

基于并行注意力机制的地面红外目标检测方法 (特邀)

赵晓枫, 徐叶斌, 吴 飞, 牛家辉, 蔡 伟, 张志利

(火箭军工程大学兵器发射理论与技术国家重点学科实验室, 陕西 西安 710025)

摘要: 地面背景下的红外目标检测是伪装防护、精确制导等领域的关键技术。针对现有基于深度学习的目标检测模型对地面背景下红外目标进行检测时容易受到复杂背景干扰、对目标关注不足, 从而导致检测准确率不高的问题, 文中提出了一种基于并行注意力机制的地面红外目标检测方法。首先, 利用卷积和注意力并行的下采样方式, 在降低模型的空间复杂度和提升训练速度的同时, 对目标特征进行聚焦和关注; 其次, 对主干网络提取的多尺度特征进行融合, 通过不同尺度信息的复用与互补抑制背景信息的干扰, 提升目标检测的准确率; 最后, 利用焦点损失函数和 CIOU 损失函数提高模型的分类与回归精度。实验结果表明, 在 Infrared-VOC 数据集上该模型的平均检测精度为 82.2%, 比 YOLOv3 提高了 6.9%, 同时模型的空间复杂度仅为 YOLOv3 的 32.6%, 训练时间为 YOLOv3 的 43.7%, 实现了模型训练效率和检测精度的提升。

关键词: 红外目标检测; 并行注意力机制; 深度学习; YOLOv3; 训练效率

中图分类号: TP391.4 **文献标志码:** A **DOI:** 10.3788/IRLA20210290

Ground infrared target detection method based on a parallel attention mechanism (*Invited*)

Zhao Xiaofeng, Xu Yebin, Wu Fei, Niu Jiahui, Cai Wei, Zhang Zhili

(Armament Launch Theory and Technology Key Discipline Laboratory of China,
Rocket Force University of Engineering, Xi'an 710025, China)

Abstract: Ground infrared target detection is a key technology in the fields of camouflage protection and precision guidance. For the current deep learning-based target detection model to detect infrared targets in the ground background, it is easy to be interfered by complex backgrounds and insufficient attention to the target, which leads to the problem of low detection accuracy. A method of ground infrared target detection based on a parallel attention mechanism was proposed. Firstly, the parallel down-sampling method of convolution and attention was used to reduce the spatial complexity of the model and increase the training speed, while focusing and paying attention to the target features. Secondly, the multi-scale features extracted by the backbone network were fused to suppress the interference of background information and improve the accuracy of target detection through the multiplexing and complementary of different scale information. Finally, the focal loss and CIOU loss were used to improve the classification and regression accuracy of the model. The experiment results showed that the average detection accuracy of the model on the Infrared-VOC dataset was 82.2%, which was 6.9% higher than YOLOv3. At the same time, the space complexity of the model was only 32.6% of YOLOv3, and the training time was 43.7% of YOLOv3. The improvement of model training efficiency and detection accuracy was achieved.

收稿日期: 2021-04-30; 修订日期: 2021-06-07

基金项目: 国家自然科学基金 (41404022); 陕西省自然科学基金面上项目 (2015JM4128)

作者简介: 赵晓枫, 男, 副教授, 硕士生导师, 博士, 主要从事兵器发射理论与技术方面的研究。

Key words: infrared target detection; parallel attention mechanism; deep learning; YOLOv3; training efficiency

0 引言

基于深度卷积神经网络 (Deep Convolutional Neural Networks, DCNN) 的计算机视觉技术在可见光图像的各研究领域取得了巨大进展。当前, 将基于深度学习的目标检测延伸到红外图像中已成为目标检测领域的研究热点, 在目标探测^[1-4]、伪装防护^[5-6]、视频监控^[7]、智能检测^[8]等领域中有着广泛的前景。

利用卷积神经网络遍历提取目标特征时, 往往会因为常规卷积不能对目标产生足够的关注, 前景背景无法得到有效区分, 从而导致目标特征提取不充分, 影响了检测精度。为了有效提取目标特征信息, 提升模型对目标的检测精度, 将注意力机制与特征提取相结合引起了研究者的关注。危水根^[9]等利用人眼视觉系统的注意机制设计了 Soft-FART 网络, 提高对红外小目标的检测精度。刘旭^[10]等采取视觉对比机制突出目标特征, 消除高相似性干扰目标, 实现了对空基复杂背景下红外弱小目标的检测。刘俊明^[11]等通过在全卷积神经网络中融合视觉显著机制分离出红外小目标, 并使用 SE^[12] 注意力机制提升模型的训练效果。上述注意力机制的使用提高了对空基红外目标的检测精度。但是在提升模型的检测效果之时也会引入额外的计算参数, 增加模型的空间复杂度, 影响训练效率。此外, 相对于空中的红外目标检测, 地面背景下的红外目标虽然具有更多的纹理和细节特征, 但是面临着更加复杂的环境背景, 在目标检测的过程中容易受到相似环境红外特征的干扰, 导致检测结果并不理想。

文中设计了基于并行注意力机制 (Parallel Attention Mechanism, PAM) 的主干网络, 对输入特征同时进行下采样和注意力聚焦, 以降低模型空间复杂度, 提升训练效率和对目标特征的关注度。对主干网络提取的多尺度特征进行自顶向下的融合, 通过多尺度的特征复用与信息互补, 达到抑制背景信息和筛选目标特征的目的。最终, 构建了基于并行注意力机制的红外目标检测模型 (PAM-YOLO), 并通过与其他目标检测模型的对比试验、逐步增加各改进模块的消融实验和与 SE 注意力机制的对比试验, 对构建的 PAM-YOLO

目标检测模型进行了分析和验证。

1 融合并行注意力机制的目标检测模型

1.1 下采样和 SE 注意力机制

卷积通过对输入图像进行仿射变化完成特征提取时, 配合不同步长的卷积进行下采样操作对图像进行尺度变换, 完成降维、扩大感受野和多尺度特征信息的提取。YOLOv3 以 DarkNet-53 作为主干网络, 进行五次下采样将输入图像缩小到原来的 1/32 大小。图 1 为进行下采样时通道的变换过程。

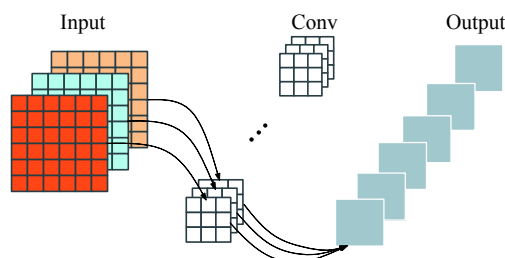


图 1 下采样的通道变换过程

Fig.1 Downsampling channel conversion process

进行下采样时遵循以下仿射变换准则:

$$Y = F(\cdot) * X + b \quad (1)$$

式中: Y 为变换后的特征; X 为变换前的特征; $F(\cdot)$ 为大小为 3×3 、步长为 2 的卷积层、归一化层和激活函数层的总函数; $*$ 为卷积操作; b 为特征通道的偏置参数, 由反向传播进行更新。每经过一次下采样, 特征图尺寸变为原来 1/2, 通道数量变为原来的 2 倍。对于输入图像而言, 目标只占图像的较小部分, 其余均为背景信息, 在进行下采样时, 背景的迭代累积会形成大量冗余信息, 从而淹没部分目标, 导致检测准确率不高。为了能够突出目标特征, SE 模块先对输入特征执行聚合空间维度的挤压操作, 将其压缩为 $1 \times 1 \times C$ 大小, 其中 1 为特征图的高和宽, C 为特征图的通道数。随后, 使用 Sigmoid 函数对压缩后的聚合特征进行激励, 产生各通道对应的权重, 并将权重赋予到原特征图对应的通道中, 完成通道注意力赋值, 生成 SE 模块的输出。图 2 为 SE 模块的工作过程。

SE 模块能够对重要信息产生关注而得到广泛的

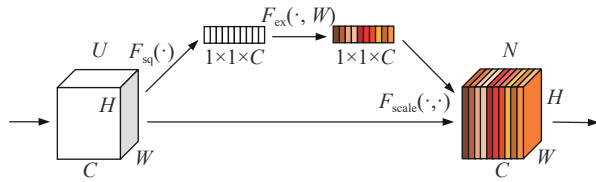


图 2 SE 注意力机制

Fig.2 SE attention mechanism

应用^[13],但是对全部通道进行聚合压缩和激励扩展会造成原始信息的丢失,并且在卷积和下采样操作之外加入注意力模块也会增加模型参数,影响训练效率。为了提升训练速度,一些轻量化的模型^[14-15]选择以牺牲部分检测精度为代价,减少参数量,缩短了训练时间。

1.2 并行注意力机制

为了能够在不增加模型参数的同时提升模型的检测精度和训练效率,文中提出了基于并行注意力机制的卷积模块 (Convolution module based on Parallel Attention mechanism, PAConv),既能够保留原始特征信息,又能对目标予以聚焦和关注。图 3 为并行注意力模块的工作过程。

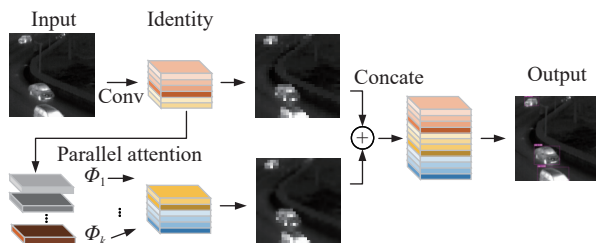


图 3 基于并行注意力机制的卷积模块

Fig.3 Convolution module based on parallel attention mechanism

并行注意力模块先使用常规卷积将输入图像仿射变换为最终输出的 1/2,得到中间层特征:

$$Y' = F'(\cdot) * X \quad (2)$$

式中: Y' 为得到的中间特征; X 为进行仿射变换的输入特征; $F'(\cdot)$ 为大小为 3×3 、步长为 2 的卷积层、归一化层和激活函数层的总函数。为了获得最终输出的 n 个通道,对得到的 Y' 使用线性映射的方式,在不改变特征尺寸的情况下进行加权操作:

$$Y_i'' = \Phi_i(Y_i') \quad \forall i = 1, \dots, \frac{n}{2} \quad (3)$$

式中: Y_i'' 为得到的加权特征层; Φ_i 为对 Y' 中第 i 层进行线性变换时,大小为 $k = 1, s = 1$ 和 $k = 3, s = 1$ 的线性层和 LeakyReLU 激活函数组成的变换函数,函数内部

参数通过反向传播进行更新; Y_i' 为得到的中间特征图的第 i 个特征通道,最后将中间层 Y' 和得到权重赋予的特征 Y'' 组合在一起,聚合成输出特征。与 SE 注意力机制相比,并行注意力模块能够同时进行下采样和注意力聚焦,不会额外引入注意力模块,能够降低模型的空间复杂度和参数,提升模型的训练效率。

1.3 基于并行注意力机制的地面红外目标检测模型

地面背景下的红外目标特征以区域化、连续性分布为主。与红外弱小目标图像相比,具有更多的纹理和细节特征,并且容易受到地面复杂环境背景的干扰。与可见光相比,红外图像为单通道的灰度图,边缘信息不明显,缺乏色彩特征和细节相关性,导致特征提取困难,影响了地面红外目标的检测的准确性。

深度卷积神经网络在进行特征提取时,不同层次的特征图所组成信息的侧重不同,浅层特征通常含有大量的细粒度信息和位置信息,高层的特征含有更抽象的语义信息^[16]。YOLOv1 使用模型的最后一层输出直接对目标进行检测。SSD 使用多尺度的特征输出,并通过 NMS 机制抑制低置信度目标,进行目标检测。为了更加充分地利用不同尺度特征的信息,Lin^[17] 提出了自上而下的金字塔特征融合机制,不仅使用模型的多尺度特征,而且促进了不同尺度特征间信息的融合。自上而下的特征融合方式在目标检测任务中取得了较好的效果,优化了目标检测任务中漏检、误检和检测准确率不高的问题,在诸多检测模型中得到了广泛的应用^[18]。因此,文中在使用基于并行注意力机制的卷积方式优化模型主干网络的基础上,使用四尺度的特征信息进行自上而下的融合,提升目标检测的效果。

为了降低模型的空间复杂度,文中使用基于并行注意力机制的红外目标检测主干网络进行六次下采样。通过下采样缩小特征图尺寸,逐步增大卷积核感受野,完成特征提取,形成 $P1 \sim P6$ 共六个尺度的输出。其中, $P5, P6$ 是使用基于并行注意力机制的卷积模块作为下采样方式得到的。改进后的主干网络能够同时进行下采样和注意力赋予,避免引入额外注意力模块,使用六次下采样进一步降低了模型空间复杂度。

通过基于并行注意机制的主干网络建立了 PAM-YOLO 目标检测模型, PAM-YOLO 利用四个不同尺度的特征信息,进行自上而下的多尺度特征信息融合

时, 首先将最后一层大小为 8×8 的输出定义为 Stage1 模块。将 Stage1 输入预测层的同时, 使用 1×1 卷积进行通道调整和上采样, 得到 16×16 大小的特征图, 再与主干网络中尺寸相同的特征连接, 得到特征融合后的 Stage2 模块。同理, 依次实现 32×32 、 64×64 尺度特征自上而下的融合, 生成相应的特征模块。最终形

成了尺度分别为 8×8 、 16×16 、 32×32 、 64×64 共四个 Stage 的融合特征。再将各 Stage 模块输入到检测网络中, 使用非极大值抑制机制 (Non-Maximum Suppression, NMS) 筛选出与真实框关联度最高的预测框, 将最优预测结果输出, 完成对目标的预测。图 4 为 PAM-YOLO 的工作原理。

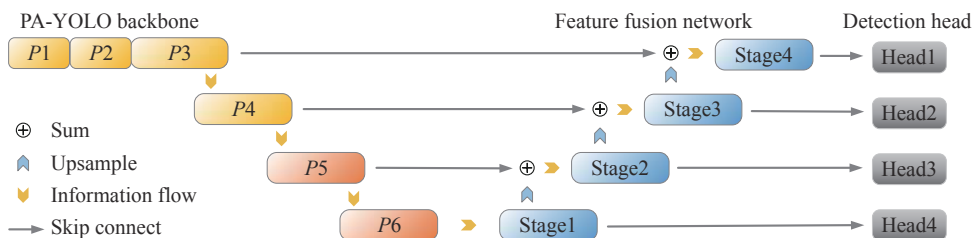


图 4 PAM-YOLO 目标检测模型

Fig.4 PAM-YOLO target detection model

同时, 为了进一步提高模型的反向传播效率与检测精度, 利用 CIoU^[19] 损失函数设计目标检测模型的回归损失函数。与交并比 (Intersection over Union, IoU) 计算方式相比, CIoU 综合考虑目标与锚框之间的距离、重叠率、尺度以及长宽比信息, 可以避免预测框与真实框不相交而导致损失函数梯度为 0 的问题。因此, CIoU 更加符合预测框的回归机制, 使边界框的生成更加稳定。将最小化中心点距离、重叠率、纵横比的惩罚项 CIoU 定义为:

$$f_{CIoU} = 1 - IoU + \frac{\rho^2(b, b^{gt})}{C^2} + \alpha v \quad (4)$$

式中: v 度量长宽比的相似性。定义为:

$$v = \frac{4}{\pi^2} \left(\arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h} \right)^2 \quad (5)$$

α 为权重函数, 定义为:

$$\alpha = \frac{v}{(1 - IoU) + v} \quad (6)$$

CIoU 损失函数定义为:

$$L_{CIoU} = 1 - f_{CIoU} \quad (7)$$

由于样本在进行训练过程中会产生大量易分辨的负样本, 同时大量简单样本和背景信息在训练过程中产生的梯度回传会影响模型对困难样本的分类准确性, 为了增强模型对困难样本的分类能力, 采用焦点损失函数^[20] 作为分类损失函数, 优化网络的分类准

确率。

$$L_{fl} = -\alpha_t (1 - p_t)^t \log(p_t) \quad (8)$$

其中

$$\alpha_t = \begin{cases} \alpha & q = 1 \\ 1 - \alpha & q = 0 \end{cases} \quad (9)$$

p_t 的定义方式如同 α_t ; q 代表指定的类别, 取 0 时为非指定类, 取 1 时为指定类; p 代表 $q = 1$ 时的概率。

2 实验结果与分析

2.1 Infrared-VOC 数据集

目前, 大型公开红外数据集较难获得。笔者课题组使用红外成像仪在不同时段、不同背景、不同气候条件下收集公交车和小汽车的原始红外图像, 制成了含有小汽车图片 1728 幅、公交车图片 572 幅、分辨率为 480×480 的 Infrared-VOC 红外数据集。与共有 9606 幅图片、分为 20 类的标准可见光数据集 VOC2007 相比, VOC2007 数据集中图片数量少于 572 幅的类别有 13 类, 且最少类的图片为 186 幅。通过对比表明, 文中构造的红外数据集可用于模型的训练和性能评估, 训练集、验证集、测试集按 7 : 1 : 2 进行划分。

2.2 评价指标

文中从平均精度均值 (mean Average Precision, mAP) 和每秒检测帧数 (Frames Per Second, FPS) 两个指标出发, 对文中提出的 PAM-YOLO 和几种主流目

标检测模型进行对比实验,并通过单类别的平均精度 (Average Precision, AP)、空间复杂度 (GFLOPS)、训练时间 (Training time) 和参数量 (Parameters) 对文中所提出的 PAM-YOLO 目标检测模型进行消融实验分析。

2.3 基于 Infrared-VOC 数据集的对比试验

表 1 为基于 Infrared-VOC 数据集的不同目标检测算法实验结果。可以看出,在大小为 512×512 分辨率的输入图像下,PAM-YOLO 目标检测模型的 mAP 值达到 82.2%,检测速度达到 46 FPS。在一阶段目标

检测算法中,该模型的 mAP 值比 SSD 高出 9.6%,比 YOLOv3 高出 6.9%,而检测速度基本保持不变;比二阶段目标检测网络中应用比较广泛的 Faster R-CNN 高出 5.1%,检测速度是其 7.1 倍,且空间复杂度仅为 50.6 GFLOPS,即 PAM-YOLO 只需要使用更少的计算资源就可以完成模型的训练。图 5 为训练过程中的损失曲线,PAM-YOLO 与 YOLOv3 均在训练 100 个 Epoch 后趋于稳定,而且 PAM-YOLO 的损失值更低,证明改进后的模型收敛效果更好,具有较好的鲁棒性。

表 1 基于 Infrared-VOC 数据集的不同目标检测模型实验结果

Tab.1 Experimental results of different target detection models based on the Infrared-VOC dataset

| Model | GFLOPS | FPS/frame·s ⁻¹ | mAP@0.5 |
|--------------|--------|---------------------------|---------|
| Faster R-CNN | 224.5 | 6.5 | 77.1% |
| SSD500 | 137 | 38 | 72.6% |
| YOLOv3 | 155.1 | 48.2 | 75.3% |
| PAM-YOLO | 50.6 | 46 | 82.2% |

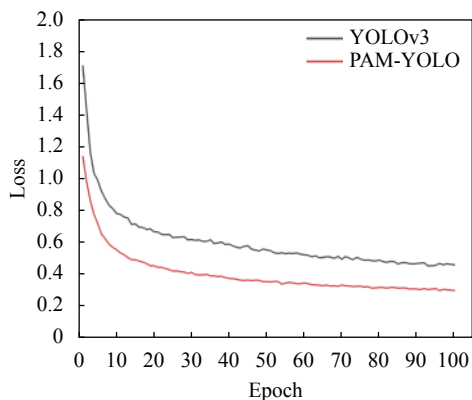


图 5 损失曲线
Fig.5 Loss curve

2.4 在 Infrared-VOC 数据集上的消融实验

为了更好地理解 PAM-YOLO 中各个改进模块对检测效果的影响,下面将研究各个模块对模型的具体贡献。表 2 为基于 Infrared-VOC 数据集以 YOLOv3 为基准模型逐步增加损失函数、并行注意力机制和多尺度特征融合进行的消融实验。

从表 2 可以看出,结合损失函数的基准模型可以使网络对目标的定位和分类更加准确,改进后模型的

mAP 值比 YOLOv3 提升了 1.8%, AP 间的差值也明显减少,由 4.7% 降低到 2.2%,且在训练过程中能够更快地收敛,使模型具有更好的鲁棒性。结合并行注意力机制后,线性变换产生的注意力聚焦和原始通道保留的信息进行互补,使 mAP 值比 YOLOv3 提升了 3.1%,同时降低了模型的空间复杂度,使得训练时间由基准模型的 2.12 h 缩短为 1.48 h,节省了训练时间和训练资源。将并行注意力机制和损失函数同时作用在基准模型上时,mAP 值比基准模型提升了 5.1%,表明损失函数和网络结构的共同优化可以有效提升目标检测模型的检测精度。最后构造的 PAM-YOLO 目标检测模型 mAP 值达到 82.2%,比基准模型 YOLOv3 高出 6.9%,同时空间复杂度仅为基准模型的 32.6%,训练时间缩短为基准模型的 43.7%。证明改进后的目标检测模型在提升目标检测效果的同时有效降低了模型的空间复杂度,在训练阶段减少了训练所需时间,有利于提升模型的训练效率;在部署阶段,有利于提升模型的更新速度。

为了进一步验证并行通道注意力机制与 SE 模块对于模型的实际作用,进行关于这两种注意力机制的

表 2 基于 Infrared-VOC 数据集的消融实验

Tab.2 Ablation experiment based on Infrared-VOC dataset

| YOLOv3 | CIoU loss and focal loss | PACov | PAM-YOLO | GFLOPS | Training time/h | mAP@0.5 | AP@0.5 | |
|--------|--------------------------|-------|----------|--------|-----------------|---------|--------|-------|
| | | | | | | | Car | Bus |
| √ | | | | 155.1 | 2.12 | 75.3% | 72.9% | 77.6% |
| √ | √ | | | 155.1 | 2.10 | 77.1% | 78.2% | 76.0% |
| √ | | √ | | 151.3 | 1.48 | 78.4% | 80.1% | 76.8% |
| √ | √ | √ | | 151.3 | 1.48 | 80.4% | 80.7% | 80.1% |
| √ | √ | √ | √ | 50.6 | 0.927 | 82.2% | 81.4% | 82.9% |

消融实验。以结合损失函数的 YOLOv3 目标检测模型为基准,在相同位置加入不同的注意力模块进行对比实验。

表 3 为 SE 模块和并行注意力模块的对比试验结果。可以发现,SE 模块通过聚合激励操作,为特征通道进行加权赋值,产生聚焦作用,与基准模型相比,mAP 值提升了 0.9%,但空间复杂度比基准模型增加

了 33.01%,参数量几乎保持不变。与基准模型相比,使用并行通道注意力机制的目标检测模型的 mAP 值提升了 3.3%,参数量减少了 4.69%,空间复杂度比基准模型降低了 2.4%,训练时间是基准模型的 70.5%,说明所提出的并行注意力机制能够在不增加空间复杂度的基础上对目标产生注意力,提升目标检测效果。

表 3 SE 模块和 PAM 模块的对比试验

Tab.3 Comparison test of SE module and PAM module

| YOLOv3 with loss | SE module | PAM module | GFLOPS | Training time/h | Parameters/M | mAP@0.5 |
|------------------|-----------|------------|--------|-----------------|--------------|---------|
| √ | | | 155.1 | 2.10 | 59.66 | 77.1% |
| √ | √ | | 206.3 | 3.14 | 59.67 | 78.0% |
| √ | | √ | 151.3 | 1.48 | 56.86 | 80.4% |

与使用 SE 模块的目标检测模型相比,使用并行通道注意力机制的目标检测模型的 mAP 提高了 2.4%,参数量减少了 4.71%,空间复杂度是使用 SE 模块的 73.4%,训练时间是使用 SE 模块的 47.1%。意味着该模型可以在同等计算资源下得到更快的训练并能取得更好的检测效果。

上述消融实验证明文中提出的并行注意力机制提高了模型的运行效率和检测精度,并未削弱地面红外车辆目标检测任务的注意力效果,同时说明文中提出的并行注意力机制更适用于地面红外车辆目标的检测。

图 6 为 YOLOv3 算法与文中改进算法在 Infrared-VOC 数据集上的检测效果对比图,其中,图 6(a)为输

入的原图,图 6(b)为 YOLOv3 算法的检测效果图,图 6(c)为文中改进算法的检测效果图。

YOLOv3 将第 2 行图中右下方的摩托车误检为小汽车,对第 3、4 行图中边缘的不完整车辆产生了漏检;在第 5、6 行图中,PAM-YOLO 能够对单张图像内相互遮挡和被其他物体遮挡的目标进行有效的检测,而 YOLOv3 则产生了较多漏检的情况。从对比图中可以看出,改进后的 PAM-YOLO 目标检测模型对于目标检测的整体置信度和检测效果均高于 YOLOv3。证明改进后的检测模型具有较高的检测置信度和较好的检测稳定性,能够实现对面背景下的红外目标进行准确地检测。

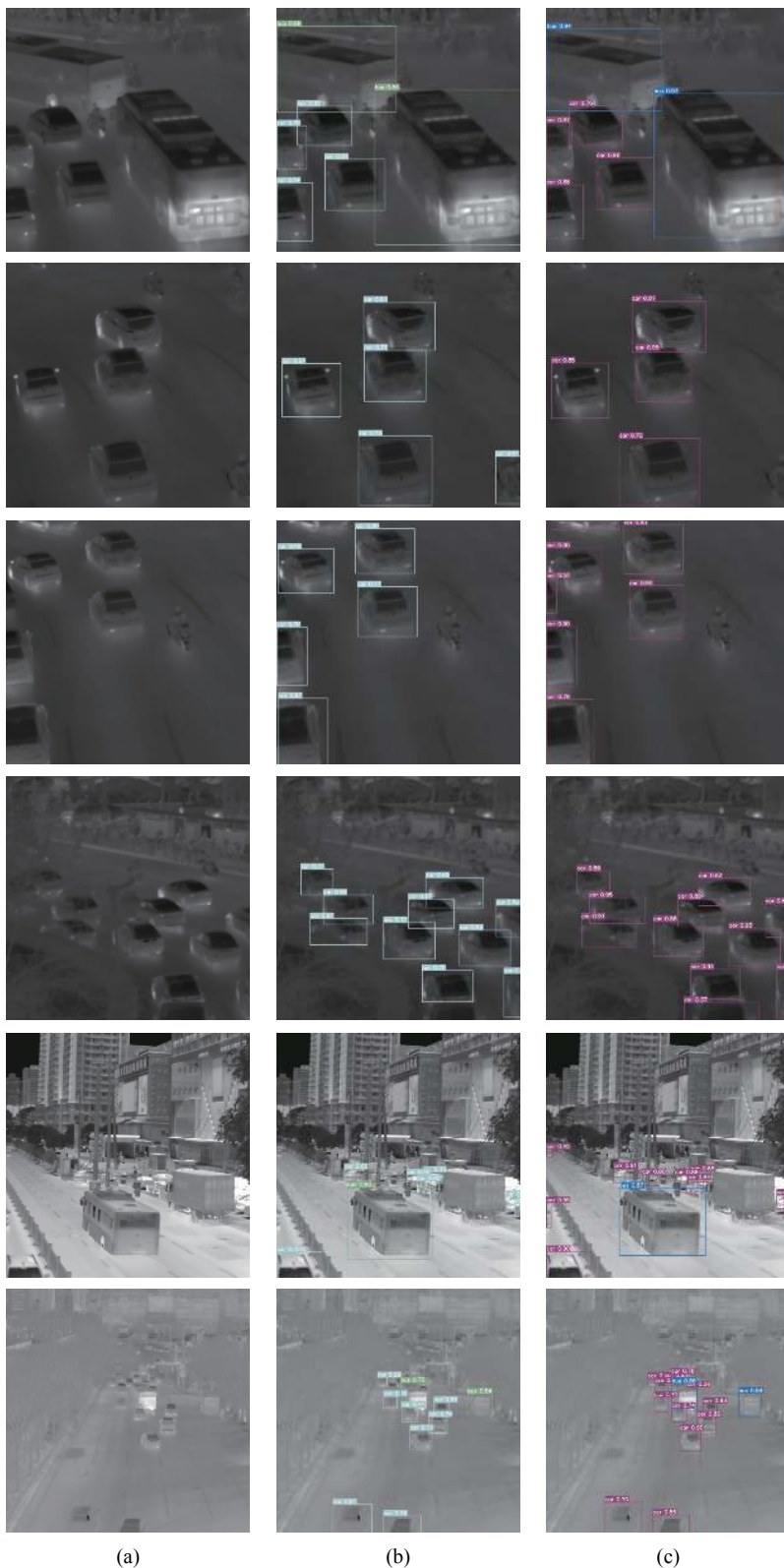


图 6 目标检测结果对比图。(a) 原图; (b) YOLOv3; (c) PAM-YOLO

Fig.6 Comparison chart of target detection results. (a) Original image; (b) YOLOv3; (c) PAM-YOLO

3 结 论

基于并行注意力机制的红外目标检测算法,通过

并行注意力机制对待检测目标进行聚焦和加权,融合四个不同尺度的特征促进不同层次间信息的互补,使

用六次下采样减少模型的空间复杂度,并结合损失函数提高分类与回归的精度,实现了提升模型的目标检测效果和优化模型结构的目的。最终构造的 PAM-YOLO 目标检测模型对地面背景下的红外目标进行检测时,在使用并行注意力机制突出目标信息提升检测精度的同时,大幅减少了模型的空间复杂度,缩短了模型所需的训练时间,并且提升了模型的训练效率和部署后的更新速度,为提升地面背景下的红外目标检测和模型训练效率提供借鉴和参考。考虑到硬件部署的需求,后续可以对文中模型在保持检测精度和检测速度的基础上进行剪枝和轻量化处理,以压缩模型的尺寸,进一步降低模型对部署端硬件性能的要求,提升模型的适用性。

参考文献:

- [1] Li Weipeng, Lu Ruitao, Yang Xiaogang, et al. An improved semi-supervised transfer learning method for infrared object detection neural network [J]. *Infrared and Laser Engineering*, 2021, 50(3): 20200511. (in Chinese)
- [2] Lu Fuxing, Chen Xin, Chen Guilin, et al. Dim and small target detection based on background adaptive multi-feature fusion [J]. *Infrared and Laser Engineering*, 2019, 48(3): 0326002. (in Chinese)
- [3] Zhang Lei, Zhu Shuai, Liu Tianyu, et al. Tracking of dense group targets based on motion grouping [J]. *Infrared and Laser Engineering*, 2020, 49(11): 20200284. (in Chinese)
- [4] Chen Faling, Ding Qinghai, Luo Haibo, et al. Anti-occlusion real time target tracking algorithm employing spatio-temporal context [J]. *Infrared and Laser Engineering*, 2021, 50(1): 20200105. (in Chinese)
- [5] Zhao Xiaofeng, Xu Mingyang, Wang Danpiao, et al. Infrared camouflage detection method for special vehicles based on improved SSD [J]. *Infrared and Laser Engineering*, 2019, 48(11): 1104003. (in Chinese)
- [6] Zhao X, Xu Y, Wu F, et al. IYOLO: Multi-scale infrared target detection method based on bidirectional feature fusion [J]. *Journal of Physics: Conference Series*, 2021, 1873(1): 12-20.
- [7] Zhang Chi, Tan Nanlin, Li Guozheng, et al. Pedestrian detection algorithm for infrared image based on multi-level features [J]. *Computer Engineering*, 2020, 46(4): 260-265. (in Chinese)
- [8] Chen Ming, Zhao Lianfei, Yuan Limin, et al. Insulator detection method based on feature selection YOLOv3 network [J]. *Infrared and Laser Engineering*, 2020, 49(S2): 20200401. (in Chinese)
- [9] Wei Shuigen, Wang Chengwei, Chen Zhen. Infrared dim target detection based on human visual mechanism[J]. *Acta Photonica Sinica*, 2021, 50(1): 0110001. (in Chinese)
- [10] Liu Xu, Cui Wennan. Infrared-image-based detection of dim and small targets using human visual contrast mechanism [J]. *Infrared Technology*, 2020, 42(6): 559-565. (in Chinese)
- [11] Liu Junming, Meng Weihua. Infrared small target detection based on fully convolutional neural network and visual saliency [J]. *Acta Photonica Sinica*, 2020, 49(7): 0710003. (in Chinese)
- [12] Hu J, Shen L, Sun G. Squeeze-and-excitation networks [C]//Proceedings of the IEEE Conference on Computer Vision And Pattern Recognition, 2018: 7132-7141.
- [13] Cai Tujian, Peng Xiaoyu, Shi Yapeng, et al. Channel attention and residual concatenation network for image super-resolution [J]. *Optics and Precision Engineering*, 2021, 29(1): 142-151. (in Chinese)
- [14] Chollet F. Xception: Deep learning with depthwise separable convolutions [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 1800-1807.
- [15] Han K, Wang Y, Tian Q, et al. Ghostnet: More features from cheap operations [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 1580-1589.
- [16] Krizhevsky A, Sutskever I, Hinton G E. Image net classification with deep convolutional neural networks [J]. *Communications of the ACM*, 2017, 60(6): 84-90.
- [17] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE Computer Society, 2017.
- [18] Qiu Xiaohua, Li Min, Deng Guangmang, et al. Multi-layer convolutional features fusion for dual-band decision-level ship recognition [J]. *Optics and Precision Engineering*, 2021, 29(1): 183-190. (in Chinese)
- [19] Zheng Z, 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, 2020, 34(7): 12993-13000.
- [20] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection [C]//Proceedings of the IEEE International Conference on Computer Vision, 2017: 2980-2988.