

Two Losses, One Goal: Balancing Conflict Gradients for Semi-supervised Semantic Segmentation

Rui Sun^{1*} Huayu Mai^{2,3*†} Wangkai Li^{2,3} Yujia Chen^{2,3} Yuan Wang³

¹Shenzhen International Graduate School, Tsinghua University

²National Key Laboratory of Deep Space Exploration, Deep Space Exploration Laboratory

³University of Science and Technology of China

{issunrui, mai556, lwkllwk, yujia_chen, wy2016}@mail.ustc.edu.cn

Abstract

Semi-supervised semantic segmentation has attracted considerable attention as it alleviates the need for extensive pixel-level annotations. However, existing methods often overlook the potential optimization conflict between supervised and unsupervised learning objectives, leading to sub-optimal performance. In this paper, we identify this under-explored issue and propose a novel Pareto Optimization Strategy (POS) to tackle it. POS aims to find a descent gradient direction that benefits both learning objectives, thereby facilitating model training. By dynamically assigning weights to the gradients at each iteration based on the model's learning status, POS effectively reconciles the intrinsic tension between the two objectives. Furthermore, we analyze POS from the perspective of gradient descent in random batch sampling and propose the Magnitude Enhancement Operation (MEO) to further unleash its potential by considering both direction and magnitude during gradient integration. Extensive experiments on challenging benchmarks demonstrate that integrating POS into existing semi-supervised segmentation methods yields consistent improvements across different data splits and architectures (CNN, Transformer), showcasing its effectiveness.

1. Introduction

Semantic segmentation, which aims to predict a specific semantic class for each pixel, has achieved remarkable success due to recent advances in deep neural networks in computer vision [8, 16]. It has widespread applications [9, 25, 30, 36, 44, 45, 47, 54], such as visual understanding [13, 55] and autonomous driving [2]. However, its data-driven nature makes it labor-intensive and time-consuming to gather the massive pixel-level annota-

tions required for training data. To alleviate this data-hunger issue, considerable research has turned attention to semi-supervised semantic segmentation. The challenge lies in effectively leveraging limited labeled data in conjunction with a large amount of unlabeled data to improve the model's generalization performance for robust segmentation.

Recently, the teacher-student network scheme has dominated this field due to its simplicity yet competitive performance. In this scheme, the student network is guided by two separate sources of supervision signals: (1) the ground truth for labeled data (yielding *supervised* loss), and (2) the pseudo labels generated by the teacher network for unlabeled data (forming *unsupervised* learning objective). Specifically, the unsupervised loss is constructed in the form of consistency regularization, that is, the teacher network generates pseudo labels with weak augmentation perturbations to instruct the counterparts from the student network under the presence of strong augmentation perturbations, following the smoothness assumption [7].

However, behind the promising performance, we reveal a risk that is ignored by previous methods: the optimization direction of student network parameters is governed by both supervised and unsupervised learning objectives. This means that the student network is expected to simultaneously minimize both learning objectives, but usually, there may not exist a set of parameters that can satisfy this goal. In other words, supervised and unsupervised objectives may conflict during the optimization process. Taking the widely-used Pascal dataset [13] under 366 partition protocols as an example, Figure 1 (a) illustrates the negative cosine similarity between gradients derived from supervised and unsupervised losses, indicating that indeed existing undesirable conflicts in optimization direction during training. Previous methods [39, 46, 60] tend to resort to heuristically assigning *equal* and *fixed* gradient weights, which are disconnected from the learning status, to instill the student network (supervised and unsupervised losses are summed

*Equal contribution

†Corresponding author

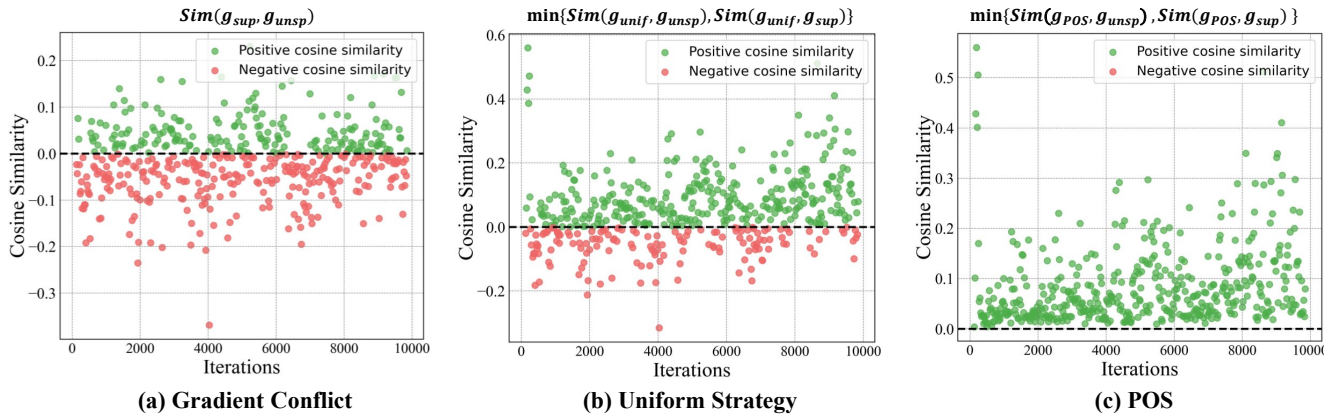


Figure 1. Gradient analysis of different optimization strategies in semi-supervised semantic segmentation. (a) Gradient Conflict: The cosine similarity between gradients derived from supervised (g_{sup}) and unsupervised (g_{unsp}) losses, indicating the existence of optimization conflicts. (b) Uniform Strategy: The minimum cosine similarity between the integrated gradient (g_{unif}) (equal and fixed weights) and the supervised (g_{sup}) or unsupervised (g_{unsp}) gradient, showing the inability to reconcile contradictory directions. (c) POS (Ours): The minimum cosine similarity between the integrated gradient from POS (g_{pos}) and the supervised (g_{sup}) or unsupervised (g_{unsp}) gradient, demonstrating POS’s ability to maintain non-negative cosine similarity and effectively balance the optimization of both objectives.

equally). We refer to this strategy as the *uniform strategy* and visualize the minimum cosine similarity between the gradients integrated by this strategy and those originating from the supervised and unsupervised losses respectively, to measure the degree of conflict. As shown in Figure 1 (b), this strategy is fragile to the variations of the two losses throughout the training process, failing to reconcile their contradictory directions and thus suffering from potential optimization risks (*i.e.*, negative cosine similarity indicating gradient conflicts). This is a corollary of the lack of sophisticated consideration and the ad hoc control of gradient weight hyperparameters, compromising the model’s capability. Therefore, it is highly desirable to balance the conflict gradients, that is, *to harmonize the two losses and strive towards one unified goal*.

In this paper, we aim to tackle the under-explored problem of optimization conflict in semi-supervised semantic segmentation, and offer a Pareto optimization perspective [6, 41], to strive to empower the integrated gradient to resonate favorably with both supervised and unsupervised losses against conflict. In specific, we prepend a dedicated Pareto Optimization Strategy (**POS**) to enable gradient integration, which aims to formally find a steep gradient direction that benefits both learning objectives to balance conflicting gradients, thereby facilitating model training. As the variations of the two losses evolve throughout the training process, POS *dynamically* assigns different weights to the gradients at each iteration in pursuit of perceiving the learning status of the model. This is distinct from previous methods that integrate gradients in a manually *fixed* manner, thus being brittle to the model’s learning process, resulting in inferior performance. Note that the integrated gradient weights derived from mathematical derivation are theoretic-

cally guaranteed to converge towards Pareto optimality by reconciling the intrinsic tension between supervised and unsupervised gradients. This is supported by the observation (depicted in Figure 1 (c)) that the integrated gradients from POS maintain *non-negative* cosine similarity with the gradients of both supervised and unsupervised losses.

Furthermore, we take a step further to understand POS from the perspective of gradient descent in random batch sampling during the semi-supervised learning process, revealing its tight connection with the model’s capability. Figure 3 illustrates the static distribution of gradient magnitudes from supervised and unsupervised losses, indicating that the unsupervised gradient magnitude is smaller than the supervised one with a smaller batch sampling covariance. We conjecture that the underlying reason might be the discrepancy in learning difficulty: training with labeled data, which has true ground truth (explicit external learning objectives), is more challenging; while training with unlabeled data relies on pseudo label predictions from the teacher network, which is inherently homologous to the student network (implicit supervision signals with common-mode biased noise), is relatively easier. This allows labeled data to dominate the training process, which is also confirmed by [52]. In light of this observation, we further analyze and empirically verify that POS may lead to a sharp minima, despite its superior performance compared to the uniform strategy, potentially limiting the further optimization of the model. Then, we propose the Magnitude Enhancement Operation (MEO) to further unleash the potential of POS, which takes the *direction* and *magnitude* into account during gradient integration to enable the model to converge to a flatter minima, thereby enhancing performance. In this way, POS, equipped with MEO, favors a better capacity to inte-

grate gradients, fulfilling conflict-free optimization conditions, coupled with enhanced magnitude for enhanced generalization, leading to improved segmentation performance.

In this work, our contributions can be summarized as follows: (1) We reveal the under-explored optimization conflict issue in semi-supervised semantic segmentation, where supervised and unsupervised learning objectives may have contradictory optimization directions. (2) We propose a novel Pareto Optimization Strategy (POS) to tackle the optimization conflict issue by finding a descent gradient direction that benefits both learning objectives. Then, we further enhance POS with the Magnitude Enhancement Operation (MEO) to improve segmentation performance. (3) Extensive experiments on challenging benchmarks demonstrate that integrating POS into existing semi-supervised segmentation methods yields consistent improvements across different data splits and architectures (CNN, Transformer), showcasing its effectiveness.

2. Related Work

Semi-Supervised Learning. Semi-supervised learning [14, 40, 66] (SSL) has emerged as a pivotal paradigm in machine learning, bridging the gap between supervised and unsupervised learning by leveraging a small amount of labeled data in conjunction with a large pool of unlabeled data. Recently, deep learning-based SSL methods have made significant progress. Their core technical mechanisms encompass two main strategies: bootstrapped label generation via prediction confidence (pseudo-labeling) [1, 5, 28, 62] and consistency regularization [27, 50, 51, 57] across perturbations. The former alleviates the scarcity of annotations by utilizing the network’s outputs on unlabeled data as proxy labels, effectively addressing the issue of limited labeled data. The latter enhances the model’s generalization robustness by enforcing stable representations across various data perturbations, such as random cropping and color jittering. Current SSL methods [3, 4, 15, 42, 59] have demonstrated the synergy between consistency regularization and pseudo-labeling, which generates pseudo-labels from weakly augmented unlabeled images for strongly augmented versions of the same images. It is worth noting that, although the aforementioned methods have achieved breakthrough progress in image classification tasks, their pixel-wise classification assumptions fundamentally conflict with the dense prediction requirements of semantic segmentation, leading to a decline in pixel-level accuracy when directly transferred.

Semi-Supervised Semantic Segmentation. Semi-supervised semantic segmentation [35, 38, 48, 49] has achieved remarkable progress through hybrid approaches combining pseudo-labeling and consistency regularization [26, 65]. Representative works like UniMatch [60] adapt FixMatch [42] with domain-specific augmentations,

establishing strong baselines for SSL. Subsequent advancements focus on four key directions: (1) Data augmentation innovation: Advanced augmentation strategies expand the effective search space. AugSeg [64] enhances RandAugment [11] with controlled randomness, while iMAS [63] introduces adaptive augmentations guided by model states. (2) Teacher network optimization: Methods like Switch [39] address EMA coupling through dual-teacher ensembling. This direction emphasizes improved knowledge distillation mechanisms. (3) External knowledge integration: LOGIC [32] employs symbolic reasoning for error mitigation, while SemiVL [20] incorporates CLIP encoders for text-conditional priors, enhancing label quality through multi-modal reasoning. (4) Consistency enhancement: Novel approaches like RankMatch [37] exploit pixel correlations, and MPMC [19] incorporates contextual class information for robust supervision. Recent studies also explore weighting strategies [43] to optimize unlabeled data utilization. Differing from prior methods, this work investigates the inherent conflict in optimizing supervised and unsupervised learning objectives in semi-supervised tasks.

3. Method

3.1. Preliminaries

Given a labeled set $\mathcal{D}^l = \{(\mathbf{x}_i^l, \mathbf{y}_i^l)\}_{i=1}^{N^l}$ and an unlabeled set $\mathcal{D}^u = \{\mathbf{x}_i^u\}_{i=1}^{N^u}$, where N^l and N^u denote the number of labeled and unlabeled images, respectively, and $N^u \gg N^l$, semi-supervised semantic segmentation aims to effectively leverage limited labeled data in conjunction with numerous unlabeled data to improve the model’s generalization performance for robust segmentation. As depicted in Figure 2, the teacher-student network scheme equipped with a teacher network f_T and a student network f_S has dominated this field. In this scheme, the student network is guided by two separate sources of supervision signals, including: (1) the ground truth for labeled data (yielding supervised loss \mathcal{L}_{sup}), and (2) the pseudo labels generated by the teacher network for unlabeled data (forming unsupervised learning objective $\mathcal{L}_{un\text{sup}}$). The teacher network can be either an exponentially moving average (EMA) version of the student network or identical to it. In specific, for the labeled data \mathbf{x}_l , the supervised loss \mathcal{L}_{sup} can be formulated as:

$$\mathcal{L}_{sup} = \frac{1}{N^l} \sum_{i=1}^{N^l} \frac{1}{HW} \sum_{j=1}^{HW} \ell_{ce}(\mathbf{y}_{ij}^l, f_S(\mathbf{x}_i^l)_j), \quad (1)$$

where H and W denote the height and width of the input image, ℓ_{ce} denotes the standard pixel-wise cross-entropy loss. For the unlabeled data, the teacher network f_T takes the weak augmentation perturbations $aug(\mathbf{x}_i^u)$ as input and

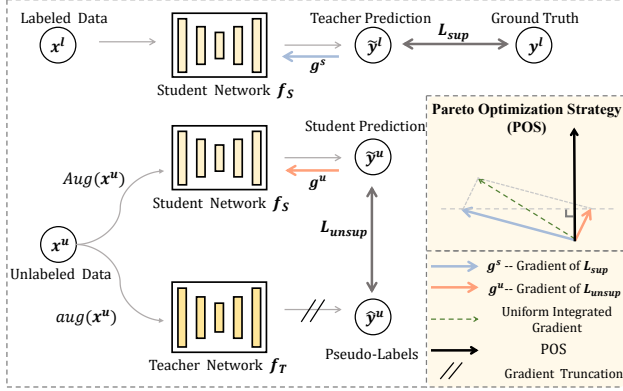


Figure 2. The framework of our proposed Pareto Optimization Strategy. The student network is guided by two sources of supervision, including the ground truth for the labeled data and the pseudo-labels generated by the teacher network for the unlabeled data. Then, we propose a novel Pareto Optimization Strategy (POS) to tackle the optimization conflict issue by finding a descent gradient direction that benefits both learning objectives. Furthermore, we take a step further to understand POS from the perspective of gradient descent in random batch sampling during the semi-supervised learning process. Finally, we enhance POS with the Magnitude Enhancement Operation (MEO) to improve segmentation performance.

generates pseudo-labels \hat{y}_{ij}^u for the student network f_S as:

$$\hat{y}_{ij}^u = \begin{cases} \arg \max f_T(\text{aug}(\mathbf{x}_i^u))_j, & c_{ij}^u > \gamma \\ \text{ignore_index}, & \text{otherwise} \end{cases}, \quad (2)$$

where $c_{ij}^u = \max f_T(\text{aug}(\mathbf{x}_i^u))_j$ denotes the confidence of the teacher prediction for j^{th} pixel and γ denotes the confidence threshold to filter unreliable pseudo-labels from training. In this way, we can obtain the unsupervised loss \mathcal{L}_{unsup} by imposing consistency regularization:

$$\mathcal{L}_{unsup} = \frac{1}{N^u} \sum_{i=1}^{N^u} \frac{1}{HW} \sum_{j=1}^{HW} \ell_{ce}(\hat{y}_{ij}^u, f_S(\text{Aug}(\mathbf{x}_i^u))_j), \quad (3)$$

where $\text{Aug}(\cdot)$ means the strong augmentation. Finally, the overall loss can be formalized as:

$$\mathcal{L} = \alpha^s \mathcal{L}_{sup} + \alpha^u \mathcal{L}_{unsup}, \quad (4)$$

where $\alpha^s + \alpha^u = 1$. Typically, the two losses are summed equally without sophisticated consideration (*i.e.*, $\alpha^s = \alpha^u = \frac{1}{2}$), which is referred to as the uniform strategy.

3.2. Pareto Optimization Strategy

Despite the promising performance of previous methods, we reveal the under-explored optimization conflict issue in semi-supervised semantic segmentation, where supervised and unsupervised learning objectives may have contradictory optimization directions, as shown in Figure 1 (a). Formally, we denote \mathbf{g}_S^s and \mathbf{g}_S^u as the gradients computed on

each mini-batch S , derived from the supervised loss \mathcal{L}_{sup} and unsupervised loss \mathcal{L}_{unsup} , respectively. To effectively balance these potentially conflicting gradients and harmonize the two losses and strive towards one unified goal, we draw inspiration from the Pareto optimization [6, 41] to strive to balance conflicting gradients. In this way, at each iteration, the gradients are assigned different weights, and the weighted combination of these gradients forms the final descent direction that is common to both objectives. Ultimately, the model parameters converge towards Pareto optimality by reconciling the intrinsic tension between supervised and unsupervised gradients, in which no objective can be advanced without harming another objective. Mathematically, the Pareto optimization process can be formulated as:

$$\begin{aligned} \min_{\alpha^s, \alpha^u \in \mathcal{R}} \quad & \|\alpha^s \mathbf{g}_S^s + \alpha^u \mathbf{g}_S^u\|^2 \\ \text{s.t.} \quad & \alpha^s, \alpha^u \geq 0, \alpha^s + \alpha^u = 1, \end{aligned} \quad (5)$$

where $\|\cdot\|^2$ represents L_2 norm. The optimization problem described in Equation 5 is equivalent to finding the minimum norm in the convex hull of the family of gradient vectors $\{\mathbf{g}_S^i\}_{i \in \{s, u\}}$. The solution to this problem, denoted as (α^s, α^u) , represents the weights assigned to the supervised and unsupervised gradients, respectively. The integrated gradient, computed as $\alpha^s \mathbf{g}_S^s + \alpha^u \mathbf{g}_S^u$, satisfies one of two conditions [12]. If the minimum-norm of the optimization problem in Equation 5 is 0, the corresponding parameters are Pareto-stationary, which is a necessary condition for Pareto-optimality. On the other hand, if the minimum-norm is non-zero, the integrated gradient provides a descent direction that is common to both supervised and unsupervised objectives. This enables the simultaneous optimization of both objectives, allowing the model to effectively balance the potentially conflicting gradients and achieve a harmonized optimization process. Note that the optimization problem in Equation 5 admits an analytical solution, which can be derived as follows:

$$\begin{cases} \alpha^u = 1, \alpha^s = 0 & \cos \beta \geq \frac{\|\mathbf{g}_S^u\|}{\|\mathbf{g}_S^s\|}, \\ \alpha^u = \frac{(\mathbf{g}_S^s - \mathbf{g}_S^u)^\top \mathbf{g}_S^s}{\|\mathbf{g}_S^u - \mathbf{g}_S^s\|^2}, \alpha^s = 1 - \alpha^u & \text{otherwise}, \\ \alpha^u = 0, \alpha^s = 1 & \cos \beta \geq \frac{\|\mathbf{g}_S^s\|}{\|\mathbf{g}_S^u\|}, \end{cases} \quad (6)$$

where β is the angle between \mathbf{g}_S^s and \mathbf{g}_S^u . During training, the gradients of the supervised and unsupervised losses are calculated separately, and thus they can be treated as independent. Consequently, the obtained weights (α^s, α^u) can be conveniently incorporated into the weighted total loss function (Equation 4). In this way, POS *dynamically* assigns different weights to the gradients at each iteration in pursuit of perceiving the learning status of the model, compared to the previous methods that integrate gradients in a *manual* and *fixed* manner, resulting in inferior performance.

3.3. Understand POS from Gradient Descent View

Furthermore, we take a step further to understand POS from the perspective of gradient descent in random batch sampling during the semi-supervised learning process. Generally speaking, during the training of the student network θ , the gradient objectives corresponding to the unsupervised/supervised losses in Equations 1 and 3 represent the full gradients, which is an ideal computation approach.

However, in practical network optimization, the optimization of both losses is instantiated as random gradient sampling at t -th mini-batches S , for example, for supervised loss, $\mathbf{g}_S^s(\theta(t)) = \frac{1}{|S|} \sum_{i=1}^{|S|} \nabla_{\theta(t)} \ell_{ce}(X_i, Y_i)$, and (X_i, Y_i) represents the i -th sample in the mini-batch. When the batch size is sufficiently large, according to the central limit theorem, \mathbf{g}_S^s is unbiased estimation of full gradient [21], $\mathbf{g}_N^s(\theta(t)) = \frac{1}{|N^l|} \sum_{i=1}^{|N^l|} \nabla_{\theta(t)} \ell_{ce}(X_i, Y_i)$, where N^l denotes the number of labeled data. For brevity, we omit the pixel-wise computations and use X_i and Y_i to represent the calculations over the entire image, which does not affect the subsequent analysis. Based on this, the gradient $\mathbf{g}_S^s(\theta(t))$ is random variables with covariance $\frac{1}{|S|} C^s$:

$$\mathbf{g}_S^s(\theta(t)) \sim \mathcal{N}\left(\mathbf{g}_N^s(\theta(t)), \frac{1}{|S|} C^s\right), \quad (7)$$

where C^s represents the batch sampling covariance. Similarly,

$$\mathbf{g}_S^u(\theta(t)) \sim \mathcal{N}\left(\mathbf{g}_N^u(\theta(t)), \frac{1}{|S|} C^u\right). \quad (8)$$

Before further investigating POS from the gradient descent perspective, we analyze the gradient magnitude distributions derived from the supervised and unsupervised learning objectives. As illustrated in Figure 3(a), the unsupervised gradient magnitude is generally smaller than the supervised one, with a smaller batch sampling covariance. We conjecture that this discrepancy might stem from the difference in learning difficulty between the two objectives. Training with labeled data, which has true ground truth (explicit external learning objectives), is more challenging. In contrast, training with unlabeled data relies on pseudo label predictions from the teacher network, which is inherently homologous to the student network (implicit supervision signals with common-mode biased noise), and is relatively easier. This allows labeled data to dominate the training process, as also confirmed by [52]. Based on this analysis, we can make the following remark:

Remark 1. *The gradient of the supervised loss tends to have a larger magnitude and larger batch sampling covariance than the unsupervised loss, statistically.*

Based on the observation in Remark 1 that $\|\mathbf{g}_S^u\| < \|\mathbf{g}_S^s\|$ holds statistically, we revisit Equation 6. When $\cos \beta \geq$

$\frac{\|\mathbf{g}_S^u\|}{\|\mathbf{g}_S^s\|}$, we have $\alpha^u > \alpha^s$. Otherwise, we can derive:

$$\begin{aligned} \alpha^u - \alpha^s &= \frac{(\mathbf{g}_S^s - \mathbf{g}_S^u)^\top \mathbf{g}_S^s}{\|\mathbf{g}_S^u - \mathbf{g}_S^s\|^2} - \left(1 - \frac{(\mathbf{g}_S^s - \mathbf{g}_S^u)^\top \mathbf{g}_S^s}{\|\mathbf{g}_S^u - \mathbf{g}_S^s\|^2}\right) \\ &> 0. \quad (\|\mathbf{g}_S^u\| < \|\mathbf{g}_S^s\|) \end{aligned} \quad (9)$$

Remark 2. *POS tends to assign a larger weight to the gradient derived from the unsupervised learning objective.*

Note that this conclusion aligns with Figure 3(b), indicating that as the variations of the two losses evolve throughout the training process, POS dynamically assigns different weights to both gradients, with a greater preference for the unsupervised gradient (i.e., $\alpha^u > \frac{1}{2}$).

Now, we are ready to further analyze POS from the gradient descent perspective. First, let's consider the uniform strategy, which integrates gradients by equally summing them. The final gradient is given by $\mathbf{h}_S(\theta(t)) = \frac{1}{2} \mathbf{g}_S^u(\theta(t)) + \frac{1}{2} \mathbf{g}_S^s(\theta(t))$. Based on Equations 7 and 8, we can derive the distribution of $\mathbf{h}_S(\theta(t))$ as follows:

$$\mathbf{h}_S(\theta(t)) \sim \mathcal{N}\left(\frac{1}{2} \mathbf{g}_N^u(\theta(t)) + \frac{1}{2} \mathbf{g}_N^s(\theta(t)), \frac{C^u + C^s}{4|S|}\right). \quad (10)$$

By applying the final gradient $\mathbf{h}_S(\theta(t))$ to update θ , we have:

$$\begin{aligned} \theta(t+1) &= \theta(t) - \eta \mathbf{h}_S(\theta(t)) \\ &= \theta(t) - \eta \mathbf{h}_N(\theta(t)) + \eta \epsilon_t, \end{aligned} \quad (11)$$

where $\mathbf{h}_N(\theta(t)) = \frac{1}{2} \mathbf{g}_N^u(\theta(t)) + \frac{1}{2} \mathbf{g}_N^s(\theta(t))$ represents the full gradient, $\eta > 0$ is the learning rate, and $\epsilon_t \sim \mathcal{N}\left(0, \frac{C^u + C^s}{4|S|}\right)$ is commonly considered as the noise term introduced by the mini-batch approximation [67]. For POS, which dynamically integrates gradients in pursuit of perceiving the learning status of the model, the parameters can be similarly updated as $\theta(t+1) = \theta(t) - \eta \mathbf{h}_S^{\text{POS}}(\theta(t))$, where $\zeta_t \sim \mathcal{N}\left(0, \frac{(\alpha^u)^2 C^u + (\alpha^s)^2 C^s}{|S|}\right)$ is the noise term.

Based on Remark 1, the supervised gradient tends to have a larger batch sampling covariance. Suppose the covariance of the unsupervised gradient and the supervised gradient satisfies $kC^u = C^s$, where $k > 1$. We can then evaluate the relative magnitude of ζ_t and ϵ_t . When the covariance of ζ_t is smaller than that of ϵ_t , it should satisfy:

$$\begin{aligned} (2\alpha^u)^2 C^u + (2\alpha^s)^2 C^s &< C^u + C^s, \\ (\alpha^u - \frac{1}{2})((4k+4)\alpha^u + 2 - 6k) &< 0. \end{aligned} \quad (12)$$

Therefore, when $\frac{1}{2} < \alpha^u < \frac{3k-1}{2k+2}$, the covariance of ζ_t is smaller than that of ϵ_t . Moreover, based on Remark 2, POS assigns a weight of $\frac{1}{2} < \alpha^u \leq 1$ to the unsupervised gradient. Consequently, we have:

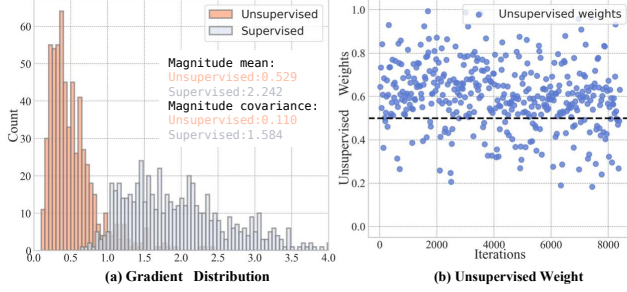


Figure 3. Visualization of gradient distributions and unsupervised weights. (a) Gradient magnitude distributions of supervised and unsupervised branches. (b) Unsupervised weights assigned by POS during training in a dynamic manner.

- (1) For $k \leq 3$, when $\frac{1}{2} < \alpha^u < \frac{3k-1}{2k+2}$, the covariance of ζ_t is smaller than that of ϵ_t ;
- (2) When $k > 3$, we have $\frac{3k-1}{2k+2} > 1$, and thus the covariance of ζ_t is always smaller than that of ϵ_t , regardless of the value of α^u .

Furthermore, as shown in Figure 3 (a), the difference in covariance between the supervised and unsupervised gradients is statistically more than 3 times in experiments. This phenomenon indicates that the scenario where the covariance of ζ_t is smaller than that of ϵ_t frequently occurs in practice. In a nutshell, we further analyze and empirically verify that POS may lead to a sharp minima, despite its superior performance compared to the uniform strategy, potentially limiting the further optimization of the model.

3.4. Magnitude Enhancement Operation

To strive to alleviate the issue discussed above, Then, we propose the Magnitude Enhancement Operation (MEO) to further unleash the potential of POS, which takes the direction and magnitude into account during gradient integration to enable the model to converge to a flatter minima. Specifically, we first solve the optimization problem in Equation 5 to obtain α^u and α^s , which can provide a conflict-free direction. Furthermore, to enhance the noise term derived from random batch sampling, we adopt a simple yet effective strategy, *i.e.*, enhancing the magnitude of the integrated gradient. We use the magnitude of the integrated gradient from the uniform strategy as a reference to adjust the magnitude of our POS gradient to a proper range:

$$\mathbf{h}_S^{\text{POS}} = \frac{\alpha^u \mathbf{g}_S^u + \alpha^s \mathbf{g}_S^s}{\|\alpha^u \mathbf{g}_S^u + \alpha^s \mathbf{g}_S^s\|} \cdot \left\| \frac{1}{2} \mathbf{g}_S^u + \frac{1}{2} \mathbf{g}_S^s \right\|, \quad (13)$$

Based on Remark 2, the smaller multimodal magnitude is associated with a larger weight. In this case, we can have $\lambda = \left\| \frac{1}{2} \mathbf{g}_S^u + \frac{1}{2} \mathbf{g}_S^s \right\| / \|\alpha^u \mathbf{g}_S^u + \alpha^s \mathbf{g}_S^s\| > 1$, which indicates that the noise strength is enhanced. In this way, POS, equipped with Magnitude Enhancement Operation (MEO), demonstrates a better capacity to integrate gradients, fulfilling conflict-free optimization conditions, coupled with en-

Table 1. Quantitative results of different SSL methods on COCO.

Method	1/512	1/256	1/128	1/64	1/32
XC-65					
<i>Sup.-only</i>	22.9	28.0	33.6	37.8	42.2
PseudoSeg _[ICLR'21]	29.8	37.1	39.1	41.8	43.6
PC2Seg _[ICCV'21]	29.9	37.5	40.1	43.7	46.1
CISC-R _[TPAMI'23]	32.1	40.2	42.3	-	-
LogicDiag _[ICCV'23]	33.1	40.3	45.4	48.8	50.5
UniMatch V1 _[CVPR'23]	31.9	38.9	44.4	48.2	49.8
UniMatch V1+Ours	34.0	40.3	45.9	49.3	50.7
DINOv2-S					
UniMatch V2 _[TPAMI'25]	39.3	45.4	53.2	55.0	57.0
UniMatch V2+Ours	40.9	46.8	54.3	56.2	57.7

Table 2. Quantitative results of different SSL methods on Pascal. We report mIoU (%) under various partition protocols.

Method	1/16 (92)	1/8 (183)	1/4 (366)	1/2 (732)	Full (1464)
ResNet-101					
<i>Sup.-only</i>	45.1	55.3	64.8	69.7	73.5
GTA-Seg _[NeurIPS'22]	70.0	73.2	75.6	78.4	80.5
PCR _[NeurIPS'22]	70.1	74.7	77.2	78.5	80.7
iMAS _[CVPR'23]	68.8	74.4	78.5	79.5	81.2
AugSeg _[CVPR'23]	71.1	75.5	78.8	80.3	81.4
Diverse CoT _[ICCV'23]	75.7	77.7	80.1	80.9	82.0
ESL _[ICCV'23]	71.0	74.0	78.1	79.5	81.8
LogicDiag _[ICCV'23]	73.3	76.7	77.9	79.4	-
DAW _[NeurIPS'24]	74.8	77.4	79.5	80.6	81.5
DDFP _[CVPR'24]	75.0	78.0	79.5	81.2	82.0
CorrMatch _[CVPR'24]	76.4	78.5	79.4	80.6	81.8
BeyondPixels _[ECCV'24]	77.3	78.6	79.8	80.8	81.7
UniMatch V1 _[CVPR'23]	75.2	77.2	78.8	79.9	81.2
UniMatch V1+Ours	77.6	78.7	79.7	80.9	81.9
DINOv2-S					
UniMatch V2 _[TPAMI'25]	79.0	85.5	85.9	86.7	87.8
UniMatch V2+Ours	80.7	86.8	87.2	87.5	88.3

hanced magnitude for improved generalization, leading to superior segmentation performance.

4. Experiments

4.1. Experimental Setup

Dataset. COCO [33] presents the most challenging scenario with 118k training and 5k validation images across 81 object categories. PASCAL VOC 2012 [13] comprises 20 object classes with 1,464 training and 1,449 validation images. This is the *classic* dataset setting. Cityscapes [10] is an urban scene understanding dataset, containing 2,975

Table 3. Quantitative results of different SSL methods on Cityscapes. We report mIoU (%) under various partition protocols.

Method	1/16 (186)	1/8 (372)	1/4 (744)	1/2 (1488)
ResNet-101				
<i>Sup.-only</i>	66.3	72.8	75.0	78.0
GTA-Seg _[NeurIPS'22]	69.4	72.0	76.1	-
PCR _[NeurIPS'22]	73.4	76.3	78.4	79.1
iMAS _[CVPR'23]	74.3	77.4	78.1	79.3
AugSeg _[CVPR'23]	75.2	77.8	79.6	80.4
Diverse CoT _[ICCV'23]	75.7	77.4	78.5	-
ESL _[ICCV'23]	75.1	77.2	78.9	80.5
LogicDiag _[ICCV'23]	76.8	78.9	80.2	81.0
DAW _[NeurIPS'24]	76.6	78.4	79.8	80.6
DDFP _[CVPR'24]	77.1	78.2	79.9	80.8
CorrMatch _[CVPR'24]	77.3	78.5	79.4	80.4
BeyondPixels _[ECCV'24]	78.5	79.2	80.9	81.3
UniMatch V1 _[CVPR'23]	76.6	77.9	79.2	79.5
UniMatch V1+Ours	77.6	78.5	79.8	80.4
DINOV2-S				
UniMatch V2 _[TPAMI'25]	80.6	81.9	82.4	82.6
UniMatch V2+Ours	81.4	82.7	83.0	83.2

training and 500 validation images. across 19 categories. The initial 30 semantic classes are re-mapped into 19 classes for the semantic segmentation task.

4.2. Implementation Details

For a fair comparison, for UniMatch V1, we use ResNet-101 [16] pretrained on ImageNet [24] as the backbone and DeepLabv3+ [8] as the decoder. the crop size is set as 321×321 for PASCAL/COCO and 801×801 for Cityscapes, respectively. We adopt stochastic gradient descent (SGD) optimizer with an initial learning rate of 0.001 for PASCAL and 0.005 for Cityscapes. And for UniMatch V2, we adopt DINOv2-S [17] as the backbone, simple DPT as the decoder. The crop size is set as 518×518 for PASCAL/COCO, and 798×798 for Cityscapes, respectively. We use the AdamW optimizer with weight decay of 0.01 for training. Polynomial Decay learning rate policy is applied throughout the whole training. The strong augmentation $Aug(\cdot)$ contains random color jitter, grayscale and Gaussian blur. The weak augmentation $aug(\cdot)$ consists of random crop, resize and horizontal flip. Further when training a baseline integrated with our method, we use the same weak and strong augmentations as used by the corresponding baseline.

4.3. Comparison with State-of-the-art Methods

We integrate POS into two representative SSL frameworks: UniMatch [60] and UniMatch V2 [61], and evaluate its per-

Table 4. Performance comparison of different components.

Configuration	Pascal		COCO	
	92	183	1/512	1/256
Baseline	75.2	77.2	31.9	38.9
Baseline+POS w/o MEO	77.0	78.1	33.4	39.8
Baseline+POS w/ MEO	77.6	78.7	34.0	40.3

Table 5. Weight Analysis on Pascal in 1/16(92).

	Fixed			Dynamic(POS)
α^u	2	1	1	-
α^s	1	1	2	-
mIoU	74.9	75.2	74.1	77.6

formance with both CNN and Transformer backbones under various partition protocols, following common practices.

Results on COCO. In Table 1, we compare our method with previous approaches, including PseudoSeg [68], PC2Seg [65], CISC-R [56] and LogicDiag [32]. Our method outperforms previous methods on all data splits, whether using CNN architectures (XC-65) or Transformer architectures (DINOv2-S). For example, under the 1/512 partition protocol, our performance of our method surpasses UniMatch by 2.1% and 1.6% on the two architectures.

Results on PASCAL. Table 2 shows the comparison of our method with the SOTA methods on PASCAL, including GTA-Seg [22], PCR [58], iMAS [63], AugSeg [64], Diverse CoT [31], ESL [34], LogicDiag [32], DAW [46], DDFP [53], CorrMatch [43], BeyondPixels [18]. Compared with the supervised-only (*Sup.-only*) model, our method achieves considerable performance improvements, demonstrating that our POS can effectively address the optimization conflict issue. Moreover, in the label-scarce scenario, *e.g.*, 1/16 (92), our approach achieves 77.6% and 80.7% mIoU with the backbone ResNet-101 and DINOv2-S, boosting the SOTA UniMatch by 2.4% and 1.7%, respectively. These superior results prove that our optimized descent gradient direction can benefit both supervised and unsupervised learning objectives.

Results on Cityscapes. Table 3 presents a performance comparison between our method and competitors on the Cityscapes dataset, demonstrating the consistently superior performance of our approach. Specifically, compared with the leading UniMatch, incorporating our POS still yields improvements of 1.0% and 0.8% on 1/16 split, demonstrating that the gradient optimization contradiction is indeed an important issue that existing methods have not explored.

Qualitative Results. We compare the qualitative results of our method with different methods on the PASCAL dataset in Figure 4. Our method can effectively distinguish closely adjacent objects and identify complete object regions (*e.g.*,

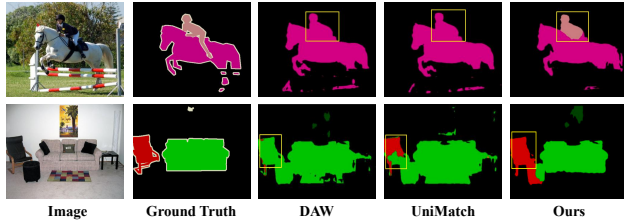


Figure 4. Qualitative comparison with different methods. Note that significant improvements are marked with yellow boxes.

the man on horseback and the chair beside the sofa).

4.4. Ablation Study

Effectiveness of different components. In Table 4, we compare the effectiveness of the proposed POS and MEO through experiments on two datasets. We validate the importance of each component by adding them in turn to the baseline [60] with CNN architecture. (1) Compared to the baseline, the utilization of POS brings obvious improvements (e.g., 1.8%, 0.9% on Pascal and 1.5%, 0.9% on COCO), indicating that POS can enable gradient integration and effectively find a steep gradient direction that benefits both supervised learning objectives and unsupervised learning objectives to balance conflicting gradients. (2) The performance boost brought by MEO demonstrates that it can further unleash the potential of POS, taking the direction and magnitude into account during gradient integration to enable the model to converge to a flatter minima. Overall, the collaboration of POS and MEO is extremely beneficial for semi-supervised semantic segmentation and enhances the performance effectively.

Effectiveness of the dynamic weight. In Table 5, we compare the different assigned weights on the Pascal dataset in the label-scarce scenario, e.g., 1/16 (92). For a fixed weight allocation, an equal ratio (e.g., 1:1) of supervised and unsupervised components achieves the best performance, indicating that any fixed bias will lead to suboptimal optimization of the model. After using POS for dynamic weight allocation, the performance improved from 75.2% to 77.6%, indicating that POS maintains non-negative cosine similarity with the gradients of both supervised and unsupervised losses, which is beneficial to the overall training process.

Gradient distribution and unsupervised weight analysis. Figure 3 (a) illustrates the gradient magnitude distributions of both supervised and unsupervised branches in the model. The unsupervised gradient tends to have a smaller magnitude compared to the supervised gradient. Moreover, the covariance of the unsupervised gradient (0.110) is significantly smaller than that of the supervised gradient (1.584). This observation aligns with Remark 1. Figure 3 (b) shows the unsupervised weights assigned by POS throughout the training iterations. This phenomenon is consistent with Re-

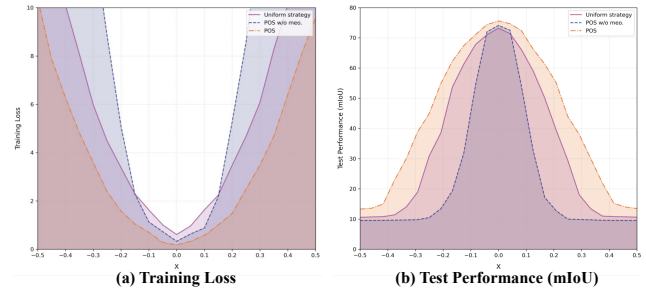


Figure 5. Visualization of training loss and test performance (mIoU). Loss landscape visualization of different gradient integration strategies.

mark 2, implying that POS tends to assign larger weights to the unsupervised gradient to achieve a balance between supervised and unsupervised learning. Besides, POS favorably manifests assignment of different weights to perceive the learning status of the model.

Loss landscape analysis. To take a closer look at the noise strength of POS during optimization derived by mini-batch random sampling. Figure 5 presents the loss landscape visualization [23, 29] of different gradient integration strategies, including the uniform strategy, POS without MEO, and POS with MEO. The training loss landscape in Figure 5 (a) reveals that POS with MEO achieves a smoother and flatter local minimum compared to the other two strategies. This observation suggests that the proposed POS with MEO approach can effectively converge to a flatter minima, thereby enhancing performance. The test performance landscape in Figure 5 (b) further validates the superiority of POS with MEO. The landscape of POS with MEO exhibits a flatter minima region around the optimal point, indicating enhanced generalization capability. In contrast, the uniform strategy demonstrates narrower peaks, implying potential overfitting issues and reduced generalization performance. By integrating gradients in a conflict-free manner and enhancing the magnitude of the integrated gradient, POS with MEO can guide the model towards a more robust and generalizable solution, ultimately leading to improved semi-supervised learning performance.

5. Conclusion

In this paper, we reveal the under-explored optimization conflict issue in semi-supervised semantic segmentation, and propose a novel Pareto Optimization Strategy (POS) to tackle the optimization conflict issue by finding a descent gradient direction that benefits both learning objectives. Then, we further enhance POS with the Magnitude Enhancement Operation (MEO) to improve segmentation performance. Extensive experimental results on challenging benchmarks show the effectiveness.

References

- [1] Eric Arazo, Diego Ortego, Paul Albert, Noel E O'Connor, and Kevin McGuinness. Pseudo-labeling and confirmation bias in deep semi-supervised learning. In *2020 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2020. 3
- [2] Saeid Asgari Taghanaki, Kumar Abhishek, Joseph Paul Cohen, Julien Cohen-Adad, and Ghassan Hamarneh. Deep semantic segmentation of natural and medical images: a review. *Artificial Intelligence Review*, 54:137–178, 2021. 1
- [3] David Berthelot, Nicholas Carlini, Ian Goodfellow, Nicolas Papernot, Avital Oliver, and Colin A Raffel. Mixmatch: A holistic approach to semi-supervised learning. *Advances in neural information processing systems*, 32, 2019. 3
- [4] David Berthelot, Rebecca Roelofs, Kihyuk Sohn, Nicholas Carlini, and Alex Kurakin. Adamatch: A unified approach to semi-supervised learning and domain adaptation. *arXiv preprint arXiv:2106.04732*, 2021. 3
- [5] Paola Cascante-Bonilla, Fuwen Tan, Yanjun Qi, and Vicente Ordonez. Curriculum labeling: Revisiting pseudo-labeling for semi-supervised learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 6912–6920, 2021. 3
- [6] Yair Censor. Pareto optimality in multiobjective problems. *Applied Mathematics and Optimization*, 4(1):41–59, 1977. 2, 4
- [7] Olivier Chapelle, Bernhard Scholkopf, and Alexander Zien. Semi-supervised learning (chapelle, o. et al., eds.; 2006)[book reviews]. *IEEE Transactions on Neural Networks*, 20(3):542–542, 2009. 1
- [8] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European conference on computer vision (ECCV)*, pages 801–818, 2018. 1, 7
- [9] Yujia Chen, Rui Sun, Wangkai Li, Huayu Mai, Naisong Luo, Yuwen Pan, and Tianzhu Zhang. Alleviate and mining: Rethinking unsupervised domain adaptation for mitochondria segmentation from pseudo-label perspective. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2025. 1
- [10] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3213–3223, 2016. 6
- [11] Ekin D Cubuk, Barret Zoph, Jonathon Shlens, and Quoc V Le. Randaugment: Practical automated data augmentation with a reduced search space. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pages 702–703, 2020. 3
- [12] Jean-Antoine Désidéri. Multiple-gradient descent algorithm (mgda) for multiobjective optimization. *Comptes Rendus Mathématique*, 350(5-6):313–318, 2012. 4
- [13] Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes (voc) challenge. *International journal of computer vision*, 88:303–338, 2010. 1, 6
- [14] Yves Grandvalet and Yoshua Bengio. Semi-supervised learning by entropy minimization. *Advances in neural information processing systems*, 17, 2004. 3
- [15] Lan-Zhe Guo and Yu-Feng Li. Class-imbalanced semi-supervised learning with adaptive thresholding. In *International Conference on Machine Learning*, pages 8082–8094. PMLR, 2022. 3
- [16] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 1, 7
- [17] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020. 7
- [18] Prantik Howlader, Srijan Das, Hieu Le, and Dimitris Samaras. Beyond pixels: Semi-supervised semantic segmentation with a multi-scale patch-based multi-label classifier. *arXiv preprint arXiv:2407.04036*, 2024. 7
- [19] Prantik Howlader, Srijan Das, Hieu Le, and Dimitris Samaras. Beyond pixels: Semi-supervised semantic segmentation with a multi-scale patch-based multi-label classifier. In *European Conference on Computer Vision*, pages 342–360. Springer, 2025. 3
- [20] Lukas Hoyer, David Joseph Tan, Muhammad Ferjad Naeem, Luc Van Gool, and Federico Tombari. Semivl: semi-supervised semantic segmentation with vision-language guidance. In *European Conference on Computer Vision*, pages 257–275. Springer, 2025. 3
- [21] Stanislaw Jastrzebski, Zachary Kenton, Devansh Arpit, Nicolas Ballas, Asja Fischer, Yoshua Bengio, and Amos Storkey. Three factors influencing minima in sgd. *arXiv preprint arXiv:1711.04623*, 2017. 5
- [22] Ying Jin, Jiaqi Wang, and Dahua Lin. Semi-supervised semantic segmentation via gentle teaching assistant. *Advances in Neural Information Processing Systems*, 35:2803–2816, 2022. 7
- [23] Nitish Shirish Keskar, Dheevatsa Mudigere, Jorge Nocedal, Mikhail Smelyanskiy, and Ping Tak Peter Tang. On large-batch training for deep learning: Generalization gap and sharp minima. *arXiv preprint arXiv:1609.04836*, 2016. 8
- [24] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 2012. 7
- [25] Huakai Lai, Guoxin Xiong, Huayu Mai, Xiang Liu, and Tianzhu Zhang. Rethinking noisy video-text retrieval via relation-aware alignment. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 9231–9241, 2025. 1
- [26] Xin Lai, Zhuotao Tian, Li Jiang, Shu Liu, Hengshuang Zhao, Liwei Wang, and Jiaya Jia. Semi-supervised semantic segmentation with directional context-aware consistency. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1205–1214, 2021. 3

- [27] Samuli Laine and Timo Aila. Temporal ensembling for semi-supervised learning. *arXiv preprint arXiv:1610.02242*, 2016. 3
- [28] Dong-Hyun Lee et al. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In *Workshop on challenges in representation learning, ICML*, page 896, 2013. 3
- [29] Hao Li, Zheng Xu, Gavin Taylor, Christoph Studer, and Tom Goldstein. Visualizing the loss landscape of neural nets. *Advances in neural information processing systems*, 31, 2018. 8
- [30] Wangkai Li, Rui Sun, Boaho Liao, Zhaoyang Li, and Tianzhu Zhang. Balanced learning for domain adaptive semantic segmentation. In *International conference on machine learning*, 2025. 1
- [31] Yijiang Li, Xinjiang Wang, Lihe Yang, Litong Feng, Wayne Zhang, and Ying Gao. Diverse cotraining makes strong semi-supervised segmentor. *arXiv preprint arXiv:2308.09281*, 2023. 7
- [32] Chen Liang, Wenguan Wang, Jiaxu Miao, and Yi Yang. Logic-induced diagnostic reasoning for semi-supervised semantic segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 16197–16208, 2023. 3, 7
- [33] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *Computer Vision—ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6–12, 2014, Proceedings, Part V 13*, pages 740–755. Springer, 2014. 6
- [34] Jie Ma, Chuan Wang, Yang Liu, Liang Lin, and Guanbin Li. Enhanced soft label for semi-supervised semantic segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1185–1195, 2023. 7
- [35] Huayu Mai, Rui Sun, Tianzhu Zhang, Zhiwei Xiong, and Feng Wu. Dualrel: Semi-supervised mitochondria segmentation from a prototype perspective. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 19617–19626, 2023. 3
- [36] Huayu Mai, Rui Sun, Yuan Wang, Tianzhu Zhang, and Feng Wu. Pay attention to target: Relation-aware temporal consistency for domain adaptive video semantic segmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 4162–4170, 2024. 1
- [37] Huayu Mai, Rui Sun, Tianzhu Zhang, and Feng Wu. Rankmatch: Exploring the better consistency regularization for semi-supervised semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3391–3401, 2024. 3
- [38] Huayu Mai, Rui Sun, and Feng Wu. Relaxed class-consensus consistency for semi-supervised semantic segmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2025. 3
- [39] Jaemin Na, Jung-Woo Ha, Hyung Jin Chang, Dongyoon Han, and Wonjun Hwang. Switching temporary teachers for semi-supervised semantic segmentation. *arXiv preprint arXiv:2310.18640*, 2023. 1, 3
- [40] Avital Oliver, Augustus Odena, Colin A Raffel, Ekin Dogus Cubuk, and Ian Goodfellow. Realistic evaluation of deep semi-supervised learning algorithms. *Advances in neural information processing systems*, 31, 2018. 3
- [41] Ozan Sener and Vladlen Koltun. Multi-task learning as multi-objective optimization. *Advances in neural information processing systems*, 31, 2018. 2, 4
- [42] Kihyuk Sohn, David Berthelot, Nicholas Carlini, Zizhao Zhang, Han Zhang, Colin A Raffel, Ekin Dogus Cubuk, Alexey Kurakin, and Chun-Liang Li. Fixmatch: Simplifying semi-supervised learning with consistency and confidence. *Advances in neural information processing systems*, 33:596–608, 2020. 3
- [43] Boyuan Sun, Yuqi Yang, Le Zhang, Ming-Ming Cheng, and Qibin Hou. Corrmatch: Label propagation via correlation matching for semi-supervised semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3097–3107, 2024. 3, 7
- [44] Rui Sun, Naisong Luo, Yuwen Pan, Huayu Mai, Tianzhu Zhang, Zhiwei Xiong, and Feng Wu. Appearance prompt vision transformer for connectome reconstruction. In *IJCAI*, pages 1423–1431, 2023. 1
- [45] Rui Sun, Huayu Mai, Naisong Luo, Tianzhu Zhang, Zhiwei Xiong, and Feng Wu. Structure-decoupled adaptive part alignment network for domain adaptive mitochondria segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 523–533. Springer, 2023. 1
- [46] Rui Sun, Huayu Mai, Tianzhu Zhang, and Feng Wu. Daw: Exploring the better weighting function for semi-supervised semantic segmentation. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023. 1, 7
- [47] Rui Sun, Yuan Wang, Huayu Mai, Tianzhu Zhang, and Feng Wu. Alignment before aggregation: trajectory memory retrieval network for video object segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1218–1228, 2023. 1
- [48] Rui Sun, Huayu Mai, Wangkai Li, Yujia Chen, Naisong Luo, Yuan Wang, and Tianzhu Zhang. Beyond confidence: Exploiting homogeneous pattern for semi-supervised semantic segmentation. In *International conference on machine learning*. PMLR, 2025. 3
- [49] Rui Sun, Huayu Mai, Wangkai Li, and Tianzhu Zhang. Towards unbiased learning in semi-supervised semantic segmentation. In *The Thirteenth International Conference on Learning Representations*, 2025. 3
- [50] Antti Tarvainen and Harri Valpola. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. *Advances in neural information processing systems*, 30, 2017. 3
- [51] Vikas Verma, Kenji Kawaguchi, Alex Lamb, Juho Kannala, Arno Solin, Yoshua Bengio, and David Lopez-Paz. Interpolation consistency training for semi-supervised learning. *Neural Networks*, 145:90–106, 2022. 3
- [52] Haonan Wang, Qixiang Zhang, Yi Li, and Xiaomeng Li. Allspark: Reborn labeled features from unlabeled in trans-

- former for semi-supervised semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3627–3636, 2024. 2, 5
- [53] Xiaoyang Wang, Huihui Bai, Limin Yu, Yao Zhao, and Jimin Xiao. Towards the uncharted: Density-descending feature perturbation for semi-supervised semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3303–3312, 2024. 7
- [54] Yuan Wang, Rui Sun, and Tianzhu Zhang. Rethinking the correlation in few-shot segmentation: A buoys view. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7183–7192, 2023. 1
- [55] Junde Wu, Rao Fu, Huihui Fang, Yu Zhang, Yehui Yang, Haoyi Xiong, Huiying Liu, and Yanwu Xu. Medsegdiff: Medical image segmentation with diffusion probabilistic model. *arXiv preprint arXiv:2211.00611*, 2022. 1
- [56] Linshan Wu, Leyuan Fang, Xingxin He, Min He, Jiayi Ma, and Zhun Zhong. Querying labeled for unlabeled: Cross-image semantic consistency guided semi-supervised semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023. 7
- [57] Qizhe Xie, Zihang Dai, Eduard Hovy, Thang Luong, and Quoc Le. Unsupervised data augmentation for consistency training. *Advances in neural information processing systems*, 33:6256–6268, 2020. 3
- [58] Haiming Xu, Lingqiao Liu, Qiuchen Bian, and Zhen Yang. Semi-supervised semantic segmentation with prototype-based consistency regularization. *Advances in Neural Information Processing Systems*, 35:26007–26020, 2022. 7
- [59] Yi Xu, Lei Shang, Jinxing Ye, Qi Qian, Yu-Feng Li, Baigui Sun, Hao Li, and Rong Jin. Dash: Semi-supervised learning with dynamic thresholding. In *International Conference on Machine Learning*, pages 11525–11536. PMLR, 2021. 3
- [60] Lihe Yang, Lei Qi, Litong Feng, Wayne Zhang, and Yinghuan Shi. Revisiting weak-to-strong consistency in semi-supervised semantic segmentation. *arXiv preprint arXiv:2208.09910*, 2022. 1, 3, 7, 8
- [61] Lihe Yang, Zhen Zhao, and Hengshuang Zhao. Unimatch v2: Pushing the limit of semi-supervised semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025. 7
- [62] Bowen Zhang, Yidong Wang, Wenxin Hou, Hao Wu, Jindong Wang, Manabu Okumura, and Takahiro Shinozaki. Flexmatch: Boosting semi-supervised learning with curriculum pseudo labeling. *Advances in Neural Information Processing Systems*, 34:18408–18419, 2021. 3
- [63] Zhen Zhao, Sifan Long, Jimin Pi, Jingdong Wang, and Luping Zhou. Instance-specific and model-adaptive supervision for semi-supervised semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 23705–23714, 2023. 3, 7
- [64] Zhen Zhao, Lihe Yang, Sifan Long, Jimin Pi, Luping Zhou, and Jingdong Wang. Augmentation matters: A simple-yet-effective approach to semi-supervised semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11350–11359, 2023. 3, 7
- [65] Yuanyi Zhong, Bodi Yuan, Hong Wu, Zhiqiang Yuan, Jian Peng, and Yu-Xiong Wang. Pixel contrastive-consistent semi-supervised semantic segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 7273–7282, 2021. 3, 7
- [66] Xiaojin Jerry Zhu. Semi-supervised learning literature survey. 2005. 3
- [67] Zhanxing Zhu, Jingfeng Wu, Bing Yu, Lei Wu, and Jinwen Ma. The anisotropic noise in stochastic gradient descent: Its behavior of escaping from sharp minima and regularization effects. *arXiv preprint arXiv:1803.00195*, 2018. 5
- [68] Yuliang Zou, Zizhao Zhang, Han Zhang, Chun-Liang Li, Xiao Bian, Jia-Bin Huang, and Tomas Pfister. Pseudoseg: Designing pseudo labels for semantic segmentation. *arXiv preprint arXiv:2010.09713*, 2020. 7