

抗倾斜的中文文本图像文件识别技术^①



周一枫, 张华熊

(浙江理工大学 信息学院, 杭州 310018)

通讯作者: 张华熊, E-mail: zhxhz@zstu.edu.cn

摘要: 针对实际应用场景中如何在大批量图像文件中快速找到中文印刷体文本图像文件进行 OCR (Optical Character Recognition) 识别的问题, 本文在笔画宽度变换算法 (SWT) 的基础上, 设计了针对中文文本固有特点的启发式规则, 并将水平投影技术与离散傅里叶变换相结合, 提出了一种适合倾斜角度在 -90° 至 90° 之间的中文印刷体文本图像文件识别技术. 实验结果显示, 在 1606 张测试集图像文件的识别中, 本文算法针对文本图像文件整体识别 F 值 (F -Measure) 为 0.95, 平均识别耗时为 0.65 s.

关键词: 笔画宽度算法; 中文文本图像文件识别; 水平投影; 离散傅里叶变换

引用格式: 周一枫, 张华熊. 抗倾斜的中文文本图像文件识别技术. 计算机系统应用, 2019, 28(1): 32-37. <http://www.c-s-a.org.cn/1003-3254/6751.html>

Anti-Tilt Chinese Text Image File Recognition Technology

ZHOU Yi-Feng, ZHANG Hua-Xiong

(School of Information Science and Technology, Zhejiang Sci-Tech University, Hangzhou 310018, China)

Abstract: In view of how to quickly find Chinese printed text image files in bulk image file for Optical Character Recognition (OCR) recognition in practical application scenarios, this study designs heuristic rules for the inherent characteristics of Chinese text, based on the Stroke Width Transform algorithm (SWT), and combines horizontal projection technology with discrete Fourier transform, a Chinese printed text image file recognition technique suitable for tilt angles between -90° and 90° is proposed. The experimental results show that in 1606 test set image files, the overall recognition F -measure of the algorithm for text image files is 0.95, and the average recognition time is 0.65 s.

Key words: Stroke Width Transform (SWT); Chinese text image file recognition; horizontal projection; discrete Fourier transform

现阶段在纸质资料数字化的大背景下, 传统的纸质资料由于占空间、不便保存、查找繁琐、易丢失等缺点正被数字化的资料所取代, 然而传统的数字化大量采用的是扫描方式, 以图像的形式进行各种资料的存档保存. 在实际应用中如何快速从大量的数字化图像文件中找出中文印刷体文本图像文件进行 OCR (Optical Character Recognition) 识别是一个现实存在的实际需求.

针对图像文本的检测技术目前主要有以下 5 种: (1)

基于边缘的方法; (2) 基于连通域的方法; (3) 基于纹理的方法; (4) 基于深度学习的方法; (5) 基于上述混合的方法. 其中 Epshtein 等人^[1]提出的笔画宽度变换算法 (Stroke Width Transform, SWT) 由于利用了文字特有的笔画宽度特征, 因而在文本检测算法中被广泛使用, Yao 等人^[2]在 SWT 算法的基础上提出了一种检测自然场景下任意倾斜角度的文本, 但该算法处理时间较长, 不太适合在批量文本图像文件检测中使用. Chen 等人^[3]在 SWT 算法基

① 基金项目: 浙江省服装个性化定制协同创新中心项目 (浙教高科[2016]63 号)

Foundation item: Customized Clothing Collaborative Innovation Center Project of Zhejiang Province (Zhejiang Higher Education Science No. [2016]63)

收稿时间: 2018-07-21; 修改时间: 2018-08-21, 2018-08-29; 采用时间: 2018-08-30; csa 在线出版时间: 2018-12-07

础之上,提出了一种利用边缘增强的 MSER 技术^[4],该方法较好的提取极值区域,但在倾斜文本的检测上,效果不太理想并且处理耗时长.还有最近这几年由于深度学习技术^[5]的快速发展,基于深度学习的技术也开始应用在图像文本检测中,其中 Zhou 等人^[6]提出了利用全卷积网络 (FCN) 模型来检测文本,在 ICDAR2015^[7]测试数据集上,取得了 F 值为 0.8072 的效果. Tian 等人^[8]提出了一种应用卷积模型 VGG16 结合双向长短时记忆网络 (Bi-LSTM) 的方法,在 ICDAR2011, ICDAR2013, ICDAR 2015 测试数据集上,分别取得了 0.84, 0.88, 0.61 的 F 值效果,然而由于神经网络的训练,往往需要大量的训练数据,才能有比较好的训练效果,有时各种数据的获取难度也较大,故也不好利用在本文的检测算法中.

在中文文本检测方面, 廖裕青等人^[9]根据汉字的结构特征,通过分别改进 MSER 和 SWT 的方法来检测中文,在自建的中文环境自然场景图像数据库上,取得了 F 值为 0.794 的效果. Ren 等人^[10]通过在 CNN 卷积神经网络的基础之上,设计了一种适合中文文本结构组件的检测器 (TSCD) 层来增加检测中文的正确率. Jiang 等人^[11]通过采用 Harris 角点检测的方法来选择初始的文字候选区域,然后基于区域特征、文本笔画宽度、颜色信息过滤非文本区域,最后再结合行扫描的方法来检测中文,在自建的中文图像数据库上,取得了 F 值为 0.703 的效果.

传统的 SWT 算法因为是以英文文字为研究对象,而中文相较英文而言,在字符笔画,结构上都更加复杂,因而在检测中文的效果上, SWT 效果并不十分理想,同时我们研究发现,在有倾斜的文本检测效果中, SWT 算法由于一般以检测水平文本为目标,在倾斜文本的检测上,效果也不太理想.

通过非固定式移动相机或者扫描仪对文本图像文件进行数字化扫描的过程中,由于相机的抖动或者扫描文本放置角度的偏斜,数字化后的文本图像难免会存在着一定的倾斜角度,传统的 OCR 识别技术对待识别图像是否存在倾斜角度,有着较强的敏感性,倾斜角度的存在不仅影响着文本图像检测的准确性,也会降低后续文本图像文件进行 OCR 识别的性能.

Sakila 等人^[12]通过对比实验,研究了现有的 4 种常见的文档图像倾斜校正的方法: (1) 基于霍夫变换的方法; (2) 基于互相关的方法; (3) 基于 K 近邻的方法; (4) 基于傅里叶变换的方法. 其中基于傅里叶变换的方

法由于利用了傅里叶变换的旋转不变性,从而在对文档图像的倾斜校正方面,拥有较好的鲁棒性. 实验结果也表明,上述方法中基于傅里叶变换^[13,14]的图像倾斜校正方法在校正正确率以及处理时间上效果都是最优的.

由于本文算法只是用于是否是文本图像文件的识别,因此识别时不需要检测出所有的文字区域,我们只需判断检测图像是否带有中文即可,同时利用中文印刷体图像文本行与空白行交替变化的水平投影分布特征,运用水平投影技术快速提取第一行文本,在 SWT 算法中新增前景文本像素数所占的特定比率特性,来检测文本行连通域,即可判断图像是否带有中文,以便决定是否进行 OCR 识别.

1 算法设计

本文图例所示的文本图像均以政府部门、企事业单位所使用的红头文件^[15]为主,提出的算法基本流程如图 1 所示,主要步骤分为图像预处理、图像倾斜检测、水平投影、边缘检测、SWT 变换这 5 个步骤.

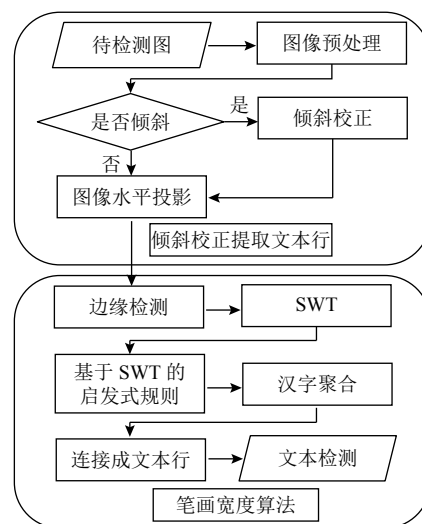


图 1 文本图像文件识别算法基本流程

a) 图像预处理. 首先对图像宽度大于 650 像素且高度大于 850 像素的图像, 缩放到 650×850 像素, 以加快图像检测速度.

b) 图像倾斜检测. 由于基于傅里叶变换的图像倾斜校正方法具有时间短、准确率高的优点. 本文对每一幅待检测图像, 进行二维离散傅里叶变换, 根据图像高低频谱交换频谱象限区域, 移频到图像中心、然后采用大津阈值法对图像进行二值化处理, 同时利用霍

夫线检测方法检测出倾斜直线与水平直线相交的倾斜角度 θ , 根据此倾斜角度 θ , 校正倾斜图像。

c) 水平投影. 将进行倾斜校正后的图像, 进行水平投影, 根据文本区域与空白区域的水平投影直方图投影分布区间来快速提取出识别区域。

d) 边缘检测. 利用 canny 算子, 提取文本行的边缘检测图。

e) 改进的 SWT 算法. 首先根据文本特有的固定笔画宽度特征, 计算每个像素的笔画宽度 (欧式距离), 得到包含文字笔画宽度信息的 SWT 图像, 然后查找文字候选区域, 对其中明显不符合中文文字的区域基于改进的 SWT 文本启发式规则进行过滤, 得到字符连通域. 对满足相关聚类条件的字符候选区域聚类成行, 形成文本行连通域, 最后对文本行连通域进行基于启发式规则的检测, 从而识别出待检测的图像是否是中文印刷体图像。

2 离散傅里叶变换

2.1 图像频谱成分

对于一副图像, 变换域的图像能量主要集中在低频部分, 其频谱分布示例图如图 2 所示。



图 2 傅里叶变换频谱分布

图 2 中, 图 2(a) 为红头文件倾斜图; 图 2(b) 为图 2(a) 带有倾斜角度的红头文件对应的傅里叶变换频谱图。

2.2 图像傅里叶变换平移效果图

图 3 为图像傅里叶变换平移效果图, 其中, 图 3(a) 为倾斜的红头文件; 图 3(b) 为傅里叶频谱平移图; 图 3(c) 为傅里叶频谱平移二值化图; 图 3(d) 为根据平移效果图计算出的倾斜角度 θ 。

通过对倾斜图像采用二维离散傅里叶变换, 来对倾斜图像在 $[-90, 90]$ 度之间进行校正, 从而获得更好的文本检测效果。

3 图像水平投影与切割

印刷体文本图像, 包括红头文件等一些文本图像, 文本行与空白行之间有着明显的明亮交替特征, 由此我们根据图像水平投影直方图的原理, 对中文印刷体图像进行水平投影处理, 提取我们的 ROI(感兴趣区域), 提高文本检测效率。

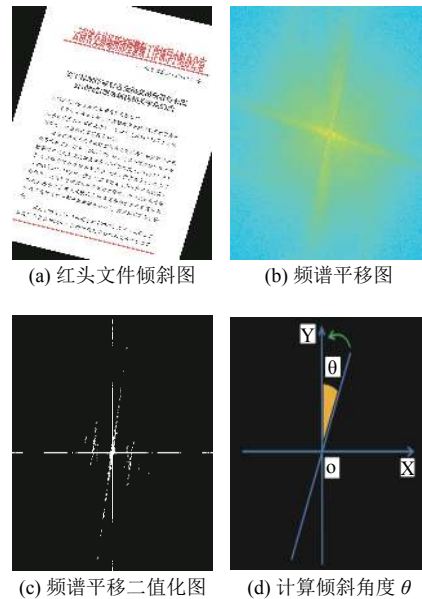


图 3 图像傅里叶变换平移

图 4 是图像水平投影直方图, 其中, 图 4(a) 为 Lena 图; 图 4(b) 为图 4(a) 的水平投影图; 图 4(c) 为红头文件; 图 4(d) 为图 4(c) 对应的水平投影直方图。



图 4 图像水平投影直方图

4 改进的笔画宽度算法

基于自然场景的文本检测,一直是国内外的研究热点,传统的自然场景文本检测算法,基本上都是以英文文本为检测对象,对于像中文这种笔画多样,结构复杂的象形类文字,检测效果往往不太理想,本文基于 SWT 算法的基础之上,利用了该算法将文本笔画宽度作为文字一个固有特征来提取文本文字,同时结合中文固有的笔画结构特征,设计了以下 4 点启发性规则:

(a) 增加文本行前景像素数占连通区域的占有比,来进一步过滤掉强干扰噪点,使检测中文文本的效果更加鲁棒,经测定当前景像素数占整体文本行连通域的比值 $SwtRule1 > 0.6$ 时,效果最优。公式如 (1) 所示:

$$SwtRule1 = \frac{S_{text}}{S_{rect}} > 0.6 \quad (1)$$

其中, S_{text} 为前景像素数面积, S_{rect} 为整体文本行连通域面积。

(b) 增加文本行连通域高宽比,公式如 (2) 所示:

$$SwtRule2 = ((s(h) \geq 10 \text{ 且 } s(w) \geq 30) \text{ 且 } \frac{s(w)}{s(h)} \geq 3) \quad (2)$$

其中, $s(h)$, $s(w)$ 分别为文本行连通域的高与宽。

(c) 针对原算法对查找文字笔画宽度的方向阈值 $\pi/6$,进行了适当的修正,我们在检测研究的过程中发现,原算法阈值 $\pi/6$,在针对英文文本这类文字形态变换简单的语言上,效果尚可,在针对中文这类形态复杂的文字上,该阈值设定过于严格,导致检索笔画宽度方向的过程中常常出现笔画越界现象,经我们针对中文印刷体图像的研究发现,当检测笔画宽度方向阈值为 $\pi/3$ 时,效果最优,公式如 (3) 所示:

$$SwtRule3 = |d_x + d_y| < \frac{\pi}{3} \quad (3)$$

其中, x 为笔画边缘像素点, d_x 为其梯度方向; y 为对应检索方向上边缘像素点, d_y 为其梯度方向。

(d) 针对中文的笔画特征,重新设定了字符连通域笔画的方差与连通域均值的比值,公式如 (4) 所示。

$$SwtRule4 = \frac{SwtVariance}{SwtMean} > 3.4 \quad (4)$$

其中, $SwtVariance$ 为字符连通域笔画的方差, $SwtMean$ 为字符连通域笔画的均值。

5 针对中文印刷体的检测算法改进

对于中文印刷体图像,我们采取通过水平投影直方图的方法来快速提取检测文本行,同时我们设定规

则排除中文印刷体中明显不是文本的投影区域(类似红头文件固有的红线),公式如下所示:

$$Result1 = proEindex - proSindex \geq 20 \quad (5)$$

其中, $proEindex$ 为水平投影图的文本行结束索引, $proSindex$ 为文本行开始索引。

为了提高检测速度,我们分别对输入图像的大小以及检测的文本区域进行了设定优化,经研究发现把输入图像宽高 ($W \times H$) 大于 650×850 像素的图像按公式 (6) 设定时,检测效果最优。

$$imageSize = \begin{cases} 650 \times 850, inputSize > 650 \\ inputSize, inputSize \leq 650 \times 850 \end{cases} \quad (6)$$

其中, 650 和 850 分别为图像的宽与高,单位为像素; $inputSize$ 为输入图像大小; $imageSize$ 为处理图像大小。

在检测文本区域上,我们采取优先选取首行文本为检测对象,同时为了保证对于倾斜角度过大的文本,图像校正之后出现文字消失,我们追加了正文头三行为检测感兴趣区域,保证了图像倾斜角度过大时,图像文字区域的文字消失给检测带来的干扰,同时也避免了对整个文本的检测,也加快了算法检测速度。

文本图像文件检测具体流程图如图 5 所示。



图 5 文本图像文件检测流程

图 5 中,图 5(a) 为待文字检测的水平图与倾斜图;图 5(b) 为经过傅里叶变换的倾斜校正图;图 5(c) 为经

过水平投影直方图提取到的检测 ROI 区域; 图 5(d) 为经过 Canny 边缘检测的效果图; 图 5(e) 为包含像素笔画宽度信息的 SWT 图; 图 5(f) 为基于启发式规则过滤的文本行; 图 5(g) 为文字检测结果输出图; 图 5(f) 与图 5(g) 是我们为了更好的说明我们的检测流程, 实体化的效果图, 在我们实际的检测算法中图 5(f) 与图 5(g) 的过程我们合为一步, 最终是否是文本图像文件的检测结果记入到日志当中。

6 实验与结果分析

6.1 中文文本图像样本库建立

本文采用的图像测试样本集为我们自己建立的红头文件样本库, 我们以每 2 度为一个倾斜角度对测试图像在[-90,90]度上进行旋转, 测试集含有水平, 以及上述倾斜角度的红头文件图像, 共计 1106 张。同时为了验证在自然场景下对图像文件的识别效果, 我们从 ImageNet2012^[16]视觉识别挑战赛 50 000 张测试集中挑选出 500 张图作为非文本类场景测试集。上述测试集图像库共计 1606 张图像。

6.2 图像倾斜校正对比实验

为了验证倾斜图像的倾斜角度与基于二维离散傅里叶变换方法检测得到的倾斜角度之间的正确性, 我们从 1106 张倾斜图像测试集中选取了 1092 张图, 按每 10 度区间作为一个测试分组, 共计 18 个分组对倾斜图像校正进行对比实验, 测试结果如表 1、表 2。

表 1 图像倾斜角度分组对比结果

测试分组	倾斜角度 (°)	准确率	平均耗时 (s)
1	0~10	0.86	0.081
2	10~20	0.92	0.076
3	20~30	0.94	0.075
4	30~40	0.95	0.076
5	40~50	0.96	0.076
6	50~60	0.97	0.076
7	60~70	0.96	0.074
8	70~80	0.89	0.073
9	80~90	0.80	0.071
10	-10~0	0.85	0.081
11	-20~-10	0.92	0.077
12	-30~-20	0.93	0.079
13	-40~-30	0.95	0.076
14	-50~-40	0.95	0.077
15	-60~-50	0.97	0.074
16	-70~-60	0.88	0.072
17	-80~-70	0.81	0.073
18	-90~-80	0.71	0.073

表 2 图像倾斜校正整体对比结果

倾斜角度 (°)	准确率	平均耗时 (s)
0~50	0.93	0.077
-50~0	0.92	0.078
-90~90	0.90	0.076

表 1 与表 2 的实验结果表明, 通过基于二维离散傅里叶变换对倾斜图像进行倾斜校正, 在各个测试分组, 常见的文本倾斜角度[-50,50]区间上, 以及在整体倾斜角度[-90,90]区间上, 都有较好的校正效果。

6.3 文本检测结果对比规则

实验采用文本检测领域常用的以准确率、召回率和 F 值为主要指标, 来评价文本检测方法的效果, 具体计算公式如下 (7), (8), (9) 所示。

准确率 (P) 的公式如下:

$$P = \frac{TP}{TP+FP} \quad (7)$$

召回率 (R) 的公式是:

$$R = \frac{TP}{TP+FN} \quad (8)$$

F 值计算公式如下:

$$F = \frac{2PR}{P+R} \quad (9)$$

其中, (1) TP : 为正确检测到的文本文件数; (2) FN : 为漏检的文本文件数; (3) FP : 为错误检出的文本文件数。

我们比较了 Epshtein^[1], Chen^[3], 以及 Zhou^[6]的算法。测试结果如表 3, 图 6, 表 4, 图 7 所示。

表 3 不同文本检测算法的准确率, 召回率和 F 值

方法	准确率	召回率	F 值
Epshtein 等 ^[1]	0.88	0.87	0.88
Chen 等 ^[3]	0.90	0.81	0.85
Zhou 等 ^[6]	0.88	0.62	0.73
本文方法	0.98	0.92	0.95

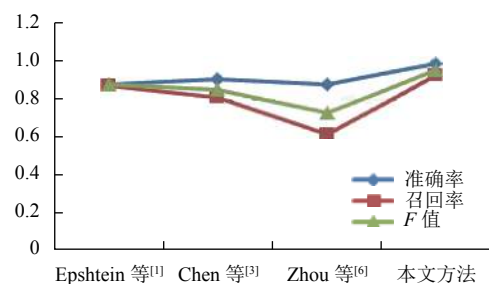
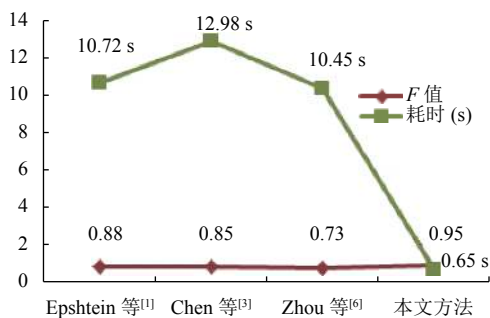


图 6 不同文本检测算法的准确率, 召回率和 F 值

表4 文本图像中文检测的 F 值与耗时结果

方法	F 值	耗时 (s)
Epshtein 等 ^[1]	0.88	10.72
Chen 等 ^[3]	0.85	12.98
Zhou 等 ^[6]	0.73	10.45
本文方法	0.95	0.65

图7 文本图像中文检测的 F 值与耗时结果

7 结束语

本文通过采用二维离散傅里叶变换的方法,对倾斜中文印刷体文本图像文件在 $[-90,90]$ 度上进行倾斜校正,通过水平投影直方图快速提取 ROI 区域,最后通过改进的 SWT 算法,来对文本图像进行中文的检测。通过上述的实验对比,可以发现本文针对中文印刷体图像文件的识别结果,无论在检测准确率、召回率和 F 值上,还是在识别文本图像的处理时间上,效果都是最优的,这也表明本文的算法具有较好的实用价值。

参考文献

- Epshtein B, Ofek E, Wexler Y. Detecting text in natural scenes with stroke width transform. Proceedings of 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Francisco, CA, USA. 2010. 2963–2970. [doi: 10.1109/CVPR.2010.5540041]
- Yao C, Bai X, Liu WY, et al. Detecting texts of arbitrary orientations in natural images. Proceedings of 2012 IEEE Conference on Computer Vision and Pattern Recognition. Providence, RI, USA. 2012. 1083–1090. [doi: 10.1109/CVPR.2012.6247787]
- Chen HZ, Tsai SS, Schroth G, et al. Robust text detection in natural images with edge-enhanced maximally stable extremal regions. Proceedings of the 2011 18th IEEE International Conference on Image Processing. Brussels, Belgium. 2011. 2609–2612. [doi: 10.1109/ICIP.2011.6116200]
- Matas J, Chum O, Urban M, et al. Robust wide-baseline stereo from maximally stable extremal regions. Image and Vision Computing, 2004, 22(10): 761–767. [doi: 10.1016/j.imavis.2004.02.006]
- Tang YB, Wu XQ. Scene text detection using superpixel-based stroke feature transform and deep learning based region classification. IEEE Transactions on Multimedia, 2018, 20(9): 2276–2288. [doi: 10.1109/TMM.2018.2802644]
- Zhou XY, Yao C, Wen H, et al. EAST: An efficient and accurate scene text detector. Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA. 2017. 2642–2651. [doi: 10.1109/CVPR.2017.283]
- Yao C, Wu JN, Zhou XY, et al. Incidental scene text understanding: Recent progresses on ICDAR 2015 robust reading competition challenge 4. arXiv: 1511.09207, 2015.
- Tian Z, Huang WL, He T, et al. Detecting text in natural image with connectionist text proposal network. Proceedings of the 14th European Conference on Computer vision. Amsterdam, The Netherlands. 2016. 56–72. [doi: 10.1007/978-3-319-46484-8_4]
- 缪裕青, 刘水清, 张万桢, 等. 自然场景图像中的中文文本检测算法. 计算机工程与设计, 2018, 39(3): 804–807. [doi: 10.16208/j.issn1000-7024.2018.03.036]
- Ren XH, Chen K, Yang XK, et al. A novel text structure feature extractor for Chinese scene text detection and recognition. Proceedings of the 2016 23rd International Conference on Pattern Recognition. Cancun, Mexico. 2016. 3380–3385. [doi: 10.1109/ICPR.2016.7900156]
- Jiang XY, Lian J, Xia ZQ, et al. Fast Chinese character detection from complex scenes. Proceedings of the 2016 6th International Conference on Image Processing Theory, Tools and Applications. Oulu, Finland. 2016. 1–4. [doi: 10.1109/IPTA.2016.7821001]
- Sakila A, Vijayarani DS. Skew detection and correction in the document image. International Journal of Innovative Research in Science, Engineering and Technology, 2017, 17457–17465. [doi: 10.15680/IJRSET.2016.0608274]
- Singh R, Kaur R. Improved skew detection and correction approach using discrete Fourier algorithm. International Journal of Soft Computing and Engineering, 2013, 3(4): 5–7.
- Kaur M, Jindal S. An integrated skew detection and correction using fast Fourier transform and DCT. International Journal of Scientific & Technology Research, 2013, 2(12): 164–169.
- 李秀勤. 我国红头文件现状分析. 合作经济与科技, 2010, (11): 60–61. [doi: 10.13665/j.cnki.hzjykyj.2010.11.004]
- Russakovsky O, Deng J, Su H, et al. Imagenet large scale visual recognition challenge. International Journal of Computer Vision, 2015, 115(3): 211–252. [doi: 10.1007/s11263-015-0816-y]