Geometric Gradient Rectification for Safe Open-Set Semi-Supervised Learning

Jiahe Chen[1*], Qian Shao[1*], Qiyuan Chen[1], Jiaying He[1,3], Jintai Chen[4], Jian Wu[1,5,6†], and Hongxia Xu[1,2†]

[1] College of Computer Science and Technology, Zhejiang University
[2] Liangzhu Laboratory and WeDoctor Cloud
[3] Shanghai Innovation Institute
[4] The Hong Kong University of Science and Technology (Guangzhou)
[5] State Key Laboratory of Transvascular Implantation Devices and TIDRI
[6] Zhejiang Key Laboratory of Medical Imaging Artificial Intelligence

Abstract. Open-set semi-supervised learning aims to leverage unlabeled data that may contain out-of-distribution outliers while maintaining performance on in-distribution classes. Existing methods mainly follow two paradigms: filtering suspicious samples or incorporating unlabeled objectives with soft weighting. We argue that both face a common trade-off: aggressive filtering can discard informative but hard ID samples, whereas utilization can introduce auxiliary gradients that conflict with supervised learning when pseudo labels are wrong. We therefore shift the focus from sample selection to gradient-level control. We propose *Geometric Gradient Rectification* (GGR), a plug-in framework that uses the supervised gradient as an anchor and projects conflicting auxiliary gradients onto an admissible region in gradient space. This makes the applied auxiliary update first-order non-opposing within the rectified coordinate block while preserving orthogonal components that may still carry useful representation signals. We further extend GGR with subspace-aware rectification to stabilize the anchor under noisy mini-batch gradients. Experiments on CIFAR and ImageNet benchmarks show that GGR improves representative OSSL baselines in most settings and yields gains in both closed-set generalization and open-set robustness. Code will be available at `https://github.com/JiaheChen2002/GGR`.

Keywords: Open-Set Semi-Supervised Learning · Gradient Rectification · Optimization Control

1 Introduction

Semi-supervised learning (SSL) [16,20,24,29] effectively scales training by leveraging unlabeled data. However, standard SSL assumes a closed-set scenario, which rarely holds in real-world deployments where unlabeled streams contain Out-of-Distribution (OOD) outliers. In this open-set semi-supervised learning (OSSL)

*Equal contribution.
†Corresponding authors.

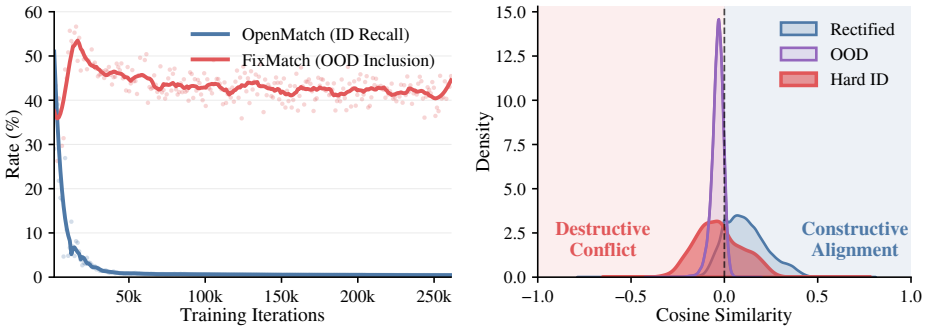

Fig. 1: Left: OSSL dilemma. A trade-off between robustness and coverage that can lead to either feature starvation or noise domination. **Right: Gradient geometry.** Density of cosine similarity between unlabeled gradients and the oracle supervised direction. **OOD:** mostly orthogonal noise; **Hard ID:** misclassified samples induce adversarial gradients; **Rectified:** GGR projects conflicting components into an anchor-aligned subspace, shifting updates toward the constructive space.

setting, naively enforcing consistency on OOD data can introduce substantial noise and degrade performance. Existing research mainly falls into two paradigms: *filtering-based* and *utilization-based* methods [6, 10, 11, 22, 25, 32]. We argue that both paradigms face a recurring trade-off between rejecting harmful samples and preserving informative but ambiguous ones.

Filtering-based methods, such as OpenMatch [22], prioritize robustness by rejecting suspicious samples. However, distinguishing between OOD noise and hard in-distribution (ID) samples is notoriously difficult in low-label regimes. To minimize false positives, these methods can become overly strict and discard valid hard ID samples, leading to feature starvation (Figure 1, left). Conversely, utilization-based approaches like FixMatch [23] and IOMatch [18] attempt to maximize data usage, often assigning outliers to a unified unknown class. This can misguide optimization: misclassifying a hard ID sample as unknown implicitly penalizes its true class probability and may weaken the learned ID manifold.

We argue that this dilemma arises because sample-level separation of OOD and hard ID data is intrinsically unreliable. Instead of hinging the pipeline on brittle categorical decisions, it is useful to examine the resulting optimization dynamics. Under softmax cross-entropy, assigning an unlabeled sample to an incorrect target can generate a gradient that opposes the oracle supervised update.

To make this concrete, we visualize the cosine similarity between unlabeled gradients and the oracle supervised direction in the right panel of Figure 1. The distributions reveal two distinct behaviors: **(i)** OOD outliers can be harmful when included in large quantities, as seen in FixMatch, by injecting variance that weakens the learning signal. Nevertheless, their gradients are often approximately orthogonal to the supervised direction, acting more like stochastic noise than systematic opposition. **(ii)** Misclassified difficult IDs are often more destructive directionally. Because they share semantics with ID classes yet are pushed toward

wrong targets, they can generate strong adversarial gradients that directly oppose the supervised objective.

These findings suggest a shift in perspective. Robustness in OSSL cannot rely solely on sample selection or category assignment. Instead, it can benefit from geometric control of optimization that suppresses harmful updates even when OOD/ID decisions are imperfect.

Motivated by this observation, we propose **G**eometric **G**radient **R**ectification (GGR), a framework that rectifies updates rather than discarding ambiguous samples or committing to potentially wrong targets. At its core, GGR performs post-hoc rectification in gradient space with respect to a supervised anchor. We further study two subspace-aware variants that summarize recent supervised descent directions in a low-dimensional basis to improve stability under noisy mini-batch gradients. In practice, this can reduce destructive interference from misguided difficult IDs, damp variance from bystander OODs, and preserve useful feature signals from ambiguous data.

Our main contributions are threefold:

- We identify the identification bottleneck in OSSL and reframe the dilemma of feature starvation versus active unlearning as a fundamental problem of gradient conflict.
- We propose GGR, a plug-in optimization framework that enforces first-order non-opposition of the applied auxiliary update within the rectified coordinate block via gradient projection, without requiring perfect OOD detection.
- Extensive experiments show that GGR improves strong OSSL baselines in most settings, particularly in low-label regimes.

2 Related Work

2.1 Standard Semi-Supervised Learning

Mainstream semi-supervised learning (SSL) is largely built upon two core mechanisms: consistency regularization and pseudo-labeling. Early works [1, 15, 19, 24] improved generalization by enforcing prediction invariance under stochastic or adversarial perturbations, laying the foundation for modern SSL. Holistic methods like MixMatch [3] and ReMixMatch [2] synthesized these principles with distribution alignment and augmentation anchoring to stabilize training. Building on these advancements, FixMatch [23] simplified the pipeline by unifying weak-to-strong consistency with confidence-based pseudo-labeling, establishing a robust baseline. Recent works further refine this framework from two perspectives: (1) Dynamic Thresholding: strategies like FlexMatch [35] and FreeMatch [27] introduce adaptive thresholds to mitigate the learning difficulty variance across classes; (2) Structure Awareness: methods such as CoMatch [17] and SimMatch [36] incorporate graph-based or contrastive objectives to enhance feature discriminability.

2.2 Open-Set Semi-Supervised Learning

OSSL addresses the challenge where the unlabeled pool contains unknown classes [8, 21, 30, 31]. Existing strategies primarily fall into two paradigms based on how they handle potential outliers: filtering and utilization.

Filtering-based. The dominant strategy prioritizes robustness by explicitly detecting and rejecting outliers. Early works like OpenMatch [22] and UASD [6] rely on confidence thresholds or entropy to filter OOD samples. To improve detection robustness in low-label regimes, recent state-of-the-art methods introduce more sophisticated selection mechanisms. ProSub [25] advances this paradigm by modeling the feature space geometry, using angles to an ID subspace to derive probabilistic outlier scores rather than relying on brittle hard thresholds; SCOMatch [28] maintains an OOD memory queue to calibrate confidence for better selection; and UAGreg [12] introduces graph regularization to smooth uncertainty estimates for soft weighting; DAC [13] leverages prediction disagreement among diverse heads to identify and filter unstable samples. Despite these advances, the paradigm still reduces the problem largely to sample selection. This creates a recurring trade-off: stricter filtering can reduce OOD noise but also suppress informative difficult ID samples, ultimately causing feature starvation.

Utilization-based. To circumvent data waste, another line of research attempts to utilize all unlabeled data by assigning outliers to a unified target. IOMatch [18] treats outliers as an additional class via a unified open-set target. Similarly, MTCF [32] and T2T [11] leverage auxiliary detection heads or self-supervised signals to harvest representation benefits from OOD samples. Although retaining difficult samples alleviates data scarcity, it also introduces a risk of misguidance. Forcing diverse difficult ID samples into a single unknown category or auxiliary targets can generate gradients that conflict with the primary supervised objective and may weaken learned ID features.

2.3 Optimization Control and Gradient Surgery

Different from sample-level selection, optimization-based methods focus on resolving task interference directly in the gradient space. Techniques like PCGrad [33] project conflicting gradients in multi-task learning to prevent negative transfer. However, standard gradient surgery assumes symmetric task importance. In OSSL, the requirement is asymmetric: the unsupervised objective should not oppose the supervised learning signal. Our GGR framework bridges this gap by keeping the supervised gradient fixed and rectifying only the auxiliary one. Instead of relying on brittle sample selection or potentially destructive joint training, we use geometric gradient rectification to control auxiliary updates directly in gradient space while preserving the supervised anchor.

3 Methodology

3.1 Problem Formulation and Gradient Conflict

We consider OSSL with a labeled set $\mathcal{D}_l = \{(x_i^l, y_i^l)\}_{i=1}^{N_l}$ over K ID classes and an unlabeled set $\mathcal{D}_u = \{x_i^u\}_{i=1}^{N_u}$ contaminated by OOD samples. A base algorithm optimizes a composite objective:

$$\min_{\theta} \; \mathcal{J}(\theta) := \mathcal{L}_s(\theta; \mathcal{D}_l) + \lambda_u \, \mathcal{L}_u(\theta; \mathcal{D}_u), \tag{1}$$

where $\mathcal{L}_s$ represents the supervised loss on labeled data and $\mathcal{L}_u$ denotes an auxiliary objective. At iteration t, let $g_s := \nabla_\theta \mathcal{L}_s(\theta_t)$ and $g_u := \nabla_\theta \mathcal{L}_u(\theta_t)$ be the stochastic gradients computed on the current labeled and unlabeled batches, respectively. We denote the Euclidean inner product as $\langle \cdot, \cdot \rangle$.

A primary failure in this formulation stems not merely from the presence of OOD samples, but from how $\mathcal{L}_u$ processes ambiguous data. Assigning a difficult ID sample to an incorrect pseudo label or an unknown category can produce an unlabeled gradient g_u that opposes the supervised descent direction g_s, potentially yielding $\langle g_s, g_u \rangle < 0$. Such destructive interference can weaken discriminative ID features by enforcing erroneous constraints.

This observation establishes a fundamental first-order non-interference requirement: the auxiliary update applied within the rectified scope should not oppose supervised progress at the current step. Formally, we seek a rectified auxiliary direction $\tilde{g}_u$ such that $\langle g_s, \tilde{g}_u \rangle \geq 0$.

3.2 Geometric Gradient Rectification (GGR)

GGR is a framework comprising three rectifiers that share the same first-order non-opposition principle: a default *vector-level rectifier* (VLR) derived in this subsection, and two subspace-aware variants—the *orthogonal subspace rectifier* (OSR) and the *conic subspace rectifier* (CSR)—presented in Section 3.3. We begin with VLR, which uses the instantaneous supervised gradient as the anchor.

Our goal is to ensure that the auxiliary update does not oppose the supervised descent direction at the current iteration. Let $g_s := \nabla_\theta \mathcal{L}_\mathrm{s}$ denote the supervised gradient and let $g_u := \nabla_\theta \mathcal{L}_\mathrm{u}$ denote the raw auxiliary gradient inside the selected surgery scope. To enforce the first-order non-interference constraint, we require the corrected auxiliary direction to satisfy

$$\langle \tilde{g}_u, g_s \rangle \geq 0. \tag{2}$$

Geometrically, this constraint defines a closed half-space induced by g_s:

$$\mathcal{H}_{\mathrm{safe}}(g_s) := \left\{ d \in \mathbb{R}^D \;\middle|\; \langle d, g_s \rangle \geq 0 \right\}. \tag{3}$$

We then obtain $\tilde{g}_u$ by the Euclidean projection of g_u onto $\mathcal{H}_{\mathrm{safe}}(g_s)$:

$$\tilde{g}_u \;=\; \arg\min_{d \in \mathbb{R}^D} \; \frac{1}{2} \|d - g_u\|_2^2 \quad \text{s.t.} \quad \langle d, g_s \rangle \geq 0. \tag{4}$$

This formulation is intentionally asymmetric: the supervised gradient g_s acts as an anchor that specifies the admissible region, while only the auxiliary gradient is modified. The same asymmetric anchoring principle carries over to the subspace variants in Section 3.3, where the anchor is replaced by a supervised subspace U rather than the instantaneous g_s; in all three GGR rectifiers the supervised direction itself is never altered. Unlike symmetric gradient surgery methods such as PCGrad [33], GGR never alters the supervised anchor.

The projection in (4) admits a closed-form solution, which we refer to as the *vector-level rectifier* (VLR):

$$\tilde{g}_u = \begin{cases} g_u - \dfrac{\langle g_u, g_s \rangle}{\|g_s\|_2^2}\, g_s, & \text{if } \langle g_u, g_s \rangle < 0 \text{ and } \|g_s\|_2 > 0, \\ g_u, & \text{otherwise.} \end{cases} \tag{5}$$

When $\langle g_u, g_s \rangle < 0$, (5) removes exactly the component of g_u that points toward $-g_s$, while preserving the orthogonal component $g_u - \frac{\langle g_u, g_s \rangle}{\|g_s\|_2^2} g_s$ that may still carry useful representation-learning signals. Algorithm 1 implements VLR together with the two subspace variants of Section 3.3, applying full rectification whenever a conflict is detected; we do not introduce any additional soft gating coefficient.

Plug-and-play update. Inside the selected surgery scope, GGR modifies a baseline OSSL method by substituting the raw auxiliary gradient g_u with its rectified counterpart $\tilde{g}_u$ before weighting and aggregation:

$$\theta_{t+1} \leftarrow \theta_t - \eta \left(g_s + \lambda_u \tilde{g}_u \right), \tag{6}$$

When rectification is applied only to a subset $\mathcal{P} \subseteq \theta$, Eq. (6) should be interpreted on the flattened coordinates of $\mathcal{P}$; parameters outside $\mathcal{P}$ follow the base update unchanged, as in Algorithm 1. Equivalently, because VLR is positively homogeneous, one may view GGR as rectifying the weighted auxiliary gradient $\lambda_u g_u$ instead; applying λ_u before or after rectification is the same. Importantly, GGR is purely a post-hoc operation in gradient space: it does not require explicit OOD detection, sample rejection, or any modification to the forward pass or the loss definitions of the underlying OSSL framework.

3.3 Rectification via Subspaces for Robust Anchoring

In practice, the minibatch supervised gradient g_s can be noisy, which may lead to unstable rectification decisions. To obtain a more robust anchor, we maintain a compact orthonormal basis $U = [u_1, \ldots, u_k] \in \mathbb{R}^{D \times k}$ (with $k \leq d$) that summarizes recent supervised directions via periodic orthonormalization. This defines a supervised subspace $\mathcal{S} = \mathrm{span}(U)$, which we use to rectify g_u at the subspace level. We consider two efficient variants:

Orthogonal subspace rectification. (OSR) This conservative option removes the entire component of g_u that lies in $\mathcal{S}$:

$$\tilde{g}_u^{\mathrm{orth}} = (I - UU^\top)\, g_u. \tag{7}$$

Since $(I - UU^\top)$ projects onto the orthogonal complement of $\mathcal{S}$, we have $\langle d, \tilde{g}_u^{\text{orth}} \rangle = 0$ for all $d \in \mathcal{S}$. In particular, if the current supervised gradient lies in $\mathcal{S}$, then $\langle g_s, \tilde{g}_u^{\text{orth}} \rangle = 0$, so the auxiliary update is exactly orthogonal to the supervised anchor and therefore non-opposing.

Conic Subspace Rectification (CSR). This less restrictive variant retains orthogonal signals while removing negative coordinates inside $\mathcal{S}$:

$$\tilde{g}_u^{\text{signed}} = g_u - U \min(U^\top g_u, 0), \tag{8}$$

where $\min(\cdot, 0)$ is applied element-wise. When U is orthonormal, (8) is exactly the Euclidean projection of g_u onto the convex cone $\mathcal{C}(U) = \{d \mid \langle d, u_i \rangle \geq 0,\ \forall i\}$. This operation removes only the directions that conflict with the stored supervised anchors, while preserving components that remain compatible. Furthermore, whenever g_s lies exactly in the conic hull of $\{u_i\}$, the rectified gradient satisfies $\langle g_s, \tilde{g}_u^{\text{signed}} \rangle \geq 0$. When g_s is only approximately represented by this cone, CSR should be interpreted as an approximate anchoring heuristic rather than a strict guarantee.

Summary. Algorithm 1 presents the complete training procedure. At each iteration, we compute g_s from labeled data and g_u from unlabeled objectives, optionally update the anchor subspace U from recent supervised gradients, and rectify g_u using either the basic VLR (5) or a subspace-level variant OSR (7) or CSR (8). The rectified auxiliary gradient is then combined with g_s for the optimizer step, while the pseudo-labeling mechanism and the objective design of the base OSSL method remain unchanged.

4 Theoretical Analysis

We begin by establishing the geometric optimality of GGR. Given a supervised gradient g_s and a raw auxiliary gradient g_u, the rectified update $\tilde{g}_u$ represents the unique minimizer for our objective. Among all possible directions that satisfy the constraint $\langle d, g_s \rangle \geq 0$, it applies the minimum possible modification to g_u measured in the Euclidean norm. This leads directly to a crucial identity regarding asymmetric alignment. The inner product $\langle g_s, \tilde{g}_u \rangle$ exactly equals $\max(0, \langle g_s, g_u \rangle)$, which is inherently greater than or equal to zero.

This alignment yields a scoped first-order non-interference statement on the selected surgery coordinates. Let $z \in \mathbb{R}^D$ denote the flattened parameter vector inside the chosen surgery scope $\mathcal{P}$, while the coordinates outside $\mathcal{P}$ are held fixed for this local analysis. We assume that the resulting supervised-loss slice is smooth with a Lipschitz constant L. Under this condition, defining the block update direction at iteration t as $v_t := g_s + \lambda_u \tilde{g}_u$, where $g_s = \nabla_z \mathcal{L}_s(z_t; \theta_{t,\setminus\mathcal{P}})$ and $\tilde{g}_u$ denotes the rectified raw auxiliary gradient on the same block, yields $\langle g_s, v_t \rangle \geq \|g_s\|_2^2$. Consequently, the partial update $z_{t+1} = z_t - \eta v_t$ yields one-step descent of the supervised objective with respect to this coordinate block under a suitable step size. If the learning rate η is bounded by $\dfrac{\langle g_s, v_t \rangle}{L\|v_t\|_2^2}$, the loss slice

Algorithm 1 Geometric Gradient Rectification as an Optimization Plug-in

Require: Labeled minibatch $\mathcal{B}_\ell$, unlabeled minibatch $\mathcal{B}_u$, model parameters θ,
Require: Supervised loss $\mathcal{L}_\mathrm{s}$, auxiliary loss $\mathcal{L}_\mathrm{u}$, unsupervised weight λ_u
Require: Surgery scope $\mathcal{P} \subseteq \theta$ (both / backbone / head), Subspace dimension $d \geq 0$
Require: Rectification mode $m \in \{\texttt{VLR}, \texttt{CSR}, \texttt{OSR}\}$
Initialize orthonormal basis $U \leftarrow \emptyset$ ▷ $U \in \mathbb{R}^{D\times k}$, $k \leq d$
for $t = 1, 2, \ldots,$ MaxIter **do**
  Compute $\mathcal{L}_\mathrm{s}(\mathcal{B}_\ell; \theta)$ and $\mathcal{L}_\mathrm{u}(\mathcal{B}_u; \theta)$ ▷ *Forward Pass*
  $g_s \leftarrow \nabla_\mathcal{P} \mathcal{L}_\mathrm{s}, \quad g_u \leftarrow \nabla_\mathcal{P} \mathcal{L}_\mathrm{u}$ ▷ *Backward Pass*
  $\mathbf{g}_s \leftarrow \mathrm{Flatten}(g_s), \quad \mathbf{g}_u \leftarrow \mathrm{Flatten}(g_u)$
  if $d > 0$ **then**
    $U \leftarrow \textsc{UpdateSubspace}(U, \mathbf{g}_s, d)$ ▷ *Subspace Maintenance*
  end if
  $\mathbf{g}_u \leftarrow \textsc{Rectify}(\mathbf{g}_u, \mathbf{g}_s, U, m)$ ▷ *Geometric Rectification*
  $g_u \leftarrow \mathrm{Unflatten}(\mathbf{g}_u)$
  $\nabla_\mathcal{P} \leftarrow g_s + \lambda_u g_u, \quad \nabla_{\theta\setminus\mathcal{P}} \leftarrow \nabla_{\theta\setminus\mathcal{P}}\big(\mathcal{L}_\mathrm{s} + \lambda_u \mathcal{L}_\mathrm{u}\big)$ ▷ *Apply Gradients*
  Update θ with the optimizer step (e.g., SGD/Adam)
end for
function $\textsc{UpdateSubspace}(U, \mathbf{g}_s, d)$
  $\mathbf{v} \leftarrow \mathbf{g}_s - U(U^\top \mathbf{g}_s)$
  if $\|\mathbf{v}\|_2 > 0$ **then**
    $U \leftarrow [U, \mathbf{v}/\|\mathbf{v}\|_2]$ ▷ *Gram-Schmidt append*
  end if
  return $\mathrm{Orthonormalize}(\, U_{[:,\,-d:]}\,)$ ▷ *Keep last d columns & apply QR*
end function
function $\textsc{Rectify}(\mathbf{g}_u, \mathbf{g}_s, U, m)$
  if $m = \texttt{VLR}$ **and** $\langle \mathbf{g}_u, \mathbf{g}_s \rangle < 0$ **and** $\|\mathbf{g}_s\|_2 > 0$ **then**
    return $\mathbf{g}_u - \frac{\langle \mathbf{g}_u, \mathbf{g}_s \rangle}{\|\mathbf{g}_s\|_2^2} \mathbf{g}_s$ ▷ *Ref: Eq.* (5)
  else if $m = \texttt{OSR}$ **and** $U \neq \emptyset$ **then**
    return $\mathbf{g}_u - U(U^\top \mathbf{g}_u)$ ▷ *Ref: Eq.* (7)
  else if $m = \texttt{CSR}$ **and** $U \neq \emptyset$ **then**
    return $\mathbf{g}_u - U \min(U^\top \mathbf{g}_u, \mathbf{0})$ ▷ *Ref: Eq.* (8)
  end if
  return $\mathbf{g}_u$ ▷ *Fallback (no conflict or empty U)*
end function

decreases according to the following relation:

$$\mathcal{L}_s(z_{t+1}; \theta_{t,\setminus\mathcal{P}}) \leq \mathcal{L}_s(z_t; \theta_{t,\setminus\mathcal{P}}) - \frac{\eta}{2}\langle g_s, v_t \rangle. \tag{9}$$

Specifically, for any non-trivial scope gradient g_s, this descent is strict for that block step. We further formalize the optimization degradation through the cumulative *applied* conflict regret over T iterations, defined on the weighted auxiliary update actually used by the optimizer, namely $g_{\mathrm{app}}^{(t)} := \lambda_u \tilde{g}_u^{(t)}$. For VLR, this applied regret is exactly zero, whereas raw pre-rectification conflicts may still occur before surgery.

To formalize the dilemma between safety and starvation, we introduce a stylized mixture model characterizing the unlabeled gradient as a combination of aligned signals from difficult IDs and conflicting components from outliers. Let $g_s^\star$ denote the non-zero population supervised gradient with a normalized direction $\hat{g}_s^\star := g_s^\star / \|g_s^\star\|_2$. We formulate the unlabeled gradient as $g_u = \beta g_s^\star + \epsilon_u$ for a scalar alignment coefficient β. For theoretical tractability, we assume the noise term ϵ_u is zero-mean and almost surely orthogonal to $g_s^\star$. This assumption is consistent with the empirical density in the right panel of Figure 1, where OOD gradients behave predominantly as orthogonal stochastic noise. Any persistent negative alignment can be absorbed into β as a negative contribution, so the single-step projection analysis still captures the destructive component that VLR removes.

Under this model, we evaluate the expected progress along the supervised direction $\hat{g}_s^\star$ for three distinct strategies. A strategy of hard rejection, representing filtering and feature starvation, yields an expected progress equal to $\|g_s^\star\|_2$. A naive utilization strategy, representing the risk of poisoning, yields an expected progress of $(1 + \lambda_u \mathbb{E}[\beta])\|g_s^\star\|_2$. Finally, our GGR approach yields an expected progress of $(1 + \lambda_u \mathbb{E}[\max(0, \beta)])\|g_s^\star\|_2$.

This formulation reveals a strict inequality when $\lambda_u > 0$ and the probability of $\beta > 0$ is strictly greater than zero. Specifically, when the expected value of β is negative, naive utilization performs strictly worse than filtering. Within this stylized model, GGR matches or exceeds the filtering baseline in expectation, while preserving positive aligned components and removing destructive ones.

5 Experiments

5.1 Experimental Setup

Benchmarks and Datasets. We evaluate our method on standard image classification SSL benchmarks under the open-set setting, following prior OSSL works [12, 13, 18, 22, 28]. Specifically, we construct OSSL tasks on CIFAR-10/100 [14] and ImageNet [7], where the labeled set contains K ID classes while the unlabeled pool is a mixture of ID and OOD samples from unknown classes. We adopt the class-splitting protocol from Kong *et al.* [13] to define seen/unseen class partitions, and evaluate multiple settings by varying (i) the number of seen/unseen classes and (ii) the amount of labeled data per seen class.

Baselines. We compare our approach against a comprehensive suite of state-of-the-art methods, categorized into two groups: (1) Standard SSL methods, including MixMatch [3], FixMatch [23], CoMatch [17], SimMatch [36], and SoftMatch [5]. Following prior work [13, 18], we exclude earlier deep SSL methods that often underperform labeled-only baselines under OOD contamination. (2) Open-Set SSL methods, including MTCF [32], OpenMatch [22], Safe-Student [10], IOMatch [18], UAGreg [12], DAC [13].

Evaluation Metrics. For closed-set evaluation, our primary metric is closed-set accuracy on the test set restricted to seen classes. This measures the core goal

Table 1: Closed-set classification accuracy on CIFAR-10/100 with varying seen/unseen class splits and labeled set sizes.

Dataset	CIFAR-10			CIFAR-100					
Class split	6 / 4			20 / 80			50 / 50		
No. of labeled samples	5	10	25	5	10	25	5	10	25
Fully Supervised	26.72±1.74	29.30±1.30	38.33±3.41	22.72±1.86	29.00±2.02	43.18±1.26	18.16±0.29	26.01±0.52	44.05±1.87
MixMatch (NIPS'19)	37.95±3.96	44.27±2.27	58.24±0.81	32.47±1.91	44.48±1.87	55.82±1.15	30.45±0.23	41.99±0.51	53.59±1.07
FixMatch (NIPS'20)	83.65±5.03	90.74±1.93	93.47±0.08	47.68±3.21	55.87±1.13	66.75±0.92	52.31±2.95	62.38±0.24	69.03±0.69
CoMatch (ICCV'21)	84.83±4.11	89.86±2.37	92.78±0.47	51.20±2.90	61.05±1.81	67.47±0.86	46.79±0.78	59.81±0.50	69.94±0.42
SimMatch (CVPR'22)	79.30±4.99	86.31±3.42	90.74±1.75	49.58±0.09	56.72±0.75	65.97±0.33	55.09±1.42	65.15±0.48	70.23±0.40
SoftMatch (ICLR'23)	80.51±5.21	83.18±1.69	86.45±1.49	53.17±2.05	58.88±1.16	66.82±0.52	55.56±0.28	63.39±0.97	69.05±0.51
MTCF (ECCV'20)	40.66±5.07	45.62±1.75	56.04±2.13	33.95±2.21	46.20±1.59	57.37±1.47	30.73±0.38	43.29±0.22	54.98±0.90
Safe-Stu (CVPR'22)	50.16±6.43	61.21±9.17	82.27±0.77	43.70±3.80	53.90±1.23	61.52±0.92	42.97±0.90	55.83±0.84	66.17±0.65
UAGreg (AAAI'24)	85.78±2.27	87.11±2.12	91.48±0.37	57.92±0.76	64.02±1.39	68.77±1.11	60.71±0.99	66.19±0.33	71.20±0.28
OpenMatch (NIPS'21)	47.26±1.20	51.21±1.82	64.73±1.78	43.72±1.24	55.78±0.02	63.78±1.01	44.55±1.56	56.19±1.53	65.65±0.52
+ **Ours**	**47.38**±1.37	**52.75**±2.65	**65.23**±1.88	**43.73**±0.70	**55.82**±0.59	**64.42**±0.90	**45.11**±0.73	**56.45**±0.62	**66.36**±0.92
IOMatch (ICCV'23)	89.87±1.26	91.16±1.47	93.61±0.29	56.48±1.86	63.00±1.51	67.32±0.89	59.49±0.99	65.33±1.11	70.43±1.29
+ **Ours**	**91.87**±1.44	**92.62**±0.93	**93.64**±0.49	**56.75**±1.45	**63.18**±1.21	**67.63**±0.30	**60.61**±0.89	**65.94**±0.47	**70.89**±0.33
DAC (TNNLS'25)	86.69±3.74	91.66±0.90	93.02±1.23	53.30±1.45	58.42±0.56	67.72±0.65	56.94±0.77	64.84±1.10	70.60±0.44
+ **Ours**	**87.63**±2.71	**92.56**±1.02	**93.25**±0.79	**53.45**±3.40	**60.10**±2.64	**68.03**±0.76	**57.19**±0.25	**65.45**±0.20	**70.68**±0.26

Table 2: Open-set classification accuracy on CIFAR-10/100 with varying seen/unseen class splits and labeled set sizes.

Dataset	CIFAR-10			CIFAR-100					
Class split	6 / 4			20 / 80			50 / 50		
No. of labeled samples	5	10	25	5	10	25	5	10	25
IOMatch (ICCV'23)	72.50±1.05	74.11±1.20	76.83±0.38	45.71±1.11	51.59±0.35	56.71±0.42	48.74±1.50	54.02±0.93	59.80±1.02
+ **Ours**	**74.13**±1.64	**75.58**±0.53	**76.86**±0.43	**45.73**±0.67	**51.68**±0.21	**56.93**±0.35	**49.93**±0.95	**54.87**±0.11	**60.18**±0.33
DAC (TNNLS'25)	72.02±3.07	76.26±0.82	77.48±0.85	49.41±0.54	55.18±0.36	64.34±0.45	54.65±0.08	62.79±1.51	68.10±0.10
+ **Ours**	**73.72**±2.07	**77.19**±0.69	**78.66**±1.61	48.94±3.81	**57.01**±2.94	**64.55**±0.80	53.77±0.31	**63.15**±0.17	**68.15**±1.18

of OSSL: leveraging an unlabeled pool contaminated with outliers to improve ID classification. For open-set evaluation, we additionally evaluate open-set classification on the test set that contains both seen and unseen classes. Following prior OSSL works [13, 18], we treat all unseen classes as a single unknown category. Because the open-set test distribution is often class-imbalanced, we adopt Balanced Accuracy (BA) [4] as the open-set metric. For each method, open-set BA is computed using the checkpoint selected by the best closed-set performance, ensuring consistent model selection across methods.

Fairness of comparisons. To ensure rigorous fairness and reproducibility, all methods are implemented and evaluated within the USB codebase [26]. We maintain strict consistency across all comparisons by using identical data splits, augmentation policies, and common hyperparameters. Regarding backbone architectures, we employ WideResNet-28-2 [34] for CIFAR benchmarks and ResNet-18 [9] for ImageNet [26]. Baseline-specific hyperparameters are initialized from original papers and tuned on the validation split, allowing us to successfully reproduce results consistent with the original reports [13, 18]. Finally, to avoid unfair comparisons due to varying convergence speeds, we report the performance of the best checkpoint selected based on the closed-set validation accuracy.

Implementation details. For our GGR framework, we adopt a unified hyperparameter configuration across all main experiments: we use VLR as the default rectification mode. To balance computational efficiency and representational stability, the gradient surgery scope ($\mathcal{P}$) is applied to the encoder backbone, meaning task-specific heads are updated via standard gradients. All method-specific hyperparameters of the base OSSL algorithms are kept identical to their baseline settings so that the effect of gradient rectification can be isolated as clearly as possible. All reported results are the mean and standard deviation over 3 independent runs with different random seeds.

5.2 Main Results

We evaluate GGR on CIFAR-10/100 and ImageNet-30 under standard open-set splits and label budgets, reporting (i) closed-set accuracy on the seen-class test set and (ii) open-set BA on the mixed test set.

CIFAR-10/100. Tables 1 and 2 summarize the closed-set and open-set performance on CIFAR-10/100. As a plug-in module, GGR improves closed-set and open-set accuracy in most settings when combined with representative filtering-based and utilization/disagreement-based methods, including OpenMatch, IOMatch, and DAC. The gains are generally larger in more challenging regimes with scarce labels and large unseen partitions. For example, relative to the corresponding baselines, GGR improves closed-set accuracy by up to +2.00 points and open-set accuracy by up to +1.83 points in the low-label settings. These results are consistent with our hypothesis that controlling destructive gradient interference can benefit OSSL, especially when pseudo-label noise and ID/OOD ambiguity are severe. For brevity, we denote the CIFAR-10 tasks with 6 labeled classes and 5, 10, or 25 labeled samples per class as CIFAR-6-30, CIFAR-6-60, and CIFAR-6-150, respectively.

ImageNet-30. Table 3 reports results on ImageNet-30 with the 20/10 split under 1% and 5% labeled ratios. Despite the increased scale and visual diversity, GGR improves strong OSSL baselines across both filtering and utilization settings on ImageNet-30. The gains are larger under the 1% labeled regime, where pseudo-label noise and ID/OOD ambiguity are more severe. These results suggest that optimization-level rectification is complementary to existing detection heuristics and can remain useful beyond small benchmarks.

Taken together, these results indicate that rectifying auxiliary gradients in a geometric manner can improve OSSL while enhancing both ID generalization and open-set recognition without relying solely on hard sample rejection.

5.3 Ablation Studies and Further Analysis

Rectification via subspaces. While VLR serves as a hyperparameter-free and efficient default rectifier at the vector level, it anchors updates dynamically to the current batch gradient g_s. In regimes with limited labels, this single-step estimate exhibits high variance, leaving the rectification susceptible to inaccurate corrections. To address this, we explore subspace-aware rectification,

Table 3: Closed-set and open-set accuracy on ImageNet-30 with the class split of 20/10.

Evaluation	Closed-Set		Open-Set	
Labeled ratio	1%	5%	1%	5%
Fully Supervised	31.87±0.97	56.08±0.88	–	–
MixMatch (NIPS'19)	51.42±1.27	76.97±0.65	–	–
FixMatch (NIPS'20)	73.92±1.74	88.82±0.88	–	–
CoMatch (ICCV'21)	83.72±0.84	91.17±0.48	–	–
SimMatch (CVPR'22)	82.78±1.59	91.88±0.48	–	–
SoftMatch (ICLR'23)	81.18±3.19	90.00±0.23	–	–
MTCF (ECCV'20)	49.33±1.50	71.62±0.46	–	–
Safe-Student (CVPR'22)	71.47±2.91	88.42±0.06	–	–
UAGreg (AAAI'24)	80.28±0.35	91.02±0.34	–	–
OpenMatch (NIPS'21)	58.55±2.10	86.72±0.91	14.75±1.18	61.51±3.63
+ Ours	**59.15**±1.00	**87.43**±0.50	**15.56**±2.38	**68.15**±1.25
IOMatch (ICCV'23)	85.18±1.80	90.32±0.52	74.88±1.56	81.03±1.01
+ Ours	**85.90**±1.07	**90.82**±0.58	**75.46**±0.60	**82.14**±1.28
DAC (TNNLS'25)	86.40±0.08	93.25±0.20	77.78±2.10	87.81±1.65
+ Ours	**87.40**±0.15	**93.33**±0.29	**78.40**±3.62	**88.74**±1.27

which summarizes recent supervised gradients into a low-dimensional basis to provide temporal smoothing. We compare two variants evaluated across subspace dimensions $d \in \{1, 5, 10, 20, 30, 40, 50\}$ on the CIFAR-100 dataset with a 50/50 seen/unseen split under two label budgets.

As Figure 2 shows, the two operators exhibit a conservativeness-utilization trade-off. OSR is the more conservative option: it projects the auxiliary gradient onto the orthogonal complement of the supervised subspace, completely eliminating any component within it. Empirically, this leads to more stable improvements in open-set balanced accuracy across a wide range of dimensions. Conversely, CSR performs a cone projection, clipping only negative coordinates while retaining positively aligned components. By preserving more auxiliary signals, it can achieve higher peak closed-set accuracy (e.g., 58.47 at $d = 10$). However, it is also more sensitive to the quality of the estimated basis; misaligned positive components can still induce destructive interference, resulting in higher variance and non-monotonic performance across different dimensions.

Selecting a moderate subspace dimension d is important for balancing conservativeness and utilization. Small dimensions (e.g., $d \in \{1, 5\}$) provide only a limited anchor and may fail to sufficiently prevent gradient conflicts under extreme label scarcity. Conversely, excessively large dimensions capture too many supervised descent directions and can suppress beneficial unsupervised signals. This over-constraining effect is particularly noticeable in easier regimes (e.g., CIFAR100-50-1250), where auxiliary guidance is already relatively reliable. Overall, a moderate d appears to capture the primary descent directions without unnecessarily discarding useful representation signals.

Table 4: Ablation study on the rectification scope. Comparison of performance when GGR is applied to the encoder backbone, task-specific heads, or both modules.

CIFAR10-6-30		
Method	Closed-Set	Open-Set
DAC	86.69±3.74	72.02±3.07
+ Ours (backbone)	87.63±2.71	73.72±2.07
+ Ours (head)	**90.38**±0.43	**75.24**±0.35
+ Ours (both)	88.47±3.02	75.06±3.55
IOMatch	89.87±1.26	72.50±1.05
+ Ours (backbone)	**91.87**±1.44	**74.13**±1.64
+ Ours (head)	90.34±0.49	72.61±0.33
+ Ours (both)	89.56±0.84	71.99±0.88

CIFAR100-50-250		
Method	Closed-Set	Open-Set
DAC	56.94±0.77	54.65±0.07
+ Ours (backbone)	57.19±0.25	53.76±0.31
+ Ours (head)	58.04±0.65	55.15±0.84
+ Ours (both)	**58.50**±1.91	**55.46**±2.43
IOMatch	59.49±0.99	48.74±1.50
+ Ours (backbone)	**60.61**±0.89	**49.93**±0.95
+ Ours (head)	60.38±0.50	49.69±0.60
+ Ours (both)	59.56±0.68	48.91±0.43

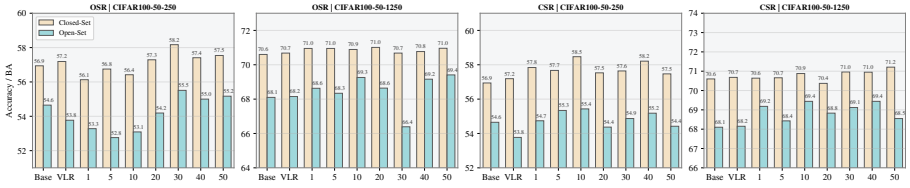

Fig. 2: Effect of subspace-aware rectification on CIFAR-100 across varying subspace dimensions d. We compare OSR and CSR projection variants under two label budgets.

In summary, VLR is an efficient default, while subspace anchoring can provide more stable behavior and temporal smoothing. The orthogonal variant offers steadier open-set improvements across dimensions, whereas the signed variant can yield slightly higher peak performance at the cost of greater sensitivity to hyperparameter tuning.

Gradient-conflict diagnostics. To empirically validate our optimization-level non-interference principle, we track gradient alignment during training. We define a *conflict event* as occurring when the auxiliary gradient ($g_u^{(t)}$) directly opposes the supervised direction ($g_s^{(t)}$), satisfying $\langle g_s^{(t)}, g_u^{(t)} \rangle < 0$. Figure 3 analyzes this hazard across three metrics: **First**, the raw conflict rate (left) tracks the moving-average frequency of conflicts before rectification. The baseline's non-trivial rate indicates that auxiliary objectives frequently generate destructive interference. Because GGR operates post-hoc in gradient space, it does not remove this underlying loss-level hazard; instead, it rectifies the update before the parameter step. **Second**, the residual conflict rate (middle) evaluates the actual applied update $g_{\text{app}}^{(t)}$. While the baseline suffers from unmitigated conflicts, GGR drives this rate close to zero. This is consistent with our first-order non-interference analysis on the rectified coordinate block: the applied auxiliary update no longer pushes against supervised progress there. **Finally**, the cumulative conflict regret (right), measured as $\sum_t \max(0, -\langle g_s^{(t)}, g_{\text{app}}^{(t)} \rangle)$, captures the accumulated optimiza-

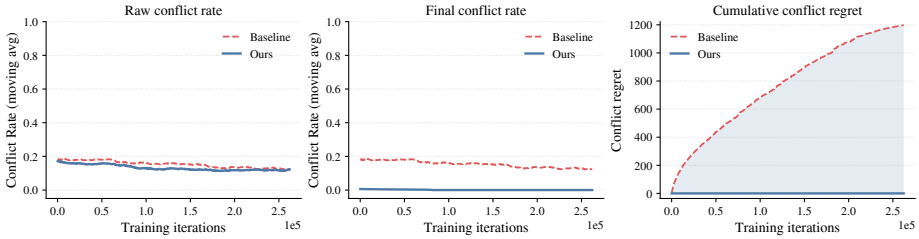

Fig. 3: Conflict diagnostics during training. Tracking the occurrence and impact of gradient conflicts between the baseline and our GGR framework.

tion damage over time. The baseline accumulates substantially more regret from counterproductive updates, whereas GGR remains nearly flat. Together, these diagnostics are consistent with the view that geometric rectification suppresses harmful update directions before they are applied.

Surgery Scope. We investigate where to apply GGR by restricting the rectification to (i) the encoder backbone, (ii) the non-backbone modules (all parameters outside the backbone; denoted as head for brevity), or (iii) both (all parameters). Table 4 reports results on CIFAR10-6-30 and CIFAR100-50-250 for two representative OSSL paradigms. For DAC, applying GGR to the head yields the largest gain on CIFAR10-6-30 in both closed-set accuracy and open-set BA, while rectifying both backbone and head performs best on the more challenging CIFAR100-50-250 setting. This suggests that, under DAC-style objectives with multiple auxiliary heads, gradient conflicts are not confined to the final decision boundary: mitigation at auxiliary branches is crucial, and full-model rectification becomes increasingly beneficial as the class mismatch grows. For IOMatch, rectifying the backbone already provides consistent improvements over the baseline across both datasets, indicating that conflict-free updates can directly benefit representation learning in utilization-based OSSL. Overall, these results support that the optimal rectification scope can depend on the base algorithm and task difficulty, and applying GGR to either the backbone or the entire model offers a robust default choice.

Efficiency and cost analysis. Table 5 reports the wall-clock time and memory cost of GGR on top of DAC. VLR adds only a modest memory increase and about 11% time overhead over the baseline. With a subspace dimension of $d = 10$, OSR keeps the overhead within $\sim$1% and CSR within $\sim$11%, while adding only a few hundred megabytes of memory. Increasing d to 50 further raises the cost without proportional accuracy gains, so VLR and the $d = 10$ subspace variants serve as the practical defaults, balancing efficiency with the accuracy benefits analyzed above.

6 Limitations

The preceding results are local and first-order. They show that the auxiliary update does not oppose the supervised descent direction at the current iterate,

Table 5: Computational cost and efficiency analysis of GGR. The relative time indicates the overhead compared to the baseline.

Method	Dim	Time(ms)	Memory(MB)	Relative Time	Accuracy(%)
DAC	–	220	9018	1.00	56.06
+ GGR (VLR)	–	244	9336	1.11	56.90
+ GGR (OSR)	10	223	9336	1.01	56.60
+ GGR (OSR)	50	293	9572	1.33	57.06
+ GGR (CSR)	10	245	9336	1.11	<u>57.26</u>
+ GGR (CSR)	50	316	9572	1.44	**57.38**

but they do not imply global optimality or monotonic decrease of the full composite objective $\mathcal{L}_s + \lambda_u \mathcal{L}_u$. In addition, when the mini-batch supervised gradient is itself highly noisy or uninformative, the vector-level anchor may become conservative or unstable; the subspace variants partially mitigate this issue via temporal smoothing. Finally, GGR is complementary to, rather than a replacement for, better open-set modeling and pseudo-labeling. Stronger OOD detection or uncertainty estimation can further improve the usefulness of the rectified auxiliary signal.

7 Conclusion

In this paper, we study OSSL from the perspective of optimization geometry. We argue that errors in ID/OOD decisions can lead either to feature starvation or to destructive gradient interference. To address this issue, we propose GGR, a plug-in framework that shifts the focus from sample-level rejection to optimization-level control. By using the supervised gradient as an anchor, GGR projects conflicting auxiliary gradients onto an admissible region and makes the applied auxiliary update first-order non-opposing within the rectified coordinate block while preserving orthogonal signals that may remain useful. Theoretical analysis and empirical results show that GGR improves strong OSSL baselines in most settings, with larger gains in challenging low-label regimes. Overall, these results suggest that geometric control of the optimization trajectory is complementary to existing detection heuristics in open-set learning.

Acknowledgements

This research was partially supported by Fundamental and Interdisciplinary Disciplines Breakthrough Plan of the Ministry of Education of China No. JYB 2025XDXM601, "Pioneer" and "Leading Goose" R&D Program of Zhejiang under Grant No. 2025C02120, and State Key Laboratory of Transvascular Implantation Devices under Grant No. SKLTID2024003.

References

1. Bachman, P., Alsharif, Q., Precup, D.: Learning with pseudo-ensembles. Advances in neural information processing systems **27** (2014)
2. Berthelot, D., Carlini, N., Cubuk, E.D., Kurakin, A., Sohn, K., Zhang, H., Raffel, C.: Remixmatch: Semi-supervised learning with distribution alignment and augmentation anchoring. arXiv preprint arXiv:1911.09785 (2019)
3. Berthelot, D., Carlini, N., Goodfellow, I., Papernot, N., Oliver, A., Raffel, C.A.: Mixmatch: A holistic approach to semi-supervised learning. Advances in neural information processing systems **32** (2019)
4. Brodersen, K.H., Ong, C.S., Stephan, K.E., Buhmann, J.M.: The balanced accuracy and its posterior distribution. In: 2010 20th international conference on pattern recognition. pp. 3121–3124. IEEE (2010)
5. Chen, H., Tao, R., Fan, Y., Wang, Y., Wang, J., Schiele, B., Xie, X., Raj, B., Savvides, M.: Softmatch: Addressing the quantity-quality trade-off in semi-supervised learning. arXiv preprint arXiv:2301.10921 (2023)
6. Chen, Y., Zhu, X., Li, W., Gong, S.: Semi-supervised learning under class distribution mismatch. In: Proceedings of the AAAI conference on artificial intelligence. vol. 34, pp. 3569–3576 (2020)
7. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. pp. 248–255. Ieee (2009)
8. Fan, Y., Kukleva, A., Dai, D., Schiele, B.: Ssb: Simple but strong baseline for boosting performance of open-set semi-supervised learning. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 16068–16078 (2023)
9. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
10. He, R., Han, Z., Lu, X., Yin, Y.: Safe-student for safe deep semi-supervised learning with unseen-class unlabeled data. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 14585–14594 (2022)
11. Huang, J., Fang, C., Chen, W., Chai, Z., Wei, X., Wei, P., Lin, L., Li, G.: Trash to treasure: Harvesting ood data with cross-modal matching for open-set semi-supervised learning. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 8310–8319 (2021)
12. Kong, H., Kim, S., Kim, H.J., Lee, S.W.: Unknown-aware graph regularization for robust semi-supervised learning from uncurated data. In: Proceedings of the AAAI conference on artificial intelligence. vol. 38, pp. 13265–13273 (2024)
13. Kong, H., Kim, S.J., Jung, G., Lee, S.W.: Diversify and conquer: Open-set disagreement for robust semi-supervised learning with outliers. IEEE Transactions on Neural Networks and Learning Systems (2025)
14. Krizhevsky, A., Hinton, G., et al.: Learning multiple layers of features from tiny images (2009), technical report, University of Toronto
15. Laine, S., Aila, T.: Temporal ensembling for semi-supervised learning. arXiv preprint arXiv:1610.02242 (2016)
16. Lee, D.H., et al.: Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In: Workshop on challenges in representation learning, ICML. vol. 3, p. 896. Atlanta (2013)
17. Li, J., Xiong, C., Hoi, S.C.: Comatch: Semi-supervised learning with contrastive graph regularization. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 9475–9484 (2021)

18. Li, Z., Qi, L., Shi, Y., Gao, Y.: Iomatch: Simplifying open-set semi-supervised learning with joint inliers and outliers utilization. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 15870–15879 (2023)
19. Miyato, T., Maeda, S.i., Koyama, M., Ishii, S.: Virtual adversarial training: a regularization method for supervised and semi-supervised learning. IEEE transactions on pattern analysis and machine intelligence **41**(8), 1979–1993 (2018)
20. Oliver, A., Odena, A., Raffel, C.A., Cubuk, E.D., Goodfellow, I.: Realistic evaluation of deep semi-supervised learning algorithms. Advances in neural information processing systems **31** (2018)
21. Ren, Y., Feng, C., Xie, X., Zhou, S.K.: Partial optimal transport based out-of-distribution detection for open-set semi-supervised learning. In: IJCAI. pp. 4851–4859 (2024)
22. Saito, K., Kim, D., Saenko, K.: Openmatch: open-set consistency regularization for semi-supervised learning with outliers. In: Proceedings of the 35th International Conference on Neural Information Processing Systems. pp. 25956–25967 (2021)
23. Sohn, K., Berthelot, D., Carlini, N., Zhang, Z., Zhang, H., Raffel, C.A., Cubuk, E.D., Kurakin, A., Li, C.L.: Fixmatch: Simplifying semi-supervised learning with consistency and confidence. Advances in neural information processing systems **33**, 596–608 (2020)
24. Tarvainen, A., Valpola, H.: Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. Advances in neural information processing systems **30** (2017)
25. Wallin, E., Svensson, L., Kahl, F., Hammarstrand, L.: Prosub: Probabilistic open-set semi-supervised learning with subspace-based out-of-distribution detection. In: European Conference on Computer Vision. pp. 129–147. Springer (2024)
26. Wang, Y., Chen, H., Fan, Y., Sun, W., Tao, R., Hou, W., Wang, R., Yang, L., Zhou, Z., Guo, L.Z., Qi, H., Wu, Z., Li, Y.F., Nakamura, S., Ye, W., Savvides, M., Raj, B., Shinozaki, T., Schiele, B., Wang, J., Xie, X., Zhang, Y.: Usb: A unified semi-supervised learning benchmark for classification. In: Thirty-sixth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (2022). https://doi.org/10.48550/ARXIV.2208.07204, https://arxiv.org/abs/2208.07204
27. Wang, Y., Chen, H., Heng, Q., Hou, W., Fan, Y., Wu, Z., Wang, J., Savvides, M., Shinozaki, T., Raj, B., et al.: Freematch: Self-adaptive thresholding for semi-supervised learning. arXiv preprint arXiv:2205.07246 (2022)
28. Wang, Z., Xiang, L., Huang, L., Mao, J., Xiao, L., Yamasaki, T.: Scomatch: Alleviating overtrusting in open-set semi-supervised learning. In: European Conference on Computer Vision. pp. 217–233. Springer (2024)
29. Yang, X., Song, Z., King, I., Xu, Z.: A survey on deep semi-supervised learning. IEEE Transactions on Knowledge and Data Engineering **35**(9), 8934–8954 (2023). https://doi.org/10.1109/TKDE.2022.3220219
30. Yang, Y., Jiang, N., Xu, Y., Zhan, D.C.: Robust semi-supervised learning by wisely leveraging open-set data. IEEE Transactions on Pattern Analysis and Machine Intelligence **46**(12), 8334–8347 (2024)
31. Yang, Y., Wei, H., Sun, Z.Q., Li, G.Y., Zhou, Y., Xiong, H., Yang, J.: S2osc: A holistic semi-supervised approach for open set classification. ACM Transactions on Knowledge Discovery from Data (TKDD) **16**(2), 1–27 (2021)
32. Yu, Q., Ikami, D., Irie, G., Aizawa, K.: Multi-task curriculum framework for open-set semi-supervised learning. In: European conference on computer vision. pp. 438–454. Springer (2020)

33. Yu, T., Kumar, S., Gupta, A., Levine, S., Hausman, K., Finn, C.: Gradient surgery for multi-task learning. Advances in neural information processing systems **33**, 5824–5836 (2020)
34. Zagoruyko, S., Komodakis, N.: Wide residual networks. arXiv preprint arXiv:1605.07146 (2016)
35. Zhang, B., Wang, Y., Hou, W., Wu, H., Wang, J., Okumura, M., Shinozaki, T.: Flexmatch: Boosting semi-supervised learning with curriculum pseudo labeling. Advances in neural information processing systems **34**, 18408–18419 (2021)
36. Zheng, M., You, S., Huang, L., Wang, F., Qian, C., Xu, C.: Simmatch: Semi-supervised learning with similarity matching. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 14471–14481 (2022)