Indirect Alignment and Relationships Preservation for Domain Generalization

Wei Wei[1] , **Zixiong Li**[1] , **Jing Yan**[1] , **Mingwen Shao**[2] and **Lin Li**[1*]
[1]Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education, School of Computer and Information Technology, Shanxi University, Taiyuan, Shanxi, China
[2]Qingdao Institute of Software, College of Computer Science and Technology, China University of Petroleum (East China), Qingdao, Shandong, China
weiwei@sxu.edu.cn, {1150789614, 541567625}@qq.com, smw278@126.com, lilynn1116@sxu.edu.cn,

Abstract

Domain generalization (DG) aims to train models on multiple source domains to generalize effectively to unseen target domains, addressing performance degradation caused by domain shifts. Many existing methods rely on direct feature alignment, which disrupts natural sequence relationships, causes misalignment and feature distortion, and leads to overfitting, especially with significant domain gaps. To tackle these issues, we propose a novel DG approach with two key modules: the Sample Difference Keeping (SDK) module, which preserves natural sequence relationships to enhance feature diversity and separability, and the Sample Consistency Alignment (SCA) module, which achieves indirect alignment by modeling inter-class and inter-domain relationships consistencies. This approach mitigates overfitting and misalignment, ensuring adaptability to significant domain gaps. Extensive experiments demonstrate that our framework consistently outperforms state-of-the-art methods.

1 Introduction

In recent years, deep learning has achieved remarkable success in fields like image recognition [Touvron *et al.*, 2022], natural language processing [Zhang *et al.*, 2023a], and object detection [Zou *et al.*, 2023]. However, these advancements typically rely on the assumption that training and testing data share the same distribution (IID assumption), which often does not hold in real-world applications. Factors such as varying devices and environments can cause data distribution shifts, leading to significant performance drops when models are tested on new data—a challenge known as domain shift. This issue severely limits the cross-domain generalization of deep learning models, particularly in fields like medical imaging and autonomous driving, where data acquisition is complex.

Domain Adaptation (DA) methods [He *et al.*, 2023] address domain shift by leveraging source domain data and a small amount of target domain data. However, these methods are limited when target domain data is unavailable. Domain Generalization (DG) trains models on multiple source domains to overcome this, enabling generalization to unseen target domains without relying on target domain data. Recently, contrastive learning-based DG methods [Tong *et al.*, 2023; Hu *et al.*, 2024; Chen *et al.*, 2023b] have gained attention for their ability to align features of same-class samples while distinctly separating those of different classes, enabling fine-grained feature representation.

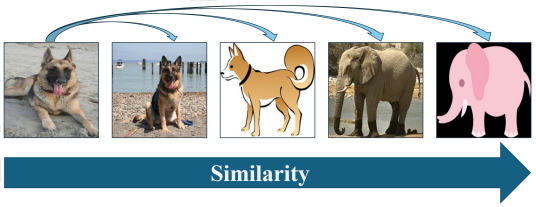


Figure 1: Illustrates the natural sequence relationships between samples across domains and class. It depicts how the similarity between samples changes progressively: (1) the highest similarity occurs between same-class samples within the same domain (e.g., two images of dogs from the "Photo" domain); (2) slightly lower similarity exists between same-class samples from different domains (e.g., a dog image from the "Photo" domain and one from the "Cartoon" domain); (3) further reduced similarity is observed between different-class samples within the same domain (e.g., a dog image and an elephant image from the "Photo" domain); (4) the lowest similarity occurs between different-class samples from different domains (e.g., a dog image from the "Photo" domain and an elephant image from the "Cartoon" domain).

However, traditional methods often focus narrowly on directly aligning same-class samples across source domains, overlooking the natural sequence relationships between data samples. This can lead to overfitting, as models may overemphasize local source-domain features while neglecting subtle yet crucial global patterns essential for generalization. For example, as shown in Figure 1, the similarity between a sample and another same-class sample within the same domain (e.g., two images of dogs from the "Photo" domain) should naturally be higher than the similarity between a sample and a same-class sample from a different domain (e.g., an image of a dog from the "Photo" domain and one from the "Cartoon" domain). Similarly, the similarity between a sample and a same-class sample from a different domain should nat-

*Corresponding author.

urally be higher than that between a sample and a different-class sample within the same domain (e.g., an image of a dog and an image of an elephant from the "Photo" domain). Finally, the similarity between a sample and a different-class sample within the same domain should naturally be higher than that between a sample and a different-class sample from different domains (e.g., an image of a dog from the "Photo" domain and an image of an elephant from the "Cartoon" domain). These sequential relationships provide critical cues for generalization that humans intuitively recognize by observing fine-grained micro-details. Ignoring these relationships often results in models that fail to capture the rich semantic nuances needed for robust cross-domain performance.

To address these challenges, we propose an innovative domain generalization framework that integrates the Sample Difference Keeping (SDK) module and the Sample Consistency Alignment (SCA) module. The main contributions of our work are as follows:

1. We introduce the SDK module, which preserves the natural sequence relationships between samples across different domains and classes. This facilitates the learning of diverse and discriminative features while reducing the risk of overfitting to domain-specific patterns.
2. We propose the SCA module, which employs a novel indirect alignment strategy based on structural relationships. This strategy ensures inter-class consistency by maintaining class separability across domains and inter-domain consistency by capturing shared patterns while preserving domain-specific characteristics. The SCA module enhances the model's generalization ability by avoiding the misalignment and feature distortion issues that are often present in direct alignment methods.
3. We conducted comprehensive evaluations on widely used domain generalization benchmarks. The results demonstrate that our framework significantly outperforms state-of-the-art methods, achieving superior cross-domain generalization and robustness against unseen domains.

2 Related Work

Domain generalization (DG) tackles distribution shifts without relying on target domain data, leveraging methods like data augmentation [Su *et al.*, 2023; Wang *et al.*, 2024], representation learning [Yu *et al.*, 2024], and advanced strategies such as meta-learning and ensemble learning [Chen *et al.*, 2023a] to enhance cross-domain generalization.

Recently, contrastive learning, a self-supervised approach, has gained attention for its ability to learn robust representations by constructing positive and negative sample pairs. Both unsupervised and supervised contrastive methods [He *et al.*, 2020; Khosla *et al.*, 2020] have demonstrated strong performance, attributed to their effective strategies for pair construction and loss design.

In DG, contrastive learning has gained traction for its simplicity and effectiveness:

- Direct Utilization and Integration: Studies [Motiian *et al.*, 2017; Kim *et al.*, 2021; Zhang *et al.*, 2023b] explore integrating contrastive learning with additional strategies to address DG challenges.
- Incorporating Global Information: Methods [Yao *et al.*, 2022; Dong *et al.*, 2022; Li *et al.*, 2023] introduce a global perspective to enhance feature alignment.
- Leveraging Sample relationships: Refining alignment by emphasizing hard positive samples [Hu *et al.*, 2024] or dataset repartitioning [Tong *et al.*, 2023].

Despite these advancements, traditional DG methods often rely on direct alignment strategies, which oversimplify complex relationships and neglect subtle features, leading to misalignment and feature distortion under significant domain gaps. Addressing these shortcomings requires methods that account for nuanced relationships while maintaining generalization.

While ranking-based methods in metric learning, such as deep metric learning [Cakir *et al.*, 2019], Ranked List Loss [Wang *et al.*, 2019], and Self-Supervised Synthesis Ranking (SSR) [Fu *et al.*, 2021], effectively preserve relative ranking information and local structures, they do not specifically address domain shifts in DG tasks. Instead, these approaches primarily focus on optimizing intra-domain ranking performance and lack mechanisms for mitigating inter-domain misalignment caused by distribution shifts.

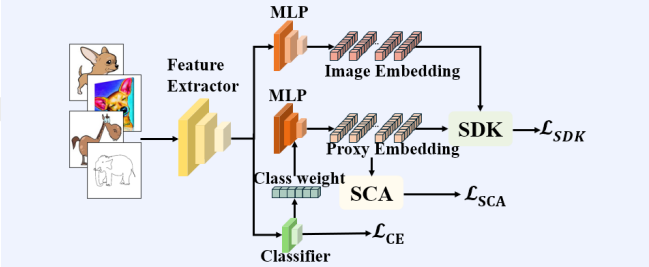
To address these challenges in DG, we propose the Sample Difference Keeping (SDK) and Sample Consistency Alignment (SCA) modules. SDK preserves natural sequence relationships to enhance feature diversity and separability across domains, while SCA employs indirect alignment to refine cross-domain consistency without the pitfalls of direct alignment, ensuring robustness and improved generalization in domain generalization tasks.

3 Proposed Method

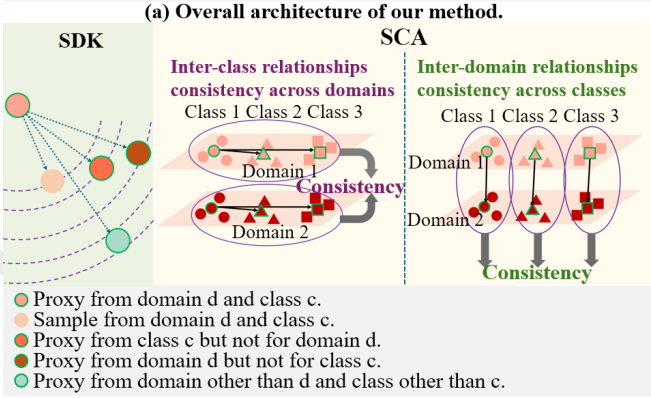
3.1 Preliminary Introduction

In the classical domain generalization setting, we assume the model has access to multiple source domains during training. These source domains form the training set $\mathcal{D} = \{D_1, D_2, \ldots, D_M\}$, where M is the number of source domains. Let $X \in \mathbb{R}^d$ denote the input feature space and $Y \in \mathbb{R}^C$ denote the class label space. Each source domain D_k contains a series of sample pairs $\{(x_i^k, y_i^k)\}_{i=1}^N$, where N represents the number of samples in this domain, $x_i^k \in X$ is the input feature, and $y_i^k \in Y$ is the class label. The goal is to train a model $F_{\phi,\theta}$, which consists of a feature extractor $f_\phi{:}X \to Z$ that extracts a universal representation Z from input data X, and a predictor $g_\theta{:}Z \to \mathbb{R}^C$ that generates predictions for C classes based on the features Z, such that the model can accurately classify data from an unknown target domain D_t without relying on the target domain, which shares the same set of class labels Y as the source domains D_k.

Our proposed domain generalization method is designed to address the limitations of direct alignment strategies in traditional approaches. These methods often disrupt the natural sequence relationships between samples, leading to overfitting, misalignment, and feature distortion, which result in



(a) Overall architecture of our method.



(b) SDK and SCA modules.

Figure 2: Illustrates the overall architecture of our method. (a) The input data is first processed by the feature extractor to produce universal features. These features are fed into both the classifier (for computing $\mathcal{L}_{\text{CE}}$) and an embedding layer (to generate feature embeddings). Classifier weights are also passed through the embedding layer to produce domain- and class-specific proxy embeddings. (b) The SDK module operates on the embeddings to preserve natural sequence relationships across samples and proxies, addressing overfitting by focusing on transferable patterns. The SCA module ensures indirect alignment by maintaining inter-class relationships consistency across domains and inter-domain relationships consistency across classes, enhancing the model's adaptability to unseen domains.

reduced performance, especially when domain discrepancies are significant. Figure 2 provides an overview of our proposed method, with specific details described in the following subsections.

3.2 Sample Diversity Keeping Module (SDK)

In domain generalization tasks, samples exhibit inherent sequence relationships based on their similarities, as shown in Figure 1. Preserving these natural sequence relationships is crucial for enhancing model generalization. However, traditional methods often neglect these relationships, leading to overfitting.

To address this, we propose the Sample Diversity Keeping (SDK) module, which explicitly constrains the relative distances between samples across domains and classes. This ensures the preservation of natural sequence relationships, enabling us to observe subtle details that aid in generalizing to new scenarios, reducing overfitting, and improving performance on unseen target domains.

To achieve this, we first project the sample features z_i and class weights w_i into feature embeddings $q_i \in \mathbb{R}^{pro}$ and proxy sets $P_i \subseteq \mathbb{R}^{pro}$, respectively. We use two different multilayer perceptrons (MLPs) as projection heads: h_z for sample features and h_w for class weights:

$$q_i = h_z(z_i), P_i = h_w(w_i), \tag{1}$$

where $P_i = \{p_i^1, p_i^2, ...p_i^M\}$ represents the set of all M proxies for class i across different source domains.

SDK maintains the natural sequence relationships between samples by constraining the relative distances between samples from different classes and domains. Specifically, to reduce computational complexity, we introduce proxy embeddings, which replace sample embeddings. We formalize the natural sequence relationships constraint as follows:

$$d(p_i^j, p_{\bar{i}}^{\bar{j}}) > d(p_i^j, p_{\bar{i}}^j) > d(p_i^j, p_i^{\bar{j}}) > d(q_i^j, p_i^j), \tag{2}$$

where p_i^j denotes the embedding of a proxy from class i and domain j, $p_i^{\bar{j}}$ is the embedding of a proxy from class i but a different domain, $p_{\bar{i}}^j$ is the embedding of a proxy from domain j but a different class, $p_{\bar{i}}^{\bar{j}}$ represents a proxy from a different class and a different domain, and q_i^j denotes the embedding of a sample from class i and domain j.

In practice, we replace the distance metric with cosine similarity $s(\cdot,\cdot)$, leading to the following modified constraint:

$$s(p_i^j, p_{\bar{i}}^{\bar{j}}) < s(p_i^j, p_{\bar{i}}^j) < s(p_i^j, p_i^{\bar{j}}) < s(q_i^j, p_i^j), \tag{3}$$

To enforce this constraint, we define the Sample Diversity Keeping (SDK) loss function using a triplet loss approach. For each triplet, we minimize the cosine similarity between the anchor and positive samples while maximizing the similarity between the anchor and negative samples, with a margin term to separate them:

$$\begin{aligned}\mathcal{L}_{\text{SDK-O}} = &\mathcal{L}_{\text{T}}\left(s\left(p_i^j, q_i^j\right), s\left(p_i^j, p_{\bar{i}}^{\bar{j}}\right), m_1\right) \\ &+\mathcal{L}_{\text{T}}\left(s\left(p_i^j, p_{\bar{i}}^{\bar{j}}\right), s\left(p_i^j, p_{\bar{i}}^j\right), m_2\right) \\ &+\mathcal{L}_{\text{T}}\left(s\left(p_i^j, p_{\bar{i}}^j\right), s\left(p_i^j, p_i^{\bar{j}}\right), m_3\right),\end{aligned} \tag{4}$$

where $\mathcal{L}_{\text{T}}$ denotes the triplet loss function, $\mathcal{L}_{\text{T}}(AP, AN, \text{m}) = max(AP - AN + \text{m}, 0)$ and AP represents the cosine similarity between the anchor and positive samples, AN represents the similarity between the anchor and negative samples, and m is a margin hyperparameter.

Because the number of triplets in metric learning can be large, resulting in heavy computational costs, we introduce an efficient max-min sample mining strategy [Schroff *et al.*, 2015] to speed up training. The modified SDK loss function, using this strategy, is given by:

$$\begin{aligned}\mathcal{L}_{\text{SDK}} = &\mathcal{L}_{\text{T}}\left(\min\left\{s\left(q_i^j, p_i^j\right)\right\}, \max\left\{s\left(p_i^j, p_i^{\bar{j}}\right)\right\}, m_1\right) \\ &+\mathcal{L}_{\text{T}}(\min\{s\left(p_i^j, p_i^{\bar{j}}\right)\}, \max\{s\left(p_i^j, p_{\bar{i}}^j\right)\}, m_2) \\ &+\mathcal{L}_{\text{T}}\left(\min\left\{s\left(p_i^j, p_{\bar{i}}^j\right)\right\}, \max\left\{s\left(p_i^j, p_{\bar{i}}^{\bar{j}}\right)\right\}, m_3\right),\end{aligned} \tag{5}$$

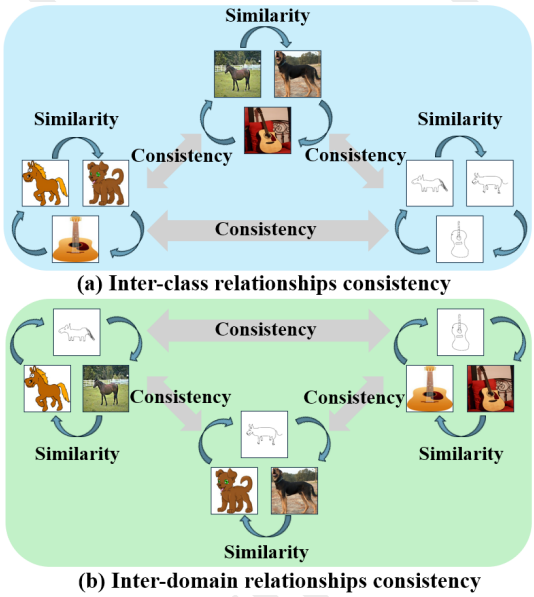


Figure 3: Illustrates the inter-class and inter-domain relationships consistency used in the proposed Sample Consistency Alignment (SCA) module. Subfigure (a) demonstrates inter-class relationships consistency, where similarity relationships between classes (e.g., dogs and horses) are preserved within each domain, ensuring that shared features can be effectively extracted. Subfigure (b) depicts inter-domain relationships consistency, highlighting that domains such as photos and cartoons exhibit higher feature similarity than sketches, which contain minimal semantic information.

By selecting the smallest and largest similarity values for each triplet, this max-min strategy significantly reduces the number of triplets, accelerating training while maintaining good generalization performance. Through the SDK module, the model learns more generalizable feature representations, which aids in generalizing to unseen target domains.

3.3 Sample Consistency Alignment Module (SCA)

Traditional domain generalization methods rely on direct alignment strategies to find common features across domains. However, these methods oversimplify relationships between samples, neglect subtle features, and frequently lead to misalignment and feature distortion, especially when domain differences are significant. These shortcomings result in performance degradation and underscore the need for an indirect alignment approach that effectively captures both domain-specific and shared features.

To address this, we analyze inter-class and inter-domain feature relationships. For example, as shown in Figure 3, within any domain, classes like dogs and horses are more similar to each other than to unrelated classes like guitars. This stable inter-class relationships facilitates extracting shared features. From an inter-domain perspective, domains such as cartoons and photos exhibit high feature similarity, while sketches, which retain less semantic content, show lower similarity. These inter-domain relationships remain consistent across classes, with photos containing the richest semantic information, followed by cartoons, and sketches exhibiting the least.

Building on these insights, we propose the Sample Consistency Alignment (SCA) module, which employs an indirect alignment strategy to leverage inter-class and inter-domain relationships. SCA consists of two key components: inter-class relationships consistency, which preserves similarity relationships between classes across domains, and inter-domain relationships consistency, which maintains similarity relationships between domains across classes. This dual consistency ensures robust cross-domain feature extraction and enhances generalization to unseen target domains.

To improve efficiency, SCA incorporates a proxy mechanism, where proxies represent samples from different classes and domains, reducing computational costs. This mechanism enables efficient consistency alignment, effectively capturing cross-domain shared features and strengthening the model To improve efficiency, we incorporate a proxy mechanism in SCA, where proxies represent samples from different classes and domains, reducing computational costs. This mechanism enables efficient consistency alignment, effectively capturing cross-domain shared features and strengthening the model's generalization performance.

Inter-class relationships consistency. Inter-class relationships consistency aims to ensure that the similarity relationships between classes remain consistent in the feature space across different domains, thereby facilitating the extraction of common features. To achieve this, we calculate the similarity matrix for the classes in each domain and minimize the Frobenius norm of the differences between these matrices.

Specifically, for domain D_i (where $i = 1, 2, ..., M$) and class c (where $c = 1, 2, ..., C$), we define the proxy embedding for class c in domain D_i as p_c^i. We then calculate the inter-class similarity matrix $\mathbf{S}_i^{dom} \in \mathbb{R}^{C\times C}$ for domain D_i, where each element $S^l_{c_1,c_2}$ represents the cosine similarity between the proxy embeddings of class c_1 and class c_2 in domain l :

$$S^l_{c_1,c_2} = s(p^l_{c_1}, p^l_{c_2}), \tag{6}$$

To ensure that the similarity matrices across the M domains are as similar as possible, we minimize the Frobenius mean difference between them, resulting in the inter-class relationships consistency loss $\mathcal{L}_{\text{SCA}-\text{C}}$:

$$\mathcal{L}_{\text{SCA}-\text{C}} = \frac{1}{M(M-1)} \sum_{i=1}^{M} \sum_{j=i+1}^{M} \parallel \mathbf{S}_i^{dom} - \mathbf{S}_j^{dom} \parallel_F^2, \tag{7}$$

where $\parallel \mathbf{S}_i^{dom} - \mathbf{S}_j^{dom} \parallel_F$ represents the Frobenius norm between matrices $\mathbf{S}_i^{dom}$ and $\mathbf{S}_j^{dom}$:

$$\parallel \mathbf{S}_i^{dom} - \mathbf{S}_j^{dom} \parallel_F = \sqrt{\sum_{c_1=1}^{C} \sum_{c_2=1}^{C} \left(S^i_{c_1,c_2} - S^j_{c_1,c_2} \right)^2}, \tag{8}$$

Inter-domain relationships consistency. Similarly, the goal is to ensure that the similarity relationships between domains remain consistent across different classes. This helps the model capture cross-domain common features and improves its adaptability to domain shifts. Analogous to inter-class relationships consistency, we define an inter-domain

Algorithm	PACS	VLCS	Office-Home	TerraIncognita	Average
CORAL [Sun and Saenko, 2016]	86.2	78.8	68.7	47.7	70.35
DANN [Ganin *et al.*, 2016]	83.7	78.6	65.9	46.7	68.73
CDANN [Li *et al.*, 2018b]	82.6	77.5	65.7	45.8	67.90
MMD [Yu *et al.*, 2024]	84.7	77.5	66.4	42.2	67.70
MLDG [Li *et al.*, 2018a]	84.9	77.2	66.8	47.8	69.18
IRM [Ahuja *et al.*, 2020]	83.5	78.6	64.3	47.6	68.50
RSC [Huang *et al.*, 2020]	85.2	77.1	65.5	46.6	68.60
SupCon [Khosla *et al.*, 2020]	88.1	79.3	70.6	50.7	72.18
MTL [Blanchard *et al.*, 2021]	84.6	77.2	66.4	45.6	68.45
ERM [Vapnik, 1999]	85.5	77.3	67.6	47.8	69.55
Mixstyle [Zhou *et al.*, 2021]	85.2	77.9	60.4	44.0	66.88
SagNet [Nam *et al.*, 2021]	86.3	77.8	68.1	48.6	70.20
SWAD [Cha *et al.*, 2021]	88.1	79.1	70.6	50.0	71.95
HTCL [Tong *et al.*, 2023]	88.6	77.6	71.3	50.9	72.10
PCL [Yao *et al.*, 2022]	88.7	78.0	71.6	52.1	72.60
IAIE-Pair[Hu *et al.*, 2024]	88.8	79.6	70.9	52.2	72.88
IAIE-Proxy[Hu *et al.*, 2024]	89.2	78.5	71.8	52.6	73.03
Ours	**89.3**	**80.4**	**72.3**	**53.9**	**73.98**

Table 1: Comparison of ACC (%) with the SOTA methods on four benchmarks.

similarity matrix $\mathbf{S}_c^{cls} \in \mathbb{R}^{M\times M}$ for each class c, where each element $S_o^{l_1,l_2}$ represents the similarity between domains l_1 and l_2 for class o:

$$S_o^{l_1,l_2} = sim(p_o^{l_1}, p_o^{l_2}), \tag{9}$$

We ensure inter-domain relationships consistency by minimizing the differences between domain similarity matrices for different classes, resulting in the inter-domain relationships consistency loss $\mathcal{L}_{\mathrm{SCA-D}}$:

$$\mathcal{L}_{\mathrm{SCA-D}} = \frac{1}{C(C-1)} \sum_{i=1}^{C} \sum_{j=i+1}^{C} \parallel \mathbf{S}_i^{cls} - \mathbf{S}_j^{cls} \parallel_F^2, \tag{10}$$

Finally, the overall Sample Consistency Alignment Loss consists of the inter-class relationships consistency loss and the inter-domain relationships consistency loss, weighted by a hyperparameter α:

$$\mathcal{L}_{\mathrm{SCA}} = \alpha \mathcal{L}_{\mathrm{SCA-C}} + \mathcal{L}_{\mathrm{SCA-D}}, \tag{11}$$

By incorporating the proxy mechanism, SCA reduces computational costs while achieving efficient consistency alignment. This dual consistency Alignment helps the model capture cross-domain common features, thereby improving its generalization to unseen target domains.

The overall objective function is defined as:

$$\mathcal{L}_{\mathrm{Total}} = \mathcal{L}_{\mathrm{CE}} + \mathcal{L}_{\mathrm{SDK}} + \mathcal{L}_{\mathrm{SCA}}. \tag{12}$$

4 Experimental Study

4.1 Datasets

To thoroughly evaluate the proposed method, we selected four challenging domain generalization datasets:

- PACS [Li *et al.*, 2017]: Contains 9,991 images, divided into 7 categories and 4 domains (Photos, Art Paintings, Cartoons, Sketches), each with its unique image style.
- Office-Home [Venkateswara *et al.*, 2017]: Contains approximately 15,500 images, divided into 65 categories and 4 domains (Art, Clipart, Product, Real World), and serves as a popular benchmark for evaluating domain generalization performance.
- VLCS [Fang *et al.*, 2013]: Contains 10,729 images, divided into 5 categories and 4 domains (VOC2007, LabelMe, Caltech101, SUN09), used for testing the model's generalization ability across different data sources.
- TerraIncognita [Beery *et al.*, 2018]: Contains 24,788 wildlife images, divided into 10 categories and 4 domains (L100, L38, L43, L46), posing a challenging task for domain generalization.

These datasets cover a wide range of image styles and domains, providing rich experimental scenarios for validating our approach.

4.2 Implementation Details

Our method builds upon SWAD [Cha *et al.*, 2021] and PCL [Yao *et al.*, 2022]. We use the PyTorch framework with a pre-trained ResNet-50 [He *et al.*, 2016] on ImageNet [Russakovsky *et al.*, 2015] as the backbone network. To ensure consistency with PCL, we employ the same projection layers. The Adam optimizer is applied with a learning rate of $5e^{-5}$. Following the standard training and evaluation procedure of SWAD, we use a batch size of 32 for each domain. A leave-one-domain strategy is adopted, where one domain serves as the target domain, and the remaining domains are used as source domains for training.

For model validation and selection, 20% of the samples from each source domain are set aside as the validation set. To ensure result stability, we run experiments with three different random seeds and report the average accuracy and standard error. For the number of iterations, the Office-Home dataset is trained for 3,000 iterations, while the other datasets are trained for 5,000 iterations. The hyperparameters are set as follows: $m_1 = 0.1$, $m_2 = 0.4$, $m_3 = 0.1$, and $\alpha = 0.01$.

Algorithm	Art	Cartoon	Photo	Sketch	Average
CORAL [Sun and Saenko, 2016]	88.3 ± 0.2	80.0 ± 0.5	97.5 ± 0.3	78.3 ± 1.3	86.2
DANN [Ganin *et al.*, 2016]	86.4 ± 0.8	77.4 ± 0.8	97.3 ± 0.4	73.5 ± 2.3	83.7
CDANN [Li *et al.*, 2018b]	84.6 ± 1.8	75.5 ± 0.9	96.8 ± 0.3	73.5 ± 0.6	82.6
MMD [Yu *et al.*, 2024]	86.1 ± 1.4	79.4 ± 0.9	96.6 ± 0.2	76.5 ± 0.5	84.7
MLDG [Li *et al.*, 2018a]	85.5 ± 1.4	80.1 ± 0.7	97.4 ± 0.3	76.6 ± 1.1	84.9
IRM [Ahuja *et al.*, 2020]	84.8 ± 1.3	76.4 ± 1.1	96.7 ± 0.6	76.1 ± 1.0	83.5
RSC [Huang *et al.*, 2020]	85.4 ± 0.8	79.7 ± 1.8	97.6 ± 0.3	78.2 ± 1.2	85.2
SupCon [Khosla *et al.*, 2020]	89.4 ± 0.5	83.5 ± 0.9	97.4 ± 0.2	82.3 ± 0.4	88.1
MTL [Blanchard *et al.*, 2021]	87.5 ± 0.8	77.1 ± 0.5	96.4 ± 0.8	77.3 ± 1.8	84.6
ERM [Vapnik, 1999]	84.7 ± 0.4	80.8 ± 0.6	97.2 ± 0.3	79.3 ± 1.0	85.5
Mixstyle [Zhou *et al.*, 2021]	86.8 ± 0.5	79.0 ± 1.4	96.6 ± 0.1	78.5 ± 2.3	85.2
SagNet [Nam *et al.*, 2021]	87.4 ± 1.0	80.7 ± 0.6	97.1 ± 0.1	80.0 ± 0.4	86.3
SWAD [Cha *et al.*, 2021]	89.3 ± 0.2	83.4 ± 0.6	97.3 ± 0.3	82.5 ± 0.5	88.1
PCL [Yao *et al.*, 2022]	90.2 ± 0.2	83.9 ± 0.6	$\mathbf{98.1 \pm 0.1}$	82.6 ± 1.2	88.7
IAIE-Pair[Hu *et al.*, 2024]	90.1 ± 0.4	83.9 ± 1.0	97.2 ± 0.3	83.8 ± 0.1	88.8
IAIE-Proxy[Hu *et al.*, 2024]	90.6 ± 0.3	84.0 ± 1.0	97.8 ± 0.1	84.1 ± 0.4	89.2
Ours	$\mathbf{90.9 \pm 0.2}$	$\mathbf{84.6 \pm 0.1}$	97.1 ± 0.3	$\mathbf{84.8 \pm 0.6}$	**89.3**

Table 2: Comparison of ACC (%) with the SOTA methods on PACS.

4.3 Comparison with State-of-the-Art Methods

We conducted a comprehensive comparison between the proposed method and state-of-the-art domain generalization methods on four standard datasets. The experimental results, presented in Tables 1 to 5, demonstrate that our method achieves superior performance across all datasets. Compared to earlier domain generalization approaches (such as ERM

Method	Caltech101	LabelMe	SUN09	VOC2007	Average
CORAL [Sun and Saenko, 2016]	98.3 ± 0.1	66.1 ± 1.2	73.4 ± 0.3	77.5 ± 1.2	78.8
DANN [Ganin *et al.*, 2016]	99.0 ± 0.3	65.1 ± 1.4	73.1 ± 0.3	77.2 ± 0.6	78.6
CDANN [Li *et al.*, 2018b]	97.7 ± 0.1	65.1 ± 1.2	70.7 ± 0.8	77.1 ± 1.5	77.5
MMD [Yu *et al.*, 2024]	97.7 ± 0.1	64.0 ± 1.1	72.8 ± 0.2	75.3 ± 3.3	77.5
MLDG [Li *et al.*, 2018a]	97.4 ± 0.2	65.2 ± 0.7	71.0 ± 1.4	75.3 ± 1.0	77.2
IRM [Ahuja *et al.*, 2020]	98.6 ± 0.1	64.9 ± 0.9	73.4 ± 0.6	77.3 ± 0.9	78.6
RSC [Huang *et al.*, 2020]	97.9 ± 0.1	62.5 ± 0.7	72.3 ± 1.2	75.6 ± 0.8	77.1
SupCon [Khosla *et al.*, 2020]	98.7 ± 0.1	63.7 ± 0.2	75.6 ± 0.5	79.0 ± 0.2	79.3
MTL [Blanchard *et al.*, 2021]	97.8 ± 0.4	64.3 ± 0.3	71.5 ± 0.7	75.3 ± 1.7	77.2
ERM [Vapnik, 1999]	98.0 ± 0.3	64.7 ± 1.2	71.4 ± 1.2	75.2 ± 1.6	77.3
Mixstyle [Zhou *et al.*, 2021]	98.6 ± 0.3	64.5 ± 1.1	72.6 ± 0.5	75.7 ± 1.7	77.9
SagNet [Nam *et al.*, 2021]	97.9 ± 0.4	64.5 ± 0.5	71.4 ± 1.3	77.5 ± 0.5	77.8
SWAD [Cha *et al.*, 2021]	98.8 ± 0.1	63.3 ± 0.3	75.3 ± 0.5	79.2 ± 0.6	79.1
PCL [Yao *et al.*, 2022]	**99.0 ± 0.1**	63.4 ± 0.6	73.8 ± 0.3	75.6 ± 1.8	78.0
IAIE-Pair[Hu *et al.*, 2024]	98.9 ± 0.1	64.0 ± 0.2	**76.1 ± 0.1**	**79.5 ± 0.3**	79.6
IAIE-Proxy[Hu *et al.*, 2024]	98.9 ± 0.1	63.0 ± 0.7	72.8 ± 1.4	79.2 ± 0.4	78.5
Ours	98.0 ± 0.3	**69.6 ± 0.6**	74.5 ± 0.8	79.4 ± 1.3	**80.4**

Table 3: Comparison of ACC (%) with the SOTA methods on VLCS.

Method	Art	Clipart	Product	Real World	Average
CORAL [Sun and Saenko, 2016]	65.3 ± 0.4	54.4 ± 0.5	76.5 ± 0.1	78.4 ± 0.5	68.7
DANN [Ganin *et al.*, 2016]	59.9 ± 1.3	53.0 ± 0.3	73.6 ± 0.7	76.9 ± 0.5	65.9
CDANN [Li *et al.*, 2018b]	61.5 ± 1.4	50.4 ± 2.4	74.4 ± 0.9	76.6 ± 0.8	65.7
MMD [Yu *et al.*, 2024]	60.4 ± 0.2	53.3 ± 0.3	74.3 ± 0.1	77.4 ± 0.6	66.4
MLDG [Li *et al.*, 2018a]	61.5 ± 0.9	53.2 ± 0.6	75.0 ± 1.2	77.5 ± 0.4	66.8
IRM [Ahuja *et al.*, 2020]	58.9 ± 2.3	52.2 ± 1.6	72.1 ± 2.9	74.0 ± 2.5	64.3
RSC [Huang *et al.*, 2020]	60.7 ± 1.4	51.4 ± 0.3	74.8 ± 1.1	75.1 ± 1.3	65.5
SupCon [Khosla *et al.*, 2020]	66.0 ± 0.2	57.0 ± 0.3	78.8 ± 0.3	80.6 ± 0.2	70.6
MTL [Blanchard *et al.*, 2021]	61.5 ± 0.7	52.4 ± 0.6	74.9 ± 0.4	76.8 ± 0.4	66.4
ERM [Vapnik, 1999]	63.1 ± 0.3	51.9 ± 0.4	77.2 ± 0.5	78.1 ± 0.2	67.6
Mixstyle [Zhou *et al.*, 2021]	51.1 ± 0.3	53.2 ± 0.4	68.2 ± 0.7	69.2 ± 0.6	60.4
SagNet [Nam *et al.*, 2021]	63.4 ± 0.2	54.8 ± 0.4	75.8 ± 0.4	78.3 ± 0.3	68.1
SWAD [Cha *et al.*, 2021]	66.1 ± 0.4	57.7 ± 0.4	78.4 ± 0.1	80.2 ± 0.2	70.6
PCL [Yao *et al.*, 2022]	67.3 ± 0.2	**59.9 ± 0.1**	78.7 ± 0.2	80.7 ± 0.1	71.6
IAIE-Pair[Hu *et al.*, 2024]	65.8 ± 0.4	57.9 ± 0.2	79.1 ± 0.3	80.8 ± 0.1	70.9
IAIE-Proxy[Hu *et al.*, 2024]	67.5 ± 0.2	58.5 ± 0.1	**79.7 ± 0.3**	**81.3 ± 0.2**	71.8
Ours	**70.2 ± 0.1**	59.5 ± 0.2	78.5 ± 0.1	81.2 ± 0.2	**72.3**

Table 4: Comparison of ACC (%) with the SOTA methods on Office-Home.

Method	L100	L38	L43	L46	Average
CORAL [Sun and Saenko, 2016]	51.6 ± 2.4	42.2 ± 1.0	57.0 ± 1.0	39.8 ± 2.9	47.7
DANN [Ganin *et al.*, 2016]	51.1 ± 3.5	40.6 ± 0.6	57.4 ± 0.5	37.7 ± 1.8	46.7
CDANN [Li *et al.*, 2018b]	47.0 ± 1.9	41.3 ± 4.8	54.9 ± 1.7	79.8 ± 2.3	45.8
MMD [Yu *et al.*, 2024]	41.9 ± 3.0	34.8 ± 1.0	57.0 ± 1.9	35.2 ± 1.8	42.2
MLDG [Li *et al.*, 2018a]	54.2 ± 3.0	44.3 ± 1.1	55.6 ± 0.3	36.9 ± 2.2	47.8
IRM [Ahuja *et al.*, 2020]	54.6 ± 1.3	39.8 ± 1.9	56.2 ± 1.8	39.6 ± 0.8	47.6
RSC [Huang *et al.*, 2020]	50.2 ± 2.2	39.2 ± 1.4	56.3 ± 1.4	40.8 ± 0.6	46.6
SupCon [Khosla *et al.*, 2020]	61.4 ± 1.0	46.0 ± 2.2	57.9 ± 0.7	37.5 ± 0.9	50.7
MTL [Blanchard *et al.*, 2021]	49.3 ± 1.2	39.6 ± 6.3	55.6 ± 1.1	37.8 ± 0.8	45.6
ERM [Vapnik, 1999]	54.3 ± 0.4	42.5 ± 0.7	55.6 ± 0.3	38.8 ± 2.5	47.8
Mixstyle [Zhou *et al.*, 2021]	54.3 ± 1.1	34.1 ± 1.1	55.9 ± 1.1	31.7 ± 2.1	44.0
SagNet [Nam *et al.*, 2021]	53.0 ± 2.9	43.0 ± 2.5	57.9 ± 0.6	40.4 ± 1.3	48.6
SWAD [Cha *et al.*, 2021]	55.4 ± 0.0	44.9 ± 1.1	59.7 ± 0.4	39.9 ± 0.2	50.0
PCL [Yao *et al.*, 2022]	58.7 ± 0.7	46.3 ± 1.5	**60.0 ± 0.7**	**43.6 ± 0.6**	52.1
IAIE-Pair[Hu *et al.*, 2024]	61.3 ± 1.0	50.0 ± 1.8	57.6 ± 0.6	40.3 ± 1.1	52.2
IAIE-Proxy[Hu *et al.*, 2024]	59.5 ± 1.4	50.9 ± 1.0	58.5 ± 0.3	41.6 ± 0.6	52.6
Ours	**61.9 ± 1.3**	**53.0 ± 1.7**	57.3 ± 1.4	43.3 ± 2.2	**53.9**

Table 5: Comparison of ACC (%) with the SOTA methods on TerraIncognita.

Model		PACS				
SCA	SDK	Art	Cartoon	Photo	Sketch	Avg
×	×	87.5 ± 1.1	84.2 ± 0.4	97.0 ± 0.8	81.0 ± 0.8	87.4
×	✓	90.8 ± 0.7	82.6 ± 1.0	96.9 ± 0.3	84.5 ± 1.1	88.7
✓	×	89.6 ± 0.5	83.2 ± 0.6	**97.2 ± 0.2**	82.4 ± 0.1	88.1
✓	✓	**90.9 ± 0.2**	**84.6 ± 0.1**	97.1 ± 0.3	**84.8 ± 0.6**	**89.3**

Table 6: Ablation experiments on PACS based on ResNet-50 ImageNet pre-training.

[Vapnik, 1999], IRM [Ahuja *et al.*, 2020], and adversarial learning-based methods like DANN [Ganin *et al.*, 2016] and CDANN [Li *et al.*, 2018b]), our method improves cross-domain feature learning and generalization capability by utilizing indirect alignment and maintaining sample diversity.

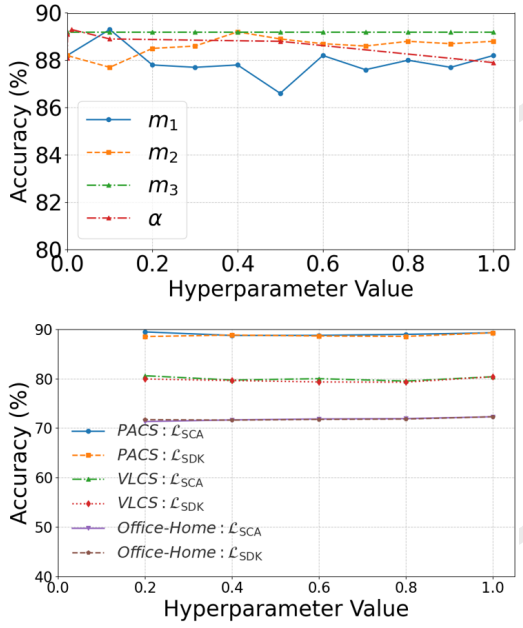


Figure 4: Hyperparameter analysis on PACS datasets.

When compared to other advanced methods based on contrastive learning, such as SelfReg [Kim *et al.*, 2021], PCL [Yao *et al.*, 2022], HTCL [Tong *et al.*, 2023], and IAIE [Hu *et al.*, 2024], our approach effectively avoids overfitting and misalignment issues that arise from direct alignment. This is achieved through the preservation of natural sequence relationships and the indirect alignment strategy. Notably, on particularly challenging domains (e.g., the "L38" domain in the TerraIncognita dataset and the "LabelMe" domain in the VLCS dataset), where the target domain exhibits significant differences from the source domains, other methods perform poorly. In contrast, our method significantly improves the model's performance in these scenarios, demonstrating its superiority and robustness in handling complex domain generalization tasks.

4.4 Ablation Studies

To better understand the contributions of the Sample Difference Keeping (SDK) and Sample Consistency Alignment (SCA) modules, we conducted ablation experiments to evaluate their effects on model performance (Figure 4).

Component Ablation. Using the PACS dataset, we first examined the individual and combined impacts of the SDK and SCA modules. Starting with the baseline containing only the classification loss, we sequentially added SDK and SCA modules. The results, shown in Table 6, reveal the following: Adding the SCA module alone significantly improved performance, confirming its role in enhancing cross-domain feature alignment through class-wise and domain-wise consistency. However, it also introduced a risk of overfitting. Adding the SDK module alone led to moderate improvement by mitigat-

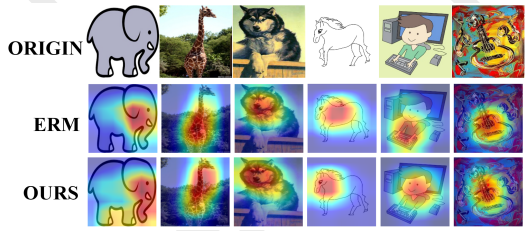


Figure 5: Grad-CAM visualization results on the PACS dataset. The first row shows the original input images. The second row shows the Grad-CAM visualization results of the baseline method (ERM), where the focus is mainly on non-discriminative regions, such as the elephant's body and face. The third row shows the Grad-CAM visualization results of our method (OURS), which accurately highlights key discriminative features, such as the elephant's trunk and the core areas in other images.

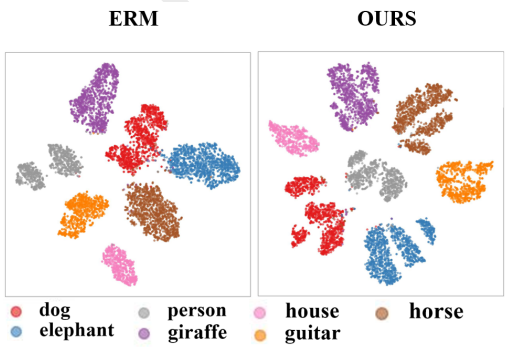


Figure 6: Visualization on the PACS dataset, with different colors representing different classes, the baseline method on the left and our method on the right.

ing reliance on domain-specific features and emphasizing relative differences between samples, though it slightly compromised cross-domain alignment. Combining SDK and SCA achieved the best performance, demonstrating a synergistic effect: SDK reduces overfitting risk from SCA by preserving natural sequence relationships, while SCA strengthens cross-domain alignment. Together, these modules enhance both generalization and stability.

Parameter Sensitivity Analysis. We further analyzed the impact of key hyperparameters, including the triplet loss margins (m_1, m_2, m_3) and the consistency loss weight α.

Sensitivity to α: Fixing margin parameters, we observed that model performance peaked when $\alpha = 0.01$, indicating that an appropriate balance between class-wise and domain-wise consistency losses is crucial.

Sensitivity to margin parameters: Fixing other hyperparameters, the model showed low sensitivity to m_1, m_2, and m_3. Notably, m_3 exhibited minimal sensitivity, and no alternative weight settings outperformed the default value of 1.

Weight Analysis of SDK and SCA Losses. We further investigated how different weightings of the SDK and SCA loss terms affect performance. Although the optimal balance between them varied slightly from one dataset to another, the overall impact on accuracy remained small. To maintain consistency and simplicity across all experiments, we therefore fixed both weights at 1.

4.5 Visualization Analysis

To intuitively demonstrate the effectiveness of our proposed method, we performed both Grad-CAM[Selvaraju *et al.*, 2017] and feature space visualizations on the PACS dataset.

Firstly, Grad-CAM visualization (Figure 5) highlights the key regions the model relies on for classification. For the cartoon domain, the baseline method (ERM) focuses on less discriminative regions, such as the elephant's body and face. In contrast, our method (OURS) effectively captures the key discriminative feature of the elephant's trunk, which plays a critical role in accurate classification. This observation indicates that our approach avoids overfitting to domain-specific features and instead learns semantically meaningful features that generalize well across domains.

Secondly, We used t-SNE [Van der Maaten and Hinton, 2008] to reduce the dimensionality of the extracted features, as shown in Figure 6. The results show that the features extracted by the ERM method are overly compact within each class, with minimal separation between different classes. This could lead to a decline in generalization ability. In contrast, our method resulted in more dispersed features with greater inter-class separation than the baseline ERM method. This is because our proposed SDK module preserves the natural sequence relationships of samples within each class, preventing excessive compression of intra-class features. Additionally, the SCA module effectively extracts cross-domain common features, further enhancing inter-class separability and resulting in clearer classification boundaries. This validates that, while ensuring intra-class consistency, our method effectively enhances inter-class diversity, thereby improving the model's cross-domain generalization ability.

5 Conclusion

In this paper, we propose a novel domain generalization framework to address the shortcomings of traditional contrastive learning methods, which often oversimplify sample relationships, leading to misalignment, feature distortion, and overfitting. The framework introduces two key modules: the Sample Difference Keeping (SDK) module, which preserves natural sequence relationships, and the Sample Consistency Alignment (SCA) module, which achieves indirect alignment through inter-class and inter-domain relationships consistency. These innovations significantly enhance feature representation quality and generalization performance. Extensive experiments on multiple benchmark datasets demonstrate the framework's effectiveness and robustness, consistently outperforming state-of-the-art methods in addressing domain discrepancies.

Despite its notable achievements, the framework has yet to fully explore hierarchical relationships between samples and lacks dynamic sample weighting strategies to better adapt to complex domain shifts. Future research will focus on integrating hierarchical sample relationships and dynamic weighting mechanisms to improve adaptability.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (No.62276160), the Fundamental Research Program of Shanxi Province (Nos.202203021211291, 202303021222030), and Fund Program for the Scientific Activities of Selected Returned Overseas Professionals in Shanxi Province (No.20240002).

References

[Ahuja *et al.*, 2020] Kartik Ahuja, Karthikeyan Shanmugam, Kush Varshney, and Amit Dhurandhar. Invariant risk minimization games. In *International Conference on Machine Learning*, pages 145–155. PMLR, 2020.

[Beery *et al.*, 2018] Sara Beery, Grant Van Horn, and Pietro Perona. Recognition in terra incognita. In *Proceedings of the European conference on computer vision (ECCV)*, pages 456–473, 2018.

[Blanchard *et al.*, 2021] Gilles Blanchard, Aniket Anand Deshmukh, Urun Dogan, Gyemin Lee, and Clayton Scott. Domain generalization by marginal transfer learning. *Journal of machine learning research*, 22(2):1–55, 2021.

[Cakir *et al.*, 2019] Fatih Cakir, Kun He, Xide Xia, Brian Kulis, and Stan Sclaroff. Deep metric learning to rank. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1861–1870, 2019.

[Cha *et al.*, 2021] Junbum Cha, Sanghyuk Chun, Kyungjae Lee, Han-Cheol Cho, Seunghyun Park, Yunsung Lee, and Sungrae Park. Swad: Domain generalization by seeking flat minima. *Advances in Neural Information Processing Systems*, 34:22405–22418, 2021.

[Chen *et al.*, 2023a] Jin Chen, Zhi Gao, Xinxiao Wu, and Jiebo Luo. Meta-causal learning for single domain generalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7683–7692, 2023.

[Chen *et al.*, 2023b] Zining Chen, Weiqiu Wang, Zhicheng Zhao, Fei Su, Aidong Men, and Yuan Dong. Instance paradigm contrastive learning for domain generalization. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(2):1032–1042, 2023.

[Dong *et al.*, 2022] Benyu Dong, Chung-Ming Own, and Qinghua Hu. Domain generalization via maximum intra-class average diameter minimization. In *2022 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2022.

[Fang *et al.*, 2013] Chen Fang, Ye Xu, and Daniel N Rockmore. Unbiased metric learning: On the utilization of multiple datasets and web images for softening bias. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1657–1664, 2013.

[Fu *et al.*, 2021] Zheren Fu, Zhendong Mao, Chenggang Yan, An-An Liu, Hongtao Xie, and Yongdong Zhang. Self-supervised synthesis ranking for deep metric learning. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(7):4736–4750, 2021.

[Ganin *et al.*, 2016] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario March, and Victor Lempitsky. Domain-adversarial training of neural networks. *Journal of machine learning research*, 17(59):1–35, 2016.

[He *et al.*, 2016] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[He *et al.*, 2020] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9729–9738, 2020.

[He *et al.*, 2023] Huan He, Owen Queen, Teddy Koker, Consuelo Cuevas, Theodoros Tsiligkaridis, and Marinka Zitnik. Domain adaptation for time series under feature and label shifts. In *International Conference on Machine Learning*, pages 12746–12774. PMLR, 2023.

[Hu *et al.*, 2024] Jiajun Hu, Lei Qi, Jian Zhang, and Yinghuan Shi. Domain generalization via inter-domain alignment and intra-domain expansion. *Pattern Recognition*, 146:110029, 2024.

[Huang *et al.*, 2020] Zeyi Huang, Haohan Wang, Eric P Xing, and Dong Huang. Self-challenging improves cross-domain generalization. In *Computer vision–ECCV 2020: 16th European conference, Glasgow, UK, August 23–28, 2020, proceedings, part II 16*, pages 124–140. Springer, 2020.

[Khosla *et al.*, 2020] Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. Supervised contrastive learning. *Advances in neural information processing systems*, 33:18661–18673, 2020.

[Kim *et al.*, 2021] Daehee Kim, Youngjun Yoo, Seunghyun Park, Jinkyu Kim, and Jaekoo Lee. Selfreg: Self-supervised contrastive regularization for domain generalization. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9619–9628, 2021.

[Li *et al.*, 2017] Da Li, Yongxin Yang, Yi-Zhe Song, and Timothy M Hospedales. Deeper, broader and artier domain generalization. In *Proceedings of the IEEE international conference on computer vision*, pages 5542–5550, 2017.

[Li *et al.*, 2018a] Da Li, Yongxin Yang, Yi-Zhe Song, and Timothy Hospedales. Learning to generalize: Meta-learning for domain generalization. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018.

[Li *et al.*, 2018b] Ya Li, Xinmei Tian, Mingming Gong, Yajing Liu, Tongliang Liu, Kun Zhang, and Dacheng Tao. Deep domain generalization via conditional invariant adversarial networks. In *Proceedings of the European conference on computer vision (ECCV)*, pages 624–639, 2018.

[Li *et al.*, 2023] Mingkang Li, Jiali Zhang, Wen Zhang, Lu Gong, and Zili Zhang. Style augmentation and domain-aware parametric contrastive learning for domain generalization. In *International Conference on Knowledge Science, Engineering and Management*, pages 211–224. Springer, 2023.

[Motiian *et al.*, 2017] Saeid Motiian, Marco Piccirilli, Donald A Adjeroh, and Gianfranco Doretto. Unified deep supervised domain adaptation and generalization. In *Proceedings of the IEEE international conference on computer vision*, pages 5715–5725, 2017.

[Nam *et al.*, 2021] Hyeonseob Nam, HyunJae Lee, Jongchan Park, Wonjun Yoon, and Donggeun Yoo. Reducing domain gap by reducing style bias. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8690–8699, 2021.

[Russakovsky *et al.*, 2015] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115:211–252, 2015.

[Schroff *et al.*, 2015] Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 815–823, 2015.

[Selvaraju *et al.*, 2017] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*, pages 618–626, 2017.

[Su *et al.*, 2023] Zixian Su, Kai Yao, Xi Yang, Kaizhu Huang, Qiufeng Wang, and Jie Sun. Rethinking data augmentation for single-source domain generalization in medical image segmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 2366–2374, 2023.

[Sun and Saenko, 2016] Baochen Sun and Kate Saenko. Deep coral: Correlation alignment for deep domain adaptation. In *Computer Vision–ECCV 2016 Workshops: Amsterdam, The Netherlands, October 8-10 and 15-16, 2016, Proceedings, Part III 14*, pages 443–450. Springer, 2016.

[Tong *et al.*, 2023] Yunze Tong, Junkun Yuan, Min Zhang, Didi Zhu, Keli Zhang, Fei Wu, and Kun Kuang. Quantitatively measuring and contrastively exploring heterogeneity for domain generalization. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 2189–2200, 2023.

[Touvron *et al.*, 2022] Hugo Touvron, Piotr Bojanowski, Mathilde Caron, Matthieu Cord, Alaaeldin El-Nouby, Edouard Grave, Gautier Izacard, Armand Joulin, Gabriel Synnaeve, Jakob Verbeek, et al. Resmlp: Feedforward networks for image classification with data-efficient training. *IEEE transactions on pattern analysis and machine intelligence*, 45(4):5314–5321, 2022.

[Van der Maaten and Hinton, 2008] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(11), 2008.

[Vapnik, 1999] Vladimir N Vapnik. An overview of statistical learning theory. *IEEE transactions on neural networks*, 10(5):988–999, 1999.

[Venkateswara *et al.*, 2017] Hemanth Venkateswara, Jose Eusebio, Shayok Chakraborty, and Sethuraman Panchanathan. Deep hashing network for unsupervised domain adaptation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5018–5027, 2017.

[Wang *et al.*, 2019] Xinshao Wang, Yang Hua, Elyor Kodirov, Guosheng Hu, Romain Garnier, and Neil M Robertson. Ranked list loss for deep metric learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5207–5216, 2019.

[Wang *et al.*, 2024] Mengzhu Wang, Yuehua Liu, Jianlong Yuan, Shanshan Wang, Zhibin Wang, and Wei Wang. Inter-class and inter-domain semantic augmentation for domain generalization. *IEEE Transactions on Image Processing*, 33:1338–1347, 2024.

[Yao *et al.*, 2022] Xufeng Yao, Yang Bai, Xinyun Zhang, Yuechen Zhang, Qi Sun, Ran Chen, Ruiyu Li, and Bei Yu. Pcl: Proxy-based contrastive learning for domain generalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7097–7107, 2022.

[Yu *et al.*, 2024] Xi Yu, Huan-Hsin Tseng, Shinjae Yoo, Haibin Ling, and Yuewei Lin. Insure: an information theory inspired disentanglement and purification model for domain generalization. *IEEE Transactions on Image Processing*, 2024.

[Zhang *et al.*, 2023a] Hanqing Zhang, Haolin Song, Shaoyu Li, Ming Zhou, and Dawei Song. A survey of controllable text generation using transformer-based pre-trained language models. *ACM Computing Surveys*, 56(3):1–37, 2023.

[Zhang *et al.*, 2023b] Jiao Zhang, Xu-Yao Zhang, Chuang Wang, and Cheng-Lin Liu. Deep representation learning for domain generalization with information bottleneck principle. *Pattern Recognition*, 143:109737, 2023.

[Zhou *et al.*, 2021] Kaiyang Zhou, Yongxin Yang, Yu Qiao, and Tao Xiang. Domain generalization with mixstyle. *arXiv preprint arXiv:2104.02008*, 2021.

[Zou *et al.*, 2023] Zhengxia Zou, Keyan Chen, Zhenwei Shi, Yuhong Guo, and Jieping Ye. Object detection in 20 years: A survey. *Proceedings of the IEEE*, 111(3):257–276, 2023.