基于 SSD 框架的自然场景盲文识别方法

吴东, 卢利琼*, 熊建芳

(岭南师范学院计算机与智能教育学院 湛江 534002)
(lulq@lingnan.edu.cn)

摘 要: 盲文是视障人士学习知识和技术的工具, 正常人通常对盲文知之甚少, 造成正常人与盲人之间的沟通障碍重重. 为此, 首先构建了自然场景盲文段图像数据集, 该数据集中包含 1 157 幅不同宽高比、不同背景的盲文段图像和对应的标签信息; 随后分析自然场景图像中盲文的特点, 并基于 SSD 框架提出自然场景盲文识别方法. 所提方法根据盲文字符尺寸小且具有固定宽高比的特点选择用于识别的特征层, 设计 CNN 结构、默认框大小、盲文字符标签、图像输入策略和损失函数, 以提高盲文字符识别的准确率; 根据盲文字符中盲文点位于边缘区域的特点设计像素层面的注意力机制, 提高盲文字符识别的回归率. 实验结果表明, 在所构建的盲文段图像数据集上, 所提方法的 H 值达到 0.903, 盲文字符检测速度为 66.22 帧/s; 与 SSD, Faster R-CNN, EAST, 以及基于 CNN 的盲文识别方法 EAST-Edge 和 UNet-Braille 相比, 该方法的盲文识别性能提升明显.

关键词: 自然场景图像; 盲文识别; SSD; 卷积神经网络
中图法分类号: TP391.41 **DOI:** 10.3724/SP.J.1089.2023-00630

Natural Scene Braille Recognition Based on SSD Framework

Wu Dong, Lu Liqiong*, and Xiong Jianfang

(*School of Computer Science and Intelligence Education, Lingnan Normal University, Zhanjiang* 534002)

Abstract: Braille is a tool for visually impaired individuals to learn knowledge and skills. Normal individuals often have little knowledge of braille, resulting in numerous communication barriers between normal and visually impaired individuals. A natural scene braille segment image dataset was first constructed, which includes 1 157 braille segment images with different aspect ratios and backgrounds, as well as corresponding label information; Subsequently, the characteristics of braille in natural scene images were analyzed, and a natural scene braille recognition method was proposed based on the SSD framework. The proposed method selects feature layers for braille recognition, designs CNN structures, default box sizes, braille character labels, image input strategies and loss functions to improve the precision of braille recognition based on the small size and fixed aspect ratio of braille characters. Moreover, a pixel level attention mechanism is designed based on the characteristic that braille points are located in the edge region of braille characters, to improve the recall of braille recognition. Experimental results show the proposed method achieves a value of 0.903 in H and a braille character recognition speed of 66.22 frame/s. Compared with SSD, Faster R-CNN, EAST, and CNN based methods named EAST-Edge and UNet-Braille on the constructed natural scene braille segment image datasets, the braille recognition performance of the proposed method has been significantly improved.

Key words: natural scene images; braille recognition; single shot multibox detector; convolutional neural network

收稿日期: 2023-09-22; 修回日期: 2024-07-22. **基金项目**: 广东省湛江市科技发展专项资金竞争性分配项目(2022A01005); 广东省哲学社会科学规划项目(GD24CJY21). **吴东**(1981—), 男, 硕士, 副教授, 主要研究方向为文本识别; **卢利琼**(1980—), 女, 博士, 副教授, 论文通信作者, 主要研究方向为计算机视觉; **熊建芳**(1980—), 女, 硕士, 讲师, 主要研究方向为模式识别.

1 相关工作

世界卫生组织数据显示，全球视力障碍者约 2.53 亿. 我国视力残疾人约 1 732 万人[1]. 视障人士通过阅读盲文书籍学习知识和技术，使用盲文记录心得、参与考试和沟通交流，盲文是他们与世界联系的纽带和桥梁. 但是，由于正常人对盲文知之甚少，导致其与视障人士之间的沟通交流障碍重重；特别是视障人士的朋友和家人，如果不懂盲文，则难以了解视障人士的学习、工作等各方面的情况. 盲文识别是利用人工智能技术自动识别图像中的盲文，帮助正常人快速地读取盲文信息，能有效地助力正常人与视障人士之间的沟通交流，亦可应用于盲文书籍电子化、盲汉文翻译等产品中.

盲文是 Louis Braille 于 1824 年在法国的一所盲人学校发明的，用触觉获取信息的文字. 盲文由盲文字符组成，而盲文字符由 6 个固定位置的盲点按不同的出现顺序组成，图 1 所示为所有 63 种盲文字符组合的示意图.

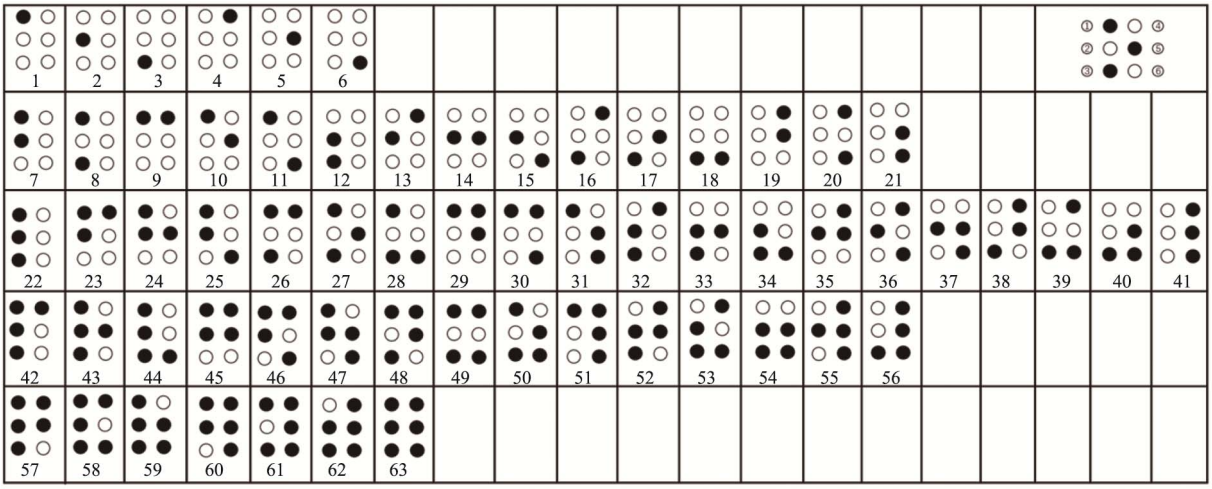


图 1 63 种盲文字符组合的示意图

与其他语言相比，盲文属于小众语言，目前对盲文识别的研究大多集中在扫描文档图像中，对自然场景中的盲文识别研究较少. 从技术上层面看，现有的盲文识别方法分为 3 类.

(1) 利用传统的图像处理手段，包括图像分割、图像角度计算、图像形态学滤波等，结合统计学方法和多层语义分析对盲文进行识别. 早期的盲文识别方法大多属于此类，这类方法针对扫描的盲文文档图像，步骤烦琐且识别准确率不高. Shahbazkia 等[2]利用二值化阈值分割的方法检测盲文后，通过语义分析方法识别盲文；Fahiem[3]提出一种联系上下文的将盲文翻译成乌尔都语的方法；Tai 等[4]通过自动确定垂直和水平方向上的旋转角度、凹痕和间距，以及图像旋转角度的关键因素，提出一种盲文识别系统；杨胜等[5]利用指纹传感器获取盲文文档中的凸点图像，然后对盲文图像进行二值化分割、形态学滤波和形状识别，获得凸点位置后进行盲文识别；江铭虎等[6]整合汉语拼音、词汇、语法及语义等知识，进行多层次的歧义校正和纠错处理，实现了盲文到汉字的互相转换；周春耕等[7]将基于统计的 Markov 模型与规则量化模型相结合，用 Viterbi 算法找到最佳语句，提高盲汉文翻译效果；李念峰等[8]对处于特殊位置的盲点进行质心替换，然后依赖信息库进行多次歧义校正，完成盲文字符的识别和翻译.

(2) 采用人工提取特征和传统机器学习相结合的方法，如支持向量机(support vector machine, SVM)等，提取盲文后进行语义识别；Li 等[9]采用人工提取 Haar 特征后，利用 SVM 获取双面盲文扫描文档图像中的盲文凸点，最后将凸点组合成盲文字符；尹佳[10]利用 SVM 检测盲点后，结合二值化图像对盲文进行识别.

(3) 引入近年来在人工智能领域大放异彩的卷积神经网络(convolutional neural network, CNN)对盲文进行检测和识别的方法，如 Li 等[11]使用改进 U-Net[12]架构的神经网络从像素级别检验某像素是否属于盲文字符，再聚合盲文像素后得到盲文字符区域；Lu 等[13]结合盲文字符的边缘特征和 Anchor-free 的思想提出了自然场景盲文字符检测方法；刘泽阳[14]利用图像的 Haar 特征和图像相似

性特征提取盲文凸点，通过提取整行盲文并输入到卷积循环神经网络(convolutional recurrent neural network，CRNN)中进行盲文字符识别；李荣瑞[15]提出通过自动校正和分割盲文图像并将其输入CNN 中进行识别的方法.

与第 1 类方法相比，第 2 类和第 3 类方法的图像处理步骤减少，识别准确率进一步提升. 特别是第 3 类基于 CNN 的盲文识别方法已经成为盲文识别的主流技术，且识别准确率较高.

现有的盲文识别方法有 2 个不足：(1) 大多数以扫描盲文文档中的盲文为研究对象，忽略了自然场景盲文图像中的盲文. 实际上，随着智能手机的普及，随时随地拍摄图像比用扫描仪扫描盲文文档要便捷得多，因此自然场景图像中的盲文识别应该成为研究热点；(2) 几乎都是先检测出盲文字符，然后针对盲文字符进行识别. 但是，在识别过程中，相邻的盲文字符之间是有语义联系的，如果在检测阶段没有将相邻的盲文字符检测出来会导致识别阶段的错误. 因此，在盲文检测阶段如果能检测出盲文段，那么在识别阶段则能更好地利用相邻盲文字符的语义关联进行识别.

基于以上分析，自然场景盲文识别可以先从自然场景图像中检测出盲文段的位置，然后从盲文段中对盲文字符进行识别，如图 2 所示.

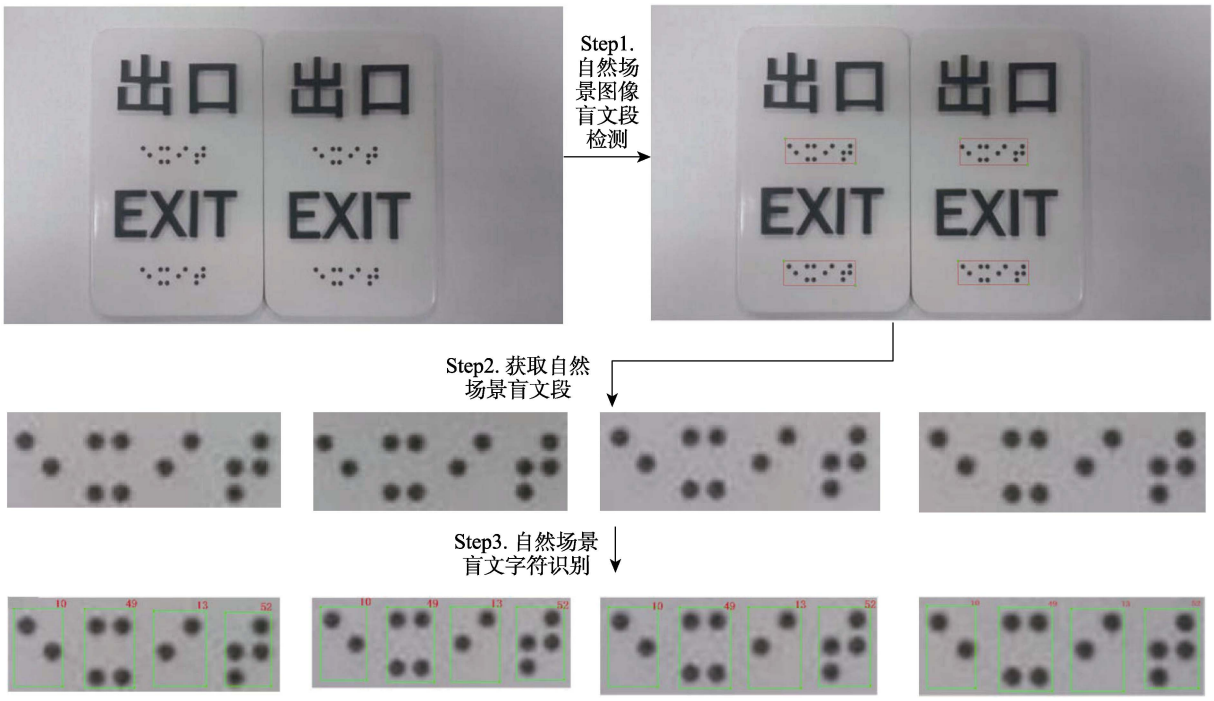


图 2　自然场景盲文识别流程图

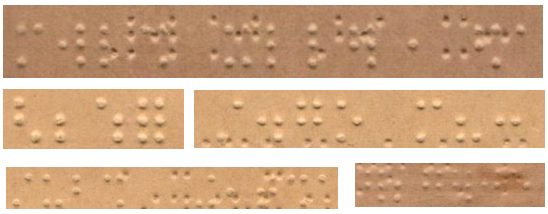
本文完成的工作是图 2 中的 Step2 和 Step3. 本文以自然场景盲文段作为研究对象，利用 CNN 有效地提取自然场景文本特征的优势[16-18]，提出基于 SSD(single shot multibox detector)框架[19]的自然场景盲文识别方法. 首先构建自然场景盲文段图像数据集，其中有 1157 幅不同宽高比、不同背景的盲文段图像，且每幅图像中盲文字符的位置和语义标签都标记在对应.xml 和.txt 文件中；然后结合自然场景图像中盲文的特点和 SSD 框架完成 CNN 结构设计，包括盲文识别特征层选择、像素级别的注意力机制(pixel attention，PA)设计、默认框大小设计、盲文字符标签和损失函数设计，以及盲文图像输入处理策略，提出自然场景盲文识别方法. 与 SSD[19]，Faster R-CNN[20]，EAST[21]，以及基于 CNN 的盲文识别方法 EAST-Edge[13]和 UNet-Braille[11]进行对比实验，结果表明，本文方法具备更好的盲文识别性能.

2　自然场景盲文段图像数据集

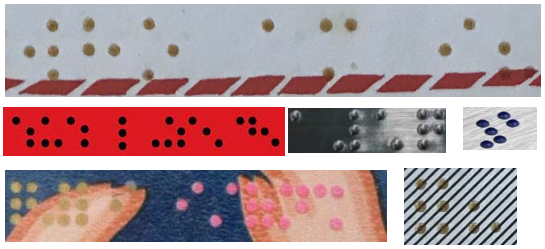
首先构造自然场景盲文段图像数据集，这些图像全部来源于手机拍摄或者网络下载，且它们具有不同的背景和不同的宽高比；盲文段中，盲文字符的大小和颜色变化较大，甚至有些盲文段图像的背景和盲文点融为一体. 该数据集中总共有 1 157 幅盲文段图像，其中，训练集中有 925 幅，测试集中

有 232 幅; 该数据集已经上传到 ScienceDB 平台, 访问链接为 https://cstr.cn/31253.11.sciencedb.j00188.00007.

与已有的扫描盲文文档图像中的背景简单、颜色一致、尺寸固定、排列方向和方式固定的盲文段相比, 自然场景图像中的盲文段在背景、颜色和尺寸、呈现方式等方面更加复杂多变. 图 3 所示为扫描盲文文档图像中的盲文段和自然场景图像中的盲文段的对比. 可以看出, 图 3a 中, 扫描图像中的盲文段在呈现形式上只有凸点和凹点 2 种形式, 颜色单一, 盲文字符大小和字符间隔基本上保持不变, 盲文字符排列方式都是水平方向, 背景都是纸质材料; 图 3b 中, 自然场景中的盲文段在呈形式上更加多样化, 水平方向和非水平方向都有, 盲文字符大小、颜色和盲文字符间隔不固定, 盲文图像背景复杂多变.



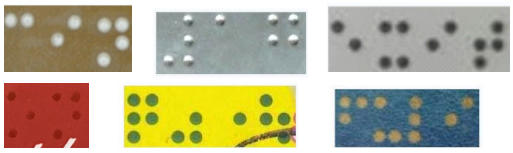
a. 扫描盲文文档图像



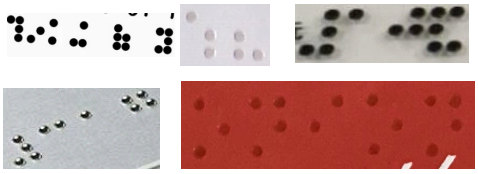
b. 自然场景盲文图像

图 3 扫描盲文文档和自然场景图像中的盲文段对比

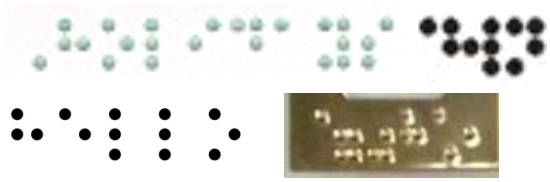
为了更好地体现自然场景中盲文段的特点, 更加详细地列举了其中不同类型盲文段的示例, 如图 4 所示. 可以看出, 呈现方式多样化, 盲文字符颜色变化大(白色、银色、黑色、红色、绿色和黄色等); 盲文尺寸(大、中、小), 盲文字符排列方向(水平、非水平)和盲文字符间隔都不固定, 而是呈变化频繁的状态; 图像背景复杂多变, 在颜色(红色、蓝色、黑色和灰色等), 材质(纸张、电梯和广告牌等)和图案(纯色、斜纹、其他图案等)上变化较大, 甚至有些盲文字符隐没在背景中; 在呈现形式上基本上都是长方形, 但是长和宽的大小变化很大, 给后续的工作包括盲文字符位置检测和盲文字符识别带来了极大的困难.



a. 不同颜色



b. 不同尺寸和排列方式



c. 不同盲文字符间隔



d. 复杂背景

图 4 自然场景中的盲文段图像复杂多变

自然场景盲文段图像数据集的目录结构如图 5 所示. character_label 文件中存放 3 种格式的盲文段图像: (1) 以 ICDAR-2015 格式存储的盲文段图像和标签文件. 该格式下, 每个.jpg 文件对应一个.txt 文件, 其中每行存储一个盲文字符矩形框的位置和识别标签, 其数据分别对应矩形框 4 个点的坐标和识别的数字标签. (2) 数据的原始格式, 存储在文件夹 org 中. 该文件夹下每个.jpg 文件对应一个 labelme 3.16.5 软件标记的.json 文件. (3) voc 格式, 存放在 voc-data 文件中. 该文件夹下根据 voc

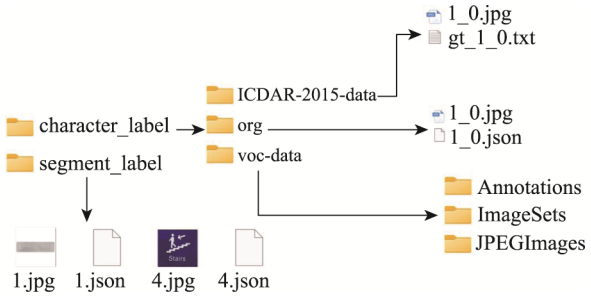


图 5 自然场景盲文段图像数据集目录结构

格式存放图像和对应的.xml 文件，在.xml 文件中标记每个盲文字符矩形框的位置和对应的数字标签信息. 此外，自然场景盲文原始图像和对应盲文段的标记.json 文件存放在文件夹 segment_ label 中.

3　自然场景盲文识别方法

自然场景盲文识别任务类似于自然场景常见的目标识别任务，自然场景中的盲文字符与自然场景中的其他目标类似，存在背景复杂、目标颜色和尺寸变化大和光线不均匀等情况[22-23]，因此，可以基于现有目标识别框架设计盲文字符识别. 进一步地分析发现，自然场景中的盲文字符和盲文段在书写和呈现形式上也有其特殊的地方: (1) 盲文段图像宽高比变化很大，而现在大多数目标识别方法的输入图像都具有固定的宽高比，如果直接将盲文段图像调整到固定大小，一定会引起图像的失真，从而影响后续的盲文字符识别任务的性能; (2) 盲文字符虽然在尺寸和颜色上有一定变化，但是在呈现形式上宽高比是固定的，如图 1 所示，每个盲文字符的高度和宽度的比例基本上是 3∶2; (3) 自然场景图像中的盲文字符大多以小尺寸的形式出现，大尺寸目标相对较少，且基本不会孤立出现，其前后都有相邻的盲文字符.

基于以上的分析，本文选定目标识别中经典的 SSD 框架作为本文自然场景盲文字符识别的基本框架. 主要有 3 点考虑: (1) SSD 在多个 CNN 特征层上进行目标识别，稍加修改后有利于识别出更多小尺寸目标; (2) SSD 通过预先设定固定大小的默认框对 CNN 进行训练，与盲文字符具有固定的宽高比可以完美契合，能够快速、准确地进行盲文字符识别; (3) SSD 属于 One-stage 目标识别方法，其识别准确率和速度均有一定保障，基于该框架设计盲文字符识别在性能上有一定的保障.

下面分别从自然场景盲文识别框架、CNN 结构设计和识别特征层选择、PA 设计、默认框大小设计、盲文字符标签和损失函数设计，以及图像输入处理策略设计 6 个方面介绍本文提出的自然场景盲文识别方法.

3.1　自然场景盲文识别框架

本文以 SSD 框架为基础架构, VGG-16 为基础结构，通过设计不同的机制和策略，提出自然场景盲文识别方法，其框架如图 6 所示. 对于 300×300×3 像素的输入图像，首先经过 VGG-16 的多阶段卷积和池化层，随后涉及 5 个阶段的卷积操作; 紧接着在特征层 Conv4_3, FC7, Conv6_2, Conv7_2 和 Conv8_2 后增加 PA 模块，生成的新特征层经过 3×3 的卷积操作后作为盲文识别的特征层; 其中，Conv4_3 和 Conv8_2 对应的特征层在每个单元格上设置 4 种不同尺寸的默认框，其余特征层中每个单元格设置 6 种不同尺寸的默认框，因此所有特征层共有 8 728 个默认框. CNN 训练过程中，计算这些默认框与真实盲文字符框的 IoU 值，如果其值大于给定的阈值，则认为是正样本; 然后对该样本与真实框在中心点坐标、高度和宽度上的偏差和类别进行预测. 在测试阶段，将检测出来的盲文字符框通过非最大化抑制 (non-maximum suppression, NMS)算法[24]过滤后得到最终的识别结果.

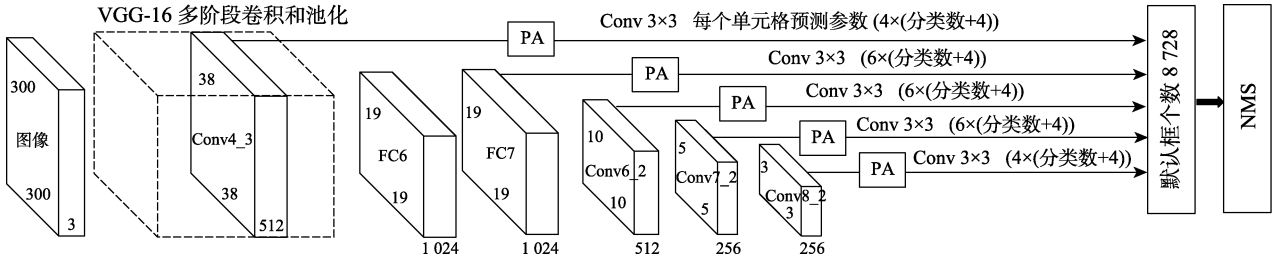


图 6　本文方法框架

3.2　CNN 结构设计和识别特征层选择

以 VGG-16 为基础结构，结合自然场景盲文字符的特点, CNN 结构设计如图 7 所示. 由于自然场景中的盲文字符以小尺寸居多且盲文字符通常紧挨着形成盲文段，本文在设计 CNN 网络结构和特征层选择时都结合了盲文字符的这个特点，因此它们与 SSD 框架不尽相同. 图 7 中，在 VGG-16 的基础上(Conv1~Conv5)增加 5 个阶段的卷积操作. 每个阶段都包含 3×3 和 1×1 的两个卷积层, FC6 和 FC7 的卷积层只用于调节通道数目; Conv6, Conv7 和 Conv8 中 1×1 的卷积层用于调节通道数目，3×3 的卷积层采用的步长为 2，用于调节特征层大小.

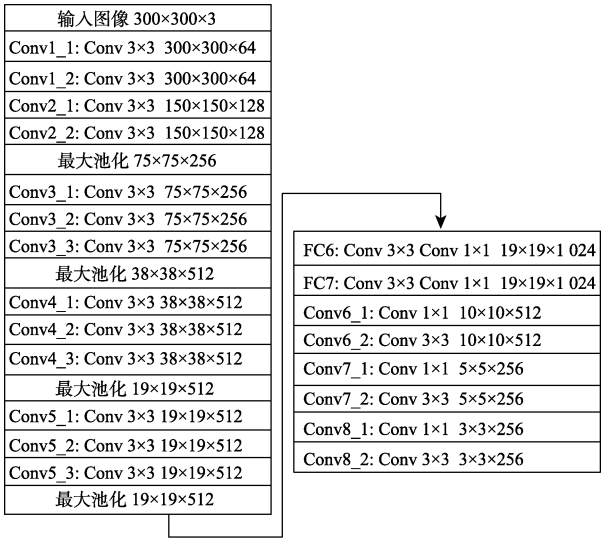


图 7 本文设计的 CNN 结构

本文在 Conv4_3, FC7, Conv6_2, Conv7_2 和 Conv8_2 这 5 个特征层上, 经 PA 模块和一个 3×3 的卷积层后进行盲文字符分类和盲文字符位置的预测, 其中, 尺寸大的特征层用于预测小目标, 尺寸小的特征层用于预测大目标. 考虑大尺寸的特征层设置较多的默认框可能会影响效率, 且利用小尺寸特征层预测位置的盲文字符不会太多, 在特征层 Conv4_3 和 Conv8_2 的每个单元格上只设置 4 个默认框, 其他特征层每个单元格上设置 6 个默认框. 根据特征层的大小和每个特征层单元格上设置的默认框个数, 可以计算出这 5 个特征层上共有 8 728 个默认框, 其计算过程如图 8 所示.

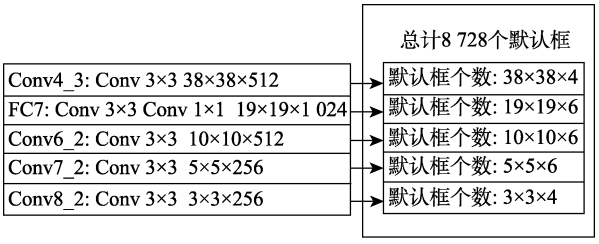


图 8 5 个特征层上默认框总数计算过程

3.3 PA 设计

与自然场景中的其他对象不同, 盲文字符由盲文点组成, 且盲文点离散地分布在盲文字符的边缘区域, 即盲文字符矩形框内, 不是所有的像素都对字符识别有贡献, 只有位于盲文点上的像素才能形成有效助力. PA 模块能够从 CNN 特征层中挖掘特定像素的特征, 有助于利用这些特征加强后续的回归和分类任务, 使得盲文字符识别更加容易.

本文在特征层 Conv4_3, FC7, Conv6_2, Conv7_2 和 Conv8-2 后面增加 PA 模块, 生成对应新的特征层后用于盲文字符矩形框的位置检测和盲文字符分类识别. 本文设计的 PA 模块中, 由 2 个 3×3 的卷积层、1 个 ReLU 层和 Sigmoid 响应函数组成的注意力层与原有的特征层进行 Element-wise 乘法操作后得到新的特征层. PA 设计的结构如图 9 所示, 后续实验结果表明, 增加 PA 模块后, 盲文识别的回归率有一定的提升.

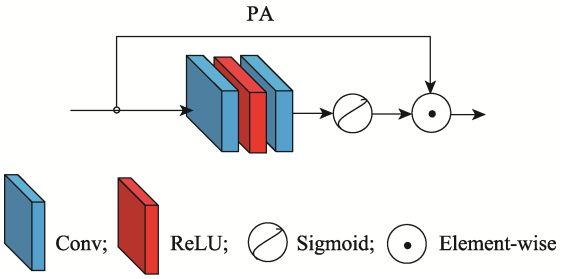


图 9 PA 结构设计

3.4 默认框大小设计

根据目标特点设计默认框的大小, 鉴于自然场景图像中通常具有盲文字符较小的特点, 本文设置 $r_{\min}=5$, $r_{\max}=75$. 如果输入图像的大小为 300×300 像素, 那么计算得到的 $s_{\min}$ =[6, 15, 66, 117, 168], $s_{\max}$ =[12,66,117,168,219]. 由于本文特征层选择和盲文字符框大小与 SSD 中的设置均不同, 因此在计算 $s_{\min}$ 和 $s_{\max}$ 过程中与 SSD 与有一定差异. 与 SSD 方法相比, 首先, $r_{\min}$ 和 $r_{\max}$ 的值变小了; 其次, 由于本文的特征层只选择了 5 个, 因此在计算过程中舍弃了 $s_{\min}$ 和 $s_{\max}$ 的最后一个值; 最后, $s_{\min}$ 和 $s_{\max}$ 的计算过程和结果有变化. 计算出 $s_{\min}$ 和 $s_{\max}$ 的值后, 再根据 r_{aspect} 的值就可以确定特征层上单元格对应的每个默认框的宽度和高度(计算默认框大小的方法与 SSD 类似). 本文中, $r_{\text{aspect}}=\left[[2],[2,3],[2,3],[2,3],[2]\right]$. 计算 $s_{\min}$ 和 $s_{\max}$ 的步骤如下:

输入. 图像大小 $d_{\min}=30$.

输出. $s_{\min}$ 和 $s_{\max}$.

Step1. 设置初始值 $r_{\min}=5$, $r_{\max}=75$.

Step2. 计算步长 $t=(r_{\max}-r_{\min})/(6-2)$.

Step3. 循环计算 $s_{\min}$ 和 $s_{\max}$.

for r in range($r_{\min}$, $r_{\max}-t$, t):

$s_{\min}$ 中增加值 $d_{\min}\times r/100$,

$s_{\max}$ 中增加值 $d_{\min}\times(r+t)/100$.

Step4. 增加 $s_{\min}$ 和 $s_{\max}$ 的第 1 个值.

$s_{\min}=[d_{\min}\times 2/100]+s_{\min}$，

$s_{\max}=[d_{\min}\times 4/100]+s_{\max}$.

Step5. 返回 $s_{\min}$ 和 $s_{\max}$.

3.5　盲文字符标签和损失函数设计

盲文字符识别标签从 2 个层面进行设计，一是盲文字符类别标签，二是盲文字符框的矩形框位置标签. 设计盲文类别标签时，按照图 1 所示的盲文字符，根据实心盲点位置的不同总共 63 个类别标签；另外，增加标签值 0 表示背景类. 盲文字符矩形框位置标签以中心点坐标(x , y)和矩形框的宽 w 和高 h 组成.

损失函数的设计与盲文字符识别标签设计呈对应关系，总的损失函数由盲文字符分类损失(包括正负样本)和盲文字符位置损失(只包括正样本) 2 部分组成，公式为

$$L(p_i,t_i)=\frac{1}{N}\left(\sum_i L_{\text{cls}}(p_i,p_i^*)+\alpha\sum_j L_{\text{reg}}(t_i,t_i^*)\right).$$

其中，N 表示正负样本的总和，正样本和负样本的比例大于等于 1∶3；与 SSD 类似，α 的值用于调节 2 种损失的比重，此处设置为 1；p_i 和 p_i^* 表示样本分类预测概率和真实概率；t 表示盲文字符预测位置与真实位置的差值，t_i^* 表示样本位置与真实位置的差值；分类损失 L_{cls} 采用 Softmax Loss 计算，该分类损失由正样本的预测概率、真实概率和负样本预测为背景的概率计算，公式为

$$\sum_i L_{\text{cls}}(p_i,p_i^*)=-\sum_{i\in\text{pos}}^{N} p_i^*\times\lg p_i-\sum_{i\in\text{neg}}\lg p_i\ .$$

位置损失 L_{reg} 采用 $L_{\text{Smooth}_{L_1}}$ 计算，公式为

$$L_{\text{reg}}(t_i,t_i^*)=\sum_{i\in(x,y,w,h)} L_{\text{Smooth}_{L_1}}(t_i-t_i^*).$$

其中，$L_{\text{Smooth}_{L_1}}(x)=\begin{cases}0.5x^2, & \text{if } |x|\leqslant 1\\ |x|-0.5, & \text{otherwise}\end{cases}$.

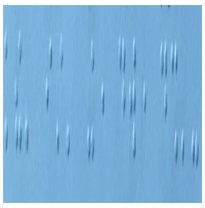
3.6　图像输入处理策略

自然场景盲文段图像在宽高比上变化极大，而用作 CNN 的输入图像大小通常是固定的，如果直接将盲文段图像重置到同一大小作为 CNN 的输入，会导致图像严重失真，盲文字符可能会出现比较严重的扭曲，识别性能不好；另外，如果直接将盲文段图像重置到同一大小作为输入，盲文字符就会在高度上几乎占满整幅图像，显然与前期分析自然场景图像中盲文字符基本都是小尺寸互相矛盾. 基于以上 2 个原因，盲文段图像不能直接作为 CNN 的输入，需要进行一定处理，并且这种处理既不能使图像失真，也不能改变盲文字符在自然场景图像中的固有特性.

本文采用的图像输入处理方案比较简单，但是非常有效. 首先将盲文段图像的宽度变成 300，在保证宽高比不变的情况下重新设置图像的高度；然后在图像的上下 2 个方向填补灰条，将图像大小扩充到 300×300 像素. 图 10 所示为盲文段图像对应的原图、直接调整成 300×300 像素的图和利用本文方法调整成 300×300 像素的图.



a. 原图



b. 直接调整到固定大小



c. 添加灰度图后的图

图 10　本文方法得到的盲文段图像

从图 10 中可以看出，本文提出的通过简单调整图像高度和增加灰度条的方法既保留了盲文字符的固有特性，也没有改变盲文字符的宽高比，符合设计初衷.

4　实验及结果分析

采用深度学习框架 Tensorflow[25]并基于 GTX 3060 实现本文方法. 在训练过程中，所有 CNN 均采用随机梯度下降进行优化. 由于盲文字符属于小尺寸目标，因此设置较小的学习率，初始学习率设为 0.000 6，最小为 0.000 006，学习率的变化过程与 SSD 方法一致. 整个训练过程分为冻结和非冻结 2 个阶段，冻结阶段的 epoch 设置为 100，batchsize 设置为 8；非冻结阶段的 epoch 设置为 1 000，batchsize 设置为 16. 图 11 所示为训练过程中，训练集和验证集的损失值变化过程. 可以看出，训练集和验证集的损失值都在逐步减少直至稳定，说明选择的实验参数是正确可靠的.

4.1　盲文识别性能评价指标

本文涉及的盲文识别有 2 类性能指标：

(1) 盲文字符检测性能指标. 采用 ICDAR 2015 评价协议，评价指标包括准确率(precision, P)、

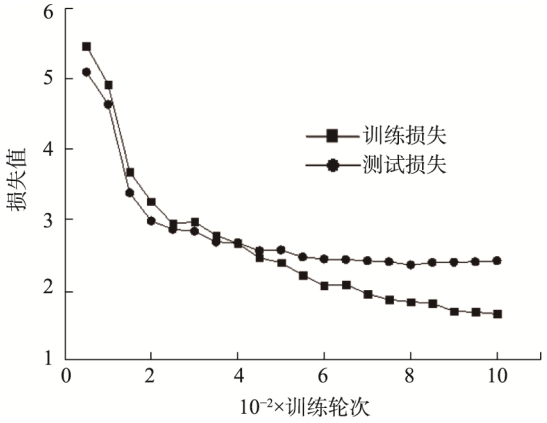


图 11 训练过程中训练集和测试集损失值对比

回归率(recall, R)和综合指标(hmean, H), 公式分别为

$$P=\frac{\mathrm{FP}}{\mathrm{FP}+\mathrm{TP}};$$

$$R=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}};$$

$$H=2\times\frac{P\times R}{P+R}.$$

其中, TP (true positive)表示检测正确的盲文字符个数; FP (false positive)表示检测错误的盲文字符个数; FN (false negative)表示未被检测出的盲文字符个数.

(2) 盲文字符检测速度的指标V, 表示每秒钟处理的图像数目, 公式为

$$V=N_{\mathrm{images}}/T_{\mathrm{images}}.$$

其中, N_{images}表示预测图像的数量; T_{images}表示所有图像预测的总时间.

4.2 盲文识别性能比较分析

在本文提出的盲文段图像数据集上, 将本文方法与下列方法进行对比: (1) 目标识别中经典的 2 种方法 SSD[19]和 Faster R-CNN[20]; (2) 自然场景文本识别中的优秀方法 EAST[21]; (3) 当前盲文识别中基于 CNN 提出的 UNet-Braille[11]和 East-Edge[13]方法. 对比实验结果如表 1 所示, 其中, 将本文方法分为不使用注意力机制的本文方法(简称为 SSD-Braille-without-Attention)和使用注意力机制的本文方法(简称为 SSD-Braille).

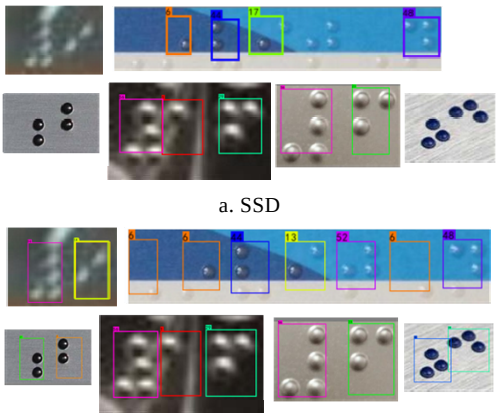
从表 1 可以看出: (1) 本文方法在检测性能指标 P, R 和 H 上均取得了最好的成绩, H 达到 0.903; (2) 与 SSD[19], Faster R-CNN[20]和 EAST[21]比, 本文方法的 H 值分别提升 0.102, 0.335 和 0.135, 性能提升非常明显; V 值比 Faster R-CNN[20], SSD[19]和 EAST[21]分别提升 58.07, 1.84 和 55.02, 说明本文方法在检测速度上占据一定优势, 甚至比 SSD[19]更快; 实验结果表明, 本文正确地分析了自然场景盲文字符的特点, CNN 结构设计和特征层选择策略、默认框大小设计, PA 设计和图像输入处理策略是正确且有效的; (3) 与 EAST-Edge[13]和 UNet-Braille[11]相比, 本文方法的 P, R 和 H 均有提升; EAST-Edge 考虑了盲文字符的边缘特征和盲文字符的固有特性, 与该方法相比, 本文方法的 P 从 0.907 提升到了 0.927, R 从 0.854 提升到 0.879; 实验结果表明, 本文方法考虑盲文字符固有特性和 PA 设计的思路更加有效; (4) 将 SSD-Braille-without-Attention 与 SSD-Braille 进行对比, 验证 PA 模块对本文方法性能提升的影响, 增加 PA 模块后, P, R 和 H 均有提升, 特别是 R 从 0.858 提升到了 0.879, 说明增加 PA 模块有助于发现更多的盲文字符, 与本文设计方法的思路一致.

表 1 7 种方法盲文识别性能对比

方法	P	R	H	V
SSD[19]	0.751	0.858	0.801	64.38
Faster R-CNN[20]	0.658	0.499	0.568	8.15
EAST[21]	0.834	0.712	0.768	11.20
UNet-Braille[11]	0.791	0.761	0.775	8.69
EAST-Edge[13]	0.907	0.854	0.879	11.20
SSD-Braille-without-Attention	0.920	0.858	0.888	
SSD-Braille	**0.927**	**0.879**	**0.903**	**66.22**

注. 粗体表示最优值.

本文方法以 SSD 为基本框架, 同时结合自然场景盲文字符在尺寸、书写形式上的特点进行重新设计. 与 SSD[19]方法相比, 本文方法有以下提升: (1) 用于盲文识别的特征层选择和特征层上单元格默认框大小和数目设置不同, 对应自然场景盲文字符尺寸小且字符通常不会孤立出现的特点, 舍弃了用于预测大尺寸目标的 Conv9 特征层, 从一定层面上减少默认框的数目, 加快了识别的速度; (2) 默认框大小设计不同, 根据自然场景盲文字符通常具有固定的宽高比和尺寸较小的特点, 重新设计了决定默认框大小的关键参数的值, 包括 $r_{\min}$, $r_{\max}$ 和 r_{aspect}; (3) 改进图像输入策略, 针对盲文段图像在呈现形式上具有宽度大而高度矮的特征, 引入添加灰度条的图像处理策略, 在保证不改变自然场景图像中盲文字符特点的情况下, 将图像调整成具有固定大小的输入, 便于 CNN 进行处理; (4) 增加 PA 模块, 挖掘盲文点的特征, 从而发现更多的盲文字符, 提升盲文字符识别的回归率. 图 12 所示为 SSD[19]和本文方法的盲文识别



a. SSD

b. 本文

图 12　2 种方法盲文识别样例效果对比

样例效果对比. 可以看出，与 SSD[19]方法相比，本文方法的改进效果非常明显.

4.3　盲文识别样例分析

为了更直观地展现本文方法的盲文识别效果，图 13 和图 14 所示分别为采用本文方法进行盲文识别时，成功和识别不完全的样例图. 从图 13 可以看出，在背景、颜色、尺寸、排列间隔等变化以及轻微倾斜的情况下，本文方法能够完整地检测出盲文字符，说明其具有良好的鲁棒性. 从图 14 可以看出，在盲文字符宽度变化大、排列方式变化大、盲文字符左右间隔不明确和边缘盲文孤点的情况下，本文方法存在盲文字符漏检.

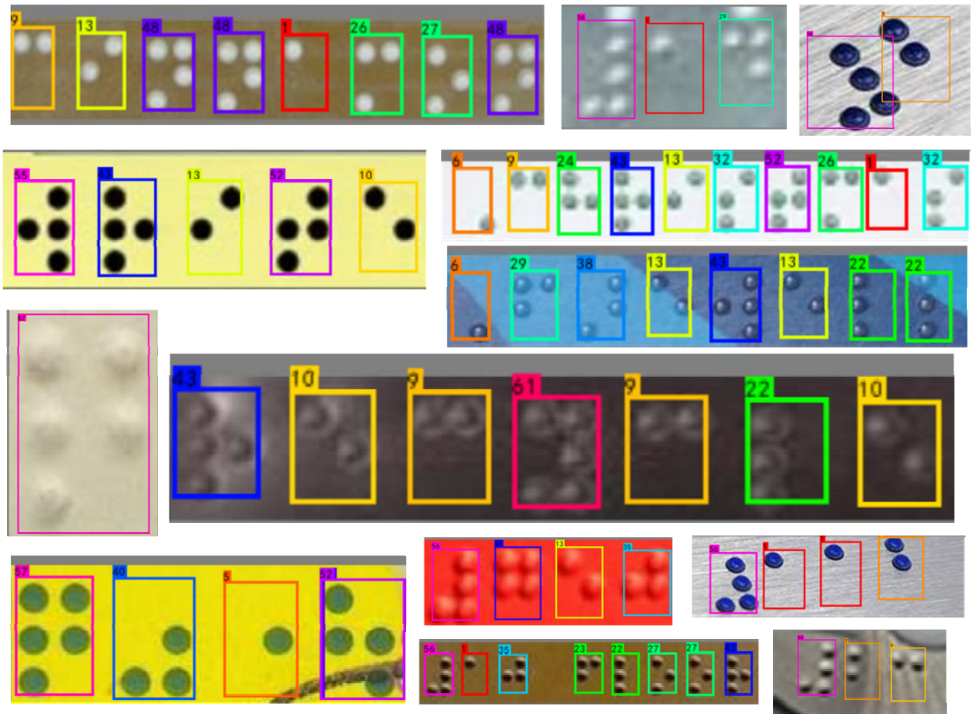


图 13　本文方法盲文识别成功样例

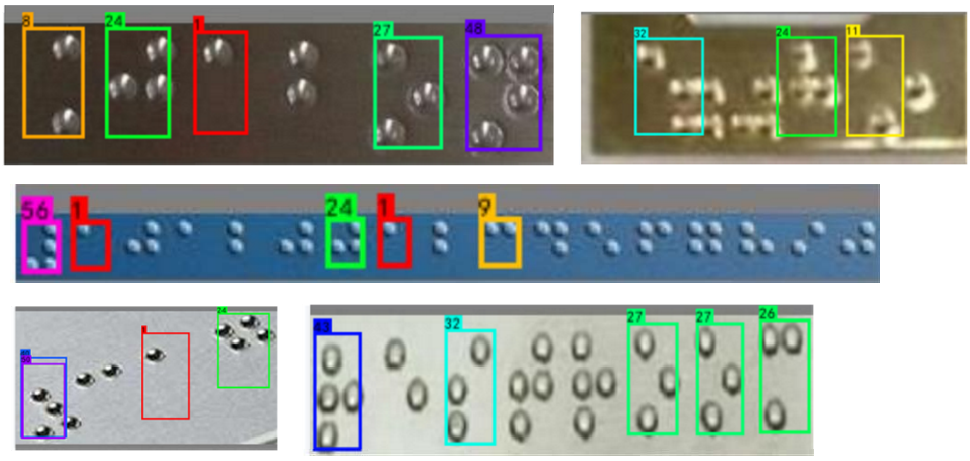


图 14　本文方法盲文识别不完整样例

5 结 语

针对正常人对盲文知之甚少，与盲人之间沟通交流障碍重重的问题，本文基于 SSD[19]框架的自然场景盲文识别方法，首先构造自然场景盲文段图像数据集；然后基于 SSD[19]框架并结合自然场景盲文的特点，提出自然场景盲文识别方法．与 SSD[19]方法相比，本文方法在盲文识别特征层选择、默认框大小设计、PA 设计和图像输入策略上，根据自然场景盲文字符的特点进行重新设计．在本文构造的盲文段图像数据集上，与 SSD[19], Faster R-CNN[20], EAST[21], EAST-Edge[13]和 UNet-Braille[11]方法进行实验对比的结果表明，本文方法识别盲文的 P，R，H 和 V 的值分别为 0.927，0.879，0.903 和 66.22，识别性能提升明显．

虽然本文方法在盲文背景、尺寸和颜色变化大的情况下能快速、准确地识别出盲文字符，但是在字符之间的宽度出现变化、盲文出现模糊和排列非水平的情况下存在漏检．未来将针对此类情况并结合更加完美的注意力机制等，进一步改进盲文识别方法．

参考文献(References):

[1] Xinhua News Agency. China will promote cross-border exchange of accessible format works[EB/OL]. [2023-09-22]. https://www.gov.cn/xinwen/2022-05/17/content_5690865.htm (in Chinese)
(新华社. 我国将推动无障碍格式版作品跨境交换[EB/OL]. [2023-09-22]. https://www.gov.cn/xinwen/2022-05/17/content_5690865.htm)

[2] Shahbazkia H R, Silva T T, Guerreiro R M. Automatic braille code translation system[C] //Proceedings of the 10th Iberoamerican Congress on Pattern Recognition. Heidelberg: Springer, 2005: 233-241

[3] Fahiem M A. A deterministic Turing machine for context sensitive translation of braille codes to Urdu text[C] //Proceedings of the 12th International Workshop on Combinatorial Image Analysis. Heidelberg: Springer, 2008: 342-351

[4] Tai Z F, Cheng S, Verma P, *et al*. Braille document recognition using Belief Propagation[J]. Journal of Visual Communication and Image Representation, 2010, 21(7): 722-730

[5] Yang Sheng, Tao Lei, He Leiying. Research on braille recognition algorithm based on fingerprint sensor[J]. Computer & Digital Engineering, 2021, 49(6): 1234-1239(in Chinese)
(杨胜, 陶磊, 贺磊盈. 基于指纹传感器的盲文识别算法研究[J]. 计算机与数字工程, 2021, 49(6): 1234-1239)

[6] Jiang Minghu, Zhu Xiaoyan, Xia Ying, *et al*. Braille translation based on multi knowledge[J]. Journal of Tsinghua University (Science and Technology), 2000, 40(9): 69-73(in Chinese)
(江铭虎, 朱小燕, 夏莹, 等. 基于多种知识的盲文翻译的研究[J]. 清华大学学报(自然科学版), 2000, 40(9): 69-73)

[7] Zhou Chungeng, Zhang Bingquan, Huang Heyan. The research and realization of Braille-Chinese machine translation system based on mixed language model[J]. Computer Engineering and Applications, 2003, 39(4): 127-130(in Chinese)
(周春耕, 张秉权, 黄河燕. 基于混合语言模型的盲汉机器翻译系统的研究与实现[J]. 计算机工程与应用, 2003, 39(4): 127-130)

[8] Li Nianfeng, Dong Yinghong, Xiao Zhiguo. Study on image processing based braille automatic identification system[J]. Manufacturing Automation, 2012, 34(3): 63-67(in Chinese)
(李念峰, 董迎红, 肖志国. 基于图像处理的盲文自动识别系统研究[J]. 制造业自动化, 2012, 34(3): 63-67)

[9] Li R Q, Liu H, Wang X D, *et al*. Effective optical braille recognition based on two-stage learning for double-sided braille image[C] //Proceedings of the 16th Pacific Rim International Conference on Artificial Intelligence. Heidelberg: Springer, 2019: 150-163

[10] Yin Jia. The key technology research on paper-mediated braille automatic recognition system[D]. Changchun: Changchun University of Science and Technology, 2011(in Chinese)
(尹佳. 纸介盲文自动识别系统关键技术研究[D]. 长春: 长春理工大学, 2011)

[11] Li R Q, Liu H, Wang X D, *et al*. Optical braille recognition based on semantic segmentation network with auxiliary learning strategy[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Los Alamitos: IEEE Computer Society Press, 2020: 2362-2368

[12] Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation[C] //Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention. Heidelberg: Springer, 2015: 234-241

[13] Lu L Q, Wu D, Xiong J F, *et al*. Anchor-free braille character detection based on edge feature in natural scene images[J]. Computational Intelligence and Neuroscience, 2022, 2022: Article No.7201775

[14] Liu Zeyang. Research on braille image recognition based on CRNN[D]. Lanzhou: Lanzhou University, 2021(in Chinese)
(刘泽阳. 基于 CRNN 的盲文图像识别研究[D]. 兰州: 兰州大学, 2021)

[15] Li Rongrui. Research on braille automatic recognition based on deep learning[D]. Kunming: Kunming University of Science and Technology, 2018(in Chinese)
(李荣瑞. 基于深度学习的盲文自动识别研究[D]. 昆明: 昆明理工大学, 2018)

[16] Lu Liqiong, Wu Dong, Wu Tao, *et al*. Mining text box location feature for anchor-free text detection in natural scene images[J]. Application Research of Computers, 2021, 38(8): 2556- 2560(in Chinese)
(卢利琼, 吴东, 吴涛, 等. 挖掘文本框位置特性的 Anchor-free 自然场景文本检测[J]. 计算机应用研究, 2021, 38(8): 2556-2560)

[17] Lu L Q, Wu D, Wu T, *et al*. Anchor-free multi-orientation text detection in natural scene images[J]. Applied Intelligence, 2020, 50(11): 3623-3637

[18] Guan T K, Gu C C, Tu J Z, *et al*. Self-supervised implicit glyph

attention for text recognition[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2023: 15285-15294

[19] Liu W, Anguelov D, Erhan D, *et al*. SSD: single shot multibox detector[C] //Proceedings of the 14th European Conference on Computer Vision. Heidelberg: Springer, 2016: 21-37

[20] Ren S Q, He K M, Girshick R, *et al*. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149

[21] Zhou X Y, Yao C, Wen H, *et al*. EAST: an efficient and accurate scene text detector[C] //Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2017: 2642-2651

[22] Yan Chengliang, Chen Guangzhu, Yi Jia, *et al*. Lightweight semantic segmentation of intelligent workshop scene objects combining multi-scale and attention mechanisms[J]. Journal of Computer-Aided Design & Computer Graphics, 2022, 34(10): 1626-1636(in Chinese)
(严成良, 陈光柱, 易佳, 等. 融合多尺度与注意力机制的智能车间场景目标轻量级语义分割[J]. 计算机辅助设计与图形学学报, 2022, 34(10): 1626-1636)

[23] Liang Haoran, Ye Lingchen, Liang Ronghua, *et al*. Text detection algorithm for natural scenes under attention supervision strategy[J]. Journal of Computer-Aided Design & Computer Graphics, 2022, 34(7): 1011-1019(in Chinese)
(梁浩然, 叶凌晨, 梁荣华, 等. 注意力监督策略下的自然场景文本检测算法[J]. 计算机辅助设计与图形学学报, 2022, 34(7): 1011-1019)

[24] Neubeck A, Van Gool L. Efficient non-maximum suppression[C] //Proceedings of the 18th International Conference on Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2006: 850-855

[25] Abadi M, Barham P, Chen J M, *et al*. TensorFlow: a system for large-scale machine learning[C] //Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation. New York: ACM Press, 2016: 265-283