DOI: 10.13382/j.jemi.B2508444

改进 YOLOv11n 的高效交通实例分割算法*

邵自强 魏利胜 武 涛
(安徽工程大学电气工程学院 芜湖 241000)

摘 要:针对交通场景下目标分割精度低和掩膜质量差的问题,提出一种改进 YOLOv11n 的高效交通实例分割算法。首先,在主干网络的 C3k2 模块中融合小波变换卷积 WTConv,构建 C3k2-WTConv 模块,以高效扩展感受野并增强低频特征提取;其次,设计特征交互增强 AIFI-LA 模块,降低快速空间金字塔池化(SPPF)的多尺度计算冗余,并提高处理长序列和保留关键特征信息能力;然后,提出特征重校准 EMCSA 模块,并嵌入至特征重组上采样算子(CARAFE)中,构建 CARAFE-EMCSA 模块重构上采样,以增强环境特征的捕获能力和特征图的整体判别性;最后,将 Soft-NMS 与 DIoU-NMS 相融合并替换原非极大值抑制算法(NMS),在保留更多高质量边界框的同时,利用相对位置信息进一步优化选择,提升边界框精度. 实验结果表明,在 Cityscapes 数据集上,与原模型相比,边界框精度 mAP@0.5 和 mAP@0.5:0.95 值分别提高了 9.2%和 8.5%,分割掩膜精度 mAP@0.5 和 mAP@0.5:0.95 值分别提高了 10.6%和 8.8%,在 BDD100K 数据集上,边界框精度 mAP@0.5 和 mAP@0.5:0.95 值分别提高了 5.1%和 7.4%,分割掩膜精度 mAP@0.5 和 mAP@0.5:0.95 值分别提高了 4.5%和 6.6%. 由此可知,所提方法在交通场景分割方面的有效性.

关键词: 交通场景;目标分割;小波变换;注意力机制;NMS 算法;YOLOv11n
中图分类号: TP391.4; U495 **文献标识码**: A **国家标准学科分类代码**: 510.40

Efficient traffic instance segmentation algorithm based on improved YOLOv11n

Shao Ziqiang Wei Lisheng Wu Tao
(School of Electrical Engineering, Anhui Polytechnic University, Wuhu 241000, China)

Abstract: To address the problems of low target segmentation accuracy and poor mask quality in traffic scenes, an improved YOLOv11n efficient traffic instance segmentation algorithm, ETIS-YOLO, is proposed. Firstly, the C3k2-WTConv module is constructed by fusing the Wavelet Transform Convolution into the C3k2 module of the backbone network to efficiently expand the receptive field and enhance low-frequency feature extraction; Secondly, the feature interaction enhancement AIFI-LA module is designed to reduce the multi-scale computational redundancy of spatial pyramid pooling-fast (SPPF) and improve its ability to handle long sequences and preserve key feature information; Additionally, the feature recalibration EMCSA module is proposed and embedded into the up-sampling operator content aware reassembly of features (CARAFE) to form a CARAFE-EMCSA module, which reconstructs the up-sampling process to enhance the capture of contextual features and the overall discriminability of feature maps; Finally, Soft-NMS and DIoU-NMS are fused and replaced with the original non-maximum suppression (NMS), which further optimizes the selection and improves the accuracy of the bounding boxes by utilizing relative position information while retaining more high-quality bounding boxes. The experimental results show that on the cityscapes dataset, the bounding box accuracy mAP@0.5 and mAP@0.5:0.95 values are improved by 9.2% and 8.5%, and the segmentation mask accuracy mAP@0.5 and mAP@0.5:0.95 values are improved by 10.6% and 8.8%, respectively, compared with the YOLOv11n model; on the BDD100K dataset, the bounding box accuracy mAP@0.5 and mAP@0.5:0.95 values are improved by 5.1% and 7.4%, and the segmentation mask accuracy mAP@0.5 and mAP@0.5:0.95 values are improved by 4.5% and 6.6%, respectively. It can be seen that the proposed method is effective in traffic scene segmentation.

收稿日期: 2025-06-06 Received Date: 2025-06-06
*基金项目:安徽省教育厅重大项目(KJ2020ZD39)、安徽省高等学校省级质量工程项目(2023cxtd057)资助

Keywords: traffic scene; target segmentation; wavelet transform; attention mechanism; NMS algorithm; YOLOv11n

0 引 言

智能交通系统环境感知能力的提升,其核心任务之一在于实现高精度与高效率的场景分割。然而,交通场景中目标的极端多尺度特性[1]、频繁的遮挡现象[2]以及复杂的背景干扰[3],共同导致了特征提取不充分、分割精度有限与掩膜边缘粗糙等关键瓶颈,严重制约了现有算法的实际应用价值。因此,场景分割在智能交通系统中具有重要的研究价值。

交通场景分割算法中,以深度学习为主要研究方向,现有的深度学习目标分割算法根据目标的处理粒度和任务目标主要可以分为语义分割算法和实例分割算法两大类。语义分割算法专注于对图像中的每个像素进行分类,以区分不同的语义类别,如全卷积网络(fully convolutional network, FCN)[4]、U-Net[5]和 DeepLab 系列[6],而实例分割算法则进一步为每个目标生成精确的分割掩码,同时区分同一类别中的不同实例,如 Mask R-CNN[7]、YOLACT[8]和 YOLO 系列[9],这些算法在交通场景分割中得到了广泛应用,并衍生出了许多改进版本,以适应不同的场景分割需求和对模型性能的提高。

针对交通场景下分割精度低问题,Zhang 等[10]提出了一种增强型 FCN 算法,通过改进传统特征交互的上下文建模方法,并嵌入至 FCN 网络头部,以增强交通场景下分割精度;Zhu 等[11]提出一种改进 U-Net 的交通场景分割算法,通过在不同网络层嵌入不同的注意力机制,并用深度可分离卷积取代所有传统卷积,以在提高分割精度的同时减少部分网络参数量;Liu 等[12]提出了一种改进 Deeplabv3+的交通场景分割算法,其认为道路场景不同水平区域的像素级分布存在较大差异,通过在网络中嵌入垂直空间注意力以提高分割精度;Fang 等[13]提出了一种改进 Mask R-CNN 的交通场景分割算法,通过选取特征提取更高效的主干网络,并加入高效的通道注意力模块,以提高分割精度;Li 等[14]提出了一种改进 YOLACT 的交通场景分割算法,通过更换更高效的主干网络等来提高分割精度。

然而,上述改进因选取模型参数量大,难以在交通场景下实现实时性分割要求。为兼顾交通场景下分割实时性与精度问题,Xia 等[15]提出了一种改进 YOLOv5n 的交通场景分割算法,采用双向跨尺度连接优化方法,细化掩码流,以提升分割精度;赵南南等[16]提出了一种 DE-YOLO 的交通场景分割算法,并通过引入高效多尺度注意力机制和可变形卷积,以提升分割精度;Gu 等[17]提出了一种改进 YOLOv8n 的交通场景分割算法,通过优化掩码解码结构提高模型实时性,并增强多尺度特征学习,在实现算法实时性同时以提升分割精度。尽管上述提出的改进 YOLO 系列的分割算法在一定程度上兼顾了交通场景分割的实时性和精度,但交通场景的复杂性对现有算法仍然具有挑战性。

上述方法主要针对在交通场景下如何提升目标分割精度问题进行研究,但研究依然存在目标分割精度低、掩膜分割质量差和模型参数大的问题。由于现有的许多实例分割方法模型结构庞大,难以满足实时性要求较高的交通场景,为此,本文选取 YOLOv11n 作为基底模型,提出了一种高效的交通实例分割算法——ETIS-YOLO,以提升交通场景分割精度和掩膜分割质量。

为了改善 C3k2 在面向复杂交通背景时,感受野扩展效率和低频特征捕捉能力方面的不足,本文通过融入小波变换卷积(wavelet transform convolution, WTConv),以提高感受野扩展效率和对低频特征捕捉能力。设计了 AIFI-LA 替换快速空间金字塔池化(spatial pyramid pooling-fast, SPPF),以改善 SPPF 的冗余计算和难以建立长距离依赖关系问题,并提高保留关键特征信息能力。提出特征重校准 EMCSA 模块,并结合特征重组上采样算子(content aware reassembly of features, CARAFE),替换并重构原上采样,在捕获和保留更多的环境特征的同时,分别对特征图进行通道和空间重校准,再次提升对特征图的整体判别能力。构建复合非极大值抑制(non-maximum suppression, NMS)算法,通过将 Soft-NMS 与 DIoU-NMS 相融合,实现在保留更多边界框的同时,利用边界框之间的距离信息,以进一步优化边界框的选择,提高边界框精度。

1 算法总体框架

YOLO 系列算法以其高效、准确和快速的特点在计算机视觉领域取得了显著成就。YOLOv11n 作为该系列典型代表,是目前最为先进和最为经典的目标分割模型之一,其由主干网络、颈部网络和分割头组成。主干网络由 Conv 模块、C3k2 模块、SPPF 模块和 C2PSA 模块组成,其主要作用是用于提取图像目标的特征信息;颈部网络的主要作用是对主干网络不同层的特征图进行特征融合;分割头的主要作用是根据颈部网络输出的多尺度特征图尺寸,对其进行目标分类、边界框回归和掩膜预测。虽然 YOLOv11n 表现出良好的分割效果,但由于其本身是为检测任务设计的,网络结构在处理像素级分割任务时,缺乏对一些细节信息的充分捕捉能力,在分割细节上存在一些不足。因此,对 YOLOv11n 进行改进,改进后的模型结构如图 1 所示。

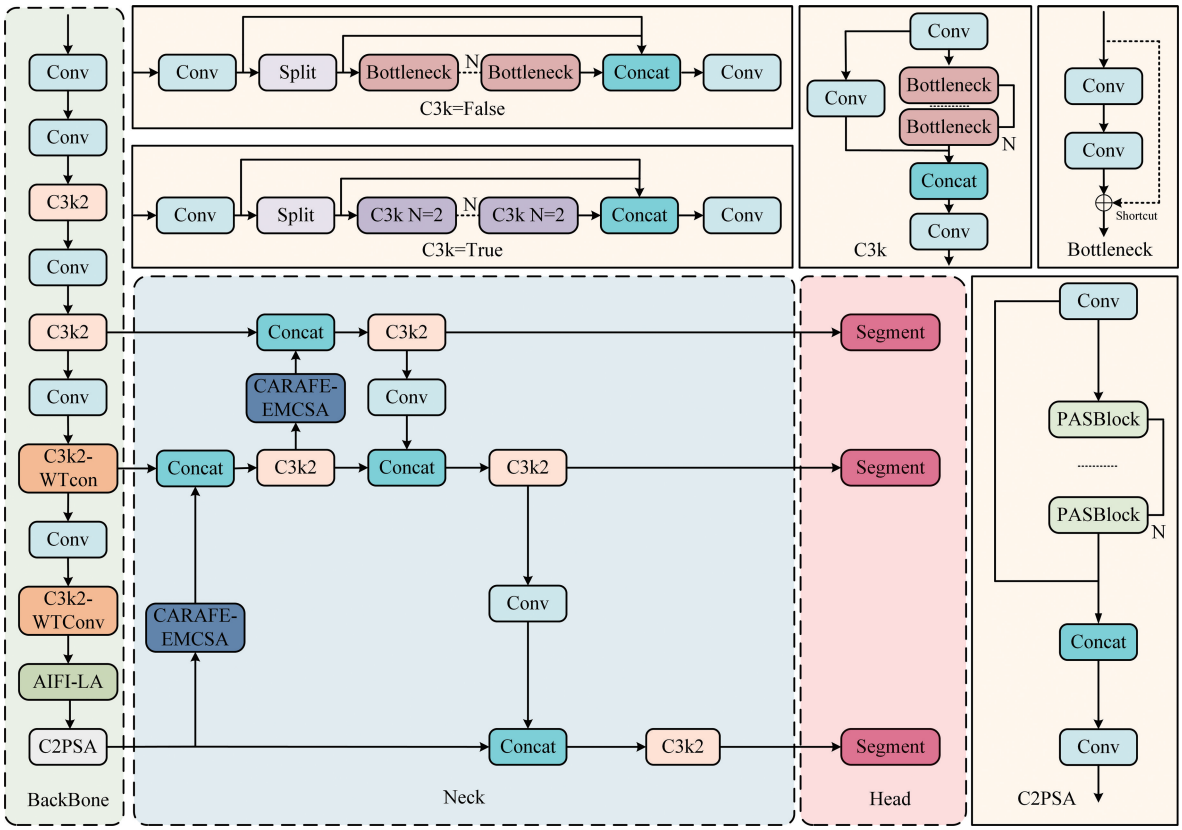


图 1　ETIS-YOLO 的网络结构

Fig. 1　The network structure of ETIS-YOLO

1.1　特征提取改进

在复杂交通场景中，更大的感受野可以帮助模型更好地理解场景的整体结构，从而提高分割的准确性。此外，复杂交通场景中存在大量的低频信息，这些信息对模型理解交通场景的整体布局至关重要。C3k2-BottleNeck 作为 YOLOv11 中的一种特征提取组件，结合了可变卷积核和通道分离策略，实现了多尺度特征融合，其通过使用不同大小的卷积核来扩展感受野，以捕捉更广泛的上下文信息，使模型在处理大物体和复杂背景的场景时表现更好。然而，C3k2-BottleNeck 的感受野扩展方式是线性增长的，效率较低。同时，传统卷积操作对低频特征的捕捉能力不足，缺乏频率域处理机制，导致其在复杂道路场景中的表现受限。

为了改善 C3k2-BottleNeck 在感受野扩展效率和低频特征捕捉能力方面的不足，改进模型通过在 C3k2-BottleNeck 位置嵌入小波变换卷积 WTConv[18] 构建 C3k2-WTConv，改进后的结构如图 2 所示。WTConv 通过小波变换对输入数据进行分解，使感受野的扩展与参数增长呈对数关系，而不是传统的平方增长，以提高了感受野扩展的效率。同时，WTConv 通过重复分解低频信息，并在多频率带进行卷积操作，提供丰富的多尺度信息，增强对低频特征的捕捉能力。为了保持初始特征的一致性，改进模型保留了骨干网的前两层不变，继续沿用 C3k2 模块，随着网络的递进，特征提取的精细度增加，对特征的判别能力提出了更高要求。因此，将骨干网中最后两个 C3k2 模块替换为 C3k2-WTConv 模块。

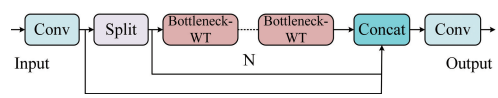


图 2　C3k2-WTConv 模块网络结构

Fig. 2　Network structure of C3k2 WTConv module

WTConv 卷积流程如图 3 所示，其核心思想是应用级联 WT 将输入信号递归分解为不同的频率分量，并对每个频率分量执行小核卷积运算，最后通过逆小波变换 *IWT* 合并结果。首先，WTConv 通过二维小波变换将一维小波变换的卷积核扩展为 4 个方向的滤波器，分别为 $\boldsymbol{f}_{LL}$、$\boldsymbol{f}_{LH}$、$\boldsymbol{f}_{HL}$ 和 $\boldsymbol{f}_{HH}$，如式(1)所示。

$$\boldsymbol{f}_{LL}=\frac{1}{2}\begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix},\quad \boldsymbol{f}_{LH}=\frac{1}{2}\begin{bmatrix}1 & -1\\ 1 & -1\end{bmatrix}$$

$$\boldsymbol{f}_{HL}=\frac{1}{2}\begin{bmatrix}1 & 1\\ -1 & -1\end{bmatrix},\quad \boldsymbol{f}_{HH}=\frac{1}{2}\begin{bmatrix}1 & -1\\ -1 & 1\end{bmatrix}\tag{1}$$

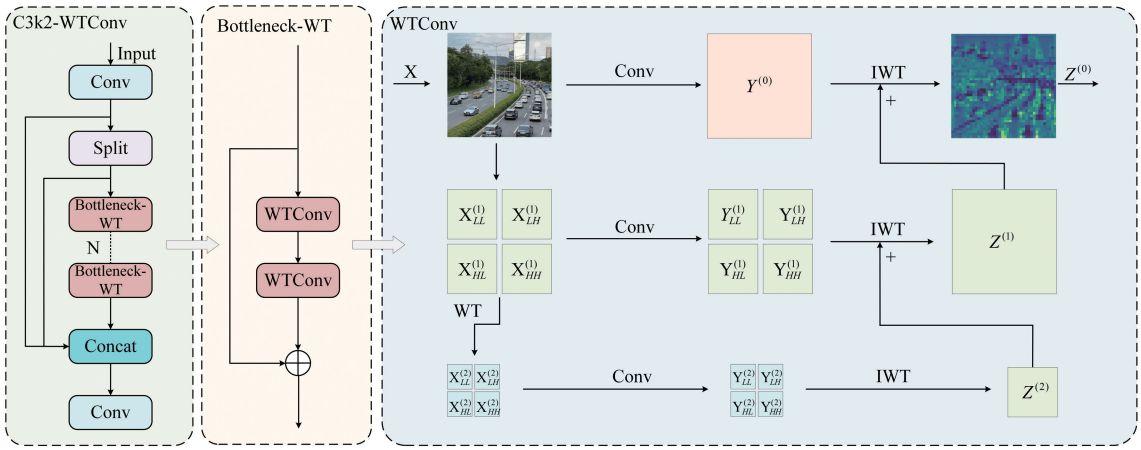


图 3 WTConv 模块网络结构

Fig. 3 Network structure of WTConv module

式中：f_{LL} 是低通滤波器；f_{LH} 是水平高通滤波器；f_{HL} 是垂直高通滤波器；f_{HH} 是对角高通滤波器。低通滤波器用于捕获图像的全局的低频信息，高通滤波器用于保留图像的局部高频细节。

其次，对于每个输入通道，卷积的输出有 4 个通道，如式(2)所示，分别为 $\boldsymbol{X}_{LL}$、$\boldsymbol{X}_{LH}$、$\boldsymbol{X}_{HL}$ 和 $\boldsymbol{X}_{HH}$，$\boldsymbol{X}_{LL}$ 是 $\boldsymbol{X}$ 的低频分量，$\boldsymbol{X}_{LH}$、$\boldsymbol{X}_{HL}$ 和 $\boldsymbol{X}_{HH}$ 分别是 $\boldsymbol{X}$ 的水平、垂直和对角高频分量，且每个通道的分辨率在每个空间维度上是输入图像 $\boldsymbol{X}$ 的一半。由于式(2)中的核构成了一个正交基，因此，逆小波变换 *IWT* 的应用可通过转置卷积 *ConvT* 获得，如式(3)所示。

$$[\boldsymbol{X}_{LL},\boldsymbol{X}_{LH},\boldsymbol{X}_{HL},\boldsymbol{X}_{HH}] = Conv([\boldsymbol{f}_{LL},\boldsymbol{f}_{LH},\boldsymbol{f}_{HL},\boldsymbol{f}_{HH}],\boldsymbol{X}) \tag{2}$$

$$\boldsymbol{X} = ConvT([\boldsymbol{f}_{LL},\boldsymbol{f}_{LH},\boldsymbol{f}_{HL},\boldsymbol{f}_{HH}],[\boldsymbol{X}_{LL},\boldsymbol{X}_{LH},\boldsymbol{X}_{HL},\boldsymbol{X}_{HH}]) \tag{3}$$

在此基础上，通过级联 *WT* 的应用，WTConv 模块可以生成多个尺度的频率分量，在小波分解的每一级，来自前一级的低频分量 $\boldsymbol{X}_{LL}^{(i-1)}$ 被进一步分解，产生一个新的低频分量 $\boldsymbol{X}_{LL}^{(i)}$，分层分解突出低频信息，增强了模型对低频特征的响应能力，如式(4)所示。

$$\boldsymbol{X}_{LL}^{(i)},\boldsymbol{X}_{LH}^{(i)},\boldsymbol{X}_{HL}^{(i)},\boldsymbol{X}_{HH}^{(i)} = WT(\boldsymbol{X}_{LL}^{(i-1)}) \tag{4}$$

然后，在每个频率分量级别，WTConv 模块执行一个小的核深度卷积，如式(5)所示。

$$\boldsymbol{Y}_{LL}^{(i)},\boldsymbol{Y}_{H}^{(i)} = Conv(\boldsymbol{W}^{(i)},(\boldsymbol{X}_{LL}^{(i)},\boldsymbol{X}_{H}^{(i)})) \tag{5}$$

式中：$\boldsymbol{W}^{(i)}$ 表示 i 级深度卷积核的权重；$\boldsymbol{Y}_{LL}^{(i)}$ 和 $\boldsymbol{Y}_{H}^{(i)}$ 分别表示卷积运算后的低频和高频输出。由于 WT 降低了每个子带的空间分辨率，小的卷积核可以覆盖图像的更大区域，以在不显著增加参数数量的情况下扩大感受野。

最后，使用 IWT 将每个频率分量的卷积结果重建到原始空间域中。利用 IWT 的线性特性，有效地整合了来自多级卷积的信息，以在实现具有大感受野的卷积操作的同时保留了多频率特征，如式(6)所示。

$$\boldsymbol{Z}^{(i)} = IWT(\boldsymbol{Y}_{LL}^{(i)} + \boldsymbol{Z}^{(i+1)},\boldsymbol{Y}_{H}^{(i)}) \tag{6}$$

式中：$\boldsymbol{Z}^{(i)}$ 表示 i 及以上级别的汇总输出。

1.2 特征交互增强模块设计

YOLOv11 采用的快速空间金字塔池化模块 SPPF，虽然其能够通过多尺度池化操作增强模型的感受野，但其固有的局部性计算方式导致难以建立长距离依赖关系，且固定尺寸的池化核限制了动态适应不同目标尺度的能力。为此，本文设计了 AIFI-LA 模块，如图 4 所示，通过剔除原有 SPPF 模块，引入内部特征尺度交互模块 AIFI[19]，并参考 Efficient ViT[20] 多尺度线性注意力设计简化线性注意力 LA 融入至 AIFI 前向通道。AIFI-LA 设计通过 AIFI 相同尺度的特征交互减少 SPPF 多尺度交互带来的计算冗余并捕获更精确的信息，通过 LA 降低计算复杂度、高效处理长序列以及保留关键特征信息，以提升模型在复杂场景中的性能。AIFI-LA 模块的工作流程可分为 4 个关键阶段。

1)对输入特征图进行序列化，由于深层 ***S5*** 特征相对浅层的 ***S3*** 和 ***S4*** 特征，具有更高级的语义特征，所以模型对输入特征图 ***S5*** 进行操作，***S5*** 的形状为[B,C,H,W]，其中 B 是批量大小，C 是通道数，H 和 W 分别是特征图的高度和宽度，为了将二维的特征图转换为适合注意力机制处理的序列形式，本文首先将特征图展平为[B,$H\times W$,C]，这种展平操作使得每个像素位置的特征向量被重新组织为序列中的一个元素，便于后续的注意力计算。

2)对展平后的特征序列进行位置编码的构建，为了引入位置信息，本文使用二维正弦和余弦位置编码，根据特征图的宽度 W、高度 H 和嵌入维度 Dim，生成位置编码，位置编码的计算基于正弦和余弦函数，能够为每个像素位置赋予独特的编码，从而让模型能够感知到像素的空间位置信息，生成的位置编码形状为[1,

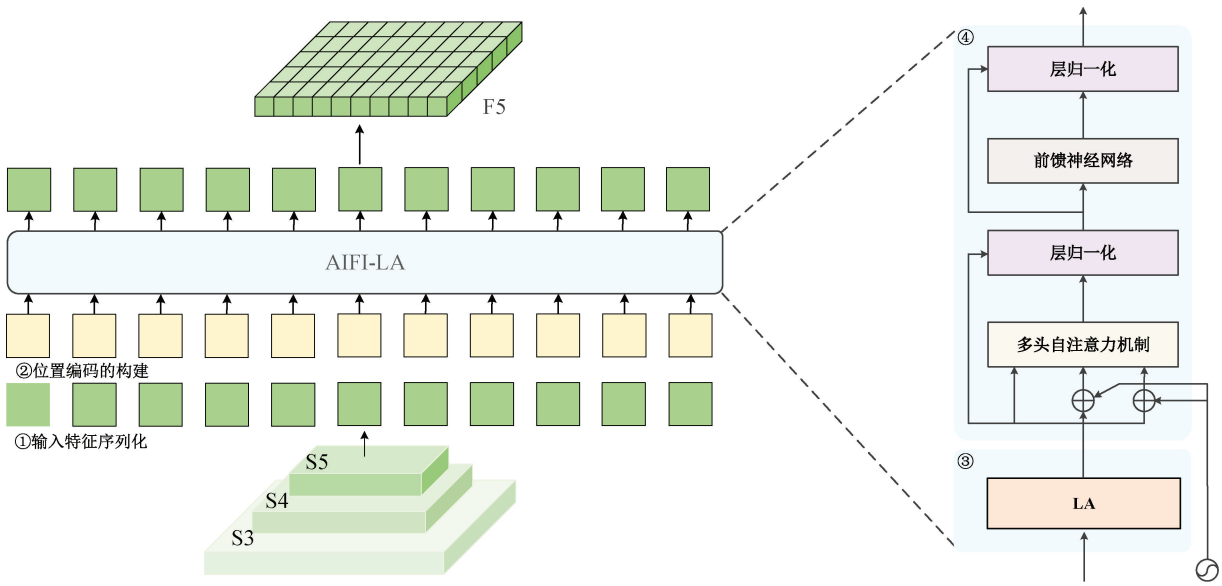


图 4　AIFI-LA 模块网络结构

Fig. 4　AIFI-LA module network architecture diagram

$H \times W$, Dim], 其与展平后的特征序列形状一致,可以直接相加。

3)对位置增强的特征序列进行线性注意力的计算,LA 线性注意力是 AIFI-LA 模块的核心部分,LA 结构图如图 5 所示。LA 首先通过 $\boldsymbol{Q}$、$\boldsymbol{K}$ 和 $\boldsymbol{V}$ 3 个线性层将输入特征映射为查询 $\boldsymbol{Q}$ 、键 $\boldsymbol{K}$ 和值 $\boldsymbol{V}$ 3 个部分,使模型能够从不同角度提取输入特征表示,如式(7)所示,其中 $\boldsymbol{W}_Q$、$\boldsymbol{W}_K$ 和 $\boldsymbol{W}_V$ 分别是查询 $\boldsymbol{Q}$ 、键 $\boldsymbol{K}$ 和值 $\boldsymbol{V}$ 的线性变换矩阵,$\boldsymbol{X}$ 是输入特征。其次,将这 3 个部分重新组织为适合多头注意力计算的形状。随后,对查询 $\boldsymbol{Q}$ 和键 $\boldsymbol{K}$ 分别进行 Softmax 归一化,使注意力权重的计算更加稳定和高效,同时减少计算复杂度,如式(8)所示。然后,通过矩阵乘法计算 $\boldsymbol{K}'$ 的转置与值 $\boldsymbol{V}$ 的积,得到全局上下文向量 ,使得模型能够更好地捕捉到输入特征中的长距离依赖关系,如式(9)所示,并将 $\boldsymbol{Q}'$ 与全局上下文向量相乘,得到最终的注意力输出,如式(10)所示,以更有效地关注到重要的特征部分。最后,将注意力输出重新组合为原始的特征形状 $[B,C,H,W]$,并通过一个投影层进行线性变换,得到最终的特征表示。

$$\begin{cases} \boldsymbol{Q} = \boldsymbol{W}_Q \cdot \boldsymbol{X}, \\ \boldsymbol{K} = \boldsymbol{W}_K \cdot \boldsymbol{X}, \\ \boldsymbol{V} = \boldsymbol{W}_V \cdot \boldsymbol{X}, \end{cases} \tag{7}$$

$$\begin{cases} \boldsymbol{Q}' = Softmax(\boldsymbol{Q}, Dim = -1) \\ \boldsymbol{K}' = Softmax(\boldsymbol{K}, Dim = -2) \end{cases} \tag{8}$$

$$Context = (\boldsymbol{K}')^{\mathrm{T}} \cdot \boldsymbol{V} \tag{9}$$

$$Output = \boldsymbol{Q}' \cdot Context \tag{10}$$

4)对线性注意力输出进行特征融合与输出,将经过线性注意力处理后的特征与原始位置编码相加,然后传递给 Transformer 编码器层,以增强特征的表达能力。首先,通过多头自注意力操作,让每个位置的特征与其他位置的特征进行交互,以增强特征的全局表达能力;然后,多头自注意力操作后的特征会通过一个前馈网络进一步增强特征的非线性表达能力;最后,进行归一化操作,将处理后的特征重新组织为原始的四维特征图形状 $[B,C,H,W]$, 以稳定训练过程并提高模型的性能。

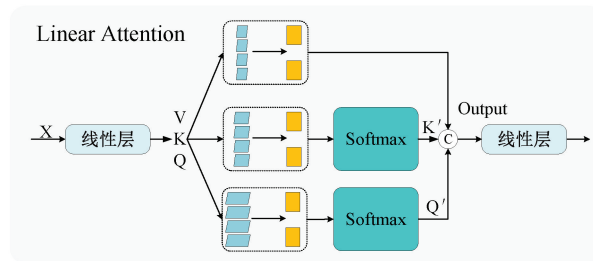


图 5　线性注意力结构图

Fig. 5　Linear attention structure diagram

1.3　特征重校准上采样设计

在 YOLOv11n 的网络架构中,上采样发生在颈部网络,上采样操作通过最近邻插值技术,将来自高层主干网络的高分辨率特征图与低层主干网络的低分辨率特征图进行融合,以恢复因下采样而丢失的空间分辨率,同时保留高层主干网络的语义信息。然而,在颈部网络中,其将特征图的尺寸放大,以便与主干网络中的特征图进行拼接,但是该部分仅通过像素点的空间分布来决定上采样,

未采用特征图的语义信息,会导致目标图像中空间信息的丢失。由于复杂交通场景中,远景处车辆与行人等构成的小目标在图像中分辨率占比较低且特征不全,这一问题在复杂道路场景分割中更为突出。

为了改善 YOLOv11n 上采样过程容易丢失小目标信息问题,改进模型采用 CARAFE 算子[21]替换颈部上采样结构,并将设计的特征重校准 EMCSA 模块融入其中,构成 CARAFE-EMCSA,其结构如图 6 所示。CARAFE-EMCSA 通过动态上采样,使其在较大的感受野内聚合上下文信息,以抑制小目标的信息丢失。同时,CARAFE-EMCSA 不是为所有的样本使用一个固定的内核,而是支持特定于实例的内容感知处理,其可以动态地生成自适应的内核,充分利用了特征图的语义信息,从而捕获和保留更多的环境特征信息。

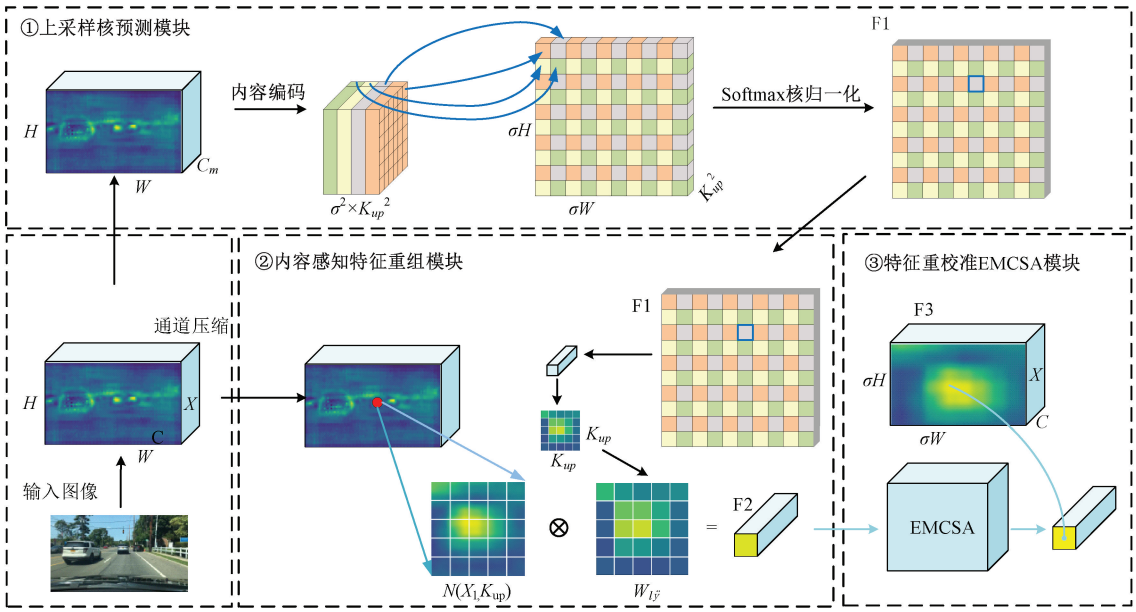


图 6　CARAFE-EMCSA 上采样算子结构

Fig. 6　Structure diagram of CARAFE-EMCSA upsampling operator

CARAFE 算子主要由上采样核预测模块和特征重组模块组成。一方面,上采样核预测模块的采样过程中,首先通过输入 $H \times W \times C$ 的特征图,特征图通过一个 1×1 卷积将其通道数压缩到 C_m,以有效减少模型参数量和计算成本,并通过卷积操作将通道数从 $H \times W \times C$ 变为 $\sigma^2 \times k_{up^2}$,完成内容编码;其次,将通道在空间维度进行展开,得到形状为 $\sigma H \times \sigma W \times k_{up^2}$ 的上采样核;最后,对上采样进行 Softmax 归一化操作,使其卷积核权重为 1,输出特征图 F1。另一方面,在特征重组模块中,其将归一化上采样核输出的特征图 F1 中每个位置映射回输入特征图中,取以之为中心的 $k_{up} \times k_{up}$ 原特征图区域,和预测点上采样核进行点积,得到输出值,且相同位置的不同通道共享同一个上采样核,得到形状为 $\sigma H \times \sigma W \times C$ 特征图 F2,有助于提升模型在小目标分割中的表现。

然而,由于 CARAFE 算子的设计主要侧重于局部区域内的内容感知特征重装配,经过 CARAFE 算子处理得到的特征图 F2 在全局上下文信息的捕捉和高频细节特征的保留方面存在不足,细节信息丢失较多。为此,提出一种特征重校准 EMCSA 模块,重新校准特征图 F2。首先,通过在通道注意力机制(efficient channel attention,

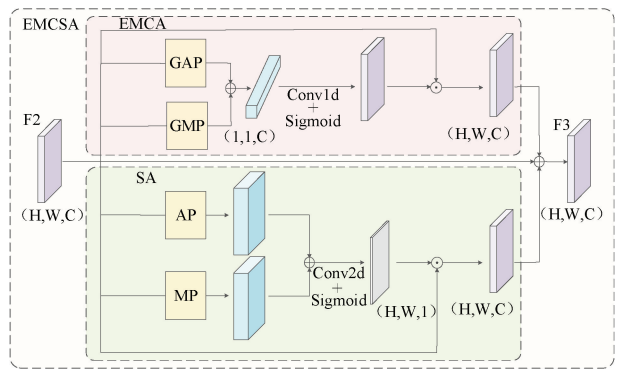


图 7　特征重校准模块

Fig. 7　Feature recalibration module

ECA)[22]内并入全局最大池化(GMP)构建 EMCA 注意力机制,然后将 EMCA 和 SA (spatial attention)空间注意力机制[23]并行设计,对特征图 F2 进行优化处理,从而增强特征图的全局语义表达能力,细化高频细节特征,提升特征图的整体判别能力,其结构图如图 7 所示,设计方案由以下两部分组成。

1) 采用 EMCA 对特征图 F2 进行通道重校准。首

先,采用全局平均池化(global average pooling,GAP)对特征图进行尺寸压缩,提取特征图的全局特征,保留通道的语义信息;同时,与 *GMP* 后的特征进行相加,以弥补全局平均池化可能导致的高频细节信息的丢失,保留重要突出特征;然后,通过卷积操作对全局特征向量进行建模,促进通道之间的信息交互,并采用 sigmoid 激活函数进行归一化处理,动态调整不同区域的重要性,生成通道注意力权重;最后,将加权后特征与初始特征进行逐元素相乘,对初始特征图进行重新筛选,过滤初始特征 F2 的冗余信息,得到最终的输出特征 Fa。这一过程如式(11)所示。

$$Fa = \sigma(Conv1d \cdot (GMP(F2) + GAP(F2))) \odot F2 \tag{11}$$

式中:σ 表示 *sigmoid* 激活函数;$\odot$ 表示逐元素相乘。

2)对特征图 F2 进行 SA 空间重校准。首先,采用平均池化(AP)对特征图进行空间尺寸压缩,提取特征图局部特征,保留空间语义信息;同时,对特征图进行最大池化(MP)保留局部显著特征;其次,将池化结果进行通道拼接,并通过卷积层处理拼接后的特征图,以动态融合平均池化和最大池化特征;然后,对拼接后特征图运用 sigmoid 激活函数生成空间注意力权重。将加权后特征与初始特征进行逐元素相乘,对初始特征图进行重新筛选,得到最终的输出特征 Fb;最后,将通道重校准后的特征图 Fa 和空间重校准后的特征图 Fb 相加,得到最终的输出特征图 F3。相比于 F2,F3 不仅在全局语义表达上更为丰富,能够更好地理解图像的整体内容,而且在局部细节特征上更为精细,能够更好地捕捉到图像中的高频信息和保留更多的环境特征。这一过程如式(12)和(13)所示。

$$Fb = \sigma(Conv2d(MP(F2) \oplus AP(F2))) \odot F2 \tag{12}$$

$$F3 = Fa + Fb \tag{13}$$

式中:$\oplus$ 表示通道维度上的拼接。

1.4　复合非极大值抑制算法

NMS 是图像处理任务中常用的一种后处理技术,用于筛选和保留最佳的边界框,以去除多余的边界框并减少误检,在使用基于区域的卷积神经网络时,NMS 是一个关键的步骤,该步骤会生成大量的边界框区域,其中许多区域表现出高度重叠,NMS 用于选择具有高可信度和非重叠区域的分割结果。然而,传统的 NMS 算法需要手动设置阈值来决定是保留还是丢弃边界框,NMS 算法可用式(14)表示。

$$s_i = \begin{cases} s_i, IoU(M,b_i) < N_t \\ 0, IoU(M,b_i)^3 \geqslant N_t \end{cases} \tag{14}$$

式中:s_i 表示目标分割中每个边界框得分;M 表示排序得分最高的边界框;b_i 表示去除 M 后剩余的边界框;N_t 是 NMS 的阈值;$IoU(M,b_i)$ 表示 M 和 b_i 之间的交集。

传统的 NMS 直接移除边界框 *IoU* 超过阈值的所有其他边界框,在复杂交通场景中,行人与车辆密集且被遮挡目标众多,NMS 算法会导致漏检问题存在。改进模型通过先引入 Soft-NMS 算法[24]以减少漏检问题,Soft-NMS 和传统 NMS 之间的关键差异在于调整边界框分数的方法,Soft-NMS 引入高斯权值,以高斯加权方式调整边界框的分数,而不是直接移除边界框 IoU 超过阈值的其他边界框,Soft-NMS 算法可用式(15)表示。

$$s_i = \begin{cases} s_i, & IoU(M,b_i) < N_t \\ s_i e^{-\frac{IoU(M,b_i)^2}{\sigma}}, & IoU(M,b_i) \geqslant N_t \end{cases} \tag{15}$$

式中:$e^{-(IoU(M,b_i)^2/\sigma)}$ 为高斯惩罚函数;σ 为超参数,用于控制边界框分数衰减的速率。当一个边界框的 IoU 低于设置的阈值时,该边界框将被保留,当一个边界框的 IoU 大于或等于设置的阈值时,Soft-NMS 将以高斯加权的方式降低边界框的分数,而不是将其设置为 0,以保留更多的边界框,降低在复杂交通场景中交通密集处车辆与行人的漏检率。

同时,为了考虑边界框之间的距离和面积信息,从而更好地处理重叠框的情况,将 Soft-NMS 与 DIoU-NMS[25]相结合,DIoU 通过引入中心点距离的惩罚项,能够更好地区分相邻目标,DIoU-NMS 算法可用式(16)表示。

$$s_i = \begin{cases} s_i, IoU - R_{DIoU}(M,b_i) < N_{td} \\ 0, IoU - R_{DIoU}(M,b_i) \geqslant N_{td} \end{cases} \tag{16}$$

式中:$DIoU = IoU - R_{DIoU}(M,b_i)$,$R_{\mathrm{DIoU}}(M,b_i)$ 表示 DIoU 损失的惩罚项。当一个边界框的 DIoU 低于设置的阈值时,该边界框将被保留,当一个边界框的 DIoU 大于或等于设置的阈值时,则认为边界框与参考框重叠较大,且与中心点距离较近,则抑制此边界框。

综上,将 Soft-NMS 与 DIoU-NMS 相结合,构成 Soft-DIoU-NMS,如式(17)所示。

$$s_i = \begin{cases} s_i, & DIoU(M,b_i) < N_t \\ s_i e^{-\frac{DIoU(M,b_i)^2}{\sigma}}, & DIoU(M,b_i) \geqslant N_t \end{cases} \tag{17}$$

式中:当一个边界框的 DIoU 低于设置的阈值时,则认为该边界框与参考框的重叠较小,中心点距离较远,该边界框的得分将被保留;当一个边界框的 *DIoU* 大于或等于设置的阈值时,边界框的得分将通过 Soft-NMS 的高斯权值来降低分数,以避免对小框或大框的过度抑制。

2　实验与分析

2.1　实验数据集

Cityscapes 实例分割数据集[26],该数据集由奔驰公司推出,是计算机视觉领域和道路交通场景分割任务中

广泛使用的权威基准数据集之一,该数据集总共8大类,含2 975张精细标注的训练集、500张精细标注的验证集和500张未标注的测试集。实验所用为有标签的训练集和验证集,并将标签转换成YOLO格式,同时由于该数据集目标复杂,实验将标签合并为人和车两大类,最后,重新将数据集随机划分为训练集、验证集和测试集3部分,随机划分比例为6∶2∶2。

BDD100K实例分割数据集[27],该数据集由伯克利大学AI实验室发布,是目前最大规模且内容最为多样的公开交通场景数据集之一,专为交通场景领域的研究和开发而设计,该数集据总共10大类,含7 000张带标签的训练集、1 000张带标签的验证集和2 000张无标签的测试集。实验所用为有标签的训练集和验证集,并将标签转换成YOLO格式,同时由于该数据集目标复杂,实验将标签合并为人和车两大类。

2.2　实验配置与评价标准

实验是在Windows10专业版上完成的,采用的GPU为NVDIA GeForce RTX 2060 SUPER,显存为8 G,CPU为AMD Ryzen 5 3600X 6-Core Processor。深度学习框架为Pytorch2.1.0,采用Python3.9.19编程,同时使用的CUDA版本为12.1。

实验参数设置如下:初始学习率为0.01,衰减系数设置为0.000 5,epochs设置为200,批处理大小为16,优化器为随机梯度下降(stochastic gradient descent,SGD)。

实验评价标准选用精确度(P)、召回率(R)、平均精确率(mean average precision,mAP)、衡量模型大小的参数量。P指的是在全部预测结果为正样本中实际是正样本目标的比例,用于衡量算法识别能力;R表示在测试集所有正样本中,被分割出来为正样本的比例,具体计算如式(18)、(19)所示。

$$P = \frac{TP}{TP + FP} \tag{18}$$

$$R = \frac{TP}{TP + FN} \tag{19}$$

式中:TP表示正确的正样本,指正确预测为正样本的数量;FP表示错误的正样本,指错误的将负样本预测为正样本的数量;FN表示错误的负样本,指错误的将正样本预测为负样本的数量。实验所用mAP包含边界框和分割掩膜,其指标有两种,分别为mAP@0.5和mAP@0.5:95,mAP@0.5表示IoU阈值为0.5时的平均分割精度,mAP@0.5:95表示IoU阈值为0.5和0.95间的平均分割精度,具体计算如式(20)~(21)下:

$$AP = \int_0^1 P(R)\,\mathrm{d}(R) \tag{20}$$

$$mAP = \frac{1}{C}\sum_{i=1}^{C} AP_i \tag{21}$$

式中:C表示复杂道路场景检测目标类别数量。

2.3　消融实验

首先,为了评估设计的AIFI-LA模块和提出的EMCSA特征重校准模块对交通场景分割的有效性,在Cityscapes分割数据集上对不同设计方法进行模块消融实验,实验结果如表1所示。

表1　模块有效性消融实验

Table 1　Module effectiveness ablation experiment

模型	边界框/%				分割掩膜/%				参数量/(×10⁶)
	P	R	mAP@0.5	mAP@0.5:0.95	P	R	mAP@0.5	mAP@0.5:0.95	
AIFI	67.3	39.3	43.5	26.1	64.0	36.5	40.7	21.3	**3.5**
+ LA	**67.9**	**40.0**	**44.1**	**26.5**	**64.8**	**36.8**	**41.3**	**21.6**	3.7
CARAFE	67.5	39.5	43.5	26.0	64.0	36.0	40.8	21.3	3.0
+ ECA	67.5	38.8	43.8	26.0	64.1	35.8	41.2	21.2	3.0
+ EMCA	67.8	**39.9**	43.8	26.3	64.5	**36.3**	41.3	21.5	3.0
+ SA	67.0	39.0	44.0	26.4	63.5	35.5	41.2	21.7	3.0
+ EMCSA	**67.8**	39.4	**44.2**	**26.5**	**64.8**	36.2	**41.6**	**21.8**	**3.0**

由表1可知,对设计的模块进行前后对比,可以看出AIFI模块结合线性注意力LA后,在参数量不显著提升的情况下,边界框平均精度mAP@0.5和mAP@0.5:0.95分别提升了0.6%和0.4%,分割掩膜平均精度mAP@0.5和mAP@0.5:0.95分别提升了0.6%和0.3%;CARAFE上采样算子在结合改进高效通道注意力EMCA和空间注意力SA组成的CARAFE-EMCSA后,相较于CARAFE单独结合ECA、EMCA和SA,精度提升更多,边界框平均精度mAP@0.5和mAP@0.5:0.95分别提升了0.7%和0.5%,分割掩膜平均精度mAP@0.5和mAP@0.5:0.95分别提升了0.8%和0.5%,验证了设计模块的有效性。

其次,为了评估不同改进方法对交通场景分割的有效性,在Cityscapes数据集上对不同改进方法进行消融实验,实验结果如表2所示。

表 2　不同模型结构的消融实验

Table 2　Ablation experiments with different model structures

模型	边界框/%				分割掩膜/%				参数量/($\times10^6$)
	P	*R*	mAP@0.5	mAP@0.5:0.95	*P*	*R*	mAP@0.5	mAP@0.5:0.95	
YOLOv11n-Seg	68.2	38.8	43.0	25.7	65.5	36.0	40.2	20.9	**2.8**
+C3k2-WTConv	68.7	39.5	44.3	27.0	65.8	35.7	41.4	22.3	2.8
+AIFI-LA	67.9	**40.0**	44.1	26.5	64.8	**36.8**	41.3	21.6	3.7
+CARAFE-EMCSA	67.8	39.4	44.2	26.5	64.8	36.2	41.6	21.8	3.0
+Soft-DIoU-NMS	68.5	38.1	49.6	31.3	66.4	35.9	47.6	26.2	2.8
+C3k-WTConv、AIFI-LA	68.5	38.8	45.0	27.4	65.9	35.8	42.6	22.8	3.7
CARAFE-EMCSA	68.3	38.9	46.0	28.5	65.7	35.8	43.7	23.8	3.8
Soft-DIoU-NMS	**69.9**	39.6	**52.2**	**34.2**	**66.9**	36.6	**50.8**	**29.7**	3.8

由表 2 可知，在 C3k2-BottleNeck 中嵌入小波变换卷积 WTConv 后，边界框平均精度 mAP@0.5 和 mAP@0.5:0.95 分别提升了 1.3%和 1.3%，分割掩膜平均精度 mAP@0.5 和 mAP@0.5:0.95 分别提升了 1.2%和 1.4%；设计 AIFI-LA 模块替换 SPPF 模块后，边界框平均精度 mAP@0.5 和 mAP@0.5:0.95 分别提升了 1.1%和 0.8%，分割掩膜平均精度 mAP@0.5 和 mAP@0.5:0.95 分别提升了 1.1%和 0.7%；设计增强型上采样 CARAFE-EMCSA 替换颈部网络原上采样后，边界框平均精度 mAP@0.5 和 mAP@0.5:0.95 分别提升了 1.2%和 0.8%，分割掩膜平均精度 mAP@0.5 和 mAP@0.5:0.95 分别提升了 1.4%和 0.9%；通过将 Soft-NMS 与 DIoU-NMS 相结合并替换原 NMS，边界框平均精度 mAP@0.5 和 mAP@0.5:0.95 分别提升了 6.6%和 5.6%，分割掩膜平均精度 mAP@0.5 和 mAP@0.5:0.95 分别提升了 7.4%和 5.3%；将 C3k-WTConv 与 AIFI-LA 结合后，边界框平均精度 mAP@0.5 和 mAP@0.5:0.95 分别提升了 2.0%和 1.7%，分割掩膜平均精度 mAP@0.5 和 mAP@0.5:0.95 分别提升了 2.4%和 1.9%；在此基础上，融入 CARAFE-EMCSA 后，边界框平均精度 mAP@0.5 和 mAP@0.5:0.95 分别提升了 3.0%和 2.8%，分割掩膜平均精度 mAP@0.5 和 mAP@0.5:0.95 分别提升了 3.5%和 2.9%；最后，结合 Soft-DIoU-NMS，在不明显提升参数量的情况下，边界框平均精度 mAP@0.5 和 mAP@0.5:0.95 分别提升了 9.2%和 8.5%，分割掩膜平均精度 mAP@0.5 和 mAP@0.5:0.95 分别提升了 10.6%和 8.8%，证明了改进算法的有效性。

改进前后的分割掩膜 mAP@0.5 训练曲线如图 8 所示，可以看出，两模型训练周期均在 150 次左右收敛，且相较于 YOLOv11n-Seg 基准模型，ETIS-YOLO 训练精度得到有效的提升，训练精度更高。

从模型结构优化，可以看出，WTConv 的引入通过小波变换增强了模型对提高感受野扩展效率和对低频特征捕捉能力；AIFI-LA 模块设计减少冗余计算并提高处理

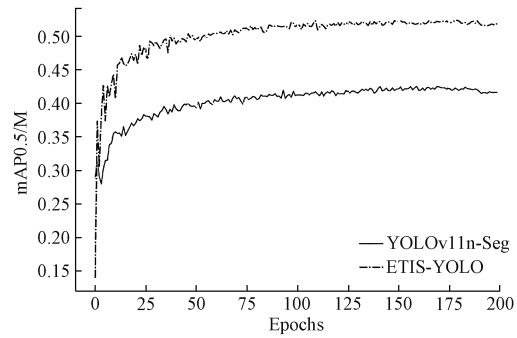
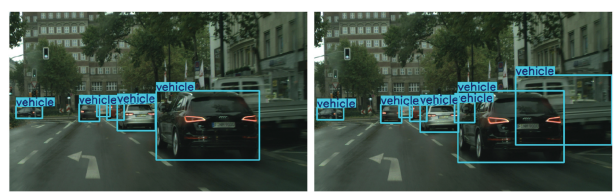


图 8　改前后训练过程对比

Fig. 8　Comparison of training processes before and after modification

长序列以及保留关键特征信息，而 CARAFE-EMCSA 上采样则通过内容感知的特征重组，显著改善了小目标的特征重建质量，并对特征图进行通道和空间重校准。从模型训练策略，可以看出，各项改进模块的组合使用产生了协同效应，当 WTConv、AIFI-LA、CARAFE-EMCSA 和复合 NMS 结合时，性能提升幅度超过了单一模块改进的简单叠加，证明了这些组件在功能上具有互补性。图 9 展示了 NMS 改进前后的推理优化对比图，可以看出，Soft-DIoU-NMS 设计优化了推理阶段边界框质量，其通过综合考虑目标间的空间关系和重叠程度，降低了传统 NMS 方法在处理密集目标时容易出现误抑制现象，从而提高边界框平均精度。



(a) NMS算法　(b) Soft-DIoU-NMS算法

(a) NMS algorithm　(b) Soft-DIoU-NMS algorithm

图 9　NMS 效果对比

Fig. 9　Comparison of NMS effects

2.4　对比实验

为了评估改进模型对交通场景分割的有效性，分别与 YOLACT、YOLOv5n-Seg、YOLOv8n-Seg、YOLOv11n-Seg 和 DE-YOLO 在 Cityscapes 实例分割数据集进行对比实验，对比实验结果如表 3 所示。

表 3　不同模型在 Cityscapes 上的对比实验

Table 3　Comparative experiments of different models on Cityscapes

模型	边界框/%				分割掩膜/%				参数量/($\times10^6$)
	P	*R*	mAP@0.5	mAP@0.5:0.95	*P*	*R*	mAP@0.5	mAP@0.5:0.95	
YOLACT	67.8	38.1	43.6	24.7	66.0	34.5	39.6	22.5	34.8
YOLOv5n-Seg	67.4	37.9	41.4	22.9	65.3	34.0	37.9	19.1	**2.0**
YOLOv8n-Seg	69.3	39.3	43.1	25.7	66.8	36.3	40.4	21.0	3.4
YOLOv11n-Seg	68.2	38.8	43.0	25.7	65.5	36.0	40.2	20.9	2.8
DE-YOLO	68.0	39.0	46.2	26.9	65.0	35.9	42.2	22.1	3.5
ETIS-YOLO	**69.9**	**39.6**	**52.2**	**34.2**	**66.9**	**36.6**	**50.8**	**29.7**	3.8

由表 3 可知，在 Cityscapes 实例分割数据集上，改进模型边界框平均精度 mAP@0.5 和 mAP@0.5:0.95 分别达到了 52.2% 和 34.2%，分割掩膜平均精度 mAP@0.5 和 mAP@0.5:0.95 分别达到了 50.8% 和 29.7%，与当前主流轻量化分割算法对比，改进模型取得了更高平均边界框精度和平均掩码精度，验证了改进模型在交通场景分割的有效性。

表 4　不同模型在 BDD100K 上的对比实验

Table 4　Comparative experiments of different models on BDD100K

模型	边界框/%				分割掩膜/%				参数量/($\times10^6$)
	P	*R*	mAP@0.5	mAP@0.5:0.95	*P*	*R*	mAP@0.5	mAP@0.5:0.95	
YOLACT	72.5	50.0	55.6	32.5	72.1	48.0	53.9	29.5	34.8
YOLOv5n-Seg	71.5	49.5	54.3	31.4	71.6	47.2	52.7	28.2	**2.0**
YOLOv8n-Seg	74.3	51.7	58.2	36.4	73.6	49.2	55.7	31.4	3.4
YOLOv11n-Seg	72.0	50.8	56.7	36.3	71.5	48.8	54.9	31.2	2.8
DE-YOLO	72.1	50.9	60.0	37.2	71.2	48.9	57.2	33.4	3.5
ETIS-YOLO	**74.6**	50.6	**61.8**	**43.7**	**74.1**	**49.7**	**59.4**	**37.8**	3.8

(a) 原图
(a) Original images

(b) YOLOv11n-Seg模型
(b) YOLOv11n-Seg model

(c) ETIS-YOLO模型
(c) ETIS-YOLO model

图 10　交通场景可视化对比

Fig. 10　Visual comparison of traffic scenes

2.5　泛化性实验

为了评估改进模型对交通场景分割的泛化性,分别与 YOLACT 、YOLOv5n-Seg、YOLOv8n-Seg、YOLOv11n-Seg 和 DE-YOLO 在 BDD100K 实例分割数据集进行对比实验,对比实验结果如表 4 所示。

由表 4 可知,在 BDD100K 实例分割数据集上,改进模型 ETIS-YOLO 边界框平均精度 mAP@0.5 和 mAP@0.5:0.95 分别达到了 61.8%和 43.7%,分割掩膜平均精度 mAP@0.5 和 mAP@0.5:0.95 分别达到了 59.4%和 37.8%,与当前主流轻量化分割算法对比,改进模型取得了更高的平均边界框精度和平均掩码精度,验证了改进模型在交通场景下分割具备一定的泛化性。

2.6　分割可视化

在上述实验的基础上,为了验证改进模型应用于实际的有效性,分别将改进前后的模型在 Cityscapes 数据集上进行可视化实验,可视化实验结果如图 10 所示。可以看出,改进后的模型在 Cityscapes 数据集上平均边界框精度更高,分割结果更精细,掩膜分割质量更好。

3　结　论

为了改善交通场景下目标分割精度低和掩膜质量差的问题,提出了一种改进 YOLOv11n 的高效交通实例分割算法——ETIS-YOLO。首先,通过重构主干网络 C3k2 模块,以提高感受野扩展效率和对低频特征捕捉能力;其次,设计简化的线性注意力 LA 嵌入 AIFI 的前向传播中,构建 AIFI-LA 模块替换主干网络 SPPF 模块,以提高处理长序列以及保留关键特征信息;在此基础上,提出特征重校准 EMCSA 模块,并嵌入至上采样算子 CARAFE 中,构建上采样算子 CARAFE-EMCSA,分别对特征图进行通道和空间重校准,再次提升对特征图的整体判别能力;最后,重构 NMS 算法来改善原模型在交通场景分割中的不足。实验表明,改进模型在不显著提升参数量的情况下,平均边界框精度和平均分割掩膜精度得到了较高的提升,且在不同场景下具备一定的泛化性。

参考文献

[1] 伍锡如, 邱涛涛, 王耀南. 改进 Mask R-CNN 的交通场景多目标快速检测与分割[J]. 仪器仪表学报, 2021, 42(7): 242-249.
WU X R, QIU T T, WANG Y N. Multi-object detection and segmentation for traffic scene based on improved Mask R-CNN [J]. Chinese Journal of Scientific Instrument, 2021, 7(42): 242-249.

[2] 贺晓东, 王春艳, 孙昊, 等. 基于局部特征与视点感知的车辆重识别算法[J]. 仪器仪表学报, 2022, 43(10): 177-184.
HE X D, WANG CH Y, SUN H, et al. Local-features and viewpoint-aware for vehicle re-identification [J]. Chinese Journal of Scientific Instrument, 2022, 43(10): 177-184.

[3] 赵红爱, 王旭智, 万旺根. 一种用于车辆图像分割的 MSSA-UNet 模型[J]. 电子测量技术, 2022, 45(8): 102-107.
ZHAO H AI, WANG X ZH, WAN W G. A MSSA-UNet model for vehicle image segmentation [J]. Electronic Measurement Technology, 2022, 45(8): 102-107.

[4] OZTURK O, SARITÜRK B, SEKER D Z. Comparison of fully convolutional networks (FCN) and U-Net for road segmentation from high resolution imageries [J]. International Journal of Environment and Geoinformatics, 2020, 7(3): 272-279.

[5] ARULANANTH T S, KUPPUSAMY P G, AYYASAMY R K, et al. Semantic segmentation of urban environments: Leveraging U-Net deep learning model for cityscape image analysis[J]. Plos One, 2024, 19(4): 1-20.

[6] LI J, JIANG F L, YANG J, et al. Lane-DeepLab: Lane semantic segmentation in automatic driving scenarios for high-definition maps[J]. Neurocomputing, 2021, 465: 15-25.

[7] WAN CH X, CHANG X N, ZHANG Q H. Improvement of road instance segmentation algorithm based on the modified Mask R-CNN[J]. Electronics, 2023, 12(22): 4699.

[8] DONG W, LIU Z Y, YANG M. FIR-YOLACT: Fusion of ICIoU and Res2Net for YOLACT on Real-Time Vehicle Instance Segmentation [J]. Computers, Materials & Continua, 2023, 77(3): 3551-3572.

[9] XUE Y, ZHAN L L, LIU ZH S, et al. SAR ship target instance segmentation based on SISS-YOLO[J]. Remote Sensing, 2025, 17(17): 3118.

[10] ZHANG D, ZHANG L Y, TANG J H. Augmented FCN: Rethinking context modeling for semantic segmentation[J]. Science China Information Sciences, 2023, 66(4): 142105.

[11] ZHU F ZH, CUI J Y, ZHU B, et al. Semantic segmentation of urban street scene images based on improved U-Net network [J]. Optoelectronics Letters, 2023, 19(3): 179-185.

[12] LIU R R, HE D ZH. Semantic segmentation based on Deeplabv3+ and attention mechanism[C]. 2021 IEEE 4th Advanced Information Management, Communicates, Electronic and Automation Control Conference, 2021: 255-259.

[13] FANG S Q, ZHANG B, HU J Y. Improved mask R-CNN multi-target detection and segmentation for autonomous driving in complex scenes [J]. Sensors, 2023, 23(8): 3853.

[14] LI X X, DUAN C, ZHI Y, et al. Instance segmentation of traffic scene based on YOLACT[C]. IOP Conference Series: Earth and Environmental Science, 2021, 032011.

[15] XIA W J, LI P Q, LI Q P, et al. TTIS-YOLO: A traffic target instance segmentation paradigm for complex road scenarios[J]. Measurement Science and Technology, 2024, 35(10): 105402.

[16] 赵南南, 高翡晨. 基于改进 YOLOv8 的交通场景实例分割算法[J]. 计算机工程, 2025, 51(1): 198-207.
ZHAO N N, GAO F CH. Improved YOLOv8-based algorithm for instance segmentation in traffic scenes[J]. Computer Engineering, 2025, 51(1): 198-207.

[17] GU X L, ZHANG G F. PSC-YOLO: A lightweight model for urban road instance segmentation[J]. Journal of Real-Time Image Processing, 2025, 22(2): 1-13.

[18] FINDER S E, AMOYAL R, TREISTER E, et al. Wavelet convolutions for large receptive fields [C]. European Conference on Computer Vision, 2024: 363-380.

[19] ZHAO Y, LV W Y, XU S L, et al. Detrs beat YOLOs on real-time object detection[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 16965-16974.

[20] CAI H, LI J Y, HU M Y, et al. Efficientvit: Multi-scale linear attention for high-resolution dense prediction[C]. International Conference on Computer Vision 2023 (ICCV 2023), 2022.

[21] WANG J Q, CHEN K, XU R, et al. Carafe: Content-aware reassembly of features[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 3007-3016.

[22] WANG Q L, WU B G, ZHU P F, et al. ECA-Net: Efficient channel attention for deep convolutional neural networks [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 11534-11542.

[23] ZHANG Q L, YANG Y B. Sa-net: Shuffle attention for deep convolutional neural networks[C]. ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing, 2021: 2235-2239.

[24] BODLA N, SINGH B, CHELLAPPA R, et al. Soft-NMS-improving object detection with one line of code[C]. IEEE International Conference on Computer Vision, 2017: 5561-5569.

[25] ZHENG ZH H, WANG P, LIU W, et al. Distance-IoU loss: Faster and better learning for bounding box regression [C]. AAAI Conference on Artificial Intelligence, 2020: 12993-13000.

[26] CORDTS M, OMRAN M, RAMOS S, et al. The cityscapes dataset for semantic urban scene understanding[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2016: 3213-3223.

[27] YU F, CHEN H F, WANG X, et al. BDD100K: A diverse driving dataset for heterogeneous multitask learning [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 2636-2645.

作者简介

邵自强,2023 年于安徽工程大学获得学士学位,现为安徽工程大学硕士研究生,主要研究方向为深度学习与交通图像处理。
E-mail: 2230321117@stu.ahpu.edu.cn

Shao Ziqiang received his B. Sc. degree from Anhui Polytechnic University in 2023. Now he is a M. Sc. candidate at Anhui Polytechnic University. His main research interests include deep learning and traffic image processing.

魏利胜(通信作者),2001 年于安徽工程大学获得学士学位,2004 年于中国航天科工集团 061 基地获得硕士学位,2009 年于上海大学获得博士学位,现为安徽工程大学教授,主要研究方向为图像识别与应用、智能化网络控制系统和仿真。
E-mail: lshwei_11@163.com

Wei Lisheng (Corresponding author) received his B. Sc. degree from Anhui Polytechnic University in 2001, M. Sc. degree from China Aerospace Science and Industry Corporation 061 Base in 2004, and Ph. D. degree from Shanghai University in 2009. Now he is a professor and M. Sc. supervisor at Anhui Polytechnic University. His main research interests include image recognition and application, intelligent network control system and simulation.