Faculty of Science and Technology

BACHELOR'S THESIS

Study program/Specialization:	Spring semester 2022
Bachelor of Science in Computer Science	Open / Restricted Access
Author(s): Tommy Fivelstad, Arkadiy Stepanov	
Faculty Supervisor: Dr. Naeem Khademi	
Thesis Title: Space Archaeology: Survey and Implementation of Deep Learning Methods for Detecting Ancient Structures	
Credits (ECTS): 20	
Keywords: Fortress Tumuli/Burial Mounds Settlement Mounds / Synthetic Aperture Radar Remote Sensing Transfer Learning Deep Learning/ R-CNN / Object Detection	Pages: 101 Stavanger, 15th of May 2022

Contents

Contents **i**

Abstract **vi**

1 Introduction **1**

1.1 Remote Sensing in Archaeology 1

1.2 Thesis Introduction 2

1.3 Problem Definition and Research Questions 3

1.4 Use Cases 3

1.5 Challenges 4

1.6 Outline 5

1.7 Former Work 6

2 Background **7**

2.1 Technical and Theoretical Background 7

2.1.1 Fortresses, Tumuli/Burial Mounds and Archaeological Settlement Mounds 7

2.1.2 Satellite and Airborne Optical Imagery 10

2.1.3 LiDAR Digital Terrain Models and Relief Models . . 10

2.1.4 Synthetic Aperture Radar - SAR 13

2.1.5 Deep Learning 18

2.1.6 Neural Networks 19

2.1.7 Neural Networks Training 21

2.1.8 Object recognition 22

2.1.9 Convolutional Neural Network (CNN) 24

2.1.10 Performance Metrics 26

2.1.11 Geographical Information Systems 28

2.2 Related Work 28

2.2.1 Tumuli/Burial Mounds 29

2.2.2 Fortress Structures 33

2.2.3 Synthetic Aperture Radar uses in Archaeology . . . 34

3 Proposed Methods **39**

3.1 Burial Mounds - Norway 39

3.2 Fortress Detection - Central Asia 42

3.3 Settlement Mound Detection - Central Asia 43

4 Implementation **48**

4.1 Data Gathering and Image Preparation 48

4.1.1 Burial Mounds - Norway 48

4.1.2 Fortresses - Central Asia 52

4.1.3 Archaeological Settlement Mounds - Central Asia . . 52

4.2 Faster R-CNN Model Training 53

4.3 Final Application 54

5 Results and Discussion **55**

5.1 Results for Tumuli/Burial Mounds 55

5.2 Discussion of Results for Tumuli/Burial Mounds 63

5.3 Results for Fortresses 67

5.4 Discussion of Results for Fortresses 67

5.5 Results for Archaeological Settlement Mounds 68

5.6 Discussion of Results for Archaeological Settlement Mounds 73

6 Conclusion **75**

6.1 Conclusion and Answers to Research Questions 75

6.2 Improvements 77

References **82**

Appendix **82**

A Github Repository and Dataset **83**

B Areas and Data for Training and Testing of Models **84**

B.1 Norway Burial Mound Data 84

B.2 Central Asia Archaeological Mound SAR Data 86

C Multi-Spectral Satellite Imagery Uses in Archaeology **88**

C.1 Multi-Spectral Satellite Imagery Technical Information . . . 88

C.2 Use of Multi-Spectral Imagery in Archaeology 90

D Sources of Free Remote Sensing Data **93**

E Free Satellite Data Technical Specifications **97**

Abstract

Remote sensing instruments are changing the nature of archaeological work. No longer are archaeological discoveries done by field work alone. Light Detection and Ranging, or LiDAR, optical imagery and different types of satellite data are giving opportunities for archaeological discoveries in areas which might be inaccessible to archaeologists. Different types of machine learning and deep learning methods are also being applied to remote sensing data, which enables automatic searches to large scale areas for detection of archaeological remains.

In this thesis faster R-CNN object detection deep learning frameworks were used to train models and apply these to three types of archaeological remains. LiDAR based Digital Terrain Models were used to identify burial mounds in Norway. Optical imagery was used to identify fortress structures in Central Asia. Synthetic Aperture Radar data, or SAR, was used to detect archaeological settlement mounds in Central Asia. The success and limitations of these models are presented.

Acknowledgments

Thanks to Dr. Naeem Khademi for guidance and support during our work. This thesis gave us the opportunity to obtain knowledge in machine and deep learning, subjects not normally available during our education. It also gave us the opportunity to learn about remote sensing use in archaeology, giving us a very challenging and interesting thesis subject. We would also like to thank Kristian Løseth at the Norwegian Directorate of Cultural Heritage for his guidance on data acquisition and information about cultural heritage data in Norway.

Chapter 1

Introduction

1.1 Remote Sensing in Archaeology

Remote sensing instruments are becoming increasingly advanced. Light Detection and Ranging instruments, LiDAR, can scan large areas in short amounts of time from airplanes, helicopters and even drones giving the ability to create very high resolution digital terrain models which are not affected by forest canopy and other vegetation. Satellite Very High Resolution optical imagery is becoming vastly available with more spectral bands like near infra-red. Synthetic Aperture Radar satellite data using electromagnetic micro-waves can scan the earth independent of time of day, penetrate clouds, and dependent on frequencies and polarization, can even penetrate the ground if the ground conditions are right.[1]

Remote sensing instruments have given opportunities for archaeologists that was previously not available. It has become possible to do large-area studies and view the landscape in a way that is not possible to do within the time and financial constraints of regular archaeological field work. It also give archaeologists the ability study areas that are not accessible to them.[2][3] Remote sensing data can also be used to monitor known archaeological sites for potential risk and damage to these sites.[2]

Deep learning and machine learning techniques are also being applied to ar-

chaeological studies to automatically detect potential archaeological structures from remote sensing data.[2][4] Manually scanning remote sensing data over large-scale areas is time consuming. Deep learning techniques applied to this data can dramatically improve the time usage by giving archaeologists positive hits for archaeological structures which can be then manually examined.

A lot of recorded remote sensing data of interest for archaeological work is not publicly available, particularly high resolution data. There is still a large amount of data publicly available, giving anyone with the interest and capability the opportunity to search for undiscovered archaeological features.

1.2 Thesis Introduction

The first aim of this thesis is to document methods of using remote sensing data for mapping and discovery of archaeological structures and areas where these methods can be used. Also, the thesis will map the available free and open remote sensing data sources found. The second aim is to apply remote sensing data together with deep learning methods to detect selected archaeological structures.

For the thesis, three archaeological structures have been selected that are suitable for detection with faster R-CNN object detection deep learning methods. The first is ancient fortress structures in Central Asia. The second is tumuli/burial mounds in parts of Norway. The third is detection of settlement mounds in Central Asia. For fortress structures high resolution freely available Mapbox images, up to 0.47m per pixel, will be used.[5] For burial mounds high resolution, 0.25m per pixel, Digital Terrain Models, DTMs, will be used. The DTMs and raw LiDAR data and freely available through the Norwegian Mapping Authority.[6] For potential settlement mounds Synthetic Aperture Radar Sentinel-1 satellite data will be used, composited and created through Google Earth Engine.

1.3 Problem Definition and Research Questions

This thesis attempts to solve the following questions:

- Can ancient fortress structures effectively be detected using high resolution optical imagery using faster R-CNN object detection?
- Can partial ancient fortress structures effectively be detected?
- How effective is detection of burial mounds/tumuli using faster R-CNN object detection using high resolution LiDAR based Digital Terrain Models?
- Will Multi-Scale Relief Model treatment of LiDAR based DTMs improve the detection of burial mounds in Norway compared to other standard relief models? Will it improve the detection of smaller burial mounds?
- How will applying the faster R-CNN model on lower resolution LiDAR based DTMs affect the detection results?
- Can Synthetic Aperture Radar Sentinel-1 data, SAR, be used with faster R-CNN object detection to detect potential settlement mounds and what are the limitations?

1.4 Use Cases

The authors have found very little information of deep learning or machine learning methods being applied to find fortress structures. Should the thesis provide a method for effectively finding potential fortress structures, without a large amount of false positives, this could be a big help for the detection and mapping of similar fortress structures in areas where optical imagery is available.

Successful attempts have been made, using deep learning and machine learning methods, to map burial mounds/tumuli.[3] However, the information the authors have found on detection of burial mounds in Norway, do only

apply the simpler relief models to prepare DTM data used for learning and detection with deep learning methods.[4] Should more advanced relief models show an improvement in the detection of burial mounds in Norway, this could be applied in future work in areas where burial mound sizes are comparable with those of Norway.

Settlement mounds have been successfully identified by using SAR data and creating Digital Elevation Models, DEMs.[2] Settlement mounds have also been successfully identified using a combination of Sentinel-1 satellite SAR data and Sentinel-2 multi-spectral data and combining this with a Random Forrest machine learning algorithm.[7] No information has been found where deep learning and object detection have been used with SAR data to detect settlement mounds. If the SAR data alone can effectively detect settlement mounds, without the need for creating Digital Elevation models, this would simplify the process of detecting potential settlement mounds and enable large area scans without much additional work, apart from creating the SAR input data in Google Earth Engine.

1.5 Challenges

One of the main challenges for fortress detection is the fact that only optical imagery is available with a very good resolution for the areas in Central Asia, from free sources. No LiDAR data or high resolution satellite SAR or Multi-Spectral imagery data with high resolution has been found to be freely available. This limits the detection in areas where fortresses are not covered by vegetation and the identification of any height differences with other background. In optical imagery for example, these differences are only indicated by varying pixel colors. Additionally, the types of fortress structures vary greatly in shape and size. The challenge is to identify and distinct the structures from other non-fortress type structures.

For burial mound detection the main challenge is to differentiate burial mounds from natural mound like structure and other round man-made structures in the LiDAR based DTMs. Many of the burial mounds found in Norway are relatively small, giving them a less distinct signature than larger mounds. The main challenge is to apply methods to increase the effectiveness of detection and discover the effectiveness towards the false

positive metric. Another challenge is that the terrain the mounds are in can have a high effect on the number of false positives and how distinct the mounds signature is to the surrounding terrain.

Detection of settlement mounds using SAR data has several challenges. One challenge is the effect of type of terrain and soil of the areas of interest, which can affect the SAR signal response. Another is the size and height of the settlement mounds which affects the SAR signal response. Furthermore other types of archaeological remains, like some fortress remains, are likely to have a similar SAR signal response to settlement mounds. Another challenge is how effective SAR data is at detecting actual archaeological settlement mounds and separating them from other types of mounds. The latter problem is unlikely to be resolved using SAR data.

1.6 Outline

Chapter 2 - Background: This chapter presents technical and theoretical background information which is used in this thesis, to solve the thesis problem definition as stated in section 1.3. Relevant related work to the problems this thesis is attempting to solve is presented in section 2.2.

Chapter 3 - Proposed Method: This chapter explains the approach taken to apply deep learning to detecting the ancient archaeological structures as defined in section 1.3.

Chapter 4 - Implementation: This chapter explains the implementation of the deep learning methods, from acquisition of training data to implementation of the trained models in a final application.

Chapter 5 - Results and Discussion: This chapter presents the results acquired from the implementation of the deep learning methods to selected archaeological structures. Discussion of the results and evaluation of suitability for practical use. Limitations of the work and possible improvements are discussed.

Chapter 6 - Conclusion: This chapter gives the conclusion of the work done in the thesis and answers the question stated in section 1.3, in addition

to proposing improvements the work done.

Appendices: The appendices will contain Github repository information and information about the data and areas used for training and testing models. In addition the appendices will contain information regarding remote sensing instruments and satellite technical information, sources of free remote sensing data gathered and uses of remote sensing data in archaeology that is not related to the specific problems attempted to be solved in this thesis.

1.7 Former Work

This thesis is partly a continuation of the work previously done at the University of Stavanger. The previous work applied convolutional neural networks, CNN, to automatically detect Quanat water systems and also attempted to apply these same principles to detect ancient fortress structures in Iran and areas around Iran. This thesis will attempt to more effectively detect ancient fortress structures in Central Asia, using faster R-CNN object detection, in addition to applying these methods to detecting tumuli/burial mounds from LiDAR data in Norway and archaeological settlement mounds using Sentinel-1 SAR data.

Chapter 2

Background

2.1 Technical and Theoretical Background

2.1.1 Fortresses, Tumuli/Burial Mounds and Archaeological Settlement Mounds

In Norway burial mounds are common features in areas of bronze age to viking age settlements. They vary in size from as small as 2m up to 77m, which is the size of the Rakne mound, the largest burial mound in Scandinavia. In the bronze age it was common to build the burial mounds mainly from stone, without any use of sand or earth material. In the Roman era from 0-400 AD, larger mounds became more common, with sizes up to 40m in diameter and height up to 6-7m, while in the period between 550-800 AD the burial customs became simpler, and fewer burial mounds where built. That changed in the Viking era, after 800 AD, where larger mounds was again becoming more common. Typical grave mounds where built by adding rocks around the grave and adding earth and sand above the rocks, after the bronze age period, although the customs for building the mounds where varied, depending on region and period. Grave fields can often be found around settlement areas of the period they belong to, with the burial field in Vang in Oppdal county being one of the largest, containing around 900 individual grave mounds in a less than 1km2 area.

Figure 2.1: Oseberg Burial Mound

Fortresses are common archaeological features in many parts of Central Asia. Throughout the historical periods, fortresses where built in many areas as defensive structures at strategical places important for the time. The Great Wall of Gorgan in the Golestan province in Northern Iran, which was built between 420-530 AD, is one example of a defensive line which contains many fortress structures along it.[8] Fortresses vary in size and shape, from square or rectangular fortresses to round fortresses and other shapes which suited the terrain upon which they were built. Today, a lot of the ancient fortress structures are only partial remains of the original structure. Some fortresses might have only partial remains of the walls. Other fortresses might only have an outline of its original structure remaining visible. The size of fortresses ranges can range from as small as a few meters to several kilometers in size.

Figure 2.2: Fortress Structure

Archaeological settlement mounds are common throughout the Middle East and Central Asia. They are mounds created over large time periods, perhaps thousands of years, due to deposition of anthropogenic materials, like broken mud bricks and other building material, in periods of settlement. The height of the settlement mounds vary a lot in size and height. Mounds above 40m in height have been registered.

Figure 2.3: Tell Uqair Settlement Mound in Irak

2.1.2 Satellite and Airborne Optical Imagery

Satellite and airborne optical imagery is widely available through many sources, including Google Earth and MapBox.[5] These services often uses a combination of satellite and airborne imagery sources for their global optical imagery data. Optical imagery satellites provides the bands visible to the human eye, red, green and blue. In addition, many of the satellites providing optical satellite imagery has additional bands, including near-infrared and panchromatic. These satellites are often termed as multi-spectral satellites. A panchromatic image is a single band image that combines the red, green and blue bands allowing for greater spatial resolution.[9] Corona images, available through the US Geographical Survey, is an example of Panchromatic imagery. Panchromatic bands can be used to enhance the lower resolution spectral bands of multi-spectral satellite data.

For this thesis the MapBox service was used to obtain optical images for the Central Asia region, as it is a free service, providing a resolution as low as 0.46m per pixel at zoom level 17.

2.1.3 LiDAR Digital Terrain Models and Relief Models

LiDAR, or Light Detection and Ranging, is a remote sensing method that uses light, or pulsed laser, to measure distances to the earth surface. The sensor typically consists of a laser, a scanner and a GPS receiver.[10] There are two types of LiDAR, topographic and bathymetric. The topographic used near-infrared laser to map land, while the bathymetric uses green light which can penetrate water, to measure seabeds and riverbeds. LiDAR data can typically be gathered from helicopters, airplanes and even drones.

The LiDAR data is presented as point cloud data where each measurement has 3D spatial coordinates - longitude, latitude and height. Point clouds can contain many different returns, and the returns gets classified based on the intensity of the returns.

The LiDAR data can be used to create Digital Elevation Models and Digital Terrain Models, DEM and DTM. By using the last returns or returns classified as ground returns from the LiDAR sensor, it is possible to measure

the earths surface, removing vegetation.

Relief models are 3D representations of part of the earths topography.

A Simplified Local Relief Model, SLRM, is a common visualisation technique used for LiDAR Digital Terrain Models. The technique uses a low-pass filter on the DTM, with a fixed size area around each data point, to create a smoothed averaged DTM. The SLRM model is created by subtracting this smoothed DTM from the original DTM, creating a height difference from average visualisation.

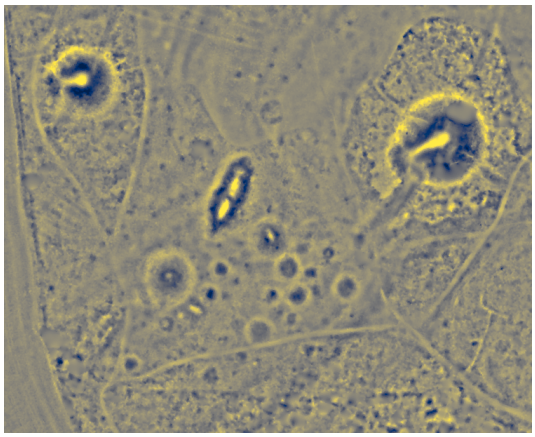


Figure 2.4: Simple Local Relief Model representation of LiDAR based Digital Terrain Model

A Slope model is another common visualisation technique used for LiDAR DTMs. It calculates the slope values, for example in degrees, between each LiDAR DTM pixel.

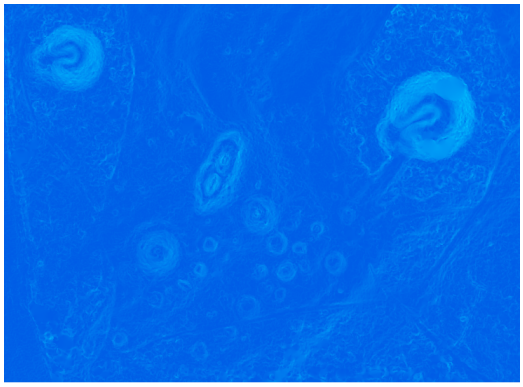


Figure 2.5: Slope representation of LiDAR based Digital Terrain Model

A Multi-Scale Relief Model, MSRM, is a visualization technique which aims to create a micro-relief to multi-scale features.[11] The technique applies several low-pass filters of different area sizes on the original DTM. The difference between low pass filtered DTMs are summed together and then averaged by the number of differences calculated, as defined by formula 2.1

$$MSRM = \frac{\sum_{i}^{n-1} LPr(i^x) - LPr((i+1)^x)}{n-i} \tag{2.1}$$

x is the scaling factor applied, giving the spacing between the radii of the filtering applied to the surface, starting at i^x.

The i and n values are calculated from equation 2.2 and 2.3.

$$i = \lfloor((f_{min} - rr)/2rr)^{\frac{1}{x}}\rfloor \tag{2.2}$$

$$n = \lceil((f_{max} - rr)/2rr)^{\frac{1}{x}}\rceil \tag{2.3}$$

f_{min} is the minimum feature size of interest, while f_{max} is the maximum feature size of interest. rr is the resolution of the DTM.

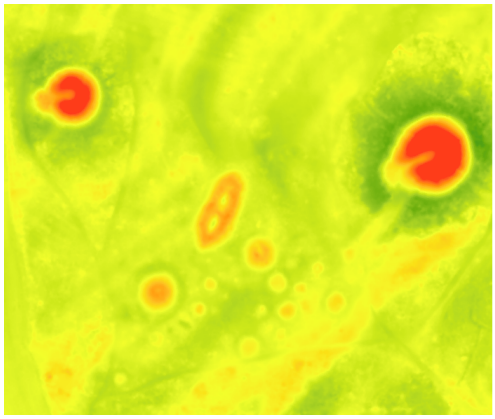


Figure 2.6: Multi-Scale Relief Model representation of LiDAR based Digital Terrain Model

None of the relief models explained here are actual representations of height in the original DTM, but methods of visualizing variation in height.

For this thesis DTM data freely available through the Norwegian Mapping Authority is used.[6] Point cloud LiDAR data is provided together with 1m resolution DTMs covering most of Norway, called the National Height Model. In addition, several higher resolution DTMs produced from different projects are being offered. The DTM data used has a spatial 0.25m resolution. 1m spatial resolution data is also used for testing faster R-CNN models, to detect the difference in results compared to the 0.25m spatial resolution data.

2.1.4 Synthetic Aperture Radar - SAR

Synthetic Aperture Radar, SAR, is a concept in which a sequence of acquisitions from a short antenna are combined to simulate the single acquisition of a larger antenna. For example, for a wavelength of 5cm a single acquisition from a satellite orbit could need as much as a 4250m long antenna to get a 10m resolution.[12].

Unlike optical satellite measurements, the SAR sensor have the ability to acquire data independent of time of day and cloud cover. The typical bands,

frequencies and wavelengths of SAR instruments are listed in table 2.1

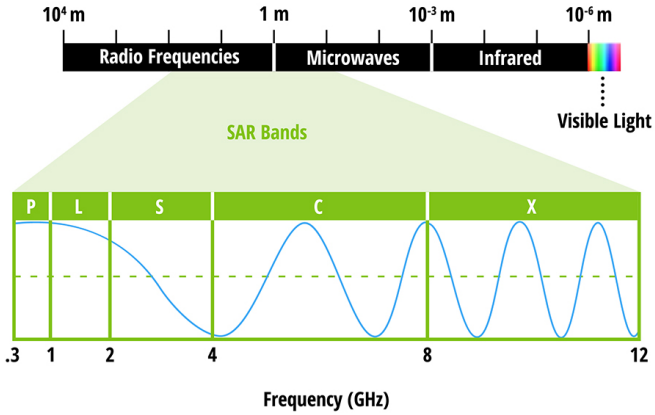


Figure 2.7: SAR Bands. Credit: NASA SAR Handbook

The electromagnetic waves that SAR instruments propagate have to some degree the ability to penetrate vegetation cover, dependent on the wavelength. The longer the wavelength, the better the ability to penetrate the vegetation cover. In dry environments, where the soil has a low dielectric constant, the electromagnetic wave can also penetrate the soil. Hence, SAR satellite data can have archaeological applications in certain environments like sand deserts and arid areas, and could detect archaeological features or structures that are buried or partially buried.

The polarization of the SAR acquisitions refers to the plane of which orientation of plane in which the electromagnetic wave oscillates. SAR satellites typically transmits linearly polarized.[12]. Depending on the satellite and data processing, one can get 4 main types of polarized acquisitions, VV, VH, HV and HH, where V stands for vertical and H stands for horizontal. The first letter represents the transmission of the electromagnetic wave and the last letter represents the reading of the wave return. Scattering of the electromagnetic wave when it hits a surface will be more visible in certain acquisition modes, dependant on the type of scattering present. Rough surface scattering, or relief scattering, will be more visible in VV acquisition mode. Volume scattering, like scattering from forest cover, will be more

visible in VH and HV acquisition modes. Double bounce scattering, which can be caused by buildings, walls or tree trunks for example, is more visible in HH acquisition mode.

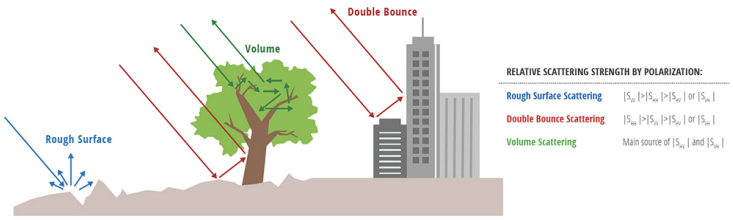


Figure 2.8: Examples of SAR scattering. Credit: NASA SAR Handbook

One major use of SAR data is to perform interferometry, or InSAR. Interferometry uses two acquisitions of the same area and combines them to produce an interferogram, which is a function of the phase difference of the returning waves for the acquisitions. The interferogram can be used to measure very accurately deformation of the surface, and typical applications are to measure deformation from earthquakes, volcanic eruptions and changes in ice sheets. Interferometry can also be used to create Digital Elevation Maps, with better spatial resolution than the freely available 30m spatial resolution STRM DEM created by NASA.[2] The Sentinel-1 SAR data has the best spatial resolution of the freely available SAR data with global coverage. However, even though it is possible to create DEMs from the data, care should be taken when producing these, as the quality is not necessarily better than the STRM DEM, due to too high time difference between acquisition or lack of proper distance between the two acquisitions required to produce the DEMs.[13]

For this thesis Sentinel-1 SAR data was used and was generated from Google Earth Engine, GEE. SAR acquisitions from October 2014 to June 2020 was composited and averaged for each area. The spatial resolution of the SAR data is 10m. Google Earth Engine provides a Level-1 Ground Range Detected product, GRD, processed to back-scatter coefficient in decibels.[14]. The data is acquired in Interferometric Wide mode.

The following pre-processing steps are used on the GEE SAR data:

- Apply orbit file
- GRD border noise removal
- Thermal noise removal
- Radiometric calibration
- Terrain correction

Table 2.1: NASA SAR Band Information[12]

Band	Frequency(s)	Wavelength	Typical Application
Ka	27-40 GHz	1.1-0.8 cm	Rarely used for SAR
K	18-27 GHz	1.7-1.1 cm	Rarely used for SAR
Ku	12-18 GHz	2.4-17 cm	Rarely used for SAR
X	8-12 GHz	3.8-2.4 cm	High Resolution SAR. Often used for urban monitoring. Has little penetration into vegetation cover.
C	4-8 GHz	7.5-3.8 cm	Most common SAR band used. Used for global mapping, like ice and land monitoring. Low to moderate vegetation penetration
S	2-4 GHz	15-7.5 cm	Increasing use for applications like global agriculture monitoring. Expands C-band applications to higher density vegetation cover.
L	1-2 GHz	30-15 cm	Medium resolution SAR. Used for geophysical monitoring, biomass and vegetation mapping. High vegetation penetration.
P	0.3-1 GHz	100-30 cm	Biomass mapping. Highest vegetation penetration.

2.1.5 Deep Learning

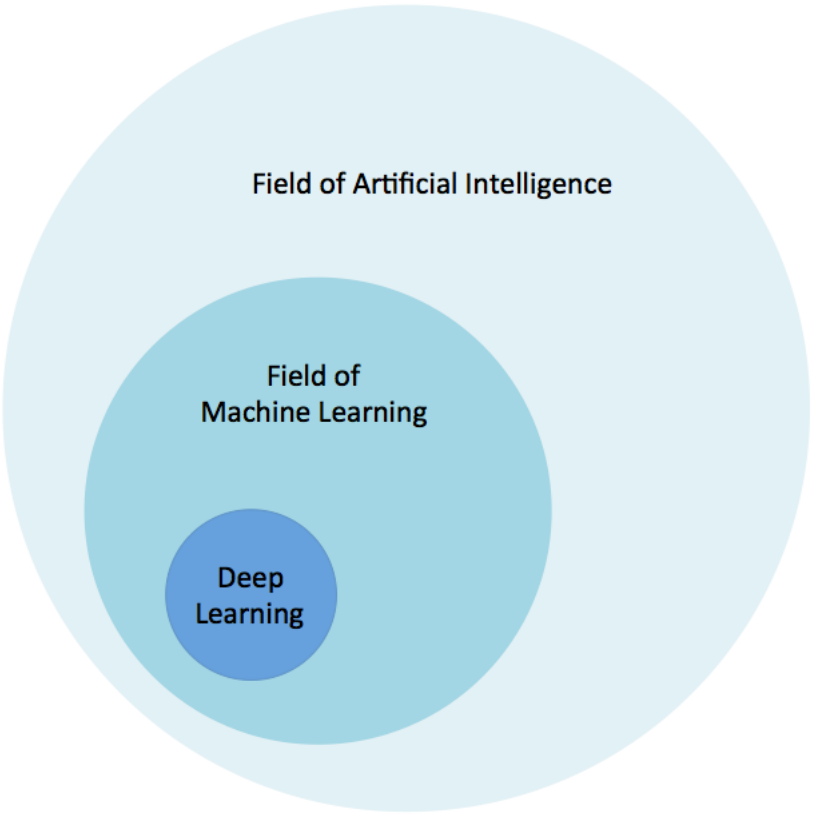


Figure 2.9: Artificial Intelligence - Machine Learning - Deep Learning relation

Artificial intelligence is a broad term referring to advanced machine intelligence. In 1956, at a conference on artificial intelligence in Dartmouth, this technology was described as follows: "Every aspect of learning or any other feature of intelligence can be so precisely described that a machine can be made to simulate it."[15] Machine learning is one of the branches of Artificial Intelligence. Basic principle of ML is the improvement of existing algorithms through various statistical methods by analysing a set of given data. Unlike programs which perform specific tasks with help of hard-coded instructions, machine learning allows the system to learn to recognize patterns and make predictions on its own.

Deep learning is a subset of machine learning and is currently becoming one of the most popular approaches. Deep neural networks show better results than alternative methods in areas such as speech recognition, processing natural language and computer vision.[16] One of the reasons for the successful application of deep neural networks is that the network automatically extracts important features from the data that are necessary to solve the problem. In alternative machine learning algorithms, features should be distinguished by people, there is a specialized area of research for that which is called feature engineering. However, when processing large amounts of data, the neural network handles with selection of features much better than a person. On the other hand, deep learning can become a very expensive operation and requires a relatively big amount of processing data for training.

2.1.6 Neural Networks

The model of artificial neural network was firstly introduced by McCulloch and Pitts in 1943. As a base for their work the authors used biological neurons.[17]. The neural network has N binary entry values $x_1, ..., x_n$ which are being summed with multiplication of weights $w_1, ..., w_n$

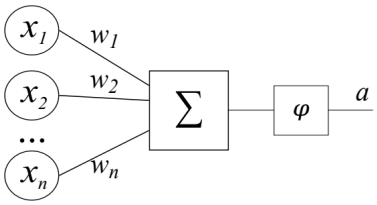


Figure 2.10: Neural network by McCulloch and Pitts

The output neural signal is defined by the function:

$$a = \varphi(\sum_{i=1}^{N} w_i * x_i) \tag{2.4}$$

where φ is the non-linear activation function. Initially, in their neural func-

tion as activation function was used the Heaviside step function 2.5

$$f(x) = \begin{cases} 1 & x > 0 \\ 0 & x \leq 0 \end{cases} \tag{2.5}$$

In the following years however it has been recommended to use other activation functions such as sigmoid function 2.6, hyperbolic tangent 2.7 and Rectified linear unit (ReLU) 2.8 [18]

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{2.6}$$

$$\tanh(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \tag{2.7}$$

$$f(x) = \begin{cases} x & x > 0 \\ 0 & x \leq 0 \end{cases} = \max\{0, x\} \tag{2.8}$$

Deep neural networks consist of multiple neural layers. A neural network receives an input, calculates the resulted value and passes it along to the next layer and so on until calculated value has reached to the final output. There does not exist a clear definition of how many layers a network must consist to be considered "deep", but a minimum requirement would be that there is at least one input/output layer and a hidden layer.

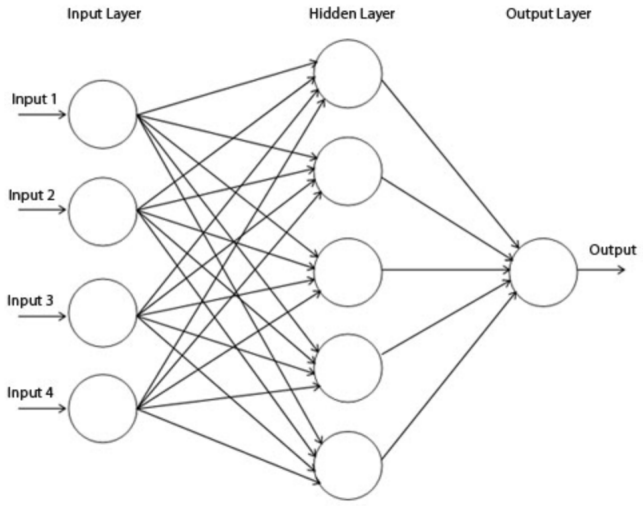


Figure 2.11: Deep neural network with one hidden layer

2.1.7 Neural Networks Training

Training of a neural network is the process of weight determination in a way that will enhance the accuracy of the function. Three approaches exist for training [19]:

- Supervised learning
- Unsupervised learning
- Reinforcement learning

During supervised learning the network receives sets of labeled input data with the correct predictions. During the training process the weights shift depending if the network's prediction coincides with the received labels. In the case of unsupervised learning the training happens without the labels. Reinforcement learning involves the existence of an external environment with which the network interacts. Learning takes place on the basis of

signals received from this environment. Those signals can be of form reward-punishment. The model starts to learn through sequential trial and error. McCulloch-Pitts neural network were not trainable. Weights for all neuron inputs had to be pre-configured beforehand.

Nowadays, error backpropagation algorithm is used for training of neural networks. The algorithm is based on the gradient decent method. The algorithm uses supervised learning which requires a labeled set as an input. During the training process in order to adjust correctly the weights of the neurons the results from the network have to be evaluated. A loss function calculates an error measurement which determines how much the output values of the network differ from the correct answers. The error value is then minimized using the gradient descent method by changes in weight values in the network. In order to assess how strongly each weight affects the output value, the partial derivatives of the error with respect to the weights are calculated. Then the weights are being adjusted with the help of the gradient and the process repeats until the output error is reduced to an acceptable value. The initial values of the weights in the network are set randomly.

2.1.8 Object recognition

Deep learning is gradually gaining popularity in object recognition. The capability of a machine identifying various objects on images or videos. Object detection tasks can be grouped into three categories.

Object classification

In object classification the neural network receives an image as an input and gives us output a list with pairs of classes and their probability scores (a percentage metric for a class to be accurate). Classification tasks can be either binary or multi-class.

Object detection

In object detection the model must identify the position and classification of the object on the image and correctly classify it. During the training process the model receives as input a list of images and coordinates which are used for locating the object of interest. During the evaluation step of an image a list of labeled bounding boxes is returned. The efficiency of the model in object detection tasks tends to be better than in classification tasks. The provided bounding boxes assist the model to identify the region of interest as opposed to image classification's training process. In the current thesis this methodology has been mainly used.

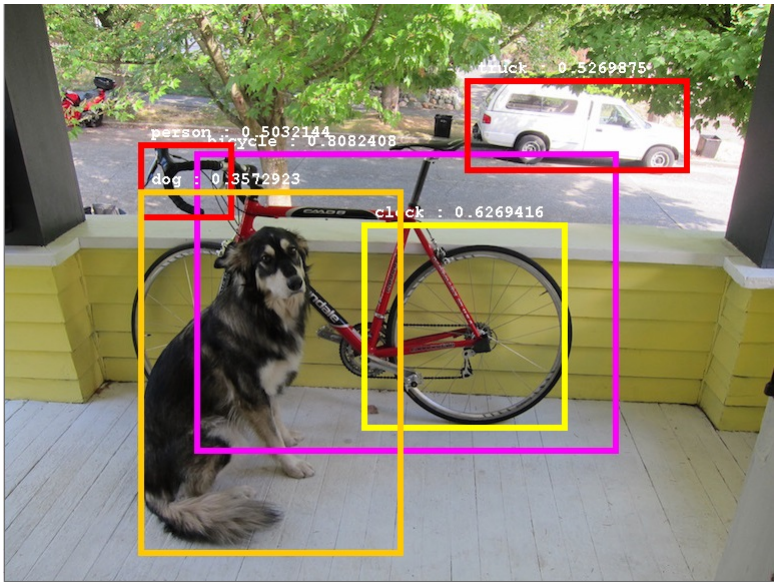


Figure 2.12: Object detection

Object segmentation

In object segmentation the model has to predict the exact pixel position of the object. Masks with pixel-wise indicators and images are being provided as input during the training process. Similar masks are given as output upon evaluation. In this case as well, the masks assist the model identifying precisely the position of the object by providing the ability to precisely to

the pixel filter target from noise.

Figure 2.13: Object segmentation

2.1.9 Convolutional Neural Network (CNN)

In figure 2.11 is shown a neural network with a property called "fully connected". This means that each and every single neuron is directly connected to all previous and following neurons. Due to the high amount of incoming parameters however, computations may become expensive and also inefficient.

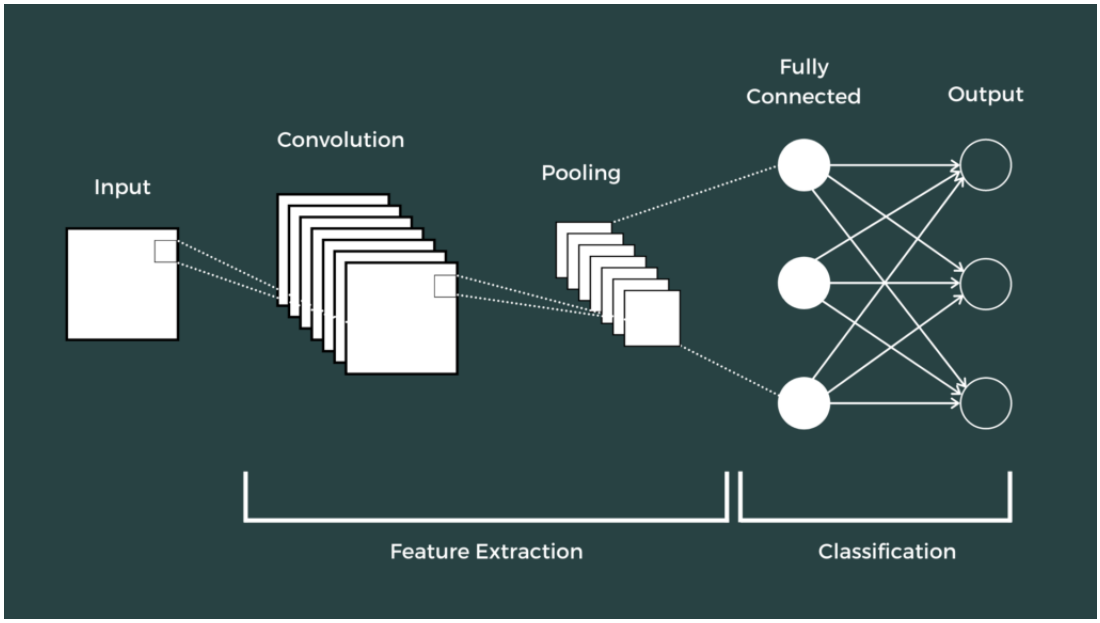


Figure 2.14: Convolutional Neural Network

CNN is an alternative architectural approach. In the early 1960s, Hubel and Wiesel studied the functionality of cat's visual cortex. One of the purpose of the experiment's was to detect neuron activity by showing different images to the cat. They found out that different neurons were responding differently, depending on the angle and placement of a line on the screen.2.15 The conclusion was that the virtual cortex has different levels of neural cells which are being activated based on the region of interest of each neuron. That means that there are simple neural cells that are being activated to individual indicators, such as angle or placement of the line, but there are also complex neural cells that are being activated based on a combination of different indicators.[20] From this was inspired the proposed CNN called neocognitron by Fukushima in 1980. They used the analogy of simple and complex cells where the former corresponds to convolution layers and the latter to sub sampling layers.[21]

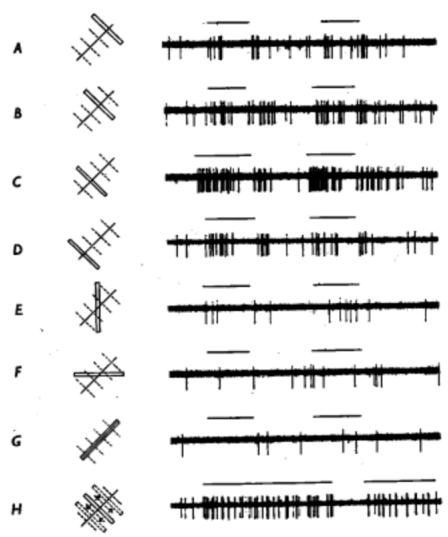


Figure 2.15: Neuron activations based on position and angle

Convolution Layers

In image processing the convolution is an operation of filtering on the image to extract various characteristics. During the process new values for each pixel of the image are being calculated taking in consideration the surrounding pixels. This calculation is done through multiplication of the pixel and the surrounding ones with a n x n matrix which often is called kernel. The values in the kernel vary and depend on the desired filtering.

2.1.10 Performance Metrics

In order to be able somehow to compare two neural networks, one must initially evaluate them. Different evaluation methods exist for different tasks and most of the times one metric is simply not enough. The classification associated metrics are being calculated from four variables.

- True positives - successful prediction of a positive classification

- False positives - unsuccessful prediction of a positive classification
- True negatives - successful prediction of a negative classification
- False negatives - unsuccessful prediction of a negative classification

Precision - it is an important metric with which it is possible to evaluate the accuracy of the neural network. However in some cases precision by itself can be a poor evaluator of the efficiency of the trained neural model. One of those cases is when the training dataset has a bad class distribution. In those cases precision should be measured for each class individually.

$$Precision = \frac{TruePositive}{TruePositive + FalsePositive} \tag{2.9}$$

Recall is another important metric. It is used to see the effectiveness of the neural network to find all the objects. In other words, the percentage of the successful predictions of positive classifications.

$$Recall = \frac{TruePositive}{TruePositive + FalseNegative} \tag{2.10}$$

In some cases, depending on the task, recall and precision can be sufficient metrics on their own. Sometimes however those metrics are equally important and one combined value is easier to argue about. F1 metric combines precision and recall and provides a meaningful value for performance measurement.

$$Precision = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{2.11}$$

It is worth mentioning that sometimes precision and recall can have a different weight on the desired result and that weight can be adjusted by adjusting the value β. In 2.11 the β values equals to one which indicates that precision and recall are equally important.

$$F1 = (1 + \beta^2) * \frac{Precision * Recall}{\beta^2 * Precision + Recall} \tag{2.12}$$

2.1.11 Geographical Information Systems

A Geographical Information System, GIS, can be defined as:

> A computer system for storing, checking and displaying data related to positions on the Earth's surface.[22]

There are several GIS software packages available which can be used to process LiDAR data, do work on DTM data, create relief models DTM data, work with satellite data and has a large range of other functionality.

Probably the most popular GIS software package is ArcGIS Pro. ArcGIS Pro is however not a free software. QGIS is another GIS software package which is free and open source, and supports scripting of tasks through Python.

Google Earth Engine[23] is an online service provided for free by Google. It provides GIS functionality and Javascript/Python scripting for task automation. A large range of free satellite data with imagery and other data, including Landsat, Sentinel and MODIS satellite data.

For this thesis, the ArcGIS Pro software package was used to work on LiDAR DTM data and SAR data.

2.2 Related Work

Many approaches are used to detect archaeological structures and features using remote sensing instruments. Machine learning and deep learning methods are also being used together with remote sensing data to automate detection of these structures and features.

As part of this thesis an overview of freely available remote sensing data, useful resources, technical specifications of common free satellite data and an overview of the use of the data in archaeology is presented. In this section is presented work related to the detection of tumuli/burial mounds, ancient fortress structures and archaeological uses for Synthetic Aperture Radar. The rest of the information is presented in the appendices.

2.2.1 Tumuli/Burial Mounds

A lot of work has been done on the detection and mapping of tumuli/burial mounds, using different AI techniques. Detection of burial mounds is mostly applicable to Digital Terrain Models created from LiDAR data.

Ø.D.Trier Et al.[4] used mostly 5 points per m2 LiDAR data to create DTMs used in detection of burial mounds, charcoal kilns and deer hunting traps. They used faster R-CNN with a VGG16 pre-trained deep neural network. The DTMs were pre-processed using a Simple Local Relief model, before being used for training, validation and testing. In a larger testing area were correctly identified 38% of burial mounds with a false positive rate was 86%. For charcoal kilns in a different vast area results were improved. 90% of the charcoal kilns were correctly identified with the false positive rate at 38%.

I. Berganzo-Besga Et al.[3] used a hybrid approach to the detection of tumuli in north-western Iberia. Their approach used 1m spatial resolution DTMs. The DTMs were pre-processed using both a Simple Local Relief Model, Slope model and a Multi-Scale Relief model. Yolo v3 was used with a Darknet-53 backbone. The average precision(AP) for the Slope model and Simple Local Relief model was approximately at 53%. The Multi-Scale Relief Model, taking into account the multi-scale nature of mounds increased the AP to about 63%. The Multi-Scale Relief Model is a visualisation technique which can be applied to any type of DTM data. See section 2.6. Furthermore, Sentinel-2 multi-spectral data was used to reduce the number of false positives. With Google Earth Engine 10-band composites were created from the Sentinel-2 data. Areas in this composite were marked that were impossible to contain mounds, like urban areas. This was used for training of a Random Forrest algorithm which excluded areas of the DTMs that should not have mounds, hence reducing the number of false positives.

W.B. Verschoof-van der Vaart Et al.[24] introduce a workflow for multi-object detection, named WODAN, which consists of three steps, pre-processing, object detection and post-processing. Burial mounds, celtic fields and charcoal kilns were the target experiment objects. The team pre-processed 0.5m spatial resolution DTM data using a Simple Local Relief Model. The object detection is faster R-CNN with a VGG16 pre-trained model. The post-processing part consisted of conversion from bounding box conversion for found objects to real-world coordinates which can be used with GIS software. They performed several experiments with different parameters, giving variable results. Overall, they conclude that faster R-CNN is one of the top performers for detection of mounds, when compared to certain other techniques used in the past. However, they did not have any success at detecting charcoal kilns, which they attribute to the lack of training data in their data set.

W.B. Verschoof-van der Vaart Et al.[25] also introduces a workflow called WODAN 2.0. The new step in the WODAN 2.0 workflow compared to their approach in the WODAN workflow, is the introduction of a location-based ranking system to help exclude false positives. This step makes the assumption that the location of archaeological features in the current terrain is not random. Effects like current land-use, erosion, drift-sand areas and other influences affects the likelihood of archaeological features being present. By taking these effects into account in a ranking system, they created a map of their test area, which categorised areas of the map between 1 and 3, 1 being the highest likelihood for finding a certain archaeological feature. By only including the highest ranking areas, they could exclude areas which likely had a higher rate of false positives.

A. Guyot Et al.[26] used a multi-scale relief approach together with a Random Forrest machine learning algorithm to detect burial mounds in parts the Bay of Quiberon and the Gulf of Morbihan in Frace. They processed high resolution LiDAR data into a 0.25m spatial resolution DTM. They present an approach called Multi-Scale Topographic Position, MSTP, which is inspired by the Integral Image algorithm, to pre-process the DTM data. They create three intermediate DTM rasters, which they call micro, meso and macro, each calculating the topographic deviation at that scale. Their scales were 1-10m, 10-100m and 100-100m, which corresponds to micro, meso and macro respectively. The mound sizes of interest in their work was between 10-100m. The micro, meso and macro rasters were combined into

a final composite raster. The MSTP composite raster was used to train a 120 tree Random Forrest algorithm. This algorithm was used to create a probability map giving the likelihood of the presence of burial mounds. Their results were good, predicting all of the known existing megalithic burial mounds in their testing area.

T.Freeland Et al.[27] used two automated feature extraction machine learning techniques from LiDAR based DTM data to detect burial mounds on the Tongatapu island in Tonga. The DTMs had a spatial resolution of 1m. The first feature extraction technique they used was object-based image analysis and the second was an inverted mound algorithm that was produced for the project, termed iMound, which uses a combination of existing visualisation techniques to create a map of mound heights. Their techniques yielded good results, with the iMound technique giving slightly better results, with a F1-score of 0.84, detecting over 10000 mounds. Their own conclusions state that the methods are "coarse", and better at identifying distribution of mounds rather than identifying all individual mounds.

Table 2.2: Overview of articles on tumuli/burial mound related work

Reference	Sensor Type(s)	Type of ML/DL	Description
[4]	LiDAR	Faster R-CNN DL	Mapping and detection of burial mounds, charcoal kilns and hunting traps in Norway.
[3]	LiDAR and Sentinel-2 Multi-spectral imagery	Yolo v3 DL and Random Forrest classifier	Mapping and detection of burial mounds in Northen Iberia using Multi-Scale Relief Model and Sentinel-2 data to exclude areas which should have no mounds.
[24]	LiDAR	Faster R-CNN DL	Mapping and detection of burial mounds, celtic fields and charcoal kilns in the Netherlands
[27]	LiDAR	Automated Feature Extraction ML	Detection of monumental earthworks/burial mounds in Tongatapu, Tonga
[26]	LiDAR	Random Forrest classifier	Detection of Neolithic burial mounds in France using multi-scale approach.
[25]	LiDAR	Faster R-CNN	Detection of burial mounds, charcoal kilns and celtic fields in the Netherlands, applying a location-based ranking system to predict likelihood of archaeological feature being present.
[11]	N/A.	N/A	Multi-Scale Relief Model. An algorithm for creating a relief model for features of multi-scale nature for any type of Digital Elevation Models and Digital Terrain Models.

2.2.2 Fortress Structures

The authors have found no information of previous work done with deep learning for detecting fortress structures. Only one article describing finding potential viking ring fortress sites with machine learning have been found. Table 2.3 shows an overview of relevant material found.

A. Monterroso-Checa Et al.[28] used LiDAR data to analyze the El Viandar Castle remains and the surrounding settlement, in the Cordoba region of Spain. The castle is placed on a hilltop and the castle and settlement ruins are in large parts covered by forest. They scanned the area with LiDAR creating 50 points per m2 point clouds, which were converted to DTMs. The data was analysed using slope and shadow maps visualization techniques in the QGIS software package. With the very high resolution data they were able to determine the castles exact location , main structures and dimensions, which would not be possible with much lower resolution LiDAR data.

D. Stott Et al.[29] used LiDAR data from the danish national airborne laser scanning project to identify possible candidate sites for viking ring fortresses in Denmark. Due to the low number of existing ring fortress structures, deep learning techniques were not viable for their task. They pre-processed the DTM data with residual relief at the scale of the expected ring fortress sizes. For identifying possible ring structures in the data, they used both template matching and a Hough circle transform. This gave them over 200000 circular features. Using several different criteria they could reduce this number, for example by excluding any circular feature with a diameter less than 45m and by comparing the maximum and minimum diameter, which would be the expected rampart width of the fortress. In addition they employed mapping data of Denmark before any large scale land reclamation and drainage altered the the environment, adding features like main communication routes and other information from the Viking era of likely terrain in which fortresses could be built. The information was used to train a Random Forest classifier that reduced the original number of circular features found. By using these techniques they identified 199 circular features that could be candidates for viking ring fortresses and reduced these down to 2 two likely candidates.

Table 2.3: Overview of articles on fortress structure detection

Reference	Sensor Type(s)	Type of ML/DL	Description
[28]	LiDAR	N/A	Detection of remains of El Viandar Castle in El Hoyo, Spain. High resolution LiDAR data was acquired by drone to detect remains of the castle on a forest covered hill.
[29]	LiDAR	Hough Circle Transform, template matching and Random Forest classifier	Use of machine learning techniques to search areas for potential sites for viking ring fortresses.

2.2.3 Synthetic Aperture Radar uses in Archaeology

D. Tapete Et al. used SAR data acquired by the italian Cosmo-SkyMed satellite to detect and monitor archaeological mounds.[2] So-called Tells are archaeological mounds generated deposition of anthropogenic material and sediments over large time periods. Tells are common archaeological features in the Middle East. They used the SAR data to create Digital Elevation Models, DEMs, for the governate of Wasit in Iraq. The DEMs are created by combining two acquisitions of the same area, using interferometry. They used the StripMap HH-polarized acquisition images with a 3m spatial resolution, creating a final DEM with a 10m spatial resolution. This spatial resolution is significantly better than globally available SRTM DEMs, which has a 30m resolution. They effectively mapped the Tells in the research area based on knowledge of their presence from other data sources, like Corona imagery, Sentinel-2 data and Soviet maps from the 1950s. 12% of the total tells found were however not possible to detect from the other sources of comparison. Tells standing out as low as 4m above the background terrain could be detected. In addition to mapping Tells, the project also used 1m resolution Enhanced Spotlight mode acquisitions to create time-series DEMs, providing a method to monitor damage to the Tells, indicating looting.. The survey was mainly performed to demonstrate the capabilities of the Cosmo-SkyMed SAR data, and not specifically as an archaeological study, but demonstrates how high resolution SAR data can be an effective

tool for archaeological monitoring and mapping.

C. Stewart Et al. used L-band SAR data from the PALSAR-1 and 2 satellites to find anthropogenic features in the North Sinai desert.[30] The study created multi-temporal filtered sigma nought back-scatter images of the study area from the SAR data. The sigma nought is also called a back-scatter coefficient, a normalized parameter indicating the strength of back-scatter.[31] Of two regions studied, one contained a large amount of linear features of interest defined by strong back-scatter, while the second region a low back-scatter anomaly of interest, near three well known archaeological sites. To extract linear features from the first area of interest an algorithm was implemented to extract the man made linear features from the back-scatter images. Standard linear feature extraction machine learning methods were not found to be good enough. The SRTM DEM was used to calculate an incidence angles for the sand dunes, and hence excluding these features. By experimentation the use of average coherence between the consecutive SAR acquisitions was found to be better at distinguishing the linear features, as the stood out more compared the the background with lower coherence. By comparing the resulting man made linear features to known existing features, the study managed to find partially buried roads and infrastructure not visible in optical imagery or other non-SAR data. For the second area of interest, the study created time-series imagery from the PALSAR-2 satellite, with a 2m spatial resolution. By looking at the De Grandi filtered intensities of the SAR attenuation from both PALSAR-1 and 2 data in different variants, the low back-scatter anomalies became more visible. By studying the area with previous archaeological work the study gave an interpretation that the anomalies might be areas of settlement that might have been surrounded by lagoons, rivers and swampland in the past, and the difference in soil moisture might be the reason for the anomalies.

D. Tapete Et al. used ENVISAT ASAR data for detection and monitoring of archaeological heritage in the Nasca region in southern Peru.[32]. The ENVISAT ASAR satellite has a 30m spatial resolution and is a C-band SAR satellite with a 5.33 GHz frequency. They created temporal averaged images if the sigma nought back-scatter coefficient. They also created filtered time series data of the back-scatter coefficient, and calculated the image ratios between each filtered image in the time series to enhance the back-scatter coefficient. By looking at this data in RGB, they could identify

back-scatter that changed from the constant back-scatter. In addition, a normalized difference vegetation index, NVDI, and a water index, NWDI, was produced from the ASTER satellite multi-spectral data to complement the SAR data. With the back-scatter coefficient data they were able to detect known Nasca lines, even with the poor resolution of the data. Puquios, which are underground water systems, similar to qanats common in the Middle East and Central Asia, were also studied using the SAR data together with the multi-spectral data. By looking at changes over time, the difference in back-scattering of the puquio paths could be observerd. The soil around functioning puquios are affected by moisture to a larger degree than the dry areas without puquios. They could observe how the back-scatter changed over time, and identify the puquios where water likely had been flowing in the period of observation. The SAR data was also used to observe the Cahuachi archaeological site, which was a major ceremonial center of the Nasca culture. The site contains large pyramid and temple structures. While the structures themselves were not very well identified with the 30m resolution SAR data, they could observe changes in the SAR data over time, and identify changes to the site. The changes were indicative of damage and looting of the site, showing that even low resolution SAR data could potentially be used for site monitoring of cultural heritage.

H.A. Orengo Et. al. performed a study of the Cholistan desert in Pakistan, using multi-temporal SAR and multi-spectral data to identify archaeological mounds from ancient settlements.[7] The area is believed to be an important settlement area in the period between 2500 and 1900 BC, called the Mature Harappan period. The study used a 5 year multi-temporal data set of Sentinel-1 SAR data and Sentinel-2 multi-spectral data, which was created in Google Earth Engine. The multi-spectral data tagged as affected by clouds were excluded when combining the data, and the data set consisted of 14 bands with VV and VH polarization in both ascending and descending mode for the SAR data and B2-B8A, B11 and B12 bands for the multi-spectral data. Creating a multi-temporal data set excludes many problems poor visibility conditions and environmental effects on the SAR data. The spatial resolution of the SAR data is 10m while the 10 bands of the multi-spectral data has spatial resolutions of 10m and 20m, deemed to be sufficient to detect the mounds. By using 25 previously mapped mounds the multi-temporal data for these mounds were used to train a 128 tree Random Forrest algorithm. The study effectively mapped the desert, finding a large number of previously unknown settlement mounds and finding settle-

ment patterns in the area. The mounds have a lot of clay material. The authors conclude that SAR signatures or multi-spectral signatures alone are not able to distinguish a mound from other areas containing naturally containing clay material. However, they believe the combination of these are enough to distinguish settlement mounds from non-anthropogenic features. Specifically they attribute the effectiveness of their method to the combination of high contrast in the SAR bands, the SAR bands ability to penetrate loose dry sand and the specific reflectance of the multi/spectral Red edge, NIR and SWIR bands for the mounds.

F.Chen Et al. present an overview of synthetic aperture radar remote sensing information for archaeology.[33] While at the time of writing, their article is 7 years old, their information is still valid. Newer SAR satellites with other SAR bands and better spatial resolution has become available, but their discussions in the article still applies to freely available SAR data. They discuss their opinions on how SAR data can be used in archaeology, and emphasizes heritage monitoring and analysis in addition to new heritage detection as the main two uses of SAR data in archaeology.

Table 2.4: Overview of articles on SAR uses in archaeology

Reference	Satellite - Band	Type of ML/DL	Description
[2]	Cosmo-SkyMed - X-band	N/A	Mapping and monitoring of "Tells", archaeological mounds from settlements, using Digital Elevation Models created from Cosmo-Skymed SAR data of variable resolution.
[30]	PALSAR 1 and 2 - L-band	N/A	Extraction of different archaeological features in SAR data, including archaeological settlements and roads in the Sinai desert.
[32]	ENVISAT ASAR - C-band	N/A	Mapping and monitoring of cultural heritage in the Nasca Region in southern Peru.
[33]	N/A	N/A	A discussion on use of SAR data in archaeology. Discusses approaches for reconnaissance of archaeological signs and how SAR interferometry can be used for the monitoring of cultural heritage sites.
[7]	Sentinel-1 - C-Band and Sentinel-2 Multi-spectral	Random Forrest classifier	Mapping and discovery of settlement mounds in the Cholistan desert using multi-temporal SAR and multi-spectral data and training a Random Forrest algorithm for generating a probability map of likely settlement mounds.

Chapter 3

Proposed Methods

3.1 Burial Mounds - Norway

For detection of burial mounds in Norway LiDAR based Digital Terrain Models is used. The Norwegian Mapping Service has over several years created a country wide 1m spatial resolution DTM which is available for download through their websites.[6] There are other projects which provide large areas of higher spatial resolution DTMs, as low as 0.25m resolution. In addition, the Norwegian Directorate for Cultural Heritage maintains a database for Norwegian cultural heritage, including positions of known burial mounds. This database is available in different formats usable with GIS software, through GeoNorge.[34].

The proposed method is to use the LiDAR based DTMs and cultural heritage database to generate training and testing data for the learning model. 4 different relief models of the DTM data will be made, Slope, Simple Local Relief Model and two Multi-Scale Relief Models with a scaling factor of 1 and 2. Once a model is trained and tested, it can be used to scan a large area by feeding a large composite image of an area identifying potential burial mounds. A user option of excluding results below and below a certain size will also be added, in addition to an option to exclude results where the bounding box ratio of the longest side against shortest side is above a certain user-defined threshold. If scanning a large area with more

than one model, the user can also compare the results between the models. The user will chose an input model result to compare and if results from this model are found in one of the other models selected, the results are kept.

Final result csv files with UTM coordinates can be used to create an ESRI shape file in QGIS (Python script) or ArcGIS Pro (Jupyter Notebok). The ESRI shape files can then be used with any GIS software supporting this format.

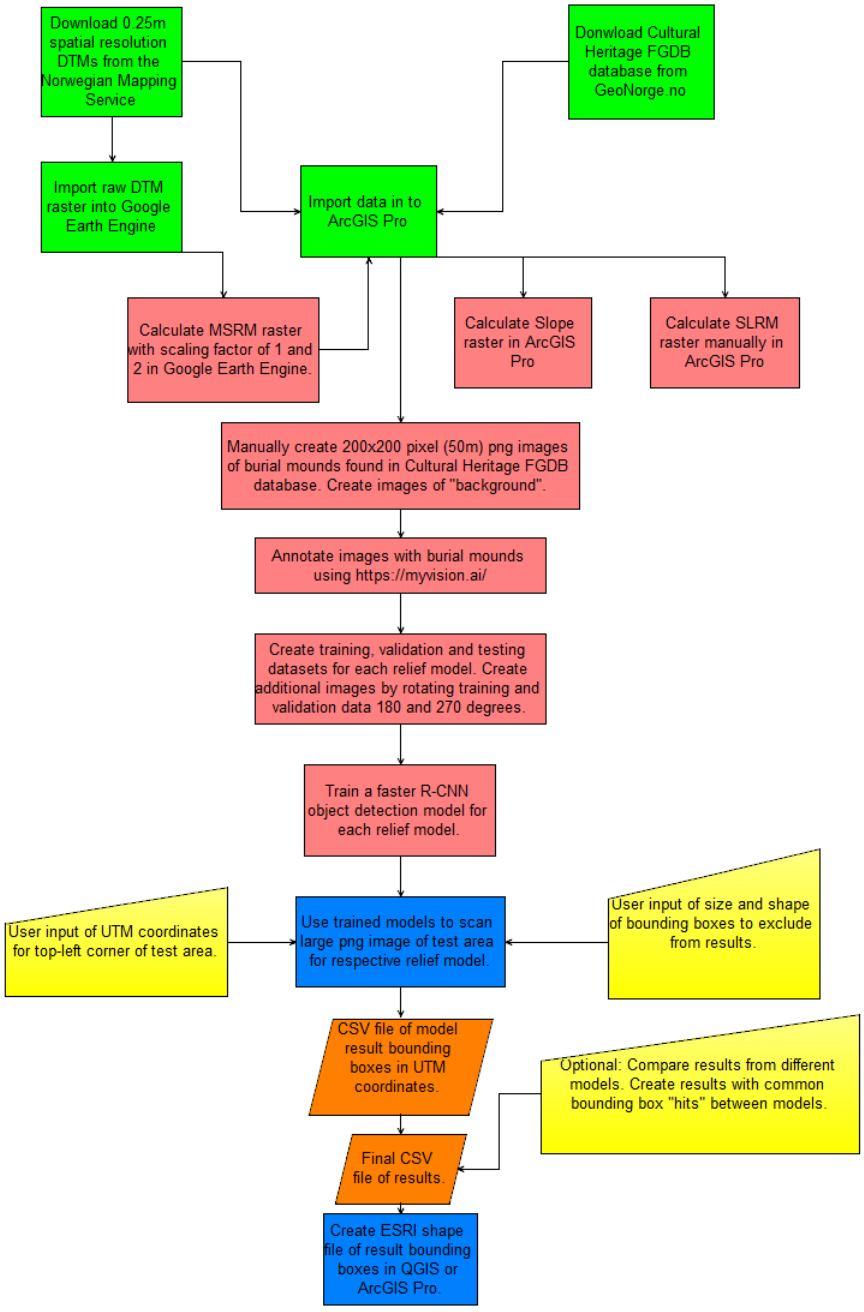


Figure 3.1: Proposed workflow for creating faster R-CNN models and applying trained models for burial mound detection.

3.2 Fortress Detection - Central Asia

The proposed method for detecting archaeological fortress structure is to use the Mapbox optical imagery to train a faster R-CNN object detection model. Fortress structures and their coordinates will be identified using vici.org. Mapbox will be used to download image tiles, which will be composited, for each fortress. Each image will be annotated and used to train the faster R-CNN object detection model. The trained model can then be used to scan a large image. A final CSV file with the resulting bounding boxes from the model will be produced with WGS1984 coordinates in decimal degrees. An ESRI shape file can be created in QGIS (Python script) or ArcGIS Pro (Jupyter Notebok) from this the results csv file. The ESRI shape files can then be used with any GIS software supporting this format.

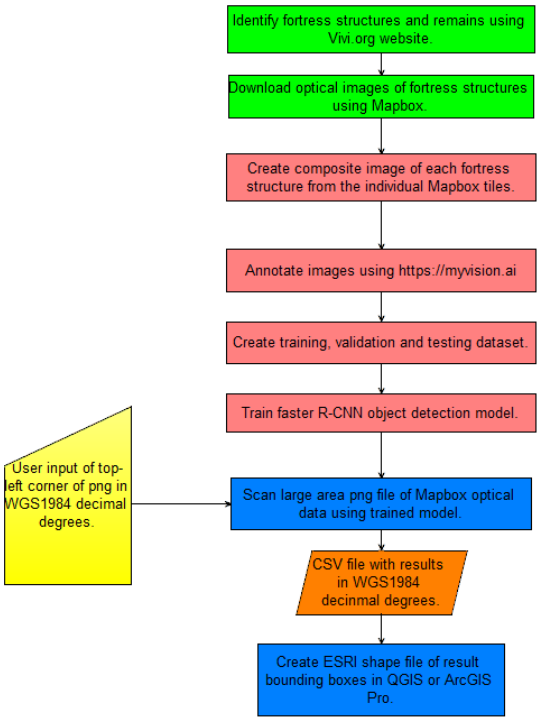


Figure 3.2: Proposed workflow for creating faster R-CNN models and applying trained models for fortress detection.

3.3 Settlement Mound Detection - Central Asia

H.A.Orengo Et. al. used multi-temporal Sentinel-1 SAR data and Sentinel-2 multi-spectral data to detect settlement mounds in the Cholistan desert in Pakistan.[7] By using their approach to generating data in Google Earth Engine, several areas in Central Asia was studied. By using vici.org, a website which maps a lot of archaeological remains identified throughout the region, the response of the data for these remains could be studied. The data is composited and averaged from acquisitions taken between October 2014 and June 2020. For SAR data in particular, this composited data removes a lot of noise and any kind of seasonal effects on the data, which might occur using only a very few acquisitions.

It was found that the SAR data in particular gave a relatively distinct response to a lot of the archaeological settlement mounds, many of which are classified as "unidentified site" in the vici.org website. By using an RGB display of the data, where red used VV polarized data in ascending mode, green used VH polarized data in ascending mode and blue used VV polarized data in descending mode, many of the mounds gave the response as seen in figure 3.3 and 3.4.

The sides of the mounds was found to give higher scattering effects than the surrounding background. One can see from observation that the left side of the mounds gave scattering more prominent in ascending mode, giving anywhere from yellow, orange through red color, while the right side of the mounds gave scattering more prominent in descending mode, giving a blue color. The intensity of the scattering, based on observation, seems to be related to the slope of the mound edges more than the actual height of the mounds. Some of the scattering response could also be related to the compaction of the soil in the mounds. Observation shows that a lot of the smaller mounds gets a dark yellow, orange through red color and transitioning to blue color immediately, as there is not much flat area in the middle of the mounds, as seen in figure 3.3. Larger mounds with more flat area, often gives a darker colour in-between the edges of the mound, as the flat area does not give as much scattering of the SAR signal. Figure 3.5 shows the SAR response and the optical image of the same square settlement mound.

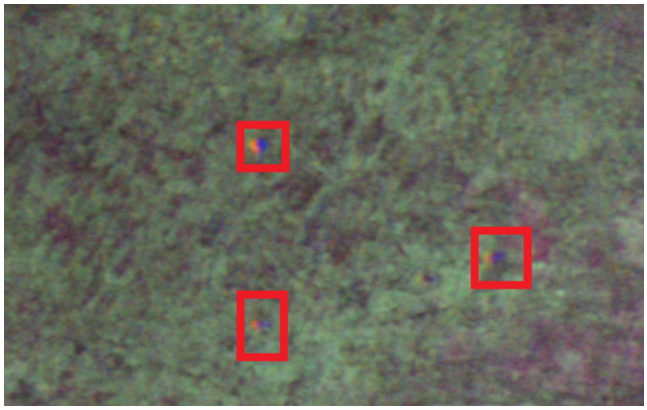


Figure 3.3: Example of SAR RGB response on settlement mounds west of Bukhara, Uzbekistan.

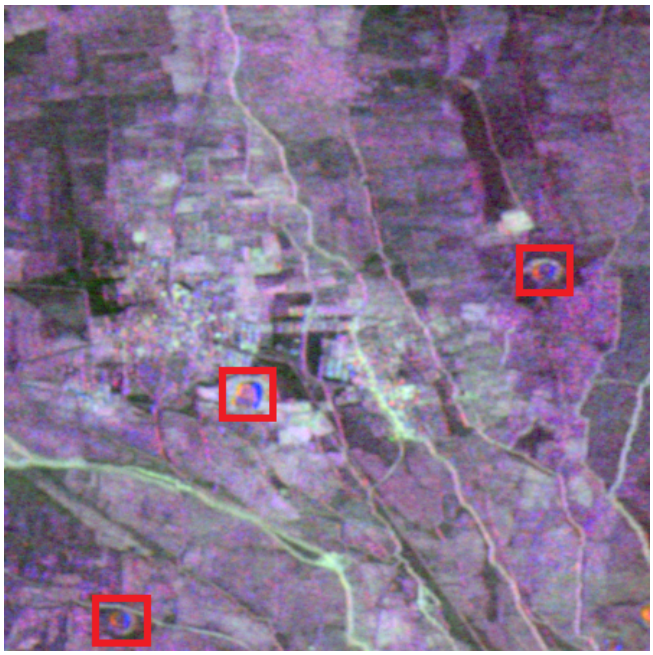
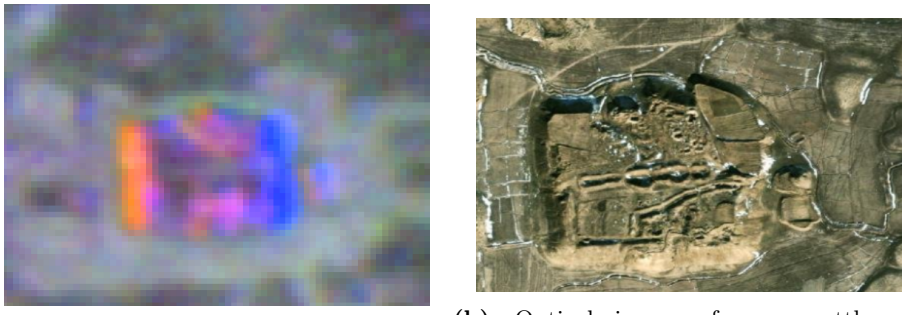


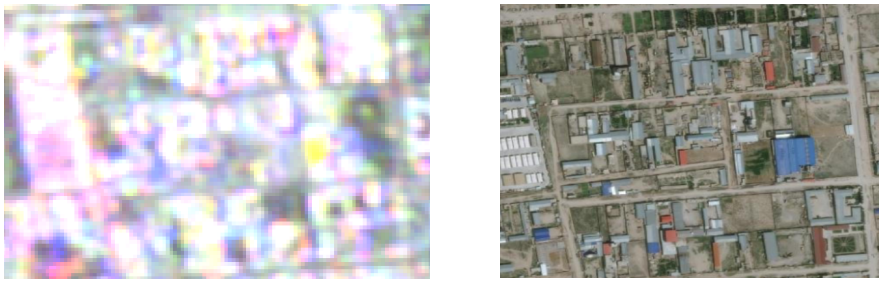
Figure 3.4: Example of SAR RGB response on settlement mounds near Masar-i-Sharif, Afghanistan.



(a) SAR RGB response of settlement mound.

(b) Optical image of same settlement mound.

Figure 3.5: SAR response and optical image of settlement mound near Sheberghan, Afghanistan.



(a) SAR RGB response

(b) Optical image

Figure 3.6: SAR RGB response and optical image of same area in Masar-i-Sharif, Afghanistan.

Modern buildings and structures generally gave a completely different SAR response than the observed settlement mounds, and very few modern buildings and structures would be confused with the response of a settlement mound. Examples are shown in figure 3.6.

It is important to note that the SAR data itself might not be able to distinguish a natural mound from an archaeological mound. Also, some ancient fortress structures or partial remains of those might also give a SAR response similar to that of a settlement mound. Hence, using faster R-CNN object detection on the SAR data, looking for SAR responses similar to those shown in this section, will find possible archaeological settlement mounds and other possible archaeological structures giving a similar response. If the model finds a SAR signature similar to those shown, which happens to be

something else than an archaeological settlement mound, the model has actually worked. The 10m spatial resolution of the SAR data sets limitations to how well a model can be trained to distinguish a small fortress from a settlement mound.

The proposed method for detecting archaeological settlement mounds is to use the SAR RGB response, as described above, to train a faster R-CNN object detection model. Multitemporal composited Sentinel-1 SAR Geotiff data will be created in Google Earth Engine. Using ArcGIS Pro, selected areas will be exported to png format and split into 448x448 pixel images. Using vici.org for the selected areas, archaeological features giving the expected SAR signature will be annotated and used to train the faster R-CNN object detection model. The trained model can then be used to scan a large image. A final CSV file with the resulting bounding boxes from the model will be produced with WGS1984 coordinates in decimal degrees. This csv file can be used to create an ESRI shape file in QGIS (Python script) or ArcGIS Pro (Jupyter Notebok). The ESRI shape files can then be used with any GIS software supporting this format.

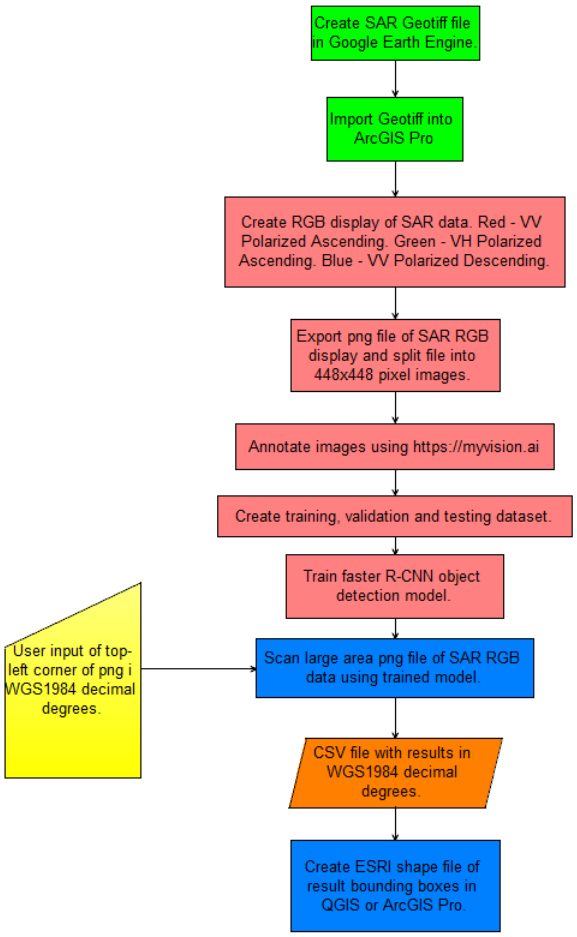


Figure 3.7: Proposed workflow for creating faster R-CNN models and applying trained models for settlement mound detection.

Chapter 4

Implementation

4.1 Data Gathering and Image Preparation

4.1.1 Burial Mounds - Norway

Data for for burial mounds in Norway was produced from 0.25m spatial resolution Digital Terrain Models downloaded from høydedata.no, which are available as part of different projects, mostly used to create the national coverage 1m spatial resolution DTM data, called NDH. All the DTMs are produced from LiDAR measurements, and were downloaded in GeoTiff format in UTM coordinates. In addition, a geospatial database in FGDB format was downloaded from GeoNorge, containing the individual cultural heritage markings maintained by the Directorate for Cultural Heritage. The process for generating training, validation and testing data for the faster R-CNN object detection training using Google Colab is shown in figure 4.1.

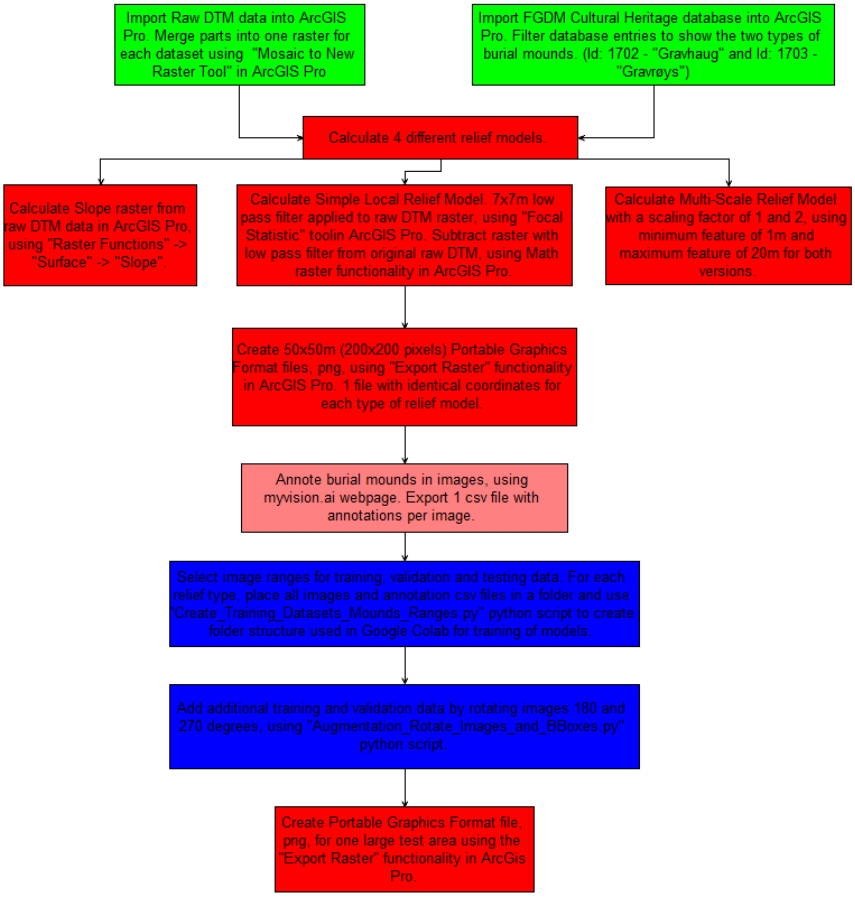


Figure 4.1: Process for generating training, validation and testing data for faster R-CNN model training in Google Colab.

The Multi-Scale Relief models were originally calculated using a python script, using the arcpy library which contains a lot of the ArcGIS Pro functionality. Due to the large space requirements and low speed of calculation (several hours per data set), a Google Earth Engine Javascript is available in the Github repository for this thesis, which can calculate and create MSRM GeoTiff files.

Once each of the 4 types of relief models were generated, burial mounds

were found from the database of cultural heritage, by overlaying these on the rasters. The database contains two different classifications of burial mounds, called "gravhaug" with id 1702 and "gravrøys" with id 1703. The markers were found to be inexact in many places, and often marking more than one burial mound, so visual identification of mounds were needed in some areas. In addition, several mound markers indicated mounds that were severely damaged or completely removed. The database often contained descriptions of the mounds gathered through surveys throughout the years.

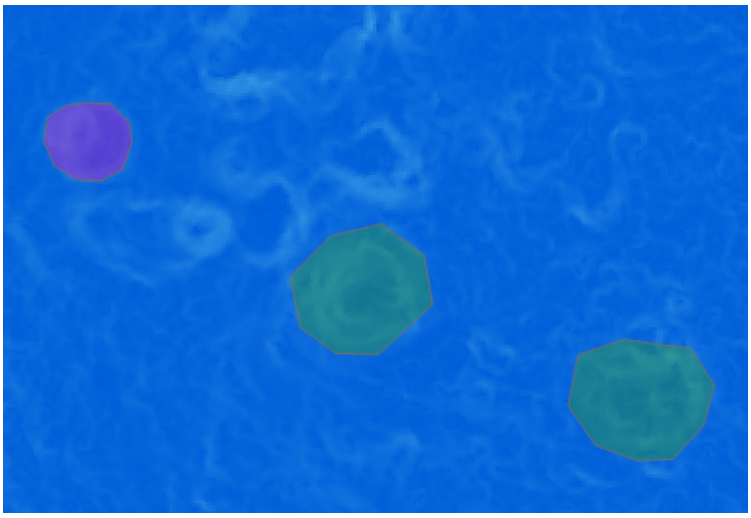


Figure 4.2: Slope display in ArcGIS Pro with burial bound markers from Cultural Heritage database.

Since many of the burial mounds are in burial fields and close together, care was taken not to have partial burial mounds at the edges of the images. The Export Raster functionality in ArcGIS Pro exports the images with the original pixel spacing as is contained in the raster, which is important. When exploring the QGIS software package as an alternative to ArcGIS Pro, export of png images seems to be dependent on zoom level and is not preserving the original 0.25m per pixel quality.

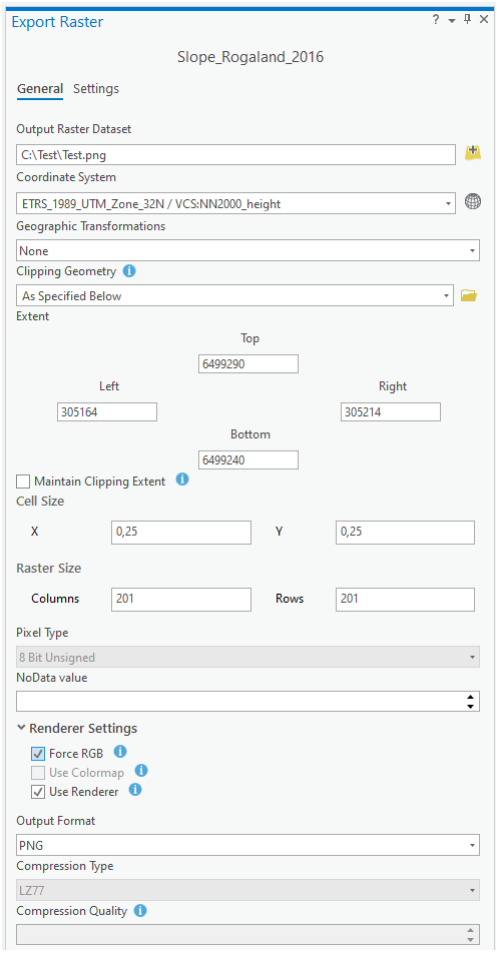


Figure 4.3: Export of png image with specific coordinates forcing RGB coloring according to scale.

To test the usage of the final trained model, a png file for each of the 4 relief types were created manually in ArcGIS Pro, for a 9km2 test area. The final application consumes this large png file, and automatically splits the image, with or without overlap and does a complete scan of the area.

4.1.2 Fortresses - Central Asia

To gather data for training of a faster R-CNN object detection model for fortresses in Central Asia, the vici.org webpage was used to find placement of fortresses throughout the region. Mapbox was used to download image tiles for the area at selected zoom levels, which was then combined to a composite image for the selected fortress and surrounding area. Each image was annotated using the myvision.ai webpage, and the annotations were exported in csv format. The data was split into training, validation and testing data for use with the faster R-CNN object detection training process in Google Colab.

4.1.3 Archaeological Settlement Mounds - Central Asia

Google Earth Engine was used to generate GeoTiff data for the Sentinel-1 Synthetic Aperture Radar, SAR. Using Javascript in GEE, SAR data acquisitions for a 5 year period was combined into 4-band composite, and exported to Google Cloud and downloaded for use in ArcGIS Pro. Figure 4.4 shows the process for creating the training and testing data used for the faster R-CNN object detection model training in Google Colab.

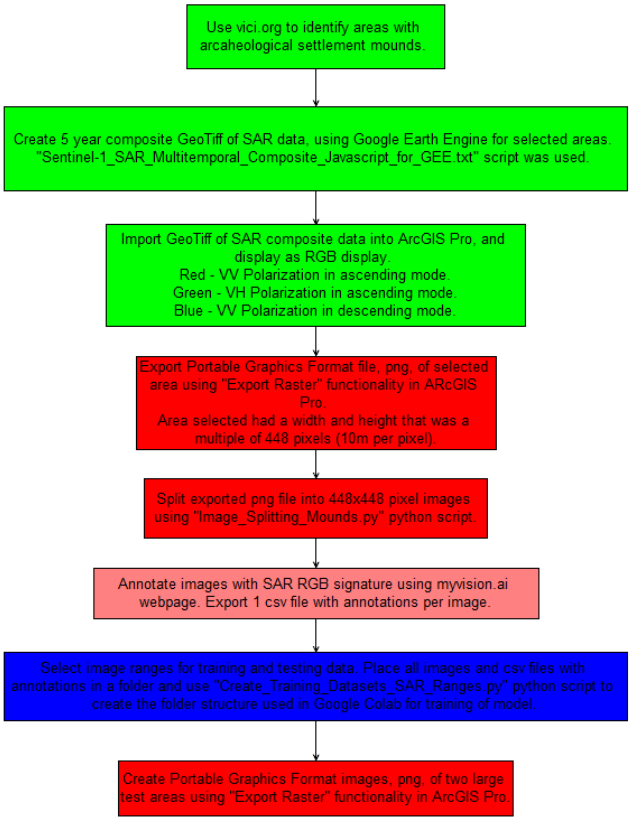


Figure 4.4: Process for generating training and testing data for faster R-CNN model training in Google Colab.

To test the usage of the final trained model, a png file for two test areas were created manually in ArcGIS Pro. The final application consumes these large png files, and automatically splits the image, with or without overlap and does a complete scan of the area.

4.2 Faster R-CNN Model Training

For the thesis, faster R-CNN object detection models were trained. The deep learning framework used was PyTorch.

The backbone of the faster R-CNN models was ResNet-50-FPN. All models were trained using Stochastic Gradient Decent optimizer, apart from the training of the fortress model, which used the ADAM optimizer.

Using Jupyter notebooks, all training was done in Google Colab. All the notebooks are available in the thesis Github repository.

4.3 Final Application

The web-based application built for applying conveniently the model was built with Python/Flask and JavaScript. The application supports scanning of large images splitting the images into selected pixel size, and a selection of overlap to use. Download of optical imagery from Mapbox is also supported. Functionality for result filtering and model cross-validation of burial mound results is available. In addition, the application supports upload of newly trained object detection models.

Chapter 5

Results and Discussion

5.1 Results for Tumuli/Burial Mounds

The faster R-CNN object detection models were trained with a learning rate scheduler, reducing the learning rate after a fixed number of epochs of training. All models were trained using a fixed training, validation and testing data sets. The details of the training, validation and testing splits are shown in table B.2. Additional training and validation images was added by rotating the images listed in the table B.2 180 and 270 degrees, increasing the number of images by a factor of 3.

The training parameters for the Slope, Simple Local Relief Model and the two Multi-Scale Relief Model are shown in table 5.1.

The two metrics found to be most relevant for evaluating the quality of the models were found to be the precision (equation 2.9) and recall (equation 2.10).

The results of the testing data is shown in table 5.2. The Slope model gives clearly the best precision and recall, while the Multi-Scale Relief models give a recall closer to the Slope model, but with a much worse precision. 33 of the 46 total images (50x50m images) used for the testing had burial mounds in them, while the remaining 13 had no mounds in them. Using

Table 5.1: Training Parameters for faster R-CNN Burial Mound Models

Model	# of Epochs	Starting LR	LR Reduction Factor	# of Epochs Before LR Reduction
Slope	120	0.005	0.1	30
SLRM	200	0.005	0.1	50
MSRM Scaling Factor 1	200	0.005	0.1	50
MSRM Scaling Factor 2	200	0.005	0.1	50

Table 5.2: Faster R-CNN Burial Mound Model Results for Testing Set

Model	Total Mounds	TP	FP	FN	Precision	Recall
Slope	79	60	17	19	0.78	0.76
SLRM	79	38	21	37	0.64	0.51
MSRM Scaling Factor 1	79	54	54	25	0.50	0.68
MSRM Scaling Factor 2	79	54	102	25	0.35	0.68

more images without any burial mounds in them in the testing data set, would affect the precision of all models and likely decrease the precision values for the testing.

To do a more practical evaluation of the final models, a 9km2 area around Klepp in Rogaland was used. Known burial mounds for the area was found using the Cultural Heritage database. The database, as in many areas, did not map individual mounds very well and some markers indicated more than one mound. The Cultural Heritage search website of the Department of Cultural Heritage, "Kulturminnesøk", was used to identify as many ground truth mounds as possible, giving a total of 76 burial mounds. Part of the area is shown in figure 5.1, with boxes showing placement of confirmed burial mounds. The area could contain more burial mounds that are not properly marked. For almost all tests, two potential burial mounds were consistently found, and are likely burial mounds and not counted as false positives in the results. These two can bee seen in figure 5.2.

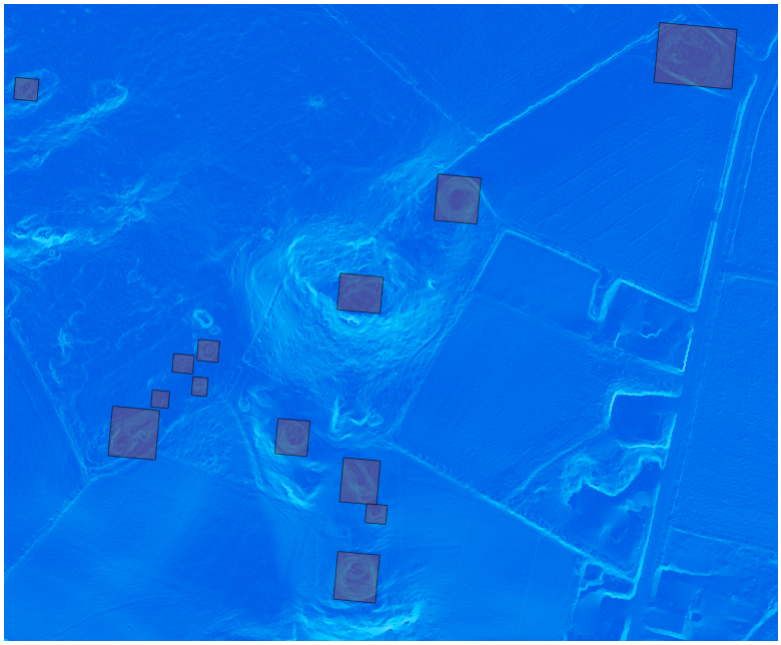


Figure 5.1: Part of Klepp test area displayed with Slope and confirmed burial mound markers.

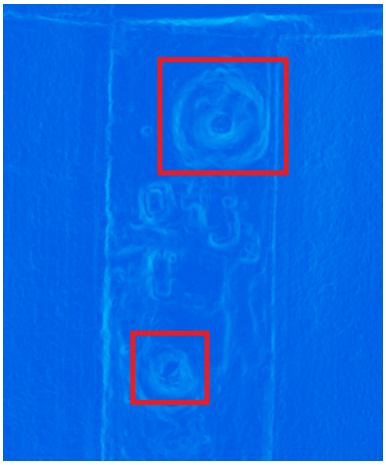


Figure 5.2: Two likely burial mounds indicated with red boxes. Data presented with Slope.

Table 5.3: Faster R-CNN Burial Mound Model Results for Klepp Test Area

Model	Total Mounds	TP	FP	Possible Burial Mounds	Precision	Recall
Slope 200px	76	38	338	2	0.10	0.50
Slope 400px	76	35	54	2	0.39	0.46
Slope 600px	76	25	23	2	0.52	0.33
SLRM 200px	76	48	1083	2	0.04	0.63
SLRM 400px	76	16	52	2	0.24	0.21
SLRM 600px	76	4	6	1	0.4	0.05
MSRM Scaling Factor 1 200pxx	76	47	1708	2	0.03	0.62
MSRM Scaling Factor 1 400pxx	76	30	533	1	0.05	0.40
MSRM Scaling Factor 1 600px	76	25	358	1	0.07	0.33
MSRM Scaling Factor 2 200px	76	49	4325	2	0.01	0.65
MSRM Scaling Factor 2 400px	76	42	1522	2	0.03	0.55
MSRM Scaling Factor 2 600px	76	30	884	2	0.03	0.40

The area was scanned with each of the 4 models, automatically splitting the large 9km2 input png format image into 200x200, 400x400 and 600x600 pixels parts, which is the equivalent of 50x50m, 100x100m and 150x150m sizes respectively. While the input images for training used 200x200 pixel images, this size increases the likelihood of a mound going across more than one image. Also, ideally a feature being searched for by the models, should not be more than 25% of the image size.

The results are shown in table 5.3. Of the 4 models, only the slope model gives a high precision while having decent recall values.

Table 5.4: Results for Klepp Test Area with Exclusion of Result Size and Shape

Model	Total Mounds	TP	FP	Possible Burial Mounds	Precision	Recall
Slope 200px	76	38	268	2	0.12	0.50
Slope 400px	76	35	52	2	0.40	0.46
Slope 600px	76	23	23	2	0.50	0.32
SLRM 200px	76	48	1009	2	0.04	0.63
SLRM 400px	76	16	50	2	0.24	0.21
SLRM 600px	76	4	4	1	0.5	0.04
MSRM Scaling Factor 1 200pxx	76	47	1518	2	0.03	0.62
MSRM Scaling Factor 1 400pxx	76	28	499	1	0.05	0.37
MSRM Scaling Factor 1 600px	76	23	295	1	0.07	0.30
MSRM Scaling Factor 2 200px	76	49	4095	2	0.01	0.65
MSRM Scaling Factor 2 400px	76	42	1300	1	0.03	0.55
MSRM Scaling Factor 2 600px	76	25	563	1	0.04	0.33

The majority of burial mounds are fairly round in shape, larger than 5m in diameter with very few burial mounds in Norway larger than 50m. To attempt to reduce the number of false positives, any mound larger than 50m and smaller than 6m were excluded from the results. The diameter was calculated as the largest of the width and height of the model bounding box result. In addition, any bounding box result with a width to height or height to width ratio above 1.7 was excluded, due to the round nature of the majority of mounds. The results are shown in table 5.4. The exclusion has very little effect on number of true positives. The maximum number of true positives removed, was from the Multi-Scale Relief model with a scaling factor of 2 with a 600 pixel size scan, where 5 true positives ended up being removed. The models with high precision had few false positives removed, while the exclusion removed many more false positives from models with poor precision.

Table 5.5: Results with Exclusion of Result Size and Shape and Common hits between Slope and other Models

Model	Total Mounds	TP	FP	Possible Burial Mounds	Precision	Recall
Slope 200px	76	36	195	2	0.16	0.47
Slope 400px	76	32	37	2	0.46	0.42

Another approach to reduce false positives metric is to cross-validate models with respect to the best model. According to the results the Slope model is considered to be the best due to it's higher precision. This calculation was done for scan of 200 and 400 pixel size, after excluding mound hits of the same sizes and ratios as presented in table 5.4 before comparison of the models. The 600 pixel scan for the 4 models was found to have a too low recall value to be relevant, although the precision was higher than the 200 and 400 pixel scans. The results are shown in table 5.5. For the 200 pixel scan, the number of false positives is reduced from 268 to 195, while removing two true positives. For the 400 pixel scan, the number of false positives is reduced from 52 to 37, while removing three true positives.

The effect of scanning the 9km2 area using overlap was also tested. Additional images overlapping the original 200x200 or 400x400 pixel size by 50% was tested. Scanning with overlap caused duplicates of bounding boxes which were excluded by added functionality. The center of each model result bounding box was compared to the other result bounding boxes and removed if the center was within one of the other bounding boxes. The overlap scans for 200 and 400 pixel size was done on the best model, which was the Slope relief model. Both were tested with and without excluding bounding boxes below 6m and above 50m size and width/height or height/width ratio above 1.7. The results can bee seen in table 5.6. The effect of using overlap is that the number of false positives are increased, but the number of true positives also increases significantly. Hence, increasing the recall and reducing the precision of the model results. Results for parts of the area using overlap and excluding bounding boxes is shown in figure 5.3

Table 5.6: Results Using Overlap

Model	Total Mounds	TP	FP	Possible Burial Mounds	Precision	Recall
Slope 200px	76	53	707	2	0.07	0.70
Slope 200px w/exclusion	76	53	558	2	0.09	0.70
Slope 400px	76	50	122	2	0.29	0.66
Slope 400px w/exclusion	76	50	117	2	0.30	0.66

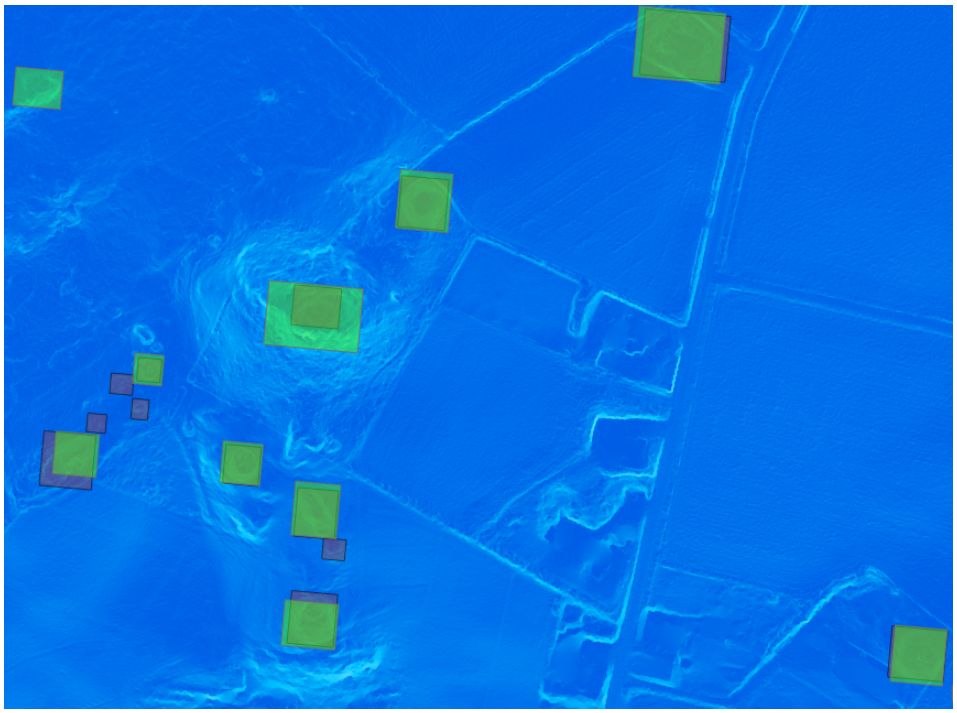


Figure 5.3: Result bounding boxes (green) overlaid Slope relief and ground truth markers. 400 pixel scan of Slope with exclusion.

All results discussed so far are based on a 0.25m spatial resolution Digital Terrain Model data. Data with this resolution is not available for all of Norway so a final test was done on the Klepp test area using 1m spatial resolution from the National Height Model DTM data. The Slope model

Table 5.7: Results Using Overlap and 1m Spatial Resolution Data

Model	Total Mounds	TP	FP	Possible Burial Mounds	Precision	Recall
Slope 0.25m w/exclusion	76	50	117	2	0.30	0.66
Slope 1m w/exclusion	76	50	215	2	0.19	0.66

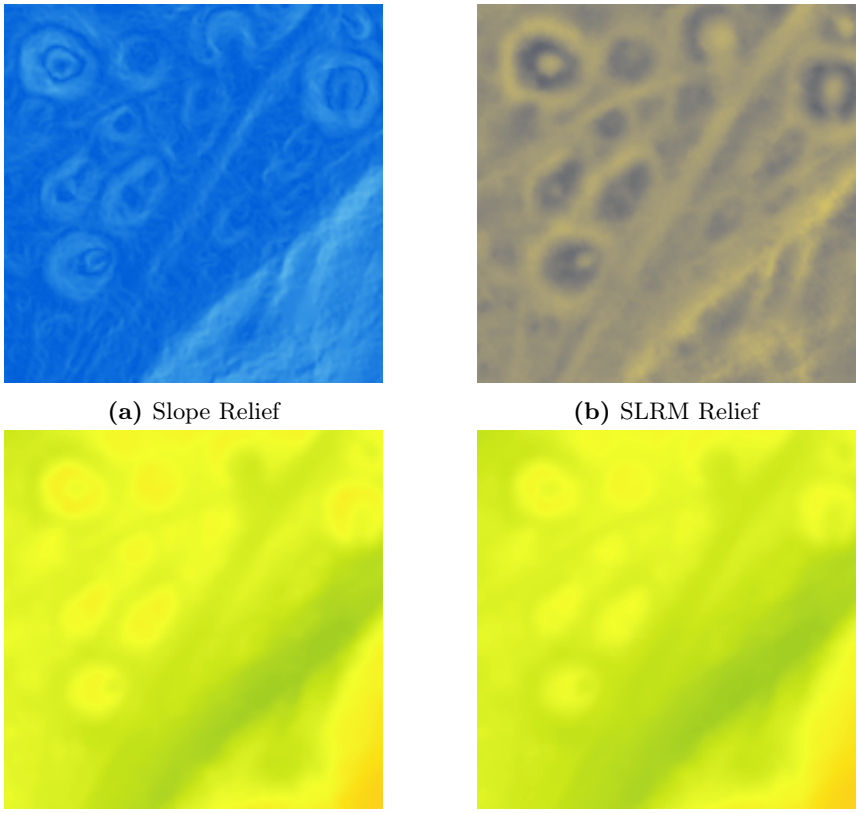
was used to compare the results between the spatial resolution. The results are shown in table 5.7. The scan of the test area was done by splitting the area into 100x100 pixels which is the equivalent of 400x400 pixels for the 0.25m spatial resolution data using a 50% overlap. The results shows that the model finds the same number of ground truth burial mounds, but the number of false positives almost doubles.

5.2 Discussion of Results for Tumuli/Burial Mounds

The testing data from model training process indicates that of the 4 models, the Slope model is the most accurate, giving the lowest number of false positives and the highest number of actual burial mounds found. The Multi-Scale Relief models have a higher recall value than the Simple Local Relief model but with a lower precision based on the same test data. The test data does not have a large amount of images without burial mounds and the numbers might not reflect accurately the precision when applying the models to larger areas.

The trends are however similar when looking at the Klepp 9km2 test area. The precision of the Slope model is much better than the SLRM and MSRM models. While the recall values might be better for SLRM and MSRM than for Slope in this test area, the precision is so low, that the actual recall values are not particularly relevant. The SLRM and MSRM models seems to be finding a huge amount of features which has some height difference above the surroundings, but not really differentiating burial mounds from other features.

As explained in section 2.1.3, the Slope model shows distinct slope changes calculated between pixels. The SLRM and MSRM models however, uses low pass filtering to enhance height differences in the Digital Terrain Models. These low pass filters are removing important details in the data which the Slope relief does not, and hence making it harder for a deep learning model to distinguish details of some burial mounds from other non-burial mound features. This can bee seen in figure 5.4. The other issue with using SLRM and MSRM relief models is that when the mound heights above the background is very low, as is the case for a lot of smaller mounds, the height difference scales of the relief models for the area calculated affects how visible the mounds are in the resulting display. For the Klepp test area, the SLRM has a height difference range about 73m and the MSRM with scaling factor 1 and 2 has height difference range which is about 2.1m and 14m respectively. The Slope model is completely independent of height range and height difference range for the area in which the slope relief is calculated. It uses a fixed scale of 0 to 90 degrees.



(a) Slope Relief

(b) SLRM Relief

(c) MSRM Scaling Factor 1 Relief

(d) MSRM Scaling Factor 2 Relief

Figure 5.4: Comparison of Relief Models

I. Berganzo-Besga Et. al. reported in their study of tumuli in Spain that their use of the Multi-Scale Relief model with a scaling factor of 2, and calculating this relief model with a minimum and maximum feature range between 1m and 19m respectively, improved their results compared to their use of Slope and SLRM relief models.[3] In their study the SLRM relief model gave a recall of 0.42 and precision of 0.85, the Slope model gave a recall of 0.46 and a precision of 0.91 and the MSRM relief model gave a recall of 0.58 and precision of 0.95, showing that the MSRM relief model had a large effect on the results. In their study, Yolo v3 was used as a object detection algorithm, which is different from the faster R-CNN used in this thesis. The results using faster R-CNN object detection on burial mounds in Norway shows a different trend for the results when comparing the models, clearly showing that the Multi-Scale Relief model is not improving results. Also, the SLRM model gives worse results than the Slope model, while in the results from Spain they give similar results. One reason for this large difference can be that over 90% of the mounds used for training in the Spanish study had a diameter over 18m, which is significantly higher than the average diameter of 9.7m of the mounds used for training in this thesis. The size distribution of mounds used for training, calculated from annotated bounding boxes, can be seen in table B.2. The much larger average diameter also makes it likely that the average height of the mounds is much higher in the Spanish study than the mounds used in this thesis. In addition, the Spanish study augmented images, using larger mounds to simulate smaller mounds due to lack of training data. This approach would not really be valid, based on the observation of burial mounds in Norway, as a signature of a large mound and small mound in Norway can be quite different in many cases.

Excluding model results where the resulting bounding boxes is above and below a certain size, and the ratio of the width/height or height/width is above a certain ratio, does have a small effect on the number of false positives for the Slope model, but is most effective when being applied to the other models with very low precision.

Using multiple models and comparing the most accurate, Slope, with the others, gives a small improvement on the precision and number of false positives, but also reduces the recall and number of true positives.

Considering the most accurate model, the Slope model, the major effect on

results is scanning a large area with or without overlap and the size of the area splitting. When excluding mounds of certain sizes and ratio, the 400 pixel scan with overlap increased the recall value from 0.46 to to 0.66 percent while reducing the precision from 0.4 to 0.29. The number of ground truth mounds found increased from 35 to 50, while the number of false positives increased from 52 to 117. The selection of parameters to use for scanning larger areas with the Slope model has a large effect on the results. Based on the results, the conclusion is that using the Slope relief model and scanning an area with overlap, splitting the area into 400x400 pixel images, gives the best results. It identifies the highest number of actual burial mounds, while keeping the number of false positives relatively low.

Ø.D.Trier Et. al. used faster R-CNN object detection for detection of burial mounds on a 67km2 area in Larvik municipality. Their results for this area was that they detected 38% of the burial mounds, finding 102 of 269 mounds. The false positive rate was 86%, having 610 false positives. Their results are based on using a Simple Local Relief Model. Comparing this to the Slope relief model and 400 pixel scan when excluding mound results with certain sizes and ratio, the Slope model gives almost a 30% improvement of the recall and a 15% improvement of the precision. However, their results are based on a different and larger area containing on average 4 mounds per km2, while the Klepp test area used in this thesis contains on average 8.4 mounds per km2, which likely affects the numbers somewhat. Due to the fact that the burial mound distribution in their test area is not known to us, a direct quantitative comparison of the models is not possible. To compare the results a manual check of each mound in the area would have to be done, and in their study they have corrected a lot of burial mound markers from the Cultural Heritage database. Another difference between their study and this thesis is that they used a VGG16 deep neural network as the backbone of faster R-CNN, while this thesis uses ResNet-50-FPN.

The comparison of 1m and 0.25m spatial resolution data shows that the number of burial mounds found might not be that highly affected, but that reducing the resolution will increase the number of false positives. If an area contains a lot of smaller burial mounds though, it is likely that using lower spatial resolution data will affect the recall value, as a lot of the important details of the smaller mounds will become less clear.

In conclusion, the Slope model gives the best results. The major problem

with detecting burial mounds using faster R-CNN object detection, is the number of false positives. A too low precision value renders the model too poor to be of any practical use, even though the recall value is high. The training data has a large variety in the burial mound signatures, and improving the models would require additional training data. Also, a lot of man-made features, like stone gathered together from farm fields, can give signatures similar to certain types of burial mounds and false positives are to be expected.

5.3 Results for Fortresses

The object detection model was trained with Adam optimizer. Adam optimizer is usually used in the initial cycles of model training process due to it's ability to give decent results without any hyper-parameter fine-tuning, like learning rate and decay.

The training process included aggregation of 300 fortresses from various countries in Central Asia and the Middle East. Their coordinates were retrieved from vici.org

Table 5.8: Results From Scan of Fortresses from various countries in Middle East

Total Images	Total Fortresses	TP	FP	Precision	Recall
54	35	12	65	0.15	0.34

5.4 Discussion of Results for Fortresses

The results were not good enough to progress towards the research of fortress detection. Due to that, the focus was turned to other methodologies to detect archaeological structures, which gave better results.

5.5 Results for Archaeological Settlement Mounds

The faster R-CNN object detection model was trained with a learning rate scheduler, reducing the learning rate after a fixed number of epochs of training. The model was trained using 120 epochs, with a starting learning rate of 0.005. The learning rate was reduced by a factor of 0.1 every 30 epochs. Due to the nature of the SAR responses for the settlement mounds, which is explained in section 3.3, no rotation augmentation was done on the images used for training, as this would change the expected SAR signal RGB response.

A total of 410 SAR signatures were used for training and validation, were 20% of these were randomly selected for validation. 70 SAR signatures were used for testing. A total of 769 448x448 pixels images were used for the training and validation, each image covering 4.48kmx4.48km. The SAR signatures are generally a very small part of the images, and areas of varied background were selected. The total area of the images that was part of training was 15435km2, while the total area used for testing was 2970km2. The details can be seen in table B.5.

The results for the testing data used during training found 54 true positives, 32 false positives and 16 false negatives. This gave a recall of 0.77 and precision of 0.63.

To further test the finished model, two test areas with quite different background SAR responses were selected. Each area was scanned by automatically splitting a png file of the entire area into 448x448 pixel images, consumed by the model. The first area was around Kandahar in Afghanistan and the second was around Bukhara in Uzbekistan. The Bukhara area is fairly flat and has a lot of populated areas, in addition to farm areas. The Kandahar area has hills and less populated areas, while containing a lot of farm land. The overall SAR RGB response in the Kandahar area is also much darker than the Bukhara area.

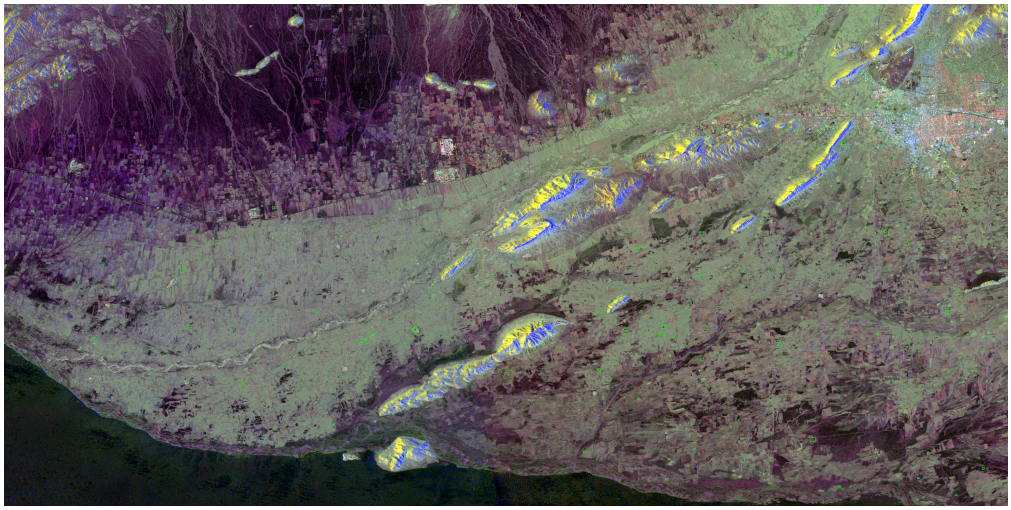


Figure 5.5: Kandahar area scanned. Green boxes are results from trained model.

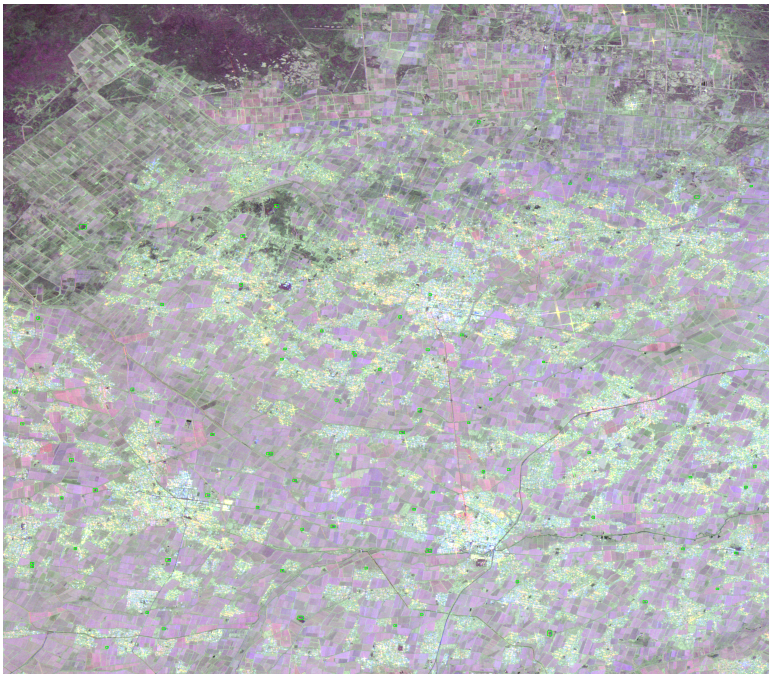


Figure 5.6: Bukhara area scanned. Green boxes are results from trained model.

Table 5.9: Results from Scans of Bukhara and Kandahar Test Areas

Area	Total Mounds	TP	FP	Precision	Recall
Kandahar	120	20	66	0.23	0.17
Bukhara	82	44	23	0.66	0.54

Table 5.10: Results from Scans of Bukhara and Kandahar Test Areas when Including Potential Settlement Mounds

Area	Total Mounds	TP	FP	Precision	Recall
Kandahar	129	29	57	0.33	0.22
Bukhara	101	63	4	0.94	0.62

For the Bukhara test area ground truth settlement mounds were found using vici.org. It should be noted that the vici.org website tags all these mounds as unidentified sites. For the Kandahar area ground truth settlement mounds were found from listings of archaeological and cultural heritage of the area made available by M.Karaucak Et. al.[35]. The results for both test areas are shown in table 5.9.

When evaluating the results from the model, the Kandahar area had 9 hits which could be potential settlement mounds not part of the ground truth data, while the Bukhara area had 19. This evaluation is of course subjective, and not an evaluation by a qualified archaeologist. Table 5.10 shows the results if these are added as actual settlement mounds.

No research has been found which uses a SAR RGB response alone to find archaeological settlement mounds. Hence, the model can not be tested directly against other research results. However, H.A.Orengo Et. al. used the same multi-temporal composite method to identify settlement mounds in the Cholistan desert in Pakistan. They combined Sentinel-1 SAR data and Sentinel-2 multi-spectral data into a 14 band composite. Using a Random Forrest classifier in Google Earth Engine, they marked 25 known mounds and areas that was not mounds to create a probability map. By using this method they identified a large number of settlement mounds throughout the desert. This same approach was used on the Bukhara test area, to see give a comparison to faster R-CNN model results. Instead of combining 14 bands, only the three SAR bands used in the RGB composite used for

training of the faster R-CNN model was used in a Random Forrest classifier in Google Earth Engine. The test area was extended further south, and 23 features from this additional area with a clear SAR RGB mound signature was marked. In addition, 20 areas without any SAR RGB signature was marked, including farm areas, modern structures and roads. The resulting probability map of the Bukhara test area is shown in figure 5.7, where the red color is a high probability close to 1, and green is a low probability close to 0.

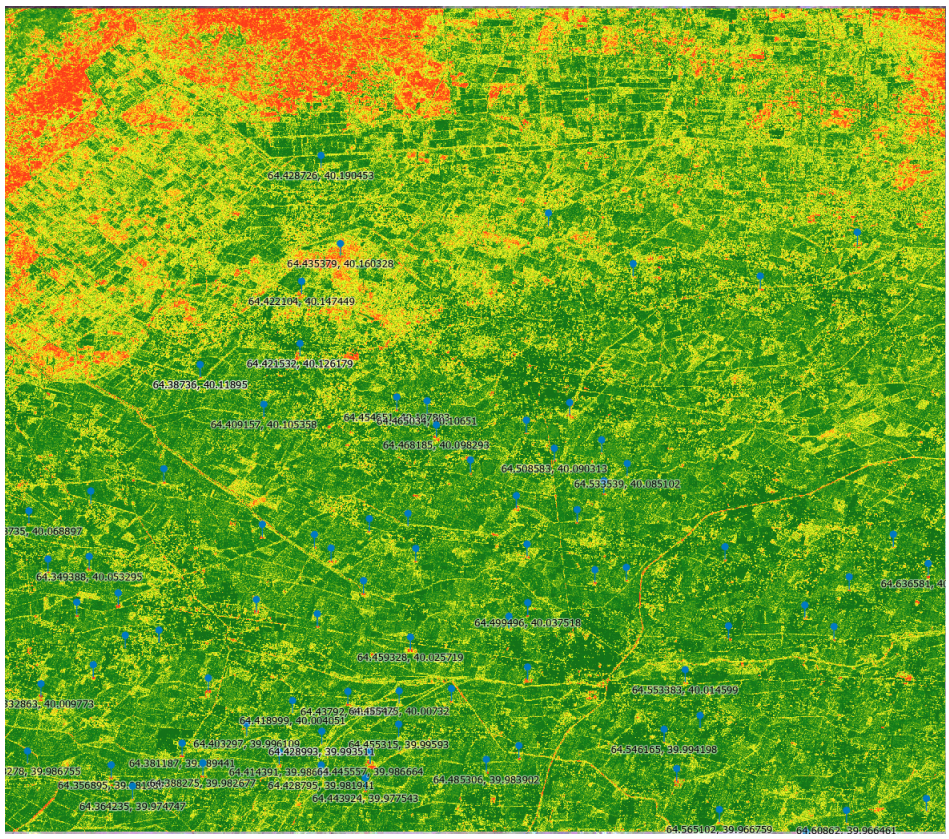
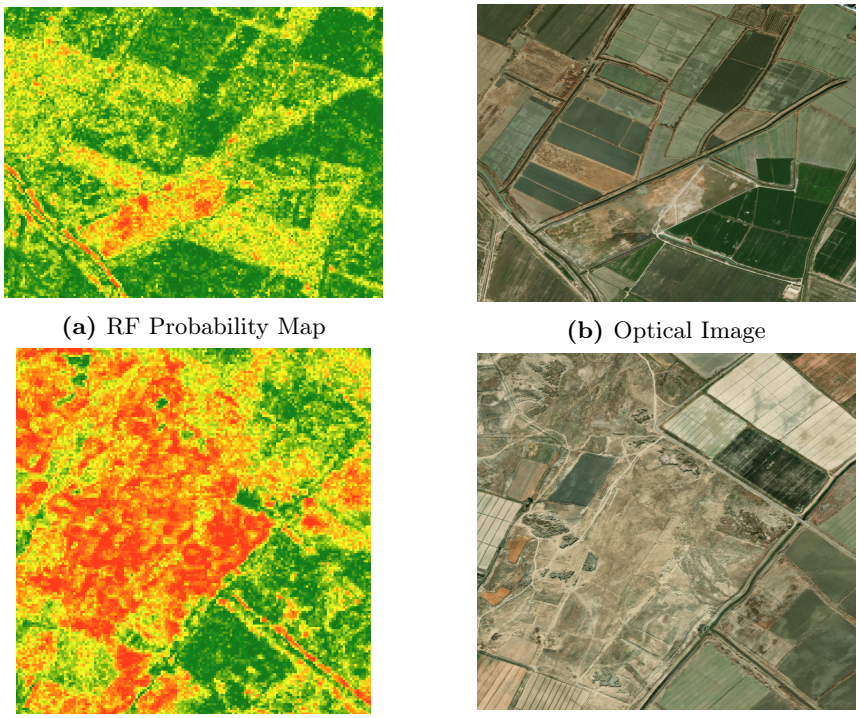


Figure 5.7: Bukhara area Random Forrest Probability Map with Ground Truth Mound Markers

From the probability map, a visual analysis was done, which is quite subjective, identifying high probability features which could be mounds. By doing this, 59 of the ground truth settlement mounds were identified, which is 15



(a) RF Probability Map

(b) Optical Image

(c) RF Probability Map

(d) Optical Image

Figure 5.8: Random Forrest Probability vs Optical Image

more than the faster R-CNN object detection model. Probable mounds, not part of the ground truth mound set, were not counted. Parts of the probability map is giving a high probability (red) for large continuous areas, in which possible mounds can not be identified. There are also a large number of high probability features which are not mounds, and a lot of high probability "noise" in the map. In addition, the probability map has to be visually analysed by a person, to identify any likely mounds, unlike the faster R-CNN object detection model. The probability map can not give a precision value for comparison with the object detection model. Two examples of high probability features which are not mounds are shown in figure 5.8

5.6 Discussion of Results for Archaeological Settlement Mounds

The results of the SAR RGB model is very different for the two areas tested, with the results from the Bukhara test area being clearly better. Of the RGB SAR features annotated and used in training and validation, only 57 of the 410 total features were from areas which gives an overall background SAR response similar to the Kandahar test area. Hence, part of the reason for the poor results for the Kandahar area is that the model could be biased. A lot of the ground truth mounds in the Kandahar area is also quite small. 52 of the 120 mounds are smaller than 0.4 ha. By evaluating each of the ground truth mound SAR RGB responses, it was found that 68 of the 120 ground truth mounds had a low likelihood of being detected, due to their size or lack of response. Doing the same evaluation of the Bukhara test area ground truth mounds it was found that only 5 had a low likelihood of being detected. This evaluation is subjective.

Figure 5.9 shows examples of mounds from the Kandahar test area considered to have a low probability of detection and a higher probability of detection. As can be seen the two mounds with low probability does not give any transition from yellow through red color on the left to a blue color on the right side of the mounds. This is very likely due to the low height of the settlement mounds creating less scattering effects than mounds with higher slope.

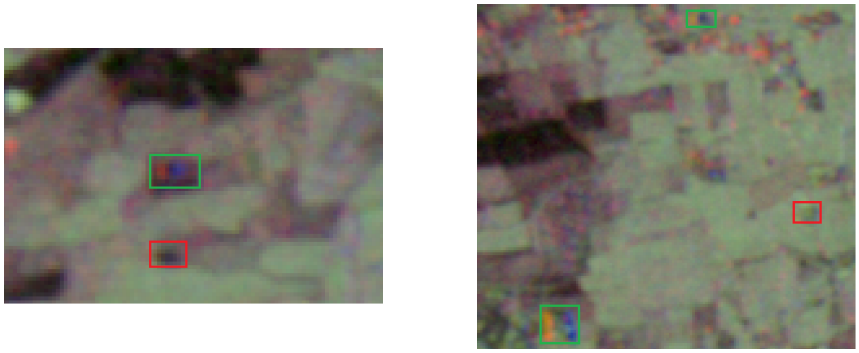


Figure 5.9: Ground Truth Mounds with Subjective Probability of Detection by Model. Red Box - Low Probability. Green Box - Higher Probability.

In conclusion, the model gives good results and detects potential archaeological settlement mounds well, dependant on type of area scanned. The model seems to give better results in fairly flat areas, where the slope of the mounds stand out from the background. Also, a brighter background SAR RGB response seems to give better results, although the testing done is limited. Areas with a darker background, like Kandahar seems to worsen results. The model is not adept at finding small mounds, with a small height above the background, giving a smaller slope on the mound sides. While precision numbers from the test areas might seem low, the amount of false positives are actually not prohibitive due to the large areas scanned. Counting both test areas, a total of 3091km2 was scanned and only 89 false positives was found, of which 28 could potentially be settlement mounds. A manual check of results for this large an area is not too time consuming.

Due to the limited amount of training data used, and the lack of training data from areas similar to the Kandahar test area, the model could likely be improved by adding more training data. The fact that additional augmentation by rotating the images is not possible, this also increases the need for additional training data.

Chapter 6

Conclusion

6.1 Conclusion and Answers to Research Questions

For burial mounds, one of the main research questions, as listed in section 1.3, was how effective a faster R-CNN object detection model was at detecting burial mounds. Another question was if the Multi-Scale Relief Model could improve results compared to other model, considering the relatively large number of smaller burial mounds in Norway compared to other places were this relief model has been applied.

The best result comes from using a Slope model for displaying the Digital Terrain Models. Up to 66% of the burial mounds were detected using this model. The number of mounds detected is however only part of the problem when using object detection for finding burial mounds. The number of false positives is the main problem, and although the Multi-Scale Relief Model and Simple Local Relief Model could give results approaching the recall values of the Slope model, the precision became very low, due to the very high number of false positives. The conclusion is that for finding burial mounds in Norway, displaying DTM data with a Slope display is the best method, as it does not remove details in the data, as the other three relief models do. This is important, as many of the burial mounds are small, and does not have much height above the background terrain. The number of false positives with a Slope model is still high, but not so high that it's use

becomes prohibitive in the detection of mounds.

The last research question for burial mounds, was if reducing the spatial resolution of the LiDAR DTM data from 0.25m to 1m would affect the detection of burial mounds. From the limited test done, it did not affect the detection of the actual burial mounds, but did affect the number of false positives, giving an 11% worse precision value.

Due to the mound sizes in the test area being fairly large, based on observation, the results could be worse than presented, in areas containing smaller mounds.

Due to the variability in mound signatures in the data, the models produced could be improved by adding more significantly more training data. The shapes and sizes of the burial mounds vary a lot.

For the detection of settlement mounds in Central Asia, the research question was if Synthetic Aperture Radar satellite data could be used to detect these mounds, and what the limitations are. Using the SAR data from the Sentinel-1 satellite, which has a 10m spatial resolution, it was from shown that it was possible to detect these mounds with a faster R-CNN object detection model.

One of the main limitations of using the SAR data, is that the height of the mounds, or rather the slope of the mound sides, seems to affect the amount of the scattering of the SAR signal. The overall background SAR signal signature might also have some effect on the results. Based on the limited testing, it is believed that the model is effective at finding potential settlement mounds given that the height and slope of the mound sides are not too small. The total area of the mounds seems to have less effect, as long as they are not so small that the resolution of the SAR data becomes an issue. Also, the model seems to work best in areas that are fairly flat and giving a brighter background SAR signal signature.

The model could likely be improved by adding more training data, particularly from areas with a different overall background SAR signature than the majority of training data used.

Detecting fortresses using optical imagery proved to be very challenging

and gave very poor results. The fortresses varied a lot in size and shape, and many of the fortress structures used were only partial remains. This variability and the lack of common features between the fortresses was the main problem. Ideally, several hundreds of images of fortresses of similar types should be used during training, and the lack of training data was the main reason for the poor results. To improve the results significantly, it is likely that several hundreds, if not thousands, more images of fortress structures are needed. Another possible approach may be using object segmentation rather than object detection due to the importance of pixel point precision to the model.

6.2 Improvements

One of the main improvements that could be done is to add more training data to the models. This applies to all the models, and in particular for training of the model for fortress detection.

Another issue, which was found after the training data had been created, was that for training the faster R-CNN object detection models, the features annotated in the images should ideally not take up more than 25% of each image. For burial mounds, many of the larger mounds take up much more than 50% of the 50x50m images. Increasing the image sizes to 100x100m or 150x150m, 400x400 pixels or 600x600 pixels respectively, could improve the training of the models.

A lot of the data gathering to create training data was manual work, and the possibility of improving this process through some automation using the arcpy python library, which contains a lot of the ArcGIS Pro functionality, could improve this process.

Bibliography

[1] F. Chen, R.Lasaponara, and N. Mancini. An overview of satellite synthetic aperture radar remote sensing in archaeology: From site detection to monitoring. *Journal of Cultural Heritage*, Volume 23, Supplement:5–11, March 2017.

[2] D. Tapete, A.Traviglia, E.Delpozzo, and F. Cigna. Regional-scale systematic mapping of archaeological mounds and detection of looting using cosmo-skymed high resolution dem and satellite imagery. *Remote Sensing*, 13, no.16, 2021.

[3] I. Berganzo-Besga, H.A. Orengo, F.Lumbreras, M.Carrero-Pazos, and J.Fonte. Hybrid msrm-based deep learning and multitemporal sentinel 2-based machine learning algorithm detects near 10k archaeological tumuli in north-western iberia. *Remote Sensing*, 13, no.20., 2021.

[4] Ø.D.Trier, J.H. Reksten, and K.Løseth. Automated mapping of cultural heritage in norway from airborne lidar data using faster r-cnn. *International Journal of applied Earth Observation and Geoinformation*, 95, March 2021.

[5] Mapbox. https://www.mapbox.com/ ,(Last accessed: 21st of February 2022).

[6] Hoydedata. https://hoydedata.no/LaserInnsyn/ ,(Last accessed: 21st of February 2022).

[7] H.A.Orengo, F.C.Conesa, A.Garcia-Molsosa, A.Lobo, A.S.Green, M.Madella, and C.A.Petrie. Automated detection of archaeological mounds using machine-learning classification of multisensor and multitemporal satellite data. *PNAS*, 117, Aug 2020.

[8] Great wall of gorgan. https://whc.unesco.org/en/tentativelists/6199/ ,(Last accessed: 27th of March 2022).

[9] Panchromatic imagery. https://eos.com/make-an-analysis/panchromatic/ ,(Last accessed: 27th of March 2022).

[10] Lidar. https://oceanservice.noaa.gov/facts/lidar.html ,(Last accessed: 28st of February 2022).

[11] H.A.Orengo and C.A.Petrie. Multi-scale relief model (msrm): a new algorithm for the visualization of subtle topographic change of variable size in digital elevation models. *Earth Surface Processes and Landforms*, 43:1361–1369, 2018.

[12] Earthdata nasa sar information. https://earthdata.nasa.gov/learn/backgrounders/what-is-sar ,(Last accessed: 5th of May 2022).

[13] A.Braun. Retrieval of digital elevation models from sentinel-1 radar data – open applications, techniques, and limitations. *Open Geosciences*, 13:532–569, 2021.

[14] Sentinel-1 preprocessing in google earth engine. https://developers.google.com/earth-engine/guides/sentinel1 ,(Last accessed: 5th of May 2022).

[15] A proposal for the dartmouth summer research project on artificial intelligence. http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html ,(Last accessed: 26th of March 2022).

[16] Deep learning. https://doi.org/10.1038/nature14539 ,(Last accessed: 26th of March 2022).

[17] Mcculloch-pitts neuron — mankind's first mathematical model of a biological neuron. https://towardsdatascience.com/mcculloch-pitts-model-5fdf65ac5dd1 ,(Last accessed: 26th of March 2022).

[18] relu, sigmoid and tanh - todays most used activation functions. https://github.com/christianversloot/machine-learning-articles/blob/main/relu-sigmoid-and-tanh-todays-most-used-activation-functions.md ,(Last accessed: 26th of March 2022).

[19] Supervised vs. unsupervised vs. reinforcement learning. https://www.educative.io/edpresso/supervised-vs-unsupervised-vs-reinforcement-learning ,(Last accessed: 26th of March 2022).

[20] Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1359523/ ,(Last accessed: 2th of April 2022).

[21] Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. https://doi.org/10.1007/BF00344251 ,(Last accessed: 2th of April 2022).

[22] Gis. https://www.nationalgeographic.org/encyclopedia/geographic-information-system-gis/ ,(Last accessed: 28st of February 2022).

[23] Google earth engine. https://earthengine.google.com/ ,(Last accessed: 28st of February 2022).

[24] W.B. Verschoof van der Vaart and K.Lambers. Learning to look at lidar: The use of r-cnn in the automated detection of archaeological objects in lidar data from the netherlands. *Journal of Computer Applications in Archaeology*, 2:31–40, March, 2019.

[25] W.B. Verschoof van der Vaart, K.Lambers, W. Kowalczyk, and Q.P.J. Bourgeois. Combining deep learning and location-based ranking for large-scale archaeological prospection of lidar data from the netherlands. *International Journal of Geo-Information*, 9, 2020.

[26] A.Guyot, L.Hubert-Moy, and T.Lorho. Detecting neolithic burial mounds from lidar-derived elevation data using a multi-scale approach and machine learning techniques. *Remote Sensing*, 10, 2018.

[27] T.Freeland, K.B.Heung, D.V.Burley, G.Clarck, and A.Knudby. Automated feature extraction for prospection and analysis of monumental earthworks from aerial lidar in the kingdom of tonga. *Journal of Archaeological Science*, 69:67–74, 2016.

[28] A. Monterroso-ChecaJosé Alejandro Conejo-Moreno, J.C. Moreno-Escribano, M. Gasparini, J.A. Conejo-Moreno, and J.L. Domínguez-Jiménez. Revealing archaeological sites under mediterranean forest canopy using lidar: El viandar castle (husum) in el hoyo (belmez-córdoba, spain). *Drones*, 5, 2021.

[29] D.Stott, S.M. Kristiansen, and S.M. Sindbæk. Searching for viking age fortresses with automatic landscape classification and feature detection. *Remote Sensing*, 11, 2019.

[30] C. Stewart, R.Montanaro, M.Sala, and P.Riccardi. Feature extraction in the north sinai desert using spaceborne synthetic aperture radar: Potential archaeological applications. *Remote Sensing*, 8, October 2016.

[31] Sigma nought - backscatter coefficient. https://sentinel.esa.int/web/sentinel/user-guides/sentinel-1-sar/definitions. ,(Last accessed: 27th of March 2022).

[32] D.Tapete, F.Cigna, N.Masini, and R.Lasaponara. Prospection and monitoring of the archaeological heritage of nasca, peru, with envisat asar. *Archaeological Prospection*, 20:133–147, May 2013.

[33] F.Chen, N.Lasaponara, and N.Masini. An overview of satellite synthetic aperture radar remote sensing in archaeology: From site detection to monitoring. *Journal of Cultural Heritage*, 23:5–11, June 2015.

[34] Cultural heritage database. https://kartkatalog.geonorge.no/metadata/kulturminner-enkeltminner/17150d2c-b50d-4792-80f4-0cb2ec5eaa79 ,(Last accessed: 27th of March 2022).

[35] N.Boroffka M.Karaucak, D.Steiniger. A remote sensing-based survey of archaeological/heritage sites near kandahar, afghanistan through publicly available satellite imagery. *PLoS ONE*, 16, Nov 2021.

[36] Multi-spectral and panchromatic images. https://www.stars-project.org/en/knowledgeportal/magazine/remote-sensing-technology/introduction/multispectral-and-panchromatic-images/ ,(Last accessed: 18th of April 2022).

[37] Gis geography: What is ndvi. https://gisgeography.com/ndvi-normalized-difference-vegetation-index/ ,(Last accessed: 18th of April 2022).

[38] A.Agapiou, D.G.Hadjimitsis, A.Georgopoulos, A.Sarris, and D.D. Alexakis. Towards an archaeological index: Identification of the spectral regions of stress vegetation due to buried archaeological remains. *Lecture Notes in Computer Science*, pages 129–138, 2012.

[39] S.Zanna and A.De Rosa. Remote sensing analyses on sentinel-2 images: Looking for roman roads in srem region (serbia). *Geosciences*, 9, Jan 2019.

[40] M.B.Rajani and K.Kasturirangan. Multispectral remote sensing data analysis and application for detecting moats around medieval settlements in south india. *Indian Society of Remote Sensing*, 42:651–657, Jan 2014.

[41] V.De Laet, G.van Loon, A.Van der Perre, and I.Deliever. Integrated remote sensing investigations of ancient quarries and road systems in the greater dayr al-barsha region, middle egypt: a study of logistics. *Journal of Archaeological Science*, 55, 2015.

Appendix A

Github Repository and Dataset

The data sets, results and code used for this thesis is available on Github, including Google Earth Engine Javascripts and Jupyter notebooks for faster R-CNN model training used in Google Colab.

Space Arachaeology Thesis Github Repository

IMPORTANT: The Github repository contains large files requiring LFS support. Install Git LFS support and clone repository using "git lfs clone" command.

Raw GeoTiff files used for generating data sets have been shared with Dr. Naeem Khademi through the thesis Teams channel, due to the large size of these raw files (450 GB).

Appendix B

Areas and Data for Training and Testing of Models

B.1 Norway Burial Mound Data

Table B.1 contains the project data set names downloaded from Hoydedata.no and the corresponding coordinates used for area selection. The project data sets can be found by selecting Projects instead of the standard height model, termed NDH. All data sets a created from 5 point per m2 LiDAR datat and has a spatial resolution of 0.25m. The data sets returned from hoydedata.no does not necessarily contain data for the entire area defined, but only what has been measured as part of the project within the area.

Table B.2 contains a breakdown of the number of images and burial mounds used for training, validation and testing of the faster R-CNN object detection models. The mound sizes are calculated from the bounding boxes added to the burial mound images and is an approximation. The mound sizes are taken as the largest of the width or height. The vast majority of mounds are considered to be round. Only 13 of the total of 595 mounds annotated have a height/width or width/height ratio larger than 1.5. Table B.3 contains an overview of the number of burial mounds taken from each data set. Table B.4 contains the coordinates for the 9km2 test area used

for the trained model finished models. These are the UTM coordinates, as listed in the ArcGIS Pro project (32V).

Images without burial mounds are sampled from various different areas, including populated areas, roads, roundabouts, hilly terrain, farms and farmland.

Table B.1: Burial Mound Data set Coordinates for Kartverket.no DTM Download

Project Dataset	Coordinate System	West (m)	East (m)	North (m)	South (m)
NDH Jæren-Randaberg-Sola 5pkt 2017	UTM 33N	−44,000	−36,000	6,545,400	6,531,800
NDH Vestfold 5pkt 2016	UTM 33N	238,400	243,200	6,593,400	6,588,000
NDH Østfold 5 pkt 2015	UTM 33N	274,400	280,000	6,583,800	6,576,000
NDH Larvik 5pkt 2017	UTM 33N	206,400	220,800	6,563,400	6,547,800
NDH Steinkjer 5pkt 2017	UTM 33N	317,022	329,462	7,076,270	7,092,983
NDH Rogaland 5pkt 2016	UTM 33N	−49,789	−-34,895	6,559,168	6,549,733

Table B.2: Breakdown of Training, Validation and Testing Data

Type	# Images	# Images with Mounds	# Mounds	Average Mound Size (m)	Less than 10m	Between 10 and 20m	Above 20m
Training	304	130	439	9.7	284	129	26
Validation	68	28	77	9.4	44	32	1
Testing	46	33	79	13.8	37	29	13
Total	418	191	595	10.2	365	190	40

Table B.3: Breakdown of Mounds from Project Datasets

Project Data set	Mounds	Training	Valdiation	Testing
NDH Jæren-Randaberg-Sola 5pkt 2017	239	207	32	0
NDH Larvik 5pkt 2017	239	201	38	0
NDH Steinkjer 5pkt 2017	38	31	7	0
NDH Vestfold 5pkt 2016	41	0	0	41
NDH Østfold 5 pkt 2015	38	0	0	38

Table B.4: Coordinates of Test Area from Rogaland Dataset

Project Data set	East UTM 32V (m)	West UTM 32V (m)	North UTM 32V (m)	South UTM 32V (m)	Area Size (km2)
NDH Rogaland 5pkt 2016	304,345	307,345	6,5202,26	6,517,226	9

B.2 Central Asia Archaeological Mound SAR Data

Table B.5 contains an overview of the SAR data sets used for training, validation and testing of the SAR faster R-CNN object detection model. The number of features annotated, which gives an expected SAR response to archaeological settlement mounds, are listed. Not all features annotated are necessarily archaeological settlement mounds, but features giving the expected SAR response. Table B.6 shows the coordinates in decimal degrees of the areas used for generating the training data and the test areas.

B.2 Central Asia Archaeological Mound SAR Data

Table B.5: Central Asia Archaeological Mound SAR Datasets

Area	Country	Number of Features	Number of Images	Model Use	Area Size
Masar-i-Sharif	Afghanistan	300	308	Training	6182km2
Farah	Afghanistan	28	168	Training	3372km2
Sheberghan	Afghanistan	27	60	Testing	1204km2
Bukhara	Uzbekistan	43	88	Testing	1766km2
Merv	Turkmenistan	52	173	Training	3472km2
Bojnurd	Iran	30	120	Training	2409km2
Kandahar	Afghanistan	120	Composite Image	Test of Final Model	1967km2
Bukhara	Uzbekistan	82	Composite Image	Test of Final Model	1124km2

Table B.6: SAR Datasets - Coordinates in Decimal Degrees

Area	Coordinate System	West (dd)	East (dd)	North (dd)	South (dd)
Masar-i-Sharif	WGS 1984	66.1766	67.223858	37.173785	36.6906
Farah	WGS 1984	61.853258	62.4166	32.5763	32.093392
Sheberghan	WGS 1984	65.5912	65.9533	36.79165	36.51
Bukhara	WGS 1984	63.69035	64.1329	40.046707	39.7248
Merv	WGS 1984	61.8631	62.4666	37.9745	37.4917
Bojnurd	WGS 1984	56.6845	57.288	37.7619	37.442
Kandahar Composite	WGS 1984	65.188	65.7513	31.6807	31.399
Bukhara Composite	WGS 1984	64.3206	64.64245	402453	39.9637

Appendix C

Multi-Spectral Satellite Imagery Uses in Archaeology

C.1 Multi-Spectral Satellite Imagery Technical Information

Multi-spectral satellite imagery sensors measures light intensity within a specific wavelength range in the electromagnetic spectrum. Typically satellites with multi-spectral sensors measures 15 or less bands, although more advanced satellite sensors are being deployed. Hyper-spectral sensors can provide more than 100 different bands.[36] In addition to the visible light in the red, green and blue bands, visible and near-infrared and short wave infrared spectral bands can be available from some satellite sensors. In addition, some satellites have panchromatic sensors, which measures the total light intensity of the red, green and blue bands. The exact wavelength range in which the satellite sensors measure bands like near-infrared and other bands might vary somewhat, and is not standardized. Often the wavelengths are chosen for the main task the satellite is built for. Hence, in terms of using the bands in for example a Normalized Difference Vegetation Index, might produce slightly different results, based on the satellite data used. The spatial resolution of different satellite multi-spectral sensor vary, from very high resolution below 1m up to 10-60m, dependant on the bands.

Pan-sharpening, also called intensity substitution, is a process which uses both the panchromatic and multi-spectral bands to enhance the resolution of an image. By combining, or fusing, 3 bands of the multi-spectral data with the higher resolution panchromatic data, a RGB composite of the data with enhanced spatial resolution can be created.

One of the main sources of free multi-spectral satellite imagery is the Sentinel-2 satellites, which contains 13 spectral bands. The satellite is part of the Copernicus program, and whose one of the main tasks is land monitoring and can used for crop monitoring. A list of some freely available multi-spectral satellite imagery and satellite technical specifications can be found in table D.1 and appendix E respectively.

For archaeology, multi-spectral data can have many uses, as some of the articles listed in table C.1 shows. One use is to identify so-called "crop marks", which are areas where vegetation is affected by possible archaeological remains right beneath the surface, which affects the growth of vegetation. These crop marks are not as pronounced in the visible spectrum as it is in the near-infrared spectrum. Healthy vegetation (chlorophyll) reflects more in the near-infrared and green spectrum and is absorbed more in the red and blue spectrum.[37] The Normalized Difference Vegetation Index, or variants of this index, can be used to identify crop marks, which might become more pronounced by studying a time series of the index over a growth period.

$$NDVI = \frac{NIR - Red}{NIR + Red} \tag{C.1}$$

Attempts have also been made to create a archaeological index, based on multi-spectral data. A.Agapiou Et. al. studies conclude that crop marks are more pronounced in the red edge and near-infra-red in the 700nm to 800nm spectrum.[38]

S.Sanni Et al. used a Normalized Archaeological Index, NAI, based on Agapiou's research, in an attempt to identify identify crop marks.[39]

$$NDVI = \frac{800nm - 700nm}{800nm + 700nm} \tag{C.2}$$

There is no universal index or use of multi-spectral data that fits into all archaeological work applying the data. Multi-spectral data might be used different ways depending on the specific use case. The Sentinel-Hub Custom Scripts web page and Remote Sensing Indices database, which is linked in table D.2, contains a large set of indices that can be used with multi-spectral data. The indices are not specific to archaeological uses.

C.2 Use of Multi-Spectral Imagery in Archaeology

S.Zanni Et al. used Sentinel-2 multi-spectral satellite data to help discover the path of a Roman road that ran from what is today Belgrade to northern Italy.[39] Parts of the path of the road was known, but several parts were not. Their research for was focused on the Srem region of Serbia, where the path of the Roman road was completely different than the modern road network, and much of the road is buried beneath farm fields. To develop a methodology for detecting the buried road, they focused on a 10km2 area using 4 acquisitions from the Sentinel-2 satellite between June and August in 2016. By using combinations of bands they attempted to identify crop marks, which are areas of poorer or different crop or vegetation growth. These crop marks can be an indication of potential archaeological features right beneath the soil, which can affect the growth of the crops. Their first attempts used the Normalized Difference Vegetation Index, NDVI, and the and the Normalized Archaeological Index, NAI. The NDVI uses the Near-Infrared (B8) and Red (B8) bands, while the NAI uses the Vegetation Red Edge bands B5 and B7. Neither approach worked satisfactory, and crop marks did not become visible. The NAI uses 20m spatial resolution bands, and the Roman road was expected to be 10-15m wide. By adding the Red (B4) and NIR (B8) bands, they same data gave a better results, and crop marks started to become visible in some areas. In addition, the Tasseled Cap transformation, a set of equations known to be good to detect archaeological remains, was also used on the data. Specifically, the Crop Coefficient 3 equation was used, and modified to fit the Sentinel-2 data, in addition to using variants of this formula. The detection of crop marks did not improve much with the Crop Coefficient 3 equations applied to the Sentinel-2 data. The study concludes that the Sentinel-2 multi-spectral data has too low resolution to be applied effectively to it's specific use case. By using higher resolution WorldView-2 data, with a spatial resolution close to 2m and

applying the Crop Coefficient 3 equations, they were able to identify crop marks and map a 70km stretch of the Roman road, where the path of the road was previously unknown.

M.B. Rajani Et al. used multi-spectral data to analyse three well know settlements in South India which had fortress walls and moats as defensive structures around them during the Hoysala period from the 12th to the 14th century AD.[40] The study used 5.8m spatial resolution multi-spectral data from the IRS-P6 satellite. Their first area studied was the Belur settlement. By looking at the individual bands and combination of bands they found that the moat around the settlement most prominent when combining the Red and NIR bands, by projecting the Red band as green and blue and the NIR band as red in an RGB composite. Looking at the individual bands the moat was less prominent. The second area studied was the Halebidu settlement. In this area parts of the moat was clearly visible and well defined by using an RGB composite of Red and NIR bands. However, other parts of the moat could not be clearly defined without using complementary data like Normalized Difference Vegetation Index and panchromatic imagery. The third site studied was Somanathapura. According to the authors, it is very difficult to identify the placement of the ancient defensive walls and moat around this settlement by traditional archaeological methods. By using the Red and NIR bands in an RGB composite, as with the Belur site, parts of the placement of the walls and moat could be found, as the walls became visible in a square around the settlement's central temple, as expected from the typical settlement layouts of the time period.

V. De Laet Et. al. performed a study of ancient road systems and quarries in the Greater Dayr al-Barsha region of Egypt, using very high resolution imagery and multi-spectral data.[41] Their aim was to study the logistics of the area in ancient Egypt, and create a road typology for the area, requiring very high resolution imagery. The study used Quickbird-2 satellite data with a spatial resolution of 0.61m. The panchromatic and multi-spectral data was fused using Gram-Schmidt spectral sharpening. This fused data allowed many more roads to be discerned than the regular multi-spectral data. However, the fused data did not did not show as many roads as the panchromatic alone could, and the authors conclusion is that the multi-spectral data, with lower spatial resolution, is of less importance when it comes to identifying the roads. For their study area they found that using Gaussian stretching and histogram equalization on the images proved most

valuable to enhance the detection of the road systems. The histogram equalization reduces contrast in the very light and very dark areas of the images, while the Gaussian stretching redistributes the pixels between the minimum and maximum values with a Gaussian distribution. The roads where put into 7 categories, identified by their spectral response in the satellite images, their width and other characteristics like the slope. The road categories were used to assess the usage of the road network for the area, identifying the likely use of them, either for transport from quarries, walking paths or other uses. With this information they study was able to put together likely transport routes and road network usage through different time periods.

Table C.1: Overview of articles on multi-spectral uses in archaeology

Reference	Satellite	Type of ML/DL	Description
[39]	Sentinel-2	N/A	Mapping of an ancient Roman road in Srem region of Serbia. Sentinel-2 multispectral data was used to identify crop marks which indicated the placement of the road beneath the soil.
[40]	IRS-P6	N/A	Analysis of three ancient settlements in South India and their surrounding moats and defensive walls using multispectral imagery.
[41]	Quickbird-2	N/A	Mapping and classification of ancient road systems connected to limestone quarries in Eqypt, using Quickbird-2 very high resolution imagery. Road were categorized by their spectral signatures, width, slope and other parameters.

Appendix D

Sources of Free Remote Sensing Data

There are many sources of free remote sensing data available. The main sources found during this thesis are listed here.

Google Earth Engine is a free service which offers the ability to download and produce data from multiple sources, upon registration. It offers some GIS capability and has scripting capability in JavaScript and Python, which allows for assembling multiple satellite data acquisitions and chosen bands from different satellites into products which can be exported to Google Drive or Google Cloud for download. It also offers machine learning options, like a Random Forrest algorithm, through APIs. The advantage of the website is that the satellite data, like the Sentinel-1 SAR data is already processed with standard parameters, and users don't have to go through this processing, which can be time consuming. Many of the data sets available are finished products of different types.

The Copernicus Services Data Hub is a free service offering the data acquired from the Copernicus program, upon registration. The data hub has an Open Data Protocol (OData) based on top of HTTPS/REST transfer protocols which allows for batch scripting of data of product download. The hub gives access to Sentinel satellite data.

The European Space Agency, ESA, offers free satellite data, which is part of the ESA charter. The data should be free for all. However, apart from the Sentinel mission data, the majority of data available is 3rd party data, and they offer only limited global coverage for much of this data. An approval process is also necessary to get access to a lot of the data, which might vary for the different 3rd party data catalogues. Sentinel data can be found through the Copernicus Services Data Hub.

The US Geographical Survey Earth explorer website contains a large selection of both satellite data and other data like declassified US government aerial imagery and Digital Elevation data, which are free upon registration. Some of the data is 3rd party data which has limited global coverage, and some data has limited access.

The NASA Earthdata website has over 33,000 data collections available, which are all open data sets and freely available, upon registration This includes a large collections of earth observation satellite data, but also data sets collected from air-based platforms.

OpenTopography is a website that offers high resolution LiDAR point clouds. Their data sets are limited in extent, and covers mostly parts of the United States. The point clouds can be used to produce high resolution Digital Elevation Maps, DEMs.

Hoydedata.no is a web page run by the Norwegian Mapping Service, which allows free access to 1m spatial resolution LiDAR based Digital Elevation Models for all of Norway. The website also contains a large number of project data, which can offer as good as 0.25m spatial resolution DEMs, all delivered as GeoTiff files. Raw point cloud data used to create the DEMs can also be requested.

Several European countries have had or are in the process of making high resolution map data based on LiDAR data. This includes The Netherlands which has 0.5m spatial resolution data, which is freely available through Google Earth Engine, Sweden and the United Kingdom, which should have least 1m spatial resolution data available. For a specific European country it is advised to search for Digital Elevation data through government websites. The data is very often freely available through these, if a country has mapped or is mapping their a large portion of their country.

For non-European countries, there seems to be limited amounts of free LiDAR data available, as the majority is acquired for specific projects and not made available through government or free data sources.

Some other useful remote sensing resources are shown in table D.2

Table D.1: Sources of Free Remote Sensing Data

Name	Link	Data Available
Google Earth Engine	GEE	Collection of DEMs, ASTER multi-spectral, Sentinel-1 SAR, Sentinel-2 multi-spectral, Sentinel-3 and 5 data, EO-1 Hyperspectral data, Landsat program collections, MODIS mulsti-spectral data and more.
Copernicus Services Data Hub	Copernicus Data Service Hub	Sentinel-1 SCopernicusAR data, Sentinel-2 multi-spectral data, Sentinel-3 and Sentinel-5P data
ESA	European Space Agency	Cosmo-SkyMED, TerraSAR-X, PALSAR and Radarsat-2 SAR data. KOMPSAT-3 multi-spectral data. Contains a large selection of 3rd party data, but most data needs approval.
US Geological Survey - Earth Explorer	EarthExplorer	Landsat collection, OrbView-3, IKONOS-2, Sentinel-2, ISRO Resourcesat, ASTER, MODIS, SIR and many other satellite data sets. Also contains aerial photography, declassified US goverment data and elevation data, like DEMs.
NASA - Earthdata	Earthdata	Sentinel, ASTER, MODIS and many other satellite datasets. Also has datasets from air-based plattforms.
Hoydedata.no	Hoydedata.no	LiDAR based DTMs and point cloud data with 1m or better spatial resolution for all of Norway.
OpenTopography	OpenTopography	LiDAR high resolution point cloud. Mostly covering parts of the US.

Table D.2: Useful Resources

Name	Link	Description
Sentinel-Hub Custom Scripts	Custom Scripts	Collection of scripts calculating different indices like NDVI and others from various common satellite data sources, which can be used in EO Browser. Also contains other useful information.
University of Twente Satellite and Sensor Database	Satellite and Sensor Database	Database of existing satellites and satellite sensors database. Can search for senor type or satellite.
Remote Sensing Indices Database	RS Indices Database	Database of remote sensing indices. Can search by sensor or application or application for a specific index.

Appendix E

Free Satellite Data Technical Specifications

Table E.1 contains an overview of technical information of some SAR satellite data. SIR-A and B, and also SRTM, are SAR sensors used by NASA on space shuttle missions. Newer SAR satellites with higher resolution, like Cosmo-SkyMed and TerraSAR-X are not listed, as the data freely available is limited. RADARSAT-1 and 2 information is not listed for the same reason.

Tables E.2,E.3,E.4,E.5 and E.6 shows the technical details of multi-spectral satellite sensors, for which the data is freely available.

The tables uses the following abbreviations:

- VIS - Visible
- NIR - Near-Infrared
- SWIR - Short-wave Infrared
- TIR - Thermal
- MIR - Mid-Infrared

- PAN - Panchromatic

The University of Twente Satellite and Sensor database, which has a link in table D.2, can be used to find technical details and information links for a large number of satellites and sensors.

Table E.1: SAR Satellite Technical Information

Satellite Sensor	Band	Polarization	Resolution	Swath Width	Organisation
Sentinel-1 Strip Map Mode	C	Dual HH+HV, VV+VH, Single HH, VV	5m	80km	ESA
Sentinel-1 Interferometric Wide Mode	C	Dual HH+HV, VV+VH, Single HH, VV	5x20m	250km	ESA
Sentinel-1 Extra Wide Mode	C	Dual HH+HV, VV+VH, Single HH, VV	20x40m	410km	ESA
Sentinel-1 Wave Mode	C	Single HH, VV	5m	60km	ESA
ENVISAT ASAR	C	All	30m	20 km	ESA
ALOS PALSAR-1	L	HH, VV, HV, VH	10-100m	70-250km	JAXA
ALOS PALSAR-2 Spotlight Mode	L	H, VV, HV	1-3m	25km	JAXA
ALOS PALSAR-2 Strip Map Mode	L	HH, VV, HV, (HH + HV), (VV+VH), (HH + HV+VV+VH)	3,6,9m	50-70km	JAXA
ALOS PALSAR-2 ScanSAR Mode	L	HH, VV, HV, (HH + HV), (VV+VH)	100m	350-490km	JAXA
SIR-A and B	L	HH	30m	60km	NIMA/NASA
SRTM	C	HH	30m	50km	NASA

Table E.2: Sentinel-2 Satellite Multi-spectral Imager Technical Information - ESA

Satellite Sensor	Band	Resolution	Wavelength (μm)	Bandwidth (μm)
MSI	1 (VIS)	60m	0.443	0.02
MSI	2-Blue (VIS)	10m	0.49	0.065
MSI	3-Green (VIS)	10m	0.56	0.035
MSI	4-Red (VIS)	10m	0.665	0.03
MSI	5 (VIS)	20m	0.705	0.015
MSI	6 (VIS)	20m	0.74	0.015
MSI	7 (VIS)	20m	0.7753	0.02
MSI	8 (NIR)	10m	0.842	0.115
MSI	8A (NIR)	20m	0.865	0.02
MSI	9 (NIR)	60m	0.943	0.02
MSI	10 (SWIR)	60m	1.375	0.02
MSI	11 (SWIR)	20m	1.61	0.09
MSI	12 (SWIR)	20m	2.193	0.18

Table E.3: Landsat-7 Satellite Enhanced Thematic Mapper Technical Information - NASA

Satellite Sensor	Band	Resolution	Wavelength (μm)
ETM	1 (VIS)	30m	0.45-0.515
ETM	2 (VIS)	30m	0.525-0.605
ETM	3 (VIS)	30m	0.63-0.69
ETM	4 (NIR)	30m	0.75-0.9
ETM	5 (SWIR)	30m	1.55-1.75
ETM	6 (TIR)	60m	10.4-12.5
ETM	7 (MWIR)	30m	2.08-2.35
ETM	PAN	15m	0.52-0.9

Table E.4: Landsat-8 and 9 Satellite Operational Land Imager Technical Information - NASA

Satellite Sensor	Band	Resolution	Wavelength (μm)
OLI	1-Coastal/Aerosol	30m	0.43-0.45
OLI	2-Blue (VIS)	30m	0.45-0.51
OLI	3-Green (VIS)	30m	0.53-0.59
OLI	4-Red (VIS)	30m	0.64-0.67
OLI	5 (NIR)	30m	0.55-0.88
OLI	6 (SWIR)	30m	1.57-1.65
OLI	7 (SWIR)	30m	2.11-2.29
OLI	8 (PAN)	15m	0.5-0.68
OLI	8-Cirrus	30m	1.36-1.38

Table E.5: Terra Satellite Advanced Spaceborne Thermal Emission and Reflection Radiometer Technical Information - NASA

Satellite Sensor	Band	Resolution	Wavelength (μm)
ASTER	1 (VIS)	15m	0.52-0.6
ASTER	2 (VIS)	15m	0.63-0.69
ASTER	3n (NIR)	15m	0.76-0.86
ASTER	3b (NIR)	15m	0.76-0.86
ASTER	4 (SWIR)	30m	1.6-1.7
ASTER	5 (SWIR)	30m	2.145-2.185
ASTER	6 (SWIR)	30m	2.185-2.225
ASTER	7 (SWIR)	30m	2.235-2.285
ASTER	8 (SWIR)	30m	2.295-2.365
ASTER	9 (SWIR)	30m	2.36-2.43
ASTER	10 (TIR)	90m	8.125-8.475
ASTER	11 (TIR)	90m	8.475-8.825
ASTER	12 (TIR)	90m	8.925-9.275
ASTER	13 (TIR)	90m	10.25-10.95
ASTER	14 (TIR)	90m	10.95-11.65

Table E.6: Orbview-3 Technical Information - Orbital Imaging Corporation

Satellite Sensor	Band	Resolution	Wavelength (μm)
Orbview-3	1 (VIS)	4m	0.45-0.52
Orbview-3	2 (VIS)	4m	0.52-0.6
Orbview-3	3 (VIS)	4m	0.625-0.695
Orbview-3	4 (NIR)	4m	0.76-0.9
Orbview-3	PAN	1m	0.45-0.9