HUB FOR OUTLIERS, SPOKES FOR INLIERS: UNIFORM LATENT SPACE CONSTRUCTION FOR DUAL-MISMATCHED SEMI-SUPERVISED LEARNING

Anonymous authors
Paper under double-blind review

ABSTRACT

Semi-supervised learning typically assumes that labeled and unlabeled data share an identical class distribution and label space. However, this setting is often violated: unlabeled data may be imbalanced and contain unknown class samples, causing mismatches in both class distribution and label space. Such dual mismatch leads to majority classes dominating the latent space and unknown class samples being overconfidently misclassified, degrading feature discriminability and pseudo-label quality. To address this, we propose a hub-spoke latent geometry, where known classes are uniformly distributed around a central hub and each class forms compact clusters around its prototype, while the hub provides an anchor for a low-evidence region specifically designed for high-uncertainty unknown class samples. Integrated with an evidence-based classifier, this geometry ultimately enhances feature discriminability and uncertainty separation by mitigating majority-class domination through structured feature organization and guiding high-uncertainty unknown class samples toward the hub. Extensive experiments show that our method outperforms state-of-the-art methods, with a maximum improvement of **3.25%** across various settings. *Code is available in the supplementary material.*

1 INTRODUCTION

Semi-supervised learning (SSL) effectively leverages large amounts of unlabeled data to improve model performance when labeled data is scarce. Representative SSL techniques, such as pseudo-labeling (Chen et al., 2018; Cascante-Bonilla et al., 2021), consistency regularization (Miyato et al., 2019; Ke et al., 2019), or a combination of both (Cheng et al., 2025; Du et al., 2024) have demonstrated considerable success by utilizing the information within unlabeled data. However, these methods typically rely on the assumption that the labeled and unlabeled data share both a consistent class distribution (i.e., both are balanced) and an identical label space (i.e., no unknown class samples exist in the unlabeled data). **In practice, however, the unlabeled data often exhibits mismatches in both class distribution and label space** (i.e., the unlabeled data is imbalanced and contains unknown class samples), as illustrated in the rightmost panel of Fig. 1. Existing SSL methods lack an effective strategy to handle these realistic and challenging settings and therefore suffer from significant performance degradation.

While existing long-tailed semi-supervised learning (LTSSL) methods tackle class distribution mismatch through approaches like auxiliary classifiers (Hou et al., 2025; Ma et al., 2024; Hou & Jia, 2025; Hou et al., 2026) and distribution alignment (Kim et al., 2020; Aimar et al., 2024), they often overlook the underlying feature distribution. We empirically observe that imbalanced class distribution causes majority classes (e.g., class 5) to dominate the feature representation space. This dominance compresses the representation space of minority classes (e.g., class 3 and class 7), leading to overlapping class boundaries (Fig. 2b left). This degradation limits the model's discriminative capability and leads to low-quality pseudo-labels. Meanwhile, to handle label space mismatch, existing methods rely heavily on softmax-based confidence scores to filter out unknown class samples to prevent negative interference (Saito et al., 2021; Yu et al., 2020) or leverage them positively (Li et al., 2023; Bae et al., 2025). However, such scores lead to overconfident misclassifications of unknown class samples as known class ones, thereby degrading pseudo-label accuracy (Fig. 2c).

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

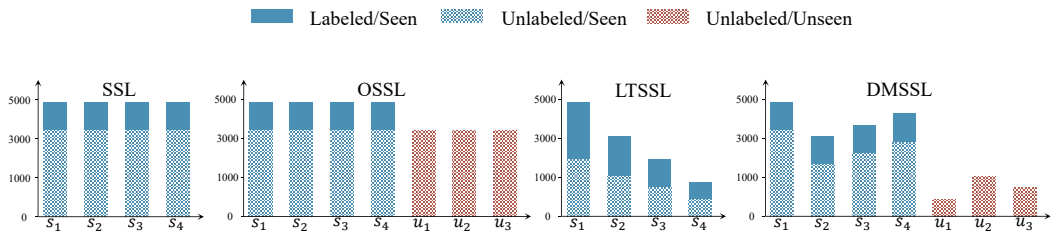


Figure 1: A comparison of the class distributions in semi-supervised learning (SSL), open-set semi-supervised learning (OSSL), long-tailed semi-supervised learning (LTSSL), and the dual-mismatched semi-supervised learning (DMSSL) scenarios. DMSSL considers a more realistic scenario where the labeled data is balanced, while the unlabeled data is imbalanced and contains samples from both known and unknown classes.

The aforementioned challenges posed by dual mismatch motivate us to explore a solution that jointly improves known class discriminability and handles unknown classes. To this end, we explicitly organize known and unknown classes in a structured latent space by constructing a uniform hub-spoke geometry (Fig. 2a), where uniformly distributed spokes represent prototypes for known classes and a central hub provides the anchor for a low-evidence region designed for high-uncertainty unknown classes. By encouraging known classes to spread uniformly around the hub and samples to cluster around their corresponding prototypes (Fig. 2b right), the geometry promotes inter-class separation and intra-class compactness, mitigating representation bias caused by class imbalance. Building upon this geometry, we further introduce an evidence-based classifier to quantify predictive uncertainty. Specifically, unlabeled samples with high uncertainty which exhibit low similarity to known classes are guided toward the central hub. Our theoretical analysis shows that these high-uncertainty candidate samples, combined with hub-centered clustering, induce a low-evidence region around the hub, providing theoretical support for this uncertainty-guided modeling (see Sec. 5 for more details). Consequently, this modeling improves the separation between known and unknown classes in the feature space and their uncertainty distributions, as illustrated in Fig. 2d.

Our main contributions can be summarized as follows:

- We identify and tackle the realistic dual-mismatched semi-supervised learning scenario, where the unlabeled data simultaneously suffers from imbalanced class distribution and the presence of unknown classes.
- We propose the hub-spoke geometry which enhances feature discriminability under class imbalance and provides a low-evidence region for unknown class samples, with a theoretical characterization showing how high-uncertainty candidates and hub-centered clustering induce this region.
- Extensive experiments on CIFAR-10, CIFAR-100, SVHN, Food-101, and ImageNet-100 datasets under various dual-mismatched conditions demonstrate that our proposed method achieves new state-of-the-art (SOTA) performance gains by up to **3.25%**.

2 Related Work

2.1 Long-Tailed Semi-Supervised Learning

Existing long-tailed semi-supervised learning (LTSSL) methods primarily aim to mitigate model bias toward majority classes during the pseudo-labeling process, and these methods can be broadly categorized into two groups. The first focuses on pseudo-label refinement and distribution alignment. For instance, DARP (Kim et al., 2020) refines biased pseudo-labels via convex optimization, while DASO (Oh et al., 2022) leverages the complementary properties of linear and semantic classifiers. Moreover, ACR (Wei & Gan, 2023) dynamically calibrates pseudo-labels based on the distance between the estimated unlabeled data distribution and anchor distributions, and CReST (Wei et al., 2021) employs a progressive distribution alignment strategy to adjust its rebalancing intensity. The second involves structural or strategic adjustments to enforce balance. ABC (Lee et al., 2021)

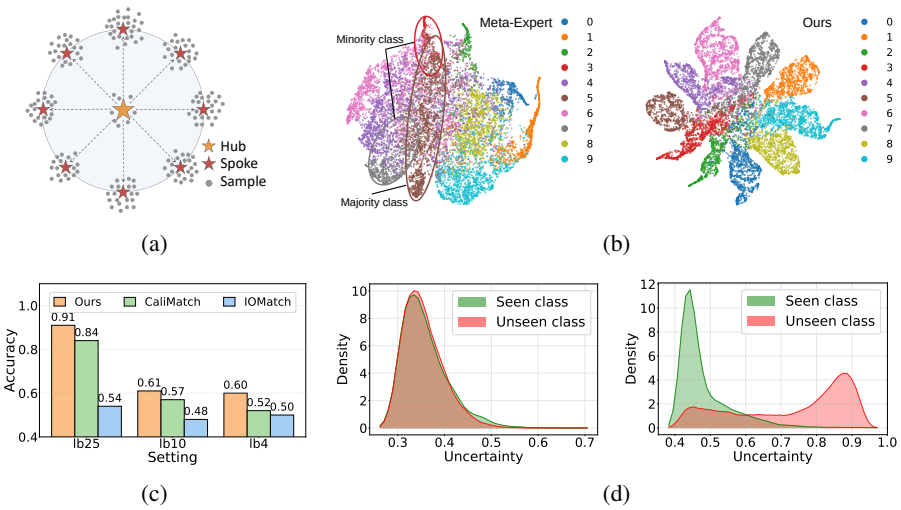


Figure 2: (a) presents the proposed hub-spoke geometry; (b) presents the t-SNE visualizations of CIFAR-10 (Krizhevsky et al., 2009) test set for Meta-Expert (Hou & Jia, 2025) and our proposed method, respectively; (c) shows a comparison of the pseudo-label accuracy on CIFAR-10 among IOMatch (Li et al., 2023), CaliMatch (Bae et al., 2025), and our method across three settings, with 4 (lb4), 10 (lb10), and 25 (lb25) labeled samples per class, respectively; (d) presents the uncertainty distributions on CIFAR-10 for the cases without guiding and with guiding high-uncertainty unknown class samples toward the central hub in the hub-spoke geometry, respectively.

introduces an auxiliary classifier to alleviate model bias toward majority classes, and CDMAD (Lee & Kim, 2024) designs a debiasing mechanism by using pattern-free images. More recently, methods such as CPE (Ma et al., 2024) train multiple experts, each specialized for a specific unlabeled data distribution, to generate higher-quality pseudo-labels. Building on this, Meta-Expert (Hou & Jia, 2025) proposes dynamic expert assignment to better incorporate individual expert strengths. **Despite their effectiveness in mitigating bias under class imbalance, these methods exhibit limited performance in realistic scenarios where unknown class samples exist in the unlabeled data.**

2.2 Open-Set Semi-Supervised Learning

A parallel research direction, known as open-set semi-supervised learning (OSSL), explicitly tackles scenarios where unlabeled data contains unknown class samples. OpenMatch (Saito et al., 2021) leverages one-vs-all classifiers to isolate unknown class samples from known class ones. IOMatch (Li et al., 2023) unifies unknown class samples into a new class via a multi-binary classifier, facilitating joint training of closed-set and open-set classifiers. Se-FOSS (Wallin et al., 2024) utilizes self-supervised learning techniques and energy-based scores to learn from all known and unknown class samples, while SCOMatch (Wang et al., 2024) explicitly selects reliable unknown class samples to form an additional class and refines the decision boundary through simultaneous closed-set and open-set self-training. Most recently, CaliMatch (Bae et al., 2025) addresses the critical issue of overconfidence by simultaneously calibrating the classifier and unknown class sample detector. **However, these methods assume a balanced class distribution and consequently perform poorly when unlabeled data is imbalanced.**

Further discussion and analysis of related work is provided in Appendix A.

3 Preliminaries

Problem formulation. This paper studies semi-supervised learning involving mismatches in both class distribution and label space between labeled and unlabeled data, which we term dual-mismatched semi-supervised learning (DMSSL). In DMSSL, we consider both a labeled dataset $\mathcal{D}_l = \{(x_i^l, y_i^l)\}_{i=1}^N$ and an unlabeled dataset $\mathcal{D}_u = \{(x_j^u, y_j^u)\}_{j=1}^M$. Here, x_i^l denotes the i-th labeled sample with its ground-truth label $y_i^l \in \{1, \ldots, C\}$, while x_j^u represents the j-th unlabeled sample with its inaccessible ground-truth label $y_j^u \in \{1, \ldots, C, \ldots, C_{IO}\}$, where $C_{IO} = C + C_O$ is the

total number of classes, including C known and C_O unknown classes. $\mathcal{D}_u$ can be decomposed as $\mathcal{D}_u = \mathcal{D}_I \cup \mathcal{D}_O$, where $\mathcal{D}_I$ and $\mathcal{D}_O$ contain known and unknown class samples, respectively. Furthermore, the labeled data $\mathcal{D}_l$ follow a balanced class distribution, whereas the unlabeled data $\mathcal{D}_u$ follow an imbalanced class distribution. Let N_c denote the number of labeled samples in class c, defining the labeled data imbalance ratio as $\gamma_l = \frac{\max_c N_c}{\min_c N_c} = 1$. Theoretically, we can assume that M_c represents the number of samples for class c in $\mathcal{D}_u$, and define its imbalance ratio as $\gamma_u = \frac{\max_c M_c}{\min_c M_c} > 1$.

Base semi-supervised learning framework. Common semi-supervised learning methods can be summarized as follows. Given a batch of B_l labeled samples $\mathcal{X}^l = \{(x_i^l, y_i^l)\}_{i=1}^{B_l}$, a weak augmentation $\Omega_w(\cdot)$ is applied to each sample before being fed to an encoder $\Phi(\cdot)$ that extracts feature representations $z = \Phi(\Omega_w(x))$, which are mapped to logits by a fully-connected layer $f(\cdot)$ to produce class probability distributions $p = f(z)$. The model's predictions on these labeled samples are supervised by the standard cross-entropy loss $\mathcal{H}(\cdot)$ against their ground-truth labels:

$$\mathcal{L}_{sup} = \frac{1}{B_l} \sum_{i=1}^{B_l} \mathcal{H}(p_i, y_i). \tag{1}$$

Given a batch of B_u samples $\mathcal{X}^u = \{x_j^u\}_{j=1}^{B_u}$, each sample is transformed by both a weak augmentation $\Omega_w(\cdot)$ and a strong augmentation $\Omega_s(\cdot)$, and the resulting two views are then processed to yield class probability distributions p^w and p^s, respectively. The unsupervised loss is then defined as the cross-entropy loss $\mathcal{H}(\cdot)$ between the prediction from the strong augmentation view and the pseudo-label generated from the weak augmentation view:

$$\mathcal{L}_{unsup} = \frac{1}{B_u} \sum_{j=1}^{B_u} \mathbb{I}(\max(p_j^w) > \tau) \mathcal{H}(p_j^s, \hat{p}_j^w), \tag{2}$$

where $\hat{p}_j^w$ is the pseudo-label for the j-th unlabeled sample in the current batch, τ is the confidence threshold, and $\mathbb{I}(\cdot)$ denotes a binary sample mask that selects pseudo-labels with confidence larger than τ. This ensures that only high-confidence pseudo-labels contribute to the training process.

4 Proposed Method

We address the dual-mismatched semi-supervised learning scenario where unlabeled data suffers from both class imbalance and the presence of unknown class samples. Our method constructs a discriminative feature representation space via a hub-spoke geometry. Meanwhile, it identifies high-uncertainty unknown class samples by leveraging evidence-based uncertainty and models them by guiding their features toward the central hub.

4.1 Hub-spoke Geometry Construction

As illustrated in Fig. 2b left, we observe that imbalanced class distribution leads to majority classes (e.g., class 5) dominating the feature representation space and compressing the feature representation space of minority classes (e.g., class 3 and class 7), thereby causing overlapping class boundaries. This limits feature discriminability and leads to low-quality pseudo-labels. To counteract this, we propose both a dual regularizer that structures the feature representation space into a hub-spoke geometry from global and local perspectives, and a class-adaptive re-weighting strategy to dynamically focus on minority classes.

Global structure regularization. To prevent majority classes from dominating the feature representation space and amplifying model bias, we enforce a globally uniform feature distribution, where all known classes are uniformly distributed around a central hub. We achieve this through two losses designed to explicitly shape the global geometry.

First, an inter-class separation loss $\mathcal{L}_{ig}$ is introduced to directly maximize the angular distances between labeled class prototypes. Formally, $\mathcal{L}_{ig}$ is defined as:

$$\mathcal{L}_{ig} = \frac{1}{|C|(|C|-1)} \sum_{i=1}^{|C|} \sum_{\substack{j=1, \\ j \neq i}}^{|C|} \left(\frac{\mu_i^l \cdot \mu_j^l}{\|\mu_i^l\|_2 \|\mu_j^l\|_2} + 1 \right)^2, \tag{3}$$

where μ_i^l denotes the labeled class prototype for the i-th class, $\cdot$ denotes the dot product, and $\|\cdot\|_2$ denotes the L_2 norm.

Second, a prototype geometry loss $\mathcal{L}_{og}$ regulates the spatial arrangement of all labeled class prototypes relative to the global labeled class prototype $\bar{\mu}^l = \frac{1}{|C|}\sum_{i=1}^{|C|}\mu_i^l$. Formally, $\mathcal{L}_{og}$ is defined as:

$$\mathcal{L}_{og} = -\log\left(\frac{1}{|C|}\sum_{i=1}^{|C|} d_i\right) + \frac{1}{|C|}\sum_{i=1}^{|C|}(d_i - \bar{d})^2, \tag{4}$$

where $d_i = \|\mu_i^l - \bar{\mu}^l\|_2$ is the distance from the i-th labeled class prototype to the global labeled class prototype, and $\bar{d} = \frac{1}{|C|}\sum_{i=1}^{|C|} d_i$ is the average distance.

Accordingly, the global structure loss $\mathcal{L}_g$ integrates these two loss items and is formulated as:

$$\mathcal{L}_g = \mathcal{L}_{ig} + \mathcal{L}_{og}. \tag{5}$$

Local structure regularization. To enforce local compactness and uniformity, we ensure that samples cluster tightly and uniformly around their respective overall class prototypes (i.e., calculated on labeled and pseudo-labeled data). This is achieved through two losses.

The intra-local loss $\mathcal{L}_{il}$ maximizes the similarity between samples and their corresponding overall class prototypes while minimizing similarity to the overall class prototypes of all other classes, enhancing intra-class compactness. Formally, $\mathcal{L}_{il}$ is defined as:

$$\mathcal{L}_{il} = -\frac{1}{B}\sum_{i=1}^{B}\log\frac{\exp(z_i \cdot \mu_{y_i}/v)}{\sum_{j=1,j\neq y_i}^{C}\exp(z_i \cdot \mu_j/v)}, \tag{6}$$

where z_i is the feature representation of x_i extracted by the encoder $\Phi(\cdot)$, μ_{y_i} is the overall class prototype for class y_i, v is the temperature parameter, and $\exp(\cdot)$ is the exponential function.

The radial uniformity loss $\mathcal{L}_{ol}$ promotes a uniform feature distribution around each labeled class prototype by minimizing pairwise cosine similarities between normalized feature representation directions within a class, preventing collapse into a narrow region. Formally, $\mathcal{L}_{ol}$ is defined as:

$$\mathcal{L}_{ol} = \frac{1}{|C|}\sum_{c=1}^{|C|}\left[\frac{1}{N_c(N_c-1)}\sum_{i=1}^{N_c}\sum_{\substack{j=1,\\ j\neq i}}^{N_c}(r_{c,i}\cdot r_{c,j})\right]^2, \tag{7}$$

where $r_{c,i} = (z_{c,i} - \mu_c^l)/\|z_{c,i} - \mu_c^l\|_2$ is the normalized feature representation direction of sample x_i in class c, and $z_{c,i}$ is the feature representation of sample x_i belonging to class c.

Accordingly, the local structure loss $\mathcal{L}_l$ integrates these two loss items and is formulated as:

$$\mathcal{L}_l = \mathcal{L}_{il} + \mathcal{L}_{ol}. \tag{8}$$

Class-adaptive re-weighting. While the dual regularizer structures the feature representation space into the hub-spoke geometry and improves its discriminability, enhancing the feature discriminability for minority classes remains a critical challenge in long-tailed semi-supervised learning. Fortunately, our empirical analysis reveals that samples from minority classes exhibit a significantly larger deviation between the overall class prototype and the labeled class prototype, as shown in Fig. 3. This observation motivates us to leverage this deviation for a class-adaptive re-weighting strategy that dynamically prioritizes samples from minority classes. Specifically, for each sample, we compute the cosine similarity between the displacement from the overall class prototype to the sample feature representation (i.e., $q_i = z_i - \mu_{y_i}$) and the overall deviation vector of the class ($\Delta_{y_i} = \mu_{y_i} - \mu_{y_i}^l$). This

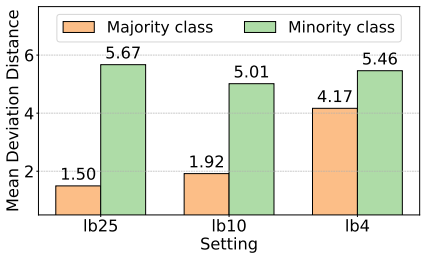


Figure 3: A comparison of the mean deviation distance between the overall class prototype and the labeled class prototype.

similarity score is then used to weight the unsupervised loss in Eq. 2. Formally, the weight w_i for sample x_i is defined as:

$$w_i = 1 + \max\left(0, \frac{q_i \cdot \Delta_{y_i}}{\|q_i\|_2 \cdot \|\Delta_{y_i}\|_2}\right), \tag{9}$$

where $\max(\cdot)$ denotes the maximum function. Since the deviation is inherently larger for minority classes, samples belonging to minority classes receive higher weights, thereby adaptively shifting the model's focus toward these classes.

4.2 Unknown Class Identification and Modeling

To effectively identify high-uncertainty samples from unknown classes, we employ an evidence-based classifier $g(\cdot)$ for uncertainty quantification. Given a sample x, the classifier produces a non-negative evidence vector e, where each component e_c indicates the support for classifying the sample into class c. This vector parameterizes a Dirichlet distribution $\mathrm{Dir}(\pi|\alpha)$ over the class probabilities, with the concentration parameters defined as $\alpha = e + 1$. The predictive uncertainty is derived from the Dirichlet strength $S = \sum_{c=1}^{C} \alpha_c$, with the uncertainty defined as $u = C/S$. A high value of u corresponds to low overall evidence, suggesting that the sample likely belongs to an unknown class.

Unknown class identification. During training, we leverage this uncertainty measure to identify potential high-uncertainty unknown class samples. For each batch, we rank samples by their uncertainty u in descending order and select the top-k most uncertain samples as candidates from unknown classes, where k is set to 10% of the current batch size (see Appendix D for the hyperparameter sensitivity analysis on k). Concretely, let u_k be the k-th largest uncertainty in the batch. We then form an evidence-based mask $\mathbb{I}(u \geq u_k)$ to select unknown class sample candidates, thus constructing the identified high-uncertainty unknown class sample set $\hat{D}_{O,k}$, which serves as the foundation for subsequent unknown class modeling.

To ensure the evidence-based classifier produces well-calibrated uncertainty estimates, we optimize it using an evidence loss $\mathcal{L}_{edl}$ that regularizes the evidence distribution to align with ground-truth labels while maintaining uncertainty awareness:

$$\mathcal{L}_{edl} = \frac{1}{B}\sum_{i=1}^{B}\left[\sum_{j=1}^{C} y_{ij}\left(\psi(S_i) - \psi(\alpha_{ij})\right) + \lambda_{kl} \cdot \mathrm{KL}(\mathrm{Dir}(\pi_i|\tilde{\alpha}_i)\|\mathrm{Dir}(\pi_i|\mathbf{1}))\right], \tag{10}$$

where $\psi(\cdot)$ is the digamma function and is defined as the logarithmic derivative of the gamma function, α_{ij} denotes the concentration parameter for class j of sample i, and $\lambda_{kl} = \min(0.6, \max(0, \frac{t_{ep}}{10}))$, where t_{ep} denotes the number of epochs since $\mathcal{L}_{edl}$ begins to be optimized during training. The adjusted parameters $\tilde{\alpha}_i = y_i + (\mathbf{1} - y_i) \cdot \alpha_i$ are obtained by removing the non-misleading evidence from the original parameters α_i.

Unknown class modeling. Building upon the established hub-spoke geometry and the identified high-uncertainty unknown class sample set $\hat{D}_{O,k}$, we introduce an unknown class sample clustering loss $\mathcal{L}_{oc}$ to explicitly model their feature distribution. The global labeled class prototype $\bar{\mu}^l$, which serves as the hub, is approximately equidistant from all known class prototypes and well separated from the corresponding spokes, making it a natural anchor for unknown class modeling. Without explicit unknown class modeling, unknown class samples may be dispersed among different known class spokes, causing feature and evidence overlap with known class samples. We therefore regularize the representations of the identified high-uncertainty unknown class samples by pulling them toward the hub. Since these samples exhibit low evidence by construction, guiding them toward the hub induces a low-evidence region around the hub, indicating the absence of known-class evidence (see Appendix 5 for theoretical analysis). As the hub remains geometrically separated from all known class spokes, this design not only preserves the discriminability of known classes but also improves the separation between known and unknown classes in the feature space and their uncertainty distributions, as illustrated in Fig. 2d. Formally, $\mathcal{L}_{oc}$ is defined as:

$$\mathcal{L}_{oc} = \frac{1}{|\hat{D}_{O,k}|}\sum_{i=1}^{|\hat{D}_{O,k}|}\|z_i^o - \bar{\mu}^l\|_2^2, \tag{11}$$

where $\hat{D}_{O,k}$ denotes the set of identified high-uncertainty unknown class samples and z_i^o represents the feature representation of the i-th identified unknown class sample.

4.3 End-to-End Training Process

Our training process is structured into three phases to ensure stable learning. First, a warm-up phase lasting sixty epochs focuses on learning stable initial feature representations using $\mathcal{L}_{warm1}$:

$$\mathcal{L}_{warm1} = \mathcal{L}_{lcon} + \lambda \left(\mathcal{L}_g + \mathcal{L}_{econ} \right), \tag{12}$$

where $\mathcal{L}_{lcon}$ and $\mathcal{L}_{econ}$ denote the contrastive losses between the weak and strong augmentation views for the classification head $f(\cdot)$ and the evidence head $g(\cdot)$, respectively.

Subsequently, a second warm-up phase lasting twenty epochs integrates the supervised and unsupervised loss to establish baseline performance and stabilize pseudo-labeling, guided by $\mathcal{L}_{warm2}$:

$$\mathcal{L}_{warm2} = \mathcal{L}_{sup} + \mathcal{L}_{unsup} + \lambda \cdot \mathcal{L}_{econ}. \tag{13}$$

Finally, the main training phase leverages the full objective function to jointly address mismatches in both class distribution and label space:

$$\mathcal{L}_{overall} = \mathcal{L}_{sup} + \mathcal{L}_{unsup} + \lambda \left(\mathcal{L}_{econ} + \mathcal{L}_g + \mathcal{L}_l + \mathcal{L}_{edl} + \mathcal{L}_{oc} \right). \tag{14}$$

Here, λ is a balancing hyperparameter set to 0.1 across all three phases. The class-adaptive weight w_i from Eq. 9 is incorporated into the $\mathcal{L}_{unsup}$ term to adaptively focus on minority classes. Additionally, the confidence-based mask in Eq. 2 is replaced by the combination of the confidence-based mask and the negated evidence-based mask. The pseudo-code is detailed in Algorithm 1 in Appendix B.

5 Theoretical Analysis

We analyze how uncertainty-guided unknown class modeling induces a low-evidence region around the hub. Let $\bar{\mu}^l = \frac{1}{C}\sum_{c=1}^{C} \mu_c^l$ denote the hub. Following Sec. 4.2, $\alpha_c(z) = 1 + g_c(z), S(z) = C + \sum_{c=1}^{C} g_c(z), u(z) = \frac{C}{S(z)}$, where $g_c(\cdot)$ indicates e_c.

Lemma 1 (Low evidence of selected candidates)**.** *Let $\widehat{\mathcal{D}}_{O,k}$ be the top-k most uncertain candidates in a training batch, and let u_k be the k-th largest uncertainty in the batch. Then, for every $z_i^o \in \widehat{\mathcal{D}}_{O,k}$,*

$$g_c(z_i^o) \le \varepsilon_{O,k}, \qquad \varepsilon_{O,k} := C\left(\frac{1}{u_k} - 1\right), \quad c = 1, \dots, C. \tag{15}$$

Proof. Since $u(z_i^o) \ge u_k$,

$$S(z_i^o) = \frac{C}{u(z_i^o)} \le \frac{C}{u_k}. \tag{16}$$

Hence

$$\sum_{j=1}^{C} g_j(z_i^o) \le C\left(\frac{1}{u_k} - 1\right). \tag{17}$$

Using $g_c(z_i^o) \ge 0$ gives Eq. 15. □

Theorem 1 (Low-evidence region around the hub)**.** *Suppose that the evidence function $g_c(\cdot)$ is L_c-Lipschitz, i.e., $|g_c(z_i) - g_c(z_j)| \le L_c \|z_i - z_j\|_2$ for any z_i, z_j. For $\mathcal{H}(r) = \{z : \|z - \bar{\mu}^l\|_2 \le r\}$, we have*

$$\sup_{z \in \mathcal{H}(r)} g_c(z) \le \varepsilon_{O,k} + L_c\left(r + \sqrt{L_{oc}}\right). \tag{18}$$

Proof. Let $n_{O,k} = |\widehat{\mathcal{D}}_{O,k}|$. From Eq. 11, we have $L_{oc} = \frac{1}{n_{O,k}}\sum_{i=1}^{n_{O,k}} \|z_i^o - \bar{\mu}^l\|_2^2$. Since L_{oc} is the average squared distance to the hub, there exists $z_*^o \in \widehat{\mathcal{D}}_{O,k}$ such that$\|z_*^o - \bar{\mu}^l\|_2^2 \le L_{oc}$, and hence

$$\|z_*^o - \bar{\mu}^l\|_2 \le \sqrt{L_{oc}}. \tag{19}$$

For any $z \in \mathcal{H}(r)$,

$$\|z - z_*^o\|_2 \le \|z - \bar{\mu}^l\|_2 + \|z_*^o - \bar{\mu}^l\|_2 \le r + \sqrt{L_{oc}}. \tag{20}$$

Since g_c is L_c-Lipschitz continuous, Lemma 1 gives

$$g_c(z) \leq g_c(z_*^o) + L_c\|z - z_*^o\|_2 \leq \varepsilon_{O,k} + L_c\left(r + \sqrt{L_{oc}}\right). \tag{21}$$

Taking the supremum over $z \in \mathcal{H}(r)$ yields Eq. 18. □

Hence, high uncertainty among the selected candidates and strong hub-centered clustering yield a quantitatively low-evidence hub region and correspondingly high evidential uncertainty.

6 Experiments

Datasets. We evaluate our method on five benchmarks (i.e., CIFAR-10 (Krizhevsky et al., 2009), CIFAR-100 (Krizhevsky et al., 2009), SVHN (Netzer et al., 2011), Food-101 (Bossard et al., 2014), and ImageNet-100 (Deng et al., 2009)). To simulate class distribution mismatch, we construct a balanced labeled dataset with different sizes (i.e., 10 and 25 samples per class), while the unlabeled dataset follows an imbalanced class distribution. For label space mismatch, we introduce unknown class samples into the unlabeled data. Concretely, we add 10 unknown classes from ImageNet-127 (Deng et al., 2009) to the CIFAR-10 and SVHN datasets, 50 from ImageNet-127 to the CIFAR-100 and Food-101 datasets, and 50 from CIFAR-100 to the ImageNet-100 dataset.

Baselines. We compare our method with three SSL algorithms, including FixMatch (Sohn et al., 2020), FreeMatch (Wang et al., 2023), and SoftMatch (Chen et al., 2023); four LTSSL algorithms, including SimPro (Du et al., 2024), CDMAD (Lee & Kim, 2024), Meta-Expert (Hou & Jia, 2025), and SC-SSL (Tian et al., 2026); and six OSSL algorithms, including OpenMatch (Saito et al., 2021), IOMatch (Li et al., 2023), SCOMatch (Wang et al., 2024), CaliMatch (Bae et al., 2025), ANEDL (Yu et al., 2024), and GGR (Chen et al., 2026). Moreover, we use the supervised learning (SL) setting as an upper-bound reference for performance. Implementation details are provided in Appendix C.

6.1 Experimental Results and Analyses

Table 1: Classification accuracy (%) on the test data of CIFAR-10, CIFAR-100, SVHN, Food-101, and ImageNet-100 with varying known/unknown classes and labeled set sizes. We report the mean with standard deviation over three runs with different random seeds.

	Dataset	CIFAR-10		CIFAR-100		SVHN		Food-101		ImageNet-100	
	Known / Unknown	10 / 10		100 / 50		10 / 10		101 / 50		100 / 50	
	Labels per class	10	25	10	25	10	25	10	25	10	25
SL	CE	78.69± 5.10	80.21± 4.85	51.01± 0.73	55.74± 0.57	88.93± 1.19	89.18± 1.05	39.79± 0.48	42.56± 0.38	44.63± 0.27	46.61± 1.16
	LA	85.45± 2.64	86.08± 2.47	56.20± 0.48	59.67± 0.45	92.31± 1.01	92.63± 1.19	45.48± 0.37	47.44± 0.20	49.61± 0.25	51.37± 0.86
SSL	FixMatch	69.57± 5.03	76.86± 4.69	39.40± 1.31	51.23± 1.29	89.71± 2.60	90.82± 1.02	15.95± 0.55	23.23± 0.25	27.44± 0.81	34.95± 0.76
	FreeMatch	64.45± 6.47	75.72± 5.27	32.56± 1.62	47.52± 1.21	89.10± 2.30	89.83± 2.54	19.64± 0.34	27.65± 0.20	27.39± 0.45	36.31± 0.97
	SoftMatch	67.97± 5.22	74.56± 2.94	34.22± 1.82	48.74± 0.46	86.34± 1.30	87.44± 2.05	19.68± 0.57	27.71± 1.07	27.13± 1.48	36.01± 1.01
LTSSL	SimPro	26.19± 3.62	68.72± 4.87	35.10± 1.24	49.68± 1.29	59.60± 5.27	83.89± 3.17	13.06± 1.23	25.18± 0.45	26.19± 0.46	36.14± 0.39
	CDMAD	58.77± 8.60	70.38± 4.41	24.47± 2.37	31.35± 3.08	81.35± 3.20	82.54± 2.09	6.35± 1.25	8.31± 2.47	14.52± 2.49	17.72± 1.43
	Meta-Expert	57.37± 7.89	69.83± 5.37	32.05± 1.54	47.54± 1.13	79.00± 2.66	86.38± 1.14	17.74± 0.39	25.32± 0.59	27.08± 0.37	34.40± 0.34
	SC-SSL	67.26± 7.13	77.23± 4.41	35.23± 1.63	49.03± 0.66	89.05± 3.71	89.63± 2.11	17.04± 0.23	23.95± 0.12	26.97± 1.18	35.37± 1.0
OSSL	OpenMatch	47.21± 5.06	68.13± 3.54	32.81± 0.19	47.31± 0.44	84.88± 5.73	90.47± 0.92	10.61± 0.92	16.00± 0.39	19.29± 0.83	25.74± 1.34
	IOMatch	67.33± 5.06	77.23± 3.11	38.90± 1.50	50.62± 1.12	87.98± 2.87	90.80± 1.50	7.62± 0.37	14.46± 0.73	9.04± 6.76	21.92± 0.04
	SCOMatch	63.41± 6.06	74.07± 3.41	30.57± 2.90	47.42± 1.81	88.52± 2.86	90.74± 0.94	13.14± 2.65	16.03± 2.88	17.36± 2.36	22.04± 1.31
	CaliMatch	51.33± 10.72	67.80± 0.16	35.99± 6.00	50.86± 5.36	55.55± 14.84	90.71± 1.66	7.22± 1.95	10.22± 5.44	15.17± 1.89	25.78± 3.46
	ANEDL	27.58± 0.79	48.21± 4.01	22.87± 3.54	37.26± 3.59	56.30± 1.85	76.12± 0.64	8.77± 0.69	15.37± 1.87	12.51± 1.56	20.23± 0.65
	GGR	61.24± 6.81	69.70± 6.69	25.69± 1.43	41.67± 2.63	88.21± 0.96	88.67± 1.54	7.25± 0.66	13.47± 0.98	13.20± 0.48	20.35± 0.91
	Ours	**71.55± 3.50**	**79.20± 2.42**	**42.65± 1.54**	**52.99± 0.61**	**90.38± 1.16**	**91.28± 0.80**	**19.82± 0.20**	**27.87± 1.12**	**27.52± 0.30**	**36.83± 0.81**

Main Results. Table 1 presents the experimental results on CIFAR-10, CIFAR-100, SVHN, Food-101, and ImageNet-100 datasets under varying known/unknown classes and labeled set sizes. As illustrated in Table 1, our method consistently outperforms all baseline methods and is robust to dual mismatch. On CIFAR-10, we outperform the recent method CaliMatch by **20.22%** and **11.40%** with 10 and 25 labeled samples per class, respectively. Moreover, the performance gain persists on CIFAR-100, where we achieve a 3.25% performance gain over the best-performing baseline FixMatch, with 10 labeled samples per class. On SVHN, while the performance gap narrows due to the dataset's lower complexity and clearer known and unknown class distinction, our method still achieves competitive performance.

On the more challenging Food-101 dataset, our method substantially outperforms Meta-Expert by 2.08% and 2.55%, respectively, despite the overall low accuracy caused by its fine-grained and numerous categories. Finally, on the highly complex ImageNet-100, with 10 labeled samples per class, our method achieves 27.52%, surpassing CaliMatch by a substantial 12.35%; with 25 labeled samples per class, we outperform the previous best FreeMatch by 0.52%. Notably, our method exhibits a lower standard deviation across three runs compared to most baseline methods. Further experimental results and analyses are provided in Appendix D.

6.2 Ablation Analyses and Discussions

The effect of each module. We conduct comprehensive ablation studies on CIFAR-10 and CIFAR-100 to evaluate the contribution of each proposed module across three settings. As shown in Table 2, progressively integrating our modules leads to consistent performance gains.

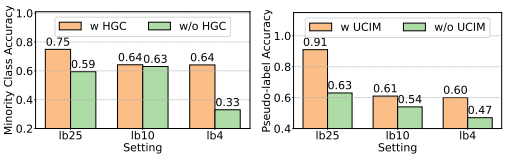
The baseline model, devoid of any proposed modules, exhibits limited capability, especially in the lb4 setting, where it attains only 38.24% and 26.50% accuracy on CIFAR-10 and CIFAR-100, respectively. Incorporating the class-adaptive re-weighting (CAW) module brings consistent improvements, with gains of up to 2.10% on CIFAR-10 and 2.50% on CIFAR-100. Introducing the local structure regularization (LSR) module yields more substantial improvements, boosting accuracy on CIFAR-10 by **10.46%** and CIFAR-100 by 2.35% in the lb4 setting.

Further adding the global structure regularization (GSR) enhances the local-only variant by an average of 0.65%. Finally, the unknown class identification and modeling (UCIM) module contributes additional gains, averaging 2.06% on CIFAR-10 and 0.45% on CIFAR-100. The full model achieves the best performance, confirming that each module addresses a unique aspect of the challenging dual-mismatch problem involved in the semi-supervised learning scenario and that their effects are complementary and cumulative.

Table 2: Classification accuracy (%) with and without the key components in the proposed method.

Module				CIFAR-10			CIFAR-100		
CAW	LSR	GSR	UCIM	lb4	lb10	lb25	lb4	lb10	lb25
				38.24	64.46	74.40	26.50	40.75	51.00
✓				40.34	64.86	76.00	27.89	43.25	53.00
✓	✓			50.80	68.05	76.95	30.24	43.25	53.49
✓	✓	✓		54.30	68.54	76.88	30.25	43.43	53.30
✓	✓	✓	✓	**56.89**	**70.96**	**78.05**	**30.43**	**44.37**	**53.53**

Moreover, we evaluate two key modules of our method. The hub-spoke geometry construction (HGC) module which combines LSR and GSR mitigates minority class performance degradation caused by class imbalance. As shown in Fig. 4a, HGC substantially improves minority class accuracy, e.g., from 33% to 64% in the lb4 setting on CIFAR-10, confirming its effectiveness in enhancing feature learning for minority classes under imbalanced distributions. In parallel, the unknown class identification and modeling (UCIM) module boosts pseudo-label accuracy from 47% to 60% in the same lb4 setting on CIFAR-10 as illustrated in Fig. 4b, demonstrating that reliable unknown class handling enhances both pseudo-label quality and robustness against dual mismatch. Further ablation analyses and discussions are provided in Appendix E.



(a) (b)

Figure 4: Ablation studies on HGC module for minority class accuracy and UCIM module for pseudo-label accuracy.

7 Conclusion

In this work, we addressed the challenging and realistic scenario of dual-mismatched semi-supervised learning, where unlabeled data exhibits an imbalanced class distribution and contains unknown class samples. Our proposed method integrates a structured hub-spoke geometry with evidence-based unknown class identification. The hub-spoke geometry ensures a discriminative and robust feature representation space, while the evidence-based classifier provides well-calibrated uncertainty estimates. We further theoretically show that the combination of high-uncertainty samples and hub-centered clustering induces a low-evidence region around the hub, providing theoretical support for our unknown class modeling strategy. Extensive experiments demonstrate that our method consistently outperforms previous state-of-the-art methods under various dual mismatch conditions.

AI USE STATEMENT

In this work, we used generative AI tools to assist in the writing of proofs and assist with translation. We have not used generative AI tools for other tasks requiring disclosure. Additionally, we used generative AI tools to edit software code and the research paper to improve readability. We have reviewed all AI-assisted work. AI-generated proof steps were manually checked for logical correctness. Translations were verified against the source text. AI-edited code was tested for correctness by two authors. All readability edits were approved by all authors. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

ETHICS STATEMENT

Our paper submission does not raise any questions regarding the Code of Ethics.

REPRODUCIBILITY STATEMENT

To ensure the reproducibility of our work, we provide a comprehensive description of the experimental setup in Sec. 6 and Appendix C. All datasets used in our experiments are publicly available, and the source code is included as supplementary material. These resources collectively enable faithful reproduction of our main experimental results.

REFERENCES

Eduardo Aguilar, Bogdan Raducanu, Petia Radeva, and Joost van de Weijer. Continual evidential deep learning for out-of-distribution detection. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 3444–3454, 2023.

Emanuel Sanchez Aimar, Nathaniel Helgesen, Yonghao Xu, Marco Kuhlmann, and Michael Felsberg. Flexible distribution alignment: Towards long-tailed semi-supervised learning with proper calibration. In *European Conference on Computer Vision (ECCV)*, volume 15112, pp. 307–327, 2024.

Jinsoo Bae, Seoung Bum Kim, and Hyungrok Do. Calimatch: Adaptive calibration for improving safe semi-supervised learning. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 2867–2876, 2025.

Lukas Bossard, Matthieu Guillaumin, and Luc Van Gool. Food-101 - mining discriminative components with random forests. In *European Conference on Computer Vision (ECCV)*, volume 8694, pp. 446–461, 2014.

Paola Cascante-Bonilla, Fuwen Tan, Yanjun Qi, and Vicente Ordonez. Curriculum labeling: Revisiting pseudo-labeling for semi-supervised learning. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, volume 35, pp. 6912–6920, 2021.

Dongdong Chen, Wei Wang, Wei Gao, and Zhi-Hua Zhou. Tri-net for semi-supervised deep learning. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 2014–2020, 2018.

Hao Chen, Ran Tao, Yue Fan, Yidong Wang, Jindong Wang, Bernt Schiele, Xing Xie, Bhiksha Raj, and Marios Savvides. Softmatch: Addressing the quantity-quality tradeoff in semi-supervised learning. In *International Conference on Learning Representations (ICLR)*, 2023.

Jiahe Chen, Qian Shao, Qiyuan Chen, Jiaying He, Jintai Chen, Jian Wu, and Hongxia Xu. Geometric gradient rectification for safe open-set semi-supervised learning. In *European Conference on Computer Vision (ECCV)*, 2026.

Bo Cheng, Jueqing Lu, Yuan Tian, Haifeng Zhao, Yi Chang, and Lan Du. Cgmatch: A different perspective of semi-supervised learning. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 15381–15391, 2025.

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 248–255, 2009.

Chaoqun Du, Yizeng Han, and Gao Huang. Simpro: A simple probabilistic framework towards realistic long-tailed semi-supervised learning. In *International Conference on Machine Learning (ICML)*, pp. 11686–11703, 2024.

Yaxin Hou and Yuheng Jia. A square peg in a square hole: Meta-expert for long-tailed semi-supervised learning. In *International Conference on Machine Learning (ICML)*, 2025.

Yaxin Hou, Bo Han, Yuheng Jia, Hui LIU, and Junhui Hou. Keep it on a leash: Controllable pseudo-label generation towards realistic long-tailed semi-supervised learning. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2025.

Yaxin Hou, Jun Ma, Hanyang Li, Bo Han, Jie Yu, and Yuheng Jia. Beyond distribution estimation: Simplex anchored structural inference towards universal semi-supervised learning. In *International Conference on Machine Learning (ICML)*, 2026.

Audun Jøsang. *Subjective Logic*. Springer, 2016.

Zhanghan Ke, Daoye Wang, Qiong Yan, Jimmy S. J. Ren, and Rynson W. H. Lau. Dual student: Breaking the limits of the teacher in semi-supervised learning. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 6728–6736, 2019.

Jaehyung Kim, Youngbum Hur, Sejun Park, Eunho Yang, Sung Ju Hwang, and Jinwoo Shin. Distribution aligning refinery of pseudo-label for imbalanced semi-supervised learning. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 33, pp. 14567–14579, 2020.

Alex Krizhevsky, Geoffrey Hinton, et al. Learning multiple layers of features from tiny images. 2009.

Hyuck Lee and Heeyoung Kim. CDMAD: class-distribution-mismatch-aware debiasing for class-imbalanced semi-supervised learning. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 23891–23900, 2024.

Hyuck Lee, Seungjae Shin, and Heeyoung Kim. ABC: auxiliary balanced classifier for class-imbalanced semi-supervised learning. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 34, pp. 7082–7094, 2021.

Zekun Li, Lei Qi, Yinghuan Shi, and Yang Gao. Iomatch: Simplifying open-set semi-supervised learning with joint inliers and outliers utilization. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 15824–15833, 2023.

Ilya Loshchilov and Frank Hutter. SGDR: stochastic gradient descent with warm restarts. In *International Conference on Learning Representations (ICLR)*, 2017.

Chengcheng Ma, Ismail Elezi, Jiankang Deng, Weiming Dong, and Changsheng Xu. Three heads are better than one: Complementary experts for long-tailed semi-supervised learning. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, volume 38, pp. 14229–14237, 2024.

Takeru Miyato, Shin-ichi Maeda, Masanori Koyama, and Shin Ishii. Virtual adversarial training: A regularization method for supervised and semi-supervised learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 41(8):1979–1993, 2019.

Yuval Netzer, Tao Wang, Adam Coates, Alessandro Bissacco, Bo Wu, and Andrew Y. Ng. Reading digits in natural images with unsupervised feature learning. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2011.

Youngtaek Oh, Dong-Jin Kim, and In So Kweon. DASO: distribution-aware semantics-oriented pseudo-label for imbalanced semi-supervised learning. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9786–9796, 2022.

Jingen Qu, Yufei Chen, Xiaodong Yue, Wei Fu, and Qiguang Huang. Hyper-opinion evidential deep learning for out-of-distribution detection. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 37, pp. 84645–84668, 2024.

Kuniaki Saito, Donghyun Kim, and Kate Saenko. Openmatch: Open-set consistency regularization for semi-supervised learning with outliers. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 34, pp. 25956–25967, 2021.

Kari Sentz and Scott Ferson. *Combination of Evidence in Dempster-Shafer Theory*. Sandia National Laboratories, 2002.

Kihyuk Sohn, David Berthelot, Nicholas Carlini, Zizhao Zhang, Han Zhang, Colin Raffel, Ekin Dogus Cubuk, Alexey Kurakin, and Chun-Liang Li. Fixmatch: Simplifying semi-supervised learning with consistency and confidence. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 33, pp. 596–608, 2020.

Senmao Tian, Xiang Wei, and Shunli Zhang. Sampling control for imbalanced calibration in semi-supervised learning. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, volume 40, pp. 25914–25922, 2026.

Erik Wallin, Lennart Svensson, Fredrik Kahl, and Lars Hammarstrand. Improving open-set semi-supervised learning with self-supervision. In *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 2356–2365, 2024.

Yidong Wang, Hao Chen, Yue Fan, Wang Sun, Ran Tao, Wenxin Hou, Renjie Wang, Linyi Yang, Zhi Zhou, Lan-Zhe Guo, Heli Qi, Zhen Wu, Yufeng Li, Satoshi Nakamura, Wei Ye, Marios Savvides, Bhiksha Raj, Takahiro Shinozaki, Bernt Schiele, Jindong Wang, Xing Xie, and Yue Zhang. USB: A unified semi-supervised learning benchmark for classification. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2022.

Yidong Wang, Hao Chen, Qiang Heng, Wenxin Hou, Yue Fan, Zhen Wu, Jindong Wang, Marios Savvides, Takahiro Shinozaki, Bhiksha Raj, Bernt Schiele, and Xing Xie. Freematch: Self-adaptive thresholding for semi-supervised learning. In *International Conference on Learning Representations (ICLR)*, 2023.

Zerun Wang, Liuyu Xiang, Lang Huang, Jiafeng Mao, Ling Xiao, and Toshihiko Yamasaki. Scomatch: Alleviating overtrusting in open-set semi-supervised learning. In *European Conference on Computer Vision (ECCV)*, volume 15109, pp. 217–233, 2024.

Chen Wei, Kihyuk Sohn, Clayton Mellina, Alan L. Yuille, and Fan Yang. Crest: A class-rebalancing self-training framework for imbalanced semi-supervised learning. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10857–10866, 2021.

Tong Wei and Kai Gan. Towards realistic long-tailed semi-supervised learning: Consistency is all you need. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3469–3478, 2023.

Qing Yu, Daiki Ikami, Go Irie, and Kiyoharu Aizawa. Multi-task curriculum framework for open-set semi-supervised learning. In *European Conference on Computer Vision (ECCV)*, volume 12357, pp. 438–454, 2020.

Yang Yu, Danruo Deng, Furui Liu, Qi Dou, Yueming Jin, Guangyong Chen, and Pheng-Ann Heng. ANEDL: adaptive negative evidential deep learning for open-set semi-supervised learning. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, volume 38, pp. 16587–16595, 2024.

Sergey Zagoruyko and Nikos Komodakis. Wide residual networks. In *Proceedings of the British Machine Vision Conference (BMVC)*, 2016.

Shenkai Zhao, Xinao Zhang, Lipeng Pan, Xiaobin Xu, and Danilo Pelusi. Evidential deep active learning for semi-supervised classification. *CoRR*, abs/2505.20691, 2025.

TABLE OF CONTENTS FOR THE APPENDIX

- A: More Related Work 14
- B: Pseudo-code of the Proposed Method 14
- C: Implementation Details 14
- D: Further Experimental Results and Analyses 15
- E: Further Ablation Analyses and Discussions 16
- F: Compute Resources 17

A More Related Work

Evidential deep learning (EDL), grounded in the Dempster-Shafer theory (Sentz & Ferson, 2002) and subjective logic (Jøsang, 2016), has emerged as a prominent framework for uncertainty quantification and unknown class sample detection. For instance, HEDL (Qu et al., 2024) introduces hyper-domain evidence to capture both certain and ambiguous classes, enhancing uncertainty estimation via an opinion projection mechanism, while CEDL (Aguilar et al., 2023) integrates EDL into continual learning, leveraging its ability to distinguish unknown class samples from past known class ones. EDALSSC (Zhao et al., 2025) proposes a semi-supervised evidential model that dynamically balances evidence for active learning sample selection. Notably, ANEDL (Yu et al., 2024) is positioned as the first to adapt EDL for OSSL, utilizing an adaptive negative optimization strategy to improve unknown class sample detection and emphasize uncertain classes. However, it does not address the realistic scenario involving both class distribution and label space mismatches between labeled and unlabeled data.

B Pseudo-code of the Proposed Method

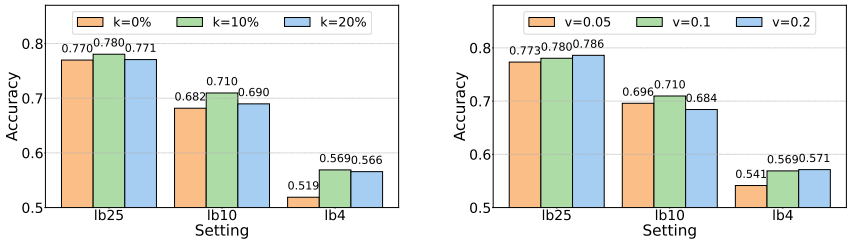
Algorithm 1 Training process of our proposed method

Require: Labeled and unlabeled datasets $\mathcal{D}_l$ and $\mathcal{D}_u$.
Ensure: Encoder $\Phi(\cdot)$, classification head $f(\cdot)$, and evidence head $g(\cdot)$.

```
Initialize the parameters of Φ(·), f(·), and g(·) randomly.
for epoch = 1, 2, ... do
    for batch = 1, 2, ... do
        if epoch < 60 then
            Calculate L_warm1 by Eq. equation 12;
        end if
        if 60 ≤ epoch < 80 then
            Calculate L_warm2 by Eq. equation 13;
        end if
        if epoch ≥ 80 then
            Construct hub-spoke geometry by Eq. equation 5 and Eq. equation 8;
            Compute class-adaptive weight w by Eq. equation 9;
            Calculate evidence loss L_edl by Eq. equation 10 for the evidence head;
            Pull identified unknown class samples toward the hub by Eq. equation 11;
            Obtain the overall loss L_overall by Eq. equation 14;
        end if
        Update network parameters via gradient descent;
    end for
end for
```

C Implementation Details

We follow the default settings and hyperparameters in USB (Wang et al., 2022). Specifically, the batch size of labeled data B_l is set to 64, while that of unlabeled data B_u is set to 7 times B_l; the confidence threshold τ is set to 0.95, and the temperature parameter v is set to 0.1. Moreover, we use the WRN-28-2 (Zagoruyko & Komodakis, 2016) architecture, the SGD optimizer with momentum 0.9, and weight decay 5e-4 for training. We use a cosine learning rate decay scheme (Loshchilov & Hutter, 2017), setting the learning rate to $\eta \cos\left(\frac{7\pi t}{16T}\right)$. Here, $\eta = 0.03$ is the initial learning rate, t is the current training step, and $T = 2^{18}$ is the total number of training steps. We repeat each experiment with three different random seeds (i.e., 0, 1, and 4) and report the mean and standard deviation of the performance. We conduct the experiments on a single NVIDIA RTX 4090 GPU using PyTorch v2.4.1.



(a) (b)

Figure 5: (a) and (b) respectively compare classification accuracy on the CIFAR-10 test data for $k \in \{0\%, 10\%, 20\%\}$ and $v \in \{0.05, 0.1, 0.2\}$ across the three settings. We clearly see that $k = 10\%$ and $k = 20\%$ yield comparable results, and performance remains stable across different values of v.

D Further Experimental Results and Analyses

D.1 Hyperparameter Sensitivity Analysis

We perform hyperparameter sensitivity analysis on three key hyperparameters: k, the percentage of high-uncertainty samples selected per batch (i.e., the number of identified unknown class samples per batch), v, the temperature scaling factor in Eq. 6, and λ, the balancing weight for losses in Eq. 14. As shown in Fig. 5a, our method is robust to the choice of k. On CIFAR-10, both $k = 10\%$ and $k = 20\%$ yield nearly identical performance and significantly outperform the $k = 0\%$ baseline across all three settings. Specifically, $k = 10\%$ improves over the baseline by +5.0%, +2.8%, and +1.0% for 4, 10, and 25 labeled samples per class, respectively. The comparable results for $k = 20\%$ (i.e., gains of +4.7%, +0.8%, and +0.1%) indicate our method's relative insensitivity to this hyperparameter. Similarly, as shown in Fig. 5b, performance remains stable across different values of v, with an average standard deviation of merely 0.99%, confirming its robustness. For λ, we evaluate three values (0.05, 0.10, 0.20) on CIFAR-10, with results reported in Table 3. The default $\lambda = 0.10$ achieves the best or near-best accuracy across all labeled settings, e.g., 56.89% (lb4) and 70.96% (lb10), while $\lambda = 0.20$ yields a slightly higher result (78.24%) only for lb25. Overall, the performance variation across different λ values remains modest, demonstrating that our approach is also robust to this hyperparameter.

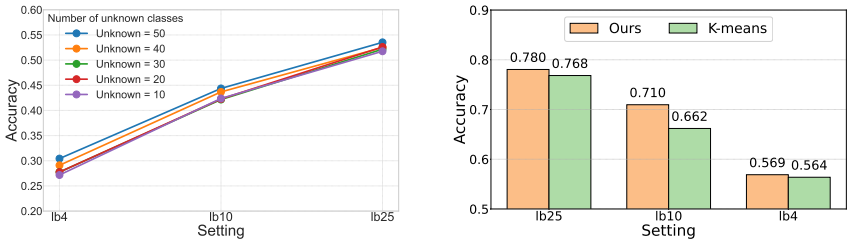
Table 3: Classification accuracy (%) with different λ values on CIFAR-10. The results show that $\lambda = 0.1$ achieves the best or near-best performance across all settings.

λ	lb4	lb10	lb25
0.05	52.74	69.18	77.43
0.10	**56.89**	**70.96**	78.05
0.20	55.34	68.43	**78.24**

D.2 Robustness to Unknown Class Diversity and Effectiveness of Unknown Class Modeling

To further evaluate the robustness and effectiveness of our proposed unknown class handling strategy, we examine the sensitivity of our method to the number of unknown classes by varying it from 50 to 10 on CIFAR-100 while keeping the total unknown class sample count fixed. As shown in Fig. 6a, our method maintains relatively stable performance across all settings despite substantial changes in the diversity of unknown classes. This robustness can be attributed to our hub-based unknown class modeling: instead of explicitly modeling each unknown category, we identify high-uncertainty samples based on their evidence and guide them toward a shared low-evidence region around the hub. Since the hub provides a common target for uncertain samples regardless of their specific unknown class, the same modeling mechanism can be applied as the number of unknown classes changes.

We further compare our unknown class modeling with a K-means clustering baseline on the identified high-uncertainty unknown class samples. For a fair comparison, we set the number of clusters K



(a) (b)

Figure 6: (a) reports classification accuracy on CIFAR-100 under different numbers of unknown classes (50, 40, 30, 20, 10) across the three lb settings; (b) compares our method with K-means clustering on the identified unknown class samples on CIFAR-10.

in K-means to the ground-truth number of unknown classes. As illustrated in Fig. 6b, our method consistently outperforms K-means across all three settings on CIFAR-10, achieving absolute gains of 0.5%, 4.8%, and 1.2% in the lb4, lb10, and lb25 settings, respectively. The superior performance, particularly in the lb10 setting, highlights the advantage of our hub-spoke geometry, which leverages evidence-based uncertainty to guide high-uncertainty unknown class samples toward a shared low-evidence hub, thereby enhancing known–unknown class discriminability and improving pseudo-label quality.

D.3 Statistical Significance

We also conduct pairwise t-tests ($\alpha = 0.05$) against thirteen baselines under different label settings to evaluate the robustness of our approach and the full results are given in Table 4. Our method achieves significant improvements over all LTSSL methods in **75%** of cases and over all OSSL methods in **76.7%** of cases, confirming the consistent superiority of our method.

Table 4: Statistical significance of performance differences assessed with pairwise t-test at a 0.05 significance level, reported as win/tie/loss counts.

	Method	lb10	lb25	Total
SSL	FixMatch	2 / 3 / 0	1 / 4 / 0	3 / 7 / 0
	FreeMatch	1 / 4 / 0	1 / 4 / 0	2 / 8 / 0
	SoftMatch	1 / 4 / 0	2 / 3 / 0	3 / 7 / 0
LTSSL	SimPro	5 / 0 / 0	3 / 2 / 0	8 / 2 / 0
	CDMAD	4 / 1 / 0	5 / 0 / 0	9 / 1 / 0
	Meta-Expert	4 / 1 / 0	4 / 1 / 0	8 / 2 / 0
	SC-SSL	2 / 3 / 0	3 / 2 / 0	5 / 5 / 0
OSSL	OpenMatch	4 / 1 / 0	5 / 0 / 0	9 / 1 / 0
	IOMatch	3 / 2 / 0	4 / 1 / 0	7 / 3 / 0
	SCOMatch	4 / 1 / 0	3 / 2 / 0	7 / 3 / 0
	CaliMatch	2 / 3 / 0	3 / 2 / 0	5 / 5 / 0
	ANEDL	5 / 0 / 0	5 / 0 / 0	10 / 0 / 0
	GGR	4 / 1 / 0	4 / 1 / 0	8 / 2 / 0
	Total	41 / 24 / 0	43 / 22 / 0	84 / 46 / 0

E Further Ablation Analyses and Discussions

To further validate the effectiveness of local structure regularization (LSR) and global structure regularization (GSR), we performed a fine-grained component-level analysis. As shown in Table 5, the integrated LSR module surpasses its individual component losses, namely $\mathcal{L}_{il}$ (LIL) and $\mathcal{L}_{ol}$ (LOL), by average margins of 1.51% on CIFAR-10 and 0.83% on CIFAR-100. Similarly, the complete GSR module achieves average improvements of 1.73% on CIFAR-10 and 0.72% on CIFAR-100 over its individual component losses $\mathcal{L}_{ig}$ (LIG) and $\mathcal{L}_{og}$ (LOG). These fine-grained ablation results further substantiate the efficacy of both LSR and GSR modules while demonstrating the complementary relationships among their individual component losses.

Table 5: Classification accuracy (%) with and without the $\mathcal{L}_{il}$ (LIL), $\mathcal{L}_{ol}$ (LOL), $\mathcal{L}_{ig}$ (LIG), and $\mathcal{L}_{og}$ (LOG) losses of the proposed method across three settings, with 4, 10, and 25 labeled samples per class. All ablation studies are conducted with the class-adaptive re-weighting (CAW) module integrated as the foundation. The datasets are CIFAR-10 and CIFAR-100, which comprise a balanced labeled dataset (imbalance ratio $\gamma_l = 1$) of only known classes and an imbalanced unlabeled dataset (imbalance ratio $\gamma_u = 100$) of both known and unknown classes.

Module				CIFAR-10			CIFAR-100		
LIL	LOL	LIG	LOG	lb4	lb10	lb25	lb4	lb10	lb25
✓				50.61	67.75	76.52	29.18	43.12	52.47
	✓			43.35	67.71	76.63	28.77	42.44	52.98
✓	✓			50.80	68.05	76.95	30.24	43.25	53.49
✓	✓	✓		52.16	67.29	76.68	29.81	42.25	52.55
✓	✓		✓	49.33	67.26	76.34	28.65	43.20	53.20
✓	✓	✓	✓	54.30	68.54	76.88	30.25	43.43	53.30

F Compute Resources

- CPU: AMD EPYC 7642 48-Core Processor × 2
- GPU: NVIDIA GeForce RTX 4090 24G × 1
- MEM: 500G
- Maximum total computing time: training + testing ≈ 20h