基于 OpenCV 与卷积神经网络的手写表格识别方法研究

卢承方，侯林杰，韦懿航，叶佳华，陈浩明
（广西民族师范学院数理与电子信息工程学院，广西 崇左 532200）

摘要：大多数单位每天会产生大量不同类型的纸质表格，采用人工方式录入纸质表格中保存信息，不仅过程繁琐，效率低下，而且信息录入的正确率也难以保证。为了提高工作效率并有效提取纸质表格中记录的手写数字信息，文章提出了基于OpenCV库的表格手写内容预处理分割方法，采用数学形态学中腐蚀膨胀运算等相关原理，搭建针对手写数字识别的卷积神经网络，对每个单元格框体内容中的手写数字进行识别，最终可以将纸质表格图片转化为电子表格。本方法的提出运用场景广泛，能极大地提高纸质表格数据提取的工作效率。

关键词：表格内容分割；轮廓查找；OpenCV；手写数字识别
中图分类号：TP391　　**文献标识码：**A　　**文章编号：**2096-9759(2022)02-0040-03

0 引言

表格因其能直观、有效的传达信息而被广泛的应用到各个场景中，人们可以快速从表格中提取有效信息，对现代化信息交互带来许多好处。虽然电子表格能高效存储和管理数据，但是目前电子表格还无法完全替代纸制表格，纸制表格在人们的工作生活中还在广泛使用。而将纸质表格转换成电子表格，才能更方便快捷的与其他类型数据交互，纸质表格内容提取是国内外研究的热点内容。在国外，Tuan AnhTran 等人提出一种使用随机旋转边界框的新形状来检测表格区域的新方法[1]，该方法不仅适用于检测倾斜的表格区域，也适用于设计文档布局分析系统。在国内，兰鹏生基于试卷图像的成绩颜色数据完成对表格内容提取与识别[2]，周壮提出可以运用于存在框线与字符交叉情况的表格框线检测算法[3]。本文研究了一种基于OpenCV的表格手写内容分割预处理方法，使用本方法可以将纸质表格图片进行单通道处理与二值化处理，以便计算机更好对图像进行滤波去噪，再通过轮廓查找得到表格框体，最后进行归一化处理并分割出各个单元格。本图像预处理方法的研究能更好的针对表格手写内容的单元格框线进行分割提取，最后搭建识别表格手写数字内容的卷积神经网络，完成表格内容批量识别的目标。

1 单通道处理

在使用手机等移动电子设备采集的彩色图像中的每一个像素的颜色通常由R、G、B三个通道分量组成，每一个通道分量的变化范围为0-255，即每一个像素点的变化范围为255*255*255，对得到的彩色图像进行单通道处理即是统一彩色图像的三通道分量，使每一个像素点变化范围缩小为0-255。进行单通道处理可以有效减少图像计算量，并且处理后的图像仍与彩色三通道图像一样能反映出整体与部分的亮度等级的分布和特征。本方法采用均值法，获取每一个像素点的三通道分量并求取平均值再赋回给原像素点的三个分量。

本方法使用 OpenCV 中的 cv2.cvtColor（image，cv2.COLOR_BGR2GRAY）函数将原始彩色三通道图像转化为如图 1 所示的灰度单通道图像。

题 号	一	二	三	四	五	六	七	八	九	十	总分
分 值	30	10	20	20	20						100
实得分	21	8	15	6	7						
统分人						核分人					

数据库操作原理

图 1 灰度单通道图像

2 图像二值化

图像二值化就是将图像以二值黑白像素的形式转化为矩阵形式，将有效的信息区域从背景区域剥离，呈现出明显的黑白效果。图像二值化最常用的方法是阈值法，通过分析选择出合适的阈值与处理过的单通道表格图像中的每一个像素点的灰度通道值相比较，超过所设定阈值的每一个像素点设置为 255，未超过阈值的像素点设置为 0。至此，通过二值化处理的图像最终呈现黑白效果。图像二值化可以降低数据维度，有效减少原始图像中存在的椒盐噪声点带来的干扰，是后续进行图像分割、特征提取等操作的重要前提。

因为需要手写内容的表格，其背景几乎不存在复杂情况，生成的表格图像没有严重的噪声影响，所以本方法采用 OpenCV 中的 cv2.adaptiveThreshold()自适应局部阈值化函数，根据表格图像中不同像素块的明暗分布，分别计算出各区域的合适的阈值，实现二值化操作。

3 中值滤波

因为设备的局限性，通过智能手机、平板电脑等移动电子设备的摄像头采集的表格图片会带有许多噪声点，过多的噪声点会降低表格框体提取的精度。因此要对图像进行滤波操作，以消除表格图像中随机存在的一些椒盐噪声点。本方法采用 OpenCV 中孔径线性尺寸参数设置为 7 的 medianBlur () 中值滤波函数消除表格图像的噪声点。中值滤波属于空间域滤波法中的一种，它对脉冲噪声滤出效果比较好，能够保证信号的边缘不丢失，对图像的细节信息有比较好的保护，所以中值滤波使用范围较广[4]。

收稿日期：2021-12-17
基金项目：广西民族师范学院大学生创新创业训练项目：基于边缘检测和卷积神经网络的试卷自动统分算法研究（S202110604071）。
作者简介：卢承方（2001-），男，广西南宁人，本科生，研究方向：计算机应用技术。

4 横竖线提取

数字图像处理中模式识别等方法需要含有手写内容的表格框体。本方法采用 OpenCV 中的 getStructuringElement()函数，为之后的数学形态学操作返回表格框体结构元素的大小和形状，再进行先腐蚀后膨胀的开操作运算识别出相应横竖线，然后采用 OpenCV 中的 addWeighted()函数对横线灰度二值图与竖线灰度二值图进行线性融合，之后对得到的单通道二值组合图像进行掩膜运算，最终形成由横线与竖线组成的表格框体图。

5 基于霍夫线变换进行图像角度校正

使用移动设备采集的表格图片可能会因为拍摄角度的影响，导致图像产生倾斜等类似的干扰情况。本算法采用霍夫线变换解决图片倾斜问题。霍夫线变换是一种变换域中提取直线的特征检测方法，该方法基于参数变换的思想，利用目标图像所处空间和霍夫空间的变换关系，从得到的直线段中提取出多条边缘线并进行拟合操作，然后根据这些直线对应的角度分布，估算出图像倾斜角的大小和方向[2]。本方法采用 OpenCV 中的 cv.HoughLines()函数在单通道二值图中查找出直线，完成对倾斜图像的角度校正。

6 单元格提取

6.1 轮廓查找

图像轮廓可以简单认为成将连续的点相连在一起的曲线且具有相同的亮度。轮廓在形状分析和物体的检测和识别中具有重要作用。轮廓查找是对单通道二值图像进行阈值化处理后，再寻找对象轮廓的方法。本方法采用 OpenCV 中的 findContours()轮廓查找函数。因为表格通常由大框体嵌套多个小框体的层级关系构成，为了更方便的分割出表格中的每一个单元格，本算法采用 RETR_TREE 的轮廓检索模式，该模式能够检索所有轮廓并为之建立一个树型结构，以外部轮廓含有内部轮廓，内部轮廓继续嵌套内部轮廓的方式检索。轮廓近似法采用 CV_CHAIN_APPROX_SIMPLE，该方法能够压缩轮廓中的水平、垂直、对角线段，仅保留轮廓的拐点位置信息。

6.2 轮廓绘制

循环遍历轮廓查找后获取的轮廓的拐点对象，并通过 OpenCV 中的 boundingRect()函数返回拐点对象的坐标信息依次保存到列表中，坐标信息包含 x,y, w,h 四个参数，本方法使用 OpenCV 的 rectangle() 函数绘制出查找到的轮廓。若拐点的位置信息用 $O_{x,y}$ 标识，则 $O_{x,y}$ 可以表示为：

$$O_{x,y}=(x,y) \quad (1)$$

(x,y)为矩阵的左上角坐标。

由此可推矩阵右下角坐标为：

$$O_{x+w,y+h}=(x+w,y+h) \quad (2)$$

w 为矩阵宽度，h 为矩阵高度。

轮廓绘制图像如图 2 所示。

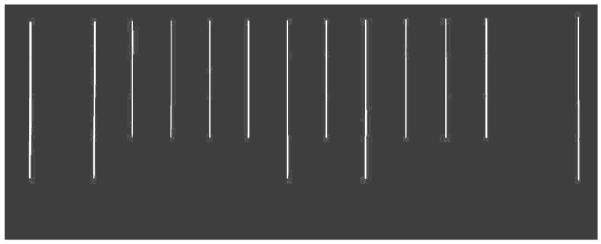

图 2 轮廓绘制图像

7 单元格切割及整合

确定了表格所在的位置后对单元格进行归一化切割操作。表格的每个单元格由矩形组成，将每个单元格区域用 L 标识。L 可表示为：

$$L=(\alpha_1,\beta_1,\alpha_2,\beta_2) \quad (3)$$

其中(α_1,β_1)为区域 L 的左上角坐标，(α_2,β_2)为区域 L 的右下角坐标。在轮廓查找中已经得到各个单元格轮廓的拐点坐标信息的集合，可以得知每一个单元格的坐标位置。从单元格区域的L_n取出(β_1,α_2)作为标识，即可分类每一个单元格的行区域，将其加入集合 $M=(L_1,L_2,L_3\cdots L_n)$。经由以上方法后，以单元格行分类，最终实现相同行单元格放置在同一集合中的整合目标。表格单元格切割整合结果如图 3 所示。

图 3 表格整合图

8 识别

8.1 数据集制作

识别实验中使用两种数据集进行合并训练，第一个数据集为 MNIST 数据集，由美国标准与技术研究所(National Institute of Standard and Technology，NIST)整理合成，该数据集中含有 70000 张样本图片，其中 60000 张作为训练数据，10000 张作为测试数据[5]，本文将 MNIST 读取并转换为 Tensorflow 深度学习框架所支持的 TFRecord 文件格式的数据集。第二个数据集为团队采集的广西民族师范学院特制纸质试卷成绩表格图片，包括有《Java 程序设计》、《数据库原理》等科目的试卷表格图片，并且采用平移、缩放与噪声干扰等数据增强方式完成对样本数量的扩充，再使用本文研究出的表格分割方法进行归一化分割操作，保存为 28×28 的 jpg 格式的图像，最后制作成 Tensorflow 深度学习框架所支持的 TFRecord 文件格式的数据集，该数据集中含有 5200 张图片，其中 3000 张作为训练数据，2200 张作为测试数据。

8.2 卷积神经网络训练

手写数字识别属于模式识别应用中的一种，进行模式识别的方法主要有 K 近邻算法(K-Nearest Neighbor[6]，K-NN)、支持向量机(Supported Vector Machine[7]，SVM)与人工神经网络等，K-NN 与 SVM 需要在数据集中进行特征向量的人工提取，

面对手写数字识别这样需要大量样本进行训练的情况时，人工特征提取效率较低且识别率不高。卷积神经网络(Convolutional Neural Networks,CNN)属于人工神经网络的一种分支，该网络能对输入的图像特征自动提取，相对 K 近邻算法和支持向量机等模式识别方法来说具有更高效率，本文采用经典卷积神经网络 AlexNet 对分割后的单元格内容进行识别训练。识别单元格内手写数字内容的网络结构如表 1 所示。

表 1 AlexNet 结构

层类	核尺寸	步长	通道数	激活函数
Input **Conv1**	**28×28** **3×3**	**4**	**1** **96**	 **ReLU**
MaxPool1	**3×3**	**2**		
Conv2	**5×5**	**1**	**256**	**ReLU**
MaxPool2	**3×3**	**2**		
Conv3	**3×3**	**1**	**384**	**ReLU**
Conv4	**3×3**	**1**	**384**	**ReLU**
Conv5	**3×3**	**1**	**256**	**ReLU**
MaxPool3	**3×3**	**2**		
FC1			**4096**	**ReLU**
FC2			**4096**	**ReLU**
FC3(Output)			**4096**	

本文搭建的 AlexNet 一共含有 1 个输入层，5 个卷积层，3 个池化层和 3 个全连接层。其中卷积层的作用是对图像进行部分特征提取，池化层的作用是降低图像维度并且保留有效信息，全连接层的作用是对图像进行全特征提取。输入层设置接收大小为 28×28 的通道数量为 1 的图像参数；五层卷积层的卷积核大小均是 3×3、步长为 1，滤波器个数依次为 96、256、384、384、256 的卷积层，对输入图像进行卷积操作；网络结构的第 1、2、5 层卷积层都设置核参数为 3×3 的最大池化层[8]；选取修正线性单元 ReLU 作为神经网络神经元的激活函数，减少卷积神经网络模型在进行模型训练时产生过拟合的可能性。加入 Dropout 技术，随机丢弃部分卷积神经网络的神经元，只保留它们的权重防止过拟合情况。加入 LRN 层，提高局部神经元中反馈较大的神经元的响应值，降低反馈较小的神经元的响应值。经卷积后得到特征矩阵的尺寸计算公式为：

$$N = \left[\frac{W}{S}\right] \tag{4}$$

其中 N 为卷积后得到的特征矩阵尺寸值，w 为输入图片尺寸的一维向量值，s 为步长数。

AlexNet 训练过程如下：(1)定义学习速率、迭代次数、输入图像的维度与图像标签维度等神经网络超参数。(2) 定义损失函数与学习步骤，并将定义的神经网络超参数初始化。(3) 将 mnist 数据集与自制的表格图片数据集加入网络中训练，计算精确度与损失值。(4)训练好模型后保存好模型响应的权重和参数。

8.3 调用模型接口

通过 Tensorflow 深度学习库中的 run()函数调用预训练模型接口，喂入如图 1 所示的测试样图，并将得到的预测值输入到 Excel 表格中，输出的 Excel 表格如图 4 所示。

	A	B	C	D	E	F	G	H	I	J	K	L
1	题号	一	二	三	四	五	六	七	八	九	十	总分
2	分值	30	10	20	20	20						
3	实得分	21	8	15	6	7						
4	统分人						核分人					

图 4 输出的 Excel 表格

9 结语

本文针对纸制表格框线分割的问题，提出基于 OpenCV 的图像预处理方法，侧重于表格图像轮廓查找的过程，采用经典的卷积神经网络 AlexNet 对模型训练，最后调用预训练模型的接口对单元格分割后的手写数字内容进行批量识别操作。从实验结果可以看到本方法可以对手写表格内容中的手写数字进行批量识别并导出成电子表格，识别率较高，能够满足一般应用的需要。下一步将针对纸质表格中手写数字粘连的情况进行研究，提高粘连的手写数字的识别率。

参考文献：

[1] Tuan AnhTran, Hong TaiTran, etc. A mixture model using Random Rotation Bounding Box to detect table region in document image[J]. Journal of Visual Communication and Image Representation,2016,39:196-208.

[2] 兰鹏生.基于数据颜色特殊性的加分表格识别系统研究[D].云南:昆明理工大学,2017.

[3] 周壮.表格识别系统中框线检测与去除的算法研究[D].辽宁:辽宁科技大学,2015.

[4] 张腾达,吕晓琪,任晓颖等.基于空间模糊核聚类的脑肿瘤图像分割方法[J]. 控制工程,2017, 24(10):12-13.

[5] 王满丽. 基于卷积神经网络的图像识别算法研究[D].北京建筑大学,2020.

[6] Tenindra Abeywickrama, Muhammad Aamir Cheema, David Taniar. k-Nearest Neighbors on Road Networks: A Journey in Experimentation and In-Memory Implementation[J/OL]. https://arxiv.org/abs/1601.01549,2016-8-10.

[7] Vapnik V.The nature of statistical learning theory[M]. Springer science & business media, 1999.

[8] 丁维龙,李涛,丁潇,余鋆,毛科技.基于改进 AlexNet 的手腕骨图像成熟等级识别[J].浙江工业大学学报,2021,49(06):614-622.

(上接第 36 页)提高服务质量。接下来，将从两个方面开展工作。首先，对现有的方案继续进行优化与完善，在保证定位系统功能的基础上，进一步提升系统的性能，提升鲁棒性和可复制性；其次，选择更多的项目或应用场景，对本文提出的方法进行应用落地或试验，让更多的场景从中受益。

参考文献：

[1] 师小波,赵丁选,孔志飞,倪涛,赵小龙.基于多传感器信息融合的车辆高精度定位技术[J].中国机械工程,2021(11).

[2] 张利,戈小中,梁子湘,何晓汉.一种 5G 高精度定位搭载自动驾驶车辆的解决方案[J]. 长江信息通信,2021(11).

[3] 魏立强,闫锐.一种基于高精度定位的 SDK 解决方案[J].长江信息通信,2021(07).

[4] 刘伟,黄瑞.北斗高精度定位技术在智能交通中的应用[J].集成电路应用,2021(07).

[5] 周小翠,席晨,赵小鹏.中移与千寻的高精度定位服务对比测试研究[J].长江信息通信,2021(05).

[6] 邬小鲁,蔡嘉,周秀.一种基于高精度定位的 SLA 服务[J].电子测试,2021(05).