

DOI:10.13232/j.cnki.jnju.2022.06.014

基于张量特征的小样本图像快速分类方法

张艳莎^{1,2}, 冯夫健^{1,2}, 王 杰^{1,2}, 潘 凤^{1,2}, 谭 棉^{1,2}, 张再军³, 王 林^{2*}

(1. 贵州民族大学数据科学与信息工程学院, 贵阳, 550025; 2. 贵州省模式识别与智能系统重点实验室, 贵州民族大学, 贵阳, 550025; 3. 黔南民族师范学院数学与统计学院, 都匀, 558000)

摘 要: 解决小样本图像分类问题最直接的方式是进行数据增强, 但目前适用于小样本图像分类的数据增强方法大都存在模型复杂、推理时间长的问题。提出一个张量特征生成器, 通过生成新的张量特征在特征空间对小样本图像进行数据增强。基于张量特征生成器, 提出一个适用于小样本图像的快速分类方法(Tensor Feature-based Faster Classification Network, TFFCN), 该方法网络结构简单, 利用残差网络提取图像的张量特征, 通过张量特征生成器对小样本图像进行数据增强, 从而训练得到一个满意的分类器对查询集图像进行分类, 解决了模型推理时间长的问题。为了验证提出模型的有效性, 选用公开数据集 miniImageNet, CUB 以及 CIFAR-FS, 对分类性能和推理时间进行对比实验。实验结果表明, TFFCN 的分类性能优于目前流行的数据增强方法, 并且, 和改进前的模型相比能有效减少模型的推理时间, 采用 ResNet18 和 ResNet12 为主干特征提取网络时, 随着生成的张量特征数量的增加, 最高可减少 49% 和 24% 的推理时间, 能更快速地完成小样本图像分类任务。

关键词: 小样本图像分类, 数据增强, 张量特征生成器, 张量特征, 推理时间

中图分类号: TP391

文献标志码: A

Tensor feature-based faster classification network for few-shot learning

Zhang Yansha^{1,2}, Feng Fujian^{1,2}, Wang Jie^{1,2}, Pan Feng^{1,2}, Tan Mian^{1,2}, Zhang Zaijun³, Wang Lin^{2*}

(1. School of Data Science and Information Engineering, Guizhou Minzu University, Guiyang, 550025, China;

2. Key Laboratory of Pattern Recognition and Intelligent System, Guizhou Minzu University, Guiyang, 550025, China;

3. School of Mathematics and Statistics, Qiannan Normal University for Nationalities, Duyun, 558000, China)

Abstract: The most straightforward way to solve the problem of few-shot image classification is data augmentation. Aiming at the problem that most of the data augmentation methods suitable for few-shot image classification are overly sophisticated and need a long inference time. We propose a tensor feature generator, which augments few-shot images in the features space by generating new tensor features. Based on the tensor feature generator, a rapid classification method for few-shot learning, Tensor Feature-based Faster Classification Network (TFFCN) is proposed. The network structure is simple, the tensor features of the image are extracted by using the residual network, and data augmentation of few-shot image with tensor feature generator, so as to train a satisfactory classifier and classify the query set images, and solve the problem of long inference time. We use public datasets miniImageNet, CUB and CIFAR-FS to verify the effectiveness of the proposed model by comparing experimental results on classification performance and inference time. Experimental results show that the classification

基金项目: 国家自然科学基金(62162012), 贵州省科技支撑计划(黔科合支撑[2021]一般531), 贵州省教育厅自然科学研究项目(黔教技[2022]015号), 贵州省科技计划(黔科合基础-ZK[2022]一般195, 黔科合基础-ZK[2022]一般550), 贵州省教育厅深化新时代教育评价改革试点项目(教学过程质量评价), 贵州省教育厅青年科技人才成长项目(黔教合KY字[2022]177号, 黔教合KY字[2021]104号, 黔教合KY字[2018]141号, 黔教合KY字[2018]140, 黔教合KY字[2021]110)

收稿日期: 2022-09-14

* 通讯联系人, E-mail: wanglin@gzmu.edu.cn

performance of the TFFCN is better than popular data augmentation methods for few-shot image classification, and the inference time is greatly reduced compared to the model before the improvement. When the backbone networks are ResNet18 and ResNet12, the inference time is almost reduced by up to 49% and 24% with the increase of the number of generated tensor features, respectively, and the TFFCN can complete the task of few-shot image classification more quickly.

Key words: few-shot image classification, data augmentation, tensor feature generator, tensor features, inference time

以深度神经网络为基础的图像分类算法通常需要大量的标注数据^[1],而很多现实场景存在局限性和特殊性,往往很难获得大量数据或大量标注数据,即使拥有足够的样本数据,对大规模的无标注样本数据进行标注也需要投入巨大的标注成本,比如一些濒危物种的图像、需要侦查的军事目标图像、危险工业产生的图像以及特殊病例的医用图像等.相反,与需要大量数据才能学习得到较好结果的神经网络模型相比,人类仅通过少量数据就能做到快速学习,受此启发,小样本学习(Few-Shot Learning)^[2-3]的概念被提出.近年来,为了解决在数据稀缺情况下的图像分类问题,小样本图像分类作为小样本学习中的分支已成为计算机视觉及其他领域讨论的热点话题.

小样本图像分类的限制在于样本量少或有监督的样本量少,基于小样本学习得到的深度神经网络分类模型容易出现过拟合的问题^[4],解决它最直接的方法是数据增强.目前已有大量学者从数据增强的角度去研究小样本图像分类问题,主要通过提出一个生成模型为小样本数据生成合成数据^[5-9]或幻觉^[10-14]新的样本数据,再利用生成的样本数据去增强原有的小样本数据,从而训练得到一个满意的分类器.例如,文献[5-8,12]运用生成对抗网络(Generative Adversarial Network, GAN)作为生成器为小样本数据生成新的样本数据,虽然生成的样本数据比较真实,但Wu et al^[13]认为基于GAN合成样本数据的方法过于复杂,因为GAN方法难以训练.虽然Chen et al^[9]和Wu et al^[13]没有运用GAN来得到生成数据和幻觉数据,但两者都需要对数据进行复杂的处理,前者需要输入一对相似的探针图像(probe image)和图库图像(gallery image),后者为了使数据服从高斯分布,需要对数据进行能量变换(power transformation).Lazarou et al^[15]提出的张量特征幻觉器

(Tensor Feature Hallucination, TFH)虽然简单,但是该模型在生成张量特征的过程中经过了张量特征变为向量特征再生成张量特征的过程,增加了模型的规模和完成小样本图像分类测试任务的时间(推理时间).

尽管各种适用于小样本学习的生成模型被提出,但大都存在模型复杂、推理时间长的问题.针对这个问题,本文在TFH^[15]的基础上提出一个基于张量特征的小样本图像快速分类方法(Tensor Feature - based Faster Classification Network, TFFCN).该方法属于利用数据增强来解决小样本图像分类的范畴,利用残差网络提取图像张量特征,通过张量特征生成器对小样本图像进行数据增强,使用增强后的张量特征对分类器进行训练,从而得到一个满意的小样本图像分类器.假设张量特征生成器直接利用张量特征生成新的张量特征,能缩减模型规模,使模型在分类准确率损失较小的情况下能更快地完成小样本图像分类任务.

本文的主要贡献:

(1)提出一个直接利用张量特征生成新的张量特征的生成器,网络结构简单.张量特征包含图像丰富的空间和结构信息,直接利用张量特征生成新的张量特征用于小样本图像特征的数据增强,不再需要从张量特征转换为向量特征再生成张量特征,使网络结构更简单.

(2)基于张量特征生成器提出TFFCN模型,在分类准确率损失较小的情况下,大幅减少模型的推理时间.在miniImageNet数据集上的实验结果表明,TFFCN模型能更快地完成小样本图像的分类任务.

1 相关工作

基于数据增强的方法在小样本图像分类中被

广泛使用,并且绝大多数基于数据增强的小样本图像分类方法都在元学习框架下进行评估.下面对元学习框架下小样本图像分类评估模式以及基于数据增强的小样本图像分类方法作简要概述.

1.1 元学习框架下的小样本图像分类 小样本学习的目标是从少量的样本数据中学习解决问题的模型.目前绝大多数的小样本图像分类方法主要基于元学习框架.在小样本学习中,元学习指从大量的先验任务中学习元知识,利用以往的先验知识来指导模型在新的任务中,即小样本任务中更快地学习^[16].在元学习框架下,小样本学习的数据集通常被分为不相交的元训练集和元测试集,两者都由大量任务组成,每个任务都包含训练数据和测试数据,为了便于区分又叫支持集和查询集.元学习框架下小样本学习的数据集划分如图1所示,它表示5-way 1-shot模式,其中一个横向的支持集和查询集就代表一个小样本图像分类任务.通常用 N -way K -shot的模式来对小样本图像分类算法进行实验评估, N -way表示在元测试集中,每个任务的支持集都有 N 个类别, K -shot表示每一类支持集有 K 个标注样本,进行准确率测试时重复数次小样本图像分类任务取平均值.



图1 小样本数据集的划分

Fig. 1 Division of few-shot datasets

1.2 基于数据增强的小样本图像分类 小样本学习的根本问题在于样本量太少,导致样本的多样性变低.面对小样本学习问题时深度学习模型会出现性能差和过拟合问题,解决小样本学习问题最直接的方法是进行数据增强来提高样本的多样性,传统的数据增强方法,如移位(translation)、旋转(rotation)、缩放(scale)、裁剪(crop)、翻转

(flip)等转换规则不具有一般性,当数据集图像为不规则形状时表现效果不佳.比如,当一个数据集的图像不是正方形时,采用旋转可能丢失图像尺寸,这一问题又需要额外的工作来解决.因此,采用传统的数据增强规则的方式不能很好地解决小样本图像分类问题^[17].

适用于小样本图像分类的数据增强方法大都借助辅助数据或辅助信息对原有的小样本数据集进行数据扩充或特征增强,这对小样本图像分类器的准确性起到至关重要的作用.目前,从数据增强的角度出发解决小样本图像分类问题已引起国内外学者的关注,GAN (Generative Adversarial Networks)能生成与真实样本分布相同的模拟样本,作为生成器被广泛应用于小样本图像分类. Mehrotra and Dukkipati^[5]利用GAN生成器为输入的支持集样本生成假的查询样本,提出一种用于单样本学习的生成对抗残差成对网络. Antoniou et al^[6]也在GAN网络的基础上为小样本学习问题提出数据增强生成对抗网络(Data Augmentation GAN, DAGAN). Zhang et al^[7]提出 Meta-GAN,进一步为特定任务生成虚假样本以补充训练样本,在Meta-GAN中,GAN和少数分类网络的组合训练使生成的示例更适合于小样本学习任务. Zhang et al^[12]提出基于条件 Wasserstein 生成对抗网络(cWGAN)的对抗性特征幻觉网络(Adversarial Feature Hallucination Network, AFHN),并运用两个新颖的正则化器对模型进行限制.但 these 方法都比较复杂,训练时需要消耗较多的资源和较长的时间,并且由于GAN网络容易出现模型崩塌,进一步加大了这些模型的训练难度.对此, Wu et al^[13]假设基集中每个类别都服从高斯分布,使用最大后验概率(Maximum A Posteriori, MAP)来估计即使只有一个例子的新类别的分布,提出一种简单有效的方法来生成样本较少的新类别模型.不同于前面的工作,蔡奇和李凡长^[18]在特征的层面上对图像进行增强,提出一种类特征增强模块(Class Feature Augmentation, CFA),将其与原型网络结合得到类特征增强原型网络(CFA-PN). Chen et al^[19]利用语义来直接合成实例特征. Cho and Kim^[20]提出一种自动选择最优增强策略的方法(Efficient Data Augmenta-

tion, EDANet), 从候选的数据增强策略中自动选择最优的组合对数据进行增强. Lazarou et al^[15] 提出一个简单的张量特征幻觉器模型 TFH, 通过利用张量特征而不是向量特征来提升小样本分类性能, 但在该模型的张量特征生成过程中, 得到原型张量特征后, 通过卷积、全连接层等操作将其变为向量特征, 再结合随机产生的噪声向量经过相应的操作得到需要的张量特征. 从张量特征过渡到向量特征再转变为张量特征这一过程, 使模型复杂繁琐, 增加了模型的规模和推理时间.

综上, 尽管不少学者对基于数据增强的小样本图像分类进行了研究, 但在该研究领域仍存在很多未能很好解决的问题, 如数据增强模型复杂、推理时间过长等. 本文对 TFH 模型进行改进, 利用张量特征直接生成需要的张量特征, 无需向

量特征过渡, 能在分类准确率损失较小的情况下缩减模型的规模, 使模型能更快地完成小样本图像分类任务.

2 算法描述

2.1 TFH 模型介绍 TFH 模型在特征空间生成张量特征来增强小样本支持集特征, 然后利用增强后的特征训练得到一个满意的分类器, 它主要由特征提取器、张量特征幻觉器和分类器组成. 支持集图像首先经过特征提取器得到相应的张量特征, 随后对每一类的张量特征相加求平均得到类原型张量特征; 接着将类原型张量特征输入张量特征幻觉器, 得到新生成的张量特征对支持集图像特征进行增强; 最后利用增强后的特征对分类器进行训练. TFH 模型的网络概览如图 2 所示.

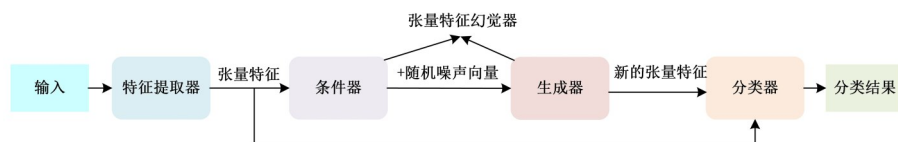


图 2 TFH 模型概览图

Fig. 2 Overview of TFH model

TFH 模型中, 张量特征幻觉器由一个条件器和生成器组成. 类原型张量特征首先输入条件器, 经过卷积、flatten 和全连接的操作变为类条件向量, 然后和服从标准正态分布的噪声向量一起输入生成器, 经过拼接、转置卷积等操作得到想要的张量特征. 但是, 从张量特征转变为向量特征再生成张量特征的过程, 会使模型的规模增大, 推理时间变长.

2.2 张量特征生成器 针对 TFH 模型在生成张量特征的过程中存在从张量特征到向量特征再到张量特征的冗余过渡阶段, 导致模型复杂、推理时间长的问題, 对其中的张量特征幻觉器进行改进, 将原有的条件器和生成器简化为一个张量特征生成器 g . 在张量特征幻觉器中, 使用 flatten 层和全连接层将张量特征转换为向量特征, 两者参数量庞大; 利用转置卷积进行上采样, 计算量也较大. 对此, 在张量特征生成器 g 中不再运用 flatten、全连接等操作将张量特征过渡到向量特征, 并使用只有较少计算量的双线性插值 (Bilinear Interpolat-

ion) 来代替转置卷积上采样的作用, 这大大简化了模型结构. 在张量特征生成器 g 中, 输入为原型张量特征 P_j , 即每一类支持集图像经过特征提取器 f 后变为张量特征, 然后使用式 (1) 将同一类支持集图像所有的张量特征相加求平均, 输出为与原型张量特征标签一致且大小维度一样的张量特征.

$$P_j = \frac{1}{K} \sum_{i=1}^K f_{\theta'}(x_i^j) \quad (1)$$

其中, $i = 1, 2, \dots, K$ 表示每类支持集图像有 i 个样本, $j = 1, 2, \dots, N$ 表示第 j 类支持集图像, x_i^j 表示第 j 类支持集图像的第 i 个样本, θ' 表示特征提取器 f 的参数.

张量特征生成器的网络结构依赖于主干网络, 如图 3 所示, 左边是主干网络为 ResNet18 时张量特征生成器的网络结构, 右边是主干网络为 ResNet12 时张量特征生成器的网络结构.

原型张量特征 P_j 首先经过两层卷积核大小为 3×3 、步长为 1 的卷积得到一个张量特征 P_j' , 然

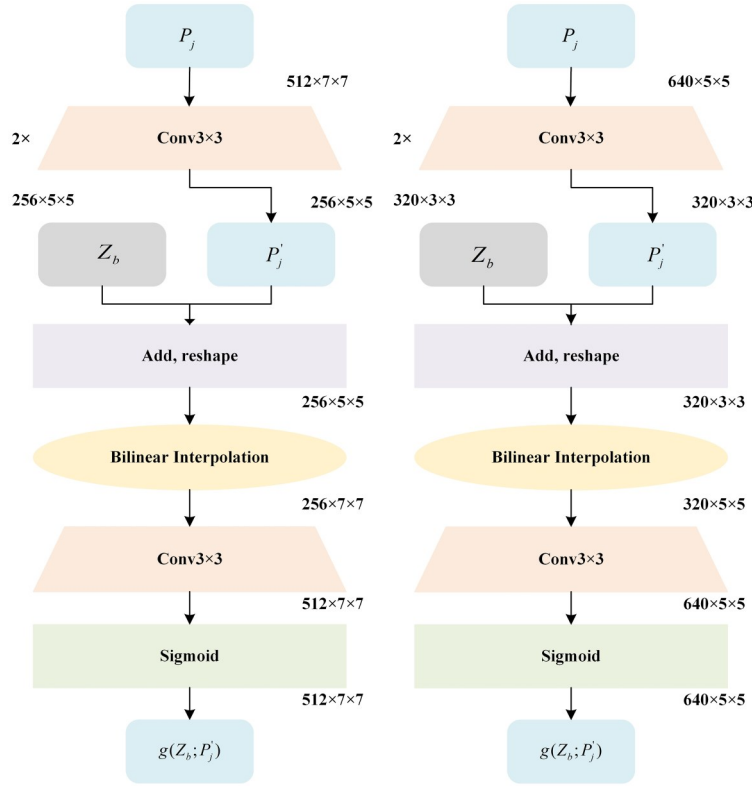
图3 以 ResNet18(左)和以 ResNet12(右)为网络主干的张量生成器 g

Fig. 3 Tensor generator with ResNet18 (left) and ResNet12 (right) as the network backbone

后将 P'_j 和随机产生的服从标准正态分布的张量噪声 Z_b 相加,再进入双线性插值操作对输入的张量特征进行上采样,最后通过一层卷积核大小为 3×3 、步长为 1 的卷积和一个 sigmoid 激活层,输出生成的张量特征 $g(Z_b; P'_j)$,生成的新的张量特征 $g(Z_b; P'_j)$ 与输入的原型张量特征 P_j 的类别标签是相同的,且大小和维度均相等. 由图 3 的张量特征生成器的网络架构可知,对于主干网络 ResNet18,张量特征生成器的张量维度变化为 $[512 \times 7 \times 7]$, $[512 \times 7 \times 7]$, $[256 \times 5 \times 5]$, $[256 \times 5 \times 5]$, $[256 \times 7 \times 7]$ 和 $[512 \times 7 \times 7]$; 对于主干网络 ResNet12,张量特征生成器的张量维度变化为 $[640 \times 5 \times 5]$, $[640 \times 5 \times 5]$, $[320 \times 3 \times 3]$, $[320 \times 3 \times 3]$, $[320 \times 5 \times 5]$ 和 $[640 \times 5 \times 5]$.

张量特征生成器 g 不再将原型张量特征转化为向量特征,而是直接对原型张量特征进行卷积后加入服从标准正态分布的张量噪声 Z_b ,利用双线性插值、卷积和激活函数等操作得到想要的张量特征. 这大大简化了模型规模,减少了模型的

参数量和计算量,使模型更简单,加快了模型的推理速度,能有效解决模型推理时间长的问题,并且它生成的张量特征使下游的分类任务性能只有较小的损失. 使用没有经过全局平均池化(Global Average Pooling, GAP)的张量特征训练张量特征生成器 g ,损失函数使用均方误差(Mean Square Error, MSE)^[21-22],如式(2)所示:

$$J(g) = \frac{1}{BN} \sum_{j=1}^N \sum_{b=1}^B \|g(Z_b; P'_j) - P_j\|^2 \quad (2)$$

其中, B 表示每一类支持集生成的张量特征数.

2.3 小样本图像快速分类网络 针对适用于小样本图像分类的数据增强方法存在模型复杂、推理时间长的问题,基于张量特征生成器 g 提出一种小样本图像快速分类网络(Tensor Feature-based Faster Classification Network, TFFCN). 它以 ResNet18 为主干网络,利用张量特征生成器 g 在特征空间生成张量特征去增强小样本支持集图像特征,使用增强后的特征去训练得到一个满意的分类器,然后对查询集图像进行分类. 该模型

主要由特征提取器 f 、张量特征生成器 g 和分类器三部分组成. 特征提取器对支持集图像和查询集图像特征进行提取, 张量特征生成器利用特征提取器得到的特征生成新的张量特征对支持集图像特征进行增强, 分类器使用增强后的图像特征进行训练, 最后对查询集图像进行分类. TFFCN模型的概览如图4所示.

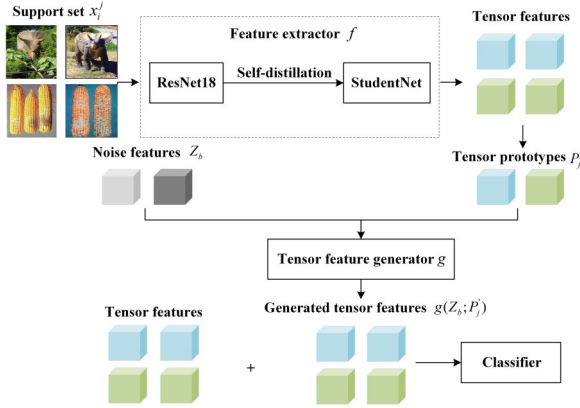


图4 TFFCN模型概览图

Fig. 4 Overview of TFFCN model

在TFFCN模型中, 支持集图像 x_i^j 首先经过特征提取器 f 得到图像的张量特征. 这里为了得到一个较好的特征提取器, 采用两阶段策略来训练特征提取器^[23]: 第一阶段, 在元训练集上对残差神经网络ResNet18进行训练, 采用标准交叉熵作为损失函数, 如式(3)所示. 第二阶段, 运用自蒸馏(self-distillation)操作. 自蒸馏指老师模型和学生模型是同一个模型, 即将第一阶段得到的模型看作老师模型, 蒸馏自己的知识来指导自己得到的模型为学生模型(StudentNet), 运用的损失函数如式(4)所示.

$$J(\theta, \phi) = \sum_{i=1}^D L_{ce}(c_\phi(\bar{f}_\theta(x_i)), y_i) + R(\phi) \quad (3)$$

$$J_{KD}(\theta', \phi') = \alpha J(\theta', \phi') + \beta KL(c_\phi(\bar{f}_{\theta'}(x_i)), c_\phi(\bar{f}_\theta(x_i))) \quad (4)$$

其中, θ 表示特征提取器的参数, 即老师模型 f_θ 的参数, ϕ 表示残差神经网络最后一层全连接层构成的分类器的参数, c_ϕ 表示残差神经网络最后一层全连接层构成的分类器, D 表示元训练集中有标签的样本总数, L_{ce} 表示交叉熵, \bar{f}_θ 表示对老师模

型 f_θ 提取的特征进行全局平均池化, y_i 表示支持集样本 x_i 对应的类别标签, R 是一个正则项. 同理, 带有上标“'”的符号分别表示自蒸馏后的参数, α, β 表示标量权重, KL 表示老师模型和学生模型之间的Kullback-Leibler散度.

通过特征提取器 f 得到支持集图像的张量特征后使用式(1)计算求得每一类的原型张量特征 P_j , 将其作为输入进入生成器 g 后先经过两层卷积, 再与随机抽取的服从标准正态分布的张量噪声 Z_b 一起生成新的张量特征 $g(Z_b; P_j')$. 经过合理的训练, TFFCN模型可以为支持集图像特征生成足够的张量特征用于分类器的训练, 使分类器能更好地分类查询集图像. 其中, 默认的分类器为原型分类器(Prototypical Classifier)^[24], 即将增强后的张量特征进行全局平均池化, 用式(5)得到一个新的类原型向量 \bar{P}_j . 查询集图像 q 经过特征提取器后, 对得到的张量特征进行全局平均池化, 运用式(6)将其分配给最近的类原型向量.

$$\bar{P}_j = \frac{1}{K+B} \left(\sum_{i=1}^K \bar{f}_{\theta'}(x_i^j) + \sum_{b=1}^B \bar{g}(Z_b; P_j') \right) \quad (5)$$

$$d = \sqrt{\sum_{n=1}^{512} (\bar{P}_n^j - \bar{f}(q_i))^2} \quad (6)$$

其中, B 表示为每一类支持集生成的张量特征数, $i=1, 2, \dots, K$ 表示每类支持集有 i 个样本, $j=1, 2, \dots, N$ 表示第 j 类支持集图像, x_i^j 表示第 j 类支持集图像的第 i 个样本, $n=1, 2, 3, \dots, 512$ 表示全局平均池化后向量的维度为512, $\bar{f}_{\theta'}$ 表示对学生模型 $f_{\theta'}$ 提取得到的特征进行全局平均池化, \bar{g} 表示对生成的新的张量特征进行全局平均池化, \bar{P}_n^j 表示增强后的图像特征中第 j 类图像特征经过全局平均池化后得到的向量特征的第 n 维, $\bar{f}(q_i)$ 表示查询集图像经过特征提取器得到的张量特征 $f(q_i)$ 经过全局平均池化操作后得到的向量特征.

3 实验结果与分析

3.1 实验设置 所有实验都使用Python为编程语言, 以深度学习框架Pytorch 1.11.0为基础, 硬件配置为Intel(R) Xeon(R) CPU @ 2.20 GHz Tesla T4 GPU 16 G内存. 实验所用模型及测试代码在 https://github.com/1997zys/zys_study/

tree/master 下载. 使用学习率为 0.05、动量为 0.9、权重衰减系数为 0.0005 的 SGD 优化器对特征提取器进行训练,训练迭代次数为 100. 对于张量特征生成器,设置 $N=5$ 个类别, $K=20$ 表示每类支持集有 20 个样本, $B=50$ 表示为每一类支持集生成 50 个新的张量特征,训练迭代 50 次,每次迭代包含 600 个小样本图像分类任务. 分别进行了 5-way 1-shot 和 5-way 5-shot 的相关实验,其中,分类准确率是重复 600 个小样本图像分类任务求得的平均值,推理时间是模型完成 600 个小样本图像分类任务花费的总时间.

为了评估 TFFCN 模型的性能,在小样本学习的公开数据集 miniImageNet, CUB, CIFAR-FS 上分别进行相关实验. 首先,在以上数据集上与目前较流行的一些算法进行分类性能的比较,并对比 TFFCN 模型与 TFH 模型的推理时间;其次进行了消融实验,分别与没有进行数据增强、直接利用 ResNet18 提取的特征和经过自蒸馏训练后得到的特征提取器提取的特征进行分类的方法进行对比,前者叫作 Baseline(1),后者叫作 Baseline-KD(2),证明提出的张量特征生成器的有效性. 具体数据集的划分如下.

miniImageNet 是从 ImageNet^[25] 分割得到的,是 ImageNet 的一个精缩版本,包含 100 个类别,每个类别含有 600 张图像. 根据 Chen et al^[26] 的划分,64 类作为基类用于训练,16 类用于验证,20 类作为新类用于测试.

CUB 是一个鸟类图像数据集,包含 200 种鸟类,共 11788 张图像. 遵循 Hilliard et al^[27] 划分为

100, 50 和 50 类分别用于训练、验证和测试.

CIFAR-FS 共有 100 个类,每个类包含 600 张图像,总计 60000 张图像. 根据 Chen et al^[28] 的划分,使用 64 类作为训练基类,16 类用于验证,20 类作为新类用于测试.

3.2 模型分类性能与推理时间对比 通过比较模型的性能来判断该模型是否适用于小样本图像分类任务. 表 1 展示了 TFFCN 模型在原型分类器下和一些流行方法在 95% 的置信区间^[29] 的分类性能对比结果. 选取流行的小样本数据增强方法[9, 12, 14, 18, 19, 20],分别在 miniImageNet, CUB 以及 CIFAR-FS 数据集上进行 5-way 1-shot 和 5-way 5-shot 的任务. 由表可见, TFFCN 模型适用于小样本图像分类,在 1-shot 和 5-shot 分类任务上的分类准确率均高于目前流行的小样本数据增强方法.

表 2 展示了当 TFFCN 模型与 TFH 模型的主干网络都为 ResNet18 时,为元测试集中的支持集样本生成不同数量的张量特征用于小样本图像分类时的推理时间与分类准确率对比结果,其中,括号内为模型 TFH 所需的推理时间和得到的分类准确率,括号外为 TFFCN 模型在相同设置下得到的推理时间和分类准确率,推理时间的单位是秒(s),黑体字为两者所需推理时间的最大差距. 由表可知, TFFCN 模型在为每一类支持集生成不同数量的张量特征时,分类准确率相对稳定. 为每类支持集生成 1000 个张量特征时,和只为每一类生成一个张量特征相比,其分类性能变好,虽然此时真实样本占比较少. 另外, TFFCN

表 1 小样本图像分类算法的对比

Table 1 Comparison of few-shot image classification algorithms

Method	Backbone	miniImageNet		CUB		CIFAR-FS	
		1-shot	5-shot	1-shot	5-shot	1-shot	5-shot
IDeMet-Net ^[9]	ResNet-18	59.14 \pm _{0.86}	74.63 \pm _{0.74}	—	—	—	—
AFHN ^[12]	ResNet-18	62.38 \pm _{0.72}	78.16 \pm _{0.56}	70.53 \pm _{1.01}	83.95 \pm _{0.63}	68.32 \pm _{0.93}	81.45 \pm _{0.87}
VI-Net ^[14]	ResNet-18	61.05	78.60	74.76	86.84	—	—
CFA-PN ^[18]	ResNet-12	60.47 \pm _{0.61}	77.82 \pm _{0.44}	—	—	—	—
Dual TriNet ^[19]	ResNet-18	58.80 \pm _{1.37}	76.71 \pm _{0.69}	69.61	84.10	63.41 \pm _{0.64}	78.43 \pm _{0.64}
EDANet+ PrototypicalNet ^[20]	ResNet-50	63.35	79.74	—	—	—	—
TFFCN	ResNet-18	63.99 \pm _{0.80}	79.95 \pm _{0.61}	75.13 \pm _{0.86}	88.08 \pm _{0.47}	73.00 \pm _{0.95}	86.52 \pm _{0.62}

模型的分类准确率比 TFH 模型最多降低 0.94%, 但所需推理时间却远远少于 TFH 模型. 在 5-way 1-shot 分类任务上 TFFCN 模型比 TFH 模型可减少高达 49% 的推理时间, 在 5-way 5-shot 分类任务上 TFFCN 模型比 TFH 模型可减少高达 47% 的推理时间.

表 3 显示了 TFFCN 模型与 TFH 模型的主干网络都为 ResNet12^[30] 时, 为元测试集中的支持集样本生成不同数量的张量特征用于小样本图像分类时的推理时间与分类准确率对比结果. 其中, 括号外是 TFFCN 模型得到的结果, 括号内是 TFH 模型得到的结果, 推理时间的单位是秒(s), 黑体字为两者所需推理时间的最大差距. 由表可见, TFFCN 模型和 TFH 模型相比, 分类准确率最高损失 0.82%, 但大幅缩减了需要的推理时间. $B=1000$ 时在 5-way 1-shot 分类任务上 TFFCN 模型的推理时间比 TFH 模型减少 24%, 在

5-way 5-shot 分类任务上 TFFCN 模型的推理时间比 TFH 模型减少 22%.

3.3 消融实验 为了验证本文提出的张量特征生成器的有效性和鲁棒性, 进行了消融实验. 表 4 展示了在三种数据集上, 不对支持集数据集进行增强的分类效果与 TFFCN 模型的分类效果的对比结果, 使用的分类器为原型分类器. 表 5 展示了在不同分类器上 TFFCN 模型与不对支持集数据进行增强的基线方法的分类效果对比.

由表 4 可见, 在三种公开数据集上 Baseline-KD(2) 的分类性能都优于 Baseline(1), 证明对特征提取器进行自蒸馏操作能增强其特征提取能力, 提升分类准确率. 更重要的, 无论是 1-shot 还是 5-shot 分类任务, TFFCN 模型的分类性能均优于未进行数据增强直接进行小样本图像分类的基线. 针对 5-way 1-shot 分类任务, 在三种数据集上 TFFCN 模型的分类准确率比 Baseline-KD(2) 分

表 2 推理时间和分类准确率的对比

Table 2 Comparison of inference time and classification accuracy

Generate B tensor feature per class	Inference times (s)		Classification accuracy	
	1-shot	5-shot	1-shot	5-shot
$B=1$	179 s (214 s)	218 s (242 s)	$62.65 \pm_{0.84}$ ($62.43 \pm_{0.85}$)	$78.97 \pm_{0.64}$ ($79.33 \pm_{0.62}$)
$B=2$	177 s (215 s)	218 s (244 s)	$63.25 \pm_{0.83}$ ($63.57 \pm_{0.87}$)	$79.00 \pm_{0.64}$ ($79.94 \pm_{0.60}$)
$B=5$	181 s (215 s)	219 s (244 s)	$63.59 \pm_{0.83}$ ($64.42 \pm_{0.83}$)	$79.34 \pm_{0.64}$ ($79.77 \pm_{0.61}$)
$B=10$	179 s (216 s)	219 s (263 s)	$63.66 \pm_{0.83}$ ($63.90 \pm_{0.81}$)	$79.21 \pm_{0.64}$ ($79.64 \pm_{0.60}$)
$B=50$	182 s (223 s)	221 s (249 s)	$63.60 \pm_{0.83}$ ($63.44 \pm_{0.86}$)	$79.08 \pm_{0.63}$ ($79.43 \pm_{0.63}$)
$B=100$	186 s (239 s)	223 s (267 s)	$63.99 \pm_{0.80}$ ($64.49 \pm_{0.84}$)	$79.95 \pm_{0.61}$ ($79.27 \pm_{0.60}$)
$B=500$	238 s (426 s)	268 s (443 s)	$63.59 \pm_{0.83}$ ($64.25 \pm_{0.85}$)	$79.32 \pm_{0.63}$ ($79.71 \pm_{0.62}$)
$B=1000$	354 s (694 s)	376 s (715 s)	$63.60 \pm_{0.83}$ ($63.76 \pm_{0.87}$)	$79.10 \pm_{0.63}$ ($79.52 \pm_{0.62}$)

表 3 推理时间和分类准确率的对比

Table 3 Comparison of inference time and classification accuracy

Generate B tensor feature per class	Inference times (s)		Classification accuracy	
	1-shot	5-shot	1-shot	5-shot
$B=1$	165 s (184 s)	188 s (204 s)	$63.45 \pm_{0.77}$ ($63.88 \pm_{0.85}$)	$79.13 \pm_{0.56}$ ($79.58 \pm_{0.59}$)
$B=2$	166 s (184 s)	189 s (204 s)	$63.72 \pm_{0.79}$ ($64.01 \pm_{0.83}$)	$79.10 \pm_{0.56}$ ($79.44 \pm_{0.61}$)
$B=5$	166 s (189 s)	193 s (207 s)	$63.79 \pm_{0.78}$ ($64.12 \pm_{0.83}$)	$79.11 \pm_{0.56}$ ($79.49 \pm_{0.61}$)
$B=10$	169 s (189 s)	195 s (209 s)	$63.83 \pm_{0.79}$ ($64.01 \pm_{0.85}$)	$79.09 \pm_{0.56}$ ($79.56 \pm_{0.63}$)
$B=50$	172 s (194 s)	200 s (215 s)	$63.31 \pm_{0.78}$ ($63.67 \pm_{0.84}$)	$78.88 \pm_{0.56}$ ($79.18 \pm_{0.65}$)
$B=100$	177 s (206 s)	204 s (227 s)	$62.84 \pm_{0.78}$ ($63.66 \pm_{0.86}$)	$78.53 \pm_{0.55}$ ($78.89 \pm_{0.65}$)
$B=500$	244 s (304 s)	268 s (319 s)	$63.84 \pm_{0.79}$ ($64.11 \pm_{0.86}$)	$78.74 \pm_{0.55}$ ($79.38 \pm_{0.65}$)
$B=1000$	334 s (442 s)	356 s (455 s)	$64.34 \pm_{0.79}$ ($64.69 \pm_{0.83}$)	$78.70 \pm_{0.55}$ ($79.48 \pm_{0.60}$)

表4 在不同数据集上的实验结果对比

Table 4 Experimental results on various datasets

Method	Backbone	miniImageNet		CUB		CIFAR-FS	
		1-shot	5-shot	1-shot	5-shot	1-shot	5-shot
Baseline(1)	ResNet-18	56.81 $\pm_{0.81}$	78.31 $\pm_{0.59}$	67.14 $\pm_{0.89}$	86.22 $\pm_{0.50}$	65.71 $\pm_{0.95}$	84.68 $\pm_{0.61}$
Baseline-KD(2)	ResNet-18	59.62 $\pm_{0.85}$	79.31 $\pm_{0.62}$	70.85 $\pm_{0.90}$	87.64 $\pm_{0.48}$	69.15 $\pm_{0.94}$	85.89 $\pm_{0.59}$
TFFCN	ResNet-18	63.99 $\pm_{0.80}$	79.95 $\pm_{0.61}$	75.13 $\pm_{0.86}$	88.08 $\pm_{0.47}$	73.00 $\pm_{0.95}$	86.52 $\pm_{0.62}$

表5 在不同分类器上的实验结果对比

Table 5 Experimental results by various classifiers

Method	Backbone	Logistic Regression				Support Vector Machine			
		miniImageNet		CUB		miniImageNet		CUB	
		1-shot	5-shot	1-shot	5-shot	1-shot	5-shot	1-shot	5-shot
Baseline(1)	ResNet-18	59.20 $\pm_{0.82}$	77.71 $\pm_{0.61}$	69.44 $\pm_{0.86}$	86.19 $\pm_{0.49}$	57.12 $\pm_{0.84}$	76.45 $\pm_{0.62}$	67.24 $\pm_{0.87}$	84.72 $\pm_{0.52}$
Baseline-KD(2)	ResNet-18	61.83 $\pm_{0.82}$	79.27 $\pm_{0.61}$	72.74 $\pm_{0.88}$	87.71 $\pm_{0.49}$	60.21 $\pm_{0.84}$	78.28 $\pm_{0.61}$	71.23 $\pm_{0.89}$	86.34 $\pm_{0.52}$
TFFCN	ResNet-18	63.92 $\pm_{0.81}$	79.84 $\pm_{0.61}$	74.82 $\pm_{0.87}$	88.43 $\pm_{0.47}$	64.01 $\pm_{0.80}$	79.50 $\pm_{0.60}$	75.40 $\pm_{0.85}$	88.44 $\pm_{0.47}$

别提高 4.37%, 4.28%, 3.85%; 针对 5-way 5-shot 分类任务, 在三种数据集上 TFFCN 模型的分

类准确率比 Baseline-KD(2) 分别提高 0.64%, 0.44%, 0.63%. 证明提出的张量特征生成器结合表 4 和表 5 可以看出, 在 Prototypical Classifier, Logistic Regression (LR), Support Vector Machine (SVM) 三个不同的分类器上, TFFCN 模型都表现出最高的分类性能. 其中, 在 LR 分类器和 SVM 分类器下的 5-way 1-shot 分类准确率, 在 miniImageNet 数据集上 TFFCN 模型分别比 Baseline-KD(2) 提高 2.09% 和 3.8%, 在 CUB 数据集上 TFFCN 模型分别比 Baseline-KD(2) 提高 2.08% 和 4.17%. 因此, 张量特征生成器有助于小样本图像分类, 同时其可以与不同分类器进行结合使用, 具有较好的鲁棒性.

表 6 为 TFFCN 模型中张量特征生成器和

表6 计算量和参数量对比

Table 6 Comparison of calculation amounts and parameter quantities

Method	计算量 (G)	参数量 (M)
TFH	20.56	24.25
TFFCN	10.15	4.72

TFH 模型中张量特征幻觉器的参数量和计算量的对比. 首先通过 python 的 thop 库得到 FLOPs 和 Params, 然后利用 $1\text{ G FLOPs} = 10^9\text{ FLOPs}$ 将

有效的, 它可以通过为支持集图像特征生成新的张量特征来达到对小样本图像进行数据增强的目的, 从而训练得到一个更好的分类器, 提高分类准确率.

计算量的单位转换为 G, 利用 $1\text{ Params} = 10^6\text{ M}$ 将参数量的单位转换为 M.

由表可见, 张量特征生成器的计算量不到张量特征幻觉器计算量的 50%, 参数量的缩减幅度更大, 仅为张量特征幻觉器参数量的 19%. 这表明提出的张量特征生成器直接用张量特征生成新的张量特征, 不再过渡为向量特征, 可以有效缩减模型的规模, 减小模型的参数量和计算量, 使网络结构更加简单, 这也间接体现了 TFFCN 模型所需推理时间更少的原因.

4 结论

针对适用于小样本图像分类的数据增强方法复杂、推理时间长的问题, 本文提出一个结构简单的张量特征生成器, 为小样本图像进行数据增强. 并基于张量特征生成器提出一个小样本图像快速分类方法 TFFCN, 能有效解决模型推理时间长的问题. 通过与目前流行的基于数据增强的小样本图像分类方法进行分类性能对比分析, 证明提出的 TFFCN 模型的小样本图像分类性能优于目前流行的基于数据增强的小样本图像分类方

法. 另外, 分别对模型的主干特征提取网络为 ResNet18 和 ResNet12 时的推理时间进行实验对比, 证明和 TFH 模型相比, TFFCN 模型在分类准确率损失较小的情况下, 大幅减少了模型的推理时间, 特别是生成的张量特征数量越多, 推理时间的缩减越明显. 同时, 也证明 TFFCN 模型在不同主干网络下都能更快地完成小样本图像分类任务.

最后, 通过消融实验证明了提出的张量特征生成器的有效性和鲁棒性, 它不仅能有效改善小样本图像分类任务的性能, 还能与多种分类器结合使用, 具有较好的鲁棒性. 并且, 张量特征生成器的结构简单, 计算量和参数量都较小, 大大缩减了模型的规模. 但 TFFCN 模型仍存在不足, 如原型张量特征的代表性不高, 未来的工作主要是集中提高原型张量特征的代表性, 使模型更加完善, 取得更好的效果.

参考文献

- [1] 鲜维富. 基于深度学习的小样本图像分类与检测算法研究与实现. 硕士学位论文. 成都: 电子科技大学, 2021. (Xian W F. Research and implementation of few-shot classification and object detection algorithm based on deep learning. Master Dissertation. Chengdu: University of Electronic Science and Technology of China, 2021.)
- [2] Jankowski N, Duch W, Grąbczewski K. Meta-learning in computational intelligence. Springer Berlin Heidelberg, 2011, 97—115.
- [3] Lake B M, Salakhutdinov R, Tenenbaum J B. One-shot learning by inverting a compositional causal process//Proceedings of the 26th International Conference on Neural Information Processing Systems. Lake Tahoe, NV, USA: Curran Associates Inc., 2015: 2526—2534.
- [4] 卢锦亮, 吴广潮, 冯夫健等. 基于联合轨迹特征的徘徊行为识别方法. 南京大学学报(自然科学), 2021, 57(5): 724—734. (Lu J L, Wu G C, Feng F J, et al. Wandering recognition method based on joint trajectory features. Journal of Nanjing University (Natural Science), 2021, 57(5): 724—734.)
- [5] Mehrotra A, Dukkipati A. Generative adversarial residual pairwise networks for one shot learning. 2017, arXiv:1703.08033.
- [6] Antoniou A, Storkey A, Edwards H. Data augmentation generative adversarial networks. 2018, arXiv:1711.04340.
- [7] Zhang R X, Che T, Ghahramani Z, et al. MetaGAN: An adversarial approach to few-shot learning//Proceedings of the 32nd International Conference on Neural Information Processing Systems. Montréal, Canada: Curran Associates Inc., 2018: 2371—2380.
- [8] Liu Y F, Zhou Y, Liu X, et al. Wasserstein GAN-based small-sample augmentation for new-generation artificial intelligence: A case study of cancer-staging data in biology. Engineering, 2019, 5(1): 156—163.
- [9] Chen Z T, Fu Y W, Wang Y X, et al. Image deformation meta-networks for one-shot learning//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, CA, USA: IEEE, 2019: 8672—8681.
- [10] Wang Y X, Girshick R, Hebert M, et al. Low-shot learning from imaginary data//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE, 2018: 7278—7286.
- [11] Hariharan B, Girshick R. Low-shot visual recognition by shrinking and hallucinating features//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE, 2017: 3037—3046.
- [12] Li K, Zhang Y L, Li K P, et al. Adversarial feature hallucination networks for few-shot learning//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, WA, USA: IEEE, 2020: 13467—13476.
- [13] Wu J Y, Dong N, Liu F, et al. Feature hallucination via maximum a posteriori for few-shot learning. Knowledge-Based Systems, 2021(225): 107129, DOI:10.1016/j.knosys.2021.107129.
- [14] Luo Q X, Wang L F, Lv J G, et al. Few-shot learning via feature hallucination with variational inference//Proceedings of 2021 IEEE Winter Conference on Applications of Computer Vision. Waikoloa, HI, USA: IEEE, 2021: 3962—3971.
- [15] Lazarou M, Stathaki T, Avrithis Y. Tensor feature hallucination for few-shot learning//Proceedings of 2022 IEEE/CVF Winter Conference on Applications of Computer Vision. Waikoloa, HI,

- USA:IEEE,2022:2050—2060.
- [16] 赵凯琳,靳小龙,王元卓.小样本学习研究综述.软件学报,2021,32(2):349—369.(Zhao K L,Jin X L,Wang Y Z. Survey on few-shot learning. Journal of Software,2021,32(2):349—369.)
- [17] 祝钧桃,姚光乐,张葛祥,等.深度神经网络的小样本学习综述.计算机工程与应用,2021,57(7):22—33.(Zhu J T,Yao G L,Zhang G X,et al. Survey of few shot learning of deep neural network. Computer Engineering and Applications,2021,57(7):22—33.)
- [18] 蔡奇,李凡长.类特征增强元学习算法.小型微型计算机系统,2022,43(2):225—230.(Cai Q,Li F C. Meta-learning with class feature augmentation. Journal of Chinese Computer Systems, 2022, 43(2): 225—230.)
- [19] Chen Z T, Fu Y W, Zhang Y D, et al. Multi-Level semantic feature augmentation for one-shot learning. IEEE Transactions on Image Processing, 2019, 28(9):4594—4605.
- [20] Cho W, Kim E. Improving augmentation efficiency for few-shot learning. IEEE Access, 2022(10): 17697—17706.
- [21] Huang H, Feng F J, Huang S Q, et al. Microscale searching algorithm for coupling matrix optimization of automated microwave filter tuning. IEEE Transactions on Cybernetics, 2022, DOI: 10.1109/TCYB.2022.3166225.
- [22] Feng F J, Huang H, Liu D, et al. Local complexity difference matting based on weight map and alpha mattes. Multimedia Tools and Applications, 2022, DOI:10.1007/s11042-022-13223-1.
- [23] Tian Y L, Wang Y, Krishnan D, et al. Rethinking few-shot image classification: A good embedding is all you need? //European Conference on Computer Vision (Online). Springer Cham, 2020:266—282.
- [24] Snell J, Swersky K, Zemel R S. Prototypical networks for few-shot learning. 2017, arXiv:1703.05175.
- [25] Deng J, Dong W, Socher R, et al. ImageNet: A large-scale hierarchical image database//2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami, FL, USA: IEEE, 2009: 248—255.
- [26] Chen Y B, Liu Z, Xu H J, et al. Meta-baseline: Exploring simple meta-learning for few-shot learning//Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal, Canada:CVPR,2021:9062—9071.
- [27] Hilliard N, Phillips L, Howland S, et al. Few-shot learning with metric-agnostic conditional embeddings. 2018, arXiv:1802.04376.
- [28] Chen W Y, Liu Y C, Kira Z, et al. A closer look at few-shot classification. 2020, arXiv:1904.04232.
- [29] 冯夫健,黄翰,吴秋霞,等.基于群体协同优化的高清图像前景遮罩提取算法.中国科学:信息科学,2020,50(3):424—437.(Feng F J,Huang H,Wu Q X,et al. An alpha matting algorithm based on collaborative swarm optimization for high-resolution images. Sci Sin Inform, 2020, 50(3):424—437.)
- [30] Zhang C, Cai Y J, Lin G S, et al. 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, WA, USA:IEEE,2020:12200—12210.

(责任编辑 杨可盛)