# AGRICULTURAL FIELD BOUNDARY DELINEATIONS IN SMALLHOLDER FARMING SYSTEMS OF SOUTHEAST ASIA USING SENTINEL-2 DATA AND CONVOLUTIONAL DEEP LEARNING MODELS

ASHIK BHARI SHIVAPRASAD
August, 2023

SUPERVISORS:

dr. Mahdi Khodadadzadeh
dr. Claudia Paris

# AGRICULTURAL FIELD BOUNDARY DELINEATIONS IN SMALLHOLDER FARMING SYSTEMS OF SOUTHEAST ASIA USING SENTINEL-2 DATA AND CONVOLUTIONAL DEEP LEARNING MODELS

ASHIK BHARI SHIVAPRASAD
Enschede, The Netherlands, August, 2023

Thesis submitted to the Faculty of Geo-information Science and Earth Observation of the University of Twente in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation.
Specialization: M-GEO

SUPERVISORS:

dr. Mahdi Khodadadzadeh
dr. Claudia Paris

THESIS ASSESSMENT BOARD:

Prof.dr. Raul Mila Zurita (chair)
Prof.dr.ir. Claudio Persello (external examiner)

# ABSTRACT

Agricultural Field Boundary (AFB) delineation is beneficial for estimating incentives as part of farming schemes. AFB delineation can help formulate innovative micro-agricultural finance programs, agricultural field statistics calculation, crop yield estimation, and other applications of precision agriculture practice(Enclona et al., 2004). Traditional methods used to monitor AFB are time-consuming and labor-intensive since they are based on human field surveys. Furthermore, the diversity of Earth Observation (EO) technology allows for data collection via a wide range of sensors with varied spatial, spectral, and temporal resolutions. Combined with the recent advancements in computer vision and machine learning algorithms, it is convenient to perform the delineation of agricultural field boundaries. Despite the obvious advantages, it is still the abundance of data created by EO sources can cause a variety of problems in processing. Through this research, we create a tailored workflow that efficiently delineates the AFB from pre-processed Sentinel-2 EO data built with seasonal statistic-based composites such as geometric median, median, and medoid with the help of CNN (U-Net) and a post-processing method based on graph-based segmentation and contour extraction for polygonization of boundary predictions.

**Keywords**

*CNN, Automated Delineation, Agricultural boundary delineation, Polygonization*

# ACKNOWLEDGEMENTS

Special thanks to my Supervisors.

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# Chapter 1

# Introduction

## 1.1 BACKGROUND

Agricultural Field Boundary (AFB) delineation is beneficial for estimating incentives as part of farming schemes. AFB delineation can help formulate innovative micro-agricultural finance programs, agricultural field statistics calculation, crop yield estimation, and other applications of precision agriculture practice (Enclona et al., 2004). Studies have shown that the inclusion of agricultural field boundary data in crop classification solutions yields better results (Janssen & Middelkoop, 1992)

Smallholding farmers own small plots of land on which they grow subsistence crops and occasionally a few varieties of cash crops, relying nearly solely on family labor(Lowder et al., 2016). According to research, "appropriate boundary information records can increase food security"(Rockson et al., 2013). Improving food security and alleviating extreme poverty is a key worldwide concern that can be achieved by aiding smallholding farmers in overcoming financial restrictions and successfully managing risks, thus increasing their capacity to invest. The social support required to narrow the poverty gap increases as impoverished people's wages improve due to higher investment returns (FAO, 2016).

Traditional methods used to monitor AFB are time-consuming and labor-intensive since they are based on human field surveys. In particular, performing field surveys is not feasible for continuously monitoring large study areas. The growth of Earth Observation (EO) data via satellite platforms has reduced the necessity for field surveys. Furthermore, the diversity of EO technology allows for data collection via a wide range of sensors with varied spatial, spectral, and temporal resolutions. Despite the obvious advantages, the abundance of data created by EO sources can cause a variety of problems in processing. As a result, suitable machine learning algorithms and tools were required to analyze massive amounts of data and provide critical insights into embedded information.

With the advancement in computer vision and biomedical science built the foundation for developing deep learning algorithms, it has been proven that deep learning technologies such as Convolutional Neural Networks (CNN) were more efficient than traditional machine learning algorithms in scientific fields that required visual analysis of images (Ronneberger et al., 2015). The advent of deep learning CNN and the availability of open-access high-resolution satellite images acquired at a global scale enabled the possibility of automatically delineating crop field boundaries. However, smallholder farms present a challenge for automated field boundary delineation due to their tiny size, asymmetrical shape, and use of mixed-cropping systems, which leave their boundaries ill-defined. For this reason, recent works focused on the use of very high-resolution

satellite images (i.e., having spatial resolution <1m) to classify field boundaries in small-scale farming areas (Persello et al., 2019). However, the main drawback of these data is that they are typically restricted to access or are not cost-effective for specific applications (Masoud et al., 2019).

Due to the lack of availability of very high-resolution satellite images, the need to evaluate the usage of medium and high-resolution satellite earth observation data to delineate AFB for smallholder farms efficiently arises.

### 1.2 RELATED WORKS AND RESEARCH GAP

Recent literature has investigated a number of methods applied to EO data for delineating agricultural fields (Gopidas & Priya, 2022; North et al., 2019; Persello et al., 2019; Waldner & Diakogiannis, 2020; M. Wang et al., 2022). Most of the techniques explored in the above research fall into two main categories: edge-based or region-based techniques. Edge-based methods aim to identify boundaries by detecting sharp changes or edges in the image. These methods typically involve edge detection algorithms such as Canny edge detection, Sobel operator, or Laplacian of Gaussian (LoG) filters. By extracting edges from satellite or aerial images, field boundaries can be identified based on the presence of strong gradient variations (Graesser & Ramankutty, 2017; Turker & Kok, 2013; Yan & Roy, 2014). In contrast, region-based methods involve segmenting the image into distinct regions or objects and then identifying the boundaries of these regions. Popular region-based algorithms include region growing, watershed segmentation, and graph-based methods. These techniques often use color, texture, or other image features to group pixels into meaningful regions and then delineate the boundaries between them (Evans et al., 2002; Garcia-Pedrero et al., 2017; Mueller et al., 2004). Although these methods can be effective for some study areas, problems may arise when applying these approaches at a large scale. Edge-based algorithms can be affected by high-frequency noise in the image, leading to false edges. Sensitivity to noise can result in incomplete boundaries, where certain field edges may not be accurately detected. Additionally, edge-based algorithms often require parameterization, and selecting the appropriate parameters can be challenging. Region-based algorithms also require parameter selection. If the parameters are not appropriately chosen, the segmentation results may be unsatisfactory. Over-segmentation can occur when fields with internal variability are divided into excessively small regions. On the other hand, under-segmentation can happen when nearby small fields are grouped together (Belgiu & Csillik, 2018).

Recently, CNN has been used to perform contour detection, which can learn relevant discriminant features from reference data automatically. CNNs can extract initial pixel values of edges, local forms, and complicated texture patterns (Bergado et al., 2016; Marmanis et al., 2016; Szegedy et al., 2015). Their effectiveness has also been demonstrated on AFB delineation for small and big-holder farms (S. Wang et al., 2022). CNN is particularly efficient on high-resolution and very high-resolution images for AFB delineation (Persello et al., 2019). However, very high-resolution images are expensive to acquire and are usually available spanning small areas. Sentinel-2 high-resolution satellite data, on the other hand, is freely available and acquired at a large scale. Nevertheless, with the spatial resolution of a sentinel-2 platform being 10m, crop boundaries delineated may have imperfections. Interestingly, some research has already been performed to increase the accuracy of AFB delineation using sentinel-2 satellite images in combination with Super-Resolution Mapping (SRM) techniques (Masoud et al., 2019) where a Fully Convolution Network (FCN) was implemented with SRM to upscale the input sentinel-2 image of 10m to 5m. This ap-

proach outperformed alternative methods for contour detection, such as the Global Probability of Boundaries (GPB) algorithm, as described by (Crommelinck, n.d.), which initiates its computation by determining the gradient. Subsequently, the program utilizes the gradient information to estimate the likelihood of boundaries and proceeds with contour extraction. The study's results demonstrate that deep learning technology is a viable strategy for delineating field boundaries using satellite data. However, SRM techniques can be computationally intensive and may require significant computational resources and processing time. The algorithms used for super-resolution mapping often involve complex mathematical operations, which can be time-consuming. SRM techniques heavily rely on high-resolution images for learning(Li et al., 2016). If the input images are of low quality, contain noise, or have artifacts, it can negatively impact the performance of the SRM technique. SRM techniques may not always yield significant improvements in image resolution or detail, especially for regions such as Cambodia with high atmospheric disturbances such as cloud presence in images. To this end, we aim to investigate the use of Sentinel-2 satellite imagery with 10-meter resolution bands for AFB delineation.

Based on literature analysis and agricultural field boundary delineation projects executed in collaboration with FAO, it is evident that existing CNN deep learning models can successfully delineate crop boundaries. However, the overall accuracy of these models tends to be somewhat compromised due to the uncertainties in the pre-processing and post-processing stages. Interestingly, recent research articles have yet to explore the utilization of temporal information to emphasize the presence of crop boundaries during the pre-processing phase. By incorporating temporal information, it is possible to address issues related to variations in image contrast and noise resulting from cloud coverage in remote sensing images. Furthermore, the boundaries predicted by existing CNN models are subsequently converted into vector data through post-processing techniques such as morphological segmentation. Morphological segmentation methods often require manual tuning of parameters such as the size of the structuring element, the threshold values, and morphological operations. This parameter selection can be subjective and may require trial-and-error or expert knowledge to achieve satisfactory results and may also result in incorrect geometric information. The lack of an automated parameter optimization process can be a limitation, especially when dealing with large datasets or when consistency across different images is required. The emphasis of the study lies in improving the pre-processing phase, model building with an efficient loss function, and post-processing techniques associated with AFB delineation using Sentinel-2 data. Through this research, a tailored workflow is built with post-processing techniques, this study aims to create a pipeline that automatically delineates AFB and saves them as vector data, eliminating the need for manual parameterization. To this extent, our implementation for post-processing is based on Graph-based segmentation followed by contour extraction. Graph-based segmentation techniques have a rich history in computer vision and image processing, dating back to several decades. Early graph-based segmentation techniques were primarily based on low-level image features like intensity, color, and texture. These methods used simple graph structures and optimization algorithms to achieve segmentation. The development of graph-cut algorithms, such as the min-cut/max-flow algorithm by (Wu & Leahy, 1993), revolutionized graph-based segmentation. However, the methods based on min cut had some limitations when applied directly to image segmentation tasks. The primary challenge is that it treats each edge independently, without considering the global context of the image, and hence was biased towards finding small components. To overcome this limitation, normalized cuts criteria based on associations proposed by (Jianbo Shi & Malik, 2000), but this technique faced mathematical complexities and was an NP-Hard computational problem (Felzenszwalb & Huttenlocher, 2004). The Felzenszwalb algorithm, a graph-based segmentation technique, is a computationally efficient and highly effective approach (Felzenszwalb & Huttenlocher, 2004). This algorithm builds upon the foundations of

previously mentioned graph-based segmentation methods. The aforementioned methodology has demonstrated advances in both the speed and quality of segmentation. Moreover, it possesses the capability to identify disparities in picture boundaries, circumventing the constraints associated with under-segmentation or over-segmentation limitations frequently observed in region-based segmentation techniques.

### 1.3 RESEARCH PROBLEMS AND OBJECTIVES

The majority of current processes that utilize deep learning models for AFB delineation exhibit shortcomings in incorporating a pre-processing stage that effectively highlights the borders of crops. Conventional workflows commonly prioritize the utilization of a single remote-sensing image from a collection of images captured over the course of one year, neglecting the temporal information that could account for the phenological variations of crops. Consequently, these workflows fail to produce a composite image that effectively highlights the distinct boundaries of crops. A persistent research difficulty revolves around the need for a proficient post-processing technique capable of extracting field boundary vector data from the field boundary prediction generated by the deep learning network. Nevertheless, it has been observed that the watershed transformation has exhibited superior performance when compared to alternative instance segmentation algorithms (S. Wang et al., 2022). However, it should be noted that the accuracy of the generated vector data for field boundaries may decrease due to the sensitivity of the watershed transformation to image noise. Minor fluctuations in pixel values have the potential to give rise to undesired segments. In addition, the integration of morphological segmentation in conjunction with the watershed transformation, an algorithm commonly employed in ImageJ software and initially outlined by (Legland et al., 2016) in their work on MorphoLibJ, represents a partially automated procedure that requires user participation. This process is not seamlessly integrated into an automated pipeline to delineate boundaries, as the output of the morphological segmentation or watershed transformation lacks geometry information. Consequently, an additional step must be undertaken to retain the geometry from the prediction image. Furthermore, the result produced by the algorithm is in the form of a raster, which needs additional processing in order to polygonize.

The study areas considered for this research are Cambodia and Vietnam. The most significant challenges with AFB delineation for such regions are the fragmented, irregular shape of smallholder farms with a size of less than 1 ha and the problem with cloud coverage during acquisition. These challenges raise the need to investigate and build a workflow based on CNN deep learning models that can efficiently generate AFB delineation results. Research output can contribute to tracking the sustainable development goals (SDG) progress through various initiatives to support smallholding farmers and to ensure food security.

Primary Objective: To develop a tailored workflow to efficiently extract AFB data from open-access EO data using convolutional deep learning models for small-scale framing areas in Cambodia and Vietnam.

The primary objective is achieved through the following specific sub-objectives:

1. Develop a pre-processing strategy that leverages temporal information from Sentinel-2 images in the study area to enhance crop boundaries.

2. Apply a specialized loss function to enhance the accuracy of existing deep learning models for generating Agriculture Field Boundary (AFB) data.
3. Implement a post-processing strategy capable of polygonizing predictions without compromising delineation performance.

The following questions will be answered through this research for the specific sub-objectives above.

- How can we ensure that the training images incorporate temporal information to effectively capture the phenology of crops, while also minimizing noise?
- How to ensure that the model does not suffer contour loss?
- How to efficiently convert the model's AFB predictions into vector data?

# Chapter 2

# Data

## 2.1 STUDY AREA AND DATASETS

The study areas for this research have been chosen based on the presence of smallholding farms in Cambodia and Vietnam. This is further supported by the inclusion of highly detailed satellite photos in figures 2.1 and 2.2 presented below. Study areas were carefully chosen to represent diverse agricultural landscapes, encompassing a range of crop types and land-use patterns. For the selection of study areas in Cambodia and Vietnam, we also accessed data from the International Production Assessment Division (IPAD) of the USDA; based on this assessment, we selected study areas for the research. The selected study areas are depicted in Map 2.3. The research covers regions characterized by rice cultivation, illustrated in Figure 2.4 ("Cambodia Production", n.d.; "Vietnam Production", n.d.). The choice of study areas is rooted in the prevalence of a particular crop's cultivation. This strategy guarantees that the compiled seasonal composites include phenological indicators, like mature crops ready for harvest, facilitating the delineation of AFB.

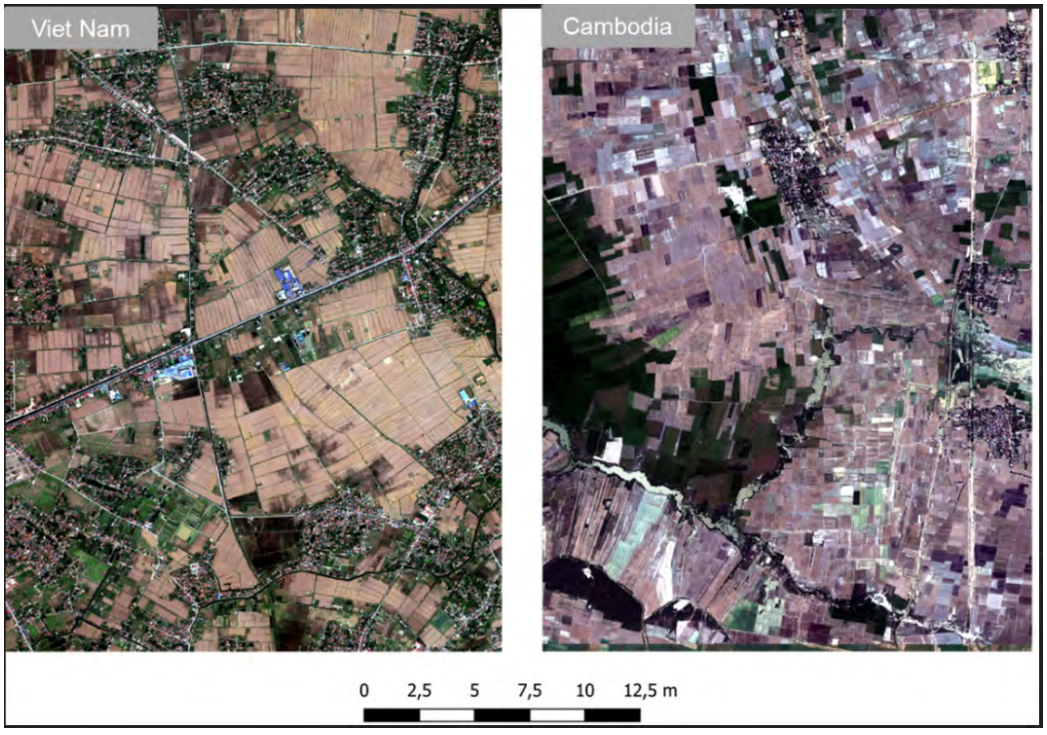



Figure (2.1) Very High Resolution (VHR) WV2 images acquired in the two pilot areas. The zooms highlight the different types of crop field boundaries present in the pilots.

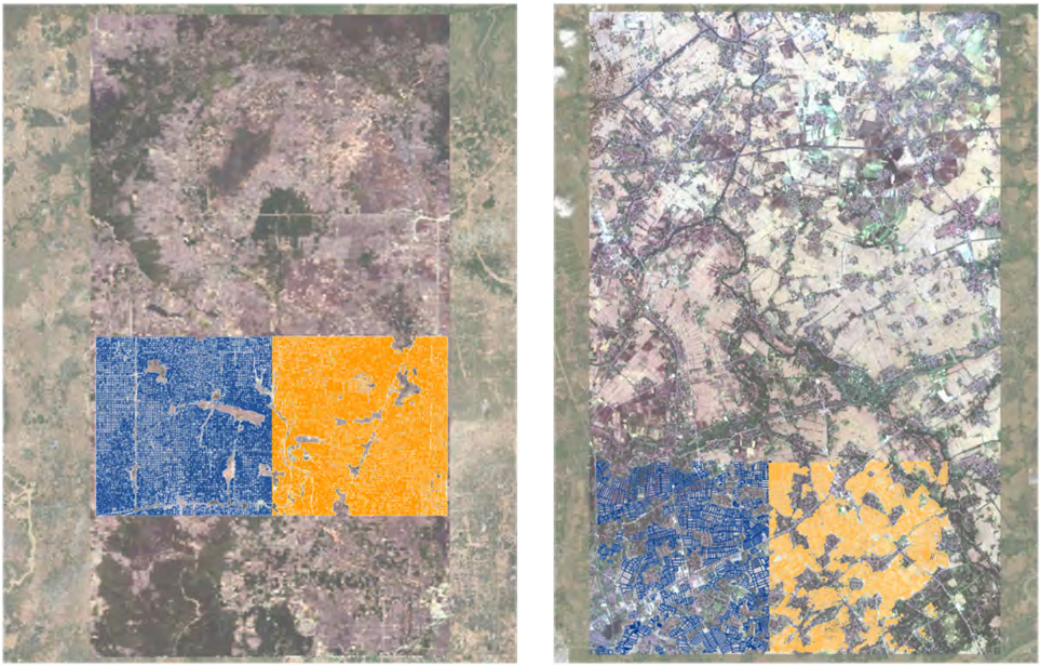


Figure (2.2) The crop boundary benchmark dataset on the left side is for Cambodia, and on the right is for Vietnam. The study areas in Cambodia and the fields in Cambodia suffer image contrast issues and are dense compared to that in Vietnam.

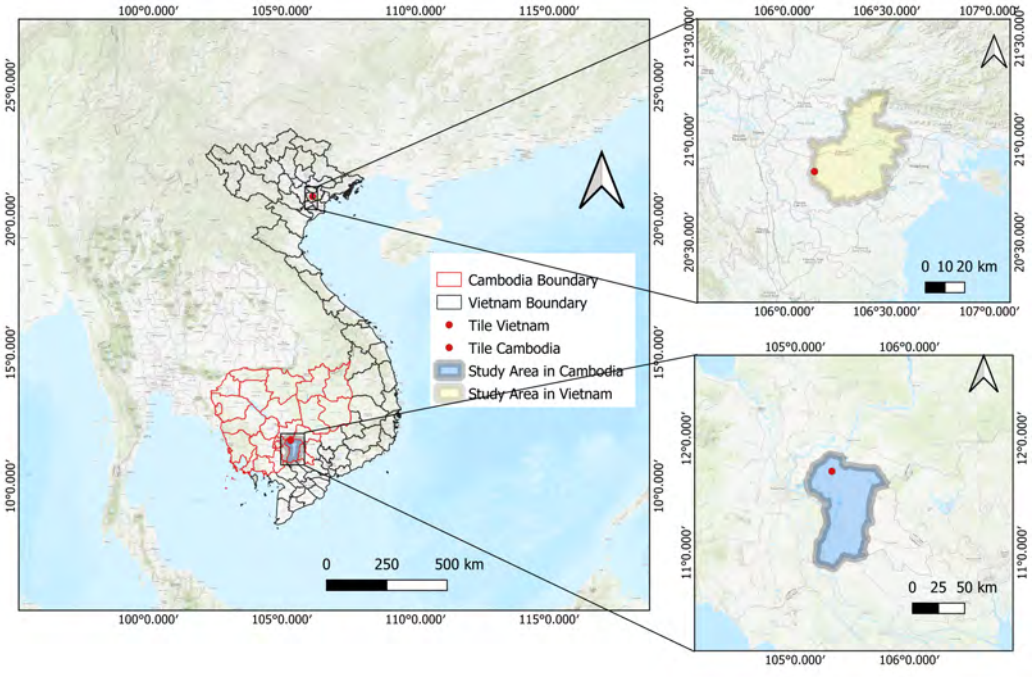
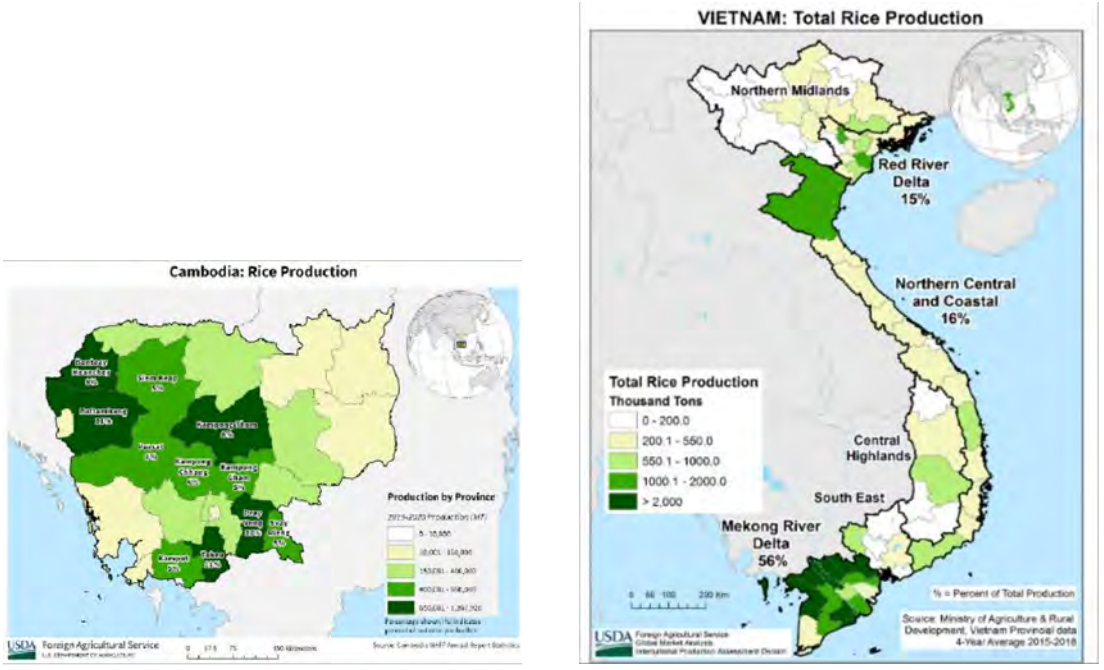



Figure (2.3) Study Areas Vietnam(Yellow) near the Red River Delta and Cambodia(Blue) near the Mekong area, and points(Red dots) representing the tiles of recognized Rice paddy growing regions for which boundary dataset must be generated using U-Net.(Units : Distance in km).




(a) Total Rice Production in Cambodia (b) Total Rice Production in Vietnam

Figure (2.4) Rice Production ("Navigo", n.d.)

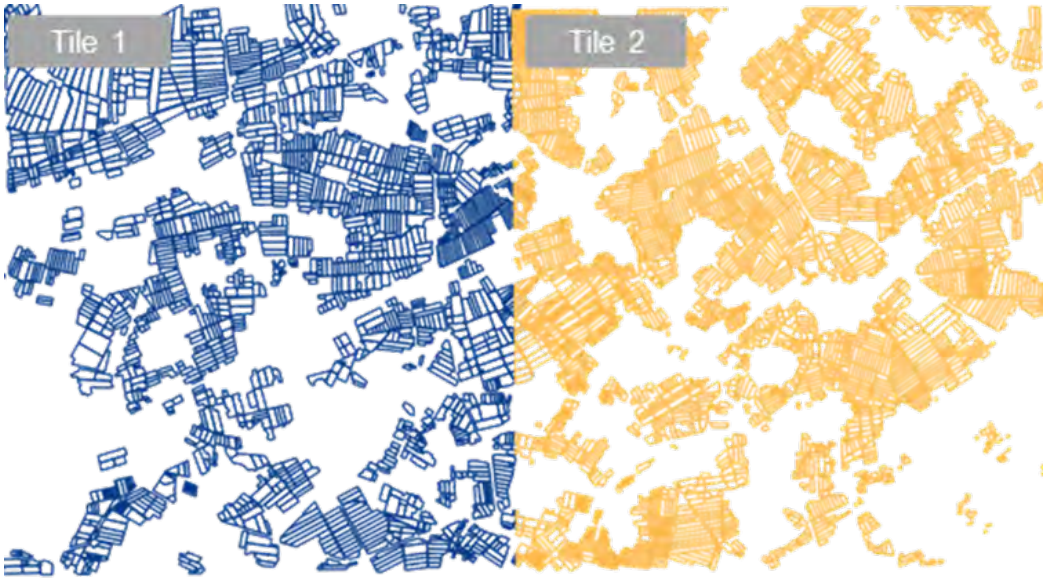



Figure (2.5) Crop boundary benchmark dataset created in Viet Nam. To test the developed DL model, the data will be divided into two spatially disjoint areas to create a training set (highlighted in yellow) independent from the validation set (highlighted in blue).

In figures 2.5 and 2.6 we can visually interpret the benchmark data that will be employed in-order to validate the deep learning model developed in this research made available through the current project supported by the Ministry of Agriculture, Forestry and Fisheries (MAFF) of Japan

and implemented by the Statistics Division of the Food and Agriculture Organization (FAO) of the United Nations (Xinyan et al., 2021). Table 2.1 below describes the number of polygons, crop size in the benchmark dataset that will be used for training and validation.

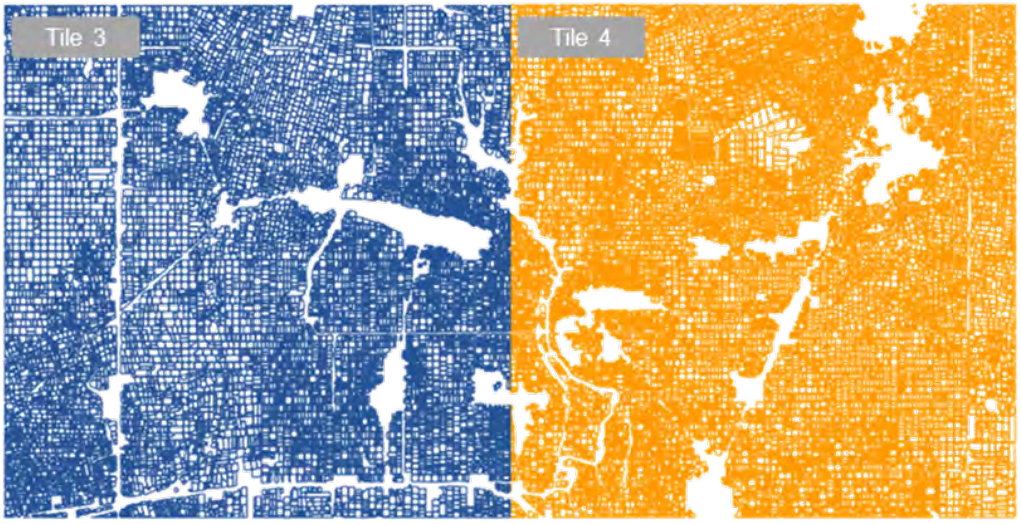



Figure (2.6) Crop boundary benchmark dataset created in Cambodia. To test the developed DL model, the data will be divided into two spatially disjoint areas to create two training sets (highlighted in yellow) independent from the validation sets (highlighted in blue).

**Table 2.1** Number of polygons, minimum, mean, and maximum values of the crop size for the training and validation data.

| Pilot area | Tile | Training | | | | Tile | Validation | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | N. | Min Crop Size | Mean Crop Size | Max Crop Size | | N. | Min Crop Size | Mean Crop Size | Max Crop Size |
| **Vietnam** | **Tile 2** | 3082 | 0.27 ha | 7.22 ha | 48.42 ha | **Tile 1** | 1839 | 1.14 ha | 12.33 ha | 105.98 ha |
| **Cambodia** | **Tile 4** | 16141 | 0.25 | 2.81 ha | 35.04 ha | **Tile 3** | 12574 | 0.29 ha | 3.68 ha | 35.42 ha |

The remote sensing data used in this research are Sentinel-2 images, which are open-access and available at a global scale. These data are acquired at 10 m spatial resolution in the spectral bands of interest (i.e., red, green, blue, and infrared). Moreover, the satellite has a short revisit time (i.e., 5 days at the equator), thus leading to a dense time series of images that can be used to overcome the cloud coverage problem present in the considered study areas. The datasets that will be used for the research are summarized in Table 2.2.

**Table 2.2** Dataset used for the study area.

| Dataset | Research-Use Description | More Info. |
|---|---|---|
| Sentinel-2 | This dataset will be used as input data for the model to output the agricultural field boundary dataset | https://scihub.copernicus.eu/ |
| Dutch Reference Dataset | This dataset with 1 million fields acquired in the Netherlands will be used to pre-train the model. | (Persello et al., 2019) |
| Local Reference Dataset | Worldview2 data for tiles that cover rice paddy fields were used to generate training data through photo interpretation | this is a local dataset (no links) |

# Chapter 3

# Methods

The methodology followed through in this research is depicted as a workflow in figure 3.1. The methodology encompasses several stages, commencing with a preprocessing and data preparation phase. Subsequently, model construction is undertaken, employing appropriate loss functions. Finally, a post-processing technique is employed to convert the model's predictions into polygons.

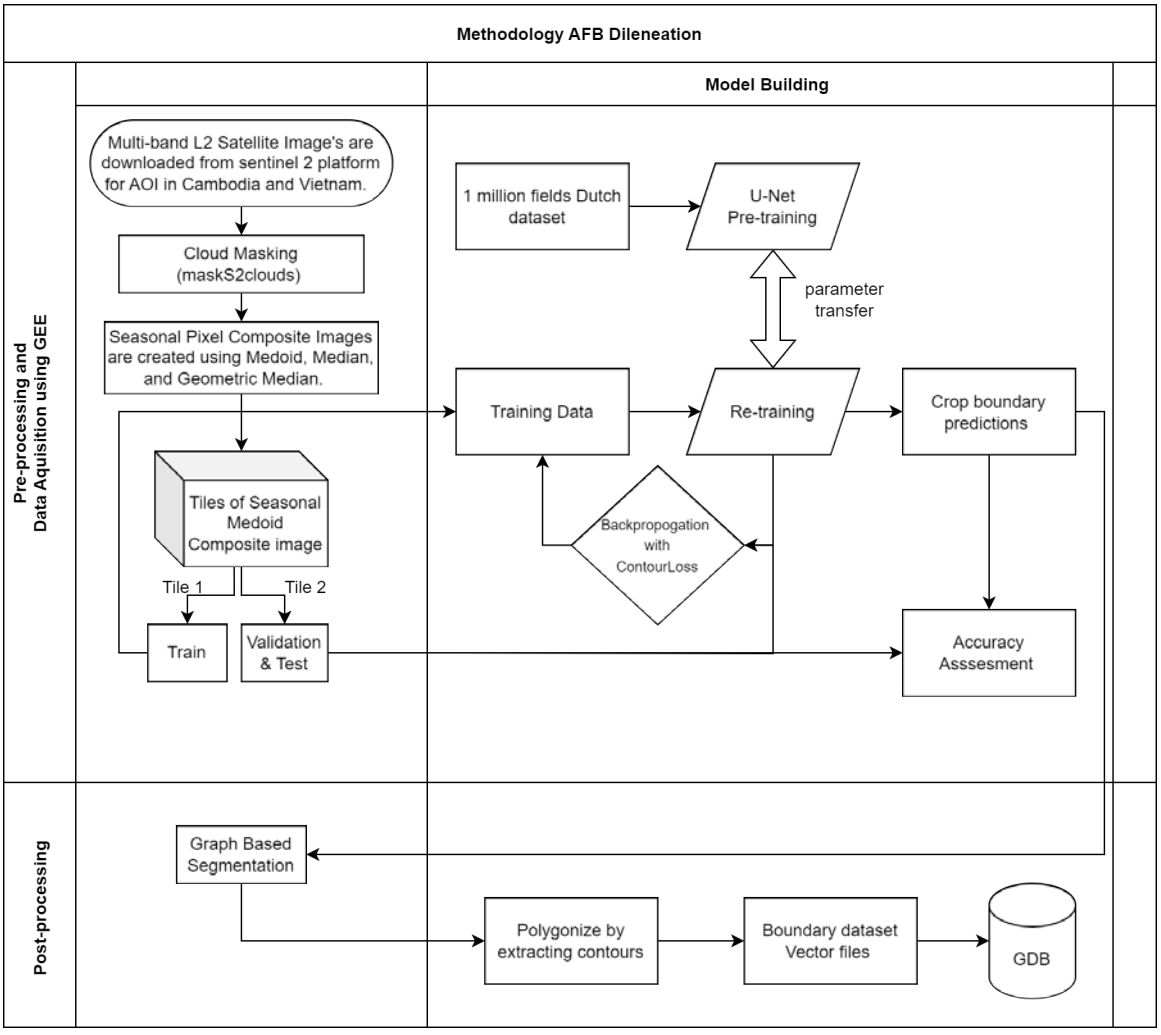



Figure (3.1) Workflow Diagram.

### 3.1 PRE-PROCESSING AND DATA PREPARATION

Using Google Earth Engine, Sentinel-2 EO data from the Surface Reflectance (S2_SR) image collection has been acquired for the year 2021, and the start date and end date for the EO remote sensing images of interest will be based on the crop seasons calendar. The EO images are chosen based on the requirements that they cover the whole study area, and correspond to the time period pertinent to the relevant agricultural season.

Data pre-processing is a crucial step to obtain high accuracy for AFB delineation. The literature review shows that images of small fields with poor boundary contrast will result in poor model performance for AFB delineation (S. Wang et al., 2022). Our focus is to preprocess data to emphasize the crop boundaries in the considered remote sensing data before applying the deep learning model. With such an effort, we can reduce uncertainty in the entire process. In our study, we are constructing statistical composite images that effectively eliminate outliers. However, it is important to note that statistical composite methods, such as the median, are susceptible to the influence of outliers. Therefore, we aim to minimize the impact of outliers on the resulting image. To achieve this, we employ a cloud masking technique on the Earth Observation (EO) images prior to constructing the composite. The sentinel-2 Level-2A dataset encompasses multiple bands, including the QA60 band, which provides detailed quality information at the pixel level for each spectral band. This inclusion is advantageous for our purposes. The QA60 spectral band offers compressed data that aids in the identification of several pixel attributes, including the presence of clouds, cloud shadows, and water bodies. Furthermore, the SCL band, also known as the Scene Classification band, furnishes data pertaining to the classification of land cover(water,vegetation and cloud pixel information ) for every individual pixel. In order to proficiently eliminate cloud-contaminated pixels, the maskS2clouds() function offered by Google Earth Engine will be utilized. The aforementioned function employs suitable bit masking techniques on the QA60 and SCL bands, facilitating the detection and mitigation of pixels affected by cloud cover and noise.

It is a common approach to consider a yearly median pixel composite satellite image to eliminate cloud cover and other noise in image (Roberts et al., 2017). This approach results in a single composite per year and effectively overcomes any noise in the EO data and is a result of loss of information on temporal dynamics. The yearly median composite treats all observations equally without considering the temporal changes that may occur throughout the year. This method effectively represents the "typical" or "central" spectral response for each pixel, but it discards valuable information about the temporal dynamics, such as seasonal variations, phenological changes, or short-term events (e.g., floods or vegetation stress). In certain applications, this loss of temporal information can limit the ability to monitor and analyze dynamic processes over time. Hence, in our research, we are utilizing a seasonal pixel-based composite instead of a yearly pixel-based composite.

Building a composite can be done in two ways: rule-based selection or statistic-based selection of pixels (Roberts et al., 2017). In the study, we will explore a statistic-based selection of pixels to build a composite.

The construction of statistical composites will rely on the careful selection of observations, which will encompass crop seasons and a combination of months within each crop season. This approach aims to prioritize the inclusion of mature crops in the dataset, hence facilitating efficient feature learning by the deep learning model and enhancing its ability to generalize effectively to new and unseen data. Crop seasons for Cambodia and Vietnam are as seen in figure 3.2 and 3.3.

**Table 3.1** Selection of Months per season based on the rice crop calendar for Cambodia and Vietnam

| Region | Dry Season | Wet Season |
|---|---|---|
| Vietnam | March, April, May | Nov, Dec, Jan |
| Cambodia | Jan to April | Oct, Nov, Dec |

Although there are multiple varieties of crops grown in Cambodia and Vietnam, we will primarily focus on Paddy crops because our study area is majorly composed of paddy fields. Study areas selected in Cambodia has two crop season known as a monsoon crop season(Wet season), starting from May until February, and a dry crop season starting from January till April. Vietnam has two crop seasons for paddy a monsoon crop season and a dry crop season. Monsoon crop season is comparatively a long cycle in comparison with dry crop season, which is shorter and takes less time to yield. Monsoon season planting is usually started from May through July, and for heavy rain and flood-prone areas, fifteen to thirty-one day's old seedlings are transplanted; harvesting mostly begins from November through January. The time frame of Sentinel-2 EO data used for experiments as part of the research, also based on the investigation and analysis of crop calendars, can be seen in Table 3.1.

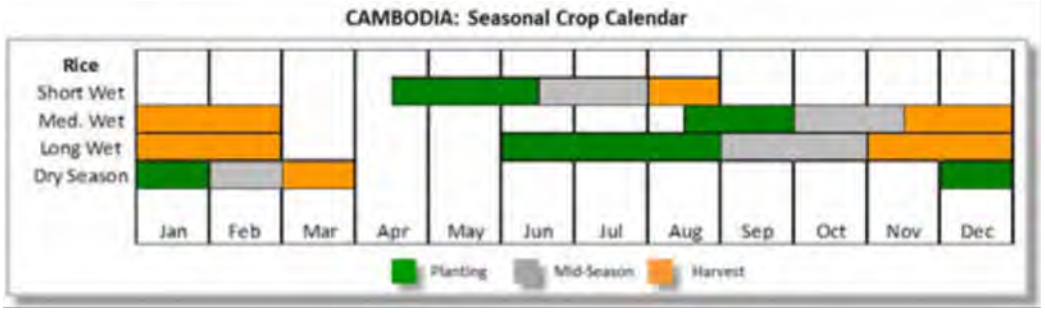



Figure (3.2) Rice Crop Calendar Cambodia ("Index.htm", n.d.)

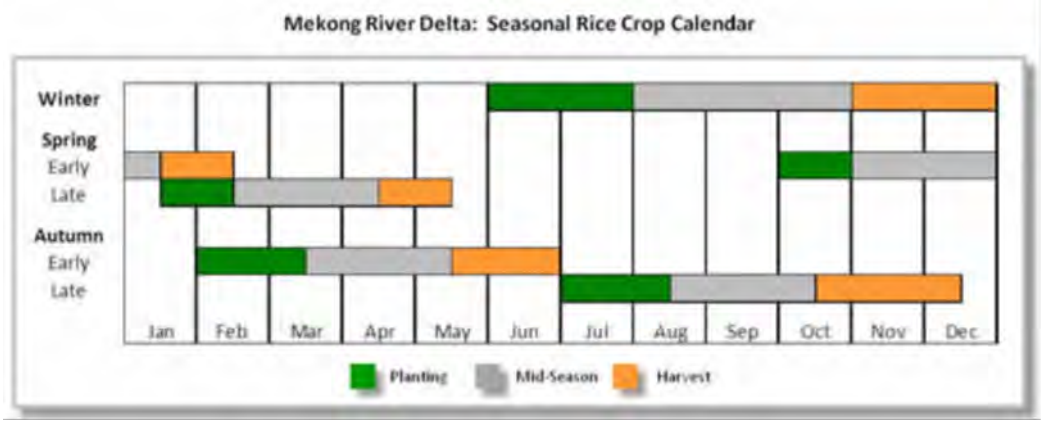



Figure (3.3) Rice Crop Calendar Vietnam ("Vietnam_Rice_Dec2012", n.d.)

Statistic-based pixel composites will enhance the overall quality of remote-sensing EO images. Statistic-based pixel composites are amongst the popular solutions to overcome random noise, sensor artifacts, and atmospheric disturbances in EO data by computing statistical measures such as Median. Pixel-based statistical composites built using median have certain limitations. It is sensitive to outliers. The median composite considers the median value for each band across all available images for the given time period. While this can reduce the influence of extreme values (outliers) in individual images, it is still sensitive to the presence of outliers. Outliers can arise due to various reasons, such as atmospheric effects, sensor noise, or data artifacts. In such cases, the median composite may not accurately represent the true spectral signature of the target area. When using the median statistical method to build a pixel composite, pixels are extracted for each band separately to form a summary statistic image (Roberts et al., 2017), but this cannot preserve the spectral relationship of the bands (Roberts et al., 2017).In order to effectively preserve spectral relationships across all spectral bands during the construction of a pixel-composite image, it is recommended to employ a high-dimensional statistic-based pixel composite method that can be applied simultaneously to all bands. One such approach is the utilization of the medoid algorithm. By utilizing an optimization technique that performs iterative searches for the data point that minimizes the sum of distances to all other data points across all dimensions, this approach differs from the median composite method, which applies the GEE reduction method on a single dimension or band at each iteration. Application of a high dimensional composite technique is demonstrated by (Flood, 2013), which is specifically designed for creating seasonal composites. In our research, we compare the results between Medoid and L1-Median, also known as the geometric median, and they are both robust. Both of these methods find the data point that minimizes the sum of the absolute differences between all other data points, with one major difference being that the medoid is an actual member of the data points (spectral reflectance pixel values) that is closest to the median value whereas Geometric median is the spatial center of the data points (Roberts et al., 2017).

The application of a seasonal monthly pixel composite built using geometric median, medoid based on the crop calendars to accentuate the crop borders will be investigated in this study, which can possibly emphasize contrast and boundaries in the images.

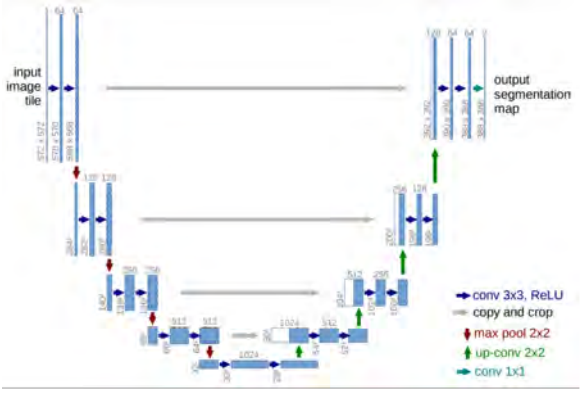



Figure (3.4) U-Net Architecture (Ronneberger et al., 2015)

## 3.2 MODEL BUILDING

The CNN model developed will be based on the U-Net architecture, as depicted in 3.4. The utilization of transfer learning methodology will be employed as a means to address the constraint of insufficient reference data available for the training of the model. Initially, the model will undergo pre-training using a dataset consisting of 1 million field data samples collected in the Netherlands. These data were made accessible by (Persello et al., 2019). Additionally, the model will be retrained using local reference data from Cambodia and Vietnam to account for the small-scale crops that are not common in the Netherlands. Previous research has provided evidence that the ability to extract features obtained from large datasets is advantageous for tasks that are interconnected and may be effectively used to a novel neural network (Yang et al., 2019). The U-Net architecture-based convolutional neural network (CNN) has been found to exhibit superior agility and effectiveness when dealing with scenarios with less supervision during the model training process (Ronneberger et al., 2015). The conventional implementation of U-Net, which typically employs a binary cross-entropy as the loss function in binary classification problems, encounters a limitation when applied to AFB delineation in regions like Cambodia. This limitation arises from the presence of fragmented fields that are very small in size and irregular in shape, resulting in the generation of partial contours in the resultant prediction image. In order to address this constraint, we have incorporated U-Net with an additional loss function, which we refer to as Contour-Loss, as inspired by a previous study (()Yuan et al., 2022).

Although the U-Net architecture-based CNN network is faster and more efficient in situations with weak supervision for training the model (Ronneberger et al., 2015), The use of U-Net for AFB delineation for areas such as Cambodia has a drawback due to fragmented fields of small size the results consist of partial or open contours. So, to overcome this limitation, we aim to integrate U-Net with another loss function and we call it Contour Loss in our research borrowed from (Yuan & Xu, 2022).

Contour Loss is a loss function built on the basic principle of attention loss, and it computes a weighted average value of cross entropy of all pixels in a batch. A batch necessarily means the neighborhood of each pixel in an image. Each pixel is evaluated based on its eight neighboring pixels on a binarized image of boundary predictions, and if the value of all eight neighboring pixels is 1, then the value of the pixel in scope is set to 1, which means the pixel is an endpoint. With such an algorithm, we can ensure that the gaps in the contours or boundaries can be appropriately

connected. Such a loss function has been implemented in a study to extract road line segments from satellite data and has outperformed in comparison with other loss functions (Yuan & Xu, 2022).

$$\text{Contour Loss} = \frac{\sum_{i=1}^{N} w_i \cdot \text{Binary Cross Entropy}(y_i, p_i)}{\sum_{i=1}^{N} w_i} \tag{3.1}$$

Where:

$$\begin{aligned} N &: \text{Number of samples in the batch} \\ w_i &: \text{Weight assigned to sample } i \\ y_i &: \text{True label of sample } i \\ p_i &: \text{Predicted probability of positive class for sample } i \\ \text{Binary Cross Entropy}(y_i, p_i) &: \text{Binary cross entropy loss for sample } i \end{aligned}$$

The algorithm of the loss function is illustrated in algorithm 1.

**Algorithm 1:** Contour Loss

**Data:** Prediction of remote image by network
**Result:** Loss function value of Contour Loss
Process the input by the softmax function to acquire the cross-entropy map $L$;
Binarize the input to acquire image $A$;
Obtain skeleton image $B$ from $A$;
Generate a matrix $C$ with the same size as $B$ and fill it with zeros;
**for** *pixel in* $B$ **do**
    **if** *There is only one pixel with a value of 1 in the 8 neighboring pixels* **then**
        Set the corresponding pixel in $C$ to 1;
Generate a matrix $W$ with the same size as $B$ and fill it with ones;
**for** *pixel in* $C$ **do**
    Obtain the sum of pixel values in the 9x9 neighborhood as $N$, where $N$ is the number of endpoints in the neighborhood;
    **if** $N \neq 0$ **then**
        Set the corresponding pixel in $W$ to $K \times N$, where $K$ is a super parameter;
**return** The mean of $W \times L$;

The U-Net is typically built with cross-entropy loss as a loss function and is not sufficient to obtain boundaries with closed segments. The contour Loss function built on the principle of attention loss can improve the accuracy of the existing U-Net model for AFB delineation.

### 3.3 POST-PROCESSING TO POLYGONIZE PREDICTIONS

Post-processing is the phase in an image analysis pipeline where additional steps are performed on the output of the CNN model to refine the results or extract meaningful information. Through

this research, we have conducted experiments on predictions using graph-based segmentation followed by contour extraction in-order to polygonize the model's predictions. Process followed to polygonize predictions a post-processing approach is described in the workflow in figure 3.5

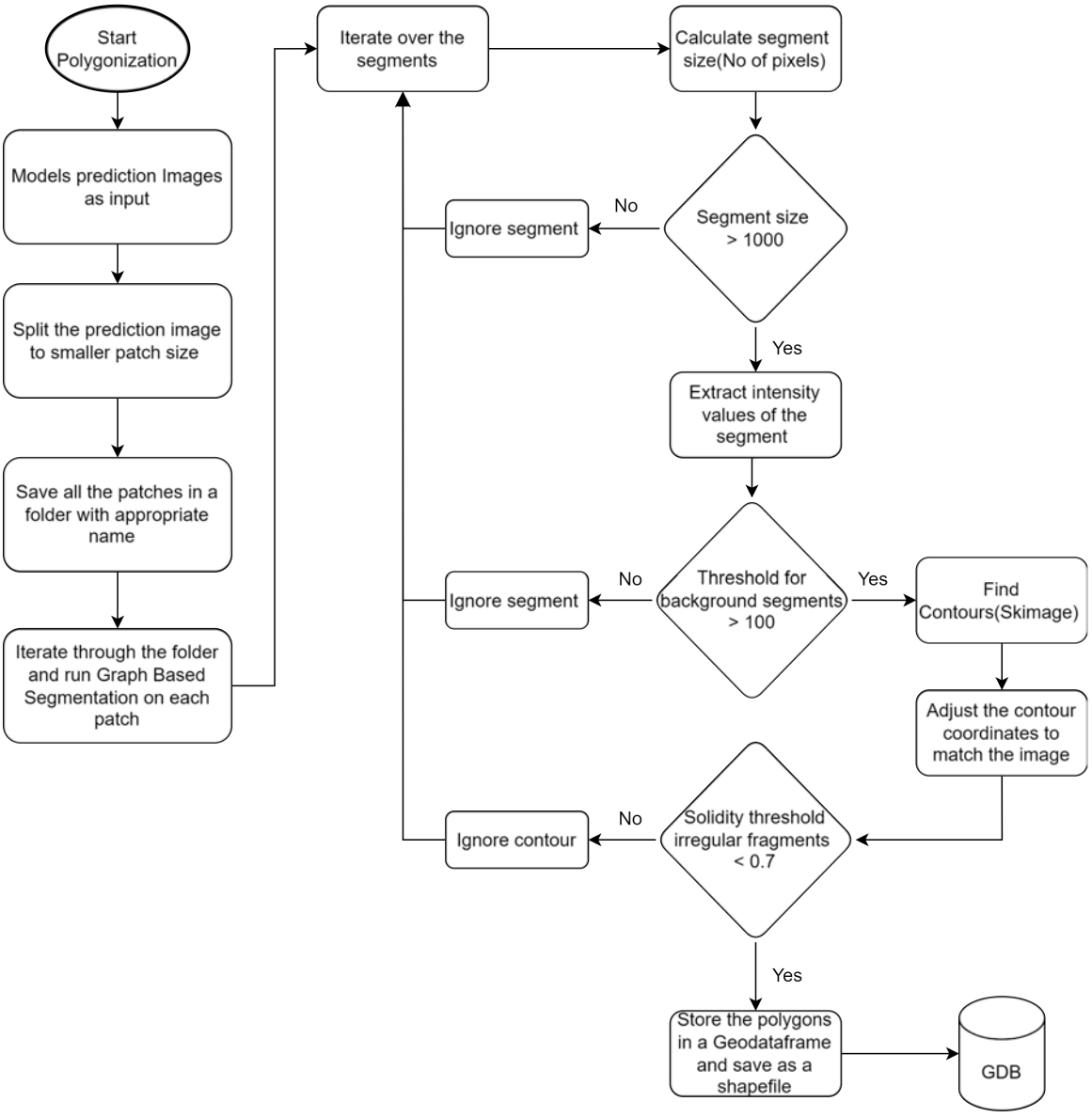



Figure (3.5) Workflow Graph-Based Segmentation for Polygonization

Graph-based segmentation based on the Felzenszwalb algorithm groups pixels into segments based on edge weights and applies a clustering algorithm in a bottom-up manner. The algorithm initially starts to find pixels in the input images and construct a graph with edges weighted based on the dissimilarity between pixels intensity values or spatial proximity. The most common disadvantage with graph-based segmentation methods other than the Felzenszwalb algorithm is that the segments created are either too fine or coarse. The research objective to polygonize boundary predictions can be best performed using Felzenszwalb's algorithm due to the predicate condition property of the algorithm that can effectively differentiate boundaries amongst the segments

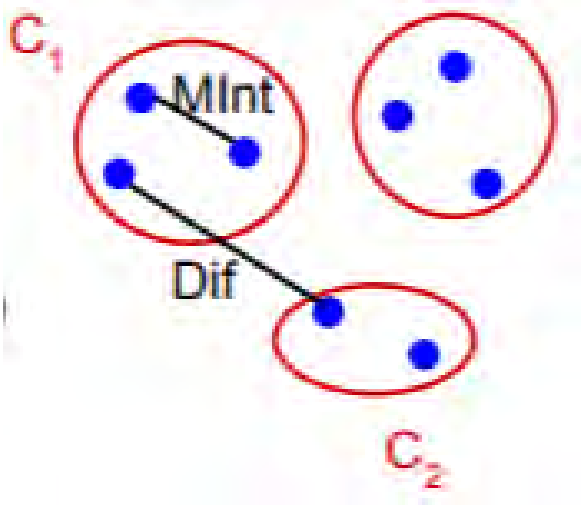



Figure (3.6) Predicate illustration

(Felzenszwalb & Huttenlocher, 2004). The predicate condition in equation 3.2 proposed by the author in (Felzenszwalb & Huttenlocher, 2004) (which is also known as pair-wise comparison predicate) compares the difference between the two components to that of inter-component difference. Only if the difference between the components is higher than the inter-component intensities will the pixels be grouped in separate segments or components. This property, therefore, is adaptive to the local characteristics of the input image.

$$D(C1, C2) = \begin{cases} true & \text{if Dif(C1,C2)} > \text{MInt(C1,C2)} \\ false & \text{otherwise} \end{cases} \tag{3.2}$$

The above equation can be much more clearly understood based on the illustration in figure 3.6

The accuracy of the algorithm and the polygonization method is estimated using the Polis metric. The Polis metric is primarily designed to measure the similarity between two polygons based on the shape and accuracy differences between the polygons and the distance of their vertices. (Avbelj et al., 2015)

Polis distance metric is calculated by estimating the distance between two polygons (vertices to nearest edge/vertex) the distances calculated between two polygons in both directions are summed and then divided by the total number of vertices of the participating polygons as visualized in 3.7 (Avbelj et al., 2015)

The pseudocode for computing the Polis metric between two closed polygons, A and B is depicted in the algorithm 2

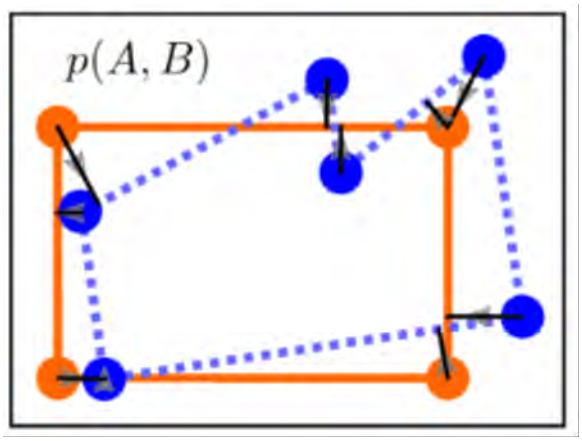



Figure (3.7) Polygon A(predicted polygon in orange) and B (Ground truth in blue) Black lines with grey arrow indicate the direction to estimate the distance between vertex and edge (Avbelj et al., 2015)

**Algorithm 2:** POLIS METRIC

**procedure** POLIS METRIC($\{A, B\}$)
  $p1, p2 \leftarrow 0$
  **for** $j = 1, ..., q$ **do**
  for every point $a$ in $A$
    $p1 \leftarrow p1 +$ MIN DIST PT2POLY$(a, B)$
  **for** $k = 1, ..., r$ **do**
  for every point $b$ in $B$
    $p2 \leftarrow p2 +$ MIN DIST PT2POLY$(b, A)$
  $p \leftarrow p1 + p2$
  **return** $p$ {POLIS distance value}
**end procedure**

# Chapter 4

# Results and Discussion

This research investigates the primary purpose of developing a customized workflow utilizing CNN-based models to accurately identify the AFB in regions such as Cambodia and Vietnam from Sentinel-2 EO images. The subsequent part presents the experimental outcomes derived from the implemented workflow in Figure 3.1 and Figure 3.5 from Chapter 3.

## 4.1 PRE-PROCESSING

Utilizing a full-year median composite image is a prevalent methodology owing to its efficacy in mitigating the influence of outliers. The computation of the median values for each band leads to the elimination of the spectral correlation present among the bands (Roberts et al., 2017). To further analyze the perseverance of spectral relationship by different composite methods, we have compared a cloud-free image from 06/09/2019 for Vietnam and 16/11/2020. Since the cloud-free image was available in 2019 and 2020, respectively, the comparison was also made on the seasonal images downloaded in the same years, except in the case of the yearly median composite. Spectral relationship-related graphs are depicted in Figure 4.2 and Figure 4.1. The spectral relationship between the bands is not preserved by median and medoid composite approaches, in contrast to the geometric median. A comparative analysis was conducted between the composites and cloud-free sentinel-2 EO images for the regions of Cambodia and Vietnam. Based on the spectral relationship graph depicted in Figure 4.1, it is evident that the spectral relationship between bands 2 (Blue) and 4 (Red), bands 2 (Blue) and 3 (Green), and bands 2 (Blue) and 8 (NIR) has experienced degradation in both the median and medoid composite as seen in Figure 4.1. However, the geometric median composite has effectively preserved this spectral relationship. Furthermore, it can be observed from the spectral relationship depicted in Figure 4.2 that the correlation between band 4 (Red) and band 8 (NIR) has deteriorated in both the median-based and medoid-based composites. However, this correlation is effectively maintained in the geometric median composite. Our research incorporates all three composites derived from the median, medoid, and geometric median. The prediction of CNN models and their corresponding accuracy are showcased in the subsequent section, utilizing images constructed from the median, medoid, and geometric median composite techniques, respectively.

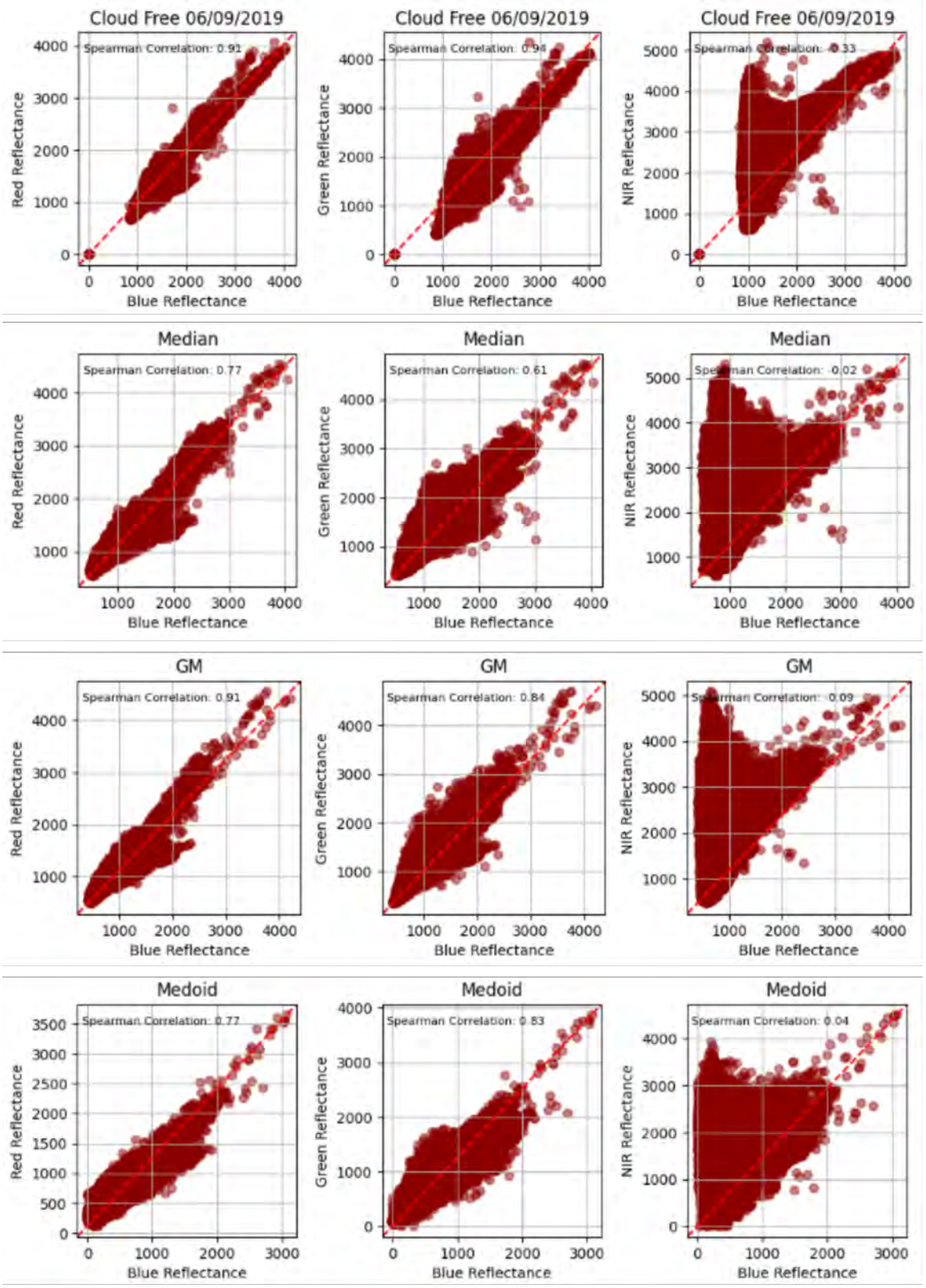



Figure (4.1) Spectral relationship Between bands of composites with the cloud-free image as reference for Vietnam.

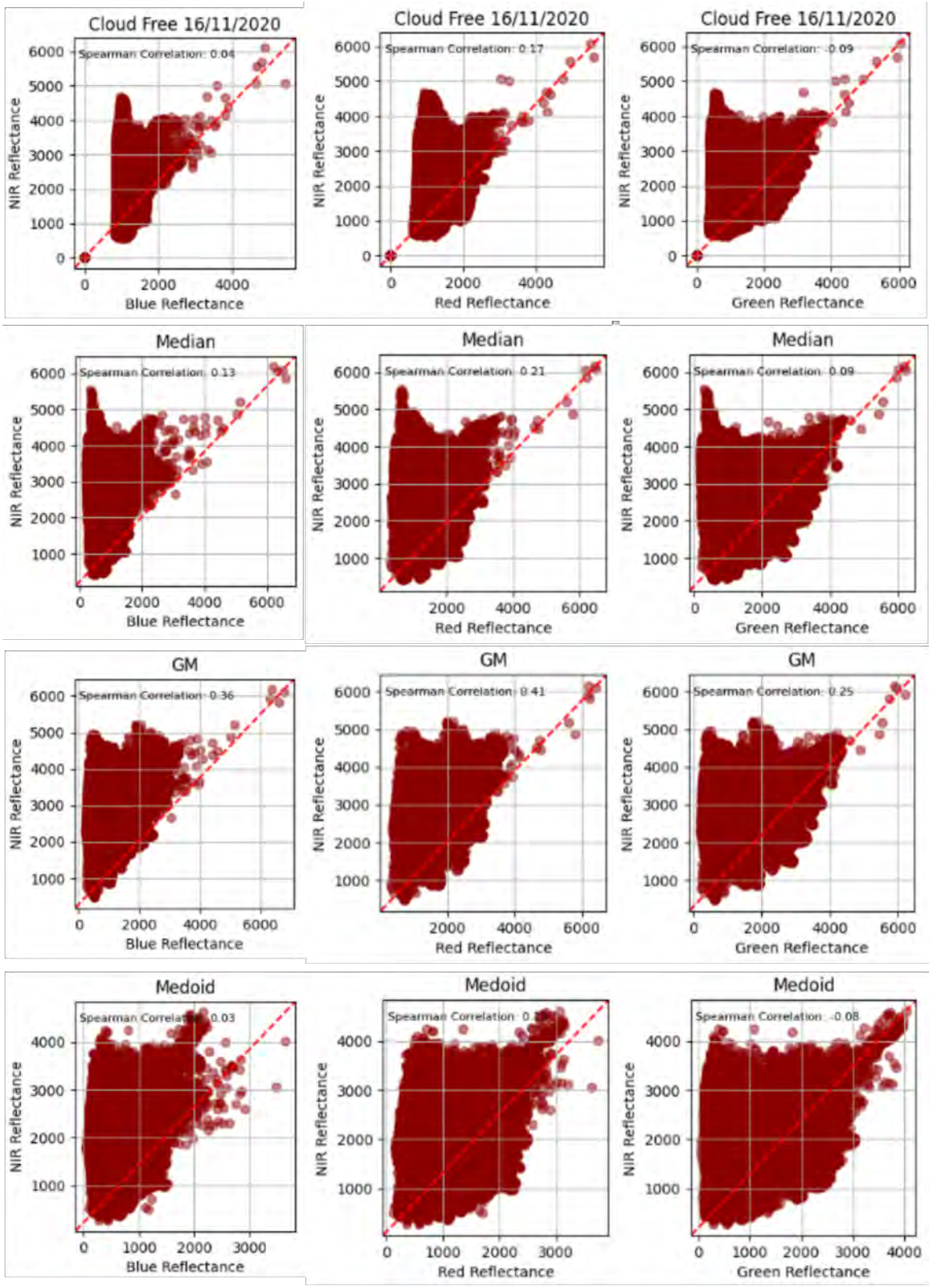



Figure (4.2) Spectral relationship Between bands of composites with the cloud-free image as reference for Cambodia.

Based on the preliminary investigation conducted in the preceding section, it has been determined that the geometric median is a composite method that effectively maintains the spectral correlation among bands. Consequently, we have devised forthcoming experiments based on the utilization of the geometric median as the foundation. Furthermore, a comparative study will be performed to evaluate the prediction outcomes and their respective accuracy, taking into account the medoid and median composites. The objective of this analysis is to identify the optimal composite method that is suitable for both the study area and the workflow. In order to achieve this objective, a comparative analysis is conducted between a yearly geometric median composite and a seasonal geometric median composite. The aim of this comparison analysis is to assess the effectiveness of the seasonal composite technique in prioritizing the emphasis on boundaries. It is noteworthy that the emphasis on boundaries can be achieved more effectively through the presence of mature crops in CNN training images. Mature crops can also be indicative of healthy vegetation. From Figure 4.3, it is evident that using a seasonal-based composite captures more healthy vegetation in our study area, such as mature harvest-ready crops. The utilization of seasonal composite-based EO images might facilitate the examination of crops, enabling the computation of indices like the Normalized Difference Vegetation Index (NDVI) or Leaf Area Index (LAI) for subsequent crop analysis. Additionally, the utilization of seasonal composite-based EO images might lead to fewer false positives for agricultural field boundaries in comparison with a yearly observation composite.

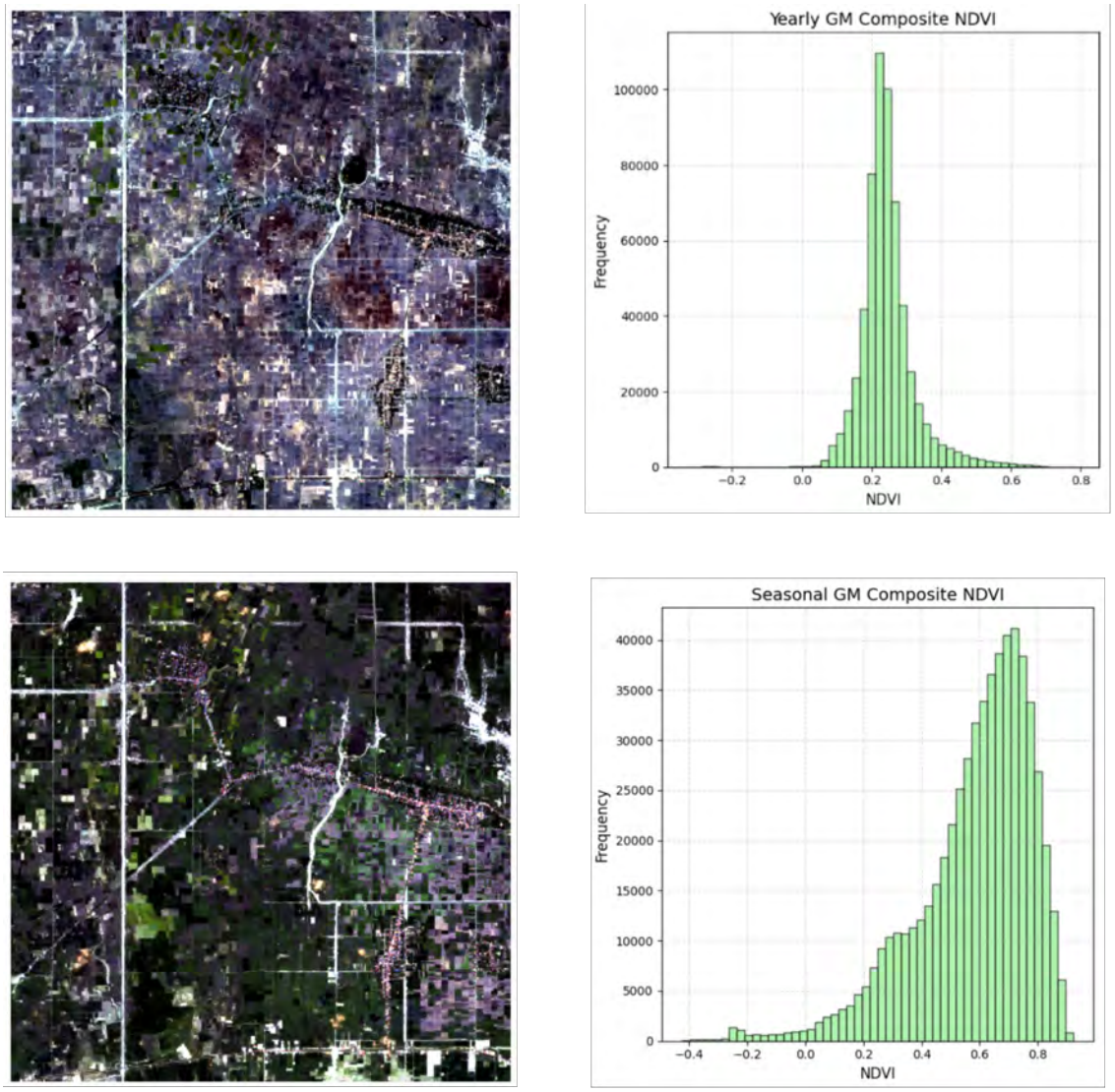



Figure (4.3) Yearly GM Composite(Top), Seasonal GM Composite(Bottom) For Cambodia

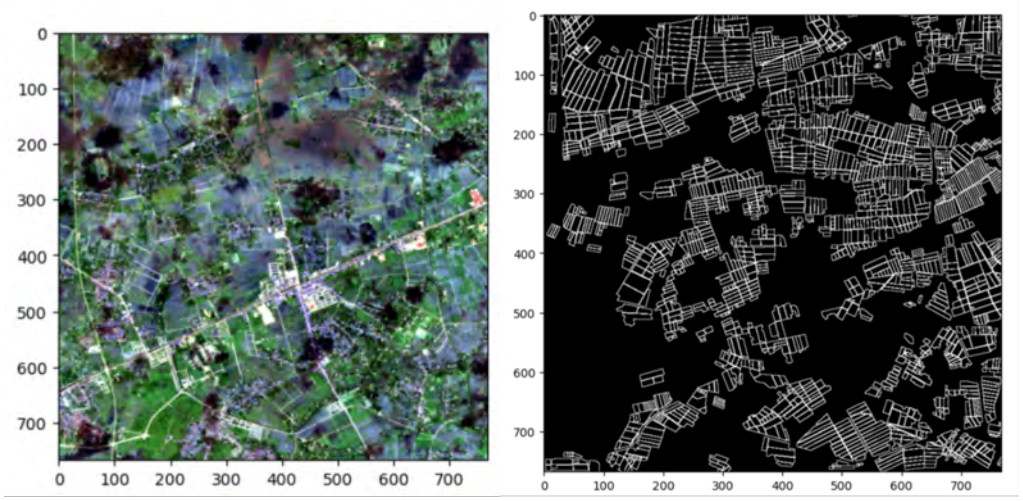
The deep learning U-Net architecture-based CNN model is utilized to make predictions regarding agricultural field boundaries in the research areas of Cambodia and Vietnam. Now, we present an analysis of the accuracy of prediction derived from trials conducted using three distinct composite approaches. The experimental findings are presented in Table 4.10. Our objective is to perform a comparative examination and discussion on the discrepancies in field prediction results. Specifically, we will concentrate on the variations between a full-year median composite and seasonal geometric median and seasonal medoid composites. The aforementioned trials, in particular, will be discussed for the wet season, spanning from September to the conclusion of November, in the regions of Cambodia and Vietnam. The months selected for the wet season are outlined in the provided table 3.1; This selection has been made based on the foundational information gained from the crop calendars in 3.3 and 3.2 Nevertheless, the outcomes of the seasonal median-based composite and other experimental trials such as the ones performed for Dry season and the respective composites, along with their related accuracy measured by the F1 score, have

been documented in the table 4.10. In this section and in its discussions, we only discuss the outcomes of trials per composite method per region for the wet season because cloud noise is more prevalent in the wet season, and hence, the overall accuracy of prediction and polygonization of predicted polygons might further deteriorate for the wet season as compared to that of the dry season. Although we do not provide qualitative analysis for dry season-based composites and their respective prediction outcomes, their results documented in the table will be summarized at the end of this section.

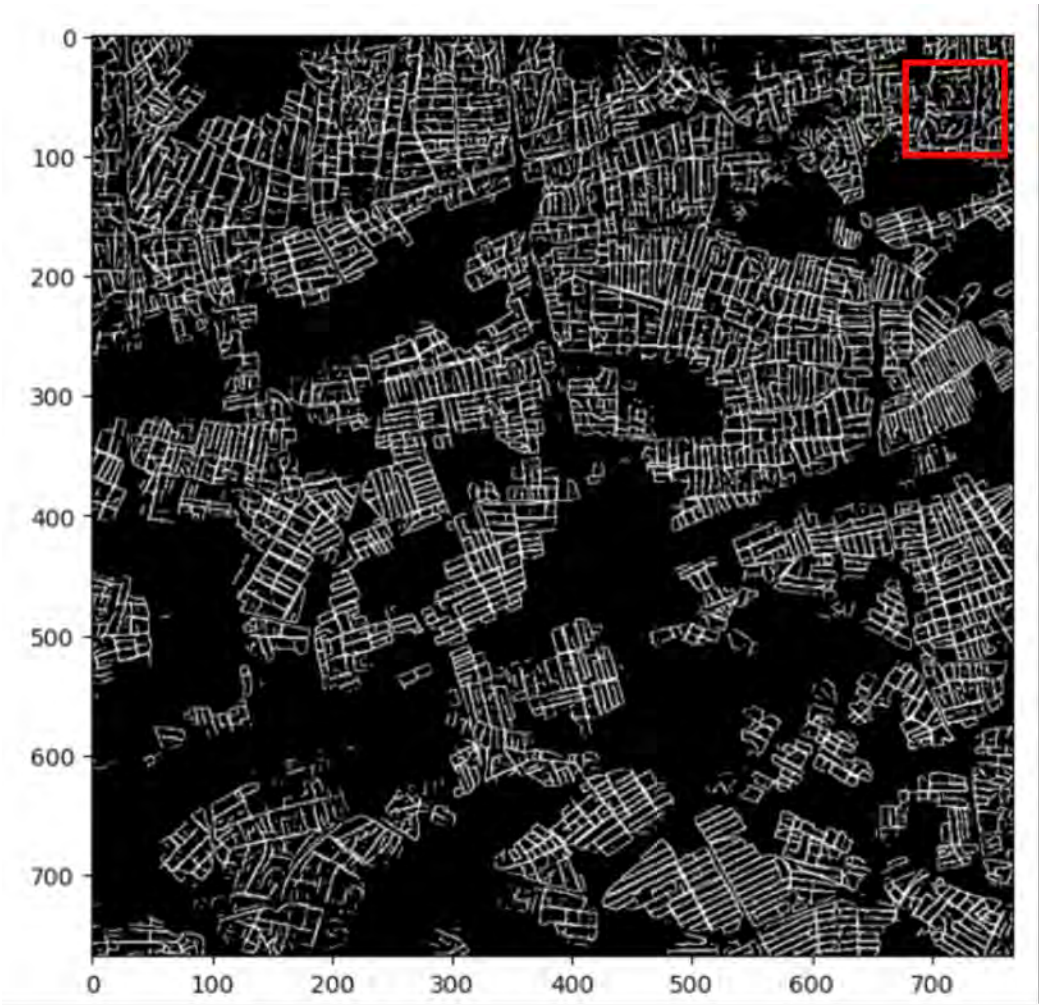
A trial was conducted involving the utilization of a seasonal geometric median composite for Vietnam as the input for training and then comparing their results with other trials for the Vietnam study area. The qualitative analysis of the training image, depicted in Figure 4.4a, revealed that the composite image still contains outliers. These outliers can be attributed to the presence of clouds in the region. Consequently, the model's training process produces prediction outcomes depicted in Figure 4.4c that exhibit partial contours(in red rectangle boxes used for highlighting) or imperfect boundaries when compared to the ground truth in Figure 4.4b, particularly in regions where cloud noise is apparent in the training image. The accuracy assessment of the trial is as represented in Table 4.2. We will be particularly interested in the F1 score because the overall score might not be the right representation of accuracy because of the majority of the presence of background-based pixels (Non-Agricultural field boundaries), which is also documented in the confusion matrix in Table 4.1. Prediction of Agricultural Field Boundaries(AFB) for the trial with seasonal geometric median-based composite for Vietnam yields an F1-score of approximately 65%.

There are two interesting findings for the composites built using seasonal observations for the Vietnam study area. First is the result of a comparison between the seasonal geometric median and that of the seasonal medoid. When comparing the prediction results of the seasonal geometric median composite-based trial with those of the seasonal medoid-based trial, it is observed that the prediction efficiency of the latter is lower than that of the former. The F1 score obtained for the seasonal medoid-based prediction is 56%. Additionally, the training images in Figure 4.6a exhibit unexpected black pixel masks. These black patches are a result of cloud noise and are partially because of the complexity involved in the computation of medoid composite, specifically in the selection of a true observation pixel that is closest to the median pixel value. The literature research reveals that the median composites exhibit sensitivity towards outliers. Consequently, the computation of the Medoid incorporates the median pixel value. This limitation of radiometric inconsistency in a medoid composite can be effectively addressed by building a medoid composite from observations spanning many months to a year, which enables the efficient removal of outliers by means of the median pixels but might not include the phenological changes of crops and not capture mature crops for boundary delineation. The Second finding is the result of a comparison between the trial based on the seasonal geometric median composite and the trial based on the seasonal median composite for predicting AFB. The analysis reveals that the predictive accuracy of the latter method is comparatively lower than that of the former method. Furthermore, it is the least effective among all the trials conducted using seasonal observations, resulting in a low F1 score of only 48%. This information can be found in Table 4.5. This low score is backed by the ineffectiveness of the seasonal median composite technique in suppressing outliers due to less number of observations available in a seasonal composite image as compared to a median composite image built from a year's observation. The influence of outliers in the seasonal median image is also apparent in the training procedure and illustrated in the validation loss graphs shown in Figure 4.10b and its corresponding gradient descent. It is important to observe that the model is exhibiting over-fitting. The observed trend in the Vietnam seasonal median composite, as depicted in Figure 4.10b, reveals an increase in validation losses and a decrease in training losses. Due to the

presence of outliers, the model exhibits a lack of generalization to unseen data.

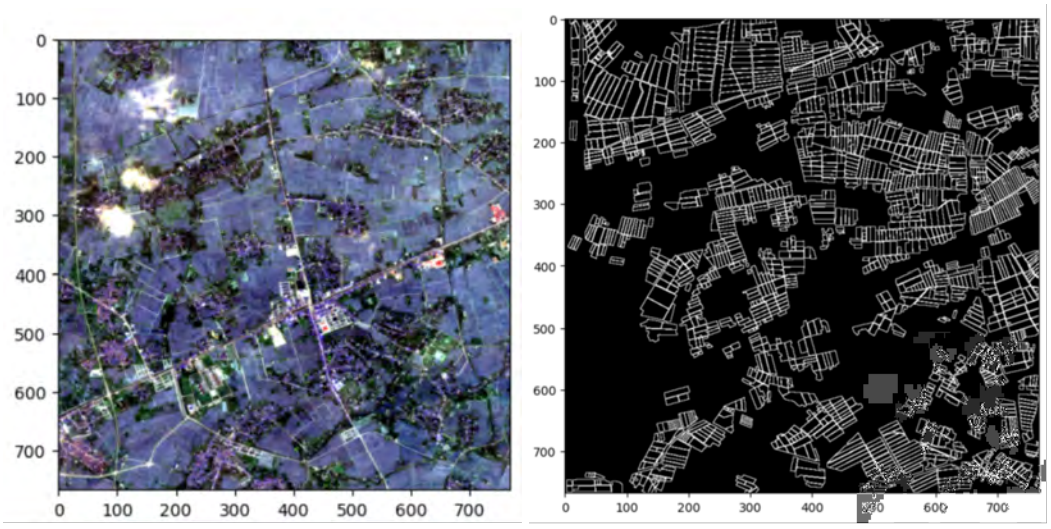



(a) Training Image (b) Ground Truth

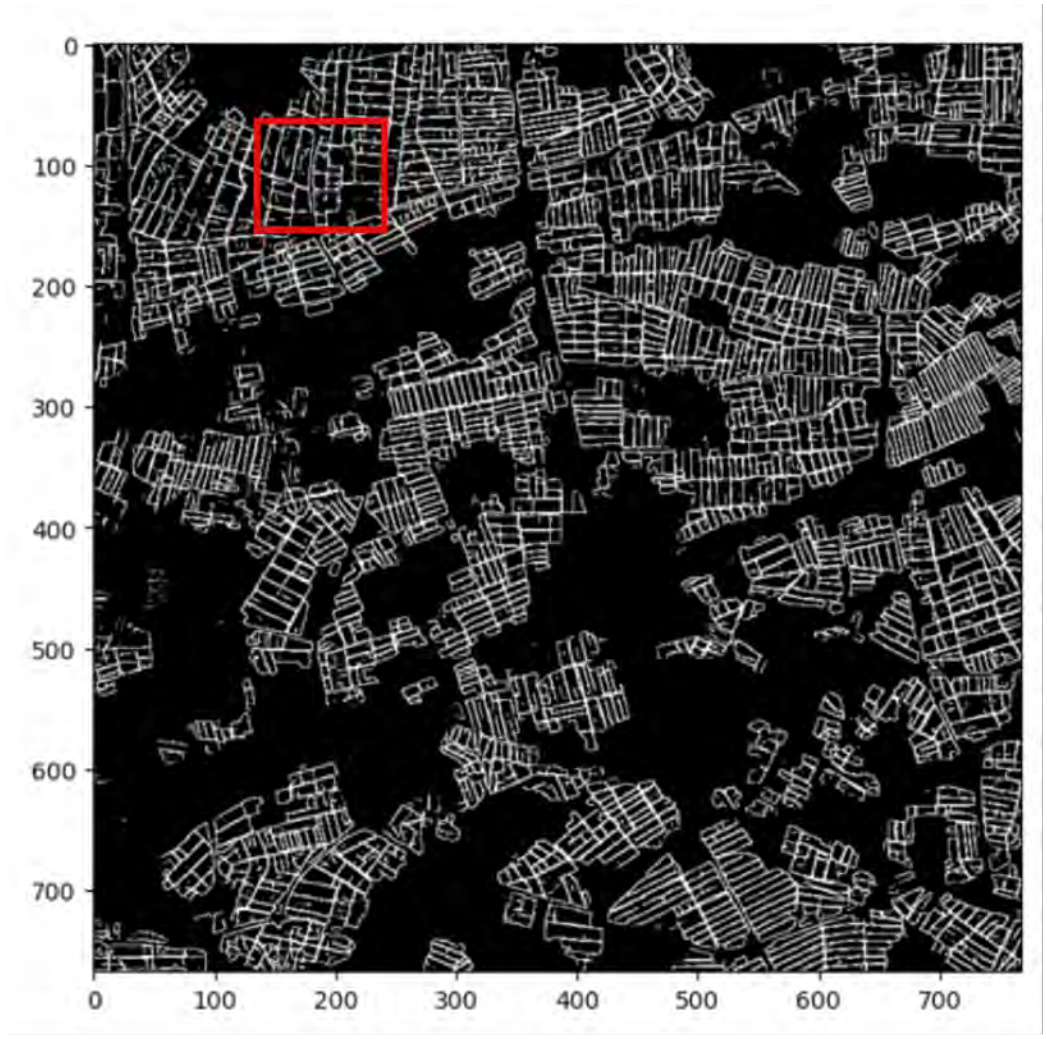



(c) Prediction

Figure (4.6) Prediction of AFB for Vietnam based on Seasonal Medoid Composite.

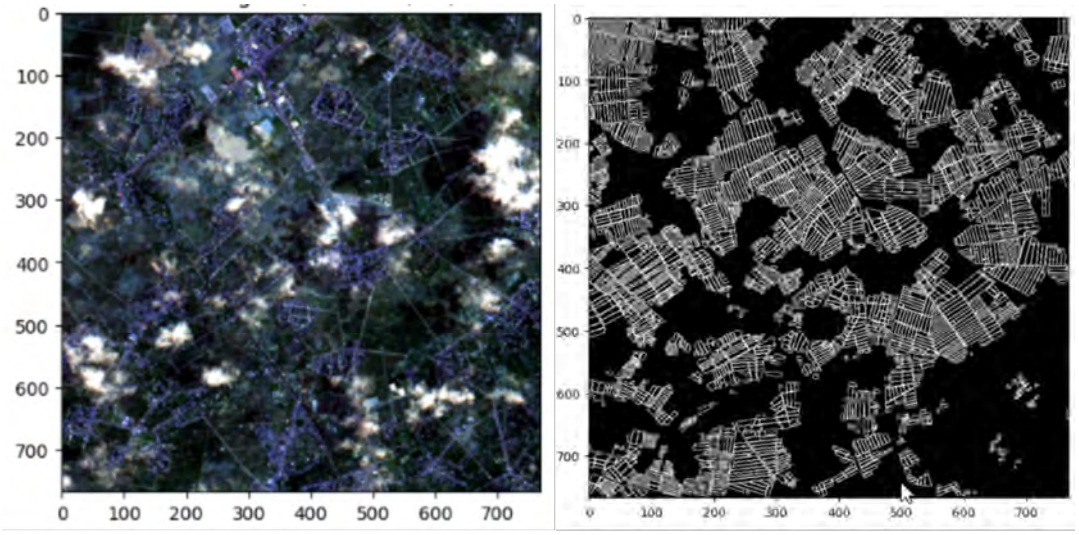



(a) Training Image (b) Ground Truth

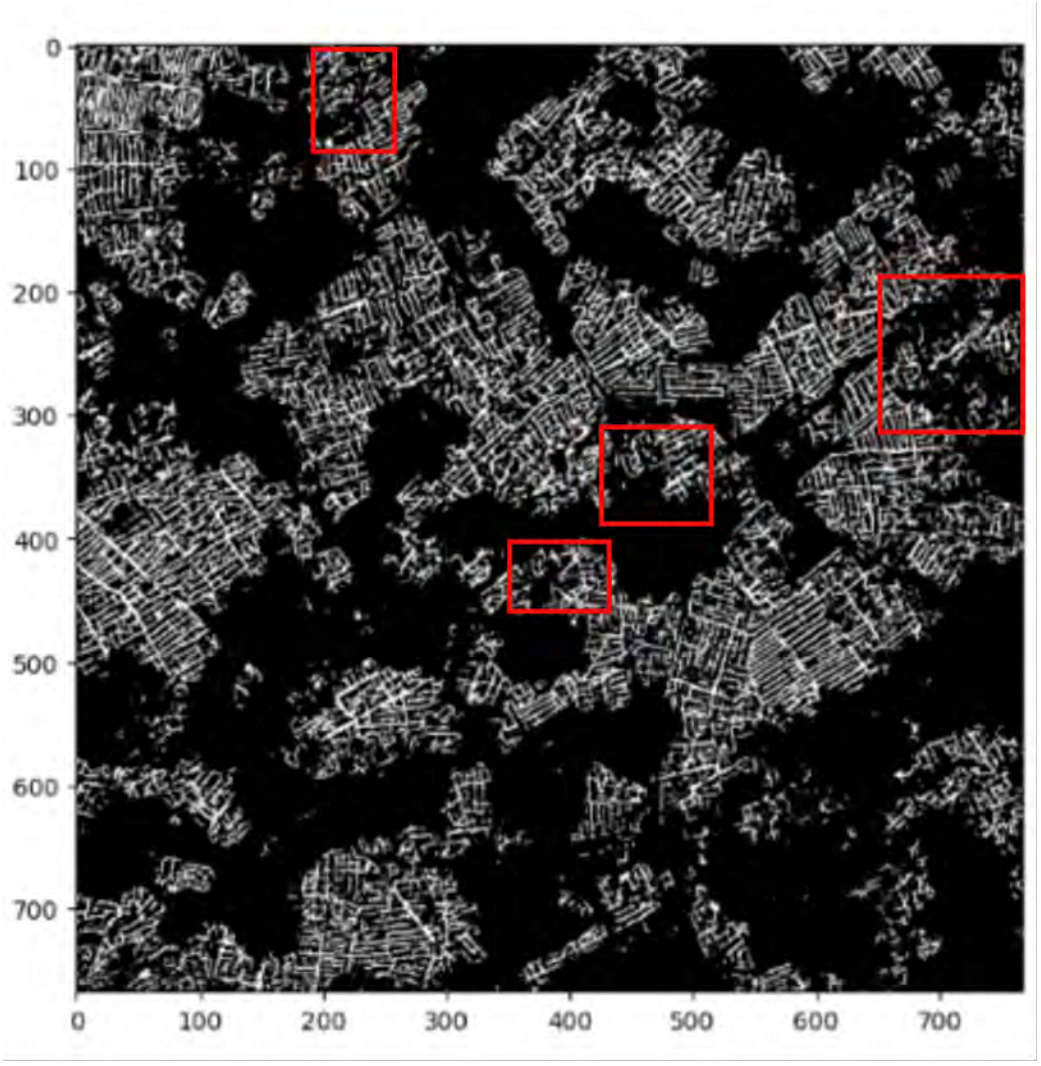



(c) Prediction

Figure (4.4) Prediction of AFB for Vietnam based on Seasonal Geometric Median Composite.




(a) Training Image (b) Ground Truth




(c) Prediction

Figure (4.5) Prediction of AFB for Vietnam based on Seasonal Median Composite.

**Table 4.1** Confusion Matrix for Vietnam based on Seasonal Geometric Median Composite.

| | Prediction Other | Prediction Field Boundary | Sum |
|---|---|---|---|
| Actual Other | 463590 (78.598%) | 25707 (4.358%) | 489297 |
| Actual Field Boundary | 39240 (6.653%) | 61287 (10.391%) | 100527 |
| Sum | 502830 | 86994 | |

**Table 4.2** Accuracy Assessment for Vietnam based on Seasonal Geometric Median Composite.

| Accuracy Assessment | Percentage |
|---|---|
| Overall Accuracy | 88.989% |
| Precision | 60.966% |
| Recall | 70.45% |
| F1 Score | 65.365% |

The research region in Cambodia was subjected to trials involving the use of a seasonal geometric median composite. The results of these trials, together with their implications for boundary delineation, are depicted in Figure 4.7. The trial results in an accuracy rate of 54.734%. The level of accuracy achieved in this study surpasses that of all other trials conducted in the Cambodia study area. Through qualitative examination, it is evident that the regions displaying incomplete boundaries(Highlighted in red boxes) in the model's prediction image in Figure 4.7c for Cambodia, as depicted in Figure 4.7, align with the locations where cloud formations are observed in the training Images in Figure 4.7a. In this discussion, we will now proceed to evaluate and compare the results obtained from a trial utilizing a seasonal geometric median composite with those obtained from trials utilizing a seasonal medoid-based composite and a seasonal median composite. The trial, including the use of seasonal medoid composite, has resulted in an accuracy rate of 48% as seen in Table 4.9. Qualitative analysis of the trial's training image in Figure 4.9a indicates the presence of radiometric inconsistencies. The primary factor contributing to the reduced precision observed in a medoid composite, in comparison to a seasonal geometric median, can be attributed to the formation of dark patches caused by the presence of clouds and the computation of the medoid. On the other hand The accuracy of the seasonal median composite is approximately 54% and is depicted in the Table 4.8. This accuracy is comparable to the findings obtained from the seasonal geometric median. The trial concluded that the accuracy of the seasonal geometric median composite was higher compared to the seasonal median composite due to the latter's sensitivity to outliers and its limited ability to remove noise efficiently. This finding also is evidence that the geometric median is more robust when compared to median.

**Table 4.3** Confusion Matrix for Vietnam based on Seasonal Medoid Composite.

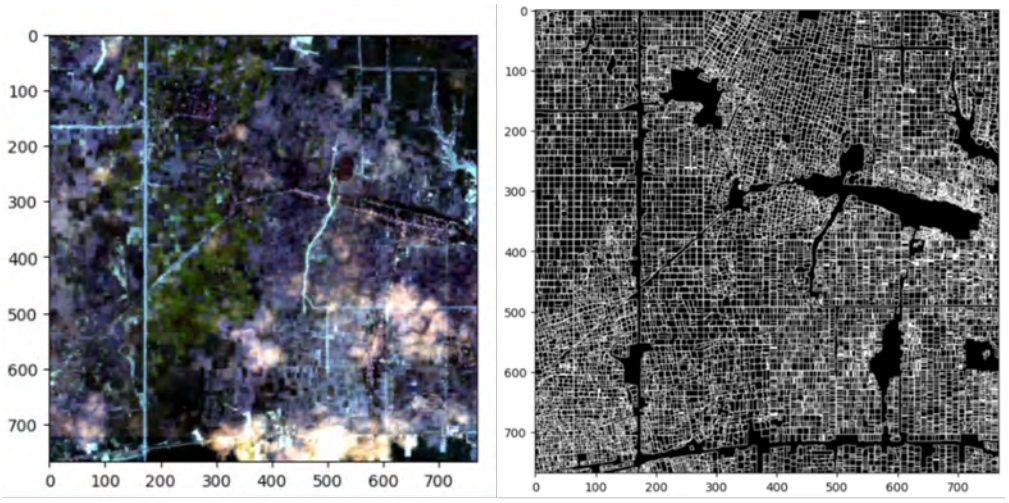
| | Prediction Other | Prediction Field Boundary | Sum |
|---|---|---|---|
| Actual Other | 456880 (77.46%) | 34742 (5.89%) | 491622 |
| Actual Field Boundary | 45950 (7.79%) | 52252 (8.859%) | 98202 |
| Sum | 502830 | 86994 | |

**Table 4.4** Accuracy Assessment for Vietnam based on Seasonal Medoid Composite.

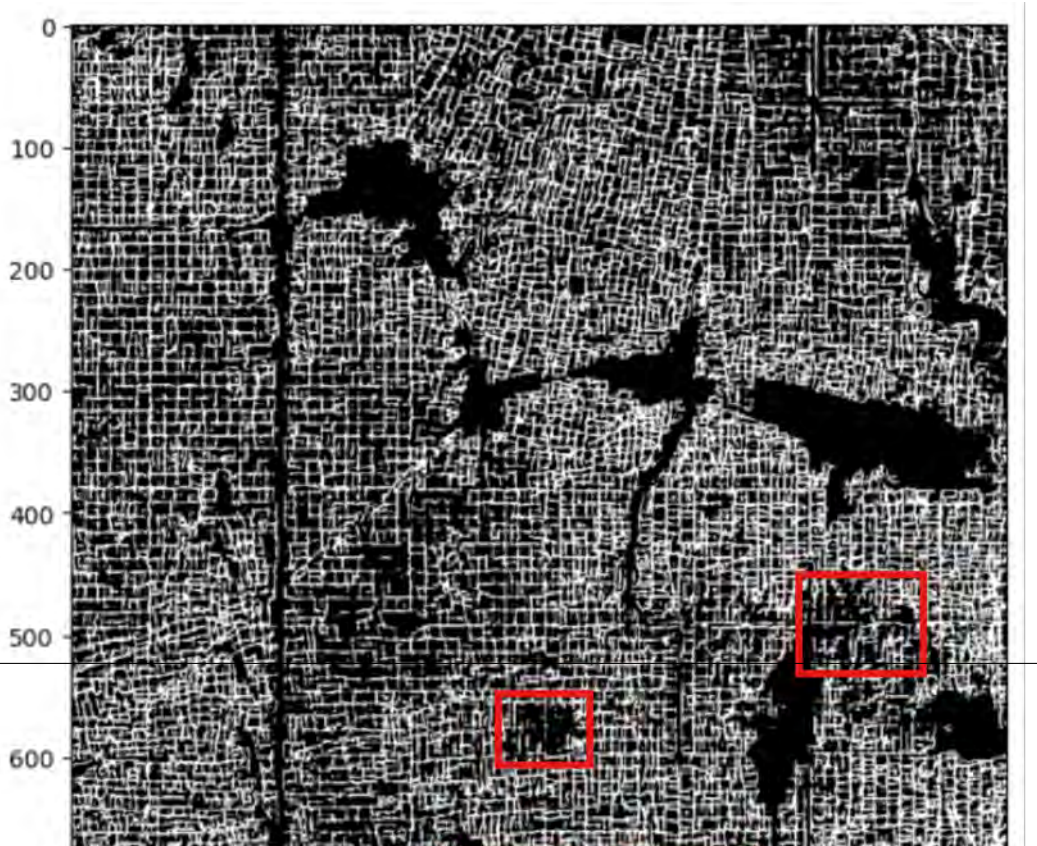
| Accuracy Assessment | Percentage |
|---|---|
| Overall Accuracy | 86.319% |
| Precision | 53.209% |
| Recall | 60.064% |
| F1 Score | 56.429% |

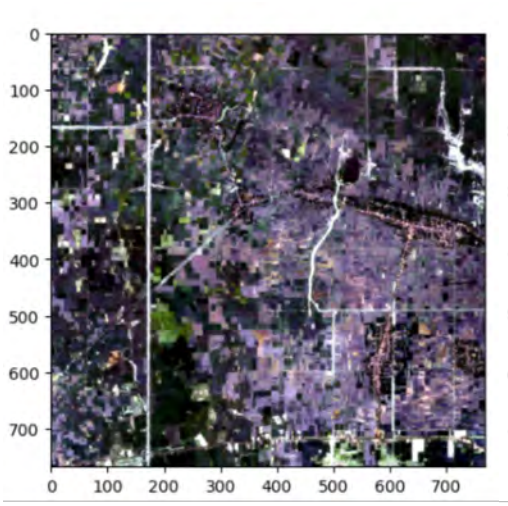
**Table 4.5** Accuracy Assessment for Vietnam based on Seasonal Median Composite.

| Accuracy Assessment | Percentage |
|---|---|
| Overall Accuracy | 81.274% |
| Precision | 55.136% |
| Recall | 43.198% |
| F1 Score | 48.443% |

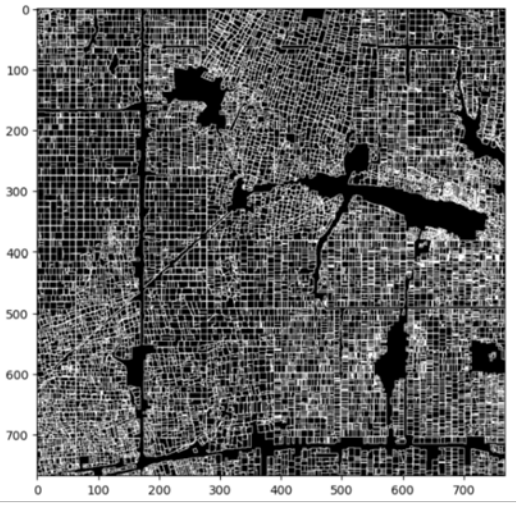



(a) Training Image (b) Ground Truth

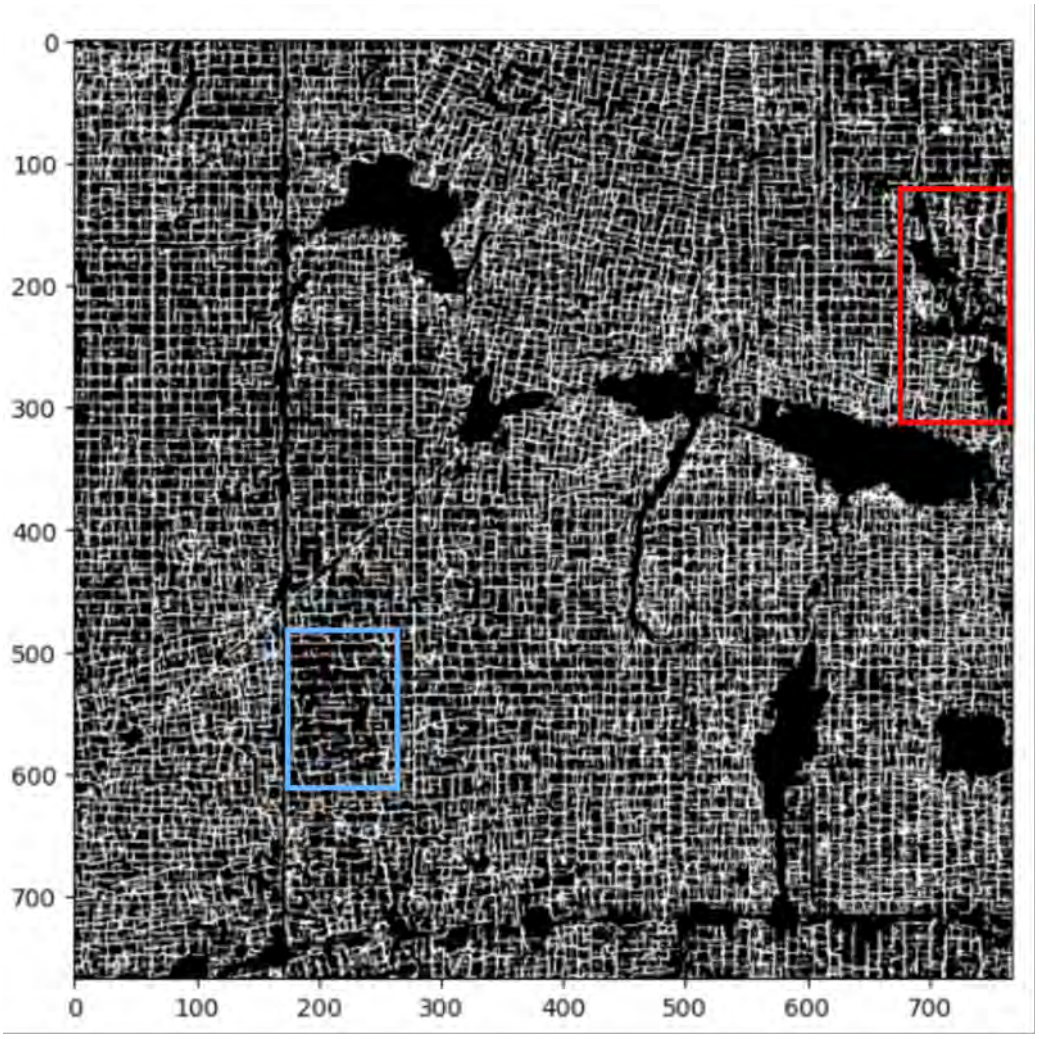





(a) Training Image

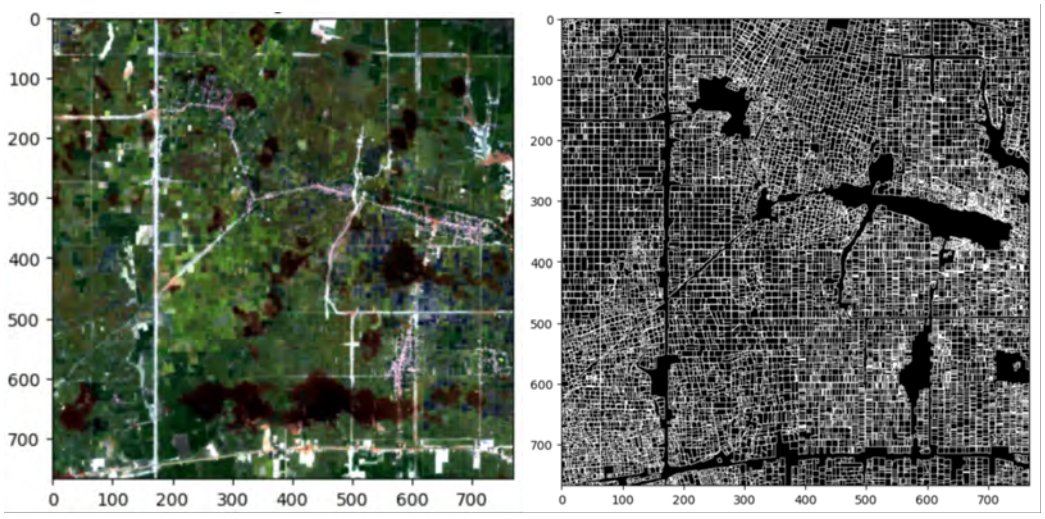



(b) Ground Truth

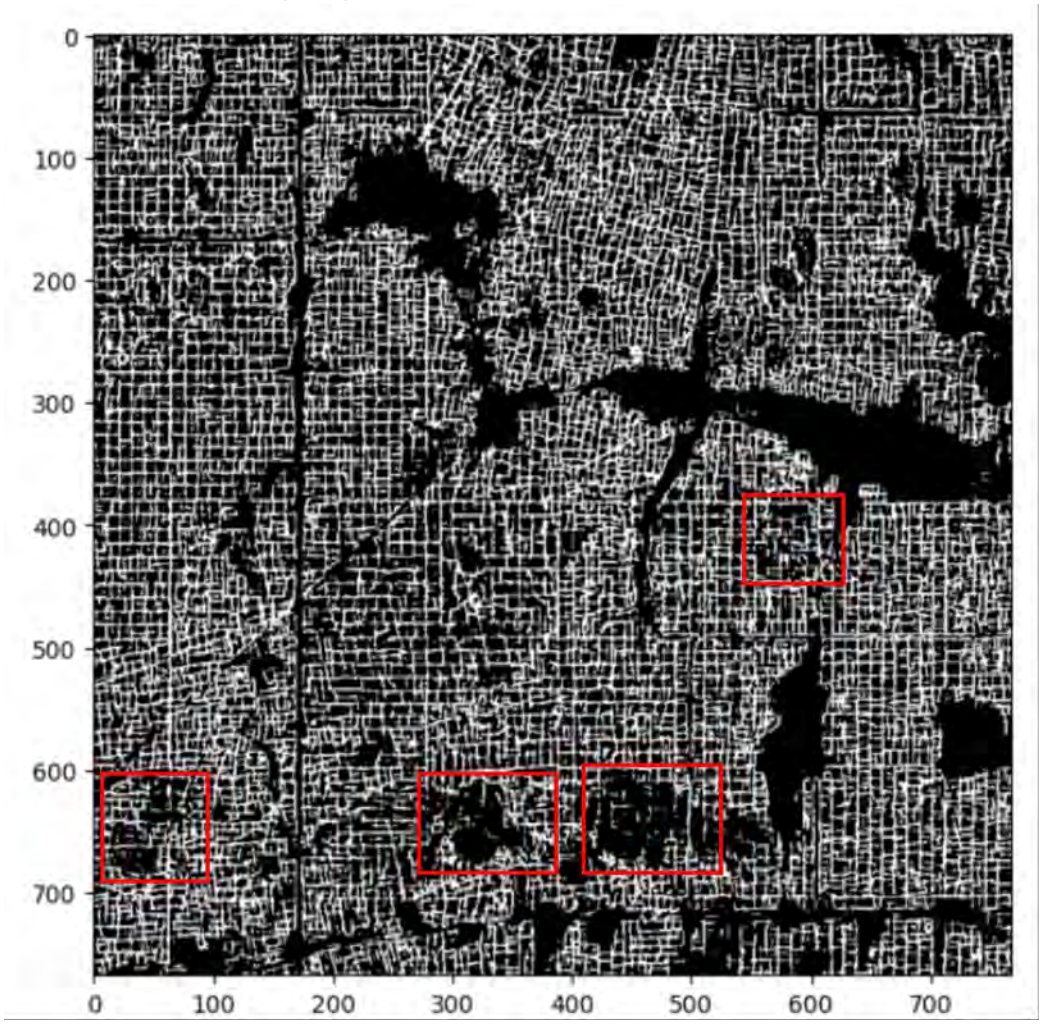



(c) Prediction

Figure (4.8) Prediction of AFB for Cambodia based on Seasonal Median Composite.




(a) Training Image (b) Ground Truth




(c) Prediction

Figure (4.9) Prediction of AFB for Cambodia based on Seasonal Medoid Composite.

**Table 4.6** Confusion Matrix for Cambodia based on Seasonal Geometric Median Composite.

| | Prediction Other | Prediction Field Boundary | Sum |
|---|---|---|---|
| Actual Other | 309303 (52.44%) | 91166 (15.45%) | 400469 |
| Actual Field Boundary | 83660 (14.18%) | 105695 (17.92%) | 189355 |
| Sum | 392963 | 196861 | |

**Table 4.7** Accuracy Assessment for Cambodia based on Seasonal Geometric Median Composite.

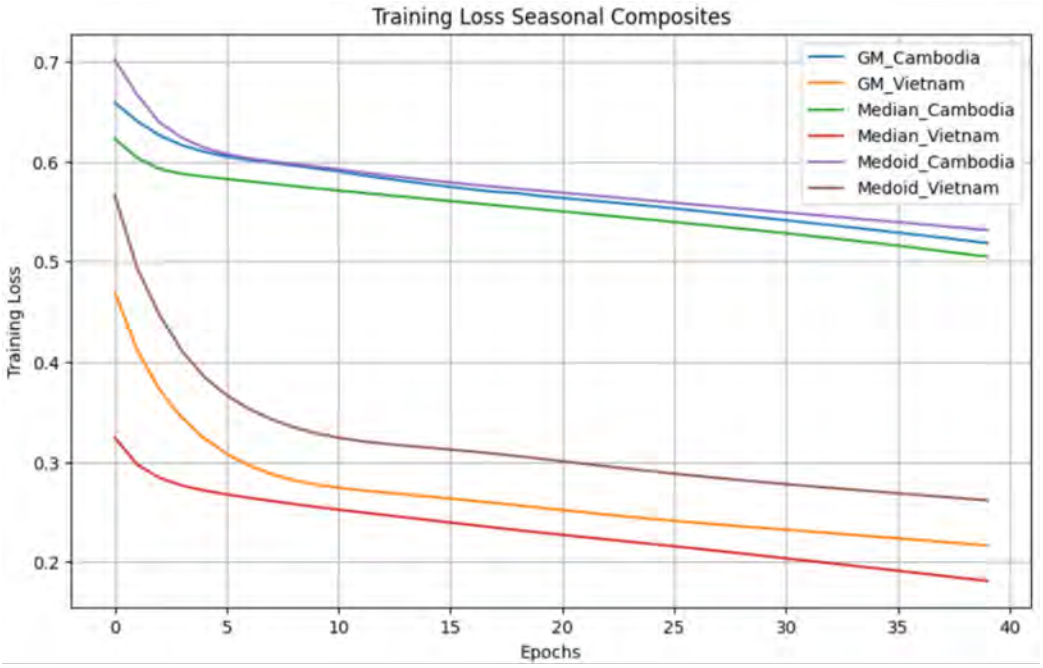
| Accuracy Assessment | Percentage |
|---|---|
| Overall Accuracy | 70.36% |
| Precision | 55.81% |
| Recall | 53.69% |
| F1 Score | 54.734% |

**Table 4.8** Accuracy Assessment for Cambodia based on Seasonal Median Composite.

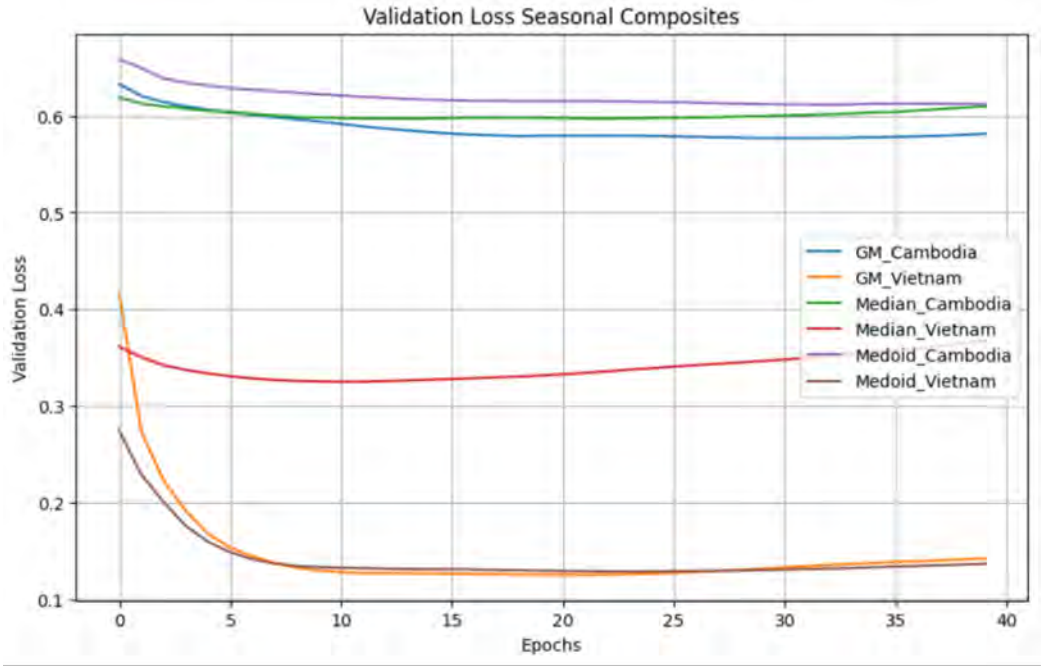
| Accuracy Assessment | Percentage |
|---|---|
| Overall Accuracy | 71.11% |
| Precision | 57.257% |
| Recall | 53.025% |
| F1 Score | 54.06% |

**Table 4.9** Accuracy Assessment for Cambodia based on Seasonal Medoid Composite.

| Accuracy Assessment | Percentage |
|---|---|
| Overall Accuracy | 81.274% |
| Precision | 55.136% |
| Recall | 43.198% |
| F1 Score | 48.443% |

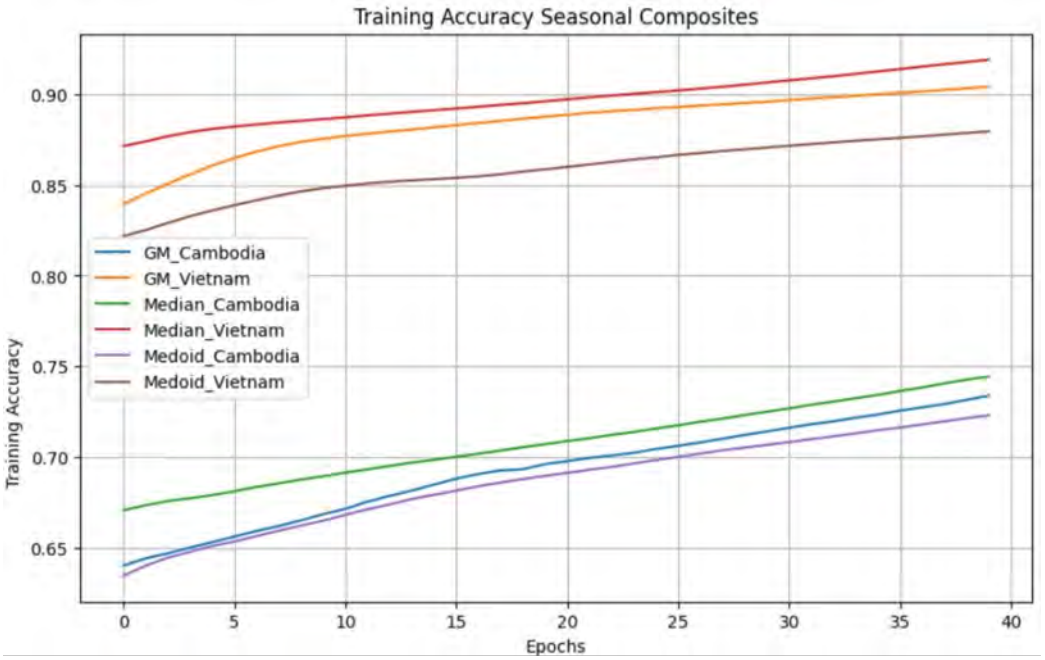



(a) Training Loss

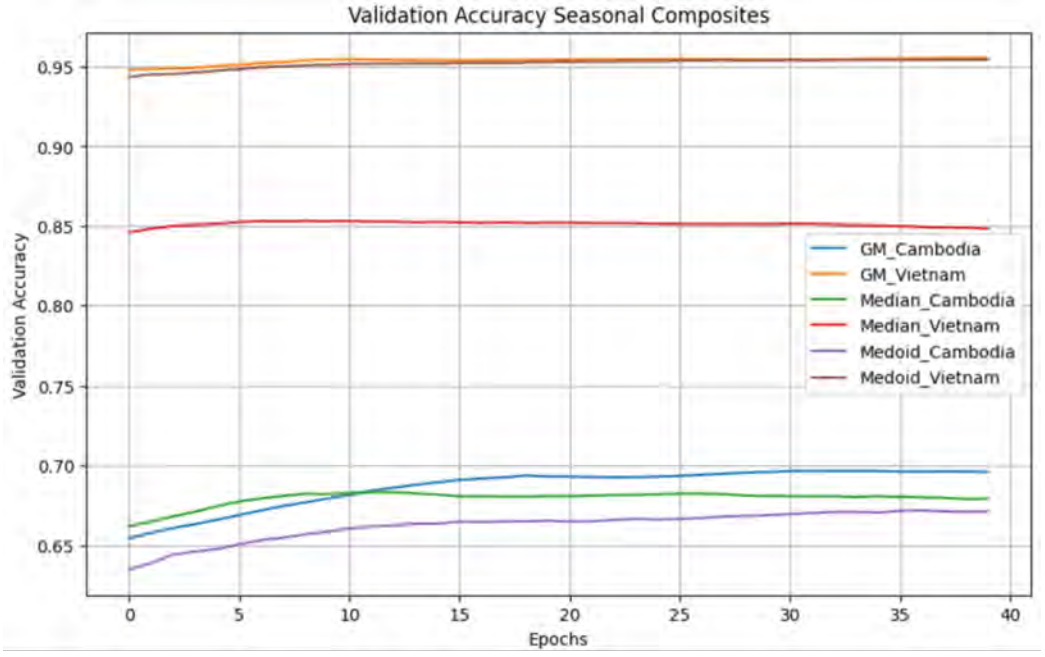



(b) Validation Loss

Figure (4.10) Training and Validation Loss for Cambodia and Vietnam




(a) Training Accuracy




(b) Validation Accuracy

Figure (4.11) Training and Validation Accuracy for Cambodia and Vietnam

**Table 4.10** Results summarized for all trials per Vietnam and Cambodia

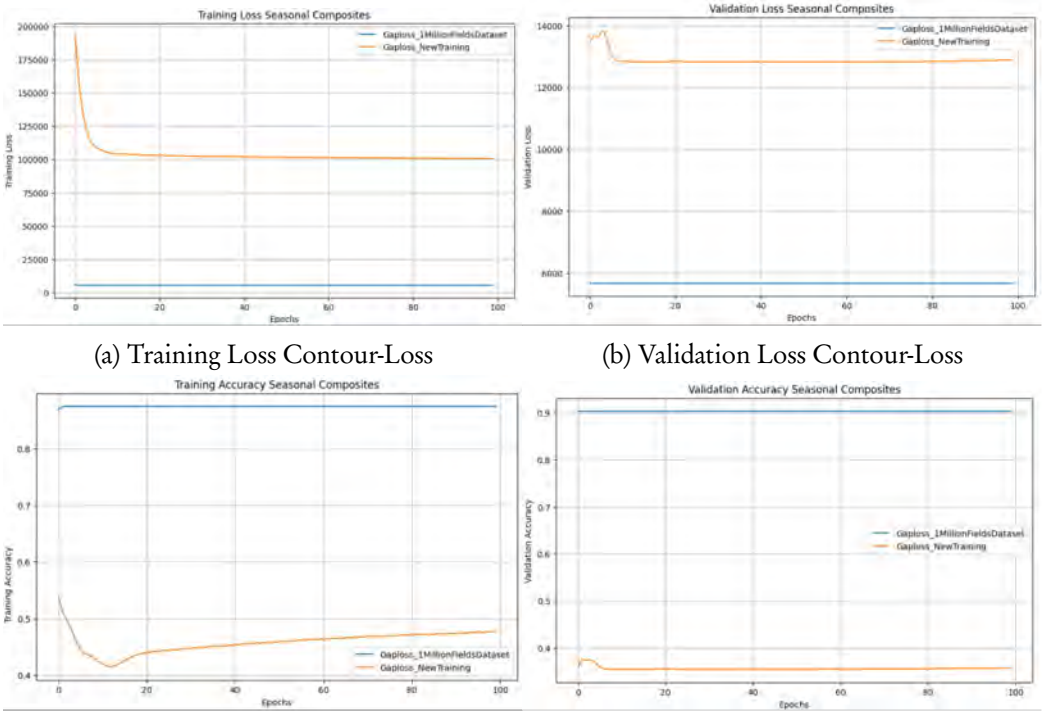
| Composite Type | Cambodia F1 Score | Vietnam F1 score |
|---|---|---|
| **Full Year Median** | 52 | 85 |
| **Dry Season Median** | 55 | 61 |
| **Dry Season GM** | 58 | 65 |
| **Dry Season Medoid** | 55 | 56 |
| **Wet Season Median** | 54 | 48 |
| **Wet Season GM** | 55 | 65 |
| **Wet Season Medoid** | 48 | 56 |

In the present study, we have made modifications to the epoch and learning rate parameters in order to achieve a harmonious equilibrium between the two variables. This adjustment is crucial for optimizing model training efficiency and mitigating the risk of over-fitting. As seen in the training and validation graphs, the U-Net model is exhibiting convergence for all of the seasonal compositing methods used except for the seasonal median composite built for Cambodia and Vietnam. All the trials for seasonal composites have been executed with 40 epochs and a learning rate of 0.01 and with Binary cross entropy as a loss function. The model tends to over-fit for the seasonal median composite-based trial for Cambodia and Vietnam due to the noise in the training images, because of which the model fails to generalize predictions for unseen data in the validation set. Also, this behavior is evident in the validation loss graph(red line in the middle and green line at the top of the graph) in Sub figure 4.10b within Figure 4.10.

Table 4.10 presents a thorough record of the conducted trials and their corresponding F1 scores. The results of the qualitative investigation suggest that the seasonal geometric median dataset has more favorable predictive performance for Vietnam and Cambodia, as supported by its higher F1 scores in comparison. Although the annual median composite fails to account for the phenological variations related to the crop and is vulnerable to outliers, it nonetheless yields the highest F1 score for Vietnam. In the case of Cambodia, it is observed that the use of the seasonal geometric median surpasses the approach involving the yearly median composite, resulting in a superior outcome. In comparison to the various composite methods employed in trials for both wet and dry seasons, it is observed that the F1 scores during the dry season are generally higher than those during the wet season. This discrepancy can primarily be attributed to the lower cloud cover experienced during the dry season in contrast to the wet season. Another intriguing discovery lies in the accuracy differences between the yearly median composite and the seasonal geometric median composite-based trials. The latter has exhibited evident enhancement in the delineation of boundaries, resulting in a score of 55% for the Wet season and 58% for the dry season. The observed improvement is limited to the research area in Cambodia, as it is characterized by uneven field shapes that present a challenge in differentiating between individual fields and their boundaries. When conducting a trial including the implementation of a seasonal Geometric Median Image, our focus is mostly on mature crops. This emphasis allows for better delineation of borders, resulting in superior quality outcomes compared to a median composite based on observations over the entire year.

## 4.2 MODEL BUILDING WITH CONTOUR-LOSS

The comparative quantitative analysis of prediction results in the previous section between Cambodia and Vietnam reveals a consistent pattern of lower accuracy for Cambodia, regardless of the statistical method employed to construct the composite. Additionally, a qualitative examination of the prediction images for Cambodia reveals the presence of partially closed boundaries or contours. The aforementioned constraint arises from the irregular shapes and inadequate demarcation of smallholder farms in Cambodia. In order to address the aforementioned constraint, an alternative loss function has been investigated in an effort to accurately anticipate limits characterized by closed contours. To facilitate the integration of this innovation, it is imperative to verify that the source model employed in the study by (Persello et al., 2019) has been equipped with this specific loss function prior to its application in training the Cambodia dataset. The utilization of binary cross-entropy as a loss function in binary classification tasks is a prevalent methodology, as demonstrated by (Persello et al., 2019).




(a) Training Loss Contour-Loss (b) Validation Loss Contour-Loss

(c) Training Accuracy Contour-Loss (d) Validation Accuracy Contour-Loss

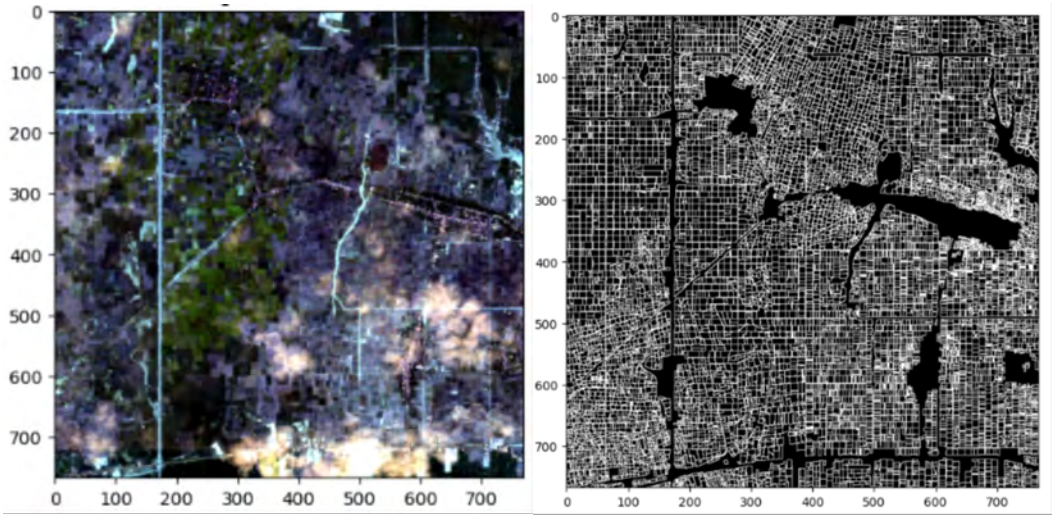
Figure (4.12) Contour-Loss Training Graphs of Gradient Descent for the Netherlands 1 million fields U-net model and U-Net training for Cambodia.

The introduction of the Contour-Loss loss function in the source model used for transfer learning led to training stagnation, impeding the model's ability to effectively extract and learn from the features in the image. This discovery remained consistent even when the hyper-parameter "learning rate" was altered, while simultaneously adjusting the number of epochs in the training process. The U-Net source model with Contour-Loss exhibits divergence with the validation loss and training loss not showing any improvements throughout the training process. This is also evident from 4.12b in Figure 4.12.

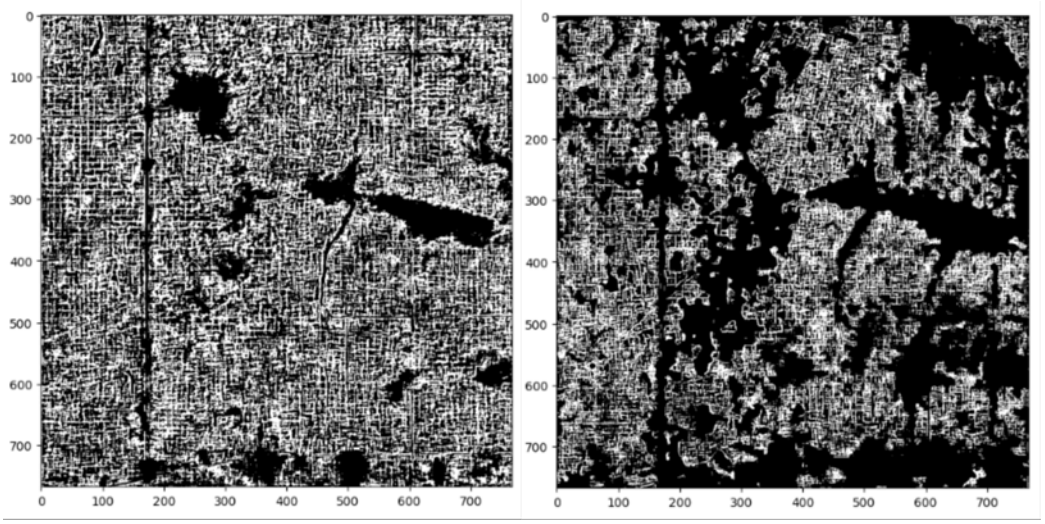
Despite the evidence provided by the training graphs and gradient descent, which suggest that the Contour-Loss loss function is not appropriate for our research, we proceeded to estimate the F1 score for predicting Cambodia using a model trained with Contour-Loss as the loss function for the Cambodia study area instead of using transfer learning techniques. The trial's outcome provides us an opportunity to compare model's performance with Contour-Loss as loss function against the model configured with binary cross entropy as loss function. The U-Net Model, trained exclusively on the Cambodia geometric median dataset and using contour-loss, exhibits a prediction F1 score of 46% as seen in Table 4.11 and the model trained with same training dataset with "binary cross entropy yields an F1 score of 43% as seen in Table 4.12 This score is the lowest among all the trials conducted to delineate the boundaries of Cambodia. Contour-Loss, has been proven to be better than binary cross entropy in model that hasn't been using transfer learning. Limitation of the research is the implementation of Contour-Loss to evaluate its performance in a trial based on transfer learning. Furthermore, the mathematical complexities introduced by the Contour-Loss loss function that led to the failure in training against the model built by (Persello et al., 2019) is yet to be investigated. Qualitative comparison of prediction image between contour-loss and binary cross entropy is as seen in the Figure 4.13. Training graphs for model training using binary cross entropy loss function is depicted in Figure 4.14.

**Table 4.11** Accuracy Assessment Cambodia using Contour-Loss in new training

| Accuracy Assessment | Percentage |
|---|---|
| Overall Accuracy | 59.26% |
| Precision | 41.266% |
| Recall | 52.124% |
| F1 Score | 46.064% |



(a) Training Image (b) Ground Truth



(c) Prediction Contour-Loss (d) Prediction Binary Cross Entropy

Figure (4.13) Prediction of AFB for Cambodia based on Seasonal Geometric Median Composite Contour-Loss Vs Binary Cross Entropy.

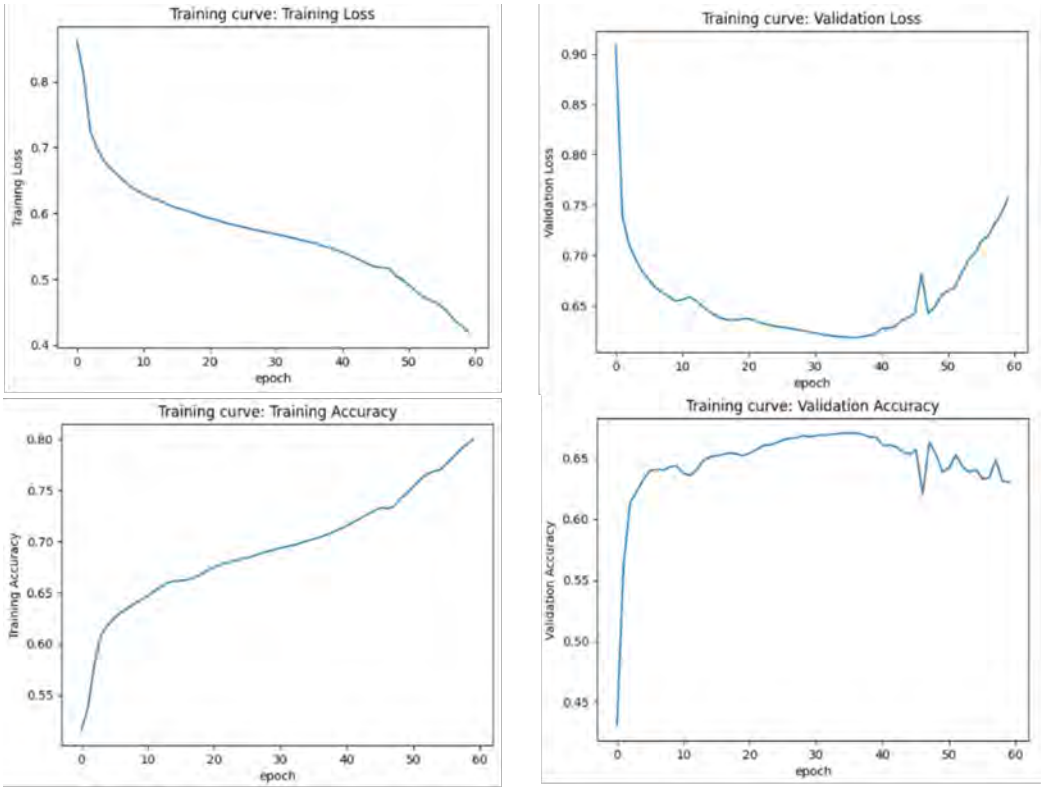



Figure (4.14) Training graphs model training with binary cross entropy loss function

**Table 4.12** Accuracy Assessment Cambodia using binary cross entropy in new training

| Accuracy Assessment | Percentage |
|---|---|
| Overall Accuracy | 65.817% |
| Precision | 48.498% |
| Recall | 39.038% |
| F1 Score | 43.257% |

## 4.3 POST-PROCESSING BASED ON GRAPH-BASED SEGMENTATION

The utilization of morphological segmentation methods, in conjunction with the watershed transformation, has been extensively employed in post-processing applications. One limitation frequently encountered with morphological segmentation and watershed transformation techniques is the potential for sub-optimal separation of objects, particularly in regions characterized by weak intensity contrasts or irregularly shaped and sized objects. This is also evident in Figure 4.15 and 4.16 for Cambodia and Vietnam, respectively. Our proposed solution can efficiently differentiate between agricultural fields as seen in Figure 4.16(b) and their boundaries as compared to the traditional approach where the boundaries overlap which is evident in Figure 4.16(a).One significant advantage of utilizing graph-based segmentation with contour extraction in post-processing is its ability to be integrated directly into the workflow, eliminating the requirement for separate software execution. This sets it apart from morphological segmentation. Additionally, the output of post-processing using graph-based segmentation is a shape file that retains the geometry information, whereas the output of morphological segmentation is a raster that lacks the appropriate geometry information.

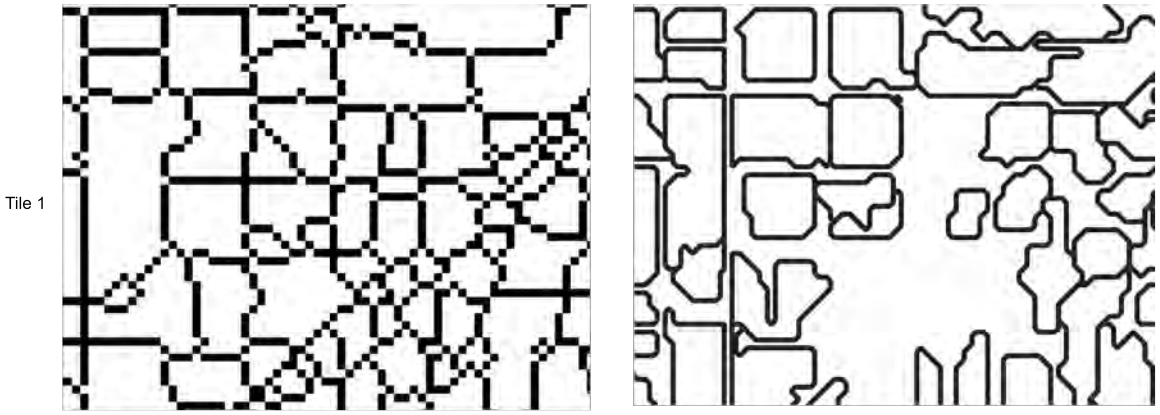



Figure (4.15) Post-processing Results for Cambodia, (a) Morphological Segmentation, (b) Graph-based segmentation with contour extraction

The utilization of graph-based segmentation, coupled with contour extraction, has demonstrated its efficacy in effectively extracting boundaries from the prediction image of the Vietnam study region. Based on qualitative analysis from Figure 4.18 the efficiency of the polygonization pipeline implemented for the research region in Cambodia has been comparatively lower than that observed for the Vietnam study area as seen . While the polygonization process applied to the prediction image is effective in capturing the contours present, it does not completely resemble the ground truth in terms of spatial accuracy, as seen in Figure 4.17. The primary reason for this distinction is that the ground truth data used in this study was derived by photo interpretation of Worldview-2 satellite images with a spatial resolution of 40cm, whereas the polygons generated through the suggested approach in this research were based on a Sentinel-2 Earth observation image with a spatial resolution of 10 meters. The dissimilarity in shape observed in Figure 4.17 can be attributed to the uncertainty introduced in the predictions derived from the sentinel-2 EO image. Another interesting finding from the qualitative analysis is that the polygons generated as part of the workflow seem to be split into two, as seen in Figure 4.17 whereas in the ground

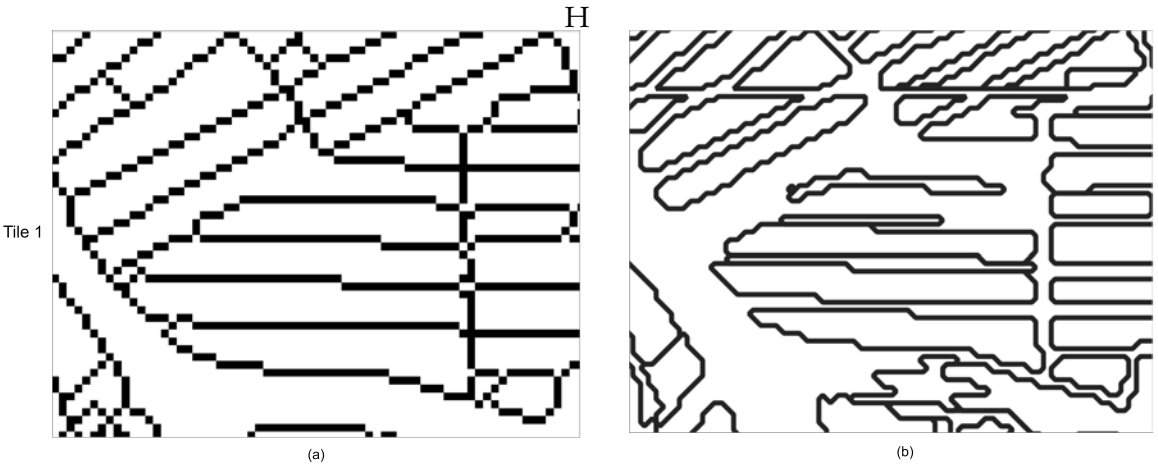



Figure (4.16) Post-processing Results for Vietnam, (a) Morphological Segmentation, (b) Graph-based segmentation with contour extraction

**Table 4.13** Results based on Polis for Vietnam and Cambodia

| Polis Results | Cambodia | Vietnam |
|---|---|---|
| **Number of matches** | 5646 | 3430 |
| **Number of reference polygons** | 12360 | 3099 |
| **Number of misses** | 6714 | -331 |
| **Number of Duplicates** | 1 | 1 |

truth, the polygon is part of a single agricultural field. This limitation is due to the process employed in polygonization, as depicted in Figure 3.5, the boundary prediction image generated by the CNN model is initially divided into smaller patches. This splitting of the prediction image enhances the efficiency of the polygonization procedure and yields polygons that exhibit greater spatial similarity to the ground truth.

Cambodia's agricultural landscape is characterized by farms of varying shapes and sizes, as well as densely inhabited fields. As a result, prediction images already incorporate boundaries that exhibit incomplete contours. Therefore, the level of polygonization in Cambodia has experienced a further decline based on qualitative analysis in Figure 4.18.

Polis a distance metric has been used to estimate the results for Cambodia and Vietnam and its results are documented in Table 4.13. Interestingly, Polis estimate detects that there are -331 missing polygons for the Vietnam study area. This is indeed true polygonization method implemented wrongly classifies background pixels as boundaries as seen in Figure 4.19. In an attempt to mitigate the issue of misclassification caused by false positives, adjustments were made to specific parameters, including the solidity threshold and the maximum size for background pixels. However, these adjustments had the unintended consequence of reducing the efficiency of polygonization for Vietnam and Cambodia.

For Cambodia, U-Net predictions are approximately 55% accurate even with seasonal geometric median composite polygonization inefficiency is backed by the low F1 score of U-Net.

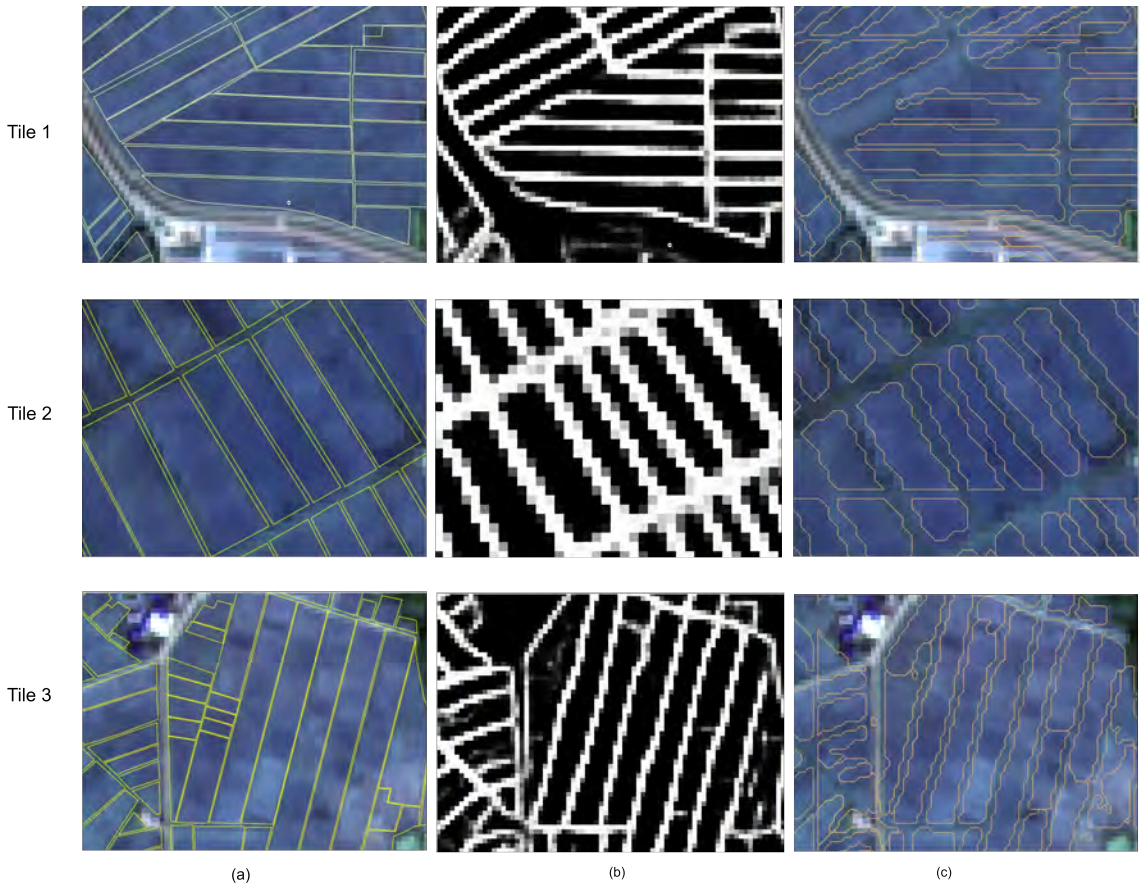



Figure (4.17) Polygonization Results for Vietnam, (a) Ground Truth, (b) Model Prediction, (c) Polygons from prediction.

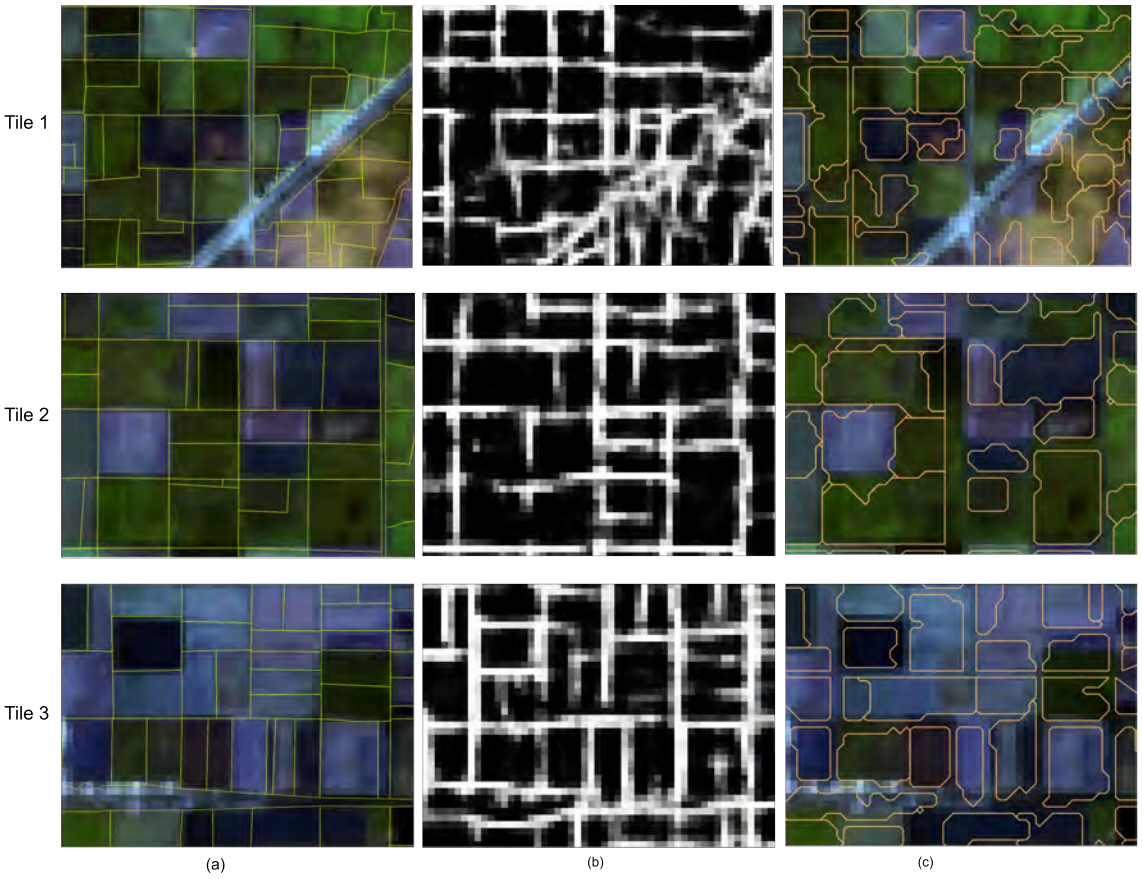



Figure (4.18) Polygonization Results for Cambodia, (a) Ground Truth, (b) Model Prediction, and (c) Polygons from prediction.

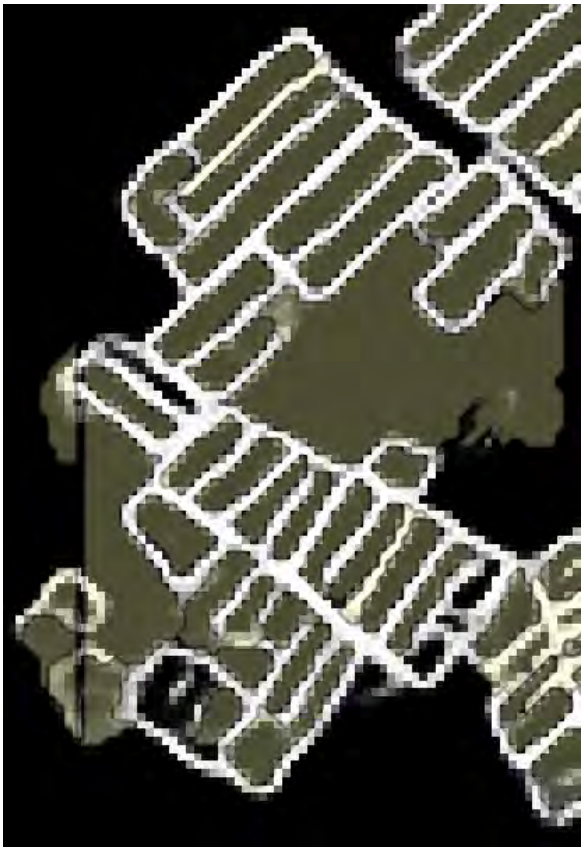


Figure (4.19) Boundary Prediction overlayed with Polygons demonstrating under-segmentation

# Chapter 5

# Conclusion

The purpose of this study was to develop a customized workflow for accurately delineating agricultural field boundaries using CNN's and Sentinel-2 EO images. The research conducted in this study has resulted in the development of a pipeline that generates polygons representing the boundaries of agricultural fields in Cambodia and Vietnam using sentinel EO images.

The pre-processing of EO images and the removal of outliers have been successfully accomplished by employing seasonal statistical composites created using the Geometric Median method, in comparison to the traditional median and medoid composites as seen in 4.10. The previously described pre-processed EO images were utilized in conjunction with an effective CNN model constructed using the UNet architecture. This implementation aimed at predicting boundaries based on EO data obtained from the Sentinel 2 platform. Furthermore, a custom loss function was employed to investigate potential methods for enhancing the accuracy and effectiveness of CNN in delineating boundaries in regions like Cambodia, where fields exhibit uneven shapes and smallholder farms lack proper demarcation. The workflow developed in the context of this research project has also integrated a post-processing stage into the pipeline. Post-processing techniques were employed to construct polygons from the prediction probability image obtained from the CNN model. This was achieved through the utilization of Graph-based segmentation in conjunction with contour extraction methods. The polygons obtained were subsequently evaluated using the Polis metric in order to determine the level of spatial similarity between the ground truth polygons and the polygons generated through the study methodology. The study successfully addressed the research inquiries and thoroughly analyzed the subsequent outcomes and findings.

Research Question 1: How to develop a pre-processing strategy that exploits the temporal information from sentinel-2 images available in the considered study area, which can potentially enhance the crop boundaries.

How to ensure training images have the proper contrast to identify boundaries of agricultural fields ?.

Research has successfully investigated the methods by which the emphasis on boundaries can be improved. Firstly, the use of seasonal images from wet and dry season months based on crop calendars and selection of specific months amongst those crop calendar months to ensure we capture the mature crops from the images EO images. Secondly, by building a statistical composite method that is more robust to outliers. From the research outcome, it was also understood that the selection of a particular technique such as geometric median, medoid or median as a composite method is not straightforward since the application also relies on the characteristics of the study area. For those with more cloud noise and outliers geometric median method was found to be the

best compositing technique that can enhance the emphasis on boundaries particularly for those observations not spanning more than a single season. With yearly observations of EO images, it was found that the median composite method can itself effectively remove outliers from the image and can lead to higher accuracy in CNN model's capacity to delineate boundaries.

Research Question 2: How to apply a loss function that can improve the accuracy of existing deep learning models to generate Agricultural field boundary data.

Research has successfully investigated the use of Contour-Loss, a custom loss function motivated by its implementation in (Yuan & Xu, 2022) as Gap loss. Although the implementation was successful the training of the CNN model used in transfer learning was not satisfactory with the Contour-Loss as loss function.The model did not exhibit convergence in training. Due to this limitation, the contour-loss loss function could not be implemented in the source model developed by (Persello et al., 2019) used for transfer learning. To prove the hypothesis that the contour-loss loss function can result in closed boundaries for the Cambodia study area which has been leading to partial boundaries without the custom loss function, the contour-loss loss function was implemented in a model without using transfer learning and training on the Cambodia training images alone. Since Cambodia is the region that needed an improvement through a loss function and Vietnam already is exhibiting a good F1 score. The training process and the gradient of descent graphs showed erratic descent in gradients whereas the F1 score of the trained model with Contour-Loss yielded a score of 46%. Whereas Binary cross entropy used as a loss function in new training yielded a score of 43% thus proving that Contour-Loss indeed is an efficient loss function that can improve the accuracy, the mathematical complexity of Contour-Loss has to be further analyzed in order to apply the loss function also in the source model built by (Persello et al., 2019) This is also a limitation of the research.

Research Question 3: How to apply a post-processing strategy able to polygonize predictions without decreasing the boundary delineation performances.

Research successfully integrates a post-processing phase within the model's execution pipeline. The initial research proposal outlined the use of frame field learning for polygonization (Girard et al., 2021) as the primary approach for the post-processing stage in delineating boundaries. However, this approach was later replaced with a graph-based segmentation technique combined with contour extraction, which is currently employed in the study. The motivation behind implementing this change was to reduce the computational load by avoiding the incorporation of an extra convolutional neural network (CNN) during the post-processing stage. Another justification for abstaining from the utilization of the frame-field framework for polygonization arises from the complex nature of implementing the frame-field learning framework within the setting of polygonization. Nevertheless, the combination of graph-based segmentation and contour extraction has exhibited the potential to transform predictions into polygonized boundaries in the form of shapefiles. The comparison between the polygonization output and the ground truth may appear less efficient due to the fact that the ground truth was derived by photo interpretation of Earth Observation (EO) photos obtained from the Worldview-2 satellite platform with a spatial resolution of 40cm. In contrast, when comparing the polygonization results to the model predictions derived from sentinel-2 EO photos with a resolution of 10 meters, it can be argued that the output obtained through polygonization using graph-based segmentation and contour extraction is reasonable. The lack of improvement in boundary predictions is a limitation of the research due to which the boundaries polygonized from predictions for Cambodia are even less efficient when compared to the ground truth for Cambodia's agricultural field boundaries.

Graph-based segmentation result for polygonization is compared with Watershed transformation-based post-processing and the findings were that Graph based segmentation yielded boundaries that were well separated whereas Watershed segmentation and morphological segmentation backed technique had overlapping boundaries. Also, the outcome of the later needed geometry properties to be restored, and the output of watershed transformation is a raster further needs to be polygonized. However, from the qualitative analysis, it was also observed that the watershed segmentation technique resulted in closed contours as compared to Graph-based segmentation used for polygonization in Cambodia. The Polis metric could not be established to compare the two techniques.

Future Research: Potential avenues for future research could involve enhancing the resolution of Sentinel-2 imagery in order to reduce pix-elation in the models' predictions, hence facilitating more efficient polygonization during post-processing. Another interesting approach in terms of improvising the methodology built in this research could be to apply graph-based segmentation for polygonization followed by watershed segmentation which might lead to a higher accuracy since watershed segmentation-based technique leads to closed contours.

# List of References

Avbelj, J., Muller, R., & Bamler, R. (2015). A Metric for Polygon Comparison and Building Extraction Evaluation. *IEEE Geoscience and Remote Sensing Letters*, *12*(1), 170–174.

Belgiu, M., & Csillik, O. (2018). Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis. *Remote Sens. Environ.*, *204*, 509–523.

Bergado, J. R., Persello, C., & Gevaert, C. (2016). A DEEP LEARNING APPROACH TO THE CLASSIFICATION OF SUB-DECIMETRE RESOLUTION AERIAL IMAGES john ray bergado , claudio persello , and caroline gevaert, 1516–1519.

Cambodia Production. (n.d.).

Crommelinck, S. (n.d.). Automating image-based cadastral boundary mapping.

Enclona, E. A., Thenkabail, P. S., Celis, D., & Diekmann, J. (2004). Within-field wheat yield prediction from IKONOS data: A new matrix approach. *Int. J. Remote Sens.*, *25*(2), 377–388.

Evans, C., Jones, R., Svalbe, I., & Berman, M. (2002). Segmenting multispectral landsat TM images into field units. *IEEE Trans. Geosci. Remote Sens.*, *40*(5), 1054–1064.

Felzenszwalb, P. F., & Huttenlocher, D. P. (2004). Efficient Graph-Based Image Segmentation - International Journal of Computer Vision.

Flood, N. (2013). Seasonal Composite Landsat TM/ETM+ Images Using the Medoid (a Multi-Dimensional Median).

Garcia-Pedrero, A., Gonzalo-Martin, C., & Lillo-Saavedra, M. (2017). A machine learning approach for agricultural parcel delineation through agglomerative segmentation. *Int. J. Remote Sens.*, *38*(7), 1809–1819.

Girard, N., Smirnov, D., Solomon, J., & Tarabalka, Y. (2021). Polygonal building extraction by frame field learning. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Gopidas, D. K., & Priya, D. R. (2022). Hybrid Segmentation Method for Boundary Delineation of Agricultural Fields in Multitemporal Satellite Image using HS-PSO-FCNN. *Materials Today: Proceedings*, *51*, 2272–2276.

Graesser, J., & Ramankutty, N. (2017). Detection of cropland field parcels from landsat imagery. *Remote Sens. Environ.*, *201*, 165–180.

Index.htm. (n.d.).

Janssen, L. L. F., & Middelkoop, H. (1992). Knowledge-based crop classification of a landsat thematic mapper image. *Int. J. Remote Sens.*, *13*(15), 2827–2837.

Jianbo Shi & Malik, J. (2000). Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *22*(8), 888–905.

Legland, D., Arganda-Carreras, I., & Andrey, P. (2016). Morpholibj: Integrated library and plugins for mathematical morphology with ImageJ. *Bioinformatics*, *32*(22), 3532–3534.

Li, X., Ling, F., Foody, G. M., & Du, Y. (2016). Improving super-resolution mapping through combining multiple super-resolution land-cover maps. *International Journal of Remote Sensing*, *37*(10), 2415–2432.

Lowder, S. K., Skoet, J., & Raney, T. (2016). The number, size, and distribution of farms, smallholder farms, and family farms worldwide. *World Dev.*, *87*, 16–29.

Marmanis, D., Datcu, M., Esch, T., & Stilla, U. (2016). Deep learning earth observation classification using ImageNet pretrained networks. *IEEE Geosci. Remote Sens. Lett.*, *13*(1), 105–109.

Masoud, K. M., Persello, C., & Tolpekin, V. A. (2019). Delineation of agricultural field boundaries from sentinel-2 images using a novel super-resolution contour detector based on fully convolutional networks. *Remote Sens. (Basel)*, *12*(1), 59.

Mueller, M., Segl, K., & Kaufmann, H. (2004). Edge- and region-based segmentation technique for the extraction of large, man-made objects in high-resolution satellite imagery. *Pattern Recognit.*, *37*(8), 1619–1628.

Navigo. (n.d.).

North, H. C., Pairman, D., & Belliss, S. E. (2019). Boundary Delineation of Agricultural Fields in Multitemporal Satellite Imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *12*(1), 237–251.

Persello, C., Tolpekin, V. A., Bergado, J. R., & de By, R. A. (2019). Delineation of agricultural fields in smallholder farms from satellite images using fully convolutional networks and combinatorial grouping. *Remote Sens. Environ.*, *231*(111253), 111253.

Roberts, D., Mueller, N., & Mcintyre, A. (2017). High-dimensional pixel composites from earth observation time series. *IEEE Trans. Geosci. Remote Sens.*, *55*(11), 6254–6264.

Rockson, G., Bennett, R., & Groenendijk, L. (2013). Land administration for food security: A research synthesis. *Land Use Policy*, *32*, 337–342.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2015). Going deeper with convolutions. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Turker, M., & Kok, E. H. (2013). Field-based sub-boundary extraction from remote sensing imagery using perceptual grouping. *ISPRS J. Photogramm. Remote Sens.*, *79*, 106–121.

Vietnam Production. (n.d.).

Vietnam_Rice_Dec2012. (n.d.).

Waldner, F., & Diakogiannis, F. I. (2020). Deep learning on edge: Extracting field boundaries from satellite images with a convolutional neural network. *Remote Sens. Environ.*, *245*(111741), 111741.

Wang, M., Wang, J., Cui, Y., Liu, J., & Chen, L. (2022). Agricultural Field Boundary Delineation with Satellite Image Segmentation for High-Resolution Crop Mapping: A Case Study of Rice Paddy. *Agronomy*, *12*(10), 2342.

Wang, S., Waldner, F., & Lobell, D. B. (2022). Unlocking large-scale crop field delineation in smallholder farming systems with transfer learning and weak supervision. *Remote Sens. (Basel)*, *14*(22), 5738.

Wu, Z., & Leahy, R. (1993). An optimal graph theoretic approach to data clustering: Theory and its application to image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *15*(11), 1101–1113.

Xinyan, F., Paris, C., Persello, C., Nelson, A., Conchedda, G., & Tubiello, F. (2021). Field boundary mapping in small scale farming using Multi-Resolution satellite data : A case study in cambodia and viet nam study areas and satellite data. *2*.

Yan, L., & Roy, D. P. (2014). Automated crop field extraction from multi-temporal web enabled landsat data. *Remote Sens. Environ.*, *144*, 42–64.

Yang, Z., Yu, W., Liang, P., Guo, H., Xia, L., Zhang, F., Ma, Y., & Ma, J. (2019). Deep transfer learning for military object recognition under small training set condition. *Neural Comput. Appl.*, *31*(10), 6469–6478.

Yuan, W., & Xu, W. (2022). GapLoss: A loss function for semantic segmentation of roads in remote sensing images. *Remote Sens. (Basel)*, *14*(10), 2422.