

低光照环境下的图像恢复与目标识别方法^{*}

任东东^{1,2}, 李金宝^{1,2}



¹(黑龙江大学 计算机科学技术学院, 黑龙江 哈尔滨 150080)

²(黑龙江省数据库与并行计算重点实验室(黑龙江大学), 黑龙江 哈尔滨 150080)

通讯作者: 李金宝, E-mail: jbli@hlju.edu.cn

摘要: 现有的低光照图像目标识别方法通常将图像恢复与目标识别任务分开处理,且在图像恢复质量和计算时间上无法满足目标识别任务的要求.针对这些问题,首先,提出一种高效的图像恢复卷积神经网络架构,通过融合不同尺度的特征图来聚合多级上下文特征信息,减少卷积层的信息冗余,提高图像恢复任务的实时性.并且设计了一个局部-全局注意力模块,通过校准各特征图的局部信息和特征通道之间的关系,提高恢复网络对噪声和图像内容的区分能力.其次,提出了一种图像恢复和目标识别任务协同处理的方法.利用目标识别的高级语义信息指导图像恢复网络学习,从而突出目标的结构和纹理等特征,使恢复结果更加适应目标识别任务.实验结果表明,该方法在图像恢复质量、计算时间和目标识别率上明显优于已有方法.

关键词: 深度学习;卷积神经网络;目标识别;低光照图像恢复;注意力机制

中文引用格式: 任东东,李金宝.低光照环境下的图像恢复与目标识别方法.软件学报, 2019,30(Suppl.(1)):94-104. <http://www.jos.org.cn/1000-9825/19010.htm>

英文引用格式: Ren DD, Li JB. Methods of image restoration and object detection in low-light environment. Ruan Jian Xue Bao/ Journal of Software, 2019,30(Suppl.(1)):94-104 (in Chinese). <http://www.jos.org.cn/1000-9825/19010.htm>

Methods of Image Restoration and Object Detection in Low-Light Environment

REN Dong-Dong^{1,2}, LI Jin-Bao^{1,2}

¹(School of Computer Science and Technology, Heilongjiang University, Harbin 150080, China)

²(Heilongjiang Provincial Key Laboratory for Database and Parallel Computing (Heilongjiang University), Harbin 150080, China)

Abstract: The existing object detection methods for low-light image usually separate image restoration from object detection tasks. In addition, the quality and computing time of image restoration cannot meet the requirements of object detection task. To solve these problems, firstly, this study proposes an efficient image restoration convolutional neural network architecture, which aggregates feature information of multi-level contexts by combining feature maps of different scales, reduces information redundancy of convolutional layers, and improves the real-time performance of image restoration. In addition, a local-global attention block is designed to improve the ability of the recovery network to distinguish between noise and image content by calibrating the local information of each feature map and the relationship between feature channels. Secondly, this study designs a solution for collaborative processing of image restoration and target recognition tasks. The high-level semantic information of target recognition is used to guide the image recovery network learning, so as to highlight the feature information such as the structure and texture of the target, and make the recovery result more suitable for the target recognition task. Experimental results show that this method is superior to the existing methods in image restoration quality, computing time and object detection rate.

Key words: deep learning; convolutional neural network; object detection; low-light image reatoration; attention mechanism

* 基金项目: 国家自然科学基金(No.61370222); 黑龙江省自然科学基金(ZD2019F003)

Foundation item: National Natural Science Foundation of China (No.61370222); Natural Science Foundation of Heilongjiang Province (ZD2019F003)

收稿时间: 2019-09-15; 采用时间: 2019-10-24

目标识别^[1,2]是计算机视觉中的基本任务之一.近年来,随着深度学习算法的不断改进^[3-5],以及数据集的不断增加,该领域得到了快速发展.然而,现有目标识别算法使用的数据集大都是在亮度适当、轮廓清晰和色彩均衡条件下收集的^[6-10],当在弱光环境,特别是极低光照环境下测试时,其识别率会大幅度降低.

低光照环境下成像主要面临低光子数和低信噪比问题,采集的图像对比度很低,且包含大量的噪声.虽然通过增加相机的感光度(ISO)可提高对比度,但相应的噪声信息也会被放大,并不能从根本上解决图像的低信噪比问题.噪声的存在不仅会影响图像的可视化效果,而且会影响后续的目标识别等高级视觉任务.

图像去噪旨在从噪声图像中恢复高质量图像信号.传统的图像去噪方法主要利用局部或非局部的图像特征^[11-13],但缺乏深层次的语义信息,限制了在复杂场景中的特征表示能力.近年来,基于卷积神经网络(CNN)的方法^[14-16]能够同时提取浅层和深层特征,在图像去噪任务上表现出优越的性能.但是,这些方法主要存在以下3个问题.首先,多数去噪网络采用以编码-解码器为基础的U型结构^[17-19],在处理高分辨率图像时计算成本会急剧增加,无法满足目标识别等高级视觉任务对实时性的要求.其次,这些网络多以单路径推理结果作为最终输出,忽略了各特征图之间的信息关系.第3,多数方法使用跳跃连接恢复编码器中丢失的图像细节.然而,跳跃连接也将浅层未去除的噪声信息带到了解码器部分,尤其在处理低信噪比图像时,无法有效分离噪声信息和图像内容,影响最终的图像恢复质量.

针对上述问题,本文提出了一种多尺度特征融合注意力网络 MSFFA(mutual-scale feature fusion attention network).如图1所示,MSFFA采用轻量级的骨干网络结构,通过融合单路卷积结构上不同尺度的特征图来聚合上下文信息,降低卷积层的信息冗余,减少计算参数.然而,浅层的特征图包含更多的噪声信息,融合后会影响整体网络的恢复性能.为了提高恢复网络对噪声和图像内容的区分能力,本文设计了一个局部-全局注意力模块 LGA-block(local global attention block).该模块有两个顺序的子模块:局部注意力模块(LA-block)和全局注意力模块(GA-block).如图2所示,LGA-block通过学习多尺度融合特征图的局部和全局信息,从分离空间和通道两个维度依次生成注意力图,然后将注意力图乘以输入要素图以进行自适应特征提取,进而优化各特征图的局部信息并重新构建通道之间的非线性关系,减弱噪声信息和图像内容在低信噪比图像上的纠缠.MSFFA不仅大幅度降低了网络的深度和复杂度,提高了图像恢复的实时性,而且其图像恢复质量明显优于目前已有的方法.

大多数基于深度学习的图像恢复方法^[18]通常只是最小化重建图像和高质量图像之间的误差,而重要的目标细节信息有时会丢失,这不但会影响图像恢复质量,而且降低了目标识别的性能.针对此问题,本文设计了一个新的低级与高级视觉任务级联的解决方案,在不改变现有数据集和目标识别算法的基础上,联合重建损失和预测损失来训练图像恢复网络,通过联合损失训练的模型不仅有利于恢复图像的感知质量,而且极大地提高了目标识别的准确性.

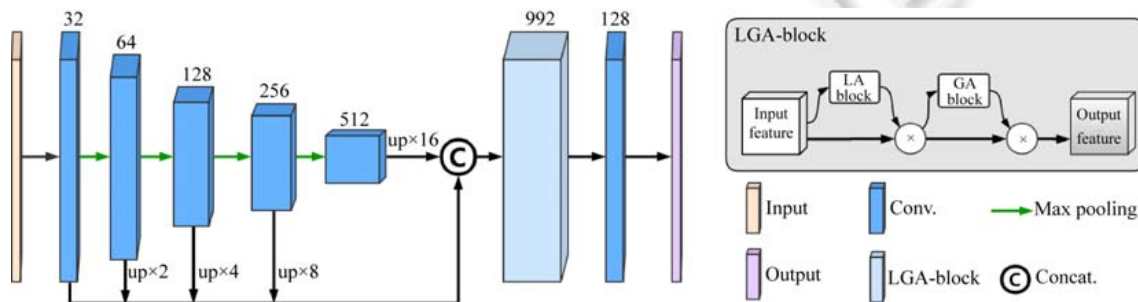


Fig.1 Architecture of multi-scale feature fusion attention network (MSFFA)

图1 多尺度特征融合注意力网络(MSFFA)整体框架结构

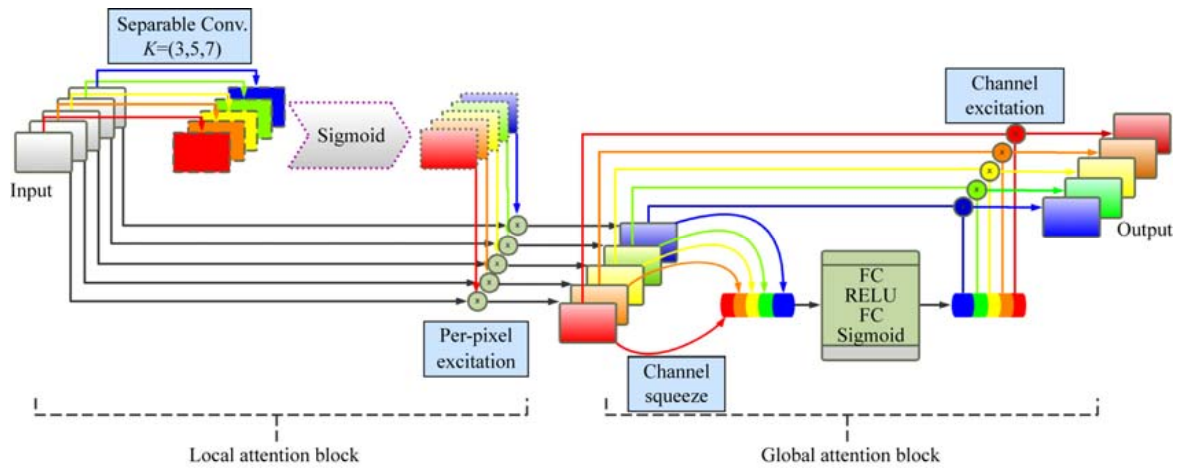


Fig.2 Local global attention block (LGA-block)

图2 局部全局注意力模块(LGA-block)

本文的主要贡献如下:

- 1) 本文提出了一个高效的图像恢复网络(MSFFA),通过融合不同卷积层的特征通道来聚合多尺度特征信息,提高信息的利用率,并使用注意力机制优化各特征图,重新构建融合层的通道间关系.与现有恢复效果最好的方法相比,MSFFA 可以实现快 4 倍以上的速度,并且具有更高的图像恢复质量.
- 2) 根据图像恢复任务的特点,本文设计了一个局部-全局注意力模块,通过学习多尺度融合特征图的局部和全局信息,从分离空间和通道两个维度生成注意力图,依次进行局部和全局信息优化,从而增强图像信号并抑制噪声等有害信息的输出,提高网络对噪声和图像内容的区分能力.
- 3) 本文通过级联图像恢复和目标检测任务,设计了一个新的训练策略,使用目标检测产生的软标签损失来指导图像恢复网络.实验结果表明,这种方式不仅获得了更高的图像恢复质量,而且更有利于目标识别任务的特征提取.

1 相关工作

1.1 图像恢复

图像恢复任务的目的是构建低质量(受损)图像与高质量图像之间的映射关系,其中特征提取和非线性表示是任务的关键.深度神经网络(DNN)以其强大的特征学习能力和函数逼近的理论性质成为这项任务的首选方法.

Jain 等人^[20]首先提出使用卷积神经网络(CNN)处理自然图像的去噪问题.Xie 等人^[21]利用多层全连接结构的自编码器进行图像去噪和图像修复.近几年,得益于深度学习算法的快速发展和硬件计算性能的提高,CNN 被广泛应用于图像恢复任务.Zhang 等人^[14]提出了一个端到端去噪神经网络(DnCNN),采用残差学习策略,在网络的隐层隐式的提取噪声图像.Mao 等人^[18]提出了一个深层的编码-解码网络框架(RED),用于图像去噪和超分辨率.Yan 等人^[22]在 U-Net 框架中引入一个特殊的移位连接层(shift-net),用于修复具有精细纹理的缺失区域.Tao 等人^[19]设计了一种尺度递归网络(SRNet)进行图像去模糊任务.

上述基于深度学习的方法已经在各种基准数据集上获得了较好的图像恢复结果,但是,高分辨率特征图在 U 型网络结构中需要较长的计算时间.还有一些方法通过限制输入图像大小^[23]或者简化网络中的冗余信道^[24]来降低计算的复杂度,从而提高运算速度,但这种方式很容易丢失高频细节信息,导致恢复图像过度平滑,并且深度较浅的网络还减弱了特征提取能力.本文采用轻量级的网络结构,通过融合不同尺度的卷积层来增强网络模型的特征利用率,减少冗余的卷积操作,提高网络的运算速度.

1.2 注意力机制

近年来,在诸如目标检测与分割等高级视觉方面,为了提高网络的特征表示能力,注意力机制受到人们的广泛关注.

Hu 等人^[25]提出了通道注意力模型 SE-block,使用全局平均池化对各通道的空间维度信息进行全局压缩,将二维的特征通道变成一个实数,以此来计算通道的注意力图.Woo 等人^[26]提出了一种卷积模块注意模型(CBAM)用于图像分类与检测,该模型在通道和空间两个维度上依次推断注意力图,然后将注意力图乘以输入要素图以进行特征校准.Kong 等人^[27]提出一种深度特征金字塔重新配置架构,该模型由全局关注和局部重构组成,以非线性方式将低级和高级语义特征融合进行目标检测.

上述方法的主要目的是通过注意力机制引导网络更加关注感兴趣的区域.然而,很少有人研究注意力机制对图像恢复任务的影响,特别是在多尺度融合的图层注意力上,几乎没有相关研究.而本文重点关注融合特征图各通道的局部噪声水平和通道之间信息的非线性关系,提高网络对噪声和图像内容的区分能力.

1.3 低级与高级视觉任务级联

目前的图像恢复等低级视觉任务多与高级视觉任务分开处理,在图像恢复过程中忽略了高级语义信息,影响后续高级视觉任务的准确率.

为了解决上述问题,Liu 等人^[28]首次提出了图像去噪与高级视觉任务级联的方案,通过联合损失来优化去噪网络,并使用去噪结果数据训练高级视觉任务.这种方法虽然在图像去噪和高级视觉任务上取得了不错的性能,但还有很大的局限性:1) 使用的训练数据是在 ILSVR2012 数据集上添加不同等级的高斯噪声合成的,然而在真实的成像系统中,图像噪声有多个来源(如暗电流噪声、短噪声、热噪声等),并且还会受到相机处理过程的影响,这与合成的高斯噪声图像差别很大;2) 高级视觉损失是由重建图像的预测值与原始标签之差产生的,而这种硬标签损失所产生的语义信息具有较低的泛化能力,并且在指导恢复网络训练时很容易产生训练过拟合或不收敛现象.

与上述方法不同,本文使用真实的低光照数据集^[29]训练和测试图像恢复网络,在固定的目标识别网络上分别使用重建图像和原始图像的预测值产生预测损失,通过联合预测损失和重建损失来训练图像恢复网络,突出目标的细节特征,更适用于目标识别任务.

2 本文方法

2.1 多尺度特征融合注意力网络(MSFFA)

为了克服目前低光照图像恢复实时性差、恢复质量低的缺点,本文提出了一种多尺度特征融合注意力网络——MSFFA,框架如图 1 所示.首先,MSFFA 网络通过对输入图像进行连续的卷积和下采样来获得不同尺寸的特征图,共进行了 5 次卷积(卷积核为 3×3)和 4 次下采样操作,卷积层的通道数量分别为 32,64,128,256,512.然后,将不同尺度的特征图上采样为相同尺寸,上采样因子分别为, $\times 1, \times 2, \times 4, \times 8, \times 16$,特征图以通道拼接(concat)的方式进行融合.最后,使用 Attention 机制重新校准融合特征图信息,并使用了一个通道输出为 128,卷积核大小为 3×3 的卷积操作进行局部信息重构.对于上采样方式,我们设计了两种方案:(1) 使用反卷积(deconvolution)来扩大特征图的尺寸.(2) 使用双线性差值法(bilinear)来扩大特征图的尺寸.经过实验,两种方案获得了相似的恢复性能,但与双线性插值法相比,反卷积需要学习大量的卷积核参数.为了降低网络的计算成本,采用双线性差值法来上采样低分辨率特征图.MSFFA 的恢复过程可以表示为

$$S = LGA(\text{cat}(L_1, L_2, L_3, L_4, L_5)) \quad (1)$$

其中, S 表示为重建图像, $LGA(\cdot)$ 代表 LGA-block 操作, $\text{cat}(\cdot)$ 代表 Concat 操作, L_i 为不同尺度卷积层的上采样特征图.

2.2 局部-全局注意力模块(LGA-block)

LGA-block 从给定的多尺度融合特征图 $F \in \mathbb{R}^{C \times W \times H}$ 作为输入依次推断出空间注意力图 $M_s \in \mathbb{R}^{C \times W \times H}$ 和通道注意力图 $M_c \in \mathbb{R}^{C \times 1 \times 1}$, 关注过程可归纳为

$$F^1 = M_s(F) \otimes F, F^2 = M_c(F^1) \otimes F^1 \tag{2}$$

其中, \otimes 表示对应元素点乘, 并相应地传递关注值: 即先沿空间维度传递关注值, 然后沿通道维度传递关注值, F^2 是最终的输出. 图 2 给出了各注意力图的详细结构.

局部注意力模块(LA-block). LA-block 利用特征图的空间信息关系生成分离空间注意力图. 分离空间注意力图主要关注各通道局部的信号和噪声水平. 为了计算空间注意力, 使用多尺度分离卷积生成各通道的局部特征描述符, 计算如下:

$$M_s(F) = \sigma(\text{sum}_{sp}(f_{sp}^{3 \times 3}(F), f_{sp}^{5 \times 5}(F), f_{sp}^{7 \times 7}(F))) \tag{3}$$

其中, σ 表示 Sigmoid 函数, $\text{sum}(\cdot)$ 为对应通道相加, $f(\cdot)$ 表示对卷积层 F 进行分离卷积操作, 卷积核大小分别为 $3 \times 3, 5 \times 5$ 和 7×7 .

全局注意力模块(GA-block). GA-block 主要分为通道压缩和特征激励两个阶段. 利用特征图的通道间关系来产生通道注意力图, 通道注意力主要关注各通道全局的信号和噪声水平. 首先, 为了有效汇集空间信息, 计算各通道的非线性关系, 压缩输入要素图的空间维度, 对每个通道实施全局平均池化操作, 具体计算如下:

$$Z_{Cr} = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W X_{Cr}(i, j) \tag{4}$$

其中, Cr 表示第 r 个通道, $X_{Cr}(i, j)$ 代表第 r 通道的第 i 行第 j 列元素.

激励阶段是由两个全连接层(FC)和一个 Relu 激活函数组成, 然后再由门控函数 Sigmoid 控制输出, 两次全连接采用中间层 16 倍降维, 即 FC(992, 62)-FC(62, 992). 激励过程可表示为

$$M_c(F) = \sigma(W_1 \delta(W_2 Z_{Cr})) \tag{5}$$

其中, $\delta(\cdot)$ 表示 Relu 激活函数, $\sigma(\cdot)$ 表示 Sigmoid 函数.

2.3 低级-高级视觉任务级联

本文设计了一个简洁高效的图像去噪与目标识别任务级联方案, 主要目的在于: 1) 以高级视觉为指导, 提高重建图像的感知质量; 2) 充分利用现有高级视觉的训练成果, 使恢复的图像更适用于目标识别任务. 如图 3 所示, 以给定的低光照图像作为输入, 首先使用恢复网络 MSFFA 进行图像重建, 然后将重建图像输入到下一个目标识别网络, 从而产生目标识别任务输出. 级联框架损失函数由重建损失和预测损失两部分组成.

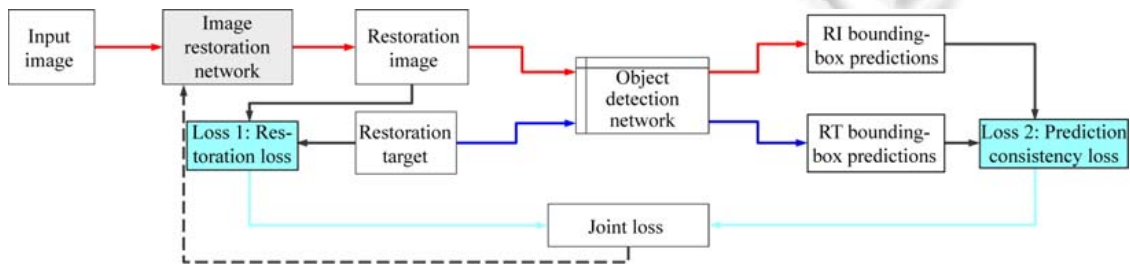


Fig.3 Overview of our proposed cascaded network

图 3 本文提出的级联网络方案整体结构图

重建损失是重建图像与高质量图像之间的平均绝对误差(MAE), 可以表示为

$$L_r(S, O) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W |F_r(S)_{i,j} - O_{i,j}| \tag{6}$$

其中, S 是输入低光照图像. O 是高质量图像. $F_r(\cdot)$ 代表图像恢复网络.

预测损失为重建图像 $F_r(S)$ 预测标签与高质量图像 O 预测标签之间的交叉熵损失,可以表示为

$$L_h(S, O) = H(\phi(F_r(S)), \phi(O)) \quad (7)$$

其中, $H(\cdot)$ 代表交叉熵损失函数, $\phi(\cdot)$ 代表目标检测网络.

联合损失定义为重建损失和预测损失之间的加权和,可以表示为

$$L(S, O) = \lambda_r L_r(S, O) + \lambda_h L_h(S, O) \quad (8)$$

其中 L 是联合损失, λ_r 和 λ_h 是用于平衡 L_r, L_h 损失的权重,在本文中分别为 10 和 1.

3 实验

在本节中,首先将介绍实验细节和评估方法,然后展示本文的方法在低光照图像恢复和目标识别任务上的性能.

3.1 实验设置

数据集:本文使用 SID^[29] 低光照数据集作为训练和测试数据. SID 图像数据由 Sony 和 Fuji 两种相机采集,共包含 5 094 张短曝光图像,每张短曝光图像都有相应的长曝光参考图像,且存在多个短曝光图像对应于相同的长曝光图像. 网络输入为亮度放大后的图像,放大比率设置为真实参照图像与相应的短曝光图像之间的曝光值之比. 在每次迭代中,随机剪裁一个 512×512 大小的补丁作为训练输入,并使用随机抖动和旋转进行数据增强.

实现细节:MSFFA 采用 ADAM^[30] 优化方法,参数 β_1 和 β_2 分别设置为 0.9 和 0.999,初始学习率为 $1e-4$, 2 000 轮迭代后学习率降为 $1e-5$,训练共进行了 3 000 轮. 实验环境为基于 Windows 10 操作系统的 Tensorflow 深度学习框架,使用 GPU 进行训练,配置 NVIDIA CUDA8.0+cuDNN6.0 深度学习库加速 GPU 运算,用于训练和测试的软件为 python 3.5,硬件配置为:Inter Core i7-7820X 处理器, Nvidia GeForce GTX 1080Ti GPU, Kingston DDR4 128GB 内存.

图像质量评价指标:现有低质量图像恢复的评价标准主要包括主观评价、客观量化和高级视觉任务检测. 主观评价即人眼观测输出图像,评定图像的质量. 目前最普遍的客观量化方法为峰值信噪比(peak signal to noise ratio,简称 PSNR),两幅图像之间的 PSNR 值(单位 dB)越高,则表明图像恢复质量越好. 高级视觉任务检测包括图像分类、目标识别和语义分割等具体任务来测试恢复图像的结果. 准确率越高,则证明恢复效果越好. 在本文中,分别使用以上 3 类方法对图像恢复结果进行了评价,并对比了图像恢复的实时性和目标检测的准确性.

3.2 实验结果

本文分别与 5 个典型的方法进行了比较:BM3D^[12]、DNCNN^[31]、EnDecoder(编码解码器网络)、CAN^[32]、Chen 等人^[29]、CBDNet^[16]. 由于 SID 为真实的低光照原始数据集,光照强度非常弱. 并且恢复过程包含盲噪声去除和色彩转换,而传统的图像去噪方法无法处理图像原始数据,所以只能转换为 sRGB 空间进行相应的处理. 基于深度学习的图像去噪和低光照增强算法虽然表现突出,但在处理极低信噪比图像时性能明显降低. 因此在定性实验结果中,我们只展示了目前在 SID 数据集上表现最好的方法:Chen 等人^[29]. 在定量实验结果中分别对比了不同方法的 PSNR、SSIM、参数量和恢复时间.

图像恢复定性结果:本文比较了不同方法在低光照图像上的恢复结果. 如图 4 所示,第 1 列是短曝光图像,第 2 列是短曝光图像进行数据放大后的结果;第 3 列和第 4 列分别是 Chen 等人^[29]方法和本文方法的恢复结果;最后一列是相应的长曝光图像. 与其他方法相比,本文方法不仅去除了噪声,而且消除了图中云彩、地面、壁画等处的伪影,以及椅子处的边界效应,恢复的结果图像更接近长曝光参考图像.

图像恢复定量结果:表 1 给出了不同方法在 SID 数据集上的定量恢复结果,分别对比了 PSNR、参数量和恢复时间 3 个指标,最好的结果字体加粗. 在 Sony 和 Fuji 两种类型的数据图像上,本文的方法取得了更好的图像恢复质量,比目前在 SID 数据集上恢复效果最好的方法 Chen 等人^[29],在 PSNR 指标上分别高出了 0.38db 和 0.89db. MSFFA 的网络参数量比编码解码器网络和 Chen 等人使用的 U-Net 网络减少了 3/4 左右. 图像恢复时间由 0.7s 缩短为 0.18s.

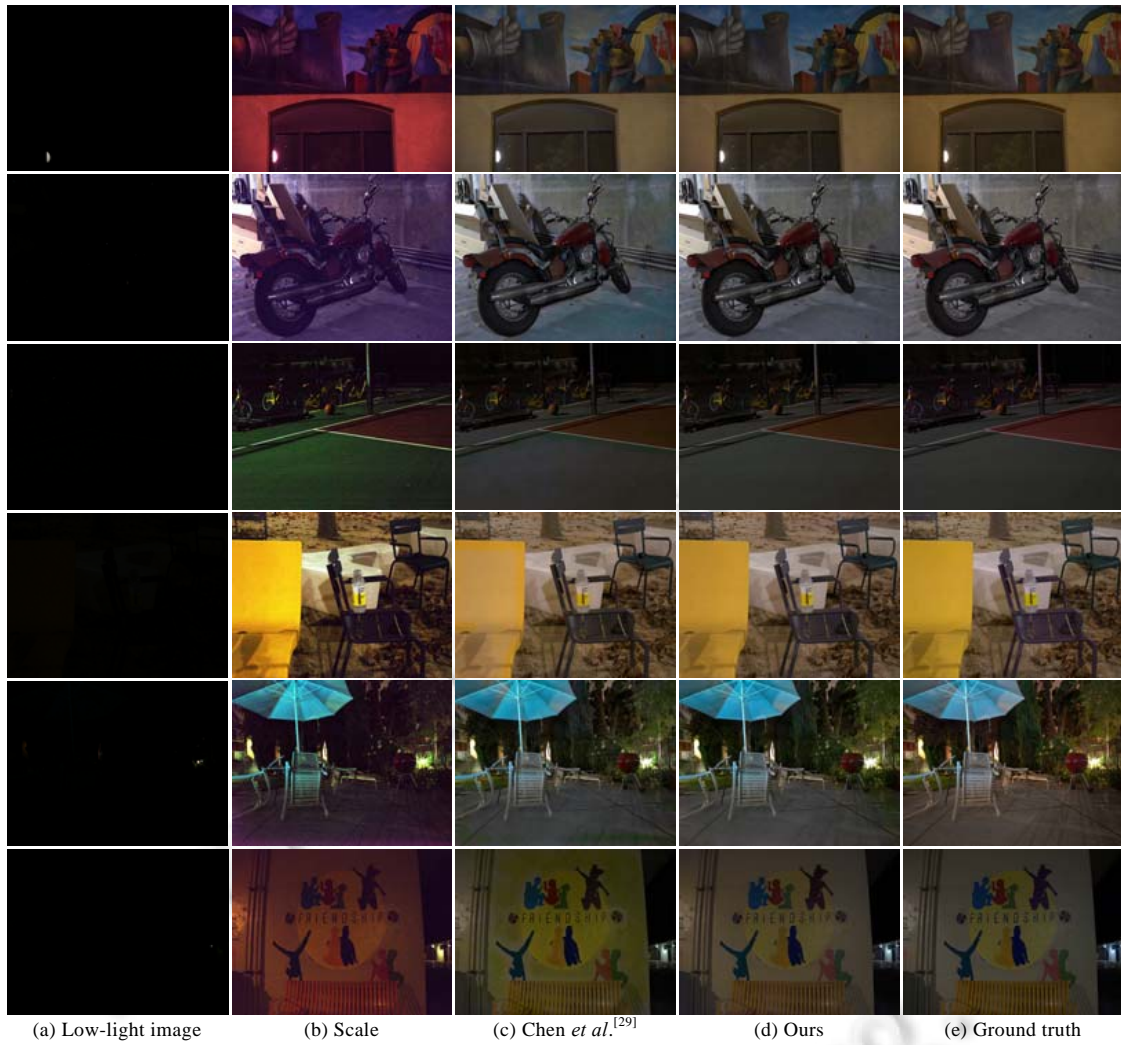


Fig.4 Low-light image restoration results of SID dataset by different methods
(The details in the results are better viewed magnified on a computer screen)

图 4 不同方法在 SID 低光照图像数据集上的恢复结果(细节信息请在计算机屏幕上查看)

Table 1 Recovery results of different methods. The evaluation indicators are average PSNR (db), parameters, and test time

表 1 不同方法的恢复结果,评价指标为平均峰值信噪比(db)、参数量和测试时间

Methods	SID/Sony	SID/Fuji	Parameters (MB)	Time (s)
BM3D	16.22	23.50	-	3.50
DNCNN	22.31	24.05	30	0.85
EnDecoder	27.99	26.20	80	0.70
CAN	27.40	25.71	36	0.38
CBDNet	26.35	25.21	91	0.85
Chen, <i>et al.</i>	28.88	26.61	89	0.78
Ours	29.26	27.50	21	0.18

目标识别结果:图 5 给出了不同的图像恢复方法在目标识别任务上的结果.采用目前表现性能较好的 YOLOv3^[33]作为目标识别方法.第 1 列为低光照噪声图像目标识别结果,第 2 列为 Chen 等人方法重建图像的目标

识别结果.第3列为本文方法的目标识别结果.第4列为长曝光参照图像的目标识别结果.从图中识别结果可知,图像中的色彩失衡和噪声严重影响目标识别结果,而本文的方法消除了图像的偏色和边界效应,利用高级语义信息更加突出目标的边缘和纹理特征,识别结果更接近正常曝光图像,并且在某些目标上具有更高的识别置信度.



Fig.5 Object detection results of low-light images restored by different methods on YOLOv3
(Please zoom in to the detailed recognition confidence)

图5 不同方法恢复的低光照图像在 YOLOv3 上的目标识别结果(详细的识别置信度请放大查看)

4 讨论

1) 卷积层数和通道数对网络性能的影响

深度卷积神经网络可以表示输入图像与目标图像之间的函数映射,随着网络深度的增加,卷积层提取的特征将更加抽象,感知区域将变得更大,更有利于网络提取全局特征,所以网络必须足够深才能有效地适应输入和输出之间的非线性关系.为此固定了采样次数即特征图尺度数量,从卷积层数和通道数角度分别做了两组对比实验.

首先,对比了每个尺度上不同数量的卷积层对图像恢复质量和参数数量的影响,层数范围为 1~3,在实验过程中 MSFFA 模块的结构保持不变,网络分别在 SID_Sony 和 SID_Fuji 两个数据集上进行了训练和测试.结果如表 2 所示,使用的卷积层数越多,网络性能越好,但参数量也大幅度增加.如当卷积层数由 1 层变为 2 层时,在 Sony 和 Fuji 两个子数据集上 PSNR 分别只增加了 0.03db 和 0.05db,但参数量却增加了将近 4 倍.为了平衡网络性能和时间以及空间成本,最终选择在每个尺度上使用 1 个卷积操作.

Table 2 Recovery results of different convolution layers number on various scales.
The evaluation indicators are average PSNR (db) and paramaters

表 2 各尺度上不同卷积层数的恢复结果.评价指标为峰值信噪比(db)和参数量

Layers number	SID/Sony	SID/Fuji	Parameters (MB)
1	29.26	27.50	21
2	29.29	27.55	57
3	29.30	27.62	93

其次,对比了卷积通道数量对图像恢复质量和参数量的影响,在保持 MSFFA 结构不变的基础上,训练和测试了卷积层上不同特征通道数对图像恢复结果的影响,结果见表 3.我们借鉴了 VGG^[34]等流行基础框架的经验,每次降采样操作都会将特征通道的数量增加 1 倍,因为我们的目标是在下采样过程中提取更具代表性的深度特征,这样可以有效地减少信息的丢失.同样,为了平衡网络性能和时间以及空间成本,最终将网络各尺度上的特征通道数选择为 32,64,128,256,512.

Table 3 Recovery results of different channel numbers on various scales.

The evaluation indicators are average PSNR (db) and paramaters

表 3 各尺度上不同通道数的恢复结果.评价指标为峰值信噪比(db)和参数量

Channels number	SID/Sony	SID/Fuji	Parmenters (MB)
(128,128,128,128,128)	28.05	27.04	9
(32,64,128,256,512)	29.26	27.50	21
(64,128,256,512,1024)	29.29	27.55	81

2) 下采样深度对网络性能的影响

下采样操作可将特征图缩小到一定尺寸,不仅可以增加感受野、有效去除噪声信息,还节省了训练/测试时间和储存空间.对比了不同的下采样深度对图像恢复质量和参数量的影响.如表 4 所示,当下采样深度由 4 层变为 5 层时,PSNR 分别增加了 0.76db 和 0.56db,参数量只增加了 20%左右.但当下采样深度由 5 层加深为 6 层时,PSNR 分别增加 0.04db 和 0.06db,参数量却增加了将近 4 倍.通过权衡网络性能和参数量,最终选用下采样深度为 5.

Table 4 Recovery results of different downsampling depths. The evaluation indicators are average PSNR (db) and paramaters

表 4 不同下采样深度的恢复结果.评价指标为峰值信噪比(db)和参数量

Down_Deep	SID/Sony	SID/Fuji	Parmenters (MB)
4	28.50	26.94	16MB
5	29.26	27.50	21MB
6	29.30	27.56	86MB

3) 不同注意力模块对网络性能的影响

近年来,注意力机制在深度学习中得到了广泛的应用,然而目前的注意力模块在生成注意力图时仅使用简单的全局信息,或在多通道注意力图时使用共享空间注意力图,这在处理不均匀噪声时限制了其性能.本文设计的 LGA-block 从空间局部和全局通道两个维度依次生成注意力图,提高了网络对融合特征图信息约束的敏感性.如表 5 所示,我们比较了 SE-block^[25]、CBAM^[26]和本文提出的 LGA-block 在低光照图像恢复上的性能.结果表明,LGA-block 在平均 PSNR 上优于其他两种方法.虽然 LGA-block 具有更多的参数量,但相对于整个网络占比很小.

Table 5 Recover results of different attention blocks. The evaluation indicators are average PSNR (db) and paramaters

表 5 不同注意力模块的恢复结果.评价指标为峰值信噪比(db)和参数量

Attention block	SID/Sony	SID/Fuji	Parmenters (MB)
SE-block	28.96	26.99	18
CBAM	29.14	27.20	20
LGA-block	29.26	27.50	21

5 结 论

本文提出了一个高效的图像恢复网络——MSFFA,通过多尺度特征融合和局部-全局注意力机制,既减少了参数量,提高了网络的运行速度,又有效地分离了噪声和图像内容.并且设计了一种新的低级视觉与高级视觉任

务协同处理的解决方案,有效地提升了低质量恢复图像的目标识别性能.实验结果表明,恢复的低光照图像,可以生成更清晰、更精细、在视觉上更可信的结果,并且具有更高的目标识别率和置信度.此外,本文提出的方法还可以推广到不同类型的低质量图像目标识别任务,如模糊图像目标识别、遮挡图像目标识别等.

References:

- [1] Ren SQ, He KM, Girshick R, Sun J. Faster R-CNN: Towards real-time object detection with region proposal networks. In: Proc. of the Annual Conf. on Neural Information Processing Systems. 2015. 91–99.
- [2] He KM, Zhang XY, Ren SQ, Sun J. Deep residual learning for image recognition. In: Proc. of the 2016 IEEE Conf. on Computer Vision and Pattern Recognition. 2016. 770–778.
- [3] Huang G, Liu Z, van der Maaten L, Weinberger KQ. Densely connected convolutional networks. In: Proc. of the 2017 IEEE Conf. on Computer Vision and Pattern Recognition. 2017. 2261–2269.
- [4] He KM, Gkioxari G, Dollár P, Girshick R. Mask R-CNN. In: Proc. of the 2017 IEEE Int'l Conf. on Computer Vision. 2017. 2980–2988.
- [5] Zhao HS, Shi JP, Qi XJ, Wang XG, Jia JY. Pyramid scene parsing network. In: Proc. of the 2017 IEEE Conf. on Computer Vision and Pattern Recognition. 2017. 6230–6239.
- [6] Neuhold G, Ollmann T, BulòSR, Kotschieder P. The mapillary vistas dataset for semantic understanding of street scenes. In: Proc. of the Int'l Conf. on Computer Vision. 2017. 5000–5009.
- [7] Cordts M, Omran M, Ramos S, Rehfeld T, Enzweiler M, Benenson R, Franke U, Roth S, Schiele B. The cityscapes dataset for semantic urban scene understanding. In: Proc. of the Conf. on Computer Vision and Pattern Recognition. 2016. 3213–3223.
- [8] Huang XY, Cheng XJ, Geng QC, Cao BB, Zhou DF, Wang P, Lin YQ, Yang RG. The ApolloScape dataset for autonomous driving. In: Proc. of the 2018 IEEE/CVF Conf. on Computer Vision and Pattern Recognition Workshops. 2018. 954–960.
- [9] Lin TY, Maire M, Belongie S, Bourdev L, Girshick R, Hays J, Perona P, Ramanan D, Zitnick CL, Dollár P. Microsoft COCO: Common objects in context. In: Proc. of the European Conf. on Computer Vision (ECCV). 2014. 740–755.
- [10] Deng J, Dong W, Socher R, Li LJ, Li K, Li FF. ImageNet: A large-scale hierarchical image database. In: Proc. of the Conf. on Computer Vision and Pattern Recognition. 2009. 248–255.
- [11] Gu SH, Zhang L, Zuo WM, Feng XC. Weighted nuclear norm minimization with application to image denoising. In: Proc. of the 2014 IEEE Conf. on Computer Vision and Pattern Recognition. 2014. 2862–2869.
- [12] Dabov K, Foi A, Katkovnik V, Egiazarian K. Image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE Trans. on Image Processing*, 2007,16(8):2080–2095.
- [13] Mairal J, Bach F, Ponce J, Sapiro G, Zisserman A. Non-local sparse models for image restoration. In: Proc. of the 2009 IEEE 12th Int'l Conf. on Computer Vision. 2009. 2272–2279.
- [14] Zhang K, Zuo WM, Chen YJ, Meng DY, Zhang L. Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising. *IEEE Trans. on Image Processing*, 2017,26(7):3142–3155.
- [15] Zhang K, Zuo WM, Zhang L. FFDNet: Toward a fast and flexible solution for CNN-based image denoising. *IEEE Trans. on Image Processing*, 2018,27(9):4608–4622.
- [16] Guo S, Yan ZF, Zhang K, Zuo WM, Zhang L. Toward convolutional blind denoising of real photographs. In: Proc. of the Conf. on Computer Vision and Pattern Recognition. 2019. 1712–1722.
- [17] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. In: *Medical Image Computing and Computer-Assisted Intervention*. 2015. 234–241.
- [18] Mao XJ, Shen CH, Yang YB. Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections. In: Proc. of the Annual Conf. on Neural Information Processing Systems. 2016. 2802–2810.
- [19] Tao X, Gao HY, Shen XY, Wang J, Jia JY. Scale-recurrent network for deep image deblurring. In: Proc. of the Conf. on Computer Vision and Pattern Recognition. 2018. 8174–8182.
- [20] Jain V, Seung HS. Natural image denoising with convolutional networks. In: Proc. of the Annual Conf. on Neural Information Processing Systems. 2008. 769–776.

- [21] Xie JY, Xu LL, Chen EH. Image denoising and inpainting with deep neural networks. In: Proc. of the Annual Conf. on Neural Information Processing Systems. 2012. 350–358.
- [22] Yan ZY, Li XM, Li M, Zuo WM, Shan SG. Shift-net: Image inpainting via deep feature rearrangement. In: Proc. of the European Conf. on Computer Vision (ECCV). 2018. 3–19.
- [23] Wu ZF, Shen CH, Anton van den Hengel. Real-time semantic image segmentation via spatial sparsity. Computing Research Repository (CoRR), 2017,abs/1712.00213.
- [24] Chollet F. Xception: Deep learning with depthwise separable convolutions. In: Proc. of the Conf. on Computer Vision and Pattern Recognition. 2017. 1800–1807.
- [25] Hu J, Shen L, Sun G. Squeeze-and-excitation networks. In: Proc. of the Conf. on Computer Vision and Pattern Recognition. 2018. 7132–7141.
- [26] Woo S, Park J, Lee JY, Kweon IS. CBAM: Convolutional block attention module. In: Proc. of the European Conf. on Computer Vision (ECCV). 2018. 3–19.
- [27] Kong T, Sun FC, Huang WB, Liu HP. Deep feature pyramid reconfiguration for object detection. In: Proc. of the European Conf. on Computer Vision (ECCV). 2018. 172–188.
- [28] Liu D, Wen BH, Jiao JB, Liu XM, Wang ZY, Huang TS. Connecting image denoising and high-level vision tasks via deep learning. Computing Research Repository (CoRR), 2018,abs/1809.01826.
- [29] Chen C, Chen QF, Xu J, Koltun V. Learning to see in the dark. In: Proc. of the Conf. on Computer Vision and Pattern Recognition. 2018. 3291–3300.
- [30] Kingma DP, Ba JL. Adam: A method for stochastic optimization. In: Proc. of the Int'l Conf. on Learning Representations. 2015.
- [31] Zhang K, Zuo WM, Chen Yj, Meng DY, Zhang L. Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising. IEEE Trans. on Image Processing, 2017,26(7):3142–3155.
- [32] Chen QF, Xu J, Koltun V. Fast image processing with fully-convolutional networks. In: Proc. of the Int'l Conf. on Computer Vision. 2017. 2516–2525.
- [33] Redmon J, Farhadi A. YOLOv3: An incremental improvement. Computing Research Repository (CoRR), 2018,abs/1804.02767.
- [34] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. In: Proc. of the Int'l Conf. on Learning Representations. 2015.



任东东(1992—),男,河北高邑人,硕士生,主要研究领域为深度学习,计算机视觉.



李金宝(1969—),男,博士,教授,博士生导师,CCF 高级会员,主要研究领域为计算机视觉,传感器网络.