

# 基于自监督约束的双尺度真实图像盲去噪算法\*

王迪, 潘金山, 唐金辉

(南京理工大学 计算机科学与工程学院, 江苏 南京 210094)

通信作者: 潘金山, E-mail: [jspan@njust.edu.cn](mailto:jspan@njust.edu.cn)



**摘要:** 现存的图像去噪算法在处理加性高斯白噪声上已经取得令人满意的效果, 然而其在未知噪声强度的真实噪声图像上泛化性能较差. 鉴于深度卷积神经网络极大地促进了图像盲去噪技术的发展, 针对真实噪声图像提出一种基于自监督约束的双尺度真实图像盲去噪算法. 首先, 所提算法借助小尺度网络分支得到的初步去噪结果为大尺度分支的图像去噪提供额外的有用信息, 以帮助后者实现良好的去噪效果. 其次, 用于去噪的网络模型由噪声估计子网络和图像非盲去噪子网络构成, 其中噪声估计子网络用于预测输入图像的噪声强度, 非盲去噪子网络则在所预测的噪声强度指导下进行图像去噪. 鉴于真实噪声图像通常缺少对应的清晰图像作为标签, 提出了一种基于全变分先验的边缘保持自监督约束和一个基于图像背景一致性的背景自监督约束, 前者可通过调节平滑参数将网络泛化到不同的真实噪声数据集上并取得良好的无监督去噪效果, 后者则可借助多尺度高斯模糊图像之间的差异信息辅助双尺度网络完成去噪. 此外, 还提出一种新颖的结构相似性注意力机制, 用于引导网络关注图像中微小的结构细节, 以便复原出纹理细节更加清晰的真实去噪图像. 相关实验结果表明在 SIDD, DND 和 Nam 这 3 个真实基准数据集上, 所提的基于自监督的双尺度盲去噪算法无论在视觉效果上还是在量化指标上均优于多种有监督图像去噪方法, 且泛化性能也得到了较为明显的提升.

**关键词:** 卷积神经网络; 真实噪声图像; 图像盲去噪; 自监督约束; 注意力机制

**中图法分类号:** TP391

中文引用格式: 王迪, 潘金山, 唐金辉. 基于自监督约束的双尺度真实图像盲去噪算法. 软件学报, 2023, 34(6): 2942–2958. <http://www.jos.org.cn/1000-9825/6512.htm>

英文引用格式: Wang D, Pan JS, Tang JH. Two-scale Real Image Blind Denoising with Self-supervised Constraints. Ruan Jian Xue Bao/Journal of Software, 2023, 34(6): 2942–2958 (in Chinese). <http://www.jos.org.cn/1000-9825/6512.htm>

## Two-scale Real Image Blind Denoising with Self-supervised Constraints

WANG Di, PAN Jin-Shan, TANG Jin-Hui

(School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China)

**Abstract:** Existing image denoising algorithms have achieved decent performance on the images with the additive white Gaussian noise (AWGN), while their generalization ability is not good on real-world images with unknown noise. Motivated by the significant progress of deep convolution neural networks (CNNs) in image denoising, a novel two-scale blind image denoising algorithm is proposed based on self-supervised constraints. Firstly, the proposed algorithm leverages the denoised results from the small-scale network branch to provide additional useful information for the large-scale image denoising, so as to achieve favorable denoised results. Secondly, the used network is composed of a noise estimation subnetwork and an image non-blind denoising subnetwork. The noise estimation subnetwork is firstly used to predict noise map, and then image denoising is carried out through the non-blind denoising subnetwork under the guidance of the corresponding noise map. In view of the fact that the real noise image lacks the corresponding clean image as the label, an edge-preserving self-supervised constraint is proposed based on the total variation (TV) priori, which generalizes the network to different real

\* 基金项目: 科技创新 2030 “新一代人工智能”重大项目 (2018AAA0102001); 国家自然科学基金 (61922043, 61872421)  
收稿时间: 2021-03-22; 修改时间: 2021-07-16, 2021-09-16; 采用时间: 2021-10-12; jos 在线出版时间: 2022-12-22  
CNKI 网络首发时间: 2022-12-26

noisy datasets by adjusting smoothing parameters. To keep the consistency of the image background, a background guidance module (BGM) is proposed, which builds a self-supervised constraint based on the information difference between multi-scale Gaussian blurred images and thus assists the network to complete image denoising. In addition, the structural similarity attention mechanism (SAM) is proposed to guide the network to pay attention to trivial structural details in images, so as to recover real denoised images with cleaner texture details. The relevant experimental results on the SIDD, DND, and Nam benchmarks indicate that the proposed self-supervised blind denoising algorithm is superior to some deep supervised denoising methods, and the generalization performance of the network is improved significantly.

**Key words:** convolutional neural network (CNN); real-world noisy images; image blind denoising; self-supervised constraint; attention mechanism

## 1 引言

日常生活中,人们观测到的图像大多都是通过数字成像设备获取的,在成像的过程中由于外部光线和设备内部元器件的干扰,原始清晰图像在传输和压缩的过程中容易遭受噪声的退化,导致观测到的图像往往是噪声图像。图像去噪作为计算机底层视觉研究中的热点问题之一,旨在从退化的噪声图像中复原出潜在的清晰图像,目前已广泛应用于自然图像、医学图像和遥感图像的处理,以及高层视觉问题的图像预处理。传统的图像去噪方法致力于处理特定的加性高斯白噪声 (additive white Gaussian noise, AWGN), 并对于未知噪声强度的真实噪声图像泛化性能较差,属于依赖噪声强度的非盲去噪。近些年来,随着深度学习在计算机视觉领域取得的重大突破,卷积神经网络的提出极大地促进了真实噪声图像盲去噪技术的发展。综上所述,图像去噪方法可以划分为两大类:基于先验的传统去噪方法和基于学习的深度去噪方法。

在基于先验的传统去噪方法中, Dabov 等人<sup>[1]</sup>提出了一种基于结构自相似性的去噪方法,其利用噪声图像块之间的局部自相似性提出三维块匹配滤波算法实现图像去噪。在结构自相似性的启发下, Dong 等人<sup>[2]</sup>提出了一种基于稀疏表示的方法,利用非局部自相似性对稀疏表示方法进行优化,取得了良好的去噪效果。鉴于基于稀疏表示的方法计算代价较大, Elad 等人<sup>[3]</sup>提出了一种基于字典编码的去噪方法,该方法采用基于稀疏和冗余表示的字典编码方法从给定的噪声图像中去除加性高斯白噪声。此外,还有一些其他基于图像先验的传统去噪方法,例如基于马尔科夫随机场 (Markov random field, MRF) 的去噪方法<sup>[4]</sup>和基于极大似然估计 (maximum likelihood estimate, MLE) 的去噪方法<sup>[5]</sup>。上述传统图像去噪方法都是针对特定的加性高斯白噪声提出的,它们严重依赖于给定的噪声强度,对于未知噪声强度的真实图像泛化性能较差。然而在现实应用中,噪声图像往往是在不同的光照条件下通过不同型号/参数的成像设备获取的,即使在同一静态场景下,也可能获取到噪声强度不同的图像,甚至出现同一图像的不同区域噪声强度亦不相同的情况。因此真实图像的噪声模型与特定的加性高斯白噪声模型存在一定程度的偏差,这就导致现存的传统去噪方法难以解决真实噪声图像的去噪问题。

为了解决上述问题,近年来,一些基于深度学习的真实图像去噪方法被相继提出。Guo 等人<sup>[6]</sup>提出一种模拟真实噪声的近似模型用于合成训练数据,并进一步提出使用真实噪声图像与合成噪声图像交替训练网络模型用于估计噪声强度并实现去噪,从而提升该方法对真实噪声图像的泛化性能。但是,该方法仍然未能解决合成噪声与真实噪声不匹配的问题。随着一些真实噪声数据集 (DND<sup>[7]</sup>, SIDD<sup>[8]</sup>) 的提出, Anwar 等人<sup>[9]</sup>提出使用真实噪声数据集对网络进行有监督训练,实验结果证明该方法对真实噪声图像表现出良好的泛化性能。然而现存的真实噪声数据集存在以下两方面的不足。第一,数据量少,目前规模最大的真实噪声数据集 SIDD 仅包含 160 个场景的图像;第二,获取到的清晰图像与真实的清晰图像有所差别,因为这些数据集中的清晰图像是通过同一场景中拍摄的多幅图像进行平均或者通过调节设备参数得到的。

基于以上分析,本文提出仅使用真实噪声图像去训练深度网络以实现自监督盲去噪。经过试验发现,同一场景下不同尺度的噪声图像所包含的信息量是不同的,小尺度噪声图像的峰值信噪比较大尺度图像更大一些。鉴于此,本文提出一种基于自监督约束的双尺度真实图像盲去噪算法,其主要贡献总结如下。

(1) 本文提出了一种基于自监督约束的双尺度真实图像盲去噪算法,并针对噪声图像的不同尺度设计一种两

阶段网络模型,解决了合成噪声与真实噪声不匹配的问题.

(2) 本文针对真实噪声图像设计了两种自监督约束,基于全变分先验的边缘保持自监督约束和基于图像背景内容一致性的背景自监督约束,前者通过调节平滑参数可将网络泛化到不同的真实噪声数据集上,后者通过构造一个多尺度的背景指导模块辅助训练网络实现更好的去噪效果.

(3) 本文提出了一种新颖的结构相似性注意力机制,通过对结构相似度图谱进行高斯拉伸,从而引导网络复原出更丰富的图像细节.

## 2 相关工作

### 2.1 基于图像先验的传统去噪方法

基于图像先验的传统去噪方法主要涉及非局部结构自相似性<sup>[1,10]</sup>、稀疏先验<sup>[2,3]</sup>、字典编码<sup>[3]</sup>、马尔科夫随机场<sup>[4]</sup>以及极大似然估计<sup>[5]</sup>等. Buades 等人<sup>[10]</sup>提出的非局部均值滤波算法 NLM 率先使用了非局部结构自相似性. 该方法将每个图像块的中心像素值替换为相似块中心像素值的加权平均得到无噪声的平滑图像. Dabov 等人<sup>[1]</sup>提出的三维块匹配滤波算法利用噪声图像中不同块之间的非局部自相似性来寻找与参考块相似的二维图像块,并将这些二维图像块组成若干三维组,经过空间变换后通过协同滤波完成去噪. 该算法是非局部结构自相似性的典型应用,其性能比 NLM 更鲁棒. 鉴于 NLM 在去除噪声的同时会过度模糊图像的边缘纹理信息,邢笑笑等人<sup>[11]</sup>则提出一种渐进非局部平均图像去噪算法,利用方差的性质对图像进行渐进式滤波来缓解边缘纹理过平滑. Burger 等人<sup>[12]</sup>使用普通多层感知机 (plain multi-layer perceptron, MLP) 实现了与文献 [1] 相当的性能. Dong 等人<sup>[2]</sup>提出的非局部集中稀疏表示方法利用非局部自相似性对稀疏表示方法进行优化,取得了良好的去噪性能. Elad 等人<sup>[3]</sup>采用基于稀疏和冗余表示的字典学习方法从给定的噪声图像中去除零均值白噪声和均匀加性高斯白噪声. Roth 等人<sup>[4]</sup>开发出一种学习图像先验的通用方法. 该方法通过在像素邻域上学习势函数来扩展传统的马尔科夫随机场 (Markov random field, MRF) 模型,并且在图像去噪任务中证明了该框架的良好性能. Zoran 等人<sup>[5]</sup>提出利用极大似然估计 (maximum likelihood estimate, MLE) 实现从图像块中学习其先验知识,并复原整幅图像的通用方法. 除了上述方法以外,基于变分自编码器的图像去噪方法也趋于流行. 例如,江泽涛等人<sup>[13]</sup>提出一种多重自编码器 MR-VAE,从特征概率分布捕获、全局重构和细节重构 3 方面由粗到精地复原出噪声更少,细节更丰富的图像. 这些基于图像先验的传统去噪方法往往过于强调模型及其可解释性,且强烈依赖于噪声强度的估计,难以适应空间变化的噪声,对于未知噪声强度的图像泛化性能较差.

### 2.2 基于深度学习的图像去噪方法

现存的基于学习的深度去噪方法可以划分为图像盲去噪和图像非盲去噪两大类. 典型的非盲去噪方法例如 DnCNN<sup>[14]</sup>、FFDNet<sup>[15]</sup>、CDCNN<sup>[16]</sup>和 EGDNet<sup>[17]</sup>主要用于处理合成的加性高斯白噪声,在真实世界的噪声图像上其泛化性能较差. 针对这一问题,一些基于学习的深度盲去噪方法被提出.

Mildenhall 等人<sup>[18]</sup>首先提出核预测算法 KPN. 该算法使用深度卷积神经网络逐像素预测空间中变化的三维去噪核,并利用从每个像素上得到的核对噪声图像进行局部配准和去噪,不仅在 AWGN 图像上实现了盲去噪,还能够泛化到真实的噪声图像中. 由于现存的真实噪声图像严重不足且拍摄难度大成本高,因此 Guo 等人<sup>[6]</sup>基于异构高斯噪声模型和相机内部处理流程提出一种真实噪声的近似模拟用于合成训练数据,并使用真实噪声图像与合成噪声图像交替训练提出的两阶段网络用于估计噪声强度并实现图像去噪. 该方法不仅缓解了真实噪声数据不足的窘境而且提升了网络对真实噪声图像的泛化性能. 随着 SIDD<sup>[8]</sup>、DND<sup>[7]</sup>等一些大规模真实噪声基准数据集的提出,Anwar 等人<sup>[9]</sup>和 Zhao 等人<sup>[19]</sup>相继提出使用真实噪声数据集来训练网络,进一步提升了网络对真实噪声图像的泛化性能. 除了上述基于数据驱动的图像盲去噪算法,基于图像先验的盲去噪算法也被提出. Yue 等人<sup>[20]</sup>提出的基于变分推理的盲去噪方法根据固有的清晰图像和噪声图像的方差构建近似后验,并在噪声先验和近似后验的指导下使用贝叶斯框架从观测到的噪声图像中直接推断潜在的噪声分布和清晰图像. 虽然这些真实场景的有监督盲去噪方法已经取得显著进展,然而由于无法获取真实的清晰图像作为监督,自监督的图像盲去噪算法依然是一个充满

挑战的研究热点。

目前的基于深度学习的自监督盲去噪方法主要依赖构造模拟真实噪声数据集、自适应域迁移和图像先验知识 3 种策略。Chen 等人<sup>[21]</sup>率先提出通过构造模拟真实噪声数据集实现图像自监督去噪的方法。其网络结构分为两个阶段,第 1 阶段借助生成对抗网络 (generative adversarial network, GAN) 学习输入图像的噪声分布,构建大量成对的噪声/清晰图像;第 2 阶段使用第 1 阶段构建的数据集来训练盲去噪网络。该方法借助数据增强的思想将无监督去噪问题转化成有监督去噪,证明了生成对抗网络在噪声建模方面的潜力。Lin 等人<sup>[22]</sup>提出基于噪声域自适应的数据增强方法,通过将源噪声图像映射到目标噪声图像来构建大规模的真实噪声数据集,并提出注意生成网络 (attention generative network),借助视觉注意机制更多地关注图像的噪声区域,使得网络在去除噪声和保持纹理之间取得平衡,避免发生过度平滑。近年来,许多基于图像先验的自监督图像盲去噪算法也被相继提出。Lehtinen 等人<sup>[23]</sup>提出一种新颖的深度去噪方法 Noise2Noise (N2N)。该方法以高噪声图像作为输入,以低噪声图像代替清晰图像作为标签来训练网络。尽管清晰的图像标签是非必要的,但该方法仍需采用不同退化程度的噪声图像对去优化网络参数,因此 N2N 被视为一种特定的无监督去噪方法。此外,在同一静态场景中获取不同强度的噪声图像对通常也是较为困难的。与 N2N 不同,Krull 等人<sup>[24]</sup>提出基于盲点网络的自监督图像去噪方法 Noise2Void (N2V)。该方法假设输入图像的噪声分布符合统计独立性,在无需清晰图像和成对噪声图像的条件下,仅使用单个噪声图像来训练盲点网络实现去噪。在盲点网络中,其感知野不包括中心像素,仅使用其邻域信息复原图像内容。然而,盲点的存在意味着信息的丢失,导致 N2V 复原的图像在质量上略逊于 N2N。鉴于此,Laine 等人<sup>[25]</sup>在 N2V 的基础上改进了盲点网络,借助最大后验概率的方法将盲点隐藏的信息开发出来,并与网络输出相结合,进而避免信息丢失,提高网络复原清晰图像的能力。Batson 等人<sup>[26]</sup>提出的自监督盲去噪框架 Noise2self (N2S) 给出与 N2V 相似的假设,即在图像的不同维度上,噪声信号表现出统计独立性,而图像内容信号则表现出某种相关性。在满足假设的条件下,N2S 使用噪声独立于图像内容的特征组进行相互预测,从每个对象的单一噪声测量中学习去噪函数实现图像盲去噪。最新的研究<sup>[27]</sup>提出一种基于目标函数变分近似的贝叶斯框架,以一种“分而治之”的策略将复杂的后验分布分解成更加简单的子分布,从而将复杂的去噪问题简单化,最终目标是寻求一个最优的先验去指导单一的卷积神经网络模型。该方法虽然取得了目前相对最佳的去噪性能,但是该方法是一种全监督的图像去噪方法且依赖于人工设计的图像先验。

上述基于深度学习的自监督图像盲去噪方法虽然取得了一定成就,但是仍然存在一些缺点,一方面是对真实噪声的建模与真实的噪声模型有所偏差;另一方面是依赖于噪声分布符合统计独立性的假设。鉴于此,本文提出一种新的基于自监督约束的双尺度真实图像盲去噪算法,所设计的自监督约束普遍适用于现有的一些深度网络,旨在挖掘噪声图像本身潜在的特征属性,从而实现真实噪声图像的盲去噪。

### 3 本文方法

本节主要介绍自监督约束下的真实图像盲去噪算法。首先,描述模型的整体结构和去噪流程;然后,详细阐述针对真实噪声图像设计的边缘保持自监督约束,背景自监督约束以及所提出的结构相似性注意力机制;最后,讨论网络模型所使用的损失函数。

#### 3.1 模型框架

本文提出自监督约束下真实噪声图像的双尺度盲去噪网络。该网络由两个尺度分支组成,且这两个分支网络共享相同的参数。每个分支网络包含一个噪声估计子网络 (Net-E) 和一个图像非盲去噪子网络 (Net-D)。Net-E 的作用是预测真实噪声图像的噪声强度,而 Net-D 的作用是在噪声强度的指导下实现真实图像的非盲去噪。图 1 展示了 Net-E 和 Net-D 的网络结构。Net-E 是一个不包括池化层 (pooling layers) 和批归一化操作 (batch normalization, BN) 的全卷积网络,包括 20 个卷积层,且除了输出层外的每个卷积层都伴随一个非线性激活函数,其结构如图 2 所示,网络参数设置如表 1 所示。Net-D 采用一个 U 型网络框架<sup>[28]</sup>,其编码器使用最大池化操作进行下采样,有助于挖掘图像的多尺度信息,扩大特征感知野,解码器使用双线性差值进行上采样用于重建去噪后的图像。Net-D 的

网络结构如图 3 所示, 其中编码器和解码器的结构和参数设置如表 2 所示, 除了最后一个卷积层, 其余每个卷积层后包含有一个非线性激活函数. 为了扩大 Net-D 的网络容量, 获取更丰富的图像细节特征, 我们在编码器和解码器中间嵌入 3 个残差密集块 (residual dense block)<sup>[29]</sup>. 这些残差密集块通过密集连接的方式提取更加丰富的局部特征, 且前一个残差密集块的状态可以传递到后续残差密集块的所有层, 从而形成连续的特征存储机制. 此外, 残差密集块的局部特征融合模块能够自适应地从当前的特征存储中学习更有效细节特征.

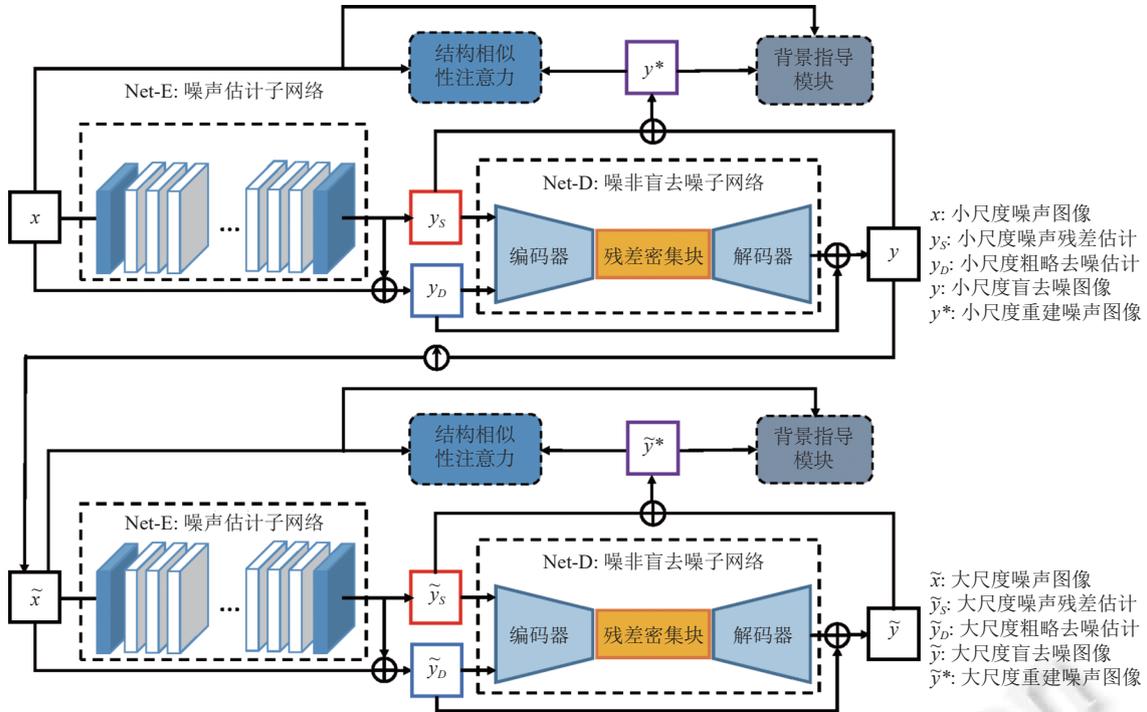


图 1 基于自监督约束的真实图像盲去噪网络结构示意图

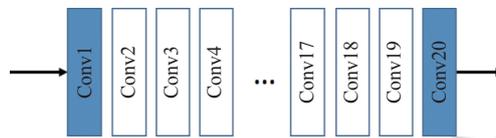


图 2 噪声估计子网络 Net-E 结构示意图

表 1 Net-E 的网络结构及参数设置

网络层	Conv1	Conv2-Conv19	Conv20
通道数	3	64	3
卷积核尺寸	3×3	3×3	3×3
步长	1	1	1
边缘填充	1	1	1

关于本文方法的去噪过程, 首先将观测到的小尺度真实噪声图像  $x \in X^n$  输入 Net-E 中估计噪声强度  $y_s = \mathcal{F}_E(x; W_E)$ , 其中  $\mathcal{F}_E(\cdot)$  表示 Net-E 的网络模型,  $W_E$  表示 Net-E 的模型参数. 生成的  $y_s$  是噪声残差图, 其尺寸与输入  $x$  相同, 因此通过元素间的“ $\oplus$ ”操作得到粗略的去噪结果  $y_D = x \oplus y_s$ . 接下来, 将 Net-E 的输出  $y_s$  和  $y_D$  在特征通道维度上进行拼接后输入到 Net-D 中得到最终的小尺度图像盲去噪结果  $y = \mathcal{F}_D(y_s, y_D; W_D)$ , 这里的  $\mathcal{F}_D(\cdot)$  指的是

Net-D 的网络模型,  $W_D$  指的是 Net-D 的模型参数. 为了使得小尺度网络分支的输出能够指导大尺度网络得到更好的去噪结果, 需要将  $y$  进行上采样后与观测到的大尺度噪声图像  $\hat{x} \in \hat{X}^n$  拼接, 再输入到大尺度网络分支中. 因此, 大尺度分支中 Net-E 估计的噪声强度图为  $\tilde{y}_S = \mathcal{F}_E(\hat{x}, y \uparrow; W_E)$ , 这里的“ $\uparrow$ ”表示图像的上采样操作, 本文使用的是双线性插值法. 然后将  $(\tilde{y}_S, \tilde{y}_D)$  输入到 Net-D 中得到图像的最终去噪结果  $\tilde{y} = \mathcal{F}_D(\tilde{y}_S, \tilde{y}_D; W_D)$ , 其中  $\tilde{y}_D = \hat{x} \oplus \tilde{y}_S$ . 本文提出的自监督双尺度图像盲去噪网络的设计主要采用残差学习 (residual learning) 的思想, 通过设计两个尺度的残差网络, 充分挖掘尺度间的信息差异, 帮助实现更好的去噪效果. 以小尺度分支网络为例, 首先将噪声图像  $x$  输入 Net-E 中得到噪声残差图  $y_S$ , 然后通过恒等映射将  $x$  和  $y_S$  进行元素间相加得到一个粗略的去噪结果  $y_D$ . 对于 Net-D, 首先使用 Net-E 的输出  $(y_S, y_D)$  作为 Net-D 的输入, 通过网络学习相应的残差映射  $\mathcal{R}(y_S, y_D; W_D)$ , 然后通过网络的跳跃连接 (skip connection) 将  $y_D$  恒等映射到 Net-D 末端得到最终的去噪结果  $y = \mathcal{R}(y_S, y_D; W_D) + y_D$ . 对于大尺度分支, 其去噪过程与小尺度分支基本相同, 唯一的差别在于大尺度分支的输入是  $(\hat{x}, y \uparrow)$ , 它融合了小尺度分支的去噪结果. 这样做是因为小尺度的噪声图像相对于大尺度的噪声图像而言, 噪声强度相对降低, 峰值信噪比相对提升, 这表明小尺度图像遭受的噪声退化程度更小一些, 可利用的信息相对更多一些, 其去噪结果能够为大尺度图像提供额外的有用信息, 有助于引导大尺度去噪网络复原出高质量的清晰图像. 本文将在第 4.3.2 节给出关于使用双尺度去噪网络结构优于单尺度去噪网络结构的相关证明.

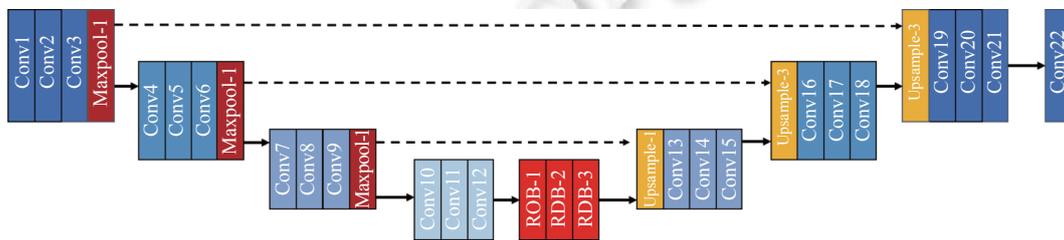


图3 图像非盲去噪子网络 Net-E 结构示意图

表2 Net-D 的网络结构及参数设置

结构	网络层	Conv1-3	Maxpool-1	Conv4-6	Maxpool-2	Conv7-9	Maxpool-3	Conv10-12
编码器	通道数	64	64	128	128	256	256	512
	卷积核	3×3	2×2	3×3	2×2	3×3	2×2	3×3
	步长	1	2	1	2	1	2	1
	边缘填充	1	0	1	0	1	0	1
结构	网络层	Upsample-1	Conv13-15	Upsample-2	Conv16-18	Upsample-3	Conv19-21	Conv22
解码器	通道数	512	256	256	128	128	64	3
	卷积核	—	3×3	—	3×3	—	3×3	3×3
	步长	—	1	—	1	—	1	1
	边缘填充	—	1	—	1	—	1	1

本文方法从输入噪声图像本身出发构造自监督约束, 并以端到端的方式训练网络. 由于缺乏清晰图像的监督约束, 当前的自监督去噪框架往往会遭遇约束不足的问题, 而且由于真实噪声在分布、强度等方面的差异, 很难学习到泛化性能较强的去噪模型. 为了解决这个问题, 本文在第 3.2-3.4 节中提出基于噪声图像内在属性的自监督约束模块.

### 3.2 边缘保持自监督约束

在缺乏清晰图像的监督下, 本文仅使用单幅噪声图像作为网络的输入并构造自监督约束. 全变分 (total variational, TV) 正则化作为一种图像先验能够有效地建模图像梯度的分布, 且梯度越小, 图像就越平滑, 噪声也就随之减弱. 因此, 为了训练网络实现真实噪声图像的自监督盲去噪, 本文首先引入 TV 作为损失函数来约束网络进

行去噪. 由于本文提出的双尺度去噪网络的两个分支共享相同的网络参数, 下面以小尺度分支为例进行阐述.

令  $\{x_i\}_{i=1}^{N_S}$  为小尺度网络的噪声输入,  $\{y_i\}_{i=1}^{N_S}$  为对应的去噪结果, 这里的  $N_S$  表示训练集真实噪声图像的数量. 首先, 根据图像梯度定义损失函数:

$$\mathcal{L}_{TV} = \frac{1}{N_S} \sum_{i=1}^{N_S} (\|\partial_h y_i\|_2 + \|\partial_v y_i\|_2) \quad (1)$$

其中,  $\partial_h$  和  $\partial_v$  分别表示水平和垂直方向的梯度算子. 该损失函数使用网络输出本身的梯度作为误差经过反向传播来约束网络, 容易导致去噪结果过于平滑, 网络训练不稳定的问题. 为了给网络提供一个稳定的约束, 本文从噪声图像的统计特性出发, 根据输入和输出图像之间的均方误差 (mean square error, MSE) 定义一个像素级的内容损失函数:

$$\mathcal{L}_{\text{pixel}} = \frac{1}{N_S} \sum_{i=1}^{N_S} \|y_i - x_i\|^2 \quad (2)$$

$\mathcal{L}_{\text{pixel}}$  以输入的噪声图像作为监督, 不仅有助于稳定训练过程, 而且有利于避免网络的输出过于平滑.

为了在去除噪声的同时, 尽可能地保留图像的边缘细节, 我们提出边缘保持平滑 (edge-preserving smooth, EPS) 约束. 该约束在公式 (1) 的基础上将噪声图像的梯度也考虑在内. 并将噪声图像作为一个潜在的标签来构造损失函数:

$$\mathcal{L}_{TV\_EPS} = \frac{\lambda}{N_S} \sum_{i=1}^{N_S} \left( \frac{\|\partial_h y_i\|_2}{\|\partial_h x_i\|^\alpha + \epsilon} + \frac{\|\partial_v y_i\|_2}{\|\partial_v x_i\|^\alpha + \epsilon} \right) \quad (3)$$

这里的  $\lambda$  是控制平滑的参数,  $\alpha$  是控制边缘细节的参数,  $\epsilon$  是一个极小正常量. 本文使用  $\mathcal{L}_{EPS}$  取代  $\mathcal{L}_{TV}$  来约束网络, 因此该去噪网络的优化目标初步定义为公式 (2) 和公式 (3) 两个正则项的加权组合:

$$\arg \min_{W_E, W_D} \left( \frac{1}{N_S} \sum_{i=1}^{N_S} \left( \|y_i - x_i\|^2 + \lambda \left( \frac{\|\partial_h y_i\|_2}{\|\partial_h x_i\|^\alpha + \epsilon} + \frac{\|\partial_v y_i\|_2}{\|\partial_v x_i\|^\alpha + \epsilon} \right) \right) \right) \quad (4)$$

其中,  $W_E$  和  $W_D$  是噪声估计子网络和图像非盲去噪子网络的参数, 平滑参数  $\lambda$  用于平衡这两个正则项之间的关系,  $\lambda$  取值越大, 网络输出就越平滑, 其取值范围是 (0, 1). 细节保持参数  $\alpha$  控制网络对输入噪声图像的梯度的敏感度,  $\alpha$  取值越大, 网络输出的细节丢失越严重, 其取值范围是 [1.2, 2.0].

根据公式 (4), 网络在去除噪声的同时能够保持部分图像细节, 本文的去噪任务由此演变为图像去噪和细节保持的折中. 关于参数  $\lambda$  和  $\alpha$  取值的分析, 本文将在第 4.3 节详细描述.

### 3.3 背景一致性自监督约束

正如前面提到的, 由于缺乏清晰图像的监督约束, 网络往往遭遇约束不足的困境. 第 3.2 节提出的边缘保持自监督约束直接利用输入的噪声图像作为标签, 网络容易遭受噪声的干扰, 导致网络的输出与输入存在内容上的偏差, 丢失一些重要的信息. 为了确保图像背景内容的一致性, 并复原重要的图像边缘细节, 本文通过构建一个多尺度的背景指导模块 (background guidance module, BGM), 从输入的噪声图像上探索基于背景内容一致性的自监督约束, 从而辅助网络实现更好的去噪效果.

在图 1 中, 小尺度分支 Net-D 输出的去噪图像  $y$  与 Net-E 估计的噪声图  $y_S$  进行元素间相加得到的噪声输出  $y^*$  与输入的噪声图像  $x$  包含相同的背景内容, 因此利用两幅噪声图像的背景构造自监督约束来指导网络训练是合理的, 图 4 为背景指导模块的具体操作过程.

为了消除噪声的干扰, 得到无噪声的背景图像, 本文采用多尺度高斯模糊核对噪声图像  $x$  和  $y^*$  进行滤波处理. 然后在梯度域中计算不同尺度高斯模糊图像对的偏差, 以此来衡量图像边缘细节等重要信息. 从图 4 可以观察到, 随着模糊核尺度  $\sigma$  增大, 对应的两幅高斯模糊图像的内容越来越接近, 图像的平均梯度误差也越来越小. 针对这一现象, 本节提出背景指导损失函数:

$$\mathcal{L}_{BG} = \sum_{\sigma=3,5,9} \lambda_\sigma \|\nabla B_\sigma(x) - \nabla B_\sigma(y^*)\|_1 \quad (5)$$

其中,  $\nabla$  表示梯度计算,  $B_\sigma$  表示尺度为  $\sigma$  高斯模糊操作,  $\lambda_\sigma$  是权重参数, 用于平衡不同尺度高斯模糊图像的梯度误差. 由于不同尺度的模糊图像所包含的背景信息有所差别, 为了能够利用更丰富的背景信息来约束网络, 该损失函数涉及  $\sigma = 3, 5, 9$  这 3 种尺度的高斯模糊图像的梯度误差, 根据实验结论将对应的权重  $\lambda_\sigma$  分别设为  $[0.01, 0.1, 1.0]$ . 本文将在第 4.3.2 节给出背景指导模块有效性证明.

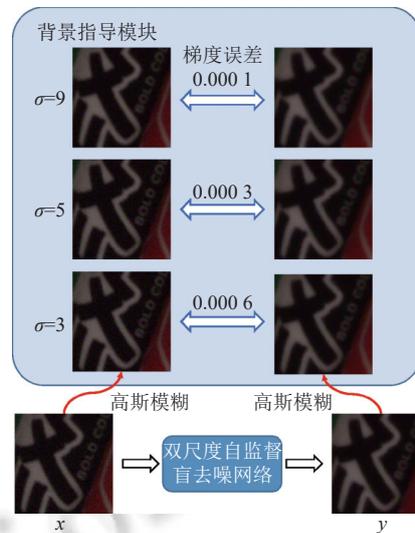


图 4 背景指导模块示意图

### 3.4 结构相似性注意力机制

图像中一些细微的结构差异对损失函数的影响较小, 经过反向传播后对网络参数产生的影响甚至可以忽略不计, 这就导致网络无法复原出这些微小的结构细节. 为了能将图像之间这些细微的结构差异利用起来, 本文提出一种结构相似性注意力机制 (SSIM attention mechanism, SAM), 并根据此机制可构建一种新的自监督约束, 其原理如图 5 所示.

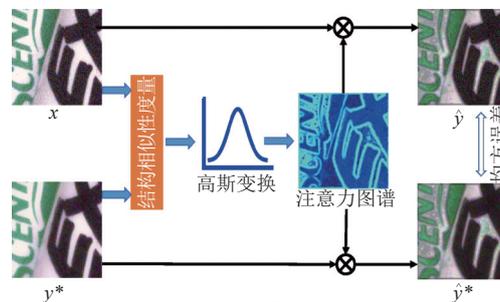


图 5 结构相似性注意力机制示意图

以小尺度网络分支为例. 首先, 利用结构相似度 (structural similarity index, SSIM) 计算输入噪声图像  $x$  和输出噪声图像  $y^*$  的结构相似性图谱:

$$M_{\text{ssim}} = \mathcal{S}(x, y^*) \tag{6}$$

其中,  $\mathcal{S}(\cdot)$  指的是结构相似性度量函数, 它利用邻域的像素值建模了  $x$  和  $y^*$  中每个对应局部区域的结构相似度, 使得最终得到  $M_{\text{ssim}}$  比绝对差异更加稳定.

由于网络对结构相似度较小的图像信息敏感度较低, 往往会导致网络忽略掉一些微小的结构细节. 鉴于此, 本文使用高斯函数对  $M_{\text{ssim}}$  进行变换, 得到其对应的结构相似性注意力图谱:

$$M_{\text{att}} = \mathcal{G}_{\eta, \mu, \delta}(M_{\text{ssim}}) \quad (7)$$

其中,  $\mathcal{G}$  指的是高斯函数,  $\eta$  是一个非负常量, 表示高斯曲线的峰值,  $\mu$  和  $\delta$  是高斯函数的均值和标准差, 分别表示其曲线峰值的位置和宽度. 对于  $M_{\text{ssim}}$  中的每一个像素点  $m_{\text{ssim}}$ , 其对应的结构相似性注意力的值计算为:

$$m_{\text{att}} = \eta \cdot \exp\left(-\frac{(m_{\text{ssim}} - \mu)^2}{\delta^2}\right) \quad (8)$$

其中,  $\eta$  的取值设为 2,  $\mu$  的取值设为 -1,  $\delta$  的取值根据网络迭代的次数  $step$  设定为:

$$\delta = \tau \cdot step + \beta \quad (9)$$

其中,  $\tau$  表示  $\delta$  的增长率,  $\beta$  表示  $\delta$  的初始值, 前者被设置为 1, 后者被设为 0. 接下来, 将  $x$  和  $y^*$  分别与  $M_{\text{att}}$  进行元素间的“ $\otimes$ ”操作得到新的输入噪声图像  $X$  和新的输出噪声图像  $Y^*$ :

$$\begin{cases} X = x \otimes M_{\text{att}} \\ Y^* = y^* \otimes M_{\text{att}} \end{cases} \quad (10)$$

最后, 利用所生成的两幅图像构造出一种新的自监督约束:

$$\mathcal{L}_{\text{SA}} = \frac{1}{N_S} \sum_{i=1}^{N_S} \|Y_i^* - X_i\|^2 \quad (11)$$

通过使用这种基于结构相似性注意力机制的自监督约束, 经过反向传播后的网络开始关注图像中一些微小的结构细节, 从而引导网络复原出结构细节更加清晰的去噪图像. 需要说明的是, 大尺度网络分支同小尺度网络分支均执行上述操作. 本文将在第 4.3.2 节给出结构相似性注意力机制有效性的验证.

### 3.5 损失函数

下面根据本文第 3.2–3.4 节所提出的自监督约束来构造网络模型的损失函数. 鉴于本文的网络框架的两个分支均采用相同的自监督约束进行训练, 因此将小尺度分支的损失函数表示为  $\mathcal{L}_c^s$ , 对应的大尺度分支的损失函数表示为  $\mathcal{L}_c^l$ ,  $c$  表示不同的损失函数.

我们首先将第 3.2 节中设计的边缘保持自监督约束作为网络的基本约束, 其损失函数定义如下:

$$\begin{cases} \mathcal{L}_{\text{EPS}}^s = \mathcal{L}_{\text{pixel}}^s + \mathcal{L}_{\text{TV\_EPS}}^s \\ \mathcal{L}_{\text{EPS}}^l = \mathcal{L}_{\text{pixel}}^l + \mathcal{L}_{\text{TV\_EPS}}^l \end{cases} \quad (12)$$

为了构造输入噪声图像  $x(\tilde{x})$  与输出噪声图像  $y^*(\tilde{y}^*)$  之间的误差, 进而对噪声估计子网络预测的噪声图产生约束, 本文增加一个额外的内容重建损失函数:

$$\begin{cases} \mathcal{L}_{\text{rec}}^s = \frac{1}{N_S} \sum_{i=1}^{N_S} \|y_i^* - x_i\|^2 \\ \mathcal{L}_{\text{rec}}^l = \frac{1}{N_S} \sum_{i=1}^{N_S} \|\tilde{y}_i^* - \tilde{x}_i\|^2 \end{cases} \quad (13)$$

其次, 根据第 3.3 节所设计的背景指导模块构建基于背景一致性的自监督约束, 其定义对应的损失函数如下:

$$\begin{cases} \mathcal{L}_{\text{BG}}^s = \sum_{\sigma=3,5,9} \lambda_{\sigma} \|\nabla B_{\sigma}(x) - \nabla B_{\sigma}(y^*)\|_1 \\ \mathcal{L}_{\text{BG}}^l = \sum_{\sigma=3,5,9} \lambda_{\sigma} \|\nabla B_{\sigma}(\tilde{x}) - \nabla B_{\sigma}(\tilde{y}^*)\|_1 \end{cases} \quad (14)$$

然后, 根据第 3.4 节提出的结构相似性注意力机制构建一种自监督约束, 其损失函数定义如下:

$$\begin{cases} \mathcal{L}_{\text{SA}}^s = \frac{1}{N_S} \sum_{i=1}^{N_S} \|Y_i^* - X_i\|^2 \\ \mathcal{L}_{\text{SA}}^l = \frac{1}{N_S} \sum_{i=1}^{N_S} \|\tilde{Y}_i^* - \tilde{X}_i\|^2 \end{cases} \quad (15)$$

最后, 根据公式 (12)–公式 (15) 得到该网络框架两个尺度分支各自总的损失函数如下:

$$\begin{cases} \mathcal{L}_{\text{total}}^s = \omega_1 \mathcal{L}_{\text{EPS}}^s + \omega_2 \mathcal{L}_{\text{rec}}^s + \omega_3 \mathcal{L}_{\text{BG}}^s + \omega_4 \mathcal{L}_{\text{SA}}^s \\ \mathcal{L}_{\text{total}}^l = \omega_1 \mathcal{L}_{\text{EPS}}^l + \omega_2 \mathcal{L}_{\text{rec}}^l + \omega_3 \mathcal{L}_{\text{BG}}^l + \omega_4 \mathcal{L}_{\text{SA}}^l \end{cases} \quad (16)$$

其中,  $\omega_1$ - $\omega_4$  表示各个损失函数的权重, 在本文所有的实验中我们将这 4 个权重均设为 1.0.

综上, 本文提出的基于双尺度自监督的真实图像盲去噪网络的总的损失函数为:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{total}}^s + \mathcal{L}_{\text{total}}^l \quad (17)$$

## 4 实验及结果分析

### 4.1 数据集

本文使用了 3 个公开的真实噪声数据集 SIDD<sup>[8]</sup>, DND<sup>[7]</sup>和 Nam<sup>[30]</sup>, 具体描述如下.

(1) SIDD 数据集: 包含由 5 台不同型号的移动手机摄像机在 10 个场景下拍摄的 200 个实例的噪声图像. 其中 160 个实例图像用作训练集, 其余的 40 个实例图像用作测试集, 其包含 1280 个尺寸为 256×256 的裁剪图像块.

(2) DND 数据集: 包含由 4 台不同型号的摄像机在 50 个场景下拍摄的 50 个实例的噪声图像, 其基准测试集提供了来自这 50 幅图像的 1000 个尺寸为 521×512 的裁剪块.

(3) Nam 数据集: 包含 3 台不同型号的摄像机在 11 个室内场景拍摄的 15 个实例的噪声图像.

本文实验使用 SIDD 中的 160 幅噪声图像作为训练集, 使用其余的 40 幅图像和 DND 以及 Nam 作为基准测试集.

### 4.2 实验设置

#### 4.2.1 网络参数设置

本文提出的自监督真实图像盲去噪网络包含两个尺度的网络分支, 且它们的网络共享相同的参数. 其中, 噪声估计子网络是一个普通的全卷积网络, 包含 20 个卷积层, 且每个卷积层的滤波器大小均为 3×3. 图像非盲去噪子网络是一个中间嵌入 3 个密集残差块的 U 型网络框架, 其编码器和解码器共包含 24 个卷积层, 每个卷积层的滤波器大小也均为 3×3. 每一个密集残差块有 4 个卷积层, 其中包括 3 个滤波器大小为 3×3 的卷积层和一个 1×1 的卷积层.

#### 4.2.2 网络训练设置

本文提出的图像盲去噪网络以端到端的方式进行训练. 在网络的训练过程中, 本文将图像的批量大小设为 16, 输入网络的初始滤波器数量为 64, 图像块尺寸设为 128×128, 网络的初始学习率设为 2E-5, 且每训练 20 轮学习率降为原来的一半, 且该网络采用 ADAM 优化器进行参数优化. 为了稳定训练过程, 加速收敛进程, 本文先单独训练 Net-E 和 Net-D 网络, 然后再联合训练这两个网络完成图像去噪. 该网络共训练 200 轮. 关于公式 (4) 中  $\lambda$  和  $\alpha$  的取值, 经过多次实验, 最终将  $\lambda$  设为 0.08,  $\alpha$  设为 1.2.

本文算法的实验设备及环境为 CPU Intel(R) Core(TM) i7-7700K@4.20 GHz 和 GPU Nvidia GTX 1080Ti 以及 64 位 Ubuntu 18.04 LTS 操作系统和 CUDA Toolkit 9.0 工具.

### 4.3 实验结果分析

#### 4.3.1 与现有方法对比

针对本文提出的基于双尺度自监督的真实图像盲去噪算法, 本节给出该算法在 SIDD, DND, Nam 真实基准数据集上与现有方法的对比实验结果. 这些现有方法包括 BM3D<sup>[1]</sup>、MLP<sup>[12]</sup>、NLM<sup>[10]</sup>、EPLL<sup>[5]</sup>、KSVD<sup>[3]</sup>、FOE<sup>[4]</sup> 等传统图像去噪方法和 TNRD<sup>[31]</sup>、DnCNN-B<sup>[14]</sup> 等深度学习去噪方法. 对于一些传统的非盲去噪方法, 本文借助文献 [32] 提出的方法来估计其噪声强度. 我们使用峰值信噪比 (peak signal-to-noise ratio, PSNR) 和结构相似度 (SSIM) 作为评价标准来衡量去噪结果的优劣.

表 3 所示为本文提出的去噪方法与现有方法在 SIDD 基准数据集上的量化对比结果. 为了保证对比结果的公平性, 我们将去噪结果统一上传到 SIDD 的线上提交系统<sup>[33]</sup>, 由官方测量其平均的 PSNR 和 SSIM 值. 根据表 3 中的量化结果可知, 本文提出的双尺度自监督图像盲去噪算法在 SIDD 真实噪声基准数据集上的 PSNR 值达到了

30.40 dB, 比有监督的深度去噪方法 TNRD 和 DnCNN-B 分别提高 5.67 dB 和 6.74 dB, 比较为先进的深度学习盲去噪方法 N2V 提高 2.72 dB, 表现出良好的去噪性能. 图 6 展示了本文方法与现存方法在视觉上的对比结果.

表 3 不同去噪方法在 SIDD 数据集上的量化结果

对比方法	MLP	BM3D	KSVD	EPLL	TNRD	DnCNN-B	WNNM	N2V	本文方法
PSNR (dB)	24.71	25.65	26.88	27.11	24.73	23.66	25.78	27.68	<b>30.40</b>
SSIM	0.641	0.685	0.842	0.870	0.643	0.583	0.809	0.668	0.769
盲/非盲去噪	非盲去噪	非盲去噪	非盲去噪	非盲去噪	非盲去噪	盲去噪	非盲去噪	盲去噪	盲去噪

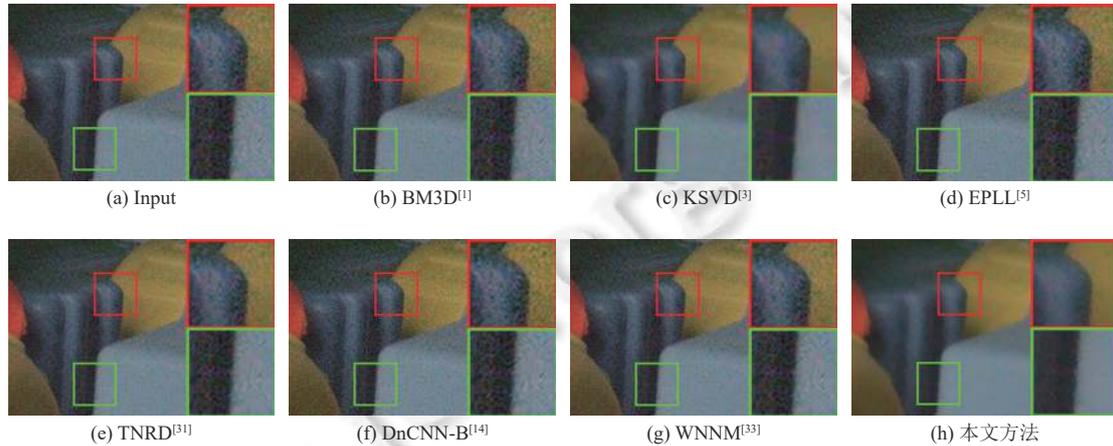


图 6 不同图像去噪方法在 SIDD 真实噪声数据集上的视觉对比结果

表 4 所示为本文提出的去噪方法与现有方法在 DND 基准数据集上的量化对比结果. 与 SIDD 类似, DND 也提供了相应的线上提交系统用于公平地测量各种对比方法的量化结果<sup>[7]</sup>. 根据表 4 的量化结果可知, 本文方法在 DND 上的 SSIM 指标明显优于对比方法, 其中比 DnCNN-B 和 TNRD 这两个有监督的深度去噪方法提高 0.082 和 0.04, PSNR 值分别提高 1.90 dB 和 0.68 dB. 由于提出 N2V 方法的作者还未将结果数据上传至 DND 数据集网页获取量化结果, 因此出于公平我们无法在该数据集上与本文方法进行对比. 图 7 展示了本文方法和对比方法在 DND 上的视觉对比实例, 包括局部放大结果. 可以看出本文方法的去噪结果比其他方法的噪声残留更少, 边缘更清晰.

除了 SIDD 和 DND 这两个数据集, 本文还使用噪声强度相对较弱的 Nam 基准数据集来评估本文提出的去噪算法. 表 5 所示为本文方法与现有方法在 Nam 数据集上的量化对比结果. 可以看出本文方法在低噪声图像上也同样适用. 图 8 展示了对应的视觉对比结果. 可以观察到本文提出的方法能够在去除噪声的同时保留更丰富的图像细节.

表 4 不同图像去噪方法在 DND 数据集上的量化结果

对比方法	PSNR (dB)	SSIM	盲/非盲去噪
MLP <sup>[12]</sup>	34.23	0.833	非盲去噪
BM3D <sup>[1]</sup>	34.51	0.851	非盲去噪
NCSR <sup>[2]</sup>	34.05	0.835	非盲去噪
EPLL <sup>[5]</sup>	33.51	0.824	非盲去噪
TNRD <sup>[31]</sup>	33.65	0.831	非盲去噪
DnCNN-B <sup>[14]</sup>	32.43	0.790	盲去噪
本文方法	34.33	<b>0.872</b>	盲去噪

表 5 不同图像去噪方法在 Nam 数据集上的量化结果

对比方法	PSNR (dB)	SSIM	盲/非盲去噪
BM3D <sup>[1]</sup>	35.19	0.906	非盲去噪
KSVD <sup>[3]</sup>	36.41	0.946	非盲去噪
EPLL <sup>[5]</sup>	33.42	0.907	非盲去噪
TNRD <sup>[31]</sup>	35.35	0.959	非盲去噪
DnCNN-B <sup>[14]</sup>	33.86	0.864	盲去噪
WNNM <sup>[34]</sup>	33.45	0.902	非盲去噪
N2V <sup>[25]</sup>	35.27	0.915	盲去噪
本文方法	<b>36.45</b>	<b>0.951</b>	盲去噪

根据上述 3 个真实噪声基准数据集的对比结果可以证明, 本文方法不论在视觉上还是量化指标上其去噪性能均优于现存大多数无监督方法及部分全监督深度学习去噪方法, 且对于不同强度噪声数据集均表现出良好的泛化性能. 本文方法的焦点不在于网络结构的设计, 而是如何挖掘真实噪声图像本身的一些特征属性来构造伪监督标签, 在没有使用任何外部监督信号的条件下实现了自监督去噪. 因此, 本文方法是一种针对真实噪声图像而提出的纯粹的自监督去噪方法, 它从一个全新的角度探索了真实场景的图像去噪问题.

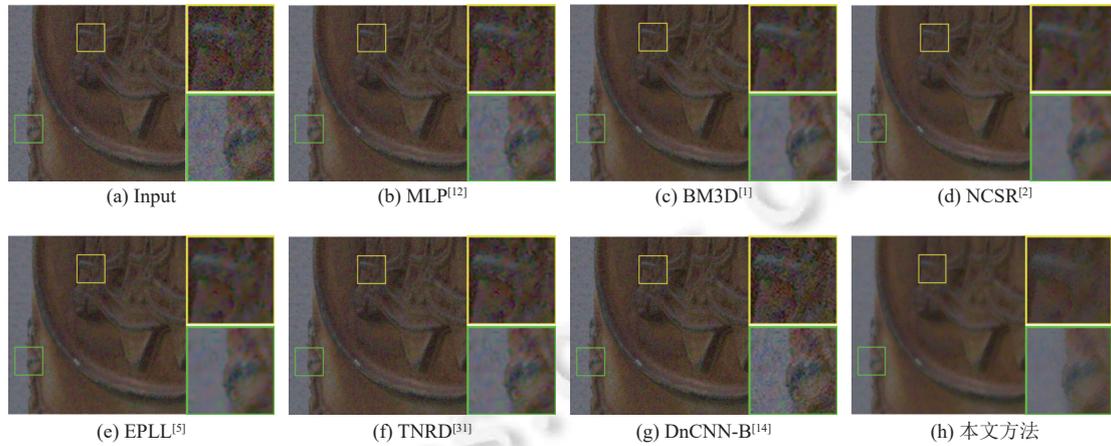


图 7 不同图像去噪方法在 DND 真实噪声数据集上的视觉对比结果

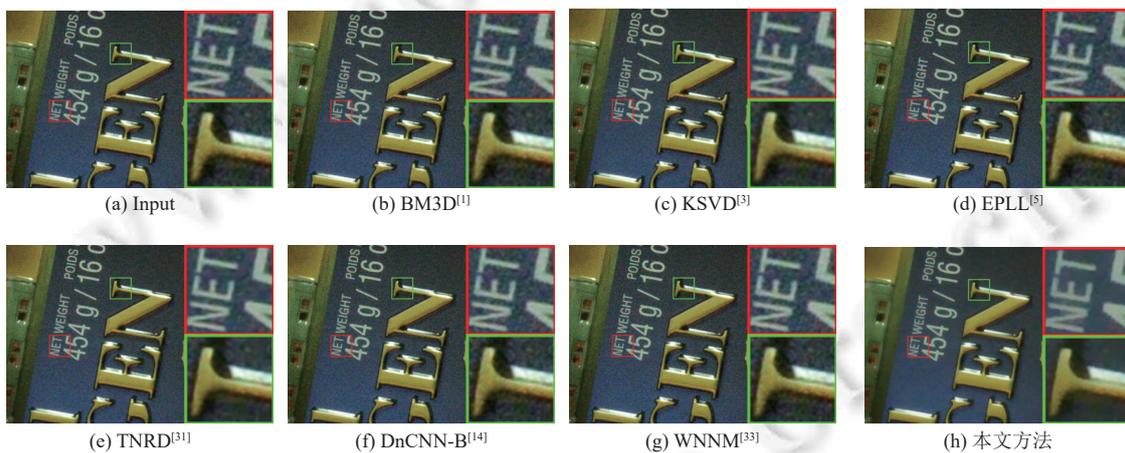


图 8 不同图像去噪方法在 Nam 真实噪声数据集上的视觉对比结果

#### 4.3.2 双尺度去噪网络与单尺度去噪网络的性能比较

本文首先对单尺度去噪网络进行了训练, 网络的输入是大尺度的噪声图像, 记录训练过程中每一轮的验证结果, 如图 9 中的橙色曲线所示. 然后, 本文提出以小尺度噪声图像作为输入的另一网络分支并使用小尺度的去噪图像经过两倍上采样之后给大尺度去噪分支提供一些辅助的监督信息, 以此来训练双尺度的去噪网络. 图 9 中的蓝色曲线表示双尺度去噪网络训练过程的收敛趋势, 根据验证结果可以看出双尺度去噪网络相比于单尺度去噪网络在性能上确实有一定的提升, 而且两个尺度网络共享参数. 因此, 本文提出的双尺度去噪网络不仅在性能上优于单尺度网络, 而且不会占用额外的存储空间.

#### 4.3.3 BGM 和 SAM 的有效性验证

本文在基于自监督的双尺度盲去噪网络的基础上增加 BGM 模块和 SAM 机制用来构建自监督约束以便实现更好的去噪性能. 下面我们在 SIDD 上分别验证这两个模块的有效性. 表 6 展示了使用不同模块的自监督

约束的网络模型对 SIDD 基准测试集的量化结果. 其中, 使用输入噪声图像与输出重建噪声图像之间的自监督损失函数  $\mathcal{L}_{rec}$  使得平均 PSNR 量化结果提高 0.10 dB. 在此基础上通过增加 BGM 模块构建新的自监督损失函数  $\mathcal{L}_{BG}$  使得平均 PSNR 值又提高 0.03 dB, 可以证明增加 BGM 模块辅助本文提出的双尺度盲去噪网络进行训练, 可以提升网络的去噪性能. 此外, 本文在增加 BGM 的基础上又提出了 SAM 机制, 用于引导网络关注图像中微小的结构细节. 从表 6 可知, 增加 SAM 模块使得网络的去噪性能得到了大幅度提升, 其平均的 PSNR 量化指标提升了 0.22 dB.

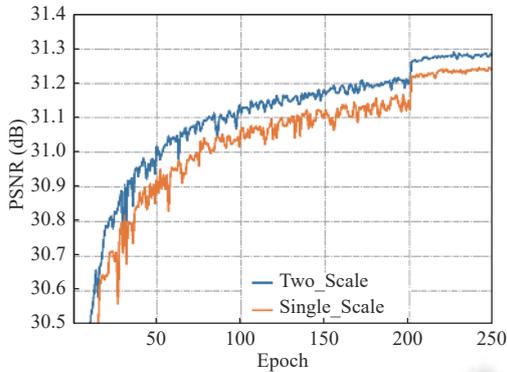


图 9 单尺度网络与双尺度网络在训练过程中的验证结果曲线图

表 6 不同自监督约束的网络在 SIDD 上的量化结果

$\mathcal{L}_{EPS}$	$\mathcal{L}_{rec}$	$\mathcal{L}_{BG}$	$\mathcal{L}_{SA}$	PSNR (dB)	SSIM
√	×	×	×	30.05	0.753
√	√	×	×	30.15	0.759
√	√	√	×	30.18	0.760
√	√	√	√	<b>30.40</b>	<b>0.769</b>

#### 4.3.4 平滑参数对网络去噪性能的影响

在边缘保持自监督约束 (公式 (12)) 中, 涉及到两个参数  $\lambda$  和  $\alpha$  用于控制去噪图像的平滑程度. 其中  $\lambda$  表示  $\mathcal{L}_{EPS}$  的平滑权重,  $\alpha$  表示网络对输入噪声图像的梯度敏感度. 接下来本节讨论不同  $\lambda$  和  $\alpha$  取值对网络去噪性能的影响. 图 10 展示了一幅真实噪声图像及其在不同  $\lambda$  和  $\alpha$  取值下网络的去噪效果的局部放大对比图. 可以观察到, 当  $\lambda=1.2$  时,  $\alpha$  取值越大, 图像的细节丢失越严重, 即  $\lambda=1.2, \alpha=0.05$  对应的去噪图像的细节保持较为完整; 当  $\alpha=0.08$  时,  $\lambda$  取值越大, 图像越平滑, 即  $\lambda=1.0, \alpha=0.08$  对应的去噪图像细节相对清晰, 但是存在轻微的噪声残余. 经过大量实验, 本文最终选取  $\lambda=0.08$  和  $\alpha=1.2$  作为最佳的平滑参数来实现图像去噪和细节保持的折中. 此外, 针对不同的真实噪声数据集, 用户可以通过调节平滑参数  $\lambda$  和  $\alpha$  来使得网络适应于该数据集的噪声特性, 以便复原出细节更加清晰的图像.

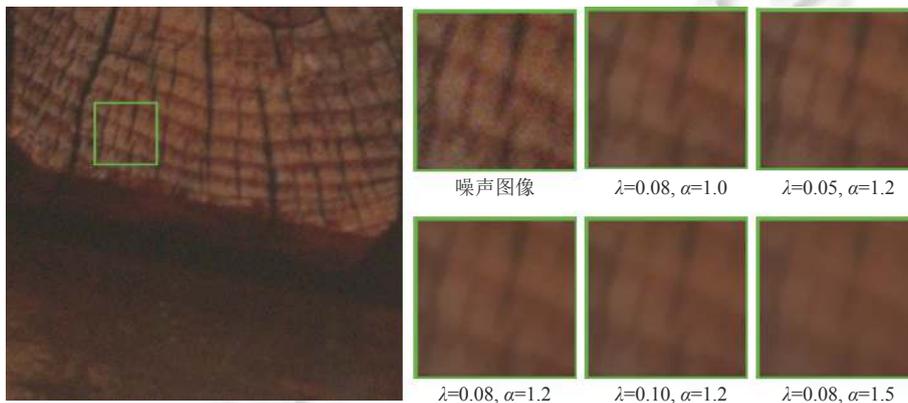


图 10 不同平滑参数  $\lambda$  和  $\alpha$  对模型的去噪效果

#### 4.3.5 生成对抗式学习对网络性能的影响

本文致力于探索适用于真实噪声图像盲去噪的自监督约束, 所提出的方法旨在训练出具有良好的去除真实噪

声能力的生成式网络. 为了探索生成对抗式学习对自监督算法性能的影响, 本文引入对抗性损失函数  $\mathcal{L}_{adv}$  进一步约束网络, 使用输出的重建噪声图像  $y^*$  ( $\tilde{y}^*$ ) 和输入的真实噪声图像  $x$  ( $\tilde{x}$ ) 来计算对抗损失:

$$\mathcal{L}_{adv} = \mathbf{E}[\log(D(x))] + \mathbf{E}[\log(1 - D(y^*))] \quad (18)$$

在这里  $D(\cdot)$  指的是判别器, 本文提出的去噪网络作为生成器. 我们不仅使用量化指标 PSNR 和 SSIM 来评估去噪结果, 还使用了非参考图像质量评估方法 NIQE<sup>[35]</sup> 和 BRISQUE<sup>[36]</sup> 从图像视觉质量方面进行了评估, 其中 NIQE 和 BRISQUE 得分越低说明图像的视觉效果越好. 实验结果如表 7 所示, 相比于未引入对抗损失的方法, 采用生成对抗学习范式的方法在 PSNR 量化指标上提高 0.18 dB, 在视觉评价指标 BRISQUE 上得分降低约 3.5, 但是比 NIQE 的得分略有上升. 综合实验结果表明采用生成对抗学习范式可以进一步提高本文方法对于真实图像的盲去噪效果.

表 7 生成对抗学习范式在量化指标和视觉质量上对网络性能的影响

方法	PSNR (dB)	SSIM	NIQE	BRISQUE
本文方法	30.40	0.769	7.215	33.546
对抗损失	30.58	0.773	7.485	30.031

#### 4.3.6 特征图谱可视化

为了更好地解释自监督约束所捕获的特征表达的有效性, 本节对训练过程中网络模型学习到的特征图谱进行了可视化. 图 11 展示的是由图像非盲去噪子网络的第 21 个卷积层 (详见图 3) 输出的某一特征图谱. 通过可视化结果可以观察到, 随着训练次数 (epoch) 的增加, 网络学习到的高频特征逐渐趋于细化, 特别是平滑区域的噪声分布, 且图像的边缘特征也逐渐由模糊变得清晰, 验证了本文所提出的自监督算法的有效性, 有助于在缺乏清晰图像监督的条件下使得网络复原出视觉效果良好的去噪图像.

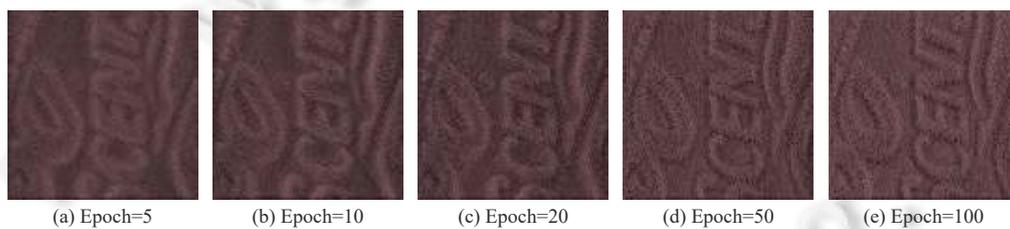


图 11 网络训练过程中的特征图谱可视化

#### 4.3.7 算法效率分析

算法的效率也是评价算法性能的重要因素之一. 表 8 给出了本文算法与现有相关的去噪方法的运行时间比较结果, 表中所有方法的运行时间均通过测试尺寸为  $512 \times 512$  的彩色图片得到. 实验结果表明本文提出的算法需要较少的运行时间.

表 8 不同图像去噪方法的运行时间对比

对比方法	NLM <sup>[10]</sup>	BM3D <sup>[11]</sup>	KSVD <sup>[3]</sup>	EPLL <sup>[5]</sup>	TNRD <sup>[31]</sup>	DnCNN-B <sup>[14]</sup>	RIDNet <sup>[9]</sup>	CBDNet <sup>[6]</sup>	本文方法
CPU/GPU	CPU	CPU	CPU	CPU	CPU	GPU	GPU	GPU	GPU
工具	Matlab	Matlab	Matlab	Matlab	Matlab	Python	Python	Python	Python
运行时间 (s)	621.9	4.9	422.7	114.93	3.64	0.058	0.190	0.400	0.057

#### 4.3.8 本文方法的局限性

根据实验结果可以发现本文提出的基于自监督约束的双尺度真实图像盲去噪算法对于暗光条件下真实图像的去噪效果不理想. 如图 12 所示, 图 12(a) 是一幅在暗光环境下拍摄的真实噪声图像, 其整体灰度值偏低且噪声强

度较大, 图 12(b) 是经过本文方法处理后的去噪结果, 可以观察到去噪后的暗光图像存在一些较为明显的噪声残留, 同时图像边缘的模糊程度也在变大. 鉴于目前暗光噪声数据集的匮乏, 我们计划在未来的工作中致力于解决暗光条件下的图像去噪问题.

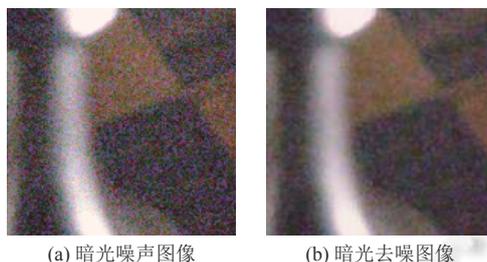


图 12 暗光条件下真实图像的去噪案例

## 5 结 论

为了解决真实噪声图像的去噪问题, 本文提出了一种基于自监督约束的双尺度真实图像盲去噪算法. 该算法仅使用真实世界的噪声图像作为训练数据, 通过构建基于全变分图像先验的边缘保持自监督约束和基于图像背景一致性的背景自监督约束, 设计了一种双尺度两阶段网络模型以实现噪声的估计和图像的去噪. 本文还提出一种结构相似性注意力机制, 可帮助网络模型复原出纹理细节更加清晰的真实图像. 通过对定性及定量的实验结果分析, 本文所提出的基于自监督约束的双尺度盲去噪算法能够针对未知噪声强度的真实图像取得较好的降噪效果, 从而验证了算法的有效性与鲁棒性. 此外, 本文提出的框架可以作为一种通用的框架运用于其他真实场景下的图像复原任务. 在未来的工作中, 我们将进一步优化算法的网络结构, 同时受生成对抗网络的启发, 我们也将进一步探索生成对抗式学习对自监督图像复原任务的重要意义.

## References:

- [1] 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. [doi: 10.1109/TIP.2007.901238]
- [2] Dong WS, Zhang L, Shi GM, Li X. Nonlocally centralized sparse representation for image restoration. *IEEE Trans. on Image Processing*, 2013, 22(4): 1620–1630. [doi: 10.1109/TIP.2012.2235847]
- [3] Elad M, Aharon M. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Trans. on Image Processing*, 2006, 15(12): 3736–3745. [doi: 10.1109/TIP.2006.881969]
- [4] Roth S, Black MJ. Fields of experts: A framework for learning image priors. In: *Proc. of the 2005 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*. San Diego: IEEE, 2005. 860–867. [doi: 10.1109/CVPR.2005.160]
- [5] Zoran D, Weiss Y. From learning models of natural image patches to whole image restoration. In: *Proc. of the 2011 Int'l Conf. on Computer Vision*. Barcelona: IEEE, 2011. 479–486. [doi: 10.1109/ICCV.2011.6126278]
- [6] Guo S, Yan ZF, Zhang K, Zuo WM, Zhang L. Toward convolutional blind denoising of real photographs. In: *Proc. of the 2019 IEEE/CVF Conf. on Computer Vision and Pattern Recognition*. Long Beach: IEEE, 2019. 1712–1722. [doi: 10.1109/CVPR.2019.00181]
- [7] Plötz T, Roth S. Benchmarking denoising algorithms with real photographs. In: *Proc. of the 2017 IEEE Conf. on Computer Vision and Pattern Recognition*. Honolulu: IEEE, 2017. 2750–2759. [doi: 10.1109/CVPR.2017.294]
- [8] Abdelhamed A, Lin S, Brown MS. A high-quality denoising dataset for smartphone cameras. In: *Proc. of the 2018 IEEE/CVF Conf. on Computer Vision and Pattern Recognition*. Salt Lake City: IEEE, 2018. 1692–1700. [doi: 10.1109/CVPR.2018.00182]
- [9] Anwar S, Barnes N. Real image denoising with feature attention. In: *Proc. of the 2019 IEEE/CVF Int'l Conf. on Computer Vision*. Seoul: IEEE, 2019. 3155–3164. [doi: 10.1109/ICCV.2019.00325]
- [10] Buades A, Coll B, Morel JM. A non-local algorithm for image denoising. In: *Proc. of the 2005 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*. San Diego: IEEE, 2005. 60–65. [doi: 10.1109/CVPR.2005.38]
- [11] Xing XX, Wang HL, Li J, Zhang XD. Asymptotic non-local means image denoising algorithm. *Acta Automatica Sinica*, 2020, 46(9):

- 1952–1960 (in Chinese with English abstract). [doi: [10.16383/j.aas.c190294](https://doi.org/10.16383/j.aas.c190294)]
- [12] Burger HC, Schuler CJ, Harmeling S. Image denoising: Can plain neural networks compete with BM3D? In: Proc. of the the 2012 IEEE Conf. on Computer Vision and Pattern Recognition. Providence: IEEE, 2012. 2392–2399. [doi: [10.1109/CVPR.2012.6247952](https://doi.org/10.1109/CVPR.2012.6247952)]
- [13] Jiang ZT, Wu X, Zhang SQ. Low-illumination image enhancement based on MR-VAE. Chinese Journal of Computers, 2021, 43(7): 1328–1339 (in Chinese with English abstract). [doi: [10.11897/SP.J.1016.2020.01328](https://doi.org/10.11897/SP.J.1016.2020.01328)]
- [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. [doi: [10.1109/TIP.2017.2662206](https://doi.org/10.1109/TIP.2017.2662206)]
- [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. [doi: [10.1109/TIP.2018.2839891](https://doi.org/10.1109/TIP.2018.2839891)]
- [16] Liu YN, Zhang SS, Wang CP, Li GY, Yang J. Image restoration based on cascading dense network in contourlet transform domain. Ruan Jian Xue Bao/Journal of Software, 2020, 31(12): 3968–3980 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/5866.htm> [doi: [10.13328/j.cnki.jos.005866](https://doi.org/10.13328/j.cnki.jos.005866)]
- [17] Liu YC, Ma W, Xu SB, Zhang XP. Edge-fidelity image denoising based on convolutional neural network. Journal of Computer-aided Design & Computer Graphics, 2020, 32(11): 1822–1831 (in Chinese with English abstract). [doi: [10.3724/SP.J.1089.2020.18170](https://doi.org/10.3724/SP.J.1089.2020.18170)]
- [18] Mildenhall B, Barron JT, Chen JW, Sharlet D, Ng R, Carroll R. Burst denoising with kernel prediction networks. In: Proc. of the 2018 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 2502–2510. [doi: [10.1109/CVPR.2018.00265](https://doi.org/10.1109/CVPR.2018.00265)]
- [19] Zhao YY, Jiang ZQ, Men AD, Ju GD. Pyramid real image denoising network. In: Proc. of the 2019 IEEE Visual Communications and Image Processing. Sydney: IEEE, 2019. 1–4. [doi: [10.1109/VCIP47243.2019.8965754](https://doi.org/10.1109/VCIP47243.2019.8965754)]
- [20] Yue ZS, Yong HW, Zhao Q, Meng DY, Zhang L. Variational denoising network: Toward blind noise modeling and removal. In: Proc. of the 2019 Annual Conf. on Neural Information Processing Systems. Vancouver, 2019. 1688–1699.
- [21] Chen JW, Chen JW, Chao HY, Yang M. Image blind denoising with generative adversarial network based noise modeling. In: Proc. of the 2018 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 3155–3164. [doi: [10.1109/CVPR.2018.00333](https://doi.org/10.1109/CVPR.2018.00333)]
- [22] Lin K, Li TH, Liu S, Li G. Real photographs denoising with noise domain adaptation and attentive generative adversarial network. In: Proc. of the 2019 IEEE/CVF Conf. on Computer Vision and Pattern Recognition Workshops. Long Beach: IEEE, 2019. 1717–1721. [doi: [10.1109/CVPRW.2019.00221](https://doi.org/10.1109/CVPRW.2019.00221)]
- [23] Lehtinen J, Munkberg J, Hasselgren J, Laine S, Karras T, Aittala M, Aila T. Noise2Noise: Learning image restoration without clean data. In: Proc. of the 35th Int'l Conf. on Machine Learning. Stockholm: PMLR, 2018. 2971–2980.
- [24] Krull A, Buchholz TO, Jug F. Noise2Void—Learning denoising from single noisy images. In: Proc. of 2019 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019. 2124–2132. [doi: [10.1109/CVPR.2019.00223](https://doi.org/10.1109/CVPR.2019.00223)]
- [25] Laine S, Karras T, Lehtinen J, Aila T. High-quality self-supervised deep image denoising. In: Proc. of the 2019 Annual Conf. on Neural Information Processing Systems. Vancouver, 2019. 6968–6978.
- [26] Batson J, Royer L. Noise2Self: Blind denoising by self-supervision. In: Proc. of the 36th Int'l Conf. on Machine Learning. Long Beach: PMLR, 2019. 524–533.
- [27] Soh JW, Cho NI. Variational deep image denoising. arXiv:2104.00965, 2021.
- [28] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. In: Proc. of the 18th Int'l Conf. on Medical Image Computing and Computer-assisted Intervention. Munich: Springer, 2015. 234–241. [doi: [10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28)]
- [29] Zhang YL, Tian YP, Kong Y, Zhong BN, Fu Y. Residual dense network for image super-resolution. In: Proc. of the 2018 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 2472–2481. [doi: [10.1109/CVPR.2018.00262](https://doi.org/10.1109/CVPR.2018.00262)]
- [30] Nam S, Hwang Y, Matsushita Y, Kim SJ. A holistic approach to cross-channel image noise modeling and its application to image denoising. In: Proc. of the 2016 IEEE Conf. on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 1683–1691. [doi: [10.1109/CVPR.2016.186](https://doi.org/10.1109/CVPR.2016.186)]
- [31] Chen YJ, Pock T. Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1256–1272. [doi: [10.1109/TPAMI.2016.2596743](https://doi.org/10.1109/TPAMI.2016.2596743)]
- [32] Chen GY, Zhu FY, Heng PA. An efficient statistical method for image noise level estimation. In: Proc. of the 2015 IEEE Int'l Conf. on Computer Vision. Santiago: IEEE, 2015. 477–485. [doi: [10.1109/ICCV.2015.62](https://doi.org/10.1109/ICCV.2015.62)]
- [33] Abdelhamed A, Lin S, Brown MS. A high-quality denoising dataset for smartphone cameras. In: Proc. of the 2018 IEEE Conf. on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 1692–1700. [doi: [10.1109/CVPR.2018.00182](https://doi.org/10.1109/CVPR.2018.00182)]

- [34] 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. Columbus: IEEE, 2014. 2862–2869. [doi: [10.1109/CVPR.2014.366](https://doi.org/10.1109/CVPR.2014.366)]
- [35] Mittal A, Soundararajan R, Bovik AC. Making a “completely blind” image quality analyzer. IEEE Signal Processing Letters, 2013, 20(3): 209–212. [doi: [10.1109/LSP.2012.2227726](https://doi.org/10.1109/LSP.2012.2227726)]
- [36] Mittal A, Moorthy AK, Bovik AC. No-reference image quality assessment in the spatial domain. IEEE Trans. on Image Processing, 2012, 21(12): 4695–4708. [doi: [10.1109/TIP.2012.2214050](https://doi.org/10.1109/TIP.2012.2214050)]

#### 附中文参考文献:

- [11] 邢笑笑, 王海龙, 李健, 张选德. 渐近非局部平均图像去噪算法. 自动化学报, 2020, 46(9): 1952–1960. [doi: [10.16383/j.aas.c190294](https://doi.org/10.16383/j.aas.c190294)]
- [13] 江泽涛, 伍旭, 张少钦. 一种基于MR-VAE的低照度图像增强方法. 计算机学报, 2021, 43(7): 1328–1339. [doi: [10.11897/SP.J.1016.2020.01328](https://doi.org/10.11897/SP.J.1016.2020.01328)]
- [16] 刘宇男, 张姗姗, 王春鹏, 李广宇, 杨建. 基于级联密集网络的轮廓波变换域图像复原. 软件学报, 2020, 31(12): 3968–3980. <http://www.jos.org.cn/1000-9825/5866.htm> [doi: [10.13328/j.cnki.jos.005866](https://doi.org/10.13328/j.cnki.jos.005866)]
- [17] 刘一畅, 马伟, 徐士彪, 张晓鹏. 基于卷积神经网络的边缘保真图像去噪算法. 计算机辅助设计与图形学学报, 2020, 32(11): 1822–1831. [doi: [10.3724/SP.J.1089.2020.18170](https://doi.org/10.3724/SP.J.1089.2020.18170)]



王迪(1992—), 女, 博士生, 主要研究领域为图像复原与增强.



唐金辉(1981—), 男, 教授, 博士生导师, CCF 杰出会员, 主要研究领域为多媒体内容分析.



潘金山(1985—), 男, 教授, 博士生导师, CCF 高级会员, 主要研究领域为图像去模糊, 图像/视频分析和增强, 计算机视觉.