

基于 YOLOv9 在雾天图像上的目标检测

沈建驰¹

SHEN Jianchi

摘要

深度学习目标检测在传统数据集上取得优异成果,但是在雾天环境下的检测任务仍然具有挑战性。雾天图片细节信息丢失,目标与背景混合,导致雾天目标检测任务难以定位与识别,容易造成漏检与误检问题。文章基于 YOLOv9 算法进行优化,结合 AOD-Net 去雾算法,在端到端的框架下,实现了从雾天图像输入到目标检测结果输出的一体化流程;提出 GNAM 模块,实现注意力权重的精细化调控,关注目标关键特征,避免误检;在 Neck 层使用 BiFPN 特征融合网络捕获目标的语义和位置信息,提高特征图质量,减少雾天环境下的漏检问题。更好地满足在雾天环境目标检测需求。实验结果表明,在 RTTS 数据集上的 mAP@0.5 和 mAP@0.5:0.95 分别达到了 82.6% 和 61.8%,能够有效应对雾天场景下的目标检测任务。

关键词

AOD-Net; 去雾; 目标检测; 注意力机制; BiFPN

doi: 10.3969/j.issn.1672-9528.2025.04.012

0 引言

目标检测作为计算机视觉任务之一,需要确定目标在给定图像中的位置以及每个目标属于哪个类别^[1]。这一任务在智能监控、自动驾驶等应用场景中扮演着关键“角色”。随着深度学习技术的发展,基于卷积神经网络(CNN)等算法的目标检测方法在多种环境中展现出卓越性能,能够在清晰的图像中实现高效的对象识别和定位。

然而,在雾天环境条件下,目标检测的效果显著下降。雾天环境下采集的图像由于大气中的微小颗粒对光线的散射作用,图像往往会出现对比度降低、细节模糊和颜色失真的问题^[2]。传统目标检测算法通常依赖于清晰的视觉信息进行特征提取和对象分类,由于目标物体的边缘和纹理特征在雾天条件下变得不清晰,以及雾霾的遮挡,导致远距离目标的可见性大幅下降,导致漏检问题。因此,探索在这种不利环境下提高目标检测的准确性和鲁棒性成为研究的关键。

近年来,为解决这一具有挑战性的问题,较为流行的方法是使用已有的图像去雾和图像增强算法进行预处理,如 He 等人^[3]提出的基于暗通道先验的去雾算法, Li 等人^[4]提出的单图像去雾深度卷积神经网络(PDR-Net)。通过去雾方法对图像进行预处理增强有雾图像的整体可见性,然后将预处理后的图像送入后续的检测网络进行目标检测。经过预处理后的图像目标清晰,细节丰富。然而额外地去雾步骤增加了计算量。还有一类则是将有雾图像直接送入目标检测模型,通过端到端的方式,深入学习输入图像到目标检测结果的映

射,根据模型的自适应能力,学习图像特征的复杂模式与规律。如 Huang 等人^[5]采用了双分支网络架构,通过共享两分支底层特征提高检测性能。Liu 等人^[6]采用生成对抗网络(GAN)生成多样化的雾天图像,扩充训练数据集,使模型在训练过程中学习更多雾天场景下的特征,从而提升对真实雾天图像的检测性能。这类方法通过大量的数据训练学习特征和规律,很难直观解释模型如何提取目标信息。

本文深度融合去雾模型和优化的检测算法,区别传统雾天图像目标检测中去雾与检测分离的模式。本文利用 AOD-Net^[7]增强 YOLOv9 的图像处理阶段,直接作用于输入的雾天图像,在端到端的框架下,实现了从雾天图像输入到目标检测结果输出的一体化流程,对图像进行实时去雾处理,去雾后的图像即刻被输入至 YOLOv9 算法,输出目标的类别、位置等信息,本文创新点如下:

- (1) 结合 AOD-Net 和 YOLOv9,在端到端的框架下,实现了从雾天图像输入到目标检测结果输出的一体化流程;
- (2) 提出 GNAM 模块,根据通道权重的重要性聚焦输入图片中的重要特征,提取更加丰富准确的特征,准确识别和定位目标;
- (3) 在颈部网络中使用双向特征金字塔网络(BiFPN),学习不同层级的 GNAM 的尺度特征信息,捕获通道扩张之前目标的位置信息,提高模型的定位精度。

1 改进后的模型整体网络

在这小节中,介绍了本文算法的基本构架如图 1 所示,考虑到雾天图片图像颜色失真细节丢失,特别是当目标与背景混合,目标边缘与轮廓模糊使得在目标检测任务难以分类与定位。因此在图像特征提取阶段之前,利用 AOD-Net 去雾

1. 南京审计大学 江苏南京 211800

算法对图像进行处理,帮助 Backbone 层更准确地提取目标的关键特征,避免因雾气造成的特征混淆。并提出一种高效的 GNAM 模块,GNAM 模块能够反映通道权重重要程度,在特征提取阶段聚焦输入特征中对目标最重要的部分,提高特征提取能力。同时随着 Backbone 层中空间信息逐步向通道传输,特征的空间压缩与通道扩展导致位置信息丢失。因此在 Neck 层中使用 BiFPN^[8]学习来自不同层级的 GNAM 的特征信息,捕获通道扩张之前目标的位置信息,通过加权特征的融合,提高模型的定位精度。

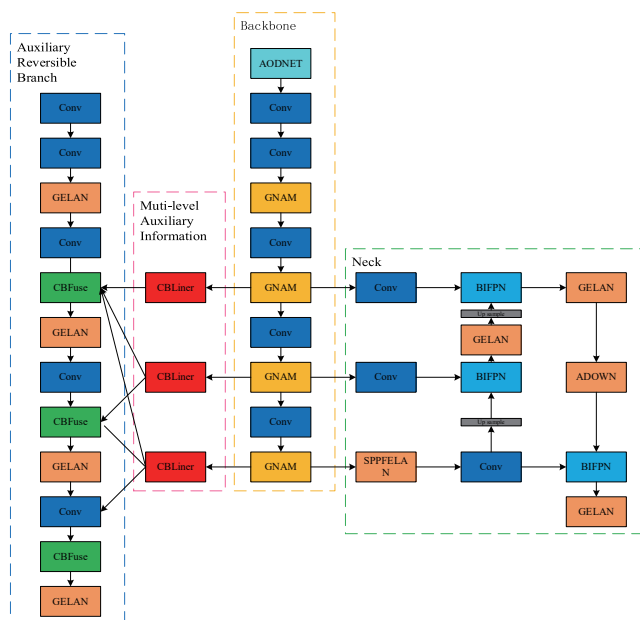


图1 改进后的 YOLOv9 网络结构

1.1 AOD-Net

针对雾天图像, Meng 等人^[9]建立了数学模型解释雾天图像成像过程以及雾天图像包含的各要素,大气散射模型公式为:

$$I(x) = J(x)t(x) + A[1 - t(x)] \quad (1)$$

式中: $I(x)$ 为雾天图像; $J(x)$ 为去雾图像; $t(x)$ 为透射率; A 为环境光强度。

AOD-Net 由 K 估计模块 (K-estimation) 和干净图像生成模块 (clean image generation) 组成,如图 2 所示。

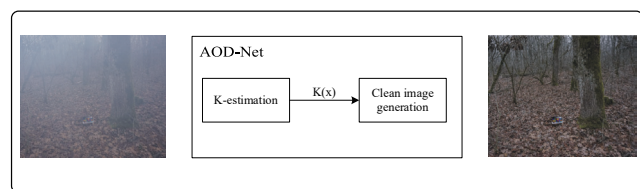


图2 AOD-Net 去雾流程

AOD-Net 首先经过 K 估计模块估算雾的深度和相对程度,该模块使用 5 个卷积层,通过融合不同大小的滤波器来形成多尺度特征,并使用 3 个连接层补偿卷积过程中的信息

损失。通过 $K(x)$ 整合 A 和 $t(x)$,最小化其单独计算的误差,用公式表示为:

$$K(x) = \frac{1}{t(x)} \frac{(I(x) - A) + (A - b)}{I(x) - 1} \quad (2)$$

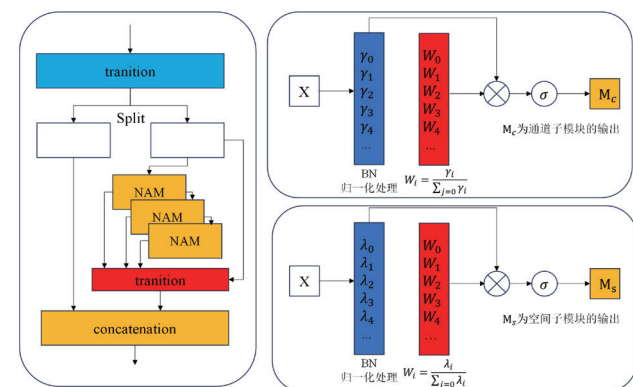
随后将干净图像生成模块,求出干净图像 $J(x)$,用公式表示为:

$$J(x) = K(x)I(x) - K(x) + b \quad (3)$$

本文将 AOD-Net 去雾网络模块作为 YOLOv9 模型接受图像后的第一个处理模块。作为整个网络的起始模块,进行去雾处理,为后续网络提取清晰的物体纹理、形状、轮廓等特征进行铺垫。

1.2 GNAM 模块

本文提出 GNAM 模块,如图 3 (a) 所示,通过将基础层的特征图分成直接连接和特征变换分支。直接连接分支最大程度地保留特征图的原始信息。特征变换分支使用 NAM 注意力机制进行特征提取,利用注意力机制聚焦图像中的重要信息。通过两分支的后续融合,实现互补,借助直接连接分支的原始信息,避免经过特征变换分支的特征信息的丢失。



(a) GNAM 示意图 (b) NAM 注意力模块

图3 GNAM 模块结构

NAM 是一种基于归一化的注意力模块,通过缩放因子计算权重方差来反映通道和空间的重要性,批量归一化公式为:

$$B_{out} = BN(B_{in}) = \gamma \frac{B_{in} - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}} + \beta \quad (4)$$

式中: B_{in} 和 B_{out} 分别为输入与输出的特征; γ 和 β 为可训练参数 (尺度和偏移); μ_B 和 σ_B 分别为平均值和标准差。NAM 的通道与空间子模块流程图如图 3 (b) 所示, X 为输入的特征图; γ_i 和 λ_i 分别为对应通道与空间子模块的缩放因子; W_i 为权重; W_C 为通道子模块的输出特征,其计算公式为:

$$M_C = \text{sigmoid}(W_\gamma(\text{BN}(X))) \quad (5)$$

W_S 为空间子模块的输出特征图，其计算公式为：

$$M_S = \text{sigmoid}(W_\lambda(\text{BN}(X))) \quad (6)$$

本文将 GNAM 模块作为 YOLOv9 的主分支负责对输入图像进行特征提取，生成多尺度的特征图。在 GNAM 的特征变换分支中使用 NAM 注意力机制，增强重要区域的特征权重，提高定位与分类的准确性。

1.3 加项双向特征金字塔网络 BiFPN

本文在 Neck 层使用加项双向特征金字塔网络 BiFPN，通过双向信息传递，使得特征图之间的信息融合更加充分。BiFPN 网络结构如图 4 所示，与传统特征融合如 FPN 相比，BiFPN 不仅具备自上而下的融合路径将高层语义信息传递至底层，同时还增加了自下而上的融合路径，将底层丰富的位置信息传导至高层。

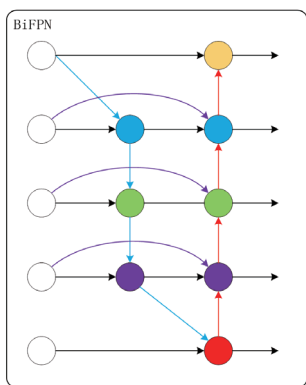


图 4 BiFPN

此外，Tan 认为不同的输入特征贡献不同，BiFPN 为每个输入特征进行加权，用公式表示为：

$$O = \sum_i \frac{w_i}{\sum_j w_j} \cdot I_i \quad (7)$$

式中： I_i 为输入特征； w_i 为可学习权重，由 ReLU 激活函数约束，并设 ϵ 为 0.000 1 避免数值不稳定。

求出加权特征后进行带权特征融合，其计算公式分别为：

$$P_i^{\text{td}} = \text{Conv} \left(\frac{w_1 \cdot P_i^{\text{in}} + w_2 \cdot \text{Resize}(P_{i-1}^{\text{in}})}{w_1 + w_2 + \epsilon} \right) \quad (8)$$

$$P_i^{\text{out}} = \text{Conv} \left(\frac{w_1' \cdot P_i^{\text{in}} + w_2' \cdot P_i^{\text{td}} + w_3' \cdot \text{Rsize}(P_{i-1}^{\text{out}})}{w_1' + w_2' + w_3' + \epsilon} \right) \quad (9)$$

式中： P_i^{td} 是第 i 层中间特征； P_i^{in} 是输入特征； P_i^{out} 是输出特征；Resize 是对特征图的上或下采样。

在 YOLOv9 的 Backbone 层中，空间信息逐步向通道传递，特征图经过空间压缩以及通道扩展会导致位置信息丢失。在雾天目标检测任务中，受雾气影响目标特征与边缘模糊，难以准确提取目标信息，在这种情况下，空间压缩和通道扩展

操作会进一步削弱目标特征的表达能力。因此在 Neck 层使用 BiFPN，通过两个阶段的特征传播，获取高层特征的语义信息以及底层特征的位置信息，提高特征融合能力，从而提高模型检测的准确性。

2 实验

2.1 实验数据集

本文旨在针对真实世界雾天环境下的目标检测，为此使用 RTTS 数据集进行实验。RTTS 是一个在雾天条件下可用的较为全面的真实世界数据集，包含 4 322 幅自然雾天图像作为训练集，以及 100 张真实场景图片作为验证集，涵盖了 5 类目标：人、自行车、汽车、公共汽车、摩托车。

2.2 实验环境设置

本文研究使用 Linux 操作系统，训练框架为 PyTorch 1.10，利用 Tesla T4 GPU 作为硬件加速平台。训练时 batch-size 设为 16，迭代次数为 100。本文实验使用精确率 P (Precision)、召回率 R (Recall)，以及不同 IoU 值下的平均精度均值 mAP (mean average precisions) 作为评价指标。

2.3 实验结果与分析

为了验证本研究的有效性，将 Fast RCNN、YOLO^[10] 等模型以及其他雾天目标检测算法在同一数据集下进行对比实验。从表 1 可知，相对其他目标检测模型，本文方法的指标 P 、 R 、mAP(IoU=0.5) 和 mmAP(IoU=0.5:0.95) 均为最优。

表 1 不同检测算法比较结果表

单位：%

方法	P	R	mAP	mmAP
Fast RCNN	66.3	53	51.3	37.8
Retinanet	67.2	54.9	53.1	39
DETR	68	56.4	56.9	42.2
YOLO5x	70.4	56	71.1	47.1
YOLOv5-Fog ^[11]	79.0	59	77.8	50.3
FPVDNet ^[12]	79.6	74.3	80	59.5
YOLOv9c	79.3	73.7	80.8	60.7
ours	81.4	77.6	82.6	61.8

为验证各改进模块的有效性，在 RTTS 数据中，以 YOLOv9 为基准进行了 4 组消融实验。实验过程中依次在模型中加入去雾模型 AOD-Net、GNAM 模块，以及 BiFPN 模块。根据实验的评价指标分析研究不同模块对实验影响，其中“√”表示模型中加入该模块，“×”表示未加入该模块。由表 2 可知，依次加入 AOD-Net、GNAM 和 BiFPN 后，与原始模型相比，各项指标均有提升， P 、 R 、mAP、mmAP 分别提升 2.1%、3.9%、1.8% 和 1.1%，可以更好完成雾天图像的目标检测任务。

表 2 消融实验对比结果表

单位: %

AOD-Net	GNAM	BiFPN	P	R	mAP	mmAP
×	×	×	79.3	73.7	80.8	60.7
√	×	×	80.3	74.3	81.4	61.1
√	√	×	79.7	77	82.4	61.5
√	√	√	81.4	77.6	82.6	61.8

实际检测结果如图 5、图 6 所示,第一列为原始模型的检测结果,第二列为改进后的模型检测结果,原始模型存在错检漏检现象,改进后的算法能够正确检测出雾天图像中的目标,且未出现漏检现象。根据消融实验可视化结果的对比,改进后的算法能很好地完成雾天图像目标检测任务,提高了模型的检测精度。



图 5 改进前检测结果



图 6 改进后检测结果

3 总结

本文提出一种结合 AOD-Net 和 YOLOv9 算法,实现了从图像输入到去雾再到目标检测结果输出的网络结构。在特征提取前进行去雾处理,为后续的网络提取清晰的特征打下基础。并提出 GNAM 模块,结合直接连接分支和特征变换分支,增强图像重要区域的特征权重提高定位与分类的精度,用于主分支的特征提取。并使用 BiFPN 网络结构用于特征融合,通过双向信息传递融合多尺度特征的语义信息和位置信息提高定位精度,避免漏检。实验表明,本文方法能有效避免原始 YOLOv9 模型在雾天图像目标检测中的漏检问题。与其他检测算法的对比,证明了本文方法的优越性。后续希望模型在不同的恶劣环境图像的目标检测中也有很好的结果,同时需要进一步探索轻量化的方法,减少因去雾模型而增加的网络参数和计算量。

参考文献:

[1] REN S Q, HE K M, SUN J, 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.

[2] LI Y, YOU S D, BROWN M S, et al. Haze visibility enhancement: a survey and quantitative benchmarking[J]. Computer

vision and image understanding, 2017,165:1-16.

[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. Piscataway: IEEE, 2009: 1956-1963.

[4] LI C Y, GUO C L, GUO J C, et al. PDR-Net: perception-inspired single image dehazing network with refinement[J]. IEEE transactions on multimedia, 2020, 22(3): 704-716.

[5] HUANG S C, LE T H, JAW D W. DSNet: joint semantic learning for object detection in inclement weather conditions[J]. IEEE transactions on pattern analysis and machine intelligence, 2021,43(8):2623-2633.

[6] LIU K H, YE Z H, GUO H Y, et al. FISS GAN: a generative adversarial network for foggy image semantic aegmentation[J].IEEE/CAA journal of automatica sinica, 2021, 8(8): 1428-1439.

[7] LI B Y, PENG X L, WANG Z Y, et al. AOD-Net: all-in-one dehazing network[C]//2017 IEEE International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2017: 4780-4788.

[8] 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 (CVPR). Piscataway: IEEE, 2020:10778-10787.

[9] MENG G F, WANG Y, DUAN J Y, et al. Efficient image dehazing with boundary constraint and contextual regularization[C]//2013 IEEE International Conference on Computer Vision. Piscataway: IEEE, 2013:617-624.

[10] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[J].2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2016: 779-788.

[11] WANG H, XU Y S, HE Y G, et al. YOLOv5-Fog: a multiobjective visual detection algorithm for fog driving scenes based on improved YOLOv5[J/OL]. IEEE transactions on instrumentation and measurement, 2022[2024-11-21]. <https://ieeexplore.ieee.org/document/9851677>.DOI:10.1109/TIM.2022.3196954.

[12] WANG Y D, GUO J C, WANG T B. Algorithm for foggy-image pedestrian and vehicle detection[J]. Journal of Xidian University, 2020, 47(4): 70-77.

【作者简介】

沈建驰(1998—),男,江苏淮安人,硕士研究生,研究方向:计算机视觉。

(收稿日期:2025-01-06)