

DOI:10.11918/202212061

基于特征过滤和自适应融合机制的不规则场景文本检测算法

杨帅磊¹, 李岐龙¹, 陈杰¹, 凡高娟¹, 张重生^{1,2}

(1. 河南省大数据分析与管理重点实验室(河南大学), 河南 开封 475001;

2. 黄河文化遗产实验室(河南大学), 河南 开封 475001)

摘要: 自然场景图像中的文本存在背景复杂、形状各异、方向多样、光线多变等特点, 为提升场景文本尤其是不规则场景文本的检测性能, 提出一种基于特征过滤和自适应特征融合机制的不规则场景文本检测网络 FGANet (feature guided adaptive network), 其特征聚合引导模块通过空洞卷积扩大感受野、提升网络的特征表达能力, 其自适应特征融合模块能够将深层的语义信息与浅层的细节信息进行综合利用, 使算法获得更强的文本感知能力。结果表明, 在场景文本检测方面, FGANet 相较于对比算法, 在 ICDAR2015、CTW1500、MSRA-TD500 和 TotalText 4 个基准数据集上的 F 值分别提升了 2.4%、1.3%、1.8% 和 1.4%, 性能提升较为显著。

关键词: 自然场景图像; 场景文本检测; 特征过滤; 自适应特征融合; 文本感知能力

中图分类号: TP183

文献标志码: A

文章编号: 0367-6234(2025)10-0165-06

Irregular scene text detection based on feature filtering and adaptive fusion mechanisms

YANG Shuailei¹, LI Qilong¹, CHEN Jie¹, FAN Gaojuan¹, ZHANG Chongsheng^{1,2}

(1. Henan Key Lab of Big Data Analysis and Processing (Henan University), Kaifeng 475001, Henan, China;

2. Lab of Yellow River Heritages (Henan University), Kaifeng 475001, Henan, China)

Abstract: Text in natural-scene images often present characteristics of complex backgrounds, varied shapes, multiple orientations and changing illumination. In order to improve detection performance for scene text, particularly irregular text, we propose the feature guided adaptive network (FGANet), an irregular-scene text detection network based on feature filtering and adaptive fusion mechanisms. In specific, FGANet designs a module that utilizes dilated convolution to enlarge the receptive and enhance the network's feature representation capability. Its adaptive feature fusion module integrates deep semantic information with shallow detailed information, enabling stronger text-awareness. Experiments results show that for scene text detection, FGANet achieves notable improvements in F -score over comparative methods on four benchmark datasets: ICDAR2015, CTW1500, MSRA-TD500, and TotalText, with gains of 2.4%, 1.3%, 1.8%, and 1.4%, respectively.

Keywords: scene text images; scene text detection; feature filtering mechanism; adaptive feature fusion mechanism; text perception ability

自然场景文本检测是计算机视觉的重要研究方向之一, 旨在自然场景图像中自动定位文本区域, 在无人驾驶、票据识别等场景中具有广泛的应用^[1]。传统的自然场景文本检测算法利用人工设计的特征提取图像特征, 当场景文本图像背景复杂、形状多变、光线多变、方向多变时^[2], 此类算法在检测性能方面存在明显的不足。随着深度学习技术 (deep learning, DL) 的兴起和繁荣, 基于深度学习的目标检

测技术和图像分割技术得到了发展。

现有的场景文本检测算法可划分为基于回归的方法和基于分割的方法两种类型。基于回归的方法, 其思想类似于通用的目标检测任务。此类场景文本检测算法, 多在 Faster-RCNN^[3] (faster region with CNN features) 或 SSD^[4] (single shot multibox detector) 等经典目标检测神经网络的基础上, 结合场景文本图像的特点进行针对性的改进, 如将 3×3

收稿日期: 2022-12-19; 录用日期: 2023-04-18; 网络首发日期: 2023-12-14

网络首发地址: <https://link.cnki.net/urlid/23.1235.T.20231214.1038.009>

基金项目: 教育部人文社会科学研究一般项目(23YJAZH210); 河南省科技攻关项目(232102211021)

作者简介: 杨帅磊(1996—), 男, 硕士研究生; 张重生(1982—)男, 教授, 博士生导师

通信作者: 凡高娟, fangaojuan@henu.edu.cn

的正方形卷积核修改为 1×5 和 5×1 的高纵横比的卷积核,然后对目标框的 4 个顶点坐标进行回归学习^[5]。然而,当场景文本形状不规则或扭曲时,此类方法不能准确定位场景文本的边界。基于分割的方法,主要思路是将场景文本当作物体进行分割,先使用语义分割方法将文本区域作为前景分割出来,再通过后处理的方式生成检测框。现有的很多场景文本分割算法如 EAST^[6],都是在全卷积网络 FCN (fully convolutional networks)^[7] 的基础上进行改进的。

Wang 等^[8]提出一种基于语义分割的场景文本检测算法 PANNet (pixel aggregation network),通过像素聚合实现场景文本检测的后处理步骤,得到最终的文本实例框。然而,该算法在特征提取阶段,由于正方形卷积核的感受野受限,无法很好地解决变长文本的检测问题。本文提出提升特征通道在特征图中的影响,并解决变长文本的特征与卷积核感受野形状不匹配问题的新思路。设计特征聚合引导机制,对特征图中的通道进行滤除的同时进行增强,提高特征质量;设计自适应特征融合机制,基于动态联合注意力增强机制,对不同尺度特征在空间与通道上赋予不同的权重,提升多尺度特征融合效果。并通过大规模实验验证了所提方法的有效性。

1 相关工作

1.1 基于边框回归的方法

受到目标检测算法的启发,一些研究者结合场景文本图像的特点,提出了一系列优秀的场景文本检测算法。如 Tian 等^[9]在两阶段算法 Faster-RCNN 的基础上提出了 CTPN (connectionist text proposal network) 算法,该算法先通过分治法的策略检测到一连串的小尺度文本框,然后,将上述小尺度文本框通过双向循环神经网络融合成一个完整的文本行。

Shi 等^[10]则在单阶段检测算法 SSD 的基础上,额外考虑了旋转角度的影响,利用角度信息的引导算法实现对不同方向文本的检测。Liao 等^[11]针对文本锚框分布不均的问题提出了 TextBoxes 算法,该算法把正方形卷积核调整为矩形卷积核,同时修改了默认锚框,提升了长文本的检测效果。文献[12]提出回归短边顶点文本检测方法,通过减少对感受野大小的依赖,使得在长文本行的检测效果上更为准确。

1.2 基于像素分割的方法

受到像素分割算法的启发,一些研究者基于分割后的前景区域预测文本边框,解决了不规则文本的检测问题。如 Zheng 等^[13]首次使用 FCN 网络对场景文本进行检测,思路是在预测出的文本区域分割图中利用 MSER 模块 (maximally stable extremal regions) 从这些文本区域中筛选候选字符,然后,连接字符区域组合成文本实例。Deng 等^[14]提出了一种基于分割图检测文本的 PixelLink 算法,该算法预测特征图上的每个像素点及其相邻像素之间 8 邻域的连通关系,然后,通过像素之间的关系组合成不同文本实例。Wang 等^[15]采用渐进尺度扩张的手段预测任意形状的文本框,该算法核心是使用不同尺度的文本核,采用模拟逐步扩展的方式,将相邻文本进行像素级别的分离,最后得到文本的检测结果。Xu 等^[16]通过预测不同文本实例的像素与像素之间的相对位置提出了 TextField 算法,利用此相对位置分离文本实例,最后,基于形态学的后处理实现文本实例的检测。

2 方法的提出

为更好地检测形状不规则、长度多变等情况的自然场景文本图像,提出了文本检测模型 FGANet,如图 1 所示。

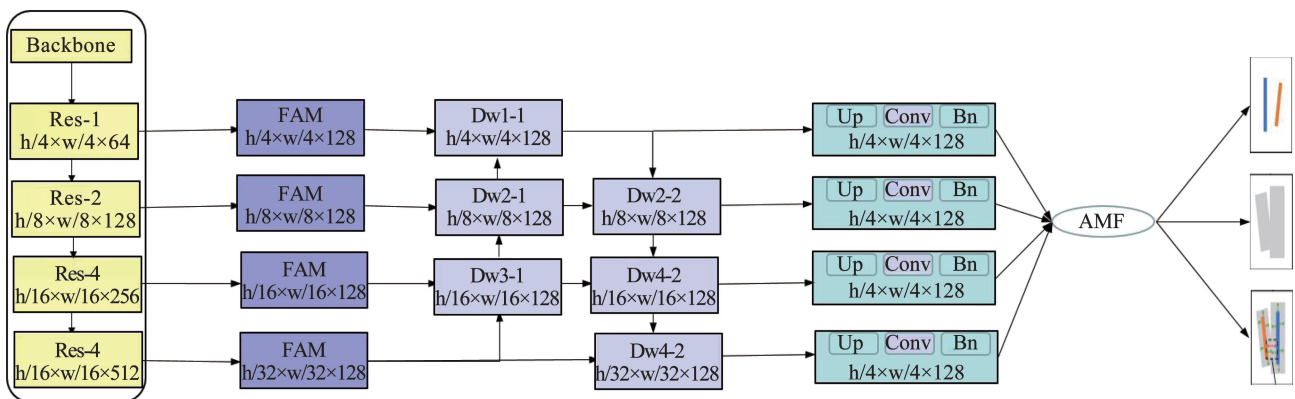


图 1 FGANet 整体网络架构

Fig. 1 Network architecture of FGANet

2.1 整体框架

FGANet 的骨干网络选择采用经典的 18 层残差神经网络 ResNet, 该网络负责对输入图像进行特征提取, 提取过程采用逐级特征提取策略, 将 ResNet 中间层产生的多尺度特征图 (F_1, F_2, F_3, F_4) 同时输出。输出的多尺度特征图 (F_1, F_2, F_3, F_4) 的通道维度数量分别为 64, 128, 256, 512。为了进行多尺度特征通道对齐, 需将不同尺度特征的通道维度数量保持一致。考虑到 ResNet-18 网络具有较弱的特征提取能力, 本文将利用特征聚合引导模块, 在尽量不损害原有特征的情况下, 减少或者增加通道数, 将多尺度特征图的通道对齐到 128 维。特征金字塔增强和融合网络负责对上述的多尺度特征进一步增强输出, 输出后的特征图经过上采样至 1/4 尺度, 将多个 1/4 尺度的特征进行拼接。但是简单的拼接操作, 不能有效地利用不同尺度的特征。为此, 引入自适应特征融合模块, 该模块通过计算不同尺度特征在特征融合中的重要程度, 生成语义特征比较丰富的特征图 F 。最终, 经过像素聚合 (PA) 方法生成最后的文本实例框。

2.2 特征聚合引导机制

特征聚合引导机制 (feature aggregation mechanism, FAM) 通过引入空洞卷积增大感受野的操作, 在有效减少通道数的同时增加感受野, 进一步弥补网络特征表达能力不强带来的缺陷。如图 2 所示, 初始的骨干网络提取的特征感受野较小, 减弱了特征图中的语义信息, 需要使用 FAM 模块进行语义增强。图 2 中, 输入 (Input) 是骨干网络提取的特征, 输出 (Output) 是特征图过滤后的输出特征。其结构共分为 4 条支路, 其中, P_1, P_2 和 P_3 是 3 个普通卷积和空洞卷积级联组成的, 普通卷积采用 1×1 与 3×3 的卷积核, 空洞卷积则依次采用 1, 3, 5 的空洞系数控制感受野的大小。同时, 为了加快模型的收敛速度, 每个卷积层添加归一化处理。

聚合引导机制/模块的计算表达可表示为

$$P_1 = D_1(C_1(I)) \quad (1)$$

$$P_2 = D_3(C_3(I)) \quad (2)$$

$$P_3 = D_5(C_5(I)) \quad (3)$$

$$P_4 = C_1(I) \quad (4)$$

$$O = \zeta(P_1, P_2, P_3, P_4) \quad (5)$$

式中: I 表示输入特征图; C 表示标准卷积, 下标分别表示卷积核的大小为 1, 3, 5; D 表示空洞卷积、归一化处理层和激活函数的非线性组合, 下标分别表示膨胀率为 1, 3, 5; ζ 表示通道拼接操作; O 表示最终输出的特征图。

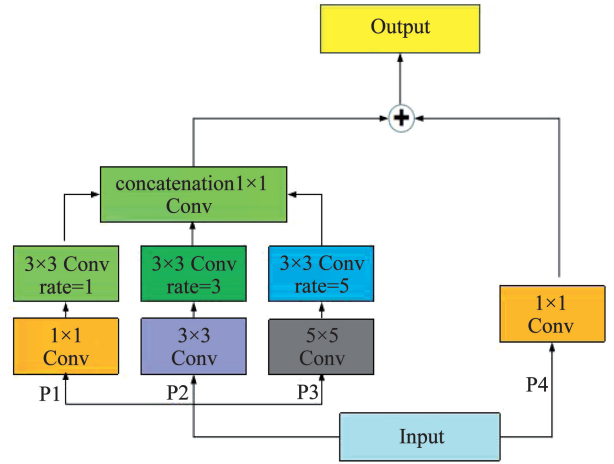


图 2 FAM 特征聚合引导机制的神经网络架构

Fig. 2 Neural network architecture of FAM mechanism

2.3 自适应特征融合机制

文本感知能力的强弱和不同层级网络提取特征的信息相关, 浅层网络提取的大尺度特征图更关注小文本实例的细节信息, 而深层网络提取的小尺度特征图能够捕获大文本实例的语义信息。将不同尺度的特征进行融合, 有利于提高网络的检测性能。

为此, 提出了自适应特征融合机制 (adaptive multi-scale feature fusion mechanism, AMF), 该机制/模块能够自适应地融合不同尺度的特征, 将深层的语义信息与浅层的细节信息进行有机整合, 减小冗余信息带来的影响。如图 3 所示, 设计的 AMF 模块主要包含两个分支, 上部分支负责实现对不同尺度特征图的信息交互, 计算出每个尺度对应的权重因子; 下部分支负责每个特征图与相对应的权重因子相乘, 拼接在一起得到语义丰富的特征图, 该模块的计算方式可以表示为

$$F_1 = C_3(\zeta(f_1, f_2, f_3, f_4)) \quad (6)$$

$$A_i = \text{Attention}(F), i = 1, 2, 3, 4 \quad (7)$$

$$F_2 = \zeta(A_i * f_i), i = 1, 2, 3, 4 \quad (8)$$

式中: F_1 代表 4 个特征图拼接卷积后的特征, ζ 表示通道拼接操作, C_3 表示 3×3 标准卷积, A_i 表示权重因子, F_2 表示该模块的最终输出特征图。

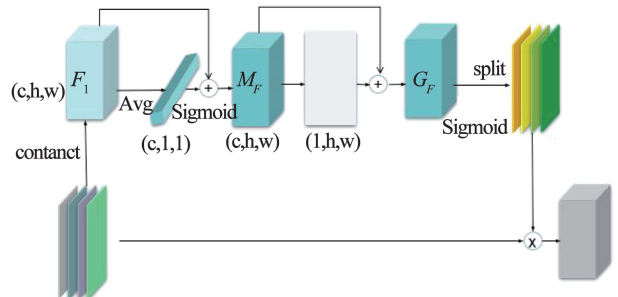


图 3 AMF 机制神经网络结构

Fig. 3 Neural network architecture of AMF mechanism

A_i 的计算方式如下:假设对于给定的输入特征图 F , 首先, 采用全局平均池化操作对空间信息进行压缩得到语义丰富的通道信息, 接着通过卷积激活对通道信息进行融合, 融合后的特征经 Sigmoid 激活函数得到每个通道的权重, 同时, 利用该权重与输入特征图 F 进行像素相加得到校准特征图 M_F 。然后, 对校准特征图 M_F 在通道上进行平均操作得到均值特征图, 此时特征图的通道数量为 1, 将均值特征图和校准特征图 M_F 进行像素相加操作一样得到全局特征图 G_F 。最后, 对得到的全局特征利用 1×1 卷积进行通道降维, 降维后的通道数量为 4, 每个通道则对应每个输入的特征图, 降维后的 4 通道特征图经 Sigmoid 函数得到每个尺度的权重 A_i , 此时, A_i 表示每个尺度特征的重要程度。

$$M_s = \sigma(C_1(C_3(\text{AvgPool}(F_1)))) \quad (9)$$

$$M_F = M_s + F_1 \quad (10)$$

$$G_F = \text{Mean}(M_F) + M_F \quad (11)$$

$$A_i = \delta(C_1(\delta(G_F))) \quad (12)$$

式中: AvgPool 表示全局平均化; δ 表示 ReLu 激活函数; σ 表示 Sigmoid 激活函数; C 表示标准卷积, 下标分别表示卷积核的大小为 1, 3; A_i 表示最后的 4 个尺度特征权重。

2.4 像素聚合模块

在文本检测的后处理阶段引入像素聚合模块可以有效对文本区域进行检测, 主要是借鉴聚类的思想, 对预测图中的文本核重建一个完整的文本实例。在该模块中, 假设聚类中心是文本核, 聚类样本是文本像素, 像素聚合模块的目的就是将文本像素聚合到对应的文本核, 然后构成一个完整的文本实例。像素聚合算法的具体过程如下:

- 1) 通过设置阈值大小, 与文本核中的像素进行比较, 判断像素的值是否大于阈值, 若大于阈值, 则当前像素属于核区域。
- 2) 遍历每个属于核区域中的像素, 采用上下左右方向(四联通)的方式对相邻像素进行聚类, 通过计算像素与该像素之间的相似度距离, 判断该相邻像素是否为该文本区域。
- 3) 不断重复上述步骤 2), 结束条件是找不到符合步骤 2) 的相邻文本像素。

3 结果及分析

3.1 评价指标

精度 (Precision, P)、召回率 (Recall, R) 和 F 值

是文本检测任务中常采用的评价指标。精度表示正确预测的样本数量与返回的样本总数的比率。召回率是正确预测的正类样本数量与数据集中所有正类样本数量的比率。在评估算法性能时, 精度和召回率难以同时提升, 故引入 F 值, 对两个指标进行权衡/折中评估, F 值是精度和召回率指标的加权平均值, 并用于评估目标检测算法的整体性能, 相关评估指标公式如下:

$$P = \text{TP} / (\text{TP} + \text{FP}) \quad (13)$$

$$R = \text{TP} / (\text{TP} + \text{FN}) \quad (14)$$

$$F = (2 \times P \times R) / (P + R) \quad (15)$$

式中: TP 表示被正确预测的正类样本数量, FN 表示被误判为负类的正样本数量, FP 表示被误判为正类的负样本数量。

3.2 ICDAR2015 实验结果分析

ICDAR2015 是常用的场景文本检测数据集之一。表 1 给出了本文设计的 FGANet 方法及基准方法 PANNet 在该数据集上的实验结果。此外, 还列出了其他经典文字检测模型的性能表现。

表 1 各方法在 ICDAR2015 数据集上的对比结果

Tab. 1 Comparison of different methods on ICDAR2015 %

方法	P	R	F
EAST	83.6	73.5	78.2
SegLink	73.1	76.8	75
PixelLink	82.9	81.7	82.3
TextSnake	84.9	80.4	82.6
PANNet	84.0	81.9	82.9
FGANet	90.6	80.6	85.3

通过表 1 可知, 本方法的精度为 90.6%, 召回率为 80.6%, F 值为 85.3%。与基准方法 PANNet 相比, FGANet 在精度上提高了 6.6%, F 值提高了 2.4%。与其他算法相比, FGANet 均获得了最佳性能。

3.3 MSRA-TD500 实验结果分析

MSRA-TD500 是一个中英文混合的场景文本检测数据集, 包含大量的多方向文本实例图像。表 2 给出了 FGANet 方法及基准方法 PANNet 在该数据集上的性能表现。可以看出, 本文提出的 FGANet 方法的精度为 90.8%, 召回率为 81.3%, F 值为 85.8%。相较于基准模型 PANNet, 精度提高了 6.4%, F 值提高了 1.3%, 提升较为显著。与其他文字检测算法相比, FGANet 也获得了最好的性能。

表 2 各方法在 MSRA-TD500 数据集上的对比结果

Tab. 2 Comparison of different methods on MSRA-TD500 %

方法	P	R	F
EAST	87.3	67.4	76.1
SegLink	86.0	70.0	77.0
PixelLink	83.0	73.2	77.8
TextSnake	83.2	73.9	78.3
PANNet	84.4	83.8	84.5
FGANet	90.8	81.3	85.8

3.4 CTW1500 实验结果分析

CTW1500 是一个曲形文本数据集, 里面有大量的弧形文本实例图像。表 3 给出了不同方法在该数据集上的检测性能表现。可以看出, FGANet 的精度达 88.2%, 召回率达 82.9%, F 值达 85.5%。与 PANNet 相比, FGANet 在精度提升的同时(1.8%), 召回率和 F 值也得到了提升, 分别提升了 1.7% 和 1.8%。相较于较新的 DB^[17] 文本检测方法, 本文所提方法在精度和 F 值方面仍有明显优势。

表 3 各方法在 CTW1500 数据集上的对比结果

Tab. 3 Comparison of different methods on CTW1500 %

方法	P	R	F
EAST	78.7	49.1	60.4
SegLink	42.3	40.0	40.8
PixelLink	82.9	81.7	82.3
TextSnake	67.9	85.3	75.6
PANNet	86.4	81.2	83.7
DB	80.2	86.9	83.4
FGANet	88.2	82.9	85.5

3.5 TotalText 实验结果分析

TotalText 是另一个经典的弯曲文本数据集, 表 4 给出了相关方法在该数据集上的文本检测性能表现。可以看出, FGANet 的精度达 92.2%, 召回率达 81.3%, F 值达 86.4%。与 PANNet 相比, FGANet 在精度、召回率和 F 值上分别提升了 2.9%、0.3% 和 1.4%。相较于较新的 DB^[17] 方法, FGANet 在精度和 F 值方面仍有明显优势。

表 4 各方法在 TotalText 数据集上的对比结果

Tab. 4 Comparison of different methods on TotalText %

方法	P	R	F
EAST	50.0	36.2	42.0
SegLink	30.3	23.8	26.7
PixelLink	—	—	—
TextSnake	82.7	74.5	78.4
PANNet	89.3	81.0	85.0
DB	82.5	87.1	84.7
FGANet	92.2	81.3	86.4

3.6 消融实验

首先, 选择 FAM 模块验证是否影响 PANNet 模型的整体性能, 验证结果如表 5 第 2 行所示, PANNet 增加了本文所提的 FAM 模块后, 其在 ICDAR2015 和 CTW1500 数据集上的性能均有提升, 精度分别提升了 1.9%、0.8%, F 值分别提升 1.2%、0.8%。因此, FAM 模块能够增强特征的信息, 提升神经网络的性能表现。其次, 对 AMF 进行消融实验, 验证结果如表 5 第 3 行所示, 当 PANNet 引入 AMF 模块后, 其性能也得到了提升, 在 ICDAR2015 和 CTW1500 数据集上精度分别提升 3.0%、0.6%, F 值分别提升 1.1%、1.2%, 因此, AMF 模块可以更加有效地融合不同尺度的特征, 增强文本区域特征并抑制背景噪声。如表 5 第 4 行所示, PANNet 模型综合使用 FAM 模块与 AMF 模块后, 能够得到更佳的 F 值, 即综合性能表现得到提升。

表 5 消融实验结果

Tab. 5 Ablation studies of different modules %

BackBone	FAM	AMF	ICDAR2015			CTW1500		
			P	R	F	P	R	F
ResNet			82.9	77.8	80.3	84.6	77.7	81.0
ResNet	✓		84.8	78.4	81.5	85.4	78.5	81.8
ResNet		✓	85.9	77.4	81.4	85.2	79.4	82.2
ResNet	✓	✓	88.0	78.3	82.9	85.7	80.5	83.0

4 结 语

针对文本形状不规则、图像背景复杂等挑战, 提出一个自然场景文本检测网络 FGANet, 该模型设计 FAM 模块和 AMF 模块, 能够在突出目标特征的同时能抑制无关的背景噪声, 且能够更精准地覆盖到目标文本区域。在多个公开数据集上的实验结果表明, 本文的方法相比其他文本检测算法, 在精度和 F 值上有较为显著的提升, 但召回率上升略有不足。因此, 在后续工作中, 将研究提升本文所提模型的召回率的方法。同时, 所提方法在对文本实例比较紧密连接一起时会表现出稍弱的检测性能, 未来也将研究如何将文字密集的场景文本实例进行更好的分离, 以进一步来提升模型的文本检测性能。

参考文献

- [1] 刘崇宇, 陈晓雪, 罗灿杰, 等. 自然场景文本检测与识别的深度学习方法[J]. 中国图象图形学报, 2021, 26(6): 1330
LIU Chongyu, CHEN Xiaoxue, LUO Canjie, et al. Deep learning method for text detection and recognition in natural scenes [J]. Journal of Image and Graphics, 2021, 26(6): 1330. DOI: 10.

- 11834/jig.210044
- [2] 王建新, 王子亚, 田莹. 基于深度学习的自然场景文本检测与识别综述[J]. 软件学报, 2020, 31(5): 1465
WANG Jianxin, WANG Ziya, TIAN Xuan. A review of natural scene text detection and recognition based on deep learning[J]. Journal of Software, 2020, 31(5): 1465. DOI:10.13328/j.cnki.jos.005988
- [3] REN Shaoqing, HE Kaiming, GIRSHICK R, et al. Towards real-time object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 39(6): 1137. DOI:10.1109/TPAMI.2016.2577031
- [4] LIU Wei, ANGUELOV D, DUMITRU E, et al. Single shot multibox detector[C]// Proceedings of the European Conference on Computer Vision. Springer: Cham, 2016; 21. DOI:10.1007/978-3-319-46448-0_2
- [5] LIAO Minghui, SHI Baoguang, BAI Xiang. A single-shot oriented scene text detector [J]. IEEE Transactions on Image Processing, 2018, 27(8): 3676. DOI:10.1109/TIP.2018.2825107
- [6] ZHOU Xiyu, YAO Cong, WANG Yuzhi, et al. An efficient and accurate scene text detector [C]// Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017; 5551. DOI:10.1109/CVPR.2017.283
- [7] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE, 2015; 3341. DOI:10.1109/CVPR.2015.7298965
- [8] WANG Wenhai, XIE Enze, SONG Xiaoge, et al. Efficient and accurate arbitrary-shaped text detection with pixel aggregation network [C]// Proceedings of the IEEE Conference on International Conference on Computer Vision. Seoul: IEEE, 2019; 8439. DOI: 10.1109/ICCV.2019.00853
- [9] 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. Springer: Cham, 2016; 56. DOI:10.1007/978-3-319-46484-8_4
- [10] 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. Honolulu: IEEE, 2017; 3482. DOI:10.1109/CVPR.2017.371
- [11] LIAO Minghui, SHI Baoguang, BAI Xiang, et al. A fast text detector with a single deep neural network [C]// Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence. San Francisco: AAAI, 2017; 4161. DOI:10.1609/aaai.v31i1.11196
- [12] 游洋彪, 石繁槐. 短边顶点回归网络: 新型自然场景文本检测器[J]. 哈尔滨工业大学学报, 2021, 53(12): 89
YOU Yangbiao, SHI Fankui. Short edge vertex regression network: a novel natural scene text detector [J]. Journal of Harbin Institute of Technology, 2021, 53(12): 89. DOI:10.11918/201908104
- [13] ZHENG Zhang, ZHANG Chengquan, WEI Shen, et al. Multi-oriented text detection with fully convolutional networks [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016; 4159. DOI: 10.1109/CVPR.2016.451
- [14] DENG Dan, LIU Haifeng, LI Xuelong, et al. PixelLink: detecting scene text via instance segmentation [C]// Proceedings of the AAAI Conference on Artificial Intelligence. New Orleans: ADAAI, 2018; 6773. DOI:10.1609/aaai.v32i1.12269
- [15] WANG Wenhai, XIE Enze, LI Xiang, et al. Shape robust text detection with progressive scale expansion network [C]// Proceedings of 2019 IEEE Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019; 9328. DOI:10.1109/CVPR.2019.00956
- [16] XU Yongchao, WANG Yukang, ZHOU Wei, et al. TextField: learning a deep direction field for irregular scene text detection [J]. IEEE Transactions on Image Processing, 2019, 28(11): 5566. DOI:10.1109/TIP.2019.2900589
- [17] LIAO Minghui, YAO Cong, BAI Xiang, 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. DOI:10.1109/TPAMI.2022.3155612

(编辑 刘彤)