

基于深度学习身份证鉴别与信息检测方法研究

焦 亮,张太红*

(新疆农业大学 计算机与信息工程学院,新疆 乌鲁木齐 830001)

摘 要:身份证信息检测成为身份证信息识别任务中的关键步骤,将深度学习中的检测算法应用于身份证文本信息检测研究。将身份证信息检测分为两部分,第一部分提出身份证照片鉴别网络,主要利用残差网络 Resnet 在身份证数据集下训练出适应于身份证照片鉴别的模型,实现身份证照片的鉴别,并且针对三种不同网络层数的 Resnet 网络进行训练,并对三种模型进行测试得到三种测试结果,得到三种不同准确率的模型;第二部分进行 CTPN 身份证文本信息检测研究,针对第一部分鉴别正确的身份证照片进行身份证图片文本信息检测,将 CTPN 网络在身份证数据集下训练出适应于身份证信息检测的模型,从而完成身份证信息的文本检测任务,并描出文本信息检测框。实验结果表明可以检测出身份证版面上相对应的文本信息并成功描出文本检测框。

关键词:深度学习;残差网络;身份证鉴别;场景文本检测;信息检测

中图分类号:TP183

文献标识码:A

文章编号:1673-629X(2020)12-0203-07

doi:10.3969/j.issn.1673-629X.2020.12.036

Research on Identity Card Identification and Information Detection Based on Deep Learning

JIAO Liang,ZHANG Tai-hong*

(School of Computer and Information Engineering,Xinjiang Agricultural University,Urumqi 830001,China)

Abstract:ID card information detection has become a key step in ID card information identification task. The detection algorithm of deep learning is applied to ID card text information detection research. ID information detection can be divided into two parts. In the first part, ID card photo identification network is put forward. A model suitable for ID card photo identification is trained by using residual network Resnet under ID card data set to realize ID card photo identification. In addition, the Resnet network with three different network layers is trained, and the three models are tested to obtain three test results and three models with different accuracy rates. In the second part, the research on ID card text information detection of CTPN is carried out. For the ID card photo identified correctly for the first part, ID card image text information detection is carried out, and a model suitable for ID card information detection is trained by CTPN network under the ID card data set, so as to complete the ID card information text detection task, and trace the text information detection box. The experiment shows that the corresponding text information on the layout of the ID card can be detected and the text detection box can be successfully traced.

Key words:deep learning;residual network;ID identification;scene text detection;information detection

0 引 言

身份证作为一种信息储存媒介在各个领域利用率越来越高,身份证版面信息作为最基本最直观的信息被应用于很多领域,是社会生活中非常重要的身份识别证件,实名认证领域需要身份证进行信息认证,例如:入学、就业、户口登记、酒店住宿、各类业务办理等等。

在一些身份证信息登记领域,通常是人工登记并录入系统,过程相对繁琐并且耗费人力。近年来深度

学习技术在图像分类、目标检测领域得到了很大的推广与应用,由于深度学习在分类任务中的准确率以及检测任务中的精确度相比传统方法都有显著提高,而且有更好的泛化能力以及对不同环境的适应性,因此,该文重点介绍了基于深度学习框架对身份证的鉴别以及身份证信息的检测,并阐述了自己的研究成果。

1 卷积神经网络(CNN)概述

卷积神经网络(convolutional neural network,

收稿日期:2020-01-18

修回日期:2020-05-20

基金项目:新疆维吾尔自治区重大科技专项(2017A01002-5)

作者简介:焦 亮(1990-),男,硕士研究生,研究方向为农业信息化;通讯作者:张太红(1965-),男,博士,教授,研究方向为数据库技术。

CNN)的核心思想是利用卷积核扫描图像,得到相应的图像特征,卷积核可以被理解为过滤器或者特征扫描器,通常称之为感受野。卷积神经网络的基本结构包括图像输入层(input layer)、卷积计算层(convolutional layer)、ReLU 激励层(rectified linear units layer)、池化层(pooling layer)、全连接层(fully-connected layer)和输出层(output layer)。

· 卷积运算:将前一层的特征图与一个卷积核进行卷积运算,卷积运算结果经过激励函数输出形成这一层神经元,从而构成这一层特征图。

· 激励函数:通常采用的激励函数为 Sigmoid 函数和 ReLU 函数,是将卷积层输出结果进行非线性映射。该文采用 ReLU 函数作为网络的激励函数,采用 ReLU 函数相比 Sigmoid 计算速度快,一部分神经元会因为 ReLU 输出为 0,网络变得稀疏使得减少了参数之间的相互依赖关系,一定程度上缓解了过拟合状况。

· 池化运算:将输入数据分割成不重叠区域,通过池化运算来降低网络的空间分辨率,池化方法包括最大值池化,选择区域内最大值。均值池化,计算区域内平均值,池化操作可以起到降维作用,减小卷积核尺寸同时保留相应特征。

· 全连接运算:经过多次卷积池化运算后的输入数据为多组数据,全连接层运算后将多组数据合并成一组数据。

· 输出:输出分类结果,利用 Softmax 做分类器。

2 基于 Resnet 身份证照片鉴别网络设计

对身份证信息检测之前首先需要对输入图像进行判别,如果判定为身份证图像再进行下一步版面文本信息的检测工作,若不是身份证则不需要进行下一步操作。身份证照片鉴别网络是基于 Resnet 的二分类任务,一类是身份证照片,一类是非身份证照片。

Resnet 网络共有五种深度的网络结构,根据网络层数不同分别分为 Resnet-18, Resnet-34, Resnet-50, Resnet-101, Resnet-152。5 种网络结构,主要将 Resnet 网络应用于身份证照片的鉴别的基础网络,这里给出其中三种网络结构,如表 1 所示。将 Resnet-50, Resnet-101, Resnet-152 作为基础网络,并在三种网络结构的基础上在同一数据集上进行训练,得到三种不同的模型,将三种模型在同一测试集下测试得到三组不同的测试结果,将三种测试结果进行对比。

表 1 Resnet 网络层数

Layer name	Resnet-50	Resnet-101	Resnet-152
Conv1	7×7, 64, stride2	7×7, 64, stride2	7×7, 64, stride2
	3×3max pool, stride2	3×3max pool, stride2	3×3max pool, stride2
	1×1, 64	1×1, 64	1×1, 64
Conv2_x	3×3, 64 ×3	3×3, 64 ×3	3×3, 64 ×3
	1×1, 256	1×1, 256	1×1, 256
	1×1, 128	1×1, 128	1×1, 128
Conv3_x	3×3, 128 ×4	3×3, 128 ×4	3×3, 128 ×8
	1×1, 512	1×1, 512	1×1, 512
	1×1, 256	1×1, 256	1×1, 256
Conv4_x	3×3, 256 ×6	3×3, 256 ×23	3×3, 256 ×36
	1×1, 1 024	1×1, 1 024	1×1, 1 024
	1×1, 512	1×1, 512	1×1, 512
Conv5_x	3×3, 512 ×3	3×3, 512 ×3	3×3, 512 ×3
	1×1, 2 048	1×1, 2 048	1×1, 2 048
Average pool, 1 000-d fc, Softmax			

3 身份证信息检测网络设计

随着科技的进步,文本检测方法也在不断发展,由传统方法发展到深度学习的方法,由简单场景下的文本信息检测到复杂文本信息检测。简单场景文本检测复杂程度不高,通常情况下简单场景的文本信息清晰,照片规整,几乎不存在光照等噪声的影响。复杂场景主要指文本信息背景复杂,图像噪声相对较多。复杂场景的文本信息检测影响因素相对较多,由于身份证

照片来自于不同环境、不同相机、不同拍照的人,照片会受到不同噪声的影响,将身份证信息检测列入复杂场景下的文本信息检测。该文主要基于深度学习框架卷积神经网络学习建立检测模型从而检测身份证文本信息。

3.1 基于形态学原理文本检测算法

利用计算机视觉中的图像形态学操作,包括腐蚀、膨胀、基本操作,即可实现简单场景下的文本检测。“膨胀”是对图像的高亮部分进行扩张,让白色区域增

加,“腐蚀”就是图像中高亮部分被蚕食让黑色区域增加。通过设置一定的阈值,经过膨胀、腐蚀一系列操作可将文字区域轮廓突出,并消除一些边框线条,再通过轮廓方法计算出文字区域位置。郭桂芳等人^[1]利用形态学原理对多方向文本检测进行研究,宾西川等人^[2]对形态学膨胀运算文本水印进行研究,李朝晖等人^[3]利用形态学原理对视频文本检测进行了研究。

3.2 最大稳定极值区域文本检测算法

最大稳定极值区域(maximally stable extremal regions, MSER)算法是2002提出的,主要是基于分水岭的思想来做图像中斑点的检测。MSER对一幅已经处理成灰度的图像进行二值化处理,这个处理的阈值从0到255递增,这个阈值的递增类似于在一片土地上做水平面的上升,随着水平面上升,高高低低凹凸不平的土地区域就会不断被淹没,这就是分水岭算法,而这个高低不同,就是图像中灰度值的不同。而在一幅含有文字的图像上,有些区域(比如文字)由于颜色(灰度值)是一致的,因此在水平面(阈值)持续增长的一段时间内都不会被覆盖,直到阈值涨到文字本身的灰度值时才会被淹没,这些区域就叫做最大稳定极值区域。张开玉等人^[4]利用MSER对自然场景倾斜文本定位算法进行了研究,邢延超等人^[5]基于MSER和NMS研究了如何检测变形文档字符,钱江等人^[6]利用最稳定极值区域与边缘增强方法对船名的定位进行了研究。

3.3 复杂场景下文本检测概述

复杂场景下的文本信息检测主要基于深度学习方法,越来越多的人将深度学习神经网络的方法应用于文本信息检测。旷世科技在2017年提出来的EAST^[7](efficient and accurate scene text detector)模型就是非常经典的文本信息检测模型。孟健等人^[8]利用改进EAST算法实现了对电厂电气设备铭牌文字的检测。Shi Baoguang等人在2017年发表于CVPR上的一篇文章提出SegLink^[9](linking segments)文本信息检测模型,针对复杂背景下文本信息进行检测。

该文主要利用CTPN^[10](detecting text in natural image with connectionist text proposal network)复杂场景下文本检测对身份证信息进行检测,并通过大量身份证数据集针对性地训练自己的模型,从而更好地进行身份证信息检测。CTPN是由Faster RCNN^[11](towards real-time object detection with region proposal networks)改进而来,加入了适应于文本序列信息检测的LSTM^[12](long short term)。邓丹^[13]利用PixelLink进行了实例分割的自然场景文本检测算法研究,PixelLink主要是做某个像素(pixel)的文本/非文本的分类预测,以及该像素的8个邻域方向是否存在连接

(link)的分类预测。

4 CTPN 身份证信息检测算法概述

CTPN是基于连接预选框的文本信息检测,CTPN模型利用RNN^[14](recurrent neural networks,循环神经网络)和CNN(convolutional neural network,卷积神经网络)。

4.1 CTPN 中卷积神经网络模型 VGG-16 网络综述

CTPN模型中通过利用VGG-16卷积输出的特征图(feature map)生成一系列预选框(proposal)进行检测。VGGNet^[15]有A-E七种结构,VGG-16是VGGNet七种结构中的一种,从A-E随着网络层数增加网络逐步变深,参数量却没有增加很多,其中D种网络模型所描述的网络结构就是VGG-16。VGGNet拥有5段卷积,每一段卷积包含2-3个卷积层,同时每段结尾都会连接一个最大池化层(max pooling),作用是对特征图进行下采样。

4.2 CTPN 中神经网络 RNN 模型综述

RNN是一类用于处理序列数据的神经网络。身份证信息检测过程中利用CNN对文本检测提取空间信息,利用RNN提取图像中文本的序列特征,在本次实验中利用LSTM模型。LSTM,长短期记忆循环神经网络是一种RNN特殊类型,在身份证照片信息检测网络模型中使用双向LSTM(bi-directional LSTM)就是将两个方向相反的LSTM连接起来。LSTM是由Hochreiter和Schmidhuber(1997)提出,在随后的工作中被广泛应用。双向LSTM是通过正向和反向两个方向输入输出更准确地判断序列信息。

双向LSTM可以看成是两层神经网络,第一层从左边作为系列的起始输入,在文本处理上可以理解成从文本的开头开始输入,而第二层则是从右边作为系列的起始输入,在文本处理上可以理解成从文本的最后一个字作为输入,反向处理与上一层做同样的处理操作,最后对得到的两个结果进行堆叠处理,模型即实现考虑上下文信息的能力。

4.3 文字定位方法与文本线构造过程

CTPN主要是针对身份证信息的检测,身份证信息主要是横向文本信息,采用一组等宽的Anchor,Anchor是固定宽度,默认宽度为16,高度采用一组10个不同的矩形框(11,16,23,33,48,68,97,139,198,283)。在x方向上,Anchor覆盖文本行每个点且不重合,为了覆盖不同高度的文本所以设置不同高度的Anchor,得到Anchor后Softmax判断Anchor是否包含文本,而后Bounding box regression修正包含文本的Anchor的中心y坐标与高度,经过这两步的处理会得到图1(a)所示的一组竖直条状text proposal。后续只

需要将这些 text proposal 用文本线构造算法连接在一起即可获得文本位置。

文本线构造 (side-refinement) 过程是就是将定位出来的宽度为 16 的“小矩形框”加以合并和归纳,得



(a)文字定位



(b)文本线构造

图1 文字定位与文本线构造过程

5 CTPN 文本信息检测整体实现过程

CTPN 文本信息检测过程如下,整体结构如图2所示。

(1)若输入 $3 * 600(h) * 900(w)$ 的图像,使用 VGG16 进行特征提取得到 cov5-3 (VGG 第5个 Block 的第三个卷积层)的 feature map 大小为 $512 * 38 * 57$ 。

(2)在得到的 feature map 上做滑动窗口的滑动操作,窗口大小(即卷积核大小) $3 * 3$ 即 $512 * 38 * 57$ 变为 $460 * 38 * 57$ 。

(3)将每一行所对应窗口对应特征输入 RNN 中,由于每个 LSTM 层是 128 个隐层,即 $57 * 38 * 4$ 608 变

到需要的文本信息的位置信息,就是将图1(a)小矩形框合并最后生成图1(b)大矩形框。主要思路就是将两个相邻的候选区 (proposal) 组成一个 pair,合并不同的 pair 直到最终无法合并为止。

为 $57 * 38 * 128$, Reverse-LSTM 同样得到 $57 * 38 * 128$,合并后最终得到结果为 $256 * 38 * 57$ 。

(4)将 RNN 结果输入到 FC 层(全连接层),FC 层是一个 $256 * 512$ 的矩阵参数,得到 $512 * 38 * 57$ 的结果。

(5)FC 层,特征输入到三个分类或者回归层中,第一个 $2k$ vertical coordinates 和第三个 k side-refinement 是用来回归 k 个 anchor 位置信息,第二个 $2k$ scores 表示的是 k 个 anchor 类别信息,判断是不是字符。使用文本构造算法得到细长的矩形框将其合并成文本的序列框,主要思想为每两个相邻的候选区域组成一个 pair,合并不同的 pair 直到无法合并。

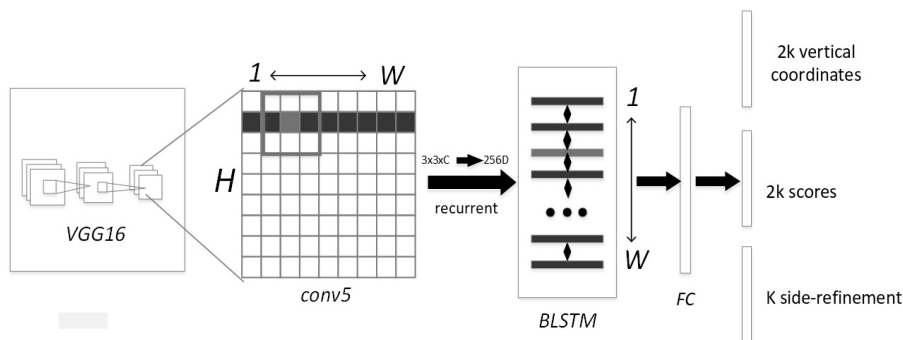


图2 CTPN 文本信息检测整体结构

6 模型评估与实验结果分析

实验环境在 Ubuntu 系统上进行搭建,实验环境配置为 GTX 1080 Ti i7-8700k 12 核 32G 250G SSD+3TB 硬盘 网速 D100/U20。

6.1 评估标准综述

目标检测的评价指标应用于身份证信息检测模型的性能评估,包括以下评价指标 (TP, TN, FP, FN, Precision, Recall, IOU, AP)。

TP (true positives), TN (true negatives) FP (false positives), FN (false negatives), GroundTruth: 有监督训练的训练集的分类准确性。

Precision (精确度), Recall (召回率), IOU (intersection-over-union, 检测框重叠程度), AP (average precision):

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{IOU} = \frac{TP}{TP + FP + FN}$$

候选框 (candidate bound) 与原标记框 (ground bound) 的交叠率或者重叠率。

不同的检测任务有不同的评估标准,通过评估标

准来判定所训练模型的好坏,也是通过评价标准中的相对应指标来衡量模型的性能。目标检测面向判定框和预测框的重叠程度 AP;以评估模型得到的 Recall 值为横轴数据,以评估模型得到的 Precision 为纵轴数据,PR 曲线下的面积即定义为 AP。由上面得到了 PR 曲线,即得到了 n 个 (P, R) 坐标点,利用这些坐标点便可以计算出 AP (average precision):

方法一:11 点插值法,这里参考的是 PASCAL VOC CHALLENGE^[16] 的计算方法。设定一组阈值, $[0, 0.1, 0.2, \dots, 1]$, 然后对于 recall 大于每一个阈值 (比如 $\text{recall} > 0.3$), 都会得到一个对应的最大 precision, 这样就计算出了 11 个 precision。AP 即为这 11 个 precision 的平均值。

方法二:PASCAL VOC CHALLENGE 自 2010 年后就换了另一种计算方法。新的计算方法假设这 N 个样本中有 M 个正例, 那么会得到 M 个 recall 值, 对于每个 recall 值, 可以计算出对应 $(r > r')$ 的最大 precision, 然后对这 M 个 precision 值取平均即得到最后的 AP 值。

6.2 数据集

本实验主体为身份证信息检测, 实验数据集包含两部分, 分别训练实验中两个网络模型。图 3 是所收集到的部分身份证图片数据。



(a) 未进行数据增强的身份证照片

(b) 数据增强的身份证照片

图3 部分身份证照片对比

6.3 身份证照片鉴别以及文本信息检测实验结果

通过实验分别获得身份证照片鉴别与身份证文本信息检测实验结果。

6.3.1 身份证照片鉴别网络实验结果分析

此次实验中测试集图片单独分离, 训练集与测试

第一部分: 身份证照片鉴别网络, 数据集包含 7 000 张图片作为训练集, 包括所收集的 3 500 张身份证照片, 如图 3(a) 所示。3 500 张其他不含身份证图片的其他图片, 100 张图片作为测试集, 包含 50 张身份证照片, 50 张非身份证照片。

相同训练集下在不同的三个不同层数的 Resnet 网络下进行训练, 包含 Resnet-50, Resnet-101, Resnet-152, 分别得到三个网络的训练模型。将训练得到的三个不同的训练模型, 在同一测试数据集下测试其对身份证照片鉴别的准确性, 在此测试数据上得到三种模型下的准确率。

第二部分: 身份证信息检测网络, 数据集包含收集的 5 000 张身份证照片, 如图 3(b) 所示。数据集利用 LabelImg 进行每张身份证的信息标签的标注工作, LabelImg 是图形图像注释工具, 用于数据集的标签制作。为了提高模型的泛化能力, 将 1 000 张身份证照片进行数据增强处理即进行不同的光照处理, 如图 3(b) 所示。部分光照处理后的身份证图片, 通过训练得到文本检测模型, 将 100 张身份证照片作为测试数据, 通过测试得到身份证照片的文本检测结果图, 以及通过训练模型对 100 张身份证照片进行检测得到的 AP 值。

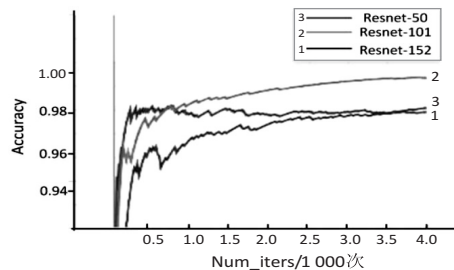
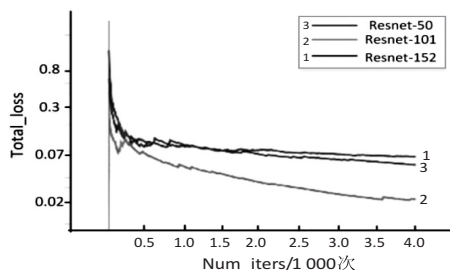


图4 训练过程损失值与准确率

集不相交, 相互独立, 在相同的数据集下, 利用 Resnet 的三种结构网络在基本训练参数相同的情况下得到了三种分类模型, 模型训练过程中的损失值曲线以及准确率如图 4 所示, 并且在 100 张图片下测试其鉴别效果, 得到 100 张图片下的准确率, 如表 2 所示。

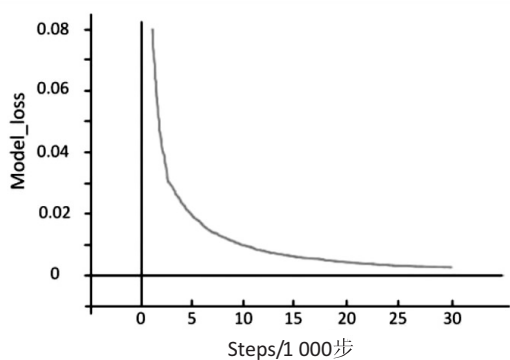
表2 测试数据集下的准确率

网络	测试图片总数	正确个数	错误个数	准确率/%
Resnet-50	100	79	21	79
Resnet-101	100	92	8	92
Resnet-152	100	72	28	72

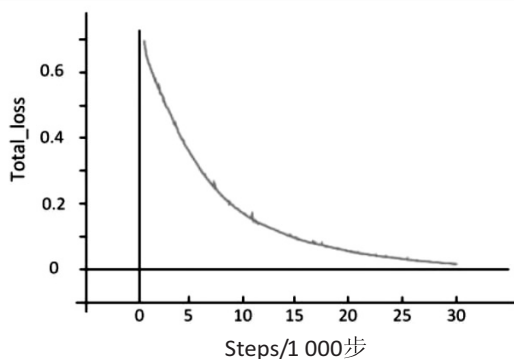
通过损失值曲线图可以看到, Resnet-101 损失降到最低最终达到收敛趋势, 通过准确率的曲线图看到 Resnet-101 在测试数据集上可以达到 92% 的准确率。Resnet-152 网络层数较多, 身份证照片鉴别网络属于二分类任务, 二分类任务在复杂网络下相对简单, 可能导致训练过程中出现过拟合的情况, 故在 Resnet-152 训练得到的模型下分类效果并不是很好, 测试数据上准确率为 72%, Resnet-50 在身份证鉴别任务中没有得到很好的效果, 可能是因为网络层数相对较少, 训练次数不足最终在测试数据集上准确率达到 79%。

6.3.2 身份证文本信息检测网络实验结果分析

在 VGG-16 作为基础网络结构的基础上通过由 5 000 张身份证照片组成的数据集上训练得到 CTPN 文本信息检测模型, 利用所得到的模型在 100 张随机身份证照片组成的测试数据集上进行测试。



(a) 训练模型损失值

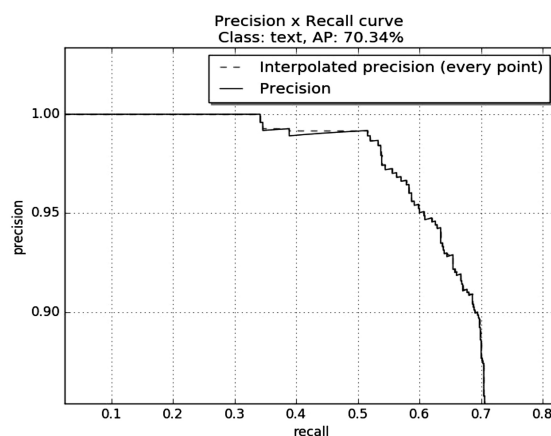


(b) 训练总损失值

图5 CTPN 训练模型损失值与总损失值

图5所示是CTPN模型损失曲线与总损失曲线, 通过损失曲线可以看到在训练30 000steps时曲线开始收敛, 在测试数据集上进行测试得到图6(a), 横坐标为召回率(recall), 纵坐标为精确度(precision), 横

纵轴所组成的曲线图即为PR图, PR图中曲线与横纵坐标轴所围成的面积即为测试数据在文本信息检测模型下的AP值, AP值所反映的即为文本检测效果, 通过图6(a)得到文本信息检测的AP值达到70.34%。图6(b)为测试数据集上部分身份证照片通过文本信息检测模型进行文本信息检测的检测效果图, 文本框所框出的区域即是身份证照片上的身份证文本信息。



(a) 测试数据 PR 图



(b) 文本信息检测结果

图6 CTPN 测试数据 PR 图与 CTPN 部分检测效果

7 结束语

文本检测技术在各个领域都取得了很大的进展, 随着科技进步实名认证的要求也会相应提高。

(1) 身份证的应用场景会越来越多, 随之而来的是不同场景下的身份证照片的信息检测, 对身份证版面信息影响的噪声也会多样化, 为了提高在身份证信息检测方向模型的适应能力, 需要进一步收集多样化的身份证数据集训练更好的模型。

(2) 在以后的研究学习中如何借助计算机手段自动生成所需要的样本图片, 在一定程度上减少样本收集的困难, 并且可以得到大量样本用于模型训练。

(3) 现有文本检测方法与深度学习其他方向研究类似,需要大量的样本,需要对大量样本进行标注,因为样本不是同一标准,故对文本信息标注是一件较为繁琐工作,日后的研究中如何在样本量不足的情况下通过其他方式改善这一状况,并且可以得到较好的训练模型。

文本检测一直是该领域一个比较热门的话题,该文主要是结合深度学习将神经网络用于身份证的分类任务以及身份证文本信息检测任务。

参考文献:

- [1] 郭桂芳,洪留荣,葛方振.基于形态学的多方向文本定位方法[J].宿州学院学报,2015,30(9):103-105.
- [2] 宾西川,裘正定,章春娥,等.一种基于数学形态学膨胀运算的文本数字水印[J].计算机安全,2009(11):4-6.
- [3] 李朝晖,余英林,黄海康.基于形态学的视频文本自动检测[J].计算机应用研究,2005,22(2):258-260.
- [4] 张开玉,邵康一,卢迪.MSER快速自然场景倾斜文本定位算法[J].哈尔滨理工大学学报,2019,24(2):81-88.
- [5] 邢延超,程雷雷,李瑞,等.基于MSER和NMS的变形文档字符检测[J].科学技术创新,2018(32):101-102.
- [6] 钱江,张桂荣,姚江,等.最稳定极值区域与边缘增强的船名定位方法[J].计算机应用与软件,2019,36(2):264-268.
- [7] ZHOU Xinyu, YAO Cong, WEN He, et al. EAST: an efficient and accurate scene text detector[C]//IEEE conference on computer vision and pattern recognition. Honolulu, Hawaii, USA: IEEE, 2017: 2642-2651.
- [8] 孟健,曾宪文,高桂革.基于改进EAST算法的电厂电气设备铭牌文字检测[J].现代计算机,2019(25):55-60.
- [9] SHI Baoguang, BAI Xiang, BELONGIE S. Detecting oriented text in natural images by linking segments[C]//2017 IEEE conference on computer vision and pattern recognition. Honolulu, Hawaii, USA: IEEE, 2017: 3482-3490.
- [10] TIAN Zhi, HUANG Weilin, HE Tong, et al. Detecting text in natural image with connectionist text proposal network[C]//14th European conference on computer vision (ACCV 2016). Amsterdam, The Netherlands: Springer International Publishing, 2016: 56-72.
- [11] REN Shaoqing, HE Kaiming, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[C]//29th annual conference on neural information processing systems. Montreal, Canada: [s. n.], 2015: 91-99.
- [12] OLAH C. Understanding LSTM networks. blog[EB/OL]. 2015-08-27. <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>.
- [13] 邓丹. PixelLink: 基于实例分割的自然场景文本检测算法[D].杭州:浙江大学,2018.
- [14] 刘礼文,俞弦.循环神经网络(RNN)及应用研究[J].科技视界,2019(32):54-55.
- [15] 顾思思,李文.基于VGGNet卷积神经网络的车辆颜色识别[J].湖南科技学院学报,2019,40(10):65-67.
- [16] JASZEWSKI M, PARAMESWARAN S, HALLENBORG E, et al. Evaluation of maritime object detection methods for full motion video applications using the PASCAL VOC challenge framework[C]//Video surveillance & transportation imaging applications. [s. l.]: International Society for Optics and Photonics, 2015.