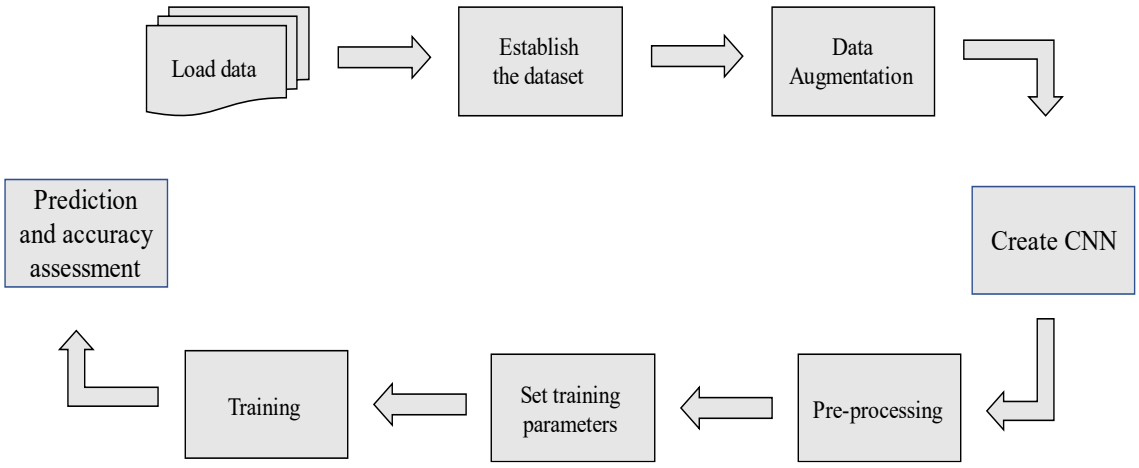


Figure 32. Phase 3 model builder workflow for training and validation

Load Data. As a starting point for building the model are the input data, those input data consist on the image dataset and the bounding box annotation file. Progressing to the next task which aims to take the input images with different range of intensity values for each image and each band, a normalization following expression (14) is performed, where Xn is the original image, X_{min} the minimum values, $X_{min}(i)$ maximum values, (i) denote the band. Normalization process target to change the range of pixel intensity values. As return normalized and standardized pixel intensity values at range [0 1] are obtained.

$$normalization(i) = \frac{Xn(i) - X_{min}(i)}{X_{max}(i) - X_{min}(i)} \tag{14}$$

Establish the dataset. In addition to this task after the normalization process, a transformation into numerical data containers such as tensors of the images, their classes and bounding box annotations is performed. Tensor attributes have the shape [*Channels, Width, Height*], moreover in **Figure 33** is shown a visual example about the transformation.

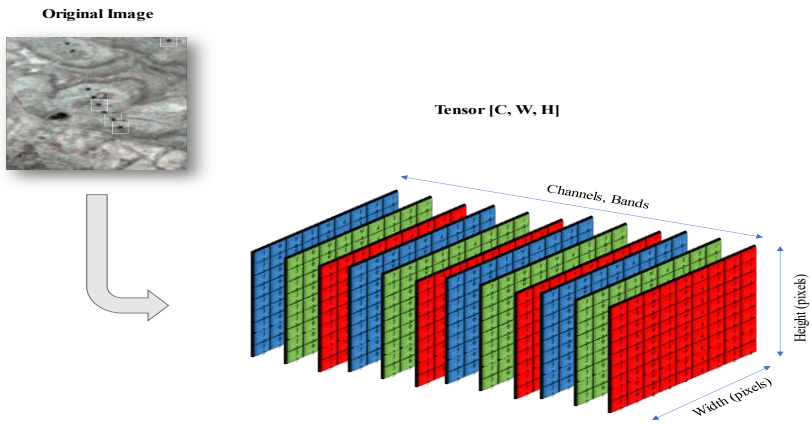


Figure 33. An example of image to tensor transformation

Data Augmentation. Further, after establishing the dataset a task named data augmentation, which is found in almost every DL model is performed. Since the DL models usually require a large amount of data to be trained in a proper way, sometimes this lack of data is compensated using the so-called data augmentation techniques, often found as a useful solution. In order to introduce more variability and increase the dataset size, without actually inserting new data for training and validation data augmentation techniques are implemented. Data augmentation help on increasing the dataset with slightly modified copies by performing augmentation of the original data. On the other hand, data augmentation techniques require a proper use, otherwise the augmented data might have a bad impact in training the DL models, which result on wrong training and low accuracy. Some of the data augmentation techniques that can be used are shown in **Figure 34**.

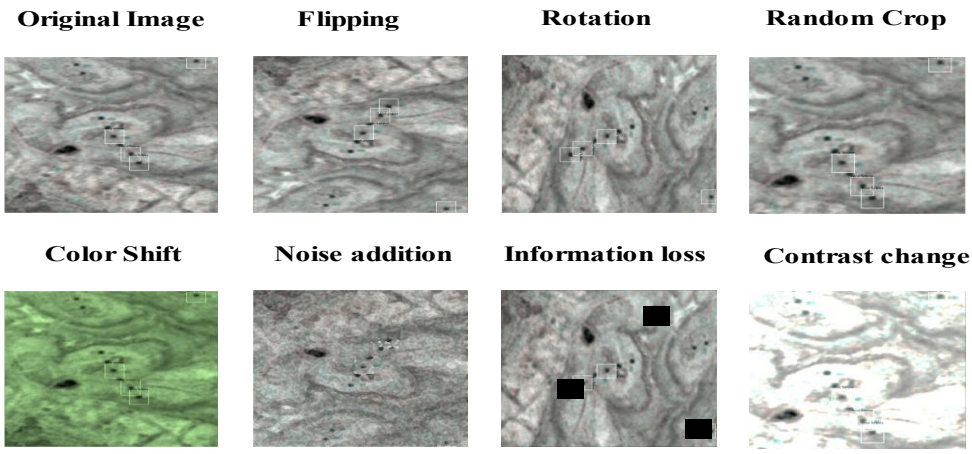


Figure 34. An example of data augmentation techniques

Create CNN. Next, the engine of the whole process is structured, the center point which is responsible for achieving the objective of the master thesis. This process has been entrusted to the state-of-the-art deep convolutional neural network Faster R-CNN (see Subsection 2.3.1). Faster R-CNN as a two-stage object detector model has been able to prove on solving complex computer vision problems. A customization on the layers of Faster R-CNN is performed starting with Generalized RCNN Transform layer which aim to implement an input and target transformation before feeding the data and adapting the model to be able working with 12 bands. In **Figure 35** a schematic overview of the design shows how the model is build, its components and different convolutional layers implemented into it.

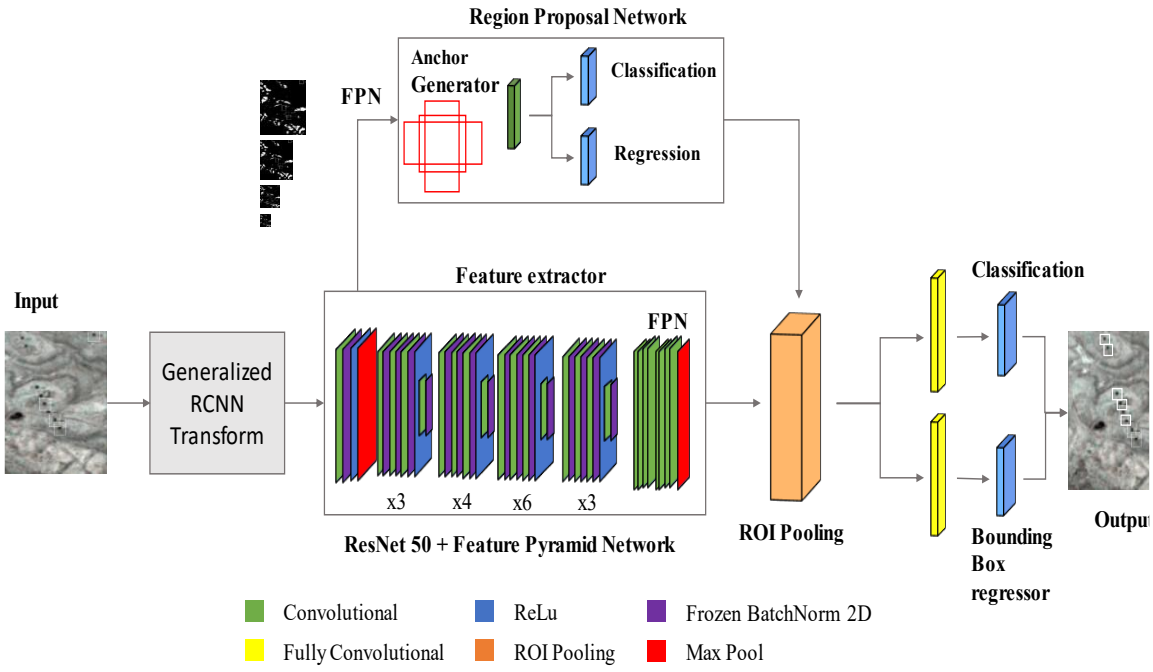


Figure 35. A design schematic overview of Faster RCNN model

As explained in Subsubsection 2.3.1.5 multiple backbone architectures from ResNet and VGG were considered for implementation in Faster R-CNN model. After conducting research, resulted that no pre-trained network exists for the purpose to improve the detection efficiency of the constructed model. Therefore, only the architectures were incorporated without being pretrained for the purpose of testing which one could have the finest performance using SAR images. When configuring the Faster R-CNN model only the ResNet backbones are considered, due to the challenging process for customizing the VGG backbones on PyTorch framework to work on 12 input channels of SAR images. The ResNet backbone which is 101 layers deep during the in-situ test is not classified as the appropriate backbone to achieve high-detection accuracy, due to the vast depth of layers. Hence, ResNet backbone of 50 layers deep was the most appropriate backbone considered for the models to extract the feature maps.

Pre-processing. The pre-processing task has an important role in the model builder, as the data from establishing the dataset are taken as a whole, thus containing a large amount of data, to ease the complex computational task those data are split into batches using Data Loaders. The function of Data Loaders is that through and iterative process to warp around the dataset and enable easy access to the sample. **Figure 36** shows in a schematic representation the mode of operation of Data Loaders.

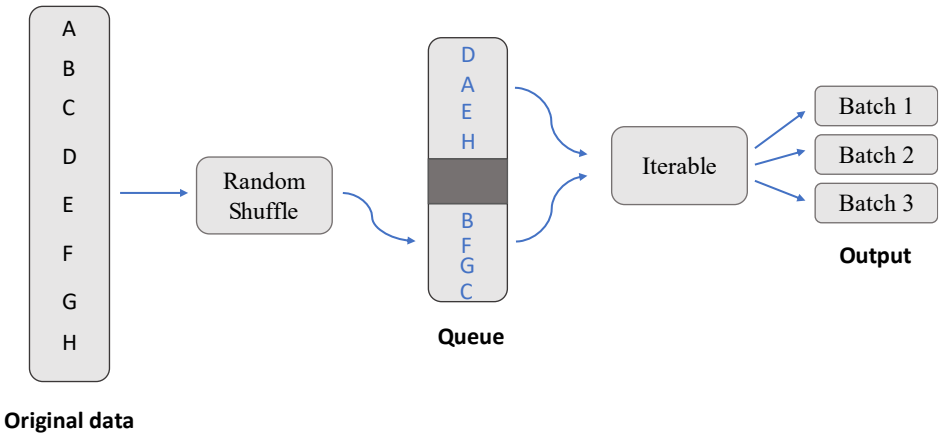


Figure 36. An example of Data Loaders

Set training parameters. Hyperparameters are settings of the constructed DL model which control the behaviour of the model. For constructing a reliable Faster R-CNN model tunning of hyperparameters is a must, since every custom input dataset to be trained comprise a unique case and thus the best fitting hyperparameters are to be found. The most important hyperparameters to be tuned are briefly introduced in Subsection 2.3.2, considering when selecting the hyperparameters the issues that might occur during the training process such as overfitting and underfitting (see Subsection 2.3.2).

Furthermore, after structuring the model for target detection application, the datasets are split into training, validation and testing samples and the hyperparameters are configured. The model is initialized and set into training mode where the initial weights are randomly chosen which consist on a bad performing neural network and wind up towards the best set of weights for maximizing the neural network accuracy. The model is set to loop through the training and validation samples for certain configured epochs, simultaneously to prevent errors during this process such as underfitting or overfitting (see Subsection 2.3.2) only the model with the best performance is being saved.

Training. On Appendix G is illustrated the algorithm build for training of the model, as well as the required parameters to start the training process.

5.1.4 Model evaluation

The last part of the **Phase 3** consists of the prediction and accuracy assessment which aim to compute the accuracy metrics such as confusion matrix, AR, AP, IoU and F1 score which are explained in details in Subsection 2.3.3. The accuracy assessment is a procedure used for evaluating the neural network model performance and computational efficiency. The model accuracy is calculated in-site during the training process on the validation step with its main components: the predicted bounding boxes and the ground truth data. The accuracy calculated for each epoch is afterwards used to update the learning rate (not used when default learning rate is defined as constant) and help the model to improve the efficiency in learning and increasing the accuracy.

In addition, **Phase 4** is strongly connected with **Phase 3** and performed as continuity of previous phase outputs. From the outputs of the trained and validated model, an algorithm is developed with the capability to perform target detection task on a set of test images with external known ground truth and finally produce again a CSV file with all the predicted bounding boxes coordinates, the confidence probability scores for the prediction and the image name where the prediction was extracted from.

5.1.5 Model testing

Last but not least, on **Phase 5** or model testing as the coordinates produced out of phase 4 are pixel coordinates the task now is through an algorithm to convert and reproduce those coordinates as geo-locations, hence to have the capability for external visualization. CSV output is extracted in the format as shown in **Table 7**, which contain information on image where the object was extracted, the class name or label of it, its geo-location and the confidence score of being correct or not.

The visual evaluation is done by means of orthoimages and ArcGIS Pro is chosen as a support software where the CSV file is added and visualized. In general, the workflow of methodology is structured in a way to make use of all available resources which lead on the automatization of the process on achieving the objectives of detection and ensure speed plus accuracy.

Table 7. Output CSV sample

Image id	labels	longitude	latitude	score
tile_001	Wind Turbine	-95.030573	16.541623	0.825686
tile_002	Wind Turbine	-94.828138	16.423719	0.904848
tile_051	Coal Power Plant	-94.828138	16.423719	0.904848
tile_052	Coal Power Plant	-94.807117	16.486197	0.938280

5.2 Experiments

To evaluate the performance on accuracy and speed of the algorithm 4 different models are presented as the case study, which are trained out of 4 different hyperparameter configurations. On one hand, SAR images are complex images to be seen with naked eye and require more professional knowledge to understand the output, on the other hand the intensity of the backscatter might change in the surrounding areas such as wetland, dry area, low vegetation area or forest. Those artefacts are considered therefore a global geographical scope is chosen during testing in order to improve and have a stable trained model.

5.2.1 Wind Turbine Model

5.2.1.1 Training and validation

During the model building phase many in-situ experiments were conducted in order to obtain the best hyperparameter configuration which are afterwards is applied for training and validation of a single model. Note that the many experiments were conducted for training and validation only for qualitative internal use, thus no evaluation of those obtained models is performed and discussed in the master thesis. Basically, extensive experiments were conducted on hyperparameters explained Subsection 2.3.2 and shown **Table 8** in order to reveal their influence of the model and making sure that the models are trained as expected. From this the final model evaluated is considered for the best achievement.

Further, the datasets utilized as input in the Wind Turbine model is comprised of GPPD, USWTDB, ESRI Brandenburg databases (see Subsection 4.1.2), which are independent open databases. The database split for each processes the model compiles is visualized in the pie chart in the upper part of **Figure 37**, in addition the lower part of **Figure 37** shows a bar chart with the number of images contributing from each particular dataset in every process.

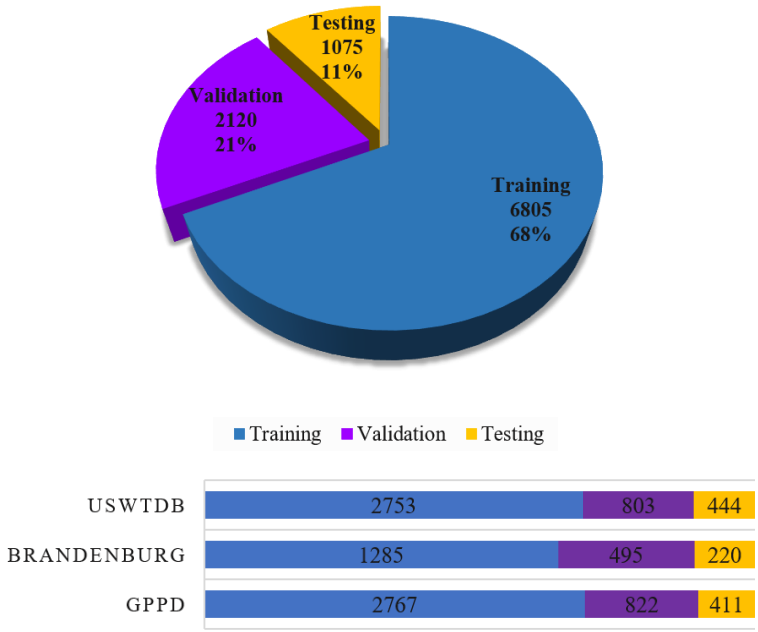


Figure 37. Dataset split for Wind Turbine Model

In order for the model to train in the most proper way, due to commonly existing high overlapping which appears in many regions wind turbines required a categorization in a single process. As it results for wind turbines to have a short distance from the neighbouring wind turbines this process requires special attention. Hence, a single wind turbine is categorized to participate only in one of the processes. This comprises of a relatively challenging task, considering that the lower the categorization scale the better accuracy can be obtained and model is trained or evaluated without inconsistencies.

5.2.1.1.1 Hyperparameter configuration

The hyperparameters to be tuned on a DL model for the purpose of object detection tasks are shown in Subsection 2.3.2. Typically, when a DL model is constructed attached are also considered a set of hyperparameters for which the model is capable to achieve a high-quality performance. **Table 8** shows the configuration of the hyperparameters for the wind turbine model based on image dataset split shown in **Figure 37**.

Table 8. Hyperparameter configuration for the Wind Turbine Model

Hyperparameter configuration	
Epochs	228
Batch size	8
Image width/height [px]	224 x 224
Starting Learning rate	0.0001
Number of workers	4
Learning rate patience	5
Anchor size	32, 68, 128, 256
Aspect ratio	0.5, 1, 2
Optimizer	ADAM
Learning rate scheduler	ReduceOnPlateau
Data Augmentation	Flipping, Rotation

As result of hyperparameter tunning, the best fitting model was sufficiently trained, thus stopped after 228 epochs due to no convergence, with the most appropriate size of the batch selected as 8. The learning rate is investigated in lower scale (0.00001) which results to no improvement after certain epochs, and in larger scale (0.1 and 0.01) which results to wrong training. As the most appropriate was considered the learning rate 0.0001, this learning rate shows some small oscillation of the accuracy curves during first epochs of training of the model, but in comparison with the other learning rates obtains the best fit for the model. This learning rate is being tracked using the evaluation metric curves and reduced using the algorithm *ReuceOnPlateau* (see Subsection 2.3.2). This algorithm tracks the learning process of the model and if no improvement is shown after 5 epochs of "patience" the learning rate is reduced with the factor 0.1.

Further, the number of workers parameter is configured as 4 simultaneous processes which means to speed the process and move the heavy calculations into GPU, thus CPU have less weight and this will result on a faster process. Moreover, the anchor sizes and their aspect ratios are configured in a way that for each pixel center location on the feature map 12 anchors of different size are generated.

As the utilization of a pre-trained backbone is not configured means that the model starts learning with zero weights, which weights afterwards at the end of an epoch are updated using optimization algorithms also shown in Subsection 2.3.2. For the Wind Turbine Model, the optimization algorithm in control of updating the weights is chosen *ADAM*, based on the capability to provide a smoother weights update compared to other algorithms.

In addition, Data Augmentation (see Subsection 5.1.3) direct an important and at the same time very sensitive process which exert influence for increasing as well as decreasing the accuracy. Hence, attempting to include several of data augmentation techniques such as Blurring, Shift Scale Rotate, Gaussian Noise, Random Brightness, Flipping, Rotation in order to artificially augment more images and increase the dataset quantity.

However, many of those data augmentation techniques did not result as sufficient for improving the accuracy of the model, thus only Flipping and Rotation of the original images are utilized.

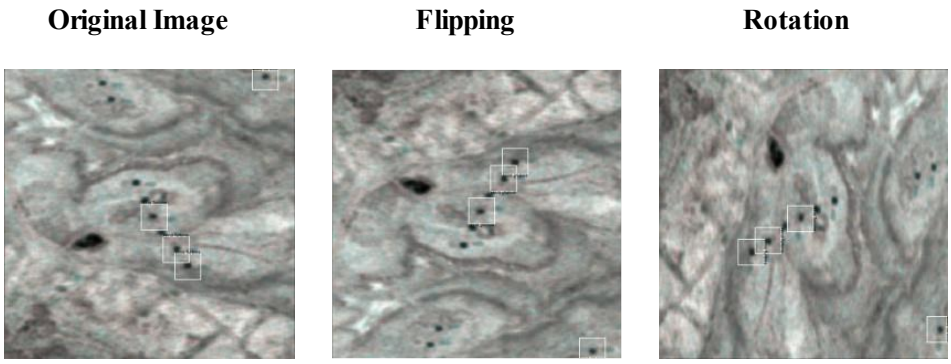


Figure 38. Data Augmentation techniques utilized in the Wind Turbine Model

Even though, the processing speed of the model is not considered primarily on this research, as the main focus is to be capable achieving high-detection and localization accuracy, however, the model appears to be very fast when compiled in the GPU, but this also depends on the type of GPU the model is running as well on how many processes the GPU generates simultaneously.

5.2.1.2 Model testing and accuracy assessment

For the wind turbine case study since the evaluation of the results consist on a relatively challenging process in order to discover the real detection capability of the model, the investigation of each of the three-input datasets separately is considered. On the other hand, this investigation helps on evaluation of datasets which have a geographical coverage from small scale to significantly large scale. This scale is measured considering the number of images utilized for the accuracy assessment as well as the geographical distribution.

Table 9. Anchor boxes from testing the Wind Turbine Model

Testing Wind Turbine Model	Records
Predictions **TOTAL**	9766
Ground Truth **TOTAL**	8489
Prediction **GPPD**	3638
Ground Truth **GPPD**	2567
Prediction **USWTDB**	3471
Ground Truth **USWTDB**	3104
Prediction **ESRI Brandenburg**	2657
Ground Truth **ESRI Brandenburg**	2818

On **Table 9** are shown the number of records used for the testing of the Wind Turbine model as ground truths and the number of records obtained by the model on the testing phase as predictions. Those records are separated using the database source of geolocations, as well a total accuracy which includes all databases is calculated.

In addition, an important parameter to configure before testing the model is the detection threshold. As the model predictions are the bounding box coordinates, a confidence score for the box being true object is generated in the output, this confidence score is set to consider only the boxes which are above 0.8 or 80 %, the rest of the boxes are not considered as the model output for the accuracy assessment.

5.2.2 Coal Power Plant Model

5.2.2.1 Training and validation

The training and validation of the Coal Power Plant model is performed based on a randomly dataset split where 72% of dataset images are used for training and 18% for validation (see **Figure 39**). As might occur areas where images overlap each other and same coal power plant appear in

multiple images, the split of the dataset is performed based on a rule that a single coal power plant is classified only for training or validation, thus not for both processes.

The dataset used for annotation of the Coal Power Plant model is obtained from GPPD contain a global distribution of the plants of different capacity generation.

Moreover, as experiments three different models are considered with different hyperparameter configuration, where the accuracy obtained from those models is evaluated afterwards. The dataset split remains the same for all experiments. For the Coal Power Plant model is decided to test the power of three different models for the only reason that for testing the model a hybrid accuracy assessment of the testing dataset is performed. This hybrid accuracy assessment makes use of ground truth boxes and externally utilizing orthoimages ArcGIS in order to validate model predictions. The decision of using a hybrid accuracy assessment is explained in Subsubsection 5.2.2.2.

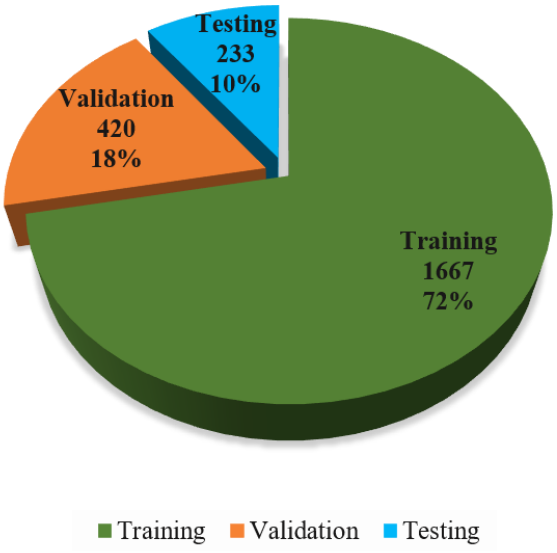


Figure 39. Coal power plant dataset split

5.2.2.1.1 Hyperparameter configuration

The hyperparameter configuration whose influence in DL models is associated with a brief introduction for their functionalities Subsection 2.3.2. For the Coal Power Plant model, the hyperparameter tuning remains mainly the same for all the models due to their capability to have a smooth converge when training. On **Table 10** are shown the hyperparameter configurations for each of the model trained.

Table 10. Hyperparameter configuration for the Coal Power Plant Models

Hyperparameter configuration			
Hyperparameter	**Model 1**	**Model 2**	**Model 3**
Epochs	150	200	150
Batch size	8	12	12
Image width/height [px]	512 x 512	512 x 512	512 x 512
Bounding box size [px]	50 x 50	50 x 50	100 x 100
Starting Learning rate	0.001	0.001	0.001
Number of workers	4	4	4
Learning rate patience	5	5	5
Anchor size	32, 68, 128, 256, 512	32, 68, 128, 256, 512	32, 68, 128, 256, 512
Aspect ratio	0.5, 1, 2	0.5, 1, 2	0.5, 1, 2
Optimizer	SGD	SGD	SGD
Momentum	0.9	0.9	0.9
Learning rate scheduler	ReduceOnPlateau	ReduceOnPlateau	ReduceOnPlateau
Data Augmentation	Flipping, Rotation	Flipping, Rotation	Flipping, Rotation

The Coal Power Plant Models are capable to attain a smooth convergence dynamic when training, which convergence decreases the loss and accuracy graph oscillations after certain epochs, this result to no further improvement on the accuracy therefore the model stops training. For model 1 and model 3 the rounds of optimizations are considered sufficient after 150 epochs, even with different batch size of 8 and 12.

On the other hand, for the 2nd model the optimization rounds are left intentionally 50 epochs more in order to expect improvement as based on research few more epochs might compass the model to reach the necessary training. For model 2 the batch size is increased to 12 because of the capability for smoother convergence in comparison to batch size 8. For the training of model 3 the larger batch size specifically 12 resulted to be the best fit, result which is conducted from the comparison of two first models.

The size of a Coal Power Plant electricity generator depends on their capacity and it may vary in range from hundreds of meters to kilometers. Therefore, due to capacity filtering inconsistencies on the input geo-location database, two different bounding box sizes are considered specifically 500m x 500m and 1000m x 1000m. Those experiments of training two different models with different input bounding box size helps to outline whether the model improves the accuracy when different input size is given or not.

Various learning rates are applied in order to find the learning rate which performs best on training of each model. From this resulted that on one hand, by applying learning rate (0.1 and 0.01) the model was not capable to perform a good training, on the other hand, learning rate (0.0001 or smaller) result on less stochasticity and thus no improvement on accuracy. Therefore, the starting learning rate of 0.001 is applied as the best fit which is reduced using the *ReduceOnPlateau* (see Subsection 2.3.2) algorithm after 5 patience epochs with no improvement on the accuracy.

Anchor sizes generated on the RPN module (see Subsubsection 2.3.1.1) at each anchor center has 5 different sizes and 3 aspect ratios as shown in **Table 10**, thus 15 anchors are generated at each center location of the feature maps in order to feed the RPN module and generate ROI's.

The calculation of gradients and update of weights after each epoch for Coal Power Plant Model is chosen *SGD* for which Subsection 2.3.2 introduces a detailed explanation for the algorithm functionality. This method of optimization is associated with a parameter named *Momentum* (see Subsection 2.3.2) which has the role to compass the gradient descent process on being smooth, since the SGD optimization algorithm is known for its stochasticity in converging.

During the in-situ hyperparameter tuning also other data augmentation techniques were applied such as Blurring, Shift Scale Rotate, Gaussian Noise, Random Brightness, Flipping, Rotation, resulting in no improvement of the accuracy, therefore, the techniques shown in **Figure 40** are considered because of their positive influence.

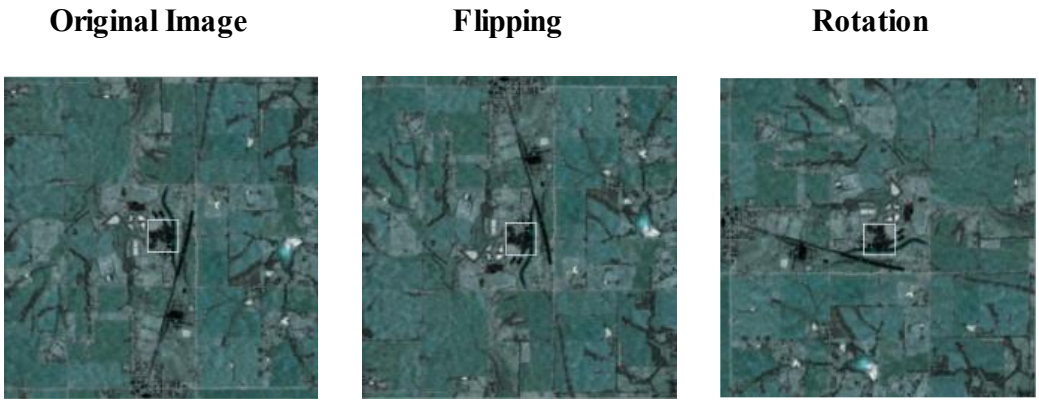


Figure 40. Data augmentation performed on Coal Power Plant Models

5.2.2.2 Coal Power plant accuracy assessment

For the performance evaluation phase of the Coal Power Plant DL model, 233 input images are separated from the dataset, who associated with the best model obtained out of training process construct the base of a very important process which indicates whether the model was trained good or not. The randomly selected 233 test images constitutes a 10% contribution on the total amount of Coal Power Plant dataset which is illustrated in the dataset split pie chart see **Figure 39**.

Table 11. Coal Power Plant model evaluation

Testing Coal Power Plant Model	Records
Predictions **Model 1**	261
Predictions **Model 2**	301
Predictions **Model 3**	284
Ground Truth	241

Furthermore, **Table 11** shows the obtained prediction records from the model and the number of ground truths from testing images. The accuracy calculation basically consists on a comparison with ground truth bounding boxes constructed based on the GPPD database of Coal Power Plants (see Subsection 4.1.2) associated with background orthoimages visualized on a GIS software. As mentioned above the Coal Power Plants depending on the capacity might cover a large area which start from hundreds of meters to kilometers this has also an impact for the model testing phase.

Consequently, as GPPD database is build upon external sources which are reported from countries or organizations, thus exist a possibility that in a geographical scope multiple coal-fired power station locations of different energy generation capacities are reported. In order to reduce the ground truth boxes misinterpretation caused by those scenarios, the utilization of orthoimages visualized in a GIS software is considered. The orthoimages assist on a more efficient manual validation of the accuracy performance of the Coal Power Plant Model. In Appendix H is shown an illustration of multi coal-fired power plants detected by the model, which in the input GPPD database utilized for annotations is missing.

5.3 Empirical results

5.3.1 Wind Turbine Model

5.3.1.1 Model training and validation results

After each epoch the model is trained and validated metrics are generated in order to track the process and if necessary, perform actions for improvement. For a more comprehensive visualization for each of the metrics graphs are generated, as model training and validation is a process which depending on dataset size requires many hours of training. Therefore, graphs help on preventing the most important and common ambiguous terms in DL techniques, model overfitting and underfitting. A more comprehensive introduction of those ambiguities is shown in the last part of Subsection 5.1.3. In addition, **Figure 41** illustrates four different graphs used for model accuracy assessment.

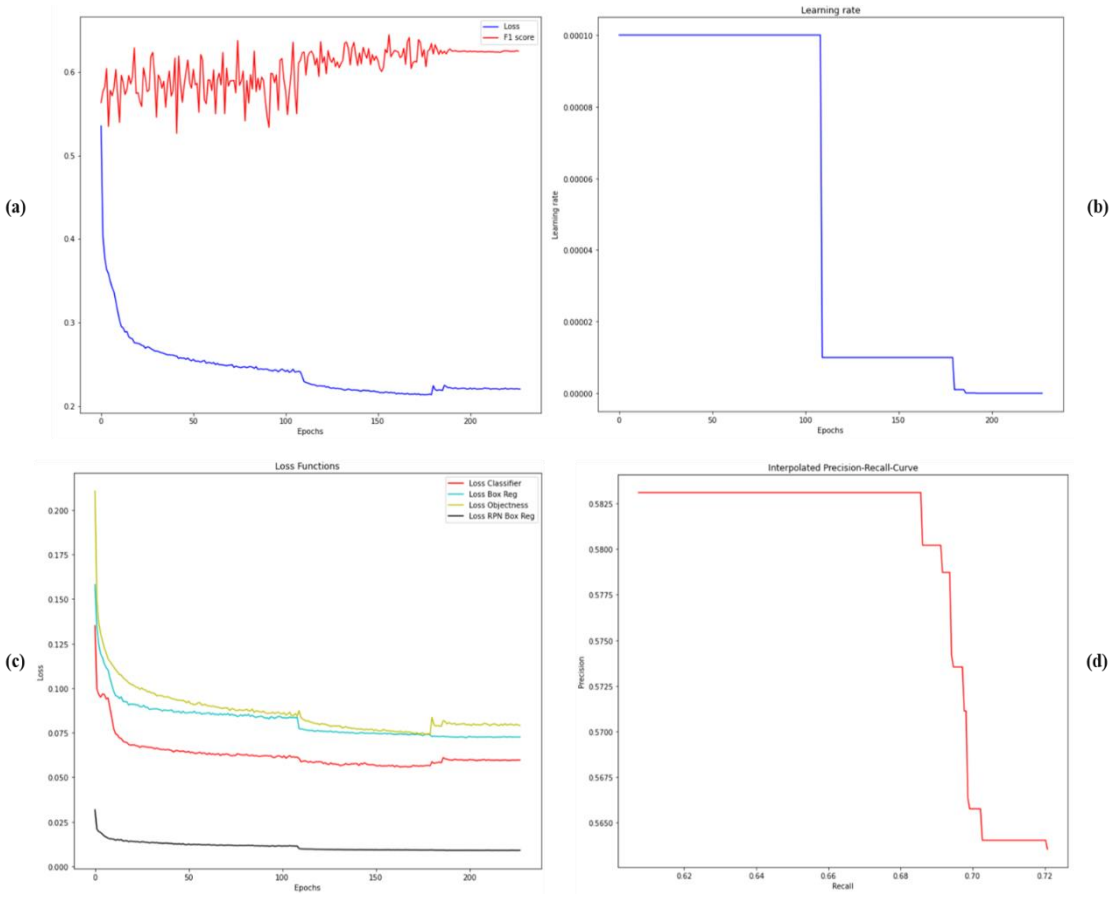


Figure 41. Wind Turbine model training and validation metrics evaluation **(a)** Model loss and F1 score curves **(b)** Model learning rate reduced based on *ReduceOnPlateau* algorithm **(c)** Other losses for classification and regression for RPN and model itself **(d)** Interpolated precision-recall curve for the wind turbine model

Figure 41 (a) shows two curves which are vital and may be considered as decision taker curves due to the information they transmit. On one hand, the blue curve is denoted as the loss curve obtained after each epoch training using multi-task loss (see Subsubsection 2.3.1.4), on the other hand the red labelled curve shows after each epoch the weighted average F1 score obtained from model Precision and Recall (see Subsection 2.3.3) when model is validated.

Further, in **Figure 41 (b)** the graph illustrates the learning rate reduction based on *ReduceOnPlateau* algorithm (see Subsection 2.3.2). Basically, this method works very efficiently on the model for the reason that the learning rate is reduced dynamically only when is needed, in contrast with other algorithms where the learning rate is predefined on which epoch a reduction is performed.

In addition, as Faster R-CNN approach utilize multiple losses such as classification or regression loss (see Subsubsection 2.3.1.4) in order to achieve best accuracy, when training the model exist the possibility to plot classification and bounding box regression loss for RPN and also for final model. **Figure 41 (c)** shows a graph for the other loss curves calculated on Faster R-CNN approach. By tracking the loss curves the most common artefacts such as overfitting and underfitting can be prevented (see Subsection 2.3.2). In addition, as well during the model training phase can be seen whether the model is following the expected synchrony as explained in Subsection 2.3.2, if not the graph shape guides where the problem occurs, afterwards fixing the issue.

Moreover, **Figure 41 (d)** shows the so-called interpolated precision-recall curve, which basically is a very useful graph for variation reductions in the precision *x* recall curve. This curve consists on interpolation precision at all recall levels, thus dividing the area under the curve in *n* number of areas. After obtaining for each validation epoch a mean Average Precision (**mAP**), mean Average Recall (**mAR**), F1 score metrics are calculated in order to evaluate their accuracy in percentage, the obtained results are shown in **Table 12** for different IoU (0.5, 0.75 and from 0.5 to 0.9 moving every 0.05) representation levels. As object detectors are trained based on ground truth boxes, for training the model in efficient way ground truth boxes are recommended to be complete. The values of validation metrics appear to be not very high for considering a very good model, and this happens for the reason due to incomplete ground truth boxes utilized as input. Therefore, the model curves are sufficient to track training performance whether the model is sufficiently learning or not based on explanation of Subsection 2.3.2.

Table 12. Wind Turbine model evaluation metrics obtained from validation

Metrics	Wind Turbine Model
mAP@.5	78.86 %
mAP@.75	67.43 %

mAP@[.5:.05:.9]	54.18 %
mAR@[.5:.05:.9]	68.28 %
F1 score	60.41 %
Training Speed [s/it]	0.32

5.3.1.2 Testing accuracy assessment results

Due to the fact that the utilized ground truth datasets are incomplete w.r.t geo-locations of every wind turbine which appear in a single image, this is more noticeable for GPPD database which was manually created. In addition, SAR imagery manifest a limitation when it comes to evaluate the predicted anchors using SAR images, due to the complexity of information on those images. Important to note is that for this reason, orthoimages visualized externally in ArcGIS software are introduced to associate the accuracy assessment process, which helps to precisely classify each predicted anchor.

On the other hand, in order to have a synchronous process between the acquisition time of orthoimages (used for visual inspection), SAR images (based on which the anchors are predicted) and ground truth boxes, the following scenario is applied. If a ground truth box exists for a predicted anchor, the prediction is classified w.r.t the ground truth box, otherwise, in occasions where no ground truth box exist the prediction is classified w.r.t the orthoimage.

Therefore, each of the predicted anchors is categorized based on evaluation metrics as TP, FP, FN samples (see Subsection 2.3.3). In addition, this categorization helps on constructing the Confusion Matrix (see Subsection 2.3.3), which is applied intentionally separately for all the input datasets that contribute on testing dataset of Wind Turbine Model (see **Figure 37**), as well these datasets are considered as one, thus calculating for the model a mean success rate. On **Table 13** are shown the categorized predicted samples as TP, FP, FN for the Wind Turbine Model based on the input test datasets.

As in testing phase participate overlapping images, a single wind turbine might appear multiple times, thus, for each time a ground truth box is generated for same wind turbine. This offers the capability to have multiple predictions for the same wind turbine, which can be afterwards optionally filtered in order to produce a more accurate final geo-location database. In this master thesis the focus is to reveal the power of the constructed methodology for the prediction of energy systems and their parameters by DL techniques and SAR images. For this reason, no filtering of

the predicted anchors is applied due to the interest on every predicted anchor. Form this can be found how many anchors is the model is capable to predict and how many from those anchors can be correctly classified without being filtered.

Nevertheless, when categorizing the model predicted anchors occur two scenarios only for Wind Turbines FN samples:

First scenario of FN samples is when for instance a single wind turbine location generates 10 ground truth boxes but the model is capable to make 5 predictions. This scenario appears to benefit from having redundant data because the model is capable to generate at least one geo-location which is sufficient for a final filtered dataset.

Second, is the scenario when for instance a single wind turbine location generates 10 ground truth boxes but the model is capable to make 0 predictions. This event is inferior compared to the first scenario but still it appears, and indeed with more influence if a filtered dataset would be considered. However, as the objective in this master thesis is to analyse the true power of the trained model and not generating a single geo-location database both scenarios of the FN samples are added together in one FN column in **Table 13**.

Table 13. Wind Turbine Model prediction samples categorization and accuracy assessment

Wind Turbine Model	TP	FP	FN	Precision [%]	Recall [%]	F1 score [%]	Test Speed [s/it]
GPPD	2803	855	313	76.66	89.95	82.75	0.32
USWTDB	2788	683	168	80.32	94.31	86.67	0.32
ESRI Brandenburg	2555	102	161	96.16	94.07	**95.10**	0.32
TOTAL	**8146**	**1640**	**642**	**83.24**	**92.69**	**87.71**	**0.32**

Furthermore, by means of the categorized samples such as TP, FP, FN first for the Wind Turbine Model is evaluated the effectiveness by examining Precision and Recall metrics (see Subsection 2.3.3), which allows on understanding based on what proportions positive predictions are correct or identified correctly. The examined metrics besides the transmission of such important information for model effectiveness, additionally assist on calculation of the weighted accuracy F1 score (see Subsection 2.3.3).

Moreover, by analysing the empirical results obtained on **Table 13**, evidently it appears that the trained model is capable to achieve an overall F1 score **87.71 %**, nevertheless, when comparing the three individual datasets is plainly visible that a larger geographical scope on image distribution, impacts on decreasing the accuracy, this due to many terrain artefacts that might appear on SAR images of large geographical scope. Additionally, Appendix I are shown some of the most common artefacts that appear when visually assessing each of the predicted anchor using orthoimages.

Next, on the upper part of **Figure 42** row **(a)** visualizes a comprehensive example of predicted anchors labelled with blue color and ground truth boxes labelled with red color, which overlay the SAR images utilized for testing phase of the model. As it can be seen on the images is relatively challenging to classify with naked eye whether the new prediction is correct or not.

However, this demonstrates that in order to have a perfect accuracy assessment process are compulsory required either very accurate input ground truth boxes or orthoimages for new prediction categorization.

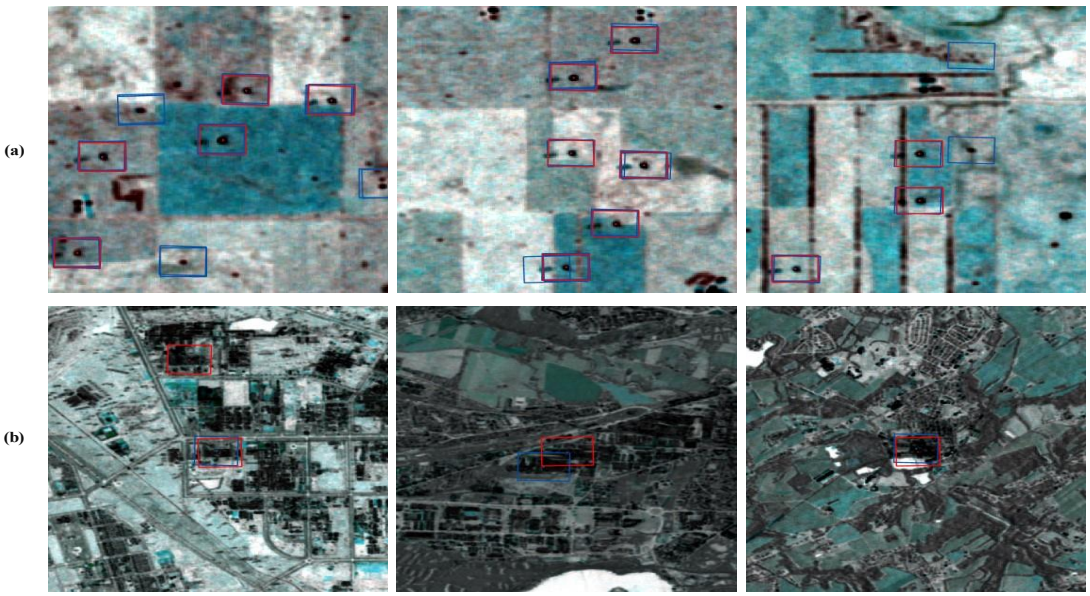


Figure 42. Illustration of predicted anchors *(blue)* and ground truth boxes *(red)* over test images **(a)** Wind Turbine test images **(b)** Coal Power Plant test images

5.3.2 Coal Power Plant Model

5.3.2.1 Model training and validation results

Although Coal Power Plant models are trained using different hyperparameter configuration, plotting the obtained empirical results in the same graph allows on a more comprehensive evaluation

of the training metrics. These comparisons create the possibility for analysing the output from models and conclude further improvements.

However, in **Figure 43** only the most important graphs are visualized in comparison with the graphs of Wind Turbine Model (see **Figure 41**), this due to the large number of curves in same graph, that might bring a difficulty on reading since for the Coal Power Plant consist of three trained models. Therefore, classification and regression graphs which generates 4 curves for each model are skipped and only the multi-task loss is shown, however, these graphs are analysed externally.

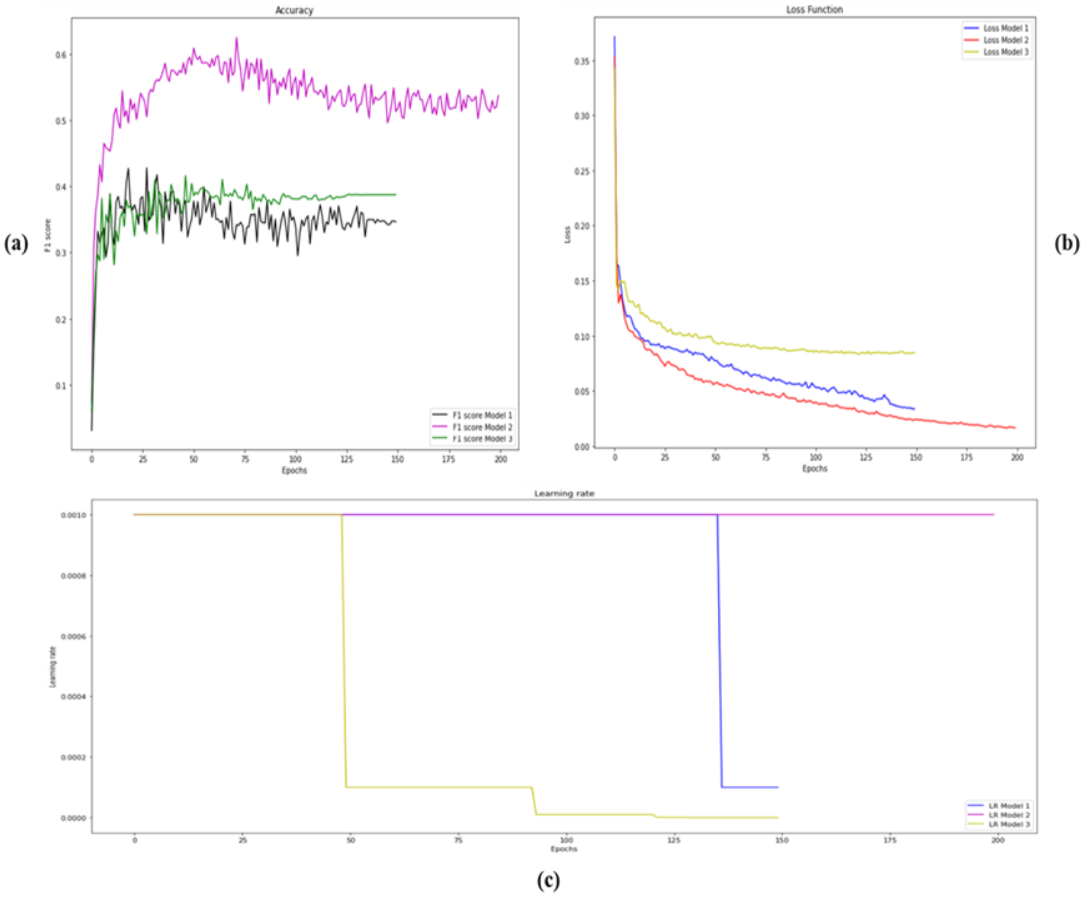


Figure 43. Coal Power Plant model training and validation metrics evaluation **(a)** F1 score obtained from training of 3 different models **(b)** Loss function for the coal power plant models **(c)** Reduction of learning rate for each model based on *ReduceOnPlateau* algorithm

In **Figure 43 (a)** it is shown a graph which includes the weighted accuracy F1 score curves for each epoch obtained from training three models using Precision and Recall (see Subsection 2.3.3). By referring to the hyperparameter configuration for the Coal Power Plant models shown in **Table**

10 is obvious the impact of larger batch size for increasing the model accuracy, this is an instance of the curve labelled with magenta color.

Nevertheless, it is interesting to note that for experiment where input bounding box size was increased did not influence the accuracy improvement, this concluded based on the graph comparison in **Figure 43 (a)** with F1 score labelled in green color.

Moreover, **Figure 43 (b)** illustrates the calculated loss after each training epoch for all the models. From the visualization can be seen that the hyperparameter configuration of Model 2 are capable to converge smoothly and obtain best loss curve with less oscillations compared to other models. The loss curve of Model 1 colored in blue (see **Figure 43 (b)**) appears with some oscillations when converging, and those oscillations are also transmitted into the accuracy curve colored in black shown in **Figure 43 (a)**. The oscillations can be attributed due to certain reasons such as the gradient descent doesn't overshoot the minima or due to high learning rate, perhaps with the last one being the reason for causing high-oscillated curves.

Furthermore, **Figure 43 (c)** illustrates the reduction of learning rate based *ReduceOnPlateau* algorithm (see Subsection 2.3.2) where is interesting to note two occasions. First occasion appears that when lowering the learning rate on Model 1 and 3, F1 score has lower oscillation but as well no improvement on the accuracy. Second occasion, is the Model 2 learning rate that remains constant due to low oscillations on loss function, which causes no reduction of learning rate, but still with small stochastic curve of F1 score.

Table 14. Coal Power Plant model evaluation metrics obtained from validation

Metrics	Model 1	Model 2	Model 3
mAP@.5	50.12 %	75.34 %	53.33 %
mAP@.75	38.29 %	55.23 %	41.98 %
mAP@[.5:.05:.9]	29.84 %	49.71 %	31.31 %
mAR@[.5:.05:.9]	42.11 %	58.57 %	46.18 %
F1 score	34.81 %	53.68 %	37.26 %
Training Speed [s/it]	0.89	1.25	1.61

In addition, by obtaining mean Average Precision (mAP), mean Average Recall (mAR), F1 score metrics their mean values can be calculated for model training phase in order to evaluate their accuracy in percentage, results are shown in **Table 14** for different IoU (0.5, 0.75 and 0.5 to 0.9 every 0.05) representation levels

5.3.2.2 Testing accuracy assessment results

In order to evaluate the real prediction power of the constructed model, an accuracy assessment in a test dataset with images that the model has not seen in earlier phases is performed. Important to note on testing images is that all images utilized on testing phase contain at least one Coal Power Plant, thus no images without Coal Power Plant exist. This accuracy assessment reveals the success rate the model is capable to achieve. Hence, for the Coal Power Plant scenario an accuracy assessment is performed on each model. The prediction CSV file where an instance of it is shown on **Table 7**, generates a confidence score of the prediction being correct object or not.

For a more sustainable result, the confidence score is filtered via a threshold where the anchors with score is above 80% are considered, the rest with lower then 80% are declassified.

On one hand each model produces a CSV output which number of prediction records is shown in **Table 11**, on the other hand **Table 11** shows the number of ground truth boxes which associated with orthoimages visualized in ArcGIS Pro assist on a more comprehensive accuracy assessment.

Furthermore, results are obtained after categorizing each of the predicted anchor for all the evaluated models utilizing the accuracy assessment methods introduced on Subsection 2.3.3, thus, with these results **Table 15** is constructed. Important to notice on the final results is that Model 3 even though has the input ground truth boxes at larger size, does not exceed the number of predicted positive samples in comparison with Model 2, but only generates less boxes, thus because the boxes are of bigger size they can be classified as FP samples instead of FN samples.

As a result, shown in **Table 15** can be seen that Model 3 has only improvement on generating less FP samples but not lowering the amount of FN samples. Last but not least, in **Figure 42 (b)** it is illustrated a scene instance for the predicted anchors *(blue color)* and ground truth boxes *(red color)* overlaying on test images.

Table 15. Coal Power Plant models testing accuracy assessment

Model	TP	FP	FN	Precision [%]	Recall [%]	F1 score [%]	Test Speed [s/it]
Model 1	226	32	21	87.59	91.49	89.50	0.43
Model 2	256	44	16	85.33	94.11	89.51	0.69
Model 3	256	27	15	**90.45**	**94.44**	**92.41**	0.69

5.4 Critical discussion and recommendation

As the models for the detection of Wind Turbines and Coal Power Plants are performed separately because their input image size differs, conclusions for each model are obtained. For the **Wind Turbine** model from the empirical results is concluded that:

By changing anchor sizes of the anchor generator on the RPN (see Subsubsection 2.3.1.1) in a single feature map layer is not sufficient to improve the accuracy. For this purpose, three different sets of anchor sizes were tested in order to generate the anchors:

- [8, 16, 32, 64, 128], this set of anchors was capable to generate 915 boxes.
- [16, 32, 64, 128, 256], this set of anchors was capable to generate 2415 boxes.
- [32, 64, 128, 256], this set of anchors was capable to generate 9766 boxes.

Even though the input of the ground truth is 20 pixels x 20 pixels, from the analysis of the predicted anchors of the model, can be concluded that the smaller anchors are not considered by the model and therefore declassified because of their small IoU.

By testing the model with overlapping images, which afterwards produces for a single wind turbine location redundant data, allows on finding a more accurate prediction.

The selection of data augmentation techniques (see last part of Subsection 5.1.3) results as a very sensitive process using SAR images, by applying wrong data augmentation in any of the processes training or validation results on decreasing the accuracy. Further research for finding proper data augmentation in order to artificially increase the image dataset help on improving the accuracy.

By introducing the other backbone of ResNet (see Subsubsection 2.3.1.5) which is 101 layers deep was not sufficient for the process to improve the accuracy.

From the outcome of Coal Power Plant model in smaller dataset but larger image size was very good in numbers, is concluded that increasing the image size helps on accuracy improvement.

Out of analysis of the empirical results for the **Coal Power Plant** model is found that:

SAR imagery combined with DL techniques manifest the capability to achieve high accuracy (also designated as F1 score in **Table 15**) of **92.41%** even when utilizing small datasets.

As the size of Coal Power Plants differs based on their electricity generation capacity an experiment for changing the input bounding box size was performed. Out of this experiment was concluded that by increasing the ground truth bounding box size from 500m x 500m into 1000m x 1000m does not directly influence the model to increase the accuracy and obtain more true positive samples. On Appendix H is shown an example which clearly state that the centroid of the bounding box in both scenarios is almost at the same location.

The selection of data augmentation also for the Coal Power Plant images results as a very sensitive process, where a wrong choice of data augmentation can easily influence bad in pertinent process where it has been applied, thus decrease the accuracy. Some of the data augmentations tested and resulted with negative effects can be mentioned: Blur, Shift Scale Rotate, Gaussian Noise, Random Brightness.

As the input geo-locations of the Coal Power Plants are obtained using the open database GPPD, results to have some object misplacement inconsistencies when it comes to the accuracy of the geo-locations. Although this inconsistency appears in the database in low scale, nevertheless the influence on the model either during training and validation or during testing are considered. Those inconsistencies became more visible when the predictions are externally visualized on a GIS software.

On Appendix H is shown an example of the inconsistency scenario where the input bounding box is not geo-located exactly into the target object, but still as this inconsistency is in a very low scale the model is capable to still predict correctly. Hence, this turns the predictions evaluation process relatively challenging for the Coal Power Plants, as the input bounding boxes are not accurate. By obtaining more accurate input bounding boxes will compass the process on a smoother training and validations without inconsistencies and allow for an automatic accuracy assessment.

6 Conclusion and outlook

The generation of a database at arbitrary geographical scales for energy infrastructure constitute a high demand for improving the efficiency on the energy transition process and not only. Hence, the already developed methodologies as shown in Subsubsection 1.1.1.2 still are limited at certain level. This research aimed at developing a novel methodology for automatic detection of energy system parameters and automatic generation of a database on detected energy systems. The proposed method is developed based on utilization of state-of-the-art technologies such as DL techniques and open EO data from Sentinel-1 SAR images.

The methodology is capable to perform well and ensure high-accuracy for different geographical scope for individual Wind Turbines and Coal-fired Power Plants detection. The proposed methodology demonstrates the capability to ensure an overall accuracy of **87.71 %** for large-scale application on detection of individual Wind Turbines. In addition, for large-scale application on detection of Coal-fired Power Plants the model demonstrates the capability of ensuring a database with **92.41%** accuracy.

However, as can be observed in general the models have also negative bias which can be reduced based on conclusions conducted in this thesis. Further, possible methodology improvements can be considered based on each model limitations (see Section 5.4)

Based on a qualitative and quantitative analysis is conducted a study in order to answer two important research questions:

- Can nowadays the state-of-the-art technologies such as SAR imagery and DL techniques provide a valuable resource for the creation of a database which contains onshore energy system parameters, thus help on filling the existing gap and assist on a more efficient energy transition process?
- Can the chosen approach ensure high-detection accuracy also for large scale applications?

The answer to the first question is yes. By analysing the outcome of the developed model is concluded that the proposed technique on this master thesis demonstrate a high potential, which will lead towards the solution for creating a database containing the energy system parameters. This solution assist on contributing on filling the accurate geo-located energy system database gap and help for efficiently improve the energy transition process.

Based on the empirical results Faster R-CNN approach as a state-of-the-art two stage detector clearly manifest the capability to ensure the high-detection accuracy when combined with SAR imagery. On the other hand, the selection of Faster R-CNN architecture to model the approach

was performed based on the investigation of existing researches, therefore an outcome comparison with state-of-the-art developments on one stage detectors e.g. YOLOX [63], Retina Net [64] or Mask R-CNN [65] might engage the interest for further improvements.

Further researches on this field would be a great help for the future planning of the energy transition. This research provides a contribution for opening new perspectives also for the identification of energy system parameters for other energy systems such as: Gas, Nuclear, Biomass, Hydropower, Solar etc.

The answer to the second question is also yes, but in order to particularly analyse the real detection capability of the model for large-scale applications, on the testing phase three different independent datasets were included. By analysing the predictions obtained out of those datasets can be concluded that the model is capable to perform high-accuracy also for large scale applications. However, here has to be considered a small decrease on the accuracy due to change on terrain reflections in SAR images.

Besides many advantages that can be mentioned when combining SAR imagery and DL techniques, there exist also limitations. An important limitation to mention is that in order to produce accurate annotation input data for the dataset, orthoimages are required in addition as this process is relatively challenging when using only SAR images.

In a broader context, in order to assess if a wrong implication exist on input bounding boxes for training, validation or testing process, orthoimages are required. In addition, an important element to note is that acquisition time of the orthoimages has to be synchronized in the temporal domain with the utilized SAR images depending on the target energy system. The inconsistencies are common for instance when detecting wind turbines as they might be displaced, demolished or even new constructed in short period of time.

References

[1] U.S. Energy Information Administration, "What is energy?," 06 2020. [Online]. Available: https://www.eia.gov/energyexplained/what-is-energy/forms-of-energy.php. [Accessed 18 11 2021].

[2] H. Ritchie and M. Roser, "Energy, Our World In Data," 2020. [Online]. Available: https://ourworldindata.org/energy. [Accessed 03 12 2021].

[3] "Report of the Conference of the Parties on its twenty-first session, held in Paris from 30 November to 13 December 2015. Addendum Part two: Action taken by the Conference of the Parties at its twenty-first session," 2016. [Online]. Available: https://www.un.org/en/development/desa/population/migration/generalassembly/docs/globalcompact/FCCC_CP_2015_10_Add.1.pdf. [Accessed 24 01 2022].

[4] Federal Ministry for Economic Affairs and Energy, "Making a success of the energy transition," 2015. [Online]. Available: https://www.bmwi.de/Redaktion/EN/Publikationen/making-a-success-of-the-energy-transition.pdf. [Accessed 25 11 2021].

[5] European Commission, "INFRASTRUCTURE ANALYTICAL GRID FOR ENERGY INFRASTRUCTURE," 2020. [Online]. Available: https://ec.europa.eu/competition/state_aid/modernisation/grid_energy_en.pdf. [Accessed 22 11 2021].

[6] Fraunhofer ISE, "Fraunhofer Institute for Solar Energy Systems," 2021. [Online]. Available: https://www.ise.fraunhofer.de/en/business-areas/power-electronics-grids-and-smart-systems/energy-system-analysis.html. [Accessed 25 11 2021].

[7] Chair of Renewable and Sustainable Energy Systems, "Application of energy system models," 2021. [Online]. Available: https://www.ei.tum.de/en/ens/research/topics/application-energy-models/. [Accessed 25 11 2021].

[8] Byers et al, "A GLOBAL DATABASE OF POWER PLANTS," World Researche Institute, 03 2021. [Online]. Available: https://files.wri.org/d8/s3fs-public/2021-07/global-power-plant-database-technical-note-v1.3.pdf. [Accessed 21 12 2021].

[9] ESA, "Sentinel-1 SAR Technical Guide," 2014. [Online]. Available: https://sentinels.copernicus.eu/web/sentinel/technical-guides/sentinel-1-sar. [Accessed 15 02 2022].

[10] J. M. Amigo, "Chapter 1.1 - Hyperspectral and multispectral imaging: setting the scene,Elsevier,Volume 32,Pages 3-16," 2020. [Online]. Available: https://doi.org/10.1016/B978-0-444-63977-6.00001-8. [Accessed 05 03 2022].

[11] Sentinel User Handbook and Exploitation Tools, ESA, "Sentinel -2 User Handbook," 2015. [Online]. Available: https://sentinel.esa.int/documents/247904/685211/Sentinel-2_User_Handbook. [Accessed 05 03 2022].

[12] Chang et al., "Ship Detection Based on YOLOv2 for SAR Imagery," 2019. [Online]. Available: https://doi.org/10.3390/rs11070786. [Accessed 10 03 2022].

[13] Xiang Zhu et al., "Deep Learning Meets SAR," 2021. [Online]. Available: arXiv:2006.10027v2. [Accessed 10 03 2022].

[14] S. Berrick and NASA Officials, "Remote Sensors," Unknown. [Online]. Available: https://earthdata.nasa.gov/learn/remote-sensors. [Accessed 25 03 2022].

[15] M. I. Skolnik, ""radar". Encyclopedia Britannica," 2021. [Online]. Available: https://www.britannica.com/technology/radar/History-of-radar. [Accessed 24 01 2022].

[16] Moreira et al, "A Tutorial on Synthetic Aperture Radar," 03 2013. [Online]. Available: https://elib.dlr.de/82313/1/SAR-Tutorial-March-2013.pdf. [Accessed 26 01 2022].

[17] A. Flores, K. Herndon, R. Thapa, E. Cherrington and P. NASA, "The SAR Handbook: Comprehensive Methodologies for Forest Monitoring and Biomass Estimation, p. 21-40," 04 2019. [Online]. Available: DOI: 10.25966/nr2c-s697. [Accessed 26 01 2022].

[18] U.S. DEPARTMENT OF COMMERCE National Telecommunications and Information Administration, "Federal Radar Spectrum Requirements Report, Chapter 3 Radars and the Electromagnetic Spectrum," 2000. [Online]. Available: https://www.ntia.doc.gov/files/ntia/publications/ntia00-40.pdf. [Accessed 25 01 2022].

[19] NASA-ISRO SAR Mission, "GET TO KNOW SAR," Unknown. [Online]. Available: https://nisar.jpl.nasa.gov/mission/get-to-know-sar/polarimetry/. [Accessed 26 01 2022].

[20] ESA, "Sentinel-1," 2014. [Online]. Available: https://sentinels.copernicus.eu/web/sentinel/missions/sentinel-1. [Accessed 27 01 2022].

[21] ESA, "Sentinel-1 SAR User Guide," 2014. [Online]. Available: https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-1-sar/acquisition-modes. [Accessed 28 01 2022].

[22] I. Goodfellow, Y. Bengio and A. Courville, "Deep Learning," 2016. [Online]. Available: https://www.deeplearningbook.org/. [Accessed 17 01 2022].

[23] O. Christopher, "Understanding LSTM Networks," 27 08 2015. [Online]. Available: https://colah.github.io/posts/2015-08-Understanding-LSTMs/. [Accessed 30 01 2022].

[24] A. Creswell, T. White, V. Dumoulin, . K. Arulkumaran, B. Sengupta and A. A. Bharath, "Generative Adversarial Networks: An Overview," 19 10 2017. [Online]. Available: arXiv:1710.07035v1. [Accessed 30 01 2022].

[25] G. Borrageiro, N. Firoozye and P. Barucca, "Online Learning with Radial Basis Function Networks," 24 12 2021. [Online]. Available: arXiv:2103.08414v6. [Accessed 30 01 2022].

[26] L. Khacef, L. Rodriguez and B. Miramond, "Improving Self-Organizing Maps with Unsupervised Feature Extraction," 04 09 2020. [Online]. Available: arXiv:2009.02174v1. [Accessed 30 01 2022].

[27] D. Bank, N. Koenigstein and R. Giryes, "Autoencoders," 03 04 2021. [Online]. Available: arXiv:2003.05991v2 , p.1-4. [Accessed 30 01 2022].

[28] L. Alzubaidi, J. Zhang and A. e. a. Humaidi, "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. J Big Data 8, 53," 2021. [Online]. Available: https://doi.org/10.1186/s40537-021-00444-8. [Accessed 30 01 2022].

[29] F.-F. Li, J. Johnson and S. Yeung, "Stanford University Lecture 11:Detection and Segmentation," 2017. [Online]. Available: http://cs231n.stanford.edu/slides/2017/cs231n_2017_lecture11.pdf#page=17. [Accessed 16 02 2022].

[30] T. Hoeser and C. Kuenze, "Object Detection and Image Segmentation with Deep Learning on Earth Observation Data: A Review-Part I: Evolution and Recent Trends, Remote Sensing, vol. 12, no. 10, p. 1667," 2020. [Online]. Available: doi: 10.3390/rs12101667. [Accessed 10 03 2022].

[31] Ren et al., "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," 06 01 2016. [Online]. Available: arXiv:1506.01497v3. [Accessed 13 02 2022].

[32] R. Girshick, "Fast R-CNN," 27 09 2015. [Online]. Available: arXiv:1504.08083v2. [Accessed 13 02 2022].

[33] Torch Contributors, "PyTorch Documentation," 2019. [Online]. Available: https://pytorch.org/docs/stable/nn.html. [Accessed 22 03 2022].

[34] Lin et al, "Feature Pyramid Networks for Object Detection," 19 04 2017. [Online]. Available: arXiv:1612.03144v2. [Accessed 27 02 2022].

[35] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," 2015. [Online]. Available: arXiv:1409.1556v6. [Accessed 27 02 2022].

[36] He et al, "Deep Residual Learning for Image Recognition," 2015. [Online]. Available: arXiv:1512.03385v1. [Accessed 27 02 2022].

[37] M. Hashemi, B. Rabus and S. Lehner, "Ocean Feature Extraction from SAR Quicklook Imagery using Convolutional Neural Networks," 06 2018. [Online]. Available: https://elib.dlr.de/117706/. [Accessed 26 03 2022].

[38] Amazon Web Services, "Amazon Machine Learning Developer Guide," Unknown. [Online]. Available: https://docs.aws.amazon.com/machine-learning/latest/dg/machinelearning-dg.pdf#model-fit-underfitting-vs-overfitting. [Accessed 15 03 2022].

[39] J. B. Diederik P. Kingma, "Adam: A Method for Stochastic Optimization," 2017. [Online]. Available: arXiv:1412.6980v9. [Accessed 18 01 2022].

[40] Ng et al, "Unsupervised Feature Learning and Deep Learning Tutorial," 2014. [Online]. Available: http://deeplearning.stanford.edu/tutorial/supervised/OptimizationStochasticGradientDescent/. [Accessed 19 01 2022].

[41] S. Ruder, "An overview of gradient descent optimization algorithms," 2017. [Online]. Available: arXiv:1609.04747v2. [Accessed 23 01 2022].

[42] R. S. Sutton, "Two problems with backpropagation and other steepest-descent learning," 1986. [Online]. Available: http://incompleteideas.net/papers/sutton-86.pdf. [Accessed 24 01 2022].

[43] G. Orr, N. Schraudolph and F. Cummins, "Willamette University, CS-449: Neural Networks," 1999. [Online]. Available: https://www.willamette.edu/~gorr/classes/cs449/intro.html.

[44] Torch Contributors, "PyTorch Documentation," 2019. [Online]. Available: https://pytorch.org/docs/stable/optim.html#. [Accessed 23 03 2022].

[45] T. Hoeser and C. Kuenze, "Object Detection and Image Segmentation with Deep Learning on Earth Observation Data: A Review—Part II: Applications," 2020. [Online]. Available: https://doi.org/10.3390/rs12183053. [Accessed 15 02 2022].

[46] Jiao et al, "A Survey of Deep Learning-based Object Detection, ArXiv," 2019. [Online]. Available: arXiv:1907.09408v2. [Accessed 05 03 2022].

[47] Alexander et al., "Mapping Electricity Infrastructure with Deep Learning," 2020. [Online]. Available: https://bassconnections.duke.edu/sites/bassconnections.duke.edu/files/site-images/mapping-electricity-infrastructure-2019.pdf. [Accessed 10 03 2022].

[48] H. Zhang and Q. Deng, "Deep Learning Based Fossil-Fuel Power Plant Monitoring in High Resolution Remote Sensing Images: A Comparative Study Remote Sensing 11, no. 9: 1117," 2019. [Online]. Available: https://doi.org/10.3390/rs11091117. [Accessed 10 03 2022].

[49] M.-Á. Manso-Callejo, C.-I. Cira, R. Alcarria and J.-J. Arranz-Justel, "Optimizing the Recognition and Feature Extraction of Wind Turbines through Hybrid Semantic Segmentation Architectures," 2020. [Online]. Available: https://doi.org/10.3390/rs12223743. [Accessed 09 03 2022].

[50] Guishen Tian et al, "Journal of Physics: Conference Series 1453 012117," 2020. [Online]. Available: doi:10.1088/1742-6596/1453/1/012117. [Accessed 09 03 2022].

[51] GEE, "Sentinel-1 Algorithms," Unknown. [Online]. Available: https://developers.google.com/earth-engine/guides/sentinel1. [Accessed 05 03 2022].

[52] F. Filipponi, "Sentinel-1 GRD Preprocessing Workflow," 2019. [Online]. Available: https://doi.org/10.3390/ECRS-3-06201. [Accessed 05 01 2022].

[53] J. Park, A. Korosov and M. Babiker, "Efficient thermal noise removal of Sentinel-1 image and its impacts on sea ice applications. 19th EGU General Assembly, EGU2017, proceedings from the conference held 23-28 April, 2017 in Vienna, Austria., p.12613," 04 2017. [Online]. Available: Bibcode: 2017EGUGA..1912613P. [Accessed 05 01 2022].

[54] Google Earth Engine, "Sentinel-1 SAR GRD: C-band Synthetic Aperture Radar Ground Range Detected, log scaling," 2021. [Online]. Available: https://developers.google.com/earth-engine/guides/sentinel1. [Accessed 05 01 2021].

[55] Russakovsky et al, "ImageNet Large Scale Visual Recognition Challenge," 2015. [Online]. Available: arXiv:1409.0575v3. [Accessed 04 01 2022].

[56] Lin et al, "Microsoft COCO: Common Objects in Context," 2015. [Online]. Available: arXiv:1405.0312v3. [Accessed 04 01 2022].

[57] Everingham et al, "The PASCAL Visual Object Classes (VOC) Challenge," 2009. [Online]. Available: https://doi.org/10.1007/s11263-009-0275-4. [Accessed 04 01 2022].

[58] World Resources Institute (WRI), "Global Power Plant Database," 03 2021. [Online]. Available: https://www.wri.org/research/global-database-power-plants. [Accessed 04 01 2022].

[59] B. D. Hoen, J. E. Diffendorfer, J. T. Rand, L. A. Kramer, C. P. Garrity and H. E. Hunt, "United States Wind Turbine Database (ver. 4.2, November 2021): U.S. Geological Survey, American Clean Power (ACP) Association, and Lawrence Berkeley National Laboratory data release," 2018. [Online]. Available: https://doi.org/10.5066/F7TX3DN0. [Accessed 04 01 2021].

[60] Esri Germany, "Windkraftanlagen - Brandenburg," 2019. [Online]. Available: https://opendata-esri-de.opendata.arcgis.com/datasets/454544115ba547e7b5e7bffd66c7ff2b_0. [Accessed 04 01 2022].

[61] Pytorch, "Pytorch," [Online]. Available: https://pytorch.org/features/ . [Accessed 02 01 2022].

[62] Tech Power Up, "NVIDIA GeForce GTX TITAN X," Unknown. [Online]. Available: https://www.techpowerup.com/gpu-specs/geforce-gtx-titan-x.c2632. [Accessed 07 01 2022].

[63] Gee et al, "YOLOX: Exceeding YOLO Series in 2021," 2021. [Online]. Available: arXiv:2107.08430v2. [Accessed 22 02 2022].

[64] Lin et al, "Retina Net: Focal Loss for Dense Object Detection," 2018. [Online]. Available: arXiv:1708.02002v2. [Accessed 22 02 2022].

[65] He et al, "Mask R-CNN," 2018. [Online]. Available: arXiv:1703.06870v3. [Accessed 22 02 2022].

[66] ESA, "Radiometric Calibration of S-1 Level-1 Products Generated," 2015. [Online]. Available: https://sentinel.esa.int/documents/247904/685163/S1-Radiometric-Calibration-V1.0.pdf. [Accessed 06 01 2022].

[67] ESA, "Level-1 SLC Products," 2015. [Online]. Available: https://sentinels.copernicus.eu/web/sentinel/technical-guides/sentinel-1-sar/products-algorithms/level-1-algorithms/single-look-complex. [Accessed 05 01 2022].

[68] European Space Agency, "Level-1 Radiometric Calibration," 2015. [Online]. Available: https://sentinels.copernicus.eu/web/sentinel/radiometric-calibration-of-level-1-products. [Accessed 05 01 2022].

[69] Farr et al, "The Shuttle Radar Topography Mission," 19 05 2007. [Online]. Available: https://doi.org/10.1029/2005RG000183. [Accessed 05 01 2022].

[70] NASA's Land Processes Distributed Active Archive Center, "New Version of the ASTER GDEM," 05 08 2019. [Online]. Available: https://earthdata.nasa.gov/learn/articles/new-aster-gdem. [Accessed 05 01 2022].

[71] F. Gotzens, H. Heinrichs, J. Hörsch and F. Hofmann, "Performing energy modelling exercises in a transparent way - The issue of data quality in power plant databases," 24 11 2018. [Online]. Available: https://doi.org/10.1016/j.esr.2018.11.004. [Accessed 25 11 2021].

Appendix

Appendix A. Schematic illustration of applied mathematical concept in a single node in the NNs**.**

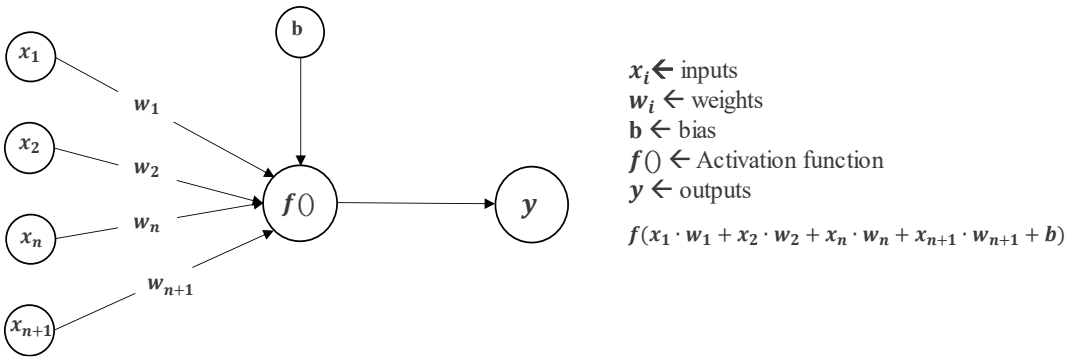


Figure 44. Illustration of single node in NNs

Appendix B. Radiometric calibration Look-up Tables.

LUT[11] is an image processing technique that allow image data to be passed through an array that maps an input pixel value to an output value. Four different calibration LUTs are provided on Level 1 of Sentinel-1 SAR which produce ß$_i^0$, σ_i^0 and γ_i or return to the Digital Number (DN) [66].

Beta β^0 yield a dataset containing the radar brightness coefficient, which is the ratio between the power transmitted and received by the antenna. It is related to slant range, and is dimensionless.

Sigma σ^0 specifies the strength of reflection in terms of the geometric cross section of a conducting sphere that would give rise to the same level of reflectivity

Gamma γ_i is typically used when calibrating the antenna. Since each range cell is equally distant from the satellite, near range and far range are equally bright, which is helpful in determining the antenna pattern in the output dataset.

The digital number of ESA Level-1 SLC product is calculated as shown in equation (15). Each image pixel is represented by a complex real in-phase component (I) and the imaginary quadrature component (Q) magnitude value and therefore contains both amplitude and phase information. Each of the I and Q value is 16 bits per pixel [67].

$$DN = \sqrt{(I^2 + Q^2)} \tag{15}$$

LUTs apply a range-dependent gain including the absolute calibration constant, for GRD products a constant offset is also applied [68].

Radiometric calibration is applied following one of the cases of equations (16), where depending on the application and the selection case of the LUTs only one of the formulas can be applied [66]:

$$\sigma_i^0 = \frac{|DN_i|^2}{A_i^2}$$

$$\beta_i^0 = \frac{|DN_i|^2}{A_i^2} \tag{16}$$

$$\gamma_0 = \frac{|DN_i|^2}{A_i^2}$$

[11] Look-up Tables - https://stemmer-imaging.com/en/knowledge-base/look-up-tables-lut/ Accessed on 05 01 2022

The Level 1 products of Sentinel-1 are provided with dedicated Calibration Annotation Data Set, who also contribute with four calibration LUTs to be used under the denominator A_i^2 [66].

- A_β used to transform the radar reflectivity into beta β^0 where the area normalization is aligned with the slant range
- A_σ used to transform the radar reflectivity into radar cross-section σ^0 where the area normalization is aligned with ground range plane
- A_γ used to transform the radar reflectivity into gamma γ where the area normalization is aligned with a plane perpendicular to slant range
- A_{dn} used to revert for the final pixel scaling, it defines the final scaling from internal SLC to the final product coded in 16 bits.

Appendix C. Levels for products Sentinel-1 satellites are capable to generate.

Level 0 products are available only in SM, IW and EW modes and they consist of raw data which include noise, internal calibration and orbit metadata. The products are compressed at variable bit rate, therefore for brighter backscatters the number of allocated bits is increased. Out of level 0 products all other levels are obtained [9].

Level 1 products are the most popular data delivered to users as two types: SLC or GRD. This means that the products are derived from level 0 raw data, thus pre-processed in range, azimuth direction and internal calibration, georeferenced using orbit metadata and attitude information [9].

- Level 1 SLC are images generated in slant range by azimuth imaging plane. They consist of 16-bit images where pixels have 65,536 tonal values, each pixel of the image contain information on amplitude and phase, making them available all four different acquisition modes (see **Figure 9**). SLC images are produces with zero Doppler geometry and georeferenced using orbit metadata and attitude information of the satellite.
- Level 1 GRD are products obtained by processing multi looking images (which causes the reduction of speckle noise) projected in ground range using ellipsoidal model which is corrected by means of terrain heights that are constant in range but varies in azimuth. The coordinates are derived from slant range coordinates using earth ellipsoid and pixel value represent only the magnitude information, hence phase information is not available.

Level 2 products consist of the derived level 1 products which contain geolocated geophysical components used for ocean surface wind measurement or wave spectrum and surface radial velocity which is the velocity of surface measured parallel to the direction of antenna transmitting the pulse in either towards or away from the radar [9].

Appendix D. Optimization algorithms utilized in deep learning models

ADAM [39] (see Subsection 2.3.2) is an optimization algorithm which has a broad usage in the neural networks. The algorithm is designed as an adaptive learning rate method, where individual learning rates of parameters are calculated and afterwards from the estimations adapt the learning rate for each weight w_t of the neural network using the first and second moment of gradient. The first moment is the estimate of mean of gradient m_t and the second moment is the uncentered variance of gradient v_t which are also denoted as the moving averages. The first and second moment of gradient are initialized as vectors of 0's, and again converging towards zero. The estimation of the first and second moment biased moving averages for a mini-batch are defined in equations (17) [39]:

$$m_t = \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot g_t$$
$$v_t = \beta_2 \cdot v_{t-1} + (1 - \beta_2) \cdot g_t \tag{17}$$

where: β_1 and β_2 are hyperparameters with default values of 0.9 and 0.999 control the exponential decay rates of the moving averages. These parameters have a very good default setting and are never changed, $g_t = \nabla f(w_{t-1} - m_{t-1})$ are gradients obtained using the loss function, t - timestep, α – learning rate, $m_{t-1}\ and\ v_{t-1}$ previous timestep moving averages.

Hence, the computation of bias corrected moving averages of first and second moment of estimate is defined in equation (19) and the update of the weights in equation (18) [39]:

$$\widehat{m_t} = \frac{m_t}{(1 - \beta_1^t)}$$
$$\widehat{v_t} = \frac{v_t}{(1 - \beta_2^t)} \tag{19}$$

$$w_t = w_{t-1} - \alpha \cdot \frac{\widehat{m_t}}{\sqrt{\widehat{v_t}} + \epsilon} \tag{18}$$

SGD [40] (see Subsection 2.3.2) algorithm is another optimization algorithm which is used for calculation of gradients and update of neural network weights. Gradient descent optimization algorithms are designed in such a way that they calculate the update parameters over the whole batch of training set, thus making the update parameters process slow. Hence, SGD was designed with the motivation behind to overcome the high cost of calculating the update parameters over the full training examples. SGD algorithm tends to randomly pick a single or few examples from the training set at each iteration to reduce the gradient computations enormously.

This computational method using mini-batches tries to speeds up the process compared to gradient descent methods and lead on a faster convergence. The mathematical formulation of the algorithm is shown in equation (20):

$$w = w_{t-1} - \alpha \cdot g_t$$
$$g_t = \nabla f(w_{t-1}, X^i, Y^i) \tag{20}$$

where: α – learning rate, w – updated weights, w_{t-1} – weights at previous timestep, g_t - gradient, t – timestep, X^i, Y^i- random sample from training set

Since the SGD algorithm is not designed to calculate the exact gradient descent of the object function, instead it estimates the gradient descent on a small batch which is also designated as local minima. In other words, it creates a stochastic process which is not always moving towards the optimal direction, because of the obtained noisy derivatives. Therefore, in order to make the weighted averages provide a better estimate and navigate efficiently towards the optima a technique named momentum is used. This technique helps on the improvement of both training speed and accuracy.

Appendix E. Reference Digital Elevation Model utilized by Google Earth Engine in order to perform terrain corrections.

SRTM 30 [69] is a joint international research of space agencies to produce a high-resolution digital elevation model of the Earth. The usage of dual radar antennas acquired radar data in interferometric mode, providing topographic data at 1 arc sec resolution or approximately 30 meters.

ASTER DEM [70] a global elevation model made available from NASA's Land Processes Distributed Active Archive Center. The first version was released on 2009 and the second version on 2011, the last release ASTER DEM v3 maintain the same tiles structure on GeoTIFF format as the previous two versions with 30-meter spatial resolution and 1° x 1° tiles. It covers land surfaces between 83° North and 83° South produced through an automatic processing of 2.3 million scenes of ASTER archives.

Appendix F. Open image databases for training Convolutional Neural Networks.

ImageNet developed by joint research of Stanford University and Princeton University by their researchers was first inspired by important needs in computer vision research. It is a large image dataset created to assist researchers to train their image classification and object detection models. The dataset spans 1000 object classes and consist of an amount of 14,197,122 images which are split as 1,281,167 training images, 50,000 validation images and 100,000 test images [55]

MS COCO dataset developed at Microsoft in 2014 is a large-scale object detection, segmentation, and captioning dataset. The dataset consists of 330 thousand images and with more then 200 thousand images labelled into 91 object categories with the 2014 release contains segmentation masks for 80 of these categories [56].

Pascal VOC was first introduced in 2005 by Everingham et al, is a dataset associated as benchmark in visual object category recognition and detection. With main challenges that run every year the stable version “VOC 2012” consist of 20 object classes. The training and validation datasets have 11,530 images containing 27,450 ROI annotated objects and 6,929 segmentations. [57]

Appendix G. An illustration of the structure of the algorithm performed for training of the model.

Training Algorithm

Require: model ← Established Faster R-CNN model
Require: training dataset ← Dataset with training samples
Require: validation dataset ← Dataset with validation samples
Require: optimizer ← Optimizer algorithm
Require: learning rate ← Learning rate scheduler, can be algorithm or constant factor
Require: device ← Parameter which decide where the model will run CPU or GPU
Require: epochs ← Total number of epochs on which the model will loop

```
for epoch in range(epochs):
    training(model, training dataset, optimizer, learning rate, device, epoch)
    validation(model, validation dataset, device)
    if best model:
        save model
```

Figure 45. Script illustration of the training and validation model

Appendix H. Coal Power Plant input bounding boxes of different size and predicted anchors and accuracy assessment.

As shown **Figure 46** with red color are labelled ground truth bounding boxes and with cyan color are bounding boxes of 1000m x 1000m size, blue color is labelled prediction bounding boxes of 500m x 500m

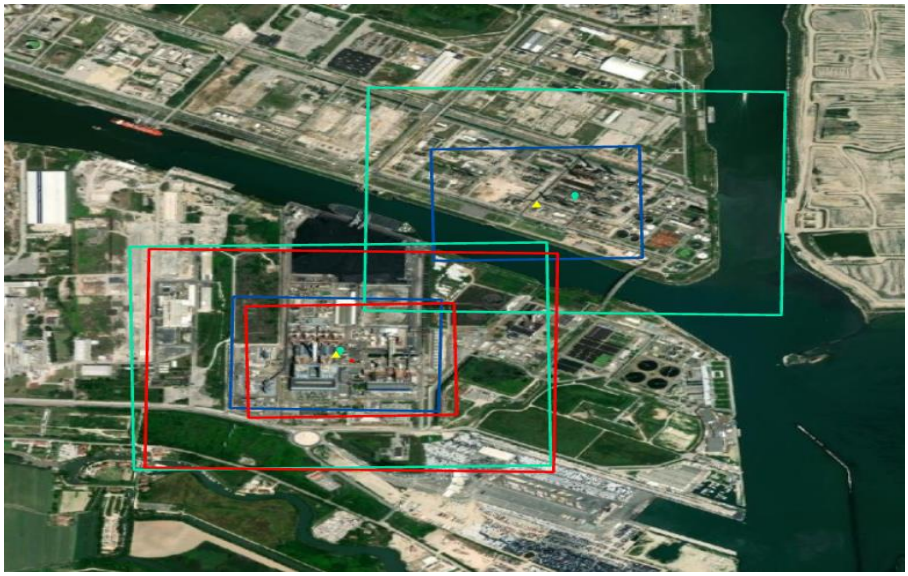


Figure 46. Coal Power Plant model predictions for different input bounding box size

In **Figure 47** with red color it is annotated the input ground truth box and with blue color the predicted anchor of the model. This can clearly show that the model predictions are better than the input ground truth at the places where inconsistencies on geo-location exist.

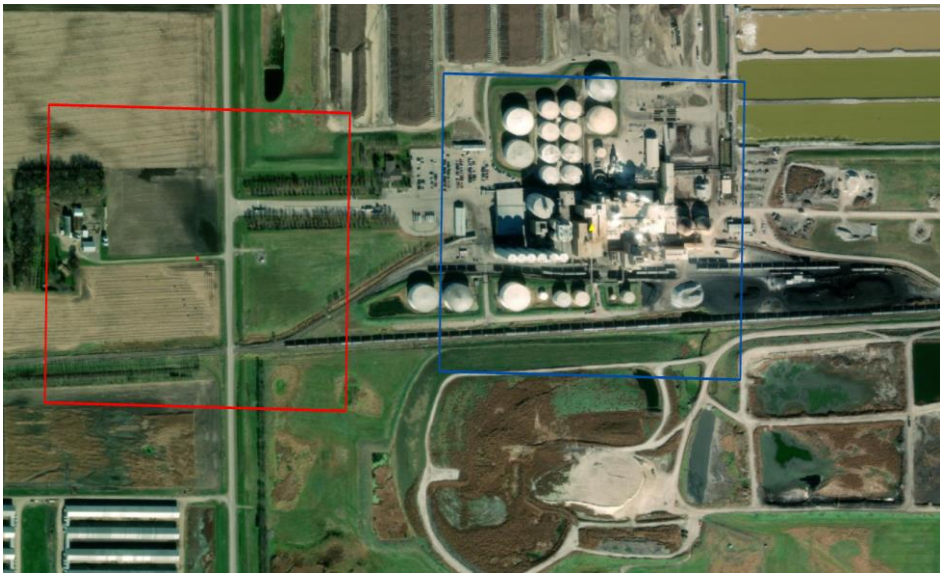


Figure 47. Illustration of Coal Power Plant ground truth misplacements

Appendix I. Wind Turbine Model artefacts during model evaluation on testing dataset.

The blue dot on images shown in **Figure 48** illustrate the artefacts which might appear on FP samples of Wind Turbine model, mainly these artefacts are shown when evaluating USWTDB where as stated in **Table 13** the number of FP samples in the USWTDB is quite high considering here that the geographical scope of this database is smaller in comparison to GPPD database where the FP samples are quite high because of different terrain conditions due to large geographical scope. The backscatter of SAR images in some particular areas is constructed very similar to the way wind turbines are build up, therefore the model is capable to detect wrong objects. Those artefacts are not shown in the ESRI Brandenburg test dataset, as the constructions similar to the examples below do not appear frequently.

Figure 48. Illustration of Wind Turbine model artefacts