Land Cover and Land Use Detection using Semi-Supervised Learning

Authors

Md Zarif Hossain (170042019)

Fahmida tasnim Lisa (170042020)

Sharmin Naj Mou (170042074)

Supervisor

Dr. Md. Hasanul Kabir

Professor

Co-supervisor

Shahriar Ivan

Lecturer

A thesis submitted to the Department of CSE
in partial fulfillment of the requirements for the degree of
Bachelor of Science in Software Engineering

Department of Computer Science and Engineering (CSE)
Islamic University of Technology (IUT)
Organization of the Islamic Cooperation (OIC)
Gazipur, Bangladesh

April, 2022

Declaration of Authorship

This is to certify that the work presented in this thesis, titled, "Land Cover and Land Use Detection using Semi-Supervised Learning", is the outcome of the investigation and research carried out by Md Zarif Hossain, Fahmida Tasnim Lisa and Sharmin Naj Mou, under the supervision of Dr. Md. Hasanul Kabir and Shahriar Ivan. It is also declared that neither this thesis nor any part of it has been submitted anywhere else for any degree, diploma or other qualifications. Information derived from the published or unpublished work of others has been acknowledged in the text and a list of references is given.

Authors:

Md Zarif Hossain	Fahmida Tasnim Lisa	Sharmin Naj Mou
Student ID: 170042019	Student ID: 170042020	Student ID: 170042074

Supervisor:

Dr. Md. Hasanul Kabir
Professor,
Department of Computer Science and Engineering
Islamic University of Technology (IUT)

Co-supervisor:

Shahriar Ivan
Lecturer,
Department of Computer Science and Engineering
Islamic University of Technology (IUT)

Abstract

There have been considerable advancements in semi-supervised learning in the remote sensing community. It is a technique that uses a small number of labeled data to train a model. Generally, deep learning networks learn from labeled data only. But since finding a huge corpus of a labeled dataset is rare and manually labeling datasets is time-consuming and expensive. And labeling remote sensing satellite images is much more challenging than typical image datasets with good accuracy. Our proposed method aims to solve the problem for labelling unlabelled data with better accuracy. We use a SSL technique with a proper class-rebalancing technique to help solve the imbalanced dataset problem. We do it by creating "artificial" labels and training a model to gain reasonable accuracy. Moreover, it is a common occurrence that datasets are typically class-imbalanced. And if they are trained using it, with a high number of samples, the model becomes biased towards the majority classes and away from minority classes having few examples. This becomes a primary problem to the poor performance of an SSL model. We use a distribution alignment strategy to iteratively redistribute the classes through re-sampling. We showed that our proposed method improve a state-of-the-art SSL method with a tweaked augmentation strategy to generate high-quality pseudo-labels, updating the labeled set handling imbalanced data through re-sampling and also can reduce model bias. This is done on various class-imbalanced satellite image datasets. This method consistently outperforms other methods and greatly reduces the need for labeled data and also solves the issue of class imbalance in datasets.

Acknowledgement

It is a great moment for us to submit our thesis work, which will bring our Bachelor of Science studies to an end. To begin, we would like to express our heartfelt gratitude to Almighty Allah for showering his blessings upon us, which helped us to successfully complete our thesis research. All praises and thanks to Almighty Allah.

Next we would like to convey our gratitude to Dr. Md. Hasanul Kabir, Professor, Department of Computer Science and Engineering, Islamic University of Technology for being our advisor and mentor. His encouragement, suggestions, and insights have been priceless for this thesis. Without his help and direction, this thesis would not have been able to follow the right research path. From the first step of thesis topic introduction, study area selection, algorithm proposal, revision, and implementation, his insightful opinion, time, and input were helpful to complete our thesis work effectively.

We would also like to extend our gratitude to Shahriar Ivan, Lecturer, Department of Computer Science and Engineering, Islamic University of Technology for his enormous help in directing us toward our objective and guiding us in times of need. We appreciate his constant and enthusiastic mentoring, constructive criticism, and excellent recommendations.

Then we would like to express our gratitude to all the respected jury members of the thesis committee for their insightful comments and criticism, which helped us better our work and all the faculty members of the Computer Science and Engineering department of Islamic University of Technology for providing us with a congenial and supportive work environment.

Finally, we want to express our gratitude to our friends and family. This job would not have been completed without their constant support and encouragement throughout the program.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In recent times there seems to be a significant abundance of satellite and aerial remote sensing images all because of the advancement of earth observational technology. Nowadays, there are numerous satellites revolving around the earth collecting huge remote sensing data. And the amount is growing exponentially every day. It has been said that there were 300 satellite launches for collecting earth satellite data in the year 2018 alone [14].

As the amount of remote sensing data grows, so does the demand for effective image processing. And at the same time, for this, numerous deep learning approaches have been created. Therefore, in the remote sensing community, there has been study for processing satellite and aerial imagery efficiently using deep learning methods [15] [16] [17]. Deep learning methods rely heavily on labeled data for training neural networks. Usually, these labeled data have been manually annotated by experts. That is why labeled data is limited. And manually annotating satellite images is difficult. This problem of manually annotating labels can be solved through semi-supervised learning(SSL). SSL trains a neural network using only a few labeled data and a large amount of unlabelled data. Semi-supervised learning combines unsupervised learning which is unlabeled training data and supervised learning which is only labeled training data. And there have been significant advancements in the last few years in the domain of semi-supervised learning(SSL). These methods save time and cost and also reach the accuracy of the fully-supervised techniques applied in the same situations [5] [3] [4]. And there have been few works where semi-supervised learning has been applied to process satellite images by Liu et al. [18] Wu and Prasad [19].

Moreover, it is seen that the datasets available are class-imbalanced i.e. majority classes have large number of samples and minority classes having few samples. Models trained on class-imbalanced data become biased Models trained on data with unequal distributions of classes become biased favouring majority classes and this becomes a primary reason for an SSL model to perform poorly. Many methods have been proposed to help reduce model bias like the work done by Buda et al. [20] where they introduce a technique for re-sampling. There have also been other works like re-weighting and averaging to solve model bias. But all of these methods rely on labeled samples. SSL on imbalanced data has not been studied extensively. It is relatively new to the scene. There have been works to handle imbalanced data by using distribution alignment with SSL algorithms by Wang et al. [21], Mayer et al. [22], Kim et al. [23], Chen et al. [24], He et al. [25], Wei et al. [9].

Figure 1.1: LULC scene classification. Images taken from AID dataset [1]

In this work, we propose a semi-supervised learning method built on a recent advancement [5] together with a distribution alignment strategy to tackle the issue of labeling land use and land cover images and handle the class imbalance problem. We compare our results with supervised learning, MSMatch [8] and FixMatch[5] with tweaked Augmentation on three datasets, EuroSAT benchmark dataset [10] [26], UC Merced Land Use (UCM) dataset [11] and WHU-RS19 [12] [13]. Our method uses a class-rebalancing strategy to re-train an improved semi-supervised learning model with a specific augmentation strategy that involves collecting pseudo-labeled data from the unlabeled set to expand the initial labeled set. Each completely trained SSL model is referred to as a generation.

To retrain the semi-supervised learning model with tweaked augmentation, the pseudo-labeled samples are taken from the unlabeled set and incorporated into the labeled set after each generation. We employ a stochastic update technique instead of updating the labeled set with all pseudo-labeled samples generated from the SSL model, in which samples are picked with a high probability(threshold crossing 95%) if they are from minority classes since they are more likely to be right predictions. The data distribution derived from the labeled set determines the updating probability. As a result, our proposed technique reduces pseudo-labeling bias and increases the accuracy of the test set.

We show that our proposed method outperforms FixMatch [5] under custom augmentation by 2.32% accuracy in the case of UCM dataset, 1% improvement in the case of EuroSAT and 2.31% in the case of WHU-RS19.

1.1 Overview

1.1.1 Semi-Supervised Learning

Semi-supervised learning is a machine learning technique that trains a model with limited amount of labelled training data and a huge amount of unlabelled data. This type

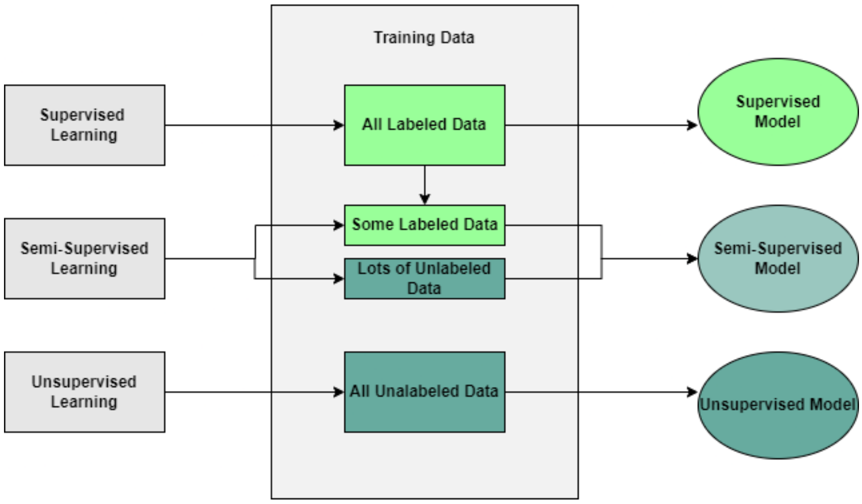

Figure 1.2: Outline of the different machine learning approaches

of learning problems are a bit challenging as they are neither supervised or unsupervised.

Every machine learning model learns from data. In supervised learning, the machine learning model learns from labeled data. That is, it predicts a label and then calculates the difference between the prediction and the actual label. The model tries to minimize the difference and then increase the accuracy. In unsupervised learning, the model tries to identify patterns, trends, or categories. So, there is no need for labels.

Semi-supervised learning merges supervised and unsupervised learning. This uses a small amount of annotated data and a huge corpus of unlabeled data. This relieves the need to have a huge amount of labeled data. Finding a good amount of labeled data is unrealistic as manual labeling takes up time and cost. Semi-supervised learning helps create artificial labels and train a classifier using the labeled data and unlabelled data with ”pseudo” labels.

How semi-supervised learning works:

- Train the model on a small amount of labelled data, similar to supervised learning, until it gives a good accuracy.
- Then use the model on unlabelled data to predict outputs i.e the pseudo-labels for the unlabelled data.

- Retrain the model with pseudo and labelled datasets together in order to decrease error and increase the model's accuracy

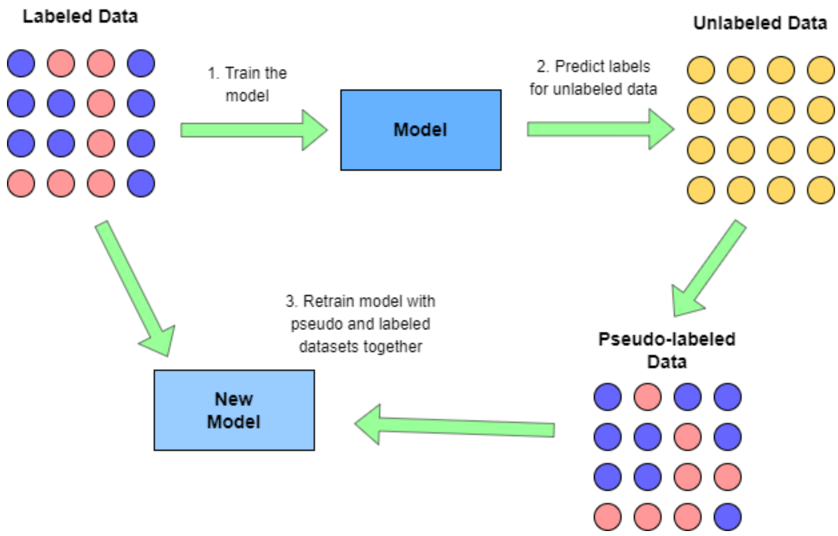

Figure 1.3: Basic technique of semi-supervised learning

1.1.2 Land Use and Land Cover Detection

Land Use/Land Cover (LULC) detection is the grouping or classification of human actions and natural compounds on the earth surface. Land cover refers to the surface cover on the ground like flora, water, urban infrastructure, forest and Land use refers to the purpose the land serves, for example, airports, highways, or agriculture. Land cover and land use (LULC) detection refer to the identification of land types namely - forests, water bodies, soil, desert, snow areas, and land use which is the human interaction with the physical environment, are - industrial areas, human habitats, crop farming, etc.

This research has extensive applications in multiple sectors. Some of the applications of Land Cover and Land Use Detection are:

- Land conservation
- Disaster Management
- Sustainable development

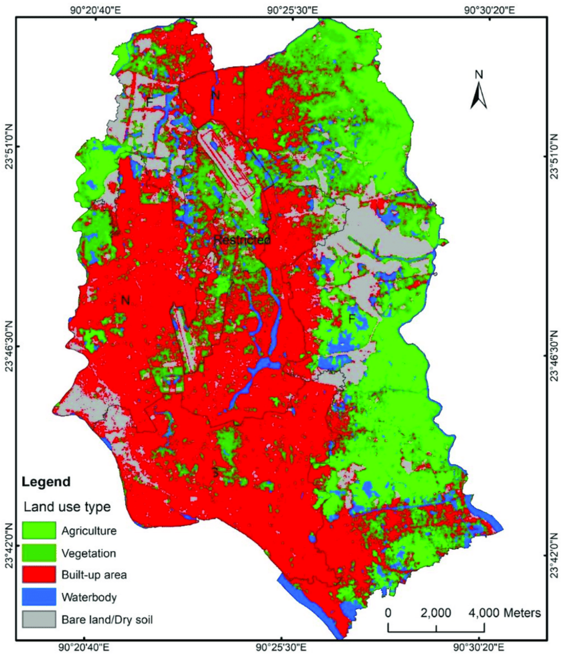

Figure 1.4: Land use land cover (LULC) map for the year of 2013, Dhaka [2]

- Land resource management
- Urban Expansion
- Legal Boundaries for Tax and Property Evaluation
- Weather Predictions
- Vegetation Mapping
- Natural Hazard Detection, etc.

1.2 Motivation

A remote sensing picture usually includes a number of different ground objects. As a result, from the method described by Cheng et al. [17], detecting a scene from a remote sensing image containing a diversity of ground objects is challenging for the model to achieve sufficient accuracy. Satellite images have much more pixel and resolution than a classical photograph. To work with these type of image data, experts need to manually annotate rigorous number of data.

Having a method which can automatically label these type of data and work with high resolution images introduces a lot of possibilities in land scene classification. Semi-supervised land scene classification can correctly generate labels for land cover and land use detection on remote sensing data with limited amount of labels and reach a reasonable accuracy. Although significant research has been done in this area, it is still a relatively new field of study.

1.3 Problem Statement

There has been sufficient work in the domain of land cover and land use in the past, but most of them have been geared toward balanced datasets. Or it required manual expert labeling. Automatically identifying land cover and land use using a small amount of labeled data is relatively new. Having explained the various applications of and cover and land use and our motivation for this domain, we can finally declare our problem statement as follows:

> The primary objective of our thesis is to develop an effective method for detecting land cover and land use from abundant remote sensing data. In particular, we aim to use unlabelled data with limited labelled data and investigate semi-supervised learning methods to achieve better performance in classifying remote sensing images.

1.4 Research Challenges

1.4.1 Challenges of Remote Sensing Image Scene Classification

In scene classification tasks, the main goal is to correctly classify the scene from the image and label it correctly. For example, correctly classifying a remote sensing image from rural to forest, crops, mountains or riverine areas, from urban to industrial, residential, commercial areas. Land cover includes the forest, river, desert, beach, pond etc. and Land use includes airports, freeways, bridge, harbour, crops farmland, residential areas, industry, resorts etc. Usually, A remote sensing image usually includes a number of different ground objects. Recognizing a scene from a remote sensing image comprising multitude of ground objects is strenuous for the model to provide adequate accuracy, as Cheng et al. [17] show in their work.

The challenges of remote sensing scene classification using semi-supervised learning is given below:

Challenges faced due specificness of satellite data: If we compare a satellite image with a classical photograph, we can see that satellite images have much more pixel and resolution than a classical photograph.

Also, a classical photograph has three channels, red green and blue but a satellite image can have upto dozens of channels including RGB, these are called multispectral(MS) images.

There has been several datasets established for scene classificaton tasks for image processing. Like EuroSat by Helber et al. [10], UCM and the Aerial Image Dataset (AID)

Challenges faced in working with satellite data:

- Big intraclass variance (variety within class)
- High interclass similarity (low similarity between classes)
- Huge variety of objects
- Dead pixels
- Cloud appearances, shadows, haze
- Color casting due to atmospheric effects
- Overexposure due to sun
- Imaging altitude variation

Ground objects appear in different style, shape and scale in remote sensing images. For example, a school buildings, bridges and airports can be of different style, shapes and scales. Also some of the images appear in different situations due to the atmospheric conditions, weather, cloud, haze etc. in the same class. It causes intra-class diversity. Sometimes satellite take images of an area at different angles and due to this sometimes the sensors cannot capture with a proper resolution. And since a huge varierty of objects can appear in a single scene, it becomes difficult for a model to provide a single label to a scene with various objects appearing in different style, shape and scale.

1.4.2 Challenges of Imbalanced Datasets

The class distribution of labeled and unlabeled data being balanced is a typical assumption made inherently during the design of datasets. However, in many real-world cases, this assumption is incorrect, and poor SSL performance is the result. In fact, in SSL, where insufficient label information prevents adjusting the unlabeled set, data imbalance presents much more complications. When pseudo-labels are created by an starting model trained on imbalanced data they can be detrimental to the model and they're biased towards the majority classes. And therefore, training with these biased pseudo-labels aggravate the bias and degrades model quality. The poor performance of conventional SSL algorithms on imbalanced data is mostly due to limited recall on minority classes. In terms of recall, the biased model trained on imbalanced data does well for majority classes, while it does worse for minority classes in terms of precision. Therefore, imbalanced datasets pose as a research challenge in our work.

1.5 Objective

To develop a semi-supervised learning method that can detect land cover and land use changes with high accuracy and do scene classification with "correct" labels that can save time and cost which would otherwise be spent on manual labelling.

1.6 Contributions

- We have worked with more diverse augmentation strategy. Tweaking the augmentation parameters helped us to get better Consistency Regularization and the better Consistency Regularization helped to get better accuracy on Satellite images.
- We have increased recall of imbalanced classes by class rebalancing.
- We have worked on Distribution Alignment to get more accurate pseudolabels.
- We have reduced model bias towards majority classes by handling data imbalance.
- Finally we have correctly labeled data with reasonable accuracy.

1.7 Organization of the Thesis

The remainder of this article is organized as follows:

Chapter 2 gives a Literature review discussing the recent advancements in Semi-Supervised Learning, different approaches of Semi-supervised learning and Distribution Alignment used in Remote Sensing Scene Classification.

Chapter 3 introduces our proposed method. This section describes the different experiments we performed in detail.

Chapter 4 presents result analysis and comparison of our proposed method with other methods.

Chapter 5 presents an overall conclusion of our thesis and discusses our future plan of work.

Chapter 2

Literature Review

The importance of labelled data in training a neural network in deep learning methods is a serious limitation. And getting labelled data for satellite imagery is vexing as satellite and aerial data are collected using various sensors and applications and these data are in various spectral bands and resolution. So training a classifier without labelled data is particularly difficult. That is why, Semi-supervised learning (SSL) comes into the picture. In the following sections, we discuss the recent advancements in semi-supervised learning techniques and the recent works done on Land cover and Land use change (LULC) detection.

2.1 Recent Advancements in Semi-Supervised Learning

Pseudo-labeling is a semi-supervised learning method where “artificial” labels are created to label the unlabelled data. First, the model gets trained with labeled data and then the labels for unlabelled data are predicted. After that, the model gets retrained with both the labeled and unlabelled data. This aids in the development of a model with minimal error and eliminates the need for time-consuming and costly human data labeling.

In the paper, Lee et al. [27] use a denoising autoencoder and dropout to handle noisy data and boost performance. This pseudo labeling technique became very popular. However, this paper has certain limitations. This method is sensitive during initial predictions and pseudo labeling works best in problems that involve clustering assumptions. Also if there are very few labels, this technique does not give good performance.

Jeong et al. [28] used a consistency regularization(CR) based method to enhance image object detection in a dataset that is abundant in unlabeled data. The authors demonstrated two object detector models: a one-stage detector and a two-stage detector. In the one-stage detector model, the object detection performance was enhanced by using a combination of consistency loss in both classification and localization of the labeled and unlabeled images and adding it to the original object detectors' classification and localization loss. They used the Jenson-Shannon divergence (JSD) as the consistency regularization loss. And they also added Background elimination(BE) to the consistency losses. They used the Mean Average Precision (mAP) as a metric. In the two-stage detector model, they use a Regional Proposal Network (RPN) network where they only pass a specific feature from the original image to it and compute a supervised RPN loss. Using the feature map from the backbone network and Region of Interest(RoI) pooling from the RPN, outputs compute the cost for network training. It is seen that training with consistency loss and also applying background elimination (BE) significantly improves the performance of object detection and provides the best mAP. It makes it much more efficient and faster than traditional Semi-supervised learning(SSL) to object detection problems. In the one-stage detector, the proposed method can be applied to both labeled and unlabeled images. It is shown to be helpful for classification and localization. However, the two-stage detector has been shown to have less improvement than the single-stage detector due to not applying consistency loss in RPN.

MixMatch by Berthelot et al. [3] is a semi-supervised learning approach that combines entropy minimization, consistency regularization, and traditional regularization. In this paper, the authors used data augmentation and sharpening function to boost performance. Data augmentation is applied to input data then "guessed" labels are generated and after that, the decision boundary is "sharpened" to improve the consistency of the artificially labeled unlabelled data. After that, the unlabelled and labeled data is shuffled and combined using MixUp algorithm (technique proposed by Zhang et al. [29]) to artificially increase data and then fed to the model to improve accuracy.

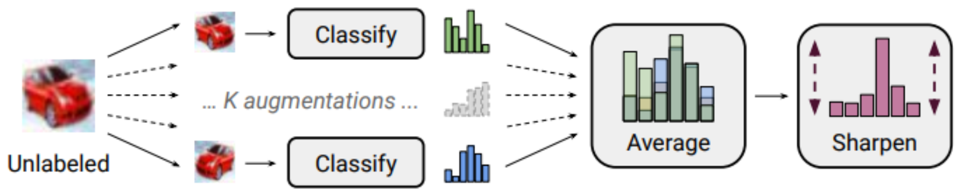

Figure 2.1: MixMatch "guessed" labelling process [3]

Using MixUp by Zhang et al. [29] and the hyperparameters seems to provide the strongest contribution to the performance of Data Augmentation and Sharpening benefits accuracy but the MixUp seems to be the most important factor. A lot less data is needed to train which is great because MixMatch uses the MixUp function to artificially increase data. Even though MixMatch consistently beats state-of-the-art methods and costs to generate the transformation, MixUp's accuracy is significantly higher.

ReMixMatch by Berthelot et al. [4] is an improved version of MixMatch by Berthelot et al. [3]. Here the consistency regularization component of MixMatch is replaced by augmentation anchoring. The middle and green graphs of the figure are the prediction for a weakly augmented image. The blue graph is the target for predictions on strong augmentations of the same image. Augmentation anchoring creates a weakly augmented version of each unlabeled input initially which is called "anchor". Then it generates multiple strongly augmented versions of the same unlabeled input. They used a variant of AutoAugment based on control theory augmentation dubbed as CTAugment. The authors use augmentation anchoring to urge each output to be near to the prediction (guessed label) for a weakly enhanced version of the identical input.

Then, distribution alignment was added to MixMatch by altering the "guessed labels". This ensures that the distribution of unlabeled data predictions matches the distribution of labeled data given. Distribution alignment was needed because the dataset's marginal class distribution was uniform. The empirical ground-truth class distribution is split by the average predicted results on unlabeled data to update the guesses label distributions.

After that, they updated labeled samples and their labels for shuffling. The weakly

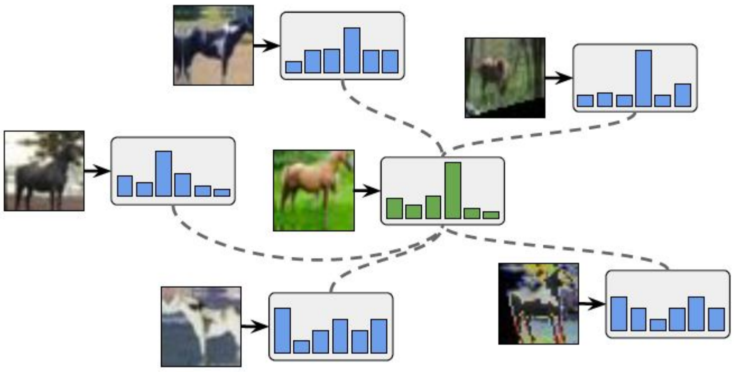

Figure 2.2: Augmentation Anchoring [4]

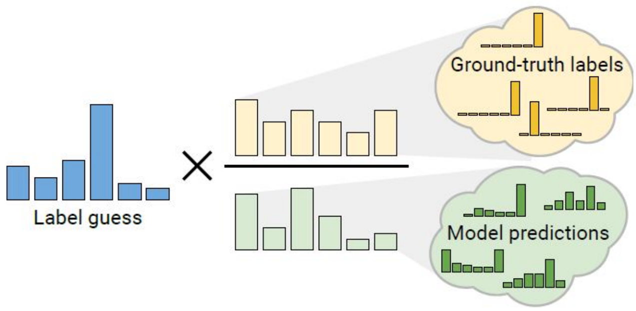

Figure 2.3: Distribution Alignment [4]

enhanced unlabeled examples are then combined with the heavily augmented unclassified example and guessed label. Lastly, They have done MixUp followed by the technique of Zhang et al. [29].

FixMatch by Sohn et al. [5] is an algorithm that brilliantly combines consistency regularization and pseudo-labeling. At first, the model gets trained by the available labeled data and gets hyper tuned.

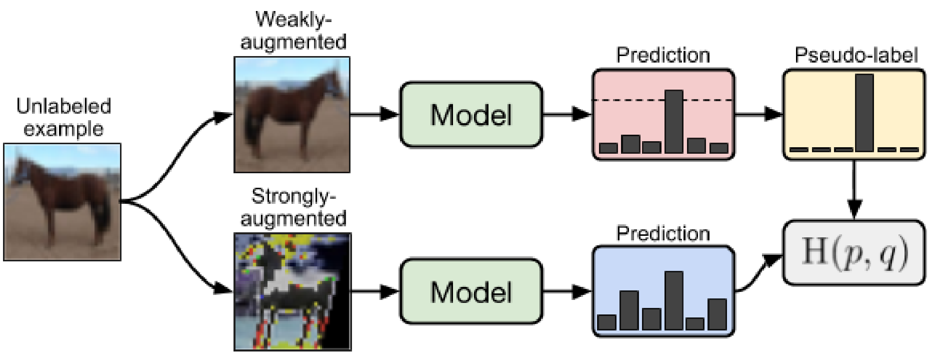

Figure 2.4: Diagram of FixMatch [5]

After the model gets hyper tuned, the next phase of the algorithm works in two pipelines, at first, a weakly-augmented image is given into the model to obtain prediction as shown in the red box in the diagram 2.4 above. Then the model assigns a probability to a class. If the probability is above a threshold (shown as a dotted line in the figure2.4), the prediction is then converted to a one-hot pseudo-label. Then, the authors use the same unlabeled image to compute the model's prediction for a strong augmentation. They train the model to make its prediction on the strongly-augmented version of the image by running a cross-entropy loss with pseudo-labeling, here they take the pseudo-label as a true label and try to minimize the loss.

For the weak augmentation strategy they chose Standard flip and Left-Right shift and for the strong augmentation strategy, they worked with RandAugment by Cubuk et al. [30] with cutOut Strategy and CTAugment with Cutout. In both RandAugment and CTAugment, they used almost similar augmentations like Autocontrast, Brightness, Contrast, Invert, Rotate, etc.

As stated above, the authors used CutOut Strategy with both augmentation strategies and in this paper, they emphasized on CutOut Strategy to have a better result. Randomly selecting space from the image and replacing it with random size pixels.

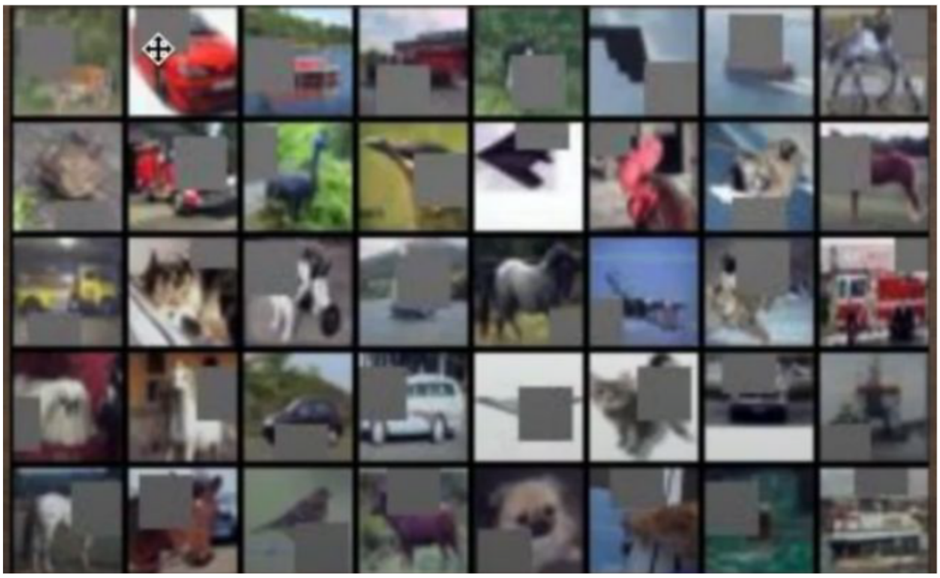

Figure 2.5: cutOut Strategy [5]

Randomly selecting space from the image and replacing it with random size pixels. Without CutOut, Fixmatch's error rate drastically increases.

To conclude, it is a very simplistic approach to SSL. Despite its simplicity, FixMatch achieves a state-of-the-art performance across varieties of benchmarks. It achieves 94.3 % accuracy on CIFAR-10 with just 250 labels for 10 classes and 88.61% accuracy with just 40 labels (4 data per class). Here, sometimes the accuracy varies on the quality of the labeled images. If the quality is too poor accuracy gets hampered. Many other SSL algorithms come at great calculation costs and complex learning algorithms that achieve such great performance. However, it has a few lackings such as the algorithm is Substantially dependent on Augmentation Strategy, and if Augmentation Strategy is replaced Error Rate significantly increased. nevertheless, Fixmatch achieves such great accuracy without having such computational cost and complexity.

2.2 Semi-supervised learning and Remote Sensing Scene Classification

2.2.1 Remote Sensing Scene Classification

There are many publicly available data on remote sensing images. And many machine learning and deep learning methods have been used to utilize these data. There has been convolutional neural network (CNN) based remote sensing image classification, auto-encoder based, and also, generative adversarial network (GAN) based remote sensing image classification. And both supervised, self-learning, and semi-supervised learning has been applied for remote sensing image classification. For instance, Othman et al. [31] suggested a remote sensing image scene classification method based on convolutional features and a sparse autoencoder in their paper. Though autoencoder-based remote sensing image classification has gotten good results, they can not completely utilize scene class information, these techniques cannot learn the optimum discriminating features to recognize distinct scene classes. In works, by Teng et al. [32] and Ma et al. [33], there have been many GAN-based remote sensing image classifications in a semi-supervised manner. In the paper by Cheng et al. [34] a new technique for learning discriminative convolutional neural networks (D-CNNs) was proposed. Cheng et al. [17] surveyed auto-encoder-based, GAN-based, and CNN-based remote sensing image classification. There have been few reports of GAN-based remote sensing image classification and CNN-based image classification is superior to it. These CNN-based techniques require annotated images for better performance and results, that is where semi-supervised learning comes in. Semi-supervised learning helps artificially sample unlabelled images to give better results.

2.2.2 Remote Sensing Image Classification using Semi-Supervised Learning

Deep neural networks, often need a high number of labeled samples for training. And these labeled samples are manually annotated by professionals. This is especially aggravating when it comes to satellite imagery or remote sensing image datasets as these are not easily labeled by humans. One way to solve this problem is using semi-supervised learning (SSL) methods. These are designed to train machine learning algorithms with a limited collection of labeled training data and a usually bigger volume of unlabeled training samples. There have been many works proposed on this. Like, Wu and Prasad [19] proposed deep learning for hyperspectral image classification by making use of deep convolutional recurrent neural networks (CRNN). So it provides high-quality pseudo-labels, leading to better deep neural network initialization.

Cenggoro et al. [6] classified imbalanced LULC data using variational semi-supervised learning. The figure illustrates the architecture of VSSL technique used to resolve the imbalance problem in the Land Use and Land Cover data of Jakarta City.

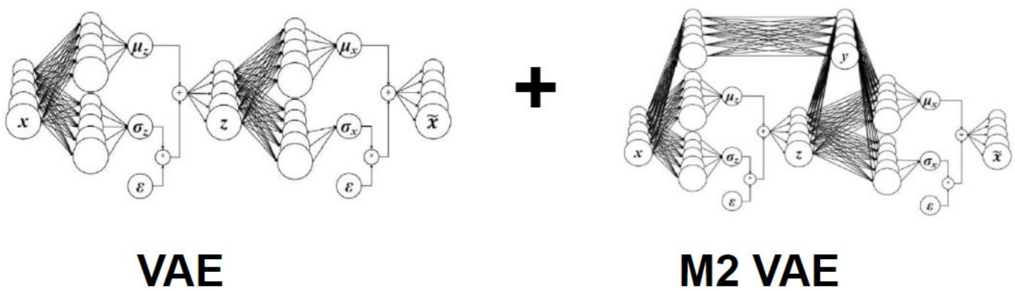

Figure 2.6: Variational Semi-Supervised Learning (VSSL) technique [6]

VSSL is used for deep generative model. It is built by combining a standard Variational Autoencoder (VAE) and a modified version of VAE called M2 VAE. Variational

Autoencoder (VAE) is a variant of autoencoder. The VSSL learns from both labeled and unlabeled data simultaneously. The VSSL makes use of a deep learning model. However, because this strategy assumes that the data is unbalanced, it may limit the model's capacity to generalize.

In this paper, Fan et al. [7] proposed Semi-MCNN for Urban Land Cover classification. Here they have used Submeter HRRS Images. They proposed a semi-supervised learning strategy in addition to multiple deep learning-based CNNs. First, an ensemble teacher model trains the training data. Then, the fine-tuned ensembled teacher model selects samples and generates a new dataset. Finally, the newly generated dataset trains a student model to get the final model.

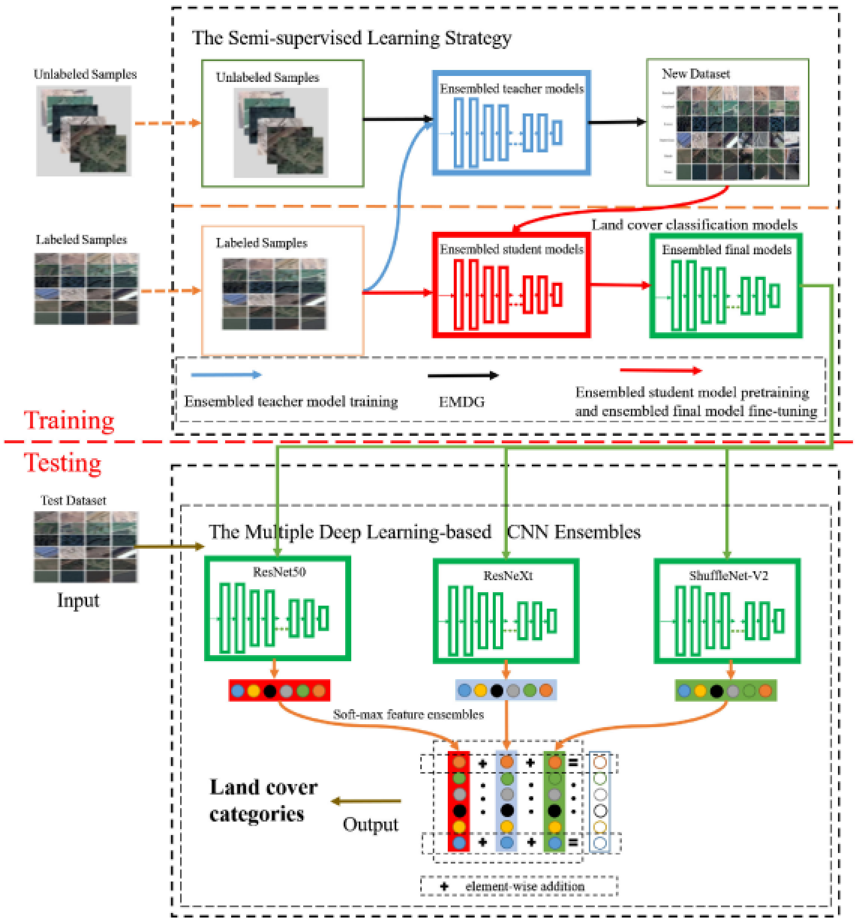

Figure 2.7: Flow of Semi-MCNN[7]

The MSMatch method by Gómez and Meoni [8] takes advantage of recent improvements as well as new neural network architectures to address the challenge of land scene classification. The authors have followed the pipeline of Fixmatch by Sohn et al. [5] on satellite image dataset. To create a pseudo-label for the image, the weakly augmented image is used.

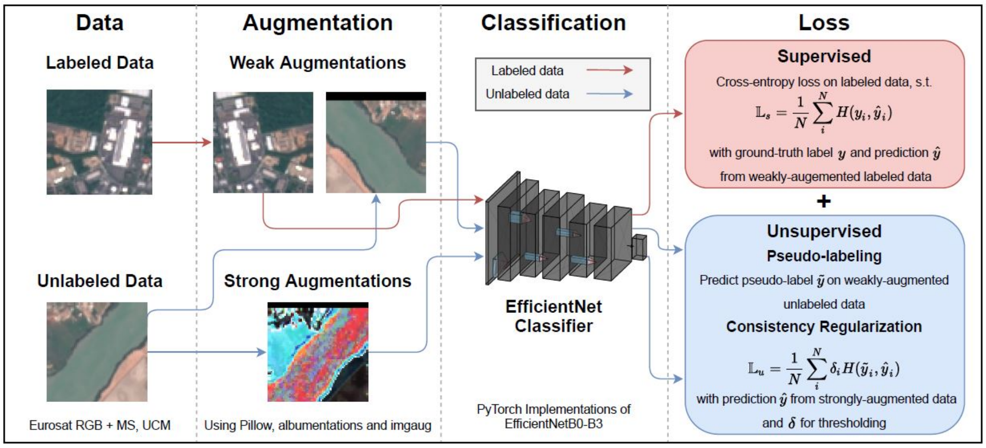

Figure 2.8: Overview of the processing pipeline of MSMatch [8]

They computed cross-entropy loss between a pseudo-label on the weakly augmented image and the model's classification of the strongly augmented image. Thus, consistency regularization is employed. Pseudo-labeling uses the model to automatically label unlabeled data and consistency regularization ensures that the model should predict the same output for similar inputs.

2.3 Distribution Alignment

Distribution Alignment (DA) by Berthelot et al. [4] is particularly well suited to class-imbalanced circumstances, aligning the distribution which the model predicted in case of unlabeled samples with the class distribution of the labeled training set. The mismatch of feature distributions between labeled and unlabeled samples is a result of SSL overfitting. The feature distribution alignment method of Mayer et al. [22] is highly effective when just a small number of tagged samples are used. This method explains why feature distribution alignment emerges and how to prevent it. The technique of Augmented Distribution Alignment by Wang et al. [21] used an adversarial training strategy inspired by domain adaptation efforts to decrease the distribution distance of labeled and unlabeled data, and generated training samples which are pseudo to solve the labeled data's small sample size issue. The feature-based refinement and augmentation method named FeatMatch by Kuo et al. [35] generates a diverse range of complex transformations based on information retrieved through clustering. By keeping features computed across iterations in the memory, this approach avoided the need for notable additional computation. These are then applied as part of the regularization loss which is consistency-based, along with standard image-based augmentation. The DASO approach by Oh et al. [36] works with the semantic pseudo-label and the linear one class. Another technique, Uncertainty-Aware Self-Distillation (UASD) by Chen et al. [24], generates soft targets that prevent catastrophic error propagation and enable learning from unconstrained unlabelled data including out-of-distribution (OOD) samples. This is based on a unified formulation that combines self-distillation and OOD filtering. He et al. [25] proposed a method named DARS which produces unbiased pseudo labels. The true class distribution of the labeled data can be matched with the pseudolabels.

2.3.1 DARP: Distribution Aligning Refinery of Pseudo-label for Imbalanced Semi-supervised Learning

In DARP by Kim et al. [23], to adjust the pseudo-labels that are produced from a biased model, a convex optimization problem is stated and a simple iterative algorithm is constructed.

With a proven guarantee, the DARP method is a procedure for solving the proposed (convex) optimization. It solves the original optimization's Lagrangian dual to obtain the unique optimal solution. When executing DARP, several minor and noisy items in the original pseudo-labels are eliminated and this improves the condition of the refined pseudo-labels even more.

2.3.2 CReST: A Class-Rebalancing Self-Training Framework for Imbalanced Semi-Supervised Learning

CReST by Wei et al. [9] retrains a baseline SSL model again and again. It uses a labeled set which is expanded by adding pseudolabeled samples from an unlabeled set.

Here, minority class pseudolabeled samples are selected more frequently by following an estimated class distribution.

A class-rebalancing selftraining strategy (CReST) frequently samples pseudo-labeled data which is gained from the unlabeled set to replace the original labeled set. To retrain an SSL model, pseudo-labeled samples from the unlabeled set are added to the labeled set after every generation (fully-trained baseline model). Samples are selected with a higher probability if they are identified as minority classes, as these are more likely to be accurate predictions. The labeled set's data distribution determines the updating probability.

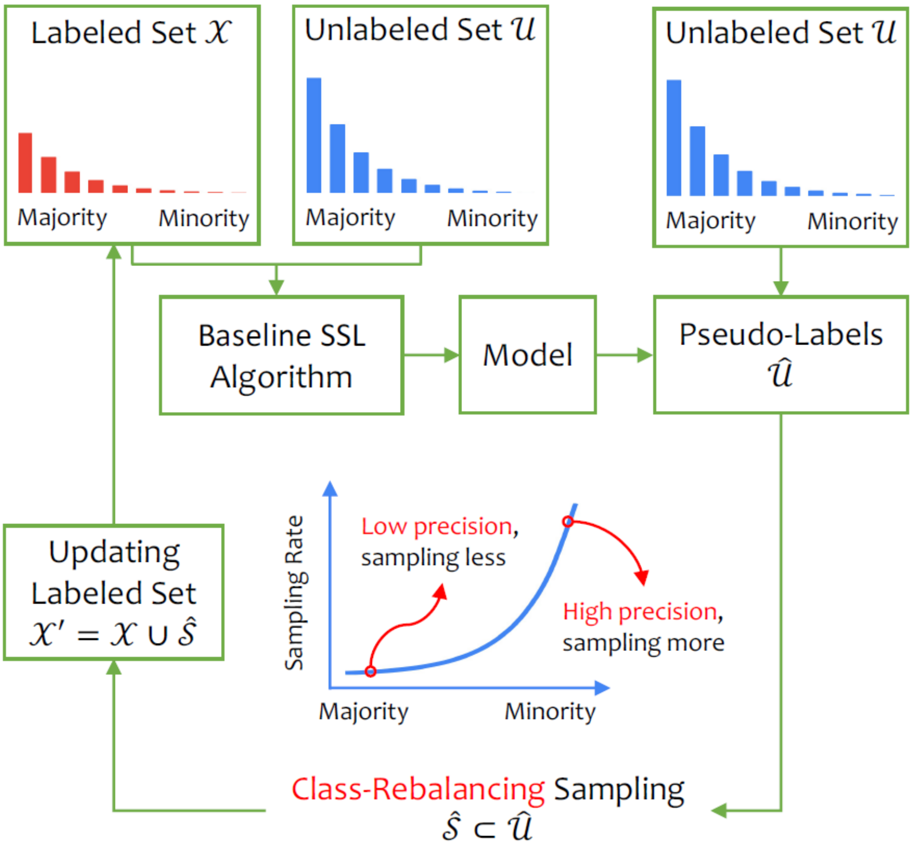

Figure 2.9: Adaptive training of CReST (Class-Rebalancing Self-Training) [9]

The proposed technique is initially tested on the long-tailed CIFAR10 (CIFAR10-LT) and long-tailed CIFAR100 (CIFAR100-LT) described in the papers by Cui et al. [37] and Cao et al. [38] respectively. The backbone was Wide ResNet-28-2 by Zagoruyko and Komodakis [39], which was based on Oliver et al. [40], Sohn et al. [5]. They also evaluated CReST on ImageNet127 by Huh et al. [41], Wu et al. [42] to verify its performance on large-scale datasets.

SSL techniques are used to leverage the unlabeled and labeled data to generate a teacher model in the very first step which is even better, rather than exclusively training on labeled data. More crucially, rather of incorporating each sample in the labeled set in the second phase, the labeled set has been enlarged with a subset. The more unlabeled samples that are determined as that class are added in the labeled set, the less frequent that class is.

CReST has been improved by introducing distribution alignment by Berthelot et al. [4] with a temperature scaling factor to adjust the alignment strength using multiple generations. This allows data distributions of prediction to be more adjusted to reduce model bias. As a result of the suggested technique, the bias is reduced, and the accuracy of balanced test set improves.

CReST has been compared to DARP by Kim et al. [23], created particularly for data imbalances. MixMatch and FixMatch are drop-in additions to basic SSL algorithms that are used in both DARP and this technique. The datasets utilized in CReST are the same ones used in DARP. On MixMatch, the model regularly achieves up to a 4.0% accuracy improvement over DARP for all three imbalance ratios, and up to a 2.4% accuracy gain on FixMatch.

But CReST is complex and time-consuming because distribution alignment with temperature scaling has been added to this method.

Chapter 3

Proposed Method

This chapter narrates our proposed methodologies in details. In the first section, we discussed about the general overview of our proposed method's pipeline. In the subsequent sections, we discussed about the different parts of our method's pipeline and how they are contributing to improve accuracy and reduce model's bias towards major classes.

3.1 General Pipeline Overview

The objective of our proposed method is to have good accuracy on imbalanced satellite dataset. Our proposed method helps to rebalance the dataset by each generation and moves towards removing model biasness to the majority classes.

Figure 3.1 illustrates a basic overview of our pipeline. At first a base SSL algorithm is used and model gets trained with available labeled data and unlabeled data. Here Fixmatch is used with our tweaked augmentation as a base SSL algorithm. Then in the next phase, After the model gets confident enough it pseudolabels the images from Unlabeled dataset. If the pseudolabel's confidence reaches the predefined threshold then we move to the last phase. In this phase, pseudolabels from the minority classes and with high precision are chosen with a higher sampling rate and pseudolabeles that are from majority classes are chosen with lower sampling rate to rebalance our dataset.

This completes one generation. The algorithm runs for several generations to remove model bias towards majority classes.

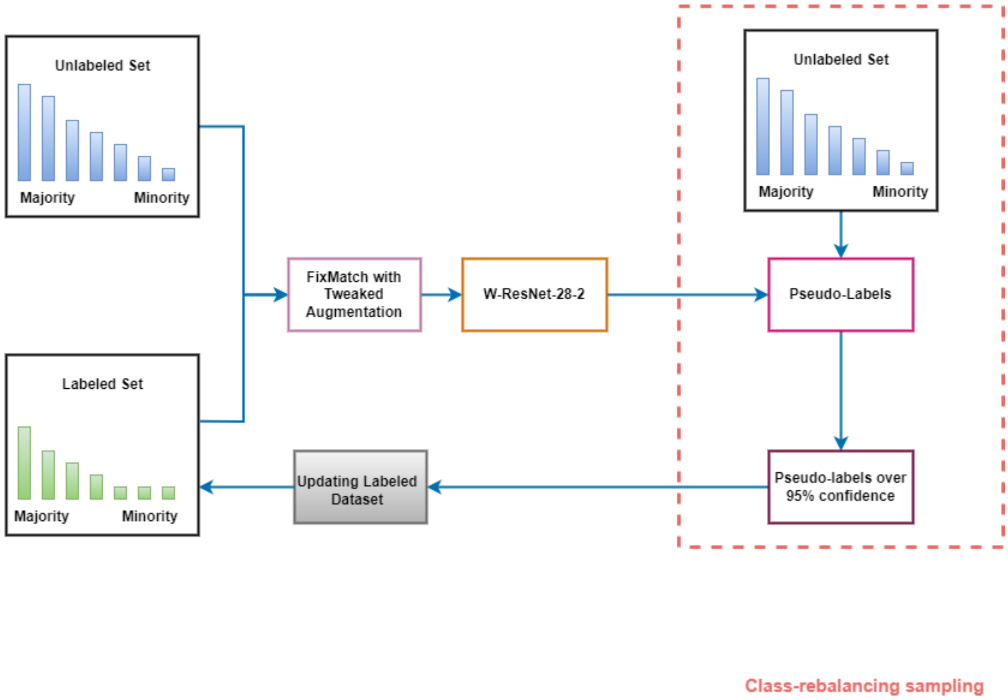

Figure 3.1: Proposed Method

3.2 Framework

Here, we will discuss about the two phases of our proposed method.

3.2.1 Baseline SSL Method: Fixmatch with Tweaked Augmentation

FixMatch algorithm is used here for leveraging unlabeled data, that brilliantly combines consistency regularization and pseudo-labeling. At first the model is trained by the labeled data which is available. Then the model is hypertuned. After that, in next phase algorithm works in two pipeline, at first a weakly-augmented image is fed into the model. This helps to obtain prediction as shown with blue line in the diagram 3.3 above. When the model assigns a probability to any class and it crosses the threshold, then the prediction is converted to a one-hot pseudo-label. We determined the

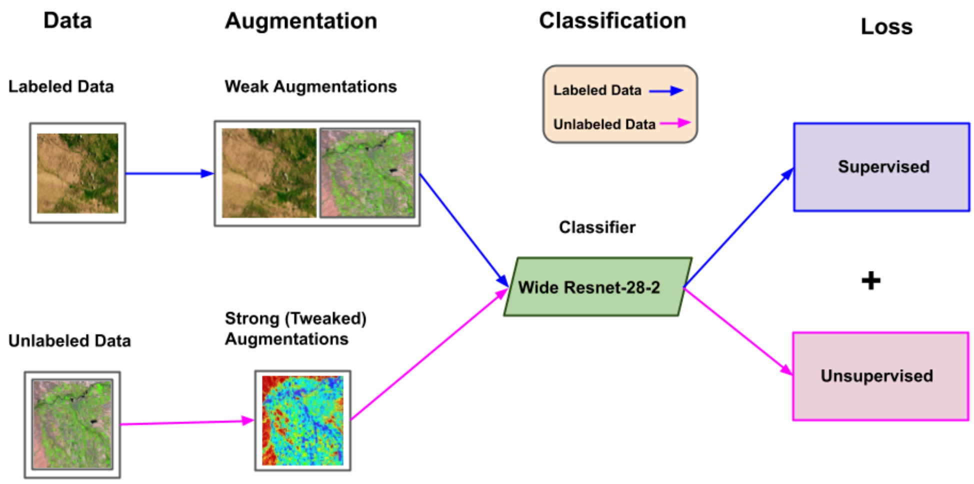

Figure 3.2: Baseline SSL Method: Fixmatch

model's prediction for a strong augmentation of the unlabeled image as shown with pink line in the diagram 3.3 above. Then we trained the model to make it's prediction on the strongly-augmented version of the image by running a Cross-entropy loss with pseudo-label, here we take the pseudo-label as true label and try to minimize the loss.

For the weak augmentation strategy we chose Standard flip and Left-Right shift and for the strong augmentation strategy we worked with RandAugment by Cubuk et al. [30]. For RandAugment we have used Autocontrast, Brightness, Contrast, Invert, Rotate etc. We have tweaked the augmentation parameters to get better result on remote sensing datasets. This helped us to get better accuracy on Satelite imagery using FixMatch.

3.2.2 Class Rebalancing

In this section, we discuss about how our proposed method is actually handling class imbalance. After having confident pseudolabels, if the pseudolabel is confident enough and if it belongs to minority classes, the pseudolabeled image gets added to the labeled set. Now, the pseudolabels from minority classes are added with higher sampling rate to increase the number of samples in minority classes of labeled set. This sampling rate is decided by predefined sampling hyperparameter tuner α = 1/3. And with the help of equation 3.1 we calculate the adaptive sampling rate.

$$\mu_l = \left(\frac{N_{L+1-l}}{N_1}\right)^\alpha \tag{3.1}$$

From the figure 3.3, we can see that the minority class pseudolabels are getting added at higher sampling rate. and from the equation 3.1 we get higher sampling rate for minority classes and lower sampling rate for majority classes. After getting added

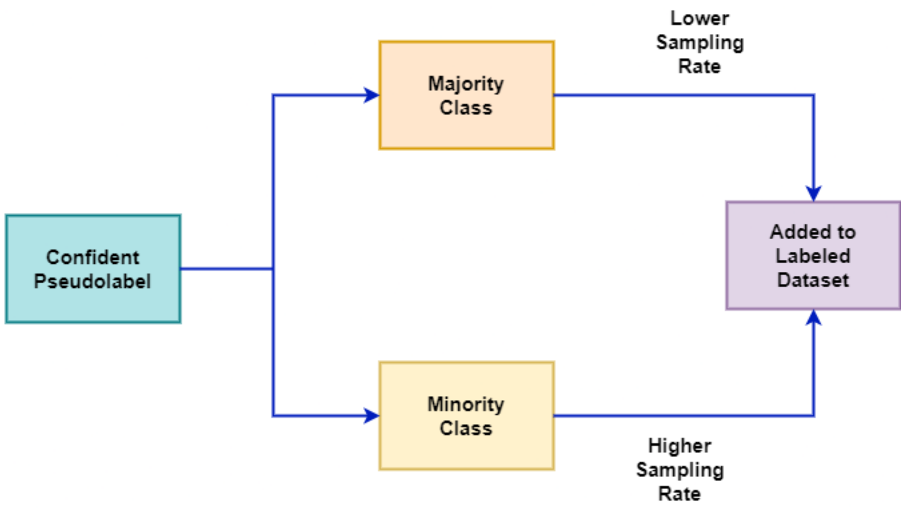

Figure 3.3: Class Rebalancing

to labeled set, the model gets retrained with the newly added data and the previous data and follows same pipeline. By finishing each iteration the algorithm completes one generation and each generation new pseudolabels are added and the model gets retrained. As new pseudolabels are added to minortiy classes the labeled set gets more balanced each generation and by going through multiple generations the labeled set gets balanced and reduces model's bias.

Chapter 4

Result Analysis

We experimented the performance of Fixmatch model on two Satelite datasets. We carried out additional experiments for evaluating the efficiency.

4.1 Datasets

4.1.1 EuroSAT

EuroSAT by Helber et al. [10] is a dataset of satellite images collected using Sentinel-2. In the Copernicus Earth observation program, it is open and fully available. This datasets covers 10 classes with in total 27,000 images. Has a resolution of 64×64 which are in RGB and also in 13-band MS format. All are labeled and geo-referenced images. The 10 classes are: Industrial Buildings, Residential Buildings, Annual Crop, Permanent Crop, River, Sea and Lake, Herbaceous Vegetation, Highway, Pasture and Forest These datasets are suitable for deep learning models.

4.1.2 UC Merced Land Use Dataset

The UCM dataset by [11] consists of 21 classes. Each class has about 100 land use images, which makes total images 2100 in the dataset having 256×256 pixels. This dataset is commonly used for scene classification.

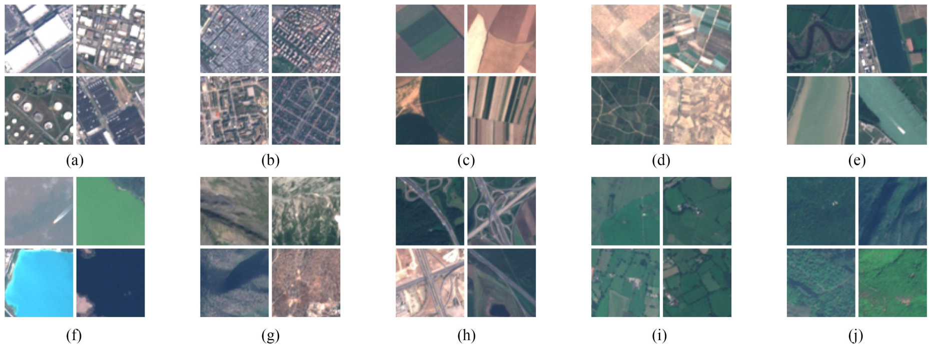

Figure 4.1: Sample images from 10 classes of EuroSAT [10]

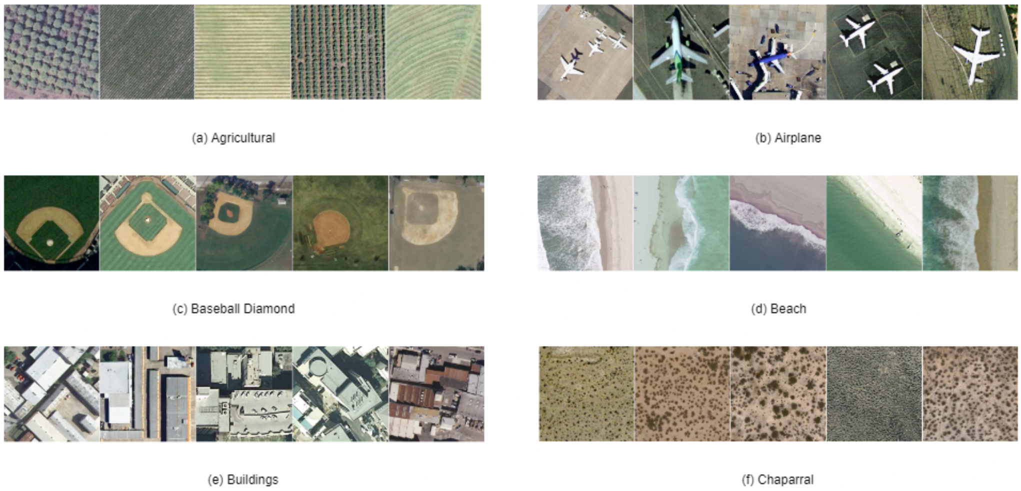

Figure 4.2: UCM contains 100 images from each of 21 land-use classes. 5 samples from some categories are shown above [11]

4.1.3 WHU-RS19

WHU-RS19 [12] [13] is a collection of high-resolution satellite photos up to 0.5 m that were extracted from Google Earth. In the image below, you can see several examples of the database. Airport, beach, bridge, commercial, desert, farmland, mountain, football field, industrial, meadow, forest, park, parking, pond, port, residential, railway station, river, and viaduct are among the 19 classifications of important scenes in high-resolution satellite images. There are around 50 to 61 samples for each class. It's worth noting that samples of images from the identical class are obtained from various places in satellite images of various resolutions, and varied sizes, orientations, and illuminations. There are near a total of 1,013 images in the WHU-RS19 dataset.

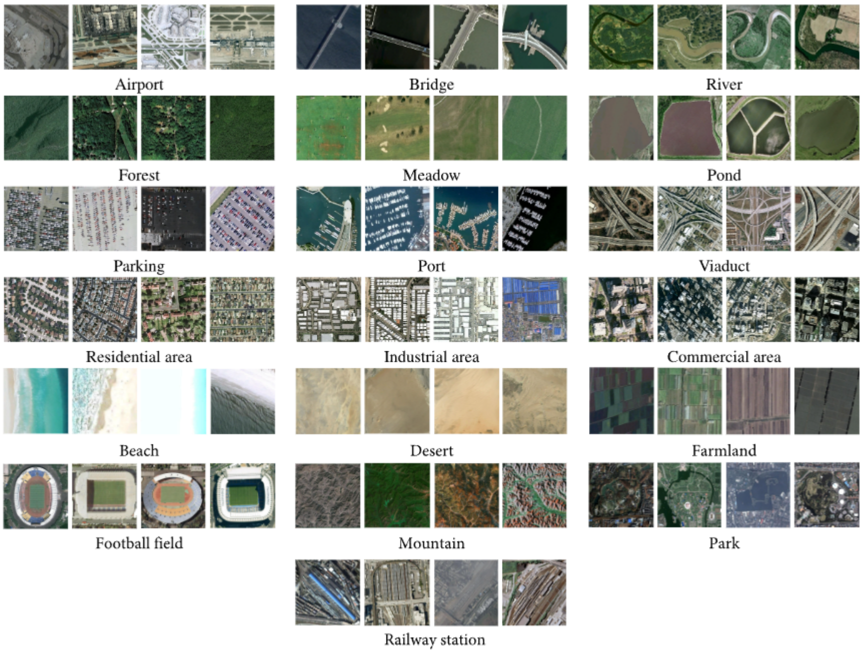

Figure 4.3: 5 samples from each of 19 categories of WHU-RS19 have been shown above [12] [13]

4.2 Performance Metrics

Precision and recall are performance indicators for pattern recognition and categorization in deep learning. These metrics produce more exact and accurate results. Some models need more precision, while others need more recall. So, it becomes difficult to understand the trade-off between precision and recall. Precision facilitates visualizing the model's reliability in classifying the model as positive. The recall evaluates how well a model can identify positive samples.

$$Precision = \frac{TruePositive}{TruePositive + FalsePositive} \tag{4.1}$$

$$Recall = \frac{TruePositive}{TruePositive + FalseNegative} \tag{4.2}$$

Precision-Recall is used as metric to evaluate the classifier to handle imbalanced data. When the proportion of true positives among positive predictions is analyzed, Precision-Recall can offer an accurate prediction of future classification performance.

4.3 Experimental Setup

PyTorch [43] is used to implement our proposed method. We experimented with Wide ResNet-28-2 [39] as feature extractor. Learning rate is calculated using this equation:

$$\mu_l = \left(\frac{N_{L+1-l}}{N_1} \right)^{\alpha} \tag{4.3}$$

.

We have used learning rate, $\mu = 0.03$, sampling rate tuner, $\alpha = 0.333$, weight decay $= 0.0005$ and imbalance ratio, $\gamma = 3$. The label fraction, β is used to measure the percentage of labeled data. For the optimizer, we have used Stochastic Gradient Descent (SGD)[44]. We have done all the experiments with only 10% labeled data. We have calculated loss with two loss functions.

The Supervised loss function:

$$\mathbb{L}_s = \frac{1}{N} \sum_{i}^{N} H(y_i, \check{y}_i) \tag{4.4}$$

The Unsupervised loss function:

$$\mathbb{L}_u = \frac{1}{N} \sum_{i}^{N} \delta_i H(\tilde{y}_i, \check{y}_i) \tag{4.5}$$

4.4 Experimental Analysis

4.4.1 Experiment with FixMatch (Tweaked Augmentation)

Fixmatch with Tweaked Augmentation

We improved upon FixMatch by Sohn et al. [5] by using our custom augmentation strategy. We experimented on three datasets-EuroSAT, UCM, and WHU-RS19 with FixMatch by [5]. At first, with Fixmatch by Sohn et al. [5] on EuroSat by Helber et al. [10] we were getting very poor results as Fixmatch's augmentation strategy wasn't for satellite images. With the augmentation strategy, the satellite images were losing too much information because of the heavy augmentations. As satellite images were already a bit hazy and most of them had fogs and other visibility issues, with heavy augmentation of Fixmatch, images were losing information. So, to solve this problem, we'll have to change Fixmatch's augmentation mechanism and parameters.

Analysis of Recall and Precision with Tweaked Fixmatch

Datasets	EuroSAT	UCM	WHU-RS19
Recall	0.888740168	0.8638202195238094	0.8337
Precision	0.911594744	0.8863465985714285	0.8768

Table 4.1: Recall and Precision with Tweaked Augmentation Strategy

After tweaking the augmentation parameters, we have calculated the recall and precision using the FixMatch (with tweaked augmentation) method.

Transformation	**RGB**	**Parameterrange**
Autocontrast		
Brightness	x	[0.1, 0.2]
Color	x	[0.05, 0.95]
Hue	x	0.1
Equalize		
Identity		
Posterize		
Shift	x	[0.1,0.2]
Rotate	x	[-30, 30]
Sharpness	x	[0.5, 1]
Shear x	x	[0.1, 0.2]
Shear y	x	[0.1, 0.2]
Solarize		
Translate x	x	[0, 1]
Translate y	x	[0, 1]

Table 4.2: Tweaked Augmentation strategy for Satellite Images

With our tweaked augmentation Table 4.2 strategy we achieved better results on the datasets. This table refers to the augmentations we used, the values give us optimal performance. In Table 4.1, we analyze the precision and recall of Fixmatch with a tweaked version of strong augmentation on the three datasets EuroSAT, UCM, and WHU-RS19.

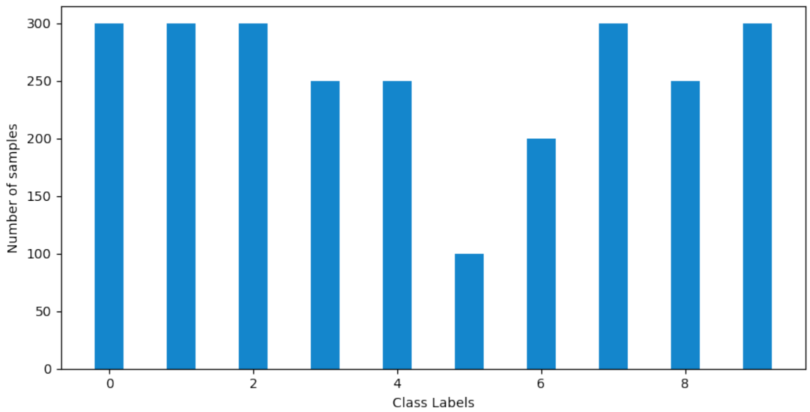

Figure 4.4: No. of samples in each class (EuroSAT)

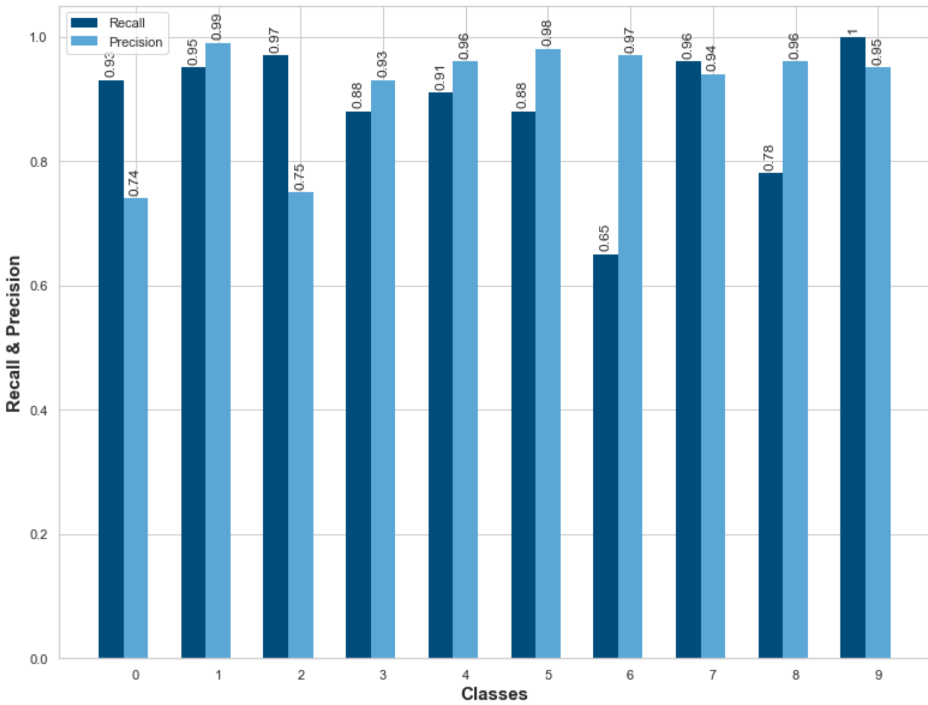

Figure 4.5: Recall and Precision (EuroSAT)

We see that the overall precision is higher in EuroSAT than in overall recall. This is because the EuroSAT dataset is imbalanced. Originally, UCM and WHU-RS19 are relatively balanced datasets. So we create an artificial imbalance on these two datasets and apply FixMatch with the tweaked augmentation on it. From the figures, we can see the recall and precision of each classes. As we can see from the table, the overall precision is higher than that of the recall of UCM and similarly, the overall precision of WHU-RS19 is higher than the overall recall. This is because these datasets are imbalanced. As a result, the model becomes biased and the generated pseudo-labels are also biased, therefore there is recall degradation.

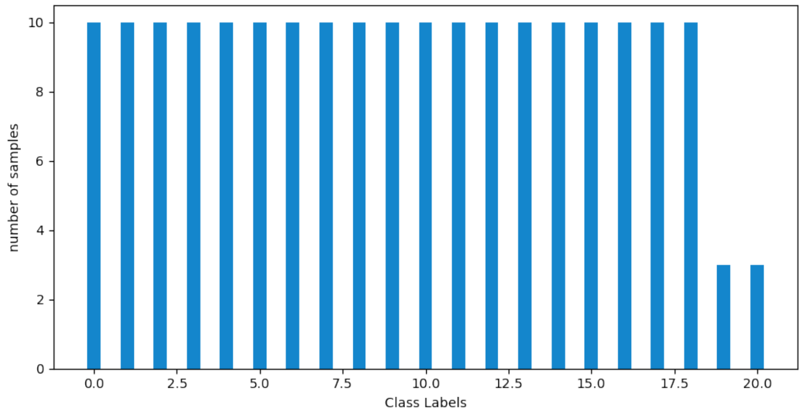

Figure 4.6: No. of samples in each class (UCM)

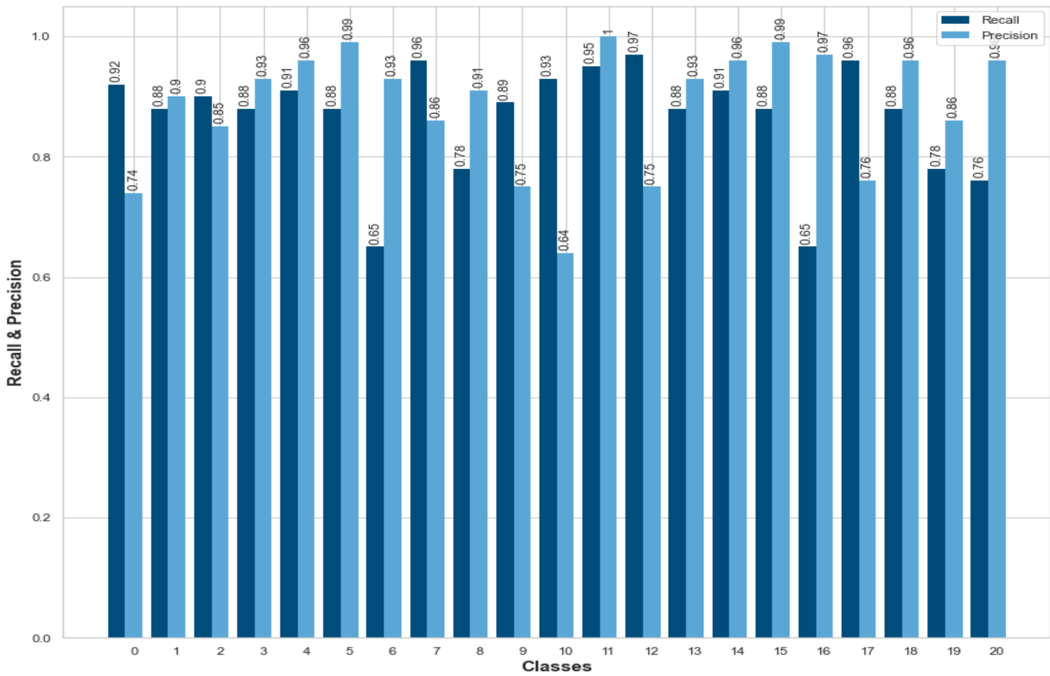

Figure 4.7: Recall and Precision (UCM)

Here, we can see in UCM dataset, class 19 and class 20 are relatively imbalanced. But their Precision is higher according to other classes.

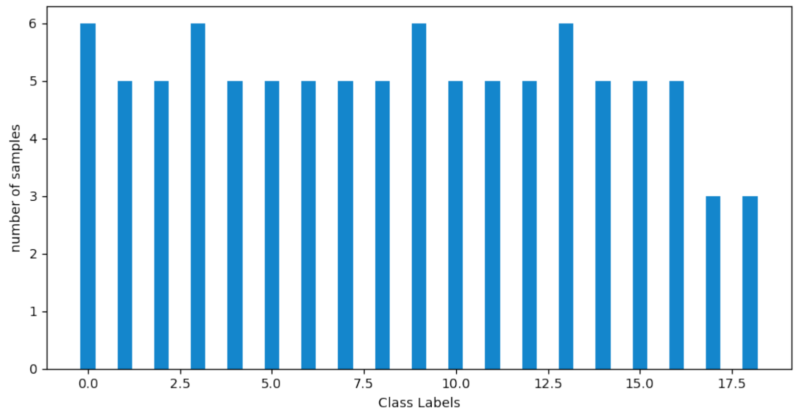

Figure 4.8: No. of samples in each class (WHU-RS19)

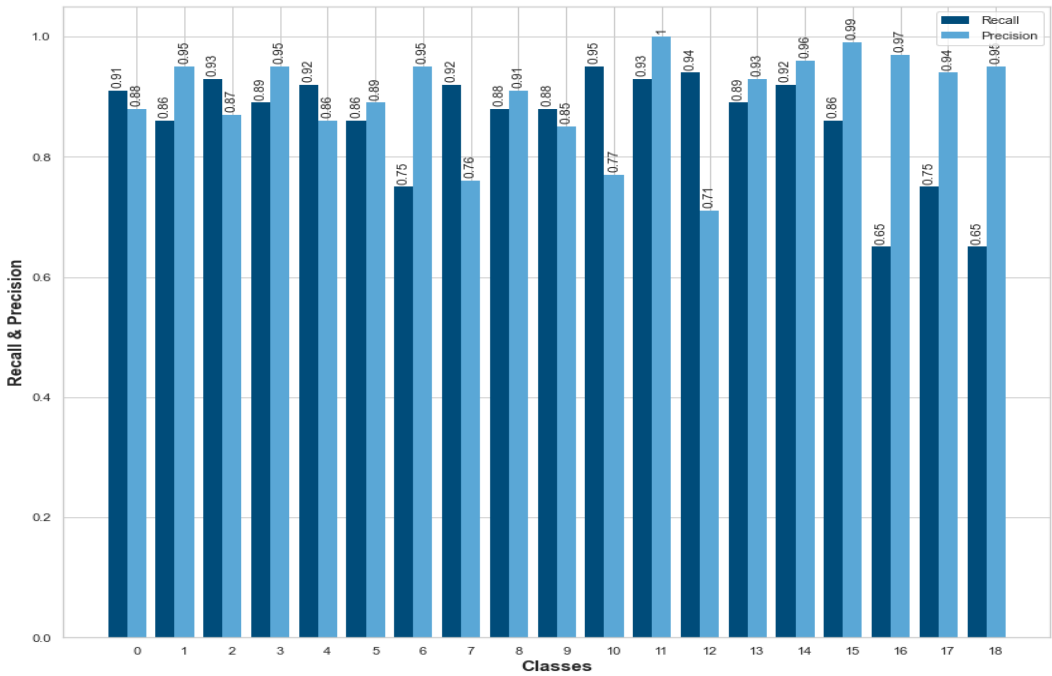

Figure 4.9: Recall and Precision (WHU-RS19)

Here, we can see in WHU-RS19 dataset, class 17 and class 18 are relatively imbalanced. Although recall is lower in those classes but Precision is higher according to other classes.

4.4.2 Experiment with Proposed Method

Precision and Recall with Proposed Method

Methods	Datasets					
	EuroSAT		UCM		WHU-RS19	
	Recall	Precision	Recall	Precision	Recall	Precision
FixMatch with Tweaked Augmentation	0.8887	0.9115	0.86382	0.8863	0.8337	0.8768
Proposed Method	0.9266	0.9211	0.8775	0.8654	0.8411	0.8592

Table 4.3: Comparison of Recall and Precision of Tweaked Augmentation Strategy and Proposed Method

In table 4.3, we see the overall precision and recall of our proposed method against the FixMatch with our tweaked augmentation. In the previous section, we discussed the precision and recall of FixMatch with Tweaked Augmentation applied on the three datasets EuroSAT, UCM and WHU-RS19. At that time, the datasets were imbalanced, and evaluating the precision and recall we could show the model bias occuring in the class-imbalanced EuroSAT, UCM and WHU-RS19 datasets. In this section, we apply our proposed method where we show that it handles model bias that is caused due to class-imbalance. Since EuroSAT is already imbalanced, we create an artificial imbalance in the other two datasets UCM and WHU-RS19. We create this imbalance using the hyperparameter called imbalance ratio (γ). We set $\gamma = 3$ for optimal performance of our proposed method. And then we apply our proposed technique on these artifically imbalanced datasets.

If there is model bias, the precision becomes high and recall becomes low. But our proposed method reduces model bias, this is shown in table 4.3 where we can see that recall and precision is very close. Like in EuroSAT, the precision is 0.9211 and recall is 0.9266. Similarly, in UCM the precision and recall is 0.8654 and 0.8775 and also for WHU-RS19 its 0.8592 and 0.8411 respectively. Precision and Recall being close to each other means there is less model bias, so the pseudo-labels generated can unbiased and high-quality. And from the table 4.3 we can see that our proposed method achieves it. And also both overall precision and recall becomes better for our proposed method than the previously applied FixMatch with tweaked augmentation. As a result, we can state that our suggested technique enhances precision and recall of classes in the datasets and thus help alleviate model bias.

4.4.3 Performance after Class-Rebalancing

Methods	Dataset									
	Balanced				Imbalanced					
	UCM		WHU-RS19		EuroSAT		UCM		WHU-RS19	
Fixmatch With Tweaked Augmentation	93.48		94.74		96.13		92.65		91.25	
Proposed Method	1st Gen	2nd Gen	1st Gen	2nd Gen	1st Gen	2nd Gen	1st Gen	2nd Gen	1st Gen	2nd Gen
	94.93	95.34	93.50	93.51	96.75	97.12	94.95	94.97	93.55	93.56

Table 4.4: Performance accuracy after Class-Rebalancing

In Table 4.4, we compare the accuracy of our proposed method with FixMatch (with our tweaked augmentation). We compare the accuracies in balanced and imbalanced datasets of the three benchmark datasets EuroSAT, UCM, and WHU-RS19 so that we can accurately evaluate whether our proposed method with class-rebalancing strategy is effective. On both balanced and unbalanced datasets, we see that our suggested technique beats FixMatch.

In the Imbalanced section from table 4.4, we apply artificial imbalance on all three datasets EuroSAT, UCM and WHU-RS19. Next, we compare the suggested method's accuracy to that of FixMatch (with tweaked augmentation), and we see that our proposed technique outperforms 1%, 2.32% and 2.31% on EuroSAT, UCM and WHU-RS19 respectively.

In the Balanced section of the table 4.4, we see our proposed method outperforms Fixmatch (with tweaked augmentation) by 1.86% in the original balanced UCM dataset. Our method also outperforms Fixmatch (with tweaked augmentation) on balanced WHU-RS19.

Our method performs better because of the class-rebalancing strategy as it incorporates high-quality pseudo-labels to the labeled set and redistributes it to the minority classes and thus balancing the classes and alleviating model bias helping achieve better accuracy.

4.4.4 Comparative Analysis with other Methods

Methods	Dataset									
	Balanced				Imbalanced					
	UCM		WHU-RS19		EuroSAT		UCM		WHU-RS19	
Supervised[10][19]	95.02		96.24		98.57		-		-	
MS Match[13]	94.13		-		96.04		-		-	
Fixmatch With Tweaked Augmentation	94.74		93.48		96.13		92.65		91.25	
Proposed Method	1st Gen	2nd Gen	1st Gen	2nd Gen	1st Gen	2nd Gen	1st Gen	2nd Gen	1st Gen	2nd Gen
	94.93	95.34	93.50	93.51	96.75	97.12	94.95	94.97	93.55	93.56

Table 4.5: Performance comparison between proposed method and other methods

In Table 4.5, we evaluate our proposed method with a fully-supervised method [9] [10], MSMatch[8] and FixMatch with our tweaked augmentation. Comparing the performance accuracies we see that our method performs (95.34%) as better as the fully-supervised method (95.02%) and our method outperforms MSMatch[8] and FixMatch with the tweaked augmentation by 1.21% and 0.6% on the UCM **balanced** dataset.

In case of imbalanced EuroSAT, our method outperforms MSMatch and FixMatch by 1.08% and 1% respectively. And it comes very close to the fully-supervised method as well.

Chapter 5

Conclusion

5.1 Summary

Land Use and Land Cover (LULC) detection is an important research area in the remote sensing community. As there are abundant satellite data available everywhere, and with time satellite images are being collected more and more every day. It's difficult to categorize it utilizing deep learning neural networks since it relies on labeled data. And there is a lack of annotated satellite images, the datasets are class-imbalanced. Our method uses a semi-supervised learning technique with a custom augmentation and a class-rebalancing distribution alignment strategy. We tweaked the augmentation strategy inspired by the FixMatch paper [5], this tweaked augmentation strategy helps improve the accuracy of the satellite image datasets as we can see from Section 4. We also added a class-rebalancing strategy that aligns the class distribution basically adding more samples to the classes with fewer samples, and thus balancing the dataset classes. This helps alleviate model bias. We demonstrated that our proposed method is superior to earlier works. Our proposed technique has better accuracy when we compare it with the supervised methods, FixMatch[5] with tweaked augmentation, MSMatch[8].

Basically, our proposed method solves the issue of manually labeling data and model bias caused by data imbalance.

5.2 Future Work

In our work so far, we have applied a popular semi-supervised learning technique to standard benchmark datasets of satellite imagery and we applied class-rebalancing distribution alignment to handle the class imbalance in the datasets. All of these datasets have few labeled images. So one of the limitations of our proposed method is that if we have datasets that have no labeled images i.e all images are unlabeled then we wouldn't be able to work with that dataset as our method requires a limited amount of labeled data. Another limitations of our work is that our method needs a huge amount of unlabeled data. So datasets having low amount of unlabeled data won't yield a good result with our proposed technique. Though we have worked with high-resolution datasets, we hope to work on higher-resolution datasets in the future. We also want to work with large-scale datasets like Million-AID [45], BigEarthNet [46]. We want to explore more how different augmentation strategies would affect the model. We also wish to incorporate our proposed method on LULC change analysis over different periods of time in a specific area. For now, we have worked with RGB satellite images, we want to work with multispectral(MS) images that contain more bands and information to see how our model can scale up to it. We hope to investigate such possibilities in the future.

References

[1] Gui-Song Xia, Jingwen Hu, Fan Hu, Baoguang Shi, Xiang Bai, Yanfei Zhong, Liangpei Zhang, and Xiaoqiang Lu. Aid: A benchmark data set for performance evaluation of aerial scene classification. *IEEE Transactions on Geoscience and Remote Sensing*, 55(7):3965–3981, 2017.

[2] Md Rahman, Ram Avtar, Ali P Yunus, Jie Dou, Prakhar Misra, Wataru Takeuchi, Netrananda Sahu, Pankaj Kumar, Brian Alan Johnson, Rajarshi Dasgupta, et al. Monitoring effect of spatial growth on land surface temperature in dhaka. *Remote Sensing*, 12(7):1191, 2020.

[3] David Berthelot, Nicholas Carlini, Ian Goodfellow, Nicolas Papernot, Avital Oliver, and Colin Raffel. Mixmatch: A holistic approach to semi-supervised learning. *arXiv preprint arXiv:1905.02249*, 2019.

[4] David Berthelot, Nicholas Carlini, Ekin D Cubuk, Alex Kurakin, Kihyuk Sohn, Han Zhang, and Colin Raffel. Remixmatch: Semi-supervised learning with distribution alignment and augmentation anchoring. *arXiv preprint arXiv:1911.09785*, 2019.

[5] Kihyuk Sohn, David Berthelot, Chun-Liang Li, Zizhao Zhang, Nicholas Carlini, Ekin D Cubuk, Alex Kurakin, Han Zhang, and Colin Raffel. Fixmatch: Simplifying semi-supervised learning with consistency and confidence. *arXiv preprint arXiv:2001.07685*, 2020.

[6] Tjeng Wawan Cenggoro, Sani M Isa, Gede Putra Kusuma, and Bens Pardamean. Classification of imbalanced land-use/land-cover data using variational semi-supervised learning. In *2017 International Conference on Innovative and Creative Information Technology (ICITech)*, pages 1–6. IEEE, 2017.

[7] Runyu Fan, Ruyi Feng, Lizhe Wang, Jining Yan, and Xiaohan Zhang. Semi-mcnn: A semisupervised multi-cnn ensemble learning method for urban land cover classification using submeter hrrs images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13:4973–4987, 2020.

[8] Pablo Gómez and Gabriele Meoni. Msmatch: Semi-supervised multispectral scene classification with few labels. *arXiv preprint arXiv:2103.10368*, 2021.

[9] Chen Wei, Kihyuk Sohn, Clayton Mellina, Alan Yuille, and Fan Yang. Crest: A class-rebalancing self-training framework for imbalanced semi-supervised learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10857–10866, 2021.

[10] Patrick Helber, Benjamin Bischke, Andreas Dengel, and Damian Borth. Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(7):2217–2226, 2019.

[11] Yi Yang and Shawn Newsam. Bag-of-visual-words and spatial extensions for land-use classification. In *Proceedings of the 18th SIGSPATIAL international conference on advances in geographic information systems*, pages 270–279, 2010.

[12] Dengxin Dai and Wen Yang. Satellite image classification via two-layer sparse coding with biased image representation. *IEEE Geoscience and Remote Sensing Letters*, 8(1):173–176, 2010.

[13] Gui-Song Xia, Wen Yang, Julie Delon, Yann Gousseau, Hong Sun, and Henri Maître. Structural high-resolution satellite image indexing. In *ISPRS TC VII Symposium-100 Years ISPRS*, volume 38, pages 298–303, 2010.

[14] Timo Wekerle, José Bezerra Pessoa, Luís Eduardo Vergueiro Loures da Costa, and Luís Gonzaga Trabasso. Status and trends of smallsats and their launch vehicles—an up-to-date review. *Journal of Aerospace Technology and Management*, 9:269–286, 2017.

[15] Lei Ma, Yu Liu, Xueliang Zhang, Yuanxin Ye, Gaofei Yin, and Brian Alan Johnson. Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS journal of photogrammetry and remote sensing*, 152:166–177, 2019.

[16] Sen Jia, Shuguo Jiang, Zhijie Lin, Nanying Li, Meng Xu, and Shiqi Yu. A survey: Deep learning for hyperspectral image classification with few labeled samples. *Neurocomputing*, 448:179–204, 2021.

[17] Gong Cheng, Xingxing Xie, Junwei Han, Lei Guo, and Gui-Song Xia. Remote sensing image scene classification meets deep learning: Challenges, methods, benchmarks, and opportunities. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13:3735–3756, 2020.

[18] Bing Liu, Xuchu Yu, Pengqiang Zhang, Xiong Tan, Anzhu Yu, and Zhixiang Xue. A semi-supervised convolutional neural network for hyperspectral image classification. *Remote Sensing Letters*, 8(9):839–848, 2017.

[19] Hao Wu and Saurabh Prasad. Semi-supervised deep learning using pseudo labels for hyperspectral image classification. *IEEE Transactions on Image Processing*, 27(3):1259–1270, 2017.

[20] Mateusz Buda, Atsuto Maki, and Maciej A Mazurowski. A systematic study of the class imbalance problem in convolutional neural networks. *Neural Networks*, 106:249–259, 2018.

[21] Qin Wang, Wen Li, and Luc Van Gool. Semi-supervised learning by augmented distribution alignment. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1466–1475, 2019.

[22] Christoph Mayer, Matthieu Paul, and Radu Timofte. Adversarial feature distribution alignment for semi-supervised learning. *Computer Vision and Image Understanding*, 202:103109, 2021.

[23] Jaehyung Kim, Youngbum Hur, Sejun Park, Eunho Yang, Sung Ju Hwang, and Jinwoo Shin. Distribution aligning refinery of pseudo-label for imbalanced semi-supervised learning. *Advances in Neural Information Processing Systems*, 33: 14567–14579, 2020.

[24] Yanbei Chen, Xiatian Zhu, Wei Li, and Shaogang Gong. Semi-supervised learning under class distribution mismatch. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 3569–3576, 2020.

[25] Ruifei He, Jihan Yang, and Xiaojuan Qi. Re-distributing biased pseudo labels for semi-supervised semantic segmentation: A baseline investigation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6930–6940, 2021.

[26] Patrick Helber, Benjamin Bischke, Andreas Dengel, and Damian Borth. Introducing eurosat: A novel dataset and deep learning benchmark for land use and land cover classification. In *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*, pages 204–207. IEEE, 2018.

[27] Dong-Hyun Lee et al. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In *Workshop on challenges in representation learning, ICML*, volume 3, page 896, 2013.

[28] Jisoo Jeong, Seungeui Lee, Jeesoo Kim, and Nojun Kwak. Consistency-based semi-supervised learning for object detection. *Advances in neural information processing systems*, 32:10759–10768, 2019.

[29] Hongyi Zhang, Moustapha Cisse, Yann N Dauphin, and David Lopez-Paz. mixup: Beyond empirical risk minimization. *arXiv preprint arXiv:1710.09412*, 2017.

[30] Ekin D. Cubuk, Barret Zoph, Jonathon Shlens, and Quoc V. Le. Randaugment: Practical automated data augmentation with a reduced search space, 2019.

[31] Esam Othman, Yakoub Bazi, Naif Alajlan, Haikel Alhichri, and Farid Melgani. Using convolutional features and a sparse autoencoder for land-use scene classification.

[32] Wenxiu Teng, Ni Wang, Huihui Shi, Yuchan Liu, and Jing Wang. Classifier-constrained deep adversarial domain adaptation for cross-domain semisupervised classification in remote sensing images. *IEEE Geoscience and Remote Sensing Letters*, 17(5):789–793, 2019.

[33] Dongao Ma, Ping Tang, and Lijun Zhao. Siftinggan: Generating and sifting labeled samples to improve the remote sensing image scene classification baseline in vitro. *IEEE Geoscience and Remote Sensing Letters*, 16(7):1046–1050, 2019.

[34] Gong Cheng, Ceyuan Yang, Xiwen Yao, Lei Guo, and Junwei Han. When deep learning meets metric learning: Remote sensing image scene classification via learning discriminative cnns. *IEEE transactions on geoscience and remote sensing*, 56(5):2811–2821, 2018.

[35] Chia-Wen Kuo, Chih-Yao Ma, Jia-Bin Huang, and Zsolt Kira. Featmatch: Feature-based augmentation for semi-supervised learning. In *European Conference on Computer Vision*, pages 479–495. Springer, 2020.

[36] Youngtaek Oh, Dong-Jin Kim, and In So Kweon. Distribution-aware semantics-oriented pseudo-label for imbalanced semi-supervised learning. *arXiv preprint arXiv:2106.05682*, 2021.

[37] Yin Cui, Menglin Jia, Tsung-Yi Lin, Yang Song, and Serge Belongie. Class-balanced loss based on effective number of samples. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9268–9277, 2019.

[38] Kaidi Cao, Colin Wei, Adrien Gaidon, Nikos Arechiga, and Tengyu Ma. Learning imbalanced datasets with label-distribution-aware margin loss. *Advances in neural information processing systems*, 32, 2019.

[39] Sergey Zagoruyko and Nikos Komodakis. Wide residual networks. *arXiv preprint arXiv:1605.07146*, 2016.

[40] Avital Oliver, Augustus Odena, Colin A Raffel, Ekin Dogus Cubuk, and Ian Goodfellow. Realistic evaluation of deep semi-supervised learning algorithms. *Advances in neural information processing systems*, 31, 2018.

[41] Minyoung Huh, Pulkit Agrawal, and Alexei A Efros. What makes imagenet good for transfer learning? *arXiv preprint arXiv:1608.08614*, 2016.

[42] Zhirong Wu, Alexei A Efros, and Stella X Yu. Improving generalization via scalable neighborhood component analysis. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 685–701, 2018.

[43] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in pytorch. 2017.

[44] Ilya Sutskever, James Martens, George Dahl, and Geoffrey Hinton. On the importance of initialization and momentum in deep learning. In *International conference on machine learning*, pages 1139–1147. PMLR, 2013.

[45] Yang Long, Gui-Song Xia, Shengyang Li, Wen Yang, Michael Ying Yang, Xiao Xiang Zhu, Liangpei Zhang, and Deren Li. On creating benchmark dataset for aerial image interpretation: Reviews, guidances, and million-aid. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14: 4205–4230, 2021.

[46] Gencer Sumbul, Jian Kang, Tristan Kreuziger, Filipe Marcelino, Hugo Costa, Pedro Benevides, Mario Caetano, and Begüm Demir. Bigearthnet dataset with a new class-nomenclature for remote sensing image understanding. *arXiv preprint arXiv:2001.06372*, 2020.