RoDA: Robust Domain Alignment for Cross-domain Retrieval Against Label Noise

Ziniu Yin [1], Yanglin Feng [1], Ming Yan [3], Xiaomin Song [4], Dezhong Peng [1,4], Xu Wang [1,2]*
[1]The College of Computer Science, Sichuan University, Chengdu, China
[2]Tianfu Engineering-oriented Numerical Simulation & Software Innovation Center, Chengdu, China
[3]Centre for Frontier AI Research (CFAR), A*STAR, Singapore
[4]Sichuan National Innovation New Vision UHD Video Technology Co., Ltd., Chengdu, China
yinzn1014@163.com, fcyzfyl@163.com, yanmingtop@gmail.com, songxiaomin@uptcsc.com , pengdz@scu.edu.cn,
wangxu.scu@gmail.com

Abstract

This paper studies the complex challenge of cross-domain image retrieval under the condition of noisy labels (NCIR), a scenario that not only includes the inherent obstacles of traditional cross-domain image retrieval (CIR) but also requires alleviating the adverse effects of label noise. To address this challenge, this paper introduces a novel Robust Domain Alignment framework (RoDA), specifically designed for the NCIR task. At the heart of RoDA is the Selective Division and Adaptive Learning mechanism (SDAL), a key component crafted to shield the model from overfitting the noisy labels. SDAL effectively learns discriminative knowledge by dividing the dataset into clean and noisy parts, subsequently rectifying the labels for the latter based on information drawn from the clean one. This process involves adaptively weighting the relabeled samples and leveraging both the clean and relabeled data to bootstrap model training. Moreover, to bridge the domain gap further, we introduce the Accumulative Class Center Alignment (ACCA), a novel approach that fosters domain alignment through an accumulative domain loss mechanism. Thanks to SDAL and ACCA, our RoDA demonstrates its superiority in overcoming label noise and domain discrepancies within the NCIR paradigm. The effectiveness and robustness of our RoDA framework are comprehensively validated through extensive experiments across three multi-domain benchmarks.

Code — https://github.com/yznovo/RoDA

Introduction

Given a query image from one domain, cross-domain image retrieval (CIR) is designed to search relevant images across different domains by leveraging the similarities of visual features. Based on the iterative advancements of image retrieval techniques and accumulation of massive volume data from different domains, CIR gradually serves as a fundamental task across a variety of applications, including e-commerce (Huang et al. 2015), social media (Yang, Zhang, and Xu 2014), medical image analytics (Zheng, Li, and Liu 2023), and beyond (Li et al. 2022b). Accomplishing the CIR task involves tackling two main challenges: learning feature representations that are both compact and discriminative within domains, and achieving alignment across domains with distinct characteristics and distributions. A considerable volume of research, represented by (Taylor and Stone 2007; Xie et al. 2019; Huang et al. 2020; Wang et al. 2020a; Hu, Zhang, and Lee 2023; Yin et al. 2024; Li et al. 2024; Wang, Du, and Zhu 2024; Iijima and Stathaki 2024) has been dedicated to solving these challenges. During these research endeavors, supervised approaches (Taylor and Stone 2007; Xie et al. 2019; Huang et al. 2020; Liu et al. 2024) have emerged as the prevalent choice due to their superior performance. Nevertheless, it should be noted that supervised CIR approaches assume the correctness of label annotations,

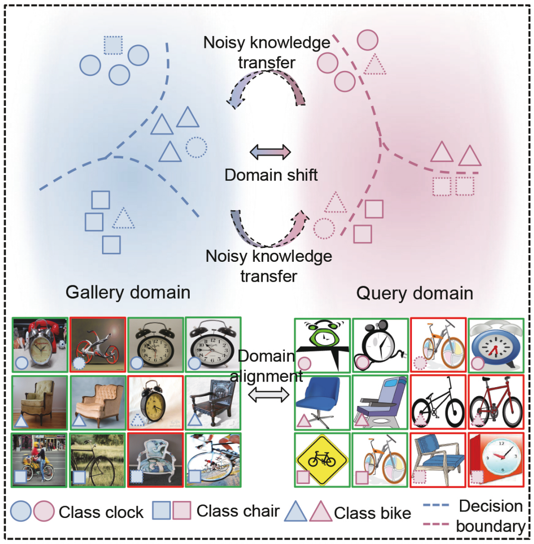


Figure 1: Visualization of cross-domain image retrieval with noisy labels (NCIR), the upper part depicts that the clean and the noisy knowledge are both transferred between the query and gallery domain, and the lower part presents some specific image samples from both domains corresponding to the upper part. Distinct colors represent various domains, while unique shapes denote different classes. Besides, dotted shapes represent samples with noisy labels within each class.

---

[*] Corresponding author

underscoring their substantial dependence on the quality of data annotations. To alleviate this dependency, unsupervised CIR methods (Wang et al. 2023b; Hu, Zhang, and Lee 2023; Yin et al. 2024; Li et al. 2024; Wang, Du, and Zhu 2024; Iijima and Stathaki 2024), which do not require labels, have been developed. However, these methods often fall short in retrieval accuracy. In reality, collecting a perfectly annotated dataset is not only laborious but also time-consuming, especially for complex, multi-domain datasets. Yet, datasets of lesser annotation quality, obtained through web crawling or crowdsourcing, are far more accessible in practical settings. It is widely acknowledged that such lower-quality labels introduce a higher incidence of incorrect labels (*i.e.*, noisy labels), significantly reducing the efficacy of current supervised approaches. To explore the negative impact of noisy labels in the CIR, this paper introduces a new paradigm called cross-domain image retrieval with noisy labels (NCIR). As illustrated in Figure 1, noisy labels (Liu and Tao 2015; Jiang et al. 2020; Bai et al. 2021; Liu et al. 2021; Nishi et al. 2021) from unrelated categories may be erroneously integrated into each category in both domains. Such noisy labels confuse learning features within domains and hinder the acquisition of domain-invariant knowledge due to the transfer of inaccurate information between domains. This, in turn, leads to a noticeable decline in retrieval performance.

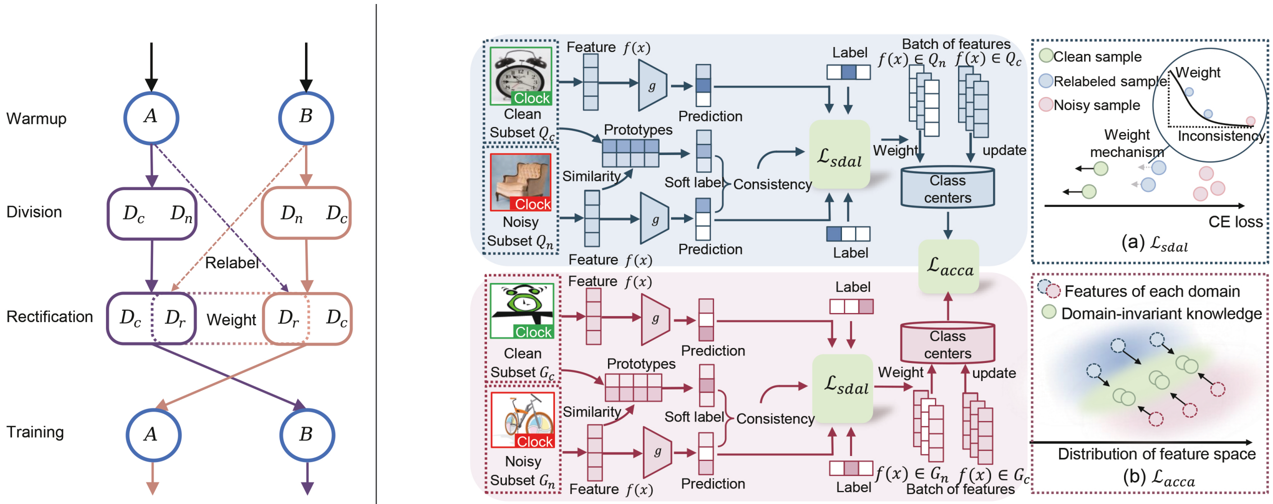
To tackle the above-mentioned challenges, this paper proposes a novel method called RoDA. In detail, RoDA incorporates the Selective Division and Adaptive Learning (SDAL) and Accumulative Class Center Alignment (ACCA) Mechanisms. Unlike existing noise label learning methods (Li, Socher, and Hoi 2020; Li et al. 2022a) that treat noisy data as unlabeled or attempt to refine noisy labels based on model predictions, our SDAL innovatively reassigns labels for the noisy data subset by leveraging insights from the clean subset after dividing the dataset into two subsets. This relabeling process involves the utilization of feature prototypes extracted from the clean subset, ensuring a reliable rectification of noisy labeled samples. Subsequently, an adaptive weighting mechanism, guided by the re-assigned labels and model predictions, directs the assignment of the weights. With relabeling and weighting, SDAL provides an effective approach to the challenges posed by label noise in cross-domain retrieval tasks. On the other hand, to address the issue of domain disparity, we introduce ACCA, a strategy designed to progressively align domains by minimizing an accumulative domain loss. This loss calculation draws upon domain class centers, which are accumulatively updated throughout the training process. ACCA utilizes not only feature insights derived from clean samples but also adaptively incorporates knowledge from relabeled samples, facilitating the refinement of domain class centers for both the query and gallery domains. Attributing to the exploitation of comprehensive reliable feature information to construct class centers, ACCA steadily diminishes the domain differences.

The main contributions of this paper are summarized below: (1) This paper notes the existence and adverse effects of noisy labels in CIR, a challenge we refer to as cross-domain image retrieval with noisy labels (NCIR). To solve the problems in NCIR, this paper proposes RoDA to achieve robust retrieval. As far as we know, this work could be the initial foray into exploring this specific challenge. (2) To mitigate the interference of noisy labels in NCIR, we introduce SDAL to achieve robust and adaptive learning. SDAL partitions the dataset into clean and noisy subsets and relabels the noisy data with reliable labels. Furthermore, an adaptive weighting mechanism assigns appropriate weights to relabeled data for effective training. (3) To bridge the gap between domains, ACCA is proposed to learn the domain-invariant knowledge. This is achieved by minimizing an accumulative domain loss that relies on the domain class centers, drawing upon representations gleaned from both clean and meticulously relabeled samples. (4) The experimental results demonstrate the effectiveness and robustness of our RoDA across a spectrum of noise conditions within NCIR tasks. Remarkably, RoDA consistently surpasses existing benchmarks, achieving superior cross-domain retrieval performance.

Related Work

Learning with Noisy Labels

Collecting a dataset with fully correct labels is a very time-consuming and labor-intensive task. Researchers are beginning to explore noisy label learning (NLL) (Liu and Tao 2015; Feng et al. 2023a; Zheng et al. 2020; Feng et al. 2023b; Liu et al. 2024) that accomplishes the model training with low-quality datasets (Natarajan et al. 2013) as a direction to mitigate the heavy dependence on label annotations (Xiao et al. 2015) and some effective methods are proposed. Currently, there are several types of approaches (Kim et al. 2019; Yi and Wu 2019; Nishi et al. 2021; Bai et al. 2021) to achieving robust training (Li et al. 2017) for noisy label learning. For example, GCE (Zhang and Sabuncu 2018), combines the cross entropy (CE) and mean absolute error (MAE) to form a robust loss function to train the model that can be seen as a generalization of CE and MAE loss. To alleviate the confirmation bias, Co-teaching (Han et al. 2018) and JoCoR (Wei et al. 2020) propose the effective learning paradigm which utilizes two models simultaneously to select the clean samples, and each network teaches the knowledge to its peer network. Due to the two networks having different learning abilities, they can neglect different label noise, which makes the models overfit less error information during training. And one more effective line for NLL is combining sample selection and semi-supervised learning (Wang et al. 2020b; Ding et al. 2018). Typically, the DivideMix (Li, Socher, and Hoi 2020), SSR (Feng, Tzimiropoulos, and Patras 2021) and NCE (Li et al. 2022a) show great results by training the model on both labeled and unlabeled data in a semi-supervised manner after the sample selection. These sample selection-based approaches mitigate susceptibility to noisy labels by filtering out samples with noisy labels. However, all current strategies have primarily been applied in single-domain scenarios. Their effectiveness diminishes significantly when applied to cross-domain tasks, due to differences in distribution and imbalances in sample categories across domains.



(a) Training Pipeline

(b) Training Process

Figure 2: The framework of our RoDA. (a) The workflow of our method. RoDA involves two peer models A and B. First, we warm up models through all training data using the $\mathcal{L}_{warmup}$ to achieve initialization. Next, RoDA accomplishes the division and relabeling to obtain the relabeled data, then adaptively assigns weights for these relabeled samples. Finally, with clean and relabeled data, RoDA implements the main training phase. (b) The main training process of RoDA. After the warmup phase, each model obtains the clean and noisy subset of query and gallery domains by selective division. Initially, RoDA utilizes the information of clean data to calculate the feature prototypes. With these prototypes, RoDA relabels the noisy data and assigns weights through the adaptive weighting mechanism after relabeling. Eventually, RoDA calculates the $\mathcal{L}_{sdal}$ with clean and weighted relabeled samples in both domains. Meanwhile, the features of clean and relabeled samples after weighting are utilized to update the domain class centers and obtain the $\mathcal{L}_{acca}$.

Cross-domain Image Retrieval

Cross-domain Image retrieval (CIR) (Li et al. 2013) is based on image retrieval (Zheng 2015), which aims to retrieve the images in the gallery domain matching the category of given image from query domain (Xie et al. 2019; Lin et al. 2016). Compared to traditional image retrieval, CIR considers the complexities of cross-domain requirements. Several effective supervised methods (Gajic and Baldrich 2018; Guo et al. 2017; Wang et al. 2019; Chen et al. 2019; Xie et al. 2019; Wang et al. 2020a; Li et al. 2022b; Wang et al. 2023a) have been introduced to bridge the domain gap, addressing the challenges inherent in this issue. For example, Bojana et.al (Gajic and Baldrich 2018) propose a three-stream Siamese architecture and use the triplet loss for training. Guo et.al (Guo et al. 2017) utilize the generative adversarial network to generate the imaginary images combining the characteristics of query and gallery domains. However, these methods are heavily dependent on the accuracy of sample labeling. Moreover, some methods even need to match samples belonging to the same categories from query and gallery domains to form training sample pairs (Taylor and Stone 2007). Beyond the supervised approaches mentioned above, there also exist some unsupervised cross-domain image retrieval (UCIR) methods as well (Taigman, Polyak, and Wolf 2016; Huang et al. 2020; Hu, Zhang, and Lee 2023; Wang et al. 2023b). Unsupervised methods are free from dependence on labels, but the corresponding results are not satisfactory. In this paper, we propose an untouched but challenging problem, *i.e.*, cross-domain image retrieval with noisy labels (NCIR). It necessitates feature learning and inter-domain alignment in the face of noisy label interference. Unlike supervised CIR, which demands accurate labels and even matching pairs, and unsupervised CIR, limited by ignoring labels, NCIR emerges as a more practical task. It strikes a balance between the need for high-quality labels and the pursuit of superior performance.

Problem Formulation

For clarity and ease of understanding, we start by providing the symbols used in this paper along with the definitions. We use $\boldsymbol{D} = \{\boldsymbol{D}_Q, \boldsymbol{D}_G\}$ to denote the cross-domain datasets. Here, $\boldsymbol{D}_Q = \{\boldsymbol{X}_Q, \boldsymbol{Y}_Q\} = \{\boldsymbol{x}_i^Q, \widetilde{y}_i^Q\}_{i=1}^{N_Q}$ and $\boldsymbol{D}_G = \{\boldsymbol{X}_G, \boldsymbol{Y}_G\} = \{\boldsymbol{x}_i^G, \widetilde{y}_i^G\}_{i=1}^{N_G}$, represent the samples of query domain and gallery domain with noisy labels respectively, where N denotes the number of samples in domains, $\boldsymbol{x}_i^d$ represents the i-th sample in domain d ($d \in \{Q, G\}$), $\widetilde{y}_i^d \in \{1, \cdots, c\}$ represents its corresponding class label which may be different from its ground truth label and c indicates the total number of categories.

Overview of Method

To address the challenges of the NCIR task, we propose Robust Domain Alignment for Cross-domain Retrieval against Label Noise (RoDA), as shown in Figure 2. RoDA contains two training steps: the warmup phase and the main training phase.

During the warmup phase, we use the original labels of the samples for supervision and employ cross entropy as the defined training objective, which is expressed as follows:

$$\mathcal{L}_{warmup} = -\frac{1}{N}\sum_{i=1}^{N} \widetilde{y}_i \log\left(\boldsymbol{g}\left(\boldsymbol{f}\left(\boldsymbol{x}_i\right)\right)\right), \tag{1}$$

where N is the number of all training samples, $\boldsymbol{f}$ and $\boldsymbol{g}$ represent the feature extractor and classifier respectively. After warmup, we can obtain the initialized models with fairly reliable parameters thanks to the memory effect of the deep neural network (Arpit et al. 2017).

In the main training phase, our RoDA incorporates three parts: selective division (SD), adaptive learning (AL), and accumulative alignment. Specifically, the Selective Division and Adaptively Learning Mechanism (SDAL) first divides the whole dataset into clean and noisy subsets through sample selection. After that, SDAL utilizes information from the clean subset to relabel the noisy subset. This process rectifies mislabeled samples, further reducing label noise and increasing the amount of reliable training data. Further, SDAL assigns weights to these relabeled samples using an adaptive weighting mechanism. Additionally, Accumulative Class Center Alignment (ACCA) is employed to eliminate the domain gap by minimizing the domain loss based on class centers. By learning knowledge within and across domains together, we effectively achieve the cross-domain image retrieval task with the interference of label noise. The overall training objective function for RoDA is formulated as follows:

$$\mathcal{L}_{roda} = \mathcal{L}_{sdal} + \lambda\mathcal{L}_{acca}, \tag{2}$$

where $\mathcal{L}_{sdal}$ and $\mathcal{L}_{acca}$ respectively define the loss terms utilized by SDAL and ACCA. λ is the trade-off parameter. In the following section, we will describe RoDA in more detail.

Selective Division and Adaptive Learning

To alleviate the erroneous information brought by label noise in cross-domain image retrieval with the noisy labels (NCIR), we propose SDAL, inspired by the methods based on sample selection (Li, Socher, and Hoi 2020; Li et al. 2022a). It splits samples into two distinct groups for processing, aimed at diminishing the impact of noisy labels. However, different from these methods, SDAL enhances this approach by rectifying the labels of all samples identified as noisy, utilizing feature prototypes derived from clean samples. Additionally, SDAL applies an adaptive weighting mechanism to these relabeled samples, further optimizing the process.

Selective Division

Considering the confirmation bias, we utilize the neighborhood information of given samples to detect noisy labels. Specifically, it takes the inconsistency between the label of given samples and their k nearest neighbors in the feature space to identify the probably corrupted labels. Given a sample $\boldsymbol{x}_i^d$, we first get its k nearest neighbors $\mathcal{N}_k^d(i)$ through the cosine similarity in the feature space:

$$\mathcal{N}_k^d(i) \leftarrow \mathbf{KNN}(\boldsymbol{x}_i^d, k), \tag{3}$$

where $\mathbf{KNN}$ is a function which returns k nearest neighbors of the given sample $\boldsymbol{x}_i^d$ from the d domain. Then, the corresponding label inconsistency LI can be calculated as follows:

$$LI_i^d = \frac{1}{k}\left(\sum_{\boldsymbol{x}\in\mathcal{N}_k^d(i)} JS\left(\widetilde{y}_i^d, softmax\left(\boldsymbol{g}\left(\boldsymbol{f}\left(\boldsymbol{x}\right)\right)\right)\right)\right), \tag{4}$$

where $\widetilde{y}_i^d$ is the original label of given sample $\boldsymbol{x}_i^d$ in d domain, JS is the Jensen-Shannon divergence. Based on LI, we can achieve the division by the hyperparameter θ, which serves as a threshold:

$$\begin{cases} \boldsymbol{x}_i^d \in \boldsymbol{D}_{cln}^d, & LI_i^d < \theta \\ \boldsymbol{x}_i^d \in \boldsymbol{D}_{noi}^d, & LI_i^d \geq \theta \end{cases}, \tag{5}$$

where the $\boldsymbol{D}_{cln}^d$ and $\boldsymbol{D}_{noi}^d$ are the clean subset and noisy subset. After the division of each epoch, we can obtain probably clean samples and take their original labels $\widetilde{y}_i$ as correct supervisions to apply the cross entropy loss for training. The equation is as follows:

$$\mathcal{L}_{cln} = -\frac{1}{N_{cln}}\sum_{i=1}^{N_{cln}} \widetilde{y}_i \log\left(\boldsymbol{g}\left(\boldsymbol{f}\left(\boldsymbol{x}_i\right)\right)\right), \tag{6}$$

where the N_{cln} indicates the number of all samples of clean data.

Adaptive Learning

After the selective division (SD), we could get a highly credible clean subset that can be used directly for training. Nevertheless, the limited size of this subset could potentially lead to insufficient representation learning (Li et al. 2019), which is more obvious when in the case of severe label noise. Some noisy label learning (NLL) methods (Li, Socher, and Hoi 2020; Li et al. 2022a; Feng, Tzimiropoulos, and Patras 2021) also consider this problem and propose to use of semi-supervised learning to utilize the noisy data as unlabeled data or to generate pseudo labels for noisy data based on the predictions of the model. However, the performance is not satisfactory in any of the above approaches due to the underutilization of information from noisy labeled data. To better leverage the noisy subset, we propose temporal feature prototype refinement (TFPR). Specifically, TFPR refines the prototypes with each epoch, incorporating both the clean data identified in the current division and the prototypes cumulated from previous divisions. The current prototypes are the mean values of all corresponding features belonging to the same categories in clean subsets. In detail, the current prototypes $\dot{\boldsymbol{V}}^d(d \in Q, G)$ is denoted as follows:

$$\dot{\boldsymbol{V}}^d = [\dot{\boldsymbol{v}}_1^d, \cdots, \dot{\boldsymbol{v}}_c^d], \tag{7}$$

where $\dot{\boldsymbol{v}}_t^d = \frac{1}{n_t}(\sum_{i=1}^{n_t} \boldsymbol{f}(\boldsymbol{x}_{i,t}^d))$, denoting the current prototype of t-th class within d domain. And $\boldsymbol{x}_{i,t}^d$ represents the i-th sample with label of t in the d domain. n_t means the number of samples with label of t. During the training, the feature prototypes $\boldsymbol{V}^d(d \in Q, G)$ are refined with every epoch:

$$\begin{cases} \boldsymbol{v}_t^d = \gamma \cdot \boldsymbol{v}_t^d + (1-\gamma)\cdot\dot{\boldsymbol{v}}_t^d, & n_t \neq 0, \\ \boldsymbol{v}_t^d = \boldsymbol{v}_t^d, & n_t = 0, \end{cases} \tag{8}$$

where $\boldsymbol{v}_t^d$ represent the prototypes cumulated during the previous training epoch and are all initialized with $\mathbf{0}$, *i.e.*, $\boldsymbol{v}_t^d = \mathbf{0}$ at the beginning. γ is the trade-off parameter. To handle situations where certain class samples are absent in the current clean subset due to severe label noise, we take this temporal accumulation way.

Once the prototypes are calculated, we can re-assign labels for the noisy data by the cosine similarities between the features of noisy labeled samples and the prototypes:

$$\hat{y}_i = softmax\left(\boldsymbol{f}\left(\boldsymbol{x}_i\right) \cdot \boldsymbol{V}^T\right), \tag{9}$$

where $\hat{y}_i$ is the newly assigned label. Note that we employ soft labels rather than hard labels here since soft labels are more robust to label noise and encapsulate richer information compared to hard ones. Additionally, we propose a novel adaptive weighting mechanism W, which relies on the consistency between the model predictions and the new soft labels:

$$W_i = \frac{(1+\beta)}{(\beta + e^{\beta \cdot ce_i})}, \tag{10}$$

where W_i is the adaptive weight for $\boldsymbol{x}_i$. ce_i is calculated as $ce_i = -\hat{y}_i \log(\boldsymbol{g}(\boldsymbol{f}(\boldsymbol{x}_i)))$. And $\beta \geq 0$ is a hyperparameter that controls the sensitivity of W_i to changes in ce_i. To elaborate further, W adaptively adjusts the weights of relabeled data based on the cross-entropy between the classifier outputs and the reassigned labels. This measure reflects the alignment between the new label and the true label for each noisy labeled sample. Consequently, a smaller cross-entropy signifies a more accurate match, prompting a corresponding increase in the assigned weights. Based on this, the training objective of relabeled samples is as follows:

$$\mathcal{L}_{rlb} = -\frac{1}{N_{rlb}} \sum_{i=1}^{N_{rlb}} \hat{y}_i \log\left(\boldsymbol{g}\left(\boldsymbol{f}\left(\boldsymbol{x}_i\right)\right)\right) \cdot W_i, \tag{11}$$

where N_{rlb} is the number of relabeled samples.

Finally, the loss function $\mathcal{L}_{sdal}$ of SDAL can be summarized as:

$$\mathcal{L}_{sdal} = \mathcal{L}_{cln} + \omega \mathcal{L}_{rlb}, \tag{12}$$

where $\omega = \frac{CurrentEpoch}{MaxEpoch}$ is a coefficient that grows linearly with training time.

Accumulative Class Center Alignment

Our SDAL effectively tackles the challenge of learning discriminative features within domains. However, the problem of discrepancy in data distribution and style between domains—referred to as the domain gap—remains. To address the issue of the domain gap between the query domain and the gallery domain, we introduce the Accumulative Class Center Alignment (ACCA) strategy. ACCA employs an accumulative domain loss that minimizes the discrepancy between the query and gallery class centers, effectively bridging the domain gap and enhancing retrieval performance. In detail, we model the class centers $\boldsymbol{M}_Q$ and $\boldsymbol{M}_G$ for the query domain and gallery domain as below:

$$\boldsymbol{M}^d = [\boldsymbol{m}_1^d, \cdots, \boldsymbol{m}_c^d], \quad (d \in \{Q, G\}). \tag{13}$$

ACCA models c class centers that are initialized with $\mathbf{0}$ for both domains. In the training process, the class centers are updated with clean and weighted relabeled features and the concrete terms are as follows:

$$\begin{cases} \boldsymbol{m}_t^d = \alpha \cdot \boldsymbol{m}_t^d + (1-\alpha)\boldsymbol{f}(\boldsymbol{x}_i), & \widetilde{y}_i = t \bigwedge \boldsymbol{x}_i \in \boldsymbol{D}_{clean}, \\ \boldsymbol{m}_t^d = \alpha \cdot \boldsymbol{m}_t^d + (1-\alpha)\boldsymbol{f}(\boldsymbol{x}_i) \cdot W_i, & \hat{y}_i = t \bigwedge \boldsymbol{x}_i \in \boldsymbol{D}_{noisy}, \end{cases} \tag{14}$$

where $\boldsymbol{m}_t^d$ denotes the t-th class center within d domain. $\hat{y}_i$ means the index of the largest item of the soft label as the specific class and α is the hyperparameter. W_i is the weight assigned by the mechanism defined in Eq. (10). The corresponding domain loss function is defined as follows:

$$\mathcal{L}_{acca} = \frac{1}{c}\left(\sum_{j=1}^{c} \left|\boldsymbol{m}_j^Q - \boldsymbol{m}_j^G\right|\right), \tag{15}$$

where c represents the number of categories. Note that the class centers adopt a batch-by-batch update strategy rather than the epoch-by-epoch way used in Eq. (8). The reason is that frequent updates allow the model to be more sensitive to capturing variations in features, contributing to the optimization of domain class centers. Correspondingly, our domain loss is updated with the adjustment of these centers. Since this cumulative loss can constantly draw the class centers of different domains closer, we can stabilize the elimination of the domain gap.

Experiment

To evaluate the effectiveness and robustness of our RoDA in cross-domain image retrieval with noisy labels (NCIR), we conduct extensive experiments on three widely used benchmark datasets.

Experimental Setup

In the experiments, we utilize two models, both of which consist of a feature extractor and a classifier. Specifically, we take the ResNet-50 network pre-trained on ImageNet as the feature extractor and a fully connected layer as the classifier. We adopt the Adam optimizer and set the learning rate to 1e-5. For evaluation, we take the mean average precision (mAP) as the measurement metric and report the best and last mAP@ALL results among total epochs. The experiment utilized three datasets: Office-31 (Saenko et al. 2010), Office-Home (Venkateswara et al. 2017), Adaptiope (Ringwald and Stiefelhagen 2021). Due to space limitations, we have placed detailed descriptions of the datasets in the supplementary materials.

During the experiments, the compared methods include Cross Entropy (Kapur and Kesavan 1992), Generalized Cross Entropy (Zhang and Sabuncu 2018), MAE (Hashidume and Morikawa 2007), Co-teaching (Han et al. 2018), JoCoR (Wei et al. 2020), DivideMix (Li, Socher, and Hoi 2020), NCE (Li et al. 2022a), LongRemix (Cordeiro et al. 2023), DISC (Li et al. 2023), Robust-DivideMix (Zhang et al. 2024) and the fully supervised Cross Entropy (with clean labels), denoted as CE-C. For comprehensive evaluations, we utilize various ratios of label noise in our experiments. Specifically, 20%, 40%,

Method	mAP	Office-Home				Office-31				Adaptiope			
		20%	40%	60%	80%	20%	40%	60%	80%	20%	40%	60%	80%
CE-C	Best	55.04	55.04	55.04	55.04	88.51	88.51	88.51	88.51	51.64	51.64	51.64	51.64
	Last	53.39	53.39	53.39	53.39	88.17	88.17	88.17	88.17	48.41	48.41	48.41	48.41
CE	Best	41.05	35.06	28.17	22.13	72.52	64.64	56.70	49.40	35.24	30.74	25.17	18.14
	Last	30.32	20.38	12.34	8.20	68.54	48.64	34.36	24.04	25.43	15.42	9.81	6.10
MAE	Best	45.17	37.90	26.62	20.94	84.94	73.98	63.79	51.05	45.41	38.54	24.24	14.10
	Last	44.49	36.15	21.21	8.79	83.93	67.79	49.96	24.96	45.10	37.99	22.50	6.76
GCE	Best	54.48	47.92	37.59	25.36	86.36	77.80	65.49	51.76	55.25	52.34	46.57	27.13
	Last	53.66	46.87	34.39	16.91	85.81	73.36	55.80	31.58	53.91	51.63	45.96	26.52
Co-teaching	Best	51.09	43.00	31.14	21.71	81.20	78.88	66.33	54.25	**58.16**	50.05	33.12	22.80
	Last	50.84	41.47	23.55	12.19	80.89	77.70	61.32	44.14	**57.49**	48.88	25.80	10.86
Divide-Mix	Best	42.25	37.82	33.86	27.05	72.99	68.45	62.27	53.37	38.73	34.61	30.86	25.89
	Last	34.59	33.54	31.30	24.34	65.16	63.82	57.84	47.56	27.52	26.34	25.60	22.64
JoCoR	Best	54.90	46.94	33.76	23.71	84.18	78.35	64.46	51.71	48.99	41.60	31.14	21.30
	Last	54.47	44.41	25.63	13.74	83.01	74.95	55.21	35.28	48.23	38.91	23.86	12.45
NCE	Best	42.73	38.59	33.81	26.63	72.73	66.67	61.32	54.03	39.73	36.80	32.54	25.14
	Last	37.06	35.06	28.72	17.38	60.38	59.05	55.51	42.29	32.84	30.33	28.08	15.00
LongRemix	Best	36.76	35.68	33.29	28.25	70.01	69.89	64.64	57.47	25.63	24.26	22.74	18.58
	Last	32.99	32.21	30.02	25.53	65.89	66.59	62.92	53.97	22.24	21.13	19.63	16.59
DISC	Best	44.22	41.29	34.06	22.50	73.78	69.64	60.65	47.66	31.68	30.57	28.25	21.00
	Last	35.83	31.56	24.86	13.43	61.43	56.29	48.55	25.44	22.99	20.19	19.19	13.56
Robust-DivideMix	Best	42.55	38.27	35.21	26.65	73.01	67.78	63.54	52.60	39.84	36.76	33.40	28.10
	Last	35.52	34.33	33.76	25.45	67.59	64.22	57.70	49.05	29.02	28.21	26.49	24.36
RoDA	Best	**57.25**	**54.13**	**48.46**	**36.99**	**88.30**	**84.73**	**76.55**	**57.64**	55.61	**52.62**	**49.92**	**39.60**
	Last	**56.35**	**53.73**	**47.75**	**36.10**	**87.33**	**84.04**	**75.50**	**54.89**	54.51	**51.79**	**48.76**	**38.25**

Table 1: The average mAP@All retrieval performance for our RoDA and other compared methods on the Office-Home, Office31 and Adaptiope datasets under different noise ratios. The best performance results among all methods are in **bold**.

60%, and 80% are applied to all three datasets to simulate datasets under different corruptions. Besides, the training batch size and max epoch are set to 16 and 50 with all methods. The hyperparameters: β, γ, θ, α and λ are set to 2, 0.5, 0.7, 0.9 and 1.0 respectively.

Comparison with the State-of-the-Art

We evaluate the effectiveness of our RoDA with all baselines on three datasets by conducting extensive NCIR experiments using the same network architecture. In detail, we conduct cross-domain image retrieval tasks on all domains of datasets. Due to space limitations, we only present the average results of NCIR tasks on three datasets, which are shown in Table 1. According to the results, we could draw the following observations: 1) The label noise that exists in the query and gallery domain could significantly impair the retrieval performance. This adverse effect is proportional to the noise ratio. 2) All compared methods are much less effective in cross-domain scenarios that perform well in traditional single-domain tasks. RoDA successfully aligns the query and gallery domain through ACCA. As a result, it achieves better results, surpassing even the performance of cross entropy applied to noise-free clean data (CE-C) in some low-noise cases. 3) Unlike the robust loss function-based methods which perform poorly in severe noise situations, RoDA maintains stable performance with high noise rates. The reason is that our method selectively utilizes the samples for adaptive learning and mitigates the overfitting of noisy label information. 4) Compared to methods based on selection and semi-supervised framework, RoDA exhibits a superior performance. Because our method makes better use of the noisy data. RoDA achieves this by rectifying the noisy labels and assigning weights to these relabeled samples for training. 5) In the setting of 20% noise in the Adaptiope dataset, RoDA exhibits a slight underperformance compared to the Co-teaching. The reason is probably that category imbalance is pronounced after division given the dataset's extensive category range. Nevertheless, RoDA significantly outperforms in terms of both performance and robustness when dealing with higher levels of noise.

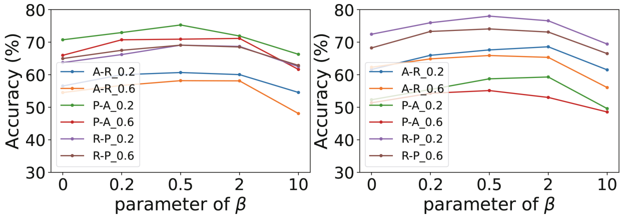
Ablation Study

We study the effect of removing different components to provide insight into what makes RoDA robust and effective for cross-domain image retrieval with noisy labels. We conduct experiments on twelve cross-domain tasks with the Office-Home dataset (Venkateswara et al. 2017) to evaluate the contribution of each part within our framework. Besides, we attach the results of training with only cross entropy for an intuitive comparison. The results are presented in Table 2. According to the table, we could reach the following conclusions: 1) RoDA without any component leads to the degradation of retrieval performance, showing the contribution of every part to our method. 2) Without AL and ACCA, the SD exhibits better retrieval performance than the CE-only case, which demonstrates that SD can effectively separate noisy label data from clean data, thus reducing the fitting of noisy label information. 3) Combining AL and SD achieves

$\mathcal{L}_{sdal}$		$\mathcal{L}_{acca}$	mAP	A⇒C	A⇒P	A⇒R	C⇒A	C⇒P	C⇒R	P⇒A	P⇒C	P⇒R	R⇒A	R⇒C	R⇒P
SD	AL	ACCA													
✓	✓	✓	Best	**49.47**	**55.45**	**53.09**	**43.00**	**61.44**	**55.53**	**52.89**	**64.78**	**70.78**	**51.42**	**58.05**	**71.34**
			Last	**49.21**	**53.58**	**52.90**	**41.44**	**59.75**	**53.77**	**50.86**	**64.76**	**69.91**	**50.89**	**57.55**	**71.34**
✓	✓		Best	47.97	53.11	51.24	42.18	59.52	54.76	52.42	63.14	69.51	50.60	56.47	69.72
			Last	46.42	51.17	50.13	40.45	59.39	53.57	51.61	63.14	69.50	49.91	56.07	69.72
✓		✓	Best	46.87	51.70	49.86	39.60	57.44	53.08	51.56	62.17	67.80	48.50	55.36	66.10
			Last	45.54	50.95	49.02	37.44	57.44	51.16	49.27	61.68	67.01	47.34	55.36	66.10
✓			Best	44.39	50.77	49.60	39.47	56.84	51.58	50.28	60.22	67.03	48.33	54.21	65.43
			Last	44.00	49.35	47.84	36.97	54.83	50.83	48.97	60.22	66.23	45.92	53.01	64.43
CE only			Best	32.99	38.96	43.71	31.80	39.29	39.68	37.99	40.59	53.46	41.05	41.34	52.29
			Last	28.58	31.91	31.54	23.15	30.24	31.25	32.19	34.47	41.02	33.74	33.74	40.62

Table 2: Ablation studies for RoDA on the Office-Home datasets under noise ratio 20%.

better results compared to using SD only, which shows that the adaptive learning of AL for the noisy part of the data can successfully extract useful feature information. 4) Utilizing ACCA on top of SD can improve retrieval effectiveness compared with SD only, indicating that ACCA effectively narrows the domain gap by aligning the domain class centers. In conclusion, the ablation study illustrates the effectiveness of each part of RoDA, demonstrating the robustness provided by each part in addressing NCIR.



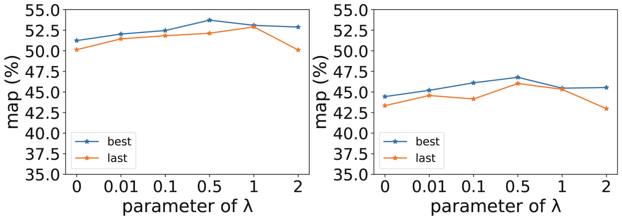
(a) Query domain

(b) Gallery domain

Figure 3: The relabel accuracy varies with the parameter β on Office-Home dataset. (a) and (b) shows the relationship between β and relabeled performance in the query domain and gallery domain with three cross-domain tasks (A⇒R, P⇒A, R⇒P) under different noise ratios (20%, 60%), respectively.

Parameter Analysis

To evaluate the robustness of our RoDA to different hyperparameter settings, namely the β and λ, we demonstrate the relabel accuracy and retrieval performance for parameter analysis on the Office-Home dataset. Considering the generalization of analysis, we set both low (ratio=0.2) and high (ratio=0.6) label noise level environments. The experimental results are shown in Figure 3 and Figure 4. From the figures, we can observe that our weighting mechanism $\boldsymbol{W}$ is robust to the selection of the hyperparameter β, *i.e.*, the RoDA achieves optimal performance when β is from 0.2 to 2. Besides, RoDA is also robust to the λ, which maintains retrieval stability when λ is within the range of 0.01 to 2.



(a) 20% noise

(b) 60% noise

Figure 4: The mAP varies with the parameter λ on Office-Home dataset. (a) shows the relationship between λ and retrieval performance in A⇒R under 20% noise. (b) depicts corresponding performance in A⇒R under 60% noise.

To further demonstrating the effectiveness of our proposed method, we conduct additional visual experiments. Due to space constraints, the detailed results and visualizations are provided in the supplementary materials. We encourage readers to refer to these materials for a comprehensive understanding of the experimental outcomes and insights.

Conclusion

This paper introduces and addresses the challenging problem of cross-domain image retrieval with noisy labels (NCIR), a scenario where both query and gallery domains suffer from label noise. To solve this challenge, we propose a novel method called RoDA, which includes two key parts: Selective Division and Adaptively Learning (SDAL) and Accumulative Class Center Alignment (ACCA). To be exact, SDAL utilizes division, relabeling, and weighting strategies to alleviate the adverse effects of label noise and ensures the robust learning of effective representation knowledge within domains. Meanwhile, ACCA utilizes an accumulative domain loss based on class centers to continuously eliminate the domain gap. Through extensive experiments on Office-Home, Office-31, and Adaptiope datasets, we show that RoDA can effectively tackle the issue of NCIR, achieving superior retrieval performance.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (62306197, 62372315), National Key Project (GJXM92579), China Postdoctoral Science Foundation (2021TQ0223, 2022M712236), Sichuan Science and Technology Planning Project (2025ZNSFSC1507, 2024YFG0007, 2024NSFTD0049, 2024ZDZX0004), Chengdu Science and Technology Project (2024-YF05-00687-SN, 2023-XT00-00004-GX, 2021-JB00-00025-GX), Postdoctoral Joint Training Program of Sichuan University (SCDXLHPY2307).

References

Arpit, D.; Jastrzebski, S.; Ballas, N.; Krueger, D.; Bengio, E.; Kanwal, M. S.; Maharaj, T.; Fischer, A.; Courville, A.; Bengio, Y.; et al. 2017. A closer look at memorization in deep networks. In *International conference on machine learning*, 233–242. PMLR.

Bai, Y.; Yang, E.; Han, B.; Yang, Y.; Li, J.; Mao, Y.; Niu, G.; and Liu, T. 2021. Understanding and improving early stopping for learning with noisy labels. *Advances in Neural Information Processing Systems*, 34: 24392–24403.

Chen, Y.-C.; Lin, Y.-Y.; Yang, M.-H.; and Huang, J.-B. 2019. Crdoco: Pixel-level domain transfer with cross-domain consistency. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 1791–1800.

Cordeiro, F. R.; Sachdeva, R.; Belagiannis, V.; Reid, I.; and Carneiro, G. 2023. Longremix: Robust learning with high confidence samples in a noisy label environment. *Pattern recognition*, 133: 109013.

Ding, Y.; Wang, L.; Fan, D.; and Gong, B. 2018. A semi-supervised two-stage approach to learning from noisy labels. In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 1215–1224. IEEE.

Feng, C.; Tzimiropoulos, G.; and Patras, I. 2021. Ssr: An efficient and robust framework for learning with unknown label noise. *arXiv preprint arXiv:2111.11288*.

Feng, Y.; Zhu, H.; Peng, D.; Peng, X.; and Hu, P. 2023a. ROAD: Robust Unsupervised Domain Adaptation with Noisy Labels. In *Proceedings of the 31st ACM International Conference on Multimedia*, 7264–7273.

Feng, Y.; Zhu, H.; Peng, D.; Peng, X.; and Hu, P. 2023b. RONO: Robust Discriminative Learning with Noisy Labels for 2D-3D Cross-Modal Retrieval. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11610–11619.

Gajic, B.; and Baldrich, R. 2018. Cross-domain fashion image retrieval. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 1869–1871.

Guo, L.; Liu, J.; Wang, Y.; Luo, Z.; Wen, W.; and Lu, H. 2017. Sketch-based image retrieval using generative adversarial networks. In *Proceedings of the 25th ACM international conference on Multimedia*, 1267–1268.

Han, B.; Yao, Q.; Yu, X.; Niu, G.; Xu, M.; Hu, W.; Tsang, I.; and Sugiyama, M. 2018. Co-teaching: Robust training of deep neural networks with extremely noisy labels. *Advances in neural information processing systems*, 31.

Hashidume, Y.; and Morikawa, Y. 2007. Lossless image coding based on minimum mean absolute error predictors. In *SICE Annual Conference 2007*, 2832–2836. IEEE.

Hu, C.; Zhang, C.; and Lee, G. H. 2023. Unsupervised Feature Representation Learning for Domain-generalized Cross-domain Image Retrieval. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 11016–11025.

Huang, J.; Feris, R. S.; Chen, Q.; and Yan, S. 2015. Cross-domain image retrieval with a dual attribute-aware ranking network. In *Proceedings of the IEEE international conference on computer vision*, 1062–1070.

Huang, Z.; Wang, H.; Xing, E. P.; and Huang, D. 2020. Self-challenging improves cross-domain generalization. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part II 16*, 124–140. Springer.

Iijima, L.; and Stathaki, T. 2024. A Multimodal Approach for Cross-Domain Image Retrieval. *arXiv preprint arXiv:2403.15152*.

Jiang, L.; Huang, D.; Liu, M.; and Yang, W. 2020. Beyond synthetic noise: Deep learning on controlled noisy labels. In *International conference on machine learning*, 4804–4815. PMLR.

Kapur, J. N.; and Kesavan, H. K. 1992. Entropy optimization principles and their applications. In *Entropy and energy dissipation in water resources*, 3–20. Springer.

Kim, Y.; Yim, J.; Yun, J.; and Kim, J. 2019. Nlnl: Negative learning for noisy labels. In *Proceedings of the IEEE/CVF international conference on computer vision*, 101–110.

Li, B.; Shi, Y.; Yu, Q.; and Wang, J. 2024. Unsupervised Cross-Domain Image Retrieval via Prototypical Optimal Transport. *arXiv preprint arXiv:2402.18411*.

Li, J.; Li, G.; Liu, F.; and Yu, Y. 2022a. Neighborhood Collective Estimation for Noisy Label Identification and Correction. In *European Conference on Computer Vision*, 128–145. Springer.

Li, J.; Socher, R.; and Hoi, S. C. 2020. Dividemix: Learning with noisy labels as semi-supervised learning. *arXiv preprint arXiv:2002.07394*.

Li, J.; Wong, Y.; Zhao, Q.; and Kankanhalli, M. S. 2019. Learning to learn from noisy labeled data. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 5051–5059.

Li, S.; Xue, Y.; Wang, Z.; and Zhou, G. 2013. Active learning for cross-domain sentiment classification. In *Twenty-Third International Joint Conference on Artificial Intelligence*. Citeseer.

Li, Y.; Han, H.; Shan, S.; and Chen, X. 2023. Disc: Learning from noisy labels via dynamic instance-specific selection and correction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 24070–24079.

Li, Y.; Yang, J.; Song, Y.; Cao, L.; Luo, J.; and Li, L.-J. 2017. Learning from noisy labels with distillation. In *Proceedings* *of the IEEE international conference on computer vision*, 1910–1918.

Li, Y.-J.; Dai, X.; Ma, C.-Y.; Liu, Y.-C.; Chen, K.; Wu, B.; He, Z.; Kitani, K.; and Vajda, P. 2022b. Cross-domain adaptive teacher for object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 7581–7590.

Lin, L.; Wang, G.; Zuo, W.; Feng, X.; and Zhang, L. 2016. Cross-domain visual matching via generalized similarity measure and feature learning. *IEEE transactions on pattern analysis and machine intelligence*, 39(6): 1089–1102.

Liu, H.; Ma, Y.; Yan, M.; Chen, Y.; Peng, D.; and Wang, X. 2024. DiDA: Disambiguated Domain Alignment for Cross-Domain Retrieval with Partial Labels. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 3612–3620.

Liu, T.; and Tao, D. 2015. Classification with noisy labels by importance reweighting. *IEEE Transactions on pattern analysis and machine intelligence*, 38(3): 447–461.

Liu, Y.-P.; Xu, N.; Zhang, Y.; and Geng, X. 2021. Label distribution for learning with noisy labels. In *Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence*, 2568–2574.

Natarajan, N.; Dhillon, I. S.; Ravikumar, P. K.; and Tewari, A. 2013. Learning with noisy labels. *Advances in neural information processing systems*, 26.

Nishi, K.; Ding, Y.; Rich, A.; and Hollerer, T. 2021. Augmentation strategies for learning with noisy labels. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 8022–8031.

Ringwald, T.; and Stiefelhagen, R. 2021. Adaptiope: A modern benchmark for unsupervised domain adaptation. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, 101–110.

Saenko, K.; Kulis, B.; Fritz, M.; and Darrell, T. 2010. Adapting visual category models to new domains. In *Computer Vision–ECCV 2010: 11th European Conference on Computer Vision, Heraklion, Crete, Greece, September 5-11, 2010, Proceedings, Part IV 11*, 213–226. Springer.

Taigman, Y.; Polyak, A.; and Wolf, L. 2016. Unsupervised cross-domain image generation. *arXiv preprint arXiv:1611.02200*.

Taylor, M. E.; and Stone, P. 2007. Cross-domain transfer for reinforcement learning. In *Proceedings of the 24th international conference on Machine learning*, 879–886.

Venkateswara, H.; Eusebio, J.; Chakraborty, S.; and Panchanathan, S. 2017. Deep hashing network for unsupervised domain adaptation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 5018–5027.

Wang, L.; Du, X.; and Zhu, Q. 2024. Semantic Feature Learning for Universal Unsupervised Cross-Domain Retrieval. *arXiv preprint arXiv:2403.05690*.

Wang, X.; Hu, P.; Liu, P.; and Peng, D. 2020a. Deep semisupervised class-and correlation-collapsed cross-view learning. *IEEE transactions on cybernetics*, 52(3): 1588–1601.

Wang, X.; Peng, D.; Hu, P.; Gong, Y.; and Chen, Y. 2023a. Cross-domain alignment for zero-shot sketch-based image retrieval. *IEEE Transactions on Circuits and Systems for Video Technology*, 33(11): 7024–7035.

Wang, X.; Peng, D.; Hu, P.; and Sang, Y. 2019. Adversarial correlated autoencoder for unsupervised multi-view representation learning. *Knowledge-Based Systems*, 168: 109–120.

Wang, X.; Peng, D.; Yan, M.; and Hu, P. 2023b. Correspondence-free domain alignment for unsupervised cross-domain image retrieval. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 10200–10208.

Wang, Z.; Jiang, J.; Han, B.; Feng, L.; An, B.; Niu, G.; and Long, G. 2020b. Seminll: A framework of noisy-label learning by semi-supervised learning. *arXiv preprint arXiv:2012.00925*.

Wei, H.; Feng, L.; Chen, X.; and An, B. 2020. Combating noisy labels by agreement: A joint training method with co-regularization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 13726–13735.

Xiao, T.; Xia, T.; Yang, Y.; Huang, C.; and Wang, X. 2015. Learning from massive noisy labeled data for image classification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2691–2699.

Xie, R.; Yu, F.; Wang, J.; Wang, Y.; and Zhang, L. 2019. Multi-level domain adaptive learning for cross-domain detection. In *Proceedings of the IEEE/CVF international conference on computer vision workshops*, 0–0.

Yang, X.; Zhang, T.; and Xu, C. 2014. Cross-domain feature learning in multimedia. *IEEE Transactions on Multimedia*, 17(1): 64–78.

Yi, K.; and Wu, J. 2019. Probabilistic end-to-end noise correction for learning with noisy labels. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 7017–7025.

Yin, Z.; Yan, J.; Xu, C.; and Deng, C. 2024. Asymmetric Mutual Alignment for Unsupervised Zero-Shot Sketch-Based Image Retrieval. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 16504–16512.

Zhang, J.; Song, B.; Wang, H.; Han, B.; Liu, T.; Liu, L.; and Sugiyama, M. 2024. Badlabel: A robust perspective on evaluating and enhancing label-noise learning. *IEEE transactions on pattern analysis and machine intelligence*.

Zhang, Z.; and Sabuncu, M. 2018. Generalized cross entropy loss for training deep neural networks with noisy labels. *Advances in neural information processing systems*, 31.

Zheng, S.; Wu, P.; Goswami, A.; Goswami, M.; Metaxas, D.; and Chen, C. 2020. Error-bounded correction of noisy labels. In *International Conference on Machine Learning*, 11447–11457. PMLR.

Zheng, Y. 2015. Methodologies for cross-domain data fusion: An overview. *IEEE transactions on big data*, 1(1): 16–34.

Zheng, Z.; Li, R.; and Liu, C. 2023. Learning robust features alignment for cross-domain medical image analysis. *Complex & Intelligent Systems*, 1–15.