

DOI: 10.19650/j.cnki.cjsi.J2514623

基于无人机视觉信息的消防炮预瞄准系统*

张晓光¹, 陈润泽¹, 赵桔贤², 李 伟²

(1. 中国矿业大学信息与控制工程学院 徐州 221116; 2. 中国矿业大学机电工程学院 徐州 221116)

摘要: 无人机可以高效感知火灾环境, 获得火场信息。为提高消防工作的智能化水平, 提出了一种基于无人机视觉信息的消防炮预瞄准系统, 预瞄准是由火场感知、位姿解算、角度调整这 3 个阶段构成的控制过程。在火场感知阶段, 针对消防场景中无人机图像存在的目标尺寸小和烟雾干扰问题, 考虑消防救援工作的实时性要求, 提出了一种联合轻量化去雾处理与目标检测的感知模型。在图像去雾方面, 考虑消防场景中烟雾具有非均匀分布和灰度值多样的特点, 改进大气散射模型, 设计一个编码器-解码器结构的神经网络求解透射图和烟雾灰度值, 实现图像质量增强。在目标检测方面, 以 YOLOv8s 为基线, 在骨干网络中, 将浅层的卷积运算替换为具有集中感受野的 PSConv 模块, 保留更多小目标信息; 将深层的卷积模块替换为 GhostConv, 采用 SimA-former 模块替代最深层 C2f 结构, 进行模型轻量化。在颈部网络特征融合阶段, 结合坐标注意力机制(CA)与小目标检测头构建高分辨率多尺度特征融合模块。基于感知的火场信息, 采用相机成像模型解算消防炮与火源的相对位姿, 据此计算消防炮应调整的水平与俯仰角度。在工业厂房外部自制消防场景进行实验, 感知模型的 mAP50 指标达到了 92.3%, 相较不使用去雾处理的 YOLOv8s 算法提升了 6.2%, 测得预瞄准的水平角度误差在 $\pm 4^\circ$, 火源距离误差在 6% 以内, 验证了所提方法的有效性和实用性。

关键词: 目标检测; 图像去雾; 无人机; 消防场景

中图分类号: TH721 **文献标识码:** A **国家标准学科分类代码:** 510.40

A fire cannon pre-aiming system based on UAV visual information

Zhang Xiaoguang¹, Chen Runze¹, Zhao Juxian², Li Wei²

(1. School of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221116, China;

2. School of Mechatronic Engineering, China University of Mining and Technology, Xuzhou 221116, China)

Abstract: UAV can efficiently perceive the fire environment and obtain the fire scene information. To improve the intelligence level of firefighting work, a fire cannon pre-aiming system based on UAV visual information is proposed. Pre-aiming is a control process consisting of three stages: Fire scene perception, pose calculation, and angle adjustment. In the stage of fire scene perception, considering the real-time requirements of firefighting work, a perception model combining lightweight object detection and dehazing processing is proposed to address the problems of small target size and smoke interference in UAV images of fire scenes. Regarding image dehazing, considering the characteristics of non-uniform distribution and diverse gray levels of haze in fire scenes, the atmospheric scattering model is improved. A neural network with an encoder-decoder structure is designed to solve transmission map and haze gray value, which significantly enhances the image quality. Regarding object detection, YOLOv8s is used as the baseline. In the backbone network, the convolution operations in shallow layers are replaced by PSConv module with a concentrated receptive field to extract more information of small targets; the convolution operations in deep layers are replaced by GhostConv, and the SimA-former module is employed to substitute the deepest C2f structure to achieve model lightweighting. During the feature fusion stage of the neck network, the coordinate attention mechanism (CA) and the small target detection head are combined to construct a high-resolution multi-scale feature fusion module. Based on the acquired fire scene information, the camera model is utilized to compute the relative position and orientation of the fire cannon and fire source. Subsequently, the required horizontal and pitch angles for the fire cannon adjustment are determined.

收稿日期: 2025-10-27 Received Date: 2025-10-27

* 基金项目: 国家自然科学基金(52204256)、江苏省重大科技示范项目(BE2023688)项目资助

Experiments were conducted in a custom-built fire scenario outside an industrial facility. The perception model achieved a mAP50 score of 92.3%, representing a 6.2% improvement over the YOLOv8s without dehazing preprocessing. The pre-alignment error in the horizontal angle was within $\pm 4^\circ$, while the distance estimation error for the fire source remained below 6%. Those results demonstrate the effectiveness and practical applicability of the proposed method.

Keywords: object detection; haze removal; UAV; fire scene

0 引言

我国火灾事故频发,智能化技术与装备的应用对提升应急救援效能具有重要战略意义^[1]。在消防救援工作中,地面设备视野通常有限,难以获得全面的火灾现场信息,而搭载云台相机和通信模块的无人机可在空中良好视角实时监测火场,获取丰富全面的火情信息并发送给地面消防设备,地面消防设备处理无人机信息,可以调整消防炮瞄准行为。消防炮瞄准分为预瞄准调整和精细控制两个环节,在灭火过程中,先进行预瞄准喷射水流,然后根据灭火效果进行动态精细角度控制,确保射流扑灭火源。本文旨在研究无人机视角下的火场信息感知,并探索基于无人机感知信息的消防炮预瞄准方法。

消防场景中非均匀烟雾降低了图像质量。图像去雾方法可以分为基于先验的方法和基于学习的方法^[2]。基于先验的去雾算法使用物理成像模型和先验统计规律估计无雾图像。大气散射模型(air scattering model, ASM)是一个经典的去雾先验模型,包含透射图和大气光值两个关键参数。文献[3]利用暗通道先验估计大气散射模型的透射图和大气光值,恢复清晰图像,但这种方法易在白色和高亮度区域造成颜色失真。文献[4]提出了一种非局部先验方法,通过 RGB(red, green and blue)空间中的颜色距离估计模糊图像的透射图。随着深度学习的发展,基于神经网络的去雾方法在近年来取得了显著研究进展。文献[5-6]使用卷积神经网络(convolution neural network, CNN)估计透射图。文献[7]采用两个 CNN 分支分别估计大气光值和透射图。文献[8-11]使用神经网络直接端到端输出去雾图像。

目标检测技术是消防场景下无人机感知火灾现场环境的重要技术。基于深度学习的目标检测方法可以分为两类。一种是基于区域建议的两阶段方法,如 R-CNN^[12]、Faster R-CNN^[13]等;两阶段方法首先在图像中生成多个候选区域,然后进行分类和检测。另一种是基于回归的单阶段方法,如单步多框检测器(single shot multibox detector, SSD)^[14]系列, YOLO(you only look once)^[15-19]系列等。单阶段方法直接对图像中的目标进行分类检测,在保证检测精度的同时保持较快速度。关键点检测算法可以分为热图法和回归法^[20],热图法预测每个关键点的概率分布图,精度较高但计算量大;回归法

直接预测关键点坐标值,模型简单且计算快速。目前一般场景的视觉检测任务已经取得了理想的研究成果,然而,无人机视角下的图像背景复杂,小目标比例高,通用视觉算法易发生错检漏检现象。针对无人机图像的小目标检测,现有研究在多个方面取得了研究成果。在数据集方面,天津大学的 AISKEYE 团队制作了 VisDrone 数据集^[21],由各种无人机拍摄,涵盖各种场景。注意力机制能够提升神经网络的特征提取能力,被广泛研究并应用于目标检测任务中^[22]。卷积块注意力模块(convolutional block attention module, CBAM)^[23]整合了空间注意力和通道注意力,提高了模型在各种环境下的检测效果。文献[24]在 YOLOv8 模型引入 BiFormer 机制对骨干网进行优化,提高了模型对关键信息的关注。文献[25]提出了基于卷积融合网络的多模态目标检测模型,将可见光和红外数据联合处理,提升了模型应对不同场景的能力。文献[26]采用模型剪枝技术,结合轻量化策略实现了精度和速度之间的良好平衡。文献[27]采用多尺度 Transformer 模块、特征融合模块和小目标检测层设计了一个感知增强网络,有效提升了小目标检测性能。

以上研究为图像去雾和目标检测提供了新颖高效的方法和思路,但对于消防场景下的视觉感知任务仍存在两方面问题:1)一般的图像去雾任务中,导致图像退化的雾通常是 220 以上的高灰度值,而燃烧产生的烟雾具有多样的灰度值,传统的图像去雾算法在消防场景的效果不佳;2)考虑到消防救援场景的实时性要求,不宜使用过于复杂的神经网络结构增加计算时间。针对消防救援实际场景的复杂性和有限的计算设备,亟需开发一种轻量而有效的实时感知模型。

针对以上问题,本文做出的工作贡献为:

1)提出一种联合轻量化目标检测与去雾处理的消防场景感知模型。首先,对输入图像先进行去雾预处理。改进大气散射模型公式使其更适用于分布着不均匀火光和烟雾的消防场景,对其中参数采用编码器-解码器神经网络结构求解。然后,对图像去雾预处理后的图像进行目标检测任务。以 YOLOv8s 为基线,改进骨干网络中的卷积层,使得性能提升的同时保持轻量化;在其颈部网络中加入一个高分辨率特征融合模块,保留更多特征图细节信息;将轻量化的注意力机制加入网络结构中,关注网络重要信息。

2) 提出一种基于无人机感知信息的消防炮预瞄准方法。无人机使用上述算法感知地面火源目标和消防设备,并基于相机成像模型处理感知结果,计算得到消防炮位姿和火源位置,进而调整消防炮的水平角度和俯仰角度,实现消防炮预瞄准。

1 基于无人机视觉信息的消防炮预瞄准系统

基于无人机视觉信息的消防炮预瞄准系统如图 1 所示。该系统由无人机和消防机器人组成,无人机装备云台相机和无线通信模块,消防机器人装备工控计算机和消防炮。该系统的灭火工作流程具体为:

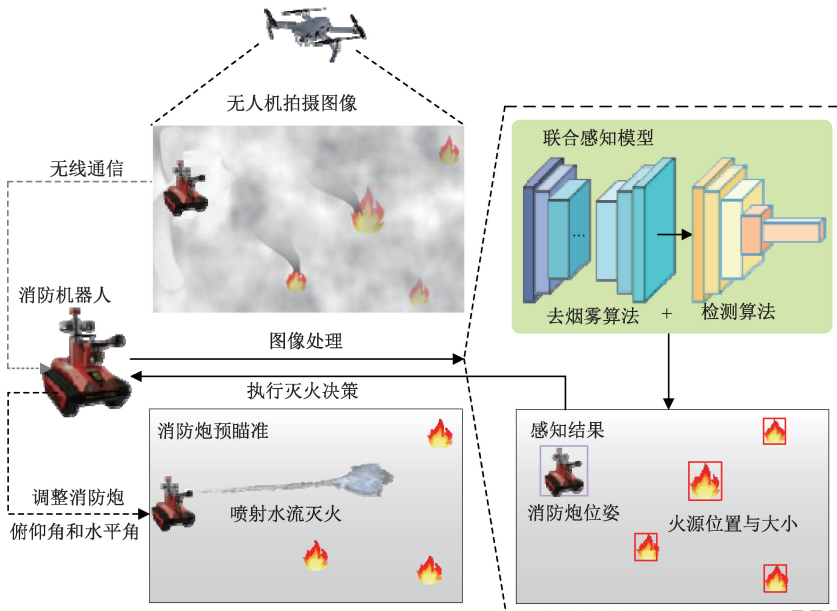


图 1 基于无人机视觉信息的消防炮预瞄准系统

Fig. 1 Firefighting cannon pre-aiming system based on UAV visual information

2 图像去烟雾算法

2.1 改进的大气散射模型

在消防场景中,火焰的燃烧通常会带来非均匀分布的光线和烟雾,在空气中形成烟雾气溶胶,射流灭火时也会产生非均匀分布的雾气,这些现象导致相机成像中的烟雾具有不确定的灰度值,图像的劣化过程如图 2 所示。

基于传统的 ASM 进行图像去雾任务通常将图像分解为清晰层和雾霾层,假设图像场景中具有均匀的雾霾分布,且雾霾层灰度值默认为固定的白色,因此不适合应用于消防场景。本文改进了大气散射模型,增加了一个待估计参数 $c(x)$ 取代大气散射模型中的 α , 如式(1)所示。

$$I(x) = J(x)t(x) + c(x)(1 - t(x)) \tag{1}$$

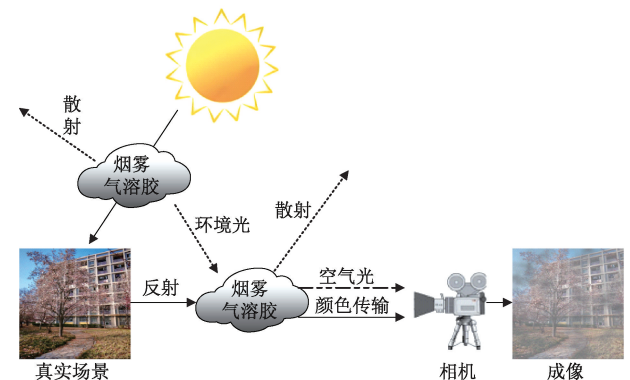


图 2 烟雾使图像劣化的过程

Fig. 2 Image degradation caused by smoke

式中: $J(x)$ 是理想的无雾霾图像; $I(x)$ 是输入的有雾霾图像; $t(x)$ 是透射图,其值表示光线经由物体反射后到达

相机的部分; $c(x)$ 表示图像中的雾霾颜色灰度值; $1-t(x)$ 表示光线被烟雾气溶胶散射或吸收。求解 $t(x)$ 和 $c(x)$,即可获得估计的去雾霾图像。

2.2 图像去烟雾神经网络

为求解 $t(x)$ 和 $c(x)$,设计了一个轻量的编码器-解码器结构的神经网络,如图3所示。编码器网络由浅层和深层构成。浅层部分参考ResNet^[28]的设计思想,缓解梯度消失和梯度爆炸问题。网络深层采用FasterNet-Block提取特征,降低计算量。解码器网络由2个分支组成,分别输出 $t(x)$ 和 $c(x)$ 。解码器网络的每个分支中网络层的结构类似于U-Net,由转置卷积层、通道拼接层和残差模块组成,解码器深层的残差模块采用FasterNet-Block,浅层的残差模块中增加了空间注意力机制(spatial attention, SA)^[23]和高效通道注意力机制(efficient channel attention, ECA)^[29],空间注意力机制可以增强图像的颜色、轮廓等底层特征的表达能,降低重建过程中图像中原本较为清晰的区域的失真;高效通道注意力机制能够使透射图更关注全局特征和语义信息,生成更自然的图像重建结果。

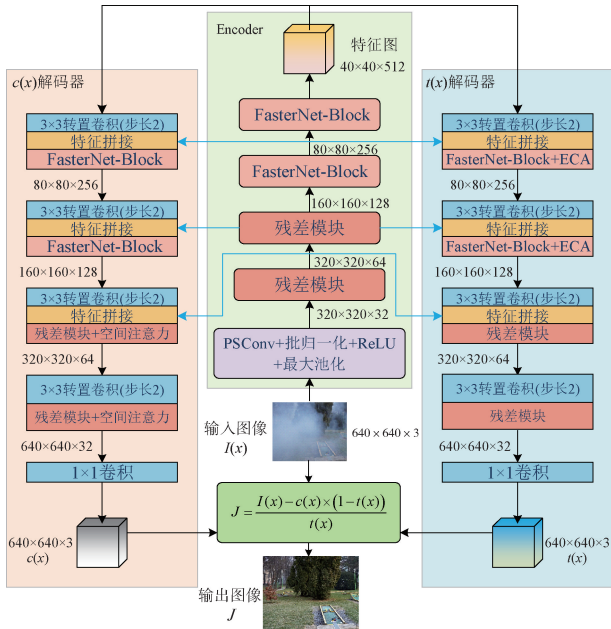


图3 图像去烟雾神经网络

Fig. 3 Image dehazing neural network

FasterNet-Block如图4所示,由部分卷积(partial convolution, PConv)^[30]和普通卷积残差连接组成,PConv对输入通道中的部分连续特征进行卷积运算, c_p 为参与卷积的通道,其余特征通过恒等映射处理,保持通道不变。

FasterNet-Block通过减少浮点运算的内存访问频率,增加了浮点运算操作的效率,具有结构简单,参数量少,

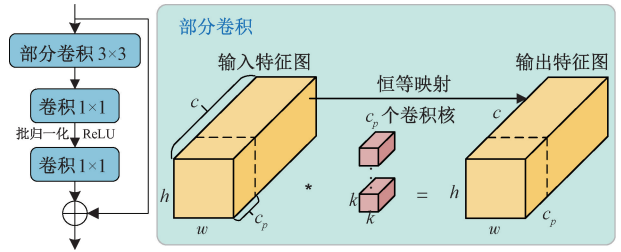


图4 基于部分卷积PConv的FasterNet-Block结构

Fig. 4 Structure of FasterNet-Block based on PConv

运算速度快的优点。FasterNet-Block的每秒浮点运算次数(floating point operations per second, FLOPs)如式(2)所示。

$$FLOPs_{FNB} = h \times w \times (k^2 + 2) \times c_p^2 \quad (2)$$

2.3 损失函数

对于2.2节中设计的去烟雾神经网络,训练过程中使用的联合损失函数由L1损失、感知损失、结构相似性损失加权组成,如式(3)所示。

$$L = \gamma_1 L_{L1} + \gamma_2 L_p + \gamma_3 L_{MS-SSIM} \quad (3)$$

式中: $\gamma_1, \gamma_2, \gamma_3$ 是超参数,设置其值为1、0.01、0.4。L1损失函数的计算公式如式(4)所示。

$$L_{L1} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (4)$$

为了提高特征空间中模型输出和真实值之间的相似性,使用感知损失函数。感知损失函数采用VGG-16作为损失网络^[31],并使用其倒数第1、第2和第3层的输出参与计算,如式(5)所示。

$$L_p = \frac{1}{3} \sum_k \frac{\phi_k^v(\hat{J}) - \phi_k^v(J)}{C_k H_k W_k} \quad (5)$$

式中: $\phi_k^v(\cdot)$ 表示VGG-16网络的第 k 层特征图,相应的特征图形状为 (C_k, H_k, W_k) 。

结构相似性损失函数(structural similarity loss function, SSIM Loss)从亮度、对比度和结构3个方面来衡量图像的相似性,如式(6)所示。

$$L_{MS-SSIM} = 1 - SSIM = 1 - \frac{(2\mu_x \mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (6)$$

式中: μ_x, μ_y 表示图像在 x 和 y 方向上的均值; σ_x, σ_y 表示图像在 x 和 y 方向上的标准差; c_1, c_2 为常数,作用是避免分母变为0。

该去雾神经网络通过上述编码器-解码器神经网络实现透射率 $t(x)$ 与烟雾灰度 $c(x)$ 的并行估计,然后基于改进的大气散射模型(式(1))将输入图像恢复为无烟雾图像。通过上述去烟雾算法,实现消防场景图像增强,为目标检测算法提供高质量数据支持。

3 目标检测

消防场景中,无人机拍摄到的火源和消防机器人目标往往尺寸较小,增大了目标及其关键点的检测难度。为提高对小目标及其关键点的实时检测性能,以 YOLOv8 模型为基线进行优化改进,改进的 YOLOv8 网络结构如图 5 所示。

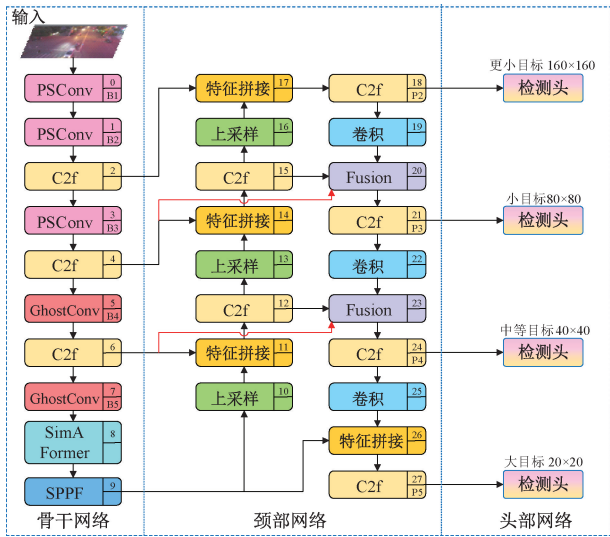


图 5 改进的 YOLOv8 网络结构

Fig. 5 Structure of improved YOLOv8 network

将骨干网络的浅层卷积运算替换为具有更集中感受野的 PSConv,以保留更多小目标及其关键点的特征信

息;将骨干网络的深层卷积运算替换为计算量更少的 GhostConv,以轻量化所改进的模型;在颈部网络中加入一个高分辨率特征融合模块,将骨干网络的信息自适应加权到颈部网络,结合新增的小目标检测头提升网络性能;将骨干网络中最深层的 C2f 模块替换为 SimAFormer,使用 L1 正则化取代 Transformer 中计算量较大的 softmax 层,降低模块计算量的同时发挥 visual Transformer 的架构优势,增强模型在复杂场景中的性能。

3.1 改进的卷积模块

小目标及其关键点由于自身所占像素更少,特征往往更加集中。基于卷积神经网络的网络模型在下采样提取特征的过程中,如果卷积层感受野较小,可能无法完全覆盖小目标,无法提取到小目标及其关键点的完整特征;如果感受野较大,又可能引入过多无关的背景信息,干扰对小目标特征的提取。PSConv^[32] 相比普通卷积,具有类似于高斯分布的感受野,更适合捕捉小目标特征,且参数量相比普通卷积更低。鉴于此,将 PSConv 应用于 YOLOv8 骨干网络浅层,替换 B1、B2 和 B3 层的卷积模块,以增强骨干网络对小目标的特征提取能力。

传统的卷积神经网络在堆叠大量卷积层时,会产生冗余的特征图,造成计算资源的浪费,这一现象在通道数更多的网络深层更加明显。GhostConv 能够生成大量的特征图,而所需的计算资源少于传统卷积操作。鉴于此,将 GhostConv 应用于 YOLOv8 骨干网络的深层,替换掉 B4 和 B5 中的卷积模块,以轻量化网络模型。

1) PSConv

PSConv 模块架构如图 6 所示。

PSConv

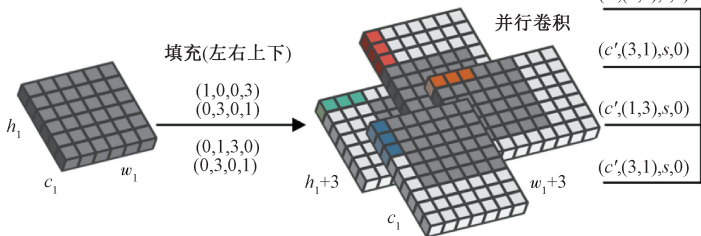
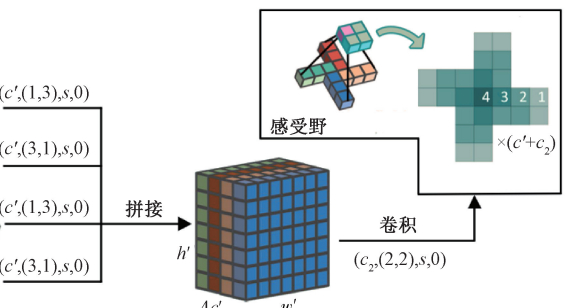


图 6 PSConv 模块

Fig. 6 PSConv module

使用不对称填充为图像的不同区域创建水平和垂直卷积核,在第 1 层执行并行卷积运算,将并行卷积的结果在通道方向上进行拼接,拼接后的张量形状参数如式(7)所示。

最终该张量由通道数为 c_2 的 2×2 卷积核进行归一化,使输出特征图的高度和宽度被调整为 h_2 和 w_2 。



$$\begin{cases} h_2 = h' - 1 = \frac{h_1}{s} \\ w_2 = w' - 1 = \frac{w_1}{s} \\ c' = \frac{c_2}{4} \end{cases} \quad (7)$$

PSConv 感受野的有效性向外减弱,图 6 的所示的 PSConv 感受野为 25,且卷积运算的数量由中心向外减少,类似于高斯分布。对于 $k=3$ 的卷积运算,PSConv 的参数量为 $7c_1c_2$ (见式(8)),少于标准卷积的 $9c_1c_2$ 参数量。

$$Params_{PSConv} = 4((c_2/4) \times c_1 \times 3 \times 1) + 4c_2c_1 = 7c_2c_1 \quad (8)$$

2) GhostConv

GhostConv 将卷积操作划分为少量卷积和轻量化线性变换操作两部分,工作方式如图 7 所示。首先,对输入特征图进行常规卷积操作。然后,将得到的固有特征图通过一系列线性变换操作生成“Ghost”特征图,这一过程显著减少了计算量。最后,将这些 Ghost 特征图与固有特征图进行拼接,形成最终的输出特征图。

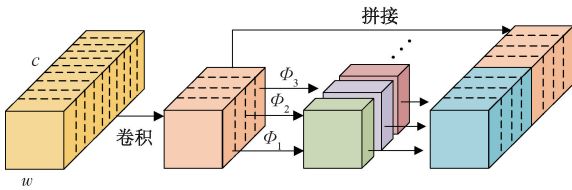


图 7 GhostConv
Fig. 7 GhostConv

3.2 高分辨率特征融合机制

YOLOv8 的颈部网络基于传统的特征金字塔-特征聚合网络 (feature pyramid network - path aggregation network, FPN-PAN) 结构进行特征融合,如图 8(a) 所示,在 FPN-PAN 结构中,小尺度特征图在网络深层经历了多次下采样后易导致特征丢失或减弱,为了进一步融合多尺度特征,改进 YOLOv8 的颈部网络如图 8(b) 所示。首先,添加了一个 160×160 的小目标检测头 (P2 层),更高分辨率的特征图能保留更多细节信息,降低错检误检概率;然后,受 EfficientDet 中 BiFPN^[33] 的启发,为骨干网络的 B3 层和 B4 层各添加一条特征融合路径,以进行多尺度自适应特征融合。

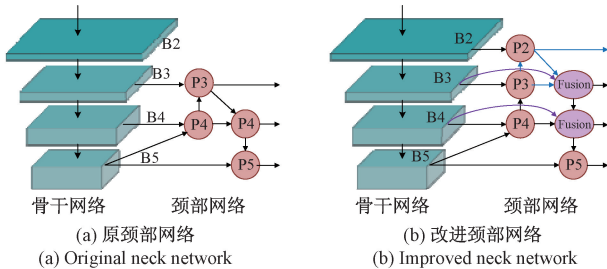


图 8 改进的 YOLOv8 颈部网络

Fig. 8 Improved YOLOv8 neck network

具有 3 路特征输入的 Fusion 模块结构如图 9 所示,其处理流程步骤为:首先,骨干网络输出的特征 X_B 经坐标注意力模块^[34] 处理得到增强特征 X'_B ;其次,颈部网络

的特征 X_1 与 X_2 进行通道维度拼接;随后通过可学习权重参数 w 与 X'_B 进行自适应加权融合。改进的坐标注意力 (coordinate attention, CA) 机制如图 10 所示,将原始模块中的平均池化替换为平均+最大池化操作以提升特征表征能力,构建的两组一维全局池化路径能有效聚合垂直/水平方向的空间特征,生成具备方向感知能力的中间特征,该中间特征经空间维度拆解后,分别生成水平方向特征图和垂直方向特征图,二者通过独立编码获得对应的空间注意力图。最终,双注意力图通过逐点相乘作用于输入特征,实现多方向长程依赖关系的建模。改进的坐标注意力机制将位置信息嵌入到通道注意力中,使网络能够更关注区域细节,同时避免产生显著的计算开销。

特征融合

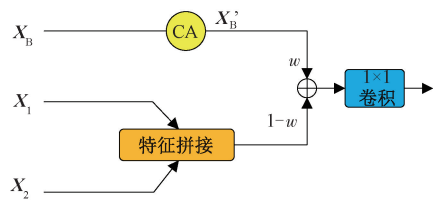


图 9 三输入特征融合模块

Fig. 9 Three-input feature fusion module

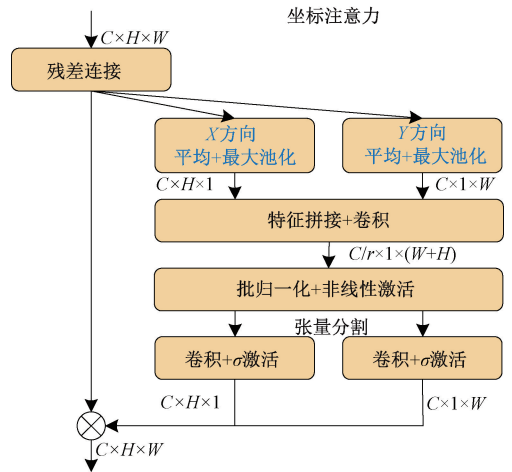


图 10 坐标注意力机制

Fig. 10 Coordinate attention mechanism

3.3 SimA-Former 模块

注意到 YOLOv8 模型中,骨干网络最深层的 C2f 模块占据了骨干网络将近 40% 的参数量。为轻量化模型,采用 SimA-Former^[35] 模块替换骨干网络最深层的 C2f, SimA-Former 模块的工作原理如图 11 所示。

SimA-Former 在 Q 矩阵和 K 矩阵相乘之前使用 L1 正则化来实现 token 之间的竞争性,取代原始 visual Transformer 中运算成本高昂的 softmax 层,即将每个通道的值除以该通道在所有 tokens 中的 L1 范数,使用这种归一化方法,自注意力块可以写为如式(9)所示。

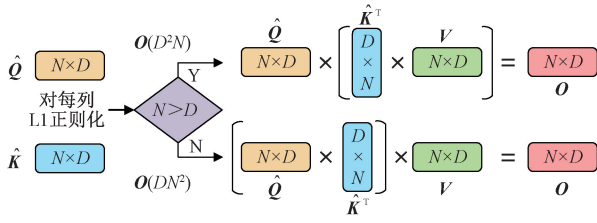


图 11 SimA-Former 计算流程

Fig. 11 The calculation of SimA-Former

$$\begin{cases} O = \hat{Q}\hat{K}^T V \\ \hat{Q}_i = \frac{Q_i}{|Q_i|_1} \\ \hat{K}_i = \frac{K_i}{|K_i|_1} \end{cases} \quad (9)$$

式中: Q_i 是 Q 的第 i 列 (令牌的第 i 个通道的值); \hat{Q} 和 \hat{K} 是归一化的查询和关键字矩阵, 与标准的视觉 Transformer 类似, 为每个注意力头使用上述方法, 将多个头的输出拼接, 最后进行输出投影。

4 消防炮预瞄准

消防炮预瞄准效果取决于对水平与俯仰角度的控制。预瞄准具体工作流程为: 无人机在高处感知火场, 对火场中火源和消防机器人进行目标检测和关键点检测, 得到的检测结果通过相机成像模型进行数学运算, 解得火源实际位置; 根据关键点检测的结果得到消防炮姿态朝向, 然后由姿态朝向和火点位置, 调整消防炮俯仰角及水平角, 对火源点喷射射流灭火。

4.1 俯仰角调整

基于相机成像模型的火源定位原理如图 12 所示。

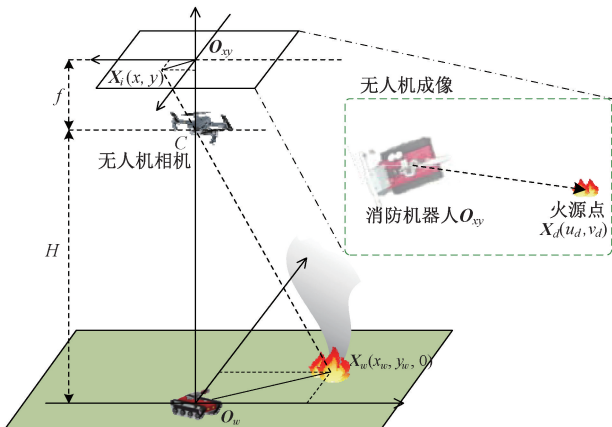


图 12 火源定位原理

Fig. 12 The demonstration of fire localization

调整无人机云台姿态, 使成像平面与地平面保持平行状态。在世界坐标系中, 设消防炮位置为原点 $O_w(0, 0, 0)$, 地面火源点坐标为 $X_w(x_w, y_w, 0)$, 将 X_w 映射到像素坐标系中火源点 $X_p(u, v)$ 的步骤为:

1) 通过刚体变换将 X_w 转换为相机坐标系下的点 $X_c(x_c, y_c, H)$, 转换公式如式 (10) 所示。

$$\begin{cases} X_c = R X_w + t \\ R = \begin{bmatrix} \cos \alpha & -\sin \alpha & 0 \\ \sin \alpha & \cos \alpha & 0 \\ 0 & 0 & 1 \end{bmatrix} \end{cases} \quad (10)$$

式中: R 为旋转矩阵, 由无人机云台相机姿态决定。当成像平面与地平面平行时, 旋转矩阵只与偏航角 α 有关; $t = [x_c, y_c, H]^T$ 为世界坐标系到相机坐标系的平移向量; H 为无人机高度。

2) 由三角形相似原理将相机坐标系下的点 X_c 投影到焦距为 f 的成像平面, 得到图像坐标 $X_i(x_i, y_i)$, 转换公式如式 (11) 所示。

$$\begin{cases} x_i = f \cdot \frac{x_c}{H} \\ y_i = f \cdot \frac{y_c}{H} \end{cases} \quad (11)$$

3) 基于像素尺寸 dx, dy 和原点偏移 (u_0, v_0) , 将图像坐标 X_i 转换为像素坐标 $X_p(u, v)$, 转换过程如式 (12) 所示。

$$\begin{cases} u = \frac{x_i}{dx} + u_0 \\ v = \frac{y_i}{dy} + v_0 \end{cases} \quad (12)$$

整理上述转换过程, 得到火源点的成像方程如式 (13) 所示。

$$H \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \underbrace{\begin{bmatrix} f/dx & 0 & u_0 \\ 0 & f/dy & v_0 \\ 0 & 0 & 1 \end{bmatrix}}_K \cdot [R | t] \cdot \begin{bmatrix} X_w \\ 1 \end{bmatrix} \quad (13)$$

式中: $[R | t]$ 为相机外参矩阵, 由无人机云台相机位姿决定; K 为相机内参矩阵。由于镜头的光学特性限制, 实际拍摄时相机会产生图像畸变, 导致成像效果失真。畸变后的火源点坐标 $X_d(u_d, v_d)$ 如式 (14) 所示。

$$\begin{cases} u_d = u \cdot \left(1 + \sum_{i=1}^3 k_i r^{2i}\right) + 2p_1 u v + p_2 (r^2 + 2u^2) \\ v_d = v \cdot \left(1 + \sum_{i=1}^3 k_i r^{2i}\right) + 2p_2 u v + p_1 (r^2 + 2v^2) \end{cases} \quad (14)$$

式中: $r = \sqrt{u^2 + v^2}$ 为点到光心的距离, 使用张正友标定法进行相机标定, 可得到相机内参矩阵 K 和畸变系数 $[k_1, k_2, p_1, p_2, k_3]$ 。

实际相机拍摄畸变后的火源点坐标 $X_d(u_d, v_d)$, 对式(14)的方程求数值近似解, 完成畸变校正, 将所得 $X_p(u, v)$ 代入式(13)的成像模型, 解算获得地面火源点的三维空间坐标 $X_w(x_w, y_w, 0)$ 。基于该三维坐标数据, 可计算消防炮车与火源点之间的实际空间距离 $O_w X_w$, 根据此距离, 结合消防炮射程及水压参数实现消防炮俯仰角调整。

4.2 水平角调整

针对图13所示消防机器人的形状, 设置如图所示的关键点, 在尾部设置 P_b 关键点, 在4个角点设置 P_{lu} 、 P_{ru} 、 P_{ld} 、 P_{rd} 这4个关键点(下标 l, r 表示左、右; u, d 表示上、下)构成凸四边形 Q , 其对角线交点 O 为消防炮位置。过点 O 作垂直 $P_{lu}P_{ru}$ 的向量 \vec{r}_1 和垂直 $P_{ld}P_{rd}$ 的向量 \vec{r}_2 , 过点 P_b 作垂直 $P_{lu}P_{ru}$ 的向量 \vec{r}_3 和垂直 $P_{ld}P_{rd}$ 的向量 \vec{r}_4 。 \vec{r}_1 、 \vec{r}_2 、 \vec{r}_3 、 \vec{r}_4 的方向均指向消防机器人前方。

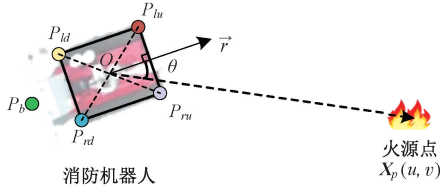


图13 消防机器人姿态朝向识别

Fig. 13 Firefighting robot pose and orientation recognition

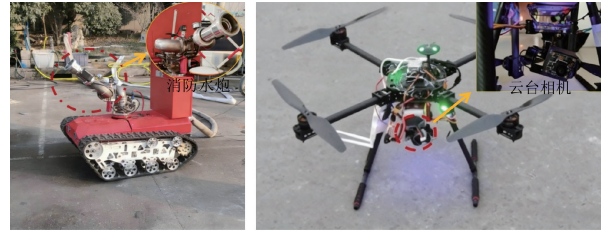
在理想的关键点检测结果中, Q 为矩形, \vec{r}_1 、 \vec{r}_2 、 \vec{r}_3 、 \vec{r}_4 这4个向量平行, 其方向均为消防机器人姿态朝向 \vec{r} 。但实际测得关键点坐标存在细小波动, 破坏了 Q 中边之间的平行关系, 导致 \vec{r}_1 、 \vec{r}_2 、 \vec{r}_3 、 \vec{r}_4 之间存在细微夹角。为降低误差, 令 $\vec{r} = \vec{r}_1 + \vec{r}_2 + \vec{r}_3 + \vec{r}_4$, 得到消防机器人姿态朝向 \vec{r} 。求 \vec{r} 和火源位置向量 $\overrightarrow{OX_p}$ 的向量夹角 θ , 根据 θ 调整消防炮水平角。

通过以上方法完成消防炮水平角和俯仰角的调整, 即可实现消防炮对火源点的预瞄准。

5 实验结果分析

5.1 实验数据集与环境配置

为进行消防炮预瞄准实验, 搭建了如图14所示的实验环境, 在工业厂房外部用汽油点燃了各种火源, 搭载消防炮的消防机器人喷射水流灭火, 无人机全程监视灭火过程, 将火场信息传输给消防机器人处理。消防机器人基于CAN总线通信协议驱动消防炮, 消防炮射流速率为30 L/s, 水压为0.11 MPa。无人机云台相机的焦距为24 mm, 分辨率为1280 pixels×720 pixels。



(a) 消防机器人
(a) Fire-fighting robot

(b) 无人机
(b) Unmanned aerial vehicle



(c) 实验场地
(c) Experimental site

图14 实验环境与设备

Fig. 14 Experimental environment and equipment

为训练火场感知模型, 本研究收集无人机拍摄的图像, 对所得图像进行旋转、缩放、剪切、拼接等数据增强操作。此外, 选取了一些其他消防场景图像以增强数据多样性。共得到了960张有效图片, 选取其中780张作为训练集, 180张作为测试集。实验的计算机配置包括 Intel® Core™ i5-13600KF 3.50 GHz CPU、NVIDIA GeForce RTX 3080 Ti GPU、32 GB RAM、Windows 10 操作系统、CUDA 版本 11.6 和 Python 3.8。

5.2 图像去烟雾

1) 训练设置

去烟雾神经网络的训练需要包含烟雾的数据集图像。根据消防场景烟雾非均匀分布和灰度值多样的特点, 采用如下方法处理数据集: 在120~255随机采样灰度值, 将烟雾层以0.1~0.8随机采样的透射率加到原图像, 在随机采样过程中使用了Perlin噪声以得到更自然的烟雾分布。

本文提出的去烟雾算法利用卷积神经网络求解先验公式参数, 因此采用类似于DehazeNet的初始化方案, 从标准正态分布中随机抽取以初始化每一层权重, 偏差设置为0。训练过程中初始学习率为0.001, 每30次迭代降低10%, 批次大小设置为4, 迭代次数设置为500, 使用Adam优化器训练神经网络, 输入神经网络的图像尺寸调整为640×640。

2) 性能评价

对于图像去雾任务, 经典的评价指标是峰值信噪比 (peak signal-to-noise ratio, PSNR) 和结构相似性

(structural similarity index, SSIM), 本文采用这两个指标定量评估去烟雾性能。较高 PSNR 值表明图像失真减少, 除雾性能优越。其计算公式如式 (15) 所示。

$$PSNR = 10 \times \lg \left(\frac{MAX_I^2}{MSE} \right) \quad (15)$$

式中: MAX_I 是图像中的最大像素值; MSE 是均方误差。结构相似性 SSIM 基于 3 个关键特征衡量去雾效果: 图像亮度、对比度和结构, 评估结构相似性, 较高 SSIM 值表示去雾图像中的失真较小, 质量较高。结构相似性的计算公式如式 (6) 所示。在模型复杂程度方面, 从参数量和 GFLOPs 两方面进行评价。GFLOPs 表示模型每秒钟执行的浮点运算次数。

将训练好的去雾网络与其他方法进行对比, 实验结果如表 1 所示。对比方法为暗通道先验法 (dark channel prior, DCP)^[3]、AODNet (all-in-one dehazing network)^[36]、GridDehazeNet^[37]、FFA-Net^[8]、TNN (two-branch neural network for dehazing)^[9]、DeHamer^[10]。

由表 1 结果可以看出, 本文所提去烟雾方法相比传统的 DCP、AODNet 和 GridDehazeNet 性能表现有较大优势。使用了双分支神经网络的 TNN 方法取得了优秀性

表 1 去烟雾算法性能对比

Table 1 Performance comparison of dehazing algorithms

方法	PSNR	SSIM	参数量	GFLOPs
原图	13.15	0.66		
DCP	15.69	0.72		
AODNet	18.78	0.74	1.76 K	1.68
GridDehazeNet	19.43	0.77	702.47 K	271.95
FFA-Net	24.32	0.81	4.46 M	287.50
TNN	29.76	0.85	50.35 M	68.36
DeHamer	26.24	0.88	29.44 M	48.93
本文	27.91	0.86	8.20 M	24.81

能。DeHamer 方法将 Transformer 和 CNN 结合使用, 使用位置嵌入模块将烟雾密度信息引入网络模型, 处理非均匀烟雾的能力更强, 但 TNN 和 DeHamer 方法计算量开销较高, 难以满足消防工作的实时需求, 本文的去烟雾算法在性能和参数量两方面取得了良好的平衡, 能够为消防炮预瞄准任务快速提供高质量图像。直观的图像去烟雾效果如图 15 所示。

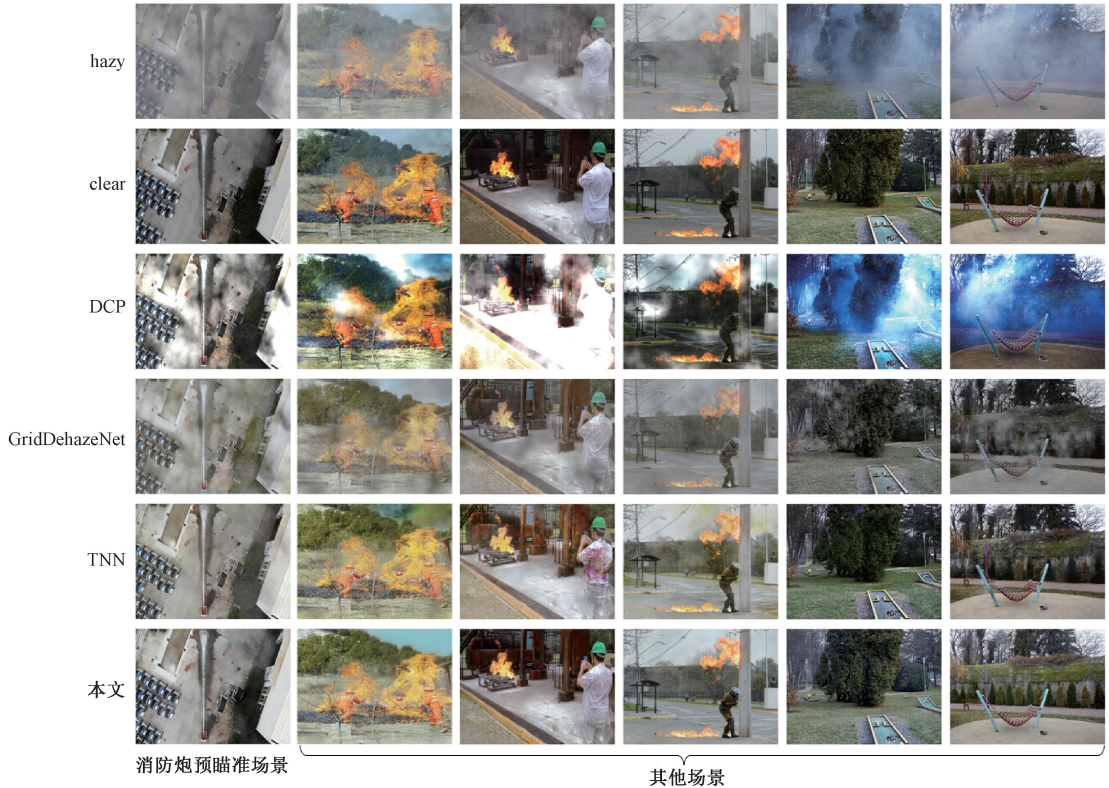


图 15 去雾算法效果对比

Fig. 15 Comparison of dehazing effects of various algorithm

可以看出, 在处理非均匀烟雾图像时, DCP 方法表现不佳, 图像色彩有较为严重的失真现象, 且未能去除黑

烟。GridDehazeNet 仍有部分烟雾未能去除。TNN 方法可以去除图像中的烟雾, 但在部分区域产生了颜色失真

现象。与其他方法相比,本文算法能够直观高效去除消防场景的非均匀烟雾。

5.3 目标检测

1) 训练设置

在自制消防场景数据集训练改进的 YOLOv8s 模型,检测目标为搭载水炮的消防机器人及其关键点、火源、射流落点和人,批次大小 Batchsize 设置为 8,初始学习率设为 0.01,终止学习率设为 0.0001,采用 Adam 优化器,输入图像的尺寸为 640×640,迭代次数为 160。

2) 性能评价

(1) 评价指标

本文在检测性能和计算量开销两方面评价模型的检测效果,选取的评价指标为精确率 (precision, P),召回率 (recall, R),平均精确率均值 (mean average precision, mAP),参数量和 GFLOPs。精确率 P、召回率 R、平均精确率 AP、平均精确率均值 mAP 的计算如式 (16) 所示。

$$\begin{cases} P = TP / (TP + FP) \\ R = TP / (TP + FN) \\ AP = \int_0^1 P(R) d(R) \\ mAP = \sum_i AP_i / N \end{cases} \quad (16)$$

式中: P 为精确率,用于评判检测为正样本的概率; R 为召回率,用于评判检测框漏检情况,其值越大模型漏检率越低; AP 为平均精度,是 $P-R$ 曲线下的面积,表示模型在不同召回率下的平均精确率; mAP 为平均精度均值,衡量了模型在多个检测类别上的平均精度,是综合评价模型检测性能的重要指标, $mAP50$ 与 $mAP50:95$,分别表示 IoU 阈值取 0.5 时的 mAP 值,取 0.5~0.95 内等间隔选取 10 个 IoU 阈值时的 mAP 取平均值; TP 、 FN 和 FP 分别为正检框、漏检框和误检框数目; N 为检测类别数。

(2) 检测效果

使用训练好的改进 YOLOv8s 模型在测试集上进行测试,在测试集中加入烟雾干扰。先使用本文所提去烟雾算法预处理测试图像,然后对去雾图像进行目标检测。与基线 YOLOv8s 对比,实验结果如表 2 所示。

表 2 检测效果
Table 2 Detection effect (%)

方法	火焰	消防机器人	人	射流	mAP50
YOLOv8s	74.2	98.7	82.7	88.6	86.1
改进 YOLOv8s	79.5	98.8	86.6	90.2	88.8
去雾+YOLOv8s	80.4	98.8	83.0	92.5	88.7
去雾+改进 YOLOv8s	85.1	99.3	88.8	95.9	92.3

可以看出在不使用去烟雾算法时,改进 YOLOv8s 模型的 $mAP50$ 指标达到了 88.8%,比基线提升了 2.7%,在火焰和人这两类目标上的性能提升更为明显,分别达到了 5.3% 和 3.9%。而无人机视角下火焰和人在图像中的大小更小,说明本文改进策略能够有效增强 YOLOv8 的检测能力。使用本文所提去烟雾算法对图像进行检测前预处理,模型检测精度均获得提升,其中对火焰、射流等目标的检测精度提升幅度尤为突出。这一现象可归因于:火焰与射流本身具有高饱和度、强对比度的颜色特征,而烟雾干扰会导致色彩失真和对比度衰减,加剧颜色特征退化。本文的去雾算法通过双参数联合估计机制,有效缓解了烟雾导致的图像质量劣化,提升了检测精度。

为了更直观呈现出本文模型的效果,在自制数据集和其他消防场景图像中选取部分具有代表性的图像,进行多维度实验验证,结果如图 16 所示。

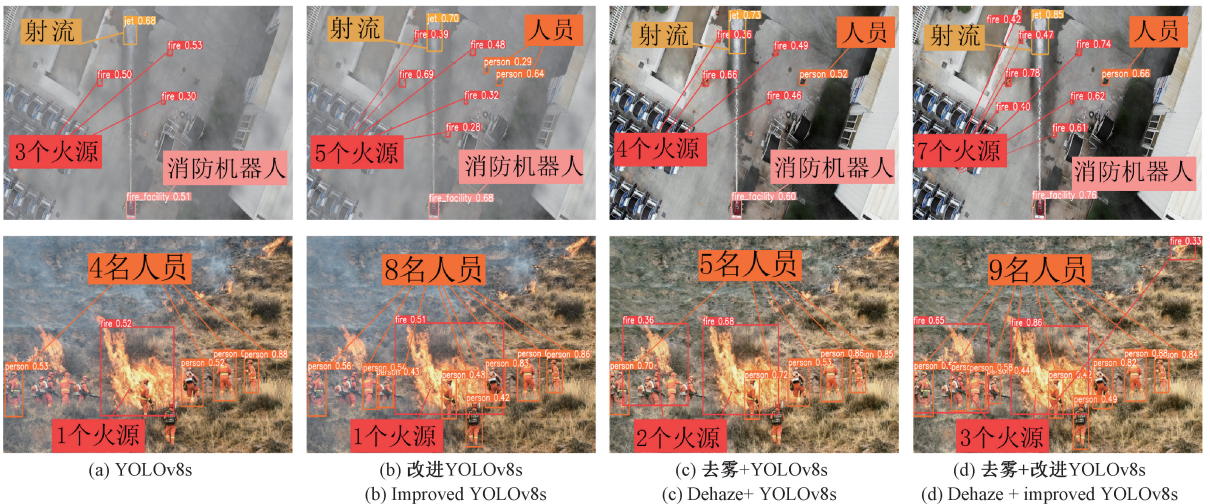


图 16 消防场景下的联合检测效果

Fig. 16 Joint detection effect in firefighting scenarios

在图 16 中自制的实验场景,包含 7 个大小不一的火源,1 辆正在喷射水流的消防车,还有 1 名监视人员,这些目标在图像中所占像素较少,检测难度较大。未应用去雾预处理时,YOLOv8s 模型漏检了 4 个火源和 1 名人员,应用了改进的 YOLOv8s 模型后,检测到了更多的目标,但仍存在漏检现象,且由于烟雾退化目标的特征,部分背景被错检为人员。应用去雾预处理后,YOLOv8s 检测到了更多目标,且检测精度提升,但仍然漏检了 3 个火源。再应用本文改进的 YOLOv8s 模型后,漏检的 3 个火源被全部检测到,对各类目标的检测精度也有所上升。在其他消防场景的图像上,未应用去雾预处理时,YOLOv8s 模型漏检了一个左侧较大火源和一些人员,应用了本文所提改进策略后,检测到了更多的人员。应用去雾预处理后,YOLOv8s 模型检测到了左侧较大火源,再应用本文改进的 YOLOv8s 模型进行检测,对火焰的检测精度提高,且检测到了更多左侧人员和右上方的一个小火源。以上的可视化实验结果直观地验证了本文方法的有效性。

实际火场中火源形态多变,为探讨本文模型的泛化性,选取一些不规则和含遮挡的火源进行检测,并与 YOLOv8s 基线模型进行对比,实验结果如图 17 所示。

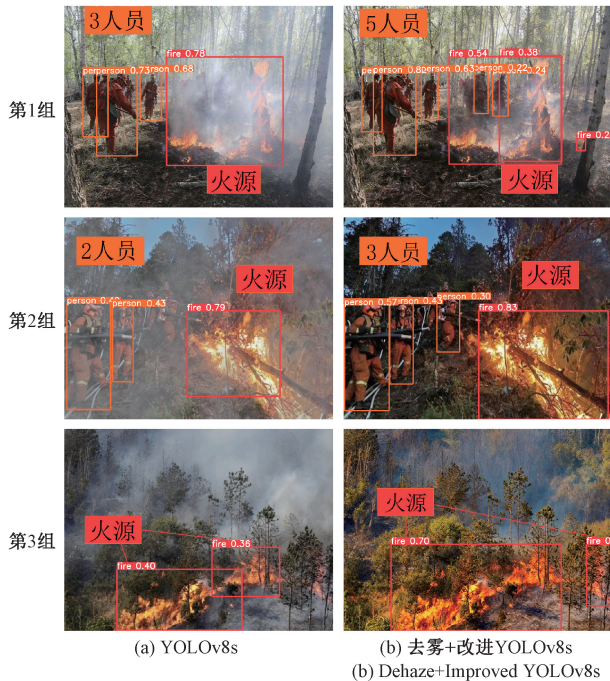


图 17 非典型火源检测效果

Fig. 17 The detection effect of atypical fire sources

可以直观看出,本文的检测模型面对遮挡和烟雾模糊情况具有更好表现,在第 1 组图像中,检测到了被烟雾遮挡的人员和小目标火源;在第 2 组图像中,检测框更好的覆盖了受树枝遮挡的火焰;在第 3 组图像中,检测到了

右侧被树木遮挡的火源。但是在火焰形状不规则的情况下,仍存在一些挑战,如第 1 组图像中,出现了检测框的重叠现象。

(3) 消融实验

为了验证本文对 YOLOv8s 基线模型所提改进的有效性,在去雾后的测试集上进行了消融实验,实验结果如表 3 所示,表中√表示使用了改进策略,表中的 A、B、C 改进分别指添加的高分辨率特征融合模块、骨干网络的改进卷积模块、SimA-Former 模块。

表 3 消融实验

Table 3 Ablation experiment for object detection

模块			mAP	mAP	参数量	GFLOPs	推理时
A	B	C	50/%	50:95/%	/M		间/ms
			88.7	54.8	11.1	28.6	7.9
√			91.5	61.0	13.4	41.3	11.2
	√		90.1	55.4	8.5	22.4	6.7
		√	88.9	55.0	9.6	24.7	7.9
√	√		92.2	62.5	10.8	35.1	8.6
√	√	√	92.3	62.8	9.3	31.2	8.6

单独加入高分辨率特征融合模块时,模型检测性能明显增加,mAP50 提高了 2.8%,mAP50:95 提高了 6.2%。但计算量也有所增加。加入改进骨干网络可以略微提升模型性能同时降低推理时间;SimA-Former 模块降低了模型的参数量和 GFLOPs,同时保持性能基本不变,这两项改进策略均是有效的轻量化方法。将改进卷积模块和颈部高分辨率特征融合模块共同使用时,模型性能进一步提升,mAP50 和 mAP50:95 相比基线分别提升了 3.5% 和 7.7%,表明 PSConv 对小目标特征的提取能力更强,能够辅助改良的颈部网络更有效地进行特征融合。最后,将改进策略结合到一起,mAP50 达到 92.3%,mAP50:95 达到 62.8%,相比基线分别提升了 3.6% 和 8.0%。参数量为 9.3 M,相比基线降低了 1.8 M,GFLOPs 和推理速度与基线几乎持平。通过上述消融实验可以看出,本文所提改进策略在保持了原 YOLOv8s 模型轻量化优势的同时增加了检测性能。

(4) 对比实验

为了进一步体现本文所提“先增强-再检测”联合感知策略的有效性,选取一些算法在自制消防场景数据集上训练,并在有烟雾干扰的测试集进行对比实验。本文感知模型为去雾算法和目标检测的级联,参数量为 8.2+9.3=17.5 M,GFLOPs 为 24.8+31.2=56.0,推理速度为 5.1+8.6=13.7 ms。所对比方法为 YOLOv5m、YOLOv8m、YOLOv9m、YOLOv10m。对比实验的结果如

表 4 所示,可以看出,本文方法相比 YOLOv5m 在检测性能上均有明显提升。相比 YOLOv8m 和 YOLOv9m,本文方法所需计算量更少,检测性能更佳。近年推出的 YOLOv10m 通过双重分配策略抛弃了对非极大值抑制 (non-maximum suppression, NMS) 的依赖,引入大核卷积和部分自注意力模块,在轻量化方面的表现与本文方法相近,但检测性能相比本文方法略有不足。对比实验结果验证了本文级联检测模型的有效性,在检测效果和模型轻量化两方面均有优秀表现。

表 4 对比实验

Table 4 Comparative experiment of object detection

方法	<i>mAP</i>	<i>mAP</i>	参数量 /M	<i>GFLOPs</i>	推理时间 /ms
	50/%	50:95/%			
YOLOv5m	86.2	57.0	21.2	49.0	15.3
YOLOv8m	89.8	58.6	25.9	78.9	14.1
YOLOv9m	90.9	59.4	20.1	76.8	13.8
YOLOv10m	90.8	60.1	15.4	59.1	12.4
本文	92.3	62.8	17.5	56.0	5.1(去雾) 8.6(检测)

5.4 消防炮预瞄准

在自制消防场景进行消防炮预瞄准实验。定性消防灭火实验效果如图 18 所示。

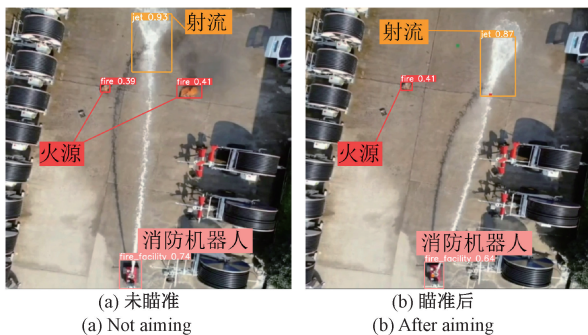


图 18 消防炮预瞄准实验

Fig. 18 Fire monitor pre-aiming experiment

由图 18 可以看出,经过预瞄准后,消防炮射流成功覆盖住右侧火源点。在定量评估实验中,采用水平角度误差和火源距离误差两方面定量评价瞄准效果,水平角度误差是指计算所得 θ 与射流完全覆盖火源时消防炮水平角度之间的误差,在 $-70^\circ \sim 70^\circ$ 每 10° 进行一次实验;火源距离误差是指计算所得火源距离与实际火源距离之间的误差,从 $15 \sim 60$ m 每隔 5 m 进行一次实验,实验结果如图 19 所示。

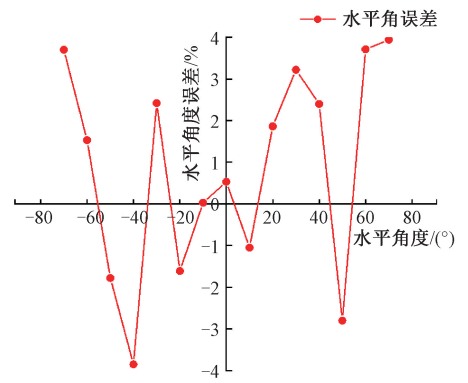
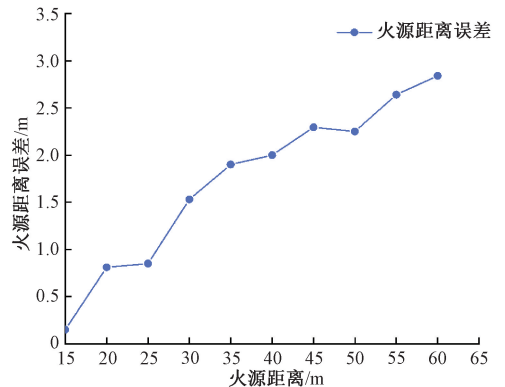
(a) 水平角度误差
(a) Error in horizontal angle(b) 火源距离误差
(b) Error in fire source location

图 19 瞄准误差分析

Fig. 19 Analysis of aiming error

由图 19 实验结果可以看出,消防炮预瞄准水平角度误差在 $\pm 4^\circ$,火源距离误差在 $0 \sim 6\%$,可以满足消防炮瞄准初调整的需求,实现预瞄准。为后续的精细化瞄准控制提供良好的初始状态。

6 结 论

本研究提出了一个基于无人机视觉信息的消防炮预瞄准系统。首先,提出一种图像去烟雾方法对无人机图像进行预处理,考虑消防场景成像过程,提出双参数可分离的改进大气散射模型,通过编码器-解码器网络联合求解透射率与烟雾灰度值,估计无烟雾图像。其次,对 YOLOv8 架构实施改进,在骨干网络中,将浅层卷积模块改为具有自适应感受野的 PSCConv,深层卷积模块改为计算量更少的 GhostConv,采用 L1 正则化约束的 SimA-former 模块替代深层 C2f 结构;在颈部网络中,应用多尺度坐标注意力机制,结合高分辨率特征图生成模块进一步强化特征提取和整合。然后,构建“去雾-检测”级联模型,对无人机图像先进行去烟雾预处理再进行检测,基

于检测结果解算消防炮与火源点的相对位姿信息,调整消防炮预瞄准角度。实验结果表明,本文感知方案在保持轻量化优势的同时具有优秀检测性能,平均精度达到了92.3%,较基线提升了6.2%,消防炮经过预瞄准调整后所喷射水流能够覆盖火源。综合以上成果,本研究提供了一套“火场感知-辅助决策”的灭火方案,为智能化、自主化消防工作提供了可靠技术参考,具有一定应用价值。

尽管本研究取得了一定进展,但仍存在一些局限性:面对浓烟、高温等强烈动态干扰时,无人机图传可靠性下降,难以支持消防机器人的决策控制;对于非典型和不规则形状的火源,模型的感知结果可能受影响。未来的研究将从增强系统可靠性和模型性能两方面继续深入,采用强化学习方法为无人机训练动态路径规划算法,自主导航至最佳观测视角以获取高质量火场信息;考虑端到端的联合训练方式,用检测的损失梯度反向校正去雾网络,提升模型性能;采用实例分割算法计算火源位置,提升对形状不规则火源的定位精度。

参考文献

- [1] DHIMAN A, SHAH N, ADHIKARI P, et al. Firefighting robot with deep learning and machine vision[J]. *Neural Computing and Applications*, 2022, 34: 2831-2839.
- [2] GUO X J, YANG Y, WANG CH Y, et al. Image dehazing via enhancement, restoration, and fusion: A survey[J]. *Information Fusion*, 2022, 86/87: 146-170.
- [3] HE K M, SUN J, TANG X O. Single image haze removal using dark channel prior[C]. 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009: 1956-1963.
- [4] BERMAN D, TREIBITZ T, AVIDAN S. Non-local image dehazing [C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition, 2016: 1674-1682.
- [5] CAI B L, XU X M, JIA K, et al. DehazeNet: An end-to-end system for single image haze removal[J]. *IEEE Transactions on Image Processing*, 2016, 25(11): 5187-5198.
- [6] REN W Q, PAN J SH, ZHANG H, et al. Single image dehazing via multi-scale convolutional neural networks with holistic edges[J]. *International Journal of Computer Vision*, 2020, 128: 240-259.
- [7] ZHANG J G, REN W Q, ZHANG SH D, et al. Hierarchical density-aware dehazing network [J]. *IEEE Transactions on Cybernetics*, 2021, 52(10): 11187-11199.
- [8] QIN X, WANG ZH L, BAI Y CH, et al. FFA-Net: Feature fusion attention network for single image dehazing[C]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020, 34(7): 11908-11915.
- [9] YU Y K, LIU H, FU M H, et al. A two-branch neural network for non-homogeneous dehazing via ensemble learning[C]. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 193-202.
- [10] GUO CH L, YAN Q X, ANWAR S, et al. Image dehazing transformer with transmission-aware 3D position embedding[C]. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 5802-5810.
- [11] GUO Y, GAO Y, LIU R W, et al. SCANet: Self-paced semi-curricular attention network for non-homogeneous image dehazing [C]. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 1885-1894.
- [12] 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, 2014: 580-587.
- [13] REN SH 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.
- [14] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector[C]. *Proceedings of the European Conference on Computer Vision*, 2016: 21-37.
- [15] 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, 2016: 779-788.
- [16] REDMON J, FARHADI A. YOLO9000: Better, faster, stronger[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition, 2017: 6517-6525.

- [17] REDMON J, FARHADI A. YOLOv3: An incremental improvement[J]. ArXiv preprint arXiv: 1804.02767, 2018.
- [18] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: Optimal speed and accuracy of object detection[J]. ArXiv preprint arXiv: 2004.10934, 2020.
- [19] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 7464-7475.
- [20] FAN ZH X, ZHU Y ZH, HE Y L, et al. Deep learning on monocular object pose detection and tracking: A comprehensive overview[J]. ACM Computing Surveys, 2022, 55(4): 1-40.
- [21] DU D W, ZHU P F, WEN L Y, et al. VisDrone-DET2019: The vision meets drone object detection in image challenge results[C]. 2019 IEEE/CVF International Conference on Computer Vision Workshop, 2019: 213-226.
- [22] GUO M H, XU T X, LIU J J, et al. Attention mechanisms in computer vision: A survey[J]. Computational Visual Media, 2022, 8(3): 331-368.
- [23] WOO S, PARK J, LEE J Y, et al. Cbam: Convolutional block attention module[C]. Proceedings of the European Conference on Computer Vision, 2018, 11211: 3-19.
- [24] WANG G, CHEN Y F, AN P, et al. UAV-YOLOv8: A small-object-detection model based on improved YOLOv8 for UAV aerial photography scenarios[J]. Sensors, 2023, 23(16): 7190.
- [25] 孙晓永, 孙备, 郭润泽, 等. 面向多无人机协同的多模态目标检测方法[J]. 仪器仪表学报, 2025, 46(2): 209-220.
- SUN X Y, SUN B, GUO R Z, et al. Multimodal target detection method for multi-UAV coordination [J]. Chinese Journal of Scientific Instrument, 2025, 46(2): 209-220.
- [26] 韩萍, 白继睿, 周杰龙, 等. 基于轻量化改进和模型剪枝的 SAR 图像飞机目标检测[J]. 仪器仪表学报, 2025, 46(9): 110-124.
- HAN P, BAI J R, ZHOU J L, et al. Lightweight SAR image aircraft target detection based on lightweight improvement and model pruning[J]. Chinese Journal of Scientific Instrument, 2025, 46(9): 110-124.
- [27] 张泽辰, 姚顺春, 梁晓俊, 等. 基于小目标感知增强的风机叶片缺陷智能检测[J]. 仪器仪表学报, 2025, 46(9): 159-172.
- ZHANG Z CH, YAO SH CH, LIANG X J, et al. Wind turbine blade intelligent defect detection based on small object perception enhancement[J]. Chinese Journal of Scientific Instrument, 2025, 46(9): 159-172.
- [28] HE F X, LIU T L, TAO D CH. Why resnet works? Residuals generalize[J]. IEEE Transactions on Neural Networks and Learning Systems, 2020, 31(12): 5349-5362.
- [29] WANG Q L, WU B G, ZHU P F, et al. ECA-Net: Efficient channel attention for deep convolutional neural networks[C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 11531-11539.
- [30] CHEN J R, KAO S H, HE H, et al. Run, don't walk: Chasing higher FLOPS for faster neural networks[C]. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 12021-12031.
- [31] JOHNSON J, ALAHI A, LI F F. Perceptual losses for real-time style transfer and super-resolution[C]. Proceedings of the European Conference on Computer Vision, 2016: 694-711.
- [32] YANG J N, LIU SH L, WU J J, et al. Pinwheel-shaped convolution and scale-based dynamic loss for infrared small target detection [J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2025, 39(9): 9202-9210.
- [33] TAN M X, PANG R M, LE Q V. Efficientdet: Scalable and efficient object detection [C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 10778-10787.
- [34] HOU Q B, ZHOU D Q, FENG J SH. Coordinate attention for efficient mobile network design[C]. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 13708-13717.
- [35] ABBASI KOOHPAYEGANI S, PIRSLAVASH H. SimA: Simple softmax-free attention for vision transformers[C].

2024 IEEE/CVF Winter Conference on Applications of Computer Vision, 2024: 2595-2605.

- [36] LI B Y, PENG X L, WANG ZH Y, et al. AOD-Net: All-in-one dehazing network [C]. 2017 IEEE International Conference on Computer Vision, 2017: 4770-4778.
- [37] LIU X H, MA Y R, SHI ZH H, et al. GridDehazeNet: Attention-based multi-scale network for image dehazing[C]. 2019 IEEE/CVF International Conference on Computer Vision, 2019: 7313-7322.

作者简介



张晓光,2011 年于中国矿业大学获得博士学位,现为中国矿业大学信息与控制学院教授,主要研究方向为智能感知、精确定位、应急通信。

E-mail:3641@cumt.edu.cn

Zhang Xiaoguang received her Ph. D. degree from China

University of Mining and Technology in 2011. She is currently a professor in the School of Information and Control Engineering at China University of Mining and Technology. Her main research interests include intelligent perception, precise positioning and emergency communication.



李伟(通信作者),2009 年于杜伊斯堡-埃森大学获得博士学位,现为中国矿业大学机电工程学院教授,主要研究方向为应急救援、视觉伺服控制和智能安全设备。

E-mail:cmeeetechologylw@163.com

Li Wei (Corresponding author) received his Ph. D. degree from the University of Duisburg-Essen, Duisburg, Germany, in 2009. He is currently a professor in the School Mechatronic Engineering at China University of Mining and Technology. His main research interests include fault diagnosis, visual servo control, and intelligent safety equipment.