

# 回转库档案实时定位中基于鱼眼镜头图像的处理识别技术

成 杰, 叶文武, 徐寅林

(南京师范大学物理科学与技术学院, 江苏 南京 210023)

**[摘要]** 介绍了一种基于鱼眼镜头图像处理识别的回转库档案盘点技术, 采用鱼眼镜头宽视角拍摄回转库中排列成行的带有数字标签的多本档案; 利用图像矫正、分割技术处理得到标签上的每一个数字图; 由构造的卷积神经网络识别出图像内容; 以标签数字列在镜头中的位序实现档案定位。该技术实时性较好, 识别准确且成本低。

**[关键词]** 鱼眼镜头, 图像识别, 卷积神经网络, 档案识别

**[中图分类号]** TH711 **[文献标志码]** A **[文章编号]** 1672-1292(2019)02-0075-07

## Processing and Recognition Technology Based on Fisheye Lens Image in Real-Time Positioning of Rotary Library Files

Cheng Jie, Ye Wenwu, Xu Yinlin

(School of Physics and Technology, Nanjing Normal University, Nanjing 210023, China)

**Abstract:** This paper introduces a rotary library file inventory technology based on fisheye lens image processing and recognition. A wide angle of view of the fisheye lens is used to capture multiple files with digital labels arranged in rows. Image correction and segmentation technology are used to process each label. Image content is recognized by constructed Convolutional Neural Network. The file position is realized by the bit sequence of the label number listed in the lens. The technology has good real-time performance, accurate identification and low cost.

**Key words:** fisheye lens, image identification, convolutional neural network, file identification

现代档案电子化已成为一种趋势, 但纸质档案仍有其不可替代的价值。随着纸质档案数量的急剧增加以及档案使用频率变高, 人工管理档案的方式逐步被淘汰。现有的高密度纸质档案管理主要采用基于射频技术和条形码识别技术的回转库、密集架存储的管理方法。其盘库方式主要有以下 3 种<sup>[1-2]</sup>: (1) 利用机械手夹持条码枪, 对排列在档案柜上的档案盒条码进行移动扫描获取档案编号及位置信息; (2) 条码枪改为射频读卡器, 移动读取贴在档案盒上的电子标签; (3) 将射频读卡器排列成一行阵列, 每个读卡器正对一个贴有电子标签的档案盒, 每个射频读卡器既读取电子标签号, 实际也获取了档案的位置信息。前两种方法系统价格便宜, 但做不到实时盘库; 后一种方法需要大量射频读卡模块, 成本高, 且易互相干扰。

本文提出了一种基于图像识别技术的回转档案库定位方法, 如图 1 所示。利用鱼眼镜头焦距小、视场大的特点, 可以使得单幅图像包含的信息量很大。一个鱼眼镜头以凝视方式工作, 即能代替多个传统摄像头, 且不需要移动摄像头进行扫描, 可在获得更多有效信息的同时降低硬件资源。回转库中存取窗口位置一般只需采用 3~4 个鱼眼摄像头阵列, 就可采集整个窗口架上存放的所有档案的数字标签图像。

收稿日期: 2019-02-20.

基金项目: 江苏省研究生科研与实践创新计划项目(SJCX19\_0186).

通讯联系人: 徐寅林, 博士, 教授, 研究方向: 精密仪器设计. E-mail: xuyinlin@njnu.edu.cn

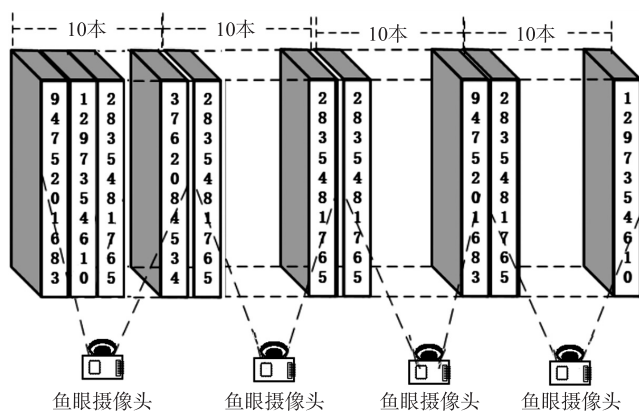


图 1 装置示意图

Fig. 1 Device schematic



图 2 鱼镜头成像模型

Fig. 2 Model of fisheye lens imaging

## 1 图像预处理

### 1.1 鱼眼畸变图像矫正

由于鱼镜头的焦距短,采集的图像有很严重的畸变,不利于内容识别,因此需对鱼眼采集到的图像进行矫正。矫正的原理是将世界坐标系点的坐标与鱼镜头的像素坐标系进行转化。系统成像模型包括世界坐标系( $x_w, y_w, z_w$ )、鱼镜头坐标系( $x_c, y_c, z_c$ )、摄像机坐标系、图像坐标系、成像平面坐标系,成像模型如图 2 所示。

鱼镜头的“桶形”畸变量较大,无法直接使用传统的线性标定,其坐标转换公式为:

$$\mathbf{X} = \mathbf{R}\mathbf{X}_w + \mathbf{T}, \quad (1)$$

式中, $\mathbf{X}_w$  代表点在世界参考系中的坐标; $\mathbf{R}$  为旋转矩阵; $\mathbf{T}$  为偏置矩阵; $\mathbf{X}$  为摄像机参考系中点坐标。

$$a = x/z, b = y/z, \quad (2)$$

$$r^2 = a^2 + b^2, \quad (3)$$

$$\theta = \tan(a/r), \quad (4)$$

式中, $a, b$  为图像坐标系中的坐标; $x, y, z$  为摄像机坐标系中点的坐标。

$$\theta_d = \theta \cdot (1 + k_1\theta^2 + k_2\theta^4 + k_3\theta^6 + k_4\theta^8), \quad (5)$$

式中, $\theta_d$  为鱼镜头的畸变系数。畸变后的点的坐标转换可表示为:

$$x' = (\theta_d/r) \cdot x, \quad (6)$$

$$y' = (\theta_d/r) \cdot y. \quad (7)$$

最后将坐标转换为摄像机像素坐标为:

$$u = \mathbf{F}_x(x' + \alpha y') + \mathbf{C}_x, \quad (8)$$

$$v = \mathbf{F}_y y' + \mathbf{C}_y, \quad (9)$$

式中, $\mathbf{F}_x, \mathbf{F}_y, \mathbf{C}_x, \mathbf{C}_y$  为鱼镜头的内部矩阵系数。

对于鱼镜头的矫正研究有多种方式<sup>[3-4]</sup>,本文直接采用 opencv 库中的 fisheye 类中的棋盘标定法对采集到的畸变图像进行矫正,矫正之后的效果如图 3 所示。

### 1.2 图像分割

图像分割的最终目的是分割出矫正图像中每位数字,并保证图像清晰,具备被识别的基本要求,因此对目标区域的定位与分割尤为重要<sup>[5]</sup>。对鱼镜头进行畸变矫正的拉伸操



图 3 矫正后图像

Fig. 3 The corrected image

作,会使得图像边缘出现较为严重的模糊. 本文采用二值化方法,期望最终可以分割出图像边缘清晰的数字图像.

由于图像在采集过程中会产生各种噪声,给档案的识别定位带来很大的麻烦. 因此,在二值化图像之前,需先采用高斯平滑及中值平滑对矫正之后的图像进行平滑处理,去除图像中的大部分噪声.

常用的二值化方法中固定阈值二值化是最常见的一种,将图像灰度值大于  $\text{thresh}$  (阈值) 的像素设为白色,小于或等于  $\text{thresh}$  (阈值) 设为黑色. 固定阈值二值化的公式为:

$$O(r,c)=\begin{cases} 255, & I(r,c)>\text{thresh}; \\ 0, & I(r,c)\leq\text{thresh}. \end{cases} \quad (10)$$

式中,  $I(r,c)$  代表图像  $I$  的第  $r$  行第  $c$  列的灰度值;  $O(r,c)$  代表该点阈值化后的灰度值. 对图像进行固定阈值二值化处理的结果如图 4 所示.

图 3 中镜头中间区域的标签白色背景与数字标签边界的灰度值差异明显,对矫正之后的图像进行固定阈值二值化处理后,数字标签的矩形轮廓清晰可见. 但对于图像边缘模糊区域,区域内的对比度变低,采用固定阈值二值化结果使得在图像边缘的标签区域的数字断断续续或已被淹没.

另外一种常用的二值化方法是自适应二值化,通常应用于光照不均匀或灰度分布不均的情况. 自适应二值化的策略是给每一个位置的灰度值设置一个对应的动态阈值:

$$O(r,c)=\begin{cases} 255, & I(r,c)>\text{thresh}(r,c); \\ 0, & I(r,c)\leq\text{thresh}(r,c). \end{cases} \quad (11)$$

这个动态阈值  $\text{thresh}(r,c)$  是图像中以  $(r,c)$  为中心点的  $7 \times 7$  的区域内的灰度平均值. 图像的自适应二值化的结果如图 5 所示,可以看出镜头两侧边缘的数字也能显现出来不被淹没.

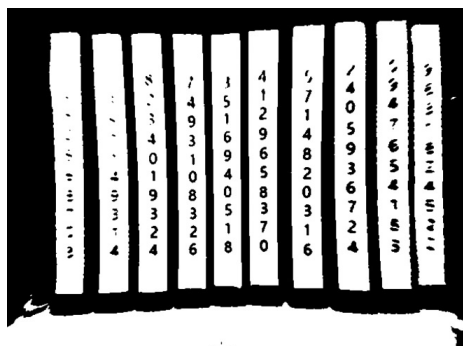


图 4 固定阈值二值化示意图

Fig. 4 Fixed threshold binarization diagram

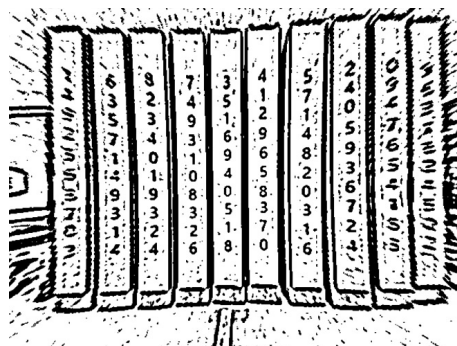


图 5 自适应阈值二值化示意图

Fig. 5 Adaptive threshold binarization diagram

但自适应阈值二值化方法会使图片中的噪声点加强显现出来,特别是标签边界线也叠加了许多干扰线条,使得边界线没有固定阈值法清晰易于分割.

根据上述分析结果,本文结合使用固定阈值和自适应阈值二值化两种方法对图像进行分割. 首先,利用固定阈值分割法只对标签区域进行分割;在获取每个标签区域边框后,再在原始图像上利用自适应二值化对标签区域内的数字进行处理,即可得到清晰的数字二值化图像. 两种二值化方法处理单个标签的效果对比如图 6 所示. 最终分割的效果如图 7 所示.

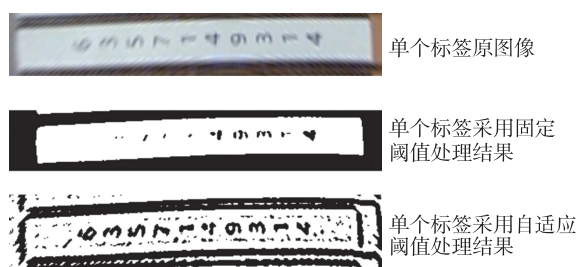


图 6 单个标签处理效果对比

Fig. 6 Comparison of individual label processing results

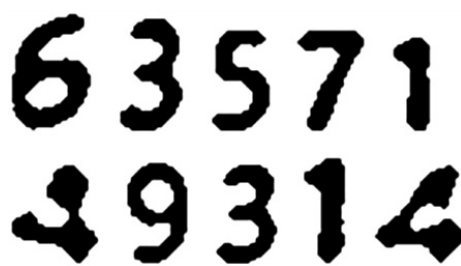


图 7 数字分割结果

Fig. 7 The result of digital segmentation



## 2 卷积神经网络字符识别方法研究

目前字符识别常用的方法有模板匹配、神经网络、支持向量机等. 模板匹配具有运行简单、识别速度快等优点,但任何光照、字体大小、清晰度及拍摄角度的不同都有可能改变识别的正确率. 为了提高正确率,实际使用的时候往往需要更多的模板,导致处理时间相应增加. 神经网络具有很好的自学性与存储知识的能力,支持向量机具有良好的训练分类能力,但这两种方法识别率都依赖于字符特征的选择与提取,而特征提取具有很大的不确定性,即人为选择的特征很难对所有字符适用,且很多较好特征的集合也未必能显著提升分类器性能.

为解决字符识别特征提取的难题,引入卷积神经网络(CNN)对字符进行识别. 卷积神经网络具有多个处理层,能对图像的特征进行逐层抽象,相比于传统的图像识别方法效果更好,对输入图像的旋转、扭曲、变形具有很好的鲁棒性,且不用预处理图像,简化了识别步骤. 卷积神经网络可以通过局部感受野、权值共享和降采样等操作来利用字符图像本身所包含的局部特征,将特征的提取融入到网络结构中,从而避免人工提取过程<sup>[6-8]</sup>. 卷积神经网络的局部连接和权值共享降低了参数量,使得训练的复杂度大大下降,减小了网络过拟合的可能性;同时,卷积层之间实现权值共享还赋予了卷积网络对平移的容忍性.

本文为了满足系统实时性的要求,参考 LeNet-5 模型,设计了一种简单但具有高正确率的卷积神经网络模型,模型采用两层卷积层和一层全连接层结构,其结构示意图如图 8 所示.

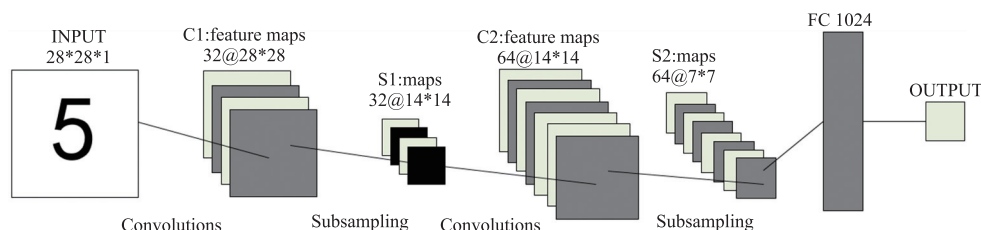


图 8 卷积神经网络结构示意图

Fig. 8 Schematic diagram of convolutional neural network structure

本文在 Tensorflow 深度学习框架下进行网络构造. 第一层卷积层采用了 32 个大小为  $5 \times 5$  的卷积核, 总共  $(5 \times 5 + 1) \times 32 = 832$  个参数,其中括号内的 1 代表了 bias. 使用  $2 \times 2$  的最大池化层,设定步长为 2 进一步缩小图像尺寸,经第一层卷积层之后输出图像大小为输入图像的一半. 第二层卷积层基本与第一层卷积层结构相同,唯一不同的是卷积核的数量变成 64 个,即在第二层卷积层会提取输入图像的 64 种特征,参数量为 1 664 个,这一层输出图像的尺寸只有最初输入图像的  $1/4$ . 全连接层含有 1 024 个隐藏节点,使用 ReLU 激活函数,输出结果为 0~9 这 10 种数字. 对于多分类问题,在输出层前可采用 Softmax 函数(或称为归一化指数函数)将其归一化,其形式如式(12)所示:

$$\sigma(z)_j = \frac{e^{z_j}}{\sum_{k=1}^K e^{z_k}}, \quad (12)$$

式中,  $j=1, \dots, K$ ;  $z_j$  为前层网络的输出结果. 卷积神经网络最终取其中概率最大值作为输出结果.

## 3 卷积神经网络数字识别实验

### 3.1 MNIST 数据集下的验证实验

首先采用目前广泛使用的 MNIST 手写数据进行验证本文构造的卷积神经网络. MNIST 数据集的训练集包含了 60 000 个样本,测试集包含了 10 000 个样本,训练后得到的最优模型在训练集和测试集的正确率如表 1 所示.

得到 MNIST 数据下拟合程度较好的 CNN 模型后,再将其应用到本文采集的鱼镜头数字图像的识别中. 利用单个鱼眼摄像头采集 10 本档案图像,并对采集的图像进行预处理得到大小为  $28 \times 28$  的二值化后的

表 1 MNIST 数据集 CNN 网络正确率

Table 1 The correct rate of CNN in MNIST data set

数据集	准确率
训练集	99.8%
测试集	99.2%

100 张数字图像进行预测,预测结果如表 2 所示.

表 2 模型预测结果

Table 2 The result of model prediction

图像数字(数量)	预测结果	图像数字(数量)	预测结果
0(8 张)	8 张识别正确	5(8 张)	8 张识别正确
1(10 张)	0 张识别正确	6(7 张)	7 张识别正确
2(9 张)	9 张识别正确	7(9 张)	0 张识别正确
3(9 张)	9 张识别正确	8(7 张)	5 张识别正确
4(12 张)	0 张识别正确	9(11 张)	1 张识别正确

从表 2 可知,训练好的卷积神经网络模型对实际标签数字图像的识别正确率仅为 47%,而 1、4、7、9 这几个数字基本不能识别.对 MNIST 数据集的图像输出可发现,MNIST 数据集中的数字与本文采用的数字字体有很大的区别,如图 9 所示. MNIST 数据集是国外手写数字,本次标签设计采用的是印刷黑体数字,因此识别的正确率很低.

### 3.2 构造黑体字数据集下的验证实验

利用现成的 MNIST 手写数据库训练的卷积神经网络识别打印的黑体数字正确率很低,因此本文对黑体数字构建数据集.首先打印 10 本档案标签,每个标签的数字全为“1”,采集鱼镜头下该 10 本档案的标签,经分割得到在鱼镜头下各个位置的 100 张数字“1”的图片,即可采集到数字“1”在图像中可能出现的所有位置.再以同样的方法得到“2”、“3”、……、“0”全部数字图片,共计 1 000 张不同数字的二值化图像,并对图像进行归一化,图像大小改为  $28 \times 28$ .对采集到的图像数据集通过增加高斯噪声及椒盐噪声的方式进行增强,其中训练集包括 1 626 张数字图像,测试集包括 557 张数字图像.

为验证本文设计的卷积神经网络结构在印刷体数字识别上的正确率,将增强后的数字标签数据集经模型分类训练,并与图像标签对比,得出网络模型的预测正确率.经 2 000 次迭代之后,模型对训练集图像的识别正确率达到了 100%.保存训练好的模型,利用测试集图像对模型进行验证识别,正确率达到 99.82%.

利用该模型对上文构建的实际环境下无污染的 100 张标签数字图像进行预测,预测的结果均正确.对档案进行定位可通过数字列在图像中的位置实现定位.

### 3.3 受污染标签的识别实验

在实际使用过程中,回转库档案柜中会有灰尘,多次借阅也可能会有油污之类落在档案盒的数字标签上.本文采用在黑体字数据集下拟合好的卷积神经网络进一步对受污染标签进行识别测试.

#### 3.3.1 数字标签添加噪声的识别

利用 opencv 库直接给已分割出来的数字图像添加不同程度的噪声信号,测试构建的卷积神经网络的识别效果.给数字添加 20%和 40%的高斯噪声或椒盐噪声,实验结果如图 10 所示.

从图 10 可知,本文构造的模型对于在原图像中添加一定比例的噪声干扰有较好的泛化能力.

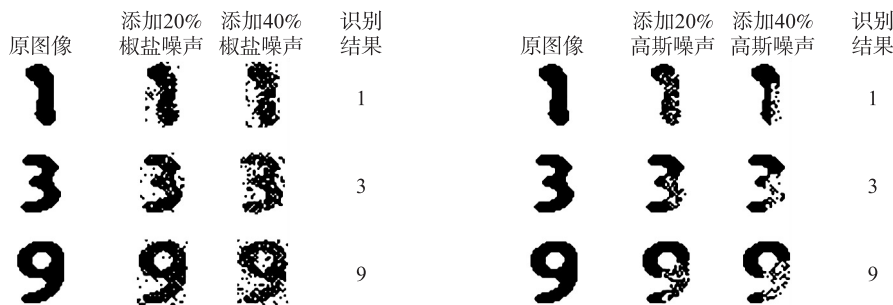


图 10 添加噪声的数字图像的识别结果

Fig. 10 Identification result of the digital image of adding noise

3.3.2 受灰尘油污污染的数字标签的识别

如图 11 所示,对一组档案盒上的部分标签增加灰尘或油污污染,检测本文提出的图像处理方法对灰尘污染的抗干扰性. 经图像预处理后的受污染标签图像的最终二值化图像如图 12 所示.



图 11 受灰尘油污污染的标签

Fig. 11 Label contaminated by dust and oil

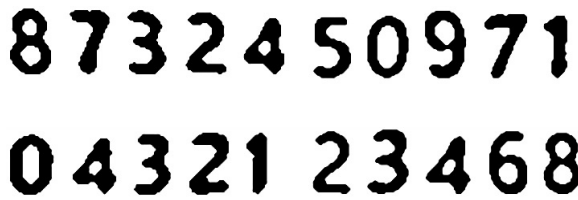


图 12 图像处理结果

Fig. 12 The result of image processing

从图 12 可知,受灰尘油污污染的标签经图像预处理后对最终分割出来的二值化图像没有影响,不会影响最终的识别效果.

3.3.3 受笔墨污染的数字标签识别

在实际档案存取过程中,可能存在各种笔墨对档案标签的污染. 采用黑色或蓝色记号笔分别对档案标签部分数字进行污染处理,经图像处理输出的二值化结果及卷积神经网络识别结果如表 3 所示.

从表 3 可以发现,当数字图像受笔墨污染情况不严重时(表中序号为 1、2、3、4、5、8 的样本),模型均可正确识别图像中的数字内容;对表中污染较为严重的样本 6 和 9,样本 6 可正确识别,而样本 9 的二值化图像更像数字“8”;样本 10 虽污染情况一般,但图像处理结果与样本 9 一样,图像特征更像数字“8”,故识别结果是数字“8”;样本 7 污染之后的图像特征与样本 3 数字“9”的特征更接近,因此识别结果为数字“9”.

表 3 受笔墨污染的识别实验结果

Table 3 Experimental results contaminated by pen and ink

序号	受污染数字图像	预处理结果	识别结果	序号	受污染数字图像	预处理结果	识别结果
1			4(正确)	6			6(正确)
2			2(正确)	7			9(错误)
3			9(正确)	8			2(正确)
4			5(正确)	9			8(错误)
5			7(正确)	10			8(错误)

通过识别实验可知,对于简单的笔墨污染,本文构建的网络有很强的泛化能力;但当污染较为严重或污染的笔墨对图像特征影响很大时,识别效果还有待改进. 实际上,本文在数字标签设计时,在标签中加入验证位,当标签最终识别结果不满足验证要求时,再人工干预对识别结果进行进一步调整.

4 结语

本文设计了图像识别技术的回转档案库定位方法,将鱼眼镜头引入到档案管理系统中,实现了较小空间下的档案图像采集;构建了黑体数字数据集,利用卷积神经网络拟合出具有很好泛化力的训练模型,可较好地满足档案定位的需求.

[参考文献] (References)

- [1] 王璐瑶. 基于 RFID 技术的实体档案管理研究[J]. 机电兵船档案,2018(2):39-41.  
WANG L Y. Research on entity archive management based on RFID technology[J]. Electromechanical ship file,2018(2):39-41.(in Chinese)
- [2] 秦苏榛. 基于物联网技术的档案馆实时自动盘库控制系统[D]. 南京:南京师范大学,2014.  
QIN S Z. Archives real-time automatic library control system based on internet of things technology[D]. Nanjing:Nanjing Normal University,2014.(in Chinese)
- [3] 郭盛威. 鱼眼图像视频校正方法研究[D]. 武汉:武汉工程大学,2017.  
GUO S W. Fisheye image video correction method research[D]. Wuhan:Wuhan Institute of Technology,2017.(in Chinese)
- [4] 冯为嘉,张宝峰,曹作良. 基于鱼镜头的全方位视觉参数标定与畸变矫正[J]. 天津大学学报,2011,44(5):417-424.  
FENG W J,ZHANG B F,CAO Z L. Omni-directional vision parameter calibration and rectification based on fisheye lens[J]. Journal of Tianjin university,2011,44(5):417-424.(in Chinese)
- [5] 朱安娜. 基于卷积神经网络的场景文本定位及多方向字符识别研究[D]. 武汉:华中科技大学,2016.  
ZHU A N. Research on convolution neural network-based scene text detection and multi-orientational character recognition [D]. Wuhan:Huazhong University of Science and Technology,2016.(in Chinese)
- [6] 余承波. 基于支持向量机(SVM)的车牌识别[D]. 淮南:安徽理工大学,2018.  
YU C B. License plate recognition based on support vector machine(SVM) [D]. Huainan:Anhui University of Science and Technology,2018.(in Chinese)
- [7] 靳涛,张永爱. TensorFlow 平台下基于深度学习的数字识别[J]. 信息技术与网络安全,2018,37(4):74-78.  
JIN T,ZHANG Y A. Digital recognition based on deep learning in Tensor Flow platform[J]. Information technology and network security,2018,37(4):74-78.(in Chinese)
- [8] 张立. 基于卷积神经网络的车牌字符识别方法研究[D]. 镇江:江苏大学,2015.  
ZHANG L. Research of license plate recognition based on convolutional neural network[D]. Zhenjiang:Jiangsu University,2015.(in Chinese).

[责任编辑:严海琳]

(上接第 49 页)

- [23] LIU N,GONG C,JIANG J,et al. Controlling odors from sewage sludge using ultrasound coupled with Fenton oxidation[J]. J Environ Manag,2016,181:124-128.
- [24] TOLEDOM M,GUTIÉRREZ C,SILES J A,et al. Full-scale composting of sewage sludge and market waste:stability monitoring and odor dispersion modeling[J]. Environmental research,2018(167):739-750.
- [25] 王强,邢秀强. 污水泵站通风设计探讨[J]. 工业安全与环保,2001,27(7):17-19.  
WANG Q,XING X Q. An approach of ventilation design for sewage pump house[J]. Industrial safety and environmental protection,2001,27(7):17-19.(in Chinese)
- [26] 张燕云,宋从波,刘茂,等. 基于 Fluent 的养猪场恶臭风险分析及应用[J]. 安全与环境学报,2015,15(1):293-296.  
ZHANG Y Y,SONG C B,LIU M,et al. Analysis of stench risk and application based on the software Fluent in the piggery [J]. Journal of safety and environment,2015,15(1):293-296.(in Chinese)
- [27] 于燕玲. 污水隧道纵向通风 CFD 仿真模拟与试验研究[D]. 天津:天津大学,2006.  
YU Y L. CFD analog simulation and experiment study on longitudinal ventilation of sewer tunnel[D]. Tianjin:Tianjin University,2006.(in Chinese)

[责任编辑:严海琳]