Debiased Learning from Naturally Imbalanced Remote Sensing Data

Chun-Hsiao Yeh[1] Xudong Wang[1] Stella X. Yu[1,2] Charles Hill[3] Zackery Steck[3] Scott Kangas[3]
[1]University of California, Berkeley [2]University of Michigan [3]Etegent Technologies, Ltd.
{daniel_yeh,xdwang}@berkeley.edu stellayu@umich.edu
{charles.hill,zackery.steck,scott.kangas}@etegent.com

Abstract

Deep learning has shown remarkable success in analyzing grounded imagery, such as consumer photos due to large-scale human annotations are available for dataset, e.g., ImageNet. However, such extensive supervision is not the case for remote sensing data.

We propose a highly effective semi-supervised approach tailored specifically for remote sensing data. Our approach encompasses two key contributions. We adapt the framework from semi-supervised learning approach, such as FixMatch, to remote sensing data by designing a set of robust strong and weak augmentations suitable for this domain. By learning from actual labeled data, combining with pseudo-labeled data, yet address the pseudo-labeling imbalance, we leverage a recently proposed debiased learning approach to mitigate the bias in pseudo-labeling.

Validated by extensive experimentation, our simple semi-supervised framework with 30% annotations delivers significant accuracy gains over the supervised learning baseline by 7.1%, and over recent supervised state-of-the-art, CDS by 2.1% on remote sensing Xview data.

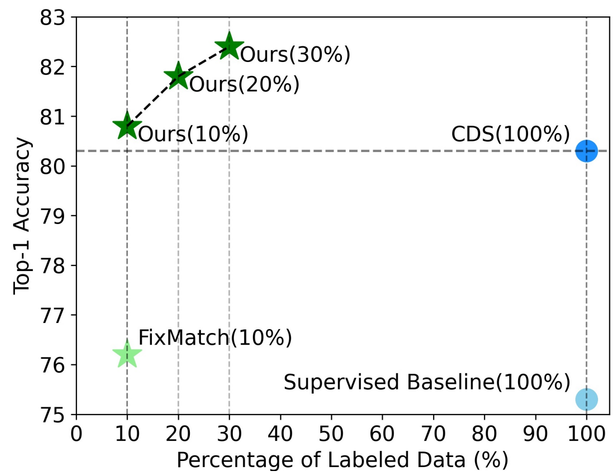

Figure 1. **Our semi-supervsed learning approach outperforms FixMatch [30], supervised baseline, and CDS [29] on remote sensing data.** By effectively addressing pseudo-labeling imbalance using debiased learning, we surpass FixMatch by 4.6% accuracy with 10% labeled data in the semi-supervised setting. Furthermore, our method showcases enhanced performance with increased labeled data, outshining the supervised learning baseline by 7.1% and surpassing the recent state-of-the-art, CDS, by 2.1% on remote sensing xView [18] data. These findings underscore the remarkable effectiveness of our approach in handling remote sensing data, leading to substantial performance improvements compared to existing methods.

1. Introduction

Deep learning has enjoyed remarkable success in analyzing natural images, primarily due to the extensive human annotation available in datasets such as ImageNet [17]. However, the scenario changes significantly when it comes to remote sensing data [18, 31, 39]. While there is an abundance of remote sensing data available, there are limited annotations due to the challenges in annotating such data. The scarcity of annotations is primarily attributable to the uncommon viewing angles and the ambiguity present in the remote sensing data, making it difficult for humans to perform annotation accurately.

Previous works, such as transfer learning approaches, have shown significant success in handling specific data domains by starting with a model pretrained on ImageNet [17] and fine-tuning it with a few samples of the target domain. These approaches have been highly effective in medical imaging [27] and self-driving [16] tasks. However, unlike these image domains, remote sensing data is characterized by a lack of high-resolution details and class-imbalanced properties, making it challenging to transfer from a pretrained model directly [10, 26, 33].

Additionally, several self-supervised learning methods [6,12,14,38] have been proposed, which can learn a representation without human-annotated data. However, learning self-supervised models without pretraining is expensive and time-consuming. Moreover, the ability to discriminate positive and negative pairs can be worse on remote sensing data than on natural images due to the ambiguity in the data.

In remote sensing scenarios with limited annotated data

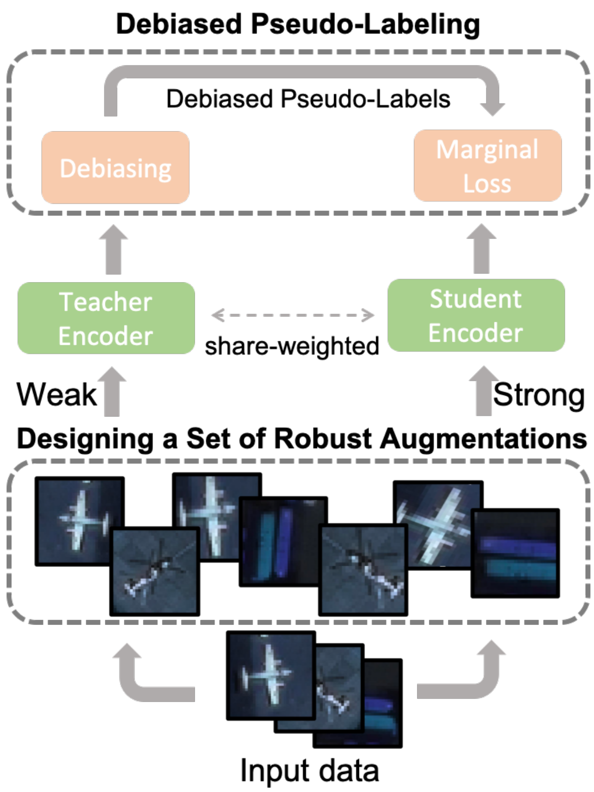

Figure 2. **Our semi-supervised learning framework.** The framework consists of two key components. The first component focuses on the design of robust strong and weak augmentations specifically tailored for remote sensing data. These augmentations are carefully crafted to enhance the performance of the framework. The second component leverages a state-of-the-art debiased learning approach [36] to effectively address the bias present in pseudo-labeling. By mitigating this bias, the framework achieves improved accuracy and reliability in handling remote sensing data.

and abundant unlabeled data, semi-supervised learning techniques hold significant promise. FixMatch [30] is a popular approach that combines weak and strong augmentations with pseudo-labeling to enhance the learning process. However, to adapt these techniques to remote sensing domains, we focus on the design of robust weak and strong augmentations specifically tailored for this context. We investigate the benefits of incorporating rotation, scaling, and horizontal flipping, aiming to improve the analysis and interpretation of remote sensing data. By thoughtfully integrating these augmentation strategies, we unlock the full potential of semi-supervised learning, enabling more effective and context-aware analysis techniques in remote sensing.

Furthermore, some previous studies have shown that machine-generated pseudo-labels often suffer from inherent imbalances [36]. This imbalance poses a challenge for learning models as it introduces a bias towards false majorities within the pseudo-labels. Similar issues have been observed in widely used datasets like CIFAR and ImageNet [36]. To address this bias problem specifically in the context of remote sensing data, we incorporate the DebiasPL [36] method, which aims to alleviate the bias associated with pseudo-labeling. In Figure 2, we illustrate the framework of our approach, and we summarize our key contributions in this work as follows:

1. We adapt the framework of semi-supervised learning approaches, such as FixMatch, to remote sensing data by designing a set of robust strong and weak augmentations specifically tailored to this domain.

2. We leverage the recently proposed DebiasPL method to mitigate the bias in pseudo-labeling by combining actual labeled data with pseudo-labeled data and addressing the imbalance in pseudo-labels.

3. We conduct extensive experiments that demonstrate the effectiveness of our approach, Validated through extensive experimentation, our simple semi-supervised framework with 30% annotations achieves significant accuracy gains over the supervised learning baseline by 7.1% and outperforms the recent supervised state-of-the-art, CDS, by 2.1% on remote sensing Xview data. This is illustrated in Figure 1.

2. Related Works

Semi-supervised learning (SSL) learns over both limited labeled data and relatively larger unlabeled data. A large portion of SSL approaches [2, 3, 19, 20] follow the self-training scheme that generates pseudo-labels to the unlabeled data based on the model learned from limited labeled data. FixMatch [30] learns to predict pseudo-labels using weakly- and strongly-augmented versions of an image and then matches both predictions via the cross-entropy loss. There are other lines of work such as consistency regularization [22,32] that apply perturbations to affect the classification loss. Transfer learning [7] approaches first learn a supervised or self-supervised model, then fine-tune the model with a supervised classifier via limited labeled data. Though previous SSL approaches achieve considerable success on natural image data and remote sensing data [37], the focus of this work is to leverage insights of SSL to remote sensing image data to develop a robust augmentation pipeline, yielding better performance.

Imbalanced-class learning learns representations that are suitable for rare classes without significantly reducing performance on majority classes. In most scenarios, the class imbalanced problem exists in real-world data [13, 34], presenting a huge challenge to deep neural networks [1]. Previous works are composed of 1) re-balancing and re-weighting approaches [9, 23] that provide more weighting to the rare classes, 2) margin-based approaches [5] that aim to impose a large margin to rare classes and have shown to be effective for the generalization of minority classes, and 3) ensemble-based approaches [35] that learn multi-expert models across classes to mitigate the data bias and variance. Unlike all the aforementioned works that mainly focus on imbalanced human annotation, we apply an approach that aims to alleviate the bias of machine annotation (i.e., pseudo-labeling) in the training data.

3. Debiased Semi-Supervised Learning

Our work provides a comprehensive review of recent advancements in addressing the data imbalance in remote sensing images. Specifically, we first review the FixMatch [30] and semi-supervised learning. We then discuss the two sources of imbalance in remote sensing data and review two methods for addressing them: debiased learning [36] to address pseudo-label imbalance and logit adjustment [21] to address training label imbalance. Finally, we present our new findings on the most effective augmentations in remote sensing data. Overall, our paper provides valuable insights for researchers and practitioners working in the field of remote sensing data.

3.1. FixMatch

FixMatch [30] is a semi-supervised learning approach based on pseudo-labeling. The two branches of inputs are based on weak- and strong-augmented images to generate augmented samples for unlabeled data. Suppose we have been given a mini-batch of labeled data $\mathcal{X} = \left\{\left(\boldsymbol{x}_i^l, y_i^l\right); i \in \left(1, \ldots, B^l\right)\right\}$ and of unlabeled data $\mathcal{U} = \left\{\boldsymbol{x}_j^u; j \in \left(1, \ldots, B^u\right)\right\}$, and the loss function of labeled data $\mathcal{X}$ can be formed as:

$$\mathcal{L}_s = \frac{1}{B^l} \sum_{i=1}^{B^l} \mathcal{H}\left(y_i^l, f\left(\mathcal{W}\left(\boldsymbol{x}_i^l\right); \boldsymbol{\theta}\right)\right) \tag{1}$$

where $\mathcal{H}(\cdot,\cdot)$ is the cross-entropy function, and $\mathcal{W}(\cdot)$ is the weak augmentation. In order to deal with the unlabeled data $\mathcal{U}$, the widely used strategy is to generate the pseudo-label by obtaining the prediction via weak augmentation: $\hat{y}_j^u = \max\left(f\left(\mathcal{W}\left(\boldsymbol{x}_j^u\right); \boldsymbol{\theta}\right)\right)$. Based on the given loss and pseudo-labels, we could express the loss $\mathcal{L}_u$ for unlabeled data that computes the distance between the one-hot pseudo-label and model predictions on the strongly-augmented image $\mathcal{S}\left(\boldsymbol{x}_j^u\right)$ as:

$$\mathcal{L}_u = \frac{1}{B^u} \sum_{j=1}^{B^u} \mathbb{1}\left(\hat{y}_j^u \geq \tau\right) \mathcal{H}\left(\hat{y}_j^u, f\left(\mathcal{S}\left(\boldsymbol{x}_j^u\right); \boldsymbol{\theta}\right)\right) \tag{2}$$

The objective function of FixMatch [30] consists of two components: $\mathcal{L} = \mathcal{L}_s + \lambda\mathcal{L}_u$. Here, $\mathcal{L}_s$ represents the loss term computed using labeled data, while $\mathcal{L}_u$ represents the loss term computed using unlabeled data. The scalar hyperparameter λ controls the relative importance of the two terms in the overall loss function.

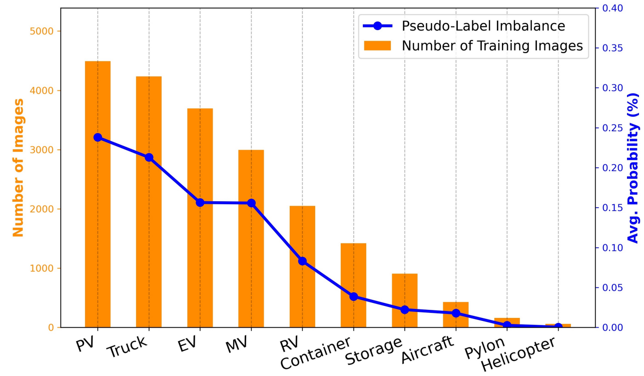

Figure 3. **Remote sensing data, such as xView [18], has two sources of imbalance.** One is training label imbalance provided by humans, the other is pseudo-label imbalance generated by semi-supervised learning framework during learning on remote sensing data. Note that the average probability distributions of FixMatch [30] are computed by averaging over all unlabeled data. The class indices are then sorted based on their corresponding average probabilities.

3.2. Sources of Imbalance in Remote Sensing Data

One challenge in using semi-supervised learning approaches based on pseudo-labeling is to deal with the imbalanced nature of both 1) the training data and 2) the pseudo-labels. The debiased learning approach [36] illustrates that imbalances exist not only in the human-annotated labels but also in the machine-generated labels, i.e., pseudo-labels.

In order to gain a deeper understanding of this issue, we analyze the distribution (Figure 3) of imbalanced training data and pseudo-labels generated by the FixMatch [30] semi-supervised learning approach on xView [18], a remote sensing dataset. Our investigation reveals that these imbalances in the data sources can introduce notable biases during the learning process, ultimately affecting the model's performance. Thus, our work aims to

1. apply the debiased learning method [36] to alleviate the imbalanced **pseudo-labels**.
2. incorporate logit adjustment [21] to mitigate the imbalanced **training labels** in the remote sensing data.

3.3. Debiased Learning (Pseudo-Labels)

Our approach applies a proposed debiased learning method, DebiasPL [36] on the semi-supervised learning task. The framework of DebiasPL is mainly based on FixMatch, and further embedded with an adaptive debiasing module and a marginal loss.

Adaptive debiasing module. DebiasPL is motivated by Causal Inference [11, 24, 28] that has been shown to be effective in mitigating the selection bias in several tasks. The DebiasPL aims to integrate causality of producing debiased predictions via counterfactual reasoning. Based on [11, 25], the debiased pseudo-labeling could be performed via the debiased logit with counterfactual reasoning:

$$\tilde{f}_i = f\left(\mathcal{W}\left(x_i\right)\right) - \lambda \log \hat{p} \tag{3}$$

$$\hat{p} \leftarrow m\hat{p} + (1-m)\frac{1}{\mu B}\sum_{k=1}^{\mu B} p_k \tag{4}$$

where m is the coefficient of momentum, $f\left(\mathcal{W}\left(\cdot\right)\right)$ refers to logits of weakly-augmented unlabeled instance.

Adaptive marginal loss. Motivated by the issue that the biases in pseudo-labels usually come from inter-class confusion, the DebiasPL aims to have a larger margin between highly biased classes by designing an adaptive marginal loss $\mathcal{L}_{\text{margin}}$ to alleviate the inter-class confusion. The marginal loss can be expressed as:

$$\mathcal{L}_{\text{margin}} = -\log \frac{e^{\left(z_{\hat{y}_i} - \Delta_{\hat{y}_i}\right)}}{e^{\left(z_{\hat{y}_i} - \Delta_{\hat{y}_i}\right)} + \sum_{k \neq \hat{y}_i}^{N} e^{\left(z_k - \Delta_k\right)}} \tag{5}$$

where $\Delta_j = \lambda \log\left(\frac{1}{\hat{p}_j}\right)$ for $j \in \{1, \ldots, N\}, z = f\left(\beta\left(x_i\right)\right)$. The $\mathcal{H}\left(\hat{y}_j^u, f\left(\mathcal{S}\left(\boldsymbol{x}_j^u\right); \boldsymbol{\theta}\right)\right)$ could be replaced as $\mathcal{L}_{\text{margin}}$, the final unsupervised loss could be updated by Eq. (2) with Eq. (3) and Eq. (5).

3.4. Logit Adjustment (Training Labels)

The performance of each majority and minority class can often be biased due to imbalanced training data. This is because the classifier may not have enough examples to learn how to distinguish the minority class from the majority class. [29] shows that the imbalanced remote sensing data also suffer from this bias, leading to poor class-wise performance. To address this issue, we aim to apply logit adjustment [21], a technique that can produce a more balanced class-wise performance during evaluation. Logit adjustment adjusts the logits of each class based on their frequency in the training data. Specifically, the adjusted logits are computed as follows:

$$\hat{y}_i = \frac{\log(\frac{p_i}{1-p_i})}{\sum_j \log(\frac{p_j}{1-p_j})} \tag{6}$$

where $\hat{y}_i$ is the adjusted output for class i, p_i is the frequency of class i in the training data, and the sum in the denominator ensures that the adjusted logits sum to 1. By applying logit adjustment, we expect to achieve more balanced performance across all classes in remote sensing data.

3.5. Our New Findings on Remote Sensing Data

Remote sensing data possesses unique properties that set it apart from natural images, such as pixel-level data and non-uniform ground sampling distance (GSD) [18]. Despite the impressive success of debiased learning in natural images [36], it remains uncertain whether these techniques can be effectively applied to remote sensing data. Furthermore, it is unclear

1. which types of data augmentation are most suitable for remote sensing data, and
2. what modifications may be necessary for existing techniques to achieve optimal performance.

These questions have yet to be fully addressed in debiased learning approach [36] or other related works [29, 30]. Our research aims to bridge these knowledge gaps by exploring the efficacy of debiased learning techniques in remote sensing data and identifying appropriate modifications and practices to achieve superior performance.

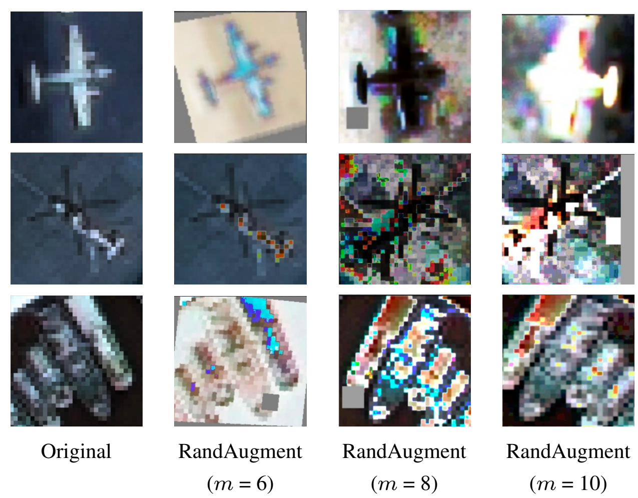

Figure 4. ***RandAugment* [8] appears to be too complicated for remote sensing data [18] with artifacts.** A series of random transformations are applied to orginal images (Column 1). By increasing the transformation magnitude in terms of parameter m, which controls transformation intensity, we can observe more pronounced and visually diverse augmentations, such as artifacts at the border of image.

Data augmentations for remote sensing datasets. In semi-supervised learning (SSL) settings, the debiased learning method [36] has been shown to be effective, using *random resize cropping* and *horizontal flipping* for weak augmentation and *RandAugment* [8] for strong augmentation on natural images. However, these augmentation techniques do not generalize well to remote sensing data due to the distinct properties of the latter, such as a lack of details in the imagery. In Figure 4, we illustrate examples by applying *RandAugment* [8] with different intensities of the transformations. *RandAugment* appears to be too complicated for remote sensing data. To address this issue, we empirically select appropriate augmentations that can improve the SSL performance on remote sensing data. Our approach includes *random resize cropping* and *horizontal flipping* for weak augmentation, and *horizontal flipping, rotation, and scaling* for strong augmentation.

Table 1 shows the augmentations that were selected for use with remote sensing data, along with the reasons for their selection. *Rotation* and *scaling* were both used to simulate variations in imaging angles and ground sampling distances, respectively, which are important factors in generalizing to remote sensing data. Additionally, horizontal flipping was included to simulate mirror-reflected scenes. *RandAugment* was not used due to potential overfitting to the training set of remote sensing data [39].

Table 1. Selected augmentations for remote sensing data, Xview [18]. The table lists the augmentations used for remote sensing data, including whether or not they were used and the reason for their selection. The augmentations include *rotation*, *scaling*, and *horizontal flipping*. The use of *RandAugment* is discussed and not used due to potential overfitting.

Augmentations	**Used**	**Reason**
RandAugment [8]	×	The use of RandAugment with high intensity of transformations may result in overfitting to the training set, which can be detrimental for generalization on remote sensing data.
Rotation	✓	random rotation (± 10 degrees) can simulate variations in remote sensing imaging angles.
Scaling	✓	scaling(0.8, 1.2) can simulate variations in ground sampling distances, which is important for generalization on remote sensing data.
Horizontal Flip	✓	horizontal flipping simulates mirror-reflected scenes in remote sensing data

The effectiveness of each augmentation method varies, and we present a table in Section 4 to illustrate the performance gain from each augmentation strategy we leveraged. Notably, the augmentations used in previous SSL works, such as debiased learning, may not always be beneficial to SSL settings in remote sensing images.

4. Experiments

We assess the effectiveness of our semi-supervised learning approach, on the imbalanced classification task using the xView remote sensing dataset by 1) comparing the instance-wise accuracy and class-wise accuracy of our model and baselines; 2) conducting several ablation studies on data augmentation, data preprocessing, and the learning recipe for optimizing the proposed method on remote sensing data. We note all experiments are carried out on a single machine equipped with four Nvidia RTX 2080 Ti GPUs.

4.1. Experimental Setup

Dataset. We conducted a thorough evaluation of the performance of our semi-supervised learning approach on the xView [18] dataset, which has large-scale multi-spectral images featuring 8-band channels obtained from satellite data.

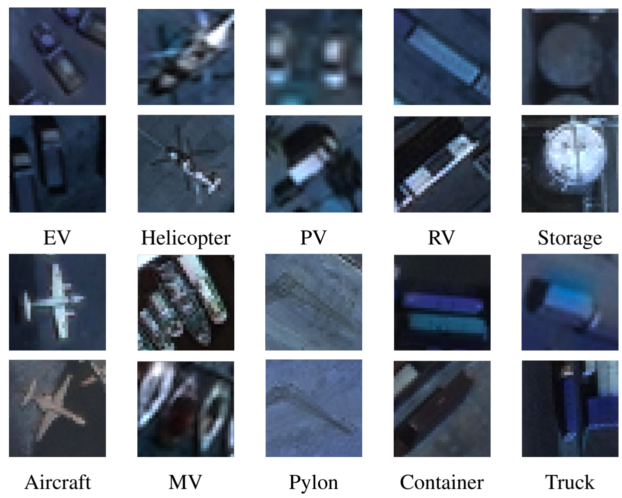

Figure 5. **xView data overview.** The data contains ten individual classes and has the properties of a highly imbalanced distribution, which is also shown in Figure 3.

To ensure consistency with previous studies [29], we preprocessed the data by selecting a subset of 10 categories out of 60 classes, namely *StorageTank, Helicopter, Pylon, Maritime Vessel (MV), ShippingContainer, Fixed-Wing Aircraft (FWAircraft), Passenger Vehicle (PV), Truck, Railway Vehicle (RV), and Engineering Vehicle (EV)*. These categories are illustrated in Figure 5. The dataset consisted of 20431 training samples, 2270 validation samples, and 63279 test samples. Notably, we selected the R, G, B bands from the 8-band channels to form RGB images.

Baseline methods. As our baseline methods, we employed FixMatch [30] for its straightforward implementation and proven performance in semi-supervised learning. In addition, we included a ResNet-18 supervised learning baseline and the recently proposed complex-valued model, CDS [29], to demonstrate the effectiveness of our semi-supervised learning approach.

Training and evaluation. In the semi-supervised learning setup, we employed ResNet-50 [15] as the backbone and adopted the same set of hyperparameters used in FixMatch

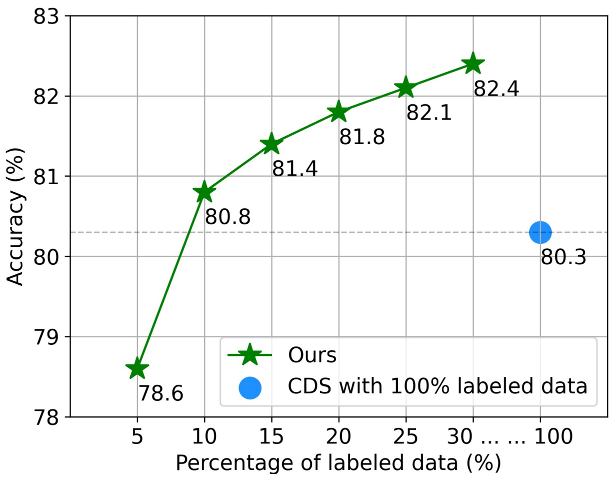

Figure 6. **Improve the performance of our semi-supervised learning framework by increasing the amount of labeled data.** The instance-wise accuracies of our method are shown with 5% to 30% of labeled data. We improve the performance and outperform the 100% labeled data trained state-of-the-art, CDS [29] by including more labeled data.

to ensure a fair comparison. We initialized the model using the pretrained ImageNet model that was trained with MoCo v2 + EMAN [4] for 800 epochs, before fine-tuning on the xView dataset for another 150 epochs. As for the supervised learning baselines and CDS [29], we utilized ResNet-18 as the backbone and did not use any data augmentation during the training phase, following their respective learning recipes. In terms of evaluation, we measured the top-1 accuracy of our models using two metrics:

1. **instance-wise accuracy:** the ratio of correctly classified instances (images) to the total number of data;
2. **class-wise accuracy:** the average accuracy of each individual class, to assess the models' generalization ability to imbalanced classes.

4.2. Instance-wise Comparisons

Table 2 shows a comparison of our approach against semi-supervised and supervised learning baselines. In our initial setup, we had our framework learn on 10% labeled data with a ResNet-50 backbone pretrained on ImageNet. As shown in Table 1, the our approach achieved 80.8% accuracy, outperforming the semi-supervised learning baseline, FixMatch, by 4.6%, and the supervised learning baseline by 5.5%, respectively.

Moreover, our method outperformed the best setting of the state-of-the-art complex-valued model for remote sensing, CDS [29], which was trained with 100% labeled data, by a 0.5% gain. In Figure 6, we demonstrate that the performance of our semi-supervised learning framework can be

Method	Pretrained	Labels (%)	Backbone	Augmented?	Top-1 (%)
Supervised [29]	×	100	RN-18	×	75.3
CDS [29]	×	100	RN-18	×	80.3
FixMatch [30]	✓	10	RN-50	✓	76.2
Ours	✓	10	RN-50	✓	**80.8**
Ours	✓	30	RN-50	✓	**82.4**

Table 2. Top-1 accuracy comparison (%) of our semi-supervised learning method, FixMatch, supervised learning baseline [29], and CDS [29] on remote sensing dataset, xView. our method consistently improves FixMatch and outperforms supervised learning baseline and CDS. Note that data augmentations in our framework include *horizontal flipping, random rotation, scaling*.

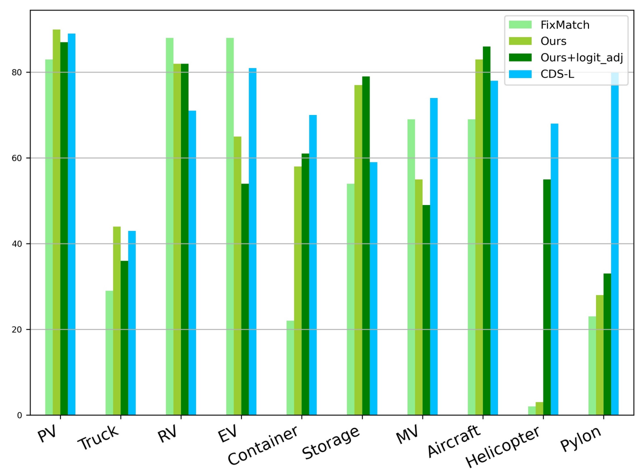

Figure 7. **Improve class-wise accuracy by including logit adjustment.** The class-wise accuracies are shown to demonstrate the results for each class, ordered from the majority class (PV) to the minority class (Pylon). To improve the performance, we incorporate the logit adjustment [21].

further improved by increasing the amount of labeled data, surpassing supervised CDS approach.

4.3. Class-wise Comparisons

Figure 7 displays the class-wise performance of the models, showcasing their ability to classify majority and minority classes. Moreover, by incorporating logit adjustment [21], the performance of our method for each class can be further boosted.

4.4. Ablation Studies

The ablation study focused on evaluating the effectiveness of selected augmentation strategies on the remote sensing dataset, xView [18]. We observed that the augmentation strategies in DebiasedPL were not directly applicable to this domain due to the complexity of *RandAugment* [8]. Therefore, we carefully selected and adapted appropriate augmentations to improve the performance of the model.

Table 3. Performance gain from selected augmentation strategy. We found that the *RandAugment* [8] strategy in DebiasedPL [36], which performed well on other image classification datasets, was too complex and not applicable to remote sensing data. Our findings involved carefully selecting and adapting augmentations to improve performance on the remote sensing dataset [18].

Weak augmentation						
None	✓	×	×	✓	×	×
Random Resize Cropping	×	✓	✓	×	✓	✓
Horizontal flipping	×	✓	✓	×	✓	✓
Strong augmentation						
ResizeCropping + Horizontal flipping	✓	✓	✓	✓	✓	✓
Rotation(± 10 degrees)	×	×	×	✓	✓	✓
Scaling(0.8, 1.2)	×	×	×	✓	×	✓
RandAugment(m=10)	✓	✓	×	×	×	×
RandAugment(m=5)	×	×	✓	×	×	×
Top-1 accuracy (%)	69.8	75.6	76.3	79.4	79.7	**80.8**

Table 3 summarizes the performance gain achieved through selected augmentations on xView. Our findings suggest that the adapted augmentations can effectively improve the performance of the model, achieving a top-1 accuracy of 80.8%, by 5.2% gain from the original DebiasPL augmentations. These results demonstrate the importance of selecting and adapting appropriate augmentations for remote sensing datasets, which can significantly improve the accuracy of the semi-supervised learning model.

5. Conclusion

We proposed a semi-supervised approach specifically designed for remote sensing data. Our paper addressed both training label and pseudo label imbalances in this domain. Our paper has two key contributions.

Firstly, we adapt the framework of FixMatch, to remote sensing data by designing robust strong and weak augmentations tailored for this context. Secondly, we leverage a recently proposed debiased learning approach to mitigate the bias in pseudo-labeling, effectively combining actual labeled data with pseudo-labeled data.

The results of our study highlight the significant potential of our straightforward semi-supervised framework, which effectively utilizes limited annotations (30%) to achieve notable performance enhancements. These findings contribute to the advancement of remote sensing data analysis and underscore the importance of developing tailored methodologies to tackle the challenge of limited annotations in this specific domain.

Acknowledgements. This research was supported, in part, by the National Geospatial Intelligence Agency / Etegent Technologies Ltd. contract HM047620C0063.

References

[1] Samy Bengio. Sharing representations for long tail computer vision problems. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, pages 1–1, 2015. 2

[2] 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. 2

[3] David Berthelot, Nicholas Carlini, Ian Goodfellow, Nicolas Papernot, Avital Oliver, and Colin A Raffel. Mixmatch: A holistic approach to semi-supervised learning. *Advances in neural information processing systems*, 32, 2019. 2

[4] Zhaowei Cai, Avinash Ravichandran, Subhransu Maji, Charless Fowlkes, Zhuowen Tu, and Stefano Soatto. Exponential moving average normalization for self-supervised and semi-supervised learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 194–203, 2021. 6

[5] 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. 3

[6] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pages 1597–1607. PMLR, 2020. 1

[7] Ting Chen, Simon Kornblith, Kevin Swersky, Mohammad Norouzi, and Geoffrey E Hinton. Big self-supervised models are strong semi-supervised learners. *Advances in neural information processing systems*, 33:22243–22255, 2020. 2

[8] Ekin D Cubuk, Barret Zoph, Jonathon Shlens, and Quoc V Le. Randaugment: Practical automated data augmentation with a reduced search space. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pages 702–703, 2020. 4, 5, 6, 7

[9] 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. 2

[10] Anthony Fuller, Koreen Millard, and James R Green. Transfer learning with pretrained remote sensing transformers. *arXiv preprint arXiv:2209.14969*, 2022. 1

[11] Sander Greenland, Judea Pearl, and James M Robins. Confounding and collapsibility in causal inference. *Statistical science*, 14(1):29–46, 1999. 4

[12] Jean-Bastien Grill, Florian Strub, Florent Altché, Corentin Tallec, Pierre Richemond, Elena Buchatskaya, Carl Doersch, Bernardo Avila Pires, Zhaohan Guo, Mohammad Gheshlaghi Azar, et al. Bootstrap your own latent-a new approach to self-supervised learning. *Advances in neural information processing systems*, 33:21271–21284, 2020. 1

[13] Agrim Gupta, Piotr Dollar, and Ross Girshick. Lvis: A dataset for large vocabulary instance segmentation. In *Pro-*

ceedings of the IEEE/CVF conference on computer vision and pattern recognition, pages 5356–5364, 2019. 2

[14] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9729–9738, 2020. 1

[15] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 5

[16] Jiman Kim and Chanjong Park. End-to-end ego lane estimation based on sequential transfer learning for self-driving cars. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 30–38, 2017. 1

[17] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84–90, 2017. 1

[18] Darius Lam, Richard Kuzma, Kevin McGee, Samuel Dooley, Michael Laielli, Matthew Klaric, Yaroslav Bulatov, and Brendan McCord. xview: Objects in context in overhead imagery. *arXiv preprint arXiv:1802.07856*, 2018. 1, 3, 4, 5, 6, 7

[19] 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. 2

[20] Junnan Li, Caiming Xiong, and Steven CH Hoi. Comatch: Semi-supervised learning with contrastive graph regularization. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9475–9484, 2021. 2

[21] Aditya Krishna Menon, Sadeep Jayasumana, Ankit Singh Rawat, Himanshu Jain, Andreas Veit, and Sanjiv Kumar. Long-tail learning via logit adjustment. *arXiv preprint arXiv:2007.07314*, 2020. 3, 4, 6

[22] Takeru Miyato, Shin-ichi Maeda, Masanori Koyama, and Shin Ishii. Virtual adversarial training: a regularization method for supervised and semi-supervised learning. *IEEE transactions on pattern analysis and machine intelligence*, 41(8):1979–1993, 2018. 2

[23] Seulki Park, Jongin Lim, Younghan Jeon, and Jin Young Choi. Influence-balanced loss for imbalanced visual classification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 735–744, 2021. 2

[24] Judea Pearl. Causal inference in statistics: An overview. *Statistics surveys*, 3:96–146, 2009. 4

[25] Judea Pearl. Direct and indirect effects. In *Probabilistic and Causal Inference: The Works of Judea Pearl*, pages 373–392. 2022. 4

[26] Rafael Pires de Lima and Kurt Marfurt. Convolutional neural network for remote-sensing scene classification: Transfer learning analysis. *Remote Sensing*, 12(1):86, 2019. 1

[27] Pranav Rajpurkar, Jeremy Irvin, Kaylie Zhu, Brandon Yang, Hershel Mehta, Tony Duan, Daisy Ding, Aarti Bagul, Curtis Langlotz, Katie Shpanskaya, et al. Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. *arXiv preprint arXiv:1711.05225*, 2017. 1

[28] Donald B Rubin. Essential concepts of causal inference: a remarkable history and an intriguing future. *Biostatistics & Epidemiology*, 3(1):140–155, 2019. 4

[29] Utkarsh Singhal, Stella X Yu, Zackery Steck, Scott Kangas, and Aaron A Reite. Multi-spectral image classification with ultra-lean complex-valued models. *arXiv preprint arXiv:2211.11797*, 2022. 1, 4, 5, 6

[30] Kihyuk Sohn, David Berthelot, Nicholas Carlini, Zizhao Zhang, Han Zhang, Colin A Raffel, Ekin Dogus Cubuk, Alexey Kurakin, and Chun-Liang Li. Fixmatch: Simplifying semi-supervised learning with consistency and confidence. *Advances in neural information processing systems*, 33:596–608, 2020. 1, 2, 3, 4, 5, 6

[31] Xian Sun, Peijin Wang, Zhiyuan Yan, Feng Xu, Ruiping Wang, Wenhui Diao, Jin Chen, Jihao Li, Yingchao Feng, Tao Xu, et al. Fair1m: A benchmark dataset for fine-grained object recognition in high-resolution remote sensing imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 184:116–130, 2022. 1

[32] Antti Tarvainen and Harri Valpola. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. *Advances in neural information processing systems*, 30, 2017. 2

[33] S Thirumaladevi, K Veera Swamy, and M Sailaja. Remote sensing image scene classification by transfer learning to augment the accuracy. *Measurement: Sensors*, 25:100645, 2023. 1

[34] Grant Van Horn, Oisin Mac Aodha, Yang Song, Yin Cui, Chen Sun, Alex Shepard, Hartwig Adam, Pietro Perona, and Serge Belongie. The inaturalist species classification and detection dataset. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8769–8778, 2018. 2

[35] Xudong Wang, Long Lian, Zhongqi Miao, Ziwei Liu, and Stella X Yu. Long-tailed recognition by routing diverse distribution-aware experts. *arXiv preprint arXiv:2010.01809*, 2020. 3

[36] Xudong Wang, Zhirong Wu, Long Lian, and Stella X Yu. Debiased learning from naturally imbalanced pseudo-labels. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14647–14657, 2022. 2, 3, 4, 7

[37] Yi Wang, Conrad M Albrecht, Nassim Ait Ali Braham, Lichao Mou, and Xiao Xiang Zhu. Self-supervised learning in remote sensing: A review. *arXiv preprint arXiv:2206.13188*, 2022. 2

[38] Zhirong Wu, Yuanjun Xiong, Stella X Yu, and Dahua Lin. Unsupervised feature learning via non-parametric instance discrimination. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3733–3742, 2018. 1

[39] Gui-Song Xia, Xiang Bai, Jian Ding, Zhen Zhu, Serge Belongie, Jiebo Luo, Mihai Datcu, Marcello Pelillo, and Liangpei Zhang. Dota: A large-scale dataset for object detection in aerial images. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3974–3983, 2018. 1, 5