光电技术及应用

DOI：10.16818/j.issn1001－5868.2024070202

基于 TSD-YOLO 的交通标志检测

井方科[1]，任红格[2*]，李　松[1]，史　涛[3]

（1. 华北理工大学 电气工程学院，河北 唐山 063210；2. 天津城建大学 控制与机械工程学院，天津 300384；
3. 天津理工大学 电气工程与自动化学院，天津 300384）

摘　要： 针对现有交通标志检测算法对小目标检测效果较差的问题，提出了一种基于级联多尺度特征融合的交通标志检测算法。首先，设计了一种新的级联多尺度特征融合网络，利用多尺度序列特征融合模块和三重特征编码模块，使模型能更好地融合交通标志的全局特征和细节特征。其次，在骨干网络中加入可变形注意力机制，使模型专注于相关区域并捕获更丰富的图像特征。最后，使用 Inner-IoU 损失函数，提升了模型的泛化性能。在 CCTSDB 数据集上的测试结果表明，改进模型的平均精度为 55.3%，较 YOLOv8s 模型提升了 3.2%；在 TT100K 和 VOC 数据集上的表现凸显了模型出色的泛化性能。

关键词： 深度学习；交通标志检测；小目标检测；多尺度特征融合；YOLOv8

中图分类号： TP391.4　**文章编号：** 1001－5868(2024)06－1014－10

Traffic Sign Detection Based on TSD-YOLO

JING Fangke[1], REN Hongge[2], LI Song[1], SHI Tao[3]

(1. School of Electrical Engineering, North China University of Science and Technology, Tangshan 063210, CHN; 2. School of Control and Mechanical Engineering, Tianjin Chengjian University, Tianjin 300384, CHN; 3. School of Electrical Engineering and Automation, Tianjin University of Technology, Tianjin 300384, CHN)

Abstract: A multiscale feature fusion-based traffic sign detection (TSD) algorithm is proposed to address the problem of poor performance of existing object detection algorithms in identifying small target traffic signs. First, a novel cascaded multiscale feature fusion network was designed, which fully utilizes the multiscale sequence feature fusion structure and triple feature encoding module, enabling the network to better integrate the detailed and global features of traffic signs. Second, deformable attention mechanisms were incorporated into the backbone network to enable the model to focus on relevant regions and capture richer image features. Finally, the use of the inner intersection over union (IoU) loss function improved the generalization performance of the TSD-YOLO model. The experimental results on the CSUST Chinese traffic sign detection benchmark (CCTSDB) dataset show that the average accuracy of the improved model was 55.3%, which is 3.2% higher than that of YOLOv8s. In addition, the performance on the visual object classes (VOC) dataset and benchmark Tsinghua-Tencent 100K (TT100K) dataset highlights the excellent generalization performance of the model.

Key words: deep learning; traffic sign detection; small object detection; multiscale feature fusion; YOLOv8

收稿日期：2024－07－02.

基金项目：国家自然科学基金项目(62103298).

* **通信作者：**任红格

0　引言

交通标志检测是自动驾驶系统中的一项基础性工作，可以传达有关道路状况和潜在危险，在减少交通事故和交通拥堵方面起着至关重要的作用。传统交通标志检测算法依赖于对交通标志的醒目颜色[1]和特定形状[2]的判定，其计算过程十分复杂，难以在实际应用中有效实施。随着卷积神经网络的快速发展，基于深度学习的目标检测算法被应用到交通标志检测中。目前常用的有单阶段和两阶段两类，单阶段目标检测算法有 YOLO 系列[3-8]和 SSD[9-11]等，两阶段目标检测算法有 RCNN 系列[12-16]。

在实际应用中，为保证驾驶员的人身安全，应尽可能早地识别出交通标志。但若此时交通标志距车太远或目标太小且特征比较模糊，则难以进行有效识别。因此，检测小目标或特征不明显的交通标志很有研究价值[17]。

Zhao 等[18]提出了一种基于亚像素卷积和特征融合的交通标志检测模型，通过构建多特征交互融合网络促进了各特征层之间的信息交互，提高了特征融合效果，强化了小目标的感知能力，但模型的实时性有待提升。Zhang 等[19]提出一种双层半复合骨干网络结构并使用协同注意力机制，减少了目标的漏检，但不能证明其改进模型在不同检测环境中的泛化性能。Zhang 等[20]提出了一种基于 YOLOv5s 的交通标志检测算法，利用区域特征增强模块来扩大感受野和改变特征维度，添加小目标检测层并与主干网络提取的局部小目标信息进行特征融合，增加了小目标的检测精度。然而，增加小目标检测层难免增加了模型的参数量。Li 等[21]对颈部网络进行改进，通过跨层链接和提高特征图分辨率更好地利用浅层特征图中包含的丰富信息来定位目标，并在 CIoU 和 EIoU 损失函数的基础上对损失函数的宽高纵横比进行优化，有效提升了模型识别的准确率，但引入了大量的参数，不便于后期的部署。Yang 等[22]在 YOLOv5 模型中增加跨层连接，重构 Concat 连接并使用 EIoU 损失函数替换 CIoU 损失函数，解决了对交通标志的错检漏检问题并提高了检测精度。但 EIoU 很难根据不同的检测器和检测任务进行自我调整，泛化能力较弱。

受上述研究的启发，本文以 YOLOv8s 为基准模型，提出了一种检测精度更高、参数量更少的 TSD-YOLO 算法。主要工作如下：

(1)在 YOLOv8 中使用了一种新的级联多尺度特征融合网络，利用联级多尺度序列特征融合结构，再结合三重特征编码模块，使模型能更好地聚焦小目标特征并进行多尺度特征融合。

(2)在骨干网络末尾加入可变形注意力模块，使模型专注于相关区域并捕获更丰富的目标特征。

(3)使用 Inner-IoU 损失函数，通过引入缩放因子比来控制辅助边界框的尺寸大小，再利用辅助边界框来计算损失，改善了原损失函数在不同检测任务中的泛化能力弱和收敛速度慢的问题。

1　YOLOv8 基本结构

YOLOv8 基本结构如图 1 所示。YOLOv8 的骨干部分与 YOLOv5 基本相同，基于跨阶段部分连接方法，将 C3 模块替换为 C2f 模块。C2f 模块借鉴了 YOLOv7 中的 ELAN 思路，将 C3 和 ELAN 相结合，形成了 C2f 模块，保证 YOLOv8 模型轻量化的同时可以获得更全面的梯度流信息。在骨干网络的末端，仍然采用 SPPF 模块，将 3 个最大值池化串行，然后将每一层级联，在保证轻量化的同时，确保了模型在不同尺度下的检测准确性。在颈部部分，YOLOv8 使用的路径聚合网络-特征金字塔网络，促进了不同尺度下特征层信息的融合。最后，YOLOv8 使用 YOLOx 的解耦头，结合了置信度和回归框，提升了模型检测的准确性。此外，YOLOv8 具有很强的灵活性，可以支持 YOLO 的大多数版本。因此，选择 YOLOv8 作为基准模型。

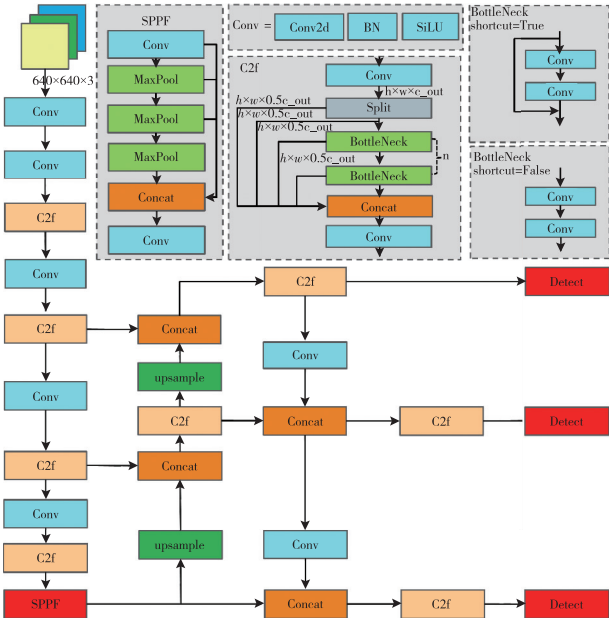


图 1　YOLOv8 结构图

2 改进模型

2.1 整体结构

TSD-YOLO 整体结构如图 2 所示。

首先，TSD-YOLO 算法使用了一种新的级联多尺度序列特征融合网络（Cascaded Multi-Scale Sequence Feature Fusion，CMSF），为小目标检测提供了互补信息。它由两个主要组件网络组成：尺度序列特征融合模块（Scale Sequence Feature Fusion，SSFF），融合了多尺度图像的全局信息或高级语义信息；三重特征编码模块（Triple Feature Encoding，TFE），可以捕获小尺寸目标更丰富的细节特征。

其次，在骨干网络末尾加入可变形注意力模块（Deformable Attention，DA），使模型更专注于感兴趣区域并捕获更多的信息特征。

最后，使用 Inner-IoU 损失函数，提升了模型的泛化性能和收敛速度。

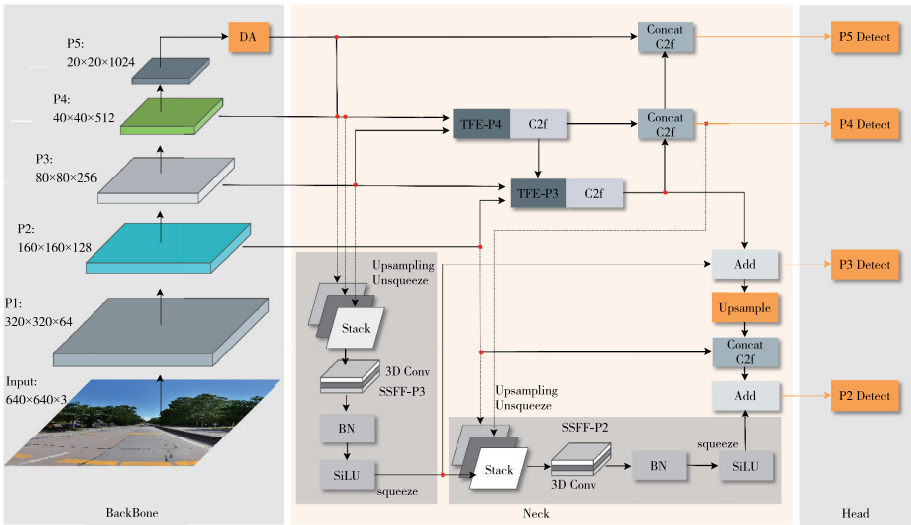


图 2 TSD-YOLO 结构图

2.2 级联多尺度序列特征融合网络

级联多尺度序列特征融合网络由 SSFF 和 TFE 模块组成[23]。

首先，高分辨率特征图级别 P2 和 P3 中包含大部分对于小目标检测至关重要的信息。因此 SSFF 模块是基于 P2 和 P3 层级设计的，分别为 SSFF-P2 和 SSFF-P3。在 YOLOv8s 的颈部网络中，使用 SSFF-P3 模块有效融合从骨干网络中提取的 P3，P4 和 P5 层特征图，这些特征图提取了交通标志各种大小和形状的不同空间尺度特征。在 SSFF-P3 模块中，P3，P4 和 P5 特征图被归一化为相同的大小，再上采样，然后堆叠在一起作为 3D 卷积的输入，以融合多尺度特征。

其次，使用 TFE-P4 模块，通过在空间维度上融合 P3，P4，P5 层三种不同尺寸的特征。将 TFE-P4 模块的输出与 P2，P3 层特征图一同传入 TFE-P3 模块进行融合，从而获得目标的细节信息。通过 PAN 结构将 TFE 模块的信息整合到各个特征支路中。再将 TFE-P3 模块的输出与 SSFF-P3 模块的多尺度特征结合到 P3 支路中，输入到 P3 的头部中。将 SSFF-P3 模块的输出作为 SSFF-P2 模块的一个输入，从骨干网络中提取的 P2 层的输出特征和 TFE-P4 模块的输出特征作为 SSFF-P2 模块的另外两个输入。使用 SSFF-P3 模块，对包含更多特征信息的特征图进行更复杂的多尺度特征融合。

最后，为了避免 P1 层特征图过大而影响检测效率，P2 层不再使用 TFE 模块拼接 P1，P2，P3 层的特征。将 TFE-P3 模块的输出与 SSFF-P3 模块的输出拼接，再上采样，最终与 P2 层特征图融合。将融合后的结果再与 SSFF-P2 模块的多尺度特征结合到 P2 支路中，输入到 P2 的头部中。将 TFE-P4 模块的输出特征与 TFE-P3 模块输出特征进行融合。融合后的结果再与 P5 特征图进行融合，分别输入到 P4，P5 检测头中。

2.2.1 尺度序列特征融合模块

特征金字塔结构仅使用求和或拼接来融合金字塔特征。然而，各种特征金字塔网络的结构不能有效地利用所有金字塔特征图之间的相关性。基于此，TSD-YOLO 提出了一种新颖的尺度序列特征融合方法，可以更好地将深层特征图的高维信息与浅

层特征图的详细信息相结合，其中在图像下采样过程中图像的大小发生变化，但尺度不变特征保持不变。尺度空间是沿着图像的尺度轴构建的，它不仅代表一个尺度，而且还代表一个对象可以具有的各种尺度的范围[24]。模糊的图像可能会丢失细节，但可以保留其图像的结构特征，如式(1)与式(2)所示：

$$F_\sigma(w,h)=G_\sigma(w,h)\times f(w,h) \tag{1}$$

$$G_\sigma(w,h)=\frac{1}{2\pi\sigma^2}e^{-(w^2+h^2)/2\sigma^2} \tag{2}$$

式中，$f(w,h)$表示高度为 h、宽度为 w 的二位输入图像，$F_\sigma(w,h)$是由一系列卷积使用二维高斯滤波器 $G_\sigma(w,h)$进行平滑处理后生成的，σ 是用于卷积的二维高斯滤波器的标准差缩放参数。

这些生成的图像具有相同的分辨率和不同的尺度。因此，不同大小的特征图可以看作是尺度空间，不同分辨率的有效特征图可以调整到相同的分辨率进行拼接。将不同尺度的特征图水平堆叠，并使用三维卷积[25]提取它们的尺度序列特征。如图 2 所示，SSFF 结构由以下组件组成：

(1)使用 1×1 卷积将 P4 和 P5 特征级别的通道数改变为 256，并且将 P3 和 P4 特征级别的通道数改变为 128。

(2)使用最近邻插值方法[26]将其大小调整为 P2 或 P3 级别的大小。

(3)使用 unsqueeze 方法增加每个特征层的维度，将其从三维张量[高度，宽度，通道]更改为四维张量[深度，高度，宽度，通道]。

(4)沿深度维度拼接四维特征图，以形成用于后续卷积的三维特征图。

(5)使用三维卷积、三维批量归一化和 SiLU[27]激活函数完成尺度序列特征提取。

2.2.2　三重特征编码模块

为了识别小尺寸目标，可以通过放大图像来参考和比较不同尺度的形状或外观变化。由于骨干网络的不同特征层具有不同的大小，常规的 FPN 融合机制仅对小尺寸的特征图进行上采样，然后将其拆分或添加到前一层的特征中，忽略了较大尺寸特征层的详细信息。因此，使用 TFE 模块将大、中、小特征进行拆分，添加大尺寸特征图，并进行特征放大以增强详细特征信息，改善了小目标的检测效果。

TFE 模块的构成如图 3 所示。

在特征编码之前，先调整特征通道的数量，使其与主尺度特征一致。在对大尺寸特征图进行处理后，其通道数调整为 1C，然后使用最大池化和平均池化的混合结构进行下采样，保留了高分辨率特征，进而保证了道路图像的有效性和多样性。对于小尺寸特征图，也使用卷积模块来调整通道数，然后使用最近邻插值方法进行上采样，这有助于保留低分辨率图像中局部特征，并防止小目标特征信息的丢失。再对大、中、小尺寸的特征图分别进行一次卷积。最后，在通道维度上拼接，如式(3)所示：

$$F_{\mathrm{TFE}}=\mathrm{Concat}(F_{\mathrm{l}},F_{\mathrm{m}},F_{\mathrm{s}}) \tag{3}$$

式中，F_{TFE} 表示 TFE 模块输出的特征图。F_{l}，F_{m} 和 F_{s} 分别表示大、中、小尺寸的特征图。F_{TFE} 是由 F_{l}，F_{m} 和 F_{s} 拼接后得到的，具有与 F_{m} 相同的分辨率，并且通道数是 F_{m} 的三倍。

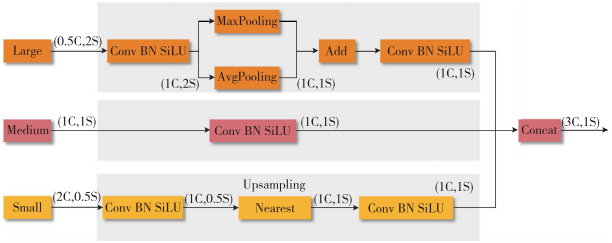


图 3　TFE 结构图

2.3　可变形自注意力模块

通过单一地扩大感受野来提高模型的表示能力会引起一些问题。一方面，使用密集关注会产生大量的计算成本，并且超出兴趣区域的不相关部分可能会影响特征质量。另一方面，使用稀疏关注会在一定程度上限制对长时关系建模的能力。因此，TSD-YOLO 使用了 DA 注意力机制[28]，其中自注意力中的键和值的位置用数据依赖的方式进行选择，这为每个查询共享移动的键和值以实现有效权衡，便于自注意力机制捕获更多的信息特征，使其专注于感兴趣区域。

如图 4 所示，可变形的注意力机制可以在特征映射中感兴趣区域的指导下建立起各位置标志之间的联系。偏移网络从查询中学习到多组可变形采样点，进而确定这些感兴趣区域。通过双线性插值法对特征映射采样，然后将采样后的特征输入投影后得到一个可变的键。最后，应用多头注意力来关注对采样键的查询，并从可变性键中聚合特征。此外，变形点可以通过相对位置偏差促进可变形注意力机制的学习。

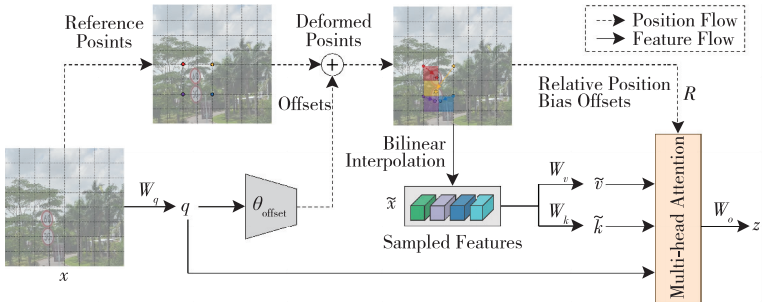


图 4 DA 注意力机制原理图

给定一个 $x \in R^{H\times W\times C}$ 的输入特征图，将满足 $p \in R^{H_G\times W_G\times 2}$ 的统一网格作为参考点。网格大小通过系数 r 在输入的特征图中进行下采样，其中 $H_G=H/r, W_G=W/r$。将成线性间隔关系的二维坐标 $\{(0,0),\cdots,(H_{G-1},W_{G-1})\}$ 作为参考点的值，然后以网格形状 $H_G\times W_G$ 为依据，将参考点归一化到[−1,+1]范围内，将(−1,−1)作为左上角坐标，将(+1,+1)作为右下角坐标。为了计算参考点的补偿，对特征映射进行线性投影，则查询标志 $q=xW_q$，然后将其输入到一个轻量级子网络 θ_{offset} 中以生成偏移量 $\Delta p=\theta_{\text{offset}}(q)$。用预定义的因子 s 对 Δp 的振幅缩放，以防止出现太大的偏移量，即 $\Delta p \leftarrow \text{stanh}(\Delta p)$，稳定了训练过程。然后于变形处对特征采样记入键和值，再对投影矩阵进行采样，如式(4)和式(5)所示：

$$q=xW_q, \tilde{k}=\tilde{x}W_k, \tilde{v}=\tilde{x}W_v \tag{4}$$

$$\Delta p=\theta_{\text{offset}}(q), \tilde{x}=\phi(x;p+\Delta p) \tag{5}$$

式中，$\tilde{k}$ 和 $\tilde{v}$ 分别为可变键的嵌入和值的嵌入。

使用双线性插值设置采样函数 ϕ，使其具有可微性：

$$\phi(z;(p_x,p_y))=\sum_{(r_x,r_y)} g(p_x,r_x)g(p_y,r_y)\cdot z[r_y,r_x,:] \tag{6}$$

式中，$g(a,b)=\max(0,1-|a-b|)$ 和 (r_x,r_y) 都对 $z\in R^{H\times W\times C}$ 进行了索引。因为 g 只在 (p_x,p_y) 附近的四个积分点上可以取到非零值，因此式(6)可以简化成上述四个积分点的加权平均值。对 q,k,v 使用多头注意力，再设置相对位置的偏移量 r。注意力头的最终输出如式(7)所示：

$$z^{(m)}=\sigma(q^{(m)}\tilde{k}^{(m)\mathrm{T}}/\sqrt{d}+\phi(\hat{B};R))\tilde{v}^{(m)} \tag{7}$$

其中，生成网络运行过程中会消耗值的特征和输出的参考点补偿值。因为每个参考点会覆盖一个 $s\times s$ 区域，则生成网络需要感知局部特征以便学习合适的补偿值，其中 s 为补偿的最大值。因此，如图 5 所示，子网络由包含了两个非线性激活函数的卷积模块组成。首先使用 5×5 的深度卷积获得局部特征，然后通过 GELU 激活函数以及 1×1 卷积，最终得到二维的补偿量。通过删除 1×1 卷积中的偏差缓解了强制性偏移。

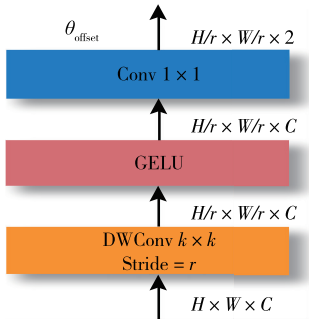


图 5 生成网络原理图

2.4 Inner-IoU 损失函数

随着检测器的快速发展，边界框回归(Bounding Box Regression，BBR)损失函数也在不断更新和优化。然而，现有的基于 IoU 的 BBR 仍然侧重于通过增加新的损失项来加速收敛，而忽略了 IoU 损失项本身的局限性。虽然理论上 IoU 损失可以有效地描述边界盒回归的状态，但在实际应用中，它不能根据不同的检测器和检测任务进行自我调整，泛化能力不强。Zhang 等[29]分析了边框回归过程与模式，基于边框回归问题自身特性，提出了在模型训练过程中使用较小的辅助边框计算损失，对高 IoU 样本的回归有增益效果，而低 IoU 样本则与之相反。

为了改善现有 IoU 损失在不同检测任务中的泛化能力弱和收敛速度慢的问题，使用辅助边界框来计算损失，加速了边界框的回归过程。在 Inner-IoU 损失函数中，引入了比例因子比，以此控制辅助边界框的比例大小。通过对不同的数据集和检测器使用不同尺度的辅助边界框，可以克服现有方法弱泛化的局限性。如图 6 所示，真实框(Ground

Truth,GT)和锚框分别记为 b^{gt} 和 b。(x_c,y_c)表示锚框中心点,(x_c^{gt},y_c^{gt})表示真实框中心点。锚框的宽度和高度分别表示为 w 和 h,真实框的宽度和高度分别表示为 w^{gt} 和 h^{gt}。变量“ratio”对应的是比例因子,其取值范围一般在[0.5,1.5]之间。Inner-IoU 的定义如式(8)～(20)所示:

$$b_l^{gt}=x_c^{gt}-\frac{w^{gt}\times ratio}{2} \tag{8}$$

$$b_r^{gt}=x_c^{gt}+\frac{w^{gt}\times ratio}{2} \tag{9}$$

$$b_t^{gt}=y_c^{gt}-\frac{h^{gt}\times ratio}{2} \tag{10}$$

$$b_b^{gt}=y_c^{gt}+\frac{h^{gt}\times ratio}{2} \tag{11}$$

$$b_l=x_c-\frac{w\times ratio}{2} \tag{12}$$

$$b_r=x_c+\frac{w\times ratio}{2} \tag{13}$$

$$b_t=y_c-\frac{h\times ratio}{2} \tag{14}$$

$$b_b=y_c+\frac{h\times ratio}{2} \tag{15}$$

$$inter1=(\min(b_r^{gt},b_r))-\max(b_l^{gt},b_l) \tag{16}$$

$$inter2=(\min(b_b^{gt},b_b))-\max(b_t^{gt},b_t) \tag{17}$$

$$inter=inter1-inter2 \tag{18}$$

$$union=(w^{gt}h^{gt})\times(ratio)^2+(wh)\times(ratio)^2-inter \tag{19}$$

$$IoU^{inner}=\frac{inter}{union} \tag{20}$$

Inner-IoU 继承了 IoU 损失的一些特性,同时也有自己的特性。与 IoU loss 一样,Inner-IoU loss 的取值范围为[0,1]。由于辅助边界框与实际边界框之间仅存在尺度上的差异,损失函数的计算方法相同。与 IoU 损失相比,当比例因子小于 1 且辅助边界框尺寸小于实际边界框时,其回归的有效范围小于 IoU 损失,但其梯度绝对值大于 IoU 损失所得的梯度,可以加速高 IoU 样本的收敛。相反,当比例因子大于 1 时,较大尺度的辅助边界框使有效回归范围扩大,使低 IoU 样本的回归效果增强。在 IoU 边界框回归损失函数中应用 Inner-IoU。如公式(21)所示:

$$L_{inner\text{-}IoU}=1-IoU^{inner} \tag{21}$$

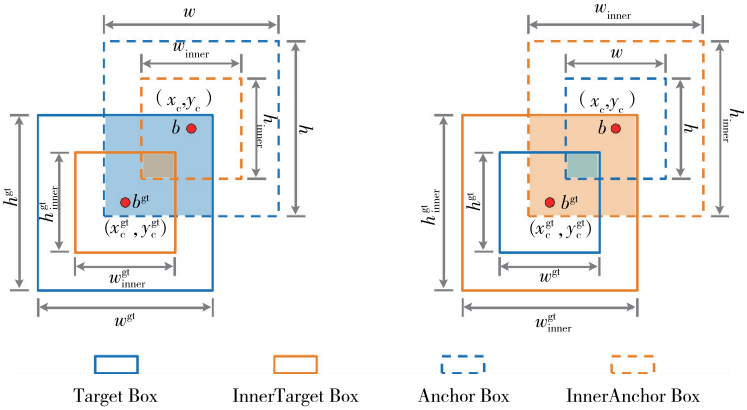


图 6　Inner-IoU 损失函数原理图

3　实验结果与分析

3.1　实验设置和评价指标

本实验使用 NVIDIA Tesla T4(16 GB)GPU,在 Windows10,conda4.10.3,CUDA11.3,python3.8.10 环境下进行。实验中使用了 SGD 优化器和余弦退火策略。将 Batch-size 设置为 16,workers 设置为 8,初始学习率设置为 0.001,共训练 300 epoch。

如式(22)所示,R 为召回率,反映目标的检出情况;如式(23)所示,P 为精度,反映边界框的正确比例;TP 反映检测对象的正确数量;FP 反映检测到不正确对象的数量;FN 反映被遗漏预测的实际框的数量。如式(24)所示,AP 反映一个类别的准确率。如式(25)所示,mAP 反映算法的准确性,对所有类别的 AP 值求平均得到 mAP。$Parms$ 为参数量,反映模型的复杂程度,如式(26)所示。其中,C_{in} 为输入通道的数量,C_{out} 为输出通道的数量,K_h 为卷积核的高度,K_w 为卷积核的宽度。

$$R=\frac{TP}{TP+FN} \tag{22}$$

$$P=\frac{TP}{TP+FP} \tag{23}$$

$$AP=\int_0^1 P(R) \tag{24}$$

$$mAP=\frac{\sum_{i=1}^{C} AP_i}{C} \tag{25}$$

$$Parms=(K_h\times K_w\times C_{in})\times C_{out}+C_{out} \tag{26}$$

3.2 数据集

实验中的CCTSDB数据集中共有17 856幅图像。在数据集中,将交通标志分为三类,分别为强制性、禁止性和警示性。实验中的数据集共保留了7 565张图像,通过比例划分,训练集共包含6 065张图像,验证集共包含1 500张图像。

TT100K原始数据集中有近一半实例为个位数的类别,这导致了数据分布的严重不平衡。因此,仅保留实例数超过100的类别,最终数据集包含45类交通标志,训练集有7 636张图像,验证集有1 908张图像。

实验中的PASCAL VOC数据集是包含了自行车、汽车、人等20个类别的目标检测数据集。数据集中的训练集共包含4 574张图像,验证集共包含1 143张图像。

3.3 对比试验

为了进一步证明TSD-YOLO模型在交通标志检测任务中的优越性,在CCTSD数据集上将TSD-YOLO模型与部分经典检测算法和流行检测算法进行了对比实验,实验结果如表1所示。

(1)经典Faster R-CNN的参数、精度、召回率和*mAP*分别为137.1×10^6,82%,78.8%和40.8%,但训练过程相对复杂,需要额外的区域生成网络,这需要大量的时间,因此其训练速度最慢,FPS为7。对于一阶段目标检测算法之一的SSD,速度更快,但召回率和检测精度最差,存在大量的漏检。出现上述情况主要是因为CCTSDB数据集在环境复杂性、光照条件、目标大小等方面对目标检测模型有很高的要求,而SSD适用于更简单的任务,不适用于这类复杂的任务。

(2)YOLOv5s,YOLOv7和YOLOv8s具有更高的精度,其中YOLOv8s的*mAP*值最高。YOLO系列算法取得良好的检测结果主要是由于使用了多尺度特征融合和跨阶段特征金字塔来提高目标检测性能。

(3)RT-DETR(Real-Time DE Detection Transformer)是在DETR的基础上对DINO检测模型进行了最高精度的改进。通过分析RT-DETR-L和RT-DETR-X模型的实验数据,收敛速度较快但训练耗时过长,实时性最差,不能满足交通标志检测模型的实时性要求。

(4)TSD-YOLO模型的精度和召回率最高,分别为89.4%和76.5%,FPS为97,满足了交通标志检测的实时性要求。与排名第二的YOLOv6s相比,精度提高了4.2%,召回率提高了3%。与排名第三的YOLOv8s相比,精度提高了1.6%,召回率提高了3.4%。此外,改进模型的*Parms*为10.1×10^6,*mAP*值为76.5%,检测效果最好。总体而言,改进模型在交通标志检测中实现了实时性和准确性之间的平衡,并获得了最佳的整体检测效果,适用于交通检测任务。

表1 不同检测算法在CCTSDB数据集上的检测性能对比

模型	$Params/10^6$	P/%	R/%	$mAP@0.5:0.95$/%	*FPS*	Mandatory/%	Prohibitory/%	Warning/%
Faster-RCNN	137.1	82	78.8	40.8	7	39.3	40.5	42.6
SSD	26.3	86.5	27.4	33.4	34	32.6	33.1	34.7
YOLOv3-tiny	12.1	82.6	47.9	36.6	137	34.8	36	39.2
YOLOv5s	9.1	90.9	69.6	51.4	123	48.5	53.7	51.9
YOLOv6s	16.3	85.2	73.5	50.1	112	45.6	53.3	51.3
YOLOv7-tiny	6.0	84.5	65.8	44.1	120	41.3	46.4	44.8
YOLOv7	37.2	80.8	60.9	43.2	74	41.9	45.4	42.3
YOLOv8s	11.1	87.8	73.1	52.1	123	49	55.1	52.3
rt-detr-l	29.3	74.1	67.1	40.2	23	41.4	36.9	42.4
rt-detr-x	67.3	77.9	59	38.7	44	35.1	37.1	43.8
TSD-YOLO	10.1	89.4	76.5	55.3	97	51.4	58.8	55.6

为了证明改进模型的泛化性能,在TT100K和VOC数据集上进行了模型性能验证。结果如表2和表3所示,TSD-YOLO模型相较于基准模型YOLOv8s,在两个数据集中实时性基本保持不变,*P*,*R*和*mAP*值都有所提升,这证明了改进模型是有效的,并且模型具有出色的泛化性能。

表 2　改进和基准模型在 TT100K 数据集上的检测性能对比

模型	*Params*/10^6	*P*/%	*R*/%	*mAP*@0.5∶0.95/%	*FPS*
YOLOv8s	11.1	85.6	76.8	65.9	99.2
TSD-YOLO	10.1	87.2	77.6	67.8	90.5

表 3　改进和基准模型在 VOC 数据集上的检测性能对比

模型	*Params*/10^6	*P*/%	*R*/%	*mAP*@0.5∶0.95/%	*FPS*
YOLOv8s	11.1	83.2	75.4	63.2	99.8
TSD-YOLO	10.1	83.7	76.6	64.3	94.5

3.4　消融试验

对 Inner-IoU 损失函数设置了九组不同的比例因子，通过改变比例因子“ratio”，使模型达到最佳检测效果，如表 4 所示。实验结果表明，ratio 为 1.1 时，该模型获得了最好的检测效果，精度为 88.4%，召回率为 74.7%，*mAP* 值为 53.6%。

表 4　Inner-IoU 中不同的比例因子对检测效果的影响

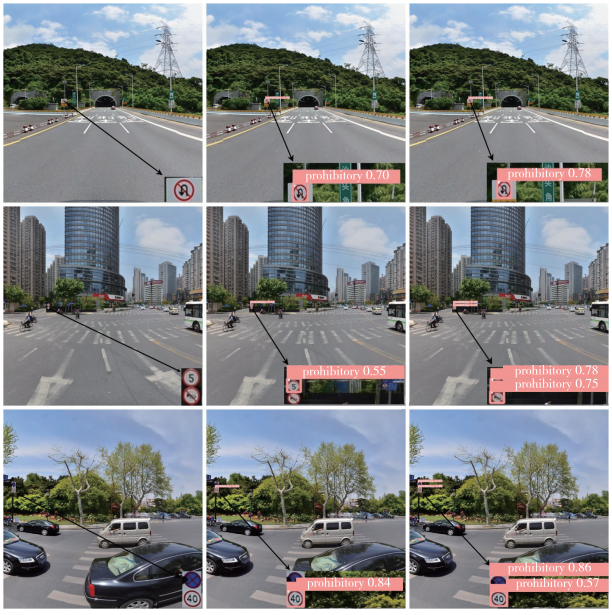
ratio	*Params*/10^6	*P*/%	*R*/%	*mAP*@0.5∶0.95/%	Mandatory/%	Prohibitory/%	Warning/%
0.7	11.1	88.7	73	53.1	50.9	54.5	53.9
0.8	11.1	85.1	73.2	53.2	49.2	55.4	55
0.9	11.1	86.5	73.5	53.5	49.9	55.6	54.9
1	11.1	91.1	72.5	53.5	50.8	55.8	53.8
1.1	11.1	88.4	74.7	53.6	49.6	56.9	54.4
1.2	11.1	87.6	73	53.2	49.9	56	53.8
1.3	11.1	89.8	71.7	53.2	50	55.3	54.2
1.4	11.1	88.8	71.9	53.2	50.3	55	54.3
1.5	11.1	89	72.8	52.4	49.3	54.8	53.2

消融实验如表 5 所示。实验结果表明，每个改进点对提升模型性能都做出了贡献。A 模型为基准模型 YOLOv8s。B 模型在 YOLOv8s 的基础上使用了 CMSF 结构，减少了参数量，同时将模型的 *mAP* 值提高了 2%。结果表明，改进后的 CMSF 结构提高了模型的多尺度特征融合能力，改善了模型的检测性能。C 模型在 YOLOv8s 的基础上使用了 DA 注意力机制，*mAP* 值增加了 0.6%，这使得模型能够专注于相关区域，提取到更多信息特征。D 模型使用 Inner-IoU 损失函数代替了 YOLOv8 模型的 CIoU 损失函数，*mAP* 值提高了 1.3%，提升了模型的泛化性能。E，F 和 G 模型是 CMSF 结构、DA 注意力机制和 Inner-IoU 损失函数的两两组合，检测效果都有所提高。H 模型是最终改进模型 TSD-YOLO，与 YOLOv8s 模型相比，不仅参数量减少了 10%，*mAP* 值还提高了 3.2%。

表 5　消融实验

模型	CMSF	DA	Inner-IoU	*Params*/10^6	*P*/%	R/%	*mAP*@0.5∶0.95/%	*FPS*
A				11.1	87.8	73.1	52.1	123
B	√			9	89.2	74.3	54.1	115
C		√		12.1	88.2	73.7	52.7	112
D			√	11.1	87.6	75.2	53.4	116
E	√	√		10.1	90.4	74.9	54.9	89.3
F	√		√	9	89.5	75.4	54.3	105
G		√	√	12.1	88.4	74.7	53.6	109
H	√	√	√	10.1	89.4	76.5	55.3	97

检测效果如图 7 所示。图 7(a)、(b)和(c)分别为原始图、YOLOv8s 模型和改进模型的检测效果图。三张原始图片中的交通标志皆尺寸极小，检测难度很高。对于第一张检测图，改进模型比 YOLOv8s 模型的 *mAP* 值更高。对于第二张和第三张检测图，改进模型与 YOLOv8s 模型相比，不仅未出现漏检，而且 *mAP* 值更高。总体而言，与 YOLOv8s 相比，改进模型具有更高的召回率和更高的检测精度，以更少的参数和更高的 *mAP* 值实现了更好的检测效果。



(a) 原图　(b) YOLOv8　(c) TSD-YOLO

图 7　检测效果图

4　总结

交通标志的尺寸差异、位置形状和复杂环境严重影响了检测效果。本文提出了一种基于 YOLOv8s 的交通标志检测模型 TSD-YOLO。该算法使用 CMSF 结构，提升了模型的多尺度特征融合能力。然后使用 DA 注意力机制，使模型专注于感兴趣区域并得到更多的信息特征。最后使用 Inner-IoU 损失函数，优化了模型的泛化性能。在 CCTSDB，TT100K 和 VOC 数据集上的实验结果证明了该算法具有较高的检测精度、速度和泛化能力，而且更轻量化，易于部署，具有实际应用价值。然

而，恶劣的天气条件和复杂的环境会严重影响特征识别工作，并对检测任务提出巨大挑战。因此，在接下来的工作中，研究主要集中在雾霾、雨雪等恶劣天气对道路目标检测的影响上，构建能够适应恶劣天气的道路目标检测模型。

参考文献：

[1] Lillo-Castellano J M, Mora-Jiménez I, Figuera-Pozuelo C, et al. Traffic sign segmentation and classification using statistical learning methods[J]. Neurocomputing, 2015, 153: 286-299.

[2] Ruta A, Li Y, Liu X. Real-time traffic sign recognition from video by class-specific discriminative features[J]. Pattern Recognition, 2010, 43(1): 416-430.

[3] Redmon J, Farhadi A. YOLOv3: An incremental improvement[J]. arXiv preprint arXiv: 1804.02767, 2018.

[4] 罗 元，王薄宇，陈 旭. 基于深度学习的目标检测技术的研究综述[J]. 半导体光电，2020，41(1)：1-10.
Luo Yuan, Wang Boyu, Chen Xu. A review of research on deep learning based object detection techniques[J]. Semiconductor Optoelectronics, 2020, 41(1): 1-10.

[5] 谭鑫平，高志辉，韩航迪，等. 基于改进YOLOv5的荧光图像细胞智能检测研究[J]. 半导体光电，2023，44(5)：709-716.
Tan Xinping, Gao Zhihui, Han Hangdi, et al. Research on intelligent detection of fluorescence image cells based on improved YOLOv5[J]. Semiconductor Optoelectronics, 2023, 44(5): 709-716.

[6] 钟运杰，李 康，莫思特，等. 基于GSR-YOLOv7的探地雷达回波图像识别的研究[J]. 半导体光电，2023，44(5)：767-774.
Zhong Yunjie, Li Kang, Mo Site, et al. Research on ground penetrating radar echo image recognition based on GSR-YOLOv7[J]. Semiconductor Optoelectronics, 2023, 44(5): 767-774.

[7] Luo W, Yuan S. Enhanced YOLOv8 for small-object detection in multiscale UAV imagery: Innovations in detection accuracy and efficiency[J]. Digital Signal Processing, 2025, 158: 104964.

[8] Mohammed A, Gang L, Yaxue P, et al. Real-time distracted driving detection based on GM-YOLOv8 on embedded systems[J]. Journal of Transportation Engineering, Part A: Systems, 2025, 151(3): 275184008.

[9] Liu W, Anguelov D, Erhan D, et al. SSD: single shot multibox detector[C]// Computer Vision-ECCV 2016, 2016: 21-37.

[10] Su R, Yue J, Li Z, et al. Detection and counting method of juvenile abalones based on improved SSD network[J]. Information Processing in Agriculture, 2024, 11(3): 325-336.

[11] Wang H, Qian H, Feng S, et al. L-SSD: lightweight SSD target detection based on depth-separable convolution[J]. Journal of Real-Time Image Processing, 2024, 21(2): 01413.

[12] 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.

[13] He K M, Gkioxari G, Dollár P, et al. Mask R-CNN[C]// 2017 IEEE International Conference on Computer Vision, 2017: 2980-2988.

[14] 代少升，刘科生，余自安. 结合视觉Transformer和CNN的道路裂缝检测方法[J]. 半导体光电，2024，45(2)：252-260.
Dai Shaosheng, Liu Kesheng, Yu Zian. A road crack detection method combining visual transformer and CNN[J]. Semiconductor Optoelectronics, 2024, 45(2): 252-260.

[15] 张 毅，龚致远，韦文闻. 基于改进Faster R-CNN模型的交通标志检测[J]. 激光与光电子学进展，2020，57(18)：173-181.
Zhang Yi, Gong Zhiyuan, Wei Wenwen. Traffic sign detection based on improved Faster R-CNN model[J]. Laser & Optoelectronics Progress, 2020, 57(18): 173-181.

[16] He K, Zhang X, Ren S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(90): 1904-1916.

[17] 井方科，任红格，李 松. 基于多尺度特征融合的小目标交通标志检测[J]. 激光与光电子学进展，2024，61(12)：372-380.
Jing Fangke, Ren Hongge, Li Song. Small target traffic sign detection based on multi-scale feature fusion[J]. Progress in Laser and Optoelectronics, 2024, 61(12): 372-380.

[18] Zhao S, Yuan Y, Wu X, et al. YOLOv7-TS: A traffic sign detection model based on sub-pixel convolution and feature fusion[J]. Sensors, 2024, 24(3): 989.

[19] Zhang X, Tian Y. Traffic sign detection algorithm based on improved YOLOv8s[J]. Engineering Letters, 2024, 32(1): 168-178.

[20] Zhang X, Tian Y. Improved YOLOv5s traffic sign detection[J]. Engineering Letters, 2023, 31(4): 1-11.

[21] 李 杰，王 峰，马 晨，等. 基于改进YOLOv5s的无人机图像识别[J]. 电光与控制，2024，31(4)：22-27，91.
Li Jie, Wang Feng, Ma Chen, et al. Drone image recognition based on improved YOLOv5s[J]. Electro Optics and Control, 2024, 31(4): 22-27, 91.

[22] 杨 祥，王华彬，董明刚. 改进YOLOv5的交通标志检测算法[J]. 计算机工程与应用，2023，59(13)：194-204.
Yang Xiang, Wang Huabin, Dong Minggang. Improving YOLOv5's traffic sign detection algorithm[J]. Computer Engineering and Applications, 2023, 59(13): 194-204.

[23] Kang M, Ting M C, Ting F F, et al. ASF-YOLO: A novel YOLO model with attentional scale sequence fusion for cell instance segmentation[J]. Image and Vision Computing,

2024，147：105057.

[24] Lindeberg T. Scale-Space Theory in Computer Vision[M]. New York：Springer，1994.

[25] Tran D，Bourdev1 L，Fergus R，et al. Learning spatiotemporal features with 3D convolutional networks [C]// Proceedings of IEEE International Conference on Computer Vision (ICCV)，2015：4489-4497.

[26] Rukundo O，Cao H. Nearest neighbor value interpolation [J]. International Journal of Advanced Computer Science and Applications，2012，3(4)：25-30.

[27] Elfwing S，Uchibe E，Doya K. Sigmoid-weighted linear units for neural network function approximation in reinforcement learning[J]. Neural Networks，2018，107：3-11.

[28] Dong Y，Zhang M，Qiu L，et al. An arrhythmia classification model based on vision transformer with deformable attention [J]. Micromachines，2023，14 (6)：1155.

[29] Zhang H，Xu C，Zhang S. Inner-IoU：More effective intersection over union loss with auxiliary bounding box[J]. arXiv preprint arXiv：2311.02877，2023.

作者简介：

井方科(1999—)，男，山东省东平县人，硕士生，主要从事智能感知与控制技术研究；

任红格(1979—)，女，河北省石家庄市人，博士，教授，主要从事目标跟踪算法研究。