

基于剪枝优化的深层胶囊网络

郑香平 梁循

(中国人民大学信息学院 北京 100872)

摘 要 卷积神经网络 (Convolutional Neural Network, CNN) 在计算机视觉任务中发展迅速, 尤其在图片分类、检索、目标检测等各个领域中发挥出优越的性能。CNN 能够直接从原始像素中经过少量预处理, 得到其有效的特征表示, 识别计算机视觉上的规律。然 CNN 采用了池化策略, 对物体的局部特征理解深刻, 而忽视了物体整体的空间结构。而胶囊网络重视物体的空间结构, 是一种新的神经网络架构。胶囊网络通过将特征叠加到向量中来表示物体的多种信息, 具备良好的拟合特征能力。使得网络架构不仅能识别出物体的局部特征, 而且能保留其整体空间特征的层次关系。然而胶囊网络在复杂数据中表现不佳。为了提高胶囊网络在图像数据的识别准确率, 本文提出一种基于剪枝优化的深层胶囊网络, 其有效地处理胶囊网络层之间的连接关系, 不仅减少了低层冗余胶囊, 还提高网络的泛化能力。新胶囊网络的剪枝策略, 对低层胶囊进行严格筛选, 确保只有重要的、活跃度大的低层胶囊重组新的胶囊层, 进行加权连接预测高层胶囊。对于图像数据, 获取的特征不足使其难以在胶囊中进一步处理。因此, 本文设计了一种新颖的胶囊块卷积层, 使用 3D 卷积核将局部的胶囊子集转化为一个更加抽象的胶囊, 获取更加全面数据的特征表示来构建更加有水平的高层网络。同时, 该层中低水平胶囊到高水平胶囊之间的跳跃连接可以减少深度模型训练过程中面临的梯度消失问题。此外, 本文在多个数据集上进行性能测试, 都取得了良好的实验效果。

关键词 深度网络; 剪枝; 胶囊网络; 卷积神经网络
中图法分类号 TP18

Deep Capsule Network Based on Pruning Optimization

ZHENG Xiang-Ping LIANG Xun

(School of Information, Renmin University of China, Beijing, 100872)

Abstract Convolutional neural network (CNN) has developed rapidly in computer vision tasks, especially in image classification, retrieval, target detection and other fields. CNN can directly preprocess the original pixels, get its effective feature representation and recognize the rules of computer vision. However, CNN adopts pool strategy, which has a deep understanding of the local characteristics of the object, but ignores the overall spatial structure of the object. Capsule network is a new neural network architecture, which attaches importance to the spatial structure of objects. Capsule network can represent various information of objects by superimposing features into vectors, and has good ability of fitting features. The network architecture can not only recognize the local features of objects, but also retain the hierarchical relationship of the overall spatial features. However, capsule network is not good at complex data. In order to improve the accuracy of image data recognition in capsule network, this paper proposes a deep capsule network based on pruning optimization, which effectively deals with the connection between the capsule network layers, not only reduces the redundant capsule layer, but also improves the generalization ability of the network. The pruning strategy of the new capsule network is to

select the low-level capsule strictly, so that only the important and active low-level capsules are reformed into the new capsule layer, and the high-rise capsules are predicted by weighted connection. For image data, the acquired features are insufficient, which makes it difficult to further process in capsule. Therefore, a novel capsule block layer is designed in this paper. The local capsule subset is transformed into a more abstract capsule by 3D convolution core, and the feature representation of more comprehensive data is obtained to construct a more level high-level network. At the same time, the jump connection between the low-level capsule and the high-level capsule can reduce the gradient disappearance problem in the training of the deep model. In addition, the performance test on multiple data sets has achieved good experimental results.

Key words Deep network; pruning; capsule network; CNN

1 引 言

近年来, 卷积神经网络在许多计算机视觉任务中获得突破性的成绩^[1,2,3,4]。它能够保留图片的基本特征, 类似人类的视觉原理, 在图片分类、检索、目标定位检测等各个领域中都表现出惊人的准确性。但是其本质来说, 只是对物体的局部特征理解的过于深刻, 而忽视了物体整体的空间结构, 归其由采用了池化策略的影响。虽说池化策略能够有效的提升网络分类的鲁棒性, 但也正是这种因素, 破坏了图片原有细微的结构, 如位置、方向、大小等。CNN 抛弃这些属性特征来最大地增加其特征检测能力, 但是在特征理解上却大打折扣。

为了打破这种固有限制, 使得网络架构不仅能识别出物体的局部特征, 而且能保留其整体空间特征的层次关系。2017 年 Sabour 等人提出了胶囊网络模型^[5], 该网络只包含一个卷积层和一个完全连接的胶囊层, 是一种全新的网络结构。我们知道, CNN 侧重的是重点记忆, 只提取数据的关键特征, 掌握特定的记忆点, 淡化物体的整体关联。这样结果往往会认识事物本身不够全面, 把相似的物体理解成同一物体, 或者是同一物体换种方式表达也不明其原理。而胶囊网络侧重的是理解记忆, 他不仅可以记忆物体的局部特征, 而且可以掌握局部特征与整体特征之间的映射关联, 是一种全新的神经网络模型框架。

胶囊网络的核心是动态路由算法。在胶囊网络中, 每一层中的每一个节点表示的是一个神经元胶囊, 属于低层的胶囊通过连接边与高层胶囊产生联系。其中连接的权值在学习过程中会发生变化的, 而引起节点之间连接强度的变化, 就叫做动态路

由。胶囊网络是把低层的胶囊与高层的胶囊全部连接, 达到全部学习的效果。但是不可避免的包含许多不必要的低层胶囊的信息, 同时加重网络学习的强度。例如, 识别一张图片人的动作分类, 通过卷积操作, 可以获得背景特征, 人的动作特征。这些特征全部组成一个个低层的胶囊, 而关于背景的这些胶囊是冗余的、无用的, 我们希望这些低层胶囊没有被激活, 真正需要的是动作特征组成的低层胶囊。过多不重要的低层胶囊与高层胶囊相连接, 高层胶囊如果没有及时地理解低层胶囊所传播的信息, 会造成网络的混乱, 引起网络的奔溃, 导致网络学习不佳。最终的结果是网络无法收敛。同时, 网络的深度是获取数据高级特征的一个重要渠道。但是相邻胶囊层完全连接的情况下, 会导致胶囊中间层学习不良。这是因为许多胶囊相连接时候, 他们的系数往往太小, 从而限制着梯度的流动, 抑制着网络学习的方向。

为了克服胶囊网络中低层胶囊过于冗余, 胶囊层中学习不良而无法获取充足特征等挑战, 本文提出了一种新的基于剪枝优化的深层胶囊网络 (Pruning Capsule Network, PCN), 该网络能够有效地选择最重要, 活跃度最大的低层胶囊来进行加权预测高层胶囊, 而把那些活跃度低的低层胶囊剪枝掉。不仅减少了低层冗余胶囊数目, 还提高了网络的泛化能力。对于图像数据, 本文设计了胶囊块卷积层, 使用 3D 卷积核将局部的胶囊子集转化为一个更加抽象的胶囊。同时, 低水平胶囊到高水平胶囊之间进行跳跃连接。经过胶囊卷积块层, 能够有效的获取复杂数据的高级特征, 而不会造成胶囊中间层因为过多胶囊使得网络学习受到抑制。本文提出的剪枝胶囊网络利用了两个关键思想: 剪枝优化和胶囊卷积块。剪枝优化能够有效地激活那些重

要的低层胶囊，同时剪枝冗余的低层胶囊，使网络更加简化；而胶囊卷积块是使网络能够获取更多的高维数据特征，同时不会抑制梯度的流动。具体做了如下贡献：

(1) 提出了一种新的基于剪枝优化的深层胶囊网络，其目的是提高胶囊网络在复杂图像数据的识别准确率。并且有效的处理胶囊网络层之间的连接关系。不仅减少了低层冗余胶囊，还提高网络的泛化能力。

(2) 新胶囊网络的剪枝策略，对低层胶囊进行严格筛选，选择的高水平胶囊进行接下来的线性组合。胶囊块卷积层的设计，实现低层胶囊与高层胶囊的跳跃连接，来获取更加全面的高维特征，同时还能有效地解决胶囊中间层的学习抑制问题。

(3) 多个基准数据集上评估网络的性能：在大大减少计算量的同时，本文提出的网络具有独特地处理图像数据。例如，CIFAR-10 数据集中，与文献[6]的方法比较，所提出的剪枝胶囊网络有着 6.89% 的性能提高。

论文的其余部分组织如下：在第 2 节中，讨论了胶囊网络的相关工作；第 3 节介绍了关键算法原理；第 4 节概述了剪枝胶囊网络体系结构。第 5 节实验结构与验证。最后，第 6 节对论文进行了总结和展望。

2 相关工作

随着大数据时代的到来和硬件计算能力的提高，深度学习在许多领域中取得了优异的成绩，突破了传统模式识别和机器学习方法的局限性。但是深度网络所面临的一个重大问题之一是网络结构越来越复杂化^[7,8,9]，这将导致面临的网络结构越来越抽象。复杂的网络虽然能带来强劲的性能，但是也伴随着一些负面的影响，如对特定数据集存在过拟合的现象。这一问题通过剪枝算法可以得到有效地缓解。文献[10]介绍了一种新的通道修剪方法以加速深度卷积神经网络，在给定的 CNN 模型使用一种迭代式两步算法，通过基于 LASSO 回归的通道选择和最小二乘重构来修剪每一层。该方法能够增强网络的兼容性。文献[11]提出了一种有效且统一的框架，在训练和推理阶段同时加速和压缩 CNN 模型，其重点放在过滤修剪上。该方法不会更

改原始的网络结构，因此现有的深度学习库可以完美地支持它。Han 等人^[12]提出了一种新颖的修剪方法，该方法优化了模糊神经网络的结构。结构学习阶段和参数训练阶段是同时执行的。结构学习方法依赖于输出的敏感性分析，参数训练算法是使用监督梯度面方法实现的。为了缩小 CNN 设计和能耗优化之间的差距，文献[13]提出了一种能量感知修剪算法，与其他的修剪方法相比，提出的逐层修剪算法通过最小化输出特征图而不是滤波器权重中的误差来进行修剪。为解决深度神经网络 (Deep Neural Network, DNN) 参数数量巨大等问题，He 等人^[14]提出了一种节点修剪方法来重构 DNN 以实现快速解码。在这种方法中，对经过完全训练的 DNN 的隐藏节点进行修剪，使其具有一定的重要性函数，并使用反向传播对经过重塑的 DNN 重新调整。这些方法都优化了神经网络结构，同时也提高了网络的泛化能力。

自胶囊网络提出以来，国内外学者对其展开了深入的研究。文献[15]提出了一种新的胶囊网络架构，称为注意力机制路由胶囊网络 (Attention Routing Capsule Network)。用注意力机制路由和胶囊激活的动态路由来进行替换。注意路由是胶囊之间通过注意模块的路由，保持空间信息的同时快速前进。文献[16]介绍一种胶囊间路由协议 (Inter Capsule Routing Protocol, ICRP)，ICRP 协议使用特征检测矩阵在胶囊网络的每次迭代中存储学习到的信息。特征检测矩阵中搜索相似特征，可以逐渐减少图像中特征检测的处理时间。Gagana 等人^[17]提出了一种称为胶囊网络的分层路由体系结构，以克服卷积神经网络的缺点。胶囊用低级和高级神经单元之间的动态路由功能替代了卷积中的平均或最大合并技术，该功能可以更好地捕获数据中的层次关系，并引入了处理等方差特性的重构正则化机制。文献[18]首次将胶囊网络扩展到图像还原任务。在编码器方面，其思想是基于编码器-解码器的结构，通过用胶囊网络代替卷积神经网络。特别是，在解码器中添加了通用对抗胶囊网络 (Capsule-GAN)，以约束参数来扩展生成空间。与此相反，我们的工作进一步简化胶囊网络，并提高胶囊网络的识别性能。

在传统的卷积神经网络中，卷积池单元缺乏处理几何变换的能力，导致了特征空间层次信息的丢

失。但是，此信息对于内容的表达却至关重要。为了解决这个问题，文献[19]提出了一种胶囊动态检测器目标检测算法。该方法用动态路由胶囊层代替了原始框架的部分卷积池层，学习空间位置特征偏移量，以便卷积单元可以通过动态路由来学习不同特征之间的位置信息，从而提高了检测算法的性能。文献[20]提出了一种多尺度胶囊网络，该网络对于图像分类中的特征表示更加鲁棒和高效。多尺度胶囊网络包括两个阶段。在第一阶段，通过多尺度特征提取获得结构和语义信息。在第二阶段，特征层次被编码为多维高层胶囊。HSU 等人^[21]提出了两个框架：胶囊图像恢复神经网络和胶囊注意与重建神经网络，以将胶囊合并到图像卷积神经网络中。第一个框架利用胶囊中编码的丰富信息来重建高分辨率图像。另外一个通过胶囊的强大分割功能来生成图像注意特征。本文的工作研究剪枝策略来优化胶囊网络，同设计胶囊块卷积层来获取更加全面的高维特征，使得网络性能有所提升。

3 关键算法原理

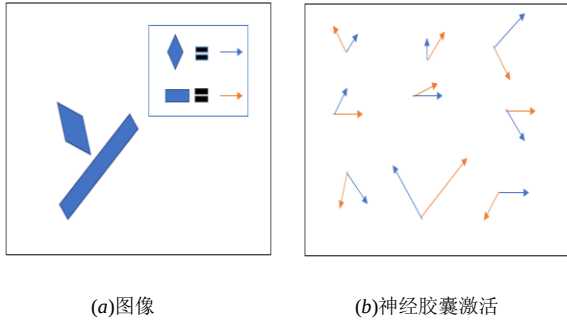


图1 向量化的神经胶囊

胶囊网络让分类器对图像中的对象有着更加深刻的理解。这是通过一组神经元组成的胶囊来实现的，实际上是一个包含多个神经元的高维活动向量。向量中的每个神经元代表实体中出现的各种属性，这些组成的向量神经元能更好的描述特征。如图1所示，图1(a)表示的是菱形和矩形存在的概率；而图1(b)表征的是这个实体存在的网络。图1(b)里面有18个高维活动的向量，即是18个胶囊。其中橙色箭头代表的是矩形胶囊存在概率，而蓝色箭头表示的是菱形胶囊存在概率。只有两个箭头比较大，表示着大概率在特定的位置上找到特定的对象，即对应图1(a)这个菱形和矩形。

3.1 耦合系数 c_{ij}

耦合系数是胶囊网络中一个重要部分，低层胶囊通过其自身的输出 $\{u_i = R^d | i=1,2,3,...,N_l\}$ ，乘以权重矩阵 W_{ij} 来计算预测向量 $\hat{u}_{j|i}$ ，如下所示：

$$\hat{u}_{j|i} = W_{ij} u_i \quad (1)$$

其中， W_{ij} 为低层第 i 个胶囊与相邻高层第 j 个胶囊之间的连接权值构成的权值矩阵。对于高级胶囊层的任意神经胶囊 j ，通过耦合系数 c_{ij} 与低层胶囊的神经胶囊 i 相连接，得到 s_j ，如下所示：

$$s_j = \sum_i c_{ij} \hat{u}_{j|i} \quad (2)$$

耦合系数 c_{ij} 表示的是高层胶囊对于低层胶囊的需求度。 c_{ij} 具有明显的几个特征。从图中看出，耦合系数 c_{ij} 是非负值的标量，即 $c_{ij} \geq 0$ 。此外，对于每一个低层胶囊 i ，与它所连接的所有高级胶囊 j ，其耦合系数之和为 1，即 $\sum_j c_{ij} = 1$ 。

3.2 动态路由算法

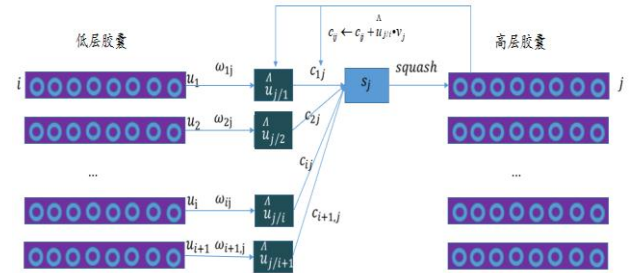


图2 神经胶囊加权连接示意图

低层胶囊代表的是对象的基本实体，使用动态路由算法将信息传递到下一层适当的高层胶囊中。动态路由算法过程如图2所示，它决定从低层胶囊的输出应该通过一个迭代过程进入下一个高层胶囊。这个机制是基于低层胶囊对高层胶囊的参数预测。预测是通过一个变换矩阵来计算的，当来自低层胶囊的多个预测与高层胶囊的输出一致时，该机制就被激活。动态路由机制如算法1所示。

算法1: ROUTING

```

0: procedure ROUTING( $\hat{\mathbf{u}}_{j|i}, r, L$ )
1: Require:
2:   for 对于在  $L$  层中的所有胶囊  $i$ , 和在  $L+1$ 
3:     层中的所有胶囊  $j$ :  $b_{ij} \leftarrow 0$ 
4:   for  $r$  in iterations do
5:     for 对于在  $L$  层中的所有胶囊  $i$ ,
6:        $c_{ij} \leftarrow \text{softmax}(b_{ij})$ 
7:     for 对于在  $L+1$  层中的所有胶囊  $j$ ,
8:        $\mathbf{s}_j = \sum_i c_{ij} \hat{\mathbf{u}}_{j|i}$ 
9:     for 对于在  $L+1$  层中的所有胶囊  $j$ ,
10:       $\mathbf{v}_j \leftarrow \text{squash}(\mathbf{s}_j)$ 
11:     for 对于在  $L$  层中的所有胶囊  $i$ , 和在
12:        $L+1$  层中的所有胶囊  $j$ :
13:        $b_{ij} \leftarrow b_{ij} + \hat{\mathbf{u}}_{j|i} \cdot \mathbf{v}_j$ 
14:   return  $\mathbf{v}_j$ 

```

4 剪枝胶囊网络的框架

在本节中，剪枝胶囊网络的结构框架如图3所示，输入的数据经过初级卷积层提取出最初的特征；然后将得到的初级特征进行胶囊块卷积层进行运算，得到更加全面的高级特征，获取更加细粒度的特征。这些特征形成的胶囊具有大小和方向，是反映特征的最小单位。但是这些低层胶囊具有以及冗余性，很多是杂乱的、不重要的低级胶囊。冗余的低级胶囊对模型的训练增加了复杂性，甚至对实验结果造成严重的干扰。因此，本文设计了剪枝层，来简化我们的胶囊层。在剪枝层中，我们需要计算每一层中的胶囊的概率（通过压缩函数可以得到），然后筛选最大活跃度的 K 个低级胶囊与高层胶囊通过动态路算法进行连接训练，该过程修正下层胶囊与上层胶囊之间的变换矩阵和激活值等参数，以形成最优整合结果。下面将详细的介绍深度剪枝胶囊网络中的每个层的重要作用。

4.1 初级卷积层

获取低层胶囊中的全部特征是胶囊网络的一个重点，胶囊网络的第一部分是由初级卷积层组成。初级卷积利用卷积核对输入数据进行滤波操作，从而对数据进行特征提取。

池化操作淡化图像中相对空间特征，这样造成

的结果是同一个物体在不同位置也能被识别，且对物体的旋转、移动、平移等并不敏感，这样大大增加了分类效果的鲁棒性，但是相应地也丢失了特征空间的整体性。因此，在卷积层中，我们不使用池化操作，而仅仅使用卷积操作。通过卷积操作获得图片数据的全部初级特征，当然这些特征是简单、粗糙的。本文通过初级卷积层，把像素转化成局部的特征输出，该层的激活函数使用 Relu。这一层主要是对图像进行预处理，抽取到合适的特征。由于一层的卷积层不足以获取合适的特征，因此在网络中又添加了一个卷积层，成为初级胶囊层，初级胶囊层属于卷积层，但传统的卷积层不也一样，其操作对象不是单个的神经元，而是粒度更大的神经胶囊单元（一组神经元组成）。初级胶囊层可以理解为以神经胶囊为对象的卷积层。

4.2 胶囊块卷积层

对于复杂的数据集，有可能获取的特征不足使其难以在胶囊中进一步处理。通过设计胶囊块卷积层，主要是获取更抽象、更细粒度的特征，使得剪枝胶囊网络能够处理相对复杂的图像数据。

在胶囊块卷积层中，假设 $\Phi^l = (c^l, m^l, m^l, n^l)$ 作为该层 l 的输入， $\Phi^{l+1} = (c^{l+1}, m^{l+1}, m^{l+1}, n^{l+1})$ 作为该层 l 的输出。其中 c^l 代表的是层 l 的胶囊数目， m^l 表示的是层 l 特征图的高度和宽度， n^l 表示的是层 l 的胶囊维度。设 3D 卷积核为 (g, g, n^l) ， $(1, 1, n^l)$ 用作 3D 卷积运算的步幅。将 $\Phi^l = (c^l, m^l, m^l, n^l)$ 进行形变为 $\hat{\Phi}^l = (m^l, m^l, c^l, n, 1)$ ，然后与 $(c^{l+1} \times n^{l+1})$ 个 3D 卷积核进行卷积。计算两个相乘得到的中间值 $(m^{l+1}, m^{l+1}, c^l, c^{l+1} \times n^{l+1})$ 。其中 $(m^{l+1}, m^{l+1}, c^l, c^{l+1} \times n^{l+1})$ 中的每个值 θ_{ijt} 可以经过如下公式得到：

$$\theta_{ijt} = \sum_p \sum_q \sum_r \hat{\Phi}^l(i-p, j-q, t-r) \cdot \ell_k(p, q, r) \quad (3)$$

使用 3D 卷积核作为转换矩阵，可以让我们使用一组较低水平的胶囊来预测较高水平的胶囊。由于深度多层思想在解决复杂数据问题具有引人注目的优势。

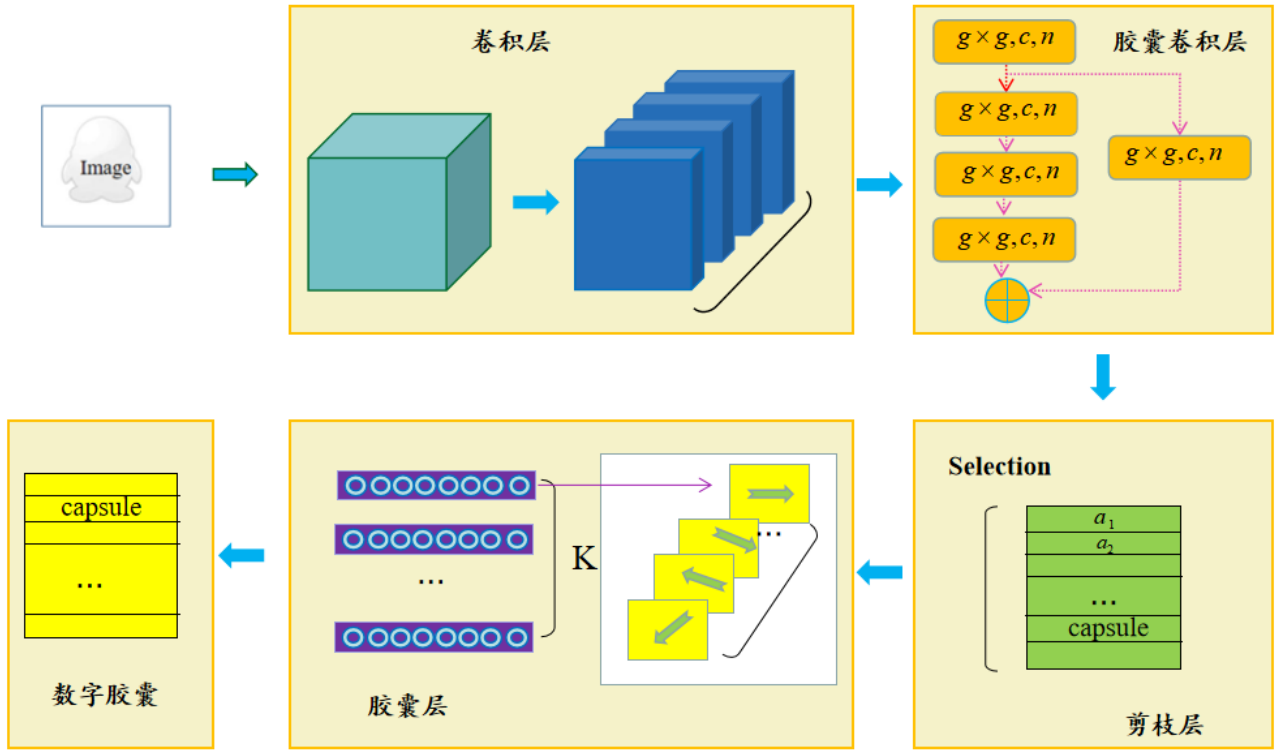


图3 剪枝胶囊网络图

因此，本文试图通过增加更多的卷积层来获取更抽象的特征。我们知道，CNN 模型通过简单地添加卷积层和神经元来捕捉不同的特征变量。然而，深度的增加很容易导致梯度消失的问题。因此，本文设计的胶囊块卷积层使用 3D 卷积思想来更加全面的对数据进行卷积。普通的胶囊网络的初始卷积层只是将像素强度转换为局部特征检测器的矢量活动，作为主胶囊的输入。本文使用 3D 卷积核将局部的胶囊子集转化为一个更加抽象的胶囊表决。例如一个局部特征实体是由彼此相连相依的胶囊组表示。

在这种情况下，胶囊块卷积层的不同水平上检测到的所有特征组合起来，作为初级胶囊的输入。本文设计的胶囊块卷积由 3 层 3D 卷积进行连接，第一层的胶囊块卷积输出分为两条路线。其中的一条路线正常卷积，向下连接；而另外一条卷积输出直接跳过第二层卷积，连接到最后一层的输出。这样设计的目的是为了减少深度模型中的梯度的消失。因为许多相邻胶囊连接时候，它们的系数往往太小，从而限制着梯度的流动，抑制着网络学习的方向。同时 $L+1$ 层中的每个胶囊张量 $\mathbf{S}$ 将具有来自

L 层的对应预测。该预测将通过如下公式定义的 3D squash 函数传递。将胶囊向量的长度限制在 0 和 1 之间，表示实体存在的可能性。

$$\begin{aligned} \hat{\mathbf{S}} &= \text{squash_3D}(\mathbf{S}) \\ &= \frac{\|\mathbf{S}\|^2}{1 + \|\mathbf{S}\|^2} \frac{\mathbf{S}}{\|\mathbf{S}\|} \end{aligned} \quad (4)$$

此外，本文设计的胶囊块卷积层直接路由由低水平胶囊到高水平胶囊之间的跳跃连接。在跳过连接后，使用元素层加法来连接两个卷积胶囊层的输出。由于胶囊用矢量表示，所以当它重复相同的胶囊时，不使用通道式连接，但元素相加减少了偏置并降低了对噪声的敏感性。同时，在胶囊块卷积层获取了更多的特征，意味着一些相关度不大的特征也随之获取。为了筛选这些特征，在剪枝层中进行处理。

4.3 胶囊剪枝层

通过胶囊块卷积层，我们得到了能表示复杂数据的特征表示，这些特征进行整组形变后，可以形

成一个个低级胶囊。然而这些低级胶囊具有冗余性，很多低级胶囊是对实验效果无关，不重要的，因此我们引入了剪枝胶囊层，来实现对低级胶囊的筛选。由于某些胶囊的活性明显高于其他胶囊。另外，对于预测的高层胶囊 v_j ，这些活性高的胶囊 u_i 往往具有相对较大的耦合系数 c_{ij} 。这是因为高层胶囊的构建主要是由高度活跃的低层胶囊贡献的。因此，而那些活性低的胶囊贡献度不大。

为确保只有重要的、活跃的胶囊输出发送到下一层高级胶囊层中。我们期望剪枝胶囊层不仅可以减少计算负担，而且提高了网络泛化能力。设低层胶囊的矢量输出为：

$$U^l = \{u_i \in D^n \mid i = 1, 2, \dots, c_l\},$$

其相应的活动水平通过 `squash` 函数获得（或实体存在的概率）表示为：

$$A^l = \{0 \leq a_i \leq 1 \mid i = 1, 2, \dots, c_l\}$$

其中 u_i 的方向表示实体的姿态， u_i 的长度表示活动级别，或实体存在的概率，并且

$$a_i = \|u_i\|$$

我们选择 K 个最活跃的胶囊。设 a_i 表示在 a 中第 i 个活跃的胶囊。对低层胶囊的所有胶囊进行排序，然后选择最活跃的 K 个低层胶囊连接下一层高层胶囊：

$$\{i \mid a_i \geq a_{(K)}\}$$

通过获取这 K 个最活跃的低层胶囊的索引值，就可以取得相应的 K 个低层胶囊，然后对其进行重新排序。使抽出的胶囊重组成新的活跃低层胶囊，进行加权连接预测高层胶囊。对于选择 K 个最重要的活跃胶囊，与高层胶囊进行连接。这个过程使用的是一种动态路由算法来实现。如何决定从低层胶囊的输出应该通过一个迭代过程进入下一个高层胶囊。这一层为胶囊网络层，进行胶囊网络的后续训练。

剪枝后的胶囊网络之所以能够减少计算负担，是因为胶囊网络从层 L 传播到层 $L+1$ 之前，通过选择 K 个最活跃的胶囊。对 L 层中的所有胶囊进行排序，然后选择最活跃的 K 个低层胶囊连接下一层高层胶囊。只有 K 个胶囊参与后续工作，而这些计算

是很少的，不会占用比较多的时间。如果 K 选择 15%，也就是说在 L 层中，只有 15% 的胶囊参与 $L+1$ 层的胶囊相连接。这极大的缩小了 L 层胶囊与 $L+1$ 层的胶囊之间的计算负担。同时也会减少网络的学习时间。深层胶囊网络对于相对复杂的任务，通过胶囊块卷积层可以获取粒度更大的神经特征，然后经过剪枝层的简化，同样降低时间的消耗，对神经网络可以更好的支持，大大节省了神经胶囊参与动态路由的运算。

4.4 损失函数

在胶囊网络中，为了让学习过程得到控制，传统的交叉熵损失函数是无法胜任此类工作的，这是因为交叉熵损失函数是独热编码，只允许存在一个分类。因此本网络使用的是边缘损失函数作为模型优化的目标函数，具体公式如下所示：

$$L_k = T_k \max(0, m^+ - \|v_k\|)^2 + \lambda(1 - T_k) \max(0, \|v_k\| - m^-)^2 \quad (5)$$

其中， k 为分类， T_k 是关于分类的函数，当且仅当 k 出现时，有 $T_k=1$ ，不存在时候， $T_k=0$ 。 $\|v_k\|$ 表示的是向量的模长，实际上就是数字 k 存在的概率，而 m^+ 和 m^- 都是阈值函数，表示胶囊之间连接的强度，大于 0.9 表示的是完全连接了，小于 0.1 认为不连接了。 λ 为稀疏系数，调整两者的比重，以调节参数调整的步伐。对于缺少数字类别的损失， λ 的下调权重会阻止初始学习缩小所有数字胶囊的活动向量的长度。因此本文使用 $\lambda=0.5$ ，总损失仅仅是所有数字胶囊损失的总和。边缘函数通过设置 m^+ 和 m^- ，来进行正负判别。当 $T_k=1$ ，因为如果 v_k 接近 m^+ ，说明这个类别概率很大，那么这个类别就是正的；当 $T_k=0$ ，如果负例 v_k 和 m^- 的距离靠近，那么这个类别就是负的，来进行学习优化的。

4.5 剪枝胶囊网络算法

深度剪枝胶囊网络的过程 `Pruning capsule` 算法的伪代码如算法 2 所示。

算法 2: `Pruning capsule`

0: Procedure: Pruning

1: Require: $\Phi^l = (c^l, m^l, n^l)$, r , and c^{l+1} , n^{l+1} 2: $\Phi^l \leftarrow \text{Reshape}(\Phi^l)$

3: 经过 3D 胶囊块卷积层

4: $V_1 \leftarrow \text{Conv3D}(\Phi^l)$ 5: $V_2 \leftarrow \text{Conv3D}(V_1)$ 6: $V_3 \leftarrow \text{Conv3D}(V_2)$ 7: $V_4 \leftarrow \text{Conv3D}(V_3)$ 8: $V_5 \leftarrow V_1 + V_4$

9: 经过 Pruning capsule 层

10: $b_{ij} = 0$ 11: for r in iterations do12: for 对于在 L 层中的胶囊 i , $K_{ij} \leftarrow \text{softmax}(b_{ij})$ 13: for 对于在 $L+1$ 层中的胶囊, $s_j \leftarrow \sum_i K_{ij} V_i$ 14: for 对于在 $L+1$ 层中的胶囊, $v_j \leftarrow \text{squash}(s_j)$ 15: 选择最优的 top K 胶囊: $\text{pro} = \text{selection}(v_j, K)$ 16: $\Phi^{l+1} = \text{capsule}(\text{pro})$ 17: return Φ^{l+1}

5 实验及结构分析

5.1 数据集介绍

MNIST^[22]数据集来自美国国家标准与技术研究所, 一个具有 10 个数字的手写体数字识别数据集。包含 60000 个训练样本和 60000 个训练标签、10000 个测试样本以及 10000 个对应的测试标签组成的。每个样本都是一张 28×28 像素的灰度手写数字图像。所有图像的标签均被初始化为 1-0 维的 one-hot 向量。MNIST 选取的部分数据集如图 4 所示。

CIFAR-10^[23]是由 32×32 像素的 RGB 图像组成。与 MNIST 一样具有 10 种类别的物体, 每个类有 6000 张图像, 总共分为 50000 张训练图像和 10000 张测试图像。数据集分为五个训练批次和一个测试批次, 每个批次 10000 张图片。CIFAR-10 部分数据集如图 5 所示。

Fashion-MNIST^[24]数据集是德国 Zalando 公司提供的衣物图像数据集, 包含 60,000 个样本的训练集和 10,000 个样本的测试集。每个样本都是 28×28 灰度图像, 与 10 个类别的标签相关联。Fashion-MNIST 部分数据集如图 6 所示。



图 4 MNIST 数据集

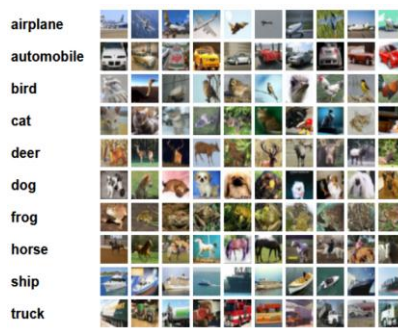


图 5 CIFAR-10 数据集



图 6 Fashion-MNIST 数据

5.2 实验分析

5.2.1 剪枝胶囊网络 K 的性能实验结果

为确保只有重要的、活跃度高的胶囊输出到下一层高级胶囊层中。本文提出的剪枝策略是概率活跃度剪枝策略。其主要步骤如下所示:

1) 在 L 层中的所有低层胶囊网络通过 $\text{squash}()$

函数: $v_j \leftarrow \text{squash}(s_j)$ 得到每个层中的低级胶囊的概率。

2) 在 L 层中筛选最高的胶囊概率 K 个: 对得到的每个胶囊的概率进行排序, 然后根据阈值条件选择活跃度大的 K 个胶囊, 获取其索引值, 重新组成初级胶囊, 进行下一层的高级胶囊连接。

$\text{pro} = \text{selection}(v_j, K)$

3) 重新训练网络剩余稀疏的最终权重。

(1) MNIST 数据集

为了充分证明剪枝胶囊网络不同的 K 值能力,我们在低级胶囊中选择最活跃的 K 个胶囊, K 值选择的是原来的初级胶囊的 10%、15%、20%、25%、30%、35%、40%、45%做的一系列实验。MNIST 数据集不存在数据增强的情况,是在原始数据的情况下进行的实验。由于误差较小,因此使用 (%) 表示误差。实验结果见表 1 所示。

表 1 不同 K 值的误差率结果

K	10%	15%	20%	25%
MNIST(%)	0.91	0.84	0.48	0.72

30%	35%	40%	45%
0.64	0.66	0.71	0.75

从表 1 可以看出,总体来说,图像的识别情况都不错。当 $K=10\%$ 的时候,出现的误差相对比较大,这是由于剪枝过多,我们选取的活跃度比较高的低级胶囊不足以表达全部的图片的特征,造成我们的模型没有完全学习图片的全部特征,因此实验测试效果不佳;而当 $K=15\%$ 的时候,我们发现训练的误差明显下降,这是因为 K 的增加有助于提升模型对图像数据的特征获取能力;而当 $K=20\%$ 的时候,模型的效果得到最佳;当 K 大于 20% 时候,发现模型的测试效果有所下降,这是因为出现冗余的低级胶囊,我们不需要那么多活跃度不高的低级胶囊。因此,只是单纯的增加低级胶囊层的胶囊是行不通的,当超过一定值时候,会出现效果下降的结果。

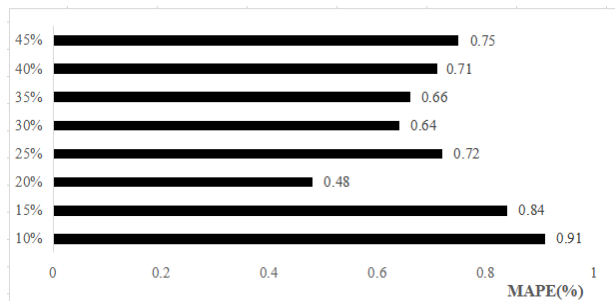


图 7 K 值柱状图误差 (%)

为了更加直观地观测其变化过程,本文使用了柱状图来表示这一现象,如图 7 所示,可以看出:剪枝胶囊网络中,选取不同的 K 值对实验结果有着

重要的影响,不同的 K 值引起的误差呈现出不同的趋势,但是在数据没有增强的情况下,实验结果也有所稳定,说明剪枝策略是一种稳定、有效的策略。在不同的 K 值中,其中 $K=10\%$ 的表现效果欠佳, $K=20\%$ 实验效果最好。说明了剪枝后的胶囊网络达到了预期的实验效果。

(2) CIFAR-10 和 Fashion-MNIST 数据集

同理,我们在 CIFAR-10 和 Fashion-MNIST 数据集中,同样在低级胶囊中选择最活跃的 K 个胶囊, K 值选择的是原来的初级胶囊的 10%、15%、20%、25%、30%、35%、40%、45%做的一系列实验。实验结果详见表 2 所示。

表 2 不同 K 值的误差率结果

K	10%	15%	20%	25%
CIFAR-10(%)	21.80	20.15	17.19	18.16
Fashion-MNIST(%)	10.09	8.41	7.20	9.32

30%	35%	40%	45%
20.79	19.15	22.13	23.22
8.54	8.28	8.18	9.42

从中可以看出,在相对复杂的数据集 CIFAR-10 和 Fashion-MNIST 中,选择 $K=20\%$,实验的误差相对较小,达到最优的效果。

5.2.2 剪枝胶囊网络深度的实验结果

处理复杂的数据集的时候,大部分胶囊网络通过增加卷积层,来对数据特征的获取;但是仅仅通过增加卷积层是不足以获取整体的数据结果,因此,本文设计了一种新颖的胶囊块卷积层,来进行图像数据的获取。在胶囊块卷积层中,低级胶囊相邻间的单个胶囊组成一个低级的实体。通过一个 3D 胶囊卷积后,分别进行后续的两条路径,一条继续经过胶囊块卷积进行 3D 卷积,进行连续 2 次胶囊卷积,另外一条只进行一次卷积,这样得到的两条路径的结果进行融合,合成一次胶囊块卷积的操作。通过如此设计的胶囊块卷积层,其目的是减少深度模型中的消失梯度。同时对复杂数据进行更抽象的特征获取。本文在相对复杂的 CIFAR-10 和 Fashion-MNIST 数据集上试验。而 MNIST 数据集属于简单数据集,因此不做多层实验。

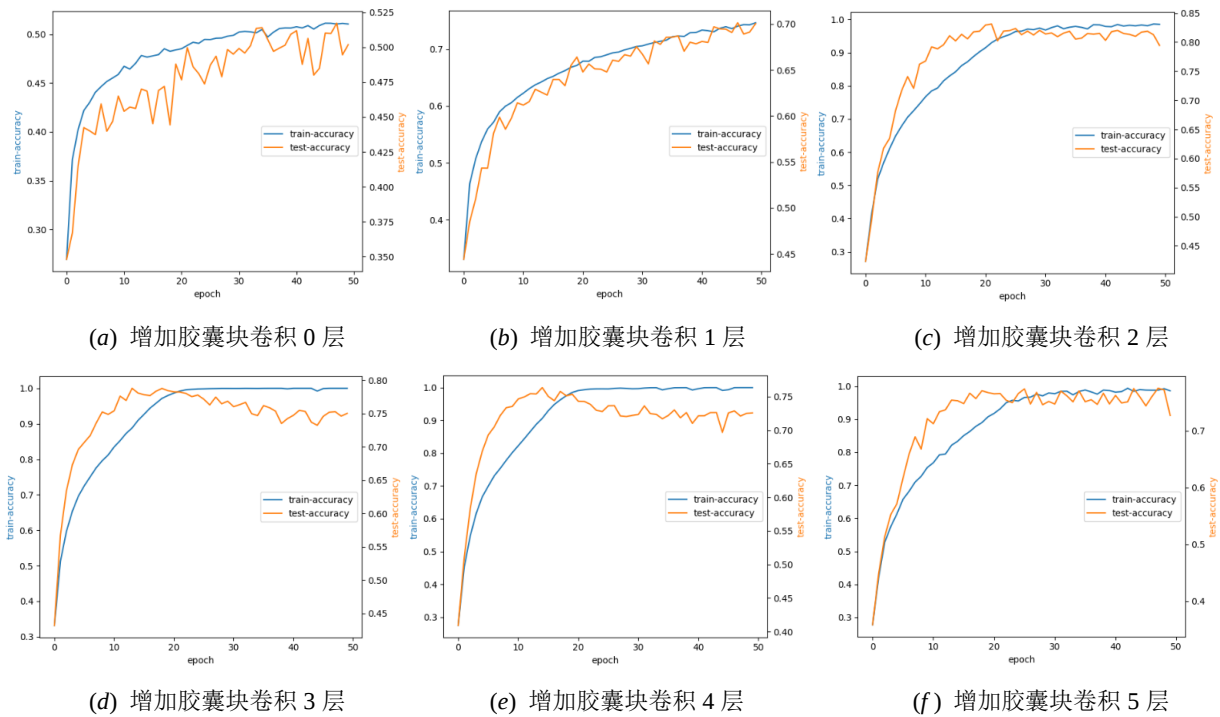


图8 CIFAR-10 胶囊卷积块层

(1) CIFAR-10 数据集

使用 CIFAR-10 数据集集中的 50000 个训练样本进行训练, 通过 50000 个训练样本对网络训练一个轮次, 采用 10000 个测试图像进行测试, 以得到当前网络对于图像的识别准确率。

在本实验中, 利用 CIFAR-10 数据集对该网络模型进行了 50 epoch 的训练。训练过程中, 该网络模型识别的变化情况如图 9 所示。使用 CIFAR-10 数据集, 是在没有进行数据增强的情况下使用。其中增加 0 层、1 层、2 层、3 层、4 层、5 层胶囊卷积层来观测实验结果。

表3 添加胶囊块层的实验误差结果

层数	Train-ACC	Test_ACC
0	0.5345	0.5322
1	0.7413	0.7011
2	0.9755	0.8281
3	0.9735	0.7874
4	0.9744	0.7658
5	0.9863	0.7567

从图 8 可以看出: 在胶囊块卷积层中, 通过不增加的情况如图 8(a)所示, 训练的准确率和测试的准确率都不是很高, 说明数据没有获得全部的特征; 图 8(b)所示, 在增加一层的胶囊块卷积层的时候,

可以看出, 训练的准确率和测试的准确率都有明显的提升; 图 8(c)所示, 在增加 2 层的胶囊块卷积层时候, 实验的训练准确度再次提高, 同时测试准确度也再次提升; 图 8(d)所示, 训练的准确度已经不再提升了, 但是测试的准确度有所下降; 图 8(e)、8(f)所示, 测试的准确度再次波动, 说明了在增加 2 层的情况下, 实验效果是最好的, 而无止境的增加 3D 胶囊块卷积层, 只会增加网络的训练时间, 而不会对实验的准确度有所提高。其具体实验数值如表 3 所示。观察通过添加胶囊块卷积层, 可以获得更加全面的特征, 但是添加过多层的胶囊块对实验效果不明显, 反而引进更多关联度不高的特征。因此可以得出结论: 超过 2 层胶囊块卷积层, 会使实验效果趋于平缓, 波动的情况, 而无准确度的提升。

为了更加直观地观测添加胶囊块卷积层的变化, 本文画出了曲线图来显示其变化的趋势, 从图 9 所示, 添加 0 层到添加 5 层胶囊块卷积层, 其准确率慢慢增加; 添加 2 层胶囊块卷积层测试性能达到最优; 超过 2 层胶囊块卷积层又呈现下降的趋势。可以证明我们的假设, 即增加胶囊块卷积层可以获取更多特征, 也提高了预测的精度。但简单的增加胶囊块卷积层的数量并不是有效的。增加 2 层的实验效果是最好的。添加胶囊块卷积层是一个变化的

过程。

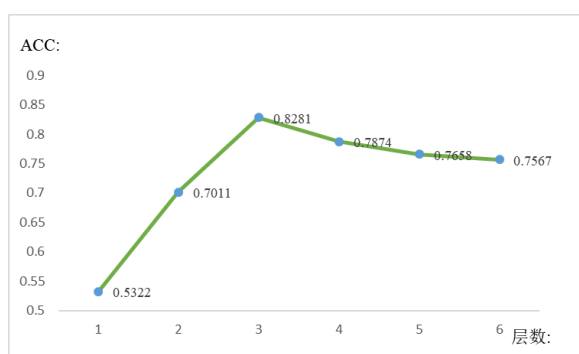


图9 胶囊卷积块层变化趋势

(2) Fashion-MNIST 数据集

使用 Fashion-MNIST 数据集中进行测试，同样的在本实验中，对该网络模型进行了 50 epoch 的训练。训练过程中，该网络模型识别的变化情况如图 11 所示。同样的，本文使用 Fashion-MNIST 数据集是在没有数据增强的情况下使用。其中增加 0 层、1 层、2 层、3 层、4 层和 5 层，观测实验。

实验结果如图 10，可以看出，在 Fashion-MNIST 数据集中，不增加胶囊块卷积层的情况如图 10(a) 所示，与 CIFAR-10 数据集一样，训练的准确率和

测试的准确率都不是很高，原始的胶囊网络无法获得数据的全部的特征；图 10(b)所示，在增加一层的胶囊块卷积层，训练的准确率和测试的准确率都有适当的提高，但是还没有达到预期的效果；图 10(c)所示，在增加 2 层胶囊块卷积层，测试的准确率到达了 90% 以上，是一个显著的提高；同时，其训练的准确度保持着稳定的波动；图 10(d)所示，添加 3 层的胶囊块卷积层，训练的准确率基本不变，但测试的准确率开始有所下降；而图 10(e)、10(f)可以看出，训练的准确率和测试的准确度在原来的基础上也是来回波动，基本达到一个平衡的状态。从中可以看出：随意的增加胶囊块卷积层是无法增加其测试的准确度，其准确率会达到一个上限，而无法继续增加。同 CIFAR-10 数据集实验结果类似，超过 2 层，会使实验效果趋于平缓，波动的情况，但是不会增加其测试的准确度。可以得出结论：增加 2 层胶囊块卷积，能够得到获取数据的全部特征，实验效果最佳，而增加超过 2 层的胶囊块卷积，不会使得实验效果更好，反而增加了模型的复杂度，因此不建议添加更多的胶囊块卷积层。

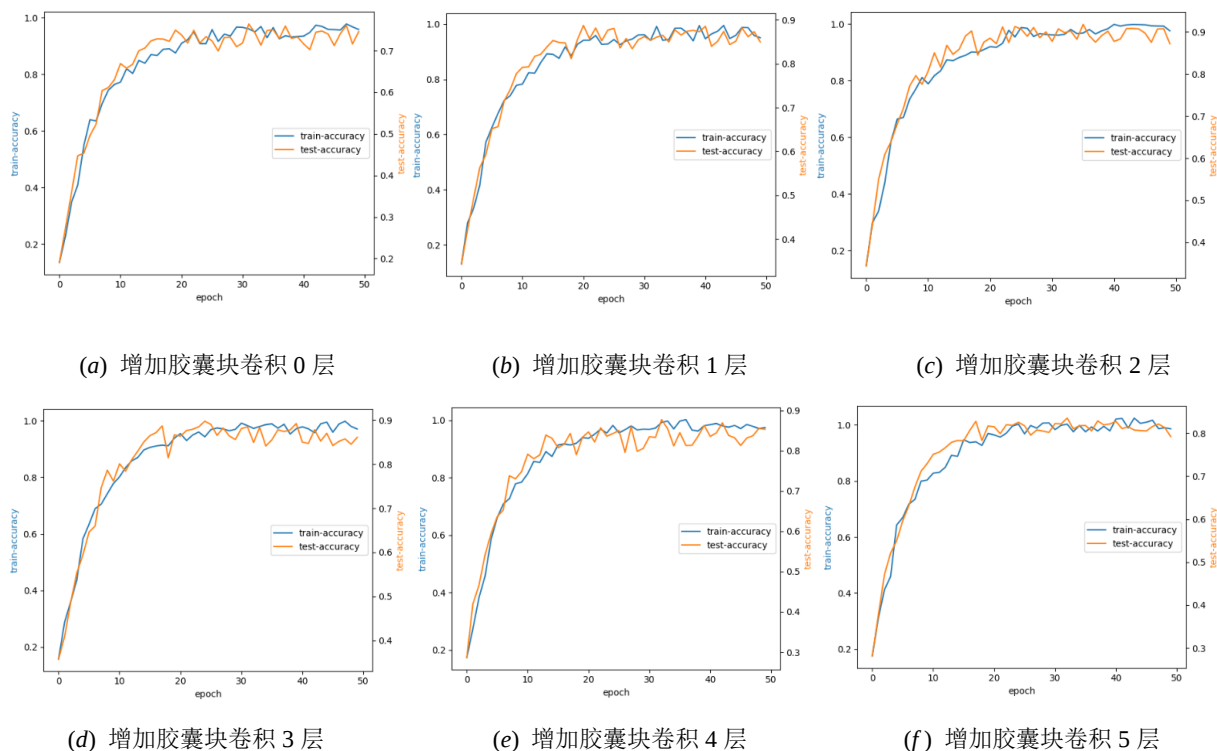


图10 Fashion-MNIST 胶囊卷积

5.2.3 动态路由的迭代次数的实验结果

本节的主要目的是验证动态路由迭代次数对实验的影响效果。在胶囊网络中,神经元被分组为胶囊的单元,路由算法通过更新步骤的迭代来确定它们之间链接的强度,发挥着举足轻重的影响。本实验中在统一其他的参数的情况下,通过改变迭代次数来验证其对网络的识别的准确率。本文分别使用迭代次数为 1、2、3、4 来验证其变化的趋势,实验结果如表 4 所示。

表 4 不同路由迭代次数下的准确率

数据集	$R=1$	$R=2$	$R=3$	$R=4$
MNIST	0.781	0.842	0.995	0.855
CIFAR-10	0.550	0.682	0.828	0.722
Fashion-MNIST	0.781	0.850	0.928	0.899

通过在 MNIST、CIFAR-10 和 Fashion-MNIST 数据集上更改迭代次数 R 的值,从表 4 可以看出,在 $R=1$ 的时候,3 个数据集的准确率都表现不佳;而在 $R=2$ 的时候,测试准确率有着明显的提高;当 $R=3$ 时候,剪枝胶囊网络的识别率是最高的,达到性能最优;而 $R=4$ 时候,又开始下降的趋势;为了更加直观的观测,本文画出了折线图,如图 11 所示,迭代次数 R 呈现着变化的趋势,得出的结果是在 $R=3$ 的时候,剪枝胶囊网络是最好的结果。由于迭代的次数不多,在各个数据集中,在 $R=3$ 的时候,网络已呈现出最佳的次数。因此不做自适应的迭代路由次数。

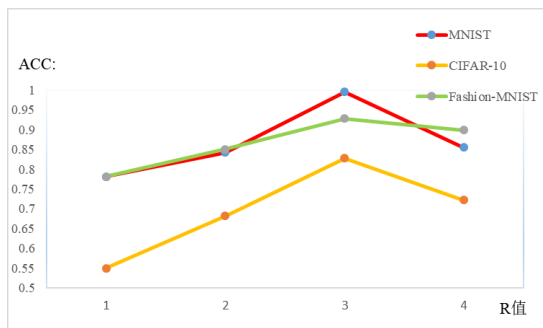


图 11 不同路由迭代次数

5.2.4 胶囊模型之间的比较

本文在几个基准数据集 (MNIST^[22]、CIFAR10^[23]和 Fashion-MNIST^[24]) 测试了我们的剪枝胶囊网络模型,并将其性能与现有的其他胶囊网

络体系结构进行了比较。实验中的数据集,本文不使用数据增强的情况,在整个实验中使用原始图像大小。

表 5 不同胶囊网络的实验结果比较

Model	CIFAR-10	F-MNIST	MNIST
Sabour <i>et al.</i> [5]	-	-	99.75%
Nair <i>et al.</i> [25]	67.53%	89.80%	99.50%
HitNet [26]	73.30%	92.30%	99.68%
Max-Min [6]	75.92%	92.07%	99.55%
MS-CapsNet [27]	75.10%	92.20%	-
MS-CapsNet+drop [27]	75.70%	92.70%	-
Mlcn2[28]	75.18%	92.63%	-
PCN-K($K=10\%$, $D=2$)	78.20%	89.91%	99.09%
PCN-K($K=20\%$, $D=2$)	82.81%	92.80%	99.52%
PCN-K($K=30\%$, $D=2$)	79.21%	91.72%	99.36%
PCN-D ($D=1$, $K=20\%$)	70.11%	89.76%	-
PCN-D ($D=2$, $K=20\%$)	82.81%	92.80%	-
PCN-D ($D=3$, $K=20\%$)	78.74%	90.18%	-

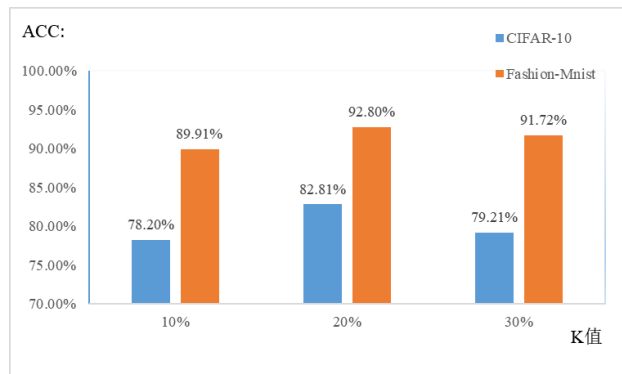


图 12 $D=2$, 不同 K 的取值结果

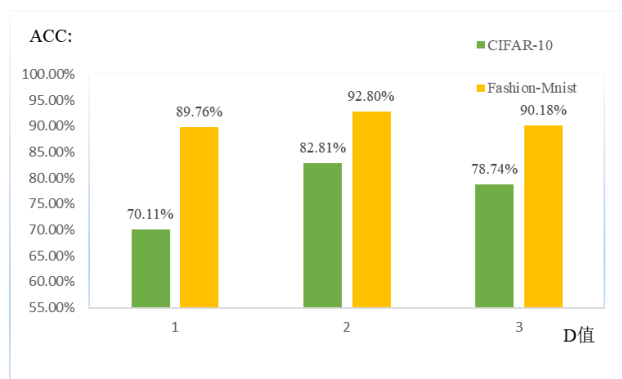


图 13 $K=20\%$, 不同 D 的取值结果

如表 5 所示，其中 PCN 中的 K 表示的是选取初级胶囊的比例，而 D 表征的是胶囊块卷积层的层数。本文的胶囊网络模型在不同的变种中，都表现着一定的稳定性。在 CIFAR10 数据集中，本文提出的剪枝胶囊网络性能超过了其他的胶囊网络模型，如果我们以最好的结果呈现，与文献[6]的方法比较，所提出的剪枝胶囊网络有着 6.89% 的性能提高；在 Fashion-MNIST 数据集，我们的性能优于文献[27]的胶囊网络结果，性能提高的结果 0.10%；而在 MNIST 数据集上，我们的结果与最优的结果相接近，相差不大；说明了剪枝胶囊网络是一种有效的、通用的胶囊网络。

从图 12 可以看出，在选取最优的活跃度时候，模型的 K 的选取对实验有影响，但是整体上不同的 K 值对数据集都有不错的表现能力，其中在深度一定时候， $K=20\%$ 取值最好，在测试的准确度达到最佳。而在 $K=10\%$ 的取得 78.20%，是由于其剪枝过多的低层胶囊导致学习不足； $K=30\%$ 时候取得 79.21% 剪枝不多，一些冗余的低层胶囊参与使得实验效果有些降低；但其总体的水准都保持在一定的水平中。而从图 13 可以看出，在 $K=20\%$ 时候，添加胶囊块卷积层对于复杂数据有着一些影响，从实验中可以看出，添加 2 个胶囊块卷积层时候，效果最佳。添加 1 层，网络还没有获得输入的全部需要的特征，因此表现效果有些欠佳；而添加 3 层，可以获取需求的数据特征，但是不会使得实验效果更好，反而增加了模型的复杂度。但是整体上都表现出不错的水准。因此，本文提出的剪枝胶囊网络在不同的数据集上都展示着不错的效果。

6 结论

胶囊网络作为一种新型的神经网络结构，在计算机领域中发挥着越来越重要的作用。由于其将传统的神经元标量提升为胶囊向量的输出，极大地增加了模型的特征表示能力。同时，近似的实体在相同的位置，由同一个胶囊感知；而相同的实体在不同的位置，是由不同的胶囊感知。另外胶囊网络的计算是并行的，可以同时预测不同实体的变换情况。这些都是 CNN 所不具备的能力。

本文提出的深层剪枝胶囊网络利用了两个关键思想：剪枝优化和深度思想。剪枝优化的应用能

够有效地实现低层无关的、冗余的胶囊剪枝掉，而保存活跃度大的、贡献显著的胶囊与高层胶囊相连接，极大地简化了胶囊网络结构；而深度思想能够处理更加复杂的图像数据，可以获取更加复杂、细粒度的特征表示，使网络能往好的方向进行。建立更加全面、健壮的网络体系结构。在几个基准数据集上评估剪枝胶囊网络的性能，本文的剪枝胶囊网络具有一定程度上提高网络的性能，超过了大部分现有的相关算法。

但目前的分类结果并没有达到我们预期的结果，今后将尝试对目前的网络做进一步的改动，以达到更高的分类准确率，例如对剪枝胶囊网络中的“向量”提升到更高维度的“张量”，是否对数据特征有着更加不一样的表示特征呢？对于动态路由算法，提出更加优秀的算法来替换，都有进一步研究的期待，是值得研究的课题。

参考文献

- [1] K. He, X. Zhang, S. Ren, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37(9): 1904-1916.
- [2] H. Shin et al. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Transactions on Medical Imaging*, 2016, 35(5): 1285-1298.
- [3] K. Zhang, W. Zuo, Y. Chen, et al. Beyond a Gaussian denoiser: residual learning of deep CNN for image denoising. *IEEE Transactions on Image Processing*, 2017, 26(7): 3142-3155.
- [4] Y. Chen, H. Jiang, C. Li, et al. Deep feature extraction and classification of hyperspectral images based on convolutional neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 2016, 54(10): 6232-6251.
- [5] Sabour, S, Frosst, N., Hinton, G. E. Dynamic routing between capsules. *Advances in Neural Information Processing Systems*, 2017: 3859-3869.
- [6] Zhao Z, Kleinhans A, Sandhu G, et al. Capsule networks with max-min normalization, CoRR, <http://arxiv.org/abs/1903.09662>, 2019.4.1.
- [7] K. He, X. Zhang, S. Ren, et al. Deep residual learning for image recognition, 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, USA, 2016:770-778.
- [8] C. Dong, C. C. Loy, K. He, et al. Image super-resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016, 38(2): 295-307.
- [9] S. Xie, R. Girshick, P. Dollár, et al. Aggregated residual transformations for deep neural networks. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, USA, 2017:5987-5995.

- [10] Y. He, X. Zhang, J. Sun. Channel pruning for accelerating very deep neural networks. 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 2017: 1398-1406.
- [11] J. Luo, J. Wu, W. Lin. ThiNet: a filter level pruning method for deep neural network compression. 2017 IEEE International Conference on Computer Vision (ICCV), Venice, 2017:5068-5076.
- [12] H. Han, J. Qiao, A self-organizing fuzzy neural network based on a growing-and-pruning algorithm. IEEE Transactions on Fuzzy Systems, 2010, 18(6): 1129-1143.
- [13] T. Yang, Y. Chen, V. Sze. Designing energy-efficient convolutional neural networks using energy-aware pruning. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, USA, 2017: 6071-6079.
- [14] T. He, Y. Fan, Y. Qian, et al. Reshaping deep neural network for fast decoding by node-pruning. 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Florence, Italy, 2014:245-249.
- [15] J. Choi, H. Seo, S. Im, et al. Attention routing between capsules. 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), Seoul, Korea, 2019:1981-1989.
- [16] S. K. Sahu, P. Kumar, A. P. Singh. Dynamic routing using inter capsule routing protocol between capsules. 2018 UKSim-AMSS 20th International Conference on Computer Modelling and Simulation (UKSim), Cambridge, UK, 2018:1-5.
- [17] B. Gagana, H. A. U. Athri, S. Natarajan. Activation function optimizations for capsule networks. 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Bangalore, India, 2018:1172-1178.
- [18] R. Wang, G. Li, D. Chu. Capsules encoder and capsgan for image inpainting. 2019 International Conference on Artificial Intelligence and Advanced Manufacturing (AIAM), Dublin, Ireland, 2019:325-328.
- [19] H. Zhang et al. Capsule dynamic network-based object detection algorithm. 2019 IEEE Symposium Series on Computational Intelligence (SSCI), Xiamen, China, 2019:2813-2818.
- [20] C. Xiang, L. Zhang, Y. Tang, et al. MS-CapsNet: a novel multi-scale capsule network. IEEE Signal Processing Letters, 2018, 25(12):1850-1854.
- [21] J. Hsu, C. Kuo, D. Chen. Image super-resolution using capsule neural networks. IEEE Access, 2020, 8:9751-9759.
- [22] Y. LeCun, C. Cortes, C. J. C. Burges. The mnist database of handwritten digits. <http://yann.lecun.com/exdb/mnist/>, 1998.2.2.
- [23] A. Krizhevsky, G. Hinton. Learning multiple layers of features from tiny images. Handbook of Systemic Autoimmune Diseases, 2009, 1(4):45-53 .
- [24] H. Xiao, K. Rasul, R. Vollgraf. Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. CoRR, <http://arxiv.org/abs/1708.07747>, 2017.8.13
- [25] P. Q. Nair, R. Doshi et al. Keselj. Pushing the limits of capsule networks. CoRR, <https://arxiv.org/abs/2103.08074>, 2021.3.23
- [26] A. Deliege, A. Cioppa, M. Van Droogenbroeck. Hitnet: a neural network with capsules embedded in a hit-or-miss layer, extended with hybrid data augmentation and ghost capsules. CoRR, <http://arxiv.org/abs/1806.06519>, 2018.8.13.
- [27] Xiang C, Zhang L, Tang Y. et al. MS-CapsNet: A novel multi-scale capsule network. IEEE Signal Processing Letters, 2018, 25(12):1850-1854.
- [28] V. M. d. Rosario, M. Breternitz, E. Borin. Efficiency and scalability of multi-lane capsule networks (MLCN)//2019 31st International Symposium on Computer Architecture and High Performance Computing (SBAC-PAD), Campo Grande, Brazil, 2019:152-159.



ZHENG Xiang-Ping, Ph.D. candidate. His research interests include image processing and machine learning.

LIANG Xun, Ph. D., professor, Ph.D. supervisor. His research interests include internet information analysis, data mining, business intelligence, and social computing.

Background

In recent years, Convolutional Neural Network (CNN) has achieved breakthrough results in many computer vision tasks. But in essence, it only understands the characteristics of objects too deeply, and ignores the overall spatial structure. This is due to the adoption of pooling strategies. The capsule network has a good ability to fit features, which is a new neural network architecture in deep learning. The pruned capsule network proposed in this paper utilizes two key ideas: pruned network and deep network. By superimposing features into vectors to represent information such as direction, position, etc., the most useful and active low-level capsules can be effectively selected during the training process to build high-level capsules, and those that have less impact or low activity Pruning off allows the network to build a more level high-level network.

Since the introduction of the capsule neural network, domestic and foreign scholars have carried out in-depth research on it. They changed the structure of the capsule network to increase its robustness. Our work explores the possibility of a deeper network composed of multiple capsule layers. The pruned network can effectively achieve the

pruning of low-level unrelated capsules, making the network more simplified; and deep networks process more complex data and obtain more data features. The performance of the pruned capsule network was evaluated on several benchmark data sets. Our pruned capsule network has improved the performance of the network to a certain extent.

This paper is supported by the Nation Natural Science Foundation of China under Grant Nos. 71531012, 62072463, 71271211, 71601013; the Natural Science Foundation of Beijing under Grant Nos. 4172032, 4174087.