

# Multi-Source Collaborative Gradient Discrepancy Minimization for Federated Domain Generalization

Yikang Wei<sup>1,2</sup>, Yahong Han<sup>1,2\*</sup>

<sup>1</sup>College of Intelligence and Computing, Tianjin University, Tianjin, China

<sup>2</sup>Tianjin Key Lab of Machine Learning, Tianjin University, Tianjin, China  
{yikang,yahong}@tju.edu.cn

## Abstract

Federated Domain Generalization aims to learn a domain-invariant model from multiple decentralized source domains for deployment on unseen target domain. Due to privacy concerns, the data from different source domains are kept isolated, which poses challenges in bridging the domain gap. To address this issue, we propose a Multi-source Collaborative Gradient Discrepancy Minimization (MCGDM) method for federated domain generalization. Specifically, we propose intra-domain gradient matching between the original images and augmented images to avoid overfitting the domain-specific information within isolated domains. Additionally, we propose inter-domain gradient matching with the collaboration of other domains, which can further reduce the domain shift across decentralized domains. Combining intra-domain and inter-domain gradient matching, our method enables the learned model to generalize well on unseen domains. Furthermore, our method can be extended to the federated domain adaptation task by fine-tuning the target model on the pseudo-labeled target domain. The extensive experiments on federated domain generalization and adaptation indicate that our method outperforms the state-of-the-art methods significantly.

## Introduction

Domain generalization (DG) (Zhou et al. 2021a; Wang et al. 2022) aims to learn a domain-invariant model from multiple labeled source domains for deployment on the unseen target domains. The conventional DG methods (Wang et al. 2020; Huang et al. 2020; Xu et al. 2021) assume that the data from different source domains are centralized for conducting domain generalization, e.g. reducing domain shift across multiple source domains to learn domain-invariant model, as shown in Figure 1(a). Considering data privacy issues (Ye et al. 2023; Huang et al. 2023a,b,c; Zhou and Konukoglu 2023), the data from different source domains are kept decentralized and only can be accessed on the isolated local clients. This data decentralization scenario brings great challenges to overcome the domain shift on the decentralized source domains.

To utilize the decentralized source domains, Federated Domain Generalization (FedDG) (Yuan et al. 2023; Wu and

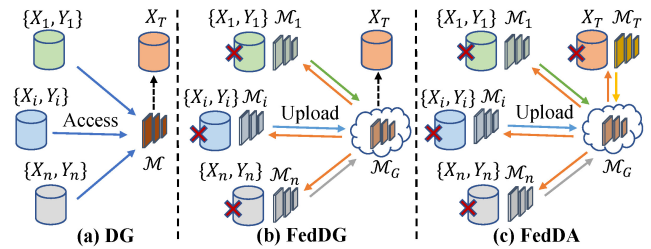


Figure 1: (a) DG assumes that the data from multiple source domains  $\{X_i, Y_i\}_{i=1}^n$  can be accessed simultaneously to learn a generalized model  $\mathcal{M}$  for deployment on the unseen domain  $X_T$ . (b) FedDG assumes that the data from different source domains are decentralized, but the local models  $\{\mathcal{M}_i\}_{i=1}^n$  of different domains can be collaboratively trained and aggregated with a parameter server. (c) FedDA assumes that an additional unlabeled target domain  $X_T$  can be accessed on server side for improving the performance.

Gong 2021; Liu et al. 2021) collaboratively trains local models on the local clients and then aggregates the local models on the server side, as shown in Figure 1(b). Since the domain shift is partly caused by the discrepancy in image styles, existing FedDG methods synthesize images with novel styles by interpolating shared style information across domains (Liu et al. 2021; Chen et al. 2023) or training data generators within isolated domains (Xu et al. 2023) to mitigate the domain shift. However, when the target domain is unseen, the generated styles may not fully cover the characteristics of the unseen domain, resulting in limited generalization performance. Additionally, sharing style information across decentralized source domains raises privacy concerns, as it may risk privacy leakage.

Unlike existing FedDG methods aiming to expand the source domains, we explore the possibility of learning the domain-invariant model by reducing the domain shift on decentralized source domains without sharing data information. The domain shift for FedDG mainly comes from two aspects. Firstly, there exists domain gap between source domains and the unseen target domain, so the local models trained on source domains tend to be domain-specific and perform poorly on the unseen target domain. Secondly, the different source domains are isolated, so the domain shift

\*Corresponding Author

Copyright © 2024, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

between isolated source domains further hinders the generalization performance on unseen target domain.

For the above domain shift problems, this paper aims to learn the intrinsic semantic information within isolated source domains and reduce the domain shift across decentralized source domains for obtaining a domain-invariant model. Inspired by the hypothesis that gradient discrepancy between domains hinders domain-invariant learning (Mansilla et al. 2021; Du et al. 2021), where conflicting gradients across domains indicate that the models optimization direction becomes domain-specific, we propose a method called Multi-source Collaborative Gradient Discrepancy Minimization (MCGDM) for FedDG. Our method contains two key components. For avoiding the domain-specific information learned within the isolated source domains, we propose intra-domain gradient matching, which minimizes the gradient discrepancy between the original images and the augmented images to learn the intrinsic semantic information. Additionally, for bridging the domain gap across decentralized source domains, we propose inter-domain gradient matching, which minimizes the gradient discrepancy between the current model and the models of other domains. Combining the intra-domain and inter-domain gradient matching, the domain shift within isolated source domains and across decentralized source domains can be mitigated effectively so that a domain-invariant model can be learned. Furthermore, the proposed MCGDM method can be extended to the Federated Domain Adaptation (FedDA) task (Feng et al. 2021; Wu and Gong 2021; Wei et al. 2022; Wei and Han 2023, 2022; Wu, Jia, and Han 2023), where the unlabeled target domain  $X_T$  can be accessed on the server side for fine-tuning the target domain model, as illustrated in Figure 1(c). The extensive experiments on multiple datasets indicate the effectiveness of our method on FedDG and FedDA tasks.

The contributions of our work can be summarized as follows:

- To avoid the domain-specific information learned by local models, we propose intra-domain gradient matching, which minimizes the gradient discrepancy between original images and augmented images for learning the intrinsic semantic information.
- To reduce the domain shift across decentralized source domains, we propose inter-domain gradient matching, which minimizes the gradient discrepancy between the current model and the models from other domains.
- By combining intra-domain gradient matching and inter-domain gradient matching, the domain shift can be reduced within isolated source domains and across decentralized source domains to generalize well on unseen target domain. Furthermore, our method can be expanded to the FedDA task. The experiments on FedDG and FedDA datasets indicate the effectiveness of our method.

## Related Work

### Domain Generalization

Conventional domain generalization utilizes multiple centralized source domains to learn a domain-invariant model

for generalizing well on unseen target domain. The existing domain generalization methods can be categorized as (1) data augmentation methods, (2) domain-invariant representation learning methods, and (3) other learning strategies, such as self-supervised learning, and ensemble learning. The data augmentation methods aim to expand the source domains for improving the generalization performance of the model on unseen domain, such as L2A-OT (Zhou et al. 2020b) and DDAIG (Zhou et al. 2020a) train a data generator to synthesize out-of-distribution images, FACT (Xu et al. 2021), MixStyle (Zhou et al. 2021b), and StyleNeo (Kang et al. 2022) interpolate the styles between different domains to generate images or features with novel styles. The domain-invariant representation learning methods aim to learn the intrinsic semantic representations or reduce the domain shift across multiple source domains for generalizing well on the unseen target domain, such as RSC (Huang et al. 2020) and CDG (Du et al. 2022) suppress the domain-specific representations to learn the intrinsic semantic representations across different domains. Furthermore, the other learning strategies such as self-supervised learning methods JGen (Carlucci et al. 2019) and EISNet (Wang et al. 2020) utilize the jigsaw auxiliary task to learn the domain-invariant representations, the ensemble learning method DAEL (Zhou et al. 2021a) distills the knowledge from the ensemble of other domains to learn domain-invariant model. Different from these methods which need to access the data from different source domains simultaneously, our method conducts multi-source collaborative federated domain generalization under the data decentralization scenario.

### Federated Domain Generalization

Federated domain generalization learns a domain-invariant global model from multiple decentralized source domains for generalizing well on unseen target domain. The existing FedDG methods mainly focus on (1) data augmentation on decentralized source domains and (2) model aggregation on the server side. The data augmentation methods e.g. ELCFS (Liu et al. 2021) and CCST (Chen et al. 2023) share the style information across decentralized source domains to generate the images with novel styles. The FADH (Xu et al. 2023) trains the data generators on the isolated source domains to synthesize images. And COPA (Wu and Gong 2021) utilizes the classifier heads from other domains to learn the domain-invariant representations with the assistance of augmented images, e.g. RandAug (Cubuk et al. 2020). These data augmentation methods focus on generating out-of-distribution images under the data decentralization scenario and conduct domain-invariant learning between the original images and augmented images. Different from these methods, our method focuses on learning domain-invariant models from the perspective of gradient discrepancy minimization and can be combined with the existing data augmentation methods. The other FedDG methods focus on the model aggregation stage on the server side for obtaining a fair model across different domains, which tends to be domain-invariant on the unseen domain. GA (Zhang et al. 2023) aggregates the local source models on the server side with different weights, and CASC (Yuan et al. 2023) aggregates the local source mod-

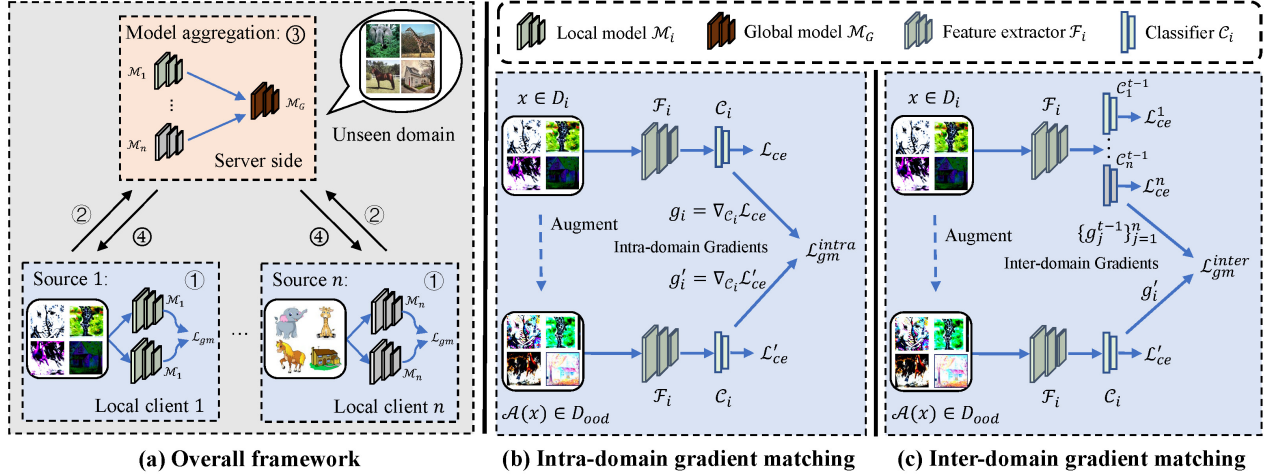


Figure 2: (a) The overall framework of our method, where the local source domain models  $\{\mathcal{M}_i\}_{i=1}^n$  are trained locally by conducting gradient matching and then are aggregated on the server side to obtain the generalizable global model  $\mathcal{M}_G$ . (b) The intra-domain gradient matching between the original images and the augmented images is conducted on local clients for learning the intrinsic semantic information within the domain. (c) The inter-domain gradient matching between the current classifier head  $\mathcal{C}_i$  and the classifier heads from other domains  $\{\mathcal{C}_j^{t-1}\}_{j=1}^n$  is conducted on the local clients for reducing the domain shift across decentralized source domains.

els with layer-wise weights. Different from these methods, we focus on learning the domain-invariant local models on decentralized source domains, which can achieve better generalization performance.

## Methodology

### Overview

For the domain generalization task, there are  $n$  source domains  $\{D_i\}_{i=1}^n$ , each domain  $D_i$  contains  $N_i$  labeled images  $\{x_i, y_i\}_{i=1}^{N_i}$ . A domain-invariant model is expected to be learned on source domains  $\{D_i\}_{i=1}^n$  for deployment on the out-of-distribution unseen target domain  $D_t$ . There exists domain shift between source domains and unseen target domain. Here we consider the covariate shift, the marginal distribution  $P(\mathcal{X})$  of the input images differs but the conditional label distribution  $P(\mathcal{Y}|\mathcal{X})$  keeps the same across source domains and unseen target domain. With this assumption, a domain-invariant model learned on source domains can generalize well on unseen target domain by reducing the domain shift.

Under the data decentralization scenario, data from different source domains  $\{D_i\}_{i=1}^n$  cannot be accessed simultaneously. In this work, we collaboratively train the decentralized source domains  $\{D_i\}_{i=1}^n$  by the federated average (McMahan et al. 2017) to obtain a domain-invariant global model for applying on the unseen target domain  $D_t$ , as shown in Figure 2(a). Specifically, the local models  $\{\mathcal{M}_i\}_{i=1}^n$  are trained on each isolated source domain  $\{D_i\}_{i=1}^n$  and then are aggregated on the server side to obtain a global model  $\mathcal{M}_G$ . Each local model  $\mathcal{M}_i$  contains a feature extractor  $\mathcal{F}_i$  and a classifier head  $\mathcal{C}_i$ . There are 4 steps to collaboratively train the decentralized source domains  $\{D_i\}_{i=1}^n$ . In step 1, the local models  $\{\mathcal{M}_i\}_{i=1}^n$  are

trained on the isolated local clients via intra-domain and inter-domain gradient matching for learning intrinsic semantic information and reducing the domain shift across decentralized source domains. In step 2 and 3, the local models  $\{\mathcal{M}_i\}_{i=1}^n$  are uploaded to the server side and are aggregated by parameter averaging to obtain a global model  $\mathcal{M}_G$ . In step 4, the global model  $\mathcal{M}_G$  and classifier heads  $\{\mathcal{C}_i^{t-1}\}_{i=1}^n$  of different domains are uploaded to local clients for the next round of training. After several rounds of training, a domain-invariant global model  $\mathcal{M}_G$  can be learned across decentralized source domains and deployed on the unseen domain  $D_t$ .

### Intra-domain Gradient Matching

As there are labeled source domain data on the isolated clients, we train the local model  $\mathcal{F}_i \circ \mathcal{C}_i$  by cross-entropy loss as follows:

$$\mathcal{L}_{ce}(x; \mathcal{F}_i \circ \mathcal{C}_i) = - \sum_{i=1}^B y_i \log \mathcal{F}_i \circ \mathcal{C}_i(x_i), \quad (1)$$

where  $B$  is the batch size of input images.

As mentioned above, it is challenging to learn the domain-invariant model on the local clients because the data from different source domains are decentralized. For example, the local model  $\mathcal{M}_i$  trained only on the isolated client tends to be domain-specific for the source domain  $D_i$ . Inspired by the existing FedDG works (Wu and Gong 2021; Zhou et al. 2021a), which can learn the domain-invariant representations by enforcing the consistency between original images and augmented images on feature space or predictions, we learn the domain-invariant model with the assistance of augmented images to filter the domain-specific information on the isolated client. Different from the existing methods (Wu

and Gong 2021), we enforce the consistency between original images and augmented images on parameter space to ensure local model updating toward the intrinsic semantic information, as shown in Figure 2(b). Specifically, we propose intra-domain gradient matching between original images and augmented images on the isolated source domains as follows:

$$\mathcal{L}_{gm}^{intra} = 1 - \text{sim}(g_i, g'_i). \quad (2)$$

The  $\text{sim}(\cdot, \cdot)$  is the cosine similarity.  $g_i$  is the gradient of classifier  $\mathcal{C}_i$  on original images:

$$g_i = \nabla_{\mathcal{C}_i} \mathcal{L}_{ce}(x; \mathcal{F}_i \circ \mathcal{C}_i), \quad (3)$$

where  $\mathcal{L}_{ce}(x; \mathcal{F}_i \circ \mathcal{C}_i)$  is the cross-entropy loss on original images.

And  $g'_i$  is the gradient of classifier  $\mathcal{C}_i$  on augmented images  $\mathcal{A}_x$ :

$$g'_i = \nabla_{\mathcal{C}_i} \mathcal{L}_{ce}(\mathcal{A}(x); \mathcal{F}_i \circ \mathcal{C}_i). \quad (4)$$

By using the intra-domain gradient matching between the original image  $x$  and augmented image  $\mathcal{A}(x)$ , the local model tends to update toward the intrinsic semantic information contained on the original image  $x$  and augmented image  $\mathcal{A}(x)$ . Following previous work (Wu and Gong 2021; Zhou et al. 2021a), the augmented images  $\mathcal{A}(x)$  are generated by RandAugment (RandAug) (Cubuk et al. 2020). Our method also can be combined with other style augmentation strategies, which is validated in the experiments section.

### Inter-domain Gradient Matching

Due to the different source domains being isolated, it is challenging to bridge the domain gap across decentralized source domains. Different from the existing FedDG methods (Liu et al. 2021; Chen et al. 2023) sharing data information across domains, we use the classifier heads  $\{\mathcal{C}_j^{t-1}\}_{j=1}^n$  from other domains as the bridges to reduce the domain shift, as shown in Figure 2(c). Specifically, we conduct inter-domain gradient matching for reducing the domain shift across decentralized source domains, where the gradient discrepancy between the current classifier head and the classifier heads from other domains are reduced:

$$\mathcal{L}_{gm}^{inter} = \sum_{j=1}^n (1 - \text{sim}(g'_i, g_j^{t-1})). \quad (5)$$

$g'_i$  is the gradient of current classifier  $\mathcal{C}_i$  on augmented images  $\mathcal{A}(x)$ :

$$g'_i = \nabla_{\mathcal{C}_i} \mathcal{L}_{ce}(\mathcal{A}(x); \mathcal{F}_i \circ \mathcal{C}_i), \quad (6)$$

$g_j^{t-1}$  is the gradient of classifier  $\mathcal{C}_j^{t-1}$  from  $j$ -th source domain on the original images:

$$g_j^{t-1} = \nabla_{\mathcal{C}_j^{t-1}} \mathcal{L}_{ce}(x; \mathcal{F}_i \circ \mathcal{C}_j^{t-1}). \quad (7)$$

By conducting inter-domain gradient matching between the current classifier head  $\mathcal{C}_i$  and the classifier heads from different domains  $\{\mathcal{C}_j^{t-1}\}_{j=1}^n$ , the domain shift can be reduced across decentralized source domains.

On the local client, the overall loss function contains classification loss on the original images and augmented images, intra-domain gradient matching loss, and inter-domain gradient matching loss as follows:

$$\mathcal{L}_{loc} = \frac{1}{2} (\mathcal{L}_{ce}(x; \mathcal{F}_i \circ \mathcal{C}_i) + \mathcal{L}_{ce}(\mathcal{A}(x); \mathcal{F}_i \circ \mathcal{C}_i)) + \lambda \mathcal{L}_{gm}^{intra} + (1 - \lambda) \mathcal{L}_{gm}^{inter}, \quad (8)$$

where the  $\lambda$  is the hyper-parameter for balancing the intra-domain gradient matching loss  $\mathcal{L}_{gm}^{intra}$  and inter-domain gradient matching loss  $\mathcal{L}_{gm}^{inter}$ .

**Remark:** Conducting gradient matching on overall model  $\mathcal{F}_i \circ \mathcal{C}_i$  or feature extractor  $\mathcal{F}_i$  will lead to large computational cost. So we only minimize the gradient discrepancy on classifier head  $\mathcal{C}_i$ , and the domain-invariant model can be learned by end-to-end training with the gradient matching loss. Furthermore, compared with sharing the data information of other domains, utilizing the classifier heads as the bridges have little risk of privacy leakage (Wu and Gong 2021; Jeong et al. 2021).

### Model Aggregation

As shown in Figure 2(a), after conducting local training e.g. 1 epoch on the isolated source domains, the local models  $\{\mathcal{M}_i\}_{i=1}^n$  from different source domains are uploaded to server-side to conduct model aggregation in Step 3:

$$\mathcal{M}_G = \sum_{i=1}^n \frac{N_i}{N_{total}} \mathcal{M}_i, \quad (9)$$

where the  $N_{total} = \sum_{i=1}^n N_i$  is the sum of all samples. After conducting model aggregation on the server side, the global model  $\mathcal{M}_G$  and the classifier heads  $\{\mathcal{C}_i^{t-1}\}_{i=1}^n$  from different source domains are uploaded to local clients for the next round training. The global model  $\mathcal{M}_G$  is used as the initial model for local training and the classifier heads  $\{\mathcal{C}_i^{t-1}\}$  from different source domains are used to collaboratively bridge the domain gap across domains. We conduct  $E$  rounds of local training on isolated source domains and model aggregation on the server side until the global model convergence.

### Extend to Federated Domain Adaptation

Moreover, our method can be extended to the FedDA task. Different from FedDG where the target domain is unseen, FedDA (Feng et al. 2021; Wei et al. 2022) assumes that unlabeled target domain data  $X_T$  can be accessed on the server side for further improving the performance on the target domain, as shown in Figure 1(c). For utilizing the unlabeled target domain, the target domain model  $\mathcal{M}_T$  is trained on the pseudo-labeled target domain data:

$$\mathcal{L}_{ce}(x; \mathcal{F}_T \circ \mathcal{C}_T) = - \sum_{i=1}^B \hat{y}_T \log \mathcal{F}_T \circ \mathcal{C}_T(x_T), \quad (10)$$

where we utilize the domain-invariant local source models  $\{\mathcal{M}_i\}_{i=1}^n$  learned by our MCGDM method to generate the

high-confident pseudo-labels  $\hat{y}_T$  on target domain following previous works (Feng et al. 2021). Then the local source domain models  $\{\mathcal{M}_i\}_{i=1}^n$  and target domain model  $\mathcal{M}_T$  are aggregated to obtain the global model  $\mathcal{M}_G$ , which is used as the initial model for the next round of training on source domains and target domain. By fine-tuning the global model  $\mathcal{M}_G$  on the pseudo-labeled target domain, the obtained target domain model  $\mathcal{M}_T$  can achieve better performance on the target domain.

## Experiments

### Datasets

**Federated Domain Generalization Datasets.** We conduct FedDG experiments on three image classification datasets including **PACS**(Li et al. 2017), **VLCS**(Ghifary et al. 2015), and **Office-Home**(Venkateswara et al. 2017). **PACS** contains 9,991 images of 7 categories from four domains: Art-painting (A), Cartoon (C), Photo (P), and Sketch (S), the data of each domain are split into 80% for training and 20% for testing. **Office-Home** contains 15,500 images of 65 categories from four domains: Artist (A), Clipart (C), Product (P), and Real-world (R), each domain is split into 90% as the training set and 10% as the test set. **VLCS** contains 10,729 images of 5 categories from four domains: Pascal (P), LabelMe (L), Caltech (C), and Sun (S), each domain is split into 80% for training and 20% for testing. The leave-one-domain-out protocol (Zhou et al. 2021b) is used to evaluate the generalization performance on one domain and train the model on the rest source domains.

**Federated Domain Adaptation Datasets.** We conduct FedDA experiments on two image classification datasets including **Digit-5** and **Office-Caltech10**. **Digit-5** contains five domains MNIST(mt), MNIST-M(mm), SVHN(sv), Syn(syn), and USPS(up) of 10 categories. **Office-Caltech10** contains four domains Amazon(A), Caltech(C), Webcam(W), and Dslr(D) of 10 categories. Following the setting of previous works (Feng et al. 2021; Wu and Gong 2021), the model is trained on the labeled source domains and an unlabeled target domain, and the results on the target domain are reported.

### Implementation Details

Following the previous works (Yuan et al. 2023; Wu and Gong 2021), we use the ResNet-18 (He et al. 2016) pre-trained on ImageNet as the backbone for PACS, Office-Home, and VLCS datasets. The SGD optimizer with momentum 0.9 and weight decay  $5e-4$  is adopted for PACS, Office-Home, and VLCS. For PACS and VLCS, the batch size is 16 and the initial learning rate is 0.001, decayed by the cosine schedule from 0.001 to 0.0001 during training. For Office-Home, the batch size is 30 and the initial learning rate is 0.002, decayed by cosine scheduled from 0.002 to 0.0001. The hyper-parameter  $\lambda$  on Equation 8 is 0.3 for PACS, 0.8 for Office-Home and VLCS.

For each local client, we train the local model 1 epoch and then upload the local model to the server side for conducting model aggregation. The local training on the client side and global model aggregation on the server side are

conducted iteratively. The total training rounds  $E$  is 40 on PACS, VLCS, and Office-Home datasets. All experiments are repeated three times with different random seeds and the mean accuracy (%) is reported.

For the federated domain adaptation datasets, we use the three-layer CNN as the backbone for Digit-5 and the ResNet-101 pre-trained on ImageNet as the backbone for Office-Caltech10 following the setting of previous work KD3A (Feng et al. 2021). The total training rounds are 50 for Digit-5 and 40 for Office-Caltech10. Following previous work (Feng et al. 2021), we generate the pseudo-labels on the unlabeled target domain by knowledge vote. We implement our method with PyTorch and use a single NVIDIA RTX3090. Moreover, we reproduce the results on PACS dataset with MindSpore framework (Huawei 2020).

Methods	A	C	P	S	Avg
DeepAll (Zhou et al. 2021a)	77.0	75.9	96.0	69.2	79.5
JiGen (Carlucci et al. 2019)	79.4	75.3	96.0	71.4	80.5
EISNet (Wang et al. 2020)	81.9	76.4	95.9	74.3	82.2
L2A-OT (Zhou et al. 2020b)	83.3	78.2	<b>96.2</b>	73.6	82.8
DDAIG (Zhou et al. 2020a)	84.2	78.1	95.3	74.7	83.1
FACT (Xu et al. 2021)	85.4	78.4	95.2	79.2	84.5
MixStyle (Zhou et al. 2021b)	84.1	78.8	<u>96.1</u>	75.9	83.7
StyleNeo (Kang et al. 2022)	84.4	79.2	94.9	83.2	85.4
RSC (Huang et al. 2020)	83.4	80.3	96.0	80.9	85.2
CDG (Du et al. 2022)	83.5	80.1	95.6	<u>83.8</u>	<u>85.8</u>
DAEL (Zhou et al. 2021a)	84.6	74.4	95.6	78.9	83.4
FedAvg (McMahan et al. 2017)	82.2	77.3	94.5	71.2	81.3
CASC (Yuan et al. 2023)	82.0	76.4	95.2	81.6	83.8
GA (Zhang et al. 2023)	83.2	76.9	94.0	82.9	84.3
ELCFS (Liu et al. 2021)	82.3	74.7	93.3	82.7	83.2
FADH (Xu et al. 2023)	83.8	77.2	94.4	<b>84.4</b>	85.0
COPA (Wu and Gong 2021)	83.3	79.8	94.6	82.5	85.1
MCGDM (ours)	<b>86.2</b>	<u>81.2</u>	96.0	81.2	<b>86.2</b>
MCGDM†(ours)	<u>86.1</u>	<b>81.4</b>	96.0	81.2	<b>86.2</b>

Table 1: Accuracy(%) on PACS dataset. We have bolded the best results and underlined the second results. † indicates the results reproduced with MindSpore framework.

### Comparison with State-of-the-Art Methods

**Results on PACS.** PACS is a dataset with a large domain gap between different domains, e.g. different image styles. As shown in Table 1, our method achieves the best average accuracy of 86.2% compared with the conventional DG methods and FedDG methods. Compared with the data augmentation FedDG methods, e.g. ELCFS, FADH, and COPA, our method focuses on learning the domain-invariant model across original images and augmented images, which achieves the 1%-3% improvement on average accuracy. Compared with the FedDG methods which calibrate the aggregation weights on the server side, e.g. GA and CASC, our method reduces the domain shift on local clients and achieves a 2% improvement on average accuracy. Furthermore, our method even outperforms the conventional DG methods which can access the data from different source domains simultaneously, e.g. data augmentation methods FACT and domain-invariant representation learning methods RSC and CDG.

Methods	A	C	P	R	Avg
DeepAll (Zhou et al. 2021a)	57.9	52.7	73.5	74.8	64.7
JiGen (Carlucci et al. 2019)	53.0	47.5	71.5	72.8	61.2
EISNet (Wang et al. 2020)	56.8	53.3	72.3	73.5	64.0
L2A-OT (Zhou et al. 2020b)	<b>60.6</b>	50.1	74.8	<b>77.0</b>	65.6
DDAIG (Zhou et al. 2020a)	59.2	52.3	74.6	<u>76.0</u>	65.5
MixStyle (Zhou et al. 2021b)	58.7	53.4	74.2	75.9	65.5
StyleNeo (Kang et al. 2022)	59.6	55.0	73.6	75.5	65.9
RSC (Huang et al. 2020)	58.4	47.9	71.6	74.5	63.1
CDG (Du et al. 2022)	59.2	54.3	<u>74.9</u>	75.7	66.0
DAEL (Zhou et al. 2021a)	59.4	55.1	74.0	75.7	<u>66.1</u>
FedAvg (McMahan et al. 2017)	58.2	51.6	73.1	73.8	64.2
GA (Zhang et al. 2023)	58.8	54.3	73.7	74.7	65.4
ELCFS (Liu et al. 2021)	57.8	54.9	71.1	73.1	64.2
CCST (Chen et al. 2023)	59.1	50.1	73.0	71.7	63.6
FADH (Xu et al. 2023)	<u>59.9</u>	<u>55.8</u>	73.5	74.9	66.0
COPA (Wu and Gong 2021)	59.4	55.1	74.8	75.0	<u>66.1</u>
MCGDM (ours)	<u>59.9</u>	<b>56.7</b>	<b>75.3</b>	75.2	<b>66.8</b>

Table 2: Accuracy(%) on Office-Home dataset. We have bolded the best results and underlined the second results.

Methods	P	L	C	S	Avg
DeepAll (Zhou et al. 2021a)	71.4	59.8	97.5	69.0	74.4
JiGen (Carlucci et al. 2019)	74.0	61.9	97.4	66.9	75.1
L2A-OT (Zhou et al. 2020b)	72.8	59.8	<u>98.0</u>	70.9	75.4
RSC (Huang et al. 2020)	<u>75.3</u>	59.8	97.0	71.5	75.9
MixStyle (Zhou et al. 2021b)	72.6	58.5	97.7	73.3	75.5
FedAvg (McMahan et al. 2017)	72.0	63.3	96.5	72.4	76.0
CASC (Yuan et al. 2023)	72.0	<u>63.5</u>	97.2	72.1	<u>76.2</u>
ELCFS (Liu et al. 2021)	71.1	59.5	96.6	<b>74.0</b>	75.3
COPA (Wu and Gong 2021)	71.5	61.0	93.8	71.7	74.5
MCGDM (ours)	<b>75.6</b>	<b>64.3</b>	<b>98.6</b>	<u>73.8</u>	<b>78.1</b>

Table 3: Accuracy(%) on VLCS dataset. We have bolded the best results and underlined the second results.

**Results on Office-Home.** As shown in Table 2, our method also achieves the best average accuracy of 66.8% compared with other FedDG methods and conventional DG methods. Due to the domain gap between isolated source domains being ignored, the existing data style augmentation methods e.g. ELCFS and CCST cannot achieve significant improvements on average accuracy. Instead, our method reduces the domain shift across isolated source domains and achieves significant improvements. Compared with COPA, which fixes the classifier heads from other domains and learns the domain-invariant feature extractor, our method achieves better performance by conducting intra-domain and inter-domain gradient matching on local clients. We also conduct ablation study about different alignment strategies in the Ablation Study section to indicate the effectiveness of our gradient matching strategy.

**Results on VLCS.** We also conduct experiments on the VLCS dataset. As shown in Table 3, our method achieves consistent improvement on Pascal, LabelMe, Caltech, and Sun domains. Compared with other FedDG methods, our method improves the average accuracy by 2%-3%. Compared with the state-of-the-art conventional DG methods, such as RSC and MixStyle, our method also achieves signif-

Baseline	RandAug	$\mathcal{L}_{gm}^{intra}$	$\mathcal{L}_{gm}^{inter}$	A	C	P	S	Avg
✓				82.2	77.3	94.5	71.2	81.3
✓	✓			85.2	78.6	<b>96.0</b>	75.8	83.9
✓	✓	✓		85.8	79.8	95.4	80.8	85.5
✓	✓		✓	85.6	79.8	94.8	<b>81.3</b>	85.4
✓	✓	✓	✓	<b>86.2</b>	<b>81.2</b>	<b>96.0</b>	81.2	<b>86.2</b>

Table 4: Accuracy(%) of each component on PACS dataset, combined with the data augmentation strategy RandAug.

icant improvements with the privacy constraints. The results on VLCS indicate the effectiveness of our method to learn the domain-invariant local models for better generalization performance.

## Ablation Study

**Contributions of Different Components** As shown in Table 4, the Baseline indicates training the decentralized source domains by FedAvg (McMahan et al. 2017) combined with MixStyle (Zhou et al. 2021b), which interpolates the features within mini-batch. By utilizing the data augmentation strategy RandAug (Cubuk et al. 2020) to expand source domains, the performance on unseen target domain can be further improved. Our method aims to reduce the domain shift so that the learned model can generalize well on unseen domain. By combining the gradient matching loss  $\mathcal{L}_{gm}^{intra}$  and  $\mathcal{L}_{gm}^{inter}$  proposed in this paper, the best average accuracy of 86.2% can be achieved.

Furthermore, we visualize the activation maps by Grad-CAM (Selvaraju et al. 2017), as shown in Figure 3. Combining with the  $\mathcal{L}_{gm}^{intra}$  and  $\mathcal{L}_{gm}^{inter}$ , the learned model can focus on the most discriminative regions on the images from the unseen domain, which will lead to better generalizable performance on the unseen domain.

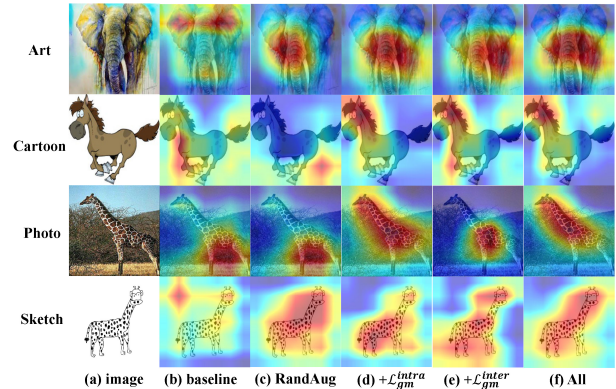


Figure 3: Visualization by Grad-CAM on unseen domain.

**Comparison with Different Alignment Strategies** We also make a comparison with different alignment strategies between the original images and augmented images: (1) reducing the Kullback-Leibler (KL) divergence on logits-level between original images and augmented images, which is used by StyleNeo (Kang et al. 2022), (2) conducting Supervised Contrastive loss (SupCon) (Khosla et al. 2020) on

Method	A	C	P	S	Avg
Baseline	85.2	78.6	96.0	75.8	83.9
+ KL (Kang et al. 2022)	85.8	78.7	96.0	78.6	84.8
+ SupCon (Khosla et al. 2020)	86.1	78.5	<b>96.2</b>	78.3	84.8
+ CoFC (Wu and Gong 2021)	85.4	79.1	96.0	76.3	84.2
+ MCGDM (ours)	<b>86.2</b>	<b>81.2</b>	96.0	<b>81.2</b>	<b>86.2</b>

Table 5: Accuracy (%) of different alignment strategies on PACS dataset.

Baseline	AM	$\mathcal{L}_{gm}^{intra}$	$\mathcal{L}_{gm}^{inter}$	A	C	P	S	Avg
✓				82.2	77.3	<b>94.5</b>	71.2	81.3
✓	✓			83.3	77.1	93.3	76.8	82.6
✓	✓	✓		83.3	78.6	93.1	81.0	84.0
✓	✓		✓	83.6	78.3	93.1	80.6	83.9
✓	✓	✓	✓	<b>84.1</b>	<b>79.0</b>	93.3	<b>81.9</b>	<b>84.6</b>

Table 6: Accuracy(%) of each component on PACS dataset, combined with other data augmentation strategy.

feature-level between the original images and augmented images, (3) Collaboration of Frozen Classifiers (CoFC) used by COPA (Wu and Gong 2021), which freezes the multiple classifier heads of other domains to learn the domain-invariant feature extractor. Different from these methods which align the features or logits between original images and augmented images, our method reduces the gradient discrepancy within isolated domains and across decentralized source domains. As shown in Table 5, our method outperforms other alignment strategies by about 2% on average accuracy and achieves large improvement on hard domains, e.g. Art, Cartoon, and Sketch.

**Combined with Other Augmentation Strategy** We also combine our gradient matching method with the data augmentation strategy Amplitude Mix (AM), which is the main component used by ELCFS and FACT to interpolate the style information across domains. Following previous work (Zhang et al. 2023), we interpolate the amplitudes of different images within mini-batch to expand source domains. As shown in Table 6, our method can improve the average accuracy by about 2% by combining with AM on PACS dataset, which can indicate the effectiveness of our method in reducing the domain shift. In this work, we use the RandAug as the default data augmentation strategy.

**Experiments on Federated Domain Adaptation** We make a comparison with the multi-source domain adaptation methods: (1) conventional unsupervised multi-source domain adaptation methods DCTN (Xu et al. 2018) and M3SDA (Peng et al. 2019a), where the source domains and target domain are centralized, (2) Source-free methods SHOT (Liang, Hu, and Feng 2020), DECISION (Ahmed et al. 2021), and CAiDA (Dong et al. 2021), where only the pre-trained source domain models can be accessed on unlabeled target domain for adaptation, (3) FedDA methods FADA (Peng et al. 2019b), KD3A (Feng et al. 2021), and COPA (Wu and Gong 2021), where the decentralized source domains and target domain are trained collaboratively.

As shown in Table 7 and Table 8, our method outperforms

Method	mt	mm	sv	syn	up	Avg
Source-only	92.3	63.7	71.5	83.4	90.7	80.3
DCTN (Xu et al. 2018)	96.2	70.5	77.6	86.8	92.8	84.8
M3SDA (Peng et al. 2019a)	98.4	72.8	81.3	89.6	96.2	87.7
SHOT (Liang, Hu, and Feng 2020)	98.2	80.2	84.5	91.1	97.1	90.2
DECISION (Ahmed et al. 2021)	99.2	93.0	82.6	97.5	97.8	94.0
CAiDA (Dong et al. 2021)	99.1	<u>93.7</u>	83.3	<b>98.1</b>	<u>98.6</u>	94.6
FADA (Peng et al. 2019b)	91.4	62.5	50.5	71.8	91.7	73.6
KD3A (Feng et al. 2021)	99.2	87.3	85.6	89.4	98.5	92.0
COPA (Wu and Gong 2021)	<u>99.4</u>	89.8	<u>91.0</u>	97.5	<b>99.2</b>	<u>95.4</u>
MCGDM (ours)	<b>99.5</b>	<b>96.9</b>	<b>92.3</b>	<u>97.8</u>	<b>99.2</b>	<b>97.1</b>

Table 7: Accuracy(%) of different multi-source domain adaptation methods on Digit-5 dataset.

Method	A	C	D	W	Avg
Source-only	86.1	87.8	98.3	99.0	92.8
DCTN (Xu et al. 2018)	92.7	90.2	99.0	99.4	95.3
M3SDA (Peng et al. 2019a)	94.5	92.2	99.2	99.5	96.4
SHOT (Liang, Hu, and Feng 2020)	96.4	96.2	98.5	99.7	97.7
DECISION (Ahmed et al. 2021)	95.9	95.9	<b>100</b>	99.6	98.0
CAiDA (Dong et al. 2021)	<u>96.8</u>	<b>97.1</b>	<b>100</b>	<u>99.8</u>	<u>98.4</u>
FADA (Peng et al. 2019b)	84.2	88.7	87.1	88.1	87.1
KD3A (Feng et al. 2021)	<b>97.4</b>	96.4	98.4	99.7	97.9
COPA (Wu and Gong 2021)	95.8	94.6	<u>99.6</u>	<u>99.8</u>	97.5
MCGDM (ours)	<b>97.4</b>	<u>96.9</u>	<b>100</b>	<b>100</b>	<b>98.6</b>

Table 8: Accuracy(%) of different multi-source domain adaptation methods on Office-Caltech10 dataset.

other methods significantly. Different from the source-free methods, e.g., SHOT, DECISION, and CAiDA, our method can collaboratively train the isolated source domains and target domain and lead to better performance. Different from the FedDA methods e.g. FADA, KD3A, and COPA, our method conducts gradient matching on the local clients to reduce the domain shift, so that the learned local models on source domains can be generalized well on the target domain. By further fine-tuning the target model on the pseudo-labeled target domain, our method can achieve state-of-the-art performance and outperform other FedDA methods significantly.

## Conclusion

In this paper, we propose a multi-source collaborative gradient discrepancy minimization method for federated domain generalization. Our main idea is to reduce the domain shift within isolated source domains and across decentralized source domains for learning the domain-invariant model, which can generalize on unseen target domain. Moreover, our method can be combined with the existing data augmentation strategies and improve the generalization performance on unseen target domain. Furthermore, we extend our method to the federated domain adaptation task by fine-tuning the target model on the pseudo-labeled target domain. The extensive experiments on federated domain generalization and adaptation datasets can validate the effectiveness of our method.

## Acknowledgments

This work is supported by the National Natural Science Foundation of China (under Grant 62376186, 61932009) and the CAAI-Huawei MindSpore Open Fund.

## References

- Ahmed, S. M.; Raychaudhuri, D. S.; Paul, S.; Oymak, S.; and Roy-Chowdhury, A. K. 2021. Unsupervised Multi-source Domain Adaptation Without Access to Source Data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 10103–10112.
- Carlucci, F. M.; D’Innocente, A.; Bucci, S.; Caputo, B.; and Tommasi, T. 2019. Domain generalization by solving jigsaw puzzles. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2229–2238.
- Chen, J.; Jiang, M.; Dou, Q.; and Chen, Q. 2023. Federated Domain Generalization for Image Recognition via Cross-Client Style Transfer. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 361–370.
- Cubuk, E. D.; Zoph, B.; Shlens, J.; and Le, Q. V. 2020. Randaugment: Practical automated data augmentation with a reduced search space. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, 702–703.
- Dong, J.; Fang, Z.; Liu, A.; Sun, G.; and Liu, T. 2021. Confident anchor-induced multi-source free domain adaptation. *Advances in Neural Information Processing Systems*, 34: 2848–2860.
- Du, D.; Chen, J.; Li, Y.; Ma, K.; Wu, G.; Zheng, Y.; and Wang, L. 2022. Cross-domain gated learning for domain generalization. *International Journal of Computer Vision*, 130(11): 2842–2857.
- Du, Z.; Li, J.; Su, H.; Zhu, L.; and Lu, K. 2021. Cross-domain gradient discrepancy minimization for unsupervised domain adaptation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 3937–3946.
- Feng, H.; You, Z.; Chen, M.; Zhang, T.; Zhu, M.; Wu, F.; Wu, C.; and Chen, W. 2021. KD3A: Unsupervised Multi-Source Decentralized Domain Adaptation via Knowledge Distillation. In *International Conference on Machine Learning*, 3274–3283. PMLR.
- Ghifary, M.; Kleijn, W. B.; Zhang, M.; and Balduzzi, D. 2015. Domain generalization for object recognition with multi-task autoencoders. In *Proceedings of the IEEE international conference on computer vision*, 2551–2559.
- He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.
- Huang, W.; Ye, M.; Shi, Z.; and Du, B. 2023a. Generalizable heterogeneous federated cross-correlation and instance similarity learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- Huang, W.; Ye, M.; Shi, Z.; Li, H.; and Du, B. 2023b. Rethinking Federated Learning With Domain Shift: A Prototype View. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 16312–16322.
- Huang, W.; Ye, M.; Shi, Z.; Wan, G.; Li, H.; Du, B.; and Yang, Q. 2023c. Federated Learning for Generalization, Robustness, Fairness: A Survey and Benchmark. arXiv:2311.06750.
- 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.
- Huawei. 2020. MindSpore. <https://www.mindspore.cn/>.
- Jeong, W.; Yoon, J.; Yang, E.; and Hwang, S. J. 2021. Federated Semi-Supervised Learning with Inter-Client Consistency & Disjoint Learning. In *International Conference on Learning Representations*.
- Kang, J.; Lee, S.; Kim, N.; and Kwak, S. 2022. Style neophile: Constantly seeking novel styles for domain generalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 7130–7140.
- Khosla, P.; Teterwak, P.; Wang, C.; Sarna, A.; Tian, Y.; Isola, P.; Maschinot, A.; Liu, C.; and Krishnan, D. 2020. Supervised contrastive learning. *Advances in neural information processing systems*, 33: 18661–18673.
- Li, D.; Yang, Y.; Song, Y.-Z.; and Hospedales, T. M. 2017. Deeper, broader and artier domain generalization. In *Proceedings of the IEEE international conference on computer vision*, 5542–5550.
- Liang, J.; Hu, D.; and Feng, J. 2020. Do we really need to access the source data? source hypothesis transfer for unsupervised domain adaptation. In *International Conference on Machine Learning*, 6028–6039. PMLR.
- Liu, Q.; Chen, C.; Qin, J.; Dou, Q.; and Heng, P.-A. 2021. Feddg: Federated domain generalization on medical image segmentation via episodic learning in continuous frequency space. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 1013–1023.
- Mansilla, L.; Echeveste, R.; Milone, D. H.; and Ferrante, E. 2021. Domain generalization via gradient surgery. In *Proceedings of the IEEE/CVF international conference on computer vision*, 6630–6638.
- McMahan, B.; Moore, E.; Ramage, D.; Hampson, S.; and y Arcas, B. A. 2017. Communication-efficient learning of deep networks from decentralized data. In *Artificial Intelligence and Statistics*, 1273–1282. PMLR.
- Peng, X.; Bai, Q.; Xia, X.; Huang, Z.; Saenko, K.; and Wang, B. 2019a. Moment matching for multi-source domain adaptation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 1406–1415.
- Peng, X.; Huang, Z.; Zhu, Y.; and Saenko, K. 2019b. Federated adversarial domain adaptation. arXiv:1911.02054.

- Selvaraju, R. R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; and Batra, D. 2017. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*, 618–626.
- 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, J.; Lan, C.; Liu, C.; Ouyang, Y.; Qin, T.; Lu, W.; Chen, Y.; Zeng, W.; and Yu, P. 2022. Generalizing to unseen domains: A survey on domain generalization. *IEEE Transactions on Knowledge and Data Engineering*.
- Wang, S.; Yu, L.; Li, C.; Fu, C.-W.; and Heng, P.-A. 2020. Learning from extrinsic and intrinsic supervisions for domain generalization. In *Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part IX*, 159–176. Springer.
- Wei, Y.; and Han, Y. 2022. Dual collaboration for decentralized multi-source domain adaptation. *Frontiers of Information Technology & Electronic Engineering*, 23(12): 1780–1794.
- Wei, Y.; and Han, Y. 2023. Exploring Instance Relation for Decentralized Multi-Source Domain Adaptation. In *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1–5.
- Wei, Y.; Yang, L.; Han, Y.; and Hu, Q. 2022. Multi-source Collaborative Contrastive Learning for Decentralized Domain Adaptation. *IEEE Transactions on Circuits and Systems for Video Technology*.
- Wu, G.; and Gong, S. 2021. Collaborative optimization and aggregation for decentralized domain generalization and adaptation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 6484–6493.
- Wu, K.; Jia, F.; and Han, Y. 2023. Domain-specific feature elimination: multi-source domain adaptation for image classification. *Frontiers of Computer Science*, 17(4): 174705.
- Xu, Q.; Zhang, R.; Zhang, Y.; Wang, Y.; and Tian, Q. 2021. A fourier-based framework for domain generalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 14383–14392.
- Xu, Q.; Zhang, R.; Zhang, Y.; Wu, Y.-Y.; and Wang, Y. 2023. Federated Adversarial Domain Hallucination for Privacy-preserving Domain Generalization. *IEEE Transactions on Multimedia*.
- Xu, R.; Chen, Z.; Zuo, W.; Yan, J.; and Lin, L. 2018. Deep cocktail network: Multi-source unsupervised domain adaptation with category shift. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3964–3973.
- Ye, M.; Fang, X.; Du, B.; Yuen, P. C.; and Tao, D. 2023. Heterogeneous federated learning: State-of-the-art and research challenges. *ACM Computing Surveys*, 56(3): 1–44.
- Yuan, J.; Ma, X.; Chen, D.; Wu, F.; Lin, L.; and Kuang, K. 2023. Collaborative Semantic Aggregation and Calibration for Federated Domain Generalization. *IEEE Transactions on Knowledge and Data Engineering*.
- Zhang, R.; Xu, Q.; Yao, J.; Zhang, Y.; Tian, Q.; and Wang, Y. 2023. Federated domain generalization with generalization adjustment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 3954–3963.
- Zhou, K.; Yang, Y.; Hospedales, T.; and Xiang, T. 2020a. Deep domain-adversarial image generation for domain generalisation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, 13025–13032.
- Zhou, K.; Yang, Y.; Hospedales, T.; and Xiang, T. 2020b. Learning to generate novel domains for domain generalization. In *Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XVI 16*, 561–578. Springer.
- Zhou, K.; Yang, Y.; Qiao, Y.; and Xiang, T. 2021a. Domain adaptive ensemble learning. *IEEE Transactions on Image Processing*, 30: 8008–8018.
- Zhou, K.; Yang, Y.; Qiao, Y.; and Xiang, T. 2021b. Domain generalization with mixstyle. arXiv:2104.02008.
- Zhou, T.; and Konukoglu, E. 2023. Fedfa: Federated feature augmentation. arXiv:2301.12995.