一种基于深度学习的增值税发票文字识别模型*

李陶深　张浩晨

广西大学计算机与电子信息学院，南宁，530004

摘　要　基于 DenseNet + GRU + CTC 网络模型，构建一种基于深度学习的增值税发票文字识别模型。该模型利用网络上开源的中文文档文字识别数据集来对 DenseNet + GRU + CTC 网络进行训练，得到最终的文本识别模型，并用于实现增值发票文本文字的识别。然后，设计并编码实现了一个增值税发票自动识别样机系统，该系统可以利用得到的模型对文本定位后的区域进行自动识别、结果分析和整理。性能测试结果说明了模型的可行性和可用性。

关键字　增值税发票，自动识别，DenseNet ，　GRU，光学字符识别

A Text Recognition Model of VAT Invoice Based on Deep Learning

Li Taoshen　　Zhang Haochen

School of Computer, Electronics & Information
Guangxi University
Nanning 530004,China
tshli@gxu.edu.cn

Abstract—Based on the DenseNet + GRU + CTC network model, a text recognition model of value-added tax (VAT) invoice based on deep learning is constructed. In this model, the DenseNet + GRU + CTC network is trained by using the open source chinese document text recognition data set on the network to obtain the final text recognition model. Then, a VAT invoice automatic recognition system is designed and coded, which can use the obtained model to automatically identify, analyze and sort out the text location area. The performance test results show the feasibility and availability of the model.

Key words—Value-added tax(VAT) invoice, automatic recognition, DenseNet, GRU, Optical character recognition（OCR）

1　引　言

随着计算机在社会生活中的全面普及，社会网络信息化的高速发展，无纸化办公的风潮愈掀愈热，计算机在财务管理上的应用也日趋普及。电子增值税发票因其更方面传播、更容易进行识别、不易损坏票面等优点，更容易进行增值税发票报销、存储以及管理等事宜。

增值税发票自动识别是一个不断更新、不断发展的研究方向，对增值税发票的不同要求，就可以开发出许多不同针对性的识别系统。但是对于所有的发票识别系统来说，最主要最迫切的就是提高文本的识别准确率和识别速度，以此最大程度地增加财务工作人员的效率，而不是将财务工作者的时间花费在将增值税发票通过手打的方法录入到计算机，然后再去整理管理账目。增值税发票自动识别系统可以帮助企业更加合理化的分配资源，提高生产效率，对于企业的财务和人员的管理都具有长远深久的影响。

美国、英国、法国等发达国家在十九世纪 80 年代就已经开始了关于增值税发票文本识别的研究，设计实现了不少应用系统，例如美国的 AcuForm 识别系统，法国的 A2iA 发票识别系统等。目前，各个国家的银行以及各大金融公司所应用的系统主要涉及增值税发票信息识别、检测、存储等方面[1]。

尽管我国在增值税发票自动识别和管理方面的研究与技术应用相对较晚，但是随着增值税发票的业务不断展开，以及我国在 OCR 文字识别技术研究方面取得的突破性进展，我国在增值税发票自动识别系统的研发与应用也取得了良好的进展[2]。文献[3]基于深度学习和图像处理算法,设计并实现一套能准确识别数字信息的发票自动识别系统.文献[4]基于国家发布的增值税发票查询网站, 提出了基于 Selenium 技术来是实现自动化网站查询，设计与实现了一套金融发票自动识别系统。文献[5]和[6]以发票文字检测和文字识别为核心，设计与开发了基于深度学习的发票识别系统，提高了发票的识别效率。文献[7] 基于 OCR 技术开发了一套发票自助识别校验系统，实现了供应商自助发

*基金资助：本文得到广西高等教育本科教学改革工程项目一般项目(2020JGA116)资助.

**通讯作者：李陶深，男，教授，tshli@gxu.edu.cn

票扫描、发票验审管理、发票校验管理、影像管理、发票状态管理及查询等功能。文献[8]设计了一款基于卷积神经网络的增值税发票自动识别系统，通过两个作用不同的卷积神经网络，用于检测目标区域和获得目标区域的具体字符信息。文献[9]以市场交易中常见的票据作为研究对象，探索票据图文识别的研究方法，并将检测与识别算法流程串联起来，实现了一个票据图文端到端的识别系统。

目前，利用 OCR 文本识别技术对增值税发票进行自动识别是增值税发票自动识别系统开发中常用的技术手段[2]。一些研究人员也开始将深度学习算法和技术用于增值税发票的文字识别，但是成果不是很多。为此，本文利用基于深度学习的网络架构模型来实现增值发票文本文字的识别，并且根据网络上开源的中文文字数据集在进行训练，实现增值税发票的文本识别功能。

2 基于深度学习的增值税发票文字识别

在对增值税发票进行的 OCR 文本检测识别中，最重要的任务是文本识别，且文本识别的准确率关乎着整个增值税发票检测识别系统的最终结果。在财务工作人员所要记录、管理的重要信息中，有汉字、符号、英文、数字以及特殊字符，难度相对较高。所以需要用深度学习算法来对增值税发票中的选框范围进行框定，然后在进行 OCR 文字识别。本文所采取的是 DenseNet + GRU +CTC 网络来对增值税发票中的文本信息进行文字识别。

2.1 DenseNet +

在深度学习领域，CNN（卷积神经网络）已经成为了最受欢迎的深度学习方法之一，尤其是 ResNet 模型的出现，更是将卷积神经网络推到了计算机视觉领域的前沿。ResNet 模型是通过在前面层和后面层之间建立“短路连接”，从而可以训练出比以往更深的 CNN 网络[10]。与之略有相同点的是 DenseNet 模型，该模型是在每两层之间建立连接，这种密集链接方式极大地减少了模型的参数与计算成本，减轻了梯度消失影响，加强了 feature 的传递效果和特征重用，显示出更加优秀的特性[11]。

在传统的卷积神经网络中，一个L层的网络中会有L个连接，但是随着网络深度的不断加深，梯度消失的问题就越来越明显。尽管人们提出了很多解决方法，但是都是以在前面层和后面层之间建立更短的连接路径为核心。在DenseNet中，每个层都会与其后面的所有层在channel维度上进行连接，一个L层的网络中的连接数量多达L(L+1)/2个，特征传递方式是直接将前面所有层的特征concat后传到下一层，而不是前面层都要有一个箭头指向后面的所有层。其表达式如下:

$$X_l = H_l([X_0, X_1, X_2, ..., X_{l-1}]) \quad (1)$$

其中，$H_l()$代表着非线性转换函数，其中包含一系列BN、ReLU、Pooling及Conv操作。这种连接方式以及输入输出之间的特殊关系，使得梯度和特征的传递更加有效。

在 DenseNet 模型中有多个DenseBlock 结构，DenseBlock结构图如1所示。

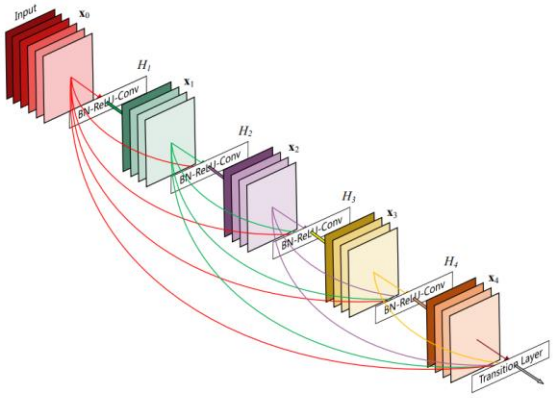

图 1 DenseBlock 结构图

DenseNet模型将总体的结构分为三个结构，分别是：模型中的最基础单元DenseLayer，用于对特征的提取；最重要的密集连接单元，进行特征的运算；在两个密集连接中的过度单元Transition。通过对这三种结构的连接叠加重复，就可以获得整个完整的DenseNet模型。

BN层（Batch Norm），在整个模型中用于加快模型训练的收敛速度，避免出现梯度消失梯度爆炸，提高训练过程的稳定性。RelU激活函数层是一个激活函数，通过RelU函数使得卷积中将负值去除，保留正值。1×1Conv层是在不改变矩阵性质的情况下，添加非线性特性，对通道数降维，减少运算量。3×3Conv层用于在卷积层中提取细小特征。2×2AvgPool层用于降低特征层宽度，降低至1/2。

在这三种核心结构中，DenseLayer层包含BN + Relu + 1×1Conv + BN + Relu + 3×3Conv，其中第i个DenseLayer层中的第一个1×1Conv层的输入通道层数是num_input_ features+(i-1) ×growth_rate，输出通道层数为bn_size×growth_rate；第二个3×3Conv的输入通道数为bn_size×growth_rate，输出通道数为growth_rate。在DenseLayer层内的特征宽度不会改变，所以stride≠2并且没有池化的情况。

密集连接DenseBlock模块就是一堆DenseLayer层

的堆合，在DenseBlock中的所有DenseLayer之中会发生密集连接。在密集连接中必须保证特征宽度不变

Transition 模块包含 BN + Relu + 1×1Conv + 2×2AvgPool，用于对不同的 DenseBlock 模块进行连接，并进行压缩特征。

2.2 GRU 网络

循环神经网络(RNN)可以对序列数据进行精确的建模，是在原始神经网络中的基础上增加一个记忆单元，并且在序列数据中记录相关信息，比普通神经网络更加适用于解决序列学习问题[11]。RNN是指将同一个网络进行多个叠加的结构，并且每一个网络可以将信息传递到下一个存储单元，运行原理是RNN基本单元的单元模块使用前一时刻的输入 X_t 和状态值 y_{t-1} 来获得该时刻的输出值 y_t [12]。

GRU网络（Gate Recurrent Unit）是RNN循环神经网络中的一种，是LSTM的一种变体网络。与LSTM网络相比，GRU网络的结构更加简单，效果更好，适用于解决RNN网络中的长依赖问题[13]。通过GRU网络来对特征进行保留，可以解决标准RNN网络的梯度消失问题。在LSTM网络中有三个门函数：输入门、输出门、遗忘门，图2给出了LSTM的模型结构图。GRU则是对LSTM网络进行了变形，简化了门控单元的数量，只有更新门和重置门。GRU的具体格式如图3所示。

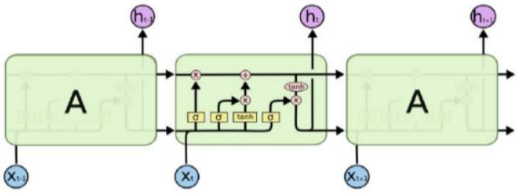

图 2 LSTM 模型结构图

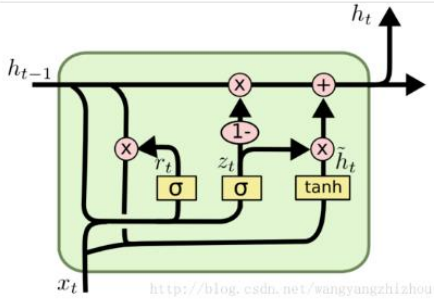

图 3 GRU 具体结构

图3中的 Z_t 和 R_t 分别表示着更新们和重置门，更新门接收当前状态 x_t 和先前隐藏状态 h_{t-1}，接收到输入信息和矩阵运 算后，sigmoid函数确定神经元是否被激活。更新门是将LSTM中的忘记门与输入门合成所得，指将控制前一瞬的信息带入到现在状态的程度，更新门的数值越大，说明前一刻的状态信息所被带入的越多[14]。重置门指控制前一瞬的信息被录入到当前的候选集$\hbar$t上，重置门的数值越小，表示前一瞬间的信息被写入的越少。GRU单元通过下列四个公式来进行数据计算：

$$Z_t=\sigma(W^z x_t+U^z h_{t-1}) \tag{2}$$

$$r_t=\sigma(W^r x_t+U^r h_{t-1}) \tag{3}$$

$$\hat{h}_t=\tanh(r_t\odot U^r h_{t-1}+ Wx_t) \tag{4}$$

$$h_t=(1-z_t)\odot\hat{h}+ z_t\, h_{t-1}) \tag{5}$$

式中，W^z 和U^z 是更新门的权重；W^r 和U^r 是重置门的权重；W 和U 是形成防止内存的网络权重；σ是sigmoid函数；tanh是双曲正切函数；算子⊙表示向量的内积。

LSTM网络与GRU网络都是使用各种Gate来对重要特征进行保留，保证这些特征在接下来的传输中不会被流失。图4给出了实验训练模型的大小，图5则给出了训练集和测试集的平均负对数概率[15]。

Unit	# of Units	# of Parameters
Polyphonic music modeling		
LSTM	36	$\approx 19.8\times10^3$
GRU	46	$\approx 20.2\times10^3$
tanh	100	$\approx 20.1\times10^3$
Speech signal modeling		
LSTM	195	$\approx 169.1\times10^3$
GRU	227	$\approx 168.9\times10^3$
tanh	400	$\approx 168.4\times10^3$

图 4 实验训练模型的大小

			tanh	GRU	LSTM
Music Datasets	Nottingham	train	3.22	2.79	3.08
		test	**3.13**	3.23	3.20
	JSB Chorales	train	8.82	6.94	8.15
		test	9.10	**8.54**	8.67
	MuseData	train	5.64	5.06	5.18
		test	6.23	**5.99**	6.23
	Piano-midi	train	5.64	4.93	6.49
		test	9.03	**8.82**	9.03
Ubisoft Datasets	Ubisoft dataset A	train	6.29	2.31	1.44
		test	6.44	3.59	**2.70**
	Ubisoft dataset B	train	7.61	0.38	0.80
		test	7.62	**0.88**	1.26

图 5 训练集和测试集的平均负对数概率

从图4与图5中可以看出，GRU的构造比LSTM少了一个gate，在训练时可以比LSTM节省时间，所以本文采用GRU网络来对经过DenseNet网络的序列数据及逆行处理。

2.3 CTC 算法

CTC(Connectionist Temporal Classification)是一种自动对齐的方式，非常适合于在OCR识别上面对文字进行对齐。因为在深度学习OCR文字识别时，需要对

文字进行分割识别，每个文字的大小宽度会有不同，所以需要使用CTC算法，来对根据固定文本框识别的信息进行提取。

CTC算法对输入到模型中的一个输入序列X，将关于这个序列X的输入时间步以及输出时间步都设为T，通过输入输出时间步T得出相互对应的输入与输出，对每个时间步T对应的序列片段X_i进行识别，从而给出该时间步内片段识别出来的文字概率集合。然后，对这些概率进行选取，就可以得到可能的输出分布，根据这个输出分布将概率最大的结果输出。损失函数是指输入的序列X，求出最大化Y后导出P(Y|X)，P(Y|X)是可以进行求导的，这样可以根据损失函数执行梯度下降算法。

在CTC算法的文本对齐中，因为会存在多个输出路径对应着一个输出结果的情况（如图6所示），将含有“CAT”的图片在CTC算法中将X分割为多个时间片来进行特征预测。如果遇到相连的两个相同字的时候，则会非常容易造成错误输出。

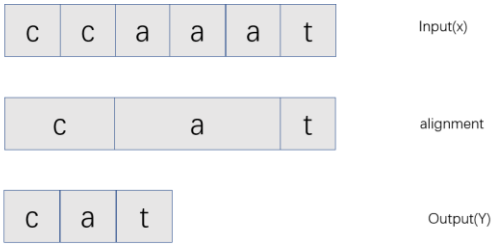

图 6 初始 CTC 算法对齐效果

为了解决初始CTC的错误识别隐患，可以在CTC算法中引进了空白字符ϵ，使得在OCR文字识别时对两个文字之间的字符间距进行识别，这样就不会对相邻的两个相同字符错误识别。为了能够在CTC算法中加入对空白字符ϵ的合并去除操作，可以做以下的改进：在CTC算法中定义一个B变换，对输出序列进行变换，变换成真实输出，比如对于下式中的state，把连续的相同字符删减为1个并删去空白符。具体实行过程如图4-13所示。

$$B(\alpha)=B(--stta-t---e)=state \tag{6}$$

因为CTC算法所得的结果是由遍历得来的，其时间复杂度是指数级的，因为有T个位置，每个位置有n种选择（字符集合的大小），那么就有n^T种可能。为了降低时间复杂度，CTC算法使用HMM中的前向后向算法来进行结果的运算。

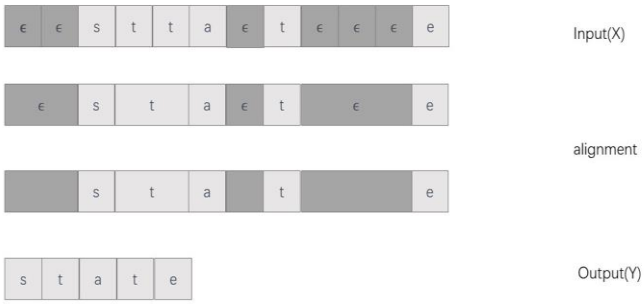

图 7 CTC 算法的改进对齐效果

在前向算法中，将长度为r的序列q，定义$q_{1:p}$和$q_{r-p:r}$分别作为序列q的前p标识和后p标识。对于一个字符序列l，定义它的前向变量为$\alpha_t(s)$为在时刻t时$l_{1:s}$的总概率，用公式表示为：

$$\alpha_t(s)=\sum_{\substack{\pi\in N^T\\ B(\pi_1:t)=l_{1:s}}}\prod_{t'=1}^{t}y_{\pi_{t'}}^{t'} \tag{7}$$

在字符序列l的最前面、最后面以及卡两个字符之间插入空格，获得新的标签序列l'，则l'的长度就是$2l+1$，为了能够计算l'前缀的概率，我们允许空格和字符之间可以转移，还有任何独立的字符之间可以转移。本文中定义所有的前缀可以从一个空格或者字符开始，也就说初始化可以定义为下列公式：

$$\begin{aligned}&\alpha_1(1)=y_b^1\\&\alpha_1(2)=y_{l1}^1\\&\alpha_1(s)=0,\forall s>2\end{aligned} \tag{8}$$

所以$\alpha_t(s)$可以通过公式(8)经过递推得到，得到的总公式如下所示：

$$\alpha_t(s)=\begin{cases}(\alpha_{t-1}(s)+\alpha_{t-1}(s-1))y_{l'_s}^t & if\ l'_s=\text{blank}\ \ or\ \ l'_{s-2}=l'_s\\(\alpha_{t-1}(s)+\alpha_{t-1}(s-1)+\alpha_{t-1}(s-2))y_{l'_s}^t & otherwise\end{cases} \tag{9}$$

在时刻T时，字符序列l的概率公式如下：

$$p(l\mid x)=\alpha_T(|l'|)+\alpha_T(|l'|-1) \tag{10}$$

在后向算法中，定义$\beta_T(s)$在时刻K时$l_{s:|l|}$的总概率，公式如下所示：

$$\beta_t(s)=\sum_{\substack{\pi\in N^T\\ B(\pi_{t:T})=l_{s:|l|}}}\prod_{t'=t}^{T}y^{t'}_{\pi_{t'}} \tag{11}$$

初始化可以定义为：

$$\begin{aligned}&\beta_T(|l'|)=y^T_b\\&\beta_T(|l'-1|)=y^T_{l_{|l|}}\\&\beta_T(s)=0,\forall s<|l'|-1\end{aligned} \tag{12}$$

所以 $\beta_T(s)$ 可以通过公式(12)经过递推得到，得到的总公式如下：

$$\beta_t(s)=\begin{cases}(\beta_{t+1}(s)+\beta_{t+1}(s+1))y^t_{l'_s} & if\ l'_s=blank\ \ or\ \ l'_{s-2}=l'_s\\(\beta_{t+1}(s)+\beta_{t+1}(s+1)+\beta_{t+1}(s+2))y^t_{l'_s} & otherwise\end{cases} \tag{13}$$

同时，本文也引入了Beam Search算法，以使得输出结果的合理性，而不是被干扰因素的存在或者识别准确率错误等导致CTC算法的错误输出。在模型中设置一个参数B，用B来对每次保留的前缀序列个数进行记录，在每个t时刻都选取出概率最大的B个字符，在对齐之后可能会有相同的输出结果，所以要将输出结果相同的前缀序列合并并且概率相加，挑出概率最大的三个结果作为下一次的输出，以此类推得出最终结果[16]。

CTC算法主要是将输入序列X进行分析文本的后验概率表示为所有表现为Y的路径之和，所进行训练的目的就是将后验概率P(Y|X)最大化。当P(Y|X)越大时，CTC算法的实现效果越好。

2.1 增值税发票自动识别系统的设计实现

基于连通域的文本定位方法和基于DenseNet + GRU + CTC架构的深度学习方法的文本识别方法，我们设计实现了一个增值税发票自动识别系统。该系统利用深度学习的方法来对拍照或者扫描获得的纸质增值税发票进行文本识别。具体实现时，主要通过研究三个深度学习算法，构建一个DenseNet + GRU + CTC网络架构，完成了整个的增值税发票OCR定位识别中的预处理工作、定位操作、文字识别操作，实现对增值税发票框选位置进行文字识别。

3 性能测试与结果分析

3.1 数据集的准备

增值税发票上的重要信息包含英文、数字、字符、汉字，所以需要使用汉字英文数字混合的数据集进行训练。本文选择公开的中文文档文字识别数据集作为训练集，该数据集约有 364 万张图片，按照 99:1 的比例被划分为训练集和检验集。该数据集中包含了中文语料库，该语料库是由字体、大小、透视、拉伸等方法获得的，共包含汉字、英文、数字字符、标点共 5990 个字符，其每个样本图片固定包含为 10 个字符。这种训练简化了数据规模，对数据集的训练十分友好，训练快速。

3.2 文本识别模型训练

完成了深度学习架构构建和OCR文字识别训练集的确定以后，即可开始进行模型训练，从而得到可以胜任对增值税发票中的候选框内的文本进行自动识别的深度学习模型。本文的实验设备是一台具有GeForce GTX 1050Ti 的 GPU的计算机，显存容量为8G。

因为所要使用的模型是DenseNet + GRU + CTC架构，所以先将这三个识别网络进行拼接封装到Train.py中。在模型训练时，将输入图像尺寸统一设置为280*32。DenseNet + GRU + CTC 架构在训练初始时的学习率为0.005，每次训练一轮后，就将学习率降低五分之三。并且在DenseNet + GRU + CTC 架构模型训练中时进行采用Adam方法进行优化。在训练时的标签长度只能设置为10，因为数据集中的每个图象均只有10个字符。

DenseNet + GRU + CTC 模型训练准确率曲线如图 8 所示，DenseNet + GRU + CTC 模型训练误差曲线如图 9 所示。从这两张图中可以看出，DenseNet + GRU + CTC 模型经过训练之后，准确率达到较高水平并且处于收敛状态，准确率为 95.36%，误差曲线降至最低处并且趋于稳定，说明了 DenseNet + GRU + CTC 模型文字识别的性能较好。

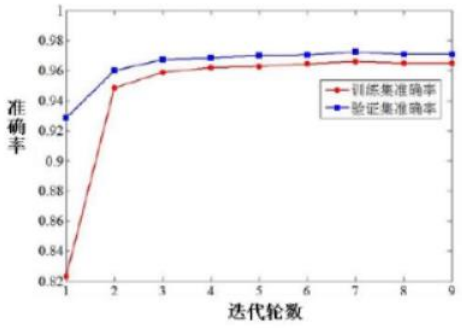

图 8 DenseNet + GRU + CTC 模型训练准确率曲线

3.3 性能测试与结果分析

利用本文提出的模型和技术方案，我们设计实现了一个增值税发票自动识别样机系统，该系统可以利用得到的模型对文本定位后的区域进行自动识别、结果分析和整理。本小节对设计实现的增值税发票定位与识别样机系统的性能进行测试，以说明模型和系统的设计是可行和有效的。测试的内容包括：文字识别

效果、文字识别运行速度等。实验时，主要是基于DenseNet + GRU + CTC架构的深度学习算法，对增值税发票文字候选框中的文本信息进行识别。

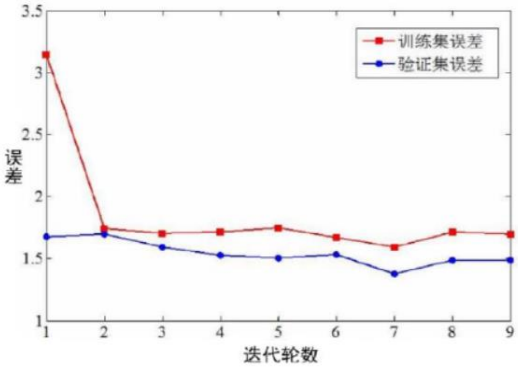

图 9 DenseNet + GRU + CTC 模型训练误差曲线

（1）文字识别结果

对图像的行文本选区域 3.7 进行基于 DenseNet + GRU + CTC 架构模型的深度学习文本识别，得到的信息如图 10 所示。

```
007368901300214130
00736890
2022年02月25日
6>0<><062+4.396<3/<37<802水米>
霸州市富立盛营金属制品有限公司
92185次-88>4/-1607450X/054准
91131081MA0FP63D9Y
48米893>2十>米1米/6<9<十48>27十/2
0316-7492222
2<水3903815/8+米8422733米水6645
河北省廊坊市霸州市堂二里镇七街红绿灯西 888米
中国农业银行般份有限公司霸州胜芳支行
50687301040006308
105338.16
810293.52
水黑色金属冶炼压延品中镀锌带钢395*0.9
4566, 3716814
Y105338.16
Y810293.52
Y915631.6 8
8玖拾壹万伍仟陆佰叁拾壹圆陆角捌分
泰皇岛顺先钢铁有限公司
911303227554512800N
昌黎县朱各庄镇前白石院村0335-7800113
```

图 10 文字识别结果

在图10中可以看出，在文字识别中，对数字简体字英文的识别效果非常好，很少有错别字出现，但是对字符的识别效果不佳，出错效率较高。将经过DenseNet + GRU + CTC模型识别得到的文字进行筛选等操作，将其中需要进行提取记录的发票代码等信息录入到Excel表格中，如图4-11所示。这说明系统的设计是基本可行的，但是功能还需要进一步改进。

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	发票代码	发票号码	开票日期	购买方名	购买方纳	销售方名	销售方纳	价税合计	发票名称	位置			
2	13002141	2022年02	霸州市富	河北省麻	911310811	秦皇岛顺	911303227	672261.2	D:/fapiaoshibie/zhizhifapiao/00.JPG				
3	13002141	2022年02	霸州市富	河北省麻	911310811	秦皇岛顺	911303227	525648.7	D:/fapiaoshibie/zhizhifapiao/IMG_7121.JPG				
4	13002141	2022年02	霸州市富	河北省麻	911310811	秦皇岛顺	911303227	773148.3	D:/fapiaoshibie/zhizhifapiao/IMG_7122.JPG				
5	13002141	2022年02	霸州市富	河北省麻	911310811	秦皇岛顺	911303227	1038354.	D:/fapiaoshibie/zhizhifapiao/IMG_7123.JPG				
6	13002141	2022年02	霸州市富	河北省麻	911310811	秦皇岛顺	911303227	1058479.	D:/fapiaoshibie/zhizhifapiao/IMG_7124.JPG				
7	13002141	2022年02	霸州市富	河北省麻	911310811	秦皇岛顺	911303227	1129756.	D:/fapiaoshibie/zhizhifapiao/IMG_7125.JPG				
8	13002141	2022年02	霸州市富	河北省麻	911310811	秦皇岛顺	911303227	1095395.	D:/fapiaoshibie/zhizhifapiao/IMG_7126.JPG				
9	13002141	2022年02	霸州市富	河北省麻	911310811	秦皇岛顺	911303227	1037330.	D:/fapiaoshibie/zhizhifapiao/IMG_7127.JPG				
10	13002141	2022年02	霸州市富	河北省麻	911310811	秦皇岛顺	911303227	1070906.	D:/fapiaoshibie/zhizhifapiao/IMG_7128.JPG				
11	13002141	2022年02	霸州市富	河北省麻	911310811	秦皇岛顺	911303227	915631.6	D:/fapiaoshibie/zhizhifapiao/IMG_7129.JPG				
12	13002141	2022年02	霸州市富	河北省麻	911310811	秦皇岛顺	911303227	1096496.	D:/fapiaoshibie/zhizhifapiao/IMG_7130.JPG				
13													

图 11 增值税发票重要信息提取效果

（2） 文字识别运行速度

在实际的文字检测中，不仅文字的检测准确率十分重要，进行识别所花费的时间也非常重要，因为财务人员要进行记录的发票数量过多，假若文字识别速度较慢，对财务人员工作的效率会有严重影响。本章利用深度学习的方法来对10张增值税发票进行文字定位与识别，具体花费时间如表1所示。从测试的结果来看，对于10张发票的文本框检测时间平均为0.753秒，对整张发票重要信息部分的识别过程平均需要1.321s，由此可见本文所设计实现的增值税发票定位与识别系统的运行速度还是较快的，对财务工作者要做的工作来说，提高了工作效率。

表1 运行速度测试

	定位时间(s)	识别时间(s)
第1张	0.76	1.365
第2张	0.74	1.345
第3张	0.78	1.384
第4张	0.75	1.325
第5张	0.756	1.362
第6张	0.758	1.345
第7张	0.742	1.452
第8张	0.73	1.236
第9张	0.75	1.652
第10张	0.74	1.256
平均值	0.753	1.321

4 结束语

本文构建了一个 DenseNet + GRU + CTC 网络模型。该模型利用 DenseNet + GRU + CTC 架构的深度学习算法，通过网络上开源的中文文档文字识别数据集来对 DenseNet + GRU + CTC 网络进行训练，得到的文本识别模型可用于实现对增值税发票文字候选框中的文本信息的自动识别。样机系统的性能测试结果表明，本文提出的模型和设计实现的增值税发票自动识别样机系统可以利用得到的最终的文本识别模型，对文本定位后的区域进行自动识别、结果分析和整理，说明了本文提出的模型和设计实现的样机系统是可行的、有效的。

参 考 文 献

[1] 李沛霖,吕巍,姚琳.基于 Tesseract 的会计票据图像识别系统设计研究[J].中国管理信息化, 2021,24(17):107-110.

[2] 张金涛. 票据号码信息识别算法的研究及系统实现[D].武汉：武汉科技大学, 2020.

[3] 黄志文.基于深度学习的发票自动识别系统的设计与实现[D].广州：广东工业大学，2018

[4] 李顿伟. 金融发票自动识别系统的研究与开发[D]. 东华大学,2018

[5] 蒋璎. 基于深度学习的发票识别系统[D].南京：南京邮电大学，2019

[6] 余子亮.基于深度学习的发票识别系统的研究与实现[D].南京：南京师范大学,2020

[7] 刘宁波，李刚，张华强. 基于 OCR 技术的发票自动识别校验系统设计[J]. 电脑知识与技术,2019,

[8] 黎贤钊.增值税发票内容自动识别系统研究[D].广州：广东工业大学，2020

[9] 陈科峻. 票据字符识别平台研究与实现[D].中国科学院大学，2021

[10] Huang G, Liu Z, Van Der Maaten L, et al. Densely Connected Convolutional Networks[C]// 2017 IEEE Conference on Computer Vision and Pattern

[11] 张永洪,孙幼政,高名岩等.百度 OCR 在房地一体户籍档案数字化中的自动著录研究[J].地矿测绘,2021,37(04):30-33

[12] 王锦涛,文晓涛,何易龙等.基于 CNN-GRU 神经网络的测井曲线预测方法[J].石油物探,2022,61(02):276-285.

[13] 万磊,余飞,鲁统伟等.基于 CEEMDAN-CNN-GRU 组合模型的短期负荷预测方法[J].河北科技大学学报,2022,43(02):154-161

[14] 张科昌,生绿伟,刘国辉等.集装箱码头箱号识别系统研发综述[J].港口装卸, 2022(02): 39-42.

[15] 曾悦,马明栋.基于 Tesseract__OCR 文字识别的研究[J].计算机技术与发展,2021, 31(11): 76-80.

[16] 颜家云,张慧源,李晨等.光学识别技术在机车检修记录单电子化中的应用[J].控制与信息技术,2021(06):77-83