

基于手机拍照的手写算式识别研究

刘兴豪¹ 陈芷妍¹ 何 滨¹ 童保鑫¹ 李文全¹

LIU Xinghao CHEN Zhiyan HE Bin TONG Baoxin LI Wenquan

摘 要

针对移动环境下手写算式识别准确率不高、效果不理想的问题,提出了基于手机拍照的手写算式识别的解决方案。首先,对上传的图片进行二值化和自适应平滑降噪,利用边缘检测和轮廓提取技术识别字符和倾斜矫正;然后,利用几何特征,采用公式定位、字符分割、字符排布树、结构解析等识别出手写算式,并通过数据集训练,提升手写算式的识别准确率;最后,设计并实现了基于三层体系架构的手写算式识别系统。实验结果表明,所设计的系统具有较高的识别准确率和较好的识别效果,具有一定的应用价值。

关键词

手写算式识别;图像预处理;手机拍照;卷积神经网络;几何特征

doi: 10.3969/j.issn.1672-9528.2024.06.015

0 引言

随着移动网络技术的快速发展,利用手机拍照将手写算式上传至网络,网络平台对手写算式解析识别转换成信息化算式,并将结果返回给用户,是当前智能化教育环境下非常普遍的学习行为^[1-3]。手写算式识别的正确与否直接影响着平台的用户体验和使用者的学习效率。现有识别技术可以对算式中的字符、符号、数字等以及组合算式进行有效识别。Aly 等人^[4]根据字符间的空间关系提出了算式中字符的识别方法。胡龙灿等人^[5]利用特征值对手写算式中的字符分类,在 Android 平台上提出基于二叉树的手写数学公式识别算法。Dai 等人^[6]通过预设数学模型和数据训练,实现了脱机手写算式识别系统。付鹏斌等人^[7]利用手写算式的几何特征和分治算法,提出了一种基于多重几何特征和卷积神经网络的脱机手写算式识别方法。王巍等人^[8]利用聚类分析提出了一种基于改进 Faster R-CNN 的基础算式检测与定位的方法。

虽然现有相关方法能对算式中的大部分字符进行有效识别,但依然存在两方面不足:一是对某些特殊数学符号识别不敏感,应用场景受限;二是对组合算式尤其是较复杂的手写算式识别准确率不高,效果不理想。因此,本文首先提出了针对手机拍照的手写算式的识别方案,然后对上传的图片预处理,并利用算式符号的几何特征提出手写算式识别模型,最后通过具体的手写算式识别系统加以测试,结果表明所提方法具有较高的识别准确率和较好的识别效果。

1 手写算式识别系统的架构设计

为了使系统具有较强的稳定性和可扩展性,算式识别系统采用了三层体系架构,具体架构如图 1 所示。其中,View 层是面向用户的图形界面,它不处理应用业务逻辑,只负责与用户的交互,接收用户图片的上传和操作请求;ViewModel 层负责处理业务逻辑,将接收到的 View 层的各种操作转换成系统能识别的指令,采用 AJAX 以 JSON 形式传送到 Model 层,并将 Model 层的处理结果返回给 View 层;Model 层负责算式的识别,通过调用算式识别算法,识别出图片中的算式,并将结果返回上层。各层之间采用 HTTP 协议进行信息传递。当用户通过 View 层的程序界面,利用手机拍照上传包含数学算式的图片后,ViewModel 层接收操作指令并将图片传送到 Model 层,Model 层对图片进行预处理、公式定位、字符分割、算式解析等一系列处理后,将识别出来的算式返回 ViewModel 层。最后,View 层将返回结果进行组织,显示在用户界面中。整个系统架构将业务表示层与逻辑处理层分开,使得系统架构具有较强的可重用性和可维护性。

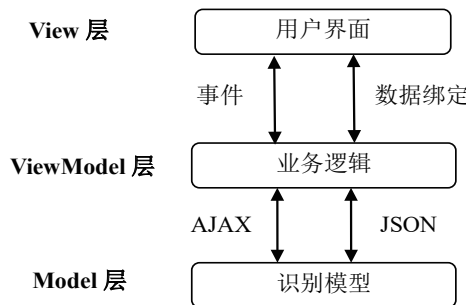


图 1 系统架构设计

1. 韶关学院信息工程学院 广东韶关 512005

[基金项目] 广东省普通高校特色创新类项目(2023KTSCX137);

韶关学院省级大学生创新创业训练计划项目(S202310576026)

2 识别模型的关键技术

根据上传手机拍照的相片，将手写算式进行识别并解析后返回给用户。其过程主要包括图像预处理、公式定位、字符分割、构建 CNN 网络结构、解析字符、构建树状图和算式结果计算。其中关键技术包括图像预处理、卷积神经网络的算式识别、模型学习和算式结果计算。

2.1 图像预处理

手机拍照的质量受光线不均、图像像素、目标模糊、拍摄角度等因素影响，在对手写算式识别之前须对图片进行预处理。预处理过程包括图片二值化处理、平滑降噪、倾斜矫正、目标识别等，最后得到可以算式识别的图像，具体如图2所示。在图像预处理中，首先采用自适应高斯阈值分割算法对图像进行二值化，将图像转化为黑白图像；然后使用 OpenCV 中图片去噪方法对二值化图像进行平滑降噪；最后利用项点提取算法纠正图像的倾斜角度，生成目标图像。

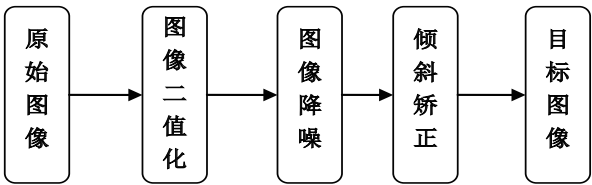


图 2 图像预处理过程

2.1.1 图像二值化

图像二值化是将灰度图片转换为非黑即白的图片，可以有效地提升手写算式识别的准确率。图像二值化处理时，需要设置一个图像阈值。手机拍照上传的图片受光线、像素等因素影响，固定的阈值不能有效处理不同环境下的图像，过大过小的阈值都会产生不同程度的噪点。因此，为了得到较理想的二值化处理效果，提出了自适应性阈值。通过分析图片的灰度值分布，依据白色背景在灰度值中所占像素较多的原理，在背景与前景中找出最适合的阈值，即根据背景与前景的灰度值分析，自适应调整阈值，在保证噪点较少的前提下，较好地分离出前景与背景。通过图像二值化处理，虽然图像噪点明显减少，但仍会存在少量噪点。因此，采用 OpenCV 中图片去噪方法，将图像中的噪点进行平滑处理，最大程度上去除图像噪点，为手写算式识别提供数据基础。

2.1.2 图像倾斜矫正

手机拍摄角度不平、用户书定算式不正都可能出现图像倾斜，它是手机拍照中不可避免的问题。为了矫正图像倾斜，首先，采用基于一阶导数 Sobel 算子的边缘检测技术^[9-10]。选用 3*3 的模板核与每个像素点做卷积和运算，并通过设定阈值提取边缘，依据二值化处理的数据将边缘与背景分离。其

次，为了提升手写算式的识别效果，在边缘检测后，采用了轮廓提取技术找出手写算式的轮廓。它以某个选定的种子点为始点，使用搜索方法找出图像中的闭合路径，以此找到边缘内部手写算式的闭合轮廓。最后，找出图像的水平线，利用倾斜角度对图像进行倾斜矫正。

2.2 手写算式的识别模型

经过图像预处理后，可以对图像中的手写算式进行识别。识别过程包括公式定位、字符分割、字符排布、结构解析等。

首先，对图像中的横向和纵向像素分布进行强度分析，当出现像素强度突然降低或升高时，则可能出现山谷或山峰。分别按横向和纵向分析，找到所有山谷，再通过截取两侧山谷坐标，确定出公式位置。

其次，根据图像轮廓确定出图像的连通区域，并依次对各连通区域采用洪水填充算法，将附近与其连通的节点填充成相同颜色，将与其不连通的节点填充成不相同颜色，然后以颜色作为主要字符的判别特征，以此将图像中的字符分割出来。字符分割示意如图3所示。

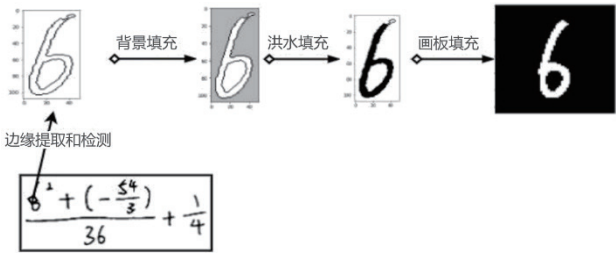


图 3 字符分割过程示意图

再次，将分割出来的字符构建字符排布树，字符位置有三种：第一种位于父字符之上，第二种位于父字符里面，第三种位于父字符下面。根据字符间的关联关系，构建字符排布树。数据存储结构为 [a,[b],[c],[d],e,g]，a 是父字符，b 是字符的上面，c 是字符的里面，d 是字符的下面，e 是 a 字符所在的坐标，g 是 a 字符预测的概率。

最后，将字符排布树解析变换成本类型。解析过程中需要结合字符背景，如对于父字符为减号，它有以下几种可能情况：一是单纯的减号，其上属结点、下属结点和中间结点均无子结点，则可直接解析成减号；二是等号，等号的特征是其父结点的三个位置有且仅有一个位置有一个子节点且该子节点为减号；三是分数，其上属结点和下属结点均有字符的存在；四是除号，其上属结点和下属结点有且仅有一个元素且该元素为点。不同的情况解析的结果是不相同的。解析的结果分为 Latex 格式和 Eval 格式，Latex 格式提供给前端展示，Eval 格式供后端算式计算，字符排布树和解析结果如图4所示。

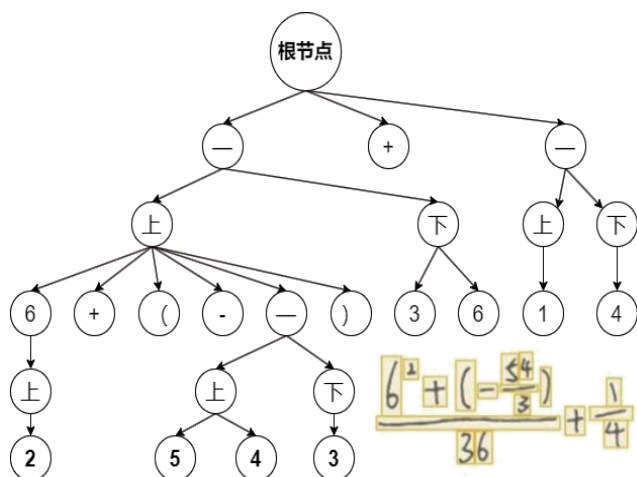


图4 字符排布树

2.3 数据训练

为了使手写算式识别模型具有较高的准确率，对识别模型进行了大量的数据训练。首先，选用优化后具有一定规模量的 MNIST+CASIA+ 自建数据集。它优化了 MNIST 数据集中的欧美版的数字书写方式的字符，并结合了 CASIA 集中适用亚洲人手写书写的数字字符，并在此基础上，增加了自手写的 600 份常用符号的数据集。其次，在选用的数据集中增加噪点、缩放、旋转和位移等方式进行数据增广，以此增加识别模型的抗过拟合度。再次，改进基于 LeNet 卷积神经网络^[11] 结构，第一层和第二层是 32 个 Filters 的 5*5 卷积层，第三层是 2*2 的池化层，第四层是丢失 50% 的 Dropout 层，第五层和第六层是 64 个 Filters 的 3*3 的卷积层，第七层是 2*2 步数为 2 的池化层，第八层是 Dropout 层，第九层是 256 个 ReLU 激活函数的神经元层，第十层是 Dropout 层，第十一层是 softmax 的分类输出层。最后，采用 ReduceLROnPlateau 优化方法对数据训练，通过一定的 epochs 后进行 learningRate 的下降，Optimaizer 选择的 RMSprop，初始 learningRate 设置为 0.001，阻碍率为 0.9，损失函数选择 categorical_crossentropy 交叉熵函数，每个 Batch_Size 是 64 个，总共进行 4 个 epochs 模型完成收敛。

3 实验测试与结果分析

3.1 实验环境及测试数据

根据上述研究，设计并实现了基于手机拍照的手写算式识别系统。该系统采用三层体系架构，对手机拍照的手写算式分别经过预处理、公式定位、算式识别等过程，基本实现了对照片中的手写算式的识别。该系统的具体环境参数如表 1 所示。

表 1 测试系统的环境参数

类别	名称	配置
硬件信息	CPU	AMD Ryzen 7 5800H with Radeon Graphics 3.20 GHz
	内存	16 GB
	显卡	NVIDIA GeForce RTX 3060
	硬盘	512 GB
软件信息	操作系统	Window 10 64 bits
	Python IDE	PyCharm 2023.1.5
	Java IDE	IntelliJ IDEA 2023.3
开发环境	Java	jdk1.8.0
	Python	Python3.8.2

测试数据来源于不同环境、不同光线、不同像素、不同手机品牌的拍照图片，字符包括小数、分数、指数、根式，算式模型包括四则模型、矩阵模型、方程模型，合计 600 多份手写算式。

3.2 测试结果分析

3.2.1 准确率测试

分别对拍照照片中的手写算式进行测试，在得到正确字符的前提下，对各种算式和算式模型进行测试。测试结果为：四则模型测试集的准确率为 99.6%，矩阵模型测试集的准确率为 99.5%，方程组模型测试集的准确率为 99.6%。

3.2.2 对比测试

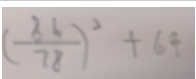
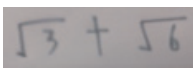
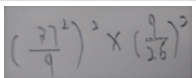
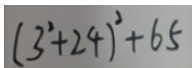
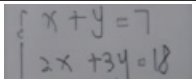
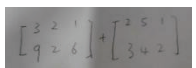
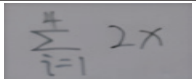
为了验证本文提出的手写算式识别模型的有效性和可用性，将它与市场流行产品 photomath、symbolab 进行了对比测试，结果如表 2 所示。

由表 2 可知，在四则模型的测试中，本文提出的模型与 photomath 能得出算式的运算结果，而 symbolab 在测试数据序号 1 和序号 3 中，仅得出算式的化简结果。在方程组模型测试中，都能得到正确的方程解。在矩阵模型测试中，本文提出的模型与 photomath 都能得到具体的矩阵运算结果，而 symbolab 仅识别出了算式，并没有进行矩阵运算。最后，在累加运算的测试中，本文模型与 symbolab 能正确识别出累加算式并进行结果运算，而 photomath 则无法识别算式。由此可知，本文提出的识别模型可以有效地对手机拍照的手写算式进行识别，并能对识别结果进行运算，且具有较好的识别效果。

4 结语

对手写算式的识别对当前移动环境下的智能化教育具有重要意义。手机拍照的手写算式受光线不均、图像像素、目标模糊、拍摄角度等因素影响，这给算式识别带来了一定难

表 2 对比测试结果

序号	测试模式	拍照照片	原式	结果 (本文方案)	结果 (photomath)	结果 (symbolab)
1	混合四则		$(\frac{86}{78})^2 + 64$	65.215 647 600 262 98	65.215 65	$\frac{99\ 193}{1521}$
2	混合四则		$\sqrt{3} + \sqrt{6}$	4.181 540 065 035 205 5	4.181 54	4.181 54
3	混合四则		$(\frac{37^2}{9})^2 \times (\frac{9}{27})^2$	2 772.427 514 792 899 3	2 772.427 51	$\frac{1\ 874\ 167}{676}$
4	混合四则		$(3^2 + 24)^2 + 65$	1154	1154	1154
5	二元一次		$\begin{cases} x + y = 7 \\ 2x + 3y = 18 \end{cases}$	$\begin{cases} x = 3 \\ y = 4 \end{cases}$	(x,y)=(3,4)	x=3, y=4
6	矩阵运算		$\begin{bmatrix} 3 & 2 & 1 \\ 9 & 2 & 6 \end{bmatrix} + \begin{bmatrix} 2 & 5 & 1 \\ 3 & 4 & 2 \end{bmatrix}$	$\begin{bmatrix} 5 & 7 & 2 \\ 12 & 6 & 8 \end{bmatrix}$	$\begin{bmatrix} 5 & 7 & 2 \\ 12 & 6 & 8 \end{bmatrix}$	$\begin{bmatrix} 3 & 2 & 1 \\ 9 & 2 & 6 \end{bmatrix} + \begin{bmatrix} 2 & 5 & 1 \\ 3 & 4 & 2 \end{bmatrix}$
7	累加运算		$\sum_{i=1}^4 2x$	20	无法识别	20

度。本文针对移动环境下手写算式识别准确率不高、效果不理想的问题，提出了基于手机拍照的手写算式识别的解决方案。该方案对预处理后的图片，利用公式定位、字符分割、字符排布树、结构解析等方法识别出手写算式，并通过数据集训练提升识别准确率。最后，通过具体的手写算式识别系统进行测试，结果表明所提方法具有较高的识别准确率和较好的识别效果。

参考文献:

[1] 沈佳伟,周宇昂,赵天宇,等.手写数学表达式识别方法研究[J].福建电脑,2021,37(7):59-61.

[2] 甘晓英,白阳,何晓栋,等.一种并行二值图像连通域标记算法[J].计算机与数字工程,2021,49(5):993-997+1002.

[3] 雷嘉兴,王伟.二维傅里叶图像预处理对DNN网络的影响研究[J].科学技术创新,2022(11):61-64.

[4] ALY W, UCHIDA S, SUZUKI M. Automatic classification of spatial relationships among mathematical symbols using geometric features[J]. IEICE transactions on information and systems, 2009, E92-D(11): 2235-2243.

[5] 胡龙灿,杨帆,樊爱军.手写数学公式的识别研究及在Android上的应用[J].计算机应用与软件,2014,31(8):28-31+44.

[6] DAI J, SUN Y, SU G, et al. Recognizing offline handwritten mathematical expressions efficiently[EB/OL]. (2019-01-10) [2024-02-22]. <https://doi.org/10.1145/3306500.3306543>.

[7] 付鹏斌,彭荆旋,杨惠荣,等.基于多重几何特征和CNN的脱机手写算式识别[J].计算机系统应用,2020,29(8):271-279.

[8] 王巍,周庆华.基于改进Faster R-CNN的基础算式检测与定位[J].智能计算机与应用,2022,12(12):164-168.

[9] 李卓璇,周亚同.改进DBNet的电商图像文字检测算法研究[J].计算机工程与科学,2023,45(11):2008-2017.

[10] 倪波,柯亨进,蔡贤涛.半监督学习下复杂背景图像边缘检测仿真[J].计算机仿真,2023,40(12):269-272+320.

[11] 孙隽姝,王树徽,杨晨雪,等.附加特征图增强的图卷积神经网络[J].计算机学报,2023,46(9):1900-1918.

【作者简介】

刘兴豪(2002—),男,广东汕尾人,本科,研究方向:人工智能。

陈芷妍(2003—),女,广东广州人,本科,研究方向:计算机应用。

何滨(2004—),男,广东韶关人,本科,研究方向:人工智能。

童保鑫(2003—),男,湖北鄂州人,本科,研究方向:人工智能

李文全(1980—),通信作者(email: 78192128@qq.com),男,江西龙南人,硕士,副教授,研究方向:数据挖掘、模糊数学。

(收稿日期: 2024-03-21)