

Anchor-free 轻量级红外目标检测方法 (特邀)

高 凡^{1,2}, 杨小冈^{2*}, 卢瑞涛², 王思宇², 高久安², 夏 海²

(1. 北京华航无线电测量研究所, 北京 100013;

2. 火箭军工程大学导弹工程学院, 陕西 西安 710025)

摘 要: 针对红外目标的特点, 提出了一种 anchor-free 轻量级红外目标检测方法, 提高了嵌入式平台对红外目标的检测能力。针对计算资源有限的平台, 提出了一种新的轻量级卷积结构, 引入非对称卷积增强标准卷积的特征表达能力, 同时有效减少参数和计算量。设计并行多路特征通道, 经过通道拼接生成丰富的特征, 结合注意力模块和 Channel Shuffle 构建轻量级特征提取单元。增加 SkipBranch 促进浅层信息向高层传递, 进一步丰富高层特征。在 FLIR 数据集进行实验验证, 设计的轻量级网络结构精度为 81.7%, 超过了 YOLOv4-tiny, 但模型参数量减少了 75.0%、计算量下降了 71.1%, 并且推理时间压缩了 91.3%, 能够满足嵌入式平台红外目标的实时检测需求。

关键词: 红外目标; 轻量化; 目标检测; 神经网络; 非对称卷积

中图分类号: TP391.4 **文献标志码:** A **DOI:** 10.3788/IRLA20220193

Anchor-free lightweight infrared object detection method (Invited)

Gao Fan^{1,2}, Yang Xiaogang^{2*}, Lu Ruitao², Wang Siyu², Gao Jiuan², Xia Hai²

(1. Beijing Huahang Radio Measurement Institute, Beijing 100013, China;

2. Missile Engineering Institute, Rocket Force University of Engineering, Xi'an 710025, China)

Abstract: According to the characteristics of infrared targets, an anchor-free lightweight infrared target detection method was proposed, which improved the detection ability of embedded platform. For the platform with limited computing resources, a new lightweight convolution structure was proposed. Asymmetric convolution was introduced to enhance the feature expression ability of standard convolution, reducing the amount of parameters and computation effectively. A lightweight feature extraction unit was constructed by designing parallel multi-feature path, which generated rich features through channel concatenation, then combining with attention module and channel shuffle. SkipBranch was added to promote the transmission of shallow information to the high level and further enrich the characteristics of the high level. Experiments on FLIR dataset showed that the accuracy of the designed lightweight network structure was 81.7%, which exceeded YOLOv4-tiny. However, the model parameters and calculation amount were reduced by 75.0% and 71.1% respectively, and the reasoning time was compressed by 91.3%, which could meet the real-time detection requirements of infrared object on embedded platform.

Key words: infrared target; lightweight; object detection; neural network; asymmetric convolution

收稿日期:2022-03-17; 修订日期:2022-04-11

基金项目:国家自然科学基金(61806209);陕西省自然科学基金(2020 JQ-490);航空科学基金(201851 U8012)

作者简介:高凡,男,硕士生,主要从事计算机视觉方面的研究。

导师(通讯作者)简介:杨小冈,男,教授,博士生导师,博士,主要从事视觉导航、目标检测、图像处理等方面的研究。

0 引言

目标检测作为计算机视觉的一个重要的分支,随着相关理论的深入研究和技术的广泛应用,取得了巨大的进步。而人工智能在实际中的应用,严重依赖高性能服务器的算力支撑。由于内存、计算等硬件的发展难以满足当前不断进步的神经网络模型庞大的存储和运算需求,模型的轻量化成为亟待解决的问题。

由于端侧和边缘设备的硬件资源有限,在模型的参数量、计算量以及推理速度与精度之间取得更加平衡的性能,才能满足实际需求。在神经网络的研究中,设计合理有效的结构来使得各层获取丰富多样的特征是网络性能提升的关键。由于参数量较少,轻量化的网络难以通过获取和复杂结构相比拟的丰富特征。

目前轻量化神经网络的设计取得了一系列成果。MobileNetv1^[1]采用深度可分离卷积替代传统卷积,并使用分辨率超参数控制输入图像分辨率和宽度超参数调节网络宽度,有效减少了模型参数量; MobileNetv2^[2]通过 inverted residual block 结构,优化 MobileNetv1 的性能; MobileNetv3^[3]引入轻量级注意力机制 SENet^[4]对通道关系重新建模,并利用神经网络结构搜索 (Neural Architecture Search, NAS) 技术,进一步提高模型性能。ShuffleNetv1^[5]提出了 pointwise group convolution 来降低计算复杂度,并引入 Channel Shuffle,以提高通道不同组内的信息流动; ShuffleNetv2^[6]针对神经网络在硬件上的实际运行情况,提出了网络设计准则并设计了新的轻量级结构,进一步提高了模型在硬件上的推理速度。SqueezeNet^[7]通过 1×1 和 3×3 的卷积共同组成 Fire 模块,有效减少了参数量。GhostNet^[8]在卷积生成特征图的基础上,又通过有效的线性运算来进行特征图扩展,减少了参数量和计算量。EfficientNet^[9]研究了网络深度、宽度和分辨率对性能指标的影响,并通过 NAS 获得了效果更好的模型; EfficientNetv2^[10]在此基础上引入了 Fused-MBConv 模块,并通过渐进式学习策略加快了训练速度。上述高效的模型大多通过深度可分离卷积和 NAS 实现,对计算资源有极高的要求,而且不能完全适用于特定的红外场景。

以 Faster R-CNN^[11]和 YOLO^[12]系列为代表的 anchor-based 算法在模型训练之前需要对数据进行聚

类分析,确定最优的锚框,而锚框的设置对模型的性能会造成影响,对于不同的任务场景又需要进行重新调整。Anchor-free 算法通过消除先验框,有效缓解了预设锚框带来的超参数干扰,简化了训练过程;同时避免了大量负样本带来的样本不均衡,大幅减少了 IoU 计算,降低了内存占用和时间消耗,适用于完成端侧实时精确目标检测任务。

针对红外图像分辨率较低、目标纹理特征不明显^[13]的问题,文中设计了一种轻量级特征提取网络,将非对称卷积和标准卷积相结合,提高对不同尺度目标特征的表达能力,降低了参数和计算量;在特征通道设置不同大小的卷积核,融合不同卷积结构的细节特征;并引入注意力机制和 Channel Shuffle 增强通道维度的特征获取和信息流动。为缓解红外图像在下采样过程中的细节丢失,采用改进的 Fcous 结构,同时提高了推理速度。通过 SkipBranch 结构直接将浅层定位信息和高层语义信息相融合,丰富高层的特征,加强轻量级结构的特征描述。实验结果表明,文中的轻量化模型具有较高的检测精度,在模型参数和计算量大幅压缩的条件下,实现了嵌入式平台红外目标实时检测。

1 轻量级特征提取网络设计

1.1 并行多特征通道轻量级卷积结构

红外成像时不同物体的红外辐射特性存在差异,如果目标重叠则会造成能量交互,使得图像中物体轮廓边缘模糊^[14]。受到 Inception^[15]网络的启发,文中设计了并行多特征通道 (Parallel Multi Feature Path, PMFP) 轻量级卷积结构,如图 1 所示。采用非对称卷积和常规卷积相结合构成单独的特征通道,通过不同大小的卷积核获取不同大小的感受野,最后对多路特征进行融合,使得模型对不同尺度的目标适应性更强,提取的信息更加丰富,减少通道的冗余信息。通过残差结构^[16]连接网络的不同层,一定程度上缓解在反向传播时出现的梯度消失问题,在实现特征复用的同时减少采用点卷积进行通道对齐带来的运算消耗,进一步提高神经网络的计算效率和融合效率。

PMFP 结构将输入特征映射为四路分支。第一路分支为 3×3 的深度可分离卷积生成的特征映射;第二路分支包括 5×1 和 1×5 的非对称卷积,以及 3×3 深度

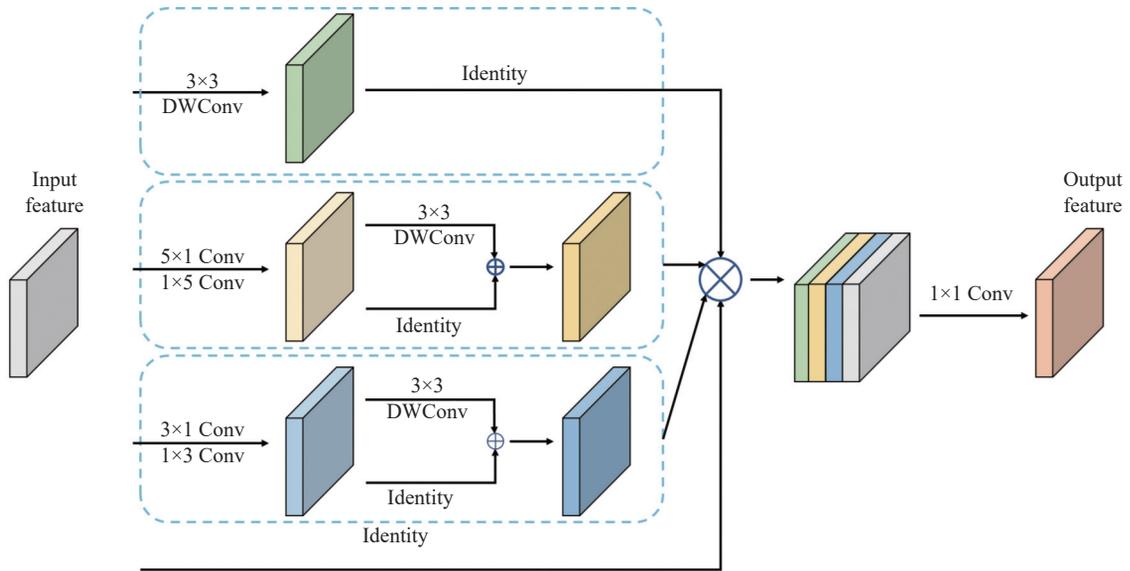


图 1 PMFP 结构

Fig.1 Structure of PMFP

可分离卷积的残差结构;第三路分支的特征通道首先经过 3×1 和 1×3 的两个非对称卷积的叠加,再串联 3×3 深度可分离卷积的残差块;第四路分支保留输入映射进行特征复用;然后将四路输出提取的不同信息进行特征融合,并通过 1×1 的卷积进行降维和通道间的编码。PMFP 通过不同的特征映射方式减少冗余信息的生成,提高模型的特征表达能力,并显著减少了模型的计算量和参数量。

一维非对称卷积能够提升标准卷积的表达能力,获得更加丰富的特征空间,无需引入额外超参数和推理时间,而且能够有效降低计算量和参数量。如图 2 所示,相比于常规卷积,这种卷积方式连接更加稀疏^[17],且感受野相同。假设特征图的尺寸为 $F \times F$,非

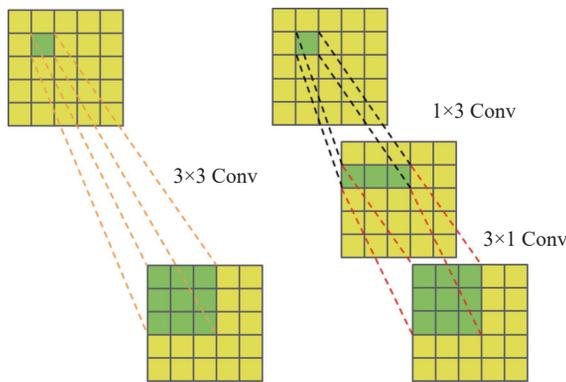


图 2 常规卷积和非对称卷积

Fig.2 Conventional convolution and asymmetric convolution

对称卷积核的大小为 $1 \times K$ 和 $K \times 1$,常规卷积核的大小为 $K \times K$,且卷积运算前后特征图的分辨率不变,则非对称卷积结构的计算量为 $2 \times K \times F \times F$,参数量为 $2K$;常规卷积的计算量为 $K \times K \times F \times F$,参数量为 $K \times K$ 。那么计算量减少为:

$$\frac{2 \times K \times F \times F}{K \times K \times F \times F} = \frac{2}{K}, K \geq 3 \quad (1)$$

参数量减少为:

$$\frac{2K}{K \times K} = \frac{2}{K}, K \geq 3 \quad (2)$$

1.2 注意力机制和 Channel Shuffle

红外目标在成像时受到外界环境的影响会导致信噪比较低,因此引入注意力机制 SimAM^[18],使得神经网络专注于输入的特定部分,增强对红外目标的检测能力。相比其他注意力结构^[19],SimAM 无需额外的参数就能够有效改善模型性能。常用的一些注意力结构 SENet 等往往需要通过额外的子网络生成注意力权值,如 SENet 结构通过全局平均池化和两个全连接层的组合来生成通道权值,并且在全连接层进行通道压缩时又会引入新的超参数,使工程实际应用中网络调优的复杂度增加。根据神经科学中信息丰富的神经元通常的放电模式与周围神经元不同的原理,SimAM 定义了能量函数:

$$e_i(\omega, b_i, y, x_i) = (y_i - \hat{t})^2 + \frac{1}{M-1} \sum_{i=1}^{M-1} (y_o - \hat{x}_i)^2 \quad (3)$$

式中: t 和 x_i 代表输入特征 $X \in \mathbb{R}^{C \times H \times W}$ 在同一通道的目标神经元和其他神经元, $\hat{t} = \omega_t t + b_t$, $\hat{x}_i = \omega_i x_i + b_i$ 是关于二者的线性变换; i 是不同的空间维度; $M = H \times W$ 表示此通道的神经元数量; ω_i 和 b_i 是线性变换的权重和偏置。

通过最小化能量函数, 找到同一通道内目标神经元和其他神经元之间的线性可分关系。对 y_i 和 y_o 采用二值简化, 并添加正则项得到新的能量函数:

$$e_i(\omega_i, b_i, y, x_i) = (1 - (\omega_i t + b_i))^2 + \frac{1}{M-1} \sum_{i=1}^{M-1} (-1 - (\omega_i x_i + b_i))^2 + \lambda \omega_i^2 \quad (4)$$

理论上, 每个通道有 M 个能量函数。公式 (4) 有解析解如下:

$$\omega_i = -\frac{2(t - \mu_i)}{(t - \mu_i)^2 + 2\sigma_i^2 + 2\lambda} \quad (5)$$

$$b_i = -\frac{1}{2}(t + \mu_i)\omega_i \quad (6)$$

式中: $\mu_i = \frac{1}{M-1} \sum_{i=1}^{M-1} x_i$ 和 $\sigma_i^2 = \frac{1}{M-1} \sum_{i=1}^{M-1} (x_i - \mu_i)^2$ 是该通道中除目标神经元 t 之外其他神经元的均值和方差。能量最小可通过下式计算:

$$e_i^* = \frac{4(\hat{\sigma}^2 + \lambda)}{(t - \hat{\mu})^2 + 2\hat{\sigma}^2 + 2\lambda} \quad (7)$$

式中: $\hat{\mu} = \frac{1}{M} \sum_{i=1}^M x_i$; $\sigma_i^2 = \frac{1}{M} \sum_{i=1}^M (x_i - \hat{\mu})^2$ 。当能量越低, 目标神经元 t 与其他神经元之间的差异越明显, 也就是其重要性越高, 即可表示为 $\frac{1}{e_i^*}$ 。通过对神经元重要性的描述, 就能进一步对特征进行增强:

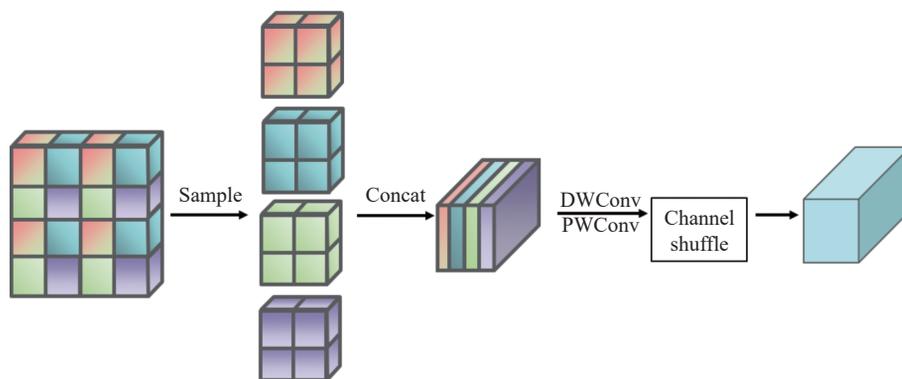


图 4 Slim-Focus 结构

Fig.4 Structure of Slim-Focus

$$\tilde{X} = \text{sigmoid}\left(\frac{1}{E}\right) \odot X \quad (8)$$

式中: E 是所有 e_i^* 在通道维度和空间维度的组合, 通过 sigmoid 函数进行数值限制。

在深度可分离卷积中的深度卷积是将输入特征图进行分组卷积, 采用不同的卷积对组内的特征图进行编码, 这种稀疏连接方式相较于对输入特征图进行全通道卷积的密集连接方式, 能够显著降低计算量。由于特征图之间缺少组间通信, 会降低网络的特征提取能力, 因此引入 Channel Shuffle 进行通道重排, 如图 3 所示, 加强通道间的信息交流, 弱化通道信息独立的影响, 提高轻量化卷积结构的性能。

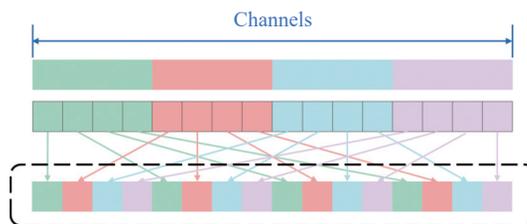


图 3 Channel Shuffle

Fig.3 Channel Shuffle

1.3 Slim-Focus 降采样

神经网络通过降采样不断增大感受野的同时, 图像的分辨率不断减小, 由于成像原理的限制, 红外目标的纹理较为模糊, 图像分辨率较低。改进 Focus 结构的 Slim-Focus 通过对输入进行隔像素采样实现特征图分辨率减小, 如图 4 所示, 将特征图的宽度和高度信息集中到了通道空间, 输入通道扩充了 4 倍, 得到的低分辨率特征图经过深度可分离卷积结构进行

空间编码,通过 Channel Shuffle 操作实现通道重组最终生成降采样特征图,减少了计算量。

图 5 为原始图像经过两次降采样之后对比结果, Slim-Focus 结构相较于最大池化降采样,能够更好地保留红外目标的纹理信息,缓解在降采样过程中的信息丢失,并弱化红外图像噪声的影响,提高定位精度。

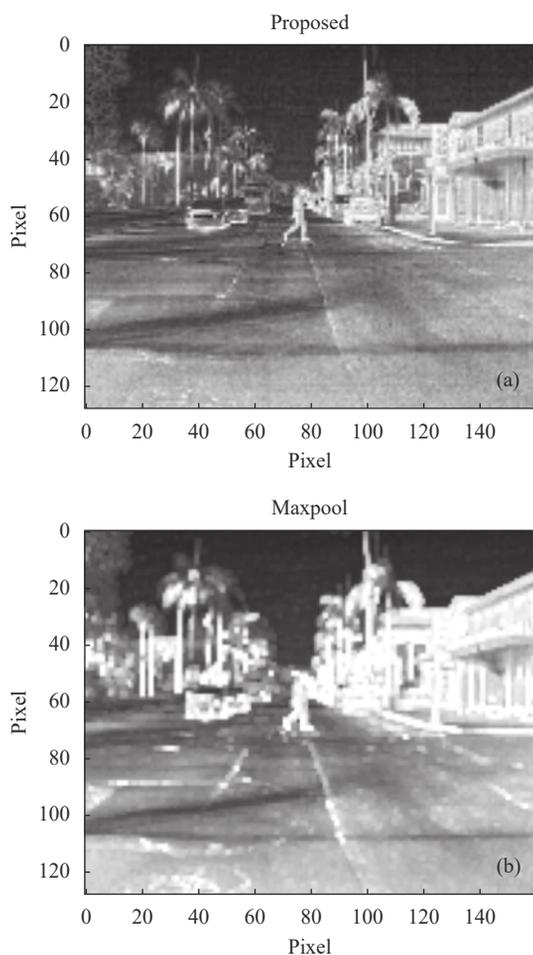


图 5 降采样效果对比

Fig.5 Comparison of down-sampling effect

1.4 SkipBranch 结构

随着神经网络层数和宽度增加,模型性能不断提高,但是在较深的结构中由于宽度的加深会出现通道冗余,一些权重参数趋于零,为提高通道利用率增加通道特征的独立性,如图 6 引入 SkipBranch 结构,促进特征提取网络底层信息向最后一层结构的流动,增强特征描述。为将不同空间维度的特征信息进行融合,首先采用自适应平均池化进行特征对齐,然后通过深度可分离卷积进行通道维度对齐,最后引入注意

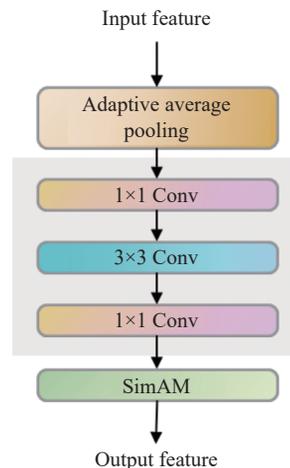


图 6 SkipBranch 结构

Fig.6 Structure of SkipBranch

力机制聚焦有效信息。

2 算法整体框架与实现

文中的轻量级模型 PMFPSNet 结构如图 7 所示,包括特征提取网络、特征融合网络和 anchor-free 算法。

2.1 PMFPSNet 结构

轻量级目标检测模型 PMFPSNet 的前端,首先由 PMFP 轻量级结构作为特征提取单元,随着神经网络层数的加深,通过 Slim-Focus 降采样模块减小特征图的分辨率增大感受野,同时 PMFP 结构的通道数逐渐增加,提取更加丰富的特征信息,然后 SkipBranch 辅助特征增强结构促进浅层信息和高层语义的融合,完成整个特征提取网络的构建。

神经网络通过不同尺度的网络结构生成丰富的特征图,浅层结构的特征图保留的目标细节信息较多,如边缘、角点和纹理等,因此对小目标的分类和定位更为重要;随着网络加深,特征图在下采样的过程中尺度减小,语义信息更加复杂,有利于大目标的识别。构建特征融合网络能够利用浅层网络的强定位信息和深层网络的强语义信息,提高网络的检测能力。FPN^[20]的低层次特征图包含准确的定位信息,在特征向高层传递的过程中路径较长,增加了高层语义访问准确定位信息的难度,这样造成特征信息丢失较为严重。

如图 8 所示,在特征融合结构中,将 PANet^[21]中的 3x3 标准卷积替换为深度可分离卷积,减少计算

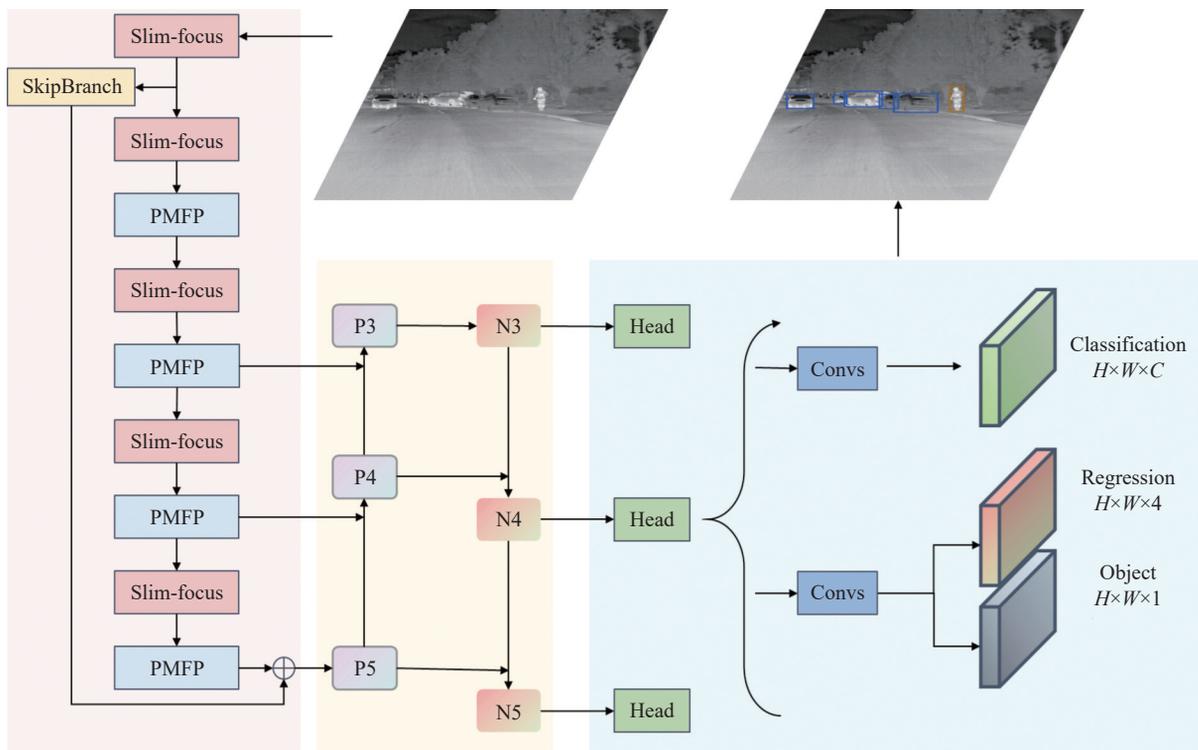


图 7 PMFPSNet 轻量级模型结构

Fig.7 Structure of lightweight model PMFPSNet

量。在 FPN 的基础上添加自下而上的路径, 通过网络的横向连接缩短底层和高层之间的信息路径, 增强融合后网络的特征表达能力。这样的多尺度融合结构能够保证获取有效的分类特征, 增强网络对小目标的定位能力, 从而提高模型对多尺度目标的检测性能。

2.2 Anchor-free 算法实现

FCOS 提出在 Regression 分支并行增加 Centerness 分支来抑制远离中心点的预测框, 在一定程度上提高了网络的性能。公式 (9) 中 (l^*, t^*, r^*, b^*) 表示 (x, y) 到 groundtruth 边框的距离, 其中 (x_0^i, y_0^i) 、 (x_1^i, y_1^i) 分别是 groundtruth 的左上角和右下角的坐标, (x, y) 是特征图上的像素位置, s 表示第 i 层特征图 $F_i \in \mathbb{R}^{H \times W \times C}$ 的缩放的步幅, 公式 (10) 用来度量预测框偏离中心点的程度:

$$\begin{cases} l^* = (x - x_0^i) / s \\ t^* = (y - y_0^i) / s \\ r^* = (x_1^i - x) / s \\ b^* = (y_1^i - y) / s \end{cases} \quad (9)$$

$$centerness^* = \sqrt{\frac{\min(l^*, r^*)}{\max(l^*, r^*)} \times \frac{\min(t^*, b^*)}{\max(t^*, b^*)}} \quad (10)$$

如图 9 所示, 特征图中的 centerness label 往往比 IoU label 小, 导致一部分正样本难以召回。此外, 由于网络的 Classification 分支和 Regression 分支相互独立, 可能导致在某些场景下分类预测值高但定位预测不一定最准确, 使得检测器的分类预测值和定位预测之间相关性较低, 导致模型检测性能下降。由 IoU 的

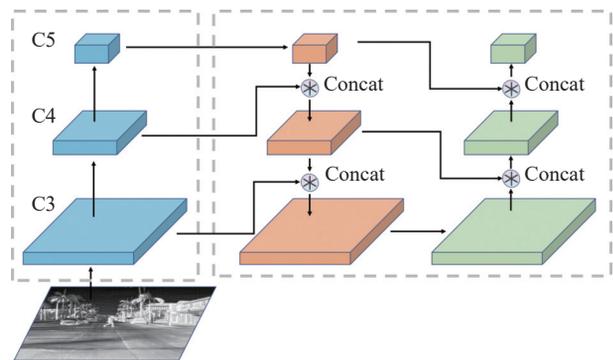


图 8 PANet 结构

Fig.8 Structure of PANet

最后采用 anchor-free 目标检测算法作为预测结构, 通过改进 FCOS^[22] 将分类预测和定位预测进行结合, 提高模型的性能。

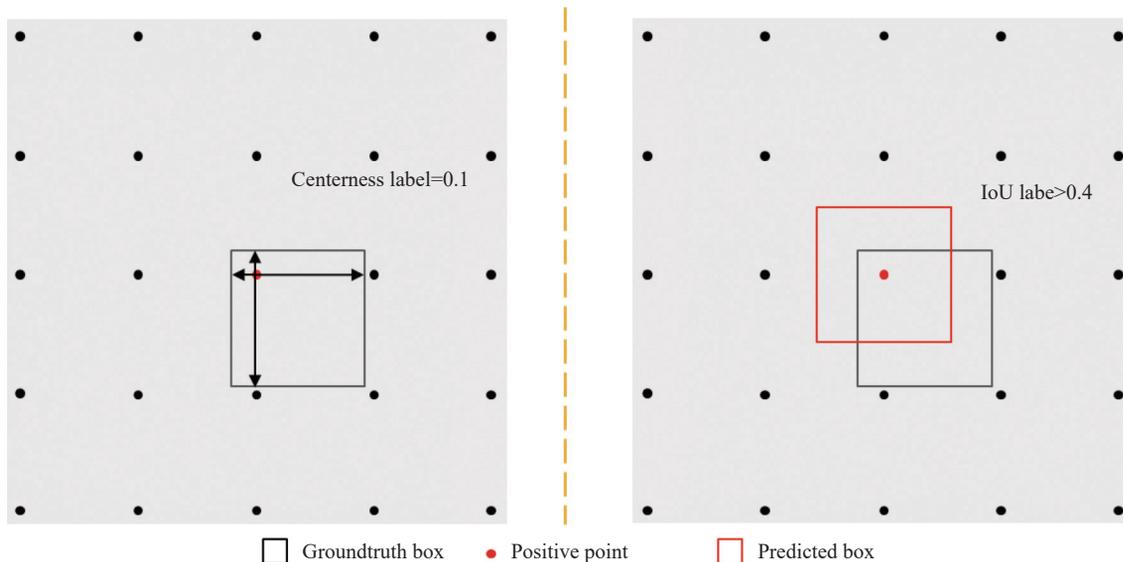


图 9 Centerness label 和 IoU label

Fig.9 Centerness label and IoU label

计算原理可知,其在一定程度上能够直接反映出定位的准确性,所以在 Classification 分支引入预测 IoU 参与模型训练,这样就能同时代表目标分类和定位精度,如公式 (11) 所示:

$$\begin{bmatrix} ct_1 \\ \vdots \\ ct_n \end{bmatrix} = \begin{bmatrix} cls_1^1 & \cdots & cls_1^c \\ \vdots & \cdots & \vdots \\ cls_n^1 & \cdots & cls_n^c \end{bmatrix} \cdot \begin{bmatrix} iou_1 \\ \vdots \\ iou_n \end{bmatrix} = \begin{bmatrix} cls_1^1 iou_1 & \cdots & cls_1^c iou_1 \\ \vdots & \cdots & \vdots \\ cls_n^1 iou_n & \cdots & cls_n^c iou_n \end{bmatrix} \quad (11)$$

式中: ct_n 表示 Classification 分支的学习目标; cls_n^c 表示第 n 个正样本的类别 c ; iou_n 表示第 n 个正样本和 groundtruth 的 IoU。

损失函数包括 IoU 损失、obj 损失和 cls 损失,如公式 (12):

$$\begin{aligned} Loss = loss_iou + loss_obj + loss_cls = & \\ & \frac{1}{N_{pos}} IoU_Loss(bbox_{pred}, bbox_{gt}) + \\ & \frac{1}{N_{pos}} BCE_Loss(obj_{pred}, obj_{gt}) + \\ & \frac{1}{N_{pos}} BCE_Loss(cls_{pred}, cls_{gt}) \end{aligned} \quad (12)$$

式中: N_{pos} 表示正样本的数量; IoU_Loss 表示 IoU 损失函数; BCE_Loss 表示交叉熵损失函数。

3 实验结果与分析

在 FLIR 数据集以行人和车辆为目标进行模型性能对比实验,算法采用 Pytorch 框架,实验环境为 Ubuntu18.04, 2 块 NVIDIA RTX 8000 显卡, batchsize 设置为 256, 起始学习率为 0.01, 并采用余弦退火策略训练 300 epochs。硬件测试平台为 Intel Core i7-10750 H 和 NVIDIA Quadro T2000, 最后在嵌入式平台 NVIDIA Jetson Xavier NX 进行实验验证。针对模型性能的评价,实验综合考虑 mAP(mean Average Precision)、参数量(Parameters)、计算量(FLOPS)以及推理延时(Delay)四个方面对不同模型进行消融实验,验证不同网络结构的性能。

3.1 特征提取网络性能对比

在特征提取网络中选择 ShuffleNetv2 和 CSPNet 与文中设计的 PMFPSNet 进行对比。ShuffleNetv2 在轻量级结构中具有较好的性能,而 CSPNet 在主流的网络模型中已被广泛使用。在通道维数相同的情况下,采用不同的特征提取网络进行模型训练比较,如表 1 所示。实验结果中,PMFPSNet 的 mAP 达到了 0.817, 参数量仅有 1.57 M, 计算量为 4.98 GFLOPS; 与 CSPNet 结构相比,精度提高了 3%, 但参数量和计算量分别下降了 30% 和 28%; 与 ShuffleNet 结构相比,具有更加明显的性能优势。

表 1 模型性能比较

Tab.1 Comparison of model performance

Model	mAP	Parameters	GFLOPS	Delay/ms
CSPNet	0.787	2.24 M	6.87	7.18
ShuffleNet	0.789	2.12 M	6.26	7.86
Maxpool	0.773	1.57 M	4.98	7.28
YOLOv4-tiny	0.811	6.27 M	17.2	84.6
PMFPSNet	0.817	1.57 M	4.98	7.34

3.2 下采样结构性能对比

在轻量级网络中 Maxpool 能够以较少的参数实现降采样, 在其他结构相同的情况下, 对 Slim-Focus 和 Maxpool 两种降采样结构性能进行比较, 如表 1 所示, 采用 Slim-Focus 能更好地保留红外目标的特征, 在降采样过程中的信息丢失更少, 比 Maxpool 方式的

精度高 4.4%, 验证了 Slim-Focus 降采样结构的有效性。

3.3 轻量级网络性能对比

将提出的 anchor-free 算法 PMFPSNet 与 anchor-based 轻量级网络性能进行对比, YOLOv4-tiny 是轻量级网络中的优秀代表, 与之相比, 在表 1 中 PMFPSNet 精度略高, 但参数量和计算量分别仅有前者的 25% 和 29%, 推理速度仅有 9%, 尽管其在网络宽度上进行了缩减, 但是由于卷积计算采用稠密方式, 计算量仍然较大, 而且网络输出层的减少限制了其性能。检测效果如图 10 所示, YOLOv4-tiny 对密集红外目标的检测存在较多的漏检, 而 PMFPSNet 对小目标的检测效果更好。综上所述, 相比其他结构, 文中的轻量级结构能够以更少的计算量和参数量实现更高的精度, 并且具有较好的推理速度, 模型性能更加优越。

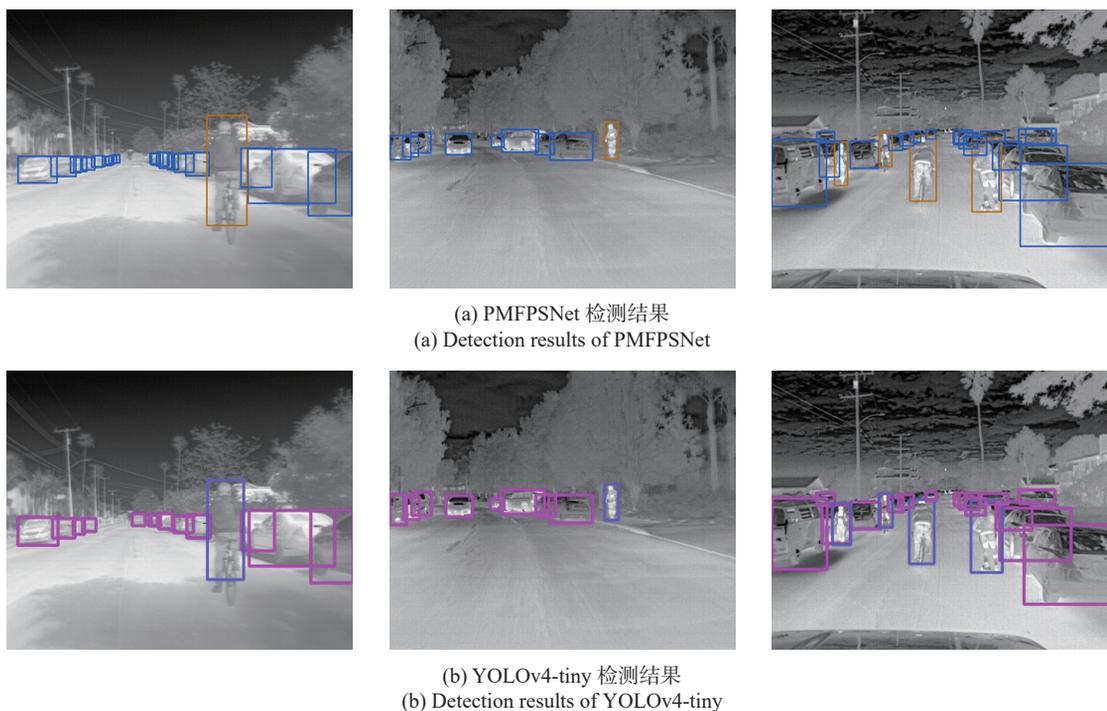


图 10 检测结果对比

Fig.10 Comparison of detection results

4 结 论

文中提出了一种 anchor-free 轻量级红外目标检测方法, 在模型 PMFPSNet 中通过并行多特征通道轻量级卷积结构 PMFP 提高特征提取单元对不同尺度目标特征的获取能力, 经过通道融合生成丰富的特征, 同时有效减少参数和计算量; 结合无参数注意力

模块 SimAM 和 Channel Shuffle 在不增加参数的情况下提高模型性能, 采用 Slim-Focus 结构改善在降采样过程中的红外特征丢失, 增加 SkipBranch 分支促进浅层信息向深层网络的流动, 提高模型在学习过程中的效率。在 FCOS 算法的基础上, 利用 IoU 分支融合定位信息和分类信息, 提高网络的精度。实验结果表

明, PMFPSNet 模型的检测精度更高, 且参数量和计算量大幅减少, 能够更好地完成嵌入式平台的红外目标实时检测任务。设计的轻量级模型 PMFPSNet 实现网络结构精简的情况下, mAP 为 81.7% 且高于其他轻量级网络。同时相较于 anchor-based 模型, 参数量和计算量分别下降 75.0% 和 71.1%, 具有更快的推理速度。

参考文献:

- [1] Howard A G, Zhu M, Chen B, et al. MobileNets: Efficient convolutional neural networks for mobile vision applications [J]. *arXiv preprint*, 2017: 1704.04861.
- [2] Sandler M, Howard A, Zhu M, et al. MobileNetV2: Inverted residuals and linear bottlenecks[C]//IEEE/CVF Conference on Computer Vision & Pattern Recognition, 2018: 4510-4520.
- [3] Howard A, Sandler M, Chen B, et al. Searching for MobileNetV3[C]//IEEE/CVF International Conference on Computer Vision, 2019: 1314-1324.
- [4] Hu Jie, Shen Li, Sun Gang, et al. Squeeze-and-excitation networks[C]//IEEE/CVF Conference on Computer Vision & Pattern Recognition, 2018: 7132-7141.
- [5] Zhang X, Zhou X, Lin M, et al. ShuffleNet: An extremely efficient convolutional neural network for mobile devices[C]//CVF Conference on Computer Vision & Pattern Recognition, 2018: 6848-6856.
- [6] Ma N, Zhang X, Zheng H T, et al. ShuffleNetV2: Practical guidelines for efficient CNN architecture design[C]//European Conference on Computer Vision, 2018, 11218: 122-138.
- [7] Iandola F N, Han S, Moskewicz M W, et al. SqueezeNet: AlexNet-level accuracy with 50 x fewer parameters and <0.5 MB model size [J]. *arXiv preprint*, 2016: 1602.07360.
- [8] Han K, Wang Y, Tian Q, et al. GhostNet: More features from cheap operations[C]//CVF Conference on Computer Vision & Pattern Recognition, 2020: 1577-1586.
- [9] Tan M X, Le Q V. EfficientNet: Rethinking model scaling for convolutional neural networks [J]. *arXiv preprint*, 2019: 1905.11946.
- [10] Tan M X, Le Q V. EfficientNetV2: Smaller models and faster training [J]. *arXiv preprint*, 2021: 2104.00298.
- [11] Ren S, He K, Girshick R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks [J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2017, 39(6): 1137-1149.
- [12] Wang Chen, Zhang Xiufeng, Liu Chao, et al. Detection method of wheel hub weld defects based on the improved YOLOv3 [J]. *Optics and Precision Engineering*, 2021, 29(8): 1942-1954. (in Chinese)
- [13] Cheng Yan, Yu Xuelian, Qian Weixian, et al. Ship wake extraction and detection from infrared remote sensing images [J]. *Infrared and Laser Engineering*, 2022, 51(2): 20210844. (in Chinese)
- [14] Wang Chunzhe, An Junshe, Jiang Xiujie, et al. Region proposal optimization algorithm based on convolutional neural networks [J]. *Chinese Optics*, 2019, 12(6): 1348-1361. (in Chinese)
- [15] Szegedy C, Ioffe S, Vanhoucke V, et al. Inception-v4, inception-ResNet and the impact of residual connections on learning[C]//AAAI Conference on Artificial Intelligence, 2017: 4278-4284.
- [16] Zhang Ruiyan, Jiang Xiujie, An Junshe, et al. Design of global-contextual detection model for optical remote sensing targets [J]. *Chinese Optics*, 2020, 13(6): 1302-1313. (in Chinese)
- [17] Li Weipeng, Yang Xiaogang, Li Chuanxiang, et al. Infrared object detection network compression using Lp normalized weight [J]. *Infrared and Laser Engineering*, 2021, 50(8): 20200510. (in Chinese)
- [18] Yang Lingxiao, Zhang Ru-Yuan, Li Lida, et al. SimAM: A simple, parameter-free attention module for convolutional neural networks[C]//International Conference on Machine Learning, 2021, 139: 11863-11874.
- [19] Ju Moran, Luo Haibo, Liu Guangqi, et al. Infrared dim and small target detection network based on spatial attention mechanism [J]. *Optics and Precision Engineering*, 2021, 29(4): 843-853. (in Chinese)
- [20] Lin T Y, Dollar P, Girshick R, et al. Feature pyramid networks for object detection[C]//IEEE Computer Society Conference on Computer Vision & Pattern Recognition, 2017: 936-944.
- [21] Liu S, Qi L, Qin H, et al. Path aggregation network for instance segmentation[C]//IEEE Conference on Computer Vision & Pattern Recognition, 2018: 8759-8768.
- [22] Tian Z, Shen C, Chen H, et al. FCOS: Fully convolutional one-stage object detection[C]//CVF International Conference on Computer Vision, 2019: 9626-9635.