

# 基于 U 形多尺度注意力方法的真实图像去噪

王新武, 陈春雨

(哈尔滨工程大学 信息与通信工程学院, 黑龙江 哈尔滨 150001)

**摘要:**针对真实世界图像去噪算法存在对上下文信息和全局信息利用不足导致的去噪效果不佳问题,提出一种 U 形金字塔注意力网络(UPCA)。U 形结构由多尺度特征模块与长距离通道注意力模块融合形成的金字塔注意力模块组成,U 形结构通过拼接操作可以将每一层的输出特征图融合,减少卷积过程以及下采样过程中图像细节特征的丢失。多尺度特征金字塔模块可以更好地利用上下文信息从而更好地恢复出干净的图像,而建立长距离依赖的通道注意力模块可以更好地利用全局信息,提高网络的去噪效果。同时在损失函数部分加入噪声项来加快训练时收敛的速度以及提高去噪效果。UPCA 网络在数据集 SIDD 和 DND 进行对比实验,验证了 UPCA 网络的可行性和先进性,同时与同样使用通道注意力的 RIDNet 相比 UPCA 网络的 PSNR/SSIM 指标提升了 0.81 dB/0.044,去噪后的效果图直观表现也更好,而且同等参数下训练所需的算力更小。

**关键词:**图像去噪;计算机视觉;真实噪声;多尺度特征;长距离通道注意力

中图分类号:TP391

文献标识码:A

文章编号:1673-629X(2024)04-0048-07

doi:10.20165/j.cnki.ISSN1673-629X.2024.0008

## Real-world Image Denoising Based on U-shaped Multi-scale Attention Method

WANG Xin-wu, CHEN Chun-yu

(School of Information and Communication Engineering, Harbin Engineering University, Harbin 150001, China)

**Abstract:**To address the issue of subpar denoising results in existing algorithms for real-world image denoising, we propose an innovative solution called the U-Shape Pyramid Channel Attention (UPCA). The U-shape structure comprises a fusion of multi-scale feature modules and long-range channel attention modules, forming a pyramid attention module. Through concatenation operations, the U-shape structure allows for the fusion of output feature maps from each layer, minimizing the loss of fine-grained image details during the convolution and downsampling processes. The multi-scale feature pyramid module effectively leverages contextual information to restore clean images, while the long-range channel attention module establishes dependencies on global information, thereby enhancing the denoising performance of the network. Additionally, we introduce a noise term in the loss function to expedite convergence during training and improve denoising efficiency. Experimental comparisons on the SIDD and DND datasets demonstrate the feasibility and superiority of the UPCA. Compared to RIDNet which also utilizes channel attention, UPCA achieves a remarkable improvement of 0.81 dB/0.044 in terms of PSNR/SSIM metrics. The visually enhanced denoised images produced by UPCA are superior, and it requires less computational power for training with the same set of parameters.

**Key words:**image denoising; computer vision; real noise; multi-scale features; long-range channel attention

## 0 引言

图像去噪作为低级视觉任务是许多高级视觉任务的基础,例如图像分类、目标检测、图像分割等。随着手机和相机在消费者和工业领域的快速发展,对被污染图像去除噪声的要求也越来越高。而在图像采集的过程中不可避免地引入噪声,如图像传感器 CCD 和

CMOS 在采集图像过程中,由于受到工作环境或者电子电路结构的影响,会引入各种噪声,如热噪声、光子噪声、暗电流噪声等。因此,从噪声图像中恢复出干净的图像不管是在工业还是生活中都是必要的。传统的去噪算法多是建立在合成的噪声上,模型往往建立在一种特定的噪声如均匀分布的高斯噪声上。传统算法

收稿日期:2023-07-30

修回日期:2023-11-30

基金项目:国家自然科学基金资助项目(61871142);中央高校基本科研业务费项目(3072020CFT0803)

作者简介:王新武(1997-),男,硕士研究生,研究方向为图像处理;通信作者:陈春雨(1974-),男,副教授,硕导,博士,研究方向为图像、视频以及语音处理。

中比较优秀的 BM3D<sup>[1]</sup> 在真实去噪效果的表现不好,因为真实世界的噪声往往是多种噪声的叠加。随着神经网络的发展,传统算法在实际的工业和生活的应用中逐渐被淘汰。最近的方法主要是基于卷积神经网络,从 DnCNN<sup>[2]</sup> 算法开始,基于神经网络的去噪算法开始明显优于传统的去噪算法。之后 FFDNet<sup>[3]</sup> 考虑将高斯噪声泛化为更加复杂的真实噪声。Zhao 等人提出了一种具有金字塔结构的真实噪声去噪网络 PRIDNet<sup>[4]</sup>,该网络分为三个阶段:噪声估计、多尺度降噪和特征融合。实验结果表明,PRIDNet 在真实噪声数据集 SIDD<sup>[5]</sup> 上取得了很好的去噪效果。Guo 等人对真实图像的噪声种类进行研究,提出了 CBDNet<sup>[6]</sup>,在真实噪声图像进行去噪并取得了较好的效果。Chen 等人提出了 GCBD<sup>[7]</sup> 网络,具体做法是用生成的含有噪声的图像样本去训练一个深层卷积神经网络,使其达到去噪目的。RIDNet<sup>[8]</sup> 是第一个将通道注意力引入图像去噪中的,在引入注意力的同时对之前基于 CNN 的两阶段模型做出改进,提出了一种端到端模型。模型可以学习噪声和在 real-noisy 图像对上,不需要单独的子网络和人为干预,同时针对之前的去噪网络在单纯叠加网络深度无法提高网络的性能问题,提出了模块化网络 EAM (Enhancement Attention Module),通过增加模块的数量来提高网络的性能。

对于三通道的彩色图像来说,真实世界噪声在每个通道上的分布是不同的<sup>[9]</sup>。而在引入通道注意力之前每个通道被同等对待,为了更好地让神经网络进行学习,去除不重要的干扰信息,在去噪的网络中加入注意力模块。RIDNet 引入的是 SENet<sup>[10]</sup> 提出的通道注意力,虽然带来了效果的提升,但存在两个具有挑战性的问题。一是无法有效地获取和利用不同尺度特征图的空间信息来丰富特征空间。前面提到的方法中,由于没有利用多尺度信息,所以只有固定的感受野,不能携带多样化的信息。由于特征并不局限于一个小区域,不同尺度的感受野可以充分地利用层次化的空间特征。当图像受到严重的噪声影响时,上下文信息有利于网络更好地恢复出原始干净图像。二是网络中的通道注意力只能有效捕获局部信息,而不能建立长距离的依赖关系。

为了解决以上问题并提升在真实世界噪声的去噪效果,该文设计了金字塔注意力模块,并使用 U 形连接的方式。同时为了让网络训练更加稳定,在网络损失的部分加入噪声的约束,最终形成了 U 形金字塔注意力网络。通过在两个测试集上的测试,UPCA 网络相比于 RIDNet 在训练速度和评价指标上都有提升。

## 1 图像去噪网络

该文设计了多尺度金字塔模块和长距离通道注意力模块,并用 U 形框架来搭建网络,同时针对噪声水平较低时噪声图像和干净图像接近造成网络优化困难的问题,增加噪声的约束。最终得到了评价指标更高、训练速度更快的真实世界噪声图像去噪网络。

### 1.1 网络结构

网络结构是类似 UNet<sup>[11]</sup> 的 U 型结构,最近一些效果不错的方法(如文献[12-15])都可以看作是 UNet 的变体。U 型结构分为编码和解码两个部分,这种结构的网络能够很好地解决由于下采样操作造成的细节丢失问题,比如边界信息,从而帮助网络更好地完成精确的定位。由编码结构到解码结构中的拼接操作可以更多地保留空间信息,这对于图像复原任务来说至关重要。受此启发,设计了 U 形网络框架并在此基础上融合多尺度金字塔模块和长距离通道注意力模块,以及为了网络训练更稳定加入了噪声约束结构,最终形成了如图 1 所示的网络整体结构。输入一张尺寸为  $128 * 128$  的噪声图像  $y$ ,图像经过  $L$  个模块,提取到特征图像  $f_L$ :

$$f_L = \text{EPCA}_L(\text{EPCA}_{L-1}(\dots(y)\dots)) \quad (1)$$

其中,EPCA (Encode Pyramid Channel Attention) 模块是 U 形网络的编码部分的金字塔注意力模块,结构如图 1(a) 所示。首先经过卷积层提升通道的维度,然后使用 PCA (Pyramid Channel Attention) 提取多尺度特征和通道注意力。经过  $L$  个 EPCA 后,将得到的特征图  $f_L$  输入 DPCA 模块进行上采样操作来还原图像的原始尺寸:

$$f_r = \text{DPCA}_L(\text{DPCA}_{L-1}(\dots(f_L)\dots)) \quad (2)$$

DPCA 结构如图 1(b),包含上采样操作、上采样特征图与下采样特征图的拼接操作以及通道注意力的计算。最后将得到的  $f_r$  送入到重构模块得到原始的尺寸:

$$N = M_r(f_r) \quad (3)$$

其中, $M_r(\cdot)$  是由一个卷积层组成的重建模块,输出的  $N$  是噪声的估计值。噪声估计值与噪声图像  $y$  进行融合操作得到去噪后的干净图像  $\hat{x}$ 。

### 1.2 U 形金字塔注意力模块

在这一节将详细介绍 U 形金字塔注意力模块,之前的图像去噪网络有两个问题。一是网络的结构通过卷积层的叠加来增强去噪效果,这导致底层特征语义信息不足,从而无法得到更好的去噪效果。二是之前的通道注意力只是关注了单个通道信息,无法建立长距离的依赖关系以及无法利用多尺度特征信息,因此对去噪效果的提升有限。为了解决这两个问题,该文

设计了一种 U 形金字塔注意力模块 (U Shape Pyramid Channel Attention, UPCA), 如图 1 所示。得益于 U 型结构同层间的长短跳跃连接, 能够使得网络在每一级的上采样过程中将同层编码器特征图在通道上进行融合。通过底层特征与高层特征的融合, 网络能够保留更多的高分辨率细节信息。同时通过金字塔注意力能

够获取不同尺度特征图的空间信息。UPCA 由  $L$  个 PCA 以及长短连接组成, 而每个 PCA 由多尺度金字塔模块 (Multiscale Pyramid Module, MPM) 和长距离注意力模块 (Long Range Attention Module, LRAM) 组成:

$$f_i = \text{MPM}(f_{i-1}) * \text{LRAM}(\text{MPM}(f_{i-1})) \quad (4)$$

$$f_i = \text{LRAM}(f) = f_m * \text{Softmax}(\text{CA}(f_m)) \quad (5)$$

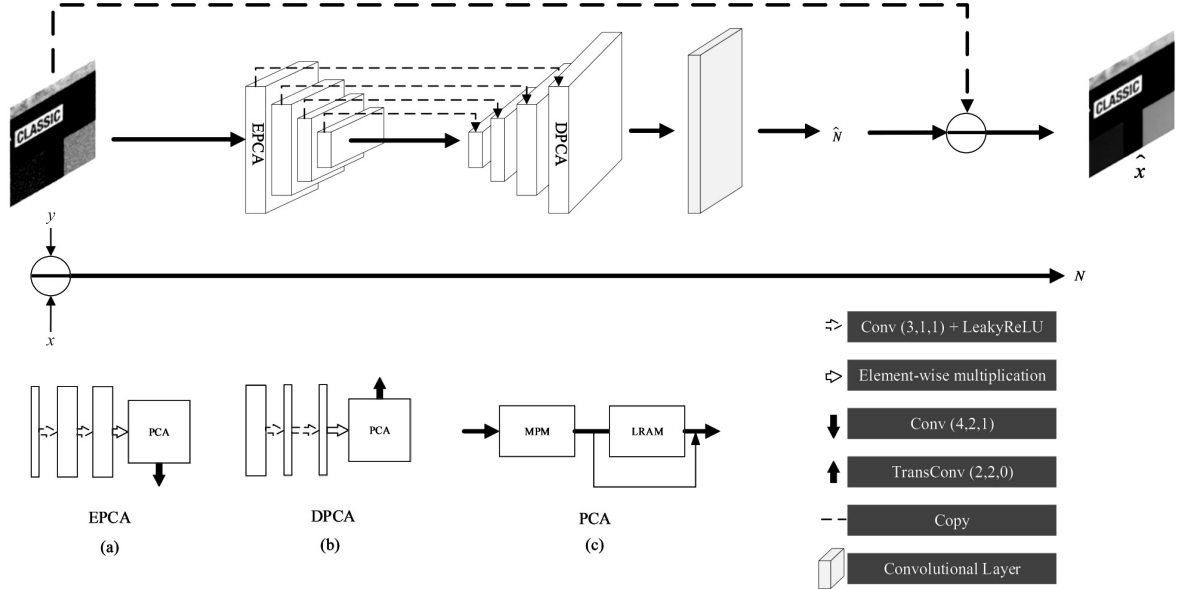


图 1 网络整体流程示意图

公式 4 中,  $f_{i-1}$  表示某层输入的特征图,  $f_i$  是融入多尺度特征注意力的特征图; 公式 5 中,  $f_m$  是 MPM 提取到的多尺度特征图,  $\text{Softmax}$  将得到的注意力分数进行重新设定, 达到获取长距离依赖的目的。MPM 将输入的特征图进行多尺度特征提取, CA 是对 MPM 提取到的多尺度特征图进行计算, 得到注意力分数, 最后经过一个  $\text{Softmax}$  操作将分组的通道注意力分数进行重新设定。这里为了兼顾训练速度和去噪效果, 经过实验测试将  $L$  设置为 5 时效果最佳。结构如图 2 所示。多尺度特征的提取将使用池化和卷积操作来实现。详细细节将在下一节展开讨论。

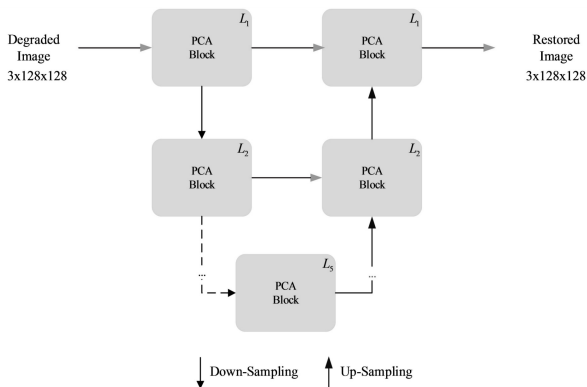


图 2 UPCA 网络模型架构

### 1.3 金字塔注意力机制

本节将对 PCA 模块中的 MPM 和 LRAM 进行详

细介绍, 并指出 RIDNet 中通道注意力的缺点以及引入多尺度通道注意力的好处。

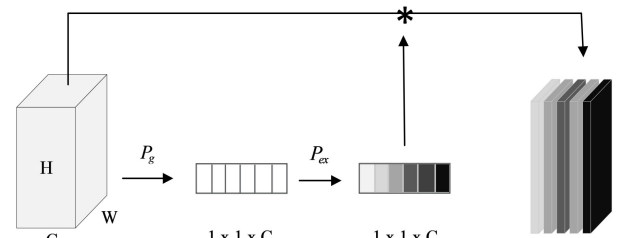


图 3 通道注意力

RIDNet 引入的是基于 SENet<sup>[10]</sup> 的通道注意力, 如图 3 所示, 直接对输入的特征图进行通道注意力计算。这种通道注意力具有两个局限性: 只能利用局部信息而不能建立长距离依赖关系和无法获取不同尺度的特征图的空间信息。这两个局限性导致在真实图像去噪的效果不好。因此, 该文设计了 MPM 和 LRAM。在通道注意力的基础上引入多尺度思想, 利用多尺度特征可以更好地提取上下文信息, 不同尺度的感受野可以充分地利用层次化的空间特征从而帮助图像恢复出原始干净图像。整体结构如图 4 所示。输入的特征图  $f$  首先经过 MPM 进行多尺度计算, 然后通过池化操作分成四种不同的尺寸, 之后得到四个子特征图  $[X_0, X_1, X_2, X_3]$ , 再分别通过卷积提取不同尺度特征图的特征信息得到  $[F_0, F_1, F_2, F_3]$ , 最后将原始输入的特



征图  $f$  和提取得到的多尺度特征图进行拼接操作得到多尺度融合特征图  $F$ 。

$$F = \text{Cat}([f, F_0, F_1, F_2, F_3]) \quad (6)$$

将得到的多尺度特征图送入 LRAM 进行通道注意力计算。首先使用  $P_g$  自适应全局池化,由于卷积层只能利用局部的信息,所以使用全局池化操作的目的是把全局空间信息压缩到通道。之后对压缩过空间信息的通道进行通道注意力计算,得到特征图  $f$  和每个  $F_i$  注意力分数  $Z_i$  之后对压缩过空间信息的通道进行通道注意力计算,得到每个  $F_i$  的注意力分数  $Z_i$ 。

$$Z_i = \sigma(P_{\text{ex}}(\max(0, P_g P_g(f, F_i)))) \quad (7)$$

其中,  $\sigma$  是一个非线性激活函数 Sigmoid,  $P_{\text{ex}}$  是两个  $1 \times 1$  的卷积用来替代全连接层,两个卷积层之间使用 LeakyReLU,最后通过一个通道乘积操作。图 4 中的  $[C_0, C_1, C_2, C_3]$  表示通道的数量。

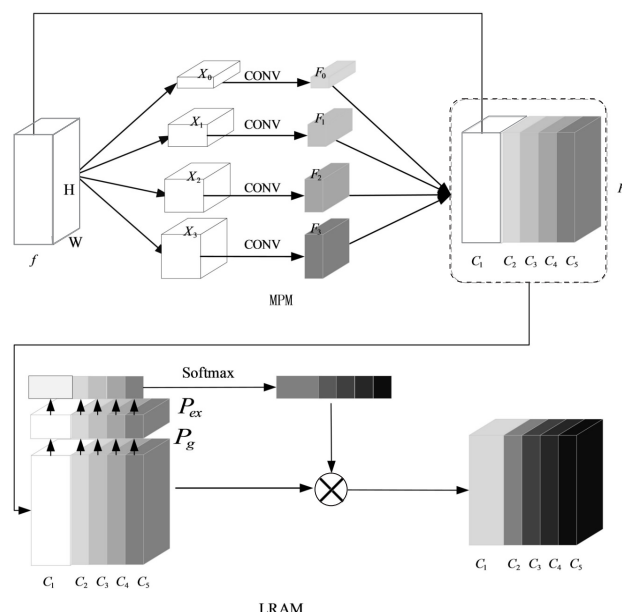


图4 PCA结构示意图

为了建立长距离的依赖关系以及利用多尺度信息之间的关系,对所有的注意力分数做 Softmax 操作,如公式 8。

$$\text{att}_i = \text{Softmax}(Z_i) = \frac{\exp(Z_i)}{\sum_{i=0}^{S-1} \exp(Z_i)} \quad (8)$$

其中,  $S$  表示注意力分数的个数,最终将重新得到的注意力分数与输入的特征图进行通道乘积操作。

#### 1.4 噪声约束模块

当噪声的水平较低时,网络的输入输出较为接近,这就导致网络的参数优化变得困难。为了解决这一问题,文中网络输出不是恢复出的干净图像而是噪声图像,这样做的原因是当噪声水平较低时即原始噪声图像和其配对的干净图像比较接近时网络的训练将变得困难,而如果输出的是噪声图像,输入原始带噪声图像

和输出纯噪声图像无论是在噪声水平较低还是正常时,两者的差距是比较大的。因此,损失函数设计如下:在两个不同的阶段分别对输出的干净图像和纯噪声图像进行约束:

$$\hat{x} = y - \text{UPCA}(y) \quad (9)$$

式中,  $\hat{x}$  表示估计出的干净图像,  $y$  是噪声图像,  $\text{UPCA}(y)$  是网络预测输出的噪声图像。网络的损失函数分为两个部分:一部分是预测的干净图像和真实图像的约束,公式如下:

$$l_{\text{mae}} = \|x - \hat{x}\|_1 \quad (10)$$

其中,  $\|\cdot\|_1$  表示的是 L1 损失函数。另一部分是真实噪声和估计噪声之间的约束,如公式 11 所示。

$$l_n = \|\text{UPCA}(y) - N\|_1 \quad (11)$$

其中,  $N$  表示真实噪声,网络的总体损失函数 loss 如公式 12 所示。

$$\text{loss} = l_{\text{mae}} + l_n = \|x - \hat{x}\|_1 + \|\text{UPCA}(y) - N\|_1 \quad (12)$$

## 2 实验结果与对比分析

### 2.1 数据集

该文使用两种不同的数据集进行训练和测试,分别是 SIDD 和 DND<sup>[16]</sup> (Darmstadt Noise Dataset),其中 DND 只能用于测试。SIDD 的全称是智能手机降噪数据集,与 DND 使用数码相机不同,因为手机本身尺寸的限制导致摄像头感光器件受到限制,所以手机摄像头相比相机更容易受到噪声的影响。SIDD 提出了一种系统性的方法用于从噪声图像中估计出无噪声图像。SIDD 使用了 5 种具有代表性的手机在不同的光照与手机设置下对大约 10 个典型场景进行了拍摄,共获取了大约 30 000 张图像。其中 24 000 张噪声和干净图像对用于训练,其余的 6 000 张成对的用于测试。为了快速训练和验证,SIDD 还提供一个中等训练集,包含了 320 对大图(用于训练)和 1 280 对  $256 \times 256$  的小图(用于验证)。文中网络在中等训练集上进行训练,分别在验证集和测试集上进行评估。

DND 通过使用 4 种商用数码相机,并对每个相机设置了不同的 ISO,得到了两两成对共 100 张的噪声图像与无噪声图像。与 SIDD 不同的是,DND 官方并没有开放无噪声图像,使得无法像 SIDD 数据集一样用于训练,但是可以使用有噪声图像,通过在官网提交去噪后的图像结果,可以得到客观评价指标。因此,该文使用 DND 进行去噪效果测试并与其他实验方法进行对比。

### 2.2 实验设置

实验模型的训练和测试在 ubuntu20.4 的系统下进行,使用 SIDD 的 sRGB 图像进行训练和测试对比实

验。由于 SIDD 数据集的原始分辨率过大无法直接进行训练,因此,将原始尺寸的图像在每个 epoch 中随机裁剪成  $128 * 128$  大小的图像块用于训练,测试时的图片尺寸为  $256 * 256$ , batch\_size 设置为 16, 总循环为 70, 每个循环进行 15 000 次的迭代。使用 PyTorch 框架 python 版本为 3.8。GPU 使用的是 12G 的 2080ti。权重的初始化参照文献[17], 网络的优化器选择 Adam, 动量参数为 (0.9, 0.999), 学习率设置为  $1e-4$ 。

### 2.3 评价指标

去噪效果的好坏可以通过主观评价指标和客观评价指标来评估,前者是通过肉眼观察,采用多人打分取平均的方式。后者主要是峰值信噪比 (Peak Signal to Noise Ratio, PSNR) 和结构相似性 (Structural Similarity, SSIM)。PSNR 表示的是原始图像像素的最大值和退化图像的噪声功率的比值,首先需要计算均方误差 (Mean Square Error, MSE), 计算公式如下:

$$MSE = \frac{1}{h * w} \sum_{i=1}^h \| I(i, j) - K(i, j) \|_2 \quad (13)$$

$$PSNR = 10 * \log_{10} \left( \frac{MAX^2}{MSE} \right) \quad (14)$$

其中,  $I(i, j)$  和  $K(i, j)$  表示尺寸为  $h * w$  的两张图像, 公式 14 中 MAX 是图像 I 中的最大像素值。SSIM 是作为 PSNR 的补充,因为在某些情景下 PSNR 和肉眼观察到的并不完全一致,有可能出现观察质量高但是 PSNR 低,或者相反的情况。这是由于人眼对于全图像素的误差并不是十分敏感,对于人眼来说,对于图像的边缘细节差异更敏感,而对平坦区域的差别没那么在意。另外人眼对于亮度差异的敏感度高于色度,这也是为什么许多图像和视频压缩方法选择损失色域的信息。为了解决这个问题,提出了基于图像结构相似性的评估方法,计算公式如下:

$$SSIM = l(i, j)^\alpha * c(i, j)^\beta * s(i, j)^\gamma \quad (15)$$

其中,  $l, c, s$  分别是亮度对比函数、对比度函数和结构相似度。

### 2.4 对比实验

由于 SIDD 的测试集没有提供无噪声图像,需要在线提交测试结果进行测试。图 5 是在线提交测试的部分结果, PSNR/SSIM 指标结果如表 1 所示。

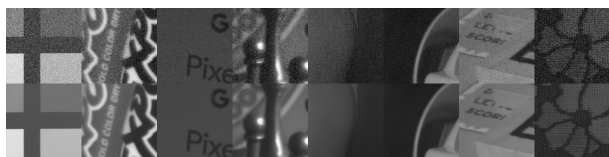


图 5 SIDD 在线测试结果

为了验证提出的算法在真实世界去噪任务上的表现,将提出的算法与最近几年去噪效果优秀的算法进行比较。对比网络包括 DnCNN<sup>[2]</sup>、FFDNet<sup>[3]</sup>、CBD-

Net<sup>[6]</sup>、UNetD<sup>[18]</sup>、RIDNet<sup>[8]</sup>、DANet<sup>[19]</sup>。在 SIDD 上的去噪结果示意图如图 6 所示。图 7 则是在 DND 上的测试示意图。在图 6 和图 7 中,对比可以看出不同方法的视觉去噪结果。FFDNet 倾向于产生过度平滑的边缘和纹理,在 DND 和 SIDD 上都是如此,不能更多地去噪声且图像的细节丢失过多。DnCNN 引入了纹理和色彩上的伪影,从图 6 中可以看出其去噪后的图像出现了原图像上没有出现过的图像,这种现象就叫作伪影。CBDNet、RIDNet 和 UNetD 在一定程度上缓解了这一现象,但是图像恢复的比较模糊,无法去除更多的噪声。DANet 和 UPCANet 在去噪效果上要明显好于前面的几种算法,但通过对比图 6 可以发现,对于文字的边缘 UPCANet 要比 DANet 恢复的更清晰,边缘的细节恢复的更好。

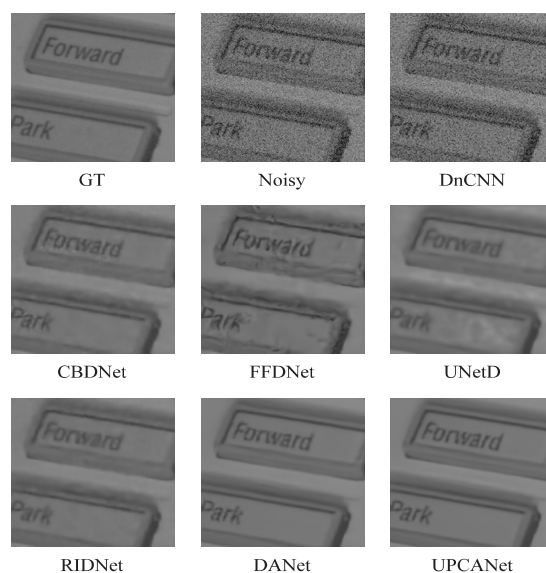


图 6 在 SIDD 上去噪的结果对比

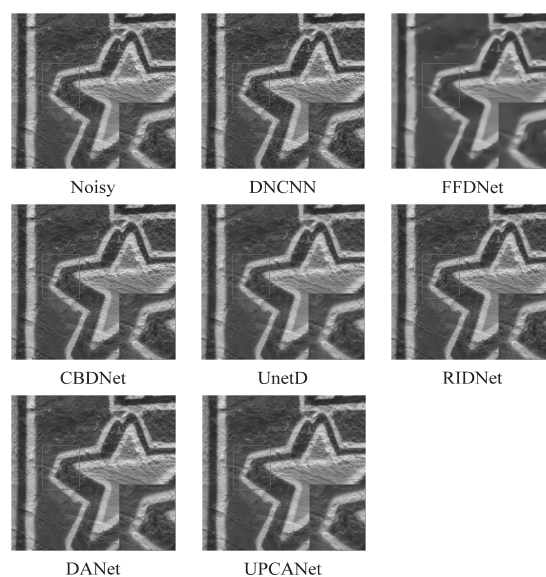


图 7 在 DND 上去噪的结果对比

从表 1 可以看出,UPCANet 的 PSNR 要比 RIDNet 的 PSNR 高 0.81 dB, SSIM 高 0.044。经过测试,在同

等批量大小和输入尺寸下UPCANet需要的算力(Multiply Accumulate Operations, MACs)是312.299G,而RIDNet则是391.815G。所以,在SIDD上UPCANet比RIDNet在去噪后的图像上更加清晰、评价指标也更高同时需要的算力更小。表2是在DND上的测试结果,包含50对有噪声和无噪声的配对图像。因为没有公开无噪声图像,因此表中列出的PSNR和SSIM是在DND官方网站上发布的数据。

表1 在SIDD上去噪的PSNR与SSIM表现

Methods	Metrics	
	PSNR/dB	SSIM
DnCNN	23.66	0.583
FFDNet	38.27	0.948
CBDNet	33.28	0.868
UNetD	38.88	0.952
RIDNet	38.71	0.913
DANet	39.25	0.955
UPCANet	<b>39.52</b>	<b>0.957</b>

从表2中可以看出,在PSNR指标上,UPCANet比RIDNet提高约0.25dB,而比单纯使用UNet去噪网络的UNetD提高了0.95dB,SSIM也要更高,说明UPCA网络在去更多的噪声情况下保留了更多的结构信息。

表2 在DND上去噪的PSNR与SSIM表现

Methods	Metrics	
	PSNR/dB	SSIM
DnCNN	32.43	0.7900
FFDNet	37.61	0.9415
CBDNet	38.06	0.9412
UNetD	38.56	0.9467
RIDNet	39.26	0.9528
DANet	39.47	<b>0.9548</b>
UPCANet	<b>39.51</b>	0.9524

UPCANet与DANet和RIDNet相比,虽然在PSNR的数值上高,但是在SSIM上分别低约0.0024和0.0004。主要是UPCA网络只在通道维度上建立了长距离依赖,而对于空间像素点的利用率不足;其次是网络多次使用池化、下采样和上采样操作,每次操作都会造成图像信息的损失。与PSNR不同,SSIM对边缘细节差异更敏感,所以导致在DND上的SSIM表现略低。但相比较在DND的SSIM单个指标的略低,在训练的速度和显存占用方面,UPCANet有着明显的优势。DANet的结构包括去噪器、生成器、判别器,所以在训练时所需要的算力(MACs)为472.857G,远高于UPCANet,同时在使用同一设置进行训练时,DANet的

显存占用为8961MB,而UPCANet只有5267MB。

所以通过以上对比可知,UPCA网络在SIDD上的评价指标和图像生成的质量上都超越其他对比的算法。而在DND上与DnCNN、FFDNet、CBDNet和UNetD相比,取得了更好的评价指标值和去噪效果。与RIDNet和DANet相比,取得了更高的PSNR值、更快的训练速度、更小的显存占用以及更低的算力。

## 2.5 消融实验

为了验证所提出的U形金字塔注意力网络去噪的有效性,针对提出的网络结构,分别设计了同时去除两种结构仅保留原始通道注意力(A),去除U形结构保留金字塔注意力结构(B);同时为了验证噪声约束结构的有效性,分别设计了去除噪声约束模块(C)和保留噪声约束模块(D)的两种实验网络。以上的消融实验同时保证网络的实验训练设置相同,使用SIDD,以PSNR评估去噪性能。结果如表3所示。图8为各模块去噪效果示意图。

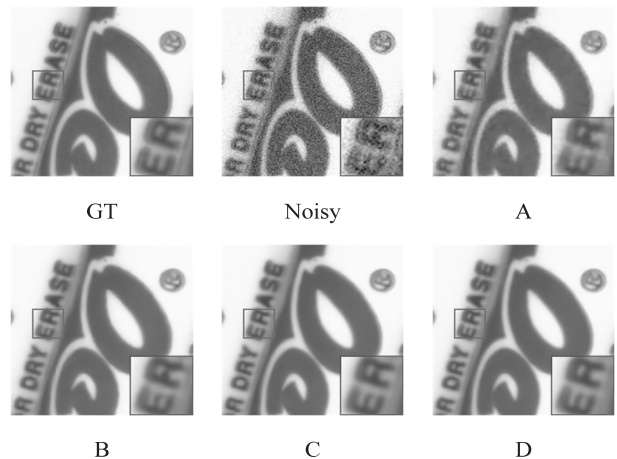


图8 各模块去噪效果示意图

从表3中可知,A,B两种网络对比可以看出有金字塔注意力结构要比单纯使用通道注意力网络提升约0.42dB。从图8中也可以看出,由于金字塔注意力结构的加入在通道维度建立了长距离依赖关系,更好地利用了多尺度特征图的空间信息,从而减少了去噪后产生的伪影,同时也去除了更多的噪声。B,C网络对比可以发现U形结构的加入要比单纯叠加网络的深度效果更好,在图8中可以看出C网络的边缘细节更加清晰,说明U形结构中的拼接操作减少了下采样带来的图像细节损失,使得PSNR提升约0.26dB。C,D分别是不加噪声约束和加入噪声约束,噪声约束的加入有利于网络的参数优化,解决了噪声水平低时训练困难的问题。从表中可以看出,在PSNR指标上比无噪声约束模块的网络高0.13dB,去噪后的图像也更加清晰。综上所述,所提出的模块和结构的去除都会带来去噪效果不同程度的降低,充分验证了模块和结



构的有效性。

表 3 在 SIDD 上不同算法得到的 PSNR 结果

Network	Component				Metrics
	CA	MPM	U 形	噪声约束	PSNR/dB
A	✓				38.71
B	✓	✓			39.13
C	✓	✓	✓		39.39
D	✓	✓	✓	✓	39.52

### 3 结束语

针对真实世界噪声图像去噪算法存在的问题,提出一种 U 形结构,减少了下采样导致的信息损失以及提高了对低维度特征图的利用。同时设计由多尺度特征模块和长距离注意力模块组成的金字塔注意力模块,在建立了长距离的依赖性关系的同时也对多尺度特征图加以利用,使得网络可以更好地利用上下文信息。最后利用噪声的约束来解决噪声水平较低时网络优化困难的问题。经过在 SIDD 和 DND 两大数据集的测试,UPCA 网络无论是在客观评价指标 PSNR/SSIM 还是在去噪的清晰度上都有提升。但该算法仍有不足和需要改进的地方。因为该算法所建立的长距离依赖关系是基于通道维度的,因此在空间像素点的长距离依赖上仍有着研究和改动的空间,这也是未来探索改进的方向。

#### 参考文献:

- [1] BUADES A, COLL B, MOREL J M. A non-local algorithm for image denoising [C]//IEEE computer society conference on computer vision and pattern recognition (CVPR'05). San Diego: IEEE, 2005: 60-65.
- [2] ZHANG K, ZUO K A, CHEN W A, et al. Beyond a Gaussian denoiser: residual learning of deep CNN for image denoising [J]. IEEE Transactions on Image Processing, 2017, 26(7): 3142-3155.
- [3] ZHANG K, ZUO W, 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.
- [4] ZHAO Y, JIANG Z, MEN A, et al. Pyramid real image denoising network [C]//IEEE visual communications and image processing (VCIP). Sydney: IEEE, 2019: 1-4.
- [5] ABDELHAMED A, LIN S, BROWN M S. A high-quality denoising dataset for smartphone cameras [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Salt Lake City: IEEE, 2018: 1692-1700.
- [6] GUO S, YAN Z, ZHANG K, et al. Toward convolutional blind denoising of real photographs [C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. Long Beach: IEEE, 2019: 1712-1722.
- [7] CHEN J, CHEN J, CHAO H, et al. Image blind denoising with generative adversarial network based noise modeling [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Salt Lake City: IEEE, 2018: 3155-3164.
- [8] ANWAR S, BARNES N. Real image denoising with feature attention [C]//Proceedings of the IEEE/CVF international conference on computer vision. Seoul: IEEE, 2019: 3155-3164.
- [9] 王瀚琛. 基于深度学习的真实图像去噪算法研究 [D]. 成都: 电子科技大学, 2021.
- [10] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Salt Lake City: IEEE, 2018: 7132-7141.
- [11] RONNEBERGER O, FISCHER P, BROX T. U-net: convolutional networks for biomedical image segmentation [C]//Medical image computing and computer-assisted intervention - MICCAI 2015: 18th international conference. Munich: Springer, 2015: 234-241.
- [12] CHEN H, WANG Y, GUO T, et al. Pre-trained image processing transformer [C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. Nashville: IEEE, 2021: 12299-12310.
- [13] ZAMIR S W, ARORA A, KHAN S, et al. Multi-stage progressive image restoration [C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. Nashville: IEEE, 2021: 14821-14831.
- [14] CHENG S, WANG Y, HUANG H, 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. Nashville: IEEE, 2021: 4896-4906.
- [15] CHO S J, JI S W, HONG J P, et al. Rethinking coarse-to-fine approach in single image deblurring [C]//Proceedings of the IEEE/CVF international conference on computer vision. Nashville: IEEE, 2021: 4641-4650.
- [16] PLOTZ T, ROTH S. Benchmarking denoising algorithms with real photographs [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Honolulu: IEEE, 2017: 1586-1595.
- [17] HE K, ZHANG X, REN S, et al. Delving deep into rectifiers: surpassing human-level performance on imagenet classification [C]//Proceedings of the IEEE international conference on computer vision. Santiago: IEEE, 2015: 1026-1034.
- [18] VO D M, NGUYEN D M, LE T P, et al. HI-GAN: a hierarchical generative adversarial network for blind denoising of real photographs [J]. Information Sciences, 2021, 570: 225-240.
- [19] YUE Z, ZHAO Q, ZHANG L, et al. Dual adversarial network: toward real-world noise removal and noise generation [C]//Computer vision - ECCV. Glasgow: Springer, 2020: 41-58.