文章编号　1004-924X(2021)09-2210-12

多维度卷积融合的密集不规则文本检测

孟月波[1,2]，石德旺[1]，刘光辉[1*]，徐胜军[1,2]，金　丹[1]
(1. 西安建筑科技大学 信息与控制工程学院，陕西 西安 710055；
2. 人工智能与数字经济广东省实验室(广州)，广东 广州 510000)

摘要：基于深度学习的自然场景文本检测算法进展显著，但对具有密集不规则排布特点的文本来说，由于其间距小、分布密集，导致特征提取困难，文本检测不全；同时，现有文本检测方法常采用的不同维度特征直接拼接的方式会导致多尺度特征融合不充分，造成语义信息的丢失。针对上述问题，本文提出一种基于多维度卷积融合的密集不规则文本检测方法。网络主体采用FPN结构，设计了文本增强模块(Text Enhancement Module，TEM)，通过引入额外全局文本映射以强化网络对文本信息的关注能力；提出了通道融合策略(Channel Fusion Strategy，CFS)，采用自底向上方式建立高低维度特征信息链，生成语义更加丰富的特征图，减少信息损失；预测阶段采用渐进式拓展文本核的方法生成文本预测结果。在DAST1500及ICDAR2015和CTW1500数据集上的实验表明，该方法其F值分别达到81.8%，83.0%及79.0%。提出算法不仅在密集不规则文本检测上表现出更好的性能，而且在一般自然场景文本(多向、曲线文本)上也具有一定竞争力。

关　键　词：密集不规则文本；深度学习；卷积神经网络；文本增强；通道融合
中图分类号：TP273　　**文献标识码**：A　　**doi**：10.37188/OPE.20212909.2210

Dense irregular text detection based on multi-dimensional convolution fusion

MENG Yue-bo[1,2], SHI De-wang[1], LIU Guang-hui[1*], XU Sheng-jun[1,2], JIN Dan[1]

(1. *College of Information and Control Engineering, Xi'an University of Architecture and Technology, Xi'an* 710055, *China*;
2. *Guangzhou Artificial Intelligence and Digital Economy Laboratory, Guangzhou* 510000, *China*)
* *Corresponding author, E-mail: guanghui*1@163.*com*

Abstract: Natural-scene text-detection algorithms based on deep learning have made significant progress; however, they only apply to texts with dense and irregular layouts. Owing to its small spacing and dense distribution, it is difficult to extract features from texts and the detection remains incomplete. Meanwhile, the existing text detection methods often use the direct splicing of different dimensional features, leading to insufficient multi-scale feature fusion and the loss of semantic information. To solve these problems, a dense irregular text detection method is proposed based on multi-dimensional convolution fusion. The net-

收稿日期：2021-02-24；**修订日期**：2021-04-10.
基金项目：国家自然科学基金面上项目(No. 51678470)；陕西省自然科学基础研究计划面上项目(No. 2020JM-473，No. 2020JM-472)；西安建筑科技大学基础研究基金项目(No. JC1703)；西安建筑科技大学自然科学基金项目(No. ZR19046)

work follows the FPN structure and utilizes a text enhancement module (TEM). By using additional global text mapping, the network pays special attention to the text information. A channel fusion strategy (CFS) is proposed, which uses the bottom-up method to establish the high-low dimension feature information chain to generate the feature map with richer semantics and reduce the information loss. In the prediction stage, text prediction results are generated through the gradual expansion of the text kernel. Experimental results on DAST1500, ICDAR2015, and CTW1500 datasets yield F values of 81.8%, 83.8%, and 79.0% respectively. The proposed algorithm not only has better performance in dense and irregular text detection but also shows a certain level of competitiveness in the case of general natural scene texts (multi-directional, curvilinear text).

Key words: dense irregular text; deep learning; convolution neural network; text enhancement; channel fusion

1　引　言

文本作为自然场景图像中，包含丰富的语义信息，有助于场景内容的分析与理解。不同于一般文本，密集不规则文本具有文字多、密集排布等特点，且彼此检测距离极小，使其检测难度较大。此类文本作为商品外包信息载体，广泛地存在于商品包装图像中，其检测对商品统计、管理、巡查至关重要，对推动无人超市的进一步发展具有重要意义。

对自然场景文本检测问题，国内外相关研究由来已久，并取得一定成果[1-2]。传统文本检测方法其核心思想是对文本低级特征进行检测，例如最大稳定极值区域(Maximally Stable Extremal Region，MSER)[3]、笔划宽度变换(Stroke Width Transform，SWT)[4]以及方向梯度直方图(Histogram Of Oriented Gridients，HOG)[5]，这些方法通常过程复杂，很大程度依赖文本质量，鲁棒性较差，复杂自然场景下文本检测效果不佳。

随着深度学习的不断发展，很多基于深度神经网络的场景文本检测方法被陆续提出，并取得良好的检测效果，逐渐成文本检测的主流方法。深度神经网络文本检测方法大致可以分为区域建议[6-11]和图像分割[12-15]的方法。基于区域建议的方法将文本视为一类特殊目标，采用诸如SSD[16]，RCNN[17]系列等目标检测算法，使用回归文本框的方式获取文本。CTPN[6]在Faster-RCNN[18]架构上提出宽度固定的水平锚框，将VGG16串联双向LSTM的联合模型预测文本，水平锚框的设计有效解决了水平文本的检测问题，由于其锚框结构固定难以处理多方向文本；SegLink[7]在SSD基础上检测文本片段并预测其链接，通过直线拟合的规则重组文本实例，以处理多方向文本，然而直线拟合规则导致曲线文本检测效果不佳；DMPNet[8]应用四边形回归来检测多方向文本，但其四边形依然难以应对形状复杂的曲线文本；针对曲线文本形状复杂，CTD[9]提出14点多边形回归曲线文本，设计非多边形抑制和多边形非最大值抑制两种后处理实现曲线文本检测，但无法解决间距较近的曲线文本检测。上述方法由于直接回归文本边界坐标的限制，应对形状复杂文本时依然存在候选框无法完整拟合文本的问题。基于图像分割的场景文本检测方法将文本检测视为广义的“分割问题”，利用全卷积网络(Fully Convolutional Network，FCN)从像素层面区分文本，由于其不受文本形状限制备受关注。文献[12]采用FCN提取文本块，通过MSER从文本块中检测候选字符，步骤繁琐使其检测过程往往比较耗时；EAST[13]针对文本检测的复杂过程，提出基于U-Net架构的FCN与非最大抑制算法的简洁框架，通过预测像素到所属文本边界距离的方式实现文本检测，然而在处理曲线文本时存在检测框冗余的现象；TextSnake[14]设计文本中心线，通过文本中心线上多个圆环预曲线文本，解决了曲线文本检测的冗余，但依然难以解决曲线文本粘连问题；2019年，PSENet[15]首次提出文字核的概念，通过渐进式拓展文本核的方式较好地解决曲线文本粘连

问题，但对密集不规则场景文本依然存在文本检测不全、特征提取能力有待进一步提升等问题。同时，现有文本检测方法通常将不同维度特征直接拼接，这样的融合方式会导致文本多尺度特征融合不充分，造成语义信息的丢失。

针对以上问题，本文提出了一种基于多维度卷积融合的密集不规则场景文本检测方法。首先，通过设计的文本增强模块（Text Enhancement Module，TEM），从原始特征空间中提取全局文本映射并将其编码，加强文本区域权重，提高网络对文本区域的关注度；然后，采用FPN网络结构提取文本的多维度特征；之后，利用提出的通道融合策略（Channel Fusion Strategy，CFS），构建不同维度特征间的链式信息关系，加强不同尺度特征间的信息联系；最后，采用渐进式拓展文本核的思想生成文本候选框。实验结果表明，该方法提升了密集不规则文本的检测性能，同时，在多向、曲线文本上也有良好的检测效果。

2　多维度卷积融合密集不规则文本检测方法

2.1　网络结构

本文方法的网络基本结构如图1所示。输入图像经一次卷积、池化操作构造初始特征空间，采用TEM模块捕获全局文本特征并增强网络关注密集文本信息的能力；以ResNet50[19]为网络骨架，利用$\{R_1,R_2,R_3,R_4\}$四个卷积块自底向上依次下采样，收集多分辨率特征信息，采用上采样方式将其与相邻卷积块的输出通过1×1卷积进行横向链接合并，得到描述不同语义信息的特征映射$\{P_1,P_2,P_3,P_4\}$；将特征映射由高到低进行不同倍数上采样操作，采用三段式CFS策略，建立多维度特征间的链式信息关系，并通过维度拼接方式生成融合特征F；通过三种不同尺寸的文本核S_1,S_2,S_3对融合特征F进行文本实例分割，采用渐进式拓展方式生成文本候选框，将图像分为文本、非文本两类，得到最终预测结果。图中，©表示维度拼接。

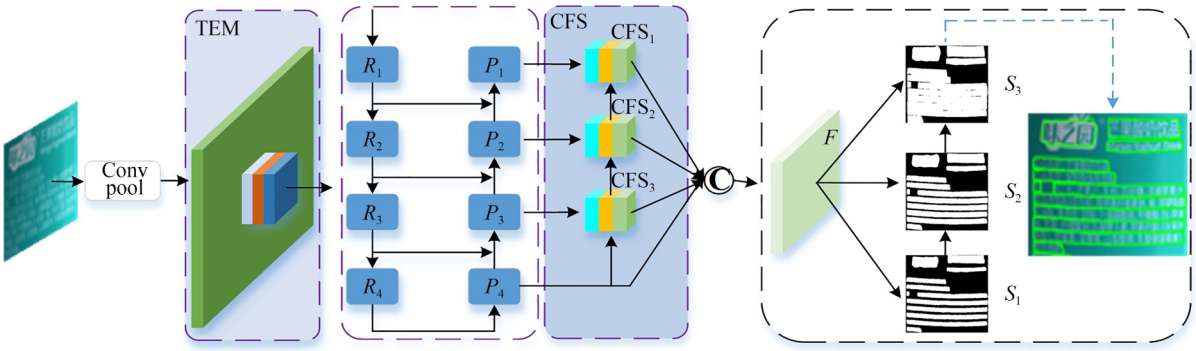


图1　网络基本结构

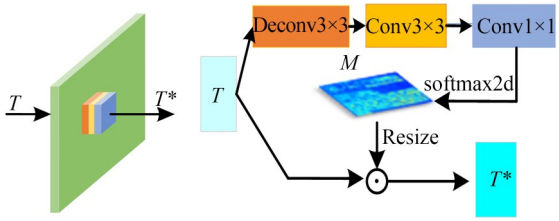
Fig. 1　Network basic structure

2.2　文本增强模块TEM

尽管PSENet在近期的文本检测模型中表现突出，其核思想有效解决了文本粘连问题，但在处理密集不规则文本时由于文本间距极小且形状复杂，使其难以捕获独立完备的文本信息，而这可能导致造成文本检测不全，导致密集不规则文本检测的召回率偏低。

本文设计的TEM模块从原始特征空间中提取粗略文本空间区域，通过生成全局文本映射M描述输入图像的文本区域概率，然后将其编码到原始特征空间获取细粒度感知映射能力，达到增强文本区域信息的目的，提高网络密集文本检测能力。TEM结构具体如图2所示，其中图2(a)为TEM结构示意图，图2(b)为TEM具体结构图。

高度为H、宽度为W的输入图像经7×7×64的卷积、3×3的池化操作后，得到张量大小为$\frac{w}{2}\times\frac{H}{2}\times 64$的高维初始特征$T$，利用公



(a) TEM结构示意图　(b) TEM具体结构图
(a) TEM structure diagram　(b) TEM specific structure diagram

图 2　TEM 模块结构

Fig. 2　TEM modular structure

式(1)计算全局文本映射 M：

$$M = Soft\max 2d\left(TEM\left(T\right)\right), \tag{1}$$

其中，$TEM(\cdot)$ 包含 3×3 反卷积、3×3 卷积及 1×1 卷积操作，全局文本映射 M 大小为 $W \times H \times 2$。

采用公式(2)，将全局文本映射 M 与初始特征 T 编码整合，得到具有显著性信息描述能力的强化特征 T^*，其大小与初始特征 T 保持一致：

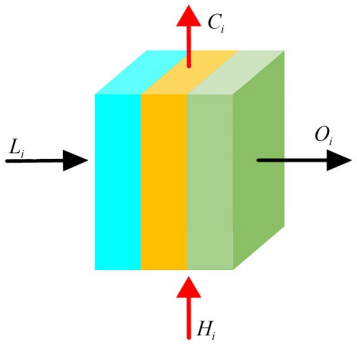
$$T^* = Resize(M) \odot T, \tag{2}$$

其中，$\odot$ 指点乘。

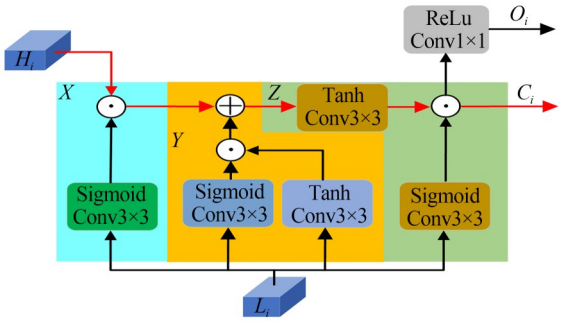
2.3　通道融合策略 CFS

在不同尺度特征图融合阶段，通常将不同维度特征图降维至统一维度后直接在维度方向进行简单叠加，这种直接叠加方式势必造成信息损失。一方面，不同阶段产生的特征图由于多次卷积导致彼此间的信息传递缺失；另一方面，由于密集文本本身过小，直接叠加会削弱高层语义信息与底层信息之间的关联性，导致局部文本漏检。本文提出的 CFS 策略通过建立高低维度特征图间的信息链，实现不同尺度特征的充分融合，生成表征信息更加丰富的特征图，减少特征损失。

本文的 CFS 策略主要分为三阶段，每阶段完成的功能一致，记为 $CFS_i(i=3,2,1)$。CFS_i 结构如图 3 所示，其中图 3(a)为 CFS_i 结构示意图，图 3(b)为 CFS_i 具体结构图。输入 L_i 为第 i 阶段的低维特征映射，获取过程如式(3)所示，即为前序模块获得的特征映射 P_i；输入 H_i 为此阶段信息链的高维特征映射，获取过程如式(4)所示；借鉴长短时记忆网络(Long Short-Term Memory，LSTM)[20]网络思想，输入 L_i 和 H_i 依次通过 X，Y，Z 三个信息筛选步骤实现信息交互与融合，输出



(a) CFS_i 结构示意图
(a) CFS_i structure diagram



(b) CFS_i 具体结构图
(b) CFS_i specific structure diagram

图 3　CFS_i 策略结构

Fig. 3　CFS_i strategy structure

该阶段的强化特征 O_i 以及融合当前高低维度特征的信息链 C_i。$\odot$ 代表点乘，$\oplus$ 代表逐元素相加。

$$L_i = P_i, i = 3, 2, 1. \tag{3}$$

$$H_i = \begin{cases} P_4 & , i = 3 \\ C_{i+1} & , i = 1, 2 \end{cases}. \tag{4}$$

三个信息筛选步骤中，步骤 X 利用低维特征信息对信息链中的高维信息进行过滤，去除无效信息和噪声，具有信息遗忘功能；步骤 Y 具有信息更新功能，提取有效特征并对信息链进行更新；步骤 Z 具有信息输出功能，得到本维度的信息链输出 C_i，并通过 1×1 的卷积获得增强特征 O_i，增强网络的非线性表达能力；更新后，信息链 C_i 中包含当前维度和当前维度之前的所有特征层的融合信息，相比原始特征映射 P_i，包含更丰富的多尺度信息。信息链输出 C_i 以及增强特征 O_i 采用公式(5)、公式(6)进行计算：

$$O_i = \Re(f_{1\times1}(C_i)), \tag{5}$$

$$C_i = F_Z^i(F_Y^i, L_i), \tag{6}$$

其中：$\Re$ 表示 Relu 激励函数，$f_{1\times1}(\cdot)$表示卷积核为 1×1 的卷积层；$F_X^i(\cdot)$，$F_Y^i(\cdot)$和 $F_Z^i(\cdot)$为第 i 阶段的信息遗忘函数、信息更新函数、信息输出函数，具体通过式(7)～式(9)进行计算：

$$F_X^i(H_i,L_i)=\sigma(f_{3\times3}(L_i))\odot H_i, \tag{7}$$

$$F_Y^i(F_X^i,L_i,L_i)=F_X^i\oplus(\sigma(f_{3\times3}(L_i))\odot\Gamma(f_{3\times3}(L_i))), \tag{8}$$

$$F_Z^i(F_Y^i,L_i)=\sigma(f_{3\times3}(L_i))\odot\Gamma(f_{3\times3}(F_Y^i)), \tag{9}$$

其中，σ，Γ 分别表示 Sigmoid 和 Tanh 激励函数，$f_{3\times3}(\cdot)$表示卷积核为 3×3 的卷积层，所有卷积之后均使用批归一化(Batch Normalization，BN)。

2.4 基于文本核渐进式拓展的文本预测过程

将 CFS 策略三个阶段输出的的强化特征 O_3，O_2，O_1进行维度拼接，生成融合特征 F。采用 PSENet[15]提出的文本核渐进式拓展方法，确定最终的文本边界，生成文本候选框，实现文本检测。具体步骤如下：

Step 1：选择三种不同尺寸的文本核 S_1，S_2，S_3 对融合特征 F 进行文本实例分割，$S_1<S_2<S_3$。同一本文实例的不同文本核分割结果形状一致、中心点一致、面积不同，大文本核对应的文本实例分割结果面积大。如图 4(a)～4(c)所示。

Step 2：以本文核 S_1 对应的文本实例分割结果作为目标文本的核心区域，按照“上下左右”四个方向将核心区域像素依次向外拓展，扩大文本区域，直至发现本文核 S_2 的实例分割边界。图 4(d)、图 4(e)、图 4(f)以 5×5 像素块为例对该过程进行了示意，灰色像素点代表可扩展范围，蓝色、桔色像素点代表 2 个文本实例，“0”表示文本核心区域。图 4(d)～4(e)为像素一轮次拓展过程，图 4(e)～4(f)为像素二轮次拓展过程。

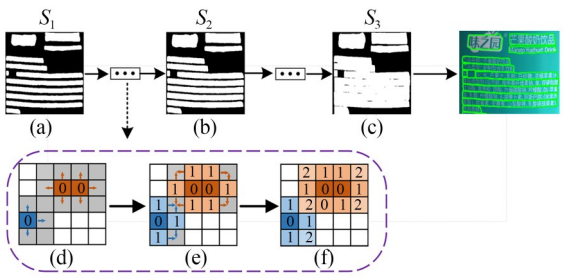


图 4　基于渐进式拓展文本核的文本候选框生成过程

Fig. 4　Text candidate box generation process based on progressive expanded text kernel

Step 3：以步骤 2 的拓展结果作为目标文本的核心区域，重复步骤 2，直至发现本文核 S_3 的实例分割边界。

3　实验与实验结果分析

3.1　数据集

DAST1500[11]：2019 年国内公开的首个密集商品包装文本数据集，该数据集数据大多来源于网络图片，包括 1 538 张图像以及 45 963 个单词级注释(包含 7 441 个不规则包围盒)，1 038 张作为训练图像，500 张作为测试图像。图像文字主要以中文为主，包含少数的英文和日文。图像尺寸任意，大多为 800×800，采用不规则点集组成的多边形描述作为文本标注形式。

ICDAR2015[21]：2015 年公开的多方向文本数据集，该数据集注重场景文本的随机性，由 1 000 张训练图像和 500 张测试图像组成，数据主要来自商场或街道。其文本在大小、方向以及清晰度存在较大差异，通过文本四个顶点对文本行进行注释。

CTW1500[9]：2017 年公开的曲线文本数据集，该数据集侧重于极具挑战性的弯曲文本，共 1 000 张训练图像及 500 张测试图像，每张图像至少包含一个文本注释，总计超过 10K 注释。数据集文本实例由 14 个点组成的多边形来表示。

3.2　评价指标

为了评估所提方法的性能，本文采用 IOU 的测评指标进行检测，其中准确率(Precision)、召回率(Recall)及 F 值(F-score)采用式(10)～式(12)进行计算：

$$Precision=\frac{TP}{TP+FP}, \tag{10}$$

$$Recall=\frac{TP}{TP+FN}, \tag{11}$$

$$F-score=\frac{2\times Precision\times Recall}{Precision+Recall}, \tag{12}$$

其中，TP，FP，FN 分别表示正确检测、错误检测、未检测的检测框个数。

3.3　实验细节

本文实验基于 Ubuntu16.04 下的 Pytorch1.3 实现，使用 CUDA10.1 及 cuDNN7.5 辅助加速计算，所有实验均在一台搭载 4 块 2080Ti 显卡的服

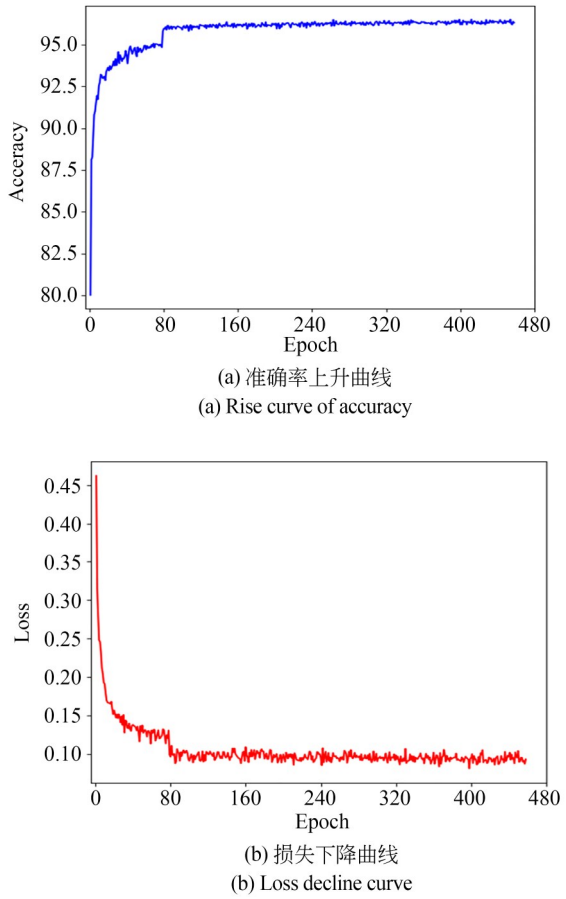
务器上进行的，其计算机CPU型号为Intel Xeon Silver 4114。主干网络ResNet50经由ImageNet[22]预训练，采用标准SGD网络优化算法，初始学习率、动量以及权重衰减系数设置为0.001，0.99，5×10^{-4}。网络采用分段学习策略，数据集DAST1500的batch size设置为8，迭代60K次，第20K次和第40K次迭代时学习率衰减为前次学习率的1/10；数据集ICDAR2015及CTW1500的batch size设置为16，迭代36K次，第12k迭代和第24k次迭代时学习率衰减为前次学习率的1/10。

为使模型充分训练，增强模型的鲁棒性，采用数据增强方法对网络训练图像进行数据扩增，具体以{0.5，1.0，2.0，3.0}的缩放比对图像进行随机缩放，在[−10°，10°]范围内对图像进行随机翻转，并对缩放后的图像进行随机裁剪（DAST1500图像尺寸裁剪为512×512，ICDAR2015、CTW1500图像尺寸裁剪为640×640）。此外，训练时采用难例样本挖掘（Online Hard Example Mining，OHEM）[23]策略，正负样本比例设定为1∶3。

DAST1500数据集训练收敛曲线如图5所示，图5(a)为准确率上升曲线，图5(b)为损失下降曲线。可以看出，训练初期准确率约为80%，损失函数值约为0.46。训练初期精度上升较快，同时损失下降较快，随着训练次数的增加，准确率从缓慢上升趋于平稳，最终稳定在96%，损失函数值由缓慢下降趋于收敛，最终稳定在0.09左右。从此参数的收敛情况分析可知，本文方法的训练结果比较理想。

3.4　网络执行过程实验

以图1中的示意图像为例，对本文所提网络各模块执行过程进行实验说明。



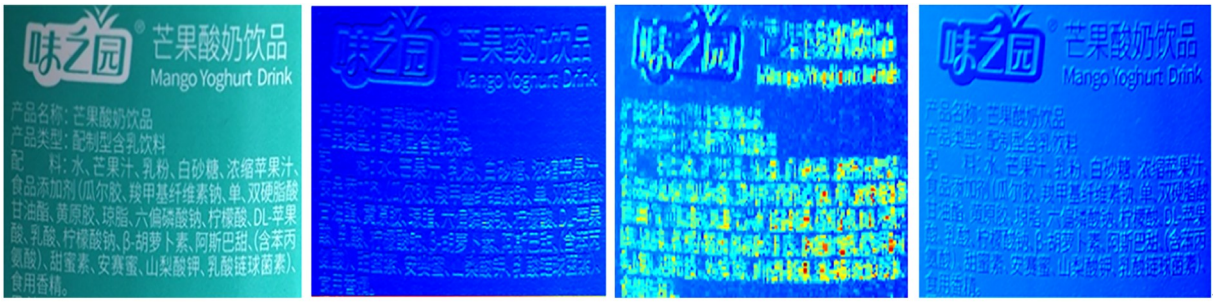
(a) 准确率上升曲线
(a) Rise curve of accuracy

(b) 损失下降曲线
(b) Loss decline curve

图5　准确率与损失曲线

Fig. 5　Accuracy and loss curves

图6(a)所示待检测图像经骨架网络ResNet50的一次卷积、最大池化层提取文本初始特征，提取结果如图6(b)所示。将提取到的初始特征送入图2所示TEM模块，强化文本区域权重。利用公式(1)计算全局文本映射M，用每一



(a) 待检测图
(a) Figure to be detected

(b) 初始特征图
(b) Initial feature map

(c) 全局映射M
(c) Global text mapping M

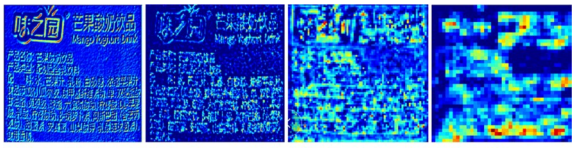
(d) 强化特征T^*
(d) Enhanced text features T^*

图6　初始特征与文本强化特征提取实验结果

Fig. 6　Experiential results of initial feature extraction and enhanced text feature extraction

个像素值表征待检测图像对应位置的文本区域概率，实验结果如图 6(c)所示；采用逐像素点乘的方式，将全局文本映射 M 与初始特征空间进行编码，生成如图 6(d)所示的文本强化特征 T^*。可以看出，经过 TEM 模块处理，网络对于文本区域的关注能力显著增强。

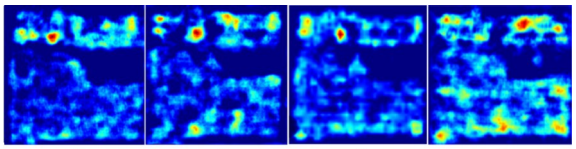
将文本强化特征 T^* 送入 FPN 网络，提取多维度特征。图 7(a)为 FPN 网络下采样过程得到的多分辨率特征 $\{R_1, R_2, R_3, R_4\}$，分辨率 $R_1 > R_2 > R_3 > R_4$；图 7(b)为 FPN 上采样过程提取到的多维度特征 $\{P_1, P_2, P_3, P_4\}$，维度 $P_1 < P_2 < P_3 < P_4$。由实验结果可以看出，获得的多分辨率特征 $\{R_1, R_2, R_3, R_4\}$ 中，高分辨率特征更关注文本的轮廓信息，低分辨率特征可以捕捉文本的语义信息。多维度特征 $\{P_1, P_2, P_3, P_4\}$ 中，维度越高，文本特征提取越全面。



R_1　R_2　R_3　R_4

(a) FPN下采样得到的多分辨率特征{R_1，R_2，R_3，R_4}

(a) Multi-resolution features obtained by sampling under FPN{R_1, R_2, R_3, R_4}



P_1　P_2　P_3　P_4

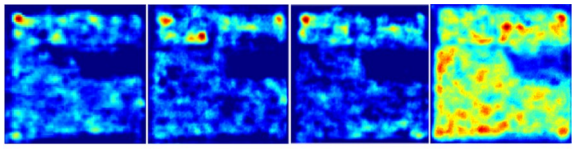
(b) FPN上采样提取到的多维度特征{P_1，P_2，P_3，P_4}

(b) Multi-resolution features obtained by sampling under FPN{P_1, P_2, P_3, P_4}

图 7　FPN 网络文本多维度特征提取实验结果

Fig. 7　Experimental results of FPN text multidimensional feature extraction

将多维度特征 $\{P_1, P_2, P_3, P_4\}$ 送入通道融合策略 CFS 模块，通过 CFS_3、CFS_2、CFS_1 三个阶段建立高低维度特征图间的链式关系，生成增强特征 $\{O_3, O_2, O_1\}$，并将其进行维度拼接，生成融合特征 F，实验结果如图 8 所示。可以看出，经三阶段 CFS 策略后，提取到的文本信息逐级完备，实现了不同尺度特征的充分融合，语义表征更加丰富。



(a) O_3　(b) O_2　(c) O_1　(d) F

图 8　道融合策略 CFS 模块实验结果

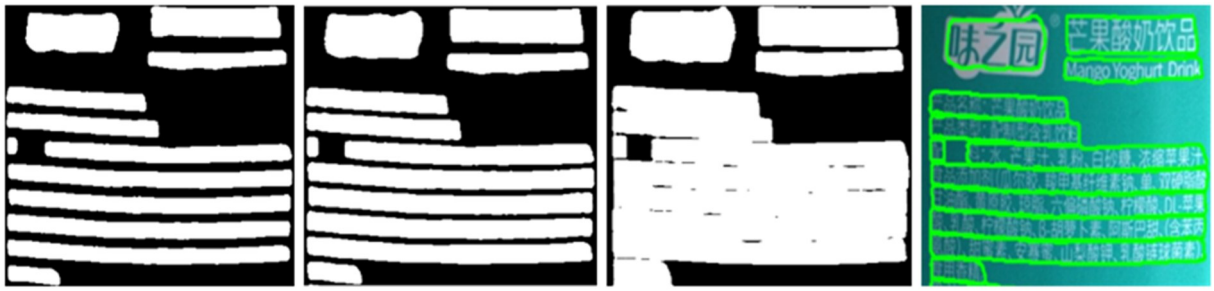
Fig. 8　Experimental results of CFS module

采用三种尺寸的文本核 S_1、S_2、S_3 对融合特征 F 进行文本实例分割，$S_1 < S_2 < S_3$，实验结果如图 9(a)、9(b)、图 9(c)所示；通过渐进式拓展方式生成文本候选框，得到最终预测结果，实验结果如图 9(d)所示。可以看出，密集文本均被正确检测，说明本文方法效果较好。

3.5　实验结果与分析

3.5.1　密集不规则文本

DAST1500 数据集实验结果如表 1 和图 10 所示。通过表 1 数据可知，本文方法在准确率、召回率和 F 值上较之前优秀的文本检测方法均有



(a) S_1分割结果　(b) S_2分割结果　(c) S_3分割结果　(d) 检测结果

(a) S_1 segmentation result　(b) S_2 segmentation result　(c) S_3 segmentation result　(d) Detection result

图 9　文本核 S_1、S_2、S_3 文本实例分割与最终检测结果

Fig. 9　Text kernel S1, S2, S3 text instance segmentation and final detection results

大幅的提升。与PSENet相比，仅采用TEM模块，可以将密集文本检测的F值提升到78.5%，其召回率提升1.6%，证明了TEM对提升检测召回率提升的良好作用；仅采用CFS策略，F值提升1.3%；既采用TEM模块也采用CFS策略，准确率、召回率和F值分别达到81.7%，81.9%和81.8%，各项评测指标的绝对提升率分别为2.9%，5.7和4.4%，本文方法在准确率、召回率及F值上均优于PSENet，证明了本文方法对密集不规则文本检测的有效性。

图10从左向右依次为样本的Ground Truth、PSENet检测结果、本文方法检测结果。如图中红色箭头所示，PSENet在极其复杂的密集场景文本下，出现了较为明显的漏检或局部的小文字漏检；本文方法因具有文本信息显著性增强能力，可以检测到更多更困难的密集文字样例，同时特征的有效融合也使局部小文本的漏检情况得到了一定的改善。综上说明本文方法整体上提升了密集不规则文本检测性能。

表1　DAST1500实验结果

Tab.1　Experimental results of DAST1500

Methed	TEM	CFS	Precision	Recall	F-score
EAST[13]			69.2	55.8	61.8
SegLink[7]			67.2	63.8	65.5
TextSnake[14]			73.6	72.1	72.8
PSENet[15]			78.8	76.2	77.4
Our Method	√		79.2	77.8	78.5
Our Method		√	79.3	78.1	78.7
Our Method	√	√	81.7	81.9	81.8

(a) 标注真值
(a) Ground truth

(b) PSENet结果
(b) Result of PSENet

(c) 本文方法结果
(c) Result of our method

图10　DAST1500检测结果

Fig10　Detection result of DAST1500

3.5.2　多方向文本

ICDAR2015数据集实验结果如表2和图11所示。通过表2数据可知，与PSENet相比，引入TEM模块，本文方法召回率提升0.4%，F值提升0.3%，进一步采用CFS，其F值提升2.1%，最终文本方法的准确率、召回率及F值达到

表 2　ICDAR2015 实验结果

Tab. 2　Experimental results of ICDAR2015

Methed	TEM	CFS	Precision	Recall	F-score
SegLink[7]			73.1	76.8	75.0
EAST[13]			83.6	73.5	78.2
PSENet[15]			81.5	79.7	80.6
TextSnake[14]			84.9	80.4	82.6
Our Method	√		81.7	80.1	80.9
Our Method		√	83.3	80.6	81.9
Our Method	√	√	84.4	80.7	83.0

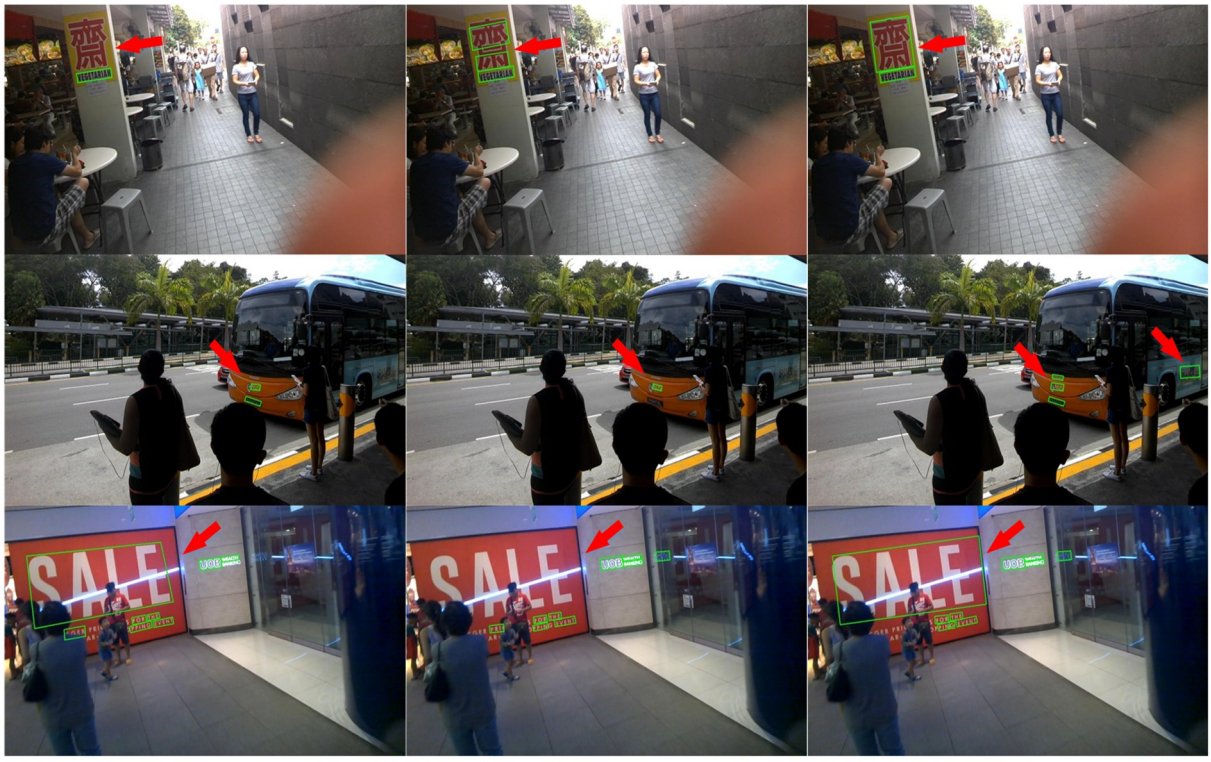
84.4%，80.7% 及 83.0%，比 PSENet 分别提升 2.9%，1.0% 及 2.4%。此外，与先前其他方法相比，文本方法在准确率与 F 值均优于之前先进的方法，召回率与其他方法相当，从而证明本文方法对多向文本检测的有效性。

通过对比图 11 中红色箭头所示区域，PSENet 对光线变化明显、模糊等复杂场景下部分文字区域检测不全甚至检测不到；文本方法充分利用文字特征，准确捕获 PSENet 错检或者漏检文字，能够关注到更具挑战性的文本样例，在一定程度上改善了自然场景多向文本的检测。

3.5.3　曲线文本

CTW1500 数据集实验结果如表 3 和图 12 所示。通过表 3 数据可知，相比 PSENet，仅采用 TEM，曲线文本检测的召回率提升 0.1%，仅采用 CFS，其准确率、召回率及 F 值分别提升了 1.1%，0.4% 和 0.7%。既采用 TEM 也采用 CFS 时，准确率、召回率及 F 值全面优于 PSENet，达到 82.1%，76.1% 和 79.0%。除此之外，本文方法在 F 值上均优于先前方法，具有从而证明本文方法对曲线文本检测的有效性。

通过对比图 12 中红色区域箭头所示，PSENet 提取图像中曲线文本特征能力不足，进而导致预测阶段图像中部分曲线文本检测不全，而本文方法可以合理利用图像，充分提取曲线文本特征，有效对其进行检测，在一定程度上提升了曲线文本检测的效果。



(a) 标注真值
(a) Ground truth

(b) PSENet结果
(b) Result of PSENet

(c) 本文方法结果
(c) Result of our method

图 11　ICDAR2015 结果

Fig. 11　Detection result of ICDAR2015

(a) 标注真值
(a) Ground truth

(b) PSENet结果
(b) Result of PSENet

(c) 本文方法结果
(c) Result of our method

图 12　CTW1500检测结果

Fig. 12　Detection result of CTW1500

表 3　CTW1500实验结果

Tab. 3　Experimental results of CTW1500

Methed	TEM	CFS	Precision	Recall	F-score
SegLink[7]			42.3	40.0	40.8
EAST[13]			78.7	49.1	60.4
TextSnake[14]			67.9	85.3	75.6
PSENet[15]			80.6	75.6	78.0
Our Method	√		80.5	75.9	78.1
Our Method		√	81.7	76.0	78.7
Our Method	√	√	82.1	76.1	79.0

4　结　论

本文提出了一种基于多维度卷积融合的密集不规则文本检测方法,该方法通过设计一种文本增强模块(TEM)提取全局文本特征,并编码进入卷积特征图以增强网络初期提取的文本特征,解决了密集不规则文本特征提取困难的问题,提高了密集文本检测的检测效果;提出了通道融合策略(CFS),在不同尺度特征间建立信息链,改善了融合不同尺度特征时直接维度拼接带来的信息损失。实验表明,该方法在密集文本数据集DAST1500上的F值达到81.8%,在倾斜文本数据集ICDA2015及曲线文本数据集CTW1500数据集上的F值分别为83.0%,79.0%,较PSENet分别提升了4.4%,2.4%和1.0%,验证了提出方法的有效性。

参考文献:

[1] 王建新,王子亚,田萱. 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(05):1465-1496.
WANG J X, WANG Z Y, TIAN X. Review of natural scene text detection and recognition based on deep learning [J]. *Journal of Software*, 2020, 31(05):1465−1496. (in Chinese)

[2] 白志程,李擎,陈鹏,郭立晴. 自然场景文本检测技术研究综述[J]. 工程科学学报,2020,42(11):1433-1448.
BAI ZH CH, LI Q, CHEN P, *et al*. Text detec-

tion in natural scenes: a literature review [J]. *Chinese Journal of Engineering*, 2020, 42(11): 1433-1448. (in Chinese)

[3] THILAGAVATHY A, CHILAMBUCHELVAN A. Fuzzy based edge enhanced text detection algorithm using MSER [J]. *Cluster Computing*, 2019, 22(05): 11681-11687.

[4] SHASWATA S, NEELOTPAL C, SOUMYADEEP K, *et al*. Multi-lingual scene text detection and language identification [J]. *Pattern Recognition Letters*, 2020, 138: 16-22.

[5] GHULAM J A, JAMAL H S, MUSSARAT Y, *et al*. A novel machine learning approach for scene text extraction [J]. *Future Generation Computer Systems*, 2018, 87: 328-340.

[6] TIAN Z, HUANG W L, H T, *et al*. Detecting text in natural image with connectionist text proposal network [C]. 14*th European conference on computer vision. Amsterdam, NETHERLANDS: Springer*, 2016: 56-72.

[7] SHI B G, BAI X, BELONGIE S. Detecting oriented text in natural images by linking segments [C]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA: IEEE*, 2017: 3482-3490.

[8] LIU Y L, JIN L W. Deep matching prior network: Toward tighter multi-oriented text detection [C]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA: IEEE*, 2017: 1962-1969.

[9] LIU Y L, JIN L W, ZHANG T S, *et al*. Curved scene text detection via transverse and longitudinal sequence connection [J]. *Pattern Recognition*, 2019, 90: 337-345.

[10] HE P, HUANG W L, HE T, *et al*. Single shot text detector with regional attention [C]. *Proceedings of the IEEE International Conference on Computer Vision. Venice, ITALY: IEEE*, 2017: 1962-1969.

[11] TANG J, YANG Z, WANG Y, *et al*. SegLink++: Detecting dense and arbitrary-shaped scene text by instance-aware component grouping [J]. *Pattern Recognition*, 2019, 96: 106954.

[12] ZHANG Z, ZHANG C Q, SHEN W, *et al*. Multi-oriented text detection with fully convolutional networks [C]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA: IEEE*, 2016: 4159-4167.

[13] ZHOU X Y, YAO C, WEN H, *et al*. East: An efficient and accurate scene text detector [C]. *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA: IEEE*, 2017: 5551-5560.

[14] SHI B G, RUAN J Q, ZHANG W, J, *et al*. TextSnake: a flexible representation for detecting text of arbitrary shapes [C]. 15*th European conference on computer vision. Munich, Germany: Springer*, 2018: 19-35.

[15] LI X, WANG W H, HOU B W, *et al*. Shape robust text detection with progressive scale expansion network [C]. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, CA, USA: IEEE*, 2019.

[16] 余永维,韩鑫,杜柳青. 基于Inception-SSD算法的零件识别[J]. 光学 精密工程,2020,28(8):1799-1809.
YU Y W, HAN X, DU L Q. Target part recognition based Inception-SSD algorithm [J]. *Opt. Precision Eng*, 2020, 28(8): 1799-1809. (in Chinese)

[17] 范丽丽,赵宏伟,赵浩宇,等. 基于深度卷积神经网络的目标检测研究综述[J]. 光学 精密工程,2020,28(5):1152-1164.
FAN L L, ZHAO H W, ZHAO H Y, *et al*. Survey of target detection based on deep convolutional neural networks [J]. *Opt. Precision Eng*, 2020, 28(05): 1152-1164. (in Chinese)

[18] REN S Q, HE K M, 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.

[19] HE K M, ZHANG X Y, REN S Q, *et al*. Identity mappings in deep residual networks [C]. 14*th European conference on computer vision. Amsterdam, NETHERLANDS: Springer*, 2016: 630-645.

[20] 杨其利,周炳红,郑伟,等. 注意力卷积长短时记忆网络的弱小目标轨迹检测[J]. 光学 精密工程,2020,28(11):2535-2548.
YANG Q L, ZHOU B H, ZHENG W, *et al*. Trajectory detection of small targets based on convolutional long short-term memory with attention mechanisms [J]. *Opt. Precision Eng*, 2020, 28(11): 2535-2548. (in Chinese)

[21] KARATZAS D, GOMEZ-BIGORDA L, NICOLAOU A, *et al*. ICDAR 2015 competition on robust reading [J]. *International Conference on Document Analysis & Recognition*. 2015:1156-1160

[22] DENG J, DONG W, SOCHER R, *et al*. ImageNet: A large-scale hierarchical image database [C]. 2009 *IEEE conference on computer vision and pattern recognition. Miami Beach, FL, USA*: *IEEE*, 2009: 248-255.

[23] SHRIVASTAVA A, GUPTA A, GIRSHICK R. Training region-based object detectors with online hard example mining [C]. *Proceedings of the IEEE conference on computer vision and pattern recognition. Las Vegas, NV, USA*: *IEEE*, 2016: 761-769.

作者简介：

孟月波(1979－)，女，陕西省西安人，副教授，西安建筑科技大学硕士生导师，2014年获得西安交通大学大学工学博士学位，主要从事计算机视觉感知与理解，人工智能与智动化系统，建筑智能化技术领域研究。E-mail: mengyuebo@163.com

通讯作者：

刘光辉(1976－)，男，陕西省西安人，副教授，西安建筑科技大学硕士生导师，2016年获得西安建筑科技大学工学博士学位，主要从事计算机视觉感知与理解，建筑智能化技术领域研究。E-mail: guanghuil@163.com