MemREIN: Rein the Domain Shift for Cross-Domain Few-Shot Learning

Yi Xu[1] , **Lichen Wang**[1] , **Yizhou Wang**[1] , **Can Qin**[1] , **Yulun Zhang**[2] and **Yun Fu**[1]
[1]Northeastern University
[2]ETH Zürich
{xu.yi, wang.lich, wang.yizhou, qin.ca}@northeastern.edu, yulun100@gmail.com, yunfu@ece.neu.edu

Abstract

Few-shot learning aims to enable models generalize to new categories (query instances) with only limited labeled samples (support instances) from each category. Metric-based mechanism is a promising direction which compares feature embeddings via different metrics. However, it always fail to generalize to unseen domains due to the considerable domain gap challenge. In this paper, we propose a novel framework, MemREIN, which considers Memorized, Restitution, and Instance Normalization for cross-domain few-shot learning. Specifically, an instance normalization algorithm is explored to alleviate feature dissimilarity, which provides the initial model generalization ability. However, naively normalizing the feature would lose fine-grained discriminative knowledge between different classes. To this end, a memorized module is further proposed to separate the most refined knowledge and remember it. Then, a restitution module is utilized to restitute the discrimination ability from the learned knowledge. A novel reverse contrastive learning strategy is proposed to stabilize the distillation process. Extensive experiments on five popular benchmark datasets demonstrate that MemREIN well addresses the domain shift challenge, and significantly improves the performance up to 16.43% compared with state-of-the-art baselines.

1 Introduction

In recent years, machine learning especially deep learning methods have made amazing achievements in the field of computer vision, image classification, semantic segmentation, etc. However, the promising performance heavily counts on the large amount of well-labeled training data, which provides comprehensive and diverse samples to cover all corner cases. Such a huge scale makes it difficult in real practice, thus leads to a new topic of few-shot learning [Wang *et al.*, 2020]. Few-shot learning aims to enable models generalize to new categories (query instances) with only limited labeled samples (support instances) from each category.

Among existing few-shot learning methods, metric-based methods have attracted significant attention because of their effectiveness and intelligibility. In general, the core idea of this kind of methods is to make classifications based on the similarity between the query images and the support images via proposed similarity measures. It usually consists of two main components: (1) feature encoder and (2) metric function. Given a task with few labeled images (support set) and unlabeled images (query set), the visual features are firstly extracted via the feature encoder and then passed through the defined metric function to determine the categories of the query images. The underlying assumption is that both training and testing are from the same dataset, namely the same domain. While, when it comes to different domains, the generalization ability of the metric-based methods greatly decreases [Chen *et al.*, 2019; Tseng *et al.*, 2020]. However, such ability to generalize to unseen domains is of great importance in practice, e.g., expensive human annotation or time-consuming data collection. As a result, considering the domain shift scenario within the few-shot learning has become an important yet challenging task.

Various unsupervised domain adaptation methods have been proposed [Yang *et al.*, 2018; Xu *et al.*, 2022] recently. These methods aim to minimize the domain gap either by learning domain-invariant representations via representation learning, projection learning, or adversarial strategies [Long *et al.*, 2015; Kumar *et al.*, 2018; Tzeng *et al.*, 2017; Kundu *et al.*, 2019]. However, these methods assume that the complete unlabeled samples from the target domain are accessible while training. We argue that this assumption may not hold in real situations, and it could leads to high computational cost in testing phase. Domain shift problem could be addressed by various domain generalization methods [Blanchard *et al.*, 2011; Muandet *et al.*, 2013]. However, these methods assume that the source and target domains share the same categories. In contrast, our goal is to recognize novel categories from the target domain with only a few (e.g., 1 or 5) of samples selected from novel categories.

As argued above, there are two main challenges in cross-domain few-shot learning task. (1) How to minimize the discrepancy between the source and target domain. (2) How to recognize novel/unseen classes with only limited samples.

To this end, we propose a novel MemREIN approach, which includes Memorized, Restitution, and Instance

Normalization as crucial modules, to "rein" the domain shift level in few-shot scenario. The core idea of MemREIN is to enhance the generalization ability while still be able to balance the discrimination ability for subsequent classification. Specifically, on the training stage, we first present an instance normalization layer operating on features with respect to samples at the channel level. This operation aims to reserve spatial feature dependency and meanwhile remove the image-specific features, i.e., alleviate the discrepancy of these training samples. In this way, the generalization ability across different samples is enhanced. Then, the filtered out features are extracted from a residual structure. Normally, the filtered out features are considered as useless feature which could be discarded. However, we consider it still contains fine-grained distinctive knowledge which could be "remembered" and "restituted". To this end, we manage to adaptively distill the long-term discriminative information from them via our proposed novel memorized approach. Then, such discriminative information is restituted to the above refined features to maintain the discrimination ability for subsequent classification. A novel reverse contrastive loss constraint in the restitution phase is designed to encourage the better separation of discriminative features and general features, which ensures the distillation process. Contributions of our work are as,

- A novel memorized and restitution strategy is proposed for discriminative information distillation. It is able to distill the long-term discriminative information from filtered out features to maintain the discrimination ability of original features for better classification.
- An instance normalization strategy is adopted to alleviate the the discrepancy across training samples, which reduces the sample-specific features and greatly enhances the overall generalization ability across features.
- A novel reverse contrastive loss is proposed to encourage the better separation of discriminative and general features, which is able to ensure the distillation process.

Our MemREIN approach is simple yet effective. It is a universal and method-agnostic approach that can be applied to various existing metric-based methods for enhancing their generalization ability to unseen domains. Extensive experiments demonstrate the effectiveness of MemREIN, which achieves consistent superior performance than existing state-of-the-art methods under the cross-domain setting.

2 Related Work

2.1 Cross-domain Few-shot Learning

Few-shot classification aims to recognize novel classes with a limited amount of labeled samples. Among these existing methods, metric-based methods have attracted significant attention and achieved promising performance. For instance, MatchingNet [Vinyals *et al.*, 2016] utilizes cosine similarity with an attention Bi-LSTM for classification and ProtoNet [Snell *et al.*, 2017] applies Euclidean distance for classification. RelationNet [Sung *et al.*, 2018] uses convolutional neural networks and GNN [Satorras and Estrach, 2018] uses the graph convolutional framework as the metric function. Although these methods are able to achieve promising performance, they always fail to generalize to unseen domains since the distributions among different domains have huge shifts. Recent work [Chen *et al.*, 2019] reveals that the accuracy of existing few-shot learning methods degrades significantly under the cross-domain setting. The motivation of our work aims to enhance the generalization ability of metric-based few-shot learning methods so that these methods can better generalize to unseen domains.

Recently, promoted by the pioneer work [Chen *et al.*, 2019], cross-domain few-shot learning problem has attracted many attentions. As an emerging task, work [Chen *et al.*, 2019] carried out a broader study and introduced a new benchmark. Some methods [Tseng *et al.*, 2020; Sun *et al.*, 2021; Phoo and Hariharan, 2020; Zou *et al.*, 2021; Islam *et al.*, 2021; Liang *et al.*, 2021; Fu *et al.*, 2021; Wang and Deng, 2021] have been proposed recently and achieved promising performance under this benchmark, which greatly promoted the development. In our work, we propose a simple yet effective method from the perspective of feature level, which is a universal and method-agnostic approach.

3 Method

3.1 Preliminaries

In the few-shot classification problem, a task T is characterized as N_w way and N_s shot, which represents the number of categories and the number of labeled samples in each category. At each iteration, the metric-based few-shot learning method randomly samples N_w categories as a task T, and then constructs a support set $S = \{(\mathcal{X}_s, \mathcal{Y}_s)\}$ and a query set $Q = \{(\mathcal{X}_q, \mathcal{Y}_q)\}$, where $\mathcal{X}$ and $\mathcal{Y}$ represent samples and labels respectively. These two sets are constructed by randomly selecting N_s and N_q samples for each of the N_w categories.

Once the data is prepared, the feature encoder E first extracts features of the samples from both support set S and query set Q. Then, the defined metric function M predicts the query samples $\mathcal{X}_q$ based on three parts: the label of support samples $\mathcal{Y}_s$, encoded query image $E(\mathcal{X}_q)$, and the encoded support images $E(\mathcal{X}_s)$, which is formulated as,

$$\hat{\mathcal{Y}}_q = M(\mathcal{Y}_s, E(\mathcal{X}_q), E(\mathcal{X}_s)). \tag{1}$$

After all, the objective of the metric-based few-shot learning method is the classification loss of the samples in the query set, which is formulated as,

$$\mathcal{L} = \mathcal{L}_{cls}(\mathcal{Y}_q, \hat{\mathcal{Y}}_q). \tag{2}$$

In this paper, we tackle the cross-domain few-shot classification problem. Given a set of few-shot classification tasks $\mathcal{T} = \{T_1, T_2, ..., T_n\}$ as a domain (dataset). At the training stage, given N accessible domains $\{\mathcal{T}_1^{seen}, \mathcal{T}_2^{seen}, ..., \mathcal{T}_N^{seen}\}$, we aim to learn a metric-based few-shot learning model with these seen domains, then the model can generalize to an unseen domain$\mathcal{T}^{unseen}$.

3.2 MemREIN Method

The core idea of our MemREIN method is to enhance the generalization ability, including the ability to balance the discrimination of metric-based few-shot learning methods, and

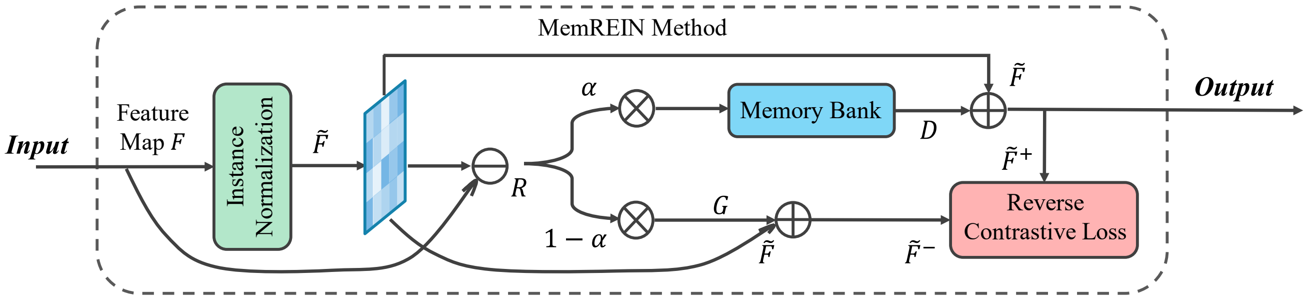


Figure 1: Framework of our MemREIN method, where ⊕, ⊖, and ⊗ denote element-wise addition, subtraction, and multiplication, respectively. With instance normalization approach, the sample-specific features of F can be reduced, and then with memorized and restitution approach, the long-term discriminative information can be distilled and restituted to refined features.

achieve promising performance on arbitrary unseen domains. The overall framework is illustrated in Figure 1.

Instance Normalization

As argued above, images with the same category from different domains have large discrepancies in many aspects e.g. , image style, color, quality. Generally speaking, the discrepancy between the source domain and the target domain hinders the generalization ability of the model to some extent.

To this end, we reduce the discrepancy cross samples by instance normalization in our proposed MemREIN method. Given the input feature map $F \in \mathbb{R}^{c\times h\times w}$ and the output feature map $\widetilde{F} \in \mathbb{R}^{c\times h\times w}$, where c, h, and w represent the number of channel, height, and width, respectively. The instance normalization is formulated as,

$$\widetilde{F} = \mathrm{IN}(F) = \gamma\left(\frac{F-\mu(F)}{\sigma(F)}\right)+\beta, \tag{3}$$

where $\mu(\cdot)$ and $\sigma(\cdot)$ denote the mean and standard deviation calculated at the channel level for each sample, $\gamma \in \mathbb{R}^c$ and $\beta \in \mathbb{R}^c$ are two trainable parameters. Instance normalization was originally used in style transfer [Dumoulin *et al.*, 2016], which is helpful to enhance the generalization ability by reducing the feature dissimilarity. It can remove instance/sample specific features out of the input, which makes more general features remained.

However, instance normalization inevitably removes some discriminative information from original feature maps [Jin *et al.*, 2020], which weakens the extracted features discrimination ability of extracted features. To address this emerging problem, we propose a memorized restitution approach to distill the discriminative information from the filtered out features and then restitute it as the final output feature maps.

Memorized Restitution

As discussed above, in order to maintain the discrimination ability of the refined features, we propose a following memorized restitution approach to distill discriminative information. We first obtain the filtered out feature R via a residual structure, which is defined as,

$$R = F - \widetilde{F}, \tag{4}$$

where $R \in \mathbb{R}^{c\times h\times w}$, denoting the features that we have filtered out via the instance normalization operation. Since instance normalization operation will inevitably remove discriminative information from the original features, we need to distill and purify discriminative features from the residual feature R to maintain the discrimination ability.

At the training stage, given the feature map R at each iteration (we omit the subscript of feature map R for brevity), we assume R consists of two parts: $D \in \mathbb{R}^{c\times h\times w}$ with relatively more discriminative information, and $G \in \mathbb{R}^{c\times h\times w}$ with relatively more general information, which is defined as,

$$\begin{cases} D(k,:,:) = \theta_k R(k,:,:), \\ G(k,:,:) = (1-\theta_k)R(k,:,:), \end{cases} \tag{5}$$

where k represents the k^{th} channel, θ_k denotes the learnable parameters to split the residual feature map R into D and G, which are channel-wise mutually exclusive.

Then, we employ the SE-like channel attention [Hu *et al.*, 2018] to define an attention vector $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, ..., \alpha_c]$ as,

$$\boldsymbol{\alpha} = \delta(W_2\eta(W_1 avepooling(R))), \tag{6}$$

where $avepooling$ represents the average pooling layer, W_1 and W_2 are parameters to be learned, δ and η are the ReLU and Sigmoid activation function, respectively.

Since there are limited labeled samples under the few-shot learning framework, it is highly possible that the model would overfit. Thus we further propose a memorized mechanism with a memory vector $M^{(l)} \in \mathbb{R}^c$ to store the long-term feature maps D, which is defined as,

$$\begin{aligned} &M^{(l)} = [M_1^{(l)}, ..., M_k^{(l)}, ..., M_c^{(l)}], \\ &M_k^{(l+1)} = D^{(l)}(k,:,:), \end{aligned} \tag{7}$$

where $M_k^{(l)} \in \mathbb{R}^{h\times w}$, (l) represents the l^{th} iteration, k denotes the k^{th} channel. At the l^{th} iteration, we concatenate the feature map D to the memory bank at the channel level, and update D as,

$$D(k,:,:) = maxpooling(concat(M_k^{(l)}, D(k,:,:))), \tag{8}$$

where $concat$ represents the concatenation operation, $maxpooling$ represents the max pooling layer.

Once we obtain the updated feature map D, we restitute it to refined feature $\widetilde{F}$ as the final output $\widetilde{F}^+$ of our proposed

MemREIN method, and we also restitute the relatively unimportant feature map G with feature $\widetilde{F}$ as the "contaminated" feature $\widetilde{F}^-$ for following loss optimization as,

$$\widetilde{F}^+ = \widetilde{F} + D, \quad \widetilde{F}^- = \widetilde{F} + G. \tag{9}$$

Reverse Contrastive Loss

Apart from the conventional cross-entropy loss defined in Equation 2, we also propose a novel reverse contrastive loss $\mathcal{L}_{rcl}$ to promote the disentanglement of feature D and feature G. Instead of directly separating feature D and G, we first combine them with $\widetilde{F}$ defined in Equation 9 for better disentanglement. The reverse contrastive loss consists of two parts: $\mathcal{L}_{rcl}^+$ and $\mathcal{L}_{rcl}^-$, e.g. , $\mathcal{L}_{rcl} = \mathcal{L}_{rcl}^+ + \mathcal{L}_{rcl}^-$. Given a mini-batch $\mathcal{X}_b = \{\mathcal{X}_1, ..., \mathcal{X}_N\}$ contains N samples at the training phase, we first randomly select one anchor sample referred as $\mathcal{X}_a$, and then we denote samples with the same category as the positive samples $\mathcal{X}_{pos}$, samples with different categories as the negative samples $\mathcal{X}_{neg}$. Note that the corresponding features of these samples are denoted with their subscripts such as $\widetilde{F}_a$, $\widetilde{F}_{pos}$, and $\widetilde{F}_{neg}$ in the following paragraphs.

We first reshape features $\widetilde{F}^+$ and $\widetilde{F}^-$ to the size of $\mathbb{R}^{chw\times 1}$ and then pass them through one fully-connected layer following the $softmax$ function to obtain the feature vectors $\widetilde{f}^+$ and $\widetilde{f}^-$. Note that these two vectors have the same size of $\in \mathbb{R}^{K\times 1}$. It is formulated as,

$$\widetilde{f}^+ = softmax\left(W^+ reshape(\widetilde{F}^+)\right), \tag{10}$$

$$\widetilde{f}^- = softmax\left(W^- reshape(\widetilde{F}^-)\right), \tag{11}$$

where W^+ and W^- are trainable parameters with the same size of $\mathbb{R}^{K\times chw}$, K is the number of classes. Then, our proposed reverse contrastive loss is defined as,

$$\mathcal{L}_{rcl}^+ = -\mathbb{E}\left[\log \frac{\exp(\widetilde{f}_a^{+\top} \widetilde{f}_{pos}^+)}{\sum_{\mathcal{X}_{pos}\in\mathcal{X}} \exp(\widetilde{f}_a^{+\top} \widetilde{f}_{neg}^+)}\right], \tag{12}$$

$$\mathcal{L}_{rcl}^- = -\mathbb{E}\left[\log \frac{\sum_{\mathcal{X}_{neg}\in\mathcal{X}} \exp(\widetilde{f}_a^{-\top} \widetilde{f}_{neg}^-)}{\exp(\widetilde{f}_a^{-\top} \widetilde{f}_{pos}^-)}\right]. \tag{13}$$

The goal of our proposed reverse contrastive loss is to promote the disentanglement of feature $\widetilde{F}^+$ and feature $\widetilde{F}^-$, where feature $\widetilde{F}^+$ contains more discriminative information and $\widetilde{F}^-$ contains more general information. Since feature $\widetilde{F}^+$ has better discrimination ability, sample features with the same category are closer and those with different identities are farther apart. Therefore, we propose $\mathcal{L}_{rcl}^+$ to encourage the features of positive samples $\widetilde{f}_{pos}^+$ to gather closer and separate the features of negative samples $\widetilde{f}_{neg}^+$ from the anchor feature. On the other hand, feature $\widetilde{F}^-$ is more general which is not capable of distinguishing samples with the same category. Therefore, we propose $\mathcal{L}_{rcl}^-$ to separate the the features of positive samples $\widetilde{f}_{pos}^-$ from both negative sample features $\widetilde{f}_{neg}^-$ and the anchor feature $\widetilde{f}_a^-$. The whole loss is defined as,

$$\mathcal{L} = \mathcal{L}_{cls} + \lambda(\mathcal{L}_{rcl}^+ + \mathcal{L}_{rcl}^-), \tag{14}$$

where λ is a hyper-parameter to control the balance of these two terms in the training phase.

4 Experiments

4.1 Experimental Setup

Baselines. Extensive experiments are made on three existing metric-based few-shot learning methods: MatchingNet [Vinyals *et al.*, 2016], RelationNet [Sung *et al.*, 2018], and GNN [Satorras and Estrach, 2018]. We compare our proposed method with following existing cross-domain few-shot learning methods: FT [Tseng *et al.*, 2020], LRP [Sun *et al.*, 2021], and ATA [Wang and Deng, 2021] to demonstrate the advantages of our proposed method.

Datesets. Five widely used datasets are used: mini-ImageNet [Ravi and Larochelle, 2017], CUB [Wah *et al.*, 2011], Cars [Krause *et al.*, 2013], Places [Zhou *et al.*, 2017], and Plantae [Van Horn *et al.*, 2018].

Setting. We take the same leave-one-out setting which is applied in other baselines. Specifically, we select one dataset among CUB, Cars, Places, and Plantae as the target domain for testing, and using the remaining three datasets along with dataset mini-ImageNet as the source domains for training.

Implementation details. We adopt the ResNet-10 [He *et al.*, 2016] as the backbone network for our feature encoder E. We insert our proposed MemREIN method after the last batch normalization layer of all the residual blocks in the feature encoder E at the training stage. Instead of optimizing from the scratch, we apply a strategy that pre-trains the feature extractor by minimizing the standard cross-entropy classification loss on the 64 training categories from the dataset mini-ImageNet and this strategy is also applied in all the baselines. In the training phase, we set $\lambda = 0.1$ and train 1000 trials for all the methods. In each trial, we randomly sample N_w categories with N_s randomly selected images for each support set, and 16 images for the query set. We use the Adam optimizer with the learning rate 0.001.

4.2 Experimental Results

Quantitative Results

Table 1 shows the results under the leave-one-out setting. We first select out one dataset as the unseen domain for testing and use the remaining three datasets as well as the dataset mini-ImageNet for training since we already use the dataset mini-ImageNet for pre-training. Note that the baseline [Tseng *et al.*, 2020] has two different training strategies, one is the "learn to learn" strategy and another is using fixed hyper-parameters. We consider the better results for comparison here, which is denoted as "+LFT" in the Table 1. The results demonstrate that our proposed MemREIN method can greatly improve the performance of all three metric-based few-shot learning methods, which reflects that our method has the capability of mitigating the domain gap problem. In addition, results show that our method consistently outperforms the "+LFT" method, which validates that our proposed

5-way 1-shot	Classification Accuracy (%)			
	CUB	Cars	Places	Plantae
MNet [Vinyals *et al.*, 2016]	37.90 ± 0.55%	28.96 ± 0.45%	49.01 ± 0.65%	33.21 ± 0.51%
MNet+LFT [Tseng *et al.*, 2020]	43.29 ± 0.59%	30.62 ± 0.48%	52.51 ± 0.67%	35.12 ± 0.54%
MNet+MemREIN (Ours)	**46.37 ± 0.50%**	**35.65 ± 0.45%**	**54.92 ± 0.64%**	**38.82 ± 0.48%**
RNet [Sung *et al.*, 2018]	44.33 ± 0.59%	29.53 ± 0.45%	47.76 ± 0.63%	33.76 ± 0.52%
RNet+LFT [Tseng *et al.*, 2020]	48.38 ± 0.63%	32.21 ± 0.51%	50.74 ± 0.66%	35.00 ± 0.52%
RNet+MemREIN (Ours)	**52.02 ± 0.52%**	**36.38 ± 0.38%**	**54.82 ± 0.57%**	**36.74 ± 0.45%**
GNN [Satorras and Estrach, 2018]	49.46 ± 0.73%	32.95 ± 0.56%	51.39 ± 0.80%	37.15 ± 0.60%
GNN+LFT [Tseng *et al.*, 2020]	51.51 ± 0.80%	34.12 ± 0.63%	56.31 ± 0.80%	42.09 ± 0.68%
GNN+MemREIN (Ours)	**54.26 ± 0.62%**	**37.55 ± 0.50%**	**59.98 ± 0.64%**	**45.69 ± 0.64%**
5-way 5-shot	Classification Accuracy (%)			
	CUB	Cars	Places	Plantae
MNet [Vinyals *et al.*, 2016]	51.92 ± 0.80%	39.87 ± 0.51%	61.82 ± 0.57%	47.29 ± 0.51%
MNet+LFT [Tseng *et al.*, 2020]	61.41 ± 0.57%	43.08 ± 0.55%	64.99 ± 0.59%	48.32 ± 0.57%
MNet+MemREIN (Ours)	**67.31 ± 0.51%**	**47.36 ± 0.48%**	**68.14 ± 0.58%**	**52.28 ± 0.52%**
RNet [Sung *et al.*, 2018]	62.13 ± 0.74%	40.64 ± 0.54%	64.34 ± 0.57%	46.29 ± 0.56%
RNet+LFT [Tseng *et al.*, 2020]	64.99 ± 0.54%	43.44 ± 0.59%	67.35 ± 0.54%	50.39 ± 0.52%
RNet+MemREIN (Ours)	**68.39 ± 0.48%**	**46.92 ± 0.50%**	**69.87 ± 0.54%**	**58.64 ± 0.50%**
GNN [Satorras and Estrach, 2018]	69.26 ± 0.68%	48.91 ± 0.67%	72.59 ± 0.67%	58.36 ± 0.68%
GNN+LFT [Tseng *et al.*, 2020]	73.11 ± 0.68%	49.88 ± 0.67%	77.05 ± 0.65%	58.84 ± 0.66%
GNN+MemREIN (Ours)	**77.54 ± 0.62%**	**56.78 ± 0.66%**	**78.84 ± 0.66%**	**65.44 ± 0.64%**

Table 1: Classification accuracy (%) of 5-way 1/5-shot tasks under the leave-one-out setting.

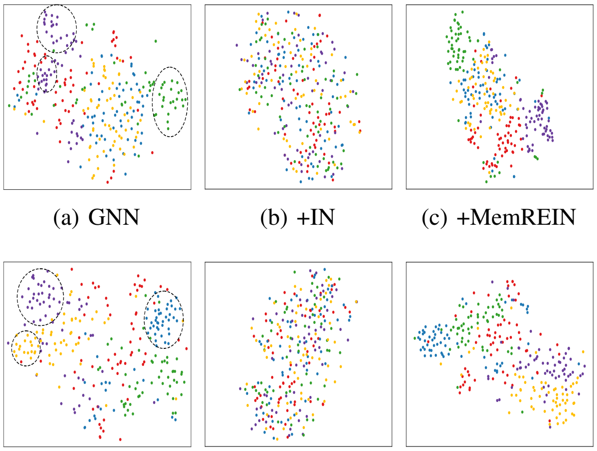
5-way 5-shot	Classification Accuracy (%)	
GNN+MemREIN	CUB	Cars
$\lambda = 0.01$	77.02 ± 0.62%	56.12 ± 0.66%
$\lambda = 0.1$	**77.54 ± 0.62%**	**56.78 ± 0.66%**
$\lambda = 0.5$	77.34 ± 0.62%	56.66 ± 0.66%
$\lambda = 1$	76.78 ± 0.64%	56.22 ± 0.66%

Table 2: Performance study on the hyper-parameter λ.

method can better capture the variation of feature distributions across multiple domains than the "+LFT" method, thus the generalization ability of extracted features are better enhanced. In particular, for dataset Cars (5-way 1-shot), the accuracy of our MNet+MemREIN is 35.65 ± 0.45%, which outperforms the accuracy of baseline MNet+LFT by 16.43%.

Qualitative Results

As illustrated in Figure 2, we employ the t-SNE algorithm to visualize features that obtained by the feature encoder "before/within/after" our MemREIN method, where each color represents one class. We take the GNN baseline under the leave-one-out setting on the dataset CUB as the example. We randomly select 5 categories with 60 samples of each category in the testing spilt of the dataset CUB. The first column indicates two examples of the features from conventional GNN baseline, The second column indicates the features that only applied the instance normalization operation, and the third column indicates the features that applied our proposed MemREIN method. As shown in the first column, there exists several rough clusters but the boundaries are unclear. After instance normalization, the overall model generalization ability of features is enhanced. In comparison with the first column and the third column, the features learned by our method are more clustered and separable, which validates the effectiveness of our novel memorized restitution approach.


(a) GNN (b) +IN (c) +MemREIN

(d) GNN (e) +IN (f) +MemREIN

Figure 2: t-SNE visualization of features extracted by encoder.

5-way 5-shot	Classification Accuracy (%)				
Variant ID	Method	CUB	Cars	Places	Plantae
1	GNN [Satorras and Estrach, 2018]	$69.26 \pm 0.68\%$	$48.91 \pm 0.67\%$	$72.59 \pm 0.67\%$	$58.36 \pm 0.68\%$
2	GNN+IN	$67.34 \pm 0.66\%$	$42.76 \pm 0.75\%$	$67.82 \pm 0.73\%$	$54.04 \pm 0.69\%$
3	**GNN+MemREIN (Ours)**	$\mathbf{77.54 \pm 0.62\%}$	$\mathbf{56.78 \pm 0.66\%}$	$\mathbf{78.84 \pm 0.66\%}$	$\mathbf{65.44 \pm 0.64\%}$
4	w/o $\mathcal{L}_{rcl}^{-}$	$75.38 \pm 0.63\%$	$55.34 \pm 0.72\%$	$78.03 \pm 0.68\%$	$65.22 \pm 0.64\%$
5	w/o $\mathcal{L}_{rcl}^{+}$	$73.02 \pm 0.62\%$	$51.45 \pm 0.64\%$	$73.26 \pm 0.66\%$	$62.22 \pm 0.64\%$
6	GNN+MemREIN w/o MB	$75.98 \pm 0.62\%$	$54.64 \pm 0.66\%$	$74.86 \pm 0.68\%$	$64.08 \pm 0.68\%$
7	GNN+MemREIN (D&G)	$76.02 \pm 0.66\%$	$55.26 \pm 0.69\%$	$78.08 \pm 0.66\%$	$64.84 \pm 0.68\%$

Table 3: Ablation study on our method. "GNN+IN" indicates that we only employ the instance normalization strategy, "w/o $\mathcal{L}_{rcl}^{-}$" indicates that we remove the $\mathcal{L}_{rcl}^{-}$ term, and "w/o $\mathcal{L}_{rcl}^{+}$" indicates that we remove the $\mathcal{L}_{rcl}^{+}$ term, "GNN+MemREIN w/o MB" represents that we remove memory bank and directly use the feature map D, and "GNN+MemREIN (D&G)" represents that the memory bank is operated both on feature map D and G (not shared).

5-shot	Classification Accuracy (%)			
	2-way	5-way	10-way	20-way
MNet [Vinyals *et al.*, 2016]	$78.46 \pm 0.78\%$	$51.92 \pm 0.80\%$	$38.22 \pm 0.38\%$	$26.17 \pm 0.24\%$
MNet+LFT [Tseng *et al.*, 2020]	$83.88 \pm 0.72\%$	$61.41 \pm 0.57\%$	$45.69 \pm 0.39\%$	$32.81 \pm 0.23\%$
MNet+MemREIN (Ours)	$\mathbf{88.68 \pm 0.68\%}$	$\mathbf{67.31 \pm 0.51\%}$	$\mathbf{49.22 \pm 0.34\%}$	$\mathbf{33.99 \pm 0.22\%}$
RNet [Sung *et al.*, 2018]	$84.25 \pm 0.72\%$	$62.13 \pm 0.74\%$	$47.15 \pm 0.40\%$	$34.52 \pm 0.24\%$
RNet+LFT [Tseng *et al.*, 2020]	$85.44 \pm 0.72\%$	$64.99 \pm 0.54\%$	$49.90 \pm 0.40\%$	$37.20 \pm 0.25\%$
RNet+MemREIN (Ours)	$\mathbf{89.12 \pm 0.66\%}$	$\mathbf{68.39 \pm 0.48\%}$	$\mathbf{52.85 \pm 0.32\%}$	$\mathbf{42.82 \pm 0.20\%}$

Table 4: Classification Accuracy (%) of our proposed method with different N_w. We consider the CUB dataset as the unseen domain under the leave-one-out setting.

Performance Study of λ

We carry out performance study on the hyper-parameter λ. We take our method under the leave-one-out setting (5-way 5-shot) and dataset CUB and Cars as the example. We set four different values $\lambda = \{0.01, 0.1, 0.5, 1\}$ and the results are shown in Table 2. It can be observed that when setting $\lambda = 0.1$, it can achieve the best performance.

Ablation Study

We carry out ablation studies of different components in our proposed method. We compare with the GNN baseline under the leave-one-out setting (5-way 5-shot) and results are shown in Table 3. Comparing the results of Variant 1 and 2, it indicates that only applying the instance normalization operation results in the decrease of the accuracy. It is reasonable because the instance normalization operation will inevitably remove some discriminative useful information. In comparison with Variant 3 and 6, it validates the effectiveness of employing the memory bank on feature D. Comparing Variant 3, 6, and 7, it indicates that when employing memory bank on feature G, it would cause performance decrease. Empirically, when applying the memory bank on the feature D and directly using feature G, it can achieve the best performance.

Different Numbers of Ways

We consider a more practical situation that N_w may be different from that at the training stage. It also reflects the generalization ability of the model and results are shown in Table 4. Note that model GNN requires the number of ways to be the same while the training and testing, thus we evaluate with method MatchingNet and RelationNet (MNet and RNet for short). The model is trained on the datasets mini-ImageNet, Cars, Places, and Plantae and evaluated on the dataset CUB with different number of ways N_w. The results indicate that our proposed method are still capable of improving the generalization ability to the unseen domain with various numbers of ways. In addition, our proposed method consistently outperforms the baseline that has considered the domain-shift issue, which validates the superiority of our method.

5 Conclusion

In this paper, we investigated the cross-domain few-shot classification problem where exists the domain gap issue. We propose a novel framework, MemREIN, which considers Memorized, Restitution, and Instance Normalization to address this issue. We first alleviate feature dissimilarity across sample features via an instance normalization algorithm to enhance the overall generalization ability. In order to avoid the loss of fine-grained discriminative knowledge between different classes, a memorized restitution approach is further proposed to adaptively remember the long-term refined knowledge and restitute the discrimination ability. Finally, a novel reverse contrastive learning strategy is proposed to stabilize the distillation process. Extensive experiments demonstrate that MemREIN well addresses the domain shift challenge, and significantly improves the performance up to 16.43% compared with state-of-the-art baselines.

References

[Blanchard *et al.*, 2011] Gilles Blanchard, Gyemin Lee, and Clayton Scott. Generalizing from several related classification tasks to a new unlabeled sample. In *NeurIPS*, 2011.

[Chen *et al.*, 2019] Wei-Yu Chen, Yen-Cheng Liu, Zsolt Kira, Yu-Chiang Frank Wang, and Jia-Bin Huang. A closer look at few-shot classification. In *ICLR*, 2019.

[Dumoulin *et al.*, 2016] Vincent Dumoulin, Jonathon Shlens, and Manjunath Kudlur. A learned representation for artistic style. In *ICLR*, 2016.

[Fu *et al.*, 2021] Yuqian Fu, Yanwei Fu, and Yu-Gang Jiang. Meta-fdmixup: Cross-domain few-shot learning guided by labeled target data. In *ACM MM*, 2021.

[He *et al.*, 2016] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, 2016.

[Hu *et al.*, 2018] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *CVPR*, 2018.

[Islam *et al.*, 2021] Ashraful Islam, Chun-Fu Chen, Rameswar Panda, Leonid Karlinsky, Rogerio Feris, and Richard J Radke. Dynamic distillation network for cross-domain few-shot recognition with unlabeled data. In *NeurIPS*, 2021.

[Jin *et al.*, 2020] Xin Jin, Cuiling Lan, Wenjun Zeng, Zhibo Chen, and Li Zhang. Style normalization and restitution for generalizable person re-identification. In *CVPR*, 2020.

[Krause *et al.*, 2013] Jonathan Krause, Michael Stark, Jia Deng, and Li Fei-Fei. 3d object representations for fine-grained categorization. In *ICCV Workshops*, 2013.

[Kumar *et al.*, 2018] Abhishek Kumar, Prasanna Sattigeri, Kahini Wadhawan, Leonid Karlinsky, Rogério Schmidt Feris, Bill Freeman, and Gregory W Wornell. Co-regularized alignment for unsupervised domain adaptation. In *NeurIPS*, 2018.

[Kundu *et al.*, 2019] Jogendra Nath Kundu, Nishank Lakkakula, and R Venkatesh Babu. UM-Adapt: Unsupervised multi-task adaptation using adversarial cross-task distillation. In *ICCV*, 2019.

[Liang *et al.*, 2021] Hanwen Liang, Qiong Zhang, Peng Dai, and Juwei Lu. Boosting the generalization capability in cross-domain few-shot learning via noise-enhanced supervised autoencoder. In *ICCV*, 2021.

[Long *et al.*, 2015] Mingsheng Long, Yue Cao, Jianmin Wang, and Michael Jordan. Learning transferable features with deep adaptation networks. In *ICML*, 2015.

[Muandet *et al.*, 2013] Krikamol Muandet, David Balduzzi, and Bernhard Schölkopf. Domain generalization via invariant feature representation. In *ICML*, 2013.

[Phoo and Hariharan, 2020] Cheng Perng Phoo and Bharath Hariharan. Self-training for few-shot transfer across extreme task differences. In *ICLR*, 2020.

[Ravi and Larochelle, 2017] Sachin Ravi and Hugo Larochelle. Optimization as a model for few-shot learning. In *ICLR*, 2017.

[Satorras and Estrach, 2018] Victor Garcia Satorras and Joan Bruna Estrach. Few-shot learning with graph neural networks. In *ICLR*, 2018.

[Snell *et al.*, 2017] Jake Snell, Kevin Swersky, and Richard Zemel. Prototypical networks for few-shot learning. In *NeurIPS*, 2017.

[Sun *et al.*, 2021] Jiamei Sun, Sebastian Lapuschkin, Wojciech Samek, Yunqing Zhao, Ngai-Man Cheung, and Alexander Binder. Explanation-guided training for cross-domain few-shot classification. In *ICPR*, 2021.

[Sung *et al.*, 2018] Flood Sung, Yongxin Yang, Li Zhang, Tao Xiang, Philip HS Torr, and Timothy M Hospedales. Learning to compare: Relation network for few-shot learning. In *CVPR*, 2018.

[Tseng *et al.*, 2020] Hung-Yu Tseng, Hsin-Ying Lee, Jia-Bin Huang, and Ming-Hsuan Yang. Cross-domain few-shot classification via learned feature-wise transformation. In *ICLR*, 2020.

[Tzeng *et al.*, 2017] Eric Tzeng, Judy Hoffman, Kate Saenko, and Trevor Darrell. Adversarial discriminative domain adaptation. In *CVPR*, 2017.

[Van Horn *et al.*, 2018] Grant Van Horn, Oisin Mac Aodha, Yang Song, Yin Cui, Chen Sun, Alex Shepard, Hartwig Adam, Pietro Perona, and Serge Belongie. The inaturalist species classification and detection dataset. In *CVPR*, 2018.

[Vinyals *et al.*, 2016] Oriol Vinyals, Charles Blundell, Timothy Lillicrap, Koray Kavukcuoglu, and Daan Wierstra. Matching networks for one shot learning. In *NeurIPS*, 2016.

[Wah *et al.*, 2011] Catherine Wah, Steve Branson, Peter Welinder, Pietro Perona, and Serge Belongie. The caltech-ucsd birds-200-2011 dataset. 2011.

[Wang and Deng, 2021] Haoqing Wang and Zhi-Hong Deng. Cross-domain few-shot classification via adversarial task augmentation. In *IJCAI*, 2021.

[Wang *et al.*, 2020] Yaqing Wang, Quanming Yao, James T Kwok, and Lionel M Ni. Generalizing from a few examples: A survey on few-shot learning. *ACM Computing Surveys*, 2020.

[Xu *et al.*, 2022] Yi Xu, Lichen Wang, Yizhou Wang, and Yun Fu. Adaptive trajectory prediction via transferable gnn. *arXiv:2203.05046*, 2022.

[Yang *et al.*, 2018] Baoyao Yang, Andy J. Ma, and Pong C. Yuen. Learning domain-shared group-sparse representation for unsupervised domain adaptation. *Pattern Recognition*, 2018.

[Zhou *et al.*, 2017] Bolei Zhou, Agata Lapedriza, Aditya Khosla, Aude Oliva, and Antonio Torralba. Places: A 10 million image database for scene recognition. *IEEE TPAMI*, 2017.

[Zou *et al.*, 2021] Yixiong Zou, Shanghang Zhang, Jianpeng Yu, Yonghong Tian, and José MF Moura. Revisiting mid-level patterns for cross-domain few-shot recognition. In *ACM MM*, 2021.