Software Engineering and Applications 软件工程与应用, 2022, 11(4), 712-720
Published Online August 2022 in Hans. http://www.hanspub.org/journal/sea
https://doi.org/10.12677/sea.2022.114074

Hans汉斯

面向图像处理的深度学习算法在文本识别中的应用

李冬妮

南京林业大学信息科学技术学院，江苏 南京

收稿日期：2022年6月8日；录用日期：2022年7月27日；发布日期：2022年8月4日

摘　要

文字识别的难点不仅表现在文字的出现形式千差万别、汉字笔画多样以及字体种类繁多；也表现在实际生活中可能出现文字被遮盖或者有复杂的背景等各种各样的情况。为更有效的进行文字识别，基于用于图文识别的卷积递归神经网络(Convolutional Recurrent Neural Network, CRNN)模型提出了一种文字识别模型，并使用Python语言和Keras进行文字识别系统的实现。首先是数据增强算法的设计；其次是特征提取网络的设计；然后是对网络的决策层设计；最后采用一个卷积层去替换最初 CRNN 模型里参数量大以及不易收敛的长短期记忆网络(Long Short-Term Memorynetworks, LSTM)层。运用此方法一方面可以提高文字的识别准确率，另一方面可以降低网络参数以及提高网络的收敛速度。实验结果显示，使用该方法设计的文字识别系统不仅可以对各种字符进行识别，而且识别准确率较高。

关键词

图像处理，深度学习，文字识别，神经网络

Application of Deep Learning Algorithms for Image Processing in Text Recognition

Dongni Li

College of Information Science and Technology, Nanjing Forestry University, Nanjing Jiangsu

Received: Jun. 8th, 2022; accepted: Jul. 27th, 2022; published: Aug. 4th, 2022

Abstract

The difficulties of character recognition are not only in the various forms of characters, various strokes of Chinese characters and various types of fonts, but also in real life, there may be various

文章引用: 李冬妮. 面向图像处理的深度学习算法在文本识别中的应用[J]. 软件工程与应用, 2022, 11(4): 712-720.
DOI: 10.12677/sea.2022.114074

situations such as text being covered or complex background. In order to carry out a more effective character recognition, a character recognition model is proposed based on the Convolutional Recurrent Neural Network (CRNN) model for character recognition, and the character recognition system is realized by Python language and Keras. The first is the design of a data enhancement algorithm; Secondly, the design of a feature extraction network; Then the design of the decision-making layer of the network; Finally, a convolution layer is used to replace the Long Short-Term Memory networks (LSTM) layer with large parameters and difficult convergence in the initial CRNN model. On one hand, this method can improve the accuracy of character recognition. On the other hand, it can reduce the network parameters and improve the convergence speed of the network. The experimental results show that the character recognition system designed by this method can not only recognize various characters, but also has high recognition accuracy.

Keywords

Image Processing, Deep Learning, Text Recognition, Neural Networks

1. 引言

深度学习算法在近些年来被频繁的应用于机器学习的各个领域中。文字识别已经成为了图像应用领域中的一个研究热点[1]。目前深度学习算法已在文字识别中取得了较好的应用[2]。最早期的文字识别工作主要的研究内容是一些数字以及英文的识别，但由于识别技术的进步，汉字识别也逐步受到了研究人员的重视。也因为汉字特征较多，对汉字的深入研究将会产生无法估量的实际价值。

深度学习的重要网络结构之一是卷积神经网络(Convolutional Neural Networks, 简称 CNN)结构[3] [4]。近些年来，卷积神经网络已在图像分割、目标检测、语音识别和图像分类等方面有了一定程度的发展。许多研究人员对其网络结构进行了改进。以前的研究表明，卷积神经网络可以用于文本识别[5]。

针对文字识别中存在的问题，本文研究使用深度学习算法来解决文字识别问题。通常来说，想要设计一个稳定且强健的神经网络结构就要使得训练样本丰富多样且足够多。所以本文研究设计了一种数据增强算法，以及基于深度学习算法的特征提取方法，即在用于图文识别的卷积递归神经网络(Convolutional Recurrent Neural Network, CRNN)模型的基础上对其特征提取网络做进一步的改进，即用一层卷积层替换长短期记忆网络(Long Short-Term Memory Networks, LSTM)层。这种改进方式一方面能够提高模型的识别准确率，另一方面可降低使用的网络参数以及提升网络模型的收敛速率。基于以上算法的改进开发了一个文字识别系统，可以有效地识别出各种各样类型的字符。

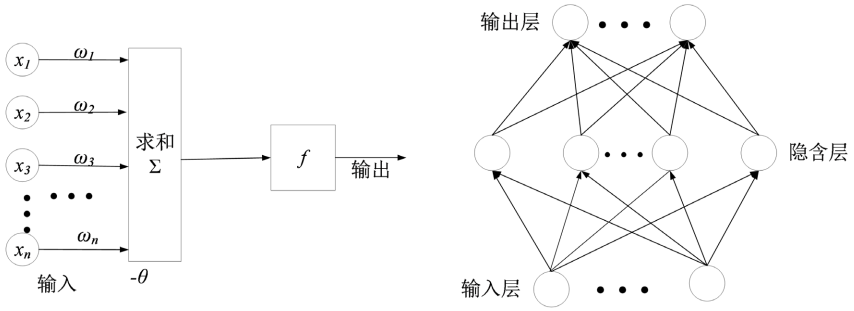
2. 材料与方法

2.1. 深度学习的基础理论

2.1.1. 传统前馈神经网络

把神经元当作一个功能逻辑器去开创人工神经元模型的研究思想，是历史上提出的第一个前馈神经网络神经元模型，即 M-P 模型[6]。该模型的基本结构图如下图 1(a)所示。感知机模型(Perceptron Model) [7]可以利用学习的方法来获得权值 ω_i 和阈值 θ [8]。如下图 1(b)所示的是一个三层神经网络结构图，也叫做

前馈神经网络[9]。



(a) (b)

Figure 1. Feed-forward neural network structure (a) single-neuron structure (M-P model); (b) three-layer feed-forward neural network structure
图 1. 前馈神经网络结构(a) 单神经元结构(M-P 模型)；(b) 三层前馈神经网络结构

记整个神经元的输出为 y，$f(\ )$是一种非线性的激活函数，则单个神经元的表达式为下式(1)：

$$y = f\left(\sum_{i=1}^{n} x_i \cdot \omega_i - \theta\right) \tag{1}$$

神经网络中的激活函数可以为模型具备非线性预测能力给予协助[10]，较常用的激活函数类型如下表1所示。

Table 1. Activation functions commonly used in neural networks
表 1. 神经网络常用的激活函数

激活函数名称	函数输入－输出关系
线性函数	$y=\delta(u-b)\quad(\delta>0)$
正线性函数	$y=\begin{cases}0 & u-b>0\\ \delta(u-b) & u-b>0\end{cases}\quad(\delta>0)$
对称 Sigmoid 函数	$y=\dfrac{\exp(\alpha)-\exp(\alpha)}{\exp(\alpha)+\exp(\alpha)}\quad(\alpha=u-b)$
Sigmoid 函数	$y=\dfrac{1}{1+\exp(-\lambda(u-b))}\quad\lambda>0$
单位阶跃函数	$y=\begin{cases}0 & u-b<0\\ 1 & u-b\geq 0\end{cases}$
对称阶跃函数	$y=\begin{cases}-1 & u-b<0\\ 1 & u-b\geq 0\end{cases}$

2.1.2. 深度学习的概念

深度学习是一种抽象数据的算法[11]，其是机器学习的一种，用于模拟数据之间的复杂关系。深度学习通过使用多层非线性信息处理，可以实现监督或无监督的模式分类、特征提取[12]、模式分析和特征转换，以研究文本和图像等其它数据。从本质上讲，深度学习是通过创建包含多个隐藏层并训练大量数据的学习模型，因此学习到更多的有用特征，从而增加了数据分类和预测的准确率。

2.2. 卷积神经网络概述

卷积神经网络[13]目前在语音、视频和图像数据领域等显示出很好的效果。如下图 2 所示的是卷积神经网络基本结构的示意图，其中 C 表示的是卷积层，S 表示的是池化层，F 表示的是全连接层。

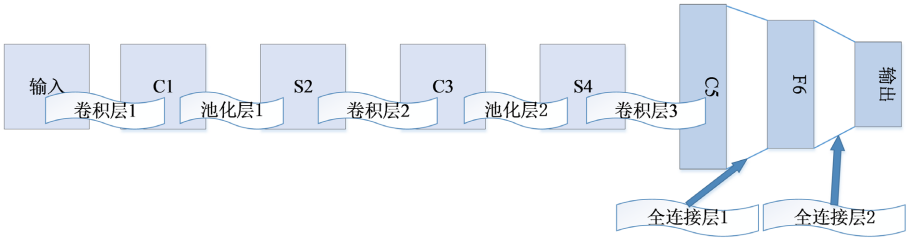


Figure 2. The basic structure of a convolutional neural network
图 2. 卷积神经网络基本结构

卷积操作是通过一定大小的卷积核作用与局部图像区域获得图像的局部信息的一种局部操作，卷积网络中的卷积核是通过网络训练获得的。图 3(a)显示的一个卷积操作示例图。

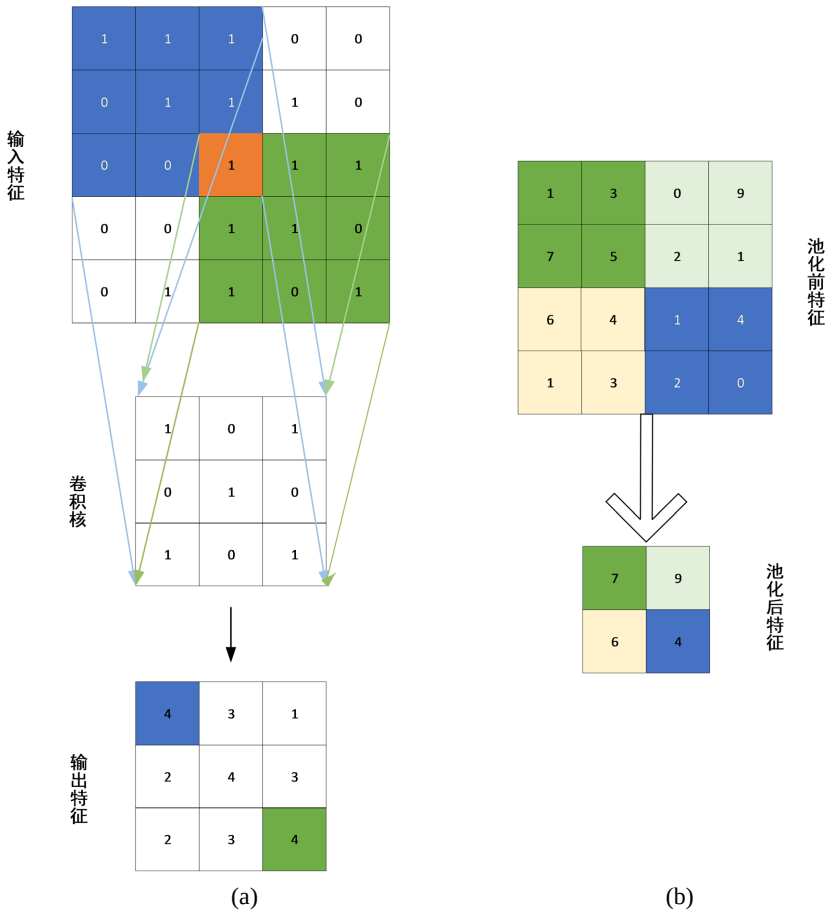


Figure 3. Operation diagram. (a) Convolutional operation diagram; (b) Maximum pooling operation diagram
图 3. 操作示意图。(a) 卷积操作示意图；(b) 最大池化操作示意图

池化(Pooling) [14]过程一般情况是在卷积过程之后，且与卷积过程类似。最大池化操作后特征高度和宽度减少一半，但通道数目不发生改变。如下图 3(b)所示的是卷积网络中的最大池化操作示例图。

2.3. CRNN 模型简介

CRNN 模型是一种可用于文字识别的模型，该模型被提出前，行业中对文字识别的方法都是通过对样本的字符切割，获得单个的字符后启动下一步的任务分类实现的。执行有序化的一种标签学习是 CRNN 模型中极为重要的一部分，即在网络训练期间，它只根据样本给出的一个序列标签，并且不用标签每个字符，从而减少对训练数据的要求。由于 CRNN 模型有其对上下文文字序列信息学习的独特性，所以可以用于文本识别。CRNN 模型流程图如下图 4 所示，该模型是由 CNN 层、LSTM 层和转录层构成的。

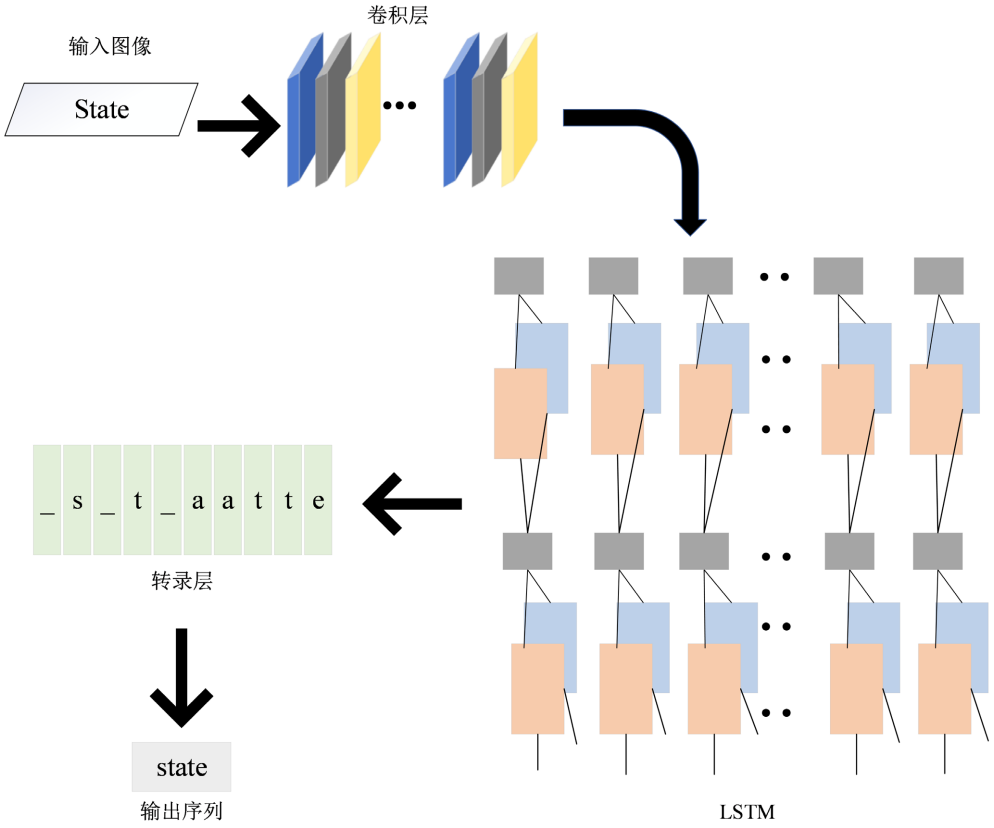


Figure 4. Schematic of the CRNN model
图 4. CRNN 模型示意图

2.4. Keras 框架简介

Keras 神经网络框架是优秀的深度学习框架中的一个，其封装性很好，对编程能力的要求不高。

在代码开发时，通过对各种深度学习框架的特点进行分析并且了解其编程语言后采用了 Keras 开发架构。该框架的练习流程能够即时监测，实现练习日志的可视化，在模拟练习流程中能够随意调节参数继续练习，以减少训练模式的实验周期。

2.5. 系统整体设计方案

由于文字识别测试样本可能存在噪声、文字模糊、仿射变换和图像扭曲等问题。因此，一个稳健的

深度学习文字识别系统的设计不仅要有一个好的网络结构以很好的学习上下文语义信息，还要能够对有限的数据进行增强，从而为深度神经网络提供质量更高的训练数据。图 5 为设计系统的整体框图。

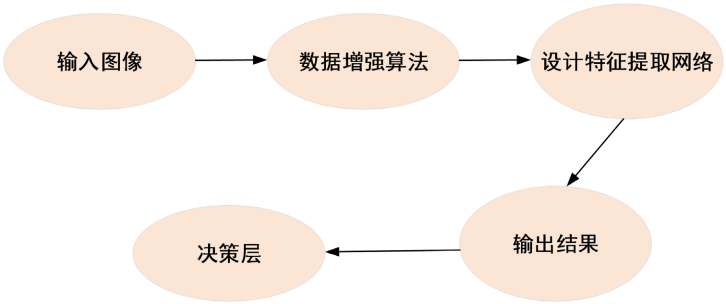


Figure 5. Overall frame diagram of the text recognition system
图 5. 文字识别系统整体框架图

2.6. 数据增强算法设计

首先透视变换所有初始的训练样本。经过深度学习模型的训练后，这些训练样本即使出现样本扭曲的情况依旧能够对其进行准确的识别。透视变换的含义是把需要的图像投至一个新的平面，变换过程如公式(2)所示：

$$\left[x',y',\omega'\right]=\left[\mu,\gamma,\omega\right]\begin{bmatrix}a_{11} & a_{12} & a_{13}\\ a_{21} & a_{22} & a_{23}\\ a_{31} & a_{32} & a_{33}\end{bmatrix} \tag{2}$$

其中原始图像的坐标为(μ,γ)，(x',y')表示的是变换之后的坐标，如公式(3)所示。

$$x=x'/\omega',\quad y=y'/\omega' \tag{3}$$

下一步，对公式(3)变形得到公式(4)和公式(5)：

$$x=x'/\omega'=\frac{a_{11}\mu+a_{21}\gamma+a_{31}}{a_{13}\mu+a_{23}\gamma+a_{33}} \tag{4}$$

$$y=y'/\omega'=\frac{a_{12}\mu+a_{22}\gamma+a_{32}}{a_{13}\mu+a_{23}\gamma+a_{33}} \tag{5}$$

因此，若要透视变换研究所需要的图像，则至少需要四对相应的点，并且可以从这四对点坐标中求出透视变换矩阵。研究对训练样本做出了透视变换操作，使训练样本集更丰富多样、网络训练更加稳健。

其次随机裁剪初始的训练样本。然后对初始的训练样本做出弯曲处理操作。

最后对初始的训练样本做出平滑处理操作，即对样本进行一定的高斯模糊处理。

2.7. 特征提取网络及决策层网络设计

如下图 6 所示的是特征提取网络设计的示意图。

如图 6 所示，在整个网络结束时，将 LSTM 层替换为卷积层用来提取特征，其次将特征送入 Softmax 激活函数进行激活以获得文字识别结果。有相关实验表明，这种方法一方面可以将网络的识别准确率提高约 4%，另一方面可以显著提升网络的训练速度，使得网络收敛难度降低，速度更快。Softmax 激活函数如公式(6)所示：

$$\sigma(z)_j=\frac{e^{e_j}}{\sum_{k=1}^{K}e^{z}k} \tag{6}$$

其中 $j=1,2,\cdots,K$ 。

文字识别系统的流程图如下图 7 所示。

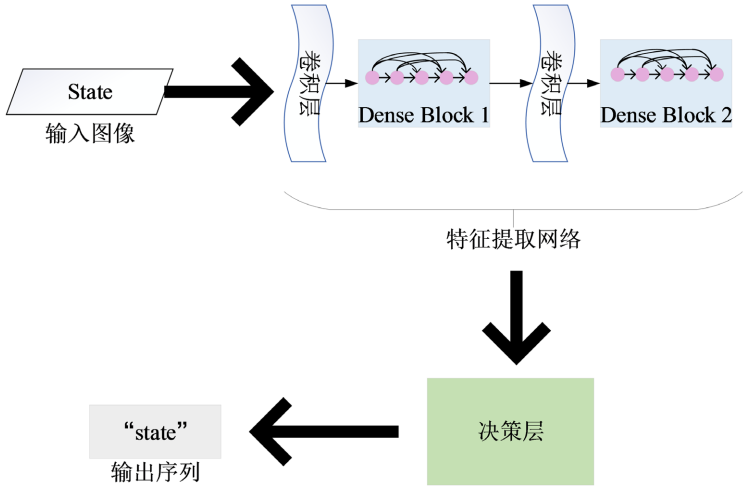


Figure 6. Schematic diagram of feature extraction network
图 6. 特征提取网络示意图

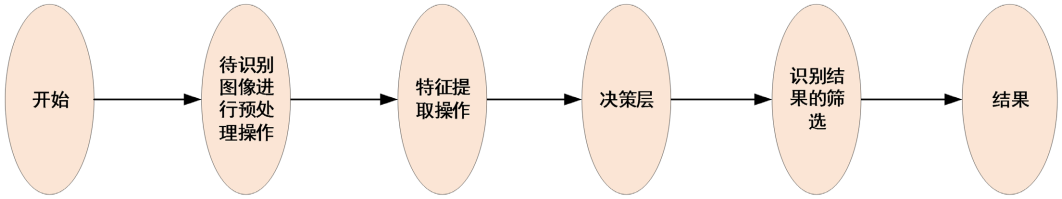


Figure 7. System overall identification flowchart
图 7. 系统整体识别流程图

3. 结果与分析

研究在各种不同的数据库上针对文字识别开展相关实验，通过对比分析不同模型在训练过程中有无使用研究中所提出的数据增强网络算法的实验结果，以此来检验提出的数据增强算法是否有效。

3.1. 汉字数据库上的实验结果

如下表 2 所示的是在汉字数据库上的实验结果。表 2 表示的分别是在训练过程没有使用提出的数据增强算法，以及训练过程中使用了提出的数据增强算法时不同模型在汉字字符数据库上的实验结果。

Table 2. Experimental results of the Chinese character database
表 2. 汉字字符数据库实验结果

序号	CRNN 模型	提出模型
无数据增强算法测试集的准确率(%)	90.13	90.97
无数据增强算法训练集的准确率(%)	91.23	92.14
有数据增强算法测试集的准确率(%)	93.24	95.37
有数据增强算法训练集的准确率(%)	94.53	96.96

从表 2 可以看出，在训练过程中没有加入数据增强算法时，CRNN 模型在测试集上的识别准确率为 90.13%，而提出的模型在测试集上的识别准确率是 90.97%。从实验结果中得到，CRNN 模型的性能在研究改进设计的特征提取网络和决策层的作用下有了明显的改善，识别准确率提升了 0.84%。在训练过程中使用了研究所提出的数据增强算法时，CRNN 模型的识别准确率在没有使用数据增强算法时的 90.13%的基础上提升了 3.11%的识别精度，达到了 93.24%；提出的模型在汉字字符数据库中的识别准确率达到了 95.37%，相比之前的 90.97%升高了 4.4%的识别精度。

从上述的实验结果可以得出结论：不论是 CRNN 模型，还是研究所提出改进深度学习模型，采用数据增强算法都对其识别准确率的提升有所帮助。同时也证明了提出的数据增强算法具有较高的有效性。

3.2. 英文数据库的实验结果

如下表 3 所示的是在英文数据库上的实验结果。

Table 3. Experimental results of English character database
表 3. 英文字符数据库实验结果

序号	CRNN 模型	提出模型
无数据增强算法测试集的准确率(%)	93.15	96.34
无数据增强算法训练集的准确率(%)	95.17	97.72
有数据增强算法测试集的准确率(%)	94.27	96.44
有数据增强算法训练集的准确率(%)	95.29	97.82

从表 3 可以看出，在训练过程中没有加入数据增强算法时，CRNN 模型在测试集上的识别准确率为 93.15%，而提出的模型在测试集上的识别准确率是 96.34%。CRNN 模型在训练过程中使用了研究所提出的数据增强算法时的识别准确率，在没有使用数据增强算法时 93.15%的识别准确率基础上提升了 1.12%，达到了 94.27%。提出的模型在英文字符数据库中的识别准确率达到了 96.44%，相比之前的 96.34%升高了 0.1%的识别精度。

同样从英文数据库的实验结果中可以得出：不论是原始的 CRNN 模型，还是研究所提出的改进深度学习模型，数据增强算法都提升了其识别准确率。即也验证了提出的数据增强算法具有有效性。

3.3. 数字数据库的实验结果

如下表 4 所示的是在数字数据库上的实验结果。

Table 4. Digital database experiment results
表 4. 数字数据库实验结果

序号	CRNN 模型	提出模型
无数据增强算法测试集的准确率(%)	95.34	97.23
无数据增强算法训练集的准确率(%)	96.45	97.31
有数据增强算法测试集的准确率(%)	96.21	98.82
有数据增强算法训练集的准确率(%)	97.65	98.90

从表 4 可得与上文同样的结论，即提出的模型在使用数据增强算法后的文字识别准确率较使用前有

所提升；提出模型整体的文字识别准确率相比较 CRNN 模型更高。综上所述，本文提出的算法以及模型在文字识别方面表现出较为优异的性能。

4. 结论

研究主要通过设计一系列改进的数据增强算法、特征提取网络以及决策层网络，提出了一种基于 CRNN 模型的文字识别系统，成功地实现了对汉字字符、英文字符和数字的高精度识别。不论是原始的 CRNN 模型，还是设计的深度学习模型，数据增强算法都对其有增加识别准确率的优点。基于深度学习算法在文本识别中应用的方法不仅可以提高文字的识别率，也提升了网络的收敛速率以及减少了网络参数的使用。但目前在设计的方法上仍有一些方面有待改进。例如，添加注意力机制在网络设计中，使网络在学习过程中可以对字符和背景区别学习，更多地去注意文字信息，而较少关注其背景信息，从而提高网络的识别精度。

参考文献

[1] 武子毅, 刘亮亮, 张再跃. 基于集成注意力层卷积神经网络的汉字识别[J]. 计算机技术与发展, 2018, 28(8): 100-103.

[2] Ha, I., Kim, H., Park, S., *et al*. (2018) Image Retrieval Using BIM and Features from Pretrained VGG Network for Indoor Localization. *Building & Environment*, **140**, 23-31. https://doi.org/10.1016/j.buildenv.2018.05.026

[3] 付发, 未建英, 张丽娜. 基于卷积网络的遥感图像建筑物提取技术研究[J]. 软件工程, 2018, 228(6): 8-11.

[4] Dolz, J., Gopinath, K., Jing, Y., *et al*. (2018) Hyper Dense-Net: A Hyper-Densely Connected CNN for Multi-Modal Image Segmentation. *IEEE Transactions on Medical Imaging*, **38**, 1116-1126. https://ieeexplore.ieee.org/document/8515234/

[5] Zhou, F., Li, X. and Li, Z. (2018) High-Frequency Details Enhancing DenseNet for Super-Resolution. *Neurocomputing*, **290**, 34-42. https://doi.org/10.1016/j.neucom.2018.02.027

[6] 夏昌新, 莫浩泓, 王成鑫, 等. 基于深度学习的图像文字识别技术研究与应用[J]. 软件导刊, 2020, 19(2): 127-131.

[7] Khened, M., Alex, V. and Krishnamurthi, G. (2019) Fully Convolutional Multi-Scale Residual DenseNets for Cardiac Segmentation and Automated Cardiac Diagnosis using Ensemble of Classifiers. *Medical Image Analysis*, **51**, 21-45. https://doi.org/10.1016/j.media.2018.10.004

[8] 李文英, 曹斌, 曹春水, 等. 一种基于深度学习的青铜器铭文识别方法[J]. 自动化学报, 2018, 44(11): 105-112.

[9] Ma, J., Shao, W., Ye, H., *et al*. (2018) Arbitrary-Oriented Scene Text Detection via Rotation Proposals. *IEEE Transactions on Multimedia*, **20**, 3111-3122.

[10] 白翔, 杨明锟, 石葆光, 等. 基于深度学习的场景文字检测与识别[J]. 中国科学: 信息科学, 2018, 48(5): 51-64.

[11] Liao, M., Zhu, Z., Shi, B., *et al*. (2018) Rotation-Sensitive Regression for Oriented Scene Text Detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, **11**, 5909-5918. https://doi.org/10.1109/CVPR.2018.00619

[12] Deng, D., Liu, H., Li, X., *et al*. (2018) Pixel Link: Detecting Scene Text via Instance Segmentation. *arXiv Preprint*, **18**, 13-15.

[13] 李新炜, 殷韶坤. 深度学习在文字识别领域的应用[J]. 电子技术与软件工程, 2018, 11(24): 40.

[14] Bai, F., Cheng, Z., Niu, Y., *et al*. (2018) Edit Probability for Scene Text Recognition. *arXiv Preprint*, **18**, 33-34. https://doi.org/10.1109/CVPR.2018.00163