

基于深度学习的轻量化红外弱小车辆目标检测算法研究

蔡仁昊 程宁 彭志勇 董施泽 安建民 金钢

Lightweight infrared dim vehicle target detection algorithm based on deep learning

Cai Renhao, Cheng Ning, Peng Zhiyong, Dong Shize, An Jianmin, Jin Gang

在线阅读 View online: <https://doi.org/10.3788/IRLA20220253>

您可能感兴趣的其他文章

Articles you may be interested in

非合作目标局部特征识别轻量化特征融合网络设计

Lightweight feature fusion network design for local feature recognition of non-cooperative target

红外与激光工程. 2020, 49(7): 20200170 <https://doi.org/10.3788/IRLA20200170>

基于改进SSD的轻量化小目标检测算法

A lightweight small object detection algorithm based on improved SSD

红外与激光工程. 2018, 47(7): 703005 <https://doi.org/10.3788/IRLA201847.0703005>

基于改进SSD的特种车辆红外伪装检测方法

Infrared camouflage detection method for special vehicles based on improved SSD

红外与激光工程. 2019, 48(11): 1104003 <https://doi.org/10.3788/IRLA201948.1104003>

基于邻域特征的红外低慢小目标检测

Infrared low, slow and small target detection based on neighborhood characteristics

红外与激光工程. 2019, 48(S1): 174 <https://doi.org/10.3788/IRLA201948.S128002>

基于视觉显著性和目标置信度的红外车辆检测技术

Infrared vehicle detection based on visual saliency and target confidence

红外与激光工程. 2017, 46(6): 604005 <https://doi.org/10.3788/IRLA201746.0604005>

二维非重构压缩感知自适应目标检测算法

Two-dimensional non-reconstruction compressive sensing adaptive target detection algorithm

红外与激光工程. 2019, 48(1): 126001 <https://doi.org/10.3788/IRLA201948.0126001>

基于深度学习的轻量化红外弱小车辆目标检测算法研究

蔡仁昊¹, 程 宁¹, 彭志勇¹, 董施泽¹, 安建民², 金 钢³

- (1. 天津航技术物理研究所, 天津 300308;
2. 空装驻天津地区第三军事代表室, 天津 300308;
3. 天津大学, 天津 300072)

摘要: 伴随高速飞行器的不断发展, 目标检测识别作为精确制导的关键一环, 需要更高实时性、高准确地进行目标定位和识别。当前, 针对装甲车辆、车辆阵地等时间敏感目标精确检测识别的需求日益迫切, 深度学习算法在特征提取及分类器设计上具备优势。文中以特定复杂背景下的小尺寸红外车辆目标为研究对象, 针对样本数据少、平台资源受限、实时性要求高、检测精度高等需求, 开展基于红外弱小车辆目标检测识别的轻量化深度学习算法研究。项目基于 YOLOv5 算法进行轻量化剪裁, 减小模型的结构, 提高实时性; 提出了混合域注意力机制模块 EPA, 该模块通过不降维的局部跨信道交互策略使算法更快速有效地关注重要通道, 抑制无效通道, 并将通道注意力机制与空间注意力机制结合, 使得算法更关注与目标相关的像素信息。提出了残差密集注意模块 (RDAB), 该模块由密集残差块与注意力机制 EPA 构成, 通过密集卷积层来提取充分的局部特征, 通过注意力机制获取更有效的通道与像素信息, 可以使得算法以较小的模型结构获得较好的检测效果。运用设计的网络对数据增广后的小尺寸红外车辆目标数据进行检测识别, 并与多种典型算法进行对比实验。由实验结果可知, 文中提出的 JH-YOLOv5-RDAB 网络检测识别效果优于其他网络, 权重大小仅为 6.6 MB, 仅为 YOLOv5s 算法模型权重的一半, 但算法检测效果更优, 与 93.7 MB 的 YOLOv5l 算法的检测效果接近, mAP50 达到 95.1%。实验结果表明: 该网络在红外弱小车辆目标检测上的优越性和可行性。

关键词: 红外车辆目标; 目标检测; 轻量化; 注意力机制; 密集卷积网络

中图分类号: TP391 **文献标志码:** A **DOI:** 10.3788/IRLA20220253

Lightweight infrared dim vehicle target detection algorithm based on deep learning

Cai Renhao¹, Cheng Ning¹, Peng Zhiyong¹, Dong Shize¹, An Jianmin², Jin Gang³

- (1. Tianjin Jinhang Institute of Technical Physics, Tianjin 300308, China;
2. The Third Military Representative Office in Tianjin, Tianjin 300308, China;
3. Tianjin University, Tianjin 300072, China)

Abstract: With the continuous development of high-speed aircraft, target detection and recognition, as a key part of precision guidance, requires higher real-time and high-accuracy target positioning and recognition. At present, the need for accurate detection and identification of time-sensitive targets such as armored vehicles and vehicle positions is increasingly urgent. Deep learning algorithms have advantages in feature extraction and classifier design. This paper takes the small-sized infrared vehicle target under a specific complex background as the research object, and develops a lightweight deep learning algorithm based on infrared dim vehicle target

收稿日期: 2022-03-20; 修订日期: 2022-04-25

作者简介: 蔡仁昊, 女, 工程师, 硕士, 主要研究方向为智能光电探测成像等。

detection and recognition to meet the needs of less sample data, limited platform resources, high real-time requirements, and high detection accuracy. The project is light-weight cut based on the YOLOv5 algorithm, reduce the structure of the model and improve the real-time performance; a hybrid domain attention mechanism module EPA is proposed, which enables the algorithm to focus on important channels more quickly and effectively through a local cross-channel interaction strategy without dimensionality reduction. Suppressing invalid channels and combining the channel attention mechanism with the spatial attention mechanism makes the algorithm pay more attention to the pixel information related to the target. The Residual Dense Attention Module (RDAB) is proposed, which is composed of dense residual blocks and attention mechanism EPA. It extracts sufficient local features through dense convolutional layers, and obtains more effective channel and pixel information through attention mechanism, which can make the algorithm obtain better detection effect. Detect and identify the small-size infrared vehicle target data after data augmentation, and compare experiments with a variety of typical algorithms. It can be seen from the experimental results that the detection and recognition effect of the JH-YOLOv5-RDAB network proposed in this paper is better than other networks, and the weight size is only 6.6 MB, which is only half of the weight of the YOLOv5s algorithm model, but the algorithm detection effect is better, and the detection effect of the algorithm is close YOLOv5l whose weight size is 93.7 MB, with mAP50 reaching 95.1%. The experimental results show the superiority and feasibility of this network in infrared dim vehicle target detection.

Key words: infrared vehicle target; target detection; lightweight; attention mechanism; dense convolutional network

0 引言

伴随新军事背景,高速飞行器不断向更高、更快、更准的方向发展。目标检测识别作为精确制导的关键一环,需要更高实时性、高准确性地进行目标定位和识别,以适应飞行器的高速背景。另一方面,当前针对装甲车辆、车辆阵地等小尺寸时敏目标的精确检测需求日益迫切,与传统图像特征手动提取方法相比,深度学习算法在特征提取及分类器设计上具备优势^[1]。而且随着红外成像技术的逐渐成熟,基于红外目标的检测方法已被广泛应用于诸多领域中。其中,红外弱小目标的检测尤其重要。

传统的目标检测算法一般先使用 HOG 或 DPM 等方法提取特征,然后采用 SVM 模型等分类器进行识别^[2]。而深度学习算法将图像映射到高维空间自动提取特征并进行分类,对目标特征提取更全面与深入。

卷积神经网络 (CNN) 的快速发展给基于深度学习目标检测算法的风靡提供了基础。2012 年, Krizhevsky^[3] 等提出的 AlexNet 模型使 ImageNet 数据集的分类准确率得到明显提高; 2014 年, Google^[4] 团队提出的 GoogLeNet 模型通过引入 Inception 模块增加网络

宽度,从而提高模型的表达能力; 同年 Simonyan^[5] 团队提出的 VGG 模型通过多个小感受野卷积核堆叠的方式代替大感受野的卷积核,从而减少网络参数; 2015 年, He^[6] 等提出了 ResNet 模型,通过引入残差块结构解决了网络因为层数增加导致的梯度消失和梯度爆炸现象。基于 CNN 构建的深度学习目标检测发展主要集中在两个方向: 一阶段算法^[7]和两阶段算法^[8], 如图 1 所示。图 1 中的目标检测方法为普通目标检测神经网络方法,可以实现高精度目标检测,但缺点是运行消耗计算资源大。

文中开展基于红外弱小车辆目标检测的轻量化深度学习算法研究,平台资源受限,需使用轻量级神经网络模型。轻量级神经网络模型虽然检测精度相对一般,但消耗计算资源较少。2020 年, Bochkovskiy 等提出轻量级模型 YOLOv4-tiny 目标检测算法,模型权重大小只有 23.1 MB; 同年,超轻量级的目标检测模型 YOLO-Fastest 被提出,其尺寸仅为 1.3 MB,是目前最小的 YOLO^[9] 模型; 2021 年,研究者就提出了一种新的基于轻量级 CNN 的目标检测模型,即基于 YOLOv3-Tiny 的 Micro-YOLO,它在保持检测性能的同时显著减少了参数数量和计算成本; 2022 年,罗禹

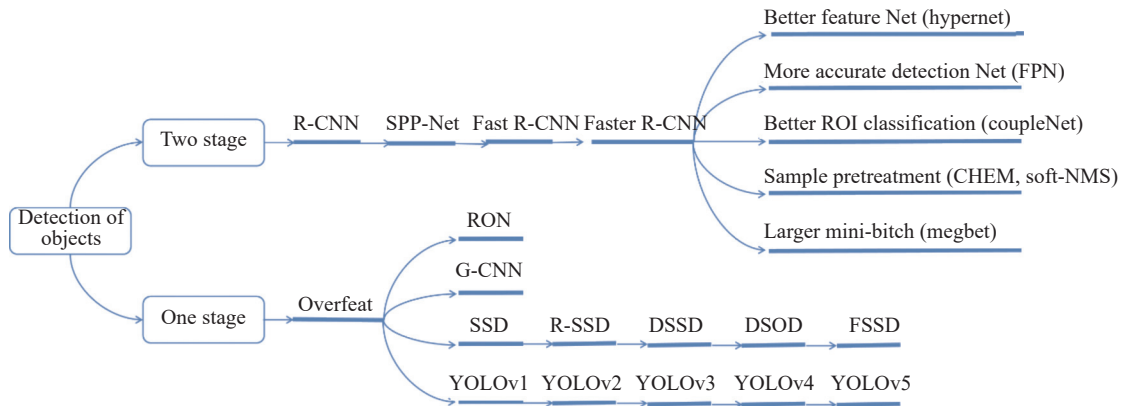


图 1 近年深度学习目标检测算法发展趋势

Fig.1 Development trend of deep learning object detection algorithms in recent years

杰等以 YOLOv4 为网络基础框架,采用轻量级 Mobile-Net 作为特征提取网络,提出了一种基于自适应空间特征融合的轻量化目标检测算法,可以部署到移动与嵌入设备实现实时检测。

目前的轻量化目标检测算法虽然在模型权重实现轻量,但算法检测效果还逊色于普通目标检测神经网络方法。而且这类轻量化目标检测算法对红外弱小目标检测效果极差。

与普通图像目标相比,红外弱小目标不是十分明显,而且只有灰度信息,两者存在巨大差异,因此用于普通目标的基于 CNN 的目标检测方法并不适用于红外弱小目标的检测。文中以特定的复杂背景下的小尺寸红外车辆目标为研究对象,不仅平台资源受限,样本数据少,而且检测精度要求高。现有的轻量化算法以及普通目标检测神经网络方法都难以满足要求。如何设计一种轻量化网络,针对红外小尺寸车辆目标,能够提取更充分的特征,提高识别准确率成为了亟待解决的问题。

针对平台资源受限,红外弱小目标不是十分明显,而且只有灰度信息的问题,文中提出了一种基于 YOLOv5 s 网络进行轻量化处理、融合注意力机制和残差密集块的目标检测识别算法 JH-YOLOv5-RDAB。该网络主要在平台资源受限情况下,对红外小尺寸目标进行检测。通过对网络轻量化剪裁,缩小结构,提高实时性;为更好的将红外弱小目标结构化特征引入 CNN 网络中,提出了混合域注意力机制模块 EPA 与基于注意力机制的残差密集块模块 (RDAB)。EPA 模块通过不降维的局部跨信道交互策略使算法更快

速有效地关注重要通道,抑制无效通道,并将通道注意力机制与空间注意力机制结合,使得算法更关注与目标相关的像素信息;RDAB 模块由密集残差块与注意力机制 EPA 构成,该模块通过密集卷积层来提取充分的局部特征,允许通过多个局部残差连接绕过较不重要的信息,使得检测算法更关注与目标相关的通道与像素位置信息,从而使得算法以较小的模型结构获得较好的检测效果。

1 红外弱小车辆目标样本建立及预处理

项目选用机载挂飞采集和无人机采集的中/长波红外车辆目标。由于典型车辆时敏目标尺寸为 10×12,目标不是十分明显,很难被肉眼发现,可视为弱小目标。笔者采集背景覆盖了沙漠、城市、田野、公路、乡村等场景,同时具备云雾干扰场景。对目标进行了多角度采集,但实验中飞行高度较固定且样本数量受限,采集图像数量不足,因此需进行数据增广处理。另一方面,受探测器限制,采集图像分辨率为 640×512,同幅图像中包含 1~25 个典型车辆时敏目标,并且在场景中存在小型建筑、树木、石墩、变压器等非典型目标混杂,具备较高的检测识别难度。

数据作为深度学习的驱动力,对于模型的训练至关重要。充足的训练数据不仅可以缓解模型在训练时的过拟合问题,而且可以进一步扩大参数搜索空间,帮助模型进一步朝着全局最优解优化。为在一定程度上模拟高速飞行其实时图像,对其进行分析,如表 1 所示。

在表中图像处理方法的基础上,为解决红外车辆

数据样本不足的问题,文中进行了单数据变形和多数据混合处理。对于单幅图像,进行了亮度变化、对比度变化、旋转、平移、放缩、翻转、裁剪、随机模糊、

边缘模糊、随机遮挡等处理,并且将多张图像及进行随机拼接操作,以丰富数据,扩充样本同时防止网络出现过拟合现象,如图 2 所示。

表 1 单幅图像的模拟图像处理方法

Tab.1 Analog image processing method for single image

Flight live image	Analog image processing
Aircraft falling at high speed	Image magnification under fixed field of view
Aircraft level flight	Image translation
Aircraft rotation	Image rotation at various angles
Aircraft shaking	Image translation
Infrared imagers are affected by temperature and weather	Image contrast, brightness changes
Aerodynamic effect of aircraft flying at high speed	Image random blur, edge blur
Aircraft is disturbed	Image random occlusion

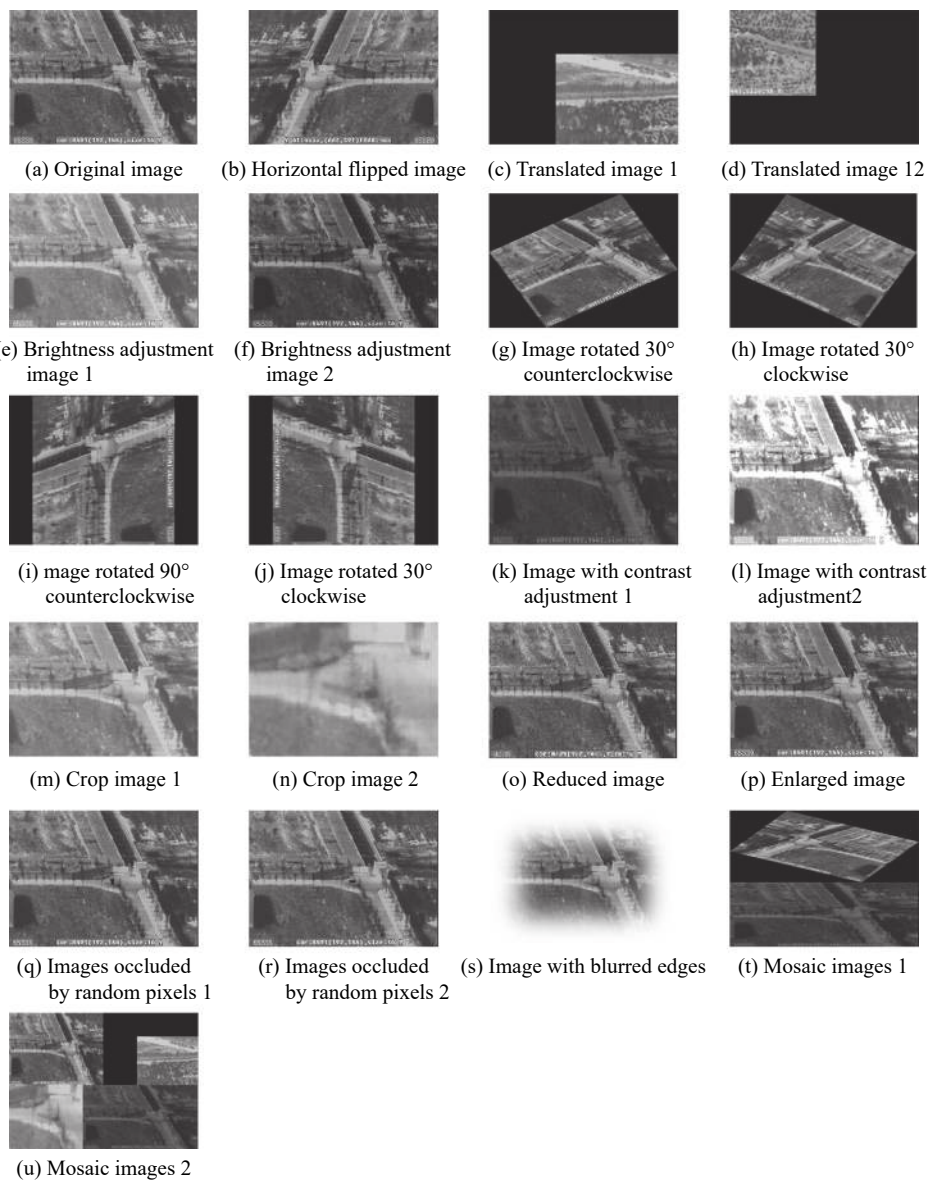


图 2 数据增广效果图

Fig.2 Effect drawing of data expansion

2 红外弱小车辆目标的轻量化检测算法设计

针对平台资源受限, 红外弱小目标不是十分明显, 而且只有灰度信息的问题, 文中以 YOLOv5 s 算法为基础, 首先通过调整系数 $depth_multiple$ 和系数 $width_multiple$ 将 YOLOv5 s 网络进行轻量化剪裁, 然后通过使用混合域注意力模块 EPA 改进 YOLOv5 算法原有的 SPP 模块、CSP 模块, 通过残差密集注意力模块 RDAB 改进 YOLOv5 算法原有的 SPP 模块、Focus 模块来对 YOLOv5 算法进行改进, 来提高红外弱小目标检测效果, 改进后的目标检测算法文中称为 JH-YOLOv5-RDAB 算法。其网络结构如图 3 所示。该部分主要包括基于 YOLOv5 的轻量化网络设计, 基于混合域注意力模块 (EPA) 设计, 基于注意力机制的残差密集块模块 (RDAB) 设计三部分内容。

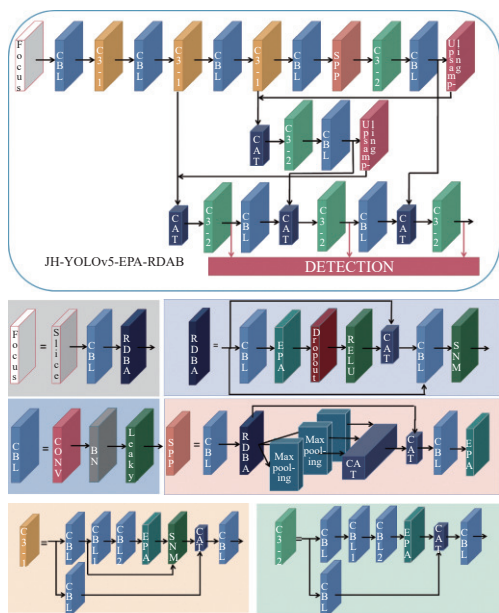


图 3 JH-YOLOv5-RDAB 结构图

Fig.3 Structure diagram of JH-YOLOv5-RDAB

2.1 基于 YOLOv5 的轻量化网络设计

神经网络由于神经元之间连接的数百万甚至更多的权重, 模型存储的是权重参数并全部参与计算, 这对嵌入式平台自身的计算资源也是一项巨大的挑战。由于平台资源受限, 需要设计一个轻量化网络, 本节以 YOLOv5 算法为基础进行轻量化网络设计。

2020 年, 由 Ultralytics 提出的 YOLOv5 算法自面世以来就受到广泛关注, 分为输入端、Backbone、Neck、

Prediction 四个部分。输入端部分算法增加了 Mosaic 数据增强和自适应锚框计算, Backbone 部分增加了 Focus 结构和 CSP 结构, Neck 部分采用 FPN+PAN 结构, Prediction 部分使用 GIOU_Loss 损失函数。YOLOv5 算法共有四种不同大小的模型: YOLOv5 s, YOLOv5 m, YOLOv5 l 和 YOLOv5 x。在 COCO 数据集上得到了较高的性能指标, 其中 YOLOv5 s 网络模型权重最小, 速度最快, 但 mAP 精度也最低, 适合大目标检测。YOLOv5 s 网络结构如图 4 所示。

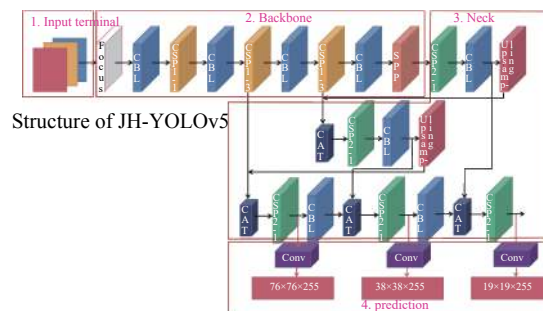


图 4 YOLOv5 s 结构图

Fig.4 Structure of YOLOv5 s

为了在轻量化与检测效果间找到一个较好的平衡, 本项目基于 YOLOv5 s 算法进行改进。其最突出的特点是结构灵活, 其引入了 $depth_multiple$ 系数来控制模型的深度和 $width_multiple$ 系数来控制卷积核的个数。系数 $depth_multiple$ 用于 backbone 中的 $number \neq 1$ 的情况下 ($number=1$ 则表示功能背景层, 比如说下采样 Conv、Focus、SPP 等), 即在 Bottleneck 层使用以控制模型的深度, 例如 YOLOv5 s 中设置为 0.33, 那么假设 YOLOv5 l 中有三个 Bottleneck, 则 YOLOv5 s 中就只有三个 Bottleneck。系数 $width_multiple$ 是用于设置卷积层的卷积核个数, 例如 YOLOv5 s 设置为 0.5, 卷积核的个数都变成了设置的一半。

同一网络结构下参数设置越小检测精度越低, 但为了达到极轻量化需求, 选择牺牲一定了检测效果来获得极低的模型尺寸。文中选用的系数 $depth_multiple$ 和系数 $width_multiple$ 分别为 0.20 和 0.25, 在 YOLOv5 s 的基础上又减少了一半的层数, 后续几节中将通过增加算子模块进一步提升检测效果。

2.2 基于混合域注意力模块 (EPA) 设计

注意力机制 (Attention Mechanism) 是一种通过模仿人脑来实现从大量输入信息中重点处理有用信息

并忽略无关信息的机制^[10],主要分为通道注意力和空间注意力。为了更准确地从红外弱小目标中提取有用信息,本节提出了一个基于混合域注意力模块(EPA),并将其融合到YOLOv5算法原有的SPP模块、CSP模块中,使得算法在有效捕获跨通道交互的信息的同时,更加关注与红外弱小目标相关的像素信息等信息特征。融合后的SPP模块、CSP模块结构图分别如图5与图6所示。

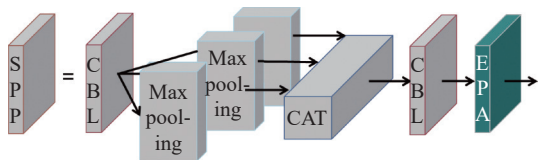


图 5 融合 EPA 模块后的 SPP 模块结构图

Fig.5 Structure diagram of SPP module after merging EPA module

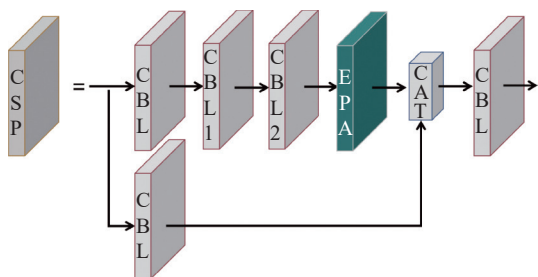


图 6 融合 EPA 模块后的 CSP 模块结构图

Fig.6 Structure diagram of CSP module after merging EPA module

近年来,将通道注意力引入卷积块,在性能改进方面显示出巨大潜力。以SENet^[11]为代表,通过学习每个卷积块的通道注意力,获得明显的性能增益。之后一些研究通过更复杂的通道依赖或结合额外的空间注意来改进,但伴随较高的模型复杂度计算负担。Qilong Wang^[12]等在CVPR2020上提出了一种针对深度CNN的高效通道注意(ECA)模块,避免了降维,有效捕获了跨通道交互的信息,其结构如图7所示。

ECA模块是通道域的注意力机制,通道注意力机

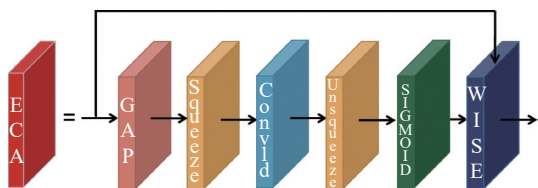


图 7 ECA 模块结构图

Fig.7 Structure diagram of ECA module

制对一个通道内的信息直接全局平均池化,其本质是给高权重在重点关注的通道上,以实现对更重要特征的最优关注,而忽略每一个通道内的局部信息,忽略关注空间上的重要位置,考虑目标像素与非目标像素的不同,这种做法其实也是比较暴力的行为。

因此在文中,对ECA模块进行改进,提出一种混合域注意力机制模块EPA,使得算法在有效捕获跨通道交互的信息的同时,更加关注与目标相关的像素信息等信息特征^[13]。其结构如图8所示。

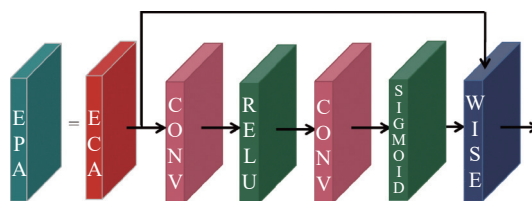


图 8 EPA 模块结构图

Fig.8 Structure diagram of EPA module

通过图8可知,EPA模块在ECA模块之后使用RELU激活函数与Sigmoid激活函数将上一步输入直接输入到两个卷积层中:

$$PA = \sigma(\text{Conv}(\delta(\text{Conv}(F^*)))) \quad (1)$$

式中: F^* 为上一步的输出; Conv 为卷积层; δ 为RELU激活函数; σ 为Sigmoid激活函数。

然后,对输入F和PA使用元素乘法:

$$\bar{F} = F * \otimes PA \quad (2)$$

2.3 基于注意力机制的残差密集块模块(RDAB)设计

红外图像类似于可见光的灰度图像,数值在[0, 16383]之间。由于成像体制的差别,两者图像在图像灰度上差异较大,但在轮廓结构上具有一致性,如图9所示。与3通道的可见光图像相比,单通道的红外图像在浅层网络的特征提取尤为重要。同时,由于需要检测的目标为弱小目标,一般卷积神经网络无法应用于小尺寸目标的识别。为此,在本节提出了可以在不同层次上突出弱小目标,能够增加对红外弱小目标特性的特征提取的基于注意力机制的残差密集块模块(Residual dense attention block, RDAB),并将其融合到YOLOv5算法原有的Focus模块与SPP模块中,使得算法可以较好的提取红外弱小目标结构化特征,并对弱小目标的尺度变化具有鲁棒性。融合后的Focus模块、SPP模块结构图分别如图10与图11所示。

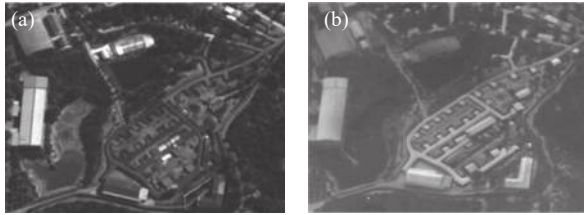


图 9 同一场景可见光图像(左)和红外图像(右)

Fig.9 Visible light image (left) and infrared image (right) of the same scene

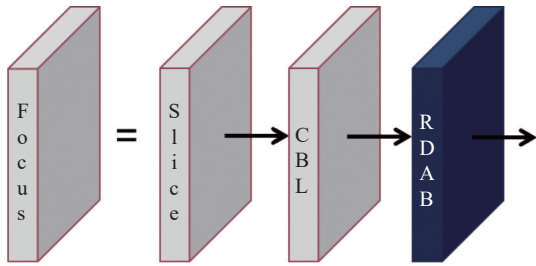


图 10 融合 RDAB 模块后的 Focus 模块结构图

Fig.10 Structure diagram of Focus module after merging RDAB module

RDAB 模块将残差密集块 (RDB) 与混合域注意力模块 (EPA) 相结合。RDAB 结构如图 12 所示。

残差密集块 RDB 的概念来自单幅图像超分辨率 (SISR)。SISR 旨在低分辨率 (LR) 测量的基础上生

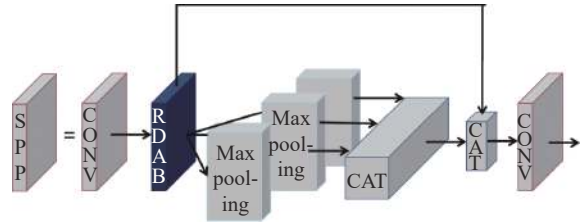


图 11 融合 RDAB 模块后的 SPP 模块结构图

Fig.11 Structure diagram of SPP module after merging RDAB module

成视觉良好的高分辨率 (HR) 图像。Yulun Zhang^[14]等在 CVPR2018 提出了一种由残差密集块构成的残差密集网络来从原图生成高分辨率图像。该网络结合残差网络与密集连接网络的特性充分利用原始 LR 图像的所有分层特征,因而能重构出高质量的图像, RDB 的结构如图 13 所示。

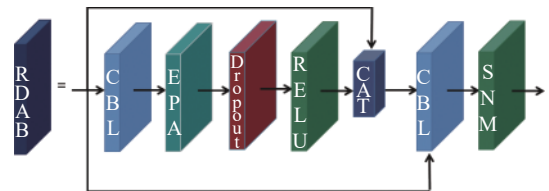


图 12 残差密集注意模块结构图

Fig.12 Structure diagram of RDAB

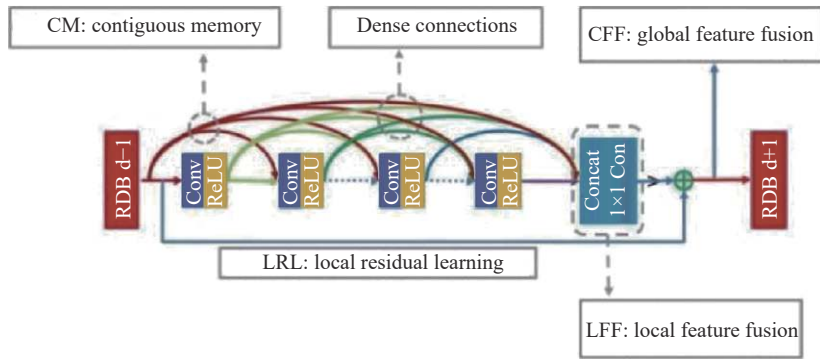


图 13 残差密集块的结构图

Fig.13 Structure diagram of RDB

RDB 通过密集卷积层来提取充分的局部特征,不仅可以通过连续记忆 (CM) 机制从前一个 RDB 读取状态,还可以通过局部密集连接充分利用其中的所有层。

混合域注意力模块 EPA 结构详见 2.2 节。由图 8 可知,在残差密集块结构的 1*1 Conv 后面加入混合域注意力模块 EPA。基于注意力机制的 RDAB 允许通过多个局部残差连接绕过较不重要的信息,使得检

测算法更关注与目标相关的通道与像素位置信息。

3 实验

3.1 实验数据和环境搭建

项目数据来源是机载挂飞采集和无人机采集的中/长波红外弱小车辆目标。初始数据为标注存在一定问题的红外车辆目标图像 5 023 张。首先进行数据

清洗保留 4986 张, 随机选取其中的 1000 张留存作为检测集, 将剩下的 3986 张样本图像作为原始训练集。按照 1 节中的数据增广方法将原始训练集进行样本扩充, 扩充后为 51818 张, 随机选取 5% 即 2590 张作为验证集, 作为训练使用, 如表 2 所示。

表 2 红外弱小车辆图像数据集分布情况

Tab.2 Distribution of infrared small and weak vehicle image dataset

Target type	[car]
Target scene	Desert, city, field, highway, village
Data Augmentation Method	Brightness change, contrast change, rotation, translation, scaling, flipping, clipping, splicing
Original data Set	5023 (Before treatment) 4986 (After treatment)
Data augmentation	1000
Training set	49228
Validation set	2590

表 3 实验环境具体参数

Tab.3 Specific parameters of the experimental environment

Experimental system	Ubuntu18.04
CPU	Inter Xeon Gold 6133
GPU	NVIDIA TITAN RTX ×4
Memory	512 GB
Development environment	Python3.7
Deep learning framework	Pytorch
CUDA	10.2
cuDNN	7.5.3

表 4 JH-YOLOv5-RDAB 与典型算法网络实验效果对比

Tab.4 Comparison of experimental results between JH-YOLOv5-RDAB and typical algorithm networks

Algorithm	Layers	Parameters	Size (Semi precision quantization)	mAP50	GFLOPs	Single test time/ms
YOLOv3	261	61497430	123.4M	81.6	154.9	5.3
YOLOv4-tiny	99	5874116	23.6M	88.2	16.1	1.8
YOLOv4	488	63937686	256.3M	94.6	141.4	8.7
YOLOv5 s	283	7063542	14.4M	93.4	16.3	2.34
YOLOv5 m	391	21056406	42.5M	94.5	50.3	3.45
YOLOv5 l	499	46631350	93.7M	95.2	114.1	4.93
YOLOv5 x	607	87244374	175.1M	95.9	217.1	8.43
YOLO-Fastest	277	356700	4.8M	32.1	0.96	1.6
JH-YOLOv5-RDAB	565	3117313	6.6M	95.1	8.4	2.52

文中使用的实验环境具体参数如表 3 所示。

3.2 JH-YOLOv5-RDAB 网络与典型网络性能对比

确定 JH-YOLOv5-RDAB 网络结构后, 在文中构建的小尺寸红外车辆目标数据集上, 进行充分训练 (epoch=500), 同时与 YOLO 系列的 YOLOv3、YOLOv4-tiny、YOLOv4、YOLOv5 s、YOLOv5 m、YOLOv5 l、YOLOv5 x、YOLO-Fastest 等 8 个算法进行了对比验证, 统计最终性能指标。图 14 所示为各算法训练结果对比结果。

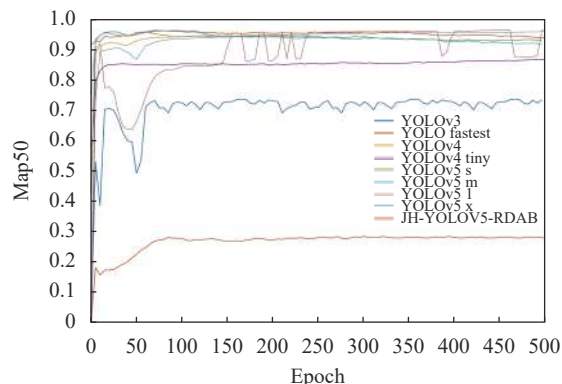


图 14 各算法训练结果对比

Fig.14 Image comparison of training results

总结对比各个算法参数量、模型大小和以 mAP50 代表的检测精度, 检测速度与模型大小正相关, 移植所用资源与参数量正相关, 得到结果如表 4 所示。mAP 用于评价检测算法在某个数据集上的平均精度, 多个类别物体检测中, 每一个类别都可以根据 recall 和 precision 绘制一条曲线, AP 就是该曲线下的

面积, mAP 是多个类别 AP 的平均值。mAP50 指的是交并比为 0.5 时 mAP 的值。而交并比是目标检测中使用的一个概念, 是产生的候选框与原标记框的交叠率, 即它们的交集与并集的比值。GFLOPs 是浮点运算次数, 可以用来衡量算法/模型复杂度, 1 GFLOPs=10 亿次浮点运算。

由表 4 可知, 提出的 JH-YOLOv5-RDAB 算法模型权重大小为 6.6 MB, 低于 YOLOv5 s 算法的一半, 而检测精度 mAP 接近 93.7 MB 的 YOLOv5 l, 达到 95.1%, 而 GFLOPs 值只有 8.4, 具有可观的优越性, 能够满足红外小尺寸车辆目标的检测识别需求, 同时其

轻量化的尺寸为后续嵌入式平台加速提供了可靠支持。但是, 由于 EPA 模块以及 RDAB 模块的使用, 使得算法含有较多的层数。由表 4 可知, 提出的 JH-YOLOv5-RDAB 算法虽然参数量少, 模型权重大小较小, 但网络层数较多, 算法含有 565 层, 仅次于 YOLOv5 x 算法的 607 层, 使得算法改写以及移植困难。

从图 15 可以看出, 文中提出的 JH-YOLOv5-RDAB 算法网络具有很好的性能指标, 能够对公路、森林、田野等复杂环境, 对像素约为 10×10 以内的红外车辆目标进行较为精确的检测识别。

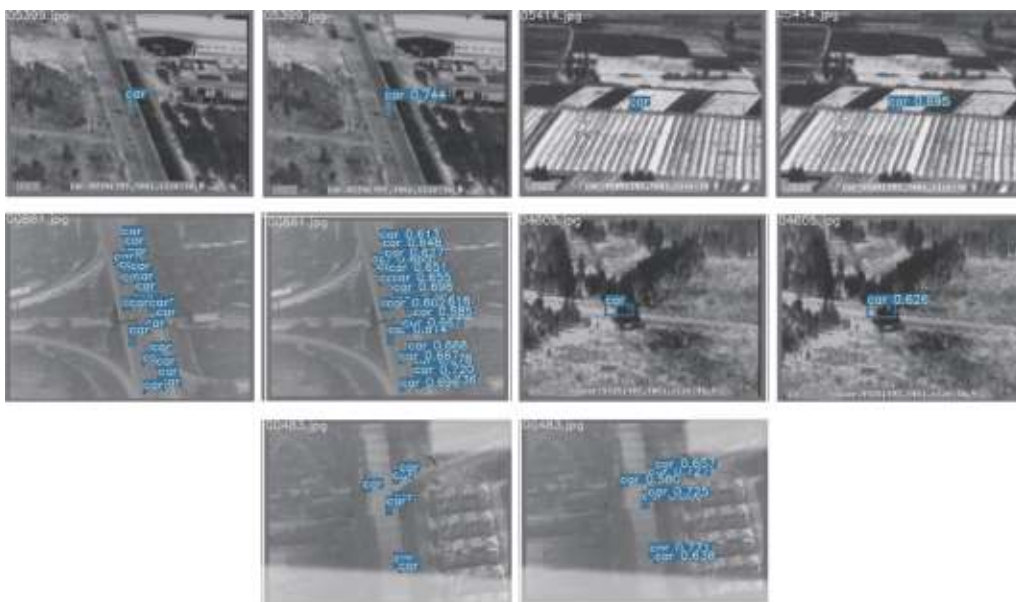


图 15 部分检测结果图像对比 (左: 实际标注图, 右: 算法检测图)

Fig.15 Image comparison of some detection results (left: actual annotation map, right: algorithm detection map)

3.3 消融实验设计

为了进一步分析每个改进点对 JH-YOLOv5-RDAB 算法的贡献, 文中进行了基于 YOLOv5 s 的轻量化剪裁的消融实验, 不同注意力机制不同融合位置的消融实验, RDAB 模块不同融合位置的消融实验, 实验结果分别如表 5、表 6、表 7 所示。

表 5 显示的是采用不同调整系数 depth_multiple 和系数 width_multiple 对 YOLOv5 s 网络进行轻量化剪裁时, 对算法性能的影响。表 5 按 epoch50 的最高 mAP 作为性能效果的验证指标, 并将各算法的性能效果与网络大小以 YOLOv5 s 为对照标准进行归一化, 即网络大小=待测网络参数量/YOLO5 s 参数量,

性能效果=待测网络检出率/YOLOv5 s 检出率。由表 5 可知在尽可能保证算法检测性能的基础上使得算法模型大小降到最低, 网络 2 最为满足所需, 以原三分之一的模型大小保留了原算法性能的约三分之二。

表 6 显示的是分别采用 ECA 与 EPA 模块在 YOLOv5 s 网络不同位置融合时, 对算法性能的影响。表 6 按 epoch50 的最高 mAP 作为性能效果的验证指标, 并将各算法的性能效果与网络大小以网络 2 为对照标准进行归一化。由表 6 可知, 融合了 EPA 模块的网络比单独融合单一注意力机制模块的网络具有更好的性能效果, 说明了在文中所用的红外车辆目标数据集中, 融合 EPA 模块的网络更适用。而将

表 5 基于 YOLOv5 s 的轻量化剪裁实验效果对比

Tab.5 Comparison of experimental effects of lightweight tailoring based on YOLOv5 s

Algorithm	Depth_multiple	width_multiple	Performance effect/times	Network size/times
YOLOv5 x	1.33	1.25	2.3	12.2
YOLOv5 l	1	1	1.8	6.5
YOLOv5 m	0.67	0.75	1.2	2.9
YOLOv5 s	0.33	0.50	1	1
Net 1	0.25	0.33	0.85	0.49
Net 2	0.2	0.25	0.63	0.35
Net 3	0.1	0.2	0.33	0.12

表 6 基于注意力机制融合的网络实验效果对比

Tab.6 Comparison of network experimental effects based on attention mechanism fusion

Fused operator	Integrated ECA		Integrated EPA	
	Performance effect/times	Network size/times	Performance effect/times	Network size/times
Net 2	1	1	1	1
CBL	1.1	1.6	1.3	2.4
SPP	1.2	1.12	1.43	1.18
CSP	1.3	1.1	1.33	1.11
Focus	1.02	1.01	1.11	1.01

表 7 基于残差密集注意模块融合的网络实验效果对比

Tab.7 Comparison of network experimental effects based on RDAB

Fused operator	Integrated RDAB	
	Performance effect/times	Network size/times
Focus	1.28	1.06
SPP+ Focus	1.35	1.08

EPA 模块与 SPP 模块、CSP 模块融合既提高了检测网络的性能指标,又不会带来过度的计算负荷和计算复杂度。

表 7 显示的是采用 RDAB 模块在 YOLOv5 s 网络不同位置融合时,对算法性能的影响。由表 7 可知,在 Focus 模块与 SPP 模块中融合 RDAB 可以进一步提升网络性能。

4 结 论

(1) 为满足特定复杂背景下的时间敏感目标快

速、精确检测识别的军事需求,针对红外车辆目标尺寸小、样本少,识别难度大等问题,提出了一种由 YOLOv5 s 进行轻量化,融合了特征注意力机制和残差密集块的深度学习检测识别算法。通过对网络轻量化剪裁,缩小结构,提高实时性;提出了混合域注意力机制模块 EPA,该模块通过不降维的局部跨信道交互策略使算法更快速有效地关注重要通道,抑制无效通道,同时,使得算法更关注与目标相关的像素信息;提出了基于注意力机制的 RDAB,由密集残差块与注意力机制 EPA 构成,该模块通过密集卷积层来提取充分的局部特征,允许通过多个局部残差连接绕过较不重要的信息,使得检测算法更关注与目标相关的通道与像素位置信息,从而使算法以较小的模型结构获得较好的检测效果。

(2) 构建基于机载挂飞采集和无人机采集的中/长波红外车辆目标数据库,并进行了亮度变化、对比度变化、旋转、平移、放缩、翻转、裁剪、随机拼接等数据增广处理。设计了不同位置和数量模块融合实验,以得到最优算法网络 JH-YOLOv5-RDAB 的网络结构。同时,在文中数据集上,将提出的 JH-YOLOv5-RDAB 算法与其他 8 种典型网络算法进行了对比实验,以 6.6 MB 的模型大小获得了 95.1% 的检测精度 (mAP50),实验结果表明了该网络的优越性和可行性。JH-YOLOv5-RDAB 算法模型小、精度高、实时性高的优势对于提升高速飞行器精确检测识别能力,提升探测系统的智能化水平具有重要意义。

参考文献:

[1] Li Xudong, Ye Mao, Li Tao. A review of object detection

- research based on convolutional neural network [J]. *Computer Application Research*, 2017, 34(10): 2881-2887. (in Chinese)
- [2] Tang Cong, Ling Yongshun, Yang Hua, et al. Visual tracking method for object detection based on deep learning [J]. *Infrared and Laser Engineering*, 2018, 47(5): 0526001. (in Chinese)
- [3] Krizhevsky A, Sutskever I, Hinton G. Imagenet classification with deep convolutional neural networks [J]. *Advances in Neural Information Processing Systems*, 2012, 25: 1097-1105.
- [4] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015: 1-9.
- [5] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[EB/OL]. (2014-09-04)[2022-03-20]. <https://arxiv.org/abs/1409.1556v4>.
- [6] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition , 2016: 770-778.
- [7] Zhou Xiaoyan, Wang Ke, Li Lingyan. A review of target detection algorithms based on deep learning [J]. *Electronic Measurement Technology*, 2017, 40(11): 89-94. (in Chinese)
- [8] 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, 2014: 580-587.
- [9] Redmon J , Divvala S , Girshick R , et al. You only look once: Unified, Real-Time Object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2016: 779-788.
- [10] Ren H, Wang X G. Review of attention mechanism [J]. *Journal of Computer Applications*, 2021, 41(S1): 1-6. (in Chinese)
- [11] Hu J, Shen L, Albanie S, et al. Squeeze-and-excitation networks [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, 42(8): 2011-2023.
- [12] Wang Q, Wu B, Zhu P, et al. ECA-Net: Efficient channel attention for deep convolutional neural networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2020: 11531-11539.
- [13] Qin X, Wang Z, Bai Y, et al. FFA-Net: Feature fusion attention network for single image dehazing[C]//Proceedings of the National Conference on Artificial Intelligence in Association for the Advancement of Artificial Intelligence, 2020: 1-9.
- [14] Zhang Y, Tian Y, Kong Y, et al. Residual dense network for image super-resolution[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 2472-2481.