

DOI:10.19651/j.cnki.cmt.2107258

基于改进 YOLOv3-Tiny 的毫米波图像目标检测^{*}

陈国平 彭之玲 黄超意 管 春

(重庆邮电大学 光电工程学院 重庆 400065)

摘要: 毫米波是一种不具有电离辐射的电磁波,其能够穿透绝缘衣物布料和对人体无害的特性使得毫米波在人体安检领域有着巨大的发展前景。将深度学习方法运用至毫米波图像目标检测领域,提出一种基于改进 YOLOv3-Tiny 的毫米波图像目标检测方法。首先,在特征提取网络中增加卷积层提升网络深度,并增加至 3 个不同尺度的预测层加强对毫米波图像目标的检测能力;然后,在特征金字塔(FPN)中引入注意力机制(CBAM),使网络更关注毫米波图像中待测目标的特征,忽略背景噪声冗余的特征。结果表明,改进后的网络平均准确率可达 93.4%,单帧检测速度为 15 ms,模型参数仅为 38.7M,为毫米波安检系统高精度、小型化的研究提供了参考价值。

关键词: 毫米波图像;目标检测;注意力机制;YOLOv3-Tiny

中图分类号: TP391 **文献标识码:** A **国家标准学科分类代码:** 510.4

Millimeter wave image object detection based on improved YOLOv3-Tiny

Chen Guoping Peng Zhiling Huang Chaoyi Guan Chun

(School of Optoelectronic Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China)

Abstract: The millimeter-wave is an electromagnetic wave without ionizing radiation. It can penetrate the insulating cloth and is harmless to the human body. These characteristics make the millimeter wave have a wide range of application prospects in the field of public safety. Apply deep learning to the field of millimeter-wave image object detection, a millimeter-wave image object detection method based on improved YOLOv3-Tiny is proposed. Firstly, add convolutional layers to the feature extraction network to increase the depth of the network and increases to 3 different scale prediction layers to enhance the detection ability of millimeter-wave image object. Then, the convolutional block attention module is introduced in the feature pyramid network to make the network pay more attention to the features of targets and ignore the characteristics of redundant background noise. The results show that the improved network has mean average accuracy up to 93.4%, single frame detection speed is 15 ms, model parameters are only 38.7M, which provides a reference value for the research of high precision and miniaturization of millimeter wave security system.

Keywords: millimeter wave image;object detection;attention mechanism;YOLOv3-Tiny

0 引言

近年来,世界上的暴恐活动层出不穷,在机场、车站等公共场所保障人民的人身财产安全显得尤为重要。现在的公共场所都有着严格的安检措施,对于行李的检查主要采用 X 光射线扫描,对于人体的检查主要采用手持金属探测仪和金属探测门^[1]。这种人体安检方法需要耗费大量人力物力,且消耗时间较多,不满足安检环境快速检测的要求,也容易给乘客造成不适。因此,借助先进技术手段进行高效且人性化的安检,成为安检领域发展的首要研究目标。毫米波的频率通常在 30~300 GHz 之间,较高

频波而言穿透性更强,能够透过衣物布料显示藏匿的手枪、管制刀具、毒品等违禁物,且毫米波对人体无害,十分适用于在安检环境中对人体进行危险物品检测^[2]。但目前的毫米波安检研究成果也存在一些不足,毫米波图像分辨率较低,目标显示不清且与背景边缘模糊,难以使用肉眼辨认^[3]。因此,研究满足安检环境的毫米波图像目标检测算法具有重要价值。

传统的目标检测方法存在许多局限性,较差的泛化能力也不满足毫米波图像中复杂目标的检测需求。深度学习模型使用卷积神经网络^[4-5](convolutional neural network,

收稿日期:2021-07-13

^{*} 基金项目:国家自然科学基金(61671091)项目资助

CNN)提取图像的特征,使得目标检测发展到了一个新阶段,在毫米波图像目标识别方面也已经有了一些研究。姚家雄等^[6]利用 CNN 自动定位毫米波图像中的违禁物品,实现了目标的定位与识别。Zhang 等^[7]提出了一种基于阈值分割的快速区域卷积神经网络,并使用迁移学习的方法证明了深度学习模型在毫米波图像上的可行性。师君等^[8]利用热图检测和 YOLO 算法分别对毫米波图像进行目标检测,证明了两种方法在毫米波图像上的有效性,其中 YOLO 的检测算法有更好的速度和精度。侯冰基等^[9]改进 Faster-RCNN 网络优化毫米波图像的检测,利用反卷积恢复高度采样的特征图,提升对小物体的检测能力。这些工作推动了毫米波安检系统的发展,但日前在实际应用中仍然存在需解决的问题。毫米波图像中存在较多的小目标,且图像分辨率较低,与背景边界不清晰,这些问题给毫米波图像的目标检测带来了一定的困难^[10]。目前的毫米波图像目标检测算法在速度和精度上的平衡还有待提升,且大都有较大的运算量,安检环境需要轻量、易于移植的目标检测算法。

针对上述问题,本文以 YOLOv3-Tiny 模型为基础网络,根据毫米波图像的特点对网络进行改进,在不增加过多参数的情况下提升毫米波图像目标检测的效率。实验数据集包括隐藏了 3 类典型危险物品(大刀、小刀、枪)的毫米波人体图像,在 Pytorch 框架上通过实验对比改进后网络与 YOLOv3-Tiny 模型、YOLOv3 模型在毫米波数据集上的检测性能,验证了网络的有效性。

1 YOLO 网络原理

YOLO 系列网络作为单阶段检测网络,直接使用原始图像及标注信息进行回归和分类,这使得它在速度上比 Fast-RCNN^[11]、Faster-RCNN^[12] 这样的两阶段算法有极大地提高。YOLOv3^[13] 网络将 Darknet-53^[14] 作为主干网络,在保持速度的同时将精度提升到了一个档次,使得 YOLOv3 成为工业界解决许多问题的首选算法。YOLOv3-Tiny 是 YOLOv3 网络的简化版本,它在 YOLOv3 的基础上去掉了残差连接和卷积层,因此运行速度比 YOLOv3 快得多,体积更小,同时也牺牲了一些精度。

YOLOv3-Tiny 是一个轻量级的网络,由 13 个卷积层、6 个最大池化层、1 个上采样层、1 个级联层和 2 个输出层组成,其结构如图 1 所示。首先,输入层通过 10 个卷积层和 6 个池化层得到第 1 输出层,当输入图像大小为 416×416 时,输出特征图大小为 13×13 。然后,第 5 卷积层的输出经过上采样与第 8 卷积层的输出进行特征融合,再经过两个卷积得到第 2 输出层,大小为 26×26 。由两个输出层进行回归预测,能够直接检测目标的种类和位置。

YOLOv3-Tiny 简单的网络模型使得检测速度大大加快,同时也降低了对硬件设备的要求,但检测精度比起 YOLOv3 网络有所下降。

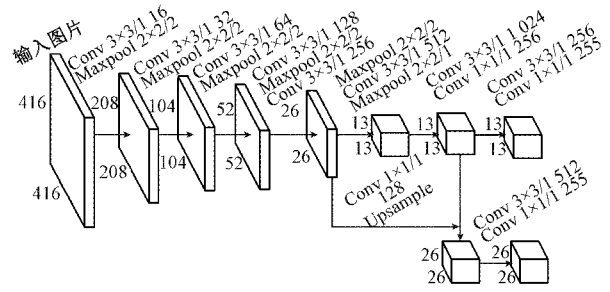


图 1 YOLOv3-Tiny 网络结构

2 改进的 YOLOv3-Tiny 网络

2.1 整体结构

YOLOv3-Tiny 网络深度较浅,对目标提取特征的能力不强,导致精度较低。为使 YOLOv3-Tiny 网络更适合毫米波图像的目标检测,主要的改进有两点:为提高 YOLOv3-Tiny 提取特征的能力,在 YOLOv3-Tiny 的特征提取网络中增加一系列 3×3 和 1×1 的卷积层,并增加一个分辨率较大的 YOLO 预测层,增强网络对小目标的检测能力;对于毫米波图像目标模糊、目标与边界不清晰、图层中存在较多冗余的背景和噪声信息的特点,在 FPN 特征金字塔中引入 CBAM 注意力机制,使网络更加关注待测目标,忽略不必要的信息,提高网络的检测精度。改进后的网络结构如图 2 所示。

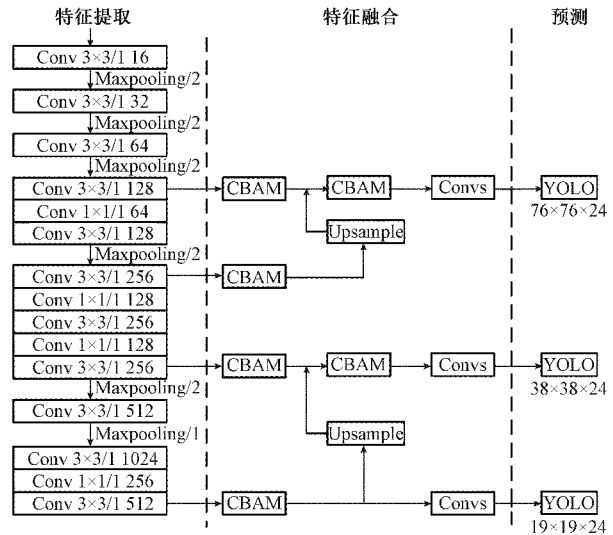


图 2 改进的 YOLOv3-Tiny 网络结构

2.2 特征提取网络改进

输入图像的分辨率影响网络提取到的信息量大小,为保证较多的图像信息输入到网络中,输入图片尺寸大小设定为 608×608 。YOLOv3-Tiny 的主干网络较浅,难以提取出更高层次的语义特征,为提高网络的性能,在 YOLOv3-Tiny 的基础上增加了一系列 3×3 和 1×1 的卷积层,改进后的特征提取网络如图 2 所示。其中, 3×3 的

卷积层目的在于提取特征,1×1 的卷积层则将特征图映射到较低的维度,避免因卷积层过多而产生大量的运算参数。这种做法可以降低网络的输入通道量,有效地在不增加过多参数的情况下提高网络的检测精度。为避免深层卷积导致毫米波图像中小目标信息的丢失,在原 YOLOv3-Tiny 网络 2 个预测层的基础上,增加一个浅层的特征融合提高对小目标的检测能力。将第 7 卷积层上采样后与第 4 卷积层进行特征融合,形成一个新的预测层,该预测层的大小为 76×76,包含更多小目标的信息。如图 2 所示,改进后的网络共有 3 个不同大小的预测分支,最后获得用于回归预测的 3 个特征图大小为 76×76,38×38 和 19×19。

YOLO 网络在预测时为不同尺度的特征图分配不同大小的预选框^[15],在训练时根据预选框不断调整最后输出预测边界框的结果。YOLO 网络的边界框预测如图 3 所示。

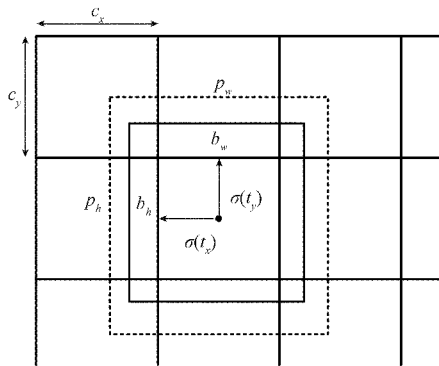


图 3 YOLO 网络的边界框预测

c_x, c_y 为特征图中左上角网格的坐标偏移量。 t_x, t_y, t_w, t_h 为每个预选框会预测的 4 个坐标。其中, t_x, t_y 表示预选框的坐标偏移量; t_w, t_h 为预选框的宽和高。最终输出预测边框的计算公式如下:

$$b_x = \sigma(t_x) + c_x \quad (1)$$

$$b_y = \sigma(t_y) + c_y \quad (2)$$

$$b_w = p_w e^{t_w} \quad (3)$$

$$b_h = p_h e^{t_h} \quad (4)$$

式中: b_x, b_y 表示最终预测框中心与左上角的距离; b_w, b_h 表示预测边框的宽和高; p_w, p_h 是先验框的宽和高,在 YOLO 网络中通过 K-means 聚类分析得到。

为适应新的数据集,使用 K-means 聚类算法^[16]对毫米波数据集中的目标进行聚类分析,得到 9 个先验框,再将这 9 个先验框均匀分配到 3 个尺度上。计算得到的 9 个先验框维度分别为 (35, 54)、(48, 42)、(59, 53)、(64, 70)、(78, 92)、(87, 64)、(54, 103)、(102, 48)、(107, 75)。分辨率较小的特征图有较大的感受野,对应大目标的先验框;分辨率较大的特征图感受野较小,对应小目标的先验框。最后,先验框与改进后 3 个特征图的对应关系如表 1 所示。

表 1 先验框与特征图对应关系

特征图大小	先验框
19×19	(54, 103), (102, 48), (107, 75)
38×38	(64, 70), (78, 92), (87, 64)
76×76	(35, 54), (48, 42), (59, 53)

2.3 引入注意力机制的特征融合层

在卷积神经网络中,浅层卷积具备更多底层细节信息,高层卷积则具备接近语义的特征信息,采用特征金字塔的方法能够融合高层和底层的特征信息,有效提升网络目标检测的能力。特征金字塔(feature pyramid network, FPN)^[17]的原理如图 4 所示,将高层特征图进行上采样后与低层特征图进行同维拼接,这种方式只是简单地在通道维度上进行拼接,不能反映出不同层特征图之间的重要性和相关性。

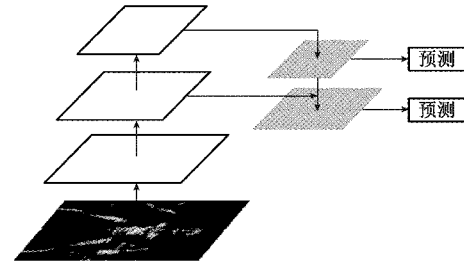


图 4 特征金字塔原理

卷积注意力机制模块(convolutional block attention module, CBAM)^[18]是一种通道维度和空间维度混合的注意力机制,它对上一层的空间和通道信息进行学习并得到相应的权重,其原理如图 5 所示。首先,对上一层特征图通道之间的关系进行学习,输入的特征图 F 的高、宽、通道数分别为 H, W, C ,经过通道注意力机制后得到一个大小为 $1 \times 1 \times C$ 的通道注意力。通道注意力表示了图层不同通道信息之间的权重,将该权重与原特征图相乘得到中间层特征图 F' ,再将 F' 输入到空间注意力机制中,得到大小为 $H \times W \times 1$ 的空间注意力。最后将空间注意力与 F' 相乘得到 F'' , F'' 为经过 CBAM 注意力机制融合后的特征图,输入到下一层网络中,公式为:

$$F' = F \times M_C(F) \quad (5)$$

$$F'' = F' \times M_S(F') \quad (6)$$

式中: M_C 表示通道注意力特征, M_S 表示空间注意力特征。

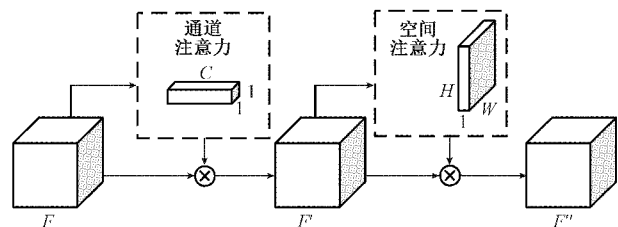


图 5 CBAM 注意力机制

在 FPN 中引入 CBAM 注意力机制,对重要特征图层

进行通道和空间上的注意力学习,改进后的特征融合层如图 6 所示。当特征提取网络输出两个不同大小的特征图进行特征融合时,先分别通过一个 CBAM 注意力机制进行权重学习,再进行上采样操作和同维拼接。接着对融合后的特征图再使用一个 CBAM 注意力机制,最后将特征图输入下一网络中。引入注意力机制的特征融合能自动学习图层中不同信息之间的重要性关系,对容易有目标的信息分配高权重,对背景噪声等冗余信息分配低权重,相比直接进行同维拼接的特征融合更关注可能出现目标的信息,能有效提升毫米波图像中目标的检测率。且 CBAM 注意力模块是一个轻量级的模块,在网络中加入并不会提升太多的参数量,能够在保持轻量的情况下使模型训练变得更加高效。

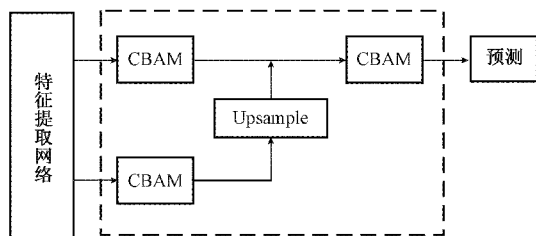


图 6 改进的特征融合层

3 实验结果与分析

3.1 实验数据集

本实验使用毫米波人体成像数据集,检测藏匿在人体身上不同位置的 3 典型危险物品(枪、小刀、大刀)。本次试验共使用了 6 000 张毫米波图像(枪、小刀、大刀各 2 000 张),每幅毫米波图片比例固定为 800 pixel×600 pixel。毫米波图像数据集按照 7:3 的比例划分为训练集和测试集,其中训练集和测试集的分布完全相同。

3.2 实验指标

实验通过单类待测目标的平均准确率(average precision, AP)、所有待测目标的平均准确率(mean average precision, mAP)、平均检测帧耗时来衡量目标检测算法的有效性。平均检测帧耗时指每张图片的平均检测时间,该指标能够直接反映模型在平台上的检测速度。模型内存大小表示在检测时的参数量大小。

AP 综合考量了查准率(precision)和召回率(recall)的影响,衡量了模型在该类别上的好坏,查准率和召回率的公式如下:

$$Precision = \frac{TP}{TP + FP} \quad (7)$$

$$Recall = \frac{TP}{TP + FN} \quad (8)$$

式中:TP、FP、FN 分别表示检测中真正例、假正例、假反例的数量。

3.3 实验结果分析

实验硬件为 Intel(R) Core(TM) i7-7820X CPU 和

NVIDIA 1080TI GPU,采用 Pytorch 深度学习框架完成实验。

实验中将训练图片的尺寸由 800×600 插值到 608×608 输入到网络中,初始阶段设置学习率为 0.001,衰减系数为 0.000 5,动量参数设置为 0.9,采用随机梯度下降法训练,随着训练次数的增加,减小学习率为 0.000 1 以达到收敛。将改进后的网络与 YOLOv3-Tiny 模型、YOLOv3 模型在毫米波数据集上的单类 AP 结果进行对比,其结果如表 2 所示。在输入图像大小和数据集相同的情况下,本文改进的模型网络对比原 YOLOv3-Tiny 网络在每个类别上的平均准确率都有所提高,其中对于小刀类别的平均准确率提升最大,比原 YOLOv3-Tiny 网络提升了 9.6%。YOLOv3-Tiny 网络对于小刀的检测精度较低,原因在于 YOLOv3-Tiny 网络结构简单,对图像提取特征的能力较弱,只有两个预测分支也缺乏对小目标的定位能力。同时小刀目标尺寸小,种类较多,在图像中显示较为模糊,能提取到的信息少,容易造成误检漏检。

表 2 单类 AP 的检测结果

模型	AP/%		
	枪	小刀	大刀
YOLOv3	96.7	89.4	97.9
YOLOv3-Tiny	87.5	78.6	89.7
本文模型	95.4	88.2	96.8

3 种网络在毫米波数据集上的检测能力比较如表 3 所示,评价指标包括模型的平均精度 mAP、平均帧检测速度、模型参数量大小。平均精度 mAP 是所有单类 AP 的均值,它衡量了网络在整个数据集上的好坏。改进后的网络增加了一些卷积层用于提取深层特征,增加的浅层特征融合层能够有效提升小目标的检测精度,融合注意力机制的特征融合层使网络忽略背景噪声等不重要的信息,能够整体提升网络的检测精度。实验证明,改进后的 YOLOv3-Tiny 网络的 mAP 为 93.4%,与原网络对比提升了 8.2%,模型大小比原网络增加了 5.6M,在不增加过多参数的情况下有效提升了检测精度。与 YOLOv3 网络相比,本文改进的网络在平均精度上比 YOLOv3 低了 1.2%,但模型大小仅为 YOLOv3 的 1/6,单帧检测耗时比 YOLOv3 网络减少了 5 ms,能够满足实际检测的需求。

表 3 不同模型检测能力比较

模型	mAP	平均帧耗时/ms	模型大小/M
YOLOv3	94.6	20	234.6
YOLOv3-Tiny	85.2	14	33.1
本文模型	93.4	15	38.7

本文改进的网络在毫米波图像数据集上的测试结果如

图 7 所示,可以看出,改进后的 YOLOv3-Tiny 网络能够有效检测出毫米波图像中的隐匿危险物品。综合考虑,本文改进的算法在不增加过多计算量的情况下对毫米波图像目标检测的精度和速度有了更好的平衡。

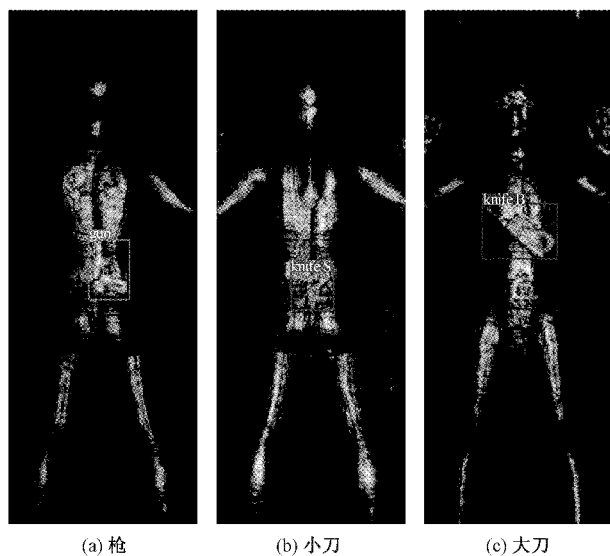


图 7 测试效果展示

4 结 论

本文针对安检环境中毫米波图像的实时目标检测,提出了一种基于改进 YOLOv3-Tiny 网络模型的毫米波图像目标检测方法。该网络模型修改了特征提取网络,通过增加一部分卷积层加强网络对特征的提取能力,并添加一个浅层的特征融合提高了对小目标的检测效果;在 FPN 特征金字塔中引入注意力机制,使网络自动学习重要特征图层上的空间关系和通道关系,减少了毫米波图像中背景噪声等冗余信息对结果的影响。实验证明了改进后的网络模型在毫米波图像数据集上的有效性,与 YOLOv3-Tiny 网络相比,本文提出的模型 mAP 提高了 8.2%,单帧检测速度为 15 ms,模型大小为 38.7M,能够满足安检环境中的检测需求,且模型参数量较小,为下一步在嵌入式设备上部署网络提供了参考。

参考文献

- [1] 蒋林华,王尉苏,童慧鑫,等. 太赫兹成像技术在人体安检领域的研究进展[J]. 上海理工大学学报, 2019, 41(1): 46-51.
- [2] 金颖康,王璞. 毫米波成像技术在人体安检领域的应用探讨[J]. 中国安防, 2019(9): 76-80.
- [3] 叶金晶,周健,孙谦晨,等. 主动毫米波成像隐私保护算法[J]. 红外与毫米波学报, 2017, 36(4): 505-512.
- [4] 郑远攀,李广阳,李晔. 深度学习在图像识别中的应用研究综述[J]. 计算机工程与应用, 2019, 55(12): 20-36.
- [5] 蓝金辉,王迪,申小盼. 卷积神经网络在视觉图像检测的研究进展[J]. 仪器仪表学报, 2020, 41(4): 167-182.
- [6] 姚家雄,杨明辉,朱玉琨,等. 利用卷积神经网络进行毫米波图像违禁物体定位[J]. 红外与毫米波学报, 2017, 36(3): 354-360.
- [7] ZHANG J, XING W, XING M, et al. Terahertz image detection with the improved faster region-based convolutional neural network[J]. Sensors, 2018, 18(7): 2327.
- [8] 师君,阙钰佳,周泽南,等. 近场毫米波三维成像与异物检测方法[J]. 雷达学报, 2019, 8(5): 578-588.
- [9] 侯冰基,杨明辉,孙晓玮. 基于改进 Faster RCNN 的毫米波图像实时目标检测[J]. 激光与光电子学进展, 2019, 56(13): 127-133.
- [10] 杜琨,王威,年丰,等. 主动毫米波图像的人体携带危险物检测研究[J]. 系统工程与电子技术, 2016, 38(6): 1462-1469.
- [11] GIRSHICK R. Fast R-CNN[C]. Proceedings of the IEEE International Conference on Computer Vision, 2015: 1440-1448.
- [12] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[C]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6): 1137-1149.
- [13] 周志锋,万旺根,王旭智. 基于 YOLO V3 框架改进的目标检测[J]. 电子测量技术, 2020, 43(18): 102-106.
- [14] 曹红燕,沈小林,刘长明,等. 改进的 YOLOv3 的红外目标检测算法[J]. 电子测量与仪器学报, 2020, 34(8): 188-194.
- [15] 张培培,王昭,王菲. 基于深度学习的图像目标检测算法研究[J]. 国外电子测量技术, 2020, 39(8): 34-39.
- [16] 黄韬,刘胜辉,谭艳娜. 基于 K-means 聚类算法的研究[J]. 计算机技术与发展, 2011, 21(7): 54-57.
- [17] LIN T, DOLLAR P, GIRSHICK R, et al. Feature Pyramid Networks for object detection[C]. IEEE Conference on Computer Vision & Pattern Recognition, 2017: 936-944.
- [18] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional block attention module[C]. European Conference on Computer Vision, 2018: 3-19.

作者简介

陈国平,教授,主要研究方向为毫米波近场成像、嵌入式系统和高性能 FPGA 并行计算。

彭之玲,硕士研究生,主要研究方向为深度学习、图像处理。

E-mail:pzlling@163.com