doi:10.11887/j.cn.201703015 http://journal.nudt.edu.cn

用卷积神经网络分类最大稳定极值区域实现汉字区域定位*

张鹏伟,张伟伟
(信息工程大学 密码工程学院,河南 郑州 450001)

摘 要:获取对应笔画级连通区的最大稳定极值区域,实施形态学闭操作融合相距较近的最大稳定极值区域,融合后最大稳定极值区域对应的单个汉字区域;利用灰度共生矩阵描述最大稳定极值矩形区域的纹理信息,将其作为卷积神经网络的输入,卷积神经网络对最大稳定极值区域进行分类,过滤非汉字部分;利用最大稳定极值区域颜色直方图的Bhattacharyya距离等特征对最大稳定极值区域进行聚类,同一类最大稳定极值区域组合得到汉字文本候选区域;再次利用卷积神经网络对候选文本区域进行分类,过滤非文本部分,剩余的就是定位到的汉字文本区域。实验结果表明,该算法对于汉字区域定位具有良好的效果。

关键词:汉字区域定位;最大稳定极值区域;卷积神经网络;深度学习;灰度共生矩阵
中图分类号:TP391.4 **文献标志码**:A **文章编号**:1001-2486(2017)03-091-06

Scene Chinese text localization by convolutional neural network classifying maximum stable extremal regions

ZHANG Pengwei, ZHANG Weiwei
(School of Cryptography Engineering, Information Engineering University, Zhengzhou 450001, China)

Abstract: Firstly, the MSERs (maximum stable extremal regions) which corresponded to Chinese strokes was extracted. The morphological close operation was used to connect the nearby MSERs. The fused MSER corresponded to Chinese characters. Gray level co-occurrence matric was used to describe the textural characteristics of the fused MSER rectangle. They were the input of CNN (convolutional neural network). The MSER rectangles were classified by CNN in order to filter none Chinese character rectangle. Then, Chinese text candidates were constructed by clustering MSER rectangles based on the features such as the color histogram Bhattacharyya distance of MSER rectangles. CNN was reused to classify Chinese text candidates to filter none Chinese text clusters. Finally, the rectangle of the remaining clusters was the Chinese text regions of natural scene image. Experiment shows that the proposed algorithm is desirable in localizing the Chinese text in natural scene images.

Key words: Chinese text localization; maximum stable extremal region; convolutional neural network; deep learning; gray level co-occurrence matric

自然场景图像除包含丰富的色彩、形状、图案等物体视觉信息外,还可能包含大量的文本信息,比如书籍封面标题、单位名称、商店名称、道路路牌、交通指示牌、街道名称、建筑物门牌号、广告牌上的文字等。这些文本信息对于基于内容的图像检索、场景理解和智能交通等应用具有重要价值,从自然场景图像中自动提取文本信息已成为研究的热点[1]。

自然场景中文本的自动提取面临着许多困难:文本存在于自然场景图像的任意位置,且与背景往往混为一体;拍摄角度多种多样,字体纷繁复杂,透视形变严重。其中,确定文本在自然场景中的位置,即文本区域定位,是场景文本自动提取的前提和基础。文本区域定位的方法主要分为两类:一是基于滑动窗口的方法[2-4],采用文字区域分类器扫描整个场景图像,时间复杂度较高;二是基于连通区域的方法[5-9],认为文本是具有相近的颜色和亮度的连通区域,连通区域被作为文本的候选区域。此类方法中,最大极值区域(Maximally Stable Extremal Regions, MSER)被广泛采用[9]。

基于连通区域的文本定位方法一般对英文文本比较有效,这是因为除i、j外的大部分英文字符直接对应一个连通区域;对于中文文本,汉字区域通常由多个连通区域构成,其定位问题更加复杂。为此,刘晓佩、潘娜等采用小波变换捕捉汉字特性[10-11],孙巧榆采用视觉关注模型获取显著图掩膜确定中文区域[12],徐琼等提取候选区域的方向梯度直方图金字塔(Gabor Pyramid of Histogram of Orientation Gradients, PHOG-Gabor)特征并采用

* 收稿日期:2016-01-07
基金项目:国家863计划资助项目(20157011012)
作者简介:张鹏伟(1978—),男,山西偏关人,工程师,博士研究生,E-mail:zhang_pw@126.com

提升树算法确定文本区域[13]。

课题组之前基于汉字特点设计了初步的汉字文本区域定位算法[14]，本文在其基础上，以提取 MSER 区域为基础，利用卷积神经网络过滤非文本 MSER 区域为核心，给出了一种改进的汉字区域定位算法。

1 汉字特点分析与定位思路

假设自然场景中的汉字文本采用印刷体汉字。相对于手写体，印刷体汉字的形体结构清晰，如图 1 所示。汉字包括了笔画、部首等两个层次。笔画居于汉字结构最低层次，是指汉字书写时不间断地一次写成的一个线条，笔画区域是连通区；部首处于汉字结构的中间层次，部首包括若干笔画，笔画要么连通，即使不连通间距也较近；汉字是中文文本基本单元，除独体字外，每个汉字都由若干个不连通的部首构成，但与汉字间的间距相比，不连通的部首间的间距也相对较小。

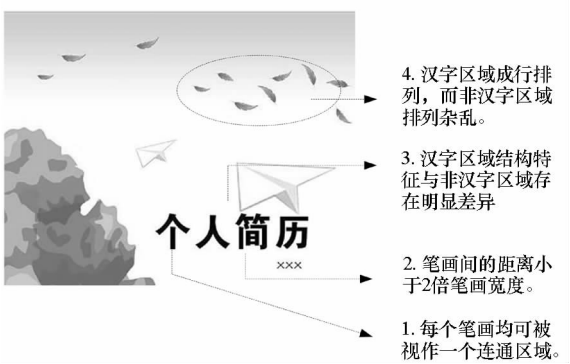


图 1 场景图像中汉字的特点

Fig. 1 Chinese characters in scene images

基于场景图像汉字中连通区域的距离关系，本文定位算法框架如图 2 所示。首先，提取 MSER 区域，它对应笔画级的连通区，通过形态学闭操作拓展连通区范围，使融合后的 MSER 区域尽可能对应单个汉字区域，将其作为汉字的候选区域；然后，用纹理信息作为 CNN 输入，CNN 分类融合后 MSER 区域，排除候选区域中的非汉字部分；最后，基于同一区域汉字具有共性的特点进行聚类，组合出汉字所在区域，利用 CNN 再次进行甄别，最终保留的区域就是最终定位到的自然场景汉字文本区域。

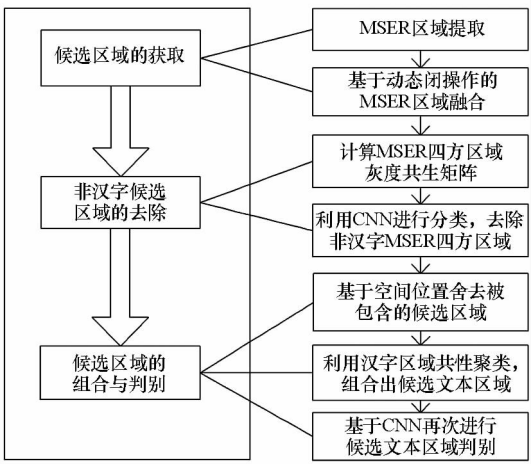


图 2 定位算法框架

Fig. 2 Chinese text localization algorithm flowchart

2 候选区域的获取

2.1 提取 MSER 区域

极值区域是内部像素点值要比外部像素点值低(或者高)的区域[15]。假设对灰度图像进行二值化，二值图像中黑色区域对应的像元集合就是极值区域。当二值化的灰度阈值从 0 依次变大到 255 时，极值区域的面积将逐渐扩大，类似于水面上升，旧的极值区域被包含到新的极值区域里面，当阈值为 255 时，整个图像成为极值区域。

MSER 区域是指在某个灰度阈值 i 的时候，区域像元数量变化最小的极值区域[15]。设 $Q_1,\cdots,Q_{i-1},Q_i,\cdots$ 为一系列由于灰度阈值升高而产生的相互包含极值区域，即 $Q_i \subset Q_{i+1}$。Δ 表示微小的灰度变化，当且仅当区域变化率 $Q(i)=|Q_{i+\Delta}-Q_{i-\Delta}|/|Q_i|$ 在 i 处取得局部极小值时，极值区域 Q_i 成为 MSER 区域。

MSER 区域是用不同灰度阈值对图像进行二值化时得到的最稳定的区域，区域内和区域外反差较大，因而一般具有明显的轮廓，此性质与自然图像中的汉字或汉字笔画区域比较吻合，因此 MSER 区域适合作为汉字初步的候选区域。

2.2 提取融合的 MSER 区域

大部分英文字符(除 i, j 外)都是“一笔画”的结构样式，对应一个完整的 MSER 区域，如 a, b, c, d 等。而汉字的组成单位是笔画，多个笔画先构成部首，部首再组合成汉字。部首中的笔画有相互连通的，如“扌”“攵”“亻”“勹”等，也有互不连通的，如“氵”“讠”“巛”等，此外，偏旁部首间一般互不连通。一个汉字往往不能对应单个 MSER 区域。

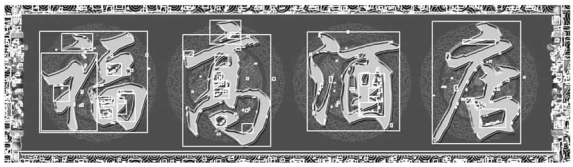
汉字一般具有四方性，这使得单个汉字中不连通的笔画间距离都很近，而汉字间的距离往往显著大于汉字内非连通的笔画间距。形态学中的闭操作能够消除狭窄的间断和细长的鸿沟，消除小的孔洞，并填补轮廓线中的断裂。因此，选择闭操作将同一汉字内的多个笔画 MSER

区域融合在一起。结构元及其参数影响融合的程度,结构元参数过小,导致融合不足,单个笔画 MSER 区域依然存在;参数过大,可能导致相邻汉字的笔画区域相互融合。基于汉字一般都是方块字的特点,采取方形作为结构元。对于图片汉字来说,笔画间的距离近似于笔画宽度,且小于相邻汉字的距离。用笔画宽度作为结构元参数,一般能够较好地进行同一汉字多个 MSER 区域的融合。

综上,为了进行汉字区域定位,首先提取图像中所有的 MSER 区域[15],然后进行闭操作进行 MSER 区域融合。融合后的 MSER 区域将对应单个汉字,但轮廓上具有不规则性,为了便于描述和分析汉字的四方特性,最后沿水平方向和垂直方向提取包络 MSER 融合区域的四方区域(下面简称为 MSER 四方区域,下面如不单独声明,MSER 四方区域均指融合后的 MSER 四方区域),用于下一步的汉字区域定位处理。图 3 给出了提取融合后 MSER 四方区域的一个例子,图 3(a)是含有汉字的场景图像,图 3(b)中各个矩形框是提取出融合前的 MSER 四方区域,图 3(c)中各个矩形框是提取出融合后的 MSER 四方区域,从"高"字可见,通过闭操作将"高"字头上的点融合进来了。

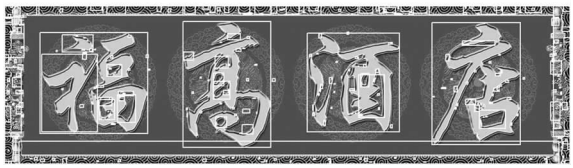
(a) 含有汉字的场景图像

(a) Example for scene images with Chinese character



(b) 闭操作融合前 MSER 四方区域(各个矩形框)

(b) Rectangle region of MSER before close operation



(c) 闭操作融合后 MSER 四方区域(各个矩形框)

(c) Rectangle region of MSER after close operation

图 3　融合后的 MSER 四方区域示例

Fig. 3　Examples for fused MSER rectangle

3　候选区域中非汉字部分的去除

3.1　计算 MSER 四方区域的灰度共生矩阵

上面提取出的 MSER 四方区域大多正好包含一个汉字,也可能包含多个汉字;但仍可能仅包含汉字的一个部分(如图 3(c)中"高"字中的小矩形框),或根本不对应是汉字区域(如图 3(c)中的上下边框部分)。包含单个汉字和多个汉字的 MSER 四方区域在纹理上具有一定相似性,可将其视为一类,记为 H_0 类。该类与对应非文本区域和汉字局部的 MSER 四方区域(这两类 MSER 四方区域记为 H_1 类)在纹理上应有较大差异,可利用纹理特征对 H_0 类和 H_1 类 MSER 四方区域进行分类,目的是滤掉 H_1 类 MSER 四方区域。

灰度共生矩阵反映不同像素相对位置的空间信息,在一定程度上反映了纹理图像中各灰度级在空间上的分布特性,是纹理分析中最经常采用的特征之一。设 I 为一幅灰度图像,其大小为 $M\times N$,灰度共生矩阵 $\boldsymbol{P}$ 定义为:

$$\boldsymbol{P}(i,j)=\#\{[(x,y),(x+a,y+b)]\mid I(x,y)=i,\ I(x+a,y+b)=j\}/(M\cdot N) \tag{1}$$

其中,#表示取集合元素数量;$i,j\in\{0,255\}$;a,b 指示了像素位置的差异。灰度共生矩阵统计了两个像素点位置的联合概率分布,是图像灰度变化的二阶统计度量,也是描述纹理结构性质特征的基本函数。

对提取出的每个 MSER 四方区域同样可以计算灰度共生矩阵 $\boldsymbol{P}$,并且无论 MSER 四方区域是长方形的(区域中包含多个汉字)还是类正方形(区域中对应单个汉字),均可以得到相同维数(256×256 维)的灰度共生矩阵 $\boldsymbol{P}$,便于分类器的处理。因此,下面将 $\boldsymbol{P}$ 作为分类器的输入,利用分类器的输出判别 MSER 四方区域属于 H_0 类还是 H_1 类。

3.2　用卷积神经网络分类 MSER 四方区域

当获得了每个待分类的 MSER 四方区域灰度共生矩阵 $\boldsymbol{P}$ 后,$i,j\in\{0,255\}$,$\boldsymbol{P}$ 的维数为 256×256,对一般的分类器而言,往往需要进一步提取灰度共生矩阵的高阶统计量以减少维数,但这往往导致了信息的丢失。最近,CNN 在图像分类上取得了巨大成功[16],它可以直接处理高维的图像矩阵数据,而不需要先进行特征提取。因此,构建输入为 256×256 维数据的 CNN,用于对 MSER 四方区域的灰度共生矩阵进行是 H_0 类和 H_1 类的判别。

参考 AlexNet[16] 设计 CNN,但为了提高训练

速度,仅采用了 6 层结构,如图 4 所示。第一层为卷积层,卷积核尺寸为 25 × 25,卷积步长为 1,即用 256 × 256 灰度共生矩阵卷积 25 × 25 的卷积核,取卷积结果的 valid 部分并进行非线性运算 ReLU,得到(256 − 25 + 1) × (256 − 25 + 1)维的特征 map,共 100 种卷积核,将得到 100 个 232 × 232 维特征 map,其中的每一个值对应一个神经元;第二层为 Max 池化层,其主要作用是在保证纹理细节被保留的前提下进行降维,为了降低运算量,取池化窗口大小为 8 × 8,步长为 8,池化处理后得到 100 个 29 × 29 维的特征 map;第三层为卷积层,卷积核尺寸为 100 × 6 × 6,共 48 种卷积核,卷积后得到 48 个 24 × 24 维的特征 map;第四层为 Max 池化层,池化窗口大小为 2 × 2,步长为 2,池化后将降维为 48 个 12 × 12 的特征 map;第五层为全连接层,共有 1000 个神经元,第六层为输出层,共有两种输出,分别对应 H_0 类和 H_1 类。

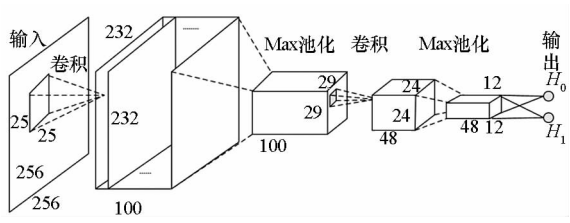


图 4 CNN 结构

Fig. 4 CNN architecture

CNN 神经元参数的有效训练是取得良好分类性能的关键,训练一方面在于训练算法,另一个方面要有大量有效的两类样本数据。训练算法方面,采用随机梯度下降(stochastic gradient descent)迭代算法[16]更新神经元参数,损失函数采用交叉熵代价函数;训练样本数据方面,提取大量 H_0 类和 H_1 类的 MSER 区域,分别计算其灰度共生矩阵作为两类样本。

4 候选区域的组合与文本区的确定

4.1 基于空间位置对候选区域取舍

经过 CNN 分类后,得到许多候选区域(即判定为 H_0 类的 MSER 区域),这其中的一些区域仍然可能存在包含关系,如图 3(c)中“福”字对应的四方候选区域就包含了偏旁对应的小候选区域,需要舍去小区域,保留大区域,大区域更可能对应整个汉字。删减步骤如下:

步骤 1:按候选区域左上角的坐标位置,从左到右、从上到下的顺序对所有区域排序,形成候选区域队列。

步骤 2:取出队头两个候选区域,若二者存在完全包含关系,即队头区域的右下角坐标位置在第二个区域右下角的右下方,则保留队头区域,第二个区域从候选区域队列中去除(即合并到大队头区域中),队列长度减 1,然后重复步骤 2。

步骤 3:若二者不存在完全包含关系,保持队列长度不变,仅将队头指向原队头下面的第二个区域,然后判断从新队头开始的队列中是否还有两个区域(含队头区域),若是(即仍可以比较)则转到步骤 2,否则退出,说明已经取舍完毕。

4.2 对候选区域进行聚类

取舍完成后的候选区域将不会存在包含关系,但仍有可能存在交叠关系,或距离很近。场景图像中的文本通常有多个汉字(字符),这些距离很近的候选区域往往正对应了同一文本区的多个汉字或汉字组成部分,需要将这些区域进行组合,进而得到完整汉字文本区域。

场景图像中属于同一文本区的汉字一般色彩纹理统一,即具有相近属性且距离较近,因此可依据色彩相近程度以及距离关系对候选区域进行聚类,聚类后属于同一类的候选区域将进行组合。

聚类算法采用 Yin 的 single-link 聚类算法[9]。在 single-link 算法中,需要计算两两候选区域的特征距离或特征差,聚类性能很大程度上依赖于选择的特征上,也就是对候选区域的描述方法上。本文沿用四种特征距离:空间距离、宽高差、顶部和底部对齐程度、笔画差[9]。在此基础上采用前续工作[14]中的颜色直方图特征矢量,增加 Bhattacharyya 距离[17]作为区域色彩相近程度的度量而不是简单的欧式距离度量。颜色直方图估算了各种色彩出现的概率分布,Bhattacharyya 距离是一种衡量两个概率分布的相似程度的距离度量,两者结合起来能更为准确地刻画候选区域汉字颜色是否相近这一特点,因而有助于提高聚类性能。

颜色直方图计算方法如下:首先把候选区域的红绿蓝(Red Green Blue, RGB)色彩空间数据转换到色度/饱和度/亮度(Hue Saturation Value, HSV)空间,然后按照人的视觉分辨能力,把色调 H 空间分成 7 份,饱和度 S 和亮度 V 空间分成 2 份,得到:

$$H_i=\begin{cases}0 & h\in(330,360]\cup[0,22],i=1\\ 1 & h\in(22,45],i=2\\ 2 & h\in(45,70],i=3\\ 3 & h\in(70,155],i=4\\ 4 & h\in(155,186],i=5\\ 5 & h\in(186,278],i=6\\ 6 & h\in(278,330],i=7\end{cases}\tag{2}$$

$$S_j=\begin{cases}0 & s\in[0,0.65],j=1\\ 1 & s\in(0.65,1],j=2\end{cases} \tag{3}$$

$$V_k=\begin{cases}0 & v\in[0,0.7],k=1\\ 1 & v\in(0.7,1],k=2\end{cases} \tag{4}$$

统计(H_i,S_j,I_k)中各个值在候选区域 $area_q$ 出现的相对次数,即:

$$P_q(i,j,k)=\#\{\delta_{m,n}\,|\,h(m,n)=i,s(m,n)=j,s(m,n)=k,\delta_{m,n}\in area_q\}/\#area_q \tag{5}$$

其中,$\forall\delta_{m,n}\in area_q$,$area_q$ 为第 q 个候选区域;$\delta_{m,n}$ 为候选区域中坐标位置为(m,n)的像素点;$\#area_q$ 表示取候选区域中像素的总个数;$(i,j,k)\in(H_i,S_j,I_k)$。$P_q(i,j,k)$就是对候选区域 $area_q$ 提取的颜色直方图。

对两个候选区域 $area_p$ 和 $area_q$ 得到的颜色直方图 $P_p(i,j,k)$、$P_q(i,j,k)$,其 Bhattacharyya 距离的计算公式为:

$$J_B(P_p,P_q)=-\ln\sum_{i,j,k}[P_p(i,j,k)\cdot P_q(i,j,k)]^{1/2} \tag{6}$$

由此得到新的颜色直方图特征距离。

4.3　确定汉字区域

对聚类算法得到的属于同一类的候选区域,提取包含所有区域的大四方区域轮廓,形成汉字文本候选区域。

对汉字文本候选区域再次提取灰度共生矩阵,并再次用前述的 CNN 进行分类,判断为 H_0 类还是 H_1 类。前面 CNN 分类用到的灰度共生矩阵是由较小面积的区域计算出的,面积小意味着计算灰度共生矩阵的元素较少,反映了汉字纹理不一定充分,而大区域则相对较好。因此,用汉字文本候选区域重新提取灰度共生矩阵,通过 CNN 再次分类来排除被错误认为是汉字文本区域的非文本区域。最终,剩余的汉字文本候选区域就是定位到汉字文本区域。

5　实验与分析

使用本文定位方法和其他中文文本定位算法对自然场景图像中的汉字区域进行定位。图 5 给出了本文算法在一些场景图像中的定位效果图。图 5 中的汉字区域均被准确地定位出来(用黑色矩形框标出),这主要得益于深度神经网络的分类能力,以及 Bhattacharyya 距离对同一区域中汉字颜色的准确度量上。

实验中对比分析了本文定位算法和其他中文文本定位算法的性能。本文算法主要面向中文文本,目前没有统一的关于自然场景中文文本分析的标准数据库。因此,利用照相机和网络获得 2000 幅包含各种不同类型的中文文本的自然场景图像,包括广告牌、图书封面、路牌、商标等,组建中文图像文本数据库(下文称作自建数据库),人工标注自建数据库中的每幅图像每个中文文本区域的矩形框坐标。

图 5　汉字区域定位结果示意图

Fig. 5　Examples for Chinese text localization

算法性能由准确率(Precision, P)和召回率(Recall, R)两个指标反映。将定位得到矩形框 e 与标注矩形框 t 的交叉面积,除以最小包含 e 和 t 矩形框(bounding boxes)的面积,这个商记为 $m(e,t)$,$0\leqslant m(e,t)\leqslant 1$。将定位算法得到的汉字区域矩形框集记为 E,标注矩形框集为 T。

根据下列公式确定算法的 P 和 R:

$$P=\frac{\sum_{e\in E}\max\{m(e,t)\,|\,t\in T\}}{|E|}$$

$$R=\frac{\sum_{t\in T}\max\{m(e,t)\,|\,e\in E\}}{|T|}$$

其中,$|E|$和$|T|$代表矩形框集中元素的个数。

表 1 给出了本文算法和潘娜算法[11]、徐琼算法[13]在自建中文文本数据库上的性能对比情况。从表中可以看出,本文算法性能优于二者,这主要因为本文算法是基于汉字的特点提出的,基于闭操作的部首融合和灰度共生矩阵 CNN 分类对于汉字区域定位是有效的。

表 1　自建数据集上的算法性能对比

Tab. 1　Performance comparison between our algorithm with others on Chinese text scene image dataset

算法	P	R
潘娜算法[11]	0.76	0.72
徐琼算法[13]	0.75	0.77
本文算法	0.826	0.788

实验还进行了西文字符的定位实验。数据库采用文档分析与识别国际会议(International Conference on Document Analysis and Recognition,

ICDAR) 2011[18]。实验结果如表 2 所示,本文算法相对于课题组先前的算法[14]有较大提高,这也是得益于 CNN 的引入;本文算法性能优于 Neumann 算法[5],接近(但低于)Yin 算法[9],其原因是基于闭操作的部首融合对于汉字是有效的,但对于西文的作用有限。

表 2 ICDAR 2011 数据集上的算法性能对比

Tab. 2 Performance comparison between our algorithm with others on ICDAR 2011 dataset

算法	*P*	*R*
张伟伟算法[14]	0.65	0.52
Neumann 算法[5]	0.73	0.64
Yin 算法[9]	0.86	0.68
本文算法	0.80	0.64

6 结论

本文利用场景图像中汉字区域的连通特点和纹理特性,以 CNN 分类过滤 MSER 区域为核心,提出了汉字区域定位算法。该算法的关键前导步骤是提取 MSER 区域,其对自然场景图像中纹理清晰的汉字文本区域具有较高的准确性,但一些成像质量较差、存在模糊的文本区域,存在被 MSER 区域提取漏检的情况。因此,如何减少漏检,提高定位算法召回率将是下一步的工作重点。此外,自动判断图像类型,根据类型特点选择不同定位策略,也将有助于定位性能的提高。

参考文献(References)

[1] Karatzas D, Shafalt F, Uchida S, et al. ICDAR 2013 robust reading competition[C]//Proceedings of 12th International Conference on Document Analysis and Recognition, 2013: 1115 – 1124.

[2] Chen X R, Yuile A L. Detecting and reading text in natural scenes[C]//Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004: 366 – 373.

[3] Pan Y F, Hou X W, Liu C L. Text localization in natural scene images based on conditional random field[C]// Proceedings of the 12th International Conference on Document Analysis and Recognition, 2009: 6 – 10.

[4] Lee J J, Lee P H, Lee S W, et al. AdaBoost for text detection in natural scene[C]//Proceedings of the International Conference on Document Analysis and Recognition, 2011: 429 – 434.

[5] Neumann L, Matas J. Real-time scene text localization and recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2012: 3538 – 3545.

[6] Neumann L, Matas J. Text localization in real-world images using efficiently pruned exhaustive search[C]//Proceedings of the International Conference on Document Analysis and Recognition, 2011: 687 – 691.

[7] Neumann L, Matas J. A method for text localization and recognition in real-world images[C]//Proceedings of the 10th Asian Conference on Computer Vision, 2010: 770 – 783.

[8] Gracia C M, Lenc K, Mimehdi M. A head-mounted device for recognizing text in natural scenes[C]//Proceedings of the 4th International Conference on Camera-based Document Analysis and Recognition, 2011: 29 – 41.

[9] Yin X C, Yin X W, Huang K Z, et al. Robust text detection in natural scene images[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014, 36(5): 970 – 983.

[10] 刘晓佩, 卢朝阳, 李静. 结合 WTLBP 特征和 SVM 的复杂场景文本定位方法[J]. 西安电子科技大学学报:自然科学版, 2012, 39(4): 103 – 108.
LIU Xiaopei, LU Zhaoyang, LI Jing. Complex scene text location method based on WTLBP and SVM[J]. Journal of Xidian University: Natural Science, 2012, 39(4): 103 – 108. (in Chinese)

[11] 潘娜. 图像中的文本定位算法研究[D]. 南京: 南京理工大学, 2013.
PAN Na. Research on text detection in images [D]. Nanjing: Nanjing University of Science and Technology, 2013. (in Chinese)

[12] 孙巧榆. 复杂背景图像的文本信息提取研究[D]. 上海: 华东师范大学, 2012.
SUN Qiaoyu. Research of the text information extraction in images with complicated background[D]. Shanghai: East China Normal University, 2012. (in Chinese)

[13] 徐琼, 干宗良, 刘峰, 等. 基于提升树的自然场景中文文本定位算法研究[J]. 南京邮电大学学报:自然科学版, 2013, 33(6): 76 – 82.
XU Qiong, GAN Zongliang, LIU Feng, et al. Chinese text localization method based on boosting tree in natural images[J]. Journal of Nanjing University of Posts and Telecommunications: Natural Science, 2013, 33(6): 76 – 82. (in Chinese)

[14] 张伟伟, 汤光明, 孙怡峰, 等. 一种针对汉字特点的场景图像中文文本定位算法[J]. 信息工程大学学报, 2015, 15(6): 729 – 736.
ZHANG Weiwei, TANG Guangming, SUN Yifeng, et al. Chinese scene text localization algorithm based on the feature of characters [J]. Journal of Information Engineering University, 2015, 15(6): 729 – 736. (in Chinese)

[15] Matas J, Chum O, Urban M, et al. Robust wide baseline stereo from maximally stable extremal regions[J]. Image & Vision Computing, 2004, 22(10): 761 – 767.

[16] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[C]// Proceedings of the 25th International Conference on Neural Information Processing Systems, 2012: 1097 – 1105.

[17] Bhattacharyya A. On a measure of divergence between two statistical populations defined by their probability distributions[J]. Bulletin of the Calcutta Mathematical Society, 1943, 35: 99 – 109.

[18] Shahab A, Shafait F, Dengel A. ICDAR 2011 robust reading competition challenge 2: reading text in scene images[C]// Proceedings of International Conference on Document Analysis and Recognition, 2011: 1491 – 1496.