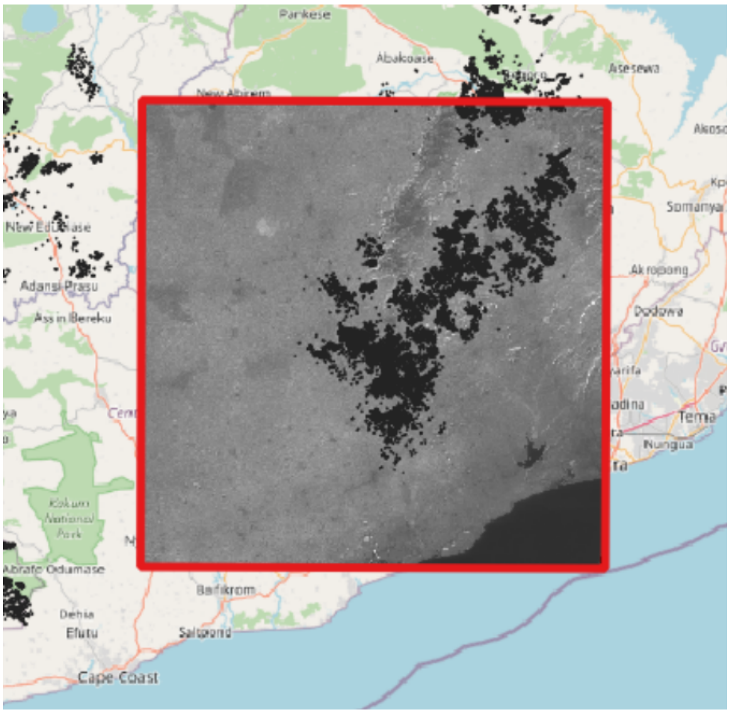
P2 Submission
Integrating radar and multi-spectral data to detect cocoa crops: a deep learning approach

Adele Therias
student #5485193
1st supervisor: Dr. Azarakhsh Rafiee
2nd supervisor: Dr. Stef Lhermitte
External supervisors: Philip van der Lugt
Thomas Vaassen (Meridia Land B.V.)

January 20, 2022



Contents

1 Introduction

1.1 Deforestation

Forests play a key role in the functioning of ecosystems at a local and global scales, providing important services such as habitats for over 80% of terrestrial biodiversity and the sequestration of around 289 Gt carbon [20]. However, forests are also under significant threat due to human activities such as commodity extraction, urbanization, and agricultural intensification, with severe and long-term impacts on ecological and human well-being including ”temperature rise, habitats depletion, extreme weather events, soil degradation, infectious diseases, rising GHGs, and environmental pollution” [20, p.21]. One of the major contributing factors to deforestation is agricultural production for export to the European Union [10]. In response, the European Union approved a law in December 2022 that aims to reduce the impact of European consumption on global deforestation by banning the import of products that are issued from deforested areas, with a particular focus on cattle, wood, palm oil, soy, cocoa and coffee [10]. Due to the due diligence requirements for companies that produce such commodities and derived products [10], the enforcement of this law will require highly accurate and timely tracking of farm extents using geodata and advanced geospatial analysis.

1.2 Cocoa production

Around 16% of the world's forested areas are located in Africa [15, p. 14], and this region faces particularly high rates of deforestation: ”the highest annual deforestation rate in 2015–2020 was in Africa (4.41 million ha)” of which 1.90 million hectares were located in Central and Western Africa” [15, p. 19]. The focus of this research is the detection of cocoa crops, of which West Africa is one of the main producing regions [4] and which are estimated to cause 7.54 % of EU-driven deforestation [10, p. 27]. The research in this thesis has emerged from a need identified by Meridia Land B.V., a company that works to improve data transparency and traceability in smallholder supply chains [21]. According to recent conversations with representatives from Meridia, the company is working with several clients who are seeking to develop more accurate ways of systematically tracking the cocoa farm extents of smallholder farmers in Ghana, where ”over a quarter of agricultural conversion stems from cocoa expansion” [6, p. 1]

1.3 Cocoa detection challenges

While the classification of multispectral satellite imagery is frequently applied to map crops and farm extents [4], cocoa presents unique challenges. First, West Africa has frequent cloud cover due to the Monsoon climate, which limits the availability of cloud-free multispectral datasets and the temporal resolution of those datasets [5]. Second, agroforestry land cover, a common practice which integrates shade trees to improve cocoa growing conditions, has a spectral signature and canopy structure similar to nearby forest [5, p. 2], and the canopy structure of cocoa can vary widely, as shown in figure 1. Researchers have addressed these challenges by using machine learning algorithms trained with Synthetic Aperture Radar (SAR) and/or multispectral datasets to identify cocoa crops. While many of these implementations use a pixel-based classification that does not consider the spatial context, recent work has applied a Convolutional Neural Network (CNN) trained with multispectral data and shows promising results in Ghana and Cote d'Ivoire [18]. The objective of this thesis is to build on this deep learning approach by using SAR data in the training of a CNN in order to test the impact of inputs on the accuracy of cocoa detection.

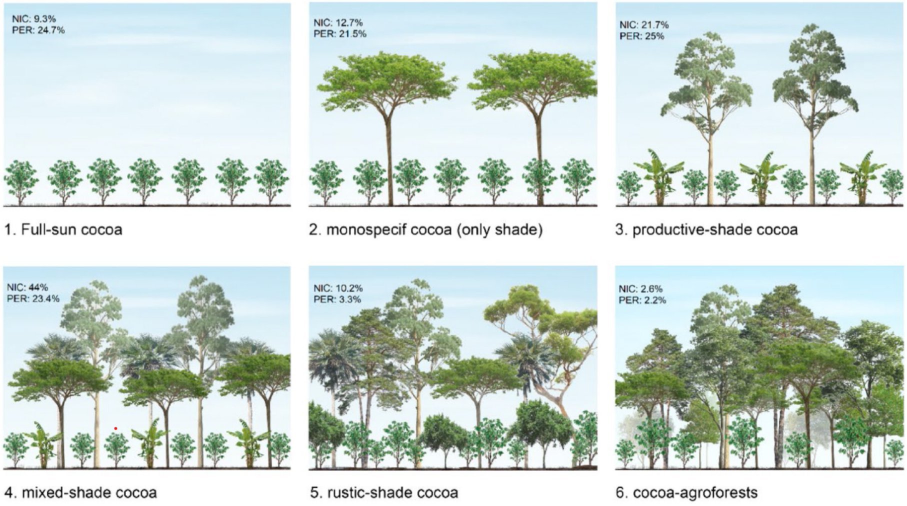


Figure 1: "Proportion of shaded cocoa typologies in Nicaragua (NIC) and Peru (PER) as expressed by in-country experts" [25, p. 5]

2 Related work

This section highlights methods that have been used to detect cocoa crops from remotely sensed data. In particular, it emphasizes two main approaches and two types of satellite data used as input. Pixel-based classification aims to assign a class to individual pixels based on their spectral attributes ([31]). In contrast, object-based classification (or segmentation) aims to identify features in an image (e.g. land parcels) before classifying them and therefore adding additional spatial context [17]. These methods have been implemented for cocoa detection based on multispectral satellite imagery (e.g. optical, Near Infrared (NIR), and Shortwave Infrared (SWIR) bands), Synthetic Aperture Radar (SAR, captured in the microwave bands), or a combination of the two data types.

2.1 Pixel-based classification of cocoa

The majority of existing cocoa detection studies focus on pixel-based classification by implementing a variety of machine learning algorithms and different combinations of datasets. Some researchers have considered only multispectral imagery (such as Landsat or Sentinel 2 data), which are classified using Maximum Likelihood Algorithm (Overall Accuracy [OA] = 82.6 %) [30], Random Forest (OA = 89.8 %) [6] and XGBoost, a type of boosted Random Forest (OA = 95.17 %) [7].

The effectiveness of multispectral data is limited considering that frequent cloud cover often impedes the collection of passively sensed data in moist tropical regions, and that agroforest cocoa parcels have spectral signatures very similar to surrounding jungle [23]. For this reason, some researchers have focused on classification methods that are based on SAR data only, which can "capture the water content (a dielectric property) and structure (a geometric property)" of land cover [23, p. 2]. SAR-based classification has been implemented using Supervised Maximum-likelihood Classifier (OA = 89 %) [27], Random Forest combined with Grey Level Co-occurrence Matrix (GLCM) (OA = 88.1 %) [23], and Multi-Layer Perceptron (MLP) Neural Networks Regression (Root Mean Square Error [RMSE] = 7.18 %) [22].

Two recent articles perform classification on a combination of SAR and multispectral data: the penetration of SAR can improve the detection of shrubs grown under a tree canopy, and differentiate between different types of trees [4]. In a 2021 study to measure the encroach-

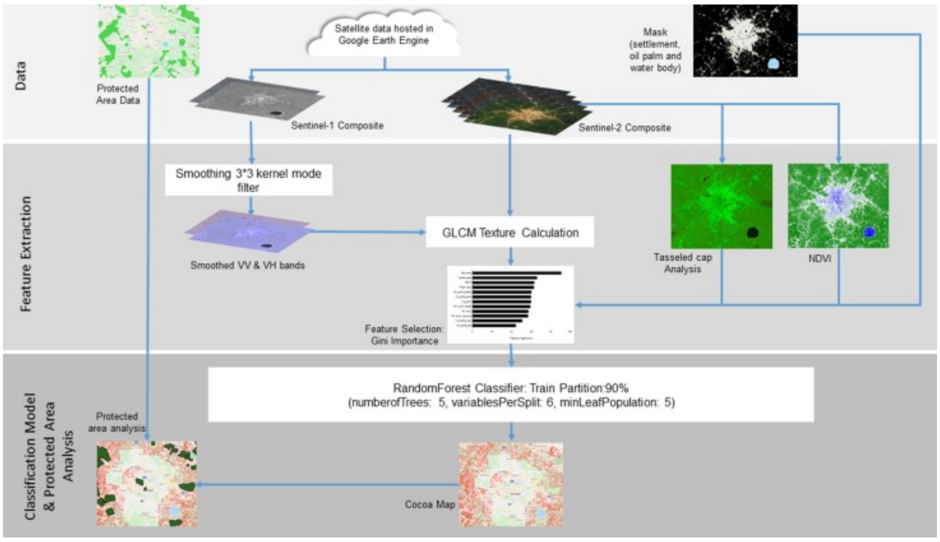


Figure 2: Classifier using SAR and Multispectral datasets (Abu et al., 2021)

ment of cocoa farms on protected forest reserves, Abu et al. created composite images of each Sentinel 1 (SAR) and Sentinel 2 (multispectral) datasets, and performed textural (Grey-Level Co-occurence Matrix [GLCM]) and spectral (Normalized Difference Vegetation Index [NDVI], Tasseled Cap index) analyses before applying a multi-feature Random Forest classifier [4], see figure 2. The producer's accuracy and user's accuracy were respectively evaluated at 82.9% and 62.2% [4]. In a 2022 paper, Tamga et al. explore the spatial distribution of classification errors with a focus on cocoa mapping in Ghana and Côte d'Ivoire [28]. The authors use Sentinel 1 datasets acquired with Interferometric Wide swath [IW] mode and with Vertical-Vertical [VV] and Vertical-Horizontal [VH] polarisations as well as red, green, blue, near-infrared and red-edge bands from Sentinel 2 multispectral data to perform textural (GLCM) and spectral analyses before applying a multi-feature Random Forest classifier [28]. The main differences between the two studies is that Tamga et al. applied many more different vegetation indices for spectral analysis (NDVI, GLI, EVI, SAVI, MSAVI, TCARI, VARI) and they calculated Shannon entropy per pixel to remove pixels with a high probability of error [28]. The higher producer's accuracy (88%) and user's accuracy (91%) compared to the work by Abu et al. is attributed to the fact that this study area is considerably smaller and focused only on a cocoa producing region [28].

As can be seen in figure 3, the results of pixel-based classifications can lead to a "salt and pepper" effect, even when smoothing filters are applied [28]. This effect does not necessarily reflect the reality of cocoa parcels on the ground and may be improved by employing a method that detects parcels of cocoa, rather than focusing on pixel-level classification. In both articles, the authors conclude that the classification output could be improved with the use of deep learning, with one paper specifically suggesting the use of "semantic image segmentation" [4].

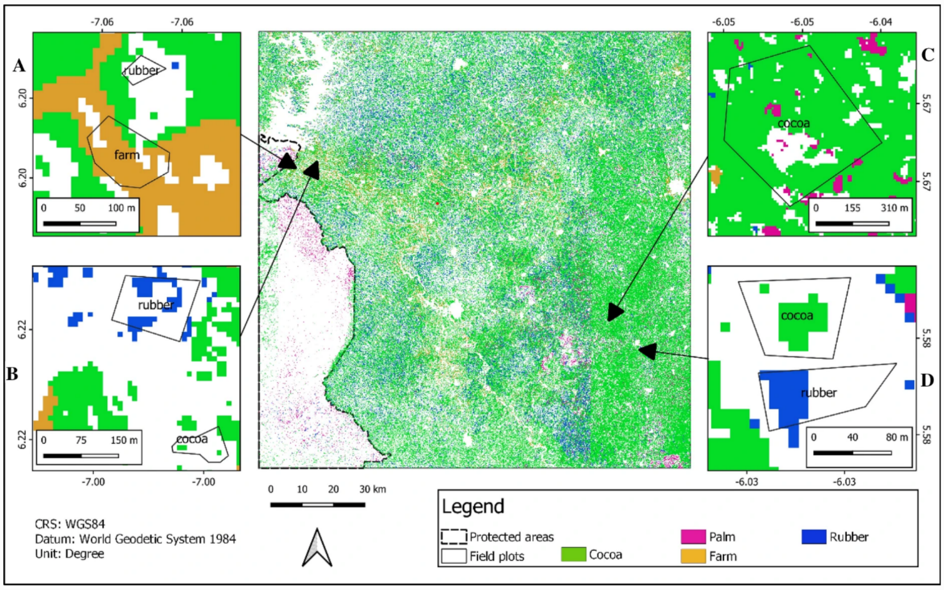


Figure 3: "Improved classification of agroforestry systems in southern Côte d'Ivoire using entropy value threshold derived from field plot" (Tamga et al., 2022, p. 9)

2.2 Object-based classification of cocoa

Beyond pixel-based classification, cocoa farm detection may be improved with methods that take into account the spatial dimension of cocoa production; in other words, detecting and classifying parcels rather than individual pixels. This approach can be implemented via algorithms that first detect image objects and then classify them: "[c]ombining spectral and backscatter pixels and image objects implies harnessing the spectral and textural capabilities of a satellite dataset and spatial patterns of the landscape and incorporating into the classification process" [5, p. 3]. One early example of such an approach combines optical (Landsat-ETM+) and dual-polarimetric radar (Envisat-ASAR) satellite data to map rice and cocoa parcels in Indonesia [11]. The authors detect image objects from a panchromatic dataset, then use a time series of co- and cross-polarized SAR datasets at a resolution of 15 meters and multispectral datasets at resolution of 30 meters for classification of each object [11]. This study makes use of an object-based nearest neighbour classifier and applies it to different combinations of datasets, and the highest OA (89%) obtained by using multispectral *and* cross-polarized SAR data [11].

Another more recent example of object-based classification is a study from 2020 which aims to detect and differentiate between open forest and agroforestry cocoa [5]. The authors first detect image objects from combined SAR and multispectral datasets using the Multiresolution Segmentation algorithm, and then apply Random Forest classification to three experimental datasets: multispectral only (OA = 79.02%), multispectral and SAR (OA = 80.49%), and finally multispectral, SAR and image objects (OA = 89.76) [5]. As shown in figure 4, not only is the accuracy of the object-based classification higher, but the visual output is also a more realistic representation of the spatial characteristics of cocoa parcels on the ground. One significant limitation of the detecting objects prior to classification is that "image objects created from the image segmentation process depends on the segmentation scale, the spatial resolution of the dataset used, and class definitions," "different datasets and mapping extent or scale will require different levels of image objects," and incorrectly defined image objects will cause the mis-classification of all pixels in that object [5, p. 11]. Deep learning, specifically Convolutional

Neural Networks, may offer an alternative to these challenges, and they have been used to detect cocoa parcels without relying on supervised classification or object detection, as described in the following section.

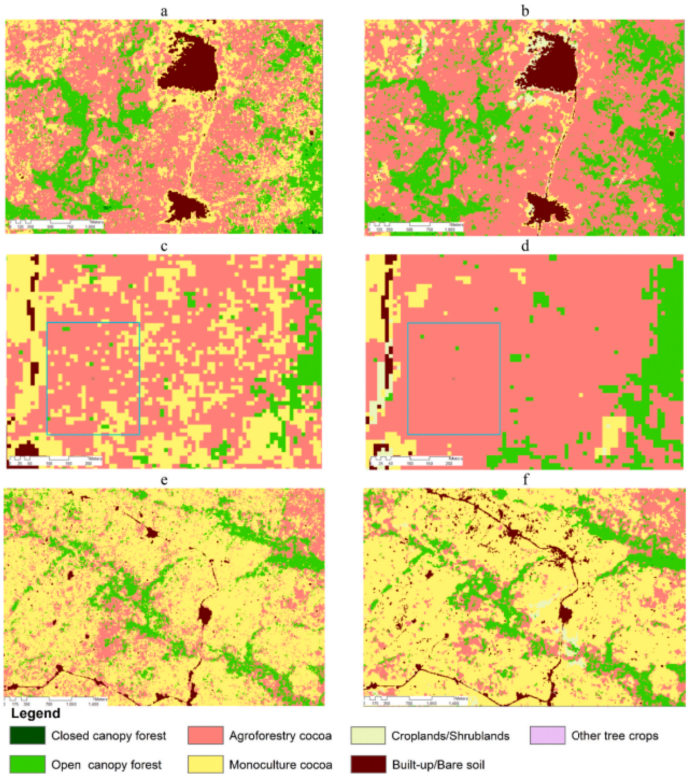


Figure 4: Visual Comparison of pixel-based and object-based classification results (Ashiagbor et al., 2020, p. 10)

2.3 Convolutional Neural Networks for cocoa parcel segmentation

Convolutional Neural Networks [CNNs] offer the possibility to detect and classify cocoa parcels by considering spectral and spatial characteristics in a deep learning architecture by automating feature selection rather than relying on supervised machine learning processes [18]. A theoretical background on CNNs can be found in section 4.1. In a 2018 Bachelor thesis, Filella adapts the U-NET architecture to detect cocoa based on Sentinel 2 imagery (using bands 2, 3, 4, 5, 6, 7, 8, 11 and 12) in Ghana and Ecuador in order to measure the impact of different inputs on the detection of full-sun and agroforestry cocoa [14]. Figure 21 in Appendix 7 provides a diagram of the U-NET, which takes as input satellite imagery and ouputs a mask indicating the class of each pixel (i.e. cocoa, non-cocoa, and unknown). The network is trained with an Adam optimizer with a mini-batch size of 32, learning rate of $10^{-}5$ and L2 regularization of $10^{-}2$.

The loss score is calculated using the Softmax cross-entropy function:

$$L_i = -\log\left(\frac{e^{s_{y_i}}}{\sum_j e^{s_j}}\right) \tag{1}$$

"where s_{yi} is the class score of the correct class and $\sum_j e^{s_j}$ is the sum of all the class scores. The Softmax function takes a vector and squeezes it into a vector with values between zero and one, that sum up to one" [14].

In order to account for an unbalanced and small dataset that over-represents "unknown" pixels and under-represents "cocoa" pixels, Filella uses the following evaluation metrics:

- Recall: "the part of the positive conditions that has been correctly predicted" [14, p. 33]
- Precision: "predictions that are correct and thus actually useful" [14, p. 34]
- Intersection Over Union (IoU): "ratio between the intersection and the union of the predictions and the conditions [...] a good a midpoint between Recall and Precision" p. 34

By training the U-NET with a multispectral image (divided into small batches) and an Ecuadorean ground-truth polygon dataset (full-sun cocoa), the final recall is 93%, the final precision is 98% , and the IoU is 92% . Results are less reliable for agroforest cocoa, shown in figure 5, for which the U-NET is trained using 12 farm polygons and a time-series of 4 multispectral images from December 2017 to January 2018. In order to measure the impacts of under-sampling and temporal data on segmentation outputs, the U-NET is trained and validated separately with temporal and non-temporal data, and different levels of undersampling. Non-temporal data provides the worst results across all metrics, with undersampling leading to an IoU of 47.2%. The use of temporal data performs better: the highest IoU of 58.2% is achieved with a batch size of 64 and a minimum of 100 cocoa pixels. Attempting to reduce batch size to increase the proportion of cocoa is not effective: the model predicts forested areas rather than cocoa crops, which is attributed to a lack of background ("non-cocoa") labels. While the author concludes that their model is effective for full-sun cocoa segmentation, the study is inconclusive regarding agroforestry cocoa. Some of the key areas of future research mentioned in this paper include training the network with a larger dataset and a longer time series, improving the labelling of non-cocoa data, detecting and processing clouds separately, and implementing a more powerful network.

In another CNN-based paper by Kalischek et al. (2022), a CNN is trained using 100,000 cocoa farm polygons, 10,000 non-cocoa polygons and a time-series of 10 Sentinel 2 images collected from each 6-month time period between October 2018 and December 2021. As part of the data pre-processing, cloud-covered samples are marked as "nodata" and the authors

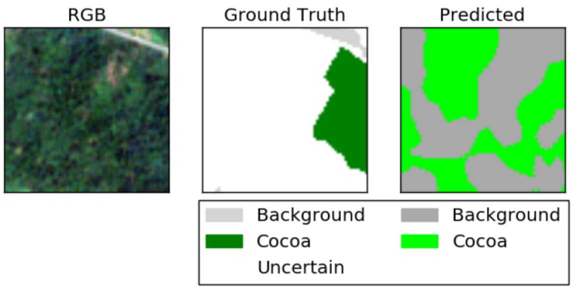


Figure 5: Visualization of validation batch for agroforest cocoa prediction (Filella, 2018, p. 46)

Minimal amount	Percentage of cocoa	Accuracy	Recall	Precision	Intersection over Union
0	2%	53.2%	0.0%	/	0.0%
100	5%	75.7%	61.1%	99.7%	58.2%
200	7%	49.8%	44.2%	99.6%	44.1%

Figure 6: "Results of the final validation using undersampling on temporal data" (Filella, 2018, p. 42)

choose input patches that are at least 10% labelled. In order to save time and provide additional data, the authors train a deep learning network with Sentinel 2 and canopy height LiDAR data, then feed the predicted height values into the CNN.

The architecture used is not openly available, however the paper indicates that it is based on the architecture described in Lang et al. (2019), shown in figure 22 in Appendix 7. The architecture developed by Kalischek et al. takes as input an image patch with 9 bands (32x32x9); there are three consecutive residual blocks with learnable 1x1 convolutional filters, six residual blocks with 3x3 depth-wise separable convolutional layers, the input of the vegetation height map, followed by two further separable residual blocks, and the final output is computed by a single convolutional layer with two 1x1 filters, whose 2-channel output is passed through a Sigmoid transformation. The CNN runs for 32,500 epochs (40,000 iterations) using Adam optimizer, with a base learning rate of 10^-5, with the Dice coefficient as loss function:

$$L = \sum_c \left(1 - \frac{2\sum_i p_{ci} g_{ci} + \epsilon}{\sum_i p_{ci} + \sum_i g_{ci} + \epsilon}\right) \tag{2}$$

"where c is the number of classes, i the pixel index, p and g the prediction and ground truth, respectively." [18, p.9]. The output of this model is a probability map that indicates, for each pixel, the probability of that pixel containing cocoa (between 0 and 1). This probability map can be converted to binary map based on desired level of confidence; in figure 7, 65% probability is used as the minimum threshold for symbolizing a pixel as "cocoa." The authors create 10 replicas of the network using the same data but with different random initializations. To evaluate the quality of the output, the authors use recall and precision, similar to Filella, as well as an additional metric called the F1-score: "the harmonic mean of precision and recall" [18, p. 2]. The results of the segmentation are shown in figure 8; compared to the pixel-based classification implemented to map cocoa at a similarly large scale [4], this deep learning approach improved "precision and recall by more than 26% and 4% respectively" [18, p. 2]. While the implementation of this CNN had good results, it requires the use of time-series datasets due to the limitations of multispectral data when dealing with cloud cover. Therefore, the authors suggest that integrating the use of SAR datasets could allow for monthly or even weekly cocoa mapping updates [18].

To conclude this section, existing research has demonstrated the potential for CNNs to detect the extent of cocoa farms based on multispectral data; however, neither of the CNN-based cocoa detection studies found in the literature integrate SAR datasets as input to their deep learning network. Therefore, this research builds on opportunities for further research that integrates a larger ground truthing dataset in the implementation of a U-NET architecture and evaluates the impact of SAR data on the segmentation of cocoa parcels.

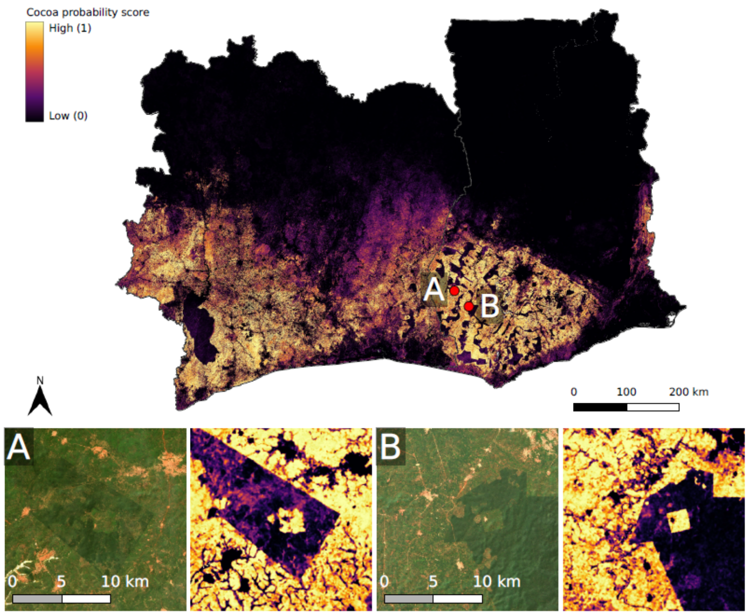
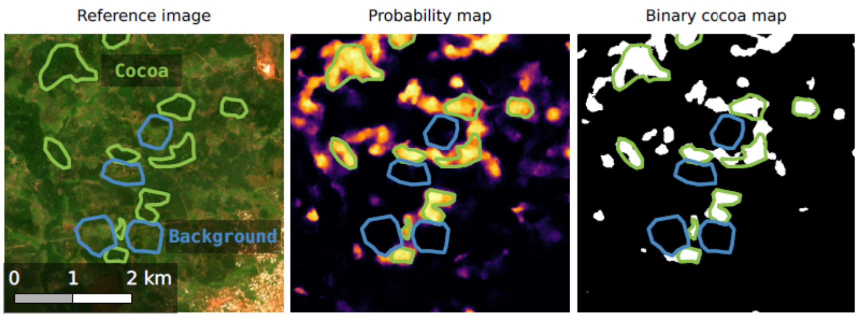


Figure 7: "Cocoa map for Côte d'Ivoire and Ghana. Probability map with 10x10m ground sampling distance." (Kalischek et al., 2022, p. 3)

(a)



(b)

Class	Precision (%)	Recall (%)	F1 (%)	Accuracy (%)
Cocoa	88.5	87.2	87.3	85.9
Non-cocoa	83.9	84.3	84.1	

Figure 8: Evaluation of cocoa segmentation (Kalischek et al., 2022, p. 4)

3 Research questions

The research conducted in this MSc thesis builds upon existing work on cocoa detection via machine learning, with a focus on advancing deep learning approaches. The overarching research question is as follows: *To what extent can a Convolutional Neural Network trained with multispectral* and *SAR datasets enable the automated detection of cocoa crops in Ghana?*

The sub-questions guiding this research include:

1. How does the combination of multispectral and SAR data affect the results of cocoa parcel segmentation trained with data from a single day?
2. How does the combination of multispectral and SAR data affect the results of cocoa parcel segmentation trained with a 6-month time series multispectral dataset?
3. What is the effect of different combinations of Sentinel 2 bands on the results of satellite imagery segmentation (i.e. Visible, NIR and/or SWIR)?
4. How might the use of different polarizations (i.e. VV or VH) affect the influence of SAR datasets on the cocoa segmentation results?
5. How might the use of seasonal SAR data affect the influence of SAR on the cocoa segmentation results?

Additional questions of interest that are beyond the scope of this MSc thesis include:

1. How does the use of high resolution (0.5-3 m) multispectral imagery in visible and NIR bands compare with segmentation results of the same bands at 10 m resolution?
2. How might the attributes of cocoa polygons affect the segmentation accuracy? (e.g. density, age, intercropping vs monocropping)
3. How might seasons of training and testing datasets affect the segmentation results? (e.g. training on Summer datasets, testing on Spring and vice-versa. . .)
4. How might the characteristics of non-cocoa ground truth polygons affect the segmentation accuracy? (e.g. number of polygons, types of land cover. . .)
5. How does the time resolution of a time series datasets affect the segmentation output (e.g. yearly, monthly, or weekly images)?

Figure 9: MuSCoW chart for research sub-questions

4 Methodology

4.1 Theoretical Background

In the field of computer vision, deep learning has been enabling unprecedented accuracy in computers' ability to derive information from images [3]. The foundation of deep learning is the Artificial Neural Network (ANN), which has been modelled after human brain. Just as a brain contains neurons that receive information and can output signals which have a varying degree of influence and inform a person's understanding or action, the ANN contains nodes which receive data, activate, and send the signal on to the next layer of nodes [19]. The complex combination of signals between nodes and layers leads to the performance of a task. For the purpose of machine vision, an ANN would take image data as input, whose pixels are fully connected to a series of hidden layers, which means that all neurons in the input layer are connected to all layers in the next layer, and so on. As shown in figure 10, each neuron applies the activation function to the input (x) and outputs a value (y) that is carried forward to the next layer. The final layer would contain neurons indicating the probability of an image belonging to each class.

Each connection between neurons (as seen in figure 11) represents a weight, which indicates the influence of the previous node activation on the next node. For example, nodes connected pixels in a specific region of the image may have more influence than other regions. The network is initialized with random weights, and is trained by inputting raw data which has already been labelled with a classification in order to otimize the weights and activations across all layers. By detecting patterns in a large number of images and generating a series of weights and activations for the neurons in the hidden layers, the network can then take a new input image, detect the patterns in its pixels, and output a probability of it belonging to each of the potential classes.

The challenge with a fully connected ANN is that the number of parameters that need to be computed grows quickly, especially when using input images that have multiple bands (e.g. R,G,B, other satellite bands). To avoid the high computation cost, a more appropriate model for image processing is the Convolutional Neural Network [CNN]. A CNN enables the detection of patterns by connecting neurons with a subset of the neurons in the previous layer, as shown in figure 12. A CNN uses a number of filters (matrices containing weights) which represent the relationship between an input and subsequent layers. As the convolutional layers are applied, the features and patterns detected become increasingly abstract. This allows the

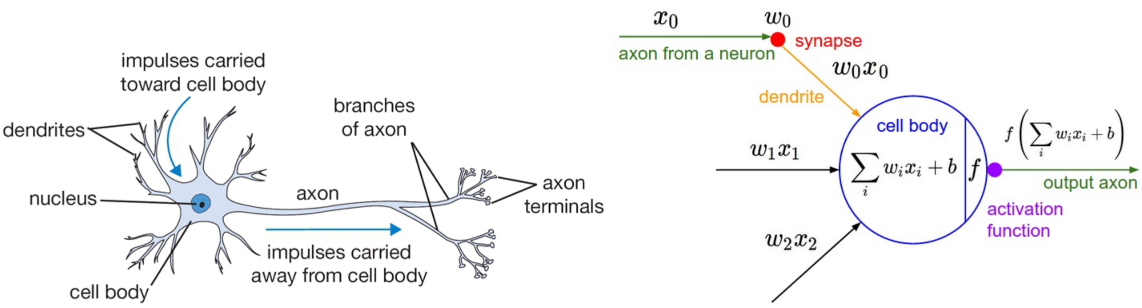


Figure 10: "A cartoon drawing of a biological neuron (left) and its mathematical model (right)." [19]

network to learn both complex and simple spatial patterns which inform the classification of an image. This type of network is useful for classifying an entire image, such as determining if an image is a cat or a dog.

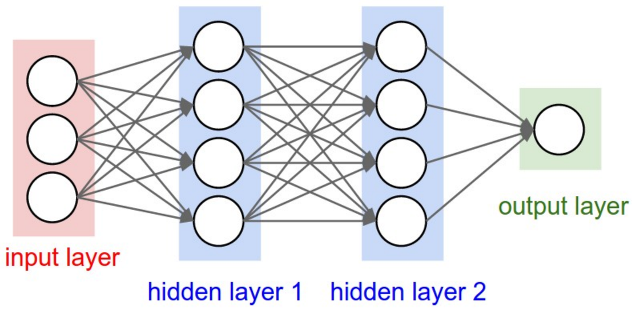


Figure 11: "A 3-layer neural network with three inputs, two hidden layers of 4 neurons each and one output layer." [19]

However, in many cases, it is equally as important to identify the location of a feature, not only its class [24]. In the case of cocoa, we not only want to know if there is cocoa in the image, but where the parcel is located. This technique is called segmentation: it involves pixel-wise classification and simultaneous detection of object instances. In their 2017 paper, Garcia-Garcia et al. provide an overview of existing datasets and methods which apply deep learning for semantic segmentation [3]. Most state-of-the-art semantic segmentation is based on the Fully Convolutional Network, an architecture developed by Long et al. in 2015 which replaces fully connected layers with convolutional layers and "output spatial maps instead of classification scores" [3, p.9]. Convolutional networks designed with segmentation in mind involve upsampling the spatial maps back to "dense per-pixel labeled outputs" [3, p.9]. This upsampling can take the form of a kind of "reverse max pooling" or "deconvolution filters."

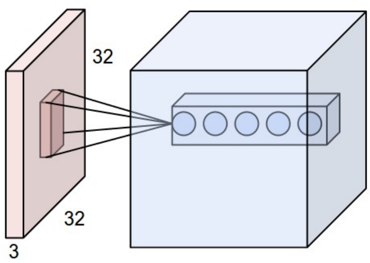


Figure 12: "Each neuron in the convolutional layer is connected only to a local region in the input volume spatially, but to the full depth (i.e. all color channels)" (Karpathy, 2018)

One popular CNN used for image segmentation is the U-NET, which was developed in 2015 by Ronneberger et al. for bio-medical image analysis and is named based on the shape of its architecture as seen in figure 13 [24]. The U-NET takes raw image data as its input, which it carries through two main components: the encoder ("contracting path") and the decoder ("expansive path") [24]. The encoder is responsible for detecting the high and low level patterns in the image as described above, therefore progressively convolving the image to a higher abstraction. Each encoder block is made up of two convolutional + ReLU activation layers followed by a max pooling layer to downsample the image and increase computation efficiency [24]. Prior to max pooling, the convolution output is saved separately for future use as it contains important contextual information that will be used in the decoding process. Each downsampling step involves reducing the spatial dimension of the image by a factor of 2, and a doubling of the feature channels [24]. The final level of the encoder (the bottleneck) includes only two convolution / activation layers and no max pooling. The decoder takes as input the output of the bottleneck and gradually up-convolves the image while bringing back the spatial context that had been saved in each level of the encoder. Each expansive step upsamples the image via a 2x2 convolution and halves the feature channels, concatenates the feature map to the results of the corresponding encoder level, then performs two convolutions [24]. The addition of spatial context allows for the features in the image to be detected and grouped into segments of a class. The final level of the decoder applies 1x1 convolutions for each class being detected, therefore producing a segmentation map for each class. The U-NET is a popular architecture due to its ability to reach a high accuracy with a relatively small number of training images [24].

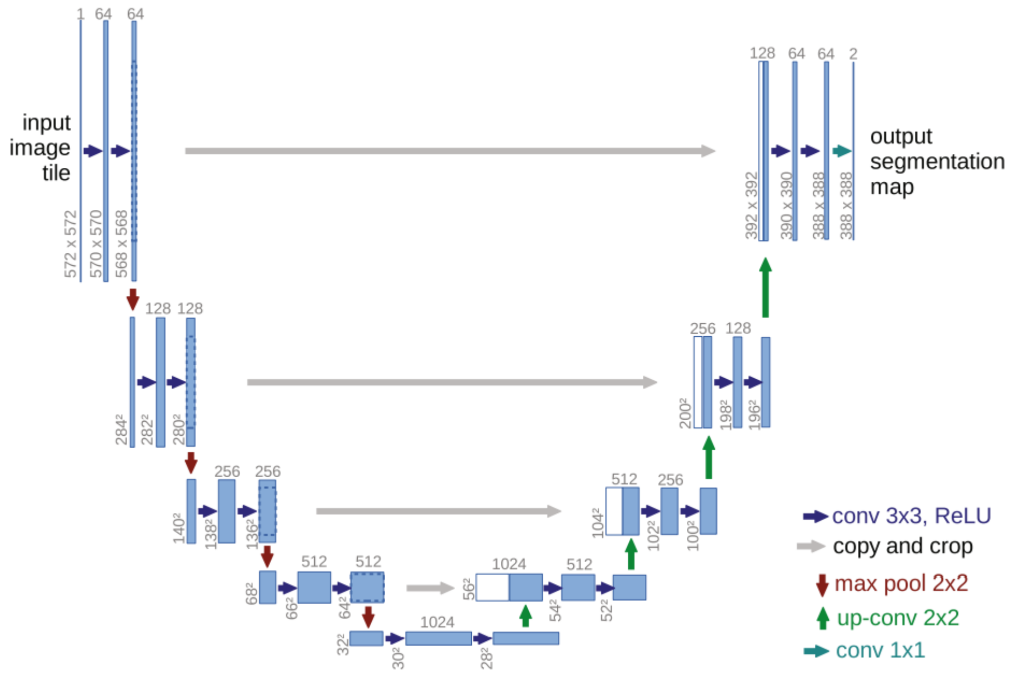


Figure 13: "U-net architecture" (Ronneberger et al., 2015, p. 235)

4.2 Data pre-processing

This thesis is concerned with the segmentation of a specific type of vegetation (cocoa) that is often surrounded by other vegetation types including forest and crops. As described earlier in this paper, two types of remotely sensed satellite data will be used in the training of a CNN for this task: Multispectral Imagery and SAR data. Multispectral imagery is obtained via passive remote sensing from the energy that is reflected from the Earth's surface due to the properties of surface objects [1]. The bands of interest are detected in the visible, Near- and Shortwave-infrared regions of the electromagnetic spectrum. Different objects reflect the sun's energy in different ways, which makes it possible to differentiate between them; for instance, the chlorophyll in vegetation causes it to reflect visible light in the green part of the spectrum which leads vegetation to appear green. This phenomenon also occurs in non-visible parts of the spectrum: healthy plants will reflect more NIR radiation than their unhealthy counterparts. Combined, these surface characteristics give land cover types unique "spectral signatures" which refers to amount of radiation they reflect in different parts of the electromagnetic spectrum (see figure 14). Different vegetation types will often have different spectral signatures, enabling, for instance, the identification of different crops. This becomes more challenging when nearby land covers have similar signatures, as is the case with cocoa and surrounding jungle [7].

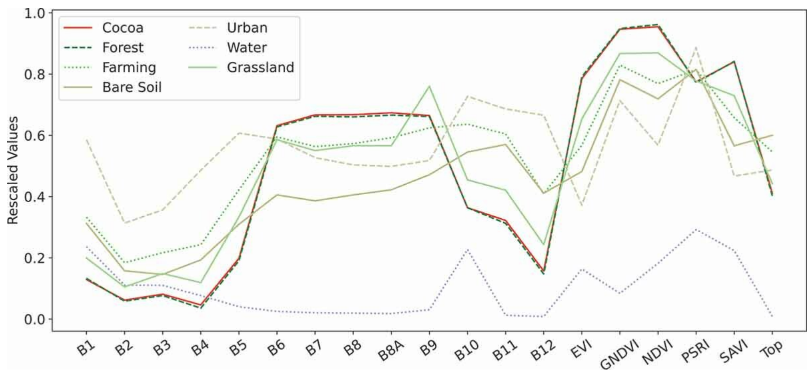


Figure 14: Example of spectral signatures from a cocoa-producing region in Brazil, showing the similarity between cocoa and forest [7]

In this case, Synthetic Aperture Radar [SAR] can provide additional information that may further differentiate between vegetation types. SAR is an active remote sensing technology that involves sending microwave pulses from the satellite to the surface of the Earth, and receiving back a signal (backscatter) with a certain phase and amplitude. The phase provides information on the distance of the surface objects from the sensor (i.e. height), whereas, as seen in figure 16, the amplitude indicates the intensity of the signal that is returned, varying based on geometry, surface roughness and water content [13]. The size of the wavelength used determines the types of surface features that will affect the backscatter: microwaves will penetrate objects smaller than their wavelength, and reflect off objects of a similar size. Sentinel 1 uses C-band, which has a 5 cm wavelength and is therefore able to penetrate tree canopies to a limited extent and capture some characteristics of the vegetation below [23]. The polarization of a radar pulse refers to the "direction of travel of an electromagnetic wave," or how the wave oscillates in relation to the surface it is imaging (horizontal or vertical) [13]. As shown in figure 15 the polarization of the pulse that is transmitted and that which is measured upon return can be the same (co-polarization, such as vertically transmitted and vertical received [VV]) or opposite (cross-polarization, such as vertical transmitted and horizontally received [VH]) [13]. In previous cocoa-related experiments, crop classification was found to be most accurate when both VV and VH data were used for supervised machine learning [23] and VV was found to most accurately predict the presence of gaps in the canopy [22].

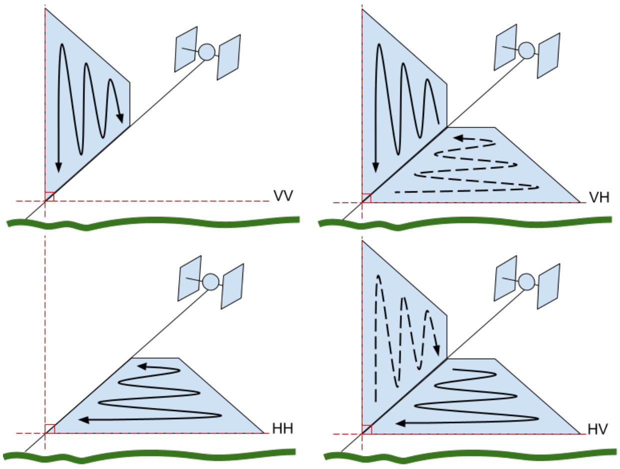


Figure 15: "SAR signals are transmitted and received either vertically (V) or horizontally (H). This gives the potential for four different polarization combinations (transmit listed first, receive second): VV, VH, HH, and HV." [13]

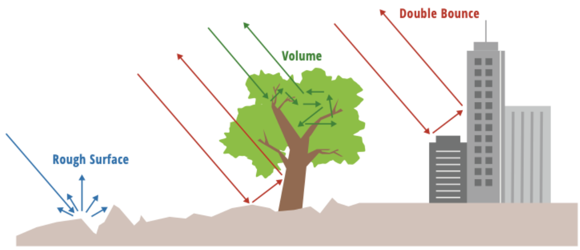


Figure 16: "Rough surfaces give bright returns due to the wide scattering. Vegetated surfaces cause volumetric scattering, which gives a darker return to the imaging platform." [13]

The quality of a dataset is of extreme importance in deep learning applications and is often one of the most complex and time-consuming aspects of a network implementation[3] The following steps will be carried out in order to prepare the datasets:

1. Collect and quality-control ground truth polygons: conduct visual checks on cocoa ground truth to check for possible misclassification, check for overlaps between cocoa and non-cocoa ground truth
2. Rasterize ground truth polygon layers into mask raster (1 = cocoa, 2 = non-cocoa)
3. Process multispectral data: resample all bands to 10 m resolution using bilinear interpolation
4. Process SAR data: project to UTM zone, resample to 10 m resolution
5. Combine multispectral bands and SAR layers: clip SAR to multispectral tile(s), create virtual raster from SAR and multispectral data, apply mask to virtual raster setting any unlabelled pixels = 0 (no data)
6. Prepare small images for input to CNN:
    a) Initialize moving window (e.g. 128 x 128 pixels)

b) Check: is there cocoa within the window? [note: to address data imbalance, potentially set minimum number of cocoa pixels [14, p. 16]]

c) If so, crop and save a copy of the class raster, save a corresponding cropped portion of the masked virtual raster.

d) Move the window slightly while maintaining some overlap in order to augment the data, and continue the process until the entire area has been visited.

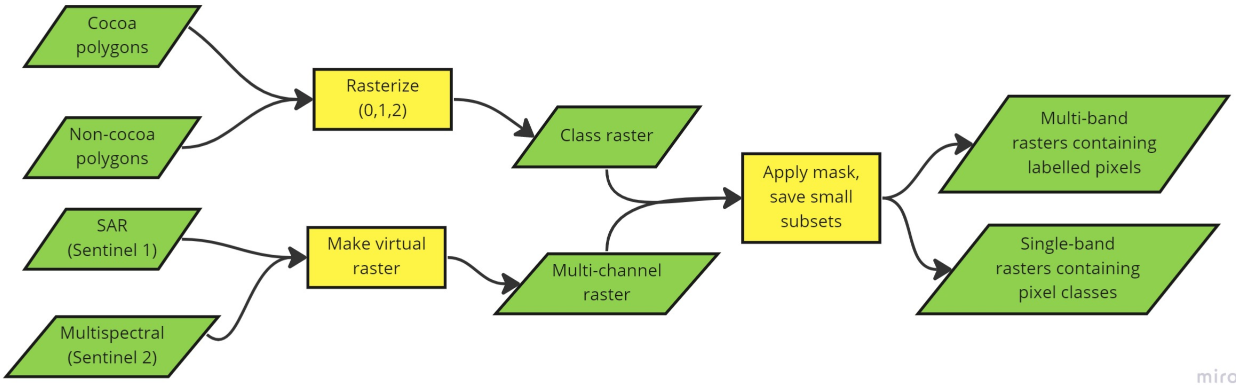


Figure 17: Steps involved in pre-processing data

4.3 Machine Learning Experiments

A U-NET will be constructed and trained with different types of datasets in order to address the research questions in quantitative and repeatable ways. The U-NET architecture will be adapted based on the cocoa segmentation work of Filella (2018) and the Jupyter notebook U-NET implementation by Bhatia (2021) with the following characteristics which may be changed over the course of the research based on the results:

- Input: 128x128x12
- Number of filters: 32
- Number of classes: 3
- Encoder: 5 blocks of two Conv Layers (3x3 filters, 'same' padding) with relu activation and HeNormal initialization, max pooling
- Decoder: 4 blocks of transpose convolution, concatenate with skip connection from encoder, two Conv layers (3x3 filters, 'same' padding)
- Model output: one Conv layer (3x3 filters, 'same' padding) followed by one 1x1 convolution layer to get image to same size as input.

MUST HAVE **1.** How does the combination of multispectral and SAR data (single day) affect the results of cocoa parcel segmentation trained with data from a single day? **2.** How does the addition of SAR data (single day) affect the results of cocoa parcel segmentation trained with 6-month time series multispectral data?	S2 [single day, all 10/20 m bands]
	S2 [single day, all 10/20 m bands] + S1 [single day, both polarizations]
	S2 [6 month series, all 10/20 m bands]
	S2 [6 month series, all 10/20 m bands] + S1 [single day, both polarizations]
SHOULD HAVE **1.** What is the effect of different combinations of Sentinel 2 bands on the results of satellite imagery segmentation (i.e. Visible, NIR and/or SWIR)? **2.** How might the use of different polarizations (i.e. VV or VV- VH) affect the influence of SAR datasets on the cocoa segmentation results? **3.** How might the use of seasonal SAR data affect the influence of SAR on the cocoa segmentation results?	S2 [single day, all 10/20 m bands] + S1 [two days, both polarizations]
	S2 [6 month series, all 10/20 m bands] + S1 [two days, both polarizations]
	S2 [single day, all 10/20 m bands] + S1 [single day, VV polarization]
	S2 [single day, all 10/20 m bands] + S1 [single day, VH polarization]
	S2 [single day]: Visible only, Visible + NIR, Visible + SWIR, Visible + SWIR +NIR
	S2 [6 month series]: Visible only, Visible + NIR, Visible + SWIR, Visible + SWIR +NIR

Figure 18: U-NET training experiments, based on priority research questions

Figure 19 summarizes the experiments that will be carried out by training the U-NET with different datasets in order to address the key research sub-questions. In each of the experiments, the network will be trained using datasets that contain at least 10% non-cocoa labels overall. The results of each experiment will be evaluated using the same metrics as Filella (2018): Recall, Precision and Intersection over Union. All experiments will first be carried out on the scale of one Sentinel 2 image (10,000 x 10,000 m) before upscaling to a larger number of tiles, and eventually aiming to apply segmentation to the the entire study area (400,000 x 600,000 m) if time allows.

4.4 "Must Have": Comparing multispectral-only to multispectral + SAR

To set a benchmark, the U-NET will be trained with Sentinel 2 single-day dataset with ¡10% cloud cover, and then it will be trained again with the single-day multispectral image stacked with a single-day SAR dataset (both polarizations) for the same date (within a 1-month timeframe). To test the impact of time-series data on this comparison, based on the work of Kalischek et al. (2022), the U-NET will be trained with Sentinel 2 time series dataset that selects 10 images with the lowest cloud cover over a 6-month period. It will then be trained again with the same time-series dataset stacked with a single-day SAR dataset.

4.5 "Should Have": Testing different combinations of multispectral bands, SAR polarizations and SAR temporal data

Building off the work of Filella (2018) which tested the influence of different bands on the segmentation of full-sun cocoa parcels in Ecuador, this thesis will conduct similar tests to quantify the importance of different bands on the output of segmentation for intercrop cocoa in Ghana by training the network with the following combinations of bands which are critical to vegetation detection:

- Visible only
- Visible + NIR
- Visible + SWIR
- Visible + NIR + SWIR

The above combinations will be tested with both the single-day dataset and the timeseries dataset to account for the influence of seasonal variation on the results. In order to evaluate the impact of SAR polarization on the segmentation, the U-NET will be trained with the single-day multispectral dataset stacked with each VV and VH polarizations separately. Considering the two distinct seasons in Ghana (the wet and dry seasons [4]), the time of year during which the SAR data is obtained may affect its influence on the segmentation, and the way(s) in which the SAR amplitude and phase change over the year may be different from surrounding vegetation. To test whether the seasonal variation in texture and water content captured by SAR may influence results, the U-NET will be trained with the single-day and timeseries multispectral data stacked each stacked with two SAR images (one from each season).

5 Preliminary Results

The U-NET architecture was adapted based on the Jupyter notebook created by Bhatia (2021) [8] to classify imagery into 3 classes (cocoa, non-cocoa and unknown) and can be viewed in this Google Colab notebook. The network was trained and tested three times, with three different datasets of thirty 128x128 px images overlapping the Sentinel 2 tile shown in figure 23 of Appendix 7. Due to the limited availability of non-cocoa ground truth at the time of writing this graduation plan, the experiments were limited to cocoa-only labels for the first two experiments.

1. Multispectral Data
2. Multispectral and SAR Data
3. Multispectral and SAR Data, including non-cocoa ground truth

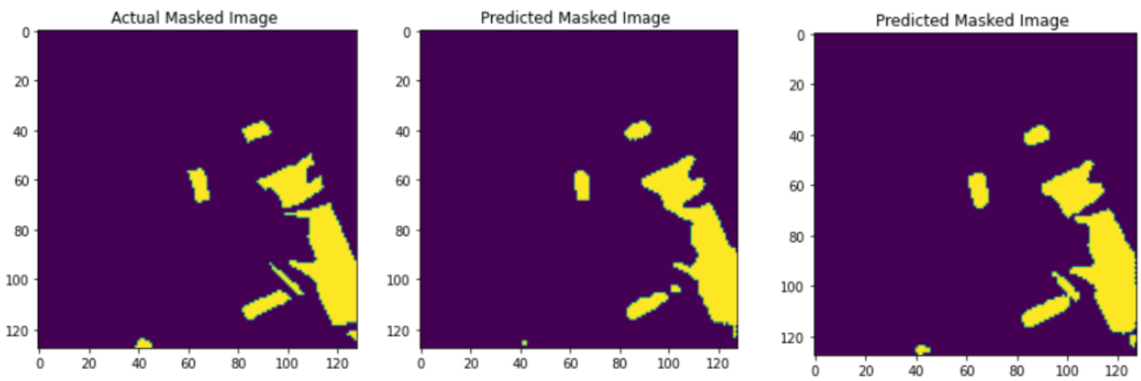


Figure 19: Visualization of initial U-NET segmentation: (1) Actual cocoa mask (2) Prediction from multispectral-only training (3) Prediction from training with multispectral and SAR

Visually, a comparison of the ground truth mask and the predicted masks for multispectral only and SAR + multispectral data show some slight difference. As can be seen in figure 19, the predicted mask from the U-NET trained with SAR seems to reflect the shape of the training polygon with more precision, although it is not possible to draw any conclusions given the small dataset used in this trial. The evaluation metric used was TensorFlow's 'Accuracy', which computes the number of times that a predicted pixel value matches the ground truth pixel value [29]. As shown in figure 24 of Appendix 7, in all three experiments, the validation accuracy reached very high values (over 0.90) from the first or second epoch. These results are likely deceptive and related to the over-representation of the "unknown" class and under-representation of the "non-cocoa" class in the training data [14]. Considering that there is a very high representation of "unknown" pixels, the number of predicted "unknown" pixels is also high, which skews the overall accuracy value. To address this challenge going forward, the output evaluation will be performed with the use of the metrics previously described (Precision, Recall and Intersection over Union) for the cocoa and non-cocoa classes only. An initial attempt was made to compute IoU for the "unknown" class and the "cocoa" class separately and the results are shown in figure 25 of Appendix 7. As expected the IoU was very high for the unknown class (0.9691) and very low for the cocoa class (0.2541). In addition, further efforts will be invested in obtaining non-cocoa ground truth labels to better differentiate between cocoa and other classes with similar spectral signatures.

6 Time planning

The initial theoretical background and data exploration has taken place prior to the graduation plan submission. Prior to the P3 mid-point check-in, all experiments will be carried out at a single tile scale. Following a discussion and evaluation of the results, the experiments will be refined and upscaled to a larger study area prior to submission of P4. Finally, the time between P4 and P5 will be used for the cleaning of code, production of illustrative visuals and refinement of the final report.

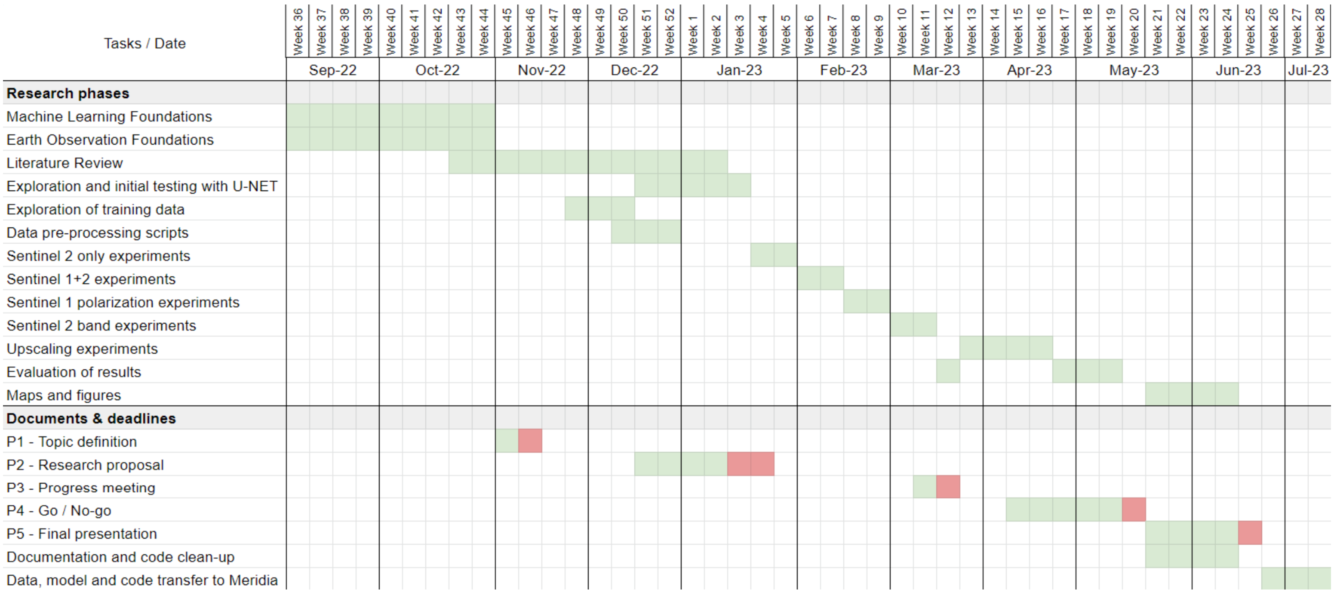


Figure 20: Gantt chart for the completion of this MSc thesis

7 Tools and datasets used

The tools used for this MSc thesis include the following:

- WEkEO: an online platform that facilitates the retrieval and analysis of data produced by Copernicus, the European Union's Earth Observation Programme. Remotely sensed data will be retrieved via the user interface or directly via the Harmonized Data Access API by specifying an area of interest and requesting specific datasets. [32]
- QGIS: an open source geographic information system that enables the visualization and analysis of geospatial datasets. QGIS will be used to conduct individual tasks such as clipping, projecting and merging vector and raster datasets. [26]
- Visual Studio Code: an Integrated Development Environment that will be used to create and run python scripts to automate data pre-processing and repetitive tasks such as splitting a large satellite datasets into small input images for the network. [9]
- Google Colaboratory: a cloud-based platform that enables the creation, sharing and running of Jupyter notebooks for Python scripting. This platform will be used to script and run the U-NET and visualize results. A "Pro" subscription enables the code to run using a Google Cloud GPU. The U-NET will be implemented using the Tensorflow library. [16]

The datasets used in this research include:

- Cocoa polygons: 3 shapefiles containing 88,975 polygons covering 1,332,112,046.91 sq. meters in Ghana. These parcels were collected by Meridia field agents, who mapped the farms alongside the farmers/farm owners or their representatives. Mapping accuracy of all datasets is around 2m on average. Many polygons contain data attributes such as trees age, farm productivity, crop composition (e.g. intercrop vs mono-crop) which are estimates by farmers or their representatives.

- Non-cocoa polygons: 1,011 polygons covering 28,473,219.38 sq. meters in Ghana. As above, these parcels were collected by Meridia field agents, who mapped the farms alongside the farmers/farm owners or their representatives. For the purpose of ground truthing, the aim is to train the network with around 8,000 non-cocoa polygons (10% of number of cocoa polygons, based on the CNN work of Kalischek et al.). Therefore, additional non-cocoa areas (including from other regions) will be labelled manually and sought from alternative sources.

- Sentinel 1 SAR data (C-band, wavelength = 5 cm): Radiometric terrain corrected, dB gamma0, VV and VV-VH polarization, IW mode [12]

- Sentinel 2 Multispectral data: All bands of 10 m or 20 m resolution, filtered for containing less than 10% cloud cover.[2]
  - Band 2: Visible (blue), 490 nm - 10 m
  - Band 3: Visible (green), 560 nm - 10 m
  - Band 4: Visible (red), 655 nm - 10 m
  - Band 5: NIR, 705 nm - 20 m
  - Band 6: NIR, 740 nm - 20 m
  - Band 7: NIR, 783 nm - 20 m
  - Band 8: SWIR, 842 nm - 10 m
  - Band 8A: SWIR, 865 nm - 20 m
  - Band 11: SWIR, 1610 nm - 20 m
  - Band 12: SWIR, 2190 nm - 20 m

Bibliography

[1] Spectral signatures, . URL https://www.esa.int/SPECIALS/Eduspace_EN/SEMPNQ3Z2OF_0.html.

[2] *Sentinel-2 MSI User Guide*, . URL https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-2-msi.

[3] S. O. V. V.-M. J. G.-R. A. Garcia-Garcia, S. Orts-Escolano. A review on deep learning techniquesapplied to semantic segmentation. *arXiv preprint*, 2017. URL https://arxiv.org/pdf/1704.06857.pdf.

[4] I.-O. Abu, Z. Szantoi, A. Brink, M. Robuchon, and M. Thiel. Detecting cocoa plantations in côte d'ivoire and ghana and their implications on protected areas. *Ecological Indicators*, 129:107863, 2021. ISSN 1470-160X. doi: https://doi.org/10.1016/j.ecolind.2021.107863. URL https://www.sciencedirect.com/science/article/pii/S1470160X21005288.

[5] G. Ashiagbor, E. K. Forkuo, W. A. Asante, E. Acheampong, J. A. Quaye-Ballard, P. Boamah, Y. Mohammed, and E. Foli. Pixel-based and object-oriented approaches in segregating cocoa from forest in the juabeso-bia landscape of ghana. *Remote Sensing Applications: Society and Environment*, 19:100349, 2020. ISSN 2352-9385. doi: https://doi.org/10.1016/j.rsase.2020.100349. URL https://www.sciencedirect.com/science/article/pii/S2352938520301178.

[6] G. Ashiagbor, W. A. Asante, E. K. Forkuo, E. Acheampong, and E. Foli. Monitoring cocoa-driven deforestation: The contexts of encroachment and land use policy implications for deforestation free cocoa supply chains in ghana. *Applied Geography*, 147:102788, 2022. ISSN 0143-6228. doi: https://doi.org/10.1016/j.apgeog.2022.102788. URL https://www.sciencedirect.com/science/article/pii/S014362282200159X.

[7] J. E. Batista, N. M. Rodrigues, A. I. R. Cabral, M. J. P. Vasconcelos, A. Venturieri, L. G. T. Silva, and S. Silva. Optical time series for the separation of land cover types with similar spectral signatures: cocoa agroforest and forest. *International Journal of Remote Sensing*, 43(9):3298–3319, 2022. doi: 10.1080/01431161.2022.2089540. URL https://doi.org/10.1080/01431161.2022.2089540.

[8] V. Bhatia. U-net implementation from scratch using tensorflow. *Medium*, July 2021. URL https://medium.com/geekculture/u-net-implementation-from-scratch-using-tensorflow-b4342266e406.

[9] V. S. Code'. Learn to code with visual studio code. URL https://code.visualstudio.com/learn.

[10] E. Commission. Proposal for a regulation of the european parliament and of the councilon the making available on the union market as well as export from the union of certaincommodities and products associated with deforestation and forest degradation andrepealing regulation (eu) no 995/2010, 2021. URL https://environment.ec.europa.eu/system/files/2021-11/COM_2021_706_1_EN_ACT_part1_v6.pdf.

[11] S. Erasmi and A. Twele. Regional land cover mapping in the humid tropics using combined optical and sar satellite data—a case study from central sulawesi, indonesia. *International Journal of Remote Sensing*, 30(10):2465–2478, 2009. doi: 10.1080/01431160802552728. URL https://doi.org/10.1080/01431160802552728.

[12] *Sentinel-1 SAR User Guide*. European Space Agency. URL https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-1-sar.

[13] A. S. Facility. Introduction to sar. URL https://hyp3-docs.asf.alaska.edu/guides/introduction_to_sar/.

[14] G. B. Filella. Cocoa segmentation in satellite images with deep learning. Master's thesis, ETH Zurich, 2018. URL https://ethz.ch/content/dam/ethz/special-interest/baug/igp/photogrammetry-remote-sensing-dam/documents/pdf/Student_Theses/BA_BonetFilella.pdf.

[15] Food and A. O. of the United Nations. Global forest resources assessment. *n/a*, 2020. doi: https://doi.org/10.4060/ca9825en. URL https://www.fao.org/3/ca9825en/ca9825en.pdf.

[16] Google. Welcome to colaboratory. URL https://colab.research.google.com/.

[17] K. S. M. M. Guillaume Rousset, Marc Despinoy. Assessment of deep learning techniques for land use land cover classification in southern new caledonia. *Remote Sensing*, 13(12), 2021. doi: 10.3390/rs13122257. URL https://hal.archives-ouvertes.fr/hal-03284887/document.

[18] N. Kalischek, N. Lang, C. Renier, R. Daudt, T. Addoah, W. Thompson, W. Blaser-Hart, R. Garrett, and J. Wegner. Satellite-based high-resolution maps of cocoa planted area for côte d'ivoire and ghana. 06 2022.

[19] A. Karpathy. Cs231n: Convolutional neural networks for visual recognition. neural networks. Website, 2018. URL https://cs231n.github.io/neural-networks-1/.

[20] R. Kumar, A. Kumar, and P. Saikia. *Deforestation and Forests Degradation Impacts on the Environment*, chapter Deforestation and Forests Degradation Impacts on the Environment, pages 19–39. Springer International Publishing, Cham, 2022. ISBN 978-3-030-95542-7. doi: 10.1007/978-3-030-95542-7_2. URL https://doi.org/10.1007/978-3-030-95542-7_2.

[21] Meridia. Field data solutions for smallholder supply chains, n.d. URL https://www.meridia.land/.

[22] F. N. Numbisi and F. Van Coillie. Does sentinel-1a backscatter capture the spatial variability in canopy gaps of tropical agroforests? a proof-of-concept in cocoa landscapes in cameroon. *Remote Sensing*, 12(24), 2020. ISSN 2072-4292. doi: 10.3390/rs12244163. URL https://www.mdpi.com/2072-4292/12/24/4163.

[23] F. N. Numbisi, F. M. B. Van Coillie, and R. De Wulf. Delineation of cocoa agroforests using multiseason sentinel-1 sar images: A low grey level range reduces uncertainties in glcm texture-based mapping. *ISPRS International Journal of Geo-Information*, 8(4), 2019. ISSN 2220-9964. doi: 10.3390/ijgi8040179. URL https://www.mdpi.com/2220-9964/8/4/179.

[24] T. B. Olaf Ronneberger, Philipp Fischer. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, 2015. URL https://link.springer.com/content/pdf/10.1007/978-3-319-24574-4.pdf?pdf=button.

[25] L. Orozco-Aguilar, A. LÃ³pez-Sampson, M. E. Leandro-MuÃ±oz, V. Robiglio, M. Reyes, M. Bordeaux, N. SepÃºlveda, and E. Somarriba. Elucidating pathways and discourses linking cocoa cultivation to deforestation, reforestation, and tree cover change

in nicaragua and peru. *Frontiers in Sustainable Food Systems*, 5, 2021. ISSN 2571-581X. doi: 10.3389/fsufs.2021.635779. URL https://www.frontiersin.org/articles/10.3389/fsufs.2021.635779.

[26] 'QGIS'. Qgis - the leading open source desktop gis. URL https://www.qgis.org/en/site/about/index.html.

[27] S. Saatchi, D. Agosti, K. Alger, J. Delabie, and J. Musinsky. Examining fragmentation and loss of primary forest in the southern bahian atlantic forest of brazil with radar imagery. *Conservation Biology*, 15(4):867–875, 2001. doi: https://doi.org/10.1046/j.1523-1739.2001.015004867.x. URL https://conbio.onlinelibrary.wiley.com/doi/abs/10.1046/j.1523-1739.2001.015004867.x.

[28] D. K. Tamga, H. Latifi, T. Ullmann, R. Baumhauer, M. Thiel, and J. Bayala. Modelling the spatial distribution of the classification error of remote sensing data in cocoa agroforestry systems. *Agroforestry Systems*, 97(1):109–119, oct 2022. doi: https://doi.org/10.1007/s10457-022-00791-2.

[29] 'TensorFlow'. tf.keras.metrics.accuracy, Nov. 2022. URL https://www.tensorflow.org/api_docs/python/tf/keras/metrics/Accuracy.

[30] D. Tutu Benefoh, G. B. Villamor, M. van Noordwijk, C. Borgemeister, W. A. Asante, and K. O. Asubonteng. Assessing land-use typologies and change intensities in a structurally complex ghanaian cocoa landscape. *Applied Geography*, 99:109–119, 2018. ISSN 0143-6228. doi: https://doi.org/10.1016/j.apgeog.2018.07.027. URL https://www.sciencedirect.com/science/article/pii/S0143622817310470.

[31] R. C. Weih and N. D. Riggan. Object-based classification vs. pixel-based classification: Comparative importance of multi-resolution imagery. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2010. URL https://www.isprs.org/proceedings/XXXVIII/4-C7/pdf/Weih_81.pdf.

[32] 'WEkEO'. About us, 2023. URL https://www.wekeo.eu/about.

Appendix A

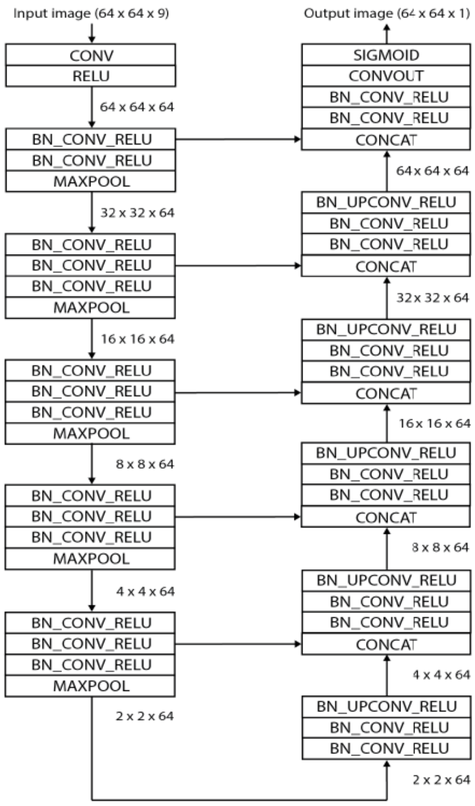


Figure 13: Modified U-Net architecture

Figure 21: "Modified U-NET Architecture" (Filella, 2018, p. 22)

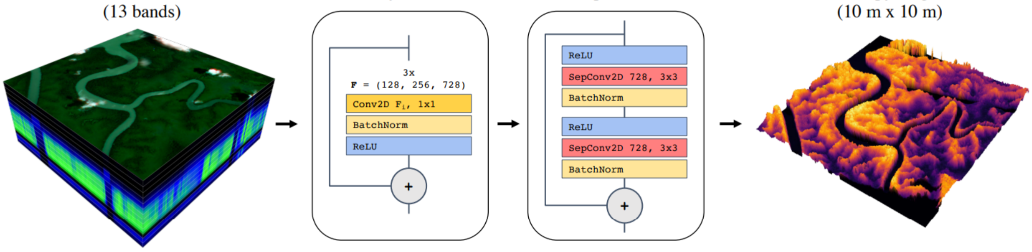


Figure 22: "CNN architecture to predict vegetation height, later adapted for Kalischek et al. cocoa research (Lang et al., 2019, p. 6)

Appendix B

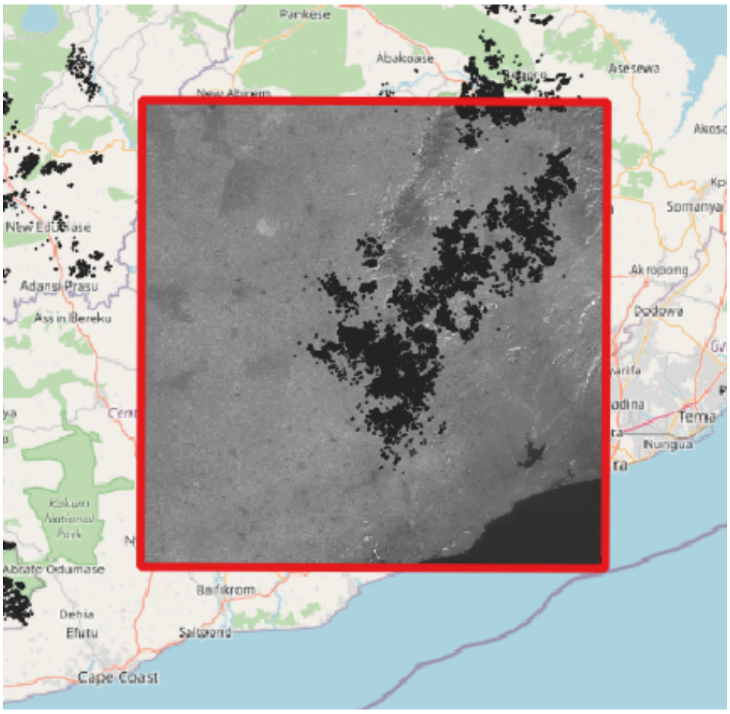


Figure 23: Extent of Sentinel 2 tile used for initial experiments (the dark polygons represent cocoa and non-cocoa ground truth polygons)

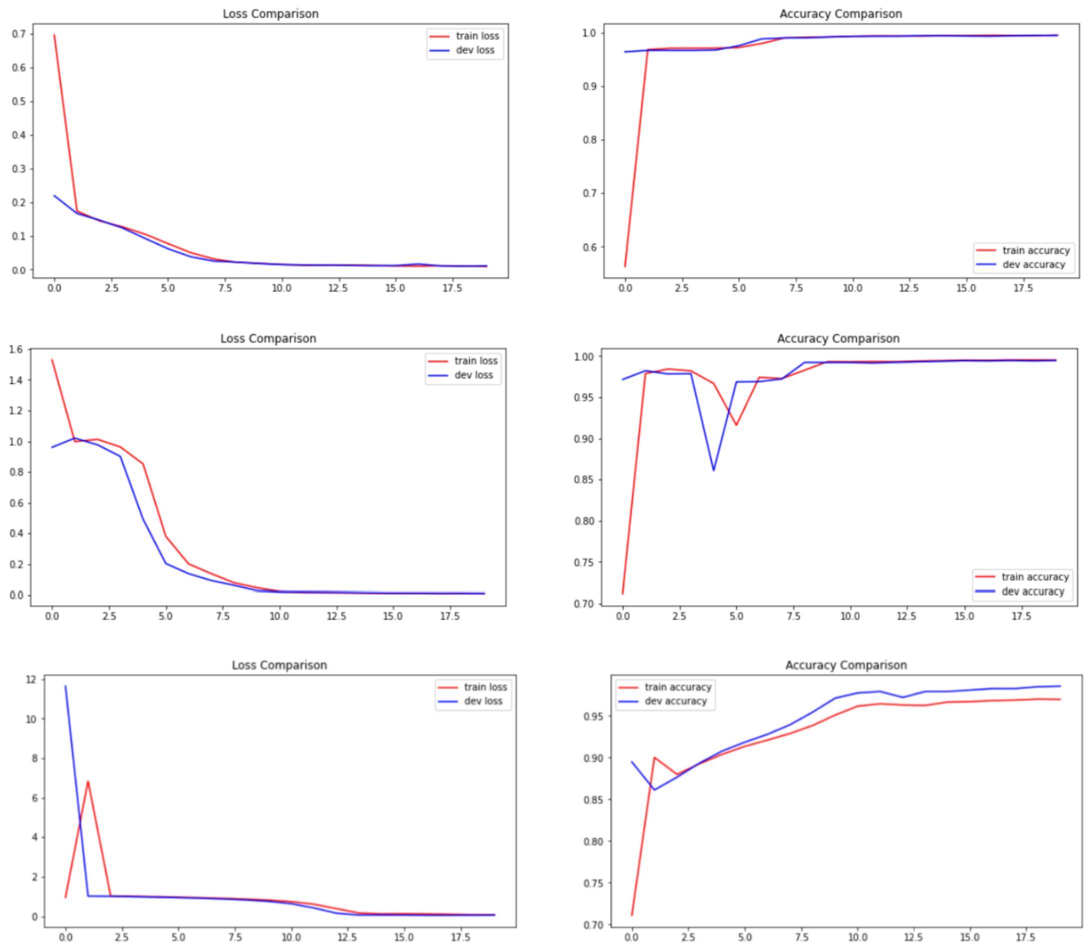


Figure 24: Plotted accuracy and loss values of initial U-NET segmentation: (1) Training with multispectral data (2) Training with multispectral and SAR data (3) Training with multispectral and SAR data with non-cocoa labels

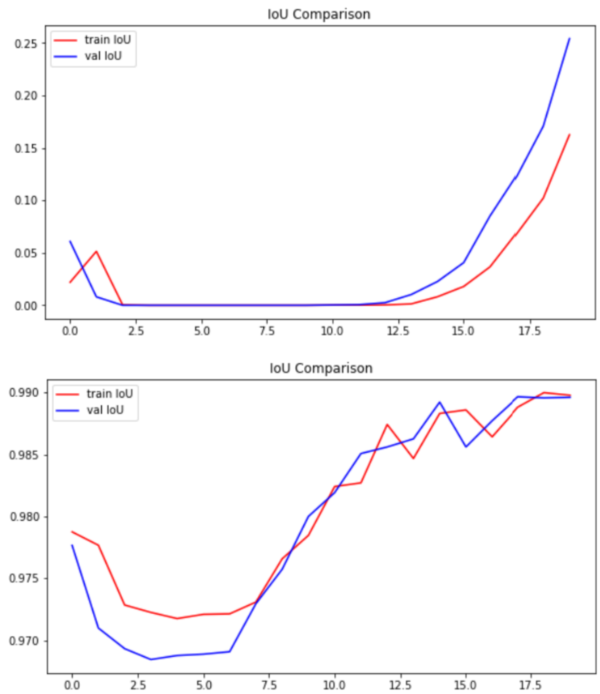


Figure 25: IoU preliminary results (1) Cocoa class (2) Unknown class