第 58 卷 第 14 期/2021 年 7 月/激光与光电子学进展 研究论文

激光与光电子学进展 先进成像

文本图像倾斜角度检测的深度卷积神经网络方法

郭从洲[1*]，李可[1]，朱奕坤[1]，童晓冲[2]，王习文[1]

[1]信息工程大学基础部，河南 郑州 450001；

[2]信息工程大学地理空间信息学院，河南 郑州 450001

摘要 文本图像倾斜校正是文字识别前端的重要预处理环节。为了克服现有方法的倾斜角度检测范围只在−90°～90°的缺点，将文本图像倾斜角度检测问题转换为倾斜角度类检测问题，利用深度卷积神经网络的分类功能，选取适当的损失函数，设计了一阶段二分类和多阶段多分类的检测结构，实现了多种文本图像倾斜角度类的检测，进而得到了文本图像的倾斜角度范围。实验结果表明，倾斜角度类的准确率、召回率和精确率都在 0.93 以上。利用经典深度学习的方法对经过倾斜校正后的文本图像进行文字识别，识别精确率比校正前有大幅度的提升。

关键词 图像处理；文本图像；倾斜角度；卷积神经网络；倾斜校正；文字识别

中图分类号 TP751.1 **文献标志码** A **doi**：10.3788/LOP202158.1410007

Deep Convolution Neural Network Method for Skew Angle Detection in Text Images

Guo Congzhou[1*], Li Ke[1], Zhu Yikun[1], Tong Xiaochong[2], Wang Xiwen[1]

[1]*Department of Basic, Information Engineering University, Zhengzhou, Henan* 450001, *China*;

[2]*School of Surveying and Mapping, Information Engineering University, Zhengzhou, Henan* 450001, *China*

Abstract Text image skew correction is an important preprocessing step in the front-end of character recognition. To overcome the disadvantage in the limited range of tilt angle detection of the existing methods which is only −90–90°, this study transforms the text image skew angle detection problem into a skew angle class detection problem. Several types of skew angle classes of text images are detected using the classification function of deep convolution neural network by selecting the appropriate loss function and designing the detection structures of one-stage two classification and multi-stage multi-classification, and then getting the tilt angle range of the text image. The experimental results show that the tilt angle class's detection accuracy, recall, and precision rates are all above 0.93. The classical deep learning method is used to recognize the text image after skew correction. Moreover, the recognition accuracy is greatly improved compared to that before the correction.

Key words image processing; text image; skew angle; convolution neural network; skew correction; character recognition

OCIS codes 100.2000; 100.4996; 100.3008

1 引 言

光学字符识别(OCR)是指一种对文本图像进行处理，获取文字信息和版面信息的过程，也是光学理论实践应用的一个重要研究方向[1]。不论是传统意义上的 OCR 技术还是基于深度学习的 OCR 技术[2]，研究者努力的方向都是获得更高的文字识别精度和准确率。随着深度学习技术的发展，人们在

收稿日期：2020-09-22；**修回日期**：2020-10-21；**录用日期**：2020-11-14

基金项目：国家自然科学基金(41671409)

通信作者：*czguo0618@sina.cn

文字识别方法上取得了很大的进步，已经将文本检测和文本识别集成到了一个模型中[3]。从单一的水平直线文字行到现在的各类曲线文字行，涉及方法都能有很好的识别结果，即便文字行出现一定程度的倾斜(30°以内)，似乎也不会影响检测精度[4]。在文字识别过程中，要求单个文字方向要保持大致上的“正立”，如果出现“倒立”的文字，识别准确率就会变得很低，文字识别准确率除了与识别方法有关以外，与文本图像中文字的倾斜方向和角度也密切相关。

借助文本图像中的文字行的方向性，常规的文本倾斜角度计算方法主要有投影法、Hough 变换法、K-最近邻簇法、Fourier 变换法及其变种方法[5-7]。这些方法的基本思路都是先对文本图像进行游长平滑和细化处理[8]，获取文本行的直线特征，利用直线的斜率来确定文本图像的倾斜角度。能够从文本行特征检测出直线的原理是，直角平面中一条直线上的所有点映射到极坐标平面后所成的正弦曲线交会于一点，通过统计计数器的局部极大值来确定出直线的斜率。由于直线的倾斜角度范围为−90°～90°，而文本图像的倾斜角度范围为−180°～180°，所以当文本图像旋转 180°成为倒立图像时，利用常规的倾斜角度检测方法检测到的倾斜角度不是 180°，而是 0°。当文本图像倾斜角度大于 90°或小于−90°时，将直线的倾斜角理解为文本图像的倾斜角的检测方法都是不准确的。这种错误的原因在于文本图像倾斜程度不仅与倾斜角度有关，还与倾斜方向有关。由于常规方法不能检测出旋转方向，因此需要寻找一个不依赖直线特征的、更加普适的文本图像倾斜角度检测方法。

本文利用深度学习的方法，将文本图像倾斜角度检测问题理解为一个特殊倾斜角分类问题，通过设计合适的深度卷积神经网络(DCNN)结构对文本行图像进行特征提取，利用 Softmax 损失函数构造最优化问题，获取倾斜类别，实现文本图像倾斜角度类的检测，进一步估计倾斜角度。对 12 类倾斜文本图像进行测试，结果显示，倾斜角度类的检测准确率、召回率和精确率都在 0.93 以上。利用深度学习方法进行文字识别测试，文字识别精确率大幅度提升，从根本上解决了现有方法检测文本图像倾斜角度有限的问题。

2 文本图像倾斜角度类检测的 DCNN 结构

倾斜的文本图像按照倾斜方向和倾斜角度可以分为不同的类型，因此可以借助 DCNN 分类的功能实现倾斜方向和角度检测。输入文本图像到设计的分类网络中，输出一组表示文本图像倾斜方向和角度的标签值，整个网络是一个由大到小、由粗到细的框架。由于文本图像的倾斜方向和倾斜角度与图像的大小没有必然的关系，所以，从文本图像中心中截取一小部分包含文本的图像，该小部分的文本行的倾斜方向和角度完全可以代表原文本图像的倾斜方向和角度，因此卷积神经网络的输入可以取成相对固定的图像大小。为了减少预测参数量和计算量，规定竖直向上方向为 0°方向，逆时针旋转度数为正，顺指针方向为负值，整体旋转角度的范围为−180°～180°，这样只需要检测倾斜角度就可以确定倾斜方向。

2.1 DCNN 结构

首先对输入的样本图像进行常规的直方图增强处理和零-均值(zero-score)标准化预处理，将文本图像的像素值调整为近似正态分布，突出文字特征，提升网络训练中的激活能力；其次为了缓解过拟合问题和减少额外连接参数，在中间层加入了多个批量标准化(BN)层，并将全局平均池化层替换为分类器中常用的完全连接层[9]；为了保持图像从大到小过程中的细节特征，采用了最大池化(MaxPooling)进行池化；卷积层(Conv)后加入线性修正单元(ReLU 函数)，可以提高收敛速度，避免梯度消失。图1显示了文本图像倾斜方向和角度检测的DCNN

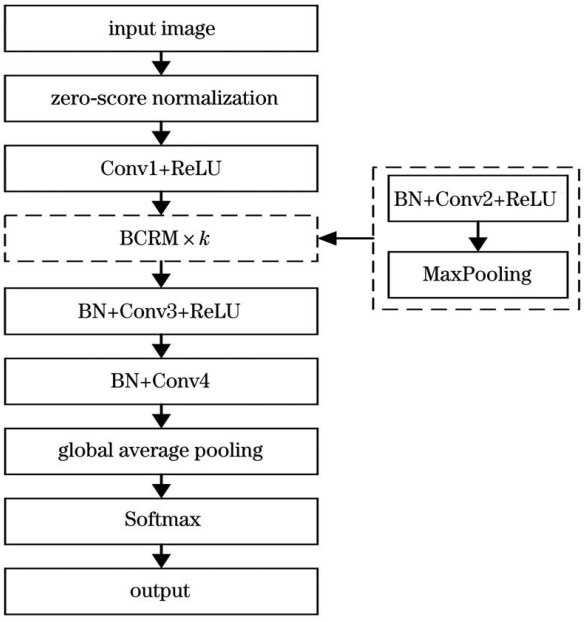

图 1 文本图像倾斜角度类检测的 DCNN 结构

Fig. 1 DCNN structure of text image tilt angle class detection

体系结构，其中 k 表示 BCRM 的数量，BCRM 表示 BN 层、Conv 层、ReLU 层与 MaxPooling 层首字母的缩写。

2.2 损失函数

损失函数是用来估量模型的预测值与真实标签值的接近程度的，它是一个非负实值函数，损失函数越小，模型的鲁棒性就越好。卷积神经网络都需要选择一个合适的损失函数，将卷积参数获取问题转换为一个优化问题。多类别分类问题的标签往往理解为概率值，所以它的损失函数主要就是针对这个概率值设定的。最常见的损失函数是交叉熵函数，它具有比方差损失函数权重更新快的优点。

假设 x 表示输入图像，λ 表示待训练的网络参数，则 DCNN 可以理解为一个关于预测文本图像 x 倾斜角度的概率函数：

$$f(x,\lambda)=P(x\in L_m|x), \tag{1}$$

式中：L_m 表示 m 个倾斜角度集合，P 表示概率。选取多分类交叉熵函数作为损失函数来迭代优化网络参数：

$$H(y,a)=-\frac{1}{N}\sum_{n=1}^{N}\sum_{m=1}^{M}\left[a_{m,n}\log(y_{m,n})+(1-a_{m,n})\log(1-y_{m,n})\right], \tag{2}$$

式中：N 表示每批次样本的数量；M 表示倾斜角度类的总数量；$a_{m,n}$ 表示第 n 个样本属于第 m 个倾斜角度类的概率；$y_{m,n}$ 表示学习训练得到的第 n 个样本 x_n 属于第 m 个标签倾斜角度类的概率，即 $y_{m,n}=f(x_n,\lambda)$。特殊地，如果只有两个质量分类等级，即 $M=2$，则

$$H(y,a)=-\frac{1}{N}\sum_{n=1}^{N}\left[a_n\log(y_n)+(1-a_n)\log(1-y_n)\right]。 \tag{3}$$

3 文本图像倾斜角度类检测

3.1 一阶段二分类检测结构

文本图像倾斜角度的检测与实际需求密切相关，比如在判断文本图像拍摄质量时只需要判断图像倾斜角度是否在某个范围之内，并不需要特别准确的倾斜角度，即使检测角度有些误差，也不影响对文本图像是否倾斜进行定性判断，这样的需求问题就是一个简单的二分类问题。二分类问题采用图 2 所示的一个固定深度的 DCNN 结构，通过对标注数据进行迭代训练，获取卷积核参数，确定一个固定分类等级数量的分类器。这种检测网络的优点是分类数量明确，训练目标明确。

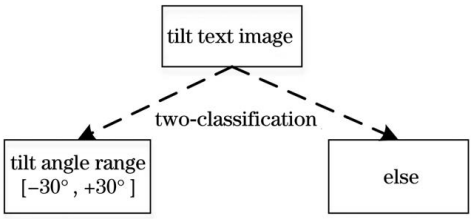

图 2 文本图像倾斜角度类二分类检测

Fig. 2 Two classification detection of text image skew angle class

3.2 多阶段多分类检测结构

在有些实际应用问题中，需要更精确的倾斜方向和角度判断，比如扫描文本图像、表格类文本图像时，利用传统 OCR 技术进行文字识别前需要进行倾斜校正，这样的需求问题是一个文本图像倾斜方向和角度定量检测问题，也是一个复杂的回归问题。倾斜角度检测得越精细，分类的等级就越多，进而要求特征提取网络也越深，训练难度也越大。为此，本文设计一种通过调整标签，从粗到细、逐层分类的卷积神经网络。该网络可以在固定深度的前提下，实现更精细的倾斜角度多分类检测。

由于任何一个大于 1 的整数都可以看作小于它的多个正整数的线性组合，所以任何一个正整数倾斜角度类数都可以由小于它的多个正整数倾斜角度类数线性组成。以倾斜角度 30°为基本单位，文本图像的倾斜角度可以分为 12 个类别，类别定义如表 1 所示。

表 1 倾斜角度类与对应角度范围值

Table 1 Tilt angle class and angle range value

Tilt classification	Tilt angle range	Tilt classification	Tilt angle range
L_1	[0, 30°)	L_7	[0, −30°)
L_2	[30°, 60°)	L_8	[−30°, −60°)
L_3	[60°, 90°)	L_9	[−60°, −90°)
L_4	[90°, 120°)	L_{10}	[−90°, −120°)
L_5	[120°, 150°)	L_{11}	[−120°, −150°)
L_6	[150°, 180°)	L_{12}	[−150°, −180°)

多阶段多分类检测流程主要包括标签替换和倾斜检测两个部分，流程如图 3 所示，具体操作过程如下。

1）标签替换 1：根据标签值将训练集数据分为两组。倾斜类别为 $L_1\sim L_6$ 的数据，其标签全部替

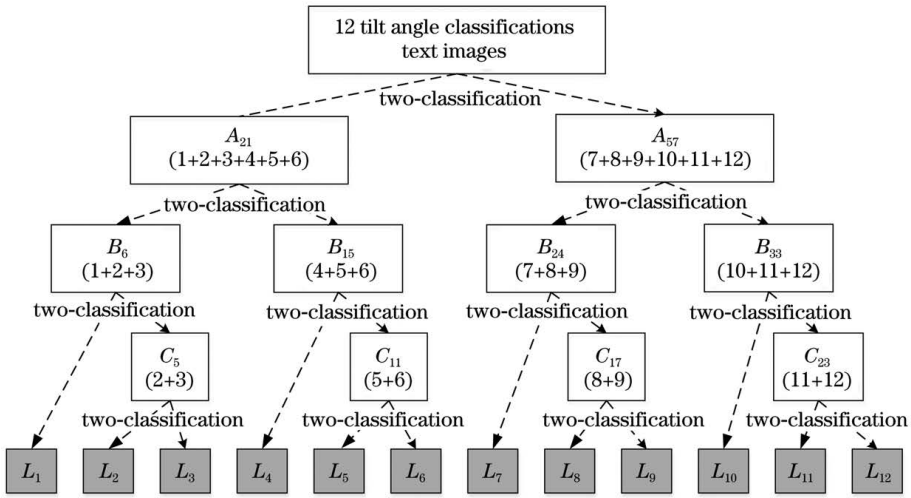

图 3　文本图像质量倾斜角度多阶段多分类检测示例 1

Fig. 3　The example 1 of text image quality skew angle multi-stage and multi classification detection

换为虚拟标签 21(1＋2＋3＋4＋5＋6＝21)，采用简单的标注方式，也可以选择其他能区分类别的数字；倾斜类别为 L_7～L_{12} 的数据，其标签全部替换为虚拟标签 57(7＋8＋9＋10＋11＋12＝57)。

2) 倾斜检测 1：利用虚拟标签值和二分类网络实现倾斜分类，再存入两个文件夹中，分别记为 A_{21} 和 A_{57}。

3) 标签替换 2：类似于步骤 1)，利用原始标签将 A_{21} 里面的数据分为两组，倾斜类别为 L_1～L_3 的数据的标签全部替换为虚拟标签 6(1＋2＋3＝6)，倾斜类别为 L_4～L_6 的数据的标签全部替换为虚拟标签 15(4＋5＋6＝15)；同理，利用原始标签将 A_{57} 里面的数据也分为两组，倾斜类别为 L_7～L_9 的数据的标签全部替换为虚拟标签 24(7＋8＋9＝24)，倾斜类别为 L_{10}～L_{12} 的数据的标签全部替换为虚拟标签 33(10＋11＋12＝33)。

4) 倾斜检测 2：利用虚拟标签值和二分类网络对 A_{21} 和 A_{57} 数据分别进行倾斜角度分类，存入 4 个文件夹，记为 B_6、B_{15}、B_{24} 和 B_{33}。

5) B_6、B_{15}、B_{24} 和 B_{33} 中数据分别有 3 个倾斜类别数据，类似于步骤 1)和 2)，利用二分类网络继续进行分类。

以上流程可以利用树状结构图进行展示，如图 3 所示，只有 2-叉树，也就是只利用二分类完成 12 类文本图像倾斜角度分类。当然，这样的分类方式还有更多的组合，比如 3-叉树(三分类)、5-叉树(五分类)等，实现文本图像倾斜角度分类，如图 4、5 所示。如果采用 3-叉树(三分类)结构，标签替换时需要生成 3 个虚拟标签对应的数据，其他类似。

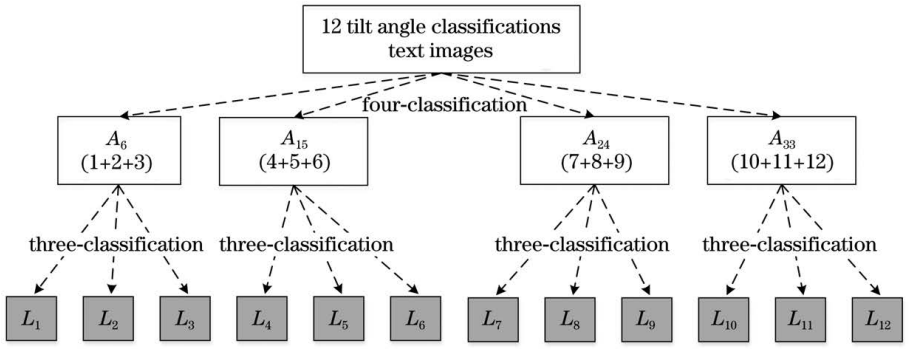

图 4　文本图像质量倾斜角度多阶段多分类检测示例 2

Fig. 4　The example 2 of text image quality skew angle multi-stage and multi classification detection

文本图像倾斜角度类多阶段多分类检测在训练过程中，每一段网络的输入都是原始训练数据，都是利用简单的分类网络实现等级分类。测试时，需要利用上一个卷积网络层的结果来选择合适的下一层网络，实现分类。多阶段多分类检测的优点是分类等级可以自由设定；网络结构简单；卷积参数相对固

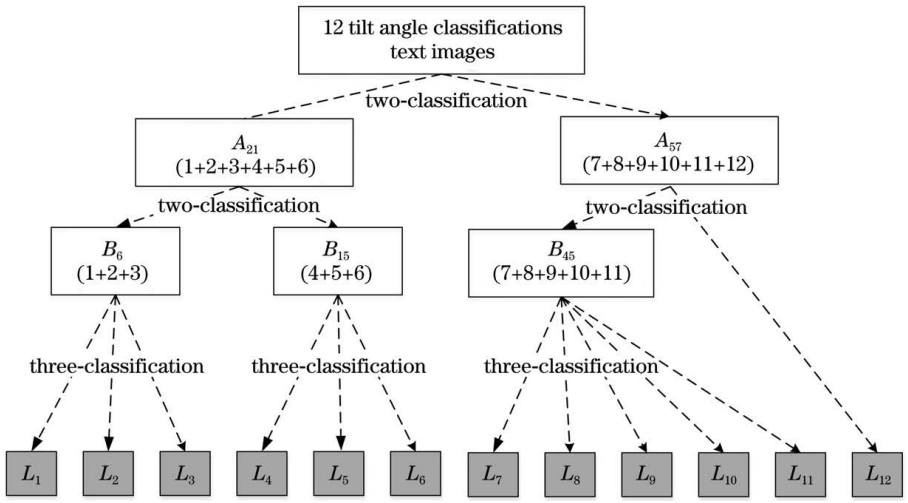

图 5　文本图像质量倾斜角度多阶段多分类检测示例 3

Fig. 5　The example 3 of text image quality skew angle multi-stage and multi classification detection

定；各段分类互不影响；可以单独训练后再组合训练。为了避免误差累计，通过内循环的方式逐层实现分类，损失函数定义为各段损失函数的和。

4　实验测试

4.1　数据集

为了验证上述方法的可行性，选取人民邮电大学出版社出版，Goodfellow等著，赵申剑等审校的《深度学习》(ISBN：9787115461476)作为数据源，将其转换为 1200 pixel×800 pixel、bmp 格式的文本图像；经过图像裁剪、增强、退化和倾斜旋转处理，共获取 12 个倾斜角度类 48000 幅文本图像，各倾斜角度类文本图像数量一致，如图 6 所示。为了测试算法的鲁棒性，倾斜旋转角度在同一类中为任意选取的。训练测试过程中选取 80%作为训练集，20%的数据作为测试集。

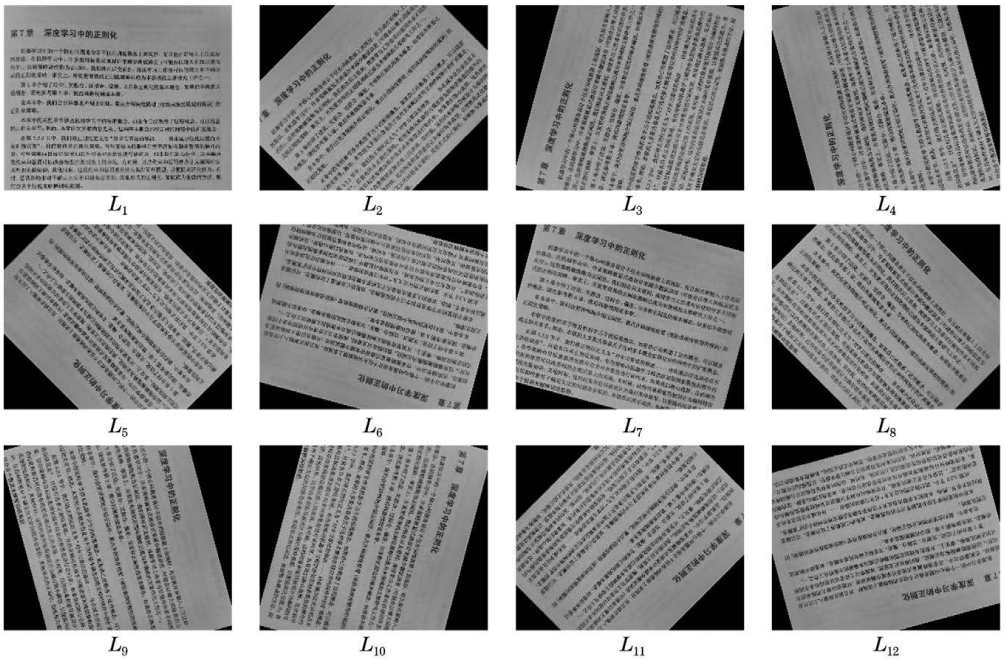

图 6　不同倾斜类别的文本图像仿真数据

Fig. 6　Text image simulation data with different skew categories

4.2　网络参数

选取图 1 的 DCNN 体系结构设置参数，以图 3 中的 2-叉数结构为例；输入文本图像的尺寸大小为 1200 pixel×800 pixel，以图像的几何中心为中心，截取 256 pixel×256 pixel 大小的图像，图像内确保有不水平的文字行；迭代次数设置为 1000，初始学

习率设置为 0.005，每迭代 50 次，学习率衰减为原来的 1/5。网络参数如表 2 所示。

表 2 网络参数

Table 2 Network parameters

Layer name	Filter size	Stride	Padding
Conv1	3×3×1×64	2	1
Conv2	3×3×64×64	2	1
Max Pooling	2×2	2	0
Conv3	3×3×64×64	2	1
Max Pooling	2×2	2	0
Conv4	3×3×64×64	2	1
Max Pooling	2×2	2	0
Conv5	3×3×64×64	2	1
Conv6	3×3×64×10	2	1

4.3 倾斜角度类结果评价

为了客观地评价深度学习在文本图像倾斜角度类检测中的性能，采用常见的数据分类评价指标，准确率(A)、召回率(R)、精确率(P)和 F1_Score，进行效果评价[10-11]。指标的定义分别为

$$A=\frac{N_{TP}+N_{TN}}{N_{TP}+N_{TN}+N_{FP}+N_{FN}}, \tag{4}$$

$$R=\frac{N_{TP}}{N_{TP}+N_{FN}}, \tag{5}$$

$$P=\frac{N_{TP}}{N_{TP}+N_{FP}}, \tag{6}$$

$$F=2\frac{P\times R}{P+R}, \tag{7}$$

式中：N_{TP} 表示将正样本预测为正样本的数量值；N_{TN} 表示将负样本预测为负样本的数量值；N_{FP} 表示负样本预测为正样本的数量值；N_{FN} 表示将正样本预测为负样本的数量值。为了更客观地反映检测精度，将检测角度的误差范围设置为−3°～3°，对没有文字行的图像进行手工剔除。

4.4 倾斜角度类检测效果分析

实验平台为 64 位 Windows 10 操作系统、Pytorch 深度学习框架；硬件配置为 Intel i7(十代)处理器、32 GB 内存和英伟达 2080TI 独立显卡。经过训练和测试，效果与精度如表 3 所示。

从表 3 可以发现：一阶段和多阶段网络结构都能够很好地实现文本图像倾斜角度类检测，准确率、召回率和精确率均达到了 0.95 以上；一阶段结构将文本图像的倾斜情况分为两个类型，准确率、召回率和精确率相对高些；多阶段结构将文本图像的倾斜情况分为了 12 个类型，三个评价指标值有所下降，计算效率相对偏低一些，这与卷积参数的结构和计算的复杂度有关，即文本图像倾斜角度类分布越细致，数据训练的时间越长，网络结构越复杂，网络参数也越多。倾斜类型的多少，对单幅文本图像倾斜角度类检测时间影响不大。

表 3 文本图像倾斜角度类(12 类)检测结果

Table 3 Detection results of text image tilt angle classification (12 classifications)

Performance index	One stage structure	Multistage structure		
	(Fig. 2)	(Fig. 3)	(Fig. 4)	(Fig. 5)
Accuracy	0.982	0.955	0.962	0.956
Recall	0.976	0.966	0.954	0.962
Precision	0.982	0.955	0.962	0.956
F1_Score	0.978	0.965	0.951	0.958
Train time /h	4.321	9.654	9.461	9.622
Test time /ms	7.139	8.112	8.121	8.110

为了测试多阶段分类的细致程度，将文本图像的倾斜角度平均分为 24 类，共计 96000 幅图像，倾斜角度范围为−15°～15°，误差范围规定为−2°～2°。训练和测试结果如表 4 所示。Fig. 2'采用类似于 Fig. 2 的 2-叉数结构；Fig. 3'和 Fig. 4'采用类似于 Fig. 3 和 Fig. 4 的结构，最后一层外采用 3-叉数结构，其他层采用 2-叉数结构。

表 4 文本图像倾斜角度类(24 类)检测结果

Table 4 Detection results of text image tilt angle classification (24 classifications)

Performance index	Multistage structure		
	(Fig. 2')	(Fig. 3')	(Fig. 4')
Accuracy	0.945	0.943	0.949
Recall	0.946	0.949	0.952
Precision	0.934	0.936	0.937
F1_Score	0.940	0.942	0.944
Train time/h	16.785	16.768	16.642
Test time/ms	13.609	13.434	13.536

对比表 3 和表 4 的数据，准确率、召回率、精确率和 F1_Score 变化不大，训练时间和测试时间明显增加，这与训练数据量、卷积网络的层数、卷积参数的数量有关。

4.5 文字识别效果对比分析

为了进一步说明文本图像倾斜角度类检测在深度学习领域的应用价值，选用 ASTER[12]、MORAN[13]和 CRNN[14]文字识别方法进行对比实验。这三种方法是基于深度学习的文字识别经典方法，代码也都是开源的。实验数据从 4.4 小节使用的 48000 幅文本图像中选取，选取时针对同一幅文本内容的图像，依照 12 个倾斜类别，各抽取 10 幅图像，共计抽取 120 幅图像，其中 L_1 表示正向数据，L_2～L_{12} 倾斜角度如图 6 所示，L_{13} 表示倒立文字。倾斜校正前指的是不对 120 幅待测试文本图像进行倾斜校正，利用三个方法直接进行文本识别；倾斜校正后指的是先利用倾斜角度类检测方法确定倾斜角度的大致范围，然后对 120 幅待测试文本图像进行倾斜校正，最后利用三个方法直接进行文本识别；由于利用深度学习方法进行文字识别本质上也是一个数据分类问题，为了能简单说明问题，对于文字识别的效果的判别，只选用精确率作为客观评价指标。

进行文字识别实验的目的是为了验证文本倾斜角度类检测的效用。文字识别测试过程中都使用三种方法训练好的参数，本文不再单独训练，倾斜校正前后的文字识别效果如表 5 所示。观察表 5 数据可发现：在倾斜校正前，当文本图像倾斜角度为 −30°～30°时（L_1 和 L_7 的角度倾斜类），三种文字识别方法的识别精确率都非常高；随着倾斜角度绝对值变大，识别精确率逐渐降低；当倾斜角度绝对值大于 60°时，文字识别精确率在 0.3 以下，识别方法几乎没有应用价值；当倾斜角度的绝对值大于 90°时，基本没有识别出正确文字。经过倾斜角度检测并倾斜校正后，图像基本保持正向，文字识别的精确率明显提升。

表 5 文本图像倾斜校正前后文字识别效果对比

Table 5 Comparison of text image recognition before and after skew correction

Tilt angle classification	Precision of ASTER		Precision of MORAN		Precision of CRNN	
	Before tilt correction	After tilt correction	Before tilt correction	After tilt correction	Before tilt correction	After tilt correction
L_1	0.966		0.974		0.977	
L_2	0.434	0.964	0.454	0.965	0.464	0.960
L_3	0.171	0.961	0.161	0.966	0.181	0.956
L_4	0.205	0.963	0.015	0.953	0.115	0.957
L_5	0	0.959	0	0.959	0	0.959
L_6	0	0.959	0	0.958	0	0.966
L_7	0.807	0.964	0.817	0.947	0.837	0.959
L_8	0.217	0.958	0.247	0.958	0.247	0.951
L_9	0.015	0.946	0.044	0.951	0.055	0.966
L_{10}	0	0.966	0	0.942	0	0.958
L_{11}	0	0.962	0	0.945	0	0.957
L_{12}	0	0.955	0	0.961	0	0.941
L_{13}	0	0.966	0	0.974	0	0.977

5 结 论

针对现有文本图像倾斜检测方法检测的倾斜角度有限的问题，利用深度卷积神经网络的数据分类的功能，构造了一阶段二分类和多阶段多分类的网络结构，通过估计倾斜角度类的办法确定更大（或更小）范围倾斜角度。该问题的解决，一方面拓展了深度学习的应用范围，另一方面为文本图像的倾斜角度检测提供了一个解决思路。在所设计的卷积神经网络结构中，涉及文本图像特征提取功能的，也可以利用 VGG 等其他常见的特征提取网络实现。本文主要将文本图像倾斜角度检测问题理解为一个数据分类问题，在后续的研究中，利用深度卷积神经网络的回归功能，结合更完善的特征提取网络实现更精

确的倾斜角度检测。

参 考 文 献

[1] Yang H J, Yan Z, Wu Z L, et al. Extraction method of interest text in image based on recurrent neural network[J]. Laser & Optoelectronics Progress, 2019, 56(24): 241501.
杨恒杰, 闫铮, 邬宗玲, 等. 基于循环神经网络的图像特定文本抽取方法[J]. 激光与光电子学进展, 2019, 56(24): 241501.

[2] Liao M H, Pang G, Huang J, et al. Mask TextSpotter v3: segmentation proposal network for robust scene text spotting[M]//Vedaldi A, Bischof H, Brox T, et al. Computer vision-ECCV 2020. Lecture notes in computer science. Cham: Springer, 2020, 12356: 706-722.

[3] Lyu P Y, Liao M H, Yao C, et al. Mask TextSpotter: an end-to-end trainable neural network for spotting text with arbitrary shapes[M]//Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11218: 71-88.

[4] Zhang Z, Zhang C Q, Shen W, et al. Multi-oriented text detection with fully convolutional networks[J]. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 4159-4167.

[5] Wu Y Q, Xie J. Document image skew detection based on the least distance fitting of the feature points[J]. Optical Technique, 2009, 35(1): 152-155.
吴一全, 谢静. 基于特征点最小距离拟合的文档图像倾斜检测[J]. 光学技术, 2009, 35(1): 152-155.

[6] Duan L, Song Y H, Zhang Y L. Layout analysis algorithm of questionnaire image[J]. Journal of Software, 2017, 28(2): 234-245.
段露, 宋永红, 张元林. 一种面向问卷图像的版面分析算法[J]. 软件学报, 2017, 28(2): 234-245.

[7] Zhou G W, Ping X J, Cheng J. Skew detection and correction method of document images based on improved Hough transform[J]. Journal of Computer Applications, 2007, 27(7): 1813-1816.
周冠玮, 平西建, 程娟. 基于改进 Hough 变换的文本图像倾斜校正方法[J]. 计算机应用, 2007, 27(7): 1813-1816.

[8] Jing L, Zhang X, Guo J X. Layout based photographs of document image skew adjusting[J]. Laser & Infrared, 2010, 40(10): 1116-1120.
荆雷, 张欣, 郭金鑫. 基于版面的拍照文档图像倾斜校正[J]. 激光与红外, 2010, 40(10): 1116-1120.

[9] Lin M, Chen Q, Yan S. Network in network[EB/OL]. (2013-12-16)[2020-09-20]. https://arxiv.org/abs/1312.4400.

[10] Zhang Y, Gong Z Y, Wei W W. Traffic sign detection based on improved Faster R-CNN model[J]. Laser & Optoelectronics Progress, 2020, 57(18): 181015.
张毅, 龚致远, 韦文闻. 基于改进 Faster R-CNN 模型的交通标志检测[J]. 激光与光电子学进展, 2020, 57(18): 181015.

[11] Luo C J, Jin L W, Sun Z H. MORAN: a multi-object rectified attention network for scene text recognition[J]. Pattern Recognition, 2019, 90: 109-118.

[12] Shi B G, Yang M K, Wang X G, et al. ASTER: an attentional scene text recognizer with flexible rectification[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 41(9): 2035-2048.

[13] Dong Y F, Zhang C T, Wang P, et al. Airplane detection of optical remote sensing images based on deep learning[J]. Laser & Optoelectronics Progress, 2020, 57(4): 041007.
董永峰, 仉长涛, 汪鹏, 等. 基于深度学习的光学遥感图像飞机检测算法[J]. 激光与光电子学进展, 2020, 57(4): 041007.

[14] Shi B G, 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.