

多分支 RA 胶囊网络及在图像分类中的应用

武霖, 孙静宇

引用本文

武霖, 孙静宇. 多分支 RA 胶囊网络及在图像分类中的应用[J]. 计算机科学, 2022, 49(6): 224-230.

WU Lin, SUN Jing-yu. Multi-branch RA Capsule Network and Its Application in Image Classification[J].

Computer Science, 2022, 49(6): 224-230.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

基于注意力机制和门控网络相结合的混合推荐系统

Hybrid Recommender System Based on Attention Mechanisms and Gating Network

计算机科学, 2022, 49(6): 158-164. <https://doi.org/10.11896/jsjcx.210500013>

基于解耦-检索-生成的图像风格化描述生成模型

Stylized Image Captioning Model Based on Disentangle-Retrieve-Generate

计算机科学, 2022, 49(6): 180-186. <https://doi.org/10.11896/jsjcx.211100129>

基于特征注意力融合网络的遥感变化检测研究

Remote Sensing Change Detection Based on Feature Fusion and Attention Network

计算机科学, 2022, 49(6): 193-198. <https://doi.org/10.11896/jsjcx.210500058>

基于样本分布损失的图像多标签分类研究

Study on Multi-label Image Classification Based on Sample Distribution Loss

计算机科学, 2022, 49(6): 210-216. <https://doi.org/10.11896/jsjcx.210300267>

基于时空图卷积和注意力模型的航拍暴力行为识别

Aerial Violence Recognition Based on Spatial-Temporal Graph Convolutional Networks and Attention Model

计算机科学, 2022, 49(6): 254-261. <https://doi.org/10.11896/jsjcx.210400272>

多分支 RA 胶囊网络及在图像分类中的应用

武霖 孙静宇

太原理工大学软件学院 太原 030024

(2669281495@qq.com)

摘要 胶囊网络是一种新型深度神经网络,采用向量表达图像特征信息,并通过引入动态路由算法解决了卷积神经网络的两个主要问题:1)无法对图像的部分与整体关系进行学习和表达;2)池化操作导致图像特征信息严重丢失。然而,CapsNet 需要学习图像的所有特征,当图像背景较复杂时,其存在提取图像特征信息不足、训练参数量大和训练效率低等问题。为此,首先设计了一种轻量级的图像特征提取器 RA 模块,用于更快、更完整地提取图像特征信息;其次,设计了两种不同深度的轻量化分支来提升网络的训练效率;最后,设计了新的压缩函数 hc-squash 来确保网络能够获取更多有用信息,并提出了多分支 RA 胶囊网络。通过在 MNIST, Fashion-MNIST, affNIST 和 CIFAR-10 这 4 个图像分类数据集中的应用,证实了多分支 RA 胶囊网络在多项性能指标上优于 CapsNet 和 MLCN,并针对所提网络设计了改进方案,以优化分类性能。

关键词: 胶囊网络; RA 模块; 深度学习; 压缩函数; 注意力机制

中图法分类号 TP391.41

Multi-branch RA Capsule Network and Its Application in Image Classification

WU Lin and SUN Jing-yu

College of Software, Taiyuan University of Technology, Taiyuan 030024, China

Abstract Capsule Network is a new type of deep neural network that uses vectors to express information of image feature and overcomes two major problems of convolutional neural networks by introducing dynamic routing algorithms. First, convolutional neural networks cannot learn and express the part-whole relationship of images. Second, pooling operations lead to serious loss of image feature information. However, CapsNet needs to learn all the features of the image, and when the image background is complex, it has the problems of insufficient information of extracted image features, large number of training parameters and low training efficiency. To this end, firstly, a lightweight image feature extractor RA module is designed to extract image feature information faster and more completely. Secondly, two different depths of lightweight branches are designed to improve the training efficiency of the network. Finally, a new compression function hc-squash is designed to ensure that the network can acquire more useful information, and a multi-branch RA (Resnet Attention) capsule network is proposed. Through the application in the four image classification datasets of MNIST, Fashion-MNIST, affNIST and CIFAR-10, it is confirmed that the multi-branch RA capsule network outperforms CapsNet and MLCN in several performance metrics, and an improvement scheme is designed for the proposed network to achieve optimised classification performance.

Keywords Capsule network, Resnet attention module, Deep learning, Squash function, Attention mechanism

1 引言

近年来,卷积神经网络(Convolutional Neural Network, CNN)已经在计算机视觉任务中取得巨大的成功(如图像分类^[1]、目标检测^[2]和实例分割^[3])。CNN 在图像特征提取的过程中,为了减少训练参数量和降低图像特征信息的维度,引入了池化操作,但池化操作不仅导致部分图像特征信息丢失,还导致 CNN 无法对图像中整体与部分的相对位置关系进行学习,降低了 CNN 的分类性能。为了弥补这些缺陷, Sabour 等^[4]提出了胶囊网络(Capsule Network, CapsNet),它抛弃了 CNN 中的池化操作,采用胶囊来存储图像特征信息,并利用

动态路由算法来获取图像的整体与部分关系。虽然 CapsNet 具有强大的表征能力,但是其仍存在许多缺陷。首先, CapsNet 的训练参数量过大,导致其训练时间较长和训练效率低;其次, CapsNet 的卷积层较少,提取图像特征的能力不足,使得 CapsNet 无法在复杂图像分类任务中获得较高的精确度。因此,许多学者对 CapsNet 展开了研究和改进。Hinton 等^[5]提出了矩阵胶囊网络,主要通过创建 EM 路由算法和矩阵胶囊来提升网络性能。Xiang 等^[6]提出了多尺度胶囊网络(Multi-scale Capsule Network, MS-CapsNet),该网络主要利用多尺度特征提取的方式来获取图像特征信息,再将特征信息编码到 CapsNet 中进行分类。Nguyen 等^[7]通过将 VGG16

和 CapsNet 相结合的方式,有效地抑制了网络的过拟合现象,并提高了分类精确度。Vanderson 等^[8]提出了多车道胶囊网络(Multi-lane Capsule Network, MLCN),其利用多个不同的平行通道进行图像特征学习,使得网络的分类性能得到提升。Xiong 等^[9]提出了 Deeper-Cap,其通过引入卷积胶囊层和创建胶囊池来达到减少训练参数数量和保留特征信息完整的目的,使得网络的分类精确度得到提升。Han 等^[10]提出了一种基于特征和空间关系的新型胶囊网络(Feature and Spatial relationship coding Capsule Network, FSc-CapsNet),其利用特征提取器和空间关系来提取特征信息和空间位置信息,并对两种信息进行组合学习,从而获得较高的分类精确度。Chang 等^[11]提出了一种多车道严格挤压得胶囊网络(Multi-lane Strict Squash Capsule Network, MLSCN),并设计了一种严格压缩函数来替代原有的压缩函数,使得 MLSCN 的性能优于 CapsNet。

当前,对胶囊网络的改进主要集中在两方面:1)在图像特征提取层中通过堆叠卷积层来提取大量图像特征信息,这种方式不但造成了训练参数数量的增大,也造成了提取特征信息的冗余,严重影响了胶囊网络的学习效率;2)对动态路由算法进行改进,虽然改进的算法能够一定程度地提升胶囊网络的学习能力,但其复杂的计算过程严重影响了胶囊网络的训练效率,延长了训练时间。为了有效地弥补这些缺陷,本文首先设计了一种轻量级的图像特征提取器 RA(Resnet Attention)模块,用于高效地提取重要图像特征信息和有效地解决图像特征信息冗余的问题;其次,引入了两种深度不同的轻量化分支结构,用于增加网络的宽度、深度以及实现网络训练参数数量的轻量化;然后,将设计的新压缩函数 hc-sauqsh 加入到主胶囊层和动态路由算法中,使得网络能够在不增加算法复杂度的前提下保留更多有用的图像特征信息,进而提升训练效率和在各种分类任务中的适用性;最后,通过设计注意力胶囊层和混合压缩函数来进一步优化多分支 RA 胶囊网络的分类性能。

2 相关工作

2.1 残差网络

残差网络以 LSTM^[12]的门机制和 Rupesh 等^[13]提出的高速公路网络(Highway Network)为基础,通过短路连接将浅层特征 x_i 直接跳跃传递给任意深度的网络层 x_L ,其结构如图 1 所示。

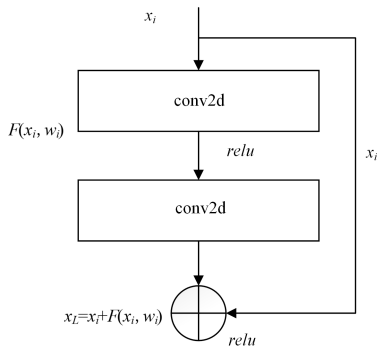


图 1 残差模块示意图

Fig. 1 Schematic diagram of residual module

其中, $F(x_i, w_i)$ 是残差模块通过残差映射学习到的残差图像特征, $x_i + F(x_i, w_i)$ 是残差模块期望学习到的图像特征。可见,残差网络能够通过学习残差图像特征来降低网络的学习难度和提升网络的训练效率。

另外,残差网络主要利用链式求导法解决了 CNN 的退化现象,使得网络具有较好的拟合性。

$$\begin{aligned} \frac{\partial \epsilon}{\partial x_L} &= \frac{\partial \epsilon}{\partial x_L} \cdot \frac{\partial x_L}{\partial x_L} \\ &= \frac{\partial \epsilon}{\partial x_L} \cdot \left(1 + \frac{\partial \sum_{i=1}^{L-1} F(x_i, w_i)}{\partial x_L} \right) \end{aligned} \quad (1)$$

其中, ϵ 是残差网络的损失; $\frac{\partial \epsilon}{\partial x_L}$ 由两部分组成,一部分是未经加权层的 $\frac{\partial \epsilon}{\partial x_L}$,另一部分是经过加权层的 $\frac{\partial \epsilon}{\partial x_L} \left(\frac{\partial \sum_{i=1}^{L-1} F(x_i, w_i)}{\partial x_L} \right)$,这两部分通过线性连接的方式将信息直接反向传播到浅层。

2.2 图像注意力机制 SENet

图像注意力机制 SENet 是一种通道域软注意力机制,其通过显式地建立特征通道之间的相互依赖关系,自动获取每个特征通道的被关注程度,再依照关注程度保留重要特征,并对分类任务作用较低的特征进行抑制。

SENet 包括 $Ftr(\cdot, \theta)$, $Fsq(\cdot)$, $Fscale(\cdot)$ 和 $Fex(\cdot, w)$ 操作,如图 2 所示。

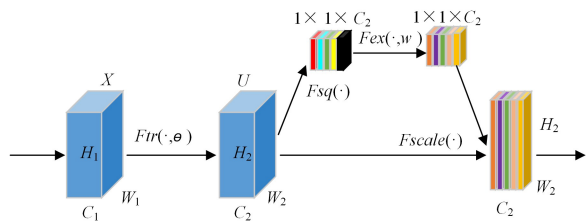


图 2 图像注意力机制 SENet 的结构图

Fig. 2 Structure diagram of image attention mechanism SENet

图 2 中, $Ftr(\cdot, \theta)$ 是卷积操作,用于提取图像特征信息; $Fsq(\cdot)$ 是 Squeeze 操作,其通过全局平均池化操作(global average pooling)将通道特征压缩成全局特征,使得网络能够通过全局感受野来获取全局特征信息; $Fex(\cdot, w)$ 是 Excitation 操作,用于获取特征通道间的依赖性并对获取的全局特征信息重标定,增强重要的特征,减弱不重要的特征,使提取的特征指向性更强; $Fscale(\cdot)$ 用于获取最终的图像特征。可以看出,SENet 能够利用图像特征通道的权重系数来对不同图像特征进行针对性的学习。

2.3 胶囊网络

Hinton 等^[14]于 2011 年首次提出了胶囊网络的概念, Sa-bour 等^[4]于 2017 年对其进行了详细阐述。胶囊网络由多层胶囊构成,每个胶囊是图像中目标实体属性(姿态、形状、颜色)的集合,其通过向量的形式独立存在,以长度代表目标实体存在的概率,以方向代表目标实体的姿态参数。同时,胶囊具有等变性,其能够通过输出长度相同但方向不同的向量对发生姿态变化的目标实体进行区识别。处于底层的胶囊被称为向量胶囊,其利用神经元对图像不同区域的属性进行学习,并将这些属性以向量的形式进行封装。处于

高层的胶囊被称为路由胶囊,其是由向量胶囊组成的,每个路由胶囊都存储着不同图像的整体特征信息和类别信息。

胶囊网络的结构如图 3 所示,其由卷积层、主胶囊层 (PrimaryCaps) 和数字胶囊层 (DigitCaps) 构成。

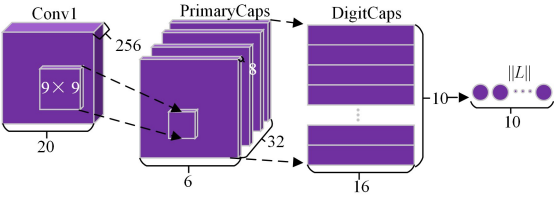


图 3 胶囊网络的结构图

Fig. 3 Structure diagram of capsule network

胶囊网络主要有两个重点,首先是压缩函数(squash),其能够将胶囊的长度压缩至[0,1]之间,用于表示目标实体存在的概率:

$$V_j = \frac{\|s_j\|^2}{1 + \|s_j\|^2} \cdot \frac{s_j}{\|s_j\|} \tag{2}$$

其中, V_j 是输出的胶囊; s_j 是输入的胶囊; $\frac{s_j}{\|s_j\|}$ 表示将 s_j 压缩至单位长度; $\frac{\|s_j\|^2}{1 + \|s_j\|^2}$ 表示对不同长度的 s_j 进行适量压缩。

其次是动态路由算法,如式(3)一式(6)所示:

$$c_i \leftarrow \text{softmax}(b_i) \tag{3}$$

其中, b_i 是向量胶囊 i 指向所有路由胶囊的可能性,其初始值是0; c_i 是向量胶囊 i 指向所有路由胶囊的概率。

$$s_j \leftarrow \sum_i c_{ij} \hat{u}_{j|i} \tag{4}$$

其中, $\hat{u}_{j|i}$ 是向量胶囊 i 对路由胶囊 j 的预测; c_{ij} 是向量胶囊 i 指向路由胶囊 j 的概率; s_j 是所有向量胶囊对路由胶囊 j 预测的总和。

$$v_j \leftarrow \text{squash}(s_j) \tag{5}$$

其中, v_j 是 s_j 经过squash()后的预测总和。

$$b_{ij} \leftarrow b_{ij} + \hat{u}_{j|i} \cdot v_j \tag{6}$$

其中, $\hat{u}_{j|i} \cdot v_j$ 是点积运算,用于判断 $\hat{u}_{j|i}$ 和 v_j 的相似度, $\hat{u}_{j|i}$ 和 v_j 越相似,二者的点积越大,向量胶囊 i 指向路由胶囊 j 的可能性就越大,反之亦然。

可见,胶囊网络能够通过动态路由算法将向量胶囊转为路由胶囊,使得网络能够学习图像的整体与部分关系。

3 多分支 RA 胶囊网络及其改进

3.1 多分支 RA 胶囊网络

多分支 RA 胶囊网络由两种深度不同的分支并行组成,如图 4 所示。首先,将图像输入一个卷积核大小为 5×5 的卷积层中进行浅层图像特征信息提取,再将结果并行输入两种分支的 RA 模块中进行深度图像特征信息提取;其次,将每种分支提取的图像特征信息分别输入各自的主胶囊层中,利用hc-squash 压缩函数将提取的图像特征信息编码成向量胶囊;然后,通过改进的动态路由算法将向量胶囊转为路由胶囊,并将每种分支中的路由胶囊分别堆叠成数字胶囊矩阵;最后,将两种分支的数字胶囊矩阵组合成一个数字胶囊矩阵,并利用欧氏距离公式计算组合后的数字胶囊矩阵中所有路由胶囊的长度,选取长度最长的路由胶囊所对应的种类作为网络最终预测的分类结果。

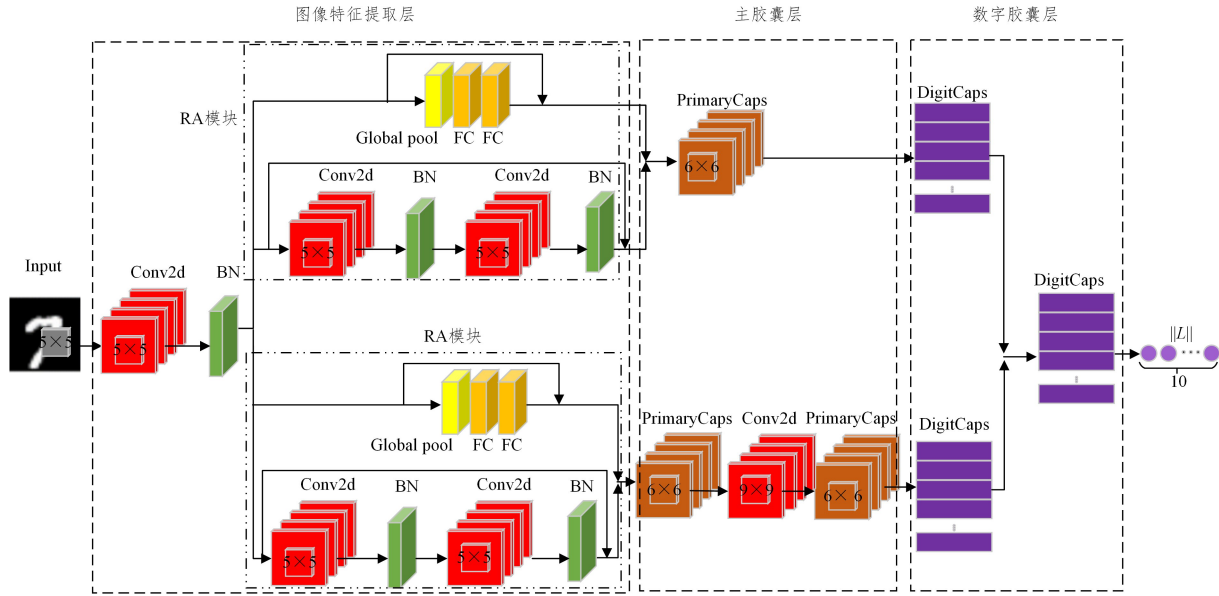


图 4 多分支 RA 胶囊网络的结构图

Fig. 4 Structure diagram of multi-branch RA capsule network

3.1.1 RA 模块

在图像特征提取层中,RA 模块由 Hu 等^[15]提出的图像注意力机制 SENet(Squeeze and Excitation Network)和 He 等^[16]提出的残差网络并行组成,其具体过程如式(7)一式(12)所示:

$$x_{res} = x_0 + \sum_{i=0}^l F(x_i, w_i) \tag{7}$$

其中, c 是通道数; x_0 是输入的图像特征; $F(x_i, w_i)$ 是残差映射获得的特征; x_{res} 是残差网络获得的图像特征。

$$u_c = v_c \cdot x_0 = \sum_{s=1}^c v_c^s x^s \tag{8}$$

其中, v_c 是一个 3D 卷积核; u_c 是通道的综合特征矩阵; v_c^s 是第 s 个特征矩阵对应的卷积核; x^s 是第 s 个特征矩阵的输入。

$$z_c = F_{sq}(u_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W u_c(i, j), z \in R^c \tag{9}$$

其中, z_c 是 c 个特征矩阵压缩得到的实数; $w_1 \in R, w_2 \in R$ 是两个全连接层的参数。

$$s = Fex(z, w) = sigmoid(w_2 Relu(w_1 z)) \tag{10}$$

其中, z 是全局特征信息; $sigmoid()$ 和 $Relu()$ 是激活函数; s 是 c 个特征矩阵的权重。

$$\tilde{x}_c = Fscale(u_c \cdot s_c) = s_c \cdot u_c \tag{11}$$

其中, $\tilde{x}_c$ 是 SENet 最终获取的特征; s_c 是权重 s 。

$$x_{sum} = x_{res} + \tilde{x}_c \tag{12}$$

其中, x_{sum} 是 RA 模块获取的图像特征。

由于残差网络和 SENet 提取的图像特征不同, 因此通过并行堆叠图像特征的方式能够获得更加完整的、重要的图像特征, 如图 5 所示。

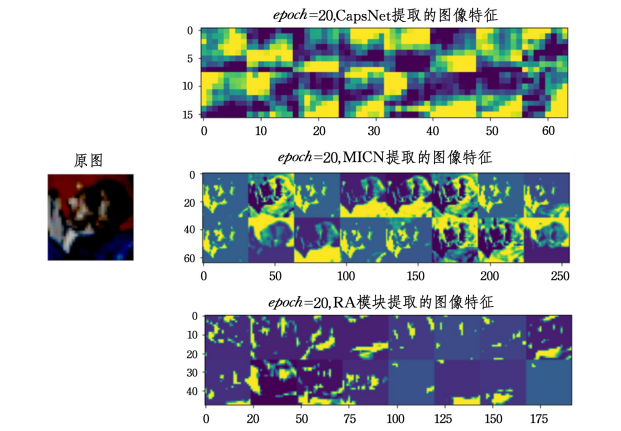


图 5 图像特征提取示意图
Fig. 5 Schematic diagram of image feature extraction

从图 5 中可以看出, 相同的 $epoch$ 下, CapsNet 和 MLCN 提取的图像特征基本是图像的全局信息, 而 RA 模块提取的图像特征更加着重于图像中的关键区域, 更加抽象化。这样不仅有效地避免了图像特征信息冗余的问题, 且更加有利于网络对图像特征的学习。

另外, RA 模块作为轻量级的图像特征提取器, 其与 CapsNet, MLCN 图像特征提取器在 CIFAR-10 数据集上训练参数数量的对比如表 1 所列。

表 1 不同图像特征提取器的参数量
Table 1 Parameter scale of different image feature extractors

	CapsNet 图像 特征提取器	MLCN 图像 特征提取器	RA 模块
训练参数量	5 372 160	1 420 519	747 008

可见, RA 模块不仅能高效地提取重要的图像特征信息, 还为分支轻量化打下了坚实的参数基础。

3.1.2 轻量化的分支

为了提升网络的分类性能, 本文引入了两种轻量化的分支。在第一种分支中, 图像特征提取层主要由 RA 模块组成, 主胶囊层由卷积核大小为 6×6 的卷积胶囊层组成, 数字胶囊

层由数字胶囊矩阵组成。第二种分支是在第一种分支的基础上, 在主胶囊层增加了卷积核大小为 9×9 的卷积层和卷积核大小为 6×6 的卷积胶囊层。

两种分支在 MNIST 数据集集中的训练参数数量如表 2 所列。

表 2 训练参数数量的对比
Table 2 Comparison of training parameter amount

模型/分支结构	训练参数量
CapsNet	8 384 929
MLCN	2 849 217
分支 1	708 288
分支 2	1 195 968

从表 2 中可知, 这种并行的轻量化分支结构不仅能够增加网络的宽度和深度, 还能够为网络的高效训练与轻量化提供重要的结构支撑。

3.1.3 hc-squash 压缩函数

压缩函数主要通过给长胶囊赋予高概率值和给短胶囊赋予低概率值的方式来保留有用的图像特征信息和抑制无用的图像特征信息。当分类任务以短胶囊居多时, squash 压缩函数无法保留足够有用的图像特征信息, 使得网络的分类性能受到限制。为此, 本文设计了 hc-squash 压缩函数来代替主胶囊层和动态路由算法中的 squash 压缩函数。

$$V_j = \frac{e^{\|u_j\|} - 1}{e^{\|u_j\|} + 1} \cdot \frac{u_j}{\|u_j\|} \tag{13}$$

其中, u_j 是 squash 向量; V_j 是 hc-squash 输出的胶囊; $\frac{e^{\|u_j\|} - 1}{e^{\|u_j\|} + 1}$ 表示对不同长度的 u_j 进行适量压缩。

squash 压缩函数与 hc-squash 压缩函数的对比如图 6 所示。

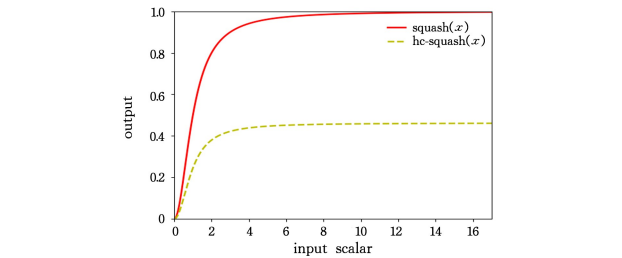


图 6 压缩函数的对比
Fig. 6 Comparison of compression function

从图 6 中可知, hc-squash 压缩函数不仅能够扩大高概率值所对应的胶囊长度范围, 为网络保留更多的长胶囊, 还能够对低概率的短胶囊进行一定程度的抑制, 使得网络不仅能够学习到更多有用的图像特征信息, 还减少了无用信息对训练的干扰, 进而提高了网络在各种分类任务中的适用性。

3.2 改进的多分支 RA 胶囊网络

虽然多分支 RA 胶囊网络的性能优于胶囊网络和 MLCN, 但是其网络深度仍然较浅, 在处理复杂图像分类任务时仍无法达到较高的分类精确度。同时, 由于 hc-squash 在主胶囊层和动态路由算法中均保留了大量胶囊, 导致网络在训练过程中学习了过多重复图像特征信息, 严重影响了网络的学习效率。因此, 针对以上问题, 本文对多分支 RA 胶囊网络

进行了改进。首先,将 SENet 加入主胶囊层中组成注意力胶囊层,用于将图像中的重要特征进一步标识出来,使得注意力胶囊层中的向量胶囊能够存储更多重要的图像特征信息,同时注意力胶囊层还能够加深网络深度,提升网络性能。其次,引入 Yang 等^[17]提出的 PA(Powered-Activation)压缩函数并将其与 hc-squash 组成混合压缩函数,其中 PA 压缩函数代替了主胶囊层中的 hc-squash。PA-squash 能够扩大低概率差值对应的胶囊区间,有效抑制较短的胶囊,极大程度地减小了无用信息对网络训练的影响。因此,使用混合压缩函数能够有效抑制主胶囊层中的无用信息,也能在动态路由中极大地保留有用信息。PA-squash 如式(14)、式(15)所示:

$$Power_n = \| \mathbf{x} \| ^n \cdot \frac{\mathbf{x}}{\| \mathbf{x} \| }$$

(14)

$$\hat{\mathbf{u}}_j = Power_n(\mathbf{u}_j)$$

(15)

其中, $\mathbf{u}_j$ 是 squash 向量; $\mathbf{x}$ 是输入的向量; $\hat{\mathbf{u}}_j$ 是 $Power_n()$ 压缩输出的胶囊。

squash 压缩函数、hc-squash 压缩函数和 PA-squash 压缩函数的对比结果如图 7 所示。

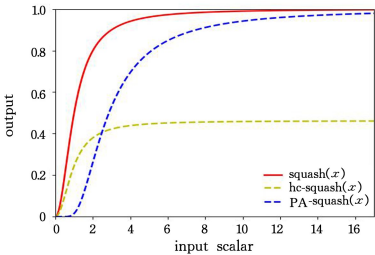


图 7 3 种压缩函数的对比图

Fig. 7 Comparison of three compression functions

4 实验及结果分析

本文将 CapsNet、MLCN、多分支 RA 胶囊网络和改进的

多分支 RA 胶囊网络在 4 个图像分类数据集 MNIST, Fashion-MNIST, affNIST 和 CIFAR-10 上进行对比实验。在网络中初始学习率为 0.0003,并采用指数衰减,衰减率为 0.95;梯度下降采用 Adam 优化器;引入早停机制(Early Stopping)和 Dropout 机制来防止过拟合现象。

4.1 实验数据集简介

本文采用的 MNIST 数据集由包含了 10 种手写数字的灰度图像组成;Fashion-MNIST 数据集由包含了 10 种服饰的灰度图像组成;affNIST 数据集由 MNIST 数据集经过 32 次仿射变换得到,本文选取第 31 次仿射变换后的 affNIST 作为数据集,并利用随机水平翻转来增加数据集的多样性;CIFAR-10 数据集由 10 种物体的 RGB 图像组成。以上 4 个公共数据集均包含一个训练集和一个测试集,数据集信息如表 3 所列。

表 3 数据集信息

Table 3 Dataset information

数据集	训练集图像数	测试集图像数	图像大小
MNIST	60 000	10 000	28×28
Fashion-MNIST	60 000	10 000	28×28
affNIST	60 000	10 000	40×40
CIFAR-10	50 000	10 000	32×32

4.2 网络性能对比与分析

为了对 CapsNet、MLCN、多分支 RA 胶囊网络和改进的多分支 RA 胶囊网络的分类性能进行评估,本文采用多个不同的评价指标对 4 种网络进行全面对比。表 4 列出了 CapsNet、MLCN、多分支 RA 胶囊网络和改进的多分支 RA 胶囊网络在 MNIST, Fashion-MNIST, affNIST 和 CIFAR-10 数据集上进行对比的结果,其中,M1 代表 CapsNet,M2 代表 MLCN,M3 代表多分支 RA 胶囊网络,M4 代表改进的多分支 RA 胶囊网络。

表 4 对比实验的结果

Table 4 Results of comparative experiments

	训练参数	训练收敛轮次	训练时间/ms	收敛精度/%	收敛损失值	评估精度/%	评估时长/s
MNIST	M1 8384929	29	155	99.32	0.2296	99.55	12
	M2 2849217	27	111	99.98	0.2320	99.49	10
	M3 1904256	22	75	99.98	0.0609	99.54	6
	M4 2003328	18	79	99.97	0.0620	99.70	7
Fashion-MNIST	M1 8384929	49	151	87.40	0.2414	88.94	11
	M2 2849217	24	108	90.40	0.2407	91.88	9
	M3 1904256	37	75	90.46	0.1084	93.35	6
	M4 2003328	40	79	99.92	0.0644	94.21	6
affNIST	M1 17320705	24	665	91.67	0.3223	96.37	14
	M2 3885297	13	126	99.96	0.2611	97.11	12
	M3 2962736	11	137	99.84	0.0625	97.73	8
	M4 2791808	5	122	99.33	0.0900	99.37	9
CIFAR-10	M1 10745251	37	271	64.89	0.3118	68.19	12
	M2 5250865	15	123	70.26	0.3148	70.18	9
	M3 2125696	17	83	80.59	0.2028	81.33	7
	M4 2224768	11	88	85.64	0.1076	82.10	7

由表 4 可知,多分支 RA 胶囊网络和改进的多分支 RA 胶囊网络在 4 个公共数据集上的多项评价指标均优于 CapsNet

和 MLCN,说明二者的学习能力和泛化能力更强。

另外,在 4 个公共数据集的测试集上对 CapsNet, ML-

CN、多分支 RA 胶囊网络和改进的多分支 RA 胶囊网络的分类性能进行测试,如图 8—图 11 所示。

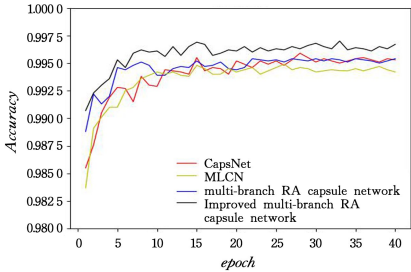


图 8 在 MNIST 测试集上的精确度对比

Fig. 8 Accuracy comparison on MNIST testset

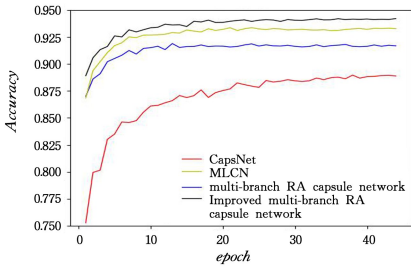


图 9 在 Fashion-MNIST 测试集上的精确度对比

Fig. 9 Accuracy comparison on Fashion-MNIST testset

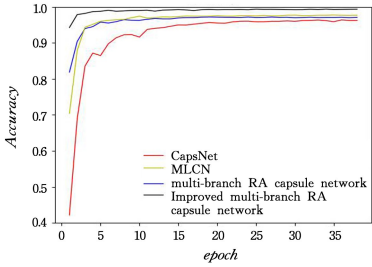


图 10 在 affNIST 测试集上的精确度对比

Fig. 10 Accuracy comparison on affNIST testset

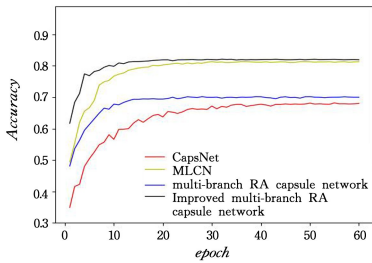


图 11 在 CIFAR-10 测试集上的精确度对比

Fig. 11 Accuracy comparison on CIFAR-10 testset

从图中可以看出,多分支 RA 胶囊网络和改进的多分支 RA 胶囊网络的起始精确度较高,说明二者具有较高的学习效率,对各种分类任务的适用性较强;另外,在图 9 和图 11 中,多分支 RA 胶囊网络和改进的多分支 RA 胶囊网络的收敛精确度均高于 CapsNet 和 MLCN,且二者仅需较少的 epoch 即可收敛至较高精确度,说明二者具有更强的泛化能力。

为了更加细致地对比多分支 RA 胶囊网络和改进的多分支 RA 胶囊网络的分类性能差异,本文引入了混淆矩阵差异分析表,如表 5 所列。

表 5 混淆矩阵差异分析

Table 5 Analysis of confusion matrix difference

	多分支 RA 胶囊网络	改进的多分支 RA 胶囊网络
MNIST	对种类 5,6,7,8 均有 1% 未被正确识别	只对 1% 的种类 6 和 9 未正确识别
Fashion-MNIST	对种类 0 和 6 的分类精确度误差较大,分别为 11%, 22%	主要对种类 0 和 6 的分类精确度分别提升了 2% 和 4%,其余种类均提升 1%
affNIST	对种类 2,5 和 6 的分类精确度误差较大,分别为 6%, 5% 和 3%	主要对种类 2,5 的分类精确度分别提升了 6% 和 4%,其余种类均提升 1%
CIFAR-10	对种类 2,3,4 和 5 的分类精确度误差较大,分别为 30%, 33%, 22% 和 27%	主要对种类 2,4,7 和 8 的分类精确度分别提升 2%, 7%, 3% 和 2%

5 与其他网络的对比

本文还与前文提及的 MS-CapsNet, Deeper-Cap, FSc-CapsNet 在 MNIST, Fashion-MNIST 和 CIFAR-10 数据集上进行了精确度的对比,结果如表 6 所列。

表 6 与其他网络的对比结果

Table 6 Comparative test with other networks

	(单位:%)		
	MNIST	Fashion-MNIST	CIFAR-10
MS-CapsNet	—	92.70	75.70
Deeper-Cap	99.84	—	81.29
FSc-CapsNet	—	94.01	80.03
多分支 RA 胶囊网络	99.54	93.35	81.33
改进的多分支 RA 胶囊网络	99.70	94.21	82.10

从表 6 中可知,多分支 RA 胶囊网络和改进的多分支 RA 胶囊网络在 MNIST 数据集上的精确度仍与 DeeperCap 存在差距,但在其他数据集上的精确度均高于其他网络,说明多分支 RA 胶囊网络和改进的多分支 RA 胶囊网络均具有较强的性能。

结束语 本文提出了多分支 RA 胶囊网络,首先,利用残差网络结构和图像注意力机制 SENet 并行组成一种轻量级的图像特征提取器 RA 模块,用于对图像特征进行提取,并引入两种不同深度的轻量化分支结构,使得网络能够提升训练效率和实现训练参数轻量化;其次,设计新的压缩函数 hc-squash,用于代替 squash 压缩函数,以保留更多图像特征信息,使网络能够充分学习图像内容;最后,在 4 个图像分类数据集上对 CapsNet、MLCN、多分支 RA 胶囊网络和改进的多分支 RA 胶囊网络的分类性能进行评估。多分支 RA 胶囊网络虽然取得了较高的精确度,但依然存在一些不足,因此设计了改进方案。首先,将 SENet 加入主胶囊层中组成注意力胶囊层,用于加深网络深度和提取更多有用的图像特征信息;其次,将 PA 压缩函数和 hc-squash 组成混合压缩函数,以实现对不同长度胶囊的兼顾,从而保证网络能够获取足够有用的图像特征信息,同时抑制更多无用的图像特征信息。虽然改进的多分支 RA 胶囊网络的精确度更高,但是其依然存在训练时间较长、训练速度较慢和无法高效处理复杂图像任务等问题。因此,在未来的工作中,将引入多种不同的神经网络结构和更多高效的图像注意力机制来增强图像特征提取器,并在动态路由算法中引入复数和聚类算法的理念来进一步提升网络的分类性能。

参 考 文 献

[1] JIANG J,LIU F,XU Y,et al. Multi-spectral RGB-NIR image classification using double-channel CNN[J]. IEEE Access, 2019,7:20607-20613.

[2] BAE S H. Object detection based on region decomposition and assembly[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2019,33(1):8094-8101.

[3] FAHIM RAHMAN A K M,RAIHAN M R,MOHIDUL ISLAM S M. Pedestrian Detection in Thermal Images Using Deep Saliency Map and Instance Segmentation[J]. International Journal of Image, Graphics and Signal Processing(IJIGSP), 2021, 13(1):40-49.

[4] SABOUR S,FROSST N,HINTON G E. Dynamic Routing Between Capsules[C]//Advances in Neural Information Processing Systems. 2017:3856-3866.

[5] HINTON G E,SABOUR S,FROSST N. Matrix capsules with EM routing[C]//International Conference on Learning Representations. 2018.

[6] XIANG C,ZHANG L,TANG Y,et al. MS-CapsNet: A novel multi-scale capsule network[J]. IEEE Signal Processing Letters, 2018,25(12):1850-1854.

[7] NGUYEN H H,YAMAGISHI J,ECHIZEN I. Capsule-forensics:Using capsule networks to detect forged images and videos [C]//2019 IEEE International Conference on Acoustics,Speech and Signal Processing (ICASSP 2019). IEEE,2019:2307-2311.

[8] DO ROSARIO V M,BORIN E,JBRETER-NITZ M. The multi-lane capsule network[J]. IEEE Signal Processing letters, 2019, 26(7):1006-1010.

[9] XIONG Y,SU G,YE S,et al. Deeper capsule network for complex data[C]// 2019 International Joint Conference on Neural Networks (IJCNN). IEEE,2019:1-8.

[10] HAN T,SUN R,SHAO F,et al. Feature and spatial relationship coding capsule network[J/OL]. Journal of Electronic Imaging. <https://doi.org/10.1117/1.JEI.29.2.023004>.

[11] CHANG S,LIU J. Multi-lane Capsule Network for classifying images with complex background[J]. IEEE Access, 2020, 8: 79876-79886.

[12] HOCHREITER S,SCHMIDHUBER J. LSTM can solve hard long time lag problems[C]//Advances in Neural Information Processing Systems. 1997:473-479.

[13] SRIVASTAVA R K,SCHMIDHUBER J,GREFF K. Highway Networks[J]. arXiv:1505.00387, 2015.

[14] HINTON G E,KRIZHEVSKY A,WANG S D. Transforming auto-encoders[C]//International Conference on Artificial Neural Networks. Berlin:Springer, 2011:44-51.

[15] HU J,SHEN L,SUN G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018:7132-7141.

[16] HE K,ZHANG X,REN S,et al. Deep Residual Learning for Image Recognition[C]//IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2016:770-778.

[17] YANG Z,WANG X. Reducing the Dilution:analysis of the information sensitiveness of capsule network and one practical solution[J]. arXiv:1903.10588v3, 2019.



WU Lin, born in 1995, postgraduate, is a member of China Computer Federation. His main research interests include image processing and recommendation system.



SUN Jing-yu, born in 1975, Ph.D, associate professor, is a member of China Computer Federation. His main research interests include collaborative web search, recommendation system and smart city.

(责任编辑:喻黎)