

基于优化 YOLOv4 算法的低空无人机检测与跟踪

赵玥萌^{1,2}, 刘会刚^{1,2*}

¹南开大学电子信息与光学工程学院薄膜光电子技术教育部工程研究中心, 天津 300350;

²天津市光电传感器与传感网络技术重点实验室, 天津 300350

摘要 随着非军事无人机的普及, 无人机检测技术已经成为安全领域的研究热点, 基于此, 提出了一种基于优化 YOLOv4 的低空无人机检测与跟踪方法, 所提方法首次将基于卷积神经网络(CNN)的检测技术与跟踪算法相结合用于低空无人机的动态检测。首先, 根据多尺度特征融合的思想对原始 YOLO 网络结构进行优化; 然后, 在此基础上结合 DeepSORT 多目标跟踪算法, 构建了低空无人机检测与跟踪模型。在自建数据集 LARotorcraft 上进行了训练和对比实验, 实验结果表明, 所提模型能够有效降低小目标漏检率, 在保证实时性的前提下, 检测平均精度达到 77.2%, 并且实现了视觉目标的稳定跟踪。

关键词 机器视觉; 目标检测; 低空无人机; YOLOv4; 目标跟踪

中图分类号 TP391.4

文献标志码 A

DOI: 10.3788/LOP202259.1215017

Detection and Tracking of Low-Altitude Unmanned Aerial Vehicles Based on Optimized YOLOv4 Algorithm

Zhao Yuemeng^{1,2}, Liu Huigang^{1,2*}

¹Engineering Research Center of Thin Film Optoelectronics Technology, Ministry of Education, College of Electronic Information and Optical Engineering, Nankai University, Tianjin 300350, China;

²Tianjin Key Laboratory of Optoelectronic Sensor and Sensing Network Technology, Tianjin 300350, China

Abstract With the popularization of nonmilitary unmanned aerial vehicles (UAVs), UAV-detection technology has become a hotspot in security research. This study proposes a low-altitude UAV-detection and -tracking method based on the optimized YOLOv4. This method combines detection technology based on convolutional neural networks with a tracking algorithm for the first time to achieve dynamic detection of low-altitude UAVs. First, the original YOLO network structure is optimized based on multiscale feature fusion. Thereafter, in combination with the DeepSORT multitarget tracking algorithm, the detection and tracking model is constructed. Training and comparative experiments are performed on the self-built LARotorcraft dataset. The experimental results show that the proposed model can effectively reduce the miss detection rate for small targets. Good real-time performance is obtained with an average detection accuracy of up to 77.2%, and stable tracking of visual targets is realized.

Key words machine vision; object detection; low-altitude unmanned aerial vehicle; YOLOv4; object tracking

1 引言

低空空域已成为空域资源开发利用的重要组

成部分^[1]。近年来,随着低空空域的开放,无人机(UAV)逐渐被应用于非军事用途,如远程监控、城市管理、农业、地质、气象、电力、抢险救灾、快递投

收稿日期: 2021-08-02; 修回日期: 2021-08-18; 录用日期: 2021-09-08

通信作者: *liuhg@nankai.edu.cn

送和视频拍摄等领域,这些无疑使得无人机市场规模增大,它的应用也变得普及化。然而,无人机的使用难以规范且事故频发,无人机的滥用则会对公共安全、个人隐私等造成严重损害,因此亟需一种有效的低空空域无人机检测与跟踪方法。

在无人机防控系统中,探测、识别与跟踪是最重要的部分,是防御的基础。目前,常用的无人机探测与跟踪技术手段包括雷达探测^[2]、声学探测^[3]和射频探测^[4]等。但由于高昂的安装和运行成本及监管限制,很难为私人或每个公共设施搭载上述探测设备。因此,针对中小型无人机,基于机器视觉的目标检测与跟踪技术是低空无人机探测与跟踪系统常用的技术之一。

目标检测是一项重要的计算机视觉任务。传统的目标检测技术大多基于手工进行特征构建^[5],其泛化能力弱,占用计算资源大。而深度学习技术的快速发展为目标检测领域带来了新的突破,基于卷积神经网络(CNN)的检测算法在提取图像深层次特征方面表现优异^[6],具有很强的泛化能力,目前主要分为两类:一类是以 R-CNN^[7]、Faster R-CNN^[8]、R-FCN^[9]为典型代表的两阶段目标检测算法,此类算法先生成候选框,再利用 CNN 完成预测,检测精度高但检测速度较慢;另一类是以 SSD^[10]、YOLO^[11-14]为典型代表的单阶段目标检测算法,此类算法直接利用检测网络进行预测,检测速度能够达到实时要求^[15]。YOLO 算法在检测速度与精度上均表现优异,并且随着版本的不断精进,在目标检测应用中展现出极大的优势^[16-17]。

实现目标跟踪的关键在于完整地分割目标、合理地提取特征并精确地识别目标。基于 CNN 的目标检测技术逐渐打破了目标检测与目标跟踪之间的壁垒,将基于 CNN 的算法检测结果作为输入可以提高跟踪的可靠性,可以在复杂环境下保证检测跟踪任务精度的同时进一步提高检测速度。Tan 等^[18]提出了一种基于 YOLO 的多目标跟踪算法,该算法同时提高了对多目标跟踪的精度和效率。Zhang 等^[19]利用 Faster R-CNN 进行目标检测,结合互信息图像配准技术进行跟踪,达到了精确跟踪多个目标的效果。Jiang 等^[20]将基于 CNN 的目标检测与基于模板匹配(TM)的目标跟踪结合,提出了一种快速的对象跟踪算法,并搭建了云台目标跟踪系统。

低空无人机具有体积小、机动性高的特点,对

其进行检测和识别需要一定的准确性和实时性。为了提高小目标无人机的实时检测精度与跟踪稳定性,本文提出了一种基于优化 YOLOv4 的目标检测与跟踪方法。首先,对自制数据集进行数据增强等预处理,使用 K-means 聚类算法重新生成最优先验框,适应低空无人机检测任务。其次,根据多尺度特征融合的思想对 YOLOv4 架构中的 PANet 特征融合模块进行改进,将第 3 个残差块的输出特征图上采样,并与第 2 个残差块的输出进行拼接,用 4 个尺度的特征图代替 3 个尺度的特征图进行预测,强化检测小目标的能力。最后,将优化的 YOLOv4 网络作为目标检测模型,结合 DeepSORT^[21]多目标跟踪方法进行动态检测,保证准确率的同时缩短了检测时间。实验结果表明,所提方法对于低空无人机目标检测的精度和速度都有明显的提升。

2 目标检测与跟踪算法

2.1 YOLOv4 算法

YOLO 系列目标检测算法将图像识别任务视作回归问题来处理,在一个卷积神经网络中进行特征提取、分类和回归。YOLOv4 是在 YOLOv3 的基础上进一步改进的,其结构如图 1 所示。它沿用了 YOLOv3 的整体架构,将主干特征提取网络 Darknet-53 替换为 CSPDarknet-53,并增加了 PANet 特征融合模块和空间金字塔池化(SPP)附加模块。

YOLOv3 的 Darknet53 网络共有 5 个残差块,5 个残差块依次包含 1,2,8,8,4 个小残差单元。而 CSPDarknet-53 在此基础上进行了两处改进:1)在 DarknetConv2D 模块中使用自正则化非单调函数 Mish 代替 Leaky ReLU 函数以提高检测准确率。Mish 激活函数的表达式为

$$\text{Mish}(x) = x \times \tanh\{\ln[1 + \exp(x)]\}。 \quad (1)$$

2)在 Darknet-53 的每个大残差块上添加 cross stage partial network(CSPNet),两个 1×1 卷积分别对输入特征进行变换,一个输入原有残差结构中,另一个则直接与输出相连接,如图 2 所示。CSPDarknet-53 是整个算法的核心,增强了 CNN 的学习能力,减小了计算量并降低了内存成本。

为了更好地提取目标的融合特征,YOLOv4 在主网络与输出层之间插入了 SPP 模块和 PANet 模块。SPP 模块添加在 CSPDarknet-53 最后一个特征层的三次卷积后,可用来分别进行 4 个不同尺度的

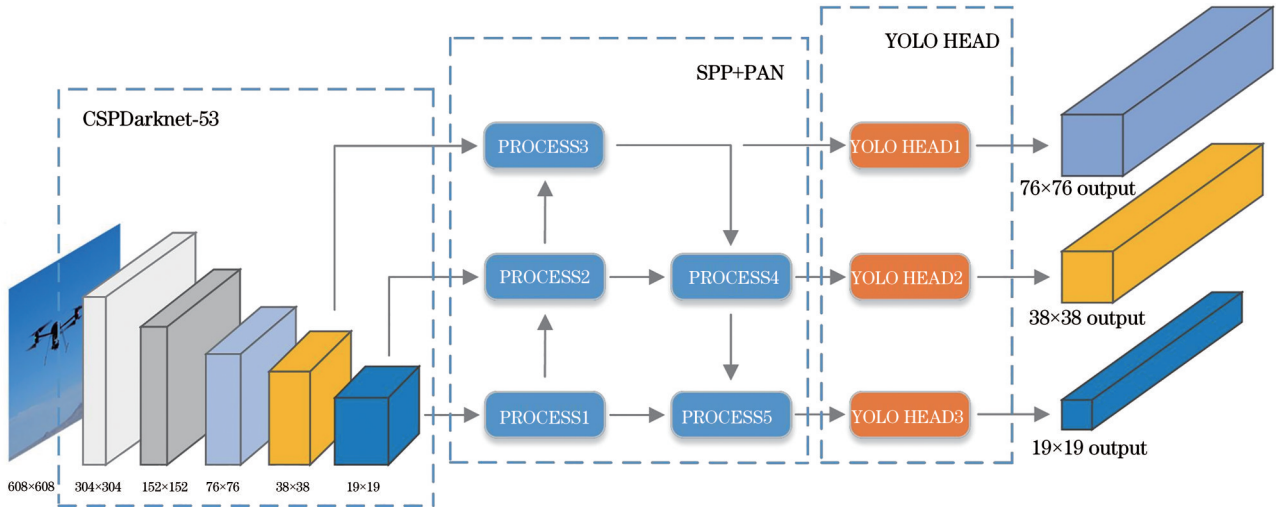


图 1 YOLOv4 结构示意图

Fig. 1 Schematic diagram of YOLOv4 structure

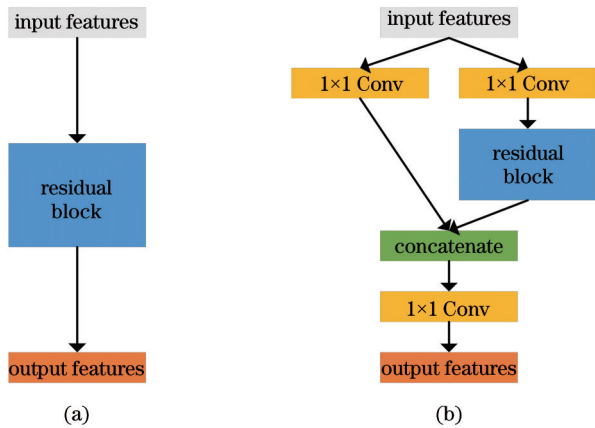


图 2 CSPNet 与原结构的对比。(a) Darknet-53; (b) CSPDarknet-53

Fig. 2 Comparison between Darknet-53 and CSPDarknet-53. (a) Darknet-53; (b) CSPDarknet-53

最大池化,池化核的大小为 $1 \times 1, 5 \times 5, 9 \times 9, 13 \times 13$ 。SPP 模块可以增加整体网络的感受野,并显著地分离重要的上下文特征。而 PANet^[22]采取了一种自底向上的路径来增强特征金字塔(FPN)^[23]结构,即先通过上采样将高层特征的语义信息传播到低层网络,再经过下采样,完成多次特征融合,从而使得低层特征更加有效地传播,进一步提高了网络的特征提取能力。

YOLOv4 还包含了很多创新的思想,例如在数据处理、网络训练、损失函数等方面也均引入了优化,降低了计算资源使用率。在 MS COCO 数据集上与 YOLOv3 相比较,在检测速度没有下降的情况下,YOLOv4 的平均精度(mAP)提高到 44%,整体性能显著提高。YOLOv4 同时实现了速度与精度

的高度平衡,综合性能优于 SSD、Faster R-CNN 等检测算法^[14]。

2.2 DeepSORT 算法

SORT^[24]是一种简单的在线实时多目标跟踪算法,该算法在图像空间中使用基于匀速运动和线性观测模型的卡尔曼滤波器进行状态估计,通过匈牙利算法对数据进行逐帧关联,并选择最优关联结果。DeepSORT 是 SORT 的改进版本,添加了基于运动信息和目标外观信息的新型数据关联度量方法,处理跟踪问题的具体步骤包括轨迹处理与状态估计、数据关联及级联匹配。

DeepSORT 算法定义了一个八维空间向量 $(u, v, \gamma, h, \dot{x}, \dot{y}, \dot{\gamma}, \dot{h})$ 用于表示某一时刻的轨迹状态,该空间包含边界框中心位置 (u, v) ,纵横比 γ ,高度 h ,参数 $\dot{x}, \dot{y}, \dot{\gamma}, \dot{h}$ 为 u, v, γ, h 相对于图像坐标的速度。

在数据关联部分,DeepSORT 使用检测结果与跟踪器预测结果之间的马氏距离 $d^{(1)}(i, j)$ 来实现对运动信息的关联。

$$d^{(1)}(i, j) = (\mathbf{d}_j - \mathbf{y}_i)^T \mathbf{S}_i^{-1} (\mathbf{d}_j - \mathbf{y}_i), \quad (2)$$

式中: \mathbf{d}_j 表示第 j 个检测框的预测位置; \mathbf{y}_i 表示第 i 个跟踪器对目标的预测位置; \mathbf{S}_i 是检测位置与平均跟踪位置之间的协方差矩阵。马氏距离仅适用于连续性运动,当运动的不确定性较高时,会造成马氏距离大量不匹配,因此又引入最小余弦距离 $d^{(2)}(i, j)$ 对目标外观信息进行关联以适应运动的不确定性。

$$d^{(2)}(i, j) = \min \{1 - \mathbf{r}_j^T \mathbf{r}_k^{(i)} | \mathbf{r}_k^{(i)} \in \mathbf{R}_i\}, \quad (3)$$

式中： r_j 是第 j 个检测框的表面特征描述符， $\|r_j\|=1$ ； $R_k = \{r_k^{(i)}\}_{k=1}^{L_k}$ 用于存放每个轨道最新的 L_k ($L_k = 100$) 个描述符。

最终的关联度量是通过加权马氏距离和最小余弦距离两个度量得到的，有

$$c_{i,j} = \lambda d^{(1)}(i,j) + (1 - \lambda) d^{(2)}(i,j) \quad (4)$$

当 $c_{i,j}$ 位于两种度量设定阈值的交集内，认为信息关联成功。当目标长时间被遮挡，滤波器预测的不确定性将大幅增加，进而破坏跟踪连续性。DeepSORT 算法提出一种级联匹配策略解决该问题，具体过程如图 3 所示。

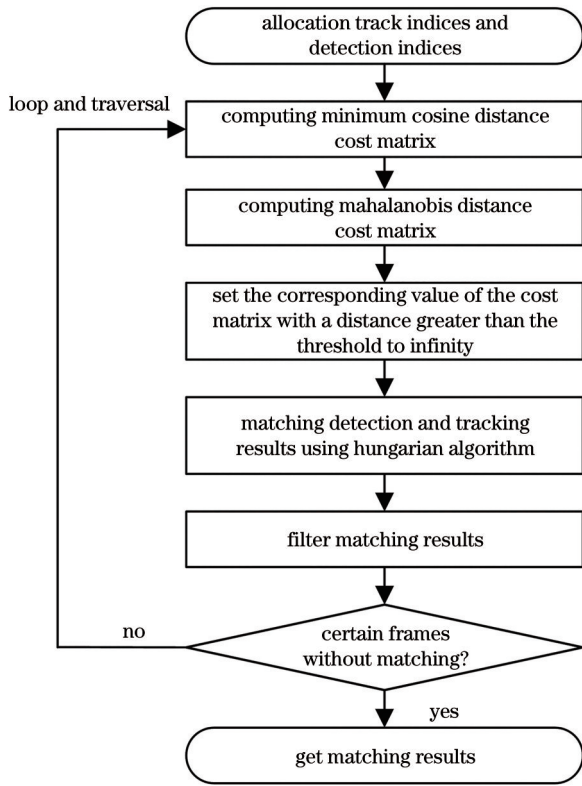


图 3 级联匹配过程

Fig. 3 Cascade matching process

DeepSORT 算法克服了 SORT 的缺点，结合深度外观信息，提高了遮挡情况下的追踪准确度，计算成本较低，可以实现实时在线跟踪。

3 基于优化 YOLOv4 的目标检测与跟踪模型

所提方法首先通过 K-means 聚类算法重新设计不同尺度的 anchor box 以适应小目标检测任务，然后修改 YOLOv4 网络的层级结构，利用多尺度特征融合提高小目标的检测精度。将优化的 YOLOv4

网络的检测结果作为 DeepSORT 多目标跟踪算法的输入，实现高稳定性和高精度的实时检测。

3.1 基于 K-means 算法的聚类分析

YOLOv4 算法中先验框的默认尺寸是利用 K-means 算法对 COCO 数据集聚类得到的^[25]。尽管该默认尺寸有一定的通用性，但对于一些自制的差异化的数据集，采用默认参数训练不仅会影响模型的精度，还会导致训练长时间不收敛。初始候选框并不适用于本实验的无人机检测任务，需要对参数进行优化，重新生成先验框。

K-means 算法是无监督的聚类算法，采用距离作为相似性指标，将 N 个样本点划分为 K 个聚类，每个聚类中心根据类中所有数据点的均值得到，目的是将相似样本尽可能地分为同一聚类，从而对检测结果产生正反馈。标准 K-means 算法选取欧氏距离函数进行计算，不同尺寸产生的误差不同，聚类结果可能产生偏离，本实验组采用交并比 (IOU) 作为衡量指标，避免误差与尺寸相关联，距离公式为

$$d(b_{\text{box}}, c_{\text{centroid}}) = 1 - R_{\text{IOU}}(b_{\text{box}}, c_{\text{centroid}}) \quad (5)$$

式中： b_{box} 表示样本； c_{centroid} 表示聚类中心； $R_{\text{IOU}}(b_{\text{box}}, c_{\text{centroid}})$ 表示聚类框与聚类中心框的交并比。

本实验组使用平均交并比 (Avg IOU) 作为评估度量对无人机数据集进行聚类分析。随着 K 增大到某值，曲线逐渐开始变得缓慢，则在该拐点附近取得最佳 K 值，从而获得最优聚类效果。聚类分析得到的 K 值与 Avg IOU 之间的关系如图 4 所示。

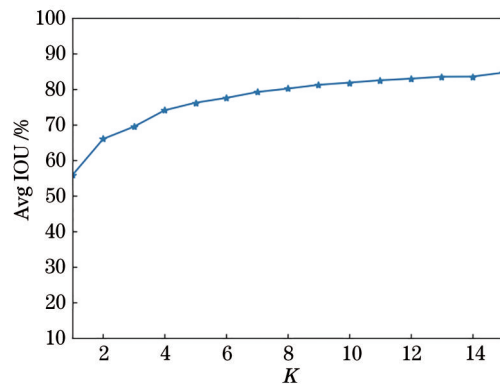


图 4 K-means 聚类分析结果

Fig. 4 K-means clustering analysis result

由图 4 可知，在 $K > 8$ 处，Avg IOU 曲线增长缓慢，因此设置 8 个先验框进行聚类，对应聚类中心的宽度和高度分别为 (12, 16)、(19, 36)、(36, 75)、(76, 55)、(128, 110)、(241, 186)、(485, 493)、

(495, 279)。

3.2 优化的YOLOv4模型

YOLOv4网络沿用了YOLOv3的分类回归层,采用3种不同尺度的特征图进行多尺度预测。假设输入图像的大小为608 pixel×608 pixel,则利用19×19,38×38,76×76三个特征层同时对输入目标进行检测。但低空无人机的体积较小,目标检测层对小目标的检测能力有限^[26],当无人机目标尺寸小于8 pixel×8 pixel时,检测准确率会大幅下降。

对于小目标检测任务,特征图一方面应具有足够高的分辨率,才能得到较小区域的目标信息;另一方面应具备较开阔的感受野,提高语义信息表征能力。为使网络能够获得更多的小目标特征信息,本实验组对YOLOv4算法的特征融合部分进行了改进,建立了一个4尺度特征融合目标检测层。在特

征提取部分添加原始网络中第2个残差块的输出特征图,利用152×152特征层进行目标检测,增加特征图分辨率,同时根据多尺度特征融合的思想,将该特征层与其他通道融合,构建新的PANet模块。

优化后的YOLOv4网络如图5所示,虚线框标出了在原有网络基础上的改进,将第3个残差块输出的76×76特征图进行二倍上采样,并连接到第2个残差块输出的特征图,实现高层语义信息向低层的传播,经卷积后得到152×152目标检测层;再将堆叠的特征图进行二倍下采样至76×76,与经过卷积的原网络八倍降采样特征图堆叠,实现自底向上特征融合,得到的特征图作为YOLO Head 2的输入。YOLO Head部分利用最后4个尺度的特征图代替3个尺度的特征图来预测无人机目标的边界框。

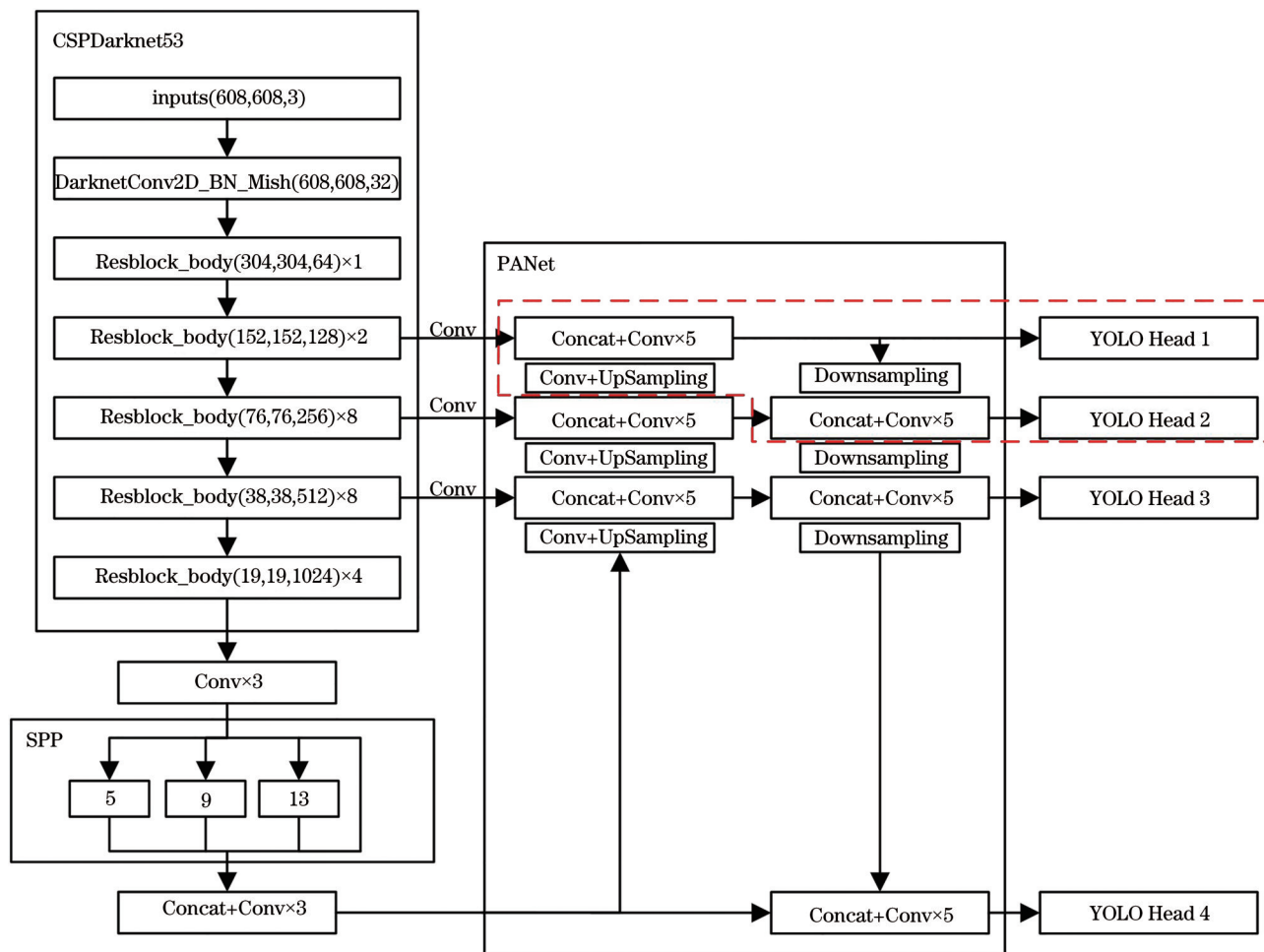


图5 优化的YOLOv4网络结构

Fig. 5 Optimized YOLOv4 network structure

为了不影响小目标分辨率,优化后的YOLOv4网络统一将输入图像剪裁成608 pixel×608 pixel的尺寸,进而在19×19,38×38,76×76,152×152四

个特征层同时对输入的无人机图像或者视频目标进行检测,在保持较高检测速度的同时,提高了对小目标的检测性能。

3.3 结合优化 YOLOv4 与 DeepSORT 的检测与跟踪模型

优化的 YOLOv4 算法能够准确有效地检测静态目标,但在实际应用场景中低空无人机机动性高,易受环境影响,需要实时稳定的动态检测,因此本实验组提出一种低空无人机的检测与跟踪模型。所提模型将优化后的 YOLOv4 检测网络作为 DeepSORT 算法的目标检测部分,两者结合进行动态检测,模型整体架构如图 6 所示。

对于输入的一段图像序列,首先将每一帧输入优化的 YOLOv4 网络以执行无人机目标检测,得到的输出为图像的一组检测框坐标(bbox)、置信度得分(conf)及类别(cls),并将这些值作为 DeepSORT 算

法的输入。接着为输入目标创建 ID,DeepSORT 算法随之创建跟踪列表,将以上检测结果写入跟踪队列以进行轨迹处理和状态估计,使用标准卡尔曼滤波器预测每个目标下一帧的状态。随后引入级联匹配策略对检测目标与跟踪目标进行匹配,分别进行运动信息与外观信息的匹配。对于匹配成功的结果,用检测到的边界框来更新目标状态并输出目标中心位置,并且对于每个轨迹,记录自上一次成功匹配以来的帧数,当大于预先设定的阈值,则判定对目标的跟踪已结束。而对于未匹配的检测器和未匹配的跟踪器,重新进行 IOU 匹配,并再次使用匈牙利算法进行匹配指派,若出现匹配成功结果则创建新的目标 ID。执行以上流程直至全部图像序列处理完成。

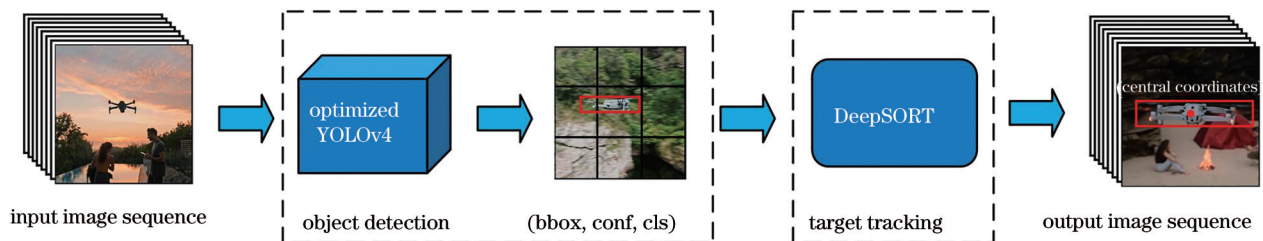


图 6 结合优化 YOLOv4 与 DeepSORT 的检测与跟踪模型

Fig. 6 Detection and tracking model based on a combination of optimized YOLOv4 and DeepSORT

4 结果与分析

优化的 YOLOv4 算法提高了对小目标的检测能力,与 DeepSORT 算法相结合可以高精度、快速地检测出目标,并提高跟踪的稳定性。为了验证所提检测跟踪模型的整体性能,将优化后的 YOLOv4 算法分别与原始 YOLOv4 算法和 YOLOv3 算法进行对比实验,然后结合 DeepSORT 进行无人机目标跟踪,测试跟踪效果。

至今尚无公开的无人机数据集,因此本实验组自行采集与构建数据集进行模型的训练和评估。构建的低空无人机数据集 LARotorcraft 由各种型号的旋翼无人机构成,包含各种飞行姿态,如盘旋、倾斜、高速起降等,共有原始图片 1006 张。采用 labelImg 图片标注工具按照 visual object classes (VOC) 数据集格式对原始图片进行人工标注。为了提升研究方法的检测精度,采取数据增强^[27]的方式对原始无人机图片进行水平镜像、模糊、添加噪音、HSV 色彩变换等处理,每张图片随机采取其中的一种图像增强方法;其次,考虑无人机飞行的不稳定性,随机对每张图片进行 $\pm 10^\circ$ 或 $\pm 20^\circ$ 的小幅

度旋转并重新调整标注框。扩充后的数据集共包含 3017 张图片,按照 9:1 的比例分割为训练集(2747 张)与测试集(270 张)。

实验在 Windows 10 操作系统下进行,采用 Keras 深度学习框架。CPU 为 Intel Xeon E5-2678 v3@2.50 GHz,内存为 12 GB,显卡为 NVIDIA GeForce GTX1080 Ti,运行库版本为 Cuda 9.0 与 CuDnn 7.3.1。

4.1 YOLO 网络训练

基于 LARotorcraft 数据集进行目标检测网络的训练,因为优化后的 YOLOv4 网络特征图尺度增加,所以需要根据特征图尺度重新进行先验框的分配,将 K-means 生成的 8 个先验框平均分配到 4 个尺度,调整后的先验框值如表 1 所示,重新聚类后平均重叠率提升了 6.52%。

对优化后的 YOLOv4 网络进行训练,采用 Adam 优化器进行优化,设置输入图像大小为 608 pixel \times 608 pixel,批大小设置为 8,动量为 0.9,衰减系数为 0.0002,最大迭代次数为 60000。初始学习率设为 0.0001,损失函数趋于稳定后,当迭代次数为 10000 和 15000 时,分别将学习率降低到

表 1 LARotorcraft 数据集的先验框值

Table 1 Value of anchor in LARotorcraft dataset

Feature Level	Size of feature map	Value of anchor
Feature Level 1	19×19	(485,493),(495,279)
Feature Level 2	38×38	(128,110),(241,186)
Feature Level 3	76×76	(36,75),(76,55)
Feature Level 4	152×152	(12,16),(19,36)

0.00005 和 0.000025,使得损失函数进一步收敛,迭代到 37000 次后,将学习率调整为 10^{-6} 。训练过程的损失曲线如图 7 所示。

优化的 YOLOv4 网络在经过 400 个 epoch 的训练后,损失值基本趋于稳定;当训练至 1200 个 epoch 后,Loss 收敛至 2 左右,训练效果较为理想。

4.2 实验结果与分析

4.2.1 目标检测对比实验

为验证优化 YOLOv4 算法对低空无人机目标的检测能力,分别对原 YOLOv3、YOLOv4 网络和优化后的网络在 LARotorcraft 数据集上进行对比实验,主要比较不同算法对低空小目标无人机的检测

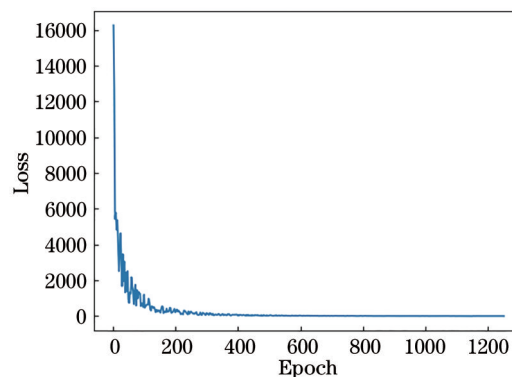


图 7 损失函数曲线

Fig. 7 Loss function curve

性能。

为保证测试结果的有效性, YOLOv3 与 YOLOv4 网络的训练均采用 Keras 深度学习框架,设定与上述相同的训练参数并迭代至损失函数收敛。在实验中,分别采用平均精度(AP,包括 AP50 和 AP75,即 IOU 为 50 和 75 的 AP 值)与检测速度进行定量分析,表 2 为 3 种算法的对比结果,其中粗体表示最优数据。

表 2 不同检测算法性能对比

Table 2 Performance comparison of different detection algorithms

Detection algorithm	AP / %	AP50 / %	AP75 / %	Detection speed / (frame·s ⁻¹)
YOLOv3	69	77.1	41.5	8.9
YOLOv4	75.6	93.2	86.5	12.7
Optimized YOLOv4	77.2	97.1	90.1	10.8

表 2 的结果表明,所提检测算法的 AP 和检测速度分别为 77.2% 和 10.8 frame·s⁻¹,相对于 YOLOv3 算法在检测精度与速度上都有大幅提升。与 YOLOv4 算法相比,优化后的 YOLOv4 算法对低空无人机目标的检测精度由 75.6% 提高到 77.2%,

检测结果如图 8 所示。在 IOU 为 0.5 的情况下,优化后的 YOLOv4 算法检测精度达到 97.1%,较 YOLOv4 算法增加了 3.9 个百分点。尽管由于特征层的增加,在检测速度上较 YOLOv4 稍有下降,但仍比 YOLOv3 算法提升了 1.9 frame·s⁻¹。

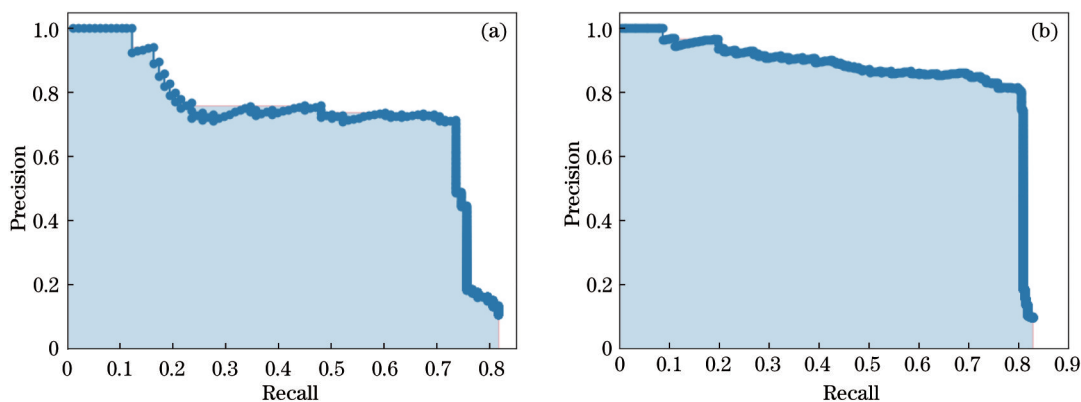


图 8 检测结果。(a) YOLOv4;(b)优化的 YOLOv4

Fig. 8 Test results. (a) YOLOv4;(b) optimized YOLOv4

选取 3 组低空小目标无人机图片,利用上述 3 种检测算法对同一场景目标进行检测,可视化检测效果如图 9 所示。从图 9 可以看出,YOLOv3 算法出现了

漏检且检测框大小具有误差,YOLOv4 算法检测出了小目标但置信度很低,而所提算法有效地检测出了小目标无人机并具有高置信度,检测精度有很大提升。

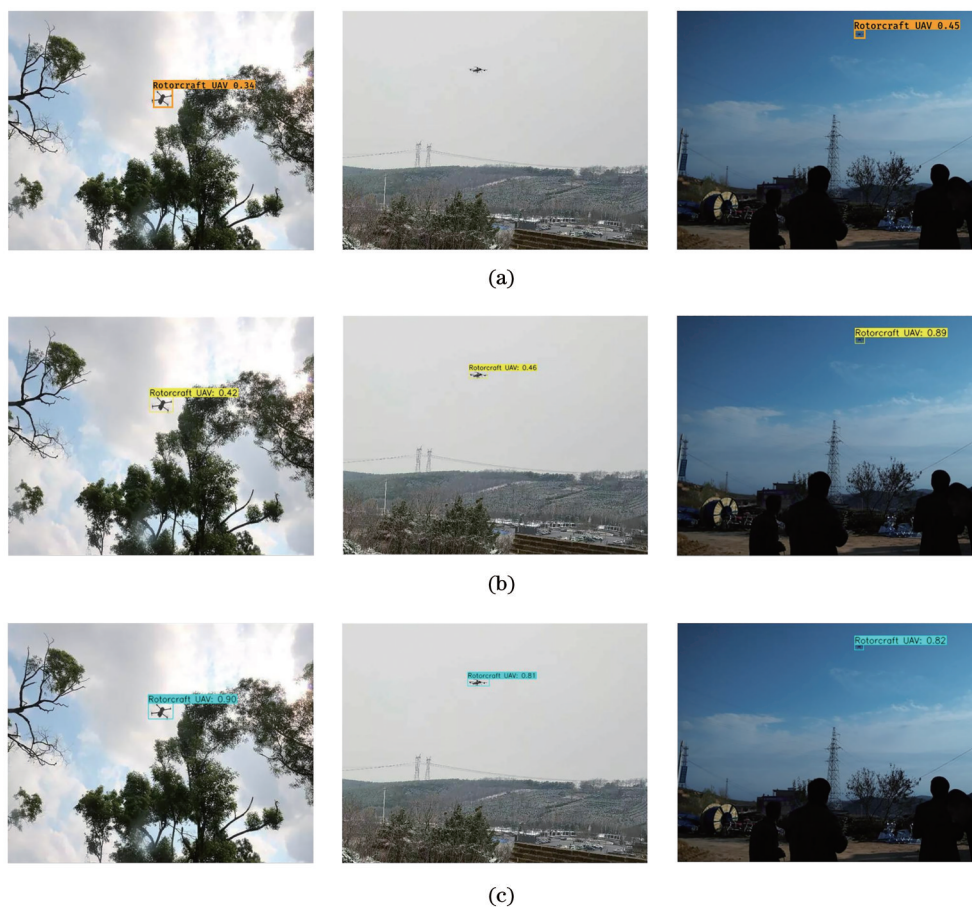


图 9 可视化小目标检测效果对比。(a) YOLOv3;(b) YOLOv4;(c) optimized YOLOv4

Fig. 9 Visual detection effects of small targets. (a) YOLOv3; (b) YOLOv4; (c) optimized YOLOv4

4.2.2 低空无人机的实时跟踪

使用分辨率为 1200×720 、帧率为 24 frame/s 且独立于训练集的视频作为测试对象。以 DJI Mini 2 为测试无人机,将视频对象输入第 3.3 节设计的检测跟踪模型中,优化 YOLOv4-DeepSORT 在测试平台达到 25 frame/s 以上的检测速度,对 56~58 三帧连续帧和 118、132、141 三帧离散帧的跟踪效果进行可视化,实线框代表检测框,虚线框代表跟踪框,效果如图 10 所示。

测试结果表明,所提算法具有较高的跟踪稳定性,在低空环境下运行效果好,且能够满足实时性要求,不仅可以准确地实现每帧中多个无人机对象的检测,而且在对同一目标进行跟踪时与之建立对应的关系,在检测框出现漏检的情况下跟踪器可以有效地框出目标对象(图 10 frame 57)。DeepSORT 跟踪器有效地解决了检测器对个别帧漏检的问题,

提升了模型的整体精度。此外,模型可以根据 DeepSORT 跟踪器提供的 ID 信息对不同的无人机进行区分并统计出现的数量。

4.2.3 结果分析

所提优化 YOLOv4 算法对小目标检测的平均召回率达到 72.2%,平均精度较 YOLOv4 与 YOLOv3 网络提高了 1.6 个百分点和 8.2 个百分点,可以有效地检测出低空场景下的小目标,避免了对小目标的漏检和错检情况。该算法与 DeepSORT 算法相结合进一步提高了逐帧检测的精确度,并且可以实现多无人机目标的跟踪与计数。所提算法可检测的最小无人机尺寸在原图中小于 $10 \text{ pixel} \times 10 \text{ pixel}$,对硬件进行优化可以满足不低于 25 frame/s 的实时性要求,实现了实时目标检测与跟踪。



图 10 优化 YOLOv4-DeepSORT 方法跟踪效果可视化

Fig. 10 Visual tracking performance of optimized YOLOv4-DeepSORT method

5 结 论

提出了一种基于优化 YOLOv4 的检测与跟踪算法,并将其应用于低空无人机目标的动态检测,实现了对低空无人机的实时检测与跟踪。针对小目标检测精度低的问题,在深度卷积神经网络 YOLOv4 的基础上对特征融合模块进行改进,将第 2 个残差块输出的特征图进行下采样与堆叠等操作,构建了 4 尺度特征图的检测网络,并且基于构建的数据集对先验框重新聚类,生成与真实目标更加匹配的候选框。在此基础上,将优化的检测网络与 DeepSORT 多目标跟踪算法相结合,对低空无人机目标实现高精度的检测与实时、稳定跟踪。将所提算法分别在 LARotorcraft 数据集与测试视频上进行了验证,实验结果表明,所提模型具有实时性和较高的准确性,泛化能力强,有广阔的应用前景。虽然所提模型运行效果较好,但多尺度特征融合的网络结构增加了计算代价,如何在降低精度的情况下进一步优化网络结构以提高检测速度与环境适应度是未来研究的重点。

参 考 文 献

- [1] 白龙, 路紫, 杜欣儒, 等. 城市区域(超)低空空域无人机活动通道划设规则与方法[J]. 地球科学进展, 2016, 31(11): 1197-1204.
Bai L, Lu Z, Du X R, et al. Rules and methods of UAV activities' aerial lanes design for(ultra) low airspace in regional areas[J]. Advances in Earth Science, 2016, 31(11): 1197-1204.
- [2] Klare J, Biallawons O, Cerutti-Maori D. Detection of UAVs using the MIMO radar MIRA-CLE Ka [C]//Proceedings of EUSAR 2016: 11th European Conference on Synthetic Aperture Radar, June 6-9, 2016, Hamburg, Germany. New York: VDE, 2016: 731-734.
- [3] Busset J, Perrodin F, Wellig P, et al. Detection and tracking of drones using advanced acoustic cameras [J]. Proceedings of SPIE, 2015, 9647: 96470F.
- [4] Nguyen P, Ravindranatha M, Nguyen A, et al. Investigating cost-effective RF-based detection of drones [C]//DroNet '16: Proceedings of the 2nd Workshop on Micro Aerial Vehicle Networks, June 26, 2016, Singapore. New York: ACM Press, 2016: 17-22.
- [5] Zou Z X, Shi Z W, Guo Y H, et al. Object detection in 20 years: a survey[EB/OL]. (2019-05-16)[2021-07-24]. <http://arxiv.org/abs/1905.05055v2>.
- [6] 王伟锋, 金杰, 陈景明. 基于感受野的快速小目标检测算法[J]. 激光与光电子学进展, 2020, 57(2): 021501.
Wang W F, Jin J, Chen J M. Rapid detection algorithm for small objects based on receptive field block[J]. Laser & Optoelectronics Progress, 2020, 57(2): 021501.
- [7] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2014, Columbus, OH, USA. New York: IEEE Press, 2014: 580-587.
- [8] Ren S Q, He K M, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.
- [9] Dai J F, Li Y, He K M, et al. R-FCN: object detection via region-based fully convolutional networks [C]//Advances in Neural Information Processing

- Systems 29: Annual Conference on Neural Information Processing Systems 2016, December 5-10, 2016, Barcelona, Spain. New York: ACM Press, 2016: 379-387.
- [10] Liu W, Anguelov D, Erhan D, et al. SSD: single shot MultiBox detector[M]//Leibe B, Matas J, Sebe N, et al. Computer vision-ECCV 2016. Lecture notes in computer science. Cham: Springer, 2016, 9905: 21-37.
- [11] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 779-788.
- [12] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 6517-6525.
- [13] Redmon J, Farhadi A. YOLOv3: an incremental improvement[EB/OL]. (2018-04-08) [2021-07-24]. <http://arxiv.org/abs/1804.02767>.
- [14] Bochkovskiy A, Wang C Y, Liao H Y M. YOLOv4: optimal speed and accuracy of object detection[EB/OL]. (2020-04-23) [2021-07-24]. <https://arxiv.org/abs/2004.10934>.
- [15] 马旗, 朱斌, 程正东, 等. 基于双通道的快速低空无人机检测识别方法[J]. 光学学报, 2019, 39(12): 1210002.
- Ma Q, Zhu B, Cheng Z D, et al. Detection and recognition method of fast low-altitude unmanned aerial vehicle based on dual channel[J]. Acta Optica Sinica, 2019, 39(12): 1210002.
- [16] 鞠默然, 罗海波, 王仲博, 等. 改进的 YOLO V3 算法及其在小目标检测中的应用[J]. 光学学报, 2019, 39(7): 0715004.
- Ju M R, Luo H B, Wang Z B, et al. Improved YOLO V3 algorithm and its application in small target detection[J]. Acta Optica Sinica, 2019, 39(7): 0715004.
- [17] 孙迎春, 潘树国, 赵涛, 等. 基于优化 YOLOv3 算法的交通灯检测[J]. 光学学报, 2020, 40(12): 1215001.
- Sun Y C, Pan S G, Zhao T, et al. Traffic light detection based on optimized YOLOv3 algorithm[J]. Acta Optica Sinica, 2020, 40(12): 1215001.
- [18] Tan L, Dong X, Ma Y X, et al. A multiple object tracking algorithm based on YOLO detection[C]//2018 11th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), October 13-15, 2018, Beijing, China. New York: IEEE Press, 2018: 18439698.
- [19] Zhang L, Fang Q. Multi-target tracking based on target detection and mutual information[C]//2020 Chinese Control and Decision Conference (CCDC), August 22-24, 2020, Hefei, China. New York: IEEE Press, 2020: 1242-1245.
- [20] Jiang M J, Shimasaki K, Hu S P, et al. A 500-fps pan-tilt tracking system with deep-learning-based object detection[J]. IEEE Robotics and Automation Letters, 2021, 6(2): 691-698.
- [21] Wojke N, Bewley A, Paulus D. Simple online and realtime tracking with a deep association metric[C]//2017 IEEE International Conference on Image Processing, September 17-20, 2017, Beijing, China. New York: IEEE Press, 2017: 3645-3649.
- [22] Liu S, Qi L, Qin H F, et al. Path aggregation network for instance segmentation[C] //2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 8759-8768.
- [23] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 936-944.
- [24] Bewley A, Ge Z Y, Ott L, et al. Simple online and realtime tracking[C]//2016 IEEE International Conference on Image Processing, September 25-28, 2016, Phoenix, AZ, USA. New York: IEEE Press, 2016: 3464-3468.
- [25] 许凯, 邓超. 基于改进 YOLOv3 的安全帽佩戴识别算法[J]. 激光与光电子学进展, 2021, 58(6): 0615002.
- Xu K, Deng C. Research on helmet wear identification based on improved YOLOv3[J]. Laser & Optoelectronics Progress, 2021, 58(6): 0615002.
- [26] 刘峰, 郭猛, 王向军. 基于跨尺度融合的卷积神经网络小目标检测[J]. 激光与光电子学进展, 2021, 58(6): 0610012.
- Liu F, Guo M, Wang X J. Small target detection based on cross-scale fusion convolution neural network[J]. Laser & Optoelectronics Progress, 2021, 58(6): 0610012.
- [27] Jeong H J, Park K S, Ha Y G. Image preprocessing for efficient training of YOLO deep learning networks [C]//2018 IEEE International Conference on Big Data and Smart Computing (BigComp), January 15-17, 2018, Shanghai, China. New York: IEEE Press, 2018: 635-637.