IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 17 April 2024, accepted 22 April 2024, date of current version 30 April 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3393440

RESEARCH ARTICLE

ConfidentMix: Confidence-Guided Mixup for Learning With Noisy Labels

RYOTA HIGASHIMOTO[1], **SOH YOSHIDA**[2], **(Member, IEEE), AND MITSUJI MUNEYASU**[2], **(Member, IEEE)**
[1]Graduate School of Science and Engineering, Kansai University, Suita, Osaka 564-8680, Japan
[2]Faculty of Engineering Science, Kansai University, Suita, Osaka 564-8680, Japan

Corresponding author: Soh Yoshida (sohy@kansai-u.ac.jp)

This work was supported by the Japan Society for the Promotion of Science (JSPS) KAKENHI, Japan, under Grant 22K18007.

ABSTRACT Deep neural networks (DNNs) have proven highly effective in various computational tasks, but their success depends largely on access to large datasets with accurate labels. Obtaining such data may be challenging and costly in real-world scenarios. Common alternatives, such as the use of search engines and crowdsourcing, often result in datasets with inaccurately labeled, or ''noisy,'' data. This noise may significantly reduce the ability of DNNs to generalize and maintain reliability. Traditional methods for learning with noisy labels mitigate this drawback by training DNNs selectively on reliable data, but they often underutilize available data. Although data augmentation techniques are useful, they do not directly solve the noisy label problem and are limited in such contexts. This paper proposes a confidence-guided Mixup named ConfidentMix, which is a data augmentation strategy based on label confidence. Our method dynamically adjusts the intensity of data augmentation according to label confidence, to protect DNNs from the detrimental effects of noisy labels and maximize the learning potential from the most reliable portions of the dataset. ConfidentMix represents a unique blend of label confidence assessment and customized data augmentation, and improves model resilience and generalizability. Our results on standard benchmarks with synthetic noise, such as CIFAR-10 and CIFAR-100, demonstrate the superiority of ConfidentMix in high-noise environments. Furthermore, extensive experiments on Clothing1M and mini-WebVision have confirmed that ConfidentMix surpasses state-of-the-art methods in handling real-world noise.

INDEX TERMS Data augmentation, deep learning, learning with noisy labels, semi-supervised learning.

I. INTRODUCTION

Deep neural networks (DNNs) have demonstrated exceptional performance in tasks such as image classification [1] and object detection [2], primarily because of the availability of large datasets with high-quality labels. However, acquiring precise labels is often difficult and expensive. Methods such as the use of search engines [3], [4] and crowdsourcing [5], [6] provide alternatives for data collection, but they tend to produce labels of lower quality. The problem of noisy labels may significantly affect DNNs because they may learn the incorrect annotations [7], leading to reduced generalization and robustness. Consequently, training deep learning models with noisy labels is a significant challenge.

Recently proposed methods for learning with noisy labels (LNL) in DNNs can be categorized into three approaches: regularization [9], [10], noisy label correction [11], [12], [13], and sample selection [14], [15]. In particular, sample selection methods, which are intended to isolate clean samples from noisy training datasets, have demonstrated impressive results. Sample selection methods are further divided into two categories: the small-loss approach and the feature representation approach. The former approach selects samples with minimal losses, according to a phenomenon known as the memorization effect [16], wherein a model initially learns simple patterns and gradually progresses to more complex patterns, such as those in data with noisy

The associate editor coordinating the review of this manuscript and approving it for publication was Nuno M. Garcia.

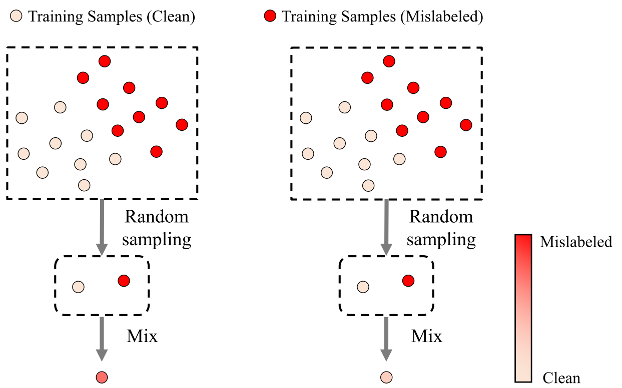

FIGURE 1. Left column: Data mixing process in Mixup [8]. Right column: Data mixing process in our proposed method. Traditional mixed data augmentation techniques that do not account for label confidence often produce mixed data with degraded labels. In contrast, our method weakens the mixing when label confidence is low, thereby preserving clean labels.

labels. The latter approach selects samples according to the similarity of feature representation. Although these methods are effective in removing noisy labels, they tend to reduce the volume of training data. A significant development in sample selection has been its integration with semi-supervised learning (SSL) [17], [18]. This strategy involves learning from large datasets that comprise a mix of few labeled samples with many unlabeled samples [19], [20], [21], [22]. It discards labels from samples that are removed during selection and generates pseudo-labels based on confident model predictions. This enables the use of the entire training dataset, thereby enhancing generalization performance and robustness to noisy labels.

Another method for learning from limited numbers of samples in the SSL context is data augmentation. Techniques such as AutoAugment [23] and RandAugment [24] have been proposed to enhance model robustness, focusing primarily on single-image transformation. However, most of these techniques fail to exploit inter-image information. Methods such as Mixup [8] and CutMix [25] were designed to address this drawback by mixing multiple data points to create new images and labels. Nonetheless, a significant concern with these techniques is their potential to corrupt clean labels, particularly in the context of LNL. This problem arises when features and labels from different images, including those with noisy labels, are indiscriminately combined. When a clean label is mixed with a noisy one, the resulting label may not accurately represent either of the original labels, thus becoming less reliable. Such mixing may introduce ambiguity and diminish the overall quality of labels, thereby impairing the learning efficiency of the model and its ability to generalize from clean, accurate data.

In this paper, we propose ConfidentMix, a confidence-guided Mixup-based data augmentation method, specifically designed for SSL. The unique feature of ConfidentMix is that it exploits pseudo-labels by assessing label confidence across the training dataset. It includes three key components: (i) *sampling reliable labels* (SRL), which uses a threshold-based method to selectively incorporate samples that have high-confidence labels; (ii) *label confidence estimation* (LCE), a process intended to evaluate label confidence to refine the augmentation strategy; and (iii) *confidence-guided mixing*, a method that adjusts the level of data mixing according to label confidence, thus maintaining the quality of clean labels and reducing the impact of noisy data.

The main difference between ConfidentMix and existing data augmentation techniques is its strategy for moderating the intensity of augmentation for pseudo-labels in noisy label situations based on estimated label confidence, as illustrated in Fig. 1. This strategy prevents the degradation of the quality of labels and pseudo-labels, to allow models to be trained with both augmented labeled data and high-quality labels. Thus, SSL-based models with ConfidentMix achieve improved generalization and robustness compared with models trained without consideration of label confidence.

The primary contributions of this paper include:

- We introduce a novel confidence label sampling technique that selectively harnesses high-confidence labels for more accurate model training.
- We develop a method to monitor and evaluate the selection of data with high-confidence labels, which effectively estimates the proportion of noisy labels and controls the distribution of random mixture weights.
- The proposed mixed data augmentation approach, ConfidentMix, can easily be integrated into existing frameworks that employ sample selection and SSL, as a consequence of its straightforward algorithm that prevents clean label corruption.
- Comprehensive experiments on widely recognized benchmark and real-world datasets in the LNL field have demonstrated that our method surpasses current state-of-the-art (SOTA) methods. In particular, our ConfidentMix-implemented SSL-based LNL method achieved a significant 15.1% improvement in scenarios characterized by high levels of noise.

The rest of this paper is organized as follows. Section II reviews related literature. Section III presents the preliminaries of our method. In Section IV, we describe the ConfidentMix method. Extensive empirical studies are reported in Section V, followed by our conclusions in Section VI.

II. RELATED WORK

A. SEMI-SUPERVISED LEARNING

SSL is an approach that learns from datasets that comprise a small number of labeled samples and a large number of unlabeled samples [26]. Recently, the predominant trend in SSL has been the use of consistency regularization. This technique enforces consistent model outputs for different transformations of the same instance. MixMatch [17] and FixMatch [18] are anchor-based methods that exploit augmentation consistency. MixMatch uses an average of sharpened model predictions for multiple weak augmentation versions as pseudo-labels, enhancing their effectiveness through the Mixup [8] strategy. In contrast, FixMatch boosts consistency regularization with strong data augmentation and applies a threshold for using only reliable pseudo-labels. Another approach involves adjusting class-specific confidence thresholds according to the difficulty of learning each class [27]. Finally, contrastive learning has been introduced to exploit instance similarity [28], [29].

B. LEARNING WITH NOISY LABELS

LNL is a serious challenge in real-world scenarios because labeled datasets often contain inaccuracies. Noisy labels may significantly degrade the performance of learning algorithms, leading to model overfitting on erroneous data and diminished generalization capabilities.

1) ROBUST LEARNING ALGORITHMS

Robust learning algorithms have been developed to handle label noise. These algorithms are intended to be insensitive to label noise; this is achieved by either modifying the loss function or adjusting the training process. The goal is to prevent models from fitting the noise in the data, thereby ensuring stable learning even when some labels are incorrect.

A phenomenon known as early learning [9] has been observed, whereby models tend to predict the true class labels correctly in the initial training phases, even in the presence of noisy data. The early-learning regularization technique, proposed by Liu et al. [9], uses a regularization term that optimizes the agreement between model predictions and an adaptive target probability, thereby lessening the impact of incorrect labels. This target probability is crafted using a temporal ensemble method [30], which relies on moving averages of model predictions.

2) NOISE DETECTION ALGORITHMS

Identifying and handling noisy labels is another important strategy. These algorithms are intended to distinguish between noisy and clean labels. They often employ a small-loss strategy that selects samples with small losses, to exploit the property that models initially predict true classes before memorizing noisy labels [9], [16]. Once they have been detected, noisy labels can be corrected or discarded to obtain a cleaner training dataset.

Sample selection is a method for selecting clean samples from noisy datasets. Many such techniques use multiple models to remove noisy samples. For instance, Jiang et al. [31] pretrain an additional model to select clean samples for the training of the main model. However, this method faces the challenge of error accumulation. To mitigate this, Co-teaching [14] proposes the training of one model using a small-loss strategy to train another. However, Co-teaching faces the problem of model convergence over time. Co-teaching+ [15] introduces decoupling [32] to force the two models to diverge. All the above approaches require careful calibration to avoid erroneous exclusion of genuine data points.

3) COMBINING SSL WITH SAMPLE SELECTION

Integrating SSL with sample selection offers a novel approach to LNL. This method trains the model using labels from a subset of samples that are considered correctly annotated, while treating the rest as unlabeled data. Research by Ding et al. [33] and Kong et al. [34] highlighted the potential of SSL in noisy environments, but its effectiveness decreases as label noise increases. DivideMix [19] is distinguished by its use of a "warm-up" phase and a Gaussian mixture model (GMM) to separate the dataset into clean labeled and noisy unlabeled samples. Fine [35] proposed the use of principal components of latent representations to filter label noise. Combined with contrastive learning, MOIT [20] identifies noisy labels by performing a k-nearest-neighbors search to quantify the agreement between feature representations and labels. LongReMix [21] extracts a reliable clean sample set through a two-stage clean sample selection process. OT-Filter [22] improves sample selection by optimizing the transport plan from the Euclidean space of feature vectors to the probability space. Other methods exist that, in addition to integrating SSL and sample selection, can be combined with regularization methods to effectively address noise challenges [36], [37]. RankMatch [38] is a recently proposed method that, to enhance the consistency between similar samples, introduces a rank contrastive loss based on the rank statistics of the principal features. Although RankMatch and ConfidentMix share the common goal of improving the robustness of LNL, they differ in their approach to sample selection and consistency regularization. RankMatch focuses on feature representation and does not explicitly integrate with SSL techniques, whereas ConfidentMix emphasizes confidence-guided mixing and the exploitation of pseudo-labels in the SSL context. As the primary focus of this paper is on the ConfidentMix method and its integration with SSL, we do not provide a quantitative comparison between ConfidentMix and RankMatch.

C. DATA AUGMENTATION

Data augmentation is intended to improve model generalization by enriching training data with transformations that preserve the semantic information of the input. Traditional image transformations include horizontal flipping, cropping, scaling, color distortion, and cut-out. Recent research has focused on finding optimal sequences of these transformations. AutoAugment [23] automates the search for the optimal augmentation sequence using predefined transformations. Although effective, its vast search space complicates training. RandAugment [24] simplifies this search space by combining random transformations, which significantly reduces computational cost. However, these approaches focus primarily on transformations on a single image and do not focus on extracting relationships and interactions among multiple images. Combining multiple images has the potential to produce richer training data by fusing new visual features and context.

1) MIXING AUGMENTATION

Recent studies have focused on data augmentation using multiple images [8], [25]. These approaches are intended to increase the diversity and novelty of visual patterns by combining images from a single class or from different classes in the training data. Because mixed data augmentation seeks to expand the training data distribution, the mixed images must remain close to the training data manifold. Mixup [8] achieves this objective by combining a pair of labels and images using convex linear interpolation. Conversely, CutMix [25] augments data by cutting a portion of one input image and overlaying it onto another. These techniques benefit both images and labels, reduce label overfitting, and enhance robustness of the model to noisy labels. However, both Mixup and CutMix were designed on the assumption that label information is correct, which could compromise clean labels in the presence of noisy labels. Additionally, models trained with corrupted labels cannot entirely eliminate noisy labels during sample selection, leaving many noisy labels in the labeled data that remain after selection. Our proposed method estimates label confidence and, according to that, carefully mixes labels with teacher labels, thereby preventing the degradation of clean labels and enhancing model generalization.

III. PRELIMINARIES

In this section, we describe SOTA methods that combine sample selection with SSL, along with Mixup, which is used in the semi-supervised phase. Because these methods form the foundation of our proposed approach, the notation introduced here will also be relevant in Section IV.

A. STATE-OF-THE-ART METHOD FOR LEARNING WITH NOISY LABELS

DivideMix and OT-Filter represent the latest techniques that combine sample selection with SSL. These methods involve dividing training data into a labeled set with clean labels and an unlabeled set with discarded noisy labels. After dividing the training data, the entire dataset is used through SSL approaches, exemplified by MixMatch. In this section, we take DivideMix as an example of a data partitioning method that uses sample selection. In DivideMix, losses are computed for all training data $D = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^{N}$, where N is the number of samples. These computed losses are then fitted to a two-component GMM to discern clean from noisy distributions, to obtain the probability w_i that $\mathbf{x}_i$ belongs to the clean distribution. The dataset is subsequently divided into

$$\mathcal{X} = \{(\mathbf{x}_i, \mathbf{y}_i)|w_i \geq \tau_w, i \in \{1, \ldots, N\}\},$$
$$\mathcal{U} = \{\mathbf{x}_i|w_i < \tau_w, i \in \{1, \ldots, N\}\}, \quad (1)$$

where τ_w denotes the threshold parameter. Subsequently, unlabeled samples in $\mathcal{U}$ are assigned pseudo-labels $\hat{\mathbf{y}}_i$ that are generated from the model's predictions, following MixMatch, to form the pseudo-labeled sample set $\mathcal{P} = \{(\mathbf{x}_i, \hat{\mathbf{y}}_i)\}_{i=1}^{N-N_x}$, where N_x is the number of labeled samples in $\mathcal{X}$. DivideMix further enhances this approach by incorporating label refinement using GMM predictive probabilities, thus improving MixMatch. The refinement of a label for sample $\mathbf{x}_i$ with network prediction $\mathbf{p}_i$ is performed as follows:

$$\bar{\mathbf{p}}_i = w_i\mathbf{y}_i + (1 - w_i)\mathbf{p}_i,$$
$$\bar{\mathbf{y}}_i = \text{sharpen}(\bar{\mathbf{p}}_i, T), \quad (2)$$

where $\text{sharpen}(\mathbf{p}, T) = \mathbf{p}^{\frac{1}{T}} / \sum_c p_c{}^{\frac{1}{T}}$, p_c denotes the score for class c, and T denotes the temperature for sharpening.

B. DATA AUGMENTATION USING MIXUP

Mixup is a data augmentation technique that ensures that mixed images reside near the training data manifold using convex linear interpolation. It combines randomly sampled image pairs $(\mathbf{x}_i, \mathbf{x}_j)$ and their corresponding label pairs $(\mathbf{y}_i, \mathbf{y}_j)$ to create new mixed images $\tilde{\mathbf{x}}_i$ and mixed labels $\tilde{\mathbf{y}}_i$ as follows:

$$\lambda' = \max(\lambda, 1 - \lambda), \quad (3)$$
$$\tilde{\mathbf{x}}_i = \lambda'\mathbf{x}_i + (1 - \lambda')\mathbf{x}_j, \quad (4)$$
$$\tilde{\mathbf{y}}_i = \lambda'\mathbf{y}_i + (1 - \lambda')\mathbf{y}_j, \quad (5)$$

where λ is a random mixing weight drawn from a beta distribution $\text{Beta}(\alpha, \alpha)$ with parameter $\alpha \in (0, \infty]$.

In methods that combine sample selection with SSL, the image and label pair $(\mathbf{x}_i, \mathbf{y}_i)$ is randomly sampled from the pseudo-labeled sample set $\mathcal{P}$, and $(\mathbf{x}_j, \mathbf{y}_j)$ is sampled from the union of the labeled and pseudo-labeled sample sets $(\mathcal{X} \cup \mathcal{P})$. However, in scenarios in which training data noise is prevalent or there are numerous classification classes, the accuracy of sample selection decreases, leaving the labeled sample set with many noisy labels. When noisy labels are sampled as $\mathbf{y}_j$, the resulting pseudo-labels are corrupted and the generated mixed labels $\tilde{\mathbf{y}}_i$ may diminish the robustness of the model to noisy labels, leading to a decrease in generalization performance.

IV. PROPOSED METHOD

This section introduces ConfidentMix, our proposed data augmentation method designed for SSL in environments with noisy labels. ConfidentMix is based on the principle of using only high-confidence labels and pseudo-labels for data augmentation. This ensures that only high-quality labels are employed in SSL.

The ConfidentMix framework, illustrated in Fig. 2, consists of three main components: (i) *SRL* (Section IV-A), which involves choosing for sampling only those labels or

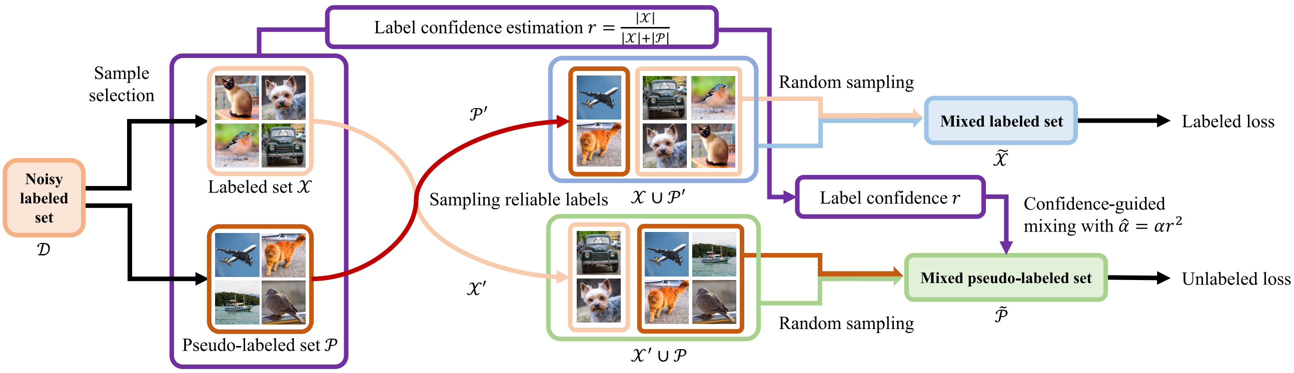

FIGURE 2. Framework of the proposed ConfidentMix method.

pseudo-labels that are considered trustworthy; (ii) *LCE* (Section IV-B), which assesses the confidence of labels in labeled data; and (iii) *confidence-guided mixing* (Section IV-C), which employs data augmentation with restricted random mixing weights according to estimated label confidence.

A. SAMPLING RELIABLE LABELS

In the context of mixing augmentation, notably that performed by the Mixup technique, a key challenge is the unintentional inclusion of noisy labels in the mixing process; this occurs when $\mathbf{y}_j$ in (5) is noisy. Mixup, which samples all training data randomly, occasionally results in the selection of noisy labels for mixing. Inspired by the approaches of [18] and [27], we define a threshold, denoted by τ, to ensure that only the samples with the most reliable labels are chosen, as follows:

$$\mathcal{X}' = \{(\mathbf{x}_i, \bar{\mathbf{y}}_i) | \max_c(\bar{\mathbf{y}}_i) > \tau, i \in \mathcal{X}\}, \tag{6}$$

$$\mathcal{P}' = \{(\mathbf{x}_i, \hat{\mathbf{y}}_i) | \max_c(\hat{\mathbf{y}}_i) > \tau, i \in \mathcal{P}\}, \tag{7}$$

where $c \in \{1, \ldots, C\}$ denotes a classification category and C is the total number of such categories. For labeled data, we can prevent the corruption of these labels by mixing them with $(\mathbf{x}_j, \mathbf{y}_j)$, which are randomly sampled from a combination of the labeled dataset with the reliable pseudo-labeled dataset $(\mathcal{X} \cup \mathcal{P}')$. For pseudo-labeled data, to avoid corruption of pseudo-labels, we use $\mathbf{y}_j$, which is randomly selected from a combined set of reliable labeled and pseudo-labeled samples $(\mathcal{X}' \cup \mathcal{P})$ during the mixing process.

Following the existing methods [18], [27], our proposed method uses a threshold to select samples with reliable labels. However, our approach differs significantly from these methods in two respects: (i) We focus on selecting reliable labels for use in mixing augmentation, rather than for samples used in backpropagation. (ii) We apply this selection process to both corrected labels and pseudo-labels.

B. LABEL CONFIDENCE ESTIMATION

The corruption of pseudo-labels in mixing augmentation is largely attributed to mixing ratios whose design does not consider label confidence. For labeled data that are still replete with noisy labels, the mixing ratio requires careful calibration. ConfidentMix approaches this task by incorporating label confidence into the design of mixing ratios. To estimate label confidence, we calculate the proportion of clean labels in the labeled data; this is also referred to as the label precision. In real-world scenarios where the clean labels are unknown, accurately determining label precision is a challenge. Therefore, we approximate label precision as the ratio of the number of labeled samples to the number of labeled and pseudo-labeled samples, as follows:

$$r = \frac{|\mathcal{X}|}{|\mathcal{X}| + |\mathcal{P}|}, \tag{8}$$

where $\mathcal{X}$ and $\mathcal{P}$ are the sets of labeled and pseudo-labeled samples, respectively. When noisy labels are present in the training data, the decrease in sample selection accuracy increases the frequency of noisy labels in the labeled data, and the r obtained by (8) decreases. Specifically, when the noise rate is high, many samples are classified as noisy, resulting in a small value of r. Conversely, when the noise rate is low, sample selection works effectively and the value of r is high because there are fewer noisy labels in the labeled data. Thus, we define r as a measure of label confidence and use it to design an appropriate mixing ratio.

These trends have been observed and validated in preliminary experiments. Fig. 3 shows the relationship between true label precision and the calculated labeled sample ratio r, using CIFAR-100 as the dataset. Further details of the

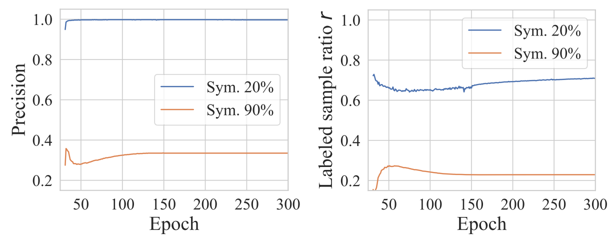

FIGURE 3. Experimental results of DivideMix on CIFAR-100 with symmetric noise.

Algorithm 1 ConfidentMix

inputs: Labeled data $\mathcal{X} = \{(\mathbf{x}_i, \bar{\mathbf{y}}_i)\}_{i=1}^{N_x}$, pseudo-labeled data $\mathcal{P} = \{(\mathbf{x}_i, \hat{y}_i)\}_{i=1}^{N-Nx}$, beta distribution parameter α, threshold τ

output: Mixed dataset $\tilde{\mathcal{X}}, \tilde{\mathcal{P}}$

$\mathcal{X}' \leftarrow$ (6) ▷ Sampling reliable labels
$\mathcal{P}' \leftarrow$ (7) ▷ Sampling reliable labels
$r \leftarrow$ (8) ▷ Label confidence estimation
$\hat{\alpha} \leftarrow \alpha r^2$
$\mathcal{D}_x \leftarrow \text{shuffle}(\mathcal{X} \cup \mathcal{P}')$
$\mathcal{D}_p \leftarrow \text{shuffle}(\mathcal{X}' \cup \mathcal{P})$
$\tilde{\mathcal{X}} = \mathit{Mixup}(\mathcal{X}, \mathcal{D}_x, \alpha)$
$\tilde{\mathcal{P}} = \mathit{Mixup}(\mathcal{P}, \mathcal{D}_p, \hat{\alpha})$ ▷ Confidence-guided mixing

experimental setups are presented in Section V. It is evident from Fig. 3 that, as the noise rate increases, both the actual and estimated label precisions decrease. It is important to note that, although the label confidence r does not have to match the label precision precisely, the two values should follow a similar trend.

C. CONFIDENCE-GUIDED MIXING

In Mixup, random mixing weights are derived from a beta distribution using the parameter α. In contrast, ConfidentMix tailors the design of these weights by incorporating the estimated label confidence r. This approach differs from that of recent modifications to the weight design of Mixup, such as those proposed by Liu et al. [39] and Mai et al. [40]. Specifically, ConfidentMix restricts the distribution of random mixing weights in accordance with label confidence. The equations for determining the restricted distribution parameters and random mixing weights are as follows:

$$\begin{aligned} &\hat{\alpha} = \alpha r^2, \\ &\lambda_c \sim \text{Beta}(\hat{\alpha}, \hat{\alpha}). \end{aligned} \tag{9}$$

Using the formulated weight λ_c, ConfidentMix generates mixed data $\tilde{\mathbf{x}}_i$ and mixed labels $\tilde{\mathbf{y}}_i$, similarly to the process described in (3–5) for Mixup. A notable aspect of the random mixing weights of ConfidentMix is their ability to preserve clean labels by considering label confidence, while simultaneously introducing diversity in the mixed data through an element of randomness. Specifically, higher label confidence leads to a larger $\hat{\alpha}$, which increases the likelihood of assigning a smaller weight to the original label $\mathbf{y}_i$ and thus facilitating more extensive data mixing. In contrast, lower label confidence results in a smaller $\hat{\alpha}$, which favors the retention of the original label $\mathbf{y}_i$. The complete mixing augmentation process is outlined in Algorithm 1.

D. APPLICATION TO SSL IN NOISY ENVIRONMENTS

Our method, ConfidentMix, generates mixed data that contain reliable labels and results in the mixed sample sets $\tilde{\mathcal{X}}$ and $\tilde{\mathcal{P}}$. This mixing augmentation integrates seamlessly with SSL. As an illustrative example, we combine MixMatch with SSL in the framework of the SOTA method [22]. The combined training process, which integrates SSL with ConfidentMix, is outlined in Algorithm 2. In this implementation, label refinement is conducted using GMM predictive probabilities and the sharpening function, as in DivideMix [19]. To strengthen consistency regularization for the clean labels maintained by our proposed method, we use RandAugment during the semi-supervised phase. This enhances the robustness of the training process. In Algorithm 2, CE and MSE represent the cross-entropy loss and mean square loss, respectively.

Algorithm 2 SSL With ConfidentMix

inputs: Labeled data $\mathcal{X} = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^{N_x}$, unlabeled data $\mathcal{U} = \{\mathbf{u}_i\}_{i=1}^{N-N_x}$, network's output softmax probability p_{model}, network parameter θ, weight for unlabeled loss λ_u, batch size B, number of augmentations K, sharpening temperature T, beta distribution parameter α, threshold τ, predictive probabilities of GMM $\mathcal{W} = \{w_i\}_{i=1}^{N_x}$

while b in B **do**
  for k in K **do**
    $\hat{\mathbf{x}}_{b,k} \leftarrow \text{augment}(\mathbf{x}_b)$
    $\hat{\mathbf{u}}_{b,k} \leftarrow \text{augment}(\mathbf{u}_b)$
  end for
  $\mathbf{p}_b \leftarrow \frac{1}{K} \sum_k p_{model}(\hat{\mathbf{x}}_{b,k} : \theta)$
  $\bar{\mathbf{p}}_b \leftarrow w_b \mathbf{y}_b + (1 - w_b)\mathbf{p}_b$
  $\bar{\mathbf{y}}_b \leftarrow \text{sharpen}(\bar{\mathbf{p}}_b, T)$
  $\mathbf{q}_b \leftarrow \frac{1}{K} \sum_k p_{model}(\hat{\mathbf{u}}_{b,k} : \theta)$
  $\hat{\mathbf{y}}_b \leftarrow \text{sharpen}(\mathbf{q}_b, T)$
end while
$\hat{\mathcal{X}} = \{(\hat{\mathbf{x}}_b, \bar{\mathbf{y}}_b)\}_{b=1}^{B}, \mathcal{P} = \{(\hat{\mathbf{u}}_b, \hat{\mathbf{y}}_b)\}_{b=1}^{B}$
$\tilde{\mathcal{X}}, \tilde{\mathcal{P}} \leftarrow \mathit{ConfidentMix}(\hat{\mathcal{X}}, \mathcal{P}, \alpha, \tau)$
$\mathcal{L}_x \leftarrow \text{CE}(\tilde{\mathcal{X}}), \mathcal{L}_u \leftarrow \text{MSE}(\tilde{\mathcal{P}})$
$\mathcal{L} \leftarrow \mathcal{L}_x + \lambda_u \mathcal{L}_u$

V. EXPERIMENTS

A. TYPES OF NOISE

Real-world noise patterns can be categorized into symmetric (Sym.) noise [41], asymmetric (Asym.) noise [42], and instance-dependent noise [43]. Symmetric noise is generated by randomly replacing labels of a specified proportion of samples with labels of other classes. Asymmetric noise is created by switching labels between similar classes, thereby reflecting class-dependent real-world label noise. Instance-dependent noise is a more complex type of label noise that depends on both the class and the specific features of each instance [44]. Our experiments considered datasets affected by each of these noise patterns. Additionally, the real-world datasets used in our experiments are among the top two in terms of noise rate according to the most recent survey [45]. Therefore, it is important to note that the complexity and size of these datasets are carefully selected benchmarks for evaluating the effectiveness of our ConfidentMix.

TABLE 1. List of λ_u values used in the CIFAR-10 and CIFAR-100 datasets.

Dataset	CIFAR-10					CIFAR-100				
Noise type	Sym.				Asym.	Sym.				Asym.
Method/Noise ratio	20%	50%	80%	90%	40%	20%	50%	80%	90%	40%
λ_u	3	25	25	50	3	25	150	150	150	150

B. DATASETS

1) CIFAR-10 AND CIFAR-100

We evaluated our method on CIFAR-10 and CIFAR-100 [46] using synthetic noise. Both datasets contain 50K training and 10K test images of size 32 × 32. CIFAR-10 and CIFAR-100 contain 10 and 100 classification classes, respectively. Because it is difficult to identify noise characteristics in real-world environments in advance, synthetic noise, whose noise rate can be controlled, is commonly used to assess learning algorithms. Hence, following previous studies [19], [22], we simulated two types of synthetic noise: symmetric and asymmetric. Symmetric noise was implemented by changing labels of a predefined percentage of the training data (20%, 50%, 80%, or 90%) to random classes. Asymmetric noise, which considers the semantic information of classes, was generated by changing labels to similar classes (e.g., truck→automobile). We conducted experiments on datasets containing 40% asymmetric noise because distinguishing classes with a noise rate exceeding 50% is practically impossible [12]. The noise rates used in our experiments were also set according to previous studies [19], [22] for a fair comparison.

2) CLOTHING1M

Clothing1M [47] is a large-scale dataset that contains over 1M training images collected from online shopping sites, annotated with noisy labels. The labels, based on text provided by sellers, include 14 classes and the estimated noise rate is 38.5% [48]. Statistical hypothesis testing conducted by Chen et al. [49] has shown this noise to be instance-dependent. This dataset also provides 14K validation and 10K test images. We sampled 1000 mini-batches from the training data, following a previous study [19].

3) WEBVISION

WebVision [3] consists of web images with noisy labels crawled from the Internet, using keywords from 1K classes of ImageNet ILSVRC12 [50]. The noise rate of these labels is estimated to be approximately 20% [3]. Following recent studies [9], [19], we used the first 50 classes of the subset of WebVision that comprises Google images, referred to as mini-WebVision, and evaluated our method on both the WebVision and ImageNet validation sets.

C. IMPLEMENTATION DETAILS

To avoid self-implementation bias, we chose DivideMix, for which official experimental code is available,[1] to combine with our ConfidentMix. Thus, most details of our experiments are as described in [19]. Our method focuses on the problem of mixing data augmentation in methods that combine sample selection with SSL. Therefore, only methods that satisfy this condition are included in our comparison.

For CIFAR-10 and CIFAR-100, we used PreActResNet18 [51] as the backbone and trained the model using a stochastic gradient descent (SGD) optimizer with a momentum of 0.9 and weight decay of 5×10^{-4}. The learning rate was initialized to 0.02 and decayed by a factor of 10 after 150 epochs. The model was trained for 300 epochs with a batch size of 128. These conditions and the hyperparameter settings were the same as those of DivideMix because we integrated ConfidentMix with DivideMix for the performance evaluation. However, for experiments on CIFAR-10 with 20% symmetric and 40% asymmetric noise, we adjusted λ_u from 0 to 3 to maximize the effect of our proposed method. We focused on Clothing1M, which has similar noise rates and the same number of classification classes as CIFAR-10, and conducted a parameter search using the validation set provided in Clothing1M to determine the optimal λ_u values for these noise rates. Table 1 lists the λ_u values used for the experiments on CIFAR-10 and CIFAR-100. For the experiments on CIFAR-10 with noise patterns other than 20% symmetric and 40% asymmetric noise, the λ_u values are the same as those used in DivideMix.

Threshold parameter τ was set using the dataset with 90% symmetric noise as a proxy validation set, following the experimental setup in [12] and [19]. We searched for the optimal τ value from the parameter set $\{0.1, 0.3, 0.5, 0.7, 0.95\}$ and selected 0.95 for CIFAR-10 and 0.7 for CIFAR-100. These selected values were then used consistently across all noise patterns.

For Clothing1M, following [19] and [22], we used a ResNet50 [1] pretrained on ImageNet as the backbone, and trained the model using an SGD optimizer with a momentum of 0.9 and weight decay of 10^{-3}. The learning rate was set to 0.002 and reduced by a factor of 10 after 40 epochs. The model was trained for 80 epochs with a batch size of 32. We set $\lambda_u = 3$ and $\tau = 0.3$, and the other hyperparameter settings were the same as those of DivideMix.

For WebVision, we used Inception-ResNet-v2 [52] as the backbone, and trained the model using an SGD optimizer with a momentum of 0.9 and weight decay of 5×10^{-4} for 100 epochs. The batch size was 32, and the learning rate was set to 0.01 and reduced by a factor of 10 after 50 epochs. We set $\lambda_u = 5$ and $\tau = 0.7$, and the remaining hyperparameter settings were the same as those

[1] https://github.com/LiJunnan1992/DivideMix

TABLE 2. Comparison of test accuracy (%) on CIFAR-10 and CIFAR-100 with symmetric and asymmetric noise. Results of baseline methods were copied from the respective papers. The best result is highlighted in bold. The differences between DivideMix and the proposed method for each noise rate are noted at the bottom.

Dataset	CIFAR-10					CIFAR-100				
Noise type	Sym.				Asym.	Sym.				Asym.
Method/Noise ratio	20%	50%	80%	90%	40%	20%	50%	80%	90%	40%
Cross-Entropy	86.8	79.4	62.9	42.7	83.2	62.0	46.7	19.9	10.1	-
Mixup ('17) [8]	95.6	87.1	71.6	52.2	-	67.8	57.3	30.8	14.6	-
Co-teaching+ ('18) [14]	89.5	85.7	67.4	47.9	71.3	65.6	51.8	27.9	13.7	-
DivideMix ('20) [19]	96.1	94.6	93.2	76.0	93.4	77.3	74.6	60.2	31.5	59.1
Fine ('21) [35]	91.0	87.3	69.4	-	89.5	70.3	64.2	25.6	-	61.7
MOIT ('21) [20]	93.1	90.0	79.0	69.6	92.0	73.0	64.6	46.6	36.0	55.0
Fine + DivideMix ('21) [35]	96.1	94.9	93.5	90.5	93.8	**79.1**	74.6	61.0	34.3	-
LongReMix ('23) [21]	96.3	95.1	93.8	79.9	94.7	77.9	75.5	62.3	34.7	59.8
OT-Filter ('23) [22]	96.0	**95.3**	**94.0**	90.5	**95.1**	76.7	73.8	61.8	**42.8**	**76.5**
ConfidentMix + DivideMix (**ours**)	**96.4**	**95.3**	93.3	**91.1**	93.6	78.0	**76.4**	**64.8**	39.3	64.1
	+0.3	+0.7	+0.1	+15.1	+0.2	+0.7	+1.8	+4.6	+7.8	+5.0

of DivideMix. For both Clothing1M and WebVision, we performed a grid search for λ_u in $\{3, 5\}$ and τ in $\{0.3, 0.7\}$ to determine the optimal combination.

D. EXPERIMENTAL RESULTS

1) RESULTS ON CIFAR-10 AND CIFAR-100

Table 2 shows the test accuracy results on CIFAR-10 and CIFAR-100 using symmetric and asymmetric noise, and compares ConfidentMix with several SOTA methods. The comparison includes traditional approaches, such as Cross-Entropy, and innovative methods, such as Mixup [8], Co-teaching+ [14], DivideMix [19], Fine [35], MOIT [20], LongReMix [21], and OT-Filter [22]. These methods represent a range of techniques, from traditional strategies to more complex mix-based approaches and recent sample selection approaches. We also evaluated the effectiveness of the proposed method as an embedded method by comparing it with Fine + DivideMix.

In this comparison, ConfidentMix achieved favorable results across all noise patterns, particularly in challenging scenarios with high noise rates or a large number of classes. On CIFAR-10 with 90% symmetric noise, our proposed method both outperformed DivideMix by 15.1% and achieved a 0.6% higher accuracy than the best performing SOTA method. On CIFAR-100 with 50% and 80% symmetric noise, it outperformed the best SOTA methods by 0.9% and 2.5%, respectively. Although OT-Filter achieved superior performance on CIFAR-100 with 90% symmetric and 40% asymmetric noise, our method clearly demonstrated its effectiveness by significantly improving the accuracy of DivideMix. Moreover, compared with the combination of DivideMix with the approach of [35], ConfidentMix was significantly superior for several noise patterns. In summary, the comparison verified the effectiveness of the proposed method as an embedded method.

2) RESULTS ON CLOTHING1M

Table 3 shows the test accuracy results on Clothing1M of our proposed method in comparison with traditional and SOTA methods. It is important to note that the DivideMix results on Clothing1M reported in the recent work [12], [22] are worse than those published in [19]. Therefore, to enable a fair comparison, we present a result obtained from our experiments using the authors' public code with the same random seeds as ours. Our experimental results demonstrate that ConfidentMix is effective in dealing with complex real-world noise patterns and outperforms the SOTA methods. This improvement highlights the robustness of ConfidentMix and its adaptability to various types of noise, particularly in complex datasets such as Clothing1M, where label noise is prevalent and diverse. The integration of ConfidentMix with DivideMix both enhances overall accuracy and underscores

TABLE 3. Test accuracy (%) on Clothing1M. The * annotation indicates results obtained using the official experimental code. Baseline results were copied from the respective papers. The best result is highlighted in bold.

Method	Test Accuracy
Cross-Entropy	69.20
Co-teaching ('18) [14]	71.70
DivideMix* ('20) [19]	74.11
Fine ('21) [35]	72.91
Fine+DivideMix ('21) [35]	74.37
LongReMix ('23) [21]	74.38
OT-Filter ('23) [22]	74.50
ConfidentMix + DivideMix (**ours**)	**74.84**

TABLE 4. Results (top1 and top5 accuracy) on mini-WebVision. Baseline results were copied from the respective papers. The best result is highlighted in bold.

Dataset	WebVision		ILSVRC12	
Method	top1	top5	top1	top5
Decoupling ('17) [32]	62.54	84.74	58.26	82.26
MentorNet ('18) [31]	63.00	81.40	57.80	79.92
Co-teaching ('18) [14]	63.58	85.20	61.48	84.70
DivideMix ('20) [19]	77.32	91.64	75.20	90.84
MOIT ('21) [20]	77.90	91.90	73.80	91.70
LongReMix ('23) [21]	78.92	92.32	-	-
ConfidentMix + DivideMix (**ours**)	**79.62**	**93.06**	**76.81**	**92.94**

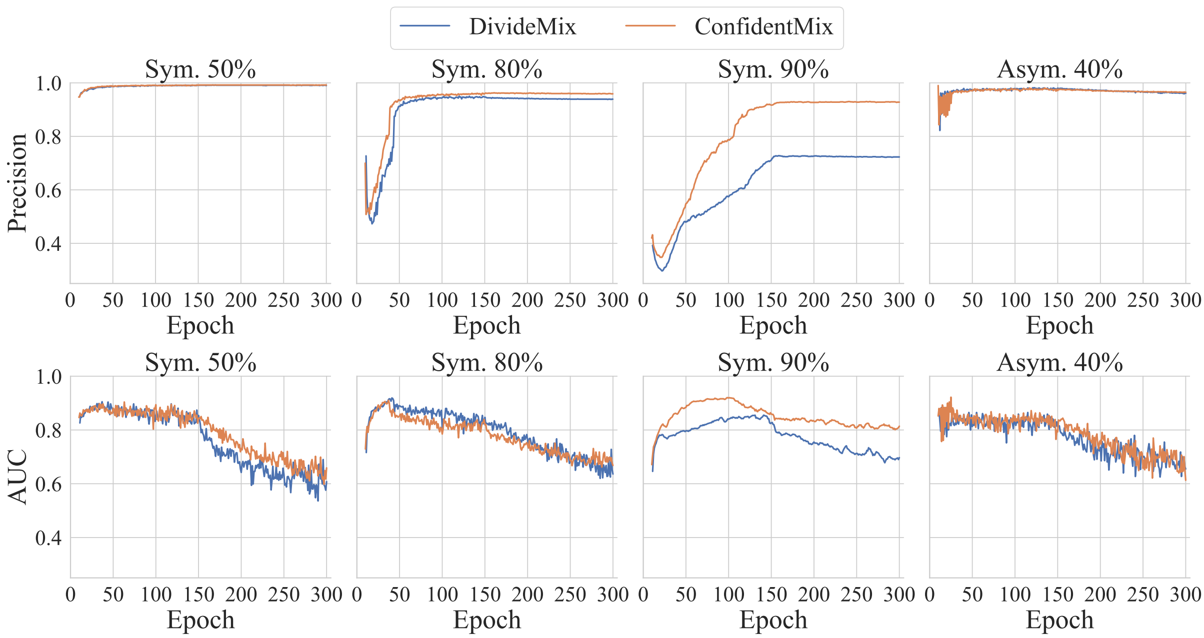

FIGURE 4. **Ablation study for the quality of sample selection on CIFAR-10 with symmetric and asymmetric noise.**

the potential benefits of combining different methodologies to effectively address challenges in training models on data with noisy labels. This achievement demonstrates the versatility and practicality of ConfidentMix in real-world applications where data quality is often unpredictable and imperfect.

3) RESULT ON WEBVISION

Table 4 presents the results of our proposed method compared with traditional and SOTA approaches on the real-world mini-WebVision dataset. In Table 4, we report the top1 and top5 classification accuracies following the conventions used in the existing literatures [19] and [21]. Top1 accuracy refers to the percentage of test instances for which the model's top predicted class matches the true class label. Top5 accuracy, by contrast, is a more relaxed metric that considers a prediction as correct if the true class label is among the model's top five predicted classes. Reporting both top1 and top5 accuracies provides a more comprehensive evaluation of the model's performance. It is worth nothing that OT-filter was excluded because the experiments in this dataset were not included in the paper. Our approach, which combines ConfidentMix with DivideMix, demonstrated superior performance, surpassing the SOTA methods. In particular, it enhanced the performance of DivideMix by 2.60% and 1.61% on mini-Webvision and ILSVRC12, respectively. This achievement is significant, particularly considering the challenging nature of the mini-WebVision dataset, which is known for its real-world complexity and noisy labels. The improvement in both the top1 and top5 accuracy metrics demonstrates that ConfidentMix effectively handles the noise and variability inherent in real-world data.

E. ABLATION STUDY

1) ANALYSIS OF LABEL NOISE DETECTION

Figs. 4 and 5 present the comparative results of ConfidentMix and DivideMix with respect to sample selection quality, using synthetic noise on the CIFAR-10 and CIFAR-100 datasets. The noise rates evaluated were 50%, 80%, and 90% for symmetric noise and 40% for asymmetric noise. The top row of each figure, labeled "Precision," plots the fraction of samples in the predicted clean labeled set that are indeed clean. The bottom row, labeled "AUC," plots the area under the receiver operating characteristic curve (AUC). This takes into account both the clean label proportion and how well the model fits to labels that are incorrectly classified as clean. ConfidentMix consistently improved the quality of sample selection in high-noise scenarios (80% or more) on both CIFAR-10 and CIFAR-100. In particular, our method suppressed fitting to noisy labels in the later stages of training (post-150 epochs) by preventing the corruption of clean labels, which greatly improved the AUC. We focused on the problem of the degradation of sample selection quality in high-noise conditions, where many noisy labels in the labeled samples could lead to the generation of corrupted mixed data with clean labels. Therefore, these results align with our hypothesis and demonstrate that our method can retain a high proportion of clean labels in high-noise environments.

This analysis shows the effectiveness of ConfidentMix in accurately distinguishing between clean and noisy labels,

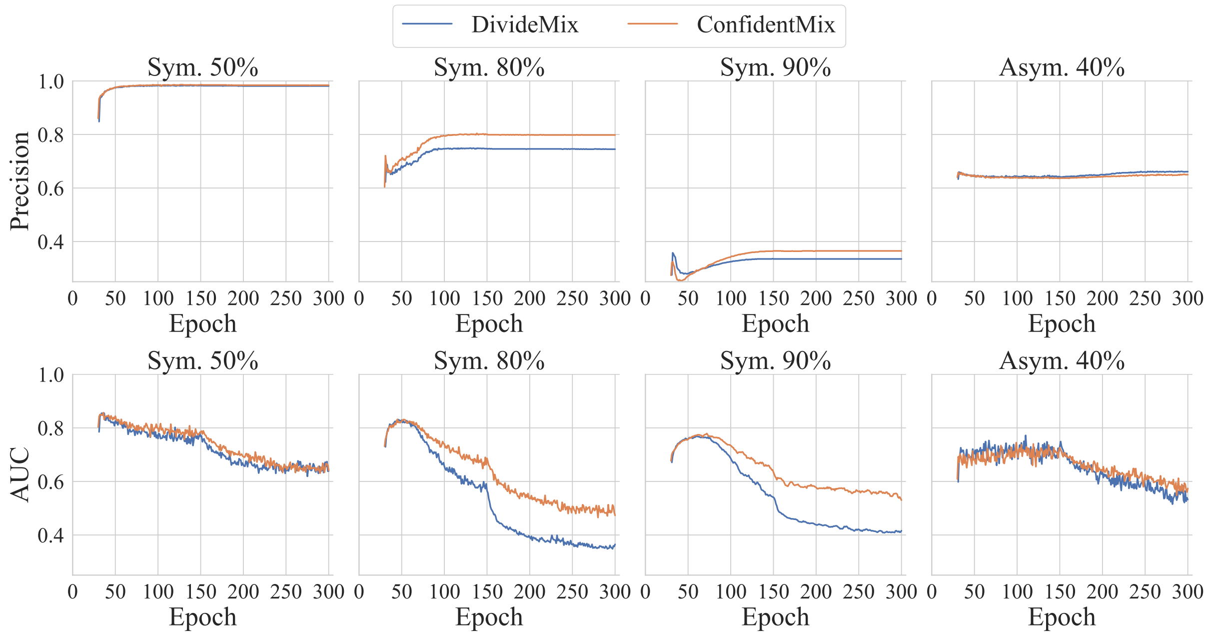

FIGURE 5. Ablation study of the quality of sample selection on CIFAR-100 with symmetric and asymmetric noise.

TABLE 5. Ablation study of the effect of each component. We abbreviate ***sampling reliable labels***, ***label confidence estimation***, and RandAugment as SRL, LCE, and RA, respectively.

Dataset	CIFAR-10		CIFAR-100		
Noise type	Sym.	Asym.	Sym.		Asym.
Method/Noise ratio	90%	40%	50%	90%	40%
ConfidentMix	91.1	93.6	**76.4**	**39.3**	**64.1**
ConfidentMix w/o LCE	90.8	**93.8**	75.8	31.1	62.2
ConfidentMix w/o SRL	**91.3**	93.6	76.2	**39.3**	61.5
DivideMix with RA	78.7	92.9	74.4	31.5	59.7

particularly in challenging high-noise conditions. The ability to maintain high precision in label noise detection is vital for the robust performance of learning algorithms in noisy environments. The findings suggest that ConfidentMix could be a valuable tool for enhancing data integrity and improving learning outcomes in scenarios where label quality cannot be guaranteed.

2) THE EFFECT OF EACH COMPONENT

In our ablation study, we analyzed the impact of the key components of our method: LCE and SRL. Additionally, we compared our method with RandAugment [24], which is known to be one of the best augmentation methods, to evaluate the effectiveness of ConfidentMix in noisy environments. The results are presented in Table 5.

LCE plays a crucial role in assessing the confidence of labels, which is vital for reducing pseudo-label corruption. Our study showed that omitting LCE often leads to decreased performance, underlining its significance in the design of weights that consider label confidence. SRL, which involves the selection of samples whose labels are highly reliable, proved to be particularly effective. This was most apparent in scenarios with asymmetric noise, where models are prone to make ambiguous predictions for similar classes, and in datasets that have fewer classification categories, such as CIFAR-10.

When SRL was excluded, we observed a minor decline in performance at high noise ratios, which suggests that the high-quality labels generated through LCE effectively surpassed the threshold set by SRL. However, the performance on CIFAR-100 with 40% asymmetric noise deteriorated significantly, confirming the importance of SRL.

In contrast to LCE and SRL, although RandAugment is a high-performance data augmentation technique, it is not tailored to noisy environments. Its inclusion with DivideMix led to performance decreases in most cases, except on CIFAR-10 with 90% symmetric noise. This comparison highlights the adaptability and robustness of ConfidentMix in handling noisy data, which is a challenging task that RandAugment does not specifically address.

3) SENSITIVITY TO HYPERPARAMETER

We evaluated the impact of the label confidence threshold parameter, denoted by τ, under 90% symmetric noise conditions on CIFAR-10 and CIFAR-100. Fig. 6 plots the variation in test accuracy as τ varied from 0.1 to 0.95. This evaluation demonstrates the robustness of ConfidentMix to variations in hyperparameters. The threshold τ is pivotal in the selection of reliable labels. It must be sufficiently large to ensure accurate selection but not so large as to be counterproductive, particularly in scenarios such as CIFAR-100, where a larger number of classes may lead to lower prediction probabilities. Thus, for these scenarios,

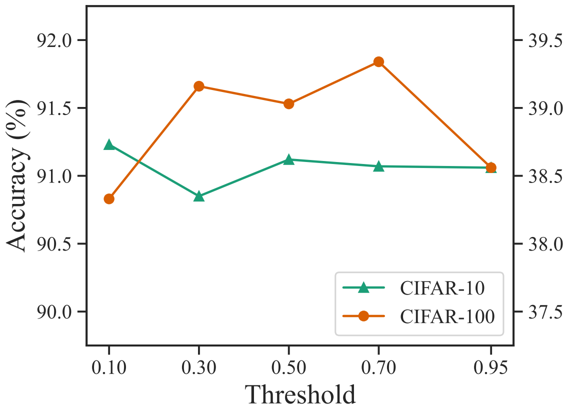

FIGURE 6. Ablation study with 90% symmetric noise for hyperparameter τ. The left and right axes represent the test accuracy achieved on CIFAR-10 and CIFAR-100, respectively.

setting τ to an excessively high value may result in the exclusion of actual clean labels and affect the overall effectiveness of the model. However, although setting τ to a small value may include a higher proportion of noisy labels, it also allows the capture of a broader range of potentially clean labels, which can be advantageous in scenarios with lower prediction certainty.

Our analysis revealed that the performance of ConfidentMix remained stable across a broad range of τ values, indicating its resilience to variations in hyperparameter settings. Nonetheless, it is crucial to calibrate τ appropriately for different datasets.

F. LIMITATION

ConfidentMix functions by estimating label confidence, which it does by monitoring fluctuations in the volume of training data that has been categorized through sample selection. The precision of this estimation hinges on the effectiveness of the sample selection algorithm. For instance, in experiments with the CIFAR-100 dataset involving 40% asymmetric noise, ConfidentMix demonstrated notable accuracy improvements when used in conjunction with DivideMix. However, its performance compared with that of OT-Filter was somewhat lower. This observation suggests that, although ConfidentMix is effective in scenarios with high noise levels, its performance on data with low to moderate noise levels largely relies on the capability of the sample selection algorithm to accurately differentiate between reliable and noisy labels. Additionally, ConfidentMix is optimized for integration with SSL-based LNL approaches. Consequently, adapting ConfidentMix to methods that do not include a division of training data (specifically those outside the purview of combining sample selection with SSL) remains a substantial challenge.

VI. CONCLUSION

In this paper, we addressed the problem of label corruption in mixing augmentation, which is a significant challenge in LNL. We introduced ConfidentMix, an innovative mixing augmentation technique designed to create training samples in which clean labels are preserved. This method can be effectively integrated with existing sample selection and SSL strategies. In contrast to typical mixing augmentation methods that rely on completely random mixing, ConfidentMix enhances the quality of labels in mixed data. It achieves this by carefully selecting samples with reliable labels and calibrating mixing weights according to the assessed label confidence. The efficacy of ConfidentMix was validated through comprehensive experiments on benchmark datasets with synthetic noise, in addition to real-world datasets. It outperformed SOTA methods on all real-world datasets and on the synthetic noise dataset for multiple noise rates. In particular, our method achieved a remarkable accuracy improvement of 15.1% in scenarios with the high noise rate of 90%.

There are two directions for future work. First, although ConfidentMix has been primarily designed and evaluated for multi-class classification problems, its key components – SRL, LCE, and confidence-guided mixing – could potentially be modified to handle multi-label classification tasks. However, modifications such as independently considering the label confidence for each label, adapting the mixing operation, and adjusting the confidence threshold would be necessary to account for the specific characteristics of multi-label noise. The extension of ConfidentMix to multi-label classification is an interesting direction for future work, as it could extend the applicability of our proposed method to a wider range of real-world problems in which instances are associated with multiple labels.

Next, while ConfidentMix has demonstrated effectiveness in various noisy label scenarios and has been specifically designed for integration with SSL techniques, there remain opportunities for future research in comparing and potentially combining our approach with other SOTA methods. For instance, RankMatch, a recently proposed method that shares the goal of improving robustness in learning with noisy labels, employs a confidence representation voting strategy for clean sample selection and a rank contrastive loss for consistency regularization. Although RankMatch and ConfidentMix have different strategies and are designed for different settings, exploring their compatibility and potential synergies could lead to further advancements in the field.

ACKNOWLEDGMENT

The authors would like to thank Edanz (https://jp.edanz.com/ac) for editing a draft of this manuscript.

REFERENCES

[1] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[2] S. Yang, P. Sun, Y. Jiang, X. Xia, R. Zhang, Z. Yuan, C. Wang, P. Luo, and M. Xu, "Objects in semantic topology," in *Proc. Int. Conf. Learn. Represent.*, 2022, pp. 1–11.

[3] W. Li, L. Wang, W. Li, E. Agustsson, and L. Van Gool, "WebVision database: Visual learning and understanding from web data," 2017, *arXiv:1708.02862*.

[4] H. Song, M. Kim, and J.-G. Lee, "SELFIE: Refurbishing unclean samples for robust deep learning," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 5907–5915.

[5] Y. Yan, R. Rosales, G. Fung, R. Subramanian, and J. Dy, "Learning from multiple annotators with varying expertise," *Mach. Learn.*, vol. 95, no. 3, pp. 291–327, Jun. 2014.

[6] X. Yu, T. Liu, M. Gong, and D. Tao, "Learning with biased complementary labels," in *Proc. Eur. Conf. Comput. Vis.*, Sep. 2018, pp. 68–83.

[7] C. Zhang, S. Bengio, M. Hardt, B. Recht, and O. Vinyals, "Understanding deep learning requires rethinking generalization," in *Proc. Int. Conf. Learn. Represent.*, 2017, pp. 1–12.

[8] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "mixup: Beyond empirical risk minimization," in *Proc. Int. Conf. Learn. Represent.*, 2018, pp. 1–9.

[9] S. Liu, J. Niles-Weed, N. Razavian, and C. Fernandez-Granda, "Early-learning regularization prevents memorization of noisy labels," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, vol. 33, 2020, pp. 20331–20342.

[10] Y. Tu, B. Zhang, Y. Li, L. Liu, J. Li, J. Zhang, Y. Wang, C. Wang, and C. R. Zhao, "Learning with noisy labels via self-supervised adversarial noisy masking," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 16186–16195.

[11] G. Zheng, A. H. Awadallah, and S. Dumais, "Meta label correction for noisy label learning," in *Proc. AAAI Conf. Artif. Intell.*, vol. 35, no. 12, 2021, pp. 11053–11061.

[12] Z. Huang, J. Zhang, and H. Shan, "Twin contrastive learning with noisy labels," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 11661–11670.

[13] Y. Tu, B. Zhang, Y. Li, L. Liu, J. Li, Y. Wang, C. Wang, and C. R. Zhao, "Learning from noisy labels with decoupled meta label purifier," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 19934–19943.

[14] B. Han, Q. Yao, X. Yu, G. Niu, M. Xu, W. Hu, I. Tsang, and M. Sugiyama, "Co-teaching: Robust training of deep neural networks with extremely noisy labels," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 31, 2018, pp. 1–7.

[15] X. Yu, B. Han, J. Yao, G. Niu, I. Tsang, and M. Sugiyama, "How does disagreement help generalization against label corruption?" in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 7164–7173.

[16] D. Arpit, "A closer look at memorization in deep networks," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 233–242.

[17] D. Berthelot, N. Carlini, I. Goodfellow, N. Papernot, A. Oliver, and C. A. Raffel, "MixMatch: A holistic approach to semi-supervised learning," in *Proc. 33rd Int. Conf. Neural Inf. Process. Syst.*, 2019, pp. 5049–5059, Art. no. 454.

[18] K. Sohn, "FixMatch: Simplifying semi-supervised learning with consistency and confidence," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2020, pp. 596–608.

[19] J. Li, R. Socher, and S. C. Hoi, "DivideMix: Learning with noisy labels as semi-supervised learning," in *Proc. Int. Conf. Learn. Represent.*, 2020, pp. 1–6.

[20] D. Ortego, E. Arazo, P. Albert, N. E. O'Connor, and K. McGuinness, "Multi-objective interpolation training for robustness to label noise," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 6602–6611.

[21] F. R. Cordeiro, R. Sachdeva, V. Belagiannis, I. Reid, and G. Carneiro, "LongReMix: Robust learning with high confidence samples in a noisy label environment," *Pattern Recognit.*, vol. 133, Jan. 2023, Art. no. 109013.

[22] C. Feng, Y. Ren, and X. Xie, "OT-filter: An optimal transport filter for learning with noisy labels," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 16164–16174.

[23] E. D. Cubuk, B. Zoph, D. Mané, V. Vasudevan, and Q. V. Le, "AutoAugment: Learning augmentation strategies from data," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 113–123.

[24] E. D. Cubuk, B. Zoph, J. Shlens, and Q. V. Le, "Randaugment: Practical automated data augmentation with a reduced search space," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2020, pp. 3008–3017.

[25] S. Yun, D. Han, S. Chun, S. J. Oh, Y. Yoo, and J. Choe, "CutMix: Regularization strategy to train strong classifiers with localizable features," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 6022–6031.

[26] Y. Chen, M. Mancini, X. Zhu, and Z. Akata, "Semi-supervised and unsupervised deep visual learning: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, no. 3, pp. 1327–1347, Mar. 2024.

[27] B. Zhang, "FlexMatch: Boosting semi-supervised learning with curriculum pseudo labeling," in *Proc. NIPS*, 2021, pp. 18408–18419.

[28] J. Li, C. Xiong, and S. C. H. Hoi, "CoMatch: Semi-supervised learning with contrastive graph regularization," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 9455–9464.

[29] M. Zheng, S. You, L. Huang, F. Wang, C. Qian, and C. Xu, "SimMatch: Semi-supervised learning with similarity matching," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 14451–14461.

[30] S. Laine and T. Aila, "Temporal ensembling for semi-supervised learning," in *Proc. Int. Conf. Learn. Represent.*, 2017, pp. 1–5.

[31] L. Jiang, Z. Zhou, T. Leung, L.-J. Li, and L. Fei-Fei, "MentorNet: Learning data-driven curriculum for very deep neural networks on corrupted labels," in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 2304–2313.

[32] E. Malach and S. Shalev-Shwartz, "Decoupling 'when to update' from 'how to update,"' in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 961–971.

[33] Y. Ding, L. Wang, D. Fan, and B. Gong, "A semi-supervised two-stage approach to learning from noisy labels," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2018, pp. 1215–1224.

[34] K. Kong, J. Lee, Y. Kwak, M. Kang, S. G. Kim, and W.-J. Song, "Recycling: Semi-supervised learning with noisy labels in deep neural networks," *IEEE Access*, vol. 7, pp. 66998–67005, 2019.

[35] T. Kim, "FINE samples for learning with noisy labels," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, vol. 34, 2021, pp. 24137–24149.

[36] D. Tanaka, D. Ikami, T. Yamasaki, and K. Aizawa, "Joint optimization framework for learning with noisy labels," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 5552–5560.

[37] E. Arazo, D. Ortego, P. Albert, N. E. O'Connor, and K. McGuinness, "Unsupervised label noise modeling and loss correction," in *Proc. Int. Conf. Learn. Represent.*, 2019, pp. 1–12.

[38] Z. Zhang, W. Chen, C. Fang, Z. Li, L. Chen, L. Lin, and G. Li, "RankMatch: Fostering confidence and consistency in learning with noisy labels," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2023, pp. 1644–1654.

[39] G. Liu, Y. Mao, H. Hailong, G. Weiguo, and L. Xuan, "Adversarial mixing policy for relaxing locally linear constraints in mixup," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2021, pp. 2998–3008.

[40] Z. Mai, G. Hu, D. Chen, F. Shen, and H. T. Shen, "MetaMixUp: Learning adaptive interpolation policy of MixUp with metalearning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 7, pp. 3050–3064, Jul. 2022.

[41] B. van Rooyen, A. Menon, and R. C. Williamson, "Learning with symmetric label noise: The importance of being unhinged," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 28, 2015, pp. 1–31.

[42] C. Scott, G. Blanchard, and G. Handy, "Classification with asymmetric label noise: Consistency and maximal denoising," in *Proc. Annu. Conf. Learn. Theory*, vol. 30, 2013, pp. 489–511.

[43] A. K. Menon, B. van Rooyen, and N. Natarajan, "Learning from binary labels with instance-dependent corruption," 2016, *arXiv:1605.00751*.

[44] A. Garg, C. Nguyen, R. Felix, T.-T. Do, and G. Carneiro, "Instance-dependent noisy label learning via graphical modelling," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2023, pp. 2287–2297.

[45] H. Song, M. Kim, D. Park, Y. Shin, and J.-G. Lee, "Learning from noisy labels with deep neural networks: A survey," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 11, pp. 8135–8153, Aug. 2023.

[46] A. Krizhevsky and G. Hinton, "Learning multiple layers of features from tiny images," M.S. thesis, Dept. Comput. Sci., Univ. Toronto, Toronto, ON, Canada, 2009. [Online]. Available: https://www.cs.toronto.edu/~kriz/learning-features-2009-TR.pdf

[47] T. Xiao, T. Xia, Y. Yang, C. Huang, and X. Wang, "Learning from massive noisy labeled data for image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 2691–2699.

[48] H. Song, M. Kim, D. Park, and J.-G. Lee, "How does early stopping help generalization against label noise?" 2019, *arXiv:1911.08059*.

[49] P. Chen, J. Ye, G. Chen, J. Zhao, and P.-A. Heng, "Beyond class-conditional assumption: A primary attempt to combat instance-dependent label noise," in *Proc. AAAI Conf. Artif. Intell.*, vol. 35, no. 13, 2021, pp. 11442–11450.

[50] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.

[51] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 630–645.

[52] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, Inception-ResNet and the impact of residual connections on learning," in *Proc. AAAI Conf. Artif. Intell.*, vol. 31, no. 1, 2017, pp. 1–7.

RYOTA HIGASHIMOTO received the B.S. and M.S. degrees in engineering from Kansai University, Japan, in 2022 and 2024, respectively. He joined Hitachi, Ltd., in 2024. His research interest includes deep learning technologies.

SOH YOSHIDA (Member, IEEE) received the B.S., M.S., and Ph.D. degrees in engineering and information science from Hokkaido University, Japan, in 2012, 2014, and 2016, respectively. He joined the Faculty of Engineering, Kansai University, in 2016. He is currently an Associate Professor. His research interests include artificial intelligence technologies related to multimedia processing and information retrieval. He is a member of ACM, IEICE, and ITE.

MITSUJI MUNEYASU (Member, IEEE) received the B.S. and M.S. degrees in system engineering from Kobe University, Japan, in 1982 and 1984, respectively, and the Ph.D. degree in engineering from Hiroshima University, Japan, in 1993. He joined Oki Electric Industry Company Ltd., Tokyo, Japan, in 1984. From 1990 to 1991, he was a Research Assistant with the Faculty of Engineering, Tottori University, Tottori, Japan. From 1991 to 2001, he was a Research Assistant and an Associate Professor with the Faculty of Engineering, Hiroshima University, Higashihiroshima, Japan. He joined the Faculty of Engineering, Kansai University, Osaka, Japan, in 2001. He is currently a Professor. His research interests include image processing theory and nonlinear digital signal processing. He is a fellow of IEICE and a member of IPSJ.