

基于扩散模型数据增广的域泛化方法

童煜钧¹, 王荷清¹, 罗悦恒¹, 宁文欣¹, 关曼丹¹, 喻雯晴¹, 黄柯彦², 张加迅², 马占宇¹

(1. 北京邮电大学人工智能学院, 北京 100876; 2. 北京空间飞行器总体设计部, 北京 100094)

摘要: 域泛化是计算机视觉领域中的一个重要且具有挑战性的问题, 该问题源于现实场景中的数据分布偏移。在实际应用中, 通常会遇到训练数据和测试数据来自不同的数据域的情况, 这种数据分布的差异会导致测试时准确率下降。因此, 提出了一种基于隐空间数据增广的域泛化方法, 与传统图像级数据增广方法不同, 该方法在隐空间中引入扩散模型, 以实现特征的精细控制和多样性生成, 从而提升模型在目标域上的泛化能力。具体来说, 基于分类器的隐式扩散模型在隐空间训练后可以条件生成准确且丰富的源域特征, 并利用高效的采样方法加速生成增广特征。实验结果表明, 新提出的方法在各种域泛化任务上取得了显著的性能提升, 在真实场景中有较好的有效性和鲁棒性。该方法的创新点在于将数据增广焦点转移到隐空间级别, 并引入扩散模型进行增广, 为解决域泛化问题提供了一种新的思路。

关键词: 域泛化; 扩散模型; 数据增广

中图分类号: TP39

文献标志码: A

doi: 10.11959/j.issn.2096-6652.202334

Data augmentation method based on diffusion model for domain generalization

TONG Yujun¹, WANG Heqing¹, LUO Yueheng¹, NING Wenxin¹, GUAN Mandan¹, YU Wenqing¹, HUANG Keyan², ZHANG Jiaxun², MA Zhanyu¹

1. School of Artificial Intelligence, Beijing University of Posts and Telecommunications, Beijing 100876, China

2. Beijing Institute of Spacecraft System Engineering, Beijing, 100094, China

Abstract: Domain generalization is an important and challenging problem in computer vision, arising from the distribution shift of real-world data. In practical applications, it is common to encounter training and testing data from different domains, and the difference in data distribution can lead to performance degradation during testing. In this paper, we propose a domain generalization method based on latent space data augmentation. Unlike traditional image-level data augmentation approaches, the method introduces a diffusion model in the latent space to achieve fine control and diversity generation of features, thereby achieving feature level data augmentation and enhancing the model's generalization ability in the target domain. Specifically, the classifier-based implicit diffusion model, trained within the latent space, can conditionally generate accurate and rich source domain features. It leverages efficient sampling techniques to expedite the generation of augmented features. Experimental results show that the method has achieved significant performance improvement in various domain generalization tasks, and has good effectiveness and robustness in real scenarios. The key innova-

收稿日期: 2023-05-05; 修回日期: 2023-08-10

通信作者: 马占宇, mazhanyu@bupt.edu.cn

基金项目: 北京市自然科学基金项目 (No.Z200002); 国家自然科学基金项目 (No.U19B2036, No.62225601); 北京邮电大学优秀青年团队项目 (No.2023QNTD02); 北京邮电大学博士生创新基金项目 (No.CX2023112)

Foundation Items: Beijing Natural Science Foundation Project (No.Z200002), The National Natural Science Foundation of China (No.U19B2036, No.62225601), Youth Innovative Research Team of BUPT (No.2023QNTD02), BUPT Excellent Ph.D. Students Foundation (No.CX2023112)

tion of this paper lies in shifting data augmentation to the latent space level and introducing the diffusion model for augmentation, providing a novel approach to address the domain generalization problem.

Key words: domain generalization, diffusion model, data augmentation

0 引言

近年来,随着计算机视觉技术的迅猛发展,人们对在现实世界中应用视觉算法的需求也日益增加。然而,现实场景中的数据分布往往会因为不同的环境、设备或者拍摄条件而产生差异^[1]。这种数据分布的偏移使得模型从一个域(源域)推广到另一个域(目标域)变得困难。例如,在一个人脸识别任务中,使用室内拍摄的人脸图像训练了一个模型,但在实际应用中需要识别室外拍摄的人脸图像。由于室内和室外环境的差异,模型在目标域上的性能往往会显著下降。针对上述问题,域泛化方法提供了很好的解决思路。域泛化方法旨在设计出一种模型,能够在未见过的目标域上保持良好的泛化能力,即使源域和目标域之间存在较大的数据分布差异。在计算机视觉领域,域泛化方法的研究引起了广泛关注,并取得了令人瞩目的成果。

在过去的研究中,已经有一些工作致力于改善域迁移现象,包括与域无关的特征空间学习^[2-6]、DG元学习^[5,7-8]和数据增广^[9-12]。数据增广被广泛应用于机器学习模型训练,被证明是改善模型泛化能力的有效手段。常见的数据增广操作包括翻转、旋转、缩放、裁剪和噪声添加等,这些操作在监督学习中被广泛采用。引入多样性的数据和噪声,可以减少过拟合问题并提高模型的泛化性能^[13-14]。本文进一步延伸思路,提出了一种基于数据增广的新策略。在数据增广过程中引入扩散模型,可以增加源域间的样本差异性,使得模型更加鲁棒,从而提升模型在目标域上的泛化能力。通过这种方式,本文方法能够更好地应对极端的域差异情况,并在域泛化任务中获得更好的性能表现。

近期,扩散模型在生成类任务中展现出了更好的性能,在图像生成、图像修复、图像编辑等领域大放异彩。条件噪声分割网络(noise conditional score network, NCSN)^[15]、去噪扩散概率模型(denoising diffusion probabilistic model, DDPM)^[8]和隐式扩散模型(latent diffusion model, LDM)^[16],具有非常灵活的建模结构,在不使用对抗性学习的情况下能够生成高质量的生成对抗网络(genera-

tive adversarial network, GAN)样本。但是,基于扩散模型的图像生成的计算量极高,需要大量的时间与计算资源,在推理过程中冗余的采样加剧了这一问题。有大量的工作在此基础上进行改进,如隐空间的扩散模型与高效跳步采样方法。本文方法利用扩散模型生成的高质量、多样的语义特征来训练模型,利用已有的高效加速手段来解决扩散模型推理采样慢的问题。本文将现有的隐空间扩散模型与高效的采样手段DPM-solver结合,将扩散模型作为一个特征“增广器”来增广已有特征,以此获得更强的域泛化模型。

具体而言,本文利用基于隐空间的扩散模型生成增广样本。在训练过程中,本文通过预训练获得样本在隐空间的初步表征。扩散模型学习该表征以生成多样的全新样本,实现特征增广,从而丰富源域的数据分布。除此以外,本文方法利用分类器指导扩散模型生成正确语义的样本。本文方法额外训练了一个增广分类器来对增广样本进行分类,以确保增广样本具有正确语义。此外,本文方法在分类器决策边界附近采样,并使生成的增广样本在重新精调模型后能被正确分类,从而提升模型在目标域上的泛化性能。为了加快采样过程,本文运用了一种快速跳步采样方法DPM-solver来调整采样步数,加快了训练速度,实现了效果与速度的平衡。

本文方法通过一个“插件式”的扩散模型,即可快速实现特征增广,可以与其他域泛化方法结合使用。通过在PACS数据集上测试并可视化增广特征,验证了本文方法的有效性和先进性。

1 相关工作

1.1 域泛化

域泛化的相关工作分为以下几类:基于学习域不变特征的方法^[2-3,6,17]、基于元学习的方法^[7,18-21]、基于自监测的方法^[16]和基于数据增广的方法^[9-12]。基于学习域不变特征的方法旨在学习一个通用的特征表征,使相同语义的样本在不同领域之间保持一致的特征。例如,文献[17]提出了一个基于自动编码器的跨域重建方法,将源域和目标域数据对齐到一个通用的分布上。文献[2,22]将深度神经网络参

数分解为域不变的部分和域特有的部分,并利用域不变参数实现更优异的泛化能力。文献[6]则提出以最大平均差(maximum mean discrepancy, MMD)为度量参数,通过减少不同领域之间的分布差异来学习域不变特征。协方关系匹配(correlation alignment, CORAL)^[5]包括了不同领域之间二阶统计差异的方法,被认为是一种有效的域对准技术。基于元学习的方法通过找到不同域之间的共有优化方向^[18]得到泛化性强的模型。基于自监督的方法学习与域无关的判别力特征,从而得到鲁棒的模型,如文献[23]提出使用拼图自监督任务实现域泛化任务。基于数据增广并利用多样的数据增加模型在不同领域的泛化能力,文献[9]将样本在不同域中转化并作为增广样本,文献[12]在特征中加入标准高斯噪声以实现特征增广。本文工作将沿用已有数据增广工作的思路,通过扩散模型生成更多增广特征。

1.2 扩散模型

扩散模型^[24]最近获得了很大的关注,在各种生成任务^[8, 25-26]以及图像到图像的转化任务^[27-30]中取得了令人震撼的先进成果。为了使模型能够生成高度相似但又具有多样性的样本,扩散模型逐步向原始数据中添加噪声建立马尔可夫链过程,直至噪声数据逐渐退化为高斯噪声,并学会反向重建^[8]。在分类器指导^[31]或交叉注意机制^[16]的帮助下,条件扩散模型更有可能应用于需要监督生成的任务^[32]。扩散模型在很多任务中都有很好的表现,但处理高维空间的数据,特别是冗余的采样过程,需要很多的计算资源。本文将强大的扩散模型作为“数据增广器”,同时采用诸多方法加速扩散采样过程。

1.3 扩散模型的加速采样

在DDPM^[8]中,生成过程和扩散过程需要相同的步数重建原始数据分布。然而,扩散模型具有解耦特性,不需要扩散和采样的步数必须一致。受生成性隐式模型的启发,文献[33]提出了具备确定性扩散特点和跳步采样的隐式采样方法DDIM。在采样过程中,根据是否添加额外的噪声,可以将扩散模型分为两类:一类是扩散随机微分方程模型,另一类是扩散常微分方程。利用线性求解器和高阶求解器,可以实现对微分方程更有效的采样。除此之外,DPM-solver^[34]认为扩散常微分方程可以被看作一种半线性的形式,这种形式可以减少离散化误

差,一个联合的不同阶数的求解器可能有更好的性能,只需15~20步即可生成高质量采样样本。

2 方法描述

本节将详细阐述基于扩散模型数据增广的域泛化方法。本文利用扩散模型的生成能力,在特征空间增广出多样且质量高的训练样本。这个过程面临两个挑战:一是如何使特征保持正确语义;二是如何减少扩散模型多步生成所带来的时间消耗。

2.1 符号标志

在域泛化问题中,有 N 个可见的带标签源域 $\{\mathcal{D}^1, \dots, \mathcal{D}^N\}$ 用于训练,每个源域都具有不同数据和标签对的分布 $\{P^j(\mathcal{X}, \mathcal{Y}) : x \in \mathcal{X}, y \in \mathcal{Y}\}_{j=1}^N$,然而,测试通常在一个具有显著域差异且未见过的目标域 $\mathcal{D}^{N+1}$ 上进行,其中 $P^j(\mathcal{X}, \mathcal{Y}) \neq P^{N+1}(\mathcal{X}, \mathcal{Y}), j=1, \dots, N$ 。域泛化的主要目标是训练一个与域无关的特征提取器 $F_\theta(\cdot)$ 和分类器 $G_\omega(\cdot)$,将在源域上训练的模型推广到未见过的新目标域 $\mathcal{D}^{N+1}$ 上。为实现数据增广,还需训练额外的扩散模型 $D_\phi(\cdot)$ 与指导扩散模型条件生成的分类器 $G_\alpha(\cdot)$ 。

2.2 方法步骤

在域泛化问题中,传统的经验风险最小化(empirical risk minimization, ERM)方法都能达到有竞争力的效果。因此,本文将传统的ERM方法与基于扩散模型的数据增广方法结合,以获得更好的域泛化效果。主要步骤如图1所示,分为3个阶段:(1)利用ERM方法预训练模型,得到样本在隐空间的初步良好表征;(2)利用上一阶段获得的隐空间特征训练条件生成的扩散模型;(3)利用扩散模型生成特征对预训练模型进行精调,以获得更好的泛化性能。

为了准确生成对应类别的增广特征,首先,利用分类器指导的条件生成扩散模型指导扩散模型的生成。具体来说,需要额外训练一个能判别加噪特征的分类器,在扩散模型采样时,通过回传梯度指导扩散模型的生成。另外,样本会在训练过程中多次被输入网络进行增广生成,多次重建过程加剧了扩散模型推理慢的问题。如何有效地避免重复且冗余的采样推理过程是本文需要解决的难点。首先,选择在隐空间中增广样本,而非增广图像本身,这极大减少了所需的计算量。其次,选择DPM-

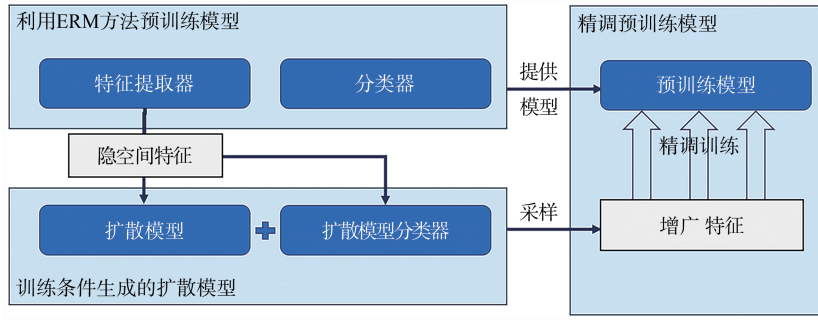


图1 基于扩散模型的数据增广方法架构

solver^[34]进行跳步采样，进一步减少了采样推理次数。

2.3 ERM

为了得到数据集的准确特征，首先使用ERM方法预训练模型，使模型能够初步生成可用于扩散模型训练的特征。ERM是一个很强的基准方法，在很多情况下甚至超越现有的最先进的方法。具体来说，ERM将各域的训练样本进行混合训练，通过交叉熵损失更新特征提取器与分类器，其中 $l(\cdot)$ 是交叉熵函数：

$$\arg \min_{\theta, \omega} \mathcal{L}_{ce} = \mathbb{E}_{x, y \in \{\mathcal{D}^1, \dots, \mathcal{D}^N\}} l(G_{\omega}(F_{\theta}(x)), y) \quad (1)$$

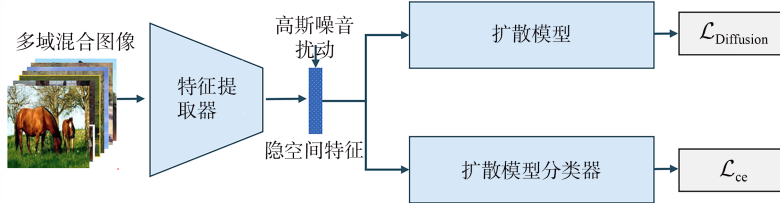


图2 基于训练条件生成的扩散模型

除此以外，额外训练了一个分类器来指导扩散模型的生成^[31]，通过对加噪特征的准确分类指导去噪过程中的去噪方向，即交叉熵损失训练，其中 $l(\cdot)$ 是交叉熵函数：

$$\arg \min_{\alpha} \mathcal{L}_{ce} = \mathbb{E}_{x, y \in \{\mathcal{D}^1, \dots, \mathcal{D}^N\}} l(G_{\alpha}(\tilde{f}_t), y) \quad (3)$$

其中，去噪模型 $D_{\phi}(\cdot)$ 采用U-Net结构， $G_{\alpha}(\cdot)$ 使用半U-Net结构。

2.5 利用扩散模型生成增广特征精调模型

如图3所示，在得到一个训练好的扩散模型之后，对噪音进行多步去噪，得到增广特征，如DDPM^[8]中的推理采样为：

2.4 基于条件生成的扩散模型训练

如图2所示，将ERM方法所提取出的特征用于扩散模型 $D_{\phi}(\cdot)$ 的训练，并额外训练能够指导噪音特征方向的分器 $G_{\alpha}(\cdot)$ 。本文选择DDPM方法作为扩散模型的训练方法^[8]。具体来说，为了能准确地将加噪特征去噪，在扩散模型的损失函数中加入噪音与过滤噪音进行L2损失匹配，更多细节可以参考文献[8]，最终的损失函数是：

$$\arg \min_{\phi} \mathcal{L}_{\text{Diffusion}} = \mathbb{E}_{t \sim [1, T], f_0, \epsilon_t} \left[\left\| \epsilon_t - \epsilon_{\phi}(\tilde{f}_t, t) \right\|^2 \right] \quad (2)$$

其中 $\tilde{f}_t$ 是加噪特征； ϵ_t 是在加噪特征中加入的噪音，从标准高斯中采样得到； t 是扩散步数，代表了加入噪音的不同幅度， t 越大噪音分量越多。

$$\dot{f}_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left(\dot{f}_t - \frac{1 - \alpha_t}{\sqrt{1 - \bar{\alpha}_t}} \epsilon_{\theta}(\dot{f}_t, t) \right) + \sigma_t z \quad (4)$$

其中 z 为采样得到的噪音， $\epsilon_{\theta}(\dot{f}_t, t)$ 为过滤出的噪音， σ_t 、 α_t 均为人为定义超参，对应去噪过程中去除噪音大小。高斯白噪音经过多步去噪，得到增广特征 $\dot{f}_0$ ，后续将其简记为 $\dot{f}$ 。

为使增广特征仍能保证应有的语义，使用预先训练好的分类器对去噪过程进行监督，控制扩散模型生成所需要的类别，即对过滤噪音进行调整与修改。根据文献[31]，对过滤噪音 $\epsilon_{\theta}(\dot{f}_t, t)$ 进行修改

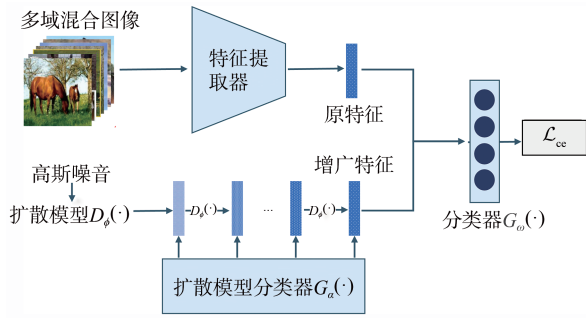


图3 利用扩散模型生成扩增特征精调模型

与监督, 得到:

$$\tilde{\epsilon}_{\theta}(\dot{f}_t, t, c) := \epsilon_{\theta}(\dot{f}_t, t) - s \cdot \sigma_t \nabla_{\dot{f}_t} \log p_a(c | \dot{f}_t, t) \quad (5)$$

其中, $\tilde{\epsilon}_{\theta}(\dot{f}_t, t, c)$ 为根据所需生成类别调整后的去噪后的噪音, $\nabla_{\dot{f}_t} \log p_a(c | \dot{f}_t, t)$ 为分类器 $G_a(\cdot)$ 根据所需类别 c 得出的调整方向, s 为调整大小。

利用分类器引导的去噪模型, 生成所需类别的扩增样本。结合扩增样本对应的类别标签 c , 用简单的ERM方法进一步精调模型, 从而得到更好的泛化效果:

$$\arg \min_{\omega} \mathbb{E}_{\dot{f}} \mathcal{L}_{cc}(G_{\omega}(\dot{f}), c) \quad (6)$$

为了加快扩散模型的采样速度, 同时不削弱扩增特征的判别性, 本文使用了高效的DPM-solver, 在10~20步就达到了良好的特征扩增效果:

$$\dot{f}_{t_i} = \frac{\alpha_{t_i}}{\alpha_{t_{i-1}}} \dot{f}_{t_{i-1}} - \sigma_{t_i} (e^{h_i} - 1) \epsilon_{\theta}(\dot{f}_{t_{i-1}}, t_{i-1}), \quad h_i = \lambda_{t_i} - \lambda_{t_{i-1}} \quad (7)$$

其中, t_i 为跳步采样的步骤, $i=0, 1, \dots, M$, M 为跳步后的采样总次数, 往往为10~20步, 具体可参考文献[34]。

3 实验分析

3.1 数据集

本文在PACS数据集^[2]上验证上述方法的有效性。PACS数据集是域自适应领域常用的图像数据集, 共有9 991张图片, 其中包含实物照片(photo, P)、艺术画(art, A)、卡通图像(cartoon, C)和简笔画(sketch, S)4个域, 每个域含有狗、大象、长颈鹿、吉他、马、房子、人7个类别。本文沿用了PACS数据集原始的训练集与验证集的划分方式。具体而言, 在训练阶段, 在4个域中选择3个

作为源域, 将剩余的一个域作为目标域。最后, 评估模型在所有拆分情况下目标域对应的准确率。

3.2 实验过程

3.2.1 数据预处理

PACS数据集的每个域包含1 600~4 000张图像, 首先将图像随机裁剪为不同的大小和宽高比, 将图像尺寸缩放为256×256; 然后进行随机改变亮度、对比度、饱和度和色调, 随机水平翻转, 随机图片灰度化等预处理, 实现初步的数据增强。

3.2.2 网络模型设计与训练参数

本文模型的主要结构为ResNet50, 扩散模型 $D_{\phi}(\cdot)$ 的网络结构为U-Net结构, 本文采用了文献[16]里的LDM-8结构, 指导扩散模型条件生成的分类器 $G_a(\cdot)$ 为半U-Net结构, 即只有LDM-8的升维过程。

训练过程中的超参数设置如下: 初始学习率为0.001, 训练迭代次数为2 000, 批次大小为16。

3.2.3 扩散模型的设定

本文的扩散模型采用LDM-8结构^[16], 使用在ImageNet上预训练的参数初始化模型权重。LDM-8嵌入交叉注意力机制以增强其底层U-Net, 从而将扩散模型转化为灵活性更高的条件图像生成器。LDM-8的输入尺寸为32×32, 扩散步数为1 000, 通道数为256。在扩散模型中加入了一个映射模块, 这样生成的特征可以与预训练好的LDM-8的输入尺寸相匹配。使用AdamW优化器对扩散模型进行微调, 使用EMA更新参数, 并将所有实验的学习率设置为 10^{-6} 。

3.3 结果分析

表1呈现了本文方法与其他基线方法在PACS数据集上分类的准确率对比结果。ERM^[1]是领域泛化方向的基本方法, 它是一个基于低秩的参数化全卷积神经网络模型, 使用所有源域混合数据直接生成与当前源域风格差异显著但与语义差异分类特征类似的图像。文献[23]提出了一种基于自监督学习的领域泛化方法Jigen, 该方法利用求解拼图游戏这一自监督辅助任务帮助学习领域不变特征。SFA-A^[12]通过在图像特征上添加高斯白噪声进行数据增强。MMLD^[35]使用多个潜在域的混合数据集进行领域泛化, 即在源域标签未知的情况下利用图像的风格特征进行聚类, 将样本迭代地划分为潜在的领域, 进而通过对抗性学习训练域不变特征提取

器。DADG^[36]是一种判别式对抗学习和元学习相结合的框架。BNE^[37]利用归一化层收集不同域的统计信息，并将其投射到共同隐空间，从而实现域泛化。本文方法在ERM模型的基础上，利用扩散模型进一步增强了域不变特征的学习。

表1 与其他领域泛化方法的准确率对比 (%)

方法	A	C	P	S	平均
ERM	78.6	74.4	96.3	72.4	80.4
Jigen	79.4	75.3	96.0	71.6	80.5
SFA-A	81.2	77.8	93.9	73.7	81.7
MMLD	81.3	77.2	96.1	72.3	81.7
DADG	79.9	76.3	94.9	70.5	80.4
BNE	78.8	78.9	94.8	79.7	83.1
本文方法(50步+2阶)	81.9	73.7	96.6	74.8	81.7
本文方法(100步+2阶)	81.8	75.4	96.6	75.1	82.2
本文方法+MMD	82.5	76.6	96.4	77.4	83.2

在4种不同的数据集划分方式下，基于扩散模型的数据增强方法的泛化准确率均优于ERM，验证了该方法的有效性。此外，本文方法实现了与其他域泛化方法相当的性能，说明扩散模型能够增广出更多高质量的潜在域特征，有效提高了分类器在未知目标域上的泛化能力。

本文方法可以选择调整采样步数来实现泛化效果与训练速度的平衡。如表1所示，本文方法在50步采样时就已实现了相当不错的性能，随着采样步数的升高，性能进一步升高。除此以外，本文方法可以实现“插件式”应用，可以与各类方法结合，在隐空间的数据增广能够与其他方法共同提升性能，如表1所示，本文方法与MMD结合进一步提升了泛化性能。

3.4 消融实验分析

本文对拥有不同参数（主要为采样次数和求解常微分方程时的阶数）的采样方法进行了消融分析，实验结果发现，在扩散步数大，高阶常微分方程的情况下，增广特征更为准确，训练后模型分类准确率更好，但同时也增加了训练时的计算量与时间。

本文评估了分类器引导对扩散模型条件生成的影响，与在生成增广特征过程中不加入分类器的引导梯度带来的变化进行比较。如表2所示，无分类器引导的方法的效果出现了下滑，这说明分类器引导能够有效帮助扩散模型生成更准确的增广特征。值得注意的是，即便没有分类器指引扩散模型条件

生成，本文方法仍好于基础的ERM方法。

表2 分类器引导对本文方法准确率的影响 (%)

方法	A	C	P	S	平均
无分类器引导	81.0	73.3	96.7	73.3	81.1
本文方法	81.8	73.7	96.6	74.8	81.7

从表3可以看出，增加模型串行计算的步数能够明显提高模型的采样结果质量，提升方法的性能。但在实验过程中，基于20步的DPM-solver扩散模型迭代3 000次用时160 min，基于50步的DPM-solver扩散模型迭代3 000次用时375 min，因此，我们应采用不同的采样步数设置以平衡采样质量与效能。此外，还可以看出DPM-solver阶数越高，方法的性能越好。

表3 扩散模型采样不同参数设置对准确率的影响 (%)

步数	阶数	A	C	P	S	平均
20	2	80.8	73.7	96.7	75.1	81.6
50	2	81.8	73.7	96.6	74.8	81.7
20	1	80.4	73.3	96.5	75.0	81.3
50	1	81.3	73.4	96.4	74.7	81.5

为了寻找到最合适的采样步数，本文在PACS数据集的艺术画模态（ART）进一步实验，对不同的采样步数所能达到的效果与训练时长进行了比较。如图4所示，随着采样步数的增长，训练时长几乎呈线性增长，当采样步数达到50步以上时，对分类准确率提升并不明显。因此，笔者推荐50步作为采样步数的默认设定。

3.5 可视化

本文分别对ERM和本文方法提取的特征进行t-SNE可视化，如图5（a）和（b）所示，不同颜色代表不同类别。结果表明，本文方法具有更好的类内类间关系，有更大的类间距，同时样本更为紧致。为了更深入地探究增广后的特征与原始特征之间的关系，将它们拼接在一起并使用t-SNE进行可视化分析。图5（c）用蓝色表示扩散模型生成的增广特征，红色代表原始特征；图5（d）用不同的颜色表示增广特征。从图5（c）中可以看出，增广后的特征与原始特征之间存在较大的相似性，但是增广后的特征的分布更加分散，这意味着增广后的特征在更广泛的特征空间中分布，使模型能够学到更多的特征信息，从而提高模型的分类准确率和泛化能

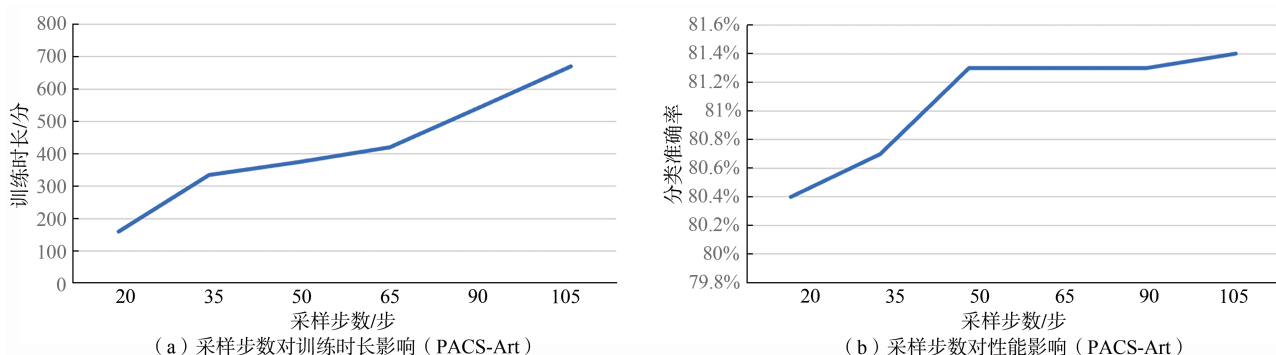


图4 不同采样步数下的训练时长与训练效果

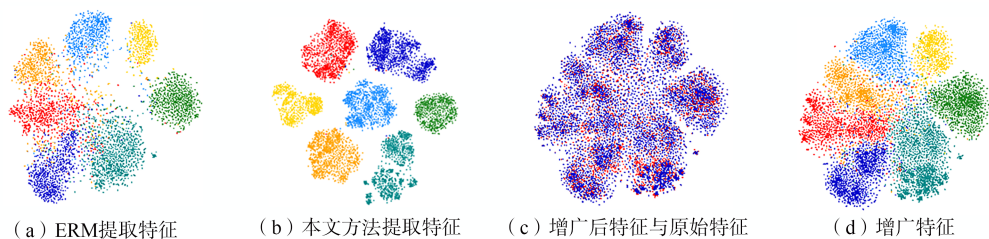


图5 ERM与本文方法的训练和测试数据特征的t-SNE可视化

力。图5(d)表明, 增广后的特征仍然具有明显的语义信息, 并且在域泛化任务中仍然具有可分性。

扩散模型对特征进行增广之后, 特征可以更加充分地利用空间, 更好地体现特征样本的多样性。同时, 增广之后的特征仍然保持明显、丰富的语义信息。这表明, 数据增广技术能够有效地增加数据集的多样性, 提高训练模型的鲁棒性, 并且能够帮助模型更好地区分不同类别之间的差异。

4 结束语

本文提出了一种基于扩散模型的数据增广方法, 目的是提高模型在不同领域之间的泛化能力。在隐空间中进行特征增广, 能够生成多样化、计算高效的增广特征, 同时保留了特征的良好语义。这种方法具有很高的灵活性, 可以轻松地整合到现有的域泛化框架中, 为解决实际应用中数据分布偏移问题提供更多可能性。本文方法采用了基于分类器指导的扩散模型, 借助标签控制生成特征的输出, 确保增广特征具有良好的语义信息。此外, 为了平衡采样速度和生成特征质量, 本文采用了跳步采样策略, 实现了高效、可调节的特征采样。

总体来说, 本文方法不仅在各种域泛化任务上取得了显著的性能提升, 而且能够作为插件与现有的域泛化方法结合, 为现有的域泛化研究提供新的

思路和解决方案。未来的研究将进一步优化和完善此方法, 使其具有更强大的泛化性和更广泛的实用性, 以更好地应对实际应用中的问题。

参考文献:

- [1] LI D, ZHANG J S, YANG Y X, et al. Episodic training for domain generalization[C]//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2019: 1446-1455.
- [2] LI D, YANG Y X, SONG Y Z, et al. Deeper, broader and artier domain generalization[C]//Proceedings of 2017 IEEE International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2017: 5543-5551.
- [3] MUANDET K, BALDUZZI D, SCHÖLKOPF B. Domain generalization via invariant feature representation[C]//Proceedings of the 30th International Conference on Machine Learning. Piscataway: IEEE Press, 2013.
- [4] TORRALBA A, EFROS A A. Unbiased look at dataset bias[C]//Proceedings of the IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2011: 1521-1528.
- [5] SUN B C, SAENKO K. Deep CORAL: correlation alignment for deep domain adaptation[C]//Proceedings of the 2016 Computer Vision - ECCV Workshops. Piscataway: IEEE Press, 2016: 443-450.
- [6] LI H L, PAN S J, WANG S Q, et al. Domain generalization with adversarial feature learning[C]//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 5400-5409.
- [7] BALAJI Y, SANKARANARAYANAN S, CHELLAPPA R. MetaReg: towards domain generalization using meta-regularization[C]//Proceedings of the 32nd International Conference on Neural Information Processing Systems. New York: ACM, 2018: 1006-1016.
- [8] HO J, JAIN A, ABBEEL P. Denoising diffusion probabilistic models[C]//

- Proceedings of the 34th International Conference on Neural Information Processing Systems. New York: ACM, 2020: 6840-6851.
- [9] ZHOU K Y, YANG Y X, HOSPEDALES T, et al. Deep domain-adversarial image generation for domain generalisation[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 13025-13032.
 - [10] ZHOU K, YANG Y, QIAO Y, et al. Domain adaptive ensemble learning[J]. IEEE Transactions on Image Processing, 2021, 30: 8008-8018.
 - [11] VOLPI R, NAMKOONG H, SENER O, et al. Generalizing to unseen domains via adversarial data augmentation[EB]. arXiv preprint, 2018, arXiv: 1805.12018.
 - [12] LI P, LI D, LI W, et al. A simple feature augmentation for domain generalization[C]//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2021: 8886-8895.
 - [13] SHORTEN C, KHOSHGOFTAAAR T M. A survey on image data augmentation for deep learning[J]. Journal of Big Data, 2019, 6(1):1-48.
 - [14] HONARVAR NAZARI N, KOVASHKA A. Domain generalization using shape representation[C]//Proceedings of the Computer Vision - ECCV 2020 Workshops. Cham: Springer, 2020: 666-670.
 - [15] SONG Y, ERMON S. Generative modeling by estimating gradients of the data distribution[EB]. arXiv preprint, 2019, arXiv: 1907.05600.
 - [16] ROMBACH R, BLATTMANN A, LORENZ D, et al. High-resolution image synthesis with latent diffusion models[C]//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2022: 10684-10695.
 - [17] GHIFARY M, KLEIJN W B, ZHANG M J, et al. Domain generalization for object recognition with multi-task autoencoders[C]//Proceedings of 2015 IEEE international Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2015: 2551-2559.
 - [18] LI D, YANG Y X, SONG Y Z, et al. Learning to generalize: meta-learning for domain generalization[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Piscataway: IEEE Press, 2018.
 - [19] DOU Q, CASTRO D C, KAMNITSAS K, et al. Domain generalization via model-agnostic learning of semantic features[EB]. arXiv preprint, 2019, arXiv: 1910.13580.
 - [20] GAO B, GOUK H, YANG Y, et al. Loss function learning for domain generalization by implicit gradient[C]//Proceedings of the 39nd International Conference on Machine Learning. [S.l.:s.n.], 2022: 7002-7016.
 - [21] LI Y, YANG Y, ZHOU W, et al. Feature-critic networks for heterogeneous domain generalization[C]//Proceedings of the 36nd International Conference on Machine Learning. [S.l.:s.n.], 2019: 3915-3924.
 - [22] KHOSLA A, ZHOU T H, MALISIEWICZ T, et al. Undoing the damage of dataset bias[C]//Proceedings of the European Conference on Computer Vision. Heidelberg: Springer, 2012: 158-171.
 - [23] CARLUCCI F M, D'INNOCENTE A, BUCCI S, et al. Domain generalization by solving jigsaw puzzles[C]//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2019: 2229-2238.
 - [24] SOHL-DICKSTEIN J, WEISS E A, MAHESWARANATHAN N, et al. Deep unsupervised learning using nonequilibrium thermodynamics [C]//Proceedings of the 32nd International Conference on Machine Learning. New York: ACM, 2015: 2256-2265.
 - [25] KINGMA D, SALIMANS T, POOLE B, et al. Variational diffusion models[EB]. arXiv preprint, 2021, arXiv: 2107.00630.
 - [26] KONG Z, PING W, HUANG J, et al. Diffwave: a versatile diffusion model for audio synthesis[EB]. arXiv preprint, 2020, arXiv:2009.09761.
 - [27] SAHARIA C, HO J, CHAN W, et al. Image super-resolution via iterative refinement[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 45(4):4713-4726.
 - [28] SINHA A, SONG J, MENG C, et al. D2c: diffusion-denoising models for few-shot conditional generation[EB]. arXiv preprint, 2021, arXiv: 2106.06819.
 - [29] HO J, SAHARIA C, CHAN W, et al. Cascaded diffusion models for high fidelity image generation[EB]. The Journal of Machine Learning Research. 2022, 3(1):2249-2281.
 - [30] SONG Y, SOHL-DICKSTEIN J, KINGMA D P, et al. Score-based generative modeling through stochastic differential equations[EB]. arXiv preprint, 2020, arXiv:2011.13456.
 - [31] DHARIWAL P, NICHOL A. Diffusion models beat GANs on image synthesis[EB]. arXiv preprint, 2021, arXiv: 2105.05233.
 - [32] SAHARIA C, CHAN W, CHANG H W, et al. Palette: image-to-image diffusion models[C]//ACM SIGGRAPH 2022 Conference Proceedings. New York: ACM, 2022: 1-10.
 - [33] SONG J, MENG C, ERMON S. Denoising diffusion implicit models[EB]. arXiv preprint, 2020, arXiv:2010.02502.
 - [34] LU C, ZHOU Y, BAO F, et al. DPM-solver: a fast ODE solver for diffusion probabilistic model sampling in around 10 steps[EB]. arXiv preprint, 2022, arXiv: 2206.00927.
 - [35] MATSUURA T, HARADA T. Domain generalization using a mixture of multiple latent domains[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Piscataway: IEEE Press, 2020: 11749-11756.
 - [36] CHEN K Y, ZHUANG D, CHANG J M. Discriminative adversarial domain generalization with meta-learning based cross-domain validation[J]. Neurocomputing, 2022, 467:418-426.
 - [37] SEGU M, TONIONI A, TOMBARI F. Batch normalization embeddings for deep domain generalization[J]. Pattern Recognition. 2023, 135:109-115.

作者简介



童煜钧 (1999-), 男, 北京邮电大学人工智能学院博士生, 主要研究方向为迁移学习、域泛化。



王荷清 (2000-), 女, 北京邮电大学人工智能学院硕士生, 主要研究方向为计算机视觉与机器学习。



罗悦恒 (2001-), 北京邮电大学模式识别实验室硕士生, 主要研究方向为计算机视觉。



宁文欣（2001- ），女，北京邮电大学人工智能学院硕士生，主要研究方向为人工智能、计算机视觉。



喻雯晴（1999- ），女，北京邮电大学人工智能学院硕士生，主要研究方向为计算机视觉、细粒度图像识别。



关曼丹（2000- ），女，北京邮电大学人工智能学院硕士生，主要研究方向为人工智能、计算机视觉、少样本学习。

黄柯彦（1977- ），男，就职于北京空间飞行器总体设计部。

张加迅（1971- ），男，就职于北京空间飞行器总体设计部。

马占宇（1982- ），男，博士，北京邮电大学人工智能学院教授、博士生导师，主要研究方向为模式识别、机器学习、计算机视觉、非高斯概率模型、贝叶斯网络。