

基于改进 CNN 的公交车内拥挤状态识别

徐明远, 崔 华, 张立恒

(长安大学 信息工程学院, 陕西 西安 710000)

摘要:针对传统的视频图像处理方法对公交车内乘客拥挤状态的检测受运动阴影、动态背景及场景光照变化等因素的影响问题,提出了一种基于改进卷积神经网络 VGG-16 的公交车内拥挤状态识别方法。该方法在 VGG-16 的模型基础上,优化全连接层层数,使用迁移学习共享 VGG-16 预训练模型的各层权值参数进行训练。相对于文中的传统图像处理方法、AlexNet 模型、GooleNet 模型以及标准 VGG-16 模型,改进的 VGG-16 模型对公交车拥挤状态的识别准确率最高,识别精度能够达到 96.1%。模型的损失值比标准 VGG-16 模型收敛得更快,模型表现得更加稳定。实验证明:改进后的 VGG-16 模型能够更好地提取公交内拥挤状态的特征,解决公交车内拥挤状态的识别问题。

关键词:图像识别;卷积神经网络;模型改进;VGG-16;公交车;拥挤状态

中图分类号:TP391.41

文献标识码:A

文章编号:1673-629X(2020)05-0032-06

doi:10.3969/j.issn.1673-629X.2020.05.007

Recognition of Crowded State in Bus Based on Improved Convolution Neural Network

XU Ming-yuan, CUI Hua, ZHANG Li-heng

(School of Informational Engineering, Chang'an University, Xi'an 710000, China)

Abstract: Aiming at the problem that the traditional video image processing method is used to detect the crowded state of passengers in the bus, such as motion shadow, dynamic background and scene illumination changes, a crowded bus state recognition method based on improved convolution neural network VGG-16 is proposed. Based on the VGG-16 model, this method optimizes the number of all-connected layers and uses the migration learning to share the weight parameters of each layer of the VGG16 pre-training model for training. Compared with the traditional image processing methods, AlexNet model, GooleNet model and standard VGG-16 model, the improved VGG-16 model has the highest recognition accuracy of bus congestion status, and the recognition accuracy can reach 96.1%. The loss value of the model converges faster than that of the standard VGG-16 model, and the model is more stable. The experiment proves that the improved VGG-16 model can better extract the characteristics of the crowded state in the bus and solve the problem of the congestion status in the bus.

Key words: image recognition; convolution neural network; model improvement; VGG-16; bus; crowded state

0 引言

随着生活水平的提高,车辆成为城市居民主要的出行手段,由此引发的道路拥堵问题困扰着许多城市。公共交通能承载城市的大量客流,缓解城市拥堵,各地都在大力提高公共交通对于居民出行的吸引力^[1]。而公交公司目前缺少可靠的乘客拥挤信息识别技术,无法根据实际的客流信息实时地调整公交车的运营策略,尤其是乘车高峰期,公交车内极其拥挤,完全不能满足乘客舒适出行的要求。因此,图像处理识别技术

被应用到公交车内乘客拥挤状态的识别中来。2012年 Garcia-Bunster G 等改正了图像的视角,通过结合线性回归模型和线性判别两个参数,使面积测算和计数达到最优映射,以此来实现公交车排队乘客计数^[2]。2013年 Daley W 等用红外技术检测公交内的乘客状态,分析车辆和乘客的位置形状分布,解决车内乘客的计数问题^[3]。2014年 Mudoi D 等利用背景差分法提取目标图像的相应区域,使用搭建的人工神经网络进行训练,并利用结果中目标的颜色、形状特征完成乘客

收稿日期:2019-07-20

修回日期:2019-11-21

网络出版时间:2020-01-10

基金项目:陕西省重点研发计划重点项目(2018ZDXM-GY-047);教育部联合基金(6141A02022610)

作者简介:徐明远(1995-),男,硕士研究生,CCF 会员(C7302G),研究方向为深度学习、图像识别;崔 华,博士,教授,研究方向为深度学习、图像识别、数据挖掘。

网络出版地址: <http://kns.cnki.net/kcms/detail/61.1450.TP.20200110.1123.056.html>

拥挤识别^[4]。以上识别方法都有着各自的不足之处,背景差分算法会受乘客的运动阴影、车厢的动态背景及车厢内光照变化等因素的干扰,而目标检测一般是检测公交车内乘客的头部特征,如果乘客染发或者乘客衣服与头发颜色相近,可能就会导致错误的判断。

目前,深度学习算法能够很好地解决图像识别的问题^[5]。文中在国内外公交车内拥挤状态识别研究的基础上,结合对公交车内图像拥挤特征的分析 and 卷积神经网络(CNN)的研究,提出了一种基于改进 VGG-16 网络的公交车内拥挤识别方法。考虑到公交公司对乘客拥挤信息识别的实际需要,建立了公交车内拥挤图像四分类数据集,经过迁移学习的权值参数共享,对改进后的 VGG-16 模型进行训练测试。该方法能较好地解决以上各个识别方法的不足,识别的结果能使公交公司及时对乘客拥挤或将要拥挤的情况采取相应的措施,实现智慧调度,提高公交出行的吸引力,从根本上缓解城市交通拥堵。

1 卷积神经网络

受到生物学家研究的启发,计算机视觉科学家开创性地提出了卷积神经网络模型^[6]。卷积神经网络与普通神经网络的不同之处是卷积神经网络采用卷积层和池化层作为特征抽取器^[7]。卷积和池化能极大地减少模型的参数和复杂度,其基本结构如图 1 所示。

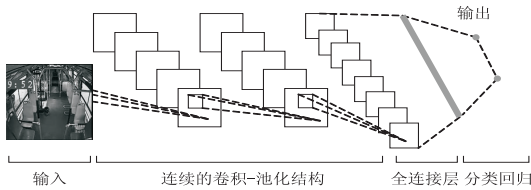


图 1 卷积神经网络结构

1.1 卷积层

在 CNN 结构中,卷积操作通过权值共享的形式减少了网络中参数的数量,同时增加了网络的泛化能力,这些权重在网络中采取随机初始化的方式,在训练中可以学习为能够检测边缘、颜色、形状的滤波器^[8]。卷积运算的公式为:

$$y^j = f_{acti}(b^j + \sum_i w^{ij} * x^i) \quad (1)$$

其中, x^i 和 y^j 是第 i 层的输入特征映射和第 j 层的输出特征映射, w^{ij} 是卷积滤波器的权重, $*$ 表示卷积运算, b^j 和 $f_{acti}(\cdot)$ 是第 j 层的偏置和激活函数。

1.2 池化层

池化层是将相似的特征进行组合,池化计算卷积输出的特征图中一个局部区域的值,通过池化操作减少卷积层输出的特征向量,防止过拟合,同时降低了数据表达的维度,并极大地减少模型参数数量和网络中

的计算量^[9]。

常见的池化方法^[10]有:最大池化(max pooling)、平均池化(mean pooling)和概率池化(stochastic pooling)。最大池化:取图像某一区域特征的最大值作为该区域的特征。平均池化:取图像某一区域特征的均值作为该区域的特征。概率池化:依据概率矩阵的大小,对矩阵区域求加权平均。三种池化方式如图 2 所示。

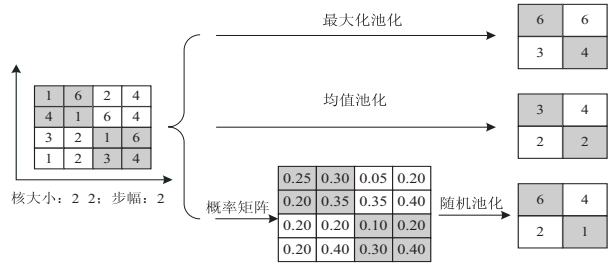


图 2 三种池化方式

此外还有一些池化方式。例如:重叠池化(overlapping pooling),空金字塔池化(spatial pyramid pooling)在卷积神经网络中也具有比较稳定的特征提取能力^[9]。

1.3 激活函数

为了解决实际问题,网络必须能逼近任意的分类函数,因此就要引入非线性激活函数^[11]。常用的激活函数包括 Sigmoid 函数、Tanh 函数和 Relu 函数。公式如下:

$$\text{sigmoid}(x) = \frac{1}{1 + e^x} \quad (2)$$

$$\text{tanh}(x) = \frac{e^{2x} - 1}{e^{2x} + 1} \quad (3)$$

$$\text{Relu}(x) = \max(x, 0) \quad (4)$$

Sigmoid 函数和 Tanh 函数都能够让神经网络逼近非线性函数,解决非线性问题,但是都可能会出现梯度消失或者梯度爆炸的情况,且计算比较复杂。而 Relu 的优点在于计算简单,梯度稳定,也可以表现出抑制的效果,所以文中所训练测试的网络模型均使用 Relu 函数作为激活函数。图 3 为上述三种激活函数的曲线图。

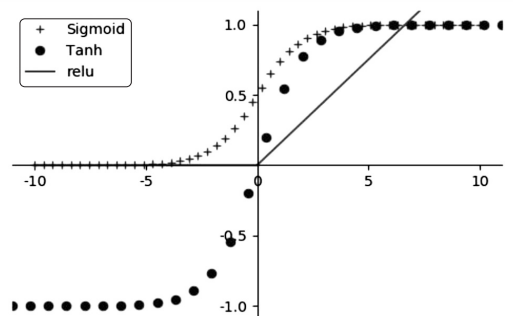


图 3 三种激活函数曲线

2 改进的 VGG-16 卷积神经网络

2.1 VGG-16 网络

VGGNet 在 AlexNet 的基础上探究了网络模型的识别性能和深度的关系,在探究过程中发现较小的卷积核和更深的网络可以明显提升模型的分类精度,使得错误识别率明显下降。VGGNet 使用了 3×3 的卷积核并将模型的层数加深,两个 3×3 卷积核堆叠在一起的视野范围相当于一个 5×5 的卷积核的视野,而三个 3×3 卷积核堆叠在一起的视野范围相当于一个 7×7 的卷积核,在同样的视野范围内采用 3×3 的卷积核可以降低参数的数量^[12]。与此同时更深的网络意味着会有更多的非线性变换,网络的学习能力会更强,可以学习到更多的特征。VGGNet 还采用多尺度训练的方法,增加了训练时的数据量,防止过拟合的同时提升了准确率^[13]。基于 VGGNet 网络结构的独特之处以及该网络在公开数据集上的优异表现,文中选择 VGG-16 网络作为解决公交车内拥挤状态识别问题的特征提取网络,网络结构如图 4 所示。

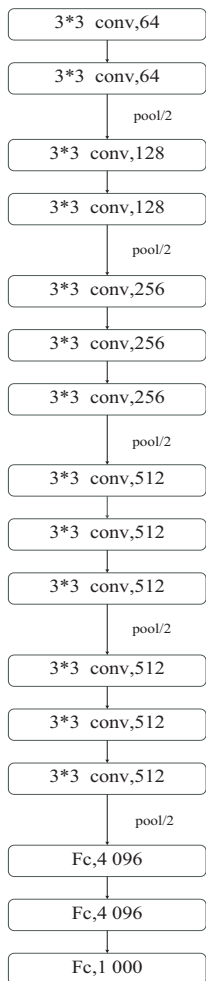


图 4 VGG-16 网络结构

2.2 改进 VGG-16 网络

为了满足对公交车内拥挤状态识别的准确性和计算效率要求,设计了改进的 VGG-16 模型,在保证准

准确率的同时降低模型的计算量。文中构造的公交车拥挤状态识别模型依然使用了标准 VGG-16 网络结构的卷积层,16 个卷积层被五个池化层分隔开,输入的公交车内乘客拥挤状态图像通过 16 个卷积层和五个池化层的计算能提取到较为明显的特征。模型最后的全连接层模拟大脑的逻辑认知过程,对卷积池化层提取的特征进行分类,得出识别标签^[14-15]。由于 VGG-16 模型的层数和参数是为 1 000 个分类类别设计的,其中的权重参数数量有 65×10^6 个,且参数较大部分分布在全连接层上,而文中要解决的问题是 4 种状态的分类识别,不需要那么多的权重参数去参与计算,因此,提出用 2 个全连接层替换原有的 3 个全连接层的改进方法设计第一个全连接层的神经元为 4 096 个,第二个全连接层即输出层的神经元个数为 4 个,降低模型参数个数和复杂度,以提高模型识别精度和计算效率。

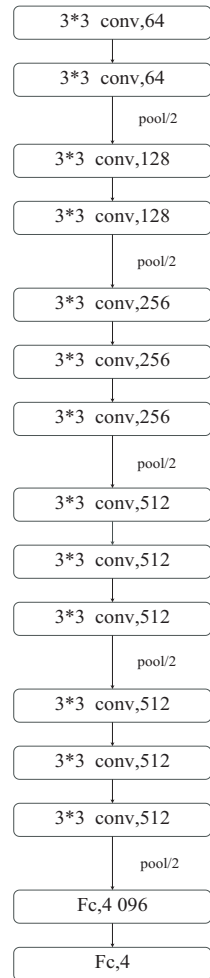


图 5 改进 VGG-16 网络结构

设计的模型如图 5 所示。所有卷积计算都采用 3×3 大小的卷积核,卷积步长固定为 1,所有的池化操作都采用 2×2 的最大池化,池化步长固定为 2,去除了一层全连接层,使得卷积神经网络可以节省大量计算和内存。模型采用 Relu 函数作为学习特征的激活

函数,由此可以使获得的特征更明显,从而达到更好的分类识别效果。

2.3 改进 VGG-16 网络的迁移学习

迁移学习是希望模型从大量的数据样本中学习到比较泛化的特征,从而补充有限样本的部分特征。在实际场景中无法收集到所有的样本,如果只用这些有限的样本训练模型,就会很容易发生过拟合现象,除此之外,如果使用的网络层数较多则可能会导致前边几层的参数很难得到更新,而通过迁移学习就可以解决以上麻烦^[16]。VGG-16 预训练网络是通过 100 万幅图像的 ImageNet 数据库中训练而成,拥有大量已训练

完成的参数和权重,特别是卷积层的权重参数已经能够对图像的曲线、边缘、轮廓特征进行较为有效的特征提取。为避免从头训练整个网络,减少网络训练时间和提高网络训练效率,可将已在 ImageNet 上训练完成的 VGG-16 网络作为文中模型的预训练模型,通过微调参数迁移学习的方式,将预训练的 VGG-16 模型权重参数迁移到改进模型,利用预训练模型的参数优化各层的模型参数,加快模型损失值的收敛速度,提高识别精度和模型的稳定性,更好地解决公交车内拥挤状态识别问题。迁移过程如图 6 所示。

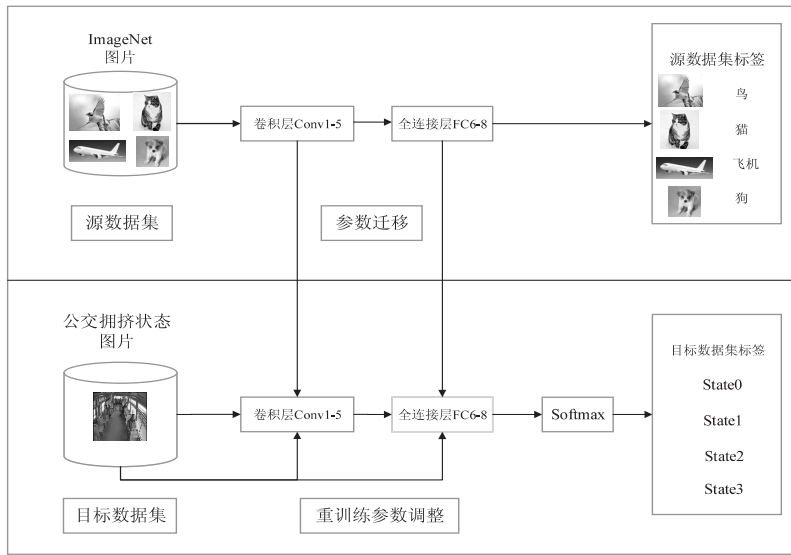


图 6 VGG 迁移学习模型示意图

3 公交车内拥挤状态数据集

基于某市公交公司实际数据,文中建立了一个公交车内拥挤状况图像数据集来评估模型的性能。这些视频数据来自公交车内的前后摄像头,从前后两个角度俯瞰整个车厢,将其中比较清晰的视频选取出来,按照 500 帧截取一次,共获得 15 194 张图片。公交车内拥挤状态如果只分为拥挤和不拥挤意义不大,无法对即将乘车的乘客和公交车的调度人员产生实际的指导价值,因此根据实际需要将公交车内的拥挤情况分为 $state_0$ 、 $state_1$ 、 $state_2$ 和 $state_3$ 四种状态。其中, $state_0$ 表示车内座位未坐满且车内走道大面积未占用的空车状态; $state_1$ 表示走道小部分被占用,乘客中有明显空隙的宽松状态; $state_2$ 表示走道大部分被占用但乘客中有少量空隙的较拥挤状态; $state_3$ 表示走道大部分或完全被占用,乘客中无空隙的拥挤状态。每张图片都是由 2 个不同的标注员独立标注的,对于相同图片的非相同标注,需要两名标注员再进行一轮评估,才能达成最终的标签。最终,15 194 张原始图像被分为四类,各类图像数量分别为 4 066 张,3 582 张,3 414 张和 4 072

张,各类状态如图 7 所示,依次为 $state_0$ 、 $state_1$ 、 $state_2$ 和 $state_3$ 。

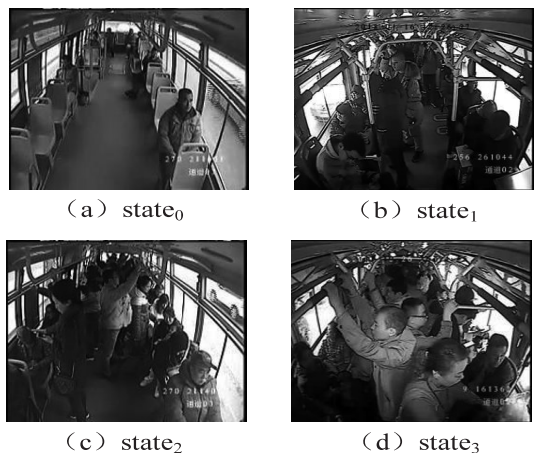


图 7 各类状态展示图

4 实验及结果

4.1 模型训练及实验结果

实验使用 Ubuntu16.04 系统,i7-7700k 处理器,24 GB 内存,采用建立好的公交车内拥挤状态四分类数据集,每一类取出 500 张作为测试样本,剩下的作为训

练样本,将每张图片归一化到 224×224 像素。网络模型训练学习率设置为 0.01,批处理大小设为 32,最大

迭代次数为 5 000 次,使用 caffe 框架进行模型训练。模型训练过程如图 8 和图 9 所示。

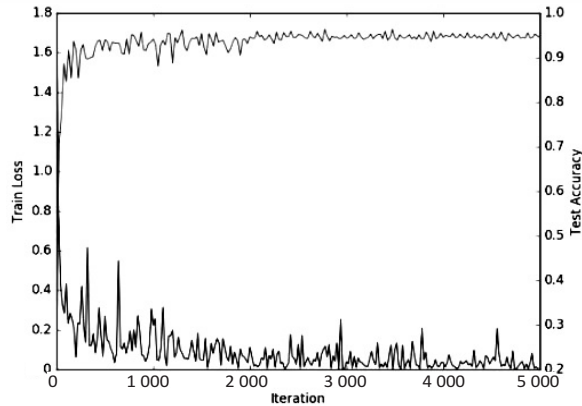


图 8 标准 VGG-16 模型训练图

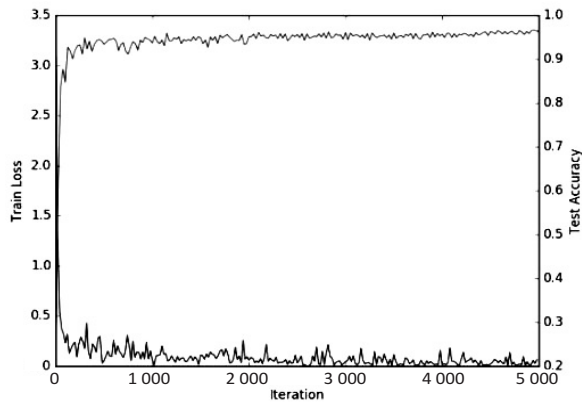


图 9 改进 VGG-16 模型训练图

(1) 两种模型训练的损失函数值都随着迭代次数的增加呈现下降趋势,且下降速度越来越慢,使用改进 VGG 训练时的初始损失函数值相比于 VGG 标准模型更小而且收敛过程中波动较小,改进的 VGG-16 最终收敛的损失函数值更低,基本上稳定在 0.04 附近,而 VGG 标准模型损失函数值收敛到 0.1 左右。

(2) 两种模型的训练准确率都随着迭代次数的增加而不断上升,且上升的速度趋于平缓,最终到达一个稳定的上限,改进后的 VGG 模型最终的训练准确率要略高于标准模型,最终稳定在 96.5% 左右,且曲线波动较小,而标准模型最终稳定在 95% 左右,曲线波

动较大,模型表现不够稳定。

4.2 实验对比分析

为进一步测试改进 VGG 模型的性能,选择传统的分区域融合算法以及 AlexNet, GoogleNet 两种深层神经网络进行对比试验。分区域融合算法是通过人工选定图像的目标区域,融合 HSV 颜色模型法,帧间差分法以及纹理分析法实现公交车的分区域拥挤判断,根据各区域的结果汇总得出乘客的拥挤状态信息。各方法使用事先建立好的测试集进行测试,测试集每一类有 500 张图片,实验结果如表 1 所示。

表 1 文中模型与其它模型方法的识别准确率对比 %

方法	state ₀	state ₁	state ₂	state ₃	平均识别准确率
分区域融合算法	97.6	71.6	85	93.4	86.9
AlexNet	99.2	81.6	92.2	97	92.5
GooleNet	100	85	93.6	96.6	93.8
VGG-16	100	86.8	95.2	98.8	95.2
文中改进模型	100	88.7	96.3	99.4	96.1

表 1 对比了不同方法各类状态的识别结果, AlexNet、GoolgeNet、VGG-16Net 与文中改进模型 4 种

卷积神经网络模型的平均识别准确率均优于传统的分区域融合算法,各种方法的平均识别准确率分别为

86.9% ,92.5% ,93.8% ,95.2% ,96.1%。在对存在模糊分类界限的 $state_1$ 和 $state_2$ 状态进型识别时,文中改进模型比传统的分区域融合算法分别提升了 17.1% 和 11.3%。在卷积神经网络模型中,标准 VGG-16 和改进模型的特征提取都是通过 3×3 滤波器的层层提取,能够更加有效地提取公交车内的拥挤状态特征,识别结果要优于 AlexNet 和 GooleNet。由于改进 VGG 模型减少了全连接层数量,大量降低了训练参数量,因此在识别精度上略高于标准 VGG。从测试结果看,文中提出的改进 VGG-16 模型比其它模型方法具有更好的识别准确率,更适合解决公交车拥挤状态识别的实际问题。

5 结束语

对公交车内拥挤识别问题进行研究,在标准 VGG-16 网络基础上,用 2 个全连接层替换原有的 3 个全连接层,并用 4 标签 Softmax 分类器替换原有的分类器,利用迁移学习对建立好的公交车拥挤状态四分类数据集进行训练和测试。

对比其它方法,改进后的 VGG 模型识别精度更高,准确率达到 96.1%,且比标准 VGG 模型收敛更快,更加稳定,能更好地解决公交车拥挤状态识别问题。下一步要对模型的卷积层和池化层的结构进行优化,提高模型对于模糊分类界限 $state_1$ 和 $state_2$ 的特征提取能力和识别精度。

参考文献:

- [1] 黄颖. 公交车乘客拥挤程度自动检测技术研究[D]. 西安:长安大学,2017.
- [2] GARCIA-BUNSTER G, TOMES-TORRID M, OBERLI C, et al. Crowded pedestrian counting at bus stops from perspective transformations of foreground areas[J]. IET Computer Vision, 2012, 6(4): 296-305.
- [3] DALEY W, USHER C, ARIF O, et al. Detection of vehicle occupants in HOV exploration lanes of image sensing for de-

tection of vehicle occupants[C]//Conference on video surveillance and transportation imaging applications. [s. l.]: [s. n.], 2013.

- [4] MUDOI D, KASHYAP P A. Vision based data extraction of vehicles in traffic[C]//International conference on signal processing and integrated networks (SPIN). Noida, India: IEEE, 2014: 202-208.
- [5] 华莉琴, 许维, 王拓, 等. 采用改进的尺度不变特征转换及多视角模型对车型识别[J]. 西安交通大学学报, 2013, 47(4): 92-99.
- [6] 周飞燕, 金林鹏, 董军. 卷积神经网络研究综述[J]. 计算机学报, 2017, 40(6): 1229-1251.
- [7] 刘雨桐, 李志清, 杨晓玲. 改进卷积神经网络在遥感图像分类中的应用[J]. 计算机应用, 2018, 38(4): 949-954.
- [8] 范望, 韩俊刚, 苟凡, 等. 卷积神经网络识别汉字验证码[J]. 计算机工程与应用, 2018, 54(3): 160-165.
- [9] TOLIAS G, SICRE R, JÉGOU H. Particular object retrieval with integral max-pooling of CNN activations[C]//Computer Vision and pattern recognition. [s. l.]: [s. n.], 2015: 1511. 05879.
- [10] MOKARIAN A, MALINOWSKI M, FRITZ M, et al. Mean box pooling: a rich image representation and output embedding for the visual madlibs task[C]//Computer Vision and pattern recognition. [s. l.]: [s. n.], 2016: 1608. 02717.
- [11] 曲之琳, 胡晓飞. 基于改进激活函数的卷积神经网络研究[J]. 计算机技术与发展, 2017, 27(12): 77-80.
- [12] GEMAN S, BIENENSTOCK E, DOURSAT R. Neural networks and the bias/variance dilemma[J]. Neural Computation, 2014, 4(1): 1-58.
- [13] 蔡娟, 蔡坚勇, 廖晓东, 等. 基于卷积神经网络的手势识别初探[J]. 计算机系统应用, 2015, 24(4): 113-117.
- [14] 王龙, 刘辉, 王彬, 等. 结合肤色模型和卷积神经网络的手势识别方法[J]. 计算机工程与应用, 2017, 53(6): 209-214.
- [15] 杨红玲, 宣士斌, 莫愿斌. 基于卷积神经网络的手势识别[J]. 计算机技术与发展, 2018, 28(7): 11-14.
- [16] 柴帅, 李壮举. 基于迁移学习的番茄病虫害检测[J]. 计算机工程与设计, 2019, 40(6): 1701-1705.