

基于胶囊网络的动态路由研究与改进

陈 珊,孙仁诚,邵峰晶,隋 毅

(青岛大学 计算机科学技术学院,山东 青岛 266071)

摘 要: 胶囊网络具有弥补卷积神经网络空间信息丢失和旋转不变性差的优点,已被广泛应用于图像分类、目标检测以及文本检测等多个领域,但胶囊网络仍存在参数量大且分类精确度低的问题。提出基于点乘注意力图卷积路由的胶囊网络分类模型。在同级胶囊之间构建连通图,通过注意力机制获取胶囊间的依赖关系,利用影响因素大的预测胶囊进行特征聚类,改变使用迭代更新高低胶囊层间耦合系数的动态路由方式,降低参数量并提升模型的分类准确率。此外,在特征提取部分加入残差网络提取更高维的特征以优化胶囊质量,在提升模型特征表达能力的同时可抑制模型过大。实验结果表明,在参数量小于多个胶囊网络变体的情况下,该模型在MNIST、FashionMNIST、CIFAR10和SVHN数据集上的精度分别达到99.74%、95.02%、91.78%和95.65%,均高于MS-CapsNet、TextCaps、AR CapsNet、FSc-CapsNet、DA-CapsNet等对比模型。

关键词: 胶囊网络;动态路由;点乘注意力机制;图卷积;图像分类

开放科学(资源服务)标志码(OSID):



中文引用格式:陈珊,孙仁诚,邵峰晶,等.基于胶囊网络的动态路由研究与改进[J].计算机工程,2022,48(5):208-214.

英文引用格式:CHEN S, SUN R C, SHAO F J, et al. Research and improvement of dynamic routing based on capsule network[J]. Computer Engineering, 2022, 48(5): 208-214.

Research and Improvement of Dynamic Routing Based on Capsule Network

CHEN Shan, SUN Rencheng, SHAO Fengjing, SUI Yi

(School of Computer Science and Technology, Qingdao University, Qingdao, Shandong 266071, China)

[Abstract] Capsule Network (CapsNet) has the advantage of compensating for the loss of spatial information and rotation invariance in convolutional neural networks. Therefore, they have been widely used in many fields such as image classification, target detection, and text detection. However, capsule networks still face some problems, such as having a large amount of parameters and low classification accuracy. This paper proposes a novel capsule network classification model that uses the dot-product attention mechanism to construct the graph convolutional layer, thereby selecting high-quality features to facilitate correct classification of the image. To obtain the spatial dependence between capsules within the same layer, the dot-multiplication attention mechanism is used to construct a fully connected graph between capsules of the same levels. The predicted capsules with large influencing factors are then used for feature clustering. In this way, the proposed model can reduce the number of parameters and improve model performance. At the same time, the residual network is added to the feature extraction process for the extraction of higher-dimensional features in order to optimize the quality of the capsule, which can not only improve the feature expression ability of the model but also restrain its size. The experimental results prove that when the number of parameters is less than that of multiple capsule network variants, the proposed model achieves 99.74%, 95.02%, 91.78% and 95.65% accuracy on the MNIST, FashionMNIST, CIFAR10, and SVHN datasets, respectively, making the proposed model more accurate than MS-CapsNet, TextCaps, AR CapsNet, FSc-CapsNet, and DA-CapsNet models.

[Key words] Capsule Network (CapsNet); dynamic routing; dot-product attention mechanism; graph convolution; image classification

DOI: 10.19678/j.issn.1000-3428.0062928

0 概述

随着人工智能技术的发展,目前卷积神经网络

(CNN)在图像识别^[1]、图像分割^[2]、文本分类^[3]等领域中获得了巨大的成功。但是随着CNN应用,发现其存在识别物体空间关系的能力不强和对物体旋转

基金项目:国家自然科学基金青年基金项目(41706198)。

作者简介:陈珊(1996—),女,硕士研究生,主研方向为动态路由机制、机器学习、深度学习;孙仁诚(通信作者),教授、博士;邵峰晶,教授、博士、博士生导师;隋毅,副教授、博士。

收稿日期:2021-10-12 修回日期:2021-11-16 E-mail:cs1421159456@163.com

后的特征识别能力较弱两大缺陷。因为CNN通过池化选取一个区域中最活跃的点或者平均聚合所有特征,这种降采样操作会丢失物体的空间信息,不利于图像的正确分类。CNN采取的解决方案是增加网络层数或数据量来提升性能,但这会造成模型参数数量和训练时间的增加。

为了弥补CNN在图像处理中丢失物体部件间的相对位置关系及旋转不变性差的缺点,SABOUR等^[4]提出了胶囊网络(Capsule Network, CapsNet)模型。CapsNet在借鉴CNN空间特征提取方面优势的基础上,提出了胶囊的结构及动态路由的特征选择方式。在胶囊网络中,低层胶囊用来刻画物体的局部特征,高层胶囊用来表达整体抽象特征,然后使用动态路由机制聚合低层胶囊信息更新高层胶囊。与CNN相比,胶囊网络仅需少量训练数据即可在小规模图像分类中取得更好的训练效果,并具有更佳的泛化性^[5-6]和可解释性^[7-8]。

然而,胶囊网络在特征提取和动态路由机制中仍有较多亟需改进之处。首先,CapsNet在特征提取部分仅使用了一层卷积提取特征,感受野太小,提取不到足够的图像信息,导致生成的初级胶囊质量不高,从而影响模型的分类精度。其次,CapsNet的动态路由机制在将低层胶囊中的信息传递到高层胶囊时使用了相同的优先级,但对于场景比较复杂的图像,动态路由则需要更多的迭代次数不断调整胶囊之间的匹配程度,才能提取到帮助图像分类的有效特征^[9]。随后,在迭代过程中,每个低层胶囊还只能隶属于一个高层胶囊,耦合系数会抑制一个有效特征对多个高层胶囊的支持,由此导致一些有效特征不能表达,最终影响图像的分类效率和准确率^[10]。最后,关于动态路由的研究均是从低层胶囊和高层胶囊之间的相似性来筛选特征,忽略了同层胶囊不同位置的像素之间的空间相关性,从而损失掉有益于图像分类的重要特征。此外,目前关于胶囊网络的大部分研究,均是在通过增加网络层数来提升性能,但是这会增加模型的参数量,从而造成模型在计算资源受限的环境下不适用。

本文提出一种基于点乘注意力的胶囊图网络(Dot-Product Attention Capsule Graph Network, DPA-CapsGraph)分类模型。该模型主要由特征提取模块、胶囊层部分以及重构部分组成,模型目标是以较小的参数量实现更多的有效特征能被高效地传递到高层胶囊,从而达到更高的分类精度,以增加模型的泛化性和鲁棒性。

1 相关工作

胶囊网络使用向量神经元而非标量表示物体在某一位置的特征,其中向量的姿态表示特征实体类别的属性,输出向量的长度表示某一类别实体存在可能性的大小。向量胶囊将物体的位置和姿态信息

封装在胶囊内部,低层胶囊用来刻画物体的具体特性,而高层胶囊用来提取物体的总体抽象特征。随机动态路由迭代更新低层胶囊和高层胶囊的耦合系数 C_{ij} ,最后通过压缩函数得到表示物体类别信息的高级胶囊。最终将向量长度最大的高级胶囊输入到重构部分,得到模型编解码后输出的图像。为提升胶囊网络在图像分类领域的分类准确率,目前的研究主要分为创新胶囊结构形式来优化胶囊质量,或提出新型动态路由机制来选择特征两个方面。

为提高胶囊网络在特征提取方面的表现能力,文献[11]通过多尺度特征提取获得结构和语义信息构建不同尺度的初级胶囊,用不同维度的转换矩阵生成相同维度的预测胶囊,增加了可学习的参数量。文献[12]通过可学习的转换矩阵将输入的特征向量正交投影到胶囊子空间上来创建胶囊,使其可以学习到更多隐层的特征信息。文献[13]通过提出使用两种类型的可变形胶囊,增加特征提取的感受野并增强胶囊对于数据的自适应性。文献[14]将多个相似的初级胶囊通过卷积操作得到张量胶囊,并通过增加胶囊层数提升模型效果。但是动态路由因为迭代更新的特性不能在网络层中并行执行,顺序执行多个胶囊层增加了模型的参数量和计算时间。文献[15]使用两个密集残差网络分别构造胶囊的姿态矩阵和预测矩阵,然后用奇异值分解检测实体在一维线性子空间上提取的特征,替换了原始的动态路由策略。

同时,关于随机动态路由的改进也得到了广泛关注,不同的动态路由表示低层胶囊聚合到高层胶囊的方式不同。文献[16]通过多组消融实验证明,随机动态路由会削弱模型的健壮性。因此,有些学者从聚类最优化的角度来改进动态路由过程。文献[17]通过最小化初级胶囊分布与高级胶囊的聚类损失和KL正则化的组合,使动态路由过程更平滑。文献[18]运用最大-最小正则化标准化耦合系数,从而允许初级胶囊可以对多个高级胶囊具有高分配概率。上述研究的动态路由都是聚焦于高层胶囊和低层胶囊之间的一致性,忽略了同层胶囊之间的内在联系,对于图像分类问题,同层胶囊之间的空间关系会影响到分类的准确性。

注意力机制^[19]则允许模型动态地捕获更多与解决问题相关的特征信息。TSAI等^[10]稍微改变了原始的动态路由,通过倒点积机制计算高水平胶囊的姿势和低水平胶囊的投票之间的一致性,并提出使用并行迭代路由取代顺序执行,在所有胶囊层之间同时执行路由过程。但是由于含有多个胶囊层并发执行动态路由机制,大幅增加了模型的训练时间,并对设备内存提出了较高的要求。文献[20]是在初级胶囊层前增加使用注意力机制的网络层,提升模型在特定数据集上的分类精度。文献[21-22]通过在网络中加入注意力模块来提升胶囊网络在特征提取

部分的能力。基于以上研究发现,注意力机制在胶囊网络中大多被用在特征提取部分。本文将点乘注意力机制用来改进胶囊层之间的路由机制,利用图卷积层式传播的思想来探究同级胶囊间所包含的大量特征对于提升分类准确性的重要性,从而充分利用同级特征来提升模型的分类效果。

本文在 MNIST、FashionMNIST、CIFAR10 和 SVHN 4 种数据集上进行了多次实验,通过和 6 个胶囊网络模型的变体进行对比,DPA-CapsGraph 不仅降低了模型的参数量,而且在 4 种数据集上均取得了更高的分类准确率。

本文的主要贡献包括以下两个部分:

1)为了优化初级胶囊的质量,使胶囊表达更多的特征信息,模型在特征提取部分加入跳跃连接来增加初级胶囊对特征的表达能力,为模型的训练提供更多的信息。

2)为减少参数量,提高特征选择的效率以及质量,本文采用图卷积的层式传播方式代替动态路由进行特征选择,可以在控制网络层数增长的同时提升模型的分类准确度。利用点乘注意力机制获得同层胶囊之间的特征依赖关系,综合评估当前层中所有胶囊内部隐藏状态之间的异同点,然后通过特征映射的方式使特征在空间上自动聚类,提高模型的分类准确率。

2 基于点乘注意力图卷积路由的胶囊网络分类模型

针对 CapsNet 的卷积层提取特征部分和动态路由部分存在的问题,本文通过借鉴残差网络和图卷积网络的思想,设计了 DPA-CapsGraph。

基于点乘注意力的图卷积路由的胶囊网络的模型框架如图 1 所示。

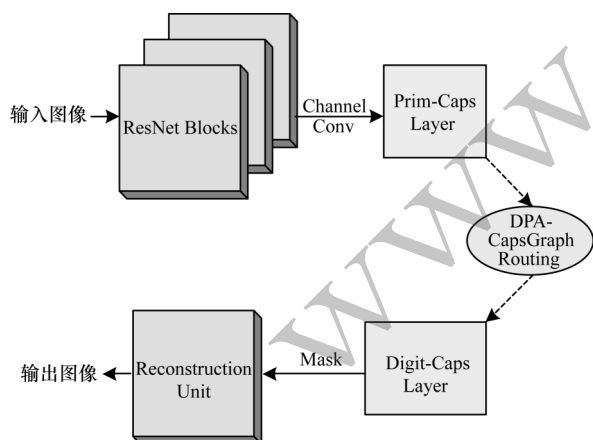


图1 DPA-CapsGraph 模型框架

Fig.1 DPA-CapsGraph model framework

DPA-CapsGraph 主要分为 3 个部分:

1)特征提取部分,利用去掉池化层的残差网络来提取既有具体的实体属性特征又有概括的抽象特征。

2)胶囊网络层,先对特征提取部分传入的特征进行通道卷积得到胶囊结构,然后通过路由过程得到高级胶囊。本文的重点是使用点乘注意力机制在同层胶囊间构建全连通图,以此获得同级胶囊中的丰富特征以及空间依赖关系,即将原始的动态路由机制改为利用注意力机制的图卷积来实现图像的正确分类。

3)重构层,将分类正确的高级胶囊重构成原始的图像。

2.1 特征提取模块

本文在特征提取部分加入恒等映射构建残差网络,训练时只需学习残差,解决了传统卷积层的特征丢失、损耗等问题,在一定程度上保护了信息的完整性。跳跃连接使得特征提取部分的输出不仅包含特定实体的属性信息,还包含图像整体结构的抽象信息。其次去掉池化层以保留实体特征之间的空间依赖关系,最后将提取到的具体属性和空间信息输入到初级胶囊层,增强胶囊的特征表达能力,优化胶囊质量,具体结构如图 2 所示。

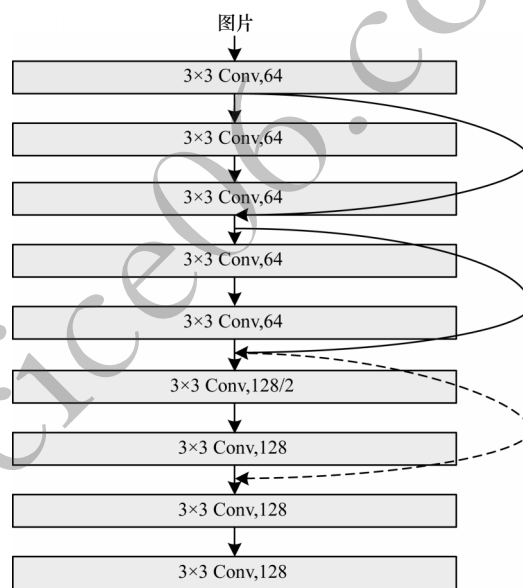


图2 特征提取部分结构

Fig.2 Structure of feature extraction part

在图 2 中,实线是在通道数量一致时使用 $Y=F(x)+x$ 的计算方式,虚线部分是通道数不一致,需要通过权重矩阵进行调整,采用 $Y=F(x)+Wx$ 的计算方式。此时的特征提取部分只需进行简单的非线性变换而无需过度参数化,在提升模型精度的同时还能抑制模型过大。

2.2 DPA-CapsGraph 路由

动态路由是在胶囊层间迭代进行特征提取的过程,当低层胶囊和高层胶囊的一致性分数较大时,高层胶囊就聚合此低层胶囊中的特征并更新自身状态,反之则抑制其一致性分数的增长。为使路由过程简单且有效,本文利用注意力分数代替迭代循环的耦合系数,不仅可以获得胶囊层全局上下文的信

息,还可以通过对高质量的胶囊赋予更大的权重来提取更多有效的特征信息。本文使用基于图注意力的特征处理方式,每个胶囊仅在特征选择期间捕获某一位置的特征,取消了模型对于数据源的依赖。胶囊层间的注意力特征提取部分如图3所示。

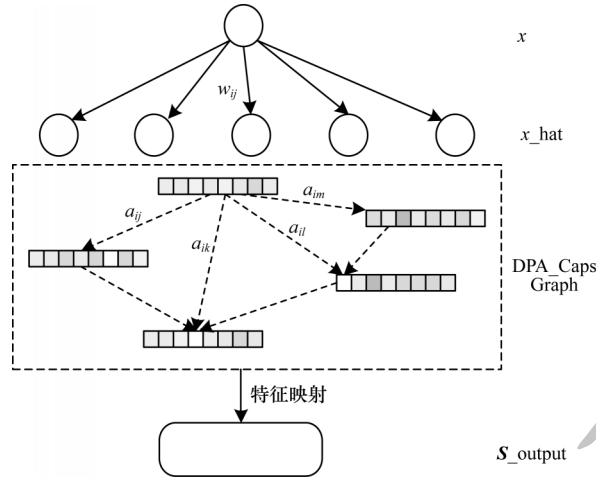


图3 DPA-CapsGraph路由过程

Fig.3 DPA-CapsGraph routing process

首先将模型在特征提取部分得到的信息进行通道维度的卷积得到初级胶囊,然后进行线性变换得到DPA-CapsGraph路由过程中的预测胶囊,即利用注意力机制的图卷积中的结点。

$$\hat{x}_{ji} = W_{ij} x_i \quad (1)$$

其中: $W_{ij} \in [8, 16]$ 是胶囊层之间做线性变化的权重矩阵,初级胶囊是8维的向量,通过线性映射得到16维的预测胶囊 \hat{x}_{ji} 。胶囊的维度越高,表示的图像信息也就越多,为胶囊间的特征选择提供更多的特征信息。

在图像分类任务中,不同的胶囊提取图像不同位置的特征,本文使用点积注意力^[21]以无监督的方式动态获取胶囊之间的相似度,为不同邻居的信息重要性赋予不同权值,在信息聚合时重点突出关键性较强的周围胶囊。

$$a_{att} = \hat{x}_{ji}^T \cdot \hat{x}_{kji} \quad (2)$$

其中: \hat{x}_{ji}^T 和 \hat{x}_{kji} 是同一层的预测胶囊,是矢量特征,矢量中的值代表图像中的不同特性,利用点乘方式的注意力机制更能表征图像不同位置之间相似程度的大小。可视化的注意力表述了初级胶囊层中周围胶囊对于中心胶囊的重要性,通过对初级胶囊的隐藏状态加权计算,完成特征权重的重新分配。

$$A = \begin{cases} 0, & a_{att} \leq 0 \\ a_{att}, & a_{att} > 0 \end{cases} \quad (3)$$

其中: A 为预测胶囊之间的相似度矩阵,将小于零的权重置为零,对相似度矩阵进行稀疏化,减少噪音信息的干扰,既可以提高分类的准确度,还能减轻模型对设备计算资源的压力,加快计算速度。

图结构在特征提取中具有关于节点度的倾向性,因为度越大的胶囊,节点的重要性越大,即节点

度越大,聚合的信息量就越大。

$$d_i = \sum_{j=1}^n a_{att_{ij}} \quad (4)$$

d_i 即注意力分数矩阵 A 的每一行元素之和,度矩阵 D 是 d_i 组成的 $n \times n$ 对角矩阵。

通过上述运算可以获得像素之间的依赖关系以及图像的结构信息,根据拉普拉斯特征映射的方式得到层式特征传播算子,将预测胶囊中包含的特征在空间上自动聚类,从而得到高级胶囊。

$$S_{output} = D^{-\frac{1}{2}} A D^{-\frac{1}{2}} \hat{X} \quad (5)$$

其中: \hat{X} 是预测胶囊的矩阵表示,即胶囊图中的结点集合; S_{output} 是经过空间映射后得到的高级胶囊。

为了提炼每个胶囊封装的通道特征信息,得到更加精确的全局特征表达,使用 Squashing() 函数得到最终的高级胶囊 v_j , 使高级胶囊的模长限制在 $[0, 1]$ 之间。在胶囊网络中,高级胶囊的长度用来表示某个类别存在的概率大小,胶囊长度越长,存在的可能性越大,反之则越小。

$$v_j = \frac{\|s_j\|}{1 + \|s_j\|^2} \cdot \frac{s_j}{\|s_j\|} \quad (6)$$

其中: 矢量 v_j 表示高级胶囊的输出,长度代表 j 类别存在的概率大小; 姿态表示物体的颜色等特征。

2.3 重构层及损失函数

取出输出 Digit-Caps 层中长度最大的高级胶囊,将其输入到三层全连接网络层进行重构,得到模型最后输出的重构图片。

本文的损失函数由编码损失和重构损失两部分构成,训练目标是损失加和最小化,公式如下:

$$L_{Loss} = L_c + L_{rec} \quad (7)$$

在模型的分类部分使用 margin loss 对每一个表征数字 c 的胶囊分别给出边缘损失函数。

$$L_k = T_c \max(0, m^+ - \|v_c\|)^2 + \lambda(1 - T_c) \max(0, \|v_c\| - m^-)^2 \quad (8)$$

其中: $m^+ = 0.9$; $m^- = 0.1$; $\lambda = 0.005$; T_c 表示 c 类别是否存在,存在取值为1,否则为0。

重构损失是原始图像和重构图像之间的差距,公式如下:

$$L_{rec} = \lambda \sum \text{MSEL}_{Loss}(\text{Pic}_i - \text{Pic}_r) \quad (9)$$

其中: Pic_i 表示原始图像; Pic_r 表示重构出来的图像,是预防模型过拟合的缩放系数。

2.4 DPA-CapsGraph 模型算法描述

DPA-CapsGraph 模型的算法学习过程如下:

算法 DPA-CapsGraph 学习算法

输入 通过卷积得到的初级胶囊 X , 训练批次 epoch

输出 表示图片类别的高级胶囊 Y

1. 初始化模型中的训练参数;

2. for 每一个训练批次:

3. 对于初级胶囊 L 层中的每一个低级胶囊进行线性变换,得到预测向量: $\hat{x}_{ji} = W_{ij} x_i$;

4. 在 L 层中构建全连通图,计算预测胶囊间的注意力分

数: $\text{att} = \hat{\mathbf{x}}_{ji}^T \cdot \hat{\mathbf{x}}_{kl}$;

5. 注意力分数矩阵稀疏化;

6. 根据经过稀疏化的注意力矩阵计算胶囊图中的度矩阵 \mathbf{D} ;

7. 预测胶囊进行特征映射得到 $L+1$ 层中的高级胶囊:

$$\mathbf{S}_{\text{output}} = \mathbf{D}^{-\frac{1}{2}} \mathbf{A} \mathbf{D}^{-\frac{1}{2}} \hat{\mathbf{X}};$$

8. $L+1$ 层中的高级胶囊进行挤压激活: $\mathbf{V}_j = \text{squashing}(\mathbf{S}_{\text{output}})$

9. Return \mathbf{V}_j

3 实验结果与分析

3.1 数据预处理

3.1.1 数据集

为了充分验证 DPA-CapsGraph 的分类效果,本文选取了 MNIST、FashionMNIST、CIFAR10 和 SVHN 4 种公开数据进行实验。FashionMNIST 和 CIFAR10 是各类衣物、动物、交通工具类的图像,其中包含更复杂的背景和物体结构信息,也更符合真实世界中的图像。MNIST 和 SVHN 都是数字数据集,而 SVHN 是来源于谷歌街景门牌号码的彩色图片,含有较多的噪音数字。本文选取的数据集类型和样本大小均是多样化的, MNIST 和 FashionMNIST 是尺寸大小为 28×28 像素的灰度图, CIFAR10 和 SVHN 是 32×32 像素的 RGB 图像。

3.1.2 实验环境及参数设置

本文模型是在 Ubuntu18.04 操作系统的平台上采用 Pytorch 框架进行开发,编程语言为 Python3.6,并使用 NVIDIA GeForce GTX 1080Ti 显卡来加速模型训练。

DPA-CapsGraph 由特征提取部分、初级胶囊层、高级胶囊层和重构层四部分构成,基于注意力机制的图卷积层间信息提取过程应用于初级胶囊层和高级胶囊层之间。在特征提取部分采用 ResNet18 的前 3 个 Block,去掉池化层共有 13 层 3×3 的卷积用来提取图像的特征,其通道数分别为 64、128 和 256。然后将特征提取部分的输出特征图经过通道卷积得到 32 个 8 维的低级向量胶囊。最后经过特征映射筛选出 10 个 16 维的高级胶囊。

3.1.3 评价指标

本文使用加权平均精确度 (Weighted Average Accuracy, WAA) 作为评价指标。在各类样本中,该指标可以更有效地评估模型的分类能力。

$$a_{\text{acc}_i} = \frac{T_{\text{TP}} + T_{\text{TN}}}{T_{\text{TP}} + F_{\text{FN}} + F_{\text{FP}} + T_{\text{TN}}} \quad (10)$$

$$W_{\text{WAP}} = \sum_{i=1}^n \frac{n_i}{N} a_{\text{acc}_i} \quad (11)$$

其中: a_{acc_i} 表示类别 i 的分类准确率; T_{TP} 是将真样本正确分类为真样本的数量; T_{TN} 是将假样本预测为假样本的数量; F_{FN} 是将真样本预测为假样本的数量; F_{FP} 是将假样本预测为真样本的数量; n_i 是类别 i 的样本总数; N 是样本总量。准确度越高,模型的分类效果越好。

3.2 实验结果

3.2.1 DPA-CapsGraph 消融实验

对 CapsNet 模型的改进包括两部分,为了准确评估各部分改进的效果,本文设置了如下消融实验对比分

析不同改进对模型分类效果的影响,如表 1 所示,其中粗体表示结果最优。消融实验共分为 3 个部分:

1) CapsNet+R。改善模型的特征提取部分来优化初级胶囊的质量,在 CapsNet 的基础上只将原始的特征提取部分更改为具有跳跃连接的残差网络。

2) CapsNet+A。优化动态路由,在 CapsNet 上仅更改动态路由部分。

3) DPA-CapsGraph。在 CapsNet 添加残差块和替换动态路由两部分。

表 1 DPA-CapsGraph 消融实验准确率

模型	MNIST	FashionMNIST	CIFAR10	SVHN
CapsNet	99.56	92.28	72.56	82.81
CapsNet+R	99.65	94.44	89.97	94.91
CapsNet+A	99.67	93.19	80.32	90.15
DPA-CapsGraph	99.74	95.02	91.78	95.65

由表 1 可知,增加特征提取部分和更换路由机制的优化方法均会提高模型的分类准确率,其中优化特征提取部分对模型的改进效果更为明显。对于 FashionMNIST 数据集来说,只改进路由机制对模型效果提升较小,但是在综合优化胶囊质量的优势之后,效果提升较大。因为 FashionMNIST 是结构较复杂的灰度图, CapsNet 中的卷积层提取不到足够的特征,所以优化胶囊质量表达更多的特征是必不可少的。对于 CIFAR10 和 SVHN 两种彩色数据集,动态路由机制的改进使模型准确率分别提升了 7.76% 和 7.34%,效果明显,消融实验证明了本文提出的路由机制利用注意力机制可以关注到特征图中距离较远的胶囊信息,得到关于图像全局的结构信息,从而使模型能够有效地选择特征进行图像分类。综合优化胶囊质量和更换路由机制两方面的优势,模型在 4 种数据集上分别提升了 0.18%、2.74%、19.22% 和 12.84%。此外, DPA-CapsGraph 在训练轮数上也有一定的表现结果,如图 4 所示。

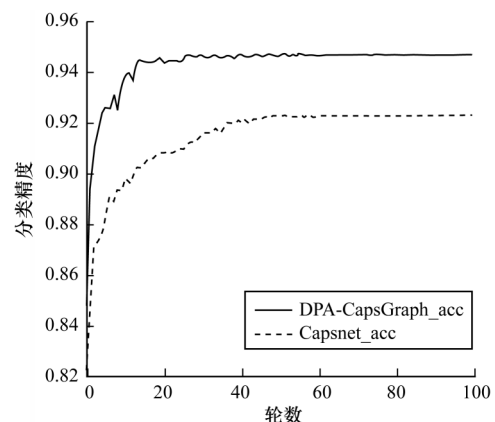


图 4 DPA-CapsGraph 在 FashionMNIST 中的分类精度

Fig.4 Classification accuracy of DPA-CapsGraph in FashionMNIST

由图 4 可知,在相同的实验环境下,DPA-CapsGraph 在训练初始阶段能够取得较好的分类精度,是因为本文模型可以提取更多的特征帮助模型进行分类。但是根据多次实验可知,DPA-CapsGraph 在每一轮训练时间上是 Capsnet 的两倍,但是在模型训练的轮数上,DPA-CapsGraph 只需更少的训练轮数就能收敛到更高的分类精度,所以总的训练时间是在可接受范围内。此外对于分类模型,适当增加训练时间来提升模型分类精度是可取的,由此证明本文提出的改进方法有一定

的实用价值。

3.2.2 DPA-CapsGraph 与其他模型的对比

为充分评估 DPA-CapsGraph 在图像分类任务中的效果,本文选取了 MS-CapsNet^[11]、TextCaps^[23]、AR CapsNet^[24]、FSc-CapsNet^[25]和 DA-CapsNet^[21]5 个对比模型。为确保对比模型的准确性,以上模型均是从胶囊的结构和动态路由两方面来改进向量胶囊网络,在实验过程中尽可能地使用了相同的参数设置,对比结果如表 2 所示。

表 2 DPA-CapsGraph 与不同模型分类准确率和参数规模对比

Table 2 Comparison of classification accuracy and parameter scale between DPA-CapsGraph and different models

模型	MNIST		FashionMNIST		CIFAR10	
	准确率/%	参数量/MB	准确率/%	参数量/MB	准确率/%	参数量/MB
CapsNet	99.56	6.9	92.28	6.9	72.56	9.5
MS-CapsNet	99.58	10.8	93.10	10.8	75.10	11.2
TextCaps	99.62	17.4	93.71	17.4	78.69	18.2
AR CapsNet	99.45	5.3	93.65	5.3	87.19	9.6
FSc-CapsNet	99.61	6.8	93.69	6.8	79.88	9.6
DA-CapsNet	100.00	7.0	93.98	7.0	85.47	9.1
DPA-CapsGraph	99.74	5.4	95.02	5.4	91.78	7.7

由表 2 可知,DPA-CapsGraph 在 MNIST、FashionMNIST 和 CIFAR10 3 种数据集上的准确率分别达到 99.74%、95.02% 和 91.78%,在除 MNIST 以外的另外两种数据集上,比表现最好的模型分别高出 1.04 和 4.59 个百分点。而且 DPA-CapsGraph 在达到最优分类准确率时,在 MNIST 和 FashionMNIST 上仅使用了 5.4 MB 的参数量,相比 CapsNet 减少了 51%,在 CIFAR10 数据集上也仅使用了 7.7 MB 的数据,参数量也减少了 36.4%。以上参数量的减少主要是因为模型在特征提取部分提取到既有实体的具体属性也有抽象的特征,保证了路由阶段所需的有效特征。所以可以通过减少初级胶囊的数量,使得模型的参数量有了一定程度的下降。综上所述,胶囊结构因包含更多的特征结构,过多的胶囊会增加模型的参数量,造成特征检测器的冗余,所以本文在提升胶囊质量的同时,在同层胶囊间用全局注意力构建图结构,筛选出图像中含有丰富特征的胶囊,从而能够在参数量小于对比模型的情况下,在 3 种数据集上均达到了更高的分类精度。DPA-CapsGraph 模型能够充分利用原始数据的信息,而且还捕捉到了同级胶囊之间的相互影响程度,由此可以证明空间依赖关系对于图像分类的准确率有一定的帮助。因此,本文的改进模型在图像分类领域是合理可行的。

4 结束语

胶囊网络可以较少的网络层和相对较少的参数量,达到标准卷积神经网络在图像分类上的效果。为进一步提高胶囊网络在图像分类领域的分类准确

率,本文通过点乘注意力构建胶囊图,获得同层胶囊之间的依赖关系。DPA-CapsGraph 不仅弥补了原始路由过程忽略同级特征的不足,而且还实现了提升模型的整体表现,在特征提取部分加入跳跃连接进行特征提取,提高了初级胶囊的特征表达能力;使用点乘注意力构建同层胶囊之间的关系图代替动态路由,提高了胶囊层间的特征选择能力。实验结果表明,DPA-CapsGraph 相比于传统胶囊网络模型,不但减少了模型的参数量,而且在 CIFAR10 等多种数据集上均取得了更高的分类准确率。但 DPA-CapsGraph 相比其他模型考虑了更多的图像信息,会存在部分冗余特征,模型在计算方面的时间相对会增加,下一步将在确保准确率的前提下精简特征,加快模型训练速度。

参考文献

[1] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.

[2] HE K M, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN [C]//Proceedings of 2017 IEEE International Conference on Computer Vision. Washington D. C. , USA: IEEE Press, 2017: 2980-2988.

[3] CONNEAU A, SCHWENK H, BARRAULT L, et al. Very deep convolutional networks for text classification[C]//Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers. Stroudsburg, USA: Association for Computational Linguistics, 2017: 213-224.

[4] SABOUR S, FROSST N, HINTON G E. Dynamic routing between capsules[EB/OL]. [2021-04-20]. <https://arxiv>.

- org/abs/1710.09829.
- [5] DE JESUS D R, CUEVAS J, RIVERA W, et al. Capsule networks for protein structure classification and prediction [EB/OL]. [2021-04-20]. <https://arxiv.org/abs/1808.07475>.
- [6] KIM Y, WANG P, ZHU Y F, et al. A capsule network for traffic speed prediction in complex road networks[C]// Proceedings of 2018 Sensor Data Fusion: Trends, Solutions, Applications. Bonn, Germany: Springer, 2018: 1-6.
- [7] AFSHAR P, MOHAMMADI A, PLATANIOTIS K N. Brain tumor type classification via capsule networks[C]// Proceedings of the 25th International Conference on Image Processing. Athens, Greece: [s. n.], 2018: 3129-3133.
- [8] JIMÉNEZ-SÁNCHEZ A, ALBARQOUNI S, MATEUS D. Capsule networks against medical imaging data challenges [C]// Proceedings of Intravascular Imaging and Computer Assisted Stenting and Large-Scale Annotation of Biomedical Data and Expert Label Synthesis. Berlin, Germany: Springer, 2018: 150-160.
- [9] MANDAL B, SARKHEL R, GHOSH S, et al. Two-phase dynamic routing for micro and macro-level equivariance in multi-column capsule networks[J]. Pattern Recognition, 2021, 109: 107595.
- [10] TSAI Y H H, SRIVASTAVA N, GOH H, et al. Capsules with inverted dot-product attention routing[EB/OL]. [2021-04-20]. <https://arxiv.org/abs/2002.04764>.
- [11] XIANG C Q, ZHANG L, TANG Y, et al. MS-CapsNet: a novel multi-scale capsule network [J]. IEEE Signal Processing Letters, 2018, 25(12): 1850-1854.
- [12] EDRAKI M, RAHNAVARD N, SHAH M. SubSpace capsule network[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 10745-10753.
- [13] LALONDE R, KHOSRAVAN N, BAGCI U. Deformable capsules for object detection[EB/OL]. [2021-04-20]. <https://arxiv.org/abs/2104.05031>.
- [14] SUN K, YUAN L M, XU H X, et al. Deep tensor capsule network[J]. IEEE Access, 2020, 8: 96920-96933.
- [15] BAHADORI M T. Spectral capsule networks[EB/OL]. [2021-04-20]. [https://gitee.com/saber5433/awesome-](https://gitee.com/saber5433/awesome-capsule-networks/blob/master/README.md)
capsule-networks/blob/master/README. md.
- [16] GU J D, TRESP V, HU H. Capsule network is not more robust than convolutional network[C]// Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE Press, 2021: 14304-14312.
- [17] WANG D L, LIU Q. An optimization view on dynamic routing between capsules[EB/OL]. [2021-04-20]. <https://openreview.net/pdf?id=HJjtFYJDf>.
- [18] ZHAO Z, KLEINHANS A, SANDHU G, et al. Capsule networks with max-min normalization[EB/OL]. [2021-04-20]. <https://arxiv.org/abs/1903.09662>.
- [19] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need?[EB/OL]. [2021-04-20]. <https://arxiv.org/abs/1706.03762>.
- [20] SALVETTI F, MAZZIA V, KHALIQ A, et al. Multi-image super resolution of remotely sensed images using residual attention deep neural networks[J]. Remote Sensing, 2020, 12(14): 2207.
- [21] HUANG W K, ZHOU F B. DA-CapsNet: dual attention mechanism capsule network[J]. Scientific Reports, 2020, 10(1): 11383.
- [22] 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.
- [23] JAYASUNDARA V, JAYASEKARA S, JAYASEKARA H, et al. TextCaps: handwritten character recognition with very small datasets [C]// Proceedings of 2019 IEEE Winter Conference on Applications of Computer Vision. Washington D. C., USA: IEEE Press, 2019: 254-262.
- [24] CHOI J, SEO H, IM S, et al. Attention routing between capsules[C]// Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Washington D. C., USA: IEEE Press, 2019: 1981-1989.
- [25] HAN T H, SUN R C, SHAO F J, et al. Feature and spatial relationship coding capsule network [J]. Journal of Electronic Imaging, 2020, 29: 1-10.