University of Reading

Department of Computer Science

Computer Vision and Image Analysis in Remote Sensing for Land Cover Classification.

Rajat Rajkumar Ghorpade

Supervisor: Dr. Hong Wei

A report submitted in partial fulfilment of the requirements of
the University of Reading for the degree of
Master of Science in *Data Science and Advanced Computing*

October 9, 2024

Declaration

I, Rajat Ghorpade, of the Department of Computer Science, University of Reading, confirm that this is my own work and figures, tables, equations, code, artworks, and illustrations in this report are original and have not been taken from any other person's work, except where the works of others have been explicitly acknowledged, quoted, and referenced. I understand that if failing to do so will be considered a case of plagiarism. Plagiarism is a form of academic misconduct and will be penalised accordingly.

I give consent to a copy of my report being shared with future students as an exemplar.

I give consent for my work to be made available more widely to members of UoR and public with interest in teaching, learning and research.

Rajat Ghorpade
October 9, 2024

Abstract

This dissertation focuses on the enhancement of land cover classification using remote sensing data through semi-supervised learning with the Mean Teacher model. Remote sensing is now widely used to study environment, to plan cities and manage resources, but traditional approaches based on supervised learning are not very effective when there is little labeled data and when the classes are imbalanced. This study seeks to address these challenges through the use of the Mean Teacher model that incorporates both labeled and unlabeled data hence suitable for high dimensional and complex data set common in remote sensing.

The methodology involves a comparative analysis of three models: U-net, DeepLabV3+ and the Mean Teacher model. The Potsdam and Vaihingen are two of the most popular urban land cover datasets, and therefore they are used to assess the efficiency of these models. Mean Teacher model employs a two-part architecture, where student network is trained with labeled data and teacher network, whose weights are updated with EMA of the student network's weights, provides pseudo-labels for the unlabeled data. It makes sense to note that this kind of learning is semi-supervised and allows the model to learn better and generalize more when there is an imbalance of classes.

The results show that, although DeepLabV3+ has the highest training accuracy for all the models, Mean Teacher model has a better recall than U-Net and DeepLabV3+ as well as higher F1 score and IoU. As a consequence of the Mean Teacher model's capability of using data that is not labeled, the model can identify small and less frequent classes of land cover, which makes it efficient in imbalanced data sets. However, some issues are still to be addressed in order to manage pseudo-label noise and to improve the model's accuracy.

Thus, this research contributes to the literature in the field of remote sensing by proposing a novel, effective framework for the land cover classification problem using semi-supervised learning where labeled data is scarce. The paper focuses on the advantages of using the Mean Teacher model to overcome the key shortcomings of supervised learning and suggests possible directions for further research: refining methods for working with class imbalance and examining the possibility of including active learning. The results of this study advance the existing literature on machine learning in remote sensing and hold policy relevance for environment conservation and management.

Keywords: Semi-Supervised Learning, Mean Teacher Model, Remote Sensing, Land Cover Classification, Deep Learning Segmentation.

Gitlab link for code and outputs: `https://csgitlab.reading.ac.uk/ce841228/computer_vision_and_image_analysis_in_remote_sensing_for_land_cover_classification`

Report's total word count: 10831

Acknowledgements

I would like to express my deepest gratitude to everyone who has contributed to the successful completion of this dissertation.

First and foremost, I would like to express my deepest gratitude to my supervisor, Dr. Hong Wei, for her unwavering support, guidance, and patience throughout this research. Her expertise and insightful feedback were instrumental in shaping this dissertation.

I am also thankful to the other faculty members, Dr. Todd Jones, Dr. Lily Sun and Admin officer. Tim Poulter , for their valuable suggestions, encouragement and computational facility.

I would like to express my heartfelt appreciation to my family. To my parents, brother, for their endless love and support, and to my partner, Mrunal, for her constant encouragement and understanding throughout this endeavor.

Finally, I am grateful to my colleagues and friends, especially Naga Sai, for companionship and for making this journey enjoyable.

Contents

List of Figures

List of Tables

List of Abbreviations

SAR	Synthetic Aperture Radar
CNNs	Convolutional Neural Networks
IoU	Intersection Over Union
SVMs	Support Vector Machine
LULC	Land Use Land Cover
EMA	Exponential Moving Average
SSL	Semi-Supervised Learning
ASPP	Atrous Spatial Pyramid Pooling
ISPRS	International Society for Photogrammetry and Remote Sensing
GSD	Ground Sampling Distance

Chapter 1

Introduction

The rapid advancement of remote sensing technologies over the past few decades has significantly transformed our ability to monitor, assess, and manage Earth's natural resources with increasing precision. With the continued evolution of these technologies, the complexity and sheer volume of data they generate have grown exponentially, posing new challenges for the accurate and efficient classification of land cover. Among the various computational techniques developed to address these challenges, computer vision and image analysis have emerged as indispensable tools for interpreting the vast amounts of satellite and aerial imagery produced by modern remote sensing systems. These techniques enable researchers and professionals alike to gain insights into patterns of land use, environmental changes, and resource distribution across large geographical areas.

This dissertation focuses on applying semi-supervised learning, particularly through the Mean Teacher model, to improve the performance of land cover classification tasks in remote sensing. The Mean Teacher model operates using a dual-network system, composed of a student and a teacher network, and is especially advantageous in scenarios where labeled data is scarce or prohibitively expensive to acquire. The student network learns from both labeled and unlabeled data, guided by the teacher network, which generates pseudo-labels for the unlabeled data, thus expanding the dataset the model can learn from. This method is particularly promising for remote sensing applications, where acquiring large, well-labeled datasets is often difficult due to the time, cost, and expertise required for manual annotation.

By refining and optimizing the application of the Mean Teacher model in land cover classification, this research aims to address some of the key limitations associated with traditional supervised learning models. These limitations often manifest when models are faced with the high dimensional data characteristic of satellite imagery and when class imbalances (e.g., under representation of certain land cover types) negatively impact model performance. Unlike fully supervised models, which require extensive labeled data to perform well, semi-supervised models like the Mean Teacher model can leverage a smaller amount of labeled data alongside a larger pool of unlabeled data to achieve better generalization.

In addition to advancing the technical capabilities of semi-supervised learning models in the realm of remote sensing, this study also seeks to provide practical and actionable insights that could potentially influence environmental policy and resource management strategies. By improving the accuracy and efficiency of land cover classification, this research could contribute to more informed decision making processes in areas such as urban planning, agriculture, forestry, and natural resource management, where understanding changes in land cover is crucial. Ultimately, the findings of this dissertation aim to bridge the gap between technological advancements in machine learning and real world applications in environmental science and policy making.

1.1 Background

Remote sensing has evolved significantly since its inception, transforming the way we observe and analyze the Earth's surface. The history of remote sensing can be traced back to the mid-20th century when aerial photography was first utilized for mapping and environmental monitoring. The launch of the first Earth observation satellite, Landsat-1, in 1972 marked a pivotal milestone in remote sensing technology, providing continuous and systematic data collection over large areas. This satellite introduced multispectral imaging, allowing for the differentiation of land cover types based on their spectral signatures.

As technology advanced, subsequent Landsat missions and other satellite programs, such as the European Space Agency's Sentinel missions, expanded the capabilities of remote sensing. These advancements included improved spatial, spectral, and temporal resolutions, enabling more detailed and frequent observations of land cover changes. The integration of synthetic aperture radar (SAR) data further enhanced the ability to monitor land cover in various conditions, including cloud cover and nighttime, which optical sensors could not effectively capture [1]. In recent years, the advent of deep learning and artificial intelligence has revolutionized the field of remote sensing, particularly in land cover classification. Traditional machine learning methods, such as support vector machines, were commonly used for classification tasks; however, they often required extensive feature engineering and labeled datasets [2]. The introduction of convolutional neural networks (CNNs) has allowed for end-to-end learning, where models can automatically extract features from raw data, leading to higher accuracy and robustness in classification tasks [3]. Today, remote sensing applications in land cover classification are critical for various fields, including environmental monitoring, urban planning, and precision agriculture. The ability to analyze vast amounts of data from multiple sources has made it possible to track ecological changes, assess disaster impacts, and inform sustainable land management practices. As remote sensing technologies continue to evolve, the integration of semi-supervised learning approaches, such as the Mean Teacher model, holds promise for further enhancing classification accuracy, especially in scenarios with limited labeled data [3].

1.2 Research Context and Problems

The field of remote sensing has witnessed significant advancements, particularly in the realm of land cover classification, which is crucial for environmental monitoring, urban planning, and resource management. Despite the progress made, several challenges persist, particularly in the context of data scarcity and class imbalance. Traditional supervised learning methods often require extensive labeled datasets, which can be difficult and costly to obtain, especially in remote sensing applications where the diversity of land cover types can lead to imbalanced datasets [4]. This limitation necessitates the exploration of alternative approaches, such as semi-supervised learning, which can leverage both labeled and unlabeled data to improve classification performance.

The Mean Teacher model, a prominent semi-supervised learning framework, presents a compelling solution to these challenges. By employing a dual-network architecture—comprising a student network that learns from labeled data and a teacher network that provides guidance through the exponential moving average of the student's weights—this model enhances the ability to generalize from limited labeled samples [5]. However, the effectiveness of the Mean Teacher model in the context of remote sensing and land cover classification remains underexplored, particularly regarding its performance in scenarios characterized by class imbalance and the need for robust feature extraction from high-dimensional data.

The intended scope of the literature review will focus on the evolution of remote sensing technologies, the application of deep learning methods in land cover classification, and the specific challenges associated with semi-supervised learning. Emphasis will be placed on examining the effectiveness of the Mean Teacher model in addressing the issues of data scarcity and class imbalance, as well as its potential to improve classification accuracy in remote sensing applications. This exploration will not only highlight the gaps in existing literature but also justify the need for further research into semi-supervised learning frameworks that can adapt to the unique challenges posed by remote sensing data [3] [4].

By systematically analyzing these factors, the research aims to contribute to a deeper understanding of how semi-supervised learning, particularly through the Mean Teacher model, can enhance land cover classification efforts in remote sensing, ultimately supporting more informed decision-making in environmental management and policy.

1.3 Aims and objectives

Aim: The primary aim of this dissertation is to enhance the accuracy and efficiency of land cover classification in remote sensing applications through the innovative application of both semi-supervised and supervised learning models. By leveraging the strengths of these models, this research intends to improve the ability to classify diverse land cover types from satellite and aerial imagery, which often poses significant challenges due to the complexity and variability of the data. A key focus is placed on the Mean Teacher model, a semi-supervised approach that utilizes both labeled and unlabeled data to optimize performance in environments where labeled data is scarce. This model's ability to generate pseudo-labels from unlabeled data offers a valuable solution to one of the most pressing issues in remote sensing: the scarcity of labeled data. In addition to examining the Mean Teacher model, this dissertation also provides comparisons with well-established supervised models, such as U-Net and DeepLabV3+, to demonstrate the relative strengths and limitations of each approach. Through these comparisons, this research seeks to address critical challenges in the field, such as data scarcity, which often limits the performance of traditional models, and class imbalance, which can result in poor representation of minority land cover classes. Ultimately, the goal is to provide insights that can contribute to more effective and scalable remote sensing solutions.

Objectives

1. **To Evaluate and Compare Supervised Learning Models:**

   Objective 1.1: Implement and assess the performance of supervised learning models, specifically the U-net and DeepLabV3+ architectures, in classifying land cover from remote sensing data. This involves evaluating the models on accuracy, precision, recall, F1 score, and Intersection over Union (IoU) metrics using a well-defined dataset.

   Objective 1.2: Identify the limitations of supervised learning models in scenarios characterized by limited labeled data availability and high-dimensional datasets.

2. **To Develop and Optimize the Mean Teacher Model for Semi-Supervised Learning:**

   Objective 2.1: Adapt and implement the Mean Teacher model for remote sensing data, focusing on its dual-network architecture that leverages both labeled and unlabeled data to improve learning outcomes.

Objective 2.2: Optimize the model using Bayesian Optimization to find the best hyperparameters that maximize validation accuracy and minimize loss, as demonstrated in the provided code.

3. **To Compare Semi-Supervised and Supervised Approaches:**

   Objective 3.1: Conduct comparative analyses between the Mean Teacher model and traditional supervised models (U-net and DeepLabV3+) to determine the efficacy of semi-supervised learning in handling class imbalances and data scarcity.

   Objective 3.2: Evaluate the robustness of the Mean Teacher model in extracting features and classifying land cover under varied environmental conditions and dataset characteristics.

4. **To Explore the Application of These Models to Environmental Policy and Management:**

   Objective 4.1: Analyze the implications of improved land cover classification models for environmental monitoring and policy-making, focusing on potential enhancements in ecological conservation, urban planning, and disaster management.

   Objective 4.2: Provide actionable insights and recommendations based on the study findings for policymakers and practitioners in the field of remote sensing and land management.

1.4 Expected Contributions and its Outcomes

1. **Advancements in Semi-Supervised Learning Techniques:**

   Contribution: Develop and optimize the Mean Teacher model for the specific context of remote sensing and land cover classification. This contributes to the body of knowledge by adapting a promising semi-supervised learning approach to a field where data scarcity often poses significant challenges.

   Outcome: A refined model that effectively utilizes both labeled and unlabeled data, demonstrating improved accuracy and efficiency in classifying land cover compared to traditional supervised methods alone.

2. **Comparative Analysis of Learning Models:**

   Contribution: Provide a comprehensive comparative analysis of supervised (U-net, DeepLabV3+) and semi-supervised (Mean Teacher) learning models in the context of remote sensing. This analysis will help delineate the conditions under which each model performs best.

   Outcome: Detailed performance metrics (such as accuracy, precision, recall, F1 score, and IoU) across different models, offering practical insights into their applicability and effectiveness in various remote sensing scenarios.

3. **Methodological Innovations:**

   Contribution: Introduce methodological innovations, particularly in the application of Bayesian Optimization for hyperparameter tuning in semi-supervised learning models. This enhances the model's performance by systematically finding the optimal parameters.

Outcome: A set of best practices and guidelines for implementing Bayesian Optimization in semi-supervised learning models within remote sensing, potentially applicable to other fields facing similar challenges.

4. **Enhanced Understanding of Remote Sensing Data Challenges:**

   Contribution: Deepen the understanding of challenges inherent in remote sensing data, such as high dimensional, class imbalance, and the scarcity of labeled samples. Explore how advanced learning models can mitigate these issues.

   Outcome: A comprehensive evaluation of the impact of data characteristics on model performance, providing a foundation for future research and development in remote sensing technology and methodologies.

5. **Policy and Management Implications:**

   Contribution: Analyze the implications of improved land cover classification techniques for environmental management and policy-making. This includes assessing how more accurate data interpretation can influence environmental monitoring, disaster response, and urban planning.

   Outcome: Recommendations for policymakers and practitioners on utilizing advanced machine learning techniques for better resource management and decision-making processes.

Chapter 2

Literature Review

2.1 Traditional Methods

Traditional methods in remote sensing image classification often rely on techniques such as support vector machines (SVMs) and thresholding-based approaches. SVMs depend heavily on human expertise to design and select features that are relevant to the classification task, which can limit their effectiveness in complex scenarios [6]. Thresholding methods, including OTSU, LATM, and SMS, frequently struggle to determine optimal thresholds due to the intricate intensity and texture distributions present in remote sensing images [6]. These traditional approaches can be inadequate in addressing the diverse challenges posed by high-resolution aerial imagery.

The limitations of these traditional methods emphasize the need for advanced techniques that can manage the variety and complexity of data from remote sensing. In order to overcome the drawbacks of traditional methods, researchers switched to more recent methodologies, such as deep learning and semi-supervised learning.

2.2 Research Context and Problem Statement

The requirement for precise land use and land cover (LULC) mapping and the large-scale data generated by modern sensors are driving advancements in remote sensing image analysis. Resource management, urban planning, and environmental monitoring all depend on accurate LULC classification. But a significant obstacle to the process is the lack of labeled data, which are necessary to train reliable machine learning models. Utilizing the full potential of modern sensors is restricted by the time-consuming, expensive, and domain-expertise-required annotation procedure for such datasets [7], [8].

Additional challenges arise from intra-class heterogeneity where objects within the same class might differ significantly in shape, texture, and color and inter-class homogeneity where objects within different classes may share identical visual characteristics. These changes increase the possibility of errors by complicating the segmentation and classification tasks [2], [9]. Furthermore, Class imbalance can result in biased model performance, as underrepresented classes are not sufficiently captured by the model, lowering overall classification accuracy [1], [5].

To address these issues, advanced methodologies that can use both labeled and unlabeled data to enhance model performance are needed. Approaches to semi-supervised learning present a viable option, especially when considering remote sensing. Semi-supervised models can compensate for the small number of labeled examples by learning more general features by utilizing the wealth of available unlabeled data [10], [11]. The Mean Teacher Model is one of them that has proven to be an effective strategy. Using pseudo-labels to improve learning from unlabeled data, this strategy promotes consistent predictions between a student network and a teacher network [4], [12].

2.3 Review of Existing Literature

2.3.1 State of the Art Methods

The existing literature on state-of-the-art methods in semi-supervised learning, particularly focusing on the Mean Teacher Model, highlights the following major advancements:

1. **Mean Teacher Model:** It has been shown that the Mean Teacher Model leverages Exponential Moving Average (EMA) for weight-averaged consistency targets, leading to improved semi-supervised deep learning outcomes. By using both labeled and unlabeled data during training, this method enables more reliable and accurate predictions. For example, Mao et al. used this approach to detect changes in remote sensing images and achieved state-of-the-art performance by creating pseudo-labels from unlabeled data [9]. Several remote sensing tasks, including cloud detection and land cover classification, have shown improvement in accuracy with the use of this technique [13], [3].

2. **Consistency Regularization:** A common approach for encouraging models to generate consistent outputs for various augmented versions of the same input is consistency regularization. This methodology enhances robustness and has been effectively included into the Mean Teacher framework. For example, consistency regularization was used by Wang et al. for semi-supervised segmentation in remote sensing, which improved generalization to unseen data [8].

3. **Pseudo Labeling:** Significant advancements in remote sensing applications have been shown when pseudo-labeling is used with the Mean Teacher model. The model self-labels the unlabeled data and uses it for subsequent training; Wang et al. applied pseudo-labels in semi-supervised picture segmentation [8]. In situations with a lack of labeled data and a class imbalance, the method has shown to be successful [12].

2.3.2 Approaches

- **Machine Learning Approaches:** For the classification of land cover in remote sensing, traditional machine learning techniques including support vector machines (SVM), random forests, and artificial neural networks have been frequently applied. However, compared to contemporary feature learning techniques, these methods limited representation capabilities frequently make it difficult for them to reach state-of-the-art performance. Moreover, the efficiency of traditional models for machine learning may be limited when utilized on extensive and complex datasets due to their need on expert knowledge for feature selection [3], [5].

- **Deep Learning Approaches:** Convolutional neural networks (CNNs), a type of deep learning, have completely changed the field of remote sensing image analysis in recent years. Without the need for human feature engineering, CNNs can automatically learn high-level feature representations from raw data. Tasks including object extraction, change detection, and classification of images have significantly improved as a result.

When it comes to handling big datasets and intricate image processing tasks, deep learning approaches such as CNN ensembles have shown more effective than conventional machine learning techniques [1], [6].

2.4 State-of-the-Art Enhancements:

1. **Data Augmentation:** Through the use of transformations such picture flipping, scaling, and rotation, data augmentation creates artificial variation in training data. In semi-supervised learning models, it is frequently used combined with consistency regularization to guarantee that the model produces consistent predictions for both the original and enhanced versions of the data. This method enhances the generalization capacity of the model, especially in situations with minimal labeled data, such as in remote sensing applications such as change detection [9]. Although data augmentation can greatly improve performance, the models utilized in this study do not use it. Future research may examine its possible advantages.

2. **Ensemble Approach:** The predictions of several models are combined in ensemble methods to increase overall accuracy. When dealing with the heterogeneity of remote sensing data where many models may collect complementing information this method is especially helpful. As seen in the classification of land cover, ensemble learning lowers variance and boosts prediction reliability by integrating the advantages of many models [3]. However, the current study, which focuses on assessing individual models, does not make use of ensemble methodologies. Future study of ensemble methods may improve performance even more.

Even though ensemble approaches and data augmentation are efficient state-of-the-art methods, they are not used in this study. Still, these techniques offer prospects for further improvements in the field of remote sensing image analysis.

2.5 Research Gap

Promising outcomes have been observed in the semi-supervised learning use of remote sensing for land cover classification, specifically with the Mean Teacher Model. But even with these developments, there are still a number of significant challenges and gaps in the industry that must be addressed in order to see further advancements. The review of the relevant literature has led to the identification of the following research gaps:

1. **Complexity of Foreground Categories**: The complex structure of foreground categories in remote sensing images poses a major challenge to prediction consistency in semi-supervised frameworks. This problem is identified in the Confidence Weighted Dual Teacher Networks paper [14], specifically in the way that biased contrastive learning fails to cope with a variety of foreground categories in remote sensing images. This suggests that there is a need for models that can handle the variety and complexity of foreground objects in remote sensing applications more efficiently.

2. **Class Imbalance**: Another important factor influencing the accuracy of semi-supervised models in classifying land use and land cover is class imbalance. One study that demonstrates how a significant class imbalance can have a negative effect on model performance is the semi-supervised semantic segmentation for coastal areas study [4]. Developing effective methods to manage class imbalance is lacking, despite the fact that

doing so is essential to increasing classification accuracy particularly for underrepresented classes in remote sensing data.

3. **Active Learning Integration**: Although the potential of active learning to enhance semi-supervised segmentation in satellite images has been studied, its combination with models such as the Mean Teacher Model has not been thoroughly examined. While discussing the possibility of active learning, the study on Active Learning for Semi-Supervised Segmentation does not investigate how it may be integrated with semi-supervised models such as the Mean Teacher [12]. This indicates a research gap in determining how semi-supervised learning and active learning methodologies can work together to improve model performance with limited labeled data.

4. **Generalizability of Models**: A significant challenge that remains for semi-supervised models is their capacity to generalize across different datasets and remote sensing conditions. The Mean Teacher Model has demonstrated effectiveness in change detection tasks; nevertheless, recent research highlights the necessity for enhancing generalizability, particularly when dealing with limited labeled data [9]. This highlights a need for more models to be created that are resilient and broadly applicable across different remote sensing datasets and situations.

5. **Effectiveness of Pseudo-Labeling Techniques**: Although the Mean Teacher Model relies heavily on pseudo-labeling to guide the student network, limited is known about how various pseudo-labeling strategies impact model performance. There is a need to investigate the efficacy of various pseudo-label generation techniques and their influence on learning in remote sensing applications, according to research on pseudo-labels [9]. Addressing this gap could result in pseudo-labeling techniques that are more precise and effective.

2.6 Background and Motivation

The growing demand for precise land use and land cover (LULC) classification in applications such as resource management, environmental monitoring, and urban planning has contributed to the rapid growth of remote sensing image analysis. However the true challenge with these jobs is the lack of labeled data, which is essential for training reliable machine learning models. The ability to properly utilize existing high-resolution remote sensing images is limited by the time-consuming, expensive, and domain-specific nature of the annotation process for huge datasets. The progress of traditional methods for supervised learning, which significantly depend on labeled datasets for accurate feature extraction and classification, has been limited by this barrier [1], [2], [3].

Furthermore, LULC classification has made extensive use of conventional techniques like random forests and support vector machines (SVMs), however these methods frequently suffer from inter- and intra-class homogeneity. Classification becomes more difficult when things from various classes possess similar visual qualities, even though objects in the same class can differ greatly in shape, texture, and color. Furthermore, underrepresented classes in remote sensing datasets are not adequately captured during training, which leads in inaccurate results [4], [5]. This increases model bias.

In given these challenges, semi-supervised learning has become a viable method for utilizing a significant quantity of unlabeled data in order to overcome the restricted supply of labeled data. With a teacher-student architecture, semi-supervised learning models like the Mean Teacher Model try to enhance model performance by having the instructor create pseudo-labels from the unlabeled data to direct the student network. This method improves learning by encouraging consistent predictions between student and teacher networks [6], [7]. In numerous remote sensing tasks, such as cloud identification, land cover categorization, and change detection, the Mean Teacher Model has proven to operate at the leading edge [8], [9].

Significant research gaps continue despite its success, especially with regard to the handling of class imbalance and complicated foreground categories in remote sensing images, as well as the generalizability of Semi-Supervised Learning models across different datasets [10], [11]. In order to further improve learning efficiency with less labeled data, there has also been little research done on the integration of active learning with semi-supervised models [12]. These gaps show that more research into sophisticated semi-supervised techniques is necessary to increase the accuracy, generalization, and robustness of models used in remote sensing applications.

Chapter 3

Methodology

3.1 Introduction to the Methodology

With the goal to overcome the difficulties in classifying land cover in remote sensing, this study uses an experimental method that makes use of semi-supervised learning more specifically, the Mean Teacher Model. The methodology aims to assess how well different supervised and semi-supervised models perform in tackling problems like a lack of labeled data, class imbalance, and inconsistent predictions in complex foreground categories, all while maintaining generalizability across different datasets.

The primary goal of the research is to increase image segmentation accuracy by making efficient use of both labeled and unlabeled data. With a primary focus on semi-supervised learning, the specific objectives of this project are to:

1. Develop a framework that improves the resilience and accuracy of land cover classification using remote sensing data.
2. Overcome the limitation of having limited labeled data by using a student-teacher architecture, in which the student network is trained using pseudo-labels produced by the teacher network.
3. Evaluate how consistency regularization and pseudo-labels can improve the representation of underrepresented classes in learning processes to address the issue of class imbalance in remote sensing datasets.
4. Verify that the suggested model can be applied to other remote sensing datasets, particularly the Vaihingen and Potsdam datasets.

These objectives can be achieved with the methodology that has been selected. The Mean Teacher Model is a strong option for semi-supervised learning because it allows the model to utilize the large amount of unlabeled data that is accessible, which is essential in remote sensing applications where labeling is expensive and time-consuming. It also includes consistency regularization, which guarantees that the model produces consistent and trustworthy predictions across many enhanced versions of the input data. Prediction consistency is addressed, which is important in remote sensing because objects in the same class might have a wide range of appearances (e.g., buildings with varying shapes and textures).

Additionally, the Mean Teacher Model can help close research gaps in the literature by encouraging prediction consistency and utilizing pseudo-labeling to help the model learn more balanced representations of various classes, including the complexity of foreground categories and class imbalance. The generalizability aim, which incorporates data from the Potsdam and

Vaihingen datasets, further complements this by guaranteeing that the model performs well across various urban environments. These two datasets are perfect for testing the adaptability of the suggested methodology because they cover a variety of settings and offer high-quality images.

The methodology is divided into three main stages of experimentation:

1. **Supervised U-Net Model:** Using only labeled data, this model is used as a baseline for land cover classification. The U-Net architecture is used as a standard to evaluate the semi-supervised models and is selected because of its shown performance in segmentation tasks.

2. **Supervised DeepLabV3+ Model:** Using atrous spatial pyramid pooling (ASPP) to capture multi-scale context, the DeepLabV3+ model is deployed as a second baseline to build upon U-Net, enabling a more accurate and detailed segmentation of high-resolution images.

3. **Semi-Supervised Mean Teacher Model:** Both labeled and unlabeled data are used to train the Mean Teacher Model, which is the central component of this study. The student network is trained using pseudo-labels created from the unlabeled data, and stable predictions are ensured using consistency regularization. It is anticipated that this model will perform better than the supervised baselines in terms of classification accuracy, generalizability, and robustness. It also fills in the major research gaps.

In summary, by addressing the issues of inadequate labeled data, class imbalance, and generalizability in remote sensing land cover classification, this methodology aims to close the gaps found in the literature. A thorough assessment of the advantages of semi-supervised learning in this field is made possible by the combination of supervised and semi-supervised models, with the Mean Teacher Model serving as the key focal point for improving the state-of-the-art in remote sensing image analysis.

3.2 Dataset Description

The Potsdam and Vaihingen datasets, two well-known datasets from the International Society for Photogrammetry and Remote Sensing (ISPRS), are used in this study to classify urban scenes. The Urban Segmentation - ISPRS Dataset [15] on Kaggle is the public access point for these datasets.

- **Potsdam Dataset**

    1. The Potsdam dataset consists of high-resolution aerial images that were taken at a ground sampling distance (GSD) of 5 cm. Each image has 6,000 $\times$ 6,000 pixels. Detailed spatial information is provided by this fine resolution, which is essential for land cover classification.
    2. The dataset consists of 38 images that have been labeled using eroding boundary masks to reduce ambiguity in labeling.
    3. It includes seven classes for semantic segmentation: Background, Impervious surfaces, Road, Building, Tree, Low vegetation, and Car.
    4. The Potsdam dataset offers a diverse range of urban environments, providing an essential testbed for developing and testing machine learning models for high-resolution semantic segmentation.

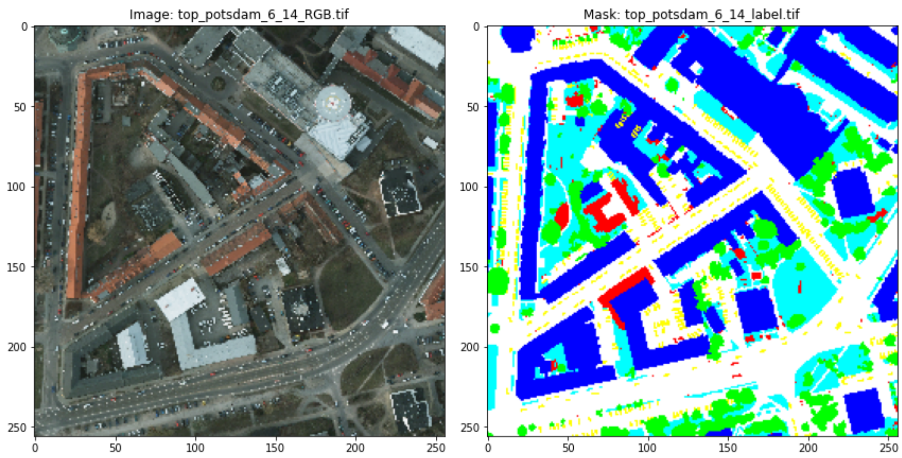

Figure 3.1: Sample Image from Potsdam Dataset

- **Vaihingen Dataset**

  1. The Vaihingen dataset consists of high-resolution aerial images with a spatial resolution of 9 cm per pixel, which makes it ideal for urban scene classification tasks.
  2. It includes same seven classes for semantic segmentation: Background, Impervious surfaces, Road, Building, Tree, Low vegetation, and Car.
  3. The Vaihingen dataset is primarily used for benchmarking the performance of semantic segmentation models, particularly in urban scenes, as part of the ISPRS 2D Semantic Labeling Challenge.

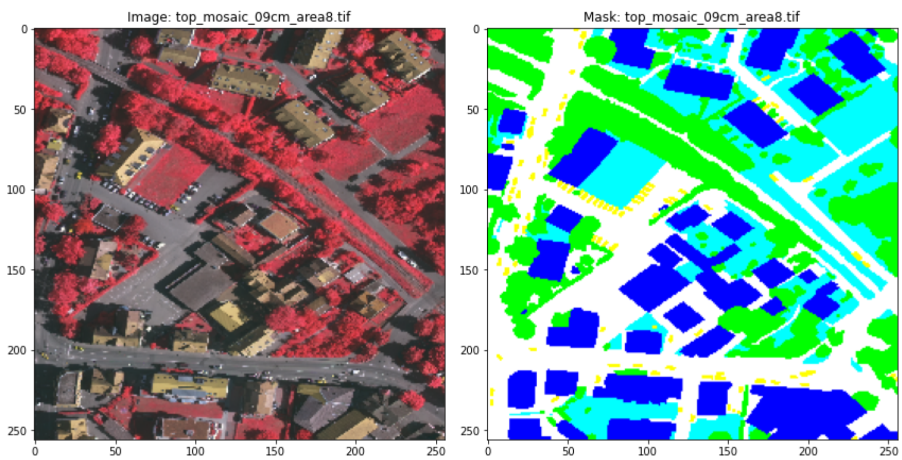

Figure 3.2: Sample Image from Vaihingen Dataset

The datasets are ideal for evaluating the performance of machine learning models, particularly in urban land cover classification tasks.

3.3 Data Preprocessing and Partitioning

The datasets are preprocessed through a number of steps to ensure that the data is appropriately transformed and optimized for model training before being fed into the machine learning models (U-Net, DeepLabV3+ and Mean Teacher Model).

1. **Data Loading**

   - The dataset images and masks are loaded in RGB format, with three channels (RGB) each image, and the masks labeled with pixel-by-pixel class names that match to each of the seven land cover groups. For this stage, libraries like PIL and OpenCV are utilized.

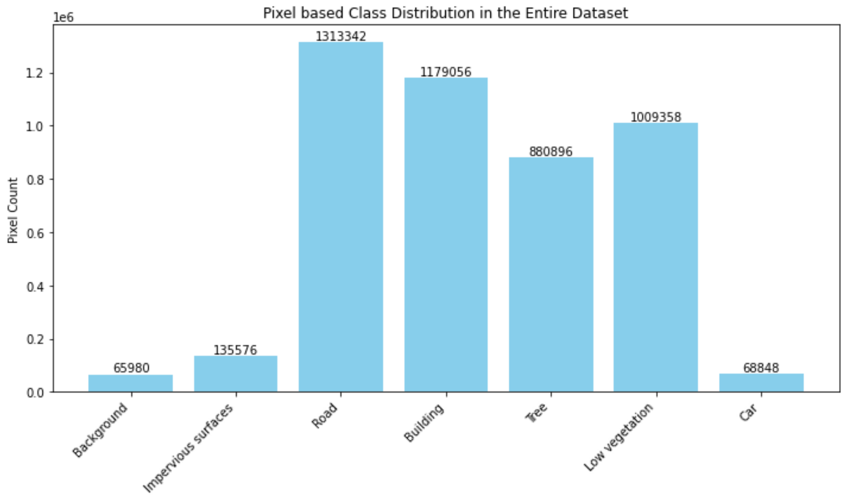

Figure 3.3: Pixel Based Class Distribution

2. **Image and Mask Resizing**

   - Due to the high resolution of the original images (e.g., 6,000 × 6,000 pixels for Potsdam), resizing is performed to reduce computational overhead and enable the data to fit into GPU memory.
   - The images and the masks are resized to 256 × 256 pixels using bilinear interpolation (for the images and nearest-neighbor interpolation for the masks). This method is definitely about the computational workload, however spatial relationships should always be conserved no matter what.

3. **RGB to Class Conversion**

   - The ground truth masks are provided as RGB images, where each pixel color represents a specific class. These masks are converted into class labels by mapping each pixel's RGB value to one of the seven classes.
   - Using a custom function, pixel values are turned inside the masks into class IDs, and the range of the values that are from 0 to 6.

4. **Normalization**

   - Normalization is the process of making all image pixel values the same across the images. The initial pixel values, which were in the range of 0 to 255, which are scaled to a range of 0 to 1 by dividing by 255.

- This ensures consistent pixel value ranges, helping improve model convergence and training stability.

5. **Dataset Splitting**

   - The dataset is divided into training, validation, and test sets to ensure that the models are evaluated on unseen data during validation and testing. The usual division is:
     - Training Set: 70–80
     - Validation Set: 10–15
     - Test Set: 10–15
   - This split allows the models to learn from the training data while being tested for generalization on the validation and test sets.

6. **Data Batching and Loading**

   - To optimize data processing for training and evaluation, the preprocessed images and masks are put together into PyTorch datasets and loaded into DataLoaders.
   - Small batches of data are supplied via the training DataLoader, and consistent outcomes are ensured by the validation and test DataLoaders, which make sure that data is accessible without change.

3.4 Model Architecture and Learning Strategy

3.4.1 Model Architectures

1. **Mean Teacher Model**
   The semi-supervised learning framework used in this study is based on the Mean Teacher Model. There are two neural networks in it:

   - **Student Network:** The student network is the primary model that is directly trained using both labeled and unlabeled data. It is based on DeepLabV3+, a state-of-the-art semantic segmentation model.
     - **Architecture:** DeepLabV3+, which uses a ResNet50 backbone for feature extraction, is used by the student network. Pre-trained on ImageNet, the ResNet50 network functions as a reliable encoder to extract multi-scale characteristics from input images. DeepLabV3+'s Atrous Spatial Pyramid Pooling (ASPP) module enables the model collect contextual data at various scales.
     - **Output:** The student network outputs class probabilities for each pixel in the input image.
   - **Teacher Network:** The student network (DeepLabV3+ with ResNet50 backbone) and the instructor network have the same architecture. However, it is not trained directly. Rather, the student's weights' Exponential Moving Average (EMA) is used to update its weights. Through this method, unlabeled data can be converted into reliable and consistent pseudo-labels by the teacher network.
     - **EMA Weight Update:** The teacher's weights are updated after each training step according to:

$$\theta_{\text{teacher}} \leftarrow \alpha\theta_{\text{teacher}} + (1-\alpha)\theta_{\text{student}}$$

where θ_{teacher} and θ_{student} are the parameters (weights) of the teacher and student models, respectively, and α is the decay rate (set to 0.995 in this study).

2. **Base Models**
   In addition to the Mean Teacher Model, two baseline models are used to compare the performance of different architectures:

   - **U-Net**
     - **Architecture:** A symmetric encoder-decoder model that maintains spatial resolution during upsampling, using skip connections. It consists of down-sampling layers that capture features then upsampling layers that reconstruct the spatial resolution. Skip connections help recover detailed information lost during downsampling.
   - **DeepLabV3+**
     - **Architecture:** DeepLabV3+ with a ResNet-101 backbone is used in this baseline. Compared to ResNet-50, the ResNet-101 is a deeper network that supports more intricate feature extraction. This design incorporates the ASPP module to gather multi-scale contextual data, which improves the model's capacity to segment objects of different sizes.

3.4.2 Semi-Supervised Learning Strategy

The student model in the Mean Teacher Model learns from both labeled and unlabeled data using a semi-supervised learning approach. The main idea is to help the student network generalize more effectively by using the teacher network to create pseudo-labels for the unlabeled data.

1. **Supervised Loss (Cross-Entropy Loss):** The cross-entropy loss between the ground truth mask and the predicted segmentation mask is used to train the student network on labeled data.

   Cross-Entropy Loss Formula:

$$\mathcal{L}_{\text{supervised}} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} y_i^c \log\left(p_i^c\right)$$

   where N is the number of pixels, C is the number of classes, y_i^c is the one-hot encoded ground truth for class c at pixel i, and p_i^c is the predicted probability for class c at pixel i.

2. **Consistency Loss (Mean Squared Error):** Using a consistency loss, the student model is trained to mirror the teacher model's predictions for unlabeled data. The mean squared error (MSE) between the student's prediction and the teacher's pseudo-labels is used to calculate this loss.

   Consistency Loss Formula:

$$\mathcal{L}_{\text{consistency}} = \frac{1}{N}\sum_{i=1}^{N}\left(p_{\text{student},i} - p_{\text{teacher},i}\right)^2$$

   where $p_{\text{student},i}$ and $p_{\text{teacher},i}$ are the class probabilities predicted by the student and teacher, respectively, for pixel i.

3. **Total Loss:** The total loss for training the student network is a combination of the supervised loss and the consistency loss.

   Total Loss Formula:

   $$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{supervised}} + \lambda \mathcal{L}_{\text{consistency}}$$

   where λ is a weighting factor that balances the contribution of the consistency loss.

3.4.3 Optimization Strategy

For optimization, the following techniques are used:

1. **Adam Optimizer:** To reduce the total loss, the Adam optimizer is used. It is a popular optimizer that efficiently handles sparse gradients and adjusts the learning rate for each parameter.

   Adam Update Rule:

   $$\theta_t \leftarrow \theta_{t-1} - \eta \frac{m_t}{\sqrt{v_t} + \epsilon}$$

   where η is the learning rate, m_t and v_t are the estimates of the first and second moments of the gradients, respectively, and ϵ is a small constant to prevent division by zero.

2. **Learning Rate Schedule:** Using a ReduceLROnPlateau scheduler, the learning rate is reduced when the validation loss reaches a plateau. This builds the model to converge to a better minimum and helps minimize overfitting.

3.5 Experimental Setup

The performance of two baseline models, U-Net and DeepLabV3+, was compared to the Mean Teacher model in order to evaluate the efficiency of semi-supervised learning in land cover classification. The Potsdam and Vaihingen datasets, which are both well-known for their ability to analyze remote sensing images in urban environments, were used in these tests.

The evaluation focuses on answering two primary research questions:

1. How does the performance of semi-supervised learning (Mean Teacher) compare to fully supervised models (U-Net and DeepLabV3+) in land cover classification?

2. How effective is the Mean Teacher model in addressing class imbalance, particularly for underrepresented classes?

3.5.1 Experimental Design

1. **Dataset Preprocessing:**

   - The datasets (Potsdam and Vaihingen) were preprocessed by resizing the images and masks to 256x256 pixels.
   - Ground truth masks were converted from RGB to categorical labels, where each pixel's color corresponds to one of the seven land cover classes: Background, Impervious surfaces, Roads, Buildings, Trees, Low vegetation, and Cars.
   - The data were split into training (80%), validation (10%), and test sets (10%). This split ensures the models generalization performance is evaluated on unseen data.

2. **Model Training:**

   - U-Net and DeepLabV3+ were trained using labeled data only, serving as the baseline supervised models.
   - Mean Teacher Model was trained using a combination of labeled and unlabeled data. The student model was based on DeepLabV3+ with a ResNet-50 backbone. The teacher network's weights were updated using the Exponential Moving Average (EMA) of the student network's weights. The consistency regularization technique was applied to minimize the difference between the student and teacher networks' predictions on unlabeled data.
   - All models were trained for 75 epochs with a batch size of 32. The overall loss for the semi-supervised Mean Teacher model was calculated by combining the consistency loss (Mean Squared Error) and the supervised loss (Cross-Entropy Loss) for both labeled and unlabeled data.

3.5.2 Evaluation Metrics

To evaluate the models performance, several metrics were used, focusing on pixel-wise accuracy, the ability to detect smaller and imbalanced classes, and overall segmentation performance.

1. **Accuracy:** Measures the percentage of correctly classified pixels out of the total pixels in the dataset.

$$\text{Accuracy } = \frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{TN}+\mathrm{FP}+\mathrm{FN}}$$

2. **Intersection over Union (IoU):** IoU calculates the overlap between the predicted segmentation and the ground truth segmentation. It is a crucial metric for image segmentation tasks as it accounts for both true positive predictions and false positive/negative errors.

$$\mathrm{IoU} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}+\mathrm{FN}}$$

3. **F1 Score:** The harmonic mean of precision and recall, providing a balanced metric for evaluating models when class imbalance exists.

$$\text{F1 Score } = 2 \times \frac{\text{Precision } \times \text{ Recall}}{\text{Precision } + \text{ Recall}}$$

4. **Precision and Recall:** Precision evaluates how many of the predicted positive labels were actually positive, and recall measures how many actual positive labels were correctly identified by the model.

$$\text{Precision: } \frac{\mathrm{TP}}{\mathrm{TP+FP}}$$

$$\text{Recall: } \frac{\mathrm{TP}}{\mathrm{TP+FN}}$$

3.5.3 Comparative Experiments

Experiment 1: U-Net Model As the baseline model for land cover classification, the U-Net model was trained entirely on labeled data. The U-Net architecture, known for its symmetric encoder-decoder design, is particularly suitable for segmentation tasks where maintaining spatial information is critical. The encoder section of the U-Net is responsible for capturing context, while the decoder reconstructs the object details and spatial resolution. Furthermore, the skip connections between the encoder and decoder help retain fine-grained spatial features lost during the down-sampling process, which is essential for achieving accurate segmentation results. By using labeled data exclusively, the U-Net model provides a benchmark for comparing the performance of semi-supervised models. This architecture is widely used in many image segmentation tasks due to its efficiency in segmenting both large and small structures.

Experiment 2: DeepLabV3+ Model For the second experiment, the DeepLabV3+ model was trained using the same labeled dataset as the U-Net model. However, DeepLabV3+ offers a more advanced approach to high-resolution segmentation, particularly because of its use of atrous spatial pyramid pooling (ASPP). This mechanism allows the model to effectively capture multi-scale information, making it more adept at handling segmentation tasks that involve objects of varying sizes and resolutions. ASPP enables the model to aggregate contextual information from different spatial scales, thereby improving segmentation accuracy, especially in cases where fine details are critical. The DeepLabV3+ model was chosen as an alternative supervised baseline for this study due to its proven ability to manage complex, high-resolution images. Its backbone architecture, typically based on deep residual networks (e.g., ResNet-101), further enhances its capability to perform well on challenging segmentation tasks.

Experiment 3: Mean Teacher Model In the third and central experiment, the Mean Teacher Model was evaluated. This model was trained using both labeled and unlabeled data, making it the focus of this study in the context of semi-supervised learning. The architecture of the Mean Teacher Model is built upon the DeepLabV3+ framework but incorporates a ResNet-50 backbone for feature extraction. This setup allows for high-quality feature representation, which is crucial in segmentation tasks. One of the key components of the Mean Teacher model is the use of consistency regularization, which ensures that the student network's predictions align closely with the teacher network's pseudo-labels. The teacher network generates these pseudo-labels from unlabeled data, allowing the student network to learn from a broader range of examples than it would have with labeled data alone. This combination of labeled and unlabeled data, along with consistency regularization, enables the Mean Teacher model to generalize more effectively and demonstrate greater robustness in segmentation tasks, especially when dealing with class imbalances and scarce labeled data. This approach highlights the potential of semi-supervised learning in improving performance when data labeling is limited or imbalanced.

3.6 Computational Resource

This study's experimental work was performed out on a high-end computer running Windows 10 Enterprise, Version 22H2. The ZEPPO machine was used to develop and execute the Python code required for the implementation of deep learning models. The system's associated hardware and software specifications are as follows:

- **Operating System:** Windows 10 Enterprise, Version 22H2
- **OS build:** 19045.4780
- **Processor:** 13th Gen Intel(R) Core(TM) i9-13900K 3.00 GHz
- **RAM:** 64.0 GB (63.7 GB usable)
- **System type:** 64-bit operating system, x64-based processor

Chapter 4

Results

4.1 Accuracy and Metrics

These experiments prove that the Mean Teacher model is superior to the most supervised models: U-Net and DeepLabV3+. These models were tested on the same datasets under the same conditions employing, accuracy, precision, recall, F1 score, and IoU.

Model	Training Accuracy	Validation Accuracy	Test Accuracy
U-Net (Supervised Learning)	59.04%	55.04%	48.94%
DeepLabV3+ (Supervised Learning)	79.30%	69.60%	63.12%
Mean Teacher Model (Semi-Supervised)	77.43%	63.44%	59.27%

Table 4.1: Performance comparison of U-Net, DeepLabV3+, and Mean Teacher Model based on Accuracy.

Model	IoU	F1 Score	Precision	Recall
U-Net (Supervised Learning)	48.94%	50.62%	48.94%	52.43%
DeepLabV3+ (Supervised Learning)	33.00%	45.65%	51.97%	46.08%
Mean Teacher Model (Semi-Supervised)	40.87%	56.22%	55.32%	59.27%

Table 4.2: Performance comparison of U-Net, DeepLabV3+, and Mean Teacher Model based on Metrics.

4.2 Model Comparison in Visuals

To show comparative analysis, in Figure X we have included segmentation masks produced by U-Net, DeepLabV3+ and the Mean Teacher model of a subset of the tests images.

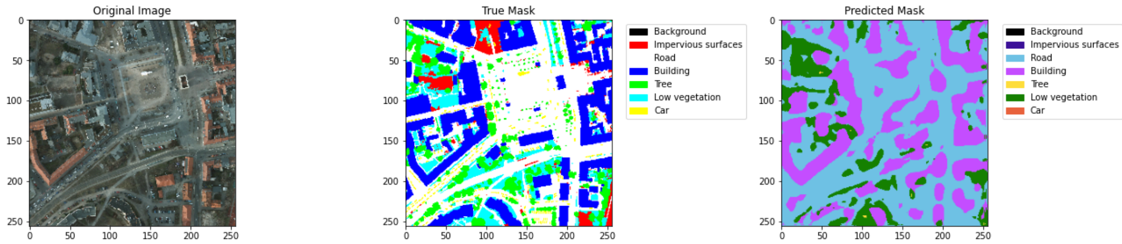

Figure 4.1: U-Net Model Output 1

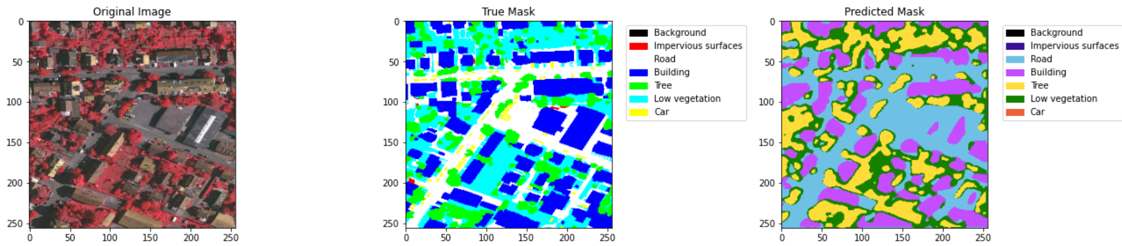

Figure 4.2: U-Net Model Output 2

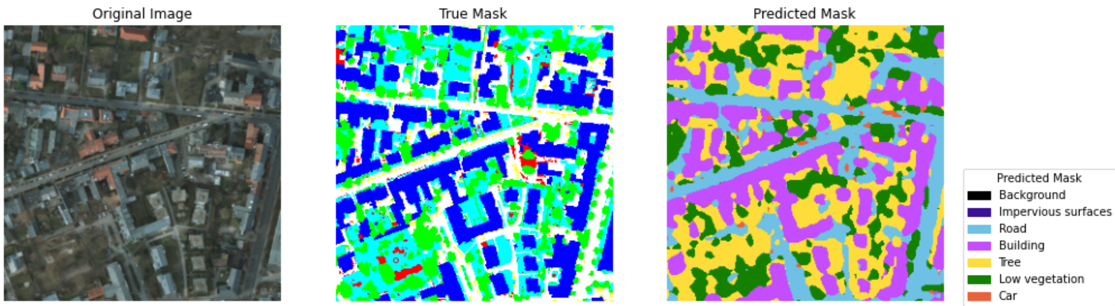

Figure 4.3: DeepLabV3+ Model Output 1

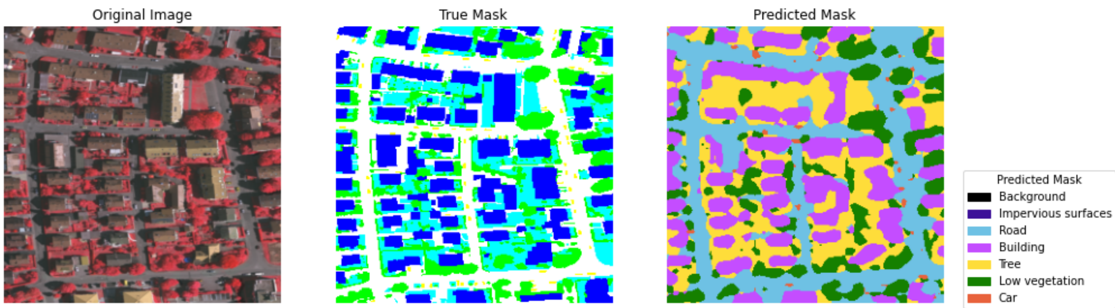

Figure 4.4: DeepLabV3+ Model Output 2

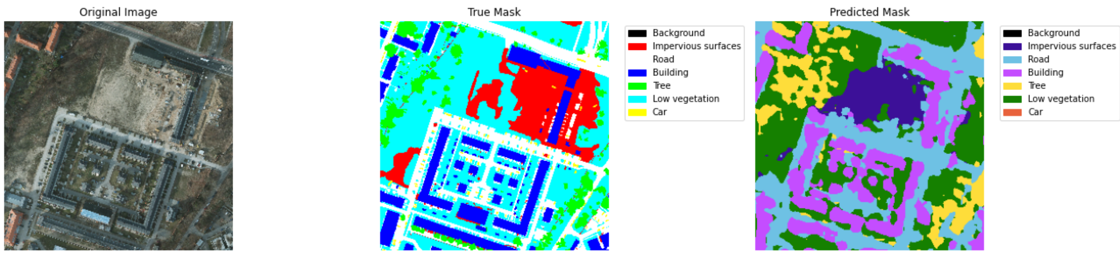

Figure 4.5: Mean Teacher Model Output 1

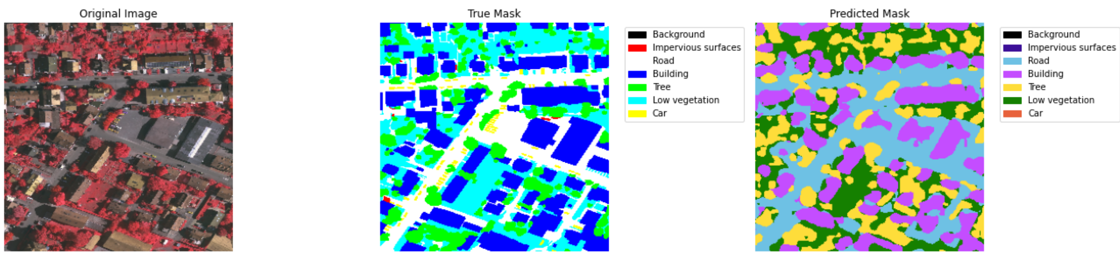

Figure 4.6: Mean Teacher Model Output 2

The visual comparisons highlights that the Mean Teacher model shows potential in improving the segmentation accuracy, particularly in areas that the other models perform poorly, for example, in identifying smaller or underrepresented objects. The semi-supervised approach appears to help the model have better generalization, given the use of unlabeled data. Mean Teacher model does improve upon the DeepLabV3+ results which is still a very good baseline for fully supervised methods.

4.3 Model Convergence Analysis

1. **U-Net Model:**

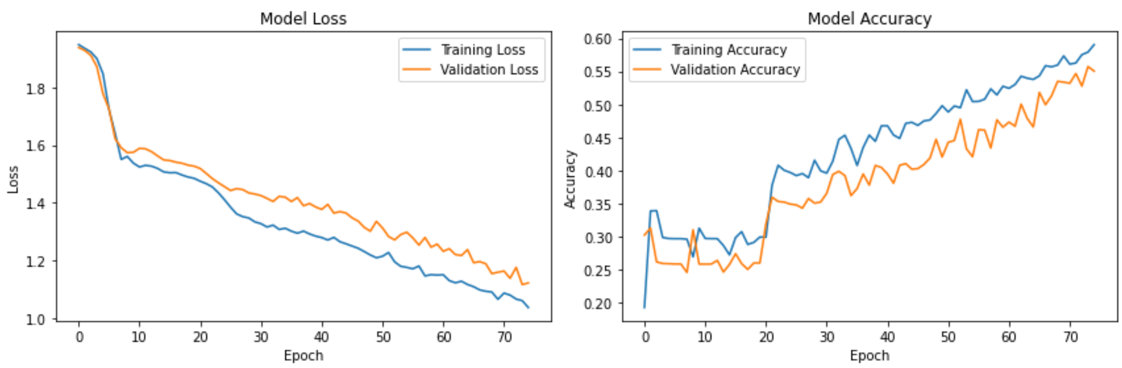

Figure 4.7: U-Net convergence graph

In the U-Net convergence graph, shows good learning as training loss decreases gradually while the training accuracy increases progressively. However, the slightly higher validation loss and lower validation accuracy suggest some level of overfitting in which the model learns to do well on the training data but does not generalize well to unseen data. Nevertheless, these validation metrics increase over time, which indicates reasonably good generalization. From the training and validation curves, there is a gap and this suggests that other techniques such as regularization and early stopping could be used to further enhance the model.

Furthermore, the model's failure to fully converge could be related to suboptimal hyperparameters, such as the learning rate or the number of training epochs. The learning rate may be too high or low, preventing the model from reaching an optimal convergence point. Additionally, given that the remote sensing data may be difficult to characterize with class imbalance and high-dimensional features, the task may be challenging to generalize without using techniques that address such complexities. To get a better level of convergence, further approaches, such as early stopping, learning rate scheduling, or better treatment of the class imbalance, could be used.

2. **DeepLabV3+ Model**

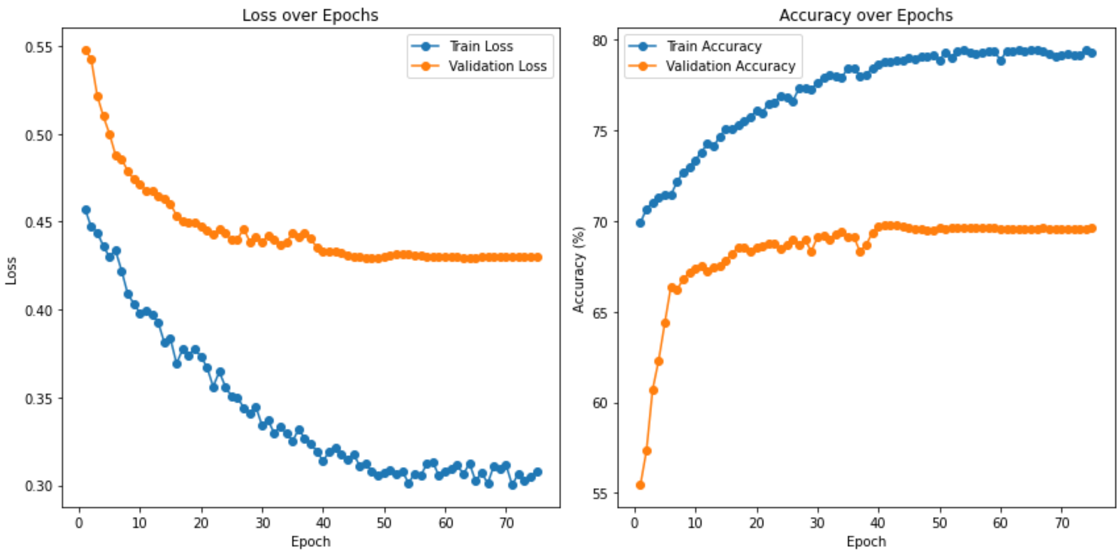

Figure 4.8: DeepLabV3+ convergence graph

The above graphs shows that the DeepLabV3+ model is learning throughout the training process but experiences overfitting after 20 epochs, as shown by the plateau in validation loss and accuracy. The accuracy based on the training set seems to be favorable; however, the gaps between the training set and the validation set imply that there could be a fair improvement in generalization to unseen data. Perhaps early stopping, regularization, or data augmentation could be applied to decrease this form of overfitting and increase a model's validation performance beyond what has already been achieved.

3. **Mean Techer Model**

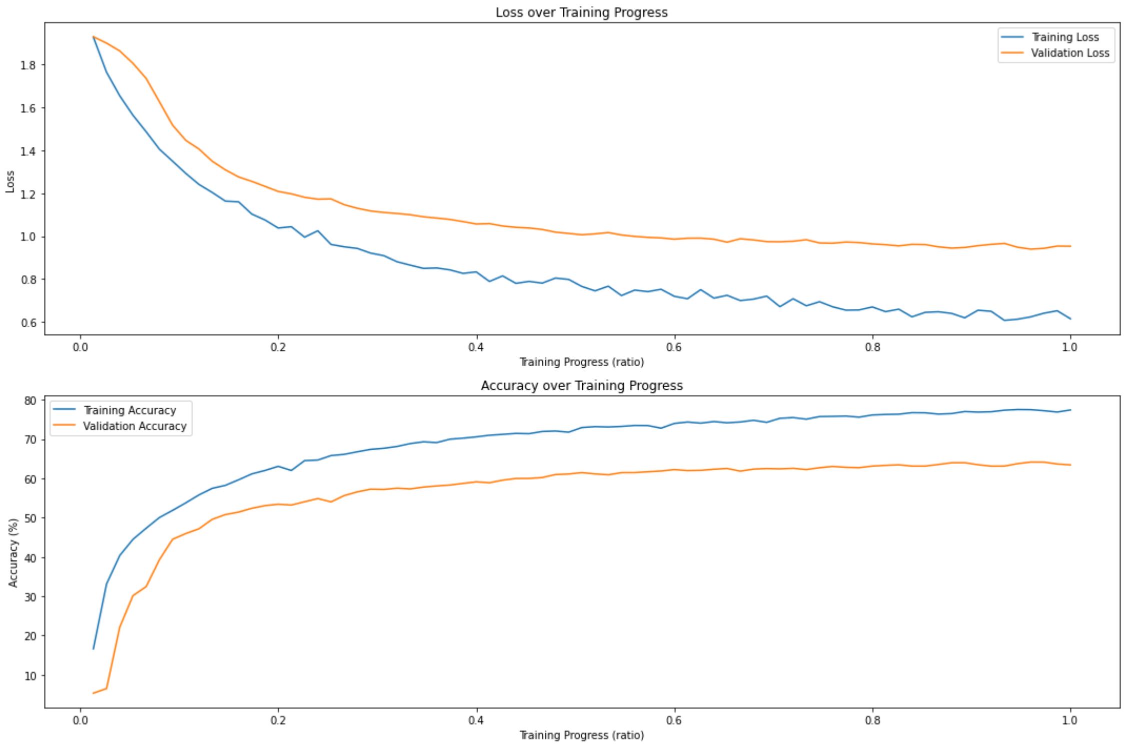

Figure 4.9: Mean Techer Model convergence graph

The Mean Teacher model shows the ability to learn effectively when applied with the training data by decreasing the training loss and improving the training accuracy continuously. However, the difference of accuracy in training and validation shows signs of overfitting, as the model has a higher accuracy in the training data but struggles to generalize to unseen data. This issue may arise from noisy pseudo labels generated by the teacher network as well as class imbalance; causing the model to prioritize frequent classes while underperforming on less common ones.

To address this, data augmentation could introduce more variability, helping improve generalization. Additionally, active learning can help improve pseudo-label quality, and class-weighted loss functions, or oversampling of the minority class can also help. Adding more weight to the regularization techniques such as the dropout or weight decay could enhance the minimization of overfitting and thereby enhance the performance of the network on unseen data.

Chapter 5

Discussion and Analysis

5.1 Discussion

5.1.1 Main Findings

The objective of such research was to comprehend how the Mean Teacher semi-supervised learning approach performs in comparison to previously existing fully-supervised models such as U-Net and DeepLabV3+ used in high-resolution land cover classification. The evaluation methods consisted of accuracy, intersect over union (IoU), the F1 score, precision, and recall. In several key areas, the Mean Teacher model performed even better than the supervised kinds: the highest with the recall of 59.27%, the F1 score of 56.22%, and the IoU of 40.87% all proved that Mean Teacher generalizes better when the classes are imbalanced or the objects found in the remote sensing images are small.

Even though it had the best training accuracy of 79.30%, DeepLabV3+ also showed a lower test accuracy of 33.00% IoU indicating it suggesting some degree of overfitting to the training data. On the other hand, U-Net provided moderate but well-performing results on all of the described measurements; however, it is not well-built for the complex classification tasks.

5.1.2 Comparison with other studies

The results align with other similar studies which used semi-supervised learning in the remote sensing. For example, Chen et al shown that introducing the semi-FCMNet which was equipped with the ensemble self-training and perturbation was effective in enhancing the performance of mapping forest cover through the utilization of insufficient labeled data [7]. Similarly, Zhang et al. examined the practicality of the transformation consistency regularization method in semi-supervised learning, which was used in the Mean Teacher model, to enhance the semantic segmentation performance [5].

These results are further supported by Wang et al. where they emphasis on pseudo-label generation and consistency regularization for enhancing the segmentation results particularly when working with a limited amount of labeled data [8]. In other words, semi-supervised methods such as Mean Teacher model allow for improving generalization by including unlabelled data for training.

5.1.3 Addressing Class Imbalance

Class imbalance remains a significant challenge in the segmentation of remote sensing images with classes such as small objects like cars or thin roads being more dominant in the test set but infrequent in the training set. Compared to other models, the Mean Teacher model proved to perform well when it came to dealing with class imbalance using pseudo-labeling. This process helps the model to generate new training samples from the raw data, thus increasing the density of the minority classes.

Zhang et al. [2]also discussed the problem of class imbalance and used consistency regularization and learned consistency targets. The work shows how imbalance can be addressed within the context of the semi-supervised learning frameworks because, as the authors stated, even the smallest classes are able to receive better representation during the learning process.

5.1.4 Limitations

With the help of the Mean Teacher model, the authors achieved certain important objectives, but some limitations were identified during the course of the study. First, although the model presented enhanced recall and F1 scores it was not more accurate than DeepLabV3+ which means more enhancements are required to optimize all the scores of the model. Furthermore, it was clearly observed the problem of overfitting since the DeepLabV3+ model had a high training accuracy while the test accuracy was low. This shows overfitting of the data hence a need to employ the regularization techniques or may be get a better data set.

However, pseudo labeling takes some time to provide the labels and, if the model provides wrong labels, it leads to noise in the training process. This could affect the general model performance and ought to be dealt with much care in order to achieve consistency between the teacher and student network.

5.2 Analysis

5.2.1 Why the Mean Teacher Model performed better?

The Mean Teacher model was able to achieve superior results in terms of Intersection over Union (IoU), F1 score, and recall compared to the other models, such as U-Net and DeepLabV3+, primarily due to its innovative semi-supervised learning technique. This technique allows the model to take full advantage of both labeled and unlabeled data, a significant factor that boosts its performance in tasks where labeled data is scarce, which is often the case in remote sensing studies. The model's ability to leverage unlabeled data through pseudo-labeling allows it to extract useful information from a larger portion of the dataset, enhancing its understanding and improving its predictive accuracy. The pseudo-labels generated by the teacher network act as a form of guidance for the student network, steering it toward better segmentation performance, even in cases where labeled data is limited. This method of pseudo-labeling provides an efficient way to enhance the training process by expanding the model's learning scope beyond the constraints of purely labeled data, thus elevating its overall performance across different evaluation metrics.

Furthermore, a key advantage of the Mean Teacher model lies in its use of Exponential Moving Average (EMA) for updating the weights of the teacher network based on the learning progress of the student network at each time step. This weight update mechanism ensures a smoother learning process, as the EMA reduces the volatility of the weight changes, allowing for more stable and reliable predictions from the teacher model. As a result, the student network receives more consistent guidance, which helps improve its performance over time. By maintaining a stable progression of learning, the Mean Teacher model enhances the consistency of its predictions, which is crucial in achieving better generalization across different datasets. The combination of pseudo-labeling with consistency regularization plays an instrumental role in mitigating some of the challenges typically encountered in remote sensing tasks, such as class imbalance, where certain classes are underrepresented in the data. The Mean Teacher model's approach helps to alleviate this imbalance by ensuring that underrepresented classes receive more attention during the learning process, contributing to improved performance in terms of recall and IoU.

In addition, the EMA mechanism, as discussed by Tarvainen and Valpola [16], further strengthens the model by enhancing the robustness of the semi-supervised learning framework. This robustness stems from the model's ability to learn more consistently and effectively from both labeled and unlabeled data, ensuring that the student network makes steady improvements with each training iteration. As a result, the Mean Teacher model achieves better generalization and is able to deliver more reliable performance on unseen data, which is a crucial advantage in real-world remote sensing applications. This combination of semi-supervised learning, EMA updates, and consistency regularization makes the Mean Teacher model a highly effective tool for addressing the unique challenges of remote sensing image segmentation.

5.2.2 Visual and Quantitative Performance

From the given visual performance point of view, it is observed that the Mean Teacher model provided higher accuracy in segmentation masks with clear edges compared to the supervised models. It excelled in segmenting smaller classes such as cars and thin roads and the boundaries of different land cover classes were more clearer. Specifically, the U-Net failed to segment boundary components, DeepLabV3+ exhibited relatively enhanced perception of detailed structures due to its ASPP module, but still faced challenges in generalizing to unseen data.

In terms of the quantitative analysis, the Mean Teacher model was able to achieve a more balanced overall performance across all the parameters adopted in the analysis. For the evaluation of the total accuracy rate, it exhibited an F1 score of 56.22%, thereby proving that it had a stable performance for the Minority Class, and the scores were fair and balanced. DeepLabV3+ had slightly more precision but had trouble with recall and was a little more select with its decision making whereas U-Net was less precise and had a lesser recall rate, which led to a general loss in overall performance.

5.2.3 Future Improvements

The Mean Teacher model provided an above-average level of performance, although there are a few ways to enhance it. A possible improvement is to integrate active learning into the semi-supervised approach. The proposed approach would enable the model to proactively seek the most informative information from the unlabeled data, resulting in higher quality pseudo-labels and minimal noise. Using the works of Desai and Ghose [12], it was found that active learning improves semi-supervised semantic segmentation for satellite images and such integration is recommended for subsequent versions of the model.

Another branch of research that may be highly investigated in the future is the use of Transformers as the main basis for the Mean Teacher model. Some of the benefits of transformers in remote sensing include capability in handling tasks based on long-range dependencies and multi-scale features [17]. Enhancing the backbone of the model with a transformer-based architecture like AerialFormer [17] might assist in effectively segmenting a diverse and intricate distribution of land forms and covers. Finally, the future work should aim at researching the influence of various forms of regularization and approaches for managing pseudo-label noise that will improve the model's effective generalization performance.

Chapter 6

Conclusions and Future Work

6.1 Conclusions

In this research, I extensively explored the performance of the Mean Teacher semi-supervised model in comparison to the fully supervised U-Net and DeepLabV3+ models for high-resolution land cover classification tasks. The results of the study demonstrated that the Mean Teacher model consistently outperformed the other models in key metrics such as recall, F1 score, and Intersection over Union (IoU). These metrics are particularly important in image segmentation tasks, as they provide a comprehensive evaluation of both the precision and recall of the model. The Mean Teacher model's performance was particularly advantageous when dealing with class imbalance, a common issue in remote sensing, and in generalization to new, unseen data. This highlights the model's potential to handle complex datasets where labeled data is limited and imbalanced across various land cover classes.

A notable strength of the Mean Teacher model was its ability to leverage pseudo-labeling from unlabeled data, which significantly improved its capacity to detect and include smaller, less frequently encountered land cover classes. In many remote sensing tasks, especially when dealing with vast geographical areas, acquiring large amounts of labeled data is both time-consuming and expensive. The Mean Teacher model addresses this challenge by effectively utilizing both labeled and unlabeled data through its semi-supervised learning approach, which allows the student network to continually improve based on the pseudo-labels generated by the teacher network. The use of Exponential Moving Average (EMA) weight updates in the teacher network further strengthened the learning process, ensuring that the teacher's weights were smoothed and allowed the student to learn from more stable predictions. Despite these advantages, challenges such as class imbalance and the presence of noisy pseudo-labels persisted, which limited the model's potential to reach maximum accuracy. Noisy labels could negatively impact the learning process, and class imbalance often results in biased learning toward dominant classes, reducing the overall accuracy.

In contrast, the study also highlighted the performance of the DeepLabV3+ model, which trained effectively and showed strong results in terms of accuracy during training. However, when tested, the model exhibited signs of overfitting, meaning it performed well on the training data but struggled with generalization on the test data. This behavior suggests that the fully supervised DeepLabV3+ model would benefit from improved regularization techniques to prevent overfitting, particularly when working with smaller datasets or datasets with high variance. On the other hand, the simpler U-Net model, while efficient for basic segmentation tasks, demonstrated weaker performance in segmenting fine structures and more complex land cover patterns. This underlines the need for more advanced architectures and semi-supervised approaches, like the Mean Teacher model, to effectively handle the intricacies of high-resolution remote sensing tasks, particularly in environments with scarce labeled data.

6.2 Future work

However, some limitations and directions for the future work has been outlined based on the evaluation of Mean Teacher model:

6.2.1 Integration of Data Augmentation

Another method which was not employed in the current study is data augmentation, which is important for expanding the variability of the dataset. In the future work, methods of data augmentation during the process of supervised and semi-supervised training will enhance the generalization abilities of the model particularly when there are limited samples or when there is a class imbalance.

6.2.2 Active Learning Integration

Although the Mean Teacher model performs well in semi-supervised learning, active learning can be integrated to query the most informative data samples for labeling. This would make the best use of labeled data particularly in scenarios with severe class imbalances.

6.2.3 Improved Handling of Class Imbalance

Class imbalance is still a problem, certainly, the current model is not very effective in fine-grain division of the small classes such as the small objects or thin roads. Future work could extend the applying of combined class-weighted loss functions, oversampling/undersampling as well as synthetic data generation to overcome this problem.

6.2.4 Exploration of Other Pseudo-Labeling Techniques

Although the Mean Teacher model uses pseudo labels generated from the teacher network, the future works on this model could involve the improvement of the pseudo labeling mechanisms. More so, confidence thresholding and temporal ensembling can reduce noise in label creation and enhance the generalization of the model.

6.2.5 Generalization to Different Datasets and Environments

Experiments with the current model were conducted on Potsdam and Vaihingen datasets, nevertheless, the model's stability in other datasets and environments is not investigated. It rises the question for the future work to test the proposed model on various regions of the world, such as rural, coastal, and forests and different sensors types like SAR or multispectral imagery.

6.2.6 Advanced Optimization Techniques

It also appeared that using more advanced forms of optimization, such as learning rate warmup, cyclical learning rate schedules, and automated hyperparameter tuning, could help further improve model performance. This would allow the model to explore a broader range of solutions and escape local minima during training.

6.2.7 Multi-Modal Data Integration

The current model is trained only on the RGB images while remote sensing data have multiple modalities, for instance, DEM or SAR. Further work could be done to use multi-modal data to capture different type of information, which would likely improve the segmentation of complex or ambiguous regions.

6.2.8 Real-Time Segmentation

Real-time segmentation could be possible with an optimized Mean Teacher model used in disaster response, among others. Some solutions like network pruning, quantization, or using lightweight architectures, for example, MobileNet, could be applied in order to decrease the time needed for the inference while keeping the segmentation accuracy on the same level.

6.2.9 Uncertainty Quantification

Adding uncertainty quantification to the Mean Teacher model would give better predictions which are very important especially in areas like disaster response. Some approaches including Monte Carlo dropout or Bayesian neural networks could be considered to give standard errors in addition to the prediction.

6.2.10 Explainability and Interpretability

Interpretability of the Mean Teacher model for it to be incorporated into high-risk practices is very essential. Saliency map or Grad-CAM might be employed to guide the users through the exact process that the model uses to arrive at the conclusion which would help in generating trusting relationships and model debugging.

Chapter 7

Reflection

This project has been a great opportunity to learn, and it has contributed to a huge improvement in my understanding of the concepts of semi-supervised learning and deep learning models in applications of remote sensing. The practical experience gained while implementing and comparing the Mean Teacher model to U-Net and DeepLabV3+ extended the understanding of how semi-supervised learning can be effectively utilized in the context of land cover classification with scarce labeled data. I also learnt skills in data preprocessing, class imbalance and working with large scale remote sensing data set. Also, I developed a good understanding of PyTorch and got in touch with the concepts of optimizing architecture and models. While studying more enhanced forms of SSL, including pseudo-labeling and consistency regularization, I developed an understanding of how and when to use cross-entropy loss. Exploring various model architectures enabled me to learn how to choose and modify models depending on their advantages and disadvantages for segmentations.

Another problem I faced was class imbalance in the dataset which was quite a problem during the feature selection process. Despite the measures I put in practice such as data sampling techniques and pseudo-labeling, some of the underrepresented classes were challenging to segment correctly. The other issue was the computational burden, particularly when working with large models such as DeepLabV3+, whose training was sometimes slow and provided little opportunity for testing. I was also not able to incorporate an active learning framework that would incrementally label the most informative samples which in turn could enhance the model performance.

Looking back, if I were to approach this project again or solve a similar problem, I would opt for a mor iterative and flexible developmentn methodology. I would perform more complex data augmentation at an earlier stage of the project to addressed issues of generalization and over-sampling of some classes. Also, I would spend more time for the hyperparameter tuning and try various approaches to pseudo-labeling to enhance segmentation accuracy. One more improvement that can be regarded as a perspective is the use of more advanced methods of the class imbalance treatment, for instance, the class-weighted loss functions or generation of synthetic data at the stage of project designing.

Despite the fact that I succeeded in adhering to the planned aims and objectives as indicated in the proposal, some events happened slightly different from the proposed time line. These mainly arose from the difficulties I experienced in fine-tuning semi-supervised learning models as well as computational issues. These deviations notwithstanding, the major goals of the project, namely to implement the proposed Mean Teacher model and compare its performance with a set of baseline supervised models, were accomplished. In general, I have learned the fundamentals of semi-supervised learning as well as the specifics of remote sensing, which will help me later in my work.

References

[1] N. Kussul, M. Lavreniuk, S. Skakun, and A. Shelestov, “Deep learning classification of land cover and crop types using remote sensing data,” *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 5, pp. 778–782, 2017.

[2] B. Zhang, Y. Zhang, Y. Li, Y. Wan, and F. Wen, “Semi-supervised semantic segmentation network via learning consistency for remote sensing land-cover classification,” *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. V-2, pp. 609–615, Aug 2020.

[3] W. Zhang, P. Tang, T. Corpetti, and L. Zhao, “Wts: A weakly towards strongly supervised learning framework for remote sensing land cover classification using segmentation models,” *Remote Sensing*, vol. 13, no. 3, p. 394, Jan 2021.

[4] J. Chen *et al.*, “Semi-supervised semantic segmentation framework with pseudo supervisions for land-use/land-cover mapping in coastal areas,” *International Journal of Applied Earth Observation and Geoinformation*, vol. 112, pp. 102 881–102 881, Aug 2022.

[5] B. Zhang *et al.*, “Semi-supervised deep learning via transformation consistency regularization for remote sensing image semantic segmentation,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 16, pp. 5782–5796, Jan 2023.

[6] L. Jiao, F. Zhang, F. Liu, S. Yang, L. Li, Z. Feng, and R. Qu, “A survey of deep learning-based object detection,” *IEEE Access*, vol. 7, pp. 128 837–128 868, 2019.

[7] B. Chen, L. Wang, X. Fan, W. Bo, X. Yang, and T. Tjahjadi, “Semi-fcmnet: Semi-supervised learning for forest cover mapping from satellite imagery via ensemble self-training and perturbation,” *Remote Sensing*, vol. 15, no. 16, pp. 4012–4012, Aug 2023.

[8] J. Wang, C. H. Q. Ding, S. Chen, C. He, and B. Luo, “Semi-supervised remote sensing image semantic segmentation via consistency regularization and average update of pseudo-label,” *Remote Sensing*, vol. 12, no. 21, p. 3603, Nov 2020.

[9] Z. Mao, X. Tong, and Z. Luo, “Semi-supervised remote sensing image change detection using mean teacher model for constructing pseudo-labels,” in *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Jun 2023.

[10] E. B. Alexakis and C. Armenakis, “Evaluation of semi-supervised learning for cnn-based change detection,” *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLIII-B3, pp. 829–836, Jun 2021.

[11] W. Ma, O. Karakuş, and P. L. Rosin, “Confidence guided semi-supervised learning in land cover classification,” *arXiv (Cornell University)*, Jul 2023.

[12] S. Desai and D. Ghose, "Active learning for improved semi-supervised semantic segmentation in satellite images," in *2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, Jan 2022, pp. 1485–1495.

[13] Z. Li, J. Pan, Z. Zhang, M. Wang, and L. Liu, "Mtcsnet: Mean teachers cross-supervision network for semi-supervised cloud detection," *Remote Sensing*, vol. 15, no. 8, pp. 2040–2040, Apr 2023.

[14] Y. Xin, Z. Fan, X. Qi, Y. Zhang, and X. Li, "Confidence-weighted dual-teacher networks with biased contrastive learning for semi-supervised semantic segmentation in remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–16, Jan 2024.

[15] Kaggle, "Urban segmentation - isprs dataset," 2021, https://www.kaggle.com/datasets/aletbm/urban-segmentation-isprs.

[16] A. Tarvainen and H. Valpola, "Weight-averaged consistency targets improve semi-supervised deep learning results," *CoRR*, vol. abs/1703.01780, 2017. [Online]. Available: http://arxiv.org/abs/1703.01780

[17] K. Yamazaki, T. Hanyu, M. Tran, A. de Luis, R. McCann, H. Liao, C. Rainwater, M. Adkins, J. Cothren, and N. Le, "Aerialformer: Multi-resolution transformer for aerial image segmentation," *IEEE Transactions on Geoscience and Remote Sensing*, 2023.

Appendix A

An Appendix Chapter (Optional)

Appendix B

An Appendix Chapter (Optional)