

基于改进 YOLOv5 的车辆红外图像多目标识别方法

左涛¹, 周慧龙², 原伟哲²

(1. 上海城投城市发展研究院有限公司, 上海 200030; 2. 西安工程大学 电子信息学院, 西安 710048)

摘要: 城乡结合部的建设是城市建设中重要的一环, 由于难以布设有效的检测设备, 该区域车辆目标的夜间监管一直是城市管理的难题; 基于无人机平台红外夜视图像多运动目标检测为解决这一难题提供了智能化路径: 一种基于改进 YOLOv5 的红外夜视条件下多运动目标识别方法, 分析了交通对象特征、车辆停放对道路红外辐射影响等, 引入了 CBAM 注意力机制, 以提取和融合空间和通道信息, 增强了网络对目标的表达能力; 综合 Efficient IOU Loss 和 Focal Loss 的优点, 使用 EIoU-Focal Loss 损失函数替换了 CIoU Loss 函数, 解决了样本不平衡、红外图像的低分辨率、噪声干扰大以及目标与背景对比度低等弊端, 提高了检测的准确性; 通过加入 DCN 动态调整卷积核的形状, 适应图像中目标的形变, 降低因形状不规则、变化较多造成的识别影响; 在公开数据集上对改进网络与经典网络进行实验和数据比较, 结果表明: 综合改进后的网络对于多目标的识别, 在 YOLOv5x 网络较高的识别结果基础上, 精度提升 3.9%, 召回率提升 4.1%, F_1 增加 4.4%。

关键词: 深度学习; 多目标识别; YOLOv5; 可变形卷积; 注意力机制

Multi-target Recognition Method for Vehicle Infrared Images Based on Improved YOLOv5

ZUO Tao¹, ZHOU Huilong², YUAN Weizhe²

(1. Shanghai Urban Development Research Institute Co., Ltd., Shanghai 200030, China;

2. School of Electronics and Information, Xi'an Polytechnic University, Xi'an 710048, China)

Abstract: The rural-urban fringe is an important part of urban construction, it is difficult to effectively deploy detection equipment, the night supervision of vehicle targets in this area always is a difficult problem for urban management. This paper provides an intelligent approach of detecting multiple moving targets for solving the problem in infrared night vision images based on UAV platforms, presents a multi-moving target recognition method based on improved YOLOv5 in infrared night vision conditions, and analyzes the characteristics of traffic objects and impact of vehicle parking on road infrared radiation, etc. Convolutional block attention module (CBAM) attention mechanism is introduced to extract and integrate spatial with channel information to enhance the expression ability of the network on the target. By combining the advantages of efficient IOU loss and focal loss, the EIoU-focal loss function is used to replace the CIoU loss function, solve the disadvantages of sample imbalance, low resolution of infrared image, large noise interference and low contrast between target and background, and improve the detection accuracy. By adding the DCN to dynamically adjust the shape of the convolution kernel, it can adapt to the deformation of the object in images, and reduce the recognition influence caused by irregular shape and many changes. Finally, experiments and data comparisons between improved network and typical network are implemented on public dataset, the results show that for multi targets recognition, the improved network based on YOLOv5 has higher recognition results, it increases the accuracy, recall rate and F_1 value by 3.9%, 4.1% and 4.4%, respectively.

Keywords: deep learning; multi-object recognition; YOLOv5; deformable convolution; attention mechanisms

0 引言

城乡结合部, 由于交通、安全基础设施的匮乏, 特别是夜间, 检测手段的落后易造成安全事件、交通数据统计、资源调配等市政工作的滞后, 因此需要寻找新的技术和方法弥补这一短板。随着红外技术的不断发展和应用, 红外图像在夜间或低光照条件下具有更好的表现, 因此在夜视或特殊环境下, 红外图像成为一种重要的数据来源。使用

红外技术可以在这些环境中提供清晰的视图, 使得目标检测能够有效进行。目前, 搭载红外目标检测设备的无人机, 可以在夜间或低光照条件下进行有效的目标检测和监控。

然而, 由于红外图像的特殊性质^[1-3], 传统的目标检测算法在红外图像中的表现并不理想。传统的目标检测对数据集有较高的要求, 比如有光照、无遮挡形态清晰等^[4-5]。红外图像中物体的特征和可见光图像存在较大差异^[6-7], 因此需要特别针对红外图像对算法进行优化和改进, 以提高

收稿日期: 2023-07-24; 修回日期: 2023-09-04。

基金项目: 国家自然科学基金(51905405); 陕西省自然科学基金基础研究计划项目(2022JM407)。

作者简介: 左涛(1982-), 男, 硕士, 工程师。

通讯作者: 周慧龙(2000-), 男, 硕士研究生。

引用格式: 左涛, 周慧龙, 原伟哲. 基于改进 YOLOv5 的车辆红外图像多目标识别方法[J]. 计算机测量与控制, 2024, 32(8): 228-235.

检测准确率和鲁棒性^[8-9]。目前国内外的红外目标检测研究主要有: 1) 基于传统算法的红外目标检测: 这类方法主要包括模板匹配、边缘检测和滤波器等^[10]。但模板匹配只适用于目标较为简单的场景、边缘检测受到红外图像质量影响较大、滤波器方法则存在易受噪声影响等问题^[11]; 2) 基于多模态信息的红外目标检测, 主要是将多种传感器获取的信息进行融合, 如红外图像和可见光图像、红外图像和雷达图像、红外图像和声波图像等; 3) 随着深度学习技术的发展, 这个方向渐渐成为目标检测的主流方法。

目前基于深度学习的目标识别方法主要分为两类: 一类是以 R-CNN、Fast R-CNN、Faster R-CNN 为代表的两阶段算法。文献 [12] 于 2014 年提出 R-CNN (Regions with Convolutional Neural Networks)。R-CNN 方法的主要思想是首先利用区域提取 Selective Search 算法产生大约 2000 个候选区域, 然后对每个区域使用预训练的卷积神经网络 (CNN) 进行特征提取, 最后使用 SVM 进行分类, 并使用回归器精细调整边界框位置。R-CNN 与传统目标检测相比, 采用了大样本下有监督预训练加小样本微调的方式解决小样本难以训练甚至过拟合等问题, 获得了显著的性能提升, 但它的计算量大, 运行速度慢, 且需要训练多个模型。为了解决 R-CNN 的计算效率低和训练过程复杂的问题, 文献 [13] 在 2015 年提出 Fast R-CNN。Fast R-CNN 引入了 ROI (Region of Interest) 池化层, 使得所有的计算可以在整张图像上进行, 而不是在每个候选区域上进行。此外, Fast R-CNN 还将分类和边界框回归整合到一个网络中, 使得模型可以在一个统一的框架下进行端到端的训练。Fast R-CNN 相比 R-CNN 有显著的速度提升, 同时精度也有所提高, 但候选区域的提取效率太低^[14]。为了解决 Fast R-CNN 中候选区域提取的效率问题, 文献 [15] 提出 Faster R-CNN。Faster R-CNN 引入了区域提议网络 (RPN, region proposal network), RPN 是一个全卷积网络, 可以在整张图像上高效地生成候选区域。RPN 与 Fast R-CNN 共享卷积特征, 因此几乎没有额外的计算开销。Faster R-CNN 不仅比 Fast R-CNN 更快, 而且在准确性上也有所提高。但 Faster R-CNN 通常会将投入训练的图像固定尺寸, 因此对不同大小尺寸的目标检测泛化性能较差^[16]。

以 R-CNN 算法为代表的 two stage 的方法检测精度越来越高, 其速度却陷入瓶颈, 一种基于回归方法的 one stage 的目标检测算法应运而生, 该算法以 SSD 和 YOLO 系列算法为代表。单阶段算法取消了候选区域生成机制, 直接在一步中预测类别和边界框, 有效提高了遮挡车辆和小物体车辆的识别精度, 能够更好适应红外图像特点和目标复杂性^[17-18]。YOLO 算法是一种常用的目标检测算法, 其以较快的速度和较高的准确率著称^[19-20]。YOLO 算法的主要思想是将目标检测问题视为一个回归问题, 直接在一个网络中预测目标的边界框和类别, 实现端到端的训练和预测。在具体实现中, YOLO 首先将输入图像划分为 $S \times S$ 的网格, 如果某个目标的中心落入某个网格, 那么这个网格就负责预测这个目标。

然而, 由于每个网格只预测一个目标, YOLO 在处理小目标和重叠目标时的性能不如两阶段的方法^[21]。此外, YOLO 对于目标形状的预测也不够准确, 因为它对所有的边界框使用了相同的损失函数, 忽略了目标形状的多样性^[22]。为了解决这些问题, 文献 [23] 提出 YOLOv2, 又称为 YOLO9000。YOLOv2 引入了锚框 (Anchor Box) 的概念, 使得每个网格可以预测多个目标。此外, YOLOv2 引入了多尺度预测, 使得模型可以更好地处理不同大小的目标, 但是当图像中存在多物体密集排列或小目标的时候, YOLOv2 的检测效果有待提升。针对 YOLOv2 存在的问题, YOLOv3 首次出现在目标识别领域中。该算法引入残差网络模块, 在低、中、高 3 个层次上分别预测目标框, 基本解决了小目标检测的问题, 在速度和精度上实现了较好的平衡。随后, 经过对 YOLOv3 的升级改造, YOLOv4 被提出。其引入了很多先进的目标检测技术, 包括 CIoU loss、PANet 和 SAM block 等, 使得它在保持高速度的同时, 达到了更高的准确性。YOLOv4 也引入了 Mish 激活函数和 CSPDarknet53 作为其主干网络, 这两项技术都有助于提升网络的性能和效率。为了进一步提高模型性能, UltralyticsLLC 公司推出 YOLOv5, 主要针对速度和易用性进行优化, 包括模型大小、推理速度和训练时间等。

2023 年, Mao^[24] 提出了一种基于 YOLOv5 和 DCN (Deformable Convolution Net) 可变形卷积的小目标检测算法, 通过 DCN 可变形卷积来提取更为准确的特征, 对浅层语义信息较弱的特征图进行有效性检查, 并对有效性的感受野进行加权处理。实验结果表明, 改进后的算法在小目标检测方面取得了较好的性能。

基于 YOLOv5 的目标检测算法在红外夜视条件下针对多运动目标的识别方面有着广阔的应用前景。近年来, 通过对 YOLOv5 算法模型的改进和优化, 红外目标检测算法的检测精度和效率得到了大幅提升, 但在红外夜视条件下多运动目标的识别方面, 还存在一些挑战和问题:

- 1) 红外夜视环境下, 多运动目标的距离、大小和运动方向都可能会发生变化, 这导致目标在图像中的尺度也会随之变化;

- 2) 在多运动目标检测中, 会出现目标遮挡或者交叠的情况, 尤其是在红外夜视环境下, 由于光照条件较差, 遮挡或交叠的目标在红外图像中的边界往往不太明显;

- 3) 由于红外图像与可见光图像在质地和特性上的差异, 使得许多针对可见光图像优化的算法在红外图像上的效果并不理想。

因此, 本文将针对上述问题, 对 YOLOv5 算法进行改进。

- 1) 引入了 CBAM 注意力机制, 以增强模型对目标的关注程度和区分能力;

- 2) 把原有的 CIoU (Compound IoU) 损失函数优化为 EIoU (Efficient IoU), 并加入 Focal 函数, 以进一步提高检测精度;

- 3) 改进了 DCN 可变形卷积, 以适应多种目标形状和运动轨迹。

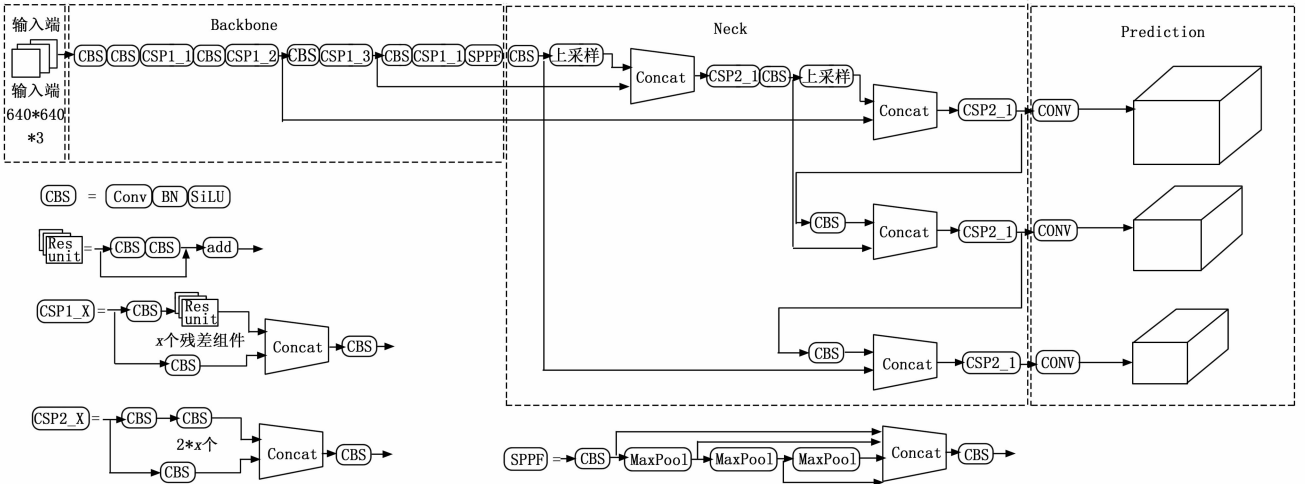


图 1 YOLOv5 网络结构图

1 基于综合改进的 YOLOv5 多目标识别方法

YOLOv5 是由 Glenn Jocher 等人在 2020 年提出的版本，其主要针对速度和易用性进行优化，包括模型大小、推理速度和训练时间等。网络结构可以分为两个部分：特征提取网络和检测网络，如图 1 所示。

特征提取网络通常采用卷积神经网络来提取图像中的特征，例如边缘、纹理、颜色等等。YOLOv5 的特征提取网络采用 CSPDarknet-53 架构，CSPDarknet 主干网络由一系列 CSP 模块组成，每个 CSP 模块都包含了一个残差块和一个 CSP 结构，其中 CSP 结构使用了跨阶段部分连接 (Cross Stage Partial Connection) 来加强网络特征提取的能力。CSPDarknet 中的 CSP 结构类似于 ResNeXt 和 EfficientNet 中的结构，可以在不增加网络深度的情况下增加网络的表示能力。其目的是从输入图像中提取出具有高维度的特征信息，以便更好地支持目标检测任务。特征提取网络通常会将图像的分辨率降低，从而能够在相对较小的内存和计算资源下处理更大的图像。其检测能力相较 YOLOv4 有了明显的提升，同时也取得了较高的速度，可以有效地提取图像特征。

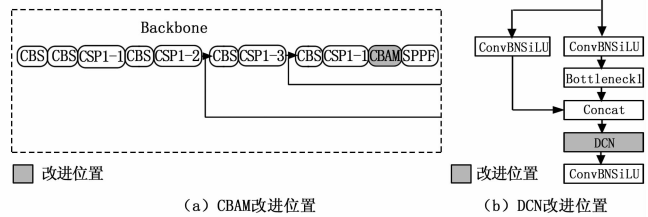
检测网络的任务是利用特征提取网络提取的特征信息来检测图像中的目标。YOLOv5 的检测网络使用了一系列卷积层、池化层、全连接层和激活函数等组件，用于对目标的类别、位置和置信度等信息的输出，以便后续的目标跟踪或者分类等任务进行处理。

在 YOLOv5 中，特征提取网络和检测网络分别采用不同的分辨率。特征提取网络的分辨率比检测网络低，这样可以大大减少计算量。在特征提取网络的末尾，引入了一个 SPP 结构，可以将不同分辨率的特征进行融合，提高网络的感受野。在检测网络中，引入了 PANet 结构，可以对不同分辨率的特征进行级联，提高检测的准确率，这两个部分协同工作，完成整个目标检测任务。

YOLOv5 有 4 个不同的版本：s、m、l 和 x，分别对应于不同的模型大小和精度。YOLOv5x 是其中具有高精度的

版本，但应用于红外图像中的小目标、重叠目标或是图像信息不全的目标检测时，检测精度较差，识别能力有限。

本文在 YOLOv5x 网络中引入 CBAM (Convolutional Block Attention Module) 注意力机制^[25]，DCN (Deformable Convolutional Networks) 可变形卷积^[26-27]来提高其对于红外目标中的小目标和形变目标的检测精度，改进位置如图 2 所示，同时，结合 EIou-Focal (Focal and Efficient IOU Loss) 损失函数，进一步提高网络对红外图像的适应能力。改进后的网络实现了红外图像多目标识别兼顾 YOLOv5x 的高精度和高识别效率。



(a) CBAM改进位置

(b) DCN改进位置

图 2 网络改进位置示意图

1.1 CBAM 卷积注意力模块

CBAM 注意力机制是在 SE (Squeeze-and-Excitation) 注意力机制的基础上提出的一种新型注意力机制^[28]，SE 通过自适应地学习每个通道的权重来提高模型的表达能力，CBAM 在 SE 注意力机制的基础上引入了空间注意力机制，以进一步提高模型的表达能力。

CBAM 注意力机制包含两个注意力模块：通道注意力模块 (Channel Attention Module) 和空间注意力模块 (Spatial Attention Module)，CBAM 整体结构如图 3 所示。

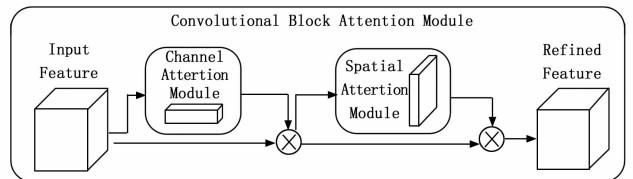


图 3 CBAM 整体结构图

通道注意力模块类似于 SE 注意力机制中的 Squeeze-and-Excitation 操作, 它通过自适应地学习每个通道的权重来提高模型的表达能力; 空间注意力模块则关注每个通道在不同空间位置上的特征表示, 进一步提高模型的准确性和鲁棒性。

假设输入特征图为 $F \in R^{C \times H \times W}$, 利用 CBAM 依此推导出一维通道注意图 $M_C \in R^{C \times 1 \times 1}$ 和二维空间注意图 $M_S \in R^{1 \times H \times W}$, 总的注意过程可以用公式概括为:

$$F' = M_C(F) \otimes F \tag{1}$$

$$F'' = M_S(F) \otimes F' \tag{2}$$

通道注意力模块用来提高卷积神经网络在通道维度上的特征提取能力, 通过对特征图在通道维度上进行统计, 得到每个通道的重要性权重, 使模型更加关注重要通道, 忽略不重要通道。简单来说就是让模型在处理特征图时能够更关注一些对于结果更重要的通道, 这样可以减少一些无用特征, 提高模型的准确性和效率, 如图 4 所示。

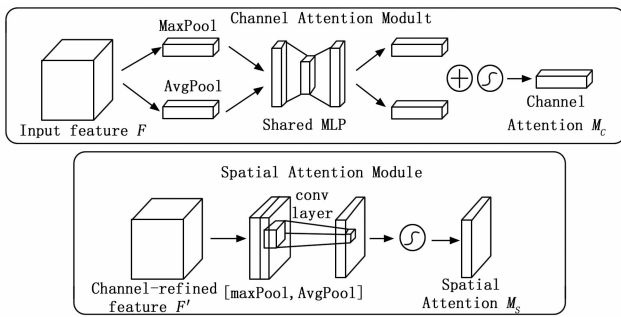


图 4 通道注意力模块和空间注意力模块注意流程图

CBAM 通道注意力模块的计算方式基于每个通道的平均值和最大值。首先对于输入的特征图, 通过对空间维度进行压缩, 得到一个通道维度的全局描述, 即平均值和最大值; 然后通过两个 MLP (多层感知机) 分别处理平均值和最大值, 生成两个通道权重向量; 最后将两个通道权重向量相加, 得到最终权重向量, 并将其应用到原始的特征图上。这样, 模型就可以在通道维度上更加关注重要的特征, 从而提高模型的性能。

空间注意力模块提高卷积神经网络在空间维度上的特征提取能力, 通过对特征图在空间维度上进行统计, 得到每个空间位置的重要性权重, 从而使得模型更加关注重要的空间位置, 忽略不重要的位置。与通道注意模块不同的是, 空间注意模块关注的是信息部分在哪。

为了计算空间注意力, 首先沿着通道轴进行平均池化 (AvgPool) 和最大池化 (MaxPool) 操作, 并将它们连接起来以生成一个有效的特征描述符。这个描述符可以用于计算空间注意力图, 通过特征间的空间关系来生成注意力图。CBAM 空间注意力模块的核心优点在于, 它能够通过对不同池化结果进行 MLP 操作, 使得模型能够充分利用每个空间位置的信息, 从而得到更加精细的空间注意力权重。与通道注意模块不同, 空间注意模块只涉及空间维度上的计

算, 因此计算量较小, 对模型的计算速度影响较小。通过这种方式, CBAM 的空间注意力模块能够有效地提高卷积神经网络在空间维度上的特征提取能力。

两种模块同时使用平均池化和最大池化, 将池化结果分别输入到两个 MLP 中。空间注意力模块中的 MLP 主要用于计算每个空间位置的权重, 而通道注意力模块中的 MLP 则用于计算每个通道的权重。并证明了这种做法比单独使用一种池化方法更具有表征力。

1.2 DCN 可变形卷积模块

DCN 可变形卷积是一种能够学习卷积核形变的卷积方式, 基本原理是在传统卷积中增加了一个偏移量, 并在卷积核中增加一个形变因子。传统的卷积操作只能在固定的采样点上执行, 而 DCN 可变形卷积则允许在不同位置执行卷积, 并且允许卷积核进行形变。在 DCN 可变形卷积中, 每个卷积核都由一个中心点和一组偏移量构成, 中心点用于控制卷积核的位置, 而偏移量则用于控制卷积核的形状。DCN 可变形卷积的卷积核可以根据像素点的特征进行形变, 从而更加精准地定位物体, 相比传统卷积, DCN 可变形卷积可以更好地适应物体的不规则形状和大小。

DCN 可变形卷积的应用场景非常广泛, 尤其是在目标检测和图像分割等领域。在目标检测中, 由于物体的形状和大小不同, 传统的卷积操作可能会导致检测的精度降低, 而 DCN 可变形卷积可以更好地适应各种不同形状和大小的物体, 从而提高了检测的准确性。在图像分割中, DCN 可变形卷积可以更准确地捕捉物体的形状和边界信息, 从而提高了分割的精度。

图 5 中, (a) 是标准卷积的规则采样网格, (b) 是在可变形卷积中具有增强偏移的变形采样位置, (c) (d) 是 (b) 的特例, 表明可变形卷积推广了尺度、各向异性纵横比和旋转的各种变换^[29]。

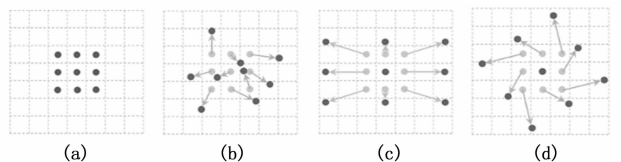


图 5 3×3 标准和可变形卷积中的采样位置示意图

DCN 可变形卷积可以更好地适应各种不同形状和大小的物体, 从而提高了检测的准确性。在图像分割中, DCN 可变形卷积可以更准确地捕捉物体的形状和边界信息, 从而提高了分割的精度, 增强了模型对目标的感知能力, 进一步提高网络的识别性能和精度。

1.3 动态缩放有效交并比损失函数

在目标检测领域中, CIoU 因考虑了检测框的中心点、宽度和高度等方面的因素, 相较于 IoU 具有更好的性能, 但其弊端在于不可导, 无法直接用于损失函数中优化。为了解决这个问题, 提出可导的基于 CIoU 的 EIoU。同时, 在本文目标识别任务中, 由于车型外形、视角等差异的存在, 识别类别不平衡, 会导致模型对于少数类别的预测效果

表 1 实验环境配置参数

参数	操作系统	CUDA	Pytorch	YOLOv5	图形处理器	中央处理器	内存
配置	Ubuntu 20.04.1	11.1	1.10.0	6.1(x)	NVIDIA RTX 4090	Intel Xeon Silver 4314	125 GB

较差,为此本文引入了 Focal Loss 损失函数^[30-31], Focal loss 具有以下几个优点:

1) 能够降低易分类样本对损失的贡献,让模型更加关注困难样本。这对于一些数据集中存在严重类别不平衡的情况特别有用。

2) 适用于各种类型的分类任务,例如二分类、多分类等等。

3) 不需要对数据集进行任何额外的处理,只需要在损失函数上进行改动即可。

4) 易于实现和计算,不需要增加额外的计算成本。

Focal EIOU 损失函数是在 Focal loss 和 EIOU 损失函数的基础上进行了改进,旨在解决 EIOU 损失函数在难易样本分布不均时的问题。具体来说, Focal EIOU 损失函数的计算公式为:

$$L_{\text{FocalEIoU}} = (1 - \text{IoU}_{\text{true,pred}})^{\gamma} L_{\text{EIoU}} = \text{IoU}^{\gamma} L_{\text{EIoU}} \quad (3)$$

式中, γ 是一个调节难易样本权重的因子。通过将 Focal loss 和 EIOU loss 结合起来, Focal EIOU 损失函数不仅能够对困难样本进行加权,还能够在计算 IoU 时避免过于乐观或悲观地预测结果,从而使得模型更加准确地预测物体的位置和大小。

YOLOv5 中采用 Focal EIOU 损失函数能够更好地适应物体的形状和大小变化,同时对于难以分类的物体,也能够更加关注,提升模型的精度。此外, Focal EIoU 损失函数的计算量相对较小,能够更快地训练模型,并且不需要对数据集进行额外的处理。

2 基于改进 YOLOv5 的目标识别与结果分析

2.1 实验环境配置

实验选用 PyTorch1.10.0 调试 YOLOv5 网络结构的源代码。采用 CUDA 11.1 适应深度学习任务中对计算效率的高要求,以支持显著的硬件加速。同时,使用 NVIDIA cuDNN 库提升 CUDA 的浮点运算性能,实验环境的具体参数配置如表 1 所示。

2.2 数据集获取

红外夜视数据集资源较少,实验所使用的数据集是来源于网络视频和公开数据集,主要内容为利用无人机搭载红外摄像机进行航拍的各类车辆、行人等的图片,以及从视频中截取的连续帧。数据集中共有 3 164 张图片,其中训练集 2 848 张,验证集 316 张,其中包含 person, small vehicles, SUV 等 11 个目标类别,如图 6 所示。图 (a) 表示训练集每个目标类别的数量; (b) 表示 center 点的位置; (c) 表示 label 的高宽。

如图 7 所示,红外图像的特性是物体颜色与其温度相关,利用红外图像进行检测能够优于普通 CCD 采集的图像。但是,仔细观察样本图可以发现深色的、类似于车辆

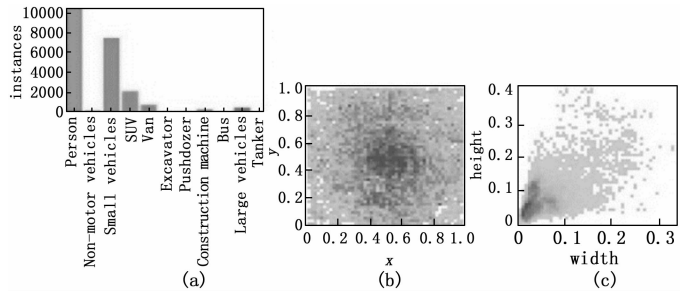


图 6 数据集及标注框信息

轮廓形状的阴影区域,这些是由于图像采集时间为傍晚,日照造成车辆驶离后的区域与暴露在阳光下路面存在显著的温度差异,形成类似阴影的深色斑块。在处理这类数据集时,标注与训练过程需要特别注意这种现象。文章训练集标注需对深色阴影区域进行仔细判别,另外,需要调整网络参数,优化模型结构,以提升模型对此类情况的识别和分辨能力。



图 7 数据集图像内容示例

2.3 网络评价指标

评价指标分为检测精度和检测速度两类。在相同的迭代次数 (epochs) 下,检测精度主要反映了一种网络模型对检测目标的识别准确程度;检测速度则反映了该网络模型进行一个训练流程的耗时长短。其中,检测精度由精度 (precision)、召回率 (recall)、 F_1 分数 (F_1 score) 等指标进行表征,这些指标的计算取决于模型对于检测目标的预测正确性和预测方向 (TP/TN/FP/FN)。

真阳性 (TP, true positive): 样本的真实类别是正例,并且模型预测的结果也是正例,预测正确。

真阴性 (TN, true negative): 样本的真实类别是负例,并且模型将其预测成为负例,预测正确。

假阳性 (FP, false positive): 样本的真实类别是负例,但是模型将其预测成为正例,预测错误。

假阴性 (FN, false negative): 样本的真实类别是正例,但是模型将其预测成为负例,预测错误。

这些指标的计算公式分别如式 (4) ~ (6) 所示:

$$precision = \frac{TP}{TP + FP} \quad (4)$$

$$recall = \frac{TP}{FP + FN} = \frac{TP}{P} = sensitive \quad (5)$$

$$F_1 Score = \frac{2 \times precision \times recall}{precision + recall} \quad (6)$$

其中: F_1 分数是分类的一个衡量标准, 是精确率和召回率的调和平均函数, 介于 0 和 1 之间, 与网络分类性能成正比。

此外, 还有多尺度精度 (mAP) 参数指标, 是基于 Precision 和 Recall 值作为两轴作图后围成的面积, 即均值平均精度。

2.4 实验结果和分析

在相同实验环境下, 对 YOLOv5 的 4 个版本进行了对比实验, 以验证本文选择 YOLOv5x 作为改进对象的合理性; 将上述模块分别加入网络模型中的 backbone 主干网络, 进行消融实验; 再分别对 YOLOv5x、YOLOv5x+CBAM、YOLOv5x+EIoU-Focal、YOLOv5x+DCN、YOLOv5x+EIoU-Focal+CBAM+DCN 进行多目标识别结果分析。

2.4.1 YOLOv5 模型实验分析

在相同环境下, 对 YOLOv5 的 4 种原始模型进行训练, 测试指标数据如表 2 和 3 所示。

表 2 YOLOv5 四种模型的速度数据

Model	YOLOv5s	YOLOv5m	YOLOv5l	YOLOv5x
Epochs	300	300	300	300
浮点运算次数/G	15.8	48.0	107.8	204.8
参数量/M	7.0	20.9	46.2	86.3
层数	157	212	267	322
运算时间/h	3.3	4.5	6.3	6.6
FPS	95	73	61	51

表 3 YOLOv5 四种模型的精度数据

Model	YOLOv5s	YOLOv5m	YOLOv5l	YOLOv5x
Epochs	300	300	300	300
Precision/%	89.3	90.7	91.1	91.0
Recall/%	82.1	81.6	82.0	82.4
F_1 分数	0.855	0.859	0.863	0.865
mAP/%	83.9	84.2	85.0	85.4

对于这 4 种模型, 我们可以从表中明显地看出差别。YOLOv5x 是 YOLOv5 中最大的网络模型, 其浮点运算次数和参数量甚至达到了 YOLOv5s 的十余倍, 是 YOLOv5l 的两倍, 但进行 300 轮训练的运算时间相较 YOLOv5l 仅仅增加了 0.3 小时, 其检测精度也是 4 种模型中最高的。

2.4.2 基于改进的 YOLOv5x 模型实验结果与分析

按照第 2 节所述, 对 YOLOv5x、YOLOv5x+CBAM、YOLOv5x+EIoU-Focal、YOLOv5x+DCN、YOLOv5x+EIoU-Focal+CBAM+DCN 进行消融实验, 结果如表 4 所示, 表中模型 1 表示 YOLOv5x, 模型 2 表示 YOLOv5x +

CBAM, 模型 3 表示 YOLOv5x+EIoU-Focal, 模型 4 表示 YOLOv5x+DCN, 模型 5 表示 YOLOv5x+EIoU-Focal+CBAM+DCN。

1) 模型实验精度结果及分析:

表 4 模型精度实验数据表

Model	模型 1	模型 2	模型 3	模型 4	模型 5
Epochs	300	300	300	300	300
Precision/%	91.0	92.8	89.5	92.1	94.9
Recall/%	82.4	80.7	85.6	83.7	86.5
F_1	0.865	0.863	0.875	0.877	0.909
mAP/%	85.4	85.9	85.4	85.9	86.4

从表 4 可以观察到, 引入 CBAM 模块后, 模型的参数量和大小都出现了不同程度的增加, 这表明引入注意力机制模块会使模型的复杂度提高, 进而导致计算效率的降低和需要的存储资源的增加。引入 CBAM 模块后, 模型在 mAP 方面取得了较好的提升, 这验证了注意力机制模块在特征提取和目标识别方面的有效性。值得注意的是, 尽管引入 CBAM 模块后模型的 mAP 提升了, 但由于召回率和精确率之间的平衡被打破, 导致了 F_1 分数的下降。此外, 单独引入 EIoU Focal, 从表中可以看出精度指标下降明显, 单独引入 DCN, 召回率提升也并不明显。

因此, 在模型中同时引入 EIoU-Focal loss、CBAM 和 DCN 模块后, 可以观察到模型相比原始的 YOLOv5x 模型在精度上有了显著的提升。同时, 实验数据也验证了 EIoU-Focal loss、CBAM、和 DCN 的组合使用能够显著提高目标检测的精度和鲁棒性。

2) 可视化训练结果及分析:

图 8 为模型训练的 F_1 曲线图, 从子图 (a)、(b) 的数据可以清晰地看到, 当单独替换为 EIoU-Focal loss 损失函数后, 虽然模型的 mAP 值和 F_1 分数在全局范围内并没有出现明显的提升, 但在模型的召回率曲线优化显著, 结果反映出识别网络对目标分类正确率有正向贡献, 识别精准度得到提高。

从子图 (c) 曲线变化可以发现, 针对红外图像的特点, 网络在加入 DCN 后, 该模块能够通过动态调整卷积核的形状, 以适应图像中目标的形变, 因此对于形状不规则、变化多端的目标具有更好的识别能力, 所以, 显著增强了对于显示不全和形变目标的识别能力, F_1 分数和 mAP 值都得到了提升。

同时, 综合分析图 8 (a) (b) (c) 曲线图, 我们观察到某些类别 (如 Non-motor vehicles、Pushdozer 等) 的曲线显示出不规则的波动和线性变化, 不规则变化的主因是数据集的类别分布不均, 对于在数据集中相对稀少的类别, 模型在训练过程中所计算的精确率和召回率可能会因为样本数量的稀缺性产生大幅度的上升或是下降; 综合改进后的网络, 如图 8 (d) 所示, 网络的 F_1 曲线中存在的折线波动已经有所改善。

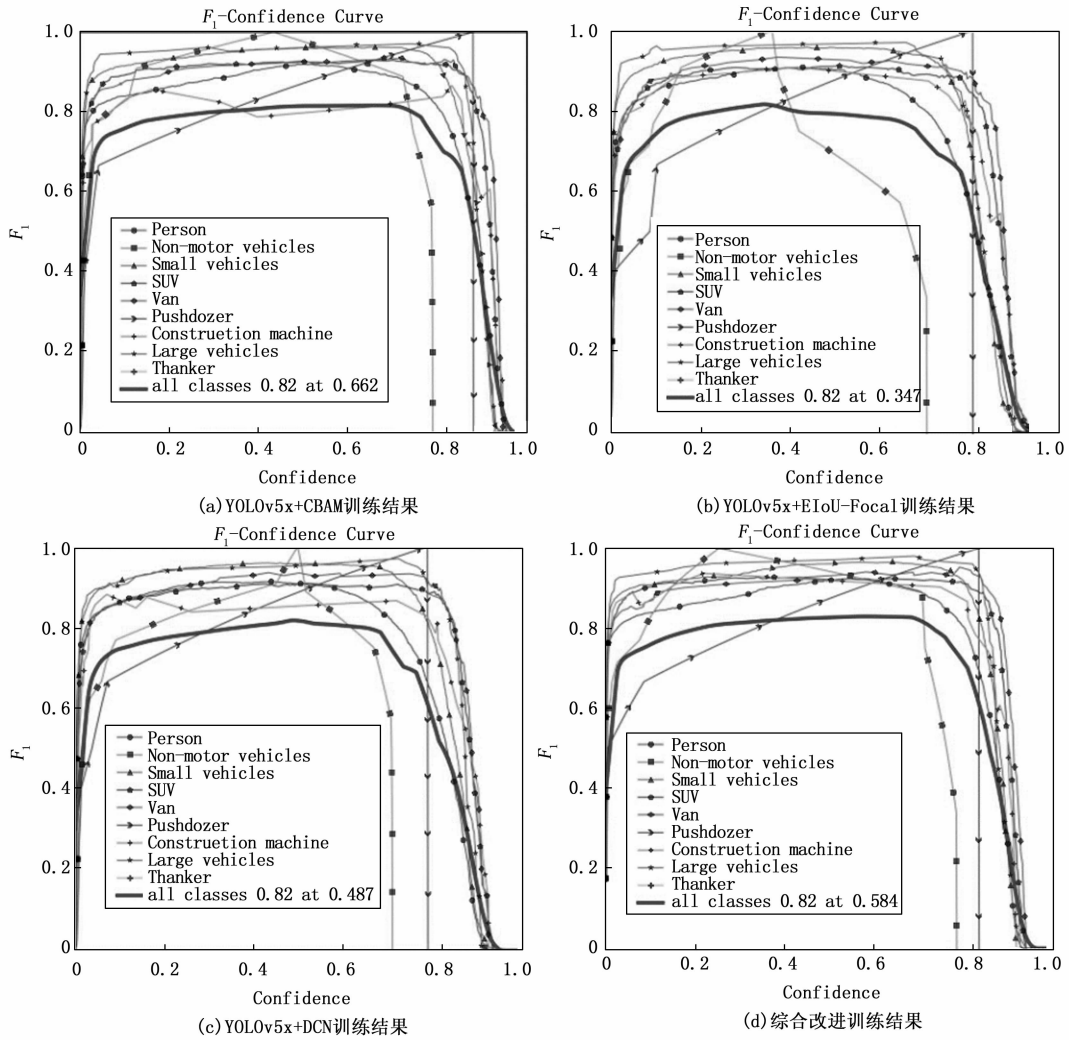


图 8 模型训练 F_1 曲线结果对比

3) 测试结果分析:

我们选择 4 种比较典型的场景进行识别结果对比分析, 如图 9 所示, 场景 1 存在大型车辆, 场景 2 存在驶离车辆后温度较低区域, 场景 3 存在静止、运动目标、场景 4 包含不同类运动目标 (机动车、非机动车)。

经过对比图 9 中 YOLOv5x+CBAM 列, 可以看到: 在所展示的特定场景中, 网络模型在添加了注意力机制模块之后, 对各类目标的平均检测精度都得到了显著的提升, 尤其是在降低小目标的误检率方面, 表现出了显著的优越性, 证明在特征注意力和空间注意力的双重机制下, 改进网络对小目标的识别和定位具有显著优势。

对比图 9 的测试结果, 加入 EIoU-Focal loss 的网络在处理易混淆目标时, 其识别与区分的能力也得到了显著提升。例如, 网络对于场景 2 中 Non-motor vehicles 和 person 的区分更为明确; 同样在场景 3 中, 网络对于 small vehicles 和 SUV 的误识别率也更低。说明 EIoU-Focal loss 具有改善模型细粒度识别能力以及提高识别正确率。同时, 由于能够有效捕捉并适应目标的局部形变信息, 加入 DCN 后, 对于 Small vehicle 类别, 改进网络表现出了更高的检

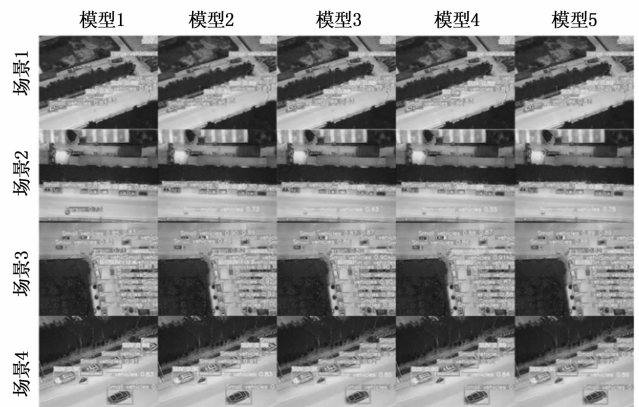


图 9 不同场景实验结果对比

测精度, 在场景 4 中, 可以看到加入 DCN 的网络可以提升对 person 和 Non-motor vehicles 这类小目标的区分能力。

通过表中最后一列 YOLOv5x+EIoU-Focal+CBAM+DCN 与其他 3 列结果图的对比, 可以看出, 综合改进后的 YOLOv5x 网络中同时应用 EIoU-Focal loss、DCN 和 CBAM 注意力机制的模型, 其检测精度显著优于其他模型。

3 结束语

为了解决城乡结合处交通目标的识别问题, 针对夜间红外图像综合改进了 YOLOv5 深度学习网络框架, 通过引入 CBAM 注意力机制、更换损失函数以及加入 DCN 可变形卷积模块, 提升了算法的整体性能。将原始网络与分别加入 CBAM 注意力机制、EIoU-Focal loss 损失函数、DCN 可变形卷积以及综合改进后的网络进行了对比, 实验结果表明, 文章所提方法能够将红外识别中类似车辆的阴影斑块造成的误检情况去除, 改进策略对网络性能的提升显著, 对红外图像中的交通目标进行有效识别。

参考文献:

- [1] XU L, ZHAO F, XU P, et al. Infrared target recognition with deep learning algorithms [J]. *Multimedia Tools and Applications*, 2023, 82 (11): 17213 - 17230.
- [2] ZHANG W, FU X, LI W. The intelligent vehicle target recognition algorithm based on target infrared features combined with lidar [J]. *Computer Communications*, 2020, 155: 158 - 165.
- [3] AKULA N V A, SARDANA H K. Optimized bag of features framework for object recognition in thermal infrared images [J]. *Journal of Electronic Imaging, SPIE*, 2018, 27 (6): 063017.
- [4] KANG J, TARIQ S, OH H, et al. A survey of deep learning-based object detection methods and datasets for overhead imagery [J]. *IEEE Access*, 2022, 10: 20118 - 20134.
- [5] KAMATH V, RENUA A. A proposed exploratory study of object detectors to learn the influence of datasets on model performance [J]. *Journal of Physics: Conference Series*, 2022, 2161 (1): 012076.
- [6] SUN L, LI Y, ZHENG M, et al. MCnet: multiscale visible image and infrared image fusion network [J]. *Signal Processing*, 2023, 208: 108996.
- [7] MA W, WANG K, LI J, et al. Infrared and visible image fusion technology and application: a review [J]. *Sensors*, 2023, 23 (2): 599.
- [8] TANG H X, LIU J, LI H Y, et al. DuReader_robust: a Chinese dataset towards evaluating robustness and generalization of machine reading comprehension in real-world applications [C] // *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*, 2021, 2: 955 - 963.
- [9] ZHANG L, WANG J, AN Z. Design of a communication device of infrared thermal imaging target recognition and tracking system [J]. *Advances in Multimedia*, 2022: 1 - 12.
- [10] ZHAO M J, LI W, LI L, et al. Single-frame infrared small-target detection: a survey [J]. *IEEE Geoscience and Remote Sensing Magazine*, 2022, 10 (2): 87 - 119.
- [11] JU M, LUO J, LIU G, et al. ISTDet: An efficient end-to-end neural network for infrared small target detection [J]. *Infrared Physics & Technology*, 2021, 114: 103659.
- [12] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C] // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016: 580 - 587.
- [13] GIRSHICK R. Fast R-CNN [C] // *Proceedings of the IEEE International Conference on Computer Vision*, 2002: 1440 - 1448.
- [14] 熊风光, 董 彪, 张 鑫, 等. 基于改进 Mask RCNN 算法的遥感建筑物检测 [J]. *计算机工程与设计*, 2023, 44 (1): 218 - 223.
- [15] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [C] // *Advances in Neural Information Processing Systems*, 2017: 91 - 99.
- [16] 王 彬, 李小曼, 赵作鹏. 基于改进 Faster RCNN 的驾驶员手持通话检测 [J]. *江苏大学学报 (自然科学版)*, 2023, 44 (3): 318 - 323.
- [17] 涂海清. 基于 Yolov5s-se 和数据增强的夜间车辆目标检测 [D]. 广州: 华南理工大学, 2021.
- [18] 林 健, 张巍巍, 张 凯, 等. 基于 YOLOv5 的红外目标检测算法 [J]. *飞控与探测*, 2022, 5 (3): 63 - 71.
- [19] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: optimal speed and accuracy of object detection [J]. *Arxiv*, 2020, 10934.
- [20] CHAI E, TA L, MA Z, et al. ERF-YOLO: A YOLO algorithm compatible with fewer parameters and higher accuracy [J]. *Image and Vision Computing*, 2021, 116: 104317.
- [21] 周晋伟, 王建平. YOLO 物体检测算法研究综述 [J]. *常州工学院学报*, 2023, 36 (1): 18 - 23.
- [22] 魏瑞歌. 基于深度学习的无人机小目标检测算法研究 [D]. 成都: 电子科技大学, 2022.
- [23] REDMON J, FARHADI A. YOLO9000: Better, faster, stronger [C] // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2022: 7263 - 7271.
- [24] 毛 涛. 基于 YOLOv5 的小目标检测算法研究 [D]. 合肥: 安徽理工大学, 2021.
- [25] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional block attention module [C] // *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018: 3 - 19.
- [26] CHEN F, WU F, XU J, et al. Adaptive deformable convolutional network [J]. *Neurocomputing*, 2021, 453: 853 - 864.
- [27] WANG H, ZHANG Y, LIU C, et al. sEMG based hand gesture recognition with deformable convolutional network [J]. *International Journal of Machine Learning and Cybernetics*, 2022, 13 (6): 1729 - 1738.
- [28] WOO S, PARK J, LEE J, et al. CBAM: Convolutional block attention module [C] // *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018: 3 - 19.
- [29] DAI J, QI H, XIONG Y, et al. Deformable Convolutional Networks [J]. *IEEE Access*, 2017: 764 - 773.
- [30] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection [J]. *IEEE Access*, 2020, 42 (2): 318 - 327.
- [31] NIU J, LI H, CHEN X, et al. An improved YOLOv5 network for detection of printed circuit board defects [J]. *Journal of Sensors*, 2023: 1 - 10.