remote sensing

MDPI

Article

Multi-Teacher Divergence Perception Network for Semi-Supervised Semantic Segmentation of Remote Sensing Images

Weiheng Zhao, Xueyan Dong *, Bangjie Li, Dongdong Guan, Xiaolong Zheng, Zhengsheng Chen and Shuaizhi Kang

Rocket Force University of Engineering, Xi'an 710025, China; 2019027008@chd.edu.cn (W.Z.); rosoldier@163.com (B.L.); gdd@whu.edu.cn (D.G.); xxgmzz@163.com (X.Z.); czsgeo@126.com (Z.C.); kangshuaitim@163.com (S.K.)
* Correspondence: 2019027009@chd.edu.cn

Highlights

What are the main findings?

- A multi-teacher adaptive disagreement-aware module converts pixel-level teacher divergence into reliability-specific pseudo-label supervision.
- An evidential uncertainty estimation module distinguishes evidence-supported predictions from overconfident predictions with insufficient class evidence.

What are the implications of the main findings?

- Joint teacher-side disagreement perception and student-side evidential uncertainty improves semi-supervised segmentation under 5% and 25% labeling ratios on Vaihingen and Potsdam.
- The framework improves difficult boundaries and small objects while retaining useful supervision from high-disagreement but reliable regions.

Abstract

Semi-supervised semantic segmentation can exploit unlabeled remote sensing images to reduce the cost of pixel-level annotation, but its performance is strongly constrained by pseudo-label reliability. Multi-teacher predictions often exhibit structured disagreement around boundaries, small objects, and interclass-confusion regions, while a conventional softmax head may assign spuriously high confidence without sufficient supporting evidence. To address these limitations, we propose a Multi-Teacher Divergence Perception Network (MTDPNet). Its Multi-Teacher Adaptive Disagreement-Aware Module (MADM) explicitly measures pixel-level teacher disagreement using the Jensen–Shannon divergence, adaptively weights teacher predictions, and applies reliability-specific supervision to different image regions. In parallel, an Evidential Uncertainty Estimation Module (EUEM) fuses multiscale student features and constructs a Dirichlet evidence distribution to identify evidence-deficient predictions and prevent their over-reinforcement. On the ISPRS Vaihingen and Potsdam datasets with 5% and 25% labeled data, MTDPNet achieves mIoU scores of 63.44%/68.18% and 50.95%/64.96%, respectively, and corresponding mF1 scores of 76.51%/80.24% and 67.01%/78.63%, yielding the best overall performance in all four settings. These results demonstrate that jointly modeling teacher disagreement and student evidential uncertainty improves the reliable use of unlabeled remote sensing images, particularly for difficult regions and small objects.

Check for updates

Academic Editors: Keyan Chen, Liqin Liu, Wei Lu, Wenyuan Li, Xiantao Jiang and Yi Xiao

Received: 5 August 2026
Revised: 20 September 2026
Accepted: 21 September 2026
Published: 24 September 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

Keywords: remote sensing images; semi-supervised semantic segmentation; multi-teacher learning; disagreement awareness; evidential uncertainty

1. Introduction

Remote sensing (RS) image semantic segmentation assigns a semantic category to every pixel and provides fundamental land-cover information for applications such as land-cover mapping, fine-grained urban management, disaster assessment, and environmental monitoring. Advances in deep neural networks have substantially improved RS scene understanding [1]. However, high-performing models commonly rely on large-scale pixel-level annotations. Compared with image-level labeling, fine-grained annotation of RS images requires domain expertise and careful treatment of large-scale variations, interclass spectral similarity, shadow occlusion, and ambiguous semantic boundaries, leading to considerable time and labor costs. Semi-supervised semantic segmentation (SSSS) offers a practical solution by jointly exploiting a small set of labeled images and a large collection of unlabeled images.

Existing SSSS methods mainly follow two lines of research: pseudo-label-based self-training and consistency regularization. Mean Teacher constructs a teacher model by applying an exponential moving average to the student parameters, thereby providing relatively stable supervision for unlabeled data [2]. FixMatch uses high-confidence pseudo-labels produced from weakly augmented inputs to constrain predictions under strong augmentation, unifying pseudo-label learning and consistency training [3]. Subsequent approaches improve self-training through prediction-stability-based sample selection [4], strengthen consistency learning with perturbed mean teachers [5], and explicitly localize erroneous pseudo-label pixels [6]. For dense prediction, Cross-Consistency Training (CCT) enforces feature consistency through multiple perturbed decoders [7], cross-pseudo supervision (CPS) uses two networks to generate pixel-level supervision for each other [8], and U^2PL shows that low-confidence predictions can still contain useful training information [9]. Although these methods broaden the use of unlabeled data, their effectiveness remains highly dependent on pseudo-label reliability. Once erroneous predictions are recurrently fed back into training, consistency regularization may amplify rather than correct the errors.

This problem is particularly pronounced in RS imagery. Large-area RS scenes exhibit strong spatial heterogeneity and often involve shifts in class distributions across images. The same class may vary substantially in scale and appearance, whereas roads, buildings, and low vegetation can show similar local spectral or textural patterns. Previous studies have improved the use of unlabeled images through class-adaptive thresholds, consistency-based self-training, label propagation and contrastive regularization, dynamic pseudo-label sampling, class-aware unlabeled-data selection for large-scale land-cover mapping, and uncertainty-aware consistency [10–16]. Nevertheless, most methods still assess pixel reliability using the maximum class probability, fixed or dynamic thresholds, or a single aggregated prediction. These criteria can reject some obvious errors but cannot fully describe the relationships among teacher predictions, and they tend to compress structured differences at boundaries, small objects, and class-confusion regions into a single confidence value.

Multi-teacher learning provides richer supervisory sources and can reduce the bias of an individual teacher [8,17]. When multiple teachers produce similar class distributions for a pixel, the resulting pseudo-label is generally more stable. Conversely, clearly separated predictions often indicate a difficult pixel located near a semantic boundary, in an interclass-confusion region, or within a minority class. Figure 1a–d illustrates complex junctions

between buildings, roads, and vegetation, where a slight localization error can induce marked disagreement across the two sides of a boundary. Vehicles in Figure 1e–h occupy only a few pixels and are easily obscured by shadows, tree crowns, and similar background textures, making them prone to omission or confusion. Such regions may exhibit both teacher disagreement and insufficient student evidence and therefore constitute typical hard regions. Existing studies improve supervisory reliability through confidence-based dual-teacher fusion, as in CWD-Teacher [17], or teacher–student prediction discrepancies, as in RS-MUCA [16]. Both approaches still collapse the multiple teacher predictions into a single fused target, so the disagreement that marks boundaries, small objects, and interclass confusion is discarded before it can be exploited. However, averaging, voting, and confidence weighting still primarily aim to produce a unified pseudo-label, and explicit inter-teacher disagreement is easily attenuated during aggregation.

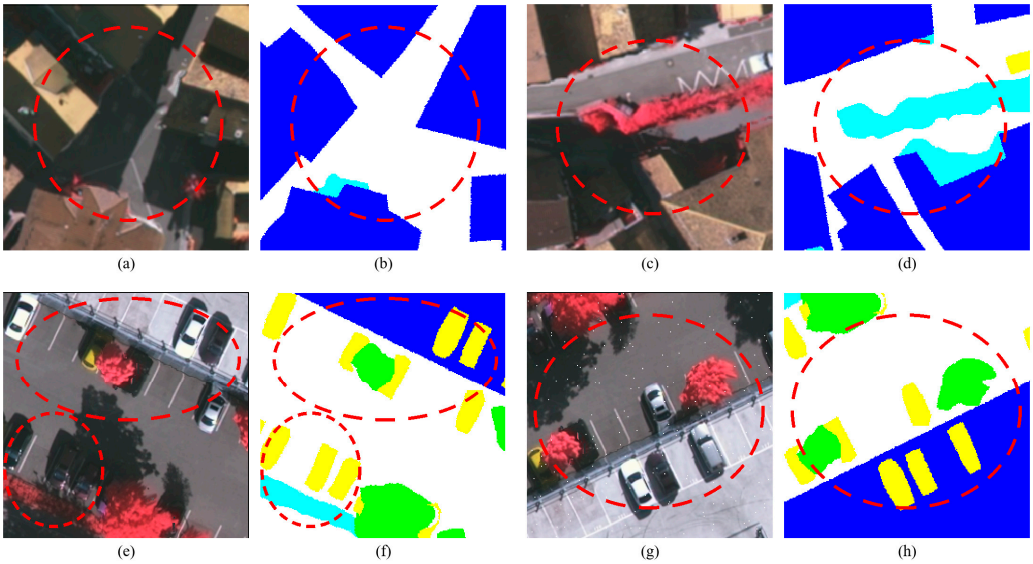

Figure 1. Representative semantic segmentation challenges in complex RS scenes. (**a**–**d**) show two image-label pairs with complex boundaries among buildings, roads, and vegetation. (**e**–**h**) show two additional image-label pairs containing small objects such as vehicles near parking areas and roads. In the label maps, white, yellow, dark blue, green, and light blue denote impervious surface, vehicle, building, vegetation, and low vegetation, respectively. Red dashed ellipses indicate representative regions.

Unlike CWD-Teacher [17], which fuses two teacher predictions by confidence weighting, and RS-MUCA [16], which aligns teacher and student features through cross-attention, MADM retains the disagreement among teacher predictive distributions as an explicit pixel-level reliability signal and uses it to regulate both the teacher contributions and the region-adaptive training of the student. This observation raises the first key question: how can multi-teacher disagreement be transformed from a by-product of fusion into an interpretable pixel-level reliability signal, and how should the supervisory strength be adapted to the degree of disagreement?

To address this problem, we develop a Multi-Teacher Adaptive Disagreement-Aware Module (MADM). MADM measures pixel-level differences among teacher predictive distributions using the symmetric and bounded Jensen–Shannon divergence (JSD) [18] and adaptively determines fusion weights according to each teacher's mean disagreement with the remaining teachers. It then combines overall disagreement with the fused confidence to partition unlabeled pixels into high-consistency, moderate-disagreement, high-disagreement-but-usable, and high-disagreement/low-confidence regions. High-consistency pixels provide stable pseudo-label supervision; moderate-disagreement pixels

receive attenuated soft-consistency constraints; high-disagreement but sufficiently confident pixels are trained conservatively with boundary-aware weighting; and highly unreliable pixels are excluded from strong pseudo-label learning. Thus, disagreement is used to regulate both teacher contributions and region-specific training rather than being uniformly averaged or discarded.

Reconstructing pseudo-label supervision only at the teacher side is insufficient to suppress the accumulation of uncertainty during semi-supervised training. A conventional softmax head outputs normalized class probabilities, but a high probability does not necessarily imply sufficient class evidence, and modern neural networks can produce poorly calibrated overconfident predictions [19]. Approximate Bayesian methods such as Monte Carlo Dropout and Monte Carlo Batch Normalization can estimate prediction variability through repeated stochastic forward passes [20,21]. RS semi-supervised methods have also characterized uncertain regions using stochastic predictions or multiscale consistency [15,16]. However, repeated sampling increases training or inference costs, whereas maximum probability, entropy, and predictive variance mainly describe the shape or variability of the output distribution. They do not directly determine whether the student has accumulated sufficient evidence for a class—a distinction that is especially important for mixed pixels, shadowed regions, small objects, and spectrally similar classes.

We therefore introduce an Evidential Uncertainty Estimation Module (EUEM). Evidential deep learning interprets network outputs as non-negative class evidence and uses a Dirichlet distribution to jointly represent class probabilities and their uncertainty [22]. EUEM fuses multiscale features from the student encoder, predicts pixel-level class evidence, and maps it to Dirichlet parameters, yielding both the expected class probabilities and an uncertainty measure determined by the total evidence strength. During training, high-confidence pixels enforce consistency between the evidential branch and the main student branch, whereas low-confidence pixels are regularized toward a uniform noninformative prior to suppress excessive evidence. EUEM can therefore distinguish reliable predictions supported by sufficient evidence from spuriously high-confidence predictions caused by softmax normalization.

Based on these observations, we propose the Multi-Teacher Divergence Perception Network (MTDPNet), which unifies teacher-side supervisory reliability and student-side evidence sufficiency within a semi-supervised framework. MADM determines whether pseudo-labels from different teachers are consistent and how strongly they should participate in training, whereas EUEM determines whether the student's decision is supported by sufficient evidence. The two modules act on complementary aspects of pseudo-label generation and student learning. Accordingly, this study addresses the following scientific question: how can multi-teacher prediction disagreement and student evidential uncertainty be jointly exploited to provide differentiated supervision for unlabeled pixels in complex RS scenes, while suppressing persistent reinforcement of erroneous pseudo-labels and retaining useful information from hard regions? Unlike prior methods that fuse teacher predictions into one target or filter pixels by a scalar confidence, MTDPNet uses teacher disagreement itself as a per-pixel supervisory attribute and student evidence strength as a second, complementary reliability criterion.

The main contributions of this work are summarized as follows.

We propose MTDPNet for semi-supervised semantic segmentation of RS images. The network models the training value of unlabeled pixels from two complementary perspectives: teacher-side pseudo-label reliability and student-side evidence sufficiency.

We develop MADM to explicitly characterize pixel-level multi-teacher disagreement using JSD and convert it into teacher reliability weights and a region-adaptive supervision strategy, enabling pixels with different reliability levels to receive differentiated training.

We design EUEM to learn Dirichlet class evidence from multiscale student features. High-confidence consistency and low-confidence uniform-prior constraints reduce the risk of reinforcing predictions that appear confident despite insufficient evidence.

2. Related Work

2.1. Semi-Supervised Semantic Segmentation

Semi-supervised semantic segmentation jointly uses a small number of labeled samples and a large number of unlabeled samples to train a pixel-level prediction model. Existing methods can be broadly grouped into pseudo-label-based self-training, consistency regularization, and combinations of the two. Pseudo-labeling converts model predictions on unlabeled samples into training targets and controls noise through confidence thresholds or sample selection. Consistency methods instead require stable predictions under different input, feature, or network perturbations. Mean Teacher constructs a teacher network by exponentially averaging student parameters and provides relatively stable consistency targets for unlabeled data [2]. FixMatch constrains predictions under strong perturbations using high-confidence pseudo-labels generated from weakly perturbed inputs [3], establishing a widely adopted paradigm for subsequent semi-supervised segmentation.

To accommodate the dense prediction nature of semantic segmentation, later studies extended these paradigms through feature perturbations, mutual network supervision, and the use of unreliable pixels. CCT attaches multiple perturbed decoders to a shared encoder and exploits unlabeled data through feature-level consistency [7]. A strong-augmentation baseline improves pseudo-label training through separate batch-normalization statistics and a self-correction loss [23]. CPS uses two independently initialized networks to provide pixel-level supervision for each other [8]. U^2PL subsequently converts low-confidence predictions into negative learning cues, showing that unreliable pixels need not be discarded [9], whereas UniMatch expands the perturbation space through image-level and feature-level dual-stream consistency [24]. Despite this mature technical foundation, performance remains dependent on pseudo-label quality. Erroneous predictions in difficult regions can be fed back by thresholding, mutual supervision, or consistency regularization, thereby inducing confirmation bias. Recent methods further refine unlabeled supervision through pruning-guided curricula [25], patch-level uncertainty localization [26], enhanced soft labels [27], collaborative logit-representation supervision [28], correlation-based label propagation [29], Transformer-specific perturbation regularization [30], and rank-aware correlation consistency [31].

2.2. Semi-Supervised Semantic Segmentation of Remote Sensing Images

Beyond the challenges of generic semi-supervised learning, SSSS of RS images must handle scene-distribution shifts, long-tailed class distributions, large object-scale variations, and complex boundary structures. Early studies mainly adapted self-training and consistency learning to these characteristics. Lu et al. proposed a linear-sampling self-training framework that dynamically allocates pseudo-label thresholds across classes to reduce the bias of a fixed threshold toward easy classes [10]. Li et al. combined consistency constraints with self-training and reweighted the unlabeled loss using pixel-level discriminative information [11]. Yang et al. connected labeled and unlabeled feature distributions through label propagation, prototype constraints, and regional contrastive regularization, improving feature separability under class imbalance [12]. These studies indicate that directly transferring natural-image semi-supervised methods is insufficient for RS data, and pseudo-label selection must account for class, region, and feature structures.

Recent research has developed targeted mechanisms for erroneous pseudo-labels, long-tailed distributions, and semantic ambiguity in RS imagery. CACT constructs

certainty maps from stochastic predictions and corrects consistency targets to mitigate uncertainty-induced ambiguity [15]. DWL decouples labeled and unlabeled predictions and weights them according to relative confidence ranks within each pseudo-class, alleviating pseudo-label errors and class imbalance [32]. DAST improves class-distribution imbalance through dynamic pseudo-label sampling, distribution matching, and adaptive threshold updates [13]. For teacher collaboration, a confidence-weighted dual-teacher network combines dual-teacher predictions with contrastive learning [17]; MUCA addresses scale variation and interclass similarity through multiscale uncertainty consistency and cross-teacher–student attention [16]; and PGNet introduces feature and boundary priors from a foundation segmentation model to improve the structural reliability of pseudo-labels [33]. Although these studies enhance RS semi-supervised learning through thresholds, class distributions, feature consistency, and prior information, a unified and fine-grained mechanism for interpreting multiple teacher predictions at the pixel level and converting them into differentiated supervision remains underexplored.

2.3. Teacher–Student Learning and Knowledge Distillation

The teacher–student paradigm is closely related to knowledge distillation. Classical knowledge distillation transfers interclass relationships through soft targets from a teacher model, enabling a student to acquire the teacher's generalization knowledge [34]. In semi-supervised learning, the teacher evolves from an offline source for model compression into an online generator of supervisory signals. Mean Teacher forms a dynamic teacher through an exponential moving average of student parameters [2], whereas CPS uses two peer networks to provide pseudo-labels for each other [8]. Both transfer predictive knowledge through unlabeled samples, but the former emphasizes temporal-ensemble stability and the latter exploits complementary network initializations. Teacher–student semi-supervised learning can thus be viewed as dynamic knowledge transfer over unlabeled data, for which the reliability and complementarity of supervisory sources are as important as teacher accuracy.

Multi-teacher designs provide more diverse supervisory sources and reduce single-teacher bias. In RS scenes, confidence-weighted dual teachers fuse two predictions according to their confidence [17], whereas MUCA reconstructs student representations using teacher features through cross-teacher–student attention [16]. Nevertheless, many multi-teacher methods still rely on averaging, voting, confidence weighting, or feature distillation to obtain one stable supervisory target. Although aggregation uses teacher complementarity, it also compresses explicit differences among teachers. For semantic boundaries, small objects, minority classes, and spectrally ambiguous regions, such disagreement often reflects local sample difficulty and supervision reliability. Simple averaging can obscure this information, whereas removing all high-disagreement pixels reduces learning opportunities in difficult regions. We therefore regard teacher disagreement as a modellable pixel-level supervisory attribute and investigate how it can regulate teacher contributions and training strength. Compared with CWD-Teacher [17], which weights a single fused pseudo-label by teacher confidence, and RS-MUCA [16], which estimates uncertainty from multi-scale student features, MADM measures inter-teacher disagreement directly from the teacher predictive distributions and converts it into region-adaptive supervision.

Interactive teacher–two-student training has also been used to exchange guided and feedback pseudo-supervision according to model readiness [35]. Another line of work relies on multi-learner mutual guidance, in which peer networks are trained simultaneously and supervise one another through exchanged pseudo-labels, as exemplified by cross-pseudo supervision between two networks (CPS) [8] and the interactive teacher–two-student exchange discussed above [35]. A recent theoretical analysis justifies mutual guidance

among peer learners and extends the principle to three mutually guiding learners under limited supervision [36]. In these mechanisms, supervision is exchanged among peers: the learners receive pseudo-labels from one another and are updated by mutual supervision. MTDPNet differs in both the source and the use of disagreement. Its teachers share the same architecture but evolve at different EMA time scales, are not trained against one another, and exchange no pseudo-labels. Their soft predictive disagreement is quantified by the Jensen–Shannon divergence and converted into teacher reliability weights and region-adaptive supervision.

2.4. Uncertainty-Aware Pseudo-Label Learning

Uncertainty estimation provides another route for assessing pseudo-label reliability. Semi-supervised methods commonly select pixels according to the maximum class probability, predictive entropy, or agreement among repeated stochastic predictions. Monte Carlo Dropout interprets stochastic dropout as approximate Bayesian inference and estimates prediction variability through multiple forward passes [20]. In RS semi-supervised segmentation, CACT constructs certainty weights from stochastic predictions [15], and MUCA estimates uncertainty at multiple feature levels to constrain teacher–student consistency [16]. These methods identify some unstable regions, but maximum probability and entropy mainly describe the shape of a normalized distribution, while sampling-based techniques require additional inference. More importantly, a high softmax probability does not necessarily indicate that sufficient class evidence has been accumulated.

Evidential deep learning characterizes reliability through evidence sufficiency rather than probability sharpness alone. Sensoy et al. interpret network outputs as non-negative class evidence and use a Dirichlet distribution to represent class probabilities and uncertainty [22]. When total evidence is low, the model retains high uncertainty instead of producing an apparently certain prediction after softmax normalization. This property is well suited to mixed pixels, shadows, and interclass-similarity regions in RS imagery. Existing RS semi-supervised methods, however, primarily use uncertainty for pseudo-label filtering, loss weighting, or feature consistency, and explicit modeling of student-side class evidence remains limited. Recently, evidential learning has attracted growing attention for label-efficient dense prediction. From softmax to Dirichlet [37] studies, there is a shift from softmax-based confidence to Dirichlet evidence for semi-supervised semantic segmentation, and Evidential Deep Partial Label Learning [38] extends evidence modeling to partial-label learning to quantify disambiguation uncertainty. EUEM belongs to this direction but occupies a distinct position: it estimates evidence sufficiency from the multiscale features of the student encoder, serves as a student-side branch complementary to the teacher-side disagreement modeling of MADM, and combines high-confidence consistency with a low-confidence uniform-prior regularizer. Maximum probability, entropy, and predictive variance characterize the shape or variability of a normalized output, whereas evidence sufficiency determines whether the student has accumulated enough class evidence; EUEM provides the latter and is therefore better suited to mixed pixels, shadows, small objects, and interclass-similarity regions. EUEM therefore learns a Dirichlet evidence representation from multiscale student features to complement teacher-side pseudo-label reliability modeling.

In summary, semi-supervised semantic segmentation has developed around pseudo-label self-training, consistency regularization, and teacher–student learning. RS-specific studies further address class imbalance, scale variation, boundary ambiguity, and interclass similarity through adaptive thresholds, contrastive learning, multiscale consistency, and prior guidance. These advances establish a strong basis for exploiting unlabeled RS images, but disagreement among multiple teacher sources and the evidence sufficiency of student

predictions are usually treated separately. In MTDPNet, MADM measures teacher-side predictive disagreement and converts it into reliability weights and region-adaptive supervision, whereas EUEM constructs a student-side Dirichlet evidence distribution. The two modules respectively address whether the pseudo-label source is reliable and whether the student prediction has sufficient evidence.

3. Materials and Methods

3.1. Materials/Study Area and Data

To evaluate MTDPNet under different levels of annotation scarcity, quantitative and qualitative experiments are conducted on the ISPRS Vaihingen and Potsdam datasets. Following the standard ISPRS benchmark protocol, the original aerial tiles of each dataset were first divided into disjoint training, validation, and test sets at the tile level. This tile-level split guarantees that patches derived from the same aerial tile never appear in more than one subset, preventing spatial data leakage. Each original image is cropped into nonoverlapping 256 × 256 patches with a stride of 256 pixels. For each dataset, 22,000 patches are generated: 4000 form the labeled training pool, 16,000 form the unlabeled training pool, 1000 labeled patches are used for validation, and the remaining 1000 labeled patches are used for testing.

To examine the effect of annotation scarcity, the unlabeled training set is fixed at 16,000 samples, while 800 or 4000 samples are selected from the labeled training pool. These settings correspond to labeled-to-unlabeled sample ratios of 5% and 25% (800/16,000 and 4000/16,000), representing severe and moderate label scarcity, respectively. Here the ratio is defined as the number of labeled samples relative to the number of unlabeled samples; the unlabeled set is fixed at 16,000 in both settings, and the labeled samples are drawn from the fixed 4000-patch labeled pool under a fixed random seed. All methods use identical data splits and evaluation protocols to ensure consistent comparisons across methods and labeling ratios. All compared methods, including MTDPNet, were re-trained under identical settings in terms of the backbone, data splits, data augmentation pipeline, initialization, and training budget, which are detailed in Section 4.1. The tile-level split and the random seed for labeled-patch sampling are fixed across all experiments (seed 42), so that every compared method and labeling ratio shares the same data partition.

3.2. Overall Objective

We construct a disagreement-aware evidential semi-supervised network for RS image semantic segmentation. The model comprises a multi-teacher pseudo-label generation branch, a student segmentation branch, and an evidential uncertainty estimation branch. Given a labeled dataset $D_L = \left\{ \left(x_n^l, y_n \right) \right\}_{n=1}^{N_L}$ and an unlabeled dataset $D_U = \{x_n^u\}_{n=1}^{N_U}$., x_n^l and x_n^u denote labeled and unlabeled RS images, respectively, y_n denotes the corresponding pixel-level annotations, and N_L and N_U are the numbers of labeled and unlabeled samples. To distinguish the number of image bands from the number of semantic classes, B denotes the number of bands and K denotes the number of classes, i.e., $x \in \mathbb{R}^{H \times W \times B}$ and $y \in \{0,1\}^{H \times W \times K}$, where H and W are the image height and width, respectively.

For labeled samples, the student is trained for standard semantic segmentation under supervised constraints. For an unlabeled sample, a weakly augmented view $a_w(x^u)$ and a strongly augmented view $a_s(x^u)$ are first constructed. The weak view is passed through multiple teachers to generate candidate pseudo-labels, whereas the strong view is fed to the student for consistency learning. The predictive probabilities of the mth teacher are $P_m^w = T_m\left(a_w(x^u); \theta_m^T\right)$, where $T_m\left(\cdot; \theta_m^T\right)$ denotes the mth teacher, θ_m^T denotes its parameters, and M is the number of teachers. The student outputs class probabilities $P_s^s = S(a_s(x^u); \theta)$ for the strongly augmented image, where $S(\cdot; \theta)$ is the student segmentation network, θ

denotes its parameters, and $P_s^s \in [0,1]^{H\times W\times K}$. Because strong augmentation may include geometric transformations such as cropping, flipping, rotation, CutMix, or ClassMix, teacher predictions are mapped synchronously to the strongly augmented view to ensure strict pixelwise alignment with the student output: $\tilde{P}{}_m^s = G_{w\to s}(P_m^w)$. Here, $G_{w\to s}(\cdot)$ is the geometric mapping from the weak to the strong view. Unless otherwise stated, subsequent teacher predictions $p_m(i)$ refer to probability distributions after this mapping.

As illustrated in Figure 2, weakly augmented views of an unlabeled image are fed to teacher networks operating at different EMA time scales. MADM uses inter-teacher JSD to perform pixel-level reliability weighting, soft pseudo-label fusion, and region partitioning, and then supervises the student prediction on the strongly augmented view in a region-adaptive manner. In parallel, EUEM fuses multiscale features from the student encoder and estimates pixel-level uncertainty through a Dirichlet evidence distribution. The two branches respectively constrain the reliability of the pseudo-label source and the evidence sufficiency of the student prediction and are jointly optimized with labeled supervision.

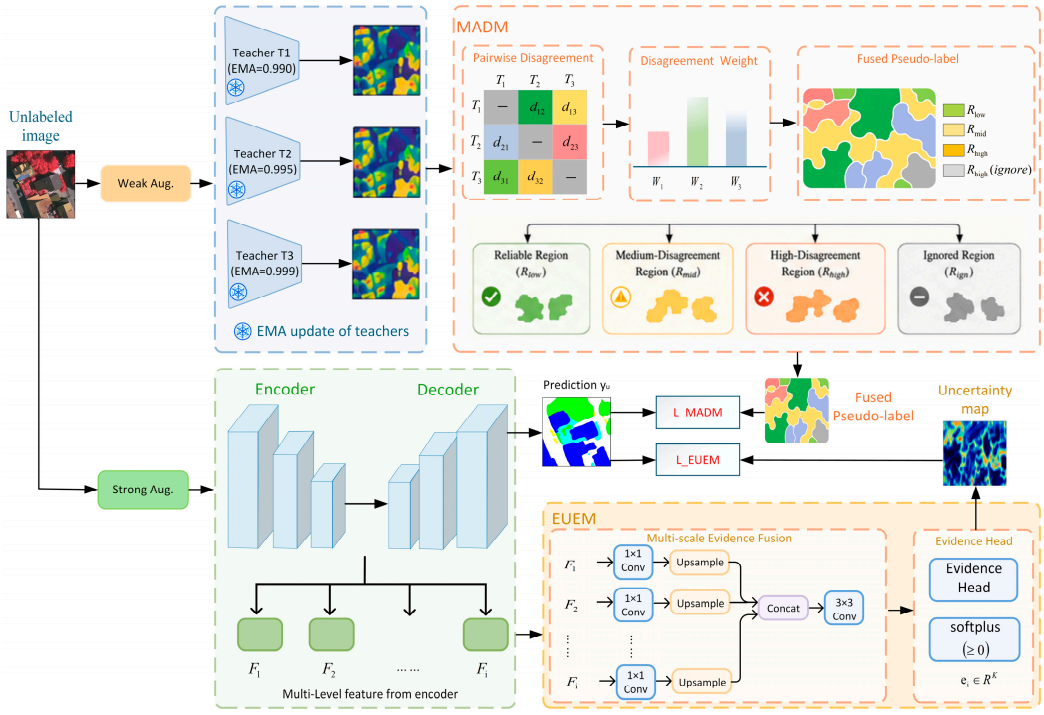

Figure 2. Overall architecture of MTDPNet. The multi-teacher branch predicts on weakly augmented views. MADM performs teacher weighting, pseudo-label fusion, and reliable-region partitioning according to JSD-based disagreement and provides differentiated supervision for the student on strongly augmented views. EUEM fuses multiscale student features and estimates pixel-level uncertainty from a Dirichlet evidence distribution. Labeled supervision, the MADM consistency loss, and the EUEM evidential constraints jointly optimize the student, while teacher parameters are updated by EMA.

Teacher weights are updated following the Mean Teacher paradigm. Teachers do not participate in gradient backpropagation; after each student update, they are updated by the exponential moving average (EMA) of the student parameters:

$$\theta_{m,t}^T = \alpha_m \theta_{m,t-1}^T + (1-\alpha_m)\theta_t, \tag{1}$$

Here, α_m is the EMA decay coefficient of the mth teacher, and θ_t denotes the current student parameters, including learnable parameters in the encoder, decoder, segmentation head, and associated normalization layers. To maintain diversity among teacher predictions, the

three teachers use decay coefficients of 0.990, 0.995, and 0.999 and independent stochastic perturbations during forward propagation. Teacher parameters receive no gradient updates, batch-normalization statistics remain frozen, and independent random masks are applied only to Dropout layers. This setting preserves prediction stability while inducing teacher diversity.

The supervised loss on labeled samples is pixelwise cross-entropy. Let Ω_l denote the set of valid labeled pixels; the supervised loss is defined as

$$\mathcal{L}_{\text{sup}} = -\frac{1}{\mathrm{N_L} \cdot |\Omega_\mathrm{l}|} \sum_{\mathrm{n=1}}^{\mathrm{N_L}} \sum_{\mathrm{i} \in \Omega_\mathrm{l}} \sum_{\mathrm{k=1}}^{\mathrm{K}} \mathrm{y_{n,i,k}} \cdot \mathrm{logP_s(n,i,k)} \tag{2}$$

where $y_{n,i,k}$ is a one-hot label. If the ground-truth class of pixel i is k, then $y_{n,i,k} = 1$; otherwise, it is 0. $P_s^l(n,i,k)$ denotes the probability that the student assigns pixel i in a labeled sample to class k. For unlabeled data, MADM provides the disagreement-aware pseudo-label consistency loss L_{MADM}, while EUEM provides the student-side evidential uncertainty regularization loss L_{EUEM}. The overall objective is

$$\mathcal{L} = \mathcal{L}_{\text{sup}} + \lambda_{\text{MADM}} \cdot \mathcal{L}_{\text{MADM}} + \lambda_{\text{EUEM}} \cdot \mathcal{L}_{\text{EUEM}}, \tag{3}$$

where λ_{MADM} and λ_{EUEM} are the weights of the MADM and EUEM losses, respectively.

3.3. Multi-Teacher Adaptive Disagreement-Aware Module

In multi-teacher semi-supervised learning, pseudo-labels are commonly generated by averaging or voting over teacher predictions. Simple averaging, however, ignores the boundary ambiguity, class confusion, and minority-class uncertainty revealed by inter-teacher disagreement. We therefore propose MADM, which explicitly models differences among teacher predictions as pixel-level reliability signals and uses them for pseudo-label fusion and region-adaptive supervision.

Inter-teacher disagreement is essentially a measure of dissimilarity among class-probability distributions. Compared with Euclidean or cosine distance, JSD is symmetric, bounded, and suitable for comparing probability distributions, making it a stable measure of the difference between two teacher predictions. A small JSD indicates similar class probabilities and a relatively reliable pseudo-label, whereas a large JSD suggests that the pixel may lie on a class boundary, in an ambiguous region, or within a minority class. We therefore use JSD to construct a pixel-level disagreement matrix among teachers.

For pixel i, let the predictive probabilities of the mth and nth teachers be $p_m(i)$ and $p_n(i)$, respectively. Their JSD is defined as

$$D_{mn}(i) = \frac{1}{2} KL(p_m(i) \parallel \bar{p}_{mn}(i)) + \frac{1}{2} KL(p_n(i) \parallel \bar{p}_{mn}(i)), \tag{4}$$

where $\bar{p}_{mn}(i) = \frac{1}{2}[p_m(i) + p_n(i)]$, and $KL(\cdot \parallel \cdot)$ denotes the Kullback–Leibler divergence. When $KL(\cdot \parallel \cdot)$ uses a base-2 logarithm, $D_{mn}(i)$ lies in $[0,1]$. A larger $D_{mn}(i)$ indicates greater predictive disagreement between the two teachers at pixel i, whereas a smaller $D_{mn}(i)$ indicates stronger agreement.

To generate reliable pseudo-labels from multi-teacher disagreement, we compute each teacher's mean disagreement with the remaining teachers and its reliability weight. Based on pairwise disagreement, d_m denotes the mean disagreement of the mth teacher with the other teachers at pixel i; a lower value indicates a more reliable teacher at that

pixel. A normalized exponential function then gives the teacher reliability weight w_m. The quantities d_m and w_m are defined as

$$d_m(i) = \frac{1}{M-1} \sum_{n \neq m} D_{mn}(i), \tag{5}$$

$$w_m(i) = \frac{\exp(-d_m(i)/\tau_d)}{\sum_{r=1}^{M} \exp(-d_r(i)/\tau_d)} \tag{6}$$

where τ_d is the softmax temperature that controls the sharpness of the mapping from teacher disagreement to reliability weights. To characterize pseudo-label reliability according to the degree of disagreement, we further compute the soft pseudo-label q_i, overall disagreement d_i, and pseudo-label confidence c_i

$$q_i = \sum_{m=1}^{M} w_m(i) p_m(i), \tag{7}$$

$$d_i = \sum_{m=1}^{M} w_m(i) d_m(i), \tag{8}$$

$$c_i = \max_k q_{i,k}, \tag{9}$$

where q_i is the disagreement-aware weighted aggregation of multiple teacher predictions, d_i is the overall disagreement after fusion, and c_i is the maximum class probability of the fused pseudo-label.

According to the overall disagreement and pseudo-label confidence, unlabeled pixels are partitioned into high-consistency, moderate-disagreement, high-disagreement-but-usable, and high-disagreement/low-confidence ignored regions. Given the disagreement threshold $0 \leq \eta_1 < \eta_2 \leq 1$, pseudo-label confidence threshold τ_p, and confidence threshold τ_h for high-disagreement regions, the regions are defined as $R_{low} = \{i \mid d_i \leq \eta_1, c_i \geq \tau_p\}$, $R_{mid} = \{i \mid \eta_1 < d_i \leq \eta_2, c_i \geq \tau_p\}$, $R_{high} = \{i \mid d_i > \eta_2, c_i \geq \tau_h\}$, $R_{ign} = \{i \mid d_i > \eta_2, c_i < \tau_h\}$. Here, R_{low} is the reliable high-consistency region, R_{mid} is the moderate-disagreement region, R_{high} is the usable high-disagreement region with sufficient pseudo-label confidence, and R_{ign} is the ignored high-disagreement/low-confidence region. This partition avoids strong pseudo-label supervision for highly unreliable pixels. The high-consistency region receives strong supervision from the fused soft pseudo-label, the moderate-disagreement region receives disagreement-attenuated soft-label consistency, and the high-disagreement region receives conservative boundary-aware supervision because it frequently contains boundaries or ambiguous classes.

To construct a boundary-aware weight, Sobel gradients are first computed on the stop-gradient fused soft pseudo-label q. Because q is a K-channel class-probability map, horizontal and vertical Sobel responses are computed separately for each class channel and then summed over all channels to obtain the semantic boundary strength:

$$B_i = \sum_{k=1}^{K} \sqrt{G_x(q_k)_i^2 + G_y(q_k)_i^2}, \tag{10}$$

where $G_x(\cdot)$ and $G_y(\cdot)$ are the horizontal and vertical Sobel operators, respectively. The response B_i is normalized to obtain the boundary-enhancement weight. The unlabeled consistency loss of MADM is then defined as

$$L_{MADM} = \sum_{r \in \{low, \text{mid}, \text{high}\}} \frac{1}{|R_r|} \sum_{i \in R_r} \omega_r(i) H(q_i, P_s^s(i)), \tag{11}$$

$$\omega_r(i) = \begin{cases} 1 & (i \in R_{low}) \\ \lambda_{mod} \cdot \exp(-d_i) & (i \in R_{mid}) \\ \lambda_{high} \cdot B_i & \left(i \in R_{high}\right) \end{cases} \tag{12}$$

where $H(q,p) = -\sum_{k=1}^{K} q_k \log(p_k + \varepsilon)$ denotes soft-label cross-entropy, and λ_{mod} and λ_{high} control the relative strengths of the moderate- and high-disagreement losses, respectively.

Pixels in R_{ign} do not contribute to the MADM pseudo-label consistency loss. Consequently, reliable regions provide stable supervision, while disagreement regions participate with weaker or boundary-enhanced but controlled supervision. This design alleviates boundary confusion, small-object mislabeling, and pseudo-label noise propagation in RS imagery.

3.4. Evidential Uncertainty Estimation Module

A conventional softmax segmentation head provides only normalized class probabilities and cannot distinguish high confidence supported by strong evidence from a high probability produced by normalizing weak evidence. Boundaries, small objects, shadows, and interclass-similarity regions in RS images often lack sufficient discriminative evidence. We therefore introduce EUEM as a parallel auxiliary branch of the student network. It predicts class evidence from multiscale encoder features and maps the evidence to Dirichlet parameters, allowing the model to estimate both class probability and the corresponding evidence strength and uncertainty.

Let the student encoder output L levels of multiscale features $F_s = \{F_s^1, F_s^2, \ldots, F_s^L\}$. To unify their channel dimensions and spatial resolutions, a 1×1 convolution first maps the channels, and the features are upsampled to a common spatial size $\hat{F}_s^l = Up\left(Conv_{1\times1}^l\left(F_s^l\right)\right)$. The aligned features are then concatenated and passed through a 3×3 convolution to obtain the fused evidential feature:

$$F_E = Conv_{3\times3}\left(\left[\hat{F}_s^1, \hat{F}_s^2, \ldots, \hat{F}_s^L\right]\right). \tag{13}$$

Let $h_E(\cdot)$ denote the evidence prediction head in EUEM, which maps the fused multiscale features to class-level evidential responses. Specifically, $h_E(\cdot)$ comprises lightweight convolutional layers: a 3×3 convolution for local spatial context, a nonlinear activation for feature enhancement, and a 1×1 convolution that maps the channels to the number of semantic classes K. For pixel i, the evidence head first outputs raw class responses and converts them into non-negative evidence $e_i = Softplus(h_E(F_E)_i)$ using Softplus. Unlike a softmax segmentation head, it estimates the evidence strength of each class instead of directly outputting normalized probabilities. The non-negative evidence is then converted to Dirichlet parameters $\alpha_{i,k} = e_{i,k} + 1$, which are always positive. When $e_i = 0$, $\alpha_{i,k} = 1$, corresponding to a uniform noninformative prior with no explicit class preference. The total Dirichlet strength at pixel i is

$$S_i = \sum_{k=1}^{K} \alpha_{i,k}. \tag{14}$$

The expected class probability under the Dirichlet distribution and the pixel-level uncertainty produced by EUEM are then obtained as

$$P_E(i,k) = \frac{\alpha_{i,k}}{S_i}, \tag{15}$$

$$u_i = \frac{K}{S_i}, \tag{16}$$

When little evidence is available, S_i approaches K and u_i approaches 1, indicating high uncertainty. As more class evidence is accumulated, S_i increases and u_i decreases, indicating a more reliable prediction.

To provide stable supervision for EUEM at the beginning of training, we first introduce a supervised evidential loss on labeled samples, enabling the evidential branch to learn class-evidence representations directly from the ground truth

$$L_E^{sup} = -\frac{1}{N_L|\Omega_l|}\sum_{n=1}^{N_L}\sum_{i\in\Omega_l}\sum_{k=1}^{K} y_{n,i,k}\log\Big(P_E^l(n,i,k)\Big). \tag{17}$$

For an unlabeled sample, the main student branch outputs class probabilities P_s^s, whereas EUEM outputs the expected Dirichlet probabilities P_E and pixel-level uncertainty u. To prevent low-confidence predictions from corrupting evidence learning, only high-confidence pixels from the main student branch are selected as consistency targets

$$M_i = \mathbf{1}\left[\max_k P_s^s(i,k) \geq \tau_e(t)\right], \tag{18}$$

where M_i is the high-confidence pixel mask and $\tau_e(t)$ is the dynamic confidence threshold for evidential consistency. By default, the threshold grows linearly from τ_{min} to τ_{max}. The lower value of $\tau_e(t) = \tau_{min} + (\tau_{max} - \tau_{min}) \cdot r(t)$. early in training retains sufficient valid supervisory pixels, whereas the higher value later in training reduces interference from low-quality predictions. The high-confidence consistency loss of EUEM is

$$L_{ec} = \frac{1}{\sum_i M_i + \varepsilon}\sum_i M_i KL(stopgrad(P_s^s(i)) \parallel P_E(i)), \tag{19}$$

Here, $stopgrad(\cdot)$ denotes the stop-gradient operation; the main-branch prediction serves as a fixed soft target and this loss does not directly update the primary segmentation head. The stop-gradient operation applies only to the main prediction path.

If EUEM shares features with the student encoder and no stop-gradient is applied to input feature F_E, the loss still backpropagates through $P_E \to h_E \to F_E$, updating the EUEM evidence head and the shared student encoder. Applying stop-gradient to F_E would instead restrict optimization to the EUEM head. We adopt the former setting, allowing evidential consistency to improve evidence-aware shared representations without directly perturbing the main prediction head.

To prevent low-confidence pixels from producing excessive evidence, a uniform-prior regularizer further encourages their Dirichlet distributions to approach a noninformative prior: $Dir(\mathbf{1})$.

$$L_{prior} = \frac{1}{\sum_i (1 - M_i) + \varepsilon}\sum_i (1 - M_i) KL(Dir(\alpha_i) \parallel Dir(\mathbf{1})), \tag{20}$$

When α_i approaches an all-ones vector, the Dirichlet distribution approaches a uniform noninformative prior, indicating insufficient evidence for assigning the pixel to any class. Minimizing this term prevents low-confidence pixels from accumulating excessive class evidence and reduces the risk of reinforcing erroneous predictions. The final EUEM loss is

$$L_{EUEM} = L_E^{sup} + L_{ec} + \lambda_p L_{prior}, \tag{21}$$

where λ_p is the weight of the low-confidence prior regularizer.

EUEM and MADM are complementary. MADM measures pseudo-label-source reliability from teacher-side predictive differences, whereas EUEM measures evidence sufficiency

from the student's evidence distribution. The former improves pseudo-label generation, and the latter represents uncertainty during student learning. Together, they mitigate confirmation bias at the levels of supervisory signals and learning states.

Algorithm 1 summarizes the training procedure. At each iteration, the supervised loss is first computed on labeled samples, and multiple EMA teachers predict on weakly augmented views. MADM then computes JSD-based disagreement and performs disagreement-aware fusion and region partitioning; EUEM estimates Dirichlet evidence to construct high-confidence consistency and low-confidence prior constraints. Finally, the joint loss updates the student and EUEM branch, after which the teacher parameters are updated using distinct EMA decay rates.

Algorithm 1 Training Procedure of MTDPNet

Input:
D_L, D_U, student $S(\cdot;\theta)$, teachers $\{T_m(\cdot;\theta_m^T)\}_{m=1}^M$, EUEM H_E.
Output:
Optimized student network S^*.
Initialize:
S, $\{T_m\}_{m=1}^M$, and H_E.
while not converged do
  $B_L \subset D_L$, $B_U \subset D_U$;
  L_{sup} on B_L;
  $x_w = A_w(B_U)$, $x_s = A_s(B_U)$;
  for $m = 1$ to M do
   $P_m = G_{w\to s}(T_m(x_w;\theta_m^T))$;
   end for
  $(P_s^s, F_s) = S(x_s;\theta)$;
  $D_{mn}(i) = \frac{1}{2}KL(p_m(i) \parallel \bar{p}_{mn}(i)) + \frac{1}{2}KL(p_n(i) \parallel \bar{p}_{mn}(i))$;
  $d_m(i) = \frac{1}{M-1}\sum_{n\neq m} D_{mn}(i)$, $w_m(i) = \frac{\exp(-d_m(i)/\tau_d)}{\sum_{r=1}^M \exp(-d_r(i)/\tau_d)}$;
  $q_i = \sum_{m=1}^M w_m(i)p_m(i)$, $d_i = \sum_m w_m(i)d_m(i)$, $c_i = \max_k q_{i,k}$;
  Divide pixels into R_{low}, R_{mid}, R_{high}, R_{ign}, and compute L_{MADM};
  $e_i = Softplus(H_E(F_s)_i)$, $\alpha_{i,k} = e_{i,k} + 1$, $S_i = \sum_{k=1}^K \alpha_{i,k}$;
  $P_E(i,k) = \frac{\alpha_{i,k}}{S_i}$, $u_i = \frac{K}{S_i}$;
  $M_i = 1[\max_k P_s^s(i,k) \geq \tau_e(t)]$;
  $L_{EUEM} = L_E^{sup} + L_{ec} + \lambda_p L_{prior}$;
  $\mathcal{L} = \mathcal{L}_{\text{sup}} + \lambda_{\text{MADM}} \cdot \mathcal{L}_{\text{MADM}} + \lambda_{\text{EUEM}} \cdot \mathcal{L}_{\text{EUEM}}$;
  Update S and H_E by gradient descent;
  Update teachers by EMA: $\theta_{m,t}^T = \alpha_m \theta_{m,t-1}^T + (1-\alpha_m)\theta_t$;
end while
return S^*

4. Results

4.1. Experimental Settings and Evaluation Metrics

All experiments are performed on an NVIDIA GeForce RTX 4090 GPU with 16 GB of memory (16,384 MiB, as verified by nvidia-smi) using PyTorch 2.9.1, torchvision 0.24.1, and CUDA 12.6. The batch size and number of data-loading workers are set to 8 and 4, respectively. In the main experiments, U-Net++ is used for the student and three EMA teachers, without ImageNet or other external pretrained weights. The student is optimized with AdamW for 30,000 iterations using an initial learning rate of 3×10^{-4} and a weight

decay of 1×10^{-4}. The supervised branch combines cross-entropy and Dice losses, with the Dice-loss weight set to 0.3. The EMA decay coefficients of the three teachers are 0.990, 0.995, and 0.999. The MADM loss weight λ_{MADM} is 0.3, and the auxiliary weights for the moderate- and high-disagreement regions are 0.2 and 0.1, respectively. According to the sensitivity study, EUEM is used as a weak evidential regularizer with $\lambda_{EUEM} = 0.1$. To make the settings reproducible, the thresholds and temperatures of the two modules are fixed as follows: the temperature τd of the teacher reliability weights is 0.1, the disagreement thresholds η_1 and η_2 are 0.2 and 0.5, the pseudo-label confidence threshold τ_p is 0.95, the confidence threshold for high-disagreement regions τ_h is 0.90, the dynamic confidence threshold of EUEM increases linearly from $\tau_{min} = 0.95$ to τmax = 0.98 over training, and the weight of the low-confidence prior regularizer λ_p is 0.1. All other settings remain unchanged except in the corresponding sensitivity experiments.

To assess statistical reliability, every comparison was repeated over three independent runs with different random seeds. The results in Tables 1 and 2 are the means over the three runs; the standard deviation of mIoU across the runs did not exceed ±0.41 in the 5% labeling setting and ±0.29 in the 25% labeling setting for any method, and the class-wise IoU and mF1 values are reported from the run with the median mIoU. Conclusions are drawn from the mean values. Performance is evaluated using classwise intersection over union (IoU), mean IoU (mIoU), classwise F1 score, and mean F1 score (mF1). For class c, IoU and F1 are computed as follows:

$$\mathrm{IoU} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP} + \mathrm{FN}}, \tag{22}$$

$$\mathrm{F1} = \frac{2\mathrm{TP}}{2\mathrm{TP} + \mathrm{FP} + \mathrm{FN}}, \tag{23}$$

where TP, FP, and FN denote the numbers of true-positive, false-positive, and false-negative pixels, respectively.

Table 1. Quantitative comparison on the ISPRS Vaihingen dataset under different labeling ratios (%).

Labeled Ratio	Model	Imp Surf.	Building	Low Veg.	Tree	Car	mIoU	mF1
5%	Baseline	72.35	80.36	60.71	54.05	30.86	59.67	73.19
	Mean Teacher [2]	73.20	78.85	64.56	49.43	35.51	60.31	73.95
	FixMatch [3]	70.53	82.68	62.29	54.16	37.54	61.44	74.97
	PGNet [33]	73.85	82.55	64.35	53.15	36.85	62.15	75.71
	CPS [8]	72.50	82.25	65.06	51.03	37.71	61.71	75.10
	U^2PL [9]	74.14	78.69	65.20	50.97	34.75	60.75	74.25
	LPCR [12]	69.97	83.32	63.13	53.40	37.58	61.48	74.98
	CWD-Teacher [17]	72.95	76.90	64.61	53.08	36.26	60.76	74.47
	DWL [32]	70.82	82.57	60.48	54.15	38.13	61.23	74.84
	RS-MUCA [16]	74.93	78.08	63.99	51.36	35.84	60.84	74.41
	MTDPNet	74.53	83.18	65.09	54.96	39.42	63.44	76.51
25%	Baseline	74.72	82.17	45.27	77.08	52.29	66.31	78.76
	Mean Teacher [2]	74.93	80.57	46.37	74.46	54.27	66.12	78.80
	FixMatch [3]	72.32	83.46	45.16	79.25	55.36	67.11	79.37
	PGNet [33]	73.25	84.35	44.65	80.85	56.05	67.83	79.98
	CPS [8]	73.99	83.73	46.63	74.82	55.23	66.88	79.31
	U^2PL [9]	75.26	79.80	46.40	75.39	52.90	65.95	78.64
	LPCR [12]	71.35	83.69	45.59	81.08	54.99	67.34	79.51
	CWD-Teacher [17]	75.12	78.06	46.86	77.55	54.46	66.41	79.03
	DWL [32]	73.00	84.74	43.74	80.63	56.34	67.69	79.67
	RS-MUCA [16]	75.30	80.44	46.44	77.03	55.24	66.89	79.34
	MTDPNet	75.59	83.23	47.23	78.32	56.54	68.18	80.24

Table 2. Quantitative comparison on the ISPRS Potsdam dataset under different labeling ratios (%).

Labeled Ratio	Model	Imp Surf.	Building	Low Veg.	Tree	Car	mIoU	mF1
5%	Baseline	64.52	42.16	39.80	39.05	51.71	47.45	63.80
	Mean Teacher [2]	64.82	43.83	43.69	34.93	53.83	48.22	64.44
	FixMatch [3]	62.31	46.82	41.58	40.82	56.02	49.51	65.82
	PGNet [33]	64.50	49.50	42.50	40.50	55.35	50.47	66.52
	CPS [8]	64.13	51.24	43.20	36.54	56.04	50.23	66.32
	U^2PL [9]	65.28	43.65	43.31	36.45	53.05	48.35	64.59
	LPCR [12]	61.75	47.46	42.42	43.06	56.06	50.15	66.47
	CWD-Teacher [17]	64.57	41.88	43.74	38.58	54.58	48.67	64.93
	DWL [32]	62.46	47.57	39.63	41.67	56.47	49.56	65.83
	RS-MUCA [16]	66.74	43.25	43.31	37.05	54.35	48.94	65.07
	MTDPNet	65.75	47.76	43.82	40.06	57.37	50.95	67.01
25%	Baseline	65.09	67.62	56.76	56.00	65.58	62.21	76.59
	Mean Teacher [2]	66.91	65.41	58.30	53.90	68.33	62.57	76.83
	FixMatch [3]	63.86	67.86	56.65	58.25	69.98	63.32	77.42
	PGNet [33]	64.50	69.00	57.80	60.50	70.10	64.38	78.18
	CPS [8]	65.67	67.27	58.26	54.96	69.99	63.23	77.32
	U^2PL [9]	68.32	64.72	58.41	54.91	67.04	62.68	76.93
	LPCR [12]	63.31	68.51	57.50	61.50	70.03	64.17	78.08
	CWD-Teacher [17]	66.96	63.76	58.65	57.85	69.38	63.32	77.45
	DWL [32]	64.25	69.85	54.94	60.34	70.67	64.01	77.90
	RS-MUCA [16]	69.03	65.03	57.12	56.22	69.05	63.29	77.37
	MTDPNet	67.30	68.80	58.89	58.49	71.32	64.96	78.63

4.2. Comparative Experiments

4.2.1. Vaihingen Dataset

MTDPNet is compared with ten representative semi-supervised methods: Mean Teacher, FixMatch, CPS, U^2PL, LPCR, CWD-Teacher, DWL, RS-MUCA, and PGNet [33]. Table 1 reports the quantitative results on Vaihingen. With 5% labeled data, MTDPNet obtains an mIoU of 63.44% and an mF1 of 76.51%, improving the supervised baseline by 3.77 and 3.32 percentage points and the best competing results by 1.29 and 0.80 points, respectively. With 25% labeled data, MTDPNet achieves 68.18% mIoU and 80.24% mF1, exceeding the supervised baseline by 1.87 and 1.48 points and the best competing results by 0.35 and 0.26 points. At the class level, MTDPNet yields the highest vehicle IoU in both settings, with gains of 8.56 and 4.25 points over the supervised baseline.

Figure 3 compares row-normalized confusion matrices of the supervised baseline and MTDPNet on Vaihingen. MTDPNet increases the correct classification rates of impervious surfaces and buildings from 88.1% and 87.8% to 90.2% and 88.8%, while reducing impervious surfaces misclassified as low vegetation from 4.3% to 2.7%. The largest improvement occurs for vehicles: the correct classification rate increases from 58.9% to 65.3%, and confusion with impervious surfaces and low vegetation decreases from 35.7% to 30.3% and from 1.7% to 0.4%, respectively. The correct classification rate of low vegetation changes slightly from 54.9% to 54.6%, although its confusion with trees decreases from 36.0% to 33.6%; the tree rate remains 93.3%.

Figure 4 presents qualitative results on Vaihingen with 25% labeled data. The columns show the input image, ground truth, Mean Teacher, FixMatch, ClassMix, PGNet, CPS, and MTDPNet. White, blue, cyan, green, and yellow denote impervious surfaces, buildings, low vegetation, trees, and vehicles, respectively; red boxes indicate representative difficult regions. MTDPNet preserves large building contours more completely and reduces interior holes and missing edges. It also produces fewer fragments and isolated artifacts where low vegetation and trees overlap and retains more continuous responses for small vehicles

near roads. Mean Teacher and FixMatch exhibit omissions or class confusion along several boundaries. ClassMix, PGNet, and CPS improve the completeness of large regions but remain locally fragmented at complex vegetation boundaries.

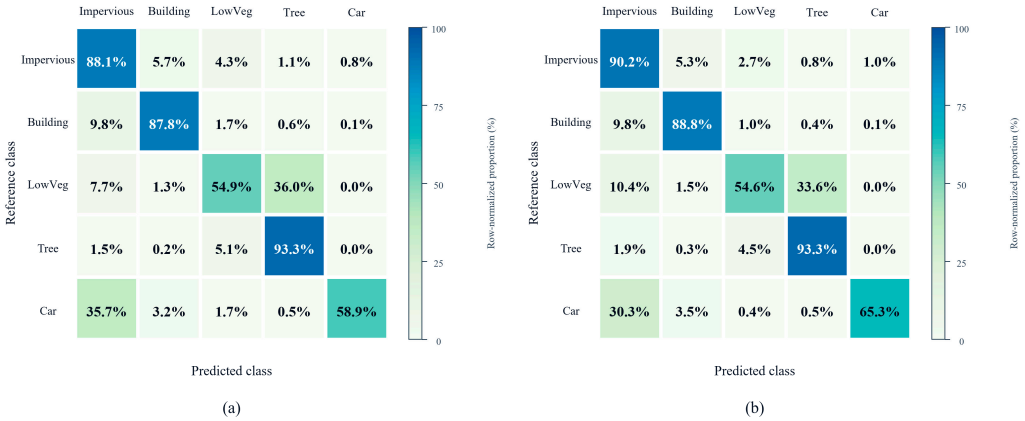

Figure 3. Row-normalized confusion matrices on the Vaihingen dataset. (**a**) Supervised baseline. (**b**) MTDPNet.

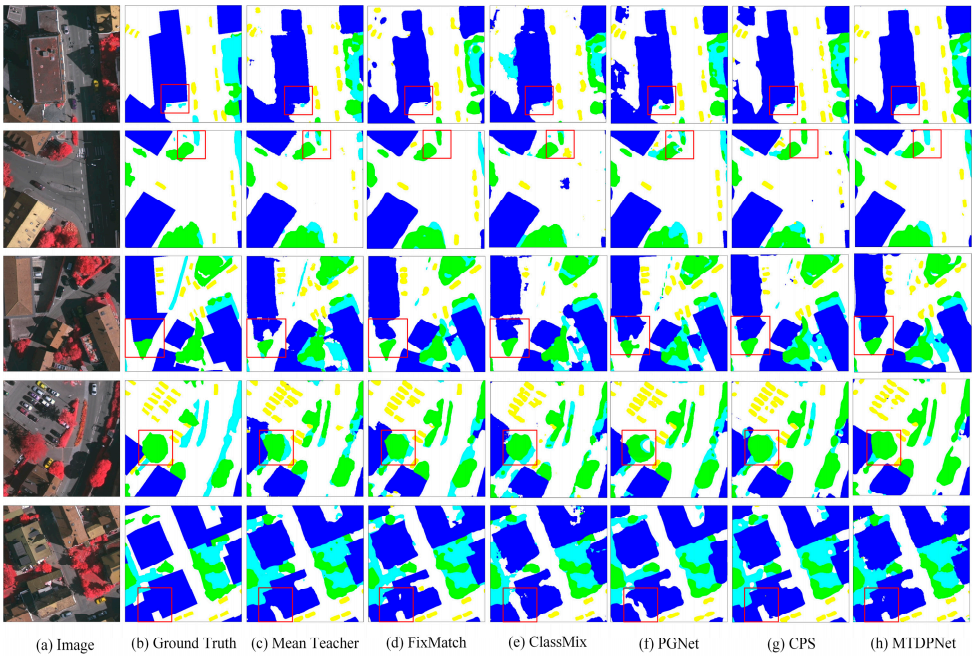

Figure 4. Semantic segmentation results on the ISPRS Vaihingen dataset with 25% labeled data. White, blue, cyan, green, and yellow denote impervious surfaces, buildings, low vegetation, trees, and vehicles, respectively. Red boxes mark the representative regions that best illustrate the segmentation quality, in which MTDPNet produces more accurate and complete predictions.

4.2.2. Potsdam Dataset

Table 2 summarizes the results on Potsdam. With 5% labeled data, MTDPNet obtains 50.95% mIoU and 67.01% mF1, improving the supervised baseline by 3.50 and 3.21 percentage points and the best competing results by 0.48 and 0.49 points. Its vehicle IoU reaches 57.37%, 0.90 points above the second-best result, while building and low-vegetation IoUs improve by 5.60 and 4.02 points over the supervised baseline. With 25% labeled data, MTDPNet reaches 64.96% mIoU and 78.63% mF1, outperforming the supervised baseline by 2.75 and 2.04 points and the best competing results by 0.58 and 0.45 points. Its vehicle IoU is 71.32%, exceeding the second-best result by 0.65 points.

Figure 5 further compares class-confusion patterns of the supervised baseline and MTDPNet on Potsdam. MTDPNet increases the correct classification rates of impervious surfaces and vehicles from 82.1% and 80.5% to 87.5% and 86.5%. For vehicles, confusion with impervious surfaces and buildings decreases from 9.7% and 4.9% to 7.8% and 2.3%, respectively. The correct tree rate increases from 74.6% to 75.9%, while confusion with low vegetation decreases from 20.5% to 18.9%. Conversely, the correct rates of buildings and low vegetation change slightly from 72.2% and 76.9% to 71.4% and 76.2%, showing that classwise gains remain heterogeneous in complex scenes.

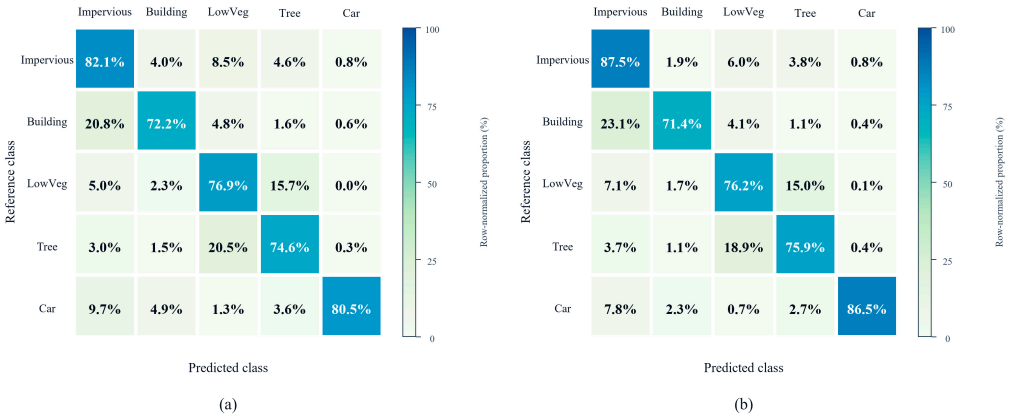

Figure 5. Row-normalized confusion matrices on the Potsdam dataset. (**a**) Supervised baseline. (**b**) MTDPNet.

Figure 6 shows qualitative results on Potsdam with 25% labeled data. The columns correspond to the input image, ground truth, U^{2}PL, LPCR, CWD-Teacher, DWL, RS-MUCA, and MTDPNet. In the red-boxed regions, MTDPNet produces more continuous predictions over large impervious-surface and low-vegetation areas and preserves clearer building boundaries. Vehicle locations and shapes near roads and buildings also better match the ground truth. U^{2}PL and LPCR exhibit local omissions or fragmentation. CWD-Teacher and DWL improve regional consistency but retain boundary shifts or confusion near class transitions. RS-MUCA generally approaches the ground truth but still misclassifies several building–vegetation boundaries.

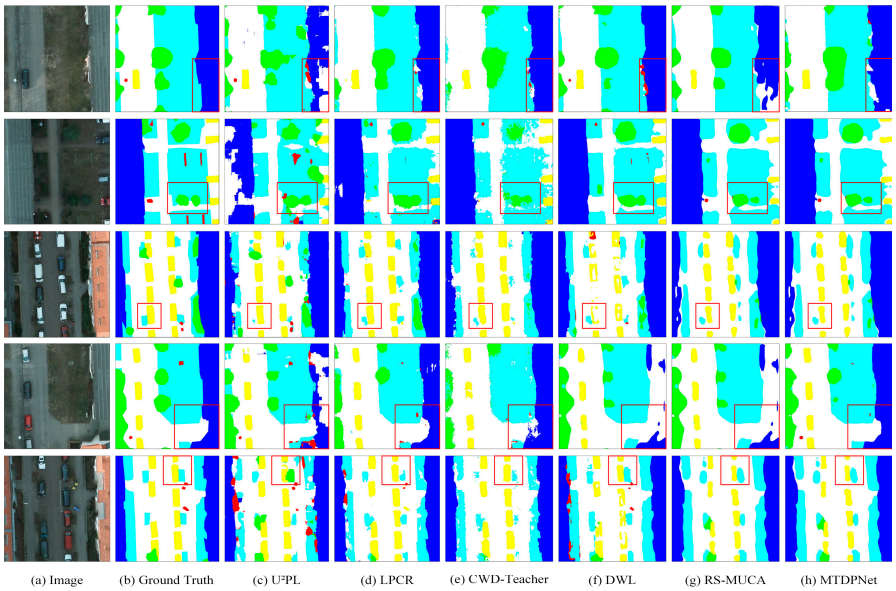

Figure 6. Semantic segmentation results on the ISPRS Potsdam dataset with 25% labeled data. White, blue, cyan, green, and yellow denote impervious surfaces, buildings, low vegetation, trees, and vehicles,

respectively. Red boxes mark the representative regions that best illustrate the segmentation quality, in which MTDPNet produces more accurate and complete predictions.

4.3. Ablation Studies

To make the ablation and sensitivity studies reproducible, the dataset, labeling ratio, split, and random seed of each table are listed below. Unless otherwise stated, all these experiments share the same tile-level split and random seed as the main experiments. The ablation studies proceed from global to component-level analyses. We first evaluate the individual and joint contributions of MADM and EUEM, then examine multi-teacher fusion, disagreement weighting, and reliable-region selection within MADM, followed by the number and EMA time scales of teachers, the EUEM weight, and the spatial distributions of teacher confidence and evidential uncertainty. Because each experiment controls different variables, conclusions are drawn only from comparisons within the same table; absolute values are not compared directly across different tables.

4.3.1. Overall Contributions of MADM and EUEM

Table 3 reports the overall module ablation. Adding MADM to the supervised baseline raises mIoU from 65.90% to 66.99%, a gain of 1.09 percentage points, and mF1 from 78.50% to 79.40%. Vehicle IoU shows the largest gain, increasing from 45.00% to 47.70% (+2.70 points). Impervious-surface, building, and low-vegetation IoUs improve by 1.10, 0.90, and 0.60 points, respectively, while tree performance remains nearly unchanged.

Table 3. Ablation of MADM and EUEM (%).

Method	MADM	EUEM	Imp Surf.	Building	Low Veg.	Tree	Car	mIoU	mF1
Baseline			76.60	85.50	67.00	55.60	45.00	65.90	78.50
MADM	√		77.70	86.40	67.60	55.50	47.70	66.99	79.40
EUEM		√	76.85	85.65	66.85	54.92	45.09	65.87	78.60
MADM + EUEM	√	√	77.60	86.10	68.40	57.40	47.30	67.37	79.70

Using EUEM alone yields 65.87% mIoU, essentially the same as the 65.90% supervised baseline, indicating that evidential uncertainty estimation cannot replace pseudo-label-based consistency learning. Combining EUEM with MADM increases mIoU to 67.37%, gains of 1.47 and 0.38 points over the supervised baseline and MADM alone, respectively, and produces an mF1 of 79.70%. Low-vegetation, tree, and vehicle IoUs exceed the baseline by 1.40, 1.80, and 2.30 points.

4.3.2. Mechanism Ablation of MADM

The overall ablation validates MADM but does not distinguish the effects of multi-teacher fusion, disagreement weighting, and reliable-region selection. Table 4 incrementally adds these components. Setting A is a single-teacher pseudo-label baseline with 65.90% mIoU. Equal-weight multi-teacher fusion in setting B increases mIoU to 66.13% (+0.23 points). Adding JSD-based disagreement weighting in setting C raises mIoU to 66.48% (+0.35 points). The complete setting D further introduces reliable-region masks and achieves 66.99% (+0.51 points).

To validate the choice of JSD as the disagreement measure, we replaced JSD inside MADM with the mean KL divergence and the cosine distance while keeping all other components unchanged. The detailed results are presented in Table 5.

Table 4. Incremental ablation of the key MADM components.

Setting	Teacher Ensemble	JSD Weighting	Region Masking	mIoU
A				65.90
B	√			66.13
C	√	√		66.48
D (MADM)	√	√	√	66.99

Table 5. Sensitivity to the disagreement measure (%).

Disagreement Measure	mIoU (%)	mF1 (%)
Cosine distance	66.52	79.05
Mean KL divergence	66.41	78.96
JSD (used in MADM)	66.99	79.40

The number of teachers and their EMA time scales jointly determine ensemble diversity and stability. The discussion first reports the trend over the teacher count with the nested EMA set and then reports the sensitivity to the EMA combination with three teachers. Any performance change in the four-teacher setting cannot be attributed solely to the teacher count, because that setting introduces the additional coefficient 0.970. As shown in Table 6, increasing the number of teachers from one to two raises mIoU from 66.45% to 66.87%; three teachers further improve it to 67.34%, 0.89 points above the single-teacher setting. With four teachers, mIoU decreases to 66.62%, 0.72 points below the three-teacher result. The mean disagreement values for two, three, and four teachers are 0.4131, 0.4102, and 0.4109, respectively. Training parameters increase from 18.07 M to 44.86 M with the number of teachers.

Table 6. Sensitivity to the Number of Teachers and EMA Decay.

Teachers	EMA Decay	mIoU (%)	Params (M)
1	0.999	66.45	18.07
2	0.995/0.999	66.87	27.00
3	0.990/0.995/0.999	**67.34**	35.93
3	0.985/0.990/0.995	67.02	35.93
3	0.999/0.999/0.999	66.53	35.93
4	0.970/0.990/0.995/0.999	66.62	44.86

4.3.3. Sensitivity to the EUEM Weight

With λMADM fixed at 0.3, Table 7 reports results for different EUEM weights. Increasing λEUEM from 0 to 0.05 changes mIoU only from 66.99% to 66.96%. At λEUEM = 0.1, mIoU reaches 67.37%, 0.38 points above the result without EUEM, and the tree and low-vegetation IoUs reach 57.40% and 68.40%, respectively. Further increasing the weight to 0.2 decreases mIoU and mF1 to 66.72% and 79.20%. We use λEUEM = 0.1 in the remaining experiments.

Table 7. Sensitivity to the EUEM loss weight (%).

Weight	Imp Surf.	Building	Low Veg.	Tree	Car	mIoU	mF1
0	77.70	86.40	67.60	55.50	47.70	66.99	79.40
0.05	77.90	86.10	68.10	55.70	47.00	66.96	79.30
0.1	77.60	86.10	68.40	57.40	47.30	67.37	79.70
0.2	77.40	86.00	67.80	55.40	47.00	66.72	79.20

Figure 7 visualizes the sensitivity to the number and EMA configuration of teachers in Figure 7a and to the EUEM weight in Figure 7b. The proposed framework performs best with moderate teacher diversity and evidential regularization. Three EMA teachers yield 67.34% mIoU, whereas adding a fourth teacher increases training cost and reduces accuracy. At λEUEM = 0.1, the model achieves 67.37% mIoU and 79.70% mF1, confirming that a weak EUEM weight better balances the evidential constraint and the primary segmentation objective.

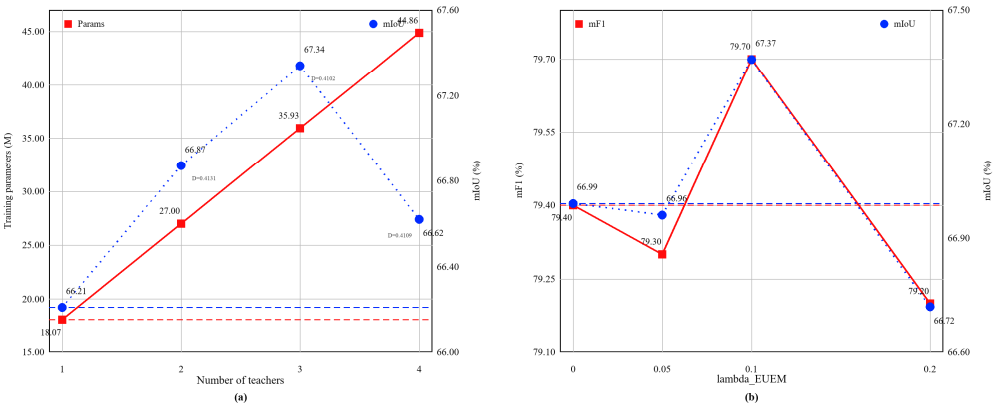

Figure 7. Parameter sensitivity. (**a**) Number of teachers and EMA configurations. (**b**) EUEM loss weight. The horizontal dashed lines mark the reference values for the single-teacher setting in (**a**) and λEUEM = 0 in (**b**).

4.3.4. Confidence and Uncertainty Visualization

To examine the spatial reliability signals provided by MADM and EUEM, Figure 8 shows the input image, ground truth, model prediction, evidential uncertainty, and teacher confidence for four representative examples. High uncertainty and low teacher confidence mainly coincide at land-cover boundaries, narrow structures, small vehicles, and ambiguous transitions between trees and low vegetation, whereas large, correctly predicted homogeneous regions show the opposite pattern. Prediction errors around vegetation boundaries, building contours, and vehicle neighborhoods are likewise associated with high uncertainty and low confidence.

4.4. Adaptability to Segmentation Networks and Complexity Analysis

To determine whether MTDPNet depends on a particular segmentation architecture, the same semi-supervised framework is applied without external pretraining to U-Net, DeepLabV3+, U-Net++, SegFormer, and a ResNet-50-based segmentation model. Complexity is measured for a single student segmentation network with an input size of $1 \times 3 \times 256 \times 256$ and excludes the EMA teachers and auxiliary EUEM branch, which are used only during training. Floating-point operations (FLOPs) include convolutions, linear layers, and attention-matrix multiplication; one multiplication and one addition are counted as two FLOPs.

As shown in Table 8, U-Net++ achieves the highest mIoU of 67.34%, outperforming the second-best SegFormer by 2.16 points. Its 8.93 M parameters are fewer than those of DeepLabV3+, SegFormer, and ResNet-50, indicating that its accuracy is not simply attributable to a larger parameter count. However, the nested U-Net++ decoder repeatedly performs convolution and feature fusion on high-resolution feature maps, resulting in the highest computational cost of 162.14 G FLOPs. SegFormer requires only 13.57 G FLOPs and is computationally more efficient, but its mIoU is 65.18%.

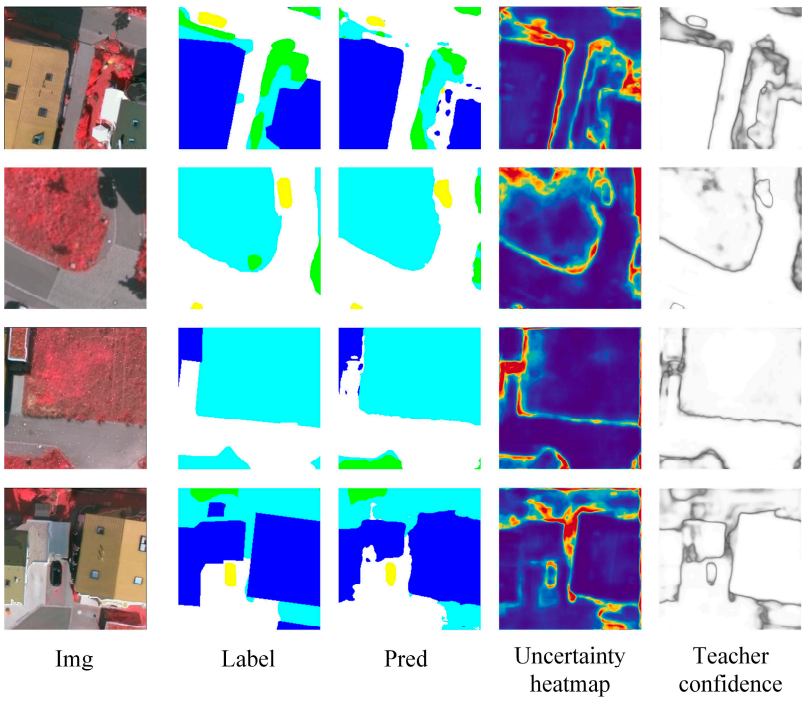

Figure 8. Spatial reliability visualization. From left to right: input image, ground truth, model prediction, evidential uncertainty, and teacher confidence. In the evidential-uncertainty maps, brighter regions indicate higher evidential uncertainty; in the teacher-confidence maps, brighter regions indicate higher teacher confidence.

Table 8. Performance and complexity of different segmentation networks.

Network	mIoU (%)	Parameters (M)	FLOPs (G)
U-Net	65.10	7.79	84.83
DeepLabV3+	62.25	60.58	75.69
U-Net++	67.34	8.93	162.14
SegFormer	65.18	14.73	13.57
ResNet50	64.85	27.45	21.36

5. Discussion

Across both benchmarks, MTDPNet achieves the highest mIoU, mF1, and vehicle IoU under both labeling ratios, and the gains are consistent rather than confined to a single image distribution. The improvement is not uniform across classes: it concentrates on vehicles and on several dominant confusion pathways, which indicates that the benefit arises from a more balanced class performance rather than from uniform superiority on every class. The larger gain under 5% labeled data shows that the framework is particularly valuable when supervision is scarce, where the reliability of pseudo-labels and the sufficiency of student evidence are the two limiting factors. The qualitative results agree with these trends: MTDPNet preserves large building contours, reduces fragmentation where low vegetation and trees overlap, and retains more continuous responses for small vehicles, indicating that the region-adaptive supervision and the evidential constraint act on the difficulties that standard consistency training handles poorly.

MADM and EUEM address different failure modes. MADM converts inter-teacher disagreement into an explicit pixel-level reliability signal and accounts for the principal part of the semi-supervised gain: the component analysis shows positive contributions from multi-teacher fusion, disagreement weighting, and reliable-region selection, with region masking providing the largest marginal gain. EUEM is designed as a weak auxiliary

regularizer on the student side rather than as an independent learning signal. Since it does not generate pseudo-labels, it is not expected to improve the baseline by itself, which is consistent with the ablation (65.87% versus 65.90%). Its contribution becomes visible when it is combined with MADM, where the mIoU increases from 66.99% to 67.37% and the difficult classes with insufficient evidence gain most (low vegetation +1.40, tree +1.80, and vehicle +2.30 points over the baseline). The two modules therefore act on complementary aspects of the same problem: whether the pseudo-label source can be trusted and whether the student prediction is supported by sufficient evidence.

Several design choices follow from the sensitivity analyses. JSD outperforms the mean KL divergence and the cosine distance as the disagreement measure, which shows that its symmetry and boundedness translate into better supervision in practice. Increasing the number of teachers is beneficial only up to a point: the nested comparison shows a monotonic gain from one to three teachers (66.45% to 67.34%), whereas the four-teacher setting, which additionally introduces the coefficient 0.970, decreases the mIoU to 66.62%, so stability and complementarity across EMA time scales matter more than the ensemble size alone. The EUEM weight behaves in the same non-monotonic way: a very small weight has limited influence, while an overly large weight competes with the main segmentation objective (66.72% at $\lambda_{EUEM} = 0.2$), and $\lambda_{EUEM} = 0.1$ gives the best trade-off.

Compared with related multi-teacher and uncertainty-aware methods, the difference lies in how the two sources of unreliability are modeled. Confidence-weighted dual-teacher fusion and cross-teacher–student attention aggregate teacher predictions into a single target, and uncertainty is typically used to filter or reweight pseudo-labels. MTDPNet instead treats inter-teacher disagreement as a supervisory attribute in its own right and combines it with student-side evidence sufficiency, so that difficult pixels receive calibrated supervision instead of being discarded or uniformly down-weighted.

The performance gains come with a training-only cost. With three teachers, the complete framework contains 35.93 M trainable parameters, about four times the 8.93 M student network, and the training time and peak memory increase by approximately 2.5× and 2× relative to the supervised baseline. Because the teachers receive no gradient and their batch-normalization statistics remain frozen, the overhead is concentrated in the forward passes of the teachers and in the multiscale fusion of EUEM. At inference, the teachers and the EUEM branch are discarded, and the deployed model remains the single student network with 8.93 M parameters and 162.14 G FLOPs. The framework also transfers to other backbones under the same training protocol, and the comparison of segmentation networks exhibits the expected accuracy–efficiency trade-off: U-Net++ gives the highest accuracy, whereas a lightweight alternative such as SegFormer reduces the computational cost substantially at a moderate accuracy loss.

Several limitations remain. Multi-teacher training increases the parameter count and the training cost, and the validation is restricted to two ISPRS benchmarks with a single sensor type and regional coverage. Future work will investigate lightweight teacher ensembles to reduce the training overhead, cross-sensor and cross-region transfer including evaluation on an additional dataset such as LoveDA, and a quantitative calibration of the estimated evidential uncertainty.

6. Conclusions

This paper presented MTDPNet to address unreliable supervision of difficult pixels in semi-supervised semantic segmentation of RS images. The framework models the training value of unlabeled pixels through complementary teacher-side disagreement and student-side evidence. MADM uses JSD to explicitly characterize pixel-level differences among teacher predictions and performs teacher weighting, soft pseudo-label fusion, and

region-adaptive supervision. EUEM constructs a Dirichlet evidence distribution from multiscale features to distinguish reliable predictions supported by sufficient evidence from spuriously high-confidence predictions with weak evidence.

Compared with existing multi-teacher and uncertainty-aware methods, MTDPNet differs in three respects: it turns teacher disagreement into an explicit pixel-level reliability signal, it estimates uncertainty through evidence sufficiency rather than probability sharpness, and it jointly models teacher-side reliability and student-side evidence, which prior work treats separately.

Across 5% and 25% labeling ratios on ISPRS Vaihingen and Potsdam, MTDPNet achieves the highest mIoU and mF1 and yields consistent gains for small classes such as vehicles. Ablation results identify MADM as the primary source of improvement, while EUEM provides complementary gains on difficult classes and overall performance as a weak evidential regularizer. Joint visualization further shows that difficult regions combine high evidential uncertainty with low teacher confidence around boundaries, small objects, and class-confusion regions, supporting the complementary roles of MADM and EUEM.

Author Contributions: W.Z. and X.D. conceived the study and designed the method. W.Z. implemented the framework and performed the experiments. B.L. and D.G. contributed to data preparation and analysis. X.Z. and Z.C. contributed to visualization and validation. S.K. contributed to the investigation. X.D. supervised the work and acquired the funding. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: No new data were created. The ISPRS Vaihingen and Potsdam datasets are publicly available from the ISPRS 2D Semantic Labeling benchmark.

Acknowledgments: During the preparation of this manuscript, the authors used a generative AI tool for language polishing. The authors reviewed and edited the content and take full responsibility for the publication.

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

References

1. Zhu, X.X.; Tuia, D.; Mou, L.; Xia, G.-S.; Zhang, L.; Xu, F.; Fraundorfer, F. Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources. *IEEE Geosci. Remote Sens. Mag.* **2017**, *5*, 8–36. [CrossRef]
2. Tarvainen, A.; Valpola, H. Mean Teachers Are Better Role Models: Weight-Averaged Consistency Targets Improve Semi-Supervised Deep Learning Results. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 1195–1204.
3. Sohn, K.; Berthelot, D.; Li, C.-L.; Zhang, Z.; Carlini, N.; Cubuk, E.D.; Kurakin, A.; Zhang, H.; Raffel, C. FixMatch: Simplifying Semi-Supervised Learning with Consistency and Confidence. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 596–608.
4. Yang, L.; Zhuo, W.; Qi, L.; Shi, Y.; Gao, Y. ST++: Make Self-Training Work Better for Semi-Supervised Semantic Segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*; IEEE: New York, NY, USA, 2022; pp. 4258–4267.
5. 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*; IEEE: New York, NY, USA, 2022; pp. 4248–4257.
6. Kwon, D.; Kwak, S. Semi-Supervised Semantic Segmentation with Error Localization Network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*; IEEE: New York, NY, USA, 2022; pp. 9957–9967.
7. 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*; IEEE: New York, NY, USA, 2020; pp. 12674–12684.
8. 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*; IEEE: New York, NY, USA, 2021; pp. 2613–2622.
9. Wang, Y.; Wang, H.; Shen, Y.; Fei, J.; Li, W.; Jin, G.; Wu, L.; Zhao, R.; Le, X. Semi-Supervised Semantic Segmentation Using Unreliable Pseudo-Labels. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*; IEEE: New York, NY, USA, 2022; pp. 4238–4247.

10. Lu, X.; Jiao, L.; Liu, F.; Yang, S.; Liu, X.; Feng, Z.; Li, L.; Chen, P. Simple and Efficient: A Semisupervised Learning Framework for Remote Sensing Image Semantic Segmentation. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5543516. [CrossRef]
11. Li, J.; Sun, B.; Li, S.; Kang, X. Semisupervised Semantic Segmentation of Remote Sensing Images with Consistency Self-Training. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5615811. [CrossRef]
12. Yang, Z.; Yan, Z.; Diao, W.; Zhang, Q.; Kang, Y.; Li, J.; Li, X.; Sun, X. Label Propagation and Contrastive Regularization for Semisupervised Semantic Segmentation of Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5609818. [CrossRef]
13. Jin, J.; Lu, W.; Yu, H.; Rong, X.; Sun, X.; Wu, Y. Dynamic and Adaptive Self-Training for Semi-Supervised Remote Sensing Image Semantic Segmentation. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5639814. [CrossRef]
14. Dong, R.; Mou, L.; Chen, M.; Li, W.; Tong, X.-Y.; Yuan, S.; Zhang, L.; Zheng, J.; Zhu, X.X.; Fu, H. Large-Scale Land Cover Mapping with Fine-Grained Classes via Class-Aware Semi-Supervised Semantic Segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*; IEEE: New York, NY, USA, 2023; pp. 16737–16747. [CrossRef]
15. Guo, Y.; Wang, F.; Xiang, Y.; You, H. Semisupervised Semantic Segmentation with Certainty-Aware Consistency Training for Remote Sensing Imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *16*, 2900–2914. [CrossRef]
16. Wang, S.; Sun, X.; Chen, C.; Hong, D.; Han, J. Semi-Supervised Semantic Segmentation for Remote Sensing Images via Multiscale Uncertainty Consistency and Cross-Teacher-Student Attention. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 5517115. [CrossRef]
17. Xin, Y.; Fan, Z.; Qi, X.; Zhang, Y.; Li, X. Confidence-Weighted Dual-Teacher Networks with Biased Contrastive Learning for Semi-Supervised Semantic Segmentation in Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 113447. [CrossRef]
18. Lin, J. Divergence Measures Based on the Shannon Entropy. *IEEE Trans. Inf. Theory* **1991**, *37*, 145–151. [CrossRef]
19. Guo, C.; Pleiss, G.; Sun, Y.; Weinberger, K.Q. On Calibration of Modern Neural Networks. In Proceedings of the 34th International Conference on Machine Learning, PMLR, Sydney, Australia, 6–11 August 2017; Volume 70, pp. 1321–1330.
20. Gal, Y.; Ghahramani, Z. Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning. In Proceedings of the 33rd International Conference on Machine Learning, PMLR, New York, NY, USA, 20–22 June 2016; Volume 48, pp. 1050–1059.
21. Teye, M.; Azizpour, H.; Smith, K. Bayesian Uncertainty Estimation for Batch Normalized Deep Networks. In Proceedings of the 35th International Conference on Machine Learning, PMLR, Stockholm, Sweden, 10–15 July 2018; Volume 80, pp. 4907–4916.
22. Sensoy, M.; Kaplan, L.; Kandemir, M. Evidential Deep Learning to Quantify Classification Uncertainty. *Adv. Neural Inf. Process. Syst.* **2018**, *31*, 3183–3193.
23. Yuan, J.; Liu, Y.; Shen, C.; Wang, Z.; Li, H. A Simple Baseline for Semi-Supervised Semantic Segmentation with Strong Data Augmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*; IEEE: New York, NY, USA, 2021; pp. 8209–8218.
24. Yang, L.; Qi, L.; Feng, L.; Zhang, W.; Shi, Y. Revisiting Weak-to-Strong Consistency in Semi-Supervised Semantic Segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*; IEEE: New York, NY, USA, 2023; pp. 7236–7246.
25. Kong, H.; Lee, G.-H.; Kim, S.; Lee, S.-W. Pruning-Guided Curriculum Learning for Semi-Supervised Semantic Segmentation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*; IEEE: New York, NY, USA, 2023; pp. 5903–5912.
26. Fang, Y.; Zhu, F.; Cheng, B.; Liu, L.; Zhao, Y.; Wei, Y. Locating Noise is Halfway Denoising for Semi-Supervised Segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*; IEEE: New York, NY, USA, 2023; pp. 16566–16576.
27. Ma, J.; Wang, C.; Liu, Y.; Lin, L.; Li, G. Enhanced Soft Label for Semi-Supervised Semantic Segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*; IEEE: New York, NY, USA, 2023; pp. 1185–1195.
28. Wang, C.; Xie, H.; Yuan, Y.; Fu, C.; Yue, X. Space Engage: Collaborative Space Supervision for Contrastive-Based Semi-Supervised Semantic Segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*; IEEE: New York, NY, USA, 2023; pp. 931–942.
29. Sun, B.; Yang, Y.; Zhang, L.; Cheng, M.-M.; Hou, Q. CorrMatch: Label Propagation via Correlation Matching for Semi-Supervised Semantic Segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*; IEEE: New York, NY, USA, 2024; pp. 3097–3107.
30. Hu, X.; Jiang, L.; Schiele, B. Training Vision Transformers for Semi-Supervised Semantic Segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*; IEEE: New York, NY, USA, 2024; pp. 4007–4017. [CrossRef]
31. Mai, H.; Sun, R.; Zhang, T.; Wu, F. RankMatch: Exploring the Better Consistency Regularization for Semi-Supervised Semantic Segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*; IEEE: New York, NY, USA, 2024; pp. 3391–3401.
32. Huang, W.; Shi, Y.; Xiong, Z.; Zhu, X.X. Decouple and Weight Semi-Supervised Semantic Segmentation of Remote Sensing Images. *ISPRS J. Photogramm. Remote Sens.* **2024**, *212*, 13–26. [CrossRef]
33. Gan, X.; Li, W.; Zhang, Y.; Long, W.; Lu, Y.; Chen, Z. Prior Information-Guided Semi-Supervised Semantic Segmentation of Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 4411016. [CrossRef]
34. Hinton, G.; Vinyals, O.; Dean, J. Distilling the Knowledge in a Neural Network. *arXiv* **2015**, arXiv:1503.02531.

35. Cho, H.; Choi, I.; Kwak, S.; Kim, W.H. Interactive Network Perturbation Between Teacher and Students for Semi-Supervised Semantic Segmentation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*; IEEE: New York, NY, USA, 2024; pp. 615–624.
36. Han, Z.; Bai, L.; Pan, B.; Ren, P. From Mutual Guide to Confucius Tri-Learning: A Theoretical Justification. *Pattern Recognit.* **2026**, *178*, 113447. [CrossRef]
37. Mai, H.; Sun, R.; Chen, Y.; Li, W.; Wang, B.; Li, A.; He, Z.; Wang, Y. From Softmax to Dirichlet: Evidential Learning for Semi-supervised Semantic Segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*; IEEE: New York, NY, USA, 2026; pp. 27697–27707.
38. Fan, J.; Li, J.; Zhong, X.; Ren, K.; Jiang, Z.; Gan, M.; Gu, T.; Huang, L. Evidential Deep Partial Label Learning to Quantify Disambiguation Uncertainty. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*; IEEE: New York, NY, USA, 2026; pp. 24770–24779.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.