MTaDCS: Moving Trace and Feature Density-based Confidence Sample Selection under Label Noise

Qingzheng Huang[1⋆], Xilin He[1⋆], Xiaole Xian[1], Qinliang Lin[1], Weicheng Xie[1,2,3†], Siyang Song[4], Linlin Shen[1,2,3], and Zitong Yu[5]

[1] Computer Vision Institute, School of Computer Science & Software Engineering, Shenzhen University
[2] Shenzhen Institute of Artificial Intelligence and Robotics for Society
[3] Guangdong Key Laboratory of Intelligent Information Processing
[4] University of Leicester
[5] Great Bay University

Abstract. Learning from noisy labels is a challenging task, as noisy labels can compromise decision boundaries and result in suboptimal generalization performance. Most previous approaches for dealing noisy labels are based on sample selection, which utilized the small loss criterion to reduce the adverse effects of noisy labels. Nevertheless, they encounter a critical limitation in being unable to effectively separate challenging samples from those that were merely mislabeled. Meanwhile, there is a lack of researches on the trace changes of samples during training. To this end, we propose a novel moving trace and feature density-based confidence sample selection strategy (called MTaDCS). Different from existing small loss-based approaches, the local feature density of samples in the latent space is explored to construct a confidence set by selectively choosing confident samples in a progressive manner in terms of moving trace. Therefore, our MTaDCS can gradually isolate noisy labels through the setting of confidence set and achieve the goal of learning discriminative features from hard samples. Extensive experiments conducted on datasets with simulated and real-world noises validate that the proposed MTaDCS outperforms the state-of-the-art methods in terms of various metrics. The code is available at https://github.com/QZ-CODER/-ECCV-24-MTaDCS.

Keywords: Noisy labels · Image classification · Sample selection · Feature density · Moving trace

1 Introduction

Deep Neural Networks (DNNs) have shown great proficiency in various real-world tasks [15, 32, 46]. This proficiency primarily results from the availability

⋆ First two authors contributed equally.
† Corresponding author: wcxie@szu.edu.cn

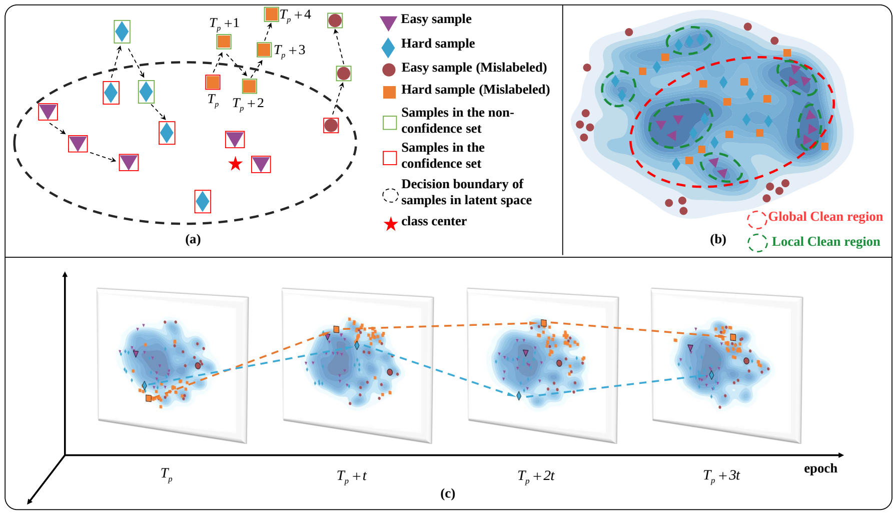

Fig. 1: (a) Different categories of sample movements in latent space: The Local Outlier Factor (LOF) [3] is employed to identify samples with local high-density features which will be added to the confidence set, T_p is a training epoch. The confidence set will be dynamically updated during training stage given different movement traces of the samples. (b) Comparison of the global and local application of LOF-based sample selection method. (c) Moving traces of easy and hard samples in the training process.

of extensive and well-annotated datasets, but acquiring such datasets can be a significant challenge. Alternatively, some large-scale image datasets are obtained with some noisy labels through methods like web crawling, as exemplified by the WebVision [21] dataset. However, noisy labels can harm the generalization capacity of a model. Therefore, given the presence of noisy labels, training the model to maintain generalization performance holds significant importance.

To reduce the potential risk of noisy labels, robust loss functions [33, 38, 52], noise transition matrix [43, 44, 48], sample selection [9, 13, 34, 39], sample weighting [37, 45] and label correction [17, 28] have been widely investigated. Among these strategies, sample selection technique has been proven to be highly effective in combating the issue of noisy labels [12, 40]. The majority of previous sample selection based methods [17, 40] utilize the small loss trick [1] to distinguish noisy labels from clean ones. However, this trick fails to effectively differentiate between hard samples and mislabeled ones as both of them exhibit relatively high classification loss values [1]. As a result, a model may encounter challenge in acquiring the discriminative features of hard samples, which is crucial for enhancing its generalization ability [6]. Meanwhile, the noisy labels will compromise the class decision boundary, resulting in its fuzziness.

As hard samples and noisy labels share a set of similar properties, including relatively high loss values and wandering on the decision boundary during training, previous sample selection methods [12, 24], which frequently apply global selection in the class-wise feature space, fail to distinguish hard samples and noisy labels. As illustrated in Fig. 1(b), global sample selection method merely focuses on globally significant samples and ignores the intra-class features relationship, which leads to a limited search space for sample selection.

To address the above concerns of: **(problem 1)** the dilemma of small loss trick [1] in differentiating between the hard samples and the mislabeled ones; and **(problem 2)** the plight of selection inefficiency by global sample selection method when dealing with hard and mislabeled samples, this paper proposes a moving trace and feature density-based confidence sample selection strategy, namely MTaDCS, based on the feature moving traces (illustrated in Fig. 1). Specifically, we hypothesize that within the class-wise feature space, easy samples, difficult samples, and mislabeled samples have different moving traces relative to the corresponding class center in the latent space when a model is only trained with correctly labeled samples. The class center is obtained from the correctly labeled samples, as shown in Fig. 1(a). To verify this hypothesis, we conduct hypothetical experiments on the CIFAR-10 dataset to guide models using a subset of correctly labeled samples, where Fig. 1(c) validates our hypotheses by providing a t-SNE visualization to track the moving trace of each kind of samples. In the following sections, we further solidify this hypothesis based on the distances between samples and class centers. Motivated from the above hypothesis that different kinds of samples show different movement traces during training, we can use the position of sample features relative to the decision boundary to distinguish hard and noisy labels. This solves the dilemma **(problem 1)** of the small loss trick [1].

Previously, Zhang et al. [51] revealed that mislabeled samples tend to distribute at the edge of the data representations. This statement is further verified by our experiments illustrated in Fig. 1, where both clean hard samples and mislabeled hard ones tend to wander on the decision boundary. Besides, we also observed that clean hard samples exhibit stronger local clustering effect compared with mislabeled hard samples, and form high-density clusters at the decision boundary, while mislabeled hard samples are sparsely distributed. Given this feature density difference between clean hard samples and mislabeled hard samples, we propose to utilize a density-based sample selection method Local Outlier Factor (LOF) [3] to the intra-class feature clusters within the decision boundary to explore clean hard samples for training, which is not well explored in previous global sample selection methods.

However, as shown in Fig. 4, the direct application of LOF for sample selection in each epoch, may result in unstable training as the vast majority of the samples are only selected very few times for training. To reduce this unstability of sample selection and maintain the consistency of leveraging previously selected trust-worthy samples, we construct a confidence set to store the selected samples via the LOF [3], while applying LOF locally to include more samples from non-confidence set for training. Additionally, we refine the confidence set by removing potential noisy labels based on the above-mentioned assumption. The contributions are summarized as three-fold:

- This paper serves as the first work revealing that the moving traces of the distances between samples and the class center can be used to distinguish easy samples, difficult samples and mislabeled samples, which drives us to solve the dilemma of small loss technique [1].
- We propose to adaptively maintain the previously selected certified confidence samples in a confidence set. Resorting to a local feature density manner, our feature density-driven sample selection mechanism largely outperforms global ones in terms of efficiency and effectiveness.
- Extensive experiments on both simulated and real-world datasets with noisy labels show the superiority of our algorithm over the state of the arts in terms of various metrics.

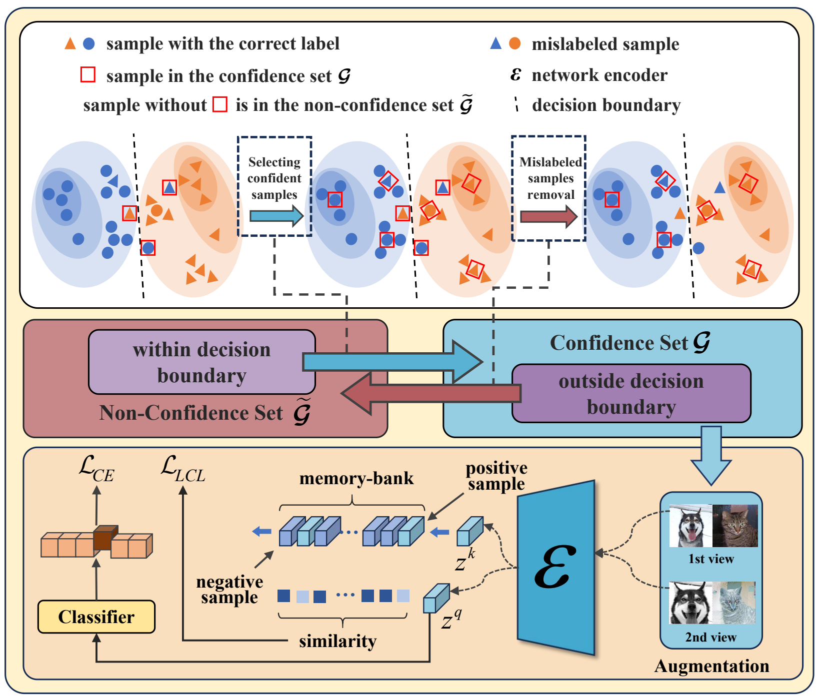

Fig. 2: Illustration of the proposed MTaDCS. After obtaining sample features from the noisy labeled dataset, the confidence set $\mathcal{G}$ is then updated using the method in Section 3.2; then the loss function in Section 3.4 is used to train the model on the confidence set $\mathcal{G}$.

2 Related Work

2.1 Learning with noisy labels

Methods for learning noisy labels can be roughly categorized as robust loss functions [23,36,38,52], noise transition matrix [43,44,48], loss correction [43,44,48], sample selection [2,9,39,49], label correction [25] and sample weighting [16,37,45]. These robust loss functions can provide better generalization performance compared to cross-entropy classification loss and have theoretical guarantees. The transition matrix bridges the class posterior probabilities for noisy labels and clean data. By accurately estimating the transformation matrix, the challenges posed by noisy labels can be effectively mitigated. Nowadays, an increasing amount of researches is dedicated to sample selection and label correction.

Label correction involves an iterative process of correcting noisy labels by leveraging model predictions and noisy labels. For example, Lu et al. [25] achieved notable results by iteratively correcting noisy labels using a self-ensemble approach. Sample selection [7, 53] reduces the adverse effect of noisy labels on a model by selecting confident samples during training, while the validity of the selection criteria is a main concern. Cordeiro et al. [7] employed a small loss criterion to select confident samples in the first stage, and then utilize them to assist model training on the entire noisy dataset in the second stage. In addition, sample weighting methods can be treated as a special form of sample selection [37, 45].

2.2 Contrastive learning

Contrastive learning has strong representation learning ability, which has been verified in many works [5, 10, 42]. MoCo [10] uses a memory queue to store consistent representations. The InfoNCE [29] can bring together data with the same examples for data augmentation and push away other negative examples. However, when noisy labels are present, using supervised contrast learning may instead degrade the generalization performance of a model. Therefore, many works [12,51] use unsupervised contrastive learning to enhance the representation but some works [20] use supervised contrastive learning after selecting confident samples, and we adopt a similar paradigm to the latter.

3 The Proposed Mothod

3.1 Preliminaries

In this sequel, lowercase letters represent scalars while lowercase boldface letters represent vectors. Let $\mathbb{I}[A]$ be the indicator of the event A, while $\mathcal{B}$ represents the symbol for a set.

Problem setup. We consider a K-class ($K \geq 2$) classification problem. Given a noisy labeled dataset $\mathcal{S} = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^{n}$, where n is the number of samples; $\boldsymbol{x}_i$ denotes the i-th sample; and $y_i \in [1, \ldots, K]$ is the corresponding label which might be incorrect. The prediction of $\boldsymbol{x}_i$ is denoted as $\hat{y}_i$; $f(\cdot)$ denotes the output representation of a backbone plus a linear layer. Meanwhile, we define a confidence set $\mathcal{G}$ and a non-confidence set $\tilde{\mathcal{G}}$, where $\mathcal{G} \cup \tilde{\mathcal{G}} = \mathcal{S}$ and $\mathcal{G} \cap \tilde{\mathcal{G}} = \emptyset$. $\mathcal{D}_c$ ($c = 1, ..., K$) represents a subset of $\mathcal{G}$ and all samples in subset $\mathcal{D}_c$ have the same label. $\tilde{\mathcal{D}}_c$ denotes a subset of $\tilde{\mathcal{G}}$ and all samples in subset $\tilde{\mathcal{D}}_c$ have the same label. A small number of manually selected samples are used to initialize the confidence set on the simulated dataset, whereas in the real scenario the confidence set is initialized using the entire dataset. T_P represents the number of pre-training epochs on the initialized confidence set.

Local outlier factor (LOF). The LOF [3] algorithm excels in accurately identifying outliers, particularly for datasets that exhibit varying densities across different regions. It detects outliers by calculating the local reachable density (LRD) of each data point, as well as the LOF of this point from the local reachable density. The outlier factor identifies the outlier degree of a data point. The larger the factor value, the higher the outlier degree. The LRD and LOF of data points are calculated as follows:

$$\begin{aligned} LRD_m(p) &= \frac{|N_m(p)|}{\sum_{o\in N_m(p)} dist_m(o,p)} \\ LOF_m(p) &= \frac{\sum_{o\in N_m(p)} LRD_m(o)}{|N_m(p)|\cdot LRD_m(p)} \end{aligned} \tag{1}$$

where the domain index $m = 10$ is a hyperparameter; $N_m(p)$ denotes the set of points within a circle centered at p, and the radius is equal to the distance from p to its m-th nearest neighbor, $|N_m(p)|$ denotes the number of points in the set $N_m(p)$ and $dist_m(o, p)$ is the m-th reachable distance from the point o to the point p. The specific details and explanations of this LOF can be found in the supplementary material.

Overview. The proposed method consists of two stages: (1) selecting confidence samples from the $\tilde{\mathcal{G}}$ according to the moving trace characteristic of each kind of samples and adding them to the $\mathcal{G}$, while removing the samples with possible noisy labels from the $\mathcal{G}$ to refine it; and (2) representation enhancement by local contrastive learning on $\mathcal{G}$. The procedure of the method is illustrated in Fig. 2. We provide the technical details of our method as follows.

3.2 Progressive hierarchical sample selection

Firstly, a preliminary training phase is conducted with the initialized confidence set $\mathcal{G}$ for T_P epochs, which enables the target model to develop a rough decision boundary. Subsequently, we train our model using the proposed MTaDCS strategy.

Selecting confident samples. By treating the feature representation of each sample as a point, we apply Eq. (1) in the latent space after each epoch training to pick confidence samples from the $\tilde{\mathcal{G}}$. Different from previous studies, we pick confidence samples that need to satisfy two conditions: 1) after each training epoch, the prediction labels of the samples are consistent with their ground truth labels; and 2) the samples have low outlier values. Specifically, we use the 1st condition to filter out easily distinguishable noisy labels to obtain $\tilde{\mathcal{P}}_c$ from $\tilde{\mathcal{D}}_c$ ($c = 1, ..., K$), which isolates most of the noisy labels that are easier to be distinguished. Furthermore, we utilize the 2nd condition to select confidence samples with higher local feature densities from $\tilde{\mathcal{P}}_c$. We then transfer the confidence samples selected from the $\tilde{\mathcal{G}}$ into the $\mathcal{G}$ as:

$$\begin{aligned} &\tilde{\mathcal{P}}_c = \left\{ f(\boldsymbol{x}_i) \mid \hat{y}_i = y_i, (\boldsymbol{x}_i, y_i) \in \tilde{\mathcal{D}}_c \right\} \\ &\mathcal{Q} = \bigcup_{c=1}^{K} LOF_m \left(\tilde{\mathcal{P}}_c \right) \\ &\mathcal{G} := \mathcal{G} \cup \mathcal{Q} \quad \text{and} \quad \tilde{\mathcal{G}} := \tilde{\mathcal{G}} - \mathcal{Q} \end{aligned} \tag{2}$$

where $\mathcal{Q}$ denotes the set of confidence samples selected using Eq. (2) from the non-confidence set $\tilde{\mathcal{G}}$. This confidence sample selection will be elaborated in the supplementary material.

Mislabeled samples removal. The decision boundaries are usually not optimal in the early stages of training, which is likely to result in the inclusion of samples with noisy labels into $\mathcal{G}$ during the process of sample selection. To remove noisy labels to refine the confidence set $\mathcal{G}$, we adopt the update as:

$$\begin{aligned} &\mathcal{T} = \{(\boldsymbol{x}_i, y_i) \mid \hat{y}_i \neq y_i, (\boldsymbol{x}_i, y_i) \in \mathcal{G}\} \\ &\mathcal{G} := \mathcal{G} - \mathcal{T} \ \text{ and } \ \tilde{\mathcal{G}} := \tilde{\mathcal{G}} \cup \mathcal{T} \end{aligned} \tag{3}$$

Specifically, samples with high density of local features are acquired as confidence samples via Eq. (2) after each training epoch, which are then incorporated into $\mathcal{G}$. Simultaneously, Eq. (3) is utilized to dynamically remove possible mislabeled samples $\mathcal{T}$ from the confidence set. In this way, we can continuously expand the confidence set while effectively isolating noisy labels.

3.3 Local Contrastive Learning

Since high-density feature clustering is required for LOF to select high-quality confidence samples from $\tilde{\mathcal{G}}$, it would be natural to attempt to craft a more compact feature space. Specifically, we extend the popular method of MoCo [10] to local contrastive learning and apply it within $\mathcal{G}$. Given each input $\boldsymbol{x}_i$, its two views ${\boldsymbol{x}_i}^q$ and ${\boldsymbol{x}_i}^k$ are generated by applying two random augmentations, which are then fed to a query encoder and a key encoder, followed by a 1-layer MLP to obtain low-dimensional embeddings $\boldsymbol{z}^q$ and $\boldsymbol{z}^k$, respectively. To enhance representation learning and expedite the aggregation of the feature space, the extended MoCo [10] also maintains a large memory bank. To keep this memory bank as consistent as possible, a key encoder uses a momentum update with a query encoder. Formally, our local supervised contrastive loss is summarized as:

$$\begin{aligned} &\mathcal{L}_{Pos} = \textstyle\sum_{j=1}^{L} \mathbb{I}\left[y_j^k = y^q\right] \exp\left(\boldsymbol{z}^q \cdot \boldsymbol{z}_j^k / \tau\right) \\ &\mathcal{L}_{Neg} = \textstyle\sum_{j=1}^{L} \mathbb{I}\left[y_j^k \neq y^q\right] \exp\left(\boldsymbol{z}^q \cdot \boldsymbol{z}_j^k / \tau\right) \\ &\mathcal{L}_{LCL} = -\tfrac{1}{D} \textstyle\sum_{j=1}^{L} \mathbb{I}\left[y_j^k = y^q\right] \frac{\exp\left(\boldsymbol{z}^q \cdot \boldsymbol{z}_j^k / \tau\right)}{\mathcal{L}_{Pos} + \mathcal{L}_{Neg}} \end{aligned} \tag{4}$$

where the memory bank has a length of L, encompassing keys represented by $z^k = \{\boldsymbol{z}_1^k, \boldsymbol{z}_2^k, \ldots, \boldsymbol{z}_L^k\}$. $\{y_1^k, y_2^k, \ldots, y_L^k\}$ are the labels of these specific keys. y^q is the label of $\boldsymbol{z}^q$. τ denotes a temperature parameter, $\cdot$ denotes dot product and D represents the number of samples with label equal to y^q in the memory bank.

The loss regularization via Eq. (4) on the confidence set $\mathcal{G}$ can largely accelerate the gathering of high-density feature clusters, which is crucial for boosting the accuracy of the LOF sample selection, while isolating the noisy labels.

3.4 Training Loss and Label Correction

Training Loss. In addition to the loss $\mathcal{L}_{LCL}$ regularized locally on the confidence set $\mathcal{G}$, a cross-entropy loss $\mathcal{L}_{CE}$ on this set is also used. Specifically, to improve the robustness of the learned representation, the Mixup operation [50] is integrated into our established framework, where a convex combination of a sample pair is performed as $\boldsymbol{x}_i = \lambda \boldsymbol{x}_{(1)} + (1-\lambda)\boldsymbol{x}_{(2)}$ to acquire the training sample $\boldsymbol{x}_i$, $\boldsymbol{x}_{(1)}$ and $\boldsymbol{x}_{(2)}$, which are two samples selected from the mini-batch and λ obeys to the $Beta$ distribution taking the value from [0,1]. Consequently, the Mixup [50] for enhancing the cross-entropy loss is formulated as:

$$\mathcal{L}_{CE} = \sum_{(\boldsymbol{x}_i, y_i) \in \mathcal{G}} \ell\left(h\left(f(\boldsymbol{x}_i)\right), y_i\right) \tag{5}$$

$$\mathcal{L} = \mathcal{L}_{CE} + \mathcal{L}_{LCL} \tag{6}$$

where $h(f(\boldsymbol{x}_i))$ is the prediction of the output class of $\boldsymbol{x}_i$, $\ell(\cdot)$ denotes the cross entropy loss function. $\mathcal{L}$ is the model training loss.

Label Correction. In the scenarios with a high noise rate of noisy labels, label correction has shown to be particularly effective. Therefore, we incorporate a self-ensemble label correction method [25] to maintain the representative capacity of learned features. Specifically, we first assign a soft label $\boldsymbol{p}_i$ $(i = 1, ..., n)$ for each sample, and then iteratively correct this label using the self-ensemble label correction method [25]. Finally, we modify the original hard label y_i according to the corrected soft label $\boldsymbol{p}_i$, which is formulated as:

$$\boldsymbol{p}_i \leftarrow (1-\alpha)\boldsymbol{p}_i + \alpha \hat{\boldsymbol{y}}_{\boldsymbol{i}} \tag{7}$$

where $\alpha = 0.01$ is the momentum and $\hat{\boldsymbol{y}}_{\boldsymbol{i}}$ is obtained by one-hot encoding of y_i. When a predicted probability specific to the other categories in $\boldsymbol{p}_i$ exceeds γ (γ $= 0.8$ is a hyperparameter), we flip the label y_i to it with the largest prediction value in $\boldsymbol{p}_i$, where samples with the flipped labels belong to the non-confidence set $\tilde{\mathcal{G}}$.

For clarity, the entire algorithm flow is presented in Algorithm 1. Following previous studies [12, 20], we employ fine-tuning after training the model, where the confidence samples selected by our method are used for this fine-tuning.

4 Experiments

4.1 Implementation Details

Simulated noisy datasets. Following previous studies [20, 22, 35, 41], we validate our method on CIFAR-10/100 [14] datasets with labels simulated by two

Algorithm 1: Training of the proposed method MTaDCS

Require: the training dataset $\mathcal{S} = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^{n}$, the confidence set $\mathcal{G}$, the non-confidence set $\tilde{\mathcal{G}}$, the pre-training epochs T_P, maximum epochs T_{max} and $\tau = 0.1$.

for $t = 1, ..., T_p$ **do**
  Pre-train the encoder network f with Eq. (6) on the confidence set $\mathcal{G}$.
end for
for $t = T_p + 1, ..., T_{max}$ **do**
  Extract deep representations of samples with $\boldsymbol{z} = f_{t-1}(\boldsymbol{x})$ on the training dataset $\mathcal{S}$.
  for $c = 1, \ldots, K$ **do**
    Execute Eq. (2) in the unconfidence set $\tilde{\mathcal{G}}$ to obtain confidence samples.
    Execute Eq. (3) in the confidence set $\mathcal{G}$ to refine the confidence set $\mathcal{G}$.
  end for
  Execute Eq. (6) to train the model on the confidence set $\mathcal{G}$.
  Execute Eq. (7) to correct the labels on the $\mathcal{S}$.
end for
Output : the optimized encoder network f.

types of noises: symmetric and asymmetric noises. Symmetric noise is generated by randomly replacing the labels of a certain percentage of the training data with all possible labels, i.e., the percentages of 20%, 50%, 80%, and 90% on two datasets in this paper. Asymmetric noise uses label flipping to mislabel data (*e.g.*, "truck" $\rightarrow$ "automobile"), which includes 40% on two datasets in this paper.

Similar to previous studies [17,19,20,22,30], we use a PreAct ResNet-18 [11] as the encoder. We adopt SGD optimizer to train our model with a momentum of 0.9, a weight decay of 10^{-4}, and a batch size of 128. The model was trained for 250 epochs, including 10 warm-up training epochs. We set the initial learning rate as 0.1, and reduce it by a factor of 10 after 125 and 200 epochs, respectively. The fine-tuning stage of the method has 70 epochs, where the learning rate is 0.001. The MoCo trick [10] is used, and the size of memory-bank is set to 10^4. The scalar temperature $\tau = 0.1$ is used. We use the data augmentation RandAug [8] in the training phase, while a standard weak data augmentation (e.g. horizontal flipping) in the fine-tuning phase.

Real-world noisy datasets. Following previous works [17,20,30], we validate our method on a real-world noisy dataset, i.e., WebVision [21]. We compare baseline methods on the preceding 50 classes of the Google image subset, called WebVision-50. For WebVision-50, we use the standard ResNet-18, and train it using SGD with a momentum of 0.9, a weight decay of 10^{-4}, and a batch size of 64. The network is trained for 130 epochs with the warm-up training having 40 epochs. We set the initial learning rate as 0.1, and reduce it by a factor of 10 after 80 and 105 epochs. The fine-tuning stage has 10 epochs, where the learning rate

Dataset	CIFAR-10						CIFAR-100				
Noise Type	Sym				Asym	Avg	Sym				Avg
Noise Rate	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	27.8
Mixup$_{(\text{ICLR'17})}$ [50]	92.3	77.6	46.7	43.9	77.7	67.6	66.0	46.6	17.6	8.1	34.6
P-correction$_{(\text{CVPR'19})}$ [47]	92.0	88.7	76.5	58.2	91.6	81.4	68.1	56.4	20.7	8.8	38.5
M-correction$_{(\text{ICML'19})}$ [1]	93.8	91.9	86.6	68.7	87.4	85.7	73.4	65.4	47.6	20.5	51.7
ELR$_{(\text{NeurIPS'20})}$ [22]	93.8	92.6	88.0	63.3	85.3	84.6	74.5	70.2	45.2	20.5	52.6
DivideMix$_{(\text{ICLR'20})}$ [17]	95.0	93.7	92.4	74.2	91.4	89.3	74.8	72.1	57.6	29.2	58.4
MOIT$_{(\text{CVPR'21})}$ [30]	93.1	90.0	79.0	69.6	92.0	84.7	73.0	64.6	46.5	36.0	55.0
RRL$_{(\text{ICCV'21})}$ [19]	<u>95.8</u>	<u>94.3</u>	92.4	75.0	91.9	89.8	**79.1**	**74.8**	57.7	29.3	60.2
Sel-CL+$_{(\text{CVPR'22})}$ [20]	95.5	93.9	89.2	81.9	<u>93.4</u>	90.7	76.5	72.4	59.6	48.8	64.3
TCL$_{(\text{CVPR'23})}$ [12]	95.0	93.9	<u>92.5</u>	<u>89.4</u>	92.6	<u>92.7</u>	<u>78.0</u>	73.3	<u>65.0</u>	<u>54.5</u>	<u>67.7</u>
MTaDCS	**95.9**	**95.4**	**92.7**	**89.6**	**94.2**	**93.5**	77.1	<u>74.1</u>	**65.7**	**61.4**	**69.5**

Table 1: Comparison with state-of-the-art methods in terms of the testing accuracy (%) on CIFAR-10 and CIFAR-100. The best and 2nd best results are marked in bold and underlined, respectively.

is 0.001. The same data augmentation techniques [8] and MoCo technique [10] as those on the CIFAR-10/100 datasets are used.

4.2 Results on Simulated Noisy Datasets

Tab. 1 compares the proposed MTaDCS with multiple baselines, where all baselines use the PreAct ResNet-18 network. It can be seen that MTaDCS achieves competitive performance in terms of various indicators compared with recent methods. It is remarkable that our method achieves greater improvement on asymmetric noise datasets than that on symmetric noise datasets. This shows that our method is more suitable for the data with the real-world noises, since the asymmetric noises are common in real-world scenario.

Following previous work [12, 20, 30], we also fine-tune the network after the training stage. Specifically, we performed 70 epochs of fine-tuning using the selected confidence samples with the mixup augmentation. Sec. 4.2 shows that MTaDCS+ outperforms other fine-tuned state-of-the-art (SOTA) methods. We also perform k-NN classification on the learned representations and show that our method retains meaningful representations better than those by a purely unsupervised learning model. Meanwhile, we provide more experimental results and analysis on data with asymmetric noisy labels in the supplementary material.

4.3 Results on the Real-World Noisy Datasets

We validate our method on the real-world dataset WebVision-50 [21], which contains images of noisy labels collected from Flickr and Google. As shown in Sec. 4.2, MTaDCS achieves the best results compared with the other SOTA methods in terms of both top-1 and top-5 accuracies on the validation sets of WebVision and ILSVRC12. Our method improves the top-1 accuracy of ILSVRC12

Dataset	CIFAR-10			CIFAR-100
Noise Type	Sym		Asym	Sym
Noise Rate	20%	50%	40%	50%
DivideMix$_{(\text{ICLR'20})}$ [17]	95.0	93.7	91.4	72.1
MOIT$_{(\text{CVPR'21})}$ [30]	93.1	90.0	92.0	64.6
MOIT+$_{(\text{CVPR'21})}$ [30]	94.1	91.8	93.3	70.6
RRL$_{(\text{ICCV'21})}$ [19]	95.8	94.3	91.9	74.8
Sel-CL+$_{(\text{CVPR'22})}$ [20]	95.5	93.9	93.5	72.4
TCL+$_{(\text{CVPR'23})}$ [12]	96.0	94.5	93.7	74.6
MTaDCS+	**96.2**	**95.5**	**94.4**	**74.8**

Table 2: (a) Comparison of state-of-the-art under fine-tuning. + indicates that the corresponding method has been fine-tuned.

Dataset	WebVision		ILSVRC12	
Methods	top1	top5	top1	top5
Forward$_{(\text{CVPR'17})}$ [31]	61.1	82.6	57.3	82.3
D2L$_{(\text{ICML'18})}$ [26]	62.6	84.0	57.8	81.3
Iterative-CV$_{(\text{ICML'19})}$ [4]	65.2	85.3	61.6	84.9
Decoupling$_{(\text{NeurIPS'17})}$ [27]	62.5	84.7	58.2	82.2
MentorNet$_{(\text{ICML'18})}$ [13]	63.0	81.4	57.8	79.9
Co-teaching$_{(\text{NeurIPS'18})}$ [9]	63.5	85.2	61.4	84.7
ELR$_{(\text{NeurIPS'20})}$ [22]	76.2	91.2	68.7	87.8
DivideMix$_{(\text{ICLR'20})}$ [17]	77.3	91.6	75.2	90.8
RRL$_{(\text{ICCV'21})}$ [19]	76.3	91.5	73.3	91.2
NGC$_{(\text{ICLR'21})}$ [18]	**79.1**	91.8	74.4	91.0
MOIT$_{(\text{CVPR'21})}$ [30]	77.9	91.9	73.8	91.7
TCL$_{(\text{CVPR'23})}$ [12]	**79.1**	92.3	75.4	92.4
MTaDCS	**79.1**	**92.4**	**76.2**	**92.6**

Table 2: (b) Accuracy (%) on the WebVision and ILSVRC2012 validation sets. The model is trained on WebVision-50.

by 1.0% over the recent SOTA method, which demonstrates that our method is better at handling real-world scenarios.

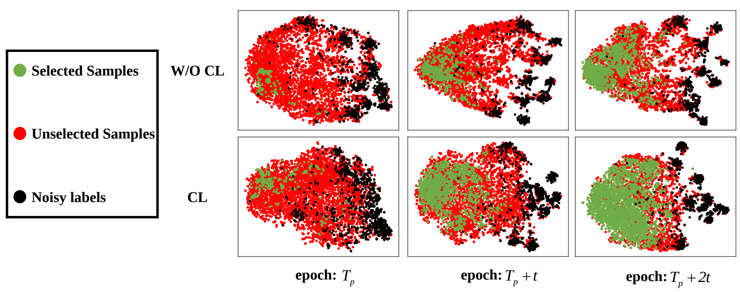

Fig. 3: The upper and lower rows depict the results obtained without (W/O) and with contrastive learning (CL), respectively. Contrastive learning is utilized to enhance sample clustering, thus expediting LOF for selecting confidence samples.

4.4 Mutually Reinforcing Effect of Local LOF-based Sample Selection and Contrastive Learning

In this section, we analyze the mutually reinforcing effect between the LOF-based sample selection applied to the non-confidence set and the contrastive learning applied to the confidence set.

Whether contrastive learning boosts LOF-based sample selection? As shown in Fig. 3, the enhanced clustering effect brought by the employed contrastive learning can promote the formation of more compact and high-density feature clusters, which is beneficial for boosting the LOF-based sample selection and improving its performance. Mutually, the setting of the confidence set can effectively reduce the risk of noisy labels during contrastive learning, i.e., the LOF applied to the non-confidence set can select more high-quality credible samples

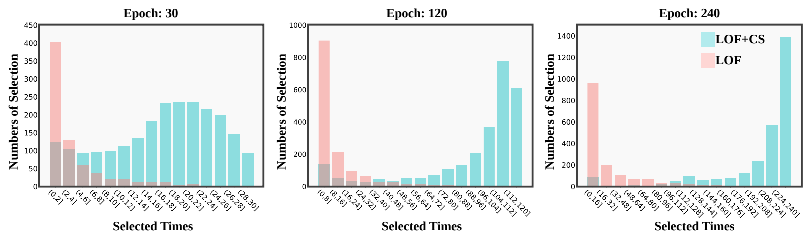

Fig. 4: The counts of samples (y-axis) against the selected times (x-axis) for training till the 30th, 120th and 240th epochs with LOF or LOF + CS, where CS represents the setting of the confidence and non-confidence sets.

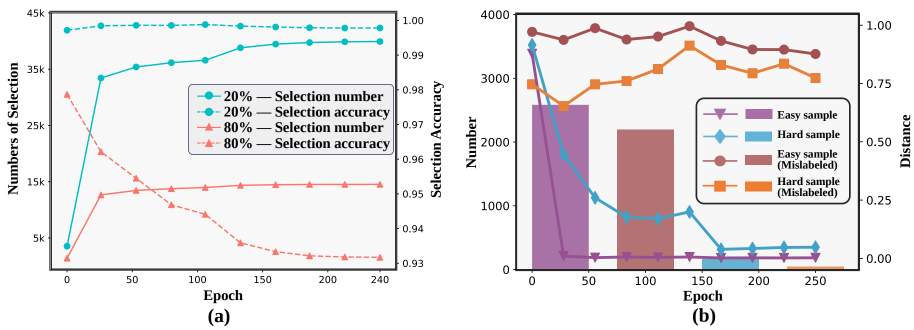

Fig. 5: (a) indicates the number of selected samples in $\mathcal{G}$ and the selection correct ratio in $\mathcal{G}$ of MTaDCS on CIFAR-10 with 20% or 80% symmetric noises (including about 40k or 14k accurately labeled samples). (b) indicates the number of samples of four types under 50% symmetric noise, as well as the distances between the class centers of the four types and the class centers of samples in the confidence set.

into the confidence set for training, thus expanding the number of samples for contrastive learning.

Whether the setting of the confidence set (CS) helps LOF explore wider sample selection? As shown in Fig. 4, the set up of confidence set can maintain the usage of certified confidence samples during training. Particularly, using LOF alone can only obtain easy samples with high feature density at each time, while using LOF + CS can obtain difficult samples with low feature density. Individually applying LOF cannot select hard sample clusters with low feature density. When the confidence set continuously includes locally high-density samples from the non-confidence set, the LOF would be able to select samples that were evolved from the state of high density to non-high density.

4.5 Analysis of the Reliability of the Confidence Set

In this section, we look into the reliability of the confidence set crafted by the proposed local LOF-based sample selection method. We conduct numerical analysis on the number of selected samples and the correct ratio on CIFAR-10 under symmetric noise rates of 20% and 80%. Experiment results are shown in Fig. 5 (a). It shows that the correct ratio can still maintain a high level as the number of selected samples increases, reflecting that MTaDCS can effectively identify correctly labeled samples as training continues.

4.6 Analysis of Hard Samples and Mislabeled Samples

This section explains how the proposed MTaDCS can discern between hard samples and noisy labeled ones. Initially, we introduce two metrics that categorize samples into four groups in terms of learning difficulty and label correctness. Subsequently, we analyze the traces of various samples in the feature space throughout the training process with MTaDCS.

Categorization metric. Samples are divided into clean easy samples, clean hard samples, mislabeled easy samples and mislabeled hard samples based on two metrics. Metric 1: Intuitively, in a robust feature space, clean easy samples are common in the confidence set, and mislabeled easy samples are more likely to be found in the non-confidence set. Therefore, throughout the training process, the appearance times of a sample in a confidence set and a non-confidence set are leveraged to distinguish between clean easy samples and mislabeled easy samples. Specifically, samples with less than 10 occurrences in the non-confidence set are classified as clean easy samples while those with less than 10 occurrences in the confidence set are deemed as mislabeled easy samples. Metric 2: The samples appear more than 10 times in the non-confidence set, but the samples that appear the most in the confidence set are defined as difficult samples. In the same way, noisy difficult samples appear more than 10 times in the confidence set but appear the most in the non-confidence set.

Experiment analysis. According to the aforementioned metrics, the number of samples in each category, as well as the distance between the feature centers of different types of samples and the labeled class centers are shown in Fig. 5 (b). It shows that correctly labeled easy and hard samples tend to move toward the class centers, whereas the mislabeled samples exhibit the opposite behavior. The different traces of correctly labeled samples and the mislabeled ones, regardless of easy or hard samples, demonstrate the effectiveness of MTaDCS in discerning between clean samples and mislabeled ones. Moreover, the different traces between hard samples and mislabeled hard samples (blue line and orange line in Fig. 5 (b), further demonstrate the superiority of MTaDCS in distinguishing between hard and noisy labels.

4.7 Ablation Study and Discussions

In this section, we conduct an ablation study to study the effect of each component in the proposed method.

Effectiveness of LOF-driven sample selection strategy. By solely training on samples selected by the local application of LOF in the class-wise feature space, model performances are largely improved compared to the baseline, demonstrating the effectiveness of the LOF-driven sample selection strategy.

Dataset						**CIFAR-10**			**CIFAR-100**
Noise Type						Sym		Asym	Sym
Noise Rate						50%	90%	40%	50%
LOF	CS	SR	CL	Mixup	LC	-	-	-	-
						59.1	18.3	77.6	39.4
✓						83.2	76.7	81.0	50.5
✓	✓					90.9	62.0	86.1	67.7
✓	✓	✓				93.0	77.3	92.1	68.2
✓	✓	✓			✓	93.2	83.2	92.0	69.3
✓	✓	✓		✓		95.1	88.4	93.5	72.2
✓	✓	✓	✓			93.4	83.2	92.1	69.8
✓	✓	✓	✓	✓	✓	95.4	89.6	94.2	74.1

Table 3: Ablation study of different components in the proposed method. LOF, CS, SR, CL, Mixup and LC denote the LOF-based sample selection method, the setting of the confidence set and non-confidence set, sample removal from the confidence set, contrastive learning, Mixup technique and the label correction, respectively.

Necessity of setting up confidence set $\mathcal{G}$. As shown in Tab. 3, setting up confidence set $\mathcal{G}$ can consistently improve performances. This improvement may be attributed to the fact that setting up $\mathcal{G}$ can optimize the objects on which the LOF algorithm [3] is executed. With confidence set $\mathcal{G}$, LOF can mine hard samples in the non-confidence set, which enables the model to learn features with stronger discrimination capacity.

Sample removal to refine $\mathcal{G}$. As shown in Tab. 3, under scenarios with a high noise rate, the initialization of the confidence set $\mathcal{G}$ may inadvertently incorporate noisy labels into the training set. Thus, it is essential to refine the confidence set $\mathcal{G}$ in training. Compared to merely setting $\mathcal{G}$ (as in the setting LOF+CS), Tab. 3 shows that the refinement process leads to a substantial performance improvement. Specifically, under a noise rate of 90% on CIFAR-10, refining $\mathcal{G}$ achieves a performance improvement of 15.3%.

Mixup. As shown in Tab. 3, the Mixup technique [50] applied on the confidence set is beneficial for learning robust representations.

Label correction. As shown in Tab. 3, by incorporating the label correction technique, the performance is improved in most situations.

5 Conclusion

In this paper, we introduce MTaDCS, a moving trace and feature density-based confidence sample selection strategy for learning in noisy labels. MTaDCS is able to select samples with high-density features in the latent space as confidence set to enhance the robustness of the model. Moreover, it can effectively distinguish hard samples and noisy labels. Extensive experiments on simulated and real-world datasets demonstrate that MTaDCS outperforms existing state-of-the-art methods in various metrics. In particular, our method achieves large improvements on real-world datasets.

6 Acknowledgement

The work was supported by the National Natural Science Foundation of China under grants no. 62276170, 82261138629, 62306061, the Science and Technology Project of Guangdong Province under grants no. 2023A1515011549, 2023A1515010688, the Science and Technology Innovation Commission of Shenzhen under grant no. JCYJ20220531101412030, Guangdong Provincial Key Laboratory under grant no. 2023B1212060076.

References

1. Arazo, E., Ortego, D., Albert, P., O'Connor, N., McGuinness, K.: Unsupervised label noise modeling and loss correction. In: International conference on machine learning. pp. 312–321. PMLR (2019) 2, 3, 4, 10
2. 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) 5
3. Breunig, M.M., Kriegel, H.P., Ng, R.T., Sander, J.: Lof: identifying density-based local outliers. In: Proceedings of the 2000 ACM SIGMOD international conference on Management of data. pp. 93–104 (2000) 2, 3, 6, 14
4. 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) 11
5. Chen, T., Kornblith, S., Norouzi, M., Hinton, G.: A simple framework for contrastive learning of visual representations. In: International conference on machine learning. pp. 1597–1607. PMLR (2020) 5
6. Cordeiro, F.R., Carneiro, G.: A survey on deep learning with noisy labels: How to train your model when you cannot trust on the annotations? In: 2020 33rd SIBGRAPI conference on graphics, patterns and images (SIBGRAPI). pp. 9–16. IEEE (2020) 2
7. Cordeiro, F.R., Sachdeva, R., Belagiannis, V., Reid, I., Carneiro, G.: Longremix: Robust learning with high confidence samples in a noisy label environment. Pattern Recognition **133**, 109013 (2023) 5
8. Cubuk, E.D., Zoph, B., Mane, D., Vasudevan, V., Le, Q.V.: Autoaugment: Learning augmentation strategies from data. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 113–123 (2019) 9, 10
9. 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) 2, 5, 11
10. He, K., Fan, H., Wu, Y., Xie, S., Girshick, R.: Momentum contrast for unsupervised visual representation learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 9729–9738 (2020) 5, 7, 9, 10
11. 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) 9
12. 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) 2, 5, 8, 10, 11

13. 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) 2, 11
14. Krizhevsky, A., Hinton, G., et al.: Learning multiple layers of features from tiny images (2009) 8
15. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems **25** (2012) 1
16. 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) 5
17. Li, J., Socher, R., Hoi, S.C.H.: Dividemix: Learning with noisy labels as semi-supervised learning. In: 8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020 (2020) 2, 9, 10, 11
18. Li, J., Xiong, C., Hoi, S.C.H.: Mopro: Webly supervised learning with momentum prototypes. In: 9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021 (2021) 11
19. Li, J., Xiong, C., Hoi, S.C.: Learning from noisy data with robust representation learning. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 9485–9494 (2021) 9, 10, 11
20. 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) 5, 8, 9, 10, 11
21. 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) 2, 9, 10
22. 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) 8, 9, 10, 11
23. Liu, Y., Guo, H.: Peer loss functions: Learning from noisy labels without knowing noise rates. In: International conference on machine learning. pp. 6226–6236. PMLR (2020) 4
24. Lu, Y., Zhang, Y., Han, B., Cheung, Y.m., Wang, H.: Label-noise learning with intrinsically long-tailed data. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 1369–1378 (2023) 2
25. Lu, Y., He, W.: Selc: Self-ensemble label correction improves learning with noisy labels. In: International Joint Conference on Artificial Intelligence (2022) 5, 8
26. Ma, X., Wang, Y., Houle, M.E., Zhou, S., Erfani, S., Xia, S., Wijewickrema, S., Bailey, J.: Dimensionality-driven learning with noisy labels. In: International Conference on Machine Learning. pp. 3355–3364. PMLR (2018) 11
27. Malach, E., Shalev-Shwartz, S.: Decoupling" when to update" from" how to update". Advances in neural information processing systems **30** (2017) 11
28. Nguyen, D.T., Mummadi, C.K., Ngo, T., Nguyen, T.H.P., Beggel, L., Brox, T.: SELF: learning to filter noisy labels with self-ensembling. In: 8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020 (2020) 2
29. Oord, A.v.d., Li, Y., Vinyals, O.: Representation learning with contrastive predictive coding. arXiv preprint arXiv:1807.03748 (2018) 5
30. 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) 9, 10, 11
31. 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) 11
32. Ren, S., He, K., Girshick, R., Sun, J.: Faster r-cnn: Towards real-time object detection with region proposal networks. Advances in neural information processing systems **28** (2015) 1
33. Song, S., Luo, Y., Tumer, T., Fu, C., Valstar, M., Gunes, H.: Loss relaxation strategy for noisy facial video-based automatic depression recognition. ACM Trans. Comput. Healthcare (mar 2024) 2
34. Song, S., Zhang, S., Schuller, B.W., Shen, L., Valstar, M.: Noise invariant frame selection: A simple method to address the background noise problem for text-independent speaker verification. In: 2018 International Joint Conference on Neural Networks (IJCNN). pp. 1–8 (2018) 2
35. Sun, H., Guo, C., Wei, Q., Han, Z., Yin, Y.: Learning to rectify for robust learning with noisy labels. Pattern Recognition **124**, 108467 (2022) 8
36. Wang, X., Hua, Y., Kodirov, E., Robertson, N.M.: Imae for noise-robust learning: Mean absolute error does not treat examples equally and gradient magnitude's variance matters. arXiv preprint arXiv:1903.12141 (2019) 4
37. Wang, Y., Liu, W., Ma, X., Bailey, J., Zha, H., Song, L., Xia, S.T.: Iterative learning with open-set noisy labels. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 8688–8696 (2018) 2, 5
38. Wang, Y., Ma, X., Chen, Z., Luo, Y., Yi, J., Bailey, J.: Symmetric cross entropy for robust learning with noisy labels. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 322–330 (2019) 2, 4
39. Wei, H., Feng, L., Chen, X., An, B.: Combating noisy labels by agreement: A joint training method with co-regularization. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 13726–13735 (2020) 2, 5
40. Wei, Q., Feng, L., Sun, H., Wang, R., Guo, C., Yin, Y.: Fine-grained classification with noisy labels. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11651–11660 (2023) 2
41. 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) 8
42. Wu, Z., Xiong, Y., Yu, S.X., Lin, D.: Unsupervised feature learning via non-parametric instance discrimination. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3733–3742 (2018) 5
43. Xia, X., Liu, T., Han, B., Wang, N., Gong, M., Liu, H., Niu, G., Tao, D., Sugiyama, M.: Part-dependent label noise: Towards instance-dependent label noise. Advances in Neural Information Processing Systems **33**, 7597–7610 (2020) 2, 4
44. 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) 2, 4
45. Xue, C., Dou, Q., Shi, X., Chen, H., Heng, P.A.: Robust learning at noisy labeled medical images: Applied to skin lesion classification. In: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019). pp. 1280–1283. IEEE (2019) 2, 5

46. Yao, Y., Chen, T., Xie, G.S., Zhang, C., Shen, F., Wu, Q., Tang, Z., Zhang, J.: Non-salient region object mining for weakly supervised semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2623–2632 (2021) 1
47. 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) 10
48. Yong, L., Pi, R., Zhang, W., Xia, X., Gao, J., Zhou, X., Liu, T., Han, B.: A holistic view of label noise transition matrix in deep learning and beyond. In: The Eleventh International Conference on Learning Representations (2022) 2, 4
49. Yu, X., Han, B., Yao, J., Niu, G., Tsang, I., Sugiyama, M.: How does disagreement help generalization against label corruption? In: International Conference on Machine Learning. pp. 7164–7173. PMLR (2019) 5
50. Zhang, H., Cissé, M., Dauphin, Y.N., Lopez-Paz, D.: mixup: Beyond empirical risk minimization. In: 6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings (2018) 8, 10, 14
51. Zhang, M., Zhao, X., Yao, J., Yuan, C., Huang, W.: When noisy labels meet long tail dilemmas: A representation calibration method. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 15890–15900 (2023) 3, 5
52. Zhang, Z., Sabuncu, M.: Generalized cross entropy loss for training deep neural networks with noisy labels. Advances in neural information processing systems **31** (2018) 2, 4
53. Zhang, Z., Chen, W., Fang, C., Li, Z., Chen, L., Lin, L., Li, G.: Rankmatch: Fostering confidence and consistency in learning with noisy labels. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 1644–1654 (2023) 5