

基于改进YOLOv8的高分辨率遥感图像目标检测算法

张霞, 乔环宇, 曹峰

山西大学计算机与信息技术学院, 山西 太原 030006

摘要

针对高分辨率遥感图像数据中目标受复杂背景干扰、目标较小且密集分布、目标多尺度且方向随机等问题, 提出了一种基于改进YOLOv8的高分遥感图像目标检测算法。首先, 融入了动态蛇形卷积, 使算法更好地检测尺度、方向不一致的目标; 其次, 为使算法能够捕获复杂背景图像中全局上下文信息, 将iRMB与Shift-Wise卷积相结合, 引入基于注意力的尺度内特征交互模块, 将双向特征金字塔网络和语义细节融合模块相结合作为算法的特征融合网络, 并设计一种改进的高效解耦检测头; 最后, 结合NWD与WIoU损失函数, 设计一个新的损失函数NWD-WIoU, 以加快算法检测时的收敛速度并提升算法对小目标的检测性能。在3个公开的遥感图像目标检测数据集上的实验结果表明, 基于改进YOLOv8的高分辨率遥感图像目标检测算法的性能优于其他主流高分辨率遥感图像目标检测算法。

关键词

深度学习; 高分辨率遥感图像; 目标检测; YOLOv8; 轻量化

中图分类号: TP751

文献标志码: A

doi:10.11959/j.issn.2096-0271.2025065

High resolution remote sensing image object detection algorithm based on improved YOLOv8

ZHANG Xia, QIAO Huanyu, CAO Feng

School of Computer and Information Technology, Shanxi University, Taiyuan 030006, China

Abstract

In view of problems such as objects are interfered by complex background, objects are small and densely distributed, objects are multi-scale and their directions are random in high resolution remote sensing image data, an object detection algorithm for high resolution remote sensing image based on improved YOLOv8 was proposed. Firstly, dynamic snake convolution was incorporated to make the algorithm detect objects with different scales and directions better; Secondly, in order to enable the algorithm to capture the global context information in the image with complex background, iRMB was combined with Shift-Wise convolution. Attention-based Intra-scale feature interaction was used. Bi-directional feature pyramid network and semantics and detail infusion module were combined as the algorithm's feature fusion network. An improved efficient decoupled detection head was designed. Finally, combining NWD and WIoU loss functions, a new loss function NWD-WIoU was designed to accelerate the convergence speed of algorithm detection

and improve the detection performance of the algorithm for small objects. Experimental results on three public remote sensing image object detection datasets show that the high resolution remote sensing image detection algorithm based on improved YOLOv8 performs better than other mainstream high resolution remote sensing image detection algorithms.

Key words

deep learning, high resolution remote sensing image, object detection, YOLOv8, lightweight

0 引言

近年来,随着遥感技术的飞速发展,遥感图像中包含的语义信息更加丰富^[1]。高分辨率遥感图像包含丰富的细节信息,这些信息能够直观反映地物目标的形状、颜色和纹理等特征,便于人眼直接观察^[2]。高分辨率遥感图像目标检测是处理遥感图像的关键技术,该技术可以从遥感图像中识别并提取出更多的地理空间信息,被应用于精准农业、城市规划,环境监测以及军事侦察等领域,在大数据时代取得了重大突破^[3]。

但卫星图像拍摄的高分辨率遥感图像与普通自然场景图像存在明显差异,高分辨率遥感图像目标检测算法仍面临很多挑战^[4]。部分高分辨率遥感图像目标与背景难以区分,甚至背景遮挡部分目标,使检测器难以有效定位目标,存在漏检或误检的问题;部分高分辨率遥感图像中存在小目标较多且密集分布的情况,导致目标检测器在特征提取过程中漏掉部分小目标的特征,出现漏检情况;部分高分辨率遥感图像中目标尺度变化大且排列方向随机,使目标检测器在检测时产生定位误差,导致目标检测算法准确性和鲁棒性较差。此外,尽管目前许多先进的目标检测算法在高分辨率遥感图像目标检测精度上有很大

提升,但其以增加算法参数量或计算量为代价。因此如何设计一个既精度高又能实现轻量化的算法是一个很有价值的研究问题。

目标检测算法分为两类,一类是传统的目标检测算法,另一类是基于深度学习的目标检测算法。目前基于深度学习的目标检测算法根据检测步骤分为两种:两阶段目标检测算法和一阶段目标检测算法^[5]。两阶段目标检测算法首先在输入图像上标注大量候选区域,在新生成的候选框中对目标区域位置及范围进行调整,然后用特征提取网络对候选目标进行特征提取,从而进行目标识别和目标分类,典型的两阶段目标检测算法包括Fast R-CNN^[6]、Faster R-CNN^[7]、Mask R-CNN^[8]、SPP-Net^[9]等。一阶段目标检测算法不需要生成候选框,通过回归分析直接生成目标类别概率并预测目标框的坐标信息,直接对图像中所有方向的物体进行分类和检测,主要包括SSD^[10]、RetinaNet^[11]、YOLO^[12-18]等。

1 YOLOv8

YOLO是当前目标检测领域性能较优的算法之一,YOLO的优势在于检测目标又快又准,且能实现实时的目标检测。YOLOv8包括5种网络结构:YOLOv8n、

YOLOv8s、YOLOv8m、YOLOv8l、YOLOv8x。其网络结构由3个部分组成：Backbone、Neck和Head。

- Backbone部分采用 CSPDarkNet (cross stage partial darknet) 结构, 由3个模块组成: Conv、C2f 和 SPPF 模块。Conv 模块产生输出特征, 加快算法的收敛速度, 保证算法的稳定性; C2f 模块使算法可以在保证轻量化的同时获得更丰富的梯度流信息; SPPF 模块用来增强特征图表示, 进一步提高算法计算速度。

- Neck部分使用路径聚合网络 (path aggregation network, PANet) 和特征金字塔网络 (feature pyramid network, FPN) 通过自顶向下与自下向顶的跨层连接对特征进行多尺度融合, 将 Backbone 部分中 SPPF 模块输出的结果作为 PANet 的输入, 经过上采样、特征融合, 最终将 PANet 的3个输出分支送入 Head 部分进行计算。

- Head部分由3个检测头组成, 每个检测头分别在3个不同的尺度上接收来自 Neck 部分的特征。YOLOv8采用了解耦检测头, 将分类分支和回归分支分离, 两条分支并行, 分别提取类别特征和位置特征, 再各用一层卷积完成分类任务和定位任务。

2 改进YOLOv8

本文对 YOLOv8 的网络结构进行了改进, 改进内容总结如下。

- 本文设计 C2f_IS 模块和 C2f_DSC 模块, 并引入 AIFI (attention-based intrascale feature interaction) 模块, 从而提升算法对多尺度目标细节信息的提取能力, 提高对目标的检测精度, 并减少算法的参数量。

- 本文设计 SDI-BiFPN 结构作为改进 YOLOv8 的特征融合网络。

- 本文设计一个新的解耦头 EDetect 作为改进 YOLOv8 的检测头。

- 本文基于 NWD 和 WIoU 损失函数设计了一个新的损失函数 NWD-WIoU, 将其与 DFL (distribution focal loss) 函数相结合, 作为 YOLOv8 的边界框回归损失函数。

改进的 YOLOv8 网络结构如图 1 所示, 其中, Upsample 表示下采样。输入图像 [3,3,1 024,1 024] 中, 从左往右第一个 3 表示批次大小为 3, 即一次处理 3 张图像; 第二个 3 表示表示通道数, 对应 RGB 的 3 个通道; 第一个 1 024 和第二个 1 024 分别表示图像的高度和宽度均为 1 024 像素。

2.1 C2f_IS 模块

反向残差移动块 (inverted residual mobile block, iRMB) ^[19] 由标准卷积、深度卷积和多头自注意力机制组成, 具有类似 CNN 高效的局部特征建模能力和类似 Transformer 的动态建模能力来学习远距离交互, 同时能够有效地扩大算法的感受野, 提升对下游任务的信息提取能力。iRMB 使用了深度卷积 (depthwise convolution, DWConv) ^[20] 模块, 该模块通过对输入通道进行卷积, 有效地减少卷积层所需的参数, 降低算法的复杂度, 加快算法的计算速度。iRMB 的网络结构如图 2 所示。其中, Q、K、V 分别代表查询 (Query)、键 (Key) 和值 (Value), 它们是自注意力机制中的核心组件, 通过动态权重分配捕捉输入序列中元素间的依赖关系, 并提取重要程度不同的语义信息; Attn Mat 为注意力机制。通过对多分支并行处理和引入注意机制, iRMB 能够充分

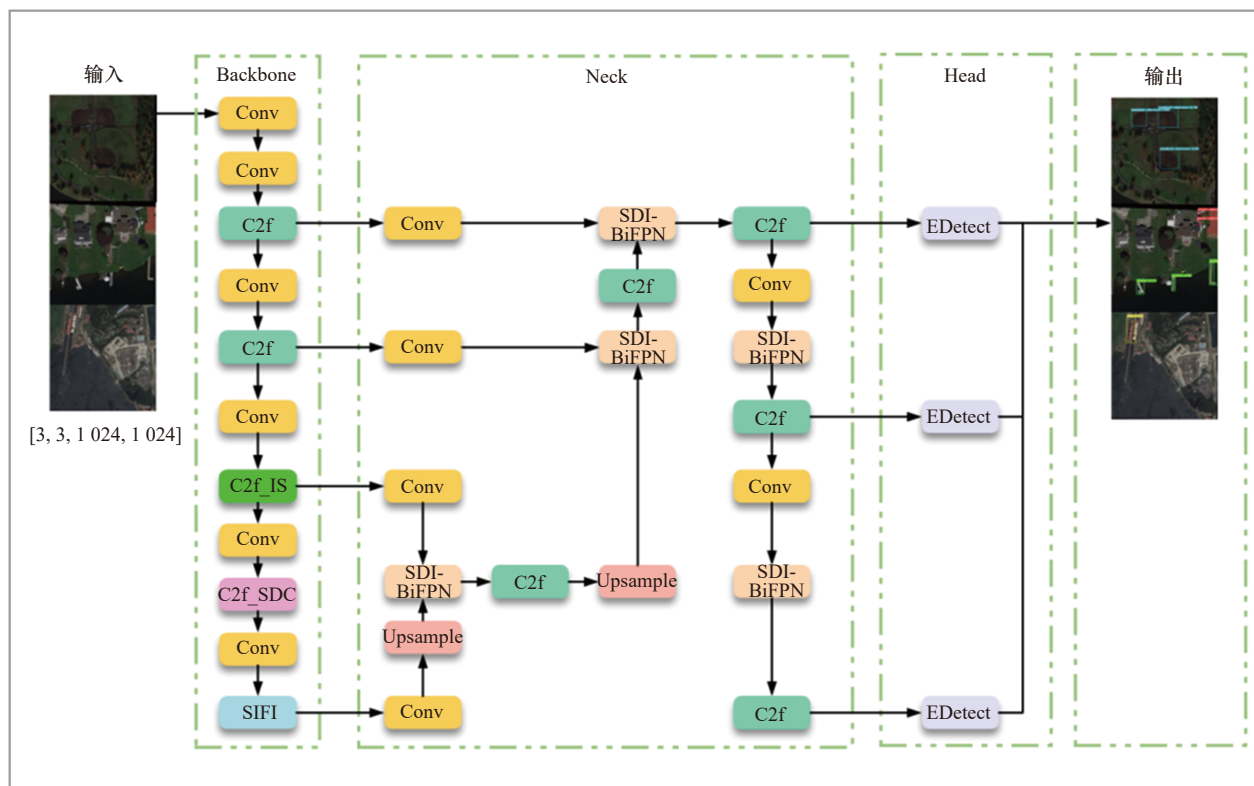


图1 改进的YOLOv8网络结构

利用多尺度、多语义特征信息来增强网络的表达能力。

大卷积核可以让算法具有更大的感受野，从而高效地读取到更多的特征信息。然而，大卷积核存在计算冗余的问题，阻碍算法更快地读取特征语义信息。移位卷积 (shift-wise conv, SWC)^[21]可以该问题，其网络结构如图3所示。SWC将大卷积核转换为一系列较小的卷积核，然后对每个卷积核进行移位操作。图3中的shift表示移位操作，一个蓝色块表示卷积核的一个通道，虚线黄色块表示通道被剪枝，移位操作后，对输出的所有通道进行拼接融合。由图3可知，在训练过程中，SWC剪枝了部分连接，通过剪枝以及后续的融合操作，SWC的总通道输出保持不变。

本文设计了基于iRMB与SWC相结合的iRMB_SWC模块，其网络结构如图4所

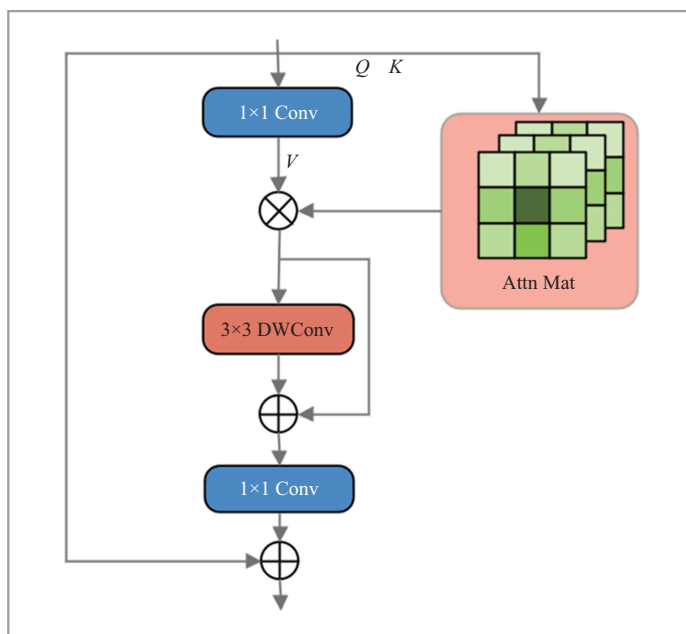


图2 iRMB的网络结构

示。本文将C2f与iRMB_SWC模块结合形成C2f-IS模块，其网络结构如图5所示。

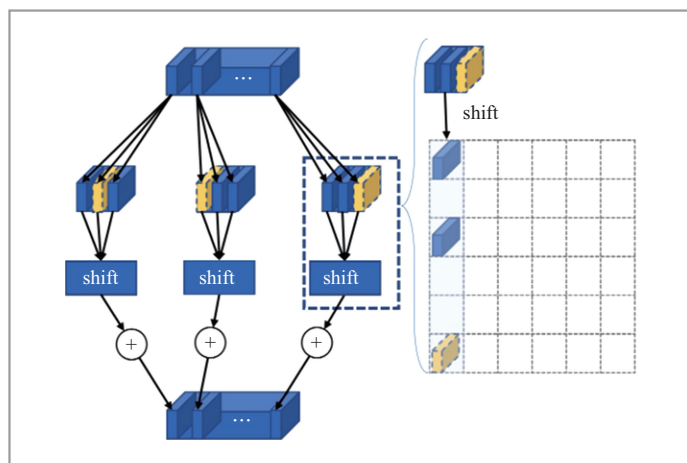


图3 SWC的网络结构

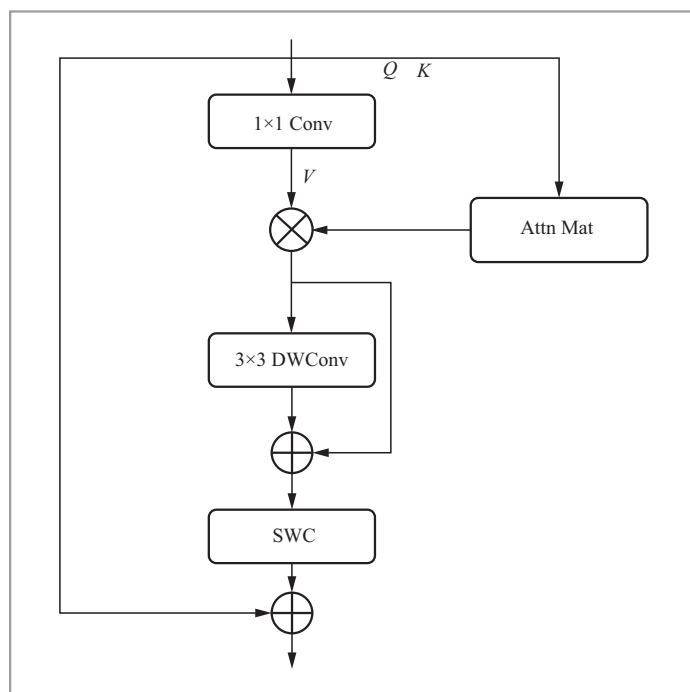


图4 iRMB_SWC模块的网络结构

其中, n 指多个 Bottleneck_IRMB_SWC 相连接, 其值不确定, YOLOv8 通过调整 n 来适应不同的应用场景。选择较小的 n (如 $n=1$ 或 $n=2$ 时), YOLOv8 会牺牲部分精度换取速度; 选择较大的 n (如 $n=3$ 或 $n=4$) 时, YOLOv8 能够充分提升模型的

特征表达能力, 但会增加模型复杂度, 运行速度变慢。C2f-IS 模块使改进的 YOLOv8 算法极大增强了对多尺度图像目标细节信息的提取能力, 并降低了计算复杂度, 减少了算法的参数数量。

2.2 C2f_DSC 模块

在传统的卷积运算中, 感受野保持不变可能使算法捕捉目标的某些细节特征时带来挑战, 特别是具有细长条状结构相关特征的目标。动态蛇形卷积 (dynamic snake convolution, DSConv)^[22] 旨在使算法更灵活地捕获图像中的复杂特征, 其通过算法学习可变形偏置可以应对上述挑战。为了防止感受野过度偏离目标, DSConv 采用了一种迭代策略, 确保每次只处理一个目标, 从而保持注意力的连续性。DSConv 的网络结构如图 6 所示。首先, 输入特征图后在网络中进行偏置学习, 即预测一组偏移量, 这些偏移量表示传统卷积网格的采样位置偏差, 针对图中划分后的每个卷积核单独预测, 以生成任意形状的采样模式; 接着, 根据预测的偏移量, 在输入特征图上进行不规则采样, 利用双线性插值等方法实现精确采样, 保证特征提取的连续性; 然后, 通过注意力机制为每个采样点分配注意力权重, 该权重经 Sigmoid 激活函数进行归一化, 用于抑制背景区域干扰; 最后, 将采样得到的特征与注意力权重进行加权聚合, 得到最终的输出。

本文将 DSConv 与 C2f 相结合, 形成 C2f_DSC 模块并将其嵌入网络中, C2f_DSC 模块的网络结构如图 7 所示。该模块能够增强 YOLOv8 中骨干网络对不同尺度特征图的适应能力, 使 YOLOv8 能够更好地捕捉不同大小和形状目标的特征信息, 从而提高算法的整体性能。

2.3 AIFI 模块

YOLOv8使用的卷积模块主要对目标的局部特征进行检测，导致对具有大尺度特征的目标进行检测时性能欠佳。基于注意力的尺度内特征交互（attention-based intrascale feature interaction, AIFI）模块^[23]将自注意操作应用于语义概念更丰富的高级特征，可以捕获图像中概念与实体之间的联系，从而能够捕获目标的全局特征，以便后续在图像中检测到具有大尺度特征的目标，增强算法的表达能力，AIFI模块的网络结构如图8所示。该模块输入维度为 $[b, c, h, w]$ 的数据。 b 表示网络中输入图像的数量， c 表示特征通道的数量， h 表示特征图的高度， w 表示特征图的宽度。Embedded Patches表示将输入的图像数据转换为适合AIFI模块处理的序列数据；Multi-Head Attention层用来计算目标之间的相似性，输入数据经过Dropout层可以减轻过度拟合；

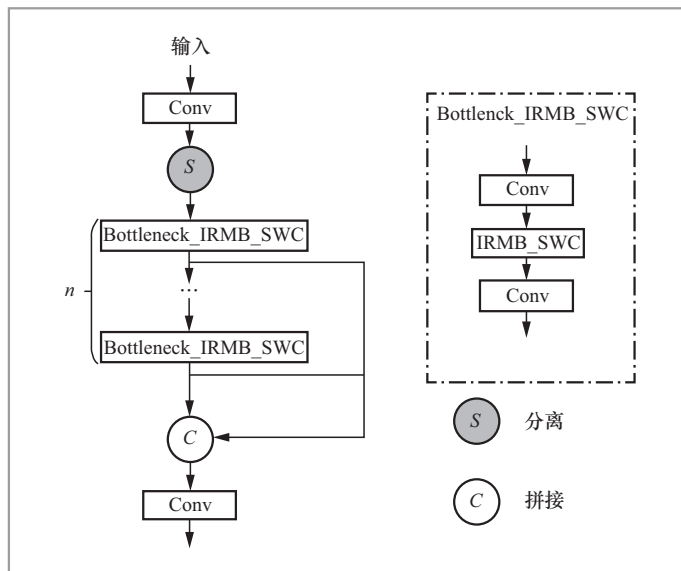


图5 C2f_IS模块的网络结构

LayerNorm层对输入数据进行归一化；Linear层对输入数据进行线性变换；Gelu（）激活函数将输入数据通过高斯误差函数进行变换，有效避免梯度消失的问题；Flatten层将从LayerNorm层输出的 $[b, c,$

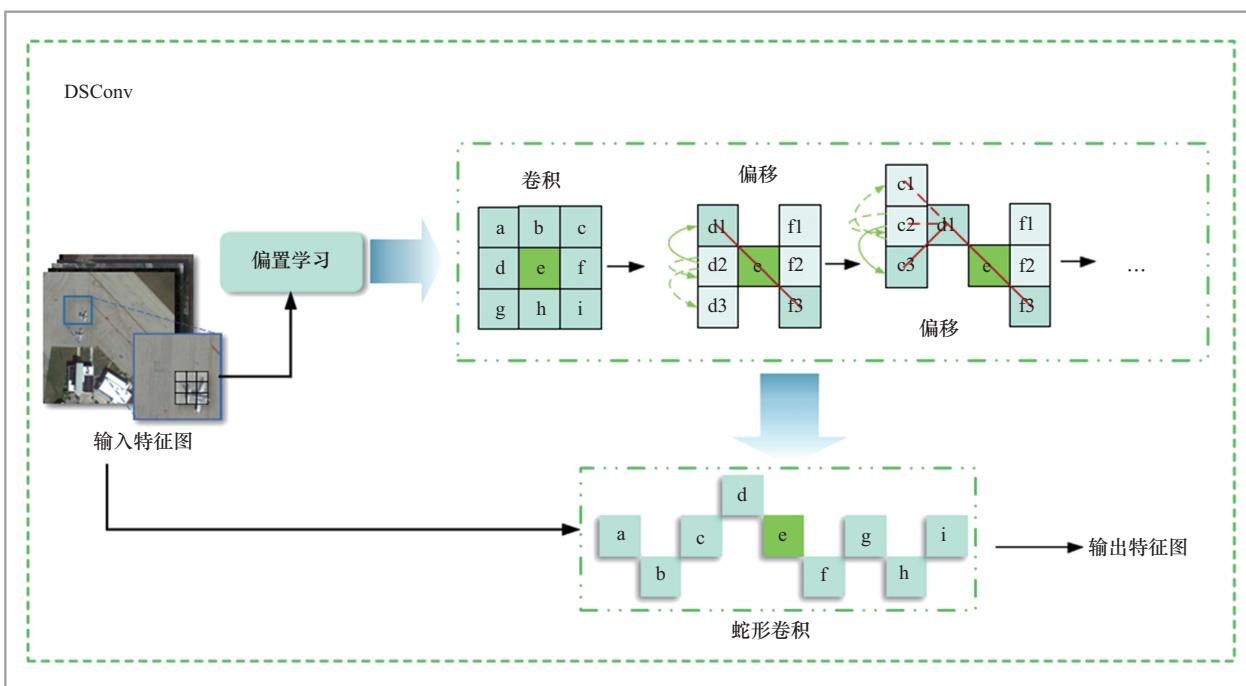


图6 DSConv的网络结构

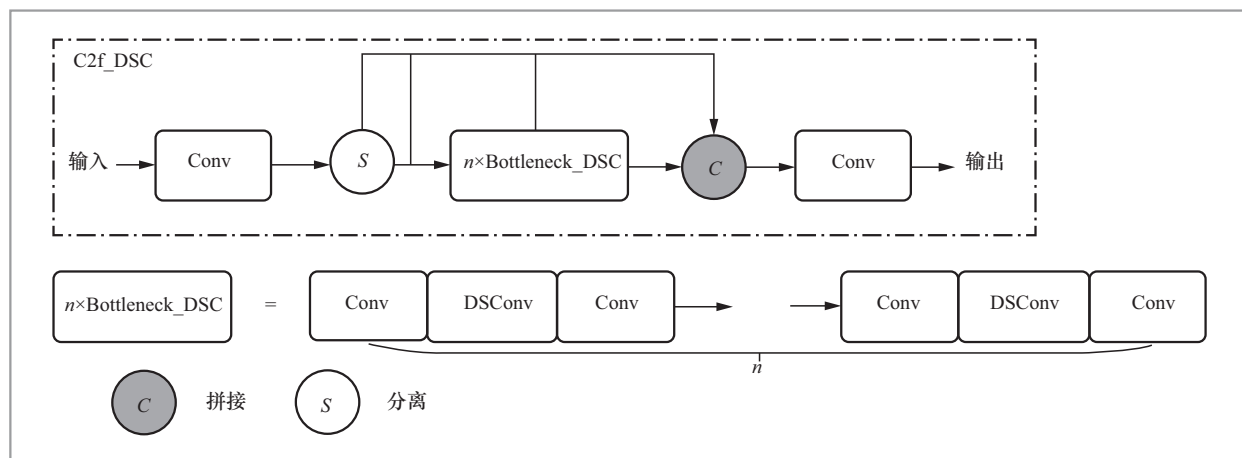


图7 C2f_DSC的网络结构

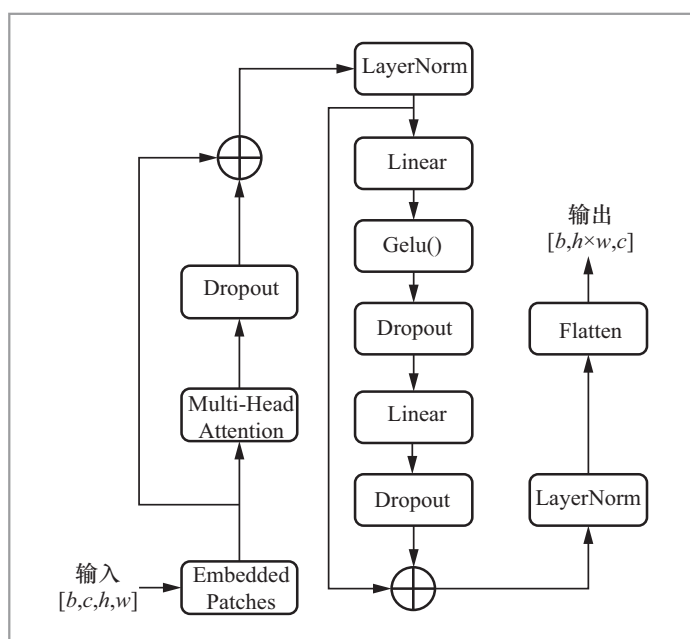


图8 AIFI模块的网络结构

$h, w]$ 转换为 $[b, h \times w, c]$, 从而实现降维。其中从一个 Dropout 层输出的数据与原始数据进行矩阵相加, 从 3 个 Dropout 层输出的数据与从一个 LayerNorm 层输出的数据同样进行矩阵相加。

本文用 AIFI 模块替换 Backbone 部分中的 SPPF 模块。为了能够有效地提取不

同通道之间的信息, 增加算法的表达能力, 本文在 AIFI 模块前加入一个卷积模块。通过引入 AIFI 模块, 改进后的 YOLOv8 在捕获目标全局信息和上下文细节方面得到了增强, 并减少了其参数量, 实现了轻量化。

2.4 SDI-BiFPN 结构

语义和细节融合 (semantics and detail infusion, SDI)^[24] 模块通过整合局部空间信息和全局通道信息, 使检测器可以从图像中更好地获取更精细、具有更多语义信息的细节。双向特征金字塔网络 (bi-directional feature pyramid network, BiFPN)^[25] 通过引入双向流动的特征信息进行多层级的特征融合和动态的特征权重分配。BiFPN 逐层融合上下采样的特征图, 同时引入横向连接和纵向连接, 使不同尺度的目标特征能够更好地融合和被利用。

本文采用 SDI-BiFPN 替代 YOLOv8 的特征融合方法 PANet, 采用 SDI 模块作为特征融合模块, SDI-BiFPN 的网络结构如图 9 所示。其中 P2、P3、P4 和 P5 分别表

示算法主干部分生成的不同尺度目标的特征。SDI模块是特征融合模块，其对前一层输出的不同特征图分别进行自适应池化和双线性插值，再用 3×3 卷积平滑特征图，然后输出。为了减轻算法复杂度，保持检测的准确性，本文采用分组混洗卷积（group shuffle convolution, GSConv）^[26]代替SDI模块中的普通卷积，改进后SDI模块的网络结构如图10所示，其中，Shuffle表示对数据进行重新排列或随机化， C 表示拼接操作。GSConv可以更好地平衡算法的准确性和速度，PWConv、DWConv主要用于减少卷积运算的参数数量。

2.5 EDetect

自校准卷积（self-calibrated convolution, SCConv）^[27]将具有学习能力的卷积核分成4个部分，每个部分卷积核的功能不同。其通过内部信息交互显式地扩展每个卷积层的视野，从而丰富输出特征。SCConv的网络结构如图11所示。其中，Identity表示按照原始特征图进行输入操作，Down表示平均池化操作；Up表示双线性插值运算；K1、K2、K3、K4表示卷积核；X1、X2表示不同的输入分

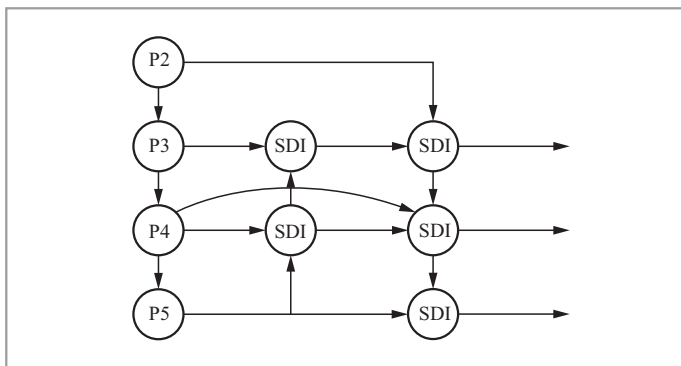


图9 SDI-BiFPN的网络结构

支； C 、 H 、 W 分别表示特征图的输入通道、输入的高度和宽度；“ $r \times$ ”表示对输入数据进行多次操作， r 值不确定，SCConv通过调整 r 来进行不同情况下的数据处理。

尽管YOLOv8的解耦头显著提高了算法的检测能力，但在处理高分辨率遥感图像中具有复杂背景和不同尺度的目标的问题时，其解耦头的特征提取能力表现较差，不能充分利用骨干网络提取高级特征。本文在YOLOv8解耦头的基础上引入SCConv，重新设计了一个新的解耦头EDetect。为了充分利用SCConv增强解耦头的特征提取能力和表达能力，本文将

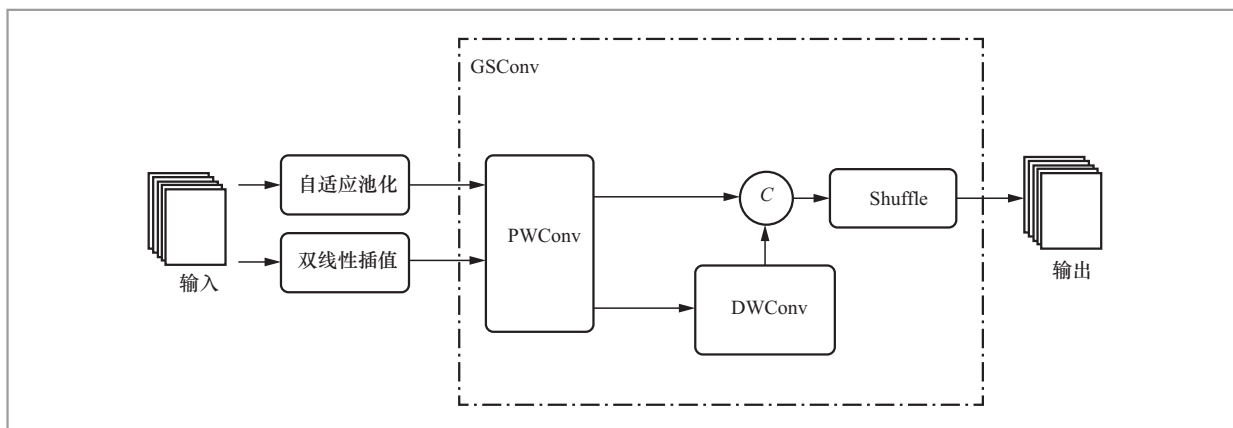


图10 改进后SDI模块的网络结构

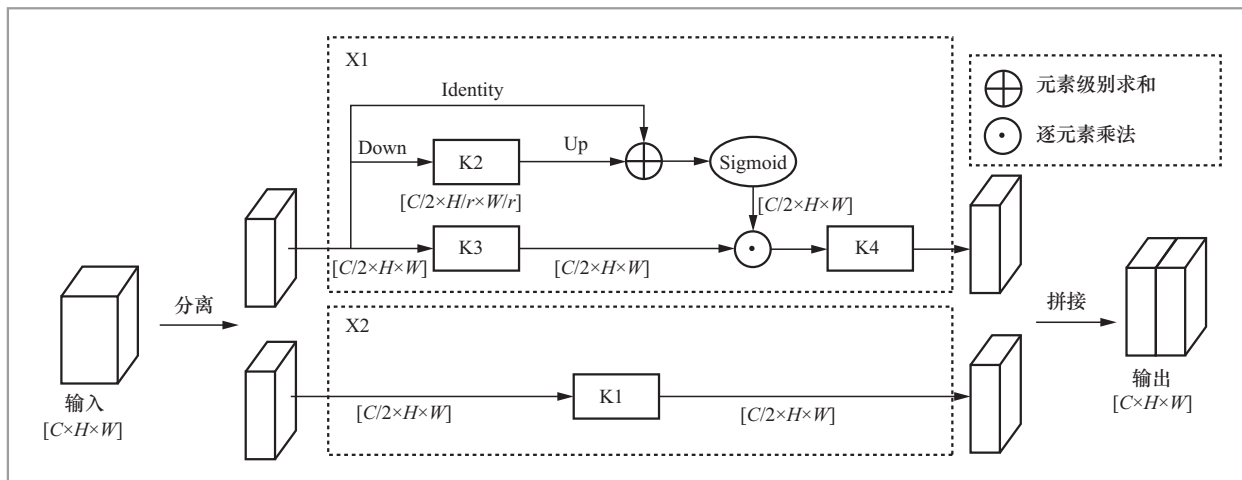


图11 SCConv的网络结构

SCConv 网络结构中的 X1 分支与解耦头相结合，并在 SCConv 后加入一个 1×1 的卷积，EDetect 的网络结构如图 12 所示，其中，Conv2d 表示二维卷积，Bbox.Loss 表示边界框损失，Cls.Loss 表示分类损失。本文通过将 YOLOv8 的解耦头替换为 EDetect，有效解决了 YOLOv8 解耦头提取不同尺度目标特征时效率低的问题。

2.6 改进损失函数

YOLOv8 采用将 CIoU^[28] 和 DFL 损失函数相结合的方法计算边界框回归损失。

CIoU 在 DIoU^[29] 的基础上通过添加纵横比，使其与 DIoU 相比算法的收敛速度更快、检测精度更高。但其加大了算法检测所需的真实框和预测框之间的误差，并且训练数据中包含负样本，增加纵横比会导致正负样本不平衡，从而降低算法的泛化性能。为了解决上述问题，本文引入具有动态非单调聚焦机制的损失函数 WIoU (wise intersection over union)^[30] 对 YOLOv8 进行优化。WIoU 设计了一种合理的动态非单调聚焦机制，该机制减少了个别样本中出现的有害梯度，使算法的收敛速度更快。WIoU 的计算式如下。

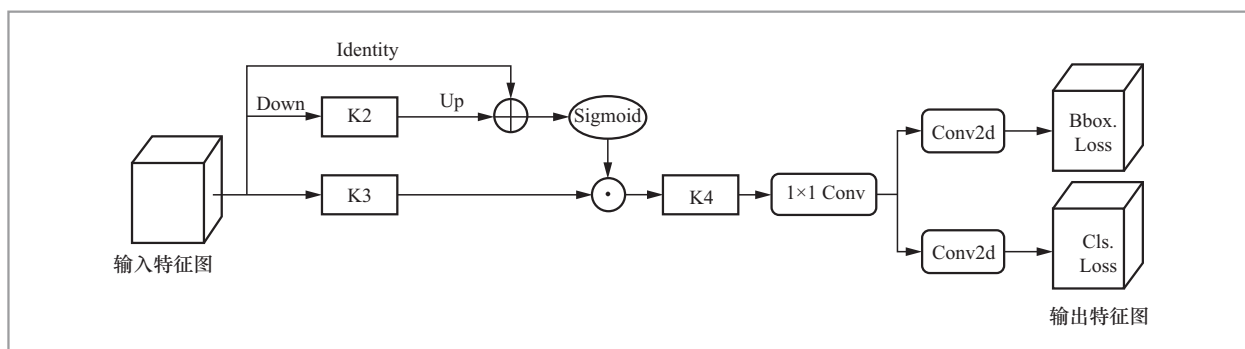


图12 EDetect的网络结构

$$L_{\text{WIoU}} = \gamma R_{\text{WIoU}} L_{\text{IoU}} \quad (1)$$

$$\gamma = \frac{\beta}{\delta \alpha^{\beta-\delta}} \quad (2)$$

$$\beta = \frac{L_{\text{IoU}}}{L_{\text{IoU}}}, \beta \in [0, +\infty] \quad (3)$$

$$R_{\text{WIoU}} = \exp\left(\frac{(x-x_{\text{gt}})^2 - (y-y_{\text{gt}})^2}{(W_g^2 + H_g^2)^*}\right) \quad (4)$$

$$L_{\text{IoU}} = 1 - \text{IoU} = 1 - \frac{W_i H_i}{S_u} \quad (5)$$

$$S_u = w \cdot h + w_{\text{gt}} \cdot h_{\text{gt}} - W_i \cdot H_i \quad (6)$$

其中, L_{WIoU} 、 L_{IoU} 分别表示 WIoU、IoU 的损失函数; γ 表示非单调聚焦系数; R_{WIoU} 表示 WIoU 的边界框回归值; β 表示离群度, 用来描述锚框的质量; α 、 δ 均表示可调节参数, 用来控制离群度大小 (实验中取 $\alpha=1.9$ 、 $\delta=3$), 离群度越小表示锚框质量越高; $\overline{L_{\text{IoU}}}$ 表示动态变量, 其可以让锚框质量的划分标准动态变化, 使 WIoU 在任何时刻采用与当前情况最符合的梯度增益分配策略。式 (4)~式 (6) 中变量的关系如图 13 所示。蓝色区域为预测框; 绿色区域为目标框; x 、 y 为预测框的中心点坐标; x_{gt} 、 y_{gt} 为目标框的中心点坐标; W_g 和 H_g 为预测框与目标框并集所构成的最小矩形框的宽和高; W_i 和 H_i 为预测框与目标框交集所构成矩形框的宽和高; w 、 h 分别为预测框的宽和高; w_{gt} 、 h_{gt} 分别为目标框的宽和高; S_u 为预测框与目标框面积之和与二者交集面积的差值。

在 WIoU 中, 每个物体框的权重值取决于其与目标框的重叠程度。重叠程度越大的物体框权重越高, 重叠程度越小的物体框权重越低。通过该方式, WIoU 能够更好地评估目标检测结果, 并且在存在正

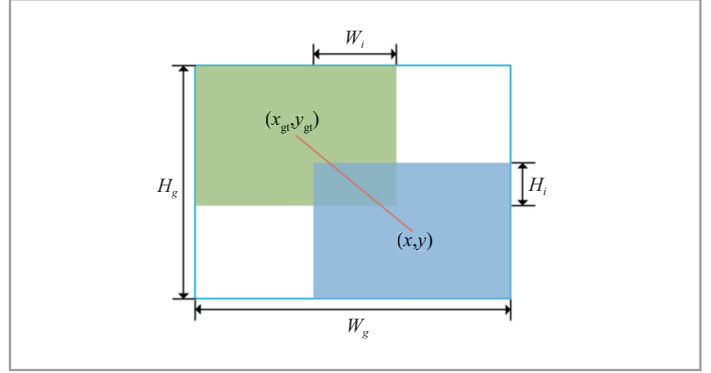


图 13 R_{WIoU} 、 L_{IoU} 中变量的关系

负样本不平衡的情况下也能给出准确的评估。

小目标一般在尺寸上只占据很小的像素, 微小目标特征信息不足, 无法反馈给网络, 导致算法只关注中、大尺寸目标的特征, 完全忽视了对微小目标特征信息的学习。为了解决该问题, 本文引入归一化瓦瑟斯坦距离 (normalized wasserstein distance, NWD) [31], 并将其运用到损失函数中, NWD 通过将预测框与目标框转换成二维高斯分布并计算二者的相似度, 可以更好地度量目标之间的相似性, 提高算法对小目标的检测性能。NWD 的计算式如下。

$$W^2(N, N_{\text{gt}}) = \left\| \left([cx, cy, \frac{w}{2}, \frac{h}{2}]^T, [cx_{\text{gt}}, cy_{\text{gt}}, \frac{w_{\text{gt}}}{2}, \frac{h_{\text{gt}}}{2}]^T \right) \right\|_2^2 \quad (7)$$

$$\text{NWD}(N, N_{\text{gt}}) = \exp\left(-\frac{\sqrt{W^2(N, N_{\text{gt}})}}{C}\right) \quad (8)$$

$$L_{\text{NWD}} = 1 - \text{NWD}(N, N_{\text{gt}}) \quad (9)$$

其中, N 、 N_{gt} 分别表示预测框与目标框的二维高斯分布; $W^2(N, N_{\text{gt}})$ 表示预测框与目标框之间的二阶瓦瑟斯坦距离度量; (cx, cy, w, h) 、 $(cx_{\text{gt}}, cy_{\text{gt}}, w_{\text{gt}}, h_{\text{gt}})$ 分别表示预测

框与目标框的坐标、框的宽度及高信息。为了获得相似度度量,将NWD指数形式归一化并获得新度量 $NWD(N, N_{gt})$,其中, C 表示数据集中根据目标尺寸设置的参数。

如果只简单地用WIoU替换CIoU,算法对小目标的检测效果较差;只简单地将NWD运用到CIoU,算法检测的收敛速度变慢,导致算法检测精度急剧下降。为了更好地发挥两个损失函数的优势,本文将WIoU与NWD相结合,设计了一个新的损失函数NWD-WIoU。

$$L_{NWD-WIoU} = r \cdot L_{WIoU} + (1-r) \cdot L_{NWD} \quad (10)$$

其中, r 表示比例因子,用来控制两类损失函数计算结果所占的权重,通过大量实验得到, $r=0.5$ 时,NWD-WIoU损失函数计算结果以及算法的收敛速度达到最优,所以 r 取0.5。

3 实验与分析

3.1 实验环境

实验的设置操作系统为Ubuntu 20.04, CUDA 版本 12.2; 显卡型号为 NVIDIA GeForce RTX 4090; 显存大小为 24 GB; 编程语言为 Python 3.10; 深度学习框架为 Pytorch 2.1.0。

3.2 数据集

DOTA^[32]是用于高分辨率遥感图像目标检测的公开数据集,包含2 806张高分辨率遥感图像,一共188 282个目标,图像尺寸介于800×800像素到4 000×4 000像素之间,分为15个类别。由于DOTA中图像尺寸较大,对显卡要求高,直接训练

不可行。因此,本文将图像尺寸裁剪为1 024×1 024像素,共得到21 046张图像。然后按8:2:2的比例将该数据集划分为训练集、验证集和测试集,其中验证集和测试集为同一数据集。

RSOD^[33]也是用于高分辨率遥感图像目标检测的公开数据集,包含飞机、油罐、操场和立交桥的高分辨率遥感图像。图像尺寸介于800像素×800像素和1 200像素×1 200像素之间,按与DOTA数据集相同的划分方式划分RSOD数据集。

DIOR^[34]是由西北工业大学提出的一种用于高分辨率遥感图像目标检测的大规模公共数据集,包含23 463幅高分辨率遥感图像和190 288个实例,图像尺寸大小为800×800像素,分为20个类别。本文按6:2:2的比例将该数据集划分为训练集、验证集和测试集。

3.3 实验参数设置

本实验统一使用SGD (stochastic gradient descent) 优化器,将批处理大小设置为8、动量设置为0.937、阈值设置为0.7、最大学习率设置为0.001、最小学习率设置为0.00001、进程数设置为8。为了得到算法在每个数据集损失值达到稳定状态所需要的迭代次数,统一设置迭代次数为400。

3.4 实验参数设置

本实验使用平均精度 (mean average precision, mAP)、每秒帧数 (frame per second, FPS)、每秒浮点运算次数 (floating point operations per second, FLOPS) 和参数量 (param) 4个指标来评估算法的性能。

当数据集中有多个目标类别需要检测

时，采用 mAP 作为目标检测性能的评价指标，其是基于正确率（precision， P ）和召回率（recall， R ）的 P-R 曲线计算的，计算式如式（11）~式（14）所示。FPS 用于评价算法的检测速度，表示每秒可以处理的图片数量。FLOPS 通常用于衡量算法在推断或训练时的计算性能。参数量指模型中所有可训练参数的总数量，包括权重和偏置，其直接关系到模型的大小和对计算资源的需求。

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

(11)

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

(12)

$$AP = \int_0^1 P(R)d(R)$$

(13)

$$mAP = \frac{1}{N} \sum_{n=1}^N AP$$

(14)

其中，TP（true positive）表示实际是正样本的分类被预测为正样本的数目；FP（false positive）表示实际是负样本的分类被预测为正样本的数目；FN（false negative）表示实际是正样本的分类被预测为负样本的数目；TN（true negative）表示实际是负样本的分类被预测为负样本的数目。 N 表示目标类别数，AP 表示某一检测类别的平均精度。

3.5 消融实验

为了验证本文提出的算法中新加入的每一个功能模块的有效性和鲁棒性，本文以 YOLOv8 为本文实验的基准算法，在 DOTA 和 RSOD 数据集上进行消融实验。实验结果见表 1、表 2。

表1 在DOTA数据集上的消融实验

功能模块						评价指标			
NWD-WIoU	AIFI	SDI-BiFPN	EDetect	C2f_DSC	C2f_IS	Param/M	FLOPS	FPS/(f·s ⁻¹)	mAP
						11.13	28.5×10 ⁹	282.4	73.18%
✓						11.13	28.5×10 ⁹	289.1	73.72%
✓	✓					10.58	27.7×10 ⁹	196.8	73.94%
✓	✓	✓				7.18	25.5×10 ⁹	157.1	74.22%
✓	✓	✓	✓			7.35	24.2×10 ⁹	166.5	74.37%
✓	✓	✓	✓	✓		8.36	24.7×10 ⁹	146.3	74.65%
✓	✓	✓	✓	✓	✓	8.22	24.4×10 ⁹	135.1	74.97%

表2 在RSOD数据集上的消融实验

功能模块						评价指标			
NWD-WIoU	AIFI	SDI-BiFPN	EDetect	C2f_DSC	C2f_IS	Param/M	FLOPS	FPS/(f·s ⁻¹)	mAP
						11.13	28.4×10 ⁹	305.2	77.38%
✓						11.13	28.4×10 ⁹	311.8	77.67%
✓	✓					10.58	27.7×10 ⁹	216.2	78.71%
✓	✓	✓				7.18	25.5×10 ⁹	188.5	77.68%
✓	✓	✓	✓			7.35	24.1×10 ⁹	169.8	80.34%
✓	✓	✓	✓	✓		8.36	24.36×10 ⁹	156.7	80.56%
✓	✓	✓	✓	✓	✓	8.21	24.3×10 ⁹	149.3	82.73%

由表1可以知，在DOTA数据集上的消融实验过程中，每新加入一个模块，mAP都有一定的提高。改进后的YOLOv8在DOTA数据集上相比YOLOv8，其FPS

表3 在DOTA数据集上不同目标检测算法性能的对比

算法	Param	FLOPS	FPS/(f·s ⁻¹)	mAP
Faster R-CNN	136.9 × 10 ⁶	804.8 × 10 ⁹	38	67.33%
SSD	25.48 × 10 ⁶	357.6 × 10 ⁹	59.7	64.92%
Retinanet	36.62 × 10 ⁶	431.2 × 10 ⁹	44.9	66.51%
YOLOv3	61.57 × 10 ⁶	154.8 × 10 ⁹	40.8	71.62%
YOLOv5s	9.12 × 10 ⁶	23.9 × 10 ⁹	222.9	72.92%
YOLOv6s	16.3 × 10 ⁶	44 × 10 ⁹	199.3	72.59%
YOLOv7-tiny	6.05 × 10 ⁶	13.1 × 10 ⁹	169.5	71.85%
YOLOv7	36.56 × 10 ⁶	103.4 × 10 ⁹	97.1	74.49%
YOLOv8s	11.13 × 10 ⁶	28.5 × 10 ⁹	277.4	73.18%
RT-DETR	20.1 × 10 ⁶	58.3 × 10 ⁹	62.5	72.15%
YOLOv9	15.59 × 10 ⁶	67.5 × 10 ⁹	104.2	74.38%
改进的YOLOv8	8.22 × 10 ⁶	24.4 × 10 ⁹	135.1	74.97%

表4 在RSOD数据集上不同目标检测算法性能的对比

算法	Param	FLOPS	FPS/(f·s ⁻¹)	mAP
YOLOv5s	9.11 × 10 ⁶	23.8 × 10 ⁹	342.6	73.54%
YOLOv6s	4.23 × 10 ⁶	11.8 × 10 ⁹	427.6	74.98%
YOLOv7-tiny	6.02 × 10 ⁶	13 × 10 ⁹	79.4	76.51%
YOLOv7	36.50 × 10 ⁶	103.2 × 10 ⁹	55.2	80.43%
YOLOv8s	11.13 × 10 ⁶	28.4 × 10 ⁹	305.2	77.38%
RT-DETR	19.88 × 10 ⁶	57 × 10 ⁹	58.2	80.39%
YOLOv9	15.58 × 10 ⁶	67.5 × 10 ⁹	31.7	77.92%
改进的YOLOv8	8.21 × 10 ⁶	24.3 × 10 ⁹	149.3	82.73%

表5 在DIOR数据集上不同目标检测算法性能的对比

算法	Param	FLOPS	FPS/(f·s ⁻¹)	mAP
YOLOv5s	9.2 × 10 ⁶	23.9 × 10 ⁹	351.3	85.37%
YOLOv6s	16.3 × 10 ⁶	44.1 × 10 ⁹	345.5	86.65%
YOLOv7-tiny	6.06 × 10 ⁶	13.2 × 10 ⁹	208.2	80.99%
YOLOv7	36.58 × 10 ⁶	103.5 × 10 ⁹	101.5	87.65%
YOLOv8s	11.13 × 10 ⁶	28.5 × 10 ⁹	302.8	86.76%
RT-DETR	19.89 × 10 ⁶	57 × 10 ⁹	59.8	86.06%
YOLOv9	15.6 × 10 ⁶	67.6 × 10 ⁹	101	87.78%
改进的YOLOv8	8.37 × 10 ⁶	24.7 × 10 ⁹	139.8	88.38%

减少147.3 f/s的情况下，mAP提升了1.79%、Param减少了2.91 × 10⁶、FLOPS减少了4.1 × 10⁹。

由表2可知，在RSOD数据集上的消融实验过程中，除了加入SDI-BiFPN模块会导致mAP下降了1.03%，当接连加入NWD-WioU、AIFI、Edetect、C2f_DSC和C2f_IS后，mAP均有所提升，分别提高了0.29%、1.04%、2.66%、0.22%和2.17%，最终mAP达到了82.73%。改进后的YOLOv8在RSOD数据集中FPS减少155.9 f/s的情况下mAP提升了5.35%，Param与FLOPS分别减少了2.92 × 10⁶和4.1 × 10⁹。

3.6 对比实验

为了进一步验证改进后YOLOv8的性能，在DOTA数据集上将其与目前主流的目标检测算法进行对比，使用Param、FLOPS、mAP、FPS指标评估算法性能，见表3。

由表3可知，本文改进的YOLOv8在DOTA数据集上的mAP大于YOLOv7、RT-DETR和YOLOv9。此外，改进的YOLOv8的Param、FLOPS远小于YOLOv7、RT-DETR和YOLOv9，其FPS也优于后三者。其进一步验证了改进YOLOv8的有效性和可行性。

为了进一步验证改进的YOLOv8在高分辨率遥感图像目标检测方面的鲁棒性，本文分别在RSOD、DIOR数据集上将其与当前主流的目标检测算法进行对比，对比结果分别见表4、表5所示。

由表4和表5可知，改进的YOLOv8的mAP比YOLOv7分别多2.3%和0.73%。且改进YOLOv8的Param仅比YOLOv7-tiny多2.31 × 10⁶，FLOPS仅比YOLOv7-

tiny 多 11.5×10^9 ，比 YOLOv5 多 0.8×10^9 。总体来看，改进的 YOLOv8 的性能优于其他主流目标检测算法，其 FPS 尽管相比 YOLOv5，YOLOv6 和 YOLOv8 相比下降，但仍满足实时目标检测算法的需求。

3.7 可视化对比试验

为了更直观地表示改进的 YOLOv8 性

能优于 YOLOv8，本文从 DOTA 数据集中选取 3 张图片，使用 YOLOv8 与改进的 YOLOv8 两个算法分别进行目标检测，检测效果如图 14 所示。

对比图 14 第一行后两张图片中标识白框的部分可知，改进的 YOLOv8 可以有效解决 YOLOv8 在检测受复杂背景信息干扰目标时的漏检问题；对比图 14 第二行后两张图片中标识白框的部分可以

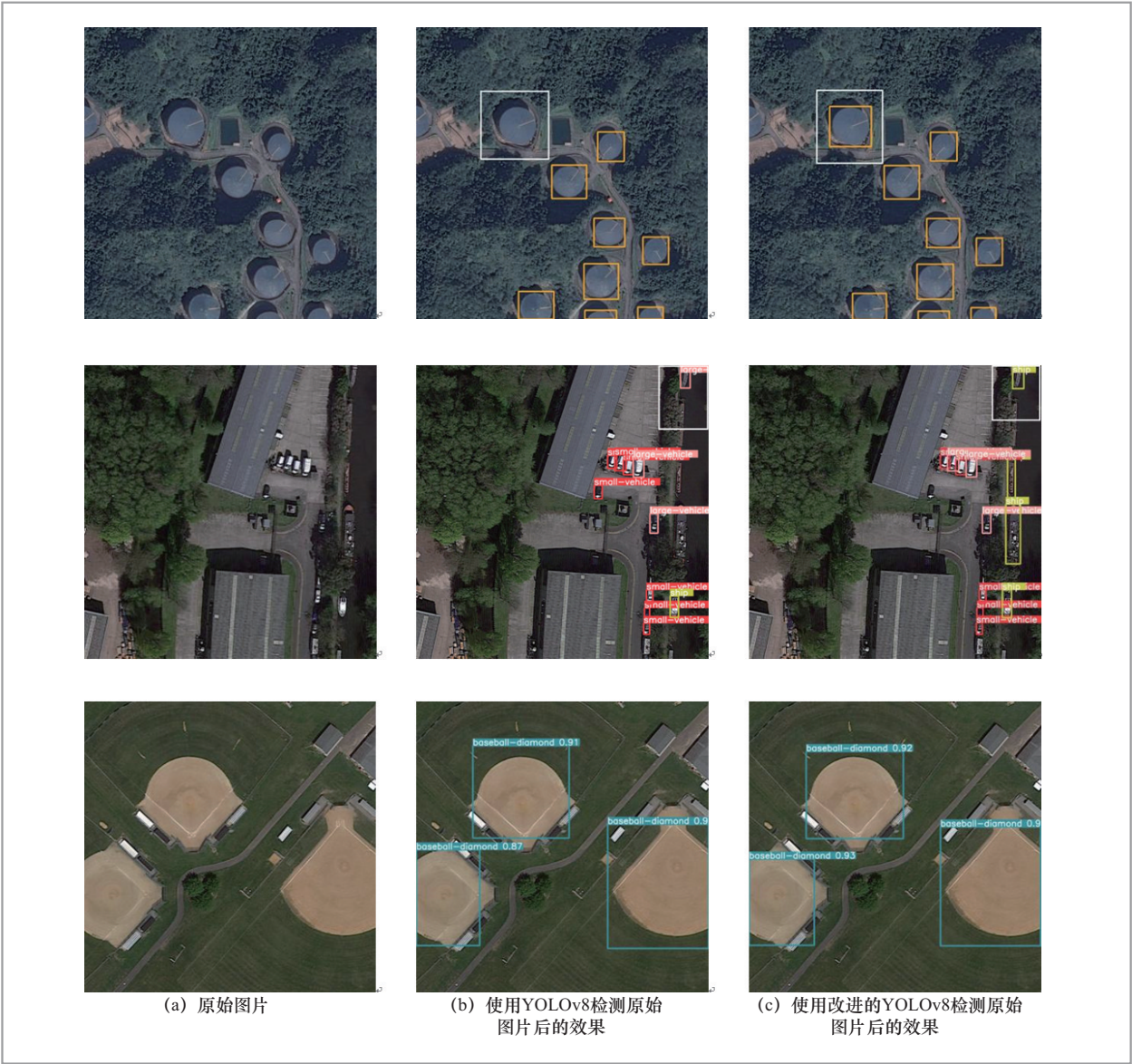


图 14 在 DOTA 数据集上 YOLOv8 和改进 YOLOv8 目标检测结果的比较

知, YOLOv8 在检测图像右上角的小目标时把船误检成大型汽车, 而改进的 YOLOv8 可以充分地捕获上下文信息从而准确识别出小目标船, 有效解决了算法检测小目标时的误检问题; 对比图 14 最后一行后两张图片左边和中间检测框可知, 改进的 YOLOv8 相比 YOLOv8 的检测精度从 0.87、0.91 分别提升到了 0.93 和 0.92。

4 结束语

针对高分遥感图像中目标受复杂背景干扰、目标较小且密集分布、目标多尺度且方向随机等问题, 本文提出了基于改进 YOLOv8 的高分遥感图像目标检测算法。该算法引入 AIFI 模块, 设计了 C2f_IS、C2F_DSC、SDI-BiFPN、EDetect 共 4 个模块以及 NWD-WIoU 损失函数, 并在 3 个高分遥感图像数据集上进行实验。实验结果表明, 改进的 YOLOv8 具有更好的目标检测性能, 充分说明了本文改进的 YOLOv8 在检测高分遥感图像目标时的有效性和鲁棒性。

尽管改进的 YOLOv8 与 RT-DETR、YOLOv7、YOLOv9 等较先进的目标检测算法相比, 其在检测精度、轻量化以及实时性方面均表现更好, 但其 FPS 减小。由表 1、表 2 可知, 每添加一个新模块会导致改进 YOLOv8 的 FPS 变小。为解决该问题, 今后的工作将集中在提高改进 YOLOv8 的 FPS 上, 利用剪枝和蒸馏技术在改进 YOLOv8 的基础上提出一个新的轻量化方法。在牺牲较小检测精度的情况下实现算法 FPS 性能的有效提高, 满足大多数高分辨率遥感图像数据集实际应用的需求。

参考文献:

- [1] 曹峰, 李文涛, 骆剑承, 等. 融合光谱度量标记迁移和 Tri-training 的高光谱遥感图像半监督分类算法[J]. 大数据, 2023, 9(6): 72-89.
CAO F, LI W T, LUO J C, et al. Semi-supervised classification algorithm for hyperspectral remote sensing images fusing spectral measure-based label transfer and tri-training[J]. Big Data Research, 2023, 9(6): 72-89.
- [2] 聂光涛, 黄华. 光学遥感图像目标检测算法综述[J]. 自动化学报, 2021, 47(8): 1749-1768.
NIE G T, HUANG H. A survey of object detection in optical remote sensing images[J]. Acta Automatica Sinica, 2021, 47(8): 1749-1768.
- [3] 袁一钦, 李浪, 姚西文, 等. 光学遥感图像目标检测数据集综述[J]. 遥感学报, 2023, 27(12): 2671-2687.
YUAN Y Q, LI L, YAO X W, et al. A comprehensive review of optical remote-sensing image object detection datasets[J]. National Remote Sensing Bulletin, 2023, 27(12): 2671-2687.
- [4] 冯长峰, 王春平, 付强, 等. 基于深度学习的光学遥感图像目标检测综述[J]. 激光与红外, 2023, 53(9): 1309-1319.
FENG C F, WANG C P, FU Q, et al. Survey of object detection in optical remote sensing images based on deep learning[J]. Laser & Infrared, 2023, 53(9): 1309-1319.
- [5] 刘伟权, 王程, 臧彧, 等. 基于遥感大数据的信息提取技术综述[J]. 大数据, 2022, 8(2): 28-57.
LIU W Q, WANG C, ZANG Y, et al. A survey on information extraction technology based on remote sensing big data[J]. Big Data Research, 2022, 8(2): 28-57.
- [6] GIRSHICK R. Fast R-CNN[C]//Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2015: 1440-

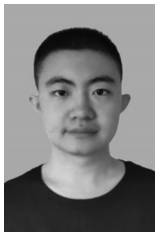
- 1448.
- [7] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.
- [8] HE K M, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN[EB]. arXiv preprint, 2017, arXiv: 1703.06870.
- [9] HE K M, ZHANG X Y, REN S Q, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904-1916.
- [10] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot MultiBox detector[C]//Proceedings of the Computer Vision - ECCV 2016. Cham: Springer, 2016: 21-37.
- [11] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]//Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2017: 2999-3007.
- [12] 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 (CVPR). Piscataway: IEEE Press, 2016: 779-788.
- [13] REDMON J, FARHADI A. YOLO9000: better, faster, stronger[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 7263-7271.
- [14] REDMON J, FARHADI A. Yolov3: an incremental improvement[EB]. arXiv preprint, 2018, arXiv: 1804.02767.
- [15] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. Yolov4: optimal speed and accuracy of object detection[EB]. arXiv preprint, 2020, arXiv: 2004.10934.
- [16] LI C Y, LI L L, JIANG H L, et al. YOLOv6: a single-stage object detection framework for industrial applications[EB]. arXiv preprint, 2022, arXiv: 2209.02976.
- [17] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE, 2023: 7464-7475.
- [18] WANG C Y, YE H I H, LIAO H Y M. YOLOv9: learning what you want to learn using programmable gradient information[EB]. arXiv preprint, 2024, arXiv: 2402.13616.
- [19] ZHANG J N, LI X T, LI J, et al. Rethinking mobile block for efficient attention-based models[EB]. arXiv preprint, 2023, arXiv: 2301.01146.
- [20] DAI J F, QI H Z, XIONG Y W, et al. Deformable convolutional networks[C]//Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2017: 764-773.
- [21] LI D C, LI L, CHEN Z Z, et al. Shift-wiseConv: small convolutional kernel with large kernel effect[EB]. arXiv preprint, 2024, arXiv: 2401.12736.
- [22] QI Y L, HE Y T, QI X M, et al. Dynamic snake convolution based on topological geometric constraints for tubular structure segmentation[C]//Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2023: 6047-6056.
- [23] ZHAO Y A, LYU W Y, XU S L, et al. DETRs beat YOLOs on real-time object detection[EB]. arXiv preprint, 2023,

- arXiv: 2304.08069.
- [24] PENG Y P, SONKA M, CHEN D Z. U-Net v2: rethinking the skip connections of U-Net for medical image segmentation[EB]. arXiv preprint, 2023, arXiv: 2311.17791.
- [25] TAN M X, PANG R M, LE Q V. Efficient-Det: scalable and efficient object detection[C]//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2020: 10778–10787.
- [26] LI H L, LI J, WEI H B, et al. Slim-neck by GSConv: a better design paradigm of detector architectures for autonomous vehicles[EB]. arXiv preprint, 2022, arXiv: 2206.02424.
- [27] LIU J J, HOU Q B, CHENG M M, et al. Improving convolutional networks with self-calibrated convolutions[C]//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2020: 10093–10102.
- [28] ZHENG Z H, WANG P, REN D W, et al. Enhancing geometric factors in model learning and inference for object detection and instance segmentation[J]. IEEE Transactions on Cybernetics, 2022, 52(8): 8574–8586.
- [29] ZHENG Z H, WANG P, LIU W, et al. Distance-IoU loss: faster and better learning for bounding box regression [C]//Proceedings of the AAAI Conference on Artificial Intelligence. New York: AAAI, 2020, 34(7): 12993–13000.
- [30] TONG Z J, CHEN Y H, XU Z W, et al. Wise-IoU: bounding box regression loss with dynamic focusing mechanism[EB]. arXiv preprint, 2023, arXiv: 2301.10051.
- [31] WANG J W, XU C, YANG W, et al. A normalized Gaussian Wasserstein distance for tiny object detection[EB]. arXiv preprint, 2021, arXiv: 2110.13389.
- [32] XIA G S, BAI X, DING J, et al. DOTA: a large-scale dataset for object detection in aerial images[EB]. arXiv preprint, 2017, arXiv: 1711.10398.
- [33] XIAO Z F, LIU Q, TANG G F, et al. Elliptic Fourier transformation-based histograms of oriented gradients for rotationally invariant object detection in remote-sensing images[J]. International Journal of Remote Sensing, 2015, 36(2): 618–644.
- [34] LI K, WAN G, CHENG G, et al. Object detection in optical remote sensing images: a survey and a new benchmark[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2020, 159: 296–307.

作者简介



张霞 (1977–), 女, 博士, 山西大学计算机与信息技术学院 (大数据学院) 副教授, 主要研究方向为人工智能、机器学习与图像处理。



乔环宇（1998-），男，山西大学计算机与信息技术学院（大数据学院）硕士生，主要研究方向为遥感图像目标检测。



曹峰（1980-），男，博士，山西大学计算机与信息技术学院（大数据学院）教授，主要研究方向为人工智能、空间数据挖掘与遥感信息处理。

收稿日期: 2024-06-05
通信作者: zhangxia@sxu.edu.cn
基金项目: 国家自然科学基金项目(No.62072291)
Foundation Item: The National Natural Science Foundation of China (No.62072291)