Maximum Spanning Tree Guided Confidence and Sparse Graph for Robust Noisy Label Learning

Gengfeng Chen[1], Xu Liu[1], Boyi Peng[1], Liangqiu Xiao[1], Weicheng Xie[1,2,4⋆], Siyang Song[5], Zitong Yu[6], Laizhong Cui[1,2], and Linlin Shen[3,4]

[1] School of Computer Science & Software Engineering, Shenzhen University, Shenzhen, China
[2] Guangdong Laboratory of Artificial Intelligence and Digital Economy (SZ), Shenzhen, China
[3] School of Artificial Intelligence, Shenzhen University, Shenzhen, China
[4] Guangdong Provincial Key Laboratory of Intelligent Information Processing, Shenzhen University, Shenzhen, China
[5] University of Leicester, Leicester, United Kingdom
[6] Great Bay University, Dongguan, China

Abstract. DNNs are highly susceptible to noisy labels, often leading to unstable training and severe performance degradation. Graph-based methods have emerged as a promising solution, utilizing sample relationships to refine labels through neighborhood aggregation. However, most existing approaches rely on dense connectivity (e.g., KNN graphs), which tends to aggregate noisy cues and cause oversmoothing. While sparse graphs offer a potential remedy, current sparse topologies often suffer from blind error propagation between samples, particularly for hard samples near class boundaries. To address these challenges, we propose Maximum Spanning Tree (MST) Guided Confidence and Sparse Graph Network (MST-GCSN). We introduce the MST as a robust sparse backbone to filter out redundant local noise while preserving a global manifold skeleton. To rectify the indiscriminate mutual influence between nodes, a Confidence-Gated Propagation (CGP) mechanism is designed to adaptively regulate information flow based on node reliability, ensuring that only high-confidence semantic signals are propagated. Building on this optimized structure, a Progressive Label Prediction (PLP) module integrates local information and global anchor-guided cues to iteratively correct labels. Extensive experiments on synthetic and real-world noisy datasets demonstrate that MST-GCSN significantly enhances robustness.

Keywords: Noisy labels · Image classification · Label correction · MST

1 Introduction

Deep neural networks (DNNs) have demonstrated superior achievements in many computer vision tasks, such as image classification [20,35]. The outstanding performance of DNNs is largely attributed to supervised training using large-scale,

⋆ Corresponding author.

high-quality, human-labeled datasets, such as ImageNet [8]. However, collecting large-scale datasets with precise annotations is both costly and time-consuming, particularly for tasks that require expert knowledge, such as medical image annotation [51]. To address this issue, researchers have turned to alternative methods, including crowd-sourcing platforms [49] and web search engines [11], to obtain more affordable label annotations. Unfortunately, these methods often lead to unavoidable noisy labels, which can degrade model performance due to the strong learning capabilities of DNNs [59]. Therefore, developing robust models that can effectively learn from noisy labels is of critical importance.

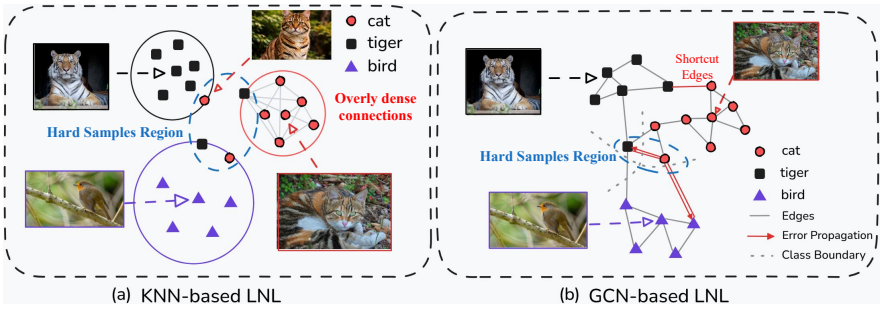

Fig. 1: Illustration of the challenges in existing feature-space methods for Learning with Noisy Labels (LNL). (a) KNN-based LNL struggles with unreliable feature similarity, causing cross-class information propagation among hard samples near class boundaries. Additionally, fixed k-neighbor graphs result in overly dense connections, which amplify local noise. (b) GCN-based LNL constructs suboptimal graph topologies containing shortcut edges that incorrectly bridge unrelated clusters (e.g., tigers and cats). Furthermore, it lacks node reliability modeling, allowing low confidence nodes (near class boundary) to propagate incorrect information to neighborhoods (Error Propagation).

Various methods [2, 3, 29, 47, 55] have been proposed to address noisy labels, which mainly fall into three categories: label correction [7, 12, 37, 53], sample selection [13, 19, 43, 57], and sample reweighting [39, 56]. Label correction methods commonly aim to rectify labels using the noise transition matrix [12,37] or model predictions [28,56]. Sample reweighting [38,51,52] is deemed as a variant of sample selection, where a soft weighting scheme replaces the binary 0/1 assignment by assigning greater weights to higher-confidence samples. Sample selection aims to divide training data into clean and noisy subsets. Loss-based methods [13, 28] leverage the small-loss assumption to identify clean samples, though they often struggle when high noise levels cause loss distributions to overlap. Alternatively, feature-based selection methods [10, 23, 24, 50] identify clean samples via representation similarity.

Despite their progress, these feature-space methods, particularly KNN-based ones [10, 17, 18, 50], still encounter two key challenges: learning from dense and unreliable similarity signals during training, as illustrated in Fig. 1(a).

Specifically, (i) unreliable neighborhood similarity: KNN-based methods rely on raw, sample-wise neighbor information derived purely from feature proximity. However, such sample-wise personalized information may not reflect true semantic relationships, especially for class boundary or hard samples, leading to locally coherent but globally inconsistent neighborhoods. As illustrated in Fig. 1(a), such samples may be incorrectly grouped with visually similar but semantically different neighbors. (ii) Overly dense neighborhood connectivity: Fixed k-neighbor graphs tend to produce overly dense local connectivity, increasing the risk of propagating noisy information and amplifying label corruption in the neighborhood. As shown by the densely connected region in Fig. 1(a), these connections may link unreliable neighbors together, which amplifies noisy signals and propagates errors within the neighborhood.

To better capture higher-order relationships beyond raw pairwise similarity while encouraging a more sparse and structured topology, graph-based modeling with message propagation, such as Graph Convolutional Networks (GCNs) [6], provides a promising direction. However, as illustrated in Fig. 1(b), existing GCN-based paradigms still encounter two challenges: (i) Suboptimal graph topology construction: the graph structure used by GCNs is typically constructed by connecting samples with their nearest neighbors based on feature similarity. Such locally greedy connections focus only on the most similar nearby samples and do not consider the global similarity structure of the data. As a result, the constructed graph may introduce shortcut edges that incorrectly connect samples from different semantic clusters, as illustrated in Fig. 1(b). (ii) Lack of node reliability modeling: GCN propagation typically assumes equal reliability for all nodes during message passing. However, samples near class boundaries are often less reliable due to ambiguous semantics or noisy labels. As illustrated in Fig. 1(b), such boundary samples may still participate in propagation with the same importance as reliable ones, allowing incorrect information to spread along the graph and accumulate through multiple propagation steps.

Facing these challenges, we observe that Maximum Spanning Tree (MST) avoids redundant local connections and reduces shortcut edges by selecting only the most informative edges that maximize global similarity. Based on this observation, we introduce the MST structure into the noisy label learning task and consequently propose **MST-GCSN**. Specifically, (i) to address the **suboptimal graph topology construction**, we build an MST over feature similarities to preserve informative connections while reducing shortcut edges between different semantic clusters. (ii) To address the lack of node reliability modeling, we propose a dual-stage propagation framework. First, Confidence-Gated Propagation (CGP) regulates feature-level information flow by suppressing uncertain nodes based on their prediction confidence. Second, Progressive Label Prediction (PLP) rectifies pseudo-labels by adaptively fusing individual, local, and global cues, where the contribution of each cue is dynamically weighted by its prediction certainty. Finally, an entropy-based confidence weight is applied to the training objective to prevent uncertain labels from dominating the optimization.

Accordingly, the key contributions of **MST-GCSN** are summarized as:

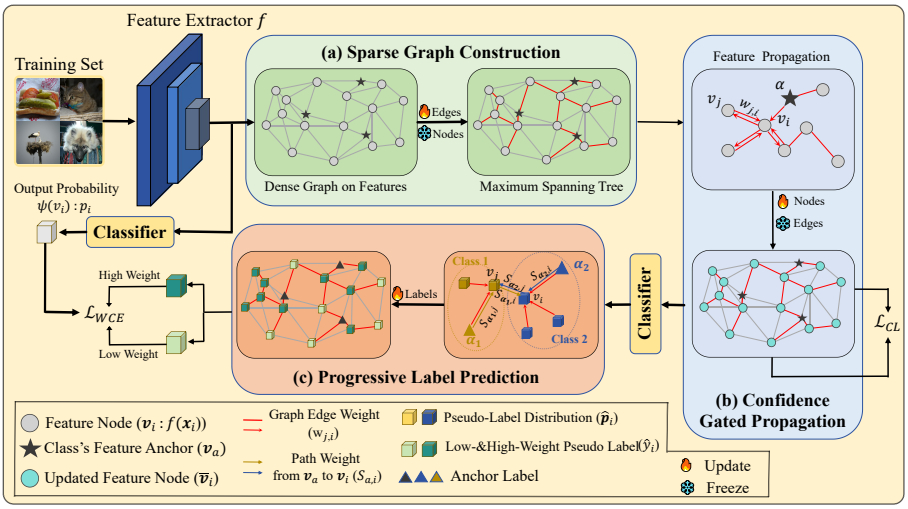

Fig. 2: The framework of our MST-GCSN. An MST graph is first constructed for each batch in the training dataset, where both the samples and the class feature anchors serve as nodes, and the joint similarity that incorporates both feature similarity and class probability distribution is used as edge weight. Further, feature state propagation between connected neighbors is introduced to refine the node representations. These updated features are then passed through a classifier to produce class probability distributions. Based on this, the Progressive Label Prediction is proposed by aggregating cues from neighboring nodes and class anchors, where each sample is subsequently assigned an adaptive weight based on the entropy of its probability distribution.

– Sparse Graph Construction (**SGC**): A dynamic MST-based graph is constructed to selectively preserve semantically reliable connections while pruning noisy edges, enabling boundary and hard samples to receive stable global guidance from high-confidence anchors.

– Confidence Gated Propagation (**CGP**): A confidence-aware gating mechanism is proposed to modulate feature propagation along MST edges, amplifying signals from reliable (high-confidence) regions and suppressing the influence of uncertain or noisy neighbors.

– Progressive Label Prediction (**PLP**): Our PLP incorporates individual, local and global label information while adjusting multi-scale signal weights based on prediction confidence, ensuring the generation of robust pseudo-labels.

– Experiments on both synthetic and real-world benchmarks demonstrate that our framework consistently outperforms state-of-the-art methods across various noise types.

2 Related Work

Researchers have explored numerous robust training strategies for learning with noisy labels. Existing LNL methods can be broadly classified into two main categories: sample selection [13,21,34]/re-weighting [44,52], label correction [41, 55,58]. Currently, the methods achieving superior performance typically leverage both selected clean labels and corrected noisy labels for training.

Sample Selection/Re-weighting. A straightforward approach for handling noisy labels is to distinguish clean samples from noisy ones. Previous sample selection methods typically treat low-loss or high-confidence samples as clean. For example, DivideMix [28] identifies the clean subset by modeling the loss distribution using a Gaussian Mixture Model (GMM). FINE [24] proposes a sample selection method based on eigenvector decomposition, where clean samples tend to be closer to the class-dominant eigenvector of the latent representations compared to noisy-label samples. Sample reweighting [9,40] can be seen as a variant of sample selection, where instead of applying a binary $0/1$ decision to include or exclude samples, a more gradual and adaptive weighting scheme is employed to reflect the reliability of each training instance. These methods assign larger weights to samples more likely to be clean, thereby mitigating the impact of noisy samples. For example, SED [39] dynamically adjusts sample weights based on a normal distribution, assigning higher weights to high-confidence samples and lower weights to low-confidence ones. However, these approaches are inherently limited by their reliance on individual sample statistics, such as loss or confidence, without considering the underlying structural relationships among samples. This often results in unreliable selection or weighting, especially for hard samples that exhibit ambiguous features or lie near decision boundaries.

Label Correction. Another intuitive approach to address noisy labels is to correct the labels of noisy samples before training the model and then use the corrected samples for model training [12, 31, 53]. Early studies, e.g. Dual-T [57] aim to correct training labels by estimating the noise transition matrix. Patrini et al. [37] introduce an additional layer to estimate the noise transition matrix. However, accurately estimating the transition matrix remains challenging in real-world scenarios. Thus, other approaches attempt to correct noisy labels by leveraging the predictions of DNNs [26, 45, 46]. PENCIL [58] introduces an end-to-end approach to directly learn label distributions for corrupted samples. Nevertheless, DNNs tend to accumulate label correction errors [13] during iterative training, where incorrect pseudo-labels are repeatedly reinforced as supervision signals. Furthermore, most existing methods ignore the structural dependencies among training samples. Specifically, they neglect both local personalized information and global guidance from high-confidence samples, which are essential for robust label refinement.

3 Method

To avoid confusion, lowercase letters denote scalars, while bold lowercase letters denote vectors.

Problem Setup. Formally, we consider a K-class ($K \geq 2$) classification problem. Given a training dataset with noisy labels $\mathcal{D}_{train} = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^{n}$, where $\boldsymbol{x}_i$ denotes the i-th sample, $y_i \in \{1, ..., K\}$ is its label which might be incorrect and n is the number of samples. In a standard supervised learning paradigm, a Deep Neural Network (DNN) typically consists of a feature extractor $f(\cdot)$ and a classifier including a softmax layer $h(\cdot)$. Given a training set $\mathcal{D}_{train}$, the model is trained to minimize the empirical risk through the cross-entropy (CE) loss:

$$\mathcal{L}_{CE} = \frac{1}{n} \sum_{i=1}^{n} \ell(\boldsymbol{p}_i, y_i), \tag{1}$$

in which $\boldsymbol{p}_i$ is the predicted probability vector of $\boldsymbol{x}_i$, ℓ denotes the cross entropy loss function. Due to the memory effect [1] of DNNs, networks tend to first fit clean samples before gradually overfitting noisy labels, which ultimately deteriorates generalization performance. Our goal is to train a robust feature extractor $f(\cdot)$ and a classifier $h(\cdot)$ on the training set $\mathcal{D}_{train}$ with noisy labels to perform prediction on the test set $\mathcal{D}_{test}$ with accurate labels.

The procedure of our method is illustrated in Fig. 2. We provide the technical details of our method as follows.

3.1 Sparse Graph Construction

First, we perform a warm-up stage on the training set $\mathcal{D}_{train}$ for T_w epochs to establish basic feature representations for our model training.

To address the dense and unreliable connectivity of KNN graphs, we introduce an MST-based approach. By integrating global semantic anchors with the MST algorithm, we extract a sparse yet robust skeleton of the data manifold. This topology preserves critical semantic connections while eliminating redundant shortcut edges that propagate label noise.

Graph nodes and similarity estimation. The graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is defined over a node set $\mathcal{V}$ that incorporates both batch samples and global class anchors. Given an input batch $\mathcal{B} = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^{B}$, the node set is constructed as $\mathcal{V} = \{f(\boldsymbol{x}_i)\}_{i=1}^{B} \cup \mathcal{A}$, resulting in $n_G = B + K$ nodes. Here, $\mathcal{A} = \{\boldsymbol{\alpha}_c\}_{c=1}^{K}$ denotes the set of global class anchors which serve as stable semantic landmarks. We define $\mathcal{I}_X = \{1, \ldots, B\}$ and $\mathcal{I}_A = \{B+1, \ldots, B+K\}$ as the indices for sample nodes and anchor nodes, respectively. To accurately model the relationships between nodes $\boldsymbol{v}_i$ and $\boldsymbol{v}_j$, we compute a fully connected similarity matrix $\boldsymbol{E} \in \mathbb{R}^{n_G \times n_G}$ as:

$$\boldsymbol{E}_{i,j} = exp(-\frac{\|\boldsymbol{v}_i - \boldsymbol{v}_j\|_2^2}{2\sigma^2}) \cdot (1 - JS(\boldsymbol{p}_i || \boldsymbol{p}_j), \tag{2}$$

where $\boldsymbol{v}_i$ denotes the feature representation of node i, which corresponds to either $f(\boldsymbol{x}_i)$ for sample nodes or $\boldsymbol{\alpha}_c$ for anchor nodes. σ is the standard deviation of all distances in each mini-batch, and $JS(\cdot||\cdot)$ denotes the Jensen–Shannon divergence.

Although $\boldsymbol{E}$ encodes comprehensive pairwise relationships, a fully connected graph introduces significant computational overhead and noisy shortcut edges.

To preserve only the most reliable semantic connections, we employ the Bernard Chazelle algorithm [4] to construct a Maximum Spanning Tree $\mathcal{G}_m = (\mathcal{V}, \mathcal{E}_m)$ based on $\boldsymbol{E}$. By maximizing the total edge weight, the MST yields a sparse backbone that connects all nodes via their strongest mutual similarities, which effectively filters out redundant local connections.

Global Anchor Update. To ensure the graph remains semantically grounded despite label noise, we maintain and update global class anchors $\mathcal{A}$ throughout the training process. (i) **Initialization**: At the start of training, we leverage the DNN's early memorization effect to initialize anchors $\boldsymbol{\alpha}_c$ via probability-weighted aggregation: $\boldsymbol{\alpha}_c = \sum_{\boldsymbol{x}_i \in \mathcal{I}_c} p_{i,c} \cdot f(\boldsymbol{x}_i) / \sum_{\boldsymbol{x}_i \in \mathcal{I}_c} p_{i,c}$, where $\mathcal{I}_c = \{\boldsymbol{x}_i \mid \arg\max \boldsymbol{p}_i = c\}$ and $p_{i,c}$ denotes the predicted softmax probability of the $\boldsymbol{x}_i$ over its c-th class. (ii) **Iterative Update**: As training progresses, we shift from probability-based initialization to structure-consistent refinement. A node is considered a confident representative of its class if its predicted label is consistent with all its neighbors. We update the anchor feature by averaging these structurally consistent nodes: $\boldsymbol{\alpha}_k = \frac{1}{|\mathcal{I}_k|} \sum_{i \in \mathcal{I}_k} \boldsymbol{v}_i$. If no such nodes are found, the previous anchor is retained to maintain stability.

3.2 Confidence-Guided Propagation and Label Prediction

To mitigate the influence of noisy mutual-sample interaction due to the neglect of node reliability, we perform a dual-stage propagation process over the constructed MST based on the node confidence. This process first refines node features through confidence-gated aggregation and then predicts pseudo labels by fusing local and global structural cues.

Node Confidence Estimation. We first quantify the reliability of each node using prediction uncertainty. Given the softmax probability $\boldsymbol{p}_i = h(\boldsymbol{v}_i)$ from the DNN, its entropy is defined as $\mathcal{H}(\boldsymbol{p}_i) = -\sum_{k=1}^{K} p_{i,k} \log p_{i,k}$. To obtain a normalized confidence score, we define it as $r_i = 1 - \mathcal{H}(\boldsymbol{p}_i)/\log K$, such that $r_i \in [0, 1]$, where nodes with more confident predictions obtain larger weights in subsequent propagation.

Confidence-Gated Propagation (CGP). To suppress noisy personalized information while encouraging intra-class compactness, we propagate states along the MST edges. The aggregation weight from node $\boldsymbol{v}_j$ to node $\boldsymbol{v}_i$ is regulated by the neighbor's confidence:

$$w_{j,i} = \frac{r_j \cdot E_{i,j}}{\sum_{k \in \mathcal{N}(i)} r_k \cdot E_{i,k}}, \tag{3}$$

where $\mathcal{N}_i$ denotes the neighbors of node $\boldsymbol{v}_i$. With the aggregation weight $w_{j,i}$, the node feature $\boldsymbol{v}_i$ is then updated as:

$$\overline{\boldsymbol{v}}_i = \gamma_i \cdot \boldsymbol{v}_i + (1 - \gamma_i) \cdot \sum_{j \in \mathcal{N}(i)} w_{j,i} \cdot \boldsymbol{v}_j, \tag{4}$$

where $\gamma_i = r_i/(r_i + \overline{r}_i)$ and $\overline{r}_i$ is the average confidence of neighbors. This mechanism ensures that confident nodes retain their original features, while uncertain nodes rely more on their reliable neighbors.

Progressive Label Prediction (PLP). After obtaining the refined features $\overline{\boldsymbol{v}}_i$, we further predict pseudo labels by fusing node-level predictions, personalized local neighborhood information and high-confidence global class-level cues. This fusion strategy ensures that labels not only preserve the inherent characteristics of the samples themselves but also align with the immediate neighbors and the overall class structure.

Specifically, since the refined feature $\overline{\boldsymbol{v}}_i$ incorporates filtered neighborhood information, its direct prediction $\overline{\boldsymbol{p}}_i$ provides a strong baseline. To further enhance robustness, we compute a local class probability estimate p_i^l by aggregating the predictions of neighboring nodes as:

$$\boldsymbol{p}_i^l = \sum_{\boldsymbol{v}_j \in \mathcal{N}(i)} w_{j,i} \cdot \overline{\boldsymbol{p}}_j, \tag{5}$$

where $\overline{\boldsymbol{p}}_j$ denotes the predicted softmax probability of $\overline{\boldsymbol{v}}_j$, and $w_{j,i}$ denotes the propagation weight from $\boldsymbol{v}_j$ to $\boldsymbol{v}_i$ in Eq. 3. This local estimation leverages the homophily property of the MST, where connected nodes are likely to share the same semantic label.

While local estimates handle cluster-wise consistency, they may still be biased if an entire local neighborhood is corrupted. To counteract this, we incorporate global semantic guidance by propagating labels from high-confidence class anchors. Unlike KNN-based methods that only consider immediate neighbors, we utilize the tree-structured topology of the MST to find the most reliable propagation paths from reliable class anchors. The global estimate $\boldsymbol{p}_i^g$ is defined as:

$$\boldsymbol{p}_i^g = \sum_{a \in \mathcal{I}_A} \left(\frac{\prod_{(u,v) \in \text{path}(a \to i)} E_{u,v}}{\sum_{o \in \mathcal{I}_A} \prod_{(u,v) \in \text{path}(o \to i)} E_{u,v}} \right) \cdot \boldsymbol{y}_a, \tag{6}$$

where $\mathcal{I}_A$ denotes the set of indices of anchor nodes and $\boldsymbol{y}_a$ is the one-hot class indicator vector associated with anchor node $\boldsymbol{v}_a$, determined by the specific class prototype. u and v are the indices of nodes along the path from anchor node $\boldsymbol{v}_a$ to sample node $\boldsymbol{v}_i$. By aggregating cues along the maximal similarity path, it naturally suppresses the influence of anchors that are semantically distant, ensuring that each sample is guided by its most relevant global prototype.

To balance these two perspectives, we first merge the local and global estimates into a unified distribution $\tilde{\boldsymbol{p}}_i$:

$$\tilde{\boldsymbol{p}}_i = \lambda \cdot \boldsymbol{p}_i^l + (1 - \lambda) \cdot \boldsymbol{p}_i^g, \tag{7}$$

where λ is a hyperparameter (set to 0.5) to control the contribution between the local estimates $\boldsymbol{p}_i^l$ and global estimates $\boldsymbol{p}_i^g$.

Finally, the refined pseudo-label distribution $\hat{\boldsymbol{p}}_i$ is generated by:

$$\hat{\boldsymbol{p}}_i = \rho_i \cdot \overline{\boldsymbol{p}}_i + (1 - \rho_i) \cdot \tilde{\boldsymbol{p}}_i, \tag{8}$$

where the fusion weight $\rho_i = \max(\overline{\boldsymbol{p}}_i)/(\max(\overline{\boldsymbol{p}}_i) + \max(\tilde{\boldsymbol{p}}_i))$ adaptively assigns more weight to the more certain distribution. This multi-level fusion effectively filters out noise at different scales, predicting reliable pseudo-labels for subsequent supervised learning.

3.3 Overall Framework

Training Loss. To supervise the model with the refined pseudo-labels while accounting for their varying reliability, we introduce an entropy-based reweighting mechanism. This adjusts each sample's influence based on its prediction confidence to prevent uncertain labels from dominating the optimization. Specifically, for each sample $\boldsymbol{x}_i$ with refined pseudo label $\hat{y}_i = \arg\max(\hat{\boldsymbol{p}}_i)$, we compute a normalized confidence weight $\hat{r}_i = 1 - \mathcal{H}(\hat{\boldsymbol{p}}_i)/\log K$ and define the weighted cross-entropy loss as:

$$\mathcal{L}_{WCE} = \frac{1}{n}\sum_{i=1}^{n} \hat{r}_i \cdot \ell(\boldsymbol{p}_i, \hat{y}_i), \tag{9}$$

where $\boldsymbol{p}_i$ is the predicted probability vector for $\boldsymbol{x}_i$.

Furthermore, to enhance representation discriminability, we incorporate a Supervised Contrastive Loss $\mathcal{L}_{CL}$ [22]. Leveraging the refined pseudo-labels $\hat{y}$ and enhanced features $\overline{\boldsymbol{v}}$, this objective pulls together positive pairs (samples with identical pseudo-labels) and repels negative pairs within the mini-batch, effectively regularizing the manifold structure captured by the MST.

Finally, the overall training objective integrates all components as:

$$\mathcal{L} = \mathcal{L}_{WCE} + \mathcal{L}_{CL}. \tag{10}$$

The complete training procedure of our method is summarized in Alg. 1.

Algorithm 1 Training of the proposed MST-GCSN

Require: Training dataset $\mathcal{D}_{train}$, feature extractor f, classifier h, warm-up epochs T_w, max iterations $T_{\max}$,

Warm-up: learn initial features and establish class anchors

for $t = 1$ **to** T_w **do**
  Pre-train feature extractor f and classifier h using Eq. (1) on $\mathcal{D}_{train}$.
  Initialize class anchors α.
end for

Main training: graph construction, feature propagation, label prediction, and model optimization

for $t = T_w + 1$ **to** $T_{\max}$ **do**
  Extract deep representations $\boldsymbol{v}_i = f(\boldsymbol{x}_i)$ for batch $\mathcal{B}$.
  Compute pairwise edge weights via Eq. (2) to form a fully connected graph.
  Construct the MST using Chazelle's MST algorithm.
  Refine graph node features using Eq. (3) and Eq. (4).
  Perform progressive label refinement via Eq. (5)-Eq. (8).
  Optimize f and h using Eq. (10).
  Update class anchors α.
end for

Output: Optimized feature extractor f and classifier h.

4 Experiment

4.1 Experimental Setup

Following previous studies [29, 32, 42, 48], we validate our method on three standard LNL benchmark datasets: CIFAR-10 [25], CIFAR-100 [25], and WebVision-50 [30]. For the CIFAR benchmarks, we adopt PreAct ResNet-18 [14] as the backbone network. All models are optimized using SGD with momentum 0.9 and weight decay 10^{-4}, with a mini-batch size of 256. We train for 300 epochs in total and employ a 5-epoch warm-up stage. We set the initial learning rate as 0.1, and reduce it by a factor of 10 after 100, 200 and 250 epochs, respectively. For WebVision-50 benchmark, we use the standard ResNet-18 as the backbone network, and train it using SGD with momentum 0.9 and weight decay 10^{-4}, with a mini-batch size of 128. The network is trained for 150 epochs in total with a 20-epoch warm-up stage. We set the initial learning rate as 0.1, and reduce it by a factor of 10 after 70, 100 and 130 epochs.

Table 1: Comparison with SOTA methods in terms of testing accuracy (%) on two datasets with synthetic noises. The best and 2nd best results are marked in **bold** and underline, respectively. **Avg** denotes the average accuracy across different noise settings for each dataset, and **Overall Avg** is the overall mean across both datasets.

Dataset	**CIFAR-10**						**CIFAR-100**					**Overall Avg**
Noisy Type	Sym				Asym	Avg	Sym				Avg	
Noise Ratio	20%	50%	80%	90%	40%		20%	50%	80%	90%		
Cross-Entropy	82.7	57.9	26.1	16.8	76.0	51.9	61.8	37.3	8.8	3.5	27.9	39.9
P-correction [58]$_{(\text{CVPR'19})}$	92.0	88.7	76.5	58.2	91.6	81.4	68.1	56.4	20.7	8.8	38.5	60.0
ELR [32]$_{(\text{NeurIPS'20})}$	95.8	94.8	93.3	78.7	93.0	91.1	77.6	73.6	60.8	33.4	61.4	76.3
DivideMix [28]$_{(\text{ICLR'20})}$	95.0	93.7	92.4	74.2	91.4	89.3	74.8	72.1	57.6	29.2	58.4	73.9
MOIT [36]$_{(\text{CVPR'21})}$	93.1	90.0	79.0	69.6	92.0	84.7	73.0	64.6	46.5	36.0	55.0	69.9
TCL [16]$_{(\text{CVPR'23})}$	95.0	93.9	92.5	89.4	92.6	92.7	78.0	73.3	65.0	54.5	67.7	80.2
PLM [62]$_{(\text{CVPR'24})}$	93.8	92.6	88.0	63.3	85.3	84.6	74.5	70.2	45.2	20.5	52.6	68.6
UNICON+HMW [61]$_{(\text{AAAI'24})}$	92.7	95.1	93.6	90.1	93.2	92.9	74.0	73.6	62.2	42.3	63.0	77.9
DULC [54]$_{(\text{AAAI'25})}$	**96.6**	96.0	**95.0**	**93.5**	95.2	**95.3**	**79.4**	76.4	67.7	52.8	69.1	82.2
PLReMix [33]$_{(\text{WACV'25})}$	96.4	95.3	94.8	92.4	94.7	94.7	77.8	**77.3**	68.4	49.4	68.2	81.5
Our MST-GCSN	**96.6**	**96.2**	94.7	93.2	**95.4**	95.2	78.4	76.7	**68.7**	**56.3**	**70.0**	**82.6**

4.2 Results on Synthetic Noisy Datasets

We conduct experiments on two synthetic noisy datasets, i.e. CIFAR-10 and CIFAR-100. Following [27, 28], we introduce two types of label noise, i.e., symmetric and asymmetric. Symmetric noise is introduced by randomly replacing the labels of a specified proportion of the training data with any possible class label. In this paper, we consider symmetric noise levels of 20%, 50%, 80%, and 90% for both datasets. In contrast, asymmetric noise is introduced through label flipping, where specific classes are misclassified into semantically similar categories

(e.g., "truck" → "automobile"). In this study, we apply asymmetric noise at a rate of 40% across both datasets.

Tab. 1 shows the comparison results with different SOTA methods under various noise types and noise rates, where all baselines use the PreAct ResNet-18 network. Notably, MST-GCSN achieves the best overall performance, obtaining the highest Overall Avg of 82.6% across both datasets. In particular, it achieves the best results on CIFAR-100 at high noise ratios and maintains strong performance under asymmetric noise, demonstrating its robustness against severe label corruption and its effectiveness across different noise types.

4.3 Results on Real-World Noisy Datasets

Our method is validated on a large-scale dataset with real-world noisy labels, i.e. WebVision [30], consisting of 2.4 million images collected through web crawling based on the 1,000 categories in ImageNet ILSVRC12 [25]. Following prior studies [13, 15, 28], we utilize only the first 50 classes from the Google Images subset to construct our training set.

As shown in Tab. 2, we further show the performance of our method on the real-world noise dataset WebVision-50. Our method achieves the best top-1 accuracy and competitive top-5 performance on WebVision-50 and ILSVRC12 compared with other state-of-the-art (SOTA) methods, revealing its superiority in handling real-world noisy label scenarios.

Table 2: Comparison with SOTA methods on WebVision and ILSVRC12. The best and 2nd best results are marked in **bold** and underline.

Dataset	WebVision		ILSVRC12	
Method	top-1	top-5	top-1	top-5
Forward [37]$_{(CVPR'17)}$	61.1	82.6	57.3	82.3
Co-teaching [13]$_{(NeurIPS'18)}$	63.5	85.2	61.4	84.7
Iterative-CV [5]$_{(ICML'19)}$	65.2	85.3	61.6	84.9
DivideMix [28]$_{(ICLR'20)}$	77.3	91.6	75.2	90.8
TCL [16]$_{(CVPR'23)}$	79.1	92.3	75.4	92.4
DPC [63]$_{(AAAI'24)}$	79.2	93.0	75.8	92.5
MTaDCS [15]$_{(ECCV'24)}$	79.1	92.4	76.2	92.6
DULC [54]$_{(AAAI'25)}$	<u>79.9</u>	**93.7**	<u>76.9</u>	<u>93.9</u>
Our MST-GCSN	**80.3**	<u>93.5</u>	**77.2**	**94.0**

Table 3: Ablation study of different components in our method. The baseline model is trained with Mixup [60] and $\mathcal{L}_{CE}$. CGP, PLP, L_{WCE}, and $\mathcal{L}_{\text{CL}}$ denote different components of our method.

Dataset						**CIFAR-10**		**CIFAR-100**	
Noise Type						Sym		Asym	Sym
Noise Rate						50%	90%	40%	50%
CGP	PLP(p_i^l)	PLP(p_i^g)	$\mathcal{L}_{WCE}$	$\mathcal{L}_{CL}$		-	-	-	-
						57.9	18.3	77.6	39.4
✓						84.1	63.1	85.7	64.0
✓	✓					90.5	83.6	90.1	66.8
✓		✓				89.3	81.9	88.7	66.2
✓	✓	✓				92.8	87.3	92.1	70.1
✓	✓	✓	✓			94.2	90.9	94.0	73.5
✓	✓	✓	✓	✓		96.2	93.2	95.4	76.7

4.4 Analysis of the Reliability of the MST

The graph structure fundamentally determines how information is propagated and how noisy labels are corrected in our framework. Therefore, it is crucial to examine whether the proposed maximum spanning tree (MST) construction provides reliable and robust connectivity under label noise. In this section, we

analyze the reliability of the MST and compare it with other graphs, which are commonly adopted in LNL methods.

Robustness and Parameter-free Property. As illustrated in Fig. 3(a), a major limitation of KNN is its sensitivity to the hyperparameter k. The optimal k value often varies significantly across different datasets and noise ratios. In contrast, our MST-based approach is parameter-free. It non-parametrically extracts the $N-1$ most significant edges that satisfy the connectivity of the entire data manifold. Empirical results in Fig. 3(a) demonstrate that MST consistently achieves a lower unreliable edge ratio (7.10%) compared to various kNN configurations (around 10%) when evaluated on all samples. Furthermore, the average weight of reliable edges (those connecting nodes with the same ground-truth label) is 0.82, which is significantly higher than that of unreliable edges (0.52), i.e., inter-class connections that link nodes from different semantic categories.

Preservation of Core Correlation Paths. The structural advantage of MST over local methods is demonstrated in Fig. 3(b). Traditional approaches like kNN and GCN are highly susceptible to unreliable local connectivity, where samples near decision boundaries (e.g., tigers vs. cats) are erroneously linked via noisy shortcut edges. In contrast, the proposed MST identifies the intrinsic skeleton of the data by capturing both local and global manifold structures. MST prioritizes high-similarity paths that link local clusters to global representative prototypes (stars), effectively bypassing high-density noise regions. Our method yields significantly sharper and more accurate class probability estimations compared to the ambiguous distributions produced by standard neighborhood aggregation.

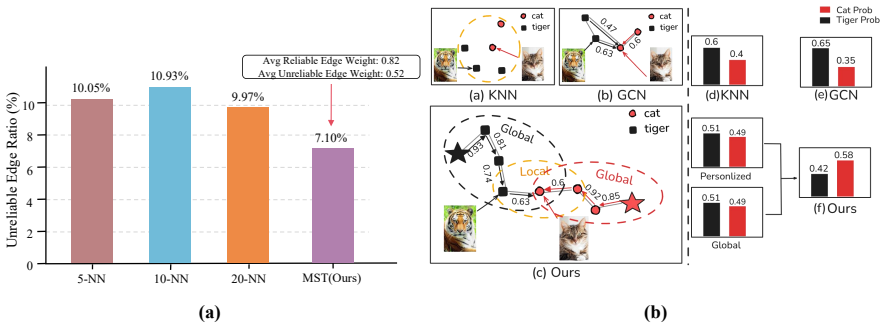

Fig. 3: Comparison of graph reliability and manifold structural modeling. (a) Quantitative analysis of the unreliable edge ratio across different KNN configurations and MST. (b) Conceptual illustration of local vs. global graph construction and its impact on class probability estimation.

4.5 Analysis of the Class Anchors

To evaluate the effectiveness of anchor extraction, we visualize the feature distribution and anchor prediction stability throughout training.

Fig. 4(a)-(b) present t-SNE visualizations under 20% and 90% noise. Despite extreme corruption, our method identifies representative class anchors (stars) at the high-density centers of semantic clusters. Fig. 4(c)-(d) track the Self-Probability of these anchors. Across different noise levels, anchor confidence converges rapidly and remains stable above 0.98 throughout training. This confirms that our mechanism successfully identifies clean prototypical samples. By propagating information from these high-confidence anchors, the framework effectively avoids local optima driven by label noise.

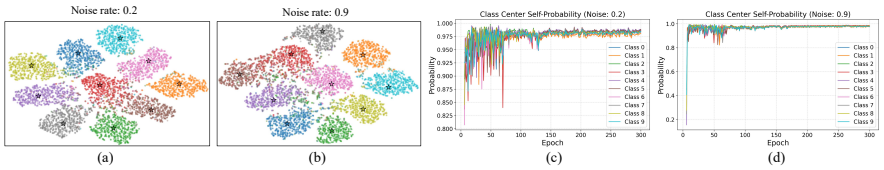

Fig. 4: Visualization of Class Anchors. (a)-(b) t-SNE plots of features under 20% and 90% noise rates, where stars represent the extracted global class anchors. (c)-(d) Evolution of anchor self-probability during training. The anchors maintain high and stable confidence even under extreme noise, providing reliable global guidance for label refinement.

4.6 Ablation Studies and Analysis

To evaluate the contribution of each component in our method, we conduct a comprehensive ablation study on the CIFAR-10 and CIFAR-100 datasets under various noise settings, as shown in Tab. 3.

Effectiveness of Confidence Gated Propagation. The most significant improvement comes from combining the MST-based topology with the CGP mechanism. For example, on CIFAR-10 with 90% symmetric noise, accuracy increases from 18.3% to 63.1%. This indicates that reliability-aware information filtering effectively prevents noisy samples from corrupting the representation space. By regulating feature propagation along the MST, CGP amplifies only high-confidence semantic signals. Visualizations in the Appendix further confirm that this propagation results in tighter intra-class clusters and more distinct class boundaries.

Effectiveness of Progressive Label Prediction. The ablation results highlight the complementarity between local and global structural cues in PLP. The local estimate ($\boldsymbol{p}_i^l$) and the anchor-guided estimate ($\boldsymbol{p}_i^g$) each provide notable improvements, while their combination consistently achieves the best performance, e.g., 92.8% on CIFAR-10 (50% Sym). This confirms that local smoothness

preserves cluster consistency, whereas global anchors provide reliable correction in heavily corrupted regions.

Effectiveness of Loss function. Tab. 3 and Fig. 5 demonstrate the importance of $\mathcal{L}_{WCE}$ for robust training. Specifically, it improves accuracy from 87.3% to 90.9% on CIFAR-10 (90% noise) and from 70.1% to 73.5% on CIFAR-100 (50% noise). As shown in Fig. 5, weights of correctly classified samples gradually approach 1.0, while incorrect ones remain below 0.4, reducing the influence of unreliable labels. Adding the supervised contrastive loss ($\mathcal{L}_{CL}$) further enhances discriminability, yielding the best result of 96.2% on CIFAR-10 (50% Sym).

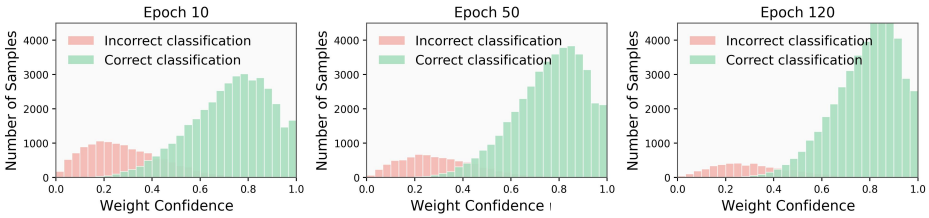

Fig. 5: Visualization of confidence evolution during training. As epochs progress, the number and confidence ($\hat{r}_i$ in Eq. 9) of correctly classified samples (green) steadily increases, while those of incorrectly classified samples (red) decreases.

5 Conclusion and Limitation

In this work, we reveal two key limitations of existing noisy label learning methods: their reliance on sample-wise personalized information and their neglect of class-level cues that govern semantic relationships. To address these limitations, we leverage the advantages of the maximum spanning tree (MST) to construct a sparse and robust topology over sample features. The proposed framework effectively explores both sample-wise and class-level relationships through a dynamically adapted MST, thereby achieving globally consistent connectivity. Furthermore, the Progressive Label Prediction (PLP) module adaptively integrates sample-wise personalized cues and class-level guidance to iteratively correct noisy labels during training.

However, our method has several limitations that warrant further investigation: (1) Sensitivity to update frequency: The framework uses a fixed period to update class anchors and prototypes. However, a static frequency may be suboptimal: rapid updates can cause early-stage instability, while slow updates may delay label correction. Implementing an adaptive scheduling mechanism could further enhance learning efficiency across diverse datasets. (2) Computational overhead: Constructing the MST for every training batch incurs non-negligible computational cost, especially on large-scale datasets (e.g., full ImageNet [8]), which may hinder scalability in real-time or resource-constrained scenarios.

6 Acknowledgements

This work was supported by the National Natural Science Foundation of China under grant nos. 62276170, 62576216, 62576076, 62306061, the Open Research Fund from Guangdong Laboratory of Artificial Intelligence and Digital Economy (SZ) under grant no. GML-KF-24-11, the Guangdong Provincial Key Laboratory under grant no. 2023B1212060076. This work was also supported by the Intelligent Computing Center of Shenzhen University.

References

1. Arpit, D., Jastrzębski, S., Ballas, N., Krueger, D., Bengio, E., Kanwal, M.S., Maharaj, T., Fischer, A., Courville, A., Bengio, Y., et al.: A closer look at memorization in deep networks. In: International conference on machine learning. pp. 233–242. PMLR (2017)
2. Bai, Y., Yang, E., Han, B., Yang, Y., Li, J., Mao, Y., Niu, G., Liu, T.: Understanding and improving early stopping for learning with noisy labels. Advances in Neural Information Processing Systems **34**, 24392–24403 (2021)
3. Berthon, A., Han, B., Niu, G., Liu, T., Sugiyama, M.: Confidence scores make instance-dependent label-noise learning possible. In: International conference on machine learning. pp. 825–836. PMLR (2021)
4. Chazelle, B.: A minimum spanning tree algorithm with inverse-ackermann type complexity. Journal of the ACM (JACM) **47**(6), 1028–1047 (2000)
5. Chen, P., Liao, B.B., Chen, G., Zhang, S.: Understanding and utilizing deep neural networks trained with noisy labels. In: International conference on machine learning. pp. 1062–1070. PMLR (2019)
6. Chen, T., Pu, T., Wu, H., Xie, Y., Liu, L., Lin, L.: Cross-domain facial expression recognition: A unified evaluation benchmark and adversarial graph learning. IEEE transactions on pattern analysis and machine intelligence **44**(12), 9887–9903 (2021)
7. Cheng, D., Liu, T., Ning, Y., Wang, N., Han, B., Niu, G., Gao, X., Sugiyama, M.: Instance-dependent label-noise learning with manifold-regularized transition matrix estimation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 16630–16639 (2022)
8. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. pp. 248–255. Ieee (2009)
9. Fang, T., Lu, N., Niu, G., Sugiyama, M.: Rethinking importance weighting for deep learning under distribution shift. Advances in neural information processing systems **33**, 11996–12007 (2020)
10. Feng, C., Tzimiropoulos, G., Patras, I.: SSR: an efficient and robust framework for learning with unknown label noise. In: British Machine Vision Conference. p. 372 (2022)
11. Fergus, R., Fei-Fei, L., Perona, P., Zisserman, A.: Learning object categories from internet image searches. Proceedings of the IEEE **98**(8), 1453–1466 (2010)
12. Goldberger, J., Ben-Reuven, E.: Training deep neural-networks using a noise adaptation layer. In: International conference on learning representations (2017)
13. Han, B., Yao, Q., Yu, X., Niu, G., Xu, M., Hu, W., Tsang, I., Sugiyama, M.: Co-teaching: Robust training of deep neural networks with extremely noisy labels. Advances in neural information processing systems **31** (2018)

14. He, K., Zhang, X., Ren, S., Sun, J.: Identity mappings in deep residual networks. In: Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part IV 14. pp. 630–645. Springer (2016)
15. Huang, Q., He, X., Xian, X., Lin, Q., Xie, W., Song, S., Shen, L., Yu, Z.: Mtadcs: Moving trace and feature density-based confidence sample selection under label noise. In: European Conference on Computer Vision. pp. 178–195. Springer (2024)
16. Huang, Z., Zhang, J., Shan, H.: Twin contrastive learning with noisy labels. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11661–11670 (2023)
17. Iscen, A., Tolias, G., Avrithis, Y., Chum, O.: Label propagation for deep semi-supervised learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 5070–5079 (2019)
18. Jiang, H., Chen, Y., Liu, L., Han, X., Zhang, X.P.: Leveraging noisy labels of nearest neighbors for label correction and sample selection. In: ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 7720–7724. IEEE (2024)
19. Jiang, L., Zhou, Z., Leung, T., Li, L.J., Fei-Fei, L.: Mentornet: Learning data-driven curriculum for very deep neural networks on corrupted labels. In: International conference on machine learning. pp. 2304–2313. PMLR (2018)
20. Jiang, X., Liu, S., Dai, X., Hu, G., Huang, X., Yao, Y., Xie, G.S., Shao, L.: Deep metric learning based on meta-mining strategy with semiglobal information. IEEE Transactions on Neural Networks and Learning Systems **35**(4), 5103–5116 (2022)
21. Karim, N., Rizve, M.N., Rahnavard, N., Mian, A., Shah, M.: Unicon: Combating label noise through uniform selection and contrastive learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 9676–9686 (2022)
22. Khosla, P., Teterwak, P., Wang, C., Sarna, A., Tian, Y., Isola, P., Maschinot, A., Liu, C., Krishnan, D.: Supervised contrastive learning. Advances in neural information processing systems **33**, 18661–18673 (2020)
23. Kim, N.r., Lee, J.S., Lee, J.H.: Learning with structural labels for learning with noisy labels. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 27610–27620 (2024)
24. Kim, T., Ko, J., Choi, J., Yun, S.Y., et al.: Fine samples for learning with noisy labels. Advances in Neural Information Processing Systems **34**, 24137–24149 (2021)
25. Krizhevsky, A., Hinton, G., et al.: Learning multiple layers of features from tiny images (2009)
26. Lee, K.H., He, X., Zhang, L., Yang, L.: Cleannet: Transfer learning for scalable image classifier training with label noise. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 5447–5456 (2018)
27. Li, J., Li, G., Liu, F., Yu, Y.: Neighborhood collective estimation for noisy label identification and correction. In: European Conference on Computer Vision. pp. 128–145. Springer (2022)
28. Li, J., Socher, R., Hoi, S.C.: Dividemix: Learning with noisy labels as semi-supervised learning. In: International Conference on Learning Representations (2020)
29. Li, S., Xia, X., Ge, S., Liu, T.: Selective-supervised contrastive learning with noisy labels. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 316–325 (2022)
30. Li, W., Wang, L., Li, W., Agustsson, E., Van Gool, L.: Webvision database: Visual learning and understanding from web data. arXiv preprint arXiv:1708.02862 (2017)

31. Liu, H., Zhang, C., Yao, Y., Wei, X.S., Shen, F., Tang, Z., Zhang, J.: Exploiting web images for fine-grained visual recognition by eliminating open-set noise and utilizing hard examples. IEEE Transactions on Multimedia **24**, 546–557 (2021)
32. Liu, S., Niles-Weed, J., Razavian, N., Fernandez-Granda, C.: Early-learning regularization prevents memorization of noisy labels. Advances in neural information processing systems **33**, 20331–20342 (2020)
33. Liu, X., Zhou, B., Yue, Z., Cheng, C.: Plremix: Combating noisy labels with pseudo-label relaxed contrastive representation learning. In: 2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). pp. 6517–6527. IEEE (2025)
34. Malach, E., Shalev-Shwartz, S.: Decoupling" when to update" from" how to update". Advances in neural information processing systems **30** (2017)
35. Mao, J., Yao, Y., Sun, Z., Huang, X., Shen, F., Shen, H.T.: Attention map guided transformer pruning for occluded person re-identification on edge device. IEEE Transactions on Multimedia **25**, 1592–1599 (2023)
36. Ortego, D., Arazo, E., Albert, P., O'Connor, N.E., McGuinness, K.: Multi-objective interpolation training for robustness to label noise. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 6606–6615 (2021)
37. Patrini, G., Rozza, A., Krishna Menon, A., Nock, R., Qu, L.: Making deep neural networks robust to label noise: A loss correction approach. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1944–1952 (2017)
38. Ren, M., Zeng, W., Yang, B., Urtasun, R.: Learning to reweight examples for robust deep learning. In: International conference on machine learning. pp. 4334–4343. PMLR (2018)
39. Sheng, M., Sun, Z., Chen, T., Pang, S., Wang, Y., Yao, Y.: Foster adaptivity and balance in learning with noisy labels. In: European Conference on Computer Vision. pp. 217–235. Springer (2024)
40. Shu, J., Xie, Q., Yi, L., Zhao, Q., Zhou, S., Xu, Z., Meng, D.: Meta-weight-net: Learning an explicit mapping for sample weighting. Advances in neural information processing systems **32** (2019)
41. Song, H., Kim, M., Lee, J.G.: Selfie: Refurbishing unclean samples for robust deep learning. In: International conference on machine learning. pp. 5907–5915. PMLR (2019)
42. Sun, H., Guo, C., Wei, Q., Han, Z., Yin, Y.: Learning to rectify for robust learning with noisy labels. Pattern Recognition **124**, 108467 (2022)
43. Sun, Z., Hua, X.S., Yao, Y., Wei, X.S., Hu, G., Zhang, J.: Crssc: salvage reusable samples from noisy data for robust learning. In: Proceedings of the 28th ACM international conference on multimedia. pp. 92–101 (2020)
44. Tu, Y., Zhang, B., Li, Y., Liu, L., Li, J., Wang, Y., Wang, C., Zhao, C.R.: Learning from noisy labels with decoupled meta label purifier. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 19934–19943 (2023)
45. Vahdat, A.: Toward robustness against label noise in training deep discriminative neural networks. Advances in neural information processing systems **30** (2017)
46. Veit, A., Alldrin, N., Chechik, G., Krasin, I., Gupta, A., Belongie, S.: Learning from noisy large-scale datasets with minimal supervision. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 839–847 (2017)
47. Wei, J., Liu, H., Liu, T., Niu, G., Sugiyama, M., Liu, Y.: To smooth or not? when label smoothing meets noisy labels. In: International Conference on Machine Learning (2022)

48. Wei, Q., Sun, H., Lu, X., Yin, Y.: Self-filtering: A noise-aware sample selection for label noise with confidence penalization. In: European Conference on Computer Vision. pp. 516–532. Springer (2022)
49. Welinder, P., Branson, S., Perona, P., Belongie, S.: The multidimensional wisdom of crowds. Advances in neural information processing systems **23** (2010)
50. Wu, P., Zheng, S., Goswami, M., Metaxas, D., Chen, C.: A topological filter for learning with label noise. Advances in neural information processing systems **33**, 21382–21393 (2020)
51. Wu, T., Dai, B., Chen, S., Qu, Y., Xie, Y.: Meta segmentation network for ultra-resolution medical images. In: Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence. pp. 544–550 (2021)
52. Xia, X., Liu, T., Han, B., Gong, M., Yu, J., Niu, G., Sugiyama, M.: Sample selection with uncertainty of losses for learning with noisy labels. In: International Conference on Learning Representations (2022)
53. Xia, X., Liu, T., Wang, N., Han, B., Gong, C., Niu, G., Sugiyama, M.: Are anchor points really indispensable in label-noise learning? Advances in neural information processing systems **32** (2019)
54. Xu, Y., Niu, X., Yang, J., Su, R., Zhang, J., Liu, S., Drew, S.: Revisiting interpolation for noisy label correction. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 39, pp. 21833–21841 (2025)
55. Yang, E., Yao, D., Liu, T., Deng, C.: Mutual quantization for cross-modal search with noisy labels. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7551–7560 (2022)
56. Yao, Y., Sun, Z., Zhang, C., Shen, F., Wu, Q., Zhang, J., Tang, Z.: Jo-src: A contrastive approach for combating noisy labels. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 5192–5201 (2021)
57. Yao, Y., Liu, T., Han, B., Gong, M., Deng, J., Niu, G., Sugiyama, M.: Dual t: Reducing estimation error for transition matrix in label-noise learning. Advances in neural information processing systems **33**, 7260–7271 (2020)
58. Yi, K., Wu, J.: Probabilistic end-to-end noise correction for learning with noisy labels. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 7017–7025 (2019)
59. Zhang, C., Bengio, S., Hardt, M., Recht, B., Vinyals, O.: Understanding deep learning requires rethinking generalization. In: International Conference on Learning Representations (2017)
60. Zhang, H., Cisse, M., Dauphin, Y.N., Lopez-Paz, D.: mixup: Beyond empirical risk minimization. In: International Conference on Learning Representations (2018)
61. Zhang, S., Li, Y., Wang, Z., Li, J., Liu, C.: Learning with noisy labels using hyperspherical margin weighting. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 38, pp. 16848–16856 (2024)
62. Zhao, R., Shi, B., Ruan, J., Pan, T., Dong, B.: Estimating noisy class posterior with part-level labels for noisy label learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 22809–22819 (2024)
63. Zong, C.C., Wang, Y.W., Xie, M.K., Huang, S.J.: Dirichlet-based prediction calibration for learning with noisy labels. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 38, pp. 17254–17262 (2024)