Modeling and Simulation 建模与仿真, 2025, 14(3), 281-293
Published Online March 2025 in Hans. https://www.hanspub.org/journal/mos
https://doi.org/10.12677/mos.2025.143222

Hans汉斯

基于改进目标检测算法的道路场景图像实例分割方法

周　洲

上海理工大学光电信息与计算机工程学院，上海

收稿日期：2025年2月20日；录用日期：2025年3月13日；发布日期：2025年3月20日

摘　要

实例分割是图像分割的重要组成部分，同时也是计算机视觉中的一个关键研究课题，广泛应用于自动驾驶和安全监控等领域。然而，由于道路场景通常具有复杂性、多样性和杂乱的特点，处理这些场景变得尤为挑战。针对道路场景图像实例分割难度大、精度低、定位不精确等问题，本文提出一种基于改进YOLOv5 (You Only Look Once version 5)的道路场景实例分割算法。以YOLOv5为基础模型，在Head模块中采用RFAConv (Receptive-Field Attention Convolution)卷积代替部分传统卷积，它全面解决了卷积核的参数共享问题，考虑到接受域中每个特征的重要性，提供了几乎可以忽略不计的计算成本和参数增量，能够更好地捕捉和融合图像特征，提升分割的精度和鲁棒性。采用ShapeIOU代替YOLOv5中原损失函数CIOU (Complete-IoU)，通过聚焦边框自身形状与自身尺度计算损失，使得边框回归更为精确，能够有效提升检测效果且优于现有方法。实验结果表明：与原模型相比，改进后的模型的分割精度mAP50 (mean Average Precision)达到了33.8%，相较于YOLOv5s，优化后的模型在分割精度上提高了1.2%，能够更加高效地完成道路场景的图像分割任务。

关键词

实例分割，YOLOv5，接受场注意力卷积，注意力机制，损失函数，特征融合

An Instance Segmentation Method for Road Scene Images Based on Improved YOLOv5

Zhou Zhou

College of Optoelectronic Information and Computer Engineering, University of Shanghai for Science and Technology, Shanghai

Received: Feb. 20th, 2025; accepted: Mar. 13th, 2025; published: Mar. 20th, 2025

文章引用: 周洲. 基于改进目标检测算法的道路场景图像实例分割方法[J]. 建模与仿真, 2025, 14(3): 281-293.
DOI: 10.12677/mos.2025.143222

Abstract

Instance segmentation is a crucial component of image segmentation and serves as a significant research area within computer vision. It finds extensive applications in various domains, including autonomous driving and security surveillance. However, due to the complexity, diversity, and cluttered nature of road scenes, handling these scenarios becomes particularly challenging. In response to the challenges of high complexity, low accuracy, and imprecise positioning in road scene image instance segmentation, this paper introduces an enhanced YOLOv5-based (You Only Look Once version 5) algorithm specifically designed for road scene instance segmentation. Taking YOLOv5 as the base model, RFAConv (Receptive-Field Attention Convolution) convolution is used in the Head module instead of part of the traditional convolution, which comprehensively solves the parameter sharing problem of convolution kernel, considers the importance of each feature in the receptive domain, provides almost negligible computational cost and parameter increment, and is able to better capture and fuse the image features to improve the segmentation accuracy and robustness. ShapeIOU is used instead of the original loss function CIOU (Complete-IoU) in YOLOv5, and the loss is calculated by focusing on the shape of the frame itself and the scale of the frame itself, which makes the frame regression more accurate, and it can effectively improve the detection effect and outperform the existing methods. The experimental results show that compared with the original model, the segmentation accuracy mAP50 of the improved model reaches 33.8%, and compared with YOLOv5s, the optimised model improves the segmentation accuracy by 1.2%, which is able to complete the image segmentation task of the road scene more efficiently.

Keywords

Instance Segmentation, YOLOv5, RFAConv, Attentional Mechanism, Loss Function, Feature Fusion

1. 引言

近年来，随着地面交通网的不断完善与社会发展带来的出行里程激增，尽管车辆为出行提供了便利，但它们也引发了许多社会问题，比如交通拥堵和频繁的交通事故，这对人们的生命财产安全造成了巨大损失。同时，自动驾驶技术的发展可以显著提高交通系统效率，解决拥堵问题，增强驾驶安全，并解放用户的通勤时间[1]，具有良好的发展前景。对道路场景进行准确检测，这对车辆能够安全正常行驶是非常重要的。

语义分割是一项核心的环境感知技术，能够实现对场景的深入理解[2]，它可以对道路场景图像中的每个像素进行分类，将其归入相应的类别，从而实现像素级的分类，它能够为自动驾驶汽车提供关键的道路情况信息，来确保车辆安全行驶。

实例分割则进一步细化，它不但要实现像素级别的分割，还必须识别并预测各个实例的具体位置。这项任务具有较高的复杂性，涉及到智慧城市建设、医学图像分析以及自动驾驶等多个领域，且在这些应用场景中扮演着关键角色[3]。尽管实例分割技术取得了一定的进展，但仍面临若干挑战。因为在实际应用中，分割对象通常具有多层次的特征，光线和拍摄角度的变化可能导致目标形状的不同。此外，复杂和杂乱的背景也会给图像实例分割带来困难，所有这些因素都使得实例分割问题更加复杂。因此对道

路场景下的实例分割算法进行研究，对于促进智能驾驶的快速发展以及减少交通问题具有重要的理论和实践意义[4]。

图像分割算法主要分为传统方法和深度学习方法两类。传统的图像分割算法基于图像的灰度[5]、彩色[6]、空间纹理[7]和几何形状[8]等特征，将图像划分为多个独立区域，在这些区域中，像素点的特征大体相似，而不同区域之间的像素特征存在明显差异。特征提取的方法可分为基于阈值、边缘和区域等分割方法。传统的分割方法通常基于图像的视觉特征，如灰度或颜色，然而这些方法在分割时容易受到光照和噪声的影响，在复杂场景中性能较差，难以满足所需的分割精度，因此适用性不强。随着大型图像检测数据集的公开和计算机性能的提升，基于深度学习的检测方法逐渐被研究者们所青睐。深度学习驱动的实例分割算法主要分为单阶段(one-stage)方法(如 YOLACT [9] (you only look at coefficients)、SOLO [10] (Segmenting Objects by Locations)、YOLO [11]系列)和两阶段(two-stage)方法(如 Mask-RCNN [12] (Region Convolutional Neural Networks))，这些方法在实例分割任务中表现出色，有效提升了分割精度。针对上述问题，本文基于 YOLOv5 网络进行改进，主要研究工作如下：

(1) 针对目前空间注意机制不能强调每个特征在接受域的重要性，不能完全解决大卷积核的参数共享问题，提出 RFAConv [13]卷积结构并替换 YOLOv5 中 FPN [14] (Feature Pyramid Network)和 PAN [15] (Path Aggregation Network)特征融合模块中的部分卷积。RFAConv 卷积不仅考虑了接受域中每个特征的重要性，而且提供了几乎可以忽略不计的计算成本和参数增量，帮助模型更好地处理不同大小、形状和复杂度的对象，提高模型在不同场景下的鲁棒性和泛化能力。

(2) 提出 ShapeIOU [16]损失函数来优化边界框回归。本文使用 ShapeIOU 损失函数来替代原模型中 CioU [17]作为边界框损失函数，通过聚焦边框自身形状与自身尺度计算损失，考虑边框其自身的形状与尺度等固有属性对边框回归的影响，使边框回归更为精确，有效提升检测效果。

(3) 本文在 Cityscapes 数据集上进行实验。设计消融实验来验证各个模块的有效性，并与其他分割模型进行对比，验证了本文所改进算法的高效性。

2. 相关工作

2.1. 基于深度学习的实例分割方法

实例分割按照算法实现框架可分为：两阶段实例分割算法和单阶段实例分割算法[18]。何凯明等人设计提出的 Mask-RCNN 的两阶段方法遵循先检测后分割的理念，它基于 Faster-RCNN [19]另加入了一个分支网络，以输出目标对象的掩码，目的是在进行目标检测的同时，分割实例对象，比起类别标签或者检测框，掩码更加细粒度地提取实例对象的空间信息[20]。

YOLACT 是首个达成实例分割的单阶段模型，它将实例分割划分为预测原型掩膜和掩膜系数这两个并行的子任务，提升了模型推理效率。后续 WANG 等提出的 SOLO 按中心位置和对象尺寸来区分并分割不同实例。SOLOv2 [21]在 SOLO 的基础上引入动态掩膜头和非极大抑制，减少了运算量，分割速度进一步提高。Poly-YOLO [22]采用类似极坐标变换的方式，获得不是 mask 的掩码输出来进行分割，不占用太大显存，让整个网络更快、更轻量。YOLOv5 后的 YOLO 系列算法的分割思想均参照自 YOLACT，其中 YOLOv5s 在工程中应用最广。以上实例分割方法的实时性符合需求，但在不同特征对目标的敏感性存在不足，导致分割精度方面仍存在局限[23]。

2.2. YOLOv5 实例分割算法

YOLOv5 属于单阶段网络，有四种型号：YOLOv5s、YOLOv5m、YOLOv5l 和 YOLOv5x，主要取决于模型的深度、宽度[24]，其中 YOLOv5s 模型的参数量最小，检测速度最快，可以最大程度地满足各种

边缘设备的部署条件，综合考虑各模型的检测速度、模型大小及检测精度，本文选择 YOLOv5s 的 7.0 版本，其网络模型包括输入端(Input)、骨干网络(Backbone)、颈部网络(Neck)和输出端(Head)四部分。

YOLOv5s 模型的输入端将输入图像调整为统一的 640 * 640 像素的网络输入，在训练过程中，YOLOv5s 使用 Mosaic 数据增强[25]、自适应锚框计算和自适应图片缩放操作。骨干网络由 CBS 模块、C3 模块和 SPPF (Spatial Pyramid Pooling Fast)模块组成。其中 CBS 模块结构如图 1 所示，是骨干网络的基础性模块，包括对输入的特征图进行卷积、批标准化和 SiLU 激活函数操作；C3 模块结构如图 2 所示，由 3 个 CBS 模块和 1 个 BottleNeck_N 层组成，主要学习残差特征，其中 CBS 模块是卷积层，Bottle Neck_N 层用于压缩并恢复通道数；SPPF 模块结构如图 3 所示，通过池化层融合多尺度特征。

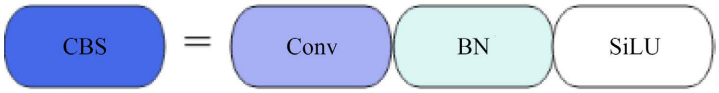


Figure 1. CBS structure chart
图 1. CBS 结构图

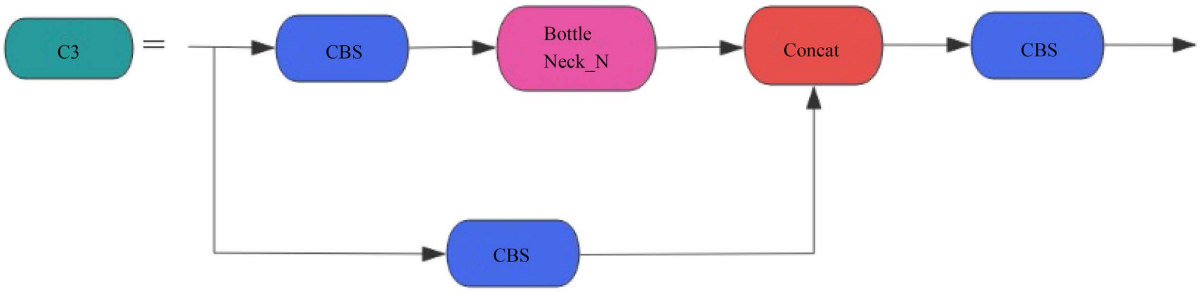


Figure 2. C3 structure chart
图 2. C3 结构图

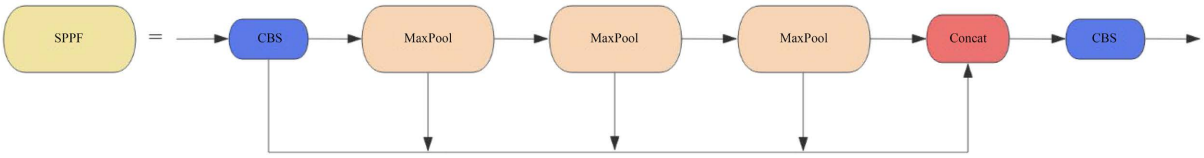


Figure 3. SPPF structure chart
图 3. SPPF 结构图

比起骨干网络，颈部网络的结构较为简单，主要包括 C3 模块、CBS 模块、上采样(Upsample)模块和拼接层(Concat)，构成了 FPN + PAN 结构，其中 FPN 层从顶部向下传达强语义特征，而 PAN 层从底部向上传达定位特征，以改善经过多层网络后的底层信息丢失的问题[26]。输出端有 3 个分支，从 Neck 模块输出的 20 * 20、40 * 40 和 80 * 80 大小的特征图去预测实例对象的类别概率、置信度和预测框坐标。ShapeIOU_Loss 被用于边界框的损失计算，同时使用非极大值抑制(NMS)找出最佳的目标检测框，消除冗余的边界框。

3. 改进 YOLOv5 道路场景实例分割模型

本文基于 YOLOv5 算法，引入实例分割思想，加入 RFAConv 卷积，创新空间注意力机制，将空间注意力机制的焦点从空间特征转移到感受野空间特征上，从而有效理解处理图像中的局部特征信息，提高特征提取和融合精确性，能够更有效地处理图像中的细节和复杂模式，在识别和定位目标时更加精准，

提升处理速度和效率，同时将原网络中的 CIOU 损失函数替换为 ShapeIOU 损失函数，优化边界框回归，有效处理细小目标和复杂边界的细节，提高模型在分割任务中的准确度。

3.1. RFAConv

注意机制使模型能够集中于显著特征上，增强了特征提取的效益和卷积神经网络捕获详细特征信息的能力，但是它也有一定的局限性，目前的空间注意机制不能强调每个特征在接受域的重要性，不能完全解决大卷积核的参数共享问题。而接受场注意(RFA)，它全面地解决了卷积核的参数共享问题，并考虑了接受域中每个特征的重要性，有效地融合多尺度特征。同时，RFA 设计的卷积运算(RFAConv)是一种开创性的方法，可以取代目前神经网络中的标准卷积运算。它提供了几乎可以忽略不计的计算成本和参数增量，同时显著提高了网络性能，本文使用 RFAConv 卷积替换 Head 模块中部分卷积。

RFAConv 卷积的接收域空间特征可以根据核大小动态生成，以 3 * 3 卷积核为例，RFAConv 整体结构如图 4 所示。其中“Spatial Feature”指的是原始的 Feature map，“接受域空间特征”是空间特征变换后的特征图，由不重叠的滑动窗口组成。当使用 3 * 3 卷积内核提取特征时，接收域空间特征中的每个 3 * 3 大小窗口代表一个接收域滑块。通过 RFA 得到的 feature map 是接受域空间特征，在“Adjust Shape”后没有重叠，如图 5 所示。该方法不仅强调了接收域滑块内不同特征的重要性，而且对接收域空间特征进行了优先排序。因此，学习到的注意图将每个接受域滑块的特征信息聚合起来，即注意力地图不再共享在每个接受域滑块。

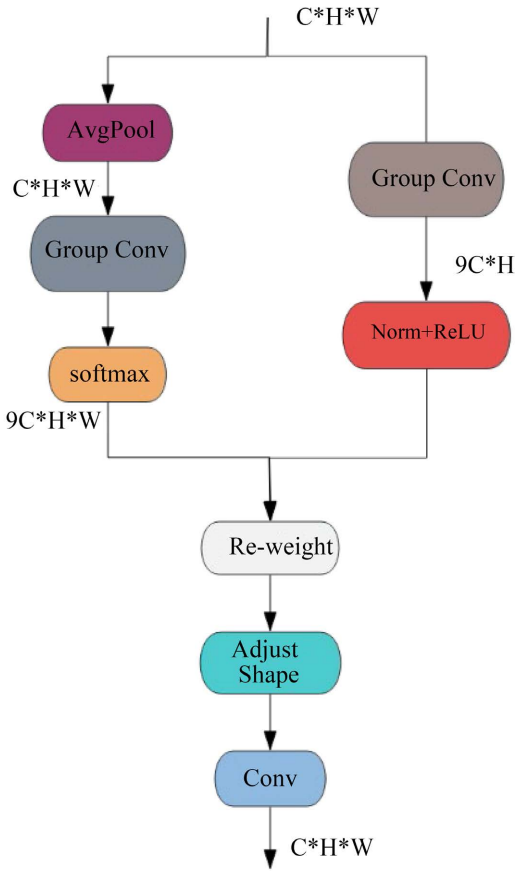


Figure 4. RFAConv structure chart
图 4. RFAConv 结构图

同时 RFAConv 利用相应大小的分组卷积来动态生成基于接受域大小的展开特征，以快速提取接受域特征，为了最小化与每个接收域特征交互时额外产生的计算开销和参数的数量，使用 AvgPool 来聚合每个接收域特征的全局信息。然后，使用 1×1 组卷积操作进行信息交互，最后使用 softmax 来强调每个特征在接受域特征中的重要性。RFA 的计算可以表示为：

$$F=\text{Softmax}\left\{g^{1\times1}\left[\text{AvgPool}\left(X\right)\right]\right\}\times\text{ReLu}\left\{\text{Norm}\left[g^{k\times k}\left(X\right)\right]\right\}=A_{rf}\times F_{rf} \tag{1}$$

式中：$g^{i\times i}$ 表示一个大小为 $i\times i$ 的分组卷积，k 表示卷积核的大小，Softmax 是归一化指数函数，AvgPool 表示平均池化，ReLU 是一种激活函数，Norm 表示归一化，X 表示输入特征图，F 是通过将注意力图 A_{rf} 与变换后的感受野空间特征 F_{rf} 相乘得到的。

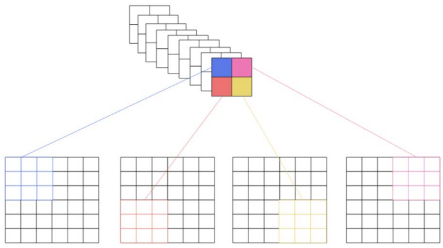


Figure 5. Adjust shape
图 5. Adjust shape

3.2. ShapeIOU

在目标检测和实例分割任务中，边界框损失函数是一种用来衡量模型预测的边界框与真实目标边界框之间差异的指标。其主要作用是帮助优化模型，使其能够准确地定位目标，并且提高目标检测的准确性。边界框损失函数会根据预测边界框和真实边界框之间的位置偏差，计算出一个损失值。模型在训练过程中，会通过最小化损失值来调整自身的参数，以使得预测的边界框更加接近真实的边界框[27]。

在 YOLOv5 原模型中使用 CIoU 作为边界框损失函数，CIoU 具体公式如下：

$$\text{CIoU}=\text{IoU}-\left[\frac{\rho^2\left(b,b^{gt}\right)}{c^2}+\alpha\upsilon\right] \tag{2}$$

$$\alpha=\frac{\upsilon}{\left(1-\text{IoU}\right)+\upsilon} \tag{3}$$

$$\upsilon=\frac{4}{\pi^2}\left[\arctan\frac{\omega_{gt}}{h_{gt}}-\arctan\left(\frac{\omega}{h}\right)\right]^2 \tag{4}$$

式中，IoU 表示交并比，用于衡量预测框与真实框的重叠程度；b 和 b^{gt} 分别为锚框和 GT 框的中心点；c 是包围预测框和真实框的最小外接框的对角线长度；α 是一个正的权衡参数，而 υ 用于衡量宽高比的一致性；ω^{gt} 和 h^{gt} 分别表示 GT 框的宽度和高度，ω 和 h 分别表示锚框的宽度和高度。

CIOU 综合考虑了预测框与真实目标框之间的位置、大小和形状等因素，以衡量二者之间的相似度。这种方法不仅惩罚边界框坐标的偏差，还包括长宽比和中心点距离，从而提升了预测精度和收敛速度。然而，CIoU 忽略了边框其自身的形状与尺度等固有属性对边框回归的影响。为弥补这方面的不足，本文使用 ShapeIOU 方法来替代 CIOU 作为边界框损失函数，ShapeIOU 能够通过聚焦边框自身形状与自身尺度计算损失，从而使得边框回归更为精确，能够有效处理不平衡数据、细小目标和复杂边界的细节，提

升分割精度且优于现有方法。如图 6 所示，b 和 b^{gt} 分别为锚框和 GT 框的中心点，$\left(x_c^{gt}, y_c^{gt}\right)$、$\left(x_c, y_c\right)$ 是其坐标，ω^{gt} 和 h^{gt} 分别表示 GT 框的宽度和高度，ω 和 h 分别表示锚框的宽度和高度。边界框的计算公式如下所示：

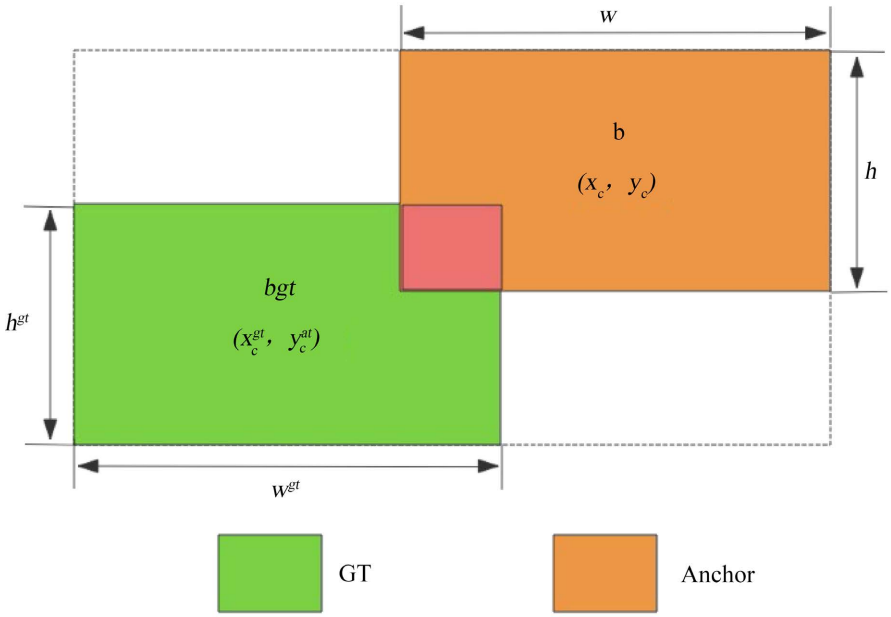


Figure 6. The graph of ShapeIOU calculated
图 6. ShapeIOU 计算图

$$\mathrm{IoU} = \frac{\left|B \cap B^{gt}\right|}{\left|B \cup B^{gt}\right|} \tag{5}$$

$$ww = \frac{2 \times \left(\omega^{gt}\right)^{\mathrm{scale}}}{\left(\omega^{gt}\right)^{\mathrm{scale}} + \left(h^{gt}\right)^{\mathrm{scale}}} \tag{6}$$

$$hh = \frac{2 \times \left(h^{gt}\right)^{\mathrm{scale}}}{\left(\omega^{gt}\right)^{\mathrm{scale}} + \left(h^{gt}\right)^{\mathrm{scale}}} \tag{7}$$

$$\mathrm{distance}^{\mathrm{shape}} = hh \times \frac{\left(x_c - x_c^{gt}\right)^2}{c^2} + ww \times \frac{\left(y_c - y_c^{gt}\right)^2}{c^2} \tag{8}$$

$$\Omega^{\mathrm{shape}} = \sum_{t=w,h} \left(1 - e^{w_t}\right)^{\theta}, \theta = 4 \tag{9}$$

$$\begin{cases} \omega_w = hh \times \dfrac{\left|w - w^{gt}\right|}{\max\left(w, w^{gt}\right)} \\ \omega_h = ww \times \dfrac{\left|h - h^{gt}\right|}{\max\left(h, h^{gt}\right)} \end{cases} \tag{10}$$

式中，B 和 B^{gt} 分别表示预测框和 GT 框；scale 为尺度因子，与数据集中目标的尺度有关，ww 与 hh 分

别为水平方向与竖直方向的权重系数，其取值与 GT 框的形状有关； $\text{distance}^{\text{shape}}$ 表示形状点 (x_c, y_c) 距离的平方， Ω^{shape} 是形状成本的定义，其对应的边框回归损失如下：

$$L_{\text{Shape-IoU}} = 1 - \text{IoU} + \text{distance}^{\text{shape}} + 0.5 * \Omega^{\text{shape}} \tag{11}$$

3.3. 改进后的 YOLOv5 模型

YOLOv5 是 YOLO 系列网络中最具代表性的网络，本文使用 YOLOv5 系列中兼顾精度和速度的 YOLOv5s-seg (7.0 版本)进行改进。改进后的网络结构图如图 7 所示。

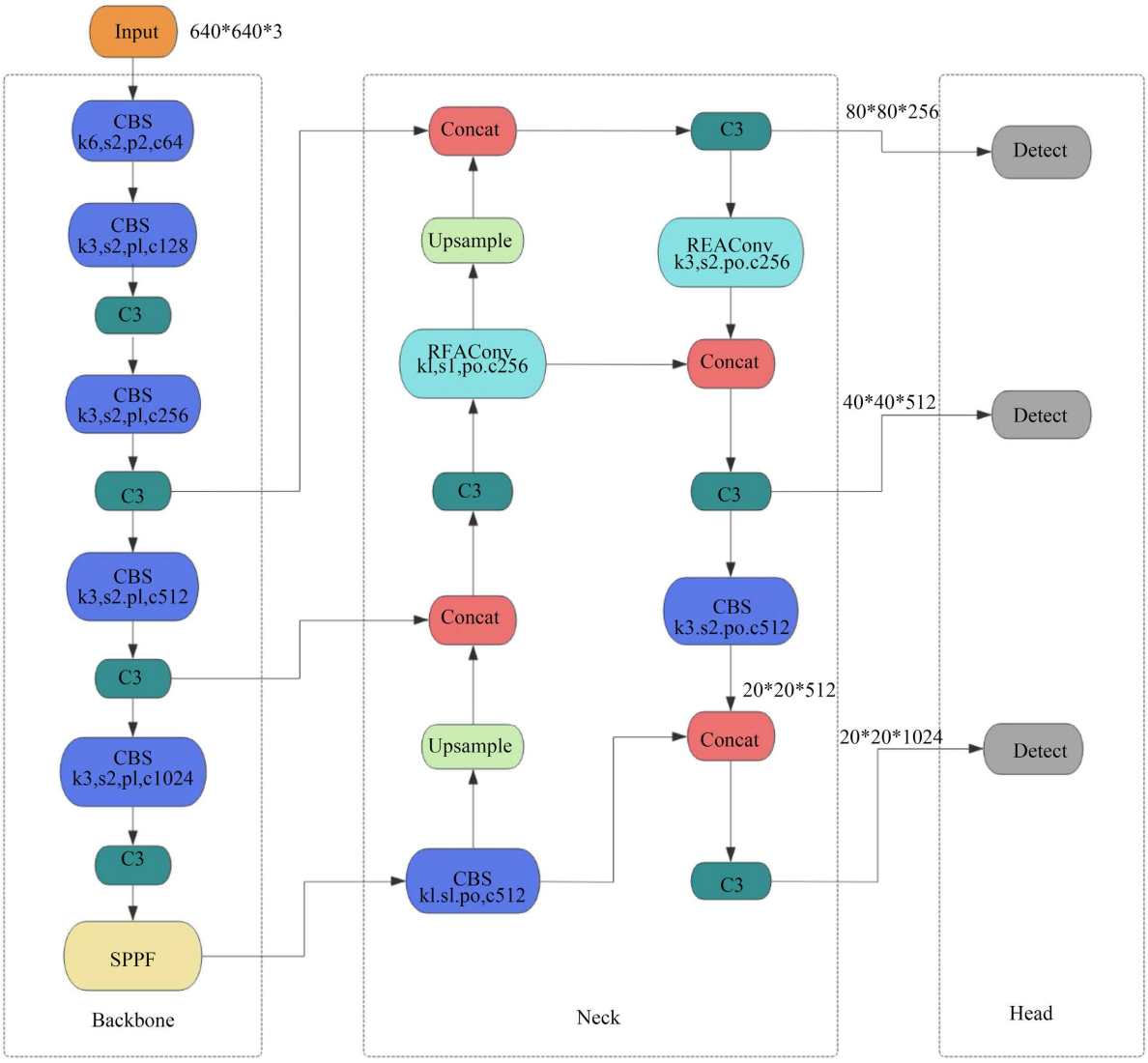


Figure 7. YOLOv5 network diagram
图 7. YOLOv5 网络图

4. 实验结果与分析

4.1. 实验环境

实验采用的深度学习框架为 Pytorch 1.9，参数环境为 CUDA 11.3，系统配置为 CPU i7 12 代 3080ti 显

卡，编译语言为 Python 3.9。训练轮次为 200，训练批次大小为 8，学习率设定为 0.01，输入图片大小为 640 * 640 像素。

4.2. 实验数据集与评估指标

4.2.1. 数据集

Cityscapes 数据集是由德国三家单位联合创建的道路场景下进行目标分割任务的数据集，该数据集以大规模图像数据、精确标注标签、高分辨率等特点被研究者广泛应用。它由 50 个不同城市、不同季节的 20,000 张粗略标注图像和 5000 张精细标注的街道场景图像组成。本文考虑到道路场景下的实例分割任务要求较高的精度，选取了 5000 张标注精细的目标图像进行实验，其中训练数据 2975 张，验证数据 500 张，还有未标注的 1525 张测试数据[28]。

4.2.2. 评价指标

本文选取了实例分割方法常用的评估指标——精准率(Precision, P)、召回率(Recall, R)、平均精度均值(mean Average Precision, mAP) [29]来评估模型的性能。其中精确率表示在所有被识别出的部位中，识别正确的部位所占的比例，体现了模型的分类准确性；召回率则是被正确识别的部位数量与测试集中所有部位数量的比值，反映了模型的目标检测全面性。平均精度值是通过预测掩码与实际目标部位的交并比(Intersection over Union, IoU)计算得到的，设置不同的 IoU 阈值会得到不同的 mAP 值，具体计算方法如下所示。当 IoU 阈值分别设为 0.50、0.75 和 0.50~0.95 时，对应的 mAP 分别表示为 mAP@0.50、mAP@0.75 和 mAP@0.50:0.95。本文选择 mAP@0.50 来衡量模型对所有类别部位的分割精度。

$$P = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{12}$$

$$R = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{13}$$

$$\mathrm{mAP} = \frac{1}{N}\sum_{n=1}^{N} \mathrm{AP}_n \tag{14}$$

式中：TP 表示预测正确实际为正的样本，FP 表示预测为正实际为负的样本，FN 表示预测为负实际为正的样本，N 为标签类别数量，AP (Average Precision)表示 PR 曲线下的面积。

4.3. 实验结果分析

4.3.1. 消融实验

为了验证各种改进策略对于模型精度提升的有效性，增强模型鲁棒性，以 YOLOv5s 为基准模型，采用消融实验来测试不同改进措施所取得的效果。结果如表 1 所示。

Table 1. Ablation results
表 1. 消融实验结果

模型	P/%	R/%	mAP@0.50/%
YOLOv5s	56.20%	30.37%	32.65%
YOLOv5s + RFAConv	56.86%	30.98%	32.78%
YOLOv5s + ShapeIOU	60.05%	29.61%	32.76%
YOLOv5s + RFAConv + ShapeIOU	56.86%	30.98%	33.82%

由表 1 可以看出，分别引入 RFAConv 卷积、ShapeIOU 损失函数均对分割精度有提升效果。RFAConv 旨在增强感受野的自适应性，使网络能够更有效地捕捉不同尺度和形状的目标特征。在实验中，我们分别使用标准卷积和 RFAConv 卷积进行对比，发现采用 RFAConv 后，模型在小目标检测和边界细节刻画方面表现更优，mAP@0.50 提升了 0.13%。而 ShapeIOU 作为改进的 IoU 计算方法，引入了对小目标和复杂边界的细节处理，通过对比标准 IoU 和 ShapeIOU 在不同目标类别上的表现，可以看到 mAP@0.50 精度提升了 0.11%。从消融实验结果来看，RFAConv 主要提升了模型的特征表达能力，使得目标边界更加清晰，而 ShapeIOU 则优化了 IoU 计算，提高了分割结果的匹配度。二者相结合后，通过在 Cityscapes 数据集上进行验证，可以看到模型的精确度 P 提升了 0.66%、召回率 R 提升了 0.61%、mAP@0.50 提升了 1.2%，使得改进后的 YOLOv5 在道路场景实例分割任务中取得了更优的性能，证明了两者结合后对提高道路场景实例分割任务有更好的效果。

4.3.2. 不同算法对比实验

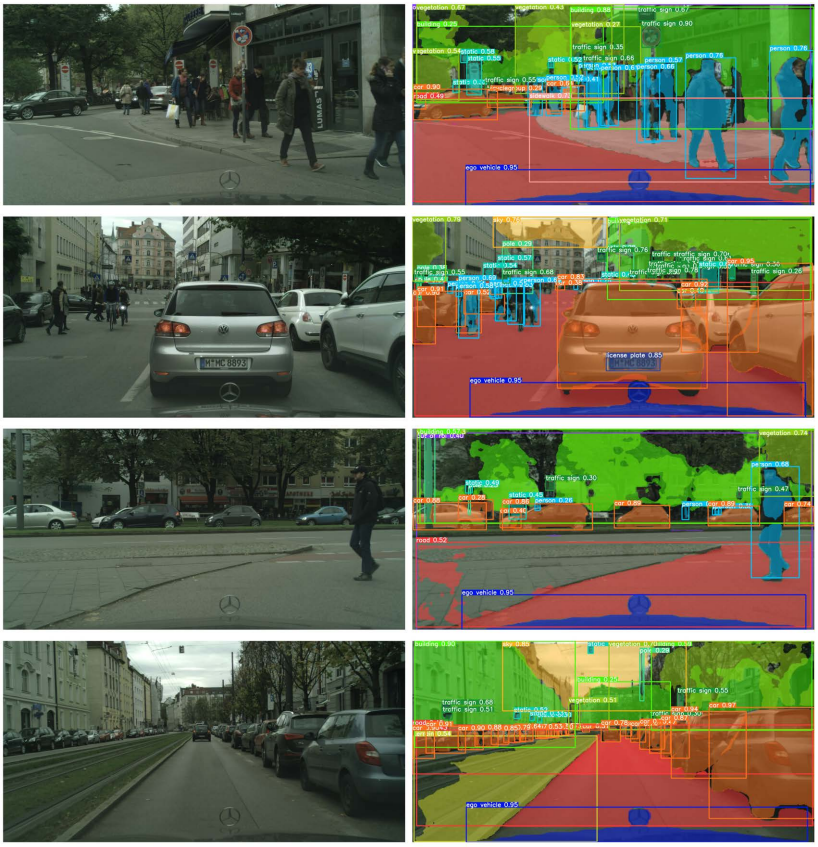
为进一步验证本文算法的优越性，设计对比试验，与 YOLOv5s 在相同数据集和实验环境下进行实验研究，以 Mask-RCNN、YOLACT 和 SOLOv2 这 3 种常见的实例分割算法作为对照组，其中 ours 表示本文提出的基于 YOLOv5 的改进模型。实验结果的评价指标对比如表 2 所示。Mask R-CNN 作为基于区域的分割方法，在分割精度上表现较好(map@0.50 = 32.70%)，但由于采用 RPN (Region Proposal Network)进行候选框生成，计算复杂度较高。YOLACT 采用一阶段方法，速度较快，但 map@0.50 仅为 21.54%，较难在小目标检测任务中保持高精度。SOLOv2 作为基于像素的分割方法，具有较好的分割效果(map@0.50 = 30.60%)，但其采用密集预测结构，而本文的改进 YOLOv5s 算法取得了最高的 map@0.50 精度，表 2 数据表明，本文所提出的 YOLOv5s 算法 map@0.50 精度达到了 33.82%，均优于 3 种对照算法，精度值达到了 56.86%，召回率达到了 30.98%，其精确率值虽然低于 Mask R-CNN，但总体上看，还是有优势的，主要归因于：RFAConv 考虑了接受域中每个特征的重要性，使网络能有效地融合多尺度特征，提升了 IoU 和边界分割精度；ShapeIOU 通过优化边界框回归，可以有效处理细小目标和复杂边界的细节，尤其对复杂形态物体(如行人、骑行者)有明显改进。通过对比试验，可知本文算法具有良好的分割性能。

Table 2. Comparison of experimental results among models in the CityScapes dataset
表 2. CityScapes 数据集上各模型对比实验结果

模型	P/%	R/%	mAP@0.50/%
YOLOv5s	56.20%	30.37%	32.65%
Mask R-CNN	59.90%	29.60%	32.70%
YOLACT	-	-	21.54%
SOLOv2	52.10%	30.30%	30.60%
ours	56.86%	30.98%	33.82%

4.3.3. 可视化结果

正文为了更直观地展示改进后的算法在 Cityscapes 数据集中的分割效果，体现指标优势，节选部分原图与分割效果图对比如图 8 所示。由图 8 可以看出，面对遮挡、形态接近、密集以及目标尺寸过小等道路场景下的诸多问题，本文改进后的算法均做出了准确识别，得到了较好的实例分割效果，验证了改进算法具有更高的检测精度。



(a) (b)

Figure 8. (a) Original test drawing; (b) segmentation effect drawing
图 8. (a) 原始测试图；(b) 分割效果图

5. 结论

本研究针对道路场景实例分割任务，提出了一种基于改进 YOLOv5s 的方法，并引入了 RFAConv 卷积和 ShapeIOU 损失函数以提升分割精度和模型的泛化能力。通过在 Cityscapes 数据集上的实验，并与 Mask R-CNN、YOLACT 和 SOLOv2 模型进行对比分析，验证了所提方法的有效性，并得出了以下主要结论：

(1) 通过引入 RFAConv 卷积，考虑到接受域中每个特征的重要性，全面解决了大卷积核的参数共享问题，在几乎不增加计算成本和参数量的情况下，同时显著提高了网络性能。

(2) 同时采用 ShapeIOU 代替 YOLOv5 中原损失函数 CIOU，考虑到边框自身的形状与尺度等固有属性对边框回归的影响，通过聚焦边框自身形状与自身尺度计算损失，使得边框回归更为精确。

实验证明，改进后的模型的分割精度 mAP@0.50 达到了 33.8%，相较于原始 YOLOv5s 模型，优化后的模型在分割精度上提高了 1.2%，能够更加高效地完成道路场景的图像分割任务，具有更好的准确性和鲁棒性，该方法可应用于自动驾驶、智能交通监控等领域，满足精度、实时性要求的道路场景实例分割任务。下一步将着重研究提升算法的实时性能，进一步改善模型性能，以更好地运用到实际中。

参考文献

[1] 林银辉. 基于卷积神经网络的道路场景分割方法研究[D]: [硕士学位论文]. 北京: 中国地质大学, 2019.

[2] 何淼楹, 崔宇超. 面向自动驾驶的交通场景语义分割[J]. 计算机应用, 2021, 41(S1): 25-30.

[3] 车子轩. 基于深度学习的城市街景图像分割方法研究[D]: [硕士学位论文]. 哈尔滨: 哈尔滨理工大学, 2023.

[4] 郭莉莉. 基于 GAN 的道路场景图像语义分割算法研究[D]: [硕士学位论文]. 绵阳: 西南科技大学, 2023.

[5] Chang, J., Fan, K. and Chang, Y. (2002) Multi-Modal Gray-Level Histogram Modeling and Decomposition. *Image and Vision Computing*, **20**, 203-216. https://doi.org/10.1016/s0262-8856(01)00095-6

[6] Dong, L., Dong, W., Feng, N., Mao, M., Chen, L. and Kong, G. (2017) Color Space Quantization-Based Clustering for Image Retrieval. *Frontiers of Computer Science*, **11**, 1023-1035. https://doi.org/10.1007/s11704-016-5538-y

[7] Xie, X., Wu, J. and Jing, M. (2013) Fast Two-Stage Segmentation via Non-Local Active Contours in Multiscale Texture Feature Space. *Pattern Recognition Letters*, **34**, 1230-1239. https://doi.org/10.1016/j.patrec.2013.04.016

[8] Pal, S., Chatterjee, S., Dey, D. and Munshi, S. (2018) Morphological Operations with Iterative Rotation of Structuring Elements for Segmentation of Retinal Vessel Structures. *Multidimensional Systems and Signal Processing*, **30**, 373-389. https://doi.org/10.1007/s11045-018-0561-9

[9] Bolya, D., Zhou, C., Xiao, F. and Lee, Y.J. (2019) YOLACT: Real-Time Instance Segmentation. 2019 *IEEE/CVF International Conference on Computer Vision* (*ICCV*), Seoul, 27 October-2 November 2019, 9156-9165. https://doi.org/10.1109/iccv.2019.00925

[10] Wang, X., Kong, T., Shen, C., Jiang, Y. and Li, L. (2020) SOLO: Segmenting Objects by Locations. In: Vedaldi, A., Bischof, H., Brox, T. and Frahm, J.M., Eds., *Computer Vision—ECCV* 2020, Springer, 649-665. https://doi.org/10.1007/978-3-030-58523-5_38

[11] Redmon, J., Divvala, S., Girshick, R. and Farhadi, A. (2016) You Only Look Once: Unified, Real-Time Object Detection. 2016 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Las Vegas, 27-30 June 2016, 779-788. https://doi.org/10.1109/cvpr.2016.91

[12] He, K., Gkioxari, G., Dollar, P. and Girshick, R. (2017) Mask R-CNN. 2017 *IEEE International Conference on Computer Vision* (*ICCV*), Venice, 22-29 October 2017, 2980-2988. https://doi.org/10.1109/iccv.2017.322

[13] Zhang, X., Liu, C., Yang, D., *et al*. (2023) RFAConv: Innovating Spatial Attention and Standard Convolutional Operation. arXiv: 2304.03198.

[14] Lin, T., Dollar, P., Girshick, R., He, K., Hariharan, B. and Belongie, S. (2017) Feature Pyramid Networks for Object Detection. 2017 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Honolulu, 21-26 July 2017, 936-944. https://doi.org/10.1109/cvpr.2017.106

[15] Liu, S., Qi, L., Qin, H., Shi, J. and Jia, J. (2018) Path Aggregation Network for Instance Segmentation. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 8759-8768. https://doi.org/10.1109/cvpr.2018.00913

[16] Zhang, H. and Zhang, S. (2023) Shape-IOU: More Accurate Metric Considering Bounding Box Shape and Scale. arXiv: 2312.17663.

[17] Zheng, Z., Wang, P., Ren, D., Liu, W., Ye, R., Hu, Q., *et al*. (2022) Enhancing Geometric Factors in Model Learning and Inference for Object Detection and Instance Segmentation. *IEEE Transactions on Cybernetics*, **52**, 8574-8586. https://doi.org/10.1109/tcyb.2021.3095305

[18] 何佳琦. 基于改进的轻量化 SOLOv2 鱼类图像实例分割方法研究[D]: [硕士学位论文]. 大连: 大连海洋大学, 2023.

[19] Girshick, R. (2015) Fast R-CNN. 2015 *IEEE International Conference on Computer Vision* (*ICCV)*, Santiago, 7-13 December 2015, 1440-1448. https://doi.org/10.1109/iccv.2015.169

[20] 徐博文. 基于深度学习的城市道路场景实例分割方法研究[D]: [硕士学位论文]. 长春: 吉林大学, 2022.

[21] Wang, X., Zhang, R., Shen, C., Kong, T. and Li, L. (2021) SOLO: A Simple Framework for Instance Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **44**, 8587-8601. https://doi.org/10.1109/tpami.2021.3111116

[22] Hurtik, P., Molek, V., Hula, J., Vajgl, M., Vlasanek, P. and Nejezchleba, T. (2022) Poly-Yolo: Higher Speed, More Precise Detection and Instance Segmentation for Yolov3. *Neural Computing and Applications*, **34**, 8275-8290. https://doi.org/10.1007/s00521-021-05978-9

[23] 赵南南, 高翡晨. 基于改进 YOLOv8 的交通场景实例分割算法[J/OL]. 计算机工程: 1-12. https://doi.org/10.19678/j.issn.1000-3428.0068677, 2024-07-13.

[24] 刘想德, 马昊. 基于 YOLOv5 的零件识别轻量化算法[J]. 组合机床与自动化加工技术, 2024(5): 100-104, 107.

[25] Bochkovskiy, A., Wang, C.Y. and Liao, H.Y.M. (2020) Yolov4: Optimal Speed and Accuracy of Object Detection. arXiv: 2004.10934.

[26] 于傲泽, 夏智权, 魏维伟, 等. 基于改进 YOLOv5 的空中目标部位快速分割算法[J]. 制导与引信, 2023, 44(4): 48-55.

[27] 马冬梅, 郭智浩, 罗晓芸. 改进 YOLOv5s-Seg 的高效实时实例分割模型[J]. 计算机工程与应用, 2024, 60(16): 258-268.

[28] 常夏宁. 基于卷积神经网络的道路场景分割算法研究[D]: [硕士学位论文]. 郑州: 河南工业大学, 2021.

[29] 高敏, 邹阳林, 曹新旺. 基于改进 YOLOv5 模型的织物疵点检测[J]. 现代纺织技术, 2023, 31(4): 155-163.