

# 基于 Transformer 的报纸版面 分割方法研究

朱一凡, 高 华, 业 宁

(南京林业大学信息科学技术学院、人工智能学院, 江苏 南京 210037)

**[摘要]** 大数据背景下信息的检索与研究对海量传统纸媒的数字化提出了挑战, 得益于不断发展的计算机视觉与人工智能方法, DETR 模型可被应用于报纸版面分割。针对原模型在版面分割中存在的检测速度慢、参数量大及分类不精准等问题, 本文提出了采用 ShuffleNet V2 轻量级主干网络的改进模型, 该方法可有效提升计算效率并减少模型参数量, 从而缓解 Transformer 结构的计算压力。同时, 通过特征金字塔结构, 该模型能够充分融合全局信息及细节信息, 显著增强多尺度目标的识别能力。此外, 该模型还引入高效通道注意力(ECA)模块来提取关键目标特征, 以此有效抑制无关背景信息, 在保证分割性能的同时实现轻量化设计。实验结果表明, 改进模型在报纸版面分割任务中的参数量为 38.5 M, 帧率(FPS)高达 47.5 img/s,  $mAP_{0.5}$  达到了 0.806。与原 DETR 模型相比, 改进模型在参数量上减少了 2.8 M, 帧率提高了 28.3 img/s,  $mAP_{0.5}$  提升了 3.2%。本文提出的模型还可以为报纸版面的 OCR 识别提供前期技术支持。

**[关键词]** 版面分割, DETR, ShuffleNet V2, 特征金字塔, ECA 通道注意力

**[中图分类号]** TP391 **[文献标志码]** A **[文章编号]** 1001-4616(2025)01-0109-10

## Research on Newspaper Layout Segmentation Method Based on Transformer

Zhu Yifan, Gao Hua, Ye Ning

(College of Information Science and Technology & Artificial Intelligence, Nanjing Forestry University, Nanjing 210037, China)

**Abstract:** The retrieval and research of information in the context of big data poses a challenge to the digitalization of massive traditional paper media. Thanks to the continuous development of computer vision and artificial intelligence methods, DETR model can be applied to newspaper layout segmentation. In view of the problems existing in the original model in layout segmentation, such as slow detection speed, large number of parameters and inaccurate classification, this paper proposes an improved model using ShuffleNet V2 lightweight backbone network, which can effectively improve computing efficiency and reduce the number of model parameters, thus easing the computing pressure of Transformer structure. At the same time, through the feature pyramid structure, the model can fully integrate the global information and detail information, and significantly enhance the recognition ability of multi-scale targets. In addition, the model also introduces Efficient Channel Attention (ECA) module to extract key target features to effectively suppress irrelevant background information and achieve lightweight design while ensuring segmentation performance. The experimental results show that the parameter number of the improved model is 38.5 M, the frame rate (FPS) is up to 47.5 img/s, and the  $mAP_{0.5}$  is up to 0.806. Compared with the original DETR model, the improved model reduces the number of parameters by 2.8 M, increases the frame rate by 28.3 img/s and improves  $mAP_{0.5}$  by 3.2%. The model proposed in this paper can provide early technical support for OCR recognition of newspaper layout.

**Key words:** layout segmentation, DETR, ShuffleNet V2, Feature Pyramid Networks (FPN), ECA

在近百年时间里,报纸经历了由纸质到信息化过程,作为一种通用信息载体,其很大程度上反应了特定时期的社会发展动态. 这对于研究社会历史进程具有不可忽视的积极作用. 然而在高度信息化的现今,传统纸媒报纸信息的检索却对研究人员提出了挑战,这对研究人文以及大数据信息仓库的建立是不利的.

2014 年“媒体融合”成为中国报业发展的主旋律,报纸的发行方式也由传统的印刷发行转为数字分发. 图 1 详细展示了 1978—2023 年全国报纸出版种类和总印数统计数据. 值得注意的是,在 2014 年“媒体融合”战略之前的纸媒报纸存量极为庞大,同时,出版种类从 1978 年的 186 种最高上涨到了 1996 年的 2 163 种,且到 2014 年一直保持在 2000 种左右. 这样数量与种类均非常庞大的传统纸媒报纸的数字化仅仅依靠人工并不经济且是低效率的.

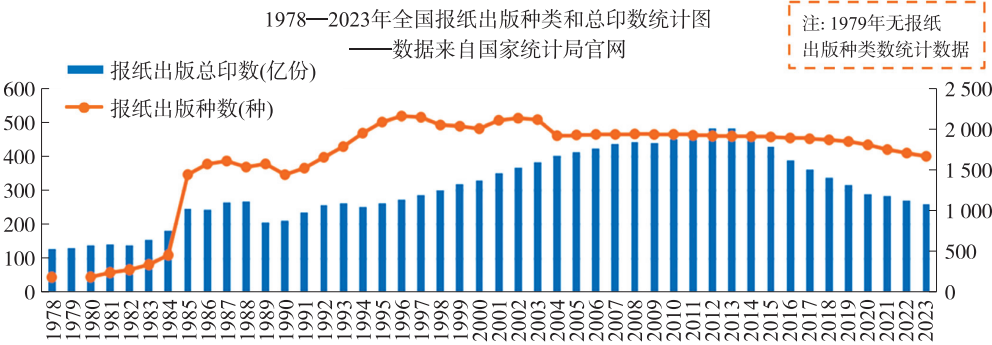


图 1 1978—2023 年全国报纸出版种类和总印数

Fig. 1 Types of newspapers published in the country and total print run, 1978—2023

计算机视觉技术、人工智能的不断发展,使得版面分割(Layout Segmentation)、光学文字识别(Optical Character Recognition, OCR)等技术可以有效推进纸媒的数字化. 通常,纸媒版面的数字化需要经历版面分割、文字识别、数据校对和数据建引这几个流程. 而版面分割作为扫描件数字化最重要的一个步骤,其通常利用目标检测(Object Detection)的方式找出图像中所有感兴趣的目标,确定它们的类别和位置,这有助于确定板块之间的关系来简化后续分析和识别阶段的复杂度.

早期,经典的文档布局分析多采用基于规则的方法<sup>[1-5]</sup>,这些方法需精心设计和调整,以适配多样文档. 其优势在于逻辑清晰,便于理解和调试. 然而,面对复杂或非标准布局的文档时,这些规则可能鲁棒性不强且不够灵活. 随着机器学习技术尤其是深度学习的兴起,现代方法倾向于使用数据驱动模型,如卷积神经网络(CNNs)、循环神经网络(RNNs)以及 Transformer 模型,这些模型能够自动学习复杂的特征和模式,从而在很多情况下超越了传统的基于规则的方法. 现在文档布局分析被视为文档对象检测问题<sup>[4-11]</sup>,即版面分割. 版面分割研究主要围绕目标检测算法展开<sup>[12-14]</sup>. 目标检测早期的传统方法是滑动窗口和人工特征提取<sup>[15]</sup>,随着 2014 年深度学习的兴起,出现了双阶段检测算法(Two-stage),主要通过选择性搜索(Selective Search)或者 Edge Boxes<sup>[16]</sup>等算法对输入图像选取可能包含检测目标的候选区域(Region Proposal)<sup>[17]</sup>,再对候选区域进行分类和位置回归以得到检测结果,代表算法有 R-CNN<sup>[18]</sup>及其变形等. 2016 年出现了单阶段检测算法(One-Stage),采用回归分析思想,也称为基于回归分析的目标检测算法,省略了候选区域生成阶段,直接得到目标分类和位置信息<sup>[19]</sup>,代表算法有 YOLO<sup>[20]</sup>系列和 SSD<sup>[21]</sup>系列. 2020 年出现的 DETR<sup>[22]</sup>模型,是 Transformer 在目标检测领域的首次应用. DETR 模型将目标检测当作一个集合预测问题,使用 Transformer 编码器对整个图像进行编码,使得模型具有全局感知能力. 相较于传统的目标检测方法,DETR 模型摒弃了预定义的锚框或候选框,也不需要非极大值抑制(NMS)来过滤重叠的框. 相反,它使用一个解码器来直接输出目标的边界框和类别,从而简化了目标检测流程. DETR 模型可以进行端到端的训练,即将图像和目标检测任务作为一个整体进行训练. 这样可以减少训练过程中的手动调整和设计,简化了模型的训练流程.

Transformer 模型已被广泛应用于多个领域,如铆接缺陷检测<sup>[23]</sup>、交通标志识别<sup>[24]</sup>、水下图像目标检测<sup>[25]</sup>、车辆目标检测<sup>[26]</sup>、地震数据断层识别<sup>[27]</sup>、情感分析<sup>[28]</sup>等. 然而,将 DETR 模型应用于版面分割领域的研究较少,本文将探索基于 Transformer 的报纸版面分割的新的检测思路. 针对 DETR 模型在版面分割中存在的检测速度慢、参数量大及分类不精准等问题,本文做出了改进. 改进模型采用 ShuffleNet V2 轻量

级主干网络提升计算效率并减少参数量,缓解 Transformer 的计算压力.同时,通过特征金字塔结构融合全局信息与细节信息,增强多尺度目标的识别能力.此外,模型还引入了 ECA 通道注意力模块,专注提取关键目标特征并抑制无关背景信息,实现性能与轻量化的平衡.实验证明,改进模型在版面分割任务中表现优异,为基于 Transformer 的目标检测提供了新的思路和方法.

## 1 模型设计

DETR 模型(图 2)的网络结构主要包含三个部分:卷积神经骨干网络、Transformer 的编码器和解码器(Encoder、Decoder)、前馈网络(FFN).首先,ResNet 卷积神经网络作为骨干网络,负责从输入图像中提取特征并缩减其尺寸,从而减轻后续 Transformer 在处理目标检测任务时的计算压力. Transformer 模型基于编码器-解码器(Encoder-Decoder)架构,通过堆叠多个编码器解码器模块来搭建深度神经网络. FFN 的主要功能是预测边界框的标准化中心坐标、高度和宽度,并使用 softmax 函数激活获得预测类标签.

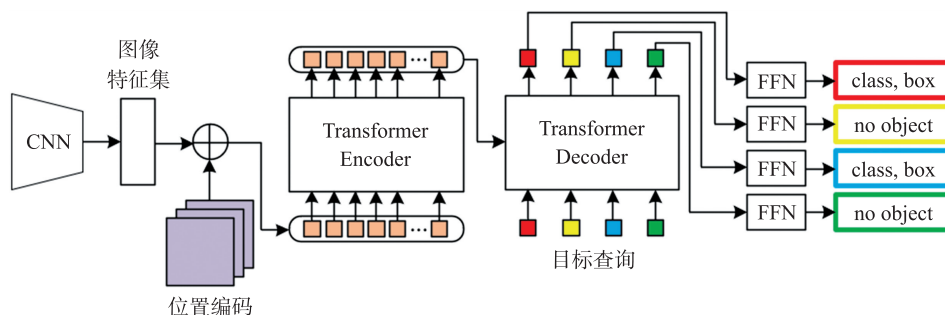


图 2 DETR 的网络结构

Fig. 2 Network structure of DETR

改进的 DETR 模型整体架构由 Backbone、特征金字塔、特征提取分支、Transformer 和 Prediction Heads 构成,如图 3 所示,将改进的模型命名为 SFE\_DETR. 该模型以轻量级主干 ShuffleNet V2 为特征金字塔的主干构造 Backbone,同时构造 CNN 特征提取分支,弥补 Transformer 中特征细节不足的问题.

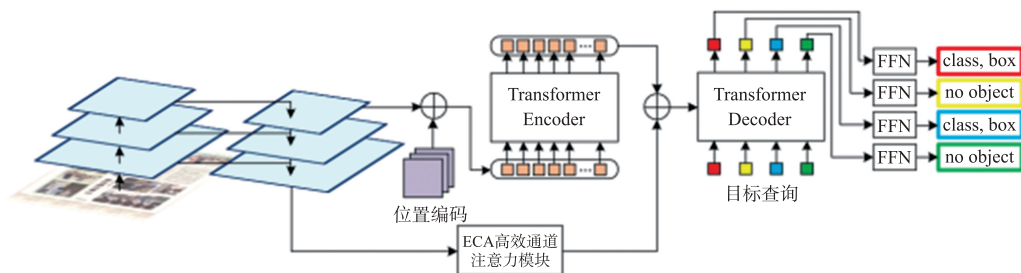


图 3 改进的 DETR 结构

Fig. 3 Improved DETR structure

通过这种设计,使得 SFE\_DETR 能够在维持计算效率的同时,提升对于复杂场景中细节特征的捕捉能力.

### 1.1 ShuffleNet V2

在深入探讨模型性能与效率之间的平衡时,本文特别关注模型的参数量以及训练速度.为了在保证模型性能的同时,减少计算资源的消耗和加快训练进程,经过仔细权衡,决定选用轻量型的 backbone 架构——ShuffleNet V2. ShuffleNet V2<sup>[29]</sup> 以其高效的计算效率和较小的模型尺寸而闻名,它能够有效地减少模型的参数量,从而在保持较高精度的同时,显著提升训练速度.这一选择不仅有助于在有限的计算资源下快速训练模型,还使得模型更易于部署在边缘设备上,为实际应用提供便利.

ShuffleNet V2 的大致结构如图 4. 该网络首先通过卷积层对输入数据进行初步的特征提取,紧接着是最大池化操作,旨在降低特征图的维度.接下来的 Stage2、Stage3 和 Stage4 为提取的不同尺度的特征图,如图 5 所示. 该网络在这三个关键阶段采取了逐层深入的策略,每个 Stage 不仅负责捕捉特定尺度的特征信息,还通过通道混洗、分组卷积等内部结构的优化,有效减少了计算量和参数冗余,从而实现了高效的特征

提取. 可以发现 Stage2 到 Stage4 包含的特征依次减少,即逐层深入策略,其中 Stage2 含有最多的特征,表明其包含较多细节信息,而 Stage4 则含有最少的特征,说明其主要捕捉更高层次的抽象信息.

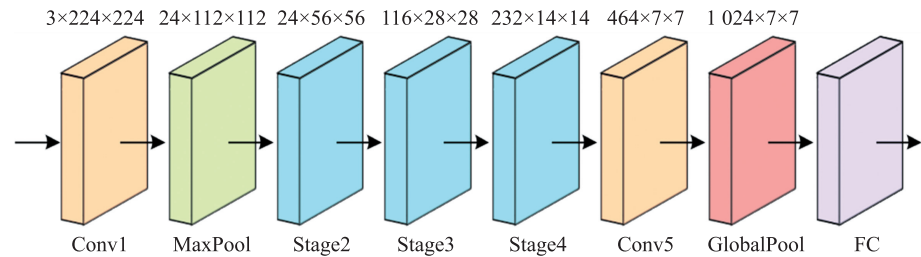


图 4 ShuffleNet V2 的结构

Fig. 4 Architecture of ShuffleNet V2

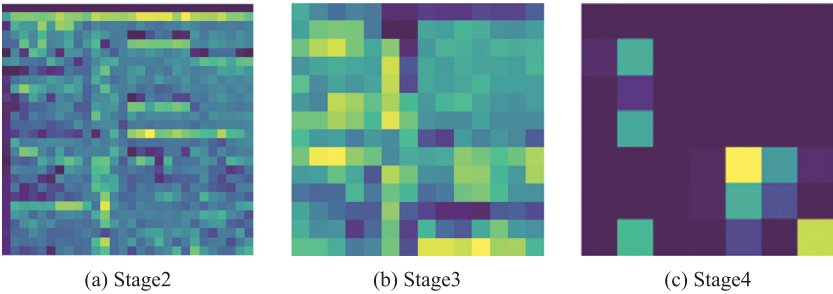


图 5 Stage2、Stage3、Stage4 的特征图

Fig. 5 Feature maps for Stage2, Stage3 and Stage4

为了有效利用这些不同尺度的特征并优化训练过程,采用特征融合操作显得尤为重要. 通过特征融合,可以整合各个阶段的特征,保证从细节到抽象各个层次的信息都被充分利用,这对于提高网络对复杂视觉任务的处理能力至关重要. 这种多尺度特征融合策略不仅增强了模型的表征能力,还有助于提升其在各种视觉识别任务中的性能表现.

1.2 FPN 特征金字塔

FPN( Feature Pyramid Networks)<sup>[30]</sup>是一个结构化的神经网络,用于提高对多尺度目标的识别能力,由三个部分组成:一个自底向上的线路、一个自顶向下的线路和横向连接( Lateral Connection). 在本研究中,自底向上的线路由 ShuffleNet V2 网络实现,利用其 Stage2、Stage3、Stage4 特征图进行特征融合,构建特征金字塔,如图 6.

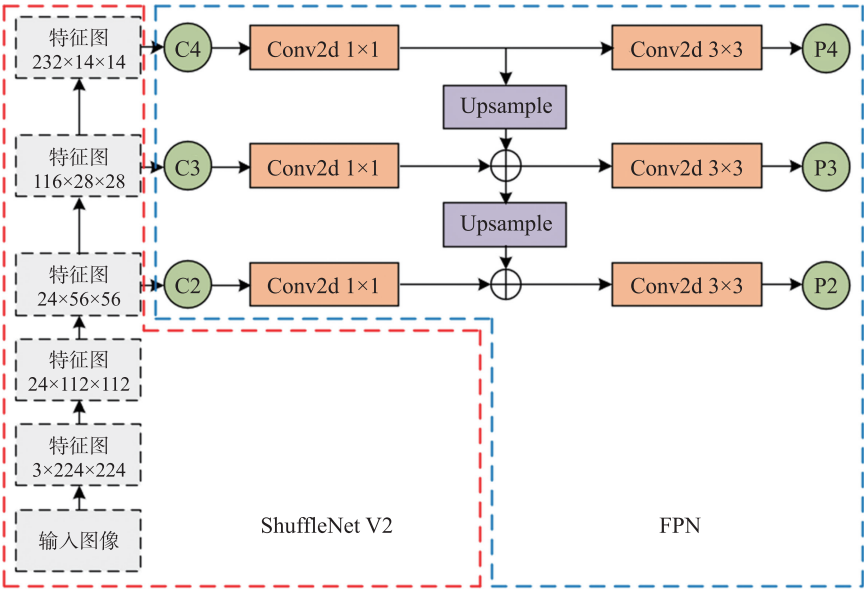


图 6 FPN 特征金字塔

Fig. 6 FPN feature pyramid



自顶向下的过程通过 2 倍的上采样 (Upsampling) 来实现,具体采用邻近插值算法进行空间尺寸的放大,确保上采样后的特征图与原特征图高宽相同。随后,采用横向链接将上采样得到的特征图与自底向上线路生成的相同尺寸的特征图进行融合 (Merge)。融合操作通过相加的方式完成。

为了消除上采样可能引入的混叠效应 (Aliasing Effect),每次融合后将结果应用  $3 \times 3$  卷积核进行卷积处理。假定融合后生成的特征图为 P2、P3、P4,这些新的特征层将与自底向上线路的 C2、C3、C4 特征层一一对应。

通过这种多层次的特征融合策略,FPN 能够有效地结合不同尺度的特征,从而增强模型对于不同尺度目标的识别能力,并丰富特征图的全局信息与细节信息,这对于复杂视觉任务的性能提升至关重要。

### 1.3 特征提取分支

在设计通道增强与编码器分支融合结构中,因为特征金字塔网络输出的不同尺寸特征图具有不同的信息密度和细节,所以构建两个主要分支:

#### (1) 小尺度特征图分支

选取 FPN 输出的小尺度特征图,这些图层虽然尺寸较小,但富含高层语义信息。将这些特征图展开成序列,输入到 Transformer 中,以学习全局上下文关系。Transformer 的能力在于通过自注意力机制,加强特征间的全局依赖,从而提取出更加丰富的语义信息。

#### (2) 大尺度特征图分支

此分支利用 FPN 输出的大尺度特征图,这些特征图含有丰富的细节信息,能够补充小尺度特征图在细节上的不足。鉴于目标检测背景的复杂性,此分支引入了 ECA (Efficient Channel Attention) 高效通道注意力模块<sup>[31]</sup>,专注于提取关键目标特征,同时抑制无关的背景信息。ECA 模块的优势在于简化的设计,它避免了传统注意力机制中的降维和升维操作,实现了高效和轻量化。

ECA 模块的结构如图 7,其核心思想是通过应用一维卷积来捕捉通道间的依赖关系,根据通道数自适应地计算一维卷积的核大小  $k$ ,计算公式如下:

$$k = \left\lceil \frac{\log_2(C)}{\gamma} + \frac{b}{\gamma} \right\rceil_{\text{odd}}, \quad (1)$$

其中  $C$  是输入特征的通道数,  $\gamma$  和  $b$  是超参数,用以精细控制卷积核的大小,从而适应不同的通道维度和特征复杂性。取绝对值并向下取整到最近的奇数是为了确保核大小是奇数。

得到核大小  $k$  后, ECA 模块将一维卷积应用于输入特征上,从而学习每个通道相对于其他通道的重要性,公式如下:

$$\text{out} = \text{Conv 1D}_k(\text{in}), \quad (2)$$

其中,  $\text{Conv 1D}_k(\text{in})$  表示核大小为  $k$  的一维卷积操作。

这种方法不仅增强了模型对重要通道的响应,还降低了对非关键信息的依赖,优化了整体的特征处理流程。

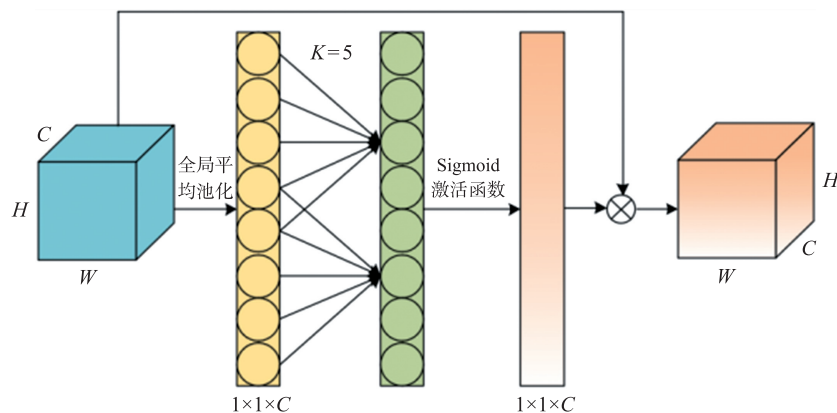


图 7 ECA 通道注意力模块

Fig. 7 ECA channel attention module

通过以上设计有效提高模型在复杂背景下对重要目标特征的识别能力,同时保持算法的运算效率,使其适合在计算资源受限的环境下使用,进一步拓宽了其应用范围。

## 2 实验设计

### 2.1 实验环境与参数设置

本文实验环境为配备 i5-13400F 处理器、32 GB 内存、一张显存为 12 GB 的 NVIDIA 3060 显卡的计算机设备。为了充分训练模型并使其参数得到精细调整,以期达到或接近最优解,设定迭代次数(Epoch)为 150 次。实验发现原模型在设定过大的批量大小(Batch Size)时会出现内存不足的错误,所以设置批量大小为 2,以确保模型能够在给定的硬件条件下稳定运行。设定学习率(Learning Rate)为 0.000 1,有助于模型在训练过程中逐步逼近最优解,同时避免过大的学习率导致的训练不稳定现象。

### 2.2 评价指标

本文评价指标选取目标检测中常见的 5 个指标:Params、FPS、mAP、mAP<sub>0.5</sub>、mAP<sub>0.75</sub>、Recall。

Params 是指网络模型中需要训练的参数总数,用来衡量模型的大小(计算空间复杂度),其计算公式为式(3):

$$Params = C_o \times (k_w \times k_h \times C_i + 1), \quad (3)$$

其中,  $C_o$  表示输出通道数,  $C_i$  表示输入通道数,  $k_w$  表示卷积核宽,  $k_h$  表示卷积核高。  $k_w \times k_h \times C_i$  表示一个卷积核的权重数量, +1 表示 bias, 括号表示一个卷积核的参数数量,  $C_o$  表示该层有  $C_o$  个卷积核。

FPS(每秒帧数)是衡量图像处理或模型推断速度的指标,通常用于评估计算机视觉应用程序的性能。FPS 表示在 1 s 内处理的图像帧数,其计算公式为式(4):

$$FPS = \frac{\text{处理的图像数量}}{\text{处理总时间}}. \quad (4)$$

mAP(mean of Average Precision)是指对所有类别的 AP 值求平均值。其中,平均精度(Average Precision, AP)表示单类别的模型平均准确度。对于目标检测任务,每一个类都可以计算出其 Precision 和 Recall,每个类都可以得到一条 P-R 曲线,曲线下的面积就是 AP 的值,计算公式如式(5):

$$AP = \int_0^1 P(r) dr. \quad (5)$$

mAP<sub>0.5</sub>指使用 IoU 阈值为 0.5 的边界框平均精度。同理, mAP<sub>0.75</sub>指使用 IoU 阈值为 0.75 的边界框平均精度。

Recall(召回率)又被称为查全率,表示预测结果为正样本中实际正样本数量占全样本中正样本的比例,计算公式如式(6):

$$Recall = \frac{TP}{TP + FN}, \quad (6)$$

其中,  $TP$  表示正确预测的正例,即真实值为正例,预测值也为正例;  $FN$  表示错误预测的反例,即真实值为正例,但被错误预测成了反例。

## 3 实验结果

### 3.1 数据集构建

为了在实际数据上验证本文的训练和测试效果,本文使用了江苏省档案馆提供的一系列《新锡山》报纸的扫描件。该数据涵盖 2010 年至 2019 年期间该报刊登的版面扫描件。为了提高训练模型的适用性,本文选取了其中具有代表性且能够全面反应数据集整体特征的 1 212 张报纸作为研究样本。在数据处理阶段,本文将报纸数据集按照 8:2 的比例随机的划分到训练集和测试集。

本文采用 LabelImg 对数据样本进行标签标注,按照传统报纸的排版和内容结构将报纸内容标签分为四大类:抬头(head)、标题(title)、正文(text)和图片(picture)。

### 3.2 对比试验

为了验证改进算法的有效性,选择了 Faster-RCNN、SSD、YOLOv3、DETR 与 SFE\_DETR 进行对比。实验

结果如表 1 所示.

表 1 实验结果  
Table 1 Experimental result

模型	Params/M	FPS/(img/s)	mAP	mAP <sub>0.5</sub>	mAP <sub>0.75</sub>	Recall
Faster-RCNN	41.4	14.2	0.563	0.725	0.618	0.642
SSD	<b>35.7</b>	42.9	0.410	0.684	0.428	0.509
YOLO v3	61.5	44.6	<b>0.568</b>	0.801	0.604	0.642
DETR	41.3	19.2	0.532	0.774	0.606	0.629
Deformable DETR	39.8	9.3	0.531	0.779	0.623	<b>0.668</b>
SFE_DETR	38.5	<b>47.5</b>	0.560	<b>0.806</b>	<b>0.629</b>	0.652

实验结果表明, SFE\_DETR 在 FPS、mAP<sub>0.5</sub>、mAP<sub>0.75</sub> 上均达到了最优性能, 尽管在 Params、mAP 以及 Recall 方面并未获得最佳表现, 但其与最优结果之间的差距并不显著, 这表明 SFE\_DETR 在多个关键指标上均展现出了出色的性能.

图 8 是一张报纸上的某个框预测的可视化. 每个框预测被表示为一个点, 其中心坐标在固定大小的空间中, 按每个图像大小归一化. 通过观察, 每个槽都有几种运行模式, 侧重于不同的区域和方框大小. 特别是, 所有槽都有预测图像范围内方框的模式 (可见图中的热力表现), 这或许与数据集中对象的分布有关. 几乎所有槽都具有预测数据集中常见的大型图像方框的模式.

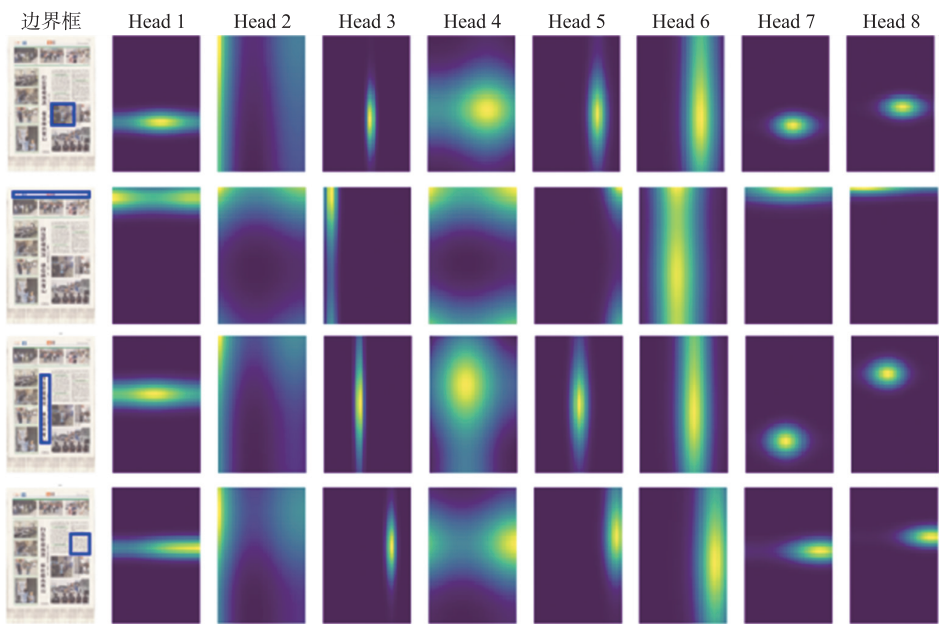


图 8 各个头部的空间注意力热图  
Fig. 8 Heat map of spatial attention for each head

SFE\_DETR 模型在迭代 150 轮过程中的损失曲线如图 9, 其中蓝色曲线是训练损失 (Train Loss), 橙色曲线是验证损失 (Val Loss). 从图中可以观察到, 随着迭代次数的增加, 训练损失和验证损失均呈现稳定的下降趋势. 且在整个训练过程中, 验证损失并没有显示出与训练损失显著背离的趋势, 即没有出现验证损失停止下降或开始上升的情况, 说明没有出现拟合 (Overfit) 的典型迹象. 同时, 训练损失和验证损失均持续下降, 表明模型在训练集和验证集上的性能都在不断提升, 未出现欠拟合 (Underfit) 的迹象. 因此, 可以判断 SFE\_DETR 模型在训练过程中没有过拟合或欠拟合现象, 显示出良好的学习能力.

为了验证 SFE\_DETR 模型的泛化性, 选取人民日报、电脑

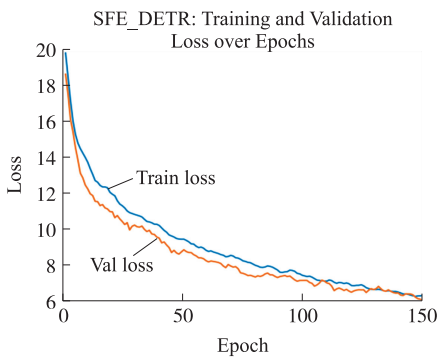


图 9 模型损失曲线  
Fig. 9 Model loss curve

报进行对比实验,实验结果见表 2.

表 2 不同数据集对比试验  
Table 2 Comparative experiments with different data sets

模型	人民日报		电脑报	
	mAP <sub>0.5</sub>	Recall	mAP <sub>0.5</sub>	Recall
Faster-RCNN	0.783	0.589	0.791	0.637
SSD	0.697	0.501	0.730	0.512
YOLO v3	0.816	0.591	0.858	0.645
DETR	0.801	0.576	0.849	<b>0.680</b>
Deformable DETR	0.816	<b>0.608</b>	0.841	0.643
SFE_DETR	<b>0.821</b>	0.598	<b>0.864</b>	0.654

可以观察到 SFE\_DETR 模型在不同类型报纸的版面分割任务中均展现出了卓越的性能,这充分表明了该模型具备良好的泛化能力.

3.3 消融实验

对算法的关键模块进行消融实验,验证算法增加模块的有效性. 包括 ShuffleNet V2、FPN 和 ECA 模块,并记录了每个模块被消融后的实验结果. 实验结果如表 3 所示.

表 3 消融实验结果  
Table 3 Ablation experimental results

模型	Params/M	FPS/(img/s)	mAP	mAP <sub>0.5</sub>	mAP <sub>0.75</sub>	Recall
DETR	41.3	19.2	0.532	0.774	0.606	0.629
+FPN	61.7	17.8	0.542	0.793	0.621	0.640
+FPN+ECA	61.7	17.6	0.576	0.813	0.655	0.663
SFE_DETR	38.5	47.5	0.560	0.806	0.629	0.652

实验结果表明,SFE\_DETR 的性能表现总体优于 DETR. 引入特征金字塔模块和特征提取分支可以有效提升模型的性能表现,但是参数量明显变大,增大了 20.4 M,其中主要是特征金字塔带来的模型参数量增加. 引入 ShuffleNet V2 后,显著降低了参数量且 FPS 显著提升,参数量减少了 23.2 M, FPS 提升了 29.9 img/s,与只引入特征金字塔模块和特征提取分支的模型对比,SFE\_DETR 的 mAP<sub>0.5</sub>并没有显著降低. 最终,SFE\_DETR 比 DETR 参数量减少 2.8 M,FPS 提高 28.3 img/s,mAP、mAP<sub>0.5</sub>、mAP<sub>0.75</sub>分别提升2.8%、3.2%、2.3%,Recall 提高了 2.3%.

最终呈现的版面分割效果图如图 10 所示,抬头用蓝色表示、标题用橙色表示、正文用绿色表示、图片用紫色表示.

4 结论

本文针对报纸版面分割,提出了 SFE\_DETR 模型. 本文特别关注模型的参数量以及训练速度. 为了在保证模型性能的同时减少计算资源的消耗和加快训练进程,经过仔细权衡,决定选用轻量型的 backbone 架构——ShuffleNet V2. 该模型通过引入 FPN 特征金字塔结构,显著增强了模型对于不同尺度目标的识别能力,同时丰富了特征图的全局信息与细节信息. 此外,该模型进一步增加了特征提取分支,



图 10 版面分割效果图

Fig. 10 Layout segmentation effect



使用 ECA 高效通道注意力模块,提取关键目标特征,同时抑制无关的背景信息. 其简化的设计保证了模型的高效性和轻量化. 实验结果表明,相较于原 DETR 网络和其他经典版面分割网络,SFE\_DETR 模型在版面分割任务上展现出了更加优越的性能,并且拥有较快的速度. 该算法在不同数据集上的验证均表现优异,具有良好的泛化性,有较好的版面分割应用前景.

#### [参考文献]

- [1] COÜASNON B,LEMAITRE A. Recognition of tables and forms[J]. Handbook of document image processing and recognition, 2019:647-677.
- [2] ZANIBBI R,BLOSTEIN D,CORDY J R. A survey of table recognition:models,observations,transformations,and inferences[J]. Document analysis and recognition,2004,7:1-16.
- [3] E SILVA A C,JORGE A M,TORGO L. Design of an end-to-end method to extract information from tables[J]. International journal of document analysis and recognition(IJDAR),2006,8:144-171.
- [4] KHUSRO S,LATIF A,ULLAH I. On methods and tools of table detection,extraction and annotation in PDF documents[J]. Journal of information science,2015,41(1):41-57.
- [5] EMBLEY D W,HURST M,LOPRESTI D, et al. Table-processing paradigms:a research survey[J]. International journal of document analysis and recognition(IJDAR),2006,8:66-86.
- [6] CESARINI F,MARINAI S,SARTI L, et al. Trainable table location in document images[C]//2002 International Conference on Pattern Recognition. Quebec, Canada:IEEE,2002,3:236-240.
- [7] YANG X,YUMER E,ASENTE P, et al. Learning to extract semantic structure from documents using multimodal fully convolutional neural networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA:IEEE,2017:5315-5324.
- [8] HE D,COHEN S,PRICE B, et al. Multi-scale multi-task fcn for semantic page segmentation and table detection[C]//2017 14th IAPR International Conference on Document Analysis and Recognition(ICDAR). Kyoto, Japan:IEEE,2017,1:254-261.
- [9] 孙皓月. 基于深度学习的文档版面分析方法研究[D]. 福建:厦门理工学院,2022.
- [10] 张洪红. 基于注意力机制的文档图像版面分析算法[D]. 山东:青岛科技大学,2023.
- [11] 杨陈慧,周小亮,张恒,等. 基于 Multi-WHFPN 与 SimAM 注意力机制的版面分割[J]. 电子测量技术,2024,47(1):159-168.
- [12] 付苗苗,邓淼磊,张德贤. 基于深度学习和 Transformer 的目标检测算法[J]. 计算机工程与应用,2023,59(1):37-48.
- [13] 李沂杨,陆声链,王继杰,等. 基于 Transformer 的 DETR 目标检测算法研究综述[J]. 计算机工程,2025:1-20.
- [14] 李建,杜建强,朱彦陈,等. 基于 Transformer 的目标检测算法综述[J]. 计算机工程与应用,2023,59(10):48-64.
- [15] ZOU Z,CHEN K,SHI Z, et al. Object detection in 20 years:a survey[J]. Proceedings of the IEEE,2023,111(3):257-276.
- [16] ZITNICK C L,DOLLÁR P. Edge boxes:locating object proposals from edges[C]//Computer Vision-ECCV 2014:13th European Conference. Zurich, Switzerland:Springer International Publishing,2014:391-405.
- [17] HU Q,ZHAI L. RGB-D image multi-target detection method based on 3D DSF R-CNN[J]. International journal of pattern recognition and artificial intelligence,2019,33(8):1954026.
- [18] GIRSHICK R,DONAHUE J,DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH, USA:IEEE,2014:580-587.
- [19] 许德刚,王露,李凡.深度学习的典型目标检测算法研究综述[J]. 计算机工程与应用,2021,57(8):10-25.
- [20] REDMON J,DIVVALA S,GIRSHICK R, et al. You only look once:unified,real-time object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Vegas, NV, USA:IEEE,2016:779-788.
- [21] LIU W,ANGUELOV D,ERHAN D, et al. Ssd:single shot multibox detector[C]//Computer Vision-ECCV 2016:14th European Conference. Amsterdam, The Netherlands:Springer International Publishing,2016:21-37.
- [22] CARION N,MASSA F,SYNNAEVE G, et al. End-to-end object detection with transformers[C]//European Conference on Computer Vision. Cham:Springer International Publishing,2020:213-229.
- [23] 李宗刚,宋秋凡,杜亚江,等. 基于改进 DETR 的机器人铆接缺陷检测方法研究[J]. 铁道科学与工程学报,2024,21(4):1690-1700.
- [24] 徐浩东. 基于 DETR 的自动驾驶汽车交通标志识别系统研究[D]. 陕西:西京学院,2022.

- [25] 崔颖,韩佳成,高山,等. 基于改进 Deformable-DETR 的水下图像目标检测方法[J]. 应用科技,2024,51(1):30-36,91.
- [26] 江志鹏,王自全,张永生,等. 基于改进 Deformable DETR 的无人机视频流车辆目标检测算法[J]. 计算机工程与科学, 2024,46(1):91-101.
- [27] 武庭润,高建虎,常德宽,等. 基于 Transformer 的地震数据断层识别[J]. 石油地球物理勘探,2024,59(6):1217-1224.
- [28] 冯程,杨海,王淑娴,等. 基于自上而下掩码生成与层叠 Transformer 的多模态情感分析[J]. 计算机工程与应用,2025: 1-11.
- [29] MA N,ZHANG X,ZHENG H T,et al. Shufflenet v2:practical guidelines for efficient cnn architecture design[C]//Proceedings of the European Conference on Computer Vision(ECCV). Munich,Germany:Springer,2018:116-131.
- [30] LIN T Y,DOLLÁR P,GIRSHICK R,et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu,HI,USA:IEEE,2017:2117-2125.
- [31] WANG Q,WU B,ZHU P,et al. ECA-Net:efficient channel attention for deep convolutional neural networks[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle,WA,USA:IEEE,2020:11534-11542.

[责任编辑:杜忆忱]