分类号：TP242.6　　学校代码：10363
密　级：公开　　学　　号：2220320129

安徽工程大学
Anhui Polytechnic University

硕士学位论文

题目　雾天条件下道路场景目标检测方法研究

论文作者　张梦雯
校内导师　杨会成（教授）
校外导师　毛德平
专业学位类别　电子信息
研究方向（领域）　控制工程

论文提交日期：2025 年 5 月 12 日

分类号：TP242.6　　单位代码：10363

密　级：公 开　　学　号:2220320129

题　　目 雾天条件下道路场景目标检测方法研究

英文并列

题　　目 **Research on Road Scene Object Detection Methods Under Foggy Conditions**

学生姓名：张梦雯

校内导师：杨会成

校外导师：毛德平

专　　业：电子信息

研究方向：控制工程

论文答辩日期：2025 年 5 月 12 日

安徽工程大学学位论文原创性声明

本人郑重声明：我恪守学术道德，崇尚严谨学风。所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已明确注明和引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的作品及成果的内容。论文为本人亲自撰写，我对所写的内容负责，并完全意识到本声明的法律后果由本人承担。

学位论文作者签名：张梦雯

日期：　2025 年 5 月 12 日

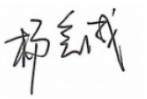
安徽工程大学学位论文版权使用授权书

学位论文作者完全了解学校有关保留、使用学位论文的规定，同意学校保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅或借阅。本人授权安徽工程大学可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论文。

保密 □，在 ____ 年解密后适用本版权书。

本学位论文属于

不保密 ☑。

学位论文作者签名：張梦雯　　指导教师签名： 

日期：　2025 年 5 月 12 日　　日期：　2025 年 5 月 12 日

雾天条件下道路场景目标检测方法研究

摘 要

道路场景目标的精确检测为智能驾驶系统的路况理解、导航决策和路径规划提供关键依据，而雾天条件下，由于空气中悬浮颗粒物的遮挡导致图像质量下降，严重影响了道路场景中目标检测的准确度。现有的图像去雾方法高度依赖图像的空间域信息，导致复原图像存在颜色失真和伪雾残留，进而影响道路场景目标的检测精度。另外，在对去雾后的图像进行目标检测时，由于小目标物体在图像中可利用特征信息较少，极易受到环境因素的干扰，进一步加剧了道路远景中小目标车辆的漏检现象。针对上述问题，本文提出了雾天条件下道路场景目标检测方法。在考虑图像空间域信息的基础上构建了基于频域增强的道路场景图像去雾模型，通过频域特征提取与自适应加权策略，融合空间信息实现高效去雾；在此基础上，提出了去雾后的道路场景图像小目标车辆检测方法，通过多尺度特征融合策略，实现了对不同尺度目标的高精度检测。本文主要研究内容和结论如下：

（1）针对现有去雾算法主要依赖图像空间信息而忽略频域信息，导致复原图像存在颜色失真和伪雾残留的问题，提出了基于频域增强的道路场景图像去雾方法。构建能够捕捉图像的空间域特征与频域特征的频域增强感知模块，灵活处理不同权重的特征信息，更加有效地去除图像中的雾气；此外，特征注意模块通过自主学习每个特征通道的重要程度，提炼显著的图像特征信息，进一步提升模型的去雾能力。在 Foggy Cityspaces 数据集上的实验结果表明，相较于基准模型，本文提出的方法峰值信噪比（Peak Signal-to-Noise Ratio, PSNR）和结构相似性（Structure Similarity Index Measure, SSIM）分别提升了 31.1413

dB 和 0.9179，其图像质量恢复效果优于现有的图像去雾方法，证明了所提出方法在去除图像雾气方面的有效性。

（2）针对去雾道路场景图像中的小目标车辆漏检问题，提出了基于特征融合的去雾道路场景图像小目标检测方法。首先利用 Focus 模块对输入特征图进行降维操作，保留重要的空间特征信息；接下来设计了轻量化的多尺度特征融合模块，通过区域残差化生成多尺度特征图与语义残差化实施多速率空洞深度可分离卷积相结合，增强对小目标物体的特征表达能力。进一步构建了双向特征金字塔模块，通过双向传播特征信息和自适应调整特征权重，更好地融合不同级别的特征信息。实验结果表明，所提方法在 Foggy Cityspaces 数据集上平均精度、精确率、召回率相比于基准模型分别提高了 1.1%、6.2%和 0.8%，实现了道路远景小目标车辆的有效检测。

综上所述，本文提出的雾天条件下道路场景目标检测方法在高效去除图像雾气的基础上，显著提高了去雾后道路场景小目标车辆的检测精度，为雾天条件下道路场景的目标检测提供了有效的技术支撑。

关键词：道路场景，图像去雾，小目标检测，频域增强，特征融合

RESEARCH ON ROAD SCENE OBJECT DETECTION METHODS UNDER FOGGY CONDITIONS

ABSTRACT

Accurate detection of road scene targets provides critical support for the understanding of road conditions, navigation decision-making, and path planning in intelligent driving systems. However, under foggy conditions, the degradation of image quality due to the occlusion of suspended particles in the air significantly affects the accuracy of target detection in road scenes. Existing image dehazing methods heavily rely on spatial domain information, resulting in color distortion and residual fog in the restored images, which further impacts the accuracy of target detection in road scenes. Additionally, when detecting targets in defogged images, small objects with limited feature information are highly susceptible to environmental interference, exacerbating the issue of missed detection for small target vehicles in distant road scenes. To address these challenges, this paper proposes a target detection method for road scenes under foggy conditions. By constructing a frequency-domain-enhanced dehazing model for road scene images that incorporates spatial domain information, the proposed method achieves efficient dehazing through frequency-domain feature extraction and adaptive weighting strategies. Furthermore, a small target detection method for defogged road scene images is proposed, which employs a multi-scale feature fusion strategy to achieve high-precision detection of targets at different scales. The main research contributions and conclusions of this paper are as follows:

(1) To address the issue of color distortion and residual fog in restored images caused by existing dehazing algorithms that primarily rely on spatial information while neglecting frequency-domain information, a frequency-domain-enhanced dehazing method for road scene images is proposed. A frequency-domain enhancement-aware module is constructed to capture both spatial and frequency-domain features of the image, enabling flexible handling of feature information with different weights and more effective removal of fog from the image. Additionally, a feature attention module autonomously learns the importance of each feature channel, refining significant image features and further enhancing the model's dehazing capability. Experimental results on the Foggy Cityscapes dataset show that the proposed method improves the Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) by 31.1413 dB and 0.9179, respectively, compared to the baseline model. The image quality restoration outperforms existing dehazing methods, demonstrating the effectiveness of the proposed approach in removing fog from images.

(2) To address the issue of missed detection of small target vehicles in defogged road scene images, a small target detection method based on feature fusion is proposed. Firstly, the Focus module is employed to conduct dimensionality reduction on the input feature map, thereby preserving critical spatial feature information. Subsequently, a lightweight multi-scale feature fusion module was developed. By integrating the generation of multi-scale feature maps via regional residualization and implementing multi-rate dilated depthwise-separable convolutions through semantic residualization, the feature representation capability for small target objects was significantly enhanced. Furthermore, a Bidirectional Feature Pyramid Network (BiFPN) module is constructed to better integrate feature information at different levels

through bidirectional feature propagation and adaptive feature weight adjustment. Experimental results show that the proposed method improves the average precision, precision, and recall by 1.1%, 6.2%, and 0.8%, respectively, on the Foggy Cityscapes dataset compared to the baseline model, achieving effective detection of small target vehicles in distant road scenes.

In summary, the proposed target detection method for road scenes under foggy conditions significantly improves the detection accuracy of small target vehicles in defogged road scenes while efficiently removing fog from images. This provides effective technical support for target detection in road scenes under foggy conditions.

KEY WORDS: road scene, image dehazing, small object detection, frequency domain enhancement, feature fusion

符号及缩略语的说明

英文缩写	英文全称	中文全称
AHE	Adaptive Histogram Equalization	自适应直方图均衡化
AOD-Net	An All-in-One Network for Dehazing and Beyond	一体化除雾网络
AP	Average Precision	平均精确度
BBox	Bounding Box	边界框
BiFPN	Bidirectional Feature Pyramid Network	双向特征金字塔网络
BN	Batch Normalization	批量归一化
CA	Channel Attention	通道注意力
CBS	Cord-Bottleneck-Short	—
CDF	Cumulative Distribution Function	累积分布函数
CLAHE	Contrast Limited Adaptive Histogram Equalization	限制对比度自适应直方图均衡化
CNN	Convolutional Neural Network	卷积神经网络
CVPR	IEEE Conference on Computer Vision and Pattern Recognition	计算机视觉与模式识别 IEEE 会议
C2PSA	Cross Stage Partial with Pyramid Squeeze Attention	—
DConv	Dilated Convolution	膨胀卷积
DDNet	Double-feature Double-motion Network	—
DWConv	Depthwise Separable Convolution	深度可分离卷积
DWR	Dilation-wise Residual	扩张残差
Fast R-CNN	Fast Region-based Convolutional Neural Network	快速区域卷积神经网络
FC	Fully Connected Layer	全连接层
FCN	Fully Convolutional Networks	全卷积神经网络
FFT	Fast Fourier Transform	快速傅里叶变换
FFN	Feed-Forward Network	前馈网络
FNN	Feed-Forward Neural Network	前馈神经结构块

英文缩写	英文全称	中文全称
FPN	Feature Pyramid Network	特征金字塔网络
GAP	Global Average Pooling	全局平均池化
GCA-Net	Gated Context Aggregation Network for Image Dehazing and Deraining	门控融合子网
IFFT	Inverse Fast Fourier Transform	快速傅里叶逆变换
ITS	Indoor Training Set	—
LRL	Local Residual Learning	局部残差学习
mAP	mean Average Precision	平均精度均值
MB-Conv	Mobile Convolution	移动卷积
MSBDN	Multi-Scale Boosted Dehazing Network with Dense Feature Fusion	—
MSE	Mean-Square Error	均方误差
NAFSSR	Stereo Image Super-Resolution Using NAFNet	—
NMS	Non-Maximum Suppression	非极大值抑制
NTIRE	New Trends in Image Restoration and Enhancement Challenges	图像恢复和增强挑战的新趋势
OTS	Outdoor Training Set	—
PA	Pixel Attention	像素注意力
PANet	Path Aggregation Network	路径聚合网络
PSA	Pointwise Spatial Attention	点式空间注意力
PSNR	Peak Signal-to-Noise Ratio	峰值信噪比
R-CNN	Region-based Convolutional Neural Network	区域卷积神经网络
ReLU	Rectified Linear Unit	线性整流函数
ResNet	Deep residual network	深度残差网络
ROI	Region of Interest	感兴趣区域
RPN	Region Proposal Network	区域建议网络
SCAM	Stereo Cross-Attention Module	双目交叉注意力模块
SE-Net	Squeeze-and-Excitation Networks	—
SIR	Simple Inverted Residual	反向残差
SOTS	Synthetic Objective Testing Set	—
SPPF	Spatial Pyramid Pooling – Fast	—

英文缩写	英文全称	中文全称
SPP-Net	Spatial Pyramid Pooling Network	空间金字塔池化网络
SSD	Single Shot MultiBox Detector	—
SSIM	Structure Similarity Index Measure	结构相似性
SVM	Support Vector Machine	多类别支持向量机
YOLO	You Only Look Once	—

目录

第 1 章 绪论

1.1 研究背景与意义

智能汽车作为战略性新兴产业，其道路场景精确感知技术的发展对我国经济与科技进步有着重大意义，目标检测作为道路场景精确感知的基础，能精准识别复杂环境目标并实现驾驶决策，是当前计算机视觉领域的热门研究方向。然而在实际应用中，视觉系统的成像效果往往受到多方面因素的干扰。光学镜头的性能、感光元件的非线性响应、设备抖动以及大气扰动等都会导致获取的图像出现不同程度的失真，特别是能见度较低、发生频率高和覆盖范围广的雾天天气，是制约视觉系统性能的重要因素[1]。当雾气存在时，拍摄出来的图像会因颗粒物的遮挡而导致原始信息丢失，并且可能伴随颜色失真等现象，这不仅影响了图像信息的有效获取，还给智能驾驶汽车的安全行驶造成严重威胁。因此，急需找到高效的方法以消除雾天对人们生活的影响。有雾与无雾道路场景图像如图 1-1。

(a)有雾图像　　(b)无雾图像

图 1-1 有雾与无雾道路场景图像

Fig 1-1 Foggy and Fog-free Road Scenes

近年来，许多研究者已提出多种去雾方法，旨在高效的消除图像中的雾气，提升图像的清晰度。然而，现有的去雾算法在如何提升图像细节信息，确保复原图像更加贴近真实自然方面还有待提高。同时，现有目标检测算法大多关注大尺

寸目标，小目标在图像中通常只占据极小的像素区域，导致小目标物体更易受到其他物体的遮挡，现有目标检测算法难以准确捕捉和识别这些小目标，从而造成小目标漏检问题。因此，对雾天条件下道路场景目标检测方法的研究具有极其重要的意义，其能够有效降低雾天交通事故的风险，精确识别道路远景中的车辆目标，为驾驶员提供及时准确的道路信息，从而有效避免潜在的危险。

综上所述，图像去雾与目标检测等关键技术研究具有重要的学术价值和实际应用意义。这些研究能够有效解决雾天环境对道路场景精确感知系统带来的不利影响，不仅提升了自动驾驶技术的可靠性，还为道路场景精确感知领域的深入推进提供有力支撑。

1.2 国内外研究现状

随着计算机视觉和人工智能的发展，道路场景感知算法研究受到了国内外许多学者的关注，涌现出了大量优秀成果。相对于一般的目标检测任务，雾天场景下的目标检测任务较少受到关注。接下来，针对本课题的研究内容，将从图像去雾和目标检测两个方面对研究现状进行阐述。

1.2.1 图像去雾算法

由于大气污染现象越发严重，空气中的悬浮颗粒对光线传播造成了显著干扰，不仅改变了光线的折射路径，还会在传输过程中发生散射和衰减现象，致使图像采集设备采集的图像质量严重下降，无法满足道路场景精确感知、目标检测等以清晰图像作为基础的专业领域需求，因此对有雾图像进行去雾清晰化处理具有非常重要的意义和应用价值。目前图像去雾方法主要可分为三大类：基于图像增强、基于图像复原和基于深度学习的图像去雾方法[2]。

（1）基于图像增强的去雾方法

基于图像增强的去雾方法是主要通过改善图像的视觉属性，如亮度、对比度等手段改善受雾气影响的图像质量。直方图均衡化的图像增强算法是由 Ketcham 在 1974 年提出的，基本思想是对图像中的灰度点做映射，使图像的灰度大致符合均匀分布，以达到图像对比度增强的目的，适用于对比度较低或亮度分布不均匀的图像，使其看起来更加清晰和自然。直方图均衡化通过拉伸像素值的分布范

围，使得图像的灰度级更加均匀分布，从而改善图像的视觉效果。通过计算原始图像的直方图后，将直方图进行归一化处理，接下来通过映射函数将归一化后的直方图转换为新的灰度级，最后应用映射函数到原始图像上，得到增强后的图像，但是存在变换后的图像灰度级减少，某些细节减少等问题。

自适应直方图均衡化（Adaptive Histogram Equalization, AHE）是图像对比度改善技术，适用于在不同区域具有不同亮度的图像，与传统直方图均衡化相比，AHE 会根据图像的局部区域来调整对比度，因而能够在保持图像局部细节的同时提升图像对比度。AHE 采用局部处理策略，通过计算每个像素点周边区域的映射函数来实现图像增强，即利用像素点周围方形区域内灰度分布的统计特性，基于累积概率分布函数（Cumulative Distribution Function, CDF）对每个像素进行均衡化处理。AHE 虽然能够提升图像的对比度表现，但会不可避免地增强图像同质区域的噪声。限制对比度自适应直方图均衡化（Contrast Limited Adaptive Histogram Equalization, CLAHE）通过引入对比度约束机制来优化传统自适应增强方法，在计算局部映射函数时，对每个像素周边区域的对比度提升幅度进行了有效控制，从而解决了 AHE 方法中噪声被过度放大的问题。CLAHE 算法虽然能够有效地增强图像的局部对比度并提高图像的视觉效果，但也存在过度增强与计算过于复杂以及适用性限制等问题。

Land 等[3]基于色彩恒常性提出视网膜皮层理论（Retinex），以颜色恒常性原理为基础，旨在模拟人类视觉系统的颜色感知机制，特别是视网膜（Retina）和皮层（Cortex）对颜色和亮度的处理过程，所以 Retinex 的思路即是去除光照的影响，保留住物体的固有属性。一幅给定的图像 $S(x,y)$ 以分解为两个不同的图像，反射图像 $R(x,y)$ 和入射图像 $L(x,y)$。

$$S(x,y)=R(x,y)\times L(x,y) \tag{1-1}$$

其中，$L(x,y)$ 表示入射光图像，$R(x,y)$ 表示物体的反射性质图像，$S(x,y)$ 表示人眼所能接收到的反射光图像[4-5]。Retinex 算法虽然可以对图像进行自适应增强，但存在亮区处理欠佳、色彩保持能力较弱以及计算较复杂等问题。因此后续有人对该理论进行了改进，提出了基于路径的 Retinex 算法[6-7]以及迭代形式的 Retinex 算法[8]，但是这些算法都存在运行时间较慢和调参复杂问题[9]。

DD-Net（Double-feature Double-motion Network, DD-Net）[10]用于超高清交

通监控的双域引导实时低光图像增强网络，采用编码器-解码器结构，能够恢复被黑暗隐藏的详细特征信息，但特征表达能力不够准确，识别精度较低。综上所述，基于图像增强的去雾方法虽然在一定程度上能够改善图像的视觉效果，但存在色彩失真、对整幅图像不同浓度雾气进行统一处理以及算法局限性等缺点。

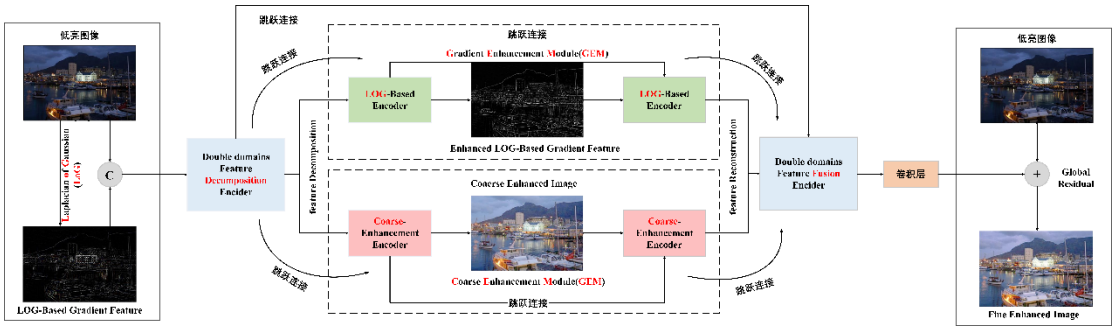


图 1-2 DD-Net 网络结构图[10]

Fig 1-2 DD-Net Network Structure

（2）基于图像复原的去雾方法

基于图像复原的去雾方法主要从光学成像的角度，分析雾天对成像过程的影响，现阶段基于图像复原的去雾算法主要依赖于大气散射模型[10-12]。Srinivasa G. Narasimhan[13]通过建立数学模型，阐述了雾天条件下图像的形成机理及其构成要素[14]，大气散射模型的公式（1-2）为:

$$I(z)=J(z)t(z)+A(1-t(z)) \tag{1-2}$$

其中 $I(z)$ 为观测到的雾气图像， A 为全球大气光， $t(z)$ 为介质透射图， $J(z)$ 为无雾气图像。式（1-2）也可表述为:

$$J(z)=\left(\left(I(z)-A\right)/t(z)\right)+A \tag{1-3}$$

对于单幅图像的去雾问题，从公式（1-2）和公式（1-3）基于精确的大气散射模型参数估计与介质透射率分布计算，理论上可实现无雾场景的有效复原。

而何凯明等人[15]提出的暗通道先验去雾理论是利用大气对光线散射的特性，具体来说，该算法首先假设在无雾情况下，像素点之间的距离越远，大气光强度越低。暗通道先验去雾理论可以通过分析像素点之间的距离和大气光强度之间的关系，来估计大气光的强度，一旦获得了大气光的强度，就可以使用该信息来消除图像中的雾气。传统的去雾算法通常会通过查找像素点周围的相似区域，并将这些区域的色彩分布进行插值，以得到清晰明亮的图像，然而这种方法往往会引入噪声和伪影，影响图像质量。暗通道去雾算法通过使用大气光强度信息，可以

更准确地估计出图像中的实际场景，从而得到更加真实、自然的去雾效果。暗通道先验去雾理论实验结果图如图 1-3 所示，图（a）为输入图像；图（b）为基于暗通道先验去雾理论的去雾效果图；图（c）为暗通道效果图；红色像素表示估算大气光的位置；图（d）为大气光估计传播图。

图 1-3 暗通道先验去雾理论去雾效果展示图[15]

Fig 1-3 Dark Channel Prior Dehazing Effect Display Diagram

暗通道的数学定义如式（1-4），其中 $J_c(y)$ 表示 J 的任意一个颜色通道，$\Omega(x)$ 表示在像素点 x 的窗口。根据暗通道先验理论得出：$J_{dark} \to 0$。接下来利用大气散射模型对大气光做归一化计算并经过运算，且引入一个恒定的参数 $\omega(0<\omega\le 1)$，得到透射率估计如式（1-5）：

$$J_{dark} = \min_{y\in\Omega(x)}\left(\min_{c\in(r,g,b)} J_c(y)\right) \tag{1-4}$$

$$t(x) = 1-\omega\min_{y\in\Omega(x)}\left(\min_c \frac{I_c(y)}{A_c}\right) \tag{1-5}$$

基于暗通道先验理论，大气光的估算通常选取图像中雾气最浓重的区域作为参考。通过分析暗通道图像的特性可以发现，雾气浓度较低的区域在暗通道图中呈现较深的色调，对应的像素值较小；反之，雾气浓度较高的区域则表现出较亮的特征，像素值相对较大。在实际计算过程中，首先从暗通道图中筛选出亮度值处于前 1%的像素区域，随后将这些高亮像素对应到原始雾化图像中，通过比较这些位置的像素亮度，最终选取最亮的像素点作为大气光强度 A 的估计基准，可得出最终的复原场景如式（1-6）：

$$J(x) = \frac{I(x)-A}{\max(t(x),t_0)} + A \tag{1-6}$$

其中 t_0 值为 0.1。虽然暗通道先验去雾理论有着较为广泛的可应用性与实践性，但由于天空等高亮场景并不符合暗通道先验理论，容易致使去雾后的图像出现色彩失真现象。

NAFSSR（Stereo Image Super-Resolution Using NAFNet, NAFSSR）[16]是一种用于双目图像超分辨率的深度学习模型，是在 CVPR（IEEE Conference on Computer Vision and Pattern Recognition, CVPR）举办的 NTIRE（New Trends in Image Restoration and Enhancement Challenges, NTIRE）2022 双目超分辨率挑战赛中，由旷视研究院提出的冠军方案。NAFSSR 不仅继承了 NAFNet 的优势，还通过引入双目交叉注意力模块（Stereo Cross-Attention Module, SCAM），实现了对双目图像中视角间特征的深度融合，从而显著提升了超分辨率图像的质量，能够以较低的系统复杂度在图像复原任务上获得较好的性能。NAFSSR 的核心在于其独特的网络结构和创新的注意力机制，该网络由两个共享权重的子网络组成，分别负责处理左右图像，两个子网络由 NAFBlock 堆叠而成，NAFBlock 是 NAFNet 中的基本模块，包括移动卷积（Mobile Convolution, MB-Conv）模块和没有非线性激活函数的前馈网络（Feed-Forward Network, FFN）模块，使得网络在保持高效性的同时能够充分提取图像中的特征信息。然而，NAFSSR 的真正创新之处在于基于 Scaled Dot-Product Attention 机制设计的 SCAM，通过计算从左视图到右视图和从右视图到左视图的双向交叉注意力，实现了跨视图特征的融合。具体来说，SCAM 首先计算左右图像特征图之间的相似度，然后利用 softmax 函数得到注意力权重，最后对特征图进行加权求和从而得到融合后的特征，使得网络能够充分融合左右视图间的特征，提高超分辨率图像的细节和清晰度。NAFSSR 具有简单而强大、有效融合信息以及解决过拟合问题等优势，但是其依赖大量训练数据且计算复杂度较高。综上所述，虽然基于图像复原去雾方法的最终效果普遍好于传统图像增强去雾方法，但该种方法恢复过程较为复杂且费时。

（3）基于深度学习的去雾方法

随着人工智能算法的日益发展，基于深度学习的图像去雾方法，尤其是以卷积神经网络（Convolutional Neural Network, CNN）[17]代表的技术已经得到了广泛应用。Li 等人[18]提出了一种基于 CNN 的图像去雾模型，称为一体化除雾网络（An All-in-One Network for Dehazing and Beyond, AOD-Net），采用了基于卷积神经网络的端到端设计，直接缩小了雾气图像与清晰图像之间的最终目标差距，直接输出去雾后的清晰图像，避免了传统方法中传输矩阵和大气光的单独估计步骤，简化了去雾过程，实现图像去噪的自适应性和最优性，提高了运行效率。其去雾原理基于大气散射模型，在保留细节的同时剥离图像中的不透明层，从而恢复图

像的真实色彩与纹理。AOD-Net 的网络结构紧凑且高效，包含了多个卷积层和残差块，能够有效地提取图像的特征并进行去噪处理，还采用了深度残差网络（Deep residual network, ResNet）的结构，有助于网络在训练过程中更好地收敛，并提高了网络的泛化能力。此外，AOD-Net 是第一个探索去雾模型如何帮助后续高层次视觉任务的模型，可以通过将自适应最优估计理论引入深度学习框架中，从而提升雾霾图像上的高层次任务性能。AOD-Net 网络虽然以端到端的设计取得了较为优秀的去雾性能但计算较复杂和资源消耗较大。

Chen 等人[19]提出门控融合子网（Gated Context Aggregation Network for Image Dehazing and Deraining, GCA-Net）引入了一种新的端到端网络结构，由于其独特的门控注意力机制，能够关注图像中的重要特征，从而更有效地增强图像的清晰度和细节表现，无论是雾气笼罩的城市景观还是雨水覆盖的画面，都能得到显著优化。GCA-Net 主要由特征提取网络和门控注意力模块两大部分构成，特征提取网络负责捕捉输入图像中的低级和高级特征，为后续的处理提供基础，而门控注意力模块通过计算不同通道和空间位置的注意力权重来识别图像中的重要区域。注意力权重被用来调整特征图，抑制不重要的信息，同时增强重要的特征，从而实现自适应的特征增强。GCA-Net 显著增强了图像的清晰度和可见性，但当操作区域比真实像素大时网络会出现较大误差且网络结构复杂，计算耗时较长。

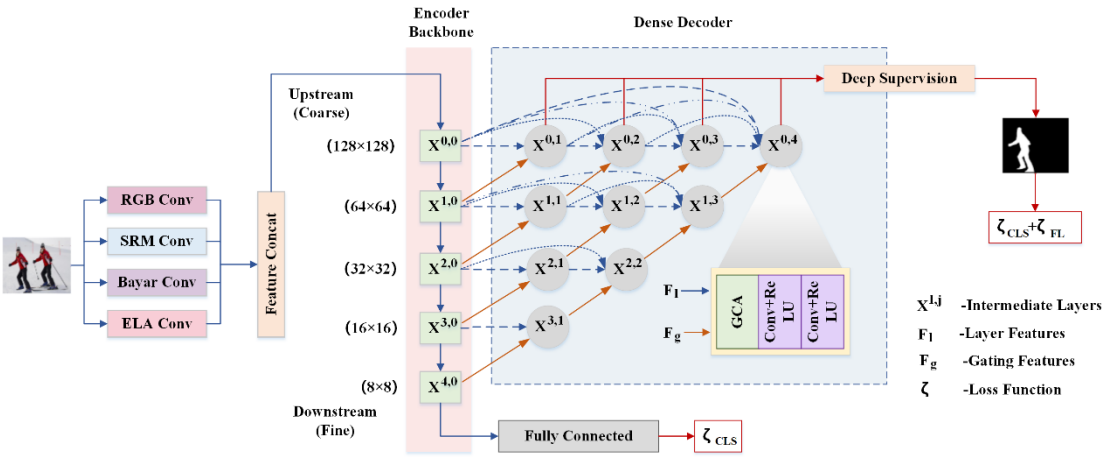


图 1-4 GCA-Net 网络结构图[19]

Fig 1-4 GCA-Net Network Structure

MSBDN（Multi-Scale Boosted Dehazing Network with Dense Feature Fusion，MSBDN）[20]通过多尺度特征融合，能够同时捕获图像的全局和局部信息，从而

提高去雾效果。MSBDN 的核心在于其多尺度增强解码器和密集特征融合模块，多尺度增强解码器通过“Strengthen-Operate-Subtract”增强策略，逐步细化去雾结果，从而得到更加清晰、真实的图像；同时密集特征融合模块利用反投影反馈方案设计，能够有效进行图像去雾并且保留高分辨率特征空间信息，实现不同层次特征的有效融合，旨在解决 U-Net 架构中空间信息丢失的问题。在具体实现上，MSBDN 采用了编码器-解码器的结构，在编码器阶段，输入的有雾图像经过多个卷积层和残差组处理，提取出图像的特征信息，传递到解码器阶段，通过多尺度增强解码器和密集特征融合模块的处理，逐步恢复出无雾图像。MSBDN 作为具有密集特征融合的多尺度增强去雾网络，在图像去雾方面展现出了卓越的性能和广泛的应用前景，但依旧出现计算复杂度较高且会出现产生伪影或伪雾等问题。基于深度学习的图像去雾方法克服了基于图像增强和图像复原的不足，在模型性能和泛化性能上均取得了很大的提升。综上所述，目前的图像去雾方法虽然可以改善图像质量、恢复图像细节以及泛化能力强，但仍存在着处理后颜色失真和伪雾残留等问题。

1.2.2 目标检测算法

目标检测是计算机视觉领域的重要分支，主要解决物体识别与位置确定两大关键问题。现有的目标检测方法可以分为两类：基于传统机器学习的目标检测方法和基于深度学习的目标检测方法。基于传统机器学习的目标检测方法通常使用分段式处理模式，检测精度较低。相比之下，基于深度学习的目标检测方法凭借端到端的检测优势而受到广泛应用，现阶段基于深度学习的目标检测方法主要分为两阶段目标检测算法和单阶段目标检测算法。两阶段目标检测算法采用分步处理策略，首先提出建议框，通过第一阶段的网络生成可能包含目标的区域建议，并预测其粗略位置、尺度及前景概率，第二阶段利用另一个网络对这些建议区域进行精修，最终输出精确的目标位置、尺寸等。两阶段目标检测算法虽然在大目标和复杂场景中有较好的检测精度，但运行速度较慢。而单阶段目标检测算法的核心是对输入图像，通过网络直接回归出目标大小、位置和类别，运行速度较快，但易出现小目标车辆漏检和误检问题。

两阶段目标检测算法通常包括以下两个阶段：第一阶段对图像进行初步检测，筛选出所有正样本，并生成感兴趣区域（Region of Interest, ROI）。这些方法能够

基于图像信息，自动地找出可能包含目标的区域。在第二阶段，通常是通过 CNN 来对前一阶段生成的 ROI 进行区域分类和位置细化，CNN 能够自动地提取图像中的高层特征，提取的特征会被送入分类器进行分类，以确定候选区域是否包含目标并进一步确定其具体类别归属。

Ross-Girshick 等人[21]提出了区域卷积神经网络（Region-based Convolutional Neural Network, R-CNN），R-CNN 算法通过候选区域生成、特征提取、分类和定位这四个步骤，能够实现对图像中目标的精确检测。在 R-CNN 中，候选区域的生成是第一步，通常使用选择性搜索算法，能够基于图像的颜色、纹理、形状等特征自动地找出可能包含目标的区域作为后续处理的输入，大大减少了计算量，同时保留了可能包含目标的重要信息。接下来是特征提取阶段。R-CNN 利用具有更强的表达能力和鲁棒性的 CNN 对每个候选区域进行特征提取，通过深度卷积网络强大的特征表达能力，R-CNN 能够捕捉到更加丰富和准确的图像信息，为后续的分类和定位任务提供了有力的支持。在分类阶段，R-CNN 首先将经过卷积神经网络提取的深层特征向量，输入到多类别支持向量机（Support Vector Machine, SVM）分类器中，能够有效地区分不同目标类别，从而实现高准确率的物体识别任务。同时，为了对初步检测得到的候选区域坐标进行精细化调整，提升定位精度，R-CNN 还训练了一个边界框（Bounding Box, BBox）回归模型。以提高目标检测的准确性，能够通过边框回归模型对框的准确位置进行修正，使得检测出的目标位置更加接近真实位置，然而 R-CNN 也存在一些局限性，对于小目标和遮挡目标的检测效果不尽如人意。

因此 He 等人[22]提出了空间金字塔池化网络（Spatial Pyramid Pooling Network, SPP-Net），SPP-Net 的核心在于其引入放置在卷积层与全连接层之间的空间金字塔池化层，主要作用是将任意尺寸的卷积特征映射转换为统一维度的特征表示。具体来说，空间金字塔池化层通过将特征图分割成多尺度的网格单元，每个网格单元内部都进行池化操作，确保了无论原始输入图像的尺寸如何变化，经过该层处理后都能输出维度一致的特征向量，满足了全连接层对固定输入尺寸的要求。SPP-Net 成功克服了传统网络对输入图像尺寸的限制，使其处理速度较 R-CNN 提升了数倍之多。然而 SPP-Net 也存在一定的局限性，在目标尺寸差异极大的情况下，其性能仍受到较大影响。

后续，Ross-Girshick 等人[23]又提出了快速区域卷积神经网络（Fast

Region-based Convolutional Neural Network, Fast R-CNN），使用两个平行的不同完全连接层分别完成分类和定位任务，解决了 R-CNN 和 SPP-Net 算法中需要单独训练 SVM 分类器的问题。与 R-CNN 不同，Fast R-CNN 对整个输入图像只进行一次卷积特征图的计算，随后进行后续的区域建议操作，显著减少了重复计算。此外，Fast R-CNN 引入了一个关键的组件 RoI 池化层，通过分割候选区域并在每个子区域内进行最大池化，接受输入的卷积特征图和候选区域列表，并从每个候选区域提取一个固定大小的特征图，从而保证不论输入候选区域的原始大小如何，输出都是固定的。Fast R-CNN 还支持端到端的训练策略，避免了 R-CNN 中分阶段的、独立的训练过程。在应用方面，Fast R-CNN 展现出了优越的性能，并且更加适合实际应用，特别是需要实时或接近实时响应的场景，但该算法不能改善区域建议算法的限制，区域建议算法的精度有限，即使经过调整，也可能无法准确覆盖目标对象的完整范围，导致在后续的分类和回归任务中出现误差。

Ren 等人[24]在 2015 年提出了 Faster R-CNN，通过加入区域建议网络（Region Proposal Network, RPN）将两者集成为一个完整的网络，可以端到端学习，既保证了准确性又提高了速度。Faster R-CNN 的结构主要包含四个部分，首先使用预训练的卷积神经网络作为特征提取器，生成一个高维特征图，捕捉输入图像中的不同层次的特征；其次 RPN 包含若干个小型卷积层，每个位置都对应一组具有不同尺度和比例的锚框，对于每一个锚框，RPN 会预测该位置是否包含一个目标和 BBox 回归参数，因此可以直接从基础网络生成的特征图中产生候选区域；接下来 ROI 池化层将不同尺寸的候选区域统一成固定大小的特征向量，确保后续全连接层能够接收相同形状的输入；最后特征向量被送入两个并行的全连接层：一个用于目标类别的判定，另一个则专注于边界框的精确回归。尽管 Faster R-CNN 在检测效率和准确率方面表现优异，但在处理小尺度目标时仍存在识别精度不足的局限性。两阶段目标检测算法通常检测精度更高，但由于复杂度的提升，在实时性方面略有不足。

单阶段（One-stage）目标检测采用端到端的处理方式，通过一次性预测目标的类别信息和空间位置，有效提升了算法效率。单阶段（One-stage）目标检测在训练过程对标记的数据进行编码，使其与 CNN 的相应输出格式一致，从而计算模型损失；在测试过程输入图像经过处理后直接生成预测结果，通过解码操作最终输出目标检测框，有效提高了检测速度。

常见的单阶段算法有 YOLO（You Only Look Once, YOLO）[25]、SSD（Single Shot MultiBox Detector, SSD）等。SSD 算法[26]的核心思想是将目标检测任务转化为一个单次前向传播的回归问题，从而实现了高效的目标检测，摒弃了传统目标检测算法中繁琐的候选区域生成和分类步骤，直接在特征图上进行目标的定位和分类预测。具体来说，SSD 通过在基础网络上添加额外的卷积层来生成不同尺度的特征图，用于检测不同大小的物体，在每个特征图上，SSD 定义了一系列的相同形状和大小的先验框，用于预测物体的位置和类别。SSD 算法通过在图像中多尺度、多比例地密集采样，结合 CNN 直接实现目标分类和位置回归，虽然有效的提升了检测速度，但由于样本分布失衡，增加了模型训练的难度，从而影响了最终的检测精度。此外，SSD 算法对小尺寸目标的识别能力明显弱于大尺寸目标。

YOLO 是一种基于深度学习的实时目标检测算法，算法的核心思想是将目标检测任务视为一个回归问题，通过单次前向传播直接预测物体的 BBox 和类别。YOLO 算法摒弃了传统目标检测算法中繁琐的候选区域生成和分类步骤，提高了检测速度和效率，同时降低了漏检率和误检率，具有背景误识别率低、泛化性能好、检测精度高的优势，因此在单阶段目标检测方面更具有代表性，如 Joseph Redmon 的《You Only Look Once : Unified, Real-Time Object Detection》为单阶段目标检测的开山鼻祖。Redmon 等人[27]提出了 YOLO 模型，将目标检测任务转化为 BBox 空间分离与类别概率预测的联合回归问题，通过构建树状结构的标签体系实现了跨数据集的类别整合，在保持较低计算成本的同时展现出优异的推理速度。随后，Redmon 等人又对 YOLO 检测方法进行各种改进增强，提出了 YOLOv2 算法，使其能够使用分层排列识别 9000 个类别以上的目标，但 YOLOv2 没有进行多尺度特征的结合预测，而且依旧无法检测到非常小的目标。YOLOv3[28]通过引入多维度检测策略和 Darknet-53 骨干网络，显著提升了对小目标的检测能力，同时利用特征金字塔网络，支持多标签分类任务，但其模型尺寸急剧增大，难以在嵌入式设备上部署。YOLOv4[29]引入 CSPDarknet53 和多种增强技术，实现了性能与速度的最佳平衡，尽管 YOLOv4 在多尺度特征融合方面有所改进，但在检测小目标时仍可能存在一定的局限性。YOLOv5[30]相较于 YOLOv4，无论是学习率调整、运行效率方面还是模型精度、速度都有显著提升，但其对于小目标、密集物体检测方面还有待提升。YOLOv6[31]更专注于工业场景的实际需求，在速

度和精度之间达到了较好的权衡，但是在光照、姿态等条件变化较大的场景下，检测精度会下降。YOLOv7[32]实现了轻量化网络结构的设计，在速度和精度上都超过 YOLOv6 及之前版本的目标检测算法，但需要大量的训练时间以及在某些复杂场景下对小目标的检测效果可能不如其他算法。YOLOv8 在复杂场景下表现优异，引入了新的注意力机制，但是复杂的网络结构和多个模块增加了模型的复杂度和训练难度。YOLOv9[33]基于 YOLOv7 进行改进，增强了自动化和模型自适应能力，在深度模型的参数数量、计算量等方面都比 YOLOv8 有所减少，但增加了模型开发的时间成本。YOLOv10[34]针对超大规模模型进行了优化，实现了高效的端到端检测，不过虽然进行了轻量化设计，但相对于一些更简单的模型来说，YOLOV10 的复杂度仍然较高。YOLO11[35]是 Ultralytics 团队于 2024 年 9 月推出的 YOLO 系列的最新版本，采用改进的骨干和颈部结构，增强了特征提取能力，从而实现更精确的目标检测，以最先进的精度、速度和效率重新定义了可能实现的目标。

单阶段目标检测算法相比两阶段目标检测算法具有较快的检测速度，在实时性要求较高的应用场景中展现出独特价值。单阶段目标检测算法以其高效性和实时性在计算机视觉领域占据了重要地位，未来随着技术的持续革新和应用场景的不断拓展，单阶段目标检测算法将迎来更加广阔的广泛的发展空间和应用前景。

1.3 主要研究内容

针对雾天条件道路场景图像中远处被雾气遮盖的机动车辆等小目标难以被检测的问题，本文提出了雾天条件下道路场景目标检测方法，主要的研究内容如下：

（1）基于频域增强的道路场景图像去雾方法

雾天条件下，空气中存在的微粒会降低图像的清晰度，阻碍对图像的进一步分析与处理。针对现有的去雾算法主要利用图像空间信息而忽略频域信息，从而导致复原图像存在颜色失真和伪雾残留的问题，提出了基于频域增强的道路场景图像去雾方法。构建的图像去雾网络模型由多个级联的组架构与特征注意模块构成。在组架构中设计了频域增强感知模块，通过捕捉图像的空间域特征与频域特征，灵活处理不同权重的特征信息，更加有效地去除图像中的雾气；特征注意模

块通过自主学习每个特征通道的重要程度，对不同比重的特征信息进行分别处理，从而提炼显著信息，提升去雾模型的细节恢复能力，有效提高图像去雾效果。实验结果验证了所提方法能够有效去除图像中的雾气，并恢复场景细节信息，图像质量恢复效果优于现有的图像去雾方法。

（2）基于特征融合的道路场景图像小目标检测方法

图像远景的车辆因占据的像素较少，目标检测算法难以提取到有效信息，从而会影响检测精度。针对去雾道路场景图像中的小目标漏检问题，提出了基于特征融合的道路场景图像小目标车辆检测方法。构建的小目标检测模型以YOLOv11 模型作为基准，首先在骨干网络中增加 Focus 模块，通过对输入特征图进行降维操作，在减少计算量的同时，保留重要的空间信息，从而增强对小目标的特征关注，提高小目标检测精度；其次设计了多尺度特征融合模块，通过设计结合区域残差化和语义残差化的两步特征提取方式，有效提高小目标检测的准确性；最后在颈部网络中设计双向特征金字塔模块，通过特征信息的双向传播，更好地融合不同级别特征信息，并且结合自适应特征调整机制，进一步提高模型的检测性能。通过实验结果表明，本文提出的方法显著减少了去雾后的图像中道路远景小目标车辆的漏检，有效提升了小目标的检测精度。

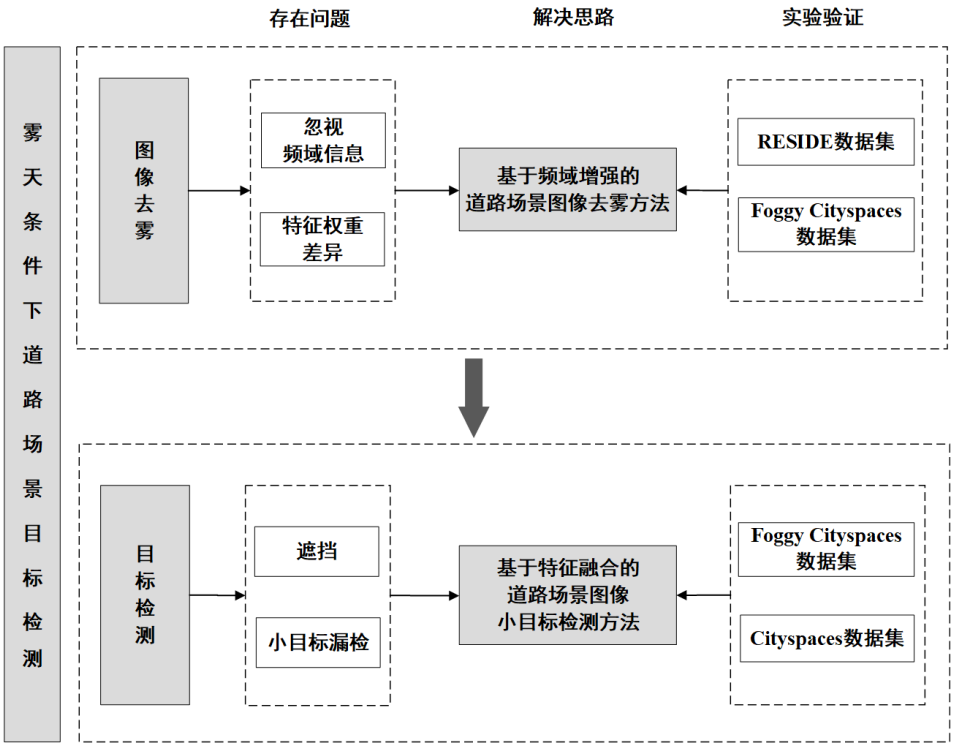


图 1-5 本文研究框架

Fig 1-5 The Research Framework of This Paper

1.4 章节组织架构

本文每一章的研究内容如下：

第一章：绪论。介绍了雾天条件下道路场景目标检测的研究背景与意义，阐述了国内外图像去雾和目标检测的研究现状，介绍了论文的研究内容和结构安排。

第二章：相关理论与相关工作。介绍了雾天图像退化机理，包括雾气的形成、雾天图像退化原因以及特性，介绍了单阶段目标检测算法理论的代表性算法 YOLO 算法中 YOLOv1、YOLOv5、YOLOv8、YOLOv11。

第三章：基于频域增强的道路场景图像去雾方法。首先介绍了基于频域增强的道路场景图像去雾模型；其次介绍了组架构模块、特征注意模块，以及频域增强感知模块；然后介绍了实验环境与评价指标、数据集介绍与预处理、以及消融实验和可视化结果分析；最后进行章节总结。

第四章：基于特征融合的道路场景图像小目标检测方法。首先介绍了去雾道路场景图像的小目标检测模型，其次介绍了 Focus 模块、多尺度特征融合模块、特征金字塔模块；然后介绍了实验环境与评价指标、数据集介绍与预处理、以及消融实验和可视化结果分析；最后进行章节总结。

第五章：总结与展望。简要总结了本文的主要工作，分析了现有的工作存在的问题，展望了未来工作方向。

第 2 章 相关理论及相关工作

2.1 图像去雾算法理论

2.1.1 雾天图像退化机理及特性

雾气由接近地球表面大气中悬浮的微小液态水滴或固态冰晶颗粒构成，其形成需要充足的水汽、冷却作用和适宜的静风或微风条件。雾气的形成过程始于充足的水汽积累，水汽主要来源于海洋、湖泊、河流等水域的蒸发以及植物的蒸腾作用，当气温降至露点温度时，空气中的水汽会凝结成微小的水滴或冰晶，从而导致地面水平能见度降低。夜间地面辐射冷却会形成的辐射雾，多发于秋冬晴夜后的清晨；冷空气接触暖湿气流则形成伴随降温的锋面雾；沿海地区因水汽丰沛易现平流雾，水体表面低温环境还能生成蒸汽雾；山谷、盆地等低洼地区空气流动受限，水汽易积聚形成山谷雾或盆地雾，而山地迎风坡则因气流抬升作用易形成上坡雾。雾气的浓度与降温幅度、水汽饱和程度直接相关，深厚逆温层可形成持续数日的浓雾，其影响不仅限于能见度下降导致的交通事故风险，还会引发监控设备成像模糊等次生问题。

雾天图像退化的主要原因归结为大气中悬浮颗粒对光线的散射、吸收以及折射作用，不仅减弱了目标物体反射光线的强度，还改变了光线的传播方向，导致成像设备接收到的光线变得模糊且对比度降低。例如光线散射，当光线穿过含有悬浮颗粒的大气时，部分光线会被悬浮颗粒散射，导致成像设备接收到的光线强度减弱且方向性变得模糊，使得图像中的细节信息变得难以分辨，降低了图像的清晰度和对比度。除了目标物体反射的光线外，周围环境中的大气光也会对成像产生干扰，导致图像中出现杂散光，进一步降低了图像的清晰度和对比度。同样还有较为明显的距离衰减，随着目标与成像设备之间距离的增加，光线在传播过程中的衰减也会加剧，导致远处物体的图像变得更加模糊，细节信息丢失严重。

雾天图像退化的特性可以从以下几个方面进行阐述：首先对比度和亮度降低，由于光线散射和吸收作用，图像整体对比度和亮度都会降低，这使得图像中的细节信息变得难以分辨，整体视觉效果变差；其次色彩失真，由于不同波长的光线被不同程度地散射和吸收，特别是蓝色和紫色等短波长的光线更容易被吸收，导

致图像中的色彩会发生偏移或失真；接下来还有细节模糊，由于光线散射和折射作用，图像中的细节信息会变得模糊，模糊程度会随着目标与成像设备之间距离的增加而加剧；最后还有空间分辨率下降，由于上述多种因素的影响，雾天图像的空间分辨率会显著下降，使得图像中的物体边缘变得模糊，难以准确识别[36]。

2.1.2 图像去雾算法频域理论

在图像去雾领域，现有的多数算法主要聚焦于空间域信息的处理，通过对图像像素的直接操作来去除雾气，提升图像的清晰度和对比度，这往往忽视了频域特征在图像去雾中的重要作用，从而引发了一系列如颜色失真以及伪雾残留的问题，而根源在于空间域与频域在处理图像信息时的本质差异，以及频域特征在揭示图像内在结构和细节方面的独特优势。空间域处理主要侧重于图像的局部特征，通过增强对比度、调整亮度等手段来改善图像的视觉效果，但在处理雾气覆盖的图像时，往往难以准确区分雾气与真实景物之间的界限，导致去雾后的图像在细节上有所损失。例如在雾气较重的区域，空间域算法会过度增强噪声或者在去除雾气的同时模糊了图像的边缘和纹理信息，使得复原后的图像显得模糊或缺乏真实感。此外，空间域处理还会因雾气对光线的影响，如亮度变化、色彩饱和度和色调的改变导致颜色失真。相比之下，频域处理则能够揭示图像在灰度变化上的不同程度，即图像的梯度大小，从而提供更多关于图像整体结构和细节的信息。在频域中，频域特征与空域先验具有强互补性，并且可以通过频域照分量，即低频分量则反映了图像的背景或平滑区域；高频分量对应反射分量，即图像的高频分量对应于图像的边缘、纹理等细节信息，因此可以通过雾气为图像的低频成分而景物的边缘和细节则对应于高频成分，从而更准确地识别出雾气覆盖的区域与真实景物之间的界限。因此，融合频域特征的去雾算法能够更有效地保留图像的细节信息，同时在去除雾气的同时保持图像的自然色彩和对比度。

在去雾算法中，快速傅里叶变换（Fast Fourier Transform, FFT）的引入为频域处理提供了强有力的支持，FFT 通过将图像信息从空间域转换到频域，揭示了图像在频域上的独特特征，从而实现了对图像信息的更深入处理。首先，FFT 能够准确地将图像从空间域转换到频率域，使得图像中的不同频率成分得以分离。雾气通常表现为低频成分的增加，而景物的边缘和细节则对应于高频成分，通过 FFT，可以清晰地观察到这些频率分量的分布情况，为后续的去雾处理提供了重

要的依据。其次，FFT 在去雾算法中实现了对图像频率成分的精确处理。在识别出雾气区域后，可以利用 FFT 对图像特征信息进行针对性的处理，通过抑制低频分量来减少雾气的影响，同时增强高频分量以保留和恢复图像的细节信息，不仅避免了空间域处理中常见的细节损失和颜色失真问题，还能够更有效地去除雾气，提升复原图像的质量和视觉效果。最后由于 FFT 能够同时处理图像的幅度和相位信息，因此在去雾过程中能够保持图像的真实性和色彩平衡，避免了空间域处理中可能出现的颜色失真和伪雾残留，使得复原后的图像更加自然和逼真。FFT 在去雾算法中的频域处理中提升了图像频率成分的处理速度，有效降低了计算复杂度，为去雾算法的高效运行奠定了坚实基础；其次，FFT 确保了图像中不同频率成分的精准识别与处理，在有效去除雾气的同时完美保留了图像的细节与色彩信息；再者，FFT 可根据具体去雾需求灵活选择，具有丰富的频域滤波手段，能够实现图像的精细处理。

快速傅里叶逆变换（Inverse Fast Fourier Transform, IFFT）是 FFT 的逆运算，将频域表示的图像信息转换回空间域，接下来使用一个简单的三维张量 A 来说明 FFT 和 IFFT 的工作原理。FFT 对张量 A 的每一个前向切片转换到频域，A 为 $3\times3\times4$ 的三维张量，如式（2-1）所示，每个元素都用 a_{ijk} 表示，其中 i 和 j 分别是第一个和第二个维度的索引，k 是第三个维度的索引，设 $A(k)$ 中的元素都是实数。

$$A=\left[\begin{bmatrix} a_{111} & a_{112} & a_{113} \\ a_{121} & a_{122} & a_{123} \\ a_{131} & a_{132} & a_{133} \end{bmatrix},\begin{bmatrix} a_{112} & a_{112} & a_{113} \\ a_{122} & a_{122} & a_{123} \\ a_{132} & a_{132} & a_{133} \end{bmatrix},\begin{bmatrix} a_{113} & a_{112} & a_{113} \\ a_{123} & a_{122} & a_{123} \\ a_{133} & a_{132} & a_{133} \end{bmatrix},\begin{bmatrix} a_{114} & a_{112} & a_{113} \\ a_{124} & a_{122} & a_{123} \\ a_{134} & a_{132} & a_{133} \end{bmatrix}\right] \tag{2-1}$$

$\hat{A}(k)$ 为经过 FFT 处理过的前向切片，[ ] 表示默认沿所有维度进行 FFT，而 3 明确指定了沿着第三个维度进行 FFT。

$$\hat{A}(k)=FFT\left(A(k),[\ ],3\right) \tag{2-2}$$

$$\hat{A}(1)=\begin{bmatrix} \hat{a}_{111} & \hat{a}_{112} & \hat{a}_{113} \\ \hat{a}_{121} & \hat{a}_{122} & \hat{a}_{123} \\ \hat{a}_{131} & \hat{a}_{132} & \hat{a}_{133} \end{bmatrix} \tag{2-3}$$

FFT 对每个前向切片的处理完成后，将这些前向切片重新组合成张量 $\tilde{A}$，通过 IFFT 将前向切片转换回空间域。$\tilde{A}$ 为 $3\times3\times4$ 的张量，保留了原始张量 A 的形状，

但其每个前向切片都已经过处理，并且已经被转换回空间域。

$$\tilde{A} = IFFT\left(\hat{A},[\ ],3\right) \tag{2-4}$$

$$\tilde{A}(1) = \begin{bmatrix} \tilde{a}_{111} & \tilde{a}_{112} & \tilde{a}_{113} \\ \tilde{a}_{121} & \tilde{a}_{122} & \tilde{a}_{123} \\ \tilde{a}_{131} & \tilde{a}_{132} & \tilde{a}_{133} \end{bmatrix} \tag{2-5}$$

2.2 单阶段目标检测算法理论

两阶段目标检测模型包含候选区域生成这一步骤，该过程往往需要消耗大量时间，而单阶段目标检测算法通过端到端的网络架构，在单一前向传播中完成目标检测任务，无需生成候选区域，通过密集的网格或锚框直接预测目标的类别和位置，从而大大提高了检测速度。单阶段目标检测算法的代表算法为 SSD 和 YOLO 系列算法。本节接下来将主要介绍 YOLO 系列算法。

YOLO 网络是一种基于深度学习的端到端目标检测算法，主要由卷积层、全连接层和输出层组成。特征提取部分采用小卷积核的卷积层，配合批量归一化（Batch Normalization, BN）和非线性激活函数来增强特征表达能力和网络稳定性；特征映射部分通过高维全连接层将特征转换为目标的位置和类别概率；输出层则使用多通道结构，每个通道分别对应不同的类别预测或位置参数估计。YOLO 网络的目标检测过程首先将输入图像划分成 $S \times S$ 个网格单元；其次 YOLO 网络预测每个网格单元 BBox 个数以及置信度，每个 BBox 由四个参数组成，分别是中心坐标（x, y）、宽度（w）和高度（h），置信度表示包含目标的可能性以及预测的准确性。接下来对每个 BBox，预测 C 个类别概率，表示该 BBox 属于不同类别的可能性。最后得到所有的预测结果后，使用非极大值抑制（Non-Maximum Suppression, NMS）算法去除冗余的 BBox，只保留最有可能包含目标的 BBox。

（1） YOLOv1 算法

YOLOv1 是单阶段目标检测算法的开山之作，YOLOv1 的网络结构主要包括输入层、卷积层、全连接层以及输出层，YOLOv1 的网络结构图 2-1 所示。输入层是由输入图像经过裁剪、归一化等方式处理得到的数据，因为目标检测需要更加精细的信息，因此将输入图像的大小固定为 448×448。YOLOv1 网络中包含

了 24 个卷积层，其中包括3×3和1×1两种卷积，主要提取输入层处理的特征信息。网络结构最后两层是全连接层，用于回归框的坐标、置信度以及分类概率。输出层是一个7×7×30的张量，表示将图像划分为7×7的网格，每个网格预测 2 个 BBox，每个 BBox 包含 4 个坐标值、1 个置信度和 20 个类别的概率，因此总共是$7\times7\times\left(2\times\left(4+1\right)+20\right)=7\times7\times30$的三维矩阵。YOLOv1 算法的结构相对简单，取得了较快的推理速度，但由于网络结构的限制，容易出现漏检和小目标的检测精度不够理想。

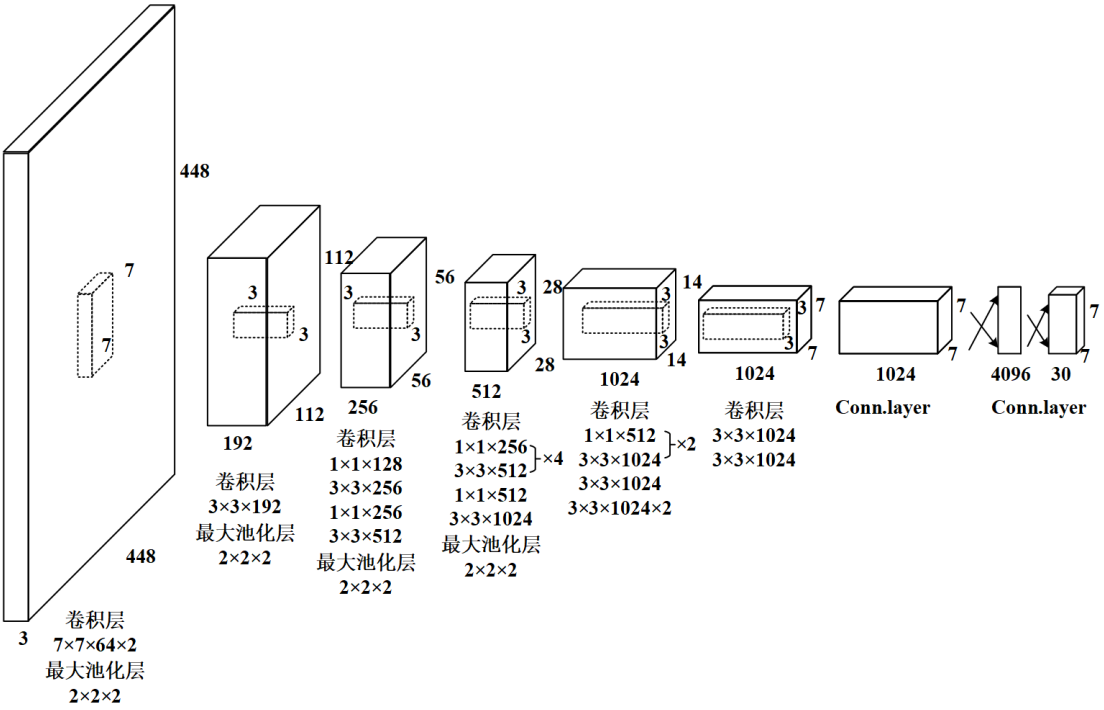


图 2-1 YOLOv1 网络结构图

Fig 2-1 YOLOv1 Network Structure

（2） YOLOv5 算法

YOLOv5 检测框架包含四个不同规模的子版本，包括 YOLOv5x、YOLOv5l、YOLOv5m 以及 YOLOv5s，其中 YOLOv5s 是 YOLOv5 系列中深度最小、特征图宽度最小的网络，因此为检测速度最高的网络，其余版本均在此基础上进行了深度和宽度的扩展。YOLOv5 在训练过程中采用标准的图像填充策略，将输入图片等比例缩放至固定尺寸，而在测试阶段则运用了优化的填充方案以提升运算效率，有效降低了冗余信息的影响。YOLOv5 网络由三部分组成：骨干网络、颈部网络和检测头，网络结构图 2-2 所示。骨干网络使用 CSPDarknet53 结构，由 CBS（Cord-Bottleneck-Short, CBS）模块、C3 模块和 SPPF（Spatial Pyramid Pooling –

Fast，SPPF）模块组成，主要对输入的图像提取特征信息。颈部网络采用特征金字塔网络（Feature Pyramid Network，FPN）和路径聚合网络（Path Aggregation Network，PANet）相结合的融合策略，首先通过上采样与更粗粒度的特征图依次融合，构建了自顶向下的特征金字塔结构，再通过下采样与上一层特征图依次融合，形成了自底向上的特征金字塔结构，二者相结合将不同主干层的特征组合在一起，进而提高了对不同尺寸目标的检测效果。检测头由损失函数和预测框筛选函数组成，输入经过处理后的特征来更好地进行分类和定位。YOLOv5 作为单阶段目标检测的经典算法，凭借其高效性和易用性被广泛采用，但存在着模型结构固化和复杂背景或密集场景中小目标易出现漏检或误检的问题。

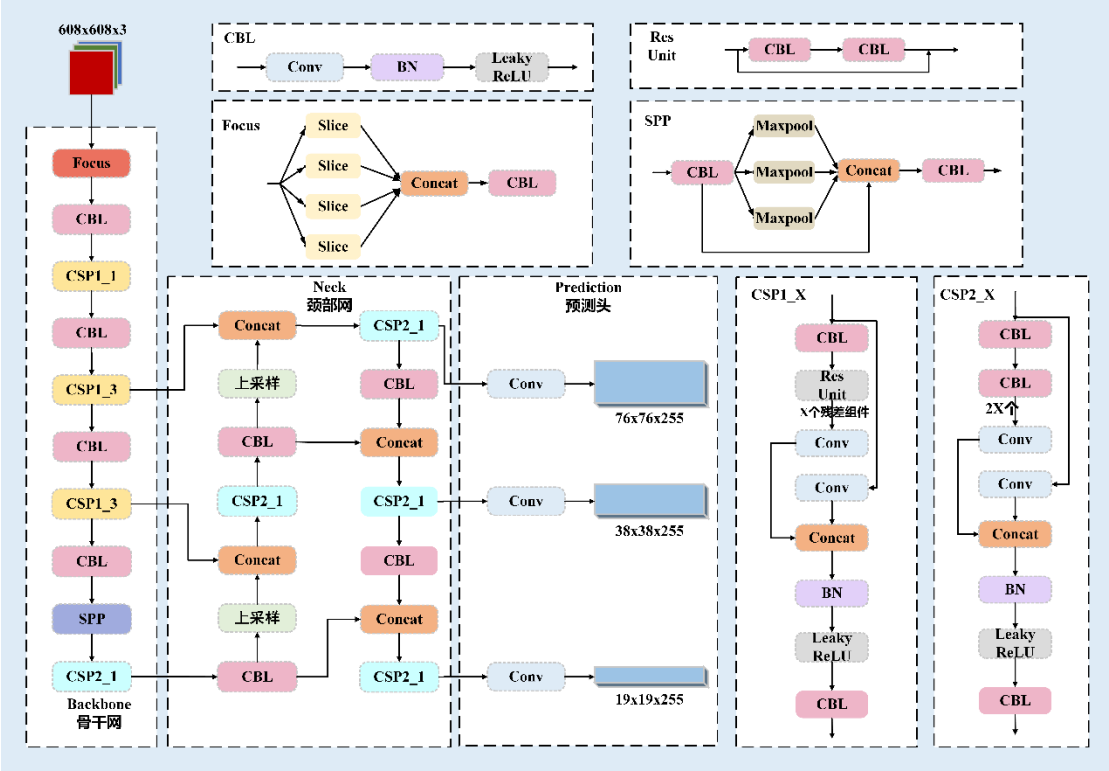


图 2-2 YOLOv5 网络结构图

Fig 2-2 YOLOv5 Network Structure

（3） YOLOv8 算法

YOLOv8 在 YOLOv5 的基础上进行了扩展，使用了无锚点检测加速 NMS 的后处理过程，能够快速精度识别和定位图像和视频中的物体，并处理图像分类和实例分割等任务，网络结构图 2-3 所示。YOLOv8 系列根据网络结构的复杂度差异，提供了 YOLOv8n、YOLOv8s、YOLOv8m、YOLOv8l、YOLOv8x 一共五种递增的模型大小，以适应不同场景需求。骨干网络部分采用 C2f 模块，相较于

YOLOv5 的 C3 模块，通过优化架构设计显著降低了参数冗余，提高了计算效率。为进一步增强特征提取能力，YOLOv8 整合了深度可分离卷积（Depthwise Separable Convolution, DWConv）和膨胀卷积（Dilated Convolution, DConv）。在检测头设计上，YOLOv8 摒弃了传统的耦合结构（Couple Head），转而采用解耦式（Decoupled Head）设计，有助于提升模型在复杂环境下的检测性能，同时通过引入无锚点机制，使模型能够更加灵活应对各种尺度和形态的目标检测任务。尽管 YOLOv8 在目标检测性能与速度平衡方面表现优异，但仍存在小目标或高密度重叠物体的检测精度较低以及模型复杂度较高的问题。

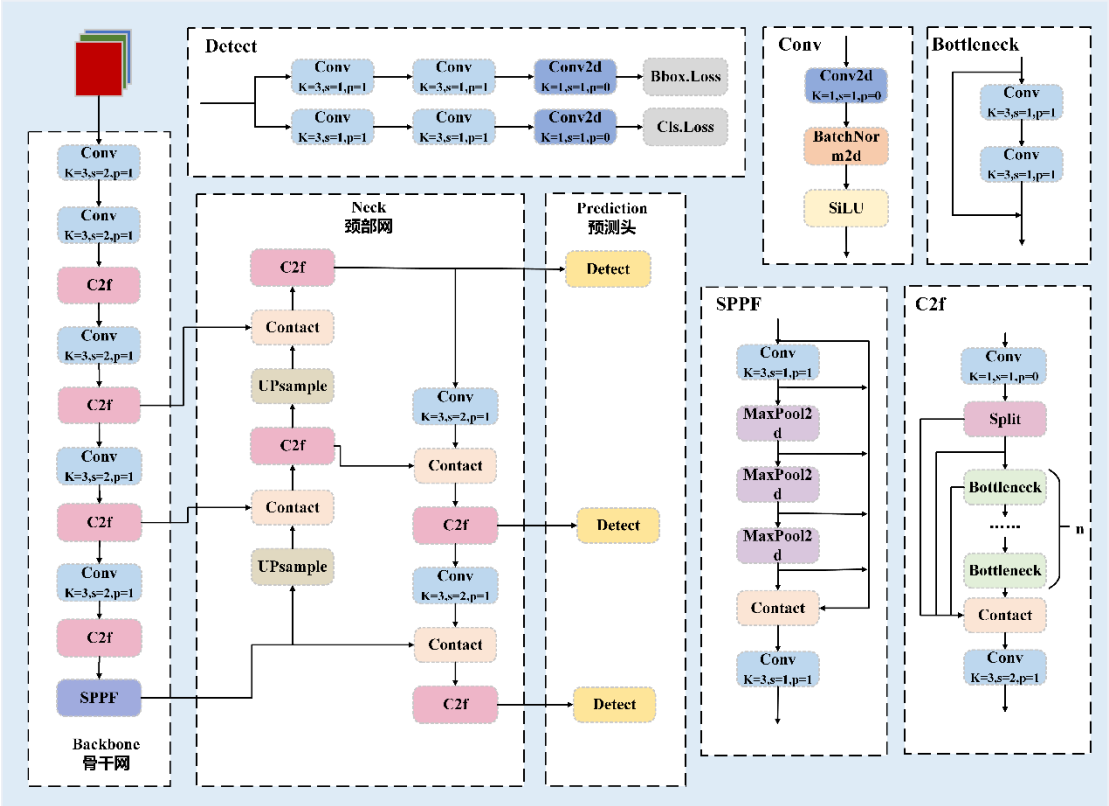


图 2-3 YOLOv8 网络结构图

Fig 2-3 YOLOv8 Network Structure

（4） YOLOv11 算法

YOLOv11 在之前 YOLO 版本的基础上引入了 C3k2 和 C2PSA（Cross Stage Partial with Pyramid Squeeze Attention, C2PSA）两个全新模块，并延续了无 NMS 的训练策略，实现了端到端的目标检测能力，进一步提高了性能和灵活性，相较于 YOLOv8，YOLOv11 在网络结构上进行了三点改进，网络结构图 2-4 所示。首先骨干网络部分使用 C3K2 模块来处理不同阶段的特征提取，C3K2 模块的结构集成了 C2f 和 C3 模块的组合，外层由 C2 结构主导，内层是一个 C3 的结构，

代码中通过设置 C3k 参数为 True 或 Fa1se 来决定其具体行为，这种可切换的设计使得模型能够根据不同的需求和场景选择合适的特征提取方式，增强了模型的灵活性和适应性，并且有效的减少了参数量；其次在 SPPF 之后，加上了一层 C2PSA 模块，结合了点式空间注意力（Pointwise Spatial Attention, PSA）结构块，通过在标准 C2f 模块中引入 PSA 结构块，能够选择性地关注输入特征中的重要部分，抑制不重要的信息，从而提高了模型对重要特征的捕捉能力，提高检测的准确性；最后检测头中卷积层更换成了 DWConv，能够减少输入通道数量，实现高精度的运算。

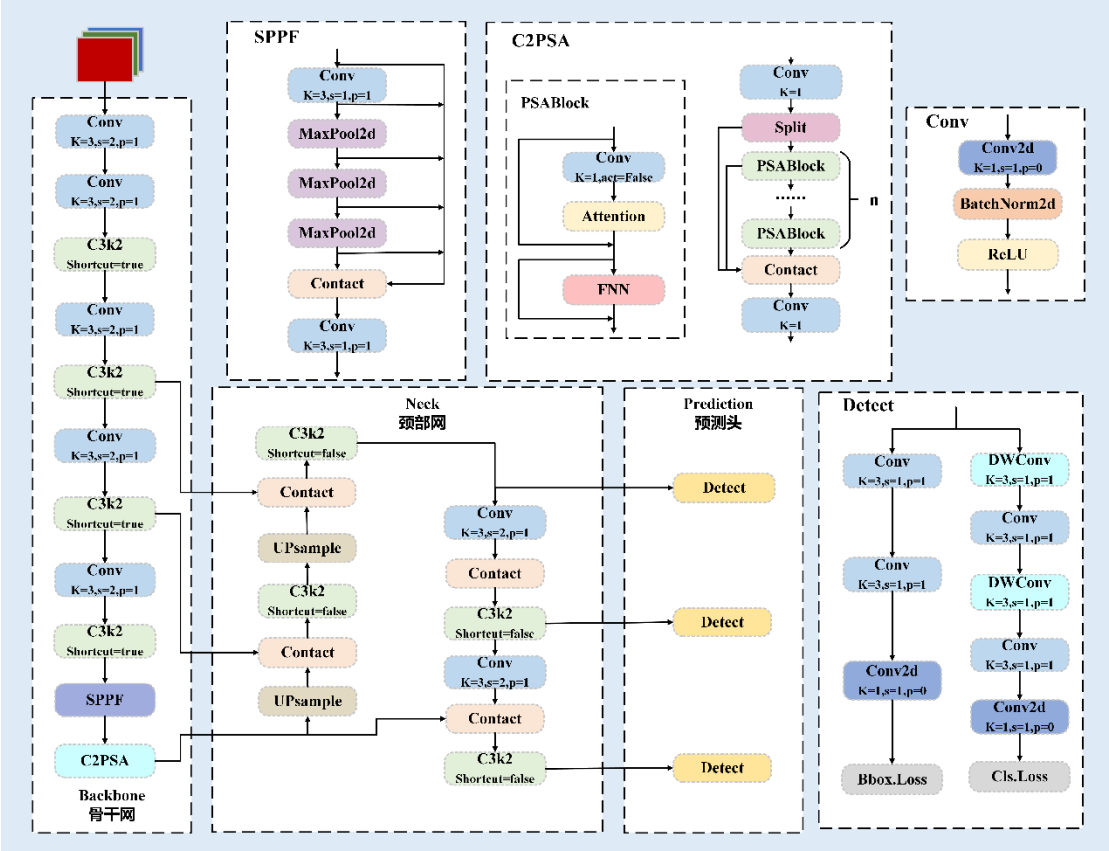


图 2-4 YOLOv11 网络结构图

Fig 2-4 YOLOv11 Network Structure

2.3 本章小结

本章首先介绍了图像去雾算法理论，包括雾天图像退化机理以及特性和图像去雾算法频域理论，介绍了 FFT 以及 IFFT 的工作原理，接下来介绍了单阶段目标检测算法理论，并展开描述了 YOLO 系列的算法结构等。

第 3 章 基于频域增强的道路场景图像去雾方法研究

3.1 引言

由于雾天空气中微粒会对光线产生吸收、辐射和散射作用，会导致许多特征信息被覆盖或模糊，物体边缘的对比度显著降低，图像远景物体的细节被模糊化或完全丢失，图像色彩信息失真等问题，因此图像去雾是视觉领域的重要研究方向，旨在改善被大气条件影响而模糊或噪声化的图像。

针对不同图像像素上雾气不均匀分布，但现有的去雾方法主要对图像中的空间特征信息做相同处理而忽略频域信息，从而导致复原图像存在颜色失真和伪雾残留的问题，本章提出基于频域增强的道路场景图像去雾方法，即 FFT-Net 算法。通过 FFT 将特征信息分解到空间域和频域中，从而可以更直观地观察特征信息的频率成分，进而实现对特征信息的精确分析和处理，提高模型的去雾效果。本章所设计的去雾模型包括两大模块，分别为组架构和特征注意模块，其中组架构中频域增强感知模块对图像特征信息进行分解，能够更好捕捉到图像的纹理和结构信息，将图像特征信息在空间域与频域分别进行处理，有效提高了图像的清晰度；特征注意模块结合了通道注意（Channel Attention, CA）模块和像素注意（Pixel Attention, PA）模块，能够更好捕捉和强调图像中不同尺度和不同空间位置的重要特征，增强了特征信息处理的灵活性。为了证明本章去雾方法的优越性，在 RESIDE 数据集和 Foggy Cityspaces 数据集进行实验验证，证明了本章方法有效地提高了图像去雾效果，在感知上看起来也接近参考基础事实。

3.2 基于频域增强的道路场景图像去雾模型构建

在本章中，主要介绍了基于频域增强的道路场景图像去雾方法，构建了结合组架构和特征注意模块的图像去雾网络模型，如图 3-1 所示。图像去雾网络模型是由以下模块结合而成，首先输入模糊有雾图像，经过一个的卷积层，接着先后经过三个组架构和前 3 个组架构输出结果的通道组合输出融合特征图，然后再经过由 CA 模块和 PA 模块组成的特征注意模块，并通过两个卷积层，最后进行相加运算后输出清晰图像。针对现有的去雾算法主要利用图像中的空间信息而忽视

频域信息的问题，改进的图像去雾网络模型可以将不同特征信息赋予不同的权重并将特征信息转换到频域空间，以便于网络达到更好的实验结果。其一在组架构中加入了 SE-Net（Squeeze-and-Excitation Networks, SE-Net）[37]网络详解，该结构块可以提高有效特征信息的权重，更好的捕捉图像中需要处理的浓雾以及高频细节等信息。其二在基础模块中设计了频域增强感知模块[38]，该模块可以将图像特征转换到傅里叶空间，通过频域来处理图像信息，从而更好地捕捉图像的频域特征。

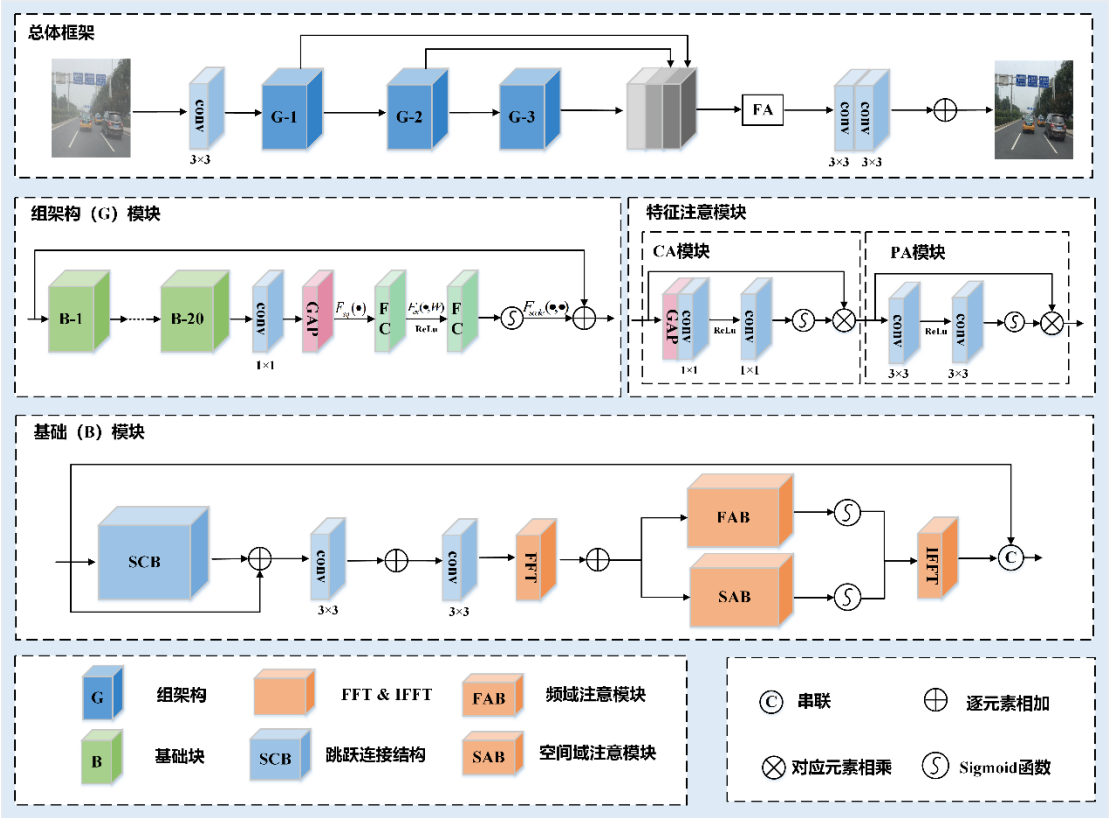


图 3-1 基于频域增强的道路场景图像去雾网络模型

Fig. 3-1 Frequency-Domain Enhancement-Based Road Scene Image Dehazing Network Model

3.2.1 组架构模块

每个组架构由 20 个基础模块和 SE-Net 结构块组成，如图 3-2 所示。每个基础块由跳跃连接结构和频域增强感知模块组成，连续的基础块增加了网络模型的深度和表现力，跳过连接结构由多个局部残差学习（Local Residual Learning, LRL）连接，可以增加信息的流动和减少经过深层网络后图像细节的丢失。

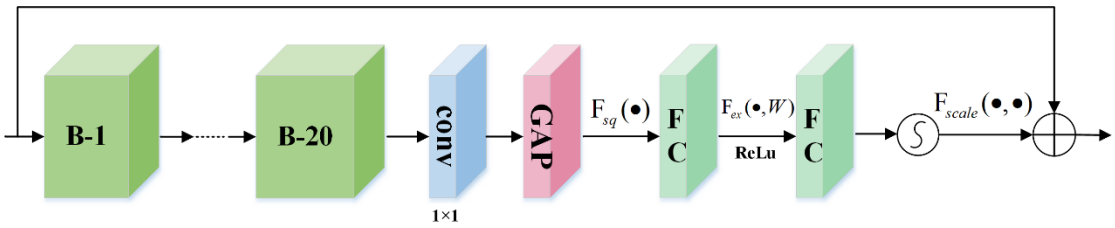


图 3-2 组架构模块

Fig. 3-2 Group Architecture Module

SE-Net 可以自主学习每个特征通道的重要程度并将图像中不同比重的特征信息进行不同的处理，因此可以更有效的处理图像中浓雾与高频细节等信息，如图 3-3 所示。首先特征经过一个卷积层，再通过 GAP 对特征图进行 Squeeze 操作得到特征向量，紧接着经过第一个全连接层（Fully Connected Layer, FC），减少参数量，然后通过 ReLU 激活函数并进行 Excitation 操作和 Reweight 操作，到达第二个 FC 层，得到带有注意力机制的权重参数，最后经过 Sigmoid 函数，并通过 Scale 操作并输出结果。两个 FC 层可以更好的融合各通道的特征信息，同时 ReLU 激活函数使 FC 层输出结果具有更多的非线性。

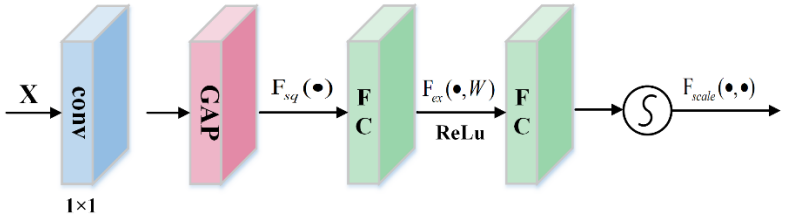


图 3-3 SE-Net 结构块

Fig. 3-3 Structure of SE-Net

Squeeze 操作：通过通道的全局平均池化，将包含全局信息的$C\times H\times W$的特征图直接压缩成一个$C\times 1\times 1$的特征向量，可以更好的利用通道间的相关性。

$$F_{sq}=\frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W}u_c\left(i,j\right) \tag{3-1}$$

式（3-1）中$u_c\left(i,j\right)$为第c个通道u_c在$\left(i,j\right)$位置的值。

Excitation 操作：基于特征通道间的相关性，每个特征通道生成一个权重，用来代表特征通道的重要程度。由原本全为白色的 C 个通道的特征，得到带有不同深浅程度颜色的特征向量，即不同的重要程度。

$$F_{ex}\left(F_{sq},W\right)=\delta\left(W_2\delta\left(W_1F_{sq}\right)\right) \tag{3-2}$$

首先通过第一个 FC 层，W_1的维度是$\frac{C}{r}\times C$，r为缩放参数，经过 ReLU 激活函

数，输出维度不变，然后再经过第二个 FC 层，和W_2相乘，W_2的维度是$C\times\frac{C}{r}$，因此输出的维度就是$C\times1\times1$。两次全连接不完全相同有利于减少通道个数从而降低计算量，并在一定程度上防止网络模型过拟合。

Reweight 操作[39]：将 Excitation 输出的权重看做每个特征通道的重要性，完成对原始特征的重校准。

Scale 操作[40]：将前面得到的归一化权重加权到每个通道的特征上。式（3-3）中s_c为权重数。

$$F_{scale}=s_c\times u_c \tag{3-3}$$

3.2.2 特征注意模块

特征注意模块由 CA 模块和 PA 模块组成，如图 3-4 所示。特征注意模块有差别的处理不同的特征的有雾区域和像素区域，可以拥有更多的灵活性。

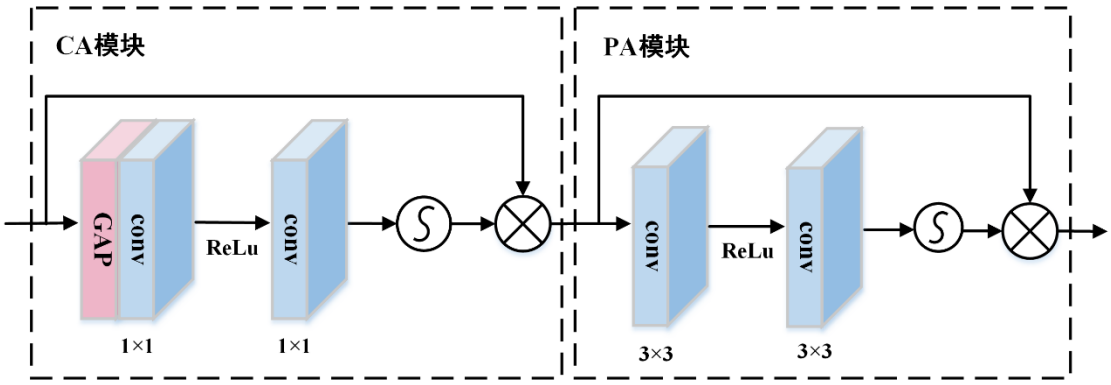


图 3-4 特征注意模块

Fig. 3-4 Feature Attention Module

CA 模块主要侧重于通过动态调整每个通道的重要性，使模型能够更有效地利用输入数据的信息。因为全局平均池化层（Global Average Pooling, GAP）不需要参数，同时因其对空间信息进行求和，因此达到减少参数量和计算量，减少过拟合的效果。所以 CA 模块中选择先通过 GAP 将通道全局空间信息转化为通道描述特征图，随后依次通过卷积层、ReLU 激活函数，最后经过 Sigmoid 函数将值映射到 0-1 的范围内，并将通道 CA 的权值与最开始的特征F_c进行对应元素相乘，可以得到不同通道的不同权值。式（3-4）中$X_c(i,j)$为第c个通道X_c在(i,j)位置的值，H_p为全局池化函数。形状为$C\times H\times W$的特征图变为$C\times1\times1$，σ为

Sigmoid 激活函数，δ 为 ReLU 函数[41]。

$$g_c = H_p(F_c) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} X_c(i, j) \tag{3-4}$$

$$CA_c = \sigma\left(conv\left(\delta\left(conv\left(g_c\right)\right)\right)\right) \tag{3-5}$$

$$F_c^* = CA_c \times F_c \tag{3-6}$$

PA 模块则是使去雾模型相比于薄雾和大块的颜色或阴影等低频细节信息，更关注于需处理的区域的信息特征，如浓雾区域和高频细节等区域。将输入 F^* 通过 ReLU 激活函数直接输入到两个卷积层中，同 CA 模块一样，经过 Sigmoid 激活函数将值映射到 0-1 的范围内，并将通道 PA 的权值与最开始的特征 F^* 进行对应元素相乘。形状为 $C \times H \times W$ 的特征图变为 $1 \times H \times W$，式（3-8）中 F^* 为 CA 模块的输出，$\hat{F}$ 表示 FA 模块的输出。

$$P_{PA} = \sigma\left(conv\left(\delta\left(conv\left(F^*\right)\right)\right)\right) \tag{3-7}$$

$$\hat{F} = F^* \times P_{PA} \tag{3-8}$$

3.2.3 频域增强感知模块

频域增强感知模块是使用 FFT 来去除特征图上一些噪声影响，加强注意力关注的区域，如图 3-5 所示。本模块可以将前面得到的特征去噪后再和最开始的特征相加。输入特征图像经过一个卷积层操作后得到特征图，再经过 FFT 获得运算后新的特征图，通过运算得到可学习的矩阵，将特征结果中的噪声去除，再通过频域上的傅里叶去噪，和空间域上的傅里叶去噪后使用 Sigmoid 激活函数将值映射到 0-1 的范围内，接着把两个特征进行合并，最后再通过 IFFT 并输出特征图像。

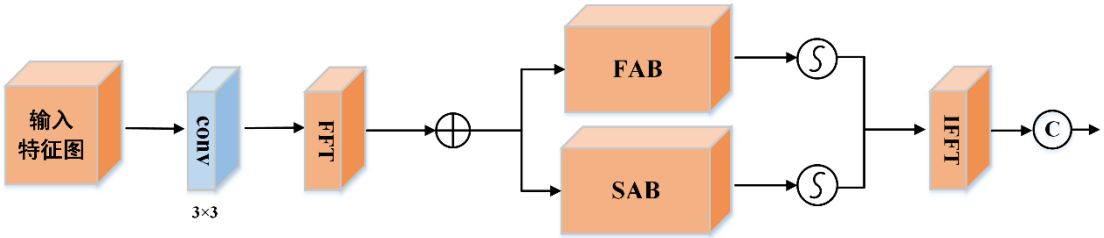


图 3-5 频域增强感知模块

Fig. 3-5 Frequency-Domain Enhancement-Aware Module

由于 FFT 可以用有限的频谱空间来表征比较复杂的时域空间，因此可有效

降低计算量；同时傅里叶变换建构出图像特征分别在空间域和频域上的联系，能够更好地捕捉到图像的纹理和结构信息；另外通过在傅里叶空间中对特征进行处理，可以更有效地去除图像中的噪声，提高图像质量和清晰度；该模块还可以灵活应用于不同场景的图像去雾任务中。

使用 FFT 后，计算复杂度从 N^2 降低到 $\log_2 N$。傅里叶变换将图像转换成实分量和虚分量的组合，虚分量是图像在频域中的描述。将图像作为 FFT 的输入，频域中的频率数与空间域中的像素数相同。如上式所述，其中 $0 \le m \le M-1, 0 \le n \le N-1$，坐标 (m,n) 处的像素由 $f(m,n)$ 给出，则坐标 x 和 y 对应的图像在频域的值为 $F(x,y)$，m 和 n 为图像的比例。执行该算法的成本非常高，但 FFT 可以自主的将二维变换转换为两个一维变换方程，一个在水平方向上，另一个在垂直方向上，最终的结果是在频率空间中执行二维变换，FFT 将 N 个点的变换表示为 $\frac{N}{2}$ 个变换的总和，与其他方法相比，具有降低计算效率和节省时间的优点。

$$F(x,y)=\sum_{m=0}^{M-1}\sum_{n=0}^{N-1} f(m,n) e^{-j2\pi\left(x\frac{m}{M}+y\frac{n}{N}\right)} \tag{3-9}$$

$$f(m,n)=\frac{1}{MN}\sum_{x=0}^{M-1}\sum_{n=0}^{N-1} F(x,y) e^{j2\pi\left(x\frac{m}{M}+y\frac{n}{N}\right)} \tag{3-10}$$

$$F(m,n)=\sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f(x,y) e^{-j2\pi x\frac{m}{M}} e^{-j2\pi y\frac{n}{N}} \tag{3-11}$$

3.3 实验设计与结果分析

3.3.1 实验环境与评价指标

本章中研究的目标检测的硬件环境为：NVIDIA GeForce RTX3090 GPU，24GB 显存，Intel 酷睿(Core)i9-14900KF@3.20GHz CPU，128GB 内存。软件环境为:Windows 10 操作系统，CUDA11.8，深度学习框架为 Pytorch2.0.0，编程语言 python 3.10.16。

PSNR 作为图像质量评估的重要量化指标，通过数学计算来反映两幅图像之间的视觉差异程度，常用于评估压缩算法对图像质量的影响，衡量生成对抗网络等模型输出结果与真实样本之间的相似度等等。不过会时常出现评价结果和人的

主观感觉有差异的情况，这是因为每个人的眼睛对视觉误差的敏感度确实会受到多种因素的影响，包括但不限于以下几个方面：不同颜色的对比度和色差对视觉误差的感知有影响；视觉对不同空间频率的图像敏感度不同，如高频细节和低频细节对视觉误差的影响不同；视觉背景的复杂程度也会影响对误差的感知；不同个体的视力、视觉敏感度、色觉和疲劳程度等生理差异也会影响对视觉误差的感知等。多种因素的综合作用使得人眼对视觉误差的感知不是绝对的，而是具有很大的主观性和情境依赖性。PSNR 的数值通常以分贝（dB）为单位，数值越高表示两幅图像之间的差异越小，图像质量越高。PSNR 常用于评估图像压缩算法、图像去雾算法以及其他图像处理技术的效果。

对于单幅图像来说，均方误差（Mean-Square Error, MSE）定义为式（3-12），PSNR 定义为式（3-13）：

$$M_{MSE}=\frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}\left[I\left(i,j\right)-K\left(i,j\right)\right]^{2} \tag{3-12}$$

$$P_{PSNR}=10\times\log_{10}\left(\frac{MAX_I^2}{M_{MSE}}\right)=20\times\log_{10}\left(\frac{MAX_I}{\sqrt{M_{MSE}}}\right) \tag{3-13}$$

式（3-12）中 I 和 K 分别表示原始图像和处理后的图像，m 和 n 是图像的宽度和高度，式（3-13）中 MAX_I 是图像像素值的最大可能值。

SSIM 为用于评估两幅图像相似度的指标，既能反映原始图像与经过各种操作后的图像之间的差异程度，也能有效评估人工智能系统生成的图像与真实场景的接近程度。SSIM 考虑了人眼对图像的感知特性，更符合人眼的直观感受。一般可以使用加权平均的方式来计算不同特征的占比，设亮度占比$\left(\alpha\right)$，结构占比$\left(\beta\right)$，对比度占比$\left(\gamma\right)$，SSIM 定义为式（3-14）：

$$S\left(x,y\right)=l\left(x,y\right)^{\alpha}\times c\left(x,y\right)^{\beta}\times s\left(x,y\right)^{\gamma} \tag{3-14}$$

SSIM 主要考量图像的三个关键特征：

（1）亮度：通过计算图像的平均灰度值，灰度值代表图像中每个像素的亮度水平，通过平均所有像素的值得到式（3-15），对比函数为式（3-16）：

$$\mu_x=\frac{1}{\mathrm{N}}\sum_{i}^{N}x_i \tag{3-15}$$

$$l\left(x,y\right)=\frac{2\mu_x\mu_y+C_1}{\mu_x^2+\mu_y^2+C_1} \tag{3-16}$$

（2）对比度：对比度通过灰度标准差来衡量。

$$\sigma_x=\left(\frac{1}{N-1}\sum_{i=1}^{N}\left(x_i-\mu_x\right)^2\right)^{\frac{1}{2}} \tag{3-17}$$

$$c\left(x,y\right)=\frac{2\sigma_x\sigma_y+C_2}{\sigma_x^2+\sigma_y^2+C_2} \tag{3-18}$$

（3）结构：通过归一化的图像对比来衡量局部的结构相似性，反映了两幅图像在局部结构上的一致性。可以用相关性系数衡量：

$$\sigma_{xy}=\frac{1}{N-1}\sum_{i=1}^{N}\left(x_i-\mu_x\right)\left(y_i-\mu_y\right) \tag{3-19}$$

$$s\left(x,y\right)=\frac{\sigma_{xy}+C_3}{\sigma_x\sigma_y+C_3} \tag{3-20}$$

$C_1C_2C_3$常量避免$l\left(x,y\right)$的值接近 0 时不稳定。$C_1=\left(K_1L\right)^2$，$C_2=\left(K_2L\right)^2$，$C_3=\frac{C_2}{2}$，常取$K_1=0.01$，$K_2=0.03$像素动态取值范围$L=2^{bitsperpixel}-1$。设$\alpha=\beta=\gamma=1$，以及$C_3=\frac{C_2}{2}$，因此公式（3-14）可以简化为：

$$S_{SSIM}\left(x,y\right)=\frac{\left(2\mu_x\mu_y+\mathrm{C}_1\right)\left(2\sigma_{\mathrm{xy}}+C_2\right)}{\left(\mu_x^2+\mu_y^2+C_1\right)\left(\sigma_x^2+\sigma_y^2+C_2\right)} \tag{3-21}$$

SSIM 的计算基于滑动窗口实现，首先从图像上取一个尺寸为$N\times N$的窗口，基于这个窗口计算 SSIM 值，然后在整个图像上滑动这个窗口，最后将所有窗口的 SSIM 值取平均，得到整张图像的 SSIM 指标，可以有效地评估图像局部区域的相似性，进而得到整个图像的质量评价。

3.3.2 数据集介绍与预处理

雾天数据集（RESIDE）[42]是一个公开的数据集，用于雾天图像处理和计算机视觉中的相关研究，旨在为学术界和工业界提供一个用于开展雾天图像处理算法评估和性能比较的标准数据集。RESIDE 数据集包含两个部分：来自真实场景的雾天图像和通过合成技术生成的雾天图像，含有 13990 个合成模糊图像，使用来自现有室内深度数据集 NYU2 和米德尔伯里立体数据库的 1,399 个清晰图像生成，图 3-6 展示了 RESIDE 数据集样张。真实场景图像是在不同雾天天气条件下的不同地点和场景中采集的，包括城市街道、乡村、海滩等，图像经过专业的摄影设备拍摄并通过真实雾天条件下的图像恢复算法去雾处理得到，为研究者提供了接近实际应用的测试环境，有助于验证算法在实际应用中的效果。另一部分是通过合成技术生成的雾天图像，可以用于测试和验证雾天图像处理算法的性能。

生成雾天图像时，研究团队考虑了不同的雾气浓度、雾气颜色和光照条件等因素，以及不同的背景场景和天气情况，可以模拟出不同的雾天气候和环境条件，使研究者能够更全面地评估算法的效果。RESIDE 数据集还包含了多个子集，如 ITS（Indoor Training Set, ITS）数据集、OTS（Outdoor Training Set, OTS）数据集、SOTS（Synthetic Objective Testing Set, SOTS）数据集等。ITS 数据集主要关注室内场景下的雾天图像处理算法训练，OTS 数据集则侧重于室外场景下的算法训练，而 SOTS 数据集则提供了用于测试的合成数据，以便评估算法在未知场景下的性能。RESIDE 数据集不仅包含了图像数据，还提供了丰富的注释信息，例如对于合成图像，数据集提供了对应的无雾图像，以便进行算法效果的对比和分析，此外 RESIDE 数据集还包含了用于训练和验证算法的分割和姿态注释等信息，提供了更多的实验资源和便利。通过使用 RESIDE 数据集，可以比较和评估不同的雾天图像处理算法的性能，提高雾天图像的处理效果和质量，同时也可以为相关领域的研究提供一个共同的基准，使得不同的研究成果更具可比性和可重复性。

图 3-6 RESIDE 数据集样张

Fig. 3-6 RESIDE Dataset Samples

Cityscapes 数据集[43]是一个专为城市场景分析和计算机视觉任务设计的大规模数据集，包含多个城市场景和不同环境条件下的图像，可用于语义分割、实例分割、物体检测等任务。Cityscapes 数据集图像由车载相机拍摄，并在 50 个不同城市的街道上收集，总共包含 5000 张图像，被拆分为 297 张用于训练，500 张用于验证，1525 张无标注信息图像用于测试，并带有高质量的像素级注释，每张图像的分辨率为1024×2048 像素，这种大规模和高分辨率的特性使得数据集能够捕捉到更多的细节信息，有助于训练出更精确的模型，图 3-7 展示了 Cityscapes 数据集样张。每张图像都经过精细的像素级别标注，标注了包括汽车、

行人、道路、建筑物等在内的多种物体和场景类别，使得模型能够学习到物体的精确边界和场景特征，从而提高模型的分割和检测精度。同时数据集涵盖了来自不同城市的图像，包括城市中心、郊区、道路、公园等多种场景，有助于模型学习到更广泛的城市场景特征，提高模型的泛化能力。Cityscapes 提供两种语义粒度的注释，即类和更高级别的类别。注释可以分为 30 个类和 8 个更高级别的类别，例如汽车、卡车、公共汽车和其他 3 个类别的类别被分组到车辆类别中，其中 19 个类别和 7 个类别用于评估。

图 3-7 Cityscapes 数据集样张

Fig. 3-7 Cityscapes Dataset Samples

Foggy Cityscapes 数据集在 Cityscapes 数据集基础上添加了雾气效果，模拟了雾天条件下的城市街景，共模拟了三个级别的大雾天气，分别为 0.005 对应 600 米的能见度范围、0.01 对应于 300 米的能见度范围和 0.02 对应 150 米的能见度范围，有助于评估算法在不同雾效条件下的性能，图 3-8 展示了 Foggy Cityscapes 数据集样张。虽然 Foggy Cityscapes 数据集涵盖了 30 多个类别，但常用类别主要为车辆、行人、骑行者、卡车、公交车、火车、摩托车和自行车等 8 个类别。

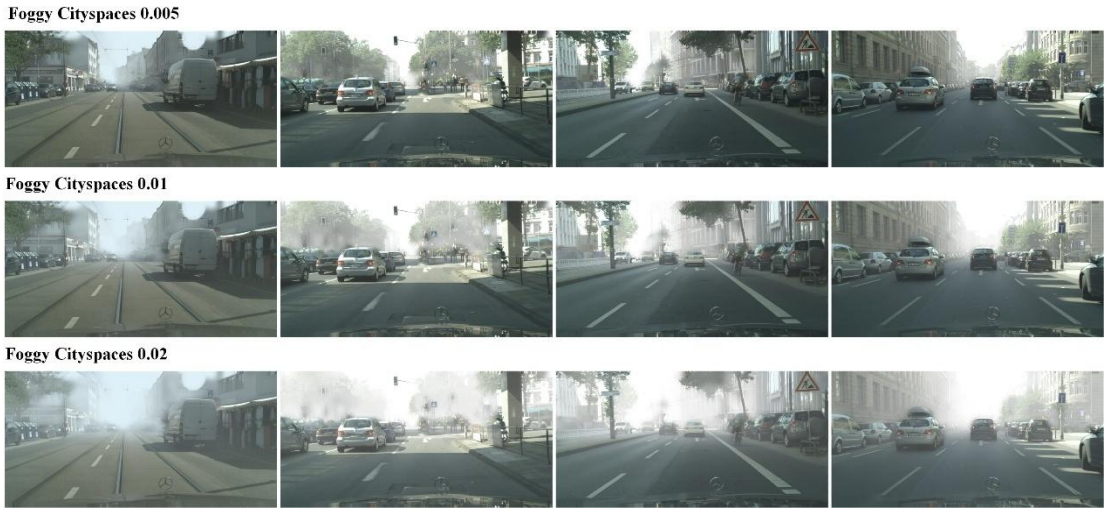


图 3-8 Foggy Cityscapes 数据集样张

Fig. 3-8 Foggy Cityscapes Dataset Samples

3.3.3 消融实验结果分析

为了评估本章方法对浓雾图像与薄雾图像的去雾结果，通过观察，将RESIDE数据集中6000张图片人工筛选出3000张浓雾数据集RSESIDE-F与3000张薄雾数据集RESIDE-M。为了证明本章提出去雾方法的优越性，在参数设置相同的情况下，本章考虑在RESIDE数据集上以FFA-Net为基准，设置了如下四组消融实验。本章主要关注这些因素：（1）特征注意模块。（2）频域增强感知模块。（3）本章改进算法 FFT-Net。为确保实验结果的准确性，每个模型均采用相同的数据划分方式与训练方式，其中模型名称后有√的表示引入了该模型，并采用了PSNR和SSIM两个广泛使用的指标[44-49]，实验结果如表3-1所示。第一组是未加入注意力机制以及频域增强感知模块的实验结果，PSNR 值为 12.3430 dB, SSIM值为0.7376。第二组是增加特征注意模块，分别提高了1.5303 dB PSNR和0.0202 SSIM。第三组是在组架构的每个基础模块里增加频域增强感知模块，PSNR和SSIM指标分别提高了11.1723 dB和0.1577。第四组结合了第二组和第三组的改进，PSNR和SSIM指标分别提高了2.3541 dB和0.0261，说明将两个模块同时结合可以更有效地去除图像中的噪声，提高图像质量和清晰度。

表3-1 消融实验结果对比

Table 3-1 Results of Ablation Experiment

CA&PA	FFT&IFFT	PSNR	SSIM
		12.3430	0.7376
√		13.8733	0.7578
	√	25.0456	0.9155
√	√	27.3997	0.9416

3.3.4 可视化结果分析

为了验证本章改进方法的性能，在RESIDE数据集上对本章改进的方法在同环境下与其他先进的图像去雾方法进行定量比较，分别与图像复原方法NAFSSR、图像增强方法 DD-Net 以及深度学习图像去雾方法 FFA-Net[49]、AOD-Net、GCA-Net和MSBDN，表3-2比较了不同方法在RESIDE数据集上的定量结果。本章改进方法的 PSNR 和 SSIM 指标以 1.3576 dB 和 0.1283 优于

NAFSSR，PSNR 和 SSIM 指标以 2.3541 dB 和 0.0261 优于 FFA-Net，PSNR 和 SSIM 指标以 2.581 dB 和 0.0253 优于 MSBDN，PSNR 和 SSIM 指标以 27.3997 dB 和 0.9416 达到最佳结果。

表 3-2 不同去雾方法在 RESIDE 数据集上的定量比较

Table 3-2 Comparison of Different Dehazing Methods on RESIDE Dataset

方法	RESIDE		RESIDE-F		RESIDE-M	
	PSNR/dB	SSIM	PSNR/dB	SSIM	PSNR/dB	SSIM
DD-Net[10]	22.8636	0.8421	20.4549	0.7599	21.0592	0.7961
NAFSSR[16]	26.0421	0.8133	22.0232	0.7045	22.5338	0.7138
FFA-Net[49]	25.0456	0.9155	21.6652	0.7641	22.8249	0.8378
AOD-Net[18]	18.1977	0.8331	14.3307	0.6897	14.1114	0.6488
GCA-Net[19]	23.5351	0.8890	17.6228	0.7462	17.8183	0.7883
MSBDN[20]	24.8187	0.9163	19.6039	0.8021	20.5799	0.8406
本文方法	**27.3997**	**0.9416**	**23.6078**	**0.8489**	**23.6878**	**0.8736**

表 3-3 不同去雾方法在 Foggy Cityspaces 数据集上的定量比较

Table 3-3 Comparison of Different Dehazing Methods on Foggy Cityspaces Dataset

方法	Foggy Cityspaces 0.005		Foggy Cityspaces 0.01		Foggy Cityspaces 0.02	
	PSNR/dB	SSIM	PSNR/dB	SSIM	PSNR/dB	SSIM
DD-Net[10]	20.6477	0.6935	19.1458	0.6737	18.1218	0.6399
NAFSSR[16]	23.7351	0.6744	21.8516	0.6231	22.3683	0.5974
FFA-Net[49]	26.3034	0.7079	24.3297	0.8026	23. 1887	0.7106
AOD-Net[18]	17.6375	0.6614	15.3913	0.6461	13.4106	0.5584
GCA-Net[19]	22.9270	0.5881	21.4771	0.5018	21.3439	0.4841
MSBDN[20]	22.9618	0.6097	22.4624	0.5837	21.5212	0.5693
本文方法	**31.1413**	**0.9179**	**26.2964**	**0.8885**	**23.2238**	**0.7764**

表 3-3 对比了多种去雾方法在不同雾气浓度的 Foggy Cityspaces 数据集上的性能表现，以 PSNR 和 SSIM 为评价指标。实验结果表明，本章方法在低浓度雾气场景中表现显著优于其他方法，PSNR 达到 31.1413 dB，SSIM 为 0.9179，分

别比次优方法 FFA-Net 提高了 4.8379 dB 和 0.2118；随着雾气浓度增加，所有方法的性能均呈现下降趋势，但本章方法仍保持相对优势，在中等浓度时 PSNR 和 SSIM 以 1.9667 dB 和 0.0859 优于第二名的 FFA-Net；在高浓度下本章方法的 PSNR 与 FFA-Net 接近，但 SSIM 仍提升出 0.0658。相比之下，传统方法如 AOD-Net 和 GCA-Net 整体表现较弱，尤其在 SSIM 指标上差距显著，这表明本章方法在去雾方面取得了有效结果，尤其适用于中低浓度雾场景。

图 3-9 RESIDE 数据集中四组雾天图像的不同方法去雾可视化比较图

Fig. 3-9 Visualization Comparison of 4 Groups of Dehazing Images based on Different Algorithms on RESIDE Dataset

为了进一步证明本章改进方法的有效性，还在 RESIDE 数据集和 Foggy Cityspaces 数据集中进行了与其他先进图像去雾方法的去雾可视化比较。图 3-9 为在 RESIDE 数据集上的去雾可视化比较图，前两行为浓雾图像，后两行为薄雾图像。可以观察到，AOD-Net 不能完全去雾，去雾效果并不理想。GCA-Net 对纹理、边缘等高频细节信息的处理并不完美，如第二行中图像深处黄色圈中的电线。而 FFA-Net 和 MSBDN 这两种以端到端的方式完成从模糊到清晰的图像转换的方法，虽然获得了具有较高 PSNR 和 SSIM 的复原图像，但存在色彩失真问题，如第一行图像深处红色圈中行驶的车身颜色和第四行绿色圈中的道路指示牌。比之下对于模糊有雾图像，本章提出的方法在产生最佳去雾效果的同时，在感知上看起来也接近参考基础事实，更能反映真实的场景细节。

如图 3-10 所示，在 Foggy Cityscapes 数据集上的去雾可视化对比结果表明，现有方法在去雾效果上均存在不同程度的局限性。具体而言，FFA-Net 在图像远景区域的去雾性能欠佳，有较为明显的雾气残留现象；AOD-Net 未能实现完全去雾；GCA-Net 则出现了显著的色彩失真问题；而 MSBDN 的去雾结果虽然有效，但整体图像色调偏暗，特别是在路面区域出现了明显的色彩失真。相比之下，本章所提出的方法在去雾效果上表现出显著优势，不仅有效去除了雾气，而且在视觉感知质量上更接近参考真值，体现了更好的去雾性能。

图 3-10 Foggy Cityspaces 数据集中四组雾天图像的不同方法去雾可视化比较图

Fig. 3-10 Visualization Comparison of 4 Groups of Dehazing Images based on Different Algorithms on Foggy Cityspaces Dataset

3.4 本章小结

本章针对现有的去雾算法难以适用于图像中雾气不均匀分布，主要利用图像中的空间信息而忽略频域信息，从而导致复原图像存在颜色失真和伪雾残留的问题，详细介绍了设计的基于频域增强的道路场景图像去雾方法，构建了一个由多个级联的组架构与特征注意模块构成的图像去雾网络模型，设计的频域增强感知

模块通过图像的空间域特征与频域特征，能够更好地捕捉到图像的纹理和结构信息，有效地去除图像中雾气的同时有效降低计算量。特征注意模块对同样的特征和像素进行非等权重处理，有效地处理了图像中雾气分布不均的问题。实验结果表明，本章所提出的方法能够灵活处理图像中的浓雾和高频信息等区域，实现了图像的有效去雾，图像质量恢复效果优于现有的图像去雾方法。

第 4 章 基于特征融合的去雾道路场景图像小目标检测方法研究

4.1 引言

目标检测是计算机视觉的一个重要任务，可以准确定位图片中的感兴趣区域，为其分配相应的类标签，但由于缺乏目标区域选择策略和手动特征选择的鲁棒性差，检测效率和准确性较低。目前基于深度学习的目标检测技术已成为主流，其中基于深度学习的单阶段目标检测算法在追求速度和准确性的同时脱颖而出。单阶段目标检测算法在输入图像后，可以直接输出检测结果，因此比其他深度学习目标检测算法具有更高的计算效率，更适合目标检测[51]。然而，由于小目标物体覆盖面积小、在图像中的特征表达能力不足，容易受到环境干扰，使模型难以准确定位小目标，造成单阶段目标检测过程中小目标的准确率低于大目标。其次，卷积神经网络的卷积幅度较大，经过连续的下采样和特征提取，输出特征图的规模会缩小，使得小目标的特征信息难以传递到深度特征图中。再次，数据集中小目标的比例较小，目前的目标检测数据集中关注的是大中型目标而较少关注小目标的类型且小目标的分布并不均匀。

本文在第三章设计了基于频域增强的道路场景图像去雾的 FFT-Net 算法，有效的去除了图像中的雾气，接下来本章针对去雾后的特征图中道路远景小目标车辆的漏检问题，提出基于特征融合的道路场景图像小目标检测方法，即 YOLOv11-DB 算法。去雾道路场景图像的小目标检测模型基于 YOLOv11 模型进行改进，首先加入的 Focus 模块能够有效地利用图像数据中的信息，在提高特征提取和目标检测的性能的同时有效提升模型的计算效率；其次设计了多尺度特征融合模块，通过区域残差化生成多尺度特征图与语义残差化实施多速率空洞深度可分离卷积相结合，可以更加高效的提取多尺度上下文信息，避免冗余的感受野，从而提高道路远景小目标车辆的检测精度；最后融入了双向特征金字塔模块，使得特征信息可以双向连接，同时采用自适应特征调整机制，可以更好地捕获多尺度特征信息。

4.2 去雾道路场景图像的小目标检测模型构建

YOLO 网络模型是基于深度学习的目标检测框架，通过预测每个像素的位置及其所属类别的概率来完成目标检测任务，显著提升了检测精度和鲁棒性。本章基于其较高的检测精度、良好的训练效率以及对复杂场景的适应能力选择使用 YOLOv11 模型作为去雾道路场景图像小目标检测的基准模型，相比较于 YOLOv8 模型，YOLOv11 将 CF2 模块改成 C3K2，同时在 SPPF 模块后面添加了一个 C2PSA 模块，并且将 YOLOv10 的 head 思想引入到 YOLO11 的 head 中，使用深度可分离的方法，减少冗余计算。

C3K2 模块的结构集成了 C2f 和 C3 模块的组合，外层由 C2f 结构主导，内层是一个 C3 的结构。C3K2 结构中以 C2f 结构为基础，将其中的特征提取层由 BottleNeck 层替换为 C3k 层，C3k 层与 C3 结构一致，不同之处在于将 BottleNeck 的卷积核大小进行了调整，转变为两个核大小为 3 的卷积层。当 C3k 结构中参数为 False 的时候，C3K2 模块为 C2F 模块，即 Bottleneck 是普通的 Bottleneck；反之当参数为 True 的时候，将 Bottleneck 模块替换成 C3 模块。

C2PSA 模块在 C2f 模块的基础上改进而来，其吸取了 YOLOv10 通过部分自注意力机制提升模型性能的经验，将多改进的多头注意力机制引入 C2f 结构中，替代 BottleNeck 层，并丢弃了中间层的输出结果，类似于 C3 的结构只输出最后一个特征提取层的结果，再与另一个分支进行拼接。多头注意力机制结合了 PSA 结构块，首先通过一个核大小为且无激活层的卷积层，随后通过前馈神经结构块（Feed-Forward Neural Network, FNN）。C2PSA 实现了更强大的注意力机制，从而提高了模型对重要特征的捕捉能力。

在雾天场景中，YOLOv11 算法仍存在着因雾气遮挡且小目标车辆在图像中的尺度通常较小，同时可能因拍摄距离、角度等因素发生变化，而出现小目标车辆漏检的情况，导致检测精度下降，因此需要优化算法对小目标检测的能力，进一步提高检测性能。针对这些问题，本章提出 YOLOv11-DB 算法，通过在 YOLOv11 骨干网络增加 Focus 模块，提升模型减小参数量，提升计算速度；同时设计了多尺度特征融合模块，有效增强了网络的多尺度特征提取能力，能够捕捉到更丰富的上下文信息，从而提高小目标检测的准确性；并在颈部网络设计了双向特征金字塔模块，通过引入双向连接机制，有效增强了特征的表达能力，并

且采用了加权融合策略，为每个输入特征分配了不同的权重，使得特征融合过程更加精准，进一步提升了小目标车辆检测的准确性。图 4-1 展示了基于特征融合的道路场景图像小目标检测模型。

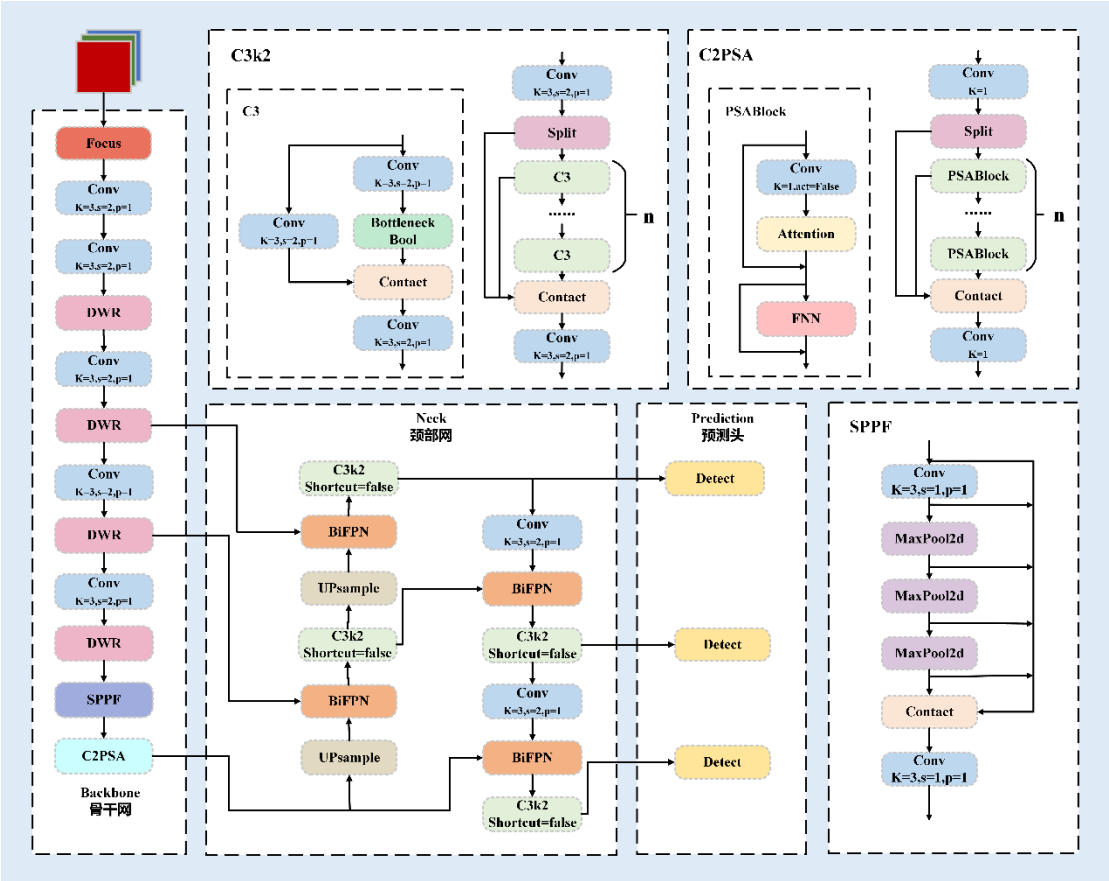


图 4-1 基于特征融合的道路场景图像小目标检测模型

Fig. 4-1 Small Target Detection Model of Road Scene Image based on Multi-scale Feature Fusion

4.2.1 Focus 模块

Focus 模块是一种用于特征提取的卷积神经网络层，通常作为首个卷积层，用于将特征提取过程中的信息压缩与整合生成更高层次的特征表示，通过对输入特征图进行下采样来降低计算复杂度。下采样过减少像素数量来实现图像尺寸的缩减，提高了模型的泛化能力。下采样即池化操作，采样有最大值采样，平均值采样，随机区域采样等，对应池化有比如最大值池化，平均值池化，随机池化等。

Focus 模块将高分辨率特征图分割成多个低分辨率子图，采用隔列采样与拼接相结合的方式进行处理，如图 4-2 所示。Focus 模块对输入特征图进行像素间隔采样，生成四张互补且信息完整的特征图，随后通过 Concat 函数连接将 W、

H 信息就集中到通道空间，在压缩特征图宽高信息的同时将输入通道数扩展为原来的四倍，在保证信息完整性的同时实现了二倍下采样，即拼接起来的特征图相对于原先的 RGB 三通道模式变成了 12 个通道，相比传统的两层卷积加一层 Bottleneck 结构，Focus 模块在有效保留关键信息的同时有效减少了参数量。

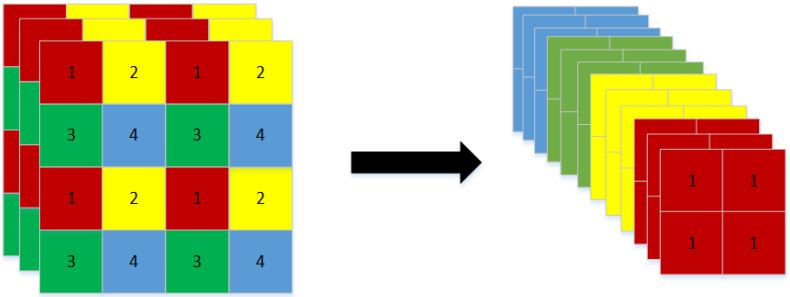


图 4-2 Focus 模块

Fig. 4-2 Focus Module

4.2.2 多尺度特征融合模块

多尺度特征融合模块主要由扩张残差[52]（Dilation-wise Residual, DWR）结构块组成，使得网络能够更灵活地适应不同尺度的特征，从而更准确地识别和分割图像中的物体，该模块在网络的高阶采用设计的两步特征提取方法，极大地提高了网络在小目标检测中的性能。反向残差（Simple Inverted Residual, SIR）模块是专门为低级设计的，从 DWR 模块调整，以满足小接受场的要求。

多尺度特征融合模块如图 4-3 所示，H 和 W 表示输入图像的尺寸，N 表示类别数量，c 表示特征图通道的基础数量，RR 和 SR 分别表示区域残差化和语义残差化，Conv 表示卷积，DConv 表示深度卷积，D-n 表示扩展率为 n 的扩展卷积，圆圈中的+表示加法运算，BN 表示批量归一化层。多尺度特征融合模块可以看做是一个编码器-解码器设置，编码器由 Stem 块，SIR 结构块的低阶段和 DWR 模块的两个高阶段组成。高阶特征提取采用 DWR 模块，采用两步多尺度上下文信息获取方法，SIR 模块是由 DWR 模块修改而来，在低阶段满足更小的接受场。Stem 块用于初始处理，SIR 模块用于从低阶段提取特征，DWR 模块用于从高阶段提取特征。上三层的输出进入解码器，解码器采用简单的全卷积神经网络（Fully Convolutional Networks, FCN）风格，来自高阶段的特征图首先进行上采样，然后与低阶段的特征图进行拼接，最后使用 Seg Head 处理拼接后的特征图以进行最终预测，最终的预测由解码器生成，并由地面真值直接监督，并且

在训练过程中不需要辅助监督。

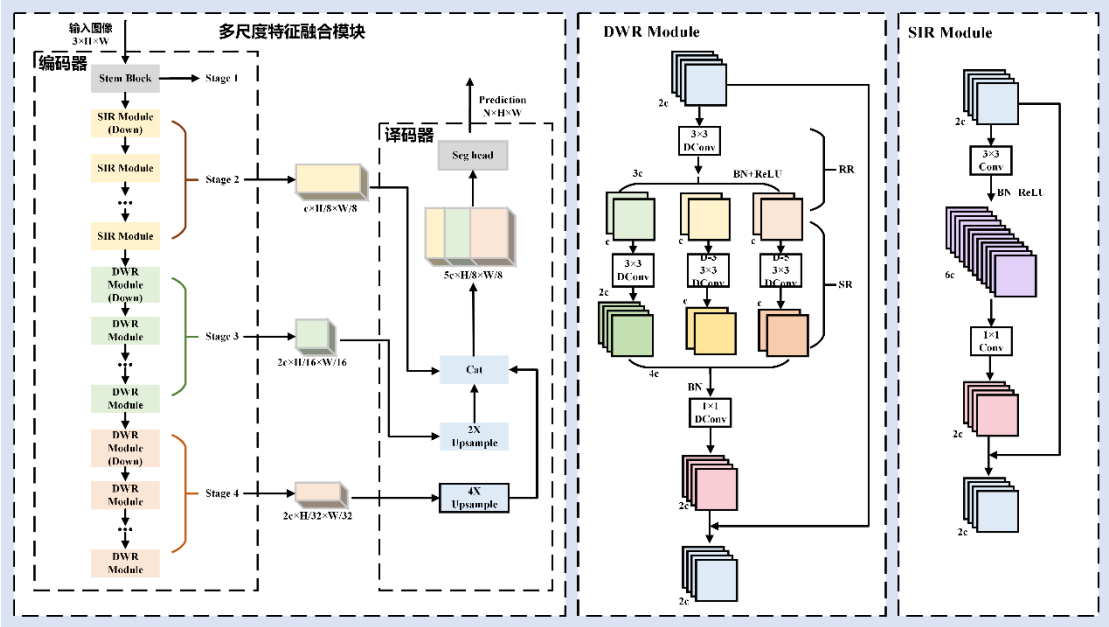


图 4-3 多尺度特征融合模块

Fig. 4-3 Multi-scale Feature Fusion Block

DWR 模块以残差方式设计，使用两步法高效地获取多尺度上下文信息，然后融合生成多尺度感受野特征图，具体而言，单步多尺度上下文信息获取方法被分解为两步法，以降低获取难度。首先从输入特征中生成关注的残差特征，称为区域残差化，即通过一个常见的3×3卷积层结合 BN 和 ReLU 层生成一系列具有不同大小的简洁区域形式特征图，3×3卷积层用于初步特征提取，ReLU 激活函数在激活区域特征并使其简洁方面发挥了重要作用。接下来采用语义残差化，每个通道特征仅应用单一期望的感受野，以避免可能的冗余感受野。实际上，第一步中可以根据第二步中感受野的大小有选择的学习所需的简洁区域特征图，以反向匹配感受野，即首先将区域特征图分为若干组，然后对不同组进行不同速率的空洞深度可分离卷积。通过两步的区域残差化-语义残差化方法，多速率深度可分离空洞卷积获取的复杂语义信息转变为在特征图上使用单一期望的感受野进行形态学滤波。区域形式的特征图使深度可分离空洞卷积的角色简化为形态学滤波，从而使学习过程有序化，因此可以更高效地获取多尺度上下文信息，在获取多尺度上下文信息后聚合多个输出，即所有特征图被连接起来。接下来，对特征图实施批量归一化操作，然后采用逐点卷积特征信息进行整合与压缩，形成最终残差，将最终残差与原始输入特征进行叠加融合，增强特征表达能力，构建出包含多层次信息的复合特征表示，有效提升了对复杂特征的提取能力。

SIR 模块从 DWR 模块设计而来，满足低阶段较小感受野大小的需求，以保持高特征提取效率。基于 DWR 模块进行了两个修改，首先移除了多分支空洞卷积结构，仅保留第一个分支以压缩感受野；其次由于贡献最小，因此移除了3×3深度可分离卷积。由于输入特征图尺寸较大且语义较弱，单通道卷积收集的信息太少，因此在低阶段一步特征提取比两步特征提取更高效，最终得到了一个继承 DWR 模块的残差和倒置瓶颈结构的简单模块。

4.2.3 双向特征金字塔模块

双向特征金字塔（Bidirectional Feature Pyramid Network, BiFPN）模块[53]可以进行多尺度融合，增加了跨尺度的链接，充分利用了全局上下文特征信息，提高了浅层特征表达能力，为每个尺度特征分配可学习的权值，调整每个尺度特征的贡献度，增强浅层特征的利用率，进一步提高了网络的检测性能，如图 4-4 所示。

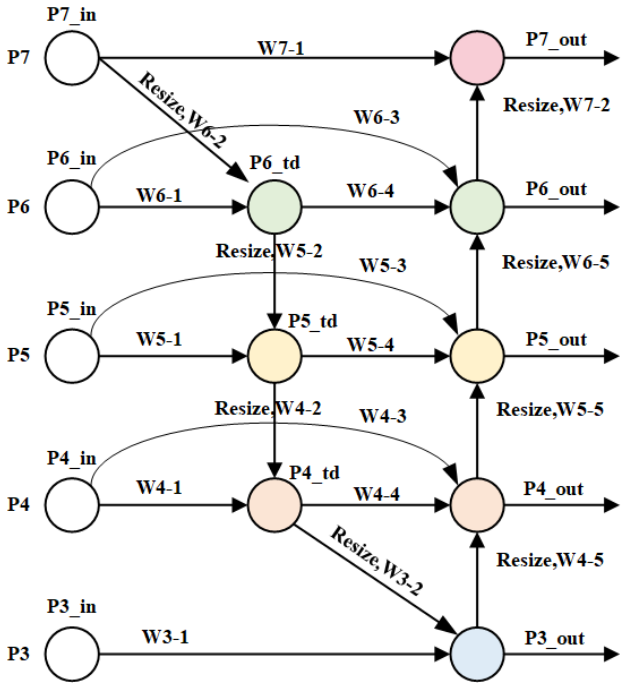


图 4-4 双向特征金字塔模块

Fig. 4-4 BiFPN Model

$P3$、$P4$、$P5$、$P6$、$P7$ 表示骨干网的输出层，每个输出层都有对应的输出特征，包含通道数，特征的大小等信息。无色的圆圈表示特征，有色的圆圈表示算子，有线连接的表示权重 w，Resize 操作表示向上向下连接即上采样或下采样。各输出层特征的加权计算公式如下：

$$P3\_out = Conv\left(\frac{P3\_in \times W31 + \mathrm{Re}side(P4\_td) \times W32}{W31 + W32 + \varepsilon}\right) \tag{4-1}$$

$$P4\_out = Conv\left(\frac{P4\_in \times W43 + P4\_td \times W44 + \mathrm{Re}side(P3\_out) \times W45}{W43 + W44 + W45 + \varepsilon}\right) \tag{4-2}$$

$$P5\_out = Conv\left(\frac{P5\_in \times W53 + P5\_td \times W54 + \mathrm{Re}side(P4\_out) \times W55}{W53 + W54 + W55 + \varepsilon}\right) \tag{4-3}$$

$$P6\_td = Conv\left(\frac{P6\_in \times W61 + \mathrm{Re}side(P71\_in) \times W62}{W61 + W62 + \varepsilon}\right) \tag{4-4}$$

$$P6\_out = Conv\left(\frac{P6\_in \times W63 + P6\_td \times W64 + \mathrm{Re}side(P5\_out) \times W65}{W63 + W64 + W65 + \varepsilon}\right) \tag{4-5}$$

$$P7\_out = Conv\left(\frac{P7\_in \times W71 + \mathrm{Re}side(P6\_out) \times W72}{W71 + W72 + \varepsilon}\right) \tag{4-6}$$

传统特征金字塔网络在处理多尺度特征时存在明显局限性，其将所有输入特征视为同等重要，采用简单的特征叠加方式进行融合，忽视了不同分辨率特征对最终输出的差异性贡献，然而在 BiFPN 中观察到由于输入特征在空间分辨率上存在显著差异，因此对输出特征的贡献是不等的。为了解决这个问题，BiFPN 提出了为每个输入添加一个额外的权重，并让网络学习每个输入特征的重要性，公式（4-7）为：

$$O = \sum_i w_i \times I_i \tag{4-7}$$

w_i 是一个可学习的权重，可以是标量--每个特征，向量--每个通道或多维张量--每个像素。$w_i \geq 0$ 是通过在每个 w_i 后应用 ReLU 激活函数来保证的，$\varepsilon = 0.0001$ 是一个极小值，以避免数值不稳定，同样每个归一化权重的值也在 0 和 1 之间，且由于没有 softmax 操作，因此效率要高得多。

双向特征金字塔模块在自顶向下和自底向上路径之间建立双向连接，允许不同尺度特征间的信息更有效地流动和融合，同时通过为每个输入特征添加权重来优化特征融合过程，使得网络可以更加重视信息量更大的特征，最后通过移除只有一个输入的节点、添加同一层级的输入输出节点之间的额外边，并将每个双向路径视为一个特征网络层[54]来优化跨尺度连接，通过简化和优化计算流程，提高了网络的运行速度和精度。

4.3 实验设计与结果分析

4.3.1 实验环境与评价指标

本章中研究的目标检测的硬件环境为：NVIDIA GeForce RTX3090 GPU，24GB 显存，Intel 酷睿(Core)i9-14900KF@3.20GHz CPU，128GB 内存。软件环境为:Windows 10 操作系统，CUDA11.8，深度学习框架为 Pytorch2.0.0，编程语言 python 3.10.16。

本章对小目标车辆进行检测，但不同场景对于小目标的定义各不相同，目前尚未形成统一的标准。现有的小目标定义方式主要分为以下两类，即基于相对尺度的定义与基于绝对尺度的定义[55]。

（1）基于相对尺度定义

在计算机视觉领域，小目标的界定主要基于目标与整幅图像的尺寸比例关系。Chen 等人提出一个针对小目标的数据集，并提出了一种量化标准：当某类目标的 BBox 面积与图像总面积之比的中值处于 0.08%到 0.58%范围内时，即可判定为小目标。以分辨率为 640 像素×480 像素的图像为例，尺寸介于 16 像素×16 像素至 42 像素×42 像素之间的即为小目标。此外，学术界对小目标的判定还存在其他广泛采用的标准：第一种方法是计算目标边界框的宽度或高度与图像对应维度的比值，若该值小于 0.1 则可判定为小目标；第二种方法是通过计算边界框面积与图像面积比值的平方根，当结果小于 0.03 时即符合小目标特征；第三种判定依据则是直接比较目标实际像素数量与图像总像素数的比例关系。这些定义方法从不同角度为小目标的识别提供了理论依据。

（2）基于绝对尺度定义

从目标在图像中的实际像素尺寸出发，学术界对小目标的界定存在多种标准。在目标检测领域，MS COCO 数据集[56]提出的 32 像素×32 像素阈值被广泛采纳。这一标准的制定依据主要包含两个方面：其一 Torralba 等人[57]的视觉认知研究表明，人类能够清晰识别彩色图像的最小分辨率恰好为 32 像素×32 像素。其二从深度学习网络结构来看，以网络结构 VGG-Net[58]为例，经过多层池化操作后，每个特征点对应的原始图像区域正好是 32 像素×32 像素，这为小目标的识别提供了理论依据。不同应用场景下的数据集根据各自特点制定了相应的标准。例如，

在航空图像分析领域和人脸检测领域，航空图像数据集 DOTA[59]与人脸检测数据集 WIDER FACE[60]中都将像素值范围在［10, 50］之间的目标定义为小目标；对于行人检测任务，行人识别数据集 CityPersons[61]中将高度小于 75 像素的行人定义为小目标；而在专门针对航空图像中小尺寸行人检测的 TinyPerson 数据集[62]中，定义标准更为精细，将 20 到 32 像素的目标划分为小目标，同时将 2 到 20 像素范围内的目标进一步归类为微小目标，这种细分标准更好地适应了特定场景下的检测需求。

评估目标检测模型的性能需要建立科学的指标体系。本章将重点阐述三个核心评价参数，即精确率（Precision）、均值平均精度（mAP）、召回率（Recall）。这些指标的量化计算需要明确几个基本概念，即真正例（True Positives, TP）、真负例（True Negatives, TN）、假正例（False Positives, FP）以及假负例（False Negatives, FP）。具体而言，TP 代表模型正确识别出的目标数量；TN 则反映了模型准确排除的非目标数量；FP 表示模型将背景错误识别为目标的次数；FN 则记录了模型未能检测到的实际目标数量。这四个基本概念构成了目标检测性能评估的基础框架，通过它们可以准确计算出各项评价指标，从而全面衡量模型的检测能力[63]。

精确率--Precision：表示分类器[分类正确的正样本]占[分类器分为正样本（不管对错）]的比例。公式（4-8）中 N 为模型预测此类的样本数。

$$\mathrm{Pr}ecision = \frac{TP}{TP + FP} = \frac{TP}{N} \tag{4-8}$$

召回率--Recall：表示分类器[分类正确的正样本]占[所有真正正样本]的比例。

$$\mathrm{Re}call = \frac{TP}{TP + FN} \tag{4-9}$$

mAP--平均精度均值（mean Average Precision, mAP）：即平均精确度(Average Precision, AP)的平均值。mAP 值越高，表明该目标检测模型在给定的数据集上的检测效果越好，AP 的计算公式（4-10）则为:

$$AP = \int_0^1 P(r)\,dr = \sum_{k=1}^{N} P(k)\Delta r(k) \tag{4-10}$$

N 为数据总量，k 为每个样本点的索引，$\Delta r(k) = r(k) - r(k-1)$，即可得到 mAP 的计算公式如下，$k$ 为类别数目，当 $k=1$ 时，mAP = AP。

$$mAP = \frac{\sum_{i=1}^{k} AP_i}{k} \tag{4-11}$$

4.3.2 数据集介绍与预处理

本章仍使用 Cityspaces 数据集、Foggy Cityspaces 数据集进行实验，对道路场景中车辆进行检测，设置了 6 个类别，分别为车辆、卡车、公交车、火车、摩托车和自行车，并且按照相对尺度定义对数据集中的小目标进行了设定。经计算，Cityspaces 数据集、Foggy Cityspaces 数据集中图像像素为 1024 像素×2048 像素分辨率图像，则 41 像素×41 像素到 110 像素×110 像素的目标应考虑为小目标。

4.3.3 消融实验结果分析

多尺度融合结构有助于提取更丰富的特征信息，而更好的解决在图像中小目标车辆漏检问题。双向特征金字塔模块通过双向的跨尺度连接，可以更好地捕捉不同尺度的目标特征并且增强特征图的表达能力和减少模型计算量，因此能够更好地提升有雾图像目标检测模型的性能。为了进一步分析本章提出 YOLO11-DB 算法的有效性和其中各个模块之间组合对原模型的提升，在参数设置相同的情况下，以 YOLO11 为基准设计了如下消融实验。

将道路场景中的车辆、卡车、公交车、火车、摩托车和自行车六个类别作为检测目标，不同改进方法的消融实验结果如表 4-1 所示，并用 √ 表示添加了该改进方法。从表中可以看出，分别引入多尺度特征融合模块和双向特征金字塔模块，模型在 P、R 和 mAP 等方面都有不同程度的提升。多尺度特征融合模块在 Cityscapes 数据集中显著提升检测性能，mAP50%和 mAP50-90%分别增加 3.5%和 2.3%，召回率提高了 7.8%，且在的 Foggy Cityscapes 数据集中精确率最高提升了 6.9%。双向特征金字塔模块则通过增强特征交互能力，在 Cityscapes 数据集中 mAP50%提高了 3.3%、精确率提高了 4.2%，并在雾气浓度最高的道路场景下 mAP50-90%提高了 2%。通过两模块的协同融合，其中 Cityscapes 数据集的 mAP50%提高了 4.3%，Foggy Cityscapes 0.01 数据集 mAP50%提高了 3.3%、mAP50-90%提高了 3.5%，且在 Foggy Cityscapes 0.02 数据集中 mAP50%提高了 2.1%、mAP50-90%提高了 2.3%，证明了本章所提出的算法有效提升了目标检测

的准确性。

表 4-1 道路场景图像的目标检测方法消融实验结果对比

Table 4-1 Comparison of Ablation Results of Target Detection Methods for Road Scene Images

	YOLOv11	DWR	BiFPN	mAP50/%	mAP50-95/%	P/%	R/%
Cityspaces	√			42.3	24.5	61	36.3
	√	√		45.8	26.8	57.7	44.1
	√		√	45.6	27.5	**65.2**	39.7
	√	√	√	**46.6**	**27.7**	58.8	**42.5**
Foggy Cityspaces 0.005	√			42.2	24.2	57.2	40
	√	√		44.7	26.6	60.7	**43.3**
	√		√	43.9	26.7	59	40.3
	√	√	√	**45.4**	**27.3**	**67.5**	40.9
Foggy Cityspaces 0.01	√			41.7	23.7	56.5	39.8
	√	√		42.5	25.5	**63.4**	37.6
	√		√	43.7	26.4	59.2	39.9
	√	√	√	**45.9**	**27.2**	62.6	**41.2**
Foggy Cityspaces 0.02	√			39.4	22.8	57.2	35.6
	√	√		40.6	24.2	54.3	36.5
	√		√	41.4	24.8	**63.3**	36.9
	√	√	√	**41.5**	**25.1**	54.4	**40.1**

接下来本章所提出的算法在 Cityspaces 数据集、Foggy Cityspaces 数据集中对六个类别的小目标车辆进行实验，如表 4-2 所示，因小目标可能由于分辨率低、遮挡、背景复杂等因素而难以被准确检测，因此检测指标 mAP 均较道路场景正常大小的目标检测指标 mAP 略低。多尺度特征融合模块在 Cityspaces 数据集中使 mAP50%和 mAP50-90%分别提高了 0.5%和 0.4%，精确率提升 2.7%，在 Foggy Cityscapes 各雾气浓度数据集中也有不同程度的 mAP 和精确率、召回率提升。而 BiPFN 模块的加入同样在 Cityspaces 数据集中提高了 mAP 指标和召回率，且在 Foggy Cityscapes 0.01 数据集中，mAP50%显著提升 1.3%，精确率和召回率也有较大幅度增长。当两个模块结合使用时，模型性能获得更大飞跃，在 Cityspaces

数据集中 mAP50%和 mAP50-90%分别提升 1.2%和 0.8%，召回率提高 3.5%，特别是在 Foggy Cityscapes 0.02 数据集中，精确率提升高达 8.6%。可以看出，本章所提出的算法有效提升了小目标车辆检测的准确性，能够更精准的定位小目标车辆。

表 4-2 道路场景图像的小目标检测方法消融实验结果对比

Table 4-2 Comparison of Ablation Results of Small Target Detection Methods for Road Scene Images

	YOLOv11	DWR	BiFPN	mAP50/%	mAP50-95/%	P/%	R/%
Cityspaces	√			21.3	10.2	37	23.9
	√	√		21.8	10.6	**39.7**	23.3
	√		√	21.5	10.7	34.9	24.2
	√	√	√	**22.5**	**11**	30.4	**27.4**
Foggy Cityspaces 0.005	√			18.8	8.6	31.9	20.8
	√	√		19.6	8.7	21.2	19.8
	√		√	19.1	8.9	**40.1**	19.8
	√	√	√	**20.7**	**9.3**	21.5	**21.9**
Foggy Cityspaces 0.01	√			15.1	7.4	**32.7**	18
	√	√		15.3	7.5	26.6	18.4
	√		√	16.4	7.5	22.6	**19.4**
	√	√	√	**17.8**	**7.7**	**32.7**	18.6
Foggy Cityspaces 0.02	√			13.4	6.1	27.1	11.4
	√	√		14.3	6.3	34	**18**
	√		√	13.6	6.3	29.6	14.6
	√	√	√	**14.9**	**7.1**	**35.7**	12.4

接下来选择在雾气浓度适中的 Foggy Cityspaces 0.01 数据集上进行雾天条件下道路场景目标检测方法的消融实验。首先使用基于频域增强的道路场景图像去雾方法对 Foggy Cityspaces 0.01 数据集图像进行去雾操作，得到的去雾道路场景图像再进行基于特征融合的目标检测算法实验。

雾天条件下道路场景目标车辆检测实验结果如表 4-3 所示。在能见度范围

300 米的有雾图像中，原有 YOLOv11 网络加入基于频域增强的道路场景图像去雾方法，即 FFT-Net 算法进行去雾后，mAP50%提高了 0.5%、mAP50-90%提高了 1%且精确率提高了 5.3%；通过基于特征融合的道路场景图像小目标检测方法，即 YOLOv11-DB 算法进行车辆检测，mAP50%提高了 3.3%、mAP50-90%提高了 3.5%且精确率提高 6.1%，召回率提高了 1.4%；最后先使用进行特征图去雾后再进行车辆检测，可以看出设计的模型在各项指标上都得到了较大的提升，mAP50%提高了 4.2%、mAP50-90%提高了 4.9%且精确率提高了 12.4%。

表 4-3 雾天条件下道路场景中目标检测方法的消融实验结果对比

Table 4-3 Comparison of Ablation Results of Target Detection Methods in Road Scenes under Foggy Conditions

YOLOv11	YOLOv11-DB	FFT-Net	mAP50/%	mAP50-95/%	P/%	R/%
√			41.7	23.7	56.5	39.8
√		√	42.2	24.7	61.7	36.3
	√		45	27.2	62.6	**41.2**
	√	√	**45.9**	**28.6**	**68.9**	38.3

表 4-4 雾天条件下道路场景中小目标车辆检测方法的消融实验结果对比

Table 4-4 Comparison of Ablation Results of Small Target Vehicles Detection Methods in Road Scenes under Foggy Conditions

YOLOv11	YOLOv11-DB	FFT-Net	mAP50/%	mAP50-95/%	P/%	R/%
√			15.1	7.4	32.7	18
√		√	16.3	7.5	**33.6**	17.2
	√		17.8	7.7	32.7	18.6
	√	√	**18.4**	**8.7**	32.8	**19**

雾天条件下道路场景小目标车辆检测实验结果如表 4-4 所示。在能见度范围 300 米的有雾图像中，原有 YOLOv11 网络加入 FFT-Net 算法进行去雾后，mAP50%提高了 1.2%、mAP50-90%提高了且精确率提高了 0.9%；通过 YOLOv11-DB 算法进行车辆检测，mAP50%提高了 2.7%，mAP50-90%提高了 0.3%且召回率提高了 0.6%；最后先使用进行特征图去雾后再进行车辆检测，可以看出改进后模型在各项指标上都得到了较大的提升，mAP50%提高了 3.4%、mAP50-90%提高了 1.3%且精确率提高了 0.1%，召回率提高了 1%。由实验可以看出改进后的方法对

道路场景上的小目标车辆检测精度均有较大提升并且复杂度并未大幅增加，证明了本章所提出算法的优越性。

4.3.4 可视化结果分析

为了验证本章所提出算法的性能，在 Foggy Cityspaces 数据集上对本章所提出的方法与其他先进的目标检测方法在同环境下进行小目标车辆检测定量比较，如表 4-5 所示。与 YOLOv11 算法相比，改进算法的 mAP50%、P、R 分别提高了 1.1%、6.2%和 0.8%，证明了所提出的算法的优越性。与其他有雾图像目标检测算法 env-YOLO[64]、DINO[65]和 YOLO-DF[66]相比，mAP50%分别提高了 2.3%、3%和 1.4%，并比检测性能较高的 YOLOv8 相比，mAP50%提高了 1.9%，可以看出本章所提出的算法有效的提升了检测精度，证明了所提出的算法的优越性。

表 4-5 不同目标检测方法在 Foggy Cityspaces 数据集上的定量比较

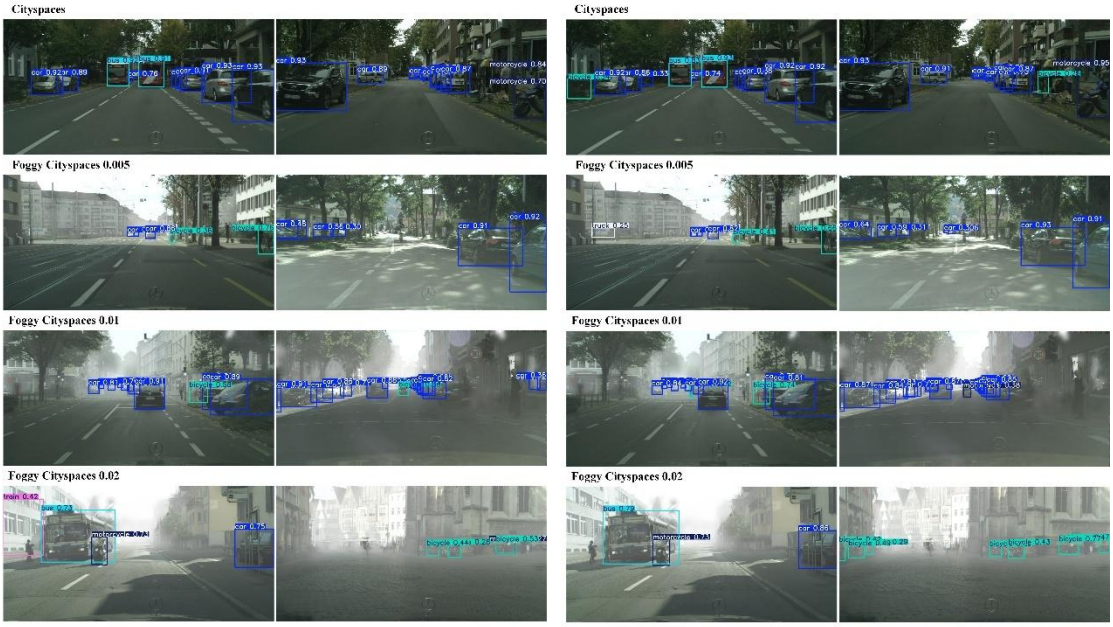
Table 4-5 Comparison of Different Target Detection Methods on Foggy Cityspaces Dataset

方法	mAP50%	P/%	R/%
env-YOLO[64]	14.9	31.1	15.2
DINO[65]	14.2	23.7	**22.4**
YOLO-DF[66]	15.8	31.3	19.7
YOLOv8	15.3	31.4	18.9
YOLOv11[35]	16.7	26.5	17.8
本文方法	**17.8**	**32.7**	18.6

为进一步验证本章设计的小目标检测模型对雾天场景下的车辆目标的检测效果，将本文算法在 Cityspaces 数据集、Foggy Cityspaces 数据集进行可视化实验。首先对道路场景车辆目标进行可视化实验，实验对比结果如图 4-5 所示。第一行为 Cityspaces 数据集，即无雾条件下道路场景车辆进行目标检测；第二行为 Foggy Cityspaces 0.005 数据集，即对雾气浓度在 600 米的能见度范围条件下道路场景车辆进行目标检测；第三行为 Foggy Cityspaces 0.01 数据集，即对雾气浓度在 300 米的能见度范围条件下道路场景车辆进行目标检测；第四行为 Foggy Cityspaces 0.02 数据集，即对雾气浓度在 150 米的能见度范围条件下道路场景车辆进行目标检测，对比图（a）则以 YOLOv11 为基准算法进行车辆目标检测，

对比图（b）则为本文算法进行车辆目标检测。

由图 4-5 可以看出，在 Cityspaces 数据集中原算法未检测到近处的自行车、远处的轿车以及将行人误检为摩托车；在 Foggy Cityspaces 0.005 数据集中原算法未检测到远处的大卡车以及图像深处的多辆小轿车；在 Foggy Cityspaces 0.01 数据集中原算法未检测到图像远景处的被遮挡的自行车以及将摩托车误检为自行车；在 Foggy Cityspaces 0.02 数据集中原算法将白色楼宇误检为火车以及未检测到被雾气遮盖的大量自行车以及被雾气和自行车严重遮挡的摩托车。总体来说，因为遮挡以及检测目标较小的原因，原算法出现大量漏检以及误检的问题，而本章所提出的算法则更有效地检测出被遮挡的目标和图像深处的小目标车辆，同时检测精度比原模型更高并且提高了检测目标的准确性。



(a) 原算法检测效果　　**(b) 改进算法检测效果**

图 4-5 道路场景目标检测算法可视化比较图

Fig. 4-5 Visual Comparison of Road Scene Target Detection Algorithms

接下来对道路场景小目标车辆进行可视化实验，实验对比结果如图 4-6 所示。可以看出，在 Cityspaces 数据集中原算法未检测到图像远景处的自行车以及将楼梯扶手和街边护栏误检为自行车；在 Foggy Cityspaces 0.005 数据集中原算法未检测到图像远景处被雾气和道路禁车柱遮挡的多辆自行车；在 Foggy Cityspaces 0.01 数据集中原算法未检测到图像远景处的被雾气遮挡的白色轿车；在 Foggy Cityspaces 0.02 数据集中原算法将行李箱误检为自行车以及未检测到被雾气遮盖的大量小轿车。总体来说，因为遮挡以及检测目标较小的原因，原算法出现大量

漏检的问题，而本章所提出的算法提高了小目标的检测能力，能够在雾气条件下检测出图像深处被遮挡的小目标车辆，有效改善了小目标车辆漏检问题，具有更好的检测性能。

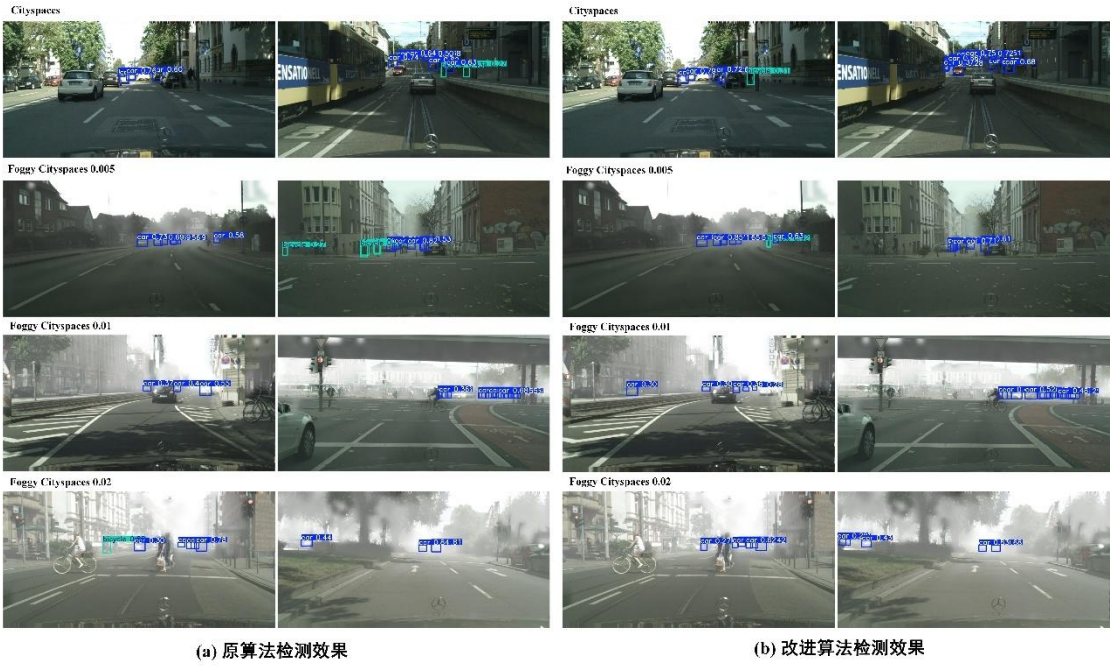


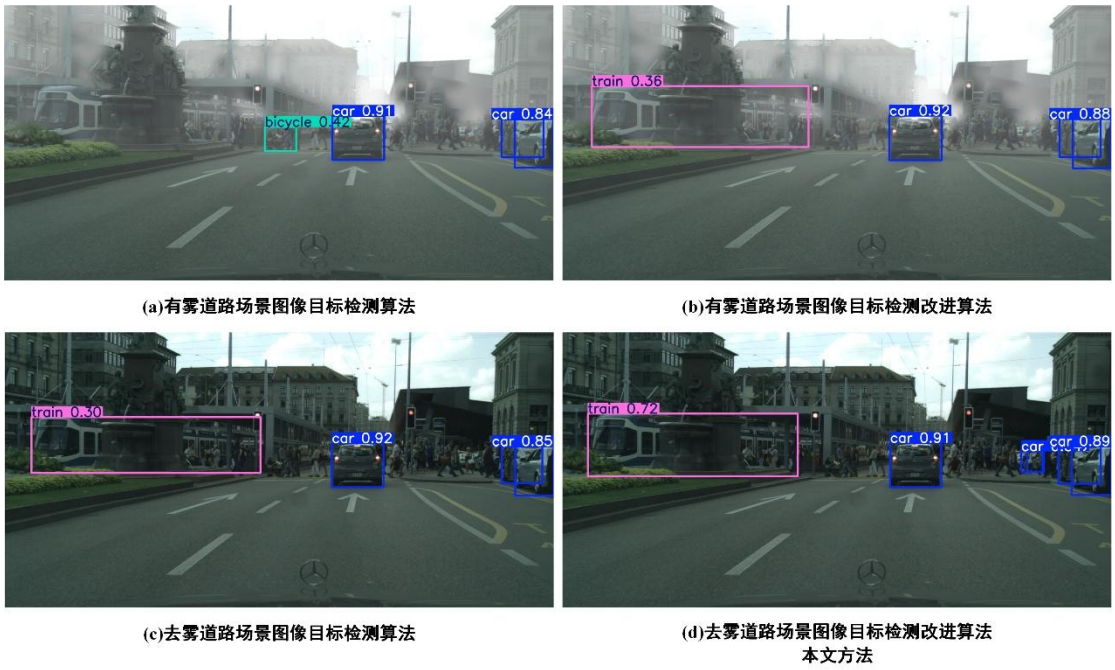
图 4-6 道路场景小目标车辆检测算法可视化比较图

Fig. 4-6 Visual Comparison of Road Scene Small Target Vehicles Detection Algorithms

最后为了验证本文提出的雾天条件下道路场景目标检测方法的有效性，首先选择在雾气浓度适中的 Foggy Cityspaces 0.01 数据集中使用基准 YOLOv11 算法进行车辆目标检测得到对比图（a）；其次使用本章所提出的检测算法进行车辆目标检测得到对比图（b）；接下来使用基于频域增强的道路场景图像去雾方法对图像进行去雾处理，再对去雾后的图像使用基准 YOLOv11 算法进行车辆目标检测得到对比图（c）；最后对去雾后的图像使用本章所提出的检测算法进行车辆目标检测得到对比图（d）。

雾天条件下道路场景目标检测算法可视化比较图如图 4-7 所示，由对比图（a）可以看到，原算法未检测出被雾气遮挡的大型火车以及将婴儿手推车误检为自行车；由对比图（b）可以看到，本章所提出的检测算法检测出被雾气遮挡的大型火车以及未有误检问题,但由于雾气和大量行人遮挡，未识别到图像远景处的小轿车；由对比图（c）可以看到，经过图像经过有效去雾，原算法清晰的检测到大型火车以及未有误检问题，但也未识别到图像远景处被大量行人遮挡的小轿车；由对比图（d）可以看到，本文算法清晰的识别了图像远景处被大量行人遮

挡的小轿车，有效降低了漏检率。

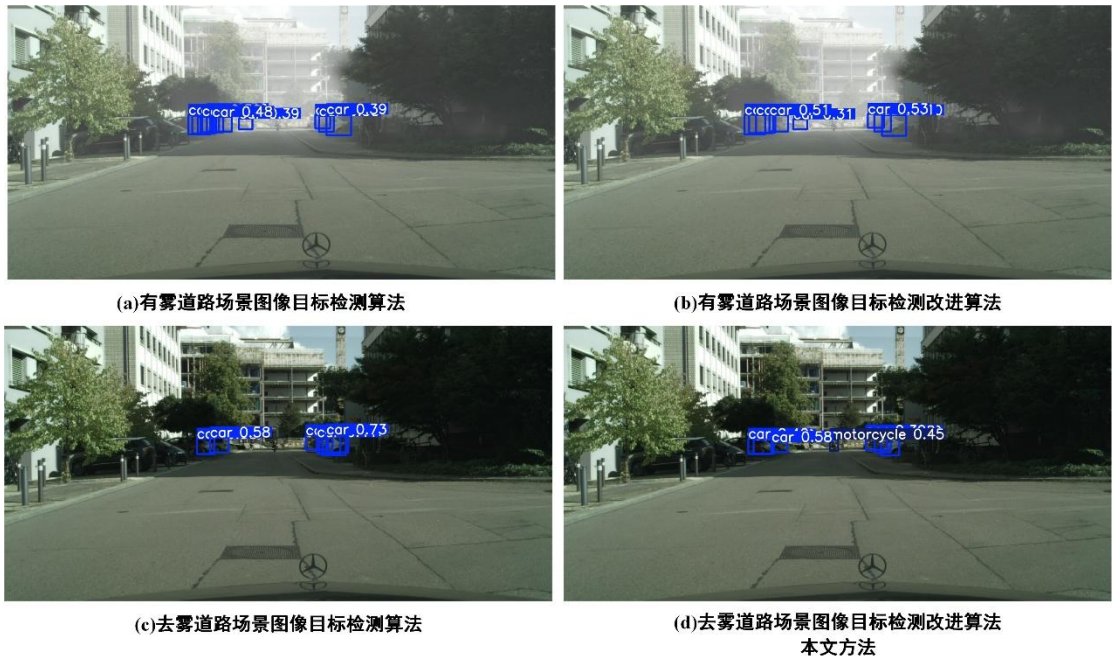


(a)有雾道路场景图像目标检测算法　(b)有雾道路场景图像目标检测改进算法

(c)去雾道路场景图像目标检测算法　(d)去雾道路场景图像目标检测改进算法 本文方法

图 4-7 雾天条件下道路场景目标检测算法可视化比较图

Fig. 4-7 Visualization Comparison of Target Detection Algorithms in Road Scene under Foggy Conditions



(a)有雾道路场景图像目标检测算法　(b)有雾道路场景图像目标检测改进算法

(c)去雾道路场景图像目标检测算法　(d)去雾道路场景图像目标检测改进算法 本文方法

图 4-8 雾天条件下道路场景小目标车辆检测算法可视化比较图

Fig. 4-8 Visualization Comparison of Small Target Vehicles Detection Algorithms in Road Scene under Foggy Conditions

雾天条件下道路场景小目标检测算法可视化比较图如图 4-8 所示，由对比图（a）可以看到，原算法未检测出被雾气遮挡的摩托车以及将图像左侧被树木遮

挡少量轿车识别为多辆轿车；由对比图（b）可以看到，本章所提出的检测算法虽未检测出被雾气遮挡的摩托车但有效降低了图像左侧被树木遮挡停放在道路旁的轿车数量；由对比图（c）可以看到，经过图像经过有效去雾，原算法仍未检测到图像远景处道路中间的摩托车且将图像右侧少量停放在道路旁的轿车识别为多辆轿车；由对比图（d）可以看到，本文算法清晰的识别了图像远景处道路中间的摩托车且正确识别了道路两侧停放的小轿车。

综合来看，本文提出的方法能够有效的对有雾图像进行图像去雾并且有效的降低了图像深处小目标车辆检测的漏检率，证明了本文方法明显提高了网络对小尺寸车辆特征的提取能力，具有更好的检测性能。

4.4 本章小结

本章针对特征图深处小目标车辆的关注度不足，从而导致道路场景小目标车辆漏检的问题，详细介绍了设计的多尺度特征融合的道路场景图像小目标检测方法。首先介绍了基准 YOLOv11 模型；其次介绍了在骨干网络设计的 Focus 模块和多尺度特征融合模块，Focus 模块采用分块-拼接-卷积的策略，对输入特征图进行降维操作，能够保留关键信息和降低计算复杂度；多尺度特征融合模块通过区域残差化和语义残差化两步特征提取方式，增强了多尺度感受野特征和提高特征提取效率，能够有效的提升小目标检测的准确性，从而获得更高的检测精度；最后在颈部网络融入的双向特征金字塔模块通过多条双向路径和加权特征融合，使得模型能够更好地融合多层特征，从而提高复杂场景下小目标检测能力。经在 Cityspaces 数据集和 Foggy Cityspaces 数据集上的消融实验和可视化实验结果表明，设计的特征融合的道路场景图像小目标检测方法能够有效处理特征图远景小目标车辆漏检问题。基于去雾后的图像，本章所提出的方法，进一步增强了对小目标车辆的检测精度，对道路场景车辆以及小目标车辆检测，均展现出了显著的性能提升。

第 5 章 总结与展望

5.1 工作总结

雾天环境会严重影响智能驾驶汽车的视觉信号，当雾气存在时，拍摄出来的图像会因雾气不均匀分布出现颜色失真且会因颗粒物的遮挡导致图像原本的信息丢失，从而导致道路远景小目标车辆漏检，因此能够有效去除图像中的雾气且准确识别图像远景的小目标车辆则尤为重要。针对上述问题，本文针对现有的去雾算法高度依赖图像中的空间信息，从而导致复原图像存在颜色失真和伪雾残留的问题和去雾道路场景图像小目标车辆的漏检问题，提出了雾天条件下道路场景目标检测方法。具体的研究成果如下：

（1）针对目前主流的图像去雾算法高度依赖图像中的空间信息，从而导致复原图像存在颜色失真和伪雾残留的问题，本文提出了基于频域增强的道路场景图像去雾方法，并构建了结合组架构和特征注意模块的图像去雾网络模型。在组架构中设计了频域增强感知模块，能够更好的捕捉图像中图像的纹理和结构信息以及提高有效特征信息的权重，实现了对图像中雾气的有效去除；特征注意模块对同样的特征和像素进行非等权重处理，有效地处理了图像中雾气分布不均的问题。通过在 Foggy Cityspaces 数据集上的实验结果表明，所提方法可以有效去除图像雾气，提高图像的清晰度。

（2）针对去雾后的图像进行目标检测时，仍存在着小目标车辆漏检的情况，本文提出了基于特征融合的道路场景图像小目标检测方法，并构建了去雾道路场景图像的小目标检测模型。设计的模型在骨干网络增加了 Focus 模块，采用分块-拼接-卷积的策略，对输入特征图进行降维操作，能够保留关键信息和降低计算复杂度，有效地提升了计算速度；其次设计了多尺度特征融合模块，通过区域残差化和语义残差化两步特征提取方式，能够捕捉到更丰富的上下文信息，有效增强了网络的多尺度特征提取能力，从而提高小目标检测的准确性；在颈部网络融入了双向特征金字塔模块，通过多条双向路径和加权特征融合，使得模型能够更好地融合多层特征，从而提高复杂场景下小目标检测能力，进一步提升了小目标车辆检测精度。通过在 Cityspaces 数据集、Foggy Cityspaces 数据集上的实验结

果表明，有效解决了雾天场景下小目标车辆检测中漏检率高的问题。同时对 Foggy Cityspaces 数据集使用本文方法先进行去雾处理，得到去雾后的图像再进行目标检测。实验结果表明，本文方法有效的去除了道路场景图像中雾气，提高了对小尺寸车辆特征的提取能力，具有更高的检测精度。

5.2 工作展望

本文所提出的雾天条件下道路场景目标检测方法有效去除了图像中的雾气颗粒，并且对目标车辆的检测精度有所提升，在一定程度上降低了小目标车辆的漏检概率，但是仍然存在需要完善和改进的方面：

（1）本文只对雾天条件下道路场景的图像进行图像清晰化研究，但雾与霾有极大概率一同出现，因此下一步将研究雾霾条件下降质图像的道路场景目标检测方法。

（2）本文所研究的图像去雾算法和目标检测算法的识别是分开的，在目标检测之前先进行去雾图像处理，然后将去雾后的图像进行目标车辆的检测，而在图像去雾的这一阶段易出现去雾效果不佳的问题，因此在今后的研究中考虑将这两部分结合起来。

参考文献

[1] 罗熙媛，相萌，刘严严，等. 面向大气颗粒物干扰的图像清晰化算法研究与展望(特邀)[J]. 红外与激光工程, 2024, 53(08): 28-49.

[2] 廖苗，陆颜，张锦，等. 基于雾线和颜色衰减先验的图像去雾方法[J]. 通信学报，2023, 44(01): 211-222.

[3] Land E H. The retinex theory of color vision[J]. Scientific American, 1977, 237(6): 108-129.

[4] 曹子钰，李洋，江雪. 多尺度 Retinex 算法在雾天图像增强中的应用[J]. 中国新通信, 2016, 18(17): 84.

[5] 李益红，周晓谊. 一种多分辨多尺度的 Retinex 彩色图像增强算法[J]. 计算机工程与应用, 2017, 53(16): 193-198.

[6] Zhang J, He F, Chen Y. A new haze removal approach for sky/river alike scenes based on external and internal clues[J]. Multimedia Tools and Applications, 2020, 79: 2085-2107.

[7] 代少升，刘琴，王斐，等. 基于路径的 Retinex 算法在红外图像增强中的应用[J]. 半导体光电, 2015, 36(03): 482-485.

[8] 张杰，周浦城，张谦. 基于迭代多尺度引导滤波 Retinex 的低照度图像增强[J]. 图学学报, 2018, 39(01): 1-11.

[9] Chen Z, Wang L, Wang C, et al. Fog image enhancement algorithm based on improved Retinex algorithm[C]//2022 3rd International Conference on Electronic Communication and Artificial Intelligence (IEWCAI). 2022: 196-199.

[10] Zhang X Y, Min F, Pan S, et al. DD-net: Dual decoder network with curriculum learning for full waveform inversion[J]. IEEE Transactions on Geoscience and Remote Sensing (TGRS). 2024, 62: 1-17.

[11] 闫文强，崔蕾. 基于改进大气散射模型的图像去雾算法[J]. 现代电子技术, 2024, 47(23): 43-48.

[12] Weng S E, Miaou S G, Christanto R, et al. Exposure correction in driving scenes using the atmospheric scattering model[C]//2024 International Conference on Consumer Electronics-Taiwan (ICCE-Taiwan). 2024: 493-494.

[13] Nayar S K, Narasimhan S G. Vision in bad weather[C]//Proceedings of the Seventh IEEE International Conference on Computer Vision (ICCV). 1999, 2: 820-827.

[14] 殷旭平，钟平，薛伟，等. 通过风格迁移的浓雾天气条件下无人机图像目标检测方法[J]. 航空兵器, 2021, 28(03): 22-30.

[15] He K, Sun J, Tang X. Single image haze removal using dark channel prior[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI). 2010, 33(12): 2341-2353.

[16] Chu X, Chen L, Yu W. Nafssr: Stereo image super-resolution using nafnet[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2022: 1239-1248.

[17] Li Z, Liu F, Yang W, et al. A survey of convolutional neural networks: analysis, applications, and prospects[J]. IEEE Transactions on Neural Networks and Learning Systems (TNNLS). 2021, 33(12): 6999-7019.

[18] 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 (ICCV). 2017: 4770-4778.

[19] Chen D, He M, Fan Q, et al. Gated context aggregation network for image dehazing and deraining[C]//2019 IEEE Winter Conference on Applications of Computer Vision (WACV). 2019: 1375-1383.

[20] Dong H, Pan J, Xiang L, et al. Multi-scale boosted dehazing network with dense feature fusion[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2020: 2157-2167.

[21] Girshick R, 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 (TPAMI). 2015, 38(1): 142-158.

[22] He K, Zhang X, Ren S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI). 2015, 37(9): 1904-1916.

[23] Girshick R. Fast R-CNN[C]//Proceedings of the IEEE International Conference on Computer Vision (ICCV). 2015: 1440-1448.

[24] 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 and Machine Intelligence (TPAMI). 2016, 39(6): 1137-1149.

[25] 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 (CVPR). 2016: 779-788.

[26] Felzenszwalb P, McAllester D, Ramanan D. A discriminatively trained, multiscale, deformable part model[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2008: 1-8.

[27] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017: 7263-7271.

[28] Farhadi A, Redmon J. Yolov3: An incremental improvement[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2018, 1804: 1-6.

[29] Bochkovskiy A, Wang C Y, Liao H Y M. Yolov4: Optimal speed and accuracy of object detection[J]. arxiv preprint arxiv:2004.10934, 2020.

[30] Jocher G, Nishimura K, Minerva T, et al. YOLOv5[EB/OL]. (2022-11-22) [2025-3-10].

https://github.com/ultralytics/yolov5.

[31] Li C, Li L, Jiang H, et al. YOLOv6: A single-stage object detection framework for industrial applications[J]. arxiv preprint arxiv:2209.02976, 2022.

[32] 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 (CVPR). 2023: 7464-7475.

[33] Wang C Y, Yeh I H, Mark Liao H Y. Yolov9: Learning what you want to learn using programmable gradient information[C]//Proceedings of the European Conference on Computer Vision (ECCV). 2024: 1-21.

[34] Wang A, Chen H, Liu L, et al. Yolov10: Real-time end-to-end object detection[J]. Advances in Neural Information Processing Systems (NIPS). 2025, 37: 107984-108011.

[35] Khanam R, Hussain M. Yolov11: An overview of the key architectural enhancements[J]. arxiv preprint arxiv:2410.17725, 2024.

[36] 吕栋腾, 杨育良. 机器视觉下的雾天车行环境感知研究综述[J]. 自动化与仪器仪表, 2022, (11): 1-6.

[37] Runyu L. Pedestrian detection based on SENet with attention mechanism[C]//2023 IEEE 7th Information Technology and Mechatronics Engineering Conference (ITOEC). 2023, 7: 616-619.

[38] Yu H, Zheng N, Zhou M, et al. Frequency and spatial dual guidance for image dehazing[C]//Proceedings of the European Conference on Computer Vision (ECCV). 2022: 181-198.

[39] Nakajima M, Tanaka K, Hashimoto T. Neural schrödinger equation: Physical law as deep neural network[J]. IEEE Transactions on Neural Networks and Learning Systems (TNNLS). 2021, 33(6): 2686-2700.

[40] 孙福临, 李振轩, 梁允泉, 等. 基于改进 YOLOv5 算法和边缘设备的电动车违规载人检测[J]. 现代计算机, 2023, 29(08): 1-11.

[41] 孙航, 方帅领, 但志平, 等. 层级特征交互与增强感受野双分支遥感图像去雾网络[J]. 遥感学报, 2023, 27(12): 2831-2846.

[42] Li B, Ren W, Fu D, et al. Benchmarking single-image dehazing and beyond[J]. IEEE Transactions on Image Processing (T-IP). 2018, 28(1): 492-505.

[43] Cordts M, Omran M, Ramos S, et al. The cityscapes dataset for semantic urban scene understanding[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016: 3213-3223.

[44] Medasani S, Shireesha B, Reddy B H, et al. Enhanced single remote sensing image dehazing via vision transformer with silencing map transmission and advanced image processing techniques[C]//2024 International Conference on Intelligent Systems for Cybersecurity (ISCS). 2024: 1-6.

[45] Martini M. A simple relationship between SSIM and PSNR for DCT-based compressed

images and video: SSIM as content-aware PSNR[C]//2023 IEEE 25th International Workshop on Multimedia Signal Processing (MMSP). 2023: 1-5.

[46] Potapova K A. Comparative analysis of clustering algorithms based on pixel image processing using the SSIM metric and visual analysis[C]//2024 6th International Conference on Control Systems, Mathematical Modeling, Automation and Energy Efficiency (SUMMA). 2024: 290-295.

[47] 陈作钧, 秦品乐, 柴锐, 等. 微光场景下的时空域 EMCCD 视频降噪算法[J].中北大学学报(自然科学版), 2023, 44(02): 168-175.

[48] Bansode A K, Hati A J, Singh S K, et al. Refining ultrasound imaging quality using SSIM & advanced artifact removal methods[C]//2024 Global Conference on Communications and Information Technologies (GCCIT). 2024: 1-5.

[49] 蒋行国, 黎明. 基于多特征融合的图像修复算法[J]. 电子测量技术, 2024, 47(18): 80-88.

[50] Qin X, Wang Z, Bai Y, et al. FFA-Net: Feature fusion attention network for single image dehazing[C]//Proceedings of the AAAI Conference on Artificial Intelligence (AAAI). 2020, 34(07): 11908-11915.

[51] 李科岑, 王晓强, 林浩, 等. 深度学习中的单阶段小目标检测方法综述[J]. 计算机科学与探索, 2022, 16(01): 41-58.

[52] Shao Z, Liu X, Wei L, et al. Foggy road scene detection based on improved YOLOv8n[C]//2024 7th International Conference on Robotics, Control and Automation Engineering (RCAE). 2024: 417-421.

[53] Tan M, Pang R, Le Q V. Efficientdet: Scalable and efficient object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2020: 10781-10790.

[54] 李曜丞, 李喆, 许永鹏, 等. 基于立体注意力机制的输电线路图像识别算法[J]. 高电压技术, 2023, 49(08): 3437-3445.

[55] 高新波, 莫梦竟成, 汪海涛, 等. 小目标检测研究进展[J]. 数据采集与处理, 2021, 36(3): 27.

[56] Lin T Y, Maire M, Belungie S, et al. Microsoft COCO: Common objects in context[C]//Proceedings of European Conference on Computer Vision (ECCV). 2014: 740-755.

[57] Torralba A, Fergus R, Freeman W T. 80 million tiny images: A large data set for nonparametric object and scene recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI). 2008, 30(11): 1958-1970.

[58] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arxiv preprint arxiv: 1409. 1556, 2014.

[59] Xia G S, Bai X, Ding J, et al. DOTA: A large-scale dataset for object detection in aerial images[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern

Recognition (CVPR). 2018: 3974-3983.

[60] Yang S, Luo P, Loy C C, et al. Wider face: A face detection benchmark[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016: 5525-5533.

[61] Zhang S, Benenson R, Schiele B. Citypersons: A diverse dataset for pedestrian detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017: 3213-3221.

[62] Yu X, Gong Y, Jiang N, et al. Scale match for tiny person detection[C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). 2020: 1257-1265.

[63] 赵锟, 余添, 周立俭, 等. 基于CNN-Transformer的街景图像分类[J]. 青岛理工大学学报, 2023, 44(03): 146-152.

[64] Wang X, Wang C. Vehicle multi-target detection in foggy scene based on foggy env-YOLO algorithm[C]//2022 IEEE 7th International Conference on Intelligent Transportation Engineering (ICITE). 2022: 451-456.

[65] Zhang H, Li F, Liu S, et al. Dino: Detr with improved denoising anchor boxes for end-to-end object detection[J]. arxiv preprint arxiv:2203.03605, 2022.

[66] Cheng L, Zhang D, Zheng Y. Road object detection in foggy complex scenes based on improved YOLOv8[J]. IEEE Access, 2024: 107420-107430.

攻读硕士学位期间的研究成果

1. 学术论文

[1] 张梦雯, 桑建斌, 贾梦波, 等. 基于快速傅里叶变换和深度学习的图像去雾方法[J/OL]. 重庆工商大学学报(自然科学版), 1-10.

[2] 桑建斌, 张梦雯, 杨思远, 等. 改进 YOLOv8n 的复杂场景行人检测算法[J]. 传感技术学报.

[3] Sang J, Zhang M, Yang H, et al. YOLOv8-GDI: A lightweight YOLOv8 for real-time Pedestrian Detection[C]//2024 7th International Conference on Robotics, Control and Automation Engineering (RCAE). IEEE, 2024: 422-428.

致谢

行文至此，落笔为终。忆往昔，于学术之林踽踽独行，于知识之海泛舟求索，其间甘苦，如鱼饮水。三年的硕士生涯有探索未知的迷茫与困惑，也有收获成果的喜悦与满足。此刻，我怀着无比感恩的心，向所有给予我帮助和支持的人表达最诚挚的谢意。

首先，感谢我的导师杨会成教授，不仅在学术上对我严格要求，在生活中也给予了我诸多关心和帮助，杨老师严谨的生活态度和开阔的视野让我明白了做学问与做人的道理，我将永远铭记在心，这些教导与鞭策会使我受益终身。在此毕业之际，向我的导师杨会成教授表示最真挚的感谢。

其次，感谢实验室老师们，老师们和蔼可亲、认真负责，不仅在学习研究上提供了很多宝贵的意见和指导，在生活中也关怀备至，在此特别感谢胡耀聪老师，徐可老师，朱文博老师。

再者，感谢我的父母，在我硕士求学的这几年里，我的家人给予了我无条件的支持和无私的爱，他们在我遇到困难时给予我鼓励和安慰，每当我遇到挫折和烦恼时，他们总是耐心地倾听我的倾诉，为我排忧解难。我会努力让自己变得更加优秀，不辜负你们对我的期望。

另外，我还要感谢我的同门，师兄师弟，师姐和师妹以及朋友等，愿我们在未来的日子里，都能在各自的领域里发光发热，实现自己的人生价值。

对百忙之中对我的论文进行审阅以及答辩的专家老师表示衷心的感谢！

张梦雯

二〇二五年于安徽芜湖