



中国图形学报

ISSN1006-8961
CN11-3758/TB



第27卷第8期（总第316期）
2022年8月16日

中国精品科技期刊
中国国际影响力优秀学术期刊
中国科技核心期刊
中文核心期刊

版权声明

凡向《中国图象图形学报》投稿，均视为同意在本刊网站及CNKI等全文数据库出版，所刊载论文已获得著作权人的授权。本刊所有图片均为非商业目的使用，所有内容，未经许可，不得转载或以其他方式使用。

Copyright

All rights reserved by Journal of Image and Graphics, Institute of Remote Sensing and Digital Earth, CAS. The content (including but not limited text, photo, etc) published in this journal is for non-commercial use.

主管单位 中国科学院
主办单位 中国科学院空天信息创新研究院
中国图象图形学学会
北京应用物理与计算数学研究所

主 编 吴一戎
编辑出版 《中国图象图形学报》编辑出版委员会
通信地址 北京市海淀区北四环西路19号
邮 编 100190
电子信箱 jig@aircas.ac.cn
电 话 010-58887035
网 址 www.cjig.cn

广告发布登记号 京朝工商广登字20170218号
总 发 行 北京报刊发行局
订 购 全国各地邮局
海外发行 中国国际图书贸易集团有限公司
(邮政信箱: 北京399信箱 邮编: 100048)
印刷装订 北京科信印刷有限公司

Journal of Image and Graphics

Title inscription: Song Jian | Monthly, Started in 1996

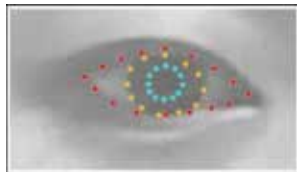
Superintended by Chinese Academy of Sciences
Sponsored by Aerospace Information Research Institute, CAS
China Society of Image and Graphics
Institute of Applied Physics and Computational Mathematics

Editor-in-Chief Wu Yirong
Editor, Publisher Editorial and Publishing Board of Journal of Image and Graphics
Address No. 19, North 4th Ring Road West, Haidian District, Beijing, P. R. China
Zip code 100190
E-mail jig@aircas.ac.cn
Telephone 010-58887035
Website www.cjig.cn

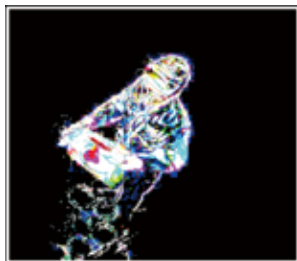
Distributed by Beijing Bureau for Distribution of Newspapers and Journals
Domestic All Local Post Offices in China
Overseas China International Book Trading Corporation
(P.O.Box 399, Beijing 100048, P.R.China)
Printed by Beijing Kexin Printing Co., Ltd.

CN 11-3758/TB
ISSN 1006-8961
CODEN ZTTXFZ

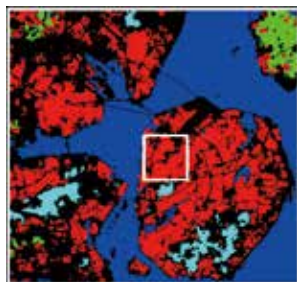
国外发行代号 M1406
国内邮发代号 82-831
国内定价 60.00元



自然光普通摄像头的眼部分割及特征点定位数据集ESLD (第2329页)



双视图三维卷积网络的工业装箱行为识别(第2368页)



多源特征自适应融合网络的高分遥感影像语义分割(第2516页)

综述

虚拟现实图像客观质量评价研究进展

周玉, 汪一, 李雷达, 高陈强, 卢兆林 2313

数据集论文

自然光普通摄像头的眼部分割及特征点定位数据集ESLD

张俊杰, 孙光民, 郑鲲, 李煜, 付晓辉, 慈康怡, 申俊杰, 孟凡超, 孔江萍, 张玥 2329

图像处理和编码

双尺度顺序填充的深度图像修复

陈东岳, 朱晓明, 马腾, 宋园园, 贾同 2344

二维码位流长度最小化算法

袁泰凌, 徐昆 2356

图像分析和识别

双视图三维卷积网络的工业装箱行为识别

胡海洋, 潘健, 李忠金 2368

面向航拍图像中工程车辆检测与识别的改进胶囊网络

钟映春, 郑海阳, 张文祥, 王波, 罗志勇 2380

用于骨架行为识别的多维特征融合注意力机制

姜权晏, 吴小俊, 徐天阳 2391

高性能整数倍稀疏网络行为识别研究

臧影, 刘天娇, 赵曙光, 杨东升 2404

面向工业零件分拣系统的低纹理目标检测

闫明, 陶大鹏, 普园媛 2418

融合策略优选和双注意力的单阶段目标检测

戴坤, 许立波, 黄世旸, 李鉴铃 2430

融合视觉词与自注意力机制的视频目标分割

季传俊, 陈亚当, 车洵 2444

基于Gaussian-Hermite矩的旋转运动模糊不变量

郭锐, 贾丽, 郝宏翔, 墨瀚林, 李华 2458

图像理解和计算机视觉

RGB-D语义分割: 深度信息的选择使用

赵经阳, 余昌黔, 桑农 2473

自纠正噪声标签的人脸美丽预测

甘俊英, 吴必诚, 翟懿奎, 何国辉, 麦超云, 白振峰 2487

面向非对称特征注意力和特征融合的太赫兹图像检测

曾文健, 朱艳, 沈韬, 曾凯, 刘英莉 2496

面向纹理平滑的方向性滤波尺度预测模型

林俊彦, 刘春晓, 章金凯, 李泓易 2506

遥感图像处理

多源特征自适应融合网络的高分遥感影像语义分割

张文凯, 刘文杰, 孙显, 许光奎, 付琨 2516

结合上下文编码与特征融合的SAR图像分割

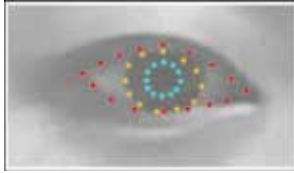
范艺华, 董张玉, 杨学志 2527

一阶全卷积遥感影像倾斜目标检测

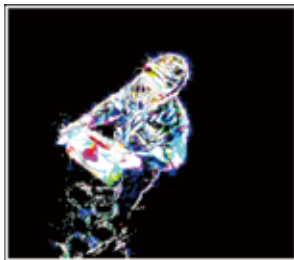
周院, 杨庆庆, 马强, 薛博维, 孔祥楠 2537

CONTENTS

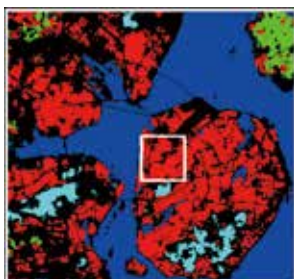
JOURNAL OF IMAGE AND GRAPHICS



ESLD: eyes segment and landmark detection in the wild (P2329)



Dual-view 3D ConvNets based industrial packing action recognition(P2368)



Multi-source features adaptation fusion network for semantic segmentation in high-resolution remote sensing images(P2516)

Review

Research progress in objective quality evaluation of virtual reality images

Zhou Yu, Wang Yi, Li Leida, Gao Chenqiang, Lu Zhaolin 2313

Dataset

ESLD: eyes segment and landmark detection in the wild

Zhang Junjie, Sun Guangmin, Zheng Kun, Li Yu, Fu Xiaohui, Ci Kangyi, Shen Junjie, Meng Fan-
chao, Kong Jiangping, Zhang Yue 2329

Image Processing and Coding

Depth image recovery based on dual-scale sequential optimized filling

Chen Dongyue, Zhu Xiaoming, Ma Teng, Song Yuanyuan, Jia Tong 2344

Minimization of bit stream length of QR codes

Yuan Tailing, Xu Kun 2356

Image Analysis and Recognition

Dual-view 3D ConvNets based industrial packing action recognition

Hu Haiyang, Pan Jian, Li Zhongjin 2368

Improved capsule network method for engineering vehicles detection and recognition in aerial ima-
ges

Zhong Yingchun, Zheng Haiyang, Zhang Wenxiang, Wang Bo, Luo Zhiyong 2380

M2FA: multi-dimensional feature fusion attention mechanism for skeleton-based action recognition

Jiang Quanyan, Wu Xiaojun, Xu Tianyang 2391

Action recognition analysis derived of integer sparse network

Zang Ying, Liu Tianjiao, Zhao Shuguang, Yang Dongsheng 2404

Texture-less object detection method for industrial components picking system

Yan Ming, Tao Dapeng, Pu Yuanyuan 2418

Single stage object detection algorithm based on fusing strategy optimization selection and dual
attention mechanism

Dai Kun, Xu Libo, Huang Shiyang, Li Yunling 2430

Visual words and self-attention mechanism fusion based video object segmentation method

Ji Chuanjun, Chen Yadang, Che Xun 2444

Rotational motion blur invariants based on Gaussian-Hermite moments

Guo Rui, Jia Li, Hao Hongxiang, Mo Hanlin, Li Hua 2458

Image Understanding and Computer Vision

RGB-D semantic segmentation: depth information selection

Zhao Jingyang, Yu Changqian, Sang Nong 2473

Self-correcting noise labels for facial beauty prediction

Gan Junying, Wu Bicheng, Zhai Yikui, He Guohui, Mai Chaoyun, Bai Zhenfeng 2487

Terahertz image detection combining asymmetric feature attention and feature fusion

Zeng Wenjian, Zhu Yan, Shen Tao, Zeng Kai, Liu Yingli 2496

Texture-smoothing-oriented directional filtering scales-predicting model

Lin Junyan, Liu Chunxiao, Zhang Jinkai, Li Hongyi 2506

Remote Sensing Image Processing

Multi-source features adaptation fusion network for semantic segmentation in high-resolution re-
mote sensing images

Zhang Wenkai, Liu Wenjie, Sun Xian, Xu Guangluan, Fu Kun 2516

The integrated contextual encoding and feature fusion SAR images segmentation method

Fan Yihua, Dong Zhangyu, Yang Xuezhi 2527

Improved one-stage fully convolutional network for oblique object detection in remote sensing ima-
gery

Zhou Yuan, Yang Qingqing, Ma Qiang, Xue Bowei, Kong Xiangnan 2537

中图法分类号: TP391 文献标识码: A 文章编号: 1006-8961(2022)08-2496-10

论文引用格式: Zeng W J, Zhu Y, Shen T, Zeng K and Liu Y L. 2022. Terahertz image detection combining asymmetric feature attention and feature fusion. Journal of Image and Graphics, 27(08): 2496-2505 (曾文健, 朱艳, 沈韬, 曾凯, 刘英莉. 2022. 面向非对称特征注意力和特征融合的太赫兹图像检测. 中国图象图形学报, 27(08): 2496-2505) [DOI: 10.11834/jig.210095]

面向非对称特征注意力和特征融合的太赫兹图像检测

曾文健, 朱艳*, 沈韬, 曾凯, 刘英莉

1. 昆明理工大学信息工程与自动化学院, 昆明 650500; 2. 昆明理工大学云南省计算机技术应用重点实验室, 昆明 650500

摘要: **目的** 太赫兹由于穿透性强、对人体无害等特性在安检领域中得到了广泛关注。太赫兹图像中目标尺寸较小、特征有限,且图像分辨率低,目标边缘信息模糊,目标信息容易和背景信息混淆,为太赫兹图像检测带来了一定困难。**方法** 本文在 YOLO(you only look once)算法的基础上提出了一种融合非对称特征注意力和特征融合的目标检测网络 AFA-YOLO(asymmetric feature attention-YOLO)。在特征提取网络 CSPDarkNet53(cross stage partial DarkNet53)中设计了非对称特征注意力模块。该模块在浅层网络中采用非对称卷积强化了网络的特征提取能力,帮助网络模型在目标特征有限的太赫兹图像中提取到更有效的目标信息;使用通道注意力和空间注意力机制使网络更加关注图像中目标的重要信息,抑制与目标无关的背景信息;AFA-YOLO 通过增加网络中低层到高层的信息传输路径对高层特征进行特征融合,充分利用到低层高分辨率特征进行小目标的检测。**结果** 本文在太赫兹数据集上进行了相关实验,相比原 YOLOv4 算法,AFA-YOLO 对 phone 的检测精度为 81.15%,提升了 4.12%,knife 的检测精度为 83.06%,提升了 3.72%。模型平均精度均值(mean average precision, mAP)为 82.36%,提升了 3.92%,漏警率(missing alarm, MA)为 12.78%,降低了 2.65%,帧率为 32.26 帧/s,降低了 4.06 帧/s。同时,本文在太赫兹数据集上对比了不同的检测算法,综合检测速度、检测精度和漏警率,AFA-YOLO 优于其他目标检测算法。**结论** 本文提出的 AFA-YOLO 算法在保证实时性检测的同时有效提升了太赫兹图像中目标的检测精度并降低了漏警率。

关键词: 太赫兹图像;小尺度目标检测;YOLOv4;非对称卷积;注意力机制;特征融合

Terahertz image detection combining asymmetric feature attention and feature fusion

Zeng Wenjian, Zhu Yan*, Shen Tao, Zeng Kai, Liu Yingli

1. Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, China;

2. Key Laboratory of Computer Technology Application of Yunnan Province, Kunming University of Science and Technology, Kunming 650500, China

Abstract: **Objective** Terahertz technology has great application potentials in related to wireless communications, biomedicine, and non-destructive testing. Some terahertz imaging features are suitable for hidden objects detection for human security inspections because terahertz waves can penetrate the substance like ceramics, plastics and cloths and are largely

收稿日期: 2021-03-01; 修回日期: 2021-04-22; 预印本日期: 2021-04-29

* 通信作者: 朱艳 zhuyan@kust.edu.cn

基金项目: 国家自然科学基金项目(61971208, 61671225, 52061020, 61702128); 云南省应用基础研究计划重点项目(2018FA034); 云南省中青年学术技术带头人后备人才项目(Shen Tao, 2018); 云南省万人计划青年拔尖人才项目(201873); 昆明理工大学人才培养项目(KKSY201703016)

Supported by: National Natural Science Foundation of China (61971208, 61671225, 52061020, 61702128)

absorbed or reflected by metals, liquids and other substances with no harmless on human body. With the development of terahertz imaging technology and the increasing flow of people in application scenarios, the use of artificial recognition of terahertz images is no longer applicable. In order to solve the problem of hidden objects and dangerous goods detection, current research have focused on using deep learning method to classify and analyze them. Due to the resolution and contrast of the terahertz image are low, the edge information of the target in the image is blurred, the target information is easily confused with the background information, and the target information is unclear in the terahertz image, and the feature information is limited. Therefore, the effectiveness issue of feature information to detect the target in terahertz image is challenged for terahertz image detection. **Method** We facilitates a target detection framework asymmetric feature attention-you only look once (AFA-YOLO) that combines asymmetric feature attention and feature fusion based on the you only look once v4 (YOLOv4) algorithm to resolve the barriers of small-scale target detection in terahertz images. cross stage partial DarkNet53 (CSPDarkNet53) is as a feature extraction network for AFA-YOLO and an asymmetric feature attention module in CSPDarkNet53 is designed. First, this module uses asymmetric convolution in the shallow network to enhance the feature extraction capabilities of the network, helping the network model to extract more effective target information from the terahertz image with limited target features; Second, the module melts convolutional block attention module (CBAM) attention force mechanism via using the channel attention mechanism to make the model pay more attention to the important information of the target in the image, suppress unrelated background information to the target, and use the spatial attention mechanism to pay attention to the position information of the target in the terahertz image, allowing the model to optimize target contexts. AFA-YOLO has carried out feature fusion operations as well, the high-level features are enhanced through increasing the information transmission path from the low-level to the high-level in the network. The high-level feature map can obtain fine-grained target appearance information and the positioning and detection of small-scale targets in terahertz images can be optimized. **Result** Our research uses the detection accuracy map, missed alarm (MA) rate and detection speed frames per second (FPS) as indicators to carry out related experiments on the terahertz data set. The detection accuracy of AFA-YOLO for the phone is 81.15% compare to the original YOLOv4 algorithm, which is an increase of 4.12%. The detection accuracy of knife is 83.06%, an increased ratio of 3.72%. The model mean average precision (mAP) is 82.36%, which is an increased ratio of 3.92%. The MA is 12.78%, which is a decreased ratio of 2.65%, and the FPS is 32.26, which is lower to 4.06. Additionally, we conduct comparative analysis of different detection algorithms on the terahertz dataset. AFA-YOLO optimized target detection algorithms in terms of recognized detection speed, detection accuracy and missed alarm rate. **Conclusion** We facilitate an AFA-YOLO detection framework that combines asymmetric feature attention and feature fusion. Our YOLOv4-based framework melts asymmetric feature attention module into the shallow network and enhances the target information on the high-level feature map. The optimal information ensures real-time detection through improving the detection accuracy of the target in the terahertz image effectively and lowering the missed alarm rate.

Key words: terahertz image; small target detection; YOLOv4; asymmetric convolution; attention mechanism; feature fusion

0 引言

太赫兹(Mittleman,2017)泛指频率在0.1~10 THz波段内的电磁波,是21世纪最具潜力的新兴技术之一,在无线通信、生物医学和无损检测都有着巨大的应用前景。由于太赫兹波对人体无害,并且可以穿透陶瓷、塑料和布料等,同时被金属、液体等物质大量吸收或反射,这些特性使得太赫兹成像非常适用于人体安检(Feng等,2020)中的隐匿物品检测。

近年来,基于卷积神经网络(convolutional neu-

ral network,CNN)的目标检测算法以其优异的检测性能和快速处理数据的能力在目标检测领域得到了广泛应用。现阶段的目标检测算法按检测流程分为两类;一类是以Faster R-CNN(region-convolutional neural network)(Ren等,2017)、Mask R-CNN(He等,2017)为代表的具有区域建议的两阶段目标检测算法,这些检测算法在检测精度上较高但是检测速度慢。另一类是以SSD(single shot multibox detector)(Liu等,2016)、YOLOv4(you only look once)(Bochkovskiy等,2020)等为代表的单阶段目标检测算法,这些检测算法能实现实时性的快速检测,但是

在检测精度上低于两阶段的目标检测算法。

基于深度学习的目标检测算法虽然在对自然光学图像的检测中取得了较好的检测效果,但由于太赫兹图像的对比度较低,图像中目标的边缘信息模糊,导致背景信息容易与目标信息混淆;同时,由于太赫兹图像中的目标尺寸较小,目标特征有限,最后导致基于深度学习的目标检测方法对太赫兹图像的检测效果较差。因此,如何利用有限的特征信息来准确检测太赫兹图像中的目标是太赫兹图像检测的难点。

为了解决上述问题,本文在 YOLOv4 的基础上提出一种融合非对称特征注意力和特征融合的目标检测网络 AFA-YOLO (asymmetric feature attention-YOLO)。AFA-YOLO 在检测模型中加入非对称卷积(Ding 等,2019)来增强模型对目标的特征提取能力,并且使用 CBAM (convolutional block attention module)(Woo 等,2018)注意力机制模块使模型关注图像中的有用特征,抑制冗余特征,从而减少背景信息对目标产生的干扰,降低漏检情况。本文的主要贡献如下:

1)以 YOLOv4 检测网络为基础,针对太赫兹图像检测提出了一种融合非对称特征注意力和特征融合的目标检测网络 AFA-YOLO,能有效提高对太赫兹图像中目标的检测精度并降低漏警率。

2)设计了一种非对称特征注意力模块,该模块通过非对称卷积增强了网络对目标的特征提取能力,利用通道注意力关注到图像中目标的重要通道信息,利用空间注意力使网络关注图像中的关键区域。

3)通过增加网络中的信息传输路径对高层特征进行特征增强,将低层高分辨率特征和具有丰富语义信息的高层特征融合,充分利用低层特征检测太赫兹图像中的小尺度目标。

4)对比了不同的目标检测算法在太赫兹数据集上的检测效果。在满足实时性检测的前提下,AFA-YOLO 在对太赫兹图像中小目标的检测上获得了最高的检测精度以及最低的漏警率,检测精度为 82.36%,漏警率为 12.78%。

1 相关工作

1.1 相关检测算法

Faster R-CNN 是以 VGG (Visual Geometry

Group)网络为基础的两阶段目标检测算法,该算法将候选区域生成、特征提取、分类和位置精修统一到一个网络结构中,真正实现了端到端的目标检测,提高了检测速度和检测精度。但是在获取区域建议的过程中,由于计算量较大,无法达到实时性检测。Mask R-CNN 对 Faster R-CNN 进行了扩展,添加了一个分支,使用现有的检测对目标进行并行的预测,并且在类别预测时做到了像素级别。

YOLO (Redmon 等,2016)是常见的单阶段检测算法之一,YOLO 将目标检测任务转化成一个回归问题,大大加快了检测速度。与 Faster R-CNN 不同的是,YOLO 算法在训练和预测过程中利用全图信息,因此可以将背景预测错误率降低一半。但 YOLO 算法对小尺度目标的检测效果不太好,随后 YOLO9000 (Redmon 和 Farhadi,2017)、YOLOv3 (Redmon 和 Farhadi,2018)等算法针对这一问题在不同程度上有了相关改进。本文中用到的 YOLOv4 算法平衡了检测精度和速度,可以在一块普通的 GPU 上完成训练,最后达到实时性检测。YOLOv4 先通过由 53 个卷积层组成的网络 CSPDarkNet53 (cross stage partial DarkNet53)进行特征提取,再由空间金字塔池化 (spatial pyramid pooling, SPP)进行多尺度特征融合,同时路径聚合网络 (path aggregation network, PANet) (Wang 等,2019)会对提取到的特征进行特征增强,最后将特征图送入检测层进行目标检测操作。

SSD 是常见的单阶段检测算法之一。主网络结构是 VGG16,并将最后两个全连接层改成了卷积层。SSD 算法与 YOLO 的不同之处是除了在最终特征图上做目标检测外,还在之前选取的 5 个特征图上进行预测,虽然提高了对小尺度目标的检测精度,但是 SSD 算法的缺点也很明显,需要人工设置先验框的尺寸大小和比例,并且对目标的召回率较低。

上述检测方法都实现了端到端的目标检测,将深度学习中的检测方法引入太赫兹图像的检测中可以大大减少人力资源的使用,同时提高安检场景中的检测效率。但上述方法并不针对太赫兹图像中的目标检测,在对太赫兹图像的检测中检测精度较低,同时存在较高的漏检。

1.2 太赫兹图像目标检测

在对太赫兹图像的检测研究中,目前研究人员通过改进相关的检测算法来提高对太赫兹图像的检

测效果。Xiao 等人(2018)提出了一种结合预处理和结构优化的框架 R-PCNN (region-preprocessing convolutional neural networks), 在 Faster R-CNN 网络中加入图像去噪和增强模块来解决传统的图像分割和定位方法对太赫兹图像检测精度和速度不够的问题。Zhang 等人(2018)分析了太赫兹图像与光学图像的相似性, 提出了基于光学特征迁移的分类方法和一种阈值分割结合 Faster R-CNN 的检测方法, 用于独立检测人体和其他物体以提高检测性能; Yang 等人(2019)通过稀疏低秩分解方法挖掘太赫兹图像的时空信息, 实现对可疑对象的高精度自动检测和识别。

1.3 小尺度目标检测

在目标检测中, 由于小尺度目标在图像中的信息较少导致检测小尺度目标要比检测中型和大型目标更困难。在 COCO (common objects in context) 评价标准中, 图像中小于或者等于 32×32 像素的物体通常归类为小尺度目标。当前研究人员针对小尺寸目标检测也展开了相关的研究。Sommer 等人(2017)用 5 个卷积层和 3 个全连接层组成的浅层神经网络来实现对航空图像中的小尺度目标检测。Gao 等人(2018)先对图像进行下采样, 然后使用强化学习动态地搜索感兴趣区域, 最后以更高的分辨率对该区域进行小尺度目标检测。陈冰曲和邓涛(2019)针对车辆数据集重新设计了 SSD

算法的区域候选框, 并在 SSD 算法的损失函数基础上增加排斥损失提高对重叠目标检测。Liu 等人(2020)提出一种基于反卷积区域的卷积神经网络来实现准确检测交通标志的问题。郭璠等人(2021)提出了一种基于语义分割的通道注意方法和空间注意方法 YOLOv3-A 来增强模型对交通标志的检测性能。

2 本文方法

2.1 AFA-YOLO 检测网络

AFA-YOLO 检测网络在 CSPDarkNet53 的浅层部分设计了一种非对称特征注意力模块, 该模块结合了非对称卷积和 CBAM 注意力机制的优点来强化网络模型对太赫兹图像中目标的特征提取能力。图 1 为 AFA-YOLO 的网络结构图, 输入图像先经过非对称特征注意力模块和特征提取网络 CSPDarkNet53 进行特征提取, 再由 SPP 流向路径聚合网络。由于小尺度目标依赖浅层网络提取的低层特征信息, 因此, AFA-YOLO 检测框架中进行了特征融合, 将经过非对称特征注意力模块后的低层高分辨率特征和高层具有丰富语义信息的特征融合, 充分利用低层特征图中目标的外观信息进行定位, 如图 1 中虚线所示, 最后网络结合不同分辨率的 YOLO 检测层进行目标检测。

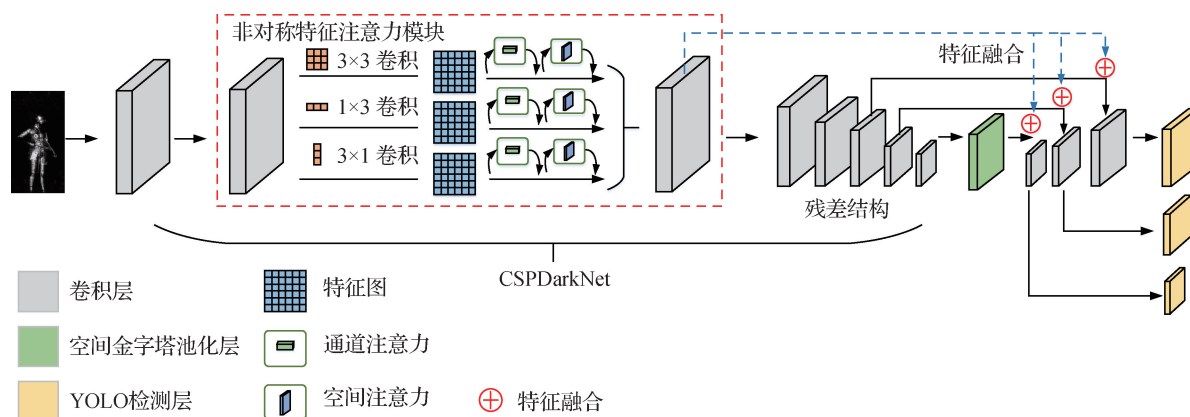


图 1 AFA-YOLO 网络结构图

Fig. 1 Structure diagram of AFA-YOLO network

2.2 非对称特征注意力模块

2.2.1 非对称卷积

在太赫兹图像中, 由于图像中目标尺寸相对较小, 特征信息有限且图像中目标的边缘信息模糊, 普

通的方形卷积核对于太赫兹图像中目标的特征提取能力不足, 不能突出卷积核提取到的目标特征。非对称卷积相比于普通方形的卷积增加了水平卷积和竖直卷积, 新增的两个卷积核增强了方形卷积核中

心骨架的特征提取能力,从而强化了方形卷积核中心的特征,弱化了边缘的特征,因此,非对称卷积能很好地帮助网络提取到目标的特征。图2为非对称卷积。

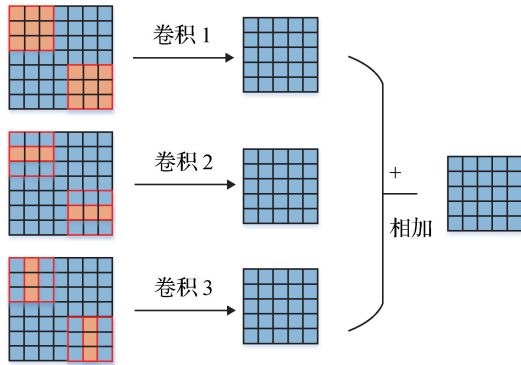


图2 非对称卷积

Fig. 2 Asymmetric convolution

在 AFA-YOLO 的网络中,本文在特征提取网络 CSPDarkNet53 第 1 个卷积层之前使用非对称卷积将现有的 3×3 卷积核替换成 3×3 、 1×3 、 3×1 这 3 个并行的卷积核,这 3 个卷积核会分别对同一个输入特征进行特征提取操作,然后将这 3 个卷积核提取到的特征进行融合后得到该卷积层的输出结果。

2.2.2 通道注意力机制

注意力机制模块的主要功能是为了增加检测模型的数据表征能力,使网络学习到图像特征中的重要信息并抑制不重要的信息。按照注意力作用的特征形式,注意力机制可以分为基于通道的注意力和基于空间的注意力。

图像的通道信息一般代表不同的特征信息,图像的通道数越多,包含的特征信息越丰富。使用通道注意力机制来选择通道中的关键特征,同时抑制与目标无关的特征,有利于将太赫兹图像中的目标信息和背景信息分离,从而减少模型的漏检情况。图3为通道注意力机制。

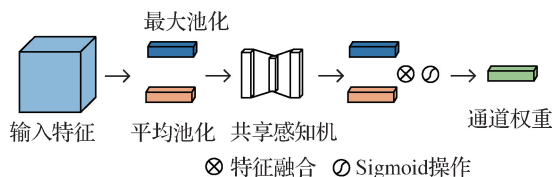


图3 通道注意力

Fig. 3 Channel attention

图3中对一个输入特征分别进行空间的全局平

均池化和最大池化后得到两个 1 维向量,然后将其输入到一个共享感知机中,再将得到的两个特征相加后经过一个 Sigmoid 激活函数得到权重系数 M_c 。最后,将得到的权重系数和原来的特征相乘即可得到经过通道注意力后的新特征。实现通道注意力的方法为

$$M_c(F) = \sigma(W_1(W_0(F_a)) + W_1(W_0(F_m))) \quad (1)$$

式中, F_a 和 F_m 分别表示原特征图经过平均池化和最大池化后的特征图, W_0 、 W_1 均为感知机的可学习参数, σ 表示 Sigmoid 函数, M_c 表示得到的通道注意力权重。

2.2.3 空间注意力机制

空间注意力关注的是图像中目标的位置信息。利用空间注意力机制可以增加目标关键区域的权重,使网络关注到太赫兹图像中的重点区域,有助于网络提取到目标关键特征。图4为空间注意力机制。

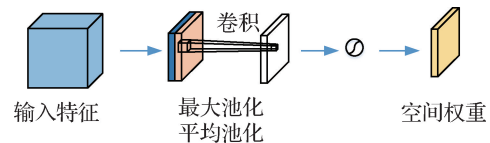


图4 空间注意力

Fig. 4 Spatial attention

经过通道注意力后得到的特征,依次经过通道维度的平均池化和最大池化操作得到两个 2 维特征 F_a^s 、 F_m^s ,将其按通道维度拼接在一起后再由一个卷积层提取特征。

在 CBAM 的空间注意力中,Woo 等人(2018)证明了使用 7×7 的卷积核比 3×3 的卷积核效果更好,因此本文同样使用 7×7 的卷积核来提取特征。然后,特征经过 Sigmoid 激活函数得到空间注意力权重 M_s 。为了保证最后得到的特征在空间维度上与输入的特征一致,将权重特征 M_s 和原始特征相乘即可得到缩放后的新特征。实现空间注意力的方法为

$$M_s(F) = \sigma(f([F_a^s; F_m^s])) \quad (2)$$

式中, F_a^s 和 F_m^s 分别表示经过平均池化和最大池化后的特征图, f 表示卷积操作, σ 表示 Sigmoid 函数, M_s 为空间注意力权重。

2.3 特征融合

神经网络通过一系列卷积和池化操作可以提取

到图像中目标的相关特征。浅层网络的感知域较小,可以学习到局部区域的特征,深层网络具有较大的感知域,能够学习到更加抽象的特征。

在 AFA-YOLO 网络中,随着网络深度的增加,神经网络提取到特征的语义信息也就越强,这些特征对物体的大小、位置和方向等敏感性更低,因此有助于检测性能的提高。但随着网络深度的增加,特征图的分辨率会逐渐降低,并且高层特征丢失了目标外观的细节信息,因此在检测太赫兹图像中这些尺寸较小的目标时需要高分辨率的低层特征来定位目标。本文通过增加网络中的信息传输路径,如图5中虚线所示,让低层特征和高层特征以 add(特征图相加,通道数不变)方式进行特征融合,充分利用高分辨率的低层特征来提高网络模型对小尺度目标的检测效果。图5为低层特征和高层特征的融合方式。

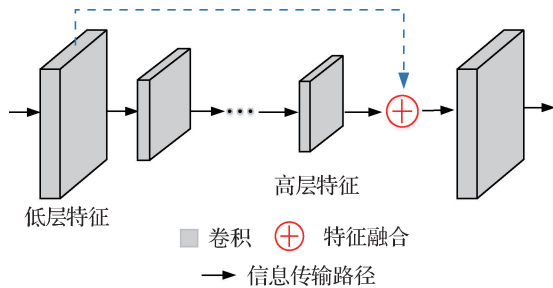


图5 特征融合

Fig.5 Feature fusion

将浅层的高分辨率特征图通过下采样的方式与不同的特征进行融合,使得高层特征图能获取到细粒度的目标外观信息,从而优化对太赫兹图像中小尺度目标的定位和检测效果。add方式的特征融合方法为

$$Z = \sum_{i=1}^c (X_i + Y_i) * K_i = \sum_{i=1}^c X_i * K_i + \sum_{i=1}^c Y_i * K_i \quad (3)$$

式中, X 和 Y 分别表示两组输入特征的通道, K 表示卷积核, Z 表示输出特征, c 为特征图通道的最大值。

3 实验结果与分析

3.1 太赫兹数据集

本文所使用的太赫兹数据集包含了 7 711 幅不同的太赫兹图像,每幅图像的尺寸为 200×380 像素,如图6所示。对收集到的 7 711 幅太赫兹图像用 LabelImg 软件进行标注,将图像中的检测目标分为两类,分别为 phone 和 knife,标注过程中发现所有目标均小于 32×32 像素。标注完成后,每一幅太赫兹图像都对应有一个 XML (extensible markup language) 格式的标注文件,文件里记录着目标的位置信息。最后,将太赫兹图像制作成 VOC (visual object classes) 类型的数据集。

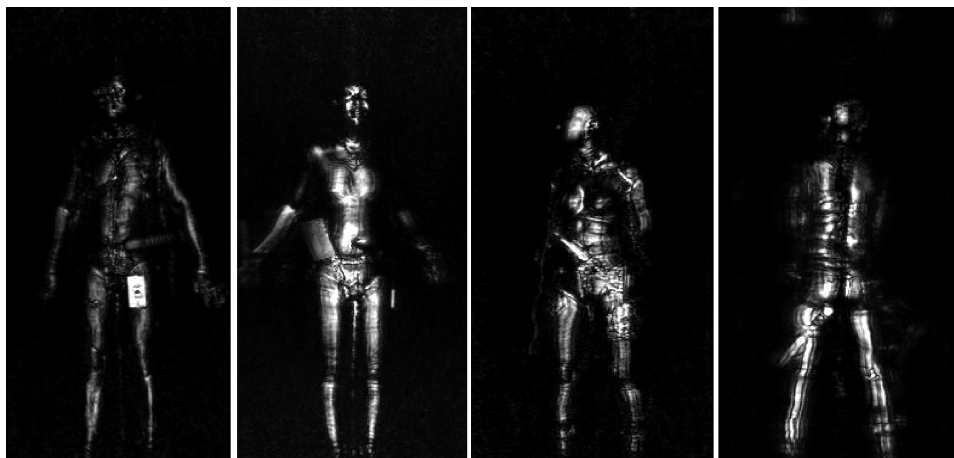


图6 太赫兹图像

Fig.6 Terahertz images

3.2 实验设置

表1为本文实验环境配置表,实验在 Ubuntu16.04 系统下进行,处理器型号为 Intel(R) Core(TM) i5-9400F CPU @ 2.90 GHz,使用 Nvidia RTX

2060 显卡,内存为 16 GB,开发语言为 python3.6,深度学习框架采用 PyTorch1.6 以及 OpenCV3 图像处理库。

在训练阶段,batchsize 批量大小为 4, momentum

表 1 实验环境配置

Table 1 Experimental environment configuration

运行环境	详细配置
系统	Ubuntu 16. 04
处理器型号	Intel (R) Core(TM) i5-9400F CPU@ 2. 90 GHz
显卡	NVIDIA Geforce RTX2060
内存	16. 0 GB
开发语言	Python3. 6
深度学习框架	Pytorch1. 6
图像处理库	OpenCV3

动量为 0. 9, decay 权值衰减为 0. 000 5, learning_rate 学习率为 0. 001, 迭代次数为 300 epoch。

3.3 评价指标

目标检测算法的评价指标有平均精度均值、漏警率以及检测速度。平均精度均值代表所有类别检测的平均准确率, 平均准确率 (average precision, AP) 由查准率 p 和召回率 r 计算得到。漏警率是衡量检测模型的重要指标之一, 尤其是对小尺度目标的检测, 判断模型是否将图像中所有目标全部检测出来。检测速度是评估一个检测器的实时性指标, 通常用每秒处理帧数表示, 帧率 (frames per second, FPS) 越大, 说明检测器的实时性越好, 当 FPS 值大于 30 帧/s 时即可做到实时性检测。各指标公式为

$$p = \frac{TP}{TP + FP} \tag{4}$$

$$r = \frac{TP}{TP + FN} \tag{5}$$

$$AP = \sum_{k=1}^n (r_{k+1} - r_k) p_k \tag{6}$$

$$mAP = \frac{1}{Q} \sum_{q=1}^Q AP(q) \tag{7}$$

$$MA = \frac{FN}{TP + FN} \tag{8}$$

式中, 查准率 p 为正确检测的正样本数占检测结果中正样本总数的比例, 召回率 r 为正确检测的正样本数占实际正样本总数的比例, 真阳性 (TP) 表示正确检测的正样本数、假阳性 (FP) 表示实际为负样本但检测为正样本的数量、假阴性 (FN) 表示实际为正样本但检测为负样本的数量, Q 表示检测的类别数, $AP(q)$ 表示 q 类别的平均准确率, MA 为漏警率, 用来衡量模型漏检程度。

3.4 实验结果与分析

3.4.1 实验结果

为了说明本文方法 AFA-YOLO 相比 YOLOv4 的优劣之处, 对比了两者的模型参数量以及模型的推理时间。从表 2 可以看出, AFA-YOLO 相比 YOLOv4, 模型的参数量大小增加了 1. 25 M, 因此推理时间也增加了 3 ms。

表 2 模型参数与推理时间

Table 2 Model parameters and inference time

模型	参数量/M	推理时间/ms
YOLOv4	62. 71	27
AFA-YOLO(本文)	63. 96	31

注: 加粗字体表示各列最优结果。

本文对训练过程进行了可视化, 其中图 7、图 8 分别对比了 AFA-YOLO 与 YOLOv4 训练时的 mAP 和 recall 的变化曲线。

图 7 中模型的 mAP 值随训练周期逐渐增大, 在训练 300 epoch 后, 其中 AFA-YOLO 的值最终稳定在 82. 36%, YOLOv4 的值最终稳定在 78. 44%, AFA-YOLO 比 YOLOv4 高了 3. 92%。图 8 中模型的召回率同样随着训练周期的增加逐渐升高, AFA-YOLO 最终稳定在 87. 22%, YOLOv4 最终稳定在 84. 57%, AFA-YOLO 比 YOLOv4 高了 2. 65%。

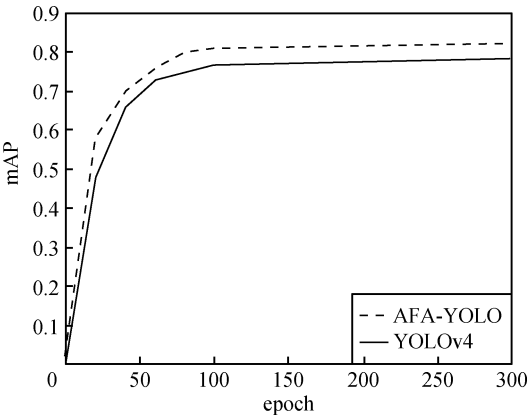


图 7 mAP 曲线

Fig. 7 mAP curves

从实验结果分析得到, 本文 AFA-YOLO 方法虽然在模型参数量上有所增加, 但是能带来更高的检测精度以及更高的召回率。

3.4.2 消融实验

为了充分探究 AFA-YOLO 中每个模块的改进

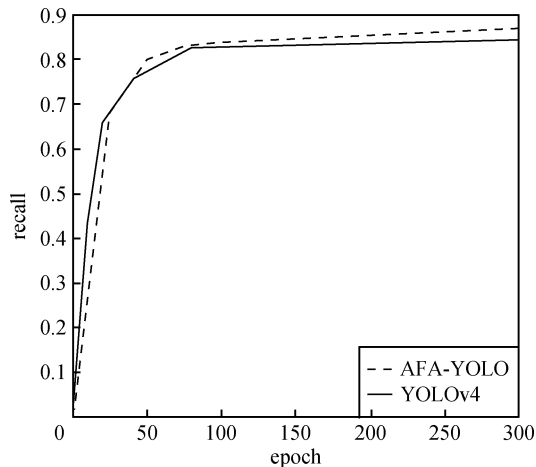


图8 召回率曲线

Fig. 8 Recall curves

对检测模型的影响,进行消融实验,对比 YOLOv4、YOLOv4 + 非对称特征注意力、YOLOv4 + 特征融合和本文 AFA-YOLO 算法。实验结果如表 3 所示。由表 3 可知,在网络中增加了非对称特征注意力模块后,mAP 提升了 1.99%, MA 降低了 2.1%,FPS 降低了 2.85 帧/s;在网络中增加了特征融合后,mAP 提升了 1.53%,MA 降低了 1.98%,FPS 降低

了 1.31 帧/s;在同时加入非对称特征注意力模块和特征融合后,mAP 提升了 3.96%,MA 降低了 2.65%,FPS 降低了 4.06 帧/s。从检测角度看,在网络中添加非对称特征注意力模块和特征融合后虽然检测速度有所降低,但是能有效让检测模型学习到太赫兹图像中目标的特征,从而能有效提升检测模型的检测精度,同时降低模型的漏警率。

表3 消融实验

Table 3 Ablation experiment

算法	mAP/%	MA/%	FPS/(帧/s)
YOLOv4	78.44	15.43	36.32
YOLOv4 + 非对称特征注意力	80.43	13.33	33.47
YOLOv4 + 特征融合	79.97	13.45	35.01
AFA-YOLO(本文)	82.36	12.78	32.26

注:加粗字体表示各列最优结果。

图 9 为 AFA-YOLO 模型对太赫兹图像检测的结果图,图中的数值代表了模型判断为该类别的置信度,从检测效果来看,AFA-YOLO 模型能很好地将太赫兹图中的目标检测出来,并且达到了较高的精度。



图9 太赫兹图像检测

Fig. 9 Terahertz image detection

3.4.3 与其他方法的对比

为了说明 AFA-YOLO 模型的有效性,本文对比了主流的目标检测算法在太赫兹数据集中的检测效果。其中,Faster R-CNN、Mask R-CNN 为两阶段目标检测算法,SSD、YOLOv3、YOLOv4 为单阶段目标检测算法。实验结果如表 4 所示,与其他检测算法

相比,AFA-YOLO 检测算法中的非对称特征注意力模块提升了对太赫兹图像中目标的检测精度,同时能有效降低漏警率,并且特征融合也有利于对太赫兹图像中的目标检测,AFA-YOLO 相比原 YOLOv4 算法虽然检测速度有所降低,但是仍能满足实时性的检测要求。

表 4 不同检测算法的比较
Table 4 Comparison of different detection algorithms

模型	AP(phone)/%	AP(knife)/%	mAP/%	MA/%	FPS/(帧/s)
Faster R-CNN	80.83	81.71	81.27	15.28	12.45
Mask R-CNN	81.06	81.92	81.49	14.76	6.40
SSD	76.87	81.55	79.21	15.89	34.01
YOLOv3	75.09	80.27	77.68	16.13	38.42
YOLOv4	77.03	79.84	78.44	15.43	36.32
AFA-YOLO(本文)	81.15	83.56	82.36	12.78	32.26

注:加粗字体表示各列最优结果。

4 结 论

随着太赫兹技术的发展,基于太赫兹成像的人体安检技术成为安检领域中的研究热点。但太赫兹图像中目标较小、特征有限,并且背景信息易对目标信息产生干扰,这成为太赫兹图像检测的难点问题。本文以 YOLOv4 算法为基础,提出了一种融合非对称特征注意力的目标检测网络 AFA-YOLO。首先,AFA-YOLO 网络在浅层网络中使用非对称卷积来增强网络的特征提取能力,能在目标特征有限的太赫兹图像中提取到有效的目标特征;其次,网络中加入了 CBAM 注意力机制,通过通道注意力和空间注意力使网络模型学习到图像中目标的重要特征,解决背景信息和目标信息干扰的问题,同时让检测模型关注图像中的重点区域;最后通过增加网络中低层到高层的信息传输路径对高层特征增强,将低层高分辨率特征和高层具有丰富语义信息的特征进行融合,充分利用低层特征检测太赫兹图像中的小尺度目标。实验结果表明,本文提出的 AFA-YOLO 方法在太赫兹数据集上的检测精度提升了 3.96%,达到了 82.36%;漏警率降低了 2.65%,为 12.78%;FPS 降低了 4.06 帧/s,为 32.26 帧/s,达到了较好的检测效果。本文与其他检测算法进行了对比,AFA-YOLO 算法在检测精度和漏警率的表现上优于其他

检测算法,但由于在网络模型中增加了非对称卷积,导致模型的参数量增多,因此 AFA-YOLO 在检测速度上慢于 YOLOv3、YOLOv4 以及 SSD 算法,但是仍然能满足实时性的检测要求。

由于本文中只有 knife 和 phone 两类检测对象,图像样本上缺乏多样性,因此在以后的研究工作中,将陆续增加检测对象,将安检中的违禁品纳入其中,并扩充太赫兹图像数据集以提高本模型的实用性和鲁棒性。

参考文献 (References)

Bochkovskiy A, Wang C Y and Liao H Y M. 2020. YOLOv4: optimal speed and accuracy of object detection [EB/OL]. [2021-03-01]. <https://arxiv.org/pdf/2004.10934.pdf>

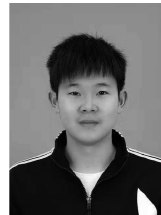
Chen B Q and Deng T. 2019. The vehicle object detection research based on improved SSD algorithm. Journal of Chongqing University of Technology (Natural Science), 33(1): 58-63, 129 (陈冰曲, 邓涛. 2019. 基于改进型 SSD 算法的目标车辆检测研究. 重庆理工大学学报(自然科学), 33(1): 58-63, 129) [DOI: 10.3969/j.issn.1674-8425(z).2019.01.009]

Ding X H, Guo Y, Ding G G and Han J G. 2019. ACNet: strengthening the kernel skeletons for powerful CNN via asymmetric convolution blocks//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South): IEEE: 1911-1920 [DOI: 10.1109/ICCV.2019.00200]

Feng H, An D Y, Tu H, Bu W H, Wang W J, Zhang Y H, Zhang H K,

- Meng X X, Wei W, Gao B X and Wu S. 2020. A passive video-rate terahertz human body imager with real-time calibration for security applications. *Applied Physics B*, 126(8): #143 [DOI: 10.1007/s00340-020-07496-3]
- Gao M F, Yu R C, Li A, Morariu V I and Davis L S. 2018. Dynamic zoom-in network for fast object detection in large images//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA; IEEE: 6926-6935 [DOI: 10.1109/CVPR.2018.00724]
- Guo F, Zhang Y X, Tang J and Li W Q. 2021. YOLOv3-A: a traffic sign detection network based on attention mechanism. *Journal on Communications*, 42(1): 87-99 (郭璠, 张泳祥, 唐璠, 李伟清. 2021. YOLOv3-A: 基于注意力机制的交通标志检测网络. *通信学报*, 42(1): 87-99) [DOI: 10.11959/j.issn.1000-436x.2021031]
- He K M, Gkioxari G, Dollár P and Girshick R. 2017. Mask R-CNN//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy; IEEE: 2980-2988 [DOI: 10.1109/ICCV.2017.322]
- Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C Y and Berg A C. 2016. SSD: single shot MultiBox detector//Proceedings of the 14th European Conference on Computer Vision. Amsterdam, the Netherlands; Springer: 21-37 [DOI: 10.1007/978-3-319-46448-0_2]
- Liu Z G, Li D Y, Ge S S and Tian F. 2020. Small traffic sign detection from large image. *Applied Intelligence*, 50(1): 1-13 [DOI: 10.1007/s10489-019-01511-7]
- Mittleman D M. 2017. Perspective: terahertz science and technology. *Journal of Applied Physics*, 122(23): #230901 [DOI: 10.1063/1.5007683]
- Redmon J, Divvala S, Girshick R and Farhadi A. 2016. You only look once: unified, real-time object detection//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Vegas, USA; IEEE: 779-788 [DOI: 10.1109/CVPR.2016.91]
- Redmon J and Farhadi A. 2017. YOLO9000: better, faster, stronger//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA; IEEE: 6517-6525 [DOI: 10.1109/CVPR.2017.690]
- Redmon J and Farhadi A. 2018. YOLOv3: an incremental improvement [EB/OL]. [2021-03-01]. <https://arxiv.org/pdf/1804.02767.pdf>
- Ren S Q, He K M, Girshick R and Sun J. 2017. Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6): 1137-1149 [DOI: 10.1109/TPAMI.2016.2577031]
- Sommer L W, Schuchert T and Beyerer J. 2017. Fast deep vehicle detection in aerial images//Proceedings of 2017 IEEE Winter Conference on Applications of Computer Vision (WACV). Santa Rosa, USA; IEEE: 311-319 [DOI: 10.1109/WACV.2017.41]
- Wang K X, Liew J H, Zou Y T, Zhou D Q and Feng J S. 2019. PANet: few-shot image semantic segmentation with prototype alignment//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South); IEEE: 9196-9205 [DOI: 10.1109/ICCV.2019.00929]
- Woo S, Park J, Lee J Y and Kweon I S. 2018. CBAM: convolutional block attention module//Proceedings of the 15th European Conference on Computer Vision. Munich, Germany; Springer: 3-19 [DOI: 10.1007/978-3-030-01234-2_1]
- Xiao H, Zhang R Y, Wang H, Zhu F, Zhang C, Dai H N and Zhou Y B. 2018. R-PCNN method to rapidly detect objects on THz images in human body security checks//2018 IEEE SmartWorld, Ubiquitous Intelligence and Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People and Smart City Innovation. Guangzhou, China; IEEE: 1777-1782 [DOI: 10.1109/SmartWorld.2018.00300]
- Yang X, Wu T, Zhang L, Yang D, Wang N N, Song B and Gao X B. 2019. CNN with spatio-temporal information for fast suspicious object detection and recognition in THz security images. *Signal Processing*, 160: 202-214 [DOI: 10.1016/j.sigpro.2019.02.029]
- Zhang J S, Xing W J, Xing M D and Sun G C. 2018. Terahertz image detection with the improved faster region-based convolutional neural network. *Sensors*, 18(7): #2327 [DOI: 10.3390/s18072327]

作者简介



曾文健, 1994年生, 男, 硕士研究生, 主要研究方向为计算机视觉和深度学习。

E-mail: zengwenjian26@163.com



朱艳, 通信作者, 女, 教授, 硕士生导师, 主要研究方向为智能检测和人工智能。

E-mail: zhuyan@kust.edu.cn

沈韬, 男, 教授, 主要研究方向为太赫兹材料无损检测、辅助驾驶和人工智能。E-mail: shentao@kmust.edu.cn

曾凯, 男, 副教授, 主要研究方向为云计算和分布式计算。

E-mail: zengkailink@sina.com

刘英莉, 女, 讲师, 主要研究方向为材料文本挖掘和深度学习。E-mail: ly12002@126.com