ISSN 2464-4617 (print)
ISSN 2464-4625 (online)
Computer Science Research Notes - CSRN
http://www.wscg.eu
WSCG 2025 Proceedings

Detect and Correct: A Selective Noise Correction Method for Learning with Noisy Labels

Yuval Grinberg
Bar Ilan University,
Immunai Inc.
juvs2k@gmail.com

Nimrod Harel
Tel Aviv University
nimrodharel@mail.tau.ac.il

Jacob Goldberger
Bar Ilan University
jacob.goldberger@biu.ac.il

Ofir Lindenbaum
Bar Ilan University
ofirlin@gmail.com

ABSTRACT

Falsely annotated samples, also known as noisy labels, can significantly harm the performance of deep learning models. Two main approaches for learning with noisy labels are global noise estimation and data filtering. Global noise estimation approximates the noise across the entire dataset using a noise transition matrix, but it can unnecessarily adjust correct labels, leaving room for local improvements. Data filtering, on the other hand, discards potentially noisy samples but risks losing valuable data. Our method identifies potentially noisy samples based on their loss distribution. We then apply a selection process to separate noisy and clean samples and learn a noise transition matrix to correct the loss for noisy samples while leaving the clean data unaffected, thereby improving the training process. Our approach ensures robust learning and enhanced model performance by preserving valuable information from noisy samples and refining the correction process. We applied our method to standard image datasets (MNIST, CIFAR-10, and CIFAR-100) and a biological scRNA-seq cell-type annotation dataset. We observed a significant improvement in model accuracy and robustness compared to traditional methods.

Keywords

Noisy Labels, Transition Matrix, Image Classification, Single-cell RNA Sequencing, Cell Annotation, Label Noise Correction, Selective Correction

1 INTRODUCTION

The performance of deep learning models is heavily influenced by the quality of their training data [30]. Noisy labels- incorrect or misleading- pose a significant challenge, as they can substantially degrade these models' accuracy and generalization capabilities. This issue is pervasive across various applications, including image classification [1] [31], natural language processing [34] [33], and speech recognition [18] [27], where training data often contains annotation errors or other noise sources.

Numerous methods have been developed to address the challenge of noisy labels. Traditional approaches include loss correction techniques that adjust the loss function to account for label noise, often relying on a noise transition matrix [19]. This matrix models the probability of label noise and adjusts the learning process accordingly. However, accurately estimating the transition matrix is not only challenging [21] but also introduces an inherent trade-off. Even with a perfectly estimated matrix, both correct and incorrect labels are affected, as the adjustments made to account for label noise can inadvertently impact clean samples, leading to performance issues.

Another avenue of research has focused on developing methods to dynamically detect and filter noisy labels. Techniques such as Co-teaching [7] and MentorNet [13], which involve training multiple networks simultaneously to filter out noisy examples based on their disagreement, have shown promise. Similarly, the O2U-Net [12], introduced in 2019, analyzes the loss behavior of each sample during the training process to identify noisy labels, demonstrating significant improvements in detecting noisy data. More recent advancements have focused on refining and integrating these techniques with sophisticated models to further enhance their effectiveness in filtering noisy labels.

All loss correction strategies based on the noise transition matrix offer a global representation of the label noise distribution [35]. While these methods capture general noise patterns, they often miss outliers due to an inherent trade-off between correcting mislabeled data and preserving true labels. Conversely, noise detection approaches that discard potentially noisy samples risk losing valuable training data [32], including both correctly labeled examples and those for which the loss could have been corrected.

Building on these advancements, we propose D&C-Net (Detect and Correct), a hybrid approach combining noise detection with targeted correction. Our method aims to identify potentially noisy samples and uses them to estimate the transition matrix more accurately. This matrix is then applied specifically to these noisy samples, enhancing the training process without altering clean data.

A notable approach for handling noisy labels is DivideMix [17], which uses a Gaussian Mixture Model (GMM) to separate clean and noisy samples, treating the latter as unlabeled data for semi-supervised learning. While effective, this approach discards mislabeled samples rather than correcting them. In contrast, our method integrates noise detection with targeted correction, using identified noisy samples to estimate a transition matrix that is selectively applied. This preserves more useful training signals while mitigating label noise, striking a better balance between robustness and data efficiency.

Our contribution goes beyond simply ensuring robust learning from clean data; it incorporates both noise detection and targeted correction strategies to mitigate the impact of noisy labels. Additionally, we introduce a loss distribution analysis using a GMM [24] to distinguish between noisy and clean labels more effectively. We evaluate our method on several benchmark datasets and compare its performance against state-of-the-art approaches, demonstrating its effectiveness in enhancing model performance under noisy conditions.

2 PROBLEM FORMULATION

Consider a dataset $\mathcal{D} = \{(x_n, y_n)\}_{n=1}^N$, where $x_n \in \mathbb{R}^d$ represents the feature vector, and $y_n \in \{1, 2, \ldots, C\}$ is the true (clean) label. However, in the presence of label noise, the observed labels $\tilde{y}_n$ (noisy labels) may differ from the true labels y_n. We denote Y and $\tilde{\text{Y}}$ as the random variables representing the true and noisy labels, respectively. Consequently, we define the noisy dataset as $\tilde{\mathcal{D}} = \{(x_n, \tilde{y}_n)\}_{n=1}^N$. This noise process can be modeled using an unknown noise transition matrix $T \in [0,1]^{C \times C}$, where each element $T_{k,l}$ represents the probability that the true label $\text{Y} = l$ is flipped to the noisy label $\tilde{\text{Y}} = k$:

$$T_{l,k} = \Pr(\tilde{\text{Y}} = k \mid \text{Y} = l), \quad \forall l, k \in \{1, 2, \ldots, C\}.$$

By definition, T is a row-stochastic matrix, meaning that each row sums to one:

$$\sum_{k=1}^{C} T_{l,k} = 1, \quad \forall l \in \{1, 2, \ldots, C\}.$$

Given a noisy dataset $\tilde{\mathcal{D}} = \{(x_n, \tilde{y}_n)\}_{n=1}^N$, our objective is to train a classifier $f_\theta(x)$ (a neural network with parameters θ) that can accurately predict the true labels despite the presence of label noise and the unknown transition matrix.

3 METHOD

Our approach (D&C-Net) separates the training process into two distinct stages, pre-training and training, to effectively tackle the challenge of noisy labels. The pre-training stage is a local process focused on detecting noisy labels within the dataset. This stage operates at a granular level, examining individual samples by analyzing their loss trajectories during training [38]. By identifying and removing these noisy labels, we set the stage for a more precise and effective learning process.

The training stage shifts to a global perspective, where the information from the pre-training stage is used to improve the overall model training. Here, we incorporate loss corrections based on the noisy labels identified earlier. This correction is applied selectively, ensuring that the model learns more accurately from the clean data while adjusting its learning for noisy samples. This two-stage strategy allows us to maintain the integrity of the training process, leading to a more robust and reliable model that can generalize well despite the presence of noise in the labels. An illustration of the entire method can be found in Figure 1.

3.1 Pre-Training

While various methods could be employed to differentiate noisy from clean labels, our pre-training procedure, inspired by Huang et al. [12], focuses on tracking the loss values of each sample across epochs to identify noise-related patterns. This is done by applying a cyclic learning rate during training [29] and computing the loss values ℓ for each sample x. The cyclic learning rate helps escape sharp local minima and allows better separation of clean and noisy samples by amplifying the differences in their loss trajectories.

Then, to estimate which labels are noisy without prior knowledge about the contamination level, we introduce a **Loss Distribution Analysis**. Additionally, we implement **Transition Matrix Initialization** to optimize the subsequent training stage. These components work together to manage noisy labels and improve the overall model performance.

Loss Distribution Analysis: Loss values, denoted as ℓ, correspond to the per-sample negative log-likelihood (i.e., the cross-entropy loss), where each ℓ_n is defined by

$$\ell_n = L(f_\theta(x_n), \tilde{y}_n).$$

And the empirical probability density function (PDF) of these loss values observed during pre-training is denoted as $g(\ell)$. Specifically, we assume $g(\ell)$ follows a mixture of two Gaussian distributions [24, 28]:

$$g(\ell) = \lambda \cdot \mathcal{N}(\ell; \mu_1, \sigma_1^2) + (1 - \lambda) \cdot \mathcal{N}(\ell; \mu_2, \sigma_2^2),$$

where $\mathcal{N}(\ell; \mu_1, \sigma_1^2)$ and $\mathcal{N}(\ell; \mu_2, \sigma_2^2)$ are the probability density functions of the clean and contaminated (i.e.,

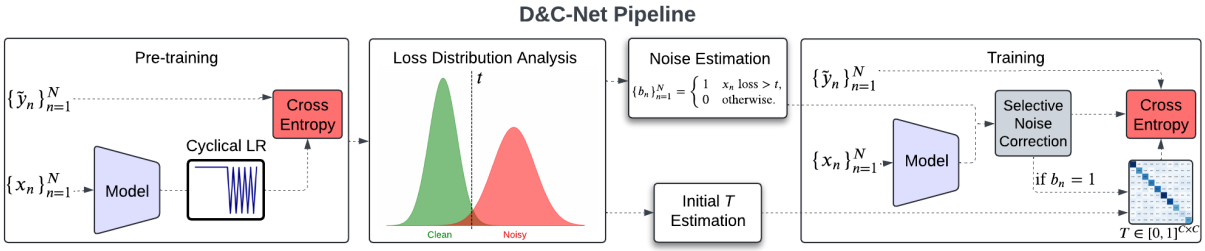

Figure 1: This figure illustrates the D&C-Net method for handling noisy labels. Initially, in the pre-training phase, a cyclical learning rate is employed to estimate two Gaussian distributions: one representing the clean loss distribution and the other representing the noisy loss distribution. Potential noisy samples are identified where the probability of being drawn from the noisy distribution exceeds a threshold t, determined by the intersection of the two distributions. Using these samples, the initial transition matrix for label correction $T \in [0,1]^{C\times C}$ is estimated. Noisy labels are flagged by a binary vector b , where $b_n = 1$ denotes a potential noisy label. During the training phase, we selectively apply T to the noisy samples, while clean samples are trained with the standard cross-entropy loss.

noisy) label loss distributions, respectively. Here, μ_1 and σ_1^2 represent the mean and variance of the clean label loss distribution, while μ_2 and σ_2^2 correspond to those of the contaminated label loss distribution. Finally, λ and $1-\lambda$ are the mixing proportions of the two distributions, where $\lambda \in (0,1)$.

All parameters, including the mixing proportion λ, as well as the mean and variance of both Gaussian components $(\mu_1, \sigma_1^2, \mu_2, \sigma_2^2)$, are estimated using Maximum Likelihood via the Expectation-Maximization (EM) algorithm [5]. Once fitted, these distributions can be compared with the empirical distribution to estimate which labels are contaminated (i.e., noisy).

We define a threshold t on the loss values ℓ, such that a sample x with a loss $\ell > t$ is classified as contaminated (i.e., noisy), while a sample with $\ell \leq t$ is considered clean. This threshold is used to separate samples that are likely mislabeled from those that are likely correctly labeled. Then, the estimated true positive (TP), estimated false negative (FN), estimated false positive (FP), and estimated true negative (TN) rates can be expressed as shown in Table 1, where t is the threshold we want to learn. The algorithm is described in Algorithm 1.

Table 1: Confusion Matrix with Gaussian Distributions

TP	**FP**
$\int_{-\infty}^{t} \lambda \cdot \mathcal{N}(x;\mu_1,\sigma_1^2)\,dx$	$\int_{-\infty}^{t} (1-\lambda) \cdot \mathcal{N}(x;\mu_2,\sigma_2^2)\,dx$
FN	**TN**
$\int_{t}^{\infty} \lambda \cdot \mathcal{N}(x;\mu_1,\sigma_1^2)\,dx$	$\int_{t}^{\infty} (1-\lambda) \cdot \mathcal{N}(x;\mu_2,\sigma_2^2)\,dx$

We then seek to find the optimal threshold t that maximizes the balance between sensitivity (SEN) and specificity (SPC) [25]:

$$\max_{-\infty<t<\infty} \left(\frac{\mathrm{SEN}(t) + \mathrm{SPC}(t)}{2} \right),$$

where sensitivity and specificity are defined as:

$$\mathrm{SEN}(t) = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}, \quad \mathrm{SPC}(t) = \frac{\mathrm{TN}}{\mathrm{TN}+\mathrm{FP}}.$$

This criterion ensures that the selected threshold minimizes the probability of misclassification by balancing the rates of true positives and true negatives, as shown in Figure 2. By optimizing the average of sensitivity and specificity, we ensure that the threshold t is neither too strict nor too lenient, leading to a well-balanced separation between clean and contaminated labels. Unlike methods that require prior knowledge of the noise rate or assume a fixed proportion of outliers, our approach estimates the contamination dynamically from the observed loss distribution. This makes our method more practical in real-world settings, where the actual noise level is often unknown.

Finally, using this optimized threshold, we create a binary vector b, indicating each sample label as either noisy or clean. The binary vector b is defined as follows:

$$b = [b_1, b_2, \ldots, b_n],$$

$$\text{where } b_n = \begin{cases} 1 & \text{if loss } \ell_n \text{ exceeds threshold } t \text{ (noisy)}, \\ 0 & \text{otherwise.} \end{cases}$$

Transition Matrix Initialization: Using all samples x_n such that $b_n = 1$, we estimate an initial transition matrix T by counting occurrences where the observed noisy label $\tilde{y}_n = k$ differs from the classifier's prediction $\hat{y}_n = l$, while the sample has been identified as noisy by the previous Loss Distribution Analysis:

$$T_{l,k} = \sum_{n=1}^{N} I(\tilde{y}_n = k,\ \hat{y}_n = l,\ b_n = 1), \quad \forall l,k \in \{1,2,\ldots,C\}$$

where $I(\cdot)$ is an indicator function that equals 1 when its condition holds and 0 otherwise. Each row is then normalized to sum to 1. After normalization, $T_{l,k}$ represents the empirical frequency of the classifier predicting

label k given the label l, only for suspected noisy labels. Since the initialization depends on classifier predictions, we assume the classifier is sufficiently accurate to provide a meaningful estimate of the noise structure.

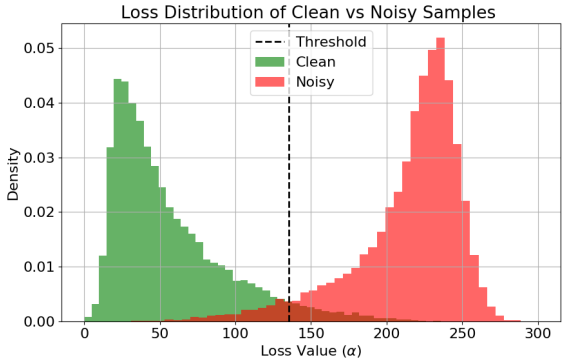

Figure 2: This figure illustrates the separation of clean and noisy examples based on their loss distributions under 50% symmetric noise. It demonstrates how our method differentiates between clean and noisy labels through loss analysis, without any user input.

3.2 Model Training

In this stage, we focus on enhancing the model's accuracy by addressing the challenges posed by noisy labels. As shown in Figure 1, we selectively apply corrections to the identified noisy data, ensuring that the model learns from the clean data without being misled by errors. This targeted approach helps the model to develop a more accurate and reliable understanding of the data, ultimately improving its overall performance.

Selective Noise Correction: During training, we modify the loss computation only for noisy samples while keeping the standard cross-entropy loss for clean samples.

For noisy samples ($b_n = 1$), we first adjust the model predictions by incorporating the transition matrix. The cross-entropy loss is then given by:

$$L(\theta; x_n, \tilde{y}_n, b_n) = \begin{cases} -\log P_\theta(\mathrm{Y} = \tilde{y}_n \mid x_n), & \text{if } b_n = 0, \\ -\log \sum_{l=1}^{C} T_{l,\tilde{y}_n} \cdot P_\theta(\mathrm{Y} = l \mid x_n), & \text{if } b_n = 1. \end{cases}$$

Here, $T_{l,\tilde{y}_n}$ represents the transition probability from the true label l to the observed noisy label, and θ represents the model parameters. The transition matrix T is continuously updated during training to adapt to evolving noise patterns in the dataset. By applying corrections only to noisy samples, we preserve the learning dynamics for clean data while improving robustness against label noise.

This hybrid approach combines noise detection with targeted correction, effectively leveraging noisy samples to improve model performance and enhance generalization.

Our approach builds on existing techniques but introduces key innovations for learning with noisy labels. Specifically, we use a **cyclical learning rate** [29] to track loss trajectories and detect noisy samples, whereas prior works used it mainly for optimization. We apply a GMM to separate noisy and clean samples, similar to DivideMix [17], but uniquely integrate it with cyclical learning for improved noise identification. Furthermore, we propose a novel **transition matrix initialization** based on this pre-training stage, which has not been done before. Finally, our **Selective Noise Correction** mechanism ensures that the transition matrix is applied only to noisy samples, avoiding unnecessary corrections to clean data- an aspect overlooked in prior noise correction methods. The algorithm is described in Algorithm 2.

Algorithm 1 Pre-Training for Noise Detection and Transition Matrix Initialization

Require: Dataset $\tilde{\mathcal{D}} = \{(x_n, \tilde{y}_n)\}_{n=1}^{N}$, model f_θ
Ensure: Updated model parameters θ, Sample noise indicator vector b, transition matrix T
1: **Training:** Train f_θ on $\tilde{\mathcal{D}}$ using a standard procedure with a scheduled cyclic learning rate, storing per-sample losses
2: **Noise Detection:** Fit a GMM to the loss distribution $g(\ell)$
3: Estimate threshold t and identify noisy samples: $b_n \leftarrow 1_{\ell_n > t}, n = 1, \ldots, N$
4: **Transition Matrix Initialization:** Initialize T using b and normalize row-wise
5: **return** θ, b, T

Algorithm 2 Detect & Correct

Require: Dataset $\tilde{\mathcal{D}} = \{(x_n, \tilde{y}_n)\}_{n=1}^{N}$, Model f_θ, transition matrix T, noisy sample indicator b
Ensure: Updated model parameters θ, corrected transition matrix T
1: **for** epoch in N_{epochs} **do**
2: **for** each $(x_n, \tilde{y}_n)$ in $\tilde{\mathcal{D}}$ **do**
3: Forward pass $f_\theta(x_n)$
4: Compute $L(\theta; x_n, \tilde{y}_n, b_n)$
5: Update transition matrix T
6: Update model parameters θ
7: **end for**
8: **end for**
9: **return** θ

4 EXPERIMENTS

In this section, we evaluate the effectiveness of our proposed method using various datasets. We report the performance of our method in comparison to other state-of-the-art methods. The code to reproduce our experiments is available at: `https://github.com/yuvalgrin/DetectAndCorrect-Net`.

Table 2: Results on MNIST, CIFAR-10, CIFAR-100 for Pair-20% noise rate.

	MNIST	**CIFAR-10**	**CIFAR-100**
	Pair-20%	Pair-20%	Pair-20%
Decoupling	96.93 ± 0.07	77.12 ± 0.30	40.12 ± 0.26
MentorNet	96.89 ± 0.04	77.42 ± 0.00	39.22 ± 0.47
Co-teaching	97.00 ± 0.06	80.65 ± 0.20	42.79 ± 0.79
Forward	98.84 ± 0.10	88.21 ± 0.48	56.12 ± 0.54
T-Revision	98.89 ± 0.08	90.33 ± 0.52	64.33 ± 0.49
DMI	98.84 ± 0.09	89.89 ± 0.45	59.56 ± 0.73
Dual T	98.86 ± 0.04	89.77 ± 0.25	67.21 ± 0.43
VolMinNet	99.01 ± 0.07	90.37 ± 0.30	**68.45 ± 0.69**
D&C-Net	**99.10 ± 0.09**	**91.86 ± 0.02**	65.89 ± 0.48

Table 3: Results on MNIST, CIFAR-10, CIFAR-100 for Sym-20% noise rate.

	MNIST	**CIFAR-10**	**CIFAR-100**
	Sym-20%	Sym-20%	Sym-20%
Decoupling	97.04 ± 0.06	77.32 ± 0.35	41.92 ± 0.49
MentorNet	97.21 ± 0.06	81.35 ± 0.23	42.88 ± 0.41
Co-teaching	97.07 ± 0.10	82.27 ± 0.07	48.48 ± 0.66
Forward	98.60 ± 0.19	85.20 ± 0.80	54.90 ± 0.74
T-Revision	98.72 ± 0.10	87.95 ± 0.36	62.72 ± 0.69
DMI	98.70 ± 0.02	87.54 ± 0.20	62.65 ± 0.39
Dual T	98.43 ± 0.05	88.35 ± 0.33	62.16 ± 0.58
VolMinNet	98.74 ± 0.08	89.58 ± 0.26	64.94 ± 0.40
D&C-Net	**98.95 ± 0.03**	**90.61 ± 0.67**	**69.28 ± 0.08**

Table 4: Results on MNIST, CIFAR-10, CIFAR-100 for Sym-50% noise rate.

	MNIST	**CIFAR-10**	**CIFAR-100**
	Sym-50%	Sym-50%	Sym-50%
Decoupling	94.58 ± 0.08	54.07 ± 0.46	22.63 ± 0.44
MentorNet	95.56 ± 0.15	73.47 ± 0.15	32.66 ± 0.40
Co-teaching	95.20 ± 0.23	75.55 ± 0.07	36.77 ± 0.52
Forward	97.77 ± 0.16	74.82 ± 0.78	41.85 ± 0.71
T-Revision	98.23 ± 0.10	80.01 ± 0.62	49.12 ± 0.22
DMI	98.12 ± 0.21	82.68 ± 0.22	52.42 ± 0.64
Dual T	98.15 ± 0.12	82.54 ± 0.19	52.49 ± 0.37
VolMinNet	98.23 ± 0.16	83.37 ± 0.25	53.89 ± 1.26
D&C-Net	**98.42 ± 0.03**	**87.30 ± 0.07**	**60.39 ± 0.08**

Table 5: Results on scRNA-seq PBMC cell-type annotation for Pair-20%, Sym-20%, and Sym-50% noise rates.

	scRNA-seq PBMC cell-type annotation		
	Pair-20%	Sym-20%	Sym-50%
VolMinNet	90.67 ± 0.77	90.85 ± 1.11	93.75 ± 0.08
D&C-Net	**93.83 ± 0.40**	**94.42 ± 0.29**	**93.93 ± 0.05**

4.1 Datasets

We performed experiments on the following datasets: MNIST [15], CIFAR-10 [14], CIFAR-100 [14] and an scRNA-seq PBMC dataset [23] [11]: A dataset consisting of scRNA-seq cell count matrix data with 161,764 cells and 31 cell types.

4.2 Experimental Setup

For each dataset, we conducted experiments under varying levels of label noise to evaluate the robustness of our method. Specifically, we introduced 20% and 50% symmetric label noise (where labels are uniformly flipped to any other class) as defined in [2], as well as 20% pair label noise (where labels are flipped to a specific incorrect class) as defined in [7]. For both CIFAR-10 and CIFAR-100, we used 200 epochs for pre-training and 80 epochs for final training, while CIFAR-10 used ResNet-18 [9] and CIFAR-100 used ResNet-34 [9]. For both MNIST and scRNA-seq PBMC cell-type annotation, we used 50 epochs for pre-training and 40 epochs for final training, while MNIST used LeNet and cell-type annotation used a three-layer fully connected model. We used a learning rate of 10^{-2} for the model and 10^{-4} for the transition matrix for all datasets and models.

4.3 Results

The results, summarized in Tables 2, 3, and 4, demonstrate that our method consistently outperforms leading baselines across various noise levels. Specifically, our approach achieves higher classification accuracy, particularly in high-noise scenarios, underscoring its robustness and effectiveness in handling noisy labels. We compared our method with several established approaches, including Decoupling [22], T-Reivision [8], MentorNet [13], Co-teaching [7], Forward [26], DMI [36], Dual T [37], and VolMinNet [20].

4.4 Single Cell RNA Sequencing Cell-Type Annotation

Single-cell RNA sequencing (scRNA-seq) enables the examination of gene expression at the single-cell level, providing insights into cellular heterogeneity. Cell-type annotation, achieved by comparing gene expression profiles to known markers or reference datasets, is crucial for uncovering this heterogeneity. However, the low signal-to-noise ratio in scRNA-seq data, along with batch effects and dropout, makes label noise a significant challenge [39], obscuring true cell-type distinctions and complicating accurate classification by learning algorithms. The results are summarized in Table 5, where our approach consistently achieves higher classification accuracy across all noise settings.

Data preparation: We utilized the Azimuth PBMC reference dataset [23]. We then used totalVI [6] to integrate the cells into a 20-dimensional latent space to

remove batch effects and facilitate the classifier training.

4.5 Ablation

We perform a detailed ablation study on D&C-Net by first evaluating it without the selective noise correction and pre-training components on all datasets, which results in a notable drop in performance, with accuracy decreasing by an absolute $\approx \mathbf{1}\% - \mathbf{6}\%$ with an average of $\approx \mathbf{2}\%$. We also examine the impact of not initializing the transition matrix T, which leads to a smaller absolute decrease in the accuracy of $\approx \mathbf{0.5}\%$ on the CIFAR-10 dataset. These results underscore the importance of both noise management and initialization strategies in achieving optimal performance.

5 DISCUSSION

D&C-Net integrates noisy label identification with selective loss correction, selectively applying the transition matrix to suspected noisy samples. Our experiments demonstrate that D&C-Net consistently improves classification accuracy across diverse datasets, including MNIST, CIFAR-10, CIFAR-100, and scRNA-seq PBMC annotation. By preserving clean data integrity and applying targeted noise correction, D&C-Net exhibits robustness, especially under high-noise conditions.

5.1 Comparisons to Existing Methods

Existing methods for learning with noisy labels generally fall into two categories: (1) *loss correction techniques* that modify the loss function to account for noise, often via a noise transition matrix, and (2) *data filtering approaches* that detect and remove noisy labels. D&C-Net balances these two strategies by identifying noisy labels while preserving valuable training information through selective correction. Unlike DivideMix [17], which treats all detected noisy labels as unlabeled data for semi-supervised learning, D&C-Net refines learning by correcting mislabels rather than discarding them. This distinction ensures that our method better retains useful information from noisy data while mitigating the risk of confirmation bias introduced by mislabeled samples.

5.2 Limitations and Future Directions

One limitation of D&C-Net is its reliance on accurate noisy label detection for transition matrix initialization. Errors in this step could lead to suboptimal noise correction, particularly in datasets with highly structured noise patterns or label-dependent noise. Future work could explore adaptive noise estimation techniques, such as iterative refinement of the transition matrix or integrating uncertainty estimation to better calibrate corrections.

Additionally, D&C-Net could benefit from complementary techniques not explored in this paper, such as:

- **Data Augmentation for Robustness:** Leveraging augmentation techniques, such as MixUp [40] or RandAugment [4], could improve model generalization by reducing reliance on individual noisy samples.
- **Pseudo-Labeling for Semi-Supervised Learning:** Incorporating pseudo-labeling strategies [16] might enable D&C-Net to refine its corrections by iteratively improving label quality.
- **Self-Supervised Pretraining:** Pretraining with self-supervised learning techniques, such as contrastive learning [3] or masked autoencoders [10], could provide a stronger feature representation before applying D&C-Net corrections.
- **Application to More Complex Noisy Scenarios:** Evaluating D&C-Net on real-world datasets with asymmetric or instance-dependent noise, such as web-scraped data or medical annotations, could further validate its effectiveness.

5.3 Impact on Biological Data Analysis

Our results on scRNA-seq PBMC annotation highlight the applicability of D&C-Net beyond traditional image datasets. Label noise in single-cell data often arises from ambiguous cell-type boundaries and batch effects, making noise correction crucial for accurate annotation. Expanding D&C-Net to integrate domain-specific biological priors, such as gene expression similarity metrics, could improve its ability to handle noise in biological datasets.

5.4 Conclusion

D&C-Net presents a novel approach to learning with noisy labels by integrating noise detection with targeted loss correction. By applying the transition matrix selectively, it achieves a strong balance between robustness and data efficiency. Future work should focus on refining noise estimation, incorporating semi-supervised learning strategies, and extending the method to more diverse and complex datasets.

ACKNOWLEDGMENT

We thank Immunai Inc. for inspiring this work, particularly in addressing noisy labels in cell-type classification using scRNA-seq datasets. We also thank Amitay Sicherman for his support and fruitful discussion.

REFERENCES

[1] Gorkem Algan and Ilkay Ulusoy. "Image classification with deep learning in the presence of noisy labels: A survey". In: *Knowledge-Based Systems* 215 (2021), p. 106771.

[2] Dana Angluin and Philip Laird. "Learning from noisy examples". In: *Machine Learning* 2 (1988), pp. 343–370.

[3] Ting Chen et al. *A Simple Framework for Contrastive Learning of Visual Representations*. 2020. arXiv: 2002.05709 [cs.LG]. URL: https://arxiv.org/abs/2002.05709.

[4] Ekin D. Cubuk et al. *RandAugment: Practical automated data augmentation with a reduced search space*. 2019. arXiv: 1909.13719 [cs.CV]. URL: https://arxiv.org/abs/1909.13719.

[5] Arthur P Dempster, Nan M Laird, and Donald B Rubin. "Maximum likelihood from incomplete data via the EM algorithm". In: *Journal of the royal statistical society: series B (methodological)* 39.1 (1977), pp. 1–22.

[6] Adam Gayoso, Zoë Steier, Romain Lopez, et al. "Joint probabilistic modeling of single-cell multi-omic data with totalVI". In: *Nature Methods* 18 (2021), pp. 272–282.

[7] Bo Han, Quanming Yao, et al. "Co-teaching: Robust training of deep neural networks with extremely noisy labels". In: *Advances in Neural Information Processing Systems (NeurIPS)*. 2018.

[8] Jiangfan Han, Ping Luo, and Xiaogang Wang. "Deep self-learning from noisy labels". In: *Proceedings of the IEEE/CVF international conference on computer vision*. 2019, pp. 5138–5147.

[9] Kaiming He et al. "Deep residual learning for image recognition". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 770–778.

[10] Kaiming He et al. *Masked Autoencoders Are Scalable Vision Learners*. 2021. arXiv: 2111.06377 [cs.CV]. URL: https://arxiv.org/abs/2111.06377.

[11] Wenpin Hou et al. "A systematic evaluation of single-cell RNA-sequencing imputation methods". In: *Genome biology* 21 (2020), pp. 1–30.

[12] Junjie Huang et al. "O2U-Net: A Simple Noisy Label Detection Approach for Deep Neural Networks". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. 2019, pp. 3326–3334.

[13] Lu Jiang et al. "MentorNet: Learning data-driven curriculum for very deep neural networks on corrupted labels". In: *Proceedings of the International Conference on Machine Learning (ICML)*. 2018, pp. 2304–2313.

[14] Alex Krizhevsky. "Learning Multiple Layers of Features from Tiny Images". In: *University of Toronto* (2009), pp. 32–33.

[15] Yann LeCun, Léon Bottou, et al. "Gradient-based learning applied to document recognition". In: *Proceedings of the IEEE* 86.11 (1998), pp. 2278–2324.

[16] Dong-Hyun Lee. "Pseudo-Label : The Simple and Efficient Semi-Supervised Learning Method for Deep Neural Networks". In: 2013. URL: https://api.semanticscholar.org/CorpusID:18507866.

[17] Junnan Li, Richard Socher, and Steven C. H. Hoi. *DivideMix: Learning with Noisy Labels as Semi-supervised Learning*. 2020. arXiv: 2002.07394 [cs.CV].

[18] Lin Li, Fuchuan Tong, and Qingyang Hong. "When Speaker Recognition Meets Noisy Labels: Optimizations for Front-Ends and Back-Ends". In: *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 30 (2022), pp. 1586–1599.

[19] Shikun Li, Xiaobo Xia, et al. In: *Advances in Neural Information Processing Systems (NeurIPS)*. Vol. 35. 2022, pp. 24184–24198.

[20] Xuefeng Li et al. "Provably End-to-end Label-Noise Learning without Anchor Points". In: *International Conference on Machine Learning (ICML)*. 2021.

[21] Yang Liu, Hao Cheng, and Kun Zhang. "Identifiability of Label Noise Transition Matrix". In: *Proceedings of the 40th International Conference on Machine Learning (ICML)*. 2023, pp. 21475–21496.

[22] Eran Malach and Shai Shalev-Shwartz. "Decoupling" when to update" from" how to update"". In: *Advances in neural information processing systems* 30 (2017).

[23] Zhen Miao, Michael S Balzer, Ziyuan Ma, et al. "Integrated analysis of multimodal single-cell data". In: *bioRxiv* (2020).

[24] Bengt Muthén and Kerby Shedden. "Finite mixture modeling with mixture outcomes using the EM algorithm". In: *Biometrics* 55.2 (1999), pp. 463–469.

[25] Nima Nouri and Steven H Kleinstein. "Optimized threshold inference for partitioning of clones from high-throughput B cell repertoire sequencing data". In: *Frontiers in immunology* 9 (2018), p. 1687.

[26] Giorgio Patrini et al. "Making deep neural networks robust to label noise: A loss correction approach". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2017), pp. 1944–1952.

[27] Mirco Ravanelli and Maurizio Omologo. "Contaminated speech training methods for robust DNN-HMM distant speech recognition". In: *arXiv preprint arXiv:1710.03538* (2017).

[28] Douglas A Reynolds et al. "Gaussian mixture models." In: *Encyclopedia of biometrics* 741.659-663 (2009).

[29] Leslie N Smith. "Cyclical learning rates for training neural networks". In: *2017 IEEE winter conference on applications of computer vision (WACV)*. IEEE. 2017, pp. 464–472.

[30] Hwanjun Song et al. "Learning from noisy labels with deep neural networks: A survey". In: *IEEE transactions on neural networks and learning systems* 34.11 (2022), pp. 8135–8153.

[31] Yisen Wang et al. "Symmetric cross entropy for robust learning with noisy labels". In: *Proceedings of the IEEE/CVF international conference on computer vision*. 2019, pp. 322–330.

[32] Jiaheng Wei et al. "Learning with noisy labels revisited: A study using real-world human annotations". In: *arXiv preprint arXiv:2110.12088* (2021).

[33] Lijun Wu et al. "Learning to teach with dynamic loss functions". In: *Advances in neural information processing systems* 31 (2018).

[34] Tingting Wu, Xiao Ding, et al. *NoisywikiHow: A Benchmark for Learning with Real-world Noisy Labels in Natural Language Processing*. 2023.

[35] Xiaobo Xia, Tongliang Liu, et al. *Are Anchor Points Really Indispensable in Label-Noise Learning?* 2019.

[36] Yilun Xu et al. "L_dmi: A novel information-theoretic loss function for training deep nets robust to label noise". In: *Advances in neural information processing systems* 32 (2019).

[37] Yu Yao et al. "Dual t: Reducing estimation error for transition matrix in label-noise learning". In: *Advances in neural information processing systems* 33 (2020), pp. 7260–7271.

[38] Kun Yi and Jianxin Wu. "Probabilistic end-to-end noise correction for learning with noisy labels". In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2019, pp. 7017–7025.

[39] Guo-Cheng Yuan et al. "Challenges and emerging directions in single-cell analysis". In: *Genome biology* 18 (2017), pp. 1–8.

[40] Hongyi Zhang et al. *mixup: Beyond Empirical Risk Minimization*. 2018. arXiv: 1710.09412 [cs.LG]. URL: https://arxiv.org/abs/1710.09412.

6 SUPPLEMENTARY MATERIAL

6.1 Statistical Significance Analysis

We compared D&C-Net with the strongest baseline, VolMinNet, using Welch's two-sample t-test, which handles unequal variances and requires only summary statistics (mean $\pm$ SD over three random seeds).

Table 6 reports the p-values: D&C-Net is significantly better ($p < 0.05$) in 3 of 6 settings, VolMinNet in 1 setting (CIFAR-100, Pair-20%), and the remaining 2 settings show no significant difference.

Table 6: Welch's t-test comparing D&C-Net to VolMinNet using summary statistics (mean $\pm$ SD over 3 seeds).

Dataset / Noise	D&C-Net	VolMinNet	p-value
CIFAR-10 / Sym-50%	87.30 ± 0.07	83.37 ± 0.25	< 0.001
CIFAR-100 / Sym-50%	60.39 ± 0.08	53.89 ± 1.26	0.012
CIFAR-10 / Pair-20%	91.86 ± 0.02	90.37 ± 0.30	0.013
CIFAR-10 / Sym-20%	90.61 ± 0.67	89.58 ± 0.26	0.100
MNIST / Sym-50%	98.42 ± 0.03	98.23 ± 0.16	0.170
CIFAR-100 / Pair-20%	65.89 ± 0.48	68.45 ± 0.69	0.0085

6.2 Computational Cost Comparison

D&C-Net introduces a single pre-training phase with a cyclical learning rate, adding roughly 1.5–2$\times$ the wall-clock time of a plain cross-entropy run while keeping memory usage virtually unchanged. This overhead is comparable to other two-stage or dual-network approaches such as Co-teaching and DivideMix, but higher than single-stage transition-matrix methods like Forward, T-Revision, and VolMinNet.