

文章编号:1007-2780(2023)12-1717-11

基于 Double-Head 的雾天图像目标检测

李任斯, 石蕴玉*, 刘翔, 汤显, 赵静文

(上海工程技术大学 电子电气工程学院, 上海 201620)

摘要:雾天环境下的图像对比度低,图像中的目标较为模糊并且其特征提取存在一定难度。现有的目标检测方法对于雾天图像的检测准确率偏低。针对上述问题,本文在 Double-Head 框架上基于图像的特征提取部分和预测头部进行改进。首先,在提取的深层特征图上添加通道和空间双维度的复合注意力机制,提高网络关注显著目标的能力;其次,将原始图像经过改进的暗通道先验以及处理后得到的先验矩阵和特征图进一步融合,获取更全面的雾天图像特征信息;最后,在预测头部引入可分离卷积,使用解耦预测头对目标进行最终的分类和回归。该方法在 RTTS 数据集上的 mAP 为 49.37%,在合成数据集 S-KITTI 和 S-COCOval 数据集上的 AP 值分别为 66.7% 和 57.7%。与其他主流算法相比,本文算法具有更高的目标检测精度。

关键词:目标检测;雾天图像;暗通道先验;注意力机制;特征融合

中图分类号:TP391.4 **文献标识码:**A **doi:**10.37188/CJLCD.2023-0089

Object detection in foggy image based on Double-Head

LI Ren-si, SHI Yun-yu*, LIU Xiang, TANG Xian, ZHAO Jing-wen

(Department of Electric and Electronic Engineering, Shanghai University of Engineering Science,
Shanghai 201620, China)

Abstract: Image contrast in the foggy environment is low, and the object is fuzzy so that it is difficult to extract features in images. The existing object detection methods has a low accuracy for detecting objects in foggy images, and the objects is fuzzy and is difficult to extract features. To solve these problems, the feature extraction and prediction head are improved on the Double-Head framework. Firstly, multi-scale salient and effective features of objects in the image are carried out by adding channel attention to the feature maps extracted from the backbone network. Secondly, the prior matrix and feature maps from the original image processing by dark channel prior method with image processing are fused to get more comprehensive feature information in foggy images. Finally, the separable convolution is introduced into the prediction head and the effective decoupled head is used to complete the classification and regression tasks. The proposed method has the mAP of 49.37% on the RTTS dataset, and the AP of 66.7% and 57.7% on the S-KITTI and S-COCOval dataset. Compared with other mainstream algorithms, this

收稿日期:2023-03-07;修订日期:2023-03-16.

基金项目:中国高校产学研创新基金(No. 2021FNB02001);上海市自然科学基金(No. 19ZR1421500)

Supported by China University Industry Research Innovation Fund(No. 2021FNB02001); Natural Science Foundation of Shanghai(No. 19ZR1421500)

*通信联系人, E-mail: yunyushi@sues.edu.cn

algorithm has higher object detection accuracy.

Key words: object detection; foggy image; dark channel prior; attention mechanism; feature fusion

1 引言

目标检测是计算机视觉中最基本、最重要的任务之一。搭配强大的主干网络,许多基于深度学习的目标检测算法在日常场景中都实现了较高的检测精度,YOLO(You Only Look Once)^[1]、SSD(Single Shot Detector)^[2]和区域卷积神经网络(Region Convolutional Neural Network, R-CNN)^[3-5]均是目标检测的经典算法。为了加强对不同特征尺度的利用, Lin 等提出的 FPN(Feature Pyramid Network)^[6]以及 Liu 等人提出的 PANet^[7]均利用特征层的自下而上和自上而下的路径级联对不同特征图提取到的信息进行学习。Libra R-CNN^[8]通过对低层次特征和高层次特征的融合增强和正负样本的平衡处理使得网络具有更高的目标检测精度。Li 等人^[9]通过特征融合以及混合注意对 SSD 算法进行改进,提升检测精度。然而在真实的雾天场景中,拍摄的图像由于各种因素其图像质量显著下降,模糊、光照不足、物体遮挡会造成图像颜色失真、细节信息缺失等问题。普通的目标检测方法在雾天环境下无法准确定位和识别出目标,其检测性能大幅降低。因此,雾天环境下的目标检测方法研究显得意义重大。

Li 等人^[10]提出了去雾算法 AODNet,并将其与检测网络 Faster R-CNN 结合进行训练,对去雾后的图像进行检测。但是这种图像预处理操作改变了图像的内部结构,且其模型的推理速度和训练速度非常慢,性能也没有得到提高。Wang 等提出了 FPVDNet^[11],在网络中加入雾浓度判别模块和注意力机制。Xie 等提出了基于知识引导的方法 KODNet^[12],通过对数据预处理和目标分布问题分析实现有雾场景下的多目标检测。但是对于雾天图像特征提取过程,上述方法存在语义信息和细节缺失问题,特别是在浓雾图像中,对于一些模糊的目标,在网络中的特征映射关系较弱,深层特征图的目标峰值很低。有关端到端的雾天图像检测算法研究较少。在图像去雾算法暗通道先验法^[13]中,作者发现统计的无雾彩色图像的非天空区域里,RGB 3 个颜色通道中

总有一个通道的像素值很低,趋向于 0;而对于有雾图像,其雾景区域的暗通道块的像素值都远大于 0,具体的识别方法见第二章理论基础。受此启发,如果能借助雾天图像自身特点,获得更加全面有效的图像信息,提高检测网络对雾区域中目标物体的信息处理能力,则对提升雾天图像检测性能具有重大作用。

本文在 Double-Head 目标检测模型框架的基础上,在特征增强方面,既融合了原始图像经过暗通道先验及处理后得到的先验权重矩阵,又加入了特征图自身的通道自注意力机制以及结合改进的解耦合预测头(Efficient Decoupled Detection Head, EDH),提出了一种新的端到端的目标检测方法来提高雾天场景中的目标检测性能。

2 理论基础

2.1 Double-Head 算法

Double-Head 算法^[14]是由 Wu 等人提出的新的二阶段检测算法。作者在网络预测头部对全连接头、全卷积头和目标的分类回归任务进行了彻底的分析,发现全连接头比全卷积头具有更高的空间敏感性。全连接头具有更强的区分完整目标和部分目标的能力,但对回归整个目标并不鲁棒。基于此,研究者提出了 Double-Head 算法,设计了专注于分类的全连接头和目标框回归的全卷积头。在 Faster R-CNN 中主要采用共享的全连接层进行最后的目标回归和分类,而在 Double-Head 中采用全连接层进行类别分类,采用全卷积层进行目标的回归。在分类预测任务中,完成分类任务的子模块由 2 个 1 024 层的全连接层构成,而回归子模块的全卷积层共包含 5 个残差卷积块和最后的池化输出 1 024 维的向量。一个普通的残差卷积块由输入依次经过一个 1×1 卷积, 3×3 卷积和 1×1 卷积后,再加上原始的输入构成,每个残差卷积块最后输出的激活函数为 ReLU。Double-Head 算法在目标检测任务中具有较高的准确率,因此本文选择其作为基础框架。

为了更加有效地利用不同尺度的图像特征,

PANet在FPN基础上利用特征层自上而下的信息,再次级联了自下而上的路径,进一步加强了网络对各个尺度的目标的提取能力。Double-Head算法同样借鉴PANet思想,级联不同尺度的特征层。本文通过添加通道和空间双维度的注意力模块对特征提取方面进行改进,并且在预测头部引入可分离卷积,利用设计的卷积模块改进原有解耦合预测头的性能。

2.2 暗通道先验法

在图像处理领域,有许多研究雾环境的光学模型。其中一个经典模型大气散射模型^[15]的方程定义如式(1)所示:

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

式中: $I(x)$ 是观测图像,与自然雾图像相同; $J(x)$ 是场景亮度; $t(x)$ 代表介质透射; A 代表全球大气光。 $t(x)$ 定义为:

$$t(x) = e^{-\beta d(x)}, \quad (2)$$

式中: β 表示大气散射系数, $d(x)$ 表示场景深度。He等将暗通道先验与大气散射模型结合提出了暗通道先验去雾算法。作者统计了大量的无雾图像,发现在每一幅图像的非天空区域里,彩色图像RGB通道中至少有一个通道的像素值非常低,几乎趋近于0。暗通道块 J 的数学表达式如式(3)所示:

$$J^{\text{dark}}(x) = \min_{c \in \{r, g, b\}} (\min_{y \in \Omega(x)} (J^c(y))), \quad (3)$$

式中: J 表示彩色图像暗通道块 J 在像素点 x 中的

一个颜色通道,即RGB通道中的一个; y 是以像素点 x 为中心的矩形区域。

根据公式(3)可以得出 $J^{\text{dark}}=0$,即非天空区域的图像块经过暗通道先验后像素值为0,暗通道先验处理后的图像从彩色图像变为灰度图像。雾天图像存在大量雾景区域,雾景区域的RGB 3个通道的像素值都不低,因此其暗通道块 J^{dark} 远远大于0。

基于暗通道先验法处理后的雾天图像可以有效地将雾景区域和前景所分离,给目标检测任务提供了有效的信息。本文通过改进暗通道先验算法提升了检测网络获取雾天图像有效信息的能力。

3 本文方法

图1是本文所提方法的结构图,基于二阶段方法 Double-Head 进行改进。在特征增强方面,网络利用原始输入图像经过暗通道先验得到的有效信息和主干网络提取的多尺度深层特征进行有效融合,得到增强的特征图。增强后的特征图经过区域建议网络^[3]和ROI池化模块^[3]处理后输送至预测头部进行最终的预测。同时,为了使分类和回归都达到更高的精度,使用解耦合检测方法作为预测头部。预测头使用全连接层进行分类,卷积模块进行回归,并在卷积模块中引入了深度可分离卷积,减少了网络模型的参数。

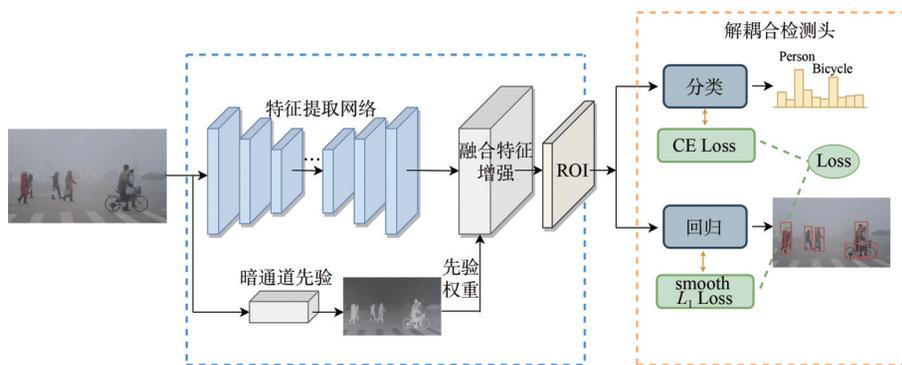


图1 提出网络的概述图

Fig. 1 Overview of the proposed network

3.1 融合暗通道先验权重的特征增强

在本文所提方法中,特征增强模块具体结构如图2所示。首先,在主干网络ResNet101^[16]和

FPN提取到的特征 $P_1 \sim P_4$ 的基础上,进一步级联了自下而上的结构,得到特征层 $N_1 \sim N_4$ 。其次,为了提高特征层 $N_1 \sim N_4$ 每个尺度的特征描述力,

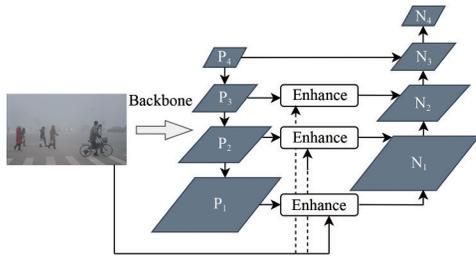


图 2 融合特征增强结构示意图

Fig. 2 Structure of the fusion feature enhancement

在 P_1 、 P_2 、 P_3 的跳跃连接中加入了 Enhance 模块。Enhance 模块主要包含两部分,如图 3 所示。第一部分为特征自注意力,即在特征层 $P_1 \sim P_4$ 输入的基础上,加上了通道池化操作,增加了通道值响应;第二部分为先验权重矩阵的融合,即将原始图像经过暗通道先验法得到的有效信息进一步融合,增强了雾天图像中前景目标和背景的分辨。

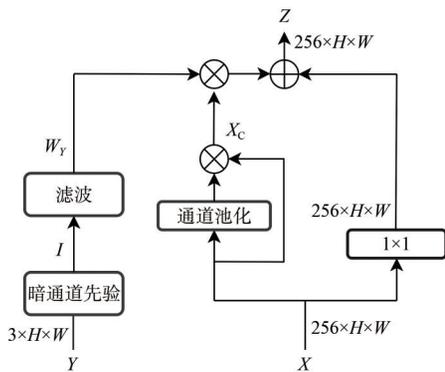


图 3 Enhance 模块的具体结构

Fig. 3 Structure of the Enhance module

在图 2 中的特征图部分,向下箭头是上采样操作,向上箭头是下采样操作,上、下采样操作分别使用双线性插值法和自适应池化法。 N_1 将下采样生成的与 N_2 具备同样尺度的特征和经过 Enhance 模块增强后的特征图在通道方向拼接起来得到 N_2 , 依此类推,获得 N_3 和 N_4 。

图 3 中的通道池化分为最大池化和平均池化,都在输入特征 $X(256 \times H \times W)$ 的通道维度上进行。最大池化在 X 的 256 维的通道上选择最大值作为表征,平均池化则是选取通道上平均为 $1 \times H \times W$ 的值作为表征。相加后系数经过 Sigmoid 激活函数与原始输入进行点积运算得到 X_c 。共享的全连接层来计算权重,记为 k_i ,在通道维度上拼接后使用 Sigmoid 函数来生成通道信息输出矩

阵 F_{oc} , 数学表达式如式(4)所示:

$$X_c = X \cdot \varphi(F_{max} + F_{avg}), \quad (4)$$

其中, $\varphi(\cdot)$ 表示 Sigmoid 激活运算。

第二部分先验权重的融合如图 3 中左边分支所示。首先对输入图像进行尺寸缩放得到 Y , 使其与图 3 中特征 X 的输入尺寸一样。然后 Y 通过暗通道先验法即利用公式(3)运算得到暗通道矩阵 I , 式(3)中 c 为 Y 的 R、G、B 3 个通道, Ω 取 5×5 大小的区域, 即

$$I = J^{dark}(Y). \quad (5)$$

接着对暗通道矩阵 I 中的所有像素值(假设其大小为 $n(0 \sim 255)$)进行取反操作, 变为 $255 - n$, 使矩阵中天空等背景区域具有较低的矩阵值, 目标等前景具备较高的矩阵表征值。此时暗通道矩阵还存在一些冗余表征和背景噪声, 需要对其进行滤波操作以增强待检测目标的矩阵响应, 同时弱化非目标等背景区域的矩阵响应, 即矩阵值的大小。滤波函数采用一阶微分算子 sobel 算子, 能有效降低图像噪声, 锐化特征, 其公式如式(6)、式(7)所示:

$$I_x = \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix} I, \quad I_y = \begin{pmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{pmatrix} I, \quad (6)$$

$$W_Y = \sqrt{I_x^2 + I_y^2}, \quad (7)$$

其中: I 为暗通道矩阵, 为灰度图像; W_Y 为输出的权重矩阵; I 和 W_Y 大小均为 $1 \times H \times W$ 。

在图 3 中, W_Y 和 X_c 进行点积运算后, 再次和经过 1×1 卷积后的输入特征 X 相加, 得到最终的输出 $Z(256 \times H \times W)$ 。

雾天图像中的目标对比度更低, 纹理特征不清晰, 普通的检测算法提取其中目标的能力较低。为了验证本文 Enhance 模块和所提方法的有效性, 如图 4 所示, 将原始的 Double-Head 方法以

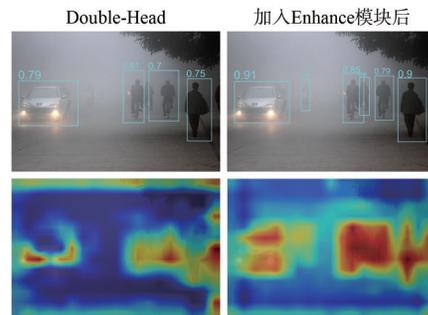


图 4 热力图对比

Fig. 4 Comparison of feature heat activation map

及加入 Enhance 模块后的特征热力图进行了可视化。特征热力图可以直观地展示网络在输入图像上的感受野和特征提取能力。从图 4 可以看出,右侧图像中汽车和行人位置的热力图颜色更红,具有更大的激活值,而背景的激活值则更小,并且像素点对应目标类别的置信度得分更高,检测结果更加准确。高质量的雾天图像特征图的提取会进一步使预测头最终对目标完成更高质量的分类和回归任务,提升网络的雾天图像目标检测性能。

3.2 高效解耦预测头 EDH

经过 RoIAlign 得到的 $7 \times 7 \times 256$ 大小的特征图最终由解耦预测头进行分类和回归,具体结构如图 5 所示。在分类任务分支中,应用两个全连接层进行分类。对于回归任务,全卷积层使用了 5 个残差模块。第一个残差模块 a 将输入从 256 维升至 1 024 维,具体内部结构对应图 6(a),另 4 个残差模块 b 如图 6(b) 所示。图 6(b) 中 5×5 表示卷积核大小为 5×5 的深度可分离卷积块,在减少网络参数数量的同时,增加了深层特征的感受野,保证了最终目标框位置回归的精度,其余为普通卷积。深度可分离卷积是轻量化网络 Xception^[17] 的主要结构,主要思想是将标准卷积

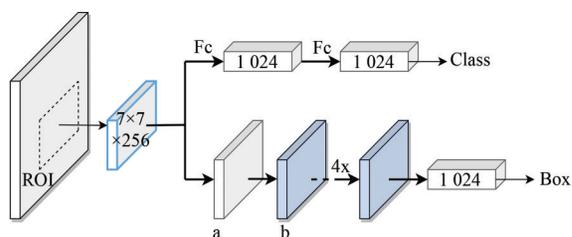


图 5 解耦预测头结构

Fig. 5 Structure of the decoupled prediction head

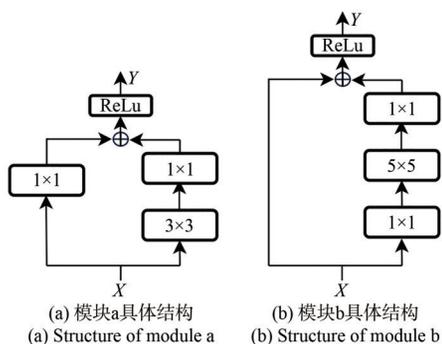


图 6 卷积模块结构图

Fig. 6 Structure of the convolution module

分解为深度卷积和逐点卷积,即在输入在每个通道独立执行空间卷积,再进行逐点卷积,将深度卷积的通道输出映射到新的通道空间。EDH 模块的最终输出是一个 $k+1$ 维向量用于分类, $k \times 4$ 维的向量用于定位,其中 k 表示类别个数。

3.3 损失函数

在区域建议网络 RPN 中,分类任务使用二值交叉熵损失。在预测头 EDH 中,分类任务使用的是交叉熵损失。对于样本训练过程中的回归任务,均使用 Smooth L_1 损失。Smooth L_1 损失函数公式如式(8)所示:

$$\text{Smooth } L_1(x, \beta) = \begin{cases} 0.5|x|^2/\beta, & \text{if } |x| < \beta, \\ |x| - 0.5\beta, & \text{otherwise,} \end{cases} \quad (8)$$

式中:对于 RPN, β 设置为 $1/9$;在 EDH 中, β 设置为 1。所以

$$L_{\text{RPNreg}} = \text{Smooth } L_1(x, 1/9), \quad (9)$$

$$L_{\text{EDHreg}} = \text{Smooth } L_1(x, 1). \quad (10)$$

RPN 的二值交叉熵损失表示为:

$$L_{\text{RPNcls}} = \frac{1}{N} \sum_i - [y_i \cdot \log p_i + (1 - y_i) \cdot \log(1 - p_i)]. \quad (11)$$

在一个训练批次中,选取部分特征图作为训练样本,其中 N 是二元分类中的样本数, y_i 是样本 i 的标签,正样本为 1,负为 0, p_i 是预测样本为正的的概率。RPN 总损失函数如式(12)所示:

$$L_{\text{RPN}} = L_{\text{RPNreg}} + L_{\text{RPNcls}}. \quad (12)$$

预测头 EDH 的交叉熵损失如式(13)所示:

$$L_{\text{EDHcls}} = -\frac{1}{N} \sum_i \sum_{c=1}^k y_{ic} \log p_{ic}, \quad (13)$$

其中: N 是多类别样本的数量; k 是类别的数量; y_{ic} 是一个符号函数,如果样本的预测类别是 c ,则取 1,否则取 0; P_{ic} 是样本 i 被预测为 c 类别的概率。整个网络的损失函数 L 如式(14)所示:

$$L = L_{\text{RPN}} + \lambda_1 L_{\text{EDHreg}} + \lambda_2 L_{\text{EDHcls}}, \quad (14)$$

式中: L_{RPN} 是区域建议网络 RPN 的损失函数; λ_1 和 λ_2 是预测头中回归损失和分类损失的权重参数,它们平衡了各部分的功能,实现了更好的检测性能。

4 实验结果与分析

4.1 实验细节

实验平台硬件为 AMD Ryzen Threadripper

1900X 8-Core Processor 3.80 GHz CPU, Nvidia GTX 2080 Ti显卡, 64G内存。

实验中,网络由随机梯度下降优化器进行训练,动量和权重衰减分别设置为0.9和0.0001。大多数二阶段检测器,包括本文提出的网络,学习率初始化为0.0025。式(14)中的 λ_1 和 λ_2 分别设置为2.0和1.5。

4.2 数据集和评估指标

本节比较了一些现有的主流目标检测方法以验证本文所提出的网络的鲁棒性和高精度。不仅在自然雾天数据集RTTS^[18]上进行了评估,还在合成雾天数据集S-KITTI和S-COCOval上进行了评估。

S-KITTI和S-COCOval为本文利用大气散射模型分别对KITTI^[19]和MS COCO数据集^[20]进行雾合成得到的合成数据集。在KITTI和MS COCO的子数据集COCO 2014 val中分别选用7481和10k余张图像进行雾合成模拟处理,应用大气散射模型和原始清晰图像合成雾天图像,以此得到式(1)中的 $I(x)$ 。在合成处理中,式(1)和式(2)中的 A 和 β 设置为0.6和0.28,以此模拟出逼近真实雾天的图像。

RTTS数据集属于自然雾天环境的目标检测数据集,图像主要来源于交通道路等场景。它包含4322副图像;带有5个类别的标签,即汽车、人、公共汽车、自行车和摩托车,包括大约41000个目标。对于KITTI数据集,使用其2D数据集进行目标检测任务,包含7481张图像,包括6个类别:汽车、面包车、行人、有轨电车、卡车和自行车。MS COCO数据集有80个类别,为了使实验检测对象一致,本文在COCO 2014 val中选择了

1万张图像,包括摩托车、汽车、公共汽车、火车、自行车和卡车等类别。

在RTTS数据集中,采用PASCAL VOC^[21]指标平均精度(mean Average Precision, mAP)来评估该方法的检测性能。准确率和召回率的计算公式如式(15)、式(16)所示:

$$P = \frac{TP}{TP + FP}, \quad (15)$$

$$R = \frac{TP}{TP + FN}, \quad (16)$$

其中:TP表示正样本被识别正确的个数,FP表示负样本被识别为正样本的个数,FN表示正样本被识别为负样本个数。将计算得到的 P 、 R 值绘制 P - R 曲线,计算曲线下的面积,即为AP值,公式如式(17)所示:

$$AP = \int_0^1 PdR. \quad (17)$$

对于多个类别 N ,平均精度mAP公式为:

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i. \quad (18)$$

在S-KITTI数据集和S-COCOval数据集中,使用指标AP、AP₇₅、AP₅₀、AP_s、AP_M和AP_L来评估目标检测的性能。其中AP指标的IoU从0.5依次增加到0.95。AP₇₅和AP₅₀是指其IoU分别设置为0.75和0.5。AP_s、AP_M和AP_L是根据图像中对象的大小设置的AP度量,分别对应图像中相对的小目标、中型目标和大目标。

4.3 RTTS实验结果

RTTS数据集的标签分布不均,在某些特定角度上的不同类别相似。本文方法在一定程度上克服了这些困难,并取得了较高的性能。如表1所示,每一行记录了不同方法的mAP值和5个类

表1 不同算法在RTTS数据集上的性能对比

Tab. 1 Performance comparison of different algorithms in the RTTS dataset

方法	骨干网络	汽车	公交车	自行车	摩托车	行人	mAP
Yolov3	DarkNet53	61.58	23.20	23.02	43.84	30.32	36.39
ATSS ^[23]	ResNet101	60.95	35.69	10.37	46.33	30.32	37.68
Dynamic R-CNN ^[24]	ResNet101	61.93	41.32	32.20	37.38	46.08	43.78
Double-Head	ResNet101	63.00	51.53	24.29	41.70	43.22	44.74
Cascade R-CNN	ResNet101	63.33	43.37	26.43	38.03	43.16	42.86
FPVDNet	ResNet101	65.47	44.10	26.50	44.31	46.61	45.40
KODNet	ResNet101	64.81	47.22	28.65	47.35	45.25	46.66
本文方法	ResNet101	66.41	50.78	33.62	50.12	46.92	49.37

别的 AP 值,其中指标中 IoU 的值设定为 0.5。

在不增加额外数据的情况下,本文方法实现了最高的 mAP 值 49.37,比 FPVDNet 高 3.97,比 KODNet 高 2.71,比 YOLOv3^[22] 高 12.98。Double-Head 方法在公交车类别中达到了最高的 AP 值 51.53,本文方法在其余类别中达到了最高的性能,汽车类别为 66.41 AP 值,自行车类别为 33.62 AP 值,摩托车、行人分别达到了 50.12 和 46.92 AP 值。本文方法在通过数据增强操作后,mAP 值可以提升至 50.33,均高于其他算法。数据增强操作主要包括对训练图像进行翻转、旋转和缩放等。通过数据增强生成的额外图像大约 1 000 张,这些图像与原始数据一起训练。除摩托车类别外,其他所有类别都处于较高的水平。结果表明,与其他方法相比,本文方法的检测准确率更高,说明本文方法能有效地提取雾天图像中的特征。

图 7(a)是从 RTTS 数据集中选择的一幅自然雾天图像,(b)、(c)、(d)是分别使用 Double-Head、Cascade R-CNN^[25] 和本文方法进行检测后的具体结果。在图 7(b)中,Double-Head 方法错误地将图片右下角的三轮车识别为汽车,并忽略了图片左侧的行人。图 7(c)的 Cascade R-CNN 方法很难检测到汽车,也忽略了左边的行人。图 7(d)为本文方法的检测结果,结果表明本文方法可以

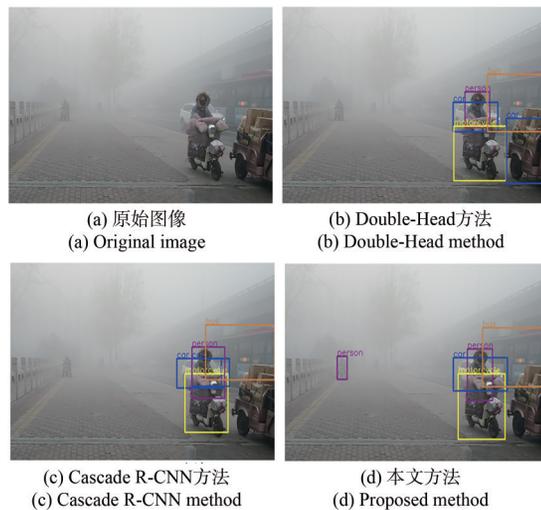


图 7 RTTS 数据集图像的检测结果示例图

Fig. 7 Examples of RTTS image detection results

克服这些问题,生成更准确的检测结果。

4.4 S-KITTI 和 S-COCOval 实验结果

S-KITTI 数据集中出现的小目标物体和遮挡目标较多,使得目标检测任务变得更具挑战性,尤其是在雾天环境中。S-KITTI 数据集实验结果如表 2 所示,除 AP_M 指标之外,所提出方法性能均优于其他方法。本文方法的 AP 、 AP_{75} 、 AP_{50} 、 AP_S 、 AP_M 和 AP_L 指标分别为 66.7、77.1、90.9、65.6、66.2 和 71.4, AP_{75} 和 AP 分别比第二高的方法高出 1.7 和 0.8。

表 2 不同算法在 S-KITTI 数据集上性能对比

Tab. 2 Performance comparison of different algorithms in the S-KITTI dataset

方法	骨干网络	AP	AP_{75}	AP_{50}	AP_S	AP_M	AP_L
FSAP ^[26]	ResNet101	62.8	69.6	89.9	63.7	63.0	66.8
ATSS	ResNet101	64.8	74.4	89.8	61.5	64.6	70.2
PANet	ResNet101	64.4	72.9	90.1	64.6	64.2	69.0
Libra R-CNN	ResNet101	64.5	74.3	89.7	63.8	63.7	69.5
Dynamic R-CNN	ResNet101	65.9	75.4	90.2	62.9	66.5	70.8
Double Head	ResNet101	65.4	74.5	90.7	65.4	64.6	69.8
本文方法	ResNet101	66.7	77.1	90.9	65.6	66.2	71.4

如表 3 所示,当主干网络使用 ResNet101 时,本文方法的 AP 、 AP_{75} 、 AP_{50} 、 AP_S 、 AP_M 和 AP_L 值分别为 57.7、65.1、79.6、41.1、51.1 和 69.3。 AP_{75} 、 AP_S 和 AP 分别比第二高的方法高 1.5、1.7 和 0.9,除了 AP_{50} 之外其余数据均为最高。Double-Head 在 AP_{50} 取得了最大值 79.8,仅比本文所提

方法高出 0.2。结果表明,本文模型在合成雾天数据集 S-COCOval 上也取得了良好的检测结果,并且应用不同的主干网络仍然具有有效性和鲁棒性。

4.5 消融实验和定量比较

表 4 对预测头的卷积部分进行了实验,通过

表 3 不同算法在 S-COCOval 数据集上性能对比

Tab. 3 Performance comparison of different algorithms in the S-COCOval dataset

方法	骨干网络	AP	AP ₇₅	AP ₅₀	AP _s	AP _M	AP _L
FSAF	ResNet101	52.5	58.4	77.6	37.5	47.1	64.3
ATSS	ResNet101	54.4	61.2	78.0	38.7	48.8	66.4
PANet	ResNet101	53.6	60.7	78.2	35.6	47.8	64.8
Libra R-CNN	ResNet101	54.1	61.4	77.8	39.2	48.1	65.3
Dynamic R-CNN	ResNet101	56.8	63.6	78.7	39.4	50.4	69.0
Double Head	ResNet101	56.2	63.2	79.8	38.7	50.5	67.2
本文方法	ResNet101	57.7	65.1	79.6	41.1	51.1	69.3

表 4 增加不同数量模块 b 后的参数和 AP 值比较

Tab. 4 Comparison of the parameters and AP values after adding different number of modules b

模块 a	模块 b	参数/M	AP
√	0	1.06	37.1
√	1	1.87	37.9(+0.8)
√	2	2.68	38.6(+1.5)
√	3	3.49	39.1(+2.0)
√	4	4.30	39.6(+2.5)
√	5	5.11	39.6(+2.5)
√	6	5.92	39.7(+2.6)

设定不同数量的卷积块得到不同的 AP 结果。表 4 中的模块 a 和模块 b 代表图 6 中的提取模块 a 和 b, 模块 a 主要是将特征图从 256 维升至 1 024 维, 数量为一个, 大小为 1.06M; 模块 b 大小为 0.81M。考虑到准确率和参数量的平衡, 本文的预测头部卷积部分最后选择一个模块 a 和 4 个模块 b。

在 RTTS 数据集上分别使用不同注意力模块进行实验, 结果如表 5 所示。在不添加注意力模块的情况下, 使用 Double-Head 方法作为基线的 mAP 值为 51.5, 分别添加注意力模块 SE(Squeeze-and-Excitation)^[27]、CA(Coordinate Attention)^[28] 和 CBAM(Convolutional Block Attention module)^[29]

表 5 使用不同注意力模块的实验结果

Tab. 5 Ablation study of proposed component

方法	注意力模块	mAP	参数量/M
	/	51.5	—
	SE	51.6	+0.41
Double-Head	CA	52.2	+0.52
	CBAM	52.7	+0.51
	本文	53.6	+0.45

后, mAP 值为 51.6、52.2 和 52.7。使用本文的注意力模块达到了最高的 mAP 值为 53.6, 参数量仅增加 0.45M。结果表明, 与其他注意力模块相比, 本文设计的注意力模块具有更高的目标检测性能。

表 6 展示了提出的特征增强模块和解耦合预测头 EDH 消融实验的结果。在 ResNet50-FPN 基线上, 对特征图应用本文的注意力模块 mAP 值提高了 0.8, 融合改进暗通道先验得到的先验权重 mAP 值提高了 2.4, 证明所提出的特征增强模块对检测雾天图像中的目标的有效性。解耦合预测头 EDH 模块将 mAP 值从 52.9 提高到 54.3, 比 ResNet50-FPN 基线高出 1.4。与基线相比, 提出的模块对网络的改进 mAP 值共提升了 3.6。在模型复杂度方面, 改进前 Double-Head 方法模型参数为 48.14 MB, 推理时间为 0.105 s; 改进后方法模型参数为 52.25 MB, 推理时间为 0.126 s。

表 6 各个模块的消融实验结果

Tab. 6 Ablation study of proposed component

主干网络	注意力	先验权重	EDH	mAP	Δ
ResNet50				52.9	—
ResNet50	√			53.7	+0.8
ResNet50		√		55.3	+2.4
ResNet50			√	54.3	+1.4
ResNet50	√	√	√	56.5	+3.6

4.6 雾天场景下目标检测结果和分析

从数据集中选取了几组代表性的图像分别使用不同的算法进行目标检测结果展示, 如图 8 所示, 从上至下依次为原始图像、本文算法、Double-Head 算法、Cascade R-CNN 算法、FPVD-

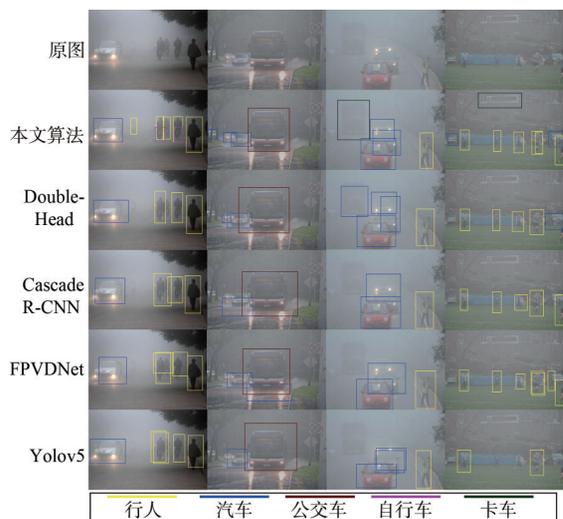


图 8 不同算法的检测结果

Fig. 8 Detection results of different algorithms in different scenarios

Net 和 YOLOv5 算法检测的结果。第一组图像中,除了本文算法外,其余算法都出现了自行车漏检的情况,且行人检测不够准确。在其他组图像中,Double-Head、Cascade R-CNN、FPVDNet 和 YOLOv5 均存在对小目标物体漏检的问题,Double-Head 算法在第三组图像中更是将卡车错误识别为汽车。本文算法在这些雾天图像中展示出了更精准的检测结果,具有更强的检测性能和鲁棒性。

本文所提方法在合成数据集不同雾浓度下的检测结果如表 7 所示。大气散射系数 β 取不同

表 7 所提方法在不同雾天程度下的实验结果

Tab.7 Experimental results of the proposed method under different foggy conditions

雾浓度	β	AP	AP ₇₅	AP ₅₀
轻度	0.1~0.2	64.2	74.1	88.2
中度	0.2~0.3	62.7	72.5	87.6
重度	0.3~0.5	59.4	69.6	85.9

参 考 文 献:

[1] REDMON J, DIVVALA S, GIRSHICK R, *et al.* You only look once: Unified, real-time object detection [C]// *Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas: IEEE, 2016: 779-788.

[2] LIU W, ANGUELOV D, ERHAN D, *et al.* SSD: single shot MultiBox detector [C]// *Proceedings of the 14th European Conference on Computer Vision*. Amsterdam: Springer, 2016: 21-37.

[3] REN S Q, HE K M, GIRSHICK R B, *et al.* Faster R-CNN: towards real-time object detection with region proposal

值分别对应不同程度的雾。在实际雾天场景下,所提方法的目标检测结果如图 9 所示。场景中的小目标车辆以及遮挡目标仍可以准确检测出来,并进行有效分类,进一步证明本文方法在实际场景中的高目标检测性能。

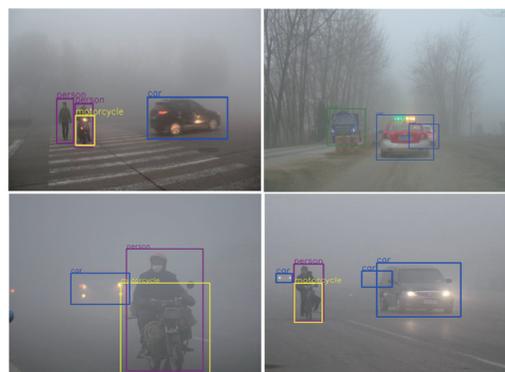


图 9 实际雾天场景检测结果图

Fig. 9 Detection results in the actual foggy scene

5 结 论

本文针对雾天场景下的目标检测任务,提出了一种新的高效且鲁棒的方法。在 Double-Head 网络上融合了改进暗通道先验法得到的先验权重以及通道和空间双维度的复合注意力,并且改进了预测头部的全卷积层,利用设计的高效解耦预测头对目标进行最终的分类和回归,提高了在雾天图像上的目标检测精度。本文算法在自然雾天数据集 RTTS 上取得了 49.37% 的 mAP 值,较改进前提升了 4.63%。在合成数据集 S-KITTI 和 S-COCOval 上 AP 值分别为 66.7% 和 57.7%。实验结果优于其他主流检测方法,进一步验证了本文算法的适用性和鲁棒性。在未来的工作中,将进一步优化网络,均衡检测精度和速率,使其应用更加灵活。

- networks [C]//*Proceedings of the 28th International Conference on Neural Information Processing Systems*. Montreal: Curran Associates, 2015: 91-99.
- [4] GIRSHICK R B, DONAHUE J, DARRELL T, *et al.* Region-based convolutional networks for accurate object detection and segmentation [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 38(1): 142-158.
- [5] GIRSHICK R. Fast R-CNN [C]//*Proceedings of the 2015 IEEE International Conference on Computer Vision*. Santiago: IEEE, 2015: 1440-1448.
- [6] LIN T Y, DOLLÁR P, GIRSHICK R, *et al.* Feature pyramid networks for object detection [C]//*Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu: IEEE, 2017: 936-944.
- [7] LIU S, QI L, QIN H F, *et al.* Path aggregation network for instance segmentation [C]//*Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City: IEEE, 2018: 8759-8768.
- [8] PANG J M, CHEN K, SHI J P, *et al.* Libra R-CNN: towards balanced learning for object detection [C]//*Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach: IEEE, 2019: 821-830.
- [9] 李静, 喻佳成, 张灵灵. 基于改进 SSD 的航拍飞机目标检测方法[J]. 液晶与显示, 2023, 38(1): 128-137.
LI J, YU J C, ZHANG L L. Aircraft target detection method based on improved SSD [J]. *Chinese Journal of Liquid Crystals and Displays*, 2023, 38(1): 128-137. (in Chinese)
- [10] LI B Y, PENG X L, ZHANG Z Y, *et al.* AOD-Net: all-in-one dehazing network [C]//*Proceedings of the 2017 IEEE International Conference on Computer Vision*. Venice: IEEE, 2017: 4780-4788.
- [11] 汪显东, 郭继昌, 王天保. 一种改进的雾天图像行人和车辆检测算法[J]. 西安电子科技大学学报, 2020, 47(4): 70-77.
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. (in Chinese)
- [12] 解宇虹, 谢源, 陈亮, 等. 真实有雾场景下的目标检测[J]. 计算机辅助设计与图形学学报, 2021, 33(5): 733-745.
XIE Y H, XIE Y, CHEN L, *et al.* Object detection in real-world hazy scene [J]. *Journal of Computer-Aided Design & Computer Graphics*, 2021, 33(5): 733-745. (in Chinese)
- [13] HE K M, SUN J, TANG X O. Single image haze removal using dark channel prior [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, 33(12): 2341-2353.
- [14] WU Y, CHEN Y P, YUAN L, *et al.* 2020. Rethinking classification and localization for object detection [C]//*Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle: IEEE, 2020: 10183-10192.
- [15] NAYAR S K, NARASIMHAN S G. Vision in bad weather [C]//*Proceedings of the 7th IEEE International Conference on Computer Vision*. Kerkyra: IEEE, 1999: 820-827.
- [16] HE K M, ZHANG X Y, REN S Q, *et al.* Deep residual learning for image recognition [C]//*Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas: IEEE, 2016: 770-778.
- [17] CHOLLET F. Xception: deep learning with depthwise separable convolutions [C]//*Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu: IEEE, 2017: 1800-1807.
- [18] LI B Y, REN W Q, FU D P, *et al.* Benchmarking single-image dehazing and beyond [J]. *IEEE Transactions on Image Processing*, 2019, 28(1): 492-505.
- [19] GEIGER A, LENZ P, URTASUN R. Are we ready for autonomous driving? The KITTI vision benchmark suite [C]//*Proceedings of 2012 IEEE Conference on Computer Vision and Pattern Recognition*. Providence: IEEE, 2012: 3354-3361.
- [20] LIN T Y, MAIRE M, BELONGIE S, *et al.* Microsoft COCO: Common objects in context [C]//*Proceedings of the 2014 13th European Conference on Computer Vision*. Zurich: Springer, 2014: 740-755.
- [21] SHETTY S. Application of convolutional neural network for image classification on Pascal VOC challenge 2012 dataset [J/OL]. *arXiv*, 2016: 1607.03785.
- [22] REDMON J, FARHADI A. YOLOv3: an incremental improvement [J/OL]. *arXiv*, 2018: 1804.02767.

- [23] ZHANG S F, CHI C, YAO Y Q, *et al.* Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection [C]//*Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle: IEEE, 2020: 9756-9765.
- [24] ZHANG H K, CHANG H, MA B P, *et al.* Dynamic R-CNN: towards high quality object detection *via* dynamic training [C]//*Proceedings of the 16th European Conference on Computer Vision*. Glasgow: Springer, 2020: 260-275.
- [25] CAI Z W, VASCONCELOS N. Cascade R-CNN: delving into high quality object detection [C]//*Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City: IEEE, 2018: 6154-6162.
- [26] ZHU C C, HE Y H, SAVVIDES M. Feature selective anchor-free module for single-shot object detection [C]//*Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach: IEEE, 2019: 840-849.
- [27] HU J, SHEN L, ALBANIE S, *et al.* Squeeze-and-excitation networks [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, 42(8): 2011-2023.
- [28] WOO S, PARK J, LEE J, *et al.* CBAM: Convolutional block attention module [C]. 2018 European Conference on Computer Vision, Springer Cham, 2018: 3-19.
- [29] HOU Q B, ZHOU D Q, FENG J S. Coordinate attention for efficient mobile network design [C]. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville: IEEE, 2021: 13708-13717.

作者简介:



李任斯(1998—),男,湖南娄底人,硕士研究生,2020年于上海工程技术大学获得学士学位,主要从事计算机视觉和图像处理方面的研究。E-mail: 751667646@qq.com



石蕴玉(1982—),女,山东济宁人,博士,讲师,2012年于上海大学获得博士学位,主要从事图像和视频处理与分析方面的研究。E-mail: yunyushi@sues.edu.cn