

基于显著性掩模混合的小样本图像分类

陈亚当, 高宇轩, 卢楚翰, 车洵

引用本文

陈亚当, 高宇轩, 卢楚翰, 车洵. [基于显著性掩模混合的小样本图像分类](#)[J]. 计算机科学, 2025, 52(6): 256-263.

CHEN Yadang, GAO Yuxuan, LU Chuhan, CHE Xun. [Saliency Mask Mixup for Few-shot Image Classification](#) [J]. Computer Science, 2025, 52(6): 256-263.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[面向开放环境的PRNU指纹提纯算法](#)

PRNU Fingerprint Purification Algorithm for Open Environment

计算机科学, 2025, 52(6): 187-199. <https://doi.org/10.11896/jsjcx.241100190>

[基于元增量学习的开放集识别方法](#)

Open Set Recognition Based on Meta Class Incremental Learning

计算机科学, 2025, 52(5): 187-198. <https://doi.org/10.11896/jsjcx.240600162>

[基于对比学习的大气湍流退化图像复原方法](#)

Restoration of Atmospheric Turbulence-degraded Images Based on Contrastive Learning

计算机科学, 2025, 52(5): 171-178. <https://doi.org/10.11896/jsjcx.240200020>

[基于图熵理论的图数据增强研究](#)

Study on Graph Data Augmentation Based on Graph Entropy Theory

计算机科学, 2025, 52(5): 149-160. <https://doi.org/10.11896/jsjcx.240200016>

[基于特征网络对比学习的图协同过滤模型研究](#)

Study on Graph Collaborative Filtering Model Based on FeatureNet Contrastive Learning

计算机科学, 2025, 52(5): 139-148. <https://doi.org/10.11896/jsjcx.240200078>

基于显著性掩模混合的小样本图像分类

陈亚当¹ 高宇轩¹ 卢楚翰¹ 车 洵²

¹ 南京信息工程大学计算机学院 南京 210044

² 南京理工大学计算机科学与工程学院 南京 210094

(adamchen@nuist.edu.cn)

摘要 小样本图像分类解决了传统图像分类在数据量不足时表现不佳的问题,其难点在于如何充分利用稀缺的样本标签数据预测真实的特征分布。一些最新方法采用随机遮挡或混合插值等数据增强方法来提高数据标签样本的多样性和泛化性,但仍然存在以下问题:1)随机遮挡具有不确定性,会出现完全遮挡或暴露前景的情况,导致样本关键信息丢失;2)由于混合插值后的数据分布过于平均,模型难以准确区分不同类别之间的差异和边界。针对上述问题,提出一种基于显著性掩模混合的数据增强方法。首先,通过视觉特征隐蔽融合和置信度裁剪选择策略,对图像关键特征信息进行自适应的筛选与保留;其次,采用视觉特征显著性融合方法,计算出图片中各个区域的重要性,引导图片融合,增加所得图片的多样性和丰富性,使类别边界更加清晰。所提方法在多个标准小样本图像分类数据集(miniImageNet, tieredImageNet, Few-shot CIFAR100 和 Caltech-UCSD Birds-200)上表现出色,优于最先进方法约 0.2%~1%,在小样本图像分类中具有显著的潜力和优势。

关键词:小样本学习;图像分类;对比学习;数据混合;数据增强;显著图

中图分类号 TP391

Saliency Mask Mixup for Few-shot Image Classification

CHEN Yadang¹, GAO Yuxuan¹, LU Chuhan¹ and CHE Xun²

¹ School of Computer Science, Nanjing University of Information Science and Technology, Nanjing 210044, China

² School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China

Abstract Few-shot image classification addresses the problem of poor performance in traditional image classification when data is scarce. The challenge lies in effectively utilizing sparse sample label data to predict the true feature distribution. To tackle this, some recent methods adopt data augmentation techniques such as random mas-king or mixed interpolation to enhance the diversity and generalization of data label samples. However, there are still the following issues: 1) Due to the uncertainty of random masking, situations where the foreground is either completely masked or exposed may occur, leading to the loss of crucial information in samples; 2) Because the data distribution after mixed interpolation tends to be overly uniform, models find it difficult to accurately distinguish differences between different classes, thus failing to effectively delineate boundaries between different categories. To address these problems, this paper proposes a data augmentation method based on Saliency Mask Mixup. Firstly, through Mask Mix(M-Mix) and Confident Clip Selector(CCS), adaptive selection and retention of key feature information in images are performed. Secondly, using Saliency Fuse(SF), the importance of various regions in the image is calculated to guide image fusion, making the resulting images more diverse and rich, thereby making category boundaries clearer. The proposed method demonstrates outstanding performance on multiple standard few-shot image classification datasets(such as miniImage-Net, tieredImageNet, Few-shot CIFAR100, and Caltech-UCSD Birds-200), outperforming state-of-the-art methods by approximately 0.2~1%. These results indicate significant potential and advantages of the proposed method in few-shot image classification.

Keywords Few-shot learning, Image classification, Contrastive learning, Date mixing, Data augmentation, Saliency map

1 引言

随着深度学习技术的发展,图像分类取得了显著的研究

进展,并在许多实际应用中取得了成功。然而,图像分类严重依赖于大量的标注训练数据,且在将所学知识用于目标类别方面存在一定困难。受人类对于小样本学习能力的启发,小

到稿日期:2024-06-20 返修日期:2024-11-21

基金项目:国家自然科学基金(62473201,62477026);江苏省重点研发计划产业前瞻与关键核心技术项目(BE2022161);无锡市产业创新研究院先导技术预研项目

This work was supported by the National Natural Science Foundation of China(62473201,62477026), Jiangsu Province Key R&D Program Industry Outlook and Key Core Technology Projects(BE2022161) and Wuxi Industrial Innovation Research Institute-Pioneering Technology Pre-Research Project.

通信作者:车洵(chexun@njust.edu.cn)

样本图像分类的概念应运而生,其目的在于调整前期从源类别中学得的知识,以适应目标类别,这些知识通常存储在深度嵌入模型中,用于支持和查询图像对的通用匹配。一种备受关注的方法是利用元学习框架来调整模型参数的初始化,以便通过少量的梯度更新使新类别的分类器达到最佳权重^[1-3]。另一方面,基于度量学习的方法^[4-7]也展现出卓越的性能。这些方法旨在通过学习样本之间的度量关系来量化嵌入空间中不同样本之间的相似性或差异性。

虽然上述方法都表现出了卓越的性能,但其难点在于背景干扰的存在。一方面,算法会对图像目标类别产生错误的数据特征归纳,即归纳偏差问题。另一方面,由于标签样本数据稀缺,所学习出的目标类别特征分布与真实类别产生偏离,即存在类别偏差问题。为了解决这些问题,Liu等^[8]通过随机遮挡策略迫使模型学习更加通用的数据规律,充分利用稀缺样本标签数据预测真实特征分布,缓解了归纳偏差问题。Mangla等^[9]和Zhuo等^[10]通过线性插值生成混合样本(其中包含了多个类别的信息),迫使模型在训练过程中学习每个类别的具体特征,从而降低过拟合的风险,缓解类别偏差的问题。

然而,上述方法仍然存在以下问题:1)因为随机遮挡的不确定性,生成的结果可能如图1(a)、图1(c)所示,完全遮挡了关键区域;也可能如图1(b)所示,只遮挡了部分背景区域,导致模型进行分类时过度依赖前景区域的特征,减弱了对背景特征的学习,从而降低了模型对物体之间关系的理解能力;2)如图1(d)所示,由于混合插值后的数据分布过于平均,模型难以准确区分不同类别之间的差异和边界。为了解决上述问题,本文提出一种基于显著性掩模混合的数据增强方法。首先,通过视觉特征隐蔽融合(M-Mix)和置信度裁剪选择策略(CCS)对图像的关键特征信息进行了自适应的筛选与保留,得到的结果如图1(e)所示,保留了图像的关键特征。其中,M-Mix将生成的融合图片作为对比学习的样本,CCS通过裁剪查询集中的图片,筛选出裁剪后置信度较高的一批图片作为最终的查询集。其次,在混合图像和标签时采用视觉特征显著性融合方法(SF)计算出图片中各个区域的重要性,引导图片融合,结果如图1(f)所示,使得到的图片更具多样性和丰富性,类别边界更加清晰。

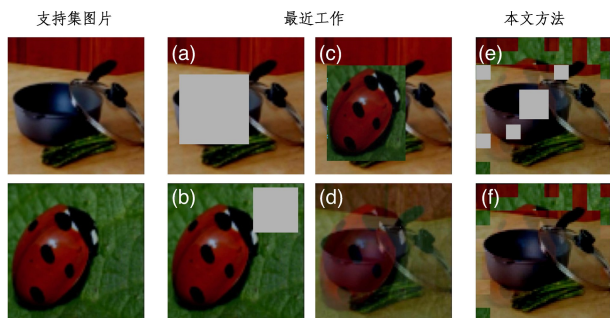


图1 数据增强方法的可视化比较

Fig. 1 Visual comparison of data augmentation methods

2 相关工作

本文的工作涉及元学习^[1]和数据增强^[8-10]在小样本学习中的应用。针对小样本图像分类中样本不足和特征表达能力

欠佳的问题,提出了一种数据增强的方法。本章回顾了与本研究相关的小样本学习领域的工作。

2.1 小样本学习

近年来,小样本学习(Few-Shot Learning)受到越来越多的关注。当前该领域的研究可以分为两大类:基于梯度的方法和基于度量的方法。

基于梯度的方法通过梯度下降来促进模型更快、更准确地适应新任务和新数据,从而增强其泛化能力和性能,目的是从各种小样本任务中提取通用知识,以快速适应新任务。Meta-Baseline^[11]通过在整体分类预训练模型的评估指标上进行元学习,能够在面对新类别或新任务时快速适应并达到较高的准确率。LSFSL^[2]利用数据中的隐含先验信息学习更泛化的特征,提高了小样本学习的鲁棒性和泛化能力。

度量学习的主要目标是提高特征空间内样本之间的相似性或距离的准确度,即最小化类内距离,同时最大化类间距离。通常,传统方法包括两个组成部分:特征提取器和分类器。ProtoNet^[4]使用神经网络学习到的嵌入空间中的每个类别的均值向量来表示该类别,即作为该类别的原型,为后续相关研究提供了重要启发。DeepEMD^[5]通过集成地球移动距离(Earth Mover's Distance, EMD)到网络中,并设计跨引用机制减轻复杂背景和类内外观变化的影响,实现了基于最优匹配的图像区域间距离度量,在少量标注样本下有效进行了小样本图像分类的端到端训练。

2.2 数据增强

数据增强是深度学习的一个关键领域,适当的数据增强可以显著提升深度网络的性能。例如,Mixup^[11]可以在多个广泛使用的数据集上提高分类性能;CutMix^[12]通过交换随机大小的图像块并使用混合标签来改进模型表现。不同于传统的拼接方法,Puzzle Mix^[13]采用更复杂的拼图策略,使图像块的混合更加自然。然而,仅切割前景而忽略背景可能导致模型缺乏上下文信息,从而影响其对图像整体内容的理解。为了解决这一问题,本文提出扩展Mixup和CutMix的方法SF。该方法在保留前景信息的同时,能更好地保留背景信息,从而减少了归纳偏差,并显著改善了小样本学习的效果。

最近,一些小样本学习研究引入了数据增强技术。PatchMix^[8]通过在训练过程中交换图像的随机大小块,并使用混合标签,来提高模型的鲁棒性和泛化能力。S2M2R^[9]通过自监督学习和自蒸馏技术,增强了模型在小样本情况下的分类性能。相比之下,本文提出的M-Mix方法采用动态调整混合比例的自适应策略,使混合后的图像更加符合实际应用场景,从而显著提升了数据增强的效果。

3 问题定义

设 X_{train} , X_{val} 和 X_{test} 分别表示训练集、验证集和测试集,而它们对应的标签分别表示为 Y_{train} , Y_{val} 和 Y_{test} 。这些集合整体表示为 $D_{train} = X_{train}, Y_{train}$; $D_{val} = X_{val}, Y_{val}$; $D_{test} = X_{test}, Y_{test}$ 。 C_{train} , C_{val} 和 C_{test} 分别表示训练集、验证集和测试集中的类别。

小样本学习与传统的监督学习范式略有不同。具体来说,训练集和测试集中的类别完全不同,意味着 C_{train} 和 C_{test} 之间没有交集($C_{train} \cap C_{test} = \emptyset$)。为了模拟数据稀缺的实际

场景,提升模型在少量训练样本下的泛化和快速学习能力,小样本学习通常采用 N-way k-shot 的任务设定。N-way k-shot 是小样本学习中的一种标准设定,表示一次任务中有 N 个类别,每个类别提供 k 个支持样本。支持集(Support Set)用于为模型提供学习样本,查询集(Query Set)用于评估模型的性能。在 N-way k-shot 设定中,支持集包含 N 个不同类别的样本,每个类别包含 k 个样本,这些样本用于指导模型在查询集中识别和分类属于相同类别的新样本。

4 本文方法

4.1 显著性掩模混合模型概述

本文提出了一个基于显著性掩模混合的算法,总体流程如图 2(a)所示。其由 3 个关键模块组成:第一个模块是视觉特征隐蔽融合模块(见图 2(b)),它优化了 infoNCE 损失以适应小样本学习,利用新颖的数据增强技术生成对比学习的样本;第二个模块是置信度裁剪选择器(见图 2(c)),其将查询图片随机剪裁成多块,利用基于置信度的方法筛选出优质的查询样本,增加了查询样本的数量,提高了训练质量;第三个

模块是视觉特征显著性融合模块(见图 2(d)),其利用显著性融合进行数据增强,在融合图片的过程中尽量保留显著区域,为模型提供更优质和更丰富的信息。

构建高质量的样本在对比学习中十分重要,如何构建信息丰富的正样本和负样本是对比学习中的一个挑战。本文期望当前正样本的标签与当前查询图片标签一致,而负样本的标签与查询图片标签不同。因此,对于每个查询实例 x_i^q ,将与其具有相同标签的支持实例作为正样本,而具有不同标签的实例作为负样本。相同的嵌入网络 Φ 被用于查询实例和支持实例。本文的损失函数为:

$$L_i = -\log \frac{\sum_{y_j=y_i^q} e^{f_j^s f_i^q}}{\sum_{y_j=y_i^q} e^{f_j^s f_i^q} + \sum_{y_k \neq y_i^q} e^{f_k^s f_i^q}} \quad (1)$$

其中, L_i 表示第 i 个查询样本的对比学习损失, f_i^q 指代第 i 个查询样本的特征表示, f_j^s 和 f_k^s 分别表示支持样本的第 j 个和第 k 个特征表示。样本 j 与查询样本 i 共享相同的标签 y_i^q , 因此样本 j 为正样本;而 k 与查询样本具有不同的标签 y_k^s , 因此样本 k 为负样本。

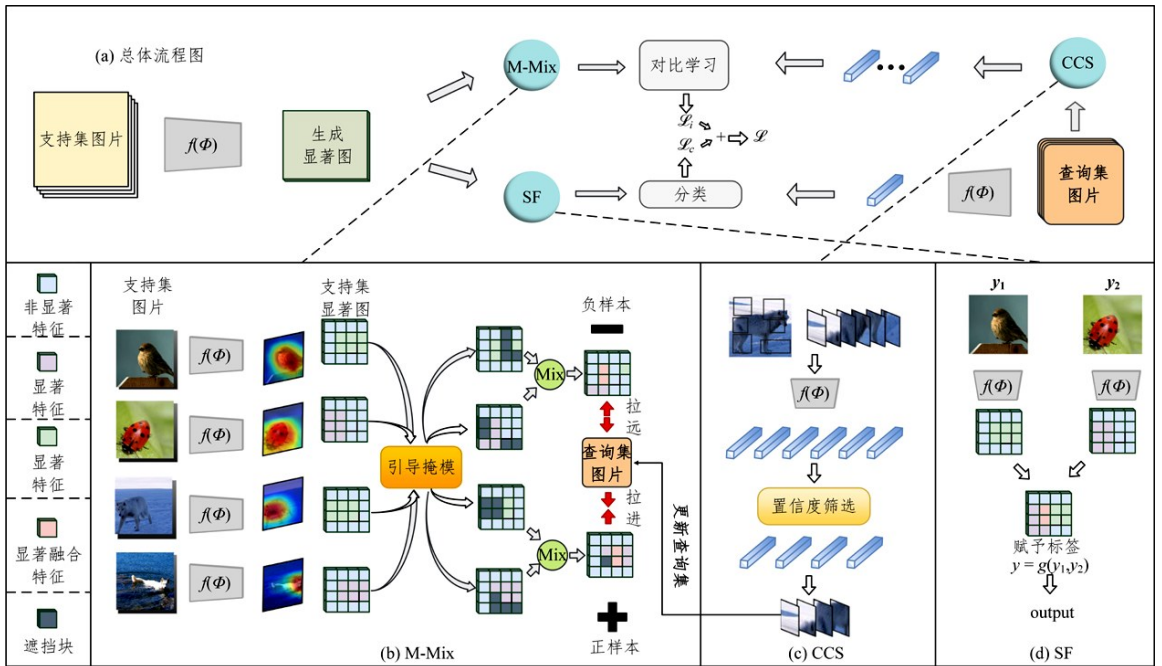


图 2 本文方法总体框架图(电子版为彩图)

Fig. 2 Overall framework of the proposed method

4.2 视觉特征隐蔽融合模块

视觉特征隐蔽融合模块通过将两个图像融合为一个样本来优化对比学习,从而在小样本分类任务中构建信息更加丰富的正负样本。图像被划分为显著块和非显著块,随机遮挡一部分块后进行融合,显著区域优先保留。通过这种方法生成对比样本后,模型能够更有效地抓取关键信息。最终,查询和支持之间的内积公式和 infoNCE 损失用于训练模型,增强其鲁棒性和泛化能力。

对比学习是一种常见的无监督学习方法,通常通过构建正样本和负样本对来训练模型。常规方法是将与自身视图相同的图片作为正样本,不同的图片作为负样本。该方法在拥有大量训练样本时效果很好,但在小样本分类任务中可能会

导致模型不够稳定。

为了解决这一问题,本文引入标签来优化对比学习。将两个图片融合成一个新的样本,在支持集中随机选择两个与当前查询图片标签相同的图像作为正样本,同时选择标签不同的图像作为负样本。对于选择的两个负样本,它们之间的标签可以不同。该方法避免了将相同标签的图片视为负样本,从而更有效地扩大了样本之间的差异。简而言之,本文通过将两个图片融合成一个新的样本,并选择标签不同的图片作为负样本,创建了更具信息量的正负样本对,从而提高了模型的鲁棒性。

融合两个图片的过程如图 2(b)所示。首先,对图像进行半随机掩模处理。用网格将每个图像 I 划分为 N 个大小均

等的块,图像 I 可以表示为块的集合 $\{B_1, B_2, \dots, B_N\}$, 并通过 $S = f_s(I)$ 计算每个块的显著程度。其中, $f_s(I)$ 是显著性检测算法。对于每个块 B_i , 计算显著程度 E_i 。显著程度可以定义为块内显著图 S 的显著值的平均值:

$$E_i = \frac{1}{|B_i|} \sum_{i=1}^N S(B_i) \quad (2)$$

其中, $|B_i|$ 表示块中 B_i 像素的数量。显著度最高的前 N_{fore} 个块被指定为显著块,并在图中用紫色和绿色表示,非显著区域则用蓝色表示。在这些指定的 N 块中,随机遮挡 M_{fore} 块,并且在非显著区域内随机遮挡 M_{back} 块,这些遮挡块在图中用黑色表示。 M_{fore} 和 M_{back} 是随机选择的,满足约束条件 $M_{fore} + M_{back} = M_{total}$, 其中 M_{total} 是一个小于 N 的固定值。

随后,将这两个正样本和两个负样本分别融合。在分完块后,融合时将每个块一一对应。本文旨在最大化显示样本的显著区域,如果两个块都为显著或不显著区域,则进行 Mixup 混合。在图 2(b) 中,粉色表示两个显著块融合后的块。如果一个区域是显著的,而另一个区域是不显著的,则显著区域直接覆盖不显著的区域。如果一个区域已经被遮挡,它将直接被另一区域覆盖。融合过程表示为: $Mix(x_i, y_i) = \lambda \cdot x_i + (1-\lambda) \cdot x_j$, 其中 λ 是从 Beta 分布中采样的随机权重, x_i 和 y_i 分别表示输入图片和标签。设 x_1 和 x_2 为两个输入图像,表示为 $\hat{x}_1 = x_1^i + x_1^m$ 和 $\hat{x}_2 = x_2^i + x_2^m$, 其中 x_1^i 和 x_2^i 表示不显著区域, x_1^m 和 x_2^m 表示显著的区域。掩模可以表示为 D 。融合式可以表示为:

$$M(x_1, x_2) = \begin{cases} D, & \text{if } (x_1^m \cap M \neq \emptyset) \text{ and } (x_2^m \cap M \neq \emptyset) \\ x_1, & \text{if } (x_1^m \cap x_2^i \neq \emptyset) \\ x_2, & \text{if } (x_1^i \cap x_2^m \neq \emptyset) \\ Mix(x_1, x_2), & \text{otherwise} \end{cases} \quad (3)$$

该操作生成的对比样本突出了差异,使模型能够优先考虑关键的图像区域,增强了模型抓取关键信息的能力。查询和支持之间的内积表达如下:

$$F = f_j^q \times M[f_j(s_1), f_j(s_2)] \quad (4)$$

其中, s_1 和 s_2 分别表示本文挑选出的两个支持样本中的两个图片。用式(3)融合后,与查询集计算内积。融合后的 infoNCE 可以写成:

$$L_i = -\log \frac{\sum_{y_j^i = y_j^q} e^F}{\sum_{y_j^i = y_j^q} e^F + \sum_{y_j^i \neq y_j^q} e^F} \quad (5)$$

4.3 置信度裁剪选择器

为了解决查询样本数量有限的问题,采用随机剪裁技术进行数据扩增,并提出置信度裁剪选择器方法。通过生成多张裁剪图像并提取其特征向量,对这些图像进行置信度排序,移除最高和最低置信度的图像,保留中间部分的图像进行对比学习,从而提高模型的鲁棒性和泛化性。

鉴于查询样本数量有限,本文采用随机剪裁技术来扩增数据。查询样本的质量和多样性直接影响着模型的鲁棒性和泛化性,利用具有代表性的查询样本有助于模型适应不同的环境和背景变化,但并非所有来自查询的裁剪图像都能达到最佳效果。某些样本或许具备信息价值,但其他一些则可能不能提供充足的信息。因此,挑选出高质量的样本至关重要。

为了更有效地利用查询数据,本文提出置信度裁剪选择器方法用于处理查询图像,如图 2(c) 所示。利用原始图像生成大约 6 张新图像,并对它们进行随机裁剪。所有裁剪后的图像都经过相同的特征提取器处理,以生成相应的特征向量:

$$\arg \min_{\gamma \in \mathbb{R}^{n \times N}} \|\tilde{X}_{cropped} - \tilde{X}_{cropped} \gamma\|_F^2 + \lambda R(\gamma) \quad (6)$$

其中, $\tilde{X}_{cropped}$ 表示图像特征的矩阵, γ 表示用于组合这些特征的线性模型的参数, $R(\gamma)$ 表示用于控制模型复杂度的正则化项。该公式的目标是通过调整参数 γ 来最小化特征矩阵 $\tilde{X}_{cropped}$ 和预测值 $\tilde{X}_{cropped} \gamma$ 之间的差异,使用平方 Frobenius 范数来衡量。最终目标是根据特定标准对图像进行排序。

随着 λ 的变化, γ 逐渐稀疏化,直到所有元素消失。本文确定了 γ 消失的 λ 值后,根据消失点对伪标记数据进行置信度排序。人类可以通过物体的一部分来识别对象,模型如果获得这种能力,则可以学习到更好的嵌入特征。为了实现半遮挡的场景,本文移除了具有最高和最低置信度的图像,并将剩余的图像用作单独的查询进行比较学习。

4.4 视觉特征显著性融合模块

视觉特征显著性融合模块采用无需掩模的数据增强策略。本文通过将图像划分为多个块,并对其中最显著的块进行标记,从而优化显著区域的融合。融合过程如图 2(d) 所示,通过不同的混合技术凸显图像的显著区域,创造出更具特色和可辨识性的图像,提升模型对关键特征的识别和理解能力。此外,本文提出了一种混合标签的生成公式,根据混合权重和混合补丁的数量生成最终标签,从而在模型训练中提供多样的视角。

SF 模块的数据增强策略不再需要 M-Mix 方法中的掩模需求。因此,在获取显著图后,将图像划分为 N_{total} 个块,并指定其中显著的 N_{fore} 块为标记块,用绿色和紫色表示。融合计算式为:

$$S(x_1, x_2) = \begin{cases} x_1, & \text{if } (x_1^m \cap x_2^i \neq \emptyset) \\ x_2, & \text{if } (x_1^i \cap x_2^m \neq \emptyset) \\ Mix(x_1, x_2), & \text{otherwise} \end{cases} \quad (7)$$

该方法凸显了图像中显著区域的重要性,如图 2(d) 所示。不同的混合技术导致图像呈现出独特的视觉特征,这种混合潜力有助于创建更具特色和可辨识性的图像,并为模型训练提供了多样的视角,从而有助于敏锐地识别和理解关键特征。混合标签的计算式如下:

$$g(y_0, y_1) = \frac{W \times W}{c(\text{minter})} (\lambda y_0 + (1-\lambda) y_1) + \frac{W \times W}{c(\neg \text{minter} \cdot m_0)} y_0 + \frac{W \times W}{c(\neg \text{minter} \cdot m_1)} y_1 \quad (8)$$

其中, $g(y_0, y_1)$ 表示混合后的标签, y_0 和 y_1 分别对应两个原始样本的标签; $W \times W$ 表示图像的尺寸或图像中的补丁数; $c(\text{minter})$ 表示激活元素的数量,即二进制掩模 m 中激活的元素数值为 1; 参数 λ 是混合权重,用于确定两个标签之间混合的程度; $c(\neg \text{minter} \cdot m_0)$ 和 $c(\neg \text{minter} \cdot m_1)$ 分别表示两个样本中非激活补丁的数量。该计算式旨在基于混合权重和混合补丁的数量生成最终的混合标签。

4.5 算法描述

算法 1 给出了显著性掩模混合算法的伪代码,该算法在

训练和测试数据集上运行,为支持图像生成显著性图。对于每个支持-查询对,对查询图像进行随机裁剪以创建新实例,然后生成特征向量并使用回归技术进行置信度评估。最后,算法针对训练集和测试集中的每一轮迭代计算 infoNEC 损失 L_i 。

算法 1 显著性掩模混合算法

输入: $D_{train}, D_{val}, D_{test}$

输出: 损失 L_i

1. FOR ALL(s, q) in (D_{train}, D_{test}) DO
2. 生成 s 的显著图
3. 通过式(3)计算 $M(x_1, x_2)$
4. 随机剪裁查询图片 q
5. 为每个裁剪的图像生成特征向量 Γ
6. 置信度筛选
7. 通过式(5)计算损失 L_i
8. END FOR

5 实验

5.1 数据集

为验证本文提出的方法,在几个广泛使用的数据集上进行了实验。miniImageNet 是来自 ImageNet 的子数据集,共有 100 个类别,每个类别有 600 个实例,这些类别被分成训练集、验证集和测试集,分别包含 64, 16 和 20 个类别。tiered-ImageNet 也是从 ImageNet 中抽样而来,它由来自 608 个类别的 779165 张图像组成,它们被分为训练集、验证集和测试集,分别包含 351, 97 和 160 个类别。Fewshot-CIFAR100 (FC-100) 数据是 CIFAR-100 的一个子集。其常见的划分是 60 个类别用于训练,20 个类别用于验证和测试。CUB 是一个细粒度分类数据集,涵盖了 200 种鸟类,用于基础类别的有 100 个,用于评估的有 50 个,剩余的 50 个用于新颖类别。

5.2 实施细节

为了与先前的工作进行公平对比,选用 ResNet12 作为本研究模型的基础架构,并严格按照 Meta-Baseline^[1] 的规范实现。模型参数的初始化采用 He-normal 方法。在优化方面,采用随机梯度下降 (Stochastic Gradient Descent, SGD),将初始学习率设定为 0.1。在 miniImageNet 数据集上进行实验时,分别在第 12000, 14000 和 16000 个 episode 时进行学习率的迭代调整。同时,在 tieredImageNet 数据集上的实验中,学习率在每 24000 个 episode 时减半。模型评估涵盖了所有实验,共计 2000 个 episode。在训练阶段,每个批次包括 4 个作为训练样本的 episode。

5.3 与其他方法对比

为了评估本文模型的性能,将其与几种先前的方法进行比较,包括 ProtoNet^[4], infoPatch^[8], LSFSL^[2], MML^[6], TALDS^[3] 和 REINet 等。这些方法或为经典的小样本学习上的正确率方法,或已经展示了最佳的结果。

表 1 列出了各方法在 miniImageNet 和 tieredImageNet 这两个数据集上的对比结果,对比方法所采用的 backbone 都是 Resnet-12。与 DeepEMD^[5] 相比,本文方法在计算显著图时计算量较小,与 REINet 和 CVET 相比也不需要额外的参数量,这也能体现出本文的方法是优于其他方法的。表 2 列出

了在 CUB-200 数据集上的对比结果,即使一些方法的骨干网络参数量大于本文方法,但本文方法效果依然优于其他方法。图 3 展示了本文方法与其他方法在 FC-100 数据集上的对比结果,红色柱状表示最佳结果,蓝色柱状表示次佳结果。从图中可以看出,本文提出的方法在 1-shot 和 5-shot 任务中的表现均优于其他对比方法,达到了最高的准确率,证明了本文方法在小样本学习任务中的有效性。

表 1 本文方法在 miniImageNet 和 tieredImageNet 数据集上与其他方法的准确率对比

方法	miniImageNet		tieredImageNet	
	1-shot	5-shot	1-shot	5-shot
infoPatch ^[8]	67.67 \pm 0.45	82.44 \pm 0.31	71.51 \pm 0.52	85.44 \pm 0.35
AMCL ^[14]	68.92 \pm 0.44	84.45 \pm 0.29	72.41 \pm 0.51	87.36 \pm 0.31
HPE ^[15]	64.12 \pm 0.33	79.82 \pm 0.29	70.15 \pm 0.33	84.54 \pm 0.26
REINet ^[16]	67.60 \pm 0.44	82.58 \pm 0.30	71.61 \pm 0.51	85.28 \pm 0.35
MixtFSL ^[17]	63.98 \pm 0.79	82.04 \pm 0.49	70.97 \pm 1.03	86.16 \pm 0.67
CVET ^[18]	70.19 \pm 0.46	84.66 \pm 0.29	72.62 \pm 0.51	86.62 \pm 0.33
LSFSL ^[2]	64.67 \pm 0.49	81.79 \pm 0.18	71.17 \pm 0.52	86.23 \pm 0.22
TALDS ^[3]	67.89 \pm 0.20	84.31 \pm 0.44	71.34 \pm 0.32	86.12 \pm 0.33
DeepEMD ^[5]	65.91 \pm 0.82	82.41 \pm 0.56	71.16 \pm 0.22	86.03 \pm 0.58
MML ^[6]	67.58 \pm 0.23	81.41 \pm 0.20	71.38 \pm 0.25	84.65 \pm 0.20
Ours	70.43\pm0.46	85.41\pm0.30	73.14\pm0.57	87.54\pm0.49

注:加粗表示精度最佳。

表 2 本文方法与在 CUB 数据集上与其他方法的准确率对比

Table 2 Accuracy of method in this paper is compared with other methods on CUB dataset

方法	Backbone	CUB	
		1-shot/%	5-shot/%
ProtoNet ^[4]	ResNet-18	71.88 \pm 0.9	86.64 \pm 0.5
AMCL ^[14]	ResNet-12	81.12 \pm 0.42	91.43 \pm 0.25
S2M2R ^[9]	ResNet-34	72.92 \pm 0.83	86.55 \pm 0.67
REINet ^[16]	ResNet-12	79.49 \pm 0.44	91.11 \pm 0.24
MixtFSL ^[17]	ResNet-18	73.94 \pm 1.1	86.01 \pm 0.5
DPGN ^[19]	ResNet-12	75.71 \pm 0.47	91.48 \pm 0.33
LRD ^[20]	ResNet-12	79.56 \pm 0.87	90.67 \pm 0.35
AA ^[21]	ResNet-18	74.22 \pm 1.09	88.65 \pm 0.55
Laplacian ^[22]	ResNet-18	80.96	88.68
setfeat ^[23]	ResNet-12	79.60 \pm 0.80	90.48 \pm 0.44
Ours	ResNet-12	81.33\pm0.59	91.73\pm0.37

注:加粗表示精度最佳。

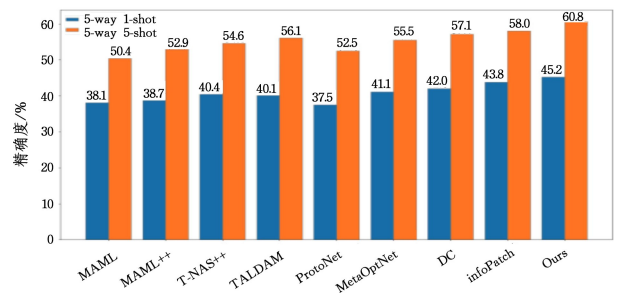


图 3 本文与其他方法在 FC-100 数据集上的对比结果 (电子版为彩图)

Fig. 3 Comparison results of the proposed method with other methods on FC-100 dataset

1) 数据增强方法对比。为了进一步验证本文提出方法的有效性,在 miniImageNet 数据集上进行了实验。除了详细的数据增强方法外,其他设置均相同。将 Augmix 和 Cutmix 添

加到本文的基线方法中。同时,利用 manifold mixup 和 IDeMeNet 的增强方法。表 3 列出了实验结果。可以看出,本文方法取得了最佳的结果。

表 3 本文方法与其他数据增强方法在 miniImageNet 数据集上的对比

Table 3 Comparison of this paper's method with other data enhancement methods on miniImageNet dataset

数据增强方法	5-way 1-shot/%	5-way 5-shot/%
mixup ^[11]	67.34	81.99
augmix ^[25]	67.95	82.07
cutmix ^[12]	66.34	82.61
IDEeMeNet ^[26]	66.59	82.48
M-mixup ^[27]	67.87	81.98
PatchMix ^[8]	68.74	83.04
Ours	70.43	85.41

2)t-SNE图。本文利用 t-SNE 图来可视化嵌入向量。具体来说,从 miniImageNet 数据集中的目标类别中构建了一个更大的 episode,并将其输入到基准模型和本文的完整模型中。图 4 给出了嵌入向量的可视化结果。从图 4(a)中可以看出,基准模型生成的图像缺乏紧凑性,导致许多类别混合在一起;而从图 4(b)中可以看出,本文方法产生的图像更加紧凑,使得每个类别的图像能够更有效地聚类。

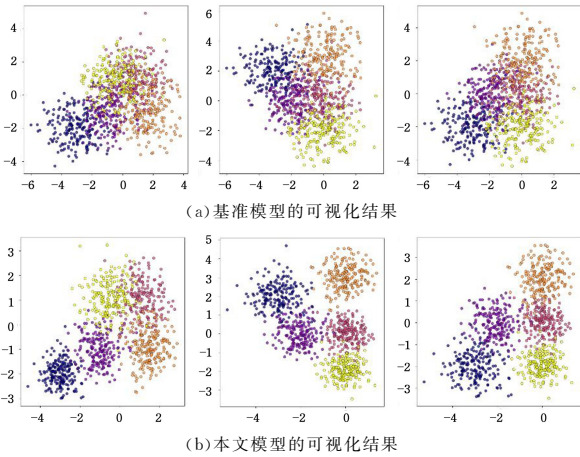


图 4 本文模型与基准模型的 t-SNE 图可视化对比

Fig. 4 Comparison of t-SNE visualizations between the proposed model and the baseline model

3)对比学习损失对比。为了验证本文改良后的对比学习方法是否有效,选择 MoCo^[24]方法进行了对比。将 MoCo 方法应用于本文的基线上,在训练阶段记录了 MoCo 方法的损失和本文的损失进行对比,如图 5 所示。

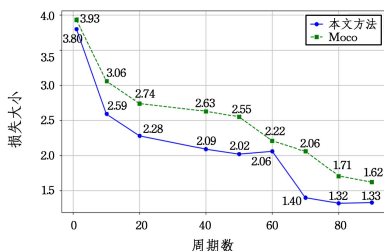


图 5 本文方法在 miniImageNet 上的损失与无监督对比学习损失的对比

Fig. 5 Comparison of the loss of the proposed method on miniImageNet with the loss of unsupervised contrastive learning

从图中可以看出,本文方法的损失收敛速度比 MoCo 方法更快,可以证明本文的对比学习方法在小样本分类中优于 MoCo 方法。

4)热力图。通过可视化空间对应关系来验证本文模型提取特征的能力。利用从支持图像中提取的特征,计算查询图像每个部分的内积。图 6 为本文模型与基准模型生成的热力图,可以看出本文模型更加准确地覆盖了前景,并在识别含噪图像中的对象方面表现出色。这说明了本文方法在空间关系理解方面优于基准模型,有效性得到了验证。

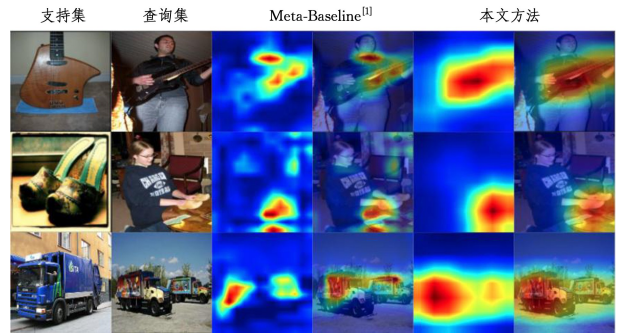


图 6 热力图可视化对比

Fig. 6 Heatmap visualization comparison

5.4 消融实验

1)模块消融。本文将 M-Mix、CCS、SF 与基准模型^[1]进行了消融研究。结果如表 4 所列,每个组件都对性能改善有所贡献。具体来说,本文在 miniImageNet 数据集上进行了这项分析,并且观察到了每个组件的重要性。虽然单独添加 CCS 并没有明显增强模型性能,但将其与 M-Mix 结合使用证明了 CCS 是有效的。采用本文提出的 SF 也显著提高了模型的泛化能力。

表 4 模型中不同模块的消融实验结果

Table 4 Results of ablation experiments for different modules of model

M-Mix	CCS	SF	1-shot/%	5-shot/%
			63.17	78.31
✓			67.33	81.93
	✓		64.07	79.39
		✓	65.21	81.44
✓	✓		69.72	83.47
✓	✓	✓	70.43	85.41

2)网格大小消融。在构建用于对比学习的支持图像时,需要使用网格来为图片分块。本文在 miniImageNet 数据集上进行了分析。为了确保实验的一致性,将图像分成方块,使用 8×8 、 10×10 和 12×12 大小的网格,将图像分别平均分成了 64、100 和 144 块。在 8×8 的网格下,给 10、16 和 20 个方块打上掩模;在 10×10 的网格下,给 20、25 和 30 个方块打上掩模;对于 12×12 的网格,使用 30、40 和 50 个掩模的方块。本文利用这些合成的图像进行对比学习,丰富了对比学习样本。结果如表 5 所列,可以看出,使用中等大小的网格可以获得更好的结果。

表5 网格分割大小的消融实验

Table 5 Ablation experiments with grid segmentation size

网格分割大小	掩模数量	精确度/%
8×8	10	69.71
	16	68.97
	20	68.58
10×10	20	68.75
	25	70.43
	30	69.84
12×12	30	69.14
	40	68.85
	50	68.43

3) 随机裁剪数量消融。在 CCS 模块中,对随机裁剪的数量进行了消融实验,分别选择 4,6,8,12 和 16 的数量,其他设置不变,在 miniImageNet 上进行了实验。实验结果如图 7 所示。当随机裁剪的数量为 6 时,效果最好,因为数量太少会导致查询集样本数量不够,数量太大虽然会让数据量增大,但容易导致过拟合。因此本文选择随机裁剪数量为 6,使实验效果达到最佳。

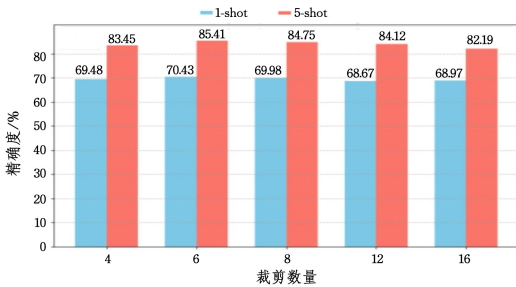


图7 CCS方法中不同裁剪数量在 miniImageNet 数据集上的对比实验

Fig. 7 Comparative experiments with different cropping quantities on miniImageNet dataset in CCS method

为了更直观地展示本文提出的方法在图像处理中的实际效果,提供了 M-Mix 和 SF 两种方法的视觉示例,如图 8 所示。M-Mix 方法通过将部分图像进行遮挡处理,增强了数据多样性,并提高了模型的鲁棒性;SF 方法通过对图像特征进行选择融合,提高了模型在复杂背景下的鲁棒性。



图8 M-Mix 和 SF 方法的视觉示例

Fig. 8 Visual examples of M-Mix and SF methods

5.5 缺陷分析

本研究存在一定局限性。首先,尽管使用的数据集相当全面,但可能缺乏真实世界变异性的某些方面,这可能会限制研究结果的泛化能力。此外,M-Mix 生成的合成图像可能与真实数据存在一定的失真或偏差,导致其在真实场景中可能会出现过拟合。在一些真实场景或医学领域^[28-31]中,本文方

法可能不适用。M-Mix 方法因融合数据而在一些分割任务^[32-35]中并不适用,无法拓展。

结束语 本文提出了一种基于显著性掩模混合的数据增强方法。M-Mix 通过引导性的遮挡保留图像的关键部分,并将这些图片融合,生成具有代表性的样本用于对比学习。该增强方法在对比学习中的应用,旨在自适应筛选与保留图像的关键特征信息,尽量避免归纳偏差,增强模型的泛化能力。CCS 通过为裁剪后查询集图片赋予置信度,筛选出一批高质量图片,为训练提供更多、更优的数据。SF 方法通过计算图片中各个区域的重要性,使生成的图片更具多样性和丰富性,令类别边界更加清晰,从而增强模型的鲁棒性。在多个标准小样本图像分类数据集上,该方法表现优异,优于当前最先进方法约 0.2%~1%。结果表明,该方法在小样本图像分类领域具有显著的潜力和优势。

参考文献

- [1] CHEN Y, LIU Z, XU H, et al. Meta-Baseline: Exploring simple meta-learning for few-shot learning[C]// 2021 IEEE/CVF International Conference on Computer Vision (ICCV). 2021:9042-9051.
- [2] PADMANABHAN D, GOWDA S, ARANI E, et al. Ls-fsl: Leveraging shape information in few-shot learning [C] // 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). 2023:4971-4980.
- [3] QIAO Q, XIE Y, ZENG Z Y, et al. Talds-net: Task aware adaptive local descriptors selection for few-shot image classification [J]. arXiv:2312.05449, 2023.
- [4] SNELL J, SWERSKY K, ZEMEL R. Prototypical networks for few-shot learning[J]. arXiv:1703.05275, 2017.
- [5] ZHANG C, CAI Y, LIN G, et al. DeepEMD: Differentiable earth mover's distance for few-shot learning[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 45(5): 5632-5648.
- [6] CHEN H, LI H, LI Y, et al. Multi-level metric learning for few-shot image recognition[C]// International Conference on Artificial Neural Networks. Cham: Springer International Publishing, 2022:243-254.
- [7] DENG G L, HUANG G H, CHEN Z Y. Category Decoupled Few-Shot Classification for Graph Neural Network[J]. Computer Engineering and Applications, 2024, 60(2): 129-136.
- [8] LIU C, FU Y, XU C, et al. Learning a few-shot embedding model with contrastive learning[C]// Proceedings of the AAAI Conference On Artificial Intelligence. 2021:8635-8643.
- [9] MANGLA P, KUMARI N, SINHA A, et al. Charting the right manifold: Manifold mixup for few-shot learning [C] // Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2020:2218-2227.
- [10] ZHUO L, FU Y, CHEN J, et al. Tgdm: Target guided dynamic mixup for cross-domain few-shot learning[C]// Proceedings of the 30th ACM International Conference on Multimedia. 2022: 6368-6376.
- [11] ZHANG H Y, CISSE M, DAUPHIN Y N, et al. mixup: Beyond

- empirical risk minimization[J]. arXiv:1710.09412,2017.
- [12] YUN S, HAN D, OH S J, et al. Cutmix: Regular-ization strategy to train strong classifiers with localizable features[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019:6023-6032.
- [13] KIM J H, CHOO W, SONG H O. Puzzle mix: Exploiting saliency and local statistics for optimal mixup[C]// International Conference on Machine Learning. PMLR, 2020:5275-5285.
- [14] PENG T, FENG L, DU Y D, et al. Meta-cosine loss for few-shot image classification[J]. Journal of Image and Graphics, 2024, 29(2):506-519.
- [15] GUO L, LIU B, LI W G, et al. A Few-Shot Image Classification Method by Hard Pairwise-Based Excitation[J]. Journal of Computer-Aided Design & Computer Graphics, 2024, 26(6):895-903.
- [16] KANG D, KWON H, MIN J, et al. Relational embedding for few-shot classification[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021:8822-8833.
- [17] AFRASIYABI A, LALONDE J F, GAGNÉ C. Mixture-based feature space learning for few-shot image classification[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021:9041-9051.
- [18] YANG Z Y, WANG J H, ZHU Y Y. Few-shot classification with contrastive learning[C]// European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2022:293-309.
- [19] YANG L, LI L, ZHANG Z, et al. Dpgn: Distribution propagation graph network for few-shot learning[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020:13390-13399.
- [20] YANG S, LIU L, XU M. Free lunch for few-shot learning: Distribution calibration[J]. arXiv:2101.06395, 2021.
- [21] AFRASIYABI A, LALONDE J F, GAGNÉ C. Associative alignment for few-shot image classification[C]// Computer Vision-ECCV 2020: 16th European Conference, Glasgow, UK, August 23-28, 2020, Proceedings, Part V 16. Springer International Publishing, 2020:18-35.
- [22] ZIKO I, DOLZ J, GRANGER E, et al. Laplacian regularized few-shot learning[C]// International Conference on Machine Learning. PMLR, 2020:11660-11670.
- [23] AFRASIYABI A, LAROCHELLE H, LALONDE J F, et al. Matching feature sets for few-shot image classification[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022:9014-9024.
- [24] HE K M, FAN H Q, WU Y X, et al. Momentum contrast for unsupervised visual representation learning[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020:9729-9738.
- [25] HENDRYCKS D, MU N, CUBUK E D, et al. Augmix: A simple data processing method to improve robustness and uncertainty [J]. arXiv:1912.02781, 2019.
- [26] CHEN Z, FU Y, WANG Y X, et al. Image deformation meta-networks for one-shot learning[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019:8680-8689.
- [27] VERMA V, LAMB A, BECKHAM C, et al. Manifold mixup: Better representations by interpolating hidden states[C]// International Conference on Machine Learning. PMLR, 2019:6438-6447.
- [28] YANG T H, GU Z H, MA L Z. Style-aware Cross Domain Few-Shot Anomaly Detection[J/OL]. <http://kns.cnki.net/kcms/detail/11.2925.TP.20240926.1526.003.html>.
- [29] CHEN Y D, ZHAO Y B, WU E H, et al. Robust Semi-supervised Video Object Segmentation with Dynamic Embedding[J]. <https://link.cnki.net/doi/10.13700/j.bh.1001-5965.2023.0354>
- [30] LIU H, TIAN Z, QIU J, et al. Survey on Few-shot for Malware Detection [J]. Journal of Software, 2024, 35(8):3785-3808.
- [31] LI F, JIA D L, YAO Y M, et al. Graph Neural Network Few Shot Image Classification Network Based on Residual and Self-attention Mechanism [J]. Computer Science, 2023, 50(S1):276-380.
- [32] CHEN Y D, HAO C Y, YANG Z X, et al. Fast Target-aware Learning for Few-shot Video Object Segmentation [J]. SCIENCE CHINA Information Sciences, 2022, 65(8):182104.
- [33] CHEN Y D, JIANG R, ZHENG Y H, et al. Dual branch multi-level semantic learning for few-shot segmentation [J]. IEEE Transactions on Image Processing, 2024, 33:1432-1447.
- [34] CHEN Y, CHEN S, YANG Z X, et al. Learning self-target knowledge for few-shot segmentation[J]. Pattern Recognition, 2024, 149:110266.
- [35] CHEN Y D, CHEN L R, YU W B, et al. Knowledge Distillation Anomaly Detection with Multi-Scale Feature Fusion[J]. Journal of Computer-Aided Design & Computer Graphics, 2022, 34(10):1542-1549.



CHEN Yadang, born in 1985, Ph.D. associate professor. His main research interests include video segmentation, video enhancement, video editing and augmented reality.



CHE Xun, born in 1985, doctoral student, professor, is a member of CCF (No. J8696S). His main research interest is model robustness and security.