王广川, 赵寿为 . 融合注意力机制的遥感图像目标检测[J]. 智能计算机与应用,2025,15(12):113-116. DOI:10.20169/j.issn.2095-2163.24030801

融合注意力机制的遥感图像目标检测

王广川, 赵寿为
(上海工程技术大学 数理与统计学院, 上海 201620)

摘 要: 针对光学遥感图像成像条件复杂,物体角度多变,易漏检、误检的特点,提出一种基于 YOLOv8n 改进的遥感图像目标检测算法,以提高遥感图像目标检测的性能。首先,在主干网络部分融合了 SimAM 注意力机制模块,以提高主干网络的特征图质量;其次,结合 ECA 注意力机制和空洞卷积构建了 EC-DCAM 注意力机制模块,融入颈部网络的输出阶段,以提高网络对多尺度特征的学习能力;最后,使用 WIoU 损失函数替换 CIoU 损失函数进行动态的损失计算。将改进的算法在 DIOR 数据集和 MAR20 数据集上进行了实验,在 DIOR 数据集上的准确率提升了 1.3%,达到了 74.7%;在 MAR20 数据集上的准确率提升了 3.8%,达到了 87.2%,改进算法的参数量和计算复杂度几乎没增加,验证了改进算法的有效性。

关键词: WIoU 损失; SimAM 注意力机制; ECA 注意力机制; 空洞卷积
中图分类号: TP751 **文献标志码**: A **文章编号**: 2095-2163(2025)12-0113-04

Remote sensing image object detection with attention mechanism

WANG Guangchuan, ZHAO Shouwei
(School of Mathematics, Physics and Statistics, Shanghai University of Engineering Science, Shanghai 201620, China)

Abstract: In view of the complex imaging conditions of optical remote sensing images, the variability of object angles, and the characteristics of easy missed detection and false detection, based on the YOLOv8n object detection algorithm, an improved remote sensing image object detection algorithm based on YOLOv8n is proposed to improve the performance of remote sensing image object detection. Firstly, the SimAM attention mechanism module is integrated into the backbone network to improve the quality of the feature map of the backbone network. Secondly, the EC-DCAM attention mechanism module is constructed by combining the ECA attention mechanism and the dilated convolution, and is integrated into the output stage of the neck network to improve the network′s ability to learn multi-scale features. Finally, the WIoU loss function is used to replace the CIoU loss function for dynamic loss calculation. The improved algorithm was tested on the DIOR dataset and MAR20 dataset, and the accuracy rate on the DIOR dataset increased by 1.3% to 74.7%; on the MAR20 dataset, the accuracy rate increased by 3.8% to 87.2%. At the same time, the number of parameters and computational complexity of the improved algorithm were almost unchanged. The experimental results verify the effectiveness of the method.

Key words: WIoU loss; SimAM attention mechanism; ECA attention mechanism; dilated convolution.

0 引 言

近年来,随着遥感技术的发展,光学遥感图像也越来越丰富。光学遥感图像包含丰富的物体空间信息,对其进行定位和分类,可以为农产品估产、国土资源普查、城市规划及防灾减灾提供不可替代的支持[1]。目前对遥感图像的目标检测主要基于深度学习的目标检测算法,可以分为单阶段目标检测算法和双阶段目标检测算法[2]。双阶段目标检测算法有 R-CNN(Regions with CNN features)、Faster RCNN、Cascade RCNN 等;单阶段目标检测算法有 YOLO(You Only Look Once, YOLO)、YOLOX、YOLOv3 等。基于深度学习的目标检测算法在自然图像的目标检测中取得了良好的效果,与自然图像相比,遥感图像背景复杂,物体的方向变化大,目标密集,这对目标检测算法在遥感图像的应用提出了挑战。程换新等[3]通过融合 EMA(Efficient Multi-Scale Attention)注意力机制并使用 WIoU(Wise

作者简介: 王广川(1992—),男,硕士,主要研究方向:计算机视觉,模式识别。
通信作者: 赵寿为(1982—),女,博士,副教授,硕士生导师,主要研究方向:应用统计分析。Email:zhaoshouwei_1029@163.com。
收稿日期: 2024-03-08

Intersection over Union）损失函数替换 CIoU（Complete Intersection over Union）损失函数以提高遥感图像目标检测算法的准确率；张秀再等[4]基于 YOLOv8 算法，融合全局注意力机制并结合可变形卷积以提升遥感图像目标检测算法的准确率。鉴于 YOLO 系列目标检测算法良好的性能，本文基于 YOLOv8 目标检测算法进行改进，以提高目标检测算法在遥感图像目标检测任务上的性能。首先，在主干网络每个输出阶段融入 SimAM（Simple，parameter-free Attention Module）注意力机制；其次，在颈部网络融合了由 ECA（Efficient Channel Attention）注意力机制和空洞卷积构成的 EC-DCAM（Efficient Channel with Dilated Convolution Attention Module）模块，以提高多尺度特征的学习能力；最后，使用 WIoU 损失函数替换 CIoU 损失函数，动态的计算损失。改进的 YOLOv8n 遥感图像目标检测算法在 DIOR 数据集和 MAR20 数据集上进行了实验，验证了改进算法的效果和泛化能力。

1 改进的 YOLOv8 目标检测算法

1.1 SimAM 注意力机制模块

注意力机制模块通常可以轻易融合到神经网络当中，优化神经网络的输出并使网络学到更多信息。常见的注意力机制如 SE（Squeeze-and-Excitiaion）注意力模块对不同通道特征图的信息进行建模，以调整神经网络通道间的权重信息[5]。受 SE 注意力机制的启发，EncNet（Context Encoding Network，EncNet）提出上下文信息编码模块，对上下文信息和物体类别概率的关系进行建模[6]。SimAM 是一种无参数的注意力机制模块，其同时考虑通道维度和空间维度的特征以生成注意力权重。神经科学理论认为，带有最多信息的神经元通常表现出对周围神经元的空间抑制性，这类神经元应该给予更多的权重。SimAM 注意力机制通过计算神经元与周围其他神经元的线性可分性，评估神经元的重要性，计算方法如下：

$$e_t^* = \frac{4(\tilde{\sigma}^2 + \lambda)}{(t - \tilde{\mu})^2 + 2\tilde{\sigma}^2 + 2\lambda} \tag{1}$$

其中：

$$\tilde{\mu} = \frac{1}{M}\sum_{i=1}^{M} x_i \tag{2}$$

$$\tilde{\sigma}^2 = \frac{1}{M}\sum_{i=1}^{M} (x_i - \tilde{\mu}) \tag{3}$$

t 是当前神经元的值；M 是单个通道上神经元个数；λ 是一个正则化常数，以防止分母为 0，本文 λ 取值为 0.000 1。

e_t^* 的值越低，神经元 t 与周围神经元的差异越大，也就越重要，因此可以用 $(e_t^*)^{-1}$ 表示神经元的权重。

1.2 EC-DCAM 模块

CBAM（Convolutional Block Attention Module）注意力模块包含 CAM（Channel Attention Module）和 SAM（Spatial Attention Module）两部分，对神经网络的通道特征和空间特征依次学习，网络输入的特征图依次通过 CAM 模块和 SAM 模块生成注意力权重，然后和原输入特征图做逐元素的乘积[7]。OUYANG 等[8]提出维持通道注意力模块的维度并进行适当的跨通道交互，可以保持通道注意力机制性能的同时降低模块复杂度。本文用 ECA 模块替代了原 CBAM 模块的通道注意力子模块，为了提高 SAM 模块的多尺度特征感知能力，引入不同空洞率的空洞卷积重新设计空间注意力模块，构建了 EC-DCAM 注意力机制模块，ECA 模块的结构如图 1 所示，改进后 EC-DCAM 模块结构如图 2 所示。

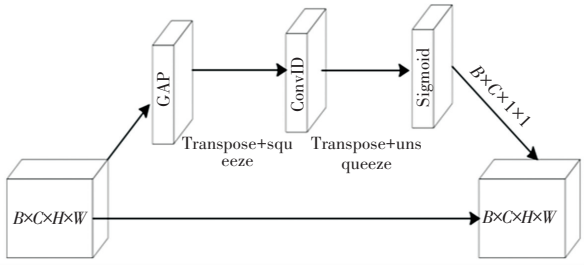

图 1 ECA 模块结构图

Fig. 1 ECA module structure diagram

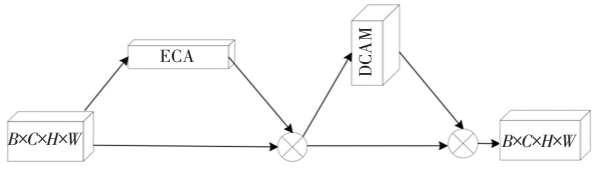

图 2 EC-DCAM 模块结构图

Fig. 2 Structure of EC-DCAM module

1.3 WIoU 损失函数

目标检测任务中，IoU（Intersection over Union，IoU）经常用于衡量预测边框和真实标注框的差异，当预测框和标注框没有交集时 IoU 恒为 0，IoU 损失恒为 1 时 IoU 损失无法衡量两者的相似程度，因此不断有工作对 IoU 的计算方式进行改进，如 WIoU、CIoU、GIoU、SIoU 等。本文使用的 WIoU 是种动态

非单调的 IoU 损失，计算方法如下：

$$\mathcal{L}_{\mathrm{WIoU}} = R_{\mathrm{IoU}}\,\mathcal{L}_{\mathrm{IoU}} \tag{4}$$

其中：

$$\mathcal{L}_{\mathrm{IoU}} = 1 - \mathrm{IoU} = 1 - \frac{W_i H_i}{S} \tag{5}$$

$$R_{\mathrm{WIoU}} = \exp\left(\frac{(x - x_{gt})^2 + (y - y_{gt})^2}{(W_g^2 + H_g^2)^*}\right) \tag{6}$$

其中，W_i，H_i 是预测框和真实框相交区域的宽高；W_g，H_g 是包围预测框和真实框的最小外接矩形的宽高；S 是最小外接矩形的面积。

$R_{\mathrm{WIoU}} \in [1, \mathrm{e})$ 可以放大普通质量锚框的 IoU 损失，$\mathcal{L}_{\mathrm{IoU}} \in [0,1]$ 可以显著降低高质量锚框的 $\mathcal{L}_{\mathrm{WIoU}}$，同时更关注锚框和标注框中心点重合时的距离。

2　实验结果分析

2.1　实验环境和数据集实验

实验环境见表 1。为了验证改进算法的性能，在 DIOR 数据集和 MAR20 军事飞机遥感图像数据集上进行了实验。DIOR 数据集是西北工业大学开源的光学遥感数据集，包含 23 463 张分辨率为 800×800 的图片，有 20 个物体类别，190 288 个标注对象。实验按照 1∶1∶2 划分训练集、验证集和测试集。MAR20 数据集是西北工业大学开源的军事飞机遥感图像目标检测数据集，包含 3 842 张图像、20 种军用飞机型号以及 22 341 个目标实例，训练集包含 1 331 张图像和 7 870 个目标实例，测试集包含 2 511 张图像和 14 471 个目标实例。

表 1　实验环境

Table 1　Experimental environment

名称	配置
操作系统	Ununtu 20.04.4
开发环境	CUDA11.3
显卡	NVIDIA RTX3090
epochs	100
学习率	0.01
图片尺寸	800×800

2.2　评价指标

遥感图像目标检测中常用的评价指标是 PASCAL VOC2007 数据集的评价指标，即 mAP（mean Average Precision，mAP），IoU 阈值为 0.5。

模型的参数量是指模型训练中需要训练的参数总量，用来衡量模型的空间复杂度，参数越小的模型越容易在边缘设备上部署。

FLOPs（Floating Point Operations，FLOPs）指浮点运算次数，用来衡量算法前向传播一次所需要执行的浮点运算次数，越低代表算法越高效。

2.3　EC-DCAM 模块消融实验

为了验证 EC-DCAM 模块的有效性，在 DIOR 数据集上进行了消融实验，结果见表 2。可以看出融入 YOLOv8n 目标检测算法后，EC-DCAM 模块对基线模型提升最多，达到 0.9，验证了构建模块的有效性。

表 2　DIOR 数据集 EC-DCAM 模块消融实验

Table 2　Experiment on EC-DCAM module of DIOR dataset

算法模块	mAP50/%	参数量/M	FLOPs/G
基线模型	73.4	3.015	6.375
+CBAM	73.8	3.026	6.378
+ECA+SAM	74.0	3.015	6.377
+ECA-DCAM	74.3	3.015	6.377

2.4 不同模块消融实验

基于 YOLOv8n 算法、SimAM 注意力模块、EC-DCAM 注意力模块和 WIoU 损失函数，本文构建了改进的目标检测算法。改进的算法在 DIOR 数据集的实验结果见表 3，可以看出改进算法的参数和浮点计算量几乎未增加，mAP50 提高了 1.3%。在 MAR20 数据集的实验结果见表 4，mAP50 提高了 3.8%，验证了改进算法的泛化能力。基线模型和改进算法在 DIOR 数据集和 MAR20 数据集训练过程的损失变化如图 3 和图 4 所示，可见改进的算法和基线模型有类似的损失变化曲线，但改进算法有更低的损失，即改进的算法有更好的收敛性。

表 3　DIOR 数据集消融实验

Table 3　Ablation experiments on DIOR dataset

算法	mAP50/%	mAP75/%	参数量/M	FLOPs/G
YOLOv8n	73.4	58.5	3.015	6.375
+SimAM	74.3	60.1	3.015	6.375
+EC-DCAM	74.3	58.5	3.015	6.377
+WIoU	74.2	58.6	3.015	6.375
本文算法	74.7	59.0	3.015	6.377

表 4　MAR20 数据集消融实验

Table 4　Ablation experiments on MAR20 dataset

算法	mAP50/%	mAP75/%	参数量/M	FLOPs/G
YOLOv8n	83.4	77.6	3.015	6.375
本文算法	87.2	80.9	3.015	6.377

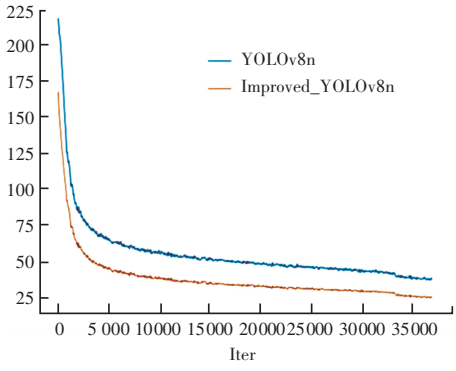

图 3 DIOR 数据集损失变化

Fig. 3 Loss change of DIOR dataset

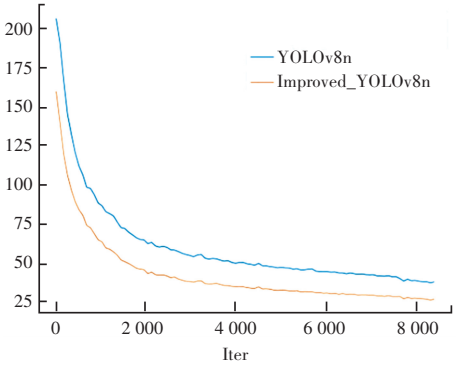

图 4 MAR20 数据集损失变化

Fig. 4 Loss variation of MAR20 dataset

3 结束语

本文结合 ECA 注意力机制和空洞卷积构建了 EC-DCAM注意力机制模块,并在DIOR数据集上验证了 EC-DCAM 模块的效果。为了提高遥感图像目标检测算法的性能,基于 YOLOv8n 目标检测算法,结合 SimAM 注意力模块、EC-DCAM 模块和 WIoU 损失函数构建了改进的遥感图像目标检测算法,在 DIOR 数据集和 MAR20 数据集上的实验结果验证了改进算法的有效性。

参考文献

[1] 徐芳. 可见光遥感图像海面目标自动检测关键技术研究[D]. 长春: 中国科学院大学(中国科学院长春光学精密机械与物理研究所), 2018.

[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.

[3] 程换新, 矫立浩, 骆晓玲, 等. 改进 YOLOv8 的遥感图像检测算法[J]. 无线电工程, 2024, 54(5): 1155-1161.

[4] 张秀再, 沈涛, 许岱. 基于改进 YOLOv8 算法的遥感图像目标检测[J]. 激光与光电子学进展, 2024, 61(10): 298-308.

[5] HU J, LI S, SUN G, et al. Squeeze-and-excitiaion networks [C]//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2018: 7132-7141.

[6] ZHANG H, DANA K, SHI J, et al. Context encoding for semantic segmentation[C]// Proceedings of 2018 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ:IEEE, 2018: 7151-7160.

[7] WOO S, PARK J, LEE J, et al. CBAM: Convolutional block attention module[C]// Proceedings of the European Conference on Computer Vision (ECCV). Cham: Springer, 2018: 3-19.

[8] OUYANG D, HE S, ZHANG G, et al. Efficient multi-scale attention module with cross-spatial learning[C]// Proceedings of ICASSP 2023- 2023 IEEE International Conference on Acoustics, Speech and Signal Processing. Piscataway, NJ:IEEE, 2023: 1-5.