RankMatch: Fostering Confidence and Consistency in Learning with Noisy Labels

Ziyi Zhang[1,2], Weikai Chen[3], Chaowei Fang[4], Zhen Li[5], Lechao Chen[6], Liang Lin[2], Guanbin Li[2,7*]
[1]National Key Laboratory of Novel Software Technology, Nanjing University, Nanjing, China
[2]Sun Yat-sen University, Guangzhou, China, [3] Tencent America
[4] Xidian Univiersity, [5] The Chinese University of Hong Kong (Shenzhen), [6] Zhejiang Lab
[7] Research Institute, Sun Yat-sen University, Shenzhen, China
zhangziyi@lamda.nju.edu.cn, liguanbin@mail.sysu.edu.cn

Abstract

Learning with noisy labels (LNL) is one of the most important and challenging problems in weakly-supervised learning. Recent advances adopt the sample selection strategy to mitigate the interference of noisy labels and use small-loss criteria to select clean samples. However, the one-dimensional loss is an over-simplified metric that fails to accommodate the complex feature landscape of various samples, and, hence, is prone to introduce classification errors during sample selection. In this paper, we propose RankMatch, a novel LNL framework that investigates additional dimensions of confidence and consistency in order to combat noisy labels. Confidence-wise, we propose a novel sample selection strategy based on confidence representation voting instead of the widely-used small-loss criterion. This new strategy is capable of increasing sample selection quantity without sacrificing labeling accuracy. Consistency-wise, instead of the widely adopted feature distance metric for measuring the consistency of inner-class samples, we advocate that the rank of principal features is a much more robust indicator. Based on this metric, we propose rank contrastive loss, which strengthens the consistency of similar samples regardless of their labels and facilitates feature representation learning. Experimental results on noisy versions of CIFAR-10, CIFAR-100, Clothing1M and WebVision have validated the superiority of our approach over existing state-of-the-art methods.

1. Introduction

The remarkable success of DNNs stems from the availability of large-scale datasets. However, it is highly laborious to obtain massive data with high-quality annotations.

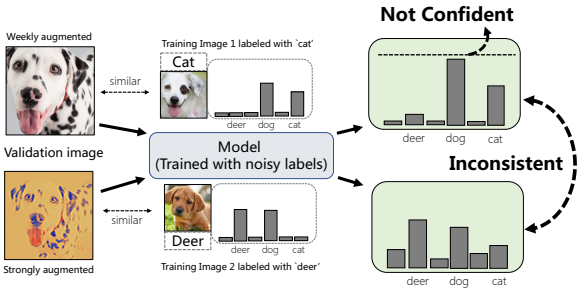

Figure 1: Motivation of RankMatch. After being trained with noisy labels, the model is validated by the augmented images. Noisy labels leads to high-entropy outputs, which lack confidence. We advocate for the use of confidence criterion for sample selection, as the confident samples are more likely to have clean labels. Moreover, The weakly augmented image is similar to training Image 1 (classified as 'dog'), while the strongly augmented one is similar to training Image 2 (classified as 'deer'). Considering the model tends to produce similar predictions for similar images, the two views of the same image have inconsistent predictions. Consequently, we introduce the rank contrastive loss to enhance the semantic consistency.

Hence, training DNNs using inexpensive samples from the search engine or machine annotation [27, 33, 11] has become an attractive alternative. However, these methods inevitably introduce erroneous labels, which cause poor performance as DNNs can easily overfit to noises [58]. Hence, learning with noisy labels (LNL) has become an important yet challenging task.

Sample selection is a widely adopted strategy in LNL that identifies clean samples from the dataset to alleviate the negative impact of noisy labels. The small-loss criterion is a popular sample selection strategy, which regards

*Corresponding author.

samples with lower loss as clean ones. However, such a one-dimensional loss is too simplified to accommodate the sophisticated distribution of high-dimensional data, such as the image features. As a result, a plethora of easy-but-noisy samples with small loss are often wrongly grouped into the clean set by using the loss-based strategy. This confirmation bias would be gradually amplified in the subsequent loss minimization training and severely impact the final performance [24, 44].

Another challenge of LNL lies in the difficulty of obtaining robust and consistent representation for samples across categories and subjects [24, 26, 19]. In semi-supervised learning, this issue can be largely addressed by enforcing consistent predictions given different views of the same sample [21]. However, in LNL, due to the existence of wrong labels, different views of the same object could be clustered into different categories, leading to inconsistent estimations (see Figure 1). Though the sample selection mechanism can identify a part of clean samples, the consistent representation learned from this set of clean data only work on simple samples with high confidence. Hence, how to achieve consistent prediction on difficult samples with low confidence remains an open question.

In this work, we follow the line of sample selection to avoid the adverse impact of noisy labels. However, instead of relying on an over-simplified 1D loss function, we propose to combat the noises by incorporating new perspectives from *confidence and consistency*. **Confidence-wise**, to ensure robust selection of clean samples, we leverage the advances of confidence learning (CL) [37] while avoiding its vulnerability to noisy network predictions. In particular, we investigate the performance of clean sample selection using fixed confidence thresholds at different noise levels. The results (Figure 4) show that a fixed high threshold leads to a much more accurate sample selection even at the presence of strong noises. Therefore, we leverage one fixed high confidence threshold for all classes to ensure the purity of confident samples. To address the lack of clean data by using a high threshold, we further propose *sample selection via confidence voting (SCV)* to ensure an ample collection of clean labels. Specifically, we generate K clusters from confident samples for each class and treat the cluster center as *confident prototype*. Each sample is considered clean if the label is consistent with the results voted by its k nearest confident prototypes. We empirically find that the proposed confidence voting mechanism can generate reliable and clean samples in sufficient amounts, ensuring both the *quality* and *quantity* of clean samples.

Consistency-wise, in addition to the classic consistency regularization, $i.e.$ enforcing consistent outputs of different views from the same sample, we propose to encourage consistent predictions between similar hard samples of the same category. To achieve this goal, we introduce *rank contrastive loss (RCL)*, a novel metric that fully leverages the rank of principal image features for robust measurement of similar samples. Empirical experiments show that the conventional l_2 distance cannot exploit the inner structure of feature representation, and, hence, fails to foster consistent clustering in the complex feature space corrupted by noisy samples. Our key observation is that while similar hard samples (with low confidence) of the same class may have a large l_2 distance in the feature space, the underlying rank of their principal features still remains consistent (see Figure 2). Therefore, our proposed RCL loss is able to promote feature consistency by encouraging correct and discriminative clustering of similar samples with low confidence.

We code our method as *RankMatch* as the two perspectives of the proposed framework can mutually benefit each other. While more reliable data segmentation based on high confidence can benefit the subsequent representation learning, the consistent representations further improve the confidence of uncertain data samples, and, in turn, strengthen our proposed confidence-based sample-selection mechanism. We validate the superiority of our method over the state-of-the-art approaches on several challenging benchmarks, including CIFAR, Clothing1M, and WebVision. Our contributions can be summarized as follows.

- A method with new state-of-the-art performance that is specially tailored for LNL problems by fostering both confidence and consistency.
- A novel sample-selection via confidence voting (SCV) strategy that generates reliable and ample clean samples for the subsequent training.
- *Rank contrastive loss* (RCL) based on the rank statistics of principal features that encourages consistency between hard samples of the same category. RCL can further benefit clean sample selection by promoting more discriminative features (see Fig. 1 in the supplementary material) for constructing intra-category clusters.

2. Related Work

Most existing methods on LNL can be categorized into two directions: 1) loss correction and 2) sample selection.

Loss Correction. For loss correction, mainstream techniques [51, 46, 29, 47, 22] estimate noise transition matrix with a small set of clean samples as prior knowledge. However, estimating the noise transition matrix is challenging, and their assumptions may be too ideal to fit real-world scenarios. To rectify the noisy labels, another line of research proposes self-training architectures where noisy labels are replaced by network predictions [43, 56]. Some methods focus on the design of noisy-tolerant loss functions. For example, [2, 48, 13, 12] leverage bounded loss functions to improve their robustness to noisy labels. Besides, there are

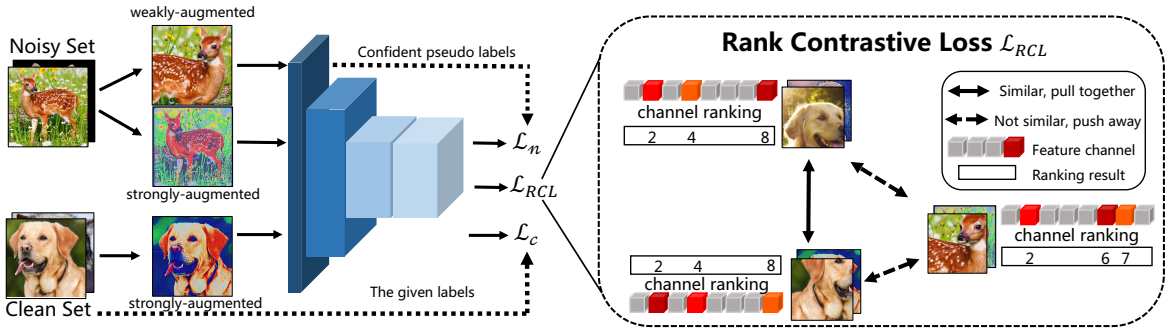

Figure 2: The training strategy of *RankMatch*. We use strong and weak augmentation to achieve consistency regularization in LNL. For the clean set, the given labels are used to regularize the network predictions on strongly augmented images via the supervised training loss L_c. For the noisy set, we first generate pseudo labels from the predictions on weakly augmented images and then use them to regularize the predictions on strongly augmented images via training loss L_n. Rank contrastive loss pulls similar samples together and pushes others away. The similarity matrix is defined by ranking feature channels instead of Euclidean distance. The final objective function is a joint loss, including a class diversity regularization term in addition.

some proposed unbiased loss function [35], information-based loss function called L_{DMI} [54], punishment regularization [32], confidence regularization [8], early learning regularization [30], contrastive representation regularization [28] and over-parameterized term [31]. However, it is arduous for loss-correction-based methods to play an advantage in practical scenarios facing high noise ratios.

Sample selection. Sample selection-based methods strive to reweight training samples and identify clean data to reduce the interference of noisy labels. Many existing methods propose introducing improved selection criteria to make the selected samples cleaner [23, 59, 3, 36, 24, 1]. Memorization effect [4] illuminates that DNNs learn clean and simple patterns faster than noisy labels, leading to a trend to consider small loss as a criterion for sample selection, which means that samples with smaller loss values are more likely to have clean labels. For instance, [3, 36, 24, 1] leverage clean data distribution as selection criterion by lower-loss component in mixture model fitted by per-sample loss. [55, 19] introduces JS-divergence between network prediction and labels as a selection criterion. [38] leverages the cross-entropy loss between disagreement distribution and labels as selection criterion. However, the loss value is just a scalar, contains limited information, and is prone to errors. For example, it is hard to distinguish between hard-to-learn clean samples and noisy samples after the warm-up stage since both have relatively large loss values. Besides, confirmation bias [44] is a common issue often encountered with the small-loss criterion. Once the noisy samples are wrongly grouped into the clean set by the small-loss criterion, their losses will keep small in the later training. It remains challenging to rectify the sample selection errors. Many methods mitigate this issue by training two networks [18, 34, 14, 57, 49, 24, 55]. But it is still a problem to get more accurate labels through the small-loss criterion. More recently, Confident Learning (CL) [37] proposes confident criterion as an alternative for sample selection. It computes a set of thresholds for every class and chooses confident samples whose network predicted highest probabilities are greater than the corresponding thresholds. However, CL directly estimates the noise transition matrix by confident samples, which is vulnerable to the network predictions and threshold setting. And it is prone to fail to perform well in large-scale cases and real-world scenarios with a high noise ratio.

Semi-supervised Learning. The Semi-supervised Learning (SSL) has seen fast progress by leveraging consistency regularization [41, 53, 5, 21, 39, 40, 45], which is used to minimize the difference in network prediction between two views of the same image. FixMatch [21] simply but effectively combines strong and weak augmentation in consistency regularization with confidently pseudo-labeling to achieve great success in semi-supervised learning setting. Most recently, [39] propose an uncertainty-aware pseudo-labels selection (UPS) framework to improve pseudo labeling accuracy. RankingMatch [45] introduces a triplet loss, where the similarity between samples is measured by the L2-norm outputs. Furthermore, it is worth noting that the rank statistics of principal features can serve as a metric to measure the pairwise similarity among samples [15]. There are similarities between Learning with Noisy Labels (LNL) problems and semi-supervised learning setting.

In semi-supervised learning settings, confident predictions have strong connection with true labels as labeled data regularizing the unlabeled data, which are not easy to over-confident confused by noisy label. And leveraging confident predictions inevitably lacks sample diversity to make pseudo labels more reliable, while difficult samples always have uncertain predictions. Our approach finds a good balance of confidence, sample diversity and consistency.

3. Method

Our method is introduced in three parts: the sample selection mechanism (Figure. 3), the consistency regularization and rank contrastive loss. Figure 2 shows the overview of *RankMatch*. The algorithm is included in the supplementary material.

Preliminaries. For the C-way image classification task with noisy labels, denoted by $\mathcal{D} = (\mathcal{X}, \widetilde{\mathcal{Y}})$ the training data, where $\mathcal{X}$ is training images, and $\widetilde{\mathcal{Y}}$ is labels which may be wrongly annotated. We denote the DNNs in training stage as $P(F(\mathbf{x}, \theta))$, where $\mathbf{x} \in \mathbb{R}^I$ is the input image, $F : \mathbb{R}^I \rightarrow \mathbb{R}^L$ is the feature extractor where I and L stand for the dimension of input space and embedding space respectively, and $P : \mathbb{R}^L \rightarrow \mathbb{R}^C$ represents the classifier. θ denotes the network parameters of the feature extractor. We use $\mathbf{f}_i$ and $\mathbf{p}_i$ as simplified forms of $F(\mathbf{x}_i, \theta)$ and $P(F(\mathbf{x}_i, \theta))$ respectively.

3.1. Sample Selection via Confidence Voting

For the sample selection, we leverage the advantage of confident learning [37] rather than the widely used small-loss criterion. We explore using a fixed confidence threshold to avoid the vulnerability of inaccurate network predictions. Figure 4 illustrates that confident predictions selected by a fixed high threshold are reliable but of a limited number. Hence, we propose a new sample selection strategy based on the confidence criterion that maintains reliability while ensuring sufficient confident samples across categories.

Confident Samples. In the setting of LNL, it is hard to select confident samples just by a single threshold when facing different levels of noise rates. Some classes are easy-to-learn, and one threshold for all classes is likely to result in class-unbalanced selection. In order to avoid data imbalance among classes, we select the the top B basic confident samples $\mathcal{I}_b$ in each class to ensure each class at least has B samples:

$$\mathcal{I}_b^c = \underset{|\hat{\mathcal{X}}|=B, \hat{\mathcal{X}} \subseteq \mathcal{X}}{\arg\max} \sum_{\mathbf{x} \in \hat{\mathcal{X}}} \mathbf{p}[c], \tag{1}$$

where $\mathbf{p}[c]$ is the c-th element of $\mathbf{p}$ $(= P(F(\mathbf{x}, \theta)))$, namely the probability value of the c-th class. And additional samples with prediction probability greater than a threshold as

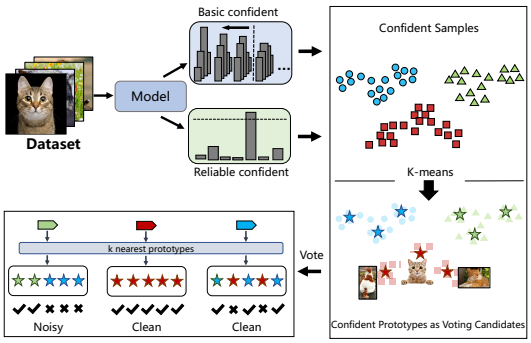

Figure 3: Sample-selection via Confidence Voting (SCV) module. We first select the most representative confident samples. Then, the selected confident samples generate prototypes as voting candidates. For each sample, its k nearest prototypes are regarded as voters to identify whether the label is clean.

the reliable confident samples:

$$\mathcal{I}_r^c = \{\mathbf{x}_i | \mathbf{p}_i[c] > \tau, \mathbf{x}_i \in \mathcal{X}\}, \tag{2}$$

The final selected confident samples is the union of the two sets $\mathcal{I}^c = \mathcal{I}_b^c \cup \mathcal{I}_r^c$.

Confident Prototype Generation. To guarantee the quality of the confident samples, we set high threshold τ and small portion B, but face the problem of low sample quantity. Although with a small number, the reliable representations of confident samples reflect some common intra-class patterns. Thus, we generate representation prototypes using the selected samples for further selection. Straightforwardly, a representative feature can be generated by averaging features of confident samples for each class:

$$\phi^c = \frac{1}{|\mathcal{I}^c|} \sum_{\mathbf{x} \in \mathcal{I}^c} F(\mathbf{x}, \theta). \tag{3}$$

Even if the confident samples include some noise during the early training stage, the prototype generation process will predominantly rely on the most clean samples, rendering it robust to noisy labels. However, the mean vector is unable to represent a class with large intra-class variations. We further diversify the intra-class patterns by generating more prototypes of each class. Specifically, we split the confident samples into K clusters for each class by K-means. For the c-th class, we define the average centers of the obtained clusters features as $\Phi^c = \{\phi_j^c\}_{j=1}^K$. Our sample-selection module is based on the generated confident prototypes Φ $(= \bigcup_{c=1}^C \Phi^c)$.

Data Segmentation by Confidence Voting. After obtaining $C \times K$ prototypes, we first find the k nearest prototypes

of $\mathbf{x}_i$ measured by the cosine similarity:

$$\mathcal{V}_i = \underset{|\hat{\Phi}|=k, \hat{\Phi}\subseteq\Phi}{\arg\max} \sum_{\phi\in\hat{\Phi}} \cos(\mathbf{f}_i, \phi). \tag{4}$$

The class label to which the most prototypes belong in $\mathcal{V}_i$ is considered as the voting result $\mathbf{y}'_i$. Clean samples $\mathcal{D}_{\text{cln}}$ are identified by checking whether $\mathbf{y}'_i$ is consistent with the provided label $\widetilde{\mathbf{y}}_i$:

$$\mathcal{D}_{\text{cln}} = \{(\mathbf{x}_i, \widetilde{\mathbf{y}}_i) | \widetilde{\mathbf{y}}_i = \mathbf{y}'_i, \forall(\mathbf{x}_i, \widetilde{\mathbf{y}}_i) \in \mathcal{D}\}, \tag{5}$$

The remaining noisy set is $\mathcal{D}_{\text{nsy}} = \mathcal{D} \setminus \mathcal{D}_{\text{cln}}$.

3.2. Consistency Regularization

This paper argues that noisy labels lead to inconsistent predictions (Figure 1). To enhance the consistency of the predictions augmented from the same image, we apply two different augmentation methods to generate two views for training images, including "weak" augmentation ($\mathcal{A}_w$) and "strong" augmentation ($\mathcal{A}_s$). Weak augmentation refers to random cropping and flipping, while strong augmentation refers to AutoAugment [9], which uses reinforcement learning to find augmentation strategies automatically and requires labeled data to learn. Note that our sample selection mechanism provides cleaner supervision signals, benefiting AutoAugment to accept a more effective augmentation policy.

Following [21], we transfer the training image $\mathbf{x}$ with the two augmentation strategies, resulting in two views: $\mathbf{v}^w = \mathcal{A}_w(\mathbf{x})$, and $\mathbf{v}^s = \mathcal{A}_s(\mathbf{x})$. The corresponding network predictions are denoted as $\mathbf{p}^w$ and $\mathbf{p}^s$, respectively. Given that the semantic information of the augmented images is not modified, we assume the two network predictions to be consistent.

For clean samples, we direct use cross-entropy loss between predictions on strongly augmented images and the original labels:

$$\mathcal{L}_c = -\frac{1}{|\mathcal{D}_{\text{cln}}|} \sum_{(\mathbf{x},\widetilde{\mathbf{y}})\in\mathcal{D}_{\text{cln}}} \widetilde{\mathbf{y}} \circ \log(\mathbf{p}^s), \tag{6}$$

where $\circ$ denotes the inner product.

For noisy samples with high prediction confidence, we rectify their labels with predictions of weakly augmented images:

$$\hat{\mathcal{D}}_{\text{nsy}} = \{(\mathbf{x}_i, \hat{\mathbf{y}}_i) | \forall \mathbf{x}_i \in \mathcal{X}_{\text{nsy}}, \max_c \{\mathbf{p}^w[c]\} > \nu, \hat{\mathbf{y}}_i = c\}, \tag{7}$$

where ν is the confidence threshold, we set it the same as that in Eq. 2 to avoid introducing another parameter. Inspired by [21], the cross entropy losses between the predictions of weakly and strongly augmented images are employed as regularizing consistency for the CNN model learning:

$$\mathcal{L}_n = -\frac{1}{|\hat{\mathcal{D}}_{\text{nsy}}|} \sum_{(\mathbf{x},\hat{\mathbf{y}})\in\hat{\mathcal{D}}_{\text{nsy}}} \hat{\mathbf{y}} \circ \log(\mathbf{p}^s). \tag{8}$$

3.3. Rank Contrastive Loss

Robust representations can benefit our samples selection strategy. However, some valuable information is excluded by Eq. 7 and Eq. 8 since the hard-to-learn noisy samples have less confident predictions and thus can not be optimized. To obtain complete and robust representations, we propose a rank contrastive loss (RCL) to strengthen the consistency of the "similar" samples while pushing "dissimilar" ones further.

The widely used L_2 distance as the similarity metric is less robust in LNL, as overfitting to noisy labels tends to introduce spurious close euclidean distance. Every convolution kernel filters out certain kinds of attributes in the input image. Similar visual patterns are prone to activate the same representation channel of response maps produced by convolution layers. Hence, the indices of feature elements rank ordered in accordance with their magnitudes, can serve as a metric for assessing the pairwise representations similarity [15]. Although noisy labels affect the feature value, we leverage rank statistics of principal features to estimate the similarity matrix. Specifically, we rank the values in all channels of $\mathbf{f}_i$ and regard the r channels with the largest activation values as the principal feature dimensions,

$$\mathcal{R}_i = \underset{|\hat{\mathcal{R}}|=r,\ \hat{\mathcal{R}}\subset\{1,2,\ldots,L\}}{\arg\max} \sum_{n\in\hat{\mathcal{R}}} \mathbf{f}_i[n], \tag{9}$$

where $\mathbf{f}_i[n]$ denote the n-th channel of feature $\mathbf{f}_i$. If two samples $\mathbf{x}_i$ and $\mathbf{x}_j$ share the same principal feature dimensions, we assume they are "similar". The similarity s_{ij} between $\mathbf{x}_i$ and $\mathbf{x}_j$ is evaluated as follows,

$$s_{ij} = \begin{cases} 1 & \text{if } \mathcal{R}_i = \mathcal{R}_j; \\ 0 & \text{otherwise.} \end{cases} \tag{10}$$

To achieve clustering in the feature space, we adopt the following binary cross-entropy loss as:

$$\mathcal{L}_{\text{RCL}} = -\frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} [s_{ij} log(\mathbf{p}_i \circ \mathbf{q}_j) + (1 - s_{ij}) log(1 - \mathbf{p}_i \circ \mathbf{q}_j)], \tag{11}$$

where $\mathbf{p}_i = P(F(\mathbf{v}_i^w, \theta))$ and $\mathbf{q}_i = P(F(\mathbf{v}_i^s, \theta))$. The train-

ing loss $\mathcal{L}_{\text{RCL}}$ can be divided into two terms:

$$\mathcal{L}_{\text{RCL}} = \overbrace{-\frac{1}{N^2}\sum_{i=1}^{N}[s_{ii}\log(\mathbf{p}_i \circ \mathbf{q}_i) + (1-s_{ii})\log(1-\mathbf{p}_i \circ \mathbf{q}_i)]}^{\text{Consistency regularization}} \\ -\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}[s_{ij}\log(\mathbf{p}_i \circ \mathbf{q}_j) + (1-s_{ij})\log(1-\mathbf{p}_i \circ \mathbf{q}_j)]. \tag{12}$$

Since $s_{ii} = 1$ always holds, organize the first term as:

$$-\frac{1}{N^2}\sum_{i=1}^{N}\log(\mathbf{p}_i \circ \mathbf{q}_i) = -\frac{1}{N^2}\sum_{i=1}^{N}\log(\mathbf{p}_i^w \circ \mathbf{p}_i^s). \tag{13}$$

Minimizing this term is equal to $\max\{\mathbf{p}_i^w \circ \mathbf{p}_i^s\}$, which pulls the outputs of strongly and weakly augmented images closer. Thus, $\mathcal{L}_{\text{RCL}}$ in the diagonal of S can be regarded as consistency regularization. Even in the early training stage, S is the identity matrix that removes the second term in Eq. 12, improving the network representation ability.

3.4. Overall Loss Function

Following [43, 24, 3], we apply the fair class diversity-promoting regularization $\mathcal{L}_{div}$ to prevent assigning most samples into few classes:

$$\mathcal{L}_{div} = \sum_{c=1}^{C}\frac{1}{C}\log(\frac{1}{C}/\frac{\sum_{i=1}^{N}\mathbf{p}_i^w[c]}{N}). \tag{14}$$

The overall loss function for network optimization is as follows:

$$\mathcal{L} = \mathcal{L}_c + \lambda_n\mathcal{L}_n + \mathcal{L}_{\text{RCL}} + \mathcal{L}_{div}. \tag{15}$$

We following [24, 14, 57, 36] to train two networks to combat the confirmation bias, detailed algorithm is included in supplementary material.

4. Experiments

4.1. Experimental Setup

We evaluate the proposed *RankMatch* on CIFAR-10, CIFAR-100 [20], as well as the real-world datasets Clothing1M [52] and WebVision [27]. Both CIFAR-10 and CIFAR-100 contain 50,000 training images and 10,000 test images of size 32×32. Following previous works [43, 25], we conduct experiments requiring noise augmentation, either symmetric or asymmetric [43, 25] label noise injection. We follow the identical implementation of generating symmetric and asymmetric noises on CIFAR datasets with DivideMix [24]. Symmetric noise is generated by randomly replacing the labels for a percentage of the training data with all possible labels. Asymmetric noise is to mimic the structure of real-world label noise, where labels are only replaced by similar classes ($e.g.$ deer→horse, dog↔cat). We adopt 18-layer PreAct Resnet [17] as the backbone. During training, we train DNNs for 300 epochs using SGD optimizer with a momentum of 0.9, with weight decay of 0.0005 for CIFAR-10 and 0.001 for CIFAR-100. The batch size is 64 for CIFAR-10 and 128 for CIFAR-100. We set the initial learning rate as 0.02 and reduce it by a factor of 10 after 150 epochs. The warm-up period is 10 epochs for CIFAR-10 and 30 epochs for CIFAR-100, respectively. For all CIFAR experiments, we set $\tau = 0.95$, $r = 5$, and choose λ_n from {0.2, 0.5, 1, 2, 10, 15} for a small validation test. More experimental details are in supplementary materials.

Clothing1M contains 1 million real-world shopping images belonging to 14 classes. We use ResNet-50 [16] with weights pretrained on ImageNet as the classification model. WebVision consists of web-crawled images with the same concepts from ImageNet ILSVRC12 [10]. We follow the previous work [24, 7] and compare baseline methods on the first 50 classes of the Google image subset using the inception-resnet v2 [42]. The training details are delineated in supplementary materials.

4.2. Experimental Results

Results on CIFAR. Table 1 illuminates *RankMatch* outperforms the state-of-the-art models across most noisy levels. RRL [26] combats noisy labels by learning robust representations and slightly outperforms our method on CIFAR-10 with 20% symmetric label noise. One possible explanation is that the massive clean labels make it easy to rectify the symmetric label noise. Thus, the sample selection strategy is less effective in this scenario but shows great advantages for a high noise ratio and real-world dataset. Benefiting from the advanced confidence sample selection mechanism, *RankMatch* boosts the averaged test accuracy of the last ten epochs from 77.4% to 92.1% on CIFAR-10 and 33.1% to 49.9% on CIFAR-100 under the extreme case of 90% noise. Even without MixMatch [6] technique, *RankMatch* achieves around **20 points** improvement (50.6% v.s. 31.5%) compared to DivideMix [24] on CIFAR-100 with 90% noisy labels. We surpass CTRR [28], the most recent contrastive learning-based method, under all noise ratios, especially on the more challenging CIFAR-100 dataset.

Results on real-world dataset. Table 2 and Table 3 demonstrate the results on Clothing1M and WebVision, respectively. Our method outperforms all the alternative methods in both noisy real-world datasets. We achieve the improvement of 0.18% over SFT [50], the most recent sample selection-based method with MixMatch [6]. Besides, the most recent baseline UNICON [19] well combines the advantages of semi-supervised learning (DivideMix-like) and contrastive learning (RRL-like) and achieves great per-

Dataset		CIFAR-10					CIFAR-100			
Method/Noise ratio		20%	50%	80%	90%	Asym. 40%	20%	50%	80%	90%
Cross-Entropy	Best	86.8	79.4	62.9	42.7	85.0	62.0	46.7	19.9	10.1
	Last	82.7	57.9	26.1	16.8	72.3	61.8	37.3	8.8	3.5
PENCIL [56]	Best	92.4	89.1	77.5	58.9	88.5	69.4	57.5	31.1	15.3
	Last	92.0	88.7	76.5	58.2	88.1	68.1	56.4	20.7	8.8
Meta-Learning [25]	Best	92.9	89.3	77.4	58.7	89.2	68.5	59.2	42.4	19.5
	Last	92.0	88.8	76.1	58.3	88.4	67.7	58.0	40.1	14.3
DivideMix [24]	Best	96.1	94.6	93.2	76.0	93.4	77.3	74.6	60.2	31.5
	Last	95.7	94.4	92.9	75.4	92.1	76.9	74.2	59.6	31.0
ELR [30]	Best	95.8	94.8	93.3	78.7	93.0	77.6	73.6	60.8	33.4
MOIT [38]	Best	94.1	91.1	75.8	70.1	93.2	75.9	70.1	51.4	24.5
RRL [26]	Last	96.4	95.3	93.3	77.4	93.3	**80.3**	76.0	61.1	33.1
CTRR [28]	Best	93.3	-	86.7	84.3	89.0	70.1	-	43.7	-
RankMatch	Best	**96.5**	**95.6**	**94.5**	**92.6**	**94.7**	79.5	**77.9**	**67.6**	**50.6**
	Last	**96.4**	**95.4**	**94.2**	**92.1**	**94.4**	79.3	**77.6**	**67.2**	**49.9**

Table 1: Comparison between RankMatch and state-of-the-art methods on CIFAR-10 and CIFAR-100 under symmetric and asymmetric noise. "Best" refers to the best test accuracy across all epochs and "Last" is the averaged test accuracy over the last 10 epochs.

Method	Test Accuracy
Cross-Entropy	69.21
PENCIL [56]	73.49
DivideMix [24]	74.76
RRL [26]	74.97
DSOS [1]	73.63
UNICON [19]	74.98
SOP [31]	73.50
SFT [50]	75.08
RankMatch	**75.22**

Table 2: Comparison with state-of-the-art methods on Clothing1M. Baseline results are copied from original papers.

Method	WebVision		ILSVRC12	
	top1	top5	top1	top5
MentorNet [18]	63.00	81.40	57.80	79.92
Co-teaching [14]	63.58	85.20	61.48	84.70
Iterative-CV [7]	65.24	85.34	61.60	84.98
DivideMix [24]	77.32	91.64	75.20	90.84
RRL [26]	77.8	91.3	74.4	90.9
DSOS [1]	77.76	92.04	74.36	90.80
SOP [31]	76.6	-	69.1	-
UNICON [19]	77.60	93.44	75.29	93.72
RankMatch	**79.91**	**93.61**	**77.39**	**94.26**

Table 3: Comparison with state-of-the-art methods on (mini) WebVision dataset. Numbers denote top-1 (top-5) accuracy (%) on the WebVision and ImageNet ILSVRC12 validation set. Results for baselines are copied from the corresponding papers.

formance gain. In particular, RankMatch achieves **0.24%** performance gain over UNICON on the challenging Clothing1M dataset. Moreover, RankMatch obtain over **2%** Top-1 accuracy improvement over state-of-the-art on both mini-WebVison and ILSVRC12 validation sets, and guarantees state-of-the-art Top-5 accuracy on WebVision and ILSVRC12, indicating the effectiveness of our approach.

4.3. Ablation Study

The role of confidence in LNL. To explore the role of the fixed confidence threshold in the LNL setting, we focus on the relationship between confident predictions and ground-truth labels. Specifically, we directly train DNNs for classification on CIFAR-10 under different noise levels and record the following statistics: confident ratio, precision, and recall of confident predictions.

From the results shown in Figure 4, we discover the following empirical rules: (1) Confident ratio and recall increase in the same trend, while precision continues to decrease, indicating that the process of overfitting to noise label can regard as the process of overconfidence. (2) The precision of confident samples maintains a high level at the early stage, which means that confident predictions selected by the high threshold are reliable after the warm-up stage.

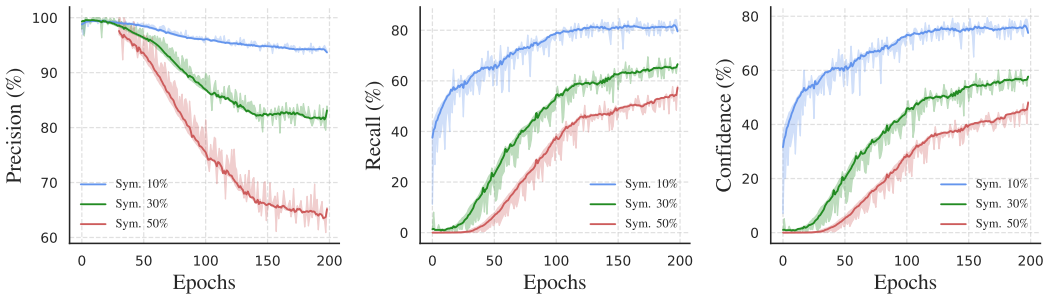

Figure 4: Some key statistics of confident samples on CIFAR-10. The confident samples are selected via the fixed confidence threshold. (a) Precision, ratio of correctly predicted samples in confident samples, maintains a high level (close to 100%) at the early stage, but continues to decline afterwards. Under Symmetry-50% noise, there is no confident samples in the early period. Hence, we record the precision after 30 epochs. (b) Recall, ratio of confident samples in correct prediction samples, maintains a low level in early stage, and gradually rises to high level. (c) Confidence ratio, ratio of confident samples in training set, has the same trend with Recall.

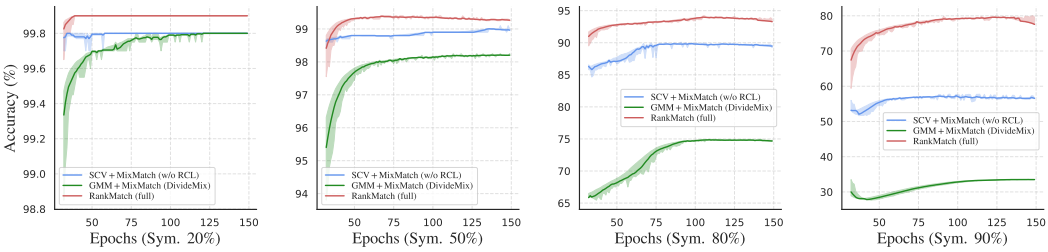

Figure 5: Accuracy (%) of the selected clean samples under different symmetric noise rates on CIFAR-100 datasets. Our full method RankMatch is compared with two baselines. One is DivideMix, denoted as GMM and MixMatch. The other is replacing GMM with our proposed Sample-selection via Confident Voting (SCV) method. Compared with GMM, SCV is more efficient to identify the clean samples. And our full model RankMatch has significant improvement in the sample selection stage under high noise rate.

(3) However, it is prone to selecting few samples when directly selecting reliable confidence samples by the high threshold because the recall is extremely low under high noise. It forms a dilemma between the quality and quantity of the selected samples.

Sample selection via confident voting (SCV) is one main reason for *RankMatch* achieving great improvement. To study and verify its effectiveness, we test the accuracy of the selected clean samples under various symmetric noises on CIFAR-100. Since the great success DivideMix has made, we regard the GMM-based sample selection method as our baseline. Figure 5 presents that the SCV-based methods are more accurate than DivideMix in all cases. Furthermore, in the extremely demanding scenario ($, i.e.$, Symmetry-80% and Symmetry-90%), SCV based method achieves tremendous improvement. Specifically, DivideMix with SCV achieves over 18% improvement than that with GMM at 80% label noise. In contrast, our full *RankMatch* achieves over 25% improvement, which implies that the SCV benefits from more robust representation training. Under the most extreme 90% label noise, our full *RankMatch* achieves over 140% selection accuracy gain, while DivideMix with SCV improves over 75%. The explanation for this result is that our full model includes consistency regularization and rank contrastive loss. Consistency regularization enhances semantic consistency and rectifies some confident-but-noisy features, which generates more robust prototypes and benefits SCV. Furthermore, rank contrastive loss promotes more discriminative features, improving confident prototype generation. The supplementary material contains other in-depth studies examining the sensitivity of high threshold τ and hyper-parameters in SCV.

Effects of *RankMatch* components. We remove the corresponding components to study the effects of RCL and network ensemble. We remove $\mathcal{L}_{\text{RCL}}$ to validate the effect of rank contrastive loss. We replace the SCV module as the sample selection strategy of DivideMix to validate the effectiveness of SCV. We remove the K-means in SCV for detailed study. As the results shown in Table 4, we find that rank contrastive loss is beneficial to *RankMatch*. The results

Dataset		CIFAR-10		CIFAR-100	
Noise ratio		50%	90%	50%	90%
RankMatch	Best	**95.6**	**92.6**	**77.9**	**50.6**
	Last	**95.4**	**92.1**	**77.6**	**49.9**
w/o SCV	Best	95.1	89.6	75.3	45.1
	Last	94.9	89.1	74.9	44.9
SCV w/o K-means	Best	95.4	91.8	76.5	47.9
	Last	95.2	91.4	75.2	47.5
w/o RCL	Best	95.2	90.8	76.1	45.6
	Last	94.9	88.7	75.2	44.5

Table 4: Ablation study results in terms of test accuracy (%) on CIFAR-10 and CIFAR-100 with symmetric label noise.

also validate the effectiveness of our proposed SCV. Diversifing the intra-class patterns by K-means is more effective under complex scenario and extreme noise ratio.

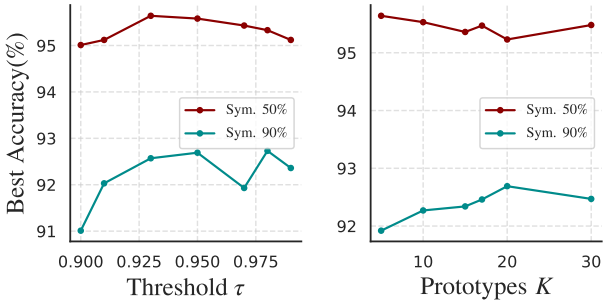

Figure 6: Sensitivity to the variance of hyperparameters. Experiments are conducted on CIFAR-10 under 50% and 90% symmetric noises.

Sensitivity Analysis. SCV introduce the confidence threshold τ and number of prototypes K. We range the number of prototypes K from 5 to 30 and vary the threshold τ from 0.90 to 0.99. Figure 6 shows that our method is robust against different choices for τ and K. More in-depth studies are included in supplementary materials.

5. Conclusions

We presented RankMatch, a novel framework for LNL that strives to combat noisy labels by enhancing confidence and consistency. Confidence-wise, we propose a novel clean sample selection strategy based on confidence representation voting rather than the small-loss criterion. Collecting votes from confident prototypes makes our method robust to noise and can sieve clean samples out with ample quantity. Consistency-wise, we introduce a novel rank contrastive loss based on the rank statistics of principal features instead of the widely-used L_2 distance. Such a loss could enhance consistency between similar samples even if they were wrongly labeled. Moreover, we leverage consistency regularization to enhance the semantic consistency. We validate our superiority over several challenging benchmarks.

Acknowledgments

This work was supported in part by the Guangdong Basic and Applied Basic Research Foundation (NO. 2020B1515020048), in part by the National Natural Science Foundation of China (NO. 61976250), in part by the Shenzhen Science and Technology Program (NO. JCYJ20220530141211024).

References

[1] Paul Albert, Diego Ortego, Eric Arazo, Noel O'Connor, and Kevin McGuinness. Addressing out-of-distribution label noise in webly-labelled data. In *Winter Conference on Applications of Computer Vision (WACV)*, 2022.

[2] Ehsan Amid, Manfred KK Warmuth, Rohan Anil, and Tomer Koren. Robust bi-tempered logistic loss based on bregman divergences. In *Advances in Neural Information Processing Systems*, pages 14987–14996, 2019.

[3] Eric Arazo, Diego Ortego, Paul Albert, Noel E. O'Connor, and Kevin McGuinness. Unsupervised label noise modeling and loss correction. In *ICML*, pages 312–321, 2019.

[4] Devansh Arpit, Stanisław Jastrzębski, Nicolas Ballas, David Krueger, Emmanuel Bengio, Maxinder S Kanwal, Tegan Maharaj, Asja Fischer, Aaron Courville, Yoshua Bengio, et al. A closer look at memorization in deep networks. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 233–242. JMLR. org, 2017.

[5] David Berthelot, Nicholas Carlini, Ekin D. Cubuk, Alex Kurakin, Kihyuk Sohn, Han Zhang, and Colin Raffel. Remixmatch: Semi-supervised learning with distribution alignment and augmentation anchoring. *arXiv preprint arXiv:1911.09785*, 2019.

[6] David Berthelot, Nicholas Carlini, Ian J. Goodfellow, Nicolas Papernot, Avital Oliver, and Colin Raffel. Mixmatch: A holistic approach to semi-supervised learning. *NeurIPS*, 2019.

[7] Pengfei Chen, Benben Liao, Guangyong Chen, and Shengyu Zhang. Understanding and utilizing deep neural networks trained with noisy labels. In *ICML*, pages 1062–1070, 2019.

[8] Hao Cheng, Zhaowei Zhu, Xingyu Li, Yifei Gong, Xing Sun, and Yang Liu. Learning with instance-dependent label noise: A sample sieve approach. *arXiv preprint arXiv:2010.02347*, 2020.

[9] Ekin D Cubuk, Barret Zoph, Dandelion Mane, Vijay Vasudevan, and Quoc V Le. Autoaugment: Learning augmentation strategies from data. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 113–123, 2019.

[10] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *CVPR*, 2009.

[11] Rob Fergus, Li Fei-Fei, Pietro Perona, and Andrew Zisserman. Learning object categories from internet image searches. *Proceedings of the IEEE*, 98(8):1453–1466, 2010.

[12] Aritra Ghosh, Himanshu Kumar, and PS Sastry. Robust loss functions under label noise for deep neural networks. In *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.

[13] Maoguo Gong, Hao Li, Deyu Meng, Qiguang Miao, and Jia Liu. Decomposition-based evolutionary multiobjective optimization to self-paced learning. *IEEE Transactions on Evolutionary Computation*, 23(2):288–302, 2018.

[14] Bo Han, Quanming Yao, Xingrui Yu, Gang Niu, Miao Xu, Weihua Hu, Ivor W. Tsang, and Masashi Sugiyama. Co-teaching: Robust training of deep neural networks with extremely noisy labels. In *NeurIPS*, pages 8536–8546, 2018.

[15] Kai Han, Sylvestre-Alvise Rebuffi, Sebastien Ehrhardt, Andrea Vedaldi, and Andrew Zisserman. Automatically discovering and learning new visual categories with ranking statistics. In *International Conference on Learning Representations (ICLR)*, 2020.

[16] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, pages 770–778, 2016.

[17] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Identity mappings in deep residual networks. In *ECCV*, pages 630–645, 2016.

[18] Lu Jiang, Zhengyuan Zhou, Thomas Leung, Li-Jia Li, and Li Fei-Fei. Mentornet: Learning data-driven curriculum for very deep neural networks on corrupted labels. In *ICML*, pages 2309–2318, 2018.

[19] Nazmul Karim, Mamshad Nayeem Rizve, Nazanin Rahnavard, Ajmal Mian, and Mubarak Shah. Unicon: Combating label noise through uniform selection and contrastive learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9676–9686, June 2022.

[20] Alex Krizhevsky and Geoffrey Hinton. Learning multiple layers of features from tiny images. Mater's thesis, University of Toronto, 2009.

[21] Alex Kurakin, Chun-Liang Li, Colin Raffel, David Berthelot, Ekin Dogus Cubuk, Han Zhang, Kihyuk Sohn, Nicholas Carlini, and Zizhao Zhang. Fixmatch: Simplifying semi-supervised learning with consistency and confidence. In *NeurIPS*, 2020.

[22] Kuang-Huei Lee, Xiaodong He, Lei Zhang, and Linjun Yang. Cleannet: Transfer learning for scalable image classifier training with label noise. In *CVPR*, pages 5447–5456, 2018.

[23] Jichang Li, Guanbin Li, Feng Liu, and Yizhou Yu. Neighborhood collective estimation for noisy label identification and correction, 2022.

[24] Junnan Li, Richard Socher, and Steven CH Hoi. Dividemix: Learning with noisy labels as semi-supervised learning. *ICLR*, 2020.

[25] Junnan Li, Yongkang Wong, Qi Zhao, and Mohan S. Kankanhalli. Learning to learn from noisy labeled data. In *CVPR*, pages 5051–5059, 2019.

[26] Junnan Li, Caiming Xiong, and Steven C.H. Hoi. Learning from noisy data with robust representation learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 9485–9494, October 2021.

[27] Wen Li, Limin Wang, Wei Li, Eirikur Agustsson, and Luc Van Gool. Webvision database: Visual learning and understanding from web data. *arXiv:1708.02862*, 2017.

[28] Yi Li, Liu Sheng, She Qi, A. Ian McLeod, and Wang Boyu. On learning contrastive representations for learning with noisy labels. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2022.

[29] Yuncheng Li, Jianchao Yang, Yale Song, Liangliang Cao, Jiebo Luo, and Li-Jia Li. Learning from noisy labels with distillation. In *ICCV*, pages 1928–1936, 2017.

[30] Sheng Liu, Jonathan Niles-Weed, Narges Razavian, and Carlos Fernandez-Granda. Early-learning regularization prevents memorization of noisy labels. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 20331–20342. Curran Associates, Inc., 2020.

[31] Sheng Liu, Zhihui Zhu, Qing Qu, and Chong You. Robust training under label noise by over-parameterization. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162, pages 14153–14172. PMLR, 17–23 Jul 2022.

[32] Yang Liu and Hongyi Guo. Peer loss functions: Learning from noisy labels without knowing noise rates. In *Proceedings of the 37th International Conference on Machine Learning*, ICML '20, 2020.

[33] Dhruv Mahajan, Ross Girshick, Vignesh Ramanathan, Kaiming He, Manohar Paluri, Yixuan Li, Ashwin Bharambe, and Laurens van der Maaten. Exploring the limits of weakly supervised pretraining. In *ECCV*, 2018.

[34] Eran Malach and Shai Shalev-Shwartz. Decoupling "when to update" from "how to update". In *NIPS*, 2017.

[35] Nagarajan Natarajan, Inderjit S Dhillon, Pradeep K Ravikumar, and Ambuj Tewari. Learning with noisy labels. In *Advances in neural information processing systems*, pages 1196–1204, 2013.

[36] Kento Nishi, Yi Ding, Alex Rich, and Tobias Höllerer. Augmentation Strategies for Learning with Noisy Labels. *arXiv e-prints*, page arXiv:2103.02130, Mar. 2021.

[37] Curtis G. Northcutt, Lu Jiang, and Isaac L. Chuang. Confident learning: Estimating uncertainty in dataset labels. *Journal of Artificial Intelligence Research (JAIR)*, 70:1373–1411, 2021.

[38] Diego Ortego, Eric Arazo, Paul Albert, Noel E. O'Connor, and Kevin McGuinness. Multi-objective interpolation training for robustness to label noise. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6606–6615, June 2021.

[39] Mamshad Nayeem Rizve, Kevin Duarte, Yogesh S Rawat, and Mubarak Shah. In defense of pseudo-labeling: An uncertainty-aware pseudo-label selection framework for semi-supervised learning. In *International Conference on Learning Representations*, 2021.

[40] Kuniaki Saito, Donghyun Kim, and Kate Saenko. Openmatch: Open-set consistency regularization for semi-supervised learning with outliers. *arXiv preprint arXiv:2105.14148*, 2021.

[41] Mehdi Sajjadi, Mehran Javanmardi, and Tolga Tasdizen. Regularization with stochastic transformations and perturbations for deep semi-supervised learning. In *Advances in Neural Information Processing Systems*, 2016.

[42] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alexander A. Alemi. Inception-v4, inception-resnet and the impact of residual connections on learning. In *AAAI*, pages 4278–4284, 2017.

[43] Daiki Tanaka, Daiki Ikami, Toshihiko Yamasaki, and Kiyoharu Aizawa. Joint optimization framework for learning with noisy labels. In *CVPR*, pages 5552–5560, 2018.

[44] Antti Tarvainen and Harri Valpola. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30, pages 1195–1204. Curran Associates, Inc., 2017.

[45] Trung Q. Tran, Mingu Kang, and Daeyoung Kim. Rankingmatch: Delving into semi-supervised learning with consistency regularization and ranking loss, 2021.

[46] Arash Vahdat. Toward robustness against label noise in training deep discriminative neural networks. In *NIPS*, pages 5601–5610, 2017.

[47] Andreas Veit, Neil Alldrin, Gal Chechik, Ivan Krasin, Abhinav Gupta, and Serge J. Belongie. Learning from noisy large-scale datasets with minimal supervision. In *CVPR*, pages 6575–6583, 2017.

[48] Yisen Wang, Xingjun Ma, Zaiyi Chen, Yuan Luo, Jinfeng Yi, and James Bailey. Symmetric cross entropy for robust learning with noisy labels. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 322–330, 2019.

[49] Hongxin Wei, Lei Feng, Xiangyu Chen, and Bo An. 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 (CVPR)*, June 2020.

[50] Qi Wei, Haoliang Sun, Xiankai Lu, and Yilong Yin. Self-filtering: A noise-aware sample selection for label noise with confidence penalization. *ECCV*, 2022.

[51] Tong Xiao, Tian Xia, Yi Yang, Chang Huang, and Xiaogang Wang. Learning from massive noisy labeled data for image classification. In *CVPR*, pages 2691–2699, 2015.

[52] Tong Xiao, Tian Xia, Yi Yang, Chang Huang, and Xiaogang Wang. Learning from massive noisy labeled data for image classification. In *CVPR*, pages 2691–2699, 2015.

[53] Qizhe Xie, Zihang Dai Dai, Eduard Hovy, Minh-Thang Luong, and Quoc V. Le. Unsupervised data augmentation for consistency training. In *NeurIPS*, 2020.

[54] Yilun Xu, Peng Cao, Yuqing Kong, and Yizhou Wang. L_dmi: An information-theoretic noise-robust loss function. *NeurIPS, arXiv:1909.03388*, 2019.

[55] Yazhou Yao, Zeren Sun, Chuanyi Zhang, Fumin Shen, Qi Wu, Jian Zhang, and Zhenmin Tang. Jo-src: A contrastive approach for combating noisy labels. *CoRR*, abs/2103.13029, 2021.

[56] Kun Yi and Jianxin Wu. Probabilistic end-to-end noise correction for learning with noisy labels. In *CVPR*, 2019.

[57] Xingrui Yu, Bo Han, Jiangchao Yao, Gang Niu, Ivor W. Tsang, and Masashi Sugiyama. How does disagreement help generalization against label corruption? In *ICML*, pages 7164–7173, 2019.

[58] Chiyuan Zhang, Samy Bengio, Moritz Hardt, Benjamin Recht, and Oriol Vinyals. Understanding deep learning requires rethinking generalization. In *ICLR*, 2017.

[59] Ganlong Zhao, Guanbin Li, Yipeng Qin, Feng Liu, and Yizhou Yu. Centrality and consistency: Two-stage clean samples identification for learning with instance-dependent noisy labels. In *ECCV*, 2022.