计算机科学
COMPUTER SCIENCE

R-YOLOv5:自动切割的旋转的文本检测模型

冉煜, 张莉

引用本文

冉煜, 张莉. R-YOLOv5:自动切割的旋转的文本检测模型[J]. 计算机科学, 2022, 49(11A): 210900185-6.

RAN Yu, ZHANG Li. R-YOLOv5:Auto-cutting,Rotated Text Detection Model[J]. Computer Science, 2022, 49(11A): 210900185-6.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.210900185

R-YOLOv5:自动切割的旋转的文本检测模型

冉 煜 张 莉
对外经济贸易大学信息学院 北京 100029
(2920763948@qq.com)

摘 要 YOLOv5模型是目前文本检测较好的模型之一,针对文本目标长度不一,文本轮廓难以精准检测以及受自然场景中文字倾斜、光影的影响文本较难检测的问题,提出了R-YOLOv5(Rotated-YOLOv5)文本检测模型。首先融入基于仿射算法的文本分割模型,将图片的文本区域等比例切割为多个单字符块,解决文本没有闭合轮廓导致的YOLOv5模型锚定框拟合效果不佳的问题;然后使用旋转卷积层、旋转池化层、改进锚定框,提出了加强角度学习的RIoU(Rotated Intersection over Union)损失函数,实现了文本旋转倾斜特征的提取。在ICDAR2019-LSVT上对原模型与改进后的模型进行实验,实验结果显示,R-YOLOv5检测效果有较明显的提升,但由于模型层数加深,训练速率与检测速率相比原模型有小幅降低。相比其他模型,由于YOLOv5自身的优点,R-YOLOv5的检测效果与检测速度均远好于其他模型。

关键词:计算机视觉;目标检测;文本检测;卷积神经网络;旋转倾斜;损失函数;YOLO
中图法分类号 TP389.1

R-YOLOv5:Auto-cutting,Rotated Text Detection Model

RAN Yu and ZHANG Li
School of Information Technology and Management,University of International Business and Economics,Beijing 100029,China

Abstract YOLOv5 model is currently one of the best models for object detection. To solve the problem of different lengths of text lines,the inclination of text,light and shadow in natural scenes,etc. the R-YOLOv5(Rotated-YOLOv5) text detection model is proposed,which improves the YOLOv5 model to deal with the weakness in text detection. Firstly,the text segmentation model based on affine algorithm is incorporated. According to the length of the string and the shape of the text area,the text area of the picture is cut into multiple single-character blocks in equal proportions to solve the problem of poor effect of YOLOv5 model caused by the text objects without closed contour lines. Then,using the rotated convolutional neural network layer,rotated max-pooling layer and improved anchor box,we propose a rotated intersection over union(RIoU) loss function that strengthens angle learning to achieve the extraction of rotation and tilt features. The original model and the improved model are tested on ICDAR2019-LSVT. Experimental results show that the detection effect of R-YOLOv5 are significantly improved. However,due to the deepening of model layers,the training efficiency and detection efficiency are slightly reduced compared with the original model. Compared with other models,due to the advantages of YOLOv5,the detection effect and efficiency of R-YOLOv5 are much better than that of other models.

Keywords Computer vision,Object detection,Text detection,Convolutional neural network,Rotation tilt,Loss function,YOLO

1 引言

随着目标检测技术的发展和应用,对图像或视频中的文本进行检测识别,形成语义表达,辅助计算机视觉的实际应用,提高各种应用场景文档分析的效率,成为了计算机视觉领域的研究热点之一,引起了国内外学者的关注,提出了一系列文本检测模型和方法,特别是深度学习在多层次特征提取与融合方面展现出极大优势后,深度学习逐渐成为当前主流的文本检测方法,主要有One-stage模式和Two-stage模式两种检测方法。One-stage模式是端对端的目标检测模型,输入图像经过模型检测直接输出检测结果(包括位置信息、目标类型等),典型模型主要有YOLO(You Only Look Once)[1]系列,包含YOLO9000[2]、YOLOv3[3]、YOLOv4[4]、YOLOv5模型与SSD(Single Shot Multibox Detector)模型[5]系列。Two-stage模式则经过两步对目标进行识别,首先生成目标候选区域,然后提取候选区域内的信息对每个候选区域经过全连接网络识别出目标信息,典型模型有R-RPN(Rotation Region Proposal Network)[6]、CTPN(Connectionist Text Proposal Network)[7]、EAST(Efficient and Accurate Scene Text Detector)[8]等。One-stage模式只需要一步便可以完成,运行速度相对更快,Two-stage模式则普遍检测较为准确。而兼顾两种模式优势的目标检测模型是最近提出的YOLOv5模型,这是目前运行速度和检测效果较好的模型之一。但YOLOv5模型仍然存在着一些问题,对于倾斜文本、非闭合轮廓的文本检测效果不佳。本文对YOLOv5模型的不足之处进行了改进,提出了针对旋转区域文本检测的R-YOLOv5模型,主要的贡献有两方面:

(1)借鉴R-RPN模型的旋转候选区域的思想,将

通信作者:张莉(zhangli_amy@uibe.edu.cn)

YOLOv5 模型的水平矩形锚定框改为旋转的平行四边形锚定框，并加强对旋转角度与倾斜角度的学习，在 YOLOv5 框架内加入旋转、倾斜卷积层与旋转、倾斜池化层以加强对角度特征的提取，从而更好地学习与检测不同方向的文本区域语义信息。

(2)借鉴字符仿射分割算法，将图片的文本行区域等比例切割为多个单字符块，解决文本行没有闭合轮廓导致的 YOLOv5 模型锚定框拟合效果不佳的问题。

2 YOLOv5 模型

Redmon 等于 2016 年 5 月发布了第一版的 YOLO 模型，将图像平均分成 7×7 个格子，每个检测框由中心点所在格子确定，利用每个格子内的框计算目标位置与每个类的置信度，筛掉置信度低于设定的阈值的格子，并对检测框进行 NMS(Non-Maximum Suppression)运算，以获得所有检测框。2016 年 12 月发表的 YOLO9000 模型中加入了批规范化，对卷积层各通道间的数据进行规范化处理，并运用 Darknet-19 加深特征提取，对于深层语义信息提取有更好的效果。同时加入了先验框的理念，给图像分出的每一格子预设一定数量的先验框，并采用聚类方法找出最佳先验框尺寸。YOLO9000 模型在训练时输入较高分辨率图像，使模型在检测时可以更好地适应高分辨率，增快检测效率。2018 年 Redmon 等提出了 YOLOv3 模型，继续加深 Darknet-19 特征提取深度，改用 Darknet-53，对深层特征有更好的提取效果，同时 YOLOv3 中首次加入了 FPN(Feature Pyramid Network)[9]，通过上采样与卷积实现跨尺度检测，对不同粒度的特征进行提取，并进行多粒度特征融合，以更好获取深层图像信息，从而实现更好的检测结果。由于 Darknet-53 结构与卷积层、残差块的交替使用很好地提升了 GPU 使用率，检测速率并未降低。

2020 年 4 月 Bochkovskiy 等发布了 YOLOv4 模型，加强了深度特征提取过程中对原特征的保留，融合了 CSPNet(Cross Stage Partial Network)结构与 Darknet-53 形成了 YOLOv4 的 Backbone 结构——CSPDarknet-53。YOLOv4 加深了对特征融合的运用，Neck 部分添加了 SPP(Spatial Pyramid Pooling)结构与 PANet(Path Aggregation Network)结构[10]，两个模块利用多尺度特征进行特征融合以加强信息提取能力。新增了 Mosaic 数据增强，使在训练时可同时训练 4 张图片，训练速度更快，同时增强了对小目标的检测能力。

YOLOv5 网络结构与 YOLOv4 几乎没有区别，主要分为 Backbone、Head 两个部分，图 1 给出了最新的 YOLOv5 4.0 版本流程示意图。Backbone 部分通过多层卷积与残差边的配合提取了 3 个尺度的特征图，然后用池化层进行初步多粒度特征融合。Head 部分利用上采样配合卷积层进行深度多粒度特征融合，增强语义信息的提取能力，最后完成检测。数据预处理运用自适应图片缩放，避免输入网络的图片有较宽灰条，有效减少了运算时的冗余数据，将检测速率再次提升一个阶段，Backbone 中新增了 Focus 模块，将原数据分割成 4 份进行叠加，并且对 YOLOv4 原网络结构进行了少量更改。

YOLO 模型根据网络的输出和标签信息进行激活函数和损失函数的设定，实现网络参数的更新。最新的 YOLOv5 4.0 版本采用了 Mish 激活函数[11]和 CIoU 损失函数。[12]

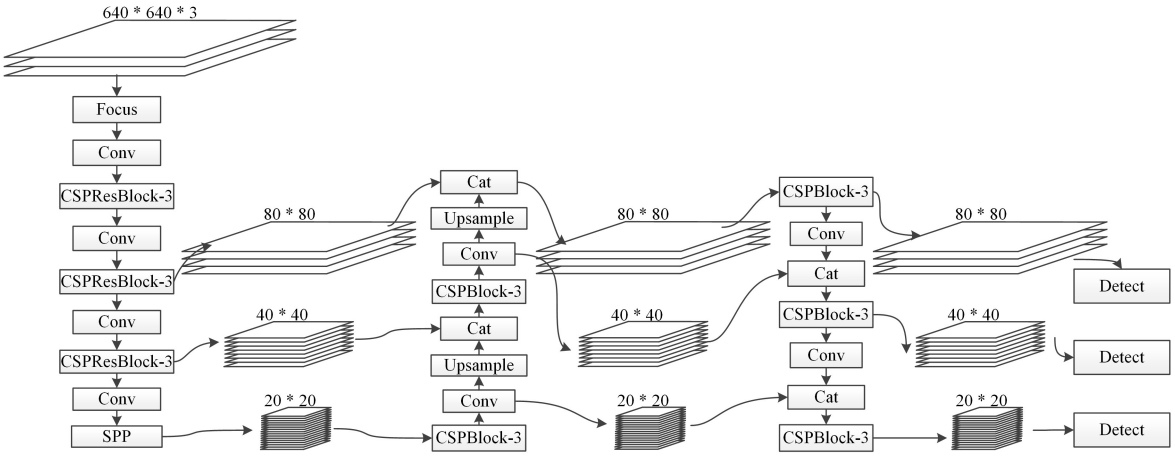


图 1 YOLOv5 4.0 模型整体流程示意图

Fig. 1 Schematic diagram of overall process of YOLOv5 4.0

YOLOv5 相比其他目标检测算法的优势十分明显，具体如下：

首先，自适应图片填充算法极大地加快了检测速率。通过最大程度地减少填充冗余数据，训练速率提升到 37%，成为了当前训练最高效的模型。

其次，模型结构几乎融合当前所有前沿研究成果并进行了一定创新。CSPDarknet-53 结构使其可以提取极深的特征，而 FPN 与 PAN 两层特征融合模型对三层特征图进行两次特征融合，使其可以很好地连结浅层与深层特征，通过特征深度融合挖掘深层语义信息。同时，最新的 Mish 激活函数与 CIoU 损失函数可以达到当前最好的学习效果，极大地增强了准确率。

YOLOv5 虽然可以针对形式多样的目标达到很好的检测效果，但由于文本行独特的旋转、非闭合轮廓特点，使得 YOLOv5 应用于文本检测时并不能很好地拟合文本区域边框。YOLOv5 输出的检测框的边界是水平垂直的，而文本区域多为不规则四边形，无法完全拟合倾斜的边界。由于字符串长度不同，不同文本区域长度与宽度的比值有很大区别。YOLOv5 训练出的有限个锚定框无法完美适配不同形状的文本区域，受限于预制锚定框的形状，检测框左侧通常囊括大片非文本区域，右侧的文本区域则很难被完全覆盖。且文本区域受现场环境影响较大，颜色、清晰度等

有较大差别。为此，本文结合 RRoI 的思想，提出了旋转倾斜卷积层、池化层，并借鉴文献[13]中提到的文本区域的分割方法，改进了 YOLOv5 模型。

3　改进的 YOLOv5 模型——R-YOLOv5

针对 YOLOv5 模型对倾斜文本、非闭合轮廓的文本检测表现不佳的情况，本文提出了针对旋转区域文本检测的 R-YOLOv5 模型。

3.1　旋转锚定框

基于旋转候选区域（RRoI）的旋转目标匹配思想[14]，将 YOLOv5 的五元标签锚定框（中心点 X 坐标、中心点 Y 坐标、宽度 w、高度 h、类别名称 c_str）加入旋转角度正切值 $\tan\alpha$ 和宽高右上夹角余切值 $\cot\beta$，将其修改为七元标签锚定框，以加强旋转、倾斜变形的锚定框（见算法 1）。

算法 1　旋转锚定框坐标计算

输入：(X,Y,w,h,tanα,cotβ)，且 X,Y,w,h∈[0,1]

输出：$((x_1,y_1),(x_2,y_2),(x_3,y_3),(x_4,y_4))$

1. $Q=\dfrac{(h-w\times\tan\alpha)\times(\cot\beta-\tan\alpha)}{1+\tan\alpha^2}$

2. $(x_1,y_1)=(X-\dfrac{w}{2},Y-\dfrac{h}{2}+(w-Q)\times\tan\alpha)$

3. $(x_2,y_2)=\left(X+\dfrac{w}{2}-Q,Y-\dfrac{h}{2}\right)$

4. $(x_3,y_3)=\left(X+\dfrac{w}{2},Y-\dfrac{h}{2}+\dfrac{Q\times\cot\beta^2}{(\cot\beta-\tan\alpha)\times(\cot\beta\tan\alpha+1)}\right)$

5. $(x_4,y_4)=\left(X-\dfrac{w}{2}+Q,Y+\dfrac{h}{2}\right)$

3.2　RIoU 损失函数

为了加强对旋转、倾斜角度的学习，在 CIoU 损失函数基础上加入角度损失项，提出使用增强角度损失的损失函数 RIoU，如式(1)所示：

$$L_{RIoU}=1-IoU+\frac{\rho^2(b,b^{gt})}{c^2}+a\vartheta+\frac{b\mu}{2\ln 3} \tag{1}$$

其中，$\alpha,\beta,\alpha^{gt},\beta^{gt}$ 分别表示预测框与目标框的旋转角度、横轴纵轴倾斜夹角。$\vartheta=\dfrac{4}{\pi^2}\left(\arctan\dfrac{w^{gt}}{h^{gt}}-arctan\dfrac{w}{h}\right)^2$，且 $a=\dfrac{\vartheta}{(1-IoU)+\vartheta}$。

μ 为新角度损失项，运用对数函数控制损失函数平滑且梯度不断放缓，如式(2)所示。b 为角度损失项的平衡参数，k 为学习率参数，用于控制损失值的大小，如式(3)所示。

$$\mu=\ln\left(\frac{|\alpha-\alpha^{gt}|}{\alpha^{gt}+\pi}+1\right)+\ln\left(\frac{|\beta-\beta^{gt}|}{\beta^{gt}+\frac{\pi}{2}}\right)+1 \tag{2}$$

$$b=\frac{k\mu}{1-IoU+a\vartheta+\mu} \tag{3}$$

3.3　旋转、倾斜卷积层与池化层

旋转、倾斜卷积层初始化、迭代计算过程是在卷积计算加入锁定矩阵 Lock，记为 0 的锁定区域不参与运算，其余部分则正常参与卷积运算。

3.3.1　初始化过程

受 R-RPN 模型的 RRoI Pooling 思路的启发，按不同维度分别使用倾斜卷积核与旋转卷积核进行运算，提取倾斜/旋转特征[7]。在卷积核尺寸为 N、旋转角度为 α 时进行旋转卷积运算，锁定矩阵 Lock 的初始化过程如算法 2 所示。

算法 2　旋转锁定矩阵 Lock 的初始化

输入：(N,α)

输出：(Lock) / * 旋转卷积层锁定矩阵 * /

1. A=round(N/1+cot(α))，B=round(N/1+tan(α))，k=A/B

2. for i in 1 to N do

3. for j in 1 to N do / * 确定计算区域 * /

4. 若(i≥−k * j+A and i≤−k * j+k * N+Bandi≤1/k * j+Aandi≥1/k * j−B/k)，则 Lock[i][j]=1；否则 Lock[i][j]=0。

在卷积核尺寸为 N、倾斜角度为 β 时进行倾斜卷积运算，锁定矩阵 Lock 的初始化过程如算法 3 所示。

算法 3　倾斜锁定矩阵 Lock 的初始化

输入：(N,β)

输出：(Lock) / * 倾斜卷积层锁定矩阵 * /

1. A=cot(β)，k=N / A

2. for i in 1 to N do

3. for j in 1 to N do //确定计算区域

4. 若(i≥−k * j+k * A and i≤−k * j+k * N)，则 Lock[i][j]=1；否则 Lock[i][j]=0。

类似地，根据池化核尺寸与旋转/倾斜角度确定池化核运算时的计算区域。

3.3.2　迭代过程

算法 4　卷积迭代流程

输入：(Picture,N,Lock) / * 旋转、倾斜卷积 * /

1. for i in 0 to Picture[cols]−N do

2. for j in 0 to Picture[rows]−N do

3. Pict_mat=Picture[i:i+N−1][j:j+N−1]/ * 卷积区域 * /

4. for k in 1 to N do / * 按行进行卷积运算 * /

5. for l in 1 to N do

6. 确定 Lock 当前行两侧空值 n1,n2 与运算区域 nc

7. end

8. 计算区域(单行)comp=Pict_mat[k][n1:N−n2−2]

9. feature_line[k][j]=Conv(comp,1 * nc)//卷积运算

10. end

11. end

12. for j in 0 to N−1 do

13. for k in 0 to N−1 do

14. feature_map[i][j]=feature_map[i][j]+feature_line[k][j]

15. / * 按列卷积结果将对应列相加，生成特征图对应行 * /

16. end

17. end

18. end

类似地，旋转、倾斜池化层使用锁定矩阵，对池化核进行切割，生成旋转(倾斜)池化核，进行池化操作时仅对锁定矩阵为 1 的区域进行池化操作。

3.4　字符仿射分割算法

由于单个字符有闭合轮廓，且所有字符形状相似，锚定框可以更好地匹配每个字符形状，增强识别的准确率。利用仿射变换思想，将文本拆分成单个字符输入模型[15]。按比例根据字符长度分割文本区域，计算每个字符位置在图像中的真实位置，建立标签数据集。字符仿射分割算法的示例如图 2 所示，首先根据锚定框的宽高比确定字符串形式，若宽高比大于 1 则为横版，小于 1 则为竖版字符串。然后根据文本字符

数量及文本方向，对文本区域进行按比例分割，根据字符串长度与左右两端比例计算字符长度比例 k，根据 k 与上下两边长度计算每个字符的分割位置。

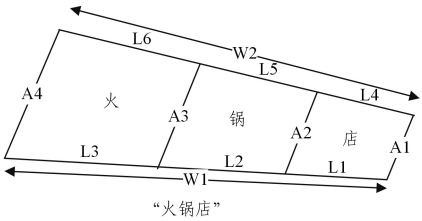


图 2 字符仿射分割算法示意图

Fig. 2 Schematic diagram of character affine segmentation algorithm

3.5 R-YOLOv5 模型

如图 3 所示，R-YOLOv5 模型在 YOLOv5 4.0 模型上添加了 R_CSPResBlock，T_CSPResBlock，R_SPP，T_SPP 模块，新结构相比原结构特征提取层更深，并加强了对不同角度的特征的提取效果。在原模型 Backbone 主干特征提取网络的每个 CSPResBlock 模块后加入旋转 CSP 残差模块与倾斜 CSP 残差模块，对旋转特征与倾斜特征进行表层提取。由于旋转、倾斜卷积层的运算量相比普通卷积层的运算量成倍数增长，为不影响 YOLOv5 整体架构的运算效率，仅在主干网络做旋转倾斜特征提取卷积计算。在进入特征金字塔结构前加入旋转 SPP 结构与倾斜 SPP 结构叠加模块，由于池化层多次重复使用体现出的旋转不变性，在模型中仅使用一次即可。

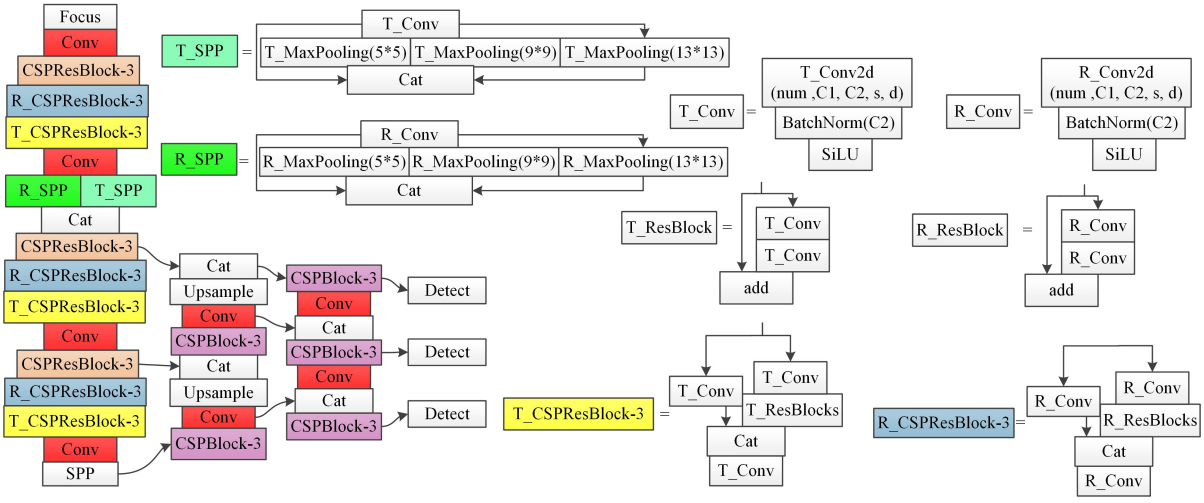


图 3 R-YOLOv5 模型的结构图

Fig. 3 R-YOLOv5 structure

在原 YOLOv5 模型的超参数中加入 angle_disfunc(角度序列分布形式，如均匀分布、正态分布、T 分布等)、rotate_num(旋转角度分割数)、tilt_num(倾斜角度分割数)、anchor_rotate(七元锚定框旋转角度个数)、anchor_tilt(七元锚定框倾斜角度个数)。通过传入 rotate_num 与 tilt_num 控制旋转卷积层、池化层与倾斜卷积层、池化层的角度个数，每个角度生成一个卷积核或池化核。将多个不同参数的旋转池化层或倾斜池化层输出结果进行叠加(分别为 5*5，9*9，13*13 池化层与原特征图)，形成多通道特征图。

YOLOv5 模型中的锚定框与损失函数被替换为前面提出的七元锚定框与 RIoU 损失函数，传入参数 anchor_rotate 与 anchor_tilt 分别控制七元锚定框旋转角度个数与倾斜角度个数，生成 anchor_rotate 与 anchor_tilt 对应的角度序列，两序列中所有角度两两组合，生成 anchor_rotate * anchor_tilt 个角度组合。在原模型中每个锚定框用 anchor_rotate * anchor_tilt 个锚定框代替，每个锚定框由其对应的角度组合进行初始化。在迭代中同样利用自适应锚定框思路，将旋转角度与倾斜角度加入锚定框的调整参数计算中，使用聚类算法选出较优的锚定框角度参数。

R-YOLOv5 模型锚定框数是 YOLOv5 模型的 anchor_rotate * anchor_tilt 倍，模型的检测速度为 YOLOv5 模型的 $\frac{1}{anchor\_rotate * anchor\_tilt}$。

4 实验及结果分析

4.1 数据预处理

本文使用 ICDAR2019-LSVT 中的全标注街景图像进行模型检验，按 7:3 比例进行训练集、测试集的划分。将 ICDAR2019 数据集的格式转化为 R-YOLOv5 使用的七元数据集格式，目标类型统一记为“text”。

4.2 评价指标

mAP 与 F-Score 是两种常见的目标检测算法评价指标。根据不同置信度阈值绘制 PR 曲线，计算曲线与坐标轴包围出的图形的面积 AP 值，并对各类计算平均值，即为 mAP 值。mAP@.5 与 mAP@.5:.95 为目前较为常用的两个指标，mAP@.5 为置信度阈值为 50%以上的 mAP 值，mAP@.5:.95 即为置信度阈值为 50%～95%的 mAP 值。

F-Score 本质上为准确率与召回率的加权调和平均数，Macro-F-Score 为各个类别的 F-Score 的算数平均值。常用的 F-Score 为 F.5，F1，F2，其中召回率的权重分别是准确率的 0.5 倍、1 倍与 2 倍。根据对召回率与准确率的不同偏好，选择不同指标。

4.3 模型训练过程

YOLOv5 模型的训练结果如表 1 所列。mAP@.5 与 mAP@.5:.95 在第 2 轮达到峰值，为 31.39%与 11.63%，后波动性降低。相比 mAP 值，F-Score 则从第 1 轮到第 5 轮

逐步提高，在第 5 轮达到峰值，因此训练在第 6 轮终止。

R-YOLOv5 训练结果如表 2 所列，在第 7 轮时 F-score 出现波动，因此轮训练终止在第 8 轮。

表 1　YOLOv5 模型的训练结果

Table 1　YOLOv5 training results

Epo	mAP.5	mAP.5:.95	F.5	F1	F2
1	0.2658	0.0898	0.2186	0.2791	0.3858
2	0.3139	0.1163	0.2659	0.3307	0.4374
3	0.2666	0.0904	0.3025	0.3572	0.4361
4	0.2138	0.0689	0.2996	0.3527	0.4287
5	0.2401	0.0800	0.3118	0.3685	0.4503
6	0.2166	0.0700	0.3046	0.3599	0.4397

表 2　R-YOLOv5 模型的训练结果

Table 2　R-YOLOv5 training results

Epo	mAP.5	mAP.5:.95	F.5	F1	F2
1	0.2572	0.0610	0.2377	0.2946	0.3873
2	0.3299	0.0945	0.2727	0.3329	0.4273
3	0.3966	0.1359	0.2931	0.3547	0.4490
4	0.4332	0.1607	0.3171	0.3783	0.4689
5	0.4473	0.1716	0.3291	0.3903	0.4794
6	0.4593	0.1807	0.3371	0.3985	0.4874
7	0.4607	0.1804	0.3362	0.3981	0.4881
8	0.4608	0.1819	0.3391	0.4005	0.4889

损失值的变化趋势从侧面反映了训练过程，两个模型训练过程的损失值如表 3 所列。由于 R-YOLOv5 模型对单个字符进行了检测，增加了目标数量与检测难度，相比原模型训练难度提高了很多，训练所需时间大幅增加，目标置信度损失值持续保持在高位，且远高于原模型对整个文本行进行检测。由于模型改进与字符仿射分割共同作用，R-YOLOv5 的检测框损失值(Box_Loss)与目标置信度损失值(Obj_Loss)均高于 YOLOv5，RIoU 检测框损失值从第 1 轮到第 4 轮缓慢降低，随后保持不变。从第 4 轮开始，锚定框基本已经定型，主要检测难点在于目标置信度的计算。

表 3　YOLOv5 与 R-YOLOv5 模型训练过程损失值对比表

Table 3　Comparison of loss values of YOLOv5 and R-YOLOv5 models during training

Epo	CIoU_BoxLoss	YOLOv5_ObjLoss	YOLOv5_TotLoss	RIoU_BoxLoss	RYOLO_ObjLoss	RYOLO_TotLoss
1	0.09450	0.07077	0.16527	0.11640	0.15350	0.26990
2	0.07630	0.07598	0.15228	0.10610	0.16390	0.27000
3	0.07326	0.07392	0.14718	0.10380	0.16200	0.2658
4	0.06887	0.07110	0.13997	0.10120	0.15880	0.2600
5	0.06604	0.07015	0.13619	0.10050	0.15830	0.25890
6	0.06396	0.06920	0.13316	0.09939	0.15740	0.25680
7	—	—	—	0.09853	0.15530	0.25380
8	—	—	—	0.09875	0.15660	0.25540

本文改进的模型总损失值持续保持在高位，训练持续进行，侧面解释了从第 1 轮到第 8 轮检测效果不断提升的原因。而 YOLOv5 总损失值持续缓慢降低，训练不断放慢，相应地从检测效果评价标准 mAP 来看，第 2 轮开始检测效果的波动性下降。

4.4　YOLOv5 与 R-YOLOv5 测试结果对比分析

利用测试数据集对前面训练好的 R-YOLOv5 和 YOLOv5 进行测试，结果如表 4 所列。

表 4　模型测试结果

Table 4　Models testing results

Model	Precision/%	Recall/%	mAP@.5/%	mAP@.5:.95/%	Macro-F.5/%	Macro-F1/%	Macro-F2/%	FPS
SegLink	71.03	19.02			45.92	30.01	22.28	0.85
EAST	75.67	19.23			47.68	30.67	22.60	0.56
YOLOv5	23.51	55.73	31.39	11.63	26.58	33.07	43.74	11.11
字符分割＋YOLOv5	30.43	57.00	45.64	17.93	33.56	39.68	48.53	10.4
R-YOLOv5	30.77	57.33	46.08	18.19	33.91	40.05	48.89	10.4

从测试结果可以看出，R-YOLOv5 模型的检测效果比 YOLOv5 模型有很大的进步，由于字符仿射分割模型将所有文本区域切分为单个字符进行训练，检测结果中每个文本框均为一个字符，有效减少了检测框中的非文本区域，同时由于对模型结构的改进，增强了特征提取能力，检测效果有显著提高，输出的检测框包含的非文本区域明显减少(见图 3)。

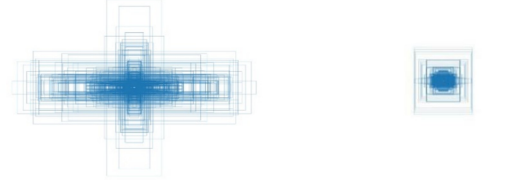
YOLOv5 由于文本目标具有开放轮廓的特点，目标的形状(宽高比)较为离散，因此锚定框形状也十分离散，以固定形状的锚定框拟合整行文本有很大难度。R-YOLOv5 中经过字符仿射分割处理，目标为单个字符，每个文字形状大致相似，都是近似方形的形状，锚定框可以很好拟合文字形状，锚定框形状较为集中，除了少量锚定框是长条形外，绝大多数锚定框均为高度略长于宽度的矩形框，高宽比趋近于 1(见图 4)。YOLOv5 与 R-YOLOv5 锚定框对比图如图 5 所示。

增加旋转角度与倾斜角度的锚定框可以加强检测准确率，代价则是极大地降低训练速度与检测速度。并且对于检测框旋转或倾斜角度的设计以及时间消耗与检测精度的取舍，仍需要根据具体场景具体分析，设计角度分布函数以达到合适的平衡点。

(a) YOLOv5　(b) R-YOLOv5

图 4　YOLOv5 与 R-YOLOv5 检测结果对比图

Fig. 4　Comparison of YOLOv5 and R-YOLOv5 testing results



(a) YOLOv5　(b) R-YOLOv5

图 5　YOLOv5 与 R-YOLOv5 锚定框对比图

Fig. 5　Comparison of anchor boxes of YOLOv5 and R-YOLOv5

结束语 本文以 YOLOv5 为基础搭建文本检测模型，针对文本目标的特点，提出了 R-YOLOv5 模型，融合了旋转锚定框、旋转倾斜卷积层与池化层和字符仿射分割算法，实验结果显示，R-YOLOv5 的改进策略对提升检测效果有很大帮助。但 R-YOLOv5 模型在一定程度上增加了训练难度，对于某些手写文本与模糊部分无法进行有效识别，且也会出现将杂乱的区域识别为文本的情况。未来将进一步对旋转区域检测进行探索，在此基础上进行优化或提出其他解决方案。

参 考 文 献

[1] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]// IEEE Conference on Computer Vision and Pattern Recognition(CVPR). Las Vegas: IEEE, 2016: 779-788.

[2] REDMON J, FARHADI A. YOLO9000: Better, Faster, Stronger [C]// IEEE Conference on Computer Vision Pattern Recognition(CVPR). Honolulu: IEEE, 2017: 6517-6525.

[3] REDMON J, FARHADI A. YOLOv3: An Incremental Improvement [EB/OL]. [2018-04-08]. https://arxiv.org/abs/1804.02767.

[4] BOCHKOVSKIY A, WANG C Y, LIAO H. YOLOv4: Optimal Speed and Accuracy of Object Detection [EB/OL]. [2020-04-23]. https://arxiv.org/abs/2004.10934.

[5] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single Shot MultiBox Detector[C]// European Conference on Computer Vision. Cham: Springer, 2016: 21-37.

[6] 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.

[7] TIAN Z, HUANG W, HE T, et al. Detecting Text in Natural Image with Connectionist Text Proposal Network[C]// European Conference on Computer Vision. Cham: Springer, 2016: 56-72.

[8] ZHOU X, YAO C, WEN H, et al. EAST: An Efficient and Accurate Scene Text Detector[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition(CVPR). Honolulu: IEEE, 2017: 2642-2651.

[9] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature Pyramid Networks for Object Detection[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition(CVPR). Honolulu: IEEE, 2017: 936-944.

[10] LIU S, QI L, QIN H, et al. Path Aggregation Network for Instance Segmentation[C]// IEEE Conference on Computer Vision Pattern Recognition(CVPR). Salt Lake City: IEEE, 2018: 8759-8768.

[11] MISRA D. Mish: A Self Regularized Non-Monotonic Neural Activation Function [EB/OL]. [2019-08-23]. https://arxiv.org/abs/1908.08681v1.

[12] ZHENG Z, WANG P, LIU W, et al. Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression [J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 12993-13000.

[13] KHALIL A, JARRAH M, AL-AYYOUB M, et al. Text detection and script identification in natural scene images using deep learning[J]. Computers & Electrical Engineering, 2021, 91(C): 107043.

[14] KESERWANI P, DHANKHAR A, SAINI R, et al. Quadbox: Quadrilateral Bounding Box Based Scene Text Detection Using Vector Regression[J]. IEEE Access, 2021, 9: 36802-36818.

[15] YANG W J, ZOU B J, LI K W, et al. A Character Flow Framework for Multi-Oriented Scene Text Detection[J]. Journal of Computer Science & Technology, 2021, 36(3): 465-477.

[16] NAGAOKA Y, MIYAZAKI T, SUGAYA Y, et al. Text Detection Using Multi-Stage Region Proposal Network Sensitive to Text Scale[J]. Sensors, 2021, 21(4): 1232.

[17] ZHAO X, ZHOU Z, LI L, et al. Scene Text Detection Based On Fusion Network[J]. International Journal of Pattern Recognition and Artificial Intelligence, 2021, 35(10): 2153005.

[18] LIU C Y, CHEN X X, LUO C J, et al. Deep learning methods for scene text detection and recognition [J]. Journal of Image and Graphics, 2021, 26(6): 1330-1367.

[19] LI H, WANG X L, XIANG X G. Scene Text Detection Based on Triple Segmentation[J]. Computer Science, 2020, 47(11): 142-147.

[20] YUAN X X, WU Q. Object Detection in Remote Sensing Images Based on Saliency Feature and Angle Information[J]. Computer Science, 2021, 48(4): 174-179.

[21] GONG F M, LIU F H, LI J J, GONG W J. Scene Text Detection and Recognition Based on Deep Learning[J]. Computer Systems & Applications, 2021, 30(8): 179-185.

[22] LIU Y J, YI X H, LI Y G, et al. Application of Scene Text Recognition Technology Based on Deep Learning: A Survey[J]. Computer Engineering and Applications, 2022, 58(4): 52-63.

[23] SHAO H L, JI Y, LIU C P, et al. Scene Text Detection Algorithm Based on Enhanced Feature Pyramid Network[J]. Computer Science, 2022, 49(2): 248-255.

[24] WANG F, HUANG J, WEN H W. Fast Text Detection Based on Improved YOLOv3 [J]. Telecommunication Engineering, 2022, 62(1): 130-137.

[25] LEI X T, HU J. Text center pixel reconstruction to achieve efficient arbitrary shape text detection[J/OL]. Computer Engineering and Applications: 1-11. [2022-02-25]. http://kns.cnki.net/kcms/detail/11.2127.TP.20220217.1622.006.html.

[26] CHEN P, LI M, ZHANG Y, WANG Z P. An End-to-End Natural Scene Text Detection and Recognition Model[J/OL]. Measurement & Control Technology: 1-7. [2022-02-25]. https://kns.cnki.net/kcms/detail/detail.aspx?doi=10.19708/j.ckjs.2021.10.276.

[27] SUN G M, GUAN S K, LI Y, et al. Handwritten text detection on test paper using improved CTPN algorithm[J]. Information Technology, 2020, 44(9): 94-98.

[28] CHEN M M, XU J H. Scene text detection model based on high resolution convolutional neural networks[J]. Computer Applications and Software, 2020, 37(10): 138-144.

[29] LIU Y, WEN J. Complex Scene Text Detection Based on Attention Mechanism[J]. Computer Science, 2020, 47(7): 135-140.

[30] KOU X C, ZHANG H R, FENG J, et al. Distortion Correction Algorithm for Complex Document Image Based on Multi-level Text Detection[J]. Computer Science, 2021, 48(12): 249-255.

RAN Yu, born in 1999, postgraduate. His main research interests include deep learning and object detection.

ZHANG Li, born in 1972, Ph.D, professor. Her main research interests include machine learning, deep learning, business intelligence and etc.