Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Dual-Guidance Framework for Semi-Supervised Semantic Segmentation via Uncertainty and Prototypes

Xi Lu

luxi@ntit.edu.cn

Nantong Institute of Technology

Byung-Won Min

Mokwon University

Research Article

Keywords: Semi-Supervised Semantic Segmentation, Uncertainty Estimation, Contrastive Learning, Prototype Learning, Pseudo-Label Refinement

Posted Date: July 14th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10321891/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Dual-Guidance Framework for Semi-Supervised Semantic Segmentation via Uncertainty and Prototypes

Xi Lu (ORCID: 0009-0003-4406-564X)[1*] and Byung-Won Min[2]

[1*]School of Information Engineering, Nantong Institute of Technology, Nantong, 226000, China.

[2]Division of Information and Communication Convergence Engineering, Mokwon University, Daejeon, 366100, Korea.

*Corresponding author(s). E-mail(s): luxi@ntit.edu.cn;
Contributing authors: minfam@mokwon.ac.kr;

Abstract

Semi-supervised semantic segmentation (SSSS) reduces the dependence on dense pixel-level annotations, but current weak-to-strong consistency methods remain vulnerable to noisy pseudo-labels and poorly separated feature representations. To address these limitations, we propose a Dual-Guidance framework that jointly refines the prediction space and regularizes the representation space. The Uncertainty-Aware Dynamic Pseudo-Label Refinement (UADPR) module estimates pixel-level uncertainty with Monte Carlo Dropout and uses it to adapt confidence thresholds and assign reliability-aware supervision weights. In parallel, the Class-Prototype Contrastive Guidance (CPCG) module maintains dynamic class prototypes and applies variance-aware contrastive learning with hard negative mining to improve intra-class compactness and inter-class separability. Prototype-guided similarity is further fused with network logits to generate more structured pseudo-labels. Experiments on PASCAL VOC 2012 and Cityscapes demonstrate the effectiveness of the proposed framework. Under the UniMatch protocol with DeepLabv3+ and a ResNet-101 backbone, our method achieves 78.16% and 82.94% mIoU on the VOC 92 and 732 labeled splits, showing robust performance in data-scarce settings.

Keywords: Semi-Supervised Semantic Segmentation, Uncertainty Estimation, Contrastive Learning, Prototype Learning, Pseudo-Label Refinement

1 Introduction

Semantic segmentation plays a pivotal role in computer vision, yet its reliance on dense, pixel-level annotations significantly limits its scalability in real-world applications. To alleviate this annotation burden, Semi-Supervised Semantic Segmentation (SSSS) has emerged as a practical paradigm. As extensively reviewed by Ran et al. [1], contemporary SSSS approaches predominantly adopt a weak-to-strong consistency framework [2–4]. These methods leverage a supervised model to generate pseudo-labels from weakly augmented unlabeled images, which then guide the learning process on strongly augmented views. While frameworks like UniMatch [4] and its recent advancements [5] have established strong baselines, they generally rely on static, high-confidence thresholds to filter out noisy predictions.

Despite their success, existing pseudo-labeling methods can suffer from confirmation bias and feature confusion, especially when dealing with complex scenes or extreme label scarcity. Relying purely on softmax confidence often misleads the model, as modern neural networks tend to be overconfident even in erroneous predictions [6]. This hard-thresholding approach directly discards low-confidence but highly informative pixels around object boundaries, leading to structural inconsistency. Furthermore, these thresholding mechanisms are fundamentally "inward-looking," restricted to the output logit space. They fail to explicitly regularize the underlying latent feature space. When facing morphologically similar classes, the lack of feature-level constraints causes the decision boundaries to blur, resulting in overlapping and poorly clustered semantic representations.

To address these limitations, we propose a unified Dual-Guidance framework that mutually aligns the prediction space and the representation space. To tackle the issue of noisy pseudo-labels, we introduce the Uncertainty-Aware Dynamic Pseudo-Label Refinement (UADPR) module. Rather than relying on rigid confidence scores, UADPR employs Monte Carlo Dropout to quantify the predictive uncertainty at the pixel level. This uncertainty measure dynamically dictates an adaptive confidence threshold and assigns continuous loss weights, ensuring the model focuses on reliable regions while cautiously evaluating ambiguous boundaries. To complement this, we propose the Class-Prototype Contrastive Guidance (CPCG) module to tackle feature confusion. CPCG maintains a global bank of dynamic class prototypes updated via Exponential Moving Average (EMA). By introducing variance-aware adaptive temperature scaling and hard negative mining, it explicitly pulls pixel representations toward their correct semantic centers while pushing them away from confusing classes. Ultimately, the prototype-guided similarity from CPCG is fused with the logits to generate robust pseudo-labels, forming a practical feedback loop between feature regularization and label refinement. Our contributions can be summarized as follows:

- We propose the UADPR module, which estimates pixel-level uncertainty via Monte Carlo Dropout to establish an adaptive confidence thresholding mechanism, effectively mitigating confirmation bias in the prediction space.
- We introduce the CPCG module, which utilizes EMA-updated dynamic prototypes and hard negative mining to perform variance-aware contrastive learning, ensuring intra-class compactness and inter-class separability in the feature space.
- Extensive experiments on standard benchmarks demonstrate that our integrated dual-guidance framework achieves competitive performance under the UniMatch/ResNet-101 setting, exhibiting particular robustness in data-scarce scenarios.

2 Related Work

2.1 Semi-Supervised Semantic Segmentation

The core objective of SSSS is to extract reliable supervisory signals from abundant unlabeled data. While early attempts explored generative adversarial networks [7], the current landscape is dominated by consistency training and pseudo-labeling. Foundational frameworks like CPS [8], AEL [9], and U^{2}PL [10] enforce prediction consistency across different network initializations or input perturbations while addressing class imbalance or unreliable pixels. Other representative works, such as PS-MT [11], employ perturbed mean teachers to enhance robustness. Recent competitive methods have increasingly focused on the quality of pseudo-labels. For instance, CorrMatch [12] explores inter-pixel correlations to propagate labels, while CAFS [13] and CSL [6] address the cognitive bias of overconfident networks by formulating class-adaptive thresholding and confidence separable learning. CW-BASS [14] further introduces confidence-weighted boundary-aware learning, highlighting the growing interest in boundary reliability for semi-supervised semantic segmentation. Despite these structural improvements in thresholding, most of these approaches still derive their thresholds directly from softmax probabilities. This reliance makes them vulnerable to the inherent calibration errors of deep networks, highlighting the need for a mechanism that explicitly models prediction uncertainty rather than merely relying on point-estimate confidences.

2.2 Uncertainty Estimation in Vision

Uncertainty estimation provides a mathematical foundation for models to gauge the reliability of their own predictions. Techniques like Monte Carlo (MC) Dropout [15] have been widely adopted to capture epistemic uncertainty. In the context of dense prediction tasks, uncertainty is particularly valuable for identifying ambiguous boundary regions. Historically utilized in active learning [16], uncertainty estimation has recently been integrated into SSSS. Works in medical and remote sensing segmentation, such as CACT [17], DUMM [18], and DUEB [19], have successfully utilized variance maps to re-weight loss functions or filter unreliable predictions.

Similarly, AmbiSSL [20] leverages diverse decoders to address annotation ambiguity. Building upon these insights, our work integrates MC Dropout into the mainstream weak-to-strong SSSS pipeline. However, unlike prior arts that use uncertainty purely as an isolated spatial mask, our UADPR module leverages batch-level uncertainty statistics to dynamically govern the global confidence threshold across training epochs, establishing a more holistic and adaptive noise-filtering strategy.

2.3 Contrastive Learning and Prototypes

To overcome the limitations of logit-space supervision, recent research has pivoted toward regularizing the latent feature space. Contrastive learning has profoundly impacted self-supervised representation learning [21], [22]. Its adaptation to dense prediction tasks, such as PixelContrast [23] and ReCo [24], enforces pixel-to-pixel consistency. However, maintaining large pixel queues is computationally exorbitant and often leads to false-negative sampling. To alleviate this, pixel-to-prototype contrastive methods, notably PCR [25], have been introduced. By aggregating features into semantic centers, prototypes provide a global context for pixel classification. Recent literature has seen a surge in this direction, including Prototype-Augmented Mean Teacher (PAMT)

[26], Dual Prototypes Contrastive learning [27], and CPDM [28]. While these methods demonstrate the power of feature-level alignment, they predominantly utilize fixed temperature parameters in their contrastive losses, ignoring the fact that intra-class variance differs drastically across semantic categories. Our proposed CPCG module addresses this specific gap by introducing a variance-aware adaptive temperature. Different from PCR-style prototype consistency and recent prototype-augmented teacher frameworks, CPCG uses uncertainty-filtered samples to stabilize prototype updates, mines hard negative prototypes to sharpen confusing class boundaries, and directly feeds prototype similarity back into pseudo-label generation. This design makes the prototype branch not only an auxiliary representation regularizer but also an explicit source of label refinement for the weak-to-strong training loop.

3 Methods

3.1 Overview

The proposed Dual-Guidance framework is built upon the weak-to-strong semi-supervised segmentation paradigm. Instead of relying only on fixed-threshold pseudo-label selection, the framework introduces two complementary guidance signals. UADPR evaluates the reliability of pseudo-labels from the prediction side, while CPCG regularizes pixel features with class prototypes and feeds prototype-aware label guidance back into training.

Given labeled and unlabeled images, the shared segmentation network produces predictions for weakly and strongly augmented views. For the weak view, UADPR estimates pixel-level uncertainty using Monte Carlo dropout and uses it to adjust the confidence threshold and reliability weight of each pseudo-label. In parallel, CPCG maintains class-wise prototypes in the feature space and measures the semantic affinity between pixel features and prototype centers. The MC-averaged prediction and prototype similarity are fused to generate a hard guided pseudo-label for subsequent supervision.

The two modules are designed to address different parts of the learning process. UADPR controls which pseudo-labels are trusted and how strongly they supervise the model, whereas CPCG improves the semantic structure of representations and provides prototype-guided label assignment.

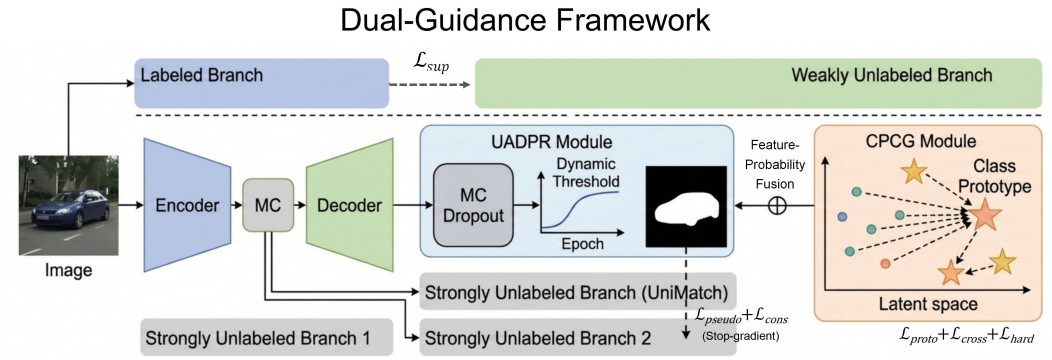


Fig. 1 Overall architecture of the proposed Dual-Guidance Framework. Built upon the weak-to-strong consistency paradigm, our framework processes unlabeled images through a shared encoder-decoder network. For the weak stream, we introduce two synergistic modules: UADPR and CPCG, which refine the prediction space and regularize the feature space, respectively. Through feature-probability fusion, they generate refined pseudo-labels ($\hat{y}_i$) that provide uncertainty-aware supervision for the strongly perturbed streams, effectively mitigating confirmation bias.

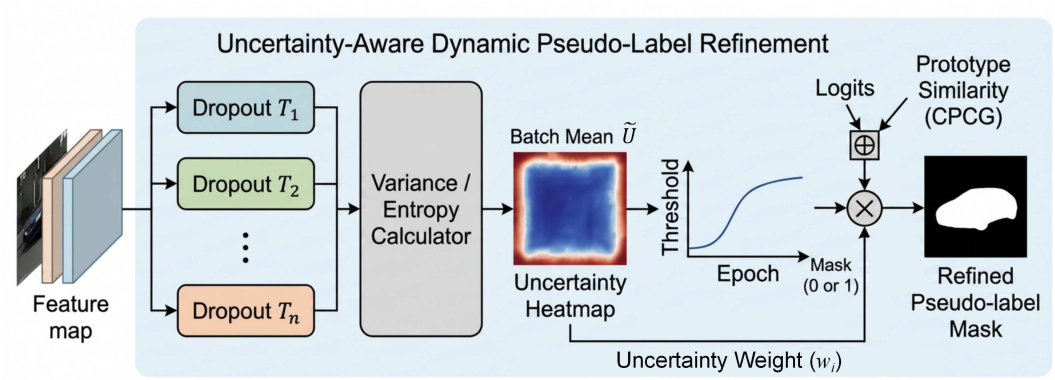


Fig. 2 Detailed illustration of UADPR. The module estimates pixel-level uncertainty from MC Dropout predictions, uses the batch-level normalized uncertainty to adjust the confidence threshold, and assigns pixel-wise reliability weights for pseudo-label supervision.

Their objectives are computed and optimized separately, allowing each module to preserve its own role while still exchanging information through the guided pseudo-label.

3.2 Uncertainty-Aware Dynamic Pseudo-Label Refinement (UADPR)

Static confidence thresholds are often too rigid for semi-supervised segmentation. At early training stages, a strict threshold may discard many informative pixels; at later stages, some ambiguous regions may still produce overconfident predictions. UADPR is designed to make pseudo-label filtering depend on both the training stage and the local uncertainty of each pixel.

For each weakly augmented unlabeled image, we perform T stochastic forward passes with Monte Carlo dropout. The averaged prediction is used as a more stable estimate of the model response:

$$\bar{p}_c(x_i) = \frac{1}{T}\sum_{t=1}^{T} p_c^{(t)}(x_i). \tag{1}$$

The uncertainty of pixel x_i is then measured by the entropy of the averaged distribution:

$$U(x_i) = -\sum_{c=1}^{C} \bar{p}_c(x_i) \log \bar{p}_c(x_i). \tag{2}$$

Because the entropy range depends on the number of classes, we use the normalized uncertainty in later weighting and consistency terms:

$$\tilde{U}(x_i) = \frac{U(x_i)}{\log C}, \quad \bar{\tilde{U}} = \frac{1}{N} \sum_{i=1}^{N} \tilde{U}(x_i). \tag{3}$$

The batch-level uncertainty is used to adjust the pseudo-label threshold over training. This makes the selection criterion less conservative when the model is still uncertain and stricter when the model becomes more stable:

$$\tau^{(k)} = \tau_0 + \alpha \cdot \sigma\left(\frac{k}{K}\right)\left(1 - \bar{\tilde{U}}\right). \tag{4}$$

where $\sigma(\cdot)$ is the sigmoid function and K is the total number of training epochs. In addition to this hard selection rule, UADPR assigns a reliability weight to each valid pixel:

$$w_i = \exp\left(-\beta \tilde{U}(x_i)\right) \cdot \left(\max_c \bar{p}_c(x_i)\right)^{\gamma}. \tag{5}$$

The pseudo-label loss therefore combines a binary confidence mask with an uncertainty-aware continuous weight. The hard guided pseudo-label $\hat{y}_i$ is provided by CPCG:

$$\mathcal{L}_{\text{pseudo}} = -\frac{1}{N} \sum_{i=1}^{N} w_i \cdot \mathbf{1}\left[\max_c \bar{p}_c(x_i) > \tau^{(k)}\right] \times \log p_{\hat{y}_i}(x_i). \tag{6}$$

Here, $\mathbf{1}[\cdot]$ is an indicator function. The mask and weight are computed from the MC-averaged prediction, while the logarithmic term supervises the current network prediction at the guided hard-label class. This design avoids treating all selected pseudo-labels equally and reduces the influence of uncertain regions without discarding useful pixels solely because they are difficult.

UADPR also applies uncertainty-aware consistency regularization between weak and strong perturbations. Pixels with lower uncertainty contribute more strongly to the consistency constraint:

$$\mathcal{L}_{\text{cons}} = \frac{1}{N} \sum_{i=1}^{N} \left(1 - \tilde{U}(x_i)\right) \cdot \left\| p(A_s(x_i)) - p(A_w(x_i)) \right\|_2^2. \tag{7}$$

An entropy minimization term further encourages confident decision boundaries:

$$\mathcal{L}_{\text{ent}} = -\frac{1}{N} \sum_{i=1}^{N} \sum_{c=1}^{C} p_c(x_i) \log p_c(x_i). \tag{8}$$

The UADPR objective is optimized for prediction-level pseudo-label refinement:

$$\mathcal{L}_{\text{UADPR}} = \mathcal{L}_{\text{sup}} + \lambda_p \mathcal{L}_{\text{pseudo}} + \lambda_c \mathcal{L}_{\text{cons}} + \lambda_e \mathcal{L}_{\text{ent}}. \tag{9}$$

3.3 Class-Prototype Contrastive Guidance (CPCG)

Although UADPR improves the reliability of pseudo-label supervision, it does not explicitly organize the feature distribution. In weakly supervised regions, features from the same class may remain scattered, and visually similar categories may be difficult to separate. CPCG addresses this problem by maintaining semantic prototypes and using them as anchors for pixel-level representation learning.

For each class, a global prototype is updated from labeled pixels and confident unlabeled pixels using an exponential moving average:

$$\mu_c^{(k)} = \eta \mu_c^{(k-1)} + (1 - \eta) \cdot \frac{1}{|S_c|} \sum_{i \in S_c} f_i. \tag{10}$$

where S_c is the set of pixels assigned to class c, f_i is the pixel feature, and η is the momentum coefficient. Stop-gradient is applied during prototype updating to keep the prototype bank stable.

A single fixed contrastive temperature may be unsuitable for all classes because different semantic categories exhibit different intra-class variance. CPCG therefore uses a class-adaptive temperature:

$$T_c = T_0 \cdot \left(1 + \delta \cdot \mathrm{Var}(\{f_i : i \in S_c\})\right). \tag{11}$$

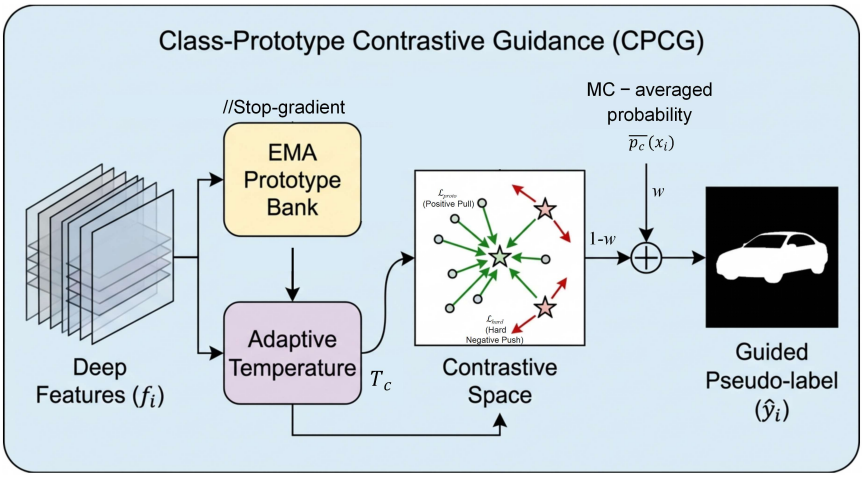


Fig. 3 Detailed illustration of CPCG. The module maintains EMA class prototypes, applies variance-aware temperature scaling and hard negative mining, and generates hard guided pseudo-labels by fusing MC-averaged predictions with prototype similarity.

With the adaptive temperature, the similarity between pixel features and class prototypes is computed as:

$$s_{i,c} = \frac{\exp(f_i^\top \mu_c / T_c)}{\sum_{j=1}^{C} \exp(f_i^\top \mu_j / T_j)}. \tag{12}$$

The prototype similarity is then fused with the MC-averaged prediction to generate a hard guided pseudo-label:

$$\hat{y}_i = \arg\max_c \left[\omega \cdot \bar{p}_c(x_i) + (1-\omega) \cdot s_{i,c}\right]. \tag{13}$$

This label assignment uses both prediction confidence and feature-level semantic affinity. It is particularly useful when a pixel has a confident prediction but its representation is closer to another semantic prototype.

CPCG first applies a prototype contrastive loss to pull each pixel feature toward its assigned prototype and repel it from other class prototypes:

$$\mathcal{L}_{\text{proto}} = -\frac{1}{N}\sum_{i=1}^{N} \log \frac{\exp(f_i^\top \mu_{\hat{y}_i}/T_{\hat{y}_i})}{D_i},$$
$$D_i = \exp(f_i^\top \mu_{\hat{y}_i}/T_{\hat{y}_i}) + \sum_{c \neq \hat{y}_i} \exp(f_i^\top \mu_c / T_c). \tag{14}$$

To prevent batch-level prototypes from drifting away from the global prototype bank, a cross-sample prototype consistency term is added:

$$\mathcal{L}_{\text{cross}} = \sum_{c=1}^{C} \left\| \mu_c^{\text{batch}} - \mu_c^{\text{global}} \right\|_2^2. \tag{15}$$

For categories with similar appearance or context, CPCG further mines the most confusing negative prototypes:

$$\mathcal{H}_i = \text{TopK}_{c \neq \hat{y}_i} \left(f_i^\top \mu_c \right). \tag{16}$$

The hard-negative loss focuses the contrastive objective on these confusing classes:

$$\mathcal{L}_{\text{hard}} = -\frac{1}{N}\sum_{i=1}^{N} \log \frac{\exp(f_i^\top \mu_{\hat{y}_i}/T_{\hat{y}_i})}{D_i^{\text{hard}}},$$
$$D_i^{\text{hard}} = \exp(f_i^\top \mu_{\hat{y}_i}/T_{\hat{y}_i}) + \sum_{c \in \mathcal{H}_i} \exp(f_i^\top \mu_c / T_c). \tag{17}$$

The CPCG objective is optimized independently from UADPR and focuses on representation-level regularization:

$$\mathcal{L}_{\text{CPCG}} = \mathcal{L}_{\text{seg}} + \lambda_{\text{proto}}\mathcal{L}_{\text{proto}} + \lambda_{\text{cross}}\mathcal{L}_{\text{cross}} + \lambda_{\text{hard}}\mathcal{L}_{\text{hard}}. \tag{18}$$

Through this design, UADPR and CPCG keep distinct optimization roles while sharing information through the guided pseudo-label. UADPR determines the reliability of pseudo-label supervision, and CPCG improves the semantic organization of pixel features that supports the label assignment.

4 Experiments and Results

4.1 Implementation Details

We evaluated the proposed Dual-Guidance framework on two standard semantic segmentation benchmarks: PASCAL VOC 2012 [29] and Cityscapes [30]. For Pascal VOC 2012, which consists of 10,582 training images across 21 classes (including the background), we followed the standard partition protocols, sampling 92, 183, 366, and 732 labeled images. Cityscapes contains 2,975 training images of urban street scenes covering 19 classes, from which we sampled 1/16, 1/8, 1/4, and 1/2 labeled proportions.

To ensure a fair comparison, we adopted UniMatch as our strong baseline and strictly followed its experimental settings. All evaluations were conducted using DeepLabv3+ [31] with an ImageNet-pretrained ResNet-101 [32] as the unified backbone network. During training, we employed the SGD optimizer with a momentum of 0.9 and a weight decay of 1×10-4. The learning rate was decayed following a poly schedule. For Pascal VOC 2012, the initial learning rate was set to 0.001, the crop size was 321×321, and the models were trained for 80 epochs. For Cityscapes, the initial learning rate was 0.005, the crop size was 769×769, and the training lasted for 240 epochs. In both datasets, each mini-batch consisted of 8 labeled images and 8 unlabeled images.

For data augmentation, the weak perturbation applied to unlabeled images included random resizing (scaling ratio from 0.5 to 2.0), random cropping, and horizontal flipping. The strong perturbation further incorporated color jittering, grayscale conversion, Gaussian blur, and CutMix [33]. Regarding our proposed modules, we set the Monte Carlo (MC) sampling times to $T = 8$ for uncertainty estimation in the UADPR module, and kept the prototype update rule, fusion strategy, and loss weights fixed across all labeled splits. During inference, we used the original image resolution for Pascal VOC, while sliding window evaluation (window size 769×769) was adopted for Cityscapes. Performance was consistently evaluated using the mean Intersection over Union (mIoU) metric.

4.2 Comparison with Representative Methods

To evaluate the effectiveness of our approach, we compared the full Dual-Guidance framework against an extensive set of well-established and recent competitive semi-supervised semantic segmentation methods. These include classical consistency regularization and pseudo-labeling milestones (CPS [8], AEL [9], ST++ [3], U^2PL [10], PS-MT [11], PCR [25]), recent advanced frameworks (AugSeg [34], CorrMatch [12], CSL [6], DDFP [35], AllSpark [36], CoVar [37]), and our robust baseline, UniMatch [4]. Our own baseline and proposed variants strictly follow the UniMatch protocol with DeepLabv3+ and ResNet-101, while externally reported results are included as representative references under comparable semi-supervised segmentation settings.

Table 1 summarizes the quantitative outcomes on the PASCAL VOC 2012 benchmark, where our method consistently yields substantial improvements across all partition splits. The performance gap is particularly striking in the extreme data-scarce regime of 92 labeled images. In this setting, our framework achieves an mIoU of 78.16%, demonstrating a considerable advancement over recent strong competitors like CSL [6] (76.80%), AllSpark [36] (76.07%), and CorrMatch [12] (76.40%), and further surpassing the highly competitive UniMatch [4] baseline by +2.98%. This upward trajectory is firmly maintained even as more labeled data becomes available; at the 732-split, the framework reaches a peak mIoU of 82.94%, exceeding UniMatch by +3.02%. This persistent performance margin suggests that the dual-guidance mechanism successfully extracts richer supervisory signals from unlabeled data, rather than merely overfitting to the limited annotations.

The strong generalization ability of this approach is further indicated by the results obtained from the complex urban environments of the Cityscapes dataset, detailed in Table 2. Under the challenging 1/16 split, the model attains

Table 1 Quantitative comparison (mIoU %) with representative methods on the Pascal VOC 2012 dataset under different labeled data splits. Our baseline and proposed method adopt DeepLabv3+ with ResNet-101 following the UniMatch protocol. The best results are highlighted in bold.

Pascal	**1/16 (92)**	**1/8 (183)**	**1/4 (366)**	**1/2 (732)**
Supervised Baseline	45.40	54.70	65.40	71.50
CPS [8]$_{\text{CVPR 2021}}$	64.07	67.42	71.71	75.88
ST++ [3]$_{\text{CVPR 2022}}$	65.20	71.00	74.60	77.30
U^{2}PL [10]$_{\text{CVPR 2022}}$	67.98	69.15	73.66	76.16
PS-MT [11]$_{\text{CVPR 2022}}$	65.80	69.58	76.57	78.42
PCR [25]$_{\text{NeurIPS 2022}}$	70.06	74.71	77.16	78.49
AugSeg [34]$_{\text{CVPR 2023}}$	71.09	75.45	78.80	80.33
DDFP [35]$_{\text{CVPR 2024}}$	74.95	78.01	79.51	81.21
AllSpark [36]$_{\text{CVPR 2024}}$	76.07	78.41	79.77	80.75
CorrMatch [12]$_{\text{CVPR 2024}}$	76.40	78.50	79.40	80.60
CSL [6]$_{\text{arXiv 2025}}$	76.80	79.60	80.30	80.90
UniMatch (Baseline) [4]$_{\text{CVPR 2023}}$	75.18	77.22	78.82	79.92
Ours	**78.16**	**80.32**	**81.42**	**82.94**

Note: Values such as 92,183 in the header of Pascal VOC 2012 represent the number of labeled images. The fractions in the header indicate the proportion of labeled images out of the total training set.

Table 2 Quantitative comparison (mIoU %) with representative methods on the Cityscapes dataset. The fractions in the header indicate the proportion of available labeled images. Our baseline and proposed method adopt DeepLabv3+ with ResNet-101 following the UniMatch protocol.

Cityscapes	**1/16 (186)**	**1/8 (372)**	**1/4 (744)**	**1/2 (1488)**
Supervised Baseline	66.30	72.80	75.00	78.00
CPS [8]$_{\text{CVPR 2021}}$	69.78	74.31	74.58	76.81
PCR [25]$_{\text{NeurIPS 2022}}$	73.41	76.31	78.40	79.11
AEL [9]$_{\text{NeurIPS 2021}}$	74.45	75.55	77.48	79.01
U^{2}PL [10]$_{\text{CVPR 2022}}$	70.30	74.37	76.47	79.05
PS-MT [11]$_{\text{CVPR 2022}}$	–	76.89	77.60	79.09
AugSeg [34]$_{\text{CVPR 2023}}$	75.22	77.82	79.56	80.43
DDFP [35]$_{\text{CVPR 2024}}$	77.10	78.19	79.88	80.82
CorrMatch [12]$_{\text{CVPR 2024}}$	77.30	78.50	79.40	80.40
CSL [6]$_{\text{arXiv 2025}}$	78.20	78.80	80.00	81.10
CoVar [37]$_{\text{arXiv 2026}}$	78.30	78.60	80.10	81.30
AllSpark [36]$_{\text{CVPR 2024}}$	78.33	79.24	80.56	81.39
UniMatch (Baseline) [4]$_{\text{CVPR 2023}}$	76.62	77.94	79.16	79.52
Ours	**79.08**	**80.32**	**81.12**	**82.12**

Note: Values such as 186,372 in the header of Cityscapes represent the number of labeled images. The fractions in the header indicate the proportion of labeled images out of the total training set.

79.08% mIoU, outperforming the UniMatch [4] baseline by +2.46% and showing favorable performance compared with recent advanced approaches such as DDFP [35] (77.10%), CSL [6] (78.20%), CoVar [37] (78.30%), and AllSpark [36] (78.33%). As the labeled data ratio increases to 1/2, our framework continues to scale effectively, reaching an mIoU of 82.12%. These results suggest that the proposed dual-guidance design remains competitive across different supervision levels, while its most important comparison is the consistent improvement over the reproduced UniMatch baseline under the same implementation protocol. Such consistent gains across a dataset characterized by high-resolution images and severe scale variations suggest that explicit feature-level regularization, combined with dynamic thresholding, provides a highly effective strategy for resolving complex spatial relationships and ambiguous class boundaries.

4.3 Ablation Studies

While the overall performance is compelling, isolating the contribution of each module reveals the underlying mechanics of the framework. We conducted systematic ablation studies on both

the Pascal VOC and Cityscapes datasets, with the quantitative breakdowns detailed in Tables 3 and 4.

Table 3 Ablation study of mIoU (%) on the Pascal VOC 2012 dataset. Absolute performance gains over the baseline are denoted in parentheses.

Split	Base	+UADPR	+CPCG	Ours
92	75.18	77.62 (+2.44)	77.46 (+2.28)	**78.16 (+2.98)**
183	77.22	79.74 (+2.52)	79.58 (+2.36)	**80.32 (+3.10)**
366	78.82	81.02 (+2.20)	80.86 (+2.04)	**81.42 (+2.60)**
732	79.92	82.48 (+2.56)	82.32 (+2.40)	**82.94 (+3.02)**

Table 4 Ablation study of mIoU (%) on the Cityscapes dataset. Absolute performance gains over the baseline are denoted in parentheses.

Split	Base	+UADPR	+CPCG	Ours
1/16	76.62	78.52 (+1.90)	78.38 (+1.76)	**79.08 (+2.46)**
1/8	77.94	79.76 (+1.82)	79.62 (+1.68)	**80.32 (+2.38)**
1/4	79.16	80.62 (+1.46)	80.48 (+1.32)	**81.12 (+1.96)**
1/2	79.52	81.68 (+2.16)	81.52 (+2.00)	**82.12 (+2.60)**

Integrating only the Uncertainty-Aware Dynamic Pseudo-Label Refinement (UADPR) module into the baseline yielded immediate benefits, lifting the mIoU by +2.44% on the VOC 92-split and +1.90% on the Cityscapes 1/16-split. This improvement likely stems from the dynamic thresholding mechanism. Early in training, the uncertainty estimation actively filters out noisy pseudo-labels, helping to mitigate confirmation bias. As training progresses, the dynamic threshold adapts to the model's growing confidence, safely reclaiming informative pixels that a static threshold would have permanently discarded.

Beyond refining the prediction space, introducing explicit feature-level constraints via the Class-Prototype Contrastive Guidance (CPCG) module appeared equally crucial. Applying CPCG alone pushed the baseline performance up by +2.28% on the VOC 92-split and +1.76% on the Cityscapes 1/16-split. These gains highlight the potential value of feature-level regularization. Without CPCG, the baseline solely relies on cross-entropy in the logit space, frequently leading to scattered feature distributions. CPCG rectifies this by explicitly pulling pixel representations toward their respective class prototypes, thereby establishing more robust and discriminative decision boundaries.

The most striking observation from these experiments, however, emerges when both modules operate simultaneously. The combined framework consistently achieved the highest performance across all settings, delivering peak absolute gains of +3.10% on VOC (183-split) and +2.60% on Cityscapes (1/2-split). This consistent advantage supports our core hypothesis regarding their synergistic effect. Specifically, UADPR provides clean, uncertainty-weighted targets that stabilize prototype updates in CPCG, while the structured feature space established by CPCG yields more accurate similarity maps, which in turn generate higher-quality pseudo-labels for UADPR. This mutual enhancement closes the loop between the representation and prediction spaces.

4.4 Qualitative Analysis and Insights

To better understand the behavior of the proposed framework, we visually analyze the intermediate feature distributions and the evolution of pseudo-labels.

We begin by observing the latent feature space using t-SNE projections in Figure 4. In the baseline distribution (left), features from different classes are noticeably scattered and intertwined.

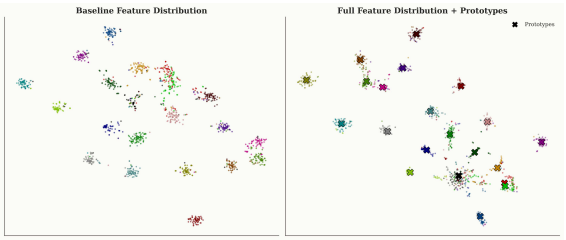


Fig. 4 t-SNE visualization of the latent feature space for the baseline and the full model. The full model shows more compact intra-class clusters and clearer inter-class margins.

In contrast, the features extracted by our full framework (right) exhibit significantly higher intra-class compactness and clearer inter-class margins. The explicitly learned class prototypes (marked as black crosses) act as strong attractors, anchoring the clusters. Furthermore, the introduction of hard negative mining in the CPCG

module visibly pushes apart morphologically similar classes, helping to prevent feature collapse in ambiguous boundary regions.

Furthermore, this more structured feature space appears to contribute to improved pseudo-label quality and final segmentation accuracy. A visual inspection of the internal learning processes (Figures 5, 6, and 7) reveals the direct impact of our modules.

As shown in the Uncertainty Estimation (Figure 5), the model accurately highlights ambiguous regions—predominantly along complex object boundaries such as the intricate spokes of the bicycles and the outline of the airplane—with high uncertainty (represented by bright heatmaps). Notably, the dynamically adjusted thresholds (e.g., $\tau = 0.939, 0.948$ displayed in the top-left corners) adaptively respond to these regions rather than applying a rigid 0.95 cutoff.

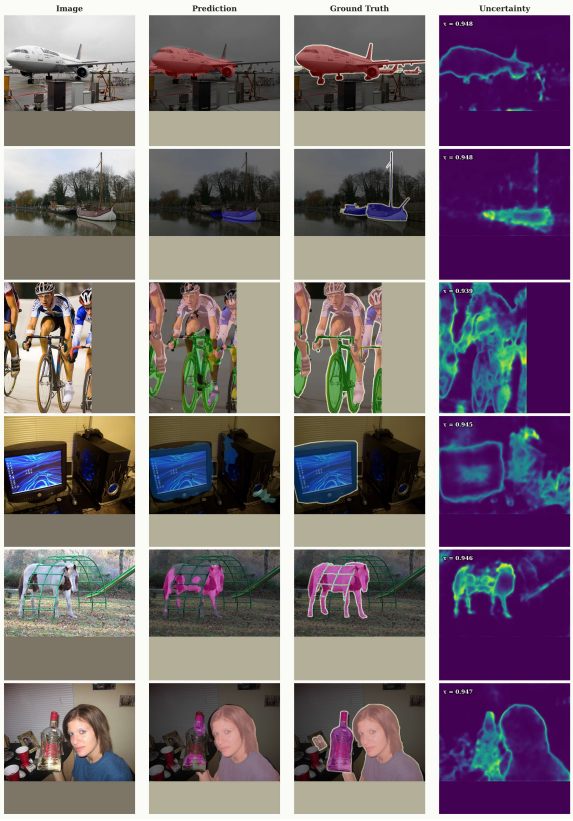


Fig. 5 Uncertainty Estimation. Highlighting ambiguous boundaries where dynamic thresholds are applied.

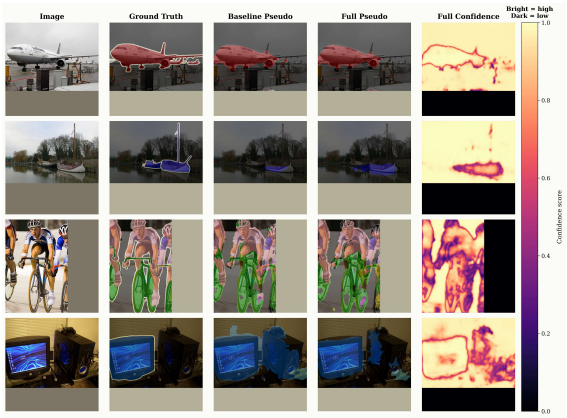


Fig. 6 Pseudo-label comparison between the UniMatch baseline and the full model. From left to right, each row shows the input image, ground truth, baseline pseudo-label, full-model pseudo-label, and full-model confidence map. In the confidence maps, brighter colors indicate higher prediction confidence, while darker colors indicate lower confidence. The full model produces more spatially coherent pseudo-labels and reduces fragmented responses around ambiguous object boundaries.

By identifying and down-weighting uncertain regions during training, UADPR alleviates the accumulation of noisy pseudo-labels. This effect can be observed in Figure 6. The UniMatch baseline often produces fragmented or jagged pseudo-masks in visually ambiguous regions, such as the monitor boundaries and the bicycle riders. In contrast, the full model yields more coherent object regions and cleaner boundary responses. The confidence maps provide further evidence for this behavior: high-confidence responses are mainly distributed over consistent object interiors, whereas low-confidence responses are concentrated around object boundaries and confusing background areas. These observations suggest that the uncertainty-aware strategy provides a more selective use of pseudo-label supervision in ambiguous regions.

Consequently, the model is encouraged to first exploit reliable core regions and then gradually extend supervision to more ambiguous boundary areas as the learned prototypes become more stable. The selected qualitative examples in Figure 7 provide visual evidence for this behavior. Compared with the UniMatch baseline, the full framework produces more spatially coherent predictions and reduces several false responses in challenging regions. These observations are consistent with the quantitative gains reported in Tables 1 and 2.

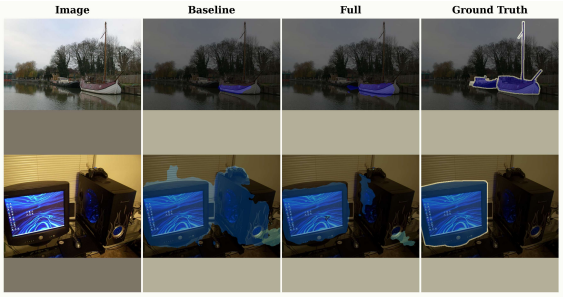


Fig. 7 Selected qualitative segmentation comparison among the input image, UniMatch baseline, full model, and ground truth.

Taken together, the qualitative results provide a consistent explanation for the quantitative improvements. The uncertainty maps show where the model should avoid over-trusting pseudo-labels, especially near thin structures, weak object contours, and visually confusing background regions. The pseudo-label comparisons further show that this selective supervision does not simply remove difficult pixels; instead, it preserves reliable object interiors while reducing noisy responses around uncertain boundaries. At the representation level, the t-SNE projections indicate that CPCG reshapes the feature space into more compact and better separated semantic groups, which helps the subsequent pseudo-label fusion produce cleaner class assignments. These observations support the intended role of the dual-guidance design: UADPR controls the reliability of supervision in the prediction space, while CPCG improves semantic consistency in the feature space. Their interaction yields pseudo-labels that are both confidence-aware and representation-aware, leading to more stable segmentation outputs under limited labeled supervision.

The visual cases also clarify why the two modules should be used together rather than independently. If uncertainty estimation is used alone, unreliable boundary pixels can be suppressed, but the latent features may still remain weakly organized for visually similar categories. If prototype contrast is used alone, the feature space can become more discriminative, but noisy pseudo-labels may still introduce incorrect prototype updates. The proposed combination reduces both risks: uncertain pixels receive lower training influence, and reliable pixels contribute to more stable prototype learning. This behavior is especially important for semi-supervised semantic segmentation, where errors in pseudo-labels can repeatedly reinforce themselves across training iterations. By coupling confidence filtering with representation guidance, the framework produces segmentation masks that are less fragmented and better aligned with object structures.

Another notable pattern is that the improvements are most visible in regions where ordinary confidence-based pseudo-labeling is likely to be unstable. Thin structures, small foreground objects, and partially occluded categories often produce mixed responses in the baseline, because a high softmax score does not necessarily imply a well-separated feature representation. In these cases, UADPR first reduces the training pressure from uncertain pixels, while CPCG encourages the remaining reliable pixels to align with class-level semantic anchors. The resulting masks therefore preserve object interiors without aggressively expanding into nearby background regions. This qualitative behavior is consistent with the ablation results: each module improves the baseline independently, but their combination provides a clearer benefit because reliability estimation and prototype alignment correct different failure modes of pseudo-label training.

5 Discussion

Our findings suggest that the UADPR and CPCG modules are complementary. By filtering noisy targets through predictive uncertainty, the framework helps to stabilize the prototype updates. This stability allows the contrastive learning module to form better-separated class clusters in the latent space. As a result, the network generates more reliable similarity maps, which directly feed back to refine the pseudo-labels. This interaction is consistent with the intended closed loop: uncertainty-aware filtering reduces noisy supervision, cleaner supervision improves prototype estimation, and prototype-guided similarity further refines pseudo-labels for subsequent weak-to-strong training.

This design also has limitations. Relying on Monte Carlo Dropout for uncertainty estimation requires multiple stochastic forward passes during training, increasing memory usage and training time. The prototype bank may be less stable

for rare or temporarily absent classes in a mini-batch, and highly confident but incorrect pseudo-labels can still affect prototype updates before being corrected by later training. These issues suggest that more efficient uncertainty estimation, class-balanced prototype maintenance, and stronger safeguards for rare categories are worthwhile directions for future work.

6 Conclusion

In conclusion, we presented a dual-guidance framework that bridges pseudo-label refinement and feature-level regularization for semi-supervised semantic segmentation. By coupling uncertainty-aware dynamic thresholding with class-prototype contrastive learning, our approach effectively mitigated the impact of noisy supervision. Evaluations across standard benchmarks demonstrated that this joint optimization yielded consistent improvements over the baseline, particularly in data-scarce scenarios. Beyond the observed performance gains, the effectiveness of this method suggests that semi-supervised segmentation can benefit from extending traditional output-space self-training to include explicit representation alignment. We anticipate that this perspective of mutually guiding the prediction and latent spaces holds potential for addressing label noise in broader label-efficient vision tasks.

Statements and Declarations

Funding. The authors received no financial support for the research, authorship, or publication of this article.

Author Contributions. Xi Lu is the corresponding author and primary contributor. Xi Lu contributed to conceptualization, methodology, experimental design, formal analysis, visualization, and writing of the original draft. Byung-Won Min contributed to supervision, validation, project administration, and review and editing of the manuscript. Both authors read and approved the final manuscript.

Data Availability. The datasets analyzed in this study are publicly available from their respective official sources. The PASCAL VOC 2012 dataset is available from the PASCAL Visual Object Classes Challenge website at http://host.robots.ox.ac.uk/pascal/VOC/voc2012/. The Cityscapes dataset is available from the Cityscapes official website at https://www.cityscapes-dataset.com/. The experimental split protocols and evaluation settings used in this study are described in the manuscript.

Competing Interests. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] L. Ran, Y. Li, G. Liang, and Y. Zhang, "Semi-supervised semantic segmentation based on pseudo-labels: A survey," IEEE Trans. Circuits Syst. Video Technol., vol. 35, no. 4, pp. 3054-3080, Apr. 2025. Available: https://arxiv.org/abs/2403.01909

[2] K. Sohn et al., "FixMatch: Simplifying semi-supervised learning with consistency and confidence," in Adv. Neural Inform. Process. Syst. (NeurIPS), 2020, pp. 596-608. Available: NeurIPS proceedings

[3] L. Yang, W. Zhuo, L. Qi, Y. Shi, and Y. Gao, "ST++: Make self-training work better for semi-supervised semantic segmentation," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2022, pp. 4268-4277. doi: https://doi.org/10.1109/CVPR52688.2022.00423

[4] L. Yang, L. Qi, L. Feng, W. Zhang, and Y. Shi, "Revisiting weak-to-strong consistency in semi-supervised semantic segmentation," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2023, pp. 7236-7246. doi: https://doi.org/10.1109/CVPR52729.2023.00699

[5] L. Yang et al., "UniMatch V2: Pushing the limit of semi-supervised semantic segmentation," IEEE Trans. Pattern Anal. Mach. Intell., 2025. doi: https://doi.org/10.1109/TPAMI.2025.3528453

[6] P. Liu and J. Liu, "When confidence fails: Revisiting pseudo-label selection in semi-supervised semantic segmentation," in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), 2025. doi: https://doi.org/10.1109/ICCV51701.2025.02031

[7] W.-C. Hung, Y.-H. Tsai, Y.-T. Liou, Y.-Y. Lin, and M.-H. Yang, "Adversarial learning for semi-supervised semantic segmentation," in Proc. Brit. Mach. Vis. Conf. (BMVC), 2018. Available: https://arxiv.org/abs/1802.07934

[8] X. Chen, Y. Yuan, G. Zeng, and J. Wang, "Semi-supervised semantic segmentation with cross pseudo supervision," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2021, pp. 2613-2622. doi: https://doi.org/10.1109/CVPR46437.2021.00264

[9] H. Hu, F. Wei, H. Hu, Q. Ye, J. Cui, and L. Wang, "Semi-supervised semantic segmentation via adaptive equalization learning," in Adv. Neural Inform. Process. Syst. (NeurIPS), 2021, pp. 22106-22118. Available: https://arxiv.org/abs/2110.05474

[10] Y. Wang et al., "Semi-supervised semantic segmentation using unreliable pseudo-labels," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2022, pp. 4248-4257. doi: https://doi.org/10.1109/CVPR52688.2022.00421

[11] Y. Liu et al., "Perturbed and strict mean teachers for semi-supervised semantic segmentation," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2022, pp. 4258-4267. doi: https://doi.org/10.1109/CVPR52688.2022.00422

[12] B. Sun, Y. Yang, L. Zhang, M.-M. Cheng, and Q. Hou, "CorrMatch: Label propagation via correlation matching for semi-supervised semantic segmentation," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2024, pp. 3097-3107. doi: https://doi.org/10.1109/CVPR52733.2024.00299

[13] J. Ju, H. Noh, Y. Wang, M. Seo, and D. G. Choi, "CAFS: Class adaptive framework for semi-supervised semantic segmentation," arXiv preprint arXiv:2303.11606, 2023. Available: https://arxiv.org/abs/2303.11606

[14] E. Tarubinga, J. Kalafatovich, and S.-W. Lee, "CW-BASS: Confidence-weighted boundary-aware learning for semi-supervised semantic segmentation," in Proc. Int. Joint Conf. Neural Netw. (IJCNN), 2025. doi: https://doi.org/10.1109/IJCNN64981.2025.11227871

[15] Y. Gal and Z. Ghahramani, "Dropout as a Bayesian approximation: Representing model uncertainty in deep learning," in Int. Conf. Mach. Learn. (ICML), 2016, pp. 1050-1059. Available: https://proceedings.mlr.press/v48/gal16.html

[16] D. Mahapatra et al., "Semi-supervised and active learning for automatic segmentation of Crohn's disease," in Med. Image Comput. Comput. Assist. Interv. (MICCAI), 2013, pp. 214-221. doi: https://doi.org/10.1007/978-3-642-40763-5_27

[17] Y. Guo, F. Wang, Y. Xiang, and H. You, "Semisupervised semantic segmentation with certainty-aware consistency training for remote sensing imagery," IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 16, pp. 2900-2914, 2023. doi: https://doi.org/10.1109/JSTARS.2023.3255553

[18] Z. Qiu, W. Gan, Z. Yang, R. Zhou, and H. Gan, "Dual uncertainty-guided multi-model pseudo-label learning for semi-supervised medical image segmentation," Math. Biosci. Eng., vol. 21, no. 2, pp. 2212-2232, 2024. doi: https://doi.org/10.3934/mbe.2024097

[19] R. S. Thakur and V. K. Kurmi, "Uncertainty and energy based loss guided semi-supervised semantic segmentation," in Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV), 2025. doi: https://doi.org/10.1109/WACV61041.2025.00780

[20] S. Kumari and P. Singh, "Annotation ambiguity aware semi-supervised medical image segmentation," in Proc. IEEE/CVF Conf.

Comput. Vis. Pattern Recognit. (CVPR), 2025, pp. 10404-10413. doi: https://doi.org/10.1109/CVPR52734.2025.00973

[21] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in Int. Conf. Mach. Learn. (ICML), 2020, pp. 1597-1607. Available: https://proceedings.mlr.press/v119/chen20j.html

[22] K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick, "Momentum contrast for unsupervised visual representation learning," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2020, pp. 9729-9738. doi: https://doi.org/10.1109/CVPR42600.2020.00975

[23] Y. Zhong, B. Yuan, H. Wu, Z. Yuan, J. Peng, and Y.-X. Wang, "Pixel contrastive-consistent semi-supervised semantic segmentation," in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), 2021, pp. 7253-7262. doi: https://doi.org/10.1109/ICCV48922.2021.00718

[24] S. Liu, S. Zhi, E. Johns, and A. J. Davison, "Bootstrapping semantic segmentation with regional contrast," in Int. Conf. Learn. Represent. (ICLR), 2021. Available: https://arxiv.org/abs/2104.04465

[25] H.-M. Xu, L. Liu, Q. Bian, and Z. Yang, "Semi-supervised semantic segmentation with prototype-based consistency regularization," in Adv. Neural Inform. Process. Syst. (NeurIPS), 2022, pp. 26007-26020. doi: https://doi.org/10.52202/068431-1886

[26] H. Zhang, P. Ma, J. Liu, J. Lian, and Y. Ma, "Prototype-augmented mean teacher for robust semi-supervised medical image segmentation," Pattern Recognit., vol. 166, p. 111722, Oct. 2025. doi: https://doi.org/10.1016/j.patcog.2025.111722

[27] T. Yue et al., "Dual prototypes contrastive learning based semi-supervised segmentation method for intelligent medical applications," Eng. Appl. Artif. Intell., vol. 154, p. 110905, Aug. 2025. doi: https://doi.org/10.1016/j.engappai.2025.110905

[28] C. Fu, Q. Zhuang, S. Lei, A. Tian, and Z. Sh, "CPDM: Class prototypes with dynamic multidomain mixed perturbation consistency for semi-supervised image segmentation," IEEE Trans. Geosci. Remote Sens., vol. 64, 2025. doi: https://doi.org/10.1109/TGRS.2025.3645047

[29] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The Pascal Visual Object Classes (VOC) Challenge," Int. J. Comput. Vis., vol. 88, no. 2, pp. 303-338, 2010. doi: https://doi.org/10.1007/s11263-009-0275-4

[30] M. Cordts et al., "The cityscapes dataset for semantic urban scene understanding," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2016, pp. 3213-3223. doi: https://doi.org/10.1109/CVPR.2016.350

[31] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in Proc. Eur. Conf. Comput. Vis. (ECCV), 2018, pp. 801-818. doi: https://doi.org/10.1007/978-3-030-01234-2_49

[32] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2016, pp. 770-778. doi: https://doi.org/10.1109/CVPR.2016.90

[33] S. Yun, D. Han, S. J. Oh, S. Chun, J. Choe, and Y. Yoo, "CutMix: Regularization strategy to train strong classifiers with localizable features," in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), 2019, pp. 6022-6031. doi: https://doi.org/10.1109/ICCV.2019.00612

[34] Z. Zhao, L. Yang, S. Long, J. Pi, L. Zhou, and J. Wang, "Augmentation matters: A simple-yet-effective approach to semi-supervised semantic segmentation," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2023, pp. 11350-11359. doi: https://doi.org/10.1109/CVPR52729.2023.01092

[35] X. Wang, H. Bai, L. Yu, Y. Zhao, and J. Xiao, "Towards the uncharted: Density-descending feature perturbation for semi-supervised semantic segmentation," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2024, pp. 3303-3312. doi: https://doi.org/10.1109/CVPR52733.2024.00318

[36] H. Wang, Q. Zhang, Y. Li, and X. Li, "AllSpark: Reborn labeled features from unlabeled in transformer for semi-supervised semantic segmentation," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2024, pp. 3627-3636. doi: https://doi.org/10.1109/CVPR52733.2024.00348

[37] J. Liu, P. Liu, and L. He, "A Confidence-Variance Theory for Pseudo-Label Selection in Semi-Supervised Learning," arXiv preprint arXiv:2601.11670, 2026. Available: https://arxiv.org/abs/2601.11670