

多级注意力特征网络的小样本学习

汪荣贵 韩梦雅 杨娟* 薛丽霞 胡敏

(合肥工业大学计算机与信息学院 合肥 230009)

摘要: 针对目前基于度量学习的小样本方法存在特征提取尺度单一, 类特征学习不准确, 相似性计算依赖标准度量等问题, 该文提出多级注意力特征网络。首先对图像进行尺度处理获得多个尺度图像; 其次通过图像级注意力机制融合所提取的多个尺度图像特征获取图像级注意力特征; 在此基础上使用类级注意力机制学习每个类的类级注意力特征。最后通过网络计算样本特征与每个类的类级注意力特征的相似性分数来预测分类。该文在Omniglot和MiniImageNet两个数据集上验证多级注意力特征网络的有效性。实验结果表明, 相比于单一尺度图像特征和均值类原型, 多级注意力特征网络进一步提高了小样本条件下的分类准确率。

关键词: 图像处理; 多尺度图像; 小样本学习; 多级注意力特征; 相似性度量

中图分类号: TN911.73; TP391.41

文献标识码: A

文章编号: 1009-5896(2020)03-0772-07

DOI: 10.11999/JEIT190242

Multi-level Attention Feature Network for Few-shot Learning

WANG Ronggui HAN Mengya YANG Juan XUE Lixia HU Min

(School of Computer and Information, Hefei University of Technology, Hefei 230009, China)

Abstract: Existing few-shot methods have problems that feature extraction scale is single, the learned class representations are inaccurate, the similarity calculation still relies on standard metrics. In order to solve the above problems, multi-level attention feature network is proposed. Firstly, the multiple scale images are obtained by scale processing, the features of multiple scale images are extracted and the image-level attention features are obtained by the image-level attention mechanism to fusion them. Then, class-level attention features are learned by using the class-level attention mechanism. Finally, the classification is performed by using the network to compute the similarity scores between features. The proposed method is evaluated on the Omniglot dataset and the MiniImagenet dataset. The experimental results show that multi-level attention feature network can further improve the classification accuracy under small sample conditions compared to the single-scale image features and average prototypes.

Key words: Image processing; Multi-scale images; Few-shot learning; Multi-level attention feature; Similarity metric

1 引言

近年来, 深度学习在人工智能任务中取得了良好的效果, 如目标检测^[1]、图像分类^[2]、机器翻译^[3]等。但是深度学习模型训练需要大规模的数据集, 且模型收敛缓慢。相比之下, 人类视觉系统能够在仅观察1个或几个实例之后识别新物体。人类视觉系统与深度学习模型之间的这种显著差距引起了

小样本学习^[4]的研究兴趣。小样本学习的目的是在每个类只有少量标注样本时构建一个分类模型, 使其具有较好的分类效果。

目前小样本学习的解决方案主要有以下3类: 数据增强、元学习和度量学习。小样本任务中每个类只有少量标签样本, 数据增强技术通过扩充数据集来缓解过拟合, 如生成样本^[5]或幻想特征^[6,7], 但是生成的样本或特征极其相似, 并没有很大地提升分类效果。元学习方法^[8-10]使用大量任务进行训练, 学习跨任务迁移知识, 实现在新的任务上可以快速学习, 因而适合小样本学习, 如学习好的初始化条件^[8], 跨任务的梯度更新策略^[9]或者通过循环神经网络(Recurrent Neural Network, RNN)构建外部记忆存储器^[10]以利用过去的经验。这些方法取得了较好的分类效果, 但是RNN会导致网络结构

收稿日期: 2019-04-11; 改回日期: 2019-09-05; 网络出版: 2019-09-17

*通信作者: 杨娟 yangjuan@hfut.edu.cn

基金项目: 国家自然科学基金(61672202), 国家自然科学基金-深圳联合基金(U1613217)

Foundation Items: The National Natural Science Foundation of China (61672202), The State Key Program of National Natural Science Foundation of China-Shenzhen Joint Foundation (U1613217)

复杂效率低下。另一类度量学习方法则是学习样本与特征之间的映射关系，将样本映射到特征空间，然后在空间内中寻找最近邻实现分类^[11-16]，并借助episode^[12]训练策略模拟测试环境以及归纳偏置思想为每个类学习一个原型表达^[13,14,16]提高小样本分类精度。具有代表性的方法是Vinyals等人^[12]提出的匹配网络，在支持集上使用注意力机制预测测试集类别，该算法可看作带权的最近邻分类器。Snell等人^[13]考虑归纳偏置的思想进一步提出原型网络，该算法学习一个度量空间，在空间内以类内样本特征的均值作为类的原型表达，计算样本特征与每个类原型的欧氏距离进行分类。

本文基于度量学习提出多级注意力特征网络，针对单一尺度图像特征忽略不同尺度图像下的细节信息以及均值类原型不能很好地评估类中各个支持集样本特征对于类原型的不同贡献，提出多级注意力特征概念。首先通过对图像进行处理获得多个尺度的图像，并提取多个尺度图像的特征使用图像级注意力机制学习图像级注意力特征，然后在类内样本的图像级注意力特征上使用类级注意力机制学习类级注意力特征。另外考虑到固定距离度量依赖学习的特征好坏导致错误分类的情况，采用网络计算特征之间的相似性关系，与特征联合学习更能捕获特征之间的相似性提高分类准确率。而且本文方法具有鲁棒性，不仅可以解决类别平衡条件下的小样本问题，还可以应用于类别不平衡条件下的小样本学习，即类内样本数量不相等的情况。

2 多级注意力特征网络

本节详细介绍所提多级注意力特征网络，本文方法使用多个尺度图像进行特征提取，通过多级注意力机制学习图像级注意力特征和类级注意力特征。在训练时使用episode训练策略，每个episode

从训练集中随机采样 N 个类构建一个支持集和一个查询集，每个类抽取 K 个样本作为支持集，称为 N -way K -shot小样本分类。支持集 $S = \{(x_i, y_i)\}_{i=1}^{N \times K}$ ，查询集 $Q = \{(x_j, y_j)\}_{j=1}^M$ ，其中 $y_i, y_j \in \{1, 2, \dots, N\}$ ， M 是查询集样本数量。下面将分为3个小节，分别从 K -shot分类、不平衡条件下的分类以及one-shot分类详细介绍本文方法。

2.1 K -Shot 分类

本文以 $N = 3, K = 5$ 为例介绍 K -shot分类任务，网络结构如图1所示，由特征模块 f 、类级注意力机制 p 以及度量模块 m 组成。首先通过特征模块 f 提取支持集图像的图像级注意力特征，然后对每个类的5个样本特征使用类级注意力机制 p 学习类级注意力特征，最后通过度量模块计算查询样本和每个类的类级注意力特征之间的相似性分数预测类别。下面将详细介绍5-shot分类中图像级注意力特征、类级注意力特征和相似性分数的学习。

2.1.1 图像级注意力特征

本文使用图像的多个尺度进行特征提取，并通过图像级注意力机制融合多个尺度图像的特征学习图像级注意力特征提高分类精度。通过实验证明使用3个尺度图像进行特征提取分类效果最好，所以本文采用3种尺度图像进行特征提取。提取图像级注意力特征的特征模块如图2所示，包含两个部分：提取不同尺度图像特征的卷积神经网络和融合多个尺度图像特征的图像级注意力机制。卷积神经网络结构如表1所示，C表示卷积层，MP表示最大池化层，由于3个分支的输入大小不同，为了方便后期在3个尺度图像特征上进行注意力操作，本文采用3个不同的网络分支获得维度相同的图像特征。图像级注意力机制是包含两个卷积块的注意力网络，以多个尺度图像特征的深度级联作为输入，学习图像级注意力特征。具体过程如下：给定一张

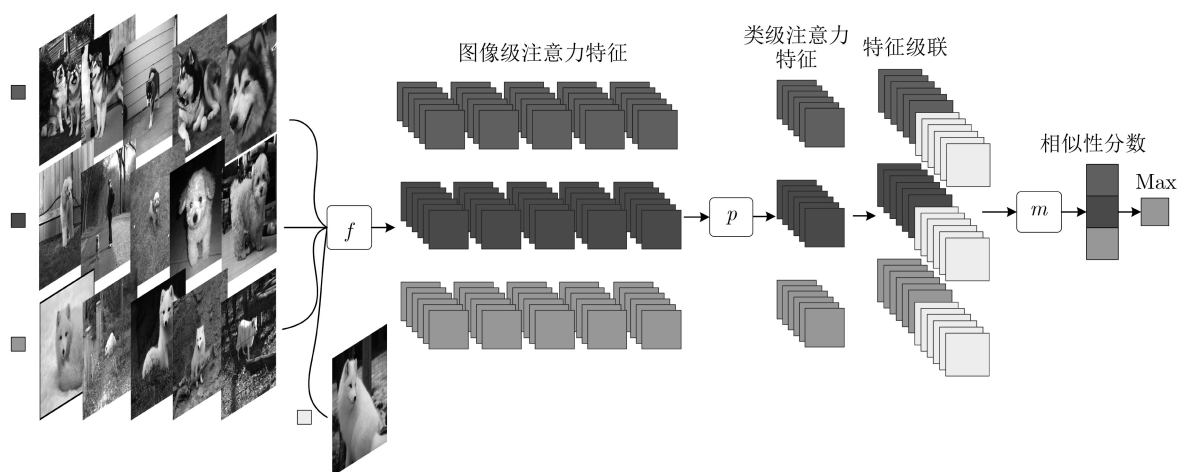


图1 5-shot 分类网络结构图

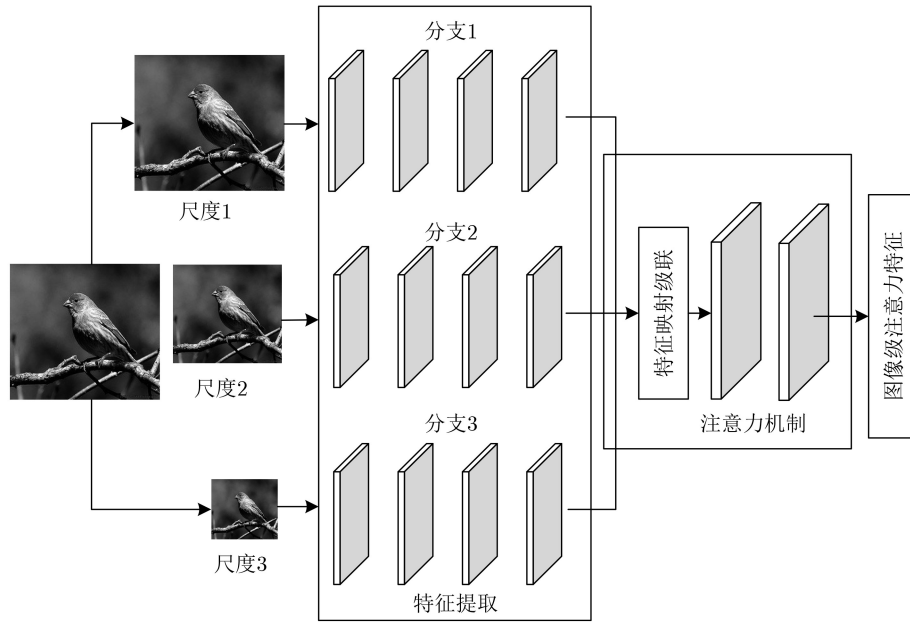


图2 特征模块

表1 不同尺度图像的特征提取网络分支结构

网络名	分支1	分支2	分支3
结构	$\begin{bmatrix} C: 3 \times 3, 64 \\ MP: 2 \times 2 \end{bmatrix}$	$\begin{bmatrix} C: 3 \times 3, 64 \\ MP: 2 \times 2 \end{bmatrix}$	$[C: 3 \times 3, 64]$
	$\begin{bmatrix} C: 3 \times 3, 64 \\ MP: 2 \times 2 \end{bmatrix}$	$[C: 3 \times 3, 64]$	$[C: 3 \times 3, 64]$
	$[C: 3 \times 3, 64]$	$[C: 3 \times 3, 64]$	$[C: 3 \times 3, 64]$
	$[C: 3 \times 3, 64]$	$[C: 3 \times 3, 64]$	$[C: 3 \times 3, 64]$

$m \times m$ 像素大小图像，首先对图像使用最大值采样方法进行多次下采样，第1次下采样即把原始图像 2×2 窗口内的图像变成1个像素，这个像素点的值是窗口内所有像素的最大值，得到 $m/2 \times m/2$ 像素大小图像，再对 $m/2 \times m/2$ 像素大小图像进行第2次下采样，得到 $m/4 \times m/4$ 像素大小图像。对 $m \times m$, $m/2 \times m/2$, $m/4 \times m/4$ 多个尺度图像使用卷积神经网络提取特征，然后在多个尺度图像特征上使用图像级注意力机制融合特征学习图像级注意力特征。

$$f(x_i) = f_{\varphi_2}(c(f_{\varphi_1}^1(x_i^1), f_{\varphi_1}^2(x_i^2), \dots, f_{\varphi_1}^l(x_i^l))) \quad (1)$$

这里 x_i^l 是图像 x_i 经过下采样之后需要通过第 l 个分支提取特征的输入图像， $f_{\varphi_1}^l$ 是卷积神经网络中的第 l 分支网络， $f_{\varphi_1}^l(x_i^l)$ 是通过第 l 个网络分支提取的图像特征， c 函数执行多个尺度特征的深度级联， f_{φ_2} 是学习图像级注意力特征的注意力网络， $f(x_i)$ 是通过整个特征模块学习到的图像 x_i 的图像级注意力特征。

2.1.2 类级注意力特征

小样本学习一个关键问题是如何从每个类少量

的一组图像中学习一个高质量的类特征。原型网络在度量空间内以类内样本特征的简单平均作为类原型。但是小样本图像数据集中类内样本差异较大，有些样本目标不明显，例如样本图片的背景较大目标较小、只包含部分目标、目标被部分遮挡等情况。这些情况下并不能很好地学习到样本的目标特征，其特征相比于样本目标明显的特征会较为偏离类原型，所以其对原型的贡献不应和样本目标明显的特征一致。本文通过类级注意力机制融合类内样本的特征，学习每个类的类级注意力特征，能够更好地表达原型。类级注意力机制是同图像级注意力机制结构相同的注意力网络，以类内样本特征的深度级联作为输入，学习每个类的类级注意力特征。通过类级注意力机制融合类内样本特征，可以学习与类最相关的部分。

$$p_n = p_{\varphi_3}(c(f(x_1), f(x_2), \dots, f(x_K))) \quad (2)$$

这里 $f(x_1), f(x_2), \dots, f(x_K)$ 是类别 n 的 K 个样本的图像级注意力特征， c 函数执行类内样本特征的深度级联， p_{φ_3} 是用于学习类级注意力特征的注意力网络。 p_n 是通过类级注意力机制融合类内样本特征学习到的第 n 个类的类级注意力特征， $n \in \{1, 2, \dots, N\}$ 。

2.1.3 相似性度量

大多数小样本学习方法使用固定的距离度量计算特征之间的相似性关系，例如欧氏或余弦距离，只学习了特征嵌入，然后在给定学习嵌入的情况下使用固定度量，在很大程度上取决于学习的特征嵌入，当嵌入网络学习的特征区分信息不充分时往往受限。本文通过灵活的函数逼近器来学习相似性，

以数据驱动的方式学习一个好的度量，而不必手动选择度量(欧氏、余弦)，可以更好地比较数据。计算相似性的度量模块包含两个卷积块，两个全连接层和1个sigmoid函数，每个卷积块后跟1个最大池化层。度量模块以查询样本的图像级注意力特征和每个类的类级注意力特征的深度级联作为输入，生成一个0~1的分数表示查询样本和每个类之间的相似性。在这里输入度量模块的特征是一个3维的特征图，而不是1维的特征向量，因为1维的特征向量只包含高级语义信息，而3维的特征图还包含着图像的部分信息，适合使用神经网络学习特征之间的关系。

$$r_{j,n} = m_{\varphi_4}(c(f(x_j), p_n)) \quad (3)$$

这里 m_{φ_4} 是计算相似性的神经网络， $r_{j,n}$ 是查询图像 x_j 和类别 n 之间的相似性分数。

2.1.4 损失函数

对于大规模图像分类或者回归任务，采用均方差损失函数(Mean Squared Error loss function, MSE)作为损失函数，目标函数为

$$J_{\text{MSE}}(\varphi) = \sum_{j=1}^M \sum_{n=1}^N (r_{j,n} - z)^2 \quad (4)$$

在小样本分类任务中，由于样本量稀疏，基本损失函数(均方差损失函数)训练的网络模型在测试时会因为数据量少产生过拟合。本文使用均方误差损失 $J_{\text{MSE}}(\varphi)$ 加上L2正则化约束项作为目标函数训练模型，增强模型的鲁棒性。损失函数为

$$J(\varphi) = J_{\text{MSE}}(\varphi) + \lambda(\|\varphi\|_2^2) \\ = \sum_{j=1}^M \sum_{n=1}^N (r_{j,n} - z)^2 + \lambda(\|\varphi\|_2^2) \quad (5)$$

其中， λ 是正则化项惩罚系数， $z = \begin{cases} 0, & y_j \neq n \\ 1, & y_j = n \end{cases}$ ， y_j 是图像 x_j 的类别标签， n 是类别标签。

反向传播梯度求导公式为

$$\frac{dJ(\varphi)}{d\varphi} = \frac{dJ_{\text{MSE}}(\varphi)}{d\varphi} + \frac{d\lambda(\|\varphi\|_2^2)}{d\varphi} = \nabla J_{\text{MSE}}(\varphi) + \lambda\varphi \quad (6)$$

加上正则化前梯度下降公式

$$\varphi = \varphi - \eta \nabla J_{\text{MSE}}(\varphi) \quad (7)$$

加上L2正则化项后梯度下降公式

$$\varphi = \varphi - \eta(\nabla J_{\text{MSE}}(\varphi) + \lambda\varphi) = (1 - \eta\lambda)\varphi - \eta \nabla J_{\text{MSE}}(\varphi) \quad (8)$$

这里 φ 是整个网络模型的可学习参数， η 是学习率。式(5)中当图像 x_j 的类别标签 $y_j = n$ 时常数项 z 为1， $y_j \neq n$ 常数项 z 为0。因此若图像 x_j 属于 n 类，只有当图像 x_j 的图像级注意力特征 $f(x_j)$ 与类 n 的类级注意力特征 p_n 相似性分数接近于1，与其他类别的类特征的相似性分数接近于0时，损失函数最小。因此，经过反向传播训练得到的网络模型会使得同类样本的相似性分数趋向于1，不同类样本的相似性分数趋向于0，使得分类更加准确。当加入正则化项后，参数的权重衰减使得网络权值不断减小，增强了模型的抗扰动能力，减轻测试数据差异情况下的过拟合。

2.2 不平衡条件下的分类

在实际生活中可能会存在类别不平衡的情况，即类内样本数量不同。本文方法能解决类别不平衡条件下的小样本分类任务，以3个类的小样本分类为例，支持集中有的类含有3个样本，有的类含有2个样本，网络结构如图3所示。首先通过特征模块 f 提取支持集样本的图像级注意力特征，然后计算每个类的类内样本特征均值再通过类级注意力机制 p 学习每个类的平衡类级注意力特征。最后通过度量模块 m 计算查询样本特征和每个类的类级注意力特征之间的相似性分数，取相似性分数最高的类别作为预测类别。

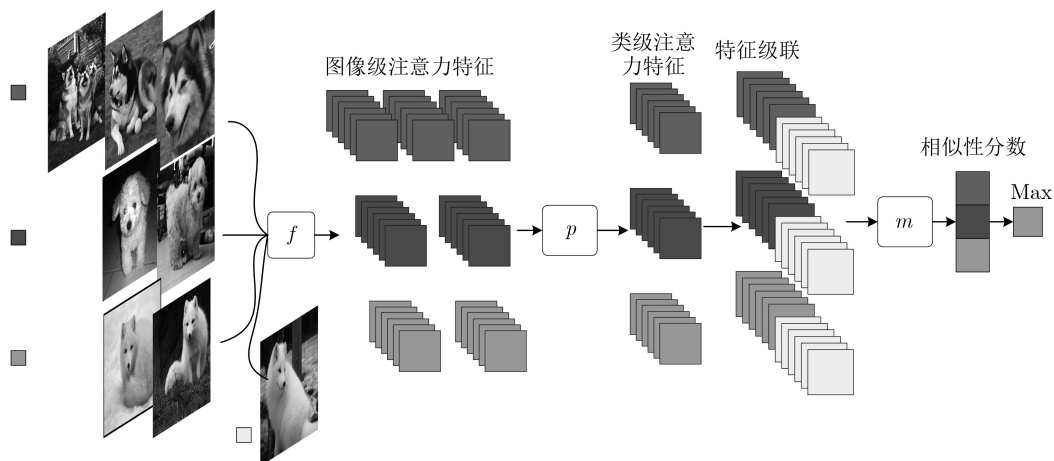


图3 类别不平衡条件下的小样本分类网络结构图

2.3 one-shot分类

one-shot分类支持集中每个类只有1张样本, 类特征即样本特征本身, 因此本文方法应用于one-shot分类任务只包含特征模块 f 和度量模块 m , 如图4所示。对于5-way 1-shot任务, 每个episode抽取5个类, 每个类抽取1张样本作为支持集, 其余样本作为查询集。给定一张查询图像, 首先通过特征模块提取支持集样本和查询图像的图像级注意力特征, 然后在深度上级联查询图像特征和每个支持集样本特征, 再通过度量模块计算特征之间的相似性分数, 取相似性分数最高的类别作为查询图像的预测类别。

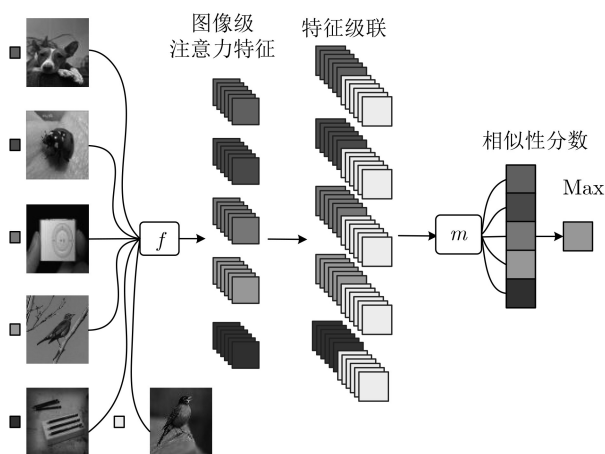


图4 one-shot 分类网络结构图

3 实验

3.1 数据集

Omniglot数据集是由50个字母表共计1623类字符组成, 每类字符包含20个样本, 由不同的人书写而成。MiniImagenet数据集是由匹配网络提出, 从ILSVRC-2012数据集中随机采样100个类, 每类600张样本组成。

3.2 实验设置

本文在Ubuntu15.04操作系统上使用开源深度学习框架PyTorch进行实验, 并使用GPU加速计

算。在所有实验中, 使用初始学习率为 10^{-3} 的Adam算法进行训练, 每10000次迭代学习率减半。在Omniglot和MiniImagenet数据集上与小样本分类的经典算法进行对比实验, 包括MANN, Meta-LSTM, MAML, 关系网络(Relation Net), 原型网络(Prototypical Net), 匹配网络(Matching Net)等。

在实验中将Omniglot数据集分为两个部分, 1200类用于训练, 423类用于测试, 所有图像大小均设置为 28×28 。并采用MAML提出的MiniImagenet数据集划分方法, 将MiniImagenet数据集分为3个部分, 64类用于训练, 16类用于验证, 20类用于测试, 所有图像大小均被设置为 84×84 。在Omniglot数据集上进行5-way 1-shot, 5-way 5-shot, 20-way 1-shot与20-way 5-shot分类任务, 以测试集随机生成的1000个episode的分类准确率取平均作为在Omniglot数据集上的分类准确率。在MiniImagenet数据集上分别进行5-way 1-shot和5-way 5-shot分类任务, 以测试集随机生成的600个episode的分类准确率取平均作为在MiniImagenet数据集上的分类准确率。

3.3 实验结果

3.3.1 Omniglot数据集

本文与小样本学习的经典算法在Omniglot数据集上的对比结果如表2所示, 由实验结果可看出本文算法在5-way 1-shot, 20-way 1-shot与20-way 5-shot分类任务上都取得了较高的分类准确率。由于MAML方法进行了模型微调, 在5-way 5-shot分类任务上的分类准确率高于本文方法, 而本文方法没有进行模型微调。

3.3.2 MiniImagenet数据集

本文与其他小样本学习方法在MiniImagenet数据集上的对比结果见表3, 由实验结果可看出本文算法在5-way 1-shot与5-way 5-shot任务上都取得了较好的性能。对于5-way 5-shot实验, 由于原型网络在训练过程中每个episode使用20个类, 每类

表2 Omniglot数据集上的小样本分类准确率(%)

方法	微调	5-way 分类准确率		20-way 分类准确率	
		1-shot	5-shot	1-shot	5-shot
MANN	否	82.8	94.9	—	—
MATCHING NETS	是	97.9	98.7	93.5	98.7
PROTOTYPICAL NETS	否	98.8	99.7	96.0	98.9
MAML	是	98.7±0.4	99.9±0.1	95.8±0.3	98.9±0.2
RELATION NET	否	99.6±0.2	99.8±0.1	97.6±0.2	99.1±0.1
本文方法	否	99.6	99.7	97.8	99.2

15个查询样本，因此分类准确率高于本文方法。而本文方法每个episode使用5个类，每类5个查询样本进行训练，样本数量比原型网络少很多，当原型网络每个episode使用5个类进行训练时，分类准确率为65.77%，低于本文方法。此外本文还在MiniImagenet数据集上进行了类别不平衡条件下的小样本分类实验，以5-way分类任务为例，将5个类的样本数量分别设为1, 2, 3, 4, 5进行测试，分类准确率为61.44%，验证了本文方法在类别不平衡条件下也可以达到很好的分类效果。

3.4 实验分析

3.4.1 类级注意力特征实验分析

本节针对2.1.2节所提类级注意力特征进行分析，与均值类原型和简单求和学习类特征的方法进行对比。从表4可以看出类级注意力特征方法使得分类准确率有所提升。分析原因可知，小样本图像分类任务类内差异大，类内有些样本类特征明显，因此每个样本在生成类特征时贡献度应当是有差异的。本文方法通过类级注意力机制融合类内样本特征，能够学习到与类最相关的部分。

3.4.2 图像级注意力特征实验分析

本节针对2.1.1节所提图像级注意力特征进行分析，与单尺度图像特征和两尺度图像特征进行对比。实验结果如表5所示，使用图像级注意力特征具有更高的分类准确率。分析原因可知，单个尺度只能提取图像的全局特征，而多个尺度既能提取图像的全局特征还能提取图像的特征细节，因此图像级注意力特征具有更多的鉴别信息，能够提高分类正确率。

表3 MiniImagenet数据集上的小样本分类准确率(%)

方法	微调	5-way分类准确率	
		1-shot	5-shot
MATCHING NETS	否	43.56±0.84	53.11±0.73
META-LEARN LSTM	否	43.44±0.77	60.60±0.71
MAML	是	48.70±1.84	63.11±0.92
PROTOTYPICAL NETS	否	49.42±0.78	68.20±0.66
RELATION NETS	否	50.44±0.82	65.32±0.70
本文方法	否	53.18±0.80	66.72±0.71
本文方法(L2正则化)	否	54.56±0.81	67.39±0.68

表4 MiniImageNet数据集上类特征方法的对比(%)

类特征	5-way 5-shot 分类准确率
本文方法(均值类原型)	65.80±0.65
本文方法(求和)	65.56±0.66
本文方法(类级注意力特征)	66.43±0.68

3.4.3 多尺度实验分析

本节针对所使用的图像金字塔多尺度机制进行分析，与特征金字塔网络的多尺度方法和不同大小卷积核分别提取图像特征再融合的多尺度方法进行比较。实验结果如表6所示，本文方法能够利用多个尺度图像下的信息，取得最好的分类效果。与特征金字塔网络的多层特征融合方法相比，分类准确率提高0.89%。与不同大小卷积核提取图像特征融合的方法相比，准确率提高1.1%。

3.4.4 相似性度量实验分析

本节验证2.1.3节使用的相似性度量方法，与欧氏距离，余弦距离度量方式进行对比实验。由于本文实验中学习的特征是一个3维的特征图，适合使用网络计算相似性关系，当使用距离度量计算相似性时需要把3维的特征图先转换为1维的特征向量，并不具有高级语义信息，因此使用欧氏距离和余弦距离效果较差，实验结果如表7所示。欧氏距离和余弦距离这类设置好固定度量方式的方法，依赖于学习到的图像特征，而本文方法通过网络计算特征间相似性分数，与特征提取一起端到端的学习更能捕获特征之间的相似性。

4 结束语

本文提出多尺度条件下的多级注意力融合特征的小样本图像分类算法，针对单一尺度图像特征忽略不同尺度图像下的细节信息提出改进的特征模

表5 MiniImageNet数据集上图像特征方法的对比(%)

图像特征	5-way 分类准确率	
	1-shot	5-shot
本文方法(单尺度特征)	52.20±0.82	66.43±0.68
本文方法(两尺度特征)	53.93±0.79	66.89±0.71
本文方法(图像级注意力特征)	54.56±0.81	67.39±0.68

表6 MiniImageNet数据集上多尺度方式对比(%)

多尺度方法	5-way 分类准确率	
	1-shot	5-shot
特征金字塔网络	53.42±0.76	66.50±0.69
不同卷积核	53.27±0.83	66.29±0.66
本文方法	54.56±0.81	67.39±0.68

表7 MiniImageNet数据集上相似性度量方法的对比(%)

度量方式	5-way 分类准确率	
	1-shot	5-shot
本文方法(欧氏距离)	48.43±0.78	63.52±0.71
本文方法(余弦相似度)	46.54±0.82	60.50±0.70
本文方法(网络计算)	54.56±0.81	67.39±0.68

块,使用多个尺度的图像进行特征提取并通过图像级注意力机制融合特征,进一步学习样本更细节的特征表达。针对类中目标不明显样本导致错误分类的情况提出类级注意力机制,使用注意力网络学习类级注意力特征的方法降低目标不明显样本对类特征的影响,提高类特征的鲁棒性从而完成正确分类。针对相似性计算使用固定度量依赖学习的特征嵌入情况提出改进的度量方法,使用网络计算相似性分数与特征嵌入联合学习提高分类准确率。本文进行详细的实验验证了多级注意力特征网络在 Omniglot和MiniImagenet数据集上的优秀表现。

参 考 文 献

- [1] GIRSHICK R. Fast R-CNN[C]. 2015 IEEE International Conference on Computer Vision, Santiago, Chile, 2015: 1440–1448. doi: [10.1109/ICCV.2015.169](https://doi.org/10.1109/ICCV.2015.169).
 - [2] HUANG Gao, LIU Zhuang, VAN DER MAATEN L, *et al.* Densely connected convolutional networks[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA, 2017: 2261–2269. doi: [10.1109/CVPR.2017.243](https://doi.org/10.1109/CVPR.2017.243).
 - [3] HE Di, XIA Yingce, QIN Tao, *et al.* Dual learning for machine translation[C]. The 30th Conference on Neural Information Processing Systems, Barcelona, Spain, 2016: 820–828.
 - [4] LI Feifei, FERGUS R, and PERONA P. One-shot learning of object categories[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, 28(4): 594–611. doi: [10.1109/TPAMI.2006.79](https://doi.org/10.1109/TPAMI.2006.79).
 - [5] MEHROTRA A and DUKKIPATI A. Generative adversarial residual pairwise networks for one shot learning[EB/OL]. <https://arxiv.org/abs/1703.08033>, 2017.
 - [6] DIXIT M, KWITT R, NIETHAMMER M, *et al.* AGA: Attribute-guided augmentation[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA, 2017: 7455–7463. doi: [10.1109/CVPR.2017.355](https://doi.org/10.1109/CVPR.2017.355).
 - [7] HARIHARAN B and GIRSHICK R. Low-shot visual recognition by shrinking and hallucinating features[C]. 2017 IEEE International Conference on Computer Vision, Venice, Italy, 2017: 3037–3046. doi: [10.1109/iccv.2017.328](https://doi.org/10.1109/iccv.2017.328).
 - [8] FINN C, ABBEEL P, and LEVINE S. Model-agnostic meta-learning for fast adaptation of deep networks[C]. The 34th International Conference on Machine Learning, Sydney, Australia, 2017: 1126–1135.
 - [9] RAVI S and LAROCHELLE H. Optimization as a model for few-shot learning[EB/OL]. <https://openreview.net/forum?id=rJY0-Kc1l>, 2017.
 - [10] SANTORO A, BARTUNOV S, BOTVINICK M, *et al.* Meta-learning with memory-augmented neural networks[C]. The 33rd International Conference on Machine Learning, New York, USA, 2016: 1842–1850.
 - [11] KOCH G. Siamese neural networks for one-shot image recognition[EB/OL]. <http://www.cs.utoronto.ca/~gkoch/files/msc-thesis.pdf>, 2015.
 - [12] VINYALS O, BLUNDELL C, LILLICRAP T, *et al.* Matching networks for one shot learning[C]. The 30th Conference on Neural Information Processing Systems, Barcelona, Spain, 2016: 3630–3638.
 - [13] SNELL J, SWERSKY K, and ZEMEL R. Prototypical networks for few-shot learning[C]. The 31st Conference on Neural Information Processing Systems, Long Beach, USA, 2017: 4080–4090.
 - [14] SUNG F, YANG Yongxin, ZHANG Li, *et al.* Learning to compare: Relation network for few-shot learning[C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018: 1199–1208. doi: [10.1109/cvpr.2018.00131](https://doi.org/10.1109/cvpr.2018.00131).
 - [15] WANG Peng, LIU Lingqiao, and SHEN Chunhua. Multi-attention network for one shot learning[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA, 2017: 6212–6220. doi: [10.1109/CVPR.2017.658](https://doi.org/10.1109/CVPR.2017.658).
 - [16] HILLIARD N, HODAS N O, and CORLEY C D. Dynamic input structure and network assembly for few-shot learning[EB/OL]. <https://arxiv.org/abs/1708.06819v1>, 2017.
- 汪荣贵: 男, 1966年生, 教授, 研究方向为智能视频处理与分析、视频大数据与云计算等。
- 韩梦雅: 女, 1996年生, 硕士生, 研究方向为深度学习、计算机视觉等。
- 杨 娟: 女, 1983年生, 讲师, 研究方向为视频信息处理、视频大数据处理技术等。
- 薛丽霞: 女, 1976年生, 副教授, 研究方向为视频大数据检索与分析。
- 胡 敏: 女, 1967年生, 教授, 研究方向为计算机视觉、数字图像处理等。