



基于注意力关系网络的无线胶囊内镜图像分类方法

安晨, 汪成亮, 廖超, 肖诗童

(重庆大学 计算机学院, 重庆 400044)

摘要: 无线胶囊内镜(WCE)技术可以检测出肠胃道异常, 计算机辅助诊断WCE图像方法由于标注图像数据量少、图像类内变异度高和类间相似等原因导致效果不佳。为此, 提出一种基于注意力关系网络的WCE图像多分类方法。将关系网络、注意力机制以及元学习训练策略相结合, 构造基于注意力机制的嵌入模块以提取WCE图像特征, 将提取后的特征进行特征映射级联后输入到关系模块, 根据关系模块输出的相似性评分判断样本所属类别, 采用元学习训练策略训练网络。实验结果表明, 该方法的分类精度高于RelationNet、MAML等小样本分类方法, 且在WCE数据集上该方法的精度高达90.28%。

关键词: 无线胶囊内镜图像; 注意力机制; 元学习; 关系网络; 卷积神经网络

开放科学(资源服务)标志码(OSID):



中文引用格式: 安晨, 汪成亮, 廖超, 等. 基于注意力关系网络的无线胶囊内镜图像分类方法[J]. 计算机工程, 2021, 47(10): 252-259, 268.

英文引用格式: AN C, WANG C L, LIAO C, et al. Wireless capsule endoscopy image classification method based on attention relational network[J]. Computer Engineering, 2021, 47(10): 252-259, 268.

Wireless Capsule Endoscopy Image Classification Method Based on Attention Relational Network

AN Chen, WANG Chengliang, LIAO Chao, XIAO Shitong

(College of Computer Science, Chongqing University, Chongqing 400044, China)

[Abstract] Wireless Capsule Endoscopy (WCE) technology can detect gastrointestinal abnormalities. However, the performance of computer-aided diagnosis based on WCE images is reduced due to the small amount of labeled image data, intra-class variation and inter-class similarity. To address the problem, an attentional relational network-based WCE image classification method is proposed. The method combines the relational network, the attention mechanism and the meta-learning training strategy. On this basis, an embedded module based on the attention mechanism is built to extract features of WCE images, and then the extracted features are input into the relation module after feature mapping cascade. The category of the samples is judged according to the similarity score output by the relation module, and the network is trained by using the meta-learning training strategy. The experimental results show that the classification accuracy of the proposed method is higher than that of RelationNet, MAML and other small sample classification methods, reaching up to 90.28% on the WCE dataset.

[Key words] Wireless Capsule Endoscopy (WCE) image; attention mechanism; meta-learning; relational network; convolutional neural network

DOI: 10.19678/j.issn.1000-3428.0059122

0 概述

肠胃道疾病对公众健康造成巨大威胁, 及时发现肠胃道异常可降低患病几率, 无线胶囊内镜(Wireless Capsule Endoscopy, WCE)^[1]能够以无创伤、无痛的方式直接检查患者的肠胃部分。一个患者的

WCE检查过程将会持续8~11 h, 摄像头以2 frame/s的速度拍摄5~8万张图像, WCE图像多达上万张, 但是, 复杂的胃肠道环境因为碎片、散焦、气泡和光线等干扰因素产生了大量质量较低的WCE图像, 可用作诊断分析的WCE图像只占所收集图像的5%。目前, 许多方法被应用于WCE图像的辅助诊断, 且多

基金项目: 国家自然科学基金(61672115); 中央高校基本科研业务费专项资金(2020CDCGJSJ040)。

作者简介: 安晨(1995—), 女, 硕士研究生, 主研方向为医疗大数据; 汪成亮, 教授、博士、博士生导师; 廖超, 博士研究生; 肖诗童, 硕士研究生。

收稿日期: 2020-07-31 **修回日期:** 2020-10-09 **E-mail:** 1242100810@qq.com

数方法都是针对单一异常进行检测,包括出血^[2]、息肉^[3]、溃疡^[4]以及克罗恩病检测^[5]。但是,真实环境中一个患者的肠胃道常具有多种异常,WCE图像的多分类在临床实践中具有重要意义,只有很少的方法被应用于WCE图像的多分类任务^[6-8]。WCE图像因数据分散、数据采集困难、隐私保护等原因通常只有少量的标注数据,与小样本自然图像分类不同,WCE图像分类还面临类间相似、类内变异度高等问题。

为了解决小样本WCE图像的多分类问题,本文提出一种基于注意力关系网络(Attention Relational Networks, ARNs)的小样本WCE图像分类方法。该方法将关系网络^[9]、注意力机制以及元学习训练策略相结合,以在少量标记样本下对WCE图像进行有效分类。构造基于注意力机制的嵌入模块提取WCE图像的特征,将提取到的查询样本特征与支持样本特征进行级联并输入到关系模块中,根据关系模块输出的关系评分得到查询样本所属的类别,采用元学习的训练策略训练网络。在多个数据集上进行实验,采用准确率、特异性和敏感度评价指标对本文模型进行评估,并将其与其他小样本学习方法以及WCE图像多分类方法进行性能比较。

1 相关研究

1.1 小样本学习方法

小样本学习能够应对标记数据不足的问题,对于训练过程中只给出极少量的标注样本的类别,模型也能够正确识别。度量学习是一种主流的小样本学习解决方案,基于度量学习的小样本学习方法通过学习一个特征映射函数,将样本投射到一个特征空间使得同类样本聚集,异类样本分离,然后比较样本在特征空间的距离以判别样本所属的类别。在度量学习研究领域,具有代表性的网络包括孪生网络^[10]、匹配网络^[11]、原型网络^[12]以及关系网络。

孪生网络模型构造不同的样本对并输入到网络进行训练,通过比较网络中输出特征的距离来判断样本对是否属于同一类,并产生相应的概率分布。匹配网络采用支持集和查询集构造不同的编码器,最终分类器的输出是支持集和查询集之间预测值的加权求和。原型网络通过学习一个度量空间,在度量空间内以每类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式距离以进行分类。关系网络采用卷积神经网络来学习如何度量2个样本间的相似性,相比于人为设定的距离度量方法,其能够较好地判断2个样本间的相似性,原因是关系网络基于训练集训练网络对于样本的比较能

力,然后利用网络进行小样本识别。

1.2 注意力机制

注意力机制被广泛应用于图像分类任务,在卷积神经网络中添加注意力机制可有效提高网络的特征表达能力。目前,图像分类领域主要有通道注意力机制、空间注意力机制以及混合注意力机制3种。SE(Squeeze&Excitation)^[13]即通道注意力机制,SE模块通过全局池化层对特征图的不同通道进行空间域压缩后沿着通道域进行重新加权,其中,权重大小表示各个通道的重要程度。空间注意力机制即在空间域内添加注意力机制,Non-local Neural Networks^[14]利用特征图中所有位置的特征加权和计算一个位置的响应。本文的注意力模块(CBAM)^[15]采用混合注意力机制,将其空间注意力和通道注意力同时加入模块中。

2 基于ARNs的WCE图像分类方法

2.1 ARNs模型结构

深度学习在计算机视觉领域取得了优异成果,然而,深度卷积神经网络参数量巨大,需要大量的标注数据进行训练,WCE图像由于标注成本大、隐私保护等原因只有少量标注数据,且WCE图像分类还存在类间相似、类内变异度高等问题。如图1所示,不同类WCE图像的差异集中在图像的部分区域,同类WCE图像类内变异度高。

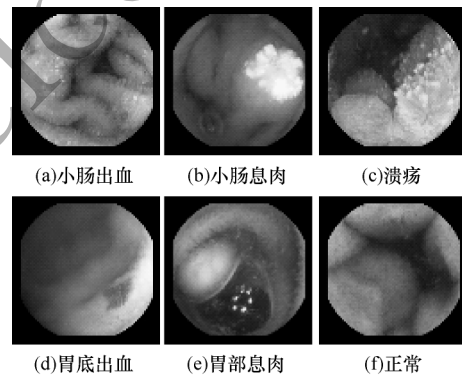


图1 不同类的WCE图像

Fig.1 WCE images of different classes

本文采用基于度量学习的小样本分类方法完成WCE图像多分类任务,其中ARNs模型结构如图2所示,整体架构分为嵌入模块和关系模块。嵌入模块是基于注意力机制的特征提取模块,其将支持样本和查询样本映射到某个特征空间以得到两者的嵌入向量。关系模块首先将查询样本和支持样本在特征空间的嵌入向量进行级联,然后将级联后的特征向量输入到卷积神经网络中,根据网络输出的关系得分来判别查询样本所属的类别。

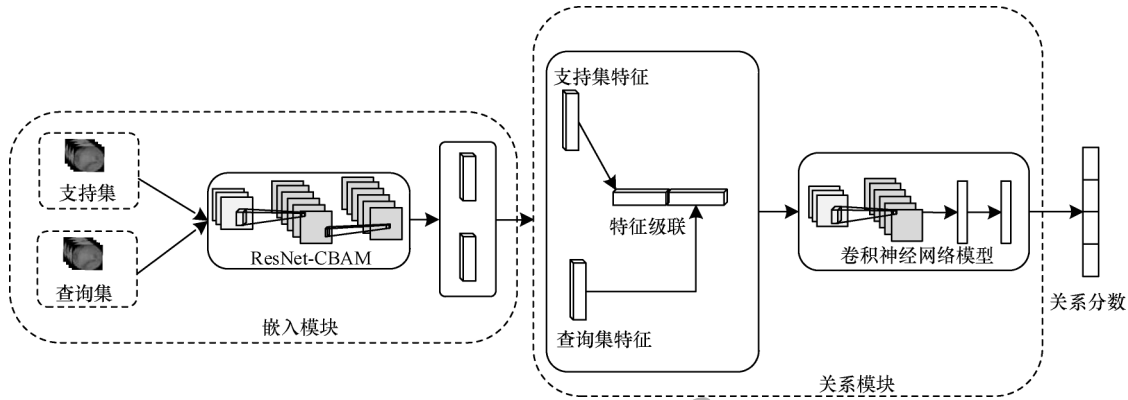


图2 ARNs模型结构

Fig.2 Structure of ARNs model

2.2 ARNs的嵌入模块

关系网络模型采用顺序连接的卷积神经网络结构作为模型的嵌入模块以进行图像特征提取,但是,顺序连接的卷积网络在卷积层数较少时难以获得具有较强表征能力的特征,随着网络的加深,在网络中反向传播的梯度会因连乘变得不稳定从而出现梯度消失的问题,导致网络的性能下降。ResNet^[16]通过加入残差块来解决深度网络中网络性能下降的问题,残差块采用跨层连接的方法,可以让浅层特征过渡到深层继续被重复地学习使用。残差块结构如图3所示,其中, x 表示输入的特征向量, $F(x, w)$ 表示残差函数,恒等映射为 $H(x, w)$ 。

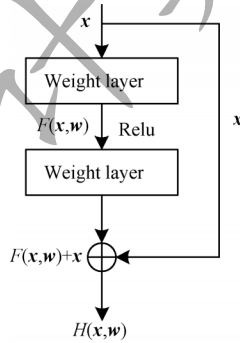


图3 残差块结构

Fig.3 Structure of residual block

残差函数表示为:

$$F(x, w) = w^2 \cdot g(w^1 x + b^1) + b^2 \quad (1)$$

其中, w^1 和 w^2 表示残差模块中第1层和第2层卷积核的权重矩阵, b^1 和 b^2 为第1层和第2层卷积核的偏置矩阵, $g(\cdot)$ 表示Relu激活函数,恒等映射为:

$$H(x, w) = g(F(x, w) + x) \quad (2)$$

残差网络的学习目标是通过恒等映射的学习,将残差函数的结果逼近于0,使得残差块的输入近似于输出,即浅层的特征能够过渡到网络的深层,从而解决梯度消失的问题,保证网络的层数增加但性能不会下降。

注意力可以利用人类视觉机制进行直观解释,人类视觉系统倾向于关注图像中辅助判断的部分信息,忽略掉不相关的信息。在计算机视觉信息处理过程中,输入图像的某些部分会比其他部分更有助于决策,在计算机视觉中添加注意力可以将有限的计算资源分配给图像中有助于预测图像类别的部分。WCE各类图像之间的差异往往只集中在图像的部分区域,并且各类图像之间的差异非常细微。准确分类WCE图像的关键在于将网络的特征提取聚焦于图像中具有足够区分度的关键区域,在WCE图像分类中,捕获到具有判别性的局部特征非常重要。为了关注WCE图像中重要的局部信息,过滤掉不重要的局部信息,本文采用CBAM注意力模块。CBAM是一个简单有效的注意力模块,可添加到前向传播的卷积神经网络模型中,并与模型一起进行端到端的训练。对于网络的中间层特征,CBAM模块分别通过空间和通道2个维度获取到3维的注意力特征图,这种分离的注意力特征图生成过程只需要少量的参数和计算资源。本文采用ResNet作为嵌入模块的主干网络,在此基础上加入CBAM模块,ResNet网络通过多层的卷积实现对WCE图像特征的提取,CBAM模块可以提高对WCE图像中关键特征的敏感度,从而提高嵌入模块的特征表达能力。

ResNet层数可以选择18、34、50、101、152等,在经过实验对比以及综合考虑网络各层的特征表示能力、计算量等因素的基础上,本文根据文献[16]的方法,将CBAM添加到ResNet18中,即在ResNet的每个残差块中插入CBAM。具体做法如图4所示,ResNet中的前一个卷积块输出的特征向量 $F \in \mathbb{R}^{C \times H \times W}$ 作为CBAM的输入首先进入CBAM的通道注意力模块。特征向量 F 在通道注意力模块中分别通过平均池化层和最大池化层获取平均池化特征向量和最大池化特征向量,平均池化可以获取到特征图中每个通道上的平均信息,最大池化考虑到通道上的显著性信息,将两者进行结合,CBAM学习到的特征更具判别性。将2个特征向量输入只包含一

个隐藏层的感知机 MLP(Multi-Layer Perceptron)中,然后将 MLP 输出的 2 个特征进行元素求和运算,并将结果经过 Sigmoid 激活函数以得到一维的通道注

意力特征向量 M_c , M_c 的数学形式如下:

$$M_c(F) = \sigma(\text{MLP}(\text{AvgPool}(F)) + \text{MLP}(\text{MaxPool}(F))) \quad (3)$$

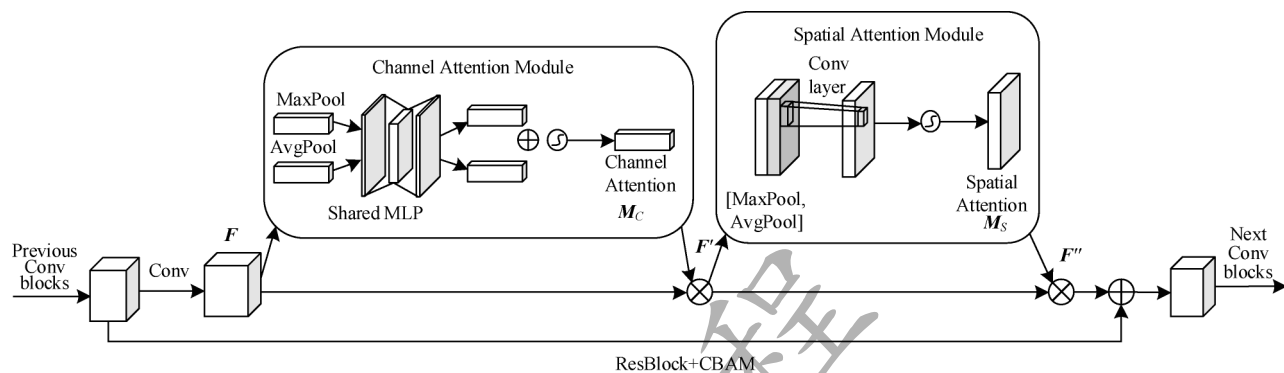


图4 ResNet中加入CBAM的结构示意图

Fig.4 Structure diagram of adding CBAM in ResNet

通道注意力可以理解为通道的权重,包含重要特征的通道权重大,包含不重要特征的通道权重小。将通道注意力特征向量 M_c 以广播的形式送到输入特征图 F 的每个通道上,即可得到中间层特征 F' :

$$F' = M_c(F) \otimes F \quad (4)$$

F' 作为 CBAM 空间注意力模块的输入特征,其首先在通道维度上分别经过最大池化层和平均池化层,将 2 个池化后的特征串联为一个特征向量,然后经过一个卷积层,最后通过一个激活函数输出得到空间注意力特征 M_s , M_s 的数学形式如下:

$$M_s(F) = \sigma(f^{7 \times 7}[\text{AvgPool}(F); \text{MaxPool}(F)]) \quad (5)$$

空间注意力特征可以理解为通道上每个像素的权重,包含重要信息的像素权重大,包含不重要信息的像素权重小。将空间注意力特征图 M_s 以广播的形式送到最开始输入的特征 F' 上,即得到整个卷积注意力模块的最终特征图 F'' :

$$F'' = M_s(F') \otimes F' \quad (6)$$

CBAM 模块分别学习通道注意力和空间注意力,通道注意力集中关注特征的含义,空间注意力关注重要特征的位置,从而使得网络聚焦于 WCE 图像的重要特征。通道注意力机制通过共享的全连接层实现,由于池化层没有引入可学习参数,因此 CBAM 模块是一个轻量级模块,减小了使用注意力机制所需要的参数量,使卷积神经网络模型的训练过程变得更加高效,因此,本文采用结合 CBAM 的 ResNet18 作为 ARNs 的嵌入模块,添加了 CBAM 的 ResNet18 在取得更好特征表达能力的同时,并没有引入过多的参数量和计算量。如图 5(a) 所示,嵌入模块主要由 4 个卷积单元组成,其中,layer1、layer2、layer3 和 layer4 都包含 2 个如图 5(b) 所示的 ResNetBlock+CBAM 模块。各层网络的卷积核参数如表 1 所示。

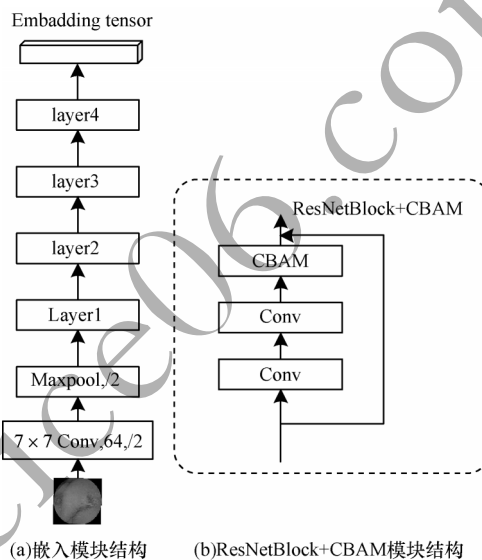


图5 嵌入模块结构

Fig.5 Embedded module structure

表1 卷积核参数

Table 1 Convolution kernel parameters

layer	卷积核参数
layer1	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$
layer2	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$
layer3	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$
layer4	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$

2.3 ARNs的关系模块

目前,基于度量学习的小样本学习方法主要采用一种固定的距离度量方法(如欧式距离和余弦距离)来衡量查询样本与支持样本之间的相似性,这种

方法主要集中构造一个可学习的嵌入模块,以适应预先指定的距离度量函数,但是,当嵌入模块学习到的特征区分信息不充分时网络的性能将受到限制,关系网络的关系模块采用卷积神经网络学习样本之间的相似性,以数据驱动的方式学习一个度量函数,相比于固定的度量函数,其可以更好地比较2个样本之间的相似性。关系模块以查询样本和支持样本的级联作为输入,生成一个0~1的分数表示查询样本与支持样本之间的相似性。本文采用关系网络中的关系模块作为 ARNs 的关系模块,其结构如图6所示,关系网络中的关系模块包含2个卷积块、2个全连接层和1个 Sigmoid 函数,每个卷积块后接1个最大池化层。关系模块的输入特征是一个3维的特征图,3维的特征图包含图像的信息,适用于卷积神经网络学习特征之间的关系,根据输入特征的大小,设置卷积块中卷积层的卷积核大小为3×3,卷积核数量为64,其后接一个归一化层、Relu非线性激活函数以及2×2的最大池化层以进行下采样。第1个全连接层后连接1个 Relu 激活函数,最后1个全连接层后添加1个 Sigmoid 激活函数,通过 Sigmoid 函数输出1个0~1的相似性评分。

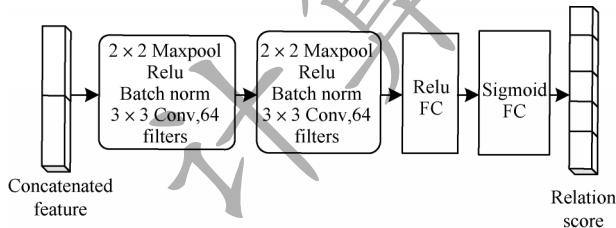


图6 关系模块结构

Fig.6 Relationship module structure

2.4 ARNs 的训练策略

元学习是一种处理小样本学习的常用方法,它包含训练阶段和测试阶段,在训练阶段通过元学习提取可传播的知识,允许模型在测试阶段执行小样本学习任务。在训练阶段,模型按照一个个训练批次来训练网络。在每个训练批次中,首先构造一个元任务,从训练集中随机抽取 c 个类别,每个类中随机抽取 k 个样本合并作为支持集,再从每个类随机抽取若干样本合并组成查询集,然后更新模型,这样的小样本学习任务通常被描述为 c -way k -shot。在测试阶段,数据分为支持集和测试集,支持集作为对比学习的样本,拥有和测试集一样的标签,在测试过程中,采用训练阶段学到的模型,通过和支持集的对比来预测查询集中的样本标签。

ARNs 模型中查询集与支持集之间的相似度是

通过关系模块输出的关系评分来进行判别的,关系评分是一个0~1的标量,0代表极不相似,1代表非常相似,此时支持样本与查询样本之间的关系评分获得可视回归问题,采用均方误差来计算损失值,损失函数表示为:

$$J_{\text{MSE}}(\varphi) = \sum_{i=1}^m \sum_{j=1}^n (r_{i,j} - z)^2 \quad (7)$$

其中, $r_{i,j}$ 为查询集样本 x_j 和支持集样本 x_i 的关系评分。在小样本 WCE 图像分类任务中,由于样本数量少,类间变异度高,均方差损失函数训练的网络容易因数据量少而产生过拟合。因此,本文采用均方误差损失函数加上 L2 正则化约束项作为目标函数以训练模型,从而增强模型的鲁棒性。损失函数定义为:

$$J_{\text{MSE}}(\varphi) = \sum_{i=1}^m \sum_{j=1}^n (r_{i,j} - z)^2 + \lambda (\|\varphi\|_2^2) \quad (8)$$

其中, λ 是正则化惩罚系数, $z = \begin{cases} 1, y_i = y_j \\ 0, y_i \neq y_j \end{cases}$, y_i 表示支持集样本标签, y_j 表示查询集样本标签。反向传播梯度求导公式为:

$$\frac{dJ(\varphi)}{d\varphi} = \frac{dJ_{\text{MSE}}(\varphi)}{d\varphi} + \frac{d\lambda(\|\varphi\|_2^2)}{d\varphi} = \nabla J_{\text{MSE}}(\varphi) + \lambda\varphi \quad (9)$$

梯度下降公式为:

$$\varphi = \varphi - \eta (\nabla J_{\text{MSE}}(\varphi) + \lambda\varphi) = (1 - \eta\lambda)\varphi - \eta \nabla J_{\text{MSE}}(\varphi) \quad (10)$$

其中, φ 表示网络中可学习的参数, η 是学习率。在网络的损失函数中,经过反向传播训练的网络模型加入了正则化项后,参数的权重衰减使得网络权重不断减小,提高了 ARNs 的抗扰动能力,减缓了测试数据差异情况下的过拟合问题。

本文采用元学习的训练策略训练 ARNs 模型,训练 ARNs 模型的目标是找到模型用于小样本 WCE 图像分类任务的合适参数,定义训练集 $D_{\text{train}} = \{(x_i, y_i), \dots, (x_N, y_N)\}$, N 是训练集的样本数量, y_i 表示样本标签, x_i 表示样本,每批次训练从 D_{train} 中随机采样 c 类别的 k 个样本,构成支持集 $S = \{(x_i, y_i)\}_{i=1}^m$, $m = c \times k$,从剩余 D_{train} 中随机抽取 n 个样本构成查询集 $Q = \{(x_j, y_j)\}_{j=1}^n$,嵌入模块的映射函数为 f_φ ,关系模块的映射函数为 g_φ 。一次网络训练的算法过程如下:

输入 训练集 $D_{\text{train}} = \{(x_i, y_i), \dots, (x_N, y_N)\}$

输出 经过一次训练的损失 J

1)从 D_{train} 中随机抽取支持集 $S = \{(x_i, y_i)\}_{i=1}^m$ 和查询集 $Q = \{(x_j, y_j)\}_{j=1}^n$ 。

2)计算支持集样本的嵌入向量 $f_\phi(x_i)$ 和查询集样本的嵌入向量 $f_\phi(x_j)$ 。

3)级联嵌入向量得到连接向量 $\text{Concat}[f_\phi(x_i), f_\phi(x_j)]$ 。

4)将连接向量输入关系模块得到关系值 $r_{i,j} = g_\phi\{\text{Concat}[f_\phi(x_i), f_\phi(x_j)]\}$, $i=1, 2, \dots, c$, 采用均方误差(MSE)计算损失值 $J_{\text{MSE}}(\phi) = \sum_{i=1}^m \sum_{j=1}^n (r_{i,j} - z)^2 + \lambda (\|\phi\|_2^2)$ 。

3 实验与结果分析

3.1 数据集和实验设置

本文采用 mini-ImageNet 数据集和 WCE 数据集进行实验。其中, mini-ImageNet 数据集是小样本学习的一个基准数据集,该数据集包含 100 类图像数据,每类由 600 幅图像组成,一共 60 000 幅图像; WCE 数据集由金山科技公司提供的 WCE 数据集和 KID^[17] 数据集组成,金山科技公司提供的 WCE 数据集由 OMOM 内镜拍摄的 WCE 图像组成,其包括胃部凹陷、胃部表浅、胃底隆起等 16 类 WCE 图像,每类包含 200~300 张 WCE 图像, KID 是一个公开的 WCE 图像数据集,由 3 个数据集组成。本次实验采用 KID 中的数据集 1 和数据集 2 混合组成 KID 数据集,且输入到 ARNs 模型的 WCE 图像大小统一为 224 像素×224 像素。

实验采用 PyTorch 深度学习框架,在 GPU 平台上运行,处理器为 Intel® Core™ i7-7700HQ CPU@2.80 GHz、GPU NVIDIA GeForce GTX 1080,运行内存为 64 GB。训练和测试阶段都是基于 c -way k -shot 的小样本学习方法进行,所有实验均采用 Adam^[18] 优化算法对模型参数进行更新,初始学习率设置为 0.001。

3.2 mini-ImageNet 数据集的小样本分类任务

本文对 mini-ImageNet 数据集进行划分,采用 64 类数据作为训练集,16 类数据作为验证集,20 类数据作为测试集,分别采用 5-way 1-shot 和 5-way 5-shot 的方法进行实验,并将 ARNs 模型的实验结果与 Baseline-linear^[19] 网络、Meta-Learner LSTM^[20] 网络、MAML 网络^[21]、匹配网络^[12]、原型网络^[13] 和关系网络^[9] 这几种小样本学习领域的主流方法进行对比,实验采用分类准确率 $A = n'/n$ 来衡量方法的性能,其中, n' 为测试集中预测结果与真实结果一致的样本个数, n 为测试集的样本总个数,实验结果如表 2 所示。

表 2 小样本分类方法性能对比

Table 2 Performance comparison of small sample

classification methods		%	
方法	微调	5-way 1-shot	5-way 5-shot
Baseline-linear ^[19]	是	40.82	52.14
Meta-Learner LSTM ^[20]	否	42.37	61.81
MAML ^[21]	是	47.48	64.01
Matching network ^[12]	是	43.74	56.13
Prototypical network ^[13]	否	48.93	67.51
RelationNet ^[9]	否	50.34	64.19
ARNs	否	53.38	67.23

从表 2 可以看出,本文提出的 ARNs 模型在 mini-ImageNet 数据集上准确率相比其他模型有所提升。与关系网络相比, ARNs 在小样本数据集上的分类效果明显提升,原因是本文引入的基于 CBAM 的 ResNet 嵌入模块提取的特征较关系网络中简单的线性连接的卷积网络具有更高的鲁棒性,加强了同类样本特征之间的相关性,降低了不同类样本特征的相似性,从而有效提高了 ARNs 模型应用于小样本学习任务的分类准确率。

3.3 WCE 数据集的小样本分类任务

对 WCE 数据集进行划分,采用金山科技公司提供的 13 类 WCE 图像作为训练集,剩下的 3 类 WCE 图像混合 KID 数据集组成测试集。测试集包含 189 张出血、157 张息肉、125 张溃疡和 193 张正常类 WCE 图像,这些 WCE 图像是临床实践中常出现的 4 类 WCE 图像。首先采用训练集对模型进行训练,同时将 4 类 WCE 图像作为测试集模拟验证在真实场景下 ARNs 模型对于少量 WCE 图像的分类能力。原则上只采用测试集中的 4 类 WCE 图像数据就可以训练 ARNs 模型的嵌入模块和关系模块,但是,测试集中标记样本很少导致这种模型的性能不能满足要求。因此,本文首先采用元学习的方法在训练集上训练 ARNs 模型,从而提取可以迁移的知识,在测试集上的 WCE 图像类中完成小样本的图像分类,分别采用 4-way 1-shot 和 4-way 5-shot 的方法进行实验,使用准确率 A 、灵敏度(sensitivity sen)和特异性(specificity spec)作为评价指标,后两者计算公式如下:

$$S_{\text{sen}} = \frac{TP}{TP + FN} \quad (11)$$

$$S_{\text{spec}} = \frac{TN}{TN + FP} \quad (12)$$

其中, TP 表示把原本属于该类的样本分成该类, TN 表示把原本属于其他类的样本分成其他类, FP 表示把原本属于其他类的样本错分成了该类, FN 表示把原本属于该类的样本错分成了其他类。

为了探究 ARNs 采用不同嵌入模块对于 WCE 图像分类效果的影响,本文分别以基于 CBAM 的 ResNet18 和基于 CBAM 的 ResNet50 作为 ARNs 模型的嵌入模块进行实验,结果如表 3 所示。

表 3 不同嵌入模块的性能对比

Table 3 Performance comparison of different embedded modules

嵌入模块	准确率/%		参数个数	训练时间/s	
	4-way 1-shot	4-way 5-shot		4-way 1-shot	4-way 5-shot
CBAM-ResNet18	81.72	90.28	13 161 024	433	793
CBAM-ResNet50	79.37	89.50	24 636 800	527	921

从表 3 可以看出,ARNs 模型采用 CBAM-ResNet18 和 CBAM-ResNet50 作为嵌入模块时的准确率相差不大,考虑到模型的参数量以及计算复杂度,本文采用 CBAM-ResNet18 作为 ARNs 模型的嵌入模块,以提取 WCE 图像特征。

图 7 所示为 WCE 图像测试集上 ARNs 模型与 RelationNet 前 6 000 次迭代(每隔 50 次)时的准确率和损失曲线,从图 7 可以看出,在小样本的 WCE 图像上,本文 ARNs 网络与 RelationNet 相比分类效果显著提升。

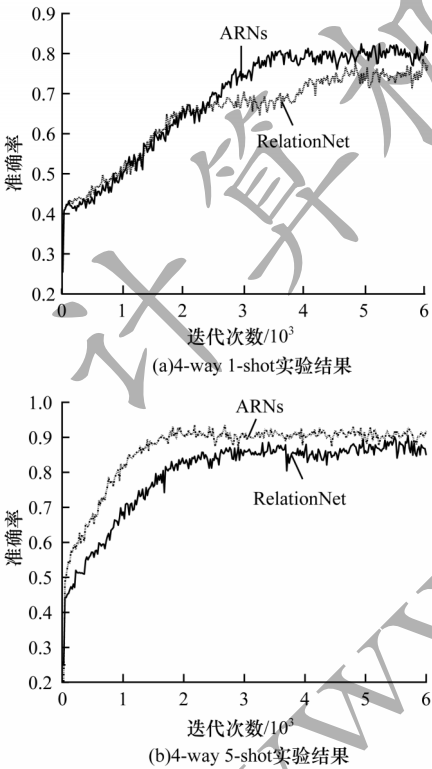


图 7 ARNs 和 RelationNet 的性能比较
Fig.7 Performance comparison between ARNs and RelationNet

WCE 图像各类之间相似,图像类内变异度高,提高 WCE 图像分类准确率的关键在于提取各类 WCE 图像之间的差异部分特征。ARNs 模型的特征提取模块采用结合注意力机制的 ResNet18,相比于关系网络的特征提取模块,其能提取到对 WCE 图像表示性更强的特征。因此,ARNs 模型在 WCE 数据集上的准确率明显高于关系网络。

表 4 所示为测试数据的混淆矩阵,其中包括各类 WCE 图像的灵敏度和特异性。从表 4 可以看出,ARNs 模型对于出血类 WCE 图像的准确率最高,因为出血类 WCE 图像的特征和其他类特征具有明显差异,ARNs 模型对溃疡类 WCE 图像的准确率较低,溃疡类图像类内变异度高,小样本分类的模型难以涵盖其特征的多样性。ARNs 模型的错误分类主要发生在溃疡类和息肉类之间,溃疡类与息肉类特征具有相似性,并且 2 类图像的类内变异度高,导致模型在 2 类 WCE 图像上的识别精度下降。ARNs 模型在 4 类 WCE 图像上的特异性值都取得了较好的结果,表示模型在 WCE 图像上的误诊率很低。

表 4 WCE 图像多异常分类混淆矩阵

Table 4 Classification confusion matrix of WCE image multiple anomalies

类别	出血	息肉	溃疡	正常	sen
出血	97.13	5.12	4.57	0.00	92.13
息肉	2.62	86.69	11.49	1.23	80.69
溃疡	2.49	9.33	84.47	3.85	76.47
正常	0.00	2.47	2.39	93.84	83.84
spec	96.47	93.15	92.73	94.32	—

类激活热图(Class Activation Mapping, CAM)是一种可视化卷积神经网络的工具,本文采用 CAM 卷积可视化技术将基于 CBAM 的 ResNet18 嵌入模块提取的特征进行热图可视化,并与关系网络的嵌入模块热图可视化的结果进行对比。如图 8 所示,采用基于 CBAM 的 ResNet18 嵌入模块提取 WCE 图像特征时能够聚焦于图像的异常区域,可以提取到图像中的差异性特征,有助于提高 WCE 图像的多分类准确率。不同 WCE 图像分类方法的准确率对比结果如表 5 所示,其中,文献[6]采用 MPEG-7v 视觉描述符分类 WCE 图像,文献[7]利用 WCE 图像的颜色和纹理特征进行图像分类,文献[8]利用词袋库技术,采用以 ResNet18 作为基础网络进行微调的深度学习多分类方法分类 WCE 图像,文献[10]利用 WCE 训练集训练关系网络。从表 5 可以看出,本文 WCE 图像多分类方法整体准确率以及在各种异常图像上的识别精度高于其他 WCE 图像分类方法,因此,本文 WCE 多异常分类方法具有有效性。

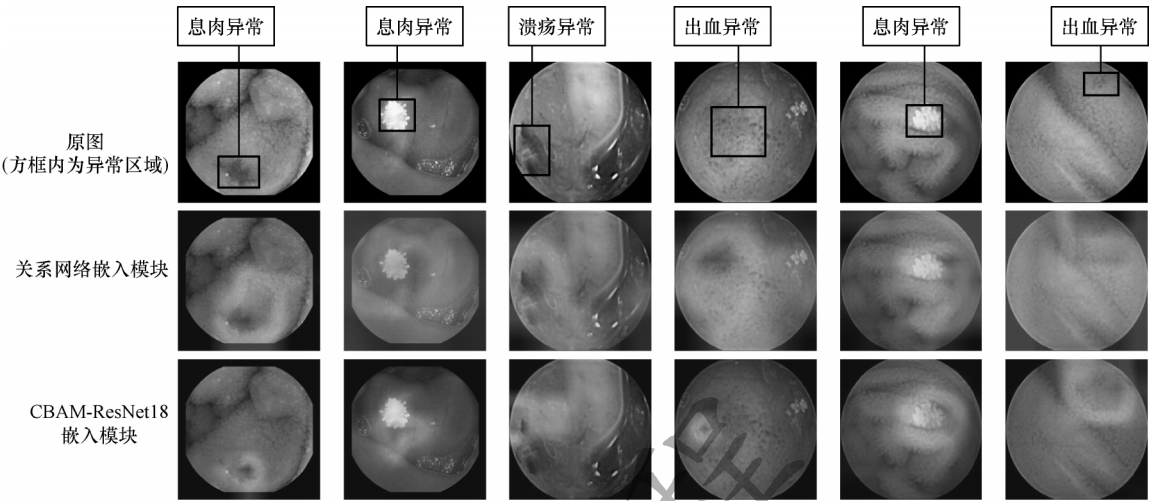


图 8 嵌入模块卷积的可视化分析

Fig.8 Visualization analysis of embedded module convolution

表 5 不同 WCE 图像分类方法准确率对比

Table 5 Comparison of the accuracy of different

WCE image classification methods					%
方法	类识别 准确率	出血类 准确率	息肉类 准确率	溃疡类 准确率	正常类 准确率
文献[6]方法	75.21	85.65	74.12	71.84	80.64
文献[7]方法	84.35	90.95	80.23	78.23	85.20
文献[8]方法	86.54	95.28	84.14	83.83	88.73
ResNet18 ^[18]	84.72	95.27	85.74	80.31	87.27
RelationNet ^[10]	86.13	94.41	85.84	81.15	90.13
ARNs	90.28	97.13	87.69	84.47	94.32

4 结束语

本文提出一种基于注意力关系网络的小样本 WCE 图像分类方法。将注意力机制、关系网络以及元学习训练策略相结合,构造基于注意力机制的特征提取模块提取 WCE 图像特征,将提取的支持样本和查询样本特征级联后输入到关系模块,关系模块输出查询样本与支持样本之间的相似性评分,根据相似性评分判别查询样本的类别。在 mini-ImageNet 数据集上的实验结果表明,该方法的准确率高于 ResNet18、RelationNet 等其他小样本分类方法。下一步将对 ARNs 关系模型中的特征级联部分和模型结构进行改进,以达到更好的 WCE 图像分类效果。

参考文献

[1] IDAN G, MERON G, GLUKHOVSKY A, et al. Wireless capsule endoscopy[J]. Nature, 2000, 405(6785): 417-427.
[2] YUAN Y, LI B, MENG M Q H. Bleeding frame and region detection in the wireless capsule endoscopy video[J]. IEEE Journal of Biomedical and Health Informatics, 2015, 20(2): 624-630.

[3] QADIR H A, BALASINGHAM I, SOLHUSVIK J, et al. Improving automatic polyp detection using CNN by exploiting temporal dependency in colonoscopy video[J]. IEEE Journal of Biomedical and Health Informatics, 2019, 24(1): 180-193.
[4] GOYAL M, REEVES N D, RAJBHANDARI S, et al. Robust methods for real-time diabetic foot ulcer detection and localization on mobile devices[J]. IEEE Journal of Biomedical and Health Informatics, 2018, 23(4): 1730-1741.
[5] JEBARANI W S L, DAISY V J. Assessment of Crohn's disease lesions in wireless capsule endoscopy images using SVM based classification [C]//Proceeding of 2013 International Conference on Signal Processing, Image Processing & Pattern Recognition. Washington D. C. , USA: IEEE Press, 2013: 303-307.
[6] COIMBRA M T, CUNHA J P S. MPEG-7 visual descriptors—contributions for automated feature extraction in capsule endoscopy[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2006, 16(5): 628-637.
[7] YUAN Y, LI B, MENG M Q H. WCE abnormality detection based on saliency and adaptive locality-constrained linear coding[J]. IEEE Transactions on Automation Science and Engineering, 2016, 14(1): 149-159.
[8] YUAN Y, YAO X, HAN J, et al. Discriminative joint-feature topic model with dual constraints for WCE classification[J]. IEEE Transactions on Cybernetics, 2017, 48(7): 2074-2085.
[9] SUNG F, YANG Y, ZHANG L, et al. Learning to compare: relation network for few-shot learning[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA: IEEE Press, 2018: 1199-1208.
[10] KOCH G R, ZEMEL R, SALAKHUTDINOV R. Siamese neural networks for one-shot image recognition [C]//Proceedings of the 32nd International Conference on Machine Learning. Washington D. C. , USA: IEEE Press, 2015: 103-123.

(上接第 259 页)

- [11] VINYALS O, BLUNDELL C, LILICRAP T, et al. Matching networks for one shot learning [C]//Proceeding of Advances in Neural Information Processing Systems. Barcelona, Spain; NIPS, 2016; 3630-3638.
- [12] SNELL J, SWERSKY K, ZEMEL R. Prototypical networks for few-shot learning [C]//Proceeding of Advances in Neural Information Processing Systems. Barcelona, Spain; NIPS, 2017; 4077-4087.
- [13] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA; IEEE Press, 2018; 7132-7141.
- [14] WANG X, GIRSHICK R, GUPTA A, et al. Non-local neural networks [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA; IEEE Press, 2018; 7794-7803.
- [15] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module [C]//Proceedings of European Conference on Computer Vision. Berlin, Germany; Springer, 2018; 3-19.
- [16] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA; IEEE Press, 2016; 770-778.
- [17] KOULAOUZIDIS A, IAKOVIDIS D K, YUNG D E, et al. KID project: an internet-based digital video atlas of capsule endoscopy for research purposes [J]. Endoscopy International Open, 2017, 5(6): 477-450.
- [18] KINGMA D P, BA J. Adam: a method for stochastic optimization [EB/OL]. [2020-05-03]. <https://arxiv.org/abs/1412.6980>.
- [19] CHEN W Y, LIU Y C, KIRA Z, et al. A closer look at few-shot classification [EB/OL]. [2020-05-03]. <http://arxiv.org/abs/1904.04232>. pdf.
- [20] RAVI S, LAROCHELLE H. Optimization as a model for few-shot learning [C]//Proceedings of the 5th International Conference on Learning Representations. Washington D. C., USA; IEEE Press, 2017; 123-135.
- [21] FINN C, ABBEEL P, LEVINE S. Model-agnostic meta-learning for fast adaptation of deep networks [EB/OL]. [2020-05-03]. <https://arxiv.org/pdf/1703.03400.pdf>.

编辑 吴云芳