

DOI:10.11918/202205119

多类场景下无人机航拍视频烟雾检测算法

王殿伟¹, 赵文博¹, 房杰¹, 许志杰²

(1. 西安邮电大学 通信与信息工程学院, 西安 710121; 2. 哈德斯菲尔德大学 计算机与工程学院, 哈德斯菲尔德 HD1 3DH, 英国)

摘要: 在无人机航拍视频烟雾检测领域中, 由于不同检测场景差异大, 导致现有烟雾检测算法经常出现检测精度低、速度慢等问题。为了解决以上问题, 建立了一个基于无人机视角的多类场景下的烟雾数据集(UAV smoke dataset, USD), 并提出了一种改进YOLOx的多类场景下无人机视频烟雾检测算法。首先, 在YOLOx网络模型中引入改进的注意力机制, 分别改进通道特征和空间特征的提取过程, 提取更加具有表征能力的烟雾特征; 然后, 提出一种双向特征融合模块, 增强多尺度特征融合模块对小目标烟雾特征的融合能力; 最后, 引入Focal-EIOU损失函数, 解决训练过程中出现正负样本不平衡, 以及预测框和真实框不相交时无法反映两个框的距离远近和重合度大小等问题。实验结果表明, 所提算法在应用于多类场景下无人机视频烟雾检测任务时具有较好的鲁棒性, 对比多个经典烟雾检测算法, 本文算法在不同数据集上的烟雾检测准确率均有不同的提升, 比如对比原有的YOLOx-s模型, 准确率提升2.7%, 召回率提升3%, 速度达到73.6帧/s。

关键词: 烟雾检测; 无人机航拍视频; 多场景; YOLOx; 注意力机制; 改进特征金字塔

中图分类号: X932; TP391.41

文献标志码: A

文章编号: 0367-6234(2023)10-0122-08

Smoke detection algorithm for UAV aerial video in multiple scenarios

WANG Dianwei¹, ZHAO Wenbo¹, FANG Jie¹, XU Zhijie²

(1. School of Communications and Information Engineering, Xi'an University of Posts and Telecommunications, Xi'an 710121, China;

2. School of Computing and Engineering, University of Huddersfield, Huddersfield HD1 3DH, UK)

Abstract: In the field of UAV smoke detection, due to the significant variations in different detection scenes, existing smoke detection algorithms often suffer from issues such as low detection accuracy and slow speed. To address these issues, in this paper we constructed a new UAV smoke dataset (USD) in multiple scenes, and proposed an improved YOLOx UAV smoke detection algorithm in multiple scenes. Firstly, we introduced an improved attention module into the YOLOx network to improve the extraction process of channel features and spatial features respectively, which can extract more representational smoke features. Then, we presented a two-way fusion network to enhance the fusion ability of multi-scale feature fusion module for small smoke target features. Finally, we utilized a Focal-EIOU loss function to address the issues such as the imbalance of positive and negative samples in the training process, and the distance and coincidence degree of two frames cannot be reflected when the prediction frame and real frame do not intersect. Experimental results show that the proposed algorithm has good robustness when applied to UAV smoke detection tasks in multiple scenarios. Compared with several classical smoke detection algorithms, the accuracy of the proposed smoke detection method on different data sets has been improved respectively. For instance, compared with the original YOLOx-s model, the accuracy was improved by 2.7%, the recall rate was improved by 3%, and the speed reached 73.6 frames per second.

Keywords: smoke detection; UAV aerial video; multiple scenarios; YOLOx; attention mechanism; improved feature pyramid

烟雾是火灾的主要伴生现象, 具有更明显的可观测性^[1], 因此开展烟雾智能检测对消防减灾具有重要的意义。无人机具有机动性强、可视范围广和行进速度快等特点^[2-3], 在烟雾检测领域中有明显的优势。因此, 开展无人机航拍视频烟雾检测技术

研究对于消防减灾具有重要的意义, 已经成为计算机视觉领域的一个新的研究热点^[4]。

传统烟雾检测算法大多是围绕着手工提取各种烟雾特征进行研究, 检测精度不高。而卷积神经网络因具有强大的特征表征能力, 已被广泛应用于烟

收稿日期: 2022-05-30; 录用日期: 2022-08-08; 网络首发日期: 2023-04-06

网络首发地址: <http://kns.cnki.net/kcms/detail/23.1235.T.20230404.1644.004.html>

基金项目: 公安部科技强警基础工作专项(2019GABJC42); 西安邮电大学研究生创新基金(CXJJZL2021022)

作者简介: 王殿伟(1978—), 男, 博士, 副教授

通信作者: 王殿伟, wangdianwei@xupt.edu.cn

雾检测领域。文献[5-6]使用 YOLO 系列算法进行烟雾检测,但在多类场景下准确率不高。文献[7]提出了烟雾检测和分割框架;文献[8]提出了两级烟雾检测算法,都提升了算法精度,但增加了计算量,导致算法检测速度慢。文献[9]提出一种新的注意机制模块,但训练数据过少导致算法性能不佳。为此,文献[10-11]利用人工合成烟雾图像解决缺乏训练数据的问题,但人工合成的烟雾图像真实性差导致算法鲁棒性低。文献[12]提出火焰和烟雾统一检测的方法;文献[13]提出基于深度可分离卷积神经网络的烟雾检测算法;文献[14]提出 $Y^dU^aV^a$ 颜色模型和改进的 MobileNetV3 网络结构,但在多类场景下烟雾检测效果都不好。文献[15]提出的 YOLOv4 算法和文献[16]提出的改进的 YOLOv5 算法都对烟雾检测算法轻量化做出了贡献;文献[17]提出的 SparseR-CNN 算法在应用于烟雾检测领域时明显提升了准确率,但算法实时性不高。文献[18]提出 YOLOx 目标检测算法,在 YOLOv3^[19] 的基础上新增了 Decoupled Head 和 Anchor-free 等结构,算法精确度显著提升,但在应用于无人机视角下的多类场景时,检测准确率降低。

针对以上算法出现的训练数据少、场景种类单一和检测精度低等问题,本文建立了一个多类场景下的无人机航拍烟雾数据集(UAV smoke dataset, USD),从注意力机制、双向特征融合模块和损失函数 3 个角度出发改进 YOLOx 算法,提升了算法在面对多类场景下无人机航拍视频的烟雾检测精度,并具有实时检测的能力。

1 无人机航拍烟雾数据集 USD 的建立

现有的无人机视频烟雾检测算法使用的训练数据集多为普通视角下的单一场景烟雾数据集,这导致训练出的网络模型不能适用于多类场景下的无人

机烟雾检测任务。目前最常使用的公共数据集有两个:1)上海师范大学袁非牛团队公布的烟雾图像和非烟雾图像数据集 VSD^[20],该数据集包含了 6 323 张各种颜色的烟雾图像,但该数据集包含的烟雾图像质量较差,分辨率低,且烟雾目标在整幅图像中占据了较大的面积,对网络模型的检测性能有很大影响;2)韩国启明大学计算机视觉与模式识别实验室(CVPR)联合美国 NIST 火灾研究实验室以及 Bilkent 信号处理小组公开的烟雾数据集 KMU^[21],其中共包含了 11 段有烟视频和 9 段非烟雾视频,加入了大量白色灯光和自然天气下的云等负样本信息,但该数据集的视频分辨率较低,且只包含森林和室内两种场景,很难训练出具有较强泛化能力的网络。

数据集的建立应包含图像信息多和场景覆盖性广两个方面,因此采用自己拍摄和网络收集等方法,创建了一个新的多类场景下的无人机航拍烟雾数据集,该数据集包含了 81 个无人机烟雾视频,对这 81 个视频进行每隔 100 帧提取一次图像,共提取 30 000 张无人机烟雾图像,并使用 LabelImg 软件对 30 000 张图像进行标注,标注方法为最大外接矩形框标注。数据集中所有视频都是由无人机在高空 50~500 m 之间拍摄完成,包含了森林、农场、居民楼、学校、工厂和海岸等 19 种场景,且视频的分辨率大小都为 1 280×720,烟雾区域更加清晰,见表 1。部分场景的烟雾图像见图 1。

表 1 USD 数据集中的场景类别
Tab. 1 Categories of scenarios in USD

场景	数量/张	场景	数量/张
森林	3 428	学校	1 235
居民楼	2 967	停车场	1 029
工厂	2 515	船舶	968
农场	1 426	海岸	876
公园	1 312	其他	14 244



图 1 USD 数据集的部分场景的烟雾图像
Fig. 1 Smoke images of some scenes from USD

由表 1 可以看出,在 USD 烟雾数据集中,森林、居民楼和工厂 3 个场景的图像占比最多,十分符合现实情况。相比于普通视角下的烟雾数据集,本文自建的 USD 烟雾数据集,包含场景多、数量多、分辨率高,图像中的烟雾形状更加多变,烟雾特征也更加丰富,与算法的应用场景非常贴切。

2 改进 YOLOx 的无人机视频烟雾检测算法

2.1 网络结构

为了更好地实现在多类场景下对无人机航拍视频的烟雾检测,设计了一个基于 YOLOx-s 的轻量级的端到端烟雾检测网络,采用改进的注意力机制模块和多尺度特征融合模块降低场景信息对烟雾检测准确率的影响,并提升网络对小目标烟雾的检测能力。网络结构见图 2。

由图 2 可知,本文的烟雾检测网络主要由 4 部分组成:1) 特征提取网络;2) 注意力机制模块;3) 双

向特征融合模块(bidirectional feature fusion module);4) 预测网络。特征提取网络由 4 个 CBS 单元(卷积层+BN 层+Silu 激活函数)、1 个 SPP 单元、1 个 Focus 单元和 3 个 CSP 结构组成。预测网络由 Decoupled Head 模块以及 Transpose 结构组成。为了提高本算法对小目标烟雾的检测能力,本文提出的双向特征融合模块在多尺度特征融合的基础上加入了一条额外支路,该支路能使 FPN 结构学习到更多的小目标烟雾特征,从而增强网络对小目标烟雾的学习能力。为了解决多类场景下算法检测准确率低的问题,引入改进的烟雾协调注意力(smoke coordinate attention, SCA)^[22],该模块是一种适用于轻量级网络的烟雾注意力机制,能使网络更加关注特征图中的烟雾区域,并对不同通道的烟雾特征以及不同大小的烟雾特征进行重新加权计算。最后本文引入了 Focal-EIOU^[23] 损失函数,解决了正负样本不平衡的问题以及预测框和真实框不相交时无法反映两个框的距离远近以及重合度大小的问题。

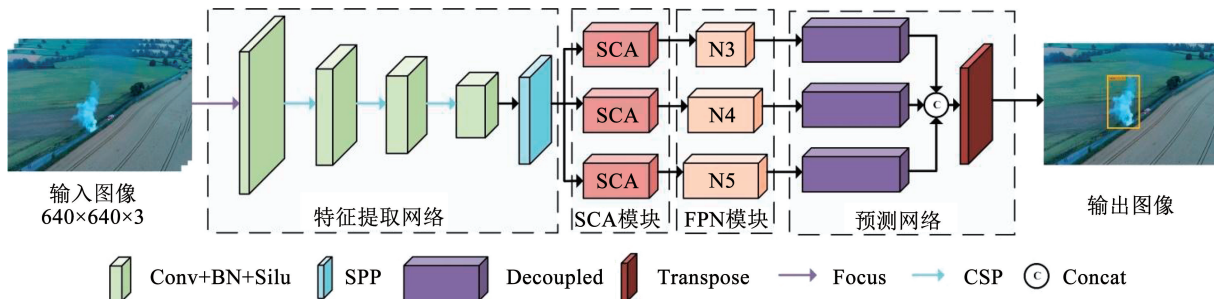


图 2 本文算法网络模型

Fig. 2 Proposed algorithm network model

2.1.1 注意力机制模块

鉴于多类场景下无人机烟雾视频图像的复杂性,本文增加改进的注意力机制模块来提升算法对烟雾特征的提取能力,其结构见图 3(a)。首先,利用 2 个不同方向的全局平均池化生成包含特征通道信息的描述符,经过 1 个拼接操作和 1 个共享的卷积层,再经过 1 个批量归一化处理,得到的特征图最终经过 1 个全连接层和 Sigmoid 激活函数。在该注意力机制模块中,本文将烟雾的位置信息嵌入到通道注意力中,并且沿着高度和宽度 2 个方向生成一对特征图,分别为方向感知图和位置感知图,最终将这一对特征图加权融合到原始输入特征图中,其加权公式为

$$y_c(i, j) = x_c(i, j) \times g_c^h(i) \times g_c^w(j) \quad (1)$$

式中: y_c 为加权融合后的特征图; x_c 为第 c 通道的原始特征图; (i, j) 为该像素点在特征图映射的坐标; h 和 w 分别为该像素点在烟雾特征图中的高度和宽度; g_c^h 和 g_c^w 分别为第 c 通道中烟雾特征图在高度方

向上和宽度方向上的权重, g_c^w 和 g_c^h 的计算公式为

$$g_c^w = \sigma(F_w(f^w)) \quad (2)$$

$$g_c^h = \sigma(F_h(f^h)) \quad (3)$$

式中: σ 为 Sigmoid 函数; F_h 和 F_w 分别为 1×1 的卷积变换; f^h 和 f^w 为输入图像在高度和宽度上生成的注意力特征图,注意力生成公式为

$$f = \delta\left(F_1\left(\left[\frac{1}{W} \sum_{0 \leq i < W} x_c(h, i), \frac{1}{H} \sum_{0 \leq j < H} x_c(j, w)\right]\right)\right) \quad (4)$$

式中: f 为生成的注意力特征图; δ 为非线性激活函数; x_c 为大小为 $H \times W \times C$ 的输入特征图; F_1 为批量归一化操作。

2.1.2 双向特征融合模块

在 YOLOx-s 的网络中,特征融合模块采用的是 FPN + PAN^[24] 结构,该结构融合了 80×80 、 40×40 和 20×20 三种不同大小的特征图。为了增强网络对烟雾小目标的检测能力,适应无人机视角下火灾早期烟雾主要为小目标的实际情况,受 BiFPN^[25] 结构的启发,本文设计了一个额外的融合支路加入到

多尺度特征融合模块,其结构见图 3(b)。双向特征融合加权计算为

$$P_4^{\text{out}} = \text{Conv} \left[\frac{w_1 \times P_4^{\text{in}} + w_2 \times \text{resize}(P_5^{\text{in}})}{w_1 + w_2 + \theta} \right] \quad (5)$$

$$N_4^{\text{out}} = \text{Conv} \left[\frac{w'_1 \times P_4^{\text{in}} + w'_2 \times P_4^{\text{out}} + w'_3 \times \text{resize}(P_5^{\text{out}})}{w'_1 + w'_2 + w'_3 + \theta} \right] \quad (6)$$

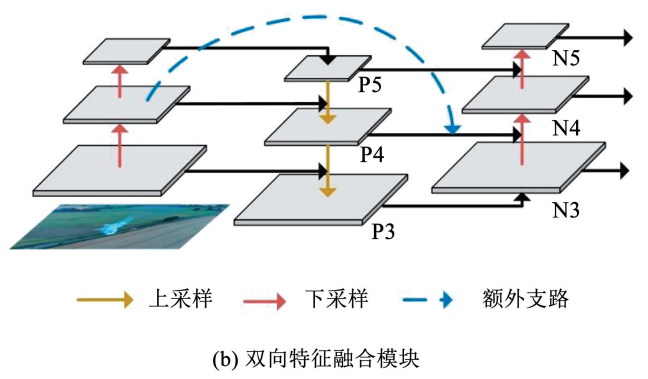
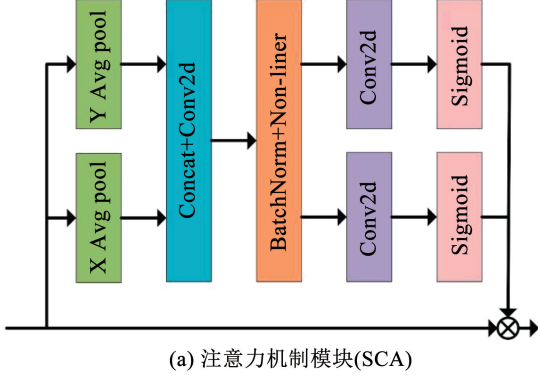


图 3 SCA 模块和双向特征融合模块

Fig. 3 SCA module and bidirectional feature fusion module

在双向特征融合模块中,本文在 PAN 结构的 N4 层额外加入了原始的 40×40 大小的特征图,新加入的特征与 P4 层的输出和 N3 层经过下采样得到的特征共同输入到 N4 层,随后再经过下采样输入到 N5 层。可以看出,这使得 N5 层相比原来获得了更多的小目标特征,并且新加入的注意力机制模块将不同层和不同大小的烟雾特征进行重新分配权重,再将得到的烟雾特征连接到一起,最终输入到 Decoupled Head 模块进行分类和检测。

2.1.3 损失函数

在无人机拍摄的烟雾视频图像中,烟雾所占图像比例远小于其他负样本信息所占的比例,这就造成了算法训练过程中存在正负样本不平衡的问题。为此,本文引入 $L_{\text{Focal-EIOU}}^{[23]}$ 损失函数,其中 L_{Focal} 损失函数可以解决训练过程中正负样本不平衡的问题。同时,在烟雾图像中,烟雾的形状多变且不规则,可能导致模型的真实框和预测框不相交的问题,为了解决这一问题,本文在计算 IOU^[26] 损失时引入 L_{EIOU} 损失函数。该损失函数考虑了预测框和真实框的重叠面积、中心点距离、长宽边长真实差,可以解决预测框和真实框不相交时无法反映两个框的距离远近以及重合度大小的问题。 $L_{\text{Focal-EIOU}}$ 损失函数计算步骤为

$$L_{\text{Focal-EIOU}} = \gamma L_{\text{IOU}} L_{\text{EIOU}} \quad (7)$$

式中: γ 为控制异常值抑制程度的参数; I_{IOU} 为真实框和预测框的交并比; L_{IOU} 为 IOU 损失; L_{EIOU} 为 EIOU 损失,其中 I_{IOU} 和 L_{EIOU} 损失计算为

$$I_{\text{IOU}} = \frac{(A \cap B)}{(A \cup B)} \quad (8)$$

式中: P_4^{out} 为 P4 层的输出特征图; P_4^{in} 和 P_5^{in} 为 P4 层和 P5 层的输入特征图; P_5^{out} 为 P5 层的输出特征图; N_4^{out} 为 N4 层的输出特征图;resize 为调整特征图大小的函数; w_1, w_2, w'_1, w'_2 和 w'_3 为训练过程中的归一化权重;学习率 θ 为 0.000 1。

$$L_{\text{EIOU}} = L_{\text{IOU}} + L_{\text{dis}} + L_{\text{asp}} \quad (9)$$

$$L_{\text{EIOU}} = 1 - I_{\text{IOU}} + \frac{\rho^2(b, b^{\text{gt}})}{c^2} + \frac{\rho^2(w, w^{\text{gt}})}{c_w^2} + \frac{\rho^2(h, h^{\text{gt}})}{c_h^2} \quad (10)$$

式中: A 和 B 分别为预测框和真实框; $L_{\text{IOU}}, L_{\text{dis}}$ 和 L_{asp} 分别为 I_{IOU} 损失、中心距离损失和宽高损失; ρ 是预测框和真实框中心点间的欧氏距离; b 和 b^{gt} 分别为预测框和真实框的中心点; w 和 w^{gt} 分别为预测框和真实框的宽度; h 和 h^{gt} 分别为预测框和真实框的高度; c_w 和 c_h 分别为两个框的最小外接框的宽度和高度。

3 实验与结果分析

3.1 实验平台、参数设置与评价指标

本文算法研究的硬件平台为 CPU (Intel (R) Core(TM) i7-12700K 3.61 GHz), GPU (Nvidia Titan XP), CUDA 版本为 11.1, 实验环境为 Python3.7 和 PyTorch 框架 1.8.1 版本。无人机型号为大疆精灵 4 Pro, 最大飞行高度 6 000 m, 拍摄视频和照片的最大分辨率为 $4\,000 \times 3\,000$ 。

本文使用自建的多类场景无人机航拍烟雾视频图像数据集 USD 作为训练数据集,按照 8:1:1 的比例随机选取了其中 24 000 张图片作为训练集,3 000 张图片作为验证集,3 000 张图片作为测试集。训练开始时,使用多尺度缩放将原始 $1\,280 \times 720$ 大小的图像缩小到 640×640 大小的图像并作为输入,训练过程中采用了随机梯度下降法 (stochastic gradient descent, SGD)。经过实验,当 Batch size 设置为 8,初

始学习率设置为 0.01,余弦衰减权重系数设置为 0.000 5时,得到的模型为最优。本文共训练了 300 个 epoch,其中每 10 个 epoch 进行一次验证,训练总耗时 52 h 34 min,最终模型大小为 70.24 MB。

本文使用准确率 P (precision)、召回率 R (recall)、类别精度 A (average precision)作为检测算法有效性的评价指标,为了更进一步的评价本文算法的适用性,另加入每秒检测帧数 F_{FPS} (frames per second)作为参考对比。

3.2 对比实验及分析

本文选用 USD 烟雾数据集中的 24 000 张和 3 000 张分别用作本文算法的训练集和测试集 A。为了验证本文算法的泛化性,利用网络收集的方法额外建立了 1 个普通视角下的测试集 B,并用现有的 2 个公开数据集制作了测试集 C 和测试集 D。测

试集 B 包含 3 864 张普通视角下的烟雾图像,测试集 C 包含袁非牛团队公开的 VSD^[20]烟雾数据集中的 2 000 张普通烟雾图像,测试集 D 包含从 CVPR 实验室联合 NIST 火灾研究实验室火灾研究部以及 Bilkent 信号处理小组公开的 KMU^[21]烟雾数据集中提取的 1 000 张烟雾图像。

本文将所提算法在自建烟雾数据集以及 2 个经典烟雾数据集上进行测试,为了验证本文自建烟雾数据集的有效性和本文所提算法在多类场景下无人机视频烟雾检测的优势,本文选用 YOLOv4^[15]、YOLOv5-s^[16]、SparseR-CNN^[17]3 种经典的目标检测算法作对照实验,另选用最新的烟雾检测模型 DeepSmoke^[7]进行对比实验。为了保证实验环境的一致性,所有算法均在相同的实验环境下运行实现,实验结果见表 2。

表 2 不同算法在 3 个数据集上的表现
Tab. 2 Performance of different algorithms on three datasets

模型	USD				VSD ^[20] /%			KMU ^[21] /%		
	P /%	R /%	A /%	$F_{FPS}/(\text{帧}\cdot\text{s}^{-1})$	P	R	A	P	R	A
YOLOv4	83.9	83.7	81.6	40.3	88.4	87.4	86.9	86.4	85.9	85.4
YOLOv5-s	85.2	84.8	83.6	103.1	89.3	88.6	87.5	89.1	88.5	88.1
SparseR-CNN	92.4	91.1	89.6	6.2	96.5	95.4	92.8	96.8	95.6	95.1
DeepSmoke	91.3	90.6	89.1	32.4	93.5	93.2	90.9	92.4	91.9	91.5
YOLOx-s	89.4	88.9	87.3	74.2	94.6	93.8	89.1	93.9	92.6	92.2
Ours	92.1	91.7	89.8	73.6	96.4	95.6	92.9	96.1	95.7	94.7

注:加粗表示最优值。

由表 2 可以看出,在 VSD^[20]和 KMU^[21]等公开数据集上,场景种类少,图像复杂度不高,本文算法取得了 96% 以上的准确率,相比于 YOLOv4、YOLOV5-s 和 DeepSmoke 算法,本文算法在准确率、召回率和类别精度都有很大的领先。在自建的 USD 烟雾数据集中,由于该数据集包含的无人机视角烟雾场景多,烟雾特征更加复杂,导致所有算法均出现了准确率降低的现象,但本文所提算法对比于原算法 YOLOx-s 仍然提升了 2.7% 的准确率和 2.5% 的精度,并且 F_{FPS} 达到了 73.6,表明本文算法在面对多类场景下的无人机航拍烟雾检测时,比原算法 YOLOx-s 和其他现有算法更有优势。本文算法的检测准确率略低于 SparseR-CNN 算法,但 SparseR-CNN 算法的 F_{FPS} 只有 6.2 帧/s,明显达不到实时检测的要求。另外,本文展示了本文算法在

USD 烟雾数据集、测试集 B、测试集 C 和测试集 D 上的部分可视化结果,见图 4 和图 5。

由图 4 的可视化检测结果可以看出,本文算法在面对不同场景下的无人机烟雾检测任务时,都能以最大框准确地标示出图像中烟雾区域外沿,证明了本文算法的有效性。从图 5 的可视化检测结果可以看出,在测试集 B 中,烟雾像素占据整幅图像近一半的像素,导致烟雾特征过于明显;在测试集 C 中,图像主要由像素大小为 100 × 100 的烟雾块组成,没有复杂的场景信息干扰;在测试集 D 中,该测试集中出现了雾天、淡烟和大量的天空区域,烟雾特征不明显,导致算法出现了漏检和误检现象,本文列出了不同算法在测试集 D 中(a)、(b)两组场景中漏检和误检的可视化结果,见图 6。



图 4 本文算法在 USD 数据集上的部分检测结果

Fig. 4 Partial detection results of proposed algorithm on USD



图 5 本文算法在不同测试集上的部分检测结果

Fig. 5 Partial detection results of proposed algorithm on different test sets

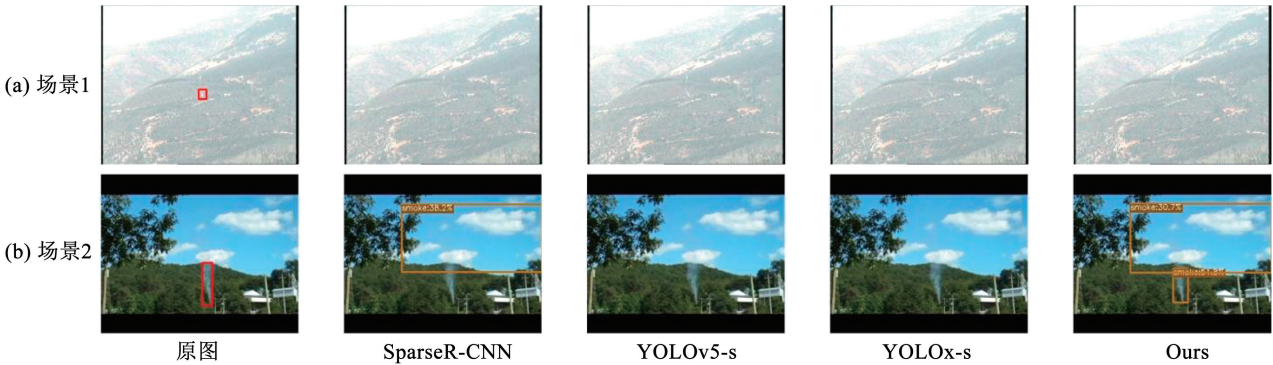


图 6 不同算法在测试集 D 上的漏检和误检结果

Fig. 6 Missed and false detection results of different algorithms on test set D

由图 6 可以看出,在图 6(a)和图 6(b)两组场景的可视化结果中,不同算法都出现了漏检和误检现象(图中烟雾区域由红框标出)。在图 6(a)场景中,出现了有雾天气,烟雾所占像素小且出现了颜色失真,导致所有算法都出现了漏检现象;在图 6(b)

场景中,出现了大量含云的天空区域,导致 SparseR-CNN 算法出现了误检现象, YOLOv5-s 算法和 YOLOx-s 算法都出现了漏检现象,本文算法虽然准确地检测出了烟雾区域,但同样也误检了天空中含云的区域,这是因为本文所用训练集 USD 中的图像

主要为无人机对地视角,包含多云的场景较少,导致算法模型对假烟物体敏感,出现误检现象。

为了评价本算法和其他算法对不同场景无人机烟雾视频的检测效果,选取 USD 数据集中 5 个不同场景的无人机烟雾视频,每个场景包含 2 000 帧的连续视频图像,对共计 10 000 帧的烟雾和非烟雾视频进行测试,测试结果见表 3。可以看出,本文算法检测准确率对比 YOLOx-s 算法在不同场景下都取得了提升。对比 DeepSmoke 算法、YOLOv5-s 算法和 YOLOv4s 算法,本文算法在 5 个不同场景下都取得了最高的检测准确率,且平均准确率达到了 92.3% 以上。

表 3 不同视频下的检测准确率对比

Tab.3 Comparison of detection accuracy for different videos %

场景类别	Ours	YOLOx-s	DeepSmoke	YOLOv5-s	YOLOv4
森林	92.7	89.7	91.6	83.6	83.1
工厂	92.5	89.5	90.5	87.8	82.7
居民楼	91.7	87.1	84.8	82.7	79.8
农场	92.6	89.6	88.9	86.9	86.1
海岸	92.0	90.1	86.7	85.8	84.3

注:加粗表示最优值。

3.3 消融实验

消融实验的对象主要包括 SCA 注意力模块、双向特征融合模块和 Focal-EIOU 损失函数,实验结果见表 4。本文算法所提出的改进对检测准确率都有所提升,其中对算法检测准确度提升最大的是 SCA 模块,检测精度提升了 1.9%,证明改进的注意力机制有效的增强了算法对烟雾特征的提取能力。单一的双向特征融合模块同样使算法检测准确度提升 1.6%,证明改进的双向特征融合模块能有效地增强 FPN 结构对小目标烟雾特征的融合能力。引进的 Focal-EIOU 损失函数也对算法检测准确度有所提升,证明该损失函数解决了网络训练过程中正负样本不平衡的问题和预测框与真实框不相交时无法反映两个框的距离远近以及重合度大小的问题。

表 4 消融实验

Tab.4 Ablation experiments

基本算法	SCA	双向特征融合模块	Focal-EIOU	A/%
YOLOx-s				86.3
YOLOx-s	✓			88.2
YOLOx-s		✓		87.9
YOLOx-s			✓	87.1
YOLOx-s	✓	✓		89.1
YOLOx-s	✓	✓	✓	89.8

4 结 论

针对现有烟雾检测算法在 multi 场景下无人机航拍视角中表现不佳的问题,本文自建了一个包含 multi 场景的无人机航拍视频烟雾数据集 USD,并提出了一种改进 YOLOx 的无人机视频烟雾检测算法,得出结论如下:

1)对现有的烟雾数据集进行分析,建立一个 multi 场景下的无人机航拍视频烟雾数据集 USD,该数据集包含了从 81 个无人机烟雾视频中提取出的 30 000 张无人机烟雾图像,总共包含了 19 类场景,在场景类别、图像质量和图像数量上远优于其他现有的烟雾数据集。

2)对现有烟雾检测算法在应用到无人机视角的 multi 场景时出现的检测率低、速度慢等问题进行了分析,针对火灾初期烟雾多为小目标的问题,提出改进的双向特征融合模块和注意力机制模块。对网络训练过程中出现的正负样本不平衡问题进行分析,并对预测框和真实框不相交时无法反映两个框的距离远近以及重合度大小的问题进行了研究,引入了 Focal-EIOU 损失函数。

3)在自建数据集 USD 以及国内外常用烟雾数据集上进行实验分析,并与现有算法对比,证明了本文各个模块能有效解决 multi 场景下无人机航拍烟雾检测领域中的常见问题,有效提升了本文算法的检测性能。

4)本文算法在面对多雾、多云和光照不良等极端场景时会出现检测准确率下降的情况,这是由于训练数据集中缺少极端场景下的烟雾图像,导致算法出现漏检和误检,下一步我们将从两个方面继续深化完善本文算法:1)扩充数据集,加入更多极端场景下的烟雾图像,提升算法在更复杂的场景中的准确率;2)研究算法的时间复杂度和空间复杂度,在不降低准确率的前提下,利用更加轻量化的网络构架,提升算法的实用性和适用性。

参考文献

[1]FRIZZIS,KAABIR,BOUCHOUICHA M, et al. Convolutional neural network for video fire and smoke detection[C]//IECON 2016-42nd Annual Conference of the IEEE Industrial Electronics Society. Florence: IEEE, 2016: 877. DOI: 10.1109/IECON.2016.7793196

[2]黄鹤,张科,陈永安,等.一种无人机航拍目标的长期跟踪算法[J].哈尔滨工业大学学报,2022,54(5):104

HUANG He,ZHANG Ke,CHEN Yongan, et al. A long-term tracking algorithm for UAV aerial photography[J]. Journal of Harbin Institute of Technology, 2022, 54(5): 104. DOI: 10.11918/202105112

[3]向锦武,阚梓,邵浩原,等.长航时无人机关键技术研究进展

[J]. 哈尔滨工业大学学报, 2020, 52(6): 57

XIANG Jinwu, KAN Zi, SHAO Haoyuan, et al. A review of key technologies for long-endurance unmanned aerial vehicle[J]. Journal of Harbin Institute of Technology, 2020, 52(6): 57. DOI: 10.11918/202004009

[4] BOUGUETTAYA A, ZARZOUR H, TABERKIT A M, et al. A review on early wildfire detection from unmanned aerial vehicles using deep learning-based computer vision algorithms[J]. Signal Processing, 2022, 190: 108309. DOI: 10.1016/2021108309

[5] SAPONARA S, ELHANASHI A, GAGLIARDI A. Real-time video fire/smoke detection based on CNN in antifire surveillance systems[J]. Journal of Real-Time Image Processing, 2021, 18(3): 889. DOI: 10.1007/s11554-020-01044-0

[6] 谢书翰, 张文柱, 程鹏, 等. 嵌入通道注意力的 YOLOv4 火灾烟雾检测模型[J]. 液晶与显示, 2021, 36(10): 1445

XIE Shuhan, ZHANG Wenzhu, CHENG Peng, et al. Firesmoke detection model based on YOLOv4 with channel attention[J]. Chinese Journal of Liquid Crystals and Displays, 2021, 36(10): 1445. DOI: 10.37188/CJLCD.2020-0312

[7] KHAN S, MUHAMMAD K, HUSSAIN T, et al. DeepSmoke: Deep learning model for smoke detection and segmentation in outdoor environments[J]. Expert Systems with Applications, 2021, 182: 115. DOI: 10.1016/J.ESWA.2021.115125

[8] WANG Zhenyu, YIN Duokun, JI Senrong. Real-time factory smoke detection based on two-stage relation-guided algorithm[J]. Scientific Reports, 2022, 12(1): 1. DOI: 10.1038/s41598-022-05523-1

[9] HE Lijun, GONG Xiaoli, ZHANG Sirou, et al. Efficient attention based deep fusion CNN for smoke detection in fog environment[J]. Neurocomputing, 2021, 434: 224. DOI: 10.1016/j.neucom.2021.01.024

[10] LI Tingting, ZHANG Changchun, ZHU Haowei, et al. Adversarial fusion network for forest fire smoke detection[J]. Forests, 2022, 13(3): 366. DOI: 10.3390/f13030366

[11] ZHANG Qixing, LIN Gaohua, ZHANG Yongming, et al. Wildland forest fire smoke detection based on Faster R-CNN using synthetic smoke images[J]. Procedia Engineering, 2018, 211: 441. DOI: 10.1016/J.PROENG.2017.12.034

[12] HOSSEINI A, HASHEMZADEH M, FARAJZADEH N. UFS-Net: A unified flame and smoke detection method for early detection of fire in video surveillance applications using CNNs[J]. Journal of Computational Science, 2022, 12(1): 101. DOI: 10.1016/j.jocs.2022.101638

[13] HUO Yinuo, ZHANG Qixing, JIA Yang, et al. A deep separable convolutional neural network for multiscale image-based smoke detection[J]. Fire Technology, 2022, 58(3): 1445. DOI: 10.1007/s10694-021-01199-7

[14] 刘通, 程江华, 华宏虎, 等. 结合 $Y^dU^aV^a$ 颜色模型和改进 MobileNetV3 的视频烟雾检测方法[J]. 国防科技大学学报, 2021, 43(5): 80

LIU Tong, CHENG Jianghua, HUA Honghu, et al. Video smoke detection method combined $Y^dU^aV^a$ color model and improved mobilenetv3[J]. Journal of National University of Defense Technology, 2021, 43(5): 80. DOI: 10.11887/j.cn.202105009

[15] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: Optimal speed and accuracy of object detection[Z]. arXiv:2004.10934, 2020. DOI: 10.48550/arXiv.2004.10934.

[16] WANG Zhong, WU Lei, LI Ting, et al. A smoke detection model based on improved YOLOv5[J]. Mathematics, 2022, 10(7): 1190. DOI: 10.3390/math10071190

[17] SUN Peizi, ZHANG Rufeng, JIANG Yi, et al. Sparse R-CNN: end-to-end object detection with learnable proposals[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S. l.]: IEEE, 2021: 14454. DOI: 10.1109/CVPR46437.2021.01422

[18] GE Zheng, LIU Songtao, WANG Feng, et al. YOLOx: Exceeding Yolo series in 2021[Z]. arXiv:2107.08430, 2021. DOI: 10.48550/arXiv.2017.08430

[19] REDMON J, FARHADI A. YOLOv3: An incremental improvement[Z]. arXiv:1804.02767, 2018. DOI: 10.48550/arXiv.1084.02767

[20] YUAN Feiniu, ZHANG Lin, XIA Xue, et al. A gated recurrent network with dual classification assistance for smoke semantic segmentation[J]. IEEE Transactions on Image Processing, 2021, 30: 4409. DOI: 10.1109/TIP.2021.3069318

[21] ZHAN Jialei, HU Yaowen, ZHOU Guoxiong, et al. A high-precision forest fire smoke detection approach based on ARGNet[J]. Computers and Electronics in Agriculture, 2022, 196: 106874. DOI: 10.1016/j.compag.106874

[22] HOU Qibin, ZHOU Daquan, FENG Jiashi. Coordinate attention for efficient mobile network design[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021: 13713. DOI: 10.1109/CVPR46437.2021.01350

[23] ZHANG Yifan, REN Weiqiang, ZHANG Zhang, et al. Focal and efficient IOU loss for accurate bounding box regression[Z]. arXiv: 2101.08158, 2021. DOI: 10.1016/j.nuecom.2022.07.042

[24] LIN Tunyi, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE/CVE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 2117. DOI: 10.1109/CVPR.2017.106

[25] TAN Mingxing, PANG Ruoming, LE Q V. EfficientDet: Scalable and efficient object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020: 10781. DOI: 10.1109/CVPR42600.2020.01079

[26] YU J, JIANG Y, WANG Z, et al. Unitbox: An advanced object detection network[C]//Proceedings of the 24th ACM international conference on Multimedia. [S. l.]: ACM, 2016: 516. DOI: 10.1145/2964284.2967274

(编辑 苗秀芝)