DOI:10.19651/j.cnki.emt.2210042

基于改进 YOLOX 的红外目标检测算法*

谌海云 余鸿皓 王海川 黄忠义
(西南石油大学电气信息学院 成都 610500)

摘 要:针对红外目标图像分辨率低,缺少纹理细节,存在复杂背景干扰导致检测精度低的问题,提出一种基于改进 YOLOX 的红外目标检测算法。首先,设计了一种有效的空间通道混合注意力模块,将其引入在特征提取主干网络 CSP-Darknet53 中,以减少网络由于远距离传输造成的精度损失;其次,为了进一步提升红外目标的检测精度,在原本加强特征提取网络 PANet 的基础上提出一种改进的路径特征融合方法;最后,为了解决红外目标中小物体预测精度低的问题,在 YOLOX 输出检测头处进行反卷积操作扩大输出特征图。在 FLIR 红外公开数据集上进行实验,实验结果表明,所提算法识别的平均精度均值(mAP)达 91.00%,相比于基准 YOLOX 网络的平均精度提升了 5.04 个百分点,对于提升红外目标的检测精度是有效的。

关键词:卷积神经网络;红外目标检测;YOLOX;注意力机制;特征融合
中图分类号:TP391 **文献标识码**:A **国家标准学科分类代码**:520.20

Object detection algorithm of thermal infrared images based on improved YOLOX

Shen Haiyun Yu Honghao Wang Haichuan Huang Zhongyi
(School of Electrical Engineering and Information, Southwest Petroleum University, Chengdu 610500, China)

Abstract: To solve the problem of low resolution of infrared target images, lack of texture details, and low detection accuracy caused by complex background interference, an infrared target detection algorithm based on improved YOLOX is proposed. First, an effective spatial channel mixed attention module is introduced into the feature extraction backbone network CSP-Darknet53 to reduce the accuracy loss of the network due to long-distance transmission; secondly, in order to further improve the detection accuracy of infrared targets, based on the original enhanced feature extraction network PANet, an improved path feature fusion method is proposed; finally, in order to solve the problem of low recognition rate of small objects in infrared targets, a deconvolution operation is performed at the YOLOX output detection-head to expand the output feature map. Experiments are carried out on the FLIR infrared public data set. The experimental results show that the mean Average Precision (mAP) of the proposed algorithm recognition reaches 91.00%, which is 5.04% percentage points higher than that of the benchmark YOLOX network, it is effective to improve the detection accuracy of infrared targets.

Keywords: convolutional neural network; thermal object detection; YOLOX; attention mechanism; feature fusion

0 引 言

红外热辐射具有较强穿透雾、尘、烟和气体的能力[1]。与可见光成像相比,红外成像具有受天气和光照变化影响小,可探测距离远等优点[2]。红外热成像技术因其独特的特点现广泛应用于医疗[3]、边境监控[4]、目标追踪[5]、汽车夜间辅助驾驶等领域[6]。红外目标检测作为红外热成像的关键技术之一,有着重要的研究意义。

随着深度学习算法,特别是卷积神经网络(CNN)的不断发展,目标检测技术在可见光图像领域取得了巨大突破。依赖于海量现实中可方便获取的可见光图片和大型可见光数据集 ImageNet[7]、MSCOCO[8]、Pascal VOC[9]训练的结果,CNN 在可见光图像领域的检测精度已经超越了人类[10]。与之相比,红外图像的成像波长长、噪声大、空间分辨率低、缺乏物体纹理、色彩和形状信息[11]。另外,红外图像现存的公开数据集较少,无法像可见光图像那样在大型

收稿日期:2022-05-21
* 基金项目:智能电网与智能控制南充市重点实验室平台建设(二期)(SXHZ053)、工业炸药智能仓储系统设计与开发项目(SXJBGS002)资助

数据集上进行有效的预训练[12]。这些缺点的存在使得基于 CNN 的红外目标检测模型在进行特征提取时的效果表现不佳,红外目标的检测精度低。如何提高红外目标检测的精度成为现代计算机视觉领域需要解决的难题。

目标检测的任务是利用算法在图像中搜寻感兴趣区域,是目标分类和定位任务的集合[13]。目前自然图像中基于卷积神经网络的目标检测技术主要分为基于锚框(anchor-based)的目标检测算法和无锚框(anchor-free)目标检测算法两种[14]。基于锚框的目标检测算法利用事先设置好的先验框对物体进行预测,比如 Faster R-CNN[15]、Cascade R-CNN[16] 等两阶段框架以及 SSD[17]、RetinaNet[18]、YOLOv3[19]、YOLOv4[20]、YOLOv5[21] 等单阶段框架。无锚框目标检测算法不使用先验框,而是采用图像中物体关键位置点信息进行预测,比如 CenterNet[22]、FCOS[23]等框架。

近几年基于卷积神经网络的红外目标检测模型中,文献[24]通过在 VGG 网络中加入级联结构的路径连接改进 SSD 网络对红外目标进行检测,但该算法网络设计较为简单,检测精度不高;文献[25]利用生成对抗网络(GAN)和特征金字塔网络(FPN)相结合设计出一种跨域融合网络对红外目标进行检测,但该模型较为复杂,训练过程较为繁琐;文献[26]在 YOLOv3 网络中加入注意力机制,提出了一种利用可见光图像和红外图像进行融合检测的模型,但该算法需要配对的红外与可见光图片,现实中难以获取;文献[27]提出一种改进的 YOLOv5 模型,在原本 CSP-Darknet53 网络中通过对浅层模块进行扩展迭代,最大限度地利用红外图像中的浅层特征进行检测,但该算法仅对大目标类别检测精度较高。

现有红外目标检测算法大都忽略了红外图像中浅层的语义信息,精度提升有限。本文提出了一种基于 YOLOX[28]的改进算法。首先设计了一种空间通道混合注意力模块将其引入在特征提取主干网络中以减小网络由于远距离传输造成的精度损失;对 PANet[29]进行改进,提出一种跨路径特征融合方式,增强特征提取提升检测精度;通过扩大检测特征输出图,提升红外目标中小物体的检测精度。

1 YOLOX 目标检测算法

YOLOX 是以 YOLOv3 作为基准模型的改进目标检测网络,与 YOLO 系列单阶段算法相比最大改进之处在于将整个目标检测器变成了 Anchor-free 的方式进行检测,在降低网络参数量的情况下提升了检测的精度,在各大公开数据集上目标检测的效果都超越了之前的 YOLO 系列算法。另外 YOLOX 将原先 YOLOv3 的检测头改为了类似 RetinaNet 的解耦头(decoupled head)结构,原本输出路径被分为两个分支,一个用于回归,另一个用于分类。

如图 1 所示为 YOLOX 网络结构,初始图像的输入尺寸为 640×640,整体网络结构以 CSP-Darknet53 作为主干特征提取网络。输入图像先进入 Focus 网络进行特征提取再送入由 1×1 和 3×3 卷积构成的 Conv2D_BN_SiLU 模块。其中 SiLU 是 YOLOX 网络中使用的激活函数,代替之前 YOLOv3 的 Leaky ReLu 激活函数。SiLU 函数可以表示为

$$y = x \times Sigmoid(x) = x \times \frac{1}{1+e^{-x}}, \tag{1}$$

其中,x 为输入张量,y 为输出张量。

随后被送入 4 个残差结构(resblock body)中生成 Dark1、Dark2、Dark3、Dark4 四个有效特征层,特征层输出尺寸分别为 160×160,80×80,40×40,20×20。其中在 Dark4 特征层处引入了空间金字塔池化(SPP)模块[30],用来增强模型对小目标的特征提取。80×80,40×40,20×20 的输出特征层被传入 PANet 中作加强特征提取,之后传入 3 个 YoloHead 中分别对图像中的大物体、中物体、小物体进行预测。

2 改进的 YOLOX 目标检测算法

2.1 LAM 混合注意力模块

传统的卷积层由于加入的卷积和平均池化操作导致卷积神经网络的感受野增长缓慢,所以导致网络无法有效地提取上下文特征信息[31]。且卷积神经网络提取特征在远距离传输的过程中会造成信息损失[32]。针对以上问题,本文在 YOLOX 的主干特征提取网络 CSP-Darknet53 的 Dark4 输出特征层后面引入一种远距离传输的混合注意力模块(long-range-transmission attention module, LAM)。设计的 LAM 模块结构如图 2 所示。

对于输入的二维张量 $x \in R^{H\times W}$,LAM 首先将张量沿着水平和垂直方向进行池化。水平和垂直的池化窗口大小分别为(H, 1)和(1, W),H 和 W 分别表示输入张量在空间位置的高度和宽度。LAM 在垂直方向的池化可以表示为:

$$y_i^h = \frac{1}{W}\sum_{0\leqslant j\leqslant W} x_{i,j}, y^h \in R^H, \tag{2}$$

同样地,水平方向的池化可以表示为:

$$y_i^v = \frac{1}{H}\sum_{0\leqslant j\leqslant H} x_{i,j}, y^v \in R^V, \tag{3}$$

由于水平和垂直池化核又长又窄可以很容易的获取远距离上下文关系,这么做有助于特征提取主干网络扩大感受野有利于对目标的识别。首先,输入的三维张量 $x \in R^{C\times H\times W}$,$C$ 代表通道数,将输入张量 x 输出为卷积核大小为 3 的水平和垂直两个方向,用于调整图像当前位置信息和相邻位置信息。输出 $y^h \in R^{C\times H}$ 从水平方向获得,输出 $y^v \in R^{C\times W}$ 从垂直方向水平获取。在扩大感受野后,最后 y^h 和 y^v 相加得到输出 $y \in R^{C\times H\times W}$:

$$y_{c,i,j} = y_{c,i}^h + y_{c,j}^v, \tag{4}$$

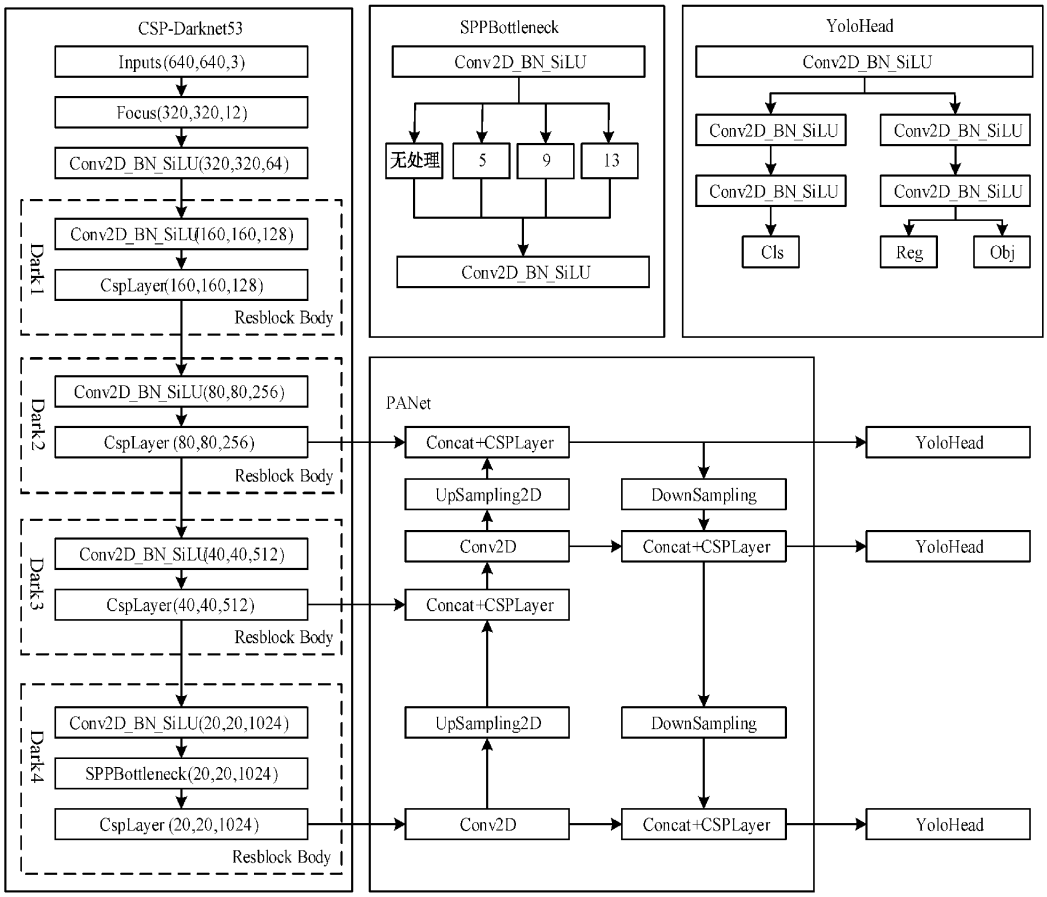


图 1　YOLOX 网络结构

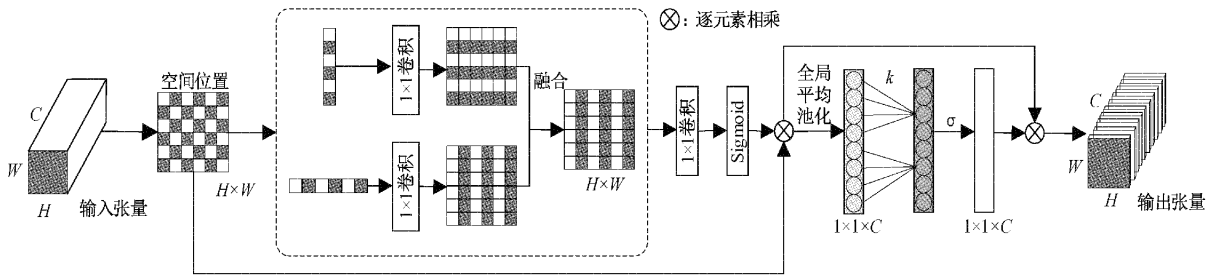


图 2　LAM 注意力模块结构

最终的输出 u 可以表示为：

$$u = Scale(x,\sigma(f(y))), \tag{5}$$

其中，f 表示 1×1 卷积，函数 $Scale(.,.)$ 代表逐元素相乘，σ 为 Sigmoid 激活函数。在经过每一层的卷积处理处理后，都使用批量标准化(batch normalization，BN)进行归一化处理，紧接着使用 ReLU 非线性激活函数进行激活。

之后，使用带状矩阵 $\boldsymbol{W}_k$ 表示 LAM 所学习到的通道注意力，同时通过卷积核大小为 k 的一维卷积来实现通道间的信息交互，带状矩阵 $\boldsymbol{W}_k$ 可以表示为

$$\boldsymbol{W}_k = \begin{bmatrix} w^{1,1} & \cdots & w^{1,k} & 0 & 0 & \cdots & \cdots & 0 \\ 0 & w^{2,2} & \cdots & w^{2,k+1} & 0 & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & 0 & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & w^{C,C-k-1} & \cdots & w^{C,C} \end{bmatrix}, \tag{6}$$

带状矩阵 $\boldsymbol{W}_k$ 包含 $C \times C$ 个参数，在权重设为 z_i 的情况下，本文所提算法仅考虑 z_i 与其 k 个邻域之间的信息交互，即

$$w_i = \sigma\left(\sum_{j=1}^{k} w_i^j z_i^j\right), z_i^j \in \Omega_i^k, \tag{7}$$

其中，Ω_i^k 表示 z_i 的 k 个邻域通道集合。带状矩阵 $\boldsymbol{W}_k$ 可以以一种有效的方式来共享所有通道的权重参数

值,即:

$$w_i = \sigma\left(\sum_{j=1}^{k} w^j z_i^j\right), z_i^j \in \Omega_i^k, \tag{8}$$

这种有效的方式主要通过一组卷积核大小为 k 的一维卷积来快速实现,它可以表示为:

$$w = \sigma(C1D_k(z)), \tag{9}$$

其中,σ 表示 Sigmoid 激活函数,$C1D$ 表示快速一维卷积,它只包含了 k 个信息参数。在 k 和 c 之间存在一个映射 ψ,当给定了通道维数 C,卷积核大小 k 可以使用下面的公式进行计算得到:

$$k = \psi(C) = \left|\frac{\log_2(C)}{2} + \frac{1}{2}\right|_{odd}, \tag{10}$$

其中,$|q|_{odd}$ 表示与 q 最接近的奇数。在 LAM 中一共设置了 256 个通道数集合其可以表示为:

$$C = [C_1, C_2, C_3, \cdots, C_{256}], \tag{11}$$

最后,通过卷积层和非线性激活函数得到权值向量,再得到最终分配的通道输出集合,其可以表示为:

$$\bar{C} = \alpha \cdot C = [\bar{C_1}, \bar{C_2}, \bar{C_3}, \cdots, \bar{C_{256}}], \tag{12}$$

其中,α 为分配权值。

在经过了 LAM 混合注意力模块后,主干特征提取网络对图像的上下文特征信息有了有效的提取,减小了网络远距离传输造成的精度损失,更能够提升 YOLOX 网络的目标检测性能。

2.2　改进的加强特征融合方式

YOLOX 中采用 PANet 网络对来自 CSP-Darknet53 主干特征提取网络提取后的特征图做加强特征提取。PANet 网络是 FPN 网络的改进体,意在尽可能多的保留图像原始浅层信息,如图 3(a)所示在原本 FPN 结构只有自上而下进行特征融合的基础上加入了一条自下而上的融合路径。但是,红外图像由于其本身分辨率低,色彩信息不像可见光图片那样丰富,且红外图像物体纹理细节信息严重丢失,使用 PANet 依然无法有效地对其进行特征提取。对此,本文提出一种基于 PANet 改进的特征融合方式,用来提高检测网络对红外图像底层特征的提取,提高整体对红外图像的检测精度。改进的加强特征融合模块结构如图 3(b)所示。

首先,输入的红外图像经过主干特征提取网络得到 4 个特征层,其中尺寸 80×80、40×40、20×20 的特征层为有效特征层被输入到图 3(a)中的 FPN 主干网络中,完成高层特征层 2 倍上采样后与低层特征层的自上而下特征融合操作;紧接着,由 FPN 网络得到的三个尺寸的特征融合层 P_1,P_3,P_2 进行低层特征层 2 倍降采样后与高层特征层自下而上的特征融合操作。与此同时,本文在 PANet 自上而下和自下而上的融合过程中加入了一种跨路径的特征融合方式,如图 4 所示。来自自上而下路径的低层特征层 P_i 通过 2 倍降采样操作后与来自自下而上路径的高层特征层 N_h 进行堆叠,堆叠过后经 1×1 卷积调整通道。经过跨路径特征融合后,红外图像的高层特征信息与低层特征信息有了交互,更有助于提高检测精度。

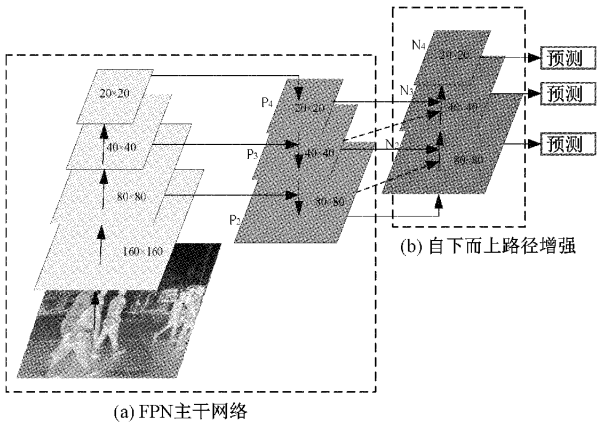


图 3　改进的加强特征融合模块结构

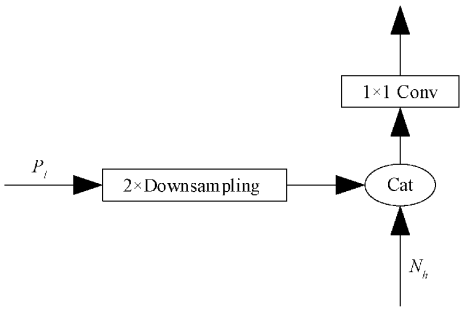


图 4　跨路径特征融合方式

2.3　扩大输出特征图

为了能够更好的对红外图像中的小目标进行预测,本文对 YOLOX 的解耦头结构进行改进,如图 5 所示。

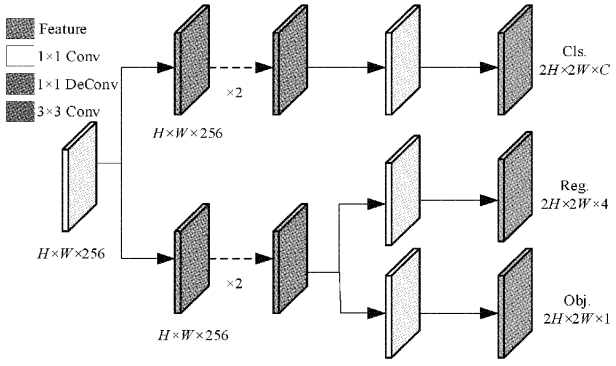


图 5　改进 YOLOX 解耦头结构

该改进的做法是对两个经过了 3×3 卷积的分支采用特征图反卷积,反卷积模块大小为 1×1。此时最后特征图输出的尺寸都×2,即原本输出 20×20、40×40、80×80 的特征图现扩大为 40×40、80×80、160×160,最后红外图像的预测原理如图 6 所示,其中 tx、ty、tw、th、th 分别代表预测框的横坐标、纵坐标、宽度和高度,Cls 表示物体预测的类别,Obj 表示要预测的图片中是否含有需要检测的类

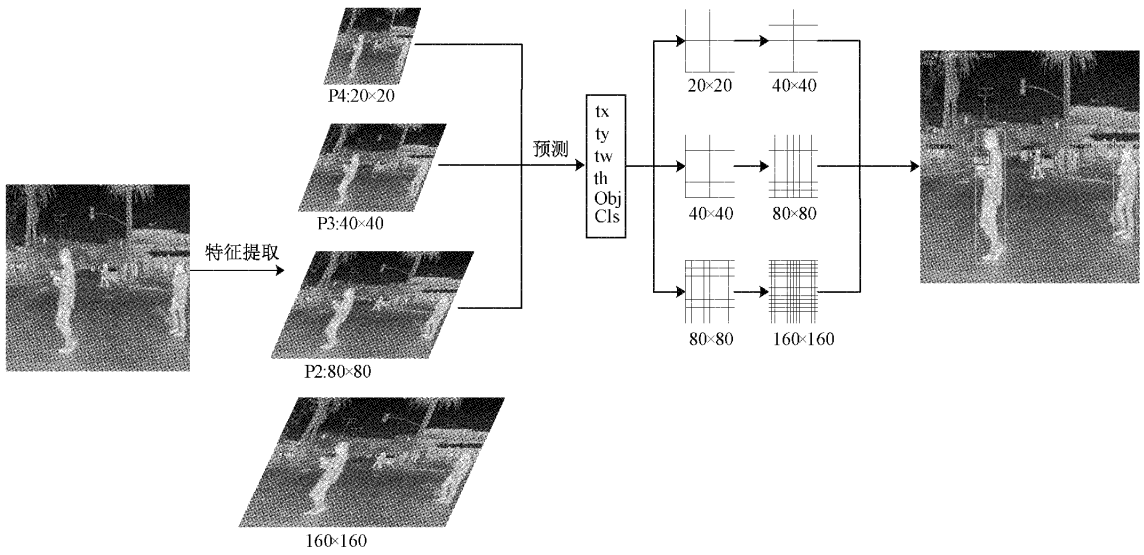


图 6　红外图像预测原理

别。经过扩大特征图操作网络能够预测出更多的目标框，并且能够很好的解决红外图像中小目标的预测问题。

2.4　改进的网路结构

本文通过对 YOLOX 算法的检测模块进行改进，在不加入额外参数量的情况下完成了对红外目标检测精度的提升，本文将所提网络命名为 YOLOX-TI（thermal image, YOLOX），如图 7 所示。YOLOX-TI 在主干特征提取网络中加入了 LAM 模块，能够很好的捕捉上下文信息，减少网络由于远距离传输造成的精度损失；对原本 PANet 网络进行改进，采用提出的跨路径融合方式，将浅层信息与高层信息进行特征融合，有效提高红外目标检测精度；将输出检测头进行反卷积操作，扩大特征图，提高对红外图像中小目标的检测精度。

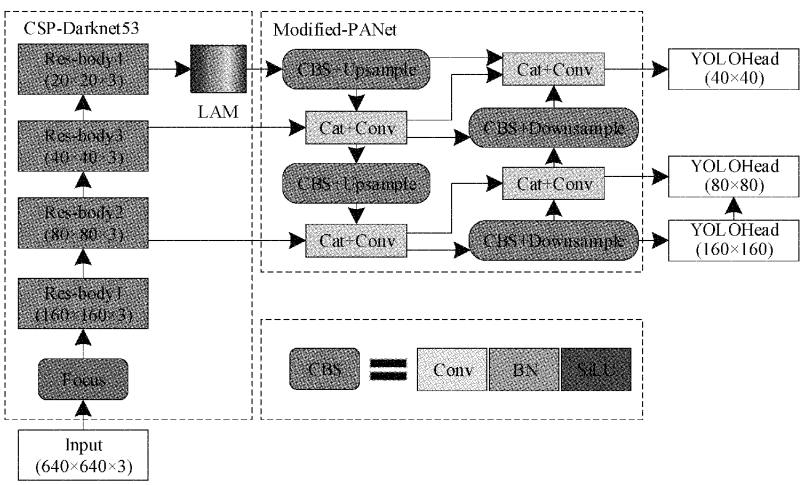


图 7　所提网络结构

3　实验结果与分析

3.1　实验数据集

本文采用 FLIR[33] 红外公开数据集进行实验。FLIR 数据集包含 12 886 张尺寸为 640×512 的红外图像，其中 10 228 张图像有目标框标注信息，包括 31 242 个行人（person）类别实例，61 763 个车辆（car）类别实例和 4 757 个自行车（bicycle）类别实例。数据集中红外图像拍摄来源自美国加利福尼亚州的圣达芭芭拉市的街道和高速公路，所有红外图像均采用 FLIR Tau2 红外车载摄像头进行拍摄获得，其中大约 60%（6 136）的图片在白天拍摄获得，大约 40%（4 092）的图片是在夜间拍摄获得。由于 YOLOX 网络对输入图像尺寸的限制[34]，实验将原本数据集图片的大小拉伸至 640×640 进行输入，在高度方向上对图片进行空白扩充，并不会影响图片中物体的大小。实验过程中将 FLIR 数据集按照 7∶2∶1 的比例划分训练集、验证集和测试集。

3.2　实验平台及实验设置细节

本文实验均在 DELL 的 Precision 5820 Tower X-Series 服务器上完成。操作系统为 Windows10 专业版，

CPU 型号为 Inter(R)Core(TM)i9-10900X CPU,主频大小 3.70 GHz,GPU 采用 NVIDIA GeForce RTX3090,显存大小为 24 GB,内存大小为 64 GB,硬盘容量为 2 TB。实验使用 Python3.8 语言进行编程和测试,选用深度学习框架 Pytorch,版本为 1.8.0,CUDA 版本为 11.5,cuDNN 版本为 8.0.5。

训练过程中,使用标准 Adam 优化器对网络参数进行优化,权重衰减率为 0.9。共计训练 100 轮,batch size 设为 16,初始学习率为 0.001,设置每训练 5 轮的学习率衰减为 0.01。训练分为冻结阶段和解冻阶段,设置前 50 轮的训练为冻结阶段,此时网络的主干部分被冻结,特征提取网络的参数不会发生变化。之后 50 轮为解冻训练阶段,此时网络的主干部分不再被冻结,网络的所有参数都会发生变化。训练过程中使用马赛克(Mosaic)数据增强方式,即将 4 张图片进行比例缩放后拼接在一起来丰富检测物体的种类,在 YOLOX-TI 网络的训练过程中前 85 轮开启 Mosaic 数据增强,最后 15 轮关闭。为了使 YOLOX 网络模型能够快速学习到有用参数,在开始训练前使用在 COCO-Train2017 上预训练的权重对 YOLOX 进行初始化。

3.3 网络模型评价指标

YOLOX-TI 红外目标检测算法使用平均精确率(AP)、平均精度均值(mAP)和检测速度作为网络模型的评价指标。其中 AP 指的是利用不同置信度得出不同查准率(Precision)和召回率(Recall),计算这些不同 Precision 和 Recall 所围成平面曲线的面积。其中 Precision 表示为分类器认为是正类并且最后预测确实是正类的部分占所有分类器认为是正类的比例,其计算表达式如下:

$$Precision = \frac{F_{TP}}{F_{TP}+F_{FP}} \times 100\%, \tag{13}$$

式中:F_{TP} 表示红外图像中正确识别该类别的个数;F_{FP} 表示红外图像中非当前需要识别的类别识别当前需要识别类别的个数。Recall 表示分类器认为是正类且最后预测确实是正类占所有确实是正类的比例,其计算表达式如下:

$$Recall = \frac{F_{TP}}{F_{TP}+F_{FN}} \times 100\%, \tag{14}$$

式中:F_{FN} 表示未被识别出的红外图像类别个数。一般,以 Recall 为坐标 x 轴,Precision 为坐标 y 轴绘制 Precision-Recall(PR)曲线,所围成曲线的面积即为 AP 值,对于连续的 PR 曲线,AP 计算公式如下:

$$F_{AP} = \int_0^1 P(R)\mathrm{d}R, \tag{15}$$

式中:$P(R)$ 表示 Recall(R)关于 Precision(P)的函数。一般来说当 AP 值越高,表示网络模型的分类效果越好。平均精度均值(mAP)是为了衡量网络模型在进行多分类任务时提出的指标,是所有类别的 AP 值叠加再求平均,用来评估一个模型识别性能在整个红外数据集上的好坏,mAP 计算公式如下:

$$F_{mAP} = \frac{1}{N}\sum_{c=1}^{c=N} F_{AP}(c) \tag{16}$$

式中:$F_{AP}(c)$ 表示类别 c 的 AP 值,N 表示总类别数。

本文使用标准 PASCAL-VOC 评价指标,即预测框与真实框的 IoU≥0.5 时的预测概率对 FLIR 红外图像数据集进行评估。

检测速度是以网络模型在单位时间每秒内能够检测的图片数量进行评估的,其单位是 frame/s。

3.4 实验结果与分析

1)实验结果

实验中使用 8 865 张红外图像作为训练集,2 998 张图像作为验证集,1 023 张图像作为测试集,输入尺寸大小均为 640×640,YOLOX-TI 训练过程中,训练和验证损失变化曲线如图 8 所示。前 50 轮由于冻结了主干网络进行训练,模型损失较大,50 轮后进行解冻训练,模型所有结构学习到有用参数,损失值在 50 轮左右迅速下降,到 100 轮损失值逐渐收敛。

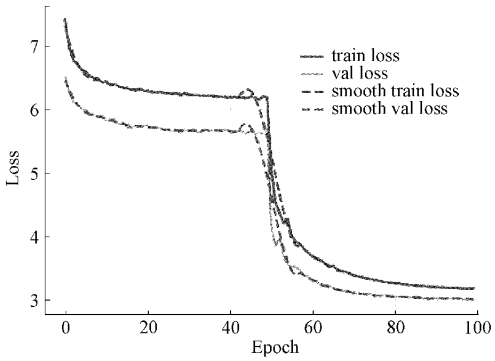


图 8 训练和验证损失曲线

共计训练 100 轮后,在测试集上进行测试,计算 person、car、bicycle 三个类别的 AP 值,得到 3 个类别的 P-R 曲线如图 9(a)~(c)所示。结果 person 类别获得了 92.99% 的 AP 值,car 类别获得了 93.44% 的 AP 值,bicycle 类别获得了 86.56%的 AP 值。最终计算 3 个类别的 mAP 值,如图 9(d)所示,得到 91.00% 的平均精度均值。

2)YOLOX-TI 与其他目标检测模型的定量结果比较

将本文所提出的红外目标检测模型 YOLOX-TI 与其他的先进目标检测算法在 FLIR 数据集上的检测结果做定量比较分析,结果如表 1 所示,粗体代表该项性能对比在所有模型中最高。

所对比的模型有原始 YOLOX 网络;基于 Anchor-based 的两阶段经典检测模型 Faster R-CNN 和 Cascade R-CNN;单阶段检测模型有 YOLOv3、YOLOv4、YOLOv5、SSD 和 RetinaNet;基于 Anchor-free 的检测模型有 CenterNet、FCOS 和原始 YOLOX。从表 1 可以看出,与两阶段模型 Faster R-CNN 和 Cascade R-CNN 相比,YOLOX-TI 在各类 AP、mAP 和检测速度上都优于前者,

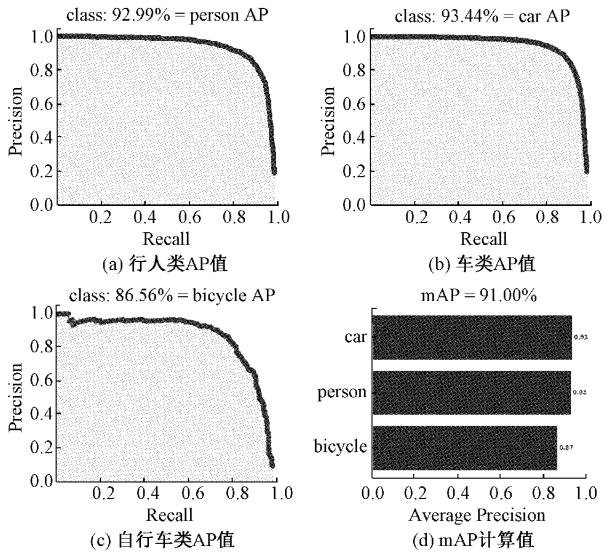


图 9　各类 AP 值以及 mAP 值计算

可以同时满足对红外图像的高精度检测和实时检测；CenterNet、FCOS 在红外图像上的检测效果要高于两阶段模型，且因为基于 Anchor-free 的模型减少了参数量，在检测速度上得到了提升，但与之后的 YOLO 系列算法以及本文提出的 YOLOX-TI 算法相比，在检测精度上还存在不足；与单阶段模型 SSD、RetinaNet、YOLOv3、YOLOv4、YOLOv5 对比，可以看出 YOLO 系列模型对红外图像检测的优越性，同时单阶段模型相比两阶段模型的检测速度更快，更具实时性。其中检测速度最快的是 YOLOv3 网络，检测速度达到 32frame/s，虽然 YOLOX-TI 网络的检测速度略低于 YOLOv3，但也能够满足现实中对红外图像的实时检测。

在所有检测网络中，YOLOX-TI 对红外图像中 car 这一大物体类别的识别精度来到 90%以上，达到 93.44%；对 person 这一中等物体类别的识别精度来到 90%以上，达到 92.99%；对 bicycle 这一小物体类别的识别精度来到了 85%以上，达到了 86.56%。其中对红外小目标的检测精度提升最为明显，与小目标检测最优异的 YOLOv4 模型相比，高出了 7.10%的 AP 值。与原始 YOLOX 网络相比，针对红外图像检测的改进效果明显，在略微降低了检测速度的情况下，YOLOX-TI 网络的 mAP 提高了 5.04 个百分点。

表 1　YOLOX-TI 与其他目标检测模型定量对比

网络结构	特征提取网络	AP/%			mAP/%	检测速度/(frame•10^{-1})
		person	car	bicycle		
Faster R-CNN	Resnet50	66.56	75.94	60.13	67.54	16
Cascade R-CNN	Resnet50	78.65	79.84	66.47	74.98	13
RetinaNet	Resnet50	64.22	72.56	52.10	62.96	26
SSD	VGG16	62.76	72.29	53.50	62.85	28
CenterNet	Resnet50	76.81	79.99	75.23	77.34	31
FCOS	Resnet50	79.43	80.10	77.65	79.06	25
YOLOv3	Darknet53	79.35	83.57	71.22	78.04	32
YOLOv4	CSP-Darknet53	85.03	88.47	78.20	83.90	28
YOLOv5	CSP-Darknet53	88.25	89.74	75.12	84.37	26
YOLOX	CSP-Darknet53	89.10	91.22	77.56	85.96	29
YOLOX-TI(本文算法)	CSP-Darknet53	**92.99**	**93.44**	**86.56**	**91.00**	27

为了能够直观地展现本文所提 YOLOX-TI 网络在红外图像上的检测效果，从 FLIR 的测试集中挑选了 3 张包含 3 个类别的红外图像，选择 YOLOv4、YOLOv5、原始 YOLOX 网络与 YOLOX-TI 网络进行检测效果对比，红外图像检测对比效果如图 10 所示。图 10 中(a)～(d) YOLOX-TI、YOLOX、YOLOv5、YOLOv4 算法的检测效果图。在图 10 第 1 列遮挡情况下红外目标的检测中，只有 YOLOX-TI 算法正确的识别出了两个小目标 bicycle 类；第 2 列复杂场景下红外目标的识别中 YOLOX-TI 和 YOLOv5 算法识别出了图中的 bicycle 类，但 YOLOX-TI 算法在检测精度上高于 YOLOv5；第 3 列光照情况下的红外图像中，YOLOv5 对图像右侧的类别进行了错误的预测，而对比剩下的 YOLOv4、原始 YOLOX 算法，YOLOX-TI 算法检测精度更好，定位更为准确。实验结果和检测效果对比证明了改进的 YOLOX-TI 算法对红外目标检测具有更好的性能。

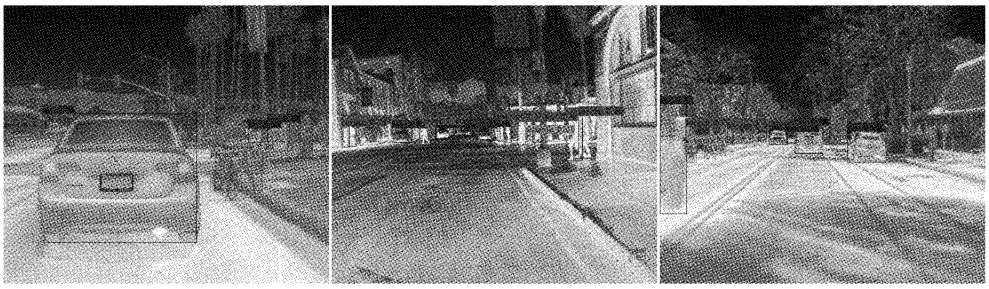
3)消融实验

为了测试 YOLOX-TI 模型中每个改进部分的有效性，设置消融实验分别将 LAM、改进 PANet、扩大输出特征图应用在原始 YOLOX 网络中。共设置 a～d 四组实验，其中 a 组实验为初始 YOLOX 网络，b～d 组为添加改进部分的 YOLOX 网络，“√”表示添加该改进部分，“×”表示不添加改进部分。消融实验结果如表 2 所示。

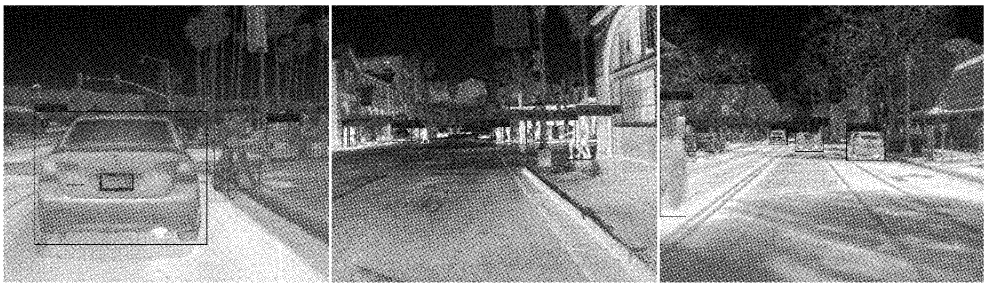
表 2 消融实验结果

实验分组	LAM	改进 PANet	扩大特征图	AP/%			mAP/%	运行速度/(frame•10^{-1})
				person	car	bicycle		
a	×	×	×	89.10	91.22	77.56	85.96	29
b	√	×	×	90.40	91.45	81.47	87.77	29
c	√	√	×	92.17	93.00	84.25	89.80	27
d	√	√	√	92.99	93.44	86.56	91.00	27

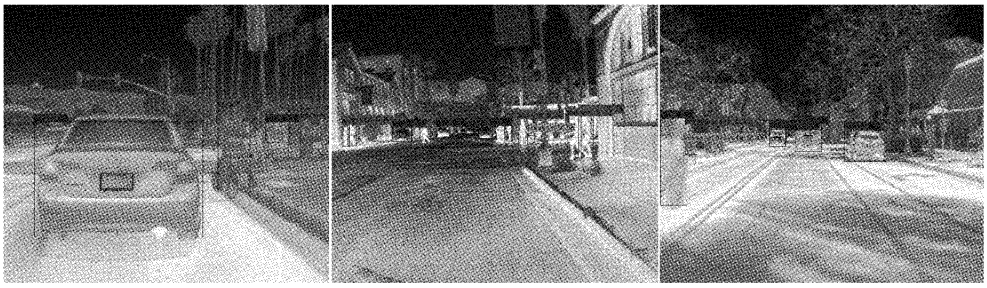
由表 2 结果可以看出,在 b 组实验中加入了 LAM 模块,各类 AP 得到提升,mAP 相比原始 YOLOX 网路提升了 1.81 个百分点,其中小目标 bicycle 的 AP 精度增长最为明显,说明 LAM 模块可以有效弥补主干特征网络在远距离传输中小目标的信息丢失,提升检测精度;在 c 组实验中加入了改进的 PANet 跨路径融合方式,其中 person、car、bicycle 类 AP 分别提升了 1.77、1.55 和 2.78 个百分点,mAP 提升了 2.03 个百分点。从此可以看出改进的 PANet 跨路径融合方式能够有效利用红外图像中高层特征信息和浅层特征信息进行融合,虽然模型计算量增加,导致运行速度有所减慢,但有效地提升了红外目标的检测精度;在 d 组实验中对 YOLOX 网络进行了扩大输出特征图操作,其中小目标 bicycle 类 AP 提升了 2.31 个百分点,mAP 提升了 1.2 个百分点。扩大特征图是专门为提升红外图像中小目标检测精度进行的,从结果看,红外图像中的小目标检测效果得到了很好的提升。综上所述,改进的 YOLOX-TI 模型相比于原始 YOLOX 网络提升效果明显,更适合用作红外目标检测。



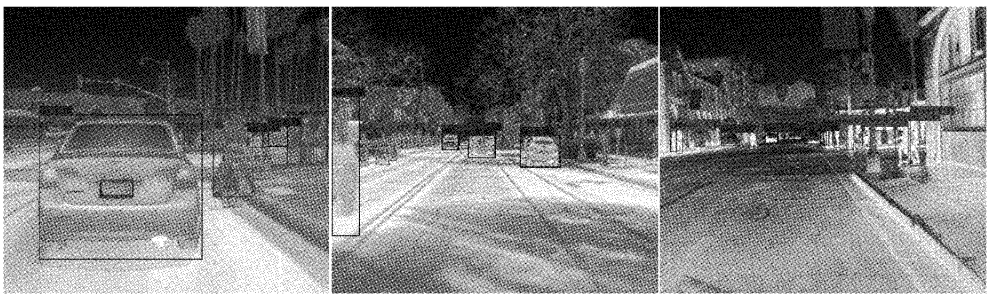
(a) YOLOX-TI



(b) YOLOX



(c) YOLOV5



(d) YOLOV4

图 10　四种算法红外目标检测效果对比

4　结　　论

为了解决红外图像因分辨率低、缺乏纹理细节造成检测精度低的问题，本文基于 YOLOX 提出了改进的红外目标检测算法 YOLOX-TI。为了解决主干特征提取网络由于远距离传输图像细节丢失造成的精度损失，加入了设计的 LAM 混合注意力模块；对 PANet 网络进行改进，提出了一种跨路径融合方式，有效利用了红外图像的高层特征信息和浅层特征信息，有效提高了模型对红外图像的检测精度；对 YOLOX 的输出头进行扩大特征图处理，有效提高模型对红外小目标的检测精度。在 FLIR 红外数据集上进行实验，结果表明本文所提 YOLOX-TI 模型在红外图像上的检测效果要优于其他目前主流的目标检测模型，相比于原始 YOLOX 网络，检测精度得到很大提升。本文所采用的的红外数据集较少，后续工作将在其他红外数据集上进行实验，进一步验证所改进的网络模型对红外目标检测的有效性。

参考文献

[1] GAUTAM A, SINGH S. Neural style transfer combined with EfficientDet for thermal surveillance[J]. The Visual Computer, 2021: 1-17.

[2] MUNIR F, AZAM S, RAFIQUE M A, et al. Exploring thermal images for object detection in underexposure regions for autonomous driving [J]. ArXiv Preprint, 2020, ArXiv: 2006.00821.

[3] 李春艳，孙韬，谢俊峰. EMF 深度学习可见光/红外图像融合算法[J]. 国外电子测量技术，2020，311(10)：25-32.

[4] 张志强，王萍，于旭东，等. 高精度红外热成像测温技术研究[J]. 仪器仪表学报，2020，41(5)：10-18.

[5] YAO T, HU J, ZHANG B, et al. Scale and appearance variation enhanced siamese network for thermal infrared target tracking[J]. Infrared Physics & Technology, 2021, 117: 103825.

[6] FAROOQ M A, CORCORAN P, ROTARIU C, et al. Object detection in thermal spectrum for advanced driver-assistance systems(ADAS)[J]. IEEE Access, 2021, 9: 156465-156481.

[7] DENG J, DONG W, SOCHER R, et al. Imagenet: A large-scale hierarchical image database[C]. 2009 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2009: 248-255.

[8] CHEN X, FANG H, LIN T Y, et al. Microsoft coco captions: Data collection and evaluation server [J]. ArXiv Preprint, 2015, ArXiv: 1504.00325.

[9] RUSSAKOVSKY O, DENG J, SU H, et al. Imagenet large scale visual recognition challenge[J]. International Journal of Computer Vision, 2015, 115(3): 211-252.

[10] WU X, SAHOO D, HOI S C H. Recent advances in deep learning for object detection[J]. Neurocomputing, 2020, 396: 39-64.

[11] LAHMYED R, EL ANSARI M, ELLAHYANI A. A new thermal infrared and visible spectrum images-based pedestrian detection system [J]. Multimedia Tools and Applications, 2019, 78(12): 15861-15885.

[12] DEVAGUPTAPU C, AKOLEKAR N, M SHARMA M, et al. Borrow from anywhere: Pseudo multi-modal object detection in thermal imagery[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2019: 0-0.

[13] 吕禾丰，陆华才. 基于 YOLOv5 算法的交通标志识别技术研究[J]. 电子测量与仪器学报，2021，250(10)：137-144.

[14] 程换新，蒋泽芹，程力，等. 基于改进 YOLOX-S 的安全帽反光衣检测算法[J]. 电子测量技术，2022，386(6)：130-135.

[15] REN S, HE K, GIRSHICK R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks [J]. Advances in Neural Information Processing Systems, 2015, 28.

[16] CAI Z, VASCONCELOS N. Cascade r-cnn: Delving into high quality object detection[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern

Recognition, 2018: 6154-6162.

[17] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector [C]. European Conference on Computer Vision, Springer, Cham, 2016: 21-37.

[18] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection [C]. Proceedings of the IEEE International Conference on Computer Vision, 2017: 2980-2988.

[19] REDMON J, FARHADI A. Yolov3: An incremental improvement [J]. ArXiv Preprint, 2018, ArXiv: 1804.02767.

[20] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. Yolov4: Optimal speed and accuracy of object detection [J]. ArXiv Preprint, 2020, ArXiv: 2004.10934.

[21] Ultralytics. YOLOv5[EB/OL]. (2020-06-03)[2021-04-15]. https://github.com/ultralytics/yolov5.

[22] ZHOU X, WANG D, KRÄHENBÜHL P. Objects as points[J]. ArXiv Preprint,2019, ArXiv:1904.07850.

[23] TIAN Z, SHEN C, CHEN H, et al. Fcos: Fully convolutional one-stage object detection [C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 9627-9636.

[24] DAI X, YUAN X, WEI X. TIRNet: Object detection in thermal infrared images for autonomous driving[J]. Applied Intelligence, 2021, 51(3): 1244-1261.

[25] 赵明,张浩然.一种基于跨域融合网络的红外目标检测方法[J]. 光子学报, 2021, 50(11): 1110001.

[26] XUE Y, JU Z, LI Y, et al. MAF-YOLO: Multi-modal attention fusion based YOLO for pedestrian detection[J]. Infrared Physics & Technology, 2021, 118: 103906.

[27] LI S, LI Y, LI Y, et al. YOLO-FIRI: Improved YOLOv5 for infrared image object detection[J]. IEEE Access, 2021, 9: 141861-141875.

[28] GE Z, LIU S, WANG F, et al. Yolox: Exceeding yolo series in 2021 [J]. ArXiv Preprint, 2021, ArXiv:2107.08430.

[29] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 8759-8768.

[30] HE K, ZHANG X, REN S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904-1916.

[31] HOU Q, ZHANG L, CHENG M, et al. Strip pooling: Rethinking spatial pooling for scene parsing [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 4003-4012.

[32] WANG Q, WU B, ZHU P, et al. ECA-Net: Efficient channel attention for deep convolutional neural networks [C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2020.

[33] GROUP F A. Flir thermal dataset for algorithm training[J]. URL: https://www.flir.co.uk/oem/adas/adas-dataset-form/(May 2019), 2018.

作者简介

谌海云,教授,硕士研究生导师,主要研究方向为图像处理、信号处理与分析。

余鸿皓(通信作者),硕士研究生,主要研究方向为计算机视觉,红外目标检测。

E-mail:1224266074@qq.com

王海川,硕士研究生,主要研究方向为无人机单目标跟踪。

黄忠义,硕士研究生,主要研究方向为多目标行人跟踪。