分类号：TP391　　学校代码：10363
密　级：公开　　学　　号：2220910109

安徽工程大学

Anhui Polytechnic University

硕士学位论文

题目　低质图像增强及目标检测算法研究

论 文 作 者　龚佳乐
校 内 导 师　窦易文(副教授)
校 外 导 师　段中华(高级工程师)
专业学位类别　计算机技术
研究方向（领域）　智能系统与应用

论文提交日期：2025 年 5 月 9 日

分类号：TP391　　单位代码：10363
密　级：公开　　学号：2220910109

题　　目　低质图像增强及目标检测算法研究

英文并列

题　　目　**Research on low quality image enhancement and object detection algorithm**

学 生 姓 名：龚佳乐
校 内 导 师：窦易文（副教授）
校 外 导 师：段中华（高级工程师）
专 业 学 位 类 别：计算机技术
研 究 方 向：智能系统与应用

论文答辩日期：　2025　年　5 月 10 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：龚佳乐

日期：2025年5月9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：[illegible]　　指导教师签名：[illegible]

日期：2025 年 5 月 9 日　　日期：2025 年 5 月 9 日

低质图像增强及目标检测算法研究

摘 要

在信息技术飞速发展浪潮的推动下，伴随着互联网越发的普及、计算机技术的提升、半导体技术和多媒体技术的飞速发展，数字图像处理技术已经渗透到各行各业，并且发挥着不可替代的作用。低质图像的处理是数字图像处理中十分重要的一部分，它在许多领域都具有重要的研究意义和目的。低质图像是指在实际应用中，由于环境、拍摄设备等因素的影响，在拍摄过程中因外界原因导致这些图像普遍存在低对比度、高噪声、色彩失真的缺点，使得图像的利用率大大降低。因此，为尽可能恢复低质图像中的信息，提高图像的细节与亮度信息，识别出图像中的重点内容，并且满足人眼的观察需求，研究一种低质图像增强及目标检测算法有着十分重要的意义。

本文将低质图像处理任务分为三个步骤，分别是图像增强、图像降噪以及目标检测。其具体内容可概括如下：

（1）基于改进 Retinex 网络的图像增强算法

网络首先会使用分解网络将输入图像分解为两部分：光照分量以及反射分量。随后，将这两部分分别输入到特征融合网络以及残差网络中，同时使用特征融合网络增强光照分量，使用残差网络增强反射分量。在特征融合网络中，在下采样阶段有两个分路，分别为上路径和下路径，其中上路径采用 1×3 和 3×1 的卷积核代替传统的 3×3 卷积核，以减少网络参数并加速训练过程，同时保持性能。对于下路径，则采用传统的 3×3 卷积。在下采样完成后，将上路径和下路径的结果进行融合，通过上采样最终生成增强后的光照分量；在残差网络中，通过残差结构对反射分量进行增强，并获得增强后的反射分量；最后，在特征融合网络以及残差网络分别增强光照分量和反射分量后，通过光照分量与反射分量相乘获得最终的增强图像。

（2）一种鲁棒的噪声去除网络

使用双层的卷积神经网络对图像进行降噪处理，图像降噪网络分为三个部分，在第一个部分中利用残差连接防止信息衰弱，保留信息；第二个部分利用扩张卷积和普通卷积交叉使用的方法得到更大的感受野，从而在不额外引入参数和计算代价的情况下，获得除噪声外更多的原始图像特征信息，从而提高网络的特征保留能力；接着将第一个部分的输出特征图和第二个部分的输出特征图通过 Concat 模块拼接融合；最后，将得到的融合特征图输入到第三个部分中，第三部分运用了注意力机制，借助该机制神经网络能够自主学习并选择性地聚焦于输入信息中的关键部分，以此提升模型的性能以及泛化能力。实验结果表明所提网络可以有效地去除图像中的各种噪声。

（3）基于 YOLOv11 的小目标检测方法

以 YOLOv11 为基础，在主干网络中融入 Swin Transformer 模块，提高网络的抗复杂背景能力的同时提升了网络的细节信息提取能力，改善对小目标的误检、漏检等问题；其次，在主干网络的特征提取过程中集成了 CBAM 注意力机制，使得网络可以更有效地从图像中学习到更有用的特征，从而提升网络的性能；最后，将网络模型中用于预测边界框的损失函数从 CIoU（Complete Intersection over Union）替换为 Inner-IoU（Inner Intersection over Union）。这种改进提升了网络对小目标的检测能力，并使得模型能够输出更精准的结果。实验结果表明，经过这一改进，网络不仅能够实现较高的检测精度，还能更有效地识别出遮挡目标以及小目标。

关键词：Retinex 算法；扩张卷积；残差连接；注意力机制；YOLOv11；Swin Transformer

Research on Low-quality Image Enhancement and Object detection Algorithms

ABSTRACT

Driven by the rapid development of information technology, coupled with the increasing popularity of the Internet, advancements in computer technology, and the swift evolution of semiconductor and multimedia technologies, digital image processing technology has penetrated into various industries and plays an irreplaceable role. The processing of low-quality images constitutes a crucial part of digital image processing, carrying significant research importance and purpose across many fields. Low-quality images refer to those captured under poor conditions due to environmental factors, quality of photographic equipment, and other external influences during the shooting process. These images are typically characterized by low contrast, high noise, and color distortion, significantly reducing their usability. Therefore, to restore information in low-quality images as much as possible, enhance image details and brightness, identify key content within the images, and meet human visual observation needs, researching algorithms for low-quality image enhancement and object recognition holds great significance.

This paper divides the task of low-quality image processing into three steps, namely image enhancement, image denoising, and object detection. The specific contents can be summarized as follows:

(1) Image enhancement algorithm based on improved Retinex network

The network first uses a decomposition network to split the input image into illumination and reflection components. Then, these two parts are input into the feature fusion network and the residual network respectively. At the same time, the feature fusion network is used to enhance the illumination component, and the residual network

is used to enhance the reflection component. In the feature fusion network, there are two branches in the downsampling stage, which are the upper path and the lower path. In the upper path, 1×3 and 3×1 convolution kernels are used to replace the traditional 3×3 convolution kernel to reduce the network parameters and accelerate the training process while maintaining the performance. For the lower path, the conventional 3×3 convolution is used. After down-sampling, the results of the upper path and the lower path were fused, and the enhanced illumination component was finally generated by up-sampling. In the residual network, the reflection component is enhanced through the residual structure, and the enhanced reflection component is obtained. After the feature fusion network and the residual network enhance the illumination component and the reflection component respectively, the final enhanced image is obtained by multiplying the illumination component and the reflection component.

(2) A robust network for noise removal

Using a Convolutional Neural Network (CNN) for image denoising, the image denoising network is divided into three parts. In the first part, residual connections are utilized to prevent information degradation and preserve information. In the second part, a method combining dilated convolutions with ordinary convolutions is employed to obtain a larger receptive field, thereby acquiring denser information without introducing additional parameters and computational costs. Subsequently, the first and second parts are concatenated (Concat) to fuse the output feature maps from both parts. Finally, the fused feature map is input into the third part, which incorporates an attention mechanism. This part enables the neural network to automatically learn and selectively focus on important information in the input, improving the model's performance and generalization capability. Experimental results demonstrate that the proposed network can effectively remove noise from images.

(3) Small Object detection method based on YOLOv11

Based on YOLOv11, the Swin Transformer module is integrated into the backbone network to improve the anti-complex background ability of the network and improve

the detail information extraction ability of the network, and improve the problems of false detection and missed detection of small road targets. Secondly, the CBAM attention mechanism was integrated in the feature extraction process of the backbone network, so that the network could learn more useful features from images more effectively, thereby improving the performance of the network. Finally, the loss function of the predicted bounding box was replaced from CIoU to Inner-IoU to enhance the small object detection ability of the network model and obtain the output result with higher detection accuracy. The experimental results show that the network can have high detection accuracy, and can better detect occluded targets and small targets.

KEY WORDS: retinex algorithm; dilated convolution; residual connection; attention mechanism; yolov11; swin transformer

主要符号说明

$S(x,y)$：实际图像

$R(x,y)$：反射图像

$L(x,y)$：照度图像

$PSNR$：峰值信噪比

$SSIM\left(x,y\right)$：结构相似度

L_{MSE}：损失函数均方误差

f_{mAP}：平均精度均值

AP：精度值

P：准确率

R：召回率

FPS：每秒检测到的图像数量

目 录

第 1 章 绪论

1.1 研究背景及意义

低质图像增强及目标检测是图像处理领域中十分重要的研究课题。在本课题中，其目的是提高低质图像的图像质量，以帮助计算机从图像中提取出感兴趣的目标对象。主要步骤包括图像增强、图像降噪以及目标检测等。

视觉作为人类感知外界的关键渠道，其重要意义毋庸置疑。数字图像作为视觉信息的承载媒介，在信息传播领域占据核心地位。相较于文字与音频，图像能够以更为丰富且直观的形式展现信息。然而，在实际运用中，图像质量的不稳定状况时常对其效能的充分发挥形成限制。在图像采集进程里，成像环境的复杂特性是致使图像质量下滑的主要缘由之一。譬如，光照条件欠佳（像光照不足或光照不均）以及恶劣天气（如雨、雪、雾）等自然因素[1]，往往会引发图像对比度降低、画面模糊以及噪声增加等问题。此外，成像设备的硬件性能，诸如遥感传感器的分辨率等，也在一定程度上限定了图像中可获取信息的上限。在图像的存储与传输阶段，压缩算法的应用以及干扰信号的存在，同样会导致噪声的增加，进而降低图像的信噪比。低质图像可以用以下公式定义：

$$L=\left\{I\in S\left|\bigvee_{k=1}^{n}\left(P_k\left(I\right)\bullet T_k\right)\right.\right\} \tag{1-1}$$

其中，L 为低质图像集合，S 为原始图像数据集，P_k 代表着第 k 个质量评价函数，T_k 代表着对应第 k 个质量函数的临界值，$\bullet$ 代表着关系运算，具体根据质量函数而定，可能为 $>,<,\geq,\leq$ 等。具体而言就是从原始图像数据集中逐个读取图像进行 k 个质量函数的质量评价，当对于 I 有一个质量函数的值高于或者低于质量函数的临界值时，则认为 I 为低质图像。

理论上，借助性能更为优越的成像设备，在理想成像条件下可获取高质量图像。然而在实际操作过程中，鉴于使用场景的复杂程度以及成像环境的多变特性，该方法常常伴随着高昂成本。所以，当前人们更多地依赖数字图像处理技术，以一种更具成本效益的方式，有针对性地改善低质量图像。其中，图像增强作为计

算机视觉领域的关键课题，致力于针对各类图像退化问题，探究如何提升图像视觉效果、恢复图像细节信息并有效去除噪声。经增强处理后的图像，不仅视觉质量显著提高，其增强后的信息对于后续高级视觉任务，如目标跟踪、图像分类及目标检测等任务的性能提升亦具有重要的促进作用。

低质图像处理一直是计算机领域的主要研究热点，其重要性体现在多个方面。首先，在科学研究领域，如天文学、地质勘探和生物医学研究中，高质量的图像对于数据的准确分析至关重要。例如，在天文学中，通过图像增强技术可以清晰地识别远处的天体；在医学影像中，低质图像的改善有助于医生更准确地诊断疾病；其次，在工业生产中，低质图像处理技术可用于质量检测和机器人视觉，提高生产效率和产品质量。此外，在安全监控领域，低质图像的增强和复原技术能够帮助监控系统更清晰地识别目标，提升公共安全。

以往，低质图像处理主要借助手工设计算法，并融合相关领域知识及图像的先验统计信息来构建模型，进而实现高质量图像的恢复。然而，由于图像退化因素的复杂多样，这些传统方法往往难以全面捕捉图像特征，无法有效提取和还原图像中的关键信息。近年来，随着深度学习技术的快速发展，研究者们逐渐转向利用深度神经网络自动提取图像特征，以克服传统方法的局限性。特别是卷积神经网络（CNN）[2]和生成对抗网络（GAN）[3]等深度学习模型的出现，使得低质图像处理技术取得了显著的进展。借助复杂的网络架构与海量训练数据，这些模型可自行习得图像特征，进而生成质量更高且更逼近真实数据的图像。伴随人工神经网络深度的持续递增以及网络结构的不断优化，其所呈现出的学习与建模能力已明显超越传统方法，成为当下低质图像处理领域应用最为广泛的技术。例如，卷积神经网络通过卷积层、池化层和全连接层对图像进行特征提取和处理，能够捕捉到图像中的高级特征。生成对抗网络借由生成器与判别器的对抗训练，产出质量更高的图像。这些深度学习方法不仅提高了图像的视觉质量，还增强了低质图像在高层视觉任务中的性能，如目标跟踪、图像分类和目标检测等。

就实际应用场景而言，低质量图像常常难以满足后续目标检测之需求。譬如，在诸如雨、雪、雾等极端天气状况下所获取的图像，其对比度偏低、细节呈现模糊，整体质量欠佳，这在极大程度上对后续计算机视觉的应用形成了限制，特别是在户外导航、交通监控以及目标检测等领域；而在夜间环境中，鉴于光照条件

不足、光源形式单一以及拍摄设备自身的局限性，所采集的图像普遍存在对比度较低、噪声水平较高以及色彩出现失真等问题，这些问题对图像的可用性产生了严重的负面影响。例如，低照度条件下拍摄的图像往往细节缺失严重，噪声干扰明显，且色彩表现不准确。因此，开发更高效的低质图像处理算法具有重要的研究意义，尤其是在军事侦察、应急救援和公共安全等领域。这些算法需要能够有效提升图像的对比度、抑制噪声，并改善色彩失真，从而提高图像在复杂环境下的可用性。本文将针对目前低质图像处理过程中仍存在的一些问题（例如，低对比度、高噪声、细节模糊以及信息丢失等）进行深入研究。

1.2 国内外研究现状

在低质图像处理领域，国内外学者致力于开发高效的算法以挖掘低质图像中的潜在信息，取得了显著的成果，这些成果已被广泛应用于多个行业。本节将聚焦于低质图像处理中的三个任务——图像增强、图像降噪和目标检测，对国内外的研究方法及发展现状进行简要的梳理和总结。

1.2.1 图像增强算法研究现状

低照度图像增强方法大致可以分为传统算法与深度学习算法，而传统算法又可以分为基于直方图均衡化（Histogram Equalization，HE）的方法以及基于 Retinex 理论的方法。

基于直方图均衡化是一种经典的图像增强方法，其目标是将图像的灰度值分布调整为均匀分布，从而提高图像的对比度，包括经典 HE 算法[4]、自适应直方图均衡化算法（Adaptive Histogram Equalization，AHE）[5]以及限制对比度自适应直方图均衡化算法（Contrast Limited Adaptive Histogram Equalization，CLAHE）[6]等。基于直方图均衡化的方法主要侧重于通过调整图像的整体灰度分布来提升图像的对比度。然而，这种方法并不专门针对低光图像的照明信息进行优化。因此，在处理低光图像时，HE 方法可能无法充分利用和增强照明信息，从而导致图像的某些部分被过度增强，而另一些部分则增强不足，因此使用基于直方图均衡化的方法面临过度增强或增强不足的风险。

与之不同的是，Retinex 理论表明，物体的颜色并非由反射光的绝对强度决定，而是由物体对不同波长光线的反射能力决定，我们观察到的图像可以分为两部分：光照和反射率，物体的颜色由反射率决定。基于 Retinex 理论，学者们已

经提出了一系列的方法。单尺度 Retinex（SSR）[7]是 Retinex 理论的早期实现形式，它通过计算图像中每个像素的局部对比度来增强图像的视觉效果。然而，SSR 在处理复杂光照条件时存在局限性。为了解决这一问题，多尺度 Retinex（MSR）[8]被提出，它结合了多个尺度的 Retinex 处理结果，从而改善了单尺度算法的不足。尽管如此，SSR 和 MSR 直接去除光照分量并将反射率作为增强后的图像输出，这可能会破坏图像的自然性，并导致图像的色彩保持能力较弱。为了解决这一问题，彩色恢复多尺度 Retinex（MSRCR）被提出。MSRCR 在 MSR 的基础上引入了彩色恢复机制，能够更好地处理彩色图像，同时保持图像的色彩恒常性，从而在增强图像视觉效果的同时，避免了因去除光照分量而导致的色彩失真问题。虽然基于 Retinex 的方法在照明增强和细节处理方面具有良好的性能，但它们无法保持影响视觉感知的自然性。

深度学习技术在图像增强领域的发展为解决传统方法的局限性提供了新的思路。基于深度学习的图像增强方法通过利用深度神经网络模型来提升图像质量、增强细节，展现出显著的优势。近年来，随着深度学习技术的不断进步，研究者们开始将卷积神经网络（CNN）与 Retinex 理论相结合，这种融合方式不仅有效弥补了传统 Retinex 方法在处理复杂图像时的不足，还显著提升了图像增强的效果和效率，为图像增强技术的发展开辟了新的方向。例如，Chen 等人[9]提出了一种基于深度学习的 Retinex 分解方法-RetinexNet，通过联合训练 Retinex 分解网络和增强网络，实现了在低光照条件下图像的有效增强，提高了图像的质量和视觉效果；Jiang 等人[10]提出了一种新的低光照图像增强网络模型 DEANet++，该网络将 Retinex 理论与深度学习结合，并将 U-Net 和密集卷积网络（DenseNet）[11]融合到一起，以此增强分解后的光照分量以及反射分量，实现了更有效的特征提取和细节恢复；Wang 等人[12]提出了一种低光照图像增强模型 RSTU-Net，该模型通过整合 Swin Transformer 和类似 U-Net 的架构，增强分解后的反射分量和光照分量，实现了对低光照图像的高效增强；陈红阳等人[13]通过将 U-Net++与 Retinex 理论网络相结合提出一种图像增强网络 RLLENet，该网络首先通过 Retinex 理论将图像分解成反射分量以及光照分量，接着使用 U-Net++对反射分量进行增强，最后将增强后的光照分量以及反射分量融合获得最终图像。这些结

合 CNN 和 Retinex 理论的基于深度学习的方法，不仅提高了图像增强的效果，还为解决复杂图像问题提供了新的思路和方法。

目前，使用深度学习方法进行图像增强已经成为了最主流的方法，深度学习方法具有自适应性、高效性、实时性、灵活性和泛化能力，是性能最好的图像增强算法。

1.2.2 图像降噪算法研究

图像降噪是低质图像处理领域的一个重要任务，也是低质图像处理中的一个难点。图像降噪旨在减少或消除图像中的噪声，以改善图像质量，有利于对图像进一步的区分和解释。

（1）传统图像降噪算法：传统图像降噪算法主要基于图像的统计特性和先验知识，可以分为空域滤波与频域滤波，其中空域滤波直接在图像的像素值上进行操作，通过邻域像素的值来计算目标像素的新值。包括均值滤波[14]、高斯滤波[15]、双边滤波[16]以及中值滤波[17]等，高斯滤波通过高斯核对图像进行平滑处理，适用于高斯噪声；算术均值滤波用邻域像素的平均值替代中心像素值，适用于脉冲噪声；中值滤波取邻域像素值的中位数替代中心像素值，对椒盐噪声效果显著；双边滤波结合空间距离和像素值差异进行加权平均，能够在降噪的同时保持边缘。而频域滤波是一种通过各种变换将图像从空间域转换到频域进行处理的方法，包括傅里叶变换[18]、小波变换[19]等，傅里叶变换通过低通滤波去除高频噪声，适用于噪声与信号频谱不重叠的情况；小波变换的核心思想是利用其对图像和噪声统计特性的差异进行区分。图像经过小波变换后，其小波系数通常具有较大的幅值，且主要集中在高频区域，而噪声的小波系数幅值相对较小，并且均匀分布于变换后的所有系数中。因此，通过设定一个阈值，可以有效分离图像信号和噪声，从而实现降噪处理，最终得到更清晰的图像。

传统图像降噪方法虽然具有简单、无需训练数据等优点，但其对复杂噪声类型和高噪声水平的适应性较差，可能无法有效去除噪声，并且在处理的过程中还会出现信息丢失的问题。

（2）基于深度学习的图像降噪方法：随着深度学习技术在图像处理领域的广泛应用，其在图像降噪方面展现出的强大能力愈发受到关注。深度学习方法能够自动学习图像中的复杂特征，从而实现高效的噪声去除。特别是卷积神经网络

（CNN），作为深度学习在图像降噪中的核心工具，已被广泛应用于该领域。CNN通过大量的图像数据进行训练，能够学习到图像的特征和模式，进而利用这些知识去除噪声。此外，深度学习方法的一个显著优势在于其计算效率，借助 GPU的强大计算能力，基于深度学习的图像降噪方法在处理速度上表现出明显的优势。这些方法通常将原始图像和含噪图像作为一组数据进行训练，使卷积神经网络能够学习到如何从含噪图像中恢复出清晰的图像。2008 年，Viren 等人[20]首次将CNN 应用于自然图像降噪。他们设计的网络包含四个隐含层，每个层有 24 个特征通道，通过 5×5 的卷积核和逐层训练方法，以最小化降噪图像与真实图像之间的误差。尽管该方法在降噪效果上没有显著超越传统算法，但它开启了深度学习在图像降噪领域的应用先河。在最近几年，Wang 等人[21]提出了一种双分支网络 DulNet，用于图像降噪。该网络包含高分辨率分支和编码器-解码器分支，通过信息交换模块实现特征共享。同时，DulNet 引入感知损失功能，以提升降噪图像的视觉质量。Zhang 等人[22]提出了一种纹理引导的卷积神经网络 TDCNN。该网络通过提取纹理特征、优化特征细节，并将其转换为干净图像。并结合感知损失和均方误差的联合损失函数进一步提升了降噪效果，保留了图像细节。何涛等人[23]提出了一种基于卷积神经网络（CNN）的暗光图像降噪网络，该网络首先通过一个卷积层对输入图像进行特征均匀化处理，接着利用 4 个暗光增强与降噪模块对图像进行处理，这些模块结合了自编码器结构与残差学习单元，以去除噪声并保留图像细节。最后，通过单核卷积操作重构出不含噪声的正常光照图像作为输出。

这些基于 CNN 的降噪方法不仅在降噪效果上取得了显著提升，而且随着GPU 等硬件技术的发展，计算速度也得到了极大提高，使得这些方法在实际应用中更具优势。

1.2.3 目标检测算法研究

（1）双阶段目标检测算法：双阶段目标检测算法（Two-Stage Object Detection）是一种将目标检测任务分为两个主要阶段的方法。第一阶段为生成候选区域，第二阶段为对这些候选区域进行分类和边界框回归。R-CNN 系列算法是双阶段目标检测算法中的主流算法，在 2014 年，Girshick 等人[24]提出了 R-CNN 算法，该算法将目标检测问题转化为对候选区域的分类问题。首先使用选择性搜索算法

（Selective Search）生成候选区域，然后提取每个候选区域的特征，最后通过 SVM 分类器进行分类，并使用边界框回归器精确定位目标。该算法虽然显著提高了目标检测的准确性，但是其训练和测试速度慢。为了解决该问题，Girshick 等人[25]提出了 Fast R-CNN 算法，该算法引入了共享卷积层，避免了对每个候选区域重复计算特征。使用 RoI Pooling 层将候选区域的特征图统一缩放到固定大小，然后通过全连接层进行分类和边界框回归。这种做法使得该算法训练和测试速度显著提升，准确率也有所提高，但其仍然依赖于外部的候选区域生成算法，这成为了速度提升的瓶颈。为了解决这个问题，Girshick 等人[26]提出了 Faster R-CNN 算法，该算法引入了 Region Proposal Network（RPN）来替代 Selective Search，实现了候选区域的自动生成。RPN 与检测网络共享卷积层，进一步提高了检测速度。

双阶段目标检测算法虽然在检测精度上具有优势，但其计算效率低、参数调整复杂和实时性差等缺点限制了其在某些应用场景中的使用。

（2）单阶段目标检测算法：单阶段目标检测算法（One-Stage Object Detection）是一种在单次前向传播中完成目标检测任务的方法。与双阶段算法不同的是，这些算法直接从图像中预测目标的类别和位置，无需生成候选区域，从而显著提高了检测速度。SSD 算法[27]是单阶段目标检测算法中较为主流的算法之一，它结合了 Faster R-CNN 和 YOLO 的优点，通过在多个尺度的特征图上进行预测，实现了高精度和高效率的目标检测。然而，SSD 算法由于在特征图上采用稀疏分布的锚点，对小目标和多目标的检测不够准确。相比之下，YOLO 系列算法取得了更好的检测性能。You Only Look Once[28]、YOLOv2[29]、YOLOv3[30]、YOLOv4[31]、YOLOv5 是 YOLO 系列的早期算法。在近年来，许多研究者在这些算法的基础上做了改进并提出了更好的 yolo 算法。在 2022 年，Wang 等人[32]在 YOLOv5 的基础上提出了 YOLOv7 算法，该算法在 YOLOv5 的基础上做了诸多改进，引入了扩展高效的层聚合网络（E-ELAN）以增强模型的特征提取能力、采用基于级联的模型缩放方法优化不同尺寸模型的参数和计算量；通过卷积重参化、Batch normalization 等技术提升训练精度；改进标签分配策略，提出由粗到细的引导式分配方法，以更好地处理多输出层的目标分配，这些改进使 YOLOv7 在实时目标检测中实现了更高的精度和效率。次年，Varghese 等人[33]在 YOLOv7 的基础

上提出了 YOLOv8，YOLOv8 采用改进后的 C2f 模块作为 Backbone，提升了特征提取效率，并且引入 PAN-FAN 结构和 SPPF 模块增强多尺度特征融合，除此之外还采用解耦头设计和 Anchor-Free 机制并在此基础上优化了损失函数，进一步提高网络的性能。时隔一年，Wang 等人[34]提出了 YOLOv9 算法，YOLOv9 主要改进为可编程梯度信息（PGI）和广义高效层聚合网络（GELAN），PGI 通过辅助可逆分支和多层次辅助信息，为目标任务提供完整的输入信息以计算目标函数，从而获得可靠的梯度信息来更新网络权重。GELAN 则是基于梯度路径规划设计，通过结合不同计算模块，优化参数、计算复杂度、准确性和推理速度。同年 5 月，Wang 等人[35]提出了 YOLOv10 算法，该算法通过一致性双重分配策略实现无 NMS 的端到端识别，优化了模型架构并引入轻量级分类头和空间-通道解耦下采样，同时结合大核卷积和部分自注意力模块增强特征提取，从而提升了检测精度和效率。同年 10 月，Khanam 等人[36]提出了 YOLOv11 算法，该算法引入了 C3k2 块、SPPF 和 C2PSA 组件。这些组件的引入增强了特征提取能力，提升了模型在多种计算机视觉任务上的性能，并以此实现了更高的平均精度。这些改进算法使得 YOLO 系列在工业应用、自动驾驶、智能安防等多个领域得到了广泛应用。

1.3 研究内容和创新点

在数字图像处理领域中，图像的质量尤为重要。低质的图像不仅影响观感，还会影响后续高级任务的应用，例如目标跟踪、图像分类以及目标检测等。因此需要先对低质图像进行增强，在进行目标检测等任务。具体而言，首先使用 Retinex 理论将图像分解为光照分量和反射分量，并使用特征融合网络分别对这两个分量进行增强并融合，从而提高对低质图像的增强能力。接着对前人提出的 CEANet 网络模型进行了优化修改，并获得更好的提取特征能力，从而获得更好的降噪效果，最后对 YOLOv11 进行改进，在改进的 YOLOv11 中加入 Swin Transformer 模块以及 CBAM 注意力模块，使得网络具有更好的特征提取能力，可以更好的捕获图像中不同尺度特征和语义信息。具体研究路线如图 1-1 所示。

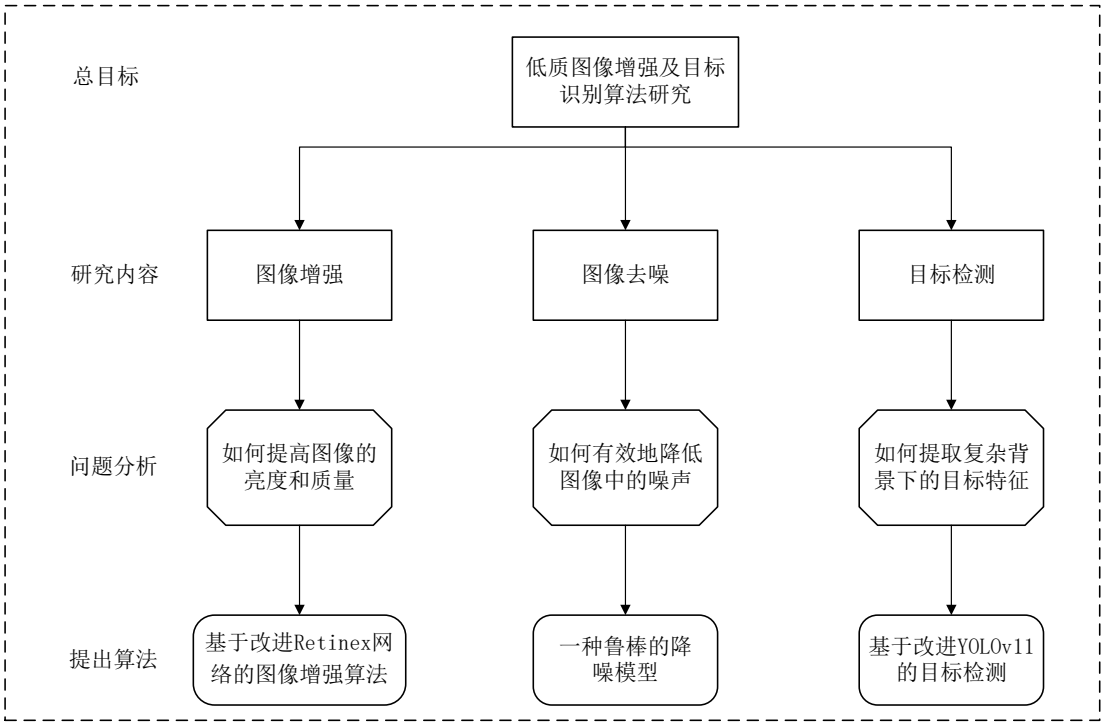


图 1-1 研究内容框架图

Fig.1-1 Research content framework diagram

1.4 本文章节安排

本文分为六个章节，其中第 3 至第 5 章为论文主体，每个章节的安排如下：

第 1 章是绪论，主要介绍了低质图像处理的研究背景与意义；介绍了当前低质图像处理的主要应用领域，对低质图像处理的流程进行一定的介绍，并分析每一步中不同经典算法的优缺点，最后介绍了本文的工作研究重点和组织架构。

第 2 章为低质图像处理相关理论及技术，本章主要介绍了低质图像增强，图像降噪以及目标检测相关的理论基础，重点介绍了 Retinex 算法，YOLO 算法以及卷积神经网络，为后续分别在此基础上进行改进做铺垫。

第 3 章提出了一种基于改进 Retinex 网络的图像增强算法，本章节主要介绍了改进的 RetinexNet 算法的设计思想、优化及求解过程，将 RetinexNet 中对光照分量以及反射分量的处理替换，使用特征融合网络处理信息，从而获得更好的结果。

第 4 章提出了一种鲁棒的噪声去除网络，本章主要介绍了图像降噪网络的结构以及处理结果。图像降噪网络分为三个部分，在第一个部分中利用残差连接防止信息衰弱，保留信息；第二个部分是利用扩张卷积和普通卷积交叉使用的方法

得到更大的感受野，从而在不额外引入参数和计算代价的情况下，获得更加密集的信息；接着将第一个部分和第二个部分通过 Concat 拼接起来，将两个部分的输出特征图融合起来；最后，将得到的融合特征图输入到第三个部分中，第三个部分使用了注意力机制，该部分使得神经网络能够自动地学习并选择性地关注输入中的重要信息，提高模型的性能和泛化能力。

第 5 章提出了面向复杂交通环境下的小目标检测方法，本章主要对于 YOLOv11 算法进行改进，首先介绍了传统的 YOLOv11 算法，接着在 YOLOv11 网络中的骨干网络加入 Swin Transformer 以及 CBAM 注意力机制，从而提高模型的特征提取能力。通过一系列实验说明了所提方法的有效性，且所提网络具有较高的鲁棒性。

第 6 章是本文的总结与展望。本章对全文的研究内容和主要工作进行了概括，详细阐述了对于低质图像处理的一系列流程，即图像增强，图像降噪，目标检测，并详细说明了对于每一个步骤的相关改进。此外，本章还分析了低质图像处理方面有待提高的内容，并对未来的研究方向进行了展望。

第 2 章 低质图像处理相关基础与预备知识介绍

本章旨在介绍低质图像处理的相关基础理论和关键技术，以便读者理解后续章节的内容。

2.1 图像增强

低质图像的增强作为图像处理的关键技术之一，可以通过计算机对图像进行处理，改善图像的质量、清晰度和视觉效果。通过增强处理，可以使低质图像变得更加逼真、细节更加清晰，提升低质图像的可视化效果和信息传达能力，以便于后续计算机视觉方法的应用。

2.1.1Retinex 理论

Retinex 理论[37]是一种模拟人类视觉系统的图像处理理论，其核心思想是将图像分解为光照分量和反射分量。Retinex 原理如图 2-1 所示：

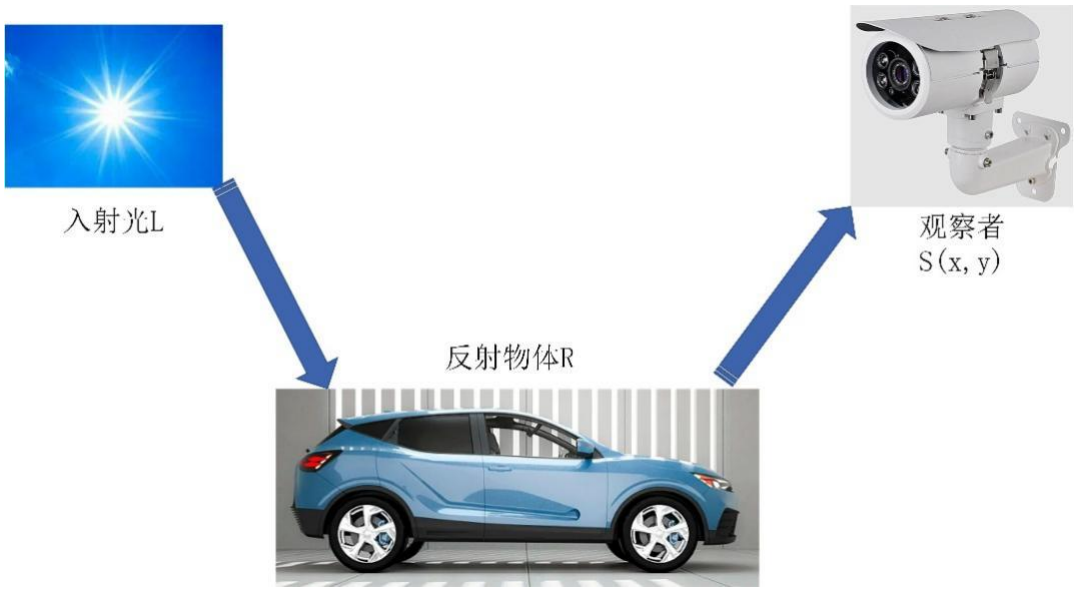


图 2-1 Retinex 物理模型示意图

Figure.2-1 Schematic diagram of the Retinex physical model

这种理论最初由 Edwin.H.Land 在 1977 年提出，旨在解释人类视觉系统如何在不同的光照条件下保持对颜色和亮度的恒常性感知。具体来说，Retinex 理论认为，我们所察觉到的图像色彩与亮度，是由物体表面的反射比率以及光照状况共同决定的。通过分离这两部分，可以更好地模拟人类视觉系统对场景的感知，从而实现图像的增强和复原。其表达式如下：

$$S(x,y)=R(x,y)\times L(x,y) \tag{2-1}$$

其中，(x,y)表示图像像素空间的二维坐标，$S(x,y)$是观察到的实际图像，$R(x,y)$是反射图像，其对图像基本属性起决定性作用，对应图像高频部分，蕴含大量图像细节信息。$L(x,y)$是照度图像，它确定图像的动态范围，对应图像的低频部分。Retinex 算法的核心要义在于从目标图像里消除照度图像的影响，提取反射光图像，进而获取图像的本质特性，以此达成图像增强的目的。

传统 Retinex 算法会把图像的乘性分解转变为对数域中的加性运算。此转换不但使计算流程得以简化，显著降低了算法的时间复杂度，而且契合人类视觉系统对信息的感知模式。借助对数运算，能够有效拉伸低动态范围的像素值，同时压缩高动态范围的像素值，进而更优地模拟人类视觉对于光照及反射的感知特性。转换后的公式呈现如下：

$$\log(S(x,y))=\log(R(x,y)\times L(x,y))=\log(R(x,y))+\log(L(x,y)) \tag{2-2}$$

Retinex 理论在图像增强领域有广泛的应用，特别是在处理光照不均匀、色偏以及低光照条件下的图像时表现出色。例如，在医学影像中，Retinex 算法可以用于矫正视网膜图像中的光照不均问题；在安防监控中，它可以帮助去除航空影像中建筑物的阴影；在自动驾驶领域，Retinex 算法可以改善夜间成像效果，提升图像的整体亮度。

2.1.2 特征融合网络

特征融合网络是一种通过整合来自不同来源或不同层次的特征，以增强模型性能和准确性的技术。其核心思想是将低层特征（高分辨率、细节丰富但语义信息弱）与高层特征（低分辨率、语义信息强但细节缺失）进行有效结合，从而提升模型对复杂模式的捕捉能力。特征融合网络从融合方式上可以分为两种，第一种方式是融合同一个路径上不同阶段的特征，也就是融合串联的特征，这样做的优点是融合方式更加灵活，但其会耗费更多的计算资源；第二种方式是融合不同子网络的输出特征，这样做的优点是可以独立训练不同的子模型，比较稳定，但是其灵活性较差，并且存在各个子模型的输出特征相关性不高的问题。特征融合最重要的是充分挖掘特征之间的互补性和相关性，从而提升模型的表现和对任务的理解能力，因此需要结合具体场景合理设计特征融合的方法。

2.2 图像降噪

图像降噪是图像处理中的一个重要环节，旨在从受噪声污染的图像中恢复出原始图像。去除噪声后，图像的细节和结构更加清晰，色彩更加自然，从而提升图像的视觉效果，使其更符合人类视觉感知的需求。除此之外，在许多图像处理和计算机视觉任务中，高质量的图像数据是关键。降噪后的图像可以更好地支持后续的处理步骤，如特征提取、目标检测、图像分割等。

2.2.1 扩张卷积

扩张卷积（Dilated Convolution）也被称为空洞卷积或者膨胀卷积，是在标准的卷积核中注入空洞，以此来增加模型的感受野[38]。在卷积神经网络中可以使用扩张卷积替换掉普通卷积，从而在不增加额外的计算代价的情况下获得更多的特征信息。相比正常的卷积操作，扩张卷积多了一个参数，也就是扩张因子 dilation rate，指的是卷积核中点的间隔数量。一般来说，增加网络的深度以及宽度是卷积神经网络较为常用的增加感受野的方法，但是，增加深度会导致网络的细节丢失，导致特征提取的效果变差，而增加宽度则会增加网络的参数数量，大大提升了网络计算的复杂度，从而导致运算时间增加。使用扩张卷积替换标准卷积就可以较好的解决这些问题，例如，扩张因子为 2 的扩张卷积的卷积核大小为(4n+1)×(4n+1)，n 为卷积神经网络的深度，假设某卷积神经网络的 n 为 5，那么该卷积神经网络的感受野为 21×21，这个感受野大小与使用标准卷积的卷积神经网络在第十层时的感受野大小相同。因此，使用扩张卷积可以有效地减少网络的深度，可以在不增加额外参数的情况下，提取到更深层次的特征信息，如图 2-2 所示。

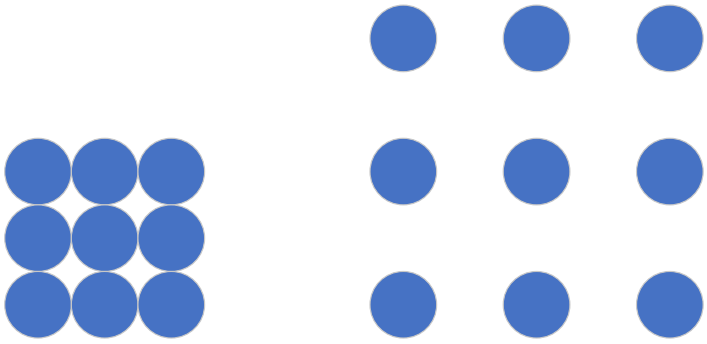


图 2-2 扩张率为 1 (左) 以及 扩张率为 2(右)

Figure.2-2 The dilation rate is 1 (left) and the dilation rate is 2(right)

图 2-2 左侧是标准卷积的卷积核，大小为 3×3。扩张卷积就是在这个标准卷积之间加入间隔，图 2-2 右侧就是对应的大小为 3×3，dilation rate 为 2 的扩张卷积核，间隔为 1，也就是相当于感受野为 5×5，虽然感受野变大了，但由于只有九个点有参数，其余点都为 0，在进行计算的过程中，参数为 0 的点都会跳过计算，因此这个扩张卷积只有 3×3 卷积的参数，却有 5×5 的感受野。

2.2.2 注意力机制

在图像降噪的过程中，网络算法的特征提取能力以及选择合适特征信息的能力是十分关键的[39-41]。然而，在背景复杂的图像中提取特征是十分困难的[42]，为了让网络可以提取到相对重要的特征信息，可以通过使用注意力模块来指导网络进行降噪。通常来说，注意力机制有两种方式可以实现，分别是不同子网络之间的注意力以及同一个网络中不同分支之间的注意力。不同子网络之间的注意力涉及多个独立的子网络，每个子网络负责处理输入数据的不同部分或特征。通过注意力机制，这些子网络之间可以相互协作，共享信息，从而提高整体的处理效果。同一个网络中不同分支之间的注意力涉及同一个网络中的多个分支，每个分支负责处理输入数据的不同特征或不同层次的特征。通过注意力机制，这些分支之间可以相互协作，共享信息，从而提高特征提取的效率和准确性。通常由通道注意力、空间注意力以及混合注意力实现，也是本文主要使用的注意力机制。

2.3 目标检测

目标检测是计算机视觉领域的一个关键任务，旨在从图像或视频中识别并精确定位特定的目标物体。这一过程通常通过边界框或分割掩码来标注目标的具体位置。目标检测不仅需要准确判断目标的类别，还要求精确确定其在图像中的位置信息。这项技术在众多领域都有着重要的应用价值，能够有效提升系统的自动化和智能化程度，同时显著增强效率和安全性。

2.3.1 YOLOv11

YOLOv11 是一种单阶段目标检测算法，算法将整个目标检测任务建模为一个端到端的神经网络模型。该网络模型由 Backbone、Neck、Head 三个部分组成，模型规模从小到大有 YOLOv11n、YOLOv11s、YOLOv11m 等多个版本。YOLOv11 的主干特征提取网络由 Conv、C3K2 以及空间金字塔池化（SPPF，Spatial Pyramid Pooling Fast）等结构组成。其中，C3K2 模块通过分割特征图并应用一系列较小

的卷积核来处理分割后的较小特征图，接着在几次卷积后合并它们，从而获得更快的处理速度以及更少的计算成本;Neck 的结构包 Conv、C3K2 以及 C2PSA 等。通过这些结构之间的相互作用可以整合不同分辨率的特征图，从而提高目标检测模型对不同尺度目标的检测性能。YOLOv11 中的 Head 是负责生成目标检测结果的部分，该部分将 3 种不同大小的特征图输入到识别模块中进行识别，这 3 种不同大小的特征图分别对应着大、中、小 3 种体型的目标，以此来实现不同尺度的目标检测。

2.3.2 Swin Transformer

Swin Transformer 是一种应用在视觉领域的 Transformer[43]结构，这种结构可以作为计算机视觉中神经网络的通用主干。Transformer 最初的作用是处理自然语言，将其应用到视觉领域的挑战主要来自于两个领域之间数据空间结构的不同以及长距离依赖导致的处理效率与模型性能下降的问题。为了解决这些问题 Swin Transformer 采用了分层式的注意力处理方法，该方法首先将图像划分成若干小块，随后在图像的多个层级上分别对这些小块执行注意力运算，以此方式来捕获图像中不同尺度的信息细节。这种分层结构有助于处理图像数据的空间结构，使得 Swin Transformer 能够更好地适应视觉任务；除此之外，Swin Transformer 还引入了窗口化的注意力机制，也就是将图像分成不重叠的块，并在每个块内计算注意力，从而实现局部性的信息提取，这有助于解决长距离依赖性问题，使得 Swin Transformer 能够更好地处理图像数据中的局部内容。此外，移位窗口技术不仅将自注意力的计算范围限定在互不重叠的局部窗口内，还通过引入跨窗口的连接来增强信息流通，进一步提升了效率。Swin Transformer 凭借其分层架构设计，能够灵活地在不同尺度上进行建模，并且拥有相对于图像大小的线性计算复杂度。这些特性使得 Swin Transformer 可以兼容很多的视觉任务，包括密集预测任务与图像分类任务，Swin Transformer 由多层感知机（MLP，Multilayer Perceptron）、层归一化（LN，LayerNorm）[44]、窗口多头自注意力层（W_MSA）以及滑动窗口多头自注意力层（SW_MSA）组成。Swin Transformer 的主要结构如图 2-3 所示。

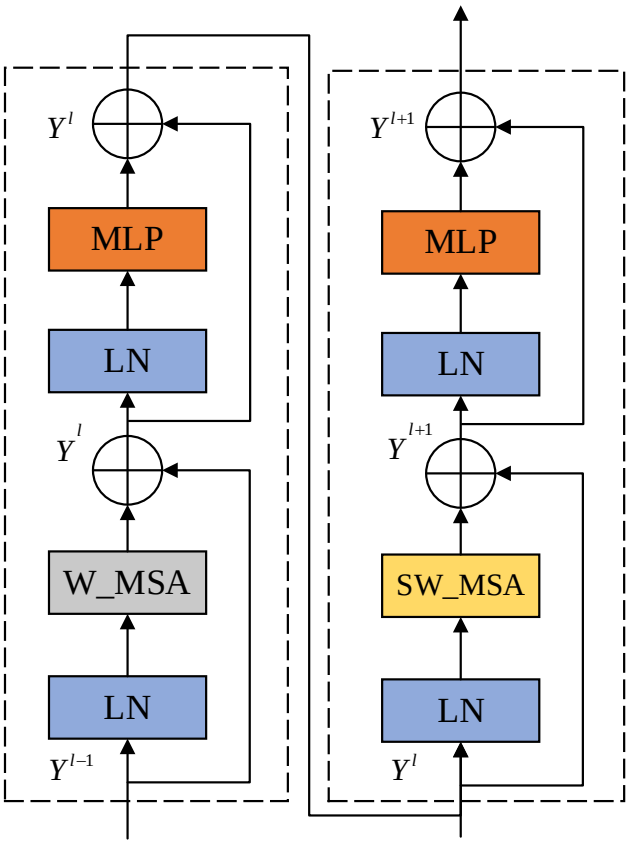


图 2-3 Swin Transformer 结构

Figure.2-3 Swin Transformer architecture

在 Swin Transformer 的窗口自注意力块中，首先对输入特征进行层归一化处理，然后通过自注意力层计算窗口内的注意力，并通过残差连接融合不同层次的特征，接着对融合后的特征再次进行层归一化，最后通过多层感知机进行非线性变换得到最终特征表示。其中，窗口自注意力的计算包括：窗口划分，线性变换得到查询向量、键向量以及值向量，计算每个窗口的自注意力，将每个窗口内得到的注意力结果融合到原始特征中。移动窗口自注意力块与窗口自注意力块类似，但是在计算自注意力时采用了滑动窗口的方式。这意味着在计算自注意力时不仅要考虑当前窗口内的信息，还需要考虑相邻窗口之间的交互关系，以更好地捕获局部区域之间的依赖关系。Swin Transformer Block 中各组件的输出表达式可以用以下公式进行表达：

$$Y^{l} = W\_MSA[LN(Y^{l-1})] + Y^{l-1} \tag{2-3}$$

$$Y^{l} = MLP(LN(Y^{l})) + Y^{l} \tag{2-4}$$

$$Y^{l+1} = SW\_MSA[LN(Y^{l})] + Y^{l} \tag{2-5}$$

$$Y^{l+1} = MLP(LN(Y^{l+1})) + Y^{l+1} \tag{2-6}$$

其中，Y^{l-1}代表 Swin Transformer 的输入，Y^{l}代表第 1 个残差后的输出，Y^{l}代表第 2 个残差后的输出，Y^{l+1}代表第 3 个残差后的输出，Y^{l+1}代表第 4 个残差后的输出，W_MSA代表窗口多头自注意力层中所做的操作，MLP代表多层感知机层中所做的操作，SW_MSA代表滑动窗口多头自注意力层中所做的操作，LN代表着层归一化。

2.3.3 边界框损失函数

在进行目标检测的过程中，预测边界框的损失函数选取是十分关键的。最初的预测边界框损失函数是 IoU，即真实边界框与预测边界框之间的交并比，然而，当真实边界框与预测边界框之间没有交集时将无法计算梯度。为了解决这个问题，Hamid 等人[45]提出了 GIoU（Generalized IoU）算法，该算法通过引入预测边界框和真实边界框的最小外接矩形来获取预测边界框、真实边界框在闭包区域中的比重，从而解决了两个目标没有交集时梯度为零的问题。尽管如此，预测边界框与真实边界框重叠时的情况仍然无法判断。为了解决这个问题，Zheng 等人[46]提出了 DIoU（Distance-IoU）算法思想，该算法通过加入预测边界框中心点与真实边界框中心点的欧式距离增加对预测边界框和真实边界框重叠度的判断。然而 DIoU 仍无法很好的表示图形长宽比的相似性，为此在 DIoU 的基础上进一步提出了 CIoU[47]（Complete-IoU），CIoU 在 DIoU 的基础上添加了长宽比的惩罚项，并且考虑了边界框的中心点距离和对角线距离，因此可以更准确地比较真实边界框与预测边界框之间的相似性。然而，CIoU 有个很大的缺点，那就是深度神经网络在更新参数时预测边界框的宽和高每次都会向相反的方向变化。为了解决这个问题，Zhang 等人[48]提出了 EIoU（Efficient-IoU）算法，该算法在 CIoU 的基础上做出了改进，将预测边界框和真实边界框的纵横比影响因子分成两个项来表达，分开计算预测边界框和真实边界框的长和宽，以此获得更好的输出结果，但是 EIoU 在小目标的检测任务中表现较差。在 2022 年，Gevorgyan 等人[49]提出了 SIoU（Soft-IoU）算法，该算法考虑了角度和尺度敏感性，通过角度损失、距离损失和形状损失三个部分来优化边界框损失，使得 SIoU 损失函数可以较好的针对图像中的小目标。但由于其对尺度变化的敏感性不足以及缺乏动态调整机制，

SIoU 无法应对目标尺度变化较大的情况。因此，zhang 等人[50]提出了 Inner-IoU 算法，该算法在 IoU 的基础上引入辅助边界框的概念，通过缩放因子生成不同大小的辅助边界框来计算损失。Inner-IoU 可以针对不同 IoU 值的样本进行优化，从而提高模型对小目标的检测能力。

2.4 客观评价指标

在图像增强以及图像降噪的客观评价中采用峰值信噪比（PSNR）以及结构相似性（SSIM）做为客观评价指标，其中峰值信噪比以及结构相似度的定义分别可以表示为：

$$PSNR = 10 \times log_{10}\left[\frac{\left(2^n - 1\right)^2}{L_{MSE}}\right] \tag{2-7}$$

$$SSIM\left(x, y\right) = \frac{\left(2\varepsilon_x\varepsilon_y + d_1\right)\left(2\sigma_{xy} + d_2\right)}{\left(\varepsilon_x{}^2 + \varepsilon_y{}^2 + d_1\right)\left(\sigma_x^2 + \sigma_y^2 + d_2\right)} \tag{2-8}$$

其中 L_{MSE} 为公式的内容，x 和 y 为 2 幅做运算的图像，ε_x 和 ε_y 为均值，σ_x^2 和 σ_y^2 为方差，σ_{xy} 为协方差，$d_1 = (k_1 L)^2$，$d_2 = (k_2 L)^2$，L 是像素值的动态范围，$k_1 = 0.01$，$k_2 = 0.03$。

在目标检测的客观评价中采用精度指标以及速度指标作为评价指标。精度指标为 mAP，指的是平均精度均值；速度指标为 FPS，指的是每秒可识别到的图像数量。其中 mAP 可以用下列公式表达：

$$f_{mAP} = \frac{1}{N}\sum_{i=1}^{N} f_{AP_i} \tag{2-9}$$

其中，N 为目标的类别数，f_{AP_i} 是指第 i 个类别准确度的平均值，其可以用以下公式表达：

$$f_{AP_i} = \frac{1}{n}\sum_{j=1}^{n} P_{i(R_j)} \tag{2-10}$$

其中，n 表示拥有召回率区间的数量，$P_{i(R_j)}$ 表示对第 i 类目标进行识别的结果中，第 j 个召回率区间下准确率的最大值，其中 P 和 R 可以用以下公式进行表达：

$$P = \frac{N_{TP}}{N_{TP} + N_{FP}} \tag{2-11}$$

$$R = \frac{N_{TP}}{N_{TP} + N_{FN}} \tag{2-12}$$

其中，N_{FN} 表示被错误识别成负样本的正样本数量，N_{FP} 表示被错误识别成正样本数量的负样本数量，N_{TP} 表示被正确识别的正样本数量。

第 3 章 基于改进 Retinex 网络的图像增强算法

图像增强是指通过一系列技术手段对图像进行处理，旨在提升图像的视觉质量或突出图像中的特定特征，以便更好地满足后续分析和处理的需求。在低质图像的处理与检测中，图像增强技术扮演着极为关键的角色，在很多领域中，低质图像增强算法都有着广泛的应用，例如：在安防监控中，由于摄像头拍摄环境复杂，光线不足、运动模糊等因素会导致图像质量较差。通过低质图像增强技术，可以提升监控图像的清晰度和细节表现，帮助监控人员更准确地识别监控画面中的内容；在工业检测领域中，在工业生产中，由于设备老化、环境复杂等原因，拍摄的图像可能存在模糊、分辨率低等问题，使用低质图像增强算法可以有效提升工业场景图像的清晰度，保留图像细节。低质图像增强算法能够有效改善低质图像的视觉表现，提升图像的对比度和清晰度，进而凸显图像中的重要特征，为后续的特征提取和目标检测任务奠定更坚实的基础。

为了获得更好的图像增强结果，研究者们采用了很多方法。早期具有代表性的方法主要是基于直方图均衡化的方法，此类方法通常通过调整直方图的分布来满足一定的约束条件，从而增强对比度[4,5]。基于直方图均衡化的方法简单有效，但也存在放大噪声或过度增强对比度等缺点。除此之外，基于 Retinex 的图像增强算法也被广泛使用，这些方法虽然简单有效，但是会出现色彩不自然等问题。除了传统算法外，目前最受欢迎的图像增强算法是深度学习算法，Huang 等人[9]构建了一种依托注意力机制与 Retinex 模型的低照度图像增强网络。此方法初始借助注意力机制对图像的光照图予以预测，并经卷积层校正色彩失真与噪声，进而获取增强后的图像。尽管该方法于处理局部光照变化及维持色彩保真度方面成效显著，然而增强后的图像里仍存在细节遗失的状况。李佳等人[51] 提出一种基于 Retinex 理论且具备双重注意力的 Transformer 增强网络——DARFormer，该网络由光照估计网络与损坏修复网络两部分构成。光照估计网络基于图像先验来估算亮度映射项，以实现低光照图像的亮度增强；损坏修复网络则通过运用带有空间注意力和通道注意力的 Transformer 架构来优化亮度增强后的图像质量，但是其增强后的结果存在自然性较差的问题。Priyadharshini 等人[52]提出了一种用于水下图像增强的深度学习方法，该方法先对水下图像进行白平衡、伽马校正和

CLAHE 预处理，改善亮度和色彩；再输入到改进的 Water-Net CNN，通过卷积层和特征转换块提取优化特征，减少色彩偏差并增强结构；最后将特征图与预处理图像融合，生成增强图像，并利用感知损失函数优化网络，提升图像质量。但是其增强后的结果存在噪声较多的问题。

本节基于 Retinex 算法，将 Retinex 算法与深度学习方法结合，提出一种用以增强低质图像的增强网络-FFRRetinexNet，主要改进有以下：

1.提出使用特征融合网络与 Retinex 相结合的方法，并使用特征融合网络处理 Retinex 分解网络分解出的光照分量。

2.使用残差网络对反射分量进行增强，从而获得更好的增强结果。

3.在特征融合网络的融合过程中加入注意力机制，使得网络可以自主的关注重要的信息，提高网络的特征提取能力。

3.1 RetinexNet 网络

在本小节中探讨了 RetinexNet 这一基础网络。作为首个将 Retinex 理论与深度学习相结合的算法，RetinexNet 为后续的图像增强算法奠定了基础，本章节所提出的增强算法也是在此基础上进行优化和改进的。图像的细节丢失和低对比度问题不仅会给人带来不好的视觉体验，还会对许多基于正常光照条件设计的计算机视觉算法的性能产生负面影响。而 RetinexNet 算法能够有效提升图像的对比度，同时增强图像的细节信息。

RetinexNet 由两个核心模块组成：分解网络（DecomNet）和增强网络（RelightNet）。分解网络用于将输入图像分解为反射分量和光照分量。在训练过程中，即使缺乏反射分量和光照分量的“真实标签”（Ground Truth），但通过一些关键的约束条件，例如低光照和正常光照图像共享的反射率一致性，以及光照分量的平滑性，分解网络依然能够顺利地进行训练。在分解网络的基础上，增强网络进一步对光照分量进行亮度增强，并对反射分量进行去噪处理。最后通过将增强后的光照分量与增强后的反射分量相乘，得到最终的增强图像。RetinexNet 的具体算法流程如图 3-1 所示。

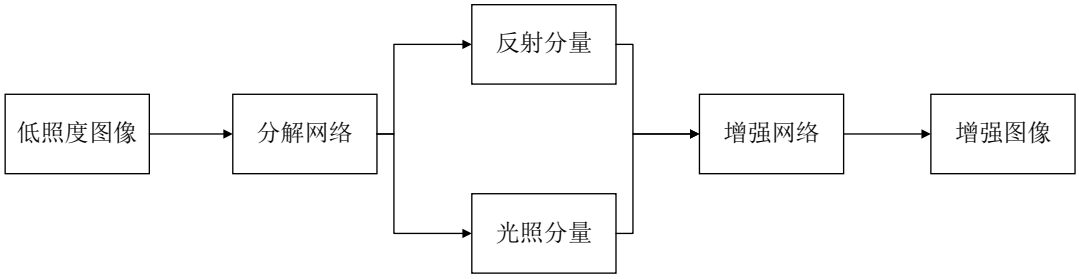


图 3-1RetinexNet 网络的算法流程

Figure.3-1Algorithm flow of RetinexNet network

RetinexNet 利用深度学习算法的强大能力，自动化地学习图像分解和增强过程，从而克服了传统方法的局限性，其网络结构如图 3-2 所示。

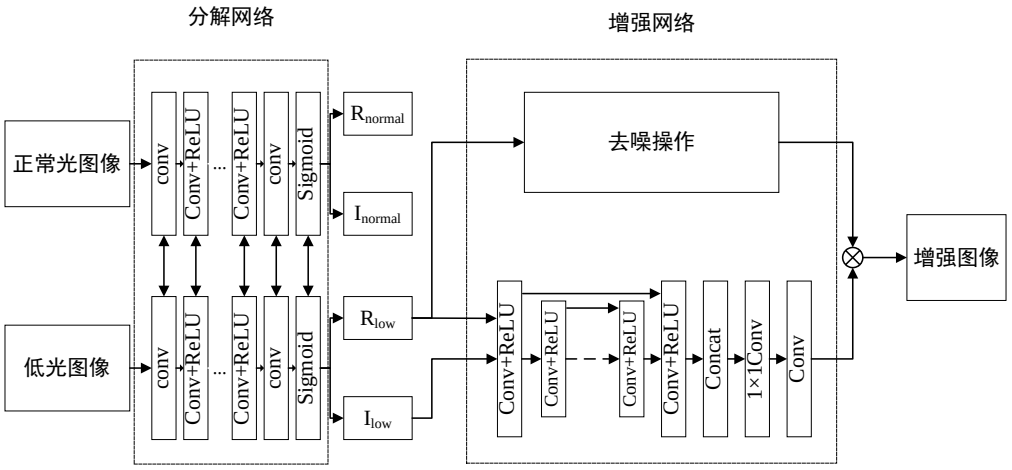


图 3-2RetinexNet 网络结构图

Figure3-2 The architecture of RetinexNet

RetinexNet 架构可细分为分解网络、增强网络与重构网络三大模块。分解网络负责解析图像中的反射分量和光照分量，增强网络致力于减弱反射分量的噪声并校正光照分量的强度，而重构网络则基于处理后的分量恢复出在正常光照下的图像。

分解网络由两个卷积层、五个含 ReLU 激活函数的卷积层及一个 Sigmoid 激活函数构成。此网络接收低/正常光照图像对作为输入，并采用共享参数策略，通过分别处理输入图像以获得低光照和正常光照下的反射及光照分量。RetinexNet 通过这四个分量之间的内在关联性来优化模型性能，其优化目标由重建损失、反射分量一致性损失和光照分量平滑损失三者构成，如下所示：

$$L_{loss} = L_{recon} + \lambda_{ij} L_{ir} + \lambda_{is} L_{is} \tag{3-1}$$

$$L_{is}=\sum_{i=low,normal}\left\|\nabla I_i\circ\exp(-\lambda_g\nabla R_i)\right\| \tag{3-2}$$

$$L_{recon}=\sum_{i=low,normal}\sum_{j=low,normal}\lambda_{ij}\left\|R_i\circ I_j-S_j\right\|_1 \tag{3-3}$$

$$L_{ir}=\left\|R_{low}-R_{normal}\right\|_1 \tag{3-4}$$

其中，L_{loss}、L_{recon}、L_{ir} 以及 L_{is} 分别为总损失、光照平滑损失、重建损失以及反射一致性损失，$\circ$ 为逐元素相乘，λ_{ij} 以及 λ_{is} 为平衡不同损失部分的权重系数；∇ 为梯度计算操作，I_i 为光照分量，R_i 为反射分量，λ_g 结构感知强度的平衡系数，用于控制梯度权重，$\exp(-\lambda_g\nabla R_i)$ 用于在反射分量梯度较大的区域放宽平滑约束；S_j 为原始图像；R_{low} 为低光照图像的反射分量，R_{normal} 正常光照图像的反射分量。

此外，在 RetinexNet 的增强网络里，针对反射分量的增强，RetinexNet 运用 BM3D 算法抑制反射分量中放大的噪声；针对光照分量的增强，RetinexNet 构建了一种多尺度光照调整网络，此网络可捕捉大范围光照分布的上下文信息，进而有助于提升其自适应调整能力。

最终，将增强后的光照分量与反射分量相乘，得到最终增强后的图像。

3.2 U-Net 网络

本章节所提网络中的特征融合网络为一种类 U-Net 结构，由 U-Net 改进而来。U-Net 由编码器（下采样路径）和解码器（上采样路径）两部分组成，网络形状呈 U 型，因此得名 U-Net。该网络可以应用在多个领域，例如；图像分割以及图像增强等，其网络结构如图 3-3 所示。

在 U-Net 网络的编码器阶段，图像在 U-Net 中通过一系列的卷积层进行特征提取，每一层都包含两个带有 ReLU 激活函数的卷积以及池化操作。随着网络的深入，特征图的尺寸逐渐减小，而特征通道的数量逐渐增加，并以此提取到更高级别的语义信息。而在解码器阶段，解码器从编码器接收的特征图开始，通过上采样操作（如反卷积）逐步增加特征图的尺寸，同时特征通道的数量也会逐渐减少。并且在每一步上采样后，通过跳跃连接将编码器相应层级的输出特征图与解码器当前层级的输入特征图进行拼接。在解码器的最后一层，使用卷积操作将特征图的通道数映射为 3，从而获得增强后的 RGB 图像。

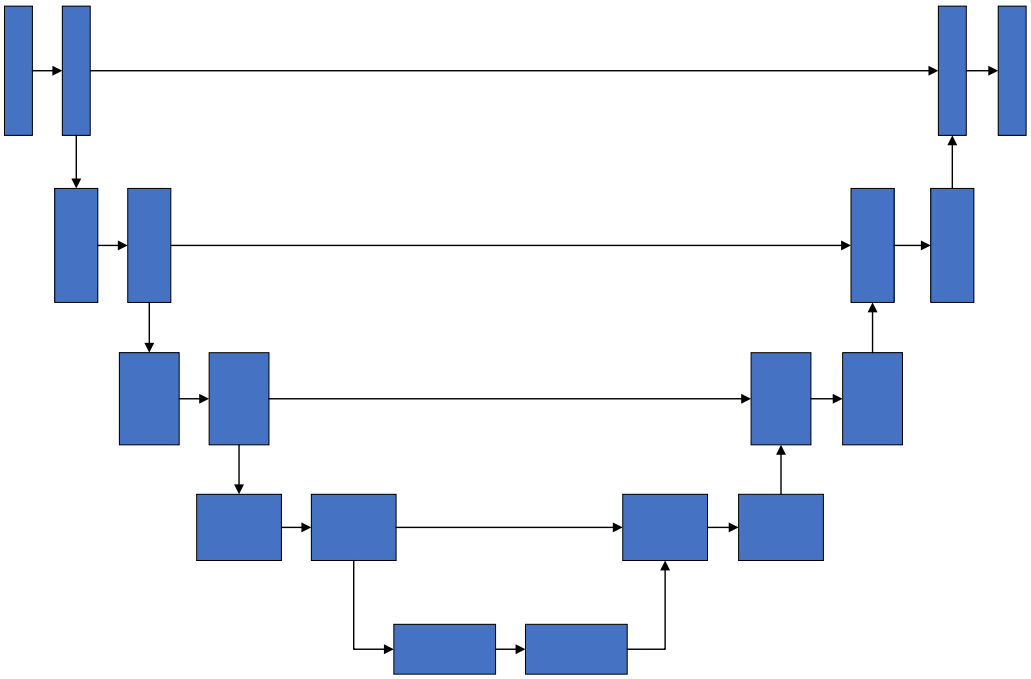


图 3-3U-Net 网络结构图

Figure.3-3The architecture of U-Net

3.3 基于改进 RetinexNet 网络的图像增强算法

在本小节所提出的网络架构中，主要涵盖了两个关键部分，即分解网络与增强网络。依据 Retinex 理论，图像能够被分解为反射分量以及照度分量。其中，反射分量主要取决于物体自身的属性，对光照强度的变化并不敏感；照度分量则由图像内的光线分布状况所决定，与物体的颜色以及纹理并无关联。基于此理论基础，网络首先借助分解网络把输入图像解析为光照分量和反射分量。随后，将这两部分分别输入到特征融合网络（FFN，Feature fusion network）以及残差网络（RN，Residual network）中，同时使用特征融合网络增强光照分量，使用残差网络增强反射分量。在特征融合网络中，在下采样阶段有两个分路，分别为上路径和下路径，上路径采用 1×3 和 3×1 的卷积核代替传统的 3×3 卷积核，以减少网络参数并加速训练过程，同时保持性能。对于下路径，则采用传统的 3×3 卷积。在完成四次下采样后，将上路径和下路径的输出结果进行融合，融合后的特征图通过上采样最终生成增强后的光照分量；在残差网络中，通过残差结构对反射分量进行增强，并获得增强后的反射分量；在特征融合网络以及残差网络分别增强光照分量和反射分量后，通过光照分量与反射分量相乘获得最终的增强图像。

3.3.1 网络结构

所提网络由两个部分组成，即分解网络和增强网络。具体结构如图 3-4 所示，其中，在分解网络中，将低照度图像及其对应的正常照度图像作为成对数据输入到网络中。网络首先通过一个 3×3 的卷积层提取图像的浅层特征。随后，利用一个包含 5 层 3×3 卷积的网络，以 Leaky ReLU 作为激活函数，将图像分解为反射分量和光照分量。接着，通过另一个 3×3 卷积层将原始图像的特征映射为反射图和光照图。最后，使用 Sigmoid 函数将反射图和光照图的像素值归一化到[0，1]范围内，以便后续处理。紧接着将分解网络输出的光照分量和反射分量输入到特征融合网络以及残差网络中，其中特征融合网络包含两条路径，分别为上路径和下路径，上路径中包括四个下采样以及十个 1×3 卷积和 3×1 卷积的组合；下路径中包括十个 3×3 卷积的组合。在分别进行处理后，将两条路径处理后不同深度的图像拼接起来，通过反卷积模块以及注意力机制模块处理融合，在融合过程中包括四个拼接过程、四个反卷积模块以及四个注意力机制模块。在融合完成后获得增强后的光照分量；而残差网络中包含 K 个残差结构，本文中 K 为 4，每个残差结构中包含四个 3×3 的卷积，在通过残差网络增强后获得增强后的反射分量；最后通过反射分量与光照分量相乘获得最终的增强图像。

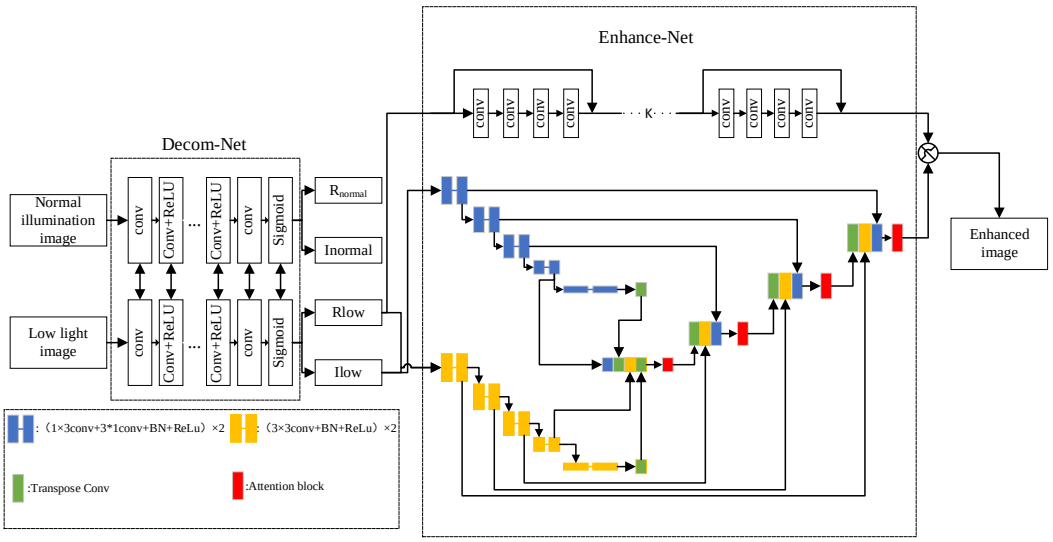


图 3-4 图像增强网络结构图

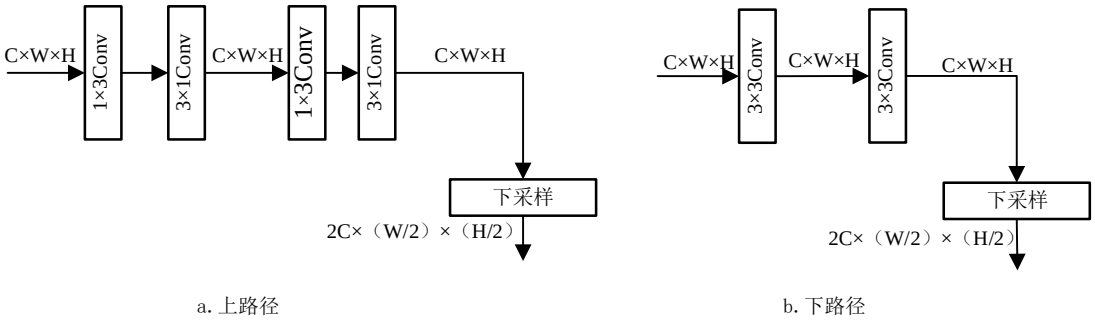
Figure.3-4 The architecture of Image enhancement network

3.3.2 特征融合网络

特征融合网络的处理过程中包括两个路径，分别是上路径和下路径。对于上路径，1×3 卷积和 3×1 卷积的组合不会改变特征图的大小和通道数，而在每两

组 1×3 卷积和 3×1 卷积的组合后有一个下采样操作，该下采样操作会使特征图的宽高缩减一倍，并使通道数翻倍，其变化如图 3-5 所示。与之相同的是，在下路径的处理过程中，3×3 卷积也不会改变特征图的大小和通道数，并且在每两组卷积后同样含有一个下采样操作，该下采样操作同样会使特征图的宽高缩减一倍，并使通道数翻倍。

在分别进行四次下采样以及十次卷积后，会分别对两条路径进行一次反卷积操作，反卷积会使特征图的宽高翻倍，通道数减半。接着分别将两个路径上反卷积后获得特征图以及第八次卷积后获得的特征图拼接起来，将拼接后的结果送入注意力机制中获得初次融合的结果。在这之后重复上述操作，在经历四次反卷积、拼接以及注意力模块后，图像恢复到初始大小，并获得增强后的光照分量。



a. 上路径　　b. 下路径

图 3-5 特征融合网络中不同路径的特征图尺寸变化

Figure.3-5 Feature map size variation in convolutional blocks

3.4 实验与分析

本文所做实验的环境设置如表 3-1 所示

表 3-1 实验环境设置

Table 3-1 Experimental Environment Configuration

设置项	设置参数
编程语言	Python3.9
CPU 配置	Intel Core i9
GPU 配置	GTX 4090 Ti
运行内存	64G
操作系统	Linux
深度学习框架	PyTorch

3.4.1 实验数据集

LOL-V1（Low-Light Dataset Version 1）是一个用于低光照图像增强研究的常用数据集，包含 500 对低照度/正常光照图像，每对图像由低光照图像及其对应的正常光照图像组成，图像为 RGB 格式，每对图像的像素大小为 600×400，这些图像大多处于极低照度条件下，涵盖了丰富的场景。在训练阶段，通常使用 485 对图像作为训练集，剩余的 15 对图像作为测试集，用于评估模型的性能。

3.4.2 消融实验

根据 Retinex 的原理，一幅图像可以表示为光照分量和反射分量的乘积。基于此，分别使用特征融合网络（FFN，Feature fusion network）以及残差网络（RN，Residual network）增强光照分量和反射分量，并将增强后的光照分量和反射分量相乘获得最终增强后的图像。通过这种方式可以有效提高网络对低质图像的增强能力。这些增强的效果将会通过消融实验进行佐证。

如表 3-2 所示，其中 FFRRetinexNet 为本节提出的增强网络；FFRRetinexNet without RN 为在 RetinexNet 的基础上只使用特征融合网络处理光照分量，不使用残差网络处理反射分量的方法；FFRRetinexNet without FFN 为在 RetinexNet 的基础上不使用特征融合网络处理光照分量，只使用残差网络处理反射分量的方法；RFFRetinexNet 是指使用特征融合网路处理反射分量，使用残差网络处理光照分量的方法。从表 3-2 可以看出，无论是特征融合网络还是残差网络都可以增强 FFRRetinexNet 的性能，而对于 RFFRetinexNet，由于特征融合网络会较多的关注特征图的边缘信息，当使用特征融合网络增强反射分量时，会导致最终增强图像的边缘信息较为明显，因此其 PSNR 与 SSIM 值也就比较低。

表 3-2 不同子网络对 FFRRetinexNet 的影响

Table 3-2 Impact of Different subnetworks on FFRRetinexNet

模型	PSNR (dB)	SSIM
FFRRetinexNet	**21.24**	**0.779**
FFRRetinexNet without RN	19.88	0.759
FFRRetinexNet without FFN	18.94	0.737
RFFRetinexNet	14.25	0.587

除了需要验证不同子网络对 FFRRetinexNet 的影响外，还需要验证特征融合网络中注意力机制（AB）的影响，具体结果如表 3-3 所示。

表 3-3 注意力机制对 FFRRetinexNet 的影响

Table 3-3 Impact of Attention Mechanism on FFRRetinexNet

模型	PSNR (dB)	SSIM
FFRRetinexNet	**21.24**	**0.779**
FFRRetinexNet without AB	19.65	0.747

观察表 3-3 可以看出，FFRRetinexNet 比去掉特征融合网络中注意力机制的 FFRRetinexNet 具有很好的客观评价指标。相比较 PSNR，FFRRetinexNet 比去掉特征融合网络中注意力机制的 FFRRetinexNet 高出 7.5%；相比较 SSIM，FFRRetinexNet 比去掉特征融合网络中注意力机制的 FFRRetinexNet 高出 4.1%.

3.4.3 对比算法

我们选择 RetinexNet、Kind、DeepUPE、U-Net 以及 LLformer 作为对比算法，从定量以及定性两方面对本节算法 FFRRetinexNet 进行性能评估。

表 3-4 不同算法的结果对比

Table 3-4Compare the results of different algorithms

模型	PSNR (dB)	SSIM
RetinexNet	17.64	0.711
U-Net	15.58	0.610
Kind	16.66	0.736
DeepUPE	17.28	0.507
LLformer	**23.41**	**0.800**
FFRRetinexNet	21.24	0.779

如表 3-4 所示，虽然相比较 LLformer 来说，本节算法表现较差。但与 RetinexNet、Kind、U-Net 以及 DeepUPE 这四个算法相比本节算法拥有较好的性能。与 RetinexNet 相比，本文算法在 PSNR 上提高了 16.9 个百分点，在 SSIM 上提高了 8.7 个百分点；与 U-Net 相比，本文算法在 PSNR 上提高了 22.6 个百分点，在 SSIM 上提高了 21.6 个百分点；与 Kind 相比，本文算法在 PSNR 上提高了 21.5 个百分点，在 SSIM 上提高了 5.5 个百分点；与 DeepUPE 相比，本文算

法在 PSNR 上提高了 18.6 个百分点，在 SSIM 上提高了 34.9 个百分点。其中，与 RetinexNet 的结果对比有力的证明了本节改进增强算法的可行性。

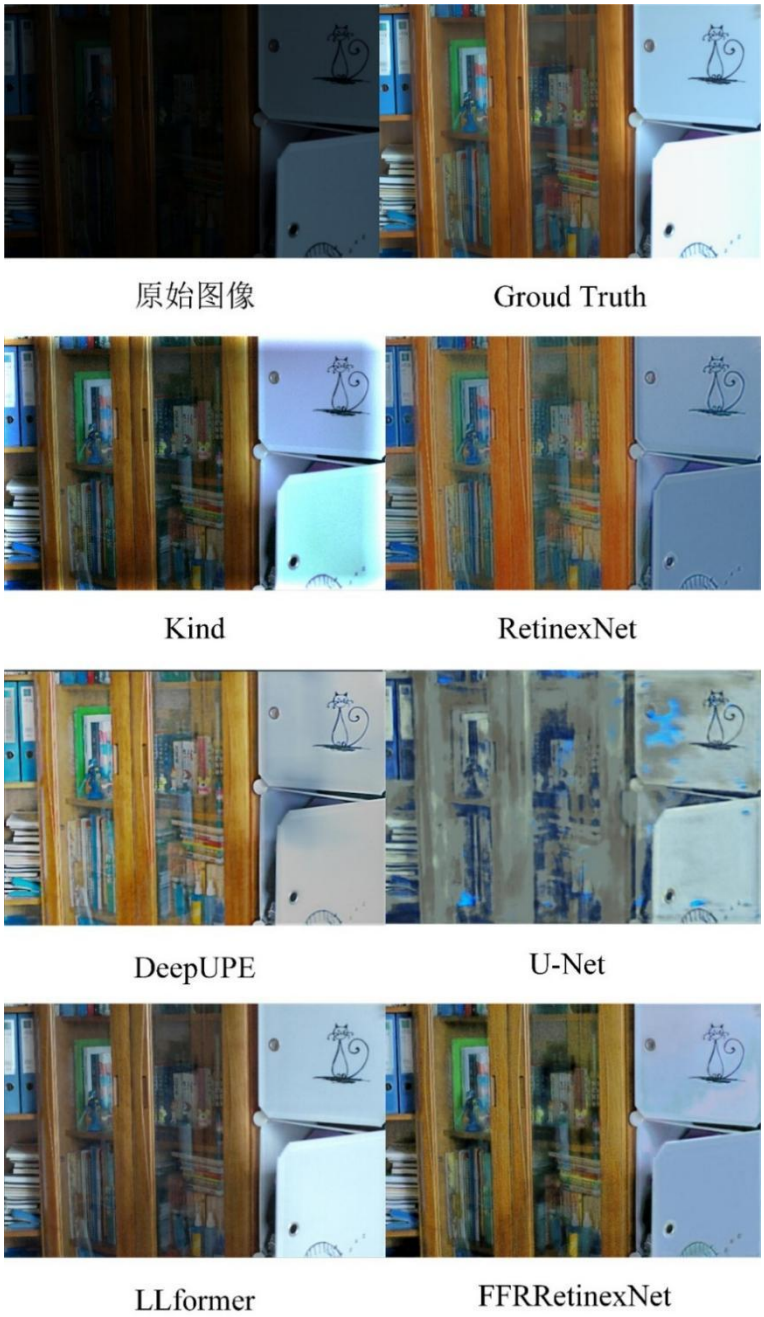


图 3-6 不同算法的结果对比

Figure.3-6 Compare the results of different algorithms

图 3-7 不同算法的结果对比

Figure.3-7 Compare the results of different algorithms

除了定量分析外，还需要对 FFRRetinexNet 进行定性分析，本节对比了 RetinexNet、U-Net、Kind、DeepUPE、LLormer 以及 FFRRetinexNet 的主观效果图，如图 3-6、图 3-7 所示，与 LLormer 相比，FFRRetinexNet 的整体对比度较差，但仔细观察可以得出，LLformer 的增强图片亮度不够自然，有些区域对比度

过高，相比之下，本算法的增强图在自然性方面具有较好的视觉效果。而通过观察 RetinexNet、Kind、U-Net 以及 DeepUPE 等算法不难发现，这些算法的增强图像普遍存在对比度差、失真以及不自然的视觉效果，与这些算法相比，FFRRetinexNet 具有较好的自然性以及对比度，拥有较好的视觉效果。

3.5 本章小结

在本节中，我们提出了 FFRRetinexNet 进行低质图像增强。FFRRetinexNet 首先将输入图像分解为光照分量和反射分量。随后，将这两个分量分别送入特征融合网络（FFN）和残差网络（RN）进行处理。在特征融合网络中，光照分量通过下采样阶段的两条路径进行处理：上路径采用 1×3 和 3×1 的卷积核代替常规的 3×3 卷积核，以降低网络参数量并加快训练速度，同时保持性能；下路径则使用传统的 3×3 卷积核。完成下采样后，将上路径和下路径的结果合并，并通过上采样生成增强后的光照分量；在残差网络中，反射分量通过残差结构进行增强，得到增强后的反射分量；最后，将增强后的光照分量与反射分量相乘，从而得到最终的增强图像。

实验结果表明，从定量和定性两方面来看，我们提出的 FFRRetinexNet 在图像增强方面是比较有效的。

第 4 章 RSARNet：一种鲁棒的噪声去除网络

图像降噪是图像处理中的一个重要且具有挑战性的任务，旨在减少图像中的噪声以提升图像质量，进而便于后续的图像分析和解释。

传统降噪方法主要分为两类：空间域滤波和频率域滤波。空间域滤波直接对图像像素进行操作，如均值滤波、高斯滤波、双边滤波和中值滤波等，其中非局部均值（NLM）算法通过捕获图像的局部和全局结构信息，在降噪的同时较好地保留了图像细节。频率域滤波则通过傅里叶变换、小波变换等方法将图像转换到频域进行噪声处理，再通过逆变换返回空间域，BM3D 算法结合了非局部降噪和小波变换，效果显著。传统方法虽实时性强，但降噪效果有限，难以保留图像细节和边缘信息。

近年来，深度学习技术在图像降噪领域取得了突破性进展。基于深度学习的降噪算法，尤其是卷积神经网络，在性能、细节处理和边缘信息保留方面优于传统方法。然而，CNN 训练中存在参数变化导致输入分布变化的问题，增加了训练复杂性。为此，Christian 等人[53]提出了批处理归一化（BN）方法，通过归一化层输入，稳定训练过程，提高网络的泛化能力和收敛速度。He 等人[54]人进一步引入残差连接，结合 BN，解决了梯度消失问题，使网络能更有效地传递浅层特征到深层。为提升网络对重要图像特征的关注度，Tian 等人[55]提出结合注意力机制的降噪方法，通过自适应计算注意力权重，加权处理图像特征，从而更好地去除噪声并保留细节。Wang 等人[56]提出了一种双分支网络 DulNet，用于图像降噪。该网络包含高分辨率分支和编码器-解码器分支，通过信息交换模块实现特征共享，并引入感知损失功能以提升降噪图像的视觉质量。Zhang 等人[57]提出了一种纹理引导的卷积神经网络 TDCNN，通过提取纹理特征、优化细节，并结合感知损失和均方误差的联合损失函数，进一步提升了降噪效果，保留了图像细节。何涛等人[58]则提出了一种基于 CNN 的暗光图像降噪网络，通过卷积层进行特征均匀化处理，利用暗光增强与降噪模块去除噪声并保留细节，最终通过单核卷积重构出干净图像。但以上算法都存在由于网络过深导致的细节丢失问题。

为解决这个问题，本节提出一种基于残差学习的深度卷积神经网络用于去除图像中的噪声。做出的改进有以下三点：

1.提出了新型的串联残差单元块，并将其嵌入至网络中，增强了网络的信息保留能力，弥补了卷积过程中丢失的细节特征。

2.采用长路径融合浅层特征以及深层特征的方式，并在特征提取的过程中使用扩张卷积替换普通卷积，可以在不额外引入参数和计算代价的情况下，获得更加密集的信息。

3.在网络中添加注意力模块，并利用注意力模块挖掘复杂背景下潜在的特征，使得网络能够自动地学习并选择性地关注输入中的重要信息，提升网络的特征提取能力，从而提升网络的降噪性能。

4.1 BRDNet 去噪网络

众所周知，端到端连接的方式是卷积神经网络成功的一个重要原因，一个卷积神经网络通常由卷积层、池化层、激活函数、初始参数设置以及基于梯度的优化方法组成。使用深度学习方法去噪是图像去噪领域中十分常见的方式，然而深度卷积神经网络也有缺点，比如性能饱和以及因深度学习去噪网络过深导致的细节丢失问题。基于此，BRDNet 网络区别于早期去噪网络向网络深度发展的方向，而是向宽度发展。BRDNet 网络是一种拥有两个子网络的卷积神经网络，这两个子网络分别对原始图像进行处理，并将处理的结果拼接起来，最后输出去噪后的图像。BRDNet 网络的结构如图 4-1 所示。

BRDNet 包含两个子网络，子网络的深度都为 17 层。这两个子网络分别为使用普通卷积进行特征提取的子网络和使用扩张卷积进行特征提取的子网络，这两个子网络同时接收网络的输入图像并处理。对于使用普通卷积进行特征提取的子网络，其 17 层的网络结构中有 16 层为带有批处理归一化（BN）以及 ReLU 激活函数的卷积层，第 17 层为卷积层，其目的是对输入图像进行初步的特征提取；而对于使用扩张卷积进行特征提取的子网络，其第 2、9、16 层为带有批处理归一化（BN）以及 ReLU 激活函数的卷积层，第 17 层为卷积层，其余层为带有 ReLU 激活函数的扩张卷积层，其目的是在不引入额外计算资源的情况下提升网络提取特征的性能。在这两个子网络处理完成后会与输入图像进行一个残差操作，目的是防止信息丢失。接着将两个子网络分别输出的特征图拼接起来，通过卷积处理后输出噪声映射图像，最后用输入图像减去噪声映射图像获得最终的去噪图像。

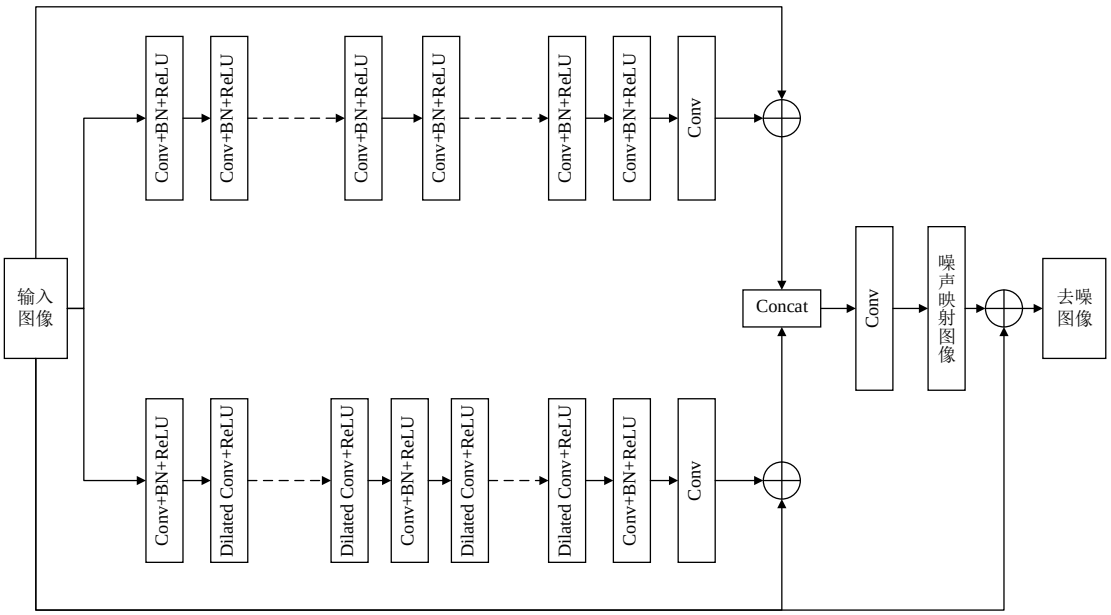


图 4-1BRDNet 网络结构图

Figure.4-1The architecture of BRDNet

4.2 改进的噪声去除网络

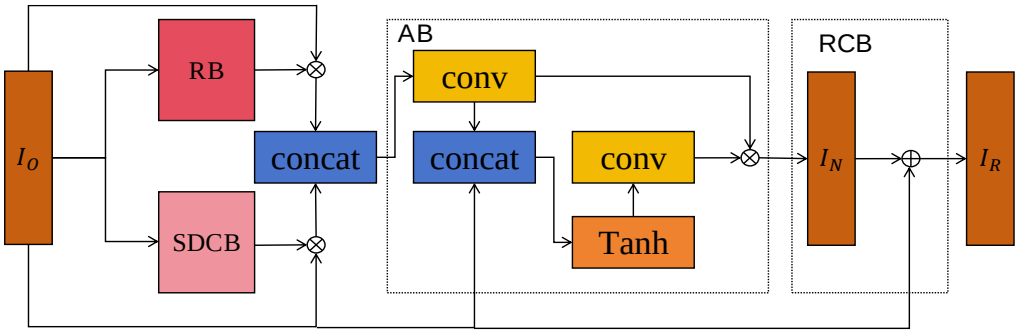
在本节中，将介绍提出的降噪网络 RSARNet，网络由 RB(残差模块)、SDCB(扩张卷积模块)、AB(注意力模块)与 RCB(重建模块)四个模块。在残差模块中，为了保留信息，防止信息衰弱，使用残差思想连接上下文信息，从而弥补卷积过程中丢失的细节特征；在扩张卷积模块中，利用扩张卷积与普通卷积交叉使用的方式提取图像特征，这有利于减少网络的深度，可以在不额外引入参数和计算代价的情况下，获得更加密集的信息，从而使得网络可以充分学习图像的局部特征和全局特征；将原始图像分别传入残差模块和扩张卷积模块中，再把获得的结果通过 Concat 拼接起来，接着把拼接后的结果传入注意力模块，通过注意力模块，可以使得网络能够自动地学习并选择性地关注输入中的重要信息，提升网络的特征提取能力，从而提升网络的降噪性能，最后通过重建模块得到最终的降噪图像。

4.2.1 网络体系结构

所提出的网络包括 RB、SDCB、AB 以及 RCB 四个模块，其结构如图 4-2 所示假设 I_O 和 I_N 分别为输入的原始图像以及残差图像，也就是噪声映射图像。那么网络的实现过程可用如下公式表达：

$$\begin{aligned} I_R &= I_O - I_N \\ I_R &= I_O - f_{AB}(f_{RB}(I_O) + f_{SDCB}(I_O)) \\ I_R &= I_O - f_{RSARNet}(I_O) \end{aligned} \tag{4-1}$$

其中 I_R 降噪后的潜在无噪图像，f_{AB}、f_{RB}、f_{SDCB} 以及 $f_{RSARNet}$ 分别代表着在对应模块中所做的操作。



I_O：原始图像 I_N：残差图像 I_R：重构图像

图 4-2 图像降噪网络结构图

Figure.4-2 Structure diagram of the image denoising network

4.2.2 残差模块

在卷积神经网络中，随着网络深度的增加，网络的浅层对于深层的影响会逐渐减弱，这也就导致网络无法更好的保留图像中的细节，从而减弱网络的特征提取能力。为了解决这一问题，本节提出残差模块(RB)，该模块充分利用残差思想连接上下文信息，从而更好的保留图像中的细节，提高网络的特征提取能力。该模块由四个残差单元块和一个卷积层串联而成，每个残差单元块的输出直接输入到后续的残差单元块中，每个残差单元块包括四个 CBL 块，其中 CBL 块是由 Conv、BN、ReLu 串联而成。整个 RB 块中合计 17 层，第 1 层的滤波器尺寸设置为 c×3×3×64，第 2 层~第 16 层的尺寸设置为 64×3×3×64，第 17 层的尺寸设置为 64×3×3×c，其中 c 为输入图像通道数，彩色图像为 3，灰度图像为 1。其具体结构如图 4-3 所示。

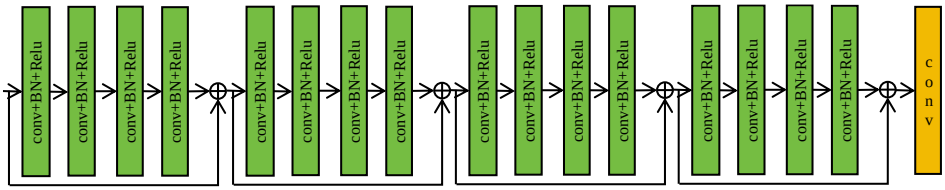


图 4-3 残差模块结构图

Figure.4-3 Structural diagram of the residual module

在相邻的残差单元块中添加跳跃连接，使用跳跃连接实现恒等映射，从而学习到图像中的更多细节，使网络拥有更强的特征提取能力，提高网络降噪的性能。

4.2.3 稀疏扩张卷积模块

利用扩张卷积在降噪性能和效率之间进行权衡，可以用更少的参数获得更多的信息，从而获得更好的降噪效果。但是随着网络深度的增加，浅层网络对于深层网络的影响将会逐渐减小，因此如果连续使用扩张卷积将无法完全映射浅层网络的特征信息。基于此，本节中提出稀疏扩张卷积模块用来降低网络的深度，可以降低计算成本和内存的消耗，还可以提升运算的效率。SDCB 的具体结构如图 4-4 所示。

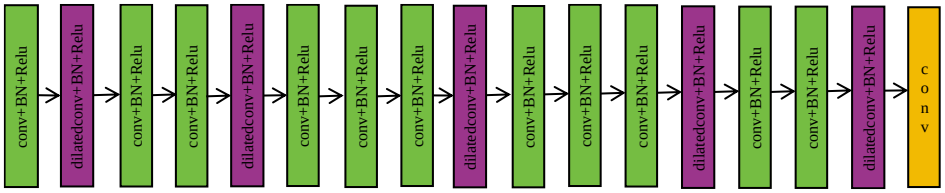


图 4-4 稀疏扩张卷积模块结构图

Figure.4-4 Structural diagram of the Sparse Dilated Convolution Block

具体来说，稀疏扩张卷积模块总共有 17 层网络，其中第 2、5、9、13、16 层设置为扩张卷积层，dilation 设置为 2，其余层为正常的 3×3 卷积，将扩张卷积和普通卷积不规则地结合在一起，且扩张卷积层数较少，符合稀疏表示。其中第 1 层的滤波器尺寸设置为 c×3×3×64，第 2 层到第 16 层的滤波器尺寸设置为 64×3×3×64，第 17 层的滤波器尺寸设置为 64×3×3×c，其中 c 为输入图像通道数，彩色图像为 3，灰度图像为 1。

4.2.4 注意力模块

分别将原始图像传入到 RB 以及 SDCB 中，将获得的结果通过 Concat 融合起来，将融合后的结果输入到 AB 中，使用 AB 模块指导网络降噪。具体来说，在 AB 中，第二个卷积层利用 1×1 的卷积，将获得的特征压缩成向量作为权重，

利用获取的权重与第一个卷积层获得的输出相乘，从而获得复杂背景中潜在的噪声特征。

4.2.5 损失函数

RSARNet 在训练时先在训练集的图像上加入高斯噪声，接着把加入噪声的图像输入到 RSARNet 中，并且通过损失函数反向传播调节网络中的参数，最终使网络收敛。在本节中，损失函数选择均方误差(MSE)[59]对网络中的参数进行训练。均方误差是预测的数据与原始数据对应点误差的平方和均值，具体公式如公式(4-2)所示：

$$L_{MSE}=\frac{1}{n}\sum_{i=1}^{n}(Y_i-\tilde{Y}_i)^2 \tag{4-2}$$

假设 x 是无噪图像，y 是有噪声的图像，当给定数据集$\{x_j,y_j\}_{j=1}^{N}$时，RASRNet 通过训练获得模型并预测残差图像 f(y)，也就是噪声映射图像。然后，可以近似通过 $x=y-f(y)$ 获得降噪图像，通过移项可以获得 $f(y)=y-x$。接着通过 Adam[60]最小化公式(4-3)中的损失函数获得最优参数：

$$L_{loss}=\frac{1}{2N}\sum_{i=1}^{N}\left\| f_{RASRNet}(I_N^i)-(I_N^i-I_C^i)\right\|^2 \tag{4-3}$$

其中 $f_{RASRNet}(I_N^i)$ 是噪声图像通过 RSARNet 处理获得的噪声映射图像，B 是一个 batch 输入的样本数量，$(I_N^i-I_C^i)$ 是指标准的噪声映射图像。在训练的过程中不断最小化 L_{loss}，减少预测的噪声映射图像与标准的噪声映射图像之间的误差，这样预测的噪声映射图像才可以更加接近标准的噪声映射图像，从而使得获得的降噪图像更加接近无噪图像，获得更好的降噪效果。

4.3 实验与分析

本文所做实验的环境设置如表 4-1 所示

表 4-1 实验环境设置

Table 4-1 Experimental Environment Configuration

设置项	设置参数
编程语言	Python3.9
CPU 配置	Intel Core i9
GPU 配置	GTX 4090 Ti
运行内存	64G
操作系统	Linux
深度学习框架	PyTorch

4.3.1 实验数据集

本章节采用的训练数据集分为灰度图像数据集和彩色图像数据集，灰度图像降噪实验的训练数据集选取的是 BSD400，一共包含了 400 张灰度图片，这些灰度图片的大小都为 180×180，包括人类、建筑以及风景等种类。为了网络训练的收敛，对图像进行不同角度的旋转、翻转以及剪裁等操作，将数据集扩充到 238336 张 40×40 的子图像块，并以此作为训练集。彩色图像降噪实验的训练集选取的则是 BSDS500 中的训练图像，使用其中的两百张训练数据作为训练集，训练集中的图片大小皆为 300×300，同样对彩色数据集进行不同角度的旋转、翻转以及剪裁等操作，以达到扩充训练集的目的。

测试数据集：为了验证网络的实用性与鲁棒性，我们采用了五个数据集来评估 RSARNet 的性能，分别是 Set12、BSD68、CBSD68、Kodak24 以及 McMaster。具体来说，Set12 和 BSD68 为灰度图像数据集，CBSD68、Kodak24 以及 McMaster 为彩色图像数据集。

4.3.2 消融实验

网络的结构对于图像处理过程中的特征提取部分十分重要，本节网络总共有四个模块，分别是残差模块(RB)、稀疏扩张卷积模块(SDCB)、注意力模块(AB)以及重建模块(RCB)。在本节将针对 RB、SDCB 以及 AB 这三个模块进行消融对比实验，从而验证这些模块的必要性。具体如表 4-2 所示。

表 4-2 当σ=25 时，各种方法在 BSD68 上的结果

Table 4-2 results of different methods on BSD68 with noise levels of 25

模型	PSNR (dB)
RSARNet	**29.23**
RSARNet without RB	28.52
RSARNet without SDCB	28.81
RSARNet without AB	28.54

从表 4-2 可以得出，完整的 RSARNet 的降噪效果最好，RB、SDCB 以及 AB 这三个模块都可以增加网络的性能。RB 使网络可以很好的利用上下文信息，保留图像中的细节；SDCB 可以用更少的参数获得更多的信息，从而获得更好的降噪效果；AB 可以提取隐藏在复杂背景中的潜在噪声。

4.3.3 对比算法

在灰度数据集上，选取了 NLM、BM3D、DnCNN、ADNet 以及 BRDNet 五种算法做对比试验，在测试集 Set12 以及 BSD68 上添加噪声等级σ分别为 15、25、50 以及 75 的高斯噪声，从定性和定量的角度分别对预测的图像进行评估，比较各算法的降噪能力。在彩色数据集上，选择了 CBM3D、DnCNN、ADNet 以及 BRDNet 四种算法做对比实验，在测试集 CBSD68、Kodak24 以及 McMaster 上添加噪声等级σ为 25 的高斯噪声，并且对各个算法预测的图像进行评估，比较各算法的降噪能力。

表 4-3 当σ=15、25、50 以及 75 时，各种方法在 Set12 上的 PSNR 结果

Table 4-3 PSNR (dB) results of different methods on Set12 with noise levels of 15，25，50 and 75

σ	NLM	BM3D	DnCNN	ADNet	BRDNet	RSARNet
15	29.95	31.77	32.47	32.83	32.84	**32.89**
25	28.21	29.98	30.12	30.42	30.42	**30.48**
50	24.46	26.67	26.99	27.17	27.22	**27.26**
75	22.39	24.83	24.09	25.35	25.52	**25.57**

表 4-3 中红色代表最优，蓝色代表次优，当测试集为 Set12 时，观察不难看出，当σ=15、25、50 以及 75 时，RSARNet 在图像降噪方面优于 NLM、BM3D、

DnCNN、ADNet 以及 BRDNet。当 σ =15 时，RSARNet 对比 DnCNN 在 PSNR 平均值上高了 0.42dB，RSARNet 对比 BM3D 在 PSNR 平均值上高了 1.12dB，对比与其他的算法在 PSNR 上也有提升；当 σ =25 时，RSARNet 对比 DnCNN 在 PSNR 平均值上高了 0.36dB，RSARNet 对比 BM3D 在 PSNR 平均值上高了 0.50dB，对比其他算法也基本都是最优；当 σ =50 时，RSARNet 对比 DnCNN 在 PSNR 平均值上高了 0.27dB，RSARNet 对比 BM3D 在 PSNR 平均值上高了 0.59dB；当 σ =75 时，RSARNet 对比其他算法均有了 0.05dB~3.18dB 的提升。同理，当测试集为 BSD68 时，观察表 4-4 不难得出，相比较其他算法，RSARNet 的 PSNR 结果平均值较为优秀。

表 4-4 当 σ =15、25、50 以及 75 时，各种方法在 BSD68 上的 PSNR 结果

Table 4-4 PSNR (dB) of different methods on BSD68 with noise levels of 15，25，50 and 75

σ	NLM	BM3D	DnCNN	ADNet	BRDNet	RSARNet
15	29.66	31.06	31.34	31.68	31.66	**31.72**
25	27.15	28.56	28.94	29.19	29.15	**29.23**
50	24.13	25.61	26.11	26.22	26.25	**26.29**
75	22.56	24.20	24.63	24.69	24.72	**24.79**

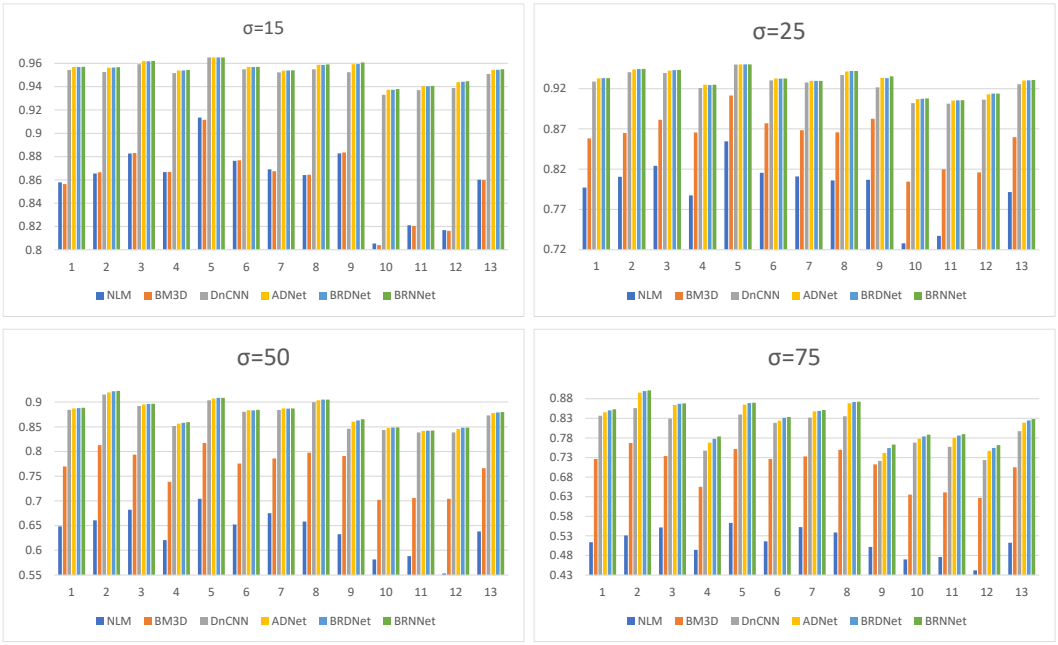


图 4-5 当 σ =15、25、50 以及 75 时，各种方法在 Set12 上的 SSIM 结果

Figure.4-5 SSIM results of different methods on Set12 with noise levels of 15，25，50 and 75.

表 4-5 当σ=15、25、50 以及 75 时，各种方法在 BSD68 上的 SSIM 结果

Table 4-5 SSIM of different methods on BSD68 with noise levels of 15，25，50 and 75

σ	NLM	BM3D	DnCNN	ADNet	BRDNet	RSARNet
15	0.82213	0.87212	0.936243	0.940522	0.940378	**0.940966**
25	0.73301	0.80121	0.895753	0.901012	0.900922	**0.901726**
50	0.57635	0.68609	0.821211	0.825489	0.827084	**0.827504**
75	0.46561	0.62211	0.770714	0.772193	0.77397	**0.777269**

如图 4-5、表 4-5 所示，无论使用 Set12 还是 BSD68，对比传统的图像降噪算法（NLM，BM3D），本节的降噪算法在 SSIM 上获得了巨大的提升，对比基于深度学习的方法（DnCNN，ADNet，BRDNet），本节的降噪算法在 SSIM 上也获得了提升。这也证明了降噪算法的有效性以及鲁棒性。

表 4-6 当噪声等级为 25 时，各种方法在 CBSD68、Kodak24 和 McMaster 数据集上的 PSNR 结果

Table4-6 PSNR(dB) results of different methods on CBSD68，Kodak24 and McMaster datasets with noise levels of 25。

datasets	CBM3D	DnCNN	ADNet	BRDNet	RSARNet
CBSD68	30.72	31.01	31.36	31.36	**31.46**
Kodak24	31.69	31.79	32.19	31.86	**32.34**
McMaster	31.66	30.71	31.71	31.07	**31.86**

对于彩色数据集，我们选择了 CBM3D、DnCNN、ADNet 以及 BRDNet 四种算法做对比实验，在测试集 CBSD68、Kodak24 以及 McMaster 上添加噪声等级 σ 为25 的高斯噪声。如表 4-6 所示与传统方法 CBM3D 相比，RSARNet 在 CBSD68 上的 PSNR 平均值增加了 0.74dB，在 Kodak24 上的 PSNR 平均值增加了 0.65dB，在 McMaster 上的 PSNR 平均值增加了 0.22dB。与基于深度学习的方法 DnCNN 相比，RSARNet 在 CBSD68 上的 PSNR 平均值增加了 0.45dB，在 Kodak24 上的 PSNR 平均值增加了 0.55dB，在 McMaster 上的 PSNR 平均值增加了 1.15dB。除此之外，与 ADNet 以及 BRDNet 两种算法对比，RSARNet 也获得了不小的提升。

除了上述的定量指标外，我们还进行了定性指标的比较。观察图 4-6 可以看出，传统的降噪方法(NLM，BM3D)的图像恢复结果相比于基于深度学习的降噪

方法(DnCNN，ADNet，BRDNet，RSARNet)的图像恢复结果在视觉效果上十分差劲。虽然 DnCNN，ADNet，BRDNet 这三个算法的图像恢复结果对比 RSARNet 的图像恢复结果在视觉效果上表现得差距不大，但是通过观察图像中的局部放大区域可以看出 RSARNet 在细节处理能力上表现更好，更能还原原图像的边缘细节以及纹理细节。

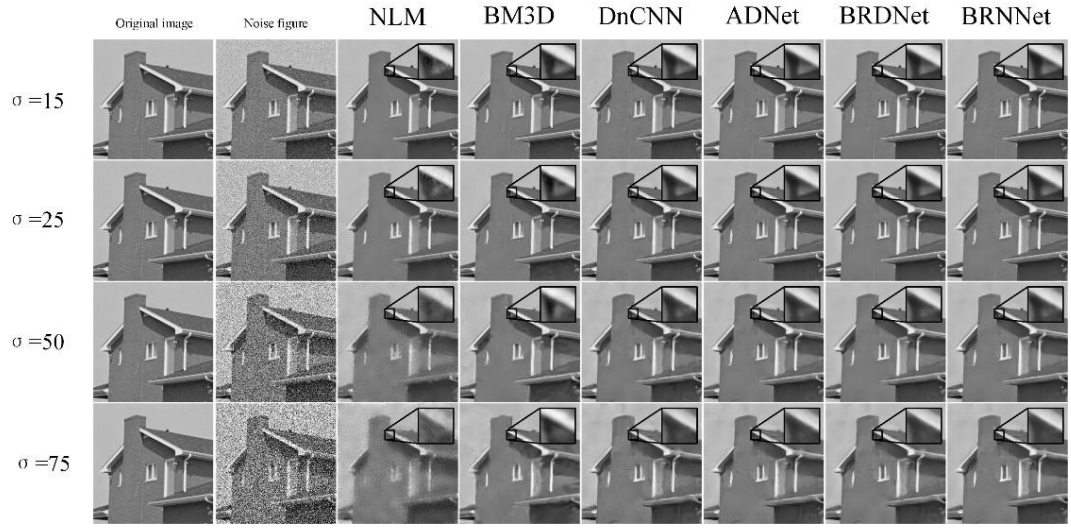


图 4-6 当 σ =15、25、50 以及 75 时，不同方法对 Set12 的降噪结果

Figure.4-6 Denoising results of different methods for Set12 with noise levels of 15，25，50 and 75

4.4 本章小结

在本节中，我们提出了 RSARNet 来进行图像降噪。RSARNet 的主要组件及其作用如下：RB（残差块）：该模块通过残差连接的方式连接上下文信息，从而保留更多的信息。残差连接能够有效地缓解深层网络中的梯度消失问题，确保网络能够学习到更丰富的特征表示。SDCB（扩张卷积块）：利用扩张卷积在降噪性能和效率之间进行权衡，可以用更少的参数获得更多的信息。扩张卷积通过增加卷积核的感受野，能够在不增加计算量的情况下捕获更广泛的上下文信息。AB（注意力块）：该模块可以提取隐藏在复杂背景中的潜在噪声信息。通过引入注意力机制，网络能够自适应地聚焦于图像中的重要区域，从而更有效地去除噪声。在执行完上述三个模块后，执行 RCB（重建块）以获得最终的干净图像。RCB 模块通过整合前面模块提取的特征，重构出不含噪声的图像。

实验结果表明，从定量和定性两方面来看，我们提出的 RSARNet 网络在图像降噪方面是非常有效的。RSARNet 不仅能够有效去除噪声，还能在降噪的同时保留图像的细节和边缘信息。

第 5 章 Swin-YOLO：面向复杂环境下的小目标检测方法

目标检测作为计算机视觉领域中的一项重要任务，其目的是在图像或视频中准确地识别和定位图像中的目标物体。传统目标检测算法主要依赖于人工目测或简单的图像处理技术，这些方法存在着检测效率低、误报率高以及对复杂缺陷的识别能力不足等问题。随着计算机视觉、图像处理和机器学习等技术的不断发展，深度神经网络在目标检测领域表现出了令人瞩目的成绩，展现出强大的性能和广泛的应用前景。在现实生活中，目标检测技术的应用极为广泛。例如：在自动驾驶领域中，目标检测技术用于检测道路上的车辆、行人、交通标志等物体。通过实时识别这些目标，自动驾驶系统能够做出准确的决策，确保行车安全；在安防监控领域中，目标检测技术用于识别监控视频中的异常行为或特定物体。例如，在公共场所的监控中，通过检测人群中的可疑人员或行为，及时发现并处理安全隐患。此外，目标检测还可以用于识别监控区域内的特定目标，如入侵者或遗留物品，提高监控系统的智能化水平。除此之外，目标检测技术还在环境保护、工业检测、智能交通等领域也有广泛的应用。

在目标检测领域，单阶段算法以其高效的实时性而著称，而在单阶段算法中，YOLO 系列算法又是效果最好、应用最广泛的算法。因此，本节选择 YOLO 系列算法作为主体网络进行目标检测。为了提升检测算法的精度和速度，刘海斌等人[61]提出了改进的 YOLOv5s 算法 YOLOv5-S，通过融合坐标注意力机制和轻量级卷积技术 GSConv，提高了检测精度和速度，并用 Focal-EIoU 替换 EIoU 以加快收敛，但在复杂背景下的表现仍有不足。因此为了解决目标检测算法中抗复杂背景干扰能力弱的问题，需要引入新的特征提取方法，提升算法的特征提取能力。Lin 等人[62]提出 Swin Transformer，该算法将 Transformer 与卷积神经网络的相关思想融合，采用移动窗口的自注意力机制，降低计算复杂度，提升图像识别精度，但其实时性受限。为了解决这个问题，Zhang 等人[63]在 YOLOv5 基础上加入 Swin Transformer，减少背景干扰，同时引入通道、空间及尺度注意力机制，增强特征提取能力，在保证了一定检测速度的基础上提升检测精度，但未验证复杂背景下的性能。2024 年，Chen 等人[64]提出 SOD-YOLOv7，该算法针对无人机图像中小目标检测问题，融合了卷积与 Swin Transformer 模块，引入注意力机制和可变形

卷积，提升了网络的多尺度检测能力。同年，Li 等人[65]提出 SRE-YOLOv8，加入 Swin Transformer 结构和残差特征增强模块，结合 ECA 注意力机制，提高复杂背景下低分辨率目标的检测性能。

上述算法虽然在不同程度上提高了目标检测模型的性能，但由于低质图像的拍摄环境大多十分复杂且多变，现有模型在目标检测过程中仍存在检测精度低、小目标漏检和误检的问题。因此，本节以 YOLOv11 为基础，提出一种在复杂环境下仍能保持较高检测精度的深度学习目标检测网络 Swin-YOLOv11。主要改进有以下：

1. 在主干网络中融入 Swin Transformer 模块，提高网络的抗复杂背景能力的同时提升了网络的细节信息提取能力，改善对道路小目标的误检、漏检等问题；

2. 在主干网络的特征提取过程中集成了 CBAM 注意力机制，使得网络可以更有效地从图像中学习到更有用的特征，从而提升网络的性能；

3. 将预测边界框的损失函数从 CIoU 替换到 Inner-IoU，增强网络模型的小目标检测能力，获得拥有更高检测精度的输出结果。

5.1 YOLOv11 网络

YOLOv11 算法是目前 YOLO 系列算法中十分先进的目标检测算法，YOLOv11 引入了改进的骨干网络和颈部架构设计，增强了特征提取能力，并提升了目标检测结果的平均精度均值（mAP）。其具体结构如图 5-1 所示。

YOLOv11 从结构上可以分为 Backbone、Neck 以及 Head，其中 BackBone 中包括五个卷积模块、两个 CSK 为 False 的 C3K2 模块（无填充）、两个 CSK 为 True 的 C3K2 模块（黄色填充）以及一个 SPPF 模块，主要作用为初步提取图像中隐藏的特征；Neck 中包括一个 C2PSA 模块、两个上采样操作、两个卷积模块、三个 CSK 为 False 的 C3K2 模块（无填充）以及一个 CSK 为 True 的 C3K2 模块（黄色填充），主要作用为接收融合 Backbone 输出的特征图，并通过整合不同分辨率的特征图，输出三个不同大小的特征图；最后将 3 种不同大小的特征图输入到 Head 中进行检测，这 3 种不同大小的特征图分别对应着大、中、小 3 种体型的目标，以此来实现不同尺度的目标检测。

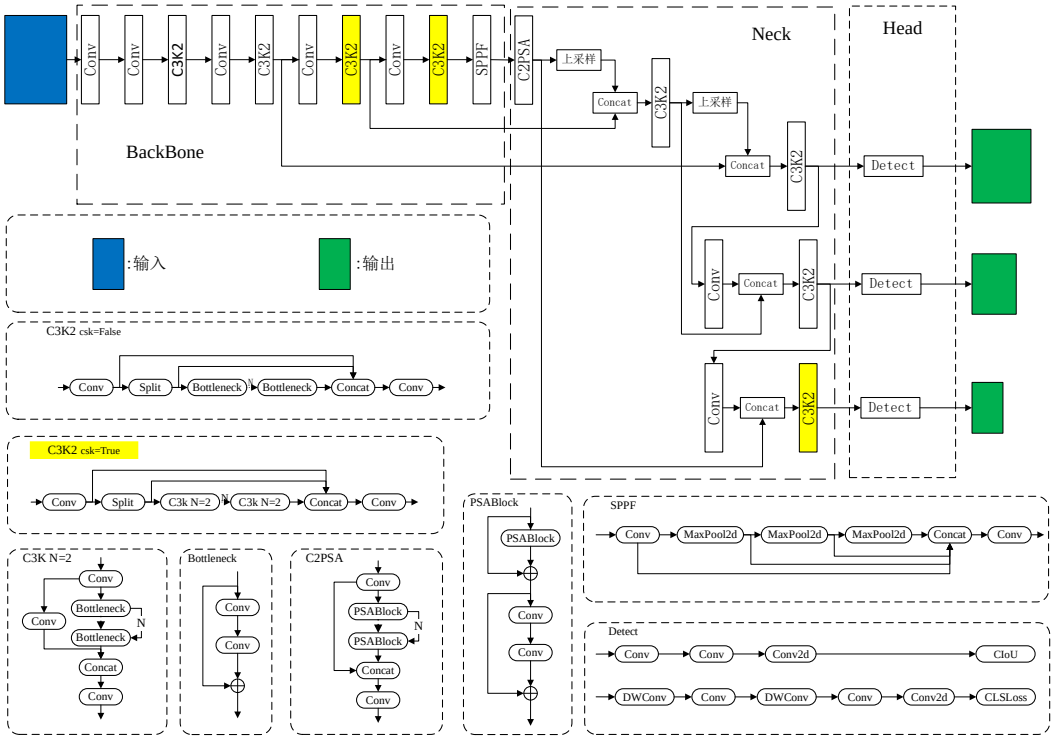


图 5-1YOLOv11 网络结构图

Figure.5-1 The architecture of YOLOv11

5.2 基于改进 YOLOv11 的目标检测

本节所提网络由三个部分组成，分别是：Swin-Backbone、Neck 以及 Head。在设计网络模型的过程中，通过分析[66]中所提出网络结构可以得知，由于 Swin Transformer 中引入了移动窗口机制和局部窗口注意力，因此在网络中使用 Swin Transformer 进行特征提取可以更好的处理大尺寸图像，增强网络提取特征以及小特征的能力，并且可以保持较低的计算复杂度。基于此，在 Swin-Backbone 以及 Swin-Head 结构中加入 Swin Transformer Block，提高网络在复杂情况下提取特征以及小特征的能力。除此之外，通过对[67]中所提出网络结构的分析，发现在网络中使用 Inner-IoU 可以使网络收敛速度加快，检测精度更好，可以使网络拥有更精准的定位效果，从而获得更好的目标检测效果。基于此，使用 Inner-IoU 替换 CIoU 作为算法的预测边界框损失函数，并以此获得更好的结果。

5.2.1 网络体系结构

所提出的网络模型由三个部分组成，即 Swin-BackBone、Neck 以及 Head，具体结构如图 5-2 所示。其中，Swin-BackBone 中分别在前两次 C3K2 后加入了 CBAM 注意力模块，这可以使网络自主的关注特征图中较为重要的部分；除此之

外，在 SPPF 前加入 Swin Transformer 结构，这可以增强主干网络的特征提取能力，提高网络模型的抗复杂背景干扰能力。Neck 保留了原 YOLOv11 中的结构，通过这些结构之间的相互作用可以整合不同分辨率的特征图，从而提高目标检测模型对不同尺度目标的检测性能；Head 则是将边界框损失函数改为了 Inner-IoU，通过这种方式可以使模型能够更细致、更专注于目标中心的性能评估，根据不同的目标自适应调整，从而提高了模型的多目标以及小目标检测能力。

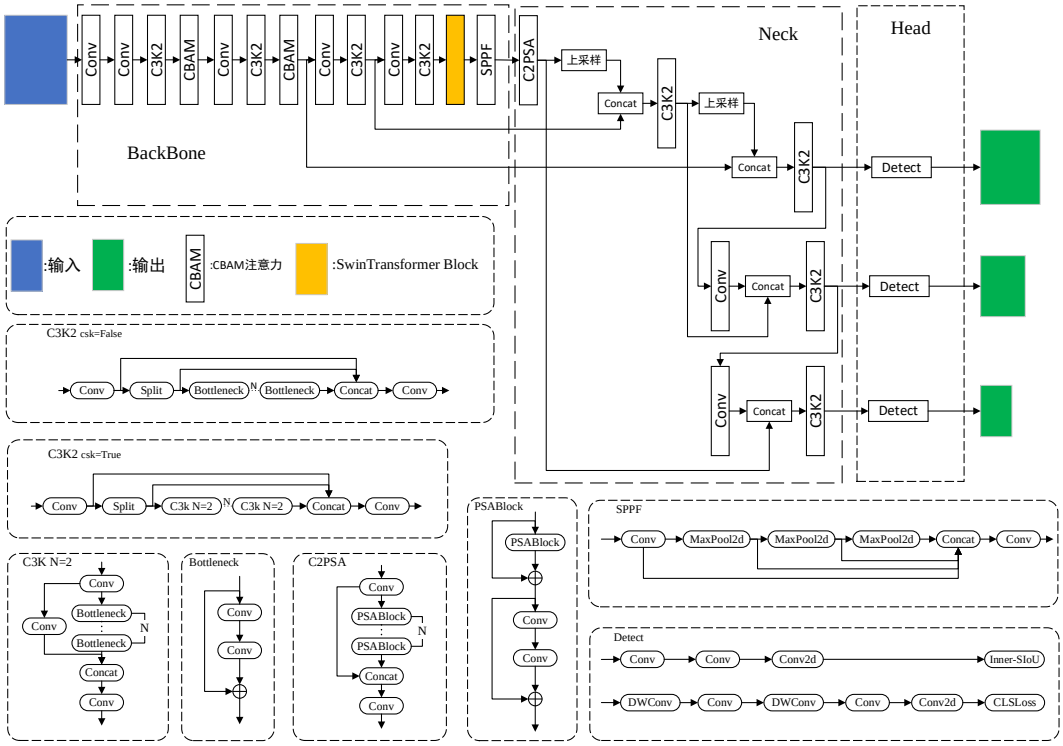


图 5-2 网络结构图

Figure.5-2 The architecture of Network

5.2.2 损失函数

Swin-YOLOv11 在训练的过程中包括多个网络迭代批次。在其中一个批次的迭代过程中，首先会从训练数据集中加载一个批量的训练样本，并且把训练样本传入到 Swin-YOLOv11 模型中进行前向传播，在此阶段，模型会针对每个预测框计算出其坐标位置、类别概率等关键信息。随后，通过对比模型的预测结果与真实标签，可以计算出损失函数的值，并依据损失函数推导出梯度，接着利用反向传播算法将这些梯度信息传回网络的每一层，以此来实现网络参数的更新，从而降低损失函数的值。这一过程会不断重复，网络参数会得到不断的更新，直到达到预先设定的训练轮数，此时损失函数会逐渐趋于收敛。本节算法选择 Inner-IoU

作为预测边界框的损失函数，Inner-IoU 通过引入一个缩放因子比率来控制辅助边界框的尺寸大小，从而为不同的数据集和检测器计算损失。Inner-IoU 损失函数定义如下：

$$L_{Inner-IoU} = 1 - IoU^{inner} \tag{5-1}$$

$$IoU^{inner} = \frac{inter}{union} \tag{5-2}$$

$$inter = (\min(b_r^{gt}, b_r) - \max(b_l^{gt}, b_l)) \times (\min(b_b^{gt}, b_b) \times \max(b_t^{gt}, b_t)) \tag{5-3}$$

$$union = (w^{gt} \times h^{gt}) \times (ratio)^2 + (w \times h) \times (ratio)^2 - inter \tag{5-4}$$

$$b_l^{gt} = x_c^{gt} - \frac{w^{gt} \times ratio}{2}, b_r^{gt} = x_c^{gt} + \frac{w^{gt} \times ratio}{2} \tag{5-5}$$

$$b_t^{gt} = y_c^{gt} - \frac{h^{gt} \times ratio}{2}, b_b^{gt} = y_c^{gt} + \frac{h^{gt} \times ratio}{2} \tag{5-6}$$

$$b_l = x_c - \frac{w \times ratio}{2}, b_r = x_c + \frac{w \times ratio}{2} \tag{5-7}$$

$$b_t = y_c - \frac{h \times ratio}{2}, b_t = y_c + \frac{h \times ratio}{2} \tag{5-8}$$

其中，(x_c^{gt}, y_c^{gt}) 代表着真实边界框内中心点的坐标，(x_c, y_c) 代表着预测边界框内中心点的坐标，$b_t^{gt}, b_b^{gt}, b_l^{gt}, b_r^{gt}$ 分别代表着真实边界框的上、下、左、右边界，b_t, b_b, b_l, b_r 分别代表着预测边界框的上、下、左、右边界，w^{gt}, h^{gt}, w, h 分别代表着真实边界框的宽和高、预测边界框的宽和高，$ratio$ 是缩放因子，用来调节辅助边界框的大小。Inner-IoU 通过辅助边界框而不是原始边界框来计算 IoU 值，即计算辅助边界框与真实边界框之间的交集和并集，以此来计算 IoU 损失。

5.2.3Swin-BackBone 模块

在目标检测中，拍摄图像含有丰富的背景信息，比如雨天、雪天、光照以及低光等。待检测目标容易受到各种复杂背景的干扰，导致特征信息衰减，无法有效的提取特征。为了解决这一问题，本节提出 Swin-BackBone 模块，该模块充分利用 Swin Transformer 以及 CBAM 的特征提取能力，以此来获得更好的细节信息，并增强网路的抗复杂背景能力。具体来说，该模块由 5 个 Conv 块、4 个 C3K2 块、1 个 Swin Transformer 块、2 个 CBAM 注意力块以及 1 个 SPPF 块组成，其

中 Conv 会对图像进行下采样处理，第 1 个 Conv 的滤波器尺寸为 3×3×3×64，第 2 个 Conv 的滤波器尺寸为 64×3×3×128，第 3 个 Conv 的滤波器尺寸为 128×3×3×256，第 4 个 Conv 的滤波器尺寸为 256×3×3×512，第 5 个 Conv 的滤波器尺寸为 512×3×3×1024。除了 Conv 之外，C3K2 块、Swin Transformer 块、CBAM 注意力块以及 SPPF 块在操作过程中不改变特征图像的尺寸。

5.3 实验与分析

本文所做实验的环境设置如表 5-1 所示

表 5-1 实验环境设置

Table 5-1 Experimental Environment Configuration

设置项	设置参数
编程语言	Python3.9
CPU 配置	Intel Core i9
GPU 配置	GTX 4090 Ti
运行内存	64G
操作系统	Linux
深度学习框架	PyTorch

5.3.1 实验数据集

本实验数据使用的数据集是 BDD100K 数据集，该数据集是目前规模最大、内容最多的公开驾驶数据集之一。数据集中的图像包含晴天、雨天以及雾天等各种天气条件，还有低光和正常光两种不同光照强度的驾驶情景。BDD100K 总计 10 万张图片，并以 7（训练集）：1（验证集）：2（测试集）的比例将数据集随机分配。

5.3.2 消融实验

在实际应用中由于复杂背景的干扰，网络的特征提取能力会被极大的影响，从而导致输出结果变差。基于此，向网络的 BackBone 中加入 Swin Transformer 模块以及 CBAM 注意力模块。通过这种方式可以提高网络的抗复杂背景能力以及信息保留能力，这些增强的效果将会通过消融实验进行佐证。

表 5-2 中 YOLOv11_S 是指在 YOLOv11 中只加入了 Swin transformer 的改进方法，YOLOv11_C 是指在 YOLOv11 中只加入了 CBAM 注意力模块的改进方

法，YOLOv11_S_C 是指在 YOLOv11 中同时加入 Swin transformer 模块以及 CBAM 注意力模块的改进方法。通过观察表中数据不难得出，虽然 YOLOv11_S_C 的 FPS 较差，但其 mAP 是最高的，比次优的 YOLOv11_S 高出 1.2%。通过综合考虑 mAP 与 FPS 这两个评价指标，本文将改进方法设置为 YOLOv11_S_C。

表 5-2 不同方法在 BDD100K 上的结果

Table 5-2 results of different methods on BDD100K

网络模型	P	R	mAP	FPS
YOLOv11	0.722	0.529	0.609	**44**
YOLOv11_S	0.71	0.551	0.623	38
YOLOv11_C	0.715	0.538	0.614	42
YOLOv11_S_C	**0.727**	**0.562**	**0.631**	35

表 5-3 损失函数对模型的影响

Table 5-3 Influence of the loss function on the model

网络模型	P	R	mAP	FPS
YOLOv11_S_C	**0.727**	0.562	0.631	**35**
Swin-YOLOv11	0.724	**0.565**	**0.637**	34

除此之外，损失函数的不同也直接影响了网络的预测输出结果，设置一个合理的损失函数不仅可以加快网络的收敛，还可以获得更好的输出结果。本文使用针对小目标检测更优的 Inner-IoU 算法替代原先的 CIoU 算法，并以此获得更好的结果。具体如表 5-3 所示。

表 5-3 中 YOLOv11_S_C 是指在 YOLOv11 中同时加入 Swin transformer 模块以及 CBAM 注意力模块但并没有改变损失函数的改进方法，Swin-YOLOv11 是指在 YOLOv11 中同时加入 Swin transformer 模块以及 CBAM 注意力模块并将损失函数改为 Inner-IoU 的改进方法。通过观察表 5-3 可以得出，这两种方法的 FPS 相差无几，但 Swin-YOLOv11 的 mAP 比 YOLOv11_S_C 的 mAP 高了 0.9%，这验证了 Inner-IoU 作为边界框损失函数的有效性。基于此，本文选择 Inner-IoU 作为边界框的损失函数。

5.3.3 对比算法

我们使用 YOLOv10、添加了 Swin Transformer 以及 CBAM 注意力机制的 YOLOv10、YOLOv11、添加了 Swin Transformer 以及 CBAM 注意力机制的 YOLOv11 作为对比算法。从定量和定性两个方面对算法做出评估。

通过观察表 5-4 中结果不难得出，相比较其他算法而言，Swin-YOLOv11 拥有较好的性能。在两个关键性能指标 mAP 与 FPS 中，Swin-YOLOv11 拥有最优的 mAP，相比 YOLOv10 的 mAP 值提升了 5.3 个百分点，相比 YOLOv11 的 mAP 值提升了 4.4 个百分点。其余算法虽然在 FPS 上表现较优，但通过综合考虑实际的复杂应用环境场景发现，失去一定的速度从而获得更高的准确率是比较有利的。因此，考虑到实际的应用场景以及种种因素，相比较其他算法，Swin-YOLOv11 算法是较为合适的。

表 5-4 不同算法的结果对比

Table 5-4 Compare the results of different algorithms

网络模型	P	R	mAP	FPS
YOLOv10	0.703	0.557	0.605	**47**
YOLOv10_S_C	0.725	0.552	0.624	42
YOLOv11	0.722	0.529	0.609	44
YOLOv11_S_C	**0.727**	0.562	0.631	35
Swin-YOLOv11	0.724	**0.565**	**0.637**	34

接着对 Swin-YOLOv11 进行定性分析，本节对比了 Swin-YOLOv11 与 YOLOv10、YOLOv11 的主观效果图，如图 5-3 所示，分别对应着雪（第一行）、低光（第二行）、雾（第三行）以及雨天（第四行）环境下的检测结果。包含三种算法的实际检测效果图，分别是 YOLOv10、YOLOv11 以及 Swin-YOLOv11。观察图 5-3 可以得出，Swin-YOLOv11 的检测结果总体上准确率更高，尤其是图像中被遮挡的目标以及小目标，比如图 5-3 第一、三行中的交通灯（traffic light）以及被遮挡的目标汽车（car）等。通过观察图 5-3 的第二、四行可以得出，当拍摄背景十分复杂，肉眼难以观察时，Swin-YOLOv11 展现出了较高的定位准确率以及检测精度。相比较其他算法，Swin-YOLOv11 的预测边界框较为合理且准确，这体现了 Swin-YOLOv11 的抗复杂背景能力。

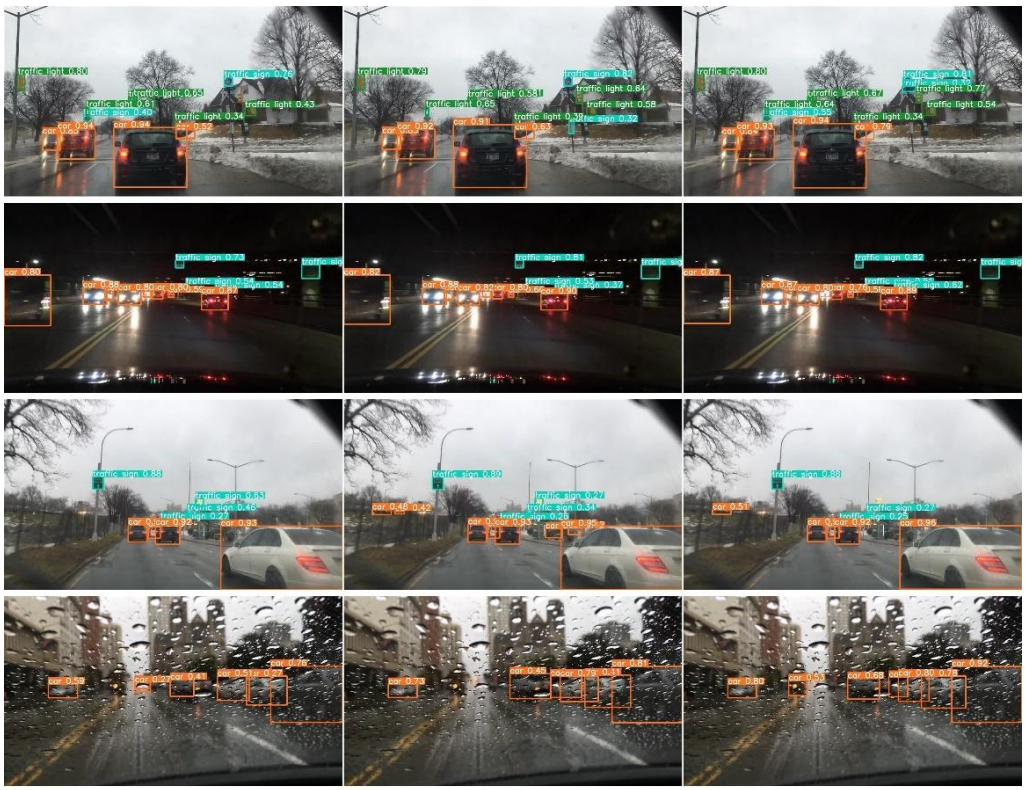


图 5-3 不同算法检测结果对比

Figure.5-3 Comparison of detection results of different algorithms

5.4 本章小结

针对现有模型在进行低质图像目标检测的过程中仍存在检测精度低、小目标漏检和误检的问题，提出一种在复杂环境下仍能保持较高检测精度的深度学习网络 Swin-YOLOv11。在主干网络中融入 Swin Transformer 模块，提高网络的抗复杂背景能力的同时提升了网络的细节信息提取能力，改善对小目标的误检、漏检等问题；其次，在主干网络的特征提取过程中集成了 CBAM 注意力机制，使得网络可以更有效地从图像中学习到更有用的特征，从而提升网络的性能；最后，将预测边界框的损失函数从 CIoU 替换到 Inner-IoU，增强网络模型的小目标检测能力，获得拥有更高检测精度的输出结果。实验结果表明，改进后的算法在 BDD100K 上的 Precision、Recall 和 mAP 都有所提高，这证明本文所提出的 Swin-YOLOv11 算法在进行复杂情况下的目标检测时是非常有效的。

在未来的研究中，将继续致力于两个方面的算法优化：一方面，持续探索提升模型鲁棒性的途径，以实现更强大的抗干扰性能，并进一步提高检测精度；另

一方面，在保证准确性的情况下，降低模型的大小，减少参数量使其可以运用到端设备上。

第 6 章 总结与展望

6.1 工作总结

本文主要探讨了低质图像增强及目标检测算法研究，提出了三种算法，分别是：基于改进 Retinex 的图像增强算法、一种鲁棒的噪声去除网络以及目标检测算法 Swin-YOLO。

首先，在第三章提出一种基于改进 Retinex 的图像增强算法，该网络首先通过分解网络将输入图像分解为光照分量和反射分量。随后，这两个分量分别被送入特征融合网络和残差网络进行处理。具体来说，特征融合网络在下采样阶段分为两条路径：上路径采用 1×3 和 3×1 的卷积核替代传统的 3×3 卷积核，以减少参数数量并加快训练速度，同时保持性能；下路径则使用常规的 3×3 卷积核。完成下采样后，将两条路径的结果进行融合，并通过上采样生成增强后的光照分量。在残差网络中，反射分量通过残差模块进行增强，从而获得增强后的反射分量。最后，将增强后的光照分量与反射分量相乘，得到最终的图像增强结果。实验表明，与其他算法相比，该算法可以有效增强低质图像，并且保留较好的自然性和对比度，具有较好的视觉效果。

其次，在第四章提出一种鲁棒的噪声去除网络，该网络使用双层卷积神经网络进行图像降噪，网络分为三部分：第一部分通过残差连接防止信息衰减；第二部分交替使用扩张卷积和普通卷积以扩大感受野；两部分特征图通过 Concat 融合后输入第三部分，该部分利用注意力机制关注重要信息，提升性能。实验表明，该网络可以通过增加宽度、减少深度的方法有效的消除因网络过深带来的细节消失等问题，与其他算法相比，该网络能在保留图像细节信息的同时更加有效地去除图像噪声。

最后，在第五章提出一种目标检测算法 Swin-YOLO，该算法通过在主干网络中融入 Swin Transformer 模块，增强了网络对复杂背景的适应能力和细节信息的提取能力，有效改善了道路小目标的误检和漏检问题；同时，集成 CBAM 注意力机制以更高效地提取有用特征，提升性能；此外，将预测边界框的损失函数从 CIoU 替换为 Inner-IoU，进一步增强了小目标的检测能力，提高了检测精度。

实验结果表明，虽然增加这些模块会是算法的 FPS 变小，但这些改进会使网络的检测精度更高，与此同时，还能更好地检测出遮挡目标和小目标，并且减少错检和误检。

6.2 研究展望

尽管本文提出的复杂环境下的低质图像处理和检测研究算法已经能够达到较好的性能，但为了进一步提升低质图像的目标分类性能，未来我们将尝试引入多模态融合技术。多模态融合技术通过整合来自不同模态（如视觉、红外、深度信息等）的数据，能够显著提升模型对复杂环境的适应性和鲁棒性。这种技术的优势在于能够捕捉不同模态之间的互补信息，从而更全面地理解图像内容，增强特征表示能力；与此同时，我们还将探索更高效的算法和更轻量级的模型，并应用在端设备上；除此之外，由于目前缺乏能够同时适用于增强、去噪以及目标检测的数据集，因此将三个创新点完全串联在一起存在困难，这也是未来研究的方向之一；最后，我们将致力于提高模型的可解释性和可控性，使其更加可信且易于调整，从而更好地满足实际应用中的需求。

参考文献

[1]刘振宇, 江海蓉, 徐鹤文. 极端天气条件下低质图像增强算法研究[J]. 计算机工程与应用, 2017, 53(08):193-198.

[2] Chang B C, Li J J, Lu R et al. Dual branch Transformer-CNN parametric filtering network for underwater image enhancement[J]. Journal of Visual Communication and Image Representation, 2024, 100:104131.

[3] Han W W, Xiao Y G, Yu Y. UM-GAN: Underground mine GAN for underground mine low-light image enhancement[J]. IET Image Processing, 2024, 18(8):2154-2160.

[4] 丁畅，董丽丽，许文海. “直方图”均衡化图像增强技术研究综述[J]. 计算机工程与应用, 2017, 53(23):12-17.

[5] Pizer S M, Amburn E P, Austin J D et al. Adaptive histogram equalization and its variations[J]. Computer Vision, Graphics, and Image Processing, 1987, 39(3): 355-368.

[6] Zuiderveld K. Contrast limited adaptive histogram equalization[J]. Graphics gems IV, 1994: 474-485.

[7] Jobson D J, Rahman Z, Woodell G A. Properties and performance of a center surround retinex[J]. IEEE Transactions on Image Processing, 1997, 6(3):451-462.

[8] Rahman Z, Jobson D J, Woodell G A. Multi-scale retinex for color image enhancement[C]. IEEE International Conference on Image Processing, Lausanne, Switzerland, 1996:1003-1006.

[9] Chen W, Wang W, Liu J et al. Deep retinex decomposition for low-light enhancement[C]. British Machine Vision Conference (BMVC), Newcastle, UK, 2018:1-12.

[10] Jiang Y L, Zhu J H, Li L L et al. A joint network for Low-light image enhancement based on retinex[J]. Cognitive Computation, 2024, 16(1):3241-3259.

[11] Gao H, Liu Z, Laurens V D M et al. Densely connected convolutional networks[C]. IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 2017:4700-4708.

[12] Wang Z X, Qing L T, Qingyi Pan et al. Retinex decomposition based lowlight image enhancement by integrating Swin transformer and U-Net−like architecture[J]. IET Image Processing, 2024, 18(11):3028-3041.

[13]陈红阳，曾上游，赵俊博. 基于 Retinex 改进的低照度图像增强网络[J].计算机应用与软件,2025,42(01):158-166.

[14] Jin Y, Jiang W Y, Shao J L, et al. An improved image denoising model based on nonlocal means filter[J]. Mathematical Problems in Engineering, 2018, 2(1):1-12.

[15] Mukund N Naragund, B. Jagadale, Vijaya Hegde. Improved image denoising with the integrated model of gaussian filter and neighshrink SURE[J]. International Journal of Engineering and Advanced Technology, 2019, 8(6):3010-3015.

[16] Xing Q W, Chen C Y, Li Z H. Progressive path tracing with bilateral-filtering-based denoising[J]. Multimedia Tools and Applications, 2020, 80(1):1529-1544.

[17] Bin Y W, Long H H. Image denoising based on adaptive sector rotation median filter[J]. Journal of Physics Conference Series, 2021, 1769(1):012056.

[18] Yousefi M I, Kschischang F R. Information transmission using the nonlinear fourier transform，Part II：Numeri- cal Methods[J]. IEEE Transactions on Information Theory, 2014,60(7):4329-4345.

[19] Tian C W, Zheng M H, Zuo W M et al. Multi-stage image denoising with the wavelet transform[J]. Pattern Recognition, 2023, 134:109050.

[20] Jain V, Seung H S. Natural image denoising with convolutional networks[C]. International Conference on Neural Information Processing Systems, Vancouver British Columbia Canada, 2008:769-776.

[21] Wang X T, Tang Y B, Cheng Y et al. DuINet: A dual-branch network with information exchange and perceptual loss for enhanced image denoising[J]. Digital Signal Processing, 2025, 156:104835.

[22] Zhang Q, Xiao J Y, Zhang S C et al. Texture-guided CNN for image denoising[J]. Multimedia Tools and Applications, 2024, 83(1):63949-63973.

[23]何涛，王超，吴贵铭. 基于卷积神经网络的暗光图像降噪算法研究[J].传感器与微系统,2023,42(12):64-67.

[24] Girshick R, Donahue J, Darrell T et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]. IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 2014:580–587.

[25] Girshick R. Fast R-CNN[C]. IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 2015:1440-1448.

[26] Ren S, He S, Girshick R et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6):1137-1149.

[27] Liu W, Auguelov D, Erhan D, et al. SSD: single shot multiBox detector[C]. European Conference on Computer Vision (ECCV), Amsterdam, The Netherlands, 2016:21-37.

[28] Redmon J, Divvala S, Girshick R and Farhadi A. You only look once: unified real-time object detection[C]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016:779-788.

[29]Redmon J and Farhadi A. YOLO9000: better faster stronger[C]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017:7263-7271.

[30] Redmon J. YOLOv3: an incremental improvement [EB/OL]., 2018-04-08/2025-02-15.

[31] Bochkovskiy A. YOLOv4: Optimal speed and accuracy of object detection [EB/OL]. https://arxiv.org/pdf/2004.10934, 2020-04-23/2025-02-15.

[32] Wang C Y, Alexey B, Liao H Y. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 2023:7464-7475.

[33] Varghese R, Sambath M. YOLOv8: a novel object detection algorithm with enhanced performance and robustness[C]. International Conference on Advances in Data Engineering and Intelligent Computing Systems, Chennai, India, 2024:1-6.

[34] Wang C Y, Yeh I, Liao H Y. YOLOv9: Learning what you want to learn using programmable gradient information[C]. Computer Vision, Milan, Italy, 2024:1-21.

[35] Wang A, Chen H, Liu L H et al. YOLOv10: real-time end-to-end object detection[C]. Conference on Neural Information Processing Systems, Vancouver, 2024:1-21.

[36] Khanam R. YOLOv11: An overview of the key architectural enhancements [EB/OL]. https://arxiv.org/pdf/2410.17725, 2024-10-23/2025-02-15.

[37] Land E H. The retinex theory of color vision[J]. Scientific American, 1977, 237(6):108-128.

[38] Yu F. Multi-scale context aggregation by dilated convolutions [EB/OL]. https://arxiv.org/pdf/1511.07122, 2015-11-23/2025-02-15.

[39] Du B, Wei Q C, Liu R. An improved quantum-behaved particle swarm optimization for endmember extraction[J]. IEEE Transactions on Geoscience and Remote Sensing, 2019, 57(8): 6003-6017.

[40] Li J X, Lu G M, Zhang B, et al. Shared linear encoder-based multikernel gaussian process latent variable model for visual classification[J]. IEEE Transactions on Cybernetics, 2021, 51(2): 534-547.

[41] Xu L, David Z, Lu G M. A novel multicamera system for high-speed touchless palm recognition[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2021, 51(3): 1534-1548.

[42] Li Y W, Chen X Z, Zhu Z. Attention-guided unified network for panoptic segmentation[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 2019:7026-7035.

[43] VASWANI A, SHAZEER N, PARMAR N. Attention is all you need[C]. Proceedings of the 31st International Conference on Neural Information Processing Systems, Red Hook, NY, United States, 2017:6000-6010.

[44] BA J L. Layer normalization [EB/OL]. https：//arxiv.org/abs/1607.06450, 2016-07-21/2025-02-15.

[45] Rezatofighi H, Tsoi N, Gwak J et al. Generalized intersection over union: A metric and a loss for bounding box regression[C]. Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 2019:658-666.

[46] Zheng Z, Wang P, Liu W et al. Distance-IoU loss:Faster and better learning for bounding box regression[J]. AAAI Conference on Artificial Intelligence, 2020, 34(07):12993-13000.

[47] Wang J, Liu M, Ss Y et al. Small target detection algorithm based on attention mechanism and data augmentation[J]. Signal, Image and Video Processing, 2024, 18(4):3837-3853.

[48] Zhang Y, Ren W, Zhang Z et al. Focal and efficient IOU loss for accurate bounding box regression[J]. Neurocomputing, 2022, 506(9):146-157.

[49] Zhora G. SIoU Loss: More Powerful Learning for Bounding Box Regression [EB/OL]. https://arxiv.org/abs/2205.12740, 2022-05-25/2025-02-15.

[50] Hao Z. Inner-IoU: More Effective Intersection over Union Loss with Auxiliary Bounding Box [EB/OL]. https://arxiv.org/abs/2311.02877, 2023-11-06/2025-02-15.

[51] 李佳,王婷,杨文杰,等.基于 Retinex 理论的双重注意力 Transformer 的低光照图像增强[J]. 计算机系统应用, 2025,:1-13

[52] Priyadharshini R A, Arivazhagan S, Pavithra K A. An ensemble deep learning approach for underwater image enhancement[J]. e-Prime - Advances in Electrical Engineering, Electronics and Energy, 2024, 9:100634.

[53] Ioffe S, Christian S. Batch normalization: Accelerating deep network training by reducing internal covariate shift[J]. Proceedings of the 32nd International Conference on International Conference on Machine Learning, 2015, 37(1):448–456.

[54] He K, Zhang X, Ren S et al. Deep residual learning for image recognition[C]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016:770-778.

[55] Tian C, Xu Y, Li Z, et al. Attention-guided CNN for image denoising[J]. Neural Networks, 2020, 124(1):117-129.

[56] Wang X T, Tang Y B, Cheng Y et al. DuINet: A dual-branch network with information exchange and perceptual loss for enhanced image denoising[J]. Digital Signal Processing, 2025, 156:104835.

[57] Zhang Q, Xiao J Y, Zhang S C et al. Texture-guided CNN for image denoising[J]. Multimedia Tools and Applications, 2024, 83(23):1-25.

[58] 何涛, 王超, 吴贵铭. 基于卷积神经网络的暗光图像降噪算法研究[J]. 传感器与微系统, 2023, 42(12):64-67.

[59] Ephraim Y, Malah D. Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator[J]. IEEE Transactions on acoustics, speech, and signal processing, 1984, 32(6), 1109–1121.

[60] Kingma D. Adam: A Method for Stochastic Optimization [EB/OL]. https://arxiv.org/pdf/1412.6980, 2014-12-22/2025-02-15.

[61] 刘海斌, 张友兵, 周奎等. 改进 YOLOv5-S 的交通标志检测算法[J]. 计算机工程与应用, 2024, 60(05):200-209.

[62] Liu Z, Lin Y, Cao Y et al. Swin transformer:Hierarchical vision transformer using shifted windows[C]. IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 2021:9992-10002.

[63] Zhang D, Zhang Z, Zhao N et al. A lightweight modified YOLOv5 network using a swin transformer for transmission-line foreign object detection[J]. Electronics, 2023, 12(18):3904-3920.

[64] Gong H, Mu T, Li Q et al. Swin-transformer-enabled YOLOv5 with attention mechanism for small object detection on satellite images[J]. Remote Sensing, 2022, 14(12):2861-2878.

[65] Zhang B, Cheng H, Zhong H et al. Real-time detection of voids in asphalt pavement based on swin-transformer-improved YOLOv5[J]. IEEE Transactions on Intelligent Transportation Systems, 2024, 25(3):2615-2626.

[66] Ma J, Wang X, Xu C et al. SF-YOLOv5: Improved YOLOv5 with swin transformer and fusion-concat method for multi-UAV detection[J]. Sage Journals, 2023, 56(7-8):1436-1445.

[67] Han J, Li Z, Cui G et al. EGS-YOLO: A fast and reliable safety helmet detection method modified based on YOLOv7[J]. Applies Sciences, 2024, 14(17):7923-7942.

攻读学位期间参与项目

[1]安徽工程大学校级预研项目（Xjky072019A02）（参与）

[2]安徽省高校协同创新项目（GXXT-2019-007）（参与）

攻读学位期间研究成果

一、发表论文

[1] 龚佳乐, 窦易文.Swin-YOLOv5：面向复杂交通环境下的小目标检测方法[J/OL].重庆工商大学学报(自然科学版), 2024, 1-13.(第一作者)

[2] Jia-Le Gong, Yiwen Dou, Yi-Ting Gao et al. RSARNet: A robust network for Gaussian noise removal[J]. Soft Computing. (在审).(第一作者)

二、专利

[1] 一种基于 SpA-YOLOv7 的跌倒检测方法（实质审查）。

[2] 基于深度学习的交通要素多目标检测方法（受理）。

致谢

时光荏苒，岁月如梭。经过数月的辛勤努力，我的毕业论文终于接近尾声。回首这段旅程，充满了挑战与收获，也离不开众多师长、同学和家人的支持与帮助。在此，我衷心感谢每一位给予我支持和鼓励的人。

首先，我要向我的导师窦易文老师致以最崇高的敬意和最诚挚的感谢。在论文写作过程中，窦易文老师不仅是我的学术引路人，更是我成长道路上的明灯。从选题的初步构思到最终成稿，窦易文老师始终以严谨的学术态度和丰富的专业知识为我指明方向。在选题阶段，老师凭借敏锐的学术洞察力，帮助我从众多方向中精准定位，确保研究的可行性和创新性；在研究过程中，老师耐心指导我如何查阅文献、整理数据，并在每一次讨论中提出关键性问题，引导我深入思考。即使在面对复杂的研究难题时，老师也总是鼓励我大胆尝试，从不放弃。老师严谨的治学态度和高尚的人格魅力，将永远激励我在学术道路上不断前行。

其次，我要感谢我的师兄-缪红超，在研究生一年级和二年级时，当我在研究过程中遇到问题时，缪师兄总是第一时间并且十分耐心的为我解答问题，帮助我成功度过了前期较为困难的研究生活。除此之外，我还要感谢我的两位师妹-高依婷和高梅，在小论文和大论文的写作过程中，她们会为我积极纠错，还会与我共同探讨学术问题，分享研究心得。与同门师兄妹的相处过程中，我不仅感受到了浓郁的学术氛围，还获得了真挚的友谊，我们一起讨论、交流的时光，将成为我研究生生活中最美好的回忆。

此外，我要感谢安徽工程大学为我提供了良好的学习环境和丰富的学术资源。在这里，我度过了充实而美好的研究生时光，并且收获了知识、友谊和成长，这将是我最大的财富。

最后，非常感谢对我论文提出宝贵意见的评审专家和老师。

再次感谢所有关心、支持和帮助我的人！未来，我将以更加饱满的热情和坚定的信念，为社会贡献出属于自己的那一份力量。