引用格式：WANG Manli, CHEN Chen, JIA Yunhong, et al. Nighttime Image Enhancement Method Baesd on Multi-scale Fusion Network[J]. Acta Photonica Sinica, 2026, 55(1):0110003
王满利,陈晨,贾运红,等.基于多尺度融合网络的夜间图像增强方法[J].光子学报,2026,55(1):0110003

基于多尺度融合网络的夜间图像增强方法

王满利[1],陈晨[1],贾运红[2],史艳楠[1],乔应旭[3]
(1 河南理工大学 物理与电子信息学院,焦作 454003)
(2 煤炭无人化开采数智技术全国重点实验室,北京 100013)
(3 河南理工大学 计算机科学与技术学院,焦作 454003)

摘 要:夜间图像增强是指对光照条件不足的图像提升亮度、增强对比度并恢复细节。针对现有夜间图像增强方法增强后纹理细节丢失、噪声放大问题,提出一种基于多尺度融合网络的夜间图像增强方法。首先,在主干网络前端引入暗区特征提取模块,通过卷积提取图像特征并生成通道注意力权重,增强与细节信息相关的通道特征,从而有效提取暗区纹理细节。其次,设计多分支网络模块,通过三个子网络并行提取不同尺度下的图像特征,实现夜间图像全局结构和局部细节的恢复。最后,提出动态权重分配机制,自适应融合各子网络的多尺度特征,强化特征表达能力,进一步提升图像增强效果。实验结果表明,本文方法在LOLv1数据集上对比Retinexformer方法PSNR和SSIM值分别提高5.4%和1.4%;在LOLv2、LIME、NPE、MEF多个基准数据集上,增强后的图像获得了更好的主观视觉效果。此外,在公开夜间人脸检测数据集DARKFACE上,经过增强后的图像人脸检测精度更高,展现了该方法在下游任务上的有效性。

关键词:夜间图像增强;暗区特征提取;多分支网络;动态权重分配;人脸检测
中图分类号:TP391.41 **文献标识码**:A **doi**:10.3788/gzxb20265501.0110003

0 引言

随着智能设备的广泛应用和人工智能技术的发展,夜间图像增强在视觉物联网[1-2]、自动驾驶[3-4]、医学影像[5-6]和目标检测[7-8]等多个领域展现出重要价值。夜间图像增强是指对夜间环境下拍摄的图像进行优化,提升其亮度和对比度。然而,在图像增强过程中如何有效保留纹理细节并抑制噪声,仍是一项具有挑战性的任务。

夜间图像增强方法[9]主要分为三类:基于分布映射的方法、基于模型优化的方法以及基于深度学习的方法。基于分布映射的方法通过重新分配图像的像素值,改善图像整体对比度,达到增强效果。代表性方法如直方图均衡化[10]、伽马校正[11]、色调映射[12]。其中,直方图均衡化通过调整像素强度分布来提升图像的整体或局部对比度;伽马校正利用非线性变换改变灰度值,以改善亮度和对比度;色调映射则通过优化动态范围进一步增强图像的亮度与对比度。然而,由于夜间图像中存在噪声,这类方法直接对原始图像进行像素分布上的调整,增强过程中噪声会被同步放大。

基于模型优化的方法通过优化物理模型,对图像的退化进行建模,进而恢复图像在不同光照条件下的真实颜色和细节。其中最具代表性的是Retinex算法[13],将图像分解为反射分量和光照分量,通过估计图像的光照分量来增强图像。Retinex算法在一定程度上提升了图像质量,但在估计光照分量过程中容易导致图像失真。

近年来,深度学习为夜间图像增强提供了全新思路。GUO Chunle等[14]提出了零参考深度曲线估计方

基金项目:国家自然科学基金(52274160, 62472145),河南理工大学博士基金(B2021-64),河南省光电检测与传感集成工程技术研究中心开放课题基金(KFZ202502)
第一作者:王满利,wml920@163.com
通讯作者:陈晨,cheeenc@163.com
收稿日期:2025-08-08;**录用日期**:2025-10-10

法(Zero Reference Deep Curve Estimation, ZRDCE),将光照增强任务转化为特定曲线的估计问题,通过训练轻量级深度网络来预测像素级高阶曲线,实现对图像动态范围的精确调整。JIANG Yifan等[15]提出了无监督对抗生成网络(Enlighten Generative Adversarial Network, EGAN),在没有成对训练数据的情况下,利用输入图像本身的信息来进行正则化,实现夜间图像的增强。MA Long等[16]提出了自校准照明网络模型(Self Calibrated Illuminatio, SCI),通过权重共享网络提取各阶段特征,结合保真度损失和平滑损失组成无监督损失函数来提升模型的泛化性。LI Chongyi等[17]设计了一种基于傅里叶变换的图像增强方法(Ultra High Definition Fourier, UHDF),将傅里叶变换嵌入到级联网络中,分别处理图像的振幅和相位分量。XU Xiaogang等[18]提出了SNRAN(Signal Noise Ration Aware Network),利用图像的信噪比先验,在低信噪比区域引导模型整合距离较远的特征信息进行修复,在高信噪比区域依赖局部特征完成修复。LUO Yu等[19]提出伪监督学习框架,利用二次曲线生成伪清晰图像,消除了对配对数据集的依赖。将夜间图像和伪清晰图像分别输入两个分解网络,通过光照损失、反射损失和互损失约束,实现对光照层和反射层的联合优化。XIE Chao等[20]提出一种残差商学习网络(Residual Quotient Learning Network, RQLN),将光照增强任务重新定义为自适应预测潜在商数,并结合残差学习进行优化,通过三通道光照建模,有效恢复了图像的颜色。FEIJOO Daniel等[21]提出基于图像分解的光照推理增强框架,通过光照建模、频域融合与注意力机制协同设计,实现了光照自适应的图像增强。然而,由于光照条件的限制,夜间图像往往存在纹理细节模糊的问题,导致现有方法在增强时容易造成细节丢失。此外,在提升图像亮度与对比度的同时,噪声也会被同步放大。

针对上述挑战,本文提出一种基于多尺度融合网络的夜间图像增强方法(Multi-scale Fusion Network, MSFNet)。首先,在主干网络前端引入暗区特征提取模块(Dark Area Feature Extraction Module, DAFEM),通过卷积提取图像特征并生成注意力权重,增强与细节信息相关的通道特征,缓解图像细节纹理丢失问题。其次,在主干网络中构建多分支网络模块(Multi-branch Network Module, MBNM),通过设计不同尺度的分支子网络提取多尺度信息。其中,大尺度分支侧重提取全局信息,恢复图像的整体结构与亮度;中等和较小尺度分支侧重提取局部信息,实现图像的精细化恢复。最后,提出动态权重分配机制(Dynamic Weight Allocation Mechanism, DWAM),以自适应方式为不同尺度特征分配权重,实现多尺度特征的有效融合,增强特征表达能力,进一步提升图像增强效果。实验结果表明,MSFNet在公开配对数据集、非配对数据集上的表现均优于现有方法。此外,MSFNet增强后的夜间人脸图像获得了更高的检测精度,说明该方法能较好的适配下游任务。

1 本文方法

1.1 网络整体框架

图1展示了现有方法增强图像与真实图像的可视化对比。结果表明,现有方法增强图像不仅存在玻璃纹理细节缺失问题,还伴随明显的噪声放大的现象。为解决现有夜间图像增强方法增强后图像纹理细节丢失、噪声放大问题,本文提出MSFNet,如图2所示。MSFNet采用U型网络[22]作为主干,通过跳跃连接在不同层级中有效提取图像的细节和上下文信息。首先,在主干网络前端引入DAFEM,用于提取图像暗区的纹理特征,缓解增强过程中纹理细节丢失的问题;主干网中的编码器和解码器均由MBNM组成,该模块包含三个专注于不同尺度图像信息的子网络,能够从多尺度感知图像内容,并在不同尺度上恢复图像信息。此外,MBNM还包含DWAM,用于自适应地为不同尺度特征分配权重,并通过加权融合多

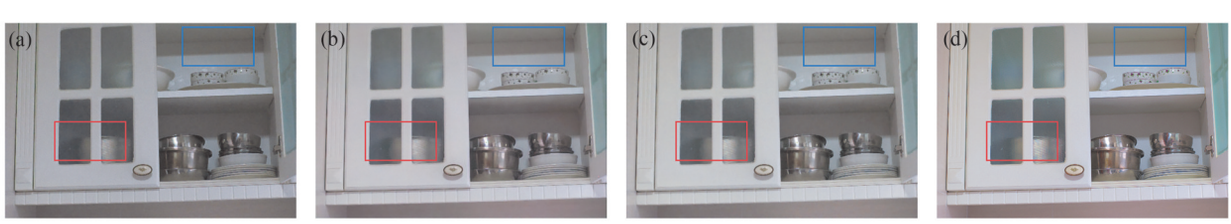


图1 现有方法增强图像可视化对比。(a)Restormer;(b)UHDFour;(c)Retinexformer;(d)GroundTruth

Fig. 1 Existing method enhance image visual comparison. (a) Restormer; (b) UHDFour; (c) Retinexformer; (d) GroundTruth

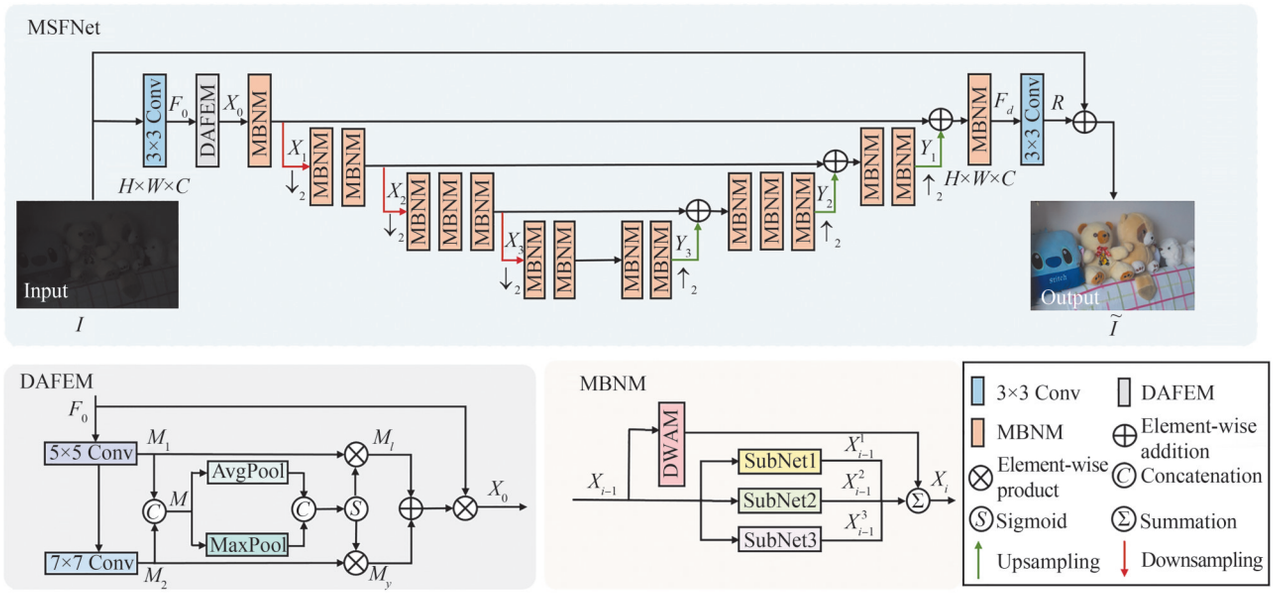


图2　网络模型整体结构

Fig. 2　The overall structure of the network model

尺度信息，进一步提升夜间图像的视觉质量。最终，经过多级编码器-解码器结构，输出增强后的夜间图像。

1.2　暗区特征提取模块

根据Retinex理论[13]，图像可以分解为物体反射分量和照明分量的乘积，但实际拍摄条件不理想，存在噪声干扰。因此，夜间图像可表示为

$$I_{\text{low}}(x,y)=R(x,y)\odot L(x,y)+N(x,y) \tag{1}$$

式中，$I_{\text{low}}(x,y)$表示夜间图像，$R(x,y)$表示反射分量，$L(x,y)$表示照明分量，$N(x,y)$为噪声。在暗区中，照明分量$L(x,y)\to 0$，导致$R(x,y)\odot L(x,y)$项被大幅抑制。与此同时，噪声项$N(x,y)$并未被同步抑制，其相对比例反而被放大。图像虽然可以保留结构轮廓信息，但纹理细节极易被噪声掩盖。为了有效提取暗区细节纹理特征，本文提出DAFEM，如图2所示。DAFEM利用不同大小卷积从两个尺度感知夜间图像特征，结合平均池化与最大池化生成互补的通道特征图，通过通道注意力机制强化与细节信息相关的通道。最后，采用残差连接融合特征，实现夜间图像细节纹理特征的强化与恢复。具体过程如下所示：

首先，输入特征F_0通过5×5的卷积分支得到特征图M_1，然后经过7×7的卷积分支进一步提取空间关系得到特征图M_2。随后，将两个特征图在通道维度拼接得到融合不同尺度语义信息的增强特征图M

$$M=\text{Conact}(M_1,M_2) \tag{2}$$

式中，Conact(·)为拼接操作。对融合特征M进行平均池化和最大池化，生成平均池化特征图Q_{avg}和最大池化特征图Q_{max}，将两者拼接后通过Sigmoid激活生成通道注意力权重SA

$$\text{SA}=\sigma\left(\text{Conact}(Q_{\text{avg}},Q_{\text{max}})\right) \tag{3}$$

式中，σ(·)表示Sigmoid激活函数。将注意力权重SA与两个尺度的特征图M_1和M_2点乘，得到空间注意力特征图M_l和M_y，从而强化与细节信息相关的通道。最后，采用残差连接融合特征得到输出X_0

$$X_0=F_0\cdot\left(M_l+M_y\right) \tag{4}$$

图3展示了特征提取效果对比。其中图3(a)为原始低光照图像，纹理难以分辨；图3(b)为3×3卷积提取的特征图，虽然部分边缘可见，但细节部分弱化；图3(c)为DAFEM提取的特征图，暗部区域纹理细节更加清晰(如图3中框出的位置)。

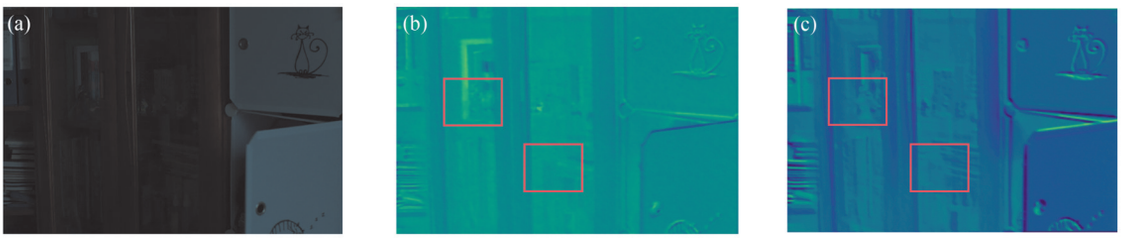


图3 特征提取效果对比。(a)输入图像;(b)3×3卷积特征提取图;(c)DAFEM特征提取图

Fig. 3 Comparison of feature extraction effects. (a) Input image; (b) 3×3 convolution feature extraction diagram; (c) DAFEM feature extraction diagram

2 多分支网络模块

夜间图像增强任务中,增强图像的整体结构和亮度需要使用较大尺度下的全局信息,实现精细恢复则需要选择中等和较小尺度的局部信息,不同尺度的信息关注图像的不同方面,采用多尺度处理方式可以帮助网络在不同尺度上有效恢复图像,从而提升增强效果。因此,本文提出MBNM,如图4所示,MBNM由三个子网络SubNet1、SubNet2、SubNet3和动态权重分配机制DWAM构成。三个子网络并行提取不同尺度下的图像特征,实现夜间图像全局结构和局部细节的恢复;DWAM根据输入特征动态地为三个子网络赋予权重,通过加权融合不同尺度信息,强化特征表达能力,进一步提升图像增强效果。

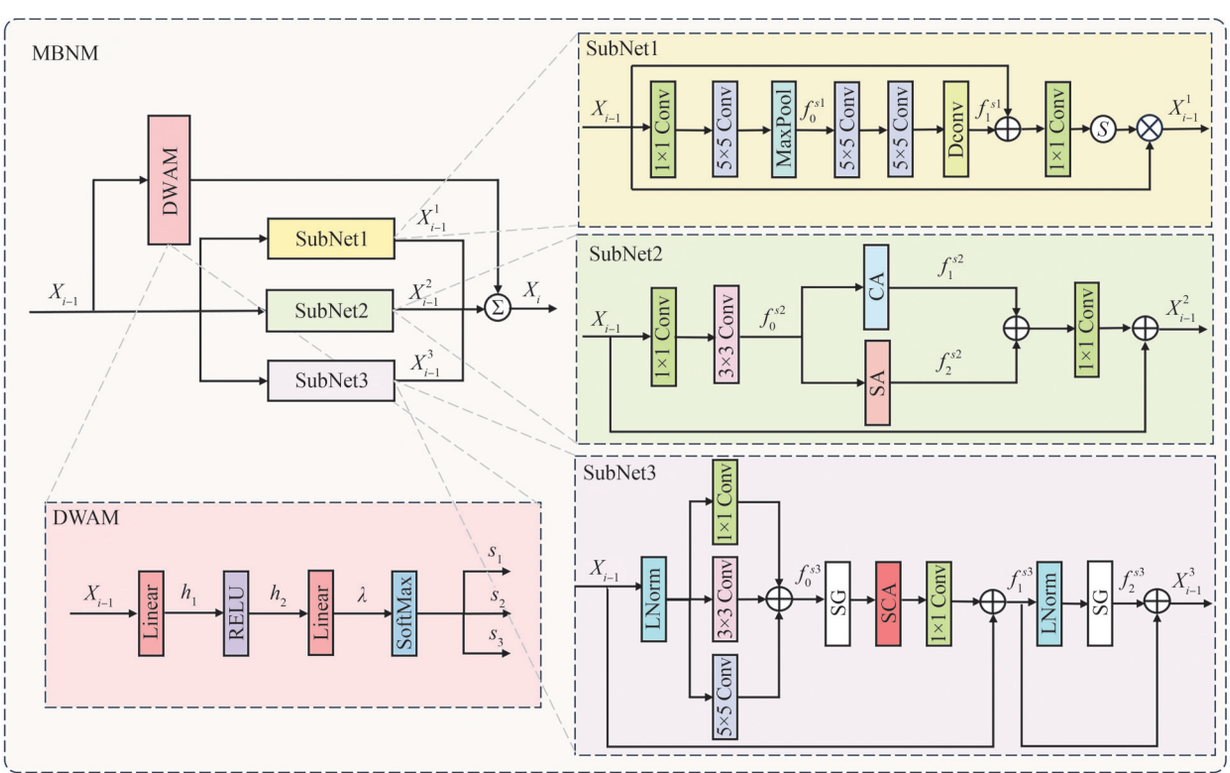


图4 多分支网络模块结构

Fig. 4 Structure of the multi-branch network module

2.1 多分支子网络

SubNet1通过大小为5×5卷积处理大尺度下的全局信息,侧重于恢复夜间图像的整体结构和亮度。给定输入特征X_{i-1},网络首先通过卷积操作来提取全局特征f_0^{s1}

$$f_0^{s1}=\phi_{\max}(\mathrm{Conv}_{5\times5}(\mathrm{Conv}_{1\times1}(X_{i-1}))) \tag{5}$$

式中,$\phi_{\max}(\cdot)$表示最大池化操作,$\mathrm{Conv}_{n\times n}(\cdot)$表示$n\times n$卷积操作。接着,通过反卷积层将特征图的尺寸恢复至原始图像大小,得到特征图f_1^{s1}

$$f_1^{s1}=\mathcal{B}(\mathrm{Conv}_{5\times5}(\mathrm{Conv}_{5\times5}(f_0^{s1}))) \tag{6}$$

式中，$\mathcal{B}(\cdot)$表示反卷积操作。将当前层的特征f_1^{s1}和来自上一层的特征X_{i-1}通过1×1卷积拼接，输入Sigmoid激活函数得到注意力权重图。最后，注意力权重图与原特征图X_{i-1}逐元素相乘，在特征图不同位置重新分配像素的权重，生成输出特征X_{i-1}^1。上述操作可表示为

$$X_{i-1}^1=X_{i-1}\cdot\sigma(\mathrm{Conv}_{1\times1}(X_{i-1}+f_1^{s1})) \tag{7}$$

SubNet2通过大小为3×3的卷积处理小尺度下的局部信息，结合通道注意力（Channel Attention，CA）和空间注意力（Spatial Attention，SA），侧重于恢复夜间图像的局部区域和细节。给定输入特征X_{i-1}，经过1×1卷积和3×3卷积后得到特征f_0^{s2}。CA模块对f_0^{s2}通道维度的特征进行加权，得到通道增强特征f_1^{s2}；SA模块则用于识别特征图中的重要空间区域，通过对每个位置上的特征进行加权，得到空间增强特征f_2^{s2}。将两个分支的增强特征进行融合，与原始输入特征X_{i-1}进行残差连接得到输出特征X_{i-1}^2

$$X_{i-1}^2=X_{i-1}+\mathrm{Conv}_{1\times1}(f_1^{s2}+f_2^{s2}) \tag{8}$$

SubNet3通过不同大小的卷积对大尺度和小尺度信息进行融合，结合简化门控机制（Simple Gate，SG）选择性增强图像区域，侧重于夜间图像的精细化恢复。如图5(a)所示，SG的输入特征分成维度为$(C/2,\ W,\ H)$的上下分支两部分。上分支是主特征分量，承载了来自前一层的特征；下分支是门控分量，为每个空间位置生成一个权重图，决定了每个位置每个通道的信息保留程度。将两者进行逐元素相乘，得到最终输出。

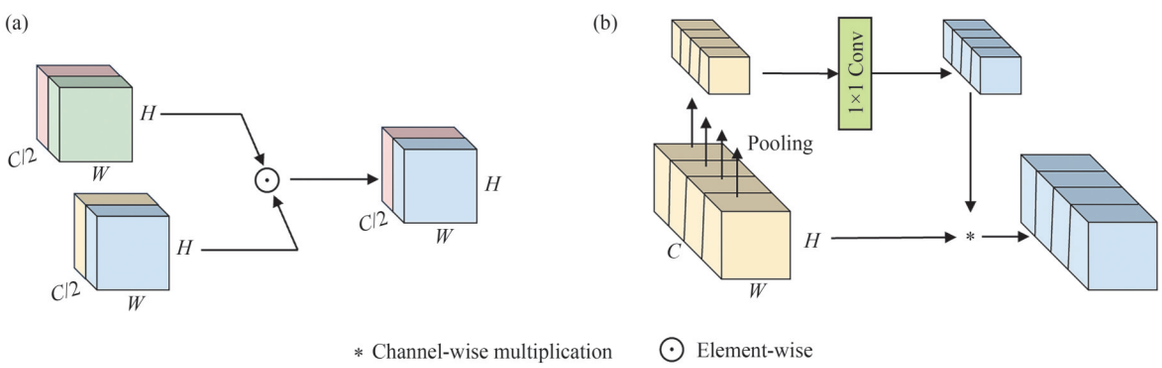


图5　简化门控和简化通道注意力结构。(a)简化门控；(b)简化通道注意力

Fig. 5　Structures of the simple gate and simplified channel attention. (a) Simple gate; (b) Simplified channel attention

首先，SubNet3对输入特征X_{i-1}进行层归一化处理，通过多个大小不同的卷积（1×1、3×3、5×5卷积）融合图像的不同尺度特征，进一步提取细节得到特征图f_0^{s3}

$$f_0^{s3}=\varphi(\mathrm{LN}(X_{i-1})) \tag{9}$$

式中，LN(·)为层归一化操作，$\varphi(\cdot)$为多卷积融合操作。随后，提出简化通道注意力机制（Simple Channel Attention，SCA），如图5(b)所示，SCA只使用一个卷积层，减少了计算量。

将特征图f_0^{s3}经过SG和SCA得到特征图f_1^{s3}。随后，通过层归一化操作和第二次SG得到细化后的特征图f_2^{s3}，将两个特征图进行加权融合得到输出特征X_{i-1}^3。f_1^{s3}和X_{i-1}^3的计算公式分别为

$$f_1^{s3}=X_{i-1}+\alpha\cdot(\mathrm{Conv}_{1\times1}(\mathrm{SCA}(\mathrm{SG}(f_0^{s3})))) \tag{10}$$

$$X_{i-1}^3=f_1^{s3}+\beta\cdot f_2^{s3} \tag{11}$$

式中，SG(·)为简化门控操作，SCA(·)为简化通道注意力操作，α、β为可学习的融合权重。

2.2　动态权重分配机制

通过重复堆叠多尺度模块实现特征交互，不同尺度的特征在交叉融合过程中容易产生大量重复信息。此外，固定结构的特征融合无法根据图像内容动态调整，存在较大局限性。因此，本文提出了动态权重分配机制DWAM，如图4所示。DWAM动态地为不同尺度信息赋予权重，通过加权处理不同尺度信息，增强特征表达能力，进一步提升夜间图像的增强效果。首先，将输入特征X_{i-1}通过线性变换降维得到h_1。然后，使用RELU激活函数得到h_2。h_1、h_2计算公式分别为

$$\begin{cases} h_1 = \mathrm{Linear}_1(X_{i-1}) = W_1 X_{i-1} + b_1 \\ h_2 = \mathrm{RELU}(h_1) \end{cases} \tag{12}$$

式中，Linear(•)是线性变换操作，RELU(•)是RELU激活函数操作，W_1和b_1为线性变换中可学习的参数。将h_2线性映射为3个通道，得到用于表示不同分支子网络重要程度的动态权重λ

$$\lambda = \mathrm{Linear}_2(h_2) = W_2 h_2 + b_2, \quad \lambda \in \mathbb{R}^{H \times W \times 3} \tag{13}$$

式中，W_2和b_2为线性变换中可学习的参数。最后，使用Softmax函数生成归一化后的动态权重分量

$$s_i = \frac{\exp(\lambda_i)}{\sum_{j=1}^{3} \exp(\lambda_j)}, \quad i = 1, 2, 3 \tag{14}$$

式中，λ_i为第i分支子网络对应的动态权重分量，s_i为归一化后第i分支子网络对应的动态权重分量。将三个分支子网络的输出特征进行加权融合，得到最终融合特征X_i

$$X_i = \sum_{j=1}^{3} s_j \cdot X_{i-1}^{j} = s_1 \cdot X_{i-1}^{1} + s_2 \cdot X_{i-1}^{2} + s_3 \cdot X_{i-1}^{3} \tag{15}$$

式中，X_{i-1}^{j}为第j个子网络的输出特征，X_i为融合后的特征。

3 实验与结果分析

3.1 实验数据集

本文方法在6个公共数据集LOLv1[23]、LOL-v2[24]、DARKFACE[25]、LIME[26]、NPE[27]和MEF[28]上进行实验。

LOLv1数据集由500对图像组成，分为485对训练图像和15对测试图像。大多数图像是室内场景，所有图像的分辨率均为400×600。

LOL-v2数据集包括两个不同的子集，即LOL-v2-real和LOL-v2-syn。LOL-v2-real是在真实场景中通过改变ISO和曝光时间捕获，包括689对训练图像和100对测试图像。LOL-v2-syn通过分析夜间图像的照明分布，从RAW图像合成夜间图像，包括900对训练图像和100对测试图像。

DARKFACE数据集包含6 000张在夜间拍摄的真实图像，拍摄地点涵盖教学楼、街道、桥梁、立交桥和公园等，所有图像均标注有人脸的边界框。

LIME、NPE和MEF是3个没有配对图像的数据集，每个数据集包含几十张在真实场景下拍摄的夜间图像。

3.2 实验环境和评价指标

本实验使用4块型号为GeForce RTX 3090的GPU进行训练，深度学习框架版本号为Pytorch1.11.0，python版本号为3.9.5。本网络训练时的图片批量大小设置为4，使用Adam优化器对模型进行200 000次迭代训练，学习率设置为0.001。

对于配对数据集，由于存在对应的正常光照图像作为参考，采用有参考图像质量评价指标：峰值信噪比(Peak Signal Noise Ratio, PSNR)、结构相似性(Structural Similarity, SSIM)、学习感知图像块相似度(Learned Perceptual Image Patch Similarity, LPIPS)。对于非配对数据集，由于缺乏对应的正常光照图像作为参考，采用无参考图像质量评价指标：自然图像质量评估(Natural Image Quality Evaluator, NIQE)。PSNR用于衡量图像的重建质量，SSIM用于评估图像结构的重建质量，LPIPS评估图像细节的恢复效果，NIQE主要验证图像的自然程度。

3.3 配对数据集对比实验

表1展示了本文方法与RUAS[29]、UFormer[30]、Restormer[31]和Retinexformer[32]等多种先进的算法在LOLv1、LOLv2-real和LOLv2-syn三个数据集上的比较结果(表现最优的结果用粗体标示)。在三个数据集上，MSFNet的PSNR值分别比Retinexformer高出1.37 dB、0.72 dB和0.50 dB；在LOLv1和LOLv2-syn数据集上SSIM值分别比Retinexformer高出0.012和0.007。MSFNet的FLOPS为18.23 G，参数量为2.14 M。相比于Retinexformer，在不明显增加计算复杂度的前提下，能够有效提升图像质量。

表1　在三个数据集上的性能对比

Table 1　Performance comparison on three different datasets

Methods	Complexity		LOLv1		LOLv2-real		LOLv2-syn	
	FLOPS(G)	Params(M)	PSNR↑(dB)	SSIM↑	PSNR↑(dB)	SSIM↑	PSNR↑(dB)	SSIM↑
EnGAN[15]	61.02	114.35	17.48	0.650	18.23	0.617	16.57	0.734
RUAS[29]	**0.83**	**0.003**	18.23	0.720	18.37	0.723	16.55	0.652
SCI[16]	2.15	0.81	14.78	0.526	17.30	0.541	17.25	0.552
UFormer[30]	12.00	5.29	16.36	0.771	18.82	0.771	19.66	0.871
Restormer[31]	144.25	26.13	22.43	0.823	19.94	0.827	21.41	0.830
SNRNet[18]	28.47	3.82	24.61	0.842	21.48	0.849	24.14	0.928
UHDFour[17]	35.62	4.95	23.09	0.821	21.78	**0.854**	23.64	0.899
Retinexformer[32]	15.57	1.61	25.16	0.845	22.80	0.840	25.67	0.930
Ours	18.23	2.14	**26.53**	**0.857**	**23.52**	0.806	**26.17**	**0.937**

为了进一步验证MSFNet的增强效果，在LOLv1、LOLv2-real和LOLv2-syn数据集上开展对比实验，选取了多种代表性算法进行比较。在图6中，EnGAN和SCI生成的图像存在明显的颗粒感，增强过程中放大了噪声；而Restormer和UHDFour生成的图像则存在边缘模糊和细节丢失问题；在图7中，RUAS生成的图像表现曝光不足，而SNRNet和Retinexformer生成的图像则出现颜色失真现象。

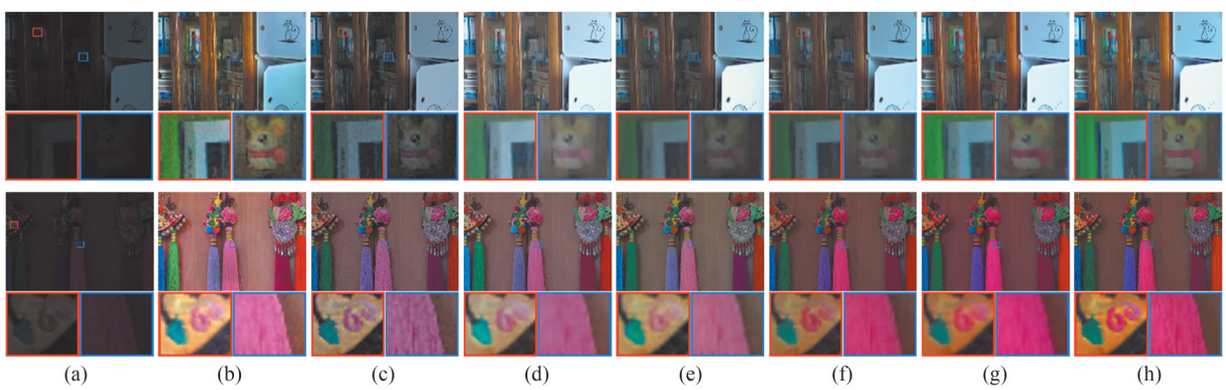


图6　LOLv1数据集上可视化对比。(a)输入；(b)EnGAN；(c) SCI；(d) Restormer；(e) UHDFour；(f) Retinexformer；(g) Ours；(h) GroundTruth

Fig. 6　Visual comparison on the LOLv1 dataset. (a) Input; (b) EnGAN; (c) SCI; (d) Restormer; (e) UHDFour; (f) Retinexformer; (g) Ours; (h) GroundTruth

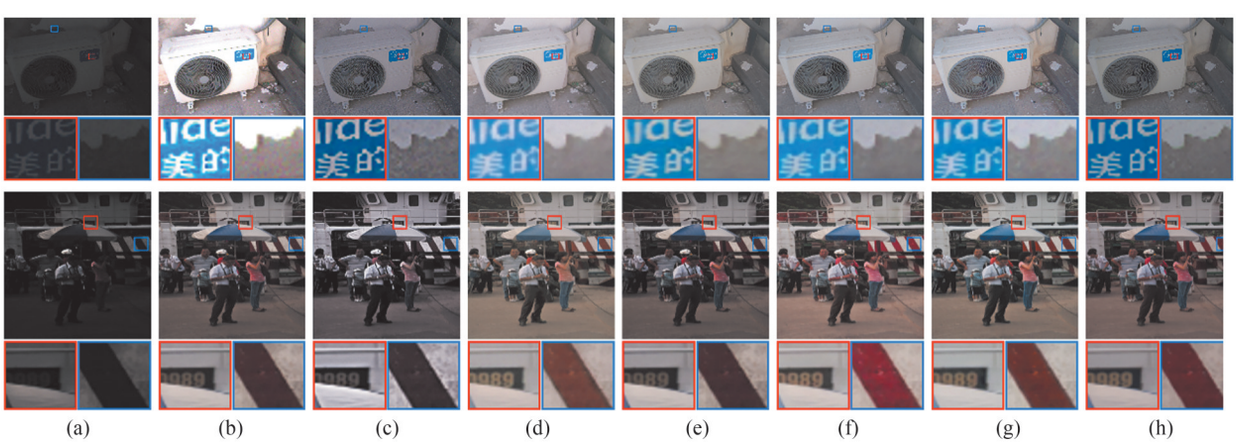


图7　LOLv2数据集上可视化对比。(a)输入；(b) RUAS；(c) SCI；(d) Restormer；(e) SNRNet；(f) Retinexformer；(g) Ours；(h)GroundTruth

Fig. 7　Visual comparison on the LOLv2 dataset. (a) Input; (b) RUAS; (c) SCI; (d) Restormer; (e) SNRNet; (f)Retinexformer; (g) Ours; (h) GroundTruth

与其他方法相比，在图6中的老鼠玩具和图7中的空调机箱字母、数字等细节部分，MSFNet展现出更高的清晰度，这表明暗区特征提取模块能够有效提取夜间图像暗区的细节纹理。同时，图6中的书本、挂饰和图7中的红色缎带增强后呈现出更优的视觉效果，验证了多分支网络模块在多尺度感知图像方面的有效性。该模块能够在图像增强过程中提取更丰富的特征信息，从而生成细节更清晰、质量更高的图像。

3.4 非配对数据集对比实验

为验证MSFNet的泛化能力，本实验将LOLv2-syn数据集上训练好的模型直接迁移至LIME、MEF和NPE三个未配对的数据集进行测试。选取了包括RetinexNet[33]、KinD[34]、ZeroDCE[14]、KinD++[35]、EnGAN[15]与SNRNet[18]在内的六种算法作为对照组，采用NIQE作为评价指标。从图8可知，MSFNet在三个非配对数据集上取得了最优的NIQE值，在LIME、MEF和NPE上的NIQE值分别低至3.74、3.91和3.36，这表明MSFNet增强后的图像在视觉自然度方面表现更优。此外，从图9的可视化结果可知，MSFNet生成的图像不仅亮度显著提升，细节也更加清晰，避免了色偏、伪影等常见问题。

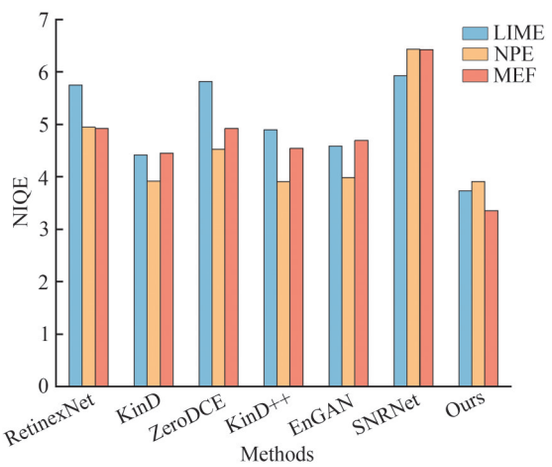


图8 客观评价指标NIQE对比

Fig. 8 Comparison of objective evaluation indicator NIQE

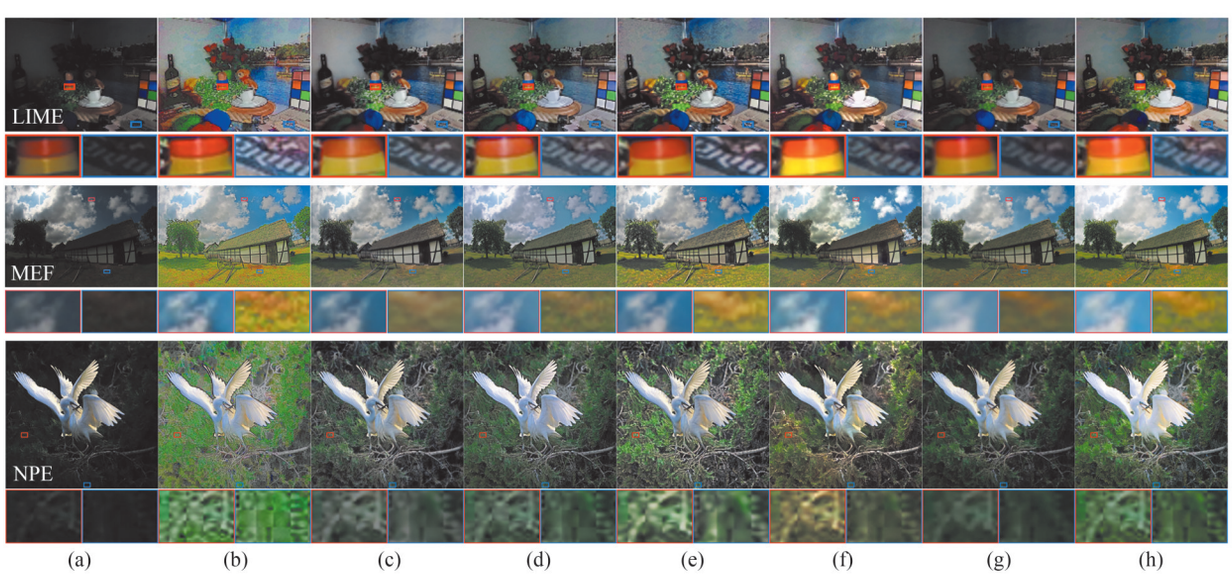


图9 LIME、MEF、NPE数据集上可视化对比。(a)输入；(b) RetinexNet；(c) KinD；(d) ZeroDCE；(e) KinD++；(f) EnGAN；(g) SNRNet；(h) Ours

Fig. 9 Visual comparison on the LIME, MEF, and NPE datasets. (a) Input; (b) RetinexNet; (c) KinD; (d) ZeroDCE; (e) KinD++; (f) EnGAN; (g) SNRNet; (h) Ours

3.5 夜间人脸检测对比实验

为评估本文方法在实际下游视觉任务中的应用效果，将MSFNet与多个先进算法在夜间人脸数据集DARKFACE上进行了对比实验。从DARKFACE数据集中随机抽取500张夜间人脸图像作为输入，选用在WIDERFACE[36]数据集上预训练的DSFD[37](Dual Shot Face Detector)模型作为人脸检测器。对不同方法增强后的图像进行统一输入，计算IoU阈值分别为0.5、0.6和0.7下的平均精度。

图10展示了夜间环境下人脸检测的可视化对比结果。对原始图像进行检测时，仅识别出1张人脸；而经过增强处理后，检测到的人脸数量明显增加。其他方法增强后的图像存在亮度不足和细节丢失等问题，导致人脸区域模糊。相比之下，MSFNet利用暗区特征提取模块，能够有效提取暗区人脸细节信息，降低噪声干扰。此外，多分支网络模块从多个尺度提取图像信息，进一步增强了模型的特征表达能力，使得增强后图像的人脸更加清晰，整体色彩更加自然。



图10 DARKFACE数据集上人脸检测可视化对比。(a)低光输入；(b) EnGAN；(c) RUAS；(d) SCI；(e) KinD；(f) Restormer；(g) UHDFour；(h) SNRNet；(i) Retinexformer；(j) Ours

Fig. 10 Visual comparison of face detection on the DARKFACE dataset. (a) Low-Input; (b) EnGAN; (c) RUAS; (d) SCI; (e) KinD; (f) Restormer; (g) UHDFour; (h) SNRNet; (i) Retinexformer; (j) Ours

从表2可以看出，MSFNet在不同IoU阈值下的平均精度均为最优；与Low-Input相比，经过MSFNet增强后的图像在不同的IoU下平均精度分别提升了6.9%、4.6%和0.7%。在夜间人脸识别和行人监控等任务中，由于光照不足，目标与背景往往难以区分，导致检测精度较低。MSFNet能够为此类下游任务生成区分度更高的增强图像，从而有效提升目标检测的准确率。

表2 人脸检测AP值对比

Table 2 Comparison of AP value for face detection

Methods	AP(%) under different IoUs		
	0.5	0.6	0.7
Low-input	19.5	6.1	0.7
EnGAN[15]	21.3	10.3	0.9
RUAS[29]	23.7	10.3	1.1
SCI[16]	23.6	9.9	0.8
KinD[34]	23.5	8.1	1.0
Restormer[31]	22.9	8.7	0.8
UHDFour[17]	23.6	9.2	1.1
SNRNet[18]	23.2	8.3	0.9
Retinexformer[32]	26.2	10.4	1.3
Ours	**26.4**	**10.7**	**1.4**

3.6 消融实验

3.6.1 不同模块的消融实验

为评估MSFNet中每个模块对整体性能的贡献，在LOLv1与LOLv2-syn两个数据集上进行消融实验。从表3中可以看出，仅使用DAFEM时，虽然PSNR和SSIM提升幅度有限，但LPIPS值由0.151下降至0.132，说明该模块能有效提取图像纹理细节。相比之下，加入MBNM后模型对图像重建的能力提升：在LOLv2-syn上，PSNR从23.15升至25.15，SSIM提升0.019。当两个模块协同工作时，性能达到最优：在LOLv1上PSNR为26.54，SSIM为0.857，分别提升3.70和0.070，LPIPS下降至0.129；在LOLv2-syn上PSNR为26.17，SSIM为0.937，分别提升3.02和0.023，LPIPS下降至0.073。DAFEM通过卷积提取图像特征生成通道注意力权重，增强与细节信息相关的通道特征，从而有效提取图像暗区纹理细节特征；MBNM提取包含全局结构、局部区域与细节纹理的多尺度信息，在增强过程中能感知丰富的多尺度信息，使得增强后的图像细节更清晰。

表3 消融实验结果

Table 3 Ablation experiment results

DAFEM	MBNM	LOLv1			LOLv2-syn		
		PSNR↑/dB	SSIM↑	LPIPS↓	PSNR↑/dB	SSIM↑	LPIPS↓
×	×	22.84	0.787	0.151	23.15	0.914	0.088
√	×	23.65	0.836	0.132	24.36	0.921	0.078
×	√	24.02	0.845	0.136	25.15	0.933	0.083
√	√	**26.53**	**0.857**	**0.129**	**26.17**	**0.937**	**0.073**

3.6.2 多分支子网络的消融实验

为评估MBNM内部各子网络在夜间图像增强任务中的具体贡献，本文针对子网络设计消融实验。包括单独使用SubNet1、SubNet2、SubNet3以及三者联合的SubNet1+2+3配置，在LOLv1、LOLv2-real和LOLv2-syn三个数据集上分别进行PSNR与SSIM指标评估。从图11中的数值结果可以看出，三个子网络协同提取特征后，图像增强性能呈现明显上升趋势。在PSNR方面，平均值从21.78逐步提升至25.41，增幅达3.63；SSIM则从0.828提升至0.867，增幅为0.039。尤其在LOLv2-syn数据集上，SubNet1+2+3的PSNR和SSIM分别达到26.17与0.937。从图12的图像增强结果可知，SubNet1通过提取全局特征，提升了整体亮度，但因忽略局部特征，导致花瓣脉络、墙面裂纹等细节信息模糊不清，存在明显的偏色问题。SubNet2对图像颜色还原更加自然，但存在边缘模糊、轮廓不清的问题。SubNet3则侧重于细节强化，清晰还原花瓣纹理和边缘结构。三者联合构建的SubNet1+2+3结构呈现出更自然的增强图像。

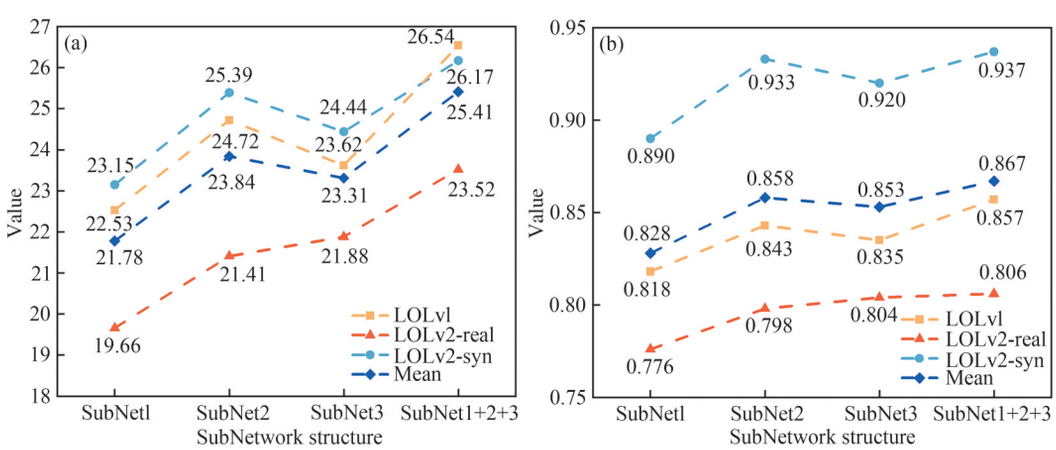


图11 MBNM中多分支子网络指标对比。(a) PSNR；(b) SSIM

Fig. 11 Comparison of multi-branch network indicators in MBNM. (a) PSNR; (b) SSIM



图12　MBNM中多分支子网络可视化对比。(a)输入；(b) SubNet1；(c) SubNet2；(d) SubNet3；(e) SubNet1+2+3；(f) GroundTruth

Fig. 12　Visual comparison of multi-branch networks in MBNM. (a) Input; (b) SubNet1; (c) SubNet2; (d) SubNet3; (e) SubNet1+2+3; (f) GroundTruth

3.6.3　网络宽度和深度的参数实验

网络宽度决定了编码器第一层和解码器最后一层的特征通道数。通过设置不同的特征通道数，在LOLv1数据集上评估网络宽度对模型性能的影响，实验结果如表4所示。随着网络宽度的增加，模型性能呈现上升趋势。宽度为16时，PSNR为24.67，SSIM为0.832，LPIPS为0.158；当宽度增加到48时，PSNR提高到26.53，SSIM提高到0.857，LPIPS降低到0.129。这表明，适度增加网络宽度可以有效增强模型的特征表达能力，从而改善图像增强效果。然而，过度增加网络宽度不仅会提高计算复杂度，还可能导致模型出现过拟合。综合权衡性能与计算开销，当网络宽度为48时，模型表现最佳。

表4　模型在不同网络宽度下的性能变化

Table 4　Model performance variation with different network widths

Width	PSNR↑/dB	SSIM↑	LPIPS↓	FLOPs/G	Params/M
16	24.67	0.832	0.158	**9.7**	**1.18**
24	25.12	0.841	0.145	12.3	1.52
32	25.89	0.849	0.138	14.7	1.78
48	**26.53**	**0.857**	**0.129**	18.2	2.14
64	26.41	0.855	0.132	23.1	2.67

通过设置不同层数的编码器与解码器，在LOLv1数据集上评估网络深度对模型性能的影响，实验结果如表5所示。随着网络深度的增加，模型性能呈现上升趋势。网络深度从2增加到4时，PSNR提升约9.8%，SSIM提升约5.2%，LPIPS下降约18.4%。但进一步增加网络深度后，性能出现下降，网络深度为

表5　模型在不同网络深度下的性能变化

Table 5　Model performance variation with different network depths

Encoder layers	Decoder layers	PSNR↑/dB	SSIM↑	LPIPS↓	FLOPs/G	Params/M
2	2	24.17	0.815	0.158	**8.4**	**0.92**
3	3	25.48	0.839	0.142	12.7	1.43
4	4	**26.53**	**0.857**	**0.129**	18.2	2.14
5	5	26.41	0.855	0.131	24.8	2.67
6	6	26.23	0.851	0.134	35.6	3.92

6时，模型参数量增长至3.92M，计算资源消耗明显增加。综合权衡性能与计算开销，当网络深度为4时，模型表现最佳。

4 结论

本文提出MSFNet，旨在解决现有方法在增强过程中常见的纹理细节丢失，噪声放大问题。首先，在主干网络前端引入暗区特征提取模块，通过卷积提取图像特征生成通道注意力权重，增强与细节信息相关的通道特征，从而有效提取暗区纹理细节。其次，设计多分支网络模块，通过三个子网络并行提取不同尺度下的图像特征，实现夜间图像全局结构和局部细节的恢复。最后，提出动态权重分配机制，自适应融合各子网络的多尺度特征，强化特征表达能力，进一步提升图像增强效果。实验结果表明，MSFNet在多个数据集上均优于现有方法；夜间人脸检测对比实验表明，MSFNet在夜间目标检测精度方面具有很大提升，验证其在下游视觉任务中的实用价值。

参考文献

[1] WANG Manli, DOU Zeya, CHEN Wei, et al. A text detection method based on multi-scale selective fusion feature pyramid and multi-semantic spatial network for visual IoT[J]. IEEE Internet of Things Journal, 2025, 12(11): 15997-16008.

[2] YOU Shuai, FENG Yujian, WANG Shitao, et al. Saliency feature learning for multi-modality person retrieval in visual internet of things[J]. IEEE Internet of Things Journal, 2025, 12(11): 15883-15896.

[3] LIU Fei, LU Zihao, LIN Xianke. Vision-based environmental perception for autonomous driving [J]. Journal of Automobile Engineering, 2025, 239(1): 39-69.

[4] ZHAO Jingyuan, ZHAO Wenyi, DENG Bo, et al. Autonomous driving system: a comprehensive survey[J]. Expert Systems with Applications, 2024, 242: 122836.

[5] AZDA R, AGHDAM E K, RAULAND A, et al. Medical image segmentation review: the success of u-net[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024, 46(12): 10076-10095.

[6] WANG Jiaji, WANG Shuihua, ZHANG Yudong. Deep learning on medical image analysis[J]. CAAI Transactions on Intelligence Technology, 2025, 10(1): 1-3.

[7] WANG Tianyu, REN Siyu, ZHANG Haiyan. Nighttime wildlife object detection based on YOLOv8-night[J]. Electronics Letters, 2024, 60(15): e13305.

[8] VIJAYAKUMAR A, VAIRAVASUNDARAM S. Yolo-based object detection models: a review and its applications[J]. Multimedia Tools and Applications, 2024, 83(35): 83535-83574.

[9] SUN Fuyan, HuaiLYU, ZongwangLYU. Review of low light image enhancement based on deep learning[J]. Application Research of Computers, 2025, 42(1): 19-27.
孙福艳, 吕淮, 吕宗旺. 基于深度学习的低光照图像增强研究综述[J]. 计算机应用研究, 2025, 42(1): 19-27.

[10] RAO B S. Dynamic histogram equalization for contrast enhancement for digital images[J]. Applied Soft Computing, 2020, 89: 106114.

[11] LI Xiangsheng, LIU Manlu, LING Qiang. Pixel-wise gamma correction mapping for low-light image enhancement[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2023, 34(2): 681-694.

[12] FENG Wei, LIU Hongdi, WU Guiming, et al. Gradient domain adaptive tone mapping algorithm based on color correction model[J]. Laser & Optoelectronics Progress, 2020, 57(8): 77-83.
冯维, 刘红帝, 吴贵铭, 等. 基于颜色校正模型的梯度域自适应色调映射算法[J]. 激光与光电子学进展, 2020, 57(8): 77-83.

[13] LI Xiaofang, WANG Weiwei, FENG Xiangchu, et al. Deep parametric Retinex decomposition model for low-light image enhancement[J]. Computer Vision and Image Understanding, 2024, 241: 103948.

[14] GUO Chunle, LI Chongyi, GUO Jichang, et al. Zero-reference deep curve estimation for low-light image enhancement [C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 1780-1789.

[15] JIANG Yifan, GONG Xinyu, LIU Ding, et al. Enlightengan: deep light enhancement without paired supervision[J]. IEEE Transactions on Image Processing, 2021, 30: 2340-2349.

[16] MA Long, MA Tengyu, LIU Risheng, et al. Toward fast, flexible, and robust low-light image enhancement[C]. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 5637-5646.

[17] LI Chongyi, GUO Chunle, ZHOU Man, et al. Embedding fourier for ultra-high-definition low-light image enhancement[J]. arXiv preprint arXiv:2302.11831, 2023.

[18] XU Xiaogang, WANG Ruixing, FU Chiwing, et al. Snr-aware low-light image enhancement[C]. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 17714-17724.

[19] LUO Yu, YOU Bijia, YUE Guanghui, et al. Pseudo-supervised low-light image enhancement with mutual learning[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2024, 34(1): 85-96.
[20] XIE Chao, FEI Linfeng, TAO Huanjie, et al. Residual quotient learning for zero-reference low-light image enhancement[J]. IEEE Transactions on Image Processing, 2025, 34: 365-378.
[21] FEIJOO D, BENITO J C, GARCIA A, et al. Darkir: robust low-light image restoration[C]. 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2025: 10879-10889.
[22] RONNEBERGER O, FISCHER P, BROX T. U-net: convolutional networks for biomedical image segmentation[C]. International Conference on Medical Image Computing and Computer-assisted Intervention, 2015: 234-241.
[23] LEE S. An efficient content-based image enhancement in the compressed domain using retinex theory[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2007, 17(2): 199-213.
[24] YANG Wenhan, WANG Wenjing, HUANG Haofeng, et al. Sparse gradient regularized deep retinex network for robust low-light image enhancement[J]. IEEE Transactions on Image Processing, 2021, 30: 2072-2086.
[25] YANG Wenhan, YUAN Ye, REN Wenqi, et al. Advancing image understanding in poor visibility environments: a collective benchmark study[J]. IEEE Transactions on Image Processing, 2020, 29: 5737-5752.
[26] GUO Xiaojie, LI Yu, LING Haibin. LIME: Low-light image enhancement via illumination map estimation[J]. IEEE Transactions on image processing, 2016, 26(2): 982-993.
[27] WANG Shuhang, ZHENG Jin, HU Haimiao, et al. Naturalness preserved enhancement algorithm for non-uniform illumination images[J]. IEEE transactions on image processing, 2013, 22(9): 3538-3548.
[28] MA Kede, ZENG Kai, WANG Zhou. Perceptual quality assessment for multi-exposure image fusion[J]. IEEE Transactions on Image Processing, 2015, 24(11): 3345-3356.
[29] LIU Risheng, MA Long, ZHANG Jiaao, et al. Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement[C]. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 10561-10570.
[30] WANG Zhendong, CUN Xiaodong, BAO Jianmin, et al. Uformer: a general u-shaped transformer for image restoration[C]. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 17683-17693.
[31] ZAMIR S W, ARORA A, KHAN S, et al. Restormer: efficient transformer for high-resolution image restoration[C]. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 5728-5739.
[32] CAI Yuanhao, BIAN Hao, LIN Jing, et al. Retinexformer: one-stage retinex-based transformer for low-light image enhancement[C]. 2023 IEEE/CVF International Conference on Computer Vision, 2023: 12504-12513.
[33] WEI Chen, WANG Wenjing, YANG Wenhan, et al. Deep retinex decomposition for low-light enhancement[C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018: 1-12.
[34] ZHANG Yonghua, ZHANG Jiawan, GUO Xiaojie. Kindling the darkness: a practical low-light image enhancer[C]. Proceedings of the 27th ACM international conference on multimedia, 2019: 1632-1640.
[35] ZHANG Yonghua, GUO Xiaojie, MA Jiayi, et al. Beyond brightening low-light images[J]. International Journal of Computer Vision, 2021, 129: 1013-1037.
[36] YANG Shuo, LUO Ping, ChenchangeLOY, et al. Wider face: a face detection benchmark[C]. 2016 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2016: 5525-5533.
[37] LI Jian, WANG Yabiao, WANG Changan, et al. DSFD: dual shot face detector[C]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 5060-5069.

Nighttime Image Enhancement Method Baesd on Multi-scale Fusion Network

WANG Manli[1], CHEN Chen[1], JIA Yunhong[2], SHI Yannan[1], QIAO Yingxu[3]
(1 *School of Physics and Electronic Information, Henan Polytechnic University, Jiaozuo* 454003, *China*)
(2 *National Key Laboratory of Digital-Intelligent Technology for Unmanned Coal Mining, Beijing* 100013, *China*)
(3 *School of Computer Science and Technology, Henan Polytechnic University, Jiaozuo* 454003, *China*)

Abstract: With the wide application of intelligent devices and the development of artificial intelligence technology, significant value has been demonstrated by nighttime image enhancement in multiple fields such as visual Internet of Things, autonomous driving, medical imaging, and object detection. Nighttime image enhancement is defined as the process of increasing the brightness, enhancing the contrast and restoring the details of images under insufficient lighting conditions. To address the issues of texture detail

loss and noise amplification in existing nighttime image enhancement methods, a nighttime image enhancement method based on a multi-scale fusion network (MSFNet) is proposed in this paper.

Firstly, in order to effectively extract the texture detail features of the dark area of the image, a dark area feature extraction module is introduced at the front end of the backbone network. The global features of the image are extracted through convolution to generate channel attention weights, enhancing the channel features related to the detail information. Secondly, multi-branch network modules are designed in the backbone network, and specific sub-networks are designed for different scales, so that rich multi-scale information can be perceived during the enhancement process and the loss of detail reconstruction at a single scale is avoided. Finally, a dynamic weight allocation mechanism is proposed in this paper, through which the multi-scale features of each sub-network are adaptively fused to enhance the feature expression ability and further improve the image enhancement effect.

On the three paired datasets of LOLv1, LOLv2-real, and LOLv2-syn, the PSNR values of MSFNet were higher than those of Retinexformer by 1.37, 0.72, and 0.50, respectively. The SSIM values on the LOLv1 and LOLv2-syn datasets were also higher than those of Retinexformer by 0.012 and 0.007, respectively. On the three unpaired datasets of LIME, MEF, and NPE, the optimal NIQE values were achieved by MSFNet, with values as low as 3.74, 3.91, and 3.36, respectively. The brightness of the enhanced images was significantly improved, and the details were rendered more clearly. Furthermore, on the publicly available nighttime face detection dataset DARKFACE, higher face detection accuracy was achieved for the enhanced images. The AP values under the three IoU thresholds of 0.5, 0.6, and 0.7 were recorded as 0.264, 0.107, and 0.014, respectively. More detail information was extracted from the images enhanced by MSFNet, thereby enhancing the discrimination between the target and the background and demonstrating the effectiveness of this method in downstream tasks.

MSFNet is proposed to address the the issues of texture detail loss and noise amplification commonly encountered during enhancement. First, a dark area feature extraction module is introduced at the front end of the backbone; channel attention weights are generated via convolutions on image features, and channel features related to fine details are emphasized, thereby enabling effective extraction of low-light texture details. Second, multi-branch network modules are designed, in which three sub-networks in parallel extract image features at different scales, allowing the recovery of global structure and local details in nighttime images. Finally, a dynamic weight allocation mechanism is proposed, by which multi-scale features from the sub-networks are adaptively fused; the feature representation capability is strengthened, and image enhancement performance is further improved. MSFNet is shown by the experimental results to outperform existing methods on multiple datasets. In comparative experiments on night face detection, a significant improvement in night target detection accuracy is achieved by MSFNet, thereby verifying its practical value in downstream visual tasks.

Key words: Nighttime image enhancement; Dark area feature extraction; Multi-branch network; Dynamic weight allocation; Face detection

OCIS Codes: 100.2000; 100.2960; 100.2980; 110.3000

CSTR: 32255.14.gzxb20265501.0110003

Foundation item: National Natural Science Foundation of China (52274160, 62472145), Doctoral Foundation of Henan Polytechnic University (B2021-64), Opening Project Foundation of Henan Province Engineering Technology Research Center for Photoelectric Detection and Sensing Integration(KFZ202502)