

基于关键点的快速红外目标检测方法

苗壮^{1,2}, 张湧^{1*}, 陈瑞敏^{1,2}, 李伟华^{1,2}

¹中国科学院上海技术物理研究所红外探测与成像技术重点实验室, 上海 200083;

²中国科学院大学电子电气与通信工程学院, 北京 100049

摘要 针对红外探测系统对目标检测的实时性要求, 提出了一种基于关键点的快速红外目标检测方法。以目标中心作为目标检测关键点, 首先设计了轻量化的特征提取网络, 之后结合红外目标较小的特点, 利用不同层次特征的空间信息和语义信息设计了相应的特征融合网络, 并最终实现目标类别、位置和尺寸信息的预测。在自建空中红外目标数据集上对模型进行了对比测试, 与 YOLOv3 等经典检测模型相比, 检测速度大幅提高, 检测精度仅略有下降; 与同类型快速检测模型 Tiny-YOLOv3 相比, 在模型尺寸压缩至 Tiny-YOLOv3 尺寸的 23.39% 的情况下, 检测精度提高了 8.9%, 在中央处理器(CPU)上运行的检测速度亦提高了 13.9 ms/frame, 检测性能显著提升, 验证了方法的有效性。

关键词 测量; 机器视觉; 深度学习; 红外目标; 目标检测; 特征融合

中图分类号 TP391

文献标志码 A

doi: 10.3788/AOS202040.2312006

Method for Fast Detection of Infrared Targets Based on Key Points

Miao Zhuang^{1,2}, Zhang Yong^{1*}, Chen Ruimin^{1,2}, Li Weihua^{1,2}

¹Key Laboratory of Infrared System Detection and Imaging Technology, Shanghai Institute of Technical Physics, Chinese Academy of Sciences, Shanghai 200083, China;

²School of Electronic, Electrical, and Communication Engineering, University of Chinese Academy of Sciences, Beijing 100049, China

Abstract Aiming at the real-time request of the infrared detection system for target detection, we propose a method for fast detection of infrared targets based on key points. Taking the target center as the key point of target detection, we first design a lightweight feature extraction network. Then, we design a corresponding feature fusion network using the spatial and semantic information of features at different levels combined with the characteristics of small infrared targets. Finally, the prediction of target category, location and size is realized. The model is comparatively tested on the self-built aerial infrared target dataset. Compared with the classic detection models such as YOLOv3, the detection speed is greatly improved and the detection accuracy is only slightly reduced. Compared with the same type of fast detection model, Tiny-YOLOv3, the detection accuracy increases by 8.9% and the detection speed running on the central processing unit (CPU) increases by 13.9 ms/frame under the condition that the model size is compressed to 23.39% of Tiny-YOLOv3's size. The detection performance is significantly improved and the effectiveness of the method is confirmed.

Key words measurement; machine vision; deep learning; infrared target; target detection; feature fusion

OCIS codes 120.1880; 110.3080; 100.4996; 150.1135

1 引 言

得益于隐蔽性好、抗干扰性强和可全天候作业等优点, 红外探测系统在军事和民用领域均得到了广泛的应用。目标检测作为红外探测系统的核心任务, 需要在定位目标的同时确定目标类别, 有效提高

检测能力是提升系统性能的关键。自 2014 年区域卷积神经网络(R-CNN)^[1]率先利用卷积神经网络(CNN)完成特征提取并在检测效果上取得重大突破以来, 以 CNN 为基础的各类检测模型相继被提出, 与之结合的红外目标检测方法已成为红外探测领域的研究热点^[2-3]。当前大部分的研究主要集中在

收稿日期: 2020-08-17; 修回日期: 2020-08-30; 录用日期: 2020-09-08

基金项目: 上海市现场物证重点实验室基金资助项目(2017xcwzk08)

* E-mail: zybyx@sina.com

于改进模型以提高红外目标的检测精度^[4-5],而忽略了检测速度的重要性,特别是在仅包含中央处理器(CPU)的应用场景下,模型复杂,计算量大且检测速度过慢。因此,随着红外系统小型化的加速发展,如何利用有限的系统资源在保证一定检测精度的前提下提高检测速度是一项十分重要且具有挑战性的研究工作。

目前,基于预选框(anchor-based)的主流目标检测模型大致可分为以 Faster R-CNN^[6] 为代表的二阶检测模型和以 YOLO (You Only Look Once)^[7-9] 为代表的一阶检测模型两大类。Faster R-CNN 作为 R-CNN 的延伸,首先需要利用区域候选网络(RPN)生成感兴趣区域(ROI),然后再对该区域进行目标分类和位置回归。虽然模型的检测精度比较高,但是以牺牲检测速度为代价,因而无法满足诸多系统对实时性的要求。相比 Faster R-CNN, YOLO 直接利用回归的思想,使用单一网络完成目标类别和位置的检测,检测速度大幅提升。通过精简网络结构,以 YOLO 为基础的快速检测模型 Tiny-YOLO^[7-9] 已成为当前小型化系统部署的首选。

本文受 ConerNet^[10] 和 CenterNet^[11] 等无预选

框(anchor-free)模型的启发,结合实际红外应用场景,设计了以目标中心为检测关键点的快速红外目标检测模型,简称 FKPD(fast key points detection)模型。与同类型快速检测模型 Tiny-YOLOv3 相比,FKPD 模型无需通过先验知识设置预选框,可直接预测目标的中心坐标和尺寸。通过采用分离卷积等方法^[12-15] 设计网络结构,将模型尺寸和计算量分别压缩至 Tiny-YOLOv3 的 23.39% 和 66.24%,在 CPU 上进行测试,单帧红外图像的推理时间可达 115.18 ms。同时,通过不同层次特征融合的方法,在保证模型检测速度的同时,实现了模型检测精度的提升。实验结果表明,在空中红外目标数据集上,检测精度提高了 8.90%。

2 FKPD 目标检测模型

如图 1 所示,FKPD 模型包含特征提取网络(FEN)、特征融合网络(FFN)和目标预测网络(TDN)三个网络。待检测图片首先在特征提取网络中完成不同层次的特征提取,之后特征融合网络对特征进行融合,最后将融合后的特征图作为目标预测网络的输入完成目标中心点位置(目标关键点)、尺寸以及类别的预测。

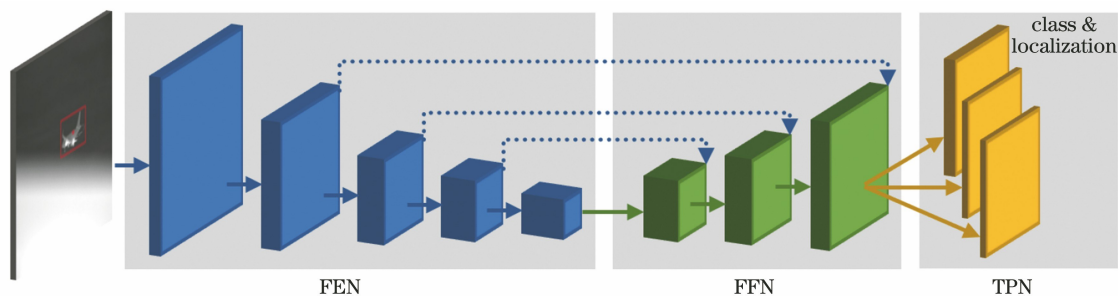


图 1 FKPD 模型框架

Fig. 1 Architecture of FKPD model

2.1 特征提取网络

特征提取网络作为检测模型的骨干网络,通常需要将其设计成一个较深的网络结构,以提高拟合能力进而提升模型的检测精度。而网络深度的增加意味着模型计算量的增加,因此对于快速检测模型来说,在保证一定检测精度的前提下降低网络的计算量是特征提取网络设计的关键要素。

图 2 为本文设计的一个轻量级的特征提取网络。待检测图片输入网络后,共计经历五个阶段的处理。每经过一个阶段的处理,输出特征的尺寸会缩小一半,而为了保留尽可能多的特征信息,输出特征的通道数会相应增加。第一阶段和第二阶段为常

规的卷积层(conv)和池化层(max pool),从第三阶段开始,为了降低网络的计算量,参考 Shufflenetv2^[15] 的单元结构进行了堆叠设计。当单元步长 s 为 2 时(block 1),输入特征首先分别经过两轮逐深度卷积(DW conv)和逐点卷积(1×1 conv)完成下采样,之后通过拼接(concat)的方式来扩充通道数,最后采用通道混洗(channel shuffle)的方式增强通道间的信息融合,所有卷积层使用批量归一化(BN)进行处理,并采用 LeakyReLU 作为激活函数;当单元步长为 1 时(block 2),输入特征首先按通道数量均分为两组(channel split),之后采用类似残差网络的结构完成逐通道和逐点卷积并进行

通道拼接和通道混洗,整个过程保持输入特征和输出特征通道数量相同。通过计算,整个网络的浮点运算量(FLOPs)约为 36.56×10^6 。由于红外目标的尺寸通常比较小,扩大感受野对于小目标检测并无优

势,因此特征提取网络中所有卷积层的卷积核尺寸均为 $3 \text{ pixel} \times 3 \text{ pixel}$ 。通过对比实验可以发现,利用空洞在不增加计算量的情况下扩大感受野,模型的检测精度会有所下降。

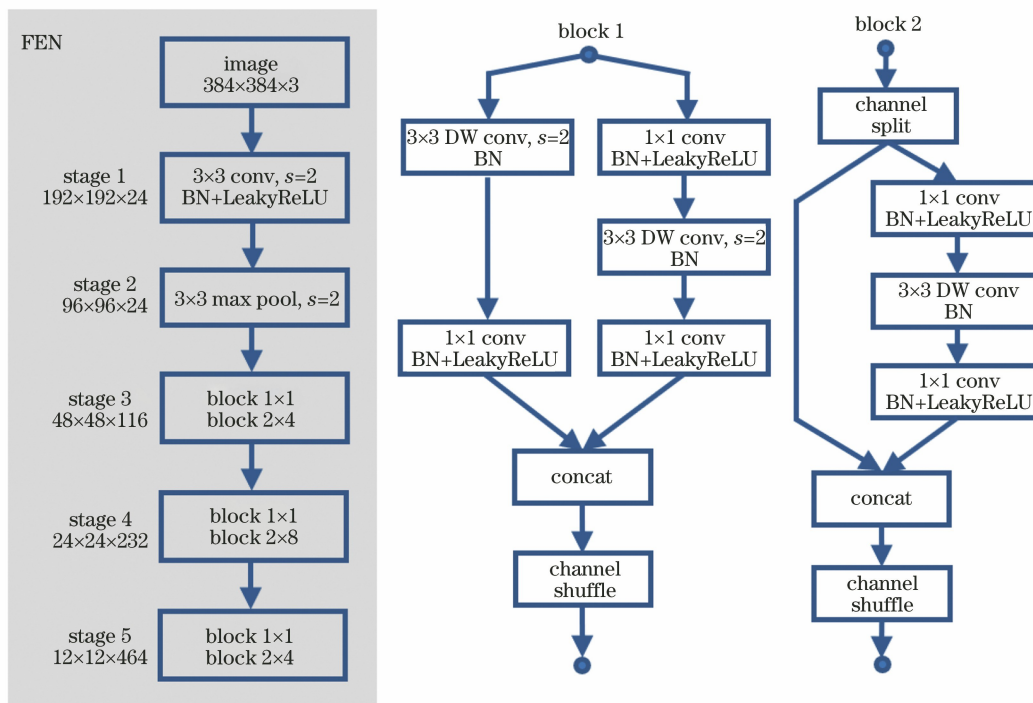


图 2 特征提取网络

Fig. 2 Feature extraction network

2.2 特征融合网络

在实际工程应用中,红外系统一般承担远距离探测的任务,因而待检测目标的尺寸通常比较小。如果直接利用特征提取网络最后一层的输出来进行预测,虽然特征图中包含了高层次的语义信息,但此时下采样率已经很大(本文为 16),小目标经过下采样后往往仅占据特征图中的几个点。而在这仅有的几个点中,由于每个点的感受野很大,因此每个特征点会包含大量目标周围区域的特征,目标自身特征占比减小,进而目标的定位精度受到影响。如果降低采样率,直接利用低层次的特征图进行预测,一个最主要的问题就是特征图的语义信息不足,进而影响目标的识别精度。因此,为了更好地利用深层特征图的语义性和浅层特征图的空间性,本文设计了图 3 所示的特征融合网络。

为了充分融合特征提取网络第二阶至第五阶所提供的特征信息,特征融合网络需要依次完成三组融合。由于每一组融合涉及到了不同尺度和通道数量的两阶特征图,因此每一组融合均需要经历三个阶段。第一阶段首先需要实现高阶特征图通道数量

的变换,考虑到前两组融合中特征图通道数量比较多,常规的卷积操作运算量比较大,因此为了降低网络的计算量,设计了图 3 所示的 block 3 的单元结构,利用分组卷积并叠加的方法,实现输入、输出特征图通道数量的减半,与直接使用标准卷积完成通道数量变换相比,利用该方式可大幅降低网络计算量。第二阶段将通道变换后的高阶特征图作为输入,通过步长为 2 的上采样(upsampling)实现高阶特征图尺度的变换;最后一个阶段,通过高低阶特征图相加的方式完成二者的融合。通过三组融合,最终特征融合网络输出的特征图尺寸为原图像尺寸的 $1/4$ (下采样率为 4),实现了不同层次特征图在语义性和空间性上的结合。实验结果表明,特征融合网络的引入有效提升了模型的检测精度且对检测速度影响很小。

2.3 目标预测网络

如图 4 所示,将特征融合网络融合后的特征图作为输入,目标预测网络共需要预测三类信息:1)特征图上每个点为目标中心点即目标关键点时的置信度得分,若共有 n 类目标需要预测,则需分别预测

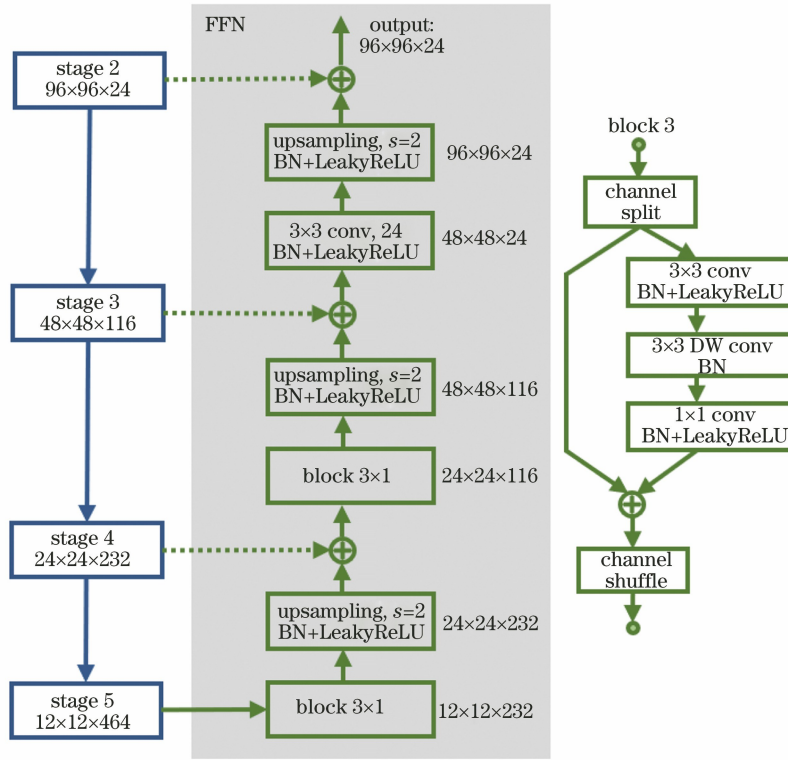


图 3 特征融合网络

Fig. 3 Feature fusion network

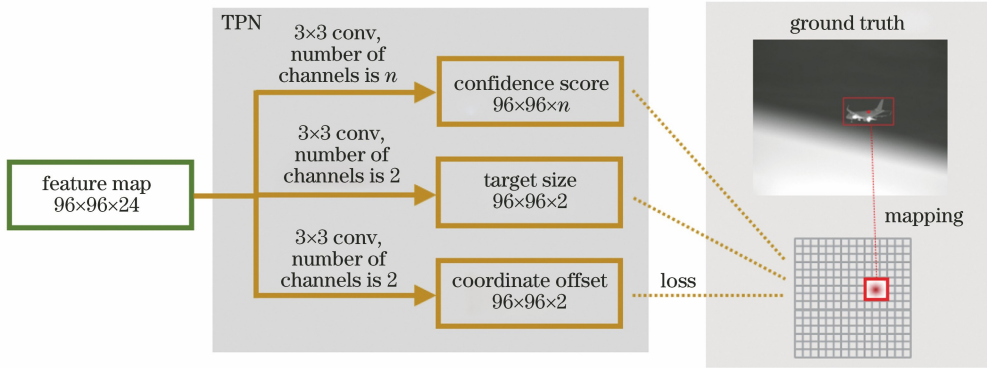


图 4 目标预测网络

Fig. 4 Target prediction network

每一类目标在特征图上对应点的置信度得分;2)目标中心点所对应目标的尺寸,即目标的宽 w 和高 h ;3)原始图像映射到特征图时,下采样引起的目标中心点坐标的偏差。

在FKPD模型训练阶段,针对这三类信息分别计算其损失函数以优化网络参数。对于目标中心点的预测,为了改善正负样本不平衡的问题,采用CornerNet中提出的修改版Focal Loss^[10,16]作为损失函数 L_k ,其表达式为

$$L_k = -\frac{1}{N} \sum_{xyc} \begin{cases} (1 - \hat{Y}_{xyc})^2 \ln(\hat{Y}_{xyc}) \\ (1 - Y_{xyc})^4 (\hat{Y}_{xyc})^2 \ln(1 - \hat{Y}_{xyc}) \end{cases}, \quad (1)$$

式中: N 为训练样本中真实目标的总数; \hat{Y}_{xyc} 为第 c 类目标中心点在特征图 (x, y) 处的置信度得分; Y_{xyc} 为第 c 类真实目标中心点在特征图 (x, y) 处的分布值,其在目标区域内的分布是以目标中心点为原点的高斯分布(最大值为1),在目标区域外取值为0。

对于目标尺寸的预测,仅需特征图上真实目标中心点处的预测值参与计算,其损失函数 L_s 的表达式为

$$L_s = \frac{1}{N} \sum_{k=1}^N |\hat{s}_k - s_k|, \quad (2)$$

式中: \hat{s}_k 为第 k 个目标的预测尺寸; s_k 为第 k 个目标的真实尺寸。

对于目标中心点坐标的偏差预测,与(2)式类似,亦仅需特征图上真实目标中心点处的预测值参与计算,其损失函数 L_o 的表达式为

$$L_o = \frac{1}{N} \sum_p \left| \hat{O}_{\tilde{p}} - \left(\frac{p}{R} - \tilde{p} \right) \right|, \quad (3)$$

式中: p 为真实目标中心点在原始图像上的坐标; R 为下采样率; \tilde{p} 为 p 经过下采样取整后在特征图上对应的坐标; $\left(\frac{p}{R} - \tilde{p} \right)$ 为下采样引起的目标中心点坐标偏差的真实值; $\hat{O}_{\tilde{p}}$ 为下采样引起的目标中心点坐标偏差的预测值。

结合(1)~(3)式,FKPD模型的总损失函数为 $L_{total} = L_k + \lambda_s L_s + \lambda_o L_o$ 。其中, λ_s 和 λ_o 分别为 L_s 和 L_o 的损失权重,在本文后续实验中二者均默认设置为 0.1。

在 FKPD 模型检测阶段,首先选择特征图中预测置信度得分最高的 m 个点作为目标中心点,要求点之间不能相连,之后根据设定的置信度阈值完成目标筛选。

3 实验与分析

3.1 训练与实验

本文所有模型均使用 Pytorch 深度学习框架来实现,并在搭载 Intel i9-9900ks 和 NVIDIA GTX1080Ti 的硬件平台上进行训练和测试。为了

验证模型的性能,本文使用了自建的红外数据集进行实验。依托地对空红外预警的应用场景,自建红外数据集取自以往外场实验数据,筛选并标注了 2758 张红外图像,其中包含飞鸟、战斗机、客机、直升机和训练机五大类共计 3000 个空中红外目标。实验中用于训练和测试的数据量比例为 7:3。同时考虑到文献中 YOLO 等相关模型大多采用 PASCAL VOC 数据集进行实验,因此为了更好地验证模型的性能,本文亦在该数据集上对 FKPD 进行了训练和测试。

在 FKPD 模型训练过程中,默认输入图片尺寸为 384 pixel×384 pixel(FKPD-384),每次训练默认 150 轮,采用 Adam 算法更新网络参数,Adam 算法的初始学习率为 1.25×10^{-3} ,当训练进行至第 70 轮和第 120 轮时,其学习率以 1/10 的速率递减。同时为了提高模型的泛化能力,输入图片在进入网络前每次都会随机进行翻转、平移和色调对比度调节等一系列预处理。表 1 对 FKPD 模型训练时的参数进行了总结。

3.2 检测精度分析

检测精度是评估快速检测模型性能的关键指标之一,通常期望模型在满足速度要求的同时,检测精度能够尽可能高,一般采用各类目标平均精度(AP)的均值 mAP(mean Average Precision)对其进行客观评价,mAP 值越大,检测精度越高。

表 1 FKPD 模型的训练参数

Table 1 Training parameters used for FKPD model

Parameter	Infrared dataset	PASCAL VOC dataset
Training epoch	150	150
Class number	5	20
Batch size	32	64
Default image size	384 pixel×384 pixel	384 pixel×384 pixel
Initial learning rate	1.25×10^{-3}	1.25×10^{-3}

表 2 为 FKPD-384 在红外数据集上的实验结果。作为对照,第一行和第二行分别为经典的标准检测模型 CenterNet-Res18 和 YOLOv3-Darknet53 的测试结果,第三行和第四行分别为目前业界广泛

应用的快速检测模型 Tiny-YOLOv3 和 FKPD-384 的测试结果。需要说明的是,表 2 中所有 YOLO 系列模型的输入图像尺寸均为 416 pixel×416 pixel,而 CenterNet-Res18 输入图像尺寸与 FKPD-384 一

表 2 红外数据集上的检测结果

Table 2 Detection results on infrared dataset

Model	mAP / %	AP / %				
		Bird	Fighter	Airliner	Helicopter	Trainer
CenterNet-Res18	88.04	76.73	88.95	94.91	90.77	88.84
YOLOv3-Darknet53	93.02	87.70	93.97	95.97	94.84	92.66
Tiny-YOLOv3	80.08	66.58	83.16	93.85	84.92	71.90
FKPD-384	88.98	79.40	90.84	95.01	90.27	89.39

致。分析表 2 中数据可知,与标准的检测模型相比,FKPD-384 模型的检测精度有一定的损失,但是考虑到计算量的巨大差距(模型尺寸仅为 YOLOv3-Darknet53 的 1/30),因而该损失是可以接受的;而与同类检测模型 Tiny-TOLOv3 相比,FKPD-384 的 mAP 值提高了 8.90%,提升幅度较大。

为了能够更加直观地观察检测效果的差异,图 5 比较了不同场景下的检测结果,其中数字为目标

的置信度。从图 5 中不难看出,相比 YOLOv3-Darknet53 等经典模型,无论是识别准确率还是定位精度,FKPD-384 均取得了比较接近的检测效果;而相比 Tiny-YOLOv3,得益于较强的特征提取能力,FKPD-384 对目标的识别明显更加精准,特别是在背景噪声比较大且目标较小的情况下,背景被识别为目标的虚警概率(False Alarm Rate, FAR)要低很多,且 FKPD-384 对目标尺寸的预测也更加准

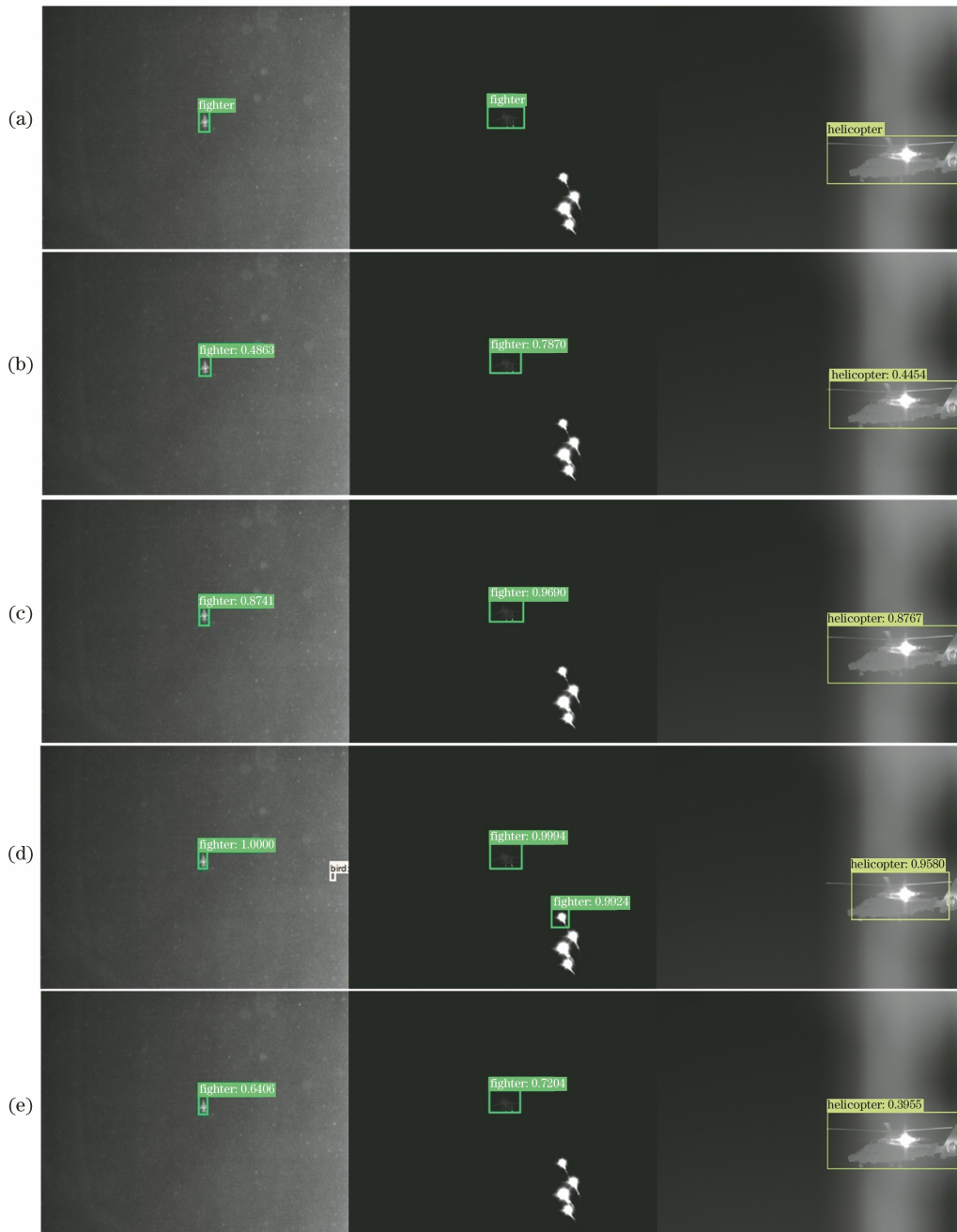


图 5 不同模型下红外目标检测效果对比。(a)标注图片;(b) CenterNet-Res18;(c) YOLOv3-Darknet53;
(d) Tiny-YOLOv3;(e) FKPD-384

Fig. 5 Comparison of infrared target detection results under different models. (a) Labeled image; (b) CenterNet-Res18;
(c) YOLOv3-Darknet53; (d) Tiny-YOLOv3; (e) FKPD-384

确,特别是一些尺寸相对较大的目标的定位误差较小。图 6 展示了更多场景下 FKPD 红外目标检测

示例,可以看到 FKPD 均比较好地完成了检测任务。



图 6 FKPD 红外目标检测示例

Fig. 6 Examples for FKPD infrared target detection

通过上述比较可知,FKPD-384 的目标虚警率较 Tiny-YOLOv3 要低很多,而作为模型在红外系统端应用的重要考核指标,表 3 对其进行了统计。其中,统计范围涵盖了测试集所有 828 张图片,统计结果为虚警目标数与检测到的目标总数之比。可以看到,FKPD-384 的虚警率与 CenterNet-Res18 和 YOLOv3-Darknet53 基本持平,而与 Tiny-YOLOv3 相比,降低了近 7.72%,印证了比较结果的准确性。

表 3 模型的虚警率

Table 3 False alarm rate of model

Model	FAR / %
CenterNet-Res18	1.23
YOLOv3-Darknet53	1.12
Tiny-YOLOv3	8.96
FKPD-384	1.24

除红外数据集外,表 4 给出了 FKPD-384 在 PASCAL VOC 数据集上的实验结果。分析表 4 中的数据可知,由于 PASCAL VOC 数据集中包含的目标种类更多(20 类),且可见光目标包含的特征信息更加复杂,因此与网络结构更深和拟合能力更强的标准模型相比,快速检测模型精度与其精度的差距比在红外数据集上更大;与 Tiny-YOLOv3 相比,FKPD-384 在检测精度上依然保持了领先,mAP 值较 Tiny-YOLOv3 提高了 3.21%。图 7 展示了不同模型在 PASCAL VOC 数据上的检测效果,可以看到,FKPD-384 的目标定位精度略好于 Tiny-YOLOv3,但与标准

模型存在一定的差距,与表 4 的实验结果一致。

表 4 PASCAL VOC 数据集上的检测结果

Table 4 Detection results on PASCAL VOC dataset

Model	mAP / %
CenterNet-Res18	68.24
YOLOv3-Darknet53	76.80
Tiny-YOLOv3	58.40
FKPD-384	61.61

3.3 检测速度分析

利用有限的系统资源,在确保检测精度的前提下,尽可能提高检测速度是快速检测方法设计的核心目标,亦是其后续在系统端应用的前提。表 5 从模型计算量、模型尺寸和 CPU 推理时间等多个方面对 FKPD 进行了对比测试。需要说明的是,为了与实际工程应用场景吻合,本节测试的模型推理时间均是在单次处理单张图像(384 pixel×384 pixel)的条件下连续处理 100 张红外图像后取平均的结果。

分析表 5 中的数据可知,在模型尺寸上,与 Tiny-YOLOv3 相比,FKPD-384 仅为 8.12 MB,压缩了近 76.61%。模型尺寸的减小意味着系统端所需的存储资源大幅减少,对于存储资源比较紧张的小型系统,这一优势十分明显。在模型计算量上,FKPD-384 的浮点运算量仅为 1.55×10^{10} ,比 Tiny-YOLOv3 压缩了近 33.76%。计算量的下降意味着系统计算资源的需求降低,特别是对于大多数仅包含 CPU 的红外系统而言,这一点尤为关键。与计



图 7 PASCAL VOC 数据集上的检测效果对比

Fig. 7 Comparison of detection effects on PASCAL VOC dataset

表 5 FKPD 的实时性分析
Table 5 Real-time analysis of FKPD

Model	Model size /	FLOPs /	CPU inference
	MB	10^{10}	time / ms
CenterNet-Res18	56.86	8.69	297.05
YOLOv3-Darknet53	246.18	27.93	844.29
Tiny-YOLOv3	34.72	2.34	129.08
FKPD-384	8.12	1.55	115.18

算量对应,在推理时间上,FKPD-384 的单帧图像检测时间可达 115.18 ms,比 Tiny-YOLOv3 提高了约 13.9 ms。结合 3.2 节检测精度的分析,FKPD-384 基本实现了在保证较好检测效果的同时提升检测速度的设计目标。

3.4 优化分析

本节旨在通过调节模型中的网络结构和参数(Ablation Experiments)来分析影响模型性能的关键因素,为下一步的优化和应用打下基础,表 6 对实验结果进行了总结,所有测试均是在红外数据集上完成的。其中,第一行为基础的 FKPD-384 的测试结果;第二行为去掉 FFN 后的测试结果;第三行为在 FEN 中增加空洞卷积(空洞数为 1)以扩大感受

野的测试结果;第四行为通过减少 FEN 第三至第五阶单元通道数(依次减少为 48、96 和 192)来进一步压缩网络的测试结果;第五行为将输入图像尺寸扩大至 512 pixel×512 pixel 的测试结果。

分析表 6 数据不难发现,去掉 FFN 后模型的检测速度基本没有变化,但是模型的检测精度下降明显,与 FKPD-384 相比 mAP 值损失了近 5.64%,对于特征信息较少的红外目标而言,需要通过融合不同层次的空间和语义信息来提高检测精度;通过增加空洞卷积来扩大特征图上特征点的感受野,这一方法对模型的检测精度并无提升,其主要原因是在以小目标为主的红外检测中,过大的感受野会导致目标对应的特征点包含大量目标外的背景信息,从而影响模型对目标的识别;通过牺牲一定的检测精度(约损失 5.01%)来进一步压缩网络,可将模型体积减少至 1.5 MB,检测时间亦仅需 67.98 ms,这一方法适用于一些对模型体量和实时性要求非常高的应用场景;通过扩大训练和检测图像的尺寸,可有效提高模型的检测精度,但是其带来的一个显著问题就是计算量的增加以及检测速度的下降,因此需要根据具体应用进行权衡。

表 6 FKPD 的消融实验
Table 6 Ablation experiment of FKPD

Method	mAP / %	Model size / MB	FLOPs / 10^{10}	CPU inference time / ms
FKPD-384(baseline)	88.98	8.12	1.55	115.18
FKPD-384 w/o FFN	84.34	8.12	1.55	114.89
FKPD-384 w Dilation	87.72	8.12	1.55	115.67
Compacted-FKPD-384	83.97	1.50	0.40	67.98
FKPD-512	90.15	8.12	2.76	174.29

4 结 论

为了满足小型红外探测系统实时检测的应用需求,提出了一种以目标中心为检测关键点的快速红外目标检测方法。在实现过程中,首先为了降低模型的计算量和提高检测速率,设计了轻量化的网路结构用于提取目标特征;其次针对红外目标尺寸小和特征少等特点,将目标不同层次特征的空间信息和语义信息进行融合,以提升目标的检测能力。在自建红外数据集上进行测试,与业界广泛应用的快速检测模型 Tiny-YOLOv3 相比,在检测精度提高 8.9% 的情况下,CPU 端的检测速度提高了 13.9 ms/frame,这验证了所提检测方法的有效性和实用性。下一阶段工作的重点将是在确保检测速度不降低的前提下,提升模型在复杂背景下的目标检测能力,同时优化模型的实现方式,为模型在系统端的部署打下基础。

参 考 文 献

- [1] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]//2014 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2014, Columbus, OH, USA. New York: IEEE, 2014: 14632381.
- [2] Xie J R, Li F M, Wei H, et al. Enhancement of single shot multibox detector for aerial infrared target detection[J]. Acta Optica Sinica, 2019, 39(6): 0615001.
谢江荣, 李范鸣, 卫红, 等. 用于空中红外目标检测的增强单发多框检测器方法[J]. 光学学报, 2019, 39(6): 0615001.
- [3] Wu S C, Zuo Z R. Small target detection in infrared images using deep convolutional neural networks[J]. Journal of Infrared and Millimeter Waves, 2019, 38(3): 371-380.
吴双忱, 左峥嵘. 基于深度卷积神经网络的红外小目标检测[J]. 红外与毫米波学报, 2019, 38(3): 371-380.
- [4] Wang D, Shen T. Research on weak and small infrared target detection algorithm under complex sky background[J]. Acta Optica Sinica, 2019, 39(5): 0512001.
王笛, 沈涛. 复杂天空背景下的红外弱小目标检测算法研究[J]. 光学学报, 2019, 39(5): 0512001.
- [5] Ma M Y, Wang D J, Sun H, et al. Infrared dim-small target detection based on robust principal component analysis and multi-point constant false alarm [J]. Acta Optica Sinica, 2019, 39(8): 0810001.
马铭阳, 王德江, 孙嵩, 等. 基于稳健主成分分析和多点恒虚警的红外弱小目标检测 [J]. 光学学报, 2019, 39(8): 0810001.
- [6] Ren S Q, He K, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.
- [7] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE, 2016: 16526932.
- [8] Redmon J, Farhadi A. YOLO9000: better, faster, stronger [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017: 17355115.
- [9] Redmon J, Farhadi A. Yolov3: an incremental improvement[EB/OL]. (2018-04-08) [2020-08-15]. <https://arxiv.org/abs/1804.02767>.
- [10] Law H, Deng J. CornerNet: detecting objects as paired keypoints [J]. International Journal of Computer Vision, 2020, 128(3): 642-656.
- [11] Zhou X Y, Wang D Q, Krähenbühl P. Objects as points [EB/OL]. (2019-05-25) [2020-08-15]. <https://arxiv.org/abs/1904.07850>.
- [12] Howard J, Zhu M L, Chen B, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications[EB/OL]. (2017-05-17) [2020-08-15]. <https://arxiv.org/abs/1704.04861>.
- [13] Sandler M, Howard J, Zhu M L, et al. Mobilenetv2: inverted residuals and linear bottlenecks [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, Utah, USA. New York: IEEE, 2018: 18311793.
- [14] Zhang X Y, Zhou X Y, Lin M X, et al. ShuffleNet: an extremely efficient convolutional neural network for mobile devices [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE, 2018: 18326147.
- [15] Ma N N, Zhang X Y, Zheng H T, et al. ShuffleNet V2: practical guidelines for efficient CNN architecture design [M]//Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision: ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11218: 122-138.
- [16] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(2): 318-327.