基于显著图融合的无人机载热红外图像目标检测方法

赵兴科[1] 李明磊[1] 张 弓[1] 黎 宁[1] 李家松[1]

摘 要 利用无人机载的热红外图像开展行人及车辆检测, 在交通监控、智能安防、防灾应急等领域中, 具有巨大的应用潜力. 热红外图像能够在夜间或者光照条件不理想的情况对场景目标清晰成像, 但也往往存在对比度低、纹理特征弱的缺点. 为此, 本文提出使用热红外图像的显著图来进行图像增强, 作为目标检测器的注意力机制, 并研究仅使用热红外图像和其显著图提高目标检测性能的方法. 此外, 针对无人机内存不足、算力有限的特点, 设计使用轻量化网络 YOLOv3-MobileNetv2 作为目标检测模型. 在实验中, 本文训练了 YOLOv3 网络作为检测的评价基准网络. 使用 BASNet 生成显著图, 通过通道替换和像素级加权融合两种方案将热红外图像与其对应的显著图进行融合增强, 比较了不同方案下 YOLOv3-MobileNetv2 模型的检测性能. 统计结果显示, 行人及车辆的平均精确度 (Average precision, AP) 相对于基准分别提升了 6.7% 和 5.7%, 同时检测速度提升了 60%, 模型大小降低了 58%. 该算法模型为开拓无人机载热红外图像的应用领域提供了可靠的技术支撑.

关键词 显著图, 无人机, 热红外图像, 目标检测, YOLOv3-MobileNetv2

引用格式 赵兴科, 李明磊, 张弓, 黎宁, 李家松. 基于显著图融合的无人机载热红外图像目标检测方法. 自动化学报, 2021, **47**(9): 2120−2131

DOI 10.16383/j.aas.c200021

Object Detection Method Based on Saliency Map Fusion for UAV-borne Thermal Images

ZHAO Xing-Ke[1] LI Ming-Lei[1] ZHANG Gong[1] LI Ning[1] LI Jia-Song[1]

Abstract Using thermal images obtained from unmanned aerial vehicles (UAV) for pedestrian and vehicle detection has great potential in the fields of traffic monitoring, intelligent security, disaster prevention, and emergency response. Thermal images can clearly observe objects at night or under bad lighting conditions, but they also have the disadvantages of low contrast and weak texture features. For these reasons, this paper proposes to use the saliency map of the thermal image for image enhancement as the attention mechanism of the object detector. The technology to improve the performance of object detection using only thermal images and their saliency maps is studied. In addition, considering the computing power of UAV platforms, a lightweight network YOLOv3-MobileNetv2 was designed as the object detection model. In the paper, YOLOv3 network is trained as a detection benchmark; BASNet is used to generate saliency maps. We fuse thermal images with their corresponding saliency maps through channel replacement and pixel-level weighted fusion schemes. In our experiments, the detection performances of YOLOv3-MobileNetv2 model with different schemes are compared. The statistical results show that the average precision (AP) of pedestrians and vehicles are increased by 6.7% and 5.7% respectively, compared with the benchmark. The detection speed is increased by 60%, while the model size is reduced by 58%. This model provides reliable technical support for the application of thermal images with UAV platforms.

Key words Saliency map, unmanned aerial vehicles (UAV), thermal image, object detection, YOLOv3-MobileNetv2

Citation Zhao Xing-Ke, Li Ming-lei, Zhang Gong, Li Ning, Li Jia-Song. Object detection method based on saliency map fusion for UAV-borne thermal images. *Acta Automatica Sinica*, 2021, **47**(9): 2120−2131

收稿日期 2020-01-13 录用日期 2020-04-07
Manuscript received January 13, 2020; accepted April 7, 2020
江苏省自然科学基金 (BK20170781), 国家自然科学基金 (41801342), 中央高校基本科研业务费 (NZ2020008XZA20016), 南京航空航天大学研究生创新基地开放基金项目 (kfjj20190415) 资助
Supported by Natural Science Foundation of Jiangsu Province (BK20170781), National Natural Science Foundation of China (41801342), Fundamental Research Funds for the Central Universities (NZ2020008XZA20016), and Funds from the Postgraduate Creative Base in Nanjing University of Aeronautics and Astronautics (kfjj20190415)
本文责任编委 徐德
Recommended by Associate Editor XU De
1. 南京航空航天大学电子信息工程学院 南京 211106
1. College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 211106

近年来, 以无人机 (Unmanned aerial vehicles, UAV) 为平台的图像获取和处理技术在交通、安防和环保等领域得到快速发展. 彩色图像在理想光照条件下, 对目标检测有很好的效果, 可以较容易地利用深度学习技术找到图像中兴趣目标的具体位置并识别其类别. 然而, 在夜间或缺乏足够光照的情况下, 基于彩色图像的目标检测往往难以达到预期效果, 容易造成漏检或误检. 红外相机非常适合在这种条件下成像, 因为它们能感应到目标物发出的辐射且不受光照条件的限制. 面向行人和车辆检测

的需求, 本文研究一种适用于以无人机为平台获取的热红外图像数据智能处理算法.

相比于彩色图像, 热红外图像有着对比度低、纹理特征弱等缺点. 因此, 在彩色图像和热红外图像之间存在明显的互补. 为了开发这种互补的潜力, 相关学者做了大量的工作来构建融合彩色和热红外图像的数据结构[1]. 但是彩色–热红外图像对并不总是可用的, 因为它们实时同步成像的成本很高, 而且数据处理前还需要图像配准准确, 图像错位还会降低检测器的性能. 这些原因促使本文探索只使用热红外图像来进行目标检测的机制. 为了解决热红外图像中目标检测的难题, 本文提出使用显著图来进行图像增强. 文献 [2] 通过在颜色、方向、运动和深度上与周围环境的不同来定义特定位置的显著性. 在一个场景中寻找显著物体可以理解为一种视觉注意机制, 它突出了给定场景中属于显著物体的像素. 因此, 本文假设显著图和热红外图像结合将帮助提升目标检测模型的性能. 为验证此假设, 首先通过训练一个 YOLOv3 目标检测模型[3] 来建立评价参考基准, 它仅使用制作的热红外图像数据集来检测目标. 然后, 使用 BASNet[4] 生成显著图融合热红外图像, 在此基础上训练目标检测模型. 此外, 由于深度显著性网络需要对显著性对象进行像素级标注, 因此实验中使用行人及车辆的像素级掩膜对制作的热红外图像数据集进行标注, 以方便对显著性目标检测的研究.

针对无人机平台内存和算力的局限性, 本文设计了 YOLOv3-MobileNetv2 网络, 利用轻量化网络 MobileNetv2[5] 替代 YOLOv3 原有的特征提取网络 DarkNet53, 在大量减少网络参数的同时显著提升运行速度. 此外, 使用 Focal loss[6] 改进 YOLOv3 原有的损失函数, 解决正负样本不均衡问题, 使得网络专注于困难样本的计算.

本文的主要贡献如下:

1) 首次将显著图用于提高无人机视角下的热红外图像目标检测性能, 通过设计不同的融合方案, 深入分析了显著图对热红外图像中行人及车辆目标检测的影响.

2) 以轻量化网络 YOLOv3-MobileNetv2 改进原有模型, 在平均精确度、模型尺寸和检测速度方面取得了很好的平衡, 即在计算资源和存储资源有限的情况下实现了最优的精度, 将卷积神经网络 (Convolutional neural network, CNN)[7] 更好地应用于无人机场景中.

1 相关工作

目前, 较少有论文探讨利用无人机结合热成像技术进行目标检测的深度学习方法. 本节回顾了在目标检测、显著性检测以及模型压缩和加速等领域的相关工作.

1.1 目标检测

在过去的 20 年里, 大量的研究工作致力于彩色图像中的行人及车辆检测. 于雪松等[8] 针对人体运动跟踪领域中的自遮挡现象, 提出了一种基于概率模型的行人四肢自遮挡检测算法. 该算法通过马尔科夫模型和椭圆肤色模型将行人四肢自遮挡状态的识别转换为计算自遮挡状态转换概率的过程. 实验表明, 该方法具有较高的准确性. Dollár 等[9] 提出了采用积分通道特征 (Integral channel feature, ICF) 和 Boosting 算法相结合的方法, 提升了车辆检测的效果. 与传统的检测算法相比, 近几年 CNN 在目标检测上取得了重大的突破. 基于 CNN的目标检测算法主要分为两大类: 一阶和二阶目标检测算法. 它们之间的主要区别在于是否存在提取候选区域的级联模块. 二阶目标检测算法中具有代表性的是 R-CNN (Region CNN) 系列检测算法[10−12], 它们通过使用级联模块可以使网络有针对性地检测疑似目标区域的物体, 但由于多了这样的级联模块, 提升精度的同时会使得模型的复杂度升高, 在检测速度上低于一阶检测算法, 不适用于无人机上的实时目标检测. 一阶目标检测算法虽然在检测精度上表现欠佳, 但其检测速度非常快. 其中最具有代表性的是 Redmon 等[13−14] 提出的 YOLO (You only look once) 系列目标检测算法, 该算法将图像划分成 $S \times S$ 的格子, 每个格子负责目标中心在该格子的目标检测, 利用回归思想同时完成检测与识别. 基于此, 本文采用 YOLOv3 算法作为无人机载热红外图像中行人及车辆目标检测识别的基础模型.

近年来, 基于无人机的交通监控系统研究十分活跃. Ruhé 等[15] 使用无人机采集了城市道路车辆信息, 结合地理信息系统 (Geographic information system, GIS) 平台, 对地面交通状况进行预测, 完成流量、车速等信息的提取. 文献 [16] 通过安装在高架平台上的摄像机模拟无人机的视角, 提出一种基于 Haar 特征的人体部位检测器. 为提高无人机道路检测的实时性和鲁棒性, 文献 [17] 提出一种基于改进的图割 (Graph cut) 算法的道路检测方法, 针对航拍图像各个区域具有不同对比度的特点, 将单一的图像全局对比度矩阵替换为局部对比度矩阵.

随着热红外成像技术的广泛应用, 越来越多的研究专注于利用热红外图像实现对行人及车辆的有效检测. 热红外图像与普通的可见光图像相比差异

明显, 热红外图像有着纹理特征不明显、成像对比度低、噪声较多等特点, 这些物理特性使得热红外场景下的目标检测一直都具有挑战性. 针对这一问题, 目前的主流方法是将热红外图像和可见光图像融合, 结合两种图像互补性优势, 获得对于场景全面准确的图像描述. 张秀伟等[18] 提出了一种基于 Co-motion 的可见光与热红外图像序列自动融合方法, 引入 Co-motion 运动统计特征来解决异源图像序列融合问题, 从而避开了异源图像相似图像特征提取和精确运动检测的难题. Li 等在文献 [19] 中设计了光照感知的神经网络, 它自适应地融合了彩色和热红外子网络, 并根据光照条件采用加权方案融合结果. 在文献 [20] 中, 作者引入了区域重建网络, 利用 CNN 对可见光与热红外数据之间的关系进行建模, 然后将这些特征输入到多尺度检测网络中进行鲁棒的目标检测.

然而, 可见光对光照变化以及其他环境影响较为敏感, 尤其是在夜间, 采用可见光获取图像的方法将完全不可用, 故可见光与热红外图像融合也无从谈起. 此外, 同时获取同一场景的可见光与热红外图像需要两种不同的传感器, 图像采集过程较为复杂, 对设备要求较高. 在此背景下, 本文研究仅使用热红外图像进行目标检测, 在克服夜间低能见度并实现全天候检测的同时, 简化检测过程, 通过算法的优化提高检测效果.

1.2 显著性检测

显著性目标检测的目的是突出图像中最明显的目标区域, 它可以引导机器视觉系统将有限的资源分配给少数几个显著区域, 为后续的视觉处理提供极大的便利. 从理论研究的层面来说, 可以把图像的显著性检测研究分为两大方向, 即数据驱动的显著性检测和目标驱动的显著性检测. 数据驱动的图像显著性区域检测算法主要关注由图像底层特征本身所引起的视觉刺激, 这类算法由内部数据驱动, 与目标任务无关. 与此相反, 目标驱动的显著性检测算法主要关注与任务相关的图像内容, 显著性检测的结果受到检测任务的决定性支配.

Itti 等[21] 最早提出认知视觉注意模型, 该模型提取场景中的特征显著图并采用线性合并的方式整合为总显著图, 以赢者通吃 (Winner takes all) 和返回抑制相结合的方式来引导视觉注意焦点的选择和转移. Hou 等[22] 提出基于频域的谱残差法, 对图像进行二维傅里叶变换后得到频域的相位谱和幅度谱, 作者认为频谱域上的统计奇异对应图像的异常区域, 因此该区域的物体显著性高. 利用深度学习技术进行显著性检测是近年来的研究趋势. He 等[23] 提出了一种新的超像素方法, 称为 Super-CNN, 可以有效地学习显著性的内部表示. 与传统的卷积网络相比, 该网络能够学习分层对比度特征, 通过多尺度网络结构检测显著性区域. Hou 等[24] 提出了一种快速的显著性检测方法, 在整体嵌套边缘检测的基础上, 增加了一种高层信息指导低层信息的跳层连接结构, 从而构建了一种简单、有效、快速的端对端的显著性物体检测网络. 张芳等[25] 设计实现了一种全卷积神经网络与低秩稀疏分解相结合的显著性检测方法, 将图像分解为代表背景的低秩矩阵和对应显著区域的稀疏噪声, 结合利用全卷积神经网络学习得到的高层语义先验知识, 检测图像中的显著区域. 本文使用了目前最先进的网络 BASNet 生成热红外图像的显著图, 并在第 3.3.1 节中进行结果评估.

1.3 模型压缩和加速

虽然现在 CNN 的特征提取能力随着网络层数的加深正在不断地提升, 但在实际工程中还需要考虑模型尺寸和模型预测速度. 深度 CNN 结构包含几十层甚至上百层的网络, 有着大量的权重参数, 如何调整其结构以在准确度、尺寸和速度之间实现最佳平衡已经成为一个很受关注的研究领域.

为了解决这个问题, 众多轻量化网络结构纷纷被提出. SqueezeNet[26] 提出了一种称作 Fire 的模块, 它分为两个部分: 一个由 1×1 卷积核构成的压缩层以及一个由 1×1 和 3×3 卷积核组成的扩张层. 通过使用这种模块, SqueezeNet 能在保持模型精度不损失的情况下达到 50 倍压缩率. ShuffleNet[27] 充分利用了分组卷积和通道混洗进一步提高模型效率, 在减少计算量的同时解决了组间信息流通问题. 而 Google 提出的 MobileNet 系列模型[28−29] 是专门针对移动和嵌入式设备开发的轻量级 CNN 结构. MobileNetv1 采用一种深度可分离卷积的高效卷积方法来提升运算速度. 深度可分离卷积将一个标准卷积分解成两步来实现, 第 1 步是深度卷积, 即对每个输入通道用单个卷积核进行卷积运算; 第 2 步是一个 1×1 卷积, 即逐点卷积, 负责通过计算输入通道间的线性组合来构建新的特征. 通过深度卷积和逐点卷积两个步骤实现卷积层, 其参数仅约为普通卷积的 1/9. 在此基础上, MobileNetv2 加入了反向残差和线性瓶颈模块构成了更高效的基本模块. 传统的残差结构特征通道维度先缩减后扩展, 即先用一个 1×1 卷积来降低通道维度, 目的是减小计算量. 而 MobileNetv2 使用深度卷积替换了 3×3 标准卷积, 虽然极大地减少了计算量和参数量, 但提

取的特征也会相对减少, 如果再进行压缩, 能提取的特征将更少, 影响模型的准确度. 因此采用反向残差结构, 先对通道进行扩展, 深度卷积能提取更多特征, 保证模型准确度. 线性瓶颈就是去掉了低维度输出层后面的非线性激活层, 目的也是为了在兼顾参数量和计算复杂度的同时实现较高的准确度. MobileNetv3 采用了新的非线性激活层 h-swish, 使用互补的网络搜索方法搜索得到轻量级的网络. 相较于之前的版本, MobileNetv3 虽然实现了性能提升, 但设计及训练过程复杂. 本文使用 MobileNetv2 改进原有的 YOLOv3 模型, 满足无人机场景下的应用需要.

2　算法

2.1　使用 YOLOv3 检测结果作为评价基准

本文采用的目标检测基础模型是 YOLOv3 模型, 首先用它来仅处理热红外图像, 进行行人及车辆的检测任务, 将这类检测的结果作为后续模型改进的评价基准. YOLOv3 将图片划分为 $S \times S$ 的网格, 各网格只负责检测中心落在该网格的目标, 每个网格需要预测三个尺度的边界框和类别信息, 一次性预测所有区域所含目标的边界框、目标置信度以及类别概率. 与之前的 YOLO 算法相比, YOLOv3 采用了精度更高的 DarkNet53 作为图像特征提取网络, 设计了目标多尺度检测结构, 对无人机视角下小目标的检测具有很好的效果. 本文在采集制作的热红外图像数据集上训练 YOLOv3 模型作为行人及车辆检测基准, 结果如表 1 所示.

2.2　使用显著图融合热红外图像改进行人及车辆检测

本文提出从热红外图像中提取显著图来使行人及车辆检测器获得更丰富的像素间关系. 在无人机航拍图像中背景复杂的情况下, 显著图的使用会使行人及车辆在图像中与周围环境的差别更大. 然而, 显著图抛弃了热红外图像中所有可用的纹理信息. 为了解决这个问题, 本文将热红外图像与其对应的显著图融合.

2.2.1　深度显著性网络 BASNet 提取显著图

深度 CNN 结构在显著目标检测 (即显著图生成) 上已有应用, 并取得了不错的性能. 但是之前的绝大多数工作都关注在兴趣目标区域的准确率上, 而不是边界的质量上. 无人机视角下目标的形态与地面视角成像时的差异较大, 因而热红外图像具有边界几何特性弱的特点. 本文提出使用显著图进行的目标检测增强, 其边界将会对图像增强效果产生较大影响. 因此, 使用更关注边界质量的 BASNet 网络[4] 作为生成显著目标的基础网络.

如图 1 所示, BASNet 是一个新的预测–优化网络. 该架构由一个类似于 U-Net 的密集监督的编译码网络和一个额外的残差优化模块组成, 分别负责显著图预测和显著图优化. 在编码–解码网络中, 前面的编码网络对图像进行特征提取, 使用池化方法得到了分辨率逐步变小的高层语义特征; 后面的解码网络部分则负责将高层语义信息逐步还原放大, 从而逐步获得大分辨率的特征图, 最终输出和原图一样大小的粗糙的显著图. 在编码网络和解码网络之间有直连结构, 将相同分辨率的特征图相加, 从

表 1　采用不同方法所得到结果的比较

Table 1　Comparison of results from different techniques

类别		指标	行人		车辆		模型大小 (MB)
			AP	FPS	AP	FPS	
使用的数据和方法	热红外图像	YOLOv3	0.836	20	0.873	20	235
		YOLOv3-MobileNetv2	0.792	32	0.826	32	97
	显著图	YOLOv3	0.771	21	0.820	21	235
		YOLOv3-MobileNetv2	0.719	34	0.761	34	97
	替换 R 通道融合	YOLOv3	0.927	20	0.932	20	235
		YOLOv3-MobileNetv2	0.880	32	0.889	32	97
	替换 G 通道融合	YOLOv3	0.938	18	0.956	18	235
		YOLOv3-MobileNetv2	0.881	30	0.899	30	97
	替换 B 通道融合	YOLOv3	0.905	19	0.972	19	235
		YOLOv3-MobileNetv2	0.857	31	0.925	31	97
	像素级加权融合	YOLOv3	0.944	20	0.978	20	235
		YOLOv3-MobileNetv2	0.903	32	0.930	32	97

而让最终的输出的特征图能够同时兼顾低级和高级的特征. 此外, 为了优化粗糙的显著图中存在的区域和边界缺陷, BASNet 设计了一个新的残差优化模块,该模块通过学习粗糙的显著图和真实值之间的残差来优化预测的输出. 这个优化模块的输出就是最终输出的显著图, 如图 1 中放大的视图展示了粗糙的和优化后的显著图.

与其他的显著图预测网络不同, BASNet 在每层损失函数的设计上, 使用了交叉熵、结构相似性损失、交并比 (Intersection over union, IoU) 损失这三种的混合损失 L , 使网络更关注于边界质量, 而不是像以前那样只关注区域精度. 损失函数的表达式为

$$L = L_{\text{bce}} + L_{\text{ssim}} + L_{\text{iou}} \tag{1}$$

其中, L_{bce} 代表交叉熵损失, 对应着像素级的监督. L_{ssim} 是结构相似性损失, 对应着区域级的监督. L_{iou} 是交并比损失, 对应着显著图级的监督. 各项损失的具体数学形式参见文献 [4].

2.2.2 使用显著图融合热红外图像

在获得显著图后 (如图 2(a) 所示), 本文设计了两类图像融合方案, 即: 1) 用显著图分别替换了热红外图像的红色 (R)、绿色 (G)、蓝色 (B) 三个通道中的一个通道, 如图 2(b)~2(d) 所示, 每个替换方式

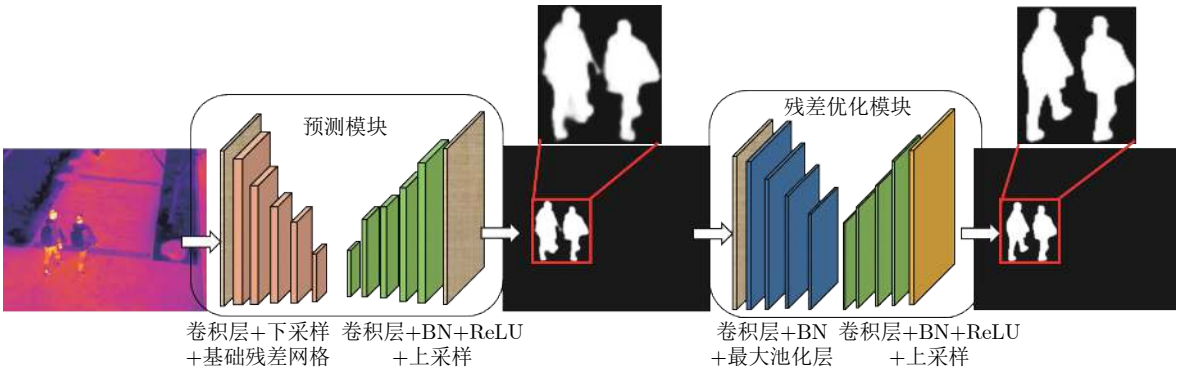


图 1 BASNet 网络结构

Fig. 1 Architecture of boundary-aware salient object detection network: BASNet

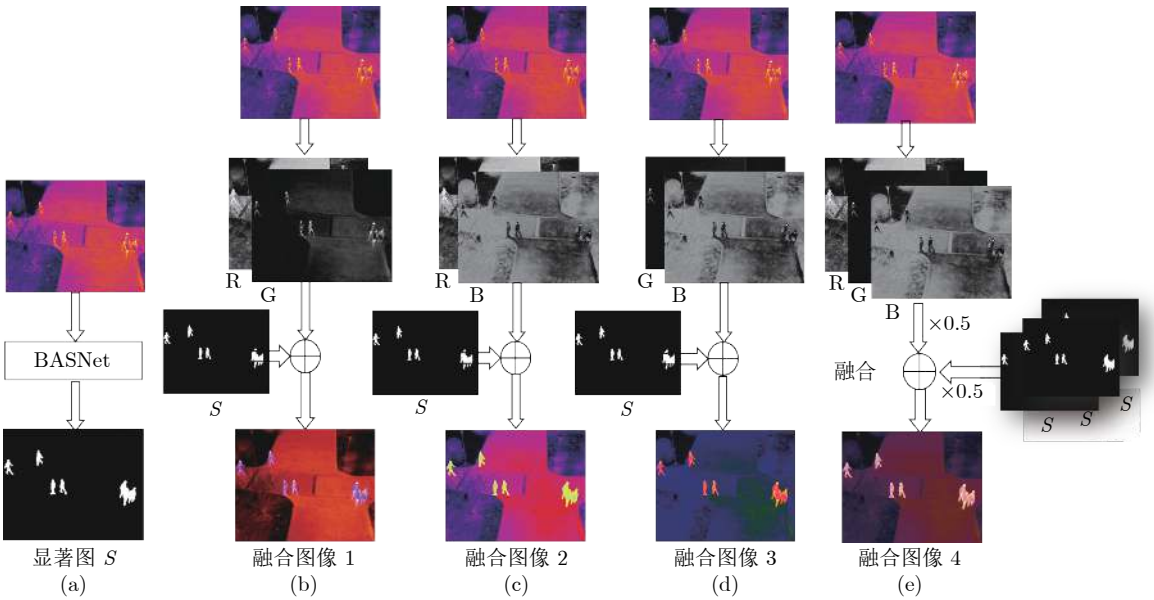


图 2 使用显著图增强热红外图像的流程 ((a) 使用 BASNet 网络生成热红外图像的显著图; (b)~(d) 分别是用显著图替换热红外图像三通道中的一个通道; (e) 将显著图与热红外图像在三个通道分别进行像素级别上直接融合)

Fig. 2 The fusion of the thermal image and its saliency map ((a) Using BASNet to generate the saliency map of a thermal image; (b) to (d) replacing each of three channels of the thermal image with the saliency map; (e) Fusion of the thermal image and the duplicated saliency maps at pixel-level)

都对应了一个新的融合图像, 这些融合图像的目标检测性能也存在差异, 实验中将对其进行分析; 2) 将显著图与热红外图像的三个通道亮度值分别在像素级别上直接按各自 0.5 的权重比例融合, 如图 2(e) 所示. 显著图与热红外图像的融合在突出图像中行人及车辆目标的同时保留了图像中的纹理信息, 如图 3 所示. 通道融合之后, 为比较显著图对行人及车辆检测的影响, 继续利用热红外图像中提取的显著图以及使用上述两类方法生成的与显著图融合的热红外图像分别训练没有改进的 YOLOv3 网络模型.

2.3　使用 MobileNetv2 改进 YOLOv3 模型

基于深度学习的目标检测算法计算开销和模型参数体量巨大, 难以部署在计算和存储资源有限的无人机平台上. 针对这一问题, 本文以轻量化目标分类网络 MobileNetv2 为基础, 提出了一种新型超轻量化目标检测网络模型. 将 MobileNetv2 的平均池化层和最后一个卷积层去掉, 使其替换 YOLOv3 网络中原有的 DarkNet53 网络. 同时, YOLOv3-MobileNetv2 沿用了 DarkNet53 的连接规则, 即分别将 MobileNetv2 中分辨率为输入图像的 8 倍和 16 倍下采样的特征层中的最后一层作为细粒度特征与检测网络中上采样之后的高级语义特征融合, 增强网络的目标检测能力[30], 这对无人机航拍图像中的小目标检测具有重要意义. 深度可分离卷积可以使网络在保证准确度的同时获得较小的模型尺寸和较低的计算复杂度. 因此, 采用深度可分离卷积替换 YOLOv3-MobileNetv2 检测网络中占据大量参数的 3×3 卷积操作. 最后根据需要识别的目标类别和先验框的尺寸对模型进行了修改, 完成了轻量化模型 YOLOv3-MobileNetv2 的网络设计. 最终的 YOLOv3-MobileNetv2 结构中只有 3 个标准卷积, 采用了 21 个深度可分离卷积模块, 结构如图 4 所示.

此外, YOLOv3 检测过程会产生两个问题. 1) 极度不平衡的正负样本比例, 其候选样本生产方式会使正负样本相差巨大, 而且大部分的负样本都是易分样本; 2) 梯度被易分样本主导, 虽然这些样本的损失值很低, 但是数量众多, 对于损失依旧有很大贡献, 从而导致收敛效果不够好. 因此, 本文在损失函数中使用 Focal loss 改进原有的交叉熵, 如式 (2) 所示.

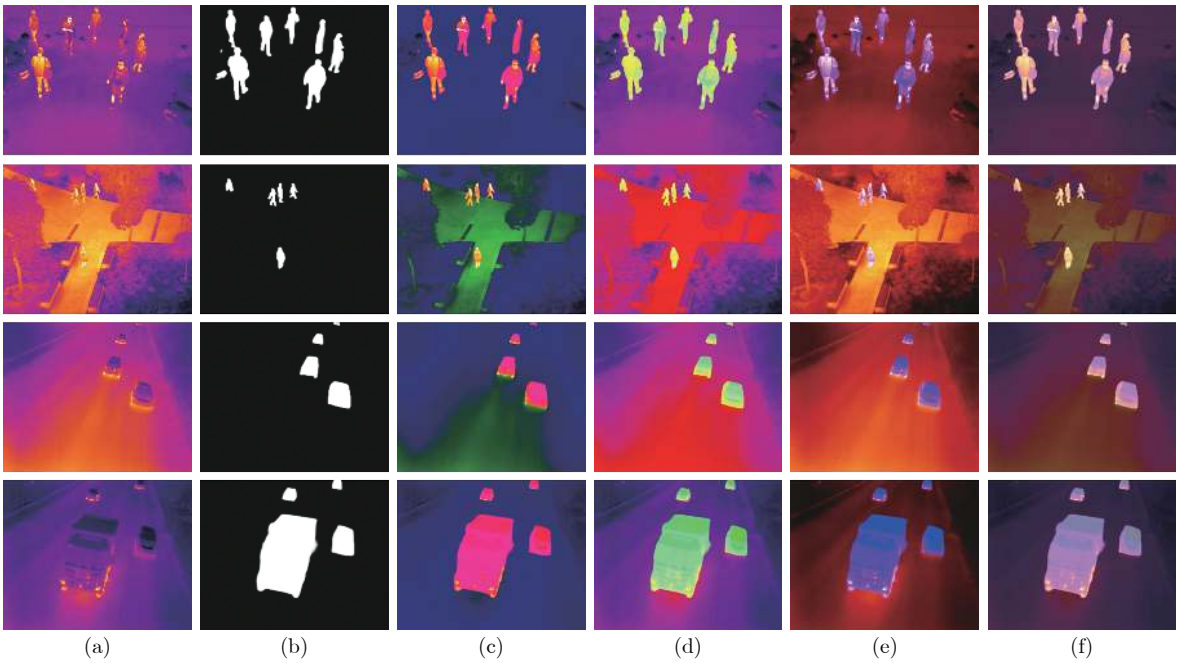


图 3　测试集中使用显著图增强的热红外行人 (第 1 行和第 2 行) 及车辆 (第 3 行和第 4 行) 图像 ((a) 原始热红外图像; (b) 显著图; (c) 使用显著图替换热红外图像 R 通道; (d) 使用显著图替换热红外图像 G 通道; (e) 使用显著图替换热红外图像 B 通道; (f) 热红外图像与显著图的三个通道分别进行像素级直接融合)

Fig. 3　Thermal images and generated saliency maps for pedestrian (top 2 rows) and vehicle (bottom 2 rows) images from the test set ((a) Original thermal images; (b) Saliency maps; (c) Replacing red channel of thermal images with saliency maps; (d) Replacing green channel of thermal images with saliency maps; (e) Replacing blue channel of thermal images with saliency maps; (f) Direct fusion of saliency maps and thermal images at pixel-level)

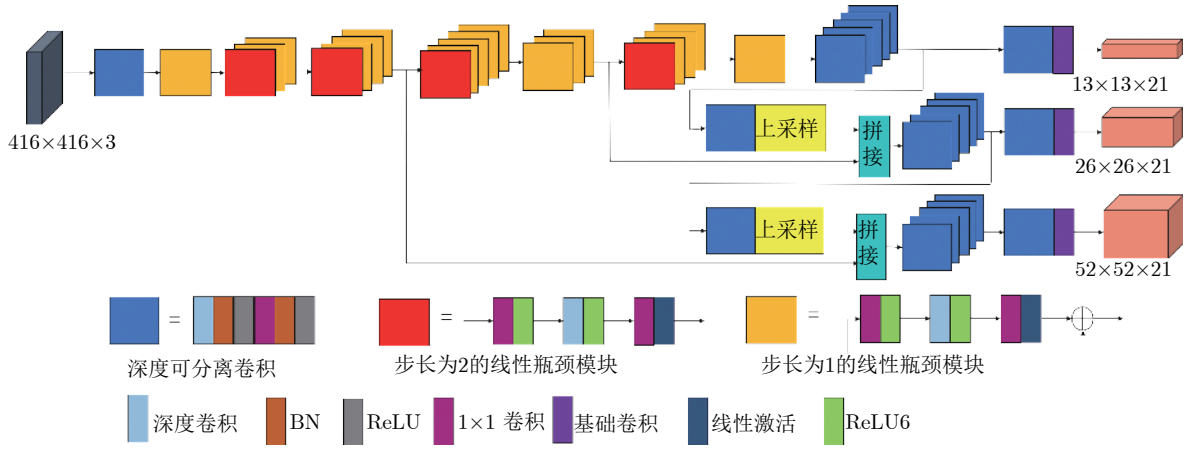


图 4 YOLOv3-MobileNetv2 网络结构图

Fig. 4 Architecture of YOLOv3-MobileNetv2

$$Fl\left(p_t\right)=-\alpha(1-p_t)^{\beta}\log\left(p_t\right) \tag{2}$$

其中, p_t 为预测的概率值, α 与 β 均为可以调节的超参数, 在本文中, α 取 0.5, β 取 2.

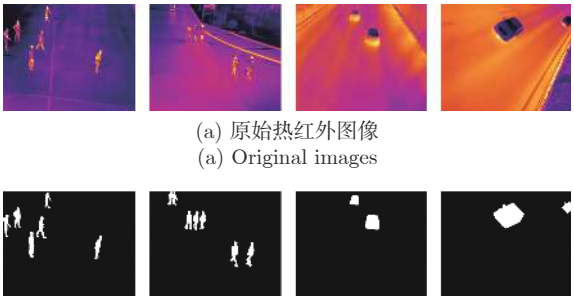
3 实验

3.1 数据集及评估方法

通常为了训练一个深度神经网络模型, 需要大量的数据样本. 然而, 目前并没有公开可用的针对无人机视角下行人及车辆的热红外数据集. 除此之外, 对图像进行显著性目标检测也需要对显著性对象进行像素级标注, 需要较大的工作量. 因此, 本文在实验环节采用现有设备制作了无人机载热红外图像行人及车辆数据集, 方便相关技术的进一步研究.

实验使用大疆无人机 DJIM600 PRO 搭载 FLIR 热红外相机 Vue Pro 采集行人及车辆热红外图像数据, 无人机飞行高度 20 ~ 40 m, 图像分辨率 640×512 像素. 值得注意的是, 热红外相机接收到的原始图像只有亮度, 为单通道灰度图像. 为方便行人及车辆目标检测的研究, 将接收到的热红外图像经过温度映射后转换为 RGB 格式三通道伪彩色图像, 单通道灰度图像中像素值 0 映射为蓝色, 像素值 255 映射为红色, 中间平滑渐变, 即使用颜色的冷暖色调来显示低温和高温区域.

无人机携带热红外成像仪分别在白天和夜间采集数据, 共获得热红外图像 2434 幅, 其中包含 3555 个行人实例和 3189 个车辆实例. 使用图像标注工具 Labelme 手动标注这些图像, 用以生成训练 BASNet 所需要的显著性目标边缘像素级标注. 同时, 使用 LabelImg 标注图像, 用以生成训练 YOLOv3 及 YOLOv3-MobileNetv2 所需要的目标边框标注. 此外, 创建了 541 幅带有类似标注的图像, 用于验证本文的深度显著性检测网络和改进的目标检测网络, 其中包含 1213 个行人实例和 667 个车辆实例. 图 5 显示了 4 组所制作数据集的示例图像和标注. 训练集和测试集每帧图像中行人及车辆的分布如图 6 所示. 第 3.3.1 节的测试结果表明, 该数据集在行人及车辆显著性目标检测中可以取得相当好的效果.



(a) 原始热红外图像

(a) Original images

(b) 像素级标注

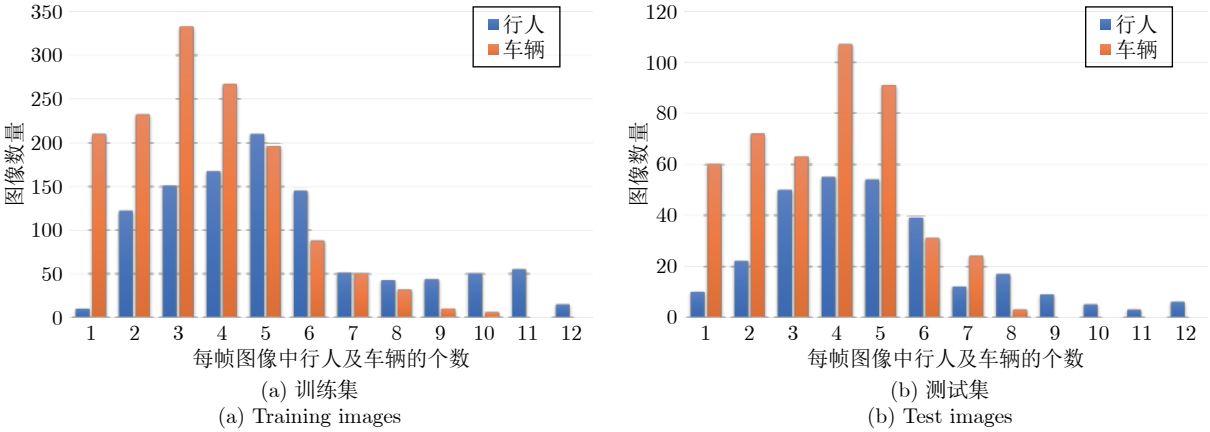
(b) Pixel level annotations

图 5 行人及车辆热红外数据集标注示例

Fig. 5 Sample annotations from pedestrian and vehicle thermal dataset

为了对行人及车辆检测结果进行评估, 本文使用 AP 值和每秒帧率 (Frame per second, FPS) 分别作为精度和速度的评价指标. 此外, 使用 F 度量值 (F-measure, F_β) 和平均绝对误差 (Mean absolute error, MAE) 来评估模型的显著性检测结果. 其中, F_β 是精确率 (Precision) 和召回率 (Recall) 在非负权重 β 下的加权调和平均值, F_β 越高则模型越好, 具体计算为

$$F_\beta=\frac{\left(1+\beta^2\right)Precision\times Recall}{\beta^2 Precision+Recall} \tag{3}$$



(a) 训练集
(a) Training images

(b) 测试集
(b) Test images

图 6　训练集和测试集中行人及车辆的分布

Fig. 6　Distribution of pedestrian and vehicle in training images and test images

其中, β^2 一般取值为 0.3.

MAE 用来直接计算模型输出的显著图与其对应的真实值之间的像素误差

$$MAE = \frac{1}{W \times H} \sum_{x=1}^{W} \sum_{y=1}^{H} \left| \bar{S}(x,y) - \bar{G}(x,y) \right| \qquad (4)$$

其中, W 和 H 分别为图像的宽和高, $\bar{S}(x,y)$ 和 $\bar{G}(x,y)$ 分别为输出的显著图和其对应的真实值二值化后的像素值.

3.2　网络模型实现细节

3.2.1　使用 YOLOv3 及 YOLOv3-MobileNetv2 分别进行行人及车辆检测

本文在有 8 GB 内存的 NVIDIA 1080ti GPU 上对 YOLOv3 及 YOLOv3-MobileNetv2 模型进行训练, 使用双线性插值将图像尺寸由 640×512 像素调整为 416×416 像素后输入网络模型. 在 Microsoft COCO数据集上预先训练 YOLOv3 及 YOLOv3-MobileNetv2 骨干网络, 并在 3.1 节中描述的热红外图像数据集上进行 100 个回合 (Epoch) 的微调. 此外, 设置批量大小 (Batch size) 为 8, 初始学习率为 0.001, 使用 Adam 优化器自适应调整学习率. IoU 阈值设置为 0.5, 经过非极大值抑制 (Non-maximum suppression, NMS) 操作后输出最终的预测结果.

3.2.2　深度显著性网络 BASNet

本文使用像素级标注的热红外图像对 BASNet 进行训练, 并保持了与原始论文中相同的网络架构. 在训练阶段, 首先将训练集中每幅图像尺寸调整为 256×256 像素, 通过随机翻转和裁剪来对训练集图像做增强操作. 使用 ResNet-34 网络的权值来初始化特征提取网络的参数, 解码网络从 0 开始训练, 学习率为 0.01. 在不使用验证集, 批量大小为 8 的情况下, 经过 6 万次迭代损失函数收敛, 整个训练过程耗时约 7 小时. 在测试阶段, 同样将输入图像尺寸调整为 256×256 像素, 并将其输入到网络中得到预测的显著图. 然后, 将降采样的显著图重新调整为原始输入图像的尺寸. 这两个调整过程都使用双线性插值.

3.3　结果和分析

3.3.1　深度显著性网络 BASNet 在热红外图像数据集上的检测效果

为了给接下来的图像显著图研究提供一个有效的支撑, 首先评估了 BASNet 在标注的无人机航拍热红外行人及车辆显著性数据集的测试集上的性能. 评估结果表明, F_β 为 0.767, MAE 为 0.008, 从中可以看出该模型的检测效果是十分优异的. BASNet 对热红外图像进行处理, 检测提取出其中的行人及车辆目标, 目标像素值为 255, 背景像素值为 0, 以二值图像的形式输出. 使用该模型生成的掩膜可以在图 3(b) 中看到. 由于热红外图像以温度作为成像基础, 因此图像中部分与目标温度相近的物体不可避免地对检测结果造成影响, 形成显著目标的误判, 例如图 3(b) 第 1 幅图像最左边区域中的高温物体被误检为显著目标. 但此类现象在实际应用中产生频率较低, 且随着训练集样本的丰富由此对目标检测造成的影响可以忽略不计.

3.3.2　目标检测的定量分析

在使用 YOLOv3 和 YOLOv3-MobileNetv2 分别训练热红外图像、显著图和通过不同的融合技术增强的热红外图像之后, 接下来对检测结果进行性

能评估. 经过比较可以发现, 显著图对行人及车辆检测精度的提升有明显效果, 而 YOLOv3-MobileNetv2 网络则显著提高了模型的检测速度. 表 1 中总结了本文所提算法的检测性能, 下一节将对一些重要的结果做进一步的分析.

3.3.2.1 使用 YOLOv3 作为检测网络

1) 只使用热红外图像. 首先使用 YOLOv3 训练热红外图像作为检测基准, 模型大小为 235 MB. 实验结果表明行人及车辆的 AP 值分别为 83.6% 和 87.3%, 检测帧率为 20 帧/s. 从实验结果中不难看出, 受限于无人机场景下内存不足、算力有限的特点, 现有的 YOLOv3 算法无论是模型尺寸还是检测速度、识别准确度都无法满足实际应用的需要, 由此证实了采用显著图融合增强热红外图像以及使用 MobileNetv2 改进 YOLOv3 算法的必要性. 此外, 由于场景中温度差别较大, 热红外图像中车辆的成像效果比行人更好, 这也是造成行人及车辆 AP 值差异的深层次原因.

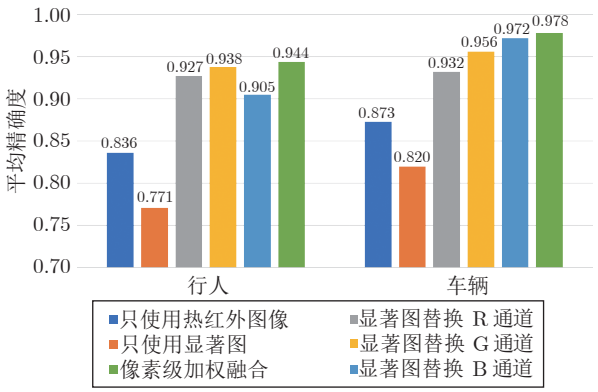
2) 只使用显著图. 只使用显著图进行行人及车辆检测, 其 AP 值分别为 77.1% 和 82.0%, 相较于基准下降了 6.5% 和 5.3%, 检测帧率为 21 帧/s. 此外, 实验中发现该方法导致了大量的误检和漏检, 影响了精度. 这表明, 虽然显著图具有一定的应用潜力, 但是它作为二值图像, 只是突出了图像中的显著物体, 并不具备任何纹理特征, 一方面使得深度卷积神经网络无法在图像中获取丰富的信息进行目标位置、大小和类别的判断; 另一方面容易造成形状大小相似的物体被错误检测为同一类目标. 尤其是当数据样本不足时, 提取显著图的误差与目标检测的误差累加, 将造成检测精度的大幅下降. 因此仅仅使用显著图来完成目标检测的任务往往达不到理想的精度要求.

3) 使用显著图增强的热红外图像. 本文设计了两种方案来完成热红外图像与显著图的融合增强. 实验结果表明, 方案 1 中使用显著图替换热红外图像的 R 通道后, 行人及车辆的 AP 值为 92.7% 和 93.2%, 分别比基准提高了 9.1% 和 5.9%, 检测帧率为 20 帧/s. 使用显著图替换热红外图像的 G 通道后, 行人及车辆的 AP 值为 93.8% 和 95.6%, 分别比基准提高了 10.2% 和 8.3%, 检测帧率为 18 帧/s. 使用显著图替换热红外图像的 B 通道后, 行人及车辆的 AP 值为 90.5% 和 97.2%, 分别比基准提高了 6.9% 和 9.9%, 检测帧率为 19 帧/s. 方案 2 中将热红外图像与显著图进行 0.5 比例的像素级加权融合后, 行人及车辆的 AP 值为 94.4% 和 97.8%, 分别比基准提高了 10.8% 和 10.5%, 检测帧率为 20 帧/s. 更直观的比较如图 7(a) 所示.

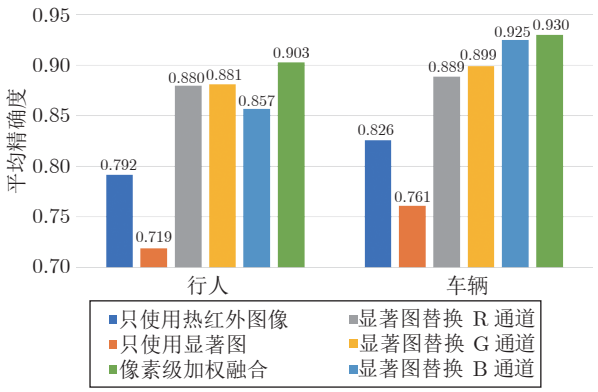
这些提升可以用图 8 中的可视化示例来解释, 图 8 中, 使用显著图增强后的热红外图像突出了场景中的行人和车辆, 帮助检测器在低对比度的情况下识别目标. 此外, 不难看出, 使用像素级加权融合的方案要优于图像通道替换方案, 这是因为图像通道替换过程中损失了部分有用的像素信息, 破坏了图像原来的结构. 而像素级加权融合是在保留热红外图像和显著图通道结构的基础上, 对行人和车辆区域进行增强, 对非目标的背景区域进行抑制. 同时, 由于输入图像尺寸没有发生变化, 因此两种方案中检测帧率都与基准大致相同.

3.3.2.2 使用 YOLOv3-MobileNetv2 作为检测网络

实验结果表明, 改用 YOLOv3-MobileNetv2 网络后, 在保持平均精确度不发生明显变化的同时大幅提高了模型的检测速度, 而网络模型的大小得到



(a) 使用 YOLOv3 检测
(a) Detection using YOLOv3



(b) 使用 YOLOv3-MobileNetv2 检测
(b) Detection using YOLOv3-MobileNetv2

图 7 不同检测模型的平均精确度比较

Fig. 7 Comparison of average precisions of different detection models

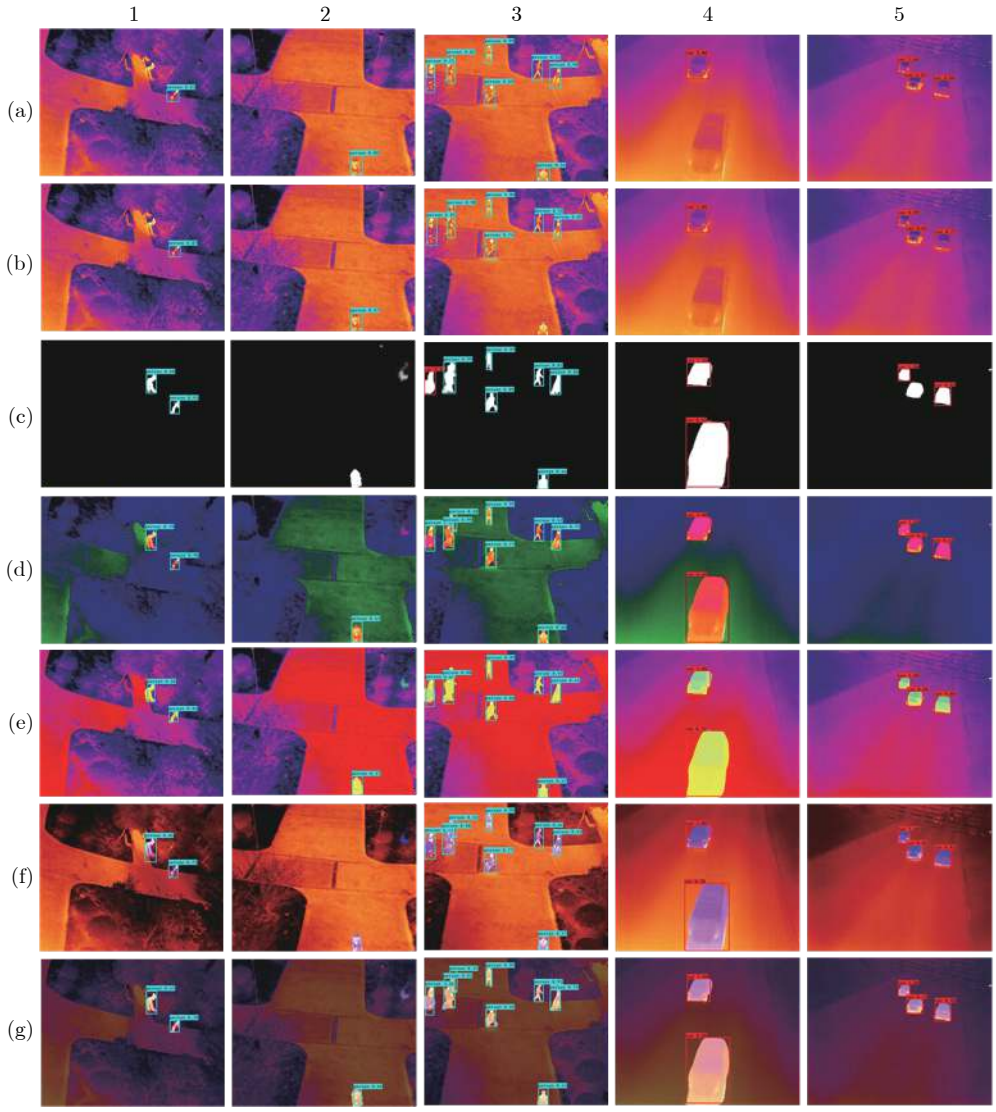


图 8　行人及车辆检测示例 (1 ~ 3 列为行人, 4、5 列为车辆) ((a) 原始热红外图像+YOLOv3; (b) 原始热红外图像+YOLOv3-MobileNetv2; (c) 显著图+YOLOv3-MobileNetv2; (d) ~ (f) 分别是使用显著图替换热红外图像 R、G、B 通道+YOLOv3-MobileNetv2; (g) 热红外图像与显著图进行像素级直接融合+YOLOv3-MobileNetv2)

Fig. 8　Sample results from pedestrian detection on images 1 ~ 3 and vehicle detection on images 4 and 5 from methods: ((a) Thermal images + YOLOv3; (b) Thermal images + YOLOv3-MobileNetv2; (c) Saliency maps + YOLOv3 -MobileNetv2; (d) ~ (f) represent replacing one of R, G, and B channel of thermal images by saliency maps + YOLOv3 -MobileNetv2; (g) Direct fusion of saliency maps and thermal images at pixel-level + YOLOv3-MobileNetv2)

了极大的精简. 这是因为 YOLOv3-MobileNetv2 沿用了 YOLOv3 多尺度预测、预测边框偏移量而不是直接预测边框的大小和位置、采用多重标签分类等优点, 以轻量化卷积替代原先数量多、参数量大的标准卷积, 从而取得速度、精度、模型大小的平衡. 在综合性能表现最好的像素级加权融合方案上, 行人及车辆的 AP 值为 90.3% 和 93.0%, 相较于未改进的 YOLOv3 网络训练热红外图像的检测基准分别提升了 6.7% 和 5.7%, 而检测帧率则由 20 帧/s 变为 32 帧/s, 提升了 60%, 模型大小由 235 MB 变为 97 MB, 减小了 58%. 其他各项方案的 AP 值及帧率、模型大小变化如表 1 和图 7(b) 所示.

3.3.3　融合显著图的目标检测分析

图 8 给出了使用上述技术在不同设置下对 5 组图像的检测结果. 通过图 8 (a) 与图 8 (b) 的比较, 可以看出YOLOv3-MobileNetv2 在检测精度上与YOLOv3相近. 由图 8 可以看到, 用于图像增强的显著图 8 (c)-1 和图 8 (c)-4 有助于捕获原始热红外图像图 8 (b)-1 和图 8 (b)-4 中出现的行人及车辆的漏检, 这显示出显著图应用于复杂场景中目标检测

的潜力. 在融合了热红外图像之后, 图 8 (c)-2 中漏检的行人在图 8 (d)-2、图 8 (e)-2 和图 8 (g)-2 中被成功捕获. 在这里需要注意的是, 使用显著图替换热红外图像 B 通道的图 8 (f)-2 依然没有检测到行人, 表明此方法在行人检测性能上略低于其他方法, 这与在表 1 中得到的结果是一致的. 使用显著图分别替换热红外图像三个通道后显著物体在图像中的对比度及显著程度是不同的, 这也是造成检测结果出现些许差异的深层次原因. 利用图 8 (b)-5 中的热红外图像信息, 图 8 (c)-5 中漏检的车辆在图 8 (d)-5~8 (g)-5 中均被成功捕获. 以上结果显示出热红外图像与显著图之间的互补性, 从而证明了本文关于融合显著图提高热红外图像目标检测精度的假设.

在图 8 (c)-3 中, 可以看到显著图中最左边的行人被误检为车辆, 而图 8 (b)-3 中的热红外图像中很少发生这种误检. 因此, 将热红外图像与显著图结合后, 检测器能够排除这种错误. 在图 8 (b)-3 中, 由于与周围环境温度相近, 热红外图像中没有检测出最下方的行人, 而图 8 (c)-3 中的显著图很好地突出了这个区域, 帮助检测器将其成功捕获 (参见图 8 (d)-3~8 (g)-3). 尤其需要关注的是, 无论是热红外图像 (图 8 (b)-3) 还是显著图像 (图 8 (c)-3), 都将左起第二处区域中两个相互重叠的行人检测为同一个人, 而通过将两种图像融合增强后, 在图 8 (d)-3、图 8 (f)-3 和图 8 (g)-3 中成功地将这两个行人分别检测出来, 表明热红外图像与显著图的结合大大提高了重叠目标的检测性能.

4 结论

面向行人和车辆两类典型目标, 本文利用无人机平台采集制作热红图像数据集, 并对其进行边界框标注和像素级标注. 利用深度学习的方法, 提取热红外图像的显著图. 将热红外图像与提取的显著图进行结合, 通过通道替换以及像素级加权等多种图像融合方案, 在热红外图像输入目标检测的深度神经网络之前进行图像增强. 融合图像为行人及车辆检测模型提供互补信息, 从而显著提高目标检测的性能. 此外, 通过使用轻量化网络 YOLOv3-MobileNetv2 替代原先的检测网络, 在保持平均精确度基本不变的前提下, 提升了模型的检测速度, 降低了模型的尺寸, 使其可以更好地应用于无人机场景下的行人及车辆检测.

References

1 Liu Zhi-Jia, Jia Peng, Xia Yin-Hui, Lin Yu, Xu Chang-Bin. Development and performance evaluation of infrared and visual image fusion technology. *Laser and Infrared*, 2019, **49**(5): 633−640
(刘智嘉, 贾鹏, 夏寅辉, 林昱, 徐长彬. 基于红外与可见光图像融合技术发展与性能评价. 激光与红外, 2019, **49**(5): 633−640)

2 Koch C, Ullman S. Shifts in selective visual attention: Towards the underlying neural circuitry. *Human Neurobiology*, 1985, **4**(4): 219−227

3 Redmon J, Farhadi A. YOLOv3: An incremental improvement [Online], available: https://arxiv.org/abs/1804.02767, April 8, 2018

4 Qin X B, Zhang Z C, Huang C Y, Gao C, Dehghan M, Jagersand M. BASNet: Boundary-aware salient object detection. In: Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, USA: IEEE, 2019. 7479−7489

5 Sandler M, Howard A, Zhu M L, Zhmoginov A, Chen L C. MobileNetV2: Inverted residuals and linear bottlenecks. In: Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Salt Lake City, USA: IEEE, 2018. 4510−4520

6 Lin T Y, Goyal P, Girshick R, He K M, Dollár P. Focal loss for dense object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, **42**(2): 318−327

7 Lecun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 1998, **86**(11): 2278−2324

8 Yu Xue-Song, Liu Jia-Feng, Tang Xiang-Long, Huang Jian-Hua. Estimating the pedestrian 3D motion indoor via hybrid tracking model. *Acta Automatica Sinica*, 2010, **36**(4): 610−615
(于雪松, 刘家锋, 唐降龙, 黄剑华. 基于概率模型的行人四肢自遮挡的检测. 自动化学报, 2010, **36**(4): 610−615)

9 Dollár P, Tu Z W, Perona P, Belongie S. Integral channel features. In: Proceedings of the 2009 British Machine Vision Conference (BMVC). London, UK: BMVA Press, 2009. 91.1−91.11

10 Girshick R, Donahue J, Darrell T, Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Columbus, USA: IEEE, 2014. 580−587

11 Girshick R. Fast R-CNN. In: Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV). Santiago, Chile: IEEE, 2015. 1440−1448

12 Ren S Q, He K M, Girshick R, Sun J. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, **39**(6): 1137−1149

13 Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: Unified, real-time object detection. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, USA: IEEE, 2016. 779−788

14 Redmon J, Farhadi A. YOLO9000: Better, faster, stronger [Online], available: https://arxiv.org/abs/1612.08242, December 25, 2016

15 Ruhé M, Kühne R, Ernst I, Zuev S, Hipp E. Air borne systems and datafusion for traffic surveillance and forecast for the soccer world cup, In: Proceedings of the 86th Annual Meeting of Transportation Research Board (TRB 2017), Washington, DC, USA, 2007.

16 Portmann J, Lynen S, Chli M, Siegwart R. People detection and tracking from aerial thermal views. In: Proceedings of the 2014 IEEE International Conference on Robotics and Automation (ICRA). Hong Kong, China: IEEE, 2014. 1794−1800

17 Dong Pei, Shi Fan-Huai. Road detection method based on small

unmanned aerial vehicle image. *Computer Engineering*, 2015, **41**(12): 36−39
(董培, 石繁槐. 基于小型无人机航拍图像的道路检测方法. 计算机工程, 2015, **41**(12): 36−39)

18 Zhang Xiu-Wei, Zhang Yan-Ning, Yang Tao, Zhang Xin-Gong, Shao Da-Pei. Automatic visual-thermal image sequence registration based on co-motion. *Acta Automatica Sinica*, 2010, **36**(9): 1220−1231
(张秀伟, 张艳宁, 杨涛, 张新功, 邵大培. 基于co-motion的可见光-热红外图像序列自动配准算法. 自动化学报, 2010, **36**(9): 1220−1231)

19 Li C Y, Song D, Tong R F, Tang M. Illumination-aware faster R-CNN for robust multispectral pedestrian detection. *Pattern Recognition*, 2019, **85**: 161−171

20 Xu D, Ouyang W L, Ricci E, Wang X G, Sebe N. Learning cross-modal deep representations for robust pedestrian detection. In: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, USA: IEEE, 2017. 5363−5371

21 Itti L, Koch C, Niebur E. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1998, **20**(11): 1254−1259

22 Hou X D, Zhang L Q. Saliency detection: A spectral residual approach. In: Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Minneapolis, USA: IEEE, 2007. 1−8

23 He S F, Lau R W H, Liu W X, Huang Z, Yang Q X. SuperCNN: A superpixelwise convolutional neural network for salient object detection. *International Journal of Computer Vision*, 2015, **115**(3): 330−344

24 Hou Q B, Cheng M M, Hu X W, Borji A, Tu Z W, Torr P. Deeply supervised salient object detection with short connections. In: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, USA: IEEE, 2017. 3203−3212

25 Zhang Fang, Wang Meng, Xiao Zhi-Tao, Wu Jun, Geng Lei, Tong Jun, et al. Saliency detection via full convolution neural network and low rank sparse decomposition. *Acta Automatica Sinica*, 2019, **45**(11): 2148−2158
(张芳, 王萌, 肖志涛, 吴骏, 耿磊, 童军, 等. 基于全卷积神经网络与低秩稀疏分解的显著性检测. 自动化学报, 2019, **45**(11): 2148−2158)

26 Iandola F N, Han S, Moskewicz M W, Ashraf K, Dally W J, Keutzer K. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and $<$ 0.5MB model size [Online], available: https://arxiv.org/abs/1602.07360, November 4, 2016

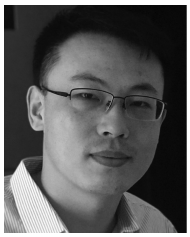
27 Zhang X Y, Zhou X Y, Lin M X, Sun J. ShuffleNet: An extremely efficient convolutional neural network for mobile devices [Online], available: https://arxiv.org/abs/1707.01083, December 7, 2017

28 Howard A G, Zhu M L, Chen B, Kalenichenko D, Wang W J, Weyand T, et al. MobileNets: Efficient convolutional neural networks for mobile vision applications [Online], available: https://arxiv.org/abs/1704.04861, April 17, 2017

29 Howard A, Sandler M, Chu G, Chen L C, Chen B, Tan M X, et al. Searching for MobileNetV3 [Online], available: https://arxiv.org/abs/1905.02244?context=cs, November 20, 2019

30 Fang Qing-Yun, Wang Zhao-Kui. Efficient object detection method based on improved YOLOv3 network for remote sensing images. *Aerospace Shanghai*, 2019, **36**(5): 21−27, 34
(方青云, 王兆魁. 基于改进YOLOv3网络的遥感目标快速检测方法. 上海航天, 2019, **36**(5): 21−27, 34)

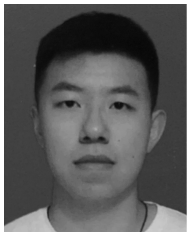
赵兴科 南京航空航天大学电子信息工程学院硕士研究生. 主要研究方向为深度学习与计算机视觉.
E-mail: zxk313@nuaa.edu.cn
(**ZHAO Xing-Ke** Master student at the College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics. His research interest covers deep learning and computer vision.)



李明磊 南京航空航天大学电子信息工程学院副教授. 主要研究方向为摄影测量与遥感和计算机视觉. 本文通信作者.
E-mail: minglei_li@126.com
(**LI Minglei** Associate professor at the College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics. His research interest covers photogrammetry and remote sensing, and computer vision. Corresponding author of this paper.)

张 弓 南京航空航天大学电子信息工程学院教授. 中国宇航学会电磁信息专业委员会委员.主要研究方向为雷达信号处理, 目标探测与识别.
E-mail: gzhang@nuaa.edu.cn
(**ZHANG Gong** Professor at the College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics, and member of Electromagnetic Information Committee of Chinese Astronautical Society. His research interest covers radar signal processing and target detection and recognition.)

黎 宁 南京航空航天大学电子信息工程学院副教授. 主要研究方向为视频图像处理, 目标检测与跟踪.
E-mail: lnee@nuaa.edu.cn
(**LI Ning** Associate professor at the College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics. Her research interest covers video image processing, and target detection and tracking.)



李家松 南京航空航天大学电子信息工程学院硕士研究生. 主要研究方向为计算机视觉与精密工业测量.
E-mail: jeasonlee_0@163.com
(**LI Jia-Song** Master student at the College of University of Aeronautics and Astronautics. His research interest covers computer vision and precision industrial measurement.)