

中图法分类号: TP391.4 文献标识码: A 文章编号: 1006-8961(2025)08-2660-15

论文引用格式: Lu T Y, Lyu F, Zhou T, Yao R and Hu F Y. 2025. Continual testing time domain adaptive image classification method. Journal of Image and Graphics, 30(8):2660-2674(陆霆洋, 吕凡, 周涛, 姚睿, 胡伏原. 2025. 持续测试时域自适应图像分类方法. 中国图象图形学报, 30(8):2660-2674)[DOI:10.11834/jig.240739]

## 持续测试时域自适应图像分类方法

陆霆洋<sup>1,2,3</sup>, 吕凡<sup>4</sup>, 周涛<sup>5</sup>, 姚睿<sup>6</sup>, 胡伏原<sup>1,2,3\*</sup>

1. 苏州科技大学电子与信息工程学院, 苏州 215009; 2. 江苏省工业智能低碳技术工程中心, 苏州 215009;
3. 苏州市智能低碳技术应用重点实验室, 苏州 215009; 4. 中国科学院自动化研究所, 北京 100089;
5. 北方民族大学计算机科学与工程学院, 银川 750021; 6. 中国矿业大学计算机科学与技术学院, 徐州 221000

**摘要:** 目的 持续测试时适应(continual test-time adaption, CTTA)旨在不使用任何源数据情况下,使源预训练模型适应持续变化的目标域。目前持续测试时适应主要依赖于自训练方法,在基于平均教师模型框架下将数据增强后样本的预测值作为伪标签,构建一致性损失函数实现模型的自训练。然而,现有方法中使用随机数据增强策略忽视了域间差异的重要性,导致模型稳定性和泛化性失衡等问题,使得在某些域间进行知识转移变得更具挑战性。为此,提出一种面向域间差异的持续测试时适应方法,聚焦于计算机视觉领域中的图像分类任务,探讨如何通过持续测试时适应技术提升模型对新域的适应能力。**方法** 首先,提出一种基于域间差异的弹性数据增强策略。通过构建表示域间特征风格的Gram矩阵,计算相邻域间的差异,选取合适的弹性因子控制数据增强的强度,在数据预处理层面考虑域间差异性,使模型能更好地适应域复杂多变的情况。其次,提出一种全局弹性对称交叉熵损失函数。将基于域间差异计算取得的弹性因子应用于伪标签生成以及一致性损失函数的构建中,在模型优化层面考虑域间差异性,增强模型对不同域变化下的理解和适应能力。最后,提出一种基于置信度的伪标签自纠错策略。在弹性数据增强下,强数据增强通过对原始数据进行较大程度的变换来实现,模型在预测过程中可能面临预测偏差的问题,而弱数据增强涉及较小程度的变换,不会显著改变基本特征,模型对其预测的置信度较高。该策略利用高置信度的弱数据增强预测值对强数据增强的预测值进行自纠错,减少误差积累现象。**结果** 在CIFAR10-C、CIFAR100-C和ImageNet-C 3个数据集上与多种先进算法进行比较,相较于基线方法CoTTA,错误率分别降低了约2.3%、2.7%和3.6%。在CIFAR10-C数据集中进行了消融实验,进一步验证了各个模块的有效性。为了符合更实际的域变化场景,在CIFAR100-C设计了域随机输入实验,结果显示本文方法在域随机输入的情况下错误率低于现有方法,对比基线平均错误率降低了3.9%,证明了本文方法可以有效地评估域间关系,并部署灵活策略以提升模型对持续变化目标域的适应能力。**结论** 本文算法平衡了模型在持续测试时适应场景中的泛化性和稳定性,有效减少了误差积累现象。

**关键词:** 持续测试时适应(CTTA); Gram矩阵; 域间差异; 全局对称交叉熵; 弹性数据增强; 伪标签自纠错; 持续学习

## Continual testing time domain adaptive image classification method

Lu Tingyang<sup>1,2,3</sup>, Lyu Fan<sup>4</sup>, Zhou Tao<sup>6</sup>, Yao Rui<sup>5</sup>, Hu Fuyuan<sup>1,2,3\*</sup>

1. School of Electronic and Information Engineering, Suzhou University of Science and Technology, Suzhou 215009, China;

收稿日期: 2024-12-12; 修回日期: 2025-01-24; 预印本日期: 2025-02-01

\* 通信作者: 胡伏原 fuyuanhu@usts.edu.cn

基金项目: 国家自然科学基金项目(62406323, 62476189); 国家资助博士后研究人员计划(GZC20232993); 中国博士后科学基金资助项目(2024M753496)

Supported by: National Natural Science Foundation of China(62406323, 62476189); Postdoctoral Fellowship Program of China Postdoctoral Science Foundation(GZC20232993); China Postdoctoral Science Foundation(2024M753496)

2. *Jiangsu Industrial Intelligent and Low-Carbon Technology Engineering Center, Suzhou 215009, China;*
3. *Suzhou Key Laboratory of Intelligent and Low-Carbon Technology Application, Suzhou 215009, China;*
4. *Institute of Automation, Chinese Academy of Sciences, Beijing 100089, China;*
5. *College of Computer Science and Engineering, North Minzu University, Yinchuan 750021, China;*
6. *School of Computer Science and Technology, China University of Mining and Technology, Xuzhou 211000, China*

**Abstract: Objective** In recent years, Deep Neural Networks have demonstrated exceptional performance in numerous computer vision tasks, including image classification, dense prediction, and image segmentation. The success of these tasks largely depends on the consistency between the test data distribution and the training data. However, in real-world application scenarios, consistency is often disrupted due to unknown changes in factors such as weather and lighting. This disruption leads to increased domain diversity and results in poor generalization and performance degradation of deployed models. The concept of continuous test-time adaptation (CTTA) has been proposed to address this challenge, with the aim of enabling the adaptation of pre-trained models to the evolving target data distribution. CTTA is designed to allow source pre-trained models to adapt to the continuously changing target domain without using any source data. Existing CTTA primarily focuses on self-training model adaptation, which employs pseudo-labels from model-predicted data-augmented samples within the mean teacher framework to achieve self-training. However, this study posits that the essence of CTTA is the transformation of inter-domain feature styles. Through experiments, this study has found that existing methods, which use random data augmentation strategies, ignore the importance of inter-domain differences. The use of simple and singular data augmentation leads to issues such as insufficient model stability and generalization, which hinders knowledge transfer across certain domains. To this end, this study proposes a CTTA method that addresses inter-domain inconsistency differences, particularly within image classification tasks in the field of computer vision. The research explores ways to improve the adaptability of models to new domains through continuous testing adaptation techniques. **Method** First, this study posits that the essence of domain variation lies in the variation in feature styles between domains, which acknowledges differences in features across target domains. Consequently, the study aims to construct a flexible data augmentation strategy based on domain differences. Calculating inter-domain differences is a key technique for measuring the distributional differences between features of different domains, which is specifically important in the field of CTTA. The Gram matrix, as a tool for assessing feature style differences, has been widely applied in this domain. The use of the Gram matrix more accurately quantifies and understands the differences in feature distribution between domains, which help calculate the appropriate elasticity factor for subsequent flexible data augmentation operations. This approach considers inter-domain differences from a data preprocessing perspective, which enables the model to adapt flexibly to continuously changing domains. Second, based on the differences in inter-domain feature styles, the study proposes a global elastic symmetric cross-entropy consistency loss function. This function incorporates the elasticity factor, which is calculated based on inter-domain differences, into the pseudo-label and loss function levels. It considers inter-domain differences from the perspective of model training optimization, which involves constructing a global elastic symmetric cross-entropy loss function that balances model generalization and stability. The elastic symmetric cross-entropy loss function is a loss function that combines differences in inter-domain feature styles, which adaptively adjusts pseudo-label outputs by constructing a flexible data augmentation strategy. The goal of this method is to achieve a balance between the generalization and stability of the model. Specifically, it enhances the adaptability of the model to new domains by dynamically adjusting the pseudo-labels and the weights between forward and backward cross-entropy based on differences in inter-domain feature styles. Finally, a confidence-based pseudo-label correction strategy is proposed. Given that the study implements controllable elastic data augmentation, data are elastically enhanced according to the degree of inter-domain differences, which produces a diverse set of outputs. However, elastic data augmentation may result in a large number of strongly augmented samples, which can obscure sample characteristics. As a result, the model has difficulty correctly predicting the true labels of strongly augmented samples, which leads to low-quality pseudo-labels and error accumulation. Therefore, using high-confidence weak data augmentation predictions to correct strong data augmentation predictions reduces the issue of low-quality pseudo-labels caused by high-intensity data augmentation during the CTTA phase. In this way, the accumulation of errors is effectively

suppressed. **Result** This study conducted comprehensive comparative experiments on the CIFAR10-C, CIFAR100-C, and ImageNet-C datasets, which were followed by comparison of results with those of various advanced algorithms. The experimental results indicate that the algorithm proposed in this study has achieved significant performance improvements over the baseline method CoTTA on these datasets. Specifically, on the CIFAR10-C, CIFAR100, and ImageNet-C datasets, the error rates were reduced by approximately 2.3%, 2.7%, and 3.6%, respectively. These results demonstrate that the algorithm can effectively enhance the robustness and accuracy of the model across datasets of varying difficulty and complexity. Further ablation experiments were conducted on the CIFAR10-C dataset to verify the effectiveness of each module within the algorithm. Ablation experiments are a method of controlling variables, where modules are added individually or in combination to test their impact on overall performance. This approach helps understand the specific contributions of each module to the performance of the algorithm, which provides a basis for algorithm optimization. Furthermore, the study designed experiments with random domain inputs to more closely align with real-world domain variation scenarios. The experimental results show that the elastic symmetric cross-entropy based on domain difference detection has a lower error rate under random domain input conditions than existing methods, which verifies the effectiveness of the approach. This enhancement in capability is crucial because models in practice often need to make predictions in constantly changing environments. In other words, the model should better adapt to unknown domain changes, which maintains high performance in practical applications. **Conclusion** The presented algorithm effectively balances the generalization and stability of the model in scenarios of continuous test-time adaptation while significantly reducing the accumulation of errors during the test-time adaptation process. This balance is crucial for machine learning models to maintain performance in dynamically changing environments. Our research focuses not only on the theoretical foundations of the algorithm but also on its effectiveness in practical applications. In future research, we plan to further explore the performance and applicability of continuous test-time adaptation algorithms in more open and complex environments. Therefore, we need to design and implement experimental setups that are closer to real-world conditions to more accurately evaluate and optimize the algorithms. Our goal is to ensure that these algorithms are not only robust in theory but also demonstrate strong adaptability and robustness in practical applications. Furthermore, we will consider the scalability and computational efficiency of the algorithms, given that these factors are crucial for the practical deployment of the algorithms in large-scale or resource-constrained environments. We hope that through these studies, we can provide more powerful and reliable tools for the fields of machine learning and artificial intelligence to address the increasingly complex challenges of the real world.

**Key words:** continual test-time adaptation (CTTA); Gram matrix; domain difference; global symmetric cross-entropy; elastic data augmentation; pseudo label self-correction; continual learning

## 0 引言

近年来,深度学习在计算机视觉、自然语言处理以及语音识别等任务中展现出了卓越的性能(王卓和翟绍军等,2024),这一成功很大程度依赖于测试数据与训练数据的分布一致性。然而在测试场景中,环境经常发生变化,导致域分布发生改变,例如随着天气、光线等场景的持续变化(Luo等,2019; Stacke等,2021;高常鑫等,2024),部署模型泛化能力不足,性能随之下降。近年来,持续测试时适应(continual test-time adaptation, CTTA)(Wang和Qu,2024; Lee等,2024; Liu等,2024; Tan等,2024; Liu等,2021)受到广泛关注,其目标是通过持续学习

(Parisi等,2019;周大蔚等,2023)未标记的测试数据,从而更新源预训练模型(Cao和Saukh,2023;Hou等,2024;Qu等,2024)适应新的数据域。出于隐私保护和实际操作的考虑,在持续测试时适应的过程中,不再访问源域数据,且每个目标域数据只能被访问一次,这使得持续测试时域适应具有较大挑战(Boudiaf等,2023;Kundu等,2020;Yang等,2021;Long等,2015)。

现有持续测试时适应的研究方法主要分为基于平均教师模型和熵最小化两类算法。熵最小化的方法,例如TENT(test entropy)(Wang等,2021)通过使用测试熵最小化更新Batchnorm层中的可训练参数来适应测试数据。GTTA(gradual test-time adaptation)(Marsden等,2024)使用混合和风格转换等技术

人为地构建中间域,最小化熵实现平滑领域之间的过渡。基于平均教师模型的方法,例如 CoTTA (continual test-time adaption) (Wang 等, 2022) 构建了平均教师模型框架,使用权重和增强平均预测解决误差积累问题,并进一步与随机恢复相结合,以促进新知识转移和减少持续域变化(李文斌等, 2024)引起的误差积累。在此基础上, PETAL (probabilistic life-long test-time adaptation with self-training prior) (Brahma 和 Rai, 2023) 进一步开发了一种数据驱动的参数恢复技术,以更好地将模型与参数空间中的原始源配置对齐,从而增强模型鲁棒性。DSS (dynamic sample selection) (Wang 等, 2024) 通过动态阈值选择来区分样本质量,仅对高置信度样本应用正学习以减少噪声,同时对所有样本实施负学习以提升模型对不确定性的适应能力,有效降低误差积累。AdaContrast (Chen 等, 2022) 利用对比学习和在线伪标签重新细化来改进特征表示,减少噪声伪标签 (Chen 等, 2024) 对自适应质量的影响。RMT (robust mean teacher) (Döbler 等, 2023) 引入了稳健的平均教师框架,通过在多个适应阶段保持稳定性以解决重复的领域转换问题。

上述 CTTA 方法中,数据增强方法得到广泛运用。数据增强 (Shorten 和 Khoshgoftaar, 2019) 可模拟出广泛的数据分布,在不直接访问目标域情况下提升模型适应新域的能力,获得更好的泛化性。然而,传统 CTTA 方法通常直接使用随机数据增强策略实现模型的自训练 (Kumar 等, 2020), 忽视了域变化间存在一定差异性 (Li 等, 2017; Jin 等, 2022; Tarvainen 和 Valpola, 2016)。具体而言,不同目标域间存在图像的颜色偏移、背景复杂度、纹理差异或光照条件等差异,这些不同可能导致简单的随机数据增强无法有效模拟目标域的真实分布 (Zhu 等, 2025), 从而影响模型的表现。因此,本文认为在持续测试时适应中,需要实时检测并分析相邻域之间的差异性,部署可控的弹性数据增强策略,达到模型可塑性和稳定性的平衡,从而保证知识的持续适应和有效保留。

本文通过比较不同数据增强方法以验证构建可控的弹性数据增强的重要性。本文认为在 CTTA 中,强数据增强通过增加样本多样性显著提升模型泛化能力,弱数据增强通过轻微变化提升模型的稳定性能力。如图 1 所示,在域 Defocus 向域 Glass 以

及域 Contrast 向域 Elastic 转变时,由于域间差异较小,给予模型过高强度的数据增强样本导致过高的泛化性 (图 1(a) 所示),使模型遗忘 (de Lange 等, 2022) 相似域的知识信息,降低模型的稳定性,因此域 Glass 和域 Elastic 的错误率升高。但是数据增强过低也会导致模型的性能减弱,在差异性过大的域间转变时 (Zoom→Snow), 弱数据增强无法给予模型多样化的数据进行自训练更新,减弱了模型的泛化能力,错误率高于其他方法。因此,重视域间的关系,构建实现模型稳定和可塑性平衡的弹性数据增强算法成为持续测试时适应的关键。

针对以上问题,本文提出一种面向域间差异的持续测试时适应方法,在数据预处理和模型训练优化角度引入基于域间差异的弹性因子,并构建伪标签自纠错算法,提升模型对当前目标域分布变化的适应。本文方法基于平均教师模型,设计了 3 个模块: 1) 基于域间差异的弹性数据增强策略。该策略通过构建表示域间特征风格的 Gram 矩阵 (Sastry 和 Oore, 2020), 计算相邻域的差异,选取合适的弹性因子控制数据增强的强度,在数据预处理考虑域间差异性,提出一种可控的弹性数据增强算法,使模型能更好地适应域复杂多变的情况。2) 全局弹性对称交叉熵损失函数。将基于域间差异性计算取得的弹性因子应用于伪标签生成以及一致性损失函数的构建中,在模型优化层面考虑域间差异性,提出一种全局弹性对称交叉熵损失函数,增强模型对不同域间变化下的理解和适应能力。3) 基于置信度的伪标签自纠错策略。在弹性数据增强下,强数据增强是通过对原始数据进行较大程度的变换来实现,模型在预测过程中可能面临预测偏差的问题,而弱数据增强涉及较小程度的变换,不会显著改变数据的基本特征,模型对其预测的置信度较高。该策略利用高置信度的弱数据增强预测值对强数据增强的预测值进行自纠错,减少持续测试时的误差积累现象。实验方面,本文在 CIFAR10-C、CIFAR100-C 和 ImageNet-C 数据集上与多种算法进行比较,各项评价指标都得到了一定程度的提高,平均错误率降低至 14.9%、29.8% 和 59.1%, 并且基于 3 个数据集设置了域随机输入,模仿更为现实的开放环境 (Wu 等, 2024), 基于基线平均错误率降低了 3.9%, 进一步验证了算法在复杂场景下的适用性。

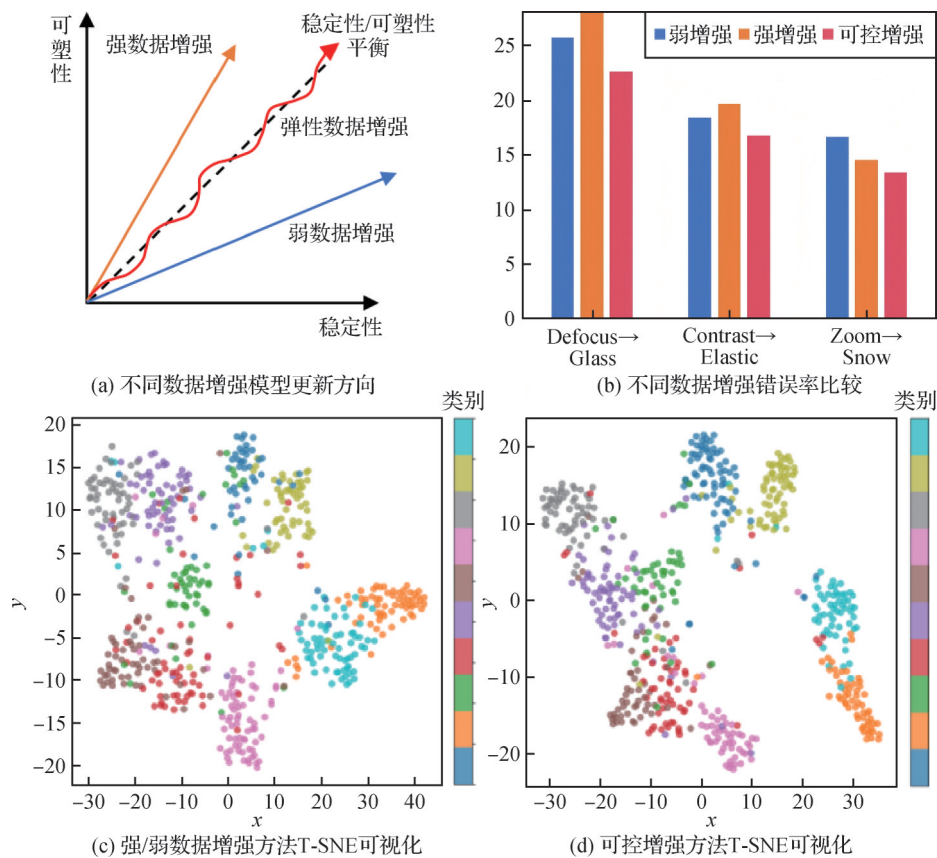


图1 不同强度数据增强的对比实验

Fig. 1 Compared experiment of different data augmentation ((a) different data augmentation model update direction; (b) error rate comparison for different data augmentation; (c) T-SNE visualization of strong/weak data augmentation method; (d) T-SNE visualization of controllable data augmentation method)

## 1 持续测试时适应的问题设定

给定一个在源域上训练的模型  $q_\theta(y^s | x^s)$ , 其参数  $\theta$  在源数据  $(x^s, y^s)$  上进行训练, 并在多个目标数据集  $D_1, D_2, \dots, D_t$  上进行测试, 其中  $D_t = \{x_i\}_{i=1}^N$ ,  $N$  表示该目标域的样本数量, 目标域的分布可以随时间变化或重复出现(例如自动驾驶场景中的晴天、阴天、雨天、雾天和夜间域)(Muandet 等, 2013; Sakaridis 等, 2021; Laskin 等, 2020)。本文的目标是在推理时以在线方式提高现有模型泛化性能。本文设定在时间步  $t$ , 提供目标域数据  $D_t$  作为输入, 模型  $q_\theta$  需要预测  $q_\theta(x_t)$  并相应地适应未来域的输入。其中  $x_t$  的数据分布是不断变化的, 并且该模型根据在线预测进行评估。本文构建平均教师模型框架下实现模型的自训练, 在时间步  $t = 1$  时, 使用源预训练模型对学生以及教师模型进行初始化。在测试期间, 通过最小化交叉熵 (cross entropy) 更新学生

$p_{\theta_t}(\theta_t \rightarrow \theta_{t+1})$ , 具体为

$$L_{ce}(q, p) = - \sum_{t=1}^T \sum_{c=1}^C q_{\theta_c}(x_t) \times \log p_{\theta_c}(x_t) \quad (1)$$

式中,  $q_{\theta_c}$  是教师模型预测  $x_t$  为类别  $c$  的概率,  $p_{\theta_c}$  是学生模型预测  $x_t$  为类别  $c$  的概率。通过构建平均教师模型框架, 模型能够在没有源数据的情况下, 通过自我优化训练提升性能。

## 2 方法

### 2.1 本文方法

本文提出一种面向域间差异的持续测试时适应方法, 本文的框架概述如图2所示。首先, 本文构建可控的弹性数据增强策略, 利用 Gram 矩阵计算变化域特征风格, 构建余弦相似度计算域间差异, 选取合适弹性调节因子部署到数据增强层面, 从数据预处理的角度考虑到了域间差异性的影响。其次, 本文提出全局弹性对称交叉熵损失函数。在

模型训练优化过程中也应考虑域间差异的影响,该方法将基于域间差异计算得到的弹性因子应用于伪标签的生成和一致性损失函数的构建,强化模型在不同域间差异下适应能力。最后,本文提出一种基于高置信度的伪标签自纠错策略,通过将弱数据增强的高置信度预测值对强数据增强对应样本和类别索引的预测值进行自纠正,减少了误差积累的现象。

### 2.2 基于域差异的弹性数据增强策略

近年来,CTTA和无监督域适应(unsupervised domain adaptation,UDA)(Patel等,2015)领域通过在训练期间利用数据增强后的模型的预测值作为伪标签,模拟人类学习过程中的自我学习纠正机制,得到广泛使用以提升模型性能。然而,本文发现给予模型随机的数据增强样本会导致一系列问题,数据的过度增强导致的稳定性问题,数据的过弱增强导致无法有效模拟目标域的真实分布,致使模型的泛化性能不足。因此,希望计算基于域间差异,构建可控的弹性数据增强策略。具体为

$$\mathbf{x}^s, \mathbf{x}^w = \mathbf{x} + \sum_{i=1}^n \omega \times \mathcal{A}(\mathbf{x}_i | \sigma_i^s, \sigma_i^w) \quad (2)$$

式中,  $\mathcal{A}(\cdot)$  表示随机函数用于对域变化样本进行数

据增强操作。 $\omega$  为由域间差异计算获取的弹性调节因子。 $\sigma_i^s$  是一个超参数,表示第  $i$  个强扰动的随机因子, $\sigma_i^w$  表示第  $i$  个弱扰动的随机因子, $\mathbf{x}^s$  和  $\mathbf{x}^w$  分别为经过强数据增强和弱数据增强后的样本。为了计算域间差异的弹性因子,本文构建 Gram 矩阵以表示变化域的特征风格(Zhang等,2023),Gram 矩阵通过计算特征映射之间的内积来量化不同特征之间的相关性,首先,给定输入图像  $\mathbf{x}_i$ , 经过特征提取网络  $\phi(\mathbf{x}_i)$ , 得到特征图。具体为

$$\mathbf{F}_i = \phi(\mathbf{x}_i) \quad (3)$$

为了使 Gram 矩阵能够捕获不同通道特征的全局分布,本文对特征图进行了展平操作  $\mathbf{F}_i = \text{Flatten}(\mathbf{x}_i)$ ,  $\mathbf{F}_i \in \mathbf{R}^{C \times (W \times H)}$ , 展平后其空间维度  $H$  和  $W$  被合并为单一维度,通道特征的空间信息被忽略,转而关注空间特征的整体分布。得到展平后的特征图用于描述特征的风格信息 Gram 矩阵,具体为

$$\mathbf{G}_F = \mathbf{F}_i \cdot \mathbf{F}_i^T, \quad \mathbf{G}_T \in \mathbf{R}^{C \times C} \quad (4)$$

式中,  $\mathbf{F}_i^T$  是域特征矩阵的转置,每一行  $\mathbf{F}_i$  是通道  $i$  的特征向量,格拉姆矩阵展开为

$$\mathbf{G}_{ij} = \sum_{k=1}^H \sum_{l=1}^W \mathbf{F}_{ikl} \cdot \mathbf{F}_{jkl} \quad (5)$$

式中,  $\mathbf{G}_{ij}$  是通道  $i$  和通道  $j$  的点积,表示通道之间的

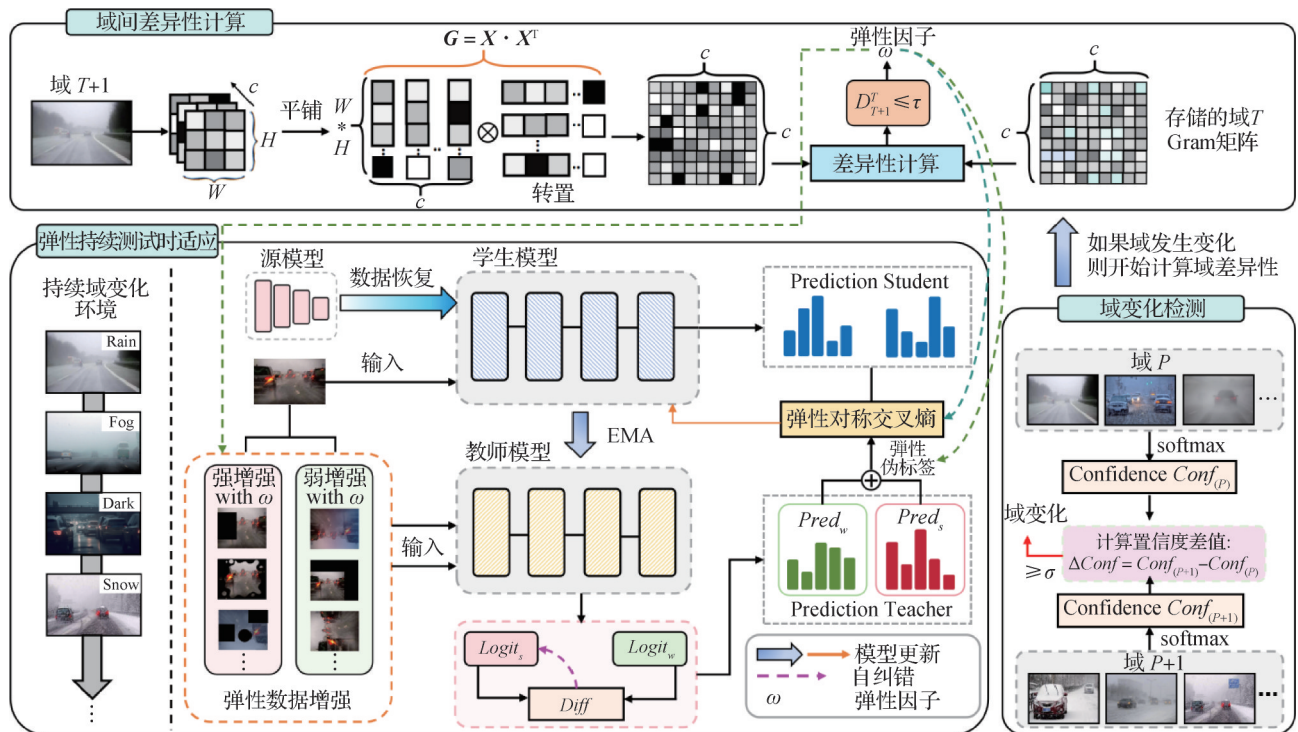


图2 面向域间差异的持续测试时适应方法研究

Fig. 2 Research on a continual test-time adaptation method for inconsistent differences between domains

相似性。不同通道  $F_i, F_j$  代表了不同种类的特征,例如纹理、边缘和颜色等,如果  $G_{ij}$  值接近 0,说明它们在空间上的特征分布几乎独立,意味着通道间没有明显的相关性。基于此,本文构建余弦相似度计算相邻变化域间的风格差异性程度。具体为

$$S_{\text{sim}}(\mathbf{G}_{F_i}, \mathbf{G}_{F_{i+1}}) = \frac{\sum_{i=1}^{c^2} \mathbf{G}_{F_i, i} \cdot \mathbf{G}_{F_{i+1}, i}}{\sqrt{\sum_{i=1}^{c^2} \mathbf{G}_{F_i, i}^2} \cdot \sqrt{\sum_{i=1}^{c^2} \mathbf{G}_{F_{i+1}, i}^2}} \quad (6)$$

式中,  $S_{\text{sim}}$  代表 Gram 矩阵的余弦相似度。本文设置超参阈值  $\tau$  判断域间风格的差异性程度,选取合适的弹性调节因子  $\omega$ 。若  $S_{\text{sim}}(\mathbf{G}_{F_i}, \mathbf{G}_{F_{i+1}}) > \tau$ ,说明相邻域相似性程度高,则  $\omega$  取 0.4;如果  $S_{\text{sim}}(\mathbf{G}_{F_i}, \mathbf{G}_{F_{i+1}}) < \tau$ ,则  $\omega$  取 0.8,选取弹性调节因子  $\omega$  用于构建可控的弹性数据增强策略,从而更好地适应复杂多变的新域。

弹性数据增强策略的核心在于根据域间差异的大小,动态调整数据增强的强度。构建基于域间差异的弹性数据增强算法,能够使模型更好地适应域复杂多变的情况。当相邻域间差异性小,表面模型已经能够较好地捕捉到域间的共同特征,需保持原有域知识,在这种情况下,模型无需过高的数据增强强度,在保留原有域知识的情况下略微增强一些泛化性即可,以免破坏模型已经学习到的稳定特征信息。这种方法有助于模型在面对新域时,能够灵活地适应并保持性能稳定。

### 2.3 全局弹性对称交叉熵

2.2 节构建了一种基于域差异性的弹性数据增强策略,旨在从样本数据预处理的角度考虑域间的差异性,平衡模型的稳定性与可塑性。本节进一步提出,在模型训练优化过程中也应考虑域间差异性的影响。为此,本文引入一种全局弹性对称交叉熵一致性损失函数。该函数通过集成基于域间差异的调节因子,使得模型能够在模仿教师模型的预测分布与关注自身预测结果之间实现动态平衡。这样的设计旨在优化模型的泛化性能,同时确保其稳定性,从而在模型训练和更新中实现更优的平衡。

伪标签是通过模型预测生成的,在 CTTA 学习中起到关键作用。将基于域间差异的弹性因子传递给伪标签,可以使这些标签更好地反映模型对不同域的理解和适应性。首先,本文赋予模型输出的伪标签以域间差异知识的弹性因子,实现使模型能够

高效地适应持续变化的新变化域。具体为

$$\tilde{q}(\mathbf{x}) = \frac{1}{N} (\omega \times \sum_{i=1}^N q(\mathbf{x}^i) + (1 - \omega) \times \sum_{i=1}^N q(\mathbf{x}^{e_i})) \quad (7)$$

式中,  $\omega$  表示弹性调节因子,通过域间差异计算获得。 $\tilde{q}(\mathbf{x})$  代表经过弹性数据增强后的目标域样本的弹性伪标签。通过将弹性因子部署于伪标签,模型可以根据域间差异性动态调整伪标签,使得模型能够灵活应对数据分布的持续变化。

为了实现控制模型训练的优化方向,本文赋予损失函数基于域间差异性的调节因子,构建全局弹性对称交叉熵损失函数。具体为

$$L_{\text{SCE}} = \sum_{c=1}^C \omega \times (\tilde{q}_c(\mathbf{x}_i) \log p_c(\mathbf{x}_i)) + \sum_{c=1}^C (1 - \omega) \times p_c(\mathbf{x}_i) \log \tilde{q}_c(\mathbf{x}_i) \quad (8)$$

式中,第 1 项对应于交叉熵损失  $L_{\text{ce}}$ ,增大权重将使学生模型更加关注模仿教师模型的预测分布,而第 2 项是反向交叉熵损失  $L_{\text{rce}}$ ,增大权重可以使学生模型更加关注其自己的预测结果。例如,若域间差异性大,则赋予第 1 项交叉熵损失  $L_{\text{ce}}$  高的权重,可以实现学生模型尽量接近教师模型的强数据增强样本的预测分布,提升模型的泛化性能。通过式(1)、式(7)和式(8)加入基于域差异性的弹性调节因子,实现持续测试时适应的弹性知识全局传递。

使用式(8)更新学生模型  $\theta_t \rightarrow \theta_{t+1}$  后,本文使用学生权重通过指数移动平均(exponential moving average, EMA)算法更新教师模型的权重,实现模型的自训练。具体为

$$\theta'_{t+1} = a \times \theta'_t + (1 - a) \times \theta_{t+1} \\ a = \min(1 - \frac{1}{t+1}, \frac{e^3 - 1}{2}) \quad (9)$$

式中,  $a$  为平均移动(Wang 等, 2022)的平滑因子,随着  $t$  增加而逐渐接近 1,因此,早期  $\theta_{t+1}$  权重更大,更倾向于学习新知识,这种动态调整可以使得算法在初期快速学习,随着  $t$  的增加,  $a$  逐渐接近 1,在后期训练阶段模型更加稳定,算法更倾向于保持已有的知识,减少了误差积累现象的产生。

此外,在实际应用中域变化不可预测。本文提出基于置信度的域变化检测机制。通过实验发现,随着域的变化,预测置信度会出现显著波动,这一现象可以用于估计目标域是否发生变化。为此,设定了一个阈值  $s$ ,持续计算预测置信度之间的差值。本文设置基于置信度的域变化检测机制,具体为

$$\Delta \text{Conf}(t+1) = \text{Conf}(t+1) - \text{Conf}(t) \quad (10)$$

式中,  $\text{Conf}(t)$  为变化域  $t$  输出的置信度, 一旦超过了设定的阈值  $s$ , 系统将触发后续计算相邻域间的差异程度的算法。通过这种方法, 可以更精确地量化域间差异性, 从而为持续域适应提供决策支持。

#### 2.4 伪标签自纠错

前文构建了可控的弹性数据增强, 根据域间差异性程度弹性地进行增强, 输出多样性的数据。然而弹性数据增强中可能存在输出大量强数据增强样本的情况, 强数据增强可能导致样本特性变得模糊, 使得模型难以正确预测经过强数据增强后样本的正确标签, 从而产生伪标签质量低下的问题, 造成误差积累的现象。此外, 弱数据增强由于变化程度较小, 模型更有可能预测正确的标签, 伪标签的可信度相对较高, 误差积累问题较轻。

因此, 本文利用高置信度的弱数据增强样本预测值对强数据增强样本预测值进行有效的修正, 提高模型对强数据增强样本的预测准确性, 减少误差积累, 提高模型整体性能。首先通过设置置信度阈值  $\varepsilon$ , 选取对应高置信度的样本标签  $n$  和类别标签  $c$ , 具体为

$$\mathbf{x}^w(n|c) = \sum_{n=1}^N \prod_{i,c} (q_{\theta_i}(\mathbf{x}^w) \geq \varepsilon) \quad (11)$$

式中,  $\Pi$  是门控函数, 用于输出弱数据增强输出置信度大于阈值  $\varepsilon$  时对应的样本索引  $n$  和对应的类别索引  $c$ 。  $N$  为弱数据增强样本的数量。通过保存的索引去寻找强数据增强对应的置信度, 对强数据增强可能错误预测的置信度进行自纠错。具体为

$$\mathbf{y}_{\text{cor}}^s = \sigma \times q_{\theta_i}(\mathbf{x}_n^w[c]) + (1 - \sigma) \times q_{\theta_i}(\mathbf{x}_n^s[c]) \quad (12)$$

式中,  $\sigma$  是自适应修正权重, 通过预测分布的 KL (Kullback-Leible) 散度计算获得, 具体为

$$\sigma = KL(q_{\theta_i}(\mathbf{x}^w) \parallel q_{\theta_i}(\mathbf{x}^s)) \quad (13)$$

预测值的 KL 散度是两个分布之间的差异度量, 如果 KL 散度较小, 表示当前样本经强数据增强的预测与经弱数据增强的预测相似, 预测的准确性较高, 需要修正的强数据增强预测值较少。通过以上操作, 实现使用高置信度的弱数据增强类别预测值对强数据增强进行优化, 实现了教师模型伪标签的修正, 减少了知识传递中模型的误差积累现象。

综上所述, 本文提出一种面向域间差异的全局弹性自纠错方法, 首先构建基于域间差异的弹性数

据增强策略。通过构建表示域间特征风格的 Gram 矩阵, 计算相邻域间的差异, 选取合适的弹性因子控制数据增强的强度。当域间差异性较低时, 策略将降低数据增强的强度, 以维护模型的稳定性; 当域间差异性较高时, 策略将增强数据增强的强度, 以强化模型的泛化能力。使模型能更好地适应域复杂多变的情况。其次, 提出一种全局弹性对称交叉熵损失函数。将基于域间差异性计算取得的弹性因子应用于伪标签生成以及一致性损失函数的构建中, 增强模型对不同域间变化下的适应能力。最后提出基于置信度的伪标签自纠错策略 (见图 3)。该策略利用高置信度的弱数据增强预测值对强数据增强的预测值进行自纠错, 减少持续测试时的误差积累现象。

算法 1 面向域间差异的持续测试时适应算法。

- 1) 初始化: 源模型  $h_{\theta}(\mathbf{x})$ ;
- 2) 输入: 每个时刻步长  $t$  和当前域的数据流  $\mathbf{x}_t$ 。
- 3) 计算域差异性以及选择数据增强调节因子  $\omega$ 。
- 4) 弹性数据增强和输出弹性伪标签 (式(1))。
- 5)  $\mathbf{y}_{\text{cor}}^s \rightarrow \mathbf{y}^s$  伪标签自纠错 (式(12))。
- 6) 通过式(8)的一致性损失更新学生模型  $p_{\theta}(\mathbf{x})$ 。
- 7) 通过指数移动平均 (EMA) 算法更新教师模型  $q_{\theta}(x)$ 。
- 8) 输出: 更新学生模型  $p_{\theta}(\mathbf{x})$  和更新教师模型  $q_{\theta}(x)$ 。
- 9) 结束。

## 3 实验部分

### 3.1 实验数据集

本文开展了一系列实验, 以验证所提方法的有效性。在图像处理领域, 对所提的方法进行了评估, 测试了 3 个基准任务:

1) CIFAR-10-C。在 CIFAR-10 数据集的基础上进行扩展, 包含了 10 个类别的  $32 \times 32$  像素彩色图像。该数据集引入了 15 种不同类型的扰动, 每种扰动均设有 5 个不同的严重程度等级, 这些干扰被应用于 CIFAR-10 的测试图像, 最终形成了一个包含 10 000 幅图像的数据集。

2) CIFAR-100-C。在 CIFAR-100 数据集的基础

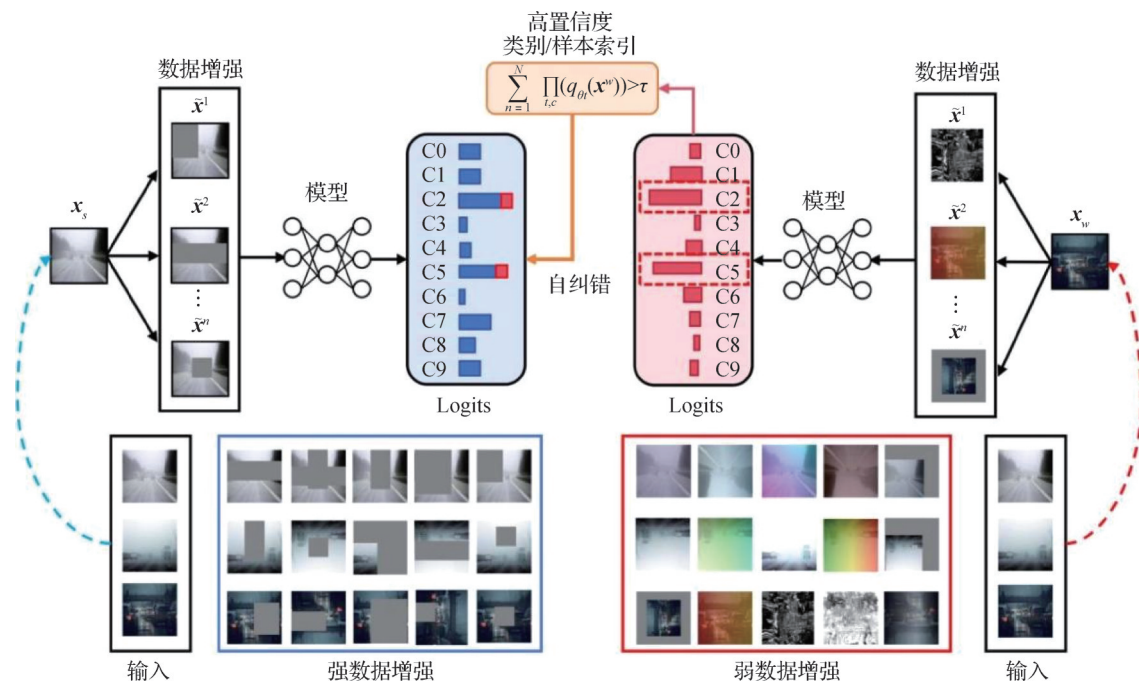


图3 基于置信度的伪标签自纠错策略

Fig. 3 Pseudo label self-correction strategy based on confidence

上进行扩展,涵盖了100个类别的 $32 \times 32$ 像素彩色图像。与CIFAR-10-C类似,它也包含了15种不同的扰动类型,每种扰动同样有5个严重程度等级,这些干扰被应用于CIFAR-100的测试图像,总共包含了10 000幅图像的数据集。

3) ImageNet-C。在ImageNet数据集的基础上进行了扩展,该数据集包含了超过1 400万幅图像,覆盖了20 000多个类别。ImageNet-C同样包含了15种不同的扰动类型,每种干扰都有5个严重程度等级,这些干扰被应用于ImageNet的测试图像集。

### 3.2 实验细节与参数

本文中所有实验均使用PyTorch进行,在单张NVIDIA RTX3090显卡进行训练。本文严格遵循CoTTA来设置实验,即源数据是不可被再次使用的。采用标准预训练的WideResNet(Zagoruyko和Kornblith, 2016)、ResNeXt-29(Yin等, 2019)和ResNet-50(Croce等, 2020)作为CIFAR10-C、CIFAR100-C和ImageNet-C的源模型。Batchsize设置为200。

### 3.3 主要对比实验

本研究旨在评估所提出的模型在CIFAR10-C、CIFAR100-C和ImageNet-C任务中的性能,并将其与包括源基线在内的6种流行方法进行比较。本文方法在3类数据集上训练源模型,并执行测试时自适应操作。在测试阶段,模型将处理15种不同的损坏类

型数据集。如表1—表3所示,直接在CIFAR10-C数据集上应用源域模型的平均误差率高达43.5%。TENT方法通过顺序适应目标域,虽然在域变化初期表现出较低的错误率,但在后期阶段却遭遇了显著的误差积累,导致在CIFAR100-C上的长期运行错误率高达60.9%。CoTTA通过构建平均教师—学生模型,并通过数据增强显著提升模型泛化能力,实现域变化测试时间下的模型的自训练,为持续测试时自适应(CTTA)方法建立了一个基准,将误差降低至16.2%。RMT(Döbler等, 2023)采用了对称交叉熵优化自训练框架,以改善梯度传播特性,但未能充分考虑域变化中特征风格转换的本质,导致在风格近似域转换时错误率不可避免地上升。DSS(Wang等, 2024)识别了数据流中的高质量与低质量样本,但未能充分考虑模型在表示不同质量数据时的动态变化趋势。Ada(Chen等, 2022)在实际模型部署中考虑了计算效率,尽管以牺牲部分性能为代价。BeCoTTA(Lee等, 2024)通过MoDE(mixture of domain low-rank experts)模块和领域自适应路由以及领域专家协同损失,实现对连续未见领域的有效适应,同时保留先前学到的知识,但忽略了域间的相互关系。PALM(pushing adaptive learning rate mechanisms)(Maharana等, 2024)通过利用模型预测不确定性和梯度空间信息来选择和调整特定层的学习

表1 不同方法在CIFAR10-C的误差率对比实验

Table 1 Comparative experiment of error rates under different methods on CIFAR10-C dataset

方法	Gau	Sho	Imp	Def	Gla	Mot	Zoo	Sno	Fro	Fog	Bri	Con	Ela	Pix	Jpe	平均
Source(源域模型)	72.3	65.7	72.9	46.9	54.3	34.8	42.0	25.1	41.3	26.0	9.3	46.7	26.6	58.5	30.3	43.5
TENT(ICLR2021)	24.9	20.6	28.6	14.4	31.1	16.5	14.1	19.1	18.6	18.6	12.2	20.3	25.7	20.8	24.9	20.7
Ada(CVPR2022)	29.1	22.5	30.0	14.0	32.7	14.1	12.0	16.6	14.9	14.4	8.1	10.0	21.9	17.7	20.0	18.5
CoTTA(CVPR2022)	24.3	21.3	26.6	11.6	27.6	12.2	10.3	14.8	14.1	12.4	7.6	10.6	18.3	13.4	17.3	16.2
RMT(CVPR2023)	24.0	20.4	25.6	12.6	25.4	14.2	12.2	15.4	15.1	14.1	10.3	13.7	<b>17.1</b>	13.5	<b>16.0</b>	16.7
DSS(WACV2024)	24.1	21.3	25.4	11.7	26.9	12.2	10.5	14.5	14.1	12.5	7.8	10.8	18.0	13.1	17.3	16.0
BeCoTTA(ICML2024)	22.9	19.1	26.9	11.2	27.5	12.7	10.4	14.7	14.3	12.4	<b>7.2</b>	<b>9.4</b>	20.9	15.2	20.2	16.3
PALM(AAAI2025)	25.9	<b>18.1</b>	<b>22.7</b>	12.4	<b>25.3</b>	13.2	10.8	<b>13.5</b>	13.6	12.2	8.5	11.9	17.9	<b>12.0</b>	16.8	15.7
本文	<b>22.5</b>	19.9	24.6	<b>11.0</b>	25.4	<b>11.6</b>	<b>10.0</b>	13.6	<b>13.5</b>	<b>11.7</b>	7.7	10.5	<b>17.1</b>	12.7	16.7	<b>14.9</b>

注:加粗字体表示各列最优效果。Gau:Ganssiam;Sho:Shot;Imp:Impluse;Def:Defocus;Gla:Glass;Mot:Motion;Zoom:Zoom;Sno:Snow;Fro:Frost;Bri:Brightness;Con:Contrast;Ela:Elastic;Pix:Pixelate;Jpe:Jpeg。

表2 不同方法在CIFAR100-C的误差率对比实验

Table 2 Comparative experiment of error rates under different methods on CIFAR100-C dataset

方法	Gau	Sho	Imp	Def	Gla	Mot	Zoo	Sno	Fro	Fog	Bri	Con	Ela	Pix	Jpe	平均
Source(源域模型)	73.0	68.0	39.4	29.3	54.1	30.8	28.8	39.5	45.8	50.3	29.5	55.1	37.2	74.7	41.2	46.4
TENT(ICLR2021)	37.2	35.8	41.7	37.9	51.2	48.3	48.5	58.4	63.7	71.1	70.4	82.3	88.0	88.5	90.4	60.9
Ada(CVPR2022)	42.4	36.8	38.6	27.7	40.1	29.1	27.6	32.9	30.8	38.2	25.9	28.3	33.9	33.3	36.2	33.4
CoTTA(CVPR2022)	40.1	37.7	39.7	26.9	38.0	27.9	26.4	32.8	31.8	40.3	24.7	26.9	32.5	28.3	33.5	32.5
RMT(CVPR2023)	40.5	36.1	36.3	27.7	<b>33.9</b>	28.5	26.4	29.0	<b>29.0</b>	<b>32.5</b>	25.1	27.4	<b>28.2</b>	<b>26.3</b>	<b>29.3</b>	30.4
DSS(WACV2024)	39.7	36.0	37.2	26.3	35.6	27.5	25.2	31.4	30.0	37.8	24.2	<b>26.0</b>	30.0	<b>26.3</b>	31.3	30.9
BeCoTTA(ICML2024)	42.1	38.0	42.2	30.2	42.9	31.7	29.8	35.1	33.9	38.5	27.9	32.0	36.7	31.6	39.9	35.5
PALM(AAAI2025)	37.3	<b>32.5</b>	<b>34.9</b>	26.2	35.3	27.5	<b>24.6</b>	<b>28.9</b>	29.2	34.1	<b>23.6</b>	27.0	31.1	26.6	34.1	30.2
本文	<b>36.7</b>	33.2	35.0	<b>25.6</b>	<b>33.9</b>	<b>27.1</b>	25.1	29.3	<b>29.2</b>	35.5	<b>23.6</b>	26.3	30.0	26.6	32.2	<b>29.8</b>

注:加粗字体表示各列最优效果。

率,以适应动态变化的数据域,但未能意识到域变化的本质是特征的变化。在与现有方法的比较分析中,本研究提出的模型在CIFAR10-C、CIFAR100-C和ImageNet-C 3个数据集上展现了卓越的性能,显著降低了平均错误率至14.9%、29.8%和59.1%。本文方法基于Gram矩阵量化域间风格差异,并实现了可控的弹性数据增强算法。此外,构建了一个全局弹性的对称交叉熵损失函数,有效地调和了持续测试时适应中模型的稳定性与可塑性之间的平衡。通过引入伪标签自纠错算法,在域变化的后期阶段保持了错误率的稳定性,避免了显著的上升趋势。进一步地,通过优化策略额外降低了2.3%的误差,将平均误差率降至14.9%。特别是在CIFAR10-C数

据集上面对Defocus、Zoom、Brightness、Motion、Fog、Contrast等损坏类型时,本文方法展现出了低于12%的平均误差率。如图4所示,本文方法在T-SNE(t-distributed stochastic neighbor embedding)可视化验证实验中也表现了较好的结果。这些结果表明,本文方法不仅实现了CTTA模型的弹性训练,而且增强了跨域知识转移的平滑性,为自适应深度学习领域提供了新的视角和方法论。

### 3.4 域随机输入实验

本文提出面向域间差异的持续测试时适应方法,为了验证方法的有效性,本节设置了域随机输入的实验,旨在评估所提出的方法在随机域输入下的CIFAR100-C任务上的性能,并将其与包括源基线在

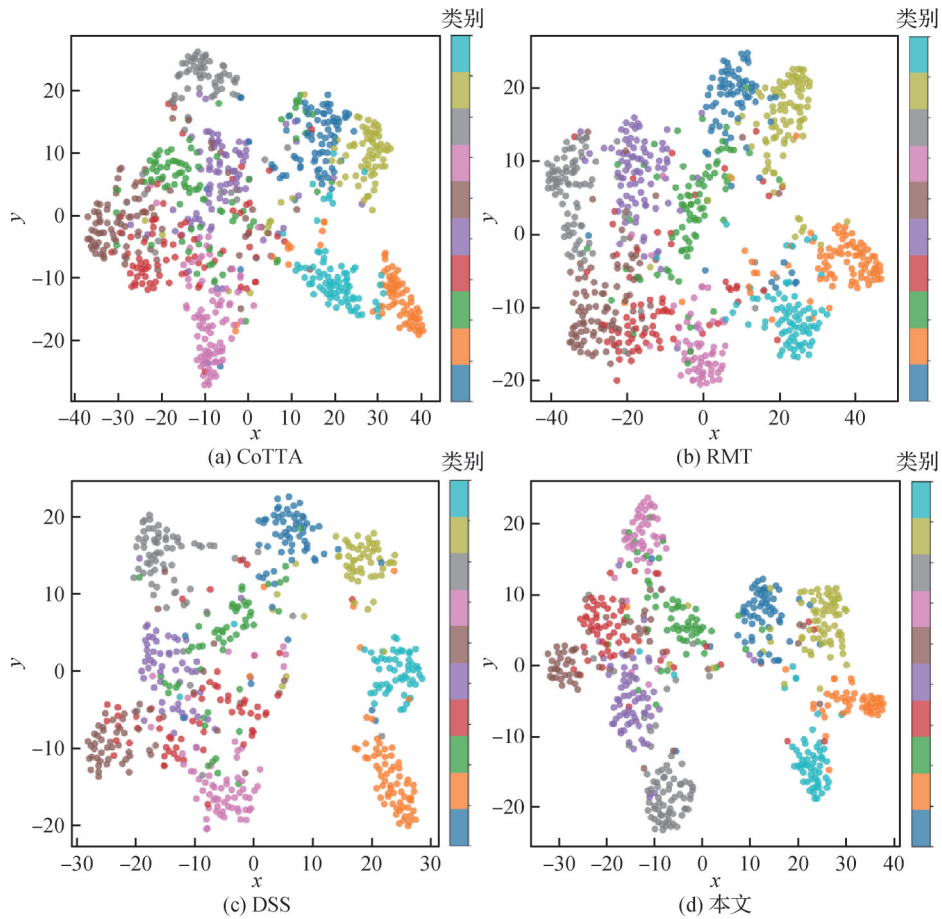


图4 T-SNE可视化实验

Fig. 4 T-SNE visualization experiment ((a) CoTTA; (b) RMT; (c) DSS; (d) ours)

表3 不同方法在 ImageNet-C 的误差率对比实验

Table 3 Comparative experiment of error rates under different methods on ImageNet-C

方法	Gau	Sho	Imp	Def	Gla	Mot	Zoo	Sno	Fro	Fog	Bri	Con	Ela	Pix	Jpe	平均
Source(源域模型)	97.8	97.1	98.2	81.7	89.8	85.2	78.0	83.5	77.1	75.9	41.3	94.5	82.5	79.3	68.6	82.0
TENT(ICLR2021)	81.6	74.6	72.7	77.6	73.8	65.6	55.3	61.6	63.0	51.7	38.2	72.1	50.8	47.4	53.3	62.6
Ada(CVPR2022)	82.9	80.9	78.4	81.4	78.7	72.9	64.0	63.5	64.5	53.5	38.4	66.7	54.6	49.3	53.0	65.6
CoTTA(CVPR2022)	84.7	82.1	80.6	81.3	79.0	68.6	57.5	60.3	60.5	48.3	36.6	66.1	<b>47.3</b>	41.2	46.0	62.7
RMT(CVPR2023)	80.2	76.4	74.5	77.1	74.4	66.2	57.6	57.0	<b>59.1</b>	48.0	39.1	60.6	<b>47.3</b>	42.5	<b>43.3</b>	60.2
DSS(WACV2024)	82.3	78.4	76.7	81.9	77.8	66.9	60.9	<b>50.8</b>	60.9	<b>47.7</b>	<b>35.4</b>	69.0	47.5	<b>40.9</b>	46.2	62.2
BeCoTTA(ICML2024)	84.1	74.3	72.2	77.4	71.9	63.4	55.1	57.2	61.2	50.7	36.4	66.1	49.2	45.6	48.4	60.4
PALM(AAAI2025)	81.1	73.3	<b>70.9</b>	77.0	71.9	62.3	<b>53.9</b>	56.7	60.7	50.4	36.3	65.9	48.1	45.3	48.0	60.1
本文	<b>78.4</b>	<b>72.5</b>	73.6	<b>74.4</b>	<b>64.1</b>	<b>56.8</b>	56.9	59.5	59.4	49.5	39.6	<b>59.9</b>	47.5	42.3	44.9	<b>59.1</b>

注:加粗字体表示各列最优效果。

内的6种流行方法进行比较。与主要对比实验相似,本文方法在CIFAR100数据集上训练源模型,并在CIFAR100-C上执行CTTA操作。在测试阶段,模型将处理15种的损坏类型数据集,进行4种域的随

机输入,然后对4种随机输入情况的错误率取均值得到最终随机域输入下的模型性能。如表4所示,本研究提出的模型在CIFAR100-C数据集上展现了卓越的性能,显著降低了平均错误率至29.6%。本

文方法在域间差异较小时,减弱了模型的泛化能力,实现了域间知识的平滑过渡,在域间差异较大时,训练了模型的泛化能力,以应对下阶段复杂多变的新域。实验结果表明,本文方法通过探究域间差异性的角度实现了持续测试时适应的弹性训练,为今后开放环境下的持续测试时自适应以及无监督学习提供了新的理论基础和方法研究。

表4 域随机输入情况下在CIFAR100-C上的对比实验

Table 4 Experiment under domain random input on CIFAR100-C dataset					
					/%
方法	Random 1	Random 2	Random 3	Random 4	平均
Source (源域模型)	45.9	44.9	44.4	42.1	44.3
TENT	65.5	64.6	65.0	62.5	63.9
Ada	36.2	36.7	36.1	34.8	35.8
CoTTA	34.2	33.8	33.9	32.6	33.5
RMT	32.1	31.3	31.0	29.9	31.0
DSS	34.7	34.2	34.0	33.1	33.8
本文	<b>30.1</b>	<b>29.4</b>	<b>29.7</b>	<b>29.2</b>	<b>29.6</b>

注:加粗字体表示各列最优效果。

### 3.5 消融实验

本文对CIFAR10-C任务场景进行了消融研究,结果如表5所示。实验结果表明,直接使用弹性数据增强算法(elastic data augmentation, EDA),在一些相似特征风格之间进行域变化时,性能有了进一步提升,例如域Jpeg、Motion、Shot,在基于EDA基础上构建全局弹性对称交叉熵(global elastic symmetric cross entropy, GE-SCE),错误率降低了1.2%,实现了提升模型的性能,这表明GE-SCE的构建已经实现了控制模型自训练优化方向,通过引入具有域差异信息的弹性因子,实现模型的全局自训练优化,可以有效地平衡模型的泛化性和稳定性能。最后将基于置信度的伪标签自纠错(pseudo label self correction, PLSC)策略整合到方法中时,发现在域变化后期阶段,错误率有了明显的降低,这表明在整个域变化阶段,强数据准确的样本的错误预测值可能已经被修正,提高了伪标签的质量,有效减少了模型中存在的误差积累现象。

### 3.6 基于伪标签自纠错实验

在本文的研究中,设置了一项可视化实验以验证伪标签自纠错模块的有效性,如图5所示。图5

表5 消融实验

Table 5 Ablation experiment

GE-SCE	EDA	PLSC	平均/%
-	-	-	16.4
√	-	-	15.4
-	√	-	15.6
-	-	√	15.6
√	√	-	15.2
-	√	√	15.4
√	-	√	15.3
√	√	√	<b>14.9</b>

注:加粗字体表示最优效果,“√”表示使用该策略,“-”表示未使用。

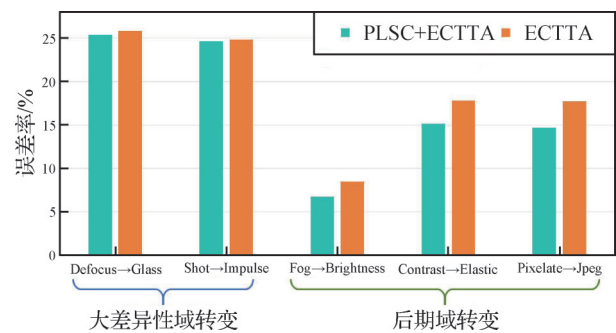


图5 伪标签自纠错实验对比可视化

Fig. 5 Visualization of pseudo label self-correction experiments

中,ECTTA(elastic continual test-time adaption)表示全局弹性持续时适应。

本研究基于提出的弹性数据增强算法,利用弱数据增强得到的高置信度预测值对强数据增强的预测结果进行纠正。在某些相似性较低的域转换过程中,一些增强策略较强的数据增强预测值信息被校正,这可能会削弱模型的泛化能力,导致模型的可塑性降低。例如,在如域Shot到域Imp,以及域Def到域Gla这样的大差异的域转换中,误差率与不使用伪标签自纠错的结果相近。此外,由于该策略在整个训练阶段持续对强数据增强的输出进行校正,提高了伪标签的质量,并减少了错误标签在模型训练中的积累。因此,在领域转换的后期阶段,采用伪标签自纠错的方法展现出了更低的误差率(如域Fog到域Jpeg的所有阶段)。这一发现表明,尽管在某些情况下模型的泛化能力受到限制,但通过持续的质量提升和误差积累的减少,伪标签自纠错策略在长期训练中能够提高模型的性能。

### 3.7 弹性因子部署对比实验

在本文的研究中,设置了一项对比实验以验证弹性因子全局部署实验的有效性。通过将基于域间差异性调节因子部署到数据增强层面、伪标签层面和对称交叉熵损失函数层面,得到弹性数据增强策略、弹性的伪标签和弹性损失函数,构建一种全局弹性对称交叉熵函数。如表6所示,单独对数据增强、伪标签和损失函数赋予弹性因子性能提升较弱,原因是在基于域间差异性的调节因子相关知识信息随着模型的传递已经被弱化。特别地,单独给SCE赋予弹性因子误差率为15.5%,性能基本上没有提升。这表明具有弹性知识的信息需要在模型传递中不断全局重放,防止信息丢失。

表6 弹性因子部署对比实验

Table 6 Comparison experiment of elastic factor deployment

DA	PL	SCE	平均误差率/%
-	-	-	15.6
√	-	-	15.3
-	√	-	15.4
-	-	√	15.5
√	√	-	15.2
-	√	√	15.2
√	-	√	15.3
√	√	√	<b>14.9</b>

注:加粗字体表示最优效果。DA(data augmentation):数据增强层面;PL(pseudo label):伪标签层面;SCE(symmetrical cross entropy):对称交叉熵损失函数层面。“√”和“-”分别表示选择和未选择对应模块。

### 3.6 弹性调节因子阈值优化实验

本实验旨在深入分析和优化弹性调节因子的阈值选择,验证其在不同域间差异下对模型性能的影响。如表7所示,通过系统地调整弹性调节因子的取值,探讨了阈值选择的合理性,并寻找最优的阈值设置,以增强模型在持续测试时适应(CTTA)中的泛化能力和稳定性。实验分为两种情况:域间差异较大 $S_{sim} > \tau$ 和域间差异较小 $S_{sim} \leq \tau$ 。当域间相似性程度较高时,选择较小的弹性因子有助于保留之前域的知识;而当域间相似性程度较低时,选择较大的弹性因子可以提高模型的泛化性能,增强其可塑性。经过多次实验,本文确定了较优的阈值 $\omega$ ,以维护在

表7 弹性因子阈值选择实验

Table 7 Experiment on threshold selection of elasticity factor

$\omega$ 取值		平均误差率/%
$S_{sim} > \tau$	$S_{sim} \leq \tau$	
0.1	0.6	15.5
0.2	0.5	15.3
0.3	0.7	15.0
0.4	0.8	<b>14.9</b>
0.5	0.9	15.1
0.6	0.7	15.4

注:加粗字体表示最优结果。

持续域变化下稳定性和可塑性的平衡。

## 4 结论

本文针对持续测试时适应,提出面向域间差异的弹性自纠错方法,旨在实现自训练模型训练的稳定性和可塑性的平衡,解决现有方法在复杂多变的变化域场景中的适应性弱的问题。首先,提出一种面向域间差异的弹性数据增强策略。该算法通过构建表示域间特征风格的Gram矩阵,计算相邻域间的差异,选取合适的弹性因子控制数据增强的强度,构建弹性数据增强算法,从数据预处理的角度考虑到域间差异性,使模型能更好地适应域复杂多变的情况。其次,引入全局弹性对称交叉熵损失函数。将基于域间差异性计算取得的弹性因子应用于伪标签生成以及一致性损失函数的构建中,从模型训练优化的角度考虑到域间差异,增强模型对不同域间变化下的理解和适应能力。再次,提出基于置信度的伪标签自纠错策略。该策略利用高置信度的弱数据增强预测值对强数据增强的预测值进行自纠错,减少持续测试时的误差积累现象。未来的研究将致力于进一步优化弹性自纠错持续测试时适应策略,提高模型在实时应用中的实用性和有效性。通过这些改进,期望该框架能够在更开放、更广泛的应用场景中发挥出色的性能。

### 参考文献(References)

Boudiaf M, Denton T, Van Merriënboer B, Dumoulin V and Triantafyllou E. 2023. In search for a generalizable method for source free

- domain adaptation//Proceedings of the 40th International Conference on Machine Learning. Honolulu, USA: PMLR: 2914-2931 [DOI: 10.48550/arXiv.2302.06658]
- Brahma D and Rai P. 2023. A probabilistic framework for lifelong test-time adaptation//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 3582-3591 [DOI: 10.1109/CVPR52729.2023.00349]
- Cao N and Saukh O. 2023. Geometric data augmentations to mitigate distribution shifts in pollen classification from microscopic images [EB/OL]. [2024-12-12]. <https://arxiv.org/pdf/2311.11029.pdf>
- Chen D, Wang D Q, Darrell T and Ebrahimi S. 2022. Contrastive test-time adaptation//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 295-305 [DOI: 10.1109/CVPR52688.2022.00039]
- Chen T C, Duan Y, Li D, Qi L, Shi Y H and Gao Y. 2024. PG-LBO: enhancing high-dimensional Bayesian optimization with pseudo-label and Gaussian process guidance//Proceedings of the 38th AAAI Conference on Artificial Intelligence and 36th Conference on Innovative Applications of Artificial Intelligence and 14th Symposium on Educational Advances in Artificial Intelligence. Vancouver, Canada: AAAI: 11381-11389 [DOI: 10.1609/aaai.v38i10.2901]
- Croce F, Andriushchenko M, Sehwag V, Debenedetti E, Flammarion N, Chiang M, Mittal P and Hein M. 2020. Robustbench: a standardized adversarial robustness benchmark [EB/OL]. [2024-12-12]. <https://arxiv.org/pdf/2010.09670.pdf>
- de Lange M, Aljundi R, Masana M, Parisot S, Jia X, Leonardis A, Slabaugh G and Tuytelaars T. 2022. A continual learning survey: defying forgetting in classification tasks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(7): 3366-3385 [DOI: 10.1109/TPAMI.2021.3057446]
- Döbler M, Marsden R A and Yang B. 2023. Robust mean teacher for continual and gradual test-time adaptation//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 7704-7714 [DOI: 10.1109/CVPR52729.2023.00744]
- Gao C X, Xu Z Z, Wu D Y, Yu C Q and Sang N. 2024. Deep learning-based real-time semantic segmentation: a survey. *Journal of Image and Graphics*, 29(5): 1119-1145 (高常鑫, 徐正泽, 吴东岳, 余昌黔, 桑农. 2024. 深度学习实时语义分割综述. *中国图象图形学报*, 29(5): 1119-1145) [DOI: 10.11834/jig.230659]
- Hou F, Yuan J, Yang Y, Liu Y, Zhang Y, Zhong C, Shi Z C, Fan J P, Rui Y and He Z Q. 2024. DomainVerse: a benchmark towards real-world distribution shifts for tuning-free adaptive domain generalization [EB/OL]. [2024-12-12]. <https://arxiv.org/pdf/2403.02714.pdf>
- Jin D, Jin Z J, Hu Z T, Vechtomova O and Mihalcea R. 2022. Deep learning for text style transfer: a survey. *Computational Linguistics*, 48(1): 155-205 [DOI: 10.1162/coli\_a\_00426]
- Kumar A, Ma T Y and Liang P. 2020. Understanding self-training for gradual domain adaptation//Proceedings of the 37th International Conference on Machine Learning. PMLR: 5468-5479 [DOI: 10.48550/arXiv.2002.11361]
- Kundu J N, Venkat N, Rahul M V and Venkatesh Babu R. 2020. Universal source-free domain adaptation//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 4544-4553 [DOI: 10.1109/CVPR42600.2020.00460]
- Laskin M, Srinivas A and Abbeel P. 2020. CURL: contrastive unsupervised representations for reinforcement learning//Proceedings of the 37th International Conference on Machine Learning. Vienna, Austria: PMLR: 5639-5650
- Lee D, Yoon J and Hwang S J. 2024. BECoTTA: input-dependent online blending of experts for continual test-time adaptation [EB/OL]. [2024-12-12]. <https://arxiv.org/pdf/2402.08712.pdf>
- Li W B, Xiong Y K, Fan Z C, Deng B, Cao F Y and Gao Y. 2024. Advances and trends of continual learning. *Journal of Computer Research and Development*, 61(6): 1476-1496 (李文斌, 熊亚锟, 范祉辰, 邓波, 曹付元, 高阳. 2024. 持续学习的研究进展与趋势. *计算机研究与发展*, 61(6): 1476-1496) [DOI: 10.7544/issn1000-1239.202220820]
- Li Y J, Fang C, Yang J M, Wang Z W, Lu X and Yang M H. 2017. Universal style transfer via feature transforms//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, USA: Curran Associates Inc.: 385-395 [DOI: 10.5555/3294771.3294808]
- Liu J M, Xu R, Yang S Q, Zhang R R, Zhang Q Z, Chen Z H, Guo Y D and Zhang S H. 2024. Continual-MAE: adaptive distribution masked autoencoders for continual test-time adaptation//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 28653-28663 [DOI: 10.1109/CVPR52733.2024.02707]
- Liu Y J, Kothari P, Van Delft B, Bellot-Gurlet B, Mordan T and Alahi A. 2021. TTT++: when does self-supervised test-time training fail or thrive//Proceedings of the 36th International Conference on Neural Information Processing Systems. Beijing, China: NIPS: 21808-21820
- Long M S, Cao Y, Wang J M and Jordan M I. 2015. Learning transferable features with deep adaptation networks//Proceedings of the 32nd International Conference on Machine Learning. Lille, France: JMLR.org: 97-105 [DOI: 10.48550/arXiv.1502.02791]
- Luo Y W, Zheng L, Guan T, Yu J Q and Yang Y. 2019. Taking a closer look at domain shift: category-level adversaries for semantics consistent domain adaptation//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 2502-2511 [DOI: 10.1109/CVPR.2019.00261]
- Maharana S K, Zhang B M and Guo Y H. 2024. PALM: pushing adaptive learning rate mechanisms for continual test-time adaptation [EB/OL]. [2024-12-12]. <https://arxiv.org/pdf/2403.10650.pdf>
- Marsden R A, Döbler M and Yang B. 2024. Introducing intermediate domains for effective self-training during test-time//Proceedings of the 34th International Joint Conference on Neural Networks. Yokohama, Japan: IEEE: 1-10 [DOI: 10.1109/IJCNN60899.2024.10651501]
- Muandet K, Balduzzi D and Schölkopf B. 2013. Domain generalization

- via invariant feature representation//Proceedings of the 30th International Conference on Machine Learning. Atlanta, Georgia, USA: JMLR: 10-18
- Parisi G I, Kemker R, Part J L, Kanan C and Wermter S. 2019. Continual lifelong learning with neural networks: a review. *Neural Networks*, 113: 54-71 [DOI: 10.1016/j.neunet.2019.01.012]
- Patel V M, Gopalan R, Li R N and Chellappa R. 2015. Visual domain adaptation: a survey of recent advances. *IEEE signal Processing Magazine*, 32(3): 53-69 [DOI: 10.1109/MSP.2014.2347059]
- Qu S Q, Zou T P, He L H, Röhrbein F, Knoll A, Chen G and Jiang C J. 2024. LEAD: learning decomposition for source-free universal domain adaptation//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 23334-23343 [DOI: 10.1109/CVPR52733.2024.02202]
- Sakaridis C, Dai D X and Van Gool L. 2021. ACDC: the adverse conditions dataset with correspondences for semantic driving scene understanding//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 10765-10775 [DOI: 10.1109/ICCV48922.2021.01059]
- Sastry C S and Oore S. 2020. Detecting out-of-distribution examples with gram matrices//Proceedings of the 37th International Conference on Machine Learning. Vienna, Austria: PMLR: 8491-8501
- Shorten C and Khoshgoftaar T M. 2019. A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1): #60 [DOI: 10.1186/s40537-019-0197-0]
- Stacke K, Eilertsen G, Unger J and Lundström C. 2021. Measuring domain shift for deep learning in histopathology. *IEEE Journal of Biomedical and Health Informatics*, 25(2): 325-336 [DOI: 10.1109/JBHI.2020.3032060]
- Tan J Y, Lyu F, Ni C G, Feng T L, Hu F L, Zhang Z, Zhao S C and Wang L. 2024. Less is more: pseudo-label filtering for continual test-time adaptation [EB/OL]. [2024-12-12]. <https://arxiv.org/pdf/2406.02609.pdf>
- Tarvainen A and Valpola H. 2016. Weight-averaged consistency targets improve semi-supervised deep learning results//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, USA Curran Associates Inc.: 1195-1204 [DOI: 10.5555/3294771.3294885]
- Wang D Q, Shelhamer E, Liu S T, Olshausen B and Darrell T. 2021. Tent: fully test-time adaptation by entropy minimization//Proceedings of 2021 International Conference on Learning Representations. Austria, Vienna: ICLR: 1-8
- Wang Q, Fink O, Van Gool L and Dai D X. 2022. Continual test-time domain adaptation//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 7201-7211 [DOI: 10.1109/CVPR52688.2022.00706]
- Wang Y S, Hong J, Cheraghian A, Rahman S, Ahmedt-Aristizabal D, Petersson L and Harandi M. 2024. Continual test-time domain adaptation via dynamic sample selection//Proceedings of 2024 IEEE/CVF Winter Conference on Applications of Computer Vision. Waikoloa, USA: IEEE: 1701-1710 [DOI: 10.1109/WACV57701.2024.00172]
- Wang Z and Qu S J. 2024. Research progress and challenges in real-time semantic segmentation for deep learning. *Journal of Image and Graphics*, 29(5): 1188-1220 (王卓, 瞿绍军. 2024. 深度学习实时语义分割研究进展和挑战. 中国图象图形学报, 29(5): 1188-1220) [DOI: 10.11834/jig.230605]
- Wu H, Zhao S J, Huang X, Wen C L, Li X and Wang C. 2024. Commonsense prototype for outdoor unsupervised 3D object detection//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 14968-14977 [DOI: 10.1109/CVPR52733.2024.01418]
- Yang S Q, Wang Y X, Van De Weijer J, Herranz L and Jui S. 2021. Generalized source-free domain adaptation//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 8978-8987 [DOI: 10.1109/ICCV48922.2021.00885]
- Yin D, Lopes R G, Shlens J, Cubuk E D and Gilmer J. 2019. A fourier perspective on model robustness in computer vision//Proceedings of the 33th International Conference on Neural Information Processing Systems. Vancouver, Canada: NIPS: 13276-13286 [DOI: 10.48550/arXiv.1906.08988]
- Zagoruyko S and Komodakis N. 2016. Wide residual networks [EB/OL]. [2024-12-12]. <https://arxiv.org/pdf/1605.07146.pdf>
- Zhang L Q, Wang Z S, He J W and Li Y X. 2023. New image processing: VGG image style transfer with gram matrix style features//Proceedings of the 5th International Conference on Artificial Intelligence and Computer Applications. Dalian, China: IEEE: 468-472 [DOI: 10.1109/ICAICA58456.2023.10405398]
- Zhou D W, Wang F Y, Ye H J and Zhan D C. 2023. Deep learning for class-incremental learning: a survey. *Chinese Journal of Computers*, 46(8): 1577-1605 (周大蔚, 汪福运, 叶翰嘉, 詹德川. 2023. 基于深度学习的类别增量学习算法综述. 计算机学报, 46(8): 1577-1605) [DOI: 10.11897/SP.J.1016.2023.01577]
- Zhu Z L, Hong X P, Ma Z H, Zhuang W J, Ma Y H, Dai Y and Wang Y W. 2025. Reshaping the online data buffering and organizing Mechanism for continual test-time adaptation//Proceedings of 2025 International Conference in European Conference on Computer Vision. Milan, Italy: ECCV: 415-433 [DOI: 10.1007/978-3-031-73007-8\_24]

## 作者简介

陆霆洋,男,硕士研究生,主要研究方向为持续学习。

E-mail:tingyanglu@post.usts.edu.cn

胡伏原,通信作者,男,教授,主要研究方向为持续学习和计算机视觉。E-mail:fuyuanhu@usts.edu.cn

吕凡,男,博士后,主要研究方向为持续学习。

E-mail:fan.lyu@cripac.ia.ac.cn

周涛,男,教授,主要研究方向为深度学习。

E-mail:zhoutaonxmu@126.com

姚睿,男,教授,主要研究方向为深度学习。

E-mail:ruiyao@cumt.edu.cn