

多光谱目标检测综述

王元喆, 梁腾飞, 曾宇乔, 金一, 李浥东

北京交通大学计算机科学与技术学院, 北京 100044

基金项目: 国家自然科学基金项目(61972030); 教育部-中移动基金项目(MCM20200203)

通信作者: 金一, yjin@bjtu.edu.cn 收稿/录用/修回: 2023-07-03/2024-01-15/2024-05-15

摘要

多光谱目标检测技术是一种利用不同波长光谱的信息来检测和识别目标的方法。该技术结合了可见光、红外、热成像等不同光谱的数据, 通过融合多光谱数据来提高目标检测的精确度和鲁棒性。在低光或其他具有挑战性的环境中, 多光谱目标检测表现出优越的性能。该技术在夜视监控、气象观测、农业、环境监测等领域有着广泛的应用。本文首先介绍了多光谱图像的定义、获取方式及常用数据集; 其次, 介绍了多光谱目标检测算法的分类并给出相关算法示例; 再次, 介绍了多光谱目标检测的应用, 重点讨论了多光谱目标检测在行人检测、农业、环境监测等领域的具体应用案例; 最后, 探讨了多光谱目标检测未来的发展方向。

关键词

多光谱
目标检测
特征提取
深度学习
图像融合

中图法分类号: TP37

文献标志码: A

Overview of Multispectral Object Detection

WANG Yuanzhe, LIANG Tengfei, ZENG Yuqiao, JIN Yi, LI Yidong

School of Computer Science and Technology, Beijing Jiaotong University, Beijing 100044, China

Abstract

Multispectral object detection technology is an approach that utilizes information derived from diverse wavelength spectra to detect and recognize objects. This technology integrates data of different spectras from visible light, infrared, and thermal imaging, and improves the accuracy and robustness of object detection. This technology performs excellently in low-light or other challenging environments. This technology is commonly applied in fields such as night vision monitoring, meteorological observation, agriculture, and environmental monitoring. First, the definition, acquisition methods, and common datasets of multispectral images are explained. Second, multispectral object detection algorithms are categorized, and examples of relevant algorithms are given. Third, the applications of the algorithms, which center on specific application cases in pedestrian detection, agriculture, and environmental monitoring, are discussed. Finally, the future directions for multispectral object detection are explored.

Keywords

multispectral;
object detection;
feature extraction;
deep learning;
image fusion

0 引言

目标检测是计算机视觉领域的基础研究之一, 旨在从图像或视频中准确地识别和定位出目标物

体。传统可见光目标检测易受光照条件限制, 例如在低光照或夜间条件下效果不佳, 同时受雾、烟、灰尘等环境影响, 检测性能会大幅下降。多光谱图像的引入可以有效解决上述问题。多光谱图像是通

过捕捉物体在不同波长范围内的反射或辐射能量而获取的图像,通常包括红外和可见光的多个波段,比起常见图像,多光谱图像记录了更广泛的光谱信息。多光谱图像对于光照变化、阴影、大气散射等环境因素具有一定的鲁棒性,并且不同波段的光谱信息在一定程度上能够减轻环境因素对目标检测造成的干扰,提高检测结果的稳定性和可靠性。多光谱图像还具有光谱信息丰富、目标区分度高、上下文信息丰富等优点。例如,随着红外摄像头的普及,行人检测往往使用红外图像作为额外光谱信息,红外图像反应物体的热辐射,因此可以在光照条件差的环境下提供有效的信息,同时红外光能更好地穿透雨雾等遮挡物,从而侦测到遮挡条件下的目标。由于多光谱目标检测相较于传统目标检测鲁棒性更好、检测精度更高,其得到越来越多的关注。多光谱图像的获取可以通过多光谱传感器或特定的遥感设备实现。这些设备能够同时捕捉不同波段的光谱信息,形成多光谱图像。多光谱图像在许多领域具有广泛的应用,包括农业^[1]、环境监测^[2-3]、地质勘探^[4]、遥感影像分析等。将多光谱图像应用到目标检测领域,可以利用不同波段的光谱信息,提供更全面、更准确的目标识别和分析,提高目标检测的精度,拓宽应用范围。

本文旨在综述基于多光谱的目标检测及其应用。首先介绍了多光谱图像的定义和相关信息,同时介绍多光谱图像的常用数据集;之后介绍多光谱目标检测的算法,引入图像融合的相关内容;最后介绍多光谱目标检测的应用领域和未来的潜在研究方向。

1 基于多光谱的目标检测

1.1 多光谱图像的定义和相关信息

为了深入了解多光谱目标检测,先要了解什么是多光谱图像。相较于传统目标检测使用的只包含红、蓝、绿三个波段的可见光图像,多光谱图像包含离散波段的图像,每个波段对应不同的光谱信息,如近红外、红外等。因此多光谱图像提供了更多维度的数据,使得目标的特征更加多样化和丰富。同时多光谱传感器通常在不同波段提供不同分辨率的数据,允许对细小目标进行更为准确的分析和检测,并且多光谱图像可以适应不同的环境和光照条件,它们对于白天和夜晚、晴天和阴天等不同条件下的目标检测具有良好的适应性。而使用多光谱图像可以实现更好的目标检测效果,更加重要的

原因是不同光谱图之间具有互补信息,例如可见光图像容易受光照影响,但可以提供更多细节;红外图像不受光照影响,但缺乏细节和纹理信息,二者互补可以取长补短。综上所述,多光谱目标检测相较于普通目标检测在获取更丰富、更准确的信息方面具有明显的优势。下面以可见光图像和红外光图像为例,详细介绍信息如何互补。

物体在可见光图像下,可以反映出颜色、形状、纹理和结构等细节信息。不同颜色的物体可能表示它们的材质、化学成分或表面特性不同。而物体的形状、大小和结构在可见光图像中往往是直接体现的。例如,建筑物和车辆都可以通过它们在图像中的形状识别出来。可见光图像还可以显示物体表面的纹理。例如植被的叶子、建筑物的窗户等。可见光图像还可以体现出反射率差异,由于物体对不同波长的光有不同的反射率,因此某些物体可能在某个波长下反射更多的光,这会导致图像中的明暗差异。

而红外热像图像捕捉了物体发射的红外辐射,波段通常位于长波红外或热红外波段。物体的温度分布和热辐射是在热成像图像中的主要信息来源。热成像图像可以显示物体表面的温度分布。不同的温度区域在图像中通常表现出不同的颜色或亮度。物体的热辐射与温度和表面特性有关,热成像图像可以显示不同物体之间的热辐射差异。有些物体可能在可见光图像中不可见,但在热成像图像中由于温度与环境不同而可见,如黑暗中的人体识别。热成像图像受环境温度、大气湍流等影响,可能出现热失真现象,需要进行校正。综上所述,可见光图像和热成像在不同方面捕捉物体的特征,它们可以相互补充,用于不同的应用领域。

1.2 常用多光谱数据集

常见的多光谱数据集如下。

KAIST 数据集^[5] 由 HWANG 等于 2015 年提出,是第 1 个多光谱行人检测基准数据集。该数据集在当今新提出的算法中依然得到了广泛使用,因此详细介绍一下此数据集。多光谱采集设备由彩色相机、热成像相机、分束器等构成。数据集共包含 95 328 个图片对,每对图片对由一张 RGB 彩色图像和一张红外图像构成,总共包含 103 128 个密集标注。数据集具体内容包括了白天、夜晚条件下的街道、校园等各种地点的交通场景图。KAIST 数据集存在的问题是存在大量冗余数据,使用前最好进行数据清理。图 1 为 KAIST 数据集示例。

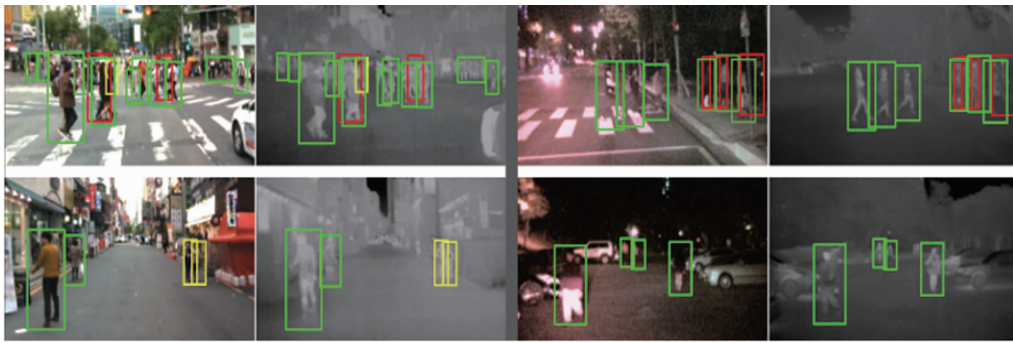


图1 KAIST 数据集示例^[5]

Fig.1 Examples of KAIST dataset^[5]

Vedai 数据集^[6] 由 RAZAKARIVONY 等于 2016 年提出, 此数据提供了用于车辆检测的航空图像。数据采集自 Utah AGRC 卫星, 共手动选择了 1 210 张 1 024 × 1 024 图像, 并将其纳入 4 个不同类型的子集。数据集中的图像背景多样化且车辆目标较小, 包括树林、城市、道路、停车场等各种背景, 车辆不限类型, 有不同方向、不同程度遮挡。数据中的车辆包括 9 种类别, 平均每幅图像有 5.5 辆车, 占总像素的约 0.7%。图 2 为 Vedai 数据集示例, 包括两张彩色图像和两张红外图像。

LLVIP 数据集^[7] 由 JIA 等于 2021 年提出。数据采集使用双目相机, 对采集到的图像进行了裁剪对齐。裁剪后的图像完全对齐, 包括标注框。LLVIP 数据集共包含 16 836 个图像对, 红外图像与可见光图像 1:1 构成。LLVIP 数据集作为弱光条件行人检测的基准数据集, 可以有效支持算法模型的训练与测试, 适合进行图像融合的研究。LLVIP 采集图像的分辨率非常高, 原始可见光图像的分辨率为 1 920 × 1 080, 红外图像的分辨率为 1 280 × 720。缺点为仅有中距离、中等大小行人图像, 对小目标

行人检测无法提供有效支持。图 3 为 LLVIP 数据集示例。



图2 Vedai 数据集示例^[6]

Fig.2 Examples of Vedai dataset^[6]



图3 LLVIP 数据集示例^[7]

Fig.3 Examples of LLVIP dataset^[7]

M3FD 数据集^[8] 由 LIU 等于 2022 年提出。数据采集使用双目光学相机和双目红外传感器,共包含 4 200 个用于图像融合、目标检测和基于图像融合的目标检测图像对及 300 个用于图像融合的独立场景的图像对。红外图像采用 24 位灰

度图,彩色图像采用 2×4 位彩色图,大部分图像分辨率为 $1\,024 \times 768$ 。数据集共包括 34 407 个手动标记标签,标签可以分为 6 种类别:人、汽车、公共汽车、摩托车、灯、卡车。图 4 为 M3FD 数据集示例。



图 4 M3FD 数据集示例^[8]

Fig.4 Examples of M3FD dataset^[8]

1.3 多光谱目标检测算法步骤

1.3.1 数据采集和预处理

使用多光谱数据是多光谱目标检测与传统目标检测的主要区别之一,因此收集多光谱数据是多光谱目标检测的第一步,也是对最终结果准确性和可靠性起到保证作用的至关重要的一步。但与普通目标检测不同,由于涉及到不同波段光谱的收集,因此多光谱目标检测首先要选择适合的传感器,如摄像机、红外摄影仪等多源传感器,这些传感器可以捕获物体光谱的不同波段或范围的信息,并且选择的波段数量和波段范围应匹配研究目标。

直接采集到的数据往往存在较多问题,例如图像噪声、分辨率低、可见光-红外光图像分辨率未对齐等,因此需要对采集到的数据进行预处理。数据预处理一般包括对齐、降噪、增强等操作,以保证收集到的数据质量。其中多光谱图像采集时最常遇到的问题便是两个相机视野大小即分辨率不同数据对齐的问题,例如 KAIST 数据集^[5]的处理过程中,由于彩色相机分辨率为 640×480 ,垂直视野为 103.6° ;热成像相机分辨率为 320×256 ,垂直视野为 39° ,二者无法直接配对,其对齐方式为牺牲可见光图像边缘区域,图像域对齐时以热成像相机为基准,这是图像对齐常使用的方法。同时,在多光谱图像采集时,由于相机位置的不同,还应进行图像校准。而对于图像噪声问题,常使用不同的滤波方

式进行去噪,如双边滤波^[9]、中值滤波^[10]等。图像增强在多光谱目标检测中可应用的方法非常广泛,如可以使用颜色增强来增强不同光谱的特征^[11],使用超分重建增强多光谱图像分辨率^[12-13]等。其余还存在如使用分束器进行可见光反射时产生的色彩失真问题。预处理阶段的最终目的为提升采集得到的数据质量,而随着摄像机的改进和红外摄像分辨率的提升,采集到的图像质量本身也在逐渐提升,因此如何调整预处理阶段的方法使其更好地作用于不断更新的数据集也成为多光谱目标检测研究的一个重要方向。

1.3.2 特征提取

在收集到多光谱数据并进行预处理后,要对得到的数据进行特征提取。传统目标检测中需要提取的特征往往包括颜色特征、纹理特征、形状特征、边缘特征等,而在多光谱目标检测中,除了上述特征,还可以提取不同光谱图独特的光谱特征。多光谱特征有助于识别物体的材料,可以提取出更多维度的数据,使特征空间更加丰富,有助于复杂场景中目标的检测和分类。例如可见光的蓝色波段更适合渗透水体,适合检测水体浊度和深度,绿色波段在植被上反射较多,适合探测植被覆盖率,红光在健康植被上被吸收较多,被反射较少,适合计算植被指数,近红外波有强烈植被反射,适合检测植被健康状况和生物量,短红外波可以穿透雾霾,适合

恶劣天气成像, 中红外波主要用于热成像, 适合检测火灾状况, 长红外波同样用于热成像, 更适合进行气象探测。特征提取具有重要研究意义, 如何更有效地提取出不同光谱图像间的相同特征和特有特征同样是多光谱目标检测中一个非常重要的研究方向。

AMZIANE 等^[14]提出的基于多光谱滤波阵列的卷积神经网络算法, 可以有效地从原始多光谱图像中直接提取纹理特征, WANG 等^[15]提出的基于几何代数的多光谱图像特征提取算法可以有效地提取出角点和图像边缘特征。

1.3.3 特征融合

特征融合是多光谱目标检测算法中最重要的一步, 其发生的阶段并不严格位于特征提取后, 直接对多光谱图像进行融合操作也可以视为特征融合操作。这里简单介绍提取出各自光谱图特征后的融合操作: 特征提取得到不同光谱图的特征后, 会发现它们之间存在有相似特征和独有特征, 其中相似特征同样存在特征冗余, 独有特征也包含对目标检测具有极大贡献的特征, 需要去除冗余特征, 提取对后续检测有作用的特征进行融合。因此如何选择特征融合算法, 才能更好地结合来自不同光谱的特征, 使其互补, 从而增强目标检测的准确性和鲁棒性是一个极重要的研究。众多团队也在特征融合这一步做出了探索。

WAGNER 等^[16]于 2016 年提出了使用深度融合卷积神经网络的多光谱行人检测算法, 首次将多光谱特征融合的想法引入到基于 CNN(convolutional neural network)的目标检测框架。LIU 等^[17]设计了 4 种卷积神经网络(ConvNet)融合架构, 分别在卷积阶段、全连接阶段、决策阶段融合两个不同分支的卷积神经网络。由于特征融合的效果往往决定了多光谱目标检测的性能, 因此多光谱目标检测算法大部分探索创新均发生于特征融合阶段。

1.3.4 目标检测

多光谱目标检测的检测阶段, 主要目的是通过分析特征提取和特征融合阶段所得到的图像特征, 完成目标识别、目标定位、目标分类等任务。多光谱目标检测算法在目标检测阶段与传统可见光目标检测算法存在许多共同之处, 因此简要介绍一些传统可见光目标检测算法。

目标检测最早可追溯到 20 世纪 70 年代, 研究人员主要关注于边缘检测、区域分割和形状分析等问题。这些技术为目标检测的发展奠定了基础。

1991 年 PENTLAND 和 TURK 提出了在图像中检测人脸的方法。21 世纪前后, 目标检测进入了机器学习时期, 以 Viola-Jones 检测器、HOG(histogram of oriented gradient)特征描述子为代表, 出现了许多基于特征的目标检测方法。早期的目标检测方法主要基于传统的图像处理技术, 依赖于手工设计的特征和规则, 限制了其在复杂场景下的性能。2014 年, GIRSHICK 等^[18]提出了 R-CNN(region-convolutional neural network)方法, 将深度学习引入目标检测领域。R-CNN 通过在候选区域上提取特征, 并使用卷积神经网络进行分类, 实现了较高的准确性。随后, Fast R-CNN^[19]和 Faster R-CNN^[20]方法相继提出, 进一步提高了检测速度和准确性。传统的目标检测方法通常采用两阶段的检测框架, 需要先提取候选框再进行分类。而近年来, 单阶段检测器成为主流, 如 YOLO(you only look once)^[21]、SSD(Single Shot multibox Detector)^[22]等。这些方法通过在不同尺度上进行多层次的特征提取和预测, 实现了更快速和准确的目标检测。目标检测领域仍在不断发展, 涌现出许多新的方法和思路。其中, 深度学习的进一步发展, 如使用注意力机制、跨域检测和跨模态检测等, 为目标检测带来了更好的性能和泛化能力。2017 年, Transformer^[23]的提出使深度学习模型不再局限于卷积网络, 目标检测同样可以基于 Transformer 架构。2020 年, Facebook 团队提出的 DETR(DEtection TRansformer)算法^[24]是第一个基于注意力机制的目标检测算法, 旨在解决 YOLO 系列处理繁琐耗时、不易优化的问题。表 1 总结了目标检测发展历程中的若干重要事件。

表 1 目标检测发展历史重要事件

Tab.1 Important events of object detection development history

年份	提出者	事件
1995	VAPNIK	SVM 分类器
2001	VIOLA、JONES	Viola-Jones 算法
2005	DALAL	HOG 特征提取
2014	GIRSHICK	R-CNN 算法 ^[18]
2015	REDMON	YOLO 算法 ^[21]
2016	LIU	SSD 算法 ^[22]
2017	REN	Faster R-CNN 算法 ^[20]
2020	CARION	DETR 算法 ^[24]

多光谱目标检测的算法大多基于传统目标检测算法, 在可见光目标检测算法框架上加入特征融合

模块或其他可以结合不同光谱图间互补信息的模块,从而适应特定的多光谱检测场景。例如 FANG 等^[25]于 2021 年提出的算法就是基于 YOLOv5 算法框架,加入 CFT(Cross-modality Fusion Transformer)模块进行多光谱目标检测;XIE 等^[53]于 2023 年提出的 YOLO-MS 同样基于 YOLOv5 算法框架,加入了特征交互模块和自注意力引导融合模块。多光谱目标检测算法的检测头往往都是直接使用较为先进的目标检测算法检测头,因此介绍多光谱目标检测算法时,往往只需介绍其创新性模块,直接基于基础框架的创新往往会先应用于单一可见光目标检测算法。

1.3.5 后处理

多光谱目标检测的后处理阶段并非必须的,在许多算法中,都隐藏了后处理阶段,将其自然地融入算法框架中,完成目标检测便结束了算法流程。如果不追求较高的效率,可以对算法的输出进行

优化处理,同样有助于提高整体目标检测的准确性。常见的后处理方式有非极大值抑制(non-maximum suppression, NMS)、阈值筛选、边界框回归等方法。非极大值抑制即对得到的检测框进行筛选,仅留下精度最高的检测框,可以有效去除重复结果,提高模型的效率。现在非极大值抑制延伸出如 Soft-NMS、Adaptive NMS 等诸多变体。阈值法是近几年较为流行的处理方式,即置信度筛选,根据置信度得分来选择检测精度最高的框,YOLO 算法便是使用了该方式。便捷框回归用于微调检测到的对象的边界框,使其更准确地包围目标。合理使用后处理可以平衡算法精度和效率。

1.4 多光谱目标检测算法

1.4.1 多光谱目标检测算法分类

多光谱目标检测算法按照融合阶段可划分为数据级融合算法、特征级融合算法和决策级融合算法,如图 5 所示。

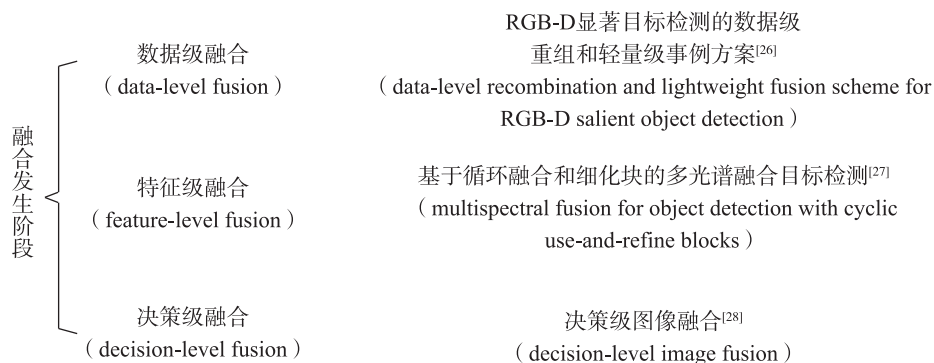


图 5 分类方式: 融合阶段

Fig.5 Classification method: Fusion stage

数据级融合直接在原始数据层面进行融合,能够最大程度地保留原始信息,但可能导致数据量大和计算复杂度高。如 WANG 等^[26]于 2020 年提出的数据层重组和轻量级融合方法,旨在更有效地融合 RGB 和深度信息,解决传统双流网络倾向于偏向 RGB 数据的限制。该方法包括将深度通道与 RGB 通道循环重组,创建新的数据表现形式。

特征级融合在特征提取后进行融合,减少了数据量,便于处理,但可能丢失一些原始数据中的信息。如 ZHANG 等^[27]于 2020 年提出的基于循环融合细化块的多光谱目标检测方法,其中的 CFR(cyclic fuse-and-refine)模块就是互补信息发生在特征层的方法,通过循环融合和细化操作,逐渐提升特

征的质量并调整特征间的互补性和一致性平衡,可以有效改善多光谱目标检测的性能。

决策级融合在做出单独决策之后融合,能够结合多种方法的优势,但可能无法充分利用数据之间的关联信息,融合损失较大。如 XIAO 等^[28]于 2020 年出版的图像融合一书中,就决策级图像融合做出了单独的介绍,并给出了使用 FLIR 及 SAR 数据集的算法流程和实现。

此外按照不同应用领域,可分类为用于交通、农业、环境监测领域等的多光谱目标检测算法,如图 6 所示。

1.4.2 多光谱目标检测算法的发展

多光谱目标检测算法在算法层面做出创新最多的领域是多光谱行人检测,其他领域更多是对方法

为基准模型,提出了基于原始图像的通道特征,每个通道特征强调图像的某些方面,例如颜色通道、梯度幅度、梯度方向、边缘信息等。多光谱 ACF 将 ACF 的基于彩色图像的 10 个增强通道拓展到热强度通道编码。算法共建议了 ACF + T、ACF + T + TM + TO 以及 ACF + T + THOG 三个拓展基线,其中 T 为热通道, TM 为热图像的归一化梯度幅值, TO 为热图像的定向梯度直方图, THOG 为增强版的 TO, 计算更多梯度方向并加入归一化步骤, 实验得出 ACF + T + THOG 性能最优, 将其命名为多光谱 ACF。该算法在 KAIST 数据集上的进行实验, 使用漏检率作为评价标准。漏检率即算法未能正确检测到的实际目标占总实际目标数量的比例, 漏检率也是多光谱行人检测中普遍使用的一个评价指标。多光谱 ACF 根据白天夜晚条件、距离远近条件及遮挡程度条件进行了对比实验, 证明了其在白天夜晚、远近距离、有无遮挡条件下性能均优于 ACF 检测器, 在一般条件下可以将漏检率降低至 64.76%。

随着多光谱行人检测的基准数据集和基线方法的提出, 多光谱行人检测任务引起了广泛研究。众多团队开始研究如何更好地融合特征。最有代表性的是 WAGNER^[16] 于 2016 年提出的使用深度融合卷积神经网络的多光谱行人检测算法, 将多光行人检测器根据融合发生的抽象级别划分为像素级、特征级和决策级。该算法建立在 R-CNN 检测框架上, 分别研究了像素级别早期融合架构(early fusion)和特征级别的后期融合(late fusion) CNN 架构。在 KAIST 数据集上的实验结果表明, 大多数情况下早期融合架构性能无法达到多光谱 ACF 的性能, 原因可能为数据匮乏、预训练过程缺乏互补信息及数据未对齐问题。而后期融合架构由于信息融合发生在与空间信息不太相关的阶段, 因此性能表现更为出色, 将漏检率降低至 43.8%, 优于当时最先进的多光谱 ACF 算法。

CHOI 等^[30] 于 2016 年针对多光谱 ACF 算法中由于手工标记特征, 而无法有效对较差环境条件下的行人进行检测的问题, 提出了一种基于全卷积神经网络^[31] 和支持向量回归的多光谱目标检测算法。在 KAIST 数据集上将漏检率降低至 47.3%。LIU 等^[17] 基于 Faster R-CNN, 探讨了 4 种多光谱深度神经网络的融合架构, 分别为早期融合、中期融合、晚期融合、得分融合。LIU 等^[17] 的研究结果表明, 相比其它的融合架构, 中期融合平衡了细节可视化特征和语义信息, 从而得到了最好的结果, 在 KAI-

ST 数据集上将漏检率降低至 36.9%。

上述方法发现, 相较于 Faster R-CNN 网络, 区域建议网络对多光谱行人的检测精度更高。KONIG 等^[32] 于 2017 年提出了基于全卷积区域建议网络的多光谱人体检测, 结合多光谱融合 RPN(region proposal network) 和增强决策树 BDT(boosted decision trees) 分类器进行了进一步的实验, 在 KAIST 数据集上将漏检率降低至 29.83%。除使用卷积神经网络提升特征融合效果外, 各团队也探寻其他途径对检测精度进行提升, 如生成对抗网络, GONG 等^[33] 于 2017 年提出了用于多光谱图像变化检测的生成对抗网络(generative adversarial network, GAN), 使用 GAN 学习两幅不同时间的多光谱图像之间的变化, 生成更准确的差异图像(DI) 以提高多光谱图像变化检测的效果和鲁棒性。

对于多光谱 ACF 提出的拓展通道特征, PARK 等^[34] 于 2018 年提出了概率融合(probabilistic fusion, PF) 网络, 该网络包含两个融合层: 通道加权融合(CWF)层和累积概率融合(APF)层。通道加权融合层的主要功能是在特征层面对不同模态的信息进行加权融合。评估每个通道的信息内容, 并根据其对检测任务的贡献进行加权, 从而有效地结合不同模态的特征。累积概率融合层致力于在提案级别上合并不同模态的概率信息。它累积各模态检测提议的概率, 以形成一个综合的、更准确的检测结果。通过这种方式, APF 层可以充分利用各模态数据的互补性质, 提高检测的精确度和鲁棒性。这两层的设计使得系统能够在复杂的场景中更好地处理不同环境条件下的行人检测问题。算法在 KAIST 数据集上将漏检率降低至 26.29%。LI 等^[35] 于 2018 年提出了一种多光谱同时检测和分割 R-CNN 算法(multispectral simultaneous detection and segmentation R-CNN, MSDS-RCNN)。该算法包含两个子网络: 用于生成行人提案的多光谱提案子网络和用于区分行人正负例样本的多光谱分类子网络。该算法将漏检率降低至 11.6%。

2019 年, 许多研究团队对多光谱目标检测算法开展了深入研究, 不仅优化特征提取和特征融合阶段, 并且还加入了对照明条件、数据对齐等问题的思考。GUAN 等^[36] 提出了用于行人检测的基于光照感知深度神经网络的多光谱数据融合方法, 共包括 3 个集成处理模块组: 照明全连接神经网络(illumination fully connected neural network, IFCNN)、照明感知双流深度卷积神经网络(illumination-aware

two-stream deep convolutional neural network, IATDC-NN)、照明感知多光谱语义分割(illumination-aware multispectral semantic segmentation, IAMSS)。该方法考虑到在白天和夜晚,行人的多光谱图像会表现出显著不同的特性,设计了多个内在子网络来捕获不同光照条件下的特性,首次利用照明信息来提高检测性能。其提出了照明感知加权机制,输入一对对齐的可见光图像及红外光图像,二者分别经过双流深度神经网络以生成多光谱语义特征图,随后日光光子网络和夜光子网络利用多光谱语义特征图进行不同光照条件下的行人检测和语义分割。生成双流特征映用作后续 IFCNN 的输入来计算照明感知权重,从而确定场景的照明条件,通过融合多个光照感知子网络的输出,得到最终的检测结果。实验结果表明,该方法在 KAIST 数据集上可将漏检率降低至 26.37%。

LI 等^[37]于 2019 年提出的照明感知 Faster R-CNN(illumination-aware faster R-CNN, IAF-RCNN)网络也同样使用照明感知机制,使用辅助的照明感知网络来评估输入图像的照明条件,基于评估结果调整颜色和热成像子网络的融合权重。ZHANG 等^[38]于 2019 年提出了跨模态交互式注意力网络(cross-modality interactive attention network, CIAN),采用交互式注意力机制来增强特征融合的效果。同年,其团队还提出了弱对齐的跨模态学习^[39]算法——AR-CNN(aligned region CNN),主要解决多光谱数据中存在的位置偏移问题,创新地应用区域特征对齐模块,预测和校正不同模态之间的位置偏移,提高了特征融合的精度。

2020 年,注意力机制在人工智能领域得到了更广泛的应用。ZHANG 等^[40]提出的基于注意力的多光谱图像多层融合行人检测(multi-layer cross-spectral fusion R-CNN, CS-RCNN)结合通道级和空间级的注意力模块,旨在更精细地调整多光谱特征的权重。同时多光谱目标检测算法开始思考可以加入独特模块来提高信息互补程度,ZHANG 等^[41]提出了一种新颖的多光谱特征融合方法,命名为循环融合细化块(cyclic fuse-and-refine module, CFR),通过循环融合和细化操作,逐渐提升特征的质量并调整特征间的互补性和一致性平衡。ZHOU 等^[42]提出了模态平衡网络(modality balance network, MB-Net),通过使用差异化模态感知融合模块和照明感知特征对齐模块来解决多光谱行人检测中的模态不平衡问题。WOLPERT 等^[43]则考虑学习行人的中心

点和尺度,而非直接预测边界框,这在小尺度或部分遮挡的行人检测中有极优效果。

2021 年,随着 ViT(vision transformer)^[44]的提出,研究人员纷纷将目光投向 Transformer 架构。ViT 凭借其高效的特征提取能力、更好的泛化能力、简化的模型架构等优势,挑战了卷积神经网络在视觉领域的统治地位。

FANG^[45]等首先将 Transformer 架构应用于多光谱目标检测领域,提出了跨模态融合 Transformer(cross-modality fusion transformer, CFT)算法。算法基于 YOLOv5 框架,但是将 YOLOv5 的特征提取网络重新设计为一个双流主干,并嵌入 CFT 模块以促进模态融合和模态交互,称为跨模态融合主干,但对于 YOLO 整体的框架并未做太大修改,如 C3 卷积模块,其精良的设计可以有效提取出图像特征,模型整体框架图如图 8 所示。Transformer 架构不同于卷积神经网络非全局的感受野,其会在特征提取阶段学习前后依赖关系并生成全局上下文信息,并且 Transformer 架构的自注意力机制可以使网络自然地同时执行模态内和模态间融合,鲁棒地检测出可见光图和热图像之间潜在的互补关系,从而提高多光谱目标检测的性能。算法分别在 FLIR、LLVIP 和 Vedai 三个数据集上进行实验,并与单一可见光、红外光模态及使用简单的双流模型进行了比较。实验结果表明 CFT 别实现了 78.5、97.5 和 85.3 mAP 50 的最佳性能。并且经过与 YOLOv3、YOLOv5 和 Faster R-CNN 等三个不同检测器的组合测试后,发现该算法在每一个检测器上,性能均得到了提升。这里的评价指标 mAP(mean Average Precision)是多光谱目标检测领域的常用评价指标,衡量的是模型对目标检测的准确性。mAP 是所有类别的平均精度(AP)的平均值。AP 计算的是不同召回率(recall)水平下的平均精度。mAP 考虑了检测的精度和召回率,是衡量目标检测性能的综合指标。较高 mAP 值意味着模型在检测准确性和召回率上都表现良好。其中 mAP50 和 mAP75 分别表示 IoU 阈值为 0.5 和 0.75 时计算的 mAP,意味着预测的边界框和真实边界框的交并比至少为 50% 和 75%。

ZHANG 等^[46]于 2021 年提出了引导式注意特征融合(guided attentive feature fusion, GAFF)算法。通过结合模态内和跨模态的注意力模块,实现多光谱特征的动态加权和融合。内部模态注意力模块专注于增强单光谱视图中的热成像或可见光特征,而跨模态注意力模块则根据特征质量的动态比较,适

应性地选择热成像或可见光特征。KIM 等^[47]于 2021 年提出了多谱域多标记行人检测器 (multi-label pedestrian detector, MLPD), 通过结合多标记学习和半配对增强策略, 有效处理多光谱图像中的非配对和配对场景。

2022 年具有代表性的几种算法依然聚焦于特征融合阶段, 并且加入局部化和注意力机制来提升检测性能, 同时更关注于实时检测。CAO 等^[48]提出了基于局部引导的跨模态特征聚合和像素级融合的多光谱行人检测方法 (locality guided cross-modal

feature aggregation and pixel-level fusion network, LG-FAPF), 创新性地增加了局部引导的跨模态特征聚合模块和像素级预测融合模块。LIU 等^[49]提出了 M3FD 数据集和目标感知的双重对抗学习与多场景多模态基准融合方法。KIM 等^[50]提出了一种用于鲁棒多光谱行人检测的不确定性引导的交叉模态学习方法, 增加了不确定性感知特征融合 (uncertainty-aware feature fusion, UFF) 和不确定性感知交叉模态引导 (uncertainty-aware cross-modal guiding, UCG) 两个模块。

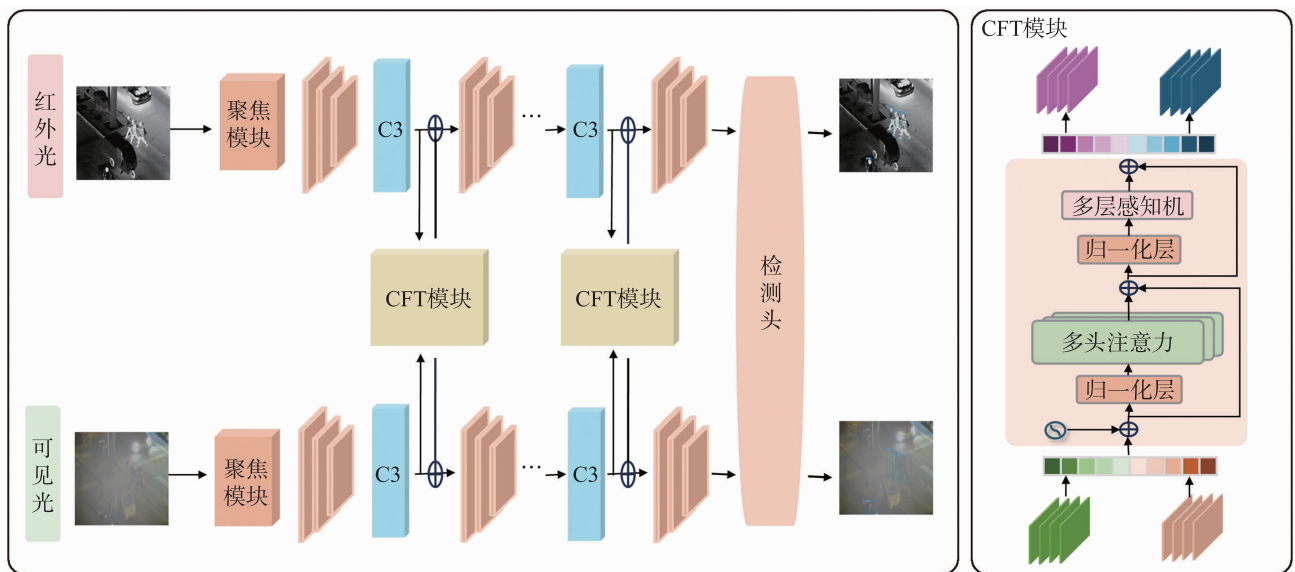


图 8 CFT 算法模型示意图^[45]

Fig.8 Schematic diagram of the CFT algorithm model^[45]

ZHANG 等^[51]于 2023 年提出了基于照明引导的多光谱目标检测算法。该方法结合照明引导的跨模态和内模态信息, 有效地提高了在不同光照条件下的目标检测性能。该算法包含了 3 个主要模块: 照明引导特征加权模块用于根据光照条件动态调整 RGB 和热成像模态特征的权重。跨模态注意力模块增强了不同模态之间的互补信息。内模态注意力模块强化了单一模态内的前景特征。通过这种方式, 其能够在各种光照条件下提供稳健的目标检测性能。

LI 等^[52]提出了多尺度跨模态同质性增强与置信度感知融合方法 (multiscale cross-modal homogeneity enhancement and confidence-aware fusion network, MCHE-CF)。MCHE-CF 网络使用了两个主要创新模块: 多尺度跨模态同质性增强模块和置信度感知融合模块。前者用于在特征融合前增强单模态

特征, 后者用于根据每种模态特征的可靠性进行特征融合。此外, 还设计了一种置信度转移损失函数, 以在训练过程中能够更加关注难以检测的样本, 从而提高检测性能。

XIE 等^[53]提出了一种基于特征交互和自注意力引导融合的多光谱目标检测方法。该方法设计了特征交互模块 (feature interaction module, FIM), 用来分析不同模态的特征以及它们之间的相关性, 以提升特征融合的有效性。此外, 该方法还设计了自注意力特征融合模块 (self-attention feature fusion module, SAFFM) 来指导不同模态特征的融合。通过自注意力机制, SAFFM 能够更加精准地识别和强调那些对目标检测更为重要的特征区域。

ZHANG 等^[54]提出了一种针对多光谱行人检测的新型目标感知融合策略, 命名为 TFDet (Target-aware Fusion for RGB-T pedestrian Detection)。该融

合策略突出了与行人相关的特征并抑制了无关的特征, 生成了更具区分性的融合特征, 背景区域的误报显著减少。该融合策略不仅更加灵活, 还实现了最先进的性能表现, 将漏检率降低至 4.47%。

对之前算法进行总结可以发现, 多光谱目标检测算法的基本框架主要有基于卷积神经网络、基于生成对抗网络和基于 Transformer 三种架构。在这 3 种基本逻辑思路之上, 可以加入照明感知机制、置信度融合机制、注意力机制等。新的多光谱目标检测算法的创新点便可以立于以上几点, 可以是基本逻辑框架的创新, 如 Vision Transformer 等重大创新, 也可以是利用某种不同光谱图独有或共有的信息实现更好的信息互补。将所有使用 KAIST 数据集进行实验的算法结果统计, 如表 2 所示, 可以看出随着模块的不断细化创新, 漏检率同样也在不断降低。

表 2 基于 KAIST 数据集的行人检测算法结果
Tab.2 Pedestrian detection results on KAIST dataset

单位: %

方法	漏检率 (全部)	漏检率 (白天)	漏检率 (夜晚)
ACF + T + THOG ^[5]	64.76	64.17	63.99
LateFusion + PreTraining ^[16]	43.80	46.15	37.00
Halfway Fusion ^[17]	36.99	36.84	35.49
Fully Convolutional ^[30]	47.31	49.31	43.75
PF networks ^[34]	31.36	31.79	30.82
MSDS-RCNN ^[35]	11.63	10.60	13.73
IATDNN + IAMSS ^[36]	26.37	27.29	24.41
IAF-RCNN ^[37]	15.73	14.55	18.26
CIAN ^[38]	27.71	30.74	21.07
AR-CNN ^[39]	9.94	9.34	8.38
CS-RCNN ^[40]	11.43	11.86	8.82
CFR 3 ^[41]	10.05	9.72	10.80
MB-Net ^[42]	8.13	8.28	7.86
GAFF ^[46]	6.48	8.35	3.46
MLPD ^[47]	7.58	7.95	6.95
LG-FAPP ^[48]	5.12	5.83	3.69
UCG ^[50]	7.89	8.00	6.95
MCHE-CF ^[52]	6.71	7.58	5.52
TFDet ^[54]	4.47	5.22	3.36

1.4.3 多光谱目标检测在农业领域中的应用

多光谱目标检测在农业领域有广泛的应用, 可以帮助农民和农业专业人员监测和管理农作物的生长状况, 优化农业生产效率。通过多光谱图像的分析, 可以获取农作物反射的不同波段信息, 如植被指数等, 用以评估农作物的健康状况, 从而使农民及时采取措施来提高农作物的产量和质量。多光谱目标检测还可以帮助评估土壤的性质和质量。通过分析土壤的光谱特征, 获取土壤的含水量、有机物含量、养分含量等信息, 用以对土壤肥力进行评估, 从而进行管理决策, 帮助农民制定适当的土壤改良和施肥计划。此外, 病虫害通常会导致农作物反射光谱变化, 基于此, 多光谱目标检测可以用于病虫害监测, 及时发现农作物的病虫害问题。

LI 等^[55]于 2016 年提出基于可见-近红外高光谱成像对双色桃果皮缺陷进行多光谱检测, 用于解决由于水果表皮的缺陷导致水果变质问题, 提升提升水果经济价值。该研究使用基于实验室的线扫描高光谱反射成像技术检测桃子表皮的缺陷, 样品检测准确率可以达到 96.6%, 证明了多光谱算法可以有效检测出双色桃果皮缺陷。

ZHOU 等^[56]于 2017 年提出使用多时相植被指数预测水稻产量, 该指数基于使用无人机收集的多光谱和数字图像。农作物的产量预测可以使国家及时调整粮食政策, 具有深刻的意义。该实验使用多旋翼 MK-Oktokopter 无人机采集图像, 经处理后选择合适的光谱进行后续实验, 实验验证了多光谱检测可以有效预测水稻生长周期和估计水稻产量。

De SILVA 等^[57]于 2023 年提出使用 Vision Transformers 对多光谱自然环境图像进行植物病害检测。算法使用现实场景的 5 种由可见光和近红外图像组成的农业植物数据集, 通过将 ViT 模型迁移学习, 病虫害识别准确率可以达到 90.02%。

1.4.4 多光谱目标检测在环境监测中的应用

多光谱目标检测在环境监测方面发挥着重要的作用。利用卫星、无人机拍摄的多光谱遥感图^[48], 并配合迁移至多光谱领域的目标检测算法, 搭配独特的多光谱特征融合模块, 可以有效检测地表的特定目标。例如, 多光谱目标检测可以对农田、森林、湖泊、城市等地表特征进行分类和监测。

NOI 等^[58]于 2017 年提出在土地覆盖问题中,

使用多光谱图像数据,结合机器学习算法,例如随机森林(random forest, RF)、 k 近邻(k -nearest neighbor)算法和支持向量机(SVM)等,可以有效完成土地覆盖的分类。该算法使用 Sentinel-2A 影像,共包含 13 个光谱带,并分别使用上述算法在经预处理后遥感数据上进行实验。实验结果显示,SVM 分类器平均产生最高的总体精度,对训练样本大小的敏感性最低,RF 和 kNN 稍次于 SVM。通过多光谱数据结合 3 种算法,土地覆盖分类准确率都在 90% ~ 95% 之间,证明了该算法可以有效地进行土地覆盖分类。

YUAN 等^[59]于 2021 年提出了基于深度学习的多光谱卫星影像水体检测分割方法。算法共提出 3 个创新组件:多通道融合模块、增强型空洞空间金字塔池模块和空间、深度转换操作模块。算法同样采用 Sentinel-2 影像数据,实验结果表明,MC-WBDN 算法性能高于其他评估模型,平均 mIoU (mean Intersection-over-Union) 达到 73.56 并且算法鲁棒性较高,可以在不同天气和环境下检测水体。

2 多光谱目标检测的发展展望

多光谱目标检测在遥感技术和图像处理领域已经取得了显著的进展,未来有以下几个研究方向:

1) 多源数据融合。多光谱目标检测技术与普通的目标检测算法的最大区别就在于可以融合来自不同光谱图的特征信息。不同来源的图像带有不同种类的信息,而不同种类信息之间的互补性可以更好地满足目标检测的需要,增强目标检测的准确性和可靠性,如高光谱、合成孔径雷达(SAR)、激光雷达(LiDAR)等。如何更好地将不同类型图像进行信息融合,是多光谱目标检测发展的一个方向。

2) 深度学习算法的研发。多光谱目标检测算法的骨干网络大多基于深度学习算法。目前为止,深度学习技术已经在图像处理和目标检测领域取得了巨大的成功,无论是卷积神经网络、生成对抗网络还是 Transformer 都已经投入了具体使用,并且每

一种算法上的研发都可以给具体的应用带来很广阔的发展空间。未来,多光谱目标检测也有望借助新的深度学习方法,提高目标检测的性能和效率。

3) 实时监测和自动化。多光谱目标检测技术的应用方面,随着计算能力的提升和算法的改进,多光谱目标检测将更加注重实时监测和自动化。这将使得交通、农业、城市规划等领域能够及时获取并处理多光谱数据,实现快速决策和精细化管理。

4) 多光谱目标检测的另一发展前景是算法的自适应和智能化。例如未来的度光谱目标检测算法可以拥有环境感知调节功能,能够根据环境变化,如光照条件、天气情况等自动调整参数,以优化检测效果,并且可以自适应地完成模态选择和融合。例如,智能化的算法可以根据光照条件选择红外外光图像进行特征互补,根据地形条件使用 3D 雷达图像进行特征互补,以提高检测的准确性和鲁棒性,并可以智能化地集成反馈机制,根据前一阶段的检测结果调整自身,以提高未来检测的准确性。

综上所述,多光谱目标检测在未来有望在数据融合、深度学习、实时监测和自动化、自适应和智能化及应用拓展等方面取得进一步的发展,为各个领域的目标识别和监测提供更准确、高效和全面的解决方案。

3 结论

本文综述了多光谱目标检测技术,重点介绍了主要的多光谱目标检测算法,并详细介绍了多光谱目标检测算法在行人检测以及其他领域的应用。多光谱目标检测算法目前已经取得了显著的进展,并且在交通、农业、环境等领域得到了具体应用,可以有效解决由于光照或环境条件导致的图像质量问题,但仍旧处在较为早期的阶段,在数据获取、数据处理、算法迁移等方面还存在或多或少的问题。随着计算机视觉领域各种技术的进步,多光谱目标检测算法本身将得到快速发展,并且将快速应用到各种领域,是目标检测算法未来的重要发展方向之一。

参考文献

- [1] ZENG L, WARDLOW B D, XIANG D, et al. A review of vegetation phenological metrics extraction using time-series, multi-spectral satellite data[J/OL]. Remote Sensing of Environment, 2020, 237[2023-06-03]. <https://www.sciencedirect.com/science/article/pii/S0034425719305309?via%3Dihub>. DOI: 10.1016/j.rse.2019.111511.

- [2] HOPKINSON C, CHASMER L, GYNAN C, et al. Multisensor and multispectral LiDAR characterization and classification of a forest environment[J]. *Canadian Journal of Remote Sensing*, 2016, 42(5): 501 – 520.
- [3] YUAN K, ZHUANG X, SCHAEFER G, et al. Deep-learning-based multispectral satellite image segmentation for water body detection[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2021, 14: 7422 – 7434.
- [4] HARTZELL P, GLENNIE C, BIBER K, et al. Application of multispectral LiDAR to automated virtual outcrop geology[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2014, 88: 147 – 155.
- [5] HWANG S, PARK J, KIM N, et al. Multispectral pedestrian detection: Benchmark dataset and baseline[C]//*IEEE Conference on Computer Vision and Pattern Recognition*. Piscataway, USA: IEEE, 2015: 1037 – 1045.
- [6] RAZAKARIVONY S, JURIE F. Vehicle detection in aerial imagery: A small target detection benchmark[J]. *Journal of Visual Communication and Image Representation*, 2016, 34: 187 – 203.
- [7] JIA X, ZHU C, LI M, et al. LLVIP: A visible-infrared paired dataset for low-light vision[C]//*IEEE/CVF International Conference on Computer Vision*. Piscataway, USA: IEEE, 2021: 3496 – 3504.
- [8] LIU J, FAN X, HUANG Z, et al. Target-aware dual adversarial learning and a multi-scenario multi-modality benchmark to fuse infrared and visible for object detection[C]//*IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Piscataway, USA: IEEE, 2022: 5802 – 5811.
- [9] YADAV A K, ROY R, KUMAR R, et al. Algorithm for de-noising of color images based on median filter[C]//*Third International Conference on Image Information Processing*. Piscataway, USA: IEEE, 2015: 428 – 432.
- [10] PENG H, RAO R, DIANAT S A. Multispectral image denoising with optimized vector bilateral filter[J]. *IEEE Transactions on Image Processing*, 2013, 23(1): 264 – 273.
- [11] HASHIMOTO N, MURAKAMI Y, BAUTISTA P A, et al. Multispectral image enhancement for effective visualization[J]. *Optics Express*, 2011, 19(10): 9315 – 9329.
- [12] YI C, ZHAO Y Q, CHAN J C W. Spectral super-resolution for multispectral image based on spectral improvement strategy and spatial preservation strategy[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2019, 57(11): 9010 – 9024.
- [13] MÜLLER M U, EKHTIARI N, ALMEIDA R M, et al. Super-resolution of multispectral satellite images using convolutional neural networks[J]. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2020, V-1-2020: 33 – 40.
- [14] AMZIANE A, LOSSON O, MATHON B, et al. MSFA-Net: A convolutional neural network based on multispectral filter arrays for texture feature extraction[J]. *Pattern Recognition Letters*, 2023, 168: 93 – 99.
- [15] WANG R, ZHANG W, SHI Y, et al. GA-ORB: A new efficient feature extraction algorithm for multispectral images based on geometric algebra[J]. *IEEE Access*, 2019, 7: 71235 – 71244.
- [16] WAGNER J, FISCHER V, HERMAN M, et al. Multispectral pedestrian detection using deep fusion convolutional neural networks[C/OL]//*24th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*. 2016, 587: 509 – 514[2023 – 06 – 19]. <https://www.esann.org/sites/default/files/proceedings/legacy/es2016 – 118. pdf>.
- [17] LIU J, ZHANG S, WANG S, et al. Multispectral deep neural networks for pedestrian detection[EB/OL]. (2016 – 11 – 08)[2023 – 06 – 25]. <https://arxiv.org/abs/1611.02644v1>. DOI: 10.48550/arXiv.1611.02644.
- [18] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//*IEEE Conference on Computer Vision and Pattern Recognition*. Piscataway, USA: IEEE, 2014: 580 – 587.
- [19] Girshick R. Fast R-CNN[C]//*IEEE International Conference on Computer Vision*. Piscataway, USA: IEEE, 2015: 1440 – 1448.
- [20] 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.
- [21] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]//*IEEE Conference on Computer Vision and Pattern Recognition*. Piscataway, USA: IEEE, 2016: 779 – 788.
- [22] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector[C]//*14th European Conference on Computer Vision*. Berlin, Germany: Springer, 2016: 21 – 37.
- [23] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//*31st International Conference on Neural Information Processing Systems*. New York, USA: ACM, 2017: 6000 – 6010.

- [24] CARION N, MASSA F, SYNNAEVE G, et al. End-to-end object detection with transformers[C]//16th European Conference on Computer Vision. Berlin, Germany: Springer, 2020: 213 – 229.
- [25] QINGYUN F, DAPENG H, ZHAOKUI W. Cross-modality fusion transformer for multispectral object detection[EB/OL]. (2021 – 10 – 30)[2023 – 06 – 08]. <https://arxiv.org/abs/2111.00273>. DOI: 10.48550/arXiv.2111.00273.
- [26] WANG X, LI S, CHEN C, et al. Data-level recombination and lightweight fusion scheme for RGB-D salient object detection[J]. IEEE Transactions on Image Processing, 2020, 30: 458 – 471.
- [27] ZHANG H, FROMONT E, LEFEVRE S, et al. Multispectral fusion for object detection with cyclic fuse-and-refine blocks[C]//IEEE International Conference on Image Processing. Piscataway, USA: IEEE, 2020: 276 – 280.
- [28] XIAO G, BAVIRISETTI D P, LIU G, et al. Decision-level image fusion[J]. Image Fusion, 2020: 149 – 170.
- [29] CHOI E J, PARK D J. Human detection using image fusion of thermal and visible image with new joint bilateral filter[C]//5th International Conference on Computer Sciences and Convergence Information Technology. Piscataway, USA: IEEE, 2010: 882 – 885.
- [30] CHOI H, KIM S, PARK K, et al. Multi-spectral pedestrian detection based on accumulated object proposal with fully convolutional networks[C]//23rd International Conference on Pattern Recognition. Piscataway, USA: IEEE, 2016: 621 – 626.
- [31] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C]//IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2015: 3431 – 3440.
- [32] KONIG D, ADAM M, JARVERS C, et al. Fully convolutional region proposal networks for multispectral person detection[C]//IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2017: 49 – 56.
- [33] GONG M, NIU X, ZHANG P, et al. Generative adversarial networks for change detection in multispectral imagery[J]. IEEE Geoscience and Remote Sensing Letters, 2017, 14(12): 2310 – 2314.
- [34] PARK K, KIM S, SOHN K. Unified multi-spectral pedestrian detection based on probabilistic fusion networks[J]. Pattern Recognition, 2018, 80: 143 – 155.
- [35] LI C, SONG D, TONG R, et al. Multispectral pedestrian detection via simultaneous detection and segmentation[EB/OL]. (2018 – 08 – 14)[2023 – 07 – 05]. <https://arxiv.org/abs/1808.04818>. DOI: 10.48550/arXiv.1808.04818.
- [36] GUAN D, CAO Y, YANG J, et al. Fusion of multispectral data through illumination-aware deep neural networks for pedestrian detection[J]. Information Fusion, 2019, 50: 148 – 157.
- [37] LI C, SONG D, TONG R, et al. Illumination-aware faster R-CNN for robust multispectral pedestrian detection[J]. Pattern Recognition, 2019, 85: 161 – 171.
- [38] ZHANG L, LIU Z, ZHANG S, et al. Cross-modality interactive attention network for multispectral pedestrian detection[J]. Information Fusion, 2019, 50: 20 – 29.
- [39] ZHANG L, ZHU X, CHEN X, et al. Weakly aligned cross-modal learning for multispectral pedestrian detection[C]//IEEE/CVF International Conference on Computer Vision. Piscataway, USA: IEEE, 2019: 5127 – 5137.
- [40] ZHANG Y, YIN Z, NIE L, et al. Attention based multi-layer fusion of multispectral images for pedestrian detection[J]. IEEE Access, 2020, 8: 165071 – 165084.
- [41] ZHANG H, FROMONT E, LEFEVRE S, et al. Multispectral fusion for object detection with cyclic fuse-and-refine blocks[C]//IEEE International Conference on Image Processing. Piscataway, USA: IEEE, 2020: 276 – 280.
- [42] Zhou K, Chen L, Cao X. Improving multispectral pedestrian detection by addressing modality imbalance problems[C]//16th European Conference on Computer Vision. Berlin, Germany: Springer, 2020: 787 – 803.
- [43] WOLPERT A, TEUTSCH M, SARFRAZ M S, et al. Anchor-free small-scale multispectral pedestrian detection[EB/OL]. (2020 – 08 – 20)[2023 – 06 – 11]. <https://arxiv.org/abs/2008.08418v1>. DOI: 10.48550/arXiv.2008.08418.
- [44] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16 × 16 words; Transformers for image recognition at scale[EB/OL]. (2021 – 06 – 03)[2023 – 06 – 21]. <https://arxiv.org/abs/2010.11929v1>.
- [45] FANG Q Y, HAN D P, WANG Z K. Cross-modality fusion transformer for multispectral object detection[EB/OL]. (2022 – 10 – 04)[2023 – 06 – 07]. <https://arxiv.org/abs/2111.00273>. DOI: 10.48550/arXiv.2111.00273.
- [46] ZHANG H, FROMONT E, LEFEVRE S, et al. Guided attentive feature fusion for multispectral pedestrian detection[C]//IEEE/CVF Winter Conference on Applications of Computer Vision. Piscataway, USA: IEEE, 2021: 72 – 80.
- [47] KIM J, KIM H, KIM T, et al. MLPD: Multi-label pedestrian detector in multispectral domain[J]. IEEE Robotics and Automa-

- tion Letters, 2021, 6(4): 7846 – 7853.
- [48] CAO Y, LUO X, YANG J, et al. Locality guided cross-modal feature aggregation and pixel-level fusion for multispectral pedestrian detection[J]. Information Fusion, 2022, 88: 1 – 11.
- [49] LIU J, FAN X, HUANG Z, et al. Target-aware dual adversarial learning and a multi-scenario multi-modality benchmark to fuse infrared and visible for object detection[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2022: 5802 – 5811.
- [50] KIM J U, PARK S, RO Y M. Uncertainty-guided cross-modal learning for robust multispectral pedestrian detection[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2021, 32(3): 1510 – 1523.
- [51] ZHANG Y, YU H, HE Y, et al. Illumination-guided RGBT object detection with inter-and intra-modality fusion[J]. IEEE Transactions on Instrumentation and Measurement, 2023, 72: 1 – 13.
- [52] LI R, XIANG J, SUN F, et al. Multiscale cross-modal homogeneity enhancement and confidence-aware fusion for multispectral pedestrian detection[J]. IEEE Transactions on Multimedia, 2023, 26: 852 – 863.
- [53] XIE Y, ZHANG L, YU X, et al. YOLO-MS: Multispectral object detection via feature interaction and self-attention guided fusion[J]. IEEE Transactions on Cognitive and Developmental Systems, 2023, 15(4): 2132 – 2143.
- [54] ZHANG X, ZHANG X, SHENG Z, et al. TFDet: Target-aware fusion for RGB-T pedestrian detection[EB/OL]. (2023 – 10 – 18)[2023 – 06 – 15]. <https://arxiv.org/abs/2305.16580>. DOI: 10.48550/arXiv.2305.16580.
- [55] LI J, CHEN L, HUANG W, et al. Multispectral detection of skin defects of bi-colored peaches based on vis-NIR hyperspectral imaging[J]. Postharvest Biology and Technology, 2016, 112: 121 – 133.
- [56] ZHOU X, ZHENG H B, XU X Q, et al. Predicting grain yield in rice using multi-temporal vegetation indices from UAV-based multispectral and digital imagery[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2017, 130: 246 – 255.
- [57] De SILVA M, Brown D. Plant disease detection using vision transformers on multispectral natural environment images[C/OL]//2023 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems. Piscataway, USA: IEEE, 2023: [2023 – 07 – 01]. <https://ieeexplore.ieee.org/document/10220517>. DOI: 10.1109/icABCD59051.2023.10220517.
- [58] THANH NOI P, KAPPAS M. Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using Sentinel-2 imagery[J/OL]. Sensors, 2017, 18(1)[2023 – 06 – 19]. <https://pubmed.ncbi.nlm.nih.gov/29271909>. DOI: 10.3390/s18010018.
- [59] YUAN K, ZHUANG X, SCHAEFER G, et al. Deep-learning-based multispectral satellite image segmentation for water body detection[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2021, 14: 7422 – 7434.
- [60] CHOI S, KIM S, PARK K, et al. Multi-spectral vehicle detection based on convolutional neural network[J]. Journal of Korea Multimedia Society, 2016, 19(12): 1909 – 1918.

作者简介

王元喆(2000 –), 男, 硕士生。研究领域为多光谱目标检测。

梁腾飞(1996 –), 男, 博士生。研究领域为多模态目标重识别。

曾宇乔(1998 –), 男, 博士生。研究领域为多模态融合、图像去噪。

金 一(1982 –), 女, 博士, 教授, 博士生导师。研究领域为模式识别, 机器学习等。

李浥东(1982 –), 男, 博士, 教授, 博士生导师。研究领域为计算机技术、高性能计算等。