

文章编号: 2095-2163(2020)10-0037-07

中图分类号: TP391.1

文献标志码: A

基于循环卷积神经网络的模块化文字识别

张焱¹, 郭梦琰¹, 王峰¹, 邱雄¹, 贺桢¹, 蔡立志², 张娟¹

(1 上海工程技术大学, 电子电气工程学院, 上海 201620; 2 上海计算机软件技术开发中心, 上海 200235)

摘要: 为了解决光学字符识别(OCR)过程中,因文字采集质量偏低导致文字识别精度低的技术问题,提出了一种将传统方法与神经网络相结合的OCR技术。本文针对图像中的文字待识别区域进行文字识别:首先将待识别图像处理成无损位图文件;随后将无损位图文件进行方向校正、去噪、字符切割等预处理操作;最后基于预处理后的文本图像进行文字识别。实验表明,本文提出的方法,降低了OCR系统处理数据的负荷、提升了识别精度。不仅节约了时间成本和硬件成本,而且可以有效的识别文字图像中的密集文字和模糊文字。

关键词: 文字图像;卷积神经网络;图像处理

Modular character recognition based on convolutional recurrent neural network

ZHANG Yan¹, GUO Mengyan¹, WANG Feng¹, QIU Xiong¹, HE Zhen¹, CAI Lizhi², ZHANG Juan¹

(1 School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai 201620, China;

2 Shanghai computer software technology development center, Shanghai 200235, China)

【Abstract】 To solve the technical problem of low accuracy of character recognition caused by the low quality of character collection in the process of optical character recognition (OCR), this paper proposes an OCR technology that combines traditional methods with neural networks. This article aims at character recognition on the areas in the text image: firstly, the text image is changed into a lossless bitmap file in which a series of preprocessing should be done, such as direction correction, denoising, and character segmentation; then the file can be identified. Experiments show that the method we proposed reduces the data processing load of OCR recognition and improves the recognition accuracy. It not only saves time and cost of hardware, but also can effectively recognize dense and fuzzy text in text images.

【Key words】 Text image recognition; Convolutional neural network; Image processing

0 引言

随着经济全球化发展和信息高速化交流时代的到来,计算机技术、人工智能技术得到快速且蓬勃的发展。人们处理文字信息的频率也飞速提高,人工智能技术已广泛应用于生产、生活的方方面面。在生活工作节奏日渐加快的同时,数字化信息渐渐被青睐,许多场所对票据管理与录入的需求日渐提升。

原始的手动整理票据方法费时费力,容易出现错误。在人工票据数字化过程中,由于工作人员扫描熟练程度不高、失误操作等各方面原因,扫描后的票据图像往往会存在倾斜、倒置等问题。直接处理这些“问题”图像,将影响文字图像识别的效率与精度,后续针对数字化成果的使用也会造成一定的困扰。因此,采用人工智能方式,自动化处理票据数据的录入、校验工作具有较大的市场需求。税单、票

据等属于私人信息,通过离线方式识别将大大避免隐私数据外泄。此外自动识别票据内容,不仅可以帮助工作人员大大提高工作效率,同时也有助于优化公司整体的经营,压缩经济支出。

目前,光学字符识别(Optical Character Recognition, OCR)技术^[1-6],能够很好的识别各种背景较清楚的印刷体文本图像。其处理的主要流程为:

- (1)待识别图像中文本区域定位;
- (2)待识别区域文字切割;
- (3)文字识别。

常见的OCR技术虽然能处理页面清晰的印刷文本图像,但是当票据模糊或者是图像噪声较大,难以直观识别图片中的文字时,如果缺失图像去噪等预处理操作,就很可能影响识别精度。

现如今,特定票据的识别工作已经成为税务工

作者简介: 张焱(1991-),男,硕士研究生,主要研究方向:图像处理、深度学习、计算机视觉;郭梦琰(1993-),女,硕士研究生,主要研究方向:图像去雾算法;王峰(1994-),男,硕士研究生,主要研究方向:图像去模糊、图像超分辨率;邱雄(1995-),男,硕士研究生,主要研究方向:三维点云的分类与分割;贺桢(1995-),女,硕士研究生,主要研究方向:计算机视觉;张娟(1975-),女,副教授,主要研究方向:计算机视觉、人工智能、软件测试;蔡立志(1972-),男,博士,主要研究方向:软件工程、信息安全。

通讯作者: 张娟 Email: zhang-j@foxmail.com

收稿日期: 2020-04-24

作中重要的一环,人工票据录入时,因工作量大等因素,常常会出现模糊或包含极大噪声的“问题”图像。为了提升“问题”图片的识别效率和精度,本文提出了基于特定票据的文本图像识别方法,以减少人工复核的工作量,并且极大提高识别效果。

1 相关工作

随着信息化进程的推进,文本数据信息化已有越来越多的应用场景。因此,基于文本图像识别的数字化任务也备受关注。数字化文本信息,对于文本数据的保存、传输、应用都有至关重要的作用。但是常见的文本数字化方法,大多采用清晰的印刷文本图像进行 OCR 识别,因此对于文本背景不清晰的图像来说,这将会大大降低识别精度。

目前,模块化 OCR 处理方法主要有手写体识别和印刷体识别两类。实现印刷字体图像识别的方法大致可以分为以下几类:基于调用百度文字识别接口实现的 OCR 方法、基于 EasyPR 接口实现的文本信息识别、基于传统算法实现文本图片的 OCR 识别和基于深度学习网络实现文本图片的 OCR 工作。

基于调用百度的文字识别接口实现文本图像的 OCR 技术^[8],是 OCR 识别中最常见的识别方式。虽然百度提供的接口能够实现高效的文字识别功能,但该方法的弊端也是十分明显。首先,该识别方式属于在线识别,对于一些需要离线的场合就无法使用;其次,百度接口对于识别次数也是有限制的,当需要批量、多次文本图像识别时,该方法也无法满足需求。

基于 EasyPR 接口实现的文本信息识别^[9],与基于百度的文字识别接口实现文本图像的 OCR 技术方法类似,通过提供的 EasyPR 接口进行文字识别。该方法接口只针对于车牌的文字识别,虽然无需庞大的文本字库,但是在日常的非车牌图像文本识别时,有很大的局限性。

基于传统算法实现文本图片的 OCR 工作^[10],是通过光学技术,检测字符亮、暗模式进行识别字符的形状。这种方法的缺点是,当检测的文本图片中存在大量的底纹或者是边框图案时,识别效果将大打折扣。

基于深度学习网络实现文本图片的 OCR 工作^[11],通过深度学习的注意力,增强网络实现文本图像的字符识别。但识别精度较低,且识别效果一般,无法作为最优的 OCR 识别技术进行推广。

常见的 OCR 技术虽然能够识别背景较清楚的印刷体文本,但是对于待识别图片区域有底纹、模糊

或是图像噪声较大时,传统的文本位置定位以及分割单个字符将会遭遇重重困难。相比于仅采用传统算法的识别方式,本文提出的传统算法与深度学习相结合的识别方式,其识别结果更加准确。不但能处理简单文本图像,而且能完美处理复杂文本图像。

2 算法描述

目前,文字图像的模块化识别方法大都基于传统方法。本文采用的是在传统算法基础上结合深度学习中的神经网络技术,实现文字识别工作,该方法有助于提高识别的效果和准确度。简单来说,本文采用的神经网络技术为卷积循环神经网络(CRNN),网络的输入为传统算法预处理的去噪图片,输出是数字化文本信息。神经网络在 OCR 操作中主要充当了字符分类器的作用,基于深度学习的文字图像识别网络在文字识别精度方面优于现有的算法且更具有商业价值。

本文算法步骤描述如下:

首先,本文通过传统算法实现单元格定位来进行数据预处理。其中包括:加密文件解密、图像方向校正、格式转换、灰度化和二值化等相关操作。由于输入数据的多样性,直接训练输入数据可能导致识别的错误率增加。

其次,将预处理之后的数据输入到 CRNN 中。先通过卷积层提取待识别图像的特征图,然后将特征图按列切分成 4 块。每块通过循环层和 softmax 函数处理并列元素对应字符的预测概率,将预测结果与类标近似软对齐,最终得到完整识别结果。

本文采用的 CRNN 是将传统算法的预处理部分与深度神经网络的 OCR 识别部分相结合,实现文字图像的识别功能。此外,相比于使用 SVM^[7](支持向量机)、逻辑回归等文字识别算法,本文的识别方法正确率将显著提高,泛化性以及鲁棒性也会得到提升。

2.1 基于 SVM 的传统算法识别

在采用 OCR 技术的识别过程中,需要对数据集进行预处理操作,其处理过程如图 1 所示。首先对有加密的图像文件进行解密(输入为图片时则跳过该步骤);其次对文本图像进行去噪处理,采用图像灰度化、二值化等操作,使得数据集高度标准化;最后把方向歪置的数据集图像进行方向校正,即可得到待识别的文本图像的预处理结果。输入文本图像的方向校正示意如图 2 所示。

公式(1)为 RGB 文本图像灰度化公式;公式(2)为图像二值化过程中,求取最佳阈值的目标函数。

$$Gray = (R \times 30 + G \times 59 + B \times 11 + 50) / 100, \quad (1)$$

$$g(t) = w_0 * (u_0 - u)^2 + w_1 * (u_1 - u)^2. \quad (2)$$

其中, w_0 为前景点占图像比例, 均值为 u_0 ; w_1 为背景点占图像比例, 均值为 u_1 ; 整个图像的均值为 $u = w_0 * u_0 + w_1 * u_1$ 。当 $g(t)$ 最大时, 对应的 t 即为最佳阈值, 然后进行二值化操作。



图 1 预处理流程图

Fig. 1 Preprocessing flow chart

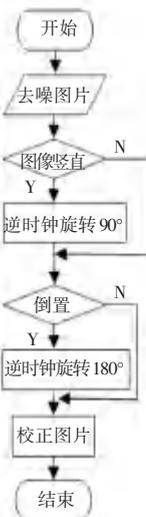


图 2 输入文本图像的方向校正示意图

Fig. 2 Schematic diagram of direction correction of input text image

Step 2 密码数据进行解密, 若为图片则跳过解密。

Step 3 图像预处理操作以及灰度化、二值化、去噪、方向校正的处理。

Step 4 角点检测, 切割出待检测文本图像。

Step 5 遍历所需识别文本, 并进行字符定位和切割。

Step 6 切割的字符块图像循环 OCR 操作。

Step 7 输出 excel 格式的文本识别结果。

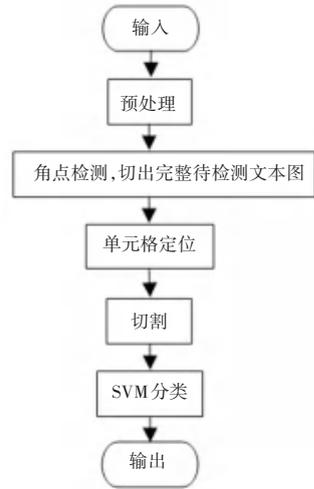


图 3 基于 SVM 的传统算法流程图

Fig. 3 Flow chart of traditional algorithm based on SVM

2.2 基于深度学习算法识别

通过对数据集的预处理, 获得了高度标准的文本图像数据; 接下来对文本图像进行角点检测, 把图像周围冗余的像素点切割掉; 通过单元格定位的算法 (YOLOv3、模板匹配等), 找到待识别文本在图像中的位置, 将位置信息存储在变量中; 再遍历存储的变量进行切割和 SVM 识别, 得到输出文本。基于 SVM 的传统算法的识别过程如图 3 所示。本文采用图像处理 OPENCV 库中 FindContours() 函数来进行角点检测。该函数传入为二值图像, 输出为检测到图像中物体最外围轮廓的坐标。具体实现过程描述为: 首先利用形态学变换——腐蚀膨胀等操作进行图像增强; 再用水平线核和垂直线核, 分别提取出文本图像中的横线和纵线, 取横线和纵线的交点为表格的四个角点, 从而定位所需文本图像; 然后利用 python 的 PIL 库, 切割出整体的轮廓, 便于后续文本图像的识别。

在基于深度学习的 OCR 任务中, 识别过程分为两步: 单字切割和分类任务。一般情况下, 人们都会将一连串文字利用投影法先切割出单个字体, 再送入 CNN 网络进行文字分类。而本文采用了一种基于深度学习的端到端文字识别方法 CRNN——不需要显式的利用传统算法来进行文字切割, 而是将文字识别问题转化为序列学习问题^[12]。虽然输入的图像尺度不同、文本长度不同, 但是经过 CNN 和 RNN 的循环处理, 并且在输出阶段采用翻译层的方法, 就可以对整个文本图像进行识别。也就是说, 文字的切割也被融入到深度学习中。如此一来, 就可以避免在文字切割环节中出错, 从而影响识别结果。

当获得四个角点坐标和切割好的轮廓图后, 再进行单个字符切割。这样就避免了 SVM 识别混淆出错的情况, 然后使用 SVM 进行文字识别。

通常语音识别问题可视为时序依赖的词汇或者短语识别问题, 这与 OCR 技术不谋而合。本文采用时序分类 (CTC) 的方法训练循环神经网络 (RNN)、构建 CRNN 算法。时序分类有助于获取字符信息间的关联特征, 且易于剔除错误文字关联。采用的 CTC 损失函数, 可增强神经网络对文字片段的理解力, 提升算法中翻译层的翻译效果。CRNN 的网络

基于 SVM 的传统识别算法步骤如下:

Step 1 输入待识别文本文件 X。

结构如图 4 所示。

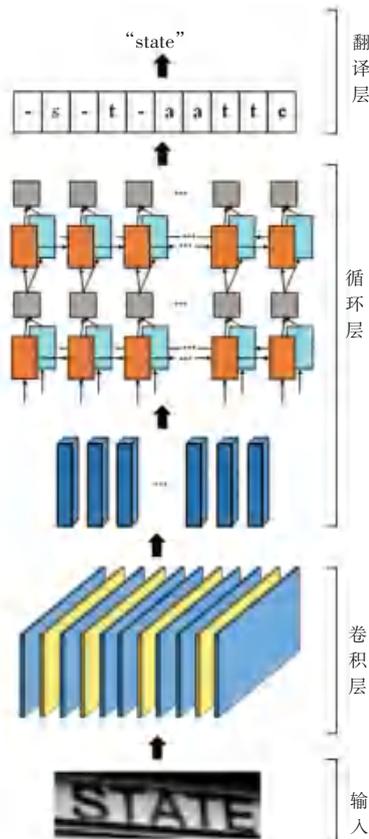


图 4 CRNN 网络结构图

Fig. 4 CRNN network structure diagram

通常解决不定长语音序列问题的实现过程为:将语音片段等声学特征,作为语音长短期记忆网络的输入,并且采用 CTC 作为损失函数提升语音识别精度。但在 OCR 文字图像识别过程中,输入为预处理的无噪声或轻噪声图片,因而需要通过 CNN 卷积网络提取浅层特征后,再输入长短期记忆网络,实现字符信息时序关联。这样不仅有助于神经网络进行图像中的特征提取,也利于实现序列化工作。此外,在传统的 OCR 识别方式中常需要进行单个字符切割后,才能进行单个字符识别。但通过本文的 CRNN 方法识别文本图像时,无需额外的算法进行切割工作,这将大大提升网络的稳定性与识别的准确性。

由图 4 可见,算法流程描述为:

首先,已预处理好的轻噪声图像被缩放到 $32 \times 100 \times 1$,输入 CRNN 网络;通过 7 层基于 VGG16 的卷积层,提取文本图像的浅层特征 ($1 \times (W/4) \times 512$)后,将特征块按列进行切割;将切割好的特征列输入双向 LSTM 中 (LSTM 参数 $T = (W/4)$, $D = 512$);经过 LSTM 后特征列变为长度为 $T \times n$ 的向量

(n 表示为文字类别);最后经过 softmax 函数处理,列向量每个元素代表对应的字符预测概率,再将 T 个预测结果通过 CTC 损失函数的指导进行识别,最终得到一个正确识别结果。

此外,CTC 为了解决二义性,提出了插入 blank 机制。如以“-”符号代表 blank。多种不同的字符排列路径均可归纳为一个识别结果,这有助于算法解决“重复识别”问题、增加泛化能力。比如字符串“state”均可由多种不同字符排列归纳获取:

$$\begin{aligned} B(\pi_1) &= B(- - stta - t - - - e) = state, \\ B(\pi_2) &= B(sst - aaa - tee -) = state, \\ B(\pi_3) &= B(- - sttaa - tee -) = state, \\ B(\pi_4) &= B(sst - aa - t - - - e) = state. \end{aligned} \quad (3)$$

从直观上理解就是,输出一张文本图像到网络中,需使输出为标签 l 的概率最大化。由于路径之间是互斥的,对于标注序列,其条件概率为所有映射到它的路径概率之和,即:

$$p(l|x) = \sum_{\pi \in B^{-1}(l)} p(\pi|x). \quad (4)$$

其中, $\pi \in B^{-1}(l)$ 的意思是可以合并成 l 的所有路径集合。通过对概率的计算,就可以通过神经网络进行反向传播更新。类似普通的分类,CTC 的损失函数 O 定义为负的最大似然。为了计算方便,对似然取对数则为:

$$O = -\ln\left(\prod_{(x,z) \in S} p(l|x)\right) = -\sum_{(x,z) \in S} \ln p(l|x). \quad (5)$$

3 实验及结果分析

本文进行了基于 SVM 的传统算法识别、经典的 LeNet 网络识别以及 CRNN 网络识别的对比实验。其中 CRNN 识别方案简述为:先用基于 SVM 预处理部分对数据集进行预处理,再将预处理结果送入 CRNN 算法进行识别工作。

3.1 实验过程

本文采用 MSRA-TD500 数据集进行训练。MSRA-TD500 数据集采用的是袖珍相机,进行实际场景拍摄的图片。其中包括标牌、门牌、警示牌等带文字标志的图片。由于是实际场景拍摄所得,所以图形背景更加复杂,对文字识别能力要求更高,更具有挑战性。

(1) 基于 SVM 的传统算法识别。首先对数据集进行预处理(其中包括:加密文件解密、方向校正、图像灰度化和二值化等);其次对预处理结果进行角点检测,切割出待检测文本图像;然后对所需识

别的内容进行定位切割,将定位结果存储在变量中;最后使用 SVM 进行 OCR 操作,输出最终所需识别文本。

(2)基于 LeNet 算法识别。首先收集字体文件,并对每一个汉字进行标签标注;此外还需适当的数据增强,如:旋转、平移等;最后将数据集送入网络进行文字多分类任务,即文本识别过程。

(3)基于深度学习的 CRNN 算法识别。首先对数据集进行预处理操作(采用传统方法);将预处理的结果送到 CRNN 算法中进行训练;经多次迭代训练后,最终得到识别结果。

SVM、LeNet、CRNN 这 3 种识别算法的执行步过程如图 6 (a)、(b)、(c) 所示。基于 SVM 的传统算法重要参数配置说明见表 1。CRNN 算法中长短期记忆网络的隐藏单元数、特征图大小、卷积核大小、移动步长等参数见表 2。

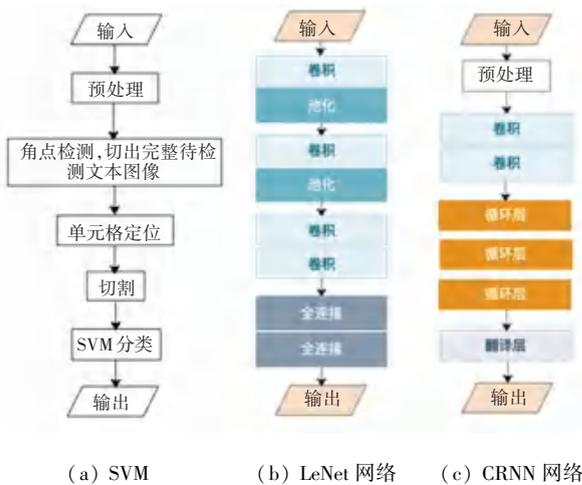


图 5 3 种识别算法执行流程

Fig. 5 Three recognition algorithm execution process

表 1 基于 SVM 传统识别算法参数描述

Tab. 1 Parameter description of traditional recognition algorithm based on SVM

参数	描述
-d degree	设置核函数中的 degree (默认 3)
-g gamma	设置核函数中的 gamma (默认 1/k)
-c cost	设置 C-SVC, epsilon-SVR 和 nu-SVR 的参数 (默认 1)
-n nu	设置 nu-SVC, SVM 和 nu-SVR 的 nu 参数 (默认 0.5)
-p epsilon	设置 epsilon-SVR 中的损失函数 p (默认 0.1)
-m cachesize	设置内存大小 (默认 100 MB)
-e epsilon	允许终止的依据 (默认 0.001)
-h shrinking	是否使用收缩启发式, 0 或 1 (默认 1)
-wi weight	设置参数 C 为 weight * C (默认 1)

表 2 基于深度学习的 CRNN 各网络层及其配置

Tab. 2 Deep learning based CRNN network layers and their configuration

网络层	配置
CTC	Transcription -
RNN	双向 LSTM 隐藏单元:256
	双向 LSTM 隐藏单元:256
	映射序列 -
CNN	卷积层 512, 卷积核:2 * 2, 步长:1, 填充:0
	池化层 核:1 * 2, 步长:2
	卷积层 512, 卷积核:3 * 3, 步长:1, 填充:1
	卷积层 512, 卷积核:3 * 3, 步长:1, 填充:1
	池化层 核:1 * 2, 步长:2
	卷积层 256, 卷积核:3 * 3, 步长:1, 填充:1
	卷积层 256, 卷积核:3 * 3, 步长:1, 填充:1
	池化层 核:2 * 2, 步长:2
	卷积层 128, 卷积核:3 * 3, 步长:1, 填充:1
	池化层 核:2 * 2, 步长:2
卷积层 64, 卷积核:3 * 3, 步长:1, 填充:1	

输入 32 * 100 灰度图像

3.2 实验结果分析

经过将 3 种网络结构通过 MSRA-TD500 的测试集进行测试,结果如图 6 所示。

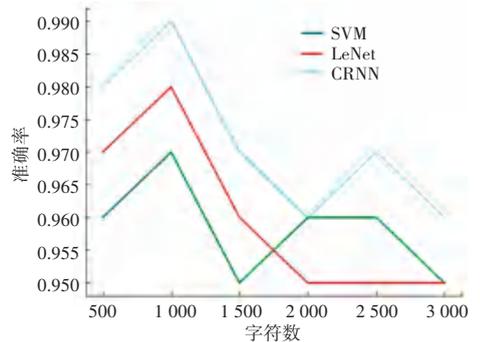


图 6 识别准确率对比

Fig. 6 Comparison of recognition accuracy

从图 6 可以看出,随着测试字符数量的增加, SVM、LeNet 网络、CRNN 网络之间识别准确率差异明显。当字符数少于 1 750 个时, LeNet 算法的识别准确率比 SVM 的识别准确率高,超过 1 750 个字符后, SVM 传统算法的准确率有少许提升,但效果仍不明显。

从图中可以看出,随着测试字符数量的增加,3 种网络之间识别准确率有较为明显的差异。在 1750 字符数以前, LeNet 算法比 SVM 的识别准确率高,而超过 1 750 字符数以后, SVM 传统算法的准确

率略有提升,但效果仍不明显。综合图6可见,本文提出的CRNN算法识别效果优异,具有较强的鲁棒性和泛化能力。当测试字符数量为3 000时,CRNN算法识别准确率最高。

本文提出的方法在实际场景中应用极为广泛,如

税单、发票、扫描文件识别等。智能识别操作可以省去大量的人工成本。图7为所需识别税单在预处理过程中,外围角点检测结果(绿色框为角点检测后,角点间的连线),然后切割出完整的待检测文本图像——即预处理结果,最后送入识别算法进行识别。

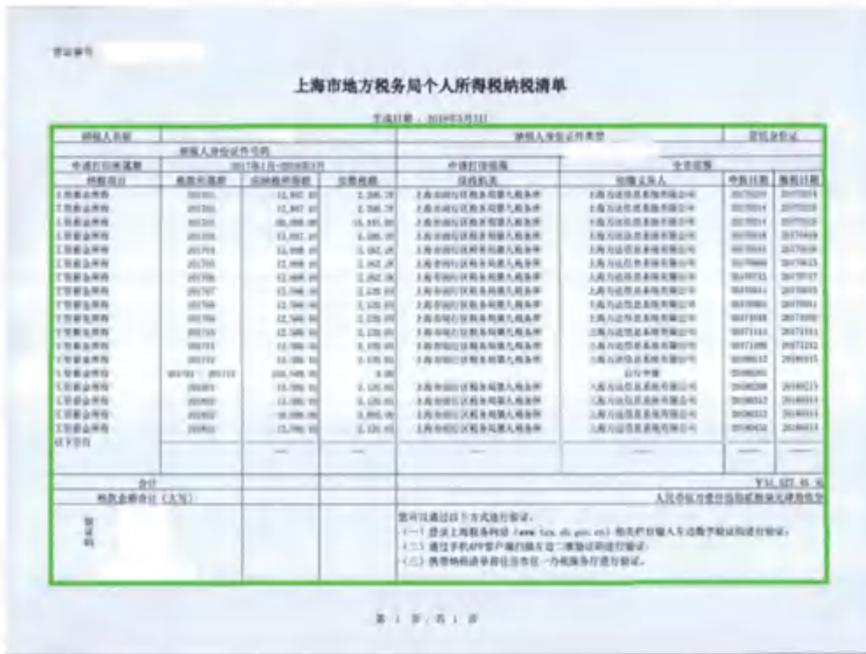
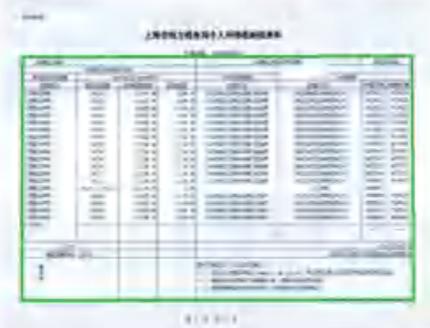


图7 所需识别税单图像示例

Fig. 7 Example of the required tax identification image



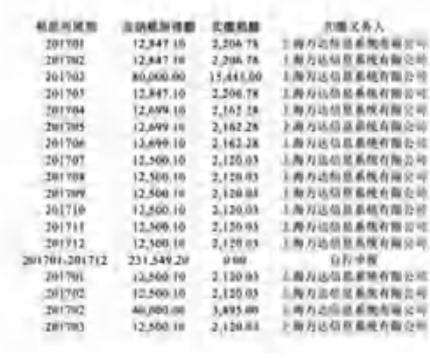
(1) 税单原图



(2) SVM 识别结果



(3) LeNet 识别结果



(4) CRNN 识别结果

图8 税单图像识别结果

Fig. 8 Results of tax bill image recognition

此应用场景采用某公司的税单作为实验测试数据。其中文件包括: 电子扫描文件、复印件二次扫描文件及 PDF 文件等, 文件大小在 500 K~2 M。图 8 是采用 3 种算法处理后的输出结果。

图中红色部分为识别错误的数据。综合实验结果可以看出, 本文方法 (CRNN 网络) 无论在公开数据集还是在实际的应用中都有较高的识别准确度。本文方法可以准确高效的识别税单中待识别内容, 不仅提高了文本图像数字化的质量和效率, 还避免了现有技术中由人工对文本图像进行检测造成效率低下、准确度差等问题。

4 结束语

本文提出了一种基于深度学习的卷积循环识别网络 (CRNN), 实现文本图像的文字识别工作。本文方法大大提升了文本图像的平均识别准确率, 并且占用较少的运算资源。(1) 由于传统算法的泛化能力和鲁棒性较差, 但切割效果显著;(2) 神经网络可以实现图像数据的有效特征提取并得到较深层的特征表示, 对数据的表达性更高。因此, 本文将二种方法的优点相结合不仅在公开数据集上效果明显, 在实际应用中也效果显著。

参考文献

[1] Optical Character Recognition; Report Summarizes Optical Character Recognition Study Findings from University Gottingen (Optical Character Recognition of Typeset Coptic Text With Neural Networks)

- [J]. Journal of Robotics & Machine Learning, 2020.
- [2] YELLAPRAGADA V, CHIANG P, MADDIKA S K. Optical Character Recognition (Ocr) Accuracy By Combining Results Across Video Frames[J]. 2018.
- [3] 汪惠权. 用于发票处理的数字图像处理系统的研究与开发[D]. 东华大学, 2018.
- [4] HOFMAN Y. Character recognition system and method for shipping containers[J]. 2018.
- [5] Miyagawa So, Bulert Kirill, Büchler Marco, et al. Optical character recognition of typeset Coptic text with neural networks [J]. Digital Scholarship in the Humanities, 2019 (Supplement_1); Supplement_1.
- [6] Natsuda Kaothanthong, Thanaruk Theeramunkong, Jinhee Chun. Improving Thai Optical Character Recognition Using Circular-Scan Histogram[C]// Iapr International Conference on Document Analysis & Recognition. IEEE, 2018.
- [7] SHANTHI N, DURAISWAMY K. A novel SVM-based handwritten Tamil character recognition system.[M]. 2010.
- [8] 张婷婷, 马明栋, 王得玉. OCR 文字识别技术的研究[J/OL]. 计算机技术与发展, 2020(03): 1-8 [2020-04-11]. <http://kns.cnki.net/kcms/detail/61.1450.TP.20191205.1113.040.html>.
- [9] 王学梅. OCR 文字识别系统的应用[J]. 现代信息科技, 2019, 3(18): 66-68.
- [10] 姜毅, 彭清涛, 徐娟. 基于图元识别的 OCR 文本图像矫正算法分析[J]. 中国新技术新产品, 2019, (2): 7-8.
- [11] 徐富勇, 余凉, 盛钟松. 基于深度学习的任意形状场景文字识别[J]. 四川大学学报(自然科学版), 2020, 57(2): 255-263.
- [12] SHI B, BAI X, YAO C. An End-to-End Trainable Neural Network for Image-Based Sequence Recognition and Its Application to Scene Text Recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(11): 2298-2304.

(上接第 36 页)

图 10 为原始 DCGAN 经过 5 000 次迭代训练后生成的织物图案, 图案比较模糊, 存在棋盘效应。图 11 为通过本文所用 NN-DCGAN 模型生成的织物图案, 在视觉上比较平滑, 图像质量相比原始 DCGAN 有所提高。本文不能证明以上两种方法是上采样的最终解决方案, 但其确实对修复棋盘效应起到一定的作用。

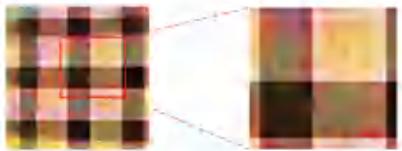


图 10 DCGAN 生成样本图

Fig. 10 Sample generated by DCGAN

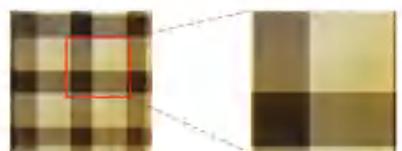


图 11 NN-DCGAN 生成样本图

Fig. 11 Sample generated by NN-DCGAN

4 结束语

本文改进了 DCGAN 的模型结构, 使用缩放卷积代替原模型中的反卷积操作。实验结果表明, 该算法可有效解决反卷积操作带来的棋盘效应问题, 同时保证了生成图像的质量。此外, 该模型在格子织物图案样本集上的实验结果显示, 生成对抗网络可用于织物图案的生成工作, 模拟设计师完成创造性任务, 对织物图案的快速创新设计具有重要的应用价值和意义。

参考文献

- [1] 林宙. 中国纺织品服装出口问题研究[D]. 暨南大学, 2005.
- [2] 黄远, 白琼, 李宏凯, 等. 基于条件生成对抗网络的图像描述生成方法[J]. 计算机辅助设计与图形学学报, 2020, 32(6): 911-918.
- [3] 邹秀芳, 朱定局. 生成对抗网络研究综述[J]. 计算机系统应用, 2019, 28(11): 1-9.
- [4] 刘柏森, 邓琛, 张雾琳. 一种改进深度卷积生成对抗网络的人脸分割方法[J]. 黑龙江工程学院学报, 2019, 33(5): 1-6.
- [5] 王万良, 李卓蓉. 生成式对抗网络研究进展[J]. 通信学报, 2018, 39(2): 135-148.
- [6] ODENA A, DUMOULIN V, OLAH C. Deconvolution and checkerboard artifacts[J]. Distill, 2016, 1(10): 3.