

YOLOX-IM:一种无人机航拍视频的轻量化 交通参数提取模型*

刘军黎 刘晓锋 邱洁 衣雨玮

(天津职业技术师范大学汽车与交通学院 天津 300222)

摘要:在无人机航拍的过程中,背景更广阔,目标的尺寸更小,种类更复杂。提出一种基于YOLOX-s的轻量化无人机航拍目标检测算法YOLOX-IM。首先,为了提高该模型检测小目标的性能,通过使用切片辅助推理(slicing aided hyper inference, SAHI)算法以及坐标修正矩阵对训练集进行预处理和数据增强。然后,在路径聚合网络(path aggregation network, PAN)中引入一个浅层特征图以及超轻量级子空间注意模块,并添加一个检测头对小物体进行检测;最后,对边界回归的损失函数进行了优化。在VisDrone2019数据集的消融实验结果表明,所提出的模型检测精度与基础YOLOX-s相比高了8.13%;模型体积4.55 MB,相较于原模型下降67.14%。利用该模型在中国天津市渌水道进行实地交通监测的交通参数提取实验,在设定的场景中,当无人机航拍高度为50 m时,该模型的交通提取参数精度最高,达到96.14%。

关键词:无人机;车辆检测与跟踪;YOLOX模型;深度学习

中图分类号: TP391 **文献标识码:** A **国家标准学科分类代码:** 520.604

YOLOX-IM: A lightweight traffic parameter extraction model for UAV aerial images

Liu Junli Liu Xiaofeng Qiu Jie Yi Yuwei

(School of Automotive and Transportation, Tianjin University of Technology and Education, Tianjin 300222, China)

Abstract: In the process of UAV aerial photography, the background is broader and the targets are smaller in size. In this paper, we propose a lightweight target detection algorithm YOLOX-IM for UAV aerial photography based on YOLOX-s. First, to improve the performance of small target detection, the training set is preprocessed and data is enhanced by using a slicing aided hyper inference (SAHI) algorithm as well as a coordinate correction matrix. Then, a shallow feature map as well as an ultra-lightweight subspace attention module are introduced in the path aggregation network (PAN), and a detection head is added for small object detection. Finally, the loss function of the boundary regression is optimized. The experimental results on the VisDrone2019 dataset show that the proposed model has 8.13% higher detection accuracy compared with the YOLOX-s; compared with the original model, the model volume is significantly reduced to 4.55 MB, which is 67.14% lower than the original model. Next, the model is used to conduct traffic parameter extraction for field traffic monitoring in Tianjin, Lushui road, China. The study indicates that the model has the highest traffic extraction parameter accuracy of 96.14% at the UAV altitude of 50 m.

Keywords: unmanned aerial vehicle; vehicle detection and tracking; YOLOX model; deep learning

0 引言

近年来,随着无人机的续航能力、悬停性能、云台平稳

性的显著提升,以及高精度定位、高带宽、低延时通信技术的使用,很多学者开展了基于无人机视频的目标检测技术研究。Wang等^[1]采用一致性目标跟踪、运动背景处理,以

收稿日期:2022-10-25

* 基金项目:天津市智能交通技术创新团队重点培养专项(XC202028)、天津市科技计划(22YDTPJC00120)、天津市高等学校科技发展基金(2021KJ018)项目资助

及神经网络目标检测等手段,从无人机视频中进行车辆检测。Chen 等^[2]开发了一个集成检测器来检测目标区域内的车辆,然后采用核化相关滤波器,从无人机视频中跟踪车辆。但在无人机航拍的过程中,视频背景更广阔,且处于动态的变化过程当中,导致传统的视频运动目标检测算法难以适用。

深度学习算法具有强大的特征提取能力,可以通过目标结构、色彩等视觉特征描述检测目标,而不依赖于目标运动信息进行目标检测,因此适用于无人机视频的目标检测与跟踪研究,从而实现交通事故的判别。目前,常用的深度学习目标检测算法分为两类,基于感兴趣区域(region of interest, ROI)的方法和基于回归的方法。基于感兴趣区域的方法将目标检测分为提取 ROI、进行目标检测分类两个阶段,有代表性的算法有 Mask R-CNN^[3]、Fast R-CNN^[4]、Faster R-CNN^[5]等。这些双阶段检测算法的优点是目标检测平均精度高、错检率和漏检率低,但也存在着推理时间长、计算量大的缺点,难以达到实时性要求。基于回归的方法又称为单阶段检测算法,以 YOLO^[6]、YOLO9000^[7]、SSD^[8]、YOLOv3^[9]、YOLOv4^[10]、YOLOv5 为代表。YOLO 算法不需要提取感兴趣位置,而是直接获得目标位置和类别。具有检测速度快、模型复杂度低等优点,能够满足实时检测的需求。Liu 等^[11]提出了路径聚合网络(path aggregation network, PAN),利用低级特征的空间信息进一步提高对小目标的检测性能。彭继慎等^[12]在 YOLO 网络中实现了模型压缩。李仁鹰等^[13]在 YOLO 网络中实现了轻量化处理。Etten 等^[14]提出了一种基于 YOLOv2 的卫星图像多尺度目标检测的方法,其使用了一个密集的预测网格来提高网络对密集的小物体的分类能力,并且使用了矩阵旋转和数据增强处理对象方向的多样性。化嫣然等^[15]采用 YOLOv3 深度学习模型,提出了一种针对遥感图像中的大中型尺寸目标检测的优化 YOLOv3-Tiny 算法,但是其对遥感图像中的小目标检测精度有待提高。Tan 等^[16]基于 YOLOv4 深度学习模型,设计了空间注意力机制,提高了车辆的检测精度,但是对小目标、有阴影遮挡的车辆检测时漏检率较高。

当前无人机航拍在交通监测领域主要面临如下难题:1)航拍背景下的小目标都处于运动状态,且运动偏移明显,小目标检测难度大。传统目标检测算法都针对中大目标进行建模设计,并且模型体积庞大,收敛速度较慢,对小目标的检测存在漏洞。2)绝大部分算法都是基于公共数据集进行仿真改良,存在以优良结果为导向的过度优化问题。并没有进行实地的交通监测实际应用去检验算法在真实环境下的目标检测能力以及交通参数提取能力。

为了解决上述问题,本文提出一种无人机航拍视频的轻量化交通参数提取模型 YOLOX-IM。

对于航拍图像复杂的背景问题,在 PAN 中加入了超轻量级子空间注意力模块(ultra-lightweight subspace attention module, ULSAM),特别突出了物体的特征,削弱了背

景信息。对于无人机航空图像中大量小目标的问题,首先利用切片辅助推理算法(slicing aided hyper inference, SAHD)以及坐标修正矩阵对高分辨率无人机航空图像进行预处理,将图像分割成多个重叠的切片,使小目标在生成的图像上占据较大的像素区域,使得浅层特征图具有丰富的空间信息。接着在路径聚合网络中添加一个检测头来处理浅层特征图中的小目标。对于轻量化和高精度问题,对边界回归损失函数进行了优化,使边界回归更快、更准确。连接 DeepSort 算法模型,搭建 YOLOX-IM-DeepSort 模型进行道路车辆检测与跟踪,利用高精度车载诊断系统(on-board diagnostics, OBD)、全球卫星导航系统(global navigation satellite system, GNSS)车载数据,开展航拍交通参数提取模型的实证研究,以给予本文模型更强的可解释性与适用性。

1 YOLOX-IM 模型的搭建

1.1 YOLOX-s 模型框架

YOLOX 深度学习框架于 2021 年提出^[17],其在目标检测比赛中优势十分明显。小目标由于分辨率低、体积小,很难被检测到。而 YOLOX-s 算法架构提高了 Anchor-free 的数量,也提高了交并比(intersection-over-union, IOU),更加强了 Feature Map 的信息,不仅如此,架构中使用的下采样与上采样结合,Feature Map 的采样分辨率提高,对小目标检测能力进一步加强。为了提高目标检测的精度与速度,本文采用 YOLOX-s 模型作为基准模型进行上述策略的改良构成 YOLOXs-IM 模型,开展无人机视频的车辆检测与分类,并结合 DeepSort 算法进行车辆跟踪。进一步实现交通参数提取模型的构建。

CSPDarknet 是 YOLOX-s 的主干特征提取网络,由 DARK1、DARK 2、DARK 3、DARK 4 和 DARK 5 构成。主干网络结构如图 1 所示。图 1 中,可见每个部分之间使用下降采样将特征图的尺寸缩小 1/2,同时将通道数量增加 1 倍。CSPDarknet 使用焦点模块,残差网络,跨级部分网络(CSPnet)和空间金字塔池(SPP)来提取图像特征。其中,焦点模块可以减少增加参数的数量,提高操作效率而不丢失信息。残差网络处理了通过增加神经网络的深度而带来的梯度消失问题。CSPnet 增强了网络学习特征的能力,同时减少了计算工作量。SPP 模块利用不同池化核的最大池化操作来提高网络的感知野。YOLOX-s 的目标检测过程如图 2 所示。首先,通过 CSPDarknet 将图像输入到主干网络进行特征提取,主干网络网络提取 3 层有效特征图 $\{C_3, C_4, C_5\}$ 。然后,将有效的特征图传递到颈部结构中,通过 PAN 对特征进行增强,得到 3 层增强的特征图 $\{N_3, N_4, N_5\}$ 。在这里,FPN 将自上而下的高级语义信息传播到浅层特征图中,在多个尺度上增强语义信息,但对对象的定位信息较弱。因此,包含自底向上的路径,通过将丰富的空间信息从浅层转移到更高层次的特征地图,增强了在多个尺度上的定位。最后,利用解耦头对 3

个增强的特征图进行分类和回归,提高了模型的收敛速度,提高了检测精度。

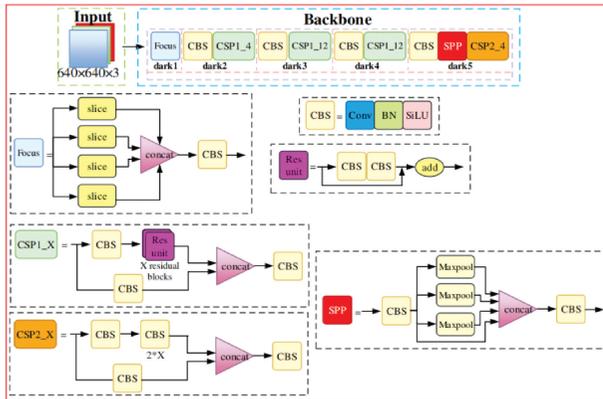


图1 YOLOX-s的主干网络结构

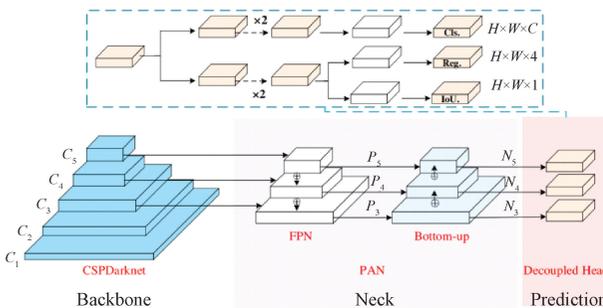


图2 YOLOX-s的目标检测过程

1.2 数据增强策略

传统的检测方法在检测小物体方面不是很有效, Akyon等^[18]提出了SAHI方法,将图像分割成多个重叠的切片,使小物体占据更多的像素在这些切片中,从而提高了模型检测小物体的能力。本文使用SAHI将训练图像裁剪成大小为160个重叠像素的640×640的切片,并去除只包含背景的切片。

数据增强可以增加训练样本的数量和多样性,提高模型的鲁棒性。光度失真和几何失真是两种最常用的数据增强方法。光度失真可以调整图像的亮度、色度、对比度和饱和度。为了执行几何失真,图像可以被随机缩放、裁剪、翻转、旋转或进行其他转换。

还有一些数据增强方法,如Cutout、Mixup、CutMix和Mosaic。Cutout随机裁剪训练图像的一个正方形区域,并用零填充切出图像;Mixup在数据集中随机选择两个样本及其相应的标签,并使用一定的比例对它们进行汇总,生成新的样本和标签;CutMix从数据集中随机选择两幅图像,然后将一幅图像的裁剪部分叠加到另一幅图像上;Mosaic随机裁剪4幅图像,然后将它们拼接成一个新的图像,丰富了图像的背景。

本文使用光度失真、几何失真、Mixup算法^[19]和Mosaic算法^[20]来进行数据增强。

1.3 增大检测头策略

由于VisDrone数据集^[21]有许多非常小的对象, YOLOX-s的3个解耦头在检测这些对象时会导致大量的遗漏和错误检测。而浅层特征图具有较小的感受野和较强的空间信息,适用于检测小物体。YOLOX-s的主干网络通过CSPDarknet提取特征,经过5轮下降采样得到5层特征图 $\{C_1, C_2, C_3, C_4, C_5\}$,通道数量增加了1倍,而特征图的大小减少了1/2。特征图 C_1 包含更详细的信息,但也包含高水平的噪声和背景信息。因此,本文提出的方法将低水平的高分辨率特征图 C_2 引入FPN中,将三级有效特征图 $\{C_3, C_4, C_5\}$ 与自上而下传输的特征进行融合,这些方法通过将强语义特征从高级转移到浅层,从而增强了多尺度上的语义表达。这些增强过的特征图 $\{P_2, P_3, P_4, P_5\}$ 以自下而上的方式将浅层特征图的强空间特征转移到更高的层次,提高了在多尺度上的定位能力。同时本文还增加了一个用于小目标检测的解耦头,改进后的网络结构如图3所示。

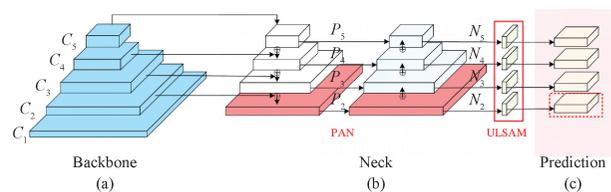


图3 YOLOX-IM主干网络架构

如图3所示, YOLOX-IM模型的主干网络结构采用CSPDarknet提取多尺度特征图,在Neck层使用PAN和ULSAM增强多尺度特征图,使FPN与 C_2 融合得到 P_2 , P_2 自下而上融合得到 N_2 。 P_2 和 N_2 用红色方块表示。在PAN结构后面添加了ULSAM结构削弱复杂的背景信息,如红框所示。在预测层中,红色虚线框表示小目标检测的解耦头部。

1.4 注意力机制策略

注意机制类似于人类视觉注意机制,它从大量特征中提取更关键的特征,增强关键特征,减少对其他非关键特征的注意,甚至忽略无用的特征,从而提高网络的性能。虽然现有的注意机制可以有效地提高网络的性能,但它增加了算力需求,不适用于轻量级网络。为此, Saini等^[22]提出了ULSAM,它可以在减少参数数量的情况下有效地学习特征子空间中的跨通道信息。ULSAM的结构如图4所示。输入的特征映射在通道上被分割成 G 个互斥的子空间,每个子空间都有 G 个特征映射,并为每个子空间推导出一个不同的注意映射。

本文提出的方法将ULSAM添加到PAN结构中,用于突出对象特征,削弱背景特征。然后,将增强的对象特征输入到解耦的头部进行分类和回归,有效地提高了检测精度。改进后的颈部结构如图4所示。

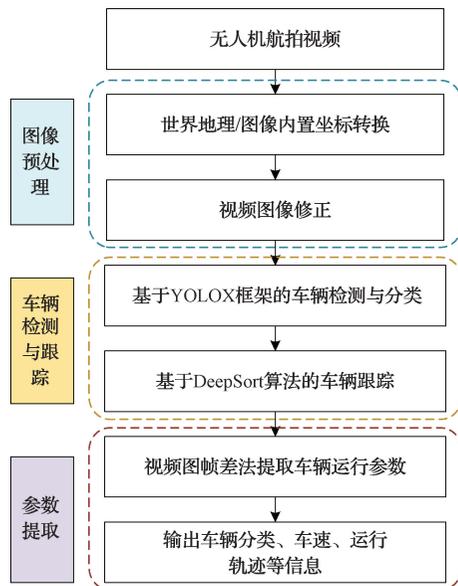


图6 无人机视频数据提取流程

叠加的坐标变化以及图像拖影。因此,需要对航拍视频进行图像预处理^[25]。

1) 坐标变化

在现实世界坐标和视频图像坐标之间,必须进行一定的变换,本文所做的是人工标记点。坐标变换可以通过现实世界中的标记坐标和图像中的人工标记坐标建立等式关系来实现,坐标变化模型为:

$$\begin{pmatrix} u \\ v \\ p \end{pmatrix} = T \cdot \begin{pmatrix} x_0 \\ y_0 \\ z_0 \end{pmatrix} = \begin{pmatrix} A & D & G \\ B & E & H \\ C & F & I \end{pmatrix} \cdot \begin{pmatrix} x_0 \\ y_0 \\ z_0 \end{pmatrix} \quad (11)$$

式中: $[u \ v \ p]$ 是真实世界中目标的坐标; $[x_0 \ y_0 \ z_0]$ 是利用参考系所做的视频中目标的坐标; T 是真实世界坐标和视频图像坐标之间的变换矩阵。在本文实验中,无人机最大高度低于 120 m,所选取的道路长度小于或等于 150 m,无人机的道路长度和飞行高度比远远大于斜率差,所以空间三维问题可以转换为平面变换计算,即矩阵 T 中的 C, F, I 分别等于 0, 0 和 1。因此,式(11)可以写为:

$$\begin{pmatrix} u \\ v \\ I \end{pmatrix} = T \cdot \begin{pmatrix} x_0 \\ y_0 \\ I \end{pmatrix} = \begin{pmatrix} A & D & G \\ B & E & H \\ 0 & 0 & I \end{pmatrix} \cdot \begin{pmatrix} x_0 \\ y_0 \\ I \end{pmatrix} \quad (12)$$

其中, A, B, D, E, G, H 是需要确定的参数。矩阵 T 至少需要 3 个对应的标记点来确定。标记点越多,变换矩阵的鲁棒性则越强。理想情况下,道路上的标记点足够多就能通过式(12)得到的矩阵 T ,构建出真实坐标和视频图像坐标之间的变换矩阵公式。

2) 图像修正

由于风力作用或无人机的自适应姿态模式的改变,无人机悬停时不可避免地会轻微移动和震动。因此,需要确

定变换关系,以统一不同视频帧之间随时间变化的坐标,并将其定义为图像修正。式(13)表示变换过程。

$$\begin{pmatrix} x_n \\ y_n \\ I \end{pmatrix} = W \cdot \begin{pmatrix} x_0 \\ y_0 \\ I \end{pmatrix} = \begin{pmatrix} a_1 & a_2 & a_3 \\ a_4 & a_5 & a_6 \\ 0 & 0 & I \end{pmatrix} \cdot \begin{pmatrix} x_0 \\ y_0 \\ I \end{pmatrix} \quad (13)$$

式中: $[x_0 \ y_0]$ 为参照坐标系 N_0 (初始点) 的坐标; $[x_n \ y_n]$ 为第 n 个坐标系的坐标; W 是变换矩阵,由 6 个参数组成; $a_1 \sim a_6$ 需要由不同图像中的固定标记来确定。因此,需要通过上述两个变换矩阵对航拍视频图像进行修正处理以提高后续目标检测的精度。

2.2 车辆检测和跟踪

图像预处理后,需要采用本文的算法检测、跟踪区域中的目标车辆,并利用视频帧数测速算法,估计目标车速。

1) 车辆检测

利用 YOLOX-IM 模型检测帧中的车辆物体,输出数据为车辆拟合度检测集 $N = \{N_i, i = 1, 2, 3\}$ 和轨迹集 $T = \{T_i, i = 1, 2, 3, \dots\}$ 。

2) 多目标跟踪

结合 YOLOX-IM 模型输出的数据集,利用 DeepSort 多目标跟踪算法对视频中的运行车辆进行跟踪。

2.3 车速估算

采用视频帧率速度估算法,对目标车辆的运行速度进行估算。在进行定位时视频帧数间隔越少越好,这样是为了减少算法的计算量及加快算法的执行速度。在实际操作中,因为需要结合车载数据采集的频率,所以本文的算法每隔 4 帧取一次速度估算值。视频帧率速度估算法如下:

$$v = \frac{\Delta x}{\Delta t} \quad (14)$$

式中: Δx 为所选帧区间之间的车辆位置相减得到移动的距离; Δt 对应的间隔时间。 $\Delta t = (\text{前一帧处理帧帧数} - \text{当前帧处理帧帧数}) / \text{帧率}$ 。视频里 Δx 为视频不同帧率下目标车辆的坐标连线距离,考虑到车辆坐标并不可能始终处于同一水平线上,仍需要结合式(12)和(13)进行变换修正。

3 实验设计

3.1 双实验思路

基于 YOLOX-IM-DeepSort 模型追踪车辆位置信息,根据车辆位置信息,提取车辆实时运行速度、跟踪识别精度。之后,开展不同算法之间的对比实验,以及地空数据的实证实验。

把 YOLOX-IM 一系列常用的目标检测模型进行对比实验,利用神经网络的常用评价指标进行目标检测比较,这为对比实验。并进行本文模型的实际应用,通过模型处理实时航拍视频得到的车辆速度称为实验样本值,通过 OBD、GNSS 记录的高精度车辆速度称为地面样本真值,接着进行实验样本值与地面样本真值的误差对比,这

一过程为对照实验。

3.2 对照实验方案步骤

实验主要包括空中无人机数据采集部分和地面车辆车载设备数据采集部分。在空中无人机部分,本实验采用的设备为大疆精灵 4RTK 无人机。在地面数据采集部分,采用了一辆安装有高精度车载设备的测试车辆,以收集验证所需的参考数据。GNSS 确保定位数据的精度,OBDD 可以通过读取车辆 CAN 来实时记录车辆行驶数据,并通过蓝牙连接到手机 APP 进行后期数据提取。实验系统如图 7 所示。



图 7 地面数据采集实验系统

实验的整个过程如图 8 所示,包括标记点设置、测试车辆和无人机设置。为了探讨无人机运用在交通监测领域设置多少高度可以实现更优的车辆检测精度,实验设置了 12 类不同环境的对照方案,其中分成了 3 种最常使用的高度,为了模拟真实道路中不同运行状态的车辆也对测试车辆速度进行分类,如表 1 所示。

表 1 12 类对照实验条件

实验组序	无人机拍摄 视频平均录 制时长/s	测试车辆 速度/ (km · h ⁻¹)	无人机 高度/m	测试道路 长度/m
1	15	40	30	50
2	10	60	30	50
3	23	10~40	30	50
4	18	10~60	30	50
5	19	40	50	75
6	13	60	50	75
7	27	10~40	50	75
8	21	10~60	50	75
9	27	40	100	150
10	24	60	100	150
11	33	10~40	100	150
12	28	10~60	100	150

4 实验验证

4.1 对比实验

1) 数据集说明

VisDrone2019 数据集由中国天津大学收集,其中包括 14 个城市不同天气和光照条件下的各种场景。该数据集提供了无人机在不同高度和位置拍摄的静止图像,图像

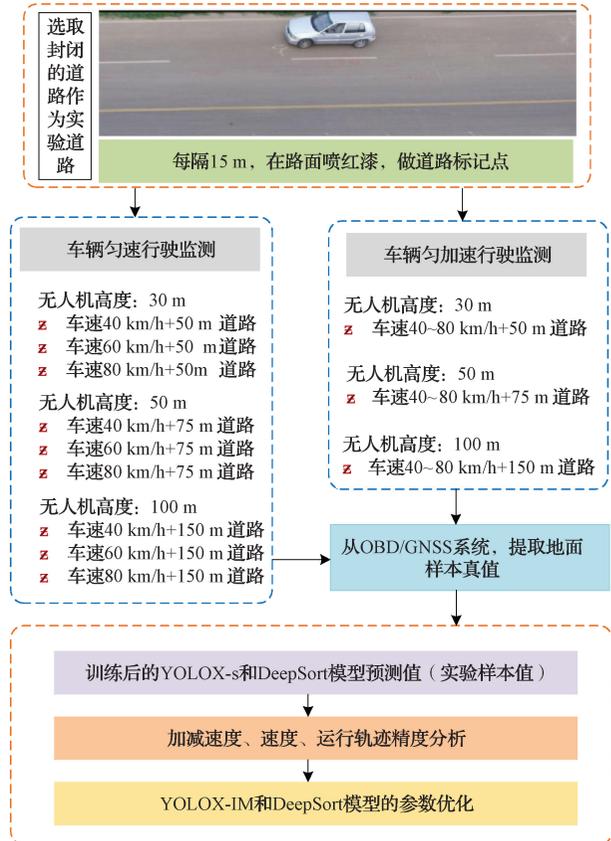


图 8 对照实验步骤流程

分辨率最高可达 2 000 × 1 500 pixels。包括 10 个类别,分别是行人、人、自行车、汽车、货车、卡车、三轮车、遮阳篷三轮车、公共汽车和汽车。其中,训练集包含 6 471 张图像,验证集包含 548 张图像,测试集开发包含 1 610 张图像。在 VisDrone 数据集中,单个图像中可以出现 902 个对象,类别和标签的分布情况如图 9 所示。为了使类别分布更加平衡,本文使用 SAHI 将 6 471 张训练图像裁剪成 640 × 640 分辨率、160 个重叠像素的切片,并去除仅包含背景的切片。接下来,使用包含较少样本类别的选定切片与原始训练图像形成新的训练集,使用 548 张图像进行验证,并通过测试 1 610 张图像获得实验结果。

2) 评价指标

使用 mAP50、AP-large、AP-mid、和 AP-small 来评估 YOLOX-IM 模型的性能。在 IoU 为 0.5 的条件下,计算 mAP50 的各类别的平均精度(AP),然后对所有类别进行平均。AP-large、AP-mid、和 AP-small 是当 IoU 在 0.5 ~ 0.95 范围内时,对 3 种不同尺度的物体(大、中、小)计算出的 AP 值。

mAP 的计算公式如下:

$$mAP = \frac{1}{n} \sum_{i=1}^n AP(i) \quad (15)$$

式中: n 为类别的总数; i 为当前类别; $AP(i)$ 为当前类别的 AP 值。

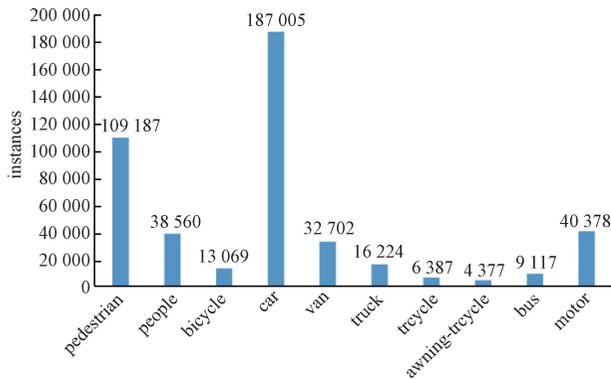


图9 每个类别的标签数量

3) 实验环境

深度学习过程在 GPU 性能较强的服务器中运行,服务器配置如下:处理器为 Intel 酷睿 I5 10400F,内存为 DDR4 32 G, GPU 为 GeForce RTX 3060 12 G。软件版本如下:Python 版本为 3.7.7, Pytorch 版本为 1.7.1, Cuda 版本为 10.1。训练采用随机梯度下降,动量值为 0.937,权重衰减因子为 5×10^{-4} ,初始学习率为 0.01,批大小设置为 4,周期设置为 150。对于几何失真,采用了随机缩放

和翻转的方法。随机缩放区间为 (0.25, 2), 输入图像在 0.25~2 随机缩放,而宽度和高度在 0.538~1.857 畸变。此外,当翻转比固定在 0.5 时,有 50% 的几率将图像向左或向右翻转。对于光度失真,色相、饱和度和亮度的失真因子分别设置为 0.1、0.7 和 0.4。在将图像从 RGB 调整为 HSV 后,改变色域。

4) 消融实验

模型消融实验结果如表 2 所示。与基线网络相比,当使用 SAHI 和坐标修正矩阵进行数据增强时,性能增强明显,大、中、小目标的检测精度分别提高了 4.7%、1.6% 和 1%。SAHI 使小物体在切片中占据更多的像素,数据的增强增加了训练样本的多样性,有效地提高了小物体的检测精度。然后,将浅层特征图 C_2 引入 FPN 进行特征融合,小目标检测精度进一步提高了 1.3%。结果表明,特征图 C_2 具有丰富的小目标空间信息,适用于无人机航空图像中的小目标检测。然后,在 PAN 阶段添加 ULSAM,得到不同的注意图,突出对象特征,弱化背景信息,再次提高了模型的检测性能。最后,将损失函数 IoU 替换为 SIoU,使边界框能够更快速、准确地定位目标,提高了网络检测小目标的精度。

表 2 改进模块的消融实验结果

基础模型 (YOLOX-s)	数据增强	C_2	ULSAM	SIoU	体积/MB	mAP50/%	AP-small	AP-mid	AP-large
✓					13.85	36.52	0.093	0.302	0.412
✓	✓				13.85	38.64	0.103	0.319	0.466
✓	✓	✓			6.7	41.88	0.114	0.331	0.485
✓	✓	✓	✓		8.2	44.03	0.126	0.354	0.478
✓	✓	✓	✓	✓	4.55	44.65	0.132	0.356	0.496

与基础 YOLOX-s 模型相比,改进的 YOLOX-IM 在体积上减少 67.14%, mAP50 增加了 8.13%, AP-small、AP-mid 和 AP-large 分别增加了 3.9%、5.4% 和 8.4%。结果表明,改进模型能够有效解决无人机航拍图像中许多小目标和复杂背景的问题,并做到了轻量化的要求。

5) 算法对比

将 YOLOX-IM 目标检测模型与 CenterNete、YOLOv3、D-A-FS SSD 和 RetinaNet 以及二阶段算法 Faster R-CNN 和 Cascade R-CNN 算法的性能进行比较。比较结果如表 3 所示,结果表明,所提出的 YOLOX-IM 模型的精度高于其他输入分辨率相同的检测模型,这说明了 YOLOX-IM 模型性能优于其他检测模型。

6) 可视化分析

本文直观地比较了基础模型 YOLOX-s 和 YOLOX-IM 在 VisDrone 数据集上的检测结果。图 10(a)所示为基础模型的检测结果,图 10(b)所示为本文改进模型的测试结果。可以看出,改进后的模型的目标检测结果比原算法的结果更丰富且准确,在图 10(b)可以明显看到更多的黄

色与深红色的小目标检测标签。所以,无论是在目标密集的环境中还是在背景复杂的环境中,本文所提出的 YOLOX-IM 目标检测模型在无人机航拍图像中对小目标更敏感,误检率和误检出率都更低。

4.2 对照实验

1) 现实实验检测

提取本文 YOLOX-IM 目标检测模型的经过训练后的权重文件,与 DeepSort 目标跟踪模型连接,构成可用于无人机航拍视频的轻量化交通参数提取的 YOLOX-IM-DeepSort 模型。对实验中的 12 类对照实验的无人机视频里的特定实验车辆进行检测与跟踪,只取输出结果中的小汽车 AP 值作为精度依据。在算法处理中,自实验车辆出现后,每 4 帧取一次置信度,检测取值过程如图 11 所示,一共选取 20 次 AP 值做对比,相关结果如图 12、13 所示。

由图 13 可知,第 6 类实验中该模型对车辆识别的置信度较其他组更高,置信度平均都高于 92%,接近 YOLOX-s 算法对于静态物体的识别置信度,也接近学习

表 3 在测试集开发上的不同算法的分类结果(mAP50)

模型	图片分辨率	所有种类精度/%	行人	人类	自行车	小汽车	三厢货车	卡车	三轮车	敞篷三轮车	公交车	摩托车
CenterNet	—	26.60	0.23	0.21	0.15	0.6	0.24	0.21	0.2	0.17	0.38	0.24
YOLOv3	768×768	41.35	—	—	—	—	—	—	—	—	—	—
	1 120×1 120	45.64	0.44	0.28	0.23	0.85	0.53	0.54	0.31	0.27	0.65	0.46
D-A-FS SSD	—	36.70	—	—	—	—	—	—	—	—	—	—
RetinaNet	—	35.59	0.27	0.13	0.14	0.59	0.50	0.54	0.25	0.30	0.59	0.24
YOLOX-s (基础模型)	640×640	36.62	0.31	0.21	0.15	0.78	0.41	0.46	0.22	0.19	0.58	0.36
YOLOX-IM (本文)	640×640	44.64	0.43	0.30	0.21	0.83	0.49	0.54	0.29	0.27	0.65	0.47
FasterR-CNN	—	33.60	—	—	—	—	—	—	—	—	—	—
CascadeR-CNN	—	43.70	0.43	0.33	0.21	0.80	0.49	0.44	0.32	0.22	0.62	0.43

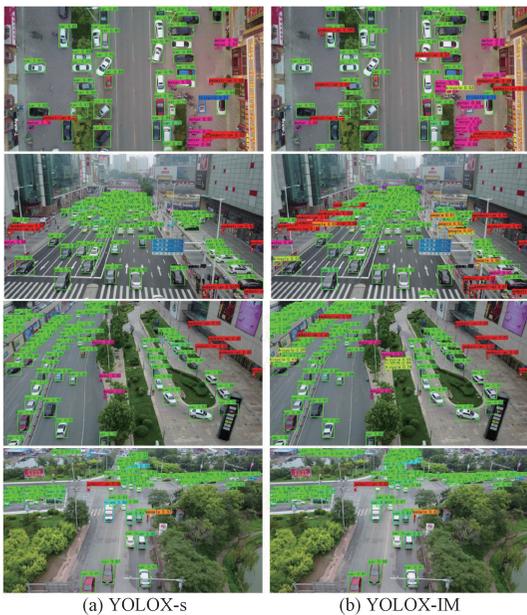


图 10 两模型可视化分析

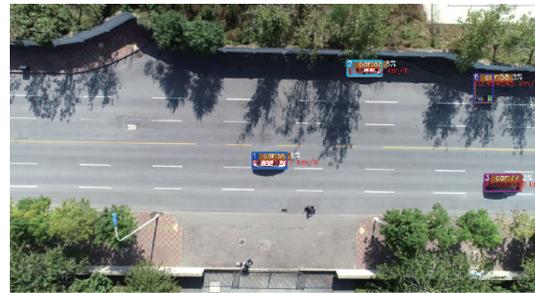
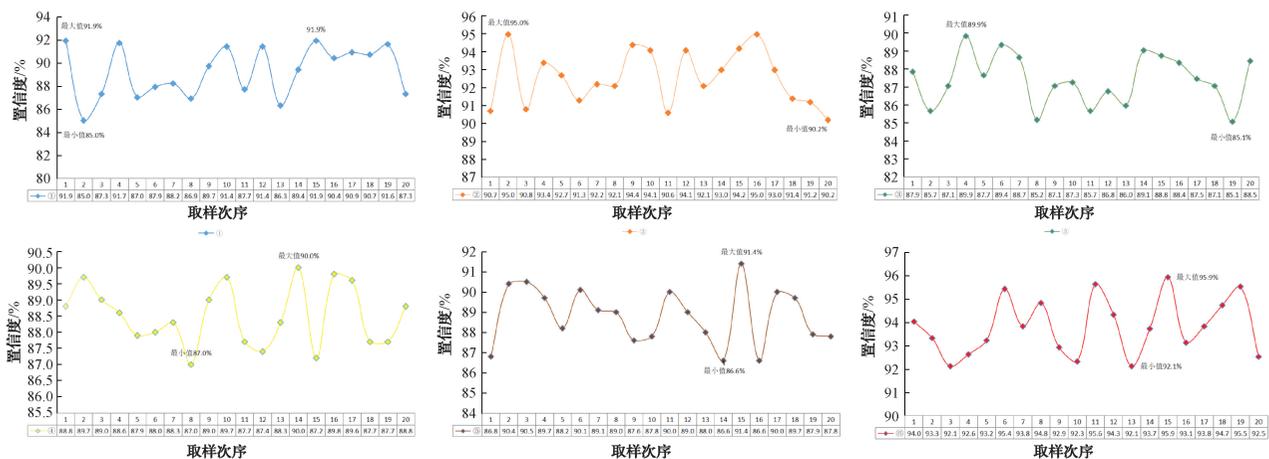


图 11 天津渌水道航拍视频检测取置信度过程

过程中的准确率。第 6 组的实验情况为无人机高度为 50 m；而第 10 组实验的置信度平均水平较低，仅在 82%~86%，第 10 组的实验情况为无人机高度为 100 m。

2) 车辆速度提取的精度

选择使用高精度 GNSS 和 OBD 数据来衡量通过视频帧率速度估算得到的无人机视频中目标车辆车速估算数据的准确性。由于 OBD 数据是在 5 Hz 的频率下记录的，而 GNSS 是在 20 Hz 的频率下产生的，因此本文对这



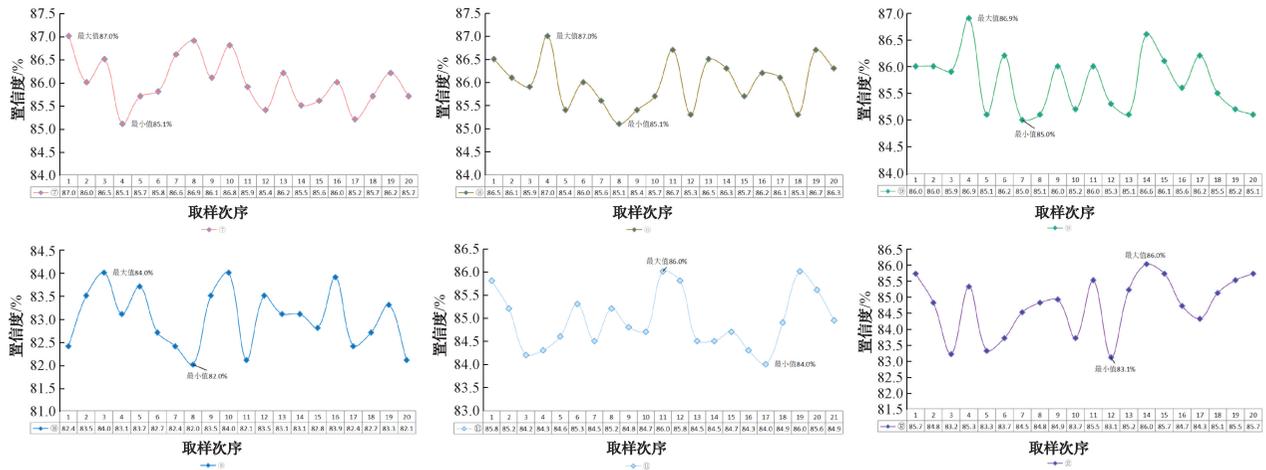


图 12 12类实验目标检测跟踪的置信度变化

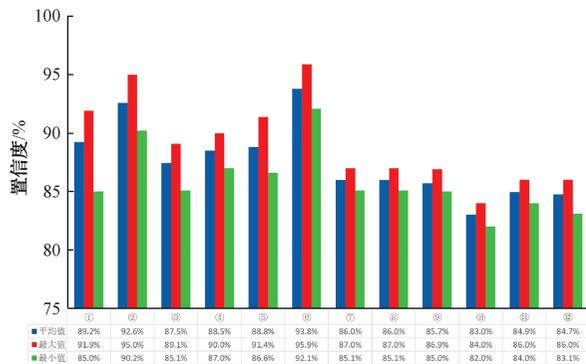


图 13 12类实验置信度最值与平均值

两个数据源进行了插值计算,经过算法处理的无人机视频中提取的速度数据也进行了类似的集成。误差比较如图 14 所示。

根据图 14 各项对比数据,采用相对误差来测量样本数据的准确性,暂不考虑引用曲线回归的拟合优度指标^[26]进行讨论。由上图结果表明,目标车辆的地面样本真值数据和实验样本值最大相对误差出现在实验①,最大相对误差为 12.83%;最小相对误差出现在实验⑥,最小相对误差为 3.86%,而其他组实验的相对误差普遍在 5%~8%。可知本文提出的基于 YOLOX-IM-DeepSort 模型以及视频帧率测速算法的无人机视频中车辆检测和

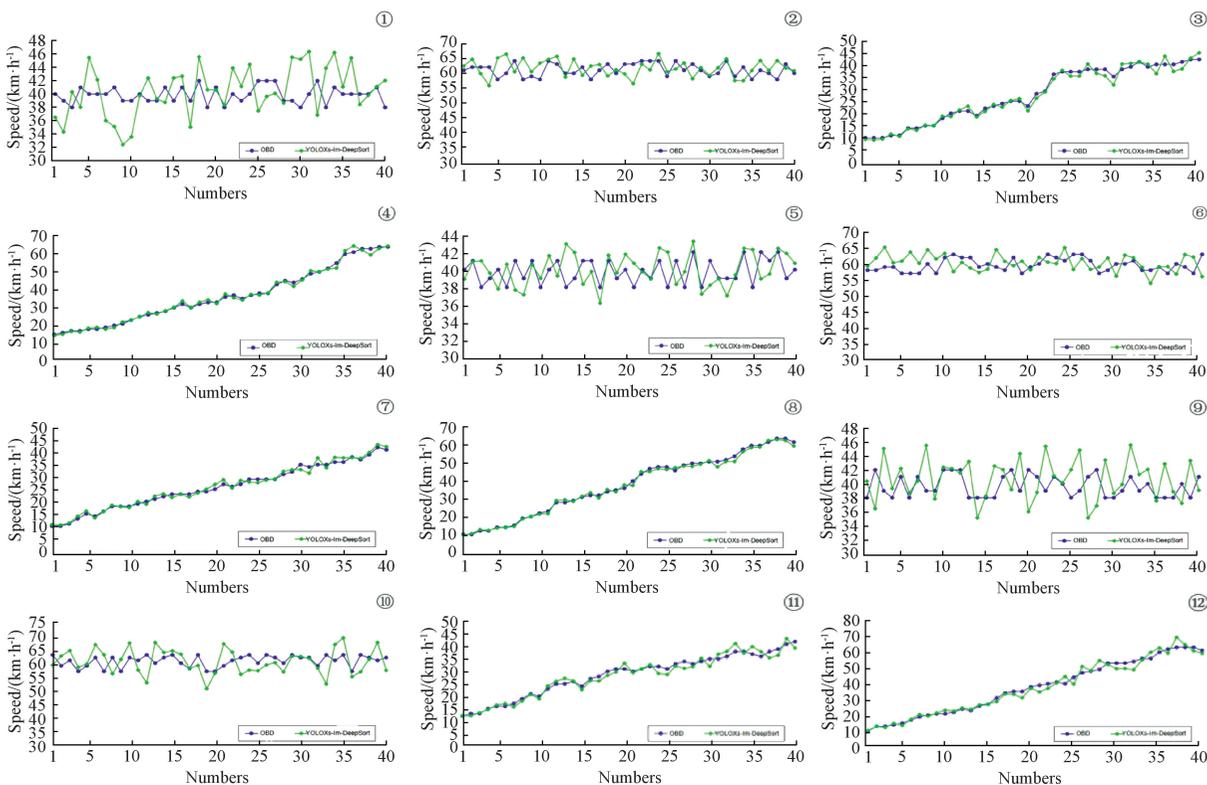


图 14 速度实验样本值与地面样本真值误差

跟踪方法最适合的无人机交通监控平台情况为无人机高度设置为 50 m 时。

5 结论

在交通监测领域,越来越多的研究机构开始引入无人机,可无人机图像具有许多小物体和复杂的背景,将传统的目标检测算法直接应用于无人机图像,会导致大量的误检测和误检测。本文提出了一种改进 YOLOX-s 的无人机航拍图像的目标检测算法 YOLOX-IM。通过对训练集进行矩阵变化预处理和使用 SAHI 进行数据增强,有效地提高了检测小目标的能力。并且,在 PAN 中引入一个包含丰富空间信息的浅层特征图 C_2 ,并与基准模型的 3 个有效特征图融合,并添加一个检测头来检测小目标。

ULSAM 还被添加来获得不同的注意力地图,突出物体信息,削弱背景信息。最后,对边界回归的损失函数进行了优化,使边界回归更快、更准确,提高了模型训练速度和预测精度。在 VisDrone 数据集上的实验结果表明,与基础 YOLOX-s 模型相比,本文改进的 YOLOX-IM 在体积上减少 67.14%,mAP50 增加了 8.13%,AP-small、AP-mid 和 AP-large 分别增加了 3.9%、5.4% 和 8.4%;与 CenterNete、YOLOv3、D-A-FS SSD 和 RetinaNet 以及二阶段算法 Faster R-CNN 和 Cascade R-CNN 算法的检测精度相比都有较大提升。

除此之外,本文基于上述优良的 YOLOX-IM 模型,连接 DeepSort 目标跟踪模型,构成可用于无人机航拍视频的轻量化交通参数提取的 YOLOX-IM-DeepSort 模型。进行实际交通监测情况下的交通参数提取实验,实验结果表明本文改良模型最佳使用交通监测环境为无人机监测平台高度设置为 50 m,此时提取的参数的总体准确率高达 96.14%。

在后续的研究中可使用本文所提出的方法收集实时的交通运行参数,为交通信号控制、交通排放、驾驶行为等提取相关的研究数据。

参考文献

- [1] WANG S J, JIANG F, ZHANG B, et al. Development of UAV-based target tracking and recognition systems [J]. IEEE Transactions on Intelligent Transportation Systems, 2020, 21(8): 3409-3422.
- [2] CHEN X Q, LI Z B, YANG Y S, et al. High-resolution vehicle trajectory extraction and denoising from aerial videos [J]. IEEE Transactions on Intelligent Transportation Systems, 2021, 22(5): 3190-3202.
- [3] HE K, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN [C]. Proceedings of the IEEE International Conference on Computer Vision, 2017: 2961-2969.
- [4] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, 580-587.
- [5] GIRSHICK R. Fast R-CNN [C]. Proceedings of the IEEE International Conference on Computer Vision, 2015: 1440-1448.
- [6] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779-788.
- [7] REDMON J, FARHADI A. YOLO9000: Better, faster, stronger [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 7263-7271.
- [8] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector [C]. European Conference on Computer Vision. Springer, 2016: 21-37.
- [9] REDMON J, FARHADI A. YOLOv3: An incremental improvement [J]. Computer Science, 2018, arXiv: 1804.02767.
- [10] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: Optimal speed and accuracy of object detection [J]. Computer Science 2020, arXiv: 2004.10934.
- [11] LIU Y, DING Z, CAO Y, et al. Multi-scale feature fusion UAV image object detection method based on dilated convolution and attention mechanism [C]. 8th International Conference on Information Technology: IoT and Smart City, 2020: 125-132.
- [12] 彭继慎, 孙礼鑫, 王凯, 等. 基于模型压缩的 ED-YOLO 电力巡检无人机避障目标检测算法 [J]. 仪器仪表学报, 2021, 42(10): 161-170.
- [13] 李仁鹰, 钱慧芳, 郭佳豪, 等. 基于 M-YOLOv4 模型的轻量化目标检测算法 [J]. 国外电子测量技术, 2022, 41(4): 15-21.
- [14] ETTEN A V. You only look twice: Rapid multi-scale object detection in satellite imagery [J]. Computer Science, 2018, arXiv: 1805.09512.
- [15] 化嫣然, 张卓, 龙赛, 等. 基于改进 YOLO 算法的遥感图像目标检测 [J]. 电子测量技术, 2020, 43(24): 87-92.
- [16] TAN L, LV X Y, LIAN X F, et al. YOLOv4_drone: UAV image target detection based on an improved YOLOv4 algorithm [J]. Computers & Electrical Engineering, 2021, 93: 107261.

- [17] GE Z, LIU S T, WANG F, et al. YOLOx: Exceeding yolo series in 2021 [J]. Computer Science, 2021, arXiv: 2107. 08430.
- [18] AKYON F C, ALTINUC S O, TEMIZEL A. Slicing aided hyper inference and fine-tuning for small object detection [J]. Computer Science, 2022, arXiv: 2202.06934.
- [19] ZHANG H, CISSE M, DAUPHIN Y N, et al. Mixup: Beyond empirical risk minimization [J]. Computer Science, 2017, arXiv: 1710.09412.
- [20] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: Optimal speed and accuracy of object detection [J]. Computer Science, 2020, arXiv: 2004.10934.
- [21] DU D, ZHU P, WEN L, et al. VisDrone-DET2019: The vision meets drone object detection in image challenge results[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, 2019.
- [22] SAINI R, JHA N K, DAS B, et al. Ulsam: Ultralightweight subspace attention module for compact convolutional neural networks[C]. Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2020: 1627-1636.
- [23] YU J, JIANG Y, WANG Z, et al. Unitbox: An advanced object detection network[C]. Proceedings of the 24th ACM international conference on Multimedia, 2016: 516-520.
- [24] 董美琳,任安虎. 基于深度学习的高速公路交通事件检测研究[J]. 国外电子测量技术, 2021, 40(10): 108-116.
- [25] 刘建蓓, 单东辉, 郭忠印, 等. 无人机视频的交通参数提取方法及验证[J]. 公路交通科技, 2021, 38(8): 149-158.
- [26] 张晶. 线性回归模型参数估计方法的分辨率[D]. 武汉: 武汉理工大学, 2020.

作者简介

刘军黎, 硕士研究生, 主要研究方向为交通场景中的目标检测、交通图像处理与识别、无人机航拍视频交通信息检测技术等。

E-mail: serafina19971009@163.com

刘晓锋(通信作者), 博士, 副教授, 主要研究方向为交通控制与优化、无人机巡航路径优化与应用、地空/车路协同交通等。

E-mail: microbreeze@126.com