文章编号：2095-2163(2022)12-0093-07 中图分类号：TP391.4 文献标志码：A

基于混合语义的图神经网络小样本图像分类方法

付炳光，杨 娟，汪荣贵，薛丽霞
（合肥工业大学 计算机与信息学院，合肥 230601）

摘 要：现有的基于图神经网络小样本分类方法，很少关注到与标签相关的语义信息。因此提出了基于混合语义的图神经网络小样本图像分类方法。使用补充词汇强化标签语义特征的表达能力，并将图像特征对齐到语义空间后，与标签语义特征混合生成实例级的混合语义特征。通过组合考虑任务上下文关系的图像特征和混合语义特征更好地描述样本，进而改善模型分类结果。在 Mini-ImageNet 和 Tiered-ImageNet 数据集上的实验结果表明，该方法对图像分类精度有明显的提高。

关键词：小样本学习；图神经网络；语义特征

Hybrid-semantic based Graph Neural Network for few-shot learning

FU Bingguang, YANG Juan, WANG Ronggui, XUE Lixia
(School of Computer Science and Information Engineering, Hefei University of Technology, Hefei 230601, China)

【Abstract】 Graph-based few-shot learning classification methods rarely pay attention to the semantic knowledge related to labels. This paper proposes hybrid-semantic based graph neural network for few-shot learning. Supplementary vocabulary is used to enhance the expression ability of label semantic features, and image features are aligned to semantic space and category semantic features are mixed to generate instance-level hybrid semantic features. By combining image features and mixed semantic features considering task context relations, the sample can be better described, and then the model classification results can be improved. The experimental results on Mini-ImageNet datasets and Tiered-ImageNet datasets show that the algorithm can significantly improve the image classification accuracy.

【Key words】 few-shot learning; graph neural network; semantic features

0 引 言

在计算机视觉的多个领域中，深度神经网络[1-2]均取得了优异的效果。深层的网络模型在训练时通常需要大量的标记数据，昂贵的数据标注成本使得模型训练成本大幅增加。同时在许多实际应用场景中，也不具备获得足够多标注样本的条件。在这种情况下，如何利用有限的标注样本获得性能较好的网络模型也随即成为一个亟待攻克的热点研究方向。基于此，小样本学习受到了广泛的关注。研究可知，元学习方法[3-4]在训练阶段和测试阶段构造相似的情节（episodes）任务，模拟人类总结任务经验的能力以使得机器从相似任务中获取通用知识并快速适应新任务，缓解了过拟合问题，成为了众多小样本学习方法的通用机制。

图神经网络[5]（graph neural networks，GNN）通过构建结构化信息的方式有效地提升深度学习模型的性能，许多研究[6-9]也开始尝试将图模型应用到小样本学习中。Garcia 等人[6]实现图模型预测值到标签值之间的后验推理，基于消息传递的想法，利用图推理将标签信息传递到没有标签的样本上，进而判别样本类型。Liu 等人[7]使用转导推理的方法，将所有无标注数据和有标注数据共同构建一个无向图，然后通过标签传播得到所有数据标签。与前面方法中图结构仅使用一组边特征表示类内相似、类间不同的节点关系不同，Kim 等人[8]构造了 2 组边特征，将节点间相似关系和不相似关系分开考虑。Ma 等人[9]使用支持样本和查询样本组合构成关系对并作为图节点，在传播和聚合节点信息过程中同时考虑节点间的相似性联系和节点内支持样本和查询样本关系。现有基于图神经网络的小样本学习方法通过构建出不同的图结构，虽然取得了优异的分类效果，但未考虑与图像相关的标签语义信息。与之不同，人们从少数样本中学习新概念时，不仅对比不同样本之间的差异，同时也考虑与之相关的文本知识。因此本文提出的方法尝试在使用图神经网络考虑图像特征间关系的同时，融入图像标签语义信息。

元学习方法的灵活性为学习新概念时引入其他

作者简介：付炳光（1994-），男，硕士研究生，主要研究方向：人工智能图像处理、小样本图像分类；杨 娟（1983-），女，博士，讲师，主要研究方向：智能视频图像处理与分析、神经网络与深度学习技术；汪荣贵（1966-），男，博士，教授，主要研究方向：深度学习、智能视频处理；薛丽霞（1976-），女，博士，副教授，主要研究方向：神经网络与深度学习技术、智能视频图像处理与分析、嵌入式多媒体技术。

收稿日期：2022-03-29

模态提供可能。不同模态蕴含的信息有互补性和一致性[10],不同模态间既含有类似的信息,同时也可能含有其他模态所欠缺的信息。在图像任务中,引入文本信息可以更全面地描述样本实例。Frederikd 等人[11]为获得更可靠的原型,通过生成对抗网络(generative adversarial networks , GAN)将语义特征对齐到图像特征空间,生成新特征改进图像类原型特征计算。Peng 等人[12]根据数据集标签在 WordNet 中的关系,由标签语义特征通过图卷积神经网络(graph convolution network, GCN)推理得到基于知识的分类权重,并与视觉分类权重融合得到新类的分类权重。Chen 等人[13]提出了 Dual TriNet,通过编码器将图像特征映射到语义空间,以随机添加高斯噪声等方式对该特征增广后,由解码器反解码形成各层特征图,由于可以无限增广,对此扩充训练特征。Li 等人[14]将标签语义特征经由多次 kNN 聚类得到多层超类语义特征,并构造底层为标签语义,上面多层为超类语义的树形结构的分层语义。如此一来,图像经由分级分类网络的同时在不同层会进行分类,训练得到良好的特征提取器。这些方法主要考虑类级别的语义信息,而忽略具体实例间的差异,一定程度上丧失了识别能力。为此,本文方法通过混合语义模块将实例级的图像特征对齐到语义空间并与其标签语义融合,为语义特征添加实例间的差异信息。此外,还通过补充语义信息增强标签语义的表达能力。

综上所述,本文提出了基于混合语义的图神经网络小样本分类方法。在常用小样本数据集上进行试验,并取得了良好的分类效果。

1 方法

1.1 问题定义

小样本图像分类目的是在仅有少量目标类标注样本的情况下,训练得到泛化性能良好的分类网络模型。通常将数据集划分为类别互不相交的训练集、测试集和查询集。同时采用 episode 训练机制[15],分为训练阶段和测试阶段,每个阶段由许多相似的 n - way k - shot 分类任务组成(现常见 5-way 1-shot 和 5-way 5-shot 两种类型)。具体地,训练阶段每个分类任务从训练集中随机抽取 n 个类,每类随机抽取 $k+q$ 张图片,构成当前任务的支持集 $S=\{(x_i, y_i), i=1,2,\cdots,nk\}$ 和查询集 $Q=\{(x_j,y_j), j=1,2,\cdots,nq\}$,其中 x_i, x_j 表示图像,y_i,y_j 表示该图像对应的标签。模型利用支持集样本的图像和标签信息判断查询样本的标签信息,并通过最小化已设计好的损失函数,反向传播更新网络模型参数达到模型训练的效果。训练阶段包含模型验证,从验证集随机采样构造 n - way k - shot 任务,检测模型泛化能力,保存最优的模型参数。最终,测试阶段在测试集上验证泛化性能。由于训练阶段和测试阶段构造类似的分类任务,由训练得到的模型能很好地迁移到训练集任务上。

1.2 基于混合语义的图神经网络模型

本文提出的模型结构如图 1 所示,包含图像特征信息传播模块、混合语义模块和决策混合模块。在每个分类任务中,图像通过图像特征提取网络得到图像特征,标签由 GloVe[16](Global Vector)计算得到标签语义特征。随后,图像特征信息传播模块使用图神经网络考虑任务上下文关系,更新得到任务相关的图像特征表示。混合语义模块利用补充语义信息和特征提取网络得到的视觉信息,增强标签特征的表达能力,得到混合语义特征。最后由决策混合模块组合图像特征和混合语义特征进行分类。

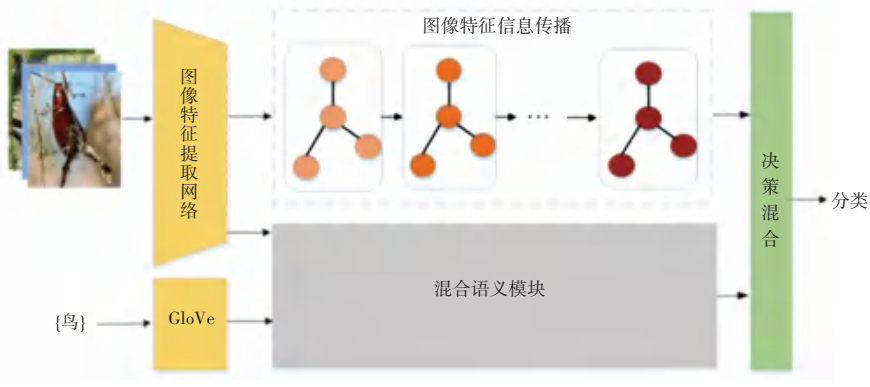

图 1 基于混合语义的图神经网络模型

Fig. 1 The model of hybrid semantic-based graph neural network

1.2.1　图像特征信息传播模块

图像特征信息传播模块使用图神经网络考虑图像特征上下文关系进而传播聚合特征信息，并更新特征表示。模块包含 L 层图像关系图 $G^l=(H^l,E^l)$，$l=1,2,\cdots,L$。第 l 层的图节点 H^l 由特征节点 h_i^l 组成，特征节点 h_i^l 与特征节点 h_j^l 间的相似度可表示为 e_{ij}^l，两节点之间的相似度越高，e_{ij}^l 越接近 1，反之接近 0。该层所有的节点相似关系构成了邻接矩阵 $\boldsymbol{E}^l$。

对于图像 x_i，输入到特征提取网络 F_{ex} 后，经过池化层和拉平层得到独热编码（one-hot coding）形式的图像特征向量 $F_{ex}(x_i)$，将其作为图中初始节点特征 $h_i^0=F_{ex}(x_i)$。初始邻接矩阵 $\boldsymbol{E}^0$ 采用公式（1）进行初始化：

$$\begin{cases} e_{ij}^0=1 & \text{实例 } i,j \text{ 标签相同} \\ e_{ij}^0=0 & \text{实例 } i,j \text{ 标签不同} \end{cases} \tag{1}$$

相同标签的支持集节点间的特征边设置为 1，而不同标签的支持集节点间设置为 0。此外，由于查询样本的标签未知，统一将支持集节点和查询节点的特征边设置为 $1/nk$。同时，根据描述节点间相似度关系的邻接矩阵，节点相互传播信息并聚集更新得到下一层节点。随后，更新节点间相似度得到下层的邻接矩阵。多层更新后得到每个样本的最终图像特征。具体地，对于第 k 层更新过程表示为：

$$h_i^k=f_h^k([h_i^{k-1},\sum_j e_{ij}^l * h_j^{k-1}]) \tag{2}$$

$$e_{ij}^k=f_e^k((h_i^k-h_j^k)^2) \tag{3}$$

式（2）中，“[，]”表示串接，对于第 $k-1$ 层的特征节点 h_i^{k-1}，将其它节点特征 h_j^{k-1} 与 h_i^{k-1} 的相似度 e_{ij}^l 作为权重累加，将累加值与 h_i^{k-1} 级联输入到神经网络 f_h^k 中，输出作为第 k 层节点特征 h_i^k。式（3）计算两节点的特征边，节点特征向量不同维度蕴含不同信息，在衡量特征相似度时有着不同的重要性。将 2 个特征向量做差取平方，由神经网络 f_e^k 判断不同维度重要性，进而计算得到边特征。式（2）和式（3）的 f_e^k，f_h^k 均由多层感知机（multilayer perceptron，MLP）和 *ReLu* 激活函数组成。

1.2.2　混合语义模块

GloVe[16] 和 Word2Vec[17] 等文本嵌入方法根据词语在语料库中的分布，将词语转换为独热编码（one-hot）表示的语义特征。语义特征不仅含有词语信息，还蕴含语料库不同词语间的联系。“卡车”与“汽车”间相对于“狗”与“汽车”间具有更强的关联性。换而言之，对于一个与“汽车”关联性很强的未知词语，该词语为“卡车”的概率比为“狗”的概率更大。据此，本节提出混合语义模块，通过文本嵌入的方法计算得到类别标签的语义特征，并引入其他词语作为补充描述，以增强其表达能力。

相比于因有限窗口大小而仅可捕捉局部信息的 Word2Vec，本文使用可以捕获全局信息的 GloVe 方法计算标签的语义特征，并将所有支持集类别作为补充词语。混合语义模块如图 2 所示。由图 2 可看到，当前任务类别标签和补充词语由 GloVe 方法计算得到标签语义特征集 $\boldsymbol{S}_t\in\mathbb{R}^{n\times d_w}$ 和补充语义特征集 $\boldsymbol{S}_a\in\mathbb{R}^{n_a\times d_w}$，这里的 d_w 表示语义特征的维度。运行时，先将 S_t 和 S_a 分别进行线性变换后输出 $\boldsymbol{Q}$ 和 $\boldsymbol{K}$，接着将 $\boldsymbol{Q}$ 和 $\boldsymbol{K}$ 做矩阵乘法之后再乘以缩放系数 $1/\sqrt{d_w}$，经过 *softmax* 函数输出注意力分数矩阵 $\boldsymbol{A}$，由此推得的各数学公式分别如下：

$$\boldsymbol{Q}=\boldsymbol{W}_Q\boldsymbol{S}_t \tag{4}$$

$$\boldsymbol{K}=\boldsymbol{W}_K\boldsymbol{S}_a \tag{5}$$

$$\boldsymbol{A}=softmax(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d_w}}) \tag{6}$$

$$\boldsymbol{S}_h=\boldsymbol{S}_t+\boldsymbol{A}\boldsymbol{S}_a \tag{7}$$

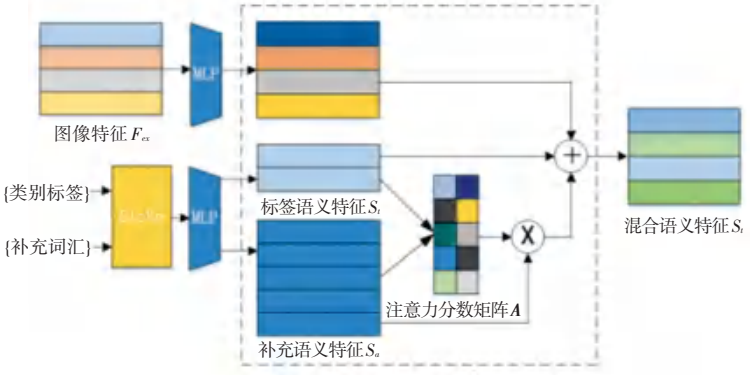

图 2　混合语义模块

Fig. 2　The hybrid semantic module

式(4)和式(5)中，$\boldsymbol{W}_Q, \boldsymbol{W}_K \in \mathbb{R}^{n\times d_w}$ 为权重矩阵，引入缩放系数 $1/\sqrt{d_w}$ 得到更为平滑的输出。式(7)将注意力分数矩阵 $\boldsymbol{A}$ 与 $\boldsymbol{S}_a$ 做矩阵乘法后，再以相加的方式与标签语义特征融合，得到增强语义特征集 $\boldsymbol{S}_h = \{s_{h1}, s_{h2}, \cdots, s_{hN}\}$。在理解特定对象时，不同词语的重要性不同。正如“卡车”与“汽车”之间的联系比“狗”与“汽车”之间的联系更强，“卡车”的语义对理解“汽车”更为重要，注意力分数矩阵 $\boldsymbol{A}$ 起到了相似的作用，反映了 $\boldsymbol{S}_a$ 中的每个语义特征补充描述 $\boldsymbol{S}_t$ 中语义特征时的重要程度。

此外，在 n - way k - shot 任务下可能有多张图片属于相同类别，不同图像特征对应同一语义特征忽略了同类图像特征之间的差异性。不同模态间往往具有相关的信息[10]，本文模型尝试将图像特征对齐映射到语义特征空间中，与混合语义特征融合构建实例级的语义特征表示。对于支持集图像特征 $F_{ex}(x_i)$，其映射得到的特征为：

$$s_{vi} = f_{mlp}(F_{ex}(x_i)) \tag{8}$$

将 s_{vi} 与实例 x_i 标签的增强语义特征 s_{hy} 融合得到该实例的混合语义特征 s_i：

$$s_i = s_{vi} + s_{hy} \tag{9}$$

1.2.3 决策融合

由图像特征信息传播模块可以得到考虑图像上下文关系的图像特征，而混合语义模块得到支持集实例的混合语义特征。不同模态间存在互补性[10]，可能包含其他模态所欠缺的信息，利用多个模态的信息有助于更好地描述实例。本方法通过式(10)组合支持集图像特征和混合语义特征：

$$h_i = f_{fusion}([h_i^L, g(s_i)]) \tag{10}$$

其中，“[,]”表示级联操作。式(10)中将实例 x_i 的混合语义特征 s_i 输入网络 g 后，与其图像特征 h_i^L，再经由网络 f_{fusion} 得到混合模态特征 h_i，g 和 f_{fusion} 均由多层感知机和 *ReLu* 激活函数构成。

1.3 损失函数

为了推理查询样本的标签，使用支持集混合模态特征和查询集图像特征计算得相似度矩阵 $\boldsymbol{E} = \{e_{ij}\}_{i,j=0}^{nk+nq}$，$e_{ij}$ 表示样本 i 和样本 j 间的相似度分数。将 $\boldsymbol{E}$ 输入到 *softmax* 函数中预测每个节点的分类 $P(\widetilde{y_i} \mid x_i)$，可由如下公式进行计算：

$$e_{ij} = f_e((h_i - h_j)^2) \tag{11}$$

$$P(\widetilde{y_j} \mid x_i) = softmax(\sum_{j=1}^{nk} onehot(y_j) \cdot e_{ij}) \tag{12}$$

其中，i，j 分别为支持样本和查询样本下标，$onehot(y_j)$ 是支持集样本 j 标签的独热编码。在给定 episode 任务下，利用最小化分类损失函数来训练模型：

$$L_{cls} = L_e(\boldsymbol{Y}_e, \boldsymbol{E}) \tag{13}$$

其中，$\boldsymbol{Y}_e$ 是真实的相似度邻接矩阵，计算 $\boldsymbol{Y}_e$ 与预测矩阵 $\boldsymbol{E}$ 之间的二值交叉熵作为分类损失函数。

图像特征信息传播模块的邻接矩阵同样也可以预测节点分类，增加式(14)的损失函数用来改善训练过程中的梯度更新，但仅用 $\boldsymbol{E}$ 作为查询样本的标签判断。式(14)的数学表述具体如下：

$$L_e = \sum_{l=1}^{L} L_e(\boldsymbol{Y}_e, \boldsymbol{E}^l) \tag{14}$$

GloVe[16]将语料库中词汇 X 在词汇 Z 出现的情况下出现的概率 $P_{X|Z}$ 与词汇 Y 在词汇 Z 出现的情况下出现的概率 $P_{Y|Z}$ 的比值 $\frac{P_{X|Z}}{P_{Y|Z}}$，称为共现概率比。当 X 与 Z 的关联性和 Y 与 Z 的关联性都很强或者都很弱时，共现概率比趋于 1，否则共现概率比趋于很大或者趋于零。通过引入第三个词汇 Z，共现概率比很好地描述了词汇 X 和词汇 Y 间的相似性。受此启发，为使公式(8) 中图像特征更好地映射到语义空间，通过计算映射后的特征与整体补充语义特征的相似度矩阵 $\boldsymbol{A}_{v,a}$，实例标签语义特征与整体补充语义特征的相似度矩阵 $\boldsymbol{A}_{t,a}$，并计算 2 个相似度矩阵之间的相似熵损失：

$$L_{KL} = \sum [\boldsymbol{A}_{t,a}\log(\boldsymbol{A}_{v,a}) - \boldsymbol{A}_{v,a}\log(\boldsymbol{A}_{t,a})] \tag{15}$$

模型的总损失如式(16)所示：

$$L = L_{cls} + \lambda_1 L_e + \lambda_2 L_{KL} \tag{16}$$

其中，λ_1，λ_2 为超参数，用于调整损失 L_e 和 L_{KL} 对网络模型训练的影响。

2 实验

2.1 数据集

为了更好地对比分析模型性能，本文在小样本学习方法常用的 Mini-ImageNet 和 Tiered-ImageNet 数据集上进行了实验。本节中所有实验均在搭载 NVIDIA GeForce TiTan X 12 GB 显卡、Intel i7 - 9700KF 处理器并具有 16 G 运行内存的 PC 机上完成，采用 Linux 版本的 PyTorch 10.2 深度学习框架实现模型的搭建。

Mini-ImageNet 数据集是 ImageNet[18] 的子集，有 100 个类别，每类由 600 张图片组成。有 2 种常见的使用方法。一种方法将 80 个类别作为训练集，

剩余的 20 个类别作为验证集。另一种方法将数据集划分为包含 64 个类别的训练集、16 个类别的验证集和 20 个类别的查询集。本文使用后一种方法。

Tiered-ImageNet 数据集同样节选于 ImageNet 数据集。不同的是该数据集比 Mini-ImageNet 包含更多的类别，也包含更多的图片数量。在规模上，包含了 608 个小类别，平均每个类别有 1 281 个样本；在语义结构上，是将数据集划分成 34 个父类别来确保类别之间的语义差距。在以往的工作中，将 20 个父类别作为训练集、对应 351 个子类别，6 个父类别作为验证集、对应 97 个子类别以及 8 个父类别作为测试集、对应 160 个子类别。

2.2　实验配置

本文分别采用 2 种流行的网络 Conv4 和 ResNet-12[15] 作为图像特征提取网络，使用 GloVe 计算语义特征。Conv4 主要由 4 个 *Conv* - *BN* - *ReLU* 块组成，每个卷积块包含一个 64 维滤波器 3×3 卷积，卷积输出分别输入到后面的批量归一化和 *ReLU* 非线性激活函数。前 2 个卷积块还包含一个 2×2 最大池化层，而末端 2 个卷积块没有最大池化层。ResNet12 主要有 4 个残差块，每层残差块由 3 层卷积层接连组成，并在残差块后添加了 2 × 2 的最大池化操作。遵循大多数现有的小样本学习工作所用的标准设置，使用 5-way 1-shot 和 5-way 5-shot 两种实验设置和提前结束策略，并将 Adam 作为学习优化器。在 Mini-ImageNet 上训练时，使用随机采样并构建 300 000 个 episode，设置 Adam 初始学习率为 0.001，每 15 000 个 episode 将学习率衰减 0.1。对于 Tiered-ImageNet 数据集，使用随机采样并构建 500 000 个 episode，设置 Adam 初始学习率为 0.001，每 20 000 个 episode 将学习率衰减 0.1。

2.3　模型对比实验和分析

本文模型与其他使用图模型和使用语义模态的小样本学习方法在 Mini-ImageNet 和 Tiered-ImageNet 数据集上的实验结果见表 1、表 2。表中，标注 N/A 表示该实验结果在原文献中并未展示出来。

表 1 给出了在 Mini-ImageNet 数据集上，本文模型与其他小样本方法在 5-way 1-shot 和 5-way 5-shot 两种任务下的实验结果。从实验结果中可以看出，本文方法明显优于当前大多数小样本学习方法。本文方法与经典小样本学习方法 Matching Network[19]、MAML[3]、Prototypical Network[20]、Relation Networks[21] 相比，准确率有明显的提升。与基于图神经网络的小样本方法相比，在 1-shot 情况下本文方法比 GNN[6] 准确率高出 5.47%，在 5-shot 情况下高出 5.15%，而与 TPN[7] 相比，本文在 1-shot 情况下准确率高出了 2.05%，5-shot 情况下高出了 2.13%。此外，与同样使用语义信息的 TriNet[20] 相比，本文模型在 1-shot 情况下高出 0.82%，但是在 5-shot情况下，TriNet[12] 的准确率高于本文模型。同样使用 Conv4 特征提取网络，与近年来最新的 FEAT[24]、MELR[25] 模型对比，本文模型虽然在 5-shot的情况下准确率略低，但在 1-shot 情况下准确率仍然高过这些基准参照模型。

表 1　在 Mini-ImageNet 数据集上不同模型的准确率

Tab. 1　Accuracy of different models on the Mini-ImageNet dataset

Model	backbone	5-way 1-shot	5-way 5-shot
Matching Network[19]	Conv4	43.56(±0.84)	55.31(±0.73)
MAML[3]	Conv4	48.70(±1.84)	63.11(±0.92)
Prototypical Network[20]	Conv4	49.42(±0.78)	68.20(±0.66)
Relation Network[21]	Conv4	50.44(±0.82)	65.32(±0.70)
PN+IFSM[22]	Conv4	N/A	66.98(±0.68)
GNN[6]	Conv4	50.33(±0.36)	66.41(±0.63)
TPN [7]	Conv4	53.75(±0.86)	69.43(±0.67)
STANet-S [23]	Conv4	53.11(±0.60)	67.16 (± 0.66)
FEAT[24]	Conv4	55.15(±0.70)	71.61(±0.63)
MELR[25]	Conv4	55.35(±0.43)	72.27(±0.35)
TriNet[13]	ResNet18	58.12(±1.37)	76.92(±0.69)
STANet-S[23]	ResNet12	58.35(±0.57)	71.07(±0.39)
本文方法	Conv4	55.80(±0.81)	71.56(±0.77)
本文方法	ResNet12	58.94(±0.91)	73.62(±0.72)

表 2　在 Tiered-ImageNet 数据集上不同模型的准确率

Tab. 2　Accuracy of different models on the Tiered - ImageNet dataset

Model	backbone	5-way 1-shot	5-way 5-shot
Matching Network[19]	Conv4	54.02(±0.00)	70.11(±0.00)
Prototypical Network[20]	Conv4	53.31(±0.89)	72.69(±0.74)
MAML [3]	Conv4	51.67(±1.81)	70.30(±0.08)
Relation Network [21]	Conv4	54.48(±0.93)	71.32(±0.70)
Soft k-means[26]	Conv4	52.39(±0.44)	69.88(±0.20)
GNN[6]	Conv4	43.56(±0.84)	55.31(±0.73)
TPN[7]	Conv4	57.53(±0.96)	72.85(±0.74)
本文方法	Conv4	57.13(±0.96)	73.21(±1.22)
本文方法	ResNet12	59.78(±0.92)	74.91(±1.25)

表 2 给出了在 Tiered-ImageNet 数据集上，本文的模型与其他小样本方法在 5-way 1-shot 和 5-way

5-shot两种任务下的实验结果。从实验结果中可以看出,本文方法明显优于当前大多数小样本学习方法。本文方法与经典小样本学习方法 Matching Network[19]、MAML[3]、Prototypical Network[20]、Relation Networks[21]相比,准确率均有较大提升。与基于图神经网络的小样本方法对比,在 1-shot 情况下本文方法比 GNN[6]准确率高出 11.47%,在 5-shot 情况下高出 16.4%;与 TPN[7]相比,本文方法在 5-shot 情况下准确率高出了 0.34%,但在 1-shot 情况下 TPN[9]有着更高的分类准确率。

在 Mini-ImageNet 和 Tiered-ImageNet 数据集上,将 5-way 1-shot 和 5-way 5-shot 两种情况进行对比可以发现随着支持集的样本数量增加,分类的效果也更好。将 Conv4 和 ResNet12 两种骨干网络进行对比发现,采用更加深层的特征提取网络能得到更高的准确率。

2.4 消融实验

本节通过在 Mini-ImageNet 数据集上进行消融实验证明本文模型的有效性以及检验部分参数对模型训练的影响。

首先,本文探究图像特征关系传播模块迭代更新层数对模型准确率的影响。图像特征关系传播模块使用图神经网络充分挖掘图像特征之间的关联信息,由多层包含特征节点和相似度邻接矩阵的相同结构组成,网络层数影响着模块的参数,对整体性能起着非常重要的作用,所以有必要对层数进行消融实验分析。选择 5-way 1-shot 作为任务设定,层数分别选择 1、2、3、4、5,模型准确率如图 3 所示。从图 3 中可以看出,当层数由 1 到 3 时,模型分类准确率有着明显提升,说明在层数较少时,增加模型的层数可以提升整体模型的分类效果。当层数从 3 到 5 时,模型分类效果有些许波动,但整体而言准确率趋于稳定,不断增加模型层数不能持续提升模型的分类准确率。因此本文在其他所有实验中,模型层设定为 3,既能得到较高的模型分类准确率,同时也避免了过多耗时的计算量。

此外,为探究混合语义模块在模型训练中发挥的作用,对混合语义模块进行消融实验。Mini-ImageNet 数据集上,混合特征模块消融实验结果见表 3。表 3 中,“仅图像”表示仅使用本文中的图神经网络进行训练分类。“标签语义”表示混合语义模块直接使用标签语义而忽略其他语义信息。“标签语义+视觉对齐语义”虽然使用补充语义信息,但是补充语义仅使用在损失函数中改进模型训练。从实验结果可以看出,引入语义信息能提高小样本图像分类的表现效果。此外,使用混合语义模块在 5-way 5-shot 任务下准确率的提高要逊色于 5-way 1-shot 任务,主要原因是在 5-shot 情况下,图像信息将更加丰富,而语义信息模型效果的提升就很有限。

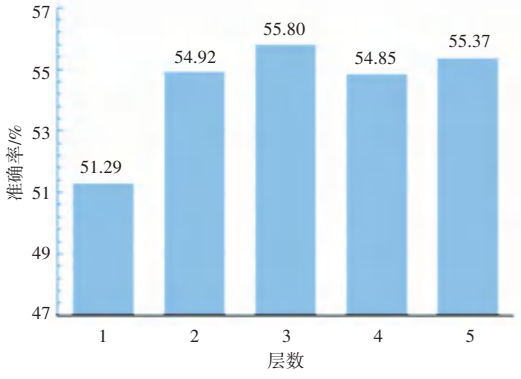

图 3 5-way 1-shot 任务下,图像信息传递模块层数对模型分类准确率的影响

Fig. 3 Influence of image information transfer module layers on model classification accuracy under 5-way 1-shot task

表 3 Mini-ImageNet 数据集上混合特征模块消融实验结果

Tab. 3 Ablation experimental results of hybrid feature module on Mini-ImageNet dataset

模型	5-way 1-shot	5-way 5-shot
仅图像	53.34(±0.45)	69.88(±0.58)
标签语义	54.42(±0.78)	70.73(±0.72)
标签语义+补充语义	55.46(±0.88)	71.21(±0.82)
标签语义+视觉对齐语义	55.14(±0.80)	70.97(±0.74)
标签语义+补充语义+视觉对齐语义	55.80(±0.81)	71.71(±0.77)

3 结束语

本文首先提出了基于混合语义的图神经网络小样本分类方法。该方法考虑实例图像特征和语义特征之间的互补性,由此得到的融合特征,能更全面描述实例信息。其中,使用图神经网络模型综合考虑支持集和查询集图像之间的关系,并使用补充语义来增强标签语义特征的表达能力,以及利用图像对齐语义特征构造了实例级语义特征。本文模型在 Mini-ImageNet 和 Tiered-ImageNet 数据集上取得了良好的分类效果。考虑到现有模型面对不同任务时,会遗忘已有的分类知识的灾难性遗忘问题,进一步扩展模型应对小样本增量学习则是未来研究工作的重点。

参考文献

[1] KIM I, BAEK W, KIM S. Spatially attentive output layer for image classification [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, 2020: 9533-9542.

[2] ZHANG Manli, ZHANG Jianhong, LU Zhiwu, et al. IEPT: Instance-level and episode-level pretext tasks for few-shot learning [C]//International Conference on Learning Representations. ICLR, 2020:1-16.

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

[4] ANTONIOU A, EDWARDS H, STORKEY A. How to train your MAML[J]. arXiv preprint arXiv:1810.09502, 2018.

[5] SCARSELLI F, GORI M, TSOI A C, et al. The graph neural network model [J]. IEEE Transactions on Neural Networks, 2008, 20(1): 61-80.

[6] GARCIA V, BRUNA J. Few-shot learning with graph neural networks[J]. arXiv preprint arXiv:1711.04043, 2017.

[7] LIU Y, LEE J, PARK M, et al. Transductive propagation network for few-shot learning [J]. arXiv preprint arXiv: 1805.10002, 2018.

[8] KIM J, KIM T, KIM S, et al. Edge-labeling graph neural network for few-shot learning [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, CA:IEEE,2019: 11-20.

[9] MA Yuqing, BAI Shihao, AN Shan, et al. Transductive relation-propagation network for few-shot learning [C]//IJCAI. Yokohama, Japan:dblp,2020, 20: 804-810.

[10] BALTRUSAITIS T, AHUJA C, MORENCY L P. Multimodal machine learning: A survey and taxonomy[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2019,41(2):423-443.

[11] FREDERIKD P, PUSCAS M, KLEIN T, et al. Multimodal prototypical networks for few-shot learning[C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. IEEE,2021: 2644-2653.

[12] PENG Zhimao, LI Zechao, ZHANG Junge, et al. Few-shot image recognition with knowledge transfer[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South):IEEE, 2019: 441-449.

[13] CHEN Zitian, FU Yanwu, ZHANG Yinda, et al. Multi-level semantic feature augmentation for one-shot learning [J]. IEEE Transactions on Image Processing,2019, 28(9):4594-4605.

[14] LI Aoxue, LUO Tiange, LU Zhiwu, et al. Large-scale few-shot learning: Knowledge transfer with class hierarchy [C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA, USA: IEEE, 2019: 7205-7213.

[15] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA:IEEE, 2016:770-778.

[16] PENNINGTON J, SOCHER R, MANNING C D.GloVe: Global vectors for word representation [C]//Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). Doha, Qatar:ACL,2014:1532-1543.

[17] GOLDBERG Y, LEVY O. word2vec Explained: deriving Mikolov et al.'s negative-sampling word-embedding method[J]. arXiv preprint arXiv:1402.3722, 2014.

[18] DENG Jia, DONG Wei, SOCHER R, et al. ImageNet: A large-scale hierarchical image database[C]//2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami Beach, Florida: IEEE, 2009: 248-255.

[19] 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.

[20] SNELL J, SWERSKY K, ZEMEL R S. Prototypical networks for few-shot learning[J]. arXiv preprint arXiv:1703.05175, 2017.

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

[22] CAI Chunhao, YUAN Minglei, LU Tong. IFSM: An iterative feature selection mechanism for few-shot image classification [C]//2020 25th International Conference on Pattern Recognition (ICPR). Milan, Italy:IEEE, 2020.

[23] YAN Shipeng, ZHANG Songyang, HE Xuming. A dual attention network with semantic embedding for few-shot learning [C]// Proceedings of the AAAI Conference on Artificial Intelligence. Honolulu,USA:AAAI,2019,33:9097-9086.

[24] YE Hanjia, HU Hexiang, ZHAN Dechuan, et al. Few-shot learning via embedding adaptation with set-to-set functions[C]// 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA: IEEE, 2020: 8805-8814.

[25] FEI Nanyi, LU Zhiwu, XIANG Tao, et al. MELR: Meta-learning via modeling episode-level relationships for few-shot learning [C]//International Conference on Learning Representations. ICLR,2020:1-20.

[26] REN Mengye, TRIANTAFILLOU E, RAVI S, et al. Meta-learning for semi-supervised few-shot classification [C]// International Conference on Learning Representations. Vancouver: ICLR, 2018:1-15.