



DOI:10.13364/j.issn.1672-6510.20210267

多尺度自路由胶囊网络的构建方法

张贤坤, 陶健伟, 董梅, 袁菁
(天津科技大学人工智能学院, 天津 300457)

摘要: 胶囊网络 (capsule network, CapsNet) 是近些年新提出的一种网络模型. 相比于传统卷积神经网络, 胶囊网络能有效编码不同特征的空间位置关系, 但是其特征提取模块难以捕捉不同尺度的特征信息. 为了进一步提升 CapsNet 的性能, 提出了一种多尺度自路由胶囊网络 (multi-scale self-routing capsule network, MSSR-CapsNet). 首先利用多分支结构改进 CapsNet 胶囊的生成过程, 增加模型多尺度特征提取能力; 然后提出了一种胶囊注意力机制, 选择性筛选出重要的低级特征; 最后在 SVHN、CIFAR-10 和 CIFAR-100 数据集上对 MSSR-CapsNet 进行实验, 分类准确率分别为 97.46%、91.22% 和 63.84%. 此外, 实验结果也验证了改进方法的高效性, 有助于提升模型的性能.

关键词: 胶囊网络; 路由算法; 多尺度表示; 注意力机制

中图分类号: TP391 文献标志码: A 文章编号: 1672-6510(2022)03-0059-08

Construction Method of Multi-Scale Self-Routing Capsule Network

ZHANG Xiankun, TAO Jianwei, DONG Mei, YUAN Jing

(College of Artificial Intelligence, Tianjin University of Science & Technology, Tianjin 300457, China)

Abstract: Capsule network (CapsNet) is a new network model proposed in recent years. Compared with the traditional convolutional neural network structure, CapsNet can effectively encode the spatial relationship of different features, but its feature extraction module is difficult to capture feature information of different scale. In order to further improve the performance of CapsNet, a multi-scale self-routing capsule network (MSSR-CapsNet) is proposed in this article. First, the multi-branch structure is used to improve the capsule generation process of CapsNet and increase the multi-scale feature extraction ability of model; then a capsule attention mechanism is proposed to selectively filter out important low-level features. Finally, on SVHN, CIFAR-10 and CIFAR-100, the classification accuracy of MSSR-CapsNet is 97.46%, 91.22% and 63.84%, respectively. In addition, the ablation experiment also verified the efficiency of the experimental method and improved the performance of the model mode.

Key words: capsule network; routing algorithm; multiple scale representation; attention mechanism

卷积神经网络 (convolutional neural network, CNN) 在目标检测和图片分类领域中取得了巨大成功, 广泛应用于计算机视觉领域^[1-4]. CNN 可以通过最大池化和一系列卷积操作简化计算, 将低维的具体特征组合成高维的抽象特征, 逐步增强每层神经元的感受野, 提高模型的特征提取能力. 但是, CNN 的池化操作会损失局部区域内一些重要的属性信息 (如纹理和位置等), 因此无法高效地建立特征间的空间位置关系. 当改变特征的位置信息时, 网络的识别能力

就会变差, 无法从根本上解决特征间的空间位置关系.

为了解决上述问题, Hinton 等^[5]提出胶囊的概念. 相较于传统的 CNN 都是用标量神经元表示特征, 胶囊则采用向量形式表示图片中的实体特征, 因此也被称为胶囊向量, 用以对特征属性例如纹理、色彩、速度、位置等进行封装和编码. 胶囊向量涵盖众多属性信息, 且具有大小和方向, 这使胶囊网络能够拟合复杂的数据特征. 在路由过程中, 低级胶囊将其输出结果发送至对该输出结果表示认同的高级胶囊,

收稿日期: 2021-12-01; 修回日期: 2022-01-28

基金项目: 天津市自然科学基金项目 (19JCYBJC1530); 天津市普通本科高校教学改革和教学质量研究计划项目 (B201005706)

作者简介: 张贤坤 (1970—), 男, 安徽芜湖人, 教授, zhxkun@tust.edu.cn

并通过迭代机制更新自身状态,这使胶囊能更好地保存并更新特征的位置与属性信息. 胶囊网络建立了特征间的位置关系,这一特性使它在小型数据集上对于目标位置和角度的变化比具有相似结构的卷积神经网络更具有鲁棒性.

Sabour 等^[6]在 2017 年提出了一种基于胶囊向量和动态路由算法的胶囊网络(dynamic routing capsule network, DR-CapsNet). 该网络包含 2 层卷积层和 1 层全连接层,其中第 1 层卷积层使用卷积操作对输入图像进行特征提取,第 2 层卷积层(初始胶囊层)将特征图编码成固定维度的胶囊向量,而全连接层输出高级胶囊用以分类图片. 动态路由算法嵌入在初始胶囊层和全连接层之间,将低级胶囊由姿态矩阵所获取的输出送入对输出结果表示确定的高级胶囊,但路由算法有复杂的迭代机制,当输入空间维度较大时,模型会面临巨大的计算负担. 为了解决这个问题, Hinton 等^[7]提出了最大期望(expectation-maximization, EM)算法,用胶囊矩阵代替胶囊向量,且在路由阶段交替使用步骤 E 和步骤 M 更新高级胶囊的状态,使胶囊网络的泛化性和鲁棒性更强,但随之导致模型的复杂度高、参数量大. Choi 等^[8]用一种注意力机制取代动态路由算法和最大期望算法中路由系数的更新过程,在减少模型参数量和复杂度的同时,能够显著提升模型分类准确率和鲁棒性能.

上述工作中,大多数仅对胶囊层的结构或者路由机制进行优化,而本研究关注的是初始胶囊层的胶囊生成阶段,以获取多尺度的属性特征,从而提升胶囊对于图片特征的提取能力. 研究^[9-11]表明,多尺度特征提取作为一种即插即用的特征重构模块,已广泛应用于各类视觉任务和自然语言处理任务中. 多尺度特征提取模块可代替一般的卷积模块和单尺度特征模块,作为网络模型的一种基本架构^[3,12],为特征提取和重构提供了一种多尺度、多层次的选择. Szegedy 等^[13]通过实验表明,多尺度特征提取模块能够将提取到的不同尺度的特征信息进行融合,使模型在分类准确率上优于单尺度特征模块.

在自路由胶囊网络^[14](self-routing capsule network, SR-CapsNet)的基础上,本文提出了多尺度自路由胶囊网络(multi-scale self-routing capsule network, MSSR-CapsNet),充分利用多尺度特征模块强大的特征重构能力和自路由胶囊网络的自路由机制. 多尺度特征提取模块有助于捕获图片中不同尺度的特征变化,而自路由算法能以极少的计算成本和模型参数路由不同尺度的低级胶囊,两者结合实现特

征提取和重构,从而提升模型分类准确率和鲁棒性. 在 SVHH、CIFAR-10 和 CIFAR-100 数据集上进行实验,与其他胶囊网络相比, MSSR-CapsNet 在性能上有明显提升,达到了较高的分类准确率.

1 自路由胶囊网络

胶囊网络作为 CNN 的一种变体,克服了 CNN 不具备旋转不变性的缺陷. 当输入特征的空间位置发生变化时,胶囊网络能利用姿态矩阵编码低层特征和高层特征间的位置关系和其他属性关系,使模型对于仿射变化和对抗攻击更具有鲁棒性^[15]. 但是,胶囊网络也存在以下不足:首先该网络仅包含 2 层卷积层和 1 层全连接层,难以提取复杂的图像特征^[16];其次在路由阶段^[6-7]需执行复杂的迭代机制,其计算开销远大于卷积神经网络. 为了进一步提升胶囊网络的性能,通常采用以下改进策略:第一,提升初始胶囊层对特征的建模能力,又称优化胶囊法;第二,改进路由算法,减少网络参数,即优化路由法. 优化胶囊法的目的是提取对任务更有利的胶囊,通过改进胶囊网络,使其更适用于某项学习任务. 其作用主要为减少动态路由算法的计算负担,针对特定任务提升模型的性能. 优化路由法主要是为了能够减少模型的参数量,通过数学方法改进动态路由算法,可能会使局部模型的性能降低,但是影响相对较小.

SR-CapsNet^[14]改进了路由系数的生成策略,提出了一种有监督、非迭代的自路由(self-routing, SR)算法,与其他胶囊网络相比^[6-7],计算效率和分类准确率更高. 如图 1 所示,在 SR-CapsNet 中,每个 L 层胶囊无需利用迭代机制协调自身和 $L + 1$ 层胶囊间的连接概率,巧妙地引入了一个可训练的路由权重矩阵 W^{route} 与对应的胶囊 u_i 相乘作为 L 层胶囊 i 连接到 $L + 1$ 层胶囊 j 的概率.

$$c_{ij} = \text{softmax}(W_i^{route} u_i)_j \tag{1}$$

其中: u_i 为 L 层胶囊, softmax 为非线性激活函数.

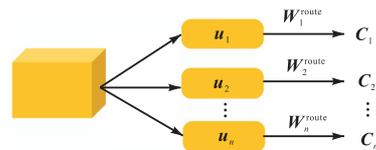


图 1 自路由算法

Fig. 1 Self-routing algorithm

对初始特征图 U_i 进行标量化,即采用 reshape 操作获取激活标量 a_i , a_i 反映了 L 层胶囊的激活概率

值;然后将连接概率和 L 层胶囊的激活标量 a_i 相乘用于产生 $L+1$ 层胶囊激活标量 a_j , 进而更新 $L+1$ 层胶囊.

$$a_j = \frac{\sum_{i \in \Omega_j} c_{ij} a_i}{\sum_{i \in \Omega_j} a_i} \quad (2)$$

其中: Ω_j 为 L 层胶囊数量, 最后进入 $L+1$ 层胶囊生成环节. 首先引入了一个可训练的姿态权重矩阵 W^{pose} 与 L 层胶囊 u_i 相乘作为 $L+1$ 层的预测胶囊 u_{ij} , 然后将预测胶囊 u_{ij} 与激活标量 a_j 作用, 进而更新 $L+1$ 层胶囊.

$$u_{ij} = W_{ij}^{\text{pose}} u_i \quad (3)$$

$$u_j = \frac{\sum_{i \in \Omega_j} c_{ij} a_i u_{ij}}{\sum_{i \in \Omega_j} a_i} \quad (4)$$

自路由胶囊网络去除了路由算法过程中复杂的迭代机制, 仅利用一个可学习路由权重矩阵编码低级特征和高级特征的空间和属性关系, 在提高胶囊网络性能的同时, 能显著降低模型的参数量和计算量, 但存在以下缺陷:

(1) 如图 1 所示, SR-CapsNet 采用的是单分支胶

囊结构, 对于图片中实体特征的提取能力不够, 难以捕获多尺度特征变化.

(2) 图 1 中仅通过 1 层自路由算法难以很好地编码低级特征和高级特征间的信息转换. 这是因为图像的特征往往复杂而又多级, 例如特征由低到高分别是线条、嘴巴、人脸, 且高级特征需要多方面融合不同位置和属性的低级特征以更新自身的状态.

2 自路由胶囊网络的改进与优化

多尺度自路由 (multi-scale self-routing, MSSR) 算法的整体结构如图 2 所示, 针对 SR 算法的不足进行了如下改进:

(1) 采用多分支胶囊结构, 在不同胶囊分支上生成不同维度的胶囊向量表征实体信息. 较之 SR-CapsNet 采用的单分支胶囊结构而言, 增强了模型的特征提取能力, 有利于路由算法生成多尺度路由信息.

(2) 在模型融合阶段, 采用胶囊注意力机制获取全局和局部的路由特征, 用以高效融合不同胶囊分支所产生的路由系数. 该胶囊注意力具有可以针对每个胶囊分支的路由系数自适应分配权重的优势, 能够产生更精确的连接概率.

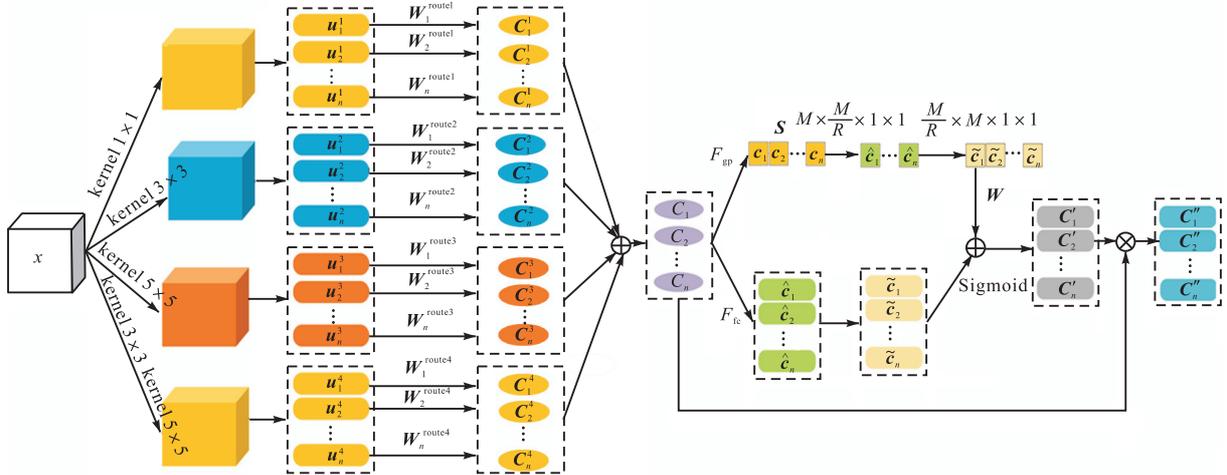


图 2 多尺度自路由算法的整体结构

Fig. 2 Overview structure of MSSR algorithm

2.1 多尺度自路由胶囊网络

文献[14]提出的架构在 SVHN 和 CIFAR-10 数据集上的表现结果尚可, 但在 CIFAR-100 这种包含复杂特征和对象的数据集上获取的分类准确率则低于预期. 这是因为 SVHN 和 CIFAR-10 数据集本身就是轻量级数据集, 包含的类别数和特征并不复杂, 模型能够很容易地提取出边缘和纹理特征用于图片分类,

而 CIFAR-100 数据集涵盖的特征较为复杂, 对模型提出了进一步要求, 需要对特征进行高层次理解和多尺度建模. 因此, 本课题组提出一种新型的多尺度胶囊网络结构, 如图 3 所示, 从左到右分别为特征映射层、多分支胶囊层、高级胶囊层.

MSSR-CapsNet 将初始胶囊层改为多分支胶囊层, 在每个胶囊分支上生成具有不同胶囊维度的胶囊

向量. 经过特征映射层和多尺度胶囊层对原始图片进行特征提取和胶囊生成, 并通过挤压函数(squash function)对每个向量神经元进行方向不变的尺度压缩, 使向量模长被压缩到 0~1, 用于表示该向量神经元被激活的概率. 在每个胶囊分支上, 通过自路由算法, 由低级胶囊通过路由权重矩阵 W^{route} 对高级胶囊进行位置预测, 最后通过胶囊注意力机制高效融合不同分支的预测结果, 选择性地激活高级胶囊.

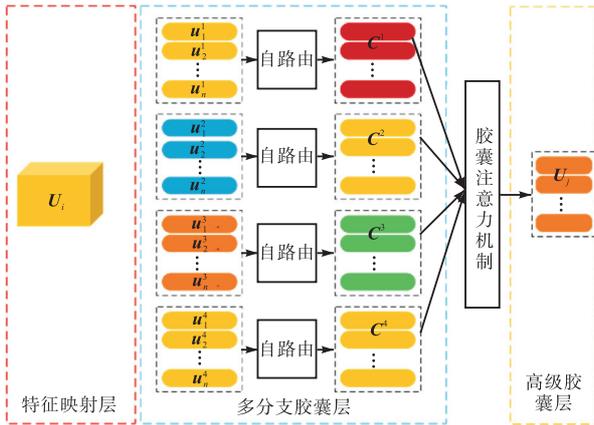


图3 多尺度自路由胶囊网络的结构
Fig.3 Structure of MSSR-CapsNet

以 CIFAR-10 数据集为例, 模型的前馈传递过程如下: 图片的大小为 32 像素 × 32 像素, 首先经过特征映射层的线性和非线性变换提取出 512 张 4 像素 × 4 像素大小的初始特征图 U . 然后多分支胶囊层对这些初始特征图进行向量化处理, 即每个胶囊分支 (在本文中默认的胶囊分支数为 4) 采用不同大小的 512 组卷积核. 前 3 个胶囊分支每组卷积核中分别包含 1×1 、 3×3 和 5×5 大小的卷积核, 对特征图以步长为 1 和填充为 0、1、2 进行深度卷积. 最后一个分支上每组中包含 3×3 和 5×5 大小的卷积核代替 7×7 大小的卷积核, 用来对特征图进行深度卷积, 之后所有胶囊分支都采用 1×1 大小的卷积核在特征图通道层面上进行点卷积. 最后采用 ReLU 激活函数, 得到不同尺度的底层特征, 即为 1×16 、 1×32 、 1×64 和 1×128 大小的胶囊, 具体结构如图 4 所示.

来自特征映射层的初始特征图 U 与图片中各实体之间的姿态矩阵 W_j^{pose} 相乘以预测高级特征 \hat{u}_{ij} , 即 $\hat{u}_{ij} = u_j W_j^{pose}$. 在多分支胶囊结构上, 针对每个胶囊分支都采用 1 次自路由机制实现低级特征和高级特征间信息交流. 得到 4 组不同尺度的路由系数 C_{ij}^t ($t=1,2,3,4$), 每组的大小均为 512×10 , 其中: 512 代表低级胶囊数量, 用 M 表示; 10 代表高级胶囊的

数量, 用 N 表示, 同时也是分类类别数. 通过一种新颖的胶囊注意力机制融合 4 组不同尺度的路由系数, 得到更精确的路由系数 C'' , 每个值反映了低级胶囊对高级胶囊的预测结果. 将该预测结果与对应的高级特征 \hat{u}_{ij} 相乘用于更新高级胶囊的状态.

$$u_j = \frac{\sum_{i \in \Omega_j} C_{ij}'' \hat{u}_{ij}}{\sum_{i \in \Omega_j} a_i} \quad (5)$$

其中: a_i 为 L 层胶囊的激活标量, u_j 为 $L+1$ 层胶囊.

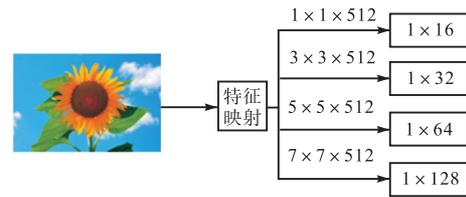


图4 多分支胶囊示意图

Fig.4 Schematic diagram of multiply branches capsule

2.2 胶囊注意力机制

本文提出一种新颖的胶囊注意力机制, 用以控制路由系数间的信息流动. 动态调整每个胶囊的路由系数值, 为每个路由系数自适应分配权重, 用以突出重要的低级特征, 而忽略无关或影响程度较小的低级特征.

首先采取直接相加的方式融合 4 组不同尺度的路由系数.

$$C = \sum_{t=1}^4 C_{ij}^t \quad (6)$$

其中: C 为融合后总的路由系数.

接下来进入胶囊注意力环节, 沿着高级胶囊维度 N 对 C 进行全局平均池化操作, 收集后计算全局路由信息统计值 (S). 具体来说, S 中第 m 个元素值的计算是通过在高级胶囊维度 N 上对 C 进行压缩操作.

$$S_m = F_{sp}(C_m) = \frac{1}{N} \sum_{j=1}^N C_m(j) \quad (7)$$

其中: S_m 为路由信息统计值.

为了实现路由信息间的交互, 引入了带有两个点卷积操作的门机制^[17], 其中所对应的卷积核尺寸分别为 $M \times \frac{M}{R} \times 1 \times 1$ 和 $\frac{M}{R} \times M \times 1 \times 1$, 其中 R 是超参数, 代表了低级胶囊数量的缩减率 (本文中默认为 32), 最终的全局路由信息统计值计算式为

$$W(S) = \beta(PWConv_2(\partial(\beta(PWConv_1(S)))))) \quad (8)$$

其中: ∂ 表示 ReLU^[18] 激活函数, β 表示批量标准化 (BN)^[19]. 同时也引入了一个统计量 Z 作为局部路由

信息统计值,对低级胶囊和高级胶囊的连接概率进行微调,产生更精确的路由系数.该操作涉及到两个全连接层 $F_1 \in \mathbb{R}^{(N \times D) \times N}$ 和 $F_2 \in \mathbb{R}^{N \times (N \times D)}$,其中两个全连接层之间的缩放维度 D 是超参数,用于提高网络的拟合性能(本文中 D 默认为 5).具体操作为

$$\mathbf{Z}(\mathbf{C}) = F_{fc}(\mathbf{C}) = F_1(\partial(F_2(\mathbf{C}))) \quad (9)$$

其中 ∂ 表示 ReLU 激活函数.在获得全局路由信息统计值 \mathbf{W} 和局部路由信息统计值 \mathbf{Z} 的基础上,胶囊注意力机制的最终输出为

$$\mathbf{C}'' = \mathbf{C} \otimes M(\mathbf{C}) = \mathbf{C} \otimes \partial(\mathbf{W}(\mathbf{S}) \oplus \mathbf{Z}(\mathbf{C})) \quad (10)$$

其中: $M(\mathbf{C})$ 表示胶囊注意力机制产生的胶囊注意力权重, \oplus 表示逐元素相加(element-wise summation), \otimes 表示逐元素相乘(element-wise product).

3 实验评估

3.1 实验设置

3.1.1 数据集

在实验中采用 SVHN 数据集、CIFAR-10 数据集和 CIFAR-100 数据集(图 5).



(a) SVHN 数据集 (b) CIFAR-10 数据集 (c) CIFAR-100 数据集
图 5 数据集展示

Fig. 5 Overview of datasets

SVHN 数据集中的图片为街景门牌号数字 0~9,其风格与 MINST 数据集相似.图片的裁剪尺度都很小,却包含了更复杂的背景特征,每张图片的大小为 32 像素 \times 32 像素,共有 99 289 张图片,其中 73 257 张图片为训练集,26 032 张图片为测试集. CIFAR-10 数据集是色彩丰富的自然图片,由 10 个不同类别的物体构成,总共有 60 000 张图片,其中 50 000 张图片为训练集,而剩余的 10 000 张为测试集.训练集中每个类别涵盖的图片数量相同,均为 6 000 张,图片大小均为 32 像素 \times 32 像素,该数据集包含的特征较为复杂,分类难度更大. CIFAR-100 数据集较之 CIFAR-10 数据集而言,包含的类别数更多,由 100 个不同种类的物体构成,总共有 60 000 张三通道的彩色图片,其中 50 000 张图片为训练集,10 000 张图片为测试集,且每个类别的图片都被划分

为 500 张训练图片和 100 张测试图片,图片大小均为 32 像素 \times 32 像素. CIFAR-100 数据集中的 100 个类分别分为 20 个超类,每个图片都带有一个自身标签和一个超类标签.该数据集包含的特征更为复杂,分类难度更大.

3.1.2 数据增强

SR-CapsNet 已经在 SVHN 数据集上取得了优异的性能表现,但是在 CIFAR-10 和 CIFAR-100 这些较为复杂的数据集下,并未取得较好的分类准确率.为了提高模型分类准确率和泛化能力,本文采取了一些数据增强方法,比如随机裁剪、随机水平翻转和归一化.

3.1.3 评价指标

为了精确衡量 MSSR-CapsNet 的分类准确率,采用准确率作为评价指标.

$$\text{准确率} = [\text{TP}/(\text{TP} + \text{FP})] \times 100\% \quad (11)$$

其中: TP 代表分类正确的样本数, FP 代表分类错误的样本数.

3.1.4 训练部署及训练策略

实验环境是基于 Ubuntu 16.04 系统,所使用的硬件环境: NVIDIA GeForce RTX3080 10 GB 显卡, 1 TB 内存(RAM), Intel Xeon E5-2650 v3 处理器.使用 PyTorch 深度学习框架对 MSSR-CapsNet 进行构建和测试.在训练期间,将 batch size 设置为 64,优化器采用 Adam,使用余弦退火算法在训练期间不断调整学习率,初始学习率为 0.1,截止学习率为 0.001,每训练 100 个 epoch 将学习率降低为原来的 0.1,训练 350 个 epoch.

3.2 实验结果

3.2.1 数据集测试结果

从分类结果和准确率这两个方面衡量模型的性能表现,统计 MSSR-CapsNet 在不同数据集上的分类结果和准确率.不同胶囊网络的分类结果对比见表 1.由表 1 可知: MSSR-CapsNet 在 SVHN 数据集和 CIFAR-10 数据集上已经取得了优异的性能表现,测试准确率均高于 90%;在最为复杂的 CIFAR-100 数据集上的测试准确率也达到了 63.84%,在同层次下已达到较高水准.

表 1 不同胶囊网络的分类结果对比

Tab. 1 Comparison of classification results of different CapsNets

数据集	TP	TF	准确率/%
SVHN	25 370	662	97.46
CIFAR-10	9 122	878	91.22
CIFAR-100	6 384	3 616	63.84

实验证明 MSSR-CapsNet 能捕获图片中不同尺度的特征信息,并且可以选择性地筛选出重要的低级特征,即使针对最复杂的 CIFAR-100 数据集,模型也能进行有效的特征提取和噪声过滤.因此, MSSR-CapsNet 可以在 SVHN、CIFAR-10 和 CIFAR-100 数据集上得到很好的分类准确率.

3.2.2 消融实验

为了验证 MSSR-CapsNet 模型的优势,对多分支胶囊结构、胶囊注意力机制和两者相融的情况进行相关实验.将不同结构的多分支胶囊网络嵌入胶囊网络的胶囊生成阶段进行实验,结果见表 2.以多分支

胶囊结构为基本单元构建多分支胶囊网络,当胶囊分支数(branches)为 0~4 时,通过增加胶囊分支的数量可以使模型有更好的特征提取能力,增强对数据的表征;当胶囊分支数增加到 5 时,随着模型参数的增加开始出现过拟合现象,导致分类准确率下降.考虑到胶囊注意力机制会选择性筛选出重要的路由信息而忽略无关的路由信息,降低模型参数冗余的同时提升分类准确率,所以在模型中加入了注意力操作(Attention).因此,最终确定多分支自路由胶囊网络结构为 4 分支胶囊数加 1 次胶囊注意力操作.

表 2 不同分支结构的胶囊网络对比结果

Tab. 2 Comparison results of capsule network with different branches

网络结构	准确率/%		
	SVHN 数据集	CIFAR-10 数据集	CIFAR-100 数据集
SR-CapNet	96.40	87.60	60.20
1 branches + SR-CapNet + 0 Attention	96.83	89.18	62.30
2 branches + SR-CapNet + 0 Attention	96.93	90.12	62.58
3 branches + SR-CapNet + 0 Attention	97.04	90.45	62.76
4 branches + SR-CapNet + 0 Attention	97.29	90.86	63.00
5 branches + SR-CapNet + 0 Attention	97.13	90.45	62.61
6 branches + SR-CapNet + 0 Attention	97.08	90.32	62.43
7 branches + SR-CapNet + 0 Attention	97.01	90.13	62.16
4 branches + SR-CapNet + 1 Attention	97.46	91.22	63.84

为了验证多分支胶囊网络和胶囊注意力的有效性,进行了消融实验(表 3).

表 3 网络结构消融实验结果

Tab. 3 Experiment results of network structure ablation

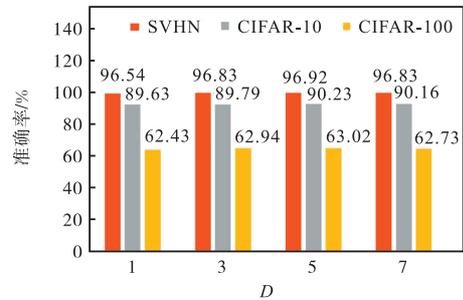
网络结构	准确率/%		
	SVHN 数据集	CIFAR-10 数据集	CIFAR-100 数据集
SR-CapsNet	96.40	87.60	60.20
Attention + SR-CapsNet	97.03	90.03	62.53
MSSR-CapsNet	97.13	90.23	62.43
Attention + MSSR-CapsNet	97.46	91.22	63.84

由表 3 可知:将胶囊注意力机制应用于路由系数的融合阶段,可以提供精确的位置预测,提高模型的分类准确率;对胶囊进行多尺度表示能有效捕获不同尺度的实体特征,并提升模型的性能.

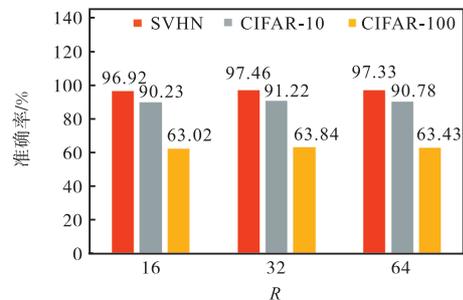
3.2.3 超参数选取实验

考虑到超参数 D 为全连接层之间的缩放维度,若选取过大会使模型参数剧增,增加计算负担,因此以奇数序列的方式依次递增 D 值.超参数 R 为低级胶囊缩放因子,模型参数数量与该值成反比,且胶囊维度的初始值为 16,因此本文选取 16 为初始值可以兼容维度变化,然后以 2 的倍率依次增加 R 值,以此探

索最佳 R 和 D 的取值,消融实验结果如图 6 所示.



(a) $R = 16$ 时不同超参数 D 的胶囊网络性能对比



(b) $D = 5$ 时不同超参数 R 的胶囊网络性能对比

图 6 消融实验对比

Fig. 6 Comparison of ablation experiments

当 D 增加到 5 时,随着模型参数的增加,各胶囊

分支间的信息交互变得冗余,导致准确率下降.当 R 增加到 64 时,随着模型参数的减少导致胶囊注意力机制难以编码重要的特征信息,导致准确率下降.因此,超参数 D 和 R 分别取 5 和 32,模型将取得最佳分类准确率.

3.2.4 训练结果对比

胶囊网络中的经典方法分别为 DR-CapsNet^[6]和

EM-CapsNet^[7],因此本文方法先与其进行对比,然后与目前最新的胶囊网络 SR-CapsNet^[14]、Deeper-CapsNet(deeper capsule network)^[20]和 Efficient-CapsNet(efficient capsule network)^[15]在不同数据集上进行对比,不同路由方法的分类准确率、参数量和每秒所执行的浮点运算次数(floating-point operations per second, FLOPs)见表 4.

表 4 不同路由方法的分类准确率和参数量

Tab. 4 Classification accuracy and parameter amount of different routing methods

方法	SVHN			CIAFR-10			CIFAR-100		
	准确率/%	参数量/MB	FLOPs	准确率/%	参数量/MB	FLOPs	准确率/%	参数量/MB	FLOPs
DR-CapsNet	96.40	5.8	73.5	57.84	5.8	73.5	30.29	31.64	174.5
EM-CapsNet	95.81	0.9	76.6	34.15	0.9	76.6	43.51	4.53	138.4
SR-CapsNet	96.40	0.9	62.2	87.60	0.9	62.2	60.20	4.78	126.7
Deeper-CapsNet	97.08	2.3	81.4	90.04	2.3	81.4	62.13	15.23	159.6
Efficient-CapsNet	97.22	1.7	65.7	90.31	1.7	65.7	62.34	5.78	129.3
MSSR-CapsNet	97.46	1.5	68.3	91.22	1.5	68.3	63.84	5.43	134.8

MSSR-CapsNet 在 SVHN 数据集上的分类准确率与上述其他方法几乎持平,但是在较为复杂的 CIFAR-10 和 CIFAR-100 数据集上优于其他方法,尤其是在 CIFAR-100 数据集上,比传统的 SR-CapsNet 提高了 3.64%.在模型参数和计算量上, MSSR-CapsNet 引入了多分支胶囊结构和注意力机制,导致模型参数量和计算量有了轻微提升,但与准确率提升的幅度相比,参数量可以忽略不计.这说明引入多分支胶囊结构和胶囊注意力机制所形成的多尺度特征融合能切实提升胶囊网络应对复杂数据集的能力.

4 结 语

在 SR-CapsNet 的基础上,提出一种新的多分支胶囊网络 MSSR-CapsNet.该网络融合了对全局和局部路由信息提取能力更强的胶囊注意力机制,在融合不同尺度路由信息时,能够选择性地突出重要的特征信息而忽略无关的特征信息.在多分支胶囊网络中,每个分支的胶囊维数都是可变的,能高效建模不同尺度的特征变化,提高模型的特征提取能力.为了验证 MSSR-CapsNet 的性能表现,分别在 SVHN 数据集、CIFAR-10 数据集和 CIFAR-100 数据集上进行实验,与传统的胶囊网络相比,模型的准确率分别达到了 97.46%、91.22% 和 63.84%,显著提高了模型在 CIFAR-100 这种复杂数据集上分类能力.但是,目前模型采用的胶囊注意力机制在融合上还不是十分完美,会有细节上的误差,分类精度还有提升空间,在

未来的工作中将会对现有的网络进行进一步改进,争取更好的性能提升.

参考文献:

- [1] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//IEEE. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE, 2017: 2117–2125.
- [2] 张媛涵,刘翔,石蕴玉,等.基于深度学习的双流程短视频分类方法[J/OL].计算机工程:1–8[2021-11-12]. <https://doi.org/10.19678/j.issn.1000-3428.0061913>.
- [3] LI X, WANG W, HU X, et al. Selective kernel networks[C]//IEEE. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE, 2019: 510–519.
- [4] JIA F, LEI Y, GUO L, et al. A neural network constructed by deep learning technique and its application to intelligent fault diagnosis of machines[J]. Neurocomputing, 2018, 272: 619–628.
- [5] HINTON G E, KRIZHEVSKY A, WANG S D. Transforming auto-encoders[C]//International Conference on Artificial Neural Networks. Berlin: Springer, 2011: 44–51.
- [6] SABOUR S, FROSST N, HINTON G E. Dynamic routing between capsules[C]//Neural Information Processing Systems. California: NIPS Proceeding, 2017: 3856–3866.
- [7] HINTON G E, SABOUR S, FROSST N. Matrix capsules with EM routing[EB/OL]. [2021-06-13]. <https://open>

- review.net/pdf?id=HJWLFGWRb.
- [8] CHOI J, SEO H, IM S, et al. Attention routing between capsules[C]//IEEE. 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW). New York: IEEE, 2019: 848–856.
- [9] 彭豪, 李晓明. 多尺度选择金字塔网络的小样本目标检测算法[J/OL]. 计算机科学与探索: 1–13[2021–11–12]. <http://kns.cnki.net/kcms/detail/11.5602.tp.20211102.1150.002.html>.
- [10] SRIVASTAVA R K, GREFF K, SCHMIDHUBER J. Highway networks[EB/OL].[2015–11–03]. <https://arxiv.org/abs/1505.00387>.
- [11] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//IEEE. IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2016: 770–778.
- [12] 张学鹏, 王远军. 采用密集多尺度卷积网络的胰腺分割方法[J/OL]. 小型微型计算机系统: 1–7[2021–11–12]. <http://kns.cnki.net/kcms/detail/21.1106.TP.20211109.1016.002.html>.
- [13] SZEGEDY C, VANHOUCKE V, IOFFE S, et al. Rethinking the inception architecture for computer vision[C]//IEEE. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE, 2016: 2818–2826.
- [14] HAHN T, PYEON M, KIM G. Self-routing capsule networks[J]. Advances in neural information processing systems, 2019, 32: 7658–7667.
- [15] MAZZIA V, SALVETTI F, CHIABERGE M. Efficient-CapsNet: capsule network with self-attention routing[EB/OL]. [2021–11–16]. <https://arxiv.org/abs/2101.12491>.
- [16] 刘林嵩, 全明磊, 吴东亮. SA-CapsNet: 自注意力胶囊网络[J]. 计算机应用研究, 2021, 38(10): 3005–3008.
- [17] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[J]. IEEE Transactions on pattern analysis and machine intelligence, 2020, 42(8): 2011–2023.
- [18] NAIR V, HINTON G E. Rectified linear units improve restricted Boltzmann machines[EB/OL]. [2021–06–13]. <https://dl.acm.org/doi/10.5555/3104322.3104425>.
- [19] IOFFE S, SZEGEDY C. Batch normalization: accelerating deep network training by reducing internal covariate shift[EB/OL]. [2021–06–13]. <https://dl.acm.org/doi/10.5555/3045118.3045167>.
- [20] RAJASEGARAN J, JAYASUNDARA V, JAYASEKARA S, et al. DeepCaps: going deeper with capsule networks[C]//IEEE. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE, 2019: 10725–10733.

责任编辑: 郎婧

(上接第 58 页)

- 哈尔滨: 黑龙江大学, 2016: 43.
- [7] 同济大学数学系. 高等数学: 上册[M]. 7 版. 北京: 高等教育出版社, 2014: 168–174.
- [8] 焦飞, 陈力航, 戴孟莲, 等. 基于圆柱凸轮机构的“8 字形”无碳小车设计[J]. 机械传动, 2018, 42(4): 71–75.
- [9] 高雨辰, 汪海溟. UG NX 10.0 三维数字化辅助产品设计[M]. 北京: 清华大学出版社, 2018: 73–78.

责任编辑: 周建军