

DOI:10.3969/j.issn.2096-8248.2024.01.010

基于改进 YOLOv5 的快速水平文本检测算法*

孙巧榆, 张静, 刘珍兵

(江苏海洋大学 电子工程学院, 江苏 连云港 222005)

摘要: 基于深度学习的文本检测算法具有强大的特征学习能力和泛化能力, 但推理速度通常较慢。针对此问题, 提出了基于改进 YOLOv5 的快速水平文本检测算法 T-YOLOv5, 通过在 SPPF (spatial pyramid pooling-fast) 模块中嵌入改进的 CAM (channel attention module) 提高网络的特征提取能力, 并在 CIoU (complete IoU) 损失中增加形状损失提高损失函数的收敛速度。所提算法在公共数据集 ICDAR2013 上 F 值达到 86.5, 推理速度达 112 FPS。实验结果表明, 在检测结果和推理速度上, 所提算法 T-YOLOv5 与现有基于深度学习的文本检测算法相比具有一定的竞争力。

关键词: 文本检测; 深度学习; YOLOv5; 场景文本

中图分类号: TP391

文献标志码: A

文章编号: 2096-8248(2024)01-0075-07

引用格式: 孙巧榆, 张静, 刘珍兵. 基于改进 YOLOv5 的快速水平文本检测算法[J]. 江苏海洋大学学报(自然科学版), 2024, 33(1): 75-81.

A Fast Horizontal Text Detection Method Based on the Improved YOLOv5

SUN Qiaoyu, ZHANG Jing, LIU Zhenbing

(School of Electronic Engineering, Jiangsu Ocean University, Lianyungang 222005, China)

Abstract: The text detection methods based on deep learning has strong feature learning ability and generalization ability, but the inference speed is usually slow. To solve this problem, a fast horizontal text detection method T-YOLOv5 based on improved YOLOv5 is proposed. By embedding an improved CAM (channel attention module) in SPPF (spatial pyramid pooling-fast) module, the feature extraction ability of the network is improved, and the shape loss is added to the CIoU (complete IoU) loss to improve the convergence speed of the loss function. The F value of the proposed method reaches 86.5 on the public dataset ICDAR2013, and the inference speed reaches 112 FPS. Experimental results show that the proposed method T-YOLOv5 is competitive with the existing text detection methods based on deep learning in terms of detection results and inference speed.

Key words: text detection; deep learning; YOLOv5; scene text

文本检测在工业生产和日常生活中具有极为广泛的应用, 如银行卡信息提取、发票信息提取、文字翻译等。作为文本图像信息提取过程的一个环节, 文本检测的实时性对信息提取速度有着重要的影

响。因此, 提高文本图像检测的速度具有一定的研究价值。

随着深度学习的发展, 目前目标检测领域的技术已经有了极大提升, 对文本检测也产生了巨大的

* 收稿日期: 2023-02-28; 修订日期: 2023-03-18

基金项目: 江苏海洋大学研究生科研创新项目(DZXS202106)

作者简介: 孙巧榆(1973—), 女, 江苏连云港人, 副教授, 博士, 研究方向为图像分析与智能系统, (E-mail)sunqy@jou.edu.cn。

影响。传统的文本检测算法多依赖于手工设计的特征,在复杂背景和任意形状文本的检测方面存在一定困难。而基于深度学习的检测算法采用神经网络进行特征提取和训练,可以自主学习文本特征,有着很强的泛化性,因此,深度学习的检测算法日益增多。最初的深度学习文本检测算法参考目标检测算法将文本作为目标进行检测,再根据文本特征进行组合,如 Tian 等^[1]提出的 CTPN(connectionist text proposal network),借鉴了 RPN(region proposal network)^[2]的思想,采用细长的垂直锚框检测文本或者文本的一部分,然后将文本框连接起来形成完整的文本区域。

基于深度学习的算法大致可以分为两类,即基于回归的方法和基于分割的方法。基于回归的方法大多数着重于回归提议区域的文本框坐标,有预设锚框的属于间接回归,无锚框的属于直接回归。Zhong 等^[3]结合 Inception 模块^[4]与 RPN^[2]思想提出 Inception RPN(inception region proposal network)直接生成文本区域建议,并通过两个不同尺寸的卷积和最大池化提取多尺寸文本的特征,对候选区域进行约束,减少了文本的漏检。Liao 等^[5]基于 SSD^[6]框架进行了改进,提出 TextBoxes,将 SSD 的全连接层换成卷积层,并重新设计了锚框的比例,使之更适用于水平文本的检测。Zhong 等^[7]提出了无锚框区域建议网络(anchor-free region proposal network, AF-RPN),将文本根据尺寸分为大、中、小三种类型,采用 FPN^[8]结构分别检测三种尺寸的文本,生成高质量的文本建议区域,然后采用滑动检测器进行分类和回归。基于分割的方法基本依赖于对像素的预测,包括基于聚类的、基于组件的和基于中心点扩展的。针对基于字符检测需要复杂的后处理和文本尺寸差异过大导致计算需求增加的问题,He 等^[9]提出了级联卷积文本网络(cascaded convolutional text network, CCTN),对粗略特征图根据文本的几何特征提取文本行,得到文本行中心和文本行分离两张特征图,然后通过裁剪放大文本区域等操作处理精细文本行。针对标准二值化不可微问题,Liao 等^[10]提出可微分二值化模块(differentiable binarization, DB)作为二值化的近似函数,更好地区分文本与非文本,并结合 FPN 结构提高检测效果,后来又提出自适应尺度融合(adaptive scale fusion, ASF)模块^[11],采用空间注意力学习不同尺度特征图的权重,增加尺度融合的鲁棒性。

与本实验相关的方法有文献^[12-13]等的方法。

王霏等^[12]提出基于 YOLOv3(you only look once v3)^[14]改进的轻量级快速文本检测算法,通过深度可分离卷积思想轻量化 Darknet-53 网络,为卷积增加偏移层实现可变感受野卷积,并在 D-IoU(distance-IoU)^[15]损失中引入宽度损失,提高了检测精度。周翔宇等^[13]提出基于 YOLO 的倾斜文本定位算法 YOLO_BOX,采用多尺度小型锚框对文本进行覆盖,再对锚框进行聚类然后绘制最小外接矩形框,再利用 NMS 和 Angle Correct 算法去除多余边框并进行角度矫正。基于深度学习的文本检测算法相较于传统检测算法具有更强的特征提取能力,但推理速度有待提高。

本文针对上述问题,提出了基于改进 YOLOv5 的快速水平文本检测算法 T-YOLOv5,以提高文本检测的推理速度,在 SPPF(spatial pyramid pooling-fast)分支上增加注意力模块来增强 SPPF 模块获取多尺度图像特征的能力,并在 CIoU(complete IoU)^[15]损失中增加形状损失,进而提升损失的收敛速度。

1 T-YOLOv5 算法设计

1.1 T-YOLOv5 网络模型

T-YOLOv5 网络主要采用 YOLOv5 网络框架,结构如图 1 所示。主干网络采用 CSPDarkNet53,整体网络由 CBS, CSP 和上采样等模块构成。图中 CSP1_x 表示该模块堆叠的 x 块,Conv 表示卷积,其后的 BN 和 SiLU 分别表示 BN 层和 SiLU 激活函数,Resx 表示残差模块,Shortcut = True 表示采用跳跃连接,Shortcut = False 表示不采用跳跃连接。

1.2 SPPFCAM * 模块

注意力机制被广泛应用于深度学习网络中,它的核心思想是使网络选择性地关注特征图的突出特征。Woo 等^[16]提出的 CBAM(convolutional block attention module)将通道注意力和空间注意力联合起来,用来学习需要被关注的信息和抑制不需要的信息。其中,通道注意力 CAM(channel attention module)的结构如图 2 所示, F 表示输入特征, M_c 表示通道注意力的输出特征。CAM 模块同时使用最大池化和平均池化,然后经过一个多感知器 MLP(multi-layer perceptron),将通道进行压缩再进行扩充,MLP 的两个输出再进行逐元素相加,使特征实现跨通道通信。

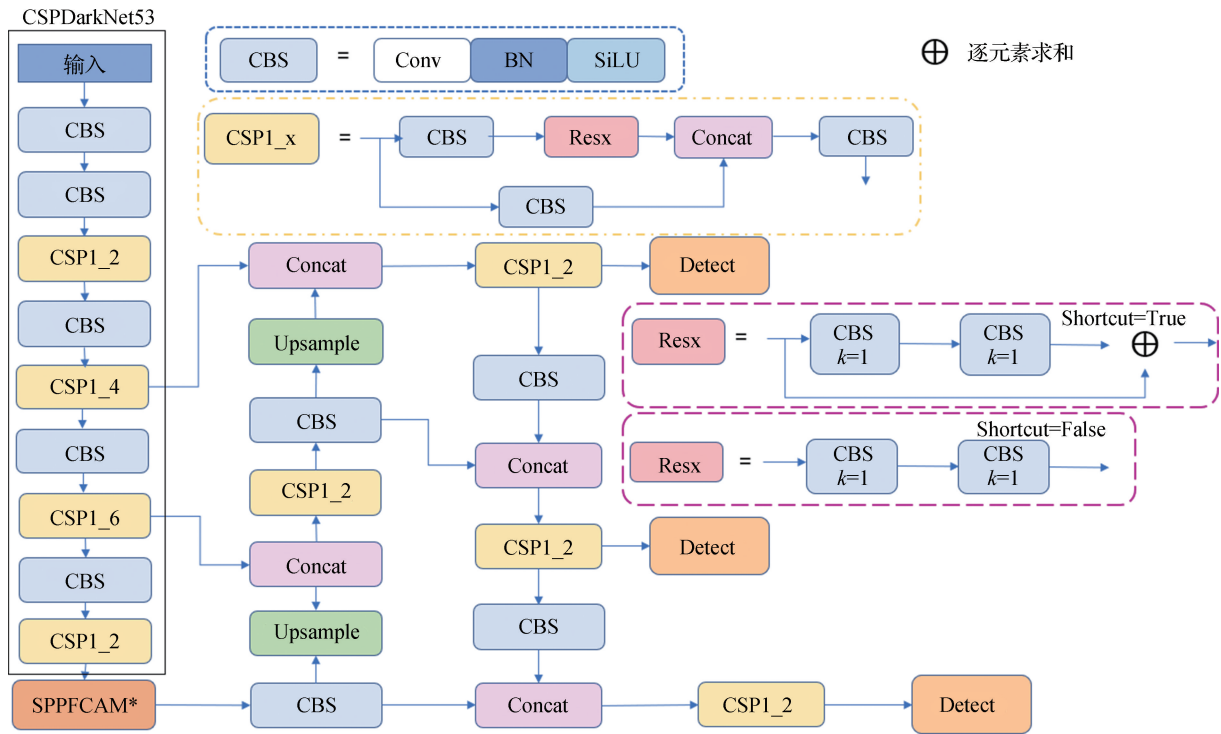


图 1 T-YOLOv5 网络结构

Fig. 1 Architecture of T-YOLOv5 network

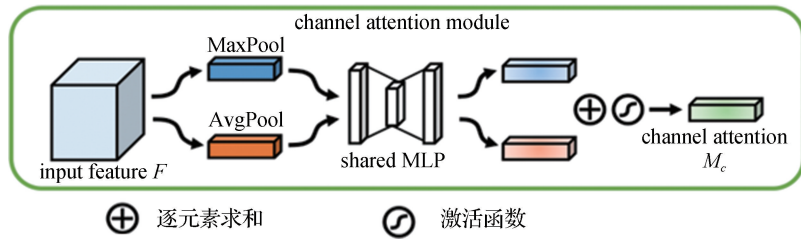


图 2 CAM 模块结构

Fig. 2 Architecture of CAM modules

CAM 注意力作为即插即用的模块,可以应用于任何 CNN 的网络中。相对于目标检测,真实框中文本的像素是较为稀疏的。因此,CAM 可以应用于 SPPF 中以增强网络对突出特征的关注和对背景信息的抑制能力,并且,CAM 注意力模块需要的参数量较少,不会为网络增加太大的计算量。为了

获得一些位置信息,模仿空间注意力 SAM(spatial attention module)在 CAM 模块中增加了均值操作,改进后的 CAM 模块称为 CAM*,其结构如图 3 所示。

添加 CAM* 的 SPPF 模块称为 SPPFCAM*,其结构如图 4 所示。

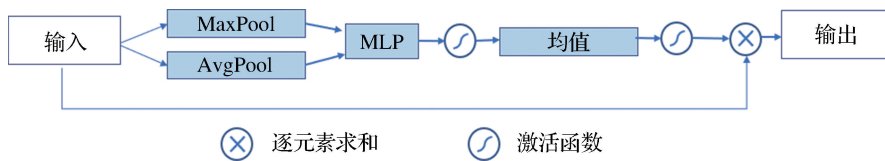


图 3 CAM* 模块结构

Fig. 3 Architecture of CAM* modules

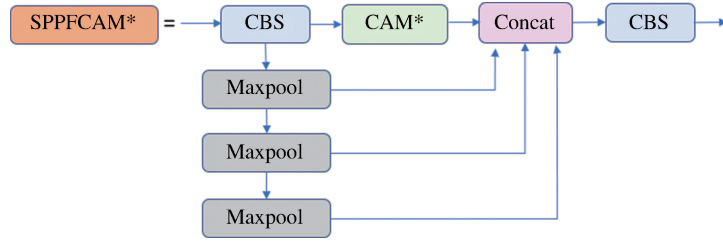


图4 SPPFCAM* 模块结构

Fig. 4 Architecture of SPPFCAM* modules

1.3 损失函数

损失函数主要由三个部分组成,包括包围盒损失(即定位损失)、分类损失和置信度损失。包围盒损失用来计算真实框与预测框之间的误差,分类损失用来衡量锚框内目标分类的准确率,置信度损失表示预测框的可信程度。

1.3.1 包围盒损失(CSIoU 损失) YOLOv5 的包围盒损失采用 CIoU^[15]。CIoU 不仅考虑了预测框与中心框之间的距离,还包含了对真实框和预测框宽高比的衡量,其公式为

$$\text{CIoU} = \text{IoU} - \frac{\rho^2(b, b^{\text{gt}})}{c^2} - \alpha_v v, \quad (1)$$

$$L_{\text{CIoU}} = 1 - \text{CIoU} = 1 - \text{IoU} + \frac{\rho^2(b, b^{\text{gt}})}{c^2} + \alpha_v v, \quad (2)$$

$$v = \frac{4}{\pi^2} \left(\arctan \frac{w^{\text{gt}}}{h^{\text{gt}}} - \arctan \frac{w}{h} \right)^2, \quad (3)$$

$$\alpha_v = \frac{v}{(1 - \text{IoU}) + v}. \quad (4)$$

其中, $\frac{\rho^2(b, b^{\text{gt}})}{c^2}$ 表示预测框和真实框中心点的欧式距离, c 表示预测框和真实框的最小外接矩形的对角线长度, w^{gt} 和 h^{gt} 分别表示真实框的宽和高, w 和 h 分别表示预测框的宽和高, α_v 是正权衡参数。为了提高损失函数的收敛速度, SIoU (Scylla-IoU)^[17] 采用另一种方式描绘预测框和真实框之间的相似度,文献中认为采用靠近 x 轴或 y 轴的方式描述包围盒损失可以加快训练。损失分为三种:角度损失、距离损失和形状损失,没有引进其他参数,公式如下:

$$\Lambda = 1 - 2 \times \sin^2 \left(\arcsin x - \frac{\pi}{4} \right), \quad (5)$$

$$x = \frac{C_h}{\sqrt{C_h^2 + C_w^2}} = \sin \alpha, \quad (6)$$

$$\Delta = \sum_{t=x, y} (1 - e^{-\gamma \rho_t}), \quad (7)$$

$$\rho_x = \left(\frac{b_x^{\text{gt}} - b_x}{w_c} \right), \rho_y = \left(\frac{b_y^{\text{gt}} - b_y}{h_c} \right), \gamma = 2 - \Lambda, \quad (8)$$

$$\Omega = \sum_{t=w, h} (1 - e^{-\omega_t})^\theta, \quad (9)$$

$$\omega_w = \frac{|\omega - \omega^{\text{gt}}|}{\max(\omega, \omega^{\text{gt}})}, \omega_h = \frac{|h - h^{\text{gt}}|}{\max(h, h^{\text{gt}})}, \quad (10)$$

$$R_{\text{SIoU}} = \Delta + \Omega, \quad (11)$$

$$L_{\text{SIoU}} = 1 - \text{IoU} + 0.5 \times R_{\text{SIoU}}. \quad (12)$$

其中: Λ 表示角度损失, Δ 表示距离损失, Ω 表示形状损失, Ω 分别考虑了真实框与预测框宽和高的真实差异; $b_x^{\text{gt}}, b_y^{\text{gt}}$ 分别表示真实框中心点 x, y 坐标, b_x, b_y 分别表示预测框中心点 x, y 坐标; C_w, C_h 分别表示真实框和预测框中心点之间的水平距离和垂直距离。角度 α 的表示如图 5 所示,若 $\alpha > 45^\circ$, 则用 $\beta = \frac{2}{\pi} - \alpha$ 代替 α , 左下角蓝框表示预测框, 右上角红框表示真实框, B 和 B^{gt} 分别表示预测框、真实框中心点, 点划线表示两个框的最小外接矩形, w_c 和 h_c 分别表示最小外接矩形的宽和高, $\theta \in [2, 6]$, 本文中 θ 取 4。

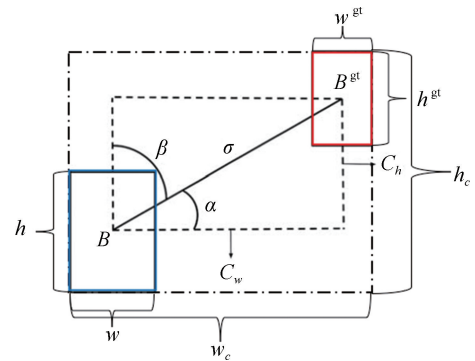


图5 SIoU 的角度表示

Fig. 5 Angular representation of SIoU

CIoU 损失中的 v 考虑了宽高比的差异,而未考虑真实框与预测框宽和高的真实差异,这有时会对包围盒损失的收敛造成阻碍。如图 6 所示,两个

预测框与真实框的宽高比是相同的,黑色预测框宽、高是真实框的 1/2,蓝色预测框宽、高是真实框的 3/4,但蓝色预测框的 CIoU 损失为 1.156,比黑色预测框损失小 0.004,两预测框损失的差值过小不利于损失的优化。

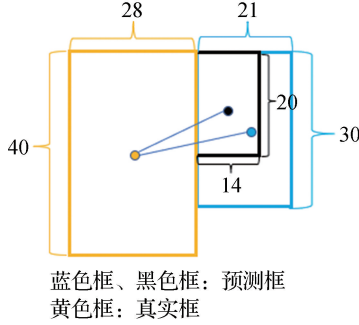


图 6 真实框外的不同预测框

Fig. 6 Different prediction box outside the real box

增加真实框与预测框宽、高的真实差异有助于包围盒损失的收敛。本文在 CIoU 基础上增加 SIoU 的形状损失 Ω ,重新调整宽高比和形状损失在 IoU 中的占比,提出了 CSIoU 损失。CSIoU 和 L_{CSIoU} 计算过程如公式(13)和(14)所示,变量含义与上文相同。

$$\text{CSIoU} = \text{IoU} - \left(\frac{\rho^2(b, b^{\text{gt}})}{c^2} + 0.9 \times \alpha_v v + 0.1 \times \Omega \right), \quad (13)$$

$$L_{\text{CSIoU}} = 1 - \text{CSIoU}. \quad (14)$$

图 6 中若采用 CSIoU 损失,两预测框损失的差值会增大,则有利于包围盒损失的收敛。

1.3.2 分类损失与置信度损失 分类损失和置信度损失与 YOLOv5 损失相同,采用二值交叉熵(Binary Cross-Entropy)计算损失,但只计算正样本的损失,其公式如下:

$$C = -\frac{1}{n} \sum_i [y \ln a + (1-j) \ln(1-a)]. \quad (15)$$

其中, i 表示样本, j 表示标签, a 表示预测的输出, n 表示样本总量。

2 实验结果及分析

2.1 实验环境与参数设置

本实验在智星云服务器上完成,采用普通性能的单片 GPU,型号为 GeForce RTX 2080 Ti(11 GiB 显

存),CUDA 版本为 10.0,Pytorch 版本为 3.7.4。YOLOv5 版本为 6.1,在网络深度适中的 YOLOv5m 模型基础上继续训练,batch size 设置为 32,epochs 为 130。默认的超参数为:采用多项式衰减学习率调度策略,初始学习率为 0.01,预热步数为 3,权值衰减设为 0.005。

2.2 数据集

消融实验和对比实验在公共数据集 ICDAR2013^[18]上进行。ICDAR2013 数据集由 462 张英文标注的场景文本图像组成,包含 229 张训练图像和 233 张测试图像,主要为水平文本,含少量倾斜文本,标签以文本行为单位,由左上角点坐标和右下角点坐标组成,改为 YOLO 格式的标签,由文本框中心点坐标、框的宽和高组成。训练集中挑出 79 张图像作为验证集,剩余图像与 ICDAR2011^[19]数据集的部分图像一起作为训练集,共 556 张训练集图像。训练集、验证集、测试集的比例约为 7:1:3。ICDAR2011 数据集主要由水平的英文电子图像组成,训练时剔除了重复的图像。

2.3 性能评价指标

文本检测算法共有 3 个评价指标,分别为准确率 P (precision)、召回率 R (recall)和综合评价指标 F (F-measure)。本文采用以下计算方式。

(1) 准确率。准确率 P 表示预测为真的标签中真实标签的个数,公式为

$$P = \frac{\text{TP}}{\text{TP} + \text{FP}}. \quad (16)$$

式中,TP 表示真实标签中预测正确的标签个数,FP 表示将错误标签预测为真实标签个数,TP+FP 表示预测为真实标签的个数。

(2) 召回率。召回率 R 表示真实标签中预测正确的标签个数,公式为

$$R = \frac{\text{TP}}{\text{TP} + \text{FN}}. \quad (17)$$

式中,FN 表示将真实标签预测错误的标签个数,TP+FN 表示所有真实标签的个数。

(3) 综合评价指标。综合评价指标 F 是 P 和 R 的加权调和平均,公式为

$$F = \frac{2 \times P \times R}{P + R}. \quad (18)$$

2.4 消融实验

(1) CSIoU 损失。为了验证 CSIoU 损失的有效性,将其与 CIoU 和 SIoU 损失进行了比较,在 ICDAR2013 数据集上的实验结果如表 1 所示。由

表1可以看出,与其他损失函数相比,CSIoU的表现最好,同时提高了 P 值和 R 值。

表1 损失函数对比实验

Table 1 Comparative experiment of loss function

损失函数	P	R	F
CIoU	89.4	82.1	85.6
SIoU	89.8	81.8	85.6
CSIoU	89.9	82.7	86.1

注:加粗字体表示该列中此数值最高。下表同。

(2) SPPFCAM * 模块。为了验证CAM * 模块的有效性,将其与CBAM及其单独的注意力模块进行了比较,在ICDAR2013数据集上的结果如表2所示。其他模块的位置与SPPFCAM * 模块中CAM * 位置相同。由表2可以看出,同时使用通道注意力CAM和空间注意力SAM对算法几乎没有提升,单独使用算法性能提升较小,增加了CAM * 模块的算法同时提高了 P 值和 R 值。同时采用SPPFCAM * 模块和CSIoU的算法 F 值相较于原算法提高了1%。

表2 模块实验对比

Table 2 Comparison of module experiments

模块	P	R	F
SPPF	89.4	82.1	85.6
SPPF+CBAM	89.2	82.6	85.8
SPPF+CAM	90.4	82.2	86.1
SPPF+SAM	89.4	82.9	86.0
SPPFCAM *	90.7	82.4	86.4
CSIoU+SPPFCAM *	91.3	82.2	86.5

2.5 结果分析

T-YOLOv5与YOLOv5m算法以及其他文本检测算法在ICDAR2013数据集上的实验结果如表3所示。

表3 在ICDAR2013数据集上的实验对比

Table 3 Experimental comparison on ICDAR2013 dataset

方法	P	R	F	推理速度/ FPS
YOLO_BOX ^[12]	86.0	77.0	81.0	—
RRD ^[20]	75.0	88.0	81.0	—
文献[21]的方法	93.3	79.3	85.8	10.4
SegLink ^[22]	87.7	83.0	85.3	20.6
YOLOv5m	89.9	82.7	85.6	119.0
T-YOLOv5	91.3	82.2	86.5	112.3

由表中可以看出,YOLOv5m算法的 F 值是85.6,推理速度达到119 FPS,改进后的算法T-YOLOv5的 F 值比改进前提高了1%,推理速度几乎不变。相较于表中的文本检测算法SegLink^[22],T-YOLOv5的速度提高了4倍,且 F 值远远高于同样改进YOLOv5算法的YOLO_BOX^[12]。本实验采用的YOLOv5m模型网络的深度和宽度在YOLOv5系列中属于中等,因此,采用YOLOv5检测文本仍然是具有潜力的。

3 结论

随着文本图像检测技术的不断提升,检测的实时性受到了越来越多的关注。本文针对深度学习文本检测算法实时性较差的问题,提出了基于改进YOLOv5的快速水平文本检测算法T-YOLOv5,在SPPF中增加改进后的通道注意力以增强网络的特征提取能力,在CIoU损失的基础上增加形状损失提出CSIoU损失,提高了损失的收敛速度。T-YOLOv5的推理速度达到了112 FPS,使得基于深度学习的水平文本检测算法的实时性得到了提升。但算法仍然存在两个方面的不足:第一,相较于目前的其他几种文本检测算法,其检测性能还有待提高;第二,文本形状需要向多方向文本和任意形状文本拓展。

参考文献:

- [1] TIAN Zhi, HUANG Weilin, HE Tong, et al. Detecting text in natural image with connectionist text proposal network[C]//Proceedings of the European Conference on Computer Vision, 2016: 56-72.
- [2] REN Shaoqing, HE Kaiming, 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.
- [3] ZHONG Zhuoyao, JIN Lianwen, HUANG Shuangping. Deeptext: a new approach for text proposal generation and text detection in natural images[C]//Proceedings of the 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2017: 1208-1212.
- [4] SZEGEDY C, LIU Wei, JIA Yangqing, et al. Going deeper with convolutions [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015: 1-9.

- [5] LIAO Minghui, SHI Baoguang, BAI Xiang, et al. Textboxes: a fast text detector with a single deep neural network[C]//Proceedings of the Thirty-first AAAI Conference on Artificial Intelligence, 2017: 4161-4167.
- [6] LIU Wei, ANGUELOV D, ERHAN D, et al. Ssd: single shot multibox detector[C]//Proceedings of the European Conference on Computer Vision, 2016: 21-37.
- [7] ZHONG Zhuoyao, SUN Lei, HUO Qiang. An anchor-free region proposal network for Faster R-CNN-based text detection approaches[J]. International Journal on Document Analysis and Recognition (IJ DAR), 2019, 22(3): 315-327.
- [8] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2117-2125.
- [9] HE Tong, HUANG Weilin, QIAO Yu, et al. Accurate text localization in natural image with cascaded convolutional text network [EB/OL]. (2016-03-31) [2023-02-03]. <https://arxiv.org/abs/1603.09423>. 2016.
- [10] LIAO Minghui, WAN Zhaoyi, YAO Cong, et al. Real-time scene text detection with differentiable binarization[C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 11474-11481.
- [11] LIAO Minghui, ZOU Zhisheng, WAN Zhaoyi, et al. Real-time scene text detection with differentiable binarization and adaptive scale fusion[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(1): 919-931.
- [12] 王霏, 黄俊, 文洪伟. 基于改进 YOLOv3 的快速文本检测[J]. 电讯技术, 2022, 62(1): 130-137.
- [13] 周翔宇, 高仲合. 基于 YOLO 的自然场景倾斜文本定位方法研究[J]. 计算机工程与应用, 2020, 56(9): 213-220.
- [14] REDMON J, FARHADI A. YOLOv3: an incremental improvement[EB/OL]. (2018-04-08) [2023-02-02]. <https://arxiv.org/abs/1804.02767>.
- [15] ZHENG Zhaohui, WANG Ping, LIU Wei, et al. Distance-IoU loss: faster and better learning for bounding box regression [C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 12993-13000.
- [16] WOO S, PARK J, LEE J Y, et al. Cbam: convolutional block attention module[C]//Proceedings of the European Conference on Computer Vision, 2018: 3-19.
- [17] GEVORGYAN Z. SIoU loss: more powerful learning for bounding box regression [J]. (2022-05-25) [2023-02-02]. <https://arxiv.org/abs/2205.12740>.
- [18] KARATZAS D, SHAFAIT F, UCHIDA S, et al. ICDAR 2013 robust reading competition[C]//Proceedings of the 2013 12th International Conference on Document Analysis and Recognition, 2013: 1484-1493.
- [19] SHAHAB A, SHAFAIT F, DENGEL A. ICDAR 2011 robust reading competition challenge 2: reading text in scene images[C]//Proceedings of the 2011 International Conference on Document Analysis and Recognition, 2011: 1491-1496.
- [20] LIAO Minghui, ZHU Zheng, SHI Baoguang, et al. Rotation-sensitive regression for oriented scene text detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 5909-5918.
- [21] LYU Pengyuan, YAO Cong, WU Wenhao, et al. Multi-oriented scene text detection via corner localization and region segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7553-7563.
- [22] SHI Baoguang, BAI Xiang, BELONGIE S. Detecting oriented text in natural images by linking segments [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2550-2558.

(责任编辑:秦海明 实习编辑:易圣杰)