Meta-Cognitive Resonance Label Correction for Instance-Dependent Noise

Gaoxia Jiang[1], **Jie Su**[1], **Senyu Hou**[1], **Jia Zhang**[1] and **Wenjian Wang**[2*]

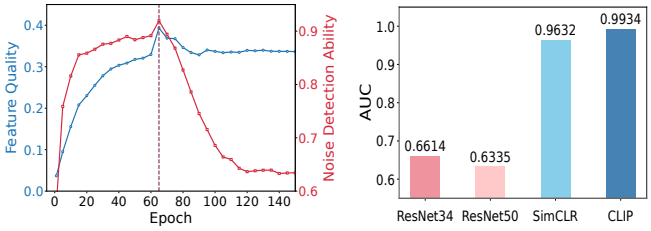
[1]School of Computer and Information Technology, Shanxi University, Taiyuan, China
[2]Key Laboratory of Data Intelligence and Cognitive Computing of Shanxi Province, Shanxi University, Taiyuan, China
{jianggaoxia, sujie1, housenyu, zhangjia8, wjwang}@sxu.edu.cn

Abstract

Training with instance-dependent label noise poses a critical challenge in deep learning, severely impeding model generalization. Existing methods often struggle to break the vicious cycle where corrupted labels degrade feature representations, which in turn impairs noise identification. To overcome this, we propose the Meta-Cognitive Resonance Label Correction (MecReC) framework. For feature representation, MecReC structurally decouples representation learning from label correction, leveraging a pre-trained model as a frozen anchor to ensure feature robustness. For label correction, we introduce a novel meta-cognitive resonance mechanism that aggregates multiple complementary signals to comprehensively assess label reliability. These collaborative signals are fed to a lightweight meta-weight network, which automatically learns whether and how to correct each label. Furthermore, our personalized confidence thresholding strategy guarantees a tighter error bound for label purification. Extensive experiments demonstrate that MecReC consistently outperforms SOTA methods, exhibiting superior robustness particularly under extreme noise conditions. Code and additional details are available at https://github.com/JieSu272/MecReC.

1 Introduction

Deep Neural Networks (DNNs) have demonstrated remarkable capabilities in visual recognition tasks, with success attributed to the availability of large-scale, high-quality annotated datasets [Chen *et al.*, 2019]. However, collecting massive datasets with precise annotations is prohibitively expensive and time-consuming. In real-world scenarios, web crawlers are frequently employed, inevitably introducing noisy labels [Fabian Benitez-Quiroz *et al.*, 2016; Yan *et al.*, 2014]. Directly training on such contaminated data leads to severe performance degradation, as high-capacity DNNs tend to overfit noisy samples by memorizing erroneous correlations rather than learning generalizable patterns [Arpit *et al.*, 2017; Zhang *et al.*, 2021]. This motivates the development of Learning with Noisy Labels (LNL) methods that are tailored to large-scale, low-quality datasets.



(a) The vicious cycle (b) Decoupling analysis

Figure 1: Empirical motivation on CIFAR-100 with 60% IDN. (a) The vicious cycle of feature degradation and noise overfitting. We monitor the training dynamics of a standard ResNet-34. (b) Decoupling analysis. We compare the noise detection AUCs obtained from features learned by end-to-end trained models (ResNet-34 and ResNet-50), a decoupled setting with self-supervised learning (SimCLR), and frozen CLIP representations.

Existing LNL methods focus on Class-Conditional Noise (CCN) and often rely on the *small-loss criterion* for sample selection [Zhao *et al.*, 2022; Li *et al.*, 2020]. However, their performance deteriorates significantly in the presence of Instance-Dependent Noise (IDN), which more faithfully reflects real-world labeling processes. For instance, in the *Clothing1M* dataset, visually similar classes induce ambiguous decision boundaries, resulting in highly asymmetric noise distributions [Chen *et al.*, 2021]. Under such challenging IDN scenarios, the failure to distinguish semantically similar instances becomes a fundamental bottleneck.

To reveal why existing methods fail on IDN, we analyze the training dynamics on CIFAR-100. We evaluate noise identification using the AUC metric based on the *small-loss criterion*. A larger AUC indicates a clearer separability between the loss distributions of noisy and clean samples, and thus a stronger noise detection capability. As shown in Fig. 1(a), we observe a vicious cycle, starting from epoch 65, the deterioration of feature quality (K-Nearest Neighbors accuracy) occurs simultaneously with a sharp decline in noise detection (AUC). This implies that noise memorization destroys the structure of the representation space, blinding the model to noise patterns and invalidating the small-loss assumption.

*Wenjian Wang is the corresponding author.

Most existing methods aggravate this issue by training feature extractors from scratch on noisy data, while jointly updating model weights and pseudo-label distributions within a unified framework [Wu *et al.*, 2021; Zheng *et al.*, 2021]. This tightly coupled training paradigm inevitably yields corrupted representations [Tu *et al.*, 2023]. Under conventional CNNs (e.g., ResNet) trained with noisy supervision, the extracted features lack sufficient robustness and tend to collapse semantically similar IDN samples together [Garg *et al.*, 2024], thereby hindering subsequent noise identification and correction. We conduct a decoupling analysis comparing End-to-End (E2E) baselines with decoupled paradigms as shown in Fig. 1(b). In E2E settings, models suffer from severe feature collapse, where higher capacity (ResNet-50) paradoxically yields lower noise detection AUC (0.63) than ResNet-34 (0.66) due to accelerated noise memorization. In contrast, decoupled approaches like SimCLR [Chen *et al.*, 2020] and frozen CLIP dramatically restore the AUC to 0.96 and 0.99, respectively. These results indicate that isolating feature learning from label corruption is critical for breaking this vicious cycle.

Building upon these insights, we propose the **Me**ta-**c**ognitive **Re**sonance Label **C**orrection (MecReC) framework. By utilizing CLIP as an anchor, MecReC effectively prevents feature collapse under noisy supervision. The main contributions are as follows.

- **Decoupled label correction paradigm.** To break the vicious cycle where noisy labels corrupt feature learning, we propose a decoupled architecture that isolates label correction from representation learning. By leveraging frozen CLIP features as an anchor, the correction process is guided by robust open-world priors rather than degenerated representations.
- **Meta-cognitive resonance signals.** Our MecReC method quantifies label reliability via meta-cognitive resonance signals derived from inter-model consensus, intra-model uncertainty, external validation, and temporal consistency. This holistic state representation enables the meta-weight network to capture fine-grained noise patterns beyond any single criterion.
- **Differentiable meta correction policy.** We design a lightweight meta-weight network to learn a differentiable label rectification policy. Unlike heuristic reweighting methods, the meta-weight network maps resonance signals to adaptive soft labels and is optimized using unbiased gradient feedback from a small clean validation set, explicitly maximizing generalization.
- **A training-coupled thresholding strategy with generalization guarantees.** The confidence threshold for deciding whether to correct a label is adaptively updated throughout training, rather than being manually specified. Under this strategy, we prove that the model trained on the corrected dataset yields a tighter error bound.

2 Related Work

Learning with Noisy Labels and Sample Selection. Learning with noisy labels is typically addressed via robust losses or sample selection. Robust loss functions, such as GCE [Zhang and Sabuncu, 2018] and MAE [Ghosh *et al.*, 2017], mitigate the impact of noisy labels by bounding gradients or balancing convergence speed with robustness. Sample selection methods, starting from Co-teaching [Han *et al.*, 2018], have evolved into hybrid frameworks. SOTA methods, including DivideMix [Li *et al.*, 2020], utilize a Gaussian mixture model to partition data for semi-supervised learning, while confidence-aware approaches like CC [Zhao *et al.*, 2022] and DISC [Li *et al.*, 2023] further refine reliability assessments using prediction probabilities. To enhance feature robustness against noise-induced corruption, recent studies such as SV-Learner [Liang *et al.*, 2024], UniCon [Karim *et al.*, 2022], and PSSCL [Zhang *et al.*, 2025] incorporate contrastive learning into the selection process. However, under high noise rates or IDN, the loss distributions of clean and noisy samples overlap significantly, rendering the small-loss assumption invalid, which inevitably discards potentially useful semantic information within the noisy set. In contrast, MecReC utilizes a soft label correction mechanism that progressively salvages effective supervisory signals from all samples via multi-dimensional consistency checks.

Decoupled Representation Learning. Traditional LNL methods often suffer from the coupling inherent in E2E frameworks, where overfitting to noisy labels corrupts feature representations. Decoupling strategies aim to break this by separating representation learning from noisy supervision. Early efforts include confusion matrix estimation [Tanno *et al.*, 2019], which jointly estimates annotation bias to decouple noise modeling from feature learning. DES [Zhang *et al.*, 2020] adopts a two-stage framework that leverages self-supervised learning for temporal decoupling of feature acquisition. More recently, optimization-level decoupling has emerged: DMLP [Tu *et al.*, 2023] prevents noisy gradient backpropagation through a decoupled meta-objective, while prototype-guided calibration [Yuan *et al.*, 2025a] employs noise-agnostic prototypes to separate representation learning from classification. However, restricted to visual expression, these approaches fail to separate visually confusing patterns under extreme noise, thereby yielding a less discriminative feature space.

Label Correction and Meta Learning. Meta-learning frameworks leverage a small clean validation set to guide the training process. MW-Net [Shu *et al.*, 2019] first introduces a meta-weight network to re-weight samples and suppress noise. Beyond re-weighting, MSLC [Wu *et al.*, 2021] treats labels as learnable parameters and purifies them via validation gradients. Recent methods further refine the correction process: DCD [Mu *et al.*, 2025] employs dynamic center distances to distinguish hard samples from noise, while DMLP [Tu *et al.*, 2023] and EMLC [Taraday and Baskin, 2023] improve efficiency and reduce confirmation bias. Yet, existing meta-correctors often rely on single-dimension feedback or static thresholds, lacking dynamic control over the correction trajectory. In contrast, our MecReC bridges these gaps by synthesizing multi-dimensional signals, i.e., consistency, entropy, and history, within a bi-level optimization framework, achieving a more stable and adaptive correction policy than prior single-view methods.

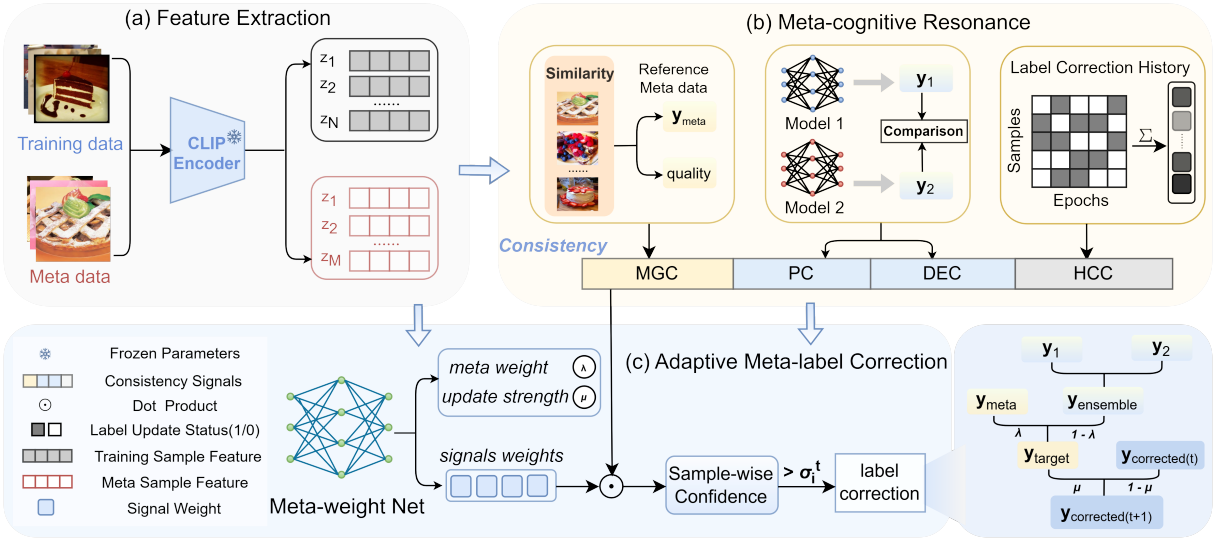


Figure 2: Our MecReC framework. The framework consists of three stages: (a) decoupled feature extraction using a frozen CLIP encoder, (b) meta-cognitive resonance signals generation, and (c) adaptive meta-label correction guided by a meta-weight network.

Algorithm 1 Meta-cognitive Resonance Label Correction

Input: Noisy training set D_{train}, clean validation set D_{val}, pre-trained feature extractor $\text{Clip}(\cdot)$, main models $f_{\theta_1}, f_{\theta_2}$, meta-weight network g_ϕ.
Output: Optimized model parameters $\theta_1^*, \theta_2^*, \phi^*$, corrected label set $\{\hat{y}_i\}_{i=1}^N$.

For all x in D_{train} and D_{val}, compute $\mathrm{z}_i = \text{Clip}(x_i)$.
for epoch $t = 1$ **to** T **do**
  for each mini-batch $B_{\text{train}} \subseteq D_{\text{train}}$ **do**
    if $t > t_{\text{warm-up}}$ **then**
      Get feature $\mathbf{z}_{\text{train}}$ and current labels $\hat{y}_{\text{train}}$.
      Calculate $\hat{y}_i^{\text{meta}}$ for each sample by Eq. (3).
      Update ϕ by solving the objective in Eq. (7).
      Compute S_i by Eqs. (1), (2), (4), (5), (6).
      Generate $\lambda_i = g(x_i, \phi)$ based on S_i.
      Update $\theta_k \leftarrow \theta_k - \alpha_{\text{main}} \nabla_{\theta_k} \mathcal{L}_{\text{main}}$, $k \in \{1, 2\}$.
      Compute confidence score Conf_i by Eq. (12).
      If $\text{Conf}_i > \sigma_i^t$, update label $\hat{y}_i$ by Eq. (14).
    end if
  end for
end for

3 Methodology

The overall framework of MecReC is shown in Fig. 2 and the pseudo-code is presented in Algorithm 1. MecReC recovers high-quality supervisory signals from noisy data through a co-learning paradigm consisting of two peer networks f_{θ_1} and f_{θ_2}, guided by a meta-weight network g_ϕ.

Let $D_{\text{train}} = \{(x_i, y_i)\}_{i=1}^N$ be the noisy training set and $D_{\text{val}} = \{(x_i^{\text{val}}, y_i^{\text{val}})\}_{i=1}^M$ be a small clean validation set.

3.1 CLIP Decoupled Feature Extraction

CLIP [Radford *et al.*, 2021] demonstrates outstanding zero-shot generalization capability through training on massive image-text pairs. We denote the pre-trained and frozen CLIP image encoder as $\text{Clip}(\cdot)$. For any input image x_i, its feature is extracted as $\mathrm{z}_i = \text{Clip}(x_i)$. These frozen features serve as inputs for both the dual main models and the meta-weight network, ensuring that the correction strategy is derived from stable and noise-invariant representations.

3.2 Meta-Cognitive Resonance Signals

MecReC performs label reliability assessment from multiple views through a mechanism termed meta-cognitive resonance. For each sample, we construct a state vector S_i by aggregating four dynamic signals, which capture label correction confidence from complementary sources.

Prediction Consistency (PC). The semantic consensus between two peer networks is measured by the cosine similarity of their predicted probability distributions:

$$\text{PC}_i = \frac{1}{2}\left(1 + \frac{p_i^{V_1} \cdot p_i^{V_2}}{\|p_i^{V_1}\| \cdot \|p_i^{V_2}\|}\right), \tag{1}$$

where $p_i^{V_1} = \text{Softmax}(f_{\theta_1}(\mathrm{z}_i))$, $p_i^{V_2} = \text{Softmax}(f_{\theta_2}(\mathrm{z}_i))$. Higher consistency indicates more reliable predictions.

Dynamic Entropy Consistency (DEC). This signal captures the synchronization of prediction uncertainty between the two peer networks:

$$\text{DEC}_i = 1 - \frac{|\text{Ent}_{V_1,i} - \text{Ent}_{V_2,i}|}{\log C}, \tag{2}$$

where $\text{Ent}_{V_k,i} = -\sum_{c=1}^{C} \mathbf{p}_{i,c}^{V_k} \log\left(\mathbf{p}_{i,c}^{V_k} + \epsilon\right)$. Here, ϵ is a small constant for numerical stability, C is the number of classes, and $\log C$ represents the maximum entropy value. Closer alignment suggests higher reliability.

Meta-Guided Consistency (MGC). This signal injects external reliable knowledge using a small clean validation set. For each training sample feature z_i, we retrieve its KNN in

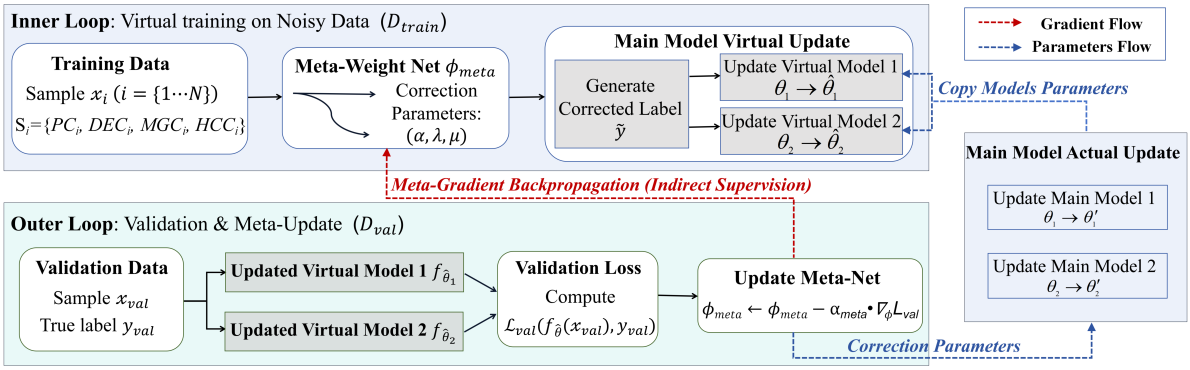


Figure 3: Bi-level optimization structure diagram. The Inner Loop performs a virtual training step on the main model using meta-generated weights. The Outer Loop evaluates this virtual update on the validation set and backpropagates the gradient through the virtual step to update the Meta-weight Network.

D_{val}, and construct a meta-guided pseudo-label by similarity-weighted aggregation:

$$y_i^{\text{meta}} = \sum_{z_j \in \text{KNN}(D_{\text{val}})} w_{i,j} \cdot y_j{}^{\text{val}}, \tag{3}$$

where $w_{i,j}$ denotes the normalized feature similarity between z_i and its j-th neighbor, satisfying $w_{i,j} \geq 0$ and $\sum_{z_j \in \text{KNN}(D_{\text{val}})} w_{i,j} = 1$.

The quality of this guidance, denoted q_i, is measured by the normalized average similarity over the first-K neighbors. Then, we compute the consistency between the average prediction $p_i = [\text{Softmax}(f_{\theta_1}(\text{z}_i)) + \text{Softmax}(f_{\theta_2}(\text{z}_i))]/2$ and the meta-guided label y_i^{meta} dynamically modulated by q_i:

$$\text{MGC}_i = q_i \cdot \frac{1}{2}\left(1 + \frac{p_i \cdot y_i{}^{\text{meta}}}{\|p_i\| \cdot \|y_i{}^{\text{meta}}\|}\right). \tag{4}$$

Historical Correction Consistency (HCC). To mitigate potential over-correction and prevent label oscillation, we introduce a historical consistency signal that tracks correction stability over time. This signal is defined based on the cumulative correction count h_i up to the current epoch t.

$$\text{HCC}_i = 1 - \frac{h_i}{t}. \tag{5}$$

A lower correction frequency implies high label stability.

Finally, the input state vector S_i is:

$$\text{S}_i = [\text{PC}_i, \text{DEC}_i, \text{MGC}_i, \text{HCC}_i]. \tag{6}$$

3.3 Meta-Guided Optimization Procedure

Bi-level Meta Loop. Formally, we cast the noise-robust training process as a bi-level optimization problem. Our objective is to learn the meta-weight network parameters ϕ, to achieve minimal loss on a clean validation set. Specifically, the outer objective minimizes the validation loss $\mathcal{L}_{\text{val}}$, while the inner objective defines a temporary, differentiable update of the main model parameters θ_k to $\hat{\theta}_k$ ($k \in \{1, 2\}$):

$$\begin{aligned} \min_{\phi} \quad & \mathcal{L}_{\text{val}}\left(f_{\hat{\theta}_k}(\text{z}_{\text{val}}), y_{\text{val}}\right) \\ \text{s. t.} \quad & \hat{\theta}_k = \theta_k - \alpha_{\text{main}} \nabla_{\theta_k} \mathcal{L}_{\text{virt}}(\phi, \theta_k; D_{\text{train}}). \end{aligned} \tag{7}$$

Computing the gradient of the outer objective with respect to ϕ, i.e., $\nabla_{\phi} \mathcal{L}_{\text{val}}$, is challenging as it requires differentiating through the inner optimization trajectory. To achieve this efficiently, we employ automatic-differentiation-based optimization techniques (implemented via the *higher* library). Fig. 3 shows the micro-circulation flow in detail.

1. Meta-weight generation. For each sample $x_i \in B_{\text{train}}$, the meta-weight network generates an instance-specific loss weight $\lambda_i = g(x_i, \phi)$ based on its input state vector S_i.

2. Virtual loss computation. We define a meta-weight virtual training loss for the current batch. To guide the model towards stability, this loss targets the current smoothed label estimate $\hat{y}_i$ (Eq. (14)).

$$\begin{aligned} \mathcal{L}_{\text{virt}}(\phi, \theta_k) = \frac{1}{|B_{\text{train}}|} \sum_{x_i \in B_{\text{train}}} [(1 - \lambda_i) \cdot \text{KL}(\text{Softmax}(f_{\theta_k}(\text{z}_i)) \\ \| \hat{y}_i) + \lambda_i \cdot \text{KL}\left(\text{Softmax}(f_{\theta_k}(\text{z}_i)) \| y_i^{\text{meta}}\right)]. \end{aligned} \tag{8}$$

3. Differentiable virtual update. Following the bi-level optimization objective in Eq. (7), a differentiable virtual update is performed on the main model parameters, yielding the virtually updated parameters, $\hat{\theta}_k = \theta_k - \alpha_{\text{main}} \nabla_{\theta_k} \mathcal{L}_{\text{virt}}$. Crucially, the computational graph of this step is retained to enable higher-order meta-gradient computation.

4. Meta-gradient calculation. The virtually updated model $f_{\hat{\theta}_k}$ is evaluated over the full clean validation D_{val}. The validation loss $\mathcal{L}_{\text{val}}(\phi)$ is computed through the virtual update path via the chain rule to capture how changes in ϕ affect validation performance by influencing $\hat{\theta}$:

$$\mathcal{L}_{\text{val}}(\phi) = \frac{1}{2} \sum_{k=1}^{2} \frac{1}{|D_{\text{val}}|} \sum_{i \in D_{\text{val}}} \text{CE}\left(f_{\hat{\theta}_k}(\text{z}_i^{\text{val}}), y_i^{\text{val}}\right). \tag{9}$$

5. Meta-weight network update. Finally, the meta-weight network parameters are updated using the computed meta-gradient: $\phi \leftarrow \phi - \alpha_{\text{meta}} \nabla_{\phi} \mathcal{L}_{\text{val}}$. This update mechanism enables the meta-weight network to learn a label-correction strategy that explicitly maximizes the generalization ability of the main models.

Main Model Actual Training. After each meta-update (updating ϕ), the main models (f_{θ_1} and f_{θ_2}) perform a real

gradient descent step on the current training batch. This process utilizes the high-quality meta weights (λ) generated by the updated meta-weight network to dynamically reweight the loss function, thereby enabling effective noise-robust training. Specifically, for each main model f_{θ_k} (with $k \in \{1, 2\}$), the real training loss $\mathcal{L}_i^{V_k}$ for the i-th sample is defined as a weighted combination of the original label loss and the meta-guided loss:

$$\begin{aligned}\mathcal{L}_i^{V_k} = (1-\lambda_i) \cdot \mathrm{KL}\left(\mathrm{Softmax}(f_{\theta_k}(\mathrm{z}_i)) \parallel \hat{y}_i\right) \\ +\lambda_i \cdot \mathrm{KL}\left(\mathrm{Softmax}(f_{\theta_k}(\mathrm{z}_i)) \parallel y_i^{\text{meta}}\right).\end{aligned} \tag{10}$$

The meta weight $\lambda_i \in [0, 1]$ acts as an adaptive balancing factor: a larger λ_i encourages the model to learn more from the meta-guided label, while a smaller λ_i prioritizes the current smoothed label.

The total training loss for the main models is defined as the average loss of the two peer networks over the current B_{train}:

$$\mathcal{L}_{\text{main}} = \frac{1}{2|B_{\text{train}}|} \sum_{x_i \in B_{\text{train}}} \left(\mathcal{L}_i^{V_1} + \mathcal{L}_i^{V_2}\right). \tag{11}$$

The parameters of the main models f_{θ_1} and f_{θ_2} are updated by θ_k: $\theta_k^{'} \leftarrow \theta_k - \alpha_{\text{main}} \nabla_{\theta_k} \mathcal{L}_{\text{main}}$, with λ_i treated as a constant during this phase.

3.4 Adaptive Meta-label Correction

To refine training targets and combat noise, we employ a soft label evolution strategy based on Meta-cognitive Resonance. In this phase, the meta-weight network $g(\cdot; \phi)$ is designed to output not only the loss reweighting factor λ_i, but also two critical parameters: a correction attention weight vector $\mathrm{a}_i \in \mathbb{R}^4$ (softmax-normalized) targeting the four resonance signals, and an instance-specific update strength $\mu_i \in [0, 1]$.

We first compute a composite confidence score by aggregating the signal vector S_i via weighted summation:

$$\mathrm{Conf}_i = \mathrm{a}_i^\top \mathrm{S}_i. \tag{12}$$

It quantifies the necessity for correction of the current label.

To ensure high-precision correction, we introduce a personalized threshold σ_i^t by Eq. (16), which is adaptively adjusted throughout training. For samples whose confidence score exceeds this threshold ($\mathrm{Conf}_i > \sigma_i^t$), we generate a refined soft label target $\hat{y}_i^{\text{target}}$. This target is a weighted combination of the model ensemble prediction p_i and the meta-guided label y_i^{meta}:

$$\hat{y}_i^{\text{target}} = \lambda_i \cdot y_i^{\text{meta}} + (1-\lambda_i) \cdot p_i. \tag{13}$$

Finally, the current smoothed label estimate is updated via an Exponential Moving Average mechanism, where μ_i is produced by the meta-weight network:

$$\hat{y}_i^{t+1} = \mu_i \cdot \hat{y}_i^{\text{target}} + (1-\mu_i) \cdot \hat{y}_i^t. \tag{14}$$

This dynamic and instance-specific update mechanism ensures that the correction process can rapidly adapt to severe label noise while maintaining training stability and robustness.

The confidence threshold ($\mathrm{Conf}_i > \sigma_i^t$) is determined based on the following theoretical findings.

Theorem 1 (Generalization guarantee)**.** *Let $D = \{(x_i, y_i)\}_{i=1}^N$ be a classification dataset with noisy labels, and let D^+ denote the corrected dataset obtained by a confidence-based label correction procedure. Let $\mathcal{B}(h; D)$ denote the Rademacher-based generalization bound of a hypothesis $h \in \mathcal{H}$ trained on dataset D. Suppose that for each corrected sample, the updated confidence satisfies*

$$\mathrm{Conf}_i^{t+1} > \mathrm{Conf}_i^t + \frac{\sigma_p - \Delta\mathrm{Acc} \cdot \mathrm{Conf}_i^{t+1}}{\mathrm{Acc}^t}, \tag{15}$$

where Conf_i^t and Acc^t denote the label confidence and model accuracy at iteration t, respectively, σ_p is a noise tolerance parameter ($\sigma_p > 0$), and the accuracy improvement in one correction step $\Delta\mathrm{Acc} = \mathrm{Acc}^{t+1} - \mathrm{Acc}^t$. Then, under the above condition, the Rademacher generalization bound is strictly reduced, i.e., $\mathcal{B}(h; D^+) < \mathcal{B}(h; D)$.

Theorem 1 shows that correcting labels for high-confidence samples, characterized by consistent signals, leads to improved generalization. Accordingly, we set the confidence threshold as

$$\sigma_i^t = \mathrm{Conf}_i^t + \frac{\sigma_p}{\mathrm{Acc}^t}, \tag{16}$$

such that $\mathrm{Conf}_i^{t+1} > \sigma_i^t$ satisfies the sufficient condition in Eq. (15) and guarantees a tighter error bound. The term $\Delta\mathrm{Acc} \cdot \mathrm{Conf}_i^{t+1}$ is omitted since accuracies across adjacent iterations are close, and this simplification introduces a conservative margin to prevent over-correction. Since the true model accuracy is unavailable during training, we approximate its evolution using a surrogate function $\mathrm{Acc}_t = 1 - \frac{c}{\tilde{t}+2c}$, where c is a positive constant, and $\tilde{t}$ is the normalized epoch. In our experiments, we set the hyperparameters as σ_p=0.15, c=0.20.

4 Experiments

4.1 Results on Synthetic Noisy Datasets

We evaluate MecReC on the standard synthetic noisy datasets CIFAR-10 and CIFAR-100. To simulate realistic scenarios, we generate two types of instance-dependent noise: Part-dependent label noise (PDL) [Xia *et al.*, 2020] and Classification-based label noise (CBL) [Chen *et al.*, 2021], with noise rates set to 20%, 40%, and 60%. Unlike traditional end-to-end approaches, we employ a pre-trained CLIP ViT-L/14 as a frozen feature extractor to leverage robust semantic representations. Following standard meta-learning-based protocols [Wu *et al.*, 2021; Shu *et al.*, 2019], we reserve 1000 clean images as the validation set, using the remaining samples for training.

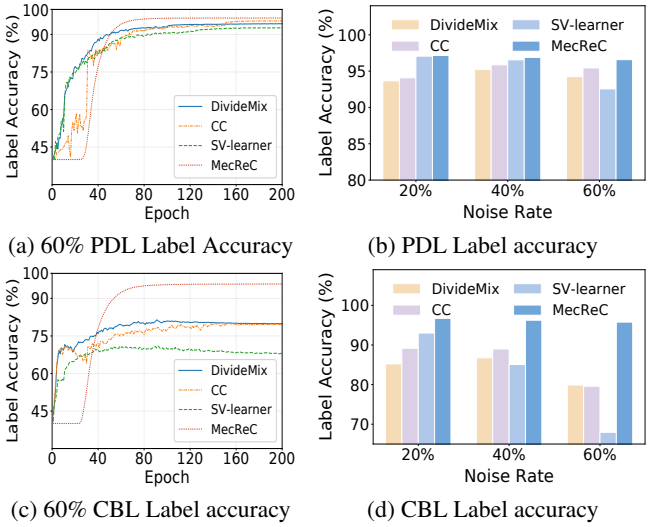
Tables 1 and 2 present the results comparisons on CIFAR-10 and CIFAR-100 under PDL and CBL noise, respectively. MecReC consistently outperforms all competing baselines across varying noise rates, with the performance gap becoming particularly pronounced in high-noise regimes. Specifically, in the challenging setting of CIFAR-100 with 60% PDL noise, the MecReC achieved 79.46% accuracy, more than 2.12% higher than the next best method. Generally, methods leveraging ViT backbones surpass conventional approaches.

Method	CIFAR-10 w/ IDN			CIFAR-100 w/ IDN		
	20%	40%	60%	20%	40%	60%
CE (Standard)	83.93 ± 0.15	67.64 ± 0.26	43.83 ± 0.33	57.35 ± 0.08	43.17 ± 0.15	24.42 ± 0.16
GCE (NeurIPS) [Zhang and Sabuncu, 2018]	89.80 ± 0.12	78.95 ± 0.15	60.76 ± 3.08	58.01 ± 0.26	45.69 ± 0.14	35.08 ± 0.23
Co-teaching (NeurIPS) [Han *et al.*, 2018]	88.87 ± 0.24	73.00 ± 1.24	62.51 ± 1.98	43.30 ± 0.39	23.21 ± 0.57	12.58 ± 0.58
JoCoR (CVPR) [Wei *et al.*, 2020]	88.78 ± 0.15	71.64 ± 3.09	63.46 ± 1.58	43.66 ± 1.32	23.95 ± 0.44	13.16 ± 0.91
DivideMix (ICLR) [Li *et al.*, 2020]	93.33 ± 0.14	95.07 ± 0.11	85.50 ± 0.71	79.04 ± 0.25	76.08 ± 0.35	46.72 ± 1.32
CC (ECCV) [Zhao *et al.*, 2022]	93.68 ± 0.12	94.97 ± 0.09	94.95 ± 0.11	79.61 ± 0.19	76.58 ± 0.25	59.40 ± 0.46
DISC (CVPR) [Li *et al.*, 2023]	96.48 ± 0.04	95.94 ± 0.04	95.05 ± 0.05	80.12 ± 0.13	78.44 ± 0.19	69.57 ± 0.14
RML (AAAI) [Li *et al.*, 2024]	96.31 ± 0.21	95.52 ± 0.32	94.75 ± 0.27	78.41 ± 0.22	76.83 ± 0.49	72.38 ± 0.44
JAL (ICCV)* [Wang *et al.*, 2025]	89.90 ± 0.14	86.78 ± 0.17	75.02 ± 0.48	67.77 ± 0.38	63.56 ± 0.18	51.69 ± 0.59
Early-Cutting (NeurIPS)* [Yuan *et al.*, 2025b]	93.40 ± 0.22	90.78 ± 0.31	88.51 ± 0.65	75.03 ± 0.23	69.94 ± 0.30	61.06 ± 0.39
PLReMix (WACV)* [Liu *et al.*, 2025]	95.91 ± 0.17	95.36 ± 0.45	94.08 ± 0.44	77.95 ± 0.17	77.67 ± 0.47	68.41 ± 0.35
DULC (AAAI)* [Xu *et al.*, 2025]	96.57 ± 0.23	94.91 ± 0.38	94.07 ± 0.25	76.42 ± 0.36	78.57 ± 0.29	68.11 ± 0.50
SV-learner (TKDE)* [Liang *et al.*, 2024]	96.68 ±0.03	96.23 ± 0.16	93.19 ± 0.22	78.97 ± 0.08	77.47 ± 0.19	68.64 ± 0.34
SIMIFEAT+ViT (ICML)* [Zhu *et al.*, 2022]	94.61 ± 0.06	93.24 ± 0.06	92.58 ± 0.11	77.96 ± 0.12	77.73 ± 0.16	76.66 ± 0.21
EPL+ViT (ICLR)* [Ko *et al.*, 2023]	95.47 ± 0.04	95.90 ± 0.07	95.11 ± 0.17	78.03 ± 0.31	77.80 ± 0.28	77.34 ± 0.24
CLIPCleaner (MM)* [Feng *et al.*, 2024]	95.51 ± 0.16	95.48 ± 0.27	95.02 ± 0.41	76.37 ± 0.42	74.98 ± 0.29	72.90 ± 0.61
LRA+ViT (NeurIPS)* [Chen *et al.*, 2023]	96.48 ± 0.14	96.44 ± 0.16	95.56 ± 0.39	79.69 ± 0.26	78.87 ± 0.19	74.53 ± 0.34
DLD+ViT (CVPR)* [Hou *et al.*, 2025]	96.98 ± 0.12	96.49 ± 0.07	83.14 ± 0.81	78.62 ± 0.11	76.03 ± 0.14	69.57 ± 0.17
MecReC (Ours)	**97.16 ± 0.08**	**97.13 ± 0.13**	**97.02 ± 0.27**	**83.05 ± 0.31**	**81.13 ± 0.57**	**79.46 ± 0.78**

Table 1: Test accuracy (%) comparison of CIFAR datasets under different levels of IDN (PDL). Results marked with * are reproduced by us. The best results are in **bold**. The second- and third-best results are underlined.

Method	CIFAR-10 w/ IDN			CIFAR-100 w/ IDN		
	20%	40%	60%	20%	40%	60%
DivideMix*	84.8	88.02	79.41	71.5	65.75	57.95
SV-learner*	87.14	86.88	71.41	72.7	66.08	55.64
CC*	88.54	91.55	81.48	71.62	65.86	56.61
DISC*	95.98	93.63	76.88	78.57	74.44	62.58
MecReC (Ours)	**97.04**	**97.03**	**96.73**	**84.05**	**79.45**	**74.84**

Table 2: Test accuracy (%) comparison of CIFAR datasets with IDN (CBL). Results marked with * are reproduced by us. The best results are in **bold**. The second-best results are underlined.



(a) 60% PDL Label Accuracy (b) PDL Label accuracy

(c) 60% CBL Label accuracy (d) CBL Label accuracy

Figure 4: Label accuracy (%) on CIFAR-10 with different noise types and ratios.

MecReC, in particular, stands out by achieving the best results within this category. A key distinction is that while competing ViT-based methods suffer from performance deterioration under rising noise levels, MecReC demonstrates remarkable robustness.

We visualize the label correction quality in Fig. 4. MecReC demonstrates the highest label accuracy and superior stability across all noise ratios. Notably, it shows a much smoother convergence trajectory under EMA guidance. This mechanism effectively stabilizes training and mitigates the memorization of corrupted labels observed in baselines.

4.2 Results on Real-world Noisy Datasets

We evaluate MecReC on three widely used real-world noisy benchmarks, covering diverse noise patterns and dataset scales. As shown in Table 3 and 4, our method achieves SOTA performance across all evaluated scenarios. On the large-scale Clothing1M and WebVision, MecReC attains accuracies of 76.04% and 85.24%, respectively, achieving the best overall performance, with particularly notable gains on WebVision. Notably, on WebVision, our model demonstrates strong generalization, achieving 83.72% accuracy on the clean ILSVRC12 validation set. Furthermore, on Animal-10N, which is characterized by challenging fine-grained label noise, our approach effectively distinguishes subtle visual differences, improving the accuracy to 90.82%. These consistent gains confirm the robustness and generalization capability of our framework in handling complex, real-world label noise.

Method	Animal-10N	Clothing1M
CE	—	68.94
GCE	81.2	—
Co-teaching	84.9	69.21
DivideMix	—	74.45
CC	—	74.40
DISC	87.3	74.79
LRA	88.6	75.70
CLIPCleaner	88.9	—
PCSE [Luo *et al.*, 2024]	85.5	71.37
DLD	89.4	75.69
CSC [Fan and Li, 2025]	89.7	—
MecReC (Ours)	**90.82**	**76.04**

Table 3: Classification accuracies (%) on Animal-10N and Clothing1M.

Method	WebVision top1	WebVision top5	ILSVRC12 top1	ILSVRC12 top5
Co-teaching	63.58	85.20	61.48	84.70
DivideMix	77.32	91.64	75.20	90.84
CC	79.36	93.64	76.08	93.86
DISC	80.28	92.28	77.44	92.28
CLIPCleaner	81.56	93.26	77.80	92.08
PLReMix	81.49	93.79	77.75	93.08
DULC	79.90	93.70	76.90	93.90
CSC	82.30	94.60	78.20	94.90
MecReC (Ours)	**85.24**	**98.56**	**83.72**	**98.84**

Table 4: Classification accuracies (%) on WebVision and ILSVRC2012 datasets.

4.3 Ablation Studies

To comprehensively validate the effectiveness and necessity of the core components within the MecReC framework, we conducted ablation studies on the CIFAR-10 dataset under a 40% PDL instance-dependent label noise setting.

Component Analysis. We compare the full MecReC model against variants where specific modules are either removed or replaced. As presented in Table 5, the baseline without label correction suffers from a low label accuracy of 60%. In contrast, the full MecReC framework achieves a remarkable label accuracy of 96.65% (+36.65%) while achieving the highest test accuracy of 97.13%. Ablating meta-weight network or resonance signals reduces the label accuracy to 80.30% and 79.40%, which strongly validates the necessity of each component in our framework for effectively correcting noisy labels.

Meta-Data Size Analysis. We analyze the impact of meta-data size on model performance in Table 6. Although a positive correlation exists between meta-set size and accuracy, the improvement is subtle. The model yields a high accuracy of 97.09% at the 1% ratio, which is comparable to the 97.16% accuracy obtained with 6% data. This suggests that MecReC can effectively extract guidance from very limited clean data.

Computational Efficiency. Fig. 5 illustrates the training efficiency on CIFAR-100 (40% CBL noise). MecReC achieves the lowest computational cost, requiring only 31.4 seconds per epoch. The reported per-epoch runtime excludes the one-time CLIP feature extraction cost, which is performed only once before training. This efficiency stems from our decoupled training strategy where features are pre-extracted and cached. By bypassing expensive backbone back-propagation, the runtime is limited to label correction and classifier optimization.

Method	Test Accuracy	Label Accuracy
w/o resonance signals (CE-loss)	95.16	79.41
w/o meta-weight network	95.98	80.34
w/o label correction	96.12	60.00
MecReC (Ours)	**97.13**	**96.65**

Table 5: Component-wise performance comparison (%).

Meta-Data Size	1%	2%	4%	6%
Accuracy	97.09	97.13	97.14	97.16
Δ	–	**+0.04**	+0.01	+0.02

Table 6: Performance comparison (%) with different meta-data size.

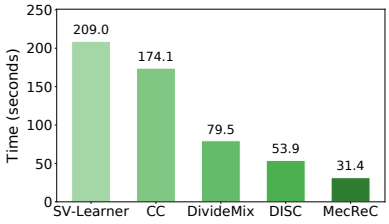


Figure 5: Comparison of training time per epoch (in seconds). All runtimes are measured on a single NVIDIA RTX 4090 GPU.

5 Conclusion

In this paper, we propose MecReC to address the challenging problem of learning with IDN label noise. We fundamentally rethink the interaction between features and labels by leveraging frozen pre-trained features to break their vicious coupling, providing a stable foundation for noise correction. We further introduced the meta-cognitive resonance, which integrates multi-perspective complementary signals to comprehensively quantify label reliability. By these components, our meta-weight network is capable of learning dynamic correction strategies in a data-driven manner. Our principled confidence thresholding strategy guarantees a tighter error bound for label purification. Extensive empirical evaluations validate the robustness and versatility of MecReC across a wide spectrum of noise types and levels.

Limitations

MecReC relies on a small clean validation set for meta-guided optimization, which may be costly in some real-world scenarios. Moreover, the surrogate accuracy function used in the adaptive thresholding strategy is an approximate estimation and may introduce deviations under different domains or data distributions.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (U21A20513, 62576201, 62476157, 62276161).

References

[Arpit *et al.*, 2017] 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 *International Conference on Machine Learning*, pages 233–242. PMLR, 2017.

[Chen *et al.*, 2019] Pengfei Chen, Ben Ben Liao, Guangyong Chen, and Shengyu Zhang. Understanding and utilizing deep neural networks trained with noisy labels. In *International Conference on Machine Learning*, pages 1062–1070. PMLR, 2019.

[Chen *et al.*, 2020] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *International Conference on Machine Learning*, pages 1597–1607. PMLR, 2020.

[Chen *et al.*, 2021] Pengfei Chen, Junjie Ye, Guangyong Chen, Jingwei Zhao, and Pheng-Ann Heng. Beyond class-conditional assumption: A primary attempt to combat instance-dependent label noise. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 11442–11450, 2021.

[Chen *et al.*, 2023] Jian Chen, Ruiyi Zhang, Tong Yu, Rohan Sharma, Zhiqiang Xu, Tong Sun, and Changyou Chen. Label-retrieval-augmented diffusion models for learning from noisy labels. *Advances in Neural Information Processing Systems*, 36:66499–66517, 2023.

[Fabian Benitez-Quiroz *et al.*, 2016] C Fabian Benitez-Quiroz, Ramprakash Srinivasan, and Aleix M Martinez. Emotionet: An accurate, real-time algorithm for the automatic annotation of a million facial expressions in the wild. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5562–5570, 2016.

[Fan and Li, 2025] Wenxiao Fan and Kan Li. Combating semantic contamination in learning with label noise. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 2870–2878, 2025.

[Feng *et al.*, 2024] Chen Feng, Georgios Tzimiropoulos, and Ioannis Patras. Clipcleaner: Cleaning noisy labels with clip. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pages 876–885, 2024.

[Garg *et al.*, 2024] Arpit Garg, Cuong Nguyen, Rafael Felix, Thanh-Toan Do, and Gustavo Carneiro. Instance-dependent noisy-label learning with graphical model based noise-rate estimation. In *European Conference on Computer Vision*, pages 372–389. Springer, 2024.

[Ghosh *et al.*, 2017] Aritra Ghosh, Himanshu Kumar, and P Shanti Sastry. Robust loss functions under label noise for deep neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31, 2017.

[Han *et al.*, 2018] 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. *Advances in Neural Information Processing Systems*, 31, 2018.

[Hou *et al.*, 2025] Senyu Hou, Gaoxia Jiang, Jia Zhang, Shangrong Yang, Husheng Guo, Yaqing Guo, and Wenjian Wang. Directional label diffusion model for learning from noisy labels. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 25738–25748, 2025.

[Karim *et al.*, 2022] 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*, pages 9676–9686, 2022.

[Ko *et al.*, 2023] Jongwoo Ko, Sumyeong Ahn, and Se-Young Yun. Efficient utilization of pre-trained model for learning with noisy labels. In *ICLR 2023 Workshop on Pitfalls of limited data and computation for Trustworthy ML*, 2023.

[Li *et al.*, 2020] Junnan Li, Richard Socher, and Steven C. H. Hoi. Dividemix: Learning with noisy labels as semi-supervised learning. In *International Conference on Learning Representations*, 2020.

[Li *et al.*, 2023] Yifan Li, Hu Han, Shiguang Shan, and Xilin Chen. Disc: Learning from noisy labels via dynamic instance-specific selection and correction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 24070–24079, 2023.

[Li *et al.*, 2024] Fengpeng Li, Kemou Li, Jinyu Tian, and Jiantao Zhou. Regroup median loss for combating label noise. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 13474–13482, 2024.

[Liang *et al.*, 2024] Xin Liang, Yanli Ji, Wei-Shi Zheng, Wangmeng Zuo, and Xiaofeng Zhu. Sv-learner: Support-vector contrastive learning for robust learning with noisy labels. *IEEE Transactions on Knowledge and Data Engineering*, 36(10):5409–5422, 2024.

[Liu *et al.*, 2025] Xiaoyu Liu, Beitong Zhou, Zuogong Yue, and Cheng Cheng. Plremix: Combating noisy labels with pseudo-label relaxed contrastive representation learning. In *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 6517–6527. IEEE, 2025.

[Luo *et al.*, 2024] Wenshui Luo, Shuo Chen, Tongliang Liu, Bo Han, Gang Niu, Masashi Sugiyama, Dacheng Tao, and Chen Gong. Estimating per-class statistics for label noise learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.

[Mu *et al.*, 2025] Chenyu Mu, Yijun Qu, Jiexi Yan, Erkun Yang, and Cheng Deng. Meta-learning dynamic center distance: Hard sample mining for learning with noisy labels.

In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 415–425, 2025.

[Radford *et al.*, 2021] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International Conference on Machine Learning*, pages 8748–8763. PMLR, 2021.

[Shu *et al.*, 2019] 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. *Advances in Neural Information Processing Systems*, 32, 2019.

[Tanno *et al.*, 2019] Ryutaro Tanno, Ardavan Saeedi, Swami Sankaranarayanan, Daniel C Alexander, and Nathan Silberman. Learning from noisy labels by regularized estimation of annotator confusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11244–11253, 2019.

[Taraday and Baskin, 2023] Mitchell Keren Taraday and Chaim Baskin. Enhanced meta label correction for coping with label corruption. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 16295–16304, 2023.

[Tu *et al.*, 2023] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 19934–19943, 2023.

[Wang *et al.*, 2025] Jialiang Wang, Xianming Liu, Xiong Zhou, Gangfeng Hu, Deming Zhai, Junjun Jiang, and Xiangyang Ji. Joint asymmetric loss for learning with noisy labels. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1947–1956, 2025.

[Wei *et al.*, 2020] 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*, pages 13726–13735, 2020.

[Wu *et al.*, 2021] Yichen Wu, Jun Shu, Qi Xie, Qian Zhao, and Deyu Meng. Learning to purify noisy labels via meta soft label corrector. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 10388–10396, 2021.

[Xia *et al.*, 2020] Xiaobo Xia, Tongliang Liu, Bo Han, Nannan Wang, Mingming Gong, Haifeng Liu, Gang Niu, Dacheng Tao, and Masashi Sugiyama. Part-dependent label noise: Towards instance-dependent label noise. *Advances in Neural Information Processing Systems*, 33:7597–7610, 2020.

[Xu *et al.*, 2025] Yuanzhuo Xu, Xiaoguang Niu, Jie Yang, Ruiyi Su, Jian Zhang, Shubo Liu, and Steve Drew. Revisiting interpolation for noisy label correction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 21833–21841, 2025.

[Yan *et al.*, 2014] Yan Yan, Rómer Rosales, Glenn Fung, Ramanathan Subramanian, and Jennifer Dy. Learning from multiple annotators with varying expertise. *Machine Learning*, 95(3):291–327, 2014.

[Yuan *et al.*, 2025a] Huiting Yuan, Tingjin Luo, Xinghao Wu, and Jie Jiang. Noise-free prototype guided representation calibration under label noise. *Knowledge-Based Systems*, page 114308, 2025.

[Yuan *et al.*, 2025b] Suqin Yuan, Lei Feng, Bo Han, and Tongliang Liu. Enhancing sample selection against label noise by cutting mislabeled easy examples. *arXiv preprint arXiv:2502.08227*, 2025.

[Zhang and Sabuncu, 2018] Zhilu Zhang and Mert Sabuncu. Generalized cross entropy loss for training deep neural networks with noisy labels. *Advances in Neural Information Processing Systems*, 31, 2018.

[Zhang *et al.*, 2020] Zizhao Zhang, Han Zhang, Sercan O Arik, Honglak Lee, and Tomas Pfister. Distilling effective supervision from severe label noise. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9294–9303, 2020.

[Zhang *et al.*, 2021] Chiyuan Zhang, Samy Bengio, Moritz Hardt, Benjamin Recht, and Oriol Vinyals. Understanding deep learning (still) requires rethinking generalization. *Communications of the ACM*, 64(3):107–115, 2021.

[Zhang *et al.*, 2025] Qian Zhang, Yi Zhu, Filipe R Cordeiro, and Qiu Chen. Psscl: A progressive sample selection framework with contrastive loss designed for noisy labels. *Pattern Recognition*, 161:111284, 2025.

[Zhao *et al.*, 2022] 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 *European Conference on Computer Vision*, pages 21–37. Springer, 2022.

[Zheng *et al.*, 2021] Guoqing Zheng, Ahmed Hassan Awadallah, and Susan Dumais. Meta label correction for noisy label learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 11053–11061, 2021.

[Zhu *et al.*, 2022] Zhaowei Zhu, Zihao Dong, and Yang Liu. Detecting corrupted labels without training a model to predict. In *International Conference on Machine Learning*, pages 27412–27427. PMLR, 2022.