

文章编号: 2095-2163(2023)08-0030-07

中图分类号: TP391

文献标志码: A

基于增大感受野的场景文本检测算法

杨 豪, 方承志, 罗佳辉, 雷 蕾

(南京邮电大学 电子与光学工程学院、柔性电子(未来技术)学院, 南京 210023)

摘要: 针对场景文本检测算法在长文本检测中的不足问题, 提出一种加入条带池化和优化损失函数的改进算法。针对文本行的长条形状, 引入 Strip Pooling, 这种池化核可以获得更大的感受野, 将带有 Strip Pooling 的 MPM (Mixed Pooling Module) 模块和 Coordinate Attention 模块插入 EAST 算法中, 使得像素点能够捕获条带位置上的远程坐标信息, 提升对长文本检测的性能。同时针对回归任务, 优化了损失函数, 使用 *Dice Loss* 和 *CloU Loss*, 得到更加精确的文本框。本文的模型在 ICDAR2013、ICDAR2015 和 MSRA-TD500 数据集上都取得了较好的检测效果。

关键词: 文本检测; EAST; 注意力机制; Strip Pooling; 金字塔池化

Scene text detection algorithm based on increasing receptive field

YANG Hao, FANG Chengzhi, LUO Jiahui, LEI Lei

(School of Electronic and Optical Engineering, School of Flexible Electronics (Future Technology), Nanjing University of Posts and Telecommunications, Nanjing 210023, China)

[Abstract] Aiming at the shortcomings of scene text detection algorithm in long text detection, an improved algorithm is proposed, which added strip pooling and optimized loss function. Strip Pooling is introduced for long bar shapes of text lines. This Pooling kernel can obtain a larger receptive field. The Mixed Pooling Module (MPM) and Coordinate Attention Module with Strip Pooling are inserted into the EAST algorithm. Thereafter, the pixel can capture the remote coordinate information on the strip position, and improve the performance of long text detection. At the same time, Loss function is optimized for regression task, and *Dice Loss* and *CloU Loss* are used to obtain more accurate text box. The model proposed in this paper has achieved good detection results on ICDAR2013, ICDAR2015 and MSRA-TD500 datasets.

[Key words] text detection; EAST; attention mechanisms; Strip Pooling; pyramid pooling

0 引言

文字作为传递信息的主要途径, 被广泛地运用在自然场景中。快速、准确地提取自然场景中的文字是一个不小的挑战, 这对后续的文本翻译、场景理解起着至关重要的作用。随着深度神经网络的兴起, 自然场景文本检测已然吸引了学界关注, 并陆续进行了相关研究。文本检测已经成为未来发展的一个热点方向。

现阶段自然场景文本检测算法已经从传统的方法转向了基于深度学习的方法。深度学习的方法的快捷之处在于: 直接检测单词, 没有过多的中间步

骤, 而且可以通过模型自行提取特征。目前基于深度学习的方法, 大致可以分成 2 类。一是基于位置回归^[1]的方法; 二是基于分割的方法。其中, 基于回归的方法源于一般的目标检测算法, 通过改变矩形框的形状来应对文本方案。Faster-RCNN^[2] 和 Ma 等学者^[3] 设计了旋转矩形框来检测那些任意方向的文本。但是回归的方法也存在一些缺陷, 因为默认每个格子只检测一个物体, 对于文本这样靠得比较近且跨度比较大的目标以及标点等小目标效果不好。例如, Shi 等学者^[4] 在 2017 年提出的 SegLink (Detecting oriented text in natural images by linking segments) 在检测跨度很大的文本行时效果不好。

基金项目: 国家自然科学基金(61977039)。

作者简介: 杨 豪(1997-), 男, 硕士研究生, 主要研究方向: 深度学习、图像处理; 方承志(1976-), 男, 博士, 副教授, 主要研究方向: 嵌入式系统、信号处理; 罗佳辉(1998-), 男, 硕士研究生, 主要研究方向: 深度学习、图像处理; 雷 蕾(2003-), 女, 本科生, 主要研究方向: 深度学习、图像处理。

通讯作者: 方承志 Email: fangchengzhi@163.com

收稿日期: 2022-08-29

基于分割的文本检测方法是将文本检测转换成语义分割问题。PixelLink^[5]就是采用了分割的方法,在预测中可以得出哪些像素属于文本。CRAFT^[6]通过评定每个字符和字符之间的联系来有效地检测文本区域。但分割的方法也存在缺陷,需要复杂的后期处理,速度较慢,且过多的后处理操作,导致在其他场景中一些参数很难被确定。例如 Li 等学者^[7]在 2019 年提出的 PSENet (Shape robust text detection with progressive scale expansion network) 需要大量的后处理去解决相邻文本实例不能准确分割的问题。

研究可知, EAST^[8] 算法可以有效地解决上述问题, 通过结合分割和回归的方法, 解决了繁琐的后处理问题, 实现了端到端的文本检测, 去掉了冗余和耗时中的中间步骤, 只包含 2 个部分, 全卷积神经网络^[9] (Fully convolutional networks, FCN) 生成文本行参数部分和非极大值抑制 (Non maximum suppression, NMS) 部分。但同时, EAST 中也展现出一些局限性, 能够检测的最大文本实例的大小与网络的感受野成正比, 从 EAST 实验结果上也反映出对于一些长度是宽度的 5 倍、甚至更长的文本, 由于跨度较大导致效果不好, 这是由于分割的方法在特征提取阶段, 使用卷积和下采样的方式提取特征, 多次的卷积和下采样会带来感受野过小的问题, 导致 EAST 算法在检测长文本时出现断裂、漏检等问题。

针对这一问题, 本文基于 EAST 算法, 引入了 Strip Pooling^[10] 的思想, 加入了 MPM (Mixed Pooling Module)^[10] 网络模块, 采用 $1 \times N$ 或 $N \times 1$ 的池化核, 在垂直和水平方向对特征图进行池化, 增大特征图感受野的同时, 可以在捕捉文本、尤其是长文本的过程中, 避免混入来自无关区域的干扰信息。同时加入了 Coordinate Attention^[11] 模块, 通过将位置信息嵌入到通道注意力中, 强化感兴趣的部分, 该模块仍然采用 $1 \times N$ 或 $N \times 1$ 的池化核, 使网络更准确地定位感兴趣的目标。在训练过程中, 为了解决正负样本不平衡问题, 引入 Dice loss^[12] 训练网络, 且重新定义了边界框的损失函数, 用 CIOU loss 代替了原来的 IoU loss, 考虑了边界框中心点的距离和宽高比, 使模型可以更快、更好地训练。

1 EAST 网络结构

EAST 是一个高效且准确的场景文本检测网络, 直接通过单一的神经网络预测出图像中的任意方向的四边形文本行, 网络只包含 FCN 阶段和 NMS 阶段, 使得网络非常简洁、优雅, 真正地实现了端到

端的文本检测。结构如图 1 所示。由图 1 可知, 主要包含特征提取层、特征融合层和输出层三大部分。对此拟做研究分述如下。

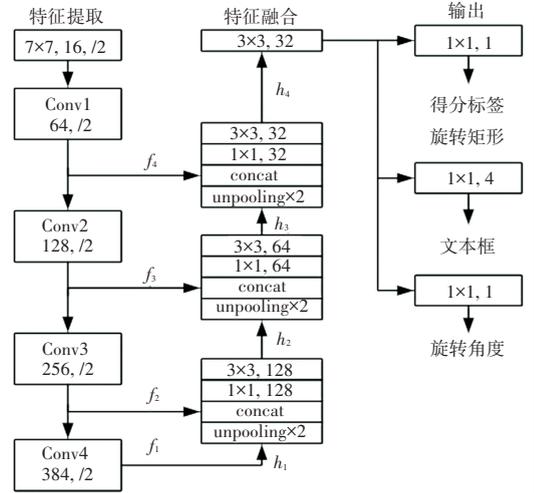


图 1 EAST 网络结构

Fig. 1 EAST network structure

(1) 特征提取层。采用 PVANet 网络, 对输入图片进行特征提取, 从 4 个卷积层抽取特征图, 卷积层的尺寸依次减半, 可以提取出不同尺寸的特征图, 用于后面的特征融合。

(2) 特征融合层。采用 U-Net 模型, 对提取的最后一层特征图进行上采样, 以便和上一层的特征图进行通道合并, 重复操作, 卷积核的个数逐层递减为 128、64、32, 最后到输出层。

(3) 输出层。通过 1×1 的卷积输出一个 1 通道的检测框置信度的分数特征图和一个 5 通道的旋转矩形特征图, 其中旋转矩形特征图包含 1 个角度特征图 and 4 个边界特征图。

2 引入 Strip Pooling 改进网络模型

本文引入 Strip Pooling 和改进损失函数来重新定义 EAST 网络, 可以有效地解决因为感受野的问题而导致检测长文本时出现错误等情况。本文的网络结构如图 2 所示。

2.1 感受野

在卷积神经网络中, 感受野是卷积神经网络每一层输出的特征图上的像素点在输入图片上映射的区域大小, 2 层 3×3 卷积的感受野演变过程如图 3 所示。2016 年 Luo 等学者^[13]提出 (Understanding the effective receptive field in deep convolutional neural networks) 实际感受野要小于理论感受野。

输入图片在一层一层的映射后, 输出的特征图受到卷积核尺寸的制约, 映射的区域大小有限, 当映

射区域不能完全覆盖长文本区域时,对长文本的检测就会出现检测框断裂、扭曲等问题,使得检测效果不尽人意,扩大感受野的大小是当前该类算法迫切需要解决的问题。2019 年提出的 (Feature pooling in scene character recognition: a comprehensive study)^[14] 中表明,池化操作可以有效地扩大感受野。

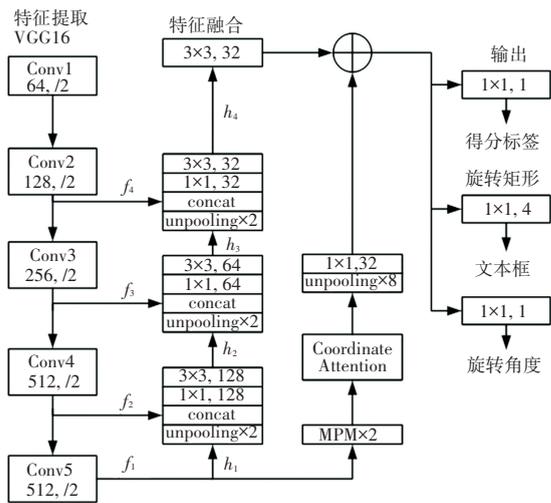


图 2 改进的 EAST 网络结构

Fig. 2 Improved EAST network structure

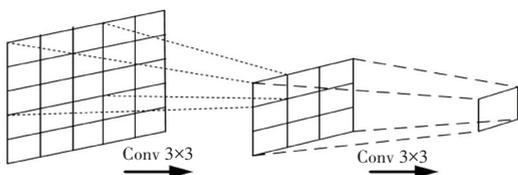


图 3 感受野演变过程

Fig. 3 Receptive field evolution process

2.2 MPM 网络模块

MPM 网络模块由一个 SPM (Strip Pooling Module)

模块和一个 PPM (Pyramid Pooling Module)^[15] 模块组成,可以同时捕获像素之间远程和短期的依赖关系。

空间池化已经被证明在像素级预测方面,对捕获远程上下文信息和扩大感受野非常有效。除了常规形状的空间池化外,针对文本文字的特殊形状,引入了新的池化策略 Strip Pooling,考虑了一个长而窄的核,即 $1 \times N$ 或 $N \times 1$ 的池化核,这可以有效地解决因使用方形池化核而引入一些不需要的干扰信息。

在扩大感受野的任务上,无论是卷积、还是池化,通常都会在一个方形窗口内对输入特征进行映射,这就对于去捕捉一些在现实场景中广泛存在的细长物体、例如长文本形成了限制。引入的 Strip Pooling 使用一个长核,沿着一个空间维度进行编码,从而能够有效地捕获该空间的远程依赖关系;在另一个空间维度上,使用一个窄核,便于捕获局部上下文信息,这可以有效地防止不相干区域的信息干扰标签的预测。在检测文本行的边缘时, 3×3 的方形池化核会突破文本行边缘从而引入不相关的背景像素, Strip Pooling 核在边缘上以一种窄核的形状进行池化,从而可以避免背景像素的干扰。Strip Pooling 核的长核部分比 3×3 、 5×5 这种方核范围长,有效地扩大了感受野,使得长的文本行也能被有效地检测。对于离散分布在图像中的文本行,方形的池化核并不能灵活地进行特征映射,长核的存在可以捕获这些孤立区域的关系,从而有效地解决这一问题。

在 Strip Pooling 的基础上,引入 SPM 模块。SPM 模块如图 4 所示,由 2 条路径组成,可以沿着水平或垂直空间维度进行编码,然后合并 2 个维度上的每个空间位置,进行特征细化。

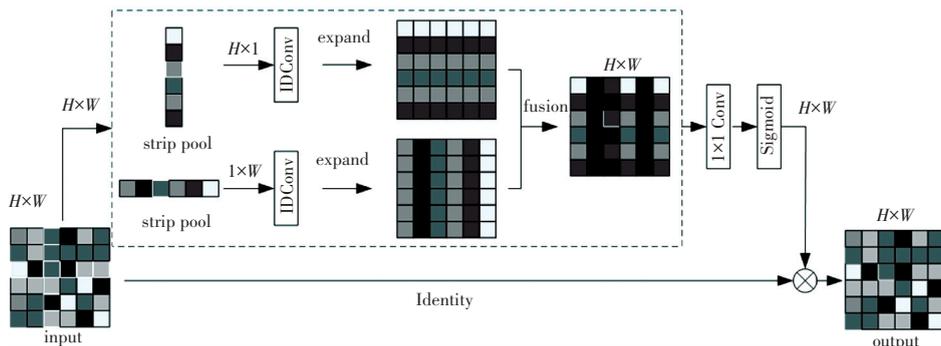


图 4 Strip Pooling Module 结构图

Fig. 4 Structure diagram of the Strip Pooling Module

SPM 沿着水平或垂直维度进行池化,输入一个二维张量 $x \in \mathbb{R}^{H \times W}$, 在 Strip Pooling 中,对一列或

一行中的所有特征值进行平均池化。因此,水平条带池化后的输出 $y^h \in \mathbb{R}^H$ 为:

$$y_i^h = \frac{1}{W} \sum_{0 \leq j \leq W} x_{i,j} \quad (1)$$

同时,垂直条带池化后的输出 $y^v \in \mathbb{R}^W$ 为:

$$y_j^v = \frac{1}{H} \sum_{0 \leq i \leq H} x_{i,j} \quad (2)$$

窄边更专注于捕获局部细节,长而窄的池化核容易建立离散区域之间的远程依赖关系,这些特性不同于依赖正方形核的传统空间池化。图4描述了SPM的结构,设 $x \in \mathbb{R}^{C \times H \times W}$ 为输入张量,将 x 输入到2个平行路径中,每个路径包含一个水平或垂直的Strip Pooling层,再经过一个核的大小为3的一维卷积层,用于调和当前位置和局部空间的特征,就可以得到 $y^h \in \mathbb{R}^{C \times H}$ 和 $y^v \in \mathbb{R}^{C \times W}$,之后在各自的方向上扩展到和输入特征图一样的大小,组合 y^h 和 y^v 得到 $y \in \mathbb{R}^{C \times H \times W}$,有助于后续得出包含更多全局信息的输出 $z \in \mathbb{R}^{C \times H \times W}$:

$$y_{c,i,j} = y_{c,i}^h + y_{c,j}^v \quad (3)$$

然后,输出 z 为:

$$z = \text{Scale}(x, \sigma(f(y))) \quad (4)$$

其中 $\text{Scale}(\cdot, \cdot)$ 是指元素级乘法; σ 是 Sigmoid 激活函数; f 是 1×1 卷积。输出特征图中的每个像素都与其有相同水平坐标或垂直坐标的像素建立了依赖关系,避免了在彼此远离的位置之间建立过多不必要的连接。

金字塔池化模块 (PPM) 可以捕获短程的依赖关系,有利于扩大感受野,但受限于标准的池化核,对长文本的处理效果不好。将 PPM 和 SPM 结合如图5所示,形成了 MPM 网络,可以通过各种池化操作聚合不同类型的上下信息,以便更有效地检测包括长文本在内的各种类型的文本文字。由图5(a)可看到,对于 PPM 模块采用一个轻量化的模块,避免引入大量参数,有2个空间池化层,池化后的特征图尺寸分别为 2×2 和 4×4 ,然后是卷积层,卷积核的尺寸为 3×3 ,用于多尺度的特征提取,还包含一个二维卷积层,用于保存原始特征图的空间信息,随后将这3条子路径合并输出。

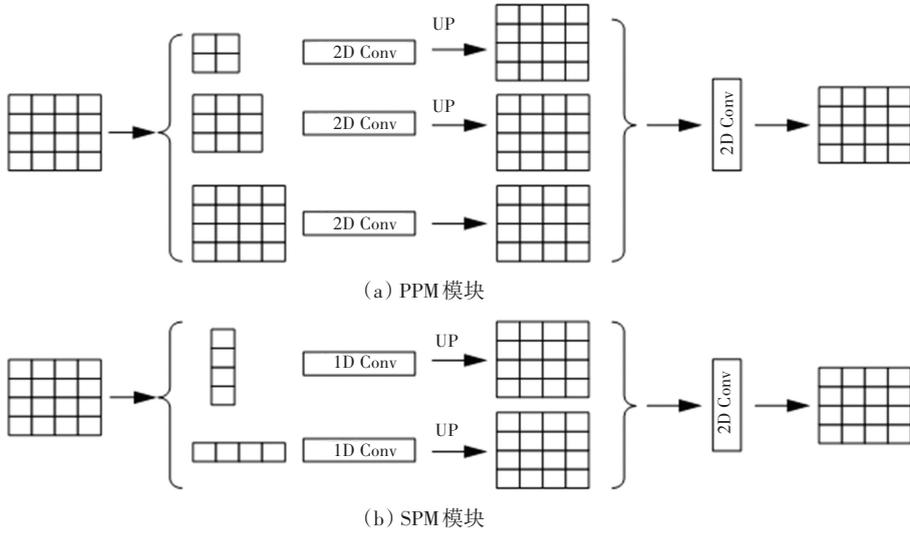


图5 PPM 模块和 SPM 模块

Fig. 5 PPM module and SPM module

MPM 是一个轻量化的模型,本文加入了2个MPM模块,且加入的参数量只有原始PPM的1/3,可以发挥更好的性能。

2.3 Coordinate Attention

Hu等学者^[16]在2020年提出的SENet (Squeeze-and excitation-networks)表明,通道注意力机制对提高模型性能具有显著效果。但忽略了位置信息,而这些位置信息对于生成空间选择性注意图非常重要,将位置信息嵌入到通道注意力中,得到Coordinate Attention,与普通二维全局池化不同之处在于位置信息的嵌入,引入Strip Pooling,将通道注

意力分解为2个一维特征进行编码,分别沿水平或垂直的空间方向聚集特征,以便更精确地检测长文本。

Coordinate Attention 模块如图6所示,可以捕获有助于最终分类决策的跨通道信息,还可以捕获方向感知和位置敏感的信息,有利于精确地定位长文本。

对于输入 X , 利用 Strip Pooling 原理,得到2个维度的输出 $z_c^h(h)$ 和 $z_c^w(w)$,这是一对具有方向感知能力的特征图,可通过通道合并、将其聚合,再通过 1×1 卷积变换函数 F_1 :

$$f = \delta(F_1([\cdot, \cdot])) \quad (5)$$

其中, $[\cdot, \cdot]$ 表示沿通道合并; δ 表示非线性激活函数; $f \in \mathbb{R}^{C/r \times (H+W)}$ 为2个方向上编码空间信息的中间特征图, r 为通道数缩减比, 以降低模型的复杂度。沿着空间维度将 f 分成独立的垂直和水平方向上的张量 $f^h \in \mathbb{R}^{C/r \times H}$ 和 $f^w \in \mathbb{R}^{C/r \times W}$, 再通过 1×1 卷积 F_h 和 F_w 扩展通道数, 以便和输入 X 的通道数一致, 得到:

$$g^h = \sigma(F_h(f^h)) \quad (6)$$

$$g^w = \sigma(F_w(f^w)) \quad (7)$$

其中, σ 为 sigmoid 激活函数。最后, 输出 Y 为:

$$y_c(i, j) = x_c(i, j) \times g_c^h(i) \times g_c^w(j)$$

这个编码过程使得 Coordinate Attention 模块可以更准确地定位长文本的位置, 整个模型就能更好地识别文本。

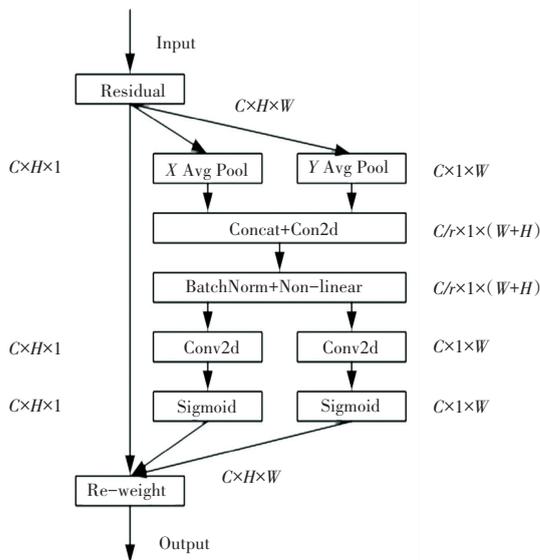


图6 Coordinate Attention 结构图

Fig. 6 Coordinate Attention structure diagram

2.4 改进损失函数

输出层由2个部分组成, 一个1通道的分数图和一个5通道的旋转矩形框, 为了优化回归任务, 替换了原有的损失函数, 得到了不错的效果。

自然场景文本检测不同于一般的物体检测, 文字占比少, 容易被背景所干扰, 导致正负样本极度不平衡, 为了解决这一问题, 使用 *Dice loss* 代替原文使用的类平衡交叉熵, 其表达式为:

$$L_{Dice} = 1 - \frac{2 \sum_{i=1}^N y_i * \hat{y}_i}{\sum_{i=1}^N y_i + \sum_{i=1}^N \hat{y}_i} \quad (8)$$

其中, y_i 与 \hat{y}_i 分别为像素 i 的标签值与预测值,

N 为像素点的总数。*Dice loss* 是一种用于评估2个样本相似性的度量函数, 对正负样本严重不平衡的场景有着不错的性能。

为了精确地定位矩形框, 重新定义了损失的计算方式, 采用 *CIoU loss*。相比常用的 *IoU loss*, *CIoU* 不仅考虑了边界框的重叠区域和中心点的距离, 还考虑了宽高比, 其表达式为:

$$CIoU = 1 - IoU + \frac{\rho^2(b, b^{gt})}{c^2} + \alpha v \quad (9)$$

其中, $\rho^2(\cdot, \cdot)$ 为欧几里得距离的平方; b 和 b^{gt} 分别为预测框和标签框的中心点坐标; c 为包含预测框和标签框的最小矩形的对角线长度; IoU 为原文中的交并比, 其公式具体如下:

$$IoU = \frac{|A \cap B|}{|A \cup B|} \quad (10)$$

其中, A 为预测框; B 为标签框; IoU 表示预测框和标签框的交集和并集的比值。 v 为预测框和标签框的宽高比的距离, 其公式具体如下:

$$v = \frac{4}{\pi^2} \left(\arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h} \right)^2 \quad (11)$$

其中, w^{gt} , h^{gt} , w , h 分别为标签框和预测框的宽和高。 α 为权重系数, 其公式具体如下:

$$\alpha = \frac{v}{(1 - IoU) + v} \quad (12)$$

这样设计的好处是 *CIoU* 不仅具有 *IoU* 的非负性的特点, 还解决了 *IoU* 的致命缺点, 当标签框和预测框的 *IoU* 为0, 网络可以继续训练。中心点距离的引入, 可以使标签框覆盖预测框或预测框覆盖标签框时, *loss* 不是一个定值, 且网络更倾向于移动预测框的位置来减少 *loss*。宽高比的引入, 使得预测框的大小更加接近标签框。旋转角度损失采用原文的损失函数。

3 实验结果与分析

3.1 实验环境

本文的实验仿真在 Linux 操作系统上进行, 开发语言为 Python3.8, 使用带有 GPU 版本的 Pytorch 深度学习框架。硬件配置 GPU 为 RTX3090, 显存为 24 GB。

3.2 实验结果

本文的实验采用 ICDAR2013、ICDAR2015 和 MSRA-TD500 数据集, ICDAR 数据集为 ICDAR 挑战赛所用的数据集。ICDAR2015 数据集包含 1 500 张图片, 其中 1 000 张图片用于训练集, 其余 500 张

图片用于测试集。MSRA-TD500 数据集包含 500 张自然场景图片。本文将改进的模型和原模型、其他模型做比较,分别从准确率 P (*Precision*)、召回率 R (*Recall*)、 F 值三个方面评估检测模型的性能。

本文加入了 2 个 MPM 模块,相比加入一个 MPM 模块性能提升了 0.9 个百分点,见表 1。轻量化的模型优势在于不需加入太多的参数,就可以更好地发挥出模型的性能。

表 1 ICDAR2015 数据集上的指标

Tab. 1 Metrics on the ICDAR2015 dataset %

方法	准确率	召回率	F - 值
1 个 MPM	83.94	81.56	82.74
2 个 MPM	84.83	82.54	83.67

加入 2 个 MPM 模块在 ICDAR2013、ICDAR2015 和 MSRA-TD500 数据集上对比其他检测模型的实验结果见表 2~表 4。由表 2~表 4 分析可知,本文的方法相较于原方法和其他方法,都有了一定的提升,且本文的方法在 ICDAR2015 数据集上 F 值达到了 83.67%,在 MSRA-TD500 数据集上 F 值达到了 75.17%。

表 2 ICDAR2013 数据集上的指标

Tab. 2 Metrics on the ICDAR2013 dataset %

方法	准确率	召回率	F - 值
EAST+VGG16	92.98	82.98	87.69
CTPN + VGG16	93.01	83.04	87.78
Pixel-link	88.62	87.53	88.16
本文方法	93.25	84.79	88.82

表 3 ICDAR2015 数据集上的指标

Tab. 3 Metrics on the ICDAR2015 dataset %

方法	准确率	召回率	F - 值
EAST+VGG16	80.46	72.75	76.41
WordSup	79.36	77.03	78.24
He et al.	82.02	80.04	81.03
TextSnake	84.92	80.41	82.63
本文方法	84.83	82.54	83.67

表 4 MSRA-TD500 数据集上的指标

Tab. 4 Metrics on the MSRA-TD500 dataset %

方法	准确率	召回率	F - 值
EAST+VGG16	81.67	61.60	70.23
He et al.	77.02	70.00	73.34
Zhang et al.	83.02	67.05	74.18
本文方法	84.53	67.69	75.17

对比原来的 EAST 算法和改进的 EAST 算法的检测效果,如图 7 所示。由图 7 可知,对于长文本的检测,效果有了不错的提升,改善了长文本检测时文本框断裂、扭曲等情况。



图 7 原图和改进效果图

Fig. 7 The original renderings and the improved effect renderings

采用本文的基于 Strip Pooling 的 EAST 算法比经典的 EAST 算法具有更高的准确率、召回率和 F 值。对比其他算法, F 值也提高了一到两个百分点,检测效果得到了提升。本文的算法在长文本的检测中展现了显著优势,对于商场标语、横幅标语、店铺名这些要求细而长的文本,改进的算法利用其优势可以快速而准确地检测出来,不会出现检测框断裂、漏检等情况;对于短小文本,改进的算法依然能够有效地检测出来,并没有损失其准确性。

4 结束语

本文基于 EAST 算法,引入 Strip Pooling 池化核优化模型,有效地解决了 EAST 模型在检测长文本时,由于感受野过小导致文本框断裂、扭曲等情况。加入 MPM 模块和 Coordinate Attention 模块,利用不同尺寸的池化核,尤其是 Strip Pooling 池化核,捕获远程和短程的信息,不仅扩大了感受野,还避免了一些背景信息对文本的干扰。实验证明,该方法有效地提升了 EAST 这类算法检测长文本的能力。

参考文献

- [1] 金灵. 基于 RetinaNet 的场景文字检测算法[J]. 计算机应用与软件, 2022, 39(02): 201-207.

- [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: 1137-1149.
- [3] MA Jianqi, SHAO Weiyuan, YE Hao, et al. Arbitrary-oriented scene text detection via rotation proposals [J]. IEEE Transactions on Multimedia, 2018, 20: 3111-3122.
- [4] SHI Baoguang, BAI Xiang, BELONGIE S. Detecting oriented text in natural images by linking segments [C] //2017 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA ;IEEE, 2017: 3482-3490.
- [5] DENG Dan, LIU Haifeng, LI Xuelong, et al. PixelLink: Detecting scene text via instance segmentation[C] //2018 AAAI Conference on Artificial Intelligence / 30th Innovative Applications of Artificial Intelligence Conference / 8th AAAI Symposium on Educational Advances in Artificial Intelligence. Louisiana, USA : AAAI, 2018: 6773-6780.
- [6] BAEK Y, LEE B, HAN D, et al. Character region awareness for text detection [C] //2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, CA, USA :IEEE, 2019: 9357-9366.
- [7] LI Xiang, WANG Wenhui, HOU Wenbo, et al. Shape robust text detection with progressive scale expansion network [C] //2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, California, USA; IEEE, 2019: 9328-9337.
- [8] ZHOU Xinyu, YAO Cong, WEN He. EAST: an efficient and accurate scene text detector[C] //2017 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA : IEEE, 2017: 2642-2651.
- [9] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39: 640-651.
- [10] HOU Qibin, ZHANG Li, CHENG Mingming, et al. Strip pooling: Rethinking spatial pooling for scene parsing [C] //2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, 2020: 4002-4011.
- [11] HOU Qibin, ZHOU Dequan, FENG Jiashi. Coordinate attention for efficient mobile network design [C] //2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, 2021: 13708-13717.
- [12] MILLETARI F, NAVAB N, AHMADI S. V-Net: Fully convolutional neural networks for volumetric medical image segmentation [C] //2016 IEEE International Conference on 3D Vision. Stanford, CA, USA: IEEE, 2016: 565-571.
- [13] LUO Wenjie, LI Yujia, URTASUN R, et al. Understanding the effective receptive field in deep convolutional neural networks [C] // 2016 Conference on Neural Information Processing Systems. Barcelona, Spain: NIPS Foundation, 2016(29): 1-9.
- [14] ZHANG Zhong, WANG Hong, LIU Shuang, et al. Feature pooling in scene character recognition: A comprehensive study [C] //2017 International Conference on Communications, Signal Processing, and Systems. Xiamen, China: ICCSPS, 2019: 2150-2157.
- [15] ZHAO Hengshuang, SHI Jianping, QI Xiaojuan, et al. Pyramid scene parsing network [C] //2017 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA : IEEE, 2017: 6230-6239.
- [16] HU Jie, SHEN Li, SUN Gang. Squeeze-and-excitation networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42: 2011-2023.

(上接第29页)

参考文献

- [1] 蔡自兴. 机器人学的发展趋势和发展战略[J]. 高技术通讯, 2001(04): 11-16.
- [2] 戈新良. 基于多种传感器信息融合的移动机器人的环境辨识[D]. 天津: 河北工业大学, 2003.
- [3] 汪宝珠. 基于SLAM的移动机器人地图构建与路径规划研究[D]. 银川: 宁夏大学, 2020.
- [4] 安石莉. 防疫消杀巡检机器人定位导航系统设计[D]. 银川: 宁夏大学, 2022.
- [5] 赵伟卓. 基于视觉与激光雷达的ROS机器人自主导航研究[D]. 沈阳: 沈阳理工大学, 2022.
- [6] 李建勇, 赵静, 刘雪梅, 等. 基于ROS的自主家庭陪伴机器人设计[J]. 电子技术应用, 2021, 47(02): 58-62.
- [7] 何琦. 一种超市购物机器人的结构设计及控制方法[J]. 内燃机与配件, 2018(09): 21-23.
- [8] 张彦俊. 机器人模块化体系与关键技术研究[D]. 上海: 上海交通大学, 2018.
- [9] 陈鑫, 秦宏伟, 陈春雨, 等. 基于Cortex-M3内核的STM32微控制器研究与电路设计[J]. 大庆师范学院学报, 2013, 33(06): 44-47.
- [10] 牛闯, 王翰章, 赵玉立, 等. 基于STC89C52单片机的循迹智能车控制系统设计[J]. 电脑知识与技术, 2013, 9(10): 2488-2492.
- [11] HUANG Jian. Development of DC motor drive based on TB6612 [C] // Proceedings of 2017 5th International Conference on Frontiers of Manufacturing Science and Measuring Technology (Computer Science and Electronic Technology International Society). [S.l.]: Atlantis Press, 2017: 714-717.
- [12] 张鹏儒. 基于STM32微控制器智能军营环境参数检测系统设计[D]. 西安: 西安电子科技大学, 2015.
- [13] JAINV, PRADHAN S K. Vibration monitoring of ball nose end mill tool during milling of sculptured surfaces using MUP6050 sensor[J]. Materials Today: Proceedings, 2020, 27(1): 2477-2486.
- [14] 任杰, 宋建涛, 李功燕. 开放式环境下变电站机器人SLAM算法研究[J]. 信息技术与网络安全, 2018, 37(03): 80-83.
- [15] 范一赢, 张慧贤, 布占伟, 等. 基于ROS的消防机器人激光雷达地图构建方法研究[J]. 工业控制计算机, 2022, 35(11): 56-58.
- [16] 李建勇, 刘雪梅, 李雪霞, 等. 基于ROS的开源移动机器人系统设计[J]. 机电工程, 2017, 34(02): 205-208.