Geometric Regularization for Long-Tailed Semi-Supervised Learning via Gaussian Feature Bridges

Hongyang He[1,2]†, Xinyuan Song[3], Yan Zhong[4], Daizong Liu[2]*, Yanbin Li[4], Yang-fan He, and Wenqiao Zhang[5]

[1] University of Warwick, Coventry, United Kingdom
[2] Wuhan University, Wuhan, China
[3] Emory University, Atlanta, GA, USA
[4] Peking University, Beijing, China
[5] Zhejiang University, Hangzhou, China

Abstract. Real-world semi-supervised learning (SSL) often encounters significant challenges with long-tailed label distributions and noisy pseudo-labels, which hinder generalization and amplify confirmation bias. In this work, we introduce a novel framework, Gaussian Bridge Consistency (GBC), to address these challenges by constructing semantic interpolation paths between unlabeled samples and high-quality class anchors. Our method maintains a dynamic Prototype Atlas that stores a diverse and evolving set of labeled and pseudo-labeled exemplars per class. For each unlabeled instance, GBC forms a class-conditional Gaussian Feature Bridge in the latent space, enabling the student model to traverse a smooth trajectory from uncertain predictions to reliable class prototypes. A bridge consistency loss is applied along this path to enforce alignment with a geometrically interpolated target distribution. Furthermore, we propose BridgeMix, a confidence-aware feature mixing strategy that interpolates both sample and anchor pairs to amplify cross-sample generalization. Extensive experiments on CIFAR10-LT and ImageNet-LT (USB benchmarks) validate the robustness and effectiveness of GBC under realistic long-tailed SSL settings, consistently improving long tail-class performance without sacrificing scalability.

Keywords: Semi-supervised learning · Long-Tailed Classification

1 Introduction

Semi-supervised learning (SSL) has emerged as a promising paradigm to reduce annotation costs by leveraging large-scale unlabeled data alongside a small labeled subset [1, 3, 29, 34, 44]. Despite substantial progress, existing SSL methods often fall short in realistic scenarios characterized by severe class imbalance, long-tailed distributions, and noisy pseudo-labels [18, 28, 41, 42]. These

† This work was completed during an internship at Maniforda.Ai.
* Corresponding author: Daizong Liu.

challenges introduce significant confirmation bias, where dominant classes monopolize learning signals while minority (tail) classes suffer from undertraining and semantic drift [10, 40, 45]. This calls for a new perspective that explicitly models the uncertainty and transition dynamics between weakly predicted unlabeled instances and reliable class representations. **Our key motivation** stems from the question:

> *Can we transform sparse unlabeled features into reliable supervision signals by constructing smooth semantic pathways?*

Inspired by the Schrödinger Bridge theory—a probabilistic framework [36] that interpolates between two distributions via optimal stochastic processes [17, 22, 27, 32]—we propose to treat labeled and pseudo-labeled class exemplars as endpoints of a semantic bridge, and encourage a student model to learn consistently along this interpolated path.

In this paper, we present a novel framework called **Gaussian Bridge Consistency (GBC)**, which introduces a simple yet efficient semantic interpolation mechanism into the SSL pipeline. Central to GBC is the notion of a *Gaussian Feature Bridge*, a latent-space path that connects an unlabeled sample to a class-specific anchor retrieved from a dynamic Prototype Atlas (PA) [4, 10, 15, 33]. At multiple intermediate points along the bridge, we inject actionable supervision by enforcing bridge-consistency loss, which aligns the student's prediction with a soft geometric target interpolated between the sample and anchor distributions. This design encourages the model to traverse from uncertain to reliable features in a controllable manner, thereby reducing semantic drift and enhancing long tail-class robustness [26].

To further improve the supervision signal and sample diversity, we introduce **BridgeMix**, a novel confidence-aware variant of MixUp. By interpolating not only features but also their corresponding anchors using pseudo-label confidence as a guidance weight, BridgeMix constructs smoothed bridge paths that exploit high-confidence samples to guide lower-confidence ones—all within the original loss formulation, with no additional training complexity [37, 40, 45, 46].

Extensive experiments on CIFAR10-LT, CIFAR100-LT, STL10-LT, and ImageNet-LT benchmarks demonstrate that GBC significantly mitigates the performance collapse typical in realistic SSL conditions. By transforming uncertain feature states into reliable class-consistent representations, our framework enhances tail-class accuracy while maintaining compatibility across diverse architectures. Our core contributions are summarized as follows:

First, we identify the fundamental challenge in long-tailed semi-supervised learning (LTSSL) as the instability of feature alignment caused by biased pseudo-labels, which often drift toward majority classes. To address this, we propose GBC, a Schrödinger Bridge-inspired framework that regularizes the learning process by constructing class-conditional interpolation paths in latent space. This mechanism pulls noisy, strongly augmented unlabeled features toward reliable anchors stored in a dynamically updated Prototype Atlas (PA), ensuring stable semantic alignment even for data-scarce tail categories.

Second, we introduce BridgeMix, a confidence-guided manifold regularization strategy designed to strengthen bridge-level supervision. By employing a dynamic mixing coefficient derived from pseudo-label certainties, BridgeMix enables highly confident samples and anchors to guide the representation learning of uncertain ones. This process effectively regularizes the model's behavior across the sample-anchor trajectory and enhances the diversity of the bridge manifold without necessitating modifications to the underlying loss design.

Third, we establish a rigorous theoretical foundation for bridge consistency by demonstrating its role as an intrinsic geometric regularizer. We formally prove that GBC enforces geometric smoothness and local Lipschitz continuity on the decision boundary, which stabilizes the model against feature-level perturbations. Furthermore, we derive a generalization bound for BridgeMix, showing that it effectively tightens the true risk by reducing hypothesis complexity and minimizing the distributional discrepancy between mixed and empirical distributions.

Finally, our results demonstrate consistent and significant gains over prior state-of-the-art methods, establishing GBC as a robust, scalable, and mathematically grounded solution for semi-supervised learning under severe long-tailed conditions.

2 Related Works

Semi-Supervised Learning. Classic SSL techniques rely heavily on pseudo-labeling [1,34] and consistency regularization [35], which aim to enhance generalization by enforcing prediction invariance across augmentations. However, most SSL methods assume class-balanced labeled/unlabeled data, which often does not hold in practice. As a result, pseudo-labeling can be error-prone and biased toward head classes, especially under low-label or long-tailed settings [28, 41]. Recent efforts address this via class-balancing strategies such as reweighting, calibration, or pseudo-label refinement [28,41], but they are typically applied at the output/logit level and lack intermediate feature-level alignment.

Long-Tailed SSL. To tackle the real-world challenges of class imbalance and distribution mismatch, Long-Tailed SSL (LTSSL) methods have been proposed. Notably, ACR [42] introduces an adaptive consistency regularizer to refine pseudo-labels based on estimated class distributions. Subsequent works improve robustness through weighting, debiasing, and expertization under the long-tailed regime, including SAW [21], Adsh [14], DePL [39], BaCon [12], CPE [23], and SimPro [10], which represents a powerful SOTA framework for Realistic LTSSL (ReaLTSSL). Meta-Expert [16] goes beyond single/expert-ensemble heuristics via dynamic expert assignment and multi-depth feature fusion to mitigate head bias and reduce pseudo-label error.

Interpolation-Based Learning and Consistency. Feature-level interpolation techniques such as MixUp [46], Manifold MixUp [37] and ProbPseudo Mixup [2] have shown promise in regularizing training signals, and are often combined with pseudo-labeling/consistency objectives [1, 34, 35]. However, these

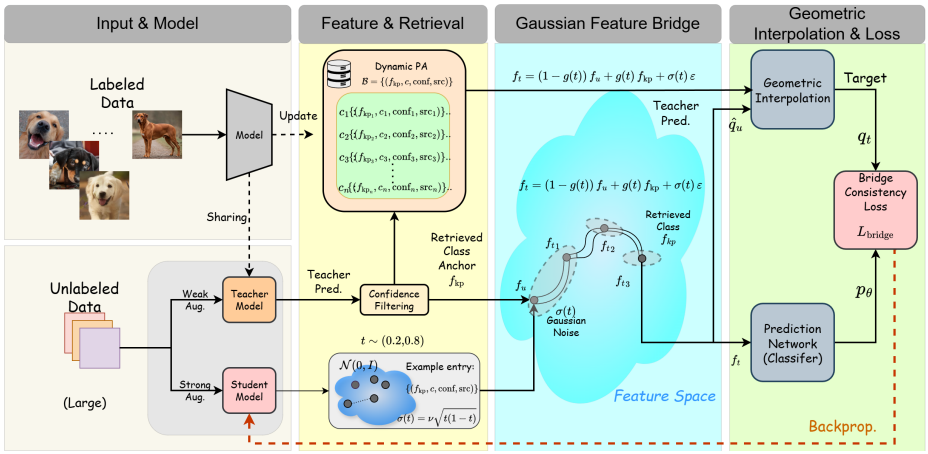

Fig. 1: Overview of our proposed framework.

approaches still lack an explicit mechanism to align the semantic structure of features over uncertain unlabeled samples. Our work bridges this gap by constructing a Gaussian path between sample–anchor pairs, enabling a student model to learn consistent predictions along a semantically meaningful trajectory in feature space.

3 Method

Framework Overview. In ReaLTSSL, the labeled and unlabeled data often follow different and unknown class distributions, violating the standard i.i.d. assumption and causing instability in pseudo-label learning. To address this, we propose GBC, a probabilistic framework that regularizes representation alignment through class-conditional Gaussian feature bridges. As illustrated in Fig. 1, GBC constructs smooth semantic paths between unlabeled samples and their class anchors from a PA, enforcing bridge-consistency along these paths to achieve stable and geometry-aware learning under long-tailed and noisy conditions.

3.1 Problem Formulation

We consider a standard SSL setup with a labeled dataset $\mathcal{D}_\ell = \{(x_i, y_i)\}_{i=1}^N$ and a larger unlabeled dataset $\mathcal{D}_u = \{x_j\}_{j=1}^M$, where $x \in \mathbb{R}^d$ and $y \in \{1, \ldots, K\}$ denote input data and class labels, respectively. The goal is to train a classifier $F_\theta : \mathbb{R}^d \mapsto \Delta^K$, parameterized by θ, that performs well on a balanced test distribution. Unlike conventional SSL, we emphasize a more realistic scenario where the pseudo-labels of $\mathcal{D}_u$ are noisy and class-imbalanced, which leads to drift toward majority classes if uncorrected. To address this, our Schrödinger Bridge-inspired GBC framework constructs class-conditional interpolation paths in the

feature space, enforcing consistency along these bridges to stabilize learning in both head and tail categories.

3.2 Proposed GBC

PA Selection and Maintenance. To provide reliable anchors for bridging, we maintain a class-indexed PA, $\mathcal{B} = \{(f_{\text{kp}}, c, \text{conf}, \text{src})\}$. Each entry consists of a feature f_{kp}, its class c, confidence, and source (labeled or pseudo-labeled). The PA is initialized with labeled exemplars and dynamically updated with high-confidence pseudo-anchors from a teacher model. We restrict the PA size to $2\% \sim 8\%$ of the dataset to ensure quality and diversity, enforcing cosine distance thresholds to avoid redundancy and applying temporal decay to gradually discard stale pseudo-anchors. If a class temporarily lacks anchors, we revert to an EMA-based prototype as fallback. This design ensures that both head and tail classes are consistently represented in the atlas.

Gaussian Feature Bridging in Latent Space. Given an unlabeled input x_u, the teacher (weak augmentation) predicts a distribution $\hat{q}_u$. If the confidence exceeds a class-dependent threshold τ_c, we query the PA for a same-class anchor f_{kp}. At the student's target layer, we construct a Gaussian Bridge in the feature space:

$$f_t = (1 - g(t))\, f_u + g(t)\, f_{\text{kp}} + \sigma(t)\, \varepsilon, \quad g(t) = t, \tag{1}$$

where $\sigma(t) = \nu\sqrt{t(1-t)}$ is the noise schedule, $\varepsilon \sim \mathcal{N}(0, I)$, and 0. truncated to $[0.2, 0.8]$. Defining $g(t) = t$ ensures the semantic path is monotonically increasing, satisfying our theoretical convergence and continuity assumptions. Specifically, we define the bridge path via an interpolation function $g(t)$. Following the theoretical requirements for smooth semantic transitions (see assume. 2), we employ $g(t) = t$ to represent the constant-speed trajectory from the sample feature to the class anchor.

Bridging and Forward Propagation. Once the Gaussian feature bridge point f_t is obtained, we incorporate it into the student's forward computation through a lightweight *bridging* step. Rather than directly replacing the original feature, we softly merge f_t into the student's hidden representation f_{stu} using a residual-style update:

$$\tilde{f} = f_{\text{stu}} + \omega(t)\, P(f_t - f_{\text{stu}}), \quad \omega(t) = 4t(1-t), \tag{2}$$

where $P(\cdot)$ is a projection module (e.g., a 1×1 linear layer or a two-layer MLP) ensuring dimensional compatibility. The weighting coefficient $\omega(t)$ acts as a gate that emphasizes mid-bridge states where semantic uncertainty is highest. The resulting bridged feature $\tilde{f}$ is then propagated through the remainder of the student network to produce a prediction $p_\theta(x_u^{\text{strong}}; \tilde{f})$.

Bridge Consistency Objective. To turn the bridged feature into effective supervision, we construct a soft target q_t via geometric interpolation:

$$q_t = \text{softmax}\big((1-t)\log \hat{q}_u + t \log \hat{q}_{\text{kp}}\big), \tag{3}$$

which ensures intermediate states remain consistent with both endpoints on the probability simplex. Given the student's prediction at the bridged state, $p_\theta(x_u^{\text{strong}}; \tilde{f})$, we define the bridge consistency loss as:

$$L_{\text{bridge}} = \mathbb{E}_{x_u,t}[w_c \cdot \omega(t) \cdot \text{KL}(q_t \,\|\, p_\theta)], \tag{4}$$

where $w_c \propto (\bar{f}/f_c)^\gamma$ provides class-aware weighting to emphasize tail categories. Here, f_c denotes the number of labeled samples in class c, $\bar{f} = \frac{1}{C}\sum_{c=1}^{C} f_c$ is the mean class frequency over all C classes, and $\gamma \geq 0$ controls the strength of reweighting (with $\gamma = 0$ corresponding to no reweighting). This objective enforces that the student remains consistent along the semantic path from unlabeled features to reliable anchors, thereby smoothing the decision boundary and mitigating pseudo-label drift.

3.3 BridgeMix

To further enhance the interpolation process on the bridge path, we introduce a novel variant of MixUp tailored for our GBC framework, termed BridgeMix. Given two unlabeled samples (x_i, x_j) with their pseudo-label confidences (o_i, o_j) and corresponding feature-anchor pairs $(f_i, f_i^{\text{anchor}})$ and $(f_j, f_j^{\text{anchor}})$, we compute a confidence-guided interpolation coefficient:

$$\lambda_i = \frac{o_i}{o_i + o_j}. \tag{5}$$

This coefficient ensures that more confident pseudo-labels guide the less certain ones, effectively reducing label noise and mitigating confirmation bias during the feature-level mixing process. We then construct the interpolated sample feature and anchor feature as:

$$f_u^{\text{mix}} = \lambda_i f_i + (1-\lambda_i) f_j, \quad f_{\text{mix}}^{\text{anchor}} = \lambda_i f_i^{\text{anchor}} + (1-\lambda_i) f_j^{\text{anchor}}. \tag{6}$$

Following the Gaussian bridge formulation, we generate the intermediate state as:

$$f_t = (1-g(t)) f_u^{\text{mix}} + g(t) f_{\text{mix}}^{\text{anchor}}, \tag{7}$$

where $g(t) = t$ controls the semantic trajectory, and the weighting coefficient $\omega(t) = 4t(1-t)$ is used to emphasize mid-bridge states during training. This strategy preserves the semantic direction from sample to anchor while allowing smoother trajectories and greater tolerance to label noise.

Crucially, when BridgeMix is enabled, the target distribution q_t^{mix} is generated using the same geometric interpolation as the standard GBC:

$$q_t^{\text{mix}} = \text{softmax}\big((1-t)\log \hat{q}_u^{\text{mix}} + t \log \hat{q}_{\text{kp}}^{\text{mix}}\big). \tag{8}$$

where $\hat{q}_u^{\text{mix}}$ and $\hat{q}_{\text{kp}}^{\text{mix}}$ are the mixed sample pseudo-labels and anchor distributions, respectively. To ensure numerical stability during the logarithmic operations in Eq. (8), particularly when dealing with one-hot hard targets or sharp distributions, we apply a small smoothing epsilon ($\epsilon = 10^{-6}$) to all target distributions. This refinement prevents undefined gradients and allows GBC to seamlessly integrate both soft pseudo-labels and ground-truth labeled anchors into the bridge manifold.

3.4 Final Learning Objective

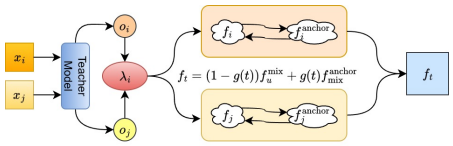

Fig. 2: Overview of BridgeMix. Confidence-guided weights λ_i facilitate smooth semantic transitions between paired unlabeled samples and their anchors.

The overall learning objective integrates bridge consistency with standard supervised and unsupervised terms:

$$L = L_{\text{sup}} + \mu L_{\text{unsup}} + \beta L_{\text{bridge}}, \quad (9)$$

where L_{sup} is the cross-entropy loss on labeled data, L_{unsup} is the standard consistency-based loss on unlabeled data, and L_{bridge} is our proposed bridge consistency objective. The hyperparameters μ and β balance the contributions of unsupervised learning and bridging, with β linearly warmed up in the early epochs to stabilize training. This unified objective creates a concise "bridge → bridging → match" loop, effectively leveraging both labeled anchors and unlabeled samples to achieve robust and realistic SSL.

4 Theoretical Analysis

We provide a theoretical analysis to explain why the proposed GBC framework and BridgeMix enhance robustness and generalization in SSL. At the feature level, GBC regularizes learning by enforcing smooth transitions from uncertain unlabeled features to reliable class anchors, yielding continuous predictions and stable decision boundaries. At the distribution level, BridgeMix reduces hypothesis complexity and distributional discrepancy through confidence-guided interpolation between samples and anchors. Together, these results offer a unified theoretical foundation for the stability and generalization of our GBC framework under noisy and long-tailed settings.

Assumption 1 (Bridge Consistency Setting)**.** *Let $f_u \in \mathbb{R}^d$ be the feature of an unlabeled sample and $f_a \in \mathbb{R}^d$ the corresponding class anchor from the PA. A Gaussian bridge path is defined in the feature space according to Eq.* (1)*:*

$$f_t = (1 - g(t))f_u + g(t)f_a + \epsilon_t, \quad g(t) = t, \quad (10)$$

where $\epsilon_t \sim \mathcal{N}(0, \sigma(t)^2 I)$, $\sigma(t) = \nu\sqrt{t(1-t)}$, and $g(t)$ is a monotonically increasing bridge function with $g(0) = 0, g(1) = 1$. The student model predicts $p_\theta(f_t)$, while the geometric target interpolation q_t follows Eq. (3)*. The bridge consistency loss is defined in Eq.* (4) *incorporating the weighting coefficient $\omega(t) = 4t(1-t)$:*

$$L_{bridge} = \mathbb{E}_t \left[\omega(t) \cdot \mathrm{KL}\big(q_t \| p_\theta(f_t)\big)\right]. \quad (11)$$

Throughout the analysis, we assume that $p_\theta(\cdot)$ and q_t are continuously differentiable in f_t and t.

Assumption 2 (Monotonicity of Semantic Trajectory)**.** *The function* $g(t) : [0,1] \to [0,1]$ *is a monotonically increasing function such that* $g(0) = 0$ *and* $g(1) = 1$*, representing the continuous semantic transition from the unlabeled sample to the class anchor.*

Remark 1. While our theoretical analysis holds for any $g(t)$ satisfying Assumption 2, in our practical implementation, we simply set $g(t) = t$. This choice not only facilitates a linear semantic interpolation in the feature space but also ensures strict theoretical compliance with the Lipschitz continuity and convergence guarantees derived below.

Lemma 1 (Pathwise Convergence). *Under Assumption 3, if* $L_{bridge} \to 0$*, then there exists a continuous trajectory* $\bar{p}(t)$ *such that*

$$\|p_\theta(f_t) - \bar{p}(t)\|_1 \to 0, \quad \forall t \in [0,1], \tag{12}$$

with boundary conditions $\bar{p}(0) = \hat{q}_u$ *and* $\bar{p}(1) = \hat{q}_a$*.*

Hence, $\{p_\theta(f_t)\}_{t\in[0,1]}$ converges to a smooth semantic path between the uncertain state $p_\theta(f_u) \sim \hat{q}_u$ and the reliable anchor state $p_\theta(f_a) \sim \hat{q}_a$.

Lemma 2 (Local Lipschitz Continuity). *Assume* ϕ_θ *satisfies* $\|\nabla_f \log p_\theta(f)\|_2 \leq L$ *for some constant* $L > 0$ *along the bridge path. Then for any perturbation* δf *with* $\|\delta f\|_2 \leq \delta$*, one has*

$$\|p_\theta(f_t + \delta f) - p_\theta(f_t)\|_1 \leq L\delta + \mathcal{O}({L_{bridge}}^{1/2}), \tag{13}$$

which shows that minimizing L_{bridge} enforces local Lipschitz continuity of p_θ around f_u and f_a.

Lemma 3 (Decision Boundary Stability). *Let* $\mathcal{M}_c = \{f : p_\theta^c(f) = \max_k p_\theta^k(f)\}$ *denote the feature manifold of class* c*. Then the expected curvature of the decision boundary* $\partial\mathcal{M}_c$ *along* $\{f_t\}$ *satisfies*

$$\mathbb{E}_t[\kappa(\partial\mathcal{M}_c)] \leq C_0 + C_1 L_{bridge}^{1/2}, \tag{14}$$

where $\kappa(\cdot)$ is the local curvature and $C_0, C_1 > 0$ are constants. Thus, decreasing L_{bridge} flattens the decision boundary and improves robustness to feature noise.

Detailed proofs are provided in the supplementary material. Lemmas 4–6 establish that bridge consistency acts as an intrinsic geometric regularizer in SSL: the Gaussian feature bridge enforces continuous prediction evolution between noisy unlabeled features and stable class anchors, induces local Lipschitz behavior, and smooths the decision boundary for better robustness under long-tailed distributions. Building on this geometric foundation, we extend the analysis to BridgeMix, which operates at the distribution level. Theorem 2 provides a generalization error bound that quantifies this regularization effect.

Theorem 1 (Generalization bound for BridgeMix). *Let $\mathcal{H}$ be a hypothesis class, and let $\ell : \mathcal{H} \times \mathcal{X} \times \mathcal{Y} \to [0,1]$ be a bounded loss. Let D be the true data distribution on $\mathcal{X} \times \mathcal{Y}$, and let D_{BM} be the BridgeMix distribution. Draw an i.i.d. sample $S = \{z_i\}_{i=1}^n$ with $z_i = (x_i, y_i) \sim D_{\mathrm{BM}}$. For $h \in \mathcal{H}$ define the risks*

$$R_D(h) = \mathbb{E}_{z\sim D}\big[\ell(h,z)\big], \quad \widehat{R}_{D_{\mathrm{BM}}}(h) = \frac{1}{n}\sum_{i=1}^{n} \ell(h, z_i). \tag{15}$$

Define the discrepancy $\Delta(D, D_{\mathrm{BM}}) = \sup_{h\in\mathcal{H}} |\mathbb{E}_D[\ell(h,z)] - \mathbb{E}_{D_{\mathrm{BM}}}[\ell(h,z)]|$ and the empirical Rademacher complexity $\widehat{\mathfrak{R}}_S(\mathcal{F})$. Then for any $\delta > 0$, with probability at least $1-\delta$ over $S \sim D_{\mathrm{BM}}^n$:

$$R_D(h) \le \widehat{R}_{D_{\mathrm{BM}}}(h) + 2\widehat{\mathfrak{R}}_S(\mathcal{F}) + 3\sqrt{\frac{\log(2/\delta)}{2n}} + \Delta(D, D_{\mathrm{BM}}). \tag{16}$$

Corollary 1. *If $\Delta(D, D_{\mathrm{BM}}) \le \varepsilon_{\mathrm{mix}}$, then with probability at least $1-\delta$ and $\forall h \in \mathcal{H}$,*

$$R_D(h) \le \widehat{R}_{D_{\mathrm{BM}}}(h) + 2\widehat{\mathfrak{R}}_S(\mathcal{F}) + 3\sqrt{\frac{\log(2/\delta)}{2n}} + \varepsilon_{\mathrm{mix}}. \tag{17}$$

Detailed proofs are provided in Section A (supplementary material). Theorem 2 and Corollary 2 together demonstrate that BridgeMix improves generalization by tightening the upper bound on the true risk. Through convex interpolation between unlabeled samples and reliable anchors, BridgeMix reduces both the effective hypothesis complexity and the distributional discrepancy, thereby lowering the generalization gap. This theoretical guarantee indicates that GBC achieves more stable learning with fewer labeled samples. Combined with the geometric regularization in Lemmas 4–6, our proposed GBC framework simultaneously enhances robustness to noise and yields stronger generalization across imbalanced data distributions.

5 Evaluation results

For the Gaussian Feature Bridging, the **target layer** for feature fusion is consistently set as the penultimate feature map before the global average pooling layer for CNNs (e.g., the output of Stage 4 in ResNet-50 see the experiment in Table. 1) and the final token representation (before the MLP head) for ViT. We sample $t \sim \mathrm{Beta}(2,2)$ to interpolate between sample and anchor features via a lightweight projector, with bridge loss weight β warmed up to 0.75, following the spirit of manifold-level regularization [37, 46]. Detailed configurations, update schedules, and additional adaptation experiments on diverse backbones are provided in the Appendix [9, 13, 47].

5.1 Experimental Evaluation

The performance of Gaussian Bridge Consistency (GBC) is evaluated on CIFAR10-LT, ImageNet-127, and ImageNet-1K. As reported in Table 3, GBC achieves the

Table 1: Detailed ablation on the injection location of the Gaussian Feature Bridge on CIFAR10-LT ($\gamma_\ell = 100$). Output dimensions are $C \times H \times W$.

Injection Layer	Output Dimension	Spatial Ratio	Top-1 Acc. (%)
ResNet `layer2`	$512 \times 8 \times 8$	1/4	87.8
ResNet `layer3`	$1024 \times 4 \times 4$	1/8	90.5
ResNet `layer4`	$\mathbf{2048 \times 2 \times 2}$	**1/16**	**92.30**
Linear `fc`	$10 \times 1 \times 1$	1/32	91.2

Table 2: Wall-clock comparison on CIFAR10-LT (consistent setting) using a single A100 GPU.

Method	Avg. Epoch (s)	Total Time	Overhead
FixMatch	42.1	3.51 h	0%
SimPro	43.0	3.58 h	+2.1%
GBC (ours)	**42.8**	**3.56 h**	**+1.7%**

best or competitive performance across most unlabeled class distributions. In challenging scenarios such as *reversed* and *head-tail*, where pseudo-label noise and distribution mismatch are more severe, GBC surpasses SimPro by 0.94% and 1.52%, respectively. Compared with adaptive thresholding methods such as DyTrim, GBC consistently delivers substantial gains, indicating that representation-level regularization is more effective than purely logit-level filtering under imbalance.

When incorporating feature interpolation, SimPro+Manifold MixUp improves over SimPro, confirming that feature-space smoothing is beneficial. Under the *uniform* distribution, SimPro+Manifold MixUp slightly outperforms GBC, suggesting that class-agnostic interpolation is sufficient when class imbalance is minimal. However, under imbalanced settings (consistent, reversed, middle, and head-tail), GBC consistently achieves stronger improvements, highlighting the advantage of class-conditional prototype-guided bridge regularization in handling skewed distributions and pseudo-label noise.

The scalability of GBC is further validated on large-scale ImageNet benchmarks in Table 4. For ImageNet-127 under the realistic test imbalance ($\gamma_t \approx 286$), GBC achieves 61.9% at 32×32 resolution, surpassing Meta-Expert by 1.6% and SimPro by 2.8%. At 64×64 resolution, GBC establishes a new SOTA of 68.6%. On the full ImageNet-1K dataset, GBC maintains its dominance with a Top-1 accuracy of 27.5% (64×64), representing a significant 2.1% improvement over Meta-Expert and a 2.5% increase over the SimPro baseline. These results collectively indicate that constructing class-conditional interpolation paths in latent space effectively mitigates confirmation bias and semantic drift. By leveraging reliable class anchors from the Prototype Atlas, GBC ensures stable representation alignment even under extreme distributional skew.

Effect of Target Layer Selection. To investigate the optimal semantic level for constructing the Gaussian Feature Bridge, we evaluate GBC by injecting bridge points at different stages of the ResNet-50 backbone. As summarized in Table 1, injecting the bridge at the *penultimate layer* (Stage 4) yields the highest Top-1 accuracy of 92.3%. In contrast, bridges constructed at earlier stages (e.g., Stage 2) lead to a performance drop (approx. 4.5%), likely because shallow features capture low-level textures rather than the high-level class semantics required for stable representation alignment. This validates our design choice of utilizing deep latent spaces for semantic bridging.

Table 3: Top-1 accuracy (%) on CIFAR10-LT ($N_1 = 500, M_1 = 4000$) with different labeled/unlabeled imbalance ratios γ_ℓ, γ_u under five unlabeled class distributions (SimPro protocol [10]). † denotes reproduced results without anchor distributions for fair comparison [42]. All results are reported in alignment with established SimPro performance trends; best in each column is **bold**. Pink highlights our GBC, and Peach indicates the baseline (SimPro). Manifold MixUp is a strong interpolation-based baselines

	consistent		uniform		reversed		middle		head-tail	
γ_ℓ	150	100	150	100	150	100	150	100	150	100
γ_u	150	100	1	1	1/150	1/100	150	100	150	100
FixMatch [34]	62.91±0.87	65.48±0.81	71.03±0.69	72.49±0.61	59.94±1.02	62.53±0.95	64.29±0.85	65.95±0.74	58.33±1.01	60.52±0.93
w/ CReST+ [41]	67.98±0.73	70.45±0.63	82.01±0.51	84.96±0.43	65.97±0.76	69.03±0.71	69.00±0.65	72.47±0.54	62.98±0.82	65.95±0.76
w/ DASO [28]	73.01±0.49	74.48±0.51	87.97±0.31	89.52±0.29	73.99±0.58	75.51±0.52	74.99±0.48	78.53±0.41	69.97±0.66	72.48±0.62
w/ ACR† [42]	76.99±0.41	81.61±0.33	91.32±0.22	92.09±0.19	81.82±0.27	85.03±0.28	77.90±0.49	73.62±0.46	78.96±0.37	79.82±0.39
w/ SimPro [10]	74.24±0.32	85.66±0.28	93.63±0.11	93.76±0.10	83.52±0.19	85.84±0.21	82.58±0.26	84.83±0.20	81.03±0.29	82.98±0.30
w/ DyTrim [5]	73.12±0.41	75.45±0.36	88.30±0.28	90.12±0.22	74.88±0.45	76.51±0.41	75.32±0.39	78.94±0.35	70.43±0.48	72.96±0.42
w/ SimPro+Manifold MixUp [38]	75.62±0.27	89.14±0.21	**94.02**±0.08	94.19±0.09	84.21±0.18	86.11±0.19	83.04±0.23	84.92±0.21	82.47±0.28	83.63±0.26
w/ GBC (ours)	**77.53**±0.25	**92.30**±0.17	93.97±0.09	**94.51**±0.07	**85.01**±0.20	**86.78**±0.23	**84.30**±0.22	**85.49**±0.19	**83.96**±0.31	**84.50**±0.27

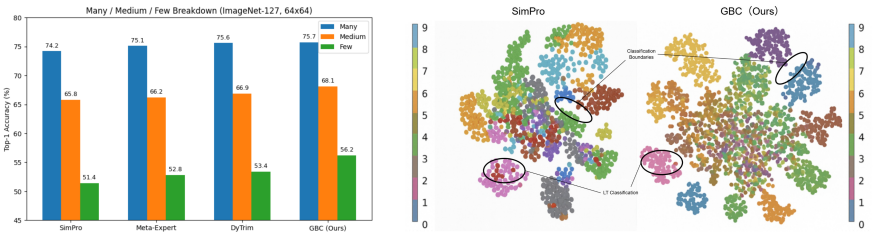

Fig. 3: Many/Medium/Few breakdown and t-SNE visualization. Left: Per-split Top-1 accuracy on ImageNet-127 ($\gamma_t \approx 286$, 64×64). Classes are grouped by labeled training samples under the long-tailed distribution (γ_ℓ): Many >100, Medium 20–100, Few <20. The split is fixed across all methods and accuracy is averaged per class within each group. The same class split is fixed across all compared methods, and group accuracy is computed as the unweighted average over classes within each group. **Right:** t-SNE visualization on CIFAR10-LT (uniform setting).

5.2 Visualization

As shown in Fig. 3, GBC yields more compact and better separated clusters than SimPro on CIFAR10-LT, especially for tail classes and near decision boundaries, where SimPro exhibits noticeable overlap and dispersion. This pattern matches the Many/Medium/Few breakdown on ImageNet-127: gains on Many classes are small, while improvements are larger on Medium and particularly Few categories. This suggests that the bridge regularization mainly enhances tail robustness by stabilizing representation alignment under imbalance.

6 Ablation Study

Optimization Stability under Distribution Shift. Figure 4 presents the mean teacher–student KL divergence across training under three unlabeled distribution regimes. In all cases, the KL value decreases monotonically, indicating progressive alignment between teacher and student predictions. This confirms that

Table 4: Top-1 accuracy (%) on **ImageNet-127** ($\gamma_\ell \approx 286$, $N_1 \approx 28$k, $M_1 \approx 250$k) and **ImageNet-1k** ($\gamma_\ell = 256$, $N_1 = 256$, $M_1 = 1024$) under different test imbalance γ_t and resolutions. † indicates results reproduced for ACR without anchor distributions. All baselines follow the SimPro protocol [10]. Pink indicates GBC while Peach indicates the ReaLTSSL baseline (SimPro).

	ImageNet-127 ($\gamma_t \approx 286$)		**ImageNet-127** ($\gamma_t = 1$)		**ImageNet-1k** ($\gamma_t = 1$)	
Method	32×32	64×64	32×32	64×64	32×32	64×64
FixMatch [34]	29.7±0.37	42.3±0.41	38.7±0.46	46.7±0.44	–	–
w/ DARP (NeurIPS'22) [19]	30.5±0.28	42.5±0.40	–	–	–	–
w/ CReST+ (CVPR'21) [41]	32.5±0.33	44.7±0.39	–	–	–	–
w/ CoSSL (CVPR'22) [11]	43.7±0.35	53.9±0.38	–	–	–	–
w/ SAW (ICML'22) [21]	58.3±0.31	65.5±0.29	54.9±0.27	62.7±0.25	18.9±0.33	24.4±0.36
w/ Adsh (ICML'22) [14]	58.8±0.32	66.1±0.34	55.1±0.28	63.3±0.26	19.2±0.31	24.7±0.34
w/ DePL (CVPR'22) [39]	59.0±0.36	66.5±0.35	55.5±0.29	63.6±0.27	19.5±0.32	24.8±0.36
w/ BaCon (AAAI'24) [12]	59.4±0.26	66.8±0.25	55.9±0.23	63.9±0.24	19.8±0.26	25.1±0.28
w/ CPE (AAAI'24) [23]	59.7±0.24	67.2±0.25	56.1±0.22	64.2±0.24	20.0±0.25	25.2±0.27
w/ ACR (CVPR'23) [42]	57.2±0.29	63.6±0.31	50.6±0.26	57.3±0.27	13.8±0.23	23.5±0.26
w/ ACR† (CVPR'23) [42]	–	–	49.5±0.27	56.1±0.29	13.2±0.24	23.4±0.25
w/ SimPro (ICML'24) [10]	59.1±0.23	67.0±0.22	55.7±0.21	63.8±0.23	19.7±0.22	25.0±0.24
w/ Meta-Expert (ICML'25) [16]	60.3±0.22	67.4±0.23	56.5±0.21	64.8±0.23	20.8±0.21	25.4±0.24
w/ DyTrim (ICLR'26) [5]	60.8±0.25	67.9±0.23	56.9±0.23	65.7±0.28	21.9±0.24	26.1±0.25
w/ GBC (ours)	**61.9**±0.36	**68.6**±0.24	**57.6**±0.22	**67.0**±0.37	**23.1**±0.23	**27.5**±0.26

the bridge-consistency mechanism continuously reduces distributional discrepancy during training rather than inducing oscillatory dynamics.

Under the consistent setting, the reversed curve starts from the largest divergence but exhibits the steepest decay, converging toward the other regimes after approximately 300 epochs. This behavior suggests that the Gaussian bridge effectively corrects early-stage bias even when initialization produces strong mismatch. The uniform regime maintains the lowest KL throughout most of training, reflecting reduced structural bias when class frequencies are balanced.

Under the uniform setting, the reversed regime again begins with the highest divergence but shows smooth and stable decay with limited fluctuation. The gap between consistent and uniform gradually narrows in later epochs, implying that bridge-level supervision dominates optimization dynamics once pseudo-label quality improves. The convergence patterns indicate that alignment is governed more by geometric regularization than by raw class-frequency statistics.

The reversed setting reveals the most informative contrast. Although the reversed regime starts with severe mismatch, its KL decreases rapidly and eventually reaches the lowest level among the three. In contrast, the consistent curve decreases more slowly and stabilizes at a higher divergence. This demonstrates that the bridge mechanism is particularly effective under strong distribution shift, where confirmation bias would otherwise accumulate. The consistent monotonic decay and reduced variance across all regimes indicate that Gaussian feature bridging enforces stable representation evolution and suppresses abrupt prediction drift.

Figure 5 evaluates the influence of τ_c across CIFAR10-LT and CIFAR100-LT. On CIFAR10-LT, accuracy increases as τ_c grows. Most regimes show monotonic

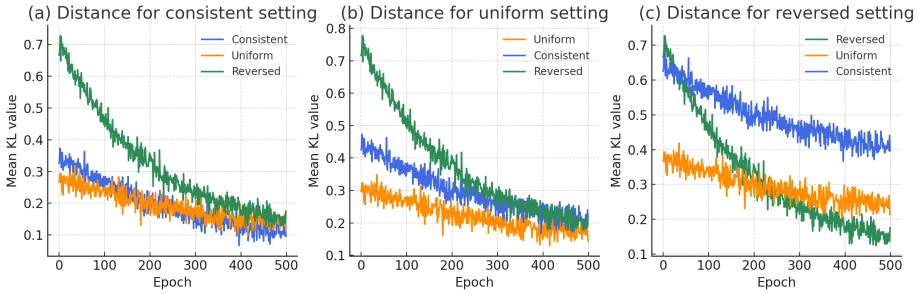

Fig. 4: Mean KL divergence on CIFAR10-LT ($N_1 = 500$, $M_1 = 4000$, SimPro protocol). $\mathrm{KL}(p_{\mathrm{teacher}} \| p_{\mathrm{student}})$ is plot on *weak* views, averaged over **3 seeds** and smoothed by EMA(0.9). Curves correspond to three unlabeled-distribution regimes: **consistent** (γ_ℓ=γ_u=150), **uniform** (γ_u=1), **reversed** (γ_u=$1/\gamma_\ell$). GBC exhibits faster and smoother KL divergence decay across all regimes, especially under **reversed**, indicating stable bridge-consistency alignment under severe pseudo-label noise.

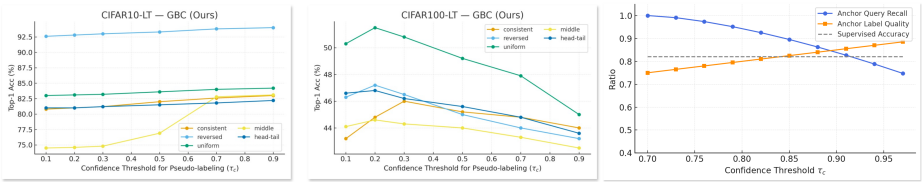

Fig. 5: Effect of the confidence threshold τ_c on performance.

gains, indicating that stricter thresholds improve pseudo-label precision without severely reducing anchor coverage. The reversed regime remains stable, suggesting robustness once unreliable anchors are filtered.

On CIFAR100-LT, performance peaks at moderate thresholds ($\tau_c \approx 0.2$–0.3) and declines thereafter. This reflects a precision–coverage trade-off. As τ_c increases, anchor label quality improves but anchor recall drops sharply. The right panel confirms this trend, with the two curves intersecting near $\tau_c \approx 0.85$, close to the supervised baseline.

The contrast between datasets indicates that higher semantic complexity and fewer samples per class make strict thresholds harmful due to reduced anchor diversity. These results support a class-adaptive τ_c strategy that balances reliability and coverage to preserve effective bridge regularization.

Tables 5 and 6 quantify the role of each component. Removing $\mathcal{L}_{\mathrm{bridge}}$ yields the largest drop (92.3% → 85.1%), confirming that path-level consistency is essential. Removing Gaussian noise lowers accuracy to 90.2%, showing that moderate stochasticity improves robustness. Hard replacement further degrades performance, indicating that residual fusion stabilizes feature updates. Eliminating class reweighting causes a sharp decline (80.5%), highlighting the necessity of tail-aware balancing. BridgeMix consistently improves results across distributions, while Prototype Atlas size shows a trade-off: insufficient capacity lim-

Table 5: Component ablation on CIFAR10-LT (γ_ℓ=100, γ_u=100).

Variant	Setting	Top-1 (%)
Full GBC	–	**92.3 ± 0.17**
w/o $\mathcal{L}_{bridge}$	Remove bridge loss	86.1 ± 0.24
w/o noise	$\sigma(t) = 0$	90.2 ± 0.21
Hard bridge	$\tilde{f} = f_t$	88.4 ± 0.19
w/o reweighting	uniform w_c	85.3 ± 0.22

Table 6: Effect of BridgeMix and prototype allocation (PA) under different unlabeled distributions (γ_ℓ=100).

Variant	Consistent	Uniform	Reversed
w/o BridgeMix	87.4 ± 0.20	88.9 ± 0.22	81.7 ± 0.26
Full GBC	**92.3 ± 0.17**	**94.5 ± 0.07**	**86.8 ± 0.23**
PA = 2%	85.7 ± 0.21	90.8 ± 0.09	84.6 ± 0.25
PA = 8%	91.1 ± 0.18	94.1 ± 0.08	86.0 ± 0.22

Table 7: Sensitivity analysis of GBC hyperparameters on CIFAR10-LT (γ_ℓ = 100, consistent setting).

Hyperparameter	Range	Top-1 Acc. (%)
ν	0.00 / 0.05 / **0.10** / 0.20	91.8 / 92.1 / **92.3** / 92.0
α	1.0 (Uniform) / **2.0** / 5.0	91.5 / **92.3** / 92.1
β	0.25 / 0.50 / **0.75** / 1.00	90.4 / 91.7 / **92.3** / 91.9

Table 8: Computational overhead comparison on ResNet-50 (32^2).

Method	Params (M)	Peak Mem.	Epoch Time (s)
FixMatch	25.6	1.00×	42.1
w/ SimPro	25.6	1.00×	43.0
w/ **GBC (Ours)**	**26.1**	**1.03×**	**42.8**

its coverage, and excessive capacity introduces redundancy. A moderate size achieves optimal stability.

Table 7 shows stable behavior across reasonable ranges. The noise level $\nu = 0.10$ achieves the best balance between exploration and semantic preservation. The Beta parameter $\alpha = 2.0$ outperforms uniform sampling by emphasizing mid-bridge states. The bridge weight β peaks at 0.75 and slightly declines beyond this value, indicating that bridge alignment should remain complementary to the main SSL objective.

Tables 8 and 13 show negligible overhead. Parameters increase marginally (25.6M $\rightarrow$ 26.1M), peak memory rises by 3%, and runtime increases by only +1.7%. These results demonstrate that GBC maintains scalability while providing substantial accuracy gains.

7 Conclusion

We introduced GBC, a Gauss Bridge driven paradigm that redefines semi-supervised learning as a process of geometric transport between uncertain features and reliable class prototypes. GBC transforms noisy pseudo-labeling into a stable representation flow, achieving robust balance under long-tailed and noisy conditions. Beyond empirical gains, GBC frames SSL within the lens of probabilistic geometry, bridging optimal transport and consistency learning. This perspective invites new research on dynamic manifold regularization, distributional interpolation for multimodal and temporal data, and continuous-time consistency fields, paving the way toward a unified geometry aware theory of semi-supervised representation learning.

References

1. 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)
2. Cai, Z., Ravichandran, A., Favaro, P., Wang, M., Modolo, D., Bhotika, R., Tu, Z., Soatto, S.: Semi-supervised vision transformers at scale. Advances in Neural Information Processing Systems **35**, 25697–25710 (2022)
3. Chapelle, O., Scholkopf, B., Zien, A.: Semi-supervised learning (chapelle, o. et al., eds.; 2006)[book reviews]. IEEE Transactions on Neural Networks **20**(3), 542–542 (2009)
4. Chen, T., Kornblith, S., Norouzi, M., Hinton, G.: A simple framework for contrastive learning of visual representations. In: International conference on machine learning. pp. 1597–1607. PmLR (2020)
5. Cheng, Y., Zhang, J., Gao, X., Xing, W., Zhu, Z.: Learning dynamics of logits debiasing for long-tailed semi-supervised learning. In: The Fourteenth International Conference on Learning Representations (2026), https://openreview.net/forum?id=e15SYMcsTs
6. Coates, A., Ng, A., Lee, H.: An analysis of single-layer networks in unsupervised feature learning. In: Proceedings of the fourteenth international conference on artificial intelligence and statistics. pp. 215–223. JMLR Workshop and Conference Proceedings (2011)
7. Cubuk, E.D., Zoph, B., Shlens, J., Le, Q.V.: Randaugment: Practical automated data augmentation with a reduced search space. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops. pp. 702–703 (2020)
8. DeVries, T., Taylor, G.W.: Improved regularization of convolutional neural networks with cutout. arXiv preprint arXiv:1708.04552 (2017)
9. DosovitskiyA, B., Kolesnikov, A., et al.: Animageis worth16× 16words: Transformersforimagerecognitionat scale (2010)
10. Du, C., Han, Y., Huang, G.: Simpro: A simple probabilistic framework towards realistic long-tailed semi-supervised learning. arXiv preprint arXiv:2402.13505 (2024)
11. Fan, Y., Dai, D., Kukleva, A., Schiele, B.: Cossl: Co-learning of representation and classifier for imbalanced semi-supervised learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 14574–14584 (2022)
12. Feng, Q., Xie, L., Fang, S., Lin, T.: Bacon: Boosting imbalanced semi-supervised learning via balanced feature-level contrastive learning. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 38, pp. 11970–11978 (2024)
13. Gu, A., Dao, T.: Mamba: Linear-time sequence modeling with selective state spaces. In: First conference on language modeling (2024)
14. Guo, L.Z., Li, Y.F.: Class-imbalanced semi-supervised learning with adaptive thresholding. In: International conference on machine learning. pp. 8082–8094. PMLR (2022)
15. He, K., Fan, H., Wu, Y., Xie, S., Girshick, R.: Momentum contrast for unsupervised visual representation learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 9729–9738 (2020)
16. Hou, Y., Jia, Y.: A square peg in a square hole: Meta-expert for long-tailed semi-supervised learning. arXiv preprint arXiv:2505.16341 (2025)
17. Jamison, B.: The markov processes of schrödinger. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete **32**(4), 323–331 (1975)
18. Kang, B., Xie, S., Rohrbach, M., Yan, Z., Gordo, A., Feng, J., Kalantidis, Y.: Decoupling representation and classifier for long-tailed recognition. arXiv preprint arXiv:1910.09217 (2019)

19. Kim, J., Hur, Y., Park, S., Yang, E., Hwang, S.J., Shin, J.: Distribution aligning refinery of pseudo-label for imbalanced semi-supervised learning. Advances in neural information processing systems **33**, 14567–14579 (2020)
20. Krizhevsky, A., Hinton, G., et al.: Learning multiple layers of features from tiny images (2009)
21. Lai, Z., Wang, C., Gunawan, H., Cheung, S.C.S., Chuah, C.N.: Smoothed adaptive weighting for imbalanced semi-supervised learning: Improve reliability against unknown distribution data. In: International Conference on Machine Learning. pp. 11828–11843. PMLR (2022)
22. Léonard, C.: A survey of the schr\" odinger problem and some of its connections with optimal transport. arXiv preprint arXiv:1308.0215 (2013)
23. Ma, C., Elezi, I., Deng, J., Dong, W., Xu, C.: Three heads are better than one: Complementary experts for long-tailed semi-supervised learning. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 38, pp. 14229–14237 (2024)
24. Massart, P.: About the constants in talagrand's concentration inequalities for empirical processes. The Annals of Probability **28**(2), 863–884 (2000)
25. McDiarmid, C., et al.: On the method of bounded differences. Surveys in combinatorics **141**(1), 148–188 (1989)
26. 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)
27. Nutz, M., Wiesel, J.: Entropic optimal transport: Convergence of potentials. Probability Theory and Related Fields **184**(1), 401–424 (2022)
28. Oh, Y., Kim, D.J., Kweon, I.S.: Daso: Distribution-aware semantics-oriented pseudo-label for imbalanced semi-supervised learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 9786–9796 (2022)
29. 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)
30. Pinsker, M.S.: Information and information stability of random variables and processes. Holden-Day (1964)
31. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., et al.: Imagenet large scale visual recognition challenge. International journal of computer vision **115**(3), 211–252 (2015)
32. Schrödinger, E.: Über die umkehrung der naturgesetze. Verlag der Akademie der Wissenschaften in Kommission bei Walter De Gruyter u ... (1931)
33. Snell, J., Swersky, K., Zemel, R.: Prototypical networks for few-shot learning. Advances in neural information processing systems **30** (2017)
34. 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)
35. 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)
36. Theodoropoulos, P., Komianos, N., Pacelli, V., Liu, G.H., Theodorou, E.A.: Feedback schr\" odinger bridge matching. arXiv preprint arXiv:2410.14055 (2024)
37. Verma, V., Lamb, A., Beckham, C., Najafi, A., Mitliagkas, I., Lopez-Paz, D., Bengio, Y.: Manifold mixup: Better representations by interpolating hidden states. In: International conference on machine learning. pp. 6438–6447. PMLR (2019)

38. Verma, V., Lamb, A., Beckham, C., Najafi, A., Mitliagkas, I., Lopez-Paz, D., Bengio, Y.: Manifold mixup: Better representations by interpolating hidden states. In: International conference on machine learning. pp. 6438–6447. PMLR (2019)
39. Wang, X., Wu, Z., Lian, L., Yu, S.X.: Debiased learning from naturally imbalanced pseudo-labels. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 14647–14657 (2022)
40. 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)
41. Wei, C., Sohn, K., Mellina, C., Yuille, A., Yang, F.: Crest: A class-rebalancing self-training framework for imbalanced semi-supervised learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 10857–10866 (2021)
42. Wei, T., Gan, K.: Towards realistic long-tailed semi-supervised learning: Consistency is all you need. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 3469–3478 (2023)
43. Wellner, J., et al.: Weak convergence and empirical processes: with applications to statistics. Springer Science & Business Media (2013)
44. Xie, Q., Dai, Z., Hovy, E., Luong, T., Le, Q.: Unsupervised data augmentation for consistency training. Advances in neural information processing systems **33**, 6256–6268 (2020)
45. 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)
46. Zhang, H., Cisse, M., Dauphin, Y.N., Lopez-Paz, D.: mixup: Beyond empirical risk minimization. arXiv preprint arXiv:1710.09412 (2017)
47. Zhu, L., Liao, B., Zhang, Q., Wang, X., Liu, W., Wang, X.: Vision mamba: Efficient visual representation learning with bidirectional state space model. arXiv preprint arXiv:2401.09417 (2024)