remote sensing

MDPI

Article

Semi-FCMNet: Semi-Supervised Learning for Forest Cover Mapping from Satellite Imagery via Ensemble Self-Training and Perturbation

Beiqi Chen [1,†], Liangjing Wang [1,†], Xijian Fan [1,*], Weihao Bo [1], Xubing Yang [1] and Tardi Tjahjadi [2]

[1] College of Information Science and Technology, Nanjing Forestry University, Nanjing 210037, China; momura@njfu.edu.cn (B.C.)
[2] School of Engineering, University of Warwick, Coventry CV4 7AL, UK
* Correspondence: xijian.fan@njfu.edu.cn
† These authors contributed equally to this work.

Abstract: Forest cover mapping is of paramount importance for environmental monitoring, biodiversity assessment, and forest resource management. In the realm of forest cover mapping, significant advancements have been made by leveraging fully supervised semantic segmentation models. However, the process of acquiring a substantial quantity of pixel-level labelled data is prone to time-consuming and labour-intensive procedures. To address this issue, this paper proposes a novel semi-supervised-learning-based semantic segmentation framework that leverages limited labelled and numerous unlabelled data, integrating multi-level perturbations and model ensembles. Our framework incorporates a multi-level perturbation module that integrates input-level, feature-level, and model-level perturbations. This module aids in effectively emphasising salient features from remote sensing (RS) images during different training stages and facilitates the stability of model learning, thereby effectively preventing overfitting. We also propose an ensemble-voting-based label generation strategy that enhances the reliability of model-generated labels, achieving smooth label predictions for challenging boundary regions. Additionally, we designed an adaptive loss function that dynamically adjusts the focus on poorly learned categories and dynamically adapts the attention towards labels generated during both the student and teacher stages. The proposed framework was comprehensively evaluated using two satellite RS datasets, showcasing its competitive performance in semi-supervised forest-cover-mapping scenarios. Notably, the method outperforms the fully supervised approach by 1–3% across diverse partitions, as quantified by metrics including mIoU, accuracy, and mPrecision. Furthermore, it exhibits superiority over other state-of-the-art semi-supervised methods. These results indicate the practical significance of our solution in various domains, including environmental monitoring, forest management, and conservation decision-making processes.

Keywords: semi-supervision; forest cover mapping; semi-supervised semantic segmentation; self-training

Citation: Chen, B.; Wang, L.; Fan, X.; Bo, W.; Yang, X.; Tjahjadi, T. Semi-FCMNet: Semi-Supervised Learning for Forest Cover Mapping from Satellite Imagery via Ensemble Self-Training and Perturbation. *Remote Sens.* **2023**, *15*, 4012. https://doi.org/10.3390/rs15164012

Academic Editors: Renjing Xu, Donghao Zhang and Jianzhe Lin

Received: 5 July 2023
Revised: 5 August 2023
Accepted: 9 August 2023
Published: 13 August 2023

1. Introduction

Forests play a vital role in the land ecosystem of the Earth. They are indispensable for conserving biodiversity, protecting watersheds, capturing carbon, mitigating climate change effects [1,2], maintaining ecological balance, regulating rainfall patterns, and ensuring the stability of large-scale climate systems [3,4]. As a result, the timely and precise monitoring and mapping of forest cover has emerged as a vital aspect of sustainable forest management and the monitoring of ecosystem transformations [5].

Traditionally, the monitoring and mapping of forest cover has primarily relied on field research and photo-interpretation techniques. However, these methods are limited by the extensive manpower required. With the advancements in remote sensing (RS) technology, the acquisition of large-scale, high-resolution forest imagery data has become possible without the need for physical contact and without causing harm to the forest environment.

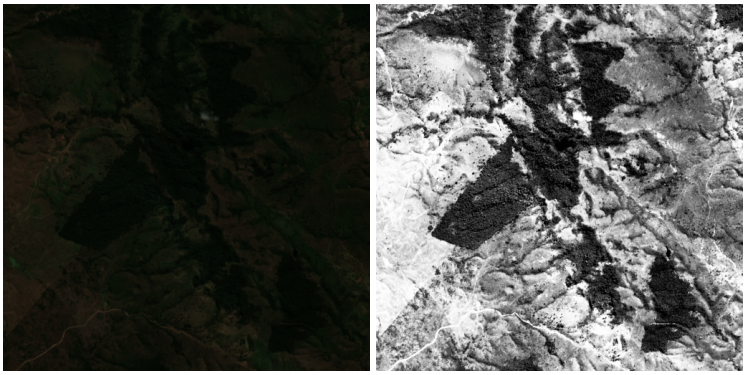
Taking advantage of RS imagery, numerous studies have proposed various methods for forest cover mapping, including decision trees [6], regression trees [7], maximum likelihood classifiers [8], random forest classification algorithms [9,10], support vector machines, spatio-temporal Markov random-field super-resolution mapping [11], and multi-scale spectral–spatial–temporal super-resolution mapping [12].

Recently, due to the growing prevalence of deep Convolutional Neural Networks (CNNs) [13] and semantic segmentation [14–16], there has been a notable shift in the research community towards utilising these techniques for forest cover mapping with RS imagery. CNNs have emerged as powerful tools for analysing two-dimensional images, employing their multi-layered convolution operations to effectively capture low-level spatial patterns (such as edges, textures, and shapes) and extract high-level semantic information. Meanwhile, semantic segmentation techniques enable the precise identification and extraction of different objects/regions in an image by classifying each image pixel into a specific semantic category, achieving pixel-level image segmentation. In the context of forest cover mapping, several existing methods have demonstrated the effectiveness of using semantic segmentation techniques. Bragagnolo et al. [17] proposed to integrate an attention block into the basic UNet network to segment the forest area using satellite imagery from South America. Flood et al. [18] also proposed a UNet-based network and achieved promising results in mapping the presence or absence of trees and shrubs in Queensland, Australia. Isaienkov et al. [19] directly employed the baseline U-Net model combined with Sentinel-2 satellite data to detect changes in Ukrainian forests. However, all these methods rely on fully supervised learning for semantic segmentation, which necessitates a substantial amount of labelled pixel data, resulting in a significant labelling expense. Semi-supervised learning [20–22] has emerged as a promising approach to address the aforementioned challenges. It involves training models using a combination of limited labelled data and a substantial amount of unlabelled data, which reduces the need for manual annotations while still improving the performance of the model. Several research studies [23,24] have explored the application of semi-supervised learning in semantic segmentation for land-cover-mapping tasks in RS. While these studies have assessed the segmentation of forests to some extent, their focus has predominantly been on forests situated in urban or semi-natural areas, limiting their performance in densely forested natural areas. Moreover, there are unique challenges associated with utilising satellite RS imagery specifically for forest cover mapping:

Challenge 1: As illustrated in Figure 1a, satellite remote sensing (RS) forest images often face problems such as variations in scene illumination and atmospheric interference during the image acquisition process. These factors can lead to colour deviations and distortions, resulting in poor colour fidelity and low contrast. Therefore, it becomes essential to employ image enhancement techniques to improve visualisation and reveal more details. This enhancement facilitates the ability of CNNs to effectively capture spatial patterns.

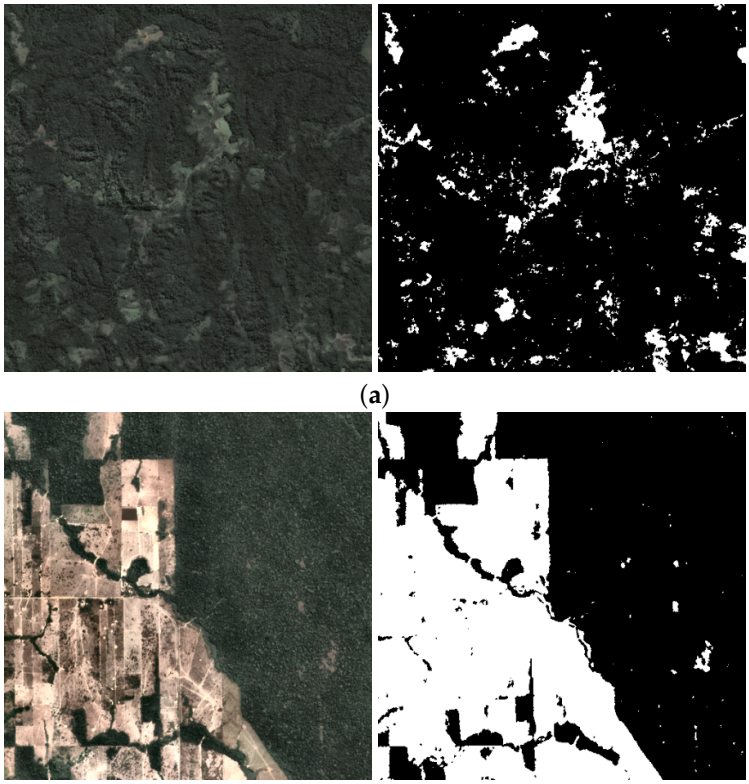
Challenge 2: Due to the high density of natural forest cover and the similar reflectance characteristics between forest targets and other non-forest targets, e.g., grass and the shadow of vegetation, the boundaries between forest and non-forest areas often become unclear. As a result, it becomes challenging to accurately distinguish and delineate the details and edges of the regions of interest, as depicted in Figure 2a.

Challenge 3: For unknown forest distributions, there are two scenarios: imbalanced (as illustrated in Figure 2) or balanced datasets. Current methods face challenges in effectively handling datasets with different distributions, resulting in poor model generalisation.



(**a**) (**b**)

Figure 1. Visualisation of images after application of different processing methods: (**a**) original image; and (**b**) image enhanced using contrast enhancement.



(**a**)

(**b**)

Figure 2. Visualisation of balanced and imbalanced samples (where black represents non-forest, and white represents forest): (**a**) balanced sample and (**b**) imbalanced sample.

In this study, we have undertaken a pioneering endeavour to integrate semi-supervised learning techniques into forest cover mapping using satellite RS imagery. We propose a novel semi-supervised semantic segmentation network called Semi-FCMNet, designed to effectively tackle the associated challenges in this task. To tackle challenge 1, which encompasses image distortion issues in RS images, we employ the concept of multi-level perturbations. Perturbations are designed at different stages, namely, at the input level, feature level, and model level. At the input level, perturbations utilising mixed augmentation techniques are employed to enhance the representation of forest features within the image. The feature-level and model-level perturbations facilitate model learning and capture valuable information during training. Importantly, these perturbations effectively counteract the overfitting of the model to noise. For challenge 2, we combine the auxiliary teacher module with the Test-Time Augmentation (TTA) approach to integrate the gener-

ated pseudo-labels through voting and multi-scale fusion, enhancing the reliability and clarity of edge information. To address challenge 3, we introduce an adaptive loss function that automatically focuses on under-learned classes and adjusts the attention towards labels generated by both the student and teacher models. This approach enables our model to achieve excellent performance on both balanced and imbalanced datasets. The adaptive loss function effectively addresses the issue of insufficient learning in certain classes and improves the capability of the model to handle diverse data distributions. Furthermore, we adopt a progressive learning approach and design a data augmentation strategy from easy to difficult, employing different intensity levels of data augmentation for models at different stages. The code is publicly available at https://github.com/baizegugugu/Semi-FCMNet

The primary contributions of this paper are summarised as follows:

1. We have designed the multi-level perturbation (MP) module, including input-, feature- and model-level perturbations at different module stages. The proposed approach incorporates perturbations at the input stage to enhance forest representation features by using mixed augmentation. Additionally, the auxiliary teacher module introduces perturbations at both the feature and model levels, allowing the model to concentrate on feature disparities and proficiently learn forest characteristics while effectively mitigating the overfitting problem to noise.
2. By integrating auxiliary teachers with the student model, the basic self-training method was enhanced. To generate more stable and reliable pseudo-labels during the pseudo-labelling phase, we introduced a novel ensemble voting (EV) module, smoothing the decision-making process for challenging boundary regions. This module leverages a combination of multiple models and adopts a strategy based on TTA and multi-model voting.
3. We have developed a simple yet effective adaptive loss (AL) that enables the model to adapt to both balanced and imbalanced data distributions while also increasing its focus on labels generated by the teacher. By incorporating AL into the training process, our model demonstrates robust performance across different data scenarios.

This paper is organised as follows. Section 2 reviews semi-supervised semantic segmentation, and Section 3 provides a detailed description of the proposed framework. Section 4.1 presents detailed information on the data distribution of the two datasets we used (Atlantic Forest and Amazon Forest), while Sections 4.2 and 4.3 introduce the relevant parameters set in our experiments and the metrics used to verify the experimental results, respectively. Section 4.4 compares and analyses our methods with the SOTA methods on the two datasets, while Section 4.5 presents the ablation experimental results for our method to validate its effectiveness. Finally, Section 5 outlines the limitations of our method, future research directions, and application prospects.

2. Related Work

2.1. Semi-Supervised Semantic Segmentation

Consistency regularisation and pseudo-labelling are two main categories of methods in the field of semi-supervised semantic segmentation [21,22].

Consistency regularisation methods aim to improve model performance by promoting consistency among diverse predictions for the same image. This is achieved by introducing perturbations to either the input images or the models themselves. For instance, the CCT approach [25] utilises an auxiliary decoder structure that incorporates multiple robust perturbations at both the feature level and decoder output stage. These perturbations are strategically employed to enforce consistency in the model predictions, ensuring that the predictions remain consistent even in the presence of perturbations. Furthermore, Liu et al. proposed PS-MT [26], which introduces innovative extensions to improve upon the mean-teacher (MT) model. These extensions include the introduction of an auxiliary teacher and the replacement of the MT's mean square error (MSE) loss with a more stringent confidence-weighted cross-entropy (Conf-CE) loss. These enhancements greatly enhance the accuracy and consistency of the predictions in the MT model. Building upon these

advancements, Abulikemu Abuduweili et al. [27] proposed a novel method that leverages adaptive consistency regularisation to effectively combine pre-trained models and unlabelled data, improving model performance. However, consistency regularisation typically relies on perturbation techniques that impose certain requirements on dataset quality and distribution, as well as numerous challenging hyperparameters to fine-tune.

Contrary to the consistency regularisation methods, pseudo-labelling techniques, exemplified by self-training [28], leverage predictions from unlabelled data to generate pseudo-labels, which are then incorporated into the training process, effectively expanding the training set and enriching the model with more information. In this context, Yi Zhu et al. [29] proposed a self-training framework for semantic segmentation that utilises pseudo-labels from unlabelled data, addressing data imbalance through centroid sampling and focusing on optimising computational efficiency. However, pseudo-labelling methods offer more stable training but have limitations in achieving substantial improvements in model performance, leading to under-utilisation of the potential of unlabelled data.

In the task of forest cover mapping using RS imagery, we propose an integrated framework that combines self-training and consistency regularisation methods. This approach effectively enhances the performance of the model by effectively utilising the abundant unlabelled data available in the context of forest RS.

2.2. Semi-Supervised Semantic Segmentation in RS

Several studies have investigated the application of semi-supervised learning in RS. Lucas et al. [23] proposed Sourcerer, a deep-learning-based technique for semi-supervised domain adaptation in land cover mapping from satellite image time-series data. Sourcerer surpasses existing methods by effectively leveraging labelled data from a source domain and adapting the model to the target domain using a novel regulariser, even with limited labelled target data. Chen et al. [30] introduced SemiRoadExNet, a novel semi-supervised road extraction network based on a Generative Adversarial Network (GAN). The network efficiently utilises both labelled and unlabelled data by generating road segmentation results and entropy maps. Zou et al. [31] introduced a novel pseudo-labelling approach for semantic segmentation, improving the training process using unlabelled or weakly labelled data. Through the intelligent combination of diverse sources and robust data augmentation, the proposed strategy demonstrates effective consistency training, showing its effectiveness for data of low or high density. On the other hand, Zhang et al. [32] proposed a semi-supervised deep learning framework for the semantic segmentation of high-resolution RS images. The framework utilises transformation consistency regularisation to make the most of limited labelled samples and abundant unlabelled data. However, the application of semi-supervised semantic segmentation methods in forest cover mapping has not been explored. In this paper, we aim to address the challenges of forest cover mapping in high-density RS imagery and investigate the application of semi-supervised semantic segmentation methods in this context.

3. Method

3.1. Problem Definition

Semi-supervised semantic segmentation is a method of semantic segmentation that uses labelled and unlabelled data. Compared to traditional fully supervised semantic segmentation methods, it does not require a large amount of labelled data to train the model, making computation more economical and efficient. The main idea of semi-supervised semantic segmentation is to enhance the generalisation ability of the model by utilising unlabelled data. Specifically, semi-supervised semantic segmentation seeks to generalise from a combined dataset consisting of pixel-wise labelled images $D^l = \{(x_i, y_i)\}_{i=1}^{M}$ unlabelled images $D^u = \{u_i\}_{i=1}^{N}$, where, typically, $N >> M$. In the majority of studies, the overall optimisation objective is formulated as

$$L = L^s + \lambda L^u, \tag{1}$$

where λ serves as a tradeoff between labelled and unlabelled data. The parameter λ can either be a fixed value or be scheduled during training. The unsupervised loss L^u is a crucial aspect that distinguishes various semi-supervised methods, whereas the supervised loss L^s typically refers to the cross-entropy loss between predictions and manually annotated masks.

3.2. Auxiliary Mean Teachers and Student Models

Although many more advanced models and methods have emerged, classical methods such as self-training and mean teachers can also perform well with improvements. To further improve the performance of the model based on the self-training method, we incorporate an improved mean-teacher mechanism. Figure 3 shows the architecture of our model. The model uses the classical encoder–decoder model, using ResNet-101 as the encoder for better pixel information extraction and restoration and using DeeplabV3+ as the decoder. All auxiliary teachers and students share the same structure, and both of the two auxiliary teachers receive exponential moving average (EMA) transfers of parameters from different epochs of the student, as shown in Figure 4, i.e.,

$$\theta_k = \gamma \cdot \theta_k + (1 - \gamma) \cdot \theta_s, \tag{2}$$

where $k \in t_{a1}, t_{a2}, \gamma \in (0, 1)$. For the training of teacher models, we update the parameters of only one of the two teachers at each training epoch.

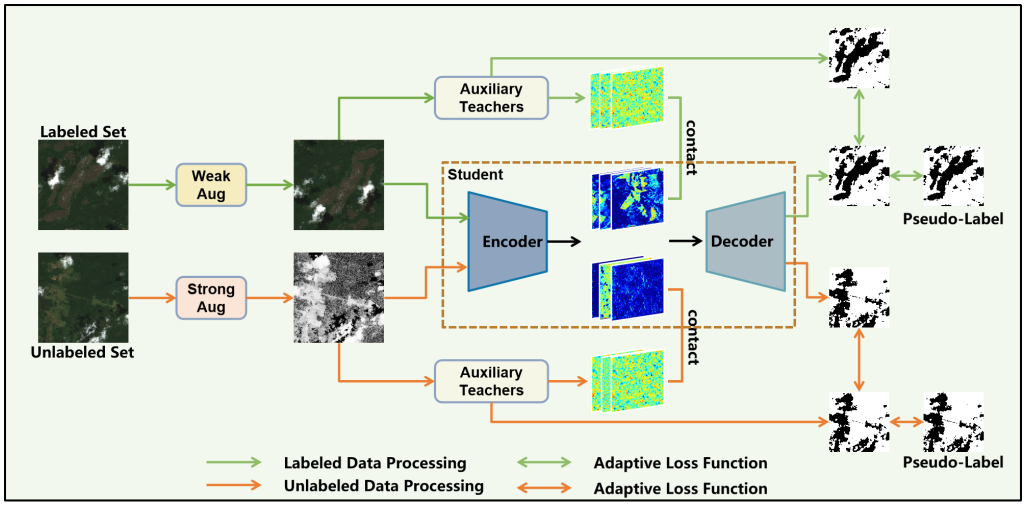


Figure 3. Overview of the proposed network pipeline.

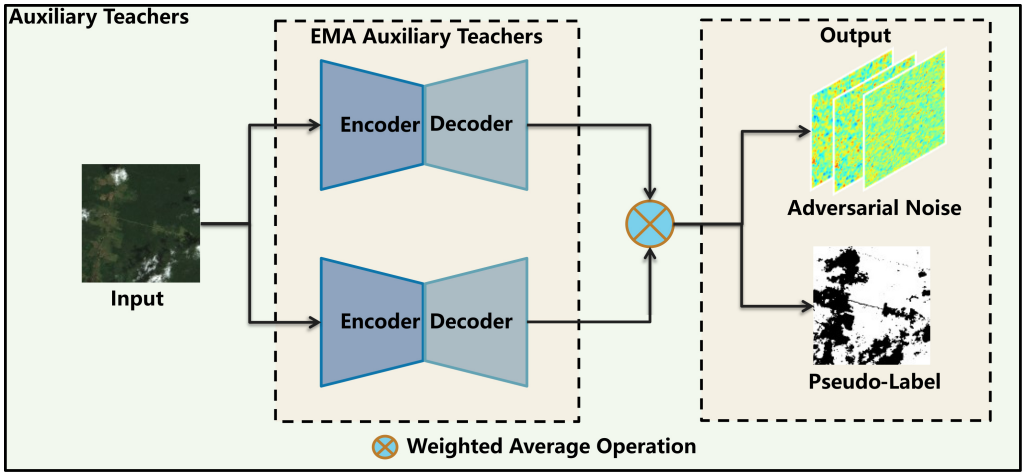


Figure 4. Overview of the proposed auxiliary teacher module.

3.3. Training Strategy: Self-Training (ST)

Currently, the perturbation method based on consistency regularisation is widely considered the most effective approach to semi-supervised learning. This method involves perturbing the image during data augmentation and improving the performance of the model by constraining the similarity between its final predicted outputs. These perturbations mostly focus on perturbing the image representation, such as adding noise, random dropout, cutout, and CutMix. However, the unprocessed RGB colour features of the forest satellite RS dataset are not obvious, and the data distribution is imbalanced. Random perturbations at the image level make it difficult for the classes with few samples to be fully learned, while the classes with many samples are constrained by the loss function, making it difficult for the model to fully learn from those samples, resulting in poor model performance. At the same time, the perturbation method based on consistency regularisation requires manually adjusting the weights of unsupervised loss and supervised loss, and it is difficult to adjust the proportion for different weights, leading to a further decline in model performance.

Therefore, in order to fully utilise the information in the dataset and reduce the settings of hyperparameters, we primarily adopt a method based on self-training, as shown in Figure 5. Firstly, we train the teacher model t^l on the labelled dataset D^l, and then t^l is used to assign pseudo-labels to the remaining unlabelled samples in D^u. Finally, the pseudo-labels are used as the labels for the unlabelled images, and the student model s is trained on the entire dataset D^a. During the training process, auxiliary teachers are introduced to more stably evaluate and predict unlabelled images.

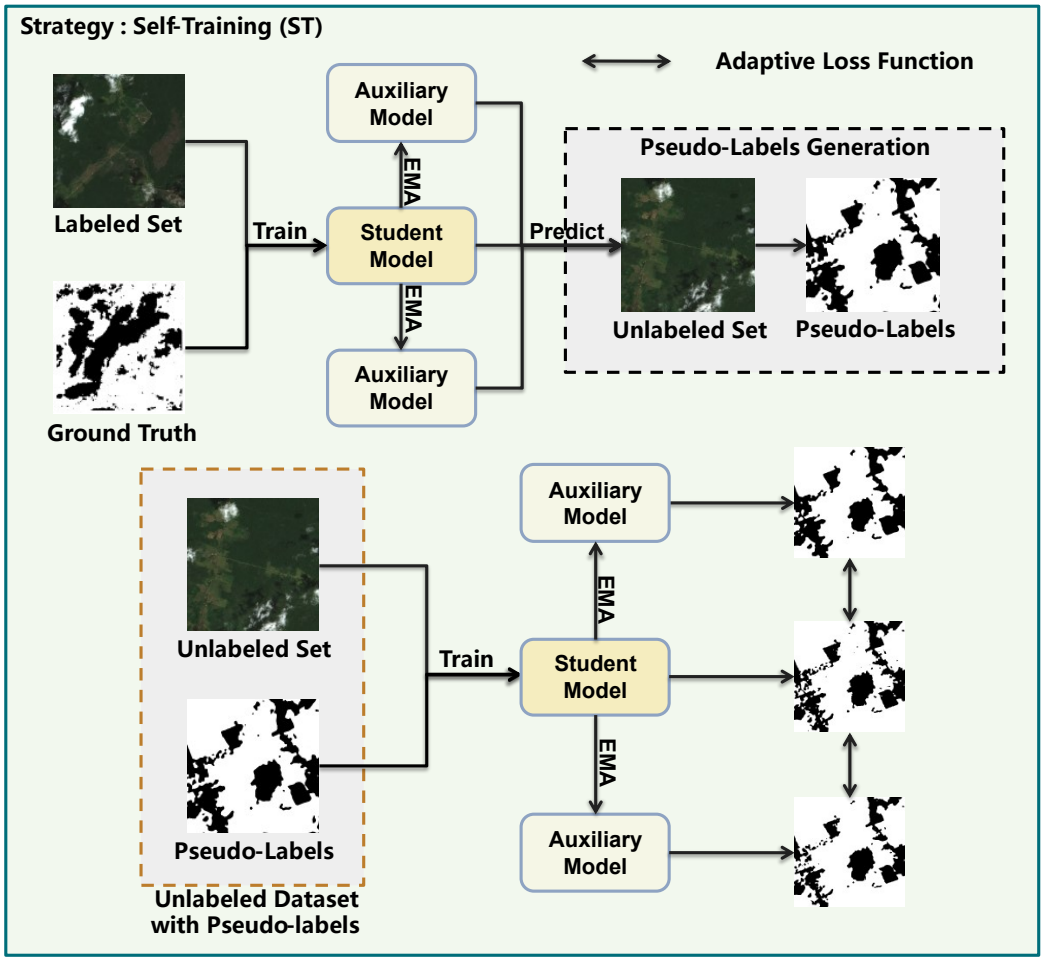


Figure 5. Overview of the proposed training strategy.

The pseudocode of our ST framework is shown in Algorithm 1.

Algorithm 1 ST with perturbation

Require: Labelled training set $D^l = \{(x_i, y_i)\}_{i=1}^{M}$,
Unlabelled training set $D^u = \{u_i\}_{i=1}^{N}$,
Strong/VAT/weak augmentations $A^s / A^V / A^w$,
Teacher/student model $t^l, t^{a1}, t^{a2}/s$
Ensure: Student model s

for minibatch $\{(x_i, y_i)\}_{i=1}^{M} \subset D^l$ **do**
  for k $\in \{1, ..., M\}$ **do**
    $x_k, y_k \leftarrow A^V(A^w(x_k, y_k))$
    $\hat{y_k} = t^l(x_k)$
    Update t^l to minimise L_{ce} and L_{focal} of $\{(\hat{y_k}, y_k)\}_{k=1}^{M}$
    Update t^{a1}, t^{a2} with EMA of s
  end for
end for
$\hat{D^u}$ =Label(D^u)
for minibatch $\{(x_i, y_i)\}_{i=1}^{B} \subset (D^l \cup \hat{D^u})$ **do**
  for k $\in \{1, ..., B\}$ **do**
    $x_k, y_k \leftarrow A^V(A^s(A^w(x_k, y_k)))$
    $\hat{y_k} = s(x_k)$
    $\hat{y_{k2}} = 0.5(t^{a1}(x_k) + t^{a2}(x_k))$
    Update s to minimise L_{ce} and L_{focal} of $\{(\hat{y_k}, y_k)\}_{k=1}^{B}$
    Update s to minimise L_{ce} and L_{focal} of $\{(\hat{y_{k2}}, \hat{y_k})\}_{k=1}^{B}$
    Update t^{a1}, t^{a2} with EMA of s
  end for
end for
return s

In this training strategy, strong augmentation refers to data transformations that alter the colour, contrast, and other properties of the image, such as colorJitter, random greyscale, blur, etc. On the other hand, weak augmentation pertains to transformations like resizing, cropping, and horizontal flipping that do not modify the main features of the original image. It is important to note that we employ weak augmentation during the supervised phase, whereas in the unsupervised phase, we adopt a combination of strong and weak augmentations. Both of these types of augmentation fall under input-level image perturbation and, together with subsequent transformations like VAT (feature-level perturbation), constitute multi-level perturbation.

3.4. Multi-Level Perturbation (MP)

In the process of model learning, the most basic self-training method will overfit the errors during iteration and reduce the performance of the student model. To better capture the intrinsic information of forest images and mitigate the risk of overfitting incorrect labels, we propose a multi-level perturbation strategy including input-level image perturbation, feature-level perturbation, and model-level perturbation.

3.4.1. Input-Level Image Perturbation

Given the limited prominence of colour features and the difficulty in learning image features, we opted for a mixed-image augmentation as input-level perturbation for fully supervised and unsupervised learning, which allows the model to prioritise the overall image rather than focusing solely on partial regions. Specifically (as shown in Figure 3), in the fully supervised stage, we applied weak augmentations, including resize, crop, and horizontal flip transformations, to the input images. In the unsupervised stage, we employed strong augmentations, including colorJitter, random greyscale, and blur, as well as random cutout and contrast/colour-filtering techniques, on the input images.

3.4.2. Feature-Level Perturbation VAT

Feature perturbation involves the creation of adversarial perturbations that pose a challenge by intentionally disrupting the cluster or low-density assumption. This is achieved by transforming the image features, computed from the model encoder, towards the classification boundaries within the feature space. One effective approach to generating such adversarial feature noise is through virtual adversarial training (VAT), which optimises a perturbation vector to maximise the divergence between correct and adversarial classifications. However, current methods estimate the adversarial noise by using the same single network where the consistency loss is applied. Thus, we suggest estimating the adversarial noise using the more accurate teachers and then applying this estimated noise to the feature of the student model, which we call VAT feature perturbation. In particular, we used the VAT feature enhancement method in both the fully supervised and semi-supervised stages and achieved significant improvements. VAT is used in the student model output, i.e.,

$$p_s(x) = decoder_s(encoder_s(x) + r_{adv}), \tag{3}$$

where p_s is the prediction result of the student model for each pixel, and $encoder_s$ and $decoder_s$ are, respectively, the encoder and decoder of the student model. The adversarial feature perturbation $r_{adv} \in Z$ is estimated from the response of the ensemble of teacher models using

$$\begin{aligned} &\max \; d(0.5 \cdot (decoder_{a1}(z^s) + decoder_{a2}(z^s)), 0.5 \cdot (decoder_{a1}(z^s + r_{adv}) + decoder_{a2}(z^s + r_{adv}))), \\ &\text{subject to } ||r_{adv}||_2 \leq \xi, \end{aligned} \tag{4}$$

where $z^s = s(x)$, and d(.) is the sum of the pixel-wise Kullback–Leibler (KL) divergence between the original and perturbed pixel predictions. Figure 6 illustrates feature-level perturbation.

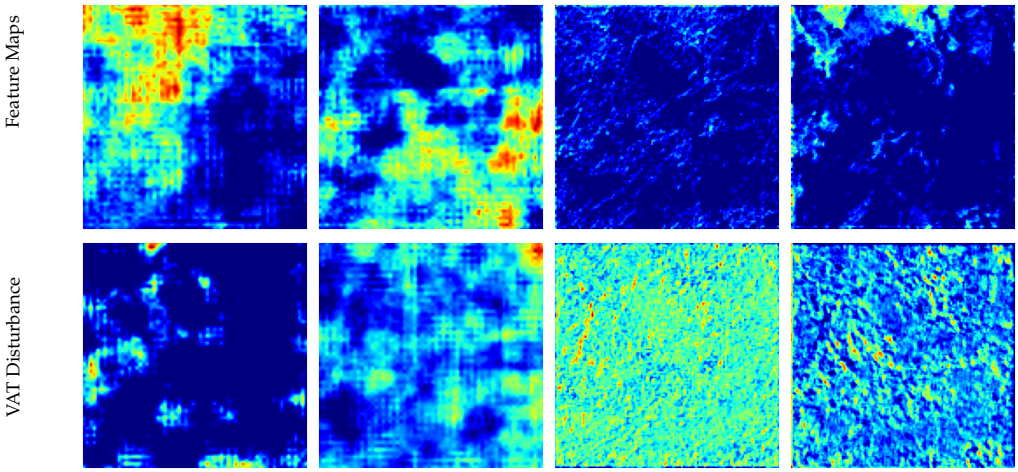


Figure 6. Visualisation results of feature maps before and after feature-level perturbation: feature maps (extracted features of the image after input perturbation by the encoder) and VAT disturbance (image with injected VAT perturbation after feature extraction).

3.4.3. Model-Level Perturbation

To enhance the perturbations of the self-training method, we introduced model-level perturbations through auxiliary teachers. In the unsupervised stage, we utilised the teacher model, which received parameters from the student model with EMA, to predict the input images. Based on the assumption of consistency regularisation, multiple models trained at different stages should have similar predictions for the same image. By measuring the differences between the labels generated by teachers and the student model using a loss function, we provided feedback to the student model and updated its parameters, thus improving the generalisation performance of the model.

It is worth noting that, based on the idea of gradually strengthening model learning, we used weak data augmentation methods and feature-level VAT perturbation instead of directly incorporating strong data augmentation methods into the supervised learning. We found from experiments that training the model from weak to strong effectively improved the performance of the supervised learning as well. Using strong data augmentation throughout all epochs may lead to a decrease in the performance of the model. However, if strong data augmentation is selectively applied in the latter half of the epochs, there is potential for improvement in the performance. Since VAT perturbation searches for pixels with more noise in the current features, it is related to the performance and is considered a weak-to-strong data augmentation method. Meanwhile, based on the idea of progressive learning, we gradually shifted the focus of the loss function towards the labels generated by the teacher model by dynamically adjusting its attention. This adjustment allowed the loss function to increasingly prioritise the labels provided by the teacher model as the training progressed, leading to more reasonable model learning.

3.5. Ensemble Voting: Pseudo-Label Generation Strategy (EV)

In a model based on the self-training paradigm, the assignment of pseudo-labels to unlabelled data using a trained teacher model plays a crucial role. However, past self-training methods have faced limitations when manually setting confidence thresholds and filtering the softmax probability results from the model. Although this approach has contributed to some improvement in the confidence of the model, it suffers from issues such as the inefficient utilisation of all pixels, heavy reliance on artificially set hyperparameters, and the presence of fuzzy segmentation boundaries.

To address these challenges and enhance the model predictions, we integrated TTA technology. TTA leverages multiple scales of images, and we uniformly resized them using bilinear interpolation to ensure consistent dimensions for predictions. In our experiments, we explored different scaling factors, including 0.5, 0.75, 1.0, 1.5, and 2.0, and weighted the predictions obtained at each scale. Furthermore, we horizontally flipped the images at each scale to augment the ability of the model to recognise objects in the image. By aggregating the predictions from all scales and applying softmax, we obtained the final prediction.

Additionally, we introduce auxiliary teachers to further support the decision-making process of the model. Following the ensemble learning vote concept, we utilise TTA technology to predict labels using the student model. Subsequently, we combine the TTA-augmented predicted labels from the auxiliary teachers and the student model, assigning appropriate weights to achieve more reliable and smoother labelling results.

Overall, our labelling method, outlined in Algorithm 2, effectively leverages TTA to enhance the model predictions and achieve improved performance in forest cover mapping. The incorporation of TTA for both auxiliary teachers and the student model contributes to better decision making and yields superior segmentation results.

Algorithm 2 Labelling

Require: Unlabelled training set $D^u = \{u_i\}_{i=1}^N$,
Test-Time Augmentation A^t,
Teacher/Auxiliary teacher model $t^l / t^{a1}, t^{a2}$
Ensure: pseudo-label $\hat{D^u}$
for minibatch $\{(x_i)\}_{i=1}^B \subset D^u$ **do**
  for k $\in \{1, ..., B\}$ **do**
    $y_{k1} = A^t(t^l(x_k))$
    $y_{k2} = A^t(t^{a1}(x_k))$
    $y_{k3} = A^t(t^{a2}(x_k))$
    $y_k = 0.5(0.5(y_{k2} + y_{k3}) + y_{k1})$
    $\hat{D^u} \leftarrow y_k$
  end for
end for
return $\hat{D^u}$

3.6. Adaptive Loss (AL)

To address the issue of dataset imbalance, we trained the model using a combination of cross-entropy loss and focal loss. Cross-entropy is given by

$$L_{ce}(y, \hat{y}) = -\sum_{c=1}^{C} y_c \log(\hat{y}_c), \tag{5}$$

where y is a one-hot vector with length C, representing the true class; y_c is the c^{th} element in the vector y; $\hat{y}$ is the predicted probability distribution vector of the model; and $\hat{y}_c$ represents the probability that the model predicts the c^{th} class. Focal loss is given by

$$L_{focal}(y, \hat{y}) = -\sum_{c=1}^{C} y_c (1 - \hat{y}_c)^{\gamma} \log(\hat{y}_c), \tag{6}$$

where y is a one-hot vector with length C, representing the true class; y_c is the c^{th} element in the vector y; $\hat{y}$ is the predicted probability distribution vector of the model; $\hat{y}_c$ represents the probability that the model predicts the c^{th} class; and γ is a hyperparameter called the focusing parameter, which is used to adjust the degree of attention that the loss function pays to the predicted probability of different classes. When $\gamma > 0$, focal loss pays more attention to the mispredicted samples, thus reducing the problem of class imbalance.

It is worth noting that focal loss was originally designed to solve the problem of extremely imbalanced positive and negative samples, as well as samples that are difficult to classify. Given that the focal loss adjusts the ratio of positive and negative sample loss adaptively based on the difficulty of the dataset, we did not remove focal loss when conducting experiments on balanced datasets. At the same time, to minimise problems caused by the manual setting of hyperparameters, we introduced an automatic coefficient,

$$L^u = \frac{totaliter - iter}{totaliter}(L_{ce}(\hat{y_k}, y_k) + L_{focal}(\hat{y_k}, y_k)) + \frac{iter}{totaliter}(L_{ce}(\hat{y_{k2}}, \hat{y_k}) + L_{focal}(\hat{y_{k2}}, \hat{y_k})), \tag{7}$$

where $\hat{y_{k2}}$ represents the output of the auxiliary teachers, $\hat{y_k}$ represents the pseudo-labels generated by the previous stage model, and y_k represents the predicted results of the student model. As training progresses, the reliability of the pseudo-labels generated during the initial training phase decreases, and the predictions from the auxiliary teachers become more accurate. Therefore, the loss for the original labelled data gradually decreases as the training progresses, while the loss for the predictions between the auxiliary teachers and the student model gradually increases. At the same time, the loss between the predicted results of the model and the pseudo-labels generated by the previous stage of the model and the auxiliary teachers can also be seen as a consistency regularisation method. Currently, we have only tried linear transformations and have not explored whether there are better adaptive adjustment coefficients.

4. Data and Experiments

4.1. Data Description

In evaluating the performance of forest cover mapping, we utilised two datasets sourced from the SentinelHub satellite image database. The details regarding the number of images and the distribution between forest and non-forest categories are presented in Table 1. Both datasets consist of four-band data, with one originating from the Amazon Rainforest and the other from the Atlantic Forest (Mata Atlantica) [33]. The geographical distribution of these biomes can be observed in Figure 7, accompanied by sample images highlighting the dataset concentration in two distinct regions.

Table 1. Number of images within each dataset, as well as the forest (F) and non-forest (NF) class balance within each dataset.

Dataset	Number of Images (F-NF Class Balance)		
	Training	Testing	Validation
RGB 3-band Amazon Forest	250 (50.0–50.0%)	20 (49.8–50.2%)	100 (47.8–52.2%)
RGB 3-band Atlantic Forest	250 (33.3–66.7%)	20 (31.5–68.5%)	100 (33.8–66.2%)

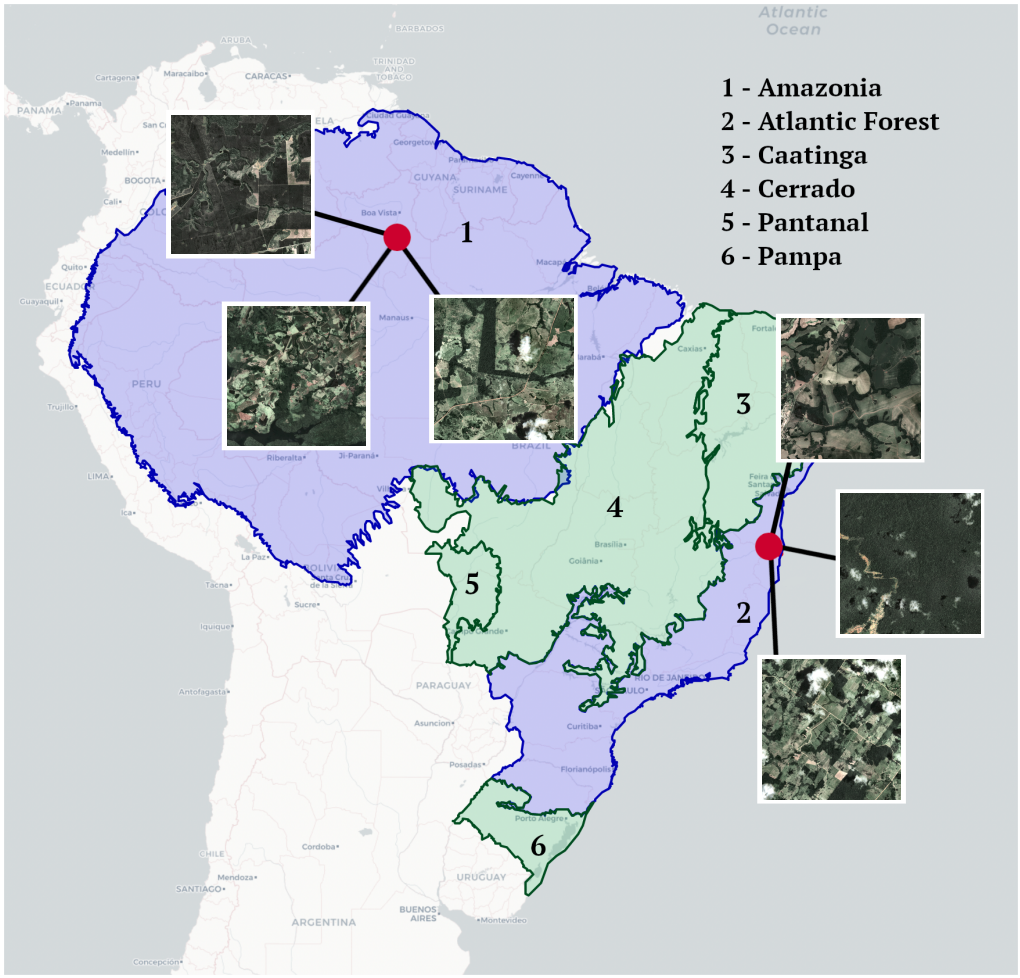


Figure 7. Map of biomes in Amazon Rainforest and Atlantic Forest.

To streamline the training process, we adopted an approach similar to a related work [34] by training solely on the RGB channels extracted from the four-band dataset. Notably, the model demonstrated favourable performance. Each image in the datasets has dimensions of (512, 512, 3), while each forest cover mapping mask is represented by (512, 512, 1).

These datasets provide insights into two real-world scenarios: one involving an imbalanced class distribution and complex images (Atlantic Forest), posing a challenging learning task, and the other involving a balanced class distribution and relatively simple images (Amazon Forest).

4.2. Experimental Settings

To ensure a fair comparison with most existing works, we maintained consistent hyperparameters between the supervised pre-training of the teacher models and the semi-supervised re-training of the student model. Specifically, we set the batch size to 8 during training with a V100-SXM2-32GB GPU. For optimisation, we used the SGD optimiser with an initial base learning rate of 0.001 for the backbones. The learning rate of the randomly

initialised segmentation head was 10 times larger than that of the backbones. Additionally, we adopted poly scheduling to decay the learning rate during the training process, i.e.,

$$lr = baselr \cdot (1 - \frac{iter}{totaliter})^{0.9}.$$

The model was trained for 80 epochs using weak data augmentations, which includes the random flipping and resizing of training images between 0.5 and 2.0. For strong data augmentations on unlabelled images, we used colorJitter with the same intensity as in [**?**], greyscale, blur (same as in [35]), and cutout with random values filled. The cutout regions were ignored in loss computation. During the pseudo-labelling phase, all unlabelled images underwent TTA, which involved five scales and horizontal flipping. The testing images were evaluated at their original resolution, and no post-processing techniques were employed. It is worth noting that to enable a fair comparison with most existing works, we have not incorporated any advanced optimisation strategies, such as OHEM in [36], auxiliary supervision in [36,37], or SyncBN, into our method.

4.3. Evaluation Metrics

To evaluate the performance of our proposed method in forest cover mapping, we measured several evaluation metrics, including IoU, mean IoU (mIoU), mean precision, mean recall, mean F1-score, and accuracy for both forest and non-forest classes, i.e.,

$$IoU = \frac{TP}{TP + FP + FN} \tag{8}$$

$$mIoU = \frac{1}{c} \sum_{i=1}^{c} \frac{TP_i}{TP_i + FP_i + FN_i} \tag{9}$$

$$mean\ precision = \frac{1}{c} \sum_{i=1}^{c} \frac{TP_i}{TP_i + FP_i} \tag{10}$$

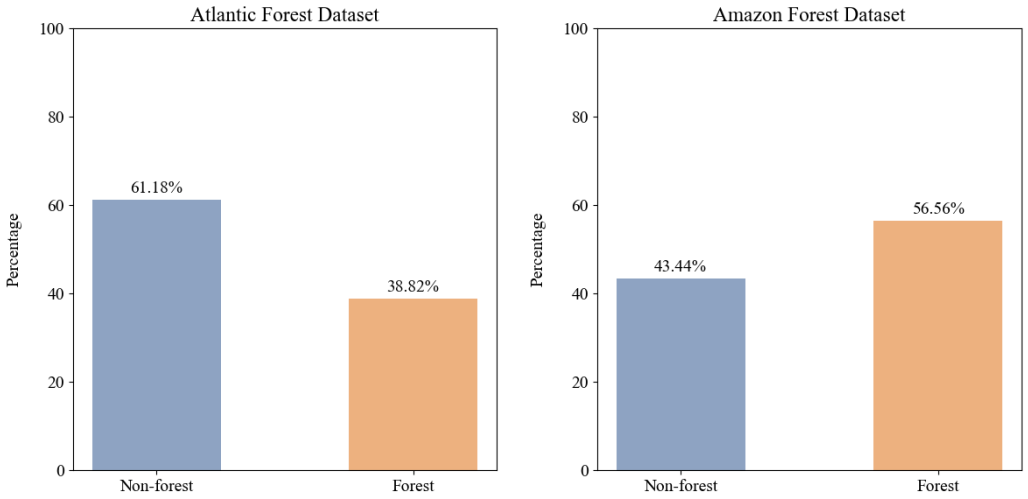
$$mean\ recall = \frac{1}{c} \sum_{i=1}^{c} \frac{TP_i}{TP_i + FN_i} \tag{11}$$

$$mean\ F1 = \frac{1}{c} \sum_{i=1}^{c} \frac{2TP_i}{2TP_i + FP_i + FN_i} \tag{12}$$

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \tag{13}$$

where c represents the number of shared classes between the benchmark datasets ($c = 2$). Figure 8 shows the class distribution of the two datasets used.

The evaluation metrics were computed using the confusion matrix generated by our semi-supervised segmentation framework, which contains the pixel numbers of true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). Specifically, mIoU and IoU measure the similarity between predicted and ground-truth forest/non-forest areas, while accuracy measures the overall percentage of correctly classified pixels. The mean precision, mean recall, and mean F1-score consider both precision and recall, which are suitable for multi-class and pixel-level classification tasks. Our results demonstrate the superior performance of our proposed method in accurately mapping forest cover, as evidenced by the higher values of these evaluation metrics.



(**a**) (**b**)

Figure 8. Class distribution of the datasets: (**a**) class distribution of the Atlantic dataset and (**b**) class distribution of the Amazon Forest dataset.

4.4. Results and Analysis

Based on the aforementioned forest RS datasets, we compared our proposed method with several SOTA semi-supervised semantic segmentation frameworks, including the feature-perturbation-based CCT [25], the multi-perturbation-based PS-MT [26], and baseline ST with TTA. We also included the supervised DeeplabV3+ [38] model using ResNet-101 [39] trained only with labelled data as a baseline. All semi-supervised methods were implemented with identical experimental conditions and settings to ensure fairness. The results demonstrate the effectiveness and superiority of our proposed method in accurately mapping forest cover. Additionally, our method has huge potential for practical applications in forest monitoring and management.

4.4.1. Comparison Results on the Atlantic Forest Dataset

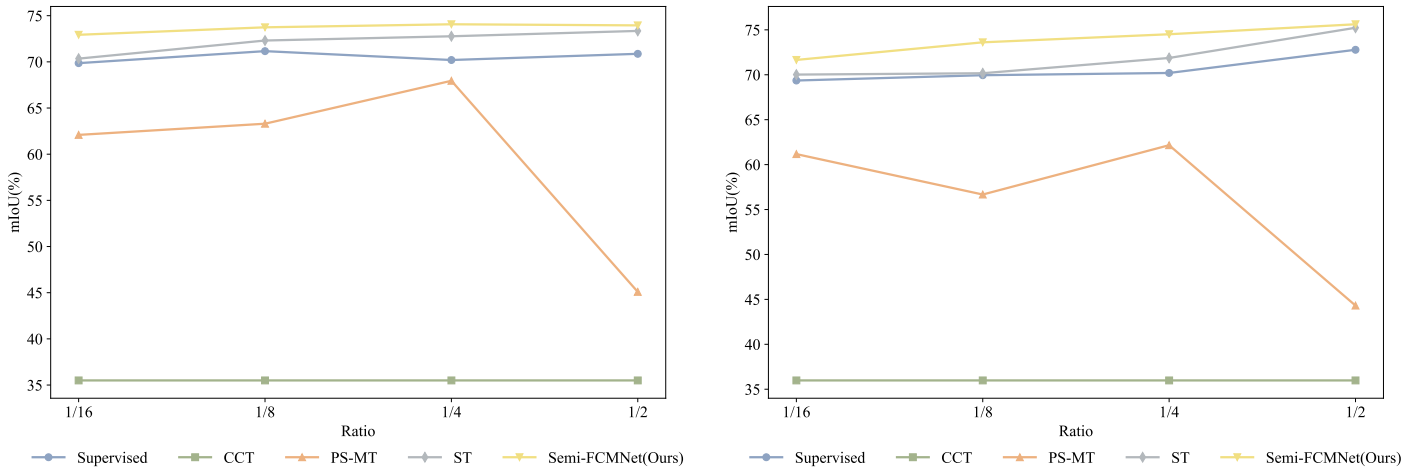
Tables 2 and 3 show comparative results with other SOTA methods on the validation and test sets of the Atlantic dataset. As introduced in Section 2, the CCT method based on image surface perturbation cannot effectively learn the information of satellite RS data, resulting in poor performance with inadequate fitting during the training process. Similarly, other semi-supervised models (e.g., PS-MT) also suffer from this problem and the instability caused by numerous hyperparameters, performing far worse than our proposed method on multiple metrics after our adjustment. Methods based on self-training and data augmentation using colour transformations exhibit stronger performance, and our method, which is an improvement on self-training, outperforms them. Meanwhile, compared with fully supervised methods, our approach also shows significant superiority.

Figure 9 graphically shows the performance of mIoU for different models. Specifically, on the validation set, except for a slightly lower forest segmentation IoU compared to the ST method at the 1/16 split, our model outperforms the other SOTA methods in all metrics. Similarly, on the test set, our method surpasses the majority of the SOTA methods in most metrics. For the important metric mIoU in forest cover segmentation, our model demonstrates superior performance across different splits, as well as on the validation and test sets. This indicates that our model is capable of effectively learning information from satellite RS images and exhibits strong robustness.

We present partial visual comparison results of all the methods in Figure 10. On the Atlantic Forest dataset, due to the data imbalance and the difficulty of image learning, it can be observed that our proposed method produces smoother boundaries and highlights more details compared to the fully supervised methods. Moreover, our method yields predictions that are closer to the ground truth compared to other semi-supervised methods.

Table 2. Comparison results with other SOTA methods on Atlantic Forest validation dataset using evaluation metrics (%). (*Denotes that we have made appropriate modifications to the model, alleviating the issue of unsuccessful training.)

		Validation				
Labelled Data	**Metric**	**SupOnly**	**CCT**	**PS-MT ***	**ST (TTA)**	**Semi-FCMNet**
1/16(16)	mIoU	69.86	35.50	62.09	70.35	**72.93**
	IoU(NF)	76.34	54.22	73.01	72.85	**79.85**
	IoU(F)	63.37	16.79	51.18	**67.84**	65.99
	Accuracy	76.30	58.09	78.96	80.79	**81.48**
	Mean precision	81.32	49.99	76.63	83.54	**83.58**
	Mean recall	83.86	49.99	75.59	83.33	**84.92**
	Mean F1-score	82.08	49.53	76.05	82.57	**84.15**
1/8(32)	mIoU	71.16	35.50	63.30	72.31	**73.74**
	IoU(NF)	78.85	54.22	77.13	79.96	**80.67**
	IoU(F)	63.46	16.79	49.47	64.65	**66.79**
	Accuracy	81.54	58.09	81.32	81.10	**82.67**
	Mean precision	82.58	49.99	82.75	83.54	**84.23**
	Mean recall	83.28	49.99	74.65	83.86	**85.27**
	Mean F1-score	82.91	49.53	76.64	83.69	**84.69**
1/4(63)	mIoU	70.20	35.50	67.95	72.77	**74.07**
	IoU(NF)	77.75	54.22	80.19	80.21	**81.21**
	IoU(F)	62.64	16.79	55.70	65.31	**66.91**
	Accuracy	79.77	58.09	82.14	83.06	**83.96**
	Mean precision	81.73	49.99	83.59	83.78	**84.66**
	Mean recall	82.95	49.99	77.98	84.28	**85.16**
	Mean F1-score	82.25	49.53	80.27	84.30	**84.90**
1/2(125)	mIoU	70.87	35.50	45.09	73.35	**73.95**
	IoU(NF)	77.50	54.22	70.41	79.33	**80.51**
	IoU(F)	64.23	16.79	19.77	67.36	**67.38**
	Accuracy	77.90	58.09	72.42	78.95	**81.65**
	Mean precision	82.00	49.99	81.63	83.65	**84.19**
	Mean recall	84.24	49.99	59.62	84.32	**85.80**
	Mean F1-score	82.77	49.53	57.82	84.48	**84.85**



(**a**) (**b**)

Figure 9. Performance of mIoU metric for different models on the Atlantic Forest dataset: (**a**) performance on the validation set of the Atlantic Forest and (**b**) performance on the test set of the Atlantic Forest.

Table 3. Comparison results with other SOTA methods on Atlantic Forest test dataset using evaluation metrics (%). (*Denotes that we have made appropriate modifications to the model, alleviating the issue of unsuccessful training.)

		Test				
Labelled Data	**Metric**	**SupOnly**	**CCT**	**PS-MT ***	**ST (TTA)**	**Semi-FCMNet**
1/16(16)	mIoU	69.36	35.98	61.17	70.02	**71.65**
	IoU(NF)	77.37	55.77	73.77	78.97	**80.60**
	IoU(F)	61.33	16.18	48.58	61.06	**62.70**
	Accuracy	76.32	59.25	78.98	79.89	**82.38**
	Mean precision	80.67	49.99	75.72	81.49	**82.94**
	Mean recall	**83.49**	49.99	74.68	82.72	83.41
	Mean F1-score	81.64	49.73	75.14	82.04	**83.16**
1/8(32)	mIoU	69.95	35.98	56.67	70.17	**73.61**
	IoU(NF)	79.73	55.77	76.21	79.19	**82.45**
	IoU(F)	60.15	16.18	37.13	61.14	**64.76**
	Accuracy	82.47	59.25	79.14	80.32	**85.40**
	Mean precision	82.09	49.99	82.81	81.66	**84.79**
	Mean recall	81.76	49.99	68.32	82.70	**84.22**
	Mean F1-score	81.92	49.73	70.32	82.13	**84.49**
1/4(63)	mIoU	70.20	35.98	62.16	71.87	**74.51**
	IoU(NF)	79.81	55.77	78.60	80.83	**83.26**
	IoU(F)	61.01	16.18	45.73	62.90	**65.76**
	Accuracy	81.93	59.25	81.87	82.81	**86.75**
	Mean precision	82.18	49.99	84.61	83.16	**85.67**
	Mean recall	82.38	49.99	72.85	83.47	**84.61**
	Mean F1-score	82.28	49.73	75.38	83.31	**85.10**
1/2(125)	mIoU	72.78	35.98	44.31	75.23	**75.62**
	IoU(NF)	81.22	55.77	72.15	83.02	**83.24**
	IoU(F)	64.33	16.18	16.47	67.43	**68.00**
	Accuracy	82.43	59.25	73.60	**84.12**	84.11
	Mean precision	83.54	49.99	84.96	85.20	**85.39**
	Mean recall	84.46	49.99	58.17	86.13	**86.49**
	Mean F1-score	83.96	49.73	56.04	85.63	**85.90**

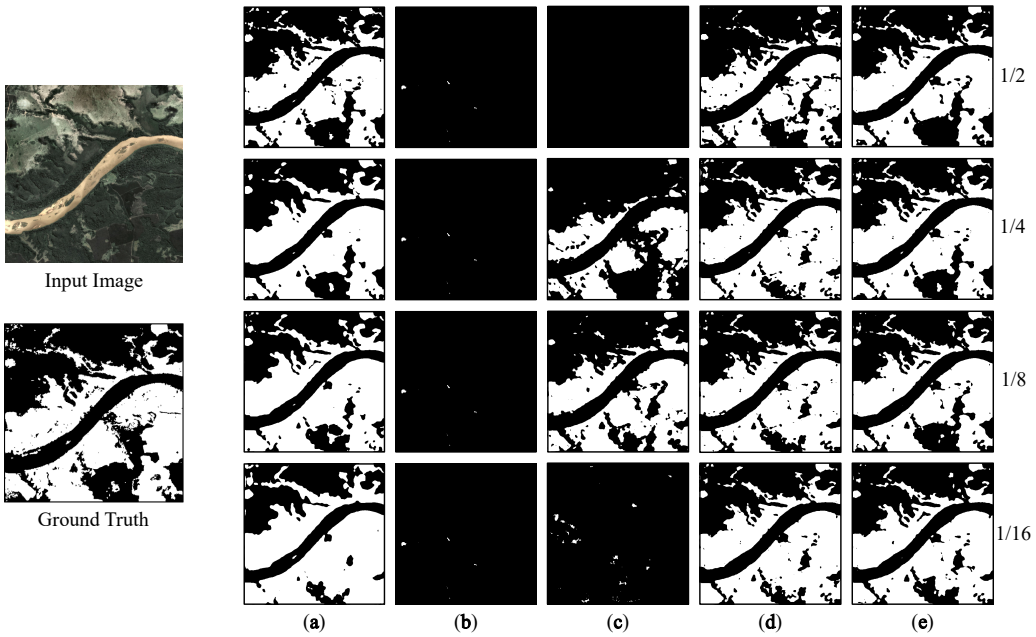


Figure 10. Visualisation of model predictions on different partitions on the Atlantic Forest dataset: (**a**) supervised; (**b**) CCT; (**c**) PS-MT; (**d**) ST; and (**e**) Semi-FCMNet.

4.4.2. Comparison Results on the Amazon Forest Dataset

Tables 4 and 5 present the comparison results on the Amazon Forest validation and test sets using our evaluation metrics of accuracy, mRecall, mPrecision, mF1, and mIoU. The tables show that on the datasets with class balance and lower image difficulty, the supervised Deeplabv3+ achieves better segmentation accuracy than the results obtained on the Atlantic Forest dataset. Similarly, the performance of PS-MT gradually improves, but due to the existence of manually adjusted semi-supervised loss ratio coefficients, the overall performance remains unstable. Additionally, the CCT method fails to learn image information due to the challenges presented by satellite RS datasets. Consistent with our expectations, the self-training method still exhibits strong stability and significantly improves performance across all evaluation metrics compared to fully supervised methods. However, as the amount of labelled data increases, the performance of the model decreases. Furthermore, the performance of the supervised method varies due to varying image difficulties, with its performance at the 1/8 partition being inferior to that achieved at the 1/16 partition. In contrast, our proposed Semi-FCMNet, which enhances the perturbation of the model, leads to performance improvements. Moreover, our method also outperforms the supervised method and other SOTA methods on the test set, indicating its strong generalisation ability.

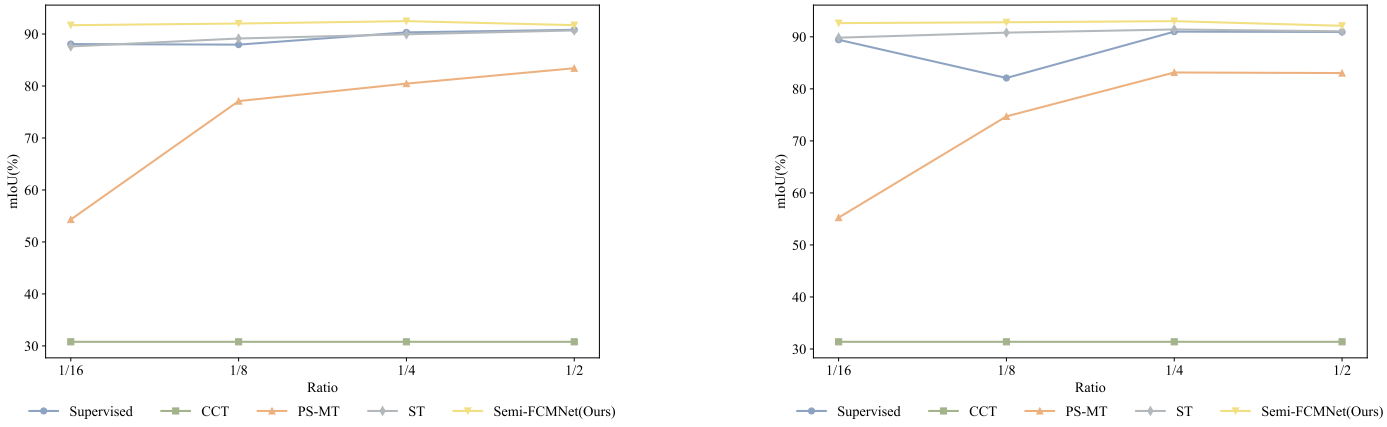
Table 4. Comparison results with other SOTA methods on the Amazon Forest validation dataset using evaluation metrics (%).

Validation						
Labelled Data	**Metric**	**SupOnly**	**CCT**	**PS-MT**	**ST (TTA)**	**Semi-FCMNet**
1/16(16)	mIoU	88.07	30.79	54.31	87.61	**91.70**
	IoU(NF)	87.63	41.24	43.66	87.05	**91.69**
	IoU(F)	88.51	20.34	64.96	88.16	**91.69**
	Accuracy	94.11	48.91	72.44	93.00	**96.73**
	Mean precision	93.64	50.00	80.48	93.41	**95.79**
	Mean recall	93.67	50.00	71.27	93.37	**95.85**
	Mean F1-score	93.65	46.10	69.76	93.39	**95.81**
1/8(32)	mIoU	87.96	30.79	77.10	89.14	**92.02**
	IoU(NF)	87.48	41.24	74.18	88.70	**91.76**
	IoU(F)	88.43	20.34	80.03	89.57	**92.28**
	Accuracy	93.78	48.91	87.31	94.40	**96.78**
	Mean precision	93.58	50.00	89.11	94.25	**95.82**
	Mean recall	93.60	50.00	86.84	94.26	**95.88**
	Mean F1-score	93.59	46.10	87.04	94.25	**95.84**
1/4(63)	mIoU	90.32	30.79	80.46	89.95	**92.48**
	IoU(NF)	90.04	41.24	79.81	89.63	**92.19**
	IoU(F)	90.59	20.34	81.11	90.26	**92.77**
	Accuracy	96.21	48.91	89.18	95.71	**96.45**
	Mean precision	94.89	50.00	89.15	94.68	**96.08**
	Mean recall	94.96	50.00	89.19	94.75	**96.10**
	Mean F1-score	94.91	46.10	89.17	94.70	**96.09**
1/2(125)	mIoU	90.81	30.79	83.42	90.69	**91.71**
	IoU(NF)	90.66	41.24	83.32	90.55	**91.49**
	IoU(F)	90.96	20.34	83.52	90.82	**91.93**
	Accuracy	**97.76**	48.91	90.96	97.89	97.28
	Mean precision	95.21	50.00	91.06	95.15	**95.66**
	Mean recall	95.29	50.00	91.10	95.23	**95.74**
	Mean F1-score	95.18	46.10	90.96	95.11	**95.67**

Table 5. Comparison results with other SOTA methods on the Amazon Forest test dataset using evaluation metrics (%).

		Test				
Labelled Data	**Metric**	**SupOnly**	**CCT**	**PS-MT**	**ST (TTA)**	**Semi-FCMNet**
1/16(16)	mIoU	89.44	31.38	55.25	89.84	**92.64**
	IoU(NF)	89.53	42.72	46.04	89.85	**92.65**
	IoU(F)	89.33	20.04	64.46	89.82	**92.62**
	Accuracy	95.75	49.91	72.73	95.17	**96.76**
	Mean precision	94.45	50.00	81.00	94.64	**96.18**
	Mean recall	94.42	50.00	72.62	94.64	**96.18**
	Mean F1-score	94.42	46.62	70.71	94.64	**96.18**
1/8(32)	mIoU	82.09	31.38	74.71	90.81	**92.80**
	IoU(NF)	81.45	42.72	72.01	90.84	**92.84**
	IoU(F)	82.71	20.04	77.42	90.76	**92.79**
	Accuracy	87.14	49.91	85.72	96.03	**96.95**
	Mean precision	90.37	50.00	87.86	95.19	**96.28**
	Mean recall	90.16	50.00	85.66	95.18	**96.27**
	Mean F1-score	90.16	46.62	85.49	95.18	**96.27**
1/4(63)	mIoU	91.00	31.38	83.15	91.42	**93.02**
	IoU(NF)	91.09	42.72	83.63	91.49	**93.00**
	IoU(F)	90.90	20.04	82.66	91.34	**93.02**
	Accuracy	96.73	49.91	90.81	**96.79**	96.62
	Mean precision	95.32	50.00	91.00	95.54	**96.38**
	Mean recall	95.29	50.00	90.82	95.52	**96.38**
	Mean F1-score	95.28	46.62	90.79	95.51	**96.38**
1/2(125)	mIoU	90.94	31.38	83.04	91.06	**92.12**
	IoU(NF)	91.08	42.72	83.86	91.25	**92.18**
	IoU(F)	90.79	20.04	82.23	90.86	**92.05**
	Accuracy	97.23	49.91	90.76	**97.60**	97.15
	Mean precision	95.32	50.00	91.27	95.44	**95.92**
	Mean recall	95.26	50.00	90.78	95.33	**95.90**
	Mean F1-score	95.25	46.62	90.73	95.32	**95.89**

Figure 11 graphically shows the performance of mIoU for different models.



(**a**) (**b**)

Figure 11. Performance of mIoU metric for different models on the Amazon Forest dataset: (**a**) performance on the validation set of the Amazon Forest and (**b**) performance on the test set of the Amazon Forest.

We also present partial visual comparison results of all methods on the Amazon Forest dataset in Figure 12. Due to the lower sample complexity of this dataset, most methods achieve good prediction results. However, it is worth noting that our proposed method generates accurate predictions at different partition ratios, as shown in the figure.

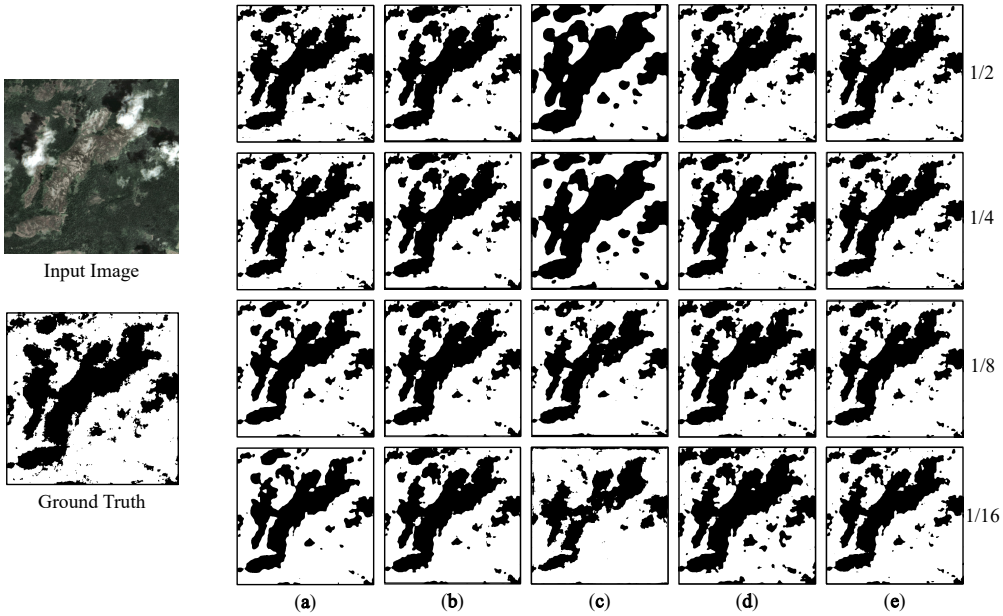


Figure 12. Visualisation of model predictions under different partitions on the Amazon Forest dataset: (**a**) supervised; (**b**) CCT; (**c**) PS-MT; (**d**) ST; and (**e**) Semi-FCMNet.

4.5. Ablation Experiments

Ablation studies were conducted to validate the effectiveness of each key component of our proposed method. Our method mainly consists of the following four core components: hybrid perturbations (including input image representation level (MP), feature level, and model level); AL; and a pseudo-label generation strategy based on TTA and multi-model voting (PGS). We present specific metric data in Tables 6 and 7, along with the visualisation results of the ablation experiments in Figures 13 and 14.

Base: When employing the basic self-training method on the imbalanced Atlantic Forest dataset with high sample learning difficulty, the performance of the model at the 1/16 partition is slightly inferior to that of the fully supervised method. However, in other partitions, the semi-supervised method showcases its superiority by exhibiting better performance on the validation and test sets compared to the fully supervised method. This outcome fully demonstrates the advantages of the semi-supervised approach. However, on the Amazon Forest dataset with lower sample learning difficulty, the basic self-training method is prone to overfitting to the noise and cannot exceed the performance of the fully supervised method.

MP: After adding the MP method, the various indicators of the model are further improved, which proves that adding perturbations to the self-training paradigm, i.e., integrating the consistency regularisation method with the self-training method, effectively improves model performance.

EV: When the pseudo-label generation strategy based on TTA and multi-model voting (EV) is added, the scores of the model on important indicators, e.g., mIoU and mF1-score, are improved in various partitions of datasets with different data distributions. It can be seen that the method of generating pseudo-labels has a significant impact on the performance of the self-training paradigm, and our pseudo-label generation method effectively improves the performance of the model.

AL: After adding the adaptive loss, the model dynamically adjusts its focus on the loss. At the beginning of the training, the model pays more attention to the predicted results of the teacher model in the previous round. However, due to the limited performance of the teacher in the previous round, the current student model cannot be effectively improved in the later stage of training. The introduction of the adaptive loss makes the model focus more on the difference between its multiple prediction results in the later stage and improve

its own performance through consistency regularisation. The improvement in various indicators reflects the correctness of our method.

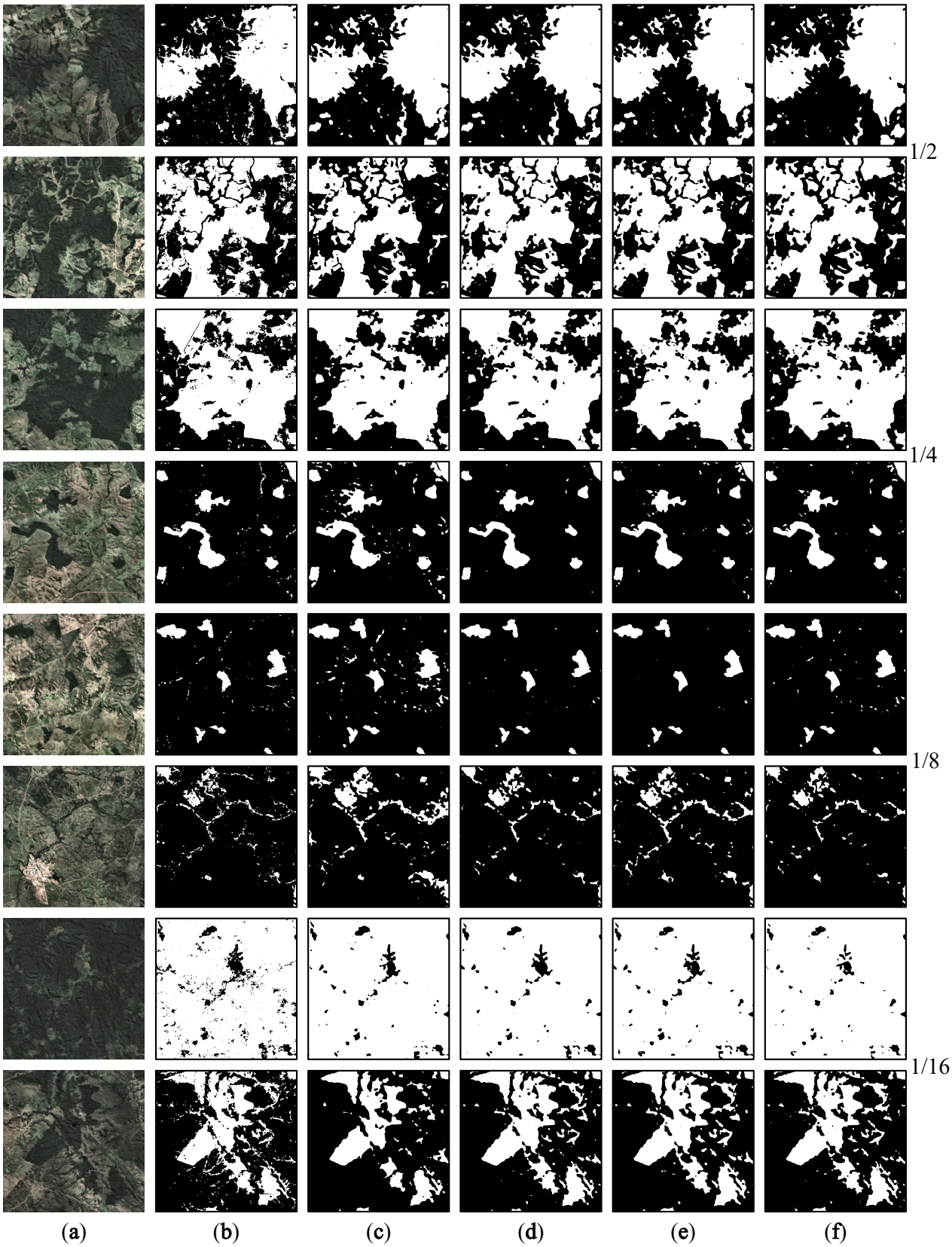


Figure 13. Visualisation of ablation experiment results on the Atlantic Forest dataset: (**a**) input image; (**b**) ground truth; (**c**) Base; (**d**) EV; (**e**) EV + MP; and (**f**) all.

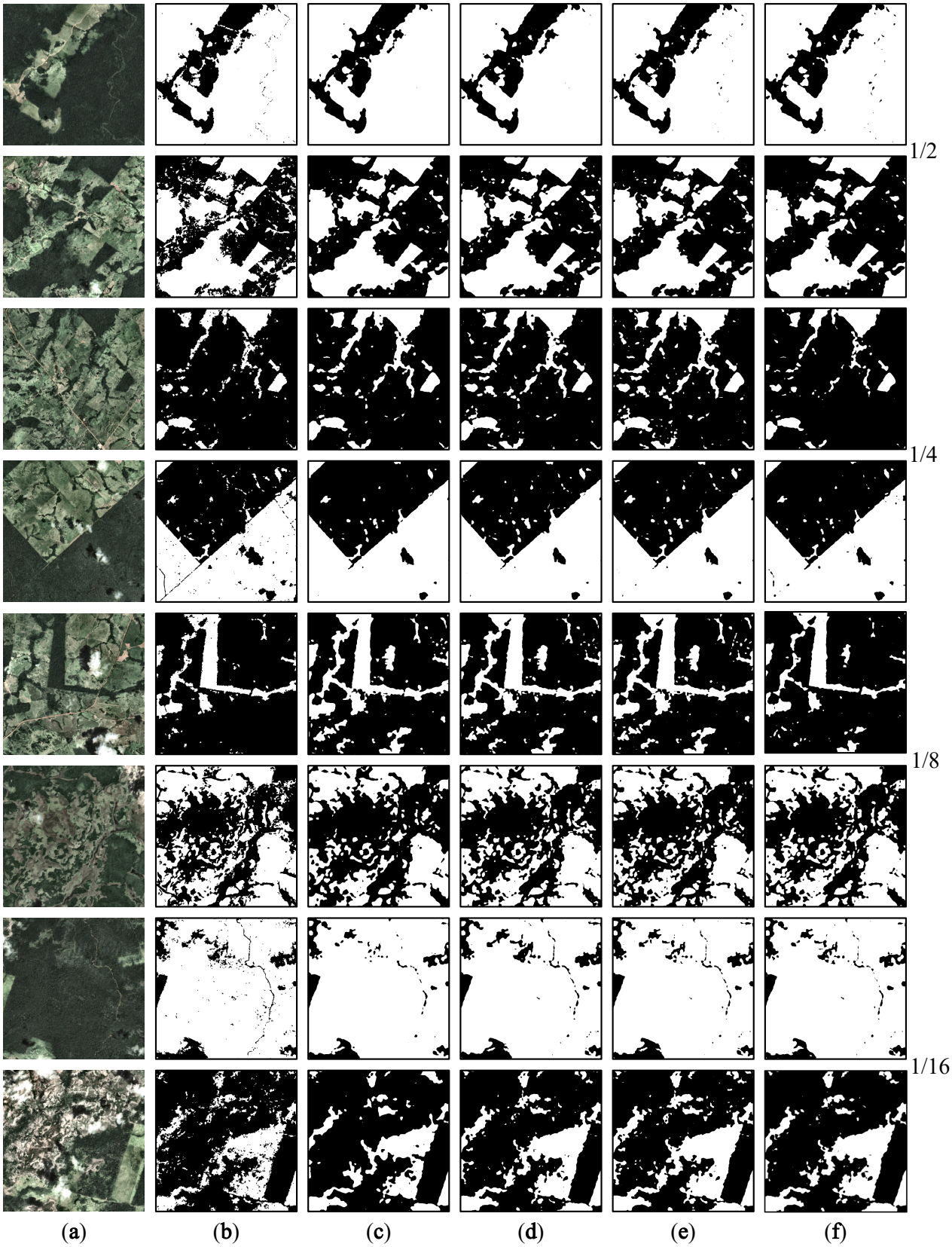


Figure 14. Visualisation of ablation experiment results on the Amazon Forest dataset: (**a**) input image; (**b**) ground truth; (**c**) Base; (**d**) EV; (**e**) EV + MP; and (**f**) all.

Table 6. Quantitative evaluation of method contributions on Atlantic Forest dataset. (✓denotes the incorporation of this module.)

Ratio	Method				Validation					Test				
	Base	MP	EV	AL	mIoU	Accuracy	mPrecision	mRecall	mF1-Score	mIoU	Accuracy	mPrecision	mRecall	mF1-Score
1/16	✓				68.87	77.70	80.68	82.37	81.33	69.59	78.28	80.97	82.90	81.76
	✓	✓			70.35	80.79	83.54	83.33	82.57	70.02	79.89	81.49	82.72	82.04
	✓	✓	✓		72.34	80.61	83.12	84.68	83.76	71.15	80.61	82.38	83.35	82.83
	✓	✓	✓	✓	**72.93**	**81.48**	**83.58**	**84.92**	**84.15**	**71.65**	**82.38**	**82.94**	**83.41**	**83.16**
1/8	✓				72.02	**83.50**	83.49	83.48	83.49	70.14	81.12	81.82	82.39	82.09
	✓	✓			72.31	81.10	83.54	83.86	83.69	70.17	80.32	81.66	82.70	82.13
	✓	✓	✓		72.81	81.79	83.56	84.72	84.07	72.91	84.07	84.04	84.01	84.02
	✓	✓	✓	✓	**73.74**	82.67	**84.23**	**85.27**	**84.69**	**73.61**	**85.40**	**84.79**	**84.22**	**84.49**
1/4	✓				70.33	81.91	82.21	82.41	82.31	69.86	82.39	82.02	81.70	81.86
	✓	✓			72.77	83.06	83.78	84.28	**89.02**	71.87	82.81	83.16	83.47	83.31
	✓	✓	✓		73.60	82.22	84.08	**85.30**	84.61	73.85	84.84	84.73	**84.62**	84.67
	✓	✓	✓	✓	**74.07**	**83.96**	**84.66**	85.16	84.90	**74.51**	**86.75**	**85.67**	84.61	**85.10**
1/2	✓				71.87	77.61	82.67	85.32	83.48	74.17	81.73	84.15	86.04	84.95
	✓	✓			73.52	80.20	83.80	**85.91**	84.58	75.33	**84.18**	85.27	86.20	85.70
	✓	✓	✓		73.35	78.95	83.65	86.22	84.48	75.23	84.12	85.20	86.13	85.63
	✓	✓	✓	✓	**73.95**	**81.66**	**84.20**	85.80	**84.85**	**75.62**	84.11	**85.39**	**86.49**	**85.90**

Table 7. Quantitative evaluation of method contributions on Amazon Forest dataset. (✓denotes the incorporation of this module.)

Ratio	Method				Validation					Test				
	Base	MP	EV	AL	mIoU	Accuracy	mPrecision	mRecall	mF1-Score	mIoU	Accuracy	mPrecision	mRecall	mF1-Score
1/16	✓				87.04	92.77	93.08	93.05	93.06	89.29	95.22	94.35	94.34	94.34
	✓	✓			87.61	93.00	93.41	93.37	93.39	89.84	95.17	94.64	94.64	94.64
	✓	✓	✓		91.60	95.99	95.60	95.63	95.61	92.27	96.46	95.98	95.98	95.98
	✓	✓	✓	✓	**91.70**	**96.73**	**95.79**	**95.85**	**95.81**	**92.64**	**96.76**	**96.18**	**96.18**	**96.18**
1/8	✓				88.73	94.86	94.00	94.06	94.02	90.09	96.06	94.81	94.79	94.78
	✓	✓			89.14	94.40	94.25	94.26	94.25	90.81	96.03	95.19	95.18	95.18
	✓	✓	✓		91.90	95.94	95.77	95.78	95.77	92.60	96.27	96.15	96.15	96.15
	✓	✓	✓	✓	**92.02**	**96.78**	**95.82**	**95.88**	**95.84**	**92.82**	**96.95**	**96.28**	**96.27**	**96.27**
1/4	✓				88.69	95.03	93.98	94.04	94.00	88.61	94.84	93.97	93.96	93.96
	✓	✓			89.95	95.71	94.68	94.75	94.70	91.42	**96.79**	95.54	95.52	95.51
	✓	✓	✓		92.14	96.09	95.90	95.91	95.91	92.85	96.58	96.29	96.29	96.29
	✓	✓	✓	✓	**92.48**	**96.45**	**96.08**	**96.10**	**96.09**	**93.02**	96.62	**96.38**	**96.38**	**96.38**
1/2	✓				89.96	97.88	94.78	94.84	94.71	90.20	**98.04**	95.03	94.86	94.84
	✓	✓			90.69	**97.89**	95.15	95.23	95.11	91.06	97.97	95.44	95.33	95.32
	✓	✓	✓		91.52	97.62	95.56	95.65	95.57	91.97	97.56	95.87	95.82	95.81
	✓	✓	✓	✓	**91.71**	97.28	**95.66**	**95.74**	**95.67**	**92.12**	97.15	**95.92**	**95.90**	**95.89**

5. Conclusions

In this study, we propose the integration of the self-training paradigm with the mean-teacher paradigm based on multiple perturbations and EV. We apply consistency regularisation and self-training pseudo-labelling techniques in the forest-cover-mapping scenario. By introducing auxiliary teachers and establishing more detailed pseudo-label generation strategies, as well as enhancing the perturbations in our model, we demonstrate excellent segmentation results. Our method demonstrates robust performance on datasets exhibiting both class imbalance and class balance. This is accomplished through the dynamic adjustment of the loss function and straightforward hyperparameter settings. Furthermore, extensive ablation experiments conducted on the validation and test sets confirm the effectiveness and robustness of our proposed method. Finally, considering the performance of

our method across various metrics and other objectives in RS-image-processing tasks, we have identified potential avenues for further research and development.

Although our proposed method has achieved promising results, we believe that there are several directions for further research in the future:

- In the pseudo-label generation strategy based on TTA and EV, we used averaging to generate pseudo-labels with higher confidence using multi-model votes. In the future, adaptive weighting generation methods will be explored.
- The adaptive loss is set for linear adjustment, and thus, the non-linear setting will be investigated.
- In view of model limitations, the easy accessibility of RGB remote sensing images, and equipment cost, this study only explored the RGB channel combination. In future work, more effective channel fusion methods can be investigated, or alternative approaches, like using the NIR band instead of the blue band, can be explored for forest cover mapping.

Author Contributions: Methodology, B.C., L.W. and X.F.; Validation, W.B.; Investigation, X.Y.; Data curation, L.W.; Writing—original draft, B.C.; Writing—review & editing, X.F. and T.T.; Visualization, X.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Key Research and Development Program (NO. 2022YFD2201005) and Natural Science Foundation of China (NO.61902187).

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Van Mantgem, P.J.; Stephenson, N.L.; Byrne, J.C.; Daniels, L.D.; Franklin, J.F.; Fulé, P.Z.; Harmon, M.E.; Larson, A.J.; Smith, J.M.; Taylor, A.H.; et al. Widespread increase of tree mortality rates in the western United States. *Science* **2009**, *323*, 521–524.
2. Zhu, Y.; Li, D.; Fan, J.; Zhang, H.; Eichhorn, M.P.; Wang, X.; Yun, T. A reinterpretation of the gap fraction of tree crowns from the perspectives of computer graphics and porous media theory. *Front. Plant Sci.* **2023**, *14*, 1109443.
3. Boers, N.; Marwan, N.; Barbosa, H.M.; Kurths, J. A deforestation-induced tipping point for the South American monsoon system. *Sci. Rep.* **2017**, *7*, 41489.
4. Li, X.; Wang, X.; Gao, Y.; Wu, J.; Cheng, R.; Ren, D.; Bao, Q.; Yun, T.; Wu, Z.; Xie, G.; et al. Comparison of Different Important Predictors and Models for Estimating Large-Scale Biomass of Rubber Plantations in Hainan Island, China. *Remote Sens.* **2023**, *15*, 3447.
5. Lewis, S.L.; Lopez-Gonzalez, G.; Sonké, B.; Affum-Baffoe, K.; Baker, T.R.; Ojo, L.O.; Phillips, O.L.; Reitsma, J.M.; White, L.; Comiskey, J.A.; et al. Increasing carbon storage in intact African tropical forests. *Nature* **2009**, *457*, 1003–1006.
6. Hansen, M.C.; Stehman, S.V.; Potapov, P.V.; Loveland, T.R.; Townshend, J.R.; DeFries, R.S.; Pittman, K.W.; Arunarwati, B.; Stolle, F.; Steininger, M.K.; et al. Humid tropical forest clearing from 2000 to 2005 quantified by using multitemporal and multiresolution remotely sensed data. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 9439–9444.
7. Sexton, J.O.; Song, X.P.; Feng, M.; Noojipady, P.; Anand, A.; Huang, C.; Kim, D.H.; Collins, K.M.; Channan, S.; DiMiceli, C.; et al. Global, 30-m resolution continuous fields of tree cover: Landsat-based rescaling of MODIS vegetation continuous fields with lidar-based estimates of error. *Int. J. Digit. Earth* **2013**, *6*, 427–448.
8. Hamunyela, E.; Reiche, J.; Verbesselt, J.; Herold, M. Using space-time features to improve detection of forest disturbances from Landsat time series. *Remote Sens.* **2017**, *9*, 515.
9. Yin, H.; Khamzina, A.; Pflugmacher, D.; Martius, C. Forest cover mapping in post-Soviet Central Asia using multi-resolution remote sensing imagery. *Sci. Rep.* **2017**, *7*, 1375.
10. Zhang, P.; Ke, Y.; Zhang, Z.; Wang, M.; Li, P.; Zhang, S. Urban land use and land cover classification using novel deep learning models based on high spatial resolution satellite imagery. *Sensors* **2018**, *18*, 3717.
11. Zhang, Y.; Li, X.; Ling, F.; Atkinson, P.M.; Ge, Y.; Shi, L.; Du, Y. Updating Landsat-based forest cover maps with MODIS images using multiscale spectral-spatial-temporal superresolution mapping. *Int. J. Appl. Earth Obs. Geoinf.* **2017**, *63*, 129–142.
12. Flores, E.; Zortea, M.; Scharcanski, J. Dictionaries of deep features for land-use scene classification of very high spatial resolution images. *Pattern Recognit.* **2019**, *89*, 32–44.
13. Alzubaidi, L.; Zhang, J.; Humaidi, A.J.; Al-Dujaili, A.; Farhan, L. Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *J. Big Data* **2021**, *8*, 53.

14. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, 5–9 October 2015; Springer: Berlin/Heidelberg, Germany, 2015; pp. 234–241.
15. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.
16. Chen, L.C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H., Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In *Computer Vision—ECCV 2018*; Lecture Notes in Computer Science; Springer International Publishing: Cham, Germany, 2018; pp. 833–851. https://doi.org/10.1007/978-3-030-01234-2_49.
17. Bragagnolo, L.; da Silva, R.; Grzybowski, J. Amazon forest cover change mapping based on semantic segmentation by U-Nets. *Ecol. Inform.* **2021**, *62*, 101279. https://doi.org/10.1016/j.ecoinf.2021.101279.
18. Flood, N.; Watson, F.; Collett, L. Using a U-net convolutional neural network to map woody vegetation extent from high resolution satellite imagery across Queensland, Australia. *Int. J. Appl. Earth Obs. Geoinf.* **2019**, *82*, 101897.
19. Isaienkov, K.; Yushchuk, M.; Khramtsov, V.; Seliverstov, O. Deep Learning for Regular Change Detection in Ukrainian Forest Ecosystem With Sentinel-2. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 364–376. https://doi.org/10.1109/JSTARS.2020.3034186.
20. Papandreou, G.; Chen, L.C.; Murphy, K.P.; Yuille, A.L. Weakly-and semi-supervised learning of a deep convolutional network for semantic image segmentation. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1742–1750.
21. Peláez-Vegas, A.; Mesejo, P.; Luengo, J. A Survey on Semi-Supervised Semantic Segmentation. *arXiv* **2023**, arXiv:cs.CV/2302.09899.
22. Van Engelen, J.E.; Hoos, H.H. A survey on semi-supervised learning. *Mach. Learn.* **2020**, *109*, 373–440.
23. Lucas, B.; Pelletier, C.; Schmidt, D.; Webb, G.I.; Petitjean, F. A bayesian-inspired, deep learning-based, semi-supervised domain adaptation technique for land cover mapping. *Mach. Learn.* **2021**, *112*, 1941–1973.
24. Hong, D.; Yokoya, N.; Ge, N.; Chanussot, J.; Zhu, X.X. Learnable manifold alignment (LeMA): A semi-supervised cross-modality learning framework for land cover and land use classification. *ISPRS J. Photogramm. Remote Sens.* **2019**, *147*, 193–205.
25. Ouali, Y.; Hudelot, C.; Tami, M. Semi-supervised semantic segmentation with cross-consistency training. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 12674–12684.
26. Liu, Y.; Tian, Y.; Chen, Y.; Liu, F.; Belagiannis, V.; Carneiro, G. Perturbed and strict mean teachers for semi-supervised semantic segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 4258–4267.
27. Abuduweili, A.; Li, X.; Shi, H.; Xu, C.Z.; Dou, D. Adaptive Consistency Regularization for Semi-Supervised Transfer Learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; pp. 6923–6932.
28. Zoph, B.; Ghiasi, G.; Lin, T.Y.; Cui, Y.; Liu, H.; Cubuk, E.D.; Le, Q. Rethinking Pre-training and Self-training. In Proceedings of the Advances in Neural Information Processing Systems, Online, 6–12 December 2020; Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., Lin, H., Eds.; Curran Associates, Inc.: San Francisco, United States, 2020; Volume 33, pp. 3833–3845.
29. Zhu, Y.; Zhang, Z.; Wu, C.; Zhang, Z.; He, T.; Zhang, H.; Manmatha, R.; Li, M.; Smola, A.J. Improving Semantic Segmentation via Efficient Self-Training. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, early access, 1. https://doi.org/10.1109/TPAMI.2021.3138337.
30. SemiRoadExNet: A semi-supervised network for road extraction from remote sensing imagery via adversarial learning. *ISPRS J. Photogramm. Remote Sens.* **2023**, *198*, 169–183. https://doi.org/10.1016/j.isprsjprs.2023.03.012.
31. Zou, Y.; Zhang, Z.; Zhang, H.; Li, C.L.; Bian, X.; Huang, J.B.; Pfister, T. Pseudoseg: Designing pseudo labels for semantic segmentation. *arXiv* **2020**, arXiv:2010.09713.
32. Zhang, B.; Zhang, Y.; Li, Y.; Wan, Y.; Guo, H.; Zheng, Z.; Yang, K. Semi-supervised Deep Learning via Transformation Consistency Regularization for Remote Sensing Image Semantic Segmentation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *16*, 5782–5796. https://doi.org/10.1109/JSTARS.2022.3203750.
33. Bragagnolo, L.; da Silva, R.V.; Grzybowski, J.M.V. Amazon and Atlantic Forest image datasets for semantic segmentation. **2021**. https://doi.org/10.5281/zenodo.4498086.
34. John, D.; Zhang, C. An attention-based U-Net for detecting deforestation within satellite sensor imagery. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *107*, 102685.
35. Chen, X.; Fan, H.; Girshick, R.; He, K. Improved baselines with momentum contrastive learning. *arXiv* **2020**, arXiv:2003.04297.
36. Hu, H.; Wei, F.; Hu, H.; Ye, Q.; Cui, J.; Wang, L. Semi-supervised semantic segmentation via adaptive equalization learning. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 22106–22118.
37. Chen, X.; Yuan, Y.; Zeng, G.; Wang, J. Semi-supervised semantic segmentation with cross pseudo supervision. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2021; pp. 2613–2622.

38. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *40*, 834–848.
39. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 770–778.