计算机科学
COMPUTER SCIENCE

基于多图特征聚合的小样本学习方法

曾武, 毛国君

引用本文

曾武, 毛国君. 基于多图特征聚合的小样本学习方法[J]. 计算机科学, 2023, 50(6A): 220400029-10.

ZENG Wu, MAO Guojun. Few-shot Learning Method Based on Multi-graph Feature Aggregation[J]. Computer Science, 2023, 50(6A): 220400029-10.

相似文章推荐（请使用火狐或 IE 浏览器查看文章）

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

基于多特征融合的GRU-LSTM大学生就业动态预测

College Students Employment Dynamic Prediction of Multi-feature Fusion Based on GRU-LSTM

计算机科学, 2023, 50(6A): 220500056-6. https://doi.org/10.11896/jsjkx.220500056

基于深度学习的超高频标签识别系统

Tag Identification for UHF RFID Systems Based on Deep Learning

计算机科学, 2023, 50(6A): 220200151-6. https://doi.org/10.11896/jsjkx.220200151

基于改进Yolov4-tiny的轻量型目标检测算法

Lightweight Target Detection Algorithm Based on Improved Yolov4-tiny

计算机科学, 2023, 50(6A): 220700006-7. https://doi.org/10.11896/jsjkx.220700006

CT影像阶段化目标检测方法研究

Study on Phased Target Detection in CT Image

计算机科学, 2023, 50(6A): 220200063-10. https://doi.org/10.11896/jsjkx.220200063

基于深度学习的摩托车车道实时检测

Real-time Detection of Motorcycle Lanes Based on Deep Learning

计算机科学, 2023, 50(6A): 220200066-5. https://doi.org/10.11896/jsjkx.220200066

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.220400029

基于多图特征聚合的小样本学习方法

曾　武[1]　毛国君[1,2]
1 福建工程学院计算机科学与数学学院　福州 350118
2 福建省大数据挖掘与应用重点实验室　福州 350118
(2201905122@smail.fjut.edu.cn)

摘　要　小样本学习可以从较少的样本中学习出各类样本的特征,但是由于低数据的问题,即样本数量较少,如何更加准确地提取图像中的重要特征信息,以及更好地学习图像中目标对象的特征和更精准地判断未标记样本与支持集类别的相似度,成为关键。提出一种基于多图特征聚合的小样本学习方法 MGFAN。具体来说,该模型通过多种数据增强方法对原图进行扩充,然后使用一个自注意力模块去获取原图以及不同扩展图之间的重要特征信息,以此获得关于图像更为准确的特征向量。其次,在模型中引入关于预测图像不同扩增方式的自监督学习任务作为辅助任务,促进模型的特征学习能力。最后,采用多个距离函数来更加精准地计算样本间的相似度。在 3 个标准数据集 miniImageNet,tieredImageNet 和 Stanford Dogs 中,使用 5-way 1-shot 以及 5-way 5-shot 实验设置中的实验表明,MGFAN 方法可以显著提高分类器的分类性能。

关键词:小样本学习;深度学习;自监督学习;特征聚合;数据扩增;自注意力
中图法分类号　TP391

Few-shot Learning Method Based on Multi-graph Feature Aggregation

ZENG Wu[1] and MAO Guojun[1,2]
1 School of Computer Science and Mathematics,Fujian University of Technology,Fuzhou 350118,China
2 Fujian Provincial Key Laboratory of Big Data Mining and Applications,Fuzhou 350118,China

Abstract　Few-shot learning can learn the characteristics of various samples from fewer samples,but due to the problem of low data,that is,the number of samples is small,how to more accurately extract the important feature information in the image,and how to better learn from the image. The characteristics of the target object and the more accurate judgment of the similarity between the unlabeled samples and the support set category become the key. A few-shot learning method MGFAN based on multi-graph feature aggregation is proposed. Specifically,the model expands the original image through various data enhancement methods,and then uses a self-attention module to obtain important feature information between the original image and different expanded images,so as to obtain more accurate features vector about the image. Secondly,the self-supervised learning task of predicting different augmentation methods of images is introduced into the model as an auxiliary task to promote the feature learning ability of the model. Finally,multiple distance functions are used to calculate the similarity between samples more accurately. Experiments in 3 standard datasets miniImageNet,tieredImageNet and Stanford Dogs using 5-way 1-shot and 5-way 5-shot experimental settings show that the MGFAN method can significantly improve the classification performance of the classifier.

Keywords　Few-shot learning,Deep learning,Self-supervised learning,Feature aggregation,Data augmentation,Self-attention

1　引言

近年来,随着科学技术的不断发展,特别是基础计算硬件的计算性能不断提高,卷积神经网络(Convolutional Neural Networks,CNN)[1]在各应用领域取得了重大的成功。其中,深度学习在人工智能领域受到了热烈的追捧,尤其是在图像分类[2]、目标检测[3]、语义分析[4]、视频检测[5]以及语音识别[6]等领域。总的来说,深度学习的成功在一定程度上归功于大型数据集(例如 ImageNet[7])以及计算能力强大的计算设备(GPU)。大型数据集中每个类别都拥有大量的数据,大量的数据有助于算法学习到性能较高的模型。然而,并不是所有的深度学习任务都可以获得较多的数据。而且,即使有大量的无标记数据,对这些数据进行标注也是非常费时费力的事情。因此,在有些任务中包含的基础数据较少,那么使用传统的神经网络模型进行训练,可能就会很容易造成模型过拟合以及收敛困难等问题。与深度神经网络需要大量的数据才能取得准确率较高模型的学习方式不同,人类可以仅依赖少量的数据,就能较好地完成对新事物的学习。例如,一个人可能只需要看到 1 张或者几张大熊猫的图片,之后去动物园或者其它地方看到大熊猫,就可以快速地辨别出这是大熊猫。

基金项目:国家自然科学基金项目(61773415);国家重点研发项目(2019YFD0900805)
This work was supported by the National Natural Science Foundation of China(61773415) and National Key Research and Development Program of China(2019YFD0900805).
通信作者:毛国君(19662092@fjut.edu.cn)

以此为启发，人们提出小样本学习[8]（Few-Shot Learning，FSL），实现对仅有少量标注的图片数据集进行分类。大多数小样本学习的主要方法为：在具有较多类别的训练任务数据集上进行训练，从而获得从少量的数据中进行学习的能力[9]，然后再把学习到的方法迁移到类别数较少的测试集中，测试学习到的能力。

在小样本学习中，设定一个元任务中包含 N 个类，且每个类只采样 K 个样本，将其称为 N-way K-shot 任务。图 1 展示了 5-way 1-shot 的小样本学习任务描述，从训练任务数据集中抽取 5 个类，每个类抽取 1 个样本作为支持集（support set），且另外抽取 2 个样本作为查询集（query set），接着在测试任务数据集中按照相同的方法抽取样本进行验证。

图 1 使用 5-way 1-shot 的小样本学习任务描述

Fig. 1 Task description of few-shot learning using 5-way 1-shot

研究人员提出了多种方法来优化小样本学习中的一些难题，而这些方法大体可以被归纳为两类，即基于度量学习或者基于元学习的小样本学习方法。度量学习的策略是：分别计算标记与无标记样本之间的特征向量，特征间距离越小，则表明二者之间相似度越高，进而预测无标记样本的类别。元学习的策略是：主张跨任务学习，根据任务的不同，动态地调整参数，可以从较好的初始点进行训练，仅需几个迭代就可以取得较好的结果[10]。

Korch 等[11]提出的孪生神经网络（Siamese Neural Networks）为一种相似性度量模型，它利用两个相同的 CNN 网络对输入的图像进行特征提取，接着将查询集中的每个图像与支持集中的图像进行成对组合，然后传入神经网络中，得到各对样本的距离，将距离最小的视为相同的类别，以此来实现图像分类。Wang 等[12]提出通过构建图像标签与图像特征的注意力机制，利用注意力机制的关注点来决定这张图像的向量表示。Snell 等[13]提出原型网络（Prototypical Networks），其核心方法为：在每个类别中抽取几个样本，接着在每个类别抽取出的样本中寻找一个原型（Prototype），这个原型点也被称为类别中心点。将原型点使用神经网络映射成特征向量，接着把同一类图像特征向量的平均值作为原型点。在此之后，同样使用神经网络对查询集中的无标签样本进行特征向量的提取，再计算这些样本的特征向量与各类原型点的平方欧氏距离，得出查询集样本与所有类别的距离，距离越小，则表示类别更为相近。模型会把无标记样本归类为距离最小的那个类别。以上方法虽然可以提高小样本学习的分类能力，且实现简单，但是仍然会受到低数据带来的问题，换句话说，因为数据样本太少，所以提取的特征不够准确。具体来说，通过 1 张或者少数几张样本提取到的特征，不能够较好地表示这个类别的类均值特征，从而导致分类结果产生偏差。主要原因有：1）使用 CNN 网络对图像进行特征提取，那么提取的特征向量中同样包含背景以及一些干扰物的特征向量，然而期望尽可能提取图像中目标对象的特征向量；2）难以尽可能多地提取一张图像中的重要特征信息。究其原因，在同一个类别中，不同的图像样本中，目标对象的大小占比、姿态及位置大都是不同的，因此在 N-way K-shot 任务中 K 越大，分类准确率往往越高。这是由于支持集中每个类别样本越多，提取的特征均值越接近各类类别的真实分布。因此，如何在低数据的情况下更好地提高每张图像的特征的表达能力是一个关键的问题。针对此难题，本文中提出：分别将原图进行水平翻转、垂直翻转和 180 度旋转来改变目标对象的姿态以及位置，以及对原图进行部分裁剪来改变目标对象的大小占比。这样，原本一张图像中只能抽取一组特征向量，经过图像扩增处理后，可以提取 5 组特征向量，之后将 5 组特征向量进行特征聚合处理，压缩成 1 组特征向量，以此来提高图像的特征表达能力。具体而言，人类可以从一张图像中快速锁定图中的关键目标对象，然后快速学习这个目标对象的一些重要特征。以此为启发，可以在图像中加入自注意力模块，对图像中较为重要的特征进行编码。换句话说，将输入的图像使用数据增强扩增为 5 张，然后使用自注意力模块分别提取这 5 张图像中的重要特征向量，之后将这些特征向量进行拼接聚合，以此来表示这张图像的特征信息，从而尽可能多地从一张图像中获取足够多的重要信息。

近年来，大多数小样本分类方法计算样本之间相似度的距离函数为平方欧氏距离。平方欧氏距离可以计算两个样本

特征向量之间各个相同维度间的距离,之后继续将这个差值进行平方计算。在理想的情况下,两个相同类别的样本之间,由于其图像中目标对象是相同的,因此,在其目标对象的像素区域之间所提取的特征向量差异是比较小的。反之,不同类别之间的差异值是比较大的,通过对这部分特征差值进行平方计算,可以继续扩大这部分的差值分数,以此来实现相同类别之间差值更小而不同类别之间差值越大的目的,从而进行更好的分类。然而,就算在同一个类别的图像中,其背景的复杂程度以及目标对象的大小占比都是不同的。例如,在同一个类别中抽取的 3 个样本,其中有 1 个样本的背景和另外 2 个的背景差距很大,且该样本目标对象的大小占比面积较小,若将这 3 个图像各个维度的特征向量差值继续进行平方计算,无疑会更加扩大这个样本与其他 2 个样本间的距离。为了缓解噪音样本对分类的影响,在判定样本相似度的距离函数中新加入曼哈顿距离函数,实现在样本距离的判定中同时使用平方欧氏距离和曼哈顿距离。曼哈顿距离函数同样可以计算两个相同维度间的差值,且不对差值进行平方计算,避免了在一些干扰噪音较大的图像中出现的误判。出于以上种种考虑,本文使用多个距离度量函数,计算经过多图特征聚合的未标记样本的特征向量与已标记样本的距离,即使用多个距离函数来度量未标记样本与原型点之间的差距,最后按照一定的权重系数进行计算,得出最终的距离。

小样本学习需要对支持集的数据进行标注,而后使用无标记的查询集数据进行验证学习。相比于小样本学习,自监督学习(Self-supervised Learning,SSL)[14]则不用对源数据进行标注,它可以从这些无标注的数据中学到可泛化的特征表示,以便于适应其他任务。这些任务在实例样本获得自监督(Self-supervised)标记(例如图片的旋转预测[15]、拼图求解[16]等)。小样本学习和自监督学习在本质上都是通过知识迁移来实现减少为目标任务搜集大量数据样本的需要。在小样本分类任务中,由于支持集中样本数量较少,因此若能更好地识别样本中的一些局部特征,将有助于模型对目标对象特征的学习,所以本文加入了关于图像变换方式的自监督学习任务来促进知识转移。换句话说,如果模型能分辨一张站立的动物图像是否发生了 180 度旋转,那么模型就必须要理解动物原有的姿态,识别到原本站立在地上的腿部此时发生了变化,出现在和原先相反的方向,以此来学习一些局部特征信息。将自监督学习任务组合在小样本学习模型中。通过将自监督学习中学到的一些知识迁移到下游任务中,有利于下行任务的进一步学习[2]。

本文的主要贡献包括:

(1)使用了多距离函数度量模块。针对部分图像中存在噪音等干扰的影响,为了更加准确地计算样本间相似度,使用多种距离函数通过加权来计算未标记样本与原型的距离,通过多种度量距离函数获取更为准确的相似度计算。

(2)使用不同的数据增强方法来生成相较于原图目标对象不同的大小占比、姿态以及位置等扩增图像。使用自注意力模块 Transformer 对原图以及扩增图的特征进行提取,再将这些特征向量进行聚合,以此来更加准确地表达图像的特征信息。

(3)在模型中加入关于预测图像变换方式的自监督学习辅助任务,目的是将自监督学习任务中学习到的知识转移到小样本分类任务中,通过促进模型更好地学习目标对象的特征来提高模型分类性能。

(4)在 3 个标准数据集上进行了大量的实验,实验结果证明本文方法可以显著提高分类器的分类性能,而且超过当前主流的小样本学习方法。

2　相关工作

2.1　小样本学习

一个人在电子产品中看到几张某种蔬菜的照片,就可以很好地归纳这种蔬菜的某些特征,之后在菜园中就可以很快认出这种蔬菜是什么。类似于这种学习方式,小样本学习任务的目标是在少量已标记的样本中学习到某类已标记样本的特征,而后以此特征为标准来对未标记的样本进行分类。近年来,许多基于深度神经网络的小样本学习方法在小样本学习任务中取得了较高的准确率,这些方法主要分为 2 种:度量学习与元学习方法。

基于度量学习的小样本学习方法,其常用方法是求取支持集和查询集图片的特征嵌入空间,之后使用相似性度量模块分别计算查询集样本与支持集样本的相似度,从而完成查询集样本的类别预测。Korch 等[11]提出孪生神经网络,其通过共享权值的对称网络来计算输入的 2 张图片之间的相似度,从而对输入的成对图片是否属于相同的类别做出预测。Snell 等[13]提出了原型网络,其中心思想是计算出支持集中每个类别的类均值特征作为此类别的原型(类中心点),之后通过度量函数计算查询集样本到原型的距离,来判断查询集样本所属的类别。Vinyals 等[17]提出的匹配网络(Matching Network)把度量学习与长短记忆网络(Long Short Term Memory,LSTM)进行联立,之后使用注意力加权过的度量方法预测样本之间的相似度。Sung 等[18]提出的关系网络(Relation Network)通过构建一个完全端到端(End-to-End)的网络来给出一个可以判定物体间关系的模型。关系网络的主要作用是提供了一种可学习的非线性分类器。

基于元学习的小样本学习方法则是主张跨任务学习,可以根据所进行任务的不同,给出适合当前训练任务的初始化参数,从而指导现有的模型在全新的任务中进行快速学习。Santoro 等[19]提出一种具有记忆增强的网络架构(Memory-Augmented Neural Networks,MANN),它是一种元学习算法,主要利用 LSTM 模块实现与其结合神经网络带有长期的记忆能力,从而实现小样本学习。Finn 等[20]提出了一种与方法无关的元学习算法(Model-agnostic Meta Learning,MAML),其主要方法是通过寻找可以快速适应新任务学习中较好的初始化参数来实现小样本学习。Ravi 等[21]提出一个新的元学习算法,其主要是基于 LSTM 的元学习框架,利用长短时记忆网络结构来学习分类器的优化参数,且在训练中采用的梯度简化方法可以有效降低学习复杂度。Li 等[22]提出的 Meta-sdg 方法可以学得较优的初始参数以及合适的学习率等。Shyam 等[23]则提出一种基于注意力机制的循环网络用于小样本学习,并且取得了不错的效果。

2.2　自监督学习

自监督学习可以利用辅助任务从大规模的无监督信息中

挖掘其本身的监督信息，像这种通过构造监督信息对网络进行训练的，可以使模型学习到对下游任务有益的表征信息。具体来说，在自监督学习中，不需要对数据进行标注，这些标注是模型从传入网络的源数据中获得的。总的学习模式主要是：模型先在辅助任务中进行预训练，之后再把学习到的知识或者经验迁移到下游任务中。自监督学习已被证实有利于多种下行任务，例如目标检测[3]、语义分析[4]以及小样本分类[8]等。

现有的自监督学习主要的不同在于辅助任务设计不同。Doersch 等[24]提出一种拼图任务，即将一张图片分为多个部分，然后对这些部分的位置进行预测，学习到的模型若是可以较好地完成这个任务，则认为模型是学得了这些图片的表征信息。Gidaris 等[15]提出通过 ConvNets 网络来预测图片的 2D 旋转，以此来学习这些图片的特征，这些简单的任务给语义特征的学习提供了很强的监督信号。Pathak 等[25]将一张图片随机去除一个部分作为训练数据，而原始数据则作为预训练的训练标签进行训练，作为图像修复的辅助任务加入下游任务中。在 PASCAL VOC2012[26]数据集的分类和检测任务中，该策略可以提高模型的性能。

3 MGFAN 的设计

3.1 问题定义

在一个小样本学习任务中，一般采用元学习的训练方式。具体而言，元学习的数据集分为训练数据集 D_{train} 和测试集 D_{test}，其中训练数据集和测试数据集的类别没有相交部分，而且在一个相同的小样本学习任务中遵循相同的 N-way K-shot 训练方式。每个元任务 T 包含有一个支持集 S(Support Set)和一个查询集 Q(Query Set)，元任务致力于通过给定的支持集 S 来对查询集 Q 中的图片进行分类。具体而言，在一个 N-way K-shot 的小样本任务中，先从训练集中采样 N 个类，并且将之记为 C，然后再从这 C 个类中分别抽取 K 个支持样本和 q 个查询样本。支持集 S、查询集 Q 和元学习任务 T 可被定义为：

$$S=\{(x_i,y_i)\mid y_i\in C,i=1,2,\cdots,N\times k\} \tag{1}$$

$$Q=\{(x_i,y_i)\mid y_i\in C,i=1,2,\cdots,N\times q\} \tag{2}$$

$$T=\{(S_i,Q_i)\}_{i=1}^{m} \tag{3}$$

其中，x_i 表示图像样本，y_i 表示其图像标签；m 表示元任务学习 T 的个数。

在元训练阶段，每个元任务中分为内环和外环。在内环中，模型通过支持集 S 来更新参数，接着在外环中通过查询集 Q 来对模型的性能进行评估，以评估的结果来判定是否更新模型的参数。

3.2 MGFAN 的模型架构

本文的总体模型架构如图 2 所示，主要由 5 个模块构成。

模块 1(数据扩增模块) 为了改善低数据的问题，提出使用数据扩增技术，对元任务中支持集以及查询集中的全部图像，针对性地采用垂直翻转、水平翻转、180°旋转以及部分裁剪方法，生成 4 张相比于原图目标对象具有不同姿态、位置以及大小占比的扩增图。

模块 2(特征提取模块) 以 ProtoNet[13]作为基准模型，使用 CNN 网络来对全部图像(原图及其扩增图)的特征进行提取。

模块 3(辅助任务模块) 在小样本学习任务中加入关于识别图像发生何种增强变换的自监督学习任务作为辅助任务，通过更好地学习图像局部特征的变化完成自监督学习，将学习到的知识转移到小样本学习任务中，促进模型更好地学习目标对象的特征。

模块 4(Transformer 多图特征聚合模块) 为了更好地提取每张图像的重要区域特征信息，将每 1 张图像通过数据增强方法扩充成 5 张图像，在这 5 张图像中都部署 Transformer 自注意力模块，利用注意力机制更新每张图像(原图及其扩增图)特征向量中重要的特征向量，接着把 5 张图像中重要的特征向量拼接聚合，以此来获取关于每个训练样本图像中更为准确以及详细的特征信息。

模块 5(多距离函数度量模块) 使用多个距离度量函数来判断样本间的距离，即在平方欧氏距离的基础上增加曼哈顿距离函数来分别计算元任务中全部查询集样本与支持集中各个原型点的距离，最后通过各自权重计算总的距离，以此来判断两者之间的相似度。

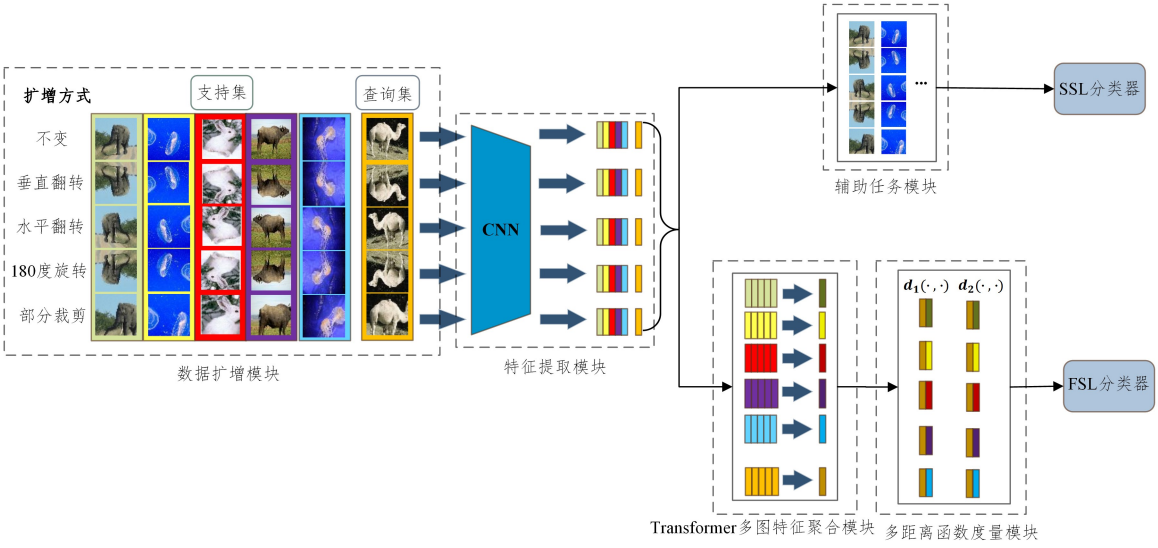

图 2 基于多图特征聚合的小样本学习方法

Fig. 2 Few-shot learning method based on multi-graph feature aggregation

3.2.1　特征提取模块

本文中选择 ProtoNet[13] 作为基础的小样本学习模型,如图 2 所示,该模型中拥有一个进行特征提取的 CNN 网络以及一个非参数分类器,其中特征提取中的参数是在元学习中获取的。具体而言,ProtoNet 根据元学习中获取的参数来更新模型,接着利用特征处理器来计算每个类的平均特征向量值嵌入,其过程可以被表示为:

$$h_c=\frac{1}{K}\sum_{(x_i,y_i)\in S} g_\sigma(x_i),\rho(y_i=c) \tag{4}$$

其中,类别 $c\in C$,g_σ 是一个特征提取器,σ 是通过元学习学到的参数,$\rho(\cdot)$ 表示指示函数。之后通过计算每个查询集样本的特征嵌入与各个类的特征嵌入的距离,来判断相对应查询集样本的类别。其相对应的损失函数定义为:

$$L_{\mathrm{fsl}}=-\frac{1}{|q|}\sum_{(x_i,y_i)\in Q}\log\frac{\exp(-d(g_\sigma(x_i),h_{y_i}))}{\sum_{c\in C}\exp(-d(g_\sigma(x_i),h_c))} \tag{5}$$

其中,$d(\cdot,\cdot)$ 表示距离函数。

3.2.2　辅助任务模块

如图 2 所示,在 MGFAN 模型中对元任务 $T=\{S,Q\}$ 中的每个图片进行一些变换,分别为图片的垂直翻转、水平翻转、180°旋转以及部分裁剪。由于这一系列的变换生成了一组扩增图片,扩增图片包括原图在内总共有 H 张样本,于是新的元集 $T^r=\{\{S^r,Q^r\mid r=0,1,\cdots,H-1\}$,其

$$L_c=-\left(\frac{1}{H(l_k+l_q)}\sum_{r=0}^{H-1}\sum_{(x_i,y_i,r_i)\in T^r}\log\frac{\exp([g_{\theta r}(g_\sigma(x_i))]_{r_i})}{\sum_{r'=0}^{H-1}\exp([g_{\theta r}(g_\sigma(x_i))]_{r'})}\right) \tag{9}$$

3.2.3　多图特征聚合模块

多图特征聚合模块是 MGFAN 方法中的关键模块。通过将原图进行水平翻转、垂直翻转、旋转 180 度以及部分裁剪生成多个扩展图,然后使用自注意力模块筛选出在不同视角下图像中的重要特征信息,最后将 5 张图的重要特征信息进行拼接聚合,以此来表示此样本的特征信息,将包含扩增图的总集表示为 T^r_{emb};

$$T^r_{\mathrm{emb}}=\{g_\sigma(x_i)\mid(x_i,y_i,r)\in T^r,r=0,\cdots,H-1;i=1,\cdots,l_k+l_q\} \tag{10}$$

首先使用特征提取网络获取原图及其扩增图的特征 T^r_{emb},接着为了获取这些扩展图片间的关联,并且将这些特征更好地聚合起来,形成类特征均值,本文使用自注意力机制来完成这一任务。具体而言,首先基于 T^r_{emb},构造特征张量 F,接着使用一个 Transformer[27] 来获取它所认为的重要特征信息。Transformer 是一种自注意力机制,它包含有缩放点积注意力(Scaled Dot-Product Attention)和多头注意力(Multi-Head Attention),其中多头注意力有助于网络捕捉到更丰富的特征信息。Transformer 会接收一组三元组的输入 (F,F,F) 作为 (Q,K,V),其中 Q,K,V 分别表示为查询(Query)、键(Key)、值(Value)。其中,$F,F^{(i)}$ 以及注意力模块被定义为:

$$F\in R^{(l_k+l_q)\times H\times d} \tag{11}$$

$$(F_Q^{(i)},F_K^{(i)},F_V^{(i)})=(F^{(i)}W_Q,F^{(i)}W_K,F^{(i)}W_V) \tag{12}$$

$$F_{\mathrm{attn}}^{(i)}=F^{(i)}+\mathrm{softmax}\left(\frac{F_Q^{(i)}(F_K^{(i)})^{\mathrm{T}}}{\sqrt{d_K}}\right)F_V^{(i)} \tag{13}$$

其中,d 表示特征维数,且 $d_K=d$。W_Q,W_K,W_V 分别表示 3 个全连接层的参数。键(Key)和值(Value)通过图片以及图片的扩增图片独立计算得出。通过 Transformer 对重点的特征向量权重进行提取后,按顺序进行特征信息拼接。拼接聚合中 S^r 和 Q^r 可被表示为:

$$S^r=\{(x_i,y_i,r)\mid y_i\in C,i=1,2,\cdots,l_k\} \tag{6}$$

$$Q^r=\{(x_i,y_i,r)\mid y_i\in C,i=1,2,\cdots,l_q\} \tag{7}$$

其中,$l_k=N\times K$,$l_q=N\times q$。经过单一元任务集的扩增后的 T^r 被表示为:

$$T^r=\{(x_i,y_i,r)\mid y_i\in C,i=1,\cdots,l_k,l_{k+1},\cdots,l_k+l_q\} \tag{8}$$

其中,前 l_k 个是在支持集 S^r 中被采样的,剩下的 l_q 个为从查询集 Q^r 中进行采样。例如,$\{S^4,Q^4\}$ 表示为对应图片进行部分裁剪的扩增集,对于一个元任务集 T^r 中的 (x_i,y_i,r_i),y_i 作为类标签参加监督学习以及它的 r_i 作为它的变换标签来进行自监督学习,在元任务集 T 生成一系列的扩增集之后,特征提取器 g_σ 会被应用于包括扩增的图像以及原图在内的每个图像中。对于图片的不同变换,采用 1 个实例级任务识别图片进行何种变换,利用自监督变换标签 r_i,模型若能较好地对这些图像的扩增方式进行识别,那么表明它可以较好地识别图像中的一些局部或者关键特征。例如,模型识别出进行水平翻转变换的图像,就要辨别图像中目标对象的姿态或者位置发生了改变。通过将自监督学习的特征进行转移,促进小样本分类任务更好地学习图像中的一些重要特征信息。在一个元任务集 T^r 中的 (x_i,y_i,r_i),考虑每个相对应的映射 $g_{\theta r}:x_i\to r_i$,$g_{\theta r}$ 表示变换分类器,θr 表示可学习参数,其 SSL 任务相对应的损失函数 L_c 表示为:

的特征表示为:

$$F_{\mathrm{all}}=[F^S,F^Q]\in R^{(l_k+l_q)\times H\times d} \tag{14}$$

$$F_{\mathrm{all}}=flatten(F_{\mathrm{attn}}) \tag{15}$$

其中,F^S,F^Q 分别表示支持集聚合特征和查询集聚合特征。$flatten(\cdot)$ 表示对 F_{attn} 的最后两个维度进行扁平操作。即对一张图像的原图以及它的不同扩增图像的重要特征片段进行连接聚合,之后将集成的特征输入到小样本分类分类器中进行分类。FSL 分类的损失函数表示为:

$$L_{all}=-\frac{1}{l_q}\sum_{i=1}^{l_q}\log\frac{\exp(-d(F_i^Q,h_{y_i}^g))}{\sum_{c\in C}\exp(-d(F_i^Q,h_c^g))} \tag{16}$$

其中,h_c^g 表示在支持集上进行计算,表示为:

$$h_c^g=\frac{1}{l_k}\sum_{i=1}^{l_k}F_i^S\cdot\rho(y_i=c) \tag{17}$$

综上,MGFAN 整体的优化函数为:

$$L=\alpha L_c+\beta L_{\mathrm{all}} \tag{18}$$

其中,α 与 β 为权重系数。

3.2.4　多距离函数度量模块

为了更加准确地计算两个样本间的相似度,使用多个距离函数来计算查询集样本与支持集中每个类别原型之间的距离,分别采用平方欧氏距离和曼哈顿距离,之后通过加权计算的方式获得最后的总差值。

设 z_i 为查询集中经过多图特征聚合模块处理后一个样本,z_i' 表示为一个类别的支持集全部样本经过多图特征聚合模块处理,之后再转化为类别原型。以此为例,需要计算一个元任务中全部查询集样本与支持集中各个原型间的距离,平方欧氏距离函数 $d_1(\cdot,\cdot)$ 表示为:

$$d_1(\cdot,\cdot)=(z_i-z_i')^2 \tag{19}$$

在理想情况下,平方欧氏距离可以扩大不同目标对象样本间

的距离，这有利于更好地区分不同的类别，但是受到数据集中一些噪音样本的干扰，这可能会对模型产生不利的影响。因此，在距离函数中新加曼哈顿距离函数。曼哈顿距离函数可以计算各个维度间的绝对距离，且不进行平方计算。曼哈顿距离函数$d_2(\cdot,\cdot)$表示为：

$$d_2(\cdot,\cdot)=|z_i-z_i'| \tag{20}$$

综上，最终的距离函数$d(\cdot,\cdot)$经过加权处理后被表示为：

$$d(\cdot,\cdot)=w_1 d_1(\cdot,\cdot)+w_2 d_2(\cdot,\cdot) \tag{21}$$

其中，w_1与w_2为权重系数。

4 实验

使用 miniImageNet[17]，tieredImageNet[28] 和 Stanford Dogs[29]这 3 个数据集对本文提出的 MGFAN 方法进行测试，并且与其他先进的方法进行了比较。

4.1 数据集和评价指标

本实验选取 3 个标准数据集对本文所提出方法进行验证。数据集的主要指标如下所示：miniImageNet 数据集的图片数量为 60000 张，类别数有 100 类，每个类别的数量为 600 张，其中训练集类数量为 64，验证集类数量为 16，测试集类数量为 20。tieredImageNet 中有 608 个类别，每个类别平均有 1282 张图像，其中训练集、验证集以及测试集类数量分别为 351，97 和 160 类。Stanford Dogs 数据集是大型数据集 ImageNet[7]的子类，其中含有 120 类犬类图片，在实验中，此数据集分别将 70，20，30 类用于模型的训练、验证以及测试。在数据传入网络进行训练时，图片的大小会被调整为 84×84。

实验采用 Top-1 平均准确率来衡量模型的性能，Top-1 准确率是指模型对一张图像进行预测，预测给出的结果中概率最大的是标签真实值，才认为是正确。

4.2 实验环境和参数设置

4.2.1 实验环境

多数实验的硬件配置为：CPU：i7-10700，运行内存 32 GB，NVIDIA GeForce RTX 3090(24 GB)。基于 Python3.7，PyTorch1.7 以及 CUDA11.1。

4.2.2 实验参数设置

在训练和测试每一个任务中，使用 N-way K-shot 的形式进行。具体而言，将数据分成 5-way 1-shot 和 5-way 1-shot 进行训练，在训练中把每个元任务作为一个 mini-batch，将 100 个任务作为 1 个 epoch，其中 MGFAN 方法总的迭代次数(epoch)为 200。置信区间为 95%，结果给出最终的 Top-1 平均准确率。

训练中选用的主要参数为：在 miniImageNet 中选用 conv4-64 和 ResNet-12 两种网络，其他数据集中采用 Conv4-64 为主干网络，其中 Conv4-64 表示输出的特征维数为 64，而 ResNet-12 输出的特征维数为 640。Conv4-64 网络中采用 Adam 优化器，ResNet-12 选用的优化器为 SGD，初始学习率(lr)为 0.0001，动量(momentum)为 0.95，权重衰减(weight_decay)为 0.0001，权重系数 $\alpha=0.6$，$\beta=1$，$w_1=0.6$，$w_2=0.4$。

4.3 对比方法

为了验证本文所提出方法的有效性，将其与 17 种流行的小样本分类方法进行对比。

(1)MatchingNet[17]：通过特征提取模块进行特征提取，使用 LSTM 网络对提取的特征进行进一步处理，接着通过距离度量网络来获取查询集样本与支持集样本的特征距离来实现小样本学习。

(2)ProtoNet[13]：求取支持集各类样本的平均特征向量，然后用这个平均向量来表示此类样本的原型，接着求取查询集样本与所有原型的距离，以此来判断查询集样本的类别。

(3)MAML[20]：在小样本学习中，模型学习一组“适应性较好”的初始参数，经过较少的迭代次数就可以使网络收敛。

(4)Relation Net[18]：提出一种非线性且可学习的相似度量方法，通过联立学习样本的特征映射与各个样本之间的距离度量模块来实现小样本学习。

(5)IMP[30]：提出无限的混合模型，不同于现有的原型网络使用单个集群表示一个类别，该方法在最近邻和原型网络中进行插值，提高了模型的鲁棒性和准确率。

(6)DN4[31]：不同于其他方法，DN4 采用的方法是比较相对应图片与类别之间的相对应的局部描述子，从而寻找与图像相似的类，以此来判定图像的类别归属。

(7)DN PARN[32]：使用一个可变形特征提取器得到支持集和查询集图像的两个对应特征图，然后对特征图分别求取互相关特征图和自相关特征图，最后对这些特征图进行级联后输入 CNN 网络中获取相应的得分。

(8)DSN-MR[33]：提出一种加入子空间概念的度量学习方法，找出各个类别相适应的子空间，之后从这个子空间中使用度量学习，对各个相对应的目标样本进行类别的预测。

(9)PCM[34]：提出一种新的网络，网络采用端到端模式，集成有双线性学习与分类映射模块。

(10)Neg-Cosine[35]：与大多数度量学习的做法不同，在度量学习中引入负边缘损失来进行小样本学习。

(11)IEPT[36]：提出一种新的框架，将实例级任务和元集级任务相结合，把自监督学习更好地结合到小样本学习中。

(12)TADAM[37]：为度量函数增加了放缩系数，增加了一个可学习的放缩系数来提高分类器的准确率。

(13)SNAIL[38]：提出了在小样本任务中结合时序空洞卷积和因果注意力，从而能够使用过往的经验。

(14)MetaOptNet[39]：利用线性分类器的优势来进行高维特征的嵌入和改进其泛化性。

(15)MTL[40]：提出使用 MTL 方法，该方法可以使用大数据集进行训练，得出 DNN 的初始权重，从而进行小样本学习。

(16)CAN[41]：将支持集和查询集样本特征进行独立提取，之后使用支持集和查询集的语义相关性来寻找目标对象；并且设计了一个元学习的计算器来计算支持集与查询集间的交叉注意力。

(17)Shot-Free[42]：提出一种不限样本数量的小样本学习方法，每个类别可以包含不同数量的样本，且在表示每个类别的表征的时候，使用显性的计算方式，而非使用均值。

4.4 实验结果与分析

4.4.1 在 miniImageNet 上的实验

在 miniImageNet 数据集中，使用 Conv4-64 以及 ResNet-12 作为主干网络进行实验，实验结果如表 1 所列。在 Conv4-64 和 ResNet-12 网络中，MGFAN 方法在 5-way 1-shot 以及 5-way 5-shot 设置中的准确率均超过表中所列出的其他

方法。相比于基线方法 ProtoNet，在 Conv4-64 网络中，本文所提出的方法在 5-way 1-shot 和 5-way 5-shot 设置中精确率分别提高了 8.20%和 6.69%，在 ResNet-12 网络中也分别提高了 4.27%和 2.85%。在 Conv4-64 网络中，在 5-way 1-shot 以及 5-way 5-shot 设置中，MGFAN 相比于最新提出的 IEPT，分别提高 1.36%和 0.98%。此外，我们注意到在 Conv4-64 网络中，本文方法的 5-way 1-shot 比 5-way 5-shot 准确率提高了 17.27%，因为随着支持集图片的增加，有标注的图片就越多，从而越有利于类均值特征的准确表达，使得小样本分类的难度就越小。

表 1　MGFAN 方法在 miniImageNet 上的实验结果

Table 1　Experimental results of MGFAN method on miniImage Net dataset

（单位：%）

方法	主干网络	5-way 1-shot	5-way 5-shot
MatchingNet[17]	Conv4-64	43.56±0.84	55.31±0.73
ProtoNet[13]	Conv4-64	49.42±0.78	68.20±0.66
MAML[20]	Conv4-64	48.70±1.84	63.10±0.92
Relation Net[18]	Conv4-64	50.04±0.80	65.30±0.70
IMP[30]	Conv4-64	49.60±0.80	68.10±0.80
DN4[31]	Conv4-64	51.24±0.74	71.02±0.64
DN PARN[32]	Conv4-64	55.22±0.84	71.55±0.66
DSN-MR[33]	Conv4-64	55.88±0.90	70.50±0.68
Neg-Cosine[35]	Conv4-64	52.84±0.76	70.41±0.66
IEPT[36]	Conv4-64	56.26±0.45	73.91±0.34
MGFAN	Conv4-64	**57.62±0.78**	**74.89±0.64**
ProtoNet[13]	ResNet-12	60.37±0.83	78.02±0.15
TADAM[37]	ResNet-12	58.50±0.30	76.70±0.38
SNAIL[38]	ResNet-12	55.71±0.99	68.88±0.92
DN4[31]	ResNet-12	54.37±0.36	74.44±0.29
MetaOptNet[39]	ResNet-12	62.64±0.61	78.63±0.46
MTL[40]	ResNet-12	61.20±1.80	75.50±0.80
CAN[44]	ResNet-12	63.85±0.48	79.44±0.34
Shot-Free[42]	ResNet-12	59.04±0.43	77.64±0.39
MGFAN	ResNet-12	**64.64±0.86**	**80.87±0.55**

注：黑体字表示在每种实验设置下的最优结果

4.4.2　在 tieredImageNet 上的实验

在 tieredImageNet 数据集中，采用 Conv4-64 作为主干网络，实验结果如表 2 所列。本文方法在 5-way 1-shot 以及 5-way 5-shot 中的准确率优于其他方法，且相比于基线方法 ProtoNet 分别提高了 1.99%和 2.55%。

表 2　tieredImageNet 数据集的实验结果

Table 2　Experimental results on tieredImageNet dataset

（单位：%）

方法	主干网络	5-way 1-shot	5-way 5-shot
MAML[20]	Conv4-64	51.67±1.81	70.30±0.08
ProtoNet[13]	Conv4-64	53.33±0.50	72.10±0.41
Relation Net[18]	Conv4-64	54.48±0.93	71.32±0.78
IMP[30]	Conv4-64	53.63±0.51	71.89±0.44
MGFAN	Conv4-64	**55.32±0.94**	**74.65±0.93**

注：黑体字表示在每种实验设置下的最优结果

4.4.3　在 Standford Dogs 上的实验

如表 3 所列，在 Standford Dogs 数据集中，使用 Conv4-64 为主干网络，相比于其他方法，MGFAN 方法取得最高的分类准确率，尤其是在 5-way 1-shot 和 5-way 5-shot 设置中，本文方法相比于基线方法 ProtoNet 准确率分别提高了 12.06%和 20.86%。结果表明，在 Stanford Dogs 数据集中，本文方法能够显著提高基线方法的分类准确率。

表 3　Stanford Dogs 数据集的实验结果

Table 3　Experimental results on Stanford Dogs dataset

（单位：%）

方法	主干网络	5-way 1-shot	5-way 5-shot
MatchingNet[17]	Conv4-64	35.80±0.99	47.50±1.03
MAML[20]	Conv4-64	44.81±0.34	56.68±0.31
ProtoNet[13]	Conv4-64	37.59±1.00	48.19±1.03
Relation Net[18]	Conv4-64	43.33±0.42	55.23±0.41
DN4[31]	Conv4-64	45.41±0.76	63.51±0.62
PCM[34]	Conv4-64	28.78±2.33	46.92±2.00
MGFAN	Conv4-64	**49.65±0.90**	**69.05±0.75**

注：黑体字表示在每种实验设置下的最优结果

在以上 3 个数据集中的实验结果表明：本文方法在大多数情况下优于现有的方法，且本文方法可以非常显著地提升基线方法的分类性能，说明 MGFAN 方法可以更好地捕捉图像中的重要特征信息，且多距离函数度量模块可以更加精确地计算未标记样本与原型点间的距离，进一步提高小样本分类的性能。

4.5　消融实验

4.5.1　不同扩增图对性能的影响

在本节中，为了验证本文提出的多图特征聚合模块中不同的扩增图对分类性能的影响，设计了 5 个实验来进行验证分析，使用的主干网络为 Conv4-64。其中，MGFAN$^{\#1}$表示只使用原图，无扩展图；MGFAN$^{\#2}$表示使用原图和原图的水平翻转扩增图；MGFAN$^{\#3}$表示使用原图、原图水平翻转和原图垂直翻转；MGFAN$^{\#4}$表示使用原图、原图水平翻转、原图垂直翻转以及原图 180 度旋转；MGFAN 表示使用原图、原图水平翻转、原图垂直翻转、原图 180°旋转以及将原图进行部分裁剪。在 miniImageNet 数据集上的实验结果如表 4 所列。

表 4　扩展图对分类性能影响的实验结果

Table 4　Experimental results of effect of extended graphs on classification performance

（单位：%）

方法	5-way 1-shot	5-way 5-shot
MGFAN$^{\#1}$(H=1)	54.83±0.85	72.16±0.72
MGFAN$^{\#2}$(H=2)	55.03±0.87	72.49±0.71
MGFAN$^{\#3}$(H=3)	55.70±0.86	73.19±0.70
MGFAN$^{\#4}$(H=4)	56.32±0.82	73.22±0.67
MGFAN(H=5)	**57.62±0.78**	**74.91±0.61**

注：黑体字表示在每种实验设置下的最优结果

由表 4 可得，相比于对原图使用多种不同几何变换实验，加入了一张对图片进行部分丢弃的 MGFAN 取得了最大的准确率提升。可以发现，随着扩增图的增加，分类准确率也在不断地提高，但是从 MGFAN$^{\#4}$之中增加了一张部分丢弃的扩增图之后，在 5-way 1-shot 设置中，MGFAN 的准确率较 MGFAN$^{\#4}$提高 1.31%，而提升速度排名第二位的 MGFAN$^{\#3}$→MGFAN$^{\#4}$准确率只提升了 0.62%；在 5-way 5-shot 设置中，MGFAN$^{\#4}$→MGFAN 准确率提升了 1.69%，而准确率提升排名第二的 MGFAN$^{\#2}$→MGFAN$^{\#3}$只提升了 0.70%。这可以说明，相比于使用传统的几何变换的扩增方法，使用了部分裁剪的图像通过将原图进行部分丢弃，可以

提供给模型一张更新的视野来提取特征，更有助于模型学习更多具有区分度的特征信息，可以较好地提升模型的分类能力，从而促使模型的分类准确率得到提升。

4.5.2 辅助任务模块对性能的影响

为了验证本文中加入的自监督学习模块对分类性能的影响，使用主干网络 Conv4-64，设计了两种实验：1）不加入辅助任务模块；2）在小样本学习中加入辅助任务模块。

表 5 列出了在 miniImageNet 数据集中，实验 1 和 2 的实验结果。

表 5 辅助任务对分类性能的影响

Table 5 Impact of auxiliary tasks on classification performance

（单位：%）

实验	L_{all}	L_c	5-way 1-shot	5-way 5-shot
1	√		56.14±0.86	73.66±0.67
2	√	√	**57.62±0.78**	**74.89±0.64**

注：黑体字表示在每种实验设置下的最优结果

在 5-way 1-shot 和 5-way 5-shot 的设置中，加入辅助任务模块的实验 2 的准确率都要明显高于未加入辅助任务模块的实验 1，分别提升了 1.48%和 1.23%。实验结果说明：自监督学习模块有助于小样本对特征的学习，改善小样本学习中监督信息的不足，通过将自监督学习任务加入小样本学习中可以促进模型获得更高的分类性能。

4.5.3 多距离函数度量模块对性能的影响

在本节中，为了验证本文提出的多距离度量函数模块对分类结果的影响，使用 Conv4-64 作为主干网络，在 3 个数据集上进行实验。表 6 中列出了不同的 w_1/w_2 权重系数对分类性能的影响。由表 6 可得，只使用平方欧氏距离函数进行计算相似度要优于只使用曼哈顿距离函数的。但是，在 $w_1=0.8, w_2=0.2$ 以及 $w_1=0.6, w_2=0.4$ 时模型的分类性能是优于只使用平方欧氏距离的，说明使用多距离函数对未知样本与原型点距离进行计算更为准确。在本文的实验中，大部分情况下使用 $w_1=0.6, w_2=0.4$ 性能是高于 $w_1=0.8, w_2=0.2$ 的，因此本文选取这个指标进行大部分实验。另外，在不同的任务中，也可以根据任务设置合适的权重比例系数，以此来使模型获取更好的分类性能。

表 6 多距离函数权重系数对分类性能的影响

Table 6 Influence of weight coefficients of multi-distance functions on classification performance

（单位：%）

w_1/w_2	miniImageNet		tieredImageNet		Standford Dogs	
	5-way 1-shot	5-way 5-shot	5-way 1-shot	5-way 5-shot	5-way 1-shot	5-way 5-shot
$w_1=1, w_2=0$	57.38±0.79	74.61±0.63	55.01±0.95	74.29±0.95	49.49±0.91	68.70±0.78
$w_1=0.8, w_2=0.2$	57.26±0.78	74.51±0.63	54.93±0.98	74.01±0.97	49.25±0.90	68.66±0.78
$w_1=0.6, w_2=0.4$	**57.62±0.78**	74.89±0.64	**55.32±0.94**	**74.65±0.93**	49.65±0.90	**69.05±0.75**
$w_1=0.4, w_2=0.6$	57.42±0.77	**74.91±0.61**	55.09±0.95	74.40±0.97	**49.75±0.89**	68.88±0.77
$w_1=0.2, w_2=0.8$	57.12±0.79	74.57±0.65	55.03±0.96	74.14±0.98	49.32±0.92	68.74±0.77
$w_1=0, w_2=1$	56.89±0.81	74.46±0.66	54.75±0.98	73.85±1.01	48.95±0.93	68.52±0.82

注：黑体字表示在每种实验设置下的最优结果

实验结果表明：合适地设置多距离度量函数中相对应的权重比例系数，可以使模型更为准确地判断未标记样本与原型点间的相似度，从而提高模型的分类性能。

4.5.4 扩增图的嵌入顺序对性能的影响

在本文中，利用原图以及原图的不同扩增图来增加关于图像的特征表达能力。为了探索多组特征向量间对次序的依赖性，设计了如下实验：在 miniImageNet 数据集中，使用 Conv4-64 作为主干网络，分别选择关于原图的 3 组特征向量进行实验，分别为原图、原图水平翻转和原图垂直翻转进行组合搭配，使用 MGFAN^{*1}（特征向量顺序为原图、原图水平翻转、原图垂直翻转）、MGFAN^{*2}（原图、垂直翻转、水平翻转）、MGFAN^{*3}（水平翻转、原图、垂直翻转）、MGFAN^{*4}（水平翻转、垂直翻转、原图）、MGFAN^{*5}（垂直翻转、原图、水平翻转）、MGFAN^{*6}（垂直翻转、水平翻转、原图）。实验结果如表 7 所列。从表 7 中可以发现，各组实验结果较为接近，属于正常的实验波动，向量排列次序对模型性能的影响较小，这是由于模型对同一个图像特征向量的提取是相同的，不论如何改变其排列顺序，都不影响其特征的提取，而且在距离度量时，也是对比相同维度间特征向量的差值，因此向量次序对其性能的影响较为有限。实验证明，模型并不依赖其特征图向量的排列顺序。

表 7 扩展图排列次序对分类性能影响的实验结果

Table 7 Experimental results on effect of extended graph arrangement order on classification performance

（单位：%）

方法	5-way 1-shot	5-way 5-shot
MGFAN^{*1}	55.70±0.86	73.19±0.70
MGFAN^{*2}	55.61±0.87	73.22±0.70
MGFAN^{*3}	55.83±0.85	73.10±0.71
MGFAN^{*4}	55.30±0.89	73.29±0.69
MGFAN^{*5}	55.48±0.89	73.08±0.73
MGFAN^{*6}	55.73±0.87	72.91±0.72

结束语 本文提出一种名为 MGFAN 的小样本分类方法，该方法首先通过多种数据增强方法获取关于图像中目标对象的不同分布的扩增图，之后通过自注意力模块聚合多个扩增图的特征向量，以此来获取更为准确的图像特征表达，接着探索了在小样本分类任务中加入关于预测图像发生何种扩增变化的自监督辅助任务对模型分类性能的影响，最后验证了多距离函数对最终分类结果的影响。实验结果表明，使用多函数距离函数有助于提升模型的分类准确率。

本文方法的计算开支会随着扩增图的增加而提高，接下来的工作中会进一步探索如何降低计算成本。

参 考 文 献

[1] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet

classification with deep convolutional neural networks[J]. Advances in Neural Information Processing Systems,2012,25.

[2] JI X,HENRIQUES J F,VEDALDI A. Invariant information clustering for unsupervised image classification and segmentation[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019:9865-9874.

[3] LIU W,ANGUELOV D,ERHAN D,et al. Ssd:Single shot multibox detector[C]// European Conference on Computer Vision. Cham:Springer,2016:21-37.

[4] CHEN L C,PAPANDREOU G,KOKKINOS I,et al. Deeplab:Semantic image segmentation with deep convolutional nets,atrous convolution,and fully connected crfs[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2017,40(4):834-848.

[5] NAM H,HAN B. Learning multi-domain convolutional neural networks for visual tracking[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016:4293-4302.

[6] HINTON G,LI D,DONG Y,et al. Deep neural networks for acoustic modeling in speech recognition:The shared views of four research groups[J]. IEEE Signal Processing Magazine,2012,29(6):82-97.

[7] RUSSAKOVSKY O,DENG J,SU H,et al. ImageNet large scale visual recognition challenge[J]. International Journal of Computer Vision,2015,115(3):211-252.

[8] FEI-FEI L,FERGUS R,PERONA P. One-shot learning of object categories[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2006,28(4):594-611.

[9] WANG Y,YAO Q,KWOK J T,et al. Generalizing from a few examples:A survey on few-shot learning[J]. ACM Computing Surveys(CSUR),2020,53(3):1-34.

[10] NICHOL A,SCHULMAN J. Reptile:a scalable metalearning algorithm[J]. arXiv:1803.02999,2018.

[11] KOCH G R,ZEMEL R,SALAKHUTDINOV R. Siamese neural networks for one-shot image recognition [C]// Proceedings of the 32nd Int conference on Machine Learning. New York:ACM,2015.

[12] WANG P,LIU L,SHEN C,et al. Multi-attention Network for One Shot Learning[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition(CVPR). IEEE,2017.

[13] SNELL J,SWERSKY K,ZEMEL R. Prototypical networks for few-shot learning [C]// Proceedings of the 31st Annual Conference on Neural Information Processing Systems. Cambridge,MA:MIT Press,2017:4077-4087.

[14] DOERSCH C,ZISSERMAN A. Multi-task self-supervised visual learning[C]// Proceedings of the IEEE International Conference on Computer Vision. 2017:2051-2060.

[15] KOMODAKIS N,GIDARIS S. Unsupervised representation learning by predicting image rotations[C]// International Conference on Learning Representations(ICLR). 2018.

[16] NOROOZI M,FAVARO P. Unsupervised learning of visual representations by solving jigsaw puzzles[C]// European Conference on Computer Vision. Cham:Springer,2016:69-84.

[17] VINYALS O,BLUNDELL C,LILLICRAP T,et al. Matching networks for one shot learning [C]// Proceedings of the 30th Annual conference on Neural Information Processing Systems. Cambridge,MA:MIT Press,2016:3630-3638.

[18] SUNG F,YANG Y,ZHANG L,et al. Learning to compare:Relation network for few-shot learning[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018:1199-1208.

[19] SANTORO A,BARTUNOV S,BOTVINICK M,et al. Meta-learning with memory-augmented neural networks [C]// Proceedings of the 33rd Int conference on Machine Learning. New York:ACM,2016:1842-1850.

[20] FINN C,ABBEEL P,LEVINE S. Model-agnostic meta-learning for fast adaptation of deep networks[C]// International Conference on Machine Learning. PMLR,2017:1126-1135.

[21] RAVI S,LAROCHELLE H. Optimization as a model for few-shot learning[C/OL] // Proceedings of the 5th Int Conference on Learning Representations. https://openreview.net/forum?id=rJY0-Kcll.

[22] LI Z,ZHOU F,CHEN F,et al. Meta-sgd:Learning to learn quickly for few-shot learning[J]. arXiv:1707.09835,2017.

[23] SHYAM P,GUPTA S,Dukkipati A. Attentive recurrent comparators[C]// International Conference on Machine Learning. PMLR,2017:3173-3181.

[24] DOERSCH C,GUPTA A,EFROS A A. Unsupervised visual representation learning by context prediction[C]// Proceedings of the IEEE International Conference on Computer Vision. 2015:1422-1430.

[25] PATHAK D,KRAHENBUHL P,DONAHUE J,et al. Context encoders:Feature learning by inpainting[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016:2536-2544.

[26] EVERINGHAM M,VAN GOOL L,WILLIAMS C,et al. The pascal visual object classes(voc) challenge[J]. International Journal of Computer Vision,2010,88(2):303-338.

[27] VASWANI A,SHAZEER N,PARMAR N,et al. Attention is all you need[J]. Advances in Neural Information Processing Systems,2017,30.

[28] REN M,TRIANTAFILLOU E,RAVI S,et al. Meta-learning for semi-supervised few-shot classification[J]. arXiv:1803.00676,2018.

[29] KHOSLA A,JAYADEVAPRAKASH N,YAO B,et al. Novel dataset for fine-grained image categorization:Stanford dogs[C]// Proc. CVPR Workshop on Fine-Grained Visual Categorization(FGVC). Citeseer,2011.

[30] ALLEN K,SHELHAMER E,SHIN H,et al. Infinite mixture prototypes for few-shot learning[C]// International Conference on Machine Learning. PMLR,2019:232-241.

[31] LI W,WANG L,XU J,et al. Revisiting local descriptor based image-to-class measure for few-shot learning[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019:7260-7268.

[32] WU Z,LI Y,GUO L,et al. Parn:Position-aware relation networks for few-shot learning[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019:6659-6667.

[33] SIMON C, KONIUSZ P, NOCK R, et al. Adaptive subspaces for few-shot learning[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 4136-4145.

[34] WEI X S, WANG P, LIU L, et al. Piecewise classifier mappings: Learning fine-grained learners for novel categories with few examples[J]. IEEE Transactions on Image Processing, 2019, 28(12): 6116-6125.

[35] LIU B, CAO Y, LIN Y, et al. Negative margin matters: Understanding margin in few-shot classification[C]// European Conference on Computer Vision. Cham: Springer, 2020: 438-455.

[36] ZHANG M, ZHANG J, LU Z, et al. IEPT: Instance-level and episode-level pretext tasks for few-shot learning[C]// International Conference on Learning Representations. 2021.

[37] ORESHKIN B, RODRÍGUEZ LÓPEZ P, LACOSTE A. Tadam: Task dependent adaptive metric for improved few-shot learning [J]. Advances in Neural Information Processing Systems, 2018, 31.

[38] MISHRA N, ROHANINEJAD M, CHEN X, et al. A simple neural attentive meta-learner[J]. arXiv: 1707. 03141, 2017.

[39] LEE K, MAJI S, RAVICHANDRAN A, et al. Meta-learning with differentiable convex optimization[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 10657-10665.

[40] SUN Q, LIU Y, CHUA T S, et al. Meta-transfer learning for few-shot learning[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 403-412.

[41] HOU R, CHANG H, MA B, et al. Cross attention network for few-shot classification[J]. Advances in Neural Information Processing Systems, 2019, 32.

[42] RAVICHANDRAN A, BHOTIKA R, SOATTO S. Few-shot learning with embedded class models and shot-free meta training [C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 331-339.

ZENG Wu, born in 1997, postgraduate. His main research interests include data augmentation and few-shot learning.

MAO Guojun, born in 1966, Ph.D, professor, is a member of China Computer Federation. His main research interests include data mining, big data and distribution computing.