

引用格式: WANG Xia, ZHANG Xin, JIAO Gangcheng, et al. Dual Residual Attention Network for ICMOS Sensing Image[J]. Acta Photonica Sinica, 2022, 51(6):0610002
王霞,张鑫,焦岗成,等. 基于双残差注意力网络的 ICMOS 图像去噪算法[J]. 光子学报, 2022, 51(6):0610002

基于双残差注意力网络的 ICMOS 图像去噪算法

王霞^{1,2}, 张鑫², 焦岗成¹, 杨晔¹, 程宏昌¹, 延波¹

(1 微光夜视技术重点实验室, 西安 710065)

(2 北京理工大学 光电学院 光电成像技术与系统教育部重点实验室, 北京 100081)

摘 要:针对微光夜视条件下增强型 CMOS(ICMOS)图像信噪比低、随机噪声明显的问题,提出了一种基于双残差注意力网络的 ICMOS 图像去噪算法。为了制作 ICMOS 噪声图像数据集,采用 ICMOS 相机在特定照度环境下拍摄静态噪声图像序列,然后采用多帧平均的方法获得每个序列对应的无噪声真值图像;其次,为了直接从噪声图像中提取噪声分量,设计了一种结合噪声残差学习和残差网络模块的双残差网络模型,并引入通道注意力机制给模型的特征图维度赋予权重,在提升模型学习能力的同时降低了模型复杂度;最后,采用网络训练所得模型对测试图像进行去噪实验。对比实验结果表明,本文提出的算法得到的峰值信噪比较经典的 BM3D 算法提升了 9.56 dB,结构相似度提升了 0.050 3。从主观效果可以看出,本文算法可以更好地去除 ICMOS 图像噪声,保留图像细节,同时,具有较高的运行效率。

关键词:微光夜视;ICMOS 图像;图像去噪;残差学习;注意力机制

中图分类号:TP391.4

文献标识码:A

doi:10.3788/gzxb20225106.0610002

0 引言

微光夜视技术是为了研究在较低光照的条件下,将所采集到的图像进行增强,传输,储存以及复现和应用的 光电技术,是近代光电子技术的重要组成部分。由于人眼的固有特性,当环境的光照度比较低的时候,人眼只能观察到物体的轮廓,对细节特征无法精确识别。增强型 CCD/CMOS(Intensified CCD/CMOS, ICCD/ICMOS)是现有的应用广泛、工作照度很低的固体微光成像器件。ICCD/ICMOS 是由像增强器和 CCD/CMOS 耦合而成。

虽然 ICMOS 可以在微光夜视条件下成像,但像增强器在增强信号的同时也放大了噪声的强度,导致获取到的图像随机噪声明显,并且噪声特性相较于传统 CMOS 成像的噪声更为复杂。由于微通道板的引入,ICMOS 图像噪声并不是独立同分布的,而是具有空间相关性的聚合状随机噪声^[1]。聚合状噪声破坏了图像原有的结构特征,这极大地增加了去噪的难度。

图像去噪是底层计算机视觉领域研究的热点问题。现有的去噪算法可以分为以下几大类:空域去噪算法、变换域去噪算法、基于稀疏表示的去噪算法、基于深度学习的去噪算法。空域去噪算法主要针对自然图像噪声在空间独立同分布的特性,采用滤波的方式去除噪声,如均值滤波、双边滤波^[2]、非局部均值(Non-Local Median, NLM)^[3]等,但这些方法并不适用于 ICMOS 图像中空间聚合状的随机噪声。变换域去噪算法首先对噪声图像进行特定的变换,然后根据变换域的特征以及噪声的性质,在变换域中对变换系数进行处理,去除噪声分量,保留信号分量,如傅里叶变换、小波变换^[4]、块匹配 3D 滤波(Block-Matching and 3D filtering, BM3D)^[5]等,其中 BM3D 是公认的效果最好的传统去噪算法。这类方法通常认为噪声为高频信息,图像信号为低频信息,但空间聚合噪声不仅仅包括高频信息,也存在低频分量,所以变换域方法很难对

基金项目:国家重点研发计划(No. 2019QY0902),国防科技重点实验室基金研究项目(No. J20190101)

第一作者:王霞(1972—),女,副教授,博士,主要研究方向为微光与超宽波段成像。Email: angelniuniu@bit.edu.cn

收稿日期:2022-01-06;录用日期:2022-02-15

<http://www.photon.ac.cn>

该噪声图像进行分离。稀疏表示的方法是将噪声图像通过某个过完备原子库进行稀疏表示,用若干个大信号来表示原始信号,利用稀疏性来将图像和噪声分离开,如奇异值分解^[6]、可学习同步稀疏编码(Learned Simultaneous Sparse Coding, LSSC)^[7]、非局部集中稀疏表示(Nonlocally Centralized Sparse Representation, NCSR)^[8]等。但这些方法在生成字典库时计算量很大,效率很低。近年来,深度学习方法已经广泛应用于图像去噪领域。深度学习方法通过从大量的噪声数据样本中学习噪声的分布特征,从而将噪声和图像信号分离。2012年,BURGER H C等^[9]发现简单的多层感知机就可以媲美BM3D算法;ZHANG Kai等^[10]在2016年提出了一个端到端的去噪卷积神经网络(Denoising Convolutional Neural Networks, DnCNN),该网络引入了残差学习,从输入图像估计对应的噪声分布。随后,在DnCNN的基础上,又提出了快速灵活的去噪网络(Fast and Flexible Denoising Convolutional Neural Network, FFDNet)^[11],可以实现对多个噪声模型同时去噪。TAI Ying等^[12]提出了一个深层记忆网络(Memory Network, MemNet),但网络结构比较复杂,难以训练。CHENG Shen等^[13]通过图像自适应映射的方法实现了去噪效果(Noise Basis Network, NBNNet)。YUE Zongsheng等^[14]提出了新的变分推理的方法(Variational Denoising Network, VDN),并把生成对抗网络引入去噪模型^[15]。ZAMIR S W等^[16]将多尺度注意力模型引入了图像去噪领域。GUO Shi等^[17]针对真实场景噪声进行仿真,并实现了真实噪声数据的盲去噪处理(Convolutional Blind Denoising, CBDnet)。

至今为止,现有的噪声图像数据集大都是基于仿真的高斯噪声和泊松噪声模型,也有一些真实场景下的可见光图像噪声数据,并没有基于ICMOS成像设备的图像数据集;其次,对ICMOS图像去噪的研究也很少,目前并没有基于深度学习的ICMOS图像去噪算法的相关研究。因此,本文首先基于实验室的直耦型ICMOS微光相机^[18]在不同照度下采集了几组暗室场景视频序列制作了ICMOS图像数据集,由于监督学习对数据的需求,训练样本中需要与噪声图像相对应的无噪声真值图像作为标签,而真实场景中很难直接获取到无噪声的标签图像,因此,本文采用了多帧平均的方法:对特定照度下的一个静态场景获取大量噪声图像序列并做多帧平均,就可以得到对应场景的无噪标签图像。其次,本文结合ICMOS的噪声分布特征设计了一个针对ICMOS图像的双残差注意力网络模型,其中,网络采用噪声残差学习策略以及残差网络模块提升网络对图像中的噪声特征提取的性能,同时,引入通道注意力机制进一步优化模型参数。实验表明,本文算法能在 $10^{-1} \sim 10^{-3}$ lx照度条件下较好地去除ICMOS图像中的噪声,并还原图像细节。

1 基于ICMOS的去噪算法模型

微光ICMOS图像是在微光场景中经过多次光电转换以及能量倍增之后获取的图像,与正常照度下获取的自然图像有较大差异性,信噪比较低,随机噪声明显,极大降低了图像整体的观感,不利于人眼的观察和识别。因此,对ICMOS图像进行噪声去除,提高图像信噪比,有利于进一步提升ICMOS成像的性能,拓宽其应用领域。

1.1 噪声原理分析

ICMOS成像的原理相较于普通的CMOS成像更加复杂,像增强器由光阴极、微通道板和荧光屏三部分组成^[19],如图1所示。首先,镜头接收场景中的光子经过光阴极光电转换后将接收到的微弱的光信号转换成电子图像,生成的电子随后被注入到微通道板中使得电子图像聚焦成像并获得能量倍增,其次,所有电子被投射到荧光屏上转换成可见的光学图像,最后由CMOS传感器捕捉荧光屏上的光信号生成图像。

由于ICMOS结构的复杂性,使得ICMOS噪声与自然图像噪声分布截然不同,其噪声模型由四个部分的噪声叠加而成:光阴极噪声、MCP噪声、荧光屏噪声以及最后的CMOS噪声。WANG Fei等^[1]对ICMOS的噪声进行了分析:1)ICMOS噪声呈现空间聚合状,与自然图像独立同分布的噪声不同,ICMOS图像噪声由于微通道板的引入具有一定的空间相关性,每一个像素点的强度都会受到周围像素点的干扰,这种噪声破坏了图像本身的结构特征,同时引入了大量无效特征;2)ICMOS噪声具有很强的随机性,低照度条件下,成像器件的颗粒噪声非常明显,整个ICMOS图像呈现闪烁噪声,与传统固定模式噪声不同,这些颗粒噪声随机性很强,大大增加了去噪的难度。

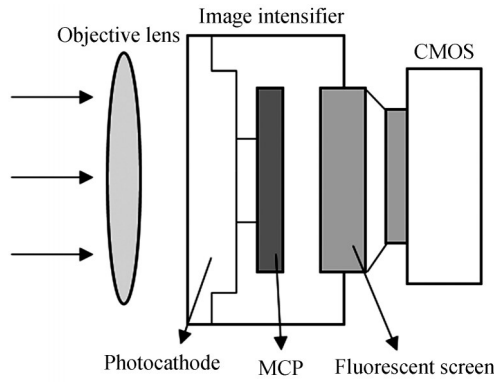


图1 ICMOS原理图
Fig.1 The pipeline of ICMOS

假设 y 表示获取到的 ICMOS 图像, x 表示与 y 对应的干净无噪声图像, n 代表图像噪声, F 代表 ICMOS 成像退化函数。那么 ICMOS 成像模型可以表示为

$$y = F(x) = x + n \quad (1)$$

因此,针对 ICMOS 图像去噪的模型可以表示为

$$x = F^{-1}(y) = y - n \quad (2)$$

式中, F^{-1} 为退化逆函数,即去噪模型。同时,从式(2)可以看出只要从图像中提取出噪声分量就可以通过残差学习策略计算出干净的无噪声图像。

1.2 双残差注意力网络模型

现有的深度学习去噪算法如 DnCNN、CBDnet 等,大都针对仿真的高斯-泊松分布噪声以及一些自然图像的真实噪声数据进行网络设计和模型训练,这些算法并不适合直接运用于 ICMOS 图像。本文针对 ICMOS 图像空间聚合状随机噪声设计了一个双残差注意力网络模型,如图 2 所示。

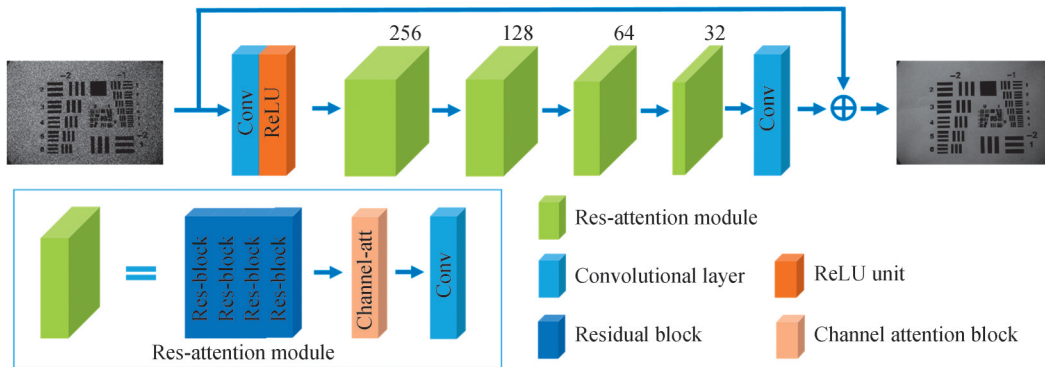


图2 算法结构图
Fig.2 The overview of the method

这里简单介绍一下网络结构,从图 2 可以看出,网络结构可以分为三个部分,首先第一部分为输入部分的卷积层和 ReLU 激活单元,这一部分负责将输入的三通道或者单通道图像转换到更高维度的特征空间从而为之后的特征提取做铺垫;第二部分是 4 个残差注意力模块,并且每一个模块的特征图的数量在不断减少,从残差注意力模块的组成可以看出,该模块由 4 个残差单元、1 个通道注意力层以及 1 个卷积层组成,每一个单元的具体结构会在后面介绍,这里引入了通道注意力机制,可以赋予中间层输出特征图不同的权重,自动获取到每个特征通道的重要程度,然后依照这个重要程度去提升有用的特征并抑制对当前任务用处不大的特征,并引导网络不断缩减特征图维度。最后一部分为输出部分的卷积层,该部分负责将残差注意力模块输出的高维度的代表噪声分量的特征图转换成与输入维度相同的噪声图像。由于 ICMOS 图像噪声具有一定的空间相关性,而下采样会导致空间信息的丢失。为了最大程度地保真图像本身的结构特征,整个

网络模型并不会改变中间特征层的尺度,保证了空间分辨率的一致性。网络所采用的卷积层都是 3×3 大小的卷积核,中间特征层的维度从前至后分别为256、128、64、32维。

传统的去噪算法模型大都是直接通过网络学习噪声图像 y 到真值图像 x 的端到端映射,也就是网络直接输出去噪后的图像。而本文提出的双残差注意力网络的任务是输入ICMOS图像 y ,输出图像中的噪声 n ,三者的关系如式(3)所示。

$$x = y + \varphi(y) = y + (-n) \quad (3)$$

式中, y 表示输入的ICMOS图像, x 表示与 y 对应的干净无噪声图像,即标签真值图像, n 表示图像噪声。算法模型采用噪声残差学习的策略,使得网络只需要着力于提取图像中的噪声分量,而不用复原原始图像的细节以及边缘信息,这样可以最大程度地减小网络的训练难度,同时在一定参数条件下提升网络的性能。最后,将网络训练的输出 $\varphi(y)$ 和输入 y 两者相加即可得到去噪后的图像 x ,此为双残差模型中的第一层残差。

为了进一步提高网络模型的性能,本文的主干网络中引入了大量的残差模块^[20],如图3所示。残差模块的映射函数表示为

$$F_{\text{res}}(s) = \text{ReLU}(f(s) + s) \quad (4)$$

式中, s 代表残差模块输入的特征图, $F_{\text{res}}(s)$ 表示整个残差模块的映射函数,ReLU表示非线性激活单元, $f(s)$ 表示中间层输出的结果。残差模块通过简单的跳跃连接实现,不会额外增加模型参数,相当于网络只学习到了输入与输出之间的残差特征,降低了网络训练的难度,同时,残差模块的引入可以较好地解决梯度弥散、梯度爆炸以及梯度退化等问题。此为双残差模型中的第二层残差。

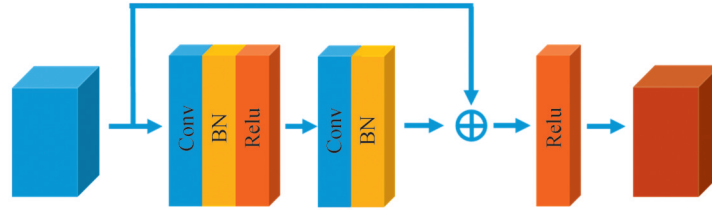


图3 残差块结构图

Fig.3 The structure of the residual block

为了进一步提升模型去噪性能,本文引入了通道注意力机制^[21],结构如图4所示。

$$F_{\text{channel}}(s) = s \cdot (H_{\text{max}}(s) + H_{\text{avg}}(s)) \quad (5)$$

式中, s 表示输入特征图, $H_{\text{max}}(s)$ 表示最大池化经过全连接层后的输出, $H_{\text{avg}}(s)$ 表示均值池化经过全连接层后的输出, $F_{\text{channel}}(s)$ 表示注意力模块的输出。输入的特征图经过最大池化和均值池化生成两个维度向量,分别经过全连接层线性优化之后相加得到输入特征图的通道注意力向量,最后与输入特征图相乘获得模块的输出特征图。通道注意力机制也就是给网络中间层特征图的不同维度赋予不同的权重,使得网络在参数优化过程中提升对重要维度的注意力,减弱对不重要维度的注意力。由于本文的特征图维度在不断缩减,通道注意力也可以引导网络在维度缩减过程中对重要特征的保留以及对冗余特征的舍弃。

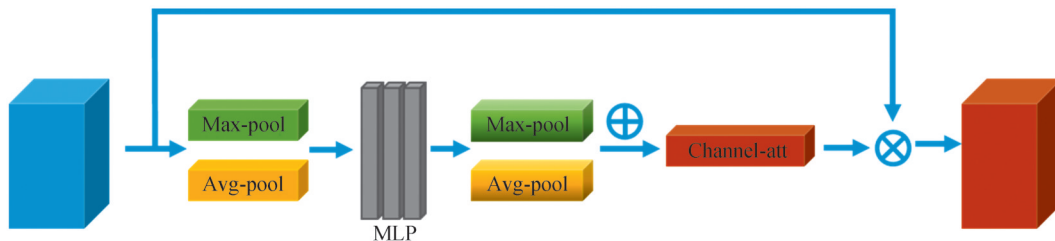


图4 通道注意力模块结构图

Fig.4 The structure of channel attention block

2 实验结果与分析

2.1 数据集的建立

现有的真实场景噪声数据大多通过调整相机ISO参数来控制噪声。比如RENOIR^[22]、DND^[23]、Nam^[24]、PolyU^[25]等数据集都是针对同一场景,拍摄低ISO图像作为真值图像,高ISO图像作为噪声图像,并调整曝光时间等相机参数使得两张图像亮度一致。但这种方法显然不适用于ICMOS图像的采集,低ISO只会让图像亮度过低从而丢失图像信息,因此,本文考虑采用多帧平均的方法,在暗室某一固定照度下对静态场景采集图像序列,然后多帧加权平均的方法合成真值图像。

实验在暗室环境下进行,并采用照度计对微光场景照度进行精确测量。本文数据集主要基于 2×10^{-1} 、 3×10^{-2} 、 2×10^{-3} lx三种照度下采用直耦型ICMOS相机进行图像序列的采集,每一个照度下采集7个场景的图像序列并做帧平均得到对应场景的真值图像。采集到的图像如图5所示,这里展示了一组静态场景的示意图,可以看出,随着照度的降低,图像的噪声强度在不断增加,亮度也在不断降低,而对应的帧平均图像可以比较好地去除图像的噪声,保留图像的细节信息,同时,能够与噪声图像的亮度保持一致。

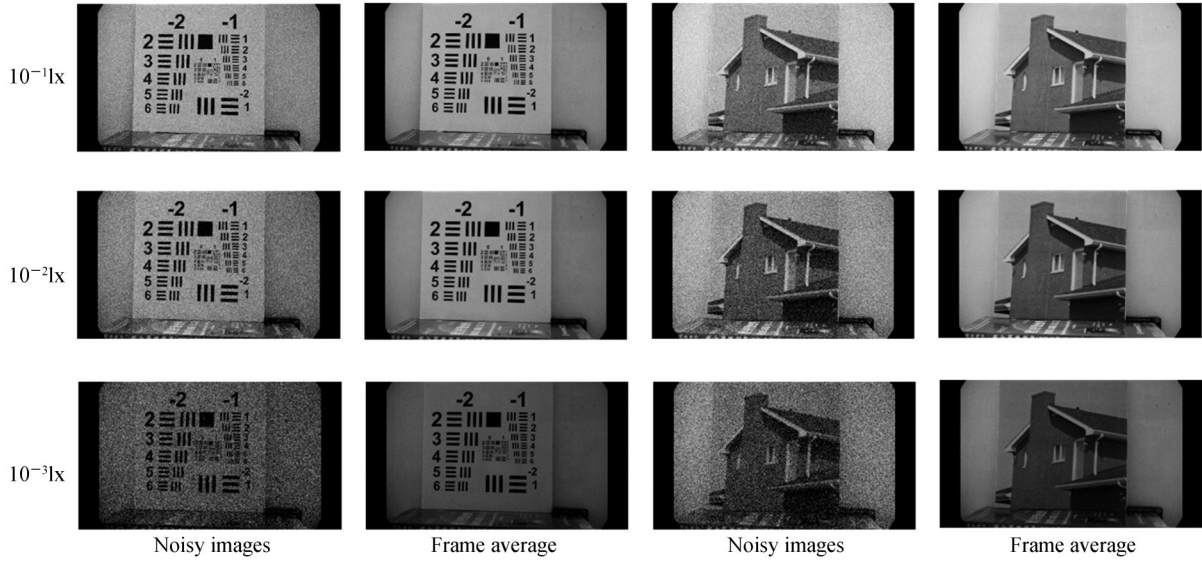


图5 不同照度下训练数据示例图

Fig.5 Examples under different illumination in the datasets

由于不同照度下的噪声强度和亮度不统一,所以本文针对不同照度下的图像分别进行训练,每种照度下采用两种静态场景作为训练集,每一个场景下采集1000张噪声图像,并与帧平均图像一一对应。

2.2 实验参数设置

由于本文去噪算法的任务为图像到图像的端到端映射,所以,本文采用的损失函数为基于图像评价指标的MSE损失和SSIM损失,具体表示为

$$\text{Loss} = \frac{1}{N} \sum_{i=1}^N \|y + \varphi(y) - x\|^2 + \text{SSIM}(y + \varphi(y), x) \quad (6)$$

式中, N 表示训练样本数, $\varphi(y)$ 为网络输出结果, x 为标签真值图像。网络采用Adam优化方法进行参数的优化,其中 $\beta_1 = 0.9$, $\beta_2 = 0.999$ 。网络权重采用Kaiming初始化,学习率的初始值为 1×10^{-3} ,并随着训练次数的增加不断衰减,每增加200轮则衰减为之前的0.9。所有图像大小统一为 1920×1080 ,但图像在输入时经过随机裁剪处理,每次输入为 256×256 大小的图像块。本文实验采用的CPU为Intel(R) Xeon(R) CPU E5-1603 v4,GPU为NVIDIA RTX3090,内存为32G。为了对比算法的性能,本文采用峰值信噪比(Peak Signal to Noise Ratio,PSNR)和结构相似度(Structural Similarity Index Measure,SSIM)客观评价指标进行估计。

2.3 实验对比效果

为了更好地评价双残差注意力网络去噪算法的优越性,本文采用几种主流的图像去噪算法与本文算法进行了对比实验,包括BM3D、非局部均值(NLM)、CBDnet、DnCNN和VDN。其中DnCNN、CBDnet以及VDN算法是基于深度学习的适用于真实图像盲去噪的主流方法,本文采用作者提供的开源代码以及预训练好的盲去噪模型参数进行对比试验。在测试时,虽然网络训练对输入图像进行了分块操作,但由于卷积网络的内在特性,训练好的模型可以处理任意大小的输入图像。图6~11展示了本文算法基于测试集中不同照度下的图像的去噪效果以及与主流算法的对比图。

从图中可以看出,非局部均值法(NLM)和基于深度学习的三种算法处理后图像仍然带有很严重的噪声干扰,主要因为这些算法对于ICMOS图像空间相关性的聚合状噪声泛化性较差,无法从输入图像中区分出噪声分量。而BM3D经典算法能够一定程度上去除图像的噪声,在 $2 \times 10^{-1} \text{lx}$ 下效果比较好,但图像细节过于平滑,导致部分细节丢失;而在照度更低的情况下,BM3D算法处理后图像会出现有比较明显的条纹状伪影,虽然图像的亮度有一定的提升,但这也比较影响图像的观感,其次,BM3D算法的时间成本非常高,计算量很大。相比较来看,本文算法在三种不同的照度下都保持着比较好地去噪效果,在最大限度上去除噪声的同时,也保留了图像的细节特征,而且,本文算法的时间成本低,效率比较高。

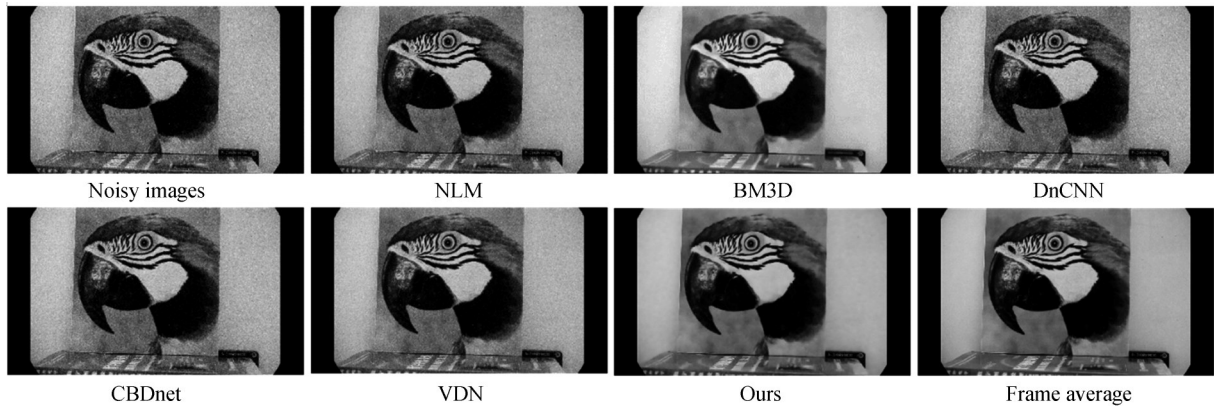


图6 $2 \times 10^{-1} \text{lx}$ 下鸚鵡图像的去噪效果
Fig.6 Denoising results of parrot image under $2 \times 10^{-1} \text{lx}$

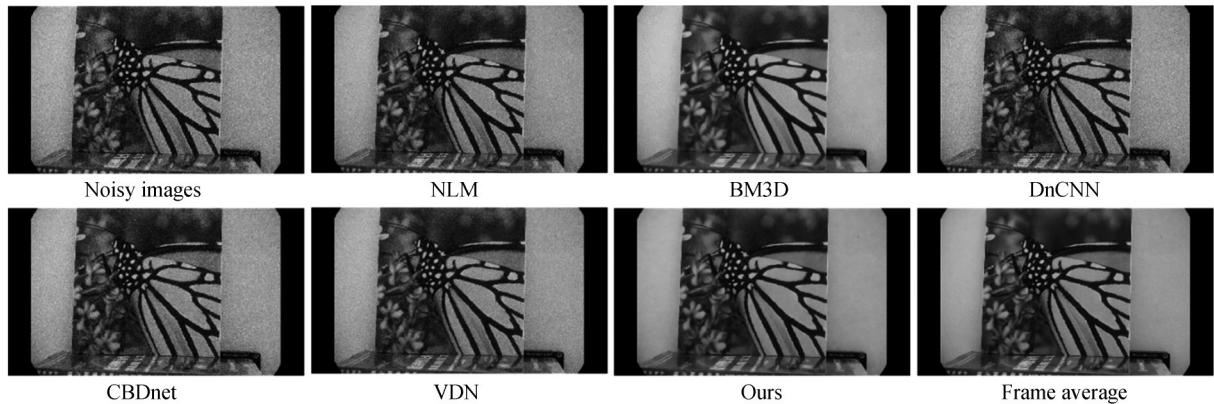


图7 $2 \times 10^{-1} \text{lx}$ 下蝴蝶图像的去噪效果
Fig.7 Denoising results of butterfly image under $2 \times 10^{-1} \text{lx}$

表1是本文算法与其他主流算法的客观评价指标数据对比,这里同样采用帧平均图像作为对比的真值图像。可以看出,本文算法在不同照度条件以及不同指标下都取得了最优越的效果。在PSNR方面,本文算法随着照度的降低始终保持着比较高的指标,而其他算法的PSNR则随着照度降低不断下降,本文算法基于所有测试图像的PSNR指标比输入噪声图像提升了56%,相比于BM3D算法领先了9.56 dB,而SSIM指标的结果也很类似,BM3D算法的SSIM指标相对于其他对比算法有比较大的领先,但仍然稍逊于本文算

法,基于所有图像测试图像的SSIM指标比输入图像提升了106%,领先了BM3D算法0.0503。为了更好地验证本文算法的运行效率,表2是不同算法处理单幅 1920×1080 大小图像的运行时间,可以看出基于深度学习的算法的运行时间要远远低于传统算法,而本文的算法速度最快,只需要0.005 s。

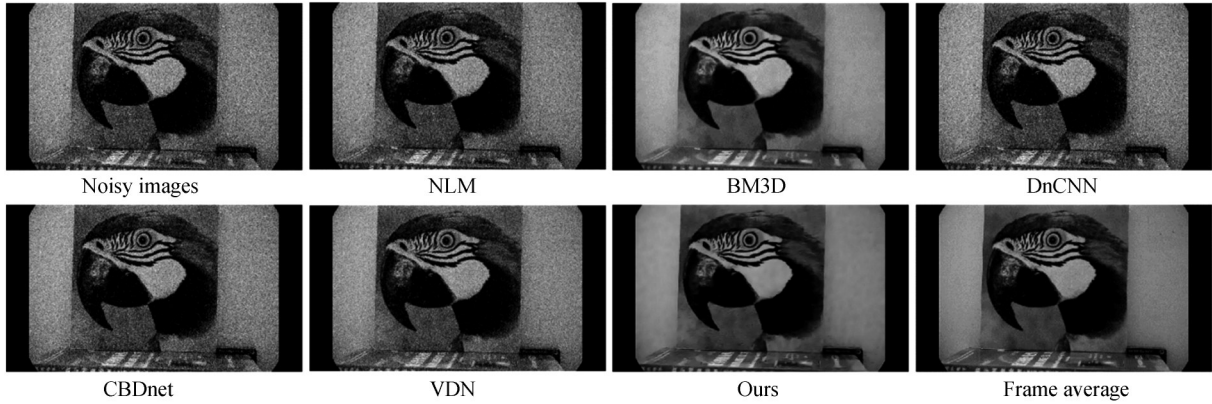


图8 3×10^{-2} lx 鸚鵡图像的去噪效果
Fig.8 Denoising results of parrot image under 3×10^{-2} lx

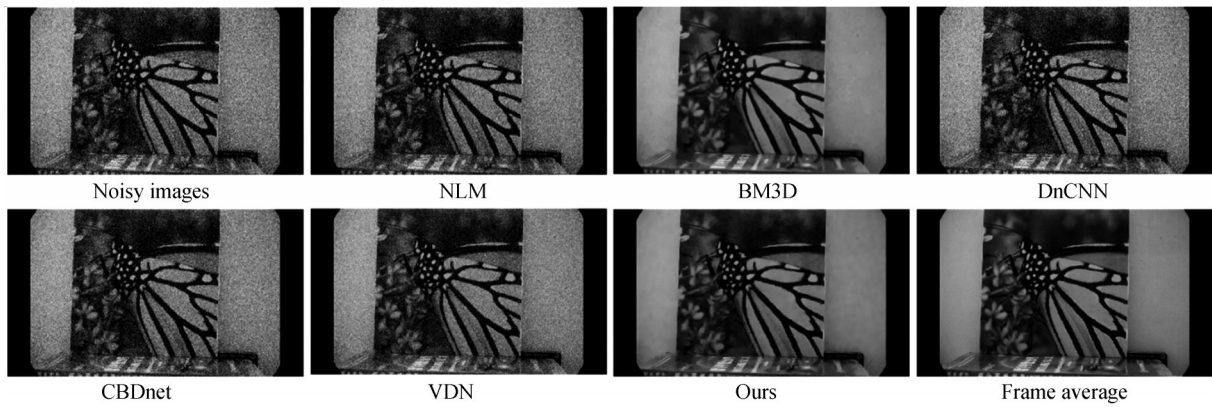


图9 3×10^{-2} lx 蝴蝶图像的去噪效果
Fig.9 Denoising results of butterfly image under 3×10^{-2} lx

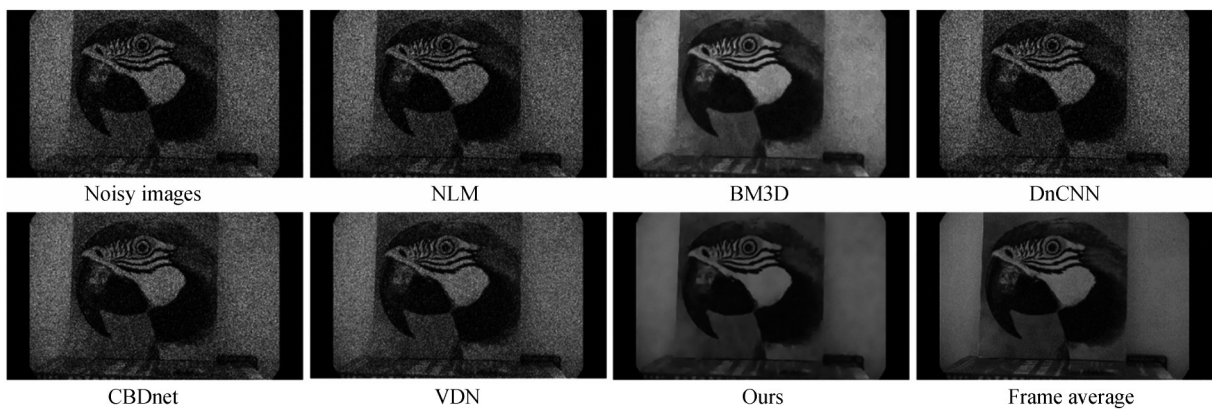


图10 2×10^{-3} lx 鸚鵡图像的去噪效果
Fig.10 Denoising results of parrot image under 2×10^{-3} lx

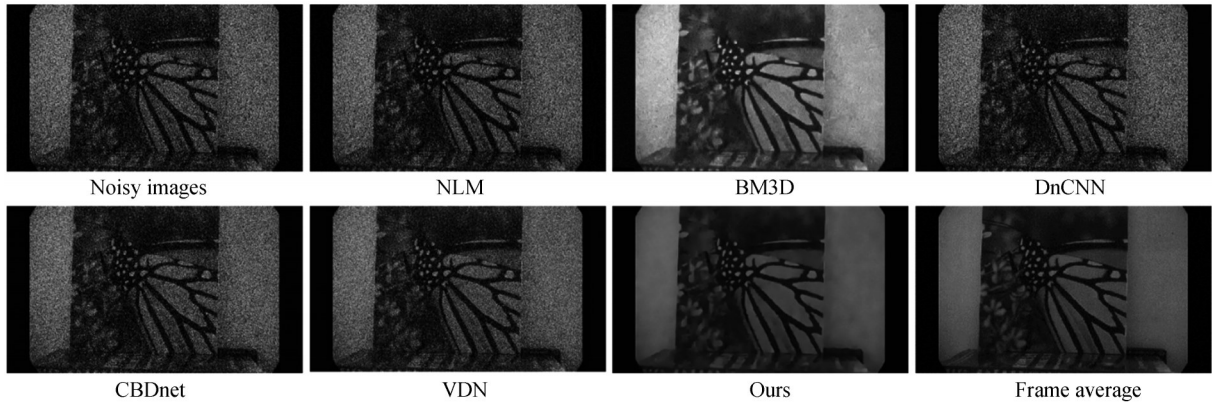


图 11 2×10^{-3} lx 蝴蝶图像的去噪效果
Fig.11 Denoising results of butterfly image under 2×10^{-3} lx

表 1 不同算法客观评价指标对比
Table 1 Comparison of objective evaluation indicators of different methods

Methods	PSNR/dB(↑)				SSIM(↑)			
	10^{-1} lx	10^{-2} lx	10^{-3} lx	Average	10^{-1} lx	10^{-2} lx	10^{-3} lx	Average
Noisy images	24.17	20.39	18.41	20.99	0.543 9	0.409 2	0.355 0	0.436 0
NLM	26.97	21.82	19.54	22.78	0.712 0	0.503 7	0.431 6	0.549 1
DnCNN	24.26	20.45	18.45	21.05	0.549 5	0.413 3	0.357 8	0.440 2
CBDnet	24.36	20.50	18.51	21.12	0.555 6	0.417 4	0.363 2	0.445 4
VDN	23.21	21.23	20.00	21.48	0.565 1	0.449 0	0.408 0	0.474 0
BM3D	25.44	26.64	17.45	23.18	0.865 7	0.878 6	0.800 1	0.848 1
Ours	33.23	32.11	32.87	32.74	0.895 0	0.886 5	0.913 7	0.898 4

表 2 不同算法处理单幅 1920×1080 图像的运行时间
Table 2 The running time of different methods for processing a single 1920×1080 image

Methods	Time cost/s	Platform
NLM	564	Python(CPU)
DnCNN	0.179	Pytorch(GPU)
CBDnet	0.01	Pytorch(GPU)
VDN	0.14	Pytorch(GPU)
BM3D	1 868	Python(CPU)
Ours	0.005	Pytorch(GPU)

3 结论

ICMOS是现有的应用广泛、工作照度最低的固体微光成像器件,但由于像增强器的引入,拍摄到的图像信噪比较低,随机噪声明显,极大降低了图像整体的观感,不利于人眼的观察和识别。本文设计了一个双残差注意力网络模型,并基于暗室环境下拍摄的不同照度下的静态场景数据建立了专门的ICMOS微光数据集。大量的实验结果表明,本文的算法能够有效地去除ICMOS图像中的随机噪声,而其他主流的去噪算法效果较差。下一步可以考虑噪声的时域特性,采用多帧融合的处理方式对ICMOS视频序列进行去噪算法的研究。

参考文献

- [1] WANG Fei, WANG Yibin, YANG Meng, et al. A denoising scheme for randomly clustered noise removal in ICCD sensing image[J]. Sensors, 2017, 17(2): 233.
- [2] TOMASI C, MANDUCHI R. Bilateral filtering for gray and color images[C]. Sixth International Conference on Computer Vision, 1998: 839-846.

- [3] BUADES A, COLL B, MOREL J M. A non-local algorithm for image denoising [C]. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005, 2: 60-65.
- [4] ZHANG Aimin. Denoising and fusion method of night vision image based on wavelet transform [J]. Electronic Measurement Technology, 2015, 38(1): 38-40.
张爱民. 一种基于小波变换的夜视图像去噪和融合方法 [J]. 电子测量技术, 2015, 38(1): 38-40.
- [5] DABOV K, FOI A, KATKOVNIK V, et al. Image denoising by sparse 3-D transform-domain collaborative filtering [J]. IEEE Transactions on Image Processing, 2007, 16(8): 2080-2095.
- [6] ELAD M, AHARON M. Image denoising via sparse and redundant representations over learned dictionaries [J]. IEEE Transactions on Image Processing, 2006, 15: 3736-3745.
- [7] MAIRAL J, BACH F, PONCE J, et al. Non-local sparse models for image restoration [C]. 2009 IEEE 12th International Conference on Computer Vision, 2009: 2272-2279.
- [8] DONG Weisheng, ZHANG Lei, SHI Guangming, et al. Nonlocally centralized sparse representation for image restoration [J]. IEEE Transactions on Image Processing, 2012, 22(4): 1620-1630.
- [9] BURGER H C, SCHULER C J, HARMELING S. Image denoising: Can plain neural networks compete with BM3D? [C]. 2012 IEEE Conference on Computer Vision and Pattern Recognition, 2012: 2392-2399.
- [10] ZHANG Kai, ZUO Wangmeng, CHEN Yunjin, et al. Beyond a Gaussian denoiser: residual learning of deep CNN for image denoising [J]. IEEE Transactions on Image Processing, 2016, 26(7): 3142-3155.
- [11] ZHANG Kai, ZUO Wangmeng, ZHANG Lei. FFDNet: Toward a fast and flexible solution for CNN-based image denoising [J]. IEEE Transactions on Image Processing, 2018, 27(9): 4608-4622.
- [12] TAI Ying, YANG Jian, LIU Xiaoming, et al. Memnet: a persistent memory network for image restoration [C]. Proceedings of the IEEE International Conference on Computer Vision, 2017: 4539-4547.
- [13] CHENG Shen, WANG Yuzhi, HUANG Haibin, et al. Nbnnet: Noise basis learning for image denoising with subspace projection [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 4896-4906.
- [14] YUE Zongsheng, YONG Hongwei, ZHAO Qian, et al. Variational denoising network: toward blind noise modeling and removal [J]. Advances in Neural Information Processing Systems, 2019: 32.
- [15] YUE Zongsheng, ZHAO Qian, ZHANG Lei, et al. Dual adversarial network: Toward real-world noise removal and noise generation [C]. European Conference on Computer Vision, 2020: 41-58.
- [16] ZAMIR S W, ARORA A, KHAN S, et al. Learning enriched features for real image restoration and enhancement [C]. European Conference on Computer Vision, 2020: 492-511.
- [17] GUO Shi, YAN Zifei, ZHANG Kai, et al. Toward convolutional blind denoising of real photographs [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 1712-1722.
- [18] GAO Tianyang, CAO Fengmei, WANG Xia, et al. Direct coupling of low light image intensifier with large size CMOS [J]. Infrared Technology, 2021, 43(6): 537-542.
高天阳, 曹峰梅, 王霞, 等. 微光像增强器与大尺寸CMOS的直接耦合 [J]. 红外技术, 2021, 43(6): 537-542.
- [19] 陈立. 微光ICCD数字图像去噪的研究 [D]. 长春: 长春理工大学, 2013.
- [20] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.
- [21] WOO S, PARK J, LEE J Y, et al. Cbam: Convolutional block attention module [C]. Proceedings of the European Conference on Computer Vision, 2018: 3-19.
- [22] ANAYA J, BARBU A. Renoir - a dataset for real low-light image noise reduction [J]. Journal of Visual Communication and Image Representation, 2018, 51: 144-154.
- [23] PLOTZ T, ROTH S. Benchmarking denoising algorithms with real photographs [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 1586-1595.
- [24] NAM S, HWANG Y, MATSUSHITA Y, et al. A holistic approach to cross-channel image noise modeling and its application to image denoising [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 1683-1691.
- [25] XU Jun, LI Hui, LIANG Zhetong, et al. Real-world noisy image denoising: a new benchmark [J]. arXiv preprint arXiv: 1804.02603, 2018.

Dual Residual Attention Network for ICMOS Sensing Image

WANG Xia^{1,2}, ZHANG Xin², JIAO Gangcheng¹, YANG Ye¹, CHENG Hongchang¹, YAN Bo¹

(1 Science and Technology on Low-Light-Level Night Vision Laboratory, Xi'an 710065, China)

(2 Key Laboratory of Optoelectronic Imaging Technology and System, Ministry of Education, School of Optics and Photonics, Beijing Institute of Technology, Beijing 100081, China)

Abstract: Low-light-level night vision technology is to explore the photoelectric technology that how to enhance, transmit, store, reproduce and apply the images captured under low light conditions. It is an important part of modern optoelectronic technology. ICCD/ICMOS (Intensified CCD/CMOS) is a solid low-light imaging device with a wide range of applications and the lowest working illuminance which is formed by coupling an image intensifier and CCD/CMOS. Although ICMOS can image under low-light night vision conditions, the image intensifier also amplifies the intensity of the noise while enhancing the signal, resulting in obvious random noise in the captured image, and the noise characteristics are more complex than that of traditional CMOS imaging. Due to the microchannel plates, ICMOS sensing image noise is not independent and identically distributed, but aggregated random noise with spatial correlation. Aggregated noise destroys the original structural features of the image, which also greatly increases the difficulty of denoising. In this paper, we propose a dual residual attention network for ICMOS sensing image denoising. There are three main ideas for our method. First, the network adopts the idea of residual learning, which means that the output of the network is the noise image, not the denoised image. Then the denoised image is achieved by subtracting the noise image from the original image. The residual learning network only needs to extract the noise component from the original image, which greatly reduces the difficulty of training the network. Secondly, we introduce four residual attention modules in our model, and the number of feature maps of each module is constantly decreasing. Each residual attention module consists of four residual blocks, one channel attention layer and one convolutional layer. The basic unit of the module is the residual block, which can effectively improve the network performance. At the same time, the introduction of the residual module can better solve the problems of gradient dispersion, gradient explosion and gradient degradation. Finally, the network introduces the channel attention layer, which can assign different weights to the output feature map of the middle layer, thereby analyzing the importance of each feature channel, and then enhancing the useful features and suppressing slight features according to this importance, and finally guide the network to continuously reduce the dimension of the feature map. Existing deep learning denoising methods mostly work for simulated Gauss-Poisson distributed noise and real noise data of some natural images. These methods can not be directly applied to ICMOS sensing images. Due to the particularity of ICMOS imaging noise, we made the ICMOS image dataset ourselves. We adopt the multi-frame averaging method to obtain the label image. The image sequence is captured from a static scene under a certain fixed illumination in the dark room, and then one label clean image of the image sequence is synthesized by a multi-frame weighted average method. The scene illuminance is accurately measured with an illuminance meter. This dataset is mainly based on three different illuminances 2×10^{-1} , 3×10^{-2} , 2×10^{-3} lx for image acquisition, and seven different static scenes are collected under each illuminance condition. Due to the inconsistency of noise intensity and brightness, we conduct model training for images under different illuminances. Two static scenes with 1 000 images are used as training sets under each illuminance. Our method applied the L1 loss as the loss function. From the subjective and objective results, it can be seen that our method has better denoising results and higher efficiency than other state-of-art methods.

Key words: Low-light-level night vision; ICMOS sensing image; Image denoising; Residual learning; Attention module

OCIS Codes: 100.2000; 100.2980; 100.3020; 150.0155