动态路由胶囊网络的可视化研究

范文豪[1],吴晓富[1],张索非[2]
(1. 南京邮电大学 通信与信息工程学院,江苏 南京 210003;
2. 南京邮电大学 物联网学院,江苏 南京 210003)

摘 要:最近提出的胶囊网络是继卷积网络(convolutional neural networks,CNN)之后的另一创新结构。相比于 CNN 特征的弱空间关联性,胶囊网络的矢量化特征则被认为能很好地表达特征之间的空间关联。然而,胶囊网络的特征应如何理解还缺乏严格的理论以及实验论证。对此,文中试图从特征可视化的角度来验证 CNN 中特征空间关联性的强弱,并探讨胶囊网络提取到的特征中是否具有空间关联。通过训练不同特征维数的胶囊网络,探究出改变特征维数对胶囊网络产生的影响。实验结果表明,相比于 CNN 特征空间的弱关联性,胶囊网络的矢量化特征呈强相关性,确实包含了所提取特征的姿态、形变等空间相关信息,并且当胶囊网络特征维数降低时提取到的空间信息会减少,使得胶囊网络复原出图像的与输入的原图像差距增大。

关键词:胶囊网络;矢量化;空间信息;可视化
中图分类号:TP391 **文献标识码**:A **文章编号**:1673-629X(2019)08-0071-05
doi:10.3969/j.issn.1673-629X.2019.08.014

Research on Visualization of Capsule Network with Dynamic Routing

FAN Wen-hao[1], WU Xiao-fu[1], ZHANG Suo-fei[2]
(1. School of Communication and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China;
2. School of Internet of Things, Nanjing University of Posts and Telecommunications, Nanjing 210003, China)

Abstract: Recently proposed capsule networks provide an alternative to convolutional neural networks (CNN). Compared with the weak spatial correlation of CNN features, the vector feature of capsule network is considered to be an effective way to express the spatial correlation between features. However, how to understand the characteristics of the capsule network still lacks strict theoretical and experimental supports. Therefore, we attempt to verify the strength of feature space association in CNN and explore whether the features extracted from the capsule network have spatial connections from the perspective of feature visualization. The influence of changing the dimension of features on the capsule network is explored by training the capsule network with different dimensions of features. The experiment shows that compared with the weak correlation of CNN feature space, the vectorization of capsule network is strongly correlated, including spatial correlation information such as attitude and deformation of extracted features. When the dimension of features is reduced, the spatial information extracted is reduced, so that the gap between the image restored by the capsule network and the input original image is increased.

Key words: capsule network; vectorization; spatial information; visualization

0 引 言

传统的卷积神经网络比较擅长检测图像特征[1-2],但由于 CNN 中神经元的输入和输出都是标量,标量只能用来表示所提取到的特征存在的可能性,不能表示特征的空间信息,这就导致 CNN 探索特征间的空间关系能力欠佳[3]。同时 CNN 中的下采样层降低了图像的空间分辨率,一些空间信息就会丢失,使得 CNN 对于输入的一些小变化不敏感。例如当输入图

收稿日期:2018-08-22 **修回日期**:2018-12-16 **网络出版时间**:2019-03-27
基金项目:国家自然科学基金(61372123,61701252)
作者简介:范文豪(1995-),男,硕士研究生,通信作者,研究方向为深度学习与计算机视觉;吴晓富,博士,教授,研究方向为信息论与编码、机器学习与计算机视觉、密码学与信息安全;张索非,博士,讲师,研究方向为图像与视频信号处理、机器学习、物联网技术等。
网络出版地址:http://kns.cnki.net/kcms/detail/61.1450.TP.20190327.1620.008.html

像为人脸时,如果将他的鼻子和嘴巴位置互换,通常 CNN 会误认为这是一张人脸。这正是由于 CNN 特征的弱空间关联性造成的[4]。

为了解决 CNN 的这一缺陷,Sara Sabour 等提出了动态路由的胶囊网络(capsule network)[5]。它与传统 CNN 的最主要区别在于,胶囊网络提出了一种由神经元组成的胶囊结构,它的输入输出都是矢量[6],不仅能够通过矢量的模长来表示某个特征出现的可能性,还能够通过矢量来表示特征的空间信息,包括位置、方向、大小、形变等。这样胶囊网络就能够学习到同一个特征的不同变体,还能很好地表达特征间的空间关联。

文中先对 CNN 中各层提取到的特征进行可视化,验证 CNN 中提取到的特征空间关联性的强弱。之后对胶囊网络中 DigiCaps 层最终提取到的矢量特征的不同维度分别进行改变,将得到特征的不同变体进行可视化[7]。通过可视化方法判断出提取到的矢量化特征是否包含空间信息,并通过不同变体的变化得出具体的空间信息,从而加深对胶囊网络所提取到的矢量化特征的理解。

可视化实验结果表明,CNN 中提取到的特征空间的关联性较弱,而胶囊网络的矢量化特征确实包含了所提取特征的姿态、形变等空间信息[8-10],能够很好地表达特征之间的空间关联。

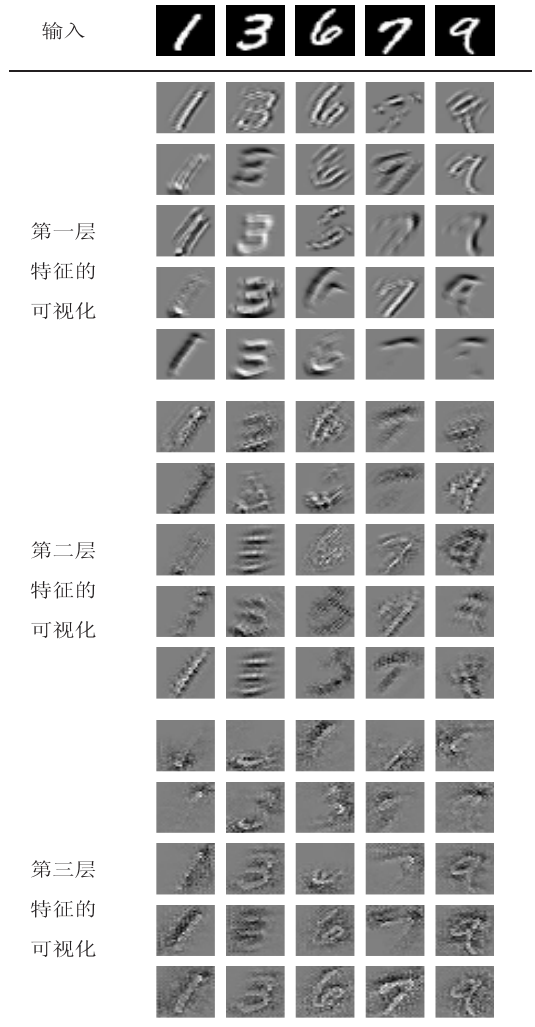
1 CNN 的特征可视化

CNN 主要由卷积层和池化层组成,对于多层的 CNN,每一层所提取到的特征都是不完全相同的。为了验证 CNN 提取到特征的空间关联性的强弱,可以通过反池化和反卷积[11-12]的操作对每一层提取到的多个特征进行可视化,从而观察所提取到的特征间空间关联性强弱。

使用 MNIST 数据集[13-14]训练一个 4 层的 CNN,其中每一层的卷积核尺寸都为 5×5,卷积时所采用的 padding 为 SAME 模式,并在第二层和第四层卷积后加上大小为 2×2 的 max pooling 层。训练完成后对 CNN 每层所提取到的特征进行可视化,可视化结果见表 1。从每层的多个特征中抽取出部分特征的可视化图像放入表中,表中包括输入图像以及 CNN 前 3 层特征的可视化。

从表 1 可以看出:CNN 第一层提取出的特征主要是图像中数字的边缘和轮廓,第二层提取到的特征主要是数字边缘的颜色分布与数字内容,第三层则能够提取到数字一些局部特征以及少量的空间信息。可视化结果表明,普通的 CNN 提取空间信息的能力较弱,并不能够学习到同一个特征的多种不同变体,所以提取到特征的空间关联性也就较弱。

表 1 CNN 特征可视化



2 动态路由胶囊网络的特征可视化

2.1 胶囊网络的结构

胶囊网络中最重要的组成部分是胶囊。胶囊是由神经元组成的一个向量,能表征检测类型的多维实体的实例化参数和存在性。

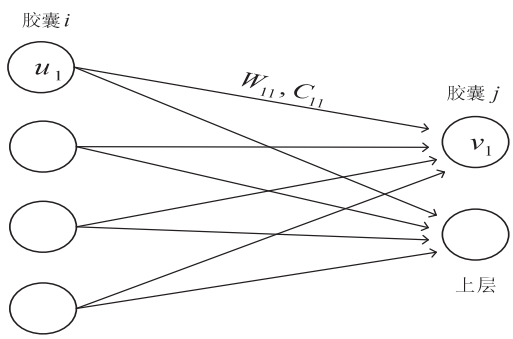


图 1 胶囊之间的转换结构

图 1 是胶囊网络中两层胶囊之间的转换关系,其中胶囊 i 和胶囊 j 都是矢量,u_i 为胶囊 i 的激活值,$\boldsymbol{W}_{ij}$,c_{ij} 为两个胶囊之间的转换矩阵和权重:

$$\hat{\boldsymbol{u}}_{j|i} = \boldsymbol{W}_{ij} u_i$$

$\hat{\boldsymbol{u}}_{j|i}$ 称为低层胶囊 i 对高层胶囊 j 的预测向量,将低层所有胶囊对高层胶囊 j 的预测向量乘权重 c_{ij} 相加得到 $\boldsymbol{s}_j$:

$$\boldsymbol{s}_j = \sum_i c_{ij}\hat{\boldsymbol{u}}_{j|i}$$

通过非线性"压缩"函数(Squash 函数)得到胶囊 j 的激活值 v_j:

$$\boldsymbol{v}_j = \frac{\| \boldsymbol{s}_j \|^2}{1 + \| \boldsymbol{s}_j \|^2} \frac{\boldsymbol{s}_j}{\| \boldsymbol{s}_j \|}$$

胶囊网络中使用动态路由[15]迭代算法,迭代步骤如下:

1. 令: $b_{ij} \leftarrow 0(\ i \in l\ , j \in l + 1)$

2. for r iterations do

$c_i \leftarrow \mathrm{softmax}(b_i)\ (i \in l)$

$\boldsymbol{s}_j \leftarrow \sum_i c_{ij}\hat{\boldsymbol{u}}_{j|i}\ (j \in l + 1)$

$\boldsymbol{v}_j \leftarrow \mathrm{squash}(\boldsymbol{s}_j)\ (j \in l + 1)$

$b_{ij} \leftarrow b_{ij} + \hat{\boldsymbol{u}}_{j|i} \cdot \boldsymbol{v}_j\ ,(\ i \in l\ , j \in l + 1)$

return $\boldsymbol{v}_j$

其中迭代次数 r 一般为 3 次。动态路由可以看成是一个类似于聚类的过程,将低层对高层的预测向量 $\hat{\boldsymbol{u}}_{j|i}$ 与高层胶囊激活值 $\boldsymbol{v}_j$ 进行点乘,如果乘积越大,权重 c_{ij} 就会越大,说明 $\hat{\boldsymbol{u}}_{j|i}$ 在 $\boldsymbol{v}_j$ 中所占比例就越大。动态路由的过程就是将 $\hat{\boldsymbol{u}}_{j|i}$ 进行聚类,聚类的标准是 $\hat{\boldsymbol{u}}_{j|i}$ 与 $\boldsymbol{v}_j$ 点乘之后结果的大小。当预测向量的模长较大或其方向与 $\boldsymbol{v}_j$ 较接近时,两者点乘结果都会偏大。

图 2 为整个胶囊网络的结构。本次实验使用的是 MNIST 数据集,图中第一层为输入层,输入图像的尺寸为 28×28×1;第二层为卷积层,使用卷积核尺寸为9×9×1×256,进行步长为 1 的 VALID 型卷积,这层把像素强度转换成局部特征检测信息;第三层为 PrimaryCaps 层,使用卷积核仍为 9×9×1×256,进行步长为 2 的 VALID 型卷积,将得到结果组合成 32×6×6 个胶囊,每一个胶囊为 8 维的向量;第四层为 DigiCaps 层,共 10 个胶囊,每个胶囊为 16 维的向量,第三层输出经转换矩阵 $\boldsymbol{W}_{ij}$ 和动态路由过程得到第四层输出;最后三层全连接层用来对图像进行复原。

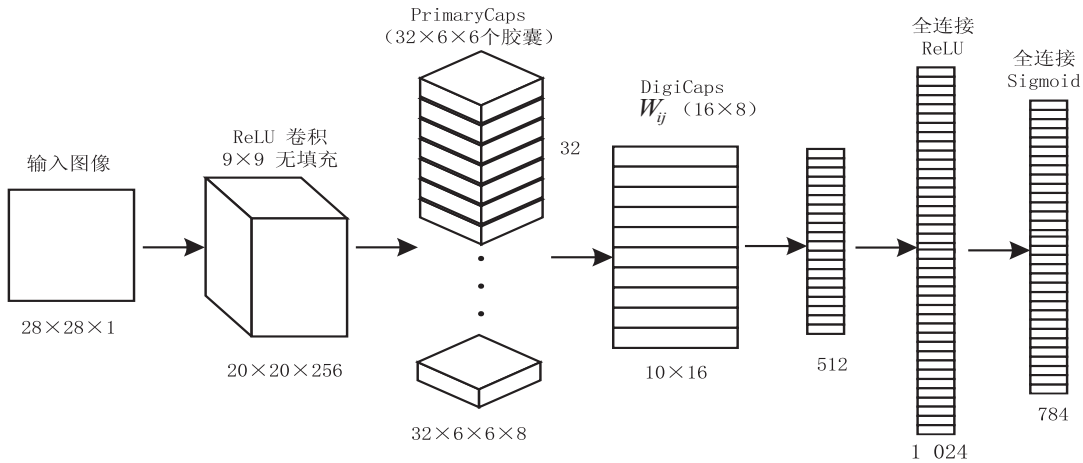


图 2　胶囊网络结构

2.2　胶囊网络特征可视化

实验使用 MNIST 数据集来训练胶囊网络[16-18],数据的预处理过程进行了简单的归一化处理,具体的训练参数见表 2。其中 DigiCaps 层的维度是会随着实验改变的。

表 2　胶囊网络主要训练参数

主要参数	数值
Number of the capsules inPrimaryCaps layer	1 152
Dimension of the vector inPrimaryCaps layer	8
Number of the capsules inDigiCaps layer	10
Dimension of the vector inDigiCaps layer	--
Epoch	300
Batch size	256

本次实验目的是研究出 DigiCaps 层胶囊提取到的矢量化特征所能代表的空间信息。实验的主要内容是改变 DigiCaps 层提取到的特征的某一维大小(其余维不变),按训练好的参数通过全连接层复原图像进行可视化,得到的复原图像会是同一个特征的不同变体,对比多个复原图像判断出所改变的那一维代表的空间信息。然后将 DigiCaps 层中胶囊特征的维数改变做同样的实验。

2.2.1　DigiCaps 层胶囊中矢量化特征维数为 16 的可视化

当 DigiCaps 层胶囊中向量维数为 16 时,按照表 2 中的参数训练图 2 的网络后得到的 loss 值见图 3。经过 300 个 epoch 之后,loss 下降到 0.003 5 左右,最终测试得到的准确率在 99.25% 左右。

参数训练完成后,通过全连接层将输入的图像进行复原。表 3 中包括初始复原图和改变特征某一维大

小得到的复原图。这里的改变某一维大小是指在特征某一维上加或减去一个小于 1 的合适的常数,多次改变大小将得到的复原图像进行对比。表 3 中初始复原图像的右下角数字代表修改的矢量化特征的维度。

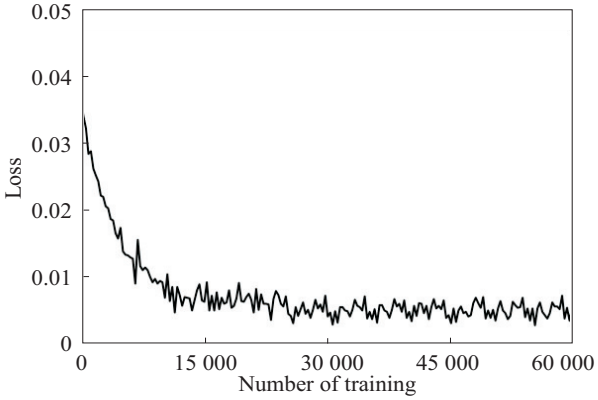


图 3 维数为 16 时网络的 loss

对比表 3 中复原图像可知,当改变特征某一维的大小后,会对数字的大小、旋转、局部形状、笔画的弯曲、粗细等产生影响。例如输入为 0 时改变特征的第 2 维发现数字的大小发生了变化,输入为 1 时改变特征的第 15 维时数字发生了旋转,输入为 2 时改变特征的第 1 维时数字的局部形状发生了改变。说明胶囊网络的矢量化特征是可以表示特征的空间信息的。从可视化结果可以看出,矢量化特征所提取到的空间信息主要是来源于同一类别的训练数据集的多样化。当 DigiCaps 层胶囊中矢量化特征维数为 16 时,对于 MNIST 数据集而言,16 个维度足够提取出输入图像的绝大部分的空间信息,并且对同一类别 16 个维度所表示的空间信息有少部分是类似的。如果同一类别的训练数据集中数字图像的变化不是非常大,那么提取到的特征的部分维度所表示的空间信息有一些会比较相似,但是对于不同类别的输入,提取到的空间信息大多数是不相同的。

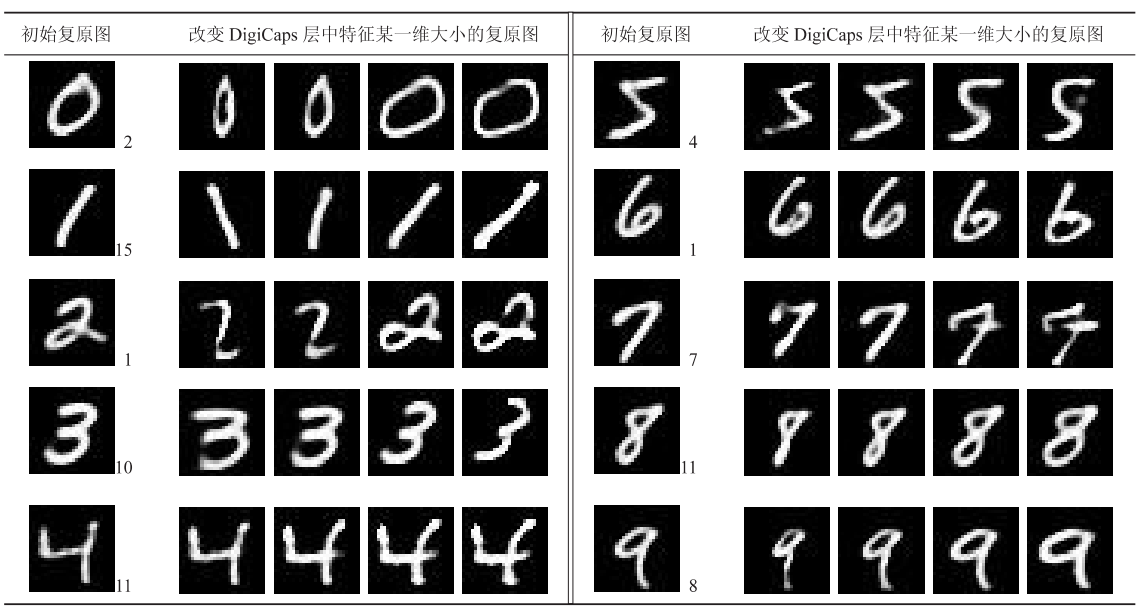
2.2.2 改变 DigiCaps 层胶囊中矢量化特征维数

将 DigiCaps 层胶囊中特征维数下降为 8 时,同样按表 2 中的参数训练图 2 网络,最终训练得到的 loss 值和维度为 16 时的 loss 值比较接近。测试得到的准确率在 98.75% 左右,相较于维度为 16 时的准确率稍有下降。

当矢量化特征维数下降为 8 时,仍然能提取到数据集的部分空间信息,它们与维数为 16 时提取到的空间信息基本相同,但是没有 16 维时提取的空间信息全面,并且在这 8 个维度所表示的空间信息中相互类似的越来越少。说明维数下降为 8 时也能够提取 MNIST 数据集的大部分空间信息,使得在复原图像过程中能够逼近原图像,维数为 8 时的 loss 值接近于维数为 16 时的 loss 值也从侧面说明了这一点。

将 DigiCaps 层胶囊中特征维数下降为 4 时,同样按表 2 中的参数训练图 2 网络,最终训练得到的 loss 值比之前的维度为 16 和 8 时都要大很多,大约在 0.008 左右,测试得到的准确率在 98.5% 左右,比之前维度为 16 和 8 时的准确率都要低。当矢量化特征维数下降为 4 时,只能够提取到数据集的一小部分空间信息,且 4 个维度所表示的空间信息一般各不相同。由于提取的空间信息较少,使得复原出的图像与原图像差距相对较大,因此矢量化特征维数为 4 时的 loss 值相对于维数为 16 和 8 时的 loss 值较大。

表 3 维数为 16 时的特征可视化



根据以上实验得出,在选取 DigiCaps 层胶囊中矢量化特征维数时,应该根据训练数据集中同一类别所有图像所包含的空间信息的多少,也就是同一类别图像的多样化来判断适合的维数。维数过少会导致提取

到的图像的空间信息较少,在复原图像时就会与原图有较大的差距,维数过多虽然能提取到更多的空间信息,但同时也会增加网络的参数量。

3　结束语

通过特征可视化的方法验证了 CNN 中提取到的特征空间关联性较弱,而胶囊网络提取到的矢量化特征确实包含了多种空间信息,这使得特征之间具有空间关联性,可解释性也更强。相比于 CNN,胶囊网络不仅可以学习到输入图像的绝大部分空间信息,同时提取到特征的多种不同变体也可以通过改变 DigiCaps 层矢量化特征得到。但是 CNN 只能够提取到输入图像的少量空间信息,能够提取到的特征的不同变体也很有限。使得对于每个类别的输入,胶囊网络比传统的卷积网络能学习到一个更加鲁棒的表示。并且随着 DigitCaps 层特征维数的降低,矢量化特征所能提取到的空间信息越来越少,胶囊网络复原出的图像与原图差距也就越大。

对于 MNIST 数据集,胶囊网络只需要更少的训练数据就可以得到较高的准确率。并且胶囊网络比较擅长分割任务,在识别高度重叠的数字时,其效果要明显好于卷积神经网络,说明胶囊网络是一个值得探索的方向。

参考文献:

[1] HUANG G, LIU Z, MAATEN L V D, et al. Densely connected convolutional networks[C]//IEEE conference on computer vision and pattern recognition. [s. l.]: IEEE, 2017:2261-2269.

[2] 李　钊,卢　苇,邢薇薇,等. CNN 视觉特征的图像检索[J]. 北京邮电大学学报,2015,38:103-106.

[3] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[C]// Proceedings of the 25th international conference on neural information processing systems. Lake Tahoe, Nevada: Curran Associates Inc, 2012:1097-1105.

[4] ZHANG Tielin, ZENG Yi, XU Bo. HCNN: a neural network model for combining local and global features towards human-like classification[J]. International Journal of Pattern Recognition & Artificial Intelligence, 2016, 30(1):1655204.

[5] SABOUR S, FROSST N, HINTON G E. Dynamic routing between capsules[C]//31st conference on neural information processing systems. Long Beach, CA, USA: [s. n.], 2017:3856-3866.

[6] HINTON G E, KRIZHEVSKY A, WANG S D. Transforming auto-encoders[C]//International conference on artificial neural networks. Berlin: Springer, 2011:44-51.

[7] ZEILER M D, FERGUS R. Visualizing and understanding convolutional networks[C]//European conference on computer vision. Berlin: Springer, 2014:818-833.

[8] 陈浩翔,蔡建明,刘铿然,等. 手写数字深度特征学习与识别[J]. 计算机技术与发展,2016,26(7):19-23.

[9] 林　列,常胜江. 用于手写体数字不变性特征提取及识别的自组织算法[J]. 电子学报,2003,31(10):1506-1509.

[10] ZHANG Tong, ZHENG Wenming, CUI Zhen, et al. A deep neural network-driven feature learning method for multi-view facial expression recognition[J]. IEEE Transactions on Multimedia, 2016, 18(12):2528-2536.

[11] 俞海宝,沈　琦,冯国灿. 在反卷积网络中引入数值解可视化卷积神经网络[J]. 计算机科学,2017,44(z1):146-150.

[12] FAKHRY A, ZENG T, JI S. Residual deconvolutional networks for brain electron microscopy image segmentation[J]. IEEE Transactions on Medical Imaging, 2017, 36(2): 447-456.

[13] 吴丽芸,王文伟,张　平,等. 手写混合字符集识别的多特征多级分类器设计[J]. 计算机应用,2005,25(12):2948-2950.

[14] 耿西伟,张　猛,沈建京. 基于结构特征分类 BP 网络的手写数字识别[J]. 计算机技术与发展,2007,17(1):130-132.

[15] OLSHAUSEN B A, ANDERSON C H, VAN ESSEN D C. A neur-obiological model of visual attention and invariant pattern recognition based on dynamic routing of information[J]. Journal of Neuroscience, 1993, 13(11):4700-4719.

[16] ZHANG K, SUN M, HAN X, et al. Residual networks of residual networks: multilevel residual networks[J]. IEEE Transactions on Circuits & Systems for Video Technology, 2016, 28(6):1303-1314.

[17] RUSSAKOVSKY O, DENG J, SU H, et al. ImageNet large scale visual recognition challenge[J]. International Journal of Computer Vision, 2015, 115(3):211-252.

[18] PELLI D G, PALOMARES M, MAJAJ N J. Crowding is unlike ordinary masking: distinguishing feature integration from detection[J]. Journal of Vision, 2004, 4(12):1136-1169.