

DOI:10.20079/j.issn.1001-893x.221014002

基于注意力和自适应特征融合的 SAR 图像飞机目标检测*

夏一帆^{1,2}, 赵凤军¹, 王樱洁¹, 王春乐¹

(1. 中国科学院空天信息创新研究院, 北京 100190; 2. 中国科学院大学 电子电气与通信工程学院, 北京 100049)

摘要:针对合成孔径雷达(Synthetic Aperture Radar, SAR)图像中飞机目标尺度多样性及背景强散射干扰的问题,提出了一种基于坐标注意力和自适应特征融合的 YOLOv4 SAR 图像飞机目标检测算法。该方法首先在主干网络引入坐标注意力机制,以增强对于飞机散射点组合结构的聚焦能力以及抗背景干扰能力。其次,在特征增强网络中引入自适应特征融合机制,提高了对不同大小飞机的特征提取能力,同时改善了 YOLOv4 算法召回率和精确率不平衡的问题。最后,通过改进的 K-Means 聚类针对飞机目标调整先验框的尺寸,提高了模型的定位精度。实验结果表明,改进算法召回率达到 91.01%,精确率达到 90.09%, $AP_{0.5}$ 达到 92.34%,分别较原 YOLOv4 算法提高 2.49%,6.56% 和 3.62%。

关键词:合成孔径雷达(SAR);飞机检测;注意力机制;特征融合

开放科学(资源服务)标识码(OSID):



微信扫描二维码
听独家语音释文
与作者在线交流
享本刊专属服务

中图分类号:TN957.52 文献标志码:A 文章编号:1001-893X(2024)03-0350-08

Aircraft Detection in SAR Images Based on Attention and Adaptive Feature Fusion

XIA Yifan^{1,2}, ZHAO Fengjun¹, WANG Yingjie¹, WANG Chunle¹

(1. Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100190, China; 2. School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: For the problems of multi-scale aircraft targets and strong scattering interference in the background in synthetic aperture radar(SAR) images, an improved YOLOv4 SAR target detection algorithm based on coordinate attention and adaptive feature fusion is proposed. Firstly, the algorithm introduces the coordinate attention mechanism in the backbone network to enhance the focusing ability on the combined structure of aircraft scattering features and the resistance to background interference. Secondly, the adaptive spatial feature fusion mechanism is implemented in the feature enhancement network to improve the feature extraction capability for aircraft of different sizes, while improving the imbalance between the recall and accuracy rates of YOLOv4. Finally, the size of the prior anchor is adjusted for aircraft targets by improved K-means clustering to improve the localization accuracy of the model. The experimental results show that the improved YOLOv4 achieves 91.01% recall, 90.09% accuracy, and 92.34% $AP_{0.5}$, which are 2.49%, 6.56%, and 3.62% better than YOLOv4 respectively.

Key words: synthetic aperture radar(SAR); aircraft detection; attention mechanism; feature fusion

* 收稿日期:2022-10-14;修回日期:2023-02-16
基金项目:国家自然科学基金资助项目(61901445)
通信作者:赵凤军 Email:fjzhao@mail.ie.ac.cn

0 引言

合成孔径雷达 (Synthetic Aperture Radar, SAR) 以其全天候、全天时、高分辨率和穿透云雾和植被的能力,具有光学传感器无法比拟的优势,在目标检测中逐步得到应用。飞机目标具有高价值和时效性的特点^[1],对检测的自动化、有效性和实时性提出了较高的要求,因而精确、鲁棒、快速的飞机目标检测算法成为当前研究的热点。

以恒虚警率 (Constant False Alarm Rate, CFAR) 等为代表的传统 SAR 目标检测算法,主要思想是将目标从背景中分割出来后,根据目标的特性建立模型进行检测和识别。传统的算法^[2-5]检测往往依赖人工设计特征,实现过程效率低且泛化能力差,对强噪声的抑制能力较弱,在复杂场景下受周围强反射地物干扰难以准确建模。

随着计算机算力的提升和深度学习的发展,基于神经网络的目标检测算法开始兴起。Girshick 等人^[6]提出了二阶段目标检测算法 R-CNN (Region Convolution Neural Network),通过选择性搜索算法获得候选区域,并使用 CNN 提取特征,然后将提取的特征输入分类器进行分类,最后利用全连接网络进行边框回归。此后的 Faster R-CNN^[7]和 Cascade R-CNN^[8]等都是基于该思想加以改进。Redmon 等人^[9]提出了一阶段目标检测算法 YOLO (You Only Look Once),其思想就是将图像划分成多个网格,然后为每一个网格同时预测边界框并给出相应概率。YOLO 系列发展迅速,迄今已有多个版本,受到了研究者们广泛认可^[10-12]。

虽然许多研究者都注意到了飞机目标在 SAR 图像中多呈现出小目标密集分布的情况,但是却往往忽视了 SAR 图像中飞机目标的多尺度的问题。过于专注于小尺度飞机的检测效果会影响模型对于大尺度飞机的整体特征的提取。此外,在飞机检测过程中,背景地物的强散射点 (如航站楼、塔台等) 分布在飞机附近会弱化飞机目标的散射,且飞机部件的散射会和地物的散射点混淆,使得飞机难以判别和准确定位。为此,本文着眼于 SAR 图像飞机检测中目标多尺度和背景强散射干扰的两大挑战,试图提出一种基于坐标注意力和自适应特征融合的 SAR 图像飞机目标检测方法。算法首先在 YOLOv4 的主干网络中引入坐标注意力机制 (Coordinate Attention, CA),有效改善了对于图像边缘目标检测效果,增强了对于飞机散射点集合的聚焦能力。同

时,在特征融合网络中将路径聚合网络 (Path Aggregation Network, PANet) 和自适应特征融合 (Adaptively Spatial Feature Fusion, ASFF) 网络结合,优化网络对飞机多层特征的提取与融合,抑制特征图中噪声叠加。此外,为了先验框能够更好地检测目标,利用基于交并比 (Intersection over Union, IOU) 距离改进的 K-means 算法对真实框进行重新聚合获取先验框。改进后的 YOLOv4 召回率达到 91.01%,精确率达到 90.09%,AP_{0.5} 达到 92.34%,显著优于原算法。

1 算法描述

1.1 YOLOv4 算法简介

YOLOv4^[13]作为一种单阶段目标检测算法,核心思想是利用整张图片作为网络的输入,直接输出检测框的位置和类别。YOLOv4 通过对前人的研究成果进行整理组合,加以创新,实现了速度和精度的平衡。在预处理阶段,YOLOv4 采用 Mosaic 数据增强方式,将 4 张图片进行随机变换组成一张图片,达到扩充数据集的目的。YOLOv4 改进自 YOLOv3,主干网络相较于 YOLOv3,在原主干网络 Darknet53 的基础上引入跨阶段局部网络 (Cross Stage Partial Network, CSP) 和 Mish 激活函数,最终获得 CSPDarknet53 网络。为了更好地利用特征信息,YOLOv4 在颈部网络引入空间金字塔池化^[14] (Spatial Pyramid Pooling, SPP) 和路径聚合网络来增强感受野的作用,获取更全面的细节特征。

1.2 坐标注意力机制

在 SAR 图像飞机检测的过程中,一方面,在 SAR 图像存在大量的噪声,这些噪声会影响模型对于目标特征的提取^[15];另一方面,飞机所在的机场中往往会有建筑物、车辆等具有强反射点的物体,这些会对飞机的检测造成干扰。为了抑制背景噪声,提高模型对于飞机目标特征表达能力,本文在主干网络中集成了坐标注意力机制^[16]。

受到人类视觉机制的启发,注意力机制倾向于通过快速扫描全局图像,关注图像中有利于判断的部分信息,并忽略掉不相关的信息。注意力机制具有“即插即用”的特点,对于通道数为 C ,高度为 H ,宽度为 W 的张量 X 进行转化后能够输出同尺寸张量 Y ,不影响输入后续的网络结构,即

$$\begin{aligned} X &= [x_1, x_2, \cdots, x_c] \in \mathbb{R}^{H \times W \times C} \rightarrow Y = \\ &[y_1, y_2, \cdots, y_c] \in \mathbb{R}^{H \times W \times C} \end{aligned} \tag{1}$$

图 1 为坐标注意力机制的结构图。坐标注意力机制分为位置信息嵌入和注意力生成两部分。

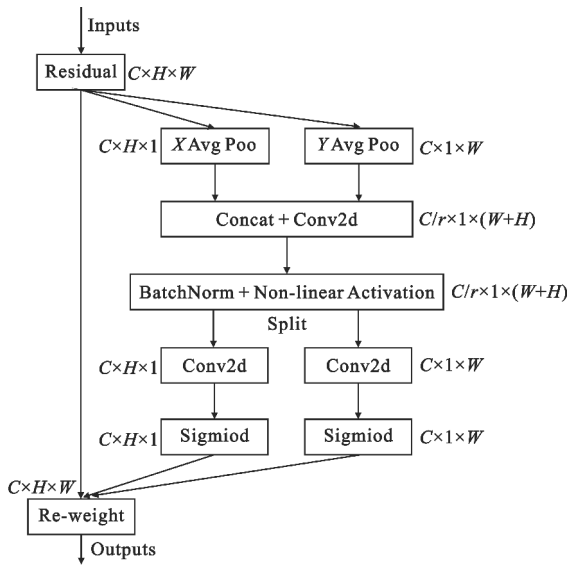


图 1 坐标注意力结构
Fig. 1 Structure of coordinate attention

对于输入的特征图 $\mathbf{X} \in \mathbb{R}^{H \times W \times C}$, 首先利用沿水平方向的全局池化和沿垂直方向的全局池化对每个通道进行编码, 获得一对嵌入位置信息的方向感知特征图。其中, 高度为 h 的第 c 通道的输出特征图 $z_c^h(h)$ 和宽度为 w 的第 c 通道的输出特征图 $z_c^w(w)$ 可以分别表示为

$$z_c^h(h) = \frac{1}{W} \sum_{0 \leq j < W} x_c(h, j) \tag{2}$$

$$z_c^w(w) = \frac{1}{H} \sum_{0 \leq j < H} x_c(j, w) \tag{3}$$

之后在注意力生成阶段, 对上面变换拼接成大小为 $C \times 1 \times (H+W)$ 的特征图。为了降低模型的复杂性和计算的开销, 采用 1×1 的卷积变换将其压缩至 C/r 维 (在实验中取 $r = 16$), 之后使用激活函数 (在本文使用 GELU 激活函数) 进行非线性激活。即

$$f = \delta(F_1(\text{concat}(z^h, z^w))) \tag{4}$$

式中: f 表示获得的特征图; δ 表示非线性激活函数; F_1 表示 1×1 的卷积变换; concat 表示拼接操作。

将获取到的特征图 f 沿空间维分解为张量 $f^h \in \mathbb{R}^{C/r \times H}$ 和 $f^w \in \mathbb{R}^{C/r \times W}$ 。利用 1×1 的卷积变换将两张量恢复至 C 维, 并利用 sigmoid 激活函数进行非线性激活, 获得注意力图:

$$g^h = \sigma(F_h(f^h)) \tag{5}$$

$$g^w = \sigma(F_w(f^w)) \tag{6}$$

最后, 将注意力图 g^w 和 g^h 和输入特征图 x 相乘获得输出:

$$y_c = x_c \times g_c^h \times g_c^w \tag{7}$$

如图 2 所示, 本文在 YOLOv4 的主干网络 CSP-Darknet53 后 3 个残差块中集成坐标注意力模块, 使不同尺度的特征图能够充分利用全局信息, 同时侧重于提升对飞机散射点集合的特征的提取能力。

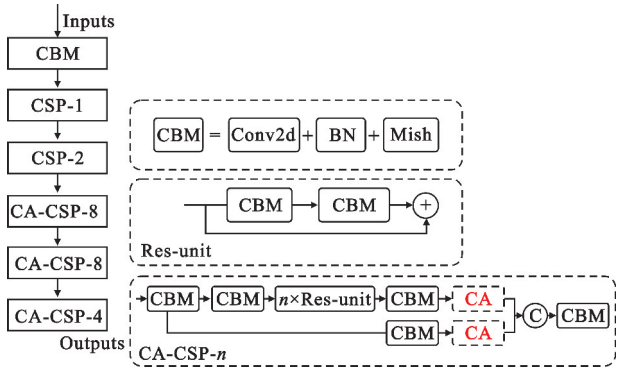


图 2 集成坐标注意力模块的主干网络 CA-CSPDarknet53
Fig. 2 Backbone CA-CSPDarknet53 with integrated coordinate attention module

对样本进行热力图实验验证, 结果如图 3 所示。对比图 3(b) 和图 3(c) 的红色区域可见, 引入注意力机制后, 网络对于飞机散射结构的聚焦能力得到增强, 目标检测效果得到有效改善。

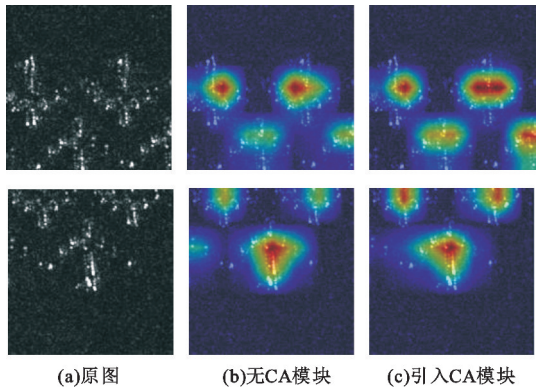


图 3 热力图对比
Fig. 3 Comparison of heatmaps

1.3 自适应特征融合技术

在特征提取与融合阶段, YOLOv4 的目标检测采用 PANet^[17] 结构。PANet 对于不同尺度的特征图通过卷积、上下采样变换成相同尺寸后进行堆叠。当数据集包含不同大小的目标时, 会出现在某一层特征图中一个对象被判定为正, 在其他层次中被判定负的矛盾。为了提高对不同大小飞机的特征提取能力, 同时降低不同特征图中的噪声叠加的影响, 本文在特征融合网络 PANet 和 YOLO 检测头之间引入自适应特征融合机制 ASFF^[18], 结构如图 4 所示。

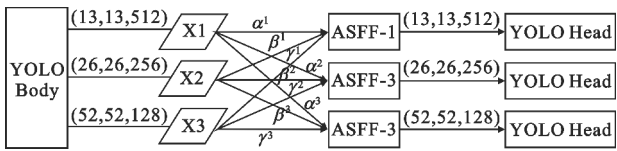


图 4 ASFF 结构
Fig. 4 Structure of ASFF

ASFF 在特征融合过程中引入的可学习的权重,能自适应地调整不同特征层之间融合的比例,抑制不同层之间的冲突信息,增强关键特征的提取能力,减少推理开销。如图 4 所示,输入图像通过 YOLO 的主干网络和 PANet 获得 3 个不同尺度的特征: $y_{ij}^l = \alpha_{ij}^l \cdot x_{ij}^{1 \rightarrow l} + \beta_{ij}^l \cdot x_{ij}^{2 \rightarrow l} + \gamma_{ij}^l \cdot x_{ij}^{3 \rightarrow l}$, $y_{ij}^l = \alpha_{ij}^l \cdot x_{ij}^{1 \rightarrow l} + \beta_{ij}^l \cdot x_{ij}^{2 \rightarrow l} + \gamma_{ij}^l \cdot x_{ij}^{3 \rightarrow l}$ 和 $y_{ij}^l = \alpha_{ij}^l \cdot x_{ij}^{1 \rightarrow l} + \beta_{ij}^l \cdot x_{ij}^{2 \rightarrow l} + \gamma_{ij}^l \cdot x_{ij}^{3 \rightarrow l}$ 。令 $y_{ij}^l = \alpha_{ij}^l \cdot x_{ij}^{1 \rightarrow l} + \beta_{ij}^l \cdot x_{ij}^{2 \rightarrow l} + \gamma_{ij}^l \cdot x_{ij}^{3 \rightarrow l}$ 表示从特征层 $y_{ij}^l = \alpha_{ij}^l \cdot x_{ij}^{1 \rightarrow l} + \beta_{ij}^l \cdot x_{ij}^{2 \rightarrow l} + \gamma_{ij}^l \cdot x_{ij}^{3 \rightarrow l}$ 经过上采样或下采样调整到 $y_{ij}^l = \alpha_{ij}^l \cdot x_{ij}^{1 \rightarrow l} + \beta_{ij}^l \cdot x_{ij}^{2 \rightarrow l} + \gamma_{ij}^l \cdot x_{ij}^{3 \rightarrow l}$ 特征层

$y_{ij}^l = \alpha_{ij}^l \cdot x_{ij}^{1 \rightarrow l} + \beta_{ij}^l \cdot x_{ij}^{2 \rightarrow l} + \gamma_{ij}^l \cdot x_{ij}^{3 \rightarrow l}$ 的特征向量,则融合公式可以表示为

$$y_{ij}^l = \alpha_{ij}^l \cdot x_{ij}^{1 \rightarrow l} + \beta_{ij}^l \cdot x_{ij}^{2 \rightarrow l} + \gamma_{ij}^l \cdot x_{ij}^{3 \rightarrow l} \quad (8)$$

式中:自适应权重 $\alpha_{ij}^l, \beta_{ij}^l$ 和 γ_{ij}^l 通过对调整到相同尺度的不同层级的特征图经过 1×1 的卷积获得。权重经过张量拼接后,利用 softmax 函数使参数的约束在 $[0,1]$ 且和为 1,即满足

$$\alpha_{ij}^l = \frac{e^{\lambda_{\alpha_{ij}}^l}}{e^{\lambda_{\alpha_{ij}}^l} + e^{\lambda_{\beta_{ij}}^l} + e^{\lambda_{\gamma_{ij}}^l}} \quad (9)$$

$$\alpha_{ij}^l + \beta_{ij}^l + \gamma_{ij}^l = 1 \quad (10)$$

$$\alpha_{ij}^l, \beta_{ij}^l, \gamma_{ij}^l \in [0,1] \quad (11)$$

PANet 获得的 3 个不同尺度的特征图经过 ASFF 处理后,获得的 3 个特征图尺寸仍与输入一致,方便输入检测头结构获得最终的分类和坐标信息。结合前文改进骨干网络 CA-CSPDarknet53,最终改进的 YOLOv4 网络结构如图 5 所示。

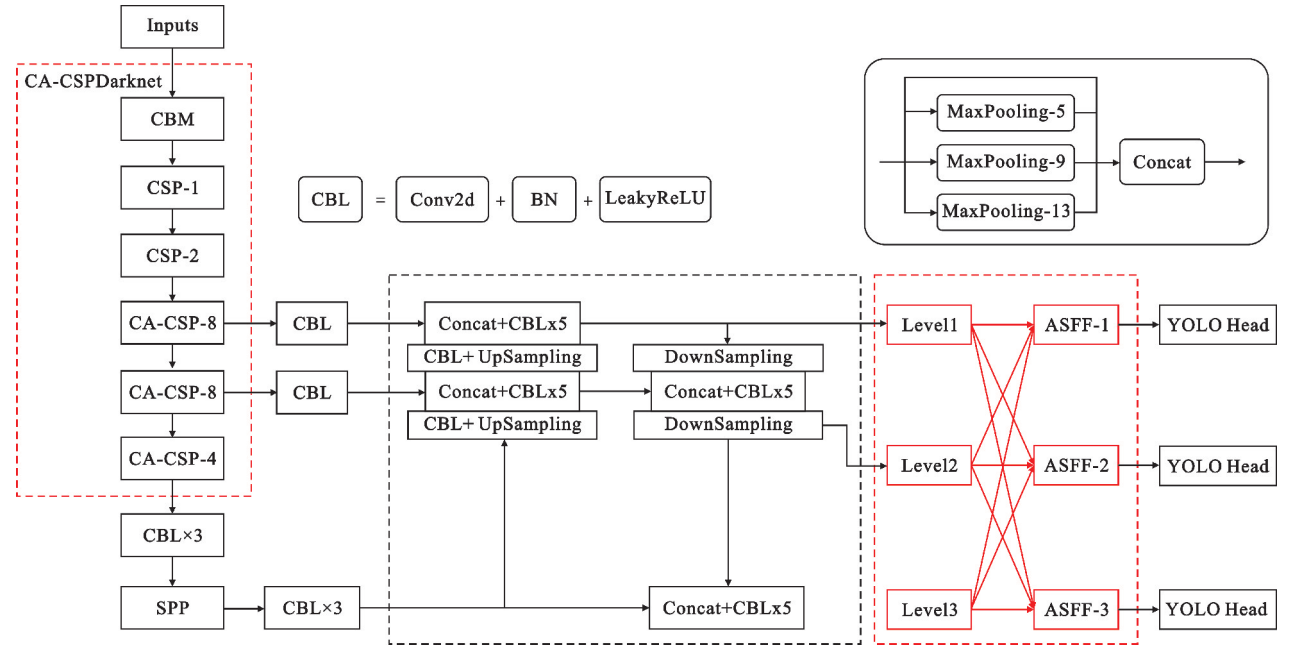


图 5 改进 YOLOv4 结构
Fig. 5 Structure of improved YOLOv4

1.4 利用改进 K-means 聚类方法获得先验框

对于 YOLO 等基于锚框的算法,锚框的设置将最终影响检测的性能。原始的 YOLO 锚框主要是针对 COCO 数据集聚类得到的,适用于检测多类别目标,对于本文要识别的对象并不适用,因此本文利用 K-mean 聚类方法获得先验框。为了避免大尺寸框在欧氏距离中分布过于离散导致获得的先验框不能覆盖到不同尺寸的目标的情况,实验中用 IoU 的距

离替换欧氏距离。IoU 距离表示为

$$D = 1 - \text{IoU}(\text{box}, \text{centriod}) \quad (12)$$

$$\text{IoU}(\text{box}, \text{centriod}) = \frac{\text{box} \cap \text{center}}{\text{box} \cup \text{center}} \quad (13)$$

式中:IoU 表示交并比运算;box 为标记框面积;centriod 为类簇中心面积。最终的聚类获得的 9 个先验框为 (18, 18), (20, 31), (29, 22), (26, 26), (29, 31), (37, 37), (53, 65), (66, 53), (107, 107)。

2 分析与验证

2.1 实验数据集

本实验采用华中科技大学公开的 SAR 图像飞机目标检测数据集^[19] (SAR Aircraft Detection Dataset, SADD)。该数据集获取自 TerraSAR-X 卫星,图像分辨率在 0.5~3 m 不等,部分数据集如图 6 所示。

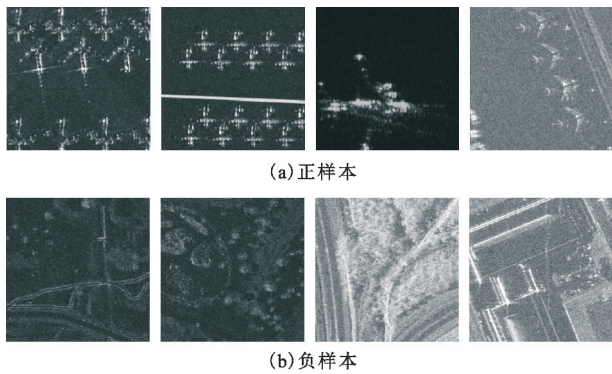


图 6 数据集展示
Fig. 6 Examples of SADD dataset

数据集共有 2 966 个大小为 224×224 的切片,包括 7 835 个飞机目标。本次实验将数据集按照 6 : 2 : 2 的比例随机划分为训练集、验证集和测试集,如表 1 所示。

表 1 数据集划分
Tab. 1 Dataset splitting

数据集	正样本	负样本	总计	飞机数
训练集	547	1 232	1 779	5 015
验证集	168	425	593	1 452
测试集	168	426	594	1 368
总计	883	2 083	2 966	7 835

2.2 评价指标

本次实验使用精确率 (Precision)、召回率 (Recall) 和类别平均精度 (Average Precision, AP)^[20] 3 项指标对模型的检测效果进行量化评价,并采用采用输入张量大小为 (224, 224, 3) 时的算法 10 亿次浮点运算次数 (Giga Floating Point Operations, GFLOPs)^[21] 和生成参数文件大小来衡量计算效率与存储占用量。

精确率表征为在所有被检测为飞机目标的结果中实际为飞机目标的概率,召回率表征为实际飞机目标中被预测为正样本的概率,两者公式如下:

Precision= $\frac{TP}{TP+FP}$ ×100% (14)

Recall= $\frac{TP}{TP+FN}$ ×100% (15)

式中:TP 表示被正确检测到的飞机目标数量;FP 表示被算法错误识别出的飞机目标数量;FN 表示没有被识别出的真正的飞机目标数量。

AP 能够反映单目标检测算法的全局性能,公式为

AP_{IoU} = $\int_0^1$ Precision (Recall) dRecall (16)

式中:Precision (Recall) 表示精确率-召回率曲线;IOU 为交并比阈值。本次实验采用 AP_{0.5} 作为评价指标。

由于在实际应用中算法的计算效率与算法复杂度、硬件环境甚至深度学习框架的版本有关,而软硬件环境往往是不同的,我们从算法的复杂度对算法的计算效率进行刻画,故采用输入张量大小为 (224, 224, 3) 时的算法浮点运算次数 (Floating Point Operations, FLOPs) 来评价计算效率,公式表示为

FLOPs = $\sum_{l=1}^D M_l^2 \times K_l^2 \times C_{l-1} \times C_l$ (17)

式中:D, M_l, K_l, C_{l-1} 和 C_l 分别代表网络总层数、第一层输出特征图尺寸、卷积核尺寸、第 l-1 层与 l 层通道数。

2.3 实验环境及训练参数

本次实验环境 GPU 为 RTX A4000,显存 16 GB;CPU 为 12 核 Intel (R) Xeon (R) Gold 5320;内存为 32 GB;操作系统为 Ubuntu 20. 04。深度学习环境配置为 TensorFlow v2. 5+ Cuda v11. 2 + cuDNN v8. 1. 1。

为公平比较,实验训练过程统一采用 SGD 算法进行优化,初始学习率设置为 0. 05,并采用 cos 退火算法进行学习率衰退,学习率最低可降低到 5 × 10⁻⁵,动量参数 momentum 设置为 0. 937,权值参数设置为 5 × 10⁻⁴,batch 设置为 32,训练 epochs 为 800,并取最好实验结果。

2.4 与其他目标检测算法对比

表 2 为本文改进算法与主流的一阶段检测算法 SSD、EfficientDet^[22]、YOLOv3^[23]、YOLOv4、YOLOv5-x 的比较,并结合图 7 进行部分测试集的可视化对比。

表 2 SAR 飞机识别网络综合性能比较
Tab. 2 Comprehensive performance comparison of SAR aircraft detection networks

算法	召回率/%	精确率/%	AP _{0.5} /%	浮点运算次数/10 ⁹	参数文件大小/MB
SSD300	85. 01	86. 99	88. 42	23. 798	91
EfficientDet-D0	82. 46	90. 53	89. 64	5. 125	16
YOLOv3	87. 00	84. 76	87. 16	19. 164	235
YOLOv4	88. 52	83. 53	88. 72	17. 493	245
YOLOv5-x	90. 88	85. 23	90. 46	26. 831	334
本文	91. 01	90. 09	92. 34	17. 497	256

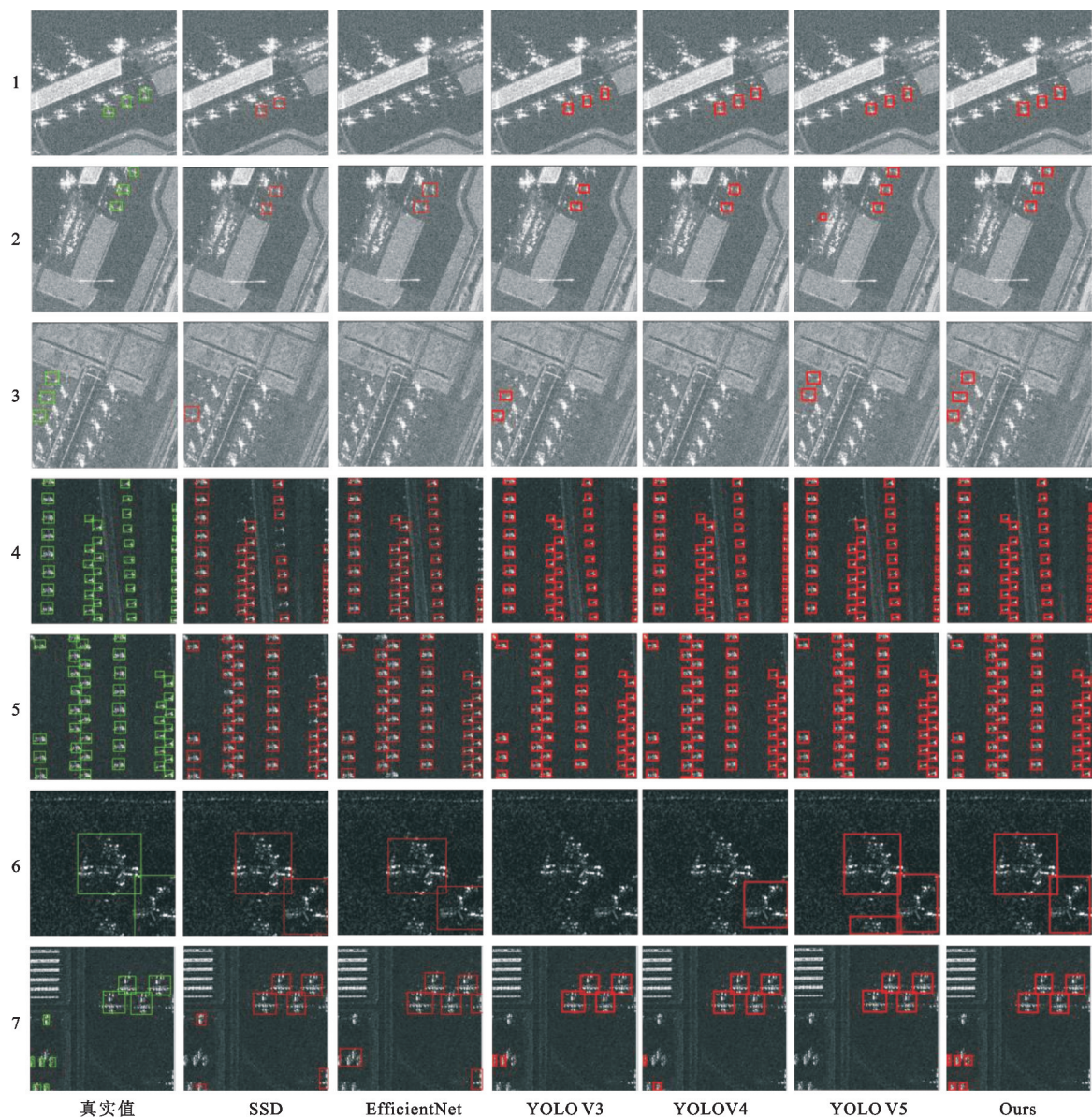


图 7 不同算法结果可视化对比

Fig. 7 Visual comparison of the results of different algorithms

通过对表 2 结果进行分析可以看出,本文的改进算法与原 YOLOv4 算法相比,引入 CA 和 ASFF 机制后网络层数增加,改进算法的 FLOPs 和参数文件大小分别增加了 4×10^6 和 11 MB,但是召回率、精确率和 AP 值分别提高 2.49%、6.56% 和 3.62%,做到了在尽可能少地引入额外参数量的情况下实现精确率、召回率和 AP 值的提升。轻量化算法 EfficientDet 虽然有较高的精确率和较低算法复杂度,但是 82.46% 的召回率无法满足算法对于检测性能的总体需求。与 YOLO 系列进行比较,改进算法除了具有最高的召回率和精确率之外,在精确率上也有了显著的提升,达到了 90.09%,较次高的 YOLOv5-x 提升了 4.86%,改善了 YOLO 系列算法

在 SAR 飞机目标检测中存在的精度低下的问题。

图 7 中组 1、组 2 和组 3 展示了 6 种算法在强散射干扰图像中的检测效果。3 组图像中,飞机呈现出以小尺寸规则分布于建筑物一侧的特点,而飞机目标附近航站楼等建筑物存在的强散射点会弱化飞机头部、机翼等部件的散射。此外,部分地物散射点也会有与飞机相近的散射特征,这使得算法容易产生漏检和虚检。如组 2 中航站楼的某处强散射点被 YOLOv5 判定为飞机,组 3 中除改进算法外其他算法都有不同程度的漏检。前 3 组图表明在背景地物强散射点干扰环境下,改进算法的检测表现要明显优于其他算法。

组 4 和组 5 中飞机呈现出小尺寸、紧密排列的

特征,同时结合组 1、组 2、组 3 中小目标的检测表现,发现 YOLO 系列对于小目标具有更好的检测性能。但是对于图 6 这种大尺寸目标, YOLOv3 和 YOLOv4 的检测效果却不如 SSD 和 EfficientDet 的检测效果。组 7 展示了不同大小的飞机目标出现在同一张图片中的检测效果,各种算法对于小尺寸飞机目标都有不同程度的漏检,其中 YOLOv5 未检测出左侧的小目标,而 EfficientDet 将多个小目标视作一个整体进行检测,改进算法则仅漏检一个目标。组 4~7 的实验表明,改进算法对于多尺寸的 SAR 图像飞机目标检测具有更显著的效果,说明针对 SAR 图像飞机尺寸多的特点引入 CA 和 ASFF 机制是有效的。

2.5 消融实验结果分析

为证明算法各种改进的有效性,在 YOLOv4 的基础上对不同改进及改进组合进行消融实验,结果如表 3 所示。

表 3 消融实验对比
Tab. 3 Ablation experiments

序号	实验	召回率/%	精确率/%	AP _{0.5} /%
1	YOLOv4	88.52	83.53	88.72
2	YOLOv4+k-means	88.89	85.93	88.87
3	YOLOv4+k-means+CA	89.40	88.62	90.41
4	YOLOv4+k-means+BAM	89.84	86.61	89.37
5	YOLOv4+k-means+GAM	87.28	87.86	90.11
6	YOLOv4+k-means+ASFF	87.06	87.57	89.23
7	YOLOv4+k-means+CA+ASFF	91.01	90.09	92.34

实验 2 对比实验 1,精确率提高 2.40%,表明改进 K-means 聚类方法获得的先验框能够更好覆盖不同尺寸的飞机目标,提高定位精度。

为了验证 CA 机制相较于其他注意力机制的优势,实验 3、实验 4 以及实验 5 在实验 2 的基础上比较了 CA 机制与常用的 BAM^[1,24] 以及 GAM^[25-26] 的性能。实验显示 3 种注意力机制都有助于提升目标检测算法的检测性能。引入 CA 和 BAM 两种机制较实验 2 在召回率上分别提高了 0.51% 和 0.95%,而引入 GAM 后召回率反而降低了 1.61%。对于精确率和 AP_{0.5} 值,引入 3 种注意力机制后相较于实验 2 都有所提升,其中实验 3 引入 CA 机制提升最高,精确率提高 2.69%,AP_{0.5} 提高 1.54%。结合图 3 的热力图,表明 CA 机制较 BAM 和 GAM 能够更有效地提高检测算法性能,增强网络对于飞机结构的提取能力。

实验 6 虽然在结果上较实验 2 召回率降低了 1.83%,但是精确率和 AP_{0.5} 值分别提高了 1.64% 和 0.36%,召回率与精确率相近,实验结果表明 ASFF 机制对于实验模型的精确率和召回率起到平衡作用。

实验 7 在综合改进机制后,召回率、精确率和 AP_{0.5} 达到了 91.01%,90.09% 和 92.34%,相较实验 1 分别提高 2.49%,6.56% 和 3.62%,改进效果显著。实验表明 CA 和 ASFF 机制结合改进算法能够满足 SAR 图像飞机目标检测的任务的需要。

3 结束语

本文着眼于 SAR 图像飞机检测中存在的目标尺度多样性和背景强散射干扰的问题,提出了一种基于注意力和自适应特征融合的 YOLOV4 SAR 图像目标检测算法 ASFF,能够缓解精确率和召回率之间不平衡的问题。此外,为了确保先验框能够更好地检测目标,本文利用基于交并比距离改进的 K-means 算法对数据集的真实框进行重新聚合获取先验框。与其他算法的对比结果表明,本文的改进算法满足了在不同条件下对飞机目标进行识别的需求。随后进行的消融实验进一步证明了本文算法各种改进的有效性。未来的工作将重点优化网络结构,实现高检测效率、低算法资源消耗。

参考文献:

[1] 赵琰,赵凌君,匡纲要. 基于注意力机制特征融合网络的 SAR 图像飞机目标快速检测[J]. 电子学报, 2021,49(9):1665-1674.

[2] LI W,ZOU B,XIN Y,et al. An improved CFAR scheme for man-made target detection in high resolution SAR images[C]//Proceedings of 2018 IEEE International Geoscience and Remote Sensing Symposium. Valencia: IEEE,2018:2829-2832.

[3] CHEN J,ZHANG B,WANG C. Backscattering feature analysis and recognition of civilian aircraft in TerraSAR-X images[J]. IEEE Geoscience and Remote Sensing Letters,2017,12(4):796-800.

[4] PAPPAS O,ACHIM A,BULL D. Superpixel-level CFAR detectors for ship detection in SAR imagery[J]. IEEE Geoscience and Remote Sensing Letters,2018,15(9):1397-1401.

[5] FU K,DOU F Z,LI H C,et al. Aircraft recognition in SAR images based on scattering structure feature and template matching[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2018,11(11):4206-4217.

[6] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus: IEEE, 2014: 580-587.

[7] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [C]//Proceedings of the 28th International Conference on Neural Information Processing Systems. Cambridge: MIT Press, 2015: 91-99.

[8] CAI Z, VASCONCELOS N. Cascade R-CNN: delving into high quality object detection[C]//Proceedings of 2018 IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 6154-6162.

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

[10] 王思雨,高鑫,孙皓,等. 基于卷积神经网络的高分辨率 SAR 图像飞机目标检测方法[J]. 雷达学报, 2017, 6(2): 195-203.

[11] 李广帅,苏娟,李义红. 基于改进 Faster R-CNN 的 SAR 图像飞机检测算法[J]. 北京航空航天大学学报, 2021, 47(1): 159-168.

[12] 赵琰,赵凌君,匡纲要. 复杂环境大场景 SAR 图像飞机目标快速检测[J]. 电波科学学报, 2020, 35(4): 594-602.

[13] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. Yolov4: optimal speed and accuracy of object detection [EB/OL]. (2020-04-23) [2022-10-14]. <https://arxiv.org/abs/2004.10934>.

[14] 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, 2014, 37(9): 1904-1916.

[15] 陈洋,张明,杨立东,等. 一种基于改进 YOLOv4 的 SAR 舰船检测算法[J]. 电子测量技术, 2022, 45(11): 120-125.

[16] HOU Q, ZHOU D, FENG J. Coordinate attention for efficient mobile network design[C]//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021: 13713-13722.

[17] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation[C]//Proceedings of 2018 IEEE

Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 8759-8768.

[18] LIU S, HUANG D, WANG Y. Learning spatial fusion for single-shot object detection [EB/OL]. (2019-11-25) [2022-10-14]. <https://arxiv.org/abs/1911.09516>.

[19] ZHANG P, XU H, TIAN T, et al. SEFEPNet: scale expansion and feature enhancement pyramid network for SAR aircraft detection with small sample dataset[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2022, 15: 3365-3375.

[20] EVERINGHAM M, VAN G L, WILLIAMS C K I, et al. The pascal visual object classes (VOC) challenge[J]. International Journal of Computer Vision, 2010, 88(2): 303-338.

[21] 李志刚,张娜. 一种轻量型 YOLOv5 交通标志识别方法[J]. 电讯技术, 2022, 62(9): 1201-1206.

[22] TAN M, PANG R, LE Q V. Efficientdet: Scalable and efficient object detection [C]//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020: 10781-10790.

[23] REDMON J, FARHADI A. Yolov3: an incremental improvement[EB/OL]. (2018-04-08) [2022-10-14]. <https://arxiv.org/abs/1804.02767>.

[24] PARK J, WOO S, LEE J Y, et al. BAM: bottleneck attention module[EB/OL]. (2018-07-18) [2022-12-16]. <https://arxiv.org/abs/1807.06514>.

[25] LIU Y, SHAO Z, HOFFMANN N. Global attention mechanism: retain information to enhance channel-spatial interactions [EB/OL]. (2021-12-10) [2022-12-16]. <https://arxiv.org/abs/1807.06514>.

[26] 李佳芯,朱卫纲,杨莹,等. 基于改进 YOLOv5 的 SAR 图像飞机目标检测[J]. 光电与控制, 2023, 30(8): 61-67.

作者简介:

夏一帆 男,1998 年生于山东日照,硕士研究生,主要研究方向为 SAR 图像目标检测。

赵凤军 男,1963 年生于河南项城,博士,研究员、博士生导师,主要从事雷达系统和射频微波技术方面的研究。

王樱洁 女,1992 年生于河南邓州,博士,助理研究员,主要从事星载合成孔径雷达干涉系统设计、时序干涉 SAR 技术应用等方面的研究。

王春乐 女,1986 年生于吉林永吉,博士,助理研究员,主要从事雷达图像特征提取、目标检测与识别技术方面的研究。