

融合类增强与多尺度自适应的小样本学习方法

董驰静, 张孙杰, 任 涵

(上海理工大学光电信息与计算机工程学院, 上海 200093)

摘要: 为了解决小样本学习存在特征信息提取不足、难以准确地捕获局部明显特征信息的问题, 提出了一种融合类增强与多尺度自适应的小样本学习方法。首先在特征的层面上对图像进行类增强, 通过将特征图的每次激活与其邻域相关联来编码丰富的语义结构, 使提取后的类内特征明显, 更利于当前的分类任务。其次通过多尺度特征生成来提取不同尺度上图像特征的低层表示。最后对每个尺度上的语义相关矩阵进行权重分配与相似元素最大化计算查询图像与各支持集类别图像之间的语义相似度, 多尺度信息进行融合后, 对目标图像进行分类。在 5-way 1-shot 和 5-way 5-shot 设置中, 该方法在 miniImageNet 数据集上的均值平均精度 (mean Average precision, mAP) 分别为 56.83% 和 75.76%, 在常用细粒度图像数据集 Stanford Cars 和 CUB-200-2011 分类基准上分别达到了 79.33% 和 93.92%、66.33% 和 85.78%, 均优于现有方法的最好结果。

关键词: 小样本学习; 类增强; 多尺度特征生成; 自适应任务注意

中图分类号: TP301 **文献标志码:** A

Few-Shot Learning Method Based on Class Enhancement and Multi-scale Adaptation

DONG Chijing, ZHANG Sunjie, REN Han

(School of Optical-Electrical and Computer Engineering, University of Shanghai for Science & Technology, Shanghai 200093, China)

Abstract: In order to solve the problems of the insufficient feature information extraction and the difficulty in capturing local obvious feature information accurately in few-shot learning, a method combining class enhancement and multi-scale adaptation is proposed. Firstly, the class enhancement is performed on the image at the level of features, and rich semantic structures are encoded by associating each activation of the feature map with its neighborhood, thus making the extracted intra class features obvious and more conducive to the current classification task. Secondly, low-level representations of image features at different scales are extracted through multi-scale feature generation. Finally, the semantic correlation matrix on each scale is weighted and similarity elements are maximized to calculate the semantic similarity between the query image and each support set category image. After the fusion of multi-scale information, the target images are classified. In the 5-way 1-shot and 5-way 5-shot settings, the mean average precision (mAP) of this method on the miniImageNet dataset is 56.83% and 75.76% respectively, and it achieves 79.33% and 93.92%, 66.33% and 85.78% on the commonly used fine grained image dataset Standard

Cars and CUB-200-2011 classification benchmarks, respectively, which are superior to the best results of the existing methods.

Key words: few-shot learning; class enhancement; multi-scale feature generation; adaptive task attention

引 言

以大规模数据集为前提,深度学习已经在图像分类、目标检测及文本分析等领域有着广泛的应用^[1]。然而其往往需要大量的训练数据才能在各种图像任务中达到令人满意的性能水平。其次在实际应用场景中,由于人工标记的成本、隐私、安全或道德的敏感性等问题,相关研究者在某种程度上无法获得大规模且具有高质量的数据及标签。当数据量过少时,模型往往会过度拟合于训练样本,而在新的类别上泛化性能下降^[2]。元学习对于新任务能够有较好的表现。因此,许多基于元学习的小样本学习^[3-6]的方法已被提出,主要有如下3类方向。

基于度量的元学习方法旨在学习将图像映射到度量空间的嵌入函数,从而基于图像对的距离来区分图像对之间的相关性。Koch等^[7]使用孪生神经网络解决 one-shot 学习问题。Snell等^[8]提出了原型网络,该网络首先假设每种类型都可以由原型表示,并且可以通过计算每个类的嵌入表示的平均值,然后使用距离函数进行分类来获得原型。事实上,研究过程中并不知道哪个距离函数最好。因此,Sung等^[9]提出一种关系网络,以通过学习获得最合适的距离度量函数。上述方法都基于图像级别的特征表示。由于样本数量的稀缺,不能很好地表示每个类别在图像级特征上的分布。最近的一些工作,如DN4^[3]和CovaMNet^[10]表明,丰富的低层特征具有更好的表示能力。然而,这些方法独立地测量查询图像和每个支持类之间的相似性,而不同时考虑整个任务。此外,基于低级特征的方法仅在单个尺度上测量查询图像和支持集之间的相似性,这可能导致当主要对象的尺度不同时分类精度较低。

基于初始化的元学习方法的主要思想是如何合理地利用现有经验或知识,在面对新任务时实现快速学习,而不是从头开始,学习如何在给定少量支持样本的情况下在线快速更新模型。Santoro等^[11]采用了长短期记忆网络(Long short-term memory, LSTM)来控制网络和外部存储模块之间的交互,通过将随机梯度下降(Stochastic gradient descent, SGD)更新规则解释为具有可训练参数的重写门控模型,描述了一种新的元学习方法。MAML(Model-agnostic meta-learning)^[12]的目的是学习良好的参数初始化,以便模型能够快速适应新任务。尽管这些基于元学习的方法在小样本学习任务中取得了优异的成绩,但它们复杂的内存寻址结构很难训练。

基于数据增强的元学习方法旨在利用生成器网络从有限的训练数据中创建新的训练样本,或者利用外部信息来支持少量学习任务,改善小样本训练中数据匮乏的问题。迁移学习方法最近表明,早期预训练和后续微调的标准迁移学习是使用深度骨干网络进行小样本学习的强大基线。

以上大多数方法存在一个缺陷:模型判断图片“异同”时,对于特征提取后的矩阵向量直接平均加权,而没有赋予当处理不同 5-way 1-shot 任务时对关键特征进行权重分配。简而言之,对于异类的 5-way 1-shot,区别特征是不同的,如查询集图像为狗类,当与支持集图像进行对比时,带有尾巴的可能为狗类,此时尾巴为关键特征。显然,具有清晰类特征的图像更易于分类。因此,在特征提取时,应对图像类特征进行增强并尽可能多地产生低层特征表示,方便之后随着分类任务的不同对不同的关键局部表示进行权重分配以达到提高模型图像分类的效果。另一方面,小样本学习一直存在难于准确地表示数据分布,使得分类模型误差较大且泛化能力差等问题,类增强模块将基本表示转换为其自相关张量,通过将特征图的每次激活与其邻域相关联来编码丰富的语义结构,形成的深度特征图使得图像的相关结构部分脱颖而出,可以很好地缓解小样本学习关于数据分布难于准确表示的问题。

本文提出了一种结合类增强的元学习方法和自适应多尺度特征融合的方法进行小样本图像分类。模型采用 Conv-4-64 前两个卷积块作为 backbone,并将类增强特征与原始图像特征进行融合之后进行特征提取。利用通道改变和全局平均池化来实现多尺度特征尽可能保留语义信息。之后经过自适应任务注意模块得到经过加权处理的相似性向量来对图像进行分类。

1 融合类增强与多尺度自适应的小样本学习方法

1.1 问题定义

在小样本学习中,通常将数据分为3组:查询集 Q 、支持集 S 和辅助集 A 。常将数据分为两类:基类和新类,其中辅助集 A 归于基类,查询集 Q 和支持集 S 为新类。新类共享图像标签,并且新类与基类的标签空间没有交集。

因此,遵循常见的少样本学习任务的定义。给定支持集 $S = \{s_i, l_i\}_{i=1}^n$,其中 $n = K \times N$,表示支持集 S 中有 N 种类别,每个类别包含 K 张图像。查询集 $\Gamma_{qz} \in \mathbf{R}^{C_z \times H_z \times W_z}$,其中 m 代表未标记样本的个数。Few-shot 分类旨在每个类只给出几个图像的情况下将未带标签的查询图像分类为目标类(支持集)。这种设置也就是 N -ways K -shot 小样本分类。情景训练机制^[4]已被证明能有效地使用辅助集 A 学习可转移知识,以更好地对查询集 Q 进行分类。因此采用 episodic training 的元学习框架来进行少样本分类。对于每个 episodic,在辅助集 A 中随机构建一个新任务,每个任务由两个子集组成:辅助支持集 A_s 和辅助查询集 A_Q 。

1.2 算法框架

本文在算法框架如图1所示,该算法主要步骤如下。

步骤1 采用 Conv-4-64 前两个卷积块作为骨干网络,并将类增强特征与原始图像特征进行融合之后进行特征提取,同时将第3个卷积块的输出通道改为128,以便提取更多的低层表示,第4个卷积块恢复64通道数减少参数量。具体结构图见图2。

步骤2 多尺度特征生成模块。生成查询图像和支持图像的多种尺度的特征,将不同尺度上的低层语义特征进行特征融合,并计算查询集图像和支持集的图像语义相关矩阵。

步骤3 自适应任务注意模块。在得到查询图像 Q 和支持集 S 之间的语义相关矩阵的基础上,与注意力权重矩阵点乘之后获取加权语义关系矩阵,通过设置不同的超参数值 k 分别获取5个维度上的语义相似性矩阵,对关键低层表示进行权重分配之后通过融合不同尺度上的类别相似性对图像进行分类。

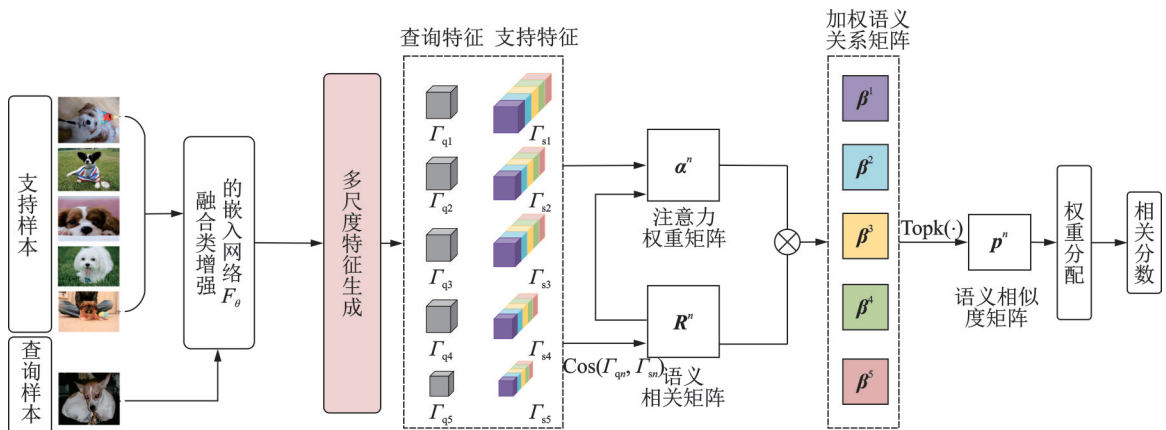


图1 融合类增强与多尺度自适应模型的结构图

Fig.1 Structure of the model based on class enhancement and multi-scale adaptation

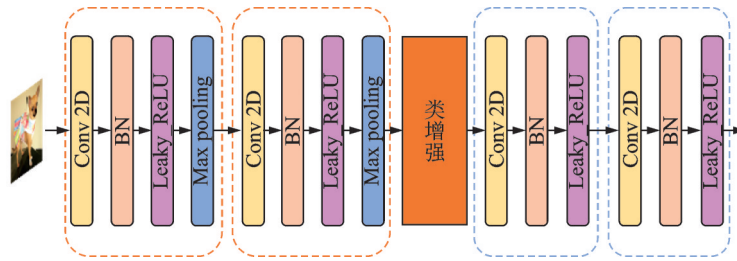


图2 特征提取器 F_θ 结构图

Fig.2 Structure of representations extractor F_θ

1.3 特征提取模块

特征提取模型的主干网络为 Conv-4-64。本文选择前两个卷积块作为骨干网络,并将类增强特征与原始图像特征进行融合之后进行特征提取,同时将第3个卷积块的输出通道改为128,以便提取更多的低层表示。传统的自注意方法^[13]通过挤压和激励得到经过加权处理的特征图,并将其输入到注意力区域网络以获得目标区域。但其缺少对空间位置信息的考虑。SCR(Self-correlational representation)^[14]基于局部自相关来提取关系特征。它的提出表明局部类特征对小样本学习的有效性。之所以可以取得这样的成果,是因为它可以学习与邻域的自相关性来抑制不相关的特征。

受此启发,本文在特征提取之后通过计算点与邻近区域的 Hadamard 积并将通道数进行暂时性压缩来达到类特征增强的效果。在计算完自相关值作为聚合的特征权重后,与输入图片进行 Concat 可以更好地保留图像自身丰富的语义信息。类特征增强原理如图3所示。为了得到类增强的特征图,对经过卷积处理得到的特征图做 ReLU-Hadamard product- (conv 3D-BN-ReLu) $\times 2$ 一系列操作,从而得到特征图类间特征表示,将类间特征与特征图相加得到最终类增强的特征图。图3给出了每个像素与其周围像素的 Hadamard 积过程,每个像素点的位移量为 $[-du, du] \times [-dv, dv]$,其中 du, dv 分别代表中心像素点长度和宽度的偏移量。当取到边缘点时,利用零填充来确保每个点都能与其邻域的 $U \times V$ 个矩阵进行 Hadamard 积操作, $U=2du+1, V=2dv+1$ 分别代表中心像素点的长度和宽度的变化范围。经过处理后,整个张量尺寸变为 $H \times W \times U \times V \times C$,将得到的类表示经过两层 3D 卷积操作后进行维度压缩至 $H \times W \times C$ 与原特征图进行残差连接得到类增强的特征图,其中 H 表示矩阵长度, W 代表矩阵宽度, C 代表矩阵的通道数。

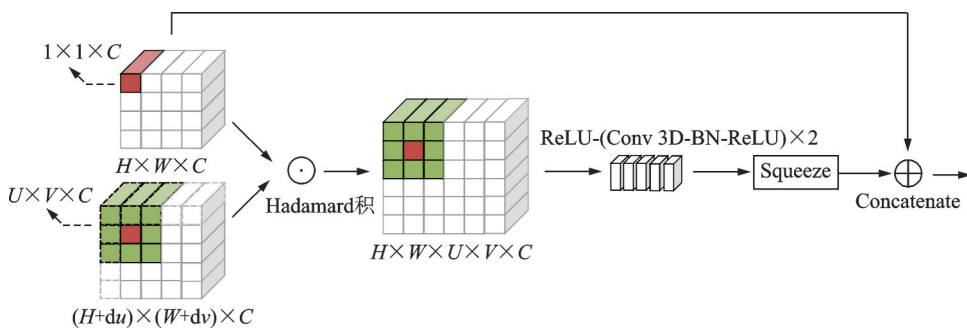


图3 类增强原理图

Fig.3 Schematic picture of class enhancement

1.4 多尺度特征生成模块

文献[15-18]证明了在比较不同尺度的对象时,集成多尺度特征对精度的有效性。集成多尺度特征提供了更多的多尺度信息,不但解决了对象的尺度失配问题,还解决了因比较对象的尺度变化而导致较差的泛化能力问题。为此,本文设计了多尺度特征生成器 $M(\cdot)$,每个尺度的池化操作如图4所示。

查询图像 A_Q^q 通过特征提取器 F_θ 得到三维向量 $F_\theta(A_Q^q) \in \mathbf{R}^{C \times H \times W}$ 。同理,支持集 S 通过特征提取器得到四维向量 $F_\theta(S) \in \mathbf{R}^{NK \times C \times H \times W}$ 。其中 N 表示一个任务中的类别数, K 表示每个类中的图像张数。

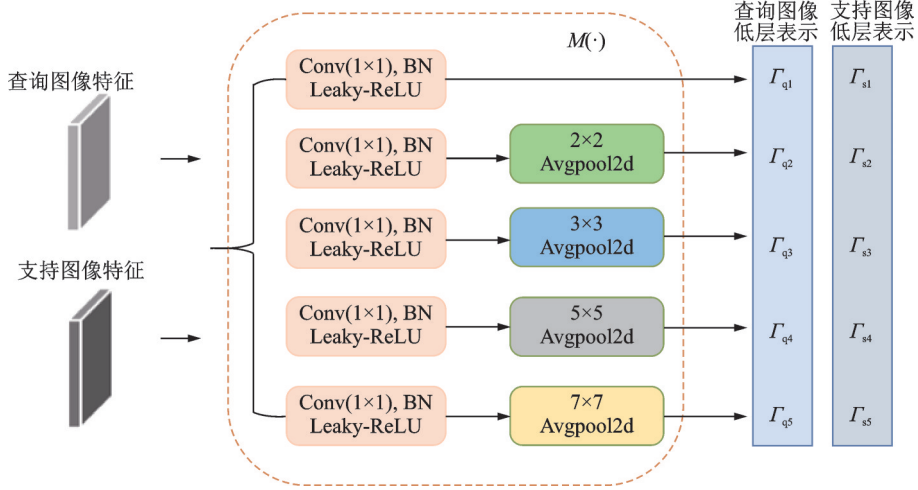


图4 多尺度特征生成器 $M(\cdot)$ 结构图

Fig.4 Architecture of multi-scale feature generator $M(\cdot)$

通过多尺度特征生成器,特征输出表示为5个三维特征 $\Gamma_{qn} \in \mathbf{R}^{C_n \times H_n \times W_n}$,其中 $n \in \{1, 2, 3, 4, 5\}$ 。将其变形为深度为 C_n ,行列数为 $[1, H_n \times W_n]$ 的低层特征信息,表示为

$$\Gamma_{qn} = [x_1, x_2, \dots, x_{H_n W_n}] \in \mathbf{R}^{C_n \times H_n W_n} \quad (1)$$

式中 x_i 表示第 i 个低层特征信息。同理支持集 S 的低层特征信息低层表示为

$$\Gamma_{sn} = [x_1, x_2, \dots, x_{NKH_n W_n}] \in \mathbf{R}^{C_n \times NKH_n W_n} \quad (2)$$

对于每个图像,低层特征信息的数量个数为 $\sum_{n=1}^5 H_n \times W_n$ 。将不同尺度上的低层语义特征进行特征融合,通过计算查询集图像和支持集的图像语义特征矩阵上每个像素点之间的余弦相关性,形成语义相关矩阵 R_i^n ,其表示查询图像与每个支持集中类别间的语义关系。

1.5 自适应任务注意模块

在单个尺度下的任务注意力权重可以表示为式(3),任务注意力权重 a^n 由所有任务注意力分数 a_i^n 组成,其中 $i \in \{1, 2, \dots, H_n W_n\}$ 。利用点积通过 a_i^n 对语义相关矩阵 R_i^n 进行加权处理。

$$a_i^n = \frac{\sum_{j=1}^{NKH_n W_n} R_{i,j}^n}{\sqrt{\sum_{i=1}^{H_n W_n} \sum_{j=1}^{NKH_n W_n} R_{i,j}^n}} \quad (3)$$

$$\beta_i^n = a_i^n \cdot R_i^n \quad (4)$$

$$\mathbf{p}^n = \sum_{i=1}^{kHW} \text{Topk}(\beta_i^n) \quad (5)$$

式中: $\text{Topk}(\cdot)$ 表示加权语义关系矩阵 β^n 的每行中的 k 个最大元素; \mathbf{p}^n 为经过加权处理后的查询图像和支持集中类别的语义相似度。

以上步骤为单个尺度上查询图像与支持集图像中各类别的语义相似度计算过程。对于最后的分类结果,需要将5个关系分数进行融合,为了解决这个问题,本文采用了一种可学习的五维 $\mathbf{w}=[w_1, w_2, w_3, w_4, w_5]$ 向量对5个尺度计算结果进行自适应整合。

$$\mathbf{p}^q = w_1 \cdot \mathbf{p}^1 + w_2 \cdot \mathbf{p}^2 + w_3 \cdot \mathbf{p}^3 + w_4 \cdot \mathbf{p}^4 + w_5 \cdot \mathbf{p}^5 \quad (6)$$

使用规范化先对5个向量大小进行平衡之后通过 1×1 卷积、膨胀值为5的一维卷积层操作,最终可以得到一个加权的五维相似性向量 \mathbf{p}^q ,并将其用于最终分类。

2 实验结果与分析

2.1 实验说明

本文实验平台是基于深度学习框架 Pytorch实现的,使用 Pycharm 编译器, GPU为双卡 8 GB显存的 NVIDIA GeForce GTX 1080Ti, CUDA采用 11.4版本,编译语言是 Python3.9。为了解决小样本学习中常见的过拟合现象,需要对训练图像进行数据增强以提高模型的泛化能力。对于数据增强,在训练期间对查询样本使用随机大小裁剪、随机颜色抖动、随机水平翻转和随机擦除等操作。

本文所用到的 benchmarks 均在 N -way K -shot的设置下进行。图片输入大小均为 84×84 。在训练阶段,本文每次随机构建 10 000 episode来训练本文提出的网络。在每个 episode中,随机从每个类选择 15个查询图像。损失函数使用交叉熵(Cross entropy, CE)损失函数,优化器为 Adam来训练网路。此外,初始学习率设置为 0.001,每 10个 epoch衰减 0.05,共设置 30个 epoch。在测试阶段,每次随机从测试集构建 600 episode,取其平均准确度和 95%置信区间。对于其他参数,本文对初始 k 的取值为 7。

2.2 实验数据集

本文首先在几个流行的数据集上进行实验,以验证所提出的网络的有效性。4个数据集的子类数量以及训练集与测试集的划分如表 1所示。miniImageNet^[4]类间差异较大,数据集更加复杂,更适合进行原型设计和实验研究。因此,本文进行的消融和对比实验均在 miniImageNet数据集上进行。CUB-200-2011是一个用于鸟类物种细粒度分类的数据集,它与 Stanford Dogs和 Stanford Cars为常用的细粒度图像数据集。为了体现实验结果的客观性,在进行训练集、验证集和测试集的划分时,采用与前人一致的类别数量划分,类别的划分具有随机性。

2.3 检测结果及分析

本文方法是基于度量学习,因此主要与该方向分支的其他方法进行比较,部分基于元学习的框架结构的实验结果作为参考使用。表 2显示了本文方法在 miniImageNet数据集上的评估结果。从表 2可以观察到,在 5-way 1-shot的设置下,本文方法相比于同样使用自适应特征的 DSN^[19]方法提升了 5.05%,准确率为 56.83%。同样地,在 5-way 5-shot的设置下,本文提出的网络比 RENet^[20]取得了明显

表 1 miniImageNet和细粒度数据集详细信息
Table 1 Details of miniImageNet and fine-grained image datasets

数据集	类别	训练集	验证集	测试集
miniImageNet	100	64	16	20
CUB-200	200	100	50	50
Stanford Dogs	120	70	20	30
Stanford Cars	196	130	17	49

表2 miniImageNet上95%置信区间的5-way 1-shot和5-way 5-shot任务的平均精度

Table 2 Average classification accuracy of 5-way 1-shot and 5-way 5-shot tasks with 95% confidence intervals on miniImageNet

方法	主干网络	度量选择	平均精度/%	
			5-way 1-shot	5-way 5-shot
MAML ^[12]	Conv-4-64	Meta	48.70±1.84	63.11±0.92
MAML+L2F ^[21]	Conv-4-32	Meta	52.10±0.50	69.38±0.46
WarpGrad ^[22]	Conv-4-32	Meta	52.30±0.80	68.40±0.66
Matching Nets ^[4]	Conv-4-64	Metric	43.56±0.84	55.31±0.73
Prototypical Nets ^[8]	Conv-4-64	Metric	49.42±0.78	68.20±0.66
Relation Nets ^[9]	Conv-4-64	Metric	50.44±0.82	65.32±0.70
DN4 ^[3]	Conv-4-64	Metric	51.24±0.74	71.02±0.64
DSN ^[19]	Conv-4-64	Metric	51.78±0.96	68.99±0.69
RENets* ^[20]	Conv-4-64	Metric	56.78±0.43	70.84±0.36
MsKPRN ^[23]	Conv-4-64	Metric	56.02 ± 0.88	72.06 ± 0.68
本文方法($k=7$)	Conv-4-64	Metric	56.83±0.39	75.76±0.42

的提高,RENet主张对类间特征进行渲染后通过交叉相关注意模块对小样本图像进行分类。为了方便对比,将RENet的backbone替换为与本文实验一致的Conv-4-64,在5-way 5-shot的设置下精度相比本文方法降低了4.92%,这说明以多尺度方式将相关图与特征图结合的框架结构丰富了查询图像和支持图像之间的深层关系,自适应权重矩阵的度量方法更利于模型分类,并获得了75.76%的准确率。相比同为多尺度特征MsKPRN的方法,准确率提升了3.70%。这都得益于本文所提出的类增强模块,可以有效地在生成多尺度特征之前先提高特定任务的类内特征表达。

另一方面,本文提出的网路在5-way 1-shot和5-way 5-shot设置下优于几种基于元学习的方法,如WarpGrad、MAML和MAML+L2F。Matching Nets、Prototypical Nets和Ration Nets作为小样本学习发展阶段的代表性度量方法,本文方法与其结果进行对比实验。

本文方法在细粒度数据集Stanford Dogs、Stanford Cars和CUB-200-2011上的平均精度见表3。这些数据被认为比标准数据集更具有挑战性,因为其类间差异相比较小。正如表2所示,本文方法在

表3 Standford Dogs, Standford Cars 和 CUB Birds数据集上的实验平均精度

Table 3 Average classification accuracy on Standford Dogs, Standford Cars and CUB Birds datasets %

方法	Stanford Dogs		Stanford Cars		CUB Birds	
	1-shot	5-shot	1-shot	5-shot	1-shot	5-shot
Matching Nets ^[4]	35.80±0.99	47.50±1.03	34.80±0.98	44.70±1.03	61.16±0.89	72.86±0.70
MAML ^[12]	44.81±0.34	58.68±0.31	47.22±0.39	61.21±0.28	55.92±0.95	72.09±0.76
Relation Nets ^[9]	43.33±0.42	55.23±0.41	47.67±0.47	60.59±0.40	62.45±0.98	76.11±0.69
DN4 ^[3]	45.41±0.76	63.51±0.62	59.84±0.80	88.65±0.44	52.79±0.86	81.45±0.70
LRPABN+cpt ^[5]	45.72±0.75	60.94±0.66	60.28±0.76	73.29±0.58	63.63±0.77	76.06±0.58
Temperature Net ^[6]	49.53±0.00	63.37±0.00	57.87±0.00	73.84±0.00	—	—
MsKPRN* ^[23]	60.12±0.89	78.12±0.54	75.64±0.84	89.88±0.46	66.29±0.95	81.94±0.65
RENets* ^[20]	59.04±0.49	72.36±0.38	66.94±0.48	75.45±0.37	69.12±0.49	85.45±0.31
本文方法($k=5$)	60.32±0.34	81.22±0.25	79.33±0.63	93.92±0.25	66.33±0.69	85.78±0.45

5-way 5-shot 的设置下实现了比 ResNet、MsKPRN 和 MATANets 更高的性能。相较于 ResNet-12, conv-4-64 的计算参数会小很多,训练结果不会过于依赖设备算力,相应地运算速度也会相对较快,但其特征提取信息不及 ResNet-12。

在 5-way 1-shot 的设置下,本文方法在 Stanford Dogs, Stanford Cars 和 CUB Birds 数据集上对比最相关的工作^[3]分别提升了 14.91%, 19.49% 和 13.54%。尤其是在 Stanford Cars 数据集上,本文方法在 5-way 5-shot, $k=5$ 的情况下准确度达到了 93.92%, 这与数据增强的相关操作密不可分。与细粒度特定方法 LRPABN 相比,本文方法在 5-way 1-shot 和 5-way 5-shot 的场景设置下在 Stanford Dogs 数据集上获得了约 15% 的改进。这些改进反应了 k 值选取的关键,合理的 k 值可以在捕获最具辨别性特征的同时避免噪声的引入。

2.4 消融研究与讨论

本节主要研究了所改进的类增强模块的有效性并对多尺度特征信息进行了分析。为了进一步研究超参数 k 对实验性能的影响,在自适应任务注意模块中,对于查询图像的每个低层表示,都需要搜索它与支持集中每个类别的低层表示之间 k 个最相似性特征,最终整合在不同尺度上计算这 5 个关键分数来进行最终的预测。因此,如何选择合适的 k 值尤为重要。为此,本文通过改变 k 值,来对比不同的 k 取值分别在 5-way 1-shot 和 5-way 5-shot 设置下对实验结果的影响。表 4 给出了在 miniImageNet 数据集上不同 k 值对实验结果的影响。从表 4 数据可以发现,在考虑参数数量和训练时间的条件下,当 k 取值为 7 时分类效果相比于其他 k 值的情况下,分类效果有小幅度的提升。需要注意的是,这并不意味着 k 的取值越大越好。通过实验发现当 $k=7$ 时,实验效果最为明显。

其次,为了进一步验证多尺度特征生成器、自相关注意模块的有效性,本文在 miniImageNet 上进行了消融实验。结果如表 5 所示。当剔除类增强模块后,在 5-way 1-shot 和 5-way 5-shot 的设置下性能分别降低了 1.72% 和 2.34%。因此类增强模块在特征提取部分发挥着重要作用,它可以帮助模型生成更具区分性的语义表征,从而提高了模型的分类性能。多尺度生成模块通过提供更多的尺度信息,不但解决了对对象的尺度失配问题,还解决了因比较对象的尺度变化而导致较差的泛化能力的问题。在其作用下,模型的精度在 5-way 1-shot 和 5-way 5-shot 的设置下分别提升了 2.51% 和 2.57%。

2.5 复杂度分析

为了帮助读者了解本文提出方法的实际运行效率和使用场景,进一步研究模型训练的代价与本文网络在参数上的优势,针对不同的网络,统一在 mini-ImageNet 数据集上进行 5-way 5-shot 实验,模型输入单个图像大小均为 (3, 84, 84), 选取参数量、准确率和单个任务图像处理时间作为指标进行分析,结果如表 6 所示。本文的骨干网络改进于 Conv-4-64。由表 6 可见,相较于经典网络 Prototypical Nets 和

表 4 miniImageNet 数据集上不同 k 值对实验平均精度的影响

k	5-way 1-shot	5-way 5-shot	%
3	55.85±0.37	74.73±0.41	
5	56.34±0.36	75.24±0.39	
7	56.83±0.39	75.76±0.42	
9	56.71±0.42	75.55±0.49	
11	56.35±0.48	75.14±0.41	

表 5 miniImageNet 数据集上对本文方法消融实验的平均精度

方法	5-way 1-shot	5-way 5-shot	%
Baseline	50.85±0.63	68.67±0.53	
去除类增强	55.11±0.61	73.42±0.62	
去除多尺度生成	54.32±0.91	73.19±0.49	
本文方法 ($k=7$)	56.83±0.39	75.76±0.42	

Relation Net,本文方法的参数量有所增加,但相对于精度方面的提升,参数数量的增加可以简单忽略。与文献[20,23]相比较,本文在性能和参数方面均具有优势。

表6 基于度量学习网络、小样本学习常用骨干网络与本文方法的参数对比

Table 6 Comparison of parameters between the metric based learning networks , the commonly used backbone networks in few-shot learning and the proposed method

方法	参数量/ 10^6	准确率/%	时间/s
Prototypical Nets ^[8]	0.113	68.20±0.66	—
Relation Net ^[9]	0.229	65.32±0.70	—
RENets* ^[20]	0.298	70.84±0.36	—
MsKPRN ^[23]	0.352	72.06 ± 0.68	0.65
Resnet12(backbone)	11.848	—	0.36
Conv-4-64(backbone)	0.178	—	0.34
本文方法	0.253	75.76±0.42	0.57

3 结束语

本文提出了一种新的小样本学习方法,用于更有效的相似性学习和更好的特征比较。本文所提出的方法以端到端的方式学习,将嵌入模型与类增强进行融合,可以更好地对类间特征进行提取。为了进一步提高分类器的可靠性,采用多尺度特征之间的深层关系来生成多尺度信息。最后的自适应任务模块将每一个训练任务中的查询图像与多个支持集图像之间的语义相关矩阵与注意力权重矩阵进行加权处理,同时将不同尺度上加权语义关系矩阵进行相关元素最大值选取来对不同任务的不同显著特征进行区分,最终将计算所得的相似性分数用于最后分类。通过广泛的分析和实验表明,本文所提出的方法可以有效改进现有小样本学习方法的性能,并且在5-way 5-shot设置下在3个常用的数据集Stanford Dogs, Stanford Cars和CUB Birds上达到了81.22%,93.92%和85.78%的平均精度,均优于一些最新方法。下一步工作可以进一步地探索小样本学习的度量学习与数据增强之间的相互作用,也可以进一步地推广到无监督任务分类中。

参考文献:

- [1] 彭云聪,秦小林,张力戈,等.面向图像分类的小样本学习算法综述[J].计算机科学,2022,49(5):1-9.
PENG Yuncong, QIN Xiaolin, ZHANG Lige, et al. A review of small sample learning algorithms for image classification[J]. Computer Science, 2022, 49(5): 1-9.
- [2] 黄友文,胡燕芳,魏国庆.结合原型校准分布的小样本学习方法[J].电子测量技术,2022,45(5):132-139.
HUANG Youwen, HU Yanfang, WEI Guoqing. A small sample learning method combining prototype calibration distribution [J]. Electronic Measurement Technology, 2022, 45(5): 132-139.
- [3] 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. [S.l.]: IEEE, 2019: 7260-7268.
- [4] VINYALS O, BLUNDELL C, LILLCRAP T, et al. Matching networks for one shot learning[J]. Advances in Neural Information Processing Systems, 2016, 29: 3630-3638.
- [5] HUANG H, ZHANG J, ZHANG J, et al. Low rank pairwise alignment bilinear network for few shot fine grained image classification[J]. IEEE Transactions on Multimedia, 2020, 23: 1666-1680.
- [6] ZHU W, LI W, LIAO H, et al. Temperature network for few shot learning with distribution aware large margin metric[J]. Pattern Recognition, 2021, 112: 107797.

- [7] KOCH G, ZEMEL R, SALAKHUTDINOV R. Siamese neural networks for one shot image recognition[C]//Proceedings of ICML Deep Learning Workshop. [S.l.]: ICML, 2015.
- [8] SNELL J, SWERSKY K, ZEMEL R. Prototypical networks for few shot learning[J]. *Advances in Neural Information Processing Systems*, 2017, 30: 4077-4087.
- [9] 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. [S.l.]: IEEE, 2018: 1199-1208.
- [10] LI W, XU J, HUO J, et al. Distribution consistency based covariance metric networks for few shot learning[C]//Proceedings of the AAAI Conference on Artificial Intelligence. [S.l.]: AAAI, 2019: 8642-8649.
- [11] SANTORO A, BARTUNOV S, BOTVINICK M, et al. Meta learning with memory augmented neural networks[C]//Proceedings of International Conference on Machine Learning. [S.l.]: PMLR, 2016: 1842-1850.
- [12] FINN C, ABBEEL P, LEVINE S. Model agnostic meta learning for fast adaptation of deep networks[C]//Proceedings of International Conference on Machine Learning. [S.l.]: PMLR, 2017: 1126-1135.
- [13] HU J, SHEN L, SUN G. Squeeze and excitation networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2018: 7132-7141.
- [14] ZHANG C, CHAM T J, CAI J. The spatially-correlative loss for various translation tasks[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2021: 16407-16417.
- [15] CHEN H, LI H, LI Y, et al. Multi scale adaptive task attention network for few shot learning[C]//Proceedings of 2022 26th International Conference on Pattern Recognition (ICPR). [S.l.]: IEEE, 2022: 4765-4771.
- [16] HAN M, WANG R, YANG J, et al. Multi scale feature network for few shot learning[J]. *Multimedia Tools and Applications*, 2020, 79: 11617-11637.
- [17] WANG X, MA B, YU Z, et al. Multi scale decision network with feature fusion and weighting for few shot learning[J]. *IEEE Access*, 2020, 8: 92172-92181.
- [18] ZHANG H, TORR P H S, KONIUSZ P. Few shot learning with multi scale self supervision[EB/OL]. (2020 01 06) [2023 03 10]. <https://arxiv.org/abs/2001.01600v1>.
- [19] 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. [S.l.]: IEEE, 2020: 4136-4145.
- [20] KANG D, KWON H, MIN J, et al. Relational embedding for few shot classification[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. [S.l.]: IEEE, 2021: 8822-8833.
- [21] BAIK S, HONG S, LEE K M. Learning to forget for meta learning[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2020: 2379-2387.
- [22] FLENNERHAG S, RUSU A, PASCANU R, et al. Meta learning with warped gradient descent[C]//Proceedings of International Conference on Learning Representations. [S.l.]: ICLR, 2020.
- [23] ABDELAZIZ M, ZHANG Z. Multi scale Kronecker product relation networks for few shot learning[J]. *Multimedia Tools and Applications*, 2022, 81(5): 6703-6722.

作者简介:



董驰静(1998-),女,硕士研究生,研究方向:小样本学习、元学习, E-mail: 2576919211@qq.com。



张孙杰(1988-),通信作者,男,博士,副教授,研究方向:智能图像处理、模糊系统控制与滤波, E-mail: zhang_sunjie@126.com。



任涵(2001-),男,本科,研究方向:传感器网络。

(编辑:张黄群)