

基于跨空间多尺度信息聚合和推理一致性的域泛化方法

黎拓新, 项凤涛, 陈君海, 张晓博, 吕云霄

(国防科技大学智能科学学院, 湖南 长沙 410073)

摘要: 在机器学习中, 通常模型假设训练数据和测试数据来自相同的分布, 然而在实际应用中, 数据分布往往存在差异, 导致域漂移问题, 进而影响模型的泛化能力。现有域泛化方法主要集中在提取域不变特征, 忽视了域特定特征对模型预测的影响。为此, 提出了一种基于跨空间多尺度信息聚合的域鉴别器, 通过多尺度信息捕捉, 有效去除域特定特征, 增强域不变特征的提取能力; 同时通过动量更新推理一致性损失函数, 利用样本类别中心的推理一致性, 进一步提高模型的鲁棒性。在多个公开数据集上进行了对比实验与分析, 结果表明, 新方法在域泛化中具有更好的性能表现, 有效缓解了域特定特征对模型性能的影响, 可为处理域漂移问题提供技术参考。

关键词: 域泛化; 迁移学习; 域漂移; 注意力机制; 推理一致性

中图分类号: TP391.4

文献标志码: A

doi: 10.11959/j.issn.2096-6652.202516

Domain generalization method based on cross-space multi-scale information aggregation and inference consistency

LI Tuoxin, XIANG Fengtao, CHEN Junhai, ZHANG Xiaobo, LYU Yunxiao

College of Intelligence Science and Technology, National University of Defense Technology, Changsha 410073, China

Abstract: In machine learning, it typically assumes that training data and testing data of models are drawn from the same distribution. However, in real-world applications, data distributions often differ, resulting in domain shift problems that adversely affect model generalization. Existing domain generalization methods primarily focus on extracting domain-invariant features while overlooking the potential impact of domain-specific features on model predictions. To address this issue, a domain discriminator based on cross-space multi-scale information aggregation was proposed. By capturing multi-scale information, domain-specific features were effectively removed and the extraction of domain-invariant features was enhanced. Additionally, the momentum update inference consistency loss function was employed to leverage the inference consistency of sample category centers, further improving model robustness. Comparative experiments and analysis conducted on multiple public datasets demonstrate that the proposed method exhibits superior performance in domain generalization, effectively mitigating the impact of domain-specific features on model performance and providing a technical reference for addressing domain shift problems.

Key words: domain generalization, transfer learning, domain drift, attention mechanism, inference consistency

0 引言

近年来, 卷积神经网络 (convolutional neural

network, CNN) 因其卓越的性能, 已成为机器学习领域的重要工具。然而, 大多数现有的机器学习算法基于分布不变假设, 即训练数据和测试数据被

收稿日期: 2025-01-21; 修回日期: 2025-04-11

通信作者: 项凤涛, xiangfengtao@nudt.edu.cn

基金项目: 国家自然科学基金项目 (No.62473371)

Foundation Item: The National Natural Science Foundation of China (No.62473371)

视为独立同分布^[1]。尽管这一假设在许多理想化场景中成立，但在实际应用中却经常失效。实际应用中，模型训练时使用的源域数据与测试时的目标域数据在分布上往往存在显著差异（即“域漂移”问题^[2]），这种差异极大地限制了模型在真实场景下的推广能力。例如，同一类别的数据在不同域中的表现形式可能存在巨大的视觉差异，不同域中图片形象差异如图1所示，这种域差异会导致训练好的模型难以在目标域上取得良好的性能。



图1 不同域中图片形象差异

为了应对域漂移问题，研究者提出了域适应（domain adaptation, DA）技术^[3-6]，其主要目标是在模型训练过程中引入部分目标域数据，利用目标域的有限标记或未标记数据优化模型性能。然而，在许多实际场景中，目标域数据往往不可获取或获取成本较高，这使得域适应技术难以在更广泛的领域中得到应用。为了解决这一问题，域泛化（domain generalization, DG）技术应运而生。域泛化旨在利用源域数据学习可迁移的知识，使模型在未知的目标域中也能取得良好的性能。相比域适应，域泛化不依赖于目标域数据，因而在许多实际应用中具有更强的适用性。

域泛化的研究历程可以追溯到早期的理论分析阶段。早期研究^[7-10]多基于核方法对问题进行建模和泛化误差分析，并通过传统的机器学习方法验证了理论预测的有效性。例如，Ye等^[11]提出了跨域特征不变性、分类任务的特征可区分性以及特征扩展函数3个概念，对域泛化问题的理论限制进行了深入探讨；Chen等^[12]通过引入基本原理的概念，扩展了经验风险最小化（empirical risk minimization, ERM）^[13]模型，并对其在域泛化任务中的性能进行

了分析。此外，Vedantam等^[14]提出了与费舍尔信息、预测熵以及最大平均差异（maximum mean discrepancy, MMD）相关的度量方法，这些度量方法可用于评估ERM模型在分布外泛化场景中的能力。

在方法层面，域泛化研究主要集中于两大方向：域对齐与数据增强。

（1）域对齐的核心思想是通过学习跨域不变性，使源域和目标域在特征空间中分布一致。例如，基于最大平均差异的度量方法可有效量化源域和目标域间的分布差异^[15]；对比损失^[16]和对抗学习^[17]则通过优化特定目标函数实现源域分布的对齐。Arjovsky等^[18]通过强制最优分类器在所有域上保持一致性，实现了跨域不变性学习；Ghifary等^[19]通过多任务自动编码器网络提取源域之间共享的特征，从而减少数据集偏差；Li等^[17]提出了一种端到端条件不变深度神经网络，以最小化跨域条件分布差异；Mahajan等^[20]假定域是通过混合因果特征和非因果特征生成的，并提出通过相同对象的表示一致性来优化特征学习；Guo等^[21]则通过丢弃域特定通道，有效实现了域不变特征的提取；倪云昊等^[22]利用不变性风险最小化的思想，提升了模型域泛化的效果。

（2）数据增强方法则致力于通过丰富源域数据的多样性来提升模型的泛化能力。例如，图像级别的数据增强技术^[23]通过引入合成样本丰富数据分布，特征级别的增强技术^[24]则侧重于扩展特征空间的覆盖范围。受集成学习思想的启发，研究者提出了多种基于域增强的集成学习方法^[25-27]，通过训练多个子模型（如特定域分类器或批量归一化层），再通过集成预测来提升模型的泛化性能。此外，童煜钧等^[28]将扩散模型应用到域泛化方法中，通过对特征的精细控制和多样性生成，提升目标域的泛化能力。

尽管现有方法在域泛化任务中取得了一定进展，仍存在显著不足。许多研究仅关注提取域不变特征，而忽视了域特定特征在模型性能提升中的潜在作用。此外，这些方法在去除域特定特征时缺乏明确的指导机制，可能导致模型在目标域上学习过多域相关信息，从而限制泛化性能。基于此，本文提出了一种结合注意力机制和域不变特征提取的域泛化方法。注意力机制的核心思想是根据每个特征通道对最终任务的贡献程度计算权重，并将其作用于原有的特征输入，从而使网络更专注于与当前任

务最相关的特征。著名的注意力机制有SENet^[29]、CBAM^[30]、EMA^[31]等。首先,本文设计了一种基于跨空间多尺度信息聚合注意力机制的域鉴别器(cross-space multi-scale attention domain discriminator, CMADD),通过多尺度并行网络捕捉短程和长程依赖,识别并去除域特定特征。同时,域鉴别器利用特征通道对域识别贡献的评分机制,主动过滤与域强相关的特征。其次,为了增强特征的鲁棒性,本文设计了基于动量更新的推理一致性损失(momentum-based inference consistency loss, MICL)函数,通过对同类样本推理一致性的约束施加损失项,确保分类决策的稳定性。最后,在PACS^[32]、VLCS^[33]、OfficeHome^[34]3个公共数据集上进行对比实验,实验结果验证了本文方法的有效性。

1 方法

1.1 问题定义

假设有 N 个分布不同的源域 $\mathcal{D}_s = \{D_s^1, D_s^2, \dots, D_s^N\}$, 对于每个域, $D_s^i = \{(x_j^i, y_j^i)\}_{j=1}^{n_i}$, 域标签为 y_n , 其中, (x_j^i, y_j^i) 表示第 i 个域中第 j 个样本的数据标签对, n_i 是 D_s^i 中的样本数量。所有源域和目标域共享相同的类标签空间, 域泛化的目标是只使用源域数据训练一个模型 F , 使其在未见过的目标域 $\mathcal{D}_t$ 上表现良好。本文方法主要由两个部分组成: CMADD和MICL。本文方法总流程如图2所示。在每个训练批次中, 随机将CMADD模块插入骨干网络的中间层中。该模块结合注意力机制与域鉴别器, 识别出对特定领域贡献最大的特征通道, 并将其去除。

为了提取更稳健的特征, MICL通过动量更新原型矩阵, 确保每个样本尽可能接近其类别中心。

1.2 基于跨空间多尺度信息聚合注意力机制的域鉴别器

深度神经网络广泛应用于机器学习的各种领域。然而, 当测试数据与训练数据的分布不一致时, 模型性能通常会显著下降, 极大地限制了深度神经网络的实际应用。在多源域的域泛化场景中, 源域上训练的模型通常包含许多特定于源域的非鲁棒特征, 当域发生变化时, 这些非鲁棒特征会在未训练的目标域上产生负面影响, 从而导致条件分布发生变化。传统的域泛化方法并未明确指导模型去除中间网络层中的域特定特征, 可能导致模型学习过多与域相关的信息。

本文方法的目标在于显式地去除域特定特征, 同时增强域不变特征的提取能力。为此, 本文提出了一种基于跨空间多尺度信息聚合注意力机制的域鉴别器, 识别并去除域特定特征通道。

在本文任务中, 需要识别并去除每个域中最具特异性的特征。注意力机制有助于完成这一任务, 因为不同通道特征对最终分类的贡献程度不同, 而每个通道特征在深度特征图中对应于输入图像的特定信息, 这些信息往往包含了域特定特征。

基于此, 本文提出了一种跨空间多尺度信息聚合的注意力机制, 具体结构如图3所示, 与坐标注意力(coordinate attention, CA)^[35]相比, 本文方法将特征通道分为多个子特征组, 使空间语义信息在每个特征组内分布均匀, 减少计算开销。同时利用多尺度卷积配合全局池化操作, 通过 1×1 和 3×3

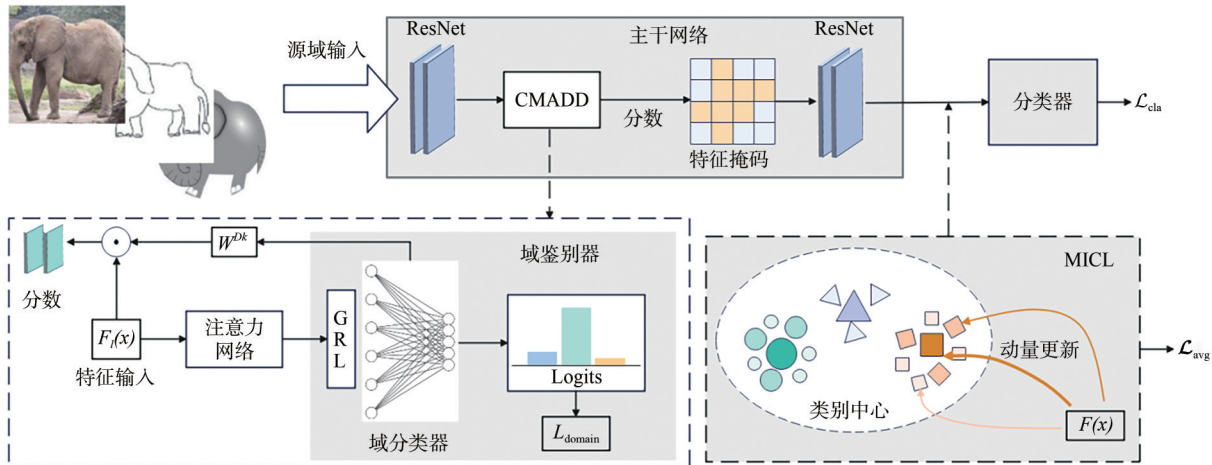


图2 本文方法总流程

两个不同大小的卷积核，并采用1D全局平均池化和全局最大池化，有效捕捉局部和全局的空间特征，建立不同方向上的短程和长程依赖。最后将不同尺度的注意力信息进行聚合，调整不同特征通道的重要性，将精确的空间结构信息保留在特征通道中，这种跨空间的信息聚合能力使模型可以在域泛化任务中更有效地捕捉域特定特征。

为了在训练过程中使模型准确地去除域特定特征，本文方法在特征提取网络的多个中间层引入了域鉴别器来定位域特定特征。给定一个来自源域 D_n 的输入 x_i^n 及其标签 y_i^n ， i 代表源域中第 i 个数据。本文方法首先提取由第 l 个中间层产生的特征 $F_l(x_i^n) \in \mathbb{R}^{C \times H \times W}$ ，其中 C 是通道数， H 和 W 分别表示高度和宽度维度。 $F_l(x_i^n)$ 后接域鉴别器 F_d^l ，以预测域标签并计算区分损失，其中 F_d 表示域鉴别器， F_d^l 表示第 l 个中间层的域鉴别器。同时，本文方法在域鉴别器之前使用梯度反转层（GRL）^[36] 来进行对抗训练，从而使特征提取器提取出更加通用的特征。

CMADD 与传统域泛化方法的不同之处在于，不断过滤掉对特定域识别贡献大的特征通道，以防止模型保留太多与域相关的特征。通过在训练过程中去除域特定特征，本文方法降低了由域偏移引起的特征通道不稳定性，并促进了域不变特征的学习。为了确定哪些特征包含特定于域的信息，本文使用中间层域鉴别器的预测结果作为衡量特征通道重要性的指标。具体来说，本文假设对域预测贡献最大的通道可能包含较多特定于域的信息。首先进行域标签预测：

$$y_k = W^{D_n} \cdot \text{GAP}\left(F_l(x_i^n)\right) \quad (1)$$

其中， x_i^n 为输入， $F_l(x_i^n)$ 为特征， W^{D_n} 为域分类器权重，GAP 代表全局平均池化。同时进行域预测损失的计算：

$$\mathcal{L}_{\text{domain}} = \text{CrossEntropy}(y_k, y_n) \quad (2)$$

其中，CrossEntropy 为交叉熵损失， y_k 和 y_n 分别为预测标签和域标签。之后通过计算正确的域预测的加权激活来量化第 j 个通道和域特定信息之间的相关性：

$$s_j = W_j^{D_n} \odot \text{GAP}\left(F_l(x_i^n)\right)_j \quad (3)$$

其中， $W_j^{D_n}$ 为域鉴别器对应于域 D_n 的全连接层权重。权重值越高，相应的特征通道对域预测的贡献越大。

在对相应的特征通道进行识别后，为了显式地减少特征图中特定于域的信息，训练过程中选择并去除对域识别贡献最大的通道。具体来说，首先计算去除特征通道的数目：

$$M = C \times P_{\text{drop}} \quad (4)$$

其中， P_{drop} 为特征通道去除率。随后，基于每个特征通道的得分生成二进制特征掩码 m ，根据掩码的值是否为1决定相应通道是否被去除。具体的特征掩码 m_j 被设置为1的概率由 s_j 决定，具有相对高 s_j 的特征通道更可能被去除。为了降低计算成本，本文采用加权随机选择（WRS）算法^[37] 来生成二进制掩码 m 。具体地，对于分数为 s_j 的第 j 个通道，本文方法首先生成随机数 $r_j \sim U(0, 1)$ 并计算关键值

$k_j = \frac{1}{s_j}$ ；然后，选择具有最大键值的 M 个值，并将掩码 m 的相应 m_j 设置为0：

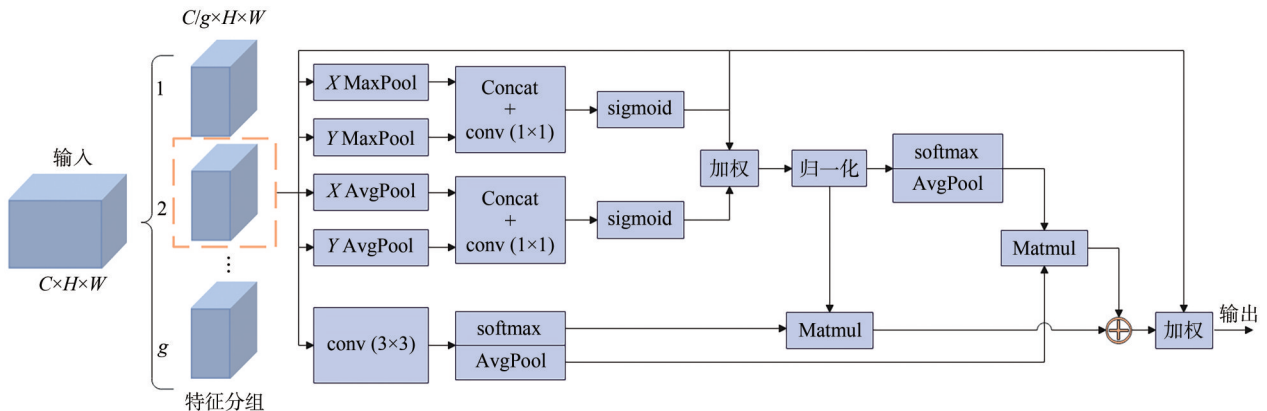


图3 跨空间多尺度信息聚合注意力示意图

$$m_j = \begin{cases} 0, & j \in \text{TOP}(\{k_1, k_2, \dots, k_C\}, M) \\ 1, & \text{其他} \end{cases} \quad (5)$$

其中, $\text{TOP}(\{k_1, k_2, \dots, k_C\}, M)$ 表示具有最大关键值的 M 个值。最后, 将特征掩码作用于原有特征得到新特征 f :

$$f = m \cdot F_i(x_i^n) \quad (6)$$

1.3 基于动量更新的推理一致性损失函数

在先前的步骤中, 通过特征掩码得到了新特征 $f \in \mathbb{R}^C$ 。如果只关注特征, 而不充分考虑与特征相对应的分类器权重, 会导致最后的分类结果不准确。例如, 一个较为重要的特征可能对应一个较小的分类权重, 这种对特征重要性的有偏不是笔者希望看到的。

与特征相比, Logits 涉及对特征加权, 在一定程度上考虑了分类权重。然而, Logits 只提供了一个粗略的分类决策值, 仅仅依靠 Logits 可能无法控制决策过程, 导致无法有效获得可靠的结果。因此, 为了得到泛化性更强的模型, 需要同时考虑特征和分类权重。假设有 K 个类别, Logits 可以表示为 $o = W^T f$, 其中 W 为分类器权重。为了更深入地研究特征和分类器权重对最终分类的影响, 可以表示由特征和分类器权重之间的元素乘积之和生成的第 K 类 Logits 值 o_k 如下:

$$o_k = \sum_{j=1}^D W_{\{j,k\}} f_j \quad (7)$$

其中, K 类的概率是所有 $W_{\{j,k\}} f_j$ 的累加, 可以被视为对 o_k 的逐元素的贡献。直观地, $W_{\{j,k\}} f_j$ 提供了一条证据来支持具体的决定。因此, 可以收集所有 $W_{\{j,k\}} f_j$ 作为矩阵 $R \in \mathbb{R}^{C \times K}$ 来表示对样本进行分类的准则, 本文称之为原型矩阵, 其形式定义如下:

$$R = \begin{bmatrix} W_{\{1,1\}} f_1 & W_{\{1,2\}} f_1 & \dots & W_{\{1,K\}} f_1 \\ W_{\{2,1\}} f_2 & W_{\{2,2\}} f_2 & \dots & W_{\{2,K\}} f_2 \\ \vdots & \vdots & \ddots & \vdots \\ W_{\{C,1\}} f_C & W_{\{C,2\}} f_C & \dots & W_{\{C,K\}} f_C \end{bmatrix} \quad (8)$$

本文主要目的是让模型可以在不同的环境下保持可靠的输出, 原型矩阵克服了基于特征和基于 Logits 分析的局限性, 从一个更细粒度的角度来表示分类决策过程, 即确保来自同一类别的样本的原型矩阵与其对应类别的平均值一致。这种方法意味

着, 对象进行分类的决策应该基于相同的推理, 而不考虑样本或域的变化, 因为决策的因果因素通常是跨样本和跨域持续存在的稳定模式。形式上, 本文施加推理一致性损失项, 以强制样本的原型矩阵接近其对应的平均值, 表示为 $\mathcal{L}_{\text{avg}}$, 该正则化术语可写为:

$$\mathcal{L}_{\text{avg}} = \frac{1}{N_b} \sum_k \sum_{\{n|y_n=k\}} \|R_n - \bar{R}_k\|^2 \quad (9)$$

其中, N_b 是一个批次中的样本数量, R_n 表示第 n 个样本的原型矩阵。 $\bar{R}_k$ 表示对应于第 k 类的原型矩阵的平均矩阵。然而, 计算 $\bar{R}_k$ 直接涉及对所有样本求平均, 这会占用大量的计算时间。因此, 本文在每个迭代步骤中以动量方式在线更新 $\bar{R}_k$:

$$\bar{R}_k^t = (1-m) \times \bar{R}_k^{t-1} + m \times \frac{1}{|y_n=k|} \sum_{\{n|y_n=k\}} R_n \quad (10)$$

其中, k 是迭代索引, m 是动量值。总的损失函数可以表示为:

$$\mathcal{L}_{\text{all}} = \mathcal{L}_{\text{domain}} + \mathcal{L}_{\text{cla}} + \lambda \mathcal{L}_{\text{avg}} \quad (11)$$

其中, $\mathcal{L}_{\text{cla}} = \text{CrossEntropy}(y_j^i, y_{\text{cla}}^i)$, y_{cla}^i 代表第 i 个域分类器预测的类别标签。

2 实验分析

2.1 数据集介绍

本文基于 PACS、OfficeHome 和 VLCS 3 个基准数据集对所提方法进行评估。(1) PACS^[32] 来自 4 个不同域的图像组成, Photo、Art、Cartoon 和 Sketch, 包括 7 个不同的类别和 9 991 张图像, 本文采用 PACS 提供的官方划分进行训练和验证。(2) OfficeHome 由 4 个不同域组成, 艺术 (Art)、剪贴画 (Clipart)、产品 (Product) 和现实世界 (RealWorld), 包含 65 个类别和约 15 500 张图像。仿照 JiGen^[38] 的设置, 本文将每个源域随机划分 90% 用于训练, 10% 用于验证。(3) VLCS 由 5 个类别的 10 729 张图像构成, 包含 4 个不同域, VOC 2007 (Pascal)、LabelMe、Caltech 和 Sun。参考 JiGen^[38] 的设置, 本文以 7:3 的比例将数据集划分为训练集和验证集。

2.2 实验设置

本文使用 ImageNet 预训练的 ResNet-18 和 ResNet-50 作为 PACS 和 OfficeHome 数据集的主干网络。对于 VLCS, 本文使用 ResNet-18 作为方法的主干网络。学习率设置为 0.002, Batchsize 为 64。

对于CMADD, 本文将特征通道去除率 P_{drop} 设置为0.33, 域鉴别器之前的梯度反转层^[36]的权重设置为0.25。对于MICL, 本文将动量值 m 设置为0.1, 损失权重 λ 设置为0.1。对于所有数据集, 本文使用留一策略进行评估, 即在每个数据集中使用一个域轮流作为目标域, 其他域作为源域。本文模型使用源域的训练集进行训练, 而超参数是基于验证集选择的。本文在所有源域的验证集上选择最佳模型并返回Top-1准确率。所有实验均采用单张RTX4070显卡在Windows11.0、Pytorch2.2.0、CUDA11.8的环境下完成。

2.3 不同数据集上方法对比

本文分别将ResNet-18和ResNet-50作为主干网络在PACS数据集上之前的域泛化工作进行了比较, 包括基于域不变特征提取的方法(I2 drop^[39]、CDG^[40]、DMG^[41]、Fishr^[42]、LRDG^[43]、RSC^[44])和其他先进的方法, 如基于数据增强的方法(MixStyle^[45]、EFDMix^[46]、ALOFT^[47]、StableNet^[48])等。

PACS数据集对比实验结果见表1和表2, 本文方法在两个主干网络上所有比较方法中获得了最高的平均准确率。值得注意的是, 本文方法在ResNet-18中的Cartoon和Sketch域中保持了最高的准确率。这是因为本文方法通过注意力机制的设计, 帮助域鉴别器分辨网络层中的域特定特征并进行去除。虽然本文方法在ResNet-50中只在Sketch域保持了最高准确率, 但取得了最高的平均准确率, 这说明本文方法中加入的推理一致性泛化损失函数为模型提供了稳健的决策依据, 使模型在面对多种不同的域时均表现出较好的泛化能力。在对比方法中, I2-Drop^[39]对模型学习表面局部纹理进行惩罚, 但忽略了存在其他形式的域特定特征的可能性。DMG^[41]假设一些域特定特征为目标域提供有用的信息, 这可能是由于目标域不可见, 所以该假设并不总是得到保证。RSC^[44]和CDG^[40]的实验结果证实了其较好的性能, 但仅适用于单层网络的特性阻碍了其进一步减少域特定信息。本文方法在所有对比方法中获得了最高的平均准确率, 这充分证明了它在提高模型泛化性上的有效性。

OfficeHome数据集的突出特征是具有大量类别和样本, OfficeHome上的对比实验结果见表3和表4, 本文方法分别在ResNet-18和ResNet-50主干网络上进行了评估, 通过与其他SOTA方法的对比证明

了本文方法的有效性。在ResNet-18主干网络中, 本文方法在Clipart和RealWorld两个域中保持了最高准确率; 同时在ResNet-50主干网络中, 本文方法超越了最新方法SWAD-NU和DMDA, 取得了71.81%的平均准确率, 并在Clipart、Product和RealWorld 3个域中保持了最高准确率, 进一步说明了本文方法的优越性能。

表1 PACS数据集在ResNet-18中对比实验结果

方法	准确率				
	Art	Cartoon	Photo	Sketch	平均
I ² -Drop ^[39]	80.27%	76.54%	96.11%	76.38%	82.23%
RSC ^[44]	83.43%	80.31%	95.99%	80.85%	85.15%
DMG ^[41]	80.38%	76.90%	93.35%	75.21%	81.46%
FACT ^[49]	85.37%	78.38%	95.15%	79.15%	84.51%
StableNet ^[48]	81.74%	79.91%	96.53%	80.50%	84.69%
MixStyle ^[45]	84.10%	78.80%	96.10%	75.90%	83.70%
EFDMix ^[46]	83.90%	79.40%	96.80%	75.00%	83.90%
LRDG ^[43]	81.88%	80.20%	95.21%	84.65%	85.48%
CDG ^[40]	83.50%	80.10%	95.60%	83.80%	85.80%
ALOFT ^[47]	84.81%	79.05%	96.11%	80.55%	85.13%
本文方法	84.53%	81.23%	96.64%	84.67%	86.77%

表2 PACS数据集在ResNet-50中对比实验结果

方法	准确率				
	Art	Cartoon	Photo	Sketch	平均
FACT ^[49]	89.63%	81.77%	96.75%	84.46%	88.15%
RSC ^[44]	87.89%	82.16%	97.92%	83.85%	87.83%
EFDMix ^[46]	90.60%	82.50%	98.10%	76.40%	86.90%
Fishr ^[42]	88.40%	78.70%	97.00%	77.80%	85.50%
CDG ^[40]	88.90%	83.50%	97.60%	84.90%	88.70%
ALOFT ^[47]	89.26%	83.11%	97.96%	84.04%	88.59%
CCFP ^[50]	87.50%	81.30%	96.40%	81.40%	86.60%
SWAD-NU ^[51]	89.8%	82.8%	97.7%	83.7%	88.5%
DMDA ^[52]	87.10%	85.20%	96.70%	83.50%	88.10%
本文方法	88.77%	84.09%	97.84%	86.20%	89.23%

表3 OfficeHome数据集在ResNet-18中对比实验结果

方法	准确率				
	Art	Clipart	Product	RealWorld	平均
RSC ^[44]	57.70%	48.58%	72.85%	74.39%	63.59%
FACT ^[49]	60.34%	54.85%	74.48%	76.55%	66.56%
MixStyle ^[45]	58.70%	53.40%	74.20%	75.90%	65.50%
SagNet ^[53]	60.20%	45.38%	70.42%	73.38%	62.34%
LRDG ^[43]	61.73%	52.43%	72.96%	75.89%	65.75%
CDG ^[40]	59.20%	54.30%	74.90%	75.70%	66.00%
本文方法	59.89%	55.74%	74.24%	76.79%	66.67%

表 4 OfficeHome 数据集在 ResNet-50 中对比实验结果

方法	准确率				
	Art	Clipart	Product	RealWorld	平均
RSC ^[44]	57.70%	51.40%	74.80%	75.10%	65.50%
SagNet ^[53]	63.40%	54.80%	75.80%	78.30%	68.10%
SWAD ^[54]	66.10%	57.70%	78.40%	80.20%	70.60%
Fishr ^[42]	63.40%	54.20%	76.40%	78.50%	68.20%
SWAD-NU ^[51]	67.5%	58.4%	78.6%	80.7%	71.3%
DMDA ^[52]	66.10%	56.10%	77.30%	77.90%	69.40%
本文方法	66.98%	60.58%	78.98%	80.86%	71.81%

VLCS 上的对比实验结果见表 5 和表 6, 本文方法分别在 ResNet-18 和 ResNet-50 主干网络上对 VLCS 数据集进行了评估。从实验结果可以看出, 本文方法超越了现有的主流域泛化方法, 同时在 ResNet-50 中平均准确率超越了最新方法 SWAD-NU^[51] 0.37%, 充分说明了本文方法的有效性。

表 5 VLCS 数据集在 ResNet-18 中对比实验结果

方法	准确率				
	Caltech	LabelMe	Pascal	Sun	平均
RSC ^[44]	95.83%	63.74%	71.86%	72.12%	75.89%
StableNet ^[48]	96.67%	65.36%	73.59%	74.97%	77.65%
DomainDrop ^[21]	98.94%	63.97%	76.36%	73.74%	78.25%
本文方法	98.58%	64.17%	76.46%	74.14%	78.34%

表 6 VLCS 数据集在 ResNet-50 中对比实验结果

方法	准确率				
	Caltech	LabelMe	Pascal	Sun	平均
RSC ^[44]	97.90%	62.50%	72.30%	75.60%	77.10%
SWAD ^[54]	98.80%	63.30%	75.30%	79.20%	79.10%
CCFP ^[50]	98.90%	64.10%	74.90%	79.90%	79.40%
SWAD-NU ^[51]	99.10%	63.60%	75.90%	80.50%	79.80%
DMDA ^[52]	98.20%	64.30%	74.40%	80.80%	79.40%
本文方法	98.78%	65.12%	77.20%	79.58%	80.17%

2.4 实验分析

2.4.1 消融实验

本文通过在 PACS 数据集上使用 ResNet-18 主干网络作为 Baseline 进行消融实验, 以验证本文方法中 CMADD 和 MICL 两个主要模块的有效性。消融实验结果见表 7, 本文实验都使用相同的参数进行训练, 相比于 Baseline, 本文最终模型提高了 4.95% 的准确率。两个模块单独加入时, 相较于原始模型分别有了 3.67% 和 3.97% 的准确率提升, 验证了这两个模块对于性能提升的有效性。两个模块都加入后性能最

好, 验证了这两个模块是相辅相成、相互促进的, 对于实现模型的上级泛化能力都是不可或缺的。

表 7 消融实验结果

方法	准确率				
	Art	Cartoon	Photo	Sketch	平均
Baseline	80.14%	76.24%	95.22%	75.68%	81.82%
CMADD	83.58%	80.27%	96.22%	83.08%	85.79%
MICL	83.16%	79.88%	95.79%	83.12%	84.67%
CMADD+MICL	84.53%	81.23%	96.64%	84.67%	86.77%

为了验证本文所提注意力机制的有效性, 将 CMADD 模块中的跨空间多尺度注意力模块替换为其他常用的注意力模块, 并在 PACS 数据集中使用 ResNet-18 作为主干网络进行对比实验, 注意力模块对比实验结果见表 8, 其中 Baseline 为 CMADD 去除注意力机制的部分。

表 8 注意力模块对比实验结果

方法	准确率				
	Art	Cartoon	Photo	Sketch	平均
Baseline	82.67%	78.22%	95.98%	81.46%	84.58%
+SE ^[29]	82.81%	78.26%	96.05%	81.89%	84.75%
+CBAM ^[30]	83.35%	79.82%	96.36%	82.58%	85.53%
+CA ^[35]	82.67%	79.54%	95.88%	82.66%	85.19%
CMADD	83.58%	80.27%	96.22%	83.08%	85.79%

可以看出, CMADD 模块相比 Baseline 显著提升了平均精度 (1.21%), 在 Art 和 Sketch 类别上的表现尤为突出 (分别提升了 0.91% 和 1.62%)。这种提升得益于 CMADD 的跨空间多尺度信息聚合机制, 能够更好地捕捉局部和全局的特征, 建立短程和长程依赖, 这使模型在域泛化任务中能够有效捕捉域特定特征。

与其他注意力模块相比, CMADD 在综合性能上更优。相比 CA 模块, CMADD 更均衡地分布了空间语义信息, 同时多尺度策略提升了对复杂特征的捕捉能力。与 CBAM 模块相比, CMADD 更擅长处理需要同时建模局部和全局特征的场景, 例如 Art 和 Sketch 类别中的图像。CBAM 虽表现接近, 但在 Sketch 类别上略逊一筹, 这说明 CMADD 的跨尺度特征聚合更适合复杂任务。总体来说, 实验结果充分验证了本文方法的有效性。

2.4.2 参数敏感性实验

本文方法涉及 3 个超参数: 特征通道去除率

P_{drop} 、动量值 m 和损失权重 λ 。本文在 PACS 数据集上使用 ResNet-18 和 ResNet-50 主干网络对超参数的不同取值展开研究, 实验结果如图4所示。本文在分析其中一个超参数值时其余超参数值均为固定值。在图4(a)中, 将特征通道去除率 P_{drop} 进行了多次取值, 发现在 $P_{\text{drop}}=0.33$ 时模型可以达到最优效果。在图4(b)和图4(c)中, 对动量值和损失函数在 0~1 进行了多次取值, 两者在取值 0.1 时训练出的模型最优。另外, 无论超参数值如何选择, 本文方法相较于只使用原主干网络进行训练都有较大幅度的提升。

2.4.3 热力图分析

本文利用 GradCAM^[55] 为基线方法 ResNet-18 和本文方法生成网络最后一层的热力图, 反映了该网络在分类时的决策依据, 以提供本文方法在提取域不变特征方面有效性的视觉证据。热力图分析如图5所示, 基线方法因为受域特定特征的影响, 往往关注错误信息; 而本文提出的方法利用注意力机

制去除域特定特征, 同时通过原型矩阵实现推理一致性, 使模型专注于同一类别的不同域对象之间的共同特征。例如, 本文方法将马的全身、狗头等作为分类的依据, 而基线方法往往会受到不同域的背景信息干扰, 从而做出错误的分类。这些结果表明, 本文方法可以有效缓解域偏移, 为其他域泛化方法提供参考。

3 结束语

本文从目前域泛化方法的普遍问题出发, 提出了一个基于跨空间多尺度信息聚合注意力机制的域鉴别器, 提高域特定特征的识别能力, 并通过去除域特定特征进行域不变特征的学习。此外, 本文通过引入原型矩阵, 设计了基于动量更新的推理一致性损失函数, 通过样本的原型矩阵与对应类别的平均值尽可能靠近来保持同一类别原型矩阵的不变性, 从而弥补了 CMADD 只关注特征而导致的输出不够可靠的问题。本文进行了广泛的对比方法和消

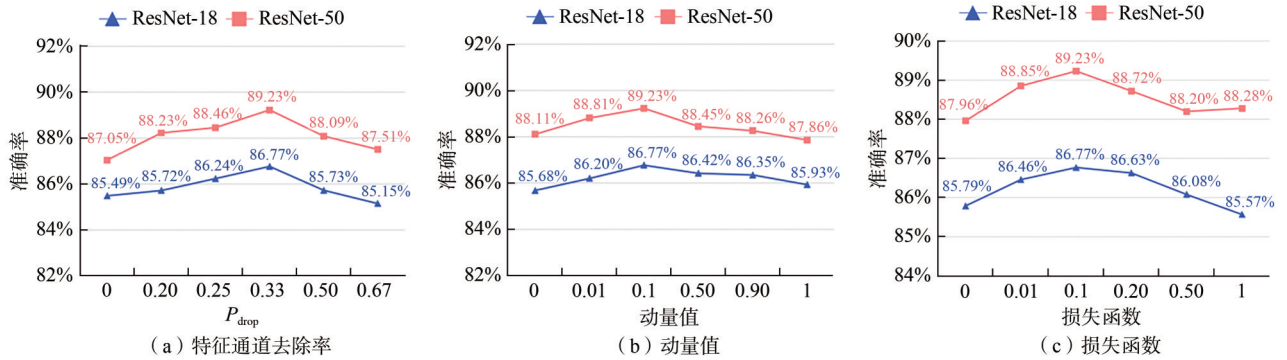


图4 实验结果

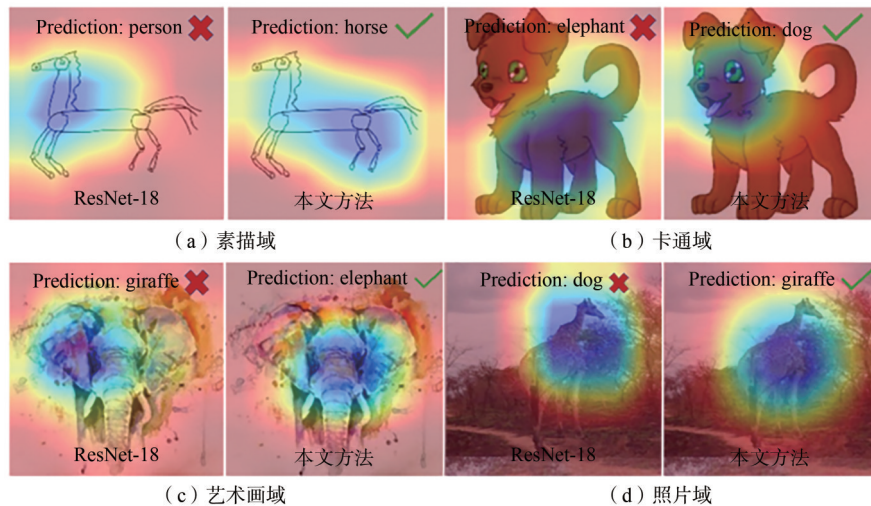


图5 热力图分析

融实验,从而证明了本文方法的优越性。尽管本文提出的方法在域泛化任务中展现了优越的性能,但仍存在一些不足。首先,设计的推理一致性损失函数仅适用于类别数量固定的情况,无法应对类别数量动态变化的任务。其次,在模型训练过程中,动量更新机制可能引入一定的随机性,导致结果不稳定。此外,当前方法在高计算资源环境下表现良好,但其计算复杂度较高,仍需优化以适应资源受限的实际应用场景。未来研究将着重于拓展该方法的适用范围,以及探索更有效的域特定特征去除机制,以提升模型的稳定性,还将尝试优化模型的计算效率,力求在不牺牲性能的前提下,降低对计算资源的需求,以实现更广泛的实际应用。通过这些改进,笔者期望在更复杂和多变的跨域场景中进一步提升模型的泛化能力。

参考文献:

- [1] KOH P W, SAGAWA S, MARKLUND H, et al. Wilds: a benchmark of in-the-wild distribution shifts[C]//Proceedings of the International Conference on Machine Learning. New York: PMLR, 2021: 5637-5664.
- [2] PAN S J, YANG Q. A survey on transfer learning[J]. IEEE Transactions on Knowledge and Data Engineering, 2010, 22(10): 1345-1359.
- [3] YOU K C, LONG M S, CAO Z J, et al. Universal domain adaptation[C]//Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2019: 2720-2729.
- [4] HUANG D J, LI J C, CHEN W K, et al. Divide and adapt: active domain adaptation via customized learning[C]//Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2023: 7651-7660.
- [5] LI Z, CAI R, CHEN G, et al. Subspace identification for multi-source domain adaptation[J]. Advances in Neural Information Processing Systems, 2023, 36: 34504-34518.
- [6] 李凤岐, 金佳玉, 杜学峰, 等. 基于域对抗自适应学习的旋翼无人机姿态稳定方法[J]. 自动化学报, 2025, 51: 1-15.
LI F Q, JIN J Y, DU X F, et al. Attitude stability method for rotor UAV based on domain adversarial self-adaptive learning[J]. Acta Automatica Sinica, 2025, 51: 1-15.
- [7] BLANCHARD G, LEE G, SCOTT C. Generalizing from several related classification tasks to a new unlabeled sample[C]//Proceedings of the 25th International Conference on Neural Information Processing Systems. Massachusetts: MIT Press, 2011: 2178-2186.
- [8] MUANDET K, BALDUZZI D, SCHÖLKOPF B. Domain generalization via invariant feature representation[C]//Proceedings of the International Conference on Machine Learning. New York: PMLR, 2013: 10-18.
- [9] DESHMUKH A, LEI Y W, SHARMA S, et al. A generalization error bound for multi-class domain generalization[EB]. 2019.
- [10] GILLES B, ANAND D A, URUN D, et al. Domain generalization by marginal transfer learning[J]. Journal of Machine Learning Research, 2021, 22(2): 1-55.
- [11] YE H, XIE C, CAI T, et al. Towards a theoretical framework of out-of-distribution generalization[J]. Advances in Neural Information Processing Systems, 2021, 34: 23519-23531.
- [12] CHEN L, ZHANG Y, SONG Y B, et al. Domain generalization via rationale invariance[C]//Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2023: 1751-1760.
- [13] SAIN S R. The nature of statistical learning theory[J]. Technometrics, 1996, 38(4): 409.
- [14] VEDANTAM R, LOPEZ-PAZ D, SCHWAB D J. An empirical investigation of domain generalization with empirical risk minimizers[J]. Advances in Neural Information Processing Systems, 2021, 34: 28131-28143.
- [15] LI H L, PAN S J, WANG S Q, et al. Domain generalization with adversarial feature learning[C]//Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 5400-5409.
- [16] MOTHIAN S, PICCIRILLI M, ADJEROH D A, et al. Unified deep supervised domain adaptation and generalization[C]//Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2017: 5716-5726.
- [17] LI Y, TIAN X, GONG M, et al. Deep domain generalization via conditional invariant adversarial networks[C]//Proceedings of the European Conference on Computer Vision (ECCV). Cham: Springer International Publishing, 2018: 624-639.
- [18] ARJOVSKY M, BOTTOU L, GULRAJANI I, et al. Invariant risk minimization[EB]. 2020.
- [19] GHIFARY M, KLEIJN W B, ZHANG M J, et al. Domain generalization for object recognition with multi-task autoencoders[C]//Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2015: 2551-2559.
- [20] MAHAJAN D, TOPLE S, SHARMA A. Domain generalization using causal matching[C]//Proceedings of the International Conference on Machine Learning. New York: PMLR, 2021: 7313-7324.
- [21] GUO J T, QI L, SHI Y H. DomainDrop: suppressing domain-sensitive channels for domain generalization[C]//Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2023: 19057-19067.
- [22] 倪云昊, 黄雷. 基于数据表示不变性的域泛化研究[J]. 图学学报, 2024, 45(4): 705-713.
NI Y H, HUANG L. Domain generalization based on data representation invariance[J]. Journal of Graphics, 2024, 45(4): 705-713.
- [23] ZHOU K Y, YANG Y X, HOSPEDALES T, et al. Deep domain-adversarial image generation for domain generalisation[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2020, 34(7): 13025-13032.
- [24] WANG Y, QI L, SHI Y H, et al. Feature-based style randomization for domain generalization[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 32(8): 5495-5509.
- [25] ZHOU K Y, YANG Y X, QIAO Y, et al. Domain adaptive ensemble learning[J]. IEEE Transactions on Image Processing, 2021, 30: 8008-8018.
- [26] SEGU M, TONIONI A, TOMBARI F. Batch normalization embeddings for deep domain generalization[J]. Pattern Recognition, 2023, 135: 109115.
- [27] SEO S, SUH Y, KIM D, et al. Learning to optimize domain specific normalization for domain generalization[C]//Proceedings of the European Conference on Computer Vision (ECCV). Cham: Springer International Publishing, 2020: 68-83.
- [28] 童煜钧, 王荷清, 罗悦恒, 等. 基于扩散模型数据增广的域泛化方法[J]. 智能科学与技术学报, 2023, 5(3): 380-388.
TONG Y J, WANG H Q, LUO Y H, et al. Domain generalization

- method based on diffusion model data augmentation[J]. Chinese Journal of Intelligent Science and Technology, 2023, 5(3): 380-388.
- [29] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 7132-7141.
- [30] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[C]//Proceedings of the European Conference on Computer Vision (ECCV). Cham: Springer International Publishing, 2018: 3-19.
- [31] OUYANG D, HE S, ZHANG G, et al. Efficient multi-scale attention module with cross-spatial learning[C]//Proceedings of the ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Piscataway: IEEE Press, 2023: 1-5.
- [32] LI D, YANG Y X, SONG Y Z, et al. Deeper, broader and artier domain generalization[C]//Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2017: 5543-5551.
- [33] TORRALBA A, EFROS A A. Unbiased look at dataset bias[C]//Proceedings of the CVPR 2011. Piscataway: IEEE Press, 2011: 1521-1528.
- [34] VENKATESWARA H, EUSEBIO J, CHAKRABORTY S, et al. Deep hashing network for unsupervised domain adaptation[C]//Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2017: 5385-5394.
- [35] HOU Q B, ZHOU D Q, FENG J S. Coordinate attention for efficient mobile network design[C]//Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2021: 13708-13717.
- [36] MATSUURA T, HARADA T. Domain generalization using a mixture of multiple latent domains[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 11749-11756.
- [37] EFRAIMIDIS P S, SPIRAKIS P G. Weighted random sampling with a reservoir[J]. Information Processing Letters, 2006, 97(5): 181-185.
- [38] CARLUCCI F M, D'INNOCENTE A, BUCCI S, et al. Domain generalization by solving jigsaw puzzles[C]//Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2019: 2224-2233.
- [39] SHI B, ZHANG D, DAI Q, et al. Informative dropout for robust representation learning: a shape-bias perspective[C]//Proceedings of the International Conference on Machine Learning. New York: PMLR, 2020: 8828-8839.
- [40] DU D P, CHEN J W, LI Y X, et al. Cross-domain gated learning for domain generalization[J]. International Journal of Computer Vision, 2022, 130(11): 2842-2857.
- [41] CHATTOPADHYAY P, BALAJI Y, HOFFMAN J. Learning to balance specificity and invariance for in and out of domain generalization[C]//Proceedings of the European Conference on Computer Vision (ECCV). Cham: Springer International Publishing, 2020: 301-318.
- [42] RAME A, DANCETTE C, CORD M. Fishr: invariant gradient variances for out-of-distribution generalization[C]//Proceedings of the International Conference on Machine Learning. New York: PMLR, 2022: 18347-18377.
- [43] DING Y Z, WANG L, LIANG B X, et al. Domain generalization by learning and removing domain-specific features[J]. Advances in Neural Information Processing Systems, 2022, 35: 24226-24239.
- [44] HUANG Z Y, WANG H H, XING E P, et al. Self-challenging improves cross-domain generalization[C]//Proceedings of the European Conference on Computer Vision (ECCV). Cham: Springer International Publishing, 2020: 124-140.
- [45] ZHOU K, YANG Y, QIAO Y, et al. Domain generalization with Mix-Style[EB]. 2021.
- [46] ZHANG Y B, LI M H, LI R H, et al. Exact feature distribution matching for arbitrary style transfer and domain generalization[C]//Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2022: 8025-8035.
- [47] GUO J T, WANG N, QI L, et al. ALOFT: a lightweight MLP-like architecture with dynamic low-frequency transform for domain generalization[C]//Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2023: 24132-24141.
- [48] ZHANG X X, CUI P, XU R Z, et al. Deep stable learning for out-of-distribution generalization[C]//Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2021: 5372-5382.
- [49] XU Q W, ZHANG R P, ZHANG Y, et al. A Fourier-based framework for domain generalization[C]//Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2021: 14383-14392.
- [50] LI C M, ZHANG D A, HUANG W J, et al. Cross contrasting feature perturbation for domain generalization[C]//Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2023: 1327-1337.
- [51] SHI Z, MING Y, FAN Y, et al. Domain generalization via nuclear norm regularization[C]//Proceedings of the Conference on Parsimony and Learning. New York: PMLR, 2024: 179-201.
- [52] LONG S C, ZHOU Q Y, YING C H, et al. Rethinking domain generalization: discriminability and generalizability[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2024, 34(11): 11783-11797.
- [53] NAM H, LEE H, PARK J, et al. Reducing domain gap by reducing style bias[C]//Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2021: 8686-8695.
- [54] CHA J, CHUN S, LEE K, et al. Swad: domain generalization by seeking flat minima[J]. Advances in Neural Information Processing Systems, 2021, 34: 22405-22418.
- [55] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-CAM: visual explanations from deep networks via gradient-based localization[C]//Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2017: 618-626.

[作者简介]



黎拓新 (2002-), 男, 国防科技大学智能科学学院硕士生, 主要研究方向为模式识别、人工智能、迁移学习。



项凤涛 (1986-), 男, 博士, 国防科技大学智能科学学院副教授, 主要研究方向为智能辅助决策、不确定性推理、智能控制。



陈君海（2001- ），男，国防科技大学智能科学学院硕士生，主要研究方向为人工智能、谣言检测。



吕云霄（1985- ），女，国防科技大学智能科学学院高级工程师、副处长，主要研究方向为导航技术。



张晓博（2002- ），男，国防科技大学智能科学学院硕士生，主要研究方向为人工智能、增量学习。