

# 光电工程

## Opto-Electronic Engineering

中文核心期刊 中国科技核心期刊  
Scopus CSCD

### 多尺度注意力与领域自适应的小样本图像识别

陈龙, 张建林, 彭昊, 李美惠, 徐智勇, 魏宇星

#### 引用本文:

陈龙, 张建林, 彭昊, 等. 多尺度注意力与领域自适应的小样本图像识别[J]. *光电工程*, 2023, **50**(4): 220232.

Chen L, Zhang J L, Peng H, et al. Few-shot image classification via multi-scale attention and domain adaptation[J]. *Opto-Electron Eng*, 2023, **50**(4): 220232.

<https://doi.org/10.12086/oe.2023.220232>

收稿日期: 2022-09-22; 修改日期: 2022-12-26; 录用日期: 2022-12-29

### 相关论文

#### 在线推断校准的小样本目标检测

彭昊, 王婉祺, 陈龙, 彭先蓉, 张建林, 徐智勇, 魏宇星, 李美惠  
*光电工程* 2023, **50**(1): 220180 doi: [10.12086/oe.2023.220180](https://doi.org/10.12086/oe.2023.220180)

#### 基于深度学习的飞机目标跟踪应用研究

赵春梅, 陈忠碧, 张建林  
*光电工程* 2019, **46**(9): 180261 doi: [10.12086/oe.2019.180261](https://doi.org/10.12086/oe.2019.180261)

#### 面向遥感图像检索的级联池化自注意力研究

吴刚, 葛芸, 储珺, 叶发茂  
*光电工程* 2022, **49**(12): 220029 doi: [10.12086/oe.2022.220029](https://doi.org/10.12086/oe.2022.220029)

#### 深度双重注意力的生成与判别联合学习的行人重识别

张晓艳, 张宝华, 吕晓琪, 谷宇, 王月明, 刘新, 任彦, 李建军  
*光电工程* 2021, **48**(5): 200388 doi: [10.12086/oe.2021.200388](https://doi.org/10.12086/oe.2021.200388)

更多相关论文见光电期刊集群网站 

 | **光电工程**  
Opto-Electronic Engineering

<http://cn.ojournal.org/oe>



 OE\_Journal



Website

DOI: 10.12086/oe.2023.220232

## 多尺度注意力与领域自适应的小样本图像识别

陈龙<sup>1,2</sup>, 张建林<sup>1\*</sup>, 彭昊<sup>1,2</sup>,  
李美惠<sup>1</sup>, 徐智勇<sup>1</sup>, 魏宇星<sup>1</sup>

<sup>1</sup>中国科学院光电技术研究所, 四川 成都 610209;

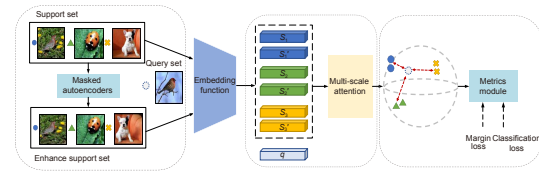
<sup>2</sup>中国科学院大学电子电气与通信工程学院, 北京 100049

**摘要:** 小样本图像识别是计算机视觉中的关键问题。针对小样本情况下度量学习方法的类别原型不准确、泛化能力差问题, 本文采用以下措施来提高小样本图像识别准确率: 第一、为减缓样本稀缺问题, 利用掩膜自编码器进行图像扩充, 提高样本复杂度。第二、设计多尺度注意力模块, 突出类别相关特征, 解决类别原型偏差大的问题。第三、提出领域自适应模块, 引入间隔损失函数, 优化嵌入函数的表征能力, 实现新类样本的精确表征, 增强模型的泛化能力。通过在多个公开数据集上进行实验表明, 本文方法可以有效提升小样本图像识别准确率。

**关键词:** 小样本图像识别; 注意力机制; 领域自适应; 相似性度量

**中图分类号:** TP391.4

**文献标志码:** A



陈龙, 张建林, 彭昊, 等. 多尺度注意力与领域自适应的小样本图像识别 [J]. 光电工程, 2023, 50(4): 220232

Chen L, Zhang J L, Peng H, et al. Few-shot image classification via multi-scale attention and domain adaptation[J]. *Opto-Electron Eng*, 2023, 50(4): 220232

## Few-shot image classification via multi-scale attention and domain adaptation

Chen Long<sup>1,2</sup>, Zhang Jianlin<sup>1\*</sup>, Peng Hao<sup>1,2</sup>, Li Meihui<sup>1</sup>, Xu Zhiyong<sup>1</sup>, Wei Yuxing<sup>1</sup>

<sup>1</sup>Institute of Optics and Electronics, Chinese Academy of Science, Chengdu, Sichuan 610209, China;

<sup>2</sup>School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Science, Beijing 100049, China

**Abstract:** Learning with limited data is a challenging field for computer visual recognition. Prototypes calculated by the metric learning method are inaccurate when samples are limited. In addition, the generalization ability of the model is poor. To improve the performance of few-shot image classification, the following measures are adopted. Firstly, to tackle the problem of limited samples, the masked autoencoder is used to enhance data. Secondly, prototypes are calculated by task-specific features, which are obtained by the multi-scale attention mechanism. The attention mechanism makes prototypes more accurate. Thirdly, the domain adaptation module is added with a margin loss function. The margin loss pushes different prototypes away from each other in the feature space. Sufficient margin space improves the generalization performance of the method. The experimental results show the proposed method achieves better performance on few-shot classification.

**Keywords:** few-shot image classification; attention mechanism; domain adaptation; similarity metric

收稿日期: 2022-09-22; 修回日期: 2022-12-26; 录用日期: 2022-12-29

基金项目: 国家自然科学基金青年科学基金资助项目 (62101529)

\*通信作者: 张建林, jlin\_zh@163.com。

版权所有©2023 中国科学院光电技术研究所

## 1 引言

近年来, 深度学习在计算机视觉<sup>[1]</sup>、机器翻译<sup>[2]</sup>、语音识别等领域应用广泛, 但其非常依赖大规模标注数据集。事实上, 在很多应用场景中搜集大量带标签数据非常困难且代价昂贵, 如医疗影像、遥感图像、安防数据等<sup>[3-4]</sup>。此外, 自然采集的数据集会呈长尾分布<sup>[5]</sup>, 随着样本类别数的增加, 为保证数据集类别均衡, 数据的采集难度将会骤增。

为使深度学习算法掌握从少量样本中学习解决问题的能力, 小样本学习已成为研究热点。通过研究训练任务的特性, 利用少量训练样本进行学习, 可以减少深度学习算法对大规模数据集的依赖, 降低采集训练数据的难度, 缩减准备数据集的高昂成本。此外, 小样本学习可以研究任务之间的潜在规律, 实现模型在新任务上快速部署。目前小样本学习的研究可分为以下四类: 机器学习方法、数据增强方法、元学习方法、度量学习方法:

a) 机器学习方法采用经验风险函数加上表示模型复杂度的正则化函数来优化模型, 模型减少估计误差以获取良好的学习性能。Li 利用变分贝叶斯模型 (variational Bayesian framework, VBF) 实现样本类别推断<sup>[6]</sup>, 从少量样本中学习类别表征信息。但是在样本稀缺的环境下, 少量样本难以充分表征数据的底层分布, 导致经验风险最小化策略不稳定且失效, 模型产生过拟合现象。

b) 数据增强方法借助辅助数据集或辅助信息对原有数据集进行数据扩充或者特征增强, 是解决样本稀缺问题的直接方法。Mehrotra 使用生成对抗网络 (generative adversarial network, GAN) 合成新样本来扩充训练数据集<sup>[7]</sup>, 提高样本复杂度。文献 [8] 中, Chen 提出利用语义信息扩充特征空间, 增强样本的判别性特征。

c) 元学习方法从大量训练任务中学习元知识, 利用知识来指导模型在新任务中更好的学习。Finn 提出模型无关的元学习方法<sup>[9]</sup> (model-agnostic meta-learning, MAML), 通过多任务的学习, 微调参数使模型收敛至一个良好的初始化参数, 在面临新任务时, 模型在很短的时间内即可收敛。Rusu 等人提出潜在嵌入空间的元学习<sup>[10]</sup> (meta-learning with latent embedding optimization, LEO), 在嵌入空间采用随机采样方式更新模型参数, 提高元学习效率。Frikha 提出二分类元学习<sup>[11]</sup> (one-class classification via meta-learning, OC-

MAML), 利用元学习方法快速训练异常分类器。

d) 度量学习方法将样本映射到公共的特征空间, 利用相似性度量函数进行分类。Vinyals 等人将度量学习与长短期记忆网络结合, 提出了匹配网络<sup>[12]</sup> (matching network), 在小样本情况下实现端到端的样本类别预测。Snell 提出原型网络<sup>[13]</sup> (prototypical network), 利用特征向量均值作为类别原型, 根据最近邻原则预测样本类别。Li 等人提出局部度量网络<sup>[14]</sup> (DN4), 使用局部描述子表征样本, 实现小样本图像细粒度识别。为消除图像背景影响, Xu 提出的 COSOC 模型<sup>[15]</sup>, 采用对比学习方法识别图像前景, 提高网络泛化能力。

小样本图像识别的训练数据与测试数据的类别不存在交集且分布不匹配。领域适应 (domain adaptation) 旨在提升数据类别领域改变后网络的泛化能力, 其目标为最大化地减小领域间的分布差异。Chen 把领域自适应模块与 Faster R-CNN 目标检测框架结合<sup>[16]</sup>, 利用领域自适应模块减少图像层次与实例层次上的域差异, 增强网络的稳定性。Gong 结合 CycleGAN 和域迁移提出了适应泛化领域流 (domain flow for adaptation and generalization, DLOW)<sup>[17]</sup>, DLOW 使用流模型学习从源域到目标域的转换, 对抗损失函数鼓励模型学习领域通用表征, 一致性损失函数确保域转换过程的一致性。DLOW 学习源域数据知识, 将其迁移到目标域数据中, 生成独有风格的图像。

现有的原型网络直接采用同类样本特征向量的均值作为类别原型, 然后利用相似性度量模块对查询集样本进行类别预测。在样本稀缺的情况下, 原型存在偏差, 导致图像识别准确率下降。此外, 模型在基类数据上进行训练, 在新类数据进行测试。两类数据没有交集, 不可避免地出现领域漂移现象。针对上述问题, 本文提出多尺度注意力与领域自适应小样本图像识别 (few-shot image classification via multi-scale attention and domain adaptation, MADA), MADA 引入多尺度注意力模块, 强化分类任务相关特征, 使类别原型表征更加准确。同时, 针对领域漂移问题, MADA 提出领域自适应模块提高嵌入函数的泛化能力。本文的主要研究工作总结如下:

1) 本文提出多尺度注意力模块, 能够强化分类任务相关的特征, 优化类别原型表示, 提升模型的图像识别能力。



2) 针对领域漂移问题, 本文引入间隔损失来构建领域自适应模块。通过对基类数据进行合理表征, 为新类数据预留间隔空间, 增强模型泛化能力。

3) 将本文方法在 miniImageNet、tieredImageNet 与 CUB 数据集进行实验, 结果表明本文方法有效提升了小样本图像识别准确率。

## 2 本文方法

### 2.1 问题定义

小样本图像识别数据集由基类数据  $D_{base}$  与新类数据  $D_{novel}$  两部分构成, 其中  $D_{base}$  与  $D_{novel}$  数据分布不同且样本类别互斥。模型利用  $D_{base}$  构建分类任务进行训练, 学习图像识别知识。其次在  $D_{novel}$  构建新的分类任务, 对模型进行测试。小样本图像识别设定为 N-way K-shot 分类任务, 图 1 所示为 5-way 1-shot 任务示意图。N-way K-shot 采用 episode 策略进行训练, 每个 episode 从训练数据集中随机抽选  $N$  个类别, 且每类抽取  $K$  个样本构建支撑集  $S = \{(x_i, y_i)\}_{i=1}^{N \times K}$ , 抽取  $M$  个样本构建查询集  $Q = \{(x_i, y_i)\}_{i=1}^{N \times M}$ 。

### 2.2 模型结构

多尺度注意力与领域自适应小样本图像识别模型结构如图 2 所示, 模型由数据扩充模块、多尺度注意力模块与领域自适应模块构成, 具体细节介绍如下。

1) 数据扩充: 采用掩膜自编码器对支撑集数据进行掩膜复原, 扩充训练样本, 提高样本复杂度。

2) 多尺度注意力: 由嵌入函数与多尺度注意力构成, 提取样本特征, 强化任务相关特征。

3) 领域自适应: 均衡样本原型距离, 优化特征空间, 增强模型的泛化能力。

#### 2.2.1 数据扩充模块

现有的数据扩充方法常采用 GAN 来扩充训练数据, 该方法需要利用数据集额外训练网络, 导致模型训练耗时长、可复用性差。本文采用掩膜自编码器<sup>[18]</sup>(masked autoencoders, MAE) 对支撑集图像  $S_{train}$  进行掩膜复原生成扩充支撑集  $S'_{train}$ 。  $S'_{train}$  中的复原图像会迫使嵌入函数  $f_{\theta}$  捕获样本更全面的特征信息用于类别预测, 有效提高了嵌入函数  $f_{\theta}$  的表征能力。此外, 采用掩膜自编码器扩充数据无需训练额外模型, 能够

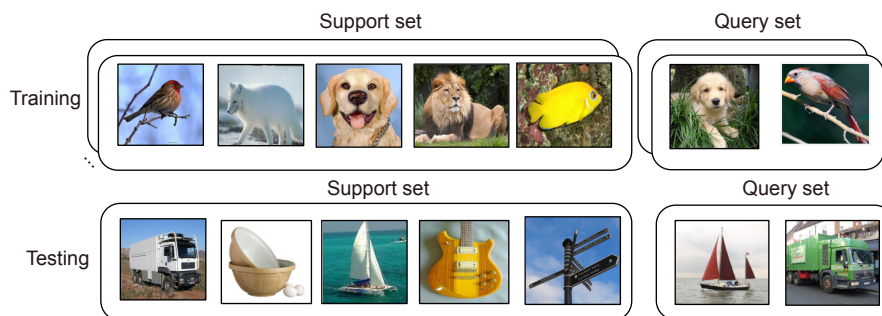


图 1 5-way 1-shot 示意图  
Fig. 1 Diagram of 5-way 1-shot

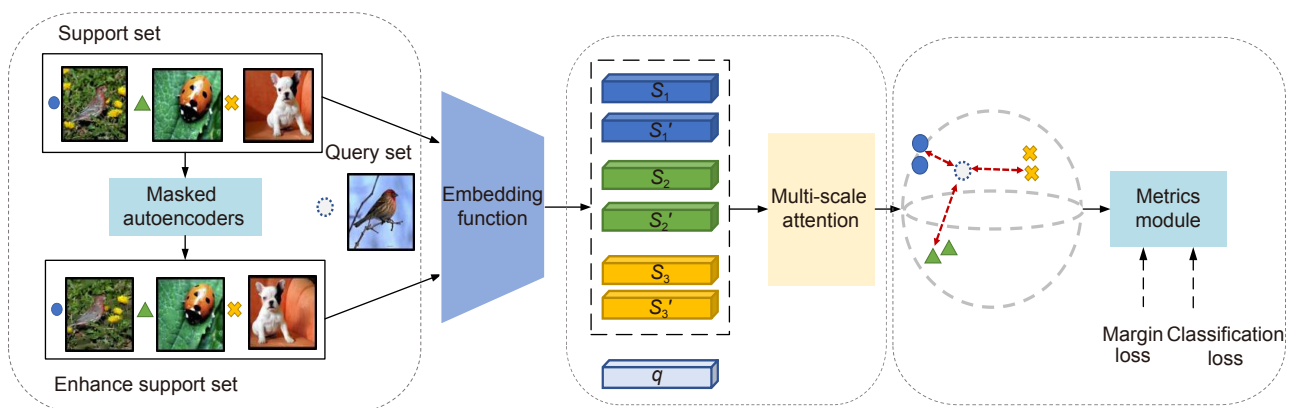


图 2 多尺度注意力与领域自适应小样本分类模型结构  
Fig. 2 The framework of few-shot image classification via multi-scale attention and domain adaptation

即插即用。本文为样本稀缺环境提供了方便快捷的数据扩充方案。

如图 3 所示, 首先利用随机掩膜模板  $M_{\text{rand}}$  对支撑集  $S_{\text{train}}$  图像随机擦除 40% 图块后可得掩膜图像  $I_M$ , 掩膜图像  $I_M$  会损失部分富有辨识度的区域。其次, 使用 MAE 对擦除区域进行像素级别重构得到复原图像  $I_E$ , 如式 (1) 所示:

$$I_E = e_\varphi(I \odot M_{\text{rand}}), \quad (1)$$

其中:  $I$  为原始图像,  $M_{\text{rand}}$  为掩膜模板,  $\odot$  为空间元素乘积,  $e_\varphi$  为掩膜自编码器映射函数。

在 N-way K-shot 中包含支撑集  $S_{\text{train}}$  和查询集  $Q_{\text{train}}$ , 采用 MAE 复原可得扩充支撑集  $S'_{\text{train}}$ , 如式 (2)~式 (4) 所示:

$$S_{\text{train}} = \{S_c\}_{c=1}^N (|S_c| = K), \quad (2)$$

$$Q_{\text{train}} = \{Q_c\}_{c=1}^N (|Q_c| = M), \quad (3)$$

$$S'_{\text{train}} = \{S'_c\}_{c=1}^N (|S'_c| = K), \quad (4)$$

其中:  $c$  为图像的类别标号,  $M$  为查询集样本数。

嵌入函数  $f_\theta$  提取  $S_{\text{train}}$ 、 $S'_{\text{train}}$  与  $Q_{\text{train}}$  中的特征向量  $s_i$ 、 $s'_i$  与  $q_i$ , 计算类别原型  $p_c$ , 如式 (5)~式 (7) 所示:

$$m_c = \frac{1}{|S_c|} \sum_{s_i \in S_c} f_\theta(s_i), c = 1, 2, \dots, N, \quad (5)$$

$$m'_c = \frac{1}{|S'_c|} \sum_{s'_i \in S'_c} f_\theta(s'_i), c = 1, 2, \dots, N, \quad (6)$$

$$p_c = \lambda m_c + (1 - \lambda) m'_c, \quad (7)$$

其中:  $m_c$ 、 $m'_c$  为原型向量,  $p_c$  为类别原型,  $\lambda$  为权重参数。

### 2.2.2 多尺度注意力模块

原型网络采用嵌入函数独立对分类样本进行特征提取, 忽略了分类任务相关特征, 弱化了图像识别的显著区域, 导致模型识别准确率降低。如图 4 所示, 本文提出多尺度注意力模块, 将全局的自注意力特征与多尺度局部感知特征相互融合, 强化分类相关特征, 优化类别原型表征。

Transformer<sup>[19]</sup> 由多层注意力机制堆叠的编码器与解码器构成, 它被广泛用于计算词向量相关性并捕捉图像感受野。本文采用 Transformer 模型的自注意力充分增强查询集样本的类别相关特征。Transformer 学习分类任务相关的空间映射, 强化图像中具有判别性的特征。Transformer 存储了查询 (query)、键 (key)、值 (value) 三元信息, 其中  $W_Q$ 、 $W_K$  与  $W_V$  是与其对应的线性映射权重, 计算特征向量的自注意力 (在公式中用  $I_{\text{Attention}}$  表示) 如式 (8)、式 (9) 所示:

$$\begin{cases} Q = W_Q^T \cdot f_\theta(x) \\ K = W_K^T \cdot f_\theta(x) \\ V = W_V^T \cdot f_\theta(x) \end{cases}, \quad (8)$$

$$I_{\text{Attention}}(f_\theta(x)) = I_s \left( \frac{QK^T}{\sqrt{d_K}} \right) V, \quad (9)$$

其中:  $f_\theta(x)$  是嵌入函数提取的特征向量,  $Q$ 、 $K$ 、 $V$  为 Transformer 的查询、键、值信息,  $W_Q$ 、 $W_K$ 、 $W_V$  是  $Q$ 、 $K$ 、 $V$  对应的线性映射权重矩阵,  $d_K$  为向量  $K$  的维度。

特征向量的维度为 640, 随着向量维度不断升高,

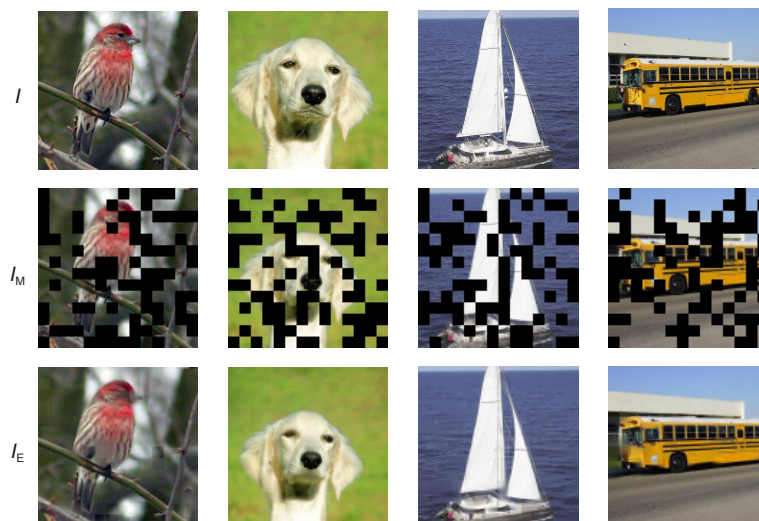


图 3 掩膜自编码器复原图像

Fig. 3 Image restoration by masked autoencoders

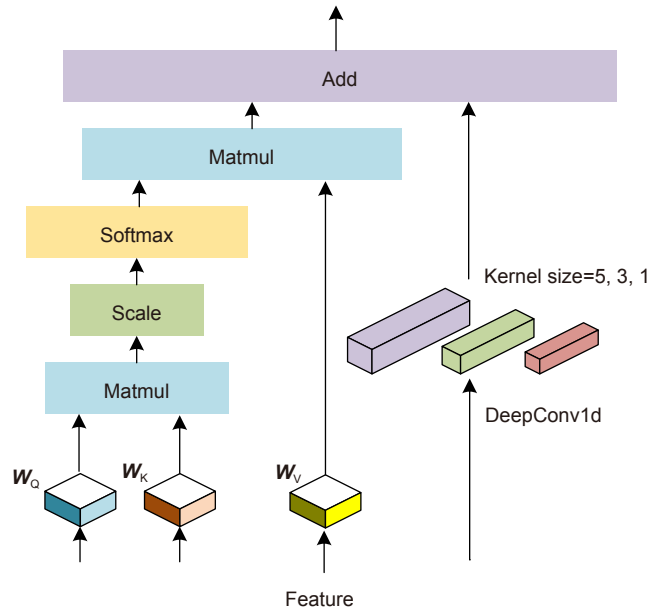


图 4 多尺度注意力模型结构  
Fig. 4 Structure of the multi-scale attention

Transformer 的全局捕捉能力会有所下降。本文提出的多尺度注意力模型将不同尺度感受野的深度卷积融入 Transformer, 缓解模型在高维向量中的性能衰减问题。计算多尺度注意力 (在公式中用  $F_{MUSA}$  表示) 特征如式 (10)、式 (11) 所示:

$$I_{Conv}(f_{\theta}(x)) = \sum_{k=1,3,5} I_{DepthConvk}(f_{\theta}(x)), \quad (10)$$

$$F_{MUSA}(f_{\theta}(x)) = I_{Attention}(f_{\theta}(x)) + I_{Conv}(f_{\theta}(x)), \quad (11)$$

其中:  $I_{Attention}(f_{\theta}(x))$  表示特征向量的注意力信息,  $I_{Conv}(f_{\theta}(x))$  表示特征向量的多尺度信息。

### 2.2.3 领域自适应模块

度量学习学习样本的类别表征和度量, 模型即可快速迁移到新的数据中。但是, 当测试数据的类别语义信息与训练数据的领域差异过大时, 模型在不可见类别图片上预测出的语义特征与图片本身的语义特征存在偏差, 导致学习了基类数据的模型无法对未见过的新类数据做出正确的预测。本文提出领域自适应模块, 优化模型对测试新类数据的表征。如图 5(a) 所示,

在模型训练过程中, 通过引入类别间隔将不同类别的样本推离, 特征向量能够均匀的分布在特征空间中, 嵌入函数对基类数据进行合理表征, 丰富样本的特征信息。如图 5(b) 所示, 在测试阶段, 模型为新类数据预留出了充足的间隔空间 (margin space), 嵌入函数可以更加准确地对新类数据进行表征, 相似性度量模块能更加准确地预测查询样本类别, 从而提高模型识别的准确率。

领域自适应模块增加异类样本在特征空间的间隔距离, 将异类样本相互推离。样本将相对均匀地分布在特征空间中, 样本特征信息更加丰富。利用类别原型距离  $D_i^{Inter}$  构建间隔损失函数  $\ell_{mrg}$ , 利用随机梯度下降算法优化嵌入函数  $f_{\theta}$  表征能力。如式 (12)~式 (14):

$$D_i^{Inter} = \sum_{j=1}^N \|p_i - p_j\|_2^2, i \neq j, \quad (12)$$

$$\ell_{mrg} = - \sum_{i=1}^N D_i^{Inter}, \quad (13)$$



图 5 领域自适应模块工作原理。(a) 模型训练阶段; (b) 模型测试阶段

Fig. 5 Operating principle of the domain adaptation module. (a) Model training process; (b) Model testing process

$$\ell = \ell_{cls} + \beta \ell_{mrg}, \quad (14)$$

其中:  $\|p_i - p_j\|_2^2$  为异类样本原型的欧氏距离,  $N$  为分类任务类别数,  $\ell_{cls}$  为交叉熵损失,  $\beta$  为权重参数。

度量模块采用余弦相似度计算查询样本与类别原型之间的相似性。如式 (15) 利用 Softmax 计算样本所属类别  $c$  概率, 通过所属类别概率可以对查询集样本进行分类预测, 分类准确率如式 (16) 所示:

$$p(y = c|x) = \frac{\exp(\langle f_\theta(x), p_c \rangle)}{\sum_{c'} \exp(\langle f_\theta(x), p_{c'} \rangle)}, \quad (15)$$

$$Acc = \frac{1}{N \times M} \sum_{i=1}^{N \times M} same(\tilde{y}_i, y_i), \quad (16)$$

其中:  $p_c$  为类别  $c$  的原型,  $\langle \cdot, \cdot \rangle$  表示计算两向量间的余弦相似度,  $N$  为类别数,  $M$  为每类查询样本数,  $y_i$  为样本真实标签,  $\tilde{y}_i$  为预测标签,  $same(\cdot, \cdot)$  标签相同为 1, 不同为 0。

### 3 实验结果

本节在三个公开数据集上进行实验, 并将依次介绍实验数据集、实验设置、实验结果分析、可视化分析、消融实验、领域自适应模块分析。

#### 3.1 数据集

MiniImageNet 数据集是从大规模 ImageNet 数据集中抽选了 60000 张图像构成的, 其中共包含 100 种类别, 每个类别有 600 张图像。在训练过程中选取其中 64 类图像作为训练集, 16 类图像作为验证集, 20 类图像作为测试集。

TieredImageNet 数据集<sup>[20]</sup> 同样取于 ImageNet 数据集, 相较于 miniImageNet 其规模更加庞大, 其中包含 608 种类别, 每个类别约有 1300 张图像样本。tieredImageNet 数据集按语义信息将 608 类数据划分为 34 类父级语义类别, 在训练过程中从 34 个父级类

别中选取 20 类作为训练集, 6 类数据作为验证集, 8 类数据为测试集。基于语义的类别划分方式使数据的语义差距较大, 对模型泛化能力提出了较高要求。

CUB 数据集全称为 Caltech-UCSD Birds-200 数据集, 是由美国加利福尼亚理工学院提供的鸟类数据库。其中包括了 200 种鸟类的 11788 张图像。CUB 数据集是目前细粒度分类识别的基准数据集。在训练过程中划分 100 类为训练数据集, 50 类为验证数据集, 其余 50 类为测试数据集。

以上所有数据集的图像统一裁剪为 84 pixels×84 pixels 输入网络进行训练。

#### 3.2 实验设置

模型性能很大程度取决于嵌入函数的特征提取能力。如表 1 所示, 本文采用 Conv4、ResNet-12 骨架网络, 其中骨架网络的中间层都使用了最大值池化 (max pooling)、ReLU 非线性激活函数和批归一化处理 (batch normalization) 操作。此外, ResNet-12 加入 dropout 层, 防止网络过拟合。所有实验均在 Ubuntu20.04 操作系统下使用深度学习框架 PyTorch1.4.0 进行。硬件配置为 Intel Xeon Platinum 8163、Nvidia GeForce RTX 3090×8。

在实验过程中, 采用 5-way 1-shot 和 5-way 5-shot 两种模式。模型在基类数据集上进行预训练。计算类别原型的过程中将  $\lambda$  设置为 0.5, 计算损失函数  $\ell$  时将参数  $\beta$  初始化为 0.05。模型训练 200 个 epoch 后计算分类准确率, 每个 epoch 中包含 100 个 episodes, 每个 episode 中每个类别抽取 15 张图像组成查询集。优化器采用随机梯度下降 (stochastic gradient descent, SGD) 更新模型参数, 初始学习率为 0.002, 每次训练 40 个 epoch 后学习率衰减一半; 测试阶段选取 10000 个 episodes, 模型评价指标为 95% 置信度的准确率。

表 1 骨架网络的模型结构  
Table 1 Structure of the backbone

模型结构	输出尺寸	ResNet-12	Conv4
卷积层1	42×42	[3×3, 64] × 3	[3×3, 64]
卷积层2	21×21	[3×3, 160] × 3	[3×3, 64]
卷积层3	10×10	[3×3, 320] × 3	[3×3, 64]
卷积层4	5×5	[3×3, 640] × 3	[3×3, 64]
池化层	1×1	5×5 Pool	5×5 Pool
参数量		50 MB	0.46 MB



### 3.3 实验结果分析

为了验证本文提出的 MADA 模型性能, 在 miniImageNet、tieredImageNet 与 CUB 数据集上进行对照实验。对照的方法有基于度量的 Matching Net<sup>[12]</sup>、Proto Net<sup>[13]</sup>、Relation Net<sup>[21]</sup>; 采用元学习机制的 MAML<sup>[9]</sup>、SNAIL<sup>[22]</sup>; 利用局部特征描述子进行匹配的 DN4<sup>[13]</sup>; 采用动态自空间进行分类, 为分类任务寻找合适的特征子空间的 DSN<sup>[23]</sup>; 采用类间遍历网络有效提取类内共有特征与类间独有特征的 CTM<sup>[24]</sup>。

在 5-way 1-shot 及 5-way 5-shot 设定下, 模型在 miniImageNet 数据集的收敛曲线与损失曲线如图 6~7 所示。模型在 miniImageNet、tieredImageNet 与 CUB 的 95% 置信度准确率如表 2~4 所示, 分析实验结果, 得到结论如下:

1) 在三个公开数据上, 本文方法在 5-way 5-shot 设定下的准确率相较于 5-way 1-shot 大约平均提升 13%, 对比模型每次学习 1 个样本, 模型学习 5 个样本即可大幅提升准确率。样本过少导致类别原型

向量受到个体偏差的影响。

2) 表 2 实验结果所示, 在 miniImageNet 数据集中, MADA 使用 ResNet-12 骨架网络比使用 Conv4 的分类准确率提升了 12.18% 与 10.65%。由于 ResNet-12 相较于 Conv4 网络结构更深、特征提取能力更强, 因此使用 ResNet-12 骨架网络可以有效提升图像分类准确率。

3) 表 2~3 实验结果所示, 相较于对照方法, MADA 在 miniImageNet、tieredImageNet 数据集中 5-way 1-shot, 5-way 5-shot 上均取得了最佳的分类精度。MADA 比 Matching Net、Proto Net 在 tieredImageNet 数据集 5-way 1-shot 实验上准确率提升 2.17%、5.02%, 5-way 5-shot 上提升 4.5%、1.7%; MADA 比 DN4 在 MiniImageNet 数据集 5-way 1-shot 与 5-way 5-shot 上准确率提升 13.08%、8.33%。Matching Net 方法采用 LSTM 学习样本特征之间的上下文信息, 由于特征向量维度过高, 导致模型的全局捕捉能力降低。MADA 将全局注意力与多尺度深度卷积结合, 提高

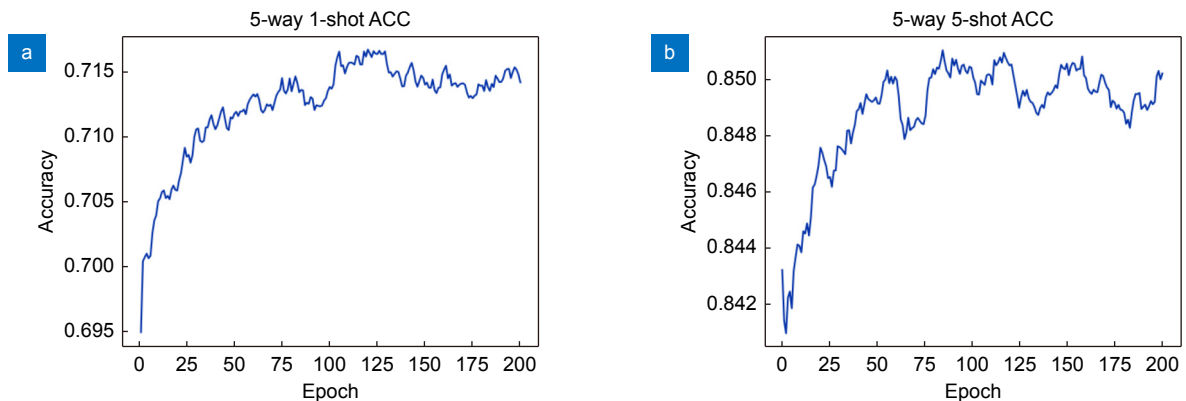


图 6 MiniImageNet 数据集模型收敛。(a) 5-way 1-shot; (b) 5-way 5-shot

Fig. 6 Model convergence on the miniImageNet dataset. (a) 5-way 1-shot; (b) 5-way 5-shot

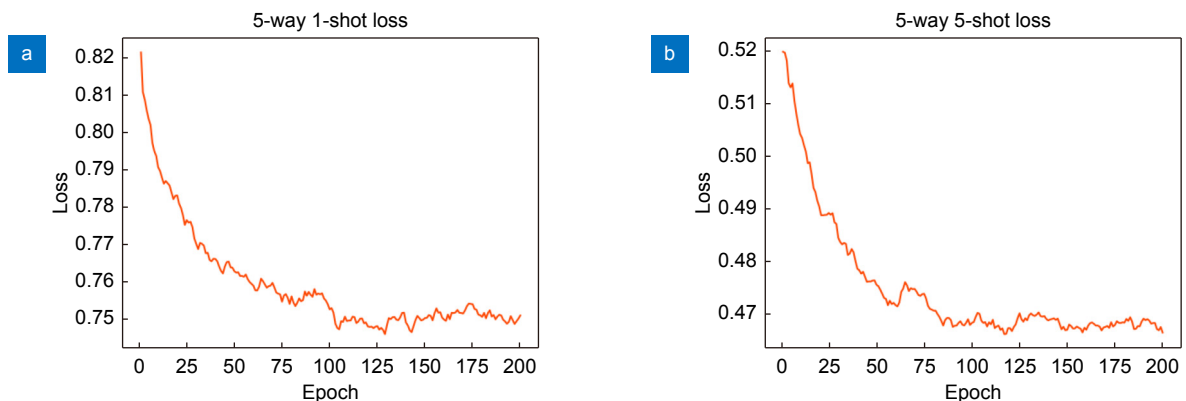


图 7 MiniImageNet 数据集模型损失。(a) 5-way 1-shot; (b) 5-way 5-shot

Fig. 7 Model loss on the miniImageNet dataset. (a) 5-way 1-shot; (b) 5-way 5-shot



表 2 MinilImageNet 数据集置信度 95% 小样本分类准确率 (episodes 为 10000)

Table 2 Few-shot classification accuracies with 95 confidence interval on the minilImageNet dataset (the number of episodes is 10000)

模型	骨架网络	5-way 1-shot	5-way 5-shot
Matching Net <sup>[12]</sup>	Conv4	43.56±0.84	55.31±0.37
Proto Net <sup>[13]</sup>	Conv4	49.42±0.78	68.20±0.66
Relation Net <sup>[21]</sup>	Conv4	50.44±0.82	65.32±0.70
MAML <sup>[9]</sup>	Conv4	48.70±1.84	63.11±0.92
DN4 <sup>[14]</sup>	Conv4	51.24±0.74	71.02±0.64
DSN <sup>[23]</sup>	Conv4	51.78±0.96	68.99±0.69
BOIL <sup>[25]</sup>	Conv4	49.61±0.16	66.45±0.37
MADA(ours)	Conv4	<b>55.27±0.20</b>	<b>72.12±0.16</b>
Matching Net <sup>[12]</sup>	ResNet-12	65.64±0.20	78.73±0.15
Proto Net <sup>[13]</sup>	ResNet-12	60.37±0.83	78.02±0.75
DN4 <sup>[14]</sup>	ResNet-12	54.37±0.36	74.44±0.29
DSN <sup>[23]</sup>	ResNet-12	62.64±0.66	78.73±0.45
SNAIL <sup>[22]</sup>	ResNet-12	55.71±0.99	68.88±0.92
CTM <sup>[24]</sup>	ResNet-12	64.12±0.82	80.51±0.14
MADA(ours)	ResNet-12	<b>67.45±0.20</b>	<b>82.77±0.13</b>

表 3 TieredImageNet 数据集置信度 95% 小样本分类准确率 (episodes 为 10000)

Table 3 Few-shot classification accuracies with 95 confidence interval on the tieredImageNet dataset (the number of episodes is 10000)

模型	骨架网络	5-way 1-shot	5-way 5-shot
Matching Net <sup>[12]</sup>	ResNet-12	68.50±0.92	80.60±0.71
Proto Net <sup>[13]</sup>	ResNet-12	65.65±0.92	83.40±0.65
MetaOpt Net <sup>[26]</sup>	ResNet-12	65.99±0.72	81.56±0.53
TPN <sup>[27]</sup>	ResNet-12	59.91±0.94	73.30±0.75
CTM <sup>[24]</sup>	ResNet-12	68.41±0.39	84.28±1.74
LEO <sup>[10]</sup>	ResNet-12	66.63±0.05	81.44±0.09
MADA(ours)	ResNet-12	<b>70.67±0.22</b>	<b>85.10±0.15</b>

表 4 CUB 数据集置信度 95% 小样本分类准确率 (episodes 为 10000)

Table 4 Few-shot classification accuracies with 95 confidence interval on the CUB dataset (the number of episodes is 10000)

模型	骨架网络	5-way 1-shot	5-way 5-shot
Matching Net <sup>[12]</sup>	Conv4	60.52±0.88	75.29±0.75
Proto Net <sup>[13]</sup>	Conv4	50.46±0.88	76.39±0.64
Relation Net <sup>[21]</sup>	Conv4	61.10±0.79	76.11±0.69
MAML <sup>[9]</sup>	Conv4	54.73±0.97	75.75±0.76
Baseline++ <sup>[28]</sup>	Conv4	60.53±0.83	79.34±0.61
DN4 <sup>[14]</sup>	Conv4	<b>66.63±0.05</b>	<b>81.44±0.09</b>
MADA(ours)	Conv4	62.12±0.24	77.63±0.17

模型在高维向量的多尺度表征能力, 有效捕捉长程信息, 增强模型在长序列上的学习能力。通过多尺度注意力捕捉样本特征间的上下文信息, 强化图像分类相关的类别特征, 从而提高小样本图像分类能力; DN4 方法采用局部特征描述子表征样本信息, 在度量阶段使用局部度量方式对目标图像进行类别预测,

miniImageNet、tieredImageNet 数据集的基类与新类之间语义信息差距大, 导致 DN4 在测试阶段泛化能力较差。MADA 的领域自适应模块能有效缓解测试阶段的语义差异, 丰富特征信息。领域自适应模块在训练阶段动态增加异类样本间隔, 将样本相对均匀地分布在特征空间中, 为测试阶段的新类数据预留充足

的表征空间, 实现新类样本的准确表征, 减缓领域漂移问题的影响, 增强模型的泛化能力, 提高鲁棒性。

4) 如表 4 数据所示, 在 CUB 鸟类分类数据集中, MADA 在 5-way 1-shot、5-way 5-shot 的准确率先大部分算法, 但低于 DN4。MADA 比 MAML 元学习方法在 5-way 1-shot、5-way 5-shot 实验上准确率提升 7.39%、1.88%; MADA 比 DN4 方法在 5-way 1-shot、5-way 5-shot 实验上准确率低 4.51%、3.81%。MAML 通过大量分类任务学习图像分类元知识, 直接利用骨架网络提取的特征训练元学习器, 导致 MAML 关注全局特征而忽略分任务中样本间的联系, 从而降低了模型的学习能力。MADA 的多尺度注意力模块强化样本间的上下文信息, 突出分类任务相关特征, 增强模型的分类能力, 从而提高模型分类准确率; DN4 方法利用局部特征描述子对图像进行度量, 关注类别的细节特征。CUB 数据集中的图像均为鸟类图像, 其总体特征大体相近, 仅在部分区域存在差距, DN4 方法的局部度量机制会捕捉到不同鸟类在喙部、足部、羽毛等富有辨识度的特征, 提高图像识别准确率。此外, CUB 数据集的基类与新类数据语义相似, MADA 的领域自适应模块的性能减弱, 导致模型在 CUB 数据集的分类准确率较低。

### 3.4 可视化分析

#### 3.4.1 CNN 特征可视化

卷积神经网络 (CNN) 由多个卷积层、池化层与激活函数构成<sup>[29]</sup>。如图 8 所示为 CNN 的中间层视觉特征可视化图像, 随着网络层次不断加深, 特征图像可视性不断减弱。CNN 的浅层网络提取图像的纹理细节特征, 中层网络提取图像轮廓形状特征<sup>[30]</sup>。随着网络层数不断加深, 特征图的分辨率变小, CNN 提取的深度特征更抽象, 特征信息更关键。多通道的深度特征向量包含了丰富的视觉判别信息, 利用深度特

征向量即可实现图像识别<sup>[31]</sup>。

#### 3.4.2 特征向量降维可视化

为了直观分析 MADA 模型为新类样本提供充足表征空间, 增强模型的表征能力。本文使用 t-SNE<sup>[32]</sup> 算法对 miniImageNet 数据在特征空间中的高维向量进行降维可视化处理<sup>[33]</sup>。如图 9(a) 所示为使用基准方法 (Baseline) 的特征分布, Baseline 直接利用嵌入函数提取样本特征, 采用距离度量函数对查询集样本进行分类预测。特征向量降维后, 5 类样本基本上被分为 5 个簇, 存在离群样本点混叠在其他类别簇中, 簇内样本分布松散, 且不同类别的簇之间的间隔较小。上述分析表明, Baseline 忽略了样本的分类相关特征, 模型的表征能力不足。

图 9(b) 为 MADA 在特征空间的数据分布情况。MADA 首先使用多尺度注意力强化分类任务相关特征, 其次引入间隔损失来动态增加异类样本间隔空间, 最终使用距离度量函数预测目标样本类别。特征向量降维处理后, 样本明显被分为 5 个簇, 离群点数量相较于 Baseline 有所减少, 簇内样本分布紧凑, 同时异类样本簇的间隔空间有所增加。上述分析表明, MADA 采用注意力机制捕捉样本特征间的语义信息, 为解决向量维数过高导致模型的捕捉能力下降, 融合多尺度深度卷积与全局自注意力来强化同类样本相关特征, 增强嵌入函数的表征能力, 减缓样本在特征空间的离群现象。此外, MADA 在训练阶段将异类样本相互推离, 为测试阶段的新类样本保留充足的表征空间, 减缓语义差异导致的领域漂移现象。

#### 3.4.3 多尺度注意力

MADA 使用多尺度注意力强化分类任务相关特征, 图 10 为 MADA 在 miniImageNet 数据集上的注意力热图。全局注意力虽然在一定程度上关注到图像主体, 并对目标主体富有判别性的特征进行强化, 但

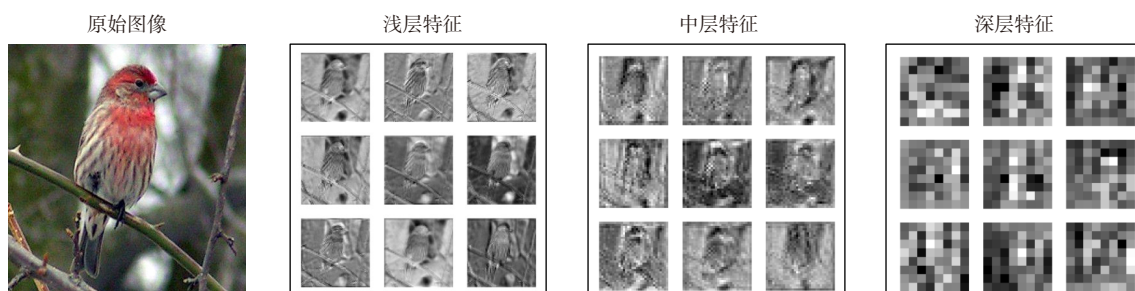


图 8 卷积神经网络视觉特征可视化

Fig. 8 Visualization of features based on convolutional neural network

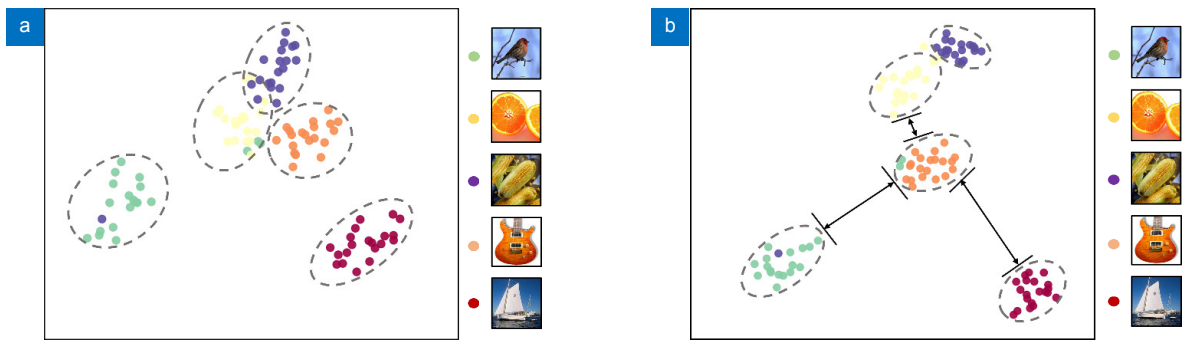


图 9 MinilImageNet 中 5 类图像特征向量可视化。(a) Baseline 方法; (b) 本文方法

Fig. 9 Visualization of image features in five classes of minilImageNet. (a) Baseline method; (b) MADA method

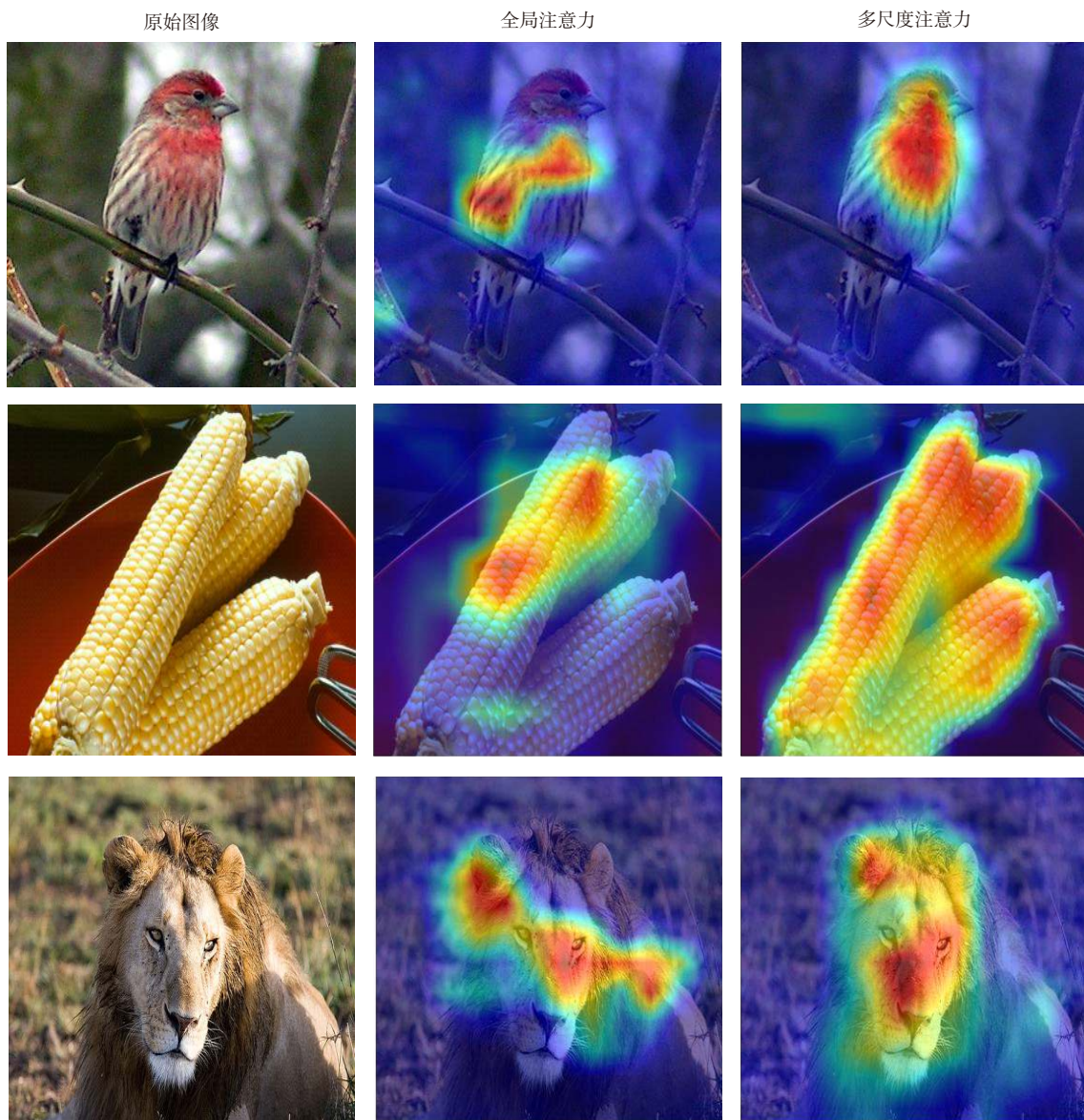


图 10 注意力热图

Fig. 10 Attention heat map



是存在部分关键区域被忽视。如图 10 中的多个玉米图像, 全局注意力的捕捉能力出现减弱; 多尺度注意力可以有效缓解由于特征向量维度过高导致的性能衰弱问题, 如图 10 中的狮子, 多尺度注意力有效增强了鬃毛、嘴部等区域。上述分析表明, 多尺度注意力将全局注意力与不同感受野的深度卷积进行融合, 有效提升模型的长程捕捉能力, 突出识别任务中目标的判别性特征, 优化类别原型表征, 从而提升模型的识别准确率。

### 3.5 消融实验

为了验证网络各个模块对模型的影响, 在 miniImageNet 数据集上使用 ResNet-12 骨架网络设计消融实验。依次在基准网络上添加多尺度注意力 (multi-scale attention, MA)、领域自适应 (domain adaptation, DA) 与数据扩充 (data enhancement, DE) 模块, 测试模型在 5-way 1-shot、5-way 5-shot 设定下的图像识别准确率。消融实验结果如表 5 所示。

实验表明本文所提出的三个模块均在一定程度上提高了模型的准确率, 数据增强模块的提升效果相对较少, 其中多尺度注意力机制显著提高了模型的精度, 通过添加 MA 模块, 准确率相较 Baseline 提升了 5.47%、3.92%, 加入 DA 模块后, 模型在 MA 基础上提升了 1.37%、0.47%。多尺度注意力与领域自适应的引入提升了模型性能, 有效增强任务相关特征,

面对未见的新类数据模型可以准确进行表征。

### 3.6 领域自适应模块分析

为了进一步探究领域自适应模型的效果, 本文在 miniImageNet 数据集保持相同条件下, 通过调整损失函数  $\ell$  中的参数  $\beta$  来检验间隔损失  $\ell_{\text{mg}}$  对模型性能的影响。实验在 5-way 1-shot、5-way 5-shot 模式下进行, 骨架网络选取 ResNet-12, 参数  $\beta$  从 0.05 开始逐步增加到 1, MADA 模型的准确率变化如图 11 所示。

实验结果表明, 在 5-way 1-shot、5-way 5-shot 条件下, 随着间隔损失函数的引入, 模型的性能开始有所提升。随着参数  $\beta$  的不断增大, 准确率呈现下降的趋势。在 5-way 1-shot 任务中, 当  $\beta$  为 0.3 时, 模型的性能最优, 在  $\beta$  继续增长后, 领域自适应模块效果减弱; 在 5-way 5-shot 任务中, 当  $\beta$  为 0.2 时, 模型性能达到最优, 且  $\beta$  持续增大后导致模型分类准确率降低。上述现象表明, 在间隔损失加入网络后, 模型在训练过程中会将异类样本在特征空间进行推离, 将所有样本相对均匀表征在特征空间中, 均匀分布的特征向量保留了更多空间信息。此外样本间保留了充足的间隔空间, 为测试数据准确的表征提供保障。当  $\beta$  持续增大时, 损失函数  $\ell$  中分类损失函数  $\ell_{\text{cls}}$  被弱化, 此外当间隔损失  $\ell_{\text{mg}}$  过大导致样本特征信息丢失, 特征表征空间退化, 模型的识别准确率降低。

表 5 在 miniImageNet 数据集上置信度 95% 小样本分类的消融实验 (episodes 为 10000)

Table 5 Ablation study of few-shot classification accuracies with 95 confidence interval on the miniImageNet (the number of episodes is 10000)

网络	MA	DA	DE	5-way 1-shot	5-way 5-shot
Baseline	×	×	×	60.37±0.83	78.02±0.75
MA	√	×	×	65.84±0.23	81.94±0.34
MADA	√	√	×	67.21±0.18	82.41±0.48
MADA+	√	√	√	67.45±0.20	82.77±0.13

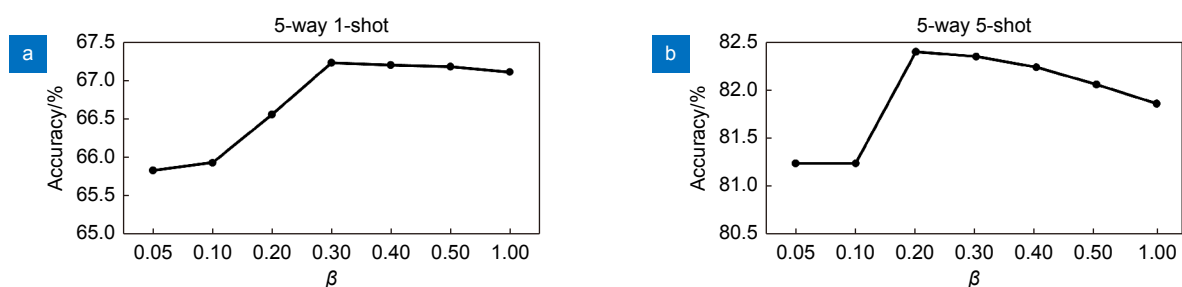


图 11 领域自适应模型性能分析。(a) 5-way 1-shot; (b) 5-way 5-shot

Fig. 11 Domain adaptation module analysis. (a) 5-way 1-shot; (b) 5-way 5-shot



## 4 结论

针对小样本情况下, 度量学习方法的类别原型存在偏差、模型泛化能力差问题。本文提出了多尺度注意力与领域自适应的小样本图像识别算法 (MADA)。首先设计多尺度注意力模块强化分类相关特征, 丰富特征向量表征信息, 优化类别原型表征; 此外, 为缓解领域漂移问题, 设计间隔损失函数构建领域自适应模块, 使模型能准确表征未见类别样本, 增强模型的泛化能力。本文在三个公开数据集上进行实验, 结果表明 MADA 方法能有效提高小样本图像识别的准确率。本文为研究小样本学习提供了数据领域自适应的研究思路, 但是仍存在部分改进空间。在未来的工作中, 可以研究如何优化相似性度量模块增强网络的分类能力, 同时可以改进嵌入函数以增强网络的学习效率。

## 参考文献

- [1] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*, 2016: 770–778. <https://doi.org/10.1109/CVPR.2016.90>.
- [2] Devlin J, Chang M W, Lee K, et al. BERT: pre-training of deep bidirectional transformers for language understanding[C]//*Proceedings of 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2019: 4171–4186. <https://doi.org/10.18653/v1/N19-1423>.
- [3] Zhao C M, Chen Z B, Zhang J L. Application of aircraft target tracking based on deep learning[J]. *Opto-Electron Eng*, 2019, **46**(9): 180261.  
赵春梅, 陈忠碧, 张建林. 基于深度学习的飞机目标跟踪应用研究[J]. *光电工程*, 2019, **46**(9): 180261.
- [4] Shi C, Chen E Q, Qi L. Ship detection from infrared video[J]. *Opto-Electron Eng*, 2018, **45**(6): 170748.  
石超, 陈恩庆, 齐林. 红外视频中的舰船检测[J]. *光电工程*, 2018, **45**(6): 170748.
- [5] Tan J R, Wang C B, Li B Y, et al. Equalization loss for long-tailed object recognition[C]//*Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020: 11659–11668. <https://doi.org/10.1109/CVPR42600.2020.01168>.
- [6] Li F F, Fergus R, Perona P. A Bayesian approach to unsupervised one-shot learning of object categories[C]//*Proceedings of the Ninth IEEE International Conference on Computer Vision*, 2003: 1134–1141. <https://doi.org/10.1109/ICCV.2003.1238476>.
- [7] Mehrotra A, Dukkipati A. Generative adversarial residual pairwise networks for one shot learning[Z]. arXiv: 1703.08033, 2017. <https://doi.org/10.48550/arXiv.1703.08033>.
- [8] Chen Z T, Fu Y W, Zhang Y D, et al. Multi-level semantic feature augmentation for one-shot learning[J]. *IEEE Trans Image Process*, 2019, **28**(9): 4594–4605.
- [9] Finn C, Abbeel P, Levine S. Model-agnostic meta-learning for fast adaptation of deep networks[C]//*Proceedings of the 34th International Conference on Machine Learning*, 2017: 1126–1135. <https://doi.org/10.5555/3305381.3305498>.
- [10] Rusu A A, Rao D, Sygnowski J, et al. Meta-learning with latent embedding optimization[C]//*Proceedings of the 7th International Conference on Learning Representations*, 2019.
- [11] Frikha A, Krompaß D, Köpken H G, et al. Few-shot one-class classification via meta-learning[J]. *Proc AAAI Conf Artif Intell*, 2021, **35**(8): 7448–7456.
- [12] Vinyals O, Blundell C, Lillicrap T, et al. Matching networks for one shot learning[C]//*Proceedings of the 30th International Conference on Neural Information Processing Systems*, 2016: 3637–3645. <https://doi.org/10.5555/3157382.3157504>.
- [13] Snell J, Swersky K, Zemel R. Prototypical networks for few-shot learning[C]//*Proceedings of the 31st Conference on Neural Information Processing Systems*, 2017: 4077–4087.
- [14] Li W B, Wang L, Huo J L, et al. Revisiting local descriptor based image-to-class measure for few-shot learning[C]//*Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019: 7253–7260. <https://doi.org/10.1109/CVPR.2019.00743>.
- [15] Luo X, Wei L H, Wen L J, et al. Rectifying the shortcut learning of background for few-shot learning[C]//*Proceedings of the 35th Conference on Neural Information Processing Systems*, 2021.
- [16] Chen Y H, Li W, Sakaridis C, et al. Domain adaptive faster R-CNN for object detection in the wild[C]//*Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018: 3339–3348. <https://doi.org/10.1109/CVPR.2018.00352>.
- [17] Gong R, Li W, Chen Y H, et al. DLOW: domain flow for adaptation and generalization[C]//*Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019: 2472–2481. <https://doi.org/10.1109/CVPR.2019.00258>.
- [18] He K M, Chen X L, Xie S N, et al. Masked autoencoders are scalable vision learners[C]//*Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022: 15979–15988. <https://doi.org/10.1109/CVPR52688.2022.01553>.
- [19] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]//*Proceedings of the 31st International Conference on Neural Information Processing Systems*, 2017: 6000–6010. <https://doi.org/10.5555/3295222.3295349>.
- [20] Ren M Y, Triantafillou E, Ravi S, et al. Meta-learning for semi-supervised few-shot classification[C]//*Proceedings of the 6th International Conference on Learning Representations*, 2018.
- [21] Sung F, Yang Y X, Zhang L, et al. Learning to compare: relation network for few-shot learning[C]//*Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018: 1199–1208. <https://doi.org/10.1109/CVPR.2018.00131>.
- [22] Mishra N, Rohaninejad M, Chen X, et al. A simple neural attentive meta-learner[C]//*Proceedings of the 6th International Conference on Learning Representations*, 2018.
- [23] Simon C, Koniusz P, Nock R, et al. Adaptive subspaces for few-shot learning[C]//*Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020: 4135–4144. <https://doi.org/10.1109/CVPR42600.2020.00419>.
- [24] Li H Y, Eigen D, Dodge S, et al. Finding task-relevant features for few-shot learning by category traversal[C]//*Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern*

- Recognition, 2019: 1–10. <https://doi.org/10.1109/CVPR.2019.00009>.
- [25] Oh J, Yoo H, Kim C H, et al. BOIL: towards representation change for few-shot learning[C]//*Proceedings of the 9th International Conference on Learning Representations*, 2021.
- [26] Lee K, Maji S, Ravichandran A, et al. Meta-learning with differentiable convex optimization[C]//*Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019: 10649–10657. <https://doi.org/10.1109/CVPR.2019.01091>.
- [27] Liu Y B, Lee J, Park M, et al. Learning to propagate labels: Transductive propagation network for few-shot learning[C]//*Proceedings of the 7th International Conference on Learning Representations*, 2019.
- [28] Chen W Y, Liu Y C, Kira Z, et al. A closer look at few-shot classification[C]//*Proceedings of the 7th International Conference on Learning Representations*, 2019.
- [29] Chen X, Peng D L, Gu Y. Real-time object detection for UAV images based on improved YOLOv5s[J]. *Opto-Electron Eng*, 2022, 49(3): 210372.
- 陈旭, 彭冬亮, 谷雨. 基于改进YOLOv5s的无人机图像实时目标检测[J]. *光电工程*, 2022, 49(3): 210372.
- [30] Li X, Li L P, Lazovik A, et al. RGB-D object recognition algorithm based on improved double stream convolution recursive neural network[J]. *Opto-Electron Eng*, 2021, 48(2): 200069.
- 李珣, 李林鹏, Lazovik A, 等. 基于改进双流卷积递归神经网络的RGB-D物体识别方法[J]. *光电工程*, 2021, 48(2): 200069.
- [31] Cao Z, Shang L D, Yin D. A weakly supervised learning method for vehicle identification code detection and recognition[J]. *Opto-Electron Eng*, 2021, 48(2): 200270.
- 曹志, 尚丽丹, 尹东. 一种车辆识别代号检测和识别的弱监督学习方法[J]. *光电工程*, 2021, 48(2): 200270.
- [32] van der Maaten L, Hinton G. Visualizing data using t-SNE[J]. *J Mach Learn Res*, 2008, 9(86): 2579–2605.
- [33] Tang B, Jin W, Li G, et al. The cloud retrieval of combining sparse representation with subspace projection[J]. *Opto-Electron Eng*, 2019, 46(10): 180627.
- 唐彪, 金炜, 李纲, 等. 结合稀疏表示和子空间投影的云图检索[J]. *光电工程*, 2019, 46(10): 180627.

## 作者简介



陈龙 (1997-), 男, 硕士研究生, 主要研究跨领域图像分类、小样本学习。

E-mail: chenlong20@mails.ucas.ac.cn



【通信作者】张建林, 男, 博士, 博士生导师, 主要研究智能图像处理, 重点以机器学习为基础进行海量图像数据的分析理解, 让机器具有类似人眼与脑的信息获取、提取与学习认知功能, 实现机器的智能化、自动化。

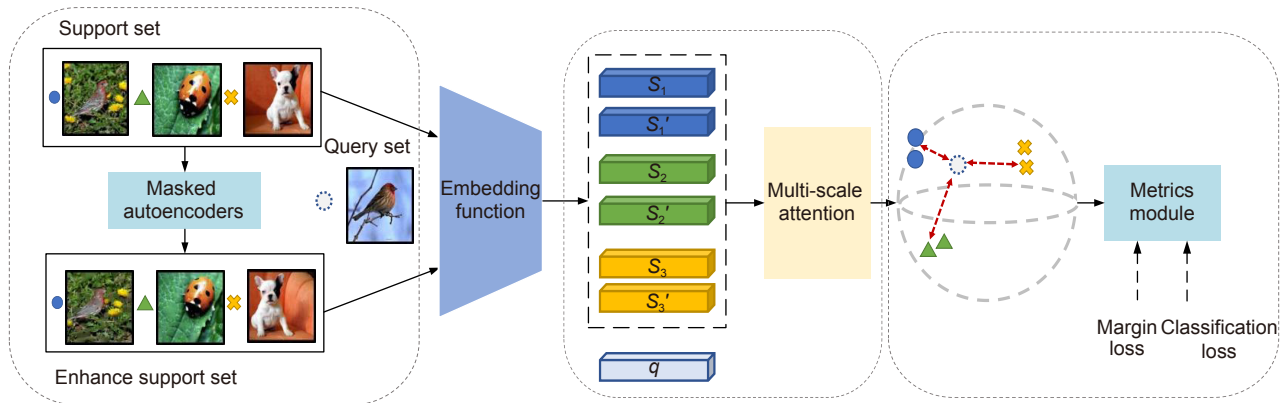
E-mail: jlin\_zh@163.com



扫描二维码, 获取PDF全文

# Few-shot image classification via multi-scale attention and domain adaptation

Chen Long<sup>1,2</sup>, Zhang Jianlin<sup>1\*</sup>, Peng Hao<sup>1,2</sup>, Li Meihui<sup>1</sup>, Xu Zhiyong<sup>1</sup>, Wei Yuxing<sup>1</sup>



Framework of multi-scale attention and domain adaptation for few-shot classification

**Overview:** In recent years, deep learning has been highly successful in the field of computer vision. However, deep learning is often hampered when the dataset is limited. Few-shot learning is proposed to tackle this problem. Few-shot learning can generalize to the new task which has a small number of samples by using prior knowledge. Metric learning is often used to solve few-shot classification, which maps samples to feature space and classifies query samples by using similarity metric functions. Prototypes calculated by metric learning are inaccurate under the condition of limited samples. Moreover, the generalization ability of the model is poor.

To improve the performance of few-shot classification, we present a general and flexible method named Multi-Scale Attention and Domain Adaptation Network (MADA). Firstly, to tackle the problem of limited samples, a masked autoencoder is used to image augmentation. Moreover, it can be inserted as a plug-and-play module into a few-shot classification. Secondly, the multi-scale attention module can adapt feature vectors extracted by embedding function to the current classification task. Multi-scale attention machine strengthens the discriminative image region by focusing on relating samples in both base class and novel class, which makes prototypes more accurate. In addition, the embedding function pays attention to the task-specific feature. Thirdly, the domain adaptation module is used to address the domain shift caused by the difference in data distributions of the two domains. The domain adaptation module consists of the metric module and the margin loss function. The margin loss pushes different prototypes away from each other in the feature space. Sufficient margin space in feature space improves the generalization performance of the method. The experimental results show the classification accuracy of the proposed method is 67.45% for 5-way 1-shot and 82.77% for 5-way 5-shot on the miniImageNet dataset. The classification accuracy is 70.57% for 5-way 1-shot and 85.10% for 5-way 5-shot on the tieredImageNet dataset. The classification accuracy of our method is better than most previous methods. After dimension reduction and visualization of features by using t-SNE, it can be concluded that domain drift is alleviated, and prototypes are more accurate. The multi-scale attention module enhanced feature representations are more discriminative for the target classification task. In addition, the domain adaptation module improves the generalization ability of the model.

Chen L, Zhang J L, Peng H, et al. Few-shot image classification via multi-scale attention and domain adaptation[J]. *Opto-Electron Eng*, 2023, 50(4): 220232; DOI: 10.12086/oe.2023.220232

Foundation item: National Natural Science Foundation of China (62101529)

<sup>1</sup>Institute of Optics and Electronics, Chinese Academy of Science, Chengdu, Sichuan 610209, China; <sup>2</sup>School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Science, Beijing 100049, China

\* E-mail: jlin\_zh@163.com