

中图法分类号: TP391 文献标识码: A 文章编号: 1006-8961(2023)07-2093-12

论文引用格式: Dong Y Y, Song B B and Sun W F. 2023. Local feature fusion network-based few-shot image classification. Journal of Image and Graphics, 28(07):2093-2104(董杨洋, 宋蓓蓓, 孙文方. 2023. 局部特征融合的小样本分类. 中国图象图形学报, 28(07):2093-2104)[DOI: 10.11834/jig.220079]

局部特征融合的小样本分类

董杨洋¹, 宋蓓蓓^{1*}, 孙文方²

1. 长安大学信息工程学院, 西安 710064; 2. 西安电子科技大学空间科学与技术学院, 西安 710126

摘要: 目的 小样本学习是一项具有挑战性的任务,旨在利用有限数量的标注样本数据对新的类别数据进行分类。基于度量的元学习方法是当前小样本分类的主流方法,但往往仅使用图像的全局特征,且模型分类效果很大程度上依赖于特征提取网络的性能。为了能够充分利用图像的局部特征以及提高模型的泛化能力,提出一种基于局部特征融合的小样本分类方法。**方法** 首先,将输入图像进行多尺度网格分块处理后送入特征提取网络以获得局部特征;其次,设计了一个基于Transformer架构的局部特征融合模块来得到包含全局信息的局部增强特征,以提高模型的泛化能力;最后,以欧几里得距离为度量,计算查询集样本特征向量与支持集类原型之间的距离,实现分类。**结果** 在小样本分类中常用的3个数据集上与当前先进的方法进行比较,在5-way 1-shot和5-way 5-shot的设置下相对次优结果,所提方法在MiniImageNet数据集上的分类精度分别提高了2.96%和2.9%,在CUB(Caltech-UCSD Birds-200-2011)数据集上的分类精度分别提高了3.22%和1.77%,而在TieredImageNet数据集上的分类精度与最优结果相当,实验结果表明了所提方法的有效性。**结论** 提出的小样本分类方法充分利用了图像的局部特征,同时改善了模型的特征提取能力和泛化能力,使小样本分类结果更为准确。

关键词: 小样本学习;度量学习;局部特征;Transformer;特征融合

Local feature fusion network-based few-shot image classification

Dong Yangyang¹, Song Beibei^{1*}, Sun Wenfang²

1. School of Information Engineering, Chang'an University, Xi'an 710064, China;

2. School of Aerospace Science and Technology, Xidian University, Xi'an 710126, China

Abstract: Objective The emerging convolutional neural network based (CNN-based) deep learning technique is beneficial for image context like its recognition, detection, segmentation and other related fields nowadays. However, the learning ability of CNN is often challenged for a large number of labeled samples. The CNN model is disturbed of over-fitting problem due to insufficient labeled sample for some categories. The collection task of labeled samples is time-consuming and costly. However, human-related perception has its ability to learn from a small number of samples. For example, it will be easily recognized other related new images in these categories even under a few pictures of each image category circumstances. To make CNN model have the learning ability similar to human, a new machine learning algorithm is concerned about more, called few-shot learning. Few-shot learning can be used to classify new image categories in terms of a limited amount of annotation data. Current metric-based meta-learning methods can be as one of the effective task for few-shot learning methods. However, it is implemented on the basis of global features, which cannot represent the image struc-

收稿日期: 2022-02-24; 修回日期: 2022-05-23; 预印本日期: 2022-05-30

* 通信作者: 宋蓓蓓 bbsong@chd.edu.cn

基金项目: 国家自然科学基金项目(62001059); 陕西省重点研发计划资助(2021GY-019)

Supported by: National Natural Science Foundation of China (62001059)

ture adequately. More local feature information is required to be involved in as well, which can provide discriminative and transferable information across categories. Furthermore, there are some local features representation-derived metric methods can be used to obtain pixel-level deep local descriptors as the local feature representation of image via removing the last global average pooling layer in CNN. However, local descriptors are depth but the classification effect is restricted by sacrificed contextual information of the image. Additionally, for the feature extraction network, due to limited labeled instances, it is challenged to learn a good feature representation and generalize new categories. To utilize the local features of image and improve the generalization ability of the model, we develop a few-shot classification method in terms of local feature fusion. **Method** First, to obtain local features, the input image is divided into $H \times W$ local blocks and then transferred to the feature extraction network. This feature representation-related method can demonstrate local information of the image and its context information. Multi-scale grid blocks are illustrated as well. Second, to learn and fuse the relationship between multiple local feature representations, we design a Transformer architecture based local feature fusion module because the self-attention mechanism in Transformer can capture and fuse the relationship between input sequences effectively. Each local feature consists of the information of other local features and it has fusion-after simultaneous global information. And, we concatenate the multiple local feature representations of each image as the final output. The feature representation of the original input image is enhanced and the generalization ability of the model can be improved after that. Finally, the Euclidean distance between the query image embedding and the support class prototype is calculated to classify the query image. Our training process is divided into two steps: pre-training and meta-training. For the pre-training stage, the Sofamax layer-attached backbone network is used to classify all images of the training set. To improve the generalization ability of the model, we use the data-augmented methods of random cropping, horizontal flipping and color jittering. After the pre-training, the backbone network in the model is initialized with pre-trained weights, and other components are then fine-tuned. For meta-learning stage, the episode training strategy is implemented for training. To make a fair comparison with other few-shot classification methods, the ResNet12 structure is used as the feature extractor of the backbone network, and the cross entropy loss of classification is optimized through stochastic gradient descent (SGD). The initial learning rate of the model is set to 5×10^{-4} , and we set 100 epochs in total, the learning rate is decreased by half every 10 epochs, 100 episodes and 600 episodes for training and validate in each epoch. The domain difference is larger since there are more samples in TieredImageNet dataset, more iteration is required to make the model convergent. Therefore, we set 200 epochs, and the learning rate is decreased by half very 20 epochs. In the test stage, to evaluate the average classification accuracy, such 5 000 episodes are selected from the test set randomly. **Result** Comparative analysis is based on three benchmark datasets in few-shot classification. For MiniImageNet dataset, each average classification accuracy is optimized by 2.96% and 2.9% under 5-way 1-shot and 5-way 5-shot settings. For CUB dataset, each of average classification accuracy is increased by 3.22% and 1.77%. For TieredImageNet dataset, the proposed method is equivalent to the state-of-the-art method in average classification accuracy. To fully verify the effectiveness of the proposed method, a large number of ablation experiments are also carried out as well. **Conclusion** We develop a local feature fusion-based method for few-shot classification. It is beneficial to make sufficient local features and the feature extraction ability and generalization ability of the model can be optimized as well. Our Transformer architecture based local feature fusion module can enhance feature representation further, which can be embedded into other few-shot classification methods potentially.

Key words: few-shot learning; metric learning; local feature; Transformer; feature fusion

0 引言

深度卷积神经网络(deep convolutional neural networks, DCNN)(He等, 2016; Krizhevsky等, 2017)在视觉、语言和语音等广泛的研究领域中均取得了

令人瞩目的成就(Gu等, 2018)。然而, CNN强大的学习能力往往依赖于大量的带标注样本以及多次的迭代训练, 当某些类别的样本数据不足或带标注样本数量过少时, CNN模型容易出现过拟合。此外, 收集和标注样本也会耗费大量的人力物力, 而且在现实的许多应用场景中(例如军事、医疗领域等), 由于隐私、

安全和伦理等因素,难以获得足够的带标注样本。

相比之下,人类具有从少量样本数据中学习的能力,仅通过一幅或几幅标注实例便可以识别新类别。为了使神经网络系统具有类似人类的这种学习能力,一种称为小样本学习(few-shot learning, FSL)的机器学习方法受到研究者的关注(Lake等,2013; Santoro等,2016)。FSL的目的是使模型通过对每类少量标注样本的训练,以实现对这些类别新图像的识别。为了解决小样本分类问题,研究者通常采用元学习范式(Thrun和Pratt,1998; Vilalta和Drissi,2002),元学习的目标是在学习多个任务的过程中学会处理新任务。

在元学习的小样本分类算法中存在两种主流方法,即基于优化的方法和基于度量的方法。在基于优化的方法中,一种途径旨在学习良好的模型初始化参数,使模型可以通过有限数量的标注样本和少量的迭代次数就可以适应新的分类任务。例如, Finn等人(2017)提出的MAML(model-agnostic meta-learning)及其多种变体(Antoniou等,2019; Sun等,2019; Jamal和Qi,2019; Rusu等,2019)。另一种途径是学习优化器,例如Ravi和Larochelle(2017)将LSTM(long short term memory)训练作为更新模型参数的优化器,取代随机梯度下降优化器。虽然这些基于优化的方法能够在有限的训练样本下实现对新类的快速适应,但在处理基类和新类之间的领域转换方面仍然存在困难。

基于度量的方法,其目的是学习一个适当的特征空间来表示样本,在该特征空间中可以使用距离度量来区分来自不同类别的样本。例如,Vinyals等人(2016)提出的匹配网络,通过计算两个样本特征向量之间的余弦距离来表示其相似性。Snell等人(2017)提出的原型网络,以支持集中每类图像特征向量的均值作为其相应的类原型来学习度量空间,并使用欧几里得距离来计算样本相似性。Sung等人(2018)提出了关系网络,用一个由CNN构成的度量模块代替固定的距离函数,从而学习一种深度非线性度量方法。基于度量的方法通过距离函数直接计算测试图像特征向量和训练图像特征向量之间的距离来分类,能够避免小样本下分类器学习优化困难的问题,但这些方法往往仅使用图像的全局特征。而局部特征可以提供跨类别的可区分和可传递的信息,基于局部特征表示的度量方法有DN4(deep

nearest neighbor neural network)(Li等,2019a)、DC+IMP(dense classification + implanting)(Lifchitz等,2019)和DeepEMD(deep earth mover's distance)(Zhang等,2020)等。他们通过去掉CNN中最后的全局平均池化层来获得像素级的深层局部描述子,以此作为图像的局部特征表示。但这种深层局部描述子会丢失图像的上下文信息。此外,当使用固定的距离函数来度量时,模型分类效果很大程度上依赖于特征提取网络的性能。

由于基于度量学习的方法简单有效而且广泛流行,基于此,为了能够有效利用图像的局部可区分信息并提高模型的特征表示能力和泛化能力,本文提出一种基于局部特征融合的小样本分类方法。1)将输入图像经过分块处理再送入特征提取网络以获得局部特征,该局部特征不仅包含用于提供跨类别的可区分的局部信息,同时还具有图像的上下文信息。2)针对特征提取网络难以学习并泛化的问题,本文设计了一个基于Transformer架构的局部特征融合模块来学习每幅图像的多个局部特征之间的关系,以获得更好的特征表示。3)在3个小样本分类的通用数据集上进行了大量实验,验证了所提算法的有效性。

1 方 法

1.1 问题定义

对于一个小样本目标分类任务,即存在一个支持集 S ,它包括 N 个不同的图像类以及每个类的 K 个带标注样本。然后给定一个查询集 Q ,它包括与支持集 S 相同类别的 N 类图像以及每类图像的 q 个未标注样本。小样本学习的目的是使模型能够根据集合 S 来对集合 Q 中的每个未标注样本进行分类,此设置称为 N -way K -shot小样本分类任务。然而,当集合 S 中每类只有很少的标注样本时,很难有效地直接学习到一个模型来对集合 Q 中的未标注样本进行分类。因此小样本学习通常借助于一个辅助集合 A ,采用元学习范式来学习迁移知识,以提高集合 Q 上的分类效果。集合 A 中包含了大量的类以及带标注样本,但是它与集合 S 不相交。通常将集合 A 称为基类或源类,集合 S 称为新类或目标类。因此与传统的机器学习不同,FSL的目标是从一组有丰富标注样本的基类(源类)中学习一个分类器,然后去

适应一组标注数据很少的新类(目标类)(Wang等, 2020)。

为了学习迁移知识,使用匹配网络中提出的情景(episode)式训练机制(Vinyals等, 2016)来训练模型。该机制通过使训练过程与测试条件相匹配,以提高模型的泛化能力。具体地,在每一次迭代训练中,通过模拟一个小样本目标分类任务,构建一个episode来训练分类模型,该episode由从辅助集合 $\mathbf{A}$ 中随机抽样的支持集 $\mathbf{A}_s$ 和查询集 $\mathbf{A}_q$ 组成。一般来说, $\mathbf{A}_s$ 和 $\mathbf{S}$ 有着相同数量的way(即类)和shot(即每类中的带标注图像数量),换句话说,在 $\mathbf{A}_s$ 中恰好有 N 个类,每类有 K 个带标注样本。在训练过程中,通

过从辅助集合 $\mathbf{A}$ 中构建数万个episode(每个episode可以看做是独立的任务)来训练分类模型,用构建的查询集的真值标签来提供监督学习。在测试阶段,利用支持集 $\mathbf{S}$ 学习得到的模型可以直接用于对查询集 $\mathbf{Q}$ 中每个未标注样本进行分类。

1.2 网络模型设计

本文提出的基于局部特征融合的小样本分类模型是一个可以端到端训练的神经网络架构。总体框架如图1所示,为5-way 1-shot小样本分类任务的示意图。所提方法主要由3个模块构成,包括局部特征提取模块、基于Transformer架构的局部特征融合模块以及度量模块。

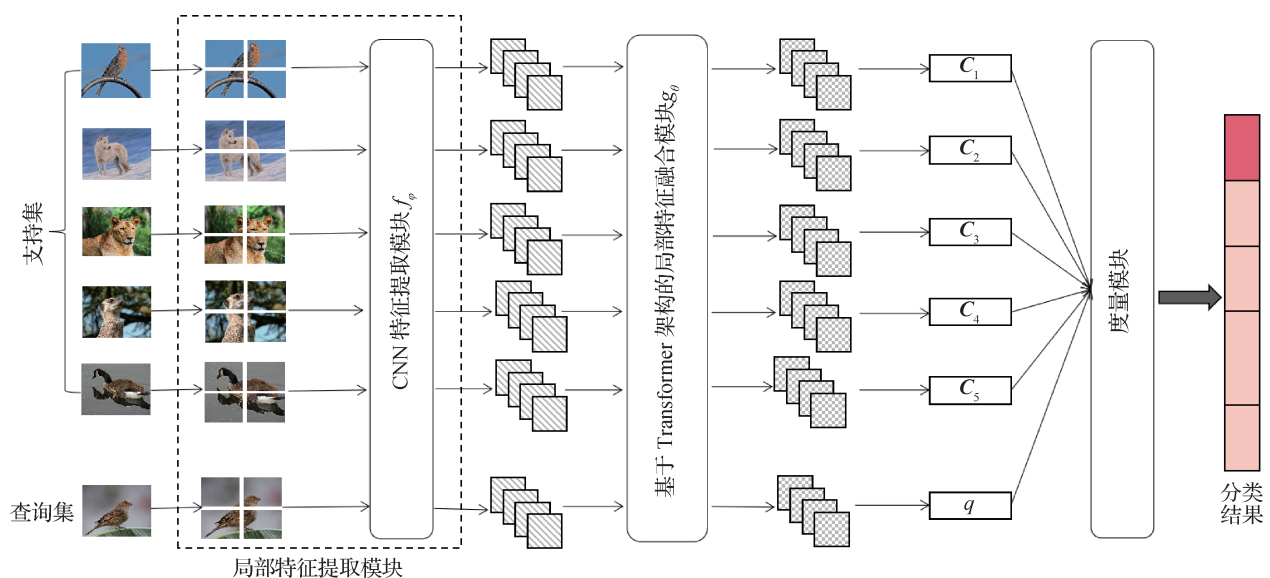


图1 模型总体框架

Fig. 1 Overall framework of the model

1.2.1 原型网络

采用基于度量的原型网络(prototypical network, ProtoNet)(Snell等, 2017)作为基本的FSL分类器。该网络由一个特征提取器和一个简单的无参分类器构成,即图1中的CNN特征提取模块 f_φ 和度量模块。训练过程只需对特征提取器的参数进行元学习。

原型网络通过CNN学习了一个输入到嵌入空间的非线性映射,各个样本通过该映射投影到同一嵌入空间中。对于支持集中每个类的样本,提取在嵌入空间中的平均值作为其类原型,并使用欧几里得距离作为度量。通过训练,使查询集样本到同类原型距离近,到其他类别原型的距离远。具体地,对于支持集 $\mathbf{S}$,原型网络先为其每个类别计算出

一个类原型 $\mathbf{C}_k$,可表示为

$$\mathbf{C}_k = \frac{1}{|\mathbf{S}_k|} \sum_{x \in \mathbf{S}_k} f_\varphi(x) \quad (1)$$

式中, x 为支持集样本, $\mathbf{S}_k$ 为支持集中类别为 k 的所有样本, $k \in [1, N]$, N 表示支持集 $\mathbf{S}$ 中的所有类别数。 f_φ 为带有可学习参数 φ 的特征提取函数,使用的是CNN结构。

测试时,对于查询样本 q ,原型网络计算其特征向量到支持集每个类原型之间的距离,再经softmax函数得到查询样本 q 所属类别的概率分布,具体计算为

$$p_\varphi(y = k | q) = \frac{\exp(-d(f_\varphi(q), \mathbf{C}_k))}{\sum_{k' \in [1, N]} \exp(-d(f_\varphi(q), \mathbf{C}_{k'}))} \quad (2)$$

式中, $d(\cdot, \cdot)$ 是距离度量函数, 采用欧几里得距离。 k 表示查询样本 q 的真实类别, y 表示查询样本 q 的预测类别, k' 为支持集中的全部类别。采用随机梯度下降(stochastic gradient descent, SGD)最小化分类的交叉熵损失来训练模型。

1.2.2 局部特征提取

为了获取图像的局部特征表示, 利用网格分块法将输入图像均匀地划分为 $H \times W$ 个局部块(图1中, 分为 $2 \times 2 = 4$ 个局部块)。网格中的每个局部块都独立地经 CNN 特征提取网络 f_φ 提取特征, 经特征提取后, 再经一个全局平均池化层, 为每个输入图像块生成 1 个一维向量, 以便作为局部特征融合模块的输入。所有局部块生成的特征向量构成了整幅图像的特征表示。具体地, 如将输入图像 x 分为 m 个局部块, 则经特征提取网络以及全局平均池化后, 获得的特征可表示为 $f_\varphi(x) \in \mathbf{R}^{m \times d}$, 其中 d 为每个局部特征的输出维度。

在具体分块时, 少量重叠网格中局部块的区域以包含上下文信息。另外, 采用金字塔结构来获得图像的多尺度信息。即根据不同的网格大小来划分局部块, 如采用 1×1 、 2×2 和 3×3 这 3 种图像分块策略(1×1 表示将整幅图像作为输入), 然后再将所有局部块送入 CNN 特征提取模块, 以获得具有上下文信息和多尺度的局部特征向量。

1.2.3 基于 Transformer 架构的局部特征融合

尽管在网格分块时采用了重叠的局部分块, 使每个局部块都包含更多的上下文信息。但是各局部块之间的特征提取仍是相互独立的, 每个特征向量仅包含一个局部块的信息, 多个局部块之间的关联信息没有得到充分利用。

Vaswani 等人(2017)提出的 Transformer 中的自注意力机制可以很好地学习并捕获输入序列之间的关系, 从而使输出获得全局性信息。因此, 利用该自注意力机制原理设计了一个基于 Transformer 架构的局部特征融合模块, 用于学习并融合各局部特征之间的关系。具体地, 将输入图像 x 分为 m 个局部块后独立地进行特征提取, 得到的每个局部特征可表示为 $f_\varphi(x_i) \in \mathbf{R}^{1 \times d}$, $i \in [1, m]$, 然后将该幅图像所有的局部特征送入局部特征融合模块。该模块首先通过 3 个不同全连接层(fully connection, FC)的值矩阵 $W_Q \in \mathbf{R}^{d \times d'}$ 、 $W_K \in \mathbf{R}^{d \times d'}$ 、 $W_V \in \mathbf{R}^{d \times d'}$ 将每个局部特征

线性映射为查询矩阵 $Q_i \in \mathbf{R}^{1 \times d'}$ 、键值矩阵 $K_i \in \mathbf{R}^{1 \times d'}$ 和值矩阵 $V_i \in \mathbf{R}^{1 \times d'}$, $i \in [1, m]$ 。然后进行自注意力的矩阵运算, 具体为

$$\tilde{g}_\theta(x_i) = \sum_{j=1}^m \text{softmax}\left(\frac{Q_i K_j^T}{\sqrt{d'}}\right) V_j, \quad \tilde{g}_\theta(x_i) \in \mathbf{R}^{1 \times d'} \quad (3)$$

$$g_\theta(x_i) = \text{LayerNorm}(f_\varphi(x_i) + \tilde{g}_\theta(x_i) W_{FC}), \quad g_\theta(x_i) \in \mathbf{R}^{1 \times d} \quad (4)$$

式中, $\sqrt{d'}$ 表示缩放系数, $W_{FC} \in \mathbf{R}^{d' \times d}$ 表示全连接层矩阵, $f_\varphi(x_i) \in \mathbf{R}^{1 \times d}$, $i \in [1, m]$ 表示输入的 m 个 d 维局部特征, $g_\theta(x_i) \in \mathbf{R}^{1 \times d}$, $i \in [1, m]$ 表示融合后的局部特征。这样每个 d 维的局部特征里融合了其他所有局部特征的信息, 从而使每个局部特征都具有全局性信息。最后将融合后的所有局部特征向量连接起来, 即 $g_\theta(x) \in \mathbf{R}^{1 \times md}$, 构成整幅输入图像 x 的特征表示。整个结构如图2所示, 表示了局部特征 $f_\varphi(x_i)$ 经特征融合后变为 $g_\theta(x_i)$ 的过程。

在每个 N -way K -shot 的小样本图像分类任务中, 对于每幅图像, 无论是支持集还是查询集, 都采用上述方法将局部特征融合后的向量作为原图的特征表示。然后遵循原型网络, 对支持集中每类图像的特征向量求取均值作为其类原型, 通过计算查询集图像特征向量到各类原型之间的欧几里得距离来进行分类, 计算方法同式(1)和式(2)。

2 实验

2.1 实验设置

2.1.1 数据集

在 3 个常用的基准数据集上进行小样本分类实验, 分别是 MiniImageNet (Vinyals 等, 2016) 数据集、TieredImageNet (Ren 等, 2018) 数据集和 CUB (Caltech-UCSD Birds-200-2011) (Wah 等, 2011) 数据集。

MiniImageNet 数据集最早在匹配网络中提出, 并成为小样本分类中广泛使用的数据集。包含 100 个类, 每个类有 600 幅图像, 这些图像是基于 ImageNet 数据集 (Russakovsky 等, 2015) 构建的。遵循 Ravi 和 Larochelle (2017) 方法中的划分, 将数据集分为 64、16 和 20 个类, 分别用于训练、验证和测试。

TieredImageNet 数据集是一个具有更多类别的大规模数据集, 也是基于 ImageNet 数据集构建的。包括 34 个大类, 608 个小类, 每个小类有近 1 200 幅

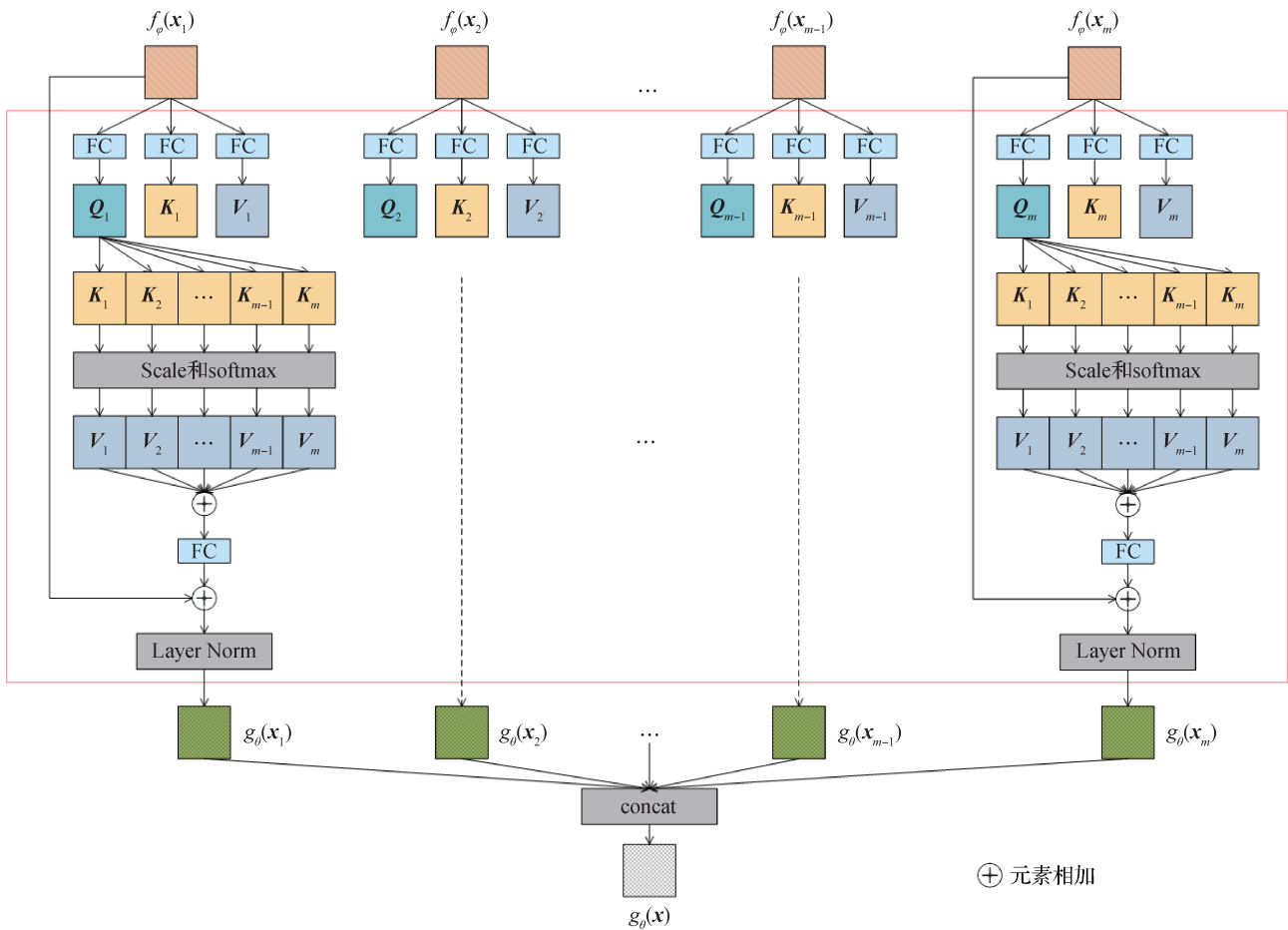


图2 基于Transformer架构的局部特征融合模块

Fig. 2 Local feature fusion module based on Transformer architecture

图像,共779 165幅图像。与MiniImageNet相比,该数据集扩大了训练和测试阶段的领域差异。遵循Hou等人(2019)方法中的划分,将608个小类分为351、97和160个类,分别用于训练、验证和测试。

CUB数据集最初被提出是用于细粒度的鸟类分类。包含200个类别,共11 788幅图像。遵循Ye等人(2018)方法中的划分,将200个类分为100、50和50个类,分别用于训练、验证和测试。此数据集分类的难点在于类之间的细微差别。

2.1.2 训练与评估

基于度量学习方法(Vinyals等,2016;Snell等,2017)中的特征提取网络较为简单,通常是4个卷积层的堆叠,难以很好地学习到复杂数据集的非线性表示。为了与当前先进的方法进行公平比较,本文采用常用的ResNet12(residual network 12)(Munkhdalai和Yu,2017)作为主干网络进行特征提取。由4个深度为3的残差块构成,每个块的内部有3个3×3内核的卷积层,在每个块的尾部应用2×2

的最大池化层。为了生成密集特征,在主干网络末端添加1个全局平均池化层,以使每个输入图像块生成1个一维向量。

整个训练过程分为预训练和元训练两个阶段。在预训练阶段,遵循Ye等人(2020)方法中的预训练策略,用主干网络附加softmax层来对训练集(例如MiniImageNet中的64个类)进行分类。对于预训练的验证,采用小样本模式对验证集进行分类。具体地,随机抽样200个N-way 1-shot的小样本学习任务(N等于验证集中的类别数,例如MiniImageNet中的16个类),每类中抽取1个和15个实例分别作为支持集和查询集来进行分类验证。在这一阶段,应用数据增强方法来提高模型的泛化能力,包括色彩抖动、随机裁剪和随机水平翻转。在元训练过程中,将模型中的主干网络使用预先训练好的权重进行初始化,然后与模型中的其他模块一起进行元训练的微调,具体采用情景式训练策略进行N-way K-shot的小样本分类任务。使用SGD优化器来优化网络,初始

学习率设置为 5×10^{-4} , 共设 100 个 epoch, 每 10 个 epoch 后学习率衰减一半, 每个 epoch 中分别设 100 个和 600 个 episodes 用于训练和验证。由于 TieredImageNet 数据集样本更多且领域差异更大, 需要更多的迭代次数才能使模型收敛, 所以设置 200 个 epoch, 每 20 个 epoch 后学习率衰减一半。

在测试阶段, 从测试集中随机抽取 5 000 个 episodes 任务, 评估其分类的平均准确率, 置信区间为 95%。在训练、验证及测试阶段, 无论输入是整幅图像还是局部图像块, 都先调整尺寸为 84×84 像素, 再输入到主干网络中提取特征。另外, 如 Oreshkin 等人(2018)所建议的, 在获得两个样本特征表示之间的距离后, 乘以一个自适应缩放因子, 以使输出的距离在合适的范围内, 可增强模型的鲁棒性。

2.2 小样本分类结果

在各项实验中, 首先将本文方法与当前先进的小样本分类方法的平均准确率进行比较。在 Mini-

ImageNet、TieredImageNet 以及 CUB 数据集上的 5-way 1-shot 和 5-way 5-shot 分类任务的结果比较如表 1 所示。表 1 中, 本文方法采用图像金字塔结构, $P_{(1, 2, 3)}$ 表示将原图分为 1×1 、 2×2 和 3×3 共 14 个局部块一起作为输入。从表 1 可以看出, 在 3 个小样本分类数据集上, 所提方法都展现出优异性能。在 5-way 1-shot 和 5-way 5-shot 的小样本分类任务上, 对于 MiniImageNet 数据集, 所提方法的结果比次优结果分别提高了 2.96% 和 2.9%; 对于 CUB 数据集, 所提方法的结果比次优的结果分别提高了 3.22% 和 1.77%; 在 TieredImageNet 数据集上, 所提方法的结果与最优的结果相当。

2.3 消融实验

为了验证网络模型中各模块的有效性以及选取最优超参数, 在 MiniImageNet 数据集上以 ResNet12 为特征提取网络进行消融实验。

2.3.1 局部特征融合模块对结果的影响

在有预训练的情况下, 分别将原图分为 1×1 、

表 1 分类结果比较
Table 1 Comparison of classification results

方法	骨干网络	/%					
		MiniImageNet		TieredImageNet		CUB	
		1-shot	5-shot	1-shot	5-shot	1-shot	5-shot
MatchingNet(Neurlps)(Vinyal等, 2016)	ResNet12	63.08 ± 0.80	75.99 ± 0.60	68.50 ± 0.92	80.60 ± 0.71	71.87 ± 0.85	85.08 ± 0.57
ProtoNet(Neurlps)(Snell等, 2017)	ResNet12	60.37 ± 0.83	78.02 ± 0.57	65.65 ± 0.92	83.40 ± 0.65	66.09 ± 0.92	82.50 ± 0.58
TADAM(Neurlps)(Oreshkin等, 2018)	ResNet12	58.50 ± 0.30	76.70 ± 0.30	—	—	—	—
AM3(Neurlps)(Xing等, 2019)	ResNet12	65.30 ± 0.49	78.10 ± 0.36	69.08 ± 0.47	82.58 ± 0.31	—	—
CAN(Neurlps)(Hou等, 2019)	ResNet12	63.85 ± 0.48	79.44 ± 0.34	69.89 ± 0.51	84.23 ± 0.37	—	—
LEO(ICLR)(Rusu等, 2019)	WRN-28-10 [†]	61.76 ± 0.08	77.59 ± 0.12	66.33 ± 0.05	81.44 ± 0.09	—	—
MetaOptNet(CVPR)(Lee等, 2019)	ResNet12	62.64 ± 0.61	78.63 ± 0.46	65.99 ± 0.72	81.56 ± 0.53	—	—
CTM(CVPR)(Li等, 2019b)	ResNet18 [†]	64.12 ± 0.82	80.51 ± 0.13	68.41 ± 0.39	84.28 ± 1.73	—	—
DeepEMD(CVPR)(Zhang等, 2020)	ResNet12	65.91 ± 0.82	<u>82.41 ± 0.56</u>	71.16 ± 0.87	86.03 ± 0.58	<u>75.65 ± 0.83</u>	88.69 ± 0.50
AFHN(CVPR)(Li等, 2020)	ResNet18 [†]	62.38 ± 0.72	78.16 ± 0.56	—	—	70.53 ± 1.01	83.95 ± 0.63
Neg-cosine(ECCV)(Liu等, 2020)	ResNet18 [†]	62.33 ± 0.82	80.94 ± 0.59	—	—	72.66 ± 0.85	<u>89.40 ± 0.43</u>
PSST(CVPR)(Chen等, 2021a)	ResNet12	64.05 ± 0.49	80.24 ± 0.45	—	—	—	—
P-Transfer(AAAI)(Shen等, 2021)	ResNet12	64.21 ± 0.77	80.38 ± 0.59	—	—	73.88 ± 0.87	87.81 ± 0.48
GLoFA(AAAI)(Lu等, 2021)	ResNet12	<u>66.12 ± 0.42</u>	81.37 ± 0.33	69.75 ± 0.33	83.58 ± 0.42	—	—
Meta-Baseline(ICCV)(Chen等, 2021b)	ResNet12	63.17 ± 0.23	79.26 ± 0.17	68.62 ± 0.27	83.74 ± 0.18	—	—
本文($P_{(1, 2, 3)}$)	ResNet12	69.08 ± 0.28	85.31 ± 0.18	<u>70.15 ± 0.32</u>	<u>85.24 ± 0.22</u>	78.87 ± 0.28	91.17 ± 0.15

注: 加粗字体表示各列最优结果, 下划线字体表示各列次优结果, “—”表示相关文献没有实验结果, 骨干网络中“†”表示比 ResNet12 更深的网络结构, 方法中具体方法后标注了文献刊发的会议。

2×2 、 3×3 和 4×4 的图像块, $P_{H \times W}$ 表示按网格将原图分为 $H \times W$ 个局部块。由于 $P_{1 \times 1}$ 时只有 1 个图像块, 仅需对这个图像块进行 CNN 特征提取而无需进行融合处理。“ $P_{H \times W}$ +无融合”表示 $H \times W$ 个块图像分别经 CNN 特征提取后, 将得到的 $H \times W$ 个局部特征矢量进行连接处理作为度量模块的输入。“ $P_{H \times W}$ +融合”表示 $H \times W$ 个块图像分别经 CNN 特征提取并基于 Transformer 架构的局部特征融合模块后, 将得到的 $H \times W$ 个局部特征矢量进行连接处理作为度量模块的输入。表 2 给出了有/无局部特征融合模块的分类结果。可以看出, 将图像分为 $H \times W$ 的局部块, 经基于 Transformer 架构的局部特征融合模块后的网络分类性能优于不经此模块的分类结果, 验证了嵌入该局部特征融合模块的有效性。

表 2 在 MiniImageNet 数据集上有/无局部特征融合模块的分类结果
Table 2 Classification results with/without local feature fusion module on MiniImageNet dataset

/%		
方法	1-shot	5-shot
$P_{1 \times 1}$	61.91 ± 0.29	80.49 ± 0.19
$P_{2 \times 2}$ + 无融合	62.90 ± 0.29	81.91 ± 0.19
$P_{3 \times 3}$ + 无融合	64.90 ± 0.29	82.75 ± 0.20
$P_{4 \times 4}$ + 无融合	61.62 ± 0.29	78.25 ± 0.21
$P_{2 \times 2}$ + 融合	67.27 ± 0.29	83.54 ± 0.19
$P_{3 \times 3}$ + 融合	68.14 ± 0.28	84.62 ± 0.18
$P_{4 \times 4}$ + 融合	67.77 ± 0.27	83.92 ± 0.19

此外, 随着分块数量的增加, 即将图像分为 2×2 和 3×3 的局部块时, 小样本分类性能均比不分块的结果要好, 且分类性能逐步增加, 但是当图像分为 4×4 的局部块时, 分类性能开始有所下降。

通过分块处理虽然可以有效利用图像的局部特征, 但是当分块数量过多时, 一个完整的目标可能被分离到多个局部块中, 目标的完整性被破坏反而对分类造成不利影响。因此需要根据分类结果对分块数量进行权衡, 图像分块并不是越多越好。

2.3.2 多尺度局部分块对结果的影响

除了将图像进行固定大小的网格分块, 还利用具有金字塔结构的多尺度相结合的局部分块进行实验, 分类结果如表 3 所示。表 3 中, $P_{(\cdot)}$ 表示多种尺度分块组合, 如 $P_{(1,2,3)}$ 表示将原图分为 1×1 、 2×2 和

3×3 共 14 个局部块一起作为输入。从表 3 可以看到, 多尺度局部特征的分类结果优于单一尺度下的分类结果。例如, 在 5-way 1-shot 和 5-way 5-shot 的小样本分类任务上, 使用 1×1 和 3×3 分块相结合的分类结果比仅使用 3×3 分块的结果分别提高了 0.45% 和 0.6%。为了能充分利用局部特征, 表 1 中均采用 $P_{(1,2,3)}$ 金字塔结构的多尺度相结合的局部分块进行实验并作为最终分类结果。

表 3 在 MiniImageNet 数据集上多尺度局部分块的结果
Table 3 Classification results of multi-scale local blocks on MiniImageNet dataset

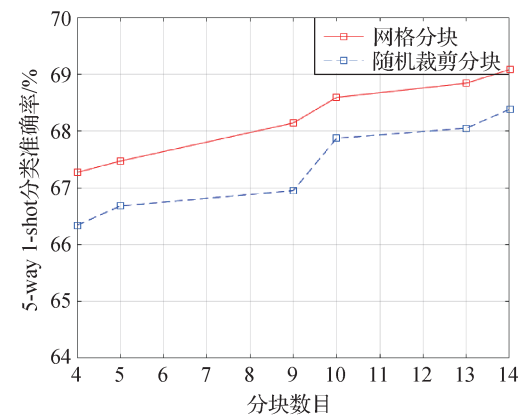
/%		
分块	1-shot	5-shot
$P_{2 \times 2}$	67.27 ± 0.29	83.54 ± 0.19
$P_{3 \times 3}$	68.14 ± 0.28	84.62 ± 0.18
$P_{(1,2)}$	67.47 ± 0.29	83.96 ± 0.18
$P_{(1,3)}$	68.59 ± 0.28	85.22 ± 0.18
$P_{(2,3)}$	68.84 ± 0.28	85.17 ± 0.18
$P_{(1,2,3)}$	69.08 ± 0.28	85.31 ± 0.18

2.3.3 局部分块方式对结果的影响

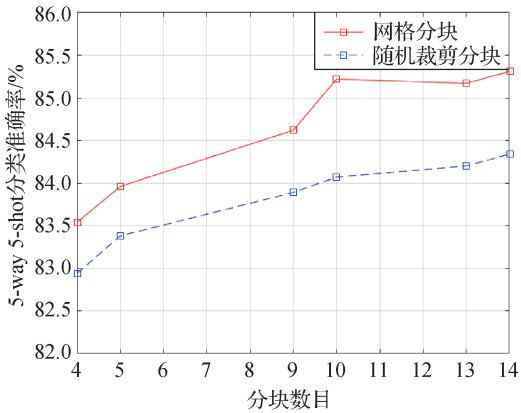
为了验证本文网格局部分块方式的合理性, 与采用随机裁剪的局部分块方式进行对比实验, 相同局部图像块数量下的对比结果如图 3 所示。可以看出, 两种方法均随着局部块数量的增加, 分类结果逐渐提升。但相同数量的局部块下, 网格分块方法的性能要比随机裁剪的效果要好。这是由于网格分块的方法中, 多局部块图像拼接起来可以包含整幅图像, 然后通过局部特征融合模块使每个局部特征都拥有了全局性信息。而随机裁剪的方式只能随机取到原图上的部分局部块, 即使通过特征融合也可能无法获得完全全局性的信息, 所以其分类性能要弱于网格分块的方法。

2.3.4 局部特征融合模块超参数的选取

在模型的局部特征融合模块中, 使用了一个浅层且简单的 Transformer 结构, 只有 1 个层 (layer) 和 1 个注意力头 (head)。而原始 Transformer (Vaswani 等, 2017) 具有更复杂的结构, 使用更多的堆叠层和更多的注意力头。为了评估 Transformer 结构超参数设置对分类结果的影响, 对网络分别使用不同层数



(a) 5-way 1-shot分类结果



(b) 5-way 5-shot分类结果

图3 采用网格分块和随机裁剪分块两种方式在 MiniImageNet 数据集上的分类结果

Fig. 3 Classification results of tasks on MiniImageNet dataset by grid partitioning and random sampling partitioning
(a) 5-way 1-shot classification results;
(b) 5-way 5-shot classification results)

表4 在 MiniImageNet 数据集上局部特征融合模块中不同层数的分类结果

Table 4 Classification results of different layer numbers in local feature fusion module on MiniImageNet dataset

层数	注意力头数	参数量/M	1-shot/%	5-shot/%
1	1	14.07	67.27 ± 0.29	83.54 ± 0.19
2	1	15.71	67.44 ± 0.29	83.53 ± 0.19
3	1	17.35	67.08 ± 0.29	83.20 ± 0.19

和不同注意力头数进行实验,表4为模型的注意力头数固定为1时不同层数的分类结果,表5为模型的层数固定为1时不同注意力头数的分类结果。

从表4可以看出,固定注意力头数不变、改变Transformer的层数时,当层数逐渐增加到3,分类性能略有下降。说明更复杂的结构并不会提高分类性能,而且随着自注意力层数的增加,网络参数增多,

优化的难度增加,训练模型收敛变得更加困难。从表5可以看出,固定层数不变、改变Transformer的注意力头数时,模型的分类性能几乎不变,这说明使用1个注意力头就可以关注到特征的全局信息,而不需要使用多个头从多个角度来进行融合处理。因此,在模型设计中,本文使用1个层(layer)和1个注意力头(head)的Transformer结构。

表5 在 MiniImageNet 数据集上局部特征融合模块中不同注意力头数的分类结果

Table 5 Classification results of different head numbers in local feature fusion module on MiniImageNet dataset

注意力头数	层数	参数量/M	1-shot/%	5-shot/%
1	1	14.07	67.27 ± 0.29	83.54 ± 0.19
5	1	14.07	67.26 ± 0.29	83.47 ± 0.19
10	1	14.07	67.26 ± 0.29	83.48 ± 0.19

2.3.5 与其他小样本分类方法结合

提出的基于Transformer架构的局部特征融合模块可看做一种图像CNN特征的增强表示方法,因此可与其他小样本分类方法结合。具体地,将所提方法与度量学习中另一经典方法——匹配网络进行结合,即通过将匹配网络中的特征提取网络替换为所提的基于局部特征融合的特征提取方法,不改变原有匹配网络中的度量方式进行结合,在 MiniImageNet 数据集上的分类结果如表6所示。

表6 在 MiniImageNet 数据集上与匹配网络结合的分类结果

Table 6 Classification results on the MiniImageNet dataset combined with the matching network

方法	1-shot	5-shot
MatchingNet	63.08 ± 0.80	75.99 ± 0.60
MatchingNet_Patch _{1x1}	64.17 ± 0.29	77.28 ± 0.20
MatchingNet_Patch _{2x2}	66.46 ± 0.28	80.30 ± 0.21

首先,用本文的预训练模型重新复现匹配网络结果,即表6中的 MatchingNet_Patch_{1x1},然后将所提局部特征融合模块加入到匹配网络中,对输入图像分为4块进行融合,结果如 MatchingNet_Patch_{2x2}所示。可以看出,在 MiniImageNet 数据集上,将所提模

块嵌入到匹配网络中的分类结果相比原匹配网络的结果有一定提升,表明了所提模块与其他小样本分类方法相结合的有效性。

2.3.6 可视化分析

为了进一步说明所提基于Transformer局部特征融合模块的有效性,对经该模块后生成的特征使用 t -SNE 方法(van der Maaten 和 Hinton, 2008)进行可视

化分析。具体地,在 MiniImageNet 数据集的测试集中,随机选取5类,每类随机选择75幅图像,提取特征的 t -SNE 可视化结果如图4所示。从图4可视化的结果可以看出,与原型网络相比,所提出的基于Transformer局部特征融合方法产生的特征具有更好的聚类性能,类内距离更小,类间距离更大,类间可区分性得到了提升。

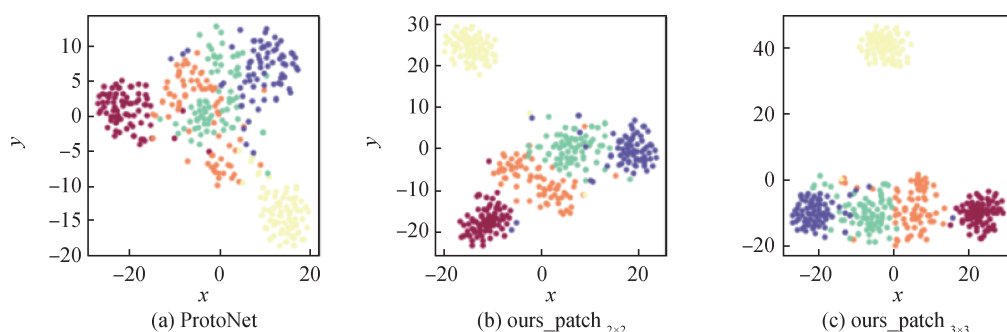


图4 MiniImageNet数据集的测试集中随机选取5类每类75幅图像特征的 t -SNE 可视化

Fig. 4 The t -SNE visualization of features for 75 randomly sampled images from 5 randomly selected test classes of MiniImageNet dataset((a)ProtoNet; (b)ours_patch_{2x2}; (c)ours_patch_{3x3})

3 结 论

为了在小样本图像识别中充分利用图像的局部特征以及增强模型的泛化能力,提出一种基于局部特征融合的小样本分类方法。通过将原图网格分块的方式来获得局部特征,通过基于Transformer架构的局部特征融合模块得到包含全局信息的局部增强特征,从而提高了模型的特征提取能力和泛化能力。与原型网络相比,所提方法在 MiniImageNet、Tiered-ImageNet 和 CUB 小样本数据集上都具有更高的分类准确率,消融实验也验证了所提方法的有效性。提出的基于Transformer架构的局部特征融合模块可视为一种图像特征的增强表示方法,简洁且通用,可以嵌入到其他小样本分类方法中。

然而,本文方法也存在一定局限性。对于图像的局部分块,随着分块数目的增加,计算量也随之增大,这是由于增加分块数目相当于增加 batch size 的大小,因此在图像分块时,要权衡分类性能与计算负担选择合适的分块数目。后续工作将探索如何更有效地从多个局部块中进行选择来提升特征融合模块的性能,以及如何与半监督、转导式小样本设置结合起来进一步提高模型的性能。

参考文献 (References)

- Antoniou A, Edwards H and Storkey A. 2019. How to train your MAM [EB/OL]. [2022-01-12]. <https://arxiv.org/pdf/1810.09502.pdf>
- Chen Y B, Liu Z, Xu H J, Darrell T and Wang X L. 2021b. Meta-baseline: exploring simple meta-learning for few-shot learning [EB/OL]. [2022-01-12]. <https://arxiv.org/pdf/2003.04390.pdf>
- Chen Z Y, Ge J X, Zhan H S, Huang S T and Wang D L. 2021a. Pareto self-supervised training for few-shot learning [EB/OL]. [2022-01-12]. <https://arxiv.org/pdf/2104.07841.pdf>
- Finn C, Abbeel P and Levine S. 2017. Model-agnostic meta-learning for fast adaptation of deep networks//Proceedings of the 34th International Conference on Machine Learning. Sydney, Australia: JMLR. org: 1126-1135
- Gu J X, Wang Z H, Kuen J, Ma L Y, Shahroudy A, Shuai B, Liu T, Wang X X, Wang G, Cai J F and Chen T. 2018. Recent advances in convolutional neural networks. Pattern Recognition, 77: 354-377 [DOI: 10.1016/j.patcog.2017.10.013]
- He K M, Zhang X Y, Ren S Q and Sun J. 2016. Deep residual learning for image recognition//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE: 770-778 [DOI: 10.1109/CVPR.2016.90]
- Hou R B, Chang H, Ma B P, Shan S G and Chen X L. 2019. Cross attention network for few-shot classification [EB/OL]. [2022-01-12]. <https://arxiv.org/pdf/1910.07677.pdf>
- Jamal M A and Qi G J. 2019. Task agnostic meta-learning for few-shot

- learning//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 11711-11719 [DOI: 10.1109/CVPR.2019.01199]
- Krizhevsky A, Sutskever I and Hinton G E. 2017. ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6): 84-90 [DOI: 10.1145/3065386]
- Lake B M, Salakhutdinov R and Tenenbaum J B. 2013. One-shot learning by inverting a compositional causal process//Proceedings of the 26th International Conference on Neural Information Processing Systems. Lake Tahoe, Nevada: Curran Associates Inc.: 2526-2534
- Lee K, Maji S, Ravichandran A and Soatto S. 2019. Meta-learning with differentiable convex optimization//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 10649-10657 [DOI: 10.1109/CVPR.2019.01091]
- Li H Y, Eigen D, Dodge S, Zeiler M and Wang X G. 2019b. Finding task-relevant features for few-shot learning by category traversal//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 1-10 [DOI: 10.1109/CVPR.2019.00009]
- Li K, Zhang Y L, Li K P and Fu Y. 2020. Adversarial feature hallucination networks for few-shot learning//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 13467-13476 [DOI: 10.1109/CVPR42600.2020.01348]
- Li W B, Wang L, Xu J L, Huo J, Gao Y and Luo J B. 2019a. Revisiting local descriptor based image-to-class measure for few-shot learning//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 7253-7260 [DOI: 10.1109/CVPR.2019.00743]
- Lifchitz Y, Avrithis Y, Picard S and Bursuc A. 2019. Dense classification and implanting for few-shot learning//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 9250-9259 [DOI: 10.1109/CVPR.2019.00948]
- Liu B, Cao Y, Lin Y T, Li Q, Zhang Z, Long M S and Hu H. 2020. Negative margin matters: understanding margin in few-shot classification//Proceedings of the 16th European Conference on Computer Vision. Glasgow, UK: Springer: 438-455 [DOI: 10.1007/978-3-030-58548-8_26]
- Lu S, Ye H J and Zhan D C. 2021. Tailoring embedding function to heterogeneous few-shot tasks by global and local feature adaptors. *Proceedings of 2021 AAAI Conference on Artificial Intelligence*, 35(10): 8776-8783 [DOI: 10.1609/aaai.v35i10.17063]
- Munkhdalai T and Yu H. 2017. Meta networks//Proceedings of the 34th International Conference on Machine Learning. Sydney, Australia: JMLR.org: 2554-2563
- Oreshkin B N, Rodriguez P and Lacoste A. 2018. TADAM: task dependent adaptive metric for improved few-shot learning//Proceedings of the 32nd International Conference on Neural Information Processing Systems. Montréal, Canada: Curran Associates Inc.: 719-729
- Ravi S and Larochelle H. 2017. Optimization as a model for few-shot learning//Proceedings of the 5th International Conference on Learning Representations. Toulon, France: ICLR: 1-11
- Ren M Y, Triantafillou E, Ravi S, Snell J, Swersky K, Tenenbaum J B, Larochelle H and Zemel R S. 2018. Meta-learning for semi-supervised few-shot classification [EB/OL]. [2022-01-12]. <https://arxiv.org/pdf/1803.00676.pdf>
- Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S A, Huang Z H, Karpathy A, Khosla A, Bernstein M, Berg A C and Li F F. 2015. ImageNet large scale visual recognition challenge. *International Journal of Computer Vision*, 115(3): 211-252 [DOI: 10.1007/s11263-015-0816-y]
- Rusu A A, Rao D, Sygnowski J, Vinyals O, Pascanu R, Osindero S and Hadsell R. 2019. Meta-learning with latent embedding optimization [EB/OL]. [2022-01-12]. <https://arxiv.org/pdf/1807.05960.pdf>
- Santoro A, Bartunov S, Botvinick M, Wierstra D and Lillicrap T. 2016. Meta-learning with memory-augmented neural networks//Proceedings of the 33rd International Conference on Machine Learning. New York, USA: JMLR.org: 1842-1850
- Shen Z Q, Liu Z C, Qin J, Savvides M and Cheng K T. 2021. Partial is better than all: revisiting fine-tuning strategy for few-shot learning [EB/OL]. [2022-01-12]. <https://arxiv.org/pdf/2102.03983.pdf>
- Snell J, Swersky K and Zemel R. 2017. Prototypical networks for few-shot learning//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, USA: Curran Associates Inc.: 4080-4090
- Sun Q R, Liu Y Y, Chua T S and Schiele B. 2019. Meta-transfer learning for few-shot learning//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 403-412 [DOI: 10.1109/CVPR.2019.00049]
- Sung F, Yang Y X, Zhang L, Xiang T, Torr P H S and Hospedales T M. 2018. Learning to compare: relation network for few-shot learning//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake, USA: IEEE: 1199-1208 [DOI: 10.1109/CVPR.2018.00131]
- Thrun S and Pratt L. 1998. Learning to learn: introduction and overview//Thrun S and Pratt L, eds. *Learning to Learn*. New York, USA: Springer: 3-17 [DOI: 10.1007/978-1-4615-5529-2_1]
- van der Maaten L and Hinton G. 2008. Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9(86): 2579-2605
- Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, Kaiser Ł and Polosukhin I. 2017. Attention is all you need//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, USA: Curran Associates Inc.: 6000-6010
- Vilalta R and Drissi Y. 2002. A perspective view and survey of meta-

- learning. *Artificial Intelligence Review*, 18(2): 77-95 [DOI: 10.1023/A:1019956318069]
- Vinyals O, Blundell C, Lillicrap T, Kavukcuoglu K and Wierstra D. 2016. Matching networks for one shot learning//*Proceedings of the 30th International Conference on Neural Information Processing Systems*. Barcelona, Spain: Curran Associates Inc.: 3637-3645
- Wah C, Branson S, Welinder P, Perona P and Belongie S. 2011. The caltech-UCSD birds-200-2011 dataset. *Computation and Neural Systems Technical Report*
- Wang Y Q, Yao Q M, Kwok J T and Ni L M. 2020. Generalizing from a few examples: a survey on few-shot learning. *ACM Computing Surveys*, 53(3): #63 [DOI: 10.1145/3386252]
- Xing C, Rostamzadeh N, Oreshkin B N and Pinheiro P O. 2019. Adaptive cross-modal few-shot learning//*Proceedings of the 33rd International Conference on Neural Information Processing Systems*. Vancouver, Canada: NIPS: 4847-4857
- Ye H J, Hu H X, Zhan D C and Sha F. 2018. Learning embedding adaptation for few-shot learning [EB/OL]. [2022-01-12]. <https://arxiv.org/pdf/1812.03664v1.pdf>
- Ye H J, Hu H X, Zhan D C and Sha F. 2020. Few-shot learning via embedding adaptation with set-to-set functions//*Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle, USA: IEEE: 8805-8814 [DOI: 10.1109/CVPR42600.2020.00883]
- Zhang C, Cai Y J, Lin G S and Shen C H. 2020. DeepEMD: few-shot image classification with differentiable earth mover's distance and structured classifiers//*Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle, USA: IEEE: 12200-12210 [DOI: 10.1109/CVPR42600.2020.01222]

作者简介

董杨洋,女,硕士研究生,主要研究方向为小样本学习。

E-mail: 1694945344@qq.com

宋蓓蓓,通信作者,女,副教授,主要研究方向为图像处理与人工智能。E-mail: bbsong@chd.edu.cn

孙文方,男,副教授,主要研究方向为遥感图像处理、人工智能和卫星系统工程。E-mail: wfsun@xidian.edu.cn