ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume X-G-2025
ISPRS Geospatial Week 2025 "Photogrammetry & Remote Sensing for a Better Tomorrow…", 6–11 April 2025, Dubai, UAE

Unique Perturbation Methods Exploitation for Semi-Supervised Remote Sensing Image Semantic Segmentation

Liang Zhou, Keyi Duan, Jinkun Dai, Xiaodan Wu, Xuming Ge, Xiaojun Li, Yuanxin Ye*

Faculty of Geosciences and Engineering, Southwest Jiaotong University, Chengdu, Sichuan, China
(zlup, duankeyi, daijinkun, wuxd)@my.swjtu.edu.cn, xuming.ge@swjtu.edu.cn, kdxiaojun@126.com, yeyuanxin@home.swjtu.edu.cn

Keywords: Remote sensing semantic segmentation, Semi-supervised learning, Perturbation space expansion.

Abstract

Deep learning has significantly improved the accuracy of remote sensing semantic segmentation, yet its effectiveness is often constrained by the limited availability of annotated training samples. Semi-supervised learning (SSL) addresses this challenge by utilizing abundant unlabeled data, reducing dependence on manual annotations. However, current consistency regularization-based SSL methods, primarily developed for natural images, struggle to produce adequate perturbation diversity for robust model training in remote sensing image segmentation. In this work, we propose FusionMatch, a novel SSL framework featuring two perturbation mechanisms - NIRPerb and PSPerb - specifically designed for remote sensing imagery. NIRPerb utilizes near-infrared spectral data to enhance perturbation diversity. PSPerb adopts differentiated pan-sharpening fusion strategies to expand the perturbation space. Extensive experiments on both a building extraction dataset and a multi-class dataset demonstrate that FusionMatch outperforms state-of-the-art SSL methods in segmentation accuracy and robustness.

1. Introduction

Recent technological progress in remote sensing (RS) has dramatically improved data acquisition capabilities (Persello et al., 2022), propelling RS image analysis to the forefront of scientific inquiry. Within this domain, semantic segmentation plays a pivotal role by performing pixel-level classification of image content, which facilitates deeper interpretation of ground objects and landscapes.

Deep learning has revolutionized remote sensing semantic segmentation (RSSS) through its exceptional feature learning capabilities (Li et al., 2024). While these data-driven approaches excel at extracting discriminative features, their application to RS imagery faces notable challenges. The inherent complexity of RS data demands robust feature extraction, yet current models typically require extensive labeled datasets for optimal performance. Unlike natural image annotation, pixel-level labeling of RS imagery requires domain expertise to interpret complex spectral-spatial features across large datasets. This annotation process creates a critical bottleneck, significantly constraining the application of deep model in RSSS.

To address this challenge, semi-supervised learning (SSL), which bridges supervised learning and unsupervised learning, has gained prominence as a promising approach attracting growing research interest. The core aim of SSL is to enhance model performance under limited labeled data by fully utilizing abundant unlabeled samples. Current SSL methodologies primarily follow two technical pathways: (1) Self-training (Bachman et al., 2014; Teh et al., 2022), which generates pseudo-labels for unlabeled data to expand the training set, and (2) Consistency Regularization (Miyato et al., 2018; Chen et al., 2021a), which enhances prediction consistency against input, feature, or network perturbations to improve feature representation. The integration of these dual strategies has also emerged as a key research focus. FixMatch (Sohn et al., 2020) introduces a weak-to-strong consistency regularization (WSCR) framework, which first generates predictions using weakly perturbed data, then filters high-confidence pseudo-labels to maintain prediction consistency when applied to strongly perturbed versions of the same input samples. Subsequent studies (Yang et al., 2023a; Bai et al., 2024) have extended this framework to advance model segmentation capabilities.

The performance improvement of the WSCR framework is highly dependent on the effective construction of the perturbation space. Introducing diverse and challenging perturbations can encourage models to learn more robust feature representations. While established weak and strong perturbation techniques in natural image processing demonstrate significant effectiveness, the complexity of RS image content imposes higher demands on perturbation design. Existing perturbation methods fail to incorporate the unique attributes of RS images, such as multi-spectral characteristics and the presence of pan-sharpening pre-processing workflows. As illustrated in Figure 1, different pan-sharpening algorithms produce fusion results exhibiting variations in spatial detail and spectral fidelity. In this work, we propose FusionMatch, a novel consistency regularization-based semi-supervised framework that integrates two novel perturbation strategies, namely NIRPerb and PSPerb. NIRPerb integrates near-infrared data to expand the spectral dimensionality of input perturbations. PSPerb utilizes the divergent fusion results generated by the pan-sharpening process as an informative perturbation source. These two perturbation strategies are specifically designed to target the characteristics of RS images. NIRPerb enhances the informational dimensionality of the input data by incorporating additional near-infrared bands. PSPerb transforms the differences in fusion results produced by various pan-sharpening methods into a valuable perturbation source. Collectively, NIRPerb and PSPerb facilitate the construction of a more comprehensive perturbation space, effectively enhancing the performance of semi-supervised semantic segmentation models.

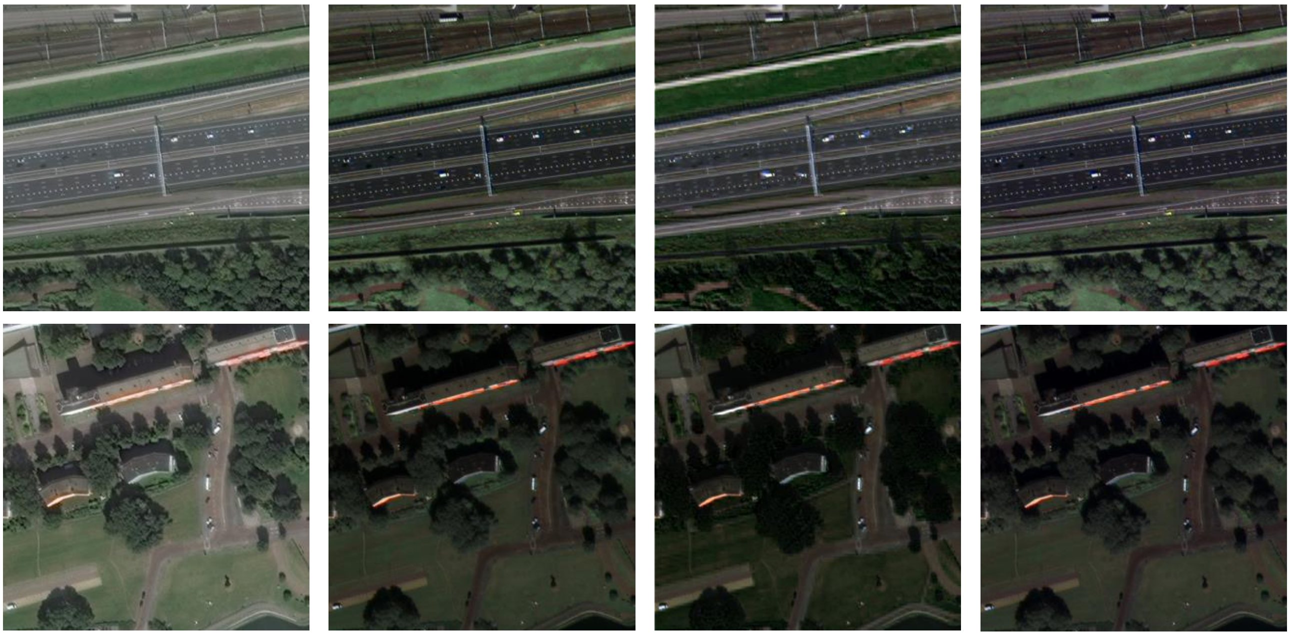

Figure 1. The results of different pan-sharpening methods. Subtle variations in spatial details and spectral fidelity can be observed in the various fusion outputs. From left to right: the pan-sharpened image provided by the SpaceNet6 (Shermeyer et al., 2020) dataset, followed by the fusion results of HIS (Chien et al., 2013), GS (Maurer et al., 2013) and PCA (Jelének et al., 2016).

2. Related Works

2.1 Semi-supervised Semantic Segmentation

The semi-supervised semantic segmentation approach artfully combines the distinctive strengths of supervised and unsupervised learning, achieving synergistic complementarity that demonstrates superior performance. Specifically, this methodology ingeniously leverages abundant and readily available unlabeled data as supplementary resources, thereby significantly reducing the excessive reliance on large-scale, high-quality annotated datasets. This characteristic holds profound practical significance, as acquiring labeled data typically demands substantial human, material, and temporal resources, whereas unlabeled data can be collected with relative ease. Among various semi-supervised learning techniques, self-training and consistency regularization methods have emerged as particularly prominent due to their unique advantages and demonstrated effectiveness.

Self-training represents a foundational semi-supervised approach where a model is initially trained on annotated samples before being applied to predict pseudo-labels for unannotated instances. These newly labeled examples are then reintegrated into the training pool to refine the model in an iterative manner (Amini et al., 2022). However, self-training is also often affected by the inherent noise in pseudo-labels and data imbalance. To address the noise in pseudo-labels, Feng et al. (2022) propose dynamic mutual training (DMT), which leverages the disagreement between two distinct models to identify and down-weight potentially erroneous pseudo-labels. For the data imbalance, particularly acute in tasks like semantic segmentation with long-tailed class distributions, Zhu et al. (2024) introduced centroid sampling within their efficient self-training (CSST) framework.

Consistency regularization is another fundamental framework in semi-supervised learning, enforcing prediction invariance under various perturbations, including input, feature, and network-level perturbations.

Beyond the conventional weak and strong perturbation techniques within the WSCR framework, diverse perturbation strategies operate at the input level. Cutout (DeVries et al., 2017) enhances regularization by randomly removing contiguous, fixed-size rectangular regions from input images, effectively simulating occlusion and forcing models to utilize broader context. Mixup (Zhang et al., 2017) synthesizes novel training samples and their corresponding soft labels through convex interpolations between randomly selected pairs of input images and their one-hot label encodings, encouraging smoother decision boundaries. Building upon these, CutMix (Yun et al., 2019) replaces a randomly cropped region from one image with the corresponding patch from another image, proportionally blending the ground truth labels based on the patch area.

Feature perturbation techniques focus on modifying intermediate layer representations rather than input data. For instance, Ouali et al. (2020) propose CCT framework, which applies diverse perturbation strategies—including feature dropout, guided masking, guided Cutout, and adversarial noise—directly to encoder outputs. Similarly, Yang et al. (2023a) propose UniMatch, which incorporates feature-level perturbations within a dual-stream architecture, specifically employing feature dropout to create auxiliary perturbation branches for unlabeled data.

Network perturbation strategies introduce architectural variations to enhance supervision signals. The cross pseudo supervision (CPS) approach (Chen et al., 2021) leverages dual segmentation networks with shared architecture but divergent initialization. These networks mutually generate pseudo-labels that enforce bidirectional consistency constraints. Extending this paradigm, guided pseudo supervision (GPS) (Cho et al., 2024) implements a tri-network framework where dual student models consolidate their predictions. This aggregated output subsequently refines the teacher network through iterative feedback mechanisms.

2.2 Semi-supervised Semantic Segmentation of RS Images

In recent years, the successful application of SSL in natural images has inspired researchers to explore its potential in RS image segmentation.

Wang et al. (2021) propose RanPaste, a semi-supervised framework for remote sensing segmentation that enhances consistency regularization through spatial mixing. The method generates composite training samples by transferring annotated regions from labeled images to unlabeled counterparts, concurrently blending ground-truth segments with pseudo-labels. Extending this work, Yang et al. (2023b) employ class-wise masks (based on ground truth) to incorporate labeled data into unlabeled images, achieving improved segmentation accuracy. Lu et al. (2023) introduce WSCL, a framework that enhances semi-supervised segmentation through a unique sparse dual-view cross-sample generation method, augmenting training data diversity. Concurrently, it employs an entropy-driven adaptive reweighting mechanism to mitigate noise within pseudo-labels. Lv et al. (2024) enforce prediction consistency across multiple diverse strong perturbations using pseudo-labels generated from a single weak augmentation. They further incorporate an adaptive confidence threshold that decays over time to selectively identify reliable pixels for training.

3. Methodology

This section elaborates on the core design of the FusionMatch framework. By innovatively integrating multi-modal near-infrared data and employing diverse pan-sharpening techniques to generate distinct fused images, FusionMatch constructs a more expansive perturbation space.

3.1 Weak-to-strong consistency regularization

The fundamental advantage of weak-to-strong consistency regularization lies in its capacity to effectively harness large-scale unlabeled image data. This paradigm enforces prediction consistency for identical input samples subjected to different perturbation intensities: pseudo-labels generated from weakly perturbed images supervise predictions from their strongly perturbed counterparts. Such multi-view training facilitates robust feature learning from data variations, ultimately improving model generalization. Thus, the design of perturbation strategies becomes crucial in semi-supervised learning. Current approaches employ diverse perturbation techniques to maximize perturbation variability. Weak perturbations typically consist of geometric transformations (e.g., rotation, scaling, cropping), whereas strong perturbations incorporate photometric modifications including grayscale conversion and stochastic adjustments of brightness, contrast, and hue.

3.2 Information Fusion Perturbation

In this work, we propose two novel input perturbation strategies based on the multi-spectral characteristics and the pan-sharpening pre-processing pipeline of RS imagery, namely NIRPerb and PSPerb. The NIRPerb enhances the spectral dimensionality of training data by fusing near-infrared band information, thereby improving the model's capability to process additional spectral features.

For technical implementation, we integrate three fusion algorithms (Awad et al., 2019; Li et al., 2020; Yang et al., 2023c) to achieve NIRPerb, which increases perturbation diversity through methodological variations in the fusion process. The implementation of NIRPerb can be expressed as:

$$x_{NIRPerb} = Fusion(x_{RGB}, x_{NIR}) \tag{1}$$

where x_{RGB} , x_{NIR} and $x_{NIRPerb}$ denote the RGB input, near-infrared band input, and perturbed output of NIRPerb, respectively. The fusion process $Fusion$ integrates NIR data into the visible spectrum. To maximize perturbation diversity, the selection of $Fusion$ follows a randomized strategy for each input pair.

Pan-sharpening of RS images exhibits dual characteristics as both an image enhancement technique and an information fusion. The diversity of fusion strategies directly influences the heterogeneity of fusion outcomes, thereby facilitating the implementation of PSPerb to significantly improve the model's capability in detecting subtle variations. The PSPerb process can be formally expressed as:

$$x_{PSPerb} = DFusion(x_{RGB}, x_{PAN}) \tag{2}$$

where x_{PAN} , x_{PSPerb} and $DFusion$ respectively denote the panchromatic image, the result image after PSPerb processing, and a pan-sharpening technique. Based on computational efficiency and methodological generalizability considerations, three established pan-sharpening techniques (Maurer et al., 2013; Chien et al., 2013; Jelsamnek et al., 2016) are employed to operationalize PSPerb

3.3 FusionMatch

FusionMatch integrates supervised and unsupervised learning within a unified framework. In the supervised branch, the model undergoes conventional training using available semantic labels to guide prediction outputs. In the unsupervised branch, information fusion perturbation is first applied to the unlabeled images. For datasets containing only multi-spectral images, NIRPerb is employed to achieve information fusion. For datasets with both panchromatic and multispectral images, PSPerb is performed before NIRPerb to attain a more pronounced perturbation effect. Subsequently, the same weak perturbation treatment is concurrently applied to the images that have undergone information fusion perturbation and the original images, generating weakly perturbed results x^{ifw} and x^{w}, respectively. x^{ifw} is then subjected to strong perturbation to obtain the perturbed version x^{ifws} . Finally, the model's output results for x^{w} and x^{ifws} are obtained. A confidence threshold is established to filter out reliable pseudo-labels from the prediction of x^{w}, which are used to supervise the prediction of x^{ifws} , thereby realizing a consistency regularization constraint from weak to strong perturbation.

4. Experiments and Results

This section begins with a concise overview of the adopted datasets, followed by a systematic presentation of the experimental implementation details and performance

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-G-2025-1085-2025 | © Author(s) 2025. CC BY 4.0 License.

evaluation metrics. Building upon this foundation, we conduct comprehensive comparative analyses to benchmark the FusionMatch model against state-of-the-art methods in semi-supervised semantic segmentation, thoroughly evaluating their relative performance.

4.1 Datasets

To establish a robust experimental framework, we evaluate our method on two publicly available benchmark datasets: the SpaceNet6 (Shermeyer et al., 2020) building extraction dataset and the ISPRS Potsdam (Rottensteiner et al., 2012) semantic labeling dataset.

The SpaceNet6 dataset provides 3,401 image collections featuring both 4-band multispectral data at 1.0-meter resolution (450×450 pixels) and higher-resolution 0.5-meter pan-sharpened products (900×900 pixels) including RGB, RGB-NIR, and panchromatic versions. A distinguishing characteristic is its comprehensive annotation of approximately 48,000 individual building footprints.

For broader validation across multiple land cover classes, we additionally employ the ISPRS Potsdam dataset, which represents one of the most widely recognized benchmarks in high-resolution urban scene analysis. The dataset contains 38 orthorectified image tiles, each measuring 6000×6000 pixels at an exceptional 0.05-meter ground sampling distance. Each tile is provided in multiple spectral configurations (IRRG, RGB, and RGBIR) to facilitate comprehensive multispectral analysis. Notably, the contiguous tile arrangement preserves complete spatial context without any inter-tile gaps, enabling more effective exploitation of neighborhood relationships during model training and evaluation.

4.2 Implementation details and metrics

The experiments are performed using an NVIDIA RTX 3090 graphics processing unit. To accommodate hardware memory limitations, images are pre-processed by cropping to 256×256 pixel dimensions. The dataset partitioning scheme allocates 70% of samples for model training, 10% for validation, and 20% for final testing. To investigate the impact of label scarcity, we create three distinct training configurations where only 1/16, 1/8, or 1/4 of the training samples contain annotations, with the remainder treated as unlabeled data.

We adopt three standard metrics for quantitative assessment. Overall accuracy (OA) is the proportion of correctly classified pixels. Intersection over union (IoU) measures the area overlap between predictions and ground truth. F1 score is the harmonic mean of precision and recall. These metrics are derived from the fundamental confusion matrix components (TP, FP, TN, FN), which are calculated as follows:

$$OA = \frac{TP + TN}{TP + FP + TN + FN} \tag{3}$$

$$IoU = \frac{TP}{TP + FP + FN} \tag{4}$$

$$F1 = \frac{2TP}{2TP + FP + FN} \tag{5}$$

For comprehensive class-wise evaluation, we report both mean IoU (mIoU) and mean F1 (mF1) scores across all C semantic categories, which can be defined as:

$$mF1 = \frac{\sum_{i=1}^{C} F1_i}{C} \tag{6}$$

$$mIoU = \frac{\sum_{i=1}^{C} IoU_i}{C} \tag{7}$$

4.3 Comparative experimental analysis

To evaluate FusionMatch's efficacy in semi-supervised RSSS, we benchmark it against four leading weak-to-strong consistency approaches: FixMatch (Sohn et al., 2020), ST++ (Yang et al., 2022), WSCL (Lu et al., 2023), and MCSS (Lv et al., 2024). Identical training configurations are employed for all methods: the DeeplabV3+ architecture [Chen et al., 2018] using ResNet101 [He et al., 2016] as backbone, trained for the same number of epochs. Performance is assessed on the SpaceNet6 dataset using OA, IoU, and F1 metrics. For the Potsdam dataset, segmentation accuracy is evaluated primarily through per-class IoU scores.

1) Comparative analysis on the SpaceNet6 dataset: The SpaceNet6 dataset exhibits a severe foreground-background pixel imbalance. To address the limited sensitivity of Overall Accuracy (OA) in distinguishing performance differences, our bar chart visualization (Figure 2) concentrates on the IoU and F1 metrics for evaluating building extraction capabilities. All semi-supervised algorithms demonstrate substantially taller bars than the supervised baseline (SupOnly), forming a distinct performance gap particularly evident in the IoU metric. This validates the enhancement effect of semi-supervised learning with limited labeled data. Our proposed FusionMatch consistently generates the tallest bars across all three metrics. Across all labeling ratios, FusionMatch's bars occupy the visually dominant rightmost position in the chart. The spatial distribution characteristics within the bar chart provide visual evidence that FusionMatch, through its multi-modal consistency learning framework, maximizes the value of labeled information across all data scales, establishing the most robust semi-supervised building extraction approach currently available.

2) Comparison results on the Potsdam dataset: Figure 3 reveals FusionMatch's consistent superiority over the supervised baseline across all labeling ratios (1/16, 1/8, 1/4). The method dominates most semantic categories, particularly excelling in impervious surfaces (IS) and buildings (Bldg) at higher annotation levels. Notably, at 1/16 ratio, FusionMatch achieves peak performance in IS, low vegetation (LV), and cars despite MCSS leading in buildings. For 1/8 labels, while ST++ shows advantages in LV/tree segmentation and MCSS in cars, FusionMatch maintains strong IS/building results. At 1/4 ratio, FusionMatch sustains leadership in IS/buildings/cars despite ST++'s tree segmentation strength and MCSS's LV performance. Comparison results highlight FusionMatch's categorical advantages over FixMatch (especially in IS/cars) and WSCL (predominantly in buildings). Though MCSS shows competitive LV/tree segmentation at certain ratios, FusionMatch delivers more balanced category-level performance overall.

https://doi.org/10.5194/isprs-annals-X-G-2025-1085-2025 | © Author(s) 2025. CC BY 4.0 License.

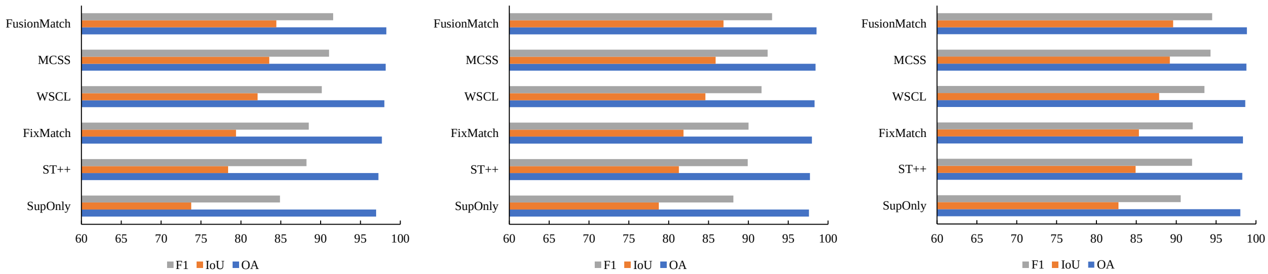

Figure 2. The experimental results on the SpaceNet6 dataset. From left to right, the label sample ratios are 1/16, 1/8, and 1/4, respectively.

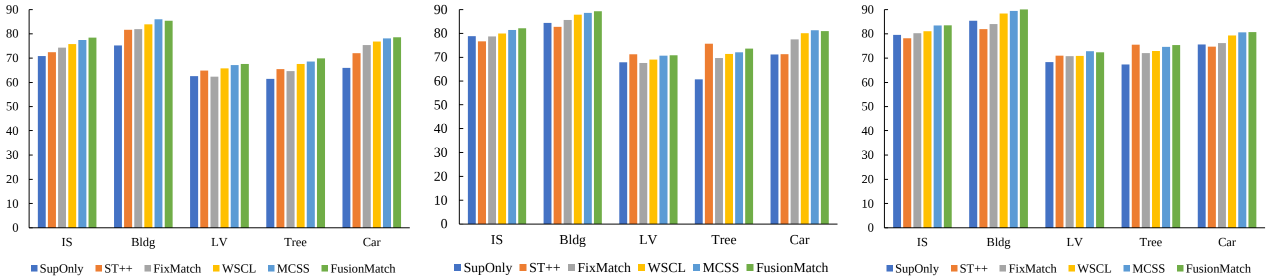

Figure 3. The experimental results on the Potsdam dataset. From left to right, the label sample ratios are 1/16, 1/8, and 1/4, respectively.

5. Conclusion

The rapid advancement of RS technologies has brought forth a critical research challenge: how to effectively harness the value of massive RS data through innovative image processing approaches. To address this challenge, we present FusionMatch, a novel semi-supervised architecture specifically designed for RS image semantic segmentation. Our framework achieves comprehensive feature learning by integrating multi-spectral information and employing differential pan-sharpening techniques, thereby significantly expanding the perturbation space of unlabeled RS data. Extensive comparative experiments demonstrate that FusionMatch substantially outperforms state-of-the-art semi-supervised methods.

Acknowledgements

This work is supported by the National Key Research and Development Program of China (2024YFC3015404), the National Natural Science Foundation of China (42271446), and the Central Guidance for Local Science and Technology Development Fund Projects of Sichuan Province (2024ZYD0084). For a more comprehensive analysis, extensive experimental results, and in-depth discussions, please refer to our extended version of this work (Zhou et al., 2025). We thank the conference reviewers for their insightful comments.

References

Amini, M. R., Feofanov, V., Pauletto, L., Hadjadj, L., Devijver, E., Maximov, Y., 2022. Self-training: A survey. arXiv preprint arXiv:2202.12040.

Awad, M., Elliethy, A., Aly, H. A., 2019. Adaptive near-infrared and visible fusion for fast image enhancement. IEEE Transactions on Computational Imaging, 6, 408-418.

Bachman, P., Alsharif, O., Precup, D., 2014. Learning with pseudo-ensembles. Advances in neural information processing systems, 27.

Bai, L., Wang, H., Zhang, X., Qin, W., Liu, B., Du, S., 2024. AP-Semi: Improving the Semi-Supervised Semantic Segmentation for VHR Images Through Adaptive Data Augmentation and Prototypical Sample Guidance. IEEE Transactions on Geoscience and Remote Sensing, 62, 1-13.

Chen, L. C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation. Proceedings of the European conference on computer vision (ECCV), 801-818.

Chen, Y., Ouyang, X., Zhu, K., Agam, G., 2021a. Complexmix: Semi-supervised semantic segmentation via mask-based data augmentation. 2021 IEEE International Conference on Image Processing (ICIP), 2264-2268.

Chen, X., Yuan, Y., Zeng, G., Wang, J., 2021b. Semi-supervised semantic segmentation with cross pseudo supervision. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR), 2613-2622.

Chien, C. L., Tsai, W. H., 2013. Image fusion with no gamut problem by improved nonlinear IHS transforms for remote sensing. IEEE Transactions on Geoscience and Remote Sensing, 52(1), 651-663.

Cho, H., Choi, I., Kwak, S., Kim, W. H., 2024. Interactive Network Perturbation between Teacher and Students for Semi-Supervised Semantic Segmentation. Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), 626-635.

DeVries, T., 2017. Improved Regularization of Convolutional Neural Networks with Cutout. arXiv preprint arXiv:1708.04552.

Feng, Z., Zhou, Q., Gu, Q., Tan, X., Cheng, G., Lu, X., Shi, J., Ma, L., 2022. DMT: Dynamic mutual training for semi-supervised learning. Pattern Recognition, 130, 108777.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), 770-778.

Jelének, J., Kopačková, V., Koucká, L., Mišurec, J., 2016. Testing a modified PCA-based sharpening approach for image fusion. Remote Sensing, 8(10), 794.

Li, J., Cai, Y., Li, Q., Kou, M., Zhang, T., 2024. A review of remote sensing image segmentation by deep learning methods. International Journal of Digital Earth, 17(1), 2328827.

Li, Z., Hu, H. M., Zhang, W., Pu, S., Li, B., 2020. Spectrum characteristics preserved visible and near-infrared image fusion algorithm. IEEE Transactions on Multimedia, 23, 306-319.

Lu, X., Jiao, L., Li, L., Liu, F., Liu, X., Yang, S., Feng, Z., Chen, P., 2023. Weak-to-strong consistency learning for semisupervised image segmentation. IEEE Transactions on Geoscience and Remote Sensing, 61, 1-15.

Lv, L., Zhang, L., 2024. Advancing Data-Efficient Exploitation for Semi-Supervised Remote Sensing Images Semantic Segmentation. IEEE Transactions on Geoscience and Remote Sensing.

Maurer, T. 2013. How to pan-sharpen images using the gram-schmidt pan-sharpen method–A recipe. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 40, 239-244.

Miyato, T., Maeda, S. I., Koyama, M., Ishii, S., 2018. Virtual adversarial training: a regularization method for supervised and semi-supervised learning. IEEE transactions on pattern analysis and machine intelligence, 41(8), 1979-1993.

Ouali, Y., Hudelot, C., Tami, M., 2020. Semi-supervised semantic segmentation with cross-consistency training. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR), 12674-12684.

Persello, C., Wegner, J. D., Hänsch, R., Tuia, D., Ghamisi, P., Koeva, M., Camps-Valls, G., 2022. Deep learning and earth observation to support the sustainable development goals: Current approaches, open challenges, and future opportunities. IEEE Geoscience and Remote Sensing Magazine, 10(2), 172-200.

Rottensteiner, F., Sohn, G., Jung, J., Gerke, M., Baillard, C., Benitez, S., Breitkopf, U., 2012. The ISPRS benchmark on urban object classification and 3D building reconstruction. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences; I-3, 1(1), 293-298.

Shermeyer, J., Hogan, D., Brown, J., Van Etten, A., Weir, N., Pacifici, F., Hansch, R., Bastidas, A., Soenen, S., Bacastow, T., Lewis, R., 2020. SpaceNet 6: Multi-sensor all weather mapping dataset. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops (CVPRW), 196-197.

Sohn, K., Berthelot, D., Carlini, N., Zhang, Z., Zhang, H., Raffel, C. A., Cubuk, E. D., Kurakin, A., Li, C. L., 2020. Fixmatch: Simplifying semi-supervised learning with consistency and confidence. Advances in neural information processing systems, 33, 596-608.

Teh, E. W., DeVries, T., Duke, B., Jiang, R., Aarabi, P., Taylor, G. W., 2022. The gist and rist of iterative self-training for semi-supervised segmentation. 19th Conference on Robots and Vision (CRV), 58-66.

Wang, J. X., Chen, S. B., Ding, C. H., Tang, J., Luo, B., 2021. RanPaste: Paste consistency and pseudo label for semisupervised remote sensing image semantic segmentation. IEEE Transactions on Geoscience and Remote Sensing, 60, 1-16.

Yang, L., Qi, L., Feng, L., Zhang, W., Shi, Y., 2023a. Revisiting weak-to-strong consistency in semi-supervised semantic segmentation. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 7236-7246.

Yang, Z., Yan, Z., Diao, W., Zhang, Q., Kang, Y., Li, J., Li, X., Sun, X., 2023b. Label propagation and contrastive regularization for semisupervised semantic segmentation of remote sensing images. IEEE Transactions on Geoscience and Remote Sensing, 61, 1-18.

Yang, B., Jiang, Z., Pan, D., Yu, H., Gui, W., 2023c. Detail-aware near infrared and visible fusion with multi-order hyper-Laplacian priors. Information Fusion, 99, 101851.

Yang, L., Zhuo, W., Qi, L., Shi, Y., Gao, Y., 2022. St++: Make self-training work better for semi-supervised semantic segmentation. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR), 4268-4277.

Yun, S., Han, D., Oh, S. J., Chun, S., Choe, J., Yoo, Y., 2019. Cutmix: Regularization strategy to train strong classifiers with localizable features. Proceedings of the IEEE/CVF international conference on computer vision (ICCV), 6023-6032.

Zhang, H., Cisse, M., Dauphin, Y. N., Lopez-Paz, D., 2017. MixUp: Beyond empirical risk minimization. arXiv preprint arXiv:1710.09412.

Zhou L., Duan K., Dai J., Ye Y., 2025. Advancing perturbation space expansion based on information fusion for semi-supervised remote sensing image semantic segmentation. Information Fusion, 117, 102830.

Zhu, Y., Zhang, Z., Wu, C., Zhang, Z., He, T., Zhang, H., Manmatha, R., Li, M., 2021. Improving semantic segmentation via efficient self-training. IEEE transactions on pattern analysis and machine intelligence, 46(3), 1589-1602.