Meta-Learning Dynamic Center Distance: Hard Sample Mining for Learning with Noisy Labels

Chenyu Mu, Yijun Qu, Jiexi Yan, Erkun Yang*, Cheng Deng
Xidian University
Xi'an, China
{cym9131,qyjtoki1,jxyan1995,erkunyang,chdeng.xd}@gmail.com

Abstract

The sample selection approach is a widely adopted strategy for learning with noisy labels, where examples with lower losses are effectively treated as clean during training. However, this clean set often becomes dominated by easy examples, limiting the model's meaningful exposure to more challenging cases and reducing its expressive power. To overcome this limitation, we introduce a novel metric called Dynamic Center Distance (DCD), which can quantify sample difficulty and provide information that critically complements loss values. Unlike approaches that rely on predictions, DCD is computed in feature space as the distance between sample features and a dynamically updated center, established through a proposed meta-learning framework. Building on preliminary semi-supervised training that captures fundamental data patterns, we incorporate DCD to further refine the classification loss, down-weighting well-classified examples and strategically focusing training on a sparse set of hard instances. This strategy prevents easy examples from dominating the classifier, leading to more robust learning. Extensive experiments across multiple benchmark datasets, including synthetic and real-world noise settings, as well as natural and medical images, consistently demonstrate the effectiveness of our method.

1. Introduction

In supervised learning, labeled data is a crucial resource for training effective models. However, acquiring high-quality labeled data is often challenging, leading to the introduction of noisy labels in many datasets [35, 44, 54]. This label noise poses significant challenges, especially as deep learning models are highly expressive and prone to overfitting—even when exposed to noisy labels [5, 7, 40, 43]. This tendency to overfit noise can severely degrade model performance, making the mitigation of label noise a key area of active research [2, 3, 29, 48, 52, 55, 56, 58].

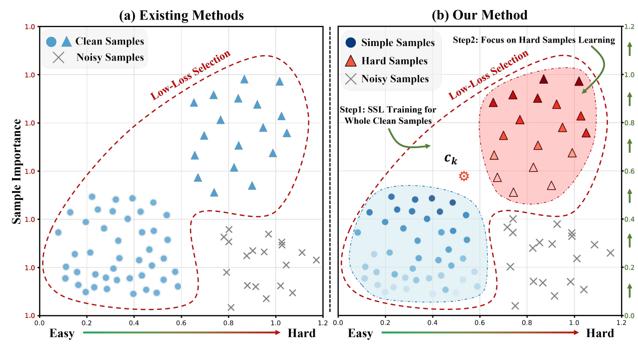


Figure 1. (a) Existing methods tend to prioritize simple, low-loss samples, potentially overlooking harder, more informative samples that could enhance model performance. (b) Our approach emphasizes the importance of selecting both easy and hard samples, resulting in a more robust model that learns from a broader range of data, mitigating overfitting and improving generalization.

Existing methods for handling label noise generally fall into two main categories: **statistically consistent** and **statistically inconsistent** approaches [30, 34, 57]. Statistically consistent methods model the label noise by estimating a noise transition matrix [12, 37]. While these methods are theoretically robust, their practical effectiveness diminishes under high noise levels or when dealing with datasets with large class counts, as accurately estimating the noise matrix becomes difficult in such conditions [16]. In contrast, statistically inconsistent approaches adopt various heuristic like memorization effect [14], co-teaching [59], label correction [63], and small-loss [13] sample selection to combat noisy labels. Most of these methods are implemented by constructing a subset of samples, often deemed less noisy. However, these methods tend to misidentify hard-to-learn samples as noisy, based on shared features such as high loss or low prediction confidence. As a result, they favor simpler, easier-to-learn samples, which can limit model learning by excluding informative ones.

*Corresponding author

Current methods for distinguishing between easy and hard samples also face limitations in high-noise scenarios [29, 49, 54]. For example, sample separation methods, such as those based on small-loss strategy [11, 17, 19], often fail to differentiate between genuinely noisy samples and those that are simply hard to learn. This misclassification leads to the exclusion of critical samples, adversely impacting model learning and generalization. Consequently, there is a growing need for advanced methods that can more accurately differentiate between noisy and informative samples, ensuring a more balanced learning process and the preservation of valuable information.

To address the aforementioned challenges, we propose DCD, an innovative method that incorporates with a meta-learning framework to dynamically differentiate easy and hard samples. DCD is computed as the Euclidean distance between each sample's feature representation and an evolving class center, which continuously adapts throughout training to capture each class's shifting feature distribution. By constructing category-specific center vectors from all samples, DCD achieves a detailed representation of each class, assigning weights based on sample distance from the center to prioritize harder, more distant examples. Additionally, to better estimate the true class centers, we introduce a weighted sample averaging technique with weights derived from a novel meta-learning approach. This weighted averaging substantially enhances the precision of class center estimates, allowing DCD to more accurately and reliably identify challenging and complex samples. In summary, these mechanisms enable the model to further focus on hard samples on the basis of the original semi-supervised training focusing on simple samples, thus improving the model's generalization in noisy label settings.

We evaluate DCD across six synthetic and real-world noisy benchmarks, including natural image classification datasets—CIFAR-10 [21], CIFAR-100 [21], Tiny-ImageNet [22], Clothing1M [54], and WebVision [26]—and a medical image segmentation dataset, PROMISE12 [28]. For synthetic noise, DCD achieves 65.3% classification accuracy on CIFAR-100 with 90% symmetric noise and 68.4% on Tiny-ImageNet with 50% symmetric noise, outperforming L2B [64] by 4.6% and 4.1%. In real-world noise settings, DCD achieves 78.34% on Clothing1M, surpassing the best baselines by 0.84%, respectively. Additionally, DCD improves performance in medical image segmentation, achieving a 1.98% gain on the PROMISE12 dataset.

The main contributions can be summarized as follows:

- We introduce DCD, a novel metric that measures sample difficulty by computing the Euclidean distance between sample features and a dynamically updated class center. This allows the model to focus on more challenging samples, improving robustness in noisy label settings.
- We propose a meta-learning framework that integrates DCD to dynamically adjust sample loss weights, allowing the model to prioritize hard examples and refine the feature centers during training. This enhances the model's ability to handle noisy labels more effectively.
- Achieves SOTA results on both synthetic/real-world noisy labels across natural and medical image datasets.

2. Related Work

Learning with Noisy Labels. Learning with noisy labels is widely studied, with methods divided into model-based and model-free approaches. Model-based methods estimate a noise transition matrix to recover an optimal classifier, but they struggle with heavy noise or many classes [13, 16, 47]. Model-free methods, such as noisy sample detection [38, 61] and pseudo-label refinement [24], aim to reduce noise impact. Despite progress, both face challenges in high-noise or data-scarce scenarios. Recent efforts in unsupervised contrastive learning [19, 25, 62] focus on learning robust features without label correction. Robust, scalable representation learning remains challenging due to noisy labels' complex ties to model generalization.

Hard Example Mining. Hard example mining is commonly used in machine learning to identify difficult samples for improved model training. Two primary approaches are employed: one for SVMs, where the working set of examples is dynamically updated to focus on hard examples violating the model's margin [10], and another for non-SVMs, where false positives are added to the training set after model convergence [6]. However, when applied to noisy label problems, these methods face challenges. Existing techniques, such as noise transition matrices or pseudo-labeling, often assume known noise transitions, which is unrealistic in practice [12, 37, 53]. To address this challenage, we introduce DCD, a metric that focuses training on harder examples by measuring distance to a dynamic class center, enhancing model performance in noisy label settings.

Meta Learning. Meta-learning has proven effective in enhancing model generalization. By leveraging a small clean set or auxiliary tasks, it optimizes model parameters and hyper-parameters efficiently, often through techniques like bi-level optimization and instance re-weighting [23, 41, 51]. These techniques enable better sample selection and weight assignment. Some methods treat label correction as a meta-process, where the model iteratively corrects noisy labels [51, 61, 63], while others prevent overfitting by learning robust loss functions [23]. Recent techniques, such as CMW-Net [42], adapt to data biases by generating adaptive sample weights, and DMLP [46] combines self-supervised learning with meta-learning for label correction. Unlike these approaches, our method integrates true and pseudo-label reweighting into meta-learning, optimizing per-sample loss weights and implicitly relabeling data.

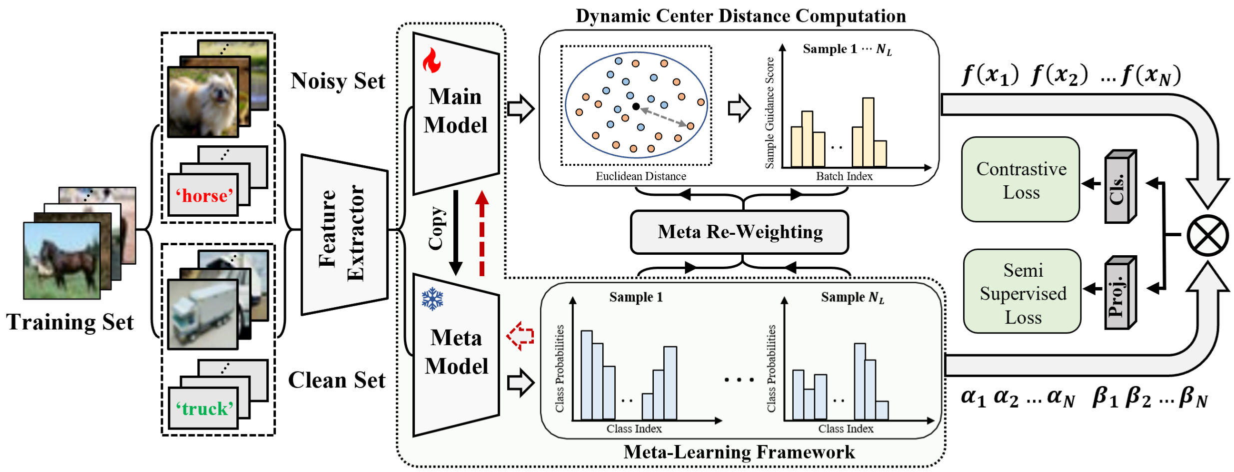


Figure 2. Overview of our model: First, sample selection splits the dataset into clean and noisy subsets. Sample importance is then computed using DCD based on distance to class centroids, which guides loss weighting. Meta-learning generates adaptive weights to refine labels. Finally, weighted features are fed into the classifier and projection head to compute loss and update the model.

3. Methodology

3.1. Preliminaries

Let the training set be $D_{\text{train}} = \{(x_i, \hat{\boldsymbol{y}}_i)\}_{i=1}^{N}$, where x_i is the i-th input image, $\hat{\boldsymbol{y}}_i$ is the corresponding observed noisy label. We aim to learn a robust model with this noisy training set, which can generalize well to unseen test examples. Specifically, we can employ a deep neural network (DNN) model $f(x;\theta)$ with parameters θ as a feature extractor and use a cross-entropy loss to obtain a noisy robust classification model:

$$\mathcal{L} = -\frac{1}{N_{cle}} \sum_{(x_i,\hat{y}_i)\in D_{cle}} \boldsymbol{t}_i \log \boldsymbol{p}_i, \tag{1}$$

where $\boldsymbol{t}_i$ is the target, which can be either the observed noisy label or a combination of noisy labels and pseudo-labels. The term $\boldsymbol{p}_i = \text{softmax}\left(f\left(x_i;\theta\right)\right)$ represents the predicted probability for the i-th sample. To handle noisy labels, existing methods try to construct a clean subset $D_{cle} \subseteq D_{train}$, where samples are assumed to be correctly labeled, and N_{cle} represents the size of this clean set. By training the model with this clean set, the impact of noisy labels can be significantly reduced.

To obtain the clean set, existing methods typically employ two strategies: (1) **sample selection**, which identifies examples with low loss values as likely correctly labeled; and (2) **label correction**, which adjusts the labels of examples exhibiting high predictive probabilities. Although these strategies differ in approach, both are grounded in shared indicators, such as low loss or high prediction confidence. Consequently, they primarily concentrate on "easy" examples, often neglecting more challenging cases [19]. Therefore, this selective focus can hinder the model's ability to capture complex patterns, ultimately constraining its representational capacity.

To address the above challenge, we propose DCD, a novel and innovative approach to tackle sample selection bias by focusing on key and representative samples despite label noise. DCD differentiates between hard-to-learn and noisy samples, prioritizing informative and crucial instances. In the next section, we introduce Meta Dynamic Center Adjustment, which refines sample selection and weight assignment to improve DCD's robustness and generalization.

3.2. Meta Dynamic Center Adjustment

In the presence of noisy labels, a major challenge is distinguishing noisy samples from genuinely informative hard samples. Traditional methods [19, 27, 39] often fail by treating all samples equally, neglecting the potential value of hard samples. To address this, our approach dynamically adjusts sample importance through Dynamic Center Distance (DCD) and meta-learning. DCD helps assess sample difficulty based on the distance to class centers, guiding the selection and weighting of samples. Meta-learning further refines this process by learning adaptive sample weights, ensuring better initialization. We elaborate on each of these components in the following.

DCD Calculation. First, we calculate the DCD to evaluate the difficulty of each sample. For a given sample x_i, its feature representation is denoted as $h(x_i)$. The DCD is defined as the Euclidean distance between this feature and the dynamic class center:

$$\text{DCD}(x_i) = \|h(x_i) - c_k\|, \tag{2}$$

where c_k represents the dynamic feature center of class k.

We can simply get c_k by averaging the features in the k-th class. However, since some examples may be wrongly labeled, directly averaging them is not an optimal solution.

To ensure that the class centers remain consistent with the true feature distribution, we propose a meta-learning framework to adaptively learn weights for each example and calculate a weighted mean as the class center. Moreover, since the sample weights are updated periodically with the changing of models, we denote the obtained centers as dynamic class centers. Assuming that the obtained weight for x_i is λ_i, then the dynamic class center for class k can be obtain with:

$$c_k = \frac{\sum_{i \in C_k} \lambda_i \cdot f(x_i;\theta)}{|C_k|}, \tag{3}$$

where C_k denotes the set of samples in class k and $|C_k|$ is the set size. The weight λ_i ensures critical samples have a greater impact on the center update and is updated with the following meta-learning framework.

Meta-Learning Driven Sample Weighting. Inspired by [38], we define the target $\boldsymbol{t}_i$ in Eq. (1) as a weighted combination of the noisy label and pseudo-label:

$$\boldsymbol{t}_i = \alpha_i \hat{\boldsymbol{y}}_i + \beta_i \tilde{\boldsymbol{y}}_i, \tag{4}$$

where $\tilde{\boldsymbol{y}}_i$ represents the pseudo-label derived from the model's current prediction, and α_i and β_i are the respective weights. In [38], the constraint $\alpha_i + \beta_i = 1$ is imposed, ensuring equal treatment for each sample. However, to account for the variability introduced by noisy labels, we relax this constraint, allowing for individual weight assignments.

To achieve precise sample selection and weight calculation, our method assigns adaptively initialized, meta-learned weights to each training sample. In each training iteration, given the dataset, these optimal weights are refined through meta-learning, representing the contributions of the observed noisy label $\hat{\boldsymbol{y}}$ and pseudo-label $\tilde{\boldsymbol{y}}$. By employing this meta-learning approach, the model dynamically adjusts α_i and β_i for each sample. Thus, the primary learning objective is defined by substituting Eq. (4) into Eq. (1):

$$\theta^*(\boldsymbol{\alpha},\boldsymbol{\beta}) = \arg\min_{\theta} \sum_{i=1}^{N} \alpha_i H(\boldsymbol{p}_i, \hat{\boldsymbol{y}}_i) + \beta_i H(\boldsymbol{p}_i, \tilde{\boldsymbol{y}}_i), \tag{5}$$

where $H(\cdot)$ is the cross-entropy loss, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are the weight vectors with α_i and β_i be the i-th element. Specifically, The weights $\alpha_i, \beta_i \geq 0$ are meta-learned based on performance on a validation set, allowing the model to dynamically adjust the relative contributions of the real label loss and pseudo-label loss. Unlike traditional guided loss methods, the sum of α_i and β_i does not have to equal 1, providing greater flexibility to maximize the model's performance.

The optimal weights are obtained by minimizing the validation loss on a clean validation set $\mathcal{D}_{\text{val}} = \{(x_i^{\text{val}}, y_i^{\text{val}})\}_{i=1}^{M}$:

$$\alpha^*, \beta^* = \arg\min_{\alpha,\beta \geq 0} \frac{1}{M} \sum_{i=1}^{M} H\left(\text{softmax}(f\left(x_i^{\text{v}}; \theta^*(\alpha,\beta)\right)), y_i^{\text{v}}\right). \tag{6}$$

Through the above dual optimization in Eq. (5) and (6), we can dynamically learn the model parameter θ and the weights for noisy labels α and pseudo-labels β. Details on the bi-level optimization are in *Supplementary Materials*.

Since α_i and β_i are weights for the noisy label and pseudo-label for x_i, the sum $(\alpha_i + \beta_i)$ can reflect the reliability of the sample x_i. Therefore, we set the weight λ_i in Eq. (3) for the sample x_i as:

$$\lambda_i = \alpha_i + \beta_i. \tag{7}$$

Based on the obtained weight, we recalculate the distance between the sample and the class center to obtain the Dynamic Center Distance:

$$\text{DCD}_{\text{meta}}(x_i) = \|h(x_i) - c_k\|. \tag{8}$$

Finally, we classify examples with a large DCD as hard examples, as they lie farther from the representation center. The importance weight for each example is then determined by normalizing the exponential of the DCD:

$$\Gamma_i = \frac{\exp\left(\text{DCD}_{\text{meta}}(x_i)\right)}{\sum_{j=1}^{N} \exp\left(\text{DCD}_{\text{meta}}(x_j)\right)}. \tag{9}$$

In Eq. (9), both the contributions of the meta-learned weights and the distance between each sample and its class center are taken into account. The DCD term modulates the relative influence of the real label loss and pseudo-label loss, while normalization based on DCD further refines these weights according to each sample's proximity to its class center. Meanwhile, we explain why λ_i is not used directly as the weight for x_i in Sec. 4.7.

3.3. Loss Function

After completing the weight calculation, we first select a clean subset with normal sample selection and label correction techniques [19]. Then, we employ a weighted cross-entropy to focus more on the hard-clean examples:

$$\mathcal{L}_{\mathcal{HF}} = -\sum_{i=1}^{N} \Gamma_i \cdot H\left(y_i, h\left(f(x_i;\theta_f);\theta_h\right)\right). \tag{10}$$

To integrate both clean and noisy samples in a semi-supervised learning (SSL) framework [4], we combine the supervised loss with additional loss terms to handle label noise and improve robustness. The overall training objective is then defined as:

$$\mathcal{L}_{\text{DCD}} = \mathcal{L}_{\mathcal{HF}} + \lambda_u \mathcal{L}_u + \lambda_r \mathcal{L}_r + \lambda_{\mathcal{C}} \mathcal{L}_{\mathcal{C}}, \tag{11}$$

where $\mathcal{C}$ is the contrastive loss. $\mathcal{L}_u$ and $\mathcal{L}_r$ handle unlabeled data and regularization, respectively. The hyperparameters λ_u, λ_r, and $\lambda_{\mathcal{C}}$ are set according to the experimental configuration in [19]. Details about the above SSL are provided in the *Supplementary Materials*.

4. Experiments

4.1. Datasets

CIFAR-10/100 [21]: CIFAR-10/100 (50K train, 10K test) use synthetic noise for robustness evaluation: symmetric noise (random class reassignment) and asymmetric noise (realistic mistakes, e.g., Truck→Auto; CIFAR-100 flips within superclasses), due to natural data's inherent noise.

Tiny-ImageNet [22]: This dataset is a smaller version of the original ImageNet in terms of the number of classes and the image resolution. There are in total 200 classes containing 500 images per class. The image size is 64 × 64.

Clothing1M [54]: Clothing1M is a large-scale real-world dataset with noisy labels. It contains 1M images from 14 different cloth-related classes. Since the labels are produced by the seller provided surrounding texts of the images, a large portion of confusing classes (e.g., Knitwear and Sweater) are mislabeled.

Webvision [26]: This dataset contains 2.4 million images (obtained from Flickr and Google) that are categorized into the same 1,000 classes as in the ImageNet ILSVRC12. Following the previous studies [24], we use the first 50 classes of the Google image subset as the training data.

PROMISE12 [28]: The dataset consists of 50 training cases of transversal T2-weighted MR images of the prostate from multiple clinical centers and vendors, with varying acquisition protocols, slice thickness, and endorectal coil usage. It includes both benign prostate conditions and prostate cancer cases.

4.2. Training Details

We utilize the PreAct ResNet18 architecture for CIFAR-10 and CIFAR-100 datasets. For these tasks, we implement the SGD optimizer with the following configuration: an initial learning rate (LR) of 0.02, a weight decay of $5e^{-4}$, a momentum of 0.9, and a batch size of 64. Each network is trained for approximately 350 epochs, with a linear learning rate decay of 0.1 every 120 epochs. A warmup period of 30 epochs is applied prior to the main training phase. We set λ_u, λ_r, and $\lambda_{\mathcal{C}}$ as 30, 1, and 0.025 for all datasets.

In the case of Tiny-ImageNet, we employ the ResNet50 architecture, starting with an initial LR of 0.01, a weight decay of $1e^{-3}$, and a batch size of 32. The network is trained for 350 epochs, with a learning rate decay of 0.1 every 100 epochs and a warmup period of 15 epochs. For the Clothing1M dataset, we adopt a similar approach, using an initial LR of 0.002 and a weight decay of $1e^{-3}$, training the model for 100 epochs with a lr-decay of 0.1 every 40 epochs. The same settings are applied for the WebVision dataset to ensure consistency in performance.

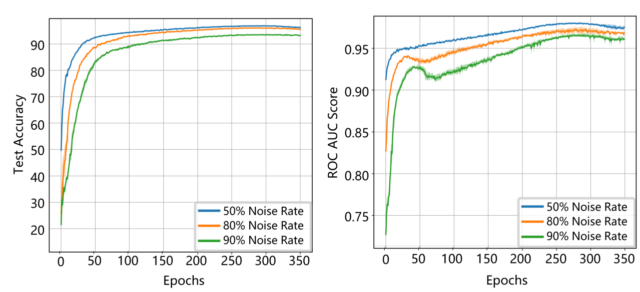
Data augmentation is performed following the Auto-augment policy, as specified in previous works. For CIFAR-10 and CIFAR-100, we use the CIFAR-10-Policy, while the ImageNet-Policy is applied to Tiny-ImageNet. Due to the transferable nature of these policies, the ImageNet-Policy is also utilized for both Clothing1M and WebVision datasets, optimizing the model's generalization capabilities across different datasets.

4.3. Clean Meta Data Construction

The clean meta-dataset D_{cle} constitutes 2% of the entire dataset in all experiments [64]. This ratio was determined through preliminary experiments on CIFAR-10 under 90% symmetric label noise, where 2% demonstrated optimal cost efficiency in balancing test accuracy gains against computational overhead. We define the Cost Efficiency (CE) metric as CE $= \Delta$Accuracy$/(\Delta$Time $\times$ ΔMemory $\times$ ΔMeta), where Δ denotes relative change versus the 1% baseline. As shown in Table 2, increasing beyond 2% yields diminishing returns in test accuracy while substantially increasing resource demands.

For datasets such as CIFAR-10/100 and Tiny-ImageNet, clean data were randomly sampled from the training set and removed, while for Clothing1M and Webvision, clean data were selected from the test set and removed accordingly. The overall data split follows an 8:1:1 ratio for training, testing, and validation.

4.4. Performance Comparisons



(a) Test Accuracy (%) (b) ROC AUC Score

Figure 3. Test accuracy (%) and ROC-AUC performance on CIFAR-10 across varying noise levels. As the model becomes more precise, the test-time performance improves accordingly. The shaded region in the figure represents the error margin from four experiments, while the solid line denotes the mean value.

Noise Mode	Symmetric noise								Asymmetric noise					
Dataset	**CIFAR-10**				**CIFAR-100**				**CIFAR-10**			**CIFAR-100**		
Method	20%	50%	80%	90%	20%	50%	80%	90%	10%	30%	40%	10%	30%	40%
Standard CE	86.8	79.4	62.9	42.7	62.0	46.7	19.9	10.1	88.8	81.7	76.1	68.1	53.3	44.5
DMix [24]	96.1	94.6	92.9	76.0	77.3	74.6	60.2	31.5	93.8	92.5	91.7	71.6	69.5	55.1
ELR [29]	95.8	94.8	93.3	78.7	77.6	73.6	60.8	33.4	95.4	94.7	93.0	77.3	74.6	73.2
UNION [19]	96.0	95.6	93.9	90.8	78.9	77.6	63.9	44.8	95.3	94.8	94.1	78.2	75.6	74.8
SFT [9]	92.6	88.5	42.1	10.5	71.9	65.5	27.0	7.9	-	-	-	-	-	-
Sel-CL+ [8]	95.5	93.9	89.2	81.9	76.5	72.4	59.6	48.8	95.6	94.5	93.4	78.7	76.4	74.2
TCL [17]	94.9	93.8	92.3	89.2	77.8	73.1	64.7	47.9	95.3	94.9	92.5	78.3	75.5	75.3
OT-Filter [11]	96.0	95.3	94.0	90.5	76.7	73.8	61.8	42.8	-	-	-	-	-	-
L2B [64]	96.7	95.6	94.8	94.4	80.1	78.1	69.6	60.7	95.1	94.7	94.0	78.5	75.4	75.1
DCD	**96.8**	**96.2**	**95.3**	**95.1**	**83.1**	**80.7**	**73.3**	**65.3**	**96.1**	**95.8**	**95.6**	**80.1**	**78.6**	**78.4**
	±0.17	±0.21	±0.24	±0.29	±0.22	±0.28	±0.35	±0.38	±0.12	±0.17	±0.22	±0.11	±0.18	±0.25
Δ	↑ 0.1	↑ 0.6	↑ 0.5	↑ 0.7	↑ 3.0	↑ 2.6	↑ 3.7	↑ 4.6	↑ 0.5	↑ 0.9	↑ 1.5	↑ 1.4	↑ 2.2	↑ 3.1

Table 1. Test accuracies (%) achieved by various techniques under both symmetric and asymmetric noise conditions. The highest accuracy values are highlighted in green, while the second-highest results are shaded in gray. This color scheme is consistently applied across all tables in the following sections. Different baselines use varying datasets. We compare each dataset with the baseline that utilized it.

Meta Data Size	Accuracy	Time/Epoch	Memory	CE
1%	94.3%	0.7 min	5.3 GB	-
2%	95.1%	0.9 min	6.9 GB	2.50
4%	95.5%	1.3 min	9.4 GB	0.20
6%	95.7%	2.1 min	11.2 GB	0.07

Table 2. Performance comparison with different meta-data size

Noise (%)	0		20		50	
Alg.	Best	Avg.	Best	Avg.	Best	Avg.
Standard CE	57.4	56.7	35.8	35.6	19.8	19.6
F-correction [36]	-	-	44.5	44.4	33.1	32.8
MentorNet [18]	-	-	45.7	45.5	35.8	35.5
Co-teaching+ [59]	52.4	52.1	48.2	47.7	41.8	41.2
M-correction [1]	57.7	57.2	57.2	56.6	51.6	51.3
NCT [39]	62.4	61.5	58.0	57.2	47.8	47.4
UNION [19]	63.1	62.7	59.2	58.4	52.7	52.4
DISC [27]	68.5	68.2	67.9	67.5	64.3	63.9
DCD	**70.3**	**69.9**	**70.2**	**69.6**	**68.4**	**67.7**
Δ	↑ 1.8	↑ 1.7	↑ 2.3	↑ 2.1	↑ 4.1	↑ 3.8

Table 3. Test accuracies (%) on Tiny-ImageNet dataset under symmetric noise settings.

CIFAR-10 and CIFAR-100 Datasets: The experimental results on CIFAR-10 and CIFAR-100 datasets, demonstrating that DCD consistently achieves the highest test accuracies across different noise levels under both symmetric and asymmetric label noise, as shown in Table 1. For CIFAR-10, DCD outperforms other methods under symmetric noise, reaching 95.1% accuracy at 90% noise level, surpassing L2B [64] by 0.7%. In CIFAR-100, DCD achieves 65.3% accuracy at 90% noise, significantly exceeding L2B [64]. Under asymmetric noise, DCD also excels, achieving 95.6% accuracy at 40% noise in CIFAR-10 and showing improvements of 1.5% to 3.1% in CIFAR-100 across various noise rates. These results highlight DCD's robustness in handling class-dependent noise. We also visualized DCD's test accuracy (%) and ROC-AUC performance under different noise levels on CIFAR-10, as shown in Figure 3a and 3b, demonstrating DCD's robustness and

Method	Backbone	Test Accuracy
Standard CE	ResNet-50	69.21
OT-Filter [11]	ResNet-50	74.51
DMix [24]	ResNet-50	74.76
ELR [29]	ResNet-50	74.81
TCL [17]	ResNet-50	74.85
UNION [19]	ResNet-50	74.98
SFT [9]	ResNet-50	75.08
SNSCL [50]	ResNet-50	75.31
L2B [64]	ResNet-50	77.50
DCD	ResNet-50	**78.34** $^{\pm 0.17}$
Δ	-	↑ 0.84

Table 4. Experimental results on Clothing1M dataset. Results for previous techniques were copied from their respective papers.

Dataset	WebVision		ILSVRC12	
Method	*Top-1*	*Top-5*	*Top-1*	*Top-5*
D2L [31]	62.68	84.00	57.80	81.36
DivideMix [24]	77.32	91.64	75.20	90.84
ELR [29]	77.78	91.68	70.29	89.76
UNION [19]	77.60	93.44	75.29	93.72
TCL [17]	79.12	92.31	75.41	92.43
Sel-CL+ [8]	79.96	92.64	76.84	93.04
DISC [27]	80.28	92.28	77.44	92.28
LSL [20]	81.40	93.00	77.00	91.84
DCD	**83.29** $^{\pm 0.28}$	**94.51** $^{\pm 0.17}$	**78.54** $^{\pm 0.31}$	**94.17** $^{\pm 0.27}$
Δ	↑ 1.89	↑ 1.07	↑ 1.10	↑ 0.45

Table 5. Experimental results on Webvision and ILSVRC12. All methods are trained on the Webvision while evaluated on both Webvsion and ILSVRC12 validation set.

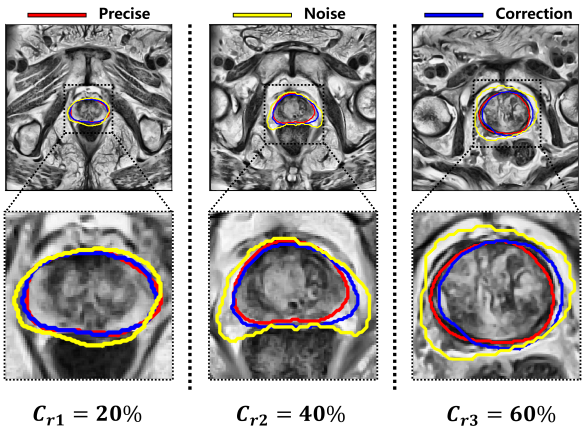


Figure 4. Experimental results under different corrupted ratios C_r of 20%, 40%, and 60% with precise (red outline), noise (yellow outline), and correction (blue outline).

Method	Dice (%)↑↑	HD (voxel)↓	ASD (voxel)↓
UNet++ [65]	73.74	11.63	3.70
NL re-weighting [32]	76.64	8.33	2.75
CRF [33]	78.00	7.45	2.57
L2B [64]	80.83	6.68	2.10
DCD	**82.81**	**6.52**	**1.78**
Δ	↑ 1.98	↓ 0.16	↓ 0.32

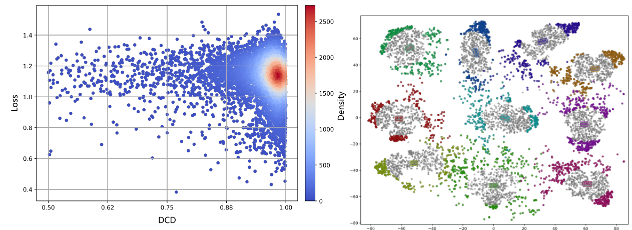
Table 6. Segmentation performance comparison under noisy-supervision on PROMISE12.

superior performance. This strong performance arises from its dynamic adaptation to noise, focusing on clean samples for learning and enhancing feature representation, ultimately improving model accuracy and robustness.

Tiny ImageNet Dataset: Table 3 compares the performance of DCD with other state-of-the-art methods. Tiny-ImageNet is challenging, and its difficulty increases with label noise. Among baselines, UNION [19] uses a consistent selection mechanism with contrastive learning, while DISC [27] applies dynamic instance selection for noise robustness. However, both methods underperform compared to DCD. As shown in Table 3, DCD achieves 1.8% to 4.1% higher accuracy than DISC, depending on label noise levels.

Clothing1M Dataset: Table 4 provides a comparison of performance on this real-world dataset with noisy labels. Our method outperforms L2B [64] by 0.84%. For Clothing1M, performance gains may fluctuate based on the warm-up period's length, as prolonged cross-entropy (CE)-based training can lead to memorization effects. In our approach, we apply a warm-up period of 2,000 steps.

WebVision Dataset: Table 5 demonstrates DCD's superior performance, establishing new state-of-the-art Top-1/Top-5 accuracy on Webvision and ILSVRC12. On Webvision, DCD outperforms LSL [20] with 1.89% and 1.51% gains in Top-1/Top-5 accuracy; on ILSVRC12, it achieves 1.10% and 0.45% improvements over prior baselines. These results confirm DCD's enhanced robustness against label noise compared to existing methods.



(a) DCD vs Loss. (b) Visualize hard examples

Figure 5. Visualizing independence of DCD and hard examples.

Method	Dice (%)↑
baseline - C_{r1}	82.01
DCD - C_{r1}	**84.72**
baseline - C_{r2}	80.83
DCD - C_{r2}	**82.81**
baseline - C_{r3}	77.70
DCD - C_{r3}	**79.68**

Table 7. L2B for segmentation under different noise levels.

4.5. Generalization to Image Segmentation

We evaluated the performance of our method under noisy supervision using the PROMISE12 dataset and compared it with several state-of-the-art approaches, including UNet++ [65], NL re-weighting [32], Mix-up [60], and L2B [64]. As summarized in Table 6, our approach outperforms all others in key evaluation metrics such as Dice score, Hausdorff Distance (HD) [45], and Average Surface Distance (ASD) [15]. Specifically, our model achieved a Dice score of 82.81%, with improvements in HD and ASD as well, demonstrating its robustness under noisy conditions.

Furthermore, we evaluate the robustness of our method under different noise levels by varying the corrupted ratios of the training set to C_{r1}, C_{r2}, and C_{r3}, which represent corruption rates of 20%, 40%, and 60%, respectively. As shown in Table 7, we compared the baseline model [64], which is trained directly on the noisy datasets, with our DCD approach across these varying noise levels. The Dice scores achieved by DCD were 84.72% for C_{r1}, 82.81% for C_{r2}, and 79.68% for C_{r3}. In contrast, the baseline scores dropped significantly from 82.01% to 77.70%. This indicates that DCD maintains high performance and is robust even as the noise levels increase, demonstrating significant improvements, particularly under severe noise conditions.

4.6. Analyzing Independence of DCD

We validated DCD as an independent measure of sample hardness through scatter plot analysis (Figure 5a), which showed no direct correlation with classification loss be-

Dataset	CIFAR10						CIFAR100					
Noise Rate	50%		80%		90%		50%		80%		90%	
Method	Best	Last	Best	Last	Best	Last	Best	Last	Best	Last	Best	Last
DCD w/o meta-pipeline	95.72	95.68	94.73	94.31	93.82	93.71	79.32	79.15	72.16	71.98	63.91	63.75
DCD w/o $\mathcal{D}$	95.81	95.75	94.65	94.37	93.03	92.83	78.48	78.27	68.74	68.51	63.30	63.14
DCD w/o CL	95.51	95.49	94.93	94.82	94.23	94.11	79.82	79.53	72.28	72.17	64.73	64.57
DCD	**96.28**	**96.17**	**95.31**	**95.25**	**95.14**	**95.07**	**80.74**	**80.69**	**73.35**	**73.24**	**65.32**	**65.19**

Table 8. Ablation study with different training settings. Both the meta-pipeline and DCD components significantly affect the performance, particularly under high noise rates. Removing contrastive loss (CL) also leads to noticeable performance degradation, especially as noise levels increase. Test results at last epoch are also shown here.

Method	20%	50%	80%
baseline (OT-Filter)	76.7	73.8	61.8
$\boldsymbol{\alpha} = \mathbf{0}$	72.9	71.9	60.9
$\boldsymbol{\beta} = \mathbf{0}$	78.3	75.8	64.8
$\boldsymbol{\alpha} + \boldsymbol{\beta} = \mathbf{1}$	81.6	78.6	72.2
$\boldsymbol{\alpha}, \boldsymbol{\beta} \geq \mathbf{0}$	**83.1**	**80.7**	**73.3**

Table 9. Ablation of α, β under diverse symmetric noise in CIFAR-100. $\boldsymbol{\alpha}, \boldsymbol{\beta} \geq \mathbf{0}$ achieves best performance.

Symmetric noise	20%	40%	60%
Direct weight (λ_i)	94.82±0.21	93.23±0.27	91.94±0.33
DCD (Γ_i)	**96.78±0.17**	**96.38±0.18**	**95.61±0.23**

Table 10. Compare Γ_i with direct weight λ_i in CIFAR-10.

tween epochs 210-220. Our method uses DCD to prioritize high-DCD samples for better generalization and down-weight low-DCD samples for increased robustness. Figure 5b illustrates this mechanism: simple samples (gray circles) are confidently classified far from decision boundaries, while hard samples (colored circles) cluster near boundaries with ambiguous class membership and lower confidence. This confirms DCD's role in identifying boundary-proximate hard samples, aligning with adversarial training to enhance model robustness.

4.7. Compare with direct weight λ_i.

We conducted additional experiments exploring the direct use of the meta-learning weights λ_i for sample selection. Experimental results show that while the direct application of these weights for sample selection yields improvements, the enhancements are relatively marginal compared to the original method. This is reflected in the reduced variance of the results, as shown in Table 10, which demonstrates that the original method provides more stable and consistent outcomes under various noise levels.

4.8. Ablation Study

Effect of Meta Learning: Removing meta-learning from DCD degrades performance across all noise levels (Table 8). At 90% noise, CIFAR-10 accuracy drops by 1.32% (best) and 1.36% (last), while CIFAR-100 shows 1.41% and 1.44% reductions respectively.

Effect of DCD: The effect of removing the Dynamic Center Distance (DCD) component is also prominent in Table 8. Without DCD, the model experiences a performance decline at each noise level. In CIFAR-10, while the reduction in accuracy is subtle, especially at moderate noise rates (50% and 80%), it becomes more pronounced at 90% noise, with a decrease of about 2.11% in the best accuracy compared to the full DCD configuration. For CIFAR-100, excluding DCD leads to a 2.02% decrease in the best accuracy at 90% noise. These results suggest that DCD plays a critical role in maintaining model accuracy, particularly under higher noise conditions, by enhancing feature learning and robustness against noisy labels.

Significance of α and β: Table 9 shows that $\alpha = 0$ causes minor degradation, while $\beta = 0$ improves over baseline but underperforms joint optimization. Constraining $\alpha + \beta = 1$ yields moderate improvements but restricts flexibility. Independent α/β optimization achieves optimal performance through maximized balance and adaptability.

5. Conclusion

We propose DCD, an innovative method that incorporates a meta-learning framework to address noisy labels and sample difficulty. By dynamically prioritizing challenging samples based on distance from evolving class centers, DCD improves training robustness. Extensive experiments across multiple benchmark datasets, including synthetic and real-world noise settings, and natural and medical images, demonstrate the effectiveness of our method, promising solution for robust learning in noisy environments.

6. Acknowledgments

Our work was supported in part by the National Natural Science Foundation of China (62132016 and 62202365), Natural Science Basic Research Program of Shaanxi (2020JC-23), Fundamental Research Funds for the Central Universities (QTZX23073), and Young Elite Scientists Sponsorship Program by CAST (2023QNRC001).

References

[1] Eric Arazo, Diego Ortego, Paul Albert, Noel O'Connor, and Kevin McGuinness. Unsupervised label noise modeling and loss correction. In *ICML*, pages 312–321, 2019. 6

[2] Yingbin Bai, Erkun Yang, Bo Han, Yanhua Yang, Jiatong Li, Yinian Mao, Gang Niu, and Tongliang Liu. Understanding and improving early stopping for learning with noisy labels. *NeurIPS*, 34:24392–24403, 2021. 1

[3] Yingbin Bai, Erkun Yang, Zhaoqing Wang, Yuxuan Du, Bo Han, Cheng Deng, Dadong Wang, and Tongliang Liu. Rsa: reducing semantic shift from aggressive augmentations for self-supervised learning. *NeurIPS*, 35:21128–21141, 2022. 1

[4] David Berthelot, Nicholas Carlini, Ian Goodfellow, Nicolas Papernot, Avital Oliver, and Colin A Raffel. Mixmatch: A holistic approach to semi-supervised learning. *NeurIPS*, 32, 2019. 4

[5] Xuan Chen, Likai Wang, Wenlong Li, Yuchen Yang, Erkun Yang, and Cheng Deng. Mitigating data imbalance and noise: A divergence-based approach with enhanced sample selection. *Neurocomputing*, 605:128269, 2024. 1

[6] Piotr Dollár, Zhuowen Tu, Pietro Perona, and Serge J Belongie. Integral channel features. In *Bmvc*, page 5, 2009. 2

[7] Chengyu Dong, Liyuan Liu, and Jingbo Shang. Label noise in adversarial training: A novel perspective to study robust overfitting. *NeurIPS*, 35:17556–17567, 2022. 1

[8] Li et al. Selective-supervised contrastive learning with noisy labels. In *CVPR*, 2022. 6

[9] Wei et al. Self-filtering: A noise-aware sample selection for label noise with confidence penalization. In *ECCV*, 2022. 6

[10] Pedro F Felzenszwalb, Ross B Girshick, David McAllester, and Deva Ramanan. Object detection with discriminatively trained part-based models. *IEEE Trans. Pattern Anal. Mach. Intell.*, 32(9):1627–1645, 2009. 2

[11] Chuanwen Feng, Yilong Ren, and Xike Xie. Ot-filter: An optimal transport filter for learning with noisy labels. In *CVPR*, pages 16164–16174, 2023. 2, 6

[12] Jacob Goldberger and Ehud Ben-Reuven. Training deep neural-networks using a noise adaptation layer. In *ICLR*, 2017. 1, 2

[13] Bo Han, Quanming Yao, Xingrui Yu, Gang Niu, Miao Xu, Weihua Hu, Ivor Tsang, and Masashi Sugiyama. Co-teaching: Robust training of deep neural networks with extremely noisy labels. *NeurIPS*, 31, 2018. 1, 2

[14] Gang Han, Wenping Guo, Haibo Zhang, Jie Jin, Xingli Gan, and Xiaoming Zhao. Sample self-selection using dual teacher networks for pathological image classification with noisy labels. *Computers in Biology and Medicine*, page 108489, 2024. 1

[15] Tobias Heimann, Bram Van Ginneken, Martin A Styner, Yulia Arzhaeva, Volker Aurich, Christian Bauer, Andreas Beck, Christoph Becker, Reinhard Beichel, György Bekes, et al. Comparison and evaluation of methods for liver segmentation from ct datasets. *IEEE Trans. Med. Imag.*, 28(8):1251–1265, 2009. 7

[16] Dan Hendrycks, Mantas Mazeika, Duncan Wilson, and Kevin Gimpel. Using trusted data to train deep networks on labels corrupted by severe noise. *NeurIPS*, 31, 2018. 1, 2

[17] Zhizhong Huang, Junping Zhang, and Hongming Shan. Twin contrastive learning with noisy labels. In *CVPR*, pages 11661–11670, 2023. 2, 6

[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 2304–2313, 2018. 6

[19] Nazmul Karim, Mamshad Nayeem Rizve, Nazanin Rahnavard, Ajmal Mian, and Mubarak Shah. Unicon: Combating label noise through uniform selection and contrastive learning. In *CVPR*, pages 9676–9686, 2022. 2, 3, 4, 5, 6, 7

[20] Noo-ri Kim, Jin-Seop Lee, and Jee-Hyong Lee. Learning with structural labels for learning with noisy labels. In *CVPR*, pages 27610–27620, 2024. 6, 7

[21] Alex Krizhevsky, Geoffrey Hinton, et al. Learning multiple layers of features from tiny images. 2009. 2, 5

[22] Ya Le and Xuan Yang. Tiny imagenet visual recognition challenge. *CS 231N*, 7(7):3, 2015. 2, 5

[23] Junnan Li, Yongkang Wong, Qi Zhao, and Mohan S Kankanhalli. Learning to learn from noisy labeled data. In *CVPR*, pages 5051–5059, 2019. 2

[24] Junnan Li, Richard Socher, and Steven CH Hoi. Dividemix: Learning with noisy labels as semi-supervised learning. *ICLR*, 2020. 2, 5, 6

[25] Junnan Li, Caiming Xiong, and Steven CH Hoi. Learning from noisy data with robust representation learning. In *ICCV*, pages 9485–9494, 2021. 2

[26] Wen Li, Limin Wang, Wei Li, Eirikur Agustsson, and Luc Van Gool. Webvision database: Visual learning and understanding from web data. *arXiv preprint arXiv:1708.02862*, 2017. 2, 5

[27] Yifan Li, Hu Han, Shiguang Shan, and Xilin Chen. Disc: Learning from noisy labels via dynamic instance-specific selection and correction. In *CVPR*, pages 24070–24079, 2023. 3, 6, 7

[28] Geert Litjens, Robert Toth, Wendy Van De Ven, Caroline Hoeks, Sjoerd Kerkstra, Bram Van Ginneken, Graham Vincent, Gwenael Guillard, Neil Birbeck, Jindang Zhang, et al. Evaluation of prostate segmentation algorithms for mri: the promise12 challenge. *MedIA*, 18(2):359–373, 2014. 2, 5

[29] Sheng Liu, Jonathan Niles-Weed, Narges Razavian, and Carlos Fernandez-Granda. Early-learning regularization prevents memorization of noisy labels. *NeurIPS*, 33:20331–20342, 2020. 1, 2, 6

[30] Sheng Liu, Zhihui Zhu, Qing Qu, and Chong You. Robust training under label noise by over-parameterization. In *ICML*, pages 14153–14172, 2022. 1

[31] Xingjun Ma, Yisen Wang, Michael E Houle, Shuo Zhou, Sarah Erfani, Shutao Xia, Sudanthi Wijewickrema, and James Bailey. Dimensionality-driven learning with noisy labels. In *ICML*, pages 3355–3364, 2018. 6

[32] Zahra Mirikharaji, Yiqi Yan, and Ghassan Hamarneh. Learning to segment skin lesions from noisy annotations. In *MICCAI*, pages 207–215, 2019. 7

[33] Miguel Monteiro, Mário AT Figueiredo, and Arlindo L Oliveira. Conditional random fields as recurrent neural networks for 3d medical imaging segmentation. *arXiv preprint arXiv:1807.07464*, 2018. 7

[34] Chenyu Mu, Erkun Yang, and Cheng Deng. Meta-guided adaptive weight learner for noisy correspondence. In *SIGIR*, pages 968–978, 2025. 1

[35] David F Nettleton, Albert Orriols-Puig, and Albert Fornells. A study of the effect of different types of noise on the precision of supervised learning techniques. *Artificial intelligence review*, 33:275–306, 2010. 1

[36] Giorgio Patrini, Alessandro Rozza, Aditya Krishna Menon, Richard Nock, and Lizhen Qu. Making deep neural networks robust to label noise: A loss correction approach. In *CVPR*, pages 1944–1952, 2017. 6

[37] Giorgio Patrini, Alessandro Rozza, Aditya Krishna Menon, Richard Nock, and Lizhen Qu. Making deep neural networks robust to label noise: A loss correction approach. In *CVPR*, pages 1944–1952, 2017. 1, 2

[38] Scott Reed, Honglak Lee, Dragomir Anguelov, Christian Szegedy, Dumitru Erhan, and Andrew Rabinovich. Training deep neural networks on noisy labels with bootstrapping. *arXiv preprint arXiv:1412.6596*, 2014. 2, 4

[39] Fahad Sarfraz, Elahe Arani, and Bahram Zonooz. Noisy concurrent training for efficient learning under label noise. In *WACV*, pages 3159–3168, 2021. 3, 6

[40] Xiaoshuang Shi, Zhenhua Guo, Kang Li, Yun Liang, and Xiaofeng Zhu. Self-paced resistance learning against overfitting on noisy labels. *Pattern Recognition*, 134:109080, 2023. 1

[41] Jun Shu, Qi Xie, Lixuan Yi, Qian Zhao, Sanping Zhou, Zongben Xu, and Deyu Meng. Meta-weight-net: Learning an explicit mapping for sample weighting. *NeurIPS*, 32, 2019. 2

[42] Jun Shu, Xiang Yuan, Deyu Meng, and Zongben Xu. Cmw-net: Learning a class-aware sample weighting mapping for robust deep learning. *IEEE Trans. Pattern Anal. Mach. Intell.*, 45(10):11521–11539, 2023. 2

[43] Sara Silva, Leonardo Vanneschi, Ana IR Cabral, and Maria J Vasconcelos. A semi-supervised genetic programming method for dealing with noisy labels and hidden overfitting. *Swarm and evolutionary computation*, 39:323–338, 2018. 1

[44] Zibo Su, Kaiqi Zhao, Zihao Fan, and Xinhua Guo. Rtpv-yolo: Real-time photovoltaic detection with uav-based thermal and rgb imaging. *IEEE Transactions on Aerospace and Electronic Systems*, pages 1–12, 2025. 1

[45] Abdel Aziz Taha and Allan Hanbury. An efficient algorithm for calculating the exact hausdorff distance. *IEEE Trans. Pattern Anal. Mach. Intell.*, 37(11):2153–2163, 2015. 7

[46] Yuanpeng Tu, Boshen Zhang, Yuxi Li, Liang Liu, Jian Li, Yabiao Wang, Chengjie Wang, and Cai Rong Zhao. Learning from noisy labels with decoupled meta label purifier. In *CVPR*, pages 19934–19943, 2023. 2

[47] Jialu Wang, Eric Xin Wang, and Yang Liu. Estimating instance-dependent label-noise transition matrix using a deep neural network. In *ICML*, 2022. 2

[48] Xiaosong Wang, Ziyue Xu, Dong Yang, Leo Tam, Holger Roth, and Daguang Xu. Learning image labels on-the-fly for training robust classification models. *arXiv preprint arXiv:2009.10325*, 2020. 1

[49] Hongxin Wei, Lue Tao, Renchunzi Xie, and Bo An. Open-set label noise can improve robustness against inherent label noise. *NeurIPS*, 34:7978–7992, 2021. 2

[50] Qi Wei, Lei Feng, Haoliang Sun, Ren Wang, Chenhui Guo, and Yilong Yin. Fine-grained classification with noisy labels. In *CVPR*, pages 11651–11660, 2023. 6

[51] Yichen Wu, Jun Shu, Qi Xie, Qian Zhao, and Deyu Meng. Learning to purify noisy labels via meta soft label corrector. In *AAAI*, pages 10388–10396, 2021. 2

[52] Zhi-Fan Wu, Tong Wei, Jianwen Jiang, Chaojie Mao, Mingqian Tang, and Yu-Feng Li. Ngc: A unified framework for learning with open-world noisy data. In *ICCV*, pages 62–71, 2021. 1

[53] Xiaobo Xia, Tongliang Liu, Nannan Wang, Bo Han, Chen Gong, Gang Niu, and Masashi Sugiyama. Are anchor points really indispensable in label-noise learning? *NeurIPS*, 32, 2019. 2

[54] 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. 1, 2, 5

[55] Erkun Yang, Tongliang Liu, Cheng Deng, Wei Liu, and Dacheng Tao. Distillhash: Unsupervised deep hashing by distilling data pairs. In *CVPR*, pages 2946–2955, 2019. 1

[56] Erkun Yang, Dongren Yao, Tongliang Liu, and Cheng Deng. Mutual quantization for cross-modal search with noisy labels. In *CVPR*, pages 7551–7560, 2022. 1

[57] Yuchen Yang, Likai Wang, Erkun Yang, and Cheng Deng. Robust noisy correspondence learning with equivariant similarity consistency. In *CVPR*, pages 17700–17709, 2024. 1

[58] Yazhou Yao, Zeren Sun, Chuanyi Zhang, Fumin Shen, Qi Wu, Jian Zhang, and Zhenmin Tang. Jo-src: A contrastive approach for combating noisy labels. In *CVPR*, pages 5192–5201, 2021. 1

[59] Xingrui Yu, Bo Han, Jiangchao Yao, Gang Niu, Ivor Tsang, and Masashi Sugiyama. How does disagreement help generalization against label corruption? In *ICML*, pages 7164–7173, 2019. 1, 6

[60] Hongyi Zhang. mixup: Beyond empirical risk minimization. *ICLR*, 2018. 7

[61] Zizhao Zhang, Han Zhang, Sercan O Arik, Honglak Lee, and Tomas Pfister. Distilling effective supervision from severe label noise. In *CVPR*, pages 9294–9303, 2020. 2

[62] Evgenii Zheltonozhskii, Chaim Baskin, Avi Mendelson, Alex M Bronstein, and Or Litany. Contrast to divide: Self-supervised pre-training for learning with noisy labels. In *WACV*, pages 1657–1667, 2022. 2

[63] Guoqing Zheng, Ahmed Hassan Awadallah, and Susan Dumais. Meta label correction for noisy label learning. In *AAAI*, pages 11053–11061, 2021. 1, 2

[64] Yuyin Zhou, Xianhang Li, Fengze Liu, Qingyue Wei, Xuxi Chen, Lequan Yu, Cihang Xie, Matthew P Lungren, and Lei Xing. L2b: Learning to bootstrap robust models for combating label noise. In *CVPR*, pages 23523–23533, 2024. 2, 5, 6, 7

[65] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: A nested u-net architecture for medical image segmentation. In *DLMIA*, pages 3–11, 2018. 7