

# 基于置信度融合的自然场景文本检测方法

蒋志鹏<sup>1</sup>, 潘坤榕<sup>1</sup>, 张国林<sup>1</sup>, 刘玉琪<sup>1</sup>, 张 瑛<sup>1</sup>, 孙科学<sup>1,2\*</sup>

(1. 南京邮电大学 电子与光学工程学院, 江苏 南京 210023;

2. 射频集成与微组装技术国家地方联合工程实验室, 江苏 南京 210023)

**摘 要:**自然场景文本检测技术已经成为计算机视觉领域重要的研究任务,在图像检索、辅助驾驶、工业检测等领域具有广泛应用。在现有的基于深度学习的自然场景文本检测方法中,非极大抑制算法在对同一个真实文本框的重复检测进行合并和筛选时,将预测框的分类置信度作为排序依据,导致那些定位更精确而分类置信度略低的预测框被抑制,从而影响检测准确率。为了提高预测框的定位精确度,文中提出基于置信度融合的自然场景文本检测方法。首先,设计了交并比网络,作为每个预测框的定位置信度;其次,在非极大抑制算法中,将定位置信度与文本分类置信度融合作为预测框排序的依据;最后,在ICDAR2011和ICDAR2013数据集上对该方法进行了实验,结果表明,该方法检测的文本框更加紧致,包含的背景区域更少,可以提高文本检测的准确率。

**关键词:**自然场景;文本检测;卷积神经网络;非极大抑制;置信度融合

中图分类号:TP391.4

文献标识码:A

文章编号:1673-629X(2021)08-0039-06

doi:10.3969/j.issn.1673-629X.2021.08.007

## Research on Scene Text Detection Based on Confidence Fusion

JIANG Zhi-peng<sup>1</sup>, PAN Kun-rong<sup>1</sup>, ZHANG Guo-lin<sup>1</sup>, LIU Yu-qi<sup>1</sup>, ZHANG Ying<sup>1</sup>, SUN Ke-xue<sup>1,2\*</sup>

(1. School of Electronic and Optical Engineering, Nanjing University of Posts and Telecommunications,  
Nanjing 210023, China;

2. Nation-Local Joint Project Engineering Lab of RF Integration & Micropackage, Nanjing 210023, China)

**Abstract:** Natural scene text detection technology has become an important research task in the field of computer vision, and has a wide range of applications in image retrieval, driver assistance, industrial detection and other fields. In the existing natural scene text detection methods based on deep learning, when the non-maximum suppression algorithm merges and filters the repeated detection of the same real text box, it uses the classification confidence of the prediction box as the sorting basis, resulting in the suppression of those prediction boxes with more accurate positioning and slightly lower classification confidence, which affects the detection accuracy. In order to improve the positioning accuracy of the prediction frame, we propose a text detection method based on confidence fusion. Firstly, the intersection and comparison network is designed as the positional reliability of each prediction box; Secondly, in the non-maximum suppression algorithm, the fusion of the positional reliability and the text classification confidence is used as the basis for ranking the prediction boxes. Finally, experiments on the ICDAR2011 and ICDAR2013 data sets show that the text box detected by this method is more compact and contains less background area, which can improve the accuracy of text detection.

**Key words:** natural scene; text detection; convolution neural network; non-maximal suppression; confidence fusion

## 0 引 言

在自然场景图像中包含大量文本,这些文本信息可以作为图像信息的说明和补充,因此从自然场景图像中定位文字区域并识别文本语义已经成为计算机视觉和文档分析领域重要的研究任务<sup>[1]</sup>;该任务在图像检索<sup>[2]</sup>、图像中敏感词检测、盲人导航<sup>[3]</sup>、辅助驾驶<sup>[4]</sup>

等领域具有广泛的应用。背景单一、颜色纹理统一的文本检测技术已经十分成熟,并且已有广泛的应用,例如身份证、发票单据等各种稿件中的文本检测与识别,但是由于自然场景背景复杂、光照不均匀、模糊遮挡等不同因素,都影响了文本检测的定位精度和召回率,给文本检测技术带来了新的挑战和难点<sup>[5-8]</sup>。

收稿日期:2020-10-09

修回日期:2021-02-22

基金项目:江苏省研究生实践创新计划(SJCX19\_0259);江苏省大学生创新训练计划(SYB2020012);南京邮电大学国自孵化项目(NY220013)

作者简介:蒋志鹏(1998-),男,硕士研究生,研究方向为智能信号处理、机器视觉与深度学习技术;孙科学,博士,教授,硕导,研究方向为电子电路设计、智能信号处理。

随着计算机硬件计算能力的提升和深度卷积神经网络在计算机视觉领域的应用,深度学习技术越来越多地应用在一般目标检测任务(SSD<sup>[9]</sup>、YOLO<sup>[10]</sup>、Faster-RCNN<sup>[11]</sup>)中,促进了自然场景文本检测任务性能的提升和应用范围的扩大。深度卷积神经网络<sup>[12]</sup>(convolutional neural network, CNN)中的卷积和池化运算对图像的平移、旋转和缩放具有较强的鲁棒性,其层层堆叠的结构能够将一些低层次的图像特征重组成一些高层次的语义特征,许多研究者将卷积神经网络应用到自然场景文本检测任务中进行特征提取。例如 2014 年, Girshick 等人<sup>[13]</sup>提出了 R-CNN 算法,将深度学习技术应用到一般目标检测技术中,随后以 R-CNN 为基础的 Fast-RCNN<sup>[14]</sup>和 Faster-RCNN<sup>[11]</sup>算法相继问世。2015 年, Jonathan 等人<sup>[15]</sup>首次提出了全卷积网络(fully convolutional networks, FCN),该网络不包含全连接层,能够实现逐像素级别的预测和分类,对细小目标的位置信息感知能力更强,并且可以接受任意尺寸的图像输入。

基于卷积神经网络的自然场景文本检测技术主要包括特征提取网络、预测网络和非极大抑制算法。在传统的自然场景文本检测方法<sup>[16-17]</sup>中,非极大抑制算法基于预测文本框的分类置信度对重复检测的预测框进行筛选和合并。然而,该过程忽略了预测框的定位精度,使得一些定位更精确而分类置信度略低的预测框可能在非极大抑制步骤中被抑制,影响文本检测的

准确率。

为了改善上述不足,文中设计了置信度融合的文本检测方法。在多任务预测网络中设计一个新的分支预测锚框与真实文本框的交并比 IOU(intersection over union)值,将该值作为预测文本框的定位置信度。在非极大抑制算法中,用分类置信度与定位置信度融合的结果取代分类置信度,保留定位更精确的预测文本框,提高文本检测的准确率。

## 1 基于置信度融合的文本检测方法

### 1.1 方法总体设计

置信度融合的文本检测方法包括特征提取网络、多任务预测网络和非极大抑制算法等三个部分,网络结构如图 1 所示。其中特征提取网络的作用是从输入图像中提取多尺度的特征图;多任务预测网络的作用是对特征图上每个预定义的锚框的文本信息进行预测;非极大抑制算法的作用是对同一文本区域重复预测的预测框进行合并和筛选。研究者通常先将分类置信度低于阈值的预测框删除,再按照分类置信度对剩下的预测框进行排序,保留分类置信度最大的预测框,剩下的预测框则会被抑制。在以上过程中,那些定位更加精确而分类置信度略低的预测框可能会被抑制。因此文中将分类置信度和定位置信度进行融合以改进非极大抑制算法。

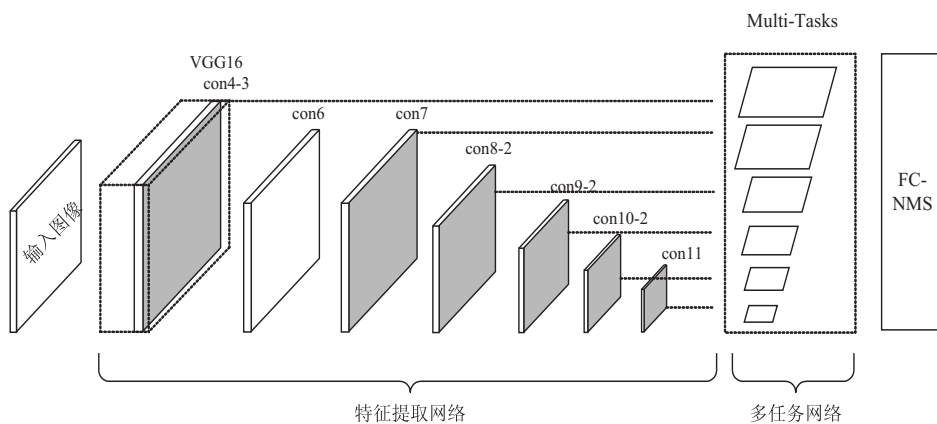


图 1 置信度融合的文本检测模型网络结构

### 1.2 基于 VGG 的特征提取网络

VGGNet 基础网络的泛化能力强、简洁实用,后续成为检测和识别任务中的主干网络。文中选择 VGGNet 网络并对其进行改进,作为特征提取网络的主干网络。

VGG-16 一共包括 13 个卷积层和 2 个全连接层,它的网络结构参数列于表 1。

特征提取网络保留 VGG-16 的 conv1 到 conv4 层,将最后的两个全连接网络改成  $3 \times 3$  的卷积层,为

conv5,并在此基础上增加 conv6 到 conv11,如图 1 灰色区域所示。其中 conv4\_3、conv7、conv8\_2、conv9\_2、conv10\_2、conv11 是文中在特征提取网络中抽取的多尺度特征图。

不同尺度的特征图具有不同的感受野,提取不同层次的特征,通常越浅层的特征图感受野越小,通常可以提取一些边缘、局部特征,能够检测面积较小的文本区域,而越深层的特征图感受野越大,通常可以提取图像的一些语义特征,可以检测面积较大的文本区域。

表 1 VGG-16 的网络结构

网络层次	操作	数量	卷积核/池化窗口	步长	填充	通道数
conv1-1	卷积+ReLU	2	3 * 3	1	1	64
	池化	1	2 * 2	2	--	64
conv2-2	卷积+ReLU	2	3 * 3	1	1	128
	池化	1	2 * 2	2	--	128
conv3-3	卷积+ReLU	3	3 * 3	1	1	256
	池化	1	2 * 2	2	--	256
conv4-3	卷积+ReLU	3	3 * 3	1	1	512
	池化	1	2 * 2	2	--	512
conv5-3	卷积+ReLU	3	3 * 3	1	1	512
	池化	1	2 * 2	2	--	512
全连接层 x2	全连接层+sigmoid	2	--	--	--	--

### 1.3 多任务预测网络

#### (1) 锚框设置。

多尺度特征图从特征提取网络输出后,文中会在特征图上密集采样锚框,设特征图的大小为  $N * N$ ,将特征图的每个像素点  $(i, j)$  视为不同横纵比的锚框的中心坐标,那么像素点  $(i, j)$  处将会产生 5 种横纵比  $a_r$  的锚框,如式(1):

$$a_r = [1, 3, 5, 7, 10] \quad (1)$$

则每个  $N * N$  的特征图中会生成  $N * N * 5$  个锚框。

由于不同输出层的特征图尺度不一样,每层的感受野大小也不同,因此每层特征图对应的锚框的面积也不一样,特征图越浅,感受野越小,锚框的面积也就越小。文中设计的特征提取网络一共输出 6 层特征图,将图 1 中从左往右的特征图依次记为特征图 1 到 6,那么第  $k$  层特征图中的锚框面积大小如式(2):

$$S_k = S_{\min} + \frac{S_{\max} - S_{\min}}{6 - 1}(k - 1), k \in [1, 6] \quad (2)$$

式中,  $S_{\min}$  表示最小锚框面积,即第一层特征图上的锚框面积;  $S_{\max}$  表示最大锚框面积,即第六层特征图上的

锚框面积;  $k$  表示特征图的层数。

每个锚框的宽和高的计算方式如式(3)和式(4):

$$w_k^a = S_k \sqrt{a_r} \quad (3)$$

$$h_k^a = S_k / \sqrt{a_r} \quad (4)$$

式中,  $w_k^a$  和  $h_k^a$  表示第  $k$  层特征图上长宽比为  $a$  的锚框的宽和高,  $S_k$  表示第  $k$  层特征图上的锚框面积,  $a_r$  表示锚框的长宽比。

#### (2) 文本框的预测和坐标计算。

接下来,多任务预测网络根据设计好的锚框预测特征图上每一个锚框的类别置信度  $\text{score}_{\text{cls}}$ 、定位置信度  $\text{score}_{\text{iou}}$  和每个锚框的坐标偏移量  $\text{offset}_{\text{loc}}$ ,分别对应如图 2 中的“预测类别”分支、“预测交并比分支”和“坐标信息”分支。

图 2 中,“交并比预测”分支和“预测类别”分支分别采用两个卷积层和两个 sigmoid 激活函数,“坐标信息”分支采用两个卷积层和 ReLU 激活函数。卷积核采用  $3 * 5$  而非  $3 * 3$  的尺寸,这种卷积核的尺寸是针对文本狭长的矩形特征设计的,这样可以产生狭长的矩形感受野,有利于处理更大横纵比的文本。

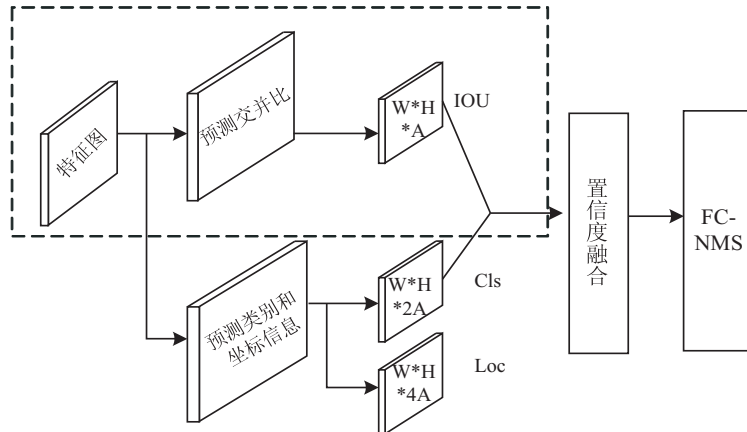


图 2 多任务预测网络

设第  $k$  层特征图的  $(i, j)$  位置处有一锚框  $b_0 = (x_0, y_0, w_0, h_0)$ , 多任务预测网络在  $(i, j)$  处输出  $(\Delta x, \Delta y, \Delta w, \Delta h, \text{score}_{\text{cls}}, \text{score}_{\text{iou}})$ ,  $\text{score}_{\text{cls}}$ 、 $\text{score}_{\text{iou}}$  为该预测文本框的分类置信度和定位置置信度。假设该预测框的  $\text{score}_{\text{cls}}$  满足阈值, 被认为是一个文本框, 那么该预测文本框的中心坐标和宽高为  $b = (x, y, w, h)$ , 计算方式如式(5):

$$\begin{cases} x = x_0 + w_0 \Delta x \\ y = y_0 + h_0 \Delta y \\ w = w_0 + \exp \Delta w \\ h = h_0 + \exp \Delta h \end{cases} \quad (5)$$

式中,  $x_0, y_0, w_0, h_0$  为锚框的中心坐标、宽和高;  $\Delta x, \Delta y, \Delta w, \Delta h$  为锚框与预测文本框之间的坐标偏移量。

#### 1.4 改进的置信度融合的非极大抑制算法

在一般非极大抑制算法(NMS)中, 当一个真实文本框存在重复检测时, 分类置信度最大的那个文本框将会被保留。然而, 由于分类置信度和定位置置信度的不匹配, 定位更准确而分类置信度偏低的文本候选框可能在 NMS 算法中被抑制, 从而影响文本检测性能。本小节在非极大抑制算法中, 用融合的分类置信度与定位置置信度取代传统的分类置信度, 改进后的 NMS 算法称为置信度融合的非极大抑制算法(FC-NMS)。

根据文献[18]中的分析, 候选文本框的 IOU 值与定位置置信度高度相关, 而与分类置信度相关性较小。考虑到传统 NMS 方法中分类置信度的作用, 与文献[18]中直接用定位置置信度取代分类置信度作为 NMS 中候选框排序的依据不同, 本小节将分类置信度  $\text{score}_{\text{cls}}$  与定位置置信度  $\text{score}_{\text{iou}}$  用不同的权重值进行融合, 得到一个融合置信度  $\text{score}_{\text{FC}}$ 。将融合置信度作为 NMS 步骤中文本框排序的依据,  $\text{score}_{\text{FC}}$  的计算方式如式(6):

$$\text{score}_{\text{FC}} = W_{\text{cls}} \times \text{score}_{\text{cls}} + W_{\text{iou}} \times \text{score}_{\text{iou}} \quad (6)$$

式中,  $W_{\text{cls}} = 0.2$ 、 $W_{\text{iou}} = 0.8$  分别表示分类置信度和定位置置信度的权重。

与传统 NMS 算法类似, 在候选框集合中, 将融合置信度  $\text{score}_{\text{FC}}$  最高的文本框记为  $A$ , 计算剩下的候选框与  $A$  的交并比 IOU 值, 计算公式如式(7):

$$\text{IOU}(A, B) = \frac{A \cap B}{A \cup B} \quad (7)$$

式中,  $A$  和  $B$  表示两个候选文本框,  $\text{IOU}(A, B)$  表示框  $A$  与框  $B$  的交集面积与并集面积之比, IOU 越大, 表示  $A$  与  $B$  重叠率越高。 $A$  与  $B$  的交集部分如图 3 所示。

图 3 中, 框  $A$  与框  $B$  的交集部分是一个矩形。若框  $B$  与框  $A$  的 IOU 值大于阈值  $Q_{\text{nms}}$ , 表明框  $B$  与框  $A$  的重叠程度较高, 将框  $B$  从候选框集合中删除, 同时更新  $A$  的分类置信度。比如要删除框  $C$ , 则框  $A$  的分

置信度重置为  $\text{score}_{\text{cls}_A}$ ,  $\text{score}_{\text{cls}_A}$  的计算公式如式(8):

$$\text{score}_{\text{cls}_A} = \max(\text{score}_{\text{cls}_A}, \text{score}_{\text{cls}_C}) \quad (8)$$

式中,  $\text{score}_{\text{cls}_A}$  表示框  $A$  的文本类别置信度,  $\max$  表示求最大值,  $\text{score}_{\text{cls}_C}$  表示框  $C$  的文本类别置信度。

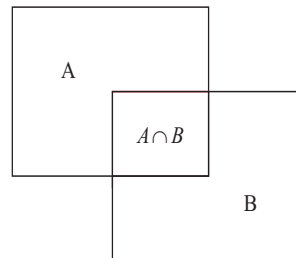


图 3 矩形框  $A$  和  $B$  的交并比示意图

置信度融合的非极大抑制算法(FC-NMS)的伪代码如算法 1 所示。

算法 1: FC-NMS。

输入:  $\text{Box} = \{b_1, b_2, \dots, b_n\}$ ,  $\text{cls}$ ,  $\text{iou}$ ,  $Q_{\text{nms}}$

$\text{Box}$  表示候选框的集合,  $b_i$  表示第  $i$  个候选框

$\text{cls}/\text{iou}/\text{FC}$ : 映射每个候选框的分类置信度, 定位置置信度和融合置信度的函数

$Q_{\text{nms}}$ : FC-NMS 的筛选阈值

输出:  $\text{Result\_Box}$ : 最终的预测文本框

```

1: Result_Box = None
2: while Box != None:
3:     box = argmax(FC)
4:     c = cls(box)
5:     delete box from Box
6:     for b_j in Box:
7:         if IOU(box, b_j) > Q_nms:
8:             c = max(c, cls(b_j))
9:             delete b_j from Box
10:        end if
11:    end for
12:    Result_Box = Result_Box ∪ {[box, c]}
13: end while
14: return Result_Box

```

## 2 文本检测器的训练

本章通过对交并比预测分支单独训练增强交并比网络的兼容性; 通过旋转、平移、缩放等手段手动变换训练集中所有的真实文本框, 从而生成候选文本框集。将该候选框集合中与真实文本框交并比小于 0.5 的候选框去除。然后从该候选集合中抽取训练数据对交并比网络进行训练。这种凭借经验增广的数据集为交并比网络带来了更好的性能和鲁棒性。

对于置信度融合的文本检测模型的初始化, 文中用预训练的 VGG-16 模型的权重参数初始化 VGG-16 部分, 用预训练的 TextBoxes 模型初始化卷积 6 ~ 11 层的权重参数。从第 11 层卷积开始往后的多任务

网络中所有的参数都用均值为 0, 标准差为 0.01 的高斯分布进行初始化。

定位置信度  $\text{score}_{\text{iou}}$  经过标准化后的取值范围为  $[-1, 1]$ 。训练和测试图像的大小均为  $700 * 700$ , 训练时的数据批量大小为 16 张图像, 迭代次数为 12 万次, 学习率的初始值设为 0.001, 在迭代 6 万次后, 学习率调整为 0.000 1, 权重衰减系数和动量分别设为 0.000 1 和 0.9。优化算法采用随机梯度下降法。本章中训练交并比网络用 IOU 损失函数<sup>[19]</sup>, 训练坐标偏移量回归任务用 smooth-L1 作为损失函数, 而文本分类任务采用交叉熵作为损失函数。

### 3 实验与结果分析

#### 3.1 数据集与评价指标

##### 3.1.1 数据集

文中采用 ICDAR2011 和 ICDAR2013 这两个水平文本数据集进行实验。ICDAR2011 包括 229 张训练图像、251 张测试图像, 对文本区域进行单词级别的标

注。ICDAR2013 包括 229 张训练图像、233 张测试图像, 对文本区域进行字符级别和单词级别的标注。这两种数据集中的图像都来自于日常生活中的拍摄, 数据样本的分布充分考虑了自然场景图像可能受到的光照不均匀、曝光过度、遮挡、模糊等影响, 覆盖了大部分复杂场景。使用这两个数据集能够对文中方法进行客观公正的评价。

##### 3.1.2 评价指标

当文本检测器输出一个预测文本框  $D$  时, 可以利用公式(7)计算  $D$  与真实文本框  $G$  的交并比  $\text{IOU}(D, G)$ , 并设置一个交并比阈值 0.7, 如果  $D$  与  $G$  的 IOU 值大于该阈值, 就认为预测出的  $D$  是与  $G$  匹配的检测正确的文本框。

按照预测文本框的正例和反例、真实文本框的正例和反例, 可以将检测结果分为四种不同的组合情况, 并据此对模型检测正确的文本框数量、检测错误的文本框数量、未检测出的文本框数量进行统计, 统计规则列于表 2。

表 2 预测框与真实文本框的匹配数量

预测值真实值	GT=True	GT=False	总和
预测=True	True Positive (TP)	False Positive (FP)	$\text{pre}_T$
预测=False	False Negative (FN)	True Negative (TN)	$\text{pre}_F$
总和	$\text{GT}_T$	$\text{GT}_F$	$S$

表 2 中, 第一列的中间两行分别表示预测结果为文本框和非文本框的情况, 第一行的中间两列表示实际情况下为文本框和非文本框的情况。预测为文本框实际也为文本框的为 True Positive (TP), 表示预测正确的文本框数量; 预测为文本框实际不是文本框的为 False Positive (FP), 表示误检的文本框数量; 预测不是文本框而实际是文本框的为 False Negative (FN), 表示漏检的文本框数量。所有预测出的文本框的数量记作  $\text{pre}_T$ , 所有实际的文本框的数量记作  $\text{GT}_T$ 。

#### 3.2 实验结果分析

基于 ICDAR2011 数据集对候选框的分类置信度和定位置信度的融合系数作了对比实验, 实验结果列于表 3。表 3 中,  $W_{\text{cls}}$  表示分类置信度的系数,  $W_{\text{iou}}$  表示定位置信度的系数。第一行实验结果表示在非极大抑制算法中仅用分类置信度作为排序依据, 即原始方法。随着定位置信度的加入和比重的增大, 文本检测的召回率逐渐提高, 但是当完全用定位置信度替代分类置信度(表 3 最后一行)时, 虽然召回率提高了, 但是准确率也有所下降, 这可能是因为提高召回率的过程中除了保留了许多正确的文本框也引入了一些误检的文本框。因此文中选取 0.2 作为分类置信度的系数、0.8 作为定位置信度的系数。

表 3 不同融合系数的实验对比

$W_{\text{cls}}$	$W_{\text{iou}}$	召回率 $R$	准确率 $P$	F 值
1	0	0.73	0.85	0.79
0.8	0.2	0.74	0.84	0.79
0.5	0.5	0.74	0.85	0.79
0.2	0.8	0.76	0.89	0.82
0	1	0.77	0.87	0.81

文中提出的置信度融合非极大抑制算法 (FC-NMS) 的文本检测方法(下文简称为文中方法)与其他方法在数据集 ICDAR2011 和 ICDAR2013 上的性能对比结果列于表 4 和表 5。

表 4 基于 ICDAR2011 的实验结果

方法	NMS	FC-NMS	召回率 $R$	准确率 $P$	F 值
NMS (Zhang et al.) <sup>[20]</sup>	✓	--	0.76	0.84	0.80
TextFlow <sup>[21]</sup>	✓	--	0.76	0.86	0.81
TextBoxes <sup>[16]</sup>	✓	--	0.74	0.86	0.80
文中方法	--	✓	0.76	0.89	0.82

由表 4 和表 5 可以看出, 文中方法与基准方法 (TextBoxes) 相比, F 值提高了 1%, 主要性能提升体现在准确率上; 在 ICDAR2011 数据集上, 比 TextBoxes 在准确率上提升了 3%; 在 ICDAR2013 数据集上, 比

TextBoxes 在准确率上提升了 2%, 这主要是因为非极大抑制算法中融合了定位置信度, 使得分类置信度较低但定位置信度较高的预测框能够保留下来。综上所述, 置信度融合的文本检测方法可以有效提高文本检测的准确率, 改善文本检测的性能。

表 5 基于 ICDAR2013 的实验结果

方法	NMS	FC-NMS	召回率 $R$	准确率 $P$	F 值
Zhang et al. [20]	✓	--	0.76	0.84	0.80
TextFlow [21]	✓	--	0.71	0.88	0.78
TextBoxes [16]	✓	--	0.74	0.86	0.80
文中方法	--	✓	0.75	0.88	0.81

#### 4 结束语

提出了一种置信度融合的自然场景文本检测方法, 使得检测的文本框更加紧致, 包含的背景区域更少, 能够有效提高自然场景文本检测的准确率。然而, 文中对新设计的交并比分支进行单独训练时, 需要自行准备训练数据, 并对数据翻转、缩放等增广操作, 这种数据准备方式可能会使数据覆盖范围受限, 从而影响模型的训练效果, 降低定位置信度预测效率。因此未来的工作可以继续探究交并比分支训练时对数据集的需求, 满足模型训练需求。

#### 参考文献:

- [1] JUNG K, KIM K I, JAIN A K. Text information extraction in images and video: a survey [J]. Pattern Recognition, 2004, 37 (5): 977-997.
- [2] SMITH J R, CHANG S F. Visual SEEK: a fully automated content-based image query system [C]//Proceedings of the fourth ACM international conference on multimedia. Boston, Massachusetts, USA: ACM, 1996: 87-98.
- [3] SHILKROT R, HUBER J, LIU C, et al. Fingerreader: a wearable device to support text reading on the go [C]//Extended abstracts on human factors in computing systems. Toronto, Ontario, Canada: ACM, 2014: 2359-2364.
- [4] GREENHALGH J, MIRMEHDI M. Recognizing text-based traffic signs [J]. IEEE Transactions on Intelligent Transportation, 2015, 16 (3): 1360-1369.
- [5] BREUEL T M, THOMAS M. The OCRopus open source OCR system [C]//Proceedings of SPIE 6815, conference on document recognition and retrieval XV. San Jose, CA, USA: DBLP, 2008.
- [6] 郭 闯, 邱晓晖. 基于 BLSTM 网络的改进 EAST 文本检测算法 [J]. 计算机技术与发展, 2020, 30 (7): 21-24.
- [7] CHIRON G, DOUCET A, COUSTATY M, et al. Impact of OCR errors on the use of digital libraries: towards a better access to information [C]//Proceedings of the 17th ACM/IEEE joint conference on digital libraries. Toronto, ON, Canada: IEEE, 2017: 249-252.
- [8] 杨雨诺, 张国林, 孙科学, 等. 基于深度学习网络的心音智能分析平台构建 [J]. 计算机技术与发展, 2019, 29 (7): 130-134.
- [9] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector [C]//14th European conference on computer vision. Amsterdam, Netherlands: Springer, 2016: 21-37.
- [10] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection [C]//2016 IEEE conference on computer vision and pattern recognition. Las Vegas, NV, USA: IEEE, 2016: 779-788.
- [11] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39 (6): 1137-1149.
- [12] LECUN Y, BOSER B, DENKER J S, et al. Backpropagation applied to handwritten zip code recognition [J]. Neural Computation, 1989, 1 (4): 541-551.
- [13] 李献军, 孙科学, 张少芳, 等. 关于复杂网络节点的加权融合感知分类算法 [J]. 计算机仿真, 2020, 37 (8): 224-227.
- [14] GIRSHICK R. Fast R-CNN [C]//IEEE international conference on computer vision. Santiago, Chile: IEEE, 2015: 1440-1448.
- [15] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation [C]//IEEE conference on computer vision and pattern recognition. Boston, MA, USA: IEEE, 2015: 3431-3440.
- [16] LIAO M, SHI B, BAI X, et al. TextBoxes: a fast text detector with a single deep neural network [C]//31st AAAI conference on artificial intelligence. San Francisco, CA: AAAI, 2017: 4161-4167.
- [17] LIAO M, SHI B, BAI X, et al. TextBoxes++: a single-shot oriented scene text detector [J]. IEEE Transactions on Image Processing, 2018, 27 (8): 3676-3690.
- [18] JIANG B, LUO R, MAO J, et al. Acquisition of localization confidence for accurate object detection [C]//The European conference on computer vision. Munich, Germany: Springer, 2018: 784-799.
- [19] YU J, JIANG Y, WANG Z, et al. UnitBox: an advanced object detection network [J]. arXiv: 1608.01471, 2016.
- [20] ZHANG Z, SHEN W, YAO C, et al. Symmetry-based text line detection in natural scenes [C]//IEEE conference on computer vision and pattern recognition. Boston, MA, USA: IEEE, 2015: 2558-2567.
- [21] TIAN S X, PAN Y F, HUANG C, et al. Text flow: a unified text detection system in natural scene images [C]//IEEE international conference on computer vision. Santiago, Chile: IEEE, 2015: 4651-4659.