[1] Farah M. Neamah
[1,*] Hadi S. Aghdasi
[1] Pedram Salehpour
[1] Alireza Sokhandan

PRDML+: A Proxy-based Robust Deep Metric Learning versus Label Noise via Self-Supervised Label Refinement Process

Journal of Electrical Systems

***Abstract:* -** Label noise in real-world datasets can negatively impact the efficacy of deep learning models. Manually correcting labels is labor-intensive and impractical for large datasets. Consequently, various methods have been developed to improve the robustness of deep models against label noise. However, most existing methods are primarily designed for classification problems and are not readily applicable to similarity learning applications without adaptation. We tackle this issue by proposing a resilient representation learning approach that models observed labels as a mixture of clean and noisy label distributions. Then, our methods identify label noise data and a semantic embedding jointly using Expectation Maximization (EM) approach. Meanwhile, it progressively modifies targets of data by incorporating a modified self-adaptive training mechanism into the EM algorithm, thereby improving the quality of the semantic embedding using the refined labels. Comprehensive evaluations on datasets containing actual or artificial noisy labels demonstrate that the proposed approach consistently surpasses peer methods experiments on datasets containing real or synthetic label noise demonstrate that our approach consistently surpasses peer methods. Furthermore, it effectively refines noisy labels during the initial training epochs.

Keywords: Robust similarity learning, Label noise, Self-adaptive training, Content-based information retrieval.

I. INTRODUCTION

By utilizing deep neural network models, Deep Metric Learning (DML) has attained eading performance in areas like content-based information retrieval, and person re-identification. DML, also known as representation learning, trains a deep neural network model $\boldsymbol{f}(\cdot,\boldsymbol{\theta})$ (parameterized by $\boldsymbol{\theta}$), which maps the input data to an embedding space in which examples from the same class should be close together, while those from different classes should be far apart [1]. Large datasets are frequently collected from the Internet through crowdsourcing and similar methods. As a result, real-world data sets often include uncertain labels, which can substantially decrease the performance of deep learning methods. Manually cleaning annotations is labor-intensive and impractical for large-scale datasets. Various methods have been proposed to address this challenge. However, these methods are primarily designed on classification and need adaptation to be applicable in DML tasks. Additionally, most research on robust similarity/distance learning against label noise has concentrated on linear metric learning [2-9].

This paper aims to overcome this gap by introducing a robust representation learning method designed to handle label noise. The proposed method called PRDML+ extends our previous work, Proxy-based Robust DML (PRDML) [10] by incorporating a modified self-adaptive training mechanism into the optimization problem and progressively modifying labels. Specifically, PRDML+ employs an Expectation Maximization (EM) approach to jointly identify the noisy label data and learn high-quality semantic embedding. In the Expectation step, the method estimates the clean probability of each label. In the Maximization step, the model's parameters are updated to enhance the likelihood of the observed data, considering the estimated clean probabilities. Simultaneously, it incorporates a self-adaptive training mechanism [11] into the optimization problem for the identified noisy labels to effectively learn useful patterns from them without any supervision signals. This mechanism progressively modifies the labels by leveraging the learned semantic embedding. By initializing the targets using the pre-trained Clip model [12] for label noise data, our label refinement process can effectively modify the noisy labels in the initial training epochs, thereby preventing the propagation of errors. As a result, PRDML+ effectively mitigates the impact of label noise, leading to improved performance in tasks that rely on accurate semantic embeddings, such as image retrieval. We conduct various experiments on several well-known image retrieval benchmarks containing real or synthetic label noise, namely Noisy-Food-101 [13], Bird-Species-200 [14], and Flowers-102 [15] to examine the performance of PRDML+. The outcomes reveal that PRDML+ surpasses current leading methods by a significant margin.

[1] Faculty of Electrical and Computer Engineering, University of Tabriz, Tabriz, Iran

*Corresponding author: aghdasi@tabrizu.ac.ir

II. RELATED WORK

Most research on robust similarity/distance learning against label noise has concentrated on linear metric learning. Yang et al. [2] estimated the probability that a given pair is similar or dissimilar using a maximum entropy optimization. Huang et al. [3] assessed the cleanness probability of a given triplet based on smooth optimization. Wang et al. [4] demonstrated the effects of label noise data to the gradient of the loss function. They then extended Neighborhood Component Analysis (NCA) [16] by reducing this effect. Instead of point estimation of the semantic distance matrix, some methods [5, 7, 17] have learned the posterior distribution of the matrix through a Bayesian framework, to mitigate the effects of noisy inputs and labels. While these methods are effective at handling noise in the input data and the overfitting problem in small datasets, they are less effective in the presence of label noise. Some studies [8, 9] have employed robust loss functions such as the rescaled hinge loss [6] to assign lower weights to potentially noisy triplets during training. The main drawbacks of these works include: 1) They have focused on linear metric learning; thus, they are not applicable to DML directly. 2) These methods have considered fine-grained relationships between data pairs or triplets, failing to capture global interactions among data items and proxies. Few studies have specifically concentrated on addressing label noise within a DML setting. Liu et al. [18] identified noisy labeled examples by computing the similarity between the examples and the identified clean instances, which are stored in a memory bank. Later, they extended the work by introducing various methods that approximate the cleanness probability of the input data [19].

III. THE PROPOSED METHOD

Let $[n] = \{1,2, \dots, n\}$, and $\mathrm{X} \in \mathcal{X} \subseteq \mathbb{R}^{d_c \times d_h \times d_w}$ $Y \in [C]$, $Y^* \in [C]$ denote the random variables of input images, latent true labels, and observed labels. Given a noisy dataset $\mathcal{D} = \{(\boldsymbol{x}_i, y_i)|\ \boldsymbol{x}_i \sim P(X) \text{ and } y_i \sim P(Y|X)\}_{i\in[N]}$, our method learns a robust semantic embedding despite label noise. The proposed method PRDML+ mitigates the effects of label noise by learning a label noise detection module and a semantic embedding simultaneously using the EM optimization mechanism. Meanwhile, our method incorporates a self-adaptive training mechanism [11] to modify the observed labels by leveraging the learned semantic embedding. Fig. 1 shows the main training steps of PRDML+. In the following sections, we provide a detailed explanation of each step of the proposed method.

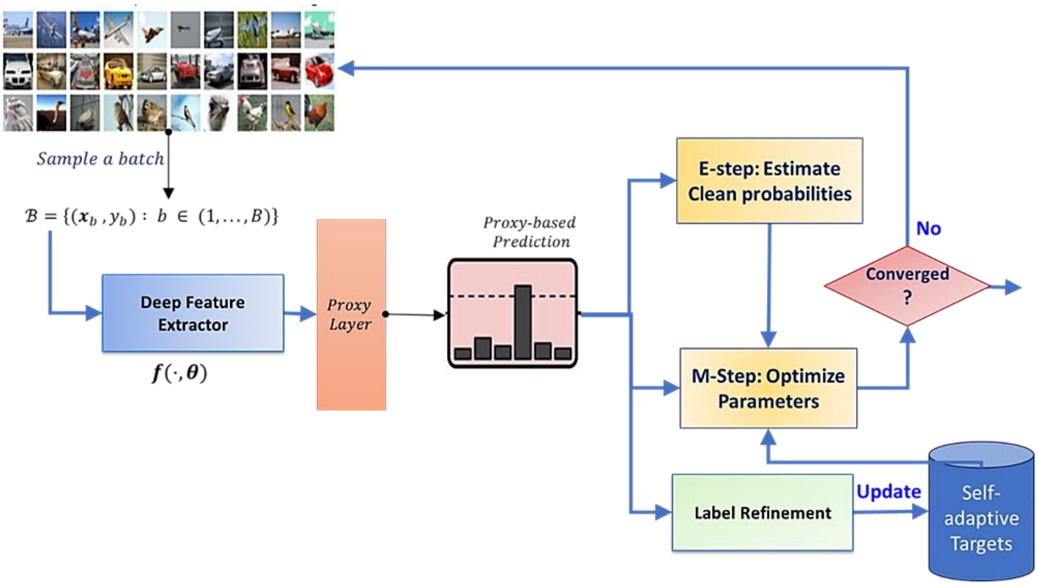


Figure 1: The overall training process of PRDML+.

A. *Probabilistic Model*

Similar to PRDML, let $\{\boldsymbol{w}_k^c\}_{k=1}^K$ denote proxies, where $\boldsymbol{w}_k^c$ represents the k-th proxy of class c. The proxy-based similarity between an input image $\boldsymbol{x}$ and class c is defined as:

$$S_p(\boldsymbol{\alpha}(\boldsymbol{x}), c) = \sum_{k=1}^{K} v^k(\boldsymbol{\alpha}(\boldsymbol{x}), c)\, \boldsymbol{f}(\boldsymbol{\alpha}(\boldsymbol{x}))^\top \boldsymbol{w}_c^k. \tag{1}$$

where, $\boldsymbol{\alpha}$ is a weak augmentation, and $v^k(\boldsymbol{\alpha}(\boldsymbol{x}), c)$ represents $\boldsymbol{w}_c^k$'s weight, defined as:

$$v^k(\boldsymbol{\alpha}(\boldsymbol{x}), c) = \frac{\exp\left(\frac{1}{\gamma}\, \boldsymbol{f}(\boldsymbol{\alpha}(\boldsymbol{x}))^\top \boldsymbol{w}_c^k\right)}{\sum_{j=1}^{K} \exp\left(\frac{1}{\gamma}\, \boldsymbol{f}(\boldsymbol{\alpha}(\boldsymbol{x}))^\top \boldsymbol{w}_c^j\right)} \tag{2}$$

Finally, we apply Softmax the matching scores $[S(\boldsymbol{\alpha}(\boldsymbol{x}), c)]_{c=1}^{C}$ to obtain the proxy-based probabilities:

$$p_{proxy}(Y^* = c|\boldsymbol{x}) = \frac{exp\left(1/\lambda\ S(\boldsymbol{\alpha}(\boldsymbol{x}), c)\right)}{\sum_{l=1}^{C} exp\ (1\ /\ \lambda\ S(\boldsymbol{\alpha}(\boldsymbol{x}), l))} \tag{3}$$

Since proxies represent the entire dataset, *remaining resilient to incorrect labels*, *training both proxies and model parameters* ensures *that* the posterior of true targets is accurately estimated by $p_{proxy}(Y^* = c|)$, meaning that:

$$p(Y^* = c|\boldsymbol{x}) \approx p_{proxy}(Y^* = c|\boldsymbol{x}) \tag{4}$$

Let π represent the probability that the observed label y is sampled from $p(Y^* = y|\boldsymbol{x})$, while with probability $1 - \pi$, it is affected by label noise generated independently of $\boldsymbol{x}$. Thus, the observed label posterior $p(y|\boldsymbol{x})$ can be expressed as:

$$p(Y = y|\boldsymbol{x}) = \pi\, p(Y^* = y|\boldsymbol{x}) + (1 - \pi)\, p_{noisy}(y) \tag{5}$$

where $p_{noisy}(y)$ is the distribution of the noisy labels. Moreover, let the latent binary variable Z indicate whether the example $(\boldsymbol{x}, y)$ is clean (1) or noisy (0). The prior distribution of Z follows a Bernoulli distribution as:

$$p(Z = z|x) = \pi^z (1 - \pi)^{1-z} \tag{6}$$

B. Training Method

Since Z is a latent variable, we use an *Expectation-Maximization* (EM)-like method to optimize the model's parameters. Let Θ indicate all parameters in our model, given the minibatch $\mathcal{M} = \{(\boldsymbol{x}_b\ , y_b) : b\ \in\ [\mathrm{B}]\}$, our method alternates between finding the expectation of Z (i.e., $\mathbb{E}[Z]_{Z\sim p(Z|X,Y)}$) and optimizing the parameters as follows:

Expectation-Step: In this stage, we fix the parameters and compute $\mathbb{E}[Z]$ as follows:

$$\mathbb{E}[Z|\boldsymbol{x}, y] \equiv p(Z = 1|\boldsymbol{x}, y) = \frac{\pi p(Y = y|Z = 1, \boldsymbol{x})}{p(Y = y|\boldsymbol{x})} = \frac{\pi p(Y^* = y|\boldsymbol{x})}{\pi\, p(Y^* = y|\boldsymbol{x}) + (1 - \pi)\, p_{noisy}(y)} \tag{7}$$

where $p_{noisy}(y)$ can be estimated by the empirical distribution of observed labels:

$$p_{noisy}(y) \approx \frac{1}{N} \sum_{i=1}^{N} \mathbb{I}(y_i = y). \tag{8}$$

where, $\mathbb{I}$ denotes the indicator function.

Maximization-Step: This step optimizes Θ by maximizing the *joint* log-likelihood $\log\left(p(Y, Z|X)\right)$ given the posterior of Z. Let $\widehat{\boldsymbol{y}}_b = (\hat{y}_{b1}, \hat{y}_{b2}, \dots, \hat{y}_{bC})$ be the model's prediction for the input $\boldsymbol{x}_i$ **,** in which $\hat{y}_{bc} = p_{proxy}(Y^* = c|\boldsymbol{x}_b)$, following [10], this step leads the following optimization problem:

$$\underset{\boldsymbol{\Phi}}{\text{minimize}}\, \mathcal{L}_M(\mathcal{B}) = \frac{1}{B} \sum_{b=1}^{B} \rho(z_b) \mathcal{L}(y_b, \widehat{\boldsymbol{y}}_b) \tag{9}$$

where $\rho(z_b) = \mathbb{E}[Z = 1|\boldsymbol{x}_b, y_b]$ and $\mathcal{L}(y_b, \widehat{\boldsymbol{y}}_b)$ is the cross-entropy loss defined as:

$$\mathcal{L}(y_b, \widehat{\boldsymbol{y}}_b) = -\ln \hat{y}_{b,y_b} = -\ln p_{proxy}(Y^* = y_b|\boldsymbol{x}_b) \tag{10}$$

In predicting the top five categories for each training image. These categories serve as potential candidates. According to [10], the Clip neural network [12] is used in predicting the top five categories for each training image. These categories serve as potential candidates for the image's true label, allowing our approach to detect label noise early in the training process.

C. Training Method

To exploit the information of identified noisy labels, for the examples with lower values of $\rho(z_b)$, we propose to incorporate a modified self-adaptive training approach [11] into our EM method. Specifically, we first initialize the noisy targets using the pre-trained Clip model[12] and obtain a target $\widetilde{\boldsymbol{y}}_i \in [0,1]^C$, for each data point $\boldsymbol{x}_i$ in the training set. Moreover, we change the optimization problem (9) to consider these targets in addition to observed labels as follows:

$$\underset{\boldsymbol{\Phi}}{\text{minimize}}\, \mathcal{L}_f(\mathcal{M}) = \frac{1}{B}\sum_{b=1}^{B} \rho(z_b)\mathcal{L}(y_b, \widehat{\boldsymbol{y}}_b) + \big(1-\rho(z_b)\big)\mathcal{L}_s(\widetilde{\boldsymbol{y}}_b, \widehat{\boldsymbol{y}}_b) \tag{11}$$

where $\mathcal{L}_s(\widetilde{\boldsymbol{y}}_b, \widehat{\boldsymbol{y}}_b)$ is the *smoothed cross-entropy loss* defined as:

$$\mathcal{L}_s(\widetilde{\boldsymbol{y}}_b, \widehat{\boldsymbol{y}}_b) = -\sum_{c\in[C]} \tilde{y}_{bc} \ln \widehat{\boldsymbol{y}}_{bc} \tag{12}$$

Finally, after optimizing the parameters, we update the targets in the input minibatch, by the Exponential-Moving-Average (EMA) mechanism as follows:

$$\widetilde{\boldsymbol{y}}_b = \lambda_{EMA}\tilde{\boldsymbol{y}}_b + (1-\lambda_{EMA})\,\widehat{\boldsymbol{y}}_b \tag{13}$$

The main steps of *PRDML+* are outlined in Algorithm 1.

Algorithm 1. Training Procedure of *PRDML+*

Input: $\mathcal{D} = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^{N}$, λ_{EMA}: target's update rate.

Output: Trained deep model: $\boldsymbol{f}(\cdot,\boldsymbol{\theta})$ and the proxy set $\{\boldsymbol{w}_k^c\ (k\in[K], c\in[\mathrm{C}])\}$

begin

1. Initialize weights of $\boldsymbol{f}(\cdot,\boldsymbol{\theta})$.
2. Initialize the self-adaptive targets $\{\widetilde{\boldsymbol{y}}_i\}_{i=1}^{N}$using the pre-trained Clip model
3. for j $= 1,2,\ldots MAX$ do
4. $\quad\mathcal{M} \leftarrow$ Sample a batch from $\{(\boldsymbol{x}_i, y_i)\}_{i=1}^{N}$
5. $\quad$Calculate $\rho(z_b)$ by (7), (8)
6. $\quad$Compute $\mathcal{L}_f(\mathcal{M})$ using (11).
7. $\quad$Update self-adaptive targets $\widetilde{\boldsymbol{y}}_b$ in the batch using (13)
8. $\quad$Backpropagate $\mathcal{L}_f(\mathcal{M})$.
9. return $\boldsymbol{f}(\cdot,\boldsymbol{\theta})$ and $\{\boldsymbol{w}_k^c\ (k\in[K], c\in[\mathrm{C}])\}$

IV. Experimental Results

This section presents experiments designed to assess *PRDML+'s performance* on datasets containing either real or artificial label noise. Additionally, we conduct comparisons with *leading* methods and analyze the effectiveness of our label noise detection and refinement modules.

A. Datasets

We evaluated PRDML+ using three image benchmarks: Noisy-Food-101 [13], Bird-Species-200 [14], and Flowers-102 [15]. The specifications and evaluation protocol for each dataset is presented in Table 1. The Noisy-Food-101 contains real-world label noise with approximately 310,009 images of food images of food dishes categorized into 101 classes. This dataset has an estimated label noise rate around 20% and only a subset of images has verified labels. Bird-Species-200 consists of 11,788 images, which are organized into 200 bird species. The images vary in scale, pose, and lighting conditions. The Flowers-102 dataset comprises 8,189 samples classified in 102 categories.

Table 1. Statistics and the evaluation protocol of the evaluated datasets.

Dataset	C	#images	Evaluation Protocol
Noisy-Food-101 [13]	196	310,009	Training set: classes 1-50, validation set: classes 51-75, test set: classes 76-101
Bird-Species-200 [14]	200	11,788	Training set: classes 1-100, validation set: classes 101-150, test set: classes 151-200
Flowers-102 [15]	102	7,370	Training set: classes 1-50, validation set: classes 51-75, test set: classes 76-102

B. Evaluation Metrics

We measured the effectiveness of PRDML+ using Precision@1 (P@1) and Mean Average Precision at K (MAP@K). For each query image $\boldsymbol{q}$, let K be the count of reference images in the same category as $\boldsymbol{q}$, and R_i indicate the count of relevant images within the top i retrieved results by a method. The $P@i$ for $\boldsymbol{q}$ is computed by the formula:

$$\mathrm{P@i}(\boldsymbol{q}) = R_i / i \tag{14}$$

P@1($\mathbf{q}$), which is also referred to as Recall@1($\mathbf{q}$), indicates whether the top-ranked image is relevant to **q** or not. $P@1$ is calculated by averaging P@1($\boldsymbol{q}$) across all queries in the dataset. The Average Precision at K for $\boldsymbol{q}$ (AP@K($\boldsymbol{q}$)) is the mean of P@i($\boldsymbol{q}$) values from position one to K:

$$\mathrm{AP@K}(\mathbf{q}) = \frac{1}{K}\sum_{i=1}^{K} \mathrm{P@i}(\boldsymbol{q}), \tag{15}$$

Finally, *MAP@K* is calculated by averaging the *AP@K* scores across all query images in the dataset:

$$\mathrm{MAP@K} = \frac{1}{n_q}\sum_{i=1}^{n_q} \mathrm{AP@K}(\boldsymbol{q}_i), \tag{16}$$

Where n_q represents the number of query images.

C. Experimental Setup

PRDML+ is developed using the PyTorch framework. Training images were initially scaled to 256 × 256, randomly cropped to 224 × 244, then horizontally flipped at random and normalized. For the test and validation set, the images were resized to 256 × 256, center cropped at 224 × 244, and subsequently normalized. To ensure a fair comparison among all evaluated methods, we used the pre-trained BN-Inception model [20] for all methods. Specifically, features from the final hidden layer were taken as the feature vector. Then, an embedding layer projects the feature vector to the embedding space, reducing the dimensionality from 1024 to 128. We compared the proposed PRDML+ with some leading methods including PRDML [10], Co-teaching+ [21], Co-teaching w/T [22], nSoftmax [22], (PRISM (MCL)) [18], and PRISM (vMF-Sim) [19]. The hyperparameters of PRDML+ were adjusted using validation set as specified in Table 2. We adjusted the Softmax temperature hyperparameter λ in Eq. (2) according to the current epoch as follows:

$$\lambda = \max\left\{\exp\left(-\frac{5 \times epoch}{n_e}\right), 0.1\right\} \tag{17}$$

Table 2. Statistics and the evaluation protocol of the evaluated datasets.

Hyperparameter	Explanation	Range
n_e	maximum number of epochs	100
lr	learning rate	$\{10^{-4}, 3 \times 10^{-4}, 5 \times 10^{-4}, 7 \times 10^{-4} 10^{-3}\}$
d_e	embedding dimension	$\{128\}$
bs	*Batch Size*	$\{32,64,128\}$
opt	*Optimizer*	$\{\mathrm{Adam}\}$
λ	Softmax temperature	$\{1\}$
λ_{EMA}	EMA decay rate	$\{0.9, 0.95, 0.99\}$

D. Performance Evaluation

This experiment evaluates the peer methods on Noisy-Food-101. The Label Noise Level (LNL) of Noisy-Food-101 is about 20%. **Error! Reference source not found.** presents the image retrieval results. The results reveal that PRDML+ effectively learns embeddings according to both MAP@K and P@1 metrics, outperforming other methods substantially. For example, PRDML+ attained a *MAP@K* of 29.53% while PRDML and Co-teaching w/T obtained a *MAP@K* of 28.77% and 22.21%, respectively. The results demonstrate PRDML+'s capability to detect real label noise and the efficacy of the label refinement process. We investigate the contribution of these mechanisms via an ablation study.

Table 3. Statistics and the evaluation protocol of the evaluated datasets.

Method	MAP@K	P@1
Co-teaching+ [21]	17.00	58.91
Co-teaching w/T [22]	22.21	68.56
nSoftmax [22]	19.51	68.11

PRISM (MCL) [18]	15.69	59.54
PRISM (vMF-Sim) [19]	21.52	68.22
PRDML[10]	28.77	75.50
PRDML+	**29.53**	**77.32**

In the second and third experiments, we applied varying levels of synthetic random label noise into the training data for the Bird-Species-200 and Flowers-102 datasets, respectively. We then train the methods on these noisy datasets. The outcomes are presented in Tables 4 and 5.

Table 1. Performance of the methods on the Bird-Species-200 dataset at various LNLs.

Method	LNL=10%		LNL=20%		LNL=50%	
	P@1	MAP@K	P@1	MAP@K	P@1	MAP@K
Co-teaching+ [21]	60.50	19.46	55.98	17.71	52.36	14.73
Co-teaching w/T [22]	60.3	20.28	59.62	20.07	56.77	17.94
nSoftmax [22]	55.21	17.37	53.64	15.88	52.26	14.13
PRISM (MCL) [18]	64.14	23.46	63.58	22.78	60.4	21.45
PRISM (vMF-Sim) [19]	65.41	23.84	64.2	23.16	61.66	20.94
PRDML [10]	66.30	28.50	65.97	26.92	65.30	26.90
PRDML+	**68.21**	**29.41**	**68.07**	**28.22**	**66.67**	**27.12**

Table 2. Performance of the methods on the Flowers-102 dataset at various LNLs.

Method	LNL=10%		LNL=20%		LNL=50%	
	P@1	MAP@K	P@1	MAP@K	P@1	MAP@K
PRISM (MCL) [18]	92.69	52.43	92.94	49.48	91.44	44.1
PRISM (vMF-Sim) [19]	93.04	51.22	93.64	48.51	92.11	43.48
PRDML[10]	96.03	53.52	94.81	53.06	94.27	50.28
PRDML+	**96.67**	**55.76**	**95.52**	**55.29**	**95.12**	**53.27**

The results indicate that PRDML+ consistently outperformed other methods across various levels of label noise in the Bird-Species-200 and the Flowers-102 datasets. For instance, PRDML+ achieved a *MAP@K* of 53.27% on the Flowers-102 dataset with a 50% label noise rate, while the second and third-best methods (PRDML and PRISM (vMF-Sim)) attained *MAP@K* scores of 50.47% and 43.76%, respectively. Additionally, PRDML+ and PRDML demonstrated greater robustness to label noise. For example, the *P@1* of PRDML+ only slightly decreased from 68.21% to 66.67% as the LNL rose from 10% to 50% on the Bird-Species-200 dataset. In contrast, PRISM (vMF-Sim) showed a more significant drop, decreasing from 65.41% to 61.66%. Likewise, on the Flowers-102 dataset, the MAP@K of PRDML+ dropped from 55.76% to 53.27% as the LNL increased from 10% to 50%, whereas PRISM (vMF-Sim) showed a substantial decrease from 51.22% to 43.48%. We observe that PRDML+ consistently slightly achieves higher performance compared to PRDML across all benchmarks. This suggests that the advancements introduced in PRDML+ offer a marginal but notable improvement in retrieval performance, particularly in fine-tuning the top result rankings.

E. Effectiveness of the Label Refinement

To assess the impact of the label refinement mechanism, we derived a variant of PRDML+ called *PRDML+ w/o label refinement* (LR), which simply discards training examples marked as label noise by the label noise detection process. **Error! Reference source not found.** compares the performance of PRDML+ and *PRDML+ w/o* LR on the Flowers-102 (LNL=50%) dataset.

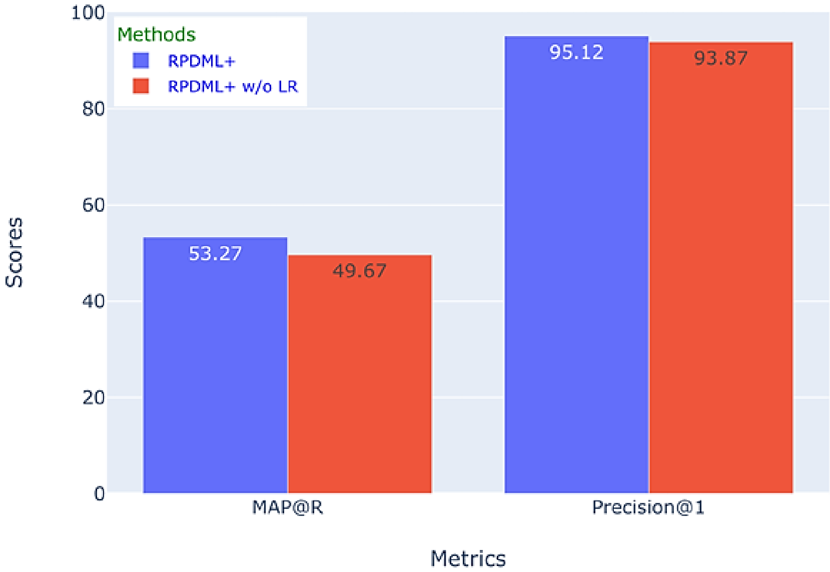


Figure 2: Comparison of PRDML+ and *PRDML+ w/o* LR on the Flowers-102 (LNL=50%) dataset.

The outcomes indicate that the label refinement process can successfully refine the noisy labels, thereby improving the learned embedding. For instance, the P@1 of PRDML+ dropped substantially from 53.27% to 49.67% by eliminating the label refinement process.

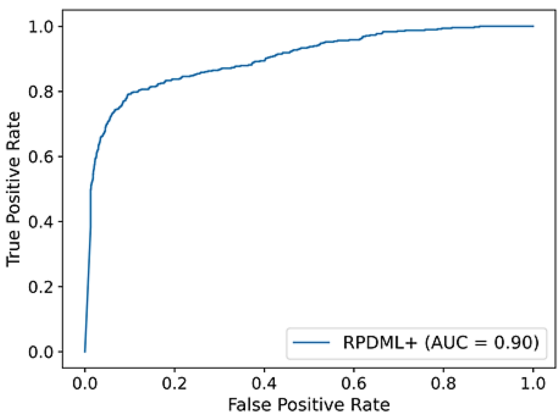
F. Efficacy of the Label Noise Detection Mechanism

This experiment assesses the efficacy of the label noise detection mechanism using standard classification criteria. Table 6 presents the performance of this mechanism at the fifth epoch for each evaluated dataset under a high LNL scenario (LNL=50%). Additionally, Figure 3 displays the corresponding ROC curves.

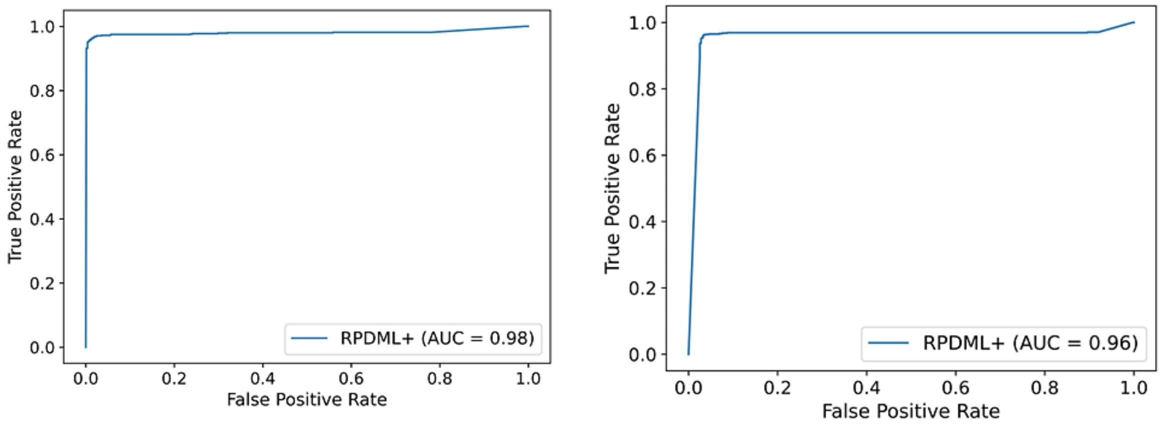
Table 6. Performance of the label noise detection mechanism at fifth epoch

Dataset	Accuracy	Recall	Precision	Specificity	F1-score
Noisy-Food-101	90.39	66.25	79.70	96.04	72.35
Bird-Species-200 (LNL=50%)	97.27	97.07	97.46	97.47	97.26
Flowers-102 (LNL=50%)	94.40	96.80	92.37	92.00	94.53

Note that the equation is centered using a center tab stop. Be sure that the symbols in your equation have been defined before or immediately following the equation. Use “(1)”, not “Eq. (1)” or “equation (1)”, except at the beginning of a sentence: “Equation (1) is . . .”



(*a*) Noisy-Food-101



(*b*) Bird-Species-200 (LNL=50%) (c) Flowers-102 (LNL=50%)

Figure 3: ROC plots of the label noise detection mechanism (epoch=5).

As the results indicate, by leveraging the top-five predictions from the pre-trained Clip model, PRDML+ can reliably identify mislabeled data, even during early training epochs, thereby enhancing the robustness of the learned embedding against label noise. Additionally, since the mislabeled images in the Noisy-Food-101 dataset present significant challenges, PRDML+'s performance on this dataset is comparatively lower than on others.

V. Conclusioin

This paper presents a robust DML method called PRDML+. PRDML+ extends our previous work [10] by incorporating a self-adaptive training mechanism into the optimization problem to address identified noisy labels. This mechanism enables our method to effectively learn useful patterns from label noise data without relying on any supervision signals. We conducted a series of experiments to evaluate the effectiveness of PRDML+ in comparison to several state-of-the-art techniques for image retrieval tasks involving both real and artificial label noise. The results demonstrate that PRDML+ develops a robust representation and consistently outperforms competing methods. Additionally, we analyzed both label refinement and noise detection performance within our method. The outcomes reveal that our label refinement approach effectively corrects noisy labels, resulting in a more accurate learned embedding. For instance, the P@1 of PRDML+ dropped significantly from 53.27% to 49.67% when the label refinement process was excluded in a high label noise level (LNL) = 50% scenario on the Flowers-102 dataset. Moreover, PRDML+'s capability for detecting noisy labels is remarkable even in the initial training epochs, which enhances the robustness of the learned embedding against label noise. For example, our method achieved an accuracy of 97.27% in detecting label noise at the fifth epoch on the Bird-Species-200 dataset (LNL=50%). In future research, we plan to investigate applying PRDML+ to other DML tasks, such as recommender systems and face recognition.

References

[1] D. Zabihzadeh, Z. Alitbi, and S. J. Mousavirad, "Ensemble of loss functions to improve generalizability of deep metric learning methods," *Multimedia Tools and Applications,* vol. 83, no. 7, pp. 21525-21549, 2024.

[2] T. Yang, R. Jin, and A. K. Jain, "Learning from noisy side information by generalized maximum entropy model," in *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, 2010: Citeseer, pp. 1199-1206.

[3] K. Huang, R. Jin, Z. Xu, and C.-L. Liu, "Robust metric learning by smooth optimization," *arXiv preprint arXiv:1203.3461,* 2012.

[4] D. Wang and X. Tan, "Robust Distance Metric Learning in the Presence of Label Noise," in *AAAI*, 2014, pp. 1321-1327.

[5] D. Wang and X. Tan, "Robust Distance Metric Learning via Bayesian Inference," *IEEE Transactions on Image Processing,* vol. 27, no. 3, pp. 1542-1553, 2018.

[6] G. Xu, Z. Cao, B.-G. Hu, and J. C. Principe, "Robust support vector machines based on the rescaled hinge loss function," *Pattern Recognition,* vol. 63, pp. 139-148, 2017.

[7] D. Zabihzadeh, R. Monsefi, and H. S. Yazdi, "Sparse Bayesian approach for metric learning in latent space," *Knowledge-Based Systems,* vol. 178, pp. 11-24, 2019.

[8] S. A. R. Al-Obaidi, D. Zabihzadeh, and H. Hajiabadi, "Robust metric learning based on the rescaled hinge loss," *International Journal of Machine Learning and Cybernetics,* vol. 11, no. 11, pp. 2515-2528, 2020.

[9] D. Zabihzadeh, A. Tuama, A. Karami-Mollaee, and S. J. Mousavirad, "Low-rank robust online distance/similarity learning based on the rescaled hinge loss," *Applied Intelligence,* vol. 53, no. 1, pp. 634-657, 2023.

[10] F. M. Neamah, H. S. Aghdasi, P. Salehpour, and A. S. Sorkhabi, "Proxy-based robust deep metric learning in the presence of label noise," *Physica Scripta,* vol. 99, no. 7, p. 076013, 2024.

[11] L. Huang, C. Zhang, and H. Zhang, "Self-adaptive training: Bridging supervised and self-supervised learning," *IEEE transactions on pattern analysis and machine intelligence,* vol. 46, no. 3, pp. 1362-1377, 2022.

[12] A. Radford *et al.*, "Learning transferable visual models from natural language supervision," in *International conference on machine learning*, 2021: PMLR, pp. 8748-8763.

[13] K.-H. Lee, X. He, L. Zhang, and L. Yang, "Cleannet: Transfer learning for scalable image classifier training with label noise," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 5447-5456.

[14] C. Wah, S. Branson, P. Welinder, P. Perona, and S. Belongie, "The Caltech-UCSD Birds-200-2011 Dataset, Computation & Neural Systems," Technical Report, CNS-TR-2011-001, 2011.

[15] M.-E. Nilsback and A. Zisserman, "Automated flower classification over a large number of classes," in *2008 Sixth Indian Conference on Computer Vision, Graphics & Image Processing*, 2008: IEEE, pp. 722-729.

[16] J. Goldberger, G. E. Hinton, S. T. Roweis, and R. R. Salakhutdinov, "Neighbourhood components analysis," in *Advances in neural information processing systems*, 2005, pp. 513-520.

[17] D. Zabihzadeh, R. Monsefi, and H. S. Yazdi, "Sparse Bayesian similarity learning based on posterior distribution of data," *Engineering Applications of Artificial Intelligence,* vol. 67, pp. 173-186, 2018.

[18] C. Liu *et al.*, "Noise-resistant deep metric learning with ranking-based instance selection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 6811-6820.

[19] C. Liu *et al.*, "Noise-Resistant Deep Metric Learning with Probabilistic Instance Filtering," *arXiv preprint arXiv:2108.01431,* 2021.

[20] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818-2826.

[21] X. Yu, B. Han, J. Yao, G. Niu, I. Tsang, and M. Sugiyama, "How does disagreement help generalization against label corruption?," in *International Conference on Machine Learning*, 2019: PMLR, pp. 7164-7173.

[22] A. Zhai, H.-Y. Wu, and U. San Francisco, "Classification is a Strong Baseline for Deep Metric Learning."