液晶与显示
Chinese Journal of Liquid Crystals and Displays

文章编号:1007-2780(XXXX)XX-0001-15

Madulous:基于多重注意力的双分支交通目标检测算法

朱　硕[1,3*],张绪康[2],曹恩琪[1],江　蕊[1],薛梓健[1]

(1.无锡学院 电子信息工程学院,江苏 无锡 214105;
2.南京信息工程大学 电子与信息工程学院,江苏 南京 210044;
3.无锡学院 江苏省通感融合光子器件及系统集成工程研究中心,江苏 无锡 214105)

摘要:随着城市智慧交通与自动驾驶的快速推进,复杂路口多目标检测面临小目标易漏检、高速移动目标易产生运动模糊、长距离遮挡识别困难以及难易样本训练失衡等问题,现有算法难以同时兼顾高精度与实时性的双重需求。本文提出基于多重注意力机制融合的双分支交通目标检测算法Madulous。算法以轻量级YOLOv8n为基础,构建DPFE双分支并行特征提取框架,训练阶段主、辅分支协同进行特征提取;在主干网络的C2F模块嵌入EMA高效多尺度注意力模块,实现空间与通道双重注意力校准;在颈部网络引入Swin Transformer模块,辅助算法融合更大范围的上下文信息;然后重构模型的分类损失函数,基于IoU动态阈值自适应提升困难样本的训练权重。对比实验表明,该算法mAP@0.5达91.5%,较基准YOLOv8n提升2.5个百分点;在EC-R3588SPC边缘平台推理速度达23.9 FPS,在雾天、红外航拍等交通场景下泛化性能显著优于主流算法。Madulous算法实现了检测精度与实时性的良好平衡,可有效支撑路口交通目标的实时监测,为降低车辆及非机动车行驶事故发生率提供了可靠的技术方案。

关　键　词:交通信息工程及控制;神经网络;Transformer;多重注意力;目标检测
中图分类号:TP394.1　　**文献标识码**:A　　**doi**:10.37188/CJLCD.2026-0080　　**CSTR**:32172.14.CJLCD.2026-0080

Madulous: a dual-branch traffic target detection algorithm based on multi-attention mechanisms

ZHU Shuo[1,3*], ZHANG Xukang[2], CAO Enqi[1], JIANG Rui[1], XUE Zijian[1]

(1. *College of Electronic Information Engineering, Wuxi University, Wuxi* 214105, *China*;
2. *College of electronic information and Communication Engineering, Nanjing University of Information Engineering, Nanjing* 210044, *China*;
3. *Jiangsu Province Engineering Research Center of Photonic Devices and System Integration for Communication Sensing Convergence, Wuxi University, Wuxi* 214105, *China*)

Abstract: With the rapid advancement of urban intelligent transportation and autonomous driving, the multi-target detection at complex intersections faces issues such as easy missed detection of small targets, easy motion blur of high-speed moving targets, difficulty in recognizing long-distance occlusions, and

收稿日期:2026-05-11;**修订日期**:2026-06-11.
基金项目:国家自然科学基金(No. 12473085)
Supported by National Natural Science Foundation of China (No. 12473085)
*通信联系人,E-mail:zshuo2011@163.com

imbalance in the training of easy and difficult samples. Existing algorithms are unable to simultaneously meet the dual requirements of high accuracy and real-time performance. This paper proposes Madulous, a two-branch traffic object detection algorithm based on the fusion of multiple attention mechanisms. Based on lightweight YOLOv8n, the DPFE dual-branch parallel feature extraction framework is constructed, and the main and auxiliary branches cooperate to extract features in the training phase. EMA efficient multi-scale attention module was embedded in the C2F module of the backbone network to realize the spatial and channel dual attention calibration. The Swin Transformer module was introduced into the neck network to assist the algorithm to fuse a wider range of context information. Then, the classification loss function of the model is reconstructed, and the training weight of difficult samples is adaptively increased based on IoU dynamic threshold. Comparative experiments show that the proposed algorithm mAP@0.5 reaches 91.5%, which is 2.5 percentage points higher than the benchmark YOLOv8n. The inference speed of EC-R3588SPC edge platform reaches 23.9 FPS, and the generalization performance is significantly better than the mainstream algorithms in traffic scenes such as fog and infrared aerial photography. Madulous algorithm achieves a good balance between detection accuracy and real-time performance, which can effectively support the real-time monitoring of traffic targets at intersections, and provides a reliable technical solution to reduce the incidence of vehicle and non-motor vehicle driving accidents.

Key words: transportation information engineering and control; neural network; transformer; multi-attention mechanism; object detection

1 引　　言

随着自动驾驶及城市智慧交通的快速发展，道路目标检测技术的重要性日益凸显。使用目标检测方法可以实时、准确地识别道路上的车辆、行人、交通标志等关键目标，及时预警潜在危险，减少交通事故的发生；道路目标检测与跟踪技术的发展也为自动驾驶系统提供必要的感知信息，不仅提高了交通安全性，还使得自动驾驶车辆能够更好地适应复杂多变的道路环境，为自动驾驶系统的决策与控制提供重要依据。目标检测是推动自动驾驶与智能交通领域发展的关键技术，对于提高交通安全性、提升交通效率以及推动智能交通系统的智能化发展具有重要意义[1-2]。

近年来，随着深度学习的快速发展，目标检测与识别技术取得了较大突破，利用深度学习实现目标检测可实时分析视频或图像数据，快速识别并定位复杂场景中的目标对象，无论是行人、车辆还是特定目标，均能实现高效检测。基于深度学习的目标检测算法现阶段主要分为One-stage算法和Two-stage算法[3]，Two-stage算法在识别时首先需要将图片生成候选框，再对候选区域进行目标检测，如区域卷积神经网络(region-based convolutional neural network, R-CNN)[4]、快速区域卷积神经网络(faster region-based convolutional neural network, Faster R-CNN)[5]和掩码区域卷积神经网络(mask region-based convolutional neural networks, Mask R-CNN)[6]。Han[7]提出了一种基于改进Faster R-CNN的实时小交通标志检测方法，使用小区域建议生成器来提取小型交通标志的特征，增强了系统对小型交通标志区域的鲁棒性。Sarumathi[8]采用Faster R-CNN和区域候选网络(region proposal network, RPN)联合完成识别任务，然后使用随机森林算法对给定数据集进行分类和回归。汪菊等[9]以Mask R-CNN为基础，在主干网络中引入极自注意力机制提升特征提取能力，在特征金字塔顶层添加高效通道注意力模块(efficient channel attention, ECA)分支以优化高层次低分辨率语义信息图，并使用余弦退火算法优化训练过程，加快模型收敛速度。Two-stage算法在目标检测任务中准确率较高，但整体速度较慢。One-stage算法与Two-stage算法相比，不但可以满足高精度的要求，而且能大大提高图片处理速度，保证结果的有效性，如单步骤多框检测器(single shot multibox detector, SSD)[10]、你只看一次系列网络(you only

look once，YOLO）[11-14]和实时 Transformer 检测器（real-time detection transformer，RT-DETR）[15]。Li等[16]以 YOLOv3 为基础，提出了用于行人检测的 YOLO-ACN 算法，算法加入卷积注意力机制（convolutional block attention module，CBAM）来优先考虑小目标检测，并替换损失函数为完整交并比损失函数（complete intersection over union，CIoU），用深度可分离卷积代替传统卷积操作，进一步提高检测小目标和障碍物的精度和速度。Guo等[17]采用特征增强和融合技术迭代提取小目标信息，整合来自不同特征层的信息，以增强SSD算法的特征提取能力，提高模型对小型车辆目标的检测精度，但网络复杂度随之增加，影响模型的推理速度。Bie等[18]采用双向特征金字塔网络实现多尺度特征的有效交互，削弱复杂环境对车辆检测的负面影响，然而模型在高算力硬件平台上的性能提升不够明显。王龙春等[19]针对自动驾驶环境与交通目标的精细化检测，基于YOLOv8架构进行了多维度的网络优化，有效提升了特征提取的效率与目标识别精度。同时，在低照度等恶劣条件下，王栋等[20]提出了CORM-YOLO算法，该算法通过图像增强模块提升了暗光环境下的特征提取能力，并降低额外的计算开销，在增强模型环境鲁棒性的同时，依然具备高效的实时检测能力。

上述算法在交通检测领域的性能与场景适配性差异明显。Two-stage算法精度高但推理速度过低、参数量较大，仅适用于离线高精度场景。One-stage算法中，SSD曾广泛应用于早期车载辅助系统，但小目标与遮挡目标漏检率较高；RT-DETR对于复杂场景的检测能力较强，但算法的计算量也限制其仅在高算力平台才能够发挥效果。因为交通检测系统多部署于资源受限的边缘计算设备，YOLO系列在精度、速度与部署便捷性上的综合平衡优势远超其他算法。所以当前城市路口实时监测、边缘车载感知、无人机交通巡检等场景均以YOLO系列算法为主流选型。

本文选取YOLOv8n作为基准模型进行改进。该模型在保持低参数量与计算复杂度的同时，能够在边缘计算平台上实现满足交通监测行业要求的实时推理速度。其采用的无锚框设计与C2F特征提取模块能够提升多尺度特征融合能力，适配交通场景中目标尺度差异大的特点。同时，YOLOv8n依托完善的开源生态，支持多平台硬件加速，可有效降低工程落地成本，其分层模块化的架构设计也为后续的特征提取与融合机制改进提供了良好的基础。

尽管YOLOv8n在实时交通检测中表现优异，但在复杂路口场景下仍存在尺度适应性差与局部特征感知不足的问题，导致在实时检测过程中，由于目标的动态变化以及目标之间的相互遮挡，导致模型的检测精度和效率下降，引发误检和漏检的问题。为解决上述问题，本文做出如下贡献：

一、提出多重注意力机制融合的双分支交通目标检测方法Madulous，以YOLOv8n为基础构建双分支网络，通过主分支完成核心特征提取与融合，辅助分支补充多级语义信息，兼顾图像全局特征与局部细节，有效减少模型在复杂交通场景下的特征信息退化问题。

二、针对高速移动目标与小目标细节特征易丢失、常规注意力机制感知范围有限的问题，将高效多尺度注意力机制（efficient multi-scale attention，EMA）嵌入至主干网络，重新组织通道维度与空间维度，并进行信息的跨维度交互，提升网络对于图像信息的细节捕捉能力。

三、对于长距离遮挡与密集分布目标难以建立空间关联的问题，在颈部网络引入移位窗口Transformer（shifted window transformer，Swin Transformer），通过多种不同下采样率，融合多尺度特征信息。

四、重构分类损失函数，通过自适应困难样本加权机制增强模型的整体泛化能力，解决交通场景中简单样本主导训练、困难样本识别能力不足的问题。

2　算法模型改进

2.1　算法整体结构

在视觉感知任务中，YOLOv8n作为单阶段目标检测器，通过无锚框设计与解耦头架构，在检测精度与端侧推理速度之间取得了良好的平衡[21]。但是算法在检测道路目标时，存在的尺度适应性差与局部特征感知不足的问题，导致出现识别精度瓶颈。为解决上述问题，本文提出基于多重注意力机制融合的双分支交通目标检测算

法 Madulous，算法以 YOLOv8n 为基础架构，在 Backbone 层级部署并行双分支特征提取框架，并于 C2F 模块输出端嵌入 EMA 模块，通过跨通道与空间维度的特征交互增强模型对关键局部特征的鉴别能力；为解决动态目标尺度变化问题，在 Neck 的下采样阶段集成 Swin Transformer 模块，利用其分层窗口自注意力机制建模长程空间依赖关系；同时重构分类损失函数，引入困难样本加权机制以提升模型泛化性能。完整网络架构如图 1 所示。

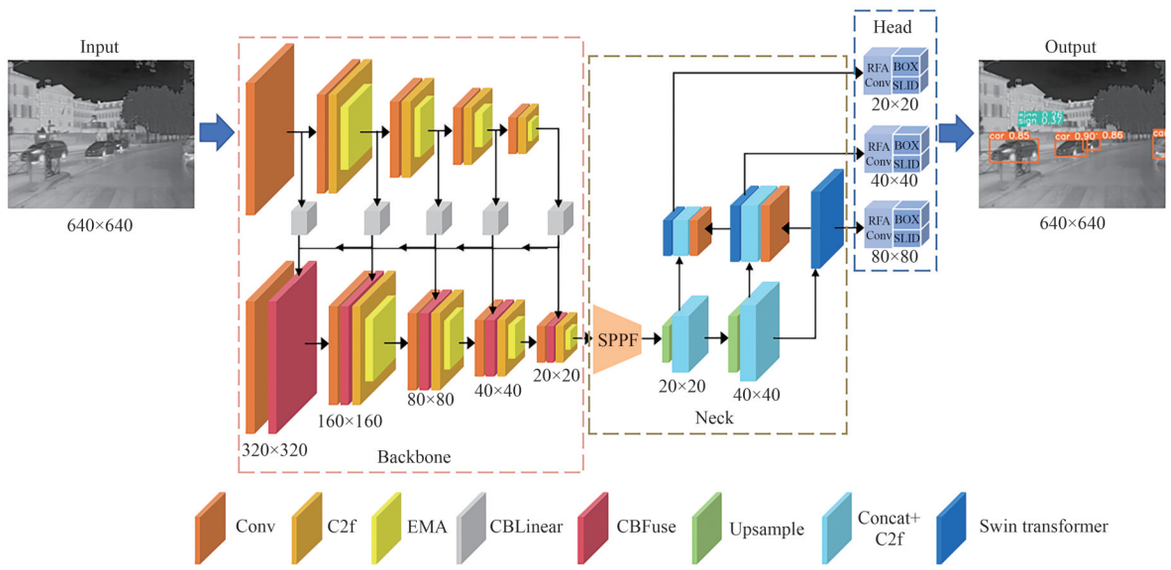


图 1 Madulous 算法结构

Fig. 1 Madulous algorithm structure

2.2 双分支并行特征提取框架

神经网络在特征提取过程中，随着网络层次加深，易出现信息瓶颈和梯度信号弱化的问题。YOLOv8n 采用单分支主干网络结构，所有层级共享同一条信息传递路径，浅层的细粒度特征与深层的语义特征无法形成有效互补，在处理复杂交通场景时，易出现深层语义强但细节丢失、浅层细节足但语义模糊的矛盾，难以同时兼顾全局上下文与局部细节的有效表征。为解决这一问题，本文提出双分支并行特征提取框架（dual-branch parallel feature extraction，DPFE）作为 Madulous 算法的核心结构。该框架旨在确保主分支高效推理的同时，通过辅助机制克服深度网络的信息损失与优化困难，显著提升对交通目标细节特征的表征能力。双分支并行特征提取框架如图 2 所示。

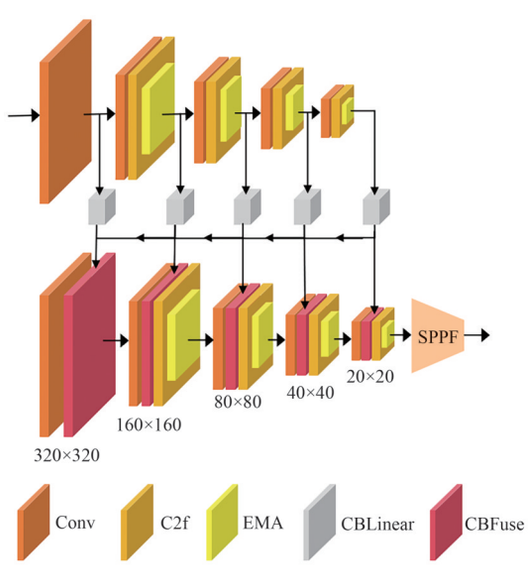


图 2 DPFE 框架

Fig. 2 DPFE framework

本框架的核心由两条并行处理路径构成：主分支与辅助可逆分支。主分支采用特征金字塔网络，承担主干特征提取任务，其网络结构与层级设计直接服务于最终的检测目标，是推理阶段唯一保留的路径，确保算法效率不受额外计算负担影响。辅助可逆分支则在训练阶段协同工作，解决网络加深导致的特征信息瓶颈及由此引发的梯度不可靠问题。该分支通过引入可逆结构设计，维持从输入数据到目标信号的完整信息流映射。

在主分支与辅助分支信息交互过程中嵌入复合主干线性模块（composite backbone linear，CBLinear）。CBLinear模块不同于简单的可逆网络堆叠模块，它对标准卷积层、批归一化层及激活函数计算流程进行优化重构，在训练阶段高效模拟出类可逆特性，保障信息在分支内流动的低损性，从而能够生成精确、指导性强的梯度信号。其计算公式如式（1）所示：

$$\mathrm{CBLinear}(X;\theta)=X+\mathrm{ReLU}(\mathrm{BN}(\mathrm{Conv}_{1\times1}(\mathrm{DWConv}_{3\times3}(X;\theta_1);\theta_2);\theta_3)),\tag{1}$$

其中，X为输入特征，θ为模块的可学习参数，ReLU(·)为激活函数，$\mathrm{DWConv}_{3\times3}(\cdot)$、$\mathrm{Conv}_{1\times1}(\cdot)$和BN(·)分别对应深度卷积、点卷积和批归一化，$\theta_1$、$\theta_2$和$\theta_3$分别为深度卷积、点卷积和批归一化的可学习参数。

基于CBLinear模块，辅助可逆分支采取双路可逆变换结构，设输入特征图为X，沿通道维度分为X_1和X_2，可逆分支的正向映射如式（2）所示：

$$\begin{cases}Y_1=X_1+\mathrm{CBLinear}(X_2;\theta_F)\\Y_2=X_2+\mathrm{CBLinear}(Y_1;\theta_G)\end{cases},\tag{2}$$

其中，θ_F为第一个CBLinear模块的可学习参数，θ_G为第二个CBLinear模块的可学习参数，均包含对应模块内卷积层与批归一化层的权重和参数。

在训练过程中，辅助可逆分支通过CBLinear模块将预测结果注入主分支的对应层级。这种梯度协同机制使主分支在更新参数时，能够有效接收来自完整信息流的可靠指导，显著增强在深层特征中提取任务关键信息的能力。

为协调主、辅分支以及后续多尺度特征金字塔之间的信息交互，并解决深度监督架构中可能存在的误差累积及目标信息断裂问题，框架进一步引入了跨分支融合模块（cross-branch fusion，CBFuse），采用通道注意力加权融合方式，作为信息融合中枢。模块首先基于主分支特征X_m和辅助分支特征X_a生成全局描述符g_m和g_a，计算公式如式（3）所示：

$$\begin{cases}g_m=\mathrm{GlobalAvgPool}(X_m)\\g_a=\mathrm{GlobalAvgPool}(X_a)\end{cases},\tag{3}$$

其中，GlobalAvgPool(·)为全局平均池化。

通过全连接网络生成自适应权重ω，计算公式如式（4）所示：

$$\omega=\sigma(\mathrm{FC}(\mathrm{ReLU}(\mathrm{FC}([g_m;g_a])))),\tag{4}$$

其中，$\sigma(\cdot)$为Sigmoid函数，FC(·)为全连接网络。

然后根据权重对两个分支的特征进行逐通道加权求和，得到融合输出如式（5）所示：

$$Y=\omega\otimes X_m+(1-\omega)\otimes X_a,\tag{5}$$

其中，⊗表示逐元素相乘符号。

CBFuse模块能够接收来自主分支不同层级、辅助分支以及前期网络阶段的特征图，根据不同交通场景自适应调整主辅分支的信息贡献比例，既保留了主分支的核心语义信息，又补充了辅助分支的细粒度细节特征。

2.3　EMA模块

交通场景中高速移动目标易产生运动模糊，且小目标占比高，导致图像关键信息与背景对比度降低、细节特征丢失，进而削弱网络对局部区域的感知能力。YOLOv8原有的C2F模块缺乏对关键区域的自适应聚焦机制，而CBAM等常规注意力机制[22]多依赖单尺度特征交互，难以同时捕获长程空间依赖与局部细粒度语义，无法有效区分目标与背景噪声，为针对性解决此问题，本文引入EMA模块[23]，EMA结构如图3所示。

EMA的核心创新在于提出一种双分支跨尺

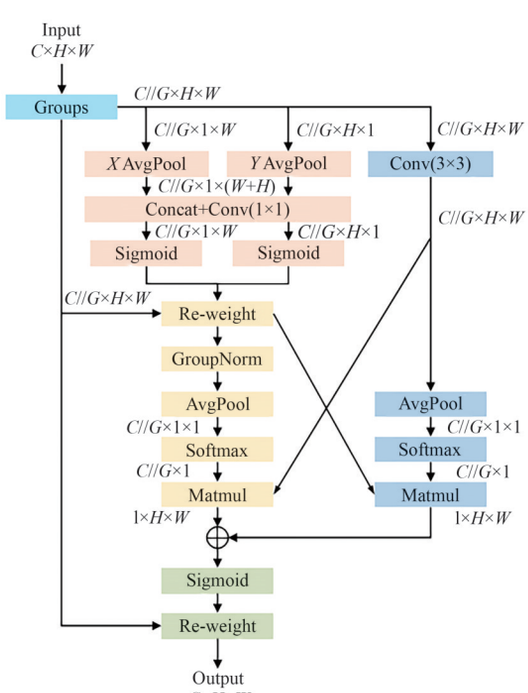


图3　EMA结构图

Fig. 3　Structure diagram of EMA

度协同与动态融合机制。该机制在特征融合阶段运作：首先将输入为$C\times H\times W$的特征图沿通道维度划分为G组子特征，然后分别分解至两条异构处理路径，划分公式X如式(6)所示：

$$X=[X_0,X_1,\cdots,X_{G-1}],X_i\in R^{C//G\times H\times W},G\ll C, \tag{6}$$

其中，X_i为第i组子特征。

第一条路径为多尺度通道交互分支，通过沿水平方向和垂直方向的一维全局池化操作分别捕获特征图在宽度和高度维度的长程依赖信息。随后将这两个方向的特征进行拼接，并通过1×1卷积运算生成一个蕴含空间维度重要性权重的特征表示。第二条路径为局部语义增强分支，采用3×3卷积核提取局部邻域内的细粒度结构特征。水平方向一维全局平均池化Z_C^H如式(7)所示，垂直方向一维全局平均池化Z_C^W如式(8)所示：

$$Z_C^H=\frac{1}{W}\sum_{i=0}^{W}x_C(H,i), \tag{7}$$

$$Z_C^W=\frac{1}{H}\sum_{j=0}^{H}x_C(j,W), \tag{8}$$

其中，H和W分别对应特征图的高度和宽度，C为通道数，$x_C(H,i)$为表示第C个通道特征图中第H行第i列的像素值，$x_C(j,W)$为表示第C个通道特征图中第j行第W列的像素值。

两条路径进入跨空间信息交互阶段，首先对各路径信息进行二维全局平均池化和Softmax函数线性变换，分别将路径输出维度转化为$R_1^{1\times1\times C//G}\times R_3^{C//G\times H\times W}$和$R_3^{1\times1\times C//G}\times R_1^{C//G\times H\times W}$。二维全局平均池$Z_C$如式(9)所示：

$$Z_C=\frac{1}{H\times W}\sum_{j=0}^{H}\sum_{i=0}^{W}x_C(i,j), \tag{9}$$

其中，$x_C(i,j)$为第C个通道特征图中第i行第j列的像素值

采取跨分支动态融合策略，将第一条路径生成的空间权重特征与第二条分支提取的局部特征进行矩阵点积运算，实现局部特征在空间维度上的自适应加权，最终输出融合了空间与通道双重注意力权重的特征图，帮助模型捕捉图像中的像素级关系，增强特征表示的能力。

2.4 引入Swin Transformer模块

交通场景中目标尺度差异大、分布密集，这对模型全局上下文建模与多尺度特征提取能力提出了更高要求。而CNN架构的下采样模块感受野有限，难以有效构建被长距离遮挡目标间的空间关联，无法充分利用大范围道路的上下文信息辅助目标识别，导致密集重叠目标的漏检率较高。为应对这一挑战，本文引入Swin Transformer[24]作为下采样结构。不同于常规Transformer在视觉任务中的全局计算方式，Swin Transformer创新的层级式设计与移位窗口机制，能够显式解决道路检测中长距离依赖建模与局部细节保留的平衡问题。模块结构如图4所示。

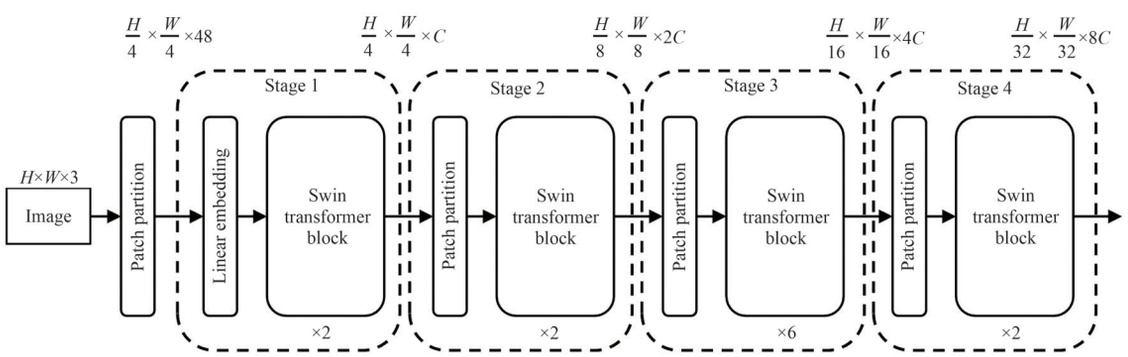


图4 Swin Transformer模块结构

Fig. 4 The architecture of a Swin Transformer

Swin Transformer采用四阶段层级式架构，先对输入特征图进行Patch Embedding处理，将图像划分为互不重叠的4×4像素块，得到通道数为48的特征张量，输入Linear Embedding模块通过线性变换将通道维度投影至C，得到初始特征。该特征随后进入由多个Swin Transformer Block组成的渐进式下采样阶段。Swin Transformer Block由窗口多头自注意力(window multi-head

self-attention，W-MSA）[25]和移位窗口多头自注意力（shifted window multi-head self-attention，SW-MSA）[26]串联构成。设输入特征图为X，进入W-MSA，通过残差连接和注意力计算W-MSA(·)得中间特征$\hat{X}^l$，然后使用多层感知机MLP(·)和层归一化LN(·)获得W-MSA模块输出X^l，具体计算公式如式(10)和式(11)所示：

$$\hat{X}^l=\text{W-MSA}(\text{LN}(X^{l-1}))+X^{l-1}, \quad (10)$$

$$X^l=\text{MLP}(\text{LN}(\hat{X}^l))+\hat{X}^l, \quad (11)$$

其中，l为模块的层数索引。

然后X^l作为输入进入SW-MSA，同样经由残差连接和注意力计算SW-MSA(·)得到中间特征$\hat{X}^{l+1}$，再通过多层感知机MLP(·)和层归一化LN(·)获得Swin Transformer Block的输出X^{l+1}，具体计算公式如式(12)和式(13)所示：

$$\hat{X}^{l+1}=\text{SW-MSA}(\text{LN}(X^l))+X^l. \quad (12)$$

$$X^{l+1}=\text{MLP}(\text{LN}(\hat{X}^{l+1}))+\hat{X}^{l+1}. \quad (13)$$

W-MSA将特征图划分为固定大小的非重叠窗口，仅在窗口内计算自注意力，大幅降低计算复杂度；而SW-MSA通过将窗口向右下角循环移位半个窗口尺寸，使相邻窗口信息交互，从而在维持计算效率的同时，有效捕捉被路面设施遮挡或分散分布的目标间关联。运用渐进式下采样策略能够为道路检测任务提供多尺度特征，其中浅层高分辨率特征保留交通标志、行人等小目标的边缘与纹理；深层低分辨率特征整合大范围道路结构信息，二者协同提升模型对复杂道路场景的鲁棒性。

2.5　分类损失函数改进

在道路目标检测中，简单样本的数量远多于困难样本，训练过程易被简单样本主导，且YOLOv8n的分类损失函数对所有样本采用统一权重，无法自适应提升困难样本的训练优先级，导致模型对遮挡目标、小目标等困难样本的识别能力不足，影响实际检测场景下的泛化性能。基于此，本文采用滑动损失函数（slide loss，SLID）。SLID提取全部边界框IoU[27]的均值μ作为临界点，据此将样本划分为简单样本和困难样本，辅以加权函数来突出边界样本的权重，其计算公式如式(14)所示：

$$f(x)=\begin{cases}1, x\leqslant\mu-0.1\\ e^{1-\mu}, \mu-0.1<x<\mu\\ e^{1-x}, x\geqslant\mu\end{cases}, \quad (14)$$

其中，阈值间隔设置为0.1，0.1的间隔宽度对应目标检测中边界样本的典型分布范围：IoU低于$\mu-0.1$的样本与真实目标重叠度极低，多为背景误检或完全不相关的区域，属于简单负样本，无需额外加权；而IoU处于$(\mu-0.1, \mu)$区间的样本为边界困难样本，这类样本与真实目标有一定重叠但定位不准确，是交通场景中遮挡目标、小目标和高速移动目标的主要分布区域，也是导致模型漏检和误检的重要原因，因此赋予高权重，以提升其训练优先级；IoU高于μ的样本为简单正样本，模型已能准确识别，采用指数衰减机制逐步降低其权重，避免易样本主导梯度更新过程。

SLID引入基于IoU的动态权重映射机制，优先获取目标框与预测框的IoU值，将其作为核心参数注入损失计算中，使得网络能够更敏锐地感知多层级的预测置信度，实现对损失值的精细化约束，借助边界框的空间回归特征提升检测的准确率及鲁棒表现。

3　实　　验

3.1　实验数据集

本文使用的交通目标检测数据集为作者团队独立采集构建，采集地点为城市平交路口，包含主干道交叉口、学校周边路口及商业综合体周边路口，采集时间覆盖工作日早高峰7:00～9:00、平峰11:00～13:00及晚高峰17:00～19:00时段。采集设备为海康威视DS-2CD3T46WD-I3网络摄像头，采集场景涵盖机动车道、非机动车道及人行横道，能够反映城市复杂路口的交通流特征。在获得原始数据后，使用Vott标注工具对图片进行人工标注，标注类别分为electrocar、bicycle和person三类，用于识别非机动车和行人目标。为满足模型的训练与性能评估需求，将筛选后的5 829幅有效图像严格遵循7:2:1的分配比例，拆分为训练、测试及验证三个数据子集以供实验调用。

本文还使用两种不同领域的数据集来验证本文模型在道路检测下的通用性。其中，选用RTTS数据集测试雾天场景下的目标识别能力，该集合内含4 322张真实雾天采集的图片，涵盖5类目标标签。此类图像中行人尺度微小、车流

密集,且常伴随人车严重遮挡现象,极大增加了目标识别的复杂性。另一个数据集是InfiRay公司开源的红外航拍人车检测数据集,该数据集共8 402张图像,有7种标注对象,小目标信息较多。上述数据集同样按照7∶2∶1的比例划分,用于模型泛化能力评估。

3.2 实验平台

为确保实验的公平性,各组测试均依托于统一的软硬件环境及超参数配置完成,平台运行环境如表1所示,训练参数如表2所示。

表1 实验配置

Tab. 1 Experimental configuration

Configuration name	Configuration information
Operating System	Windows10
CPU	Intel Core i7 10750H
GPU	NVIDIA GeForce RTX 2070
Programming languages	Python3.9
Algorithm framework	Pytorch1.12
Acceleration environment	CUDA11.3

表2 训练参数

Tab. 2 Training parameters

Parameter names	Parameter information
Learning rate	0.01
Weight decay coefficient	0.000 5
Image size	640×640
Batch size	8
Epoch	300

3.3 评价指标

本节针对道路目标检测任务建立多维度的评估标准。通过计算准确率(Precision)、召回率(Recall)和平均精度均值(mean average precision, mAP)来验证算法的检测水平;利用模型参数规模(Params)及其浮点数(GFLOPs)来侧面评估网络架构的复杂程度。

其中,准确率衡量所有被系统判定为正类的样本中,真实目标所占的概率;而召回率则反映了数据集中所有真实存在的标签被算法成功找出的概率。数学表达式如式(15)与式(16)所示:

$$\text{Precision}=\frac{\text{TP}}{\text{TP}+\text{FP}}, \tag{15}$$

$$\text{Recall}=\frac{\text{TP}}{\text{TP}+\text{FN}}, \tag{16}$$

其中,TP代表被网络成功捕获并准确定位的真实目标总数;FP指代模型将背景或干扰因素错判为既定目标的次数;而FN则对应漏检数量,说明原本存在于图像中的真实目标未能被算法有效识别的情况。

平均精度AP是将Recall作为横轴,Precision作为纵轴所形成的二维图形的面积,用于衡量神经网络的整体性能。平均精度均值mAP是所有类别AP的平均值,用于表示网络模型预测的识别精度。计算公式如式(17)、式(18)所示:

$$\text{AP}=\int_0^1 P(R)\mathrm{d}R, \tag{17}$$

$$\text{mAP}=\frac{1}{m}\sum_{i=1}^{m}\text{AP}_i, \tag{18}$$

其中,m代表类别个数。

模型的空间复杂度通常由Params来衡量,该指标直观反映了网络架构的繁冗程度,结构越庞大,其在部署时占的存储资源便越多。GFLOPs被用于量化算法在执行推理或训练阶段的计算开销,该数值越高,说明模型对底层硬件的算力需求越苛刻,直接影响实时检测的帧率。

3.4 消融实验评估

为验证Madulous中各改进模块的有效性,在相同的数据集和实验环境下进行消融实验。实验以YOLOv8n作为基准网络,采用控制变量法逐步添加各个改进模块,通过对比评估指标的变化来分析各模块对模型整体性能的贡献,实验结果如表3所示。

基线模型YOLOv8n的mAP@0.5为89%,准确率和召回率分别为83.6%和83.1%,模型的参数量和浮点计算量分别为3M和8.2G。引入DPFE框架,mAP@0.5提升至90.6%,准确率提升了2.3%,说明DPFE利用辅助分支有效克服深度网络在特征提取时易产生的信息瓶颈与梯度弱化问题,使得模型获得更为丰富的底层特征信息。单独添加EMA模块,mAP@0.5达到89.6%,与SENet等依赖单尺度特征交互的常规注意力机制不同,EMA采用双分支跨尺度协同与动态融合策略,有效引导模型聚焦重要目标区域,增强了对局部细节的感知能力。实验D在颈部网络引入Swin Transformer模块,mAP@0.5

表3　消融实验结果

Tab. 3　Results of ablation experiments

Number	Model	mAP@0.5/%	*P*/%	*R*/%	GFLOPs	Params
A	YOLOv8n	89	83.6	83.1	8.2	3
B	YOLOv8n+DPFE	90.6	85.9	84.9	18.3	5.7
C	YOLOv8n+EMA	89.6	85.4	83.5	8.3	3.1
D	YOLOv8n+Swin Transformer	89.4	84.6	82.8	9	3.3
E	YOLOv8n+DPFE+EMA	91.1	88	85.3	18.4	5.8
F	YOLOv8n+DPFE+Swin Transformer	90.8	86.6	83.9	19.1	6
G	YOLOv8n+DPFE+EMA+Swin Transformer	91.2	87.8	83.6	19.4	6.1
H	Madulous	91.5	88.9	85.5	19.8	6.2

达到89.4%,表明其层级式移位窗口机制比常规下采样操作更能兼顾长程空间依赖与局部细节,提升了网络对道路场景中多尺度目标的适应性。实验E将DPFE与EMA结合,使得模型在主干网络获得丰富特征信息的基础上,进一步通过注意力机制过滤背景噪声,mAP@0.5跃升至91.1%,准确率达到88%,证明并行特征提取与跨空间维度注意力交互的联合效果。实验F则融合了DPFE与Swin Transformer,mAP@0.5提升至90.8%。实验G同时集成上述三个模块,通过多层次信息融合与多尺度特征交互的互补作用,进一步将mAP@0.5推升至91.2%。最终的Madulous算法继续进行损失函数的对比与重构,以SLID替换原有的分类损失函数,通过设置均值阈值,能够有效引导模型关注遮挡、小尺寸等困难样本,提升整体检测性能。最终改进模型的mAP@0.5达到91.5%,准确率达到88.9%,尽管模型整体的浮点计算量增加至19.8G、参数量增加至6.2M,但在复杂多变的交通场景下换取了目标检测精度、小目标识别率与整体模型鲁棒性的显著提升。

3.5　对比实验评估

为评估Madulous模型的实际检测效果,本节在统一的数据划分与软硬件平台下开展对比实验。参与比对模型包含SSD[28]、Faster R-CNN[29]、RT-DETR[30]、RT-DETRv2-R18[31]、DECO[32]以及YOLO系列算法如YOLOv8n[33]、YOLOv9t[34]、YOLOv10n[35]。

从表4的实验结果可以得出,SSD算法的检测精度最差,该算法在41.1M高参数量的情况下,mAP@0.5仅有74.3%。Faster R-CNN凭借二阶段架构取得了84.7%的平均精度,却受限于二阶段设计而导致资源占用极度冗余。RT-DETR模型的mAP@0.5提升至86.5%,超越前述一、二阶段模型,但其105.2G的计算量和29.2M的参数量,使其在实际工程应用中面临极高的算力门槛。RT-DETRv2-R18作为RT-DETR的改进版本,采用ResNet18轻量级骨干优化编码器-解码器结构,在参数量20.2M、计算量60G的条件

表4　对比实验

Tab. 4　Contrast experiment

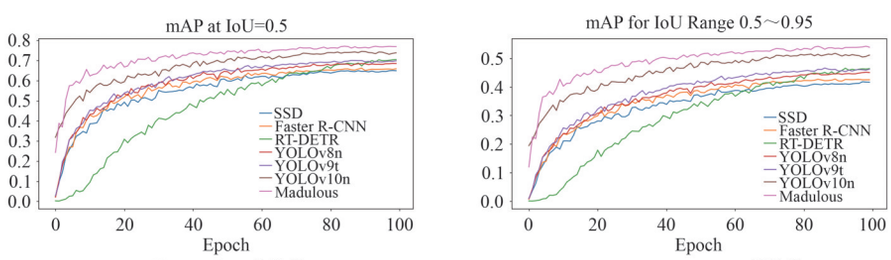
Method	mAP@0.5/%	*P*/%	*R*%	GFLOPs	Params
SSD	74.3	80.9	78.5	145.3	41.1
Faster R-CNN	84.7	81.6	80.6	167.3	72
RT-DETR	86.5	86.2	81	105.2	29.2
YOLOv8n	89	83.6	83.1	8.2	3
YOLOv9t	90.3	86.7	85.7	12.1	2.8
YOLOv10n	90.5	86.9	82.7	8.4	2.8
RT-DETRv2-R18	89.7	85.4	82.3	60	20.2
DECO	90.8	87.8	84.6	32	11.2
本文算法	91.5	88.9	85.5	19.8	6.2

下，mAP@0.5达到89.7%，相较于RT-DETR算法，在降低计算开销的同时实现了精度提升。DECO为2025年提出的卷积式Query-based检测架构，在参数量11.2M、计算量32.0G的情况下，mAP@0.5达到90.8%。在YOLO系列中，YOLOv8算法开发程度较高，本文选择的对比算法为YOLOv8n算法，该算法参数量为3M，浮点数为8.2G，mAP@0.5为89%，YOLOv9算法选择结构较为精简的YOLOv9t模型，mAP@0.5达到90.3%。对于2024年提出的YOLOv10算法，本文选择的是YOLOv10n版本，mAP@0.5的值为90.5%。本文算法在参数量为6.2M，浮点数为19.8G的前提下，准确率和召回率分别为88.9%和85.5%，识别精度达到91.5%，检测性能优于其他算法。与SSD和Faster R-CNN相比，本文算法计算复杂度更小，能够更为高效地完成检测任务，与YOLO系列算法相比，本文算法的辅助分支能够提升模型的特征提取能力，在增加少量计算量的同时提升模型的检测性能。

此外，本文选择数据集为交通路口的行人和车辆数据集，含有多尺度目标信息，且目标信息会因密集而产生相互遮挡等问题，引入EMA模块能够对双分支网络提取到的大量特征信息进行筛选，根据信息的重要程度给予不同的关注度，结合Swin Transformer模块，以提高对于密集重叠目标的检测效果，排除干扰信息，改进分类损失函数，为困难样本赋予更高的权重，保证模型的整体泛化能力。

3.6 泛化性实验评估

考虑到实际道路环境的复杂多变，单一场景难以全面检验算法的鲁棒性。因此，选择RTTS雾天开源交通数据集与InfiRay旗下的红外航拍数据集进行泛化性实验评估。评估的可视化对比如图5，图6所示，图表通过横坐标对应的映射训练周期与纵坐标代表的检测精度，直观呈现了算法的实际表现。

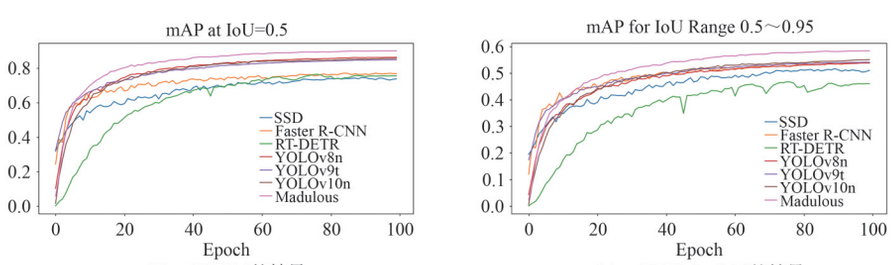


(a) mAP@0.5的结果
(a) Results for mAP@0.5

(b) mAP@0.5～0.95的结果
(b) Results for mAP@0.5～0.95

图5 RTTS数据检测结果

Fig. 5 Detection results of RTTS data



(a) mAP@0.5的结果
(a) Results for mAP@0.5

(b) mAP@0.5～0.95的结果
(b) Results for mAP@0.5～0.95

图6 红外航拍人车数据检测结果

Fig. 6 Detection results of infrared aerial data of person and vehicle

实验结果表明，在具有动态环境噪声干扰的RTTS雾天数据集中，各主流目标检测模型的性能均受到不同程度的影响。如图5所示的对比分析可知，虽然YOLOv10n在该场景下表现出相对较高的检测精度曲线，但仍低于Madulous模型精度曲线，说明Madulous模型通过多重注意力机制，有效提取全局上下文信息与局部细节特征的协同表征，在复杂多目标场景下保持更强的鲁棒

性。针对InfiRay公司开源的红外航拍人车检测数据集，其图像特征主要表现为密集分布的小尺度目标，图6实验数据显示，YOLO系列模型在此类数据上的检测精度趋于稳定，其mAP曲线高于SSD和Faster R-CNN，Madulous模型采用双分支并行网络架构，实现多层次特征提取与多尺度信息融合。综合以上两组验证结果可知，Madulous算法在应对雾天低能见度以及红外微小目标这两类挑战时，均维持了稳定且高效的识别水准，表明该模型在处理复杂多变的实际交通目标检测任务时，具备较强的环境适应性与泛化能力。

3.7　可视化实验评估

本节通过可视化实验直观对比不同算法的检测表现，如图7至图10所示，图中内容分别对

(a)原始图像
(a) Original image
(b) YOLOv10n
(b) YOLOv10n
(c) Madulous
(c) Madulous

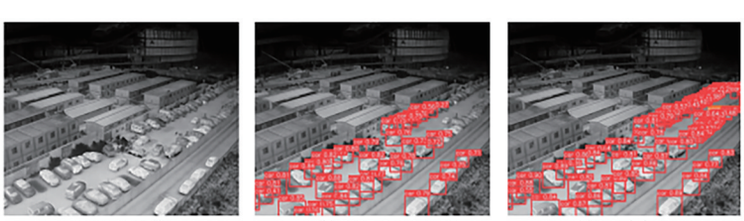
图7　非机动车及行人检测效果对比

Fig. 7　Performance comparison of non-motorized vehicle and pedestrian detection

(a)原始图像
(a) Original image
(b) YOLOv10n
(b) YOLOv10n
(c) Madulous
(c) Madulous

图8　雾天交通检测效果对比

Fig. 8　Performance comparison under foggy traffic conditions



(a)原始图像
(a) Original image
(b) YOLOv10n
(b) YOLOv10n
(c) Madulous
(c) Madulous

图9　无人机航拍检测效果对比

Fig. 9　Performance comparison using UAV aerial imagery

(a)原始图像
(a) Original image
(b) YOLOv10n
(b) YOLOv10n
(c) Madulous
(c) Madulous

图10　极端困难场景检测效果对比

Fig. 10　Performance comparison in extremely difficult scenes

应原始图像、YOLOv10n模型的检测结果，以及Madulous模型的检测结果。

从图7中可以看出，该非机动车与行人检测场景包含尺度差异明显的目标，其中远处部分行人在图像中的像素占比极小，且受到复杂的街道背景干扰，YOLOv10n在处理此类微小尺寸对象时存在明显的感知盲区，导致了2处远处行人目标的直接遗漏，且对非机动车的检测置信度偏低。为此，Madulous架构部署EMA注意力机制，加强对局部区域的权重分配，提升网络筛选和识别小目标信息的能力。从结果可见，Madulous不仅成功召回了远处极小的行人目标，有效消除了漏检，而且对非机动车目标的检测置信度提升至0.83以上，边界框的定位也更加紧密。

在图8的雾天场景中，能见度受限导致图像整体对比度降低，目标纹理信息变弱、边缘轮廓模糊，极易与背景产生特征混淆。在这种退化条件下，YOLOv10n遗漏了1处被遮挡的摩托车和行人目标，导致了预测框的异常重叠。相比之下，Madulous通过重构分类损失函数，自适应增强了模型对于困难样本的惩罚权重，有效剥离了雾天噪声信息与目标真实特征。算法准确识别并区分了所有重叠与被遮挡的目标，不仅检测到先前网络的漏检目标，且整体检测置信度较YOLOv10n得到进一步的提升。

图9为无人机航拍视角的道路检测结果，该场景的车辆目标高密度聚集，单张图像包含大量目标，且多处目标存在车身相互遮挡。YOLOv10n在处理这种重叠的车流时，大量漏检了远处遮挡车辆目标。而Madulous利用双分支并行特征提取框架，充分提取图像的多元化信息特征，并引入Swin Transformer通过层级式下采样融合全局上下文语义，面对拥挤的航拍车流，展现了对于密集目标的分割能力，漏检车辆数大幅降低，且在严重遮挡区域依然维持了极高的召回率。

此外，针对难以准确识别的极端困难场景，图10展示了伴随强光源散射的浓雾环境检测对比。在这种极端低能见度下，车辆的物理轮廓几乎完全丧失，背景噪声极强。YOLOv10n未能感知到画面右侧远处的车辆，出现了严重的漏检；而Madulous虽然凭借多重注意力机制的局部特征聚焦与全局上下文感知，成功定位到了右侧被浓雾隐蔽的两辆汽车，但其检测置信度偏低，且对于极远处仅保留微弱光晕的潜在目标依然难以完成有效识别。这种因极端物理特征缺失导致的置信度下降，暴露出模型在应对超出常规训练分布的极端退化图像时，其特征辨识能力仍存在一定局限性。综上，Madulous在复杂道路场景中表现出了优于基线模型的综合性能与鲁棒性，而针对极端恶劣环境下目标特征的深度恢复与识别，仍是未来持续优化的重要方向。

3.8 边缘计算平台部署实验评估

为进一步验证Madulous在真实端侧场景下的实用性，本文选取EC-R3588SPC与NVIDIA Jetson Orin NX两款主流边缘平台，对YOLOv8n、YOLOv9t、YOLOv10n及Madulous进行推理性能测试。模型输入尺寸统一为640×640，FPS统计均包含图像预处理、模型推理、NMS及后处理全流程耗时，各模型重复测试100轮，取FPS平均值，模型训练权重均统一转换为对应平台专用格式，统一使用FP16半精度优化，结果如表5所示。

表5 部署实验结果

Tab. 5 Deploy experimental results

Edge computing platforms	Method	FPS	Size
EC-R3588SPC	YOLOv8n	25.5	26
	YOLOv9t	24.6	28.3
	YOLOv10n	25.1	27.1
	Madulous	23.9	29.4
NVIDIA Jetson Orin NX	YOLOv8n	43.6	26
	YOLOv9t	42.5	28.3
	YOLOv10n	43.1	27.1
	Madulous	41.8	29.4

通过RKNN工具链将模型权重转换为硬件适配格式后，在EC-R3588SPC平台上的测试结果表明基准模型YOLOv8n的推理速度达25.5 FPS，而Madulous模型的实时帧率为23.9 FPS。尽管模型体积增加13.1%，但帧率仅下降6.3%。在算力更强的Jetson Orin NX平台中，Madulous展现出更显著的性能优势，推理速度提升至41.8 FPS，较同平台YOLOv8n仅降低4.1%，显著优于EC-R3588SPC平台的相对性能差距，说明硬件算力对模型部署效率同样有所影响，即高端平台可更高效补偿算法复杂度的增加。

上述实验证明，算法的实测帧率能够适配城

市平交路口的交通流特征。在EC-R3588SPC平台23.9 FPS的推理速度下，单帧处理间隔约为41.8毫秒；在Jetson Orin NX平台41.8 FPS的推理速度下，单帧处理间隔进一步缩短至23.9毫秒。而城市平交路口场景中，机动车最高行驶速度通常不超过60 km/h（约16.7 m/s），非机动车最高行驶速度不超过25 km/h（约6.9 m/s），在此速度范围内，目标在两帧之间的最大位移均远小于单个交通目标的物理尺寸，不会出现目标在两帧之间移出视野或因运动轨迹断裂而导致的漏检和误检问题，能够保证对运动目标的连续稳定检测。

综上可以得出，Madulous在资源受限的嵌入式环境中实现了精度与效率的平衡，其部署适应性满足智能交通系统对实时目标检测的技术要求。

4　结　　论

道路目标识别与检测是自动驾驶技术、城市智能交通发展的关键之一，针对现有算法在复杂交通场景下存在的检测精度不足与鲁棒性欠佳等情况，提出基于多重注意力机制融合的双分支交通目标检测方法Madulous，构建双分支检测网络，实现目标空间信息与语义特征的提取，将EMA注意力模块内嵌于Backbone中，引导网络自适应聚焦局部特征，增强算法对微小边缘及图像细节的感知能力；在Neck部分添加Swin Transformer模块，利用层级式网络机制建立空间长程依赖关系，增强网络对多尺度目标的特征表征能力；重构分类损失函数，强调边界处的样本，在平衡难易样本权重的同时，改善算法对困难样本的定位与分类性能。

实验结果表明，改进后的模型在交通数据集上mAP@0.5达到91.5%，准确率和召回率分别达到88.9%和85.5%；在EC-R3588SPC和NVIDIA Jetson Orin NX边缘计算平台上推理速度分别达到23.9 FPS和41.8 FPS，同时在雾天、红外航拍等复杂交通场景中，展现出优于主流检测算法的环境适应性与鲁棒性。这充分证明本文提出的改进方法有效可行，能够在不同交通路口准确识别各类道路目标，为道路安全保障提供了可靠的技术支撑。

为满足更高标准的自动驾驶端侧落地需求，后续将重点开展模型轻量化与量化压缩研究，进一步挖掘算法潜力，在保持检测精度的同时持续提升推理效率，并拓展算法在夜间、雨雪等极端天气场景下的适应能力。

参　考　文　献：

[1] 吴一全，童康. 基于深度学习的无人机航拍图像小目标检测研究进展[J]. 航空学报，2025，46(3)：030848.
WU Y Q, TONG K. Research advances on deep learning-based small object detection in UAV aerial images [J]. *Acta Aeronautica et Astronautica Sinica*, 2025, 46(3): 030848. (in Chinese)

[2] 金黎威，徐望明，李垚翔. 基于自适应切片辅助推理的航拍图像目标检测方法[J]. 液晶与显示，2025，40(3)：472-480.
JIN L W, XU W M, LI Y X. Object detection method for aerial images based on adaptive slicing aided inference [J]. *Chinese Journal of Liquid Crystals and Displays*, 2025, 40(3): 472-480. (in Chinese)

[3] 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(6): 1137-1149.

[4] REDMON J, DIVVALA S, GIRSHICK R, *et al*. You only look once: unified, real-time object detection [C]// *Proceedings of the* 2016 *IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas: IEEE, 2016: 779-788.

[5] LIU W, ANGUELOV D, ERHAN D, *et al*. SSD: single shot multibox detector [J/OL]. *arXiv*, 2015: 1512.02325.

[6] QI S H, SONG X F, SHANG T F, *et al*. MSFE-YOLO: an improved YOLOv8 network for object detection on drone view [J]. *IEEE Geoscience and Remote Sensing Letters*, 2024, 21: 6013605.

[7] SUN T, YANG S Q, LIU H Y, *et al*. MIS-YOLOv8: an improved algorithm for detecting small objects in UAV aerial photography based on YOLOv8 [J]. *IEEE Transactions on Instrumentation and Measurement*, 2025, 74: 5020212.

[8] 雷帮军,余翱,吴正平,等. 改进YOLOv8n的无人机航拍小目标检测算法[J]. 现代电子技术,2025,48(3):26-34.
LEI B J, YU A, WU Z P, *et al*. Improved small object detection algorithm based on YOLOv8n for UAV aerial photography [J]. *Modern Electronics Technique*, 2025, 48(3): 26-34. (in Chinese)

[9] 孔垂乐,孟昱煜,火久元,等. 改进YOLOv11的无人机海上小目标检测算法[J]. 计算机工程与应用,2026,62(1):151-161.
KONG C L, MENG Y Y, HUO J Y, *et al*. Improved UAV maritime small target detection algorithm for YOLOv11 [J]. *Computer Engineering and Applications*, 2026, 62(1): 151-161. (in Chinese)

[10] 罗可心,李松江,王鹏,等. 面向无人机影像小目标检测的轻量化算法[J]. 液晶与显示,2026,41(2):253-266.
LUO K X, LI S J, WANG P, *et al*. Lightweight algorithm for small object detection in UAV images [J]. *Chinese Journal of Liquid Crystals and Displays*, 2026, 41(2): 253-266. (in Chinese)

[11] 彭晏飞,赵涛,陈炎康,等. 基于上下文信息与特征细化的无人机小目标检测算法[J]. 计算机工程与应用,2024,60(5):183-190.
PENG Y F, ZHAO T, CHEN Y K, *et al*. UAV small object detection algorithm based on context information and feature refinement [J]. *Computer Engineering and Applications*, 2024, 60(5): 183-190. (in Chinese)

[12] 李云红,张富星,苏雪平,等. 增强上下文特征交互的实时无人机影像分割算法[J]. 北京航空航天大学学报,2026,52(3):668-677.
LI Y H, ZHANG F X, SU X P, *et al*. Real-time UAV image segmentation algorithm with enhanced contextual feature interaction [J]. *Journal of Beijing University of Aeronautics and Astronautics*, 2026, 52(3): 668-677. (in Chinese)

[13] WANG H Y, YU Y T, TANG Z X. FDM-RTDETR: a multi-scale small target detection algorithm [J]. *IEEE Access*, 2025, 13: 88747-88761.

[14] 陈崇杨,彭力,杨杰龙. 基于特征增强与上下文融合的无人机小目标检测算法[J]. 计算机科学,2025,52(11):131-140.
CHEN C Y, PENG L, YANG J L. UAV small object detection algorithm based on feature enhancement and context fusion [J]. *Computer Science*, 2025, 52(11): 131-140. (in Chinese)

[15] 张志豪,厉小润,陈淑涵. 基于改进YOLO11的无人机航拍图像小目标检测算法[J]. 液晶与显示,2025,40(6):915-930.
ZHANG Z H, LI X R, CHEN S H. Small object detection algorithm in UAV aerial images based on improved YOLO11 [J]. *Chinese Journal of Liquid Crystals and Displays*, 2025, 40(6): 915-930. (in Chinese)

[16] KHANAM R, HUSSAIN M. YOLOv11: an overview of the key architectural enhancements [J/OL]. *arXiv*, 2024: 2410.17725.

[17] 吕学涵,李富,祁铭瑞,等. 基于改进YOLOv11s的无人机小目标检测算法[J]. 液晶与显示,2025,40(11):1744-1756.
LV X H, LI F, QI M R, *et al*. Target detection algorithm based on improved YOLOv11s UAV aerial image [J]. *Chinese Journal of Liquid Crystals and Displays*, 2025, 40(11): 1744-1756. (in Chinese)

[18] 张志豪,杜丽霞,侯越,等. 跨层注意力交互下的多特征交叉无人机图像检测[J]. 光学 精密工程,2024,32(24):3616-3631.
ZHANG Z H, DU L H, HOU Y, *et al*. Multi-feature cross UAV image detection algorithm under cross-layer attentional interaction [J]. *Optics and Precision Engineering*, 2024, 32(24): 3616-3631. (in Chinese)

[19] CAI X H, LAI Q X, WANG Y W, *et al*. Poly kernel inception network for remote sensing detection[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle: IEEE, 2024: 27706-27716.

[20] ZHANG H, ZHANG S J. Focaler-IoU: more focused intersection over union loss [J/OL]. *arXiv*, 2024: 2401.10525.

[21] DU D W, ZHU P F, WEN L Y, *et al*. VisDrone-DET2019: the vision meets drone object detection in image challenge results [C]//2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW). Seoul: IEEE, 2019: 213-226.

[22] 朱嘉浩,黄俊,刘兆金. 基于YOLOv11的小目标检测模型[J]. 激光杂志,2025,46(11):49-56.
ZHU J H, HUANG J, LIU Z J. Small object detection model based on YOLOV11 [J]. *Laser Journal*, 2025, 46(11): 49-56.

[23] 景婷婷,曹玉东,陈鑫,等. 改进YOLOv11的无人机航拍小目标检测算法[J]. 计算机工程与应用,2026,62(2):138-148.(in Chinese)
JING T T, CAO Y D, CHEN X, *et al*. Improved YOLOv11 algorithm for aerial small target detection in UAVs [J]. *Computer Engineering and Applications*, 2026, 62(2): 138-148. (in Chinese)

[24] 梁秀满,张永胜,吴楠,等. 基于RSC-YOLO的无人机目标检测算法[J]. 红外技术,2025,47(12):1491-1501.
LIANG X M, ZHANG Y S, WU N, *et al*. UAV object detection algorithm based on RSC-YOLO [J]. *Infrared Technology*, 2025, 47(12): 1491-1501. (in Chinese)

[25] BAOLONG N, ZHANG C Y, SHI Y Z, *et al*. DeBiFormer: vision transformer with deformable agent bi-level routing attention [M]//NUGENT R, MOBASHERI D A, NAKAHARA JR A J, *et al*. *Lecture Notes in Computer Science*. Singapore: Springer, 2024: 445-462.

[26] HU S, GAO F, ZHOU X W, *et al*. Hybrid convolutional and attention network for hyperspectral image denoising [J]. *IEEE Geoscience and Remote Sensing Letters*, 2024, 21: 5504005.

[27] WU T Y, TANG S, ZHANG R, *et al*. CGNet: a light-weight context guided network for semantic segmentation [J]. *IEEE Transactions on Image Processing*, 2021, 30: 1169-1179.

[28] GAO P, LU J S, LI H S, *et al*. Container: context aggregation network [J/OL]. *arXiv*, 2021: 2106.01401.

[29] GUO M H, LU C Z, HOU Q B, *et al*. SegNeXt: rethinking convolutional attention design for semantic segmentation [C]//*Proceedings of the* 36*th Conference on Neural Information Processing Systems*. New Orleans: Curran Associates, Inc., 2022: 1140-1156.

[30] MOSTOFA M, FERDOUS S N, RIGGAN B S, *et al*. Joint-SRVDNet: joint super resolution and vehicle detection network [J]. *IEEE Access*, 2020, 8: 82306-82319.

[31] XIA G S, BAI X, DING J, *et al*. DOTA: a large-scale dataset for object detection in aerial images [C]//*Proceedings of* 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City: IEEE, 2018: 3974-3983.

作者简介:

朱　硕,女,博士,副教授,2014年于中国科学院大学获得博士学位,主要研究方向为计算机视觉,智能感知等。E-mail:zshuo2011@163.com