

基于前馈卷积神经网络的多模块 自监督降噪模型

郭晓新^{1,2}, 李广宇^{1,2}, 杨广奇^{1,2}

(1. 吉林大学 计算机科学与技术学院, 长春 130012; 2. 吉林大学 符号计算与知识工程教育部重点实验室, 长春 130012)

摘 要: 针对降噪过程中成对的噪声图像和干净图像难以获取的问题, 提出了一种基于前馈降噪卷积神经网络的多模块自监督降噪模型(MMS-DNet)。该模型采用改进的DnCNN架构, 通过添加结合原图下采样和全局噪声水平估计的多元输入模块, 生成多张输入子图的同时弥补了生成过程中的信息丢失; 通过添加子像素卷积模块, 避免了图像重建过程中冗余信息的引入。实验结果表明: 本文模型在灰度数据集BSD68和Set12上的效果优于现有的自监督学习模型, 在彩色数据集CBSD68和Kodak24上同样表现出良好的效果。

关键词: 图像降噪; 前馈神经网络; 自监督; 多模块

中图分类号: TP751 **文献标志码:** A **文章编号:** 1671-5497(2026)09-2467-09

DOI: 10.13229/j.cnki.jdxbgxb.20250138

Multi-modular self-supervised image denoising network based on feed-forward convolutional neural network

GUO Xiao-xin^{1,2}, LI Guang-yu^{1,2}, YANG Guang-qi^{1,2}

(1. College of Computer Science and Technology, Jilin University, Changchun 130012, China; 2. Key Laboratory of Symbol Computation and Knowledge Engineering of Ministry of Education, Jilin University, Changchun 130012, China)

Abstract: In order to solve the problem that pairwise noisy images and clean images are difficult to obtain in the process of denoising, this paper proposes a multi-modular self-supervised image denoising network based on feed-forward convolutional neural network (MMS-DNet). The algorithm uses the improved DnCNN architecture. By adding a multivariate input module that combines original image down-sampling and global noise level estimation, multiple input sub-images are generated while compensating for the information loss in the generation process. The introduction of redundant information in the image reconstruction process is avoided by adding a sub-pixel convolution module. Experimental results show that the network outperforms existing self-supervised learning algorithms on the BSD68 and Set12 datasets of intensity images, and also achieves good results on CBSD68 and Kodak24 datasets of color images.

Key words: image denoising; feed-forward neural network; self-supervised; multi-modular

收稿日期: 2025-02-24.

基金项目: 国家自然科学基金项目(82071995); 吉林省自然科学基金项目(20250102219JC).

作者简介: 郭晓新(1974-), 男, 教授, 博士, 研究方向: 计算机图形学, 医学影像学, 计算机视觉.

E-mail: guoxx@jlu.edu.cn

0 引言

图像降噪作为图像恢复的关键子问题,旨在从给定的噪声图像中去除无用的结构化或非结构化噪声信息,同时尽可能地保持原有图像的结构和特征,以恢复成高质量的干净图像。一个优秀的图像降噪模型不仅能有效提升图像的视觉表现质量,还可以作为其他计算机视觉任务的重要预处理步骤,例如图像恢复^[1]、图像分类等^[2]。通过高效的降噪处理,后续的图像分析和处理步骤能够在更准确和稳定的基础上进行,从而提高整体系统的性能和可靠性。

目前,主流的降噪模型可以分为两类:传统的降噪方法和基于深度学习的降噪方法。传统的降噪方法通过分析目标图像的噪声规律,采用滤波^[3]、低秩先验^[4]等数学方法处理图像降噪问题。Buades等^[5]提出非局部滤波算法(NLM),该方法使用像素与其邻域作为一个图像块,首先比较该图像块与整个图像其他图像块的相似性,再对局部和全局的相似度进行加权平均计算得到一个新的像素值来代替原始图像中的噪声像素;Dabov等^[6]先在图像中搜索相似的图像块,然后将这些相似块组合成三维数组,经过变换和硬阈值处理后在该数组上应用协同滤波来减少噪声。

相比于传统的图像降噪方法,基于深度学习的方法展现出更好的降噪效果和更强大的学习能力。Zhang等^[7]提出了前馈降噪卷积神经网络DnCNN,将堆叠的残差块加入神经网络学习图像的复杂特征和噪声分布,从而实现高效的降噪效果;Valsesia等^[8]提出一种基于图卷积操作的端到端的可训练神经网络架构,通过引入轻量级边缘条件卷积解决了特定图卷积的梯度消失和参数过多的问题,在合成高斯噪声和真实噪声上表现出了优异的性能。

然而,在实际的神经网络训练时收集大量成对的噪声图像和干净图像往往十分困难,并且成本高昂,因此只使用噪声图像降噪的自监督训练方法也逐渐应用到降噪任务中。Lehtinen等^[9]发现由同一张原始图像生成的两张噪声图像可以取代原本的干净噪声图像对实现同样的效果,巧妙解决了清晰样本难以获取的问题。由于生成对抗网络(GAN)^[10]具有生成多样化图像的特性,Chen等^[11]使用GAN来估计输入图像的噪声分布并生成噪声图像,并将其与输入图像成对作为卷积神

经网络的训练数据集对给定噪声进行降噪;Chen等^[12]直接应用Transformer^[13]作为主要结构在规模较大的ImageNet^[14]数据集上做预训练,并添加多个提取特征的头部和映射特征的尾部,在去噪、去雨等任务上取得了较好的效果;Pang^[15]等使用矩阵变换操作将单张输入图像变换成一对噪声图像输出给神经网络,仅使用小型数据集便达到了与有监督方法相同的降噪效果。

本文的主要贡献如下:①改进了DnCNN,并以此作为主体网络结构;②针对单尺度输入易丢失细节和固定噪声水平下模型灵活性不足的问题,设计了结合全局噪声水平估计与原图下采样形成的多元输入模块;③为达到减少计算量的同时保持图像重建效果的目的,采用子像素卷积作为上采样输出模块,设计了一种多模块自监督降噪模型(MMS-DNet)。实验表明,该模型在无需干净图像监督的条件下可有效完成降噪任务,同时作为其他视觉任务的关键预处理模块,通过提升输入数据信噪比显著增强目标检测、语义分割等下游任务的鲁棒性与准确度。此外,在医学影像分析和自动驾驶等领域,不仅通过减少噪声干扰提高了真实场景下决策的准确性,还验证了该模型在跨域应用中的泛化能力,证明了该模型的应用价值。

1 本文方法

1.1 网络模型

为了提高自监督图像降噪模型的效果,本文提出了一个多模块自监督降噪模型(Multi-modular self-supervised image denoising network, MMS-DNet),它由3个主要部分组成。第一部分是由多张独立子图和全局噪声水平估计组成的多元输入模块。第二部分是前馈神经网络,利用它作为主体结构来捕捉图像的噪声模式,并且可以替换为其他具有降噪效果的网络。最后,引入子像素卷积模块来实现降噪后图像的重建,网络结构如图1所示。

1.2 多元输入模块

1.2.1 原始图像下采样

降噪神经网络的训练往往需要大量的干净-噪声图像对,Noise2Noise^[9]首次证明了不需要干净图像,仅使用同一图像生成的两幅噪声图像也可以完成同样的降噪任务,即训练神经网络最小化如下的 l_2 损失:

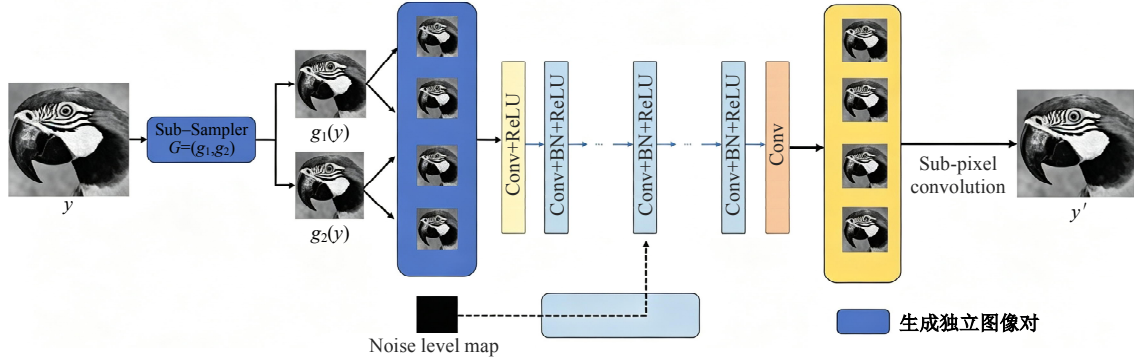


图1 模型架构图

Fig. 1 Model architecture diagram

$$E \left\{ \left\| f_{\theta}(y) - y' \right\|_2^2 \right\} \quad (1)$$

式中: y 和 y' 为添加噪声后独立生成的两张噪声图像; $f_{\theta}(\cdot)$ 为降噪神经网络。

为了达到自监督的目的,本文引入多元输入模块,由一张原始图像生成多张子图作为输入来解决没有干净图像的问题。首先,为了满足生成图像对条件独立的同时相同位置像素差距尽可能小,模型采用图2所示的方式:对于给定的宽为 W 、高为 H 的原始图像 y ,经下采样器(Down-sampler)两步处理,第一步,使用 2×2 的正方形区域将图像分为 $(W/2) \times (H/2)$ 个子像素块;第二步,对于每个 2×2 子像素块,随机选取相邻的一对像素并将其分别分配给两个子图,重复 $(W/2) \times (H/2)$ 次后即得到两个大小为 $(W/2) \times (H/2)$ 的子图 $G=(g_1(y), g_2(y))$ 。当给定的噪声图像 y 满足条件像素独立时,生成的子图 $(g_1(y), g_2(y))$ 也满足独立性,又由于子图 $(g_1(y), g_2(y))$ 由成对的相邻像素生成,因此可认为子图的相似性也得以满足。

为了进一步获取前馈神经网络的输入子图,受 R2R^[15] 启发,本文采用矩阵变换的方式将经上述下采样器生成的两张子图各自独立构建两个独立噪声子图、即如果噪声分布服从 $n \sim N(0, \sigma^2 I)$,使用

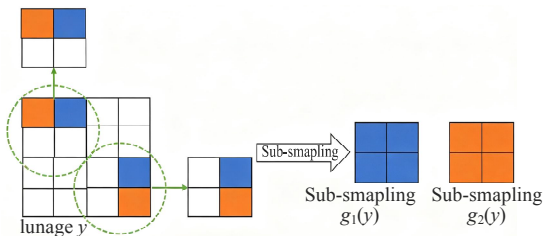


图2 下采样器

Fig. 2 Down-sampler

如下方法进行采样:

$$y' = y + D^T z, y'' = y - D^{-1} z \quad (2)$$

通过以上两个步骤,可以由单张原始噪声图像获得4张下采样噪声子图,并且它们相互独立。首先经过下采样器,原始图像 y 分为两个子图:

$$y_1 = g(y) = g(x) + g(n) \stackrel{\Delta}{=} x_1 + n_1 \quad (3)$$

$$y_2 = g(y) = g(x') + g(n') \stackrel{\Delta}{=} x_2 + n_2 \quad (4)$$

式中: g 为下采样操作,子图满足 $x \approx E[y_1 + y_2]$,即原始信号可通过子图恢复。

对于经过矩阵变换生成的子图,可视为添加噪声 ϵ_i ,即:

$$y_{1a} = y_1 + \epsilon_1, y_{1b} = y_1 + \epsilon_2 \quad (5)$$

$$y_{2a} = y_2 + \epsilon_3, y_{2b} = y_2 + \epsilon_4 \quad (6)$$

其中所有噪声项满足:

$$\begin{cases} E[\epsilon_i] = 0 \\ \text{Cov}(\epsilon_i, \epsilon_j) = 0, i \neq j \\ \text{Cov}(\epsilon_i, n) = 0 \end{cases} \quad (7)$$

设降噪网络为 $f_{\theta}(\cdot)$,本文模型目标为最小化以下损失函数:

$$L(\theta) = E \left[\left\| f_{\theta}(y_{1a}) - y_{1b} \right\|^2 + \left\| f_{\theta}(y_{2a}) - y_{2b} \right\|^2 \right] \quad (8)$$

展开单子图损失项,并代入 $y_1 = x_1 + n_1$,原式变为:

$$\begin{aligned} & E \left[\left\| f_{\theta}(y_{1a}) - y_{1b} \right\|^2 \right] = \\ & E \left[\left\| f_{\theta}(y_1 + \epsilon_1) - (y_1 + \epsilon_2) \right\|^2 \right] = \\ & E \left[\left\| f_{\theta}(x_1 + n_1 + \epsilon_1) - (x_1 + n_1 + \epsilon_2) \right\|^2 \right] = \\ & E \left[\left\| f_{\theta}(x_1 + n_1 + \epsilon_1) - x_1 \right\|^2 + \left\| n_1 + \epsilon_2 \right\|^2 + \right. \\ & \quad \left. 2 \langle f_{\theta}(y_{1a}) - x_1, n_1 + \epsilon_2 \rangle \right] \quad (9) \end{aligned}$$

式中:第一项 $E\|f_{\theta}(x_1+n_1+\epsilon_1)-x_1\|^2$ 为网络对干净子图 x_1 的估计误差;第二项 $E\|n_1+\epsilon_2\|^2=\sigma_n^2+\sigma_{\epsilon}^2$ 为原始噪声与新增噪声的总方差;第三项交叉项的期望为0。

因此,单子图的损失可以简化为:

$$E\left[\|f_{\theta}(y_{1a})-y_{1b}\|^2\right]=E\left[\|f_{\theta}(y_{1a})-x_1\|^2\right]+\sigma_n^2+\sigma_{\epsilon}^2 \quad (10)$$

总损失为子图损失之和:

$$L(\theta)=E\left[\|f_{\theta}(y_{1a})-x_1\|^2\right]+E\left[\|f_{\theta}(y_{2a})-x_2\|^2\right]+2(\sigma_n^2+\sigma_{\epsilon}^2) \quad (11)$$

忽略常数项,优化目标等价于:

$$\min_{\theta} E\left[\|f_{\theta}(y_{1a})-x_1\|^2+\|f_{\theta}(y_{2a})-x_2\|^2\right] \quad (12)$$

根据子图构造性质 $x \approx x_1+x_2$,网络在子图域学习到 $f_{\theta}(y_{ia}) \approx x_i$,则重建图像为:

$$f_{\theta}(y_{1a})+f_{\theta}(y_{2a}) \approx x_1+x_2 \quad (13)$$

进一步地,总损失可收敛到与传统监督学习损失相近的形式:

$$E\left[\|f_{\theta}(y_{1a}+y_{2a})-x\|^2\right] \quad (14)$$

因此,本文模型损失函数的期望与监督学习的MSE目标等价,并且子图提供的独立监督信号可降低训练方差,提升鲁棒性。

1.2.2 全局噪声估计

为了提高模型的通用性,MMS-DNet引入了噪声水平映射图,与原始图像生成的4张子图共

同作为前馈神经网络的输入。噪声水平映射图作为一个附加的输入通道,用于指示图像中每个像素的噪声强度,即噪声水平映射图的每个像素值表示相应图像像素的噪声方差。

由文献[21]可知,神经网络降噪模型的降噪过程可以抽象为如下公式:

$$\hat{x}=F(y,\sigma,\lambda,\theta) \quad (15)$$

式中: y 为噪声图像; σ 为噪声水平; λ 为平衡参数; θ 代表其他参数。

参数 λ 的选择会直接影响生成的图像质量,太小时图像会保留噪点,太大则会抹平图像细节。而平衡参数 λ 与噪声水平 σ 呈正相关,因此设置噪声水平 σ 同样可起到平衡控制图像整体降噪效果和保留边缘细节的作用。

针对整幅图像与单个像素维度不匹配的问题,在本文中将噪声水平 σ 拉伸为噪声水平映射图,根据手动添加的不同水平噪声($\sigma=15, 25, 50$),对于噪声图的每个像素以相同的噪声估计填充整个噪声图,以符合整个图像的均匀噪声分布。通过添加噪声水平映射图,保证了降噪的同时保留原始图像细节,并使单个MMS-DNet能处理不同数据集相同级别的噪声。

1.3 前馈神经网络

本文改进了DnCNN^[7]作为深层架构的主体。DnCNN结合了残差学习和卷积神经网络的优势,在有监督学习的降噪任务上表现出色,其网络结构如图3所示。

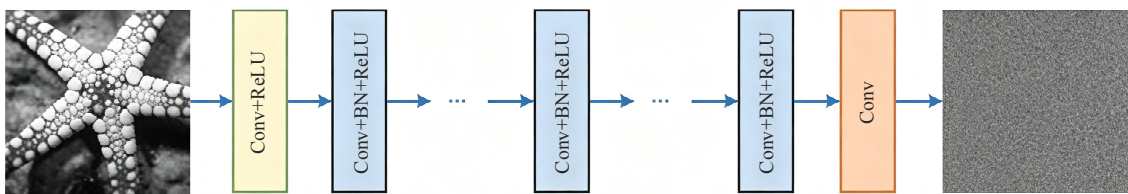


图3 DnCNN的网络结构

Fig. 3 Model architecture of DnCNN

具体地,DnCNN每层由3种主要操作组成:卷积(Conv)、修正线性单元(ReLU)、批量归一化(BN),其中批量归一化可以加快降噪模型的训练速度并提高降噪性能,修正线性单元可以帮助图像结构与噪声观察分离。本文的前馈神经网络模块保持了原始结构,首层采用“卷积-激活”组合模块实现特征初始映射,中间层使用“卷积-批归一化-激活”的复合计算单元,末层通过单一卷积

核完成信号重建,每次卷积后采用零填充(Zero-padding)保证每一层的输入、输出尺寸保持一致。由于添加了额外的全局噪声估计作为输入,为了兼顾整体网络的性能和效率,本文针对灰度图像和彩色图像设置了不同的层数和通道数,具体信息如表1所示。

1.4 子像素卷积

本文采用子像素卷积替代传统的反卷积进行

表1 不同图像网络结构设置

Table 1 Model architecture of different images

项目	网络层数	通道数	噪声范围	子图大小
灰度图	14	64	[15, 50]	70×70
彩色图	12	96	[15, 50]	50×50

图像重建。反卷积处理过程中容易引入过多的人工冗余信息,并且通过卷积运算实现上采样的效率不如子像素卷积通过像素重排的方式,而且开销也更大。

子像素卷积的过程如图4所示,对于给定的分辨率为 $n \times n$ 的原始图像,若想得到一个大小为 $3n \times 3n$ 的图像,即需要将原图等比例放大3倍,首先使用 1×1 大小的卷积核对原图进行卷积操作,生成 $3^2 = 9$ 大小的特征图,然后将这个相同尺寸特征图进行像素的重新排列,拼成一个放大3倍的高分辨率图像。



图4 子像素卷积

Fig. 4 Sub-pixel convolution

像素重排实际上是一个内存操作,不涉及复杂的数学计算,因此子像素卷积的计算量实际存在于 1×1 卷积部分,计算复杂度如下:

$$FLOPs = H_{in} \times W_{in} \times r^2 \times K^2 \quad (16)$$

式中: H 和 W 为输入图像尺寸; r 为放大因子; K 为卷积核尺寸。

采用子像素卷积作为上采样方法,避免引入人工冗余信息的同时减少了计算量,节省了整体开销。

2 实验结果与分析

2.1 实验环境及数据集

本文模型使用Pytorch深度学习框架,以PyCharm为实验平台实现,实验均在Ubuntu系统上运行,网络训练所使用的GPU型号为NVIDIA GTX3070。

对于灰度图像,本文选取BSD400数据集作为训练数据集。BSD400^[16]数据集包含400张分辨率为 180×180 的灰度图。测试数据集使用BSD68和Set12。BSD68包含68张从BSD300选出的大小不一的灰度图,Set12^[17]包含6张分辨率为 256×256 的灰度图和6张 512×512 的灰度图。

对于彩色图像,随机从ImageNet选取了400张图像作为训练数据集。测试数据集使用CBSD68和Kodak24^[18]。CBSD68是BSD68数据集的彩色版本,而Kodak24由原始Kodak数据集中心裁剪出的24张尺寸 500×500 的彩色图构成。

2.2 评价指标

本文采用峰值信噪比(PSNR)^[19]和结构相似度(SSIM)^[20]作为降噪效果的评价指标。PSNR值越大,表示降噪后的图像噪声越小,图像质量越接近原始图像。PSNR的公式如下:

$$PSNR = 10 \times \log \left(\frac{MAX_i^2}{MSE} \right) \quad (17)$$

式中: MAX_i^2 为像素最大值的平方,在8位图像的情况下为255;MSE为原始图像与处理后的图像之间对应像素值的平均差异的平方。

SSIM用来衡量两幅图像之间的相似度,特别关注图像的亮度、对比度和结构特征的相似性。SSIM值越接近1,表示图像相似性越高,降噪效果越好。SSIM的公式如下:

$$SSIM = \frac{(2\mu_Y \mu_{Y'} + c_1)(2\sigma_{YY'} + c_2)}{(\mu_Y^2 + \mu_{Y'}^2 + c_1)(\sigma_Y^2 + \sigma_{Y'}^2 + c_2)} \quad (18)$$

式中: Y 和 Y' 分别为原始干净图像和降噪网络处理后的图像; μ_Y 和 $\mu_{Y'}$ 分别为两张图像的均值; σ_Y^2 和 $\sigma_{Y'}^2$ 分别为两张图像的方差; $\sigma_{YY'}$ 为两张图像的协方差; c_1 和 c_2 为常数。

SSIM的值越大,两张图像的相似度就越高,最大值为1,当取值大于0.9时,就可以认为两张图像非常相似。

2.3 图像预处理

为了实现扩充数据集的目的,本文使用文献[21]的方法扩充训练集,训练时每个epoch从训练数据集中随机裁剪出 $128 \times 8\,000$ 个子图。如表1所示,对于灰度图像patch size设置为 70×70 ,彩色图像patch size设置为 50×50 。对于噪声图像,通过在干净图像上添加特定噪声水平($\sigma = 15, 25, 50$)的零均值AWGN生成。

本文模型采用MSE损失函数,使用自适应力矩估计(Adam)来优化损失函数,初始学习率设置为0.001,当损失函数停止减少时调整为0.0001。整个模型一共训练150个epoch,批量大小为128个。

2.4 综合降噪效果对比实验

为验证MMS-DNet模型的降噪效果,将本文

模型与目前主流的几种有监督和自监督模型的性能进行比较。为确保模型比较的公平性,本文中的模型结果均来自于原始论文中的数据。对于未在本文数据集上验证的方法,实验结果采用原始论文提供的代码进行再现。比较模型包括:基于非局部自相似性先验的两种有代表性的传统降噪模型 BM3D^[6]和 WNNM^[22],经典的有监督模型 DnCNN^[7],以及几种自监督模型 N2N^[9]、Blind2Unblind^[23]、SwinIA^[24]。

2.4.1 灰度图像实验

表2展示了MMS-DNet模型与其他模型在BSD68数据集上的效果对比,在BSD68数据集上本文模型优于其他自监督模型:当噪声水平 $\sigma=15$ 时,PSNR为31.66 dB,达到有监督模型的性能基准; $\sigma=25$ 时PSNR/SSIM分别提升0.11 dB/0.02; $\sigma=50$ 时虽略低于SwinIA,但较DnCNN仍保持0.07的PSNR优势。图5的视觉对比进一步验证,本文方法在边缘保持与纹理平滑方面更接近真实图像,尤其在复杂纹理区域减少了过度平滑导致的细节损失。

表2 在BSD68数据集上的不同噪声强度对比实验

方法	dB/-		
	$\sigma=15$	$\sigma=25$	$\sigma=50$
BM3D	31.07/0.82	28.57/0.80	25.62/0.69
WNNM	31.37/0.82	28.83/0.81	25.87/0.70
DnCNN	31.72/0.83	29.23/0.83	26.23/0.72
N2N	30.89/0.85	28.86/0.82	25.77/0.70
Blind2Unblind	31.44/0.89	28.99/0.82	26.09/0.72
SwinIA	31.07/0.86	29.17/0.80	26.61/0.71
MMS-DNet	31.66/0.87	29.28/0.82	26.30/0.72

表3展示了在Set12数据集上的实验结果对比,可知,本文在Set12数据集上 $\sigma=25$ 时PSNR和SSIM值分别为30.47 dB和0.88, $\sigma=50$ 时PSNR和SSIM值分别为27.41 dB和0.79,在对比的所有模型中取得了最好水平。在 $\sigma=15$ 时降噪效果略差于DnCNN等监督模型,说明本文提出的模型在低噪声等级下捕捉图像细节的能力较弱,可能由模型的架构做了轻量化处理导致。图6为不同方法在Set12数据集上的定性比较,可以看出本方法保留了更多的原始图像信息,减少了噪声对图像细节的影响。

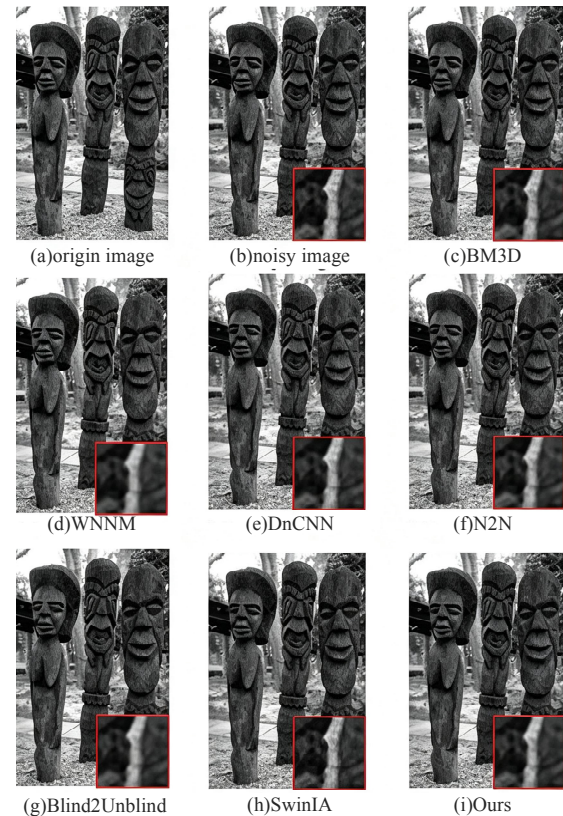


图5 在BSD68数据集上噪声等级 $\sigma=25$ 的降噪效果对比

Fig. 5 Image denoising results by different methods on BSD68 dataset

表3 在Set12数据集上的不同噪声强度对比实验

方法	dB/-		
	$\sigma=15$	$\sigma=25$	$\sigma=50$
BM3D	32.37/0.87	29.97/0.85	26.72/0.77
WNNM	32.70/0.85	30.26/0.86	27.05/0.78
DnCNN	32.86/0.91	30.43/0.87	27.18/0.78
N2N	32.60/0.90	30.02/0.85	26.83/0.77
Blind2Unblind	32.46/0.90	30.09/0.85	26.01/0.78
SwinIA	30.37/0.86	28.72/0.82	26.03/0.74
MMS-DNet	32.75/0.89	30.47/0.88	27.41/0.79

2.4.2 彩色图像实验

对于彩色图像,表4展示了CBSD68含噪测试图像经各模型降噪后的PSNR和SSIM值。从表中可以看出,本文模型在 $\sigma=15$ 和 $\sigma=25$ 时,相比于自监督模型,PSNR分别提升了0.20 dB和0.35 dB,相比于有监督模型,只低了0.02 dB。在 $\sigma=50$ 时,本文模型取得了最好效果,PSNR达到了27.95 dB,相较于自监督和有监督模型分别提升了0.07 dB和0.03 dB。在SSIM评价指标上,本文均达到了最高值。图7展示了在CBSD68数

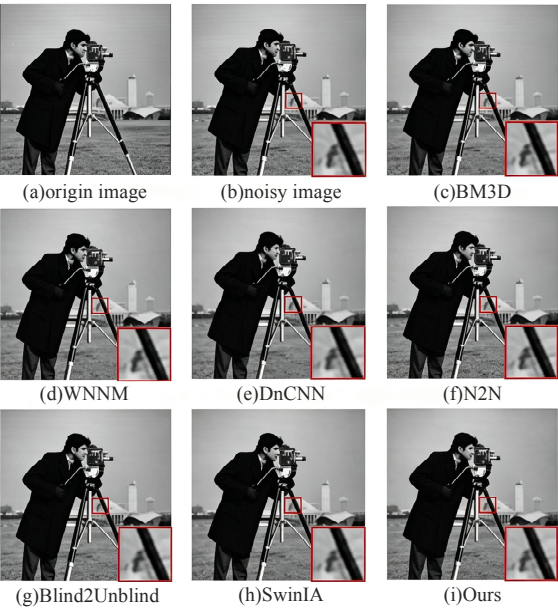


图 6 在 Set12 数据集上噪声等级 $\sigma=25$ 的降噪效果对比

Fig. 6 Image denoising results by different methods on Set12 dataset $\sigma=25$

表 4 在 CBSD68 数据集上的不同噪声强度对比实验

Table 4 Different noise level comparison on CBSD68 dataset

方法	$\sigma=15$	$\sigma=25$	$\sigma=50$
	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
BM3D	33.52/0.93	30.71/0.86	27.38/0.79
WNNM	33.67/0.93	31.02/0.87	27.79/0.79
DnCNN	33.89/0.93	31.23/0.88	27.92/0.79
N2N	— / —	30.79/0.87	— / —
Blind2Unblind	33.67/0.92	30.86/0/86	27.88/0.78
SwinIA	33.45/0.91	28.40/0.79	27.34/0.72
MMS-DNet	33.87/0.93	31.21/0.88	27.95/0.79

据集上的不同网络输出图像降噪效果对比图,可以看出,N2N降噪效果较差,图中仍存在部分噪点,DnCNN和Blind2Unblind整体效果较好,但边缘细节的处理较差,本文模型在整体和边缘细节的处理能力上均表现良好。

表 5 展示了 Kodak24 含噪测试图像经各模型降噪后的 PSNR 和 SSIM 值。在该数据集上,本文模型整体优于有监督模型,即在 3 种噪声强度下,PSNR 提升了 0.08~0.15 dB。在 $\sigma=25$ 时,PSNR 略差于自监督模型 Blind2Unblind,但 SSIM 仍达到了同等水平。图 8 为在 Kodak24 数据集上降噪效果的定性比较。从整体来看,DnCNN、Blind2Unblind 的降噪效果与本文模型相当,但从局部区域来看,本文模型更能还原原始图像的边缘细节和纹理细节。

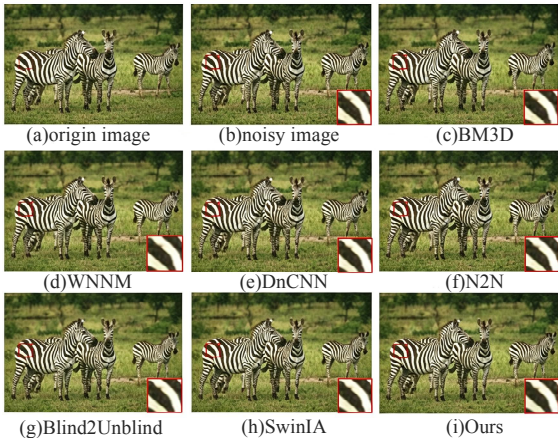


图 7 在 CBSD68 数据集上噪声等级 $\sigma=25$ 的降噪效果对比

Fig. 7 Image denoising results by different methods on CBSD68 $\sigma=25$

表 5 在 Kodak24 数据集上的不同噪声强度对比实验

Table 5 Different noise level comparison on Kodak24 dataset

方法	$\sigma=15$	$\sigma=25$	$\sigma=50$
	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
BM3D	34.28/0.92	31.68/0.87	28.46/0.78
WNNM	34.35/0.92	32.00/0.88	28.74/0.79
DnCNN	34.48/0.92	32.03/0.88	28.85/0.80
N2N	— / —	32.08/0.88	— / —
Blind2Unblind	34.12/0.90	32.27/0.88	28.92/0.81
SwinIA	33.50/0.87	30.12/0.82	28.33/0.75
MMS-DNet	34.62/0.92	32.11/0.88	29.00/0.80

2.5 消融实验

为了验证所添加模块的有效性,在添加了噪声强度 $\sigma=25$ 的 BSD68 数据集上进行消融实验。先将仅使用 DnCNN 模块的网络结构作为主干结构进行训练,得到一个降噪结果,再逐步将多元输入模块和子像素卷积上采样模块加入主干结构得到对应的降噪结果,并与添加其他上下采样模块的网络进行对比,结果如表 6 所示。

由表 6 可知,实验 1 在没有任何模块添加、网络结构仅为原始 DnCNN 的情况下,可得到的 PSNR 为 29.23 dB;实验 2 在实验 1 网络结构的基础上额外在首尾分别添加了下采样和上采样模块,PSNR 反而略微下降,表明单独引入普通的上下采样层并不一定学习到图像的局部细节,提升降噪效果;实验 3 和 4 分别使用多元输入模块和子像素卷积替代实验 2 的下采样和上采样模块,PSNR 都有所提升,分别证明了构造独立噪声子图和减少引入人工冗余信息的必要性;实验 5 即

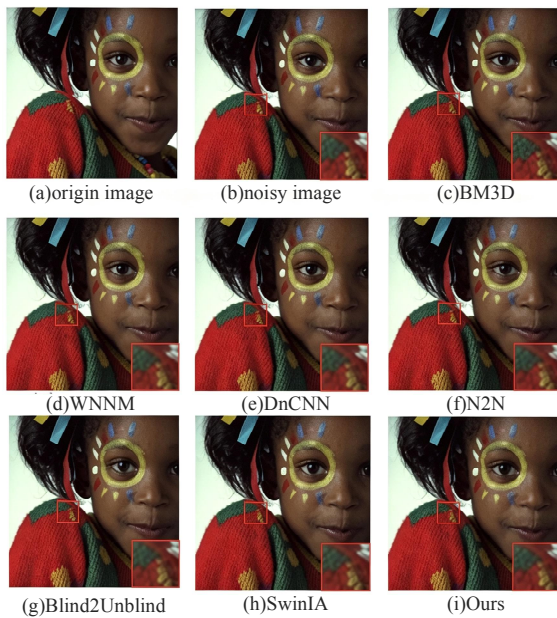


图 8 在 Kodak24 数据集上噪声等级 $\sigma=25$ 的降噪效果对比

Fig. 8 Image denoising results by different methods on Kodak24 $\sigma=25$

表 6 不同条件下消融实验

Table 6 Ablation experiment under different conditions

实验序号	架构模块	PSNR/dB
1	DnCNN	29.23
2	DnCNN+池化下采样+反池化	29.19
3	DnCNN+多元输入模块+反池化	29.25
4	DnCNN+池化下采样+子像素卷积	29.24
5	DnCNN+多元输入模块+子像素卷积	29.28

为本文 MMS-DNet 模型的实现,表明所有模块结合可以最大限度保留图像的原始信息,达到最佳的降噪效果。

2.6 模型可迁移性实验

本文主体的网络结构采用了 DnCNN 并做了微调,由于 MMS-DNet 各模块的独立性,主体网络结构可替换为其他可以有效处理降噪任务的神经网络,即 MMS-DNet 具有可迁移性,在 BSD68 数据集上针对不同噪声水平进行消融实验,实验结果(PSNR)如表 7 所示。

表 7 不同主体网络消融实验

Table 7 Ablation experiment with different backbone

主体网络	$\sigma=15$	$\sigma=25$	$\sigma=50$
DnCNN	31.66	29.28	26.30
U-Net	30.79	28.99	26.16
VDSR	30.77	28.64	26.12

特别地,由于 U-Net 在医学图像领域的分割任务上效果表现优异,在以 U-Net 为主要网络结构时,本文同样尝试针对脑部 MRI 图像做了降噪验证,从图 9 可以看到,在脑部的边缘部分噪声消除效果比较明显,表明本文的研究模型在医学图像领域也有一定的研究价值。

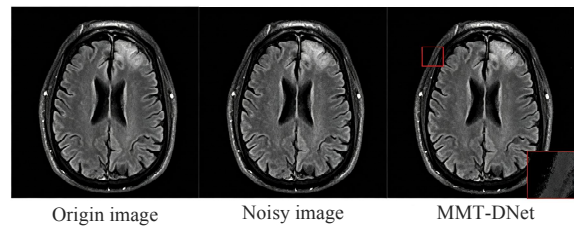


图 9 在脑部 MRI 图像上的降噪效果

Fig. 9 Image denoising results on MRI image

3 结束语

本文提出了一个多模块自监督降噪模型 MMS-DNet。在模型中,引入多元输入模块提取原始图像信息,其中下采样模块利用单幅原始图像生成多张降噪图像用于输入,额外添加的噪声估计图弥补了生成过程中的信息丢失;加入了子像素卷积,有助于避免图像上采样重建过程中的冗余信息输入,提升了降噪效果。实验结果表明:该网络能实现没有干净图像时的图像降噪任务,在峰值信噪比和结构相似性的指标上达到了有监督降噪模型的效果,有效提升了降噪准确率,有助于后续其他计算机视觉任务的进行。

参考文献:

- [1] Su J, Xu B, Yin H. A survey of deep learning approaches to image restoration[J]. Neurocomputing, 2022, 487: 46-65.
- [2] Liu D, Wen B X, Liu X M, et al. When image denoising meets high-level vision tasks: A deep learning approach[C]//Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, Stockholm, Sweden, 2018: 842-848.
- [3] Tomasi C, Manduchi R. Bilateral filtering for gray and color images[C]//Proceedings of Sixth International Conference on Computer Vision, Bombay, India, 1998: 839-846.
- [4] Gu S H, Zhang L, Zuo W M, et al. Weighted nuclear norm minimization with application to image denoising[C]//Proceedings of the IEEE Conference on Com-

- puter Vision and Pattern Recognition, Columbus, USA, 2014: 2862-2869.
- [5] 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, San Diego, USA, 2005: 60-65.
- [6] 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.
- [7] Zhang K, Zuo W M, Gu S H, et al. Learning deep CNN denoiser prior for image restoration[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA, 2017: 2808-2817.
- [8] Valsesia D, Fracastoro G, Magli E. Deep graph-convolutional image denoising[J]. IEEE Transactions on Image Processing, 2020, 29: 8226-8237.
- [9] Lehtinen J, Munkberg J, Hasselgren J, et al. Noise2Noise: learning image restoration without clean data[C]//Proceedings of International Conference on Machine Learning, Stockholm, USA, 2018: 2965-2974.
- [10] Goodfellow I, Pougetabadi J, Mirza M, et al. Generative adversarial networks[J]. Communications of the ACM, 2020, 63(11): 139-144.
- [11] Chen J W, Chen J W, Chao H Y, et al. Image blind denoising with generative adversarial network based noise modeling[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018: 3155-3164.
- [12] Chen H T, Wang Y H, Guo T Y, et al. Pre-trained image processing Transformer[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, USA, 2021: 12294-12305.
- [13] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. Advances in Neural Information Processing Systems, 2017, 30: 5998-6008.
- [14] Deng J, Dong W, Socher R, et al. ImageNet: A large-scale hierarchical image database[C]//2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, USA, 2009: 248-255.
- [15] Pang T Y, Zheng H, Quan Y H, et al. Recorrupted-to-recorrupted: Unsupervised deep learning for image denoising[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, USA, 2021: 2043-2052.
- [16] Martin D, Fowlkes C, Tal D, et al. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics[C]//Proceedings Eighth IEEE International Conference on Computer Vision, Vancouver, Canada, 2001: 416-423.
- [17] Zhang K, Zuo W M, Chen Y J, et al. Beyond a gaussian denoiser: residual learning of deep CNN for image denoising[J]. IEEE Transactions on Image Processing, 2017, 26: 3142-3155.
- [18] R Franzen . Kodak lossless true color image suite [EB/OL]. [2026-08-23]. <http://r0k.us/graphics/kodak/>
- [19] Huynh-Thu Q, Ghanbari M. Scope of validity of PSNR in image/video quality assessment[J]. Electronics Letters, 2008, 44(13): 800-801.
- [20] Wang Z, Bovik A C, Sheikh H R, et al. Image quality assessment: from error visibility to structural similarity[J]. IEEE Transactions on Image Processing, 2004, 13(4): 600-612.
- [21] Zhang K, Zuo W M, Zhang L. FFDNet: toward a fast and flexible solution for CNN-based image denoising[J]. IEEE Transactions on Image Processing, 2018, 27(9): 4608-4622.
- [22] Gu S H, Zhang L, Zuo W M, et al. Weighted nuclear norm minimization with application to image denoising[C]//IEEE Conference on Computer Vision and Pattern Recognition, Columbus, USA, 2014: 2862-2869.
- [23] Wang Z J, Liu J Z, Li G Q, et al. Blind2unblind: Self-supervised image denoising with visible blind spots[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, USA, 2022: 2017-2026.
- [24] Mikhail P, Chizhov P, Parts L. SwinIA: self-supervised blind-spot image denoising with zero convolutions[C]//2025 IEEE/CVF Winter Conference on Applications of Computer Vision, Tucson, Arizona, USA, 2023: 7071-7080.