计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
2024,33(11):111−120 [doi: 10.15888/j.cnki.csa.009699] [CSTR: 32024.14.csa.009699]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

基于注意力与量化感知的航拍红外目标检测①

周　进[1], 裴晓芳[1,2]

[1](南京信息工程大学 电子与信息工程学院, 南京 210044)
[2](无锡学院 电子信息工程学院, 无锡 214105)
通信作者: 裴晓芳, E-mail: xiaofangpei@163.com

摘　要: 针对航拍场景下红外目标对比度低、识别精度差、检测难度大等问题, 提出一种基于注意力与量化感知的航拍红外目标检测算法. 首先, 利用 DCNv2 替代 ELAN 模块中的 3×3 卷积, 构建了 DC-ELAN 模块, 有效提升了模型捕捉局部和全局特征的能力, 进而强化了网络的特征表达能力; 其次, 通过巧妙地将 SE 注意力机制融入 SPPCSPC 模块和 ELAN 模块中, 设计出了 SE-SPPCSPC 模块和 SE-ELAN 模块, 有助于增强特征图的空间自注意力, 模型能够更好地关注目标区域; 此外, 引入 QARepVGG 模块, 提升模型的量化感知能力并增强其对量化误差的鲁棒性; 最后, 引入 DyHead 模块, 该模块可以根据输入图像的不同动态调整检测头, 提高模型对不同大小、形状目标的检测能力, 从而进一步提高红外目标检测的准确性和鲁棒性. 实验结果表明, 相较于原模型, 改进后的 YOLOv7-tiny 模型在计算量未增长的情况下, *mAP*@0.5 值提升了 3.4%, *mAP*@0.5:0.95 值提升了 4.8%, 显著提高了模型检测精度.

关键词: 红外目标检测; 可变形卷积; 注意力机制; 量化感知训练; 目标检测头

引用格式: 周进,裴晓芳.基于注意力与量化感知的航拍红外目标检测.计算机系统应用,2024,33(11):111–120. http://www.c-s-a.org.cn/1003-3254/9699.html

Aerial Infrared Target Detection Based on Attention and Quantization Awareness

ZHOU Jin[1], PEI Xiao-Fang[1,2]

[1](School of Electronics and Information Engineering, Nanjing University of Information Science & Technology, Nanjing 210044, China)
[2](School of Electronic and Information Engineering, Wuxi University, Wuxi 214105, China)

Abstract: Aiming at the problems of low contrast, poor recognition accuracy, and difficult detection of infrared targets in aerial scenes, this study proposes an aerial infrared target detection algorithm based on attention and quantization awareness. Firstly, the DC-ELAN module is constructed by using DCNv2 to replace the 3×3 convolution in the ELAN module, which effectively improves the ability of the model to capture local and global features, and then strengthens the feature representation ability of the network. Secondly, by cleverly integrating the SE attention mechanism into the SPPCSPC module and the ELAN module, the SE-SPPCSPC module and the SE-ELAN module are designed, which helps to enhance the spatial self-attention of the feature map, and the model can better focus on target areas. In addition, the QARepVGG module is introduced to improve the quantization awareness of the model and enhance its robustness to quantization errors. Finally, the DyHead module is introduced, which can dynamically adjust the detection head according to different input images, improve the detection ability of the model to targets of different sizes and shapes, and further improve the accuracy and robustness of infrared target detection. Experimental results show that compared with the

① 基金项目: 国家自然科学基金青年项目 (42205078); 苏高教会“高质量公共课教学改革研究”专项课题 (2022JDKT138); 高校哲学社会科学研究一般项目 (2022SJYB0979); 江苏职业教育研究立项课题一般项目 (XHYBLX2023282); 无锡学院教改课题 (XYJG2023002, XYJG2023023); 2023 江苏省大学生创新创业训练计划 (202313982007Z)
收稿时间: 2024-04-28; 修改时间: 2024-06-28; 采用时间: 2024-07-04; csa 在线出版时间: 2024-09-24
CNKI 网络首发时间: 2024-09-25

original model, the improved YOLOv7-tiny model has 3.4% and 4.8% increases in *mAP*@0.5 and *mAP*@0.5:0.95 values without increasing the amount of calculation, which significantly improves model detection accuracy.

Key words: infrared target detection; deformable convolution; attention mechanism; quantization-aware training; target detection head

随着无人机技术和航拍设备的快速发展, 航拍红外图像作为一种重要的信息获取方式, 已逐渐广泛应用于军事、安防、应急救援等领域. 航拍红外图像能够提供独特的信息和视角, 对于目标检测和识别具有重要的意义. 与可见光图像相比, 首先, 红外图像的对比度较低[1], 目标与背景的区分度往往较小, 这使得在红外图像中准确检测目标变得尤为困难; 其次, 红外图像中的噪声干扰较大, 这进一步降低了目标检测的准确性; 此外, 航拍场景下的红外图像还受到拍摄角度、拍摄高度、气象条件以及遮挡等多种因素的影响, 使得目标检测更加困难. 因此, 研究开发高效准确的红外目标检测算法对于充分利用航拍红外图像的优势具有重要意义.

在过去的几年中, 深度学习技术的快速发展为图像处理和计算机视觉领域带来了巨大的进步. 目标检测作为图像处理的一个重要分支, 也取得了显著的进展. 在 2012 年, AlexNet[2]在 ImageNet 图像分类任务中取得了突破性成果, 开启了深度学习在目标检测领域的应用[3]. 随后, 一些早期的目标检测算法, 如 R-CNN[4]、Fast R-CNN[5]和 Faster R-CNN 等, 逐渐发展起来. 这些算法主要基于卷积神经网络 (convolutional neural network, CNN)[6]和区域生成网络 (region proposal network, RPN), 通过提取候选区域的特征并使用分类器进行分类, 实现了较高的目标检测精度. 随着 YOLO (you only look once) 系列[7–9]的推出, 深度学习目标检测技术取得了重要突破. 此外, 还有许多其他具有代表性的目标检测网络模型, 如 SSD (single shot multibox detector)[10]、RetinaNet[11]和 DETR (detection Transformer)[12]等, 这些算法都在不同领域发挥着各自的优势和特点.

针对航拍场景下的红外目标检测问题, 越来越多的研究者投入到相关研究工作中. Gong 等人[13]针对红外图像检测速度难以达到实时性要求的问题, 提出了一种基于轻量型检测网络 YOLOv3-tiny 的红外图像检测算法, 为了提升检测精度, 他们不仅优化了网络结构以加深其层次, 还巧妙地采用了 K-means 聚类算法对先验框进行了精确调整, 尽管如此, 由于当前检测网络的深度仍有所欠缺, 其检测精度尚存提升空间. Zhu 等人[14]成功地集成了 Transformer 预测头和卷积注意力模块 (convolutional block attention module, CBAM)[15]至 YOLOv5 中, 尽管这一举措显著提升了模型检测效果, 但随之而来的问题是模型参数规模庞大, 导致网络推理速度降低. Hu 等人[16]巧妙地结合各种运动形态的目标模式, 成功地融合了不同维度的红外目标特征, 这确实在增强目标特征的同时抑制了背景噪声, 但由于该方法涉及不同维度的特征融合, 需要大量的计算资源来进行特征提取和融合, 这可能导致在实际应用中, 特别是在需要实时处理的场景中, 难以满足性能要求.

本文在前人研究的基础上, 聚焦于航拍场景下红外目标检测, 以 YOLOv7-tiny[17]网络模型为基准, 提出一种新的网络结构, 具体改进如下.

(1) 引入可变形卷积第 2 版 (deformable ConvNets version 2, DCNv2)[18], 提出将 DCNv2 应用于 YOLOv7-tiny 骨干网络的 ELAN (efficient layer aggregation network) 模块中, 将其中的 3×3 卷积替换为 DCNv2, 构建 DC-ELAN 模块, 可以有效地捕获局部和全局特征, 增强网络的特征表达能力, 从而提高红外目标检测的准确性.

(2) 对 YOLOv7-tiny 网络模型的 Neck 部分进行改进, 在 SPPCSPC 模块和 ELAN 模块中融入 SE (squeeze-and-excitation) 注意力机制[19], 提出 SE-SPPCSP 模块和 SE-ELAN 模块, 有助于增强特征图的空间自注意力, 使其能够更好地关注到目标区域, 从而显著提升目标检测的精度.

(3) 在 Neck 部分融入 QARepVGG[20]模块, 能够提升模型的量化感知能力, 增强对量化误差的鲁棒性, 从而优化目标检测的性能.

(4) 对模型的 Head 部分进行改进, 引入 DyHead (dynamic head)[21]模块, 可以根据输入图像的不同动态调整检测头, 提高模型对不同大小和形状目标的检测能力, 从而增强模型的泛化性能.

1 本文方法

1.1 概述

YOLOv7-tiny 是 YOLO 系列目标检测算法中的一个轻量级版本, 它是在 YOLOv7 的基础上进行了一些改进, 在保持较高检测精度的同时, 减小了模型大小和计算复杂度, 适用于资源受限的场景.

YOLOv7-tiny 网络模型主要由 4 个部分组成, 分别是输入部分、Backbone 部分、Neck 部分和 Head 部分. 输入部分首先对输入图像进行预处理操作, 例如调整大小以适应模型的输入要求, 并进行归一化处理[22]; Backbone 部分作为特征提取的核心部分, 采用了轻量级的设计, 它由一系列的卷积层、池化层等构成, 通过优化网络结构和参数数量来减小计算负担; Neck 部分负责融合 Backbone 网络提取的特征, 并生成用于目标检测的特征图, 采用了一些特殊的结构, 如空间金字塔池化, 以增强特征的表达能力, 通过 Neck 部分的处理, 模型能够捕获多尺度的目标信息; Head 部分基于 Neck 部分输出的特征图进行目标检测和分类, 包含一组卷积层和全连接层, 用于回归目标边界框的坐标和分类目标的类别, 通过回归和分类的损失函数进行联合训练, 模型能够同时优化定位和分类的性能.

1.2 改进的 YOLOv7-tiny 网络模型

本文在 YOLOv7-tiny 网络模型的基础上, 对其进行改进, 在 Backbone 部分提出新的 DC-ELAN 模块, 增强网络的特征表达能力; 在 Neck 部分提出 SE-SPPCSPC 模块和 SE-ELAN 模块, 以及引入 QARepVGG 模块, 提高红外目标检测精度; 在 Head 部分使用带有自注意力检测的 DyHead 检测头, 增强模型的泛化性能, 改进后的 YOLOv7-tiny 网络结构图如图 1 所示.

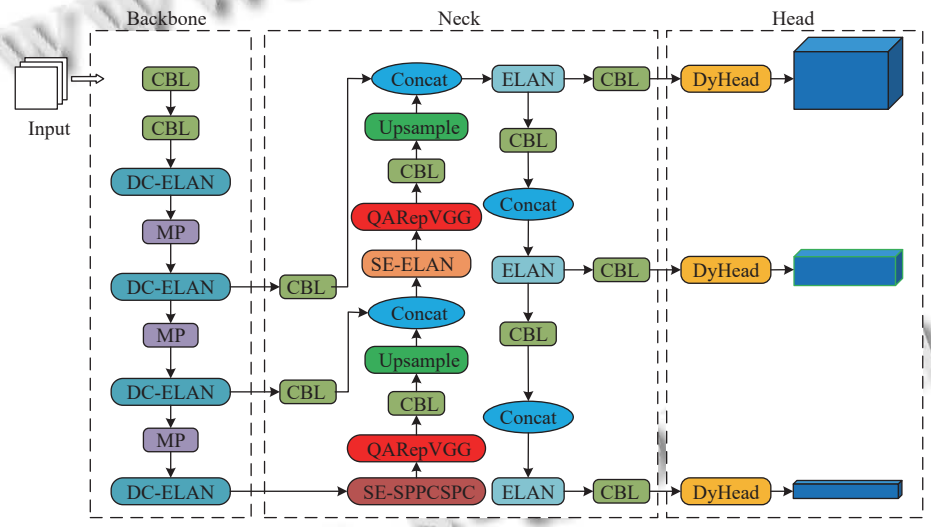


图 1 改进的 YOLOv7-tiny 网络结构图

1.3 可变形卷积替换标准卷积

在航拍场景下, 由于目标往往存在较大的形变, 采用标准卷积在处理这种场景时往往表现不佳. 为了解决这个问题, 本文提出采用 DCNv2 替换标准卷积, 以适应不同形状和大小的目标.

DCNv2 是一种创新性的卷积方式, 它通过学习目标物体的偏移量, 能够自适应地调整卷积核的位置. DCNv2 的实现过程可以分为两个阶段, 如图 2 所示. 通过一个卷积层来学习目标物体的偏移量, 在这一阶段, 网络接收输入图像并使用 DCNv2 来提取特征, 每个像素点位置上的卷积核都可以根据学习到的偏移量进行移动, 使得网络能够逐渐适应不同目标的形状和大小, 一旦学习到了目标物体的偏移量, 这些偏移量将被用于调整卷积核的位置. 在第 2 个阶段, 根据学习到的偏移量将卷积核移动到正确的位置上进行卷积操作, 它能够自适应地调整卷积核的位置, 从而更好地适应不同形状、大小的目标物体.

YOLOv7-tiny 网络模型的 ELAN 模块使用标准卷积来进行特征提取, ELAN 模块结构图如图 3 所示. 本文提出新的 DC-ELAN 模块, 提高网络的特征提取能力, DC-ELAN 模块结构图如图 4 所示.

1.4 SE 注意力机制模块

在航拍场景下, 由于红外图像的特殊性, 目标往往存在于图像的局部区域. SE 注意力机制是一种通道类

型的注意力机制, 它在通道维度上增加注意力机制, 通过挤压和激励操作建模通道间的相互依赖性来提取更丰富的特征信息, 提高网络表达的质量, SE 注意力机制结构图[19]如图 5 所示.

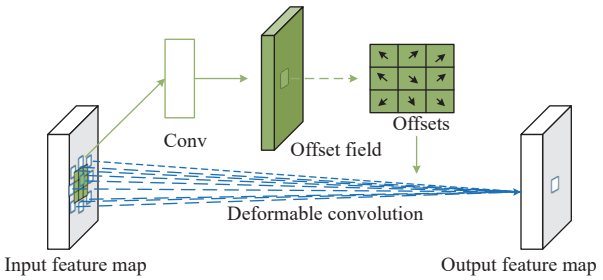


图 2 DCNv2 的实现过程

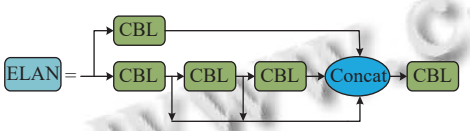


图 3 ELAN 模块结构图

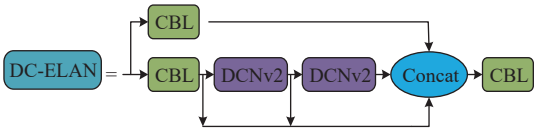


图 4 DC-ELAN 模块结构图

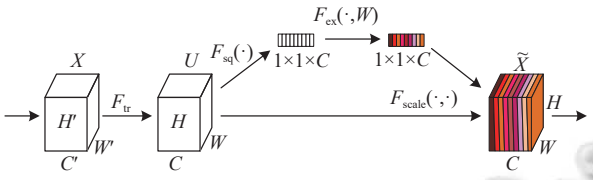


图 5 SE 注意力机制结构图

图 5 中, X为原始输入特征图, W'为原始输入宽度, H'为原始输入高度, C'为原始输入通道数, U为经过F_{tr} (transformation) 操作后的特征图, W为操作后的宽度, H为操作后的高度, C为操作后的通道数.

在挤压步骤中, SE 注意力机制使用全局平均池化操作将输入特征图压缩成一个$1\times1\times C$的特征向量, 从而提取出输入特征的全局信息. 挤压步骤计算所用数学公式如下所示:

$$z_c = F_{\text{sq}}(u_c) = \frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W} u_c(i,j) \tag{1}$$

其中, u_c为二维矩阵, $u_c(i,j)$为第i行第j列元素的特征向量.

在激励步骤中, 经过全局平均池化操作后, 将得到的$1\times1\times C$的特征向量通过一个全连接层, 将其映射到一个较小的向量, 然后再通过另一个全连接层将其映射回原始的维度. 在这个过程中, 使用 Sigmoid 函数将每个元素压缩到 0–1 之间, 并将其与原始输入特征图相乘, 得到加权后的特征图. 激励步骤计算所用数学公式如下所示:

$$s = F_{\text{ex}}(z,W) = \sigma(g(z,W)) = \sigma(W_2\delta(W_1 z)) \tag{2}$$

其中, $W_1 z$为第 1 个全连接操作, $\delta(\cdot)$为 ReLU 激活函数, $W_2\delta(\cdot)$为第 2 个全连接操作, $\sigma(\cdot)$为 Sigmoid 函数.

接着通过 Sigmoid 函数将每个通道的权重归一化到[0, 1]范围内, 然后通过一个 Scale 的操作将归一化后的权重加权到每个通道的特征上[23]. Scale 操作计算所用数学公式如下所示:

$$\tilde{X}_c = F_{\text{scale}}(u_c, s_c) = s_c u_c \tag{3}$$

其中, s_c为各通道重要程度的权重.

本文将 SE 注意力机制融入 Neck 部分的 SPPCSPC 模块和 ELAN 模块中, SPPCSPC 模块结构图如图 6 所示, 提出 SE-SPPCSPC 模块和 SE-ELAN 模块. SE-SPPCSPC 模块结构图如图 7 所示, SE-ELAN 模块结构图如图 8 所示.

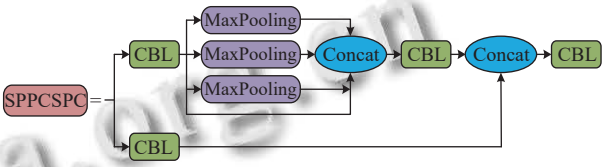


图 6 SPPCSPC 模块结构图

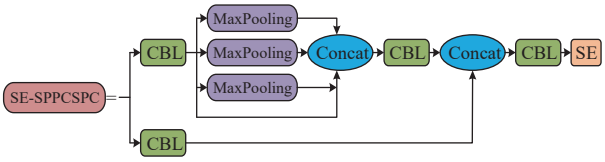


图 7 SE-SPPCSPC 模块结构图

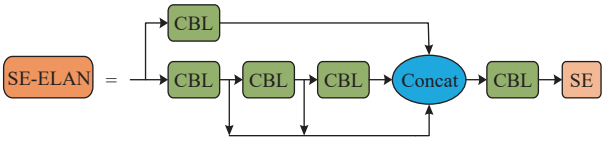


图 8 SE-ELAN 模块结构图

1.5 QARepVGG 模块

在航拍场景下的红外目标检测任务中, 检测的目标

是从复杂的航拍图像中准确地检测出红外目标. 为了实现这一目标, 本文采用量化感知训练 (quantization-aware training) 方法, 将量化过程融入 RepVGG 网络的训练中, 形成 QARepVGG 模块, RepVGG 模块与 QARepVGG 模块结构对比图[20]如图 9 所示.

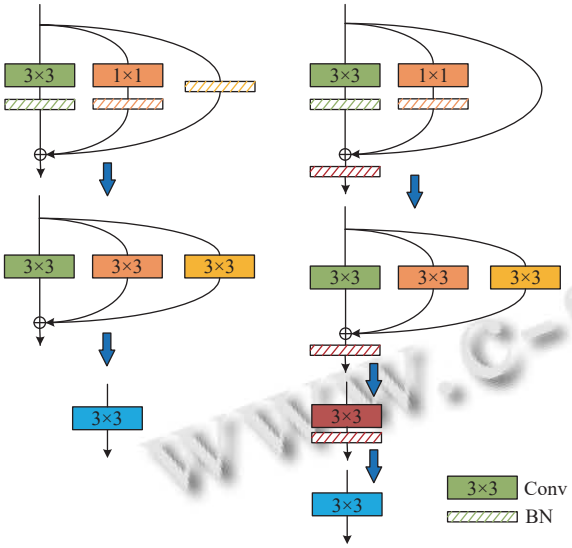


图 9 RepVGG 模块与 QARepVGG 模块结构对比图

QARepVGG 模块相对于 RepVGG 模块进行了多方面的改进, 这些改进主要集中在权重和激活分布的差异、优化设计和量化友好特性上. 具体来说, QARepVGG 在权重和激活分布方面保持了与 RepVGG 的基本差异, 这种差异使得它能够更好地适应不同的输入特征图的大小和形状, 从而减少不必要的计算和内存消耗; 优化设计使得 QARepVGG 可以方便地集成到现有的神经网络架构中, 并能够提供出色的后量化性能, 同时减少了对计算资源和内存的消耗; 此外, QARepVGG 还具有量化友好特性, 使得它在处理大量数据时更加高效和可靠, 同时减少了对计算资源和内存的消耗. 这些改进让 QARepVGG 具有更高效的计算能力, 更好的泛化性能以及更低的内存消耗, 从而全面提升模型的整体性能.

1.6 DyHead 检测头模块

红外目标在航拍图像中往往呈现出较小的尺寸和较低的对比度, 为了解决这些问题, 本文采用 DyHead 作为新的目标检测头, 它通过将尺度感知注意力模块、空间感知注意力模块和任务感知注意力模块统一在一个框架中, 提高目标检测的性能, DyHead 模块结构图如图 10 所示.

在检测层上给定特征张量$F \in R^{L\times S\times C}$, 将注意力函数转换为 3 个连续的注意力, 每个注意力只关注一个角度, 该注意力函数计算所用数学公式如下所示:

$$W(F)=\pi_C(\pi_S(\pi_L(F)\cdot F)\cdot F)\cdot F \tag{4}$$

其中, $\pi_L(\cdot)$、$\pi_S(\cdot)$、$\pi_C(\cdot)$分别是适用于维度 L、S 和 C 的 3 个不同的注意力函数.

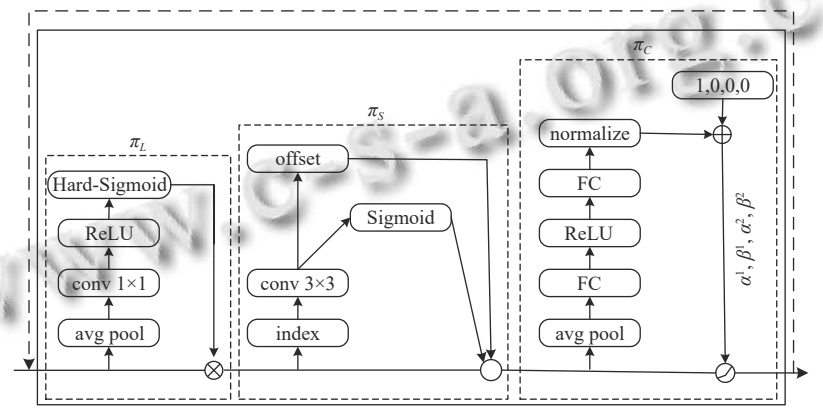


图 10 DyHead 模块结构图

(1) 尺度感知注意力模块$\pi_L(\cdot)$

该模块主要解决目标尺寸变化对检测的影响, 由于红外目标在图像中的尺寸通常较小, 尺度感知模块通过并行计算多个尺度感受野, 从不同尺度上提取特征信息, 即使目标尺寸变化, 也可以在多个尺度上进行有效的特征提取, 提高检测的准确性. 尺度感知注意力模块计算所用数学公式如下所示:

$$\pi_L(F)\cdot F=\sigma\left(f\left(\frac{1}{SC}\sum_{S,C}F\right)\right)\cdot F \tag{5}$$

其中, $f(\cdot)$函数是一个由 1×1 卷积层近似模拟的线性函数, 它的输入是特征图, 输出是经过线性变换后的特

征图; $\sigma(x)$函数是 Hard-Sigmoid 函数, 它的输入是一个实数, 输出是一个 0–1 之间的实数.

(2) 空间感知注意力模块$\pi_S(\cdot)$

该模块关注于目标在图像中的空间位置变化对检测的影响, 由于红外目标的位置往往不固定, 空间感知模块利用空洞卷积和注意力机制, 将上下文信息融入特征提取过程中, 即使目标位置发生变化, 也可以有效地进行检测, 提高鲁棒性. 空间感知注意力模块计算所用数学公式如下所示:

$$\pi_S(F)\cdot F=\frac{1}{L}\sum_{l=1}^{L}\sum_{k=1}^{K}\omega_{l,k}\cdot F(l;p_k+\Delta p_k;c)\cdot\Delta m_k \tag{6}$$

其中, K 为稀疏采样位置的数量, $p_k+\Delta p_k$是由自学习空间偏移量Δp_k转移的位置, 聚焦于一个辨别区域, Δm_k是位置p_k上的自学习重要标量.

(3) 任务感知注意力模块$\pi_C(\cdot)$

该模块将任务的先验知识融入目标检测过程中, 通过引入一个轻量级的神经网络, 任务感知模块将先验知识和特征提取过程进行有机结合, 可以针对不同的任务需求, 自适应地调整特征提取的侧重点, 提高目标检测的性能. 空间感知注意力模块计算所用公式如下所示:

$$\pi_C(F)\cdot F=\max\left(\alpha^1(F)\cdot F_c+\beta^1(F),\alpha^2(F)\cdot F_c+\beta^2(F)\right) \tag{7}$$

其中, F_c是第 c 个通道的特征切片, $\left[\alpha^1,\beta^1,\alpha^2,\beta^2\right]^{\mathrm{T}}=\theta(\cdot)$是学习控制激活阈值的超函数.

2 实验结果与分析

2.1 数据集处理

实验数据集来自 infiRay 公司提供的开源数据库中的红外航拍人车检测数据集, 该数据集从不同的视角高度采集了 11 045 张红外图像, 包含不同场景下不同尺度的人、小汽车、公交车、自行车、骑自行车的人、卡车等 6 类红外目标, 本文按照 7:1:2 的比例划分为训练集、验证集、测试集, 分别有 7 731、1 105、2 209 张图像.

2.2 实验环境及参数配置

为了保证实验的公平性, 本研究中涉及的所有实验均在相同的实验配置和训练参数下进行, 实验配置如表 1 所示, 训练参数如表 2 所示.

表 1 实验配置

配置名称	配置信息
操作系统	Windows 11
CPU	Intel(R) Core I7-13700KF
GPU	NVIDIA GeForce RTX 4070Ti
深度学习框架	PyTorch
CUDA	11.3
语言	Python 3.8.5

表 2 训练参数

参数名称	参数信息
学习率	0.01
动量参数	0.937
衰减系数	0.000 5
输入图像的尺寸	640×640
批量大小	16
训练周期	300

2.3 模型评价指标

(1) 平均精度均值 (mean average precision, *mAP*)

mAP 是衡量模型在多个类别上的平均性能, *mAP*@0.5 表示 IoU (intersection over union) 阈值设置为 0.5 时, 模型的平均精度均值. *mAP*@0.5:0.95 表示 IoU 阈值从 0.5 到 0.95 的范围内, 计算每个类别的 *AP*, 并取平均值. *mAP* 计算所用数学公式如下所示:

$$mAP=\frac{\sum AP}{Num(cls)} \tag{8}$$

(2) GFLOPs

GFLOPs 是每秒执行的 10 亿次浮点运算, 常作为衡量 GPU 性能的指标. 它表示模型或算法在 GPU 上每秒可以执行的浮点运算次数, 可以用来评估模型或算法的并行计算能力和计算效率.

2.4 实验与分析

2.4.1 ELAN 改进对比实验

为了更好地验证本文提出的 DC-ELAN 模块的有效性, 分别用 PConv、DSConv、ODConv、DySnakeConv 以及 DCNv2 这 5 种卷积替换 ELAN 模块中的 3×3 卷积, 基准模型采用标准卷积, 实验结果如表 3 所示.

表 3 ELAN 改进对比实验

Model	*mAP*@0.5 (%)	*mAP*@0.5:0.95 (%)	GFLOPs
Baseline	85.0	50.1	13.2
+PConv	84.1	48.9	10.1
+DSConv	84.8	50.1	9.4
+ODConv	84.9	49.6	9.5
+DySnakeConv	85.1	49.8	17.7
+DCNv2	85.9	50.9	10.9

由表 3 可以看出, 使用 PConv、DSConv、ODConv 后, 计算量虽显著减少, 但 *mAP*@0.5 值均出现不同程度的下降, 表明这些卷积方法虽在降低计算复杂度上有所成效, 但在全局特征提取或特征捕捉能力方面尚需进一步优化提升. 使用 DySnakeConv 后, *mAP*@0.5 值得到提升, 但计算量也随之增加, 表明其动态蛇形结构在处理图像时动态调整感受野的形状, 从而增强了对目标非规则形状的适应性, 但要实现灵活的感受野调整, 则需要更多计算资源, 导致参数和计算步骤增多, 计算量上升. 使用 DCNv2 后, *mAP*@0.5 值显著增加且计算量更低, 表明其引入的可学习卷积核变换使模型能动态调整参数, 这种动态性对大小、形状、姿态各异的目标具有强适应性, 同时, DCNv2 优化计算效率, 减少冗余, 实现高性能与低计算量的平衡, 因此 DC-ELAN 模块更适用于本实验提高检测精度.

2.4.2 注意力机制对比实验

为了更好地关注到目标区域, 在 SPPCSPC 模块和 ELAN 模块中分别融入 CBAM、BiFormer、CA、ECA、EMA、ACMix、SimAM 以及 SE 这 8 种注意力机制做性能对比实验, 基准模型未融入注意力机制, 实验结果如表 4 所示.

表 4 注意力机制对比实验

Model	*mAP*@0.5 (%)	*mAP*@0.5:0.95 (%)	GFLOPs
Baseline	85.0	50.1	13.2
+CBAM	85.2	50.2	13.3
+BiFormer	83.8	48.4	39.2
+CA	85.3	49.9	13.5
+ECA	84.7	50.4	13.2
+EMA	85.4	50.8	13.5
+ACMix	85.1	49.8	14.4
+SimAM	84.9	50.3	13.2
+SE	85.6	50.5	13.2

由表 4 可以看出, 融入 BiFormer、ECA、SimAM 注意力机制, 计算量没有降低, *mAP*@0.5 值却有所减少, 表明这些注意力机制引入了额外的复杂性, 导致计算量增加, 同时, 航拍红外图像目标小、分辨率和对比度低, 使得提取有效信息更具挑战, 这些注意力机制更关注全局或大目标特征, 对小目标处理不佳, 从而降低了 *mAP*@0.5 性能. 融入 CBAM、CA、EMA、ACMix 注意力机制, 虽然计算量略有增加, 但 *mAP*@0.5 值均有所提高, 表明它们能有效提升模型对关键信息的关注能力, 契合红外图像中目标小、背景复杂的特点, 注意力机制的引入虽然增加了模型的复杂度, 但同时也为模型提供了更多的信息来优化其决策过程. 融入 SE 注意力机制, 在计算量并未增加的情况下, *mAP*@0.5 值明显提升, 表明 SE 注意力机制能够自适应地调整特征通道, 使模型更加聚焦于关键信息, 通过挤压和激励操作, 有效地增强了与目标相关的特征通道, 抑制了无关通道, 从而提高检测性能.

2.4.3 QARepVGG 与 RepVGG 对比实验

为了更好地提升模型的量化感知能力, 增强对量化误差的鲁棒性, 在实验中分别融入 QARepVGG 模块和 RepVGG 模块, 实验结果如表 5 所示.

表 5 QARepVGG 与 RepVGG 对比实验

Model	*mAP*@0.5 (%)	*mAP*@0.5:0.95 (%)	GFLOPs
Baseline	85.0	50.1	13.2
+RepVGG	85.7	51.0	14.3
+QARepVGG	85.9	51.4	14.3

由表 5 可以看出, 融入 QARepVGG 和 RepVGG 模块的计算量相同, 但 QARepVGG 的 *mAP*@0.5 值提升更多, 表明 QARepVGG 专注于提升量化感知能力, 能有效处理量化误差, 具有更优的结构或设计特性, 使其在红外目标检测任务中表现更佳, 而 RepVGG 虽高效易训, 但在量化感知方面不如 QARepVGG 精细.

2.4.4 DyHead 检测头个数对比实验

为了更好地验证 DyHead 重复堆叠 4 次可以达到最好的性能检测, 本实验分别测试了堆叠 1、2、4、6、8 次该模块的性能表现, 0 次为基准模型, 实验结果如表 6 所示.

表 6 DyHead 堆叠个数对比实验

No.	*mAP*@0.5 (%)	*mAP*@0.5:0.95 (%)	GFLOPs
0	85.0	50.1	13.2
1	86.2	51.9	12.6
2	87.2	53.2	13.2
4	87.6	54.1	14.3
6	87.3	54.3	15.5
8	87.8	54.8	16.7

由表 6 可以看出, 相较于堆叠 1 次、2 次和 6 次, DyHead 堆叠 4 次时, *mAP*@0.5 值提升最为显著. 尽管堆叠 8 次时 *mAP*@0.5 值仍有所增加, 但计算量的显著增长却不利于模型的轻量化. 因此, 从性价比和检测精度的综合考虑来看, DyHead 堆叠 4 次显然是一个更优的选择, 能够更有效地提升检测性能.

2.5 模型对比实验

为了全面验证本文改进的 YOLOv7-tiny 模型在不同模型中的优越性, 在相同的实验环境和参数配置下, 将其与 RetinaNet、YOLOv3、YOLOv4s、YOLOv5s、

YOLOv6s、YOLOv7-tiny 以及 YOLOv8s 模型进行了对比实验, 实验结果如表 7 所示.

表 7 模型对比实验 (%)

Model	*mAP*@0.5	*mAP*@0.5:0.95
RetinaNet	82.3	49.1
YOLOv3	80.2	48.4
YOLOv4s	81.4	48.9
YOLOv5s	84.3	49.5
YOLOv6s	85.2	51.3
YOLOv7-tiny	85.0	50.1
YOLOv8s	86.5	51.9
Ours	88.4	54.9

根据表 7 中的详细数据对比, 本文改进后的 YOLOv7-tiny 模型在 *mAP*@0.5 指标上的卓越性能得以凸显. 在相同的实验环境和参数配置下, 该模型展现出了显著的 *mAP*@0.5 值提升, 相较于 RetinaNet、YOLOv3、YOLOv4s、YOLOv5s、YOLOv6s、原版的 YOLOv7-tiny 以及 YOLOv8s 等经典模型, 分别实现了 6.1%、8.2%、7.0%、4.1%、3.2%、3.4% 和 1.9% 的提升. 这一显著的性能提升不仅充分验证了本文模型改进策略的有效性, 更彰显出改进后的模型在红外目标任务检测精度上的显著进步.

2.6 消融实验

为了探究本实验模型改进的有效性, 将不同模块在数据集上检测效果做消融实验进行评估, 实验结果如表 8 所示.

表 8 消融实验

DCNv2	SE	QARepVGG	DyHead	*mAP*@0.5 (%)	*mAP*@0.5:0.95 (%)	GFLOPs
—	—	—	—	85.0	50.1	13.2
√	—	—	—	85.9	50.9	10.9
—	√	—	—	85.6	50.5	13.2
—	—	√	—	85.9	51.4	14.3
—	—	—	√	87.6	54.1	14.3
√	—	√	—	86.5	52.2	12.0
√	—	—	√	87.8	54.4	12.1
√	—	√	√	88.1	54.7	13.1
√	√	√	√	88.4	54.9	13.1

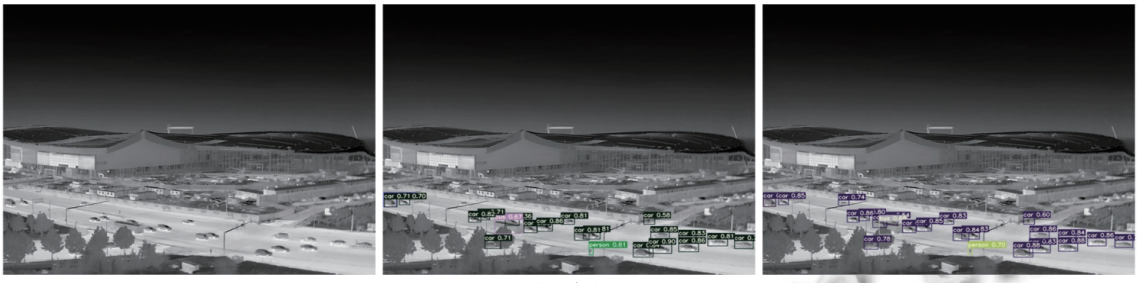
由表 8 可以看出, 通过与基准模型对比, 各模块的引入均带来了不同程度的性能提升. 增加 DCNv2 后, 不仅计算量得到有效降低, *mAP*@0.5 值还提升了 0.9%, 证明了 DC-ELAN 模块的轻量化优势, 以及其对于处理不同形状和大小目标检测任务的卓越适应性, 由于红外目标往往具有多样的形态和尺寸, DCNv2 的可变形卷积特性使其能够灵活调整卷积核的形状和大小, 从而更精确地捕捉目标的特征, 进而提升检测精度. SE 注意力机制的加入使得 *mAP*@0.5 值提升了 0.6%, SE 注意力机制能够自动学习和分配不同特征通道的权重, 使模型更加关注红外目标的特征, 通过强调关键特征并抑制不相关噪声, SE 机制显著提高了模型的特征提取能力, 从而提升了检测精度. 引入 QARepVGG 模块后, *mAP*@0.5 值提升了 0.9%, QARepVGG 的量化感知训练能力能够在模型量化过程中减少量化误差对性能的影响, 通过增强模型的量化鲁棒性, QARepVGG 优化了红外目标检测的性能, 使得模型在保持轻量级的同时依然保持出色的检测精度. DyHead 模块的加入使得 *mAP*@0.5 值显著提升了 2.6%, DyHead 的自注意力机制使得 DyHead 能够自适应地学习红外目标的特征, 从而更好地捕捉红外目标的外观和纹理特征, 通过提高模型对目标的识别能力, DyHead 显著提升了红外目标检测的准确性. 当同时增加 2 个或 3 个模块时, *mAP*@0.5 值有了更大的提升, 这证明了不同模块之间的协同作用能够进一步增强模型的性能. 特别是当 DCNv2、SE、QARepVGG 和 DyHead 这 4 个模块同时作用于基准模型时, 在计算量并未增加的情况下, 各项评价指标均得到了显著提升, 其中 *mAP*@0.5 值提升了 3.4%, 这一显著成果充分证实了本实验模型改进的有效性, 以及各模块在提高模型检测精度方面的卓越贡献.

2.7 检测结果

为了检测改进后的 YOLOv7-tiny 网络模型在真实航拍场景下红外目标检测效果, 选取小目标、密集、遮挡、模糊 4 类场景进行对比分析, 左侧一列是真实图片, 中间一列是 YOLOv7-tiny 模型检测结果, 右侧一列是改进后的 YOLOv7-tiny 模型检测结果, 检测结果对比如图 11 所示.

根据图 11 的展示, 可以清晰地观察到改进后的 YOLOv7-tiny 模型在不同航拍场景下的优越性能. 在小目标场景中, 由于目标拍摄距离较远, 识别难度增加, 但改进后的模型却能成功检测出更小的目标, 展现出了其卓越的细节捕捉能力. 在密集场景中, 这里目标分布较为集中, 改进后的模型凭借其出色的处理能力, 有效降低漏检概率, 确保了目标的全面捕获. 在遮挡场景中, 即便在目标被部分遮挡的情况下, 改进后的模型依然能够精准地识别出目标, 显示出了其强大的抗遮挡能力. 在模糊场景中的目标由于拍摄条件限制, 轮廓较

为模糊, 改进后的模型相较于原模型, 能够检测出更多的模糊目标, 显示出其在复杂环境下的稳健性和准确性. 综上所述, 本文改进后的 YOLOv7-tiny 模型在处理不同航拍场景下的红外目标检测难题时, 展现出了卓越的性能和精准度, 为红外目标检测提供了新的解决方案.



(a) 小目标场景

(b) 密集场景

(c) 遮挡场景

(d) 模糊场景

图 11 检测结果对比

3 结论

针对航拍场景下红外目标对比度低、识别精度差、检测难度高等一系列挑战, 本文提出了一系列改进方法. 其中, DC-ELAN 模块的引入, 使得模型能够更有效地捕获目标的局部和全局特征, 进而显著增强了网络的特征表达能力. 同时, 通过巧妙地在 SPPCSPC 模块和 ELAN 模块中融入 SE 注意力机制, 构建了 SE-SPPCSP 模块和 SE-ELAN 模块, 这些模块有助于增强特征图的空间自注意力, 使模型能够更加精准地聚焦于目标区域. 此外, 还引入了 QARepVGG 模块, 以提升模型的量化感知能力, 并有效增强其对量化误差的鲁棒性, 这一改进使得模型在应对复杂多变的航拍场景时, 能够保持稳定的性能. 最后, 通过引入 DyHead 模块, 实现了根据输入图像的不同动态调整检测头的功

能, 从而提高了模型对不同大小和形状目标的检测能力. 实验结果表明, 本文提出的方法在航拍场景下红外目标检测任务中展现出了较高的准确性和鲁棒性. 在未来的研究中可以进一步研究如何将其他先进的特征提取方法和深度学习技术应用于这一领域, 以进一步提升检测的准确性和鲁棒性, 为航拍场景下的红外目标检测任务提供更加可靠和高效的解决方案.

参考文献

1 明英, 蒋晶珏. 基于柯西分布的视频图像序列背景建模和运动目标检测. 光学学报, 2008, 28(3): 587–592. [doi: 10.3321/j.issn:0253-2239.2008.03.035]

2 Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Communications of the ACM, 2017, 60(6): 84–90. [doi: 10.1145/3065386]

3 张涛, 杨小冈, 卢孝强, 等. Dense RFB 和 LSTM 遥感图像舰船目标检测. 遥感学报, 2022, 26(9): 1859–1871.

4 Girshick R, Donahue J, Darrell T, *et al*. Rich feature hierarchies for accurate object detection and semantic segmentation. Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus: IEEE, 2014. 580–587.

5 Girshick R. Fast R-CNN. Proceedings of the 2015 IEEE International Conference on Computer Vision. Santiago: IEEE, 2015. 1440–1448.

6 LeCun Y, Bengio Y, Hinton G. Deep learning. Nature, 2015, 521(7553): 436–444. [doi: 10.1038/nature14539]

7 Redmon J, Farhadi A. YOLO9000: Better, faster, stronger. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017. 6517–6525.

8 Redmon J, Farhadi A. YOLOv3: An incremental improvement. arXiv:1804.02767, 2018.

9 Bochkovskiy A, Wang CY, Liao HYM. YOLOv4: Optimal speed and accuracy of object detection. arXiv:2004.10934, 2020.

10 Liu W, Anguelov D, Erhan D, *et al*. SSD: Single shot multibox detector. Proceedings of the 14th European Conference on Computer Vision. Amsterdam: Springer International Publishing, 2016. 21–37.

11 Lin TY, Goyal P, Girshick R, *et al*. Focal loss for dense object detection. Proceedings of the 2017 IEEE International Conference on Computer Vision. Venice: IEEE, 2017. 2999–3007.

12 Carion N, Massa F, Synnaeve G, *et al*. End-to-end object detection with Transformers. Proceedings of the 16th European Conference on Computer Vision. Glasgow: Springer International Publishing, 2020. 213–229.

13 Gong J, Zhao JH, Li F, *et al*. Vehicle detection in thermal images with an improved YOLOv3-tiny. Proceedings of the 2020 IEEE International Conference on Power, Intelligent Computing and Systems. Shenyang: IEEE, 2020. 253–256.

14 Zhu XK, Lyu SC, Wang X, *et al*. TPH-YOLOv5: Improved YOLOv5 based on Transformer prediction head for object detection on drone-captured scenarios. Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision Workshops. Montreal: IEEE, 2021. 2778–2788.

15 Woo S, Park J, Lee JY, *et al*. CBAM: Convolutional block attention module. Proceedings of the 15th European Conference on Computer Vision (ECCV). Munich: Springer, 2018. 3–19.

16 Hu XD, Wang XQ, Yang X, *et al*. An infrared target intrusion detection method based on feature fusion and enhancement. Defence Technology, 2020, 16(3): 737–746. [doi: 10.1016/j.dt.2019.10.005]

17 Wang CY, Bochkovskiy A, Liao HYM. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE, 2023. 7464–7475.

18 Zhu XZ, Hu H, Lin S, *et al*. Deformable ConvNets v2: More deformable, better results. Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019. 9300–9308.

19 Hu J, Shen L, Sun G. Squeeze-and-excitation networks. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Lake City: IEEE, 2018. 7132–7141.

20 Chu XX, Li L, Zhang B. Make RepVGG greater again: A quantization-aware approach. Proceedings of the 38th AAAI Conference on Artificial Intelligence. Vancouver: AAAI Press, 2024. 11624–11632.

21 Dai XY, Chen YP, Xiao B, *et al*. Dynamic head: Unifying object detection heads with attentions. Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021. 7369–7378.

22 李情芸. 高分辨率机载 SAR 图像中飞机目标的检测与识别技术研究 [硕士学位论文]. 武汉: 华中科技大学, 2015.

23 邝孝伟. 基于注意力机制的肺结节分类模型研究与实现 [硕士学位论文]. 长春: 吉林大学, 2020.

(校对责编: 孙君艳)