

一种基于YOLOX_s的雾天场景目标检测方法

姜铮铮, 张欣, 胡世哲, 吴云鹏

引用本文

姜铮铮, 张欣, 胡世哲, 吴云鹏. 一种基于YOLOX_s的雾天场景目标检测方法[J]. 计算机科学, 2024, 51(7): 206-213.

LOU Zhengzheng, ZHANG Xin, HU Shizhe, WU Yunpeng. [Foggy Weather Object Detection Method Based on YOLOX_s](#) [J]. Computer Science, 2024, 51(7): 206-213.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[针对系统调用的基于语义特征的多方面信息融合的主机异常检测框架](#)

Host Anomaly Detection Framework Based on Multifaceted Information Fusion of SemanticFeatures for System Calls

计算机科学, 2024, 51(7): 380-388. <https://doi.org/10.11896/jsjcx.230400023>

[基于深度确定性策略梯度与注意力Critic的多智能体协同清障算法](#)

Multi-agent Cooperative Algorithm for Obstacle Clearance Based on Deep Deterministic PolicyGradient and Attention Critic

计算机科学, 2024, 51(7): 319-326. <https://doi.org/10.11896/jsjcx.230600129>

[融合多图卷积与层级池化的文本分类模型](#)

Text Classification Method Based on Multi Graph Convolution and Hierarchical Pooling

计算机科学, 2024, 51(7): 303-309. <https://doi.org/10.11896/jsjcx.230400164>

[基于联合学习的语言粒度融合的重叠事件抽取方法](#)

Overlap Event Extraction Method with Language Granularity Fusion Based on Joint Learning

计算机科学, 2024, 51(7): 287-295. <https://doi.org/10.11896/jsjcx.230700118>

[基于外部先验和自先验注意力的图像描述生成方法](#)

Image Captioning Generation Method Based on External Prior and Self-prior Attention

计算机科学, 2024, 51(7): 214-220. <https://doi.org/10.11896/jsjcx.230600167>

一种基于 YOLOX_s 的雾天场景目标检测方法

娄铮铮 张欣 胡世哲 吴云鹏

郑州大学计算机与人工智能学院 郑州 450000

(iezzlou@zzu.edu.cn)

摘要 文中提出了一个基于深度可分离卷积和注意力机制的雾天目标检测模型,旨在实现在雾天场景中对目标的快速、准确检测。该模型由去雾模块和检测模块组成,并在训练过程中共同训练。为确保模型在雾天场景中检测的准确性和实时性,在去雾模块方面,采用 AODNet 对输入图像进行去雾处理,以降低雾对图像中待检测目标的干扰,在检测模块中使用改进后的 YOLOX_s 模型,输出目标的分类置信度和位置坐标。为提升网络的检测性能,在 YOLOX_s 基础上采用深度可分离卷积和注意力机制来提高特征提取能力,扩大特征图感受野。所提模型能提高有雾场景中模型的检测精度,且不增加模型参数量和计算量。实验结果表明,所提模型在 RTTS 数据集和合成有雾目标检测数据集上均表现出色,有效提高了模型在雾天场景中的检测精度。与基准模型相比,平均精度(mAP@50_95)分别提升了 1.9% 和 2.37%。

关键词: 雾天场景;目标检测;图像去雾;深度可分离卷积;注意力机制

中图分类号 TP391

Foggy Weather Object Detection Method Based on YOLOX_s

LOU Zhengzheng, ZHANG Xin, HU Shizhe and WU Yunpeng

School of Computer and Artificial Intelligence, Zhengzhou University, Zhengzhou 450000, China

Abstract This paper proposes a foggy weather object detection model based on depth-wise separable convolution and attention mechanism, aiming to achieve fast and accurate detection of objects in foggy scenes. The model consists of a dehazing module and a detection module, which are jointly trained during the training process. To ensure the accuracy and real-time performance of the model in foggy scenes, the dehazing module adopts AODNet to perform dehazing processing on input images, reducing the interference of fog on the detected objects in the images. In the detection module, an improved version of the YOLOX_s model is used to output the confidence scores and position coordinates of the detected objects. To enhance the detection performance of the network, depth-wise separable convolution and attention mechanism are employed on the basis of YOLOX_s to improve the feature extraction capability and expand the receptive field of the feature maps. The proposed model can improve the detection accuracy of the model in foggy scenes without increasing the model parameters and computational complexity. Experimental results demonstrate that the proposed model performs excellently on the RTTS dataset and the synthesized foggy object detection dataset, effectively enhancing the detection accuracy in foggy weather scenarios. Compared to the baseline model, the average precision (mAP@50_95) is improved by 1.9% and 2.37% respectively.

Keywords Foggy scene, Object detection, Image dehazing, Depthwise separable convolution, Attention mechanism

1 引言

雾天场景下的目标检测旨在从有雾图像或视频等环境中定位和识别室外存在的目标,是目标检测方向的一个重要分支。雾天场景下的目标检测在城市监控、道路监控和自动驾驶等领域有着广泛的用途。近年来,卷积神经网络的飞速发展和大规模目标检测数据集的出现,使得基于卷积神经网络的目标检测方法迅速成为目标检测的主流方法之一。例如, Girshick 等提出的 R-CNN^[1] 将卷积神经网络运用于目标

检测中,通过卷积神经网络增强特征的提取能力,极大提升了目标检测的准确度。Redmon 等针对 R-CNN 系列目标检测算法检测时间长的问题,将目标检测任务视为一个回归问题,设计了一个基于卷积的端到端模型 YOLO^[2]。YOLO 算法将整张图片划分成多个网格,每个网格预测一组目标的位置和类别,对于每个网格,算法会输出包含目标的边界框,以及这个边界框所包含目标的类别概率,每个边界框还有一个置信度,用于表示这个边界框是否包含目标。YOLO 使用单个网格来预测所有目标的位置和类别,可以实现端到端的训练

到稿日期:2023-04-03 返修日期:2023-09-10

基金项目:国家自然科学基金青年科学基金(62002330,62206254)

This work was supported by the Young Scientists Fund of the National Natural Science Foundation of China(62002330,62206254).

通信作者:吴云鹏(ieypwu@zzu.edu.cn)

和推理,具有更快的检测速度和更高的实时性。Liu 等^[3]通过多个大小和纵横比不同的边界框来检测不同大小和形状的目标,提升模型对不同尺寸目标的检测精度。Ge 等^[4]针对目标检测正负样本不均衡和回归任务与分类任务矛盾的问题,设计了一个标签分配策略并将解耦头应用到 YOLO 系列中,从而降低了目标检测预测头的复杂度,提升了模型的检测精度。虽然基于卷积的方法在大部分通用数据集上表现出良好的性能,在检测精度和检测速度取得了很好的平衡,可以在精度没有明显下降的同时保持良好的实时性。但是,如何在恶劣环境下特别是雾天场景中更好地提取待检测目标的位置信息和语义信息,以及如何更好地区分雾天场景中的不同目标仍是目标检测任务中面临的重要问题。

目前大部分目标检测模型都使用由高质量图像组成的数据集进行训练,即目标清晰的数据集。然而真实环境比较复杂,会存在许多影响图像成像质量的因素,如恶劣天气等。若直接将这类模型运用到实际场景中,往往无法达到令人满意的检测效果,主要是因为真实环境中各种天气信息与待检测目标信息在成像时相互影响,导致成像质量下降,待检测目标与背景难以区分,进而导致检测性能下降。目前解决上述问题的方法主要分为两大类。第一类是使用现有的一些图像去雾或图像增强方法对输入图像进行预处理,消除天气信息和噪声,然后将预处理后的图像输入目标检测模型中进行训练和推理。Li 等提出一个端到端去雾网络 AODNet^[5],并将 AODNet 与 Faster R-CNN^[6]相结合进行雾天场景的目标检测,但是这类方法需要使用复杂的图像恢复网络,会影响目标检测模型的推理速度。第二类是使用网络对去雾任务和检测任务共同学习训练。Huang 等^[7]使用两个子网联合学习目标检测任务和图像恢复任务,由于二者会共享一些特征层,因此在训练过程中很难通过调整参数来平衡检测任务和去雾任务的权重。

为解决上述问题,本文提出了一种基于联合优化的雾天场景目标检测网络。该网络由去雾模块和检测模块两部分组成,从有雾图像中提取特征进行图像去雾,然后将去雾后的图像用于目标检测。为了让去雾后的图像能更有利于后续的目标检测任务,将用于去雾的卷积神经网络和用于目标检测的卷积神经网络联合优化,同时学习去雾任务和检测任务。去雾模块旨在从有雾图像中学习对目标检测任务最佳的无雾图像;检测模块旨在从无雾图像中定位和识别目标。最后,本文使用联合优化策略来同时训练 AOD-Net 和改进的 YOLOX_s,以实现目标检测任务和图像去雾任务的联合学习,并在 RTTS 数据集和合成有雾场景目标检测数据集上验证 DA_YOLOX 的有效性。

2 相关工作

2.1 目标检测

目标检测是计算机视觉领域不可或缺的一部分,被广泛应用于自动驾驶、交通监控和机器人视觉等领域。传统的目标检测算法使用滑动窗口方法生成大量的候选区域,然后使用手工设计的特征提取方法对每个候选区域进行特征提取,并将这些特征送入分类器进行分类,但这种方法需要耗费

大量时间,且手工设计的特征无法满足现实场景的需求,因此现在越来越多的基于深度学习的目标检测方法得到了应用。

近年来,随着深度学习的快速发展,基于深度学习的目标检测方法已经能很好地平衡检测精度与速度,可以在保证检测精度的同时,满足实时应用的需求。目前,基于深度学习的目标检测方法主要分为两大类,即基于回归的一阶段方法和基于候选框的二阶段方法。基于回归的一阶段方法使用端到端的方式,直接根据输入图像预测物体位置信息和类别信息,检测速度很快,实时性较好,但精度不如基于候选框的二阶段目标检测算法。2015 年,Redmon 等提出 YOLO 算法^[2],将目标检测作为一个回归问题,使用神经网络直接从输入图像中预测边界框和其对应的类别概率,实现一个端到端目标检测模型。2016 年,Liu 等提出 SSD 算法^[3],在 YOLO 算法回归思想的基础上,有效结合多尺度检测的思想和 Anchor 机制,显著提高了目标的检测效果。2017 年,Redmon 等提出 YOLOv2 算法^[8],通过引入多尺度特征提取、AnchorBox、特征融合和改进的损失函数,提升了模型的检测精度和速度,在多个数据集上取得了优秀的性能。2018 年,Redmon 等提出 YOLOv3 算法^[9],使用 ResNet^[10]中跳跃连接的思想,在提升精度的同时还减少了模型计算量,并使用 3 个不同尺度的特征图来进行目标检测,提高了对不同尺度大小物体的检测精度。2020 年,Bochkovskiy 等提出 YOLOv4 算法^[11],引入 Mosaic 数据增强、CSPNet^[12]等思想对 YOLOv3 进行了一系列改进,检测精度和检测速度都得到了明显提升。2021 年,Ge 等提出 YOLOX 算法^[4],将解耦头引入 YOLO 系列,通过 anchor free 机制和标签分配策略进一步提升目标检测精度并在检测精度和速度上取得了很好的平衡。2022 年,Li 等提出 YOLOv6 算法^[13],该算法通过自我蒸馏策略和通道级蒸馏等技术,对目标检测的量化方案进行了优化。此外,YOLOv6 根据不同的应用场景,设计了不同规模的网络结构,以在速度和准确性之间实现最佳平衡。2023 年,Wang 等提出 YOLOv7 目标检测算法^[14],其采用了多种可训练的 bag-of-freebies 方法,显著提高了模型的检测精度,但未增加推理成本。同时,YOLOv7 还引入了扩展和复合缩放方法,更有效地利用了参数和计算资源。

2.2 雾天场景下的目标检测

相对于一般的目标检测任务,雾天场景下的目标检测任务受到的关注较少。现有的解决方案可以分为两种:第一种方法是使用传统的或深度学习的图像增强方法或图像去雾方法对输入图像进行预处理,然后将处理后的图像送入目标检测模型中进行训练;第二种方法是将目标检测模型与去雾模型相结合或使用领域自适应任务,同时学习训练目标检测任务和图像去雾任务,这种方法能够在不影响图像质量的情况下,提高目标检测模型在雾天场景下的性能。例如,2017 年,Li 等将 AODNet^[5]与 Faster R-CNN 共同训练优化,有效提升了模型在合成的有雾场景下的目标检测精度;2020 年,Huang 等^[7]使用两个子网同时学习图像增强任务和检测任务,显著提升了在恶劣环境下的目标检测精度;2021 年,Xie 等^[15]提出了基于知识引导的目标检测框架和基于图像去雾和目标检测的联合学习框架,通过先验知识和模型的联合

优化来提升真实有雾场景下的目标检测性能;2021年,Hnewa等提出了一种多尺度领域自适应YOLO(MS-DAY-OLO)框架^[16],该框架将多个领域自适应路径和相应的领域分类器引入YOLOv4目标检测器,以生成领域不变的特征,在具有挑战性的天气条件下目标检测性能有明显提升;2022年,Liu等^[17]提出了一个自适应增强图像的可微分图像处理模块,考虑到恶劣的天气条件,模型使用小型卷积神经网络学习可微分图像处理模块中的超参数,使用YOLOv3作为检测网络,以端到端的方式共同学习两个网络,以增强图像进行检测,该方法在雾天和低光场景中取得了良好的性能。2023年,Li等^[18]针对雾天交通场景下的目标检测任务,提出了一个双主干网络,该网络以并行方式,有效提取两个网络各自在不同特征层的信息,并在不同网络层实现两个网络的通道拼接,提升了雾天交通场景下目标检测的精度。

由于第一种方法是去雾任务和检测任务无关联的方法,经过网络去雾后的图像可能会出现伪影、颜色失真等问题。因此,这种方法并不总能提升目标检测精度。而第二种方法的模型较为复杂且难以平衡去雾网络和检测网络权重。针对上述问题,本文提出了一种雾天场景目标检测模型DA_YOLOX,使用一个联合优化策略将去雾网络和检测网络结合起来,在保持模型大小基本不变的情况下,提高了雾天场景下模型的检测精度。

3 本文方法

3.1 整体概述

雾天场景中图像对比度下降、图像前景背景难以区分、图像成像质量降低,使得传统目标检测方法难以准确识别目标。而现有的雾天场景目标检测方法大多是去雾和检测无关联的方法,这类方法忽略了图像中雾气对目标的相互作用,导致目标检测模型的检测效果并不一定能够得到提升。针对上述问题,本文提出了一种基于联合优化的雾天目标检测模型DA_YOLOX,为保证雾天目标检测模型检测的实时性与准确性,本文选择使用在检测精度和检测速度两方面有着较好平衡的YOLOX_s目标检测模型作为检测模块的基准模型,去雾模块使用网络结构轻巧、简单高效的图像去雾模型AODNet。为实现雾天场景下的目标检测,本文通过对YOLOX_s的骨干网络和预测头进行改进,提升网络的特征提取能力,但没有增加模型的参数量和计算量。然后将优化后的检测网络与AODNet结合在一起,使用相同的损失函数,共同训练优化。去雾和检测联合优化网络的大致工作流程如图1所示,首先输入图像经过由AODNet组成的去雾模块得到输入图像对应的无雾特征图,然后将无雾特征图作为检测模块的输入送入检测模块中,最终得到待检测目标的坐标和位置信息。

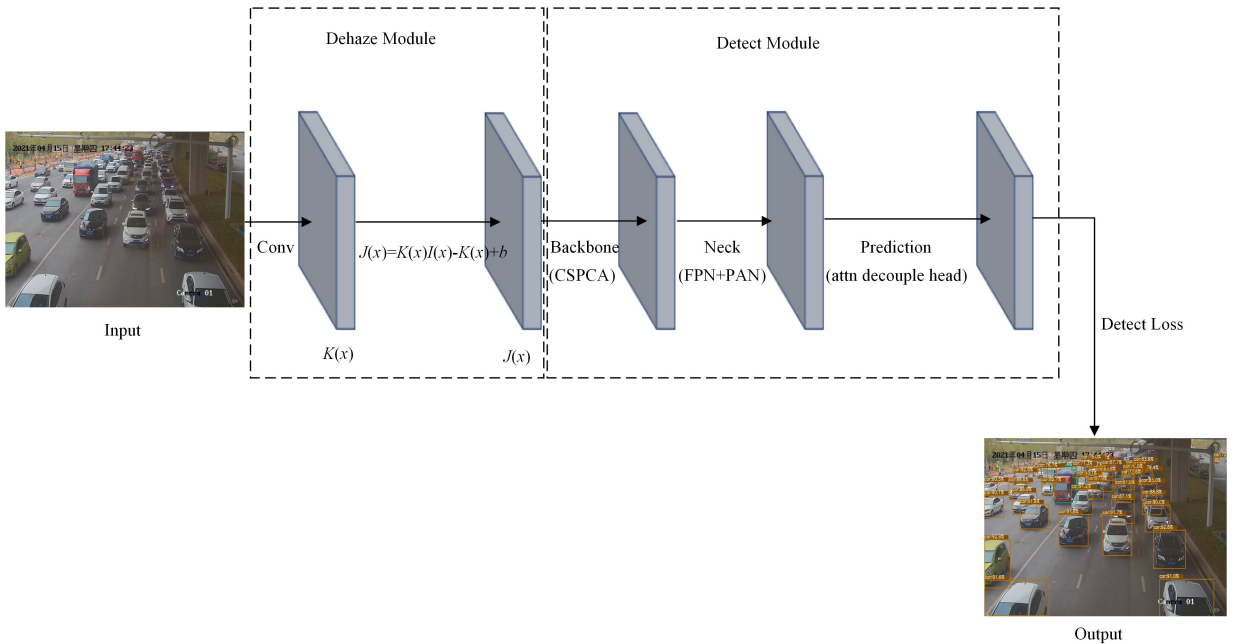


图1 DA_YOLOX模型结构图

Fig. 1 DA_YOLOX model structure diagram

3.2 检测模块

为了在雾天场景下更有效地提取图像特征,同时保持模型检测的实时性,本文选取YOLOX_s作为检测模块的基准模型。针对YOLOX_s网络深度较浅,在处理复杂场景如雾天时,存在对特征图中重要区域的关注不足的问题,本文从骨干网络和预测头两方面对模型进行改进,旨在提升模型在特征提取和目标检测方面的性能。与基准模型相比,改进后的模型精度提升1%左右,同时模型参数量减少了 0.51×10^6 ,

仅为 8.43×10^6 ,但仍保持相当的计算复杂度,改进后的检测模块如图2所示。整个检测模块分为骨干网、颈部网和预测头3部分。骨干网采用本文设计的基于深度可分离卷积卷积和注意力机制的特征增强模块CSPCA提取图像特征;颈部网采用FPN结构和PANet结构对骨干网中不同尺度的特征图进行融合;预测头采用注意力解耦头对颈部网融合后的特征图进行回归和分类,给出目标的边界框、目标的类别和置信度。接下来对各部分的改进进行详细阐述。

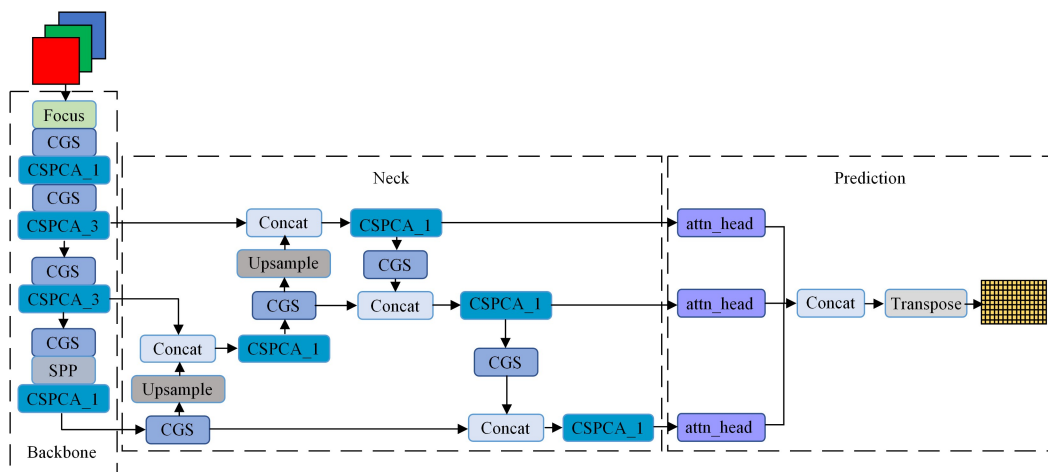


图2 检测模块结构图

Fig. 2 Structure diagram of detection module

3.2.1 CSPCA 模块

受注意力机制和深度可分离卷积的启发,本文设计了 CSPCA 模块,它采用 7×7 的深度可分离卷积,以减小模型大小,扩大特征图的感受野。通过跨层连接,将不同层次的特征融合,以提取全局上下文信息。CSPCA 还引入了注意力机制,以对融合后的特征图进行加权处理,进一步提高特征的表达力,CSPCA 模块结构如图 3 所示。CSPCA 模块使用两个 1×1 卷积对特征图进行压缩,然后对其中一个分支使用 BottleNeck 进行特征提取,帮助网络获取更丰富的信息。本文使用深度可分离卷积和跨层连接来改进 BottleNeck。改进后的 BottleNeck 将输入特征图分成两个分支,其中一个分支使用 7×7 的逐通道卷积生成多个特征图,紧接着使用逐点

卷积合并特征图,使用 1×1 卷积调整通道数。最后将两个分支的特征图进行合并得到最终的输出特征图。新的 BottleNeck 结构通过卷积核大小为 7×7 的深度卷积使卷积操作覆盖更多的输入特征图区域,从而扩大特征图感受野。此外,CSPCA 模块还引入了注意力机制,以增强特征图的表达力。具体来说,CSPCA 模块对融合后的特征图使用全局平均池化来降维,从而得到各个通道的均值,然后使用 1×1 卷积学习每个通道的权重,最后将这些权重向量应用于融合后的特征图,以生成增强的特征表示。CSPCA 模块利用 7×7 的深度可分离卷积扩大特征图的感受野,并使用注意力机制来加强网络对特征图中重要区域的关注程度。这种方法可以有效提升模型的特征提取能力,同时模型参数量和计算量均略微减少。

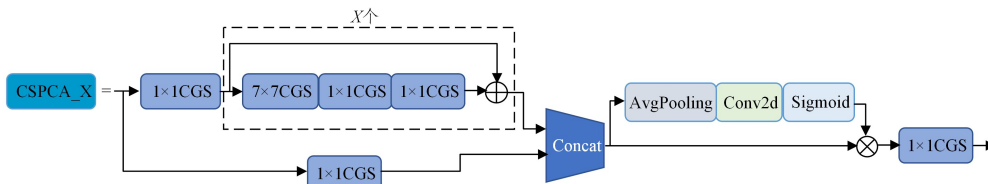


图3 CSPCA 模块结构图

Fig. 3 Structure diagram of CSPCA module

3.2.2 注意力解耦头

在目标检测任务中,分类子任务和回归子任务之间的矛盾是一个众所周知的问题。YOLOX 通过从其他先进的目标检测模型中吸取经验,将解耦头成功应用于 YOLO 系列模型中,以提高精度^[19]。虽然 YOLOX 使用解耦头将分类子任务和回归子任务分开并独立学习,但仍缺乏针对特定任务的学习。针对上述问题,本文将注意力机制和卷积操作相结合设计了一个注意力解耦头,其结构如图 4 所示。该解耦头由 3 个子头组成,分别用于目标类别预测、置信度回归和定位回归。不同的分支使用池化和卷积的操作组合来进行分类和回归任务。

具体来说,注意力解耦头首先将颈部网络融合后不同尺度的特征图送入对应尺度的预测头进行预测,融合后的特征图通过 1×1 卷积调整特征图通道数,再分别送入分类分支和回归分支。在这两个分支中,使用平均池化操作对特征图降维,然后使用 1×1 卷积操作学习权重,将学习到的权重与原特征图

相乘得到增强后的特征图。增强后的特征图经过两个 3×3 卷积提取高维特征,分类分支提取输入特征图的纹理信息,定位回归分支提取输入特征图的位置信息,置信度回归分支提取输入特征图的背景信息。最后使用 1×1 卷积调整特征图通道数,使用 Concat 操作拼接各个分支的结果。总的来说,注意力解耦头将注意力机制和卷积操作相结合,注意力机制根据不同的输入特征,动态地给予它们不同的权重,从而更加关注重要的特征。而卷积操作能够在局部区域内提取特征,并且在不同位置共享权重。将注意力机制和卷积操作结合在一起,可以让模型更好地捕捉目标对象的特征,增强模型的表达能力和泛化性能,使模型更加适用于各种实际场景。由于分类和回归任务需要不同的特征,因此解耦头使用不同的分支来学习,从而根据任务需求自适应地学习特征。这有助于模型更准确地区分和定位目标,并提高在复杂场景中的目标检测效果。

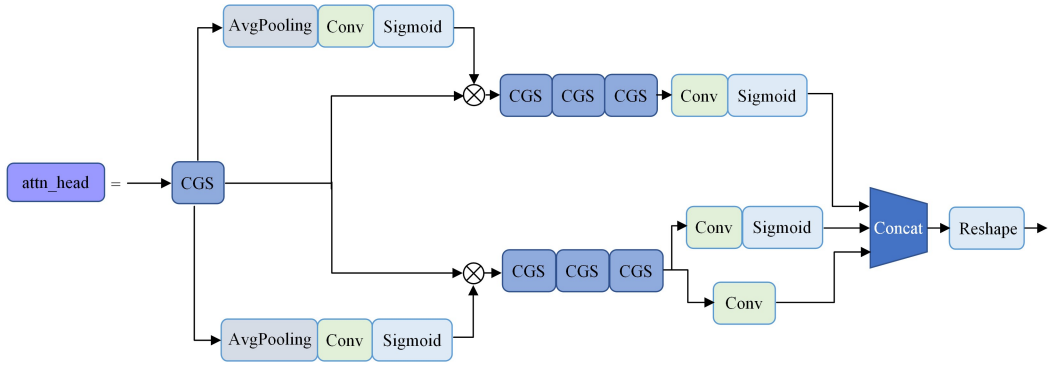


图4 注意力解耦头

Fig. 4 Attention decoupling head

3.2.3 CGS 模块

在雾天交通场景中,雾霾等环境因素会影响捕获到的图像,导致图像像素值分布不均匀且复杂,由于YOLOX_s中使用的CBS模块的性能受到输入数据分布和统计特性的影响,对于分布差异较大的雾天场景图像数据,CBS模块的效果并不理想。为了应对图像像素值分布复杂和不均匀的问题,本文采用Group Normalization^[20](GN)替代CBS模块中的Batch Normalization(BN),从而获得CGS模块。CGS模块包含Conv2d, Group Normalization和SiLU 3个组件。其中Group Normalization是一种数据归一化方法,相对于传统的Batch Normalization,它对数据分布和统计特征的依赖性较小。CGS模块采用GN方法对输入数据进行归一化处理,具体来说,它将每个mini-batch的输入数据按通道C和组数G进行分组,然后计算每组数据的均值和方差并进行标准化。接着,对于每组通道,将输出特征图中所有空间位置的值相加,乘以一个缩放因子,加上一个平移因子进行线性变换,以增强特征的表达能力。最后,将各组数据拼接在一起,得到最终的输出特征图,用于下一层的前向传播计算。GN能够在更细粒度的层面对不同通道的数据进行归一化,相比BN更适合处理像素层面的变化。CBS模块中的归一化是在batch维度上进行的,batch维度的归一化会将来自不同图像的像素进行混合,忽略了不同图像之间的差异性。而CGS模块是在channel维度上进行归一化的,其能够将不同通道的数据进行独立处理,避免了不同图像之间的像素混合,提高了模型的泛化性能。因此,与CBS模块相比,CGS模块在处理复杂像素值分布的雾天场景图像时表现更为优秀,同时,其对小批量数据表现更为稳定,不受样本顺序的影响。

3.3 联合优化

在高质量图像上训练得到的目标检测模型未能考虑到雾天场景中待检测目标与雾之间的相互作用,从而导致雾天场景中模型检测精度下降。为此,本文设计了一种联合优化策略,该策略将去雾网络和检测网络结合起来共同训练。这种方法旨在增强模型在雾天交通场景下的检测能力,从而减少漏检和误检的情况发生。

本文设计的模型联合优化策略包括去雾模块、检测模块和学习用的损失函数,其中去雾模块和检测模块均使用基于深度学习的去雾网络和检测网络。整个模型联合优化策略的输入为有雾或无雾图像以及图像对应的目标检测标签。

有雾图像或无雾图像进入去雾模块,进行去雾过程的学习。去雾模块末端得到去雾后的图像,接着将去雾后的图像以及对应图像的目标检测标签送入检测模块,进行检测过程的学习。检测模块与去雾模块均使用检测损失进行网络的训练。最终检测模块输出有雾图像中目标的分类置信度以及位置的偏移量。本文使用AODNet作为去雾网络,使用上述改进后的目标检测模型作为检测网络,采用联合优化策略来提高雾天场景下目标检测的精度。

AODNet是一种端到端的去雾模型,它将大气散射模型中的两个参数大气光值A和传输函数 $t(x)$ 集成为一个参数 $K(x)$,使用一个由少量卷积层组成的多尺度特征融合网络来估计 $K(x)$,然后将得到的 $K(x)$ 代入式(1)中,得到对应的无雾图像。其中 $K(x)$ 表示对全局大气光值A和透射率 $t(x)$ 的联合估计,如式(2)所示, b 为常数偏置,默认值为1。

$$J(x) = K(x)I(x) - K(x) + b \quad (1)$$

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

AODNet由K estimating模块和图像生成模块两部分组成,K estimating模块由一个小型卷积神经网络组成,其包含5个卷积层,通过特征融合操作将不同卷积层的输出特征图结合起来,K estimating模块从输入图像中自适应的学习参数 $K(x)$;图像生成模块将K estimating模块预测的 $K(x)$ 代入式(1)中得到输入图像对应的无雾图像。AODNet是一个轻量级去雾网络,模型的计算量和参数量都非常小,整体非常简单高效。本文基于雾天场景下目标检测网络的准确性和实时性,选择使用AODNet作为去雾网络与检测网络进行融合。

改进后的目标检测模型在原有模型的基础上减少了模型计算量,增强了网络的特征表示能力,提升了目标检测的精度。在联合优化策略中,有雾图像首先经过输入端进行图像缩放和增强操作扩充数据集;然后被送入去雾模块即AODNet中,学习图像的纹理信息、细节信息以及场景深度信息,并得到参数 $K(x)$,将 $K(x)$ 代入变换后的大气散射模型中,得到去雾后的特征图;最后,将去雾后的特征图送入改进后的检测网络中学习待检测目标的位置信息和语义信息,输出待检测目标的检测框信息和类别信息。去雾模块和检测模块均以目标的分类损失和回归损失学习网络,去雾模块以弱监督的方式学习参数 $K(x)$ 。这种方法可以有效地减少雾气对

图像造成的影响,使网络提取更多有用的特征,以提高雾天目标检测精度。

4 实验

4.1 数据集

在雾天场景下,由于图像质量差、环境复杂,因此很难获取与该场景对应的无雾图像,这给目标检测任务带来了挑战。现有的数据集或只包含用于图像去雾任务的图像,没有目标标注,或只包含在无雾情况下拍摄的带有目标标注的图像。目前专门针对真实雾天场景的目标检测数据集是 RESIDE RTTS 数据集^[21],它包含 4322 张自然有雾图像,其中包含人、自行车、汽车、公共汽车和摩托车 5 个类别的目标。但由于该数据集的规模相对较小,难以满足深度学习模型训练所需要的大量数据,因此影响了模型的训练效果和鲁棒性。

针对上述问题,本文在 PASCAL VOC 目标检测数据集^[22]上使用大气散射模型人工合成一个雾天场景目标检测数据集。为使生成的雾天目标检测数据集更加贴近真实的雾天场景,本文首先对 PASCAL VOC 数据集进行类别筛选,即从 VOC2007 和 VOC2012 的训练集和验证集中筛选出人、汽车、自行车、摩托车和公共汽车这 5 个类别的图像。然后随机选择 0.05~0.15 之间的数赋值给散射系数 β ,随机选择 0~1 之间的实数赋值给大气光值 A 。接着将二者代入大气散射模型中生成不同雾气浓度和不同亮度的图像。最终,本文合成的有雾目标检测数据集共包含 21 579 张图像。在训练阶段,本文从中随机选取 70% 的图像作为训练集,另外 30% 的图像作为验证集。最终,在 RTTS 数据集和合成有雾目标检测数据集上验证所提出的模型。

4.2 实验设置

为公平比较,所有实验的设置均相同,实验使用 pytorch 框架在一块显存为 10GB 的 GeForce RTX3080 进行训练和测试,训练时的批次大小都为 8,共迭代 300 个 epoch,训练时使用随机梯度下降方法,初始学习率为 0.01,动量系数为 0.9,权重衰减系数为 0.0005。其中训练时输入图像大小为 640×640。

4.3 实验结果与分析

4.3.1 消融实验

本文通过在合成的有雾目标检测数据集上进行实验,评估了 CSPCA 模块、注意力解耦头、联合优化策略等改进措施对目标检测模型性能的影响。实验使用 YOLOX_s 作为基础网络,并记录了不同实验组合的结果,如表 1 所列。其中 dehaze 是去雾和检测联合优化后的目标检测模型,train_dehaze 是初始化去雾模块权重后的目标检测模型。

根据表 1 可以看出,在合成有雾目标检测数据集进行训练时,使用 CSPCA 模块可以提高目标检测模型的精度,从 51.88% 提升到 52.78%,同时,模型参数数量和计算量都有所减少。当在模型中同时使用 CSPCA 模块和注意力解耦头时,模型的检测精度进一步提高至 53.06%。而使用 CGS 模块之后,模型的精度又有了一定程度的提升,达到了 53.56%。但在将去雾网络和检测网络进行联合优化时,模型的检测精度有所下降,为 52.92%。而当使用初始化权重的去雾网络

进行联合优化时,模型的精度进一步得到提升,增长至 53.78%。这是因为去雾网络和检测网络联合优化时,去雾网络以半监督的方式进行学习,去雾模块可能会缺乏足够的知识来帮助提高检测网络的性能。然而,当使用训练好的去雾模型对去雾网络权重进行初始化,然后将去雾网络和检测网络联合优化时,去雾网络已经学习到去雾相关特征,因此在联合优化过程中,去雾模块更加注重学习有助于提高检测网络性能的图像特征,从而提高了目标检测模型的精度。总的来说,本文做出的改进可以有效提高雾天交通场景中的目标检测模型的精度。

表 1 消融实验结果

Table 1 Ablation experimental results

方法	mAP@50_95/%	参数量	计算量
YOLOX_s	51.88	8.94×10^6	26.77×10^9
+CSPCA	52.78	8.39×10^6	23.41×10^9
+attnhead	53.06	8.46×10^6	23.99×10^9
+CGS	53.56	8.43×10^6	23.83×10^9
+dehaze	52.92	8.43×10^6	25.28×10^9
+train_dehaze	53.78	8.43×10^6	25.28×10^9

4.3.2 对比实验

为验证 DA_YOLOX 模型的有效性,本文将 DA_YOLOX 与其他主流目标检测模型分别在合成有雾目标检测数据集与 RTTS 数据集上进行比较。为保证实验的公平性,本文使用人工合成的雾天场景目标检测数据集对几个先进的目标检测算法进行训练,每个算法都训练 300 个 epoch。实验结果如表 2 所列。其中 DA_YOLOX(detect)是仅包含检测模块的 DA_YOLOX,DA_YOLOX 是本文提出的目标检测模型,DA_YOLOX(weight)是初始化去雾模块权重的 DA_YOLOX。

表 2 合成有雾目标检测数据集实验结果

Table 2 Experimental results of synthetic foggy target detection dataset

方法	mAP@50_95/%	参数量	计算量
YOLOv3_U	52.10	58.60×10^6	154.60×10^9
YOLOv5_n	41.60	1.60×10^6	4.20×10^9
YOLOv5_s	47.90	6.70×10^6	16.00×10^9
YOLOv5_m	52.70	19.90×10^6	48.30×10^9
YOLOv5_l	55.70	44.10×10^6	108.30×10^9
YOLOX_s	51.88	8.94×10^6	26.77×10^9
DA_YOLOX(detect)	53.56	8.43×10^6	23.83×10^9
DA_YOLOX	52.92	8.43×10^6	25.28×10^9
DA_YOLOX(weight)	53.78	8.43×10^6	25.28×10^9

从表 2 可知,在合成有雾目标检测数据集上,本文提出的 DA_YOLOX 模型与 YOLOX_s 基准模型相比 mAP 有了明显的提高,同时模型大小略微减小。DA_YOLOX 模型在与其他目标检测模型的比较中表现出了性能优势。特别是在模型大小相近的 YOLOv5_s 和 YOLOv5_m 等模型上,DA_YOLOX 的检测效果更好。与 YOLOv3_U 相比,DA_YOLOX 精度更高,模型参数数量和计算量更小。与大型模型 YOLOv5_l 相比,DA_YOLOX 在精度上低于 YOLOv5_l,但 DA_YOLOX 模型的参数数量和计算量却远远小于 YOLOv5_l。

本文使用 RTTS 数据集对训练好的 DA_YOLOX 模型进行了 10 轮微调,并与其他主流目标检测模型进行了比较。结果如表 3 所列,当 IOU 为 0.5 时,DA_YOLOX(detect),DA_

YOLOX 和 DA_YOLOX(weight) 的检测精度分别提高了 0.65%, 1.83% 和 1.99%, 与其他大小类似的目标检测模型相比检测精度更高。这得益于 DA_YOLOX 模型采用 CSP-

CA 模块和注意力解耦头, 从而提升了特征提取能力, 帮助模型获取到了更丰富的特征信息。此外, 联合优化策略将去雾网络和检测网络相结合, 提高了雾天场景下模型的检测精度。

表 3 RTTS 数据集实验结果
Table 3 RTTS dataset experimental results

方法	mAP@50_95/%	mAP@50/%	参数量	计算量
YOLOv3_U	46.90	65.70	58.60×10^6	154.60×10^9
YOLOv5_n	40.80	59.40	1.60×10^6	4.20×10^9
YOLOv5_s	42.90	61.50	6.70×10^6	16.00×10^9
YOLOv5_m	47.10	65.10	19.92×10^6	48.30×10^9
YOLOv5_l	51.30	69.60	44.02×10^6	108.30×10^9
YOLOX_s	47.05	64.49	8.94×10^6	26.77×10^9
DA_YOLOX(detect)	47.57	65.14	8.43×10^6	23.83×10^9
DA_YOLOX	48.79	66.32	8.43×10^6	25.28×10^9
DA_YOLOX(weight)	49.42	66.48	8.43×10^6	25.28×10^9

为了比较本文模型和其他主流目标检测模型的性能, 本文在 RTTS 测试集中随机选择了一张图像, 并展示了它们的检测结果, 如图 5 所示。从图中可以发现 DA_YOLOX(detect) 的检测效果更好且分类置信度更高, 明显优于 YOLOX_s 和其他主流目标检测模型。此外, DA_YOLOX 和 DA_YOLOX(weight) 相较于其他目标检测模型在检测遮挡目标方面表现更加出色, 可以成功检测被遮挡的车辆等。从图 5 的检测结果可以看出, DA_YOLOX(detect), DA_YOLOX 和 DA_YOLOX(weight) 模型在雾天交通场景下表现较好。值得一提的是, DA_YOLOX 和 DA_YOLOX(weight) 模型相较于其他模型在雾天场景下的检测效果更佳, 且能够成功检测被遮挡的目标。

检测数据集。并在 RTTS 数据集和合成有雾目标检测数据集上进行实验, 实验结果表明, 在薄雾和中雾情况下, 该模型的检测结果表现出色。然而, 在浓雾情况下, 该模型的检测性能相对较差。为了进一步提升模型在浓雾情况下的检测性能, 未来的研究方向将集中在领域自适应方面, 以改进模型在复杂雾天环境下的适应能力和泛化能力。

参考文献

- [1] 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.
- [2] 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.
- [3] LIU W, ANGUELOV D, ERHAN D, et al. Ssd: Single shot multibox detector[C]//Computer Vision + ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, Part I 14. Springer International Publishing, 2016:21-37.
- [4] GE Z, LIU S, WANG F, et al. Yolox: Exceeding yolo series in 2021[J]. arXiv:2107.08430, 2021.
- [5] LI B, PENG X, WANG Z, et al. Aod-net: All-in-one dehazing network[C]//Proceedings of the IEEE International Conference on Computer Vision. 2017:4770-4778.
- [6] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6): 1137-1149.
- [7] HUANG S C, LE T H, JAW D W. DSNNet: Joint semantic learning for object detection in inclement weather conditions[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 43(8): 2623-2633.
- [8] REDMON J, FARHADIA. YOLO9000: better, faster, stronger [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017:7263-7271.
- [9] REDMON J, FARHADI A. Yolov3: An incremental improvement[J]. arXiv:1804.02767, 2018.
- [10] HE K, ZHANG X, REN S, et al. Deep residual learning for



图 5 检测结果比较

Fig. 5 Comparison of detection results

结束语 本文针对雾天场景中目标检测模型性能下降的问题, 提出了一个雾天场景目标检测模型 DA_YOLOX。该模型包括去雾模块和检测模块, 去雾模块使用 AODNet 对输入图像进行去雾, 检测模块使用改进后的 YOLOX_s 对去雾后的特征图进行特征提取。通过共同训练优化, 增加去雾模块与检测模块的相关性, 使模型在雾天场景仍具有良好的检测效果。本文为解决雾天场景目标检测数据集较少的问题, 使用大气散射模型和 VOC 数据集合成了一个雾天场景目标

- image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016:770-778.
- [11] BOCHKOVSKIY A,WANG C Y,LIAO H Y M. Yolov4: Optimal speed and accuracy of object detection[J]. arXiv: 2004. 10934,2020.
- [12] WANG C Y,LIAO H Y M,WU Y H, et al. CSPNet: A new backbone that can enhance learning capability of CNN[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2020:390-391.
- [13] LI C,LI L,JIANG H, et al. YOLOv6: A single-stage object detection framework for industrial applications[J]. arXiv: 2209. 02976,2022.
- [14] 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]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023:7464-7475.
- [15] XIE Y H,XIE Y,CHEN Y. Object Detection in Real Misty Scenes [J]. Journal of Computer Aided Design and Graphics, 2021,33(5):733-745.
- [16] HNEWA M,RADHA H. Multiscale domain adaptive yolo for cross-domain object detection[C]// 2021 IEEE International Conference on Image Processing(ICIP). IEEE,2021:3323-3327.
- [17] LIU W,REN G, YU R, et al. Image-adaptive YOLO for object detection in adverse weather conditions[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2022:1792-1800.
- [18] LI X X,QIANG J,LIU W J, et al. Research on Traffic Object Detection Method in Fog Based on Dual Backbone Network[J]. Journal of Chongqing Technology and Business University (Natural Science Edition),2023,40(4):25-34.
- [19] XU S,WANG X, LV W, et al. PP-YOLOE: An evolved version of YOLO[J]. arXiv:2203. 16250,2022.
- [20] WU Y,HE K. Group normalization [C] // Proceedings of the European Conference on Computer Vision(ECCV). 2018:3-19.
- [21] LI B,REN W,FU D, et al. Benchmarking single-image dehazing and beyond[J]. IEEE Transactions on Image Processing, 2018, 28(1):492-505.
- [22] EVERINGHAM M,VAN GOOL L,WILLIAMS C K I, et al. The pascal visual object classes(voc)challenge[J]. International Journal of Computer Vision,2010,88:303-338.



LOU Zhengzheng, born in 1984, Ph.D, associate professor, is a member of CCF (No. 42111M). His main research interests include data mining, IB methods, intelligent traffic signal control and so on.



WU Yunpeng, born in 1987, Ph.D, associate professor, is a member of CCF (No. 42109M). His main research interests include pattern recognition, computer vision, computer graphics and so on.

(责任编辑:何杨)