

〈图像处理与仿真〉

基于红外与可见光图像的目标检测算法

邝楚文¹, 何 望²

(1. 惠州经济职业技术学院, 广东 惠州 516057; 2. 华中科技大学 计算机科学与技术学院, 湖北 武汉 430074)

摘要: 针对现有基于可见光的目标检测算法存在的不足, 提出了一种红外和可见光图像融合的目标检测方法。该方法将深度可分离卷积与残差结构相结合, 构建并列的高效率特征提取网络, 分别提取红外和可见光图像目标信息; 同时, 引入自适应特征融合模块以自主学习的方式融合两支路对应尺度的特征, 使两类图像信息互补; 最后, 利用特征金字塔结构将深层特征逐层与浅层融合, 提升网络对不同尺度目标的检测精度。实验结果表明, 所提网络能够充分融合红外和可见光图像中的有效信息, 并在保障精度与效率的前提下实现目标识别与定位; 同时, 在实际变电站设备检测场景中, 该网络也体现出较好的鲁棒性和泛化能力, 可以高效完成检测任务。

关键词: 目标检测; 红外与可见光图像; 深度学习; 自适应融合

中图分类号: TP391.41 文献标志码: A 文章编号: 1001-8891(2022)09-0912-08

Object Detection Algorithm Based on Infrared and Visible Light Images

KUANG Chuwen¹, HE Wang²

(1. *Huizhou Economics and Polytechnic College, Huizhou 516057, China*;

2. *School of Computer Science & Technology, Huazhong University of Science and Technology, Wuhan 430074, China*)

Abstract: A target detection method based on infrared and visible image fusion is proposed to overcome the shortcomings of the existing target detection algorithms based on visible light. In this method, depth separable convolution and the residual structure are combined to construct a parallel high-efficiency feature extraction network to extract the object information of infrared and visible images, respectively. Simultaneously, the adaptive feature fusion module is introduced to fuse the features of the corresponding scales of the two branches through autonomous learning such that the two types of image information are complementary. Finally, the deep and shallow features are fused layer by layer using the feature pyramid structure to improve the detection accuracy of different scale targets. Experimental results show that the proposed network can completely integrate the effective information in infrared and optical images and realize target recognition and location on the premise of ensuring accuracy and efficiency. Moreover, in the actual substation equipment detection scene, the network shows good robustness and generalization ability and can efficiently complete the detection task.

Key words: object detection, infrared and visible light image, deep learning, adaptive fusion

0 引言

图像目标识别检测技术作为数字图像处理和模式识别领域中的一个重要分支, 广泛地应用于自动驾驶、医疗图像、工业检测、智能机器人、智能视频监

控等众多场景^[1-2]。目标检测是通过分析目标特征信息, 对视频或图像中感兴趣目标区域进行定位提取, 并准确识别出各个区域的目标类别及其对应的包围框^[3]。近年来, 随着深度学习在图像处理领域的不断突破, 目标检测技术也得到了长足的进步^[4]。目前,

收稿日期: 2021-11-29; 修订日期: 2022-01-28。

作者简介: 邝楚文 (1984-), 男, 汉族, 广东珠海人, 讲师, 研究方向: 计算机科学与技术, 人工智能。E-mail: 1952707159@qq.com。

基金项目: 国家自然科学基金项目 (61972169)。

大多数目标检测算法主要基于可见光图像,可见光图像虽然包含丰富的纹理和细节信息,但通常实际场景中各目标所处环境错综复杂,造成目标存在遮挡、尺度变化范围大、光照不均以及噪声干扰等情况,使目标检测技术的落地仍存在较大挑战^[5]。而红外图像主要利用热辐射能量成像,受光照影响较少,但图像对比度低,目标纹理结构等特征损失严重,较大地限制了其在目标检测领域中的应用。因此,研究一种基于可见光与红外图像相结合的目标检测方法,可以有效实现性能互补,同时降低光照、雨雾等噪声干扰,对目标检测技术的发展以及实际场景的应用都有较大的促进作用。

传统的目标检测算法流程主要为图像预处理、候选框提取、特征提取、目标分类以及后处理几个步骤,尽管在各个领域得到了广泛的应用,但存在候选框众多、特征设计复杂、算法迁移性差等问题^[5]。为了缓解传统算法的弊端,研究者将深度学习方法应用于目标检测,通过端到端的训练方式,使目标检测精度得到大幅提升。尽管目前大多数目标检测研究工作是基于可见光图像,但也有部分研究者对可见光和红外图像融合检测进行了探索。郝永平等^[6]人利用双通道深度残差卷积网络分别对可见光和红外图像进行特征提取,并通过计算特征欧氏距离结合注意力机制实现对感兴趣区域的目标检测。李舒涵等^[7]人提出了一种基于红外与可见光融合的交通标志检测方法,通过对两种图像中的目标进行粗定位后再结合卷积网络对定位框进行决策融合,实现交通标志高效检测。Xiao等^[8]人利用差分最大损失函数指导红外和可见光两个支路的卷积网络提取目标特征,并设计特征增强和级联语义扩展模块提升对不同尺度目标的检测。Banuls等^[9]人提出了一种基于决策级融合的目标检测算法,

利用改进 YOLOv3 网络对可见光和红外图像进行分别检测后再进行加权融合提升目标检测效果。可见,基于深度学习的红外与可见光融合的检测方法可以有效提升目标检测效果,但大多数方法采用分别提取特征后再融合检测,未能充分利用两类图像中目标特征进行信息互补。

针对上述情况,本文在结合前人研究成果的基础上,提出了一种基于深度可分离卷积的特征交叉融合目标检测网络模型。为了网络保障效率,该模型以深度可分离卷积作为基本特征提取单元,构建轻量级的双支路特征提取结构,分别提取可见光和红外图像中目标信息。同时,利用双支路特征交叉融合结构,对各阶段提取的特征进行充分融合互补,并结合多层特征跳层融合实现不同尺度目标高精度识别检测。

1 目标检测网络

1.1 整体框架

本文所设计的红外可见光融合目标检测网络整体结构如图 1 所示。该网络由特征提取模块、特征融合模块以及检测模块 3 部分构成。针对红外和可见光图像,特征提取部分由两个并列的相同结构支路构成,主要采用深度可分离卷积作为特征提取基本单元,结合 LeakyReLU 激活层、最大池化层、上采样等操作,对红外和可见光图像特征信息由浅到深的高效提取。特征融合模块通过线性组合方式对两支路池化层特征进行建模,利用自主学习方式实现红外与可见光信息共享,使两支路提取的特征进行互补,提升网络特征多样性。检测模块利用多个不同尺度的深层特征以逐层上采样融合的操作构建特征金字塔预测结构,使网络在不同尺度上都具有强语义信息,保障网络对不同尺度的目标实现准确检测。

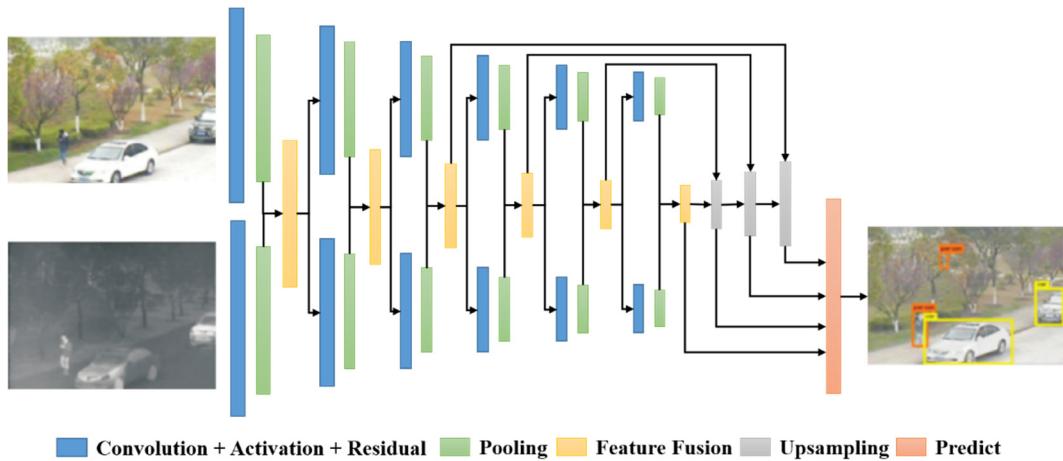


图 1 双支路自适应目标检测网络整体结构

Fig.1 The overall structure of the dual-branch adaptive object detection network

1.2 特征提取模块

特征提取作为目标检测的首要任务,其提取的特征好坏直接决定目标检测模型的优劣。对于传统目标检测而言,特征主要靠人工设计,如SIFT(Scale Invariant Feature Transform)、HOG(Histogram of Oriented Gradient)、Haar、DPM(Deformable Part Model)等^[10],通过提取滑动窗口中相应特征并利用机器学习进行分类,实现目标检测。而基于深度学习的目标检测方法扩大了特征提取的范畴,利用端到端的训练学习方式自动学习目标特征,避免了人工设计特征的局限。因此,基于深度学习的检测算法通常能够获得比传统方法更优的检测效果。基于此,本文借鉴目前经典的深度学习网络,设计了适用于红外和可见光图像的并列双支路特征提取网络。

为有效提取图像中各目标浅层和深层特征,本文所构建的特征提取结构采用多个不同特征尺度的子模块串联堆叠构成,详细结构如表1所示。

表1 特征提取结构

Table 1 Feature extraction structure

Stage	Layer structure	Repetitions	Output size
Original	RGB, 3	1	448×448
	Conv 3×3, 13	1	224×224
Init	Max pooling 2×2, 3	1	112×112
	DWconv 3×3, 32	1	112×112
Stage1	Residual	1	112×112
	DWconv 3×3, 64	2	56×56
Stage2	Residual	2	56×56
	DWconv 3×3, 128	4	28×28
Stage3	Residual	4	28×28
	DWconv 3×3, 256	4	14×14
Stage4	Residual	4	14×14
	DWconv 3×3, 512	2	7×7
Stage5	Residual		

特征提取结构由init模块和多个stage模块组成,如图2所示。init模块如图2(a)所示,该结构主要对原图进行特征预处理,采用步长为2的并列卷积和池化两条支路提取目标显著特征,降低图像维度的同时也过滤了部分噪声,保障后续结构对特征的深入提取。stage模块如图2(b)所示,主要利用卷积层、激活层以残差结构方式构建,不同stage之间通过步长为2的2×2池化操作进行降维。由于需要分别对红外和

可见光图像进行特征提取,为避免网络计算量过大,stage模块采用深度可分离卷积替代传统卷积来提取特征,有效减少了网络参数并降低了计算量。深度可分离卷积相对于传统卷积虽然提取的特征信息有所降低,但双支路的信息融合可以较好地弥补特征缺失等问题。同时,stage模块引入残差结构来避免网络层数过深造成训练时出现梯度消失、梯度爆炸等问题,并以LeakyReLU作为激活函数,如式(1)所示,降低神经元“坏死”概率,使网络更快收敛。

$$f(x) = \begin{cases} \alpha x & x < 0 \\ x & x \geq 0 \end{cases} \quad (1)$$

式中: α 是偏移量,为一个较小数值的超参数,默认设置为0.02。由LeakyReLU公式可以看出,当输入小于零时,函数的输出不为零,求导之后导数为固定值,从而避免进入负区间神经元不学习的问题。

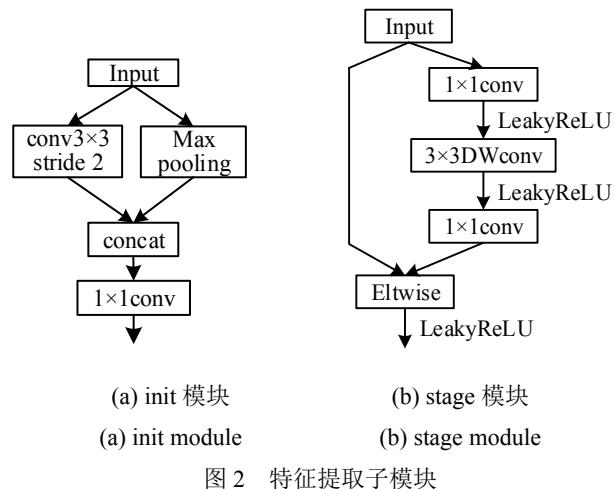


图2 特征提取子模块

Fig.2 Feature extraction submodules

1.3 特征融合模块

通常,可见光图像包含丰富的颜色、纹理等信息,能够提供较多的细节信息,但容易受到光照强度、天气等影响;而红外图像利用目标的热辐射能量大小进行成像,不受光照影响,但图像对比度较低,会损失部分目标的纹理、结构等外观特征。因此,通过融合红外和可见光图像信息,可以更好地增强和发现目标^[11-12]。基于此,本文在特征提取结构基础上,设计了特征交叉融合模块,使提取的红外和可见光图像信息互补,融合结构如图3所示。

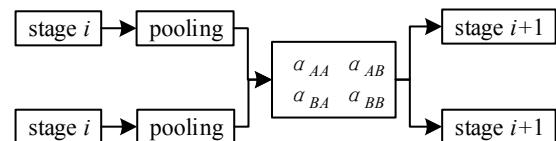


图3 特征融合模块

Fig.3 Feature fusion modules

融合结构考虑到网络运行效率, 主要将特征提取过程中每个尺度的最后一层进行融合, 即 init 模块最后一层和 stage 模块之间的池化层(图 3 中未画出 init 层, i 取值为 1,2,3,4)。同时, 融合结构采用自主学习的线性加权方式来替代特征信息直接相加, 避免融合时引入过多噪声。具体融合计算过程如式(2)所示。

$$\begin{bmatrix} y_A^{ij} \\ y_B^{ij} \end{bmatrix} = \begin{bmatrix} \alpha_{AA} & \alpha_{AB} \\ \alpha_{BA} & \alpha_{BB} \end{bmatrix} \begin{bmatrix} x_A^{ij} \\ x_B^{ij} \end{bmatrix} \quad (2)$$

自主学习过程中权重迭代计算如式(3)(4)所示:

$$\begin{bmatrix} \frac{\partial L}{\partial x_A^{ij}} \\ \frac{\partial L}{\partial x_B^{ij}} \end{bmatrix} = \begin{bmatrix} \alpha_{AA} & \alpha_{BA} \\ \alpha_{AB} & \alpha_{BB} \end{bmatrix} \begin{bmatrix} \frac{\partial L}{\partial y_A^{ij}} \\ \frac{\partial L}{\partial y_B^{ij}} \end{bmatrix} \quad (3)$$

$$\frac{\partial L}{\partial \alpha_{AA}} = \frac{\partial L}{\partial y_A^{ij}} x_A^{ij}, \quad \frac{\partial L}{\partial \alpha_{AB}} = \frac{\partial L}{\partial y_B^{ij}} x_A^{ij} \quad (4)$$

式中: x_A^{ij} 为待融合可见光特征图上坐标位置为 (i, j) 的像素值; 同理, x_B^{ij} 为待融合红外特征图上坐标位置为 (i, j) 的像素值; $\alpha_{AA}, \alpha_{AB}, \alpha_{BA}, \alpha_{BB}$ 为待学习的权重; y 为融合后的对应特征图位置的输出, 表示利用误差反向传播更新权重。由式(2)可以看出, 当 $\alpha_{AA} = \alpha_{BB} = 1, \alpha_{AB} = \alpha_{BA} = 0$ 时, 表明该层特征信息并不融合, 通过自主学习的方式决定红外和可见光图像特征信息的融合程度, 进而达到最优组合。

1.4 检测模块

对于基于深度学习的目标检测算法而言, 目前的检测思路可根据预设框大致分为两类: Anchor Based 和 Anchor-free 检测模型^[4]。Anchor Based 需要事先预设多个包围框, 预测时通过微调包围框实现检测。Anchor Based 又分为 Two-stage 和 One-stage 检测模型, Two-stage 模型检测思路主要分为 3 个部分: 候选区域生成、基于卷积神经网络的特征提取和目标定位分类, 典型的网络有 R-CNN^[13]系列网络; One-stage 模型则舍弃了候选框的生成过程, 直接在特征图上对预设框进行微调实现检测, 典型的网络有 YOLO(You Only Look Once) 系列^[14]、SSD (Single Shot MultiBox Detector)^[15]等网络。而 Anchor-free 模型则舍弃所有包围框, 直接根据提取的特征信息对目标进行定位识别, 典型网络有 FCOS (Fully Convolutional One-Stage)^[16]、LSNet (Light-Weight Siamese Network)^[17]等网络。考虑到不同检测模型的效率和精度, 本文基于 One-stage 模型设计了特征金字塔检测模块, 如图 4 所示。

检测模块以特征融合结构的输出作为输入, 由于特征融合的输出为红外和可见光两个通道, 故采用

concat 操作对两路特征拼接后作为检测输入, 即图 4 中 fuse 部分。同时, 考虑到现实场景中各目标大小差异十分明显, 本文采用多个不同维度的融合特征(fuse2~fuse5), 以自上而下的方式构建特征金字塔检测结构, 该结构首先将融合后的深层特征通过点卷积调整特征通道与浅层特征一致; 然后, 上采样至浅层特征尺度大小, 并与浅层特征拼接; 再将拼接后的特征信息进行卷积操作, 充分融合深层特征信息; 将 fuse5~fuse2 的特征依次重复上述上采样拼接融合操作, 使检测模块能充分获取全局以及局部特征信息; 最后, 分别利用 4 个尺度的特征对目标类别和位置进行预测, 并将预测结果通过极大值抑制算法筛选出最优目标检测框, 进而实现对各目标的精准检测。而对于预设框的设置, 利用 k-means 算法对训练集中的标注框利用聚合, 自动生成一组适用于对应场景的预设框。

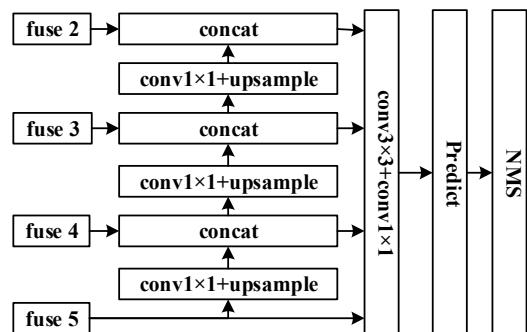


图 4 金字塔检测结构

Fig.4 Pyramid detection structure

2 实验与结果分析

为有效验证本文所提的方法, 实验分别采用搭载 Titan Xp 的高性能电脑以及 NVIDIA Jetson TX2 嵌入式平台进行实验。网络采用 caffe 深度学习框架进行构建, 通过公开数据集和实际变电站数据集分别对网络的可行性和实际落地性进行测试。为方便与其他同类网络对比, 网络训练时的超参数主要参照文献[8-9,14]进行设置: 采用小批次梯度下降方法(mini batch Stochastic Gradient Descent)优化网络参数, 并利用 Momentum 动量算法加速优化, 动量参数设置为 0.9; 初始学习率设置为 0.001, 学习策略采用 step 模式, 即迭代训练 3 万次之后每迭代 1 万次学习率降低十分之一; 权值衰减系数设为 0.0005, 防止过拟合; 对于损失函数, 由于目标检测是多任务模型, 因此, 训练时的损失为分类损失、置信度损失以及回归框损失的叠加, 具体的损失函数参考 YOLO 网络^[14]计算方式。

对于所提网络的精度及效率评价分别采用均值平均精度(mAP)和网络每秒处理图像数量(FPS)来

评价。均值平均精度指各个类别的平均精度的平均值,而针对目标尺寸不同,又分为 mAP_s 、 mAP_m 、 mAP_l 来衡量网络对小中大目标的检测效果,以边界框面积 32^2 和 96^2 作为区分边界。

$$Precession_C = \frac{N(\text{True Positives})_C}{N(\text{Total Objects})_C} \quad (5)$$

$$AP_C = \frac{\sum \text{Precision}_C}{N(\text{Total Image})_C} \quad (6)$$

$$mAP = \frac{\sum AP_C}{N(\text{Classes})} \quad (7)$$

$$FPS = N / \sum_k^N T_k \quad (8)$$

式(5)(6)(7)中: C 为类别; $N(\text{True Positives})_C$ 指一张图像中的 C 类目标正确预测的数量; $N(\text{Total Object})_C$ 指该图像中 C 类目标总数; Precision_C 指该图像中 C 类目标检测精度; $N(\text{Total Image})_C$ 指训练集中含 C 类目标的图像数量; $N(\text{Classes})$ 表示样本中类别数, AP_C 指所有图像中 C 类目标的平均精度。式(8)为效率评价指标: N 为图像数量, T_k 指网络处理第 k 张图像所消耗的时间。

2.1 标准数据集测试

为有效验证所提方法的可行性,实验采用李成龙教授团队构建的 RGBT210 数据集^[18]作为标准数据进行测试。RGBT210 数据集由具有相同的成像参数红外和可见光摄像机在 210 种场景下采集的图像构成。该数据集包含约 21 万张图像,涵盖约 20 多种目标在不同时间段、不同光照强度下的红外和可见光图像对。由于该数据集较大,且大多数图像相似,为了能快速验证所提网络,从中筛选出了 5000 张相似度较低的图像进行测试。筛选出的图像包含汽车、行人、自行车、狗、风筝等 10 种类别,将其归一化图像尺寸为 512×448 ,并以 7:1:2 的比例构建训练、验证及测试集,在搭载 Titan Xp 的电脑上进行训练测试。

实验首先对所提网络的单条支路进行测试,即删

除融合模块和红外支路,只利用可见光支路进行目标检测。网络训练时设置 batchsize 为 4, 经过约 10 万次迭代收敛后,与同类型的目标检测网络对比如表 2 所示。

表 2 可见光网络测试结果对比

Table 2 Comparison of visible network test results

Network	FPS	Test accuracy/ (%)			
		mAP	mAP_s	mAP_m	mAP_l
Faster RCNN ^[13]	25	71.9	52.2	73.2	82.6
YOLO ^[14]	66	67.5	49.1	70.6	78.9
Shuffle+SSD ^[15]	113	65.6	46.1	68.3	75.3
Visible-light branch	93	67.3	48.5	70.1	77.6

由表 2 可以看出,本文所构建的可见光单支路网络与目前主流的高精度 (Faster RCNN) 和高效率 (Shuffle+SSD) 目标检测网络相比,较好地平衡了网络性能。同时,网络以深度可分离卷积替代传统卷积,并引用了残差、LeakyReLU 激活函数等网络构建策略,与同类型网络 (YOLO) 相比,所提网络以较小的精度损失来换取网络效率大幅提升。但相对于 Faster RCNN,由于所提网络为单步检测,并且深度可分离卷积相对传统卷积损失了部分特征信息,所以造成精度有所降低。为进一步验证红外与可见光双支路结构的特征互补性以及所提特征融合结构的有效性,实验分别测试了可见光、红外以及融合后的网络性能,同时,针对两支路的融合结构,分别测试了直接相加 (Eltwise)、拼接 (Concat) 以及本文所提的自适应融合结构,实验结果如表 3 和图 5 所示。

表 3 不同结构测试结果

Table 3 Different dilation rates test results

Network	FPS	Test accuracy / (%)			
		mAP	mAP_s	mAP_m	mAP_l
Infrared branch	94	61.0	45.3	64.2	71.8
Visible-light branch	93	67.3	48.5	70.1	77.6
Eltwise Fusion	81	70.4	51.1	73.8	80.1
Concat Fusion	79	71.6	52.3	74.2	81.6
This paper	78	73.8	54.3	75.1	83.2





(a) 可见光原图 (b) 红外原图 (c) 可见光支路检测 (d) 红外支路检测 (e) 本文
(a) Visible light original image (b) Infrared original image (c) Visible light branch detection (d) Infrared branch detection (e) This paper

图 5 单支路与融合支路目标检测结果

Fig.5 Single branch and fusion branch object detection results

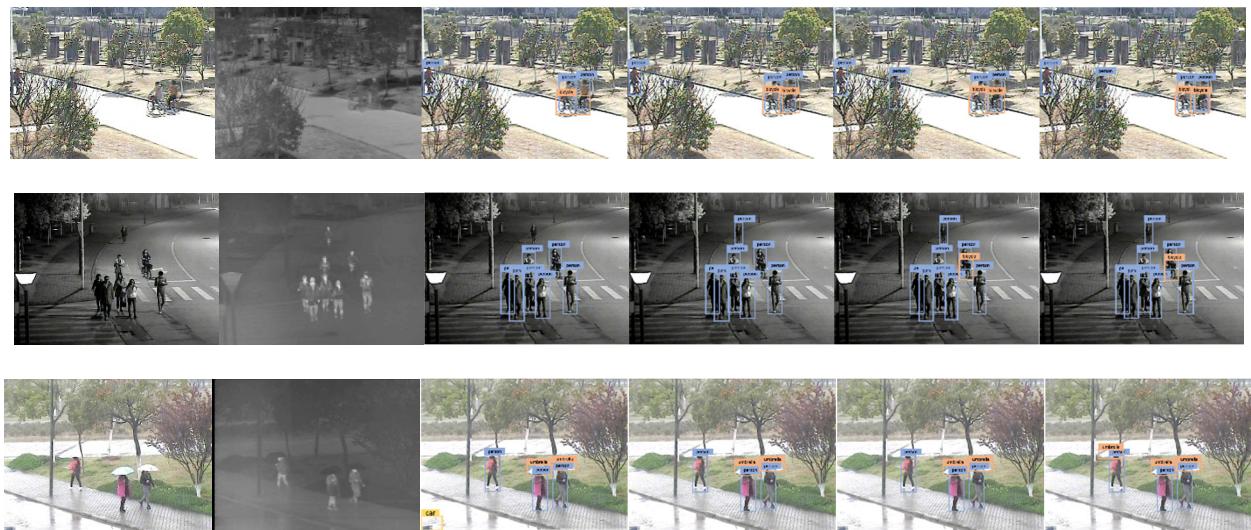
由表 3 和图 5 可以看出, 红外图像由于缺少较多细节信息, 与可见光图像相比, 目标检测效果较差; 而可见光受光照等影响, 也造成部分目标无法识别。通过两支路融合的方式, 使图像特征信息更加丰富, 网络检测精度也有较大提升, 但由于引入了新的支路, 网络效率有一定的下降。同时, 不同的融合方式, 对网络性能也有一定的影响, Eltwise 和 Concat 融合方式虽然增加了特征信息, 但也引入了较多噪声, 而本文所提方法较好地缓解了噪声的引入, 使两支路的有效信息能更好的互补, 检测精度也更高。为更好地验证所提网络的优势, 实验分别与基于多级分层检测^[7]、基于特征级融合^[8]以及基于决策级融合^[19]的红外与可见

光联合检测网络进行了测试对比, 实验结果如表 4 所示, 检测效果如图 6 所示。

表 4 同类型网络测试结果对比

Table 4 Comparison of test results of the same type of network

Network	FPS	Test accuracy(%)			
		AP	AP _s	AP _m	AP _l
Literature[7]	83	68.1	49.5	71.0	76.3
Literature [8]	51	74.0	54.0	76.5	84.1
Literature [19]	73	72.2	51.2	74.6	81.5
This paper	78	73.8	54.3	75.1	83.2



(a) 可见光图 (b) 红外图 (c) 文献[7] (d) 文献[8] (e) 文献[19] (f) 本文
(a) Visible light image (b) Infrared image (c) Literature [7] (d) Literature [8] (e) Literature [19] (f) This paper

图 6 红外-可见光网络检测效果对比

Fig.6 Comparison of infrared-visible network detection effects

根据上述结果可以看出, 与同类型的红外-可见光联合检测网络相比, 虽然融合方式提高目标的显著性, 但也引入了更多的噪声, 如图 6 文献[7]中错误检测出 car。同时, 对于复杂背景并且红外图像无法检测的目标, 各算法检测效果较差, 如图 6 第 3 排中雨伞的检测。尽管本文所提网络未能在所有指标中达到最

优, 但基本保持高精度和高效率的完成目标检测。同时, 由于自适应融合的方式可以较好地避免噪声的引入, 所以网络对浅层信息也进行了融合, 使网络能更好地识别检测出图像中的小目标, 并且对存在遮挡的目标也有较好的检测效果。

2.2 变电站场景测试

标准数据集的测试结果有效地验证了所提网络的可行性,为了进一步验证该网络的泛化能力以及实际落地性,本文采用实际变电站场景下的设备图像数据,对网络进行测试验证。在实际变电站场景中,大多数设备需要巡检机器人进行红外测温来监测设备状况,但由于实际环境复杂,如何避免其他设备干扰,准确定位出待测设备对机器人而言极为重要。基于此,本文以搭载 Jetson TX2 变电站巡检机器人为测试平台,通过机器人搭载的可见光和红外相机对设备图像进行采集,并根据两相机参数,将采集的红外和可见光图像进行裁剪使各目标对齐。实验主要采集了断路器、绝缘子、冷控箱、变压器等 6 种变电站设备图像,涵盖了不同大小的目标,从中筛选出 5000 张重

复率较低的图像，利用公开标注工具 LabelImg 对图像中的各个目标进行手动标注，并调整图像尺寸为 512×448 大小。将处理后的图像以 7:1:2 的比例构建训练、验证和测试集，在 Jetson TX2 平台上进行训练测试，测试结果如表 5 所示，检测效果如图 7 所示。

表 5 可见光网络测试结果对比

Table 5 Comparison of visible network test results

Network	FPS	Test accuracy /(%)			
		AP	AP _S	AP _m	AP _I
Literature[7]	21	73.2	54.1	76.6	81.4
Literature [8]	13	78.2	58.9	81.4	88.7
Literature [19]	17	76.9	56.2	78.3	86.6
This paper	20	78.1	59.1	80.8	88.4

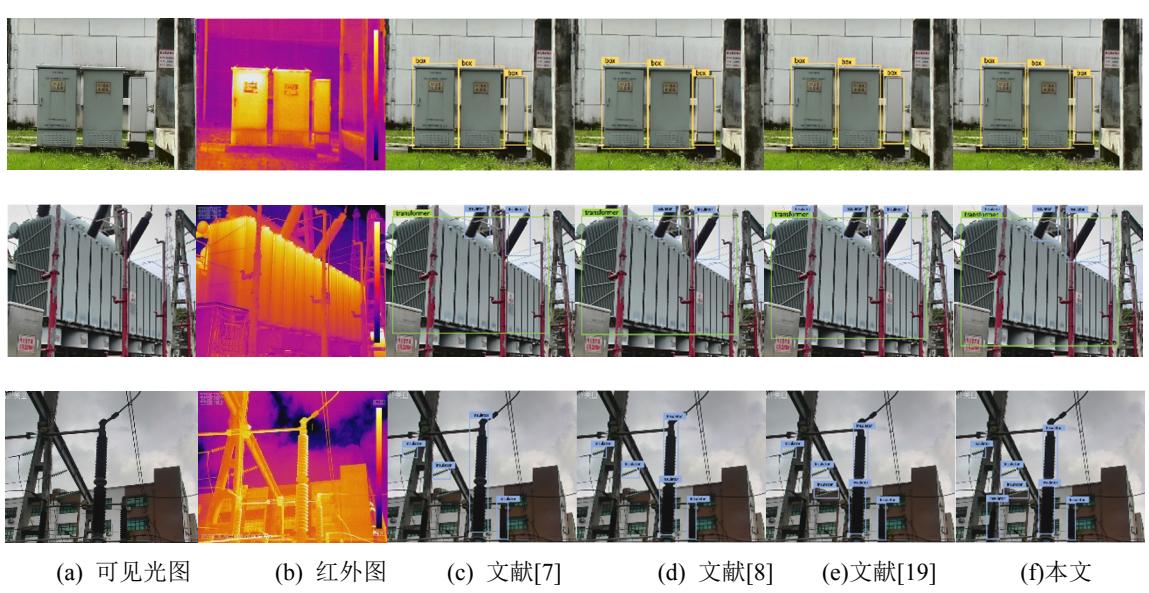


图 7 变电站设备检测效果对比

Fig.7 Comparison of substation equipment detection effects

根据实际场景中的实验结果可以看出,由于测试平台的计算性能较低,各网络的效率都相应地下降到较低水平,但总体来看,本文所提网络与同类型高效率网络相比,基本达到相同的检测效率。尽管效率仍相对较低,但相对于巡检速度较慢的机器人而言,也基本能满足其实时检测的需求。同时,在检测精度方面,由于实际场景中的数据相对于标准数据集较为简单,各方法在检测精度上都有所提升,对于简单的设备环境,各算法基本都能实现准确检测,如图 7 第一排中对变电箱的检测;但对于相对复杂的环境,如图 7 第二、三排中,存在目标过大、各目标相互影响、背景复杂的情况,使各算法检测效果受到较大影响。

但通过整体对比，可以看出本文所提网络的检测效果基本与高精度网络不相上下，并且可以达到高效率网络的检测效率。综上可见，本文所提方法具有更强的鲁棒性及泛化能力，可以很好地适用于巡检机器人对变电站设备的检测，使机器人能够保持高精度高效率的完成检测任务。

3 结论

本文针对目前单一可见光目标检测算法存在的局限,提出了一种融合红外和可见光图像特征信息的深度神经网络目标检测方法。该方法以深度可分离卷积为基本特征提取单元,结合高效率 LeakyReLU 激

活函数和残差结构构建并列的红外-可见光特征提取支路; 为有效融合红外和可见光图像中目标特征信息, 引入了自适应特征融合模块, 利用自主学习的方式对两支路中的有效信息进行融合, 避免了过多噪声的引入; 同时, 为保证网络对不同大小目标的准确检测, 采用特征金字塔结构方式融合不同维度的特征信息, 并利用多个尺度特征分别对目标进行预测, 提升了对不同大小目标的检测效果。通过在标准数据集以及实际变电站设备检测场景中的实验结果表明, 本文所提网络有效地平衡了检测精度与效率, 可以较好地应用于变电站巡检机器人完成设备检测任务。尽管本文所提网络在一定程度上提高了目标检测效果, 但仍有较多值得深入研究的地方, 后续将尝试引入注意力机制来进一步提升检测精度, 并利用模型剪枝或知识蒸馏等策略, 提高网络的检测效率。

参考文献:

[1] 孙怡峰, 吴疆, 黄严严, 等. 一种视频监控中基于航迹的运动小目标检测算法[J]. 电子与信息学报, 2019, 41(11): 2744-2751.
SUN Yifeng, WU Jiang, HUANG Yan, et al. A track based moving small target detection algorithm in video surveillance [J]. *Journal of Electronics and Information*, 2019, 41(11): 2744-2751.

[2] LIN C, LU J, GANG W, et al. Grainingness-aware deep feature learning for pedestrian detection[J]. *IEEE Transactions on Image Processing*, 2020, 29: 3820-3834.

[3] 范丽丽, 赵宏伟, 赵浩宇, 等. 基于深度卷积神经网络的目标检测研究综述[J]. 光学精密工程, 2020, 28(5): 161-173.
FAN Lili, ZHAO Hongwei, ZHAO Haoyu, et al. Overview of target detection based on deep convolution neural network[J]. *Optical Precision Engineering*, 2020, 28(5): 161-173.

[4] 赵永强, 饶元, 董世鹏, 等. 深度学习目标检测方法综述[J]. 中国图象图形学报, 2020, 28(4): 5-30.
ZHAO Yongqiang, RAO yuan, DONG Shipeng, et al. Overview of deep learning target detection methods[J]. *Chinese Journal of Image and Graphics*, 2020, 28(4): 5-30.

[5] 罗会兰, 彭珊, 陈鸿坤. 目标检测难点问题最新研究进展综述[J]. 计算机工程与应用, 2021, 57(5): 36-46.
LUO Huilan, PENG Shan, CHEN Hongkun. Overview of the latest research progress on difficult problems of target detection[J]. *Computer Engineering and Application*, 2021, 57(5): 36-46.

[6] 郝永平, 曹昭睿, 白帆, 等. 基于兴趣区域掩码卷积神经网络的红外-可见光图像融合与目标识别算法研究[J]. 光子学报, 2021, 50(2): 15-16.
HAO Yongping, CAO Zhaorui, BAI fan, et al. Research on infrared visible image fusion and target recognition algorithm based on region of interest mask convolution neural network[J]. *Acta Photonica Sinica*, 2021, 50(2): 15-16.

[7] 李舒涵, 许宏科, 武治宇. 基于红外与可见光图像融合的交通标志检测[J]. 现代电子技术, 2020, 43(3): 45-49.
LI Shuhan, XU Hongke, WU Zhiyu. Traffic sign detection based on infrared and visible image fusion [J]. *Modern Electronic Technology*, 2020, 43(3): 45-49.

[8] XIAO X, WANG B, MIAO L, et al. Infrared and visible image object detection via focused feature enhancement and cascaded semantic extension[J]. *Remote Sensing*, 2021, 13(13): 2538.

[9] Banuls A, Madow A, Vazquez-Martin R, et al. Object detection from thermal infrared and visible light cameras in search and rescue scenes[C]// 2020 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR). IEEE, 2020: 380-386.

[10] 李章维, 胡安顺, 王晓飞. 基于视觉的目标检测方法综述[J]. 计算机工程与应用, 2020, 56(8): 7-15.
LI Zhangwei, HU Anshun, WANG Xiaofei. Overview of vision based target detection methods[J]. *Computer Engineering and Application*, 2020, 56(8): 7-15.

[11] 汪廷. 红外图像与可见光图像融合研究与应用[D]. 西安: 西安理工大学, 2019.
WANG Ting. Research and Application of Infrared Image and Visible Image Fusion[D]. Xi'an: Xi'an University of Technology, 2019.

[12] XIANG X, LV N, YU Z, et al. Cross-modality person re-identification based on dual-path multi-branch network[J]. *IEEE Sensors Journal*, 2019, 19(23): 11706-11713.

[13] 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.

[14] Bochkovskiy A, WANG C Y, LIAO H. YOLOv4: Optimal speed and accuracy of object detection[J/OL]. *Arxiv Preprint Arxiv*, <https://arxiv.org/abs/2004.10934>.

[15] LIU W, Anguelov D, Erhan D, et al. SSD: Single Shot MultiBox Detector[C]// European Conference on Computer Vision, 2016: 21-37.

[16] TIAN Z, SHEN C, CHEN H, et al. FCOS: Fully convolutional one-stage object detection[C]// 2019 IEEE/CVF International Conference on Computer Vision (ICCV). IEEE, 2020: 9626-9635.

[17] DUAN K, XIE L, QI H, et al. Location-sensitive visual recognition with cross-IOU loss[J/OL]. *Arxiv Preprint Arxiv*, <https://arxiv.org/abs/2104.04899>.

[18] LI C, ZHAO N, LU Y, et al. Weighted sparse representation regularized graph learning for RGB-T object tracking[C]// ACM on Multimedia Conference. ACM, 2017: 1856-1864.

[19] 白玉, 侯志强, 刘晓义, 等. 基于可见光图像和红外图像决策级融合的目标检测算法[J]. 空军工程大学学报, 2020, 21(6): 53-59.
BAI Yu, HOU Zhiqiang, LIU Xiaoyi, et al. Target detection algorithm based on decision level fusion of visible and infrared images[J]. *Journal of Air Force Engineering University*, 2020, 21(6): 53-59.