

# 算法详细说明文档——攀登者

马云钊、赵乐平、张慎民、黄梓沫、梁傲

2023 年 2 月 8 日

## 目录

|          |   |          |
|----------|---|----------|
| <b>1</b> | <b>整体介绍</b>                                     | <b>2</b> |
| <b>2</b> | <b>算法开源信息概述</b>                                 | <b>2</b> |
| <b>3</b> | <b>算法实现的功能/拟解决的问题</b>                           | <b>2</b> |
| <b>4</b> | <b>算法设计</b>                                     | <b>2</b> |
| 4.1      | 系统算法设计 . . . . .                                | 2        |
| 4.2      | 子模块（功能）算法设计 . . . . .                           | 3        |
| 4.2.1    | Multi-Dconv Head Transposed Attention . . . . . | 3        |
| 4.2.2    | Gated-Dconv Feed-forward Network . . . . .      | 4        |
| 4.2.3    | Progressive Learning . . . . .                  | 5        |
| <b>5</b> | <b>测试结果</b>                                     | <b>5</b> |
| 5.1      | 雨滴数据集结果 . . . . .                               | 5        |
| 5.2      | 其他场景结果 . . . . .                                | 6        |
| <b>6</b> | <b>数据集及程序指令说明</b>                               | <b>7</b> |

## 1 整体介绍

组名：攀登者

小组成员：马云钊、赵乐平、张慎民、黄梓沫、梁傲

任务编号：2

任务名称：图像优化与增强技术

任务概述：基于环视摄像头采集的数据，在摄像头镜头表面有雨滴或泥点时，可去除拍摄图像中的雨滴或泥点，输出清晰、干净的图像。

任务功能性要求（关键指标）：PSNR > 25; SSIM > 0.8

## 2 算法开源信息概述

Restormer 针对 Transformer 在高分辨率图像复原中存在的难点，提出了 MDTA 与 GDFN 两种改进，极大程度上缓解了计算量与 GPU 缓存占用问题。所提方案刷新了多个图像复原任务的行业最佳性能。

论文地址：<https://arxiv.org/pdf/2111.09881.pdf> [1]

## 3 算法实现的功能/拟解决的问题

近年来，随着科技的不断发展，车载摄像头识别技术在汽车辅助驾驶领域的应用越来越广泛，其优越的拍摄质量和便捷性使得它成为了各大厂商的首选。然而，在雨天或沙尘暴天气车载摄像头镜头表面经常会有雨滴或泥点等污染物，导致图像变得模糊不清，影响识别效果、导致汽车辅助驾驶可用度下降。

针对这一问题，我们团队利用 Restormer 模型，该模型能够基于摄像头采集的数据，在摄像头镜头表面有雨滴或泥点时，可以有效去除拍摄图像中的雨滴或泥点，输出清晰、干净的图像。此外，我们团队还利用此模型对其他场景的应用，实现了对单一图像的运动模糊的去除、错焦图片的焦点去除以及夜间照片的噪点消除，从而保证了摄像头在更多工况下的可用性。

我们采用的模型利用了先进的图像处理技术，并结合了模型训练的方法，使得算法具有了极高的准确性和可靠性。经过实际验证，我们在多个雨滴数据集上对本模型和行业内其他解决方案进行了对比评测，证明了我们采用的模型具有行业内最佳的实践效果，并满足了任务功能性要求。

## 4 算法设计

本小组活用开源的 Restormer 模型，在多个雨滴数据集上对本模型和行业内其他解决方案进行对比评测，通过本模型实现了行业内的最佳实践，本组对本模型的运行效果进行了测试和验证。我们同时还针对其他场景（如：图像去污、单图像运动去模糊、图像去噪、图像去焦点）对模型进行了效果的测试和验证，验证了模型在其他场景下的可用性，我们小组还录制了一个演示视频，以展示模型出色的效果。

### 4.1 系统算法设计

由于卷积神经网络 (CNN) 在从大规模数据中学习可泛化的图像先验方面表现良好，因此这些模型已广泛应用于图像恢复和相关任务。最近，另一类神经架构 Transformers 在自然语言和高级

视觉任务上表现出显著的性能提升。虽然 Transformer 模型减轻了 CNN 的缺点（即有限的感受野和对输入内容的不适应），但其计算复杂度随空间分辨率成二次方增长，因此无法应用于大多数涉及高分辨率图像的图像恢复任务。本小组活用的 Restormer 模型是一种基于 Transformer 的图像修复模型，它能够对全局连接进行建模，并且能够适用于较大的图像。为了缓解计算瓶颈，本模型对 multi-head SA layer 和多尺度分层模块（multi-scale hierarchical module）进行了重新设计，以实现更低的计算需求。Restormer 的核心模块包括：multi-Dconv head transposed attention (MDTA) 和 gated-Dconv feed-forward network (GDFN)。这两个模块将在子模块设计中详细展开。

Restormer 模型总体流程如下图所示，模型采用类似 UNet 的结构，按着 1/2, 1/4, 1/8 进行下采样，在中间添加 skip connection。每个 Transformer block 由两个部分串联组成：MDTA 和 GDFN。给定一个退化图像  $I \in R^{H \times W \times 3}$ ，Restormer 的处理流程如下：

- 首先，采用卷积得到底层特征  $F_0 \in R^{H \times W \times 3}$ ，
- 然后，这些底层特征将经由 4 级对称编解码器处理变换为深层特征  $F_d \in R^{H \times W \times 2C}$ （每级编解码器包含多个 Transformer 模块，编码器分层降低空间分辨率，同时扩展通道，解码器则降低通道数，提升空间分辨率），注：为辅助重建过程，编解特征将与解码特征通过跳过连接进行 concat 并后接  $1 \times 1$  卷积降维；
- 其次，深层特征将经由提炼模块在进一步丰富特征多样性；
- 最后，通过卷积层处理生成残差图像  $R \in R^{H \times W \times 3}$ ，与输入图像相加即可得到复原图像  $\hat{I} = I + R$

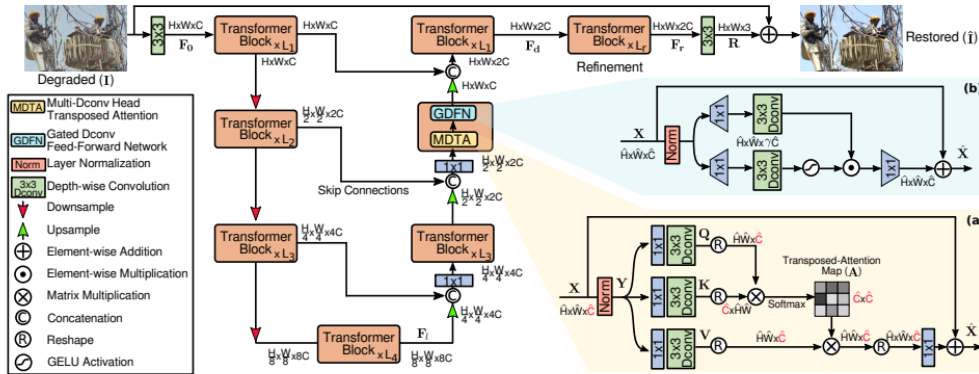


Figure 2. Architecture of Restormer for high-resolution image restoration. Our Restormer consists of multiscale hierarchical design incorporating efficient Transformer blocks. The core modules of Transformer block are: (a) multi-Dconv head transposed attention (MDTA) that performs (spatially enriched) query-key feature interaction across channels rather the spatial dimension, and (b) Gated-Dconv feed-forward network (GDFN) that performs controlled feature transformation, i.e., to allow useful information to propagate further.

图 1: Restormer 模型整体架构图

## 4.2 子模块（功能）算法设计

本模型引入了一种 multi-Dconv head transposed attention (MDTA) 模块和 gated-Dconv feed-forward network (GDFN) 以及一种渐进式的训练方式，以下是详细介绍。

### 4.2.1 Multi-Dconv Head Transposed Attention

模型提出了一种 multi-Dconv head transposed attention (MDTA) 模块，该模块能够聚合局部和非局部像素交互，并且足够有效地处理高分辨率图像，以取代具有线性复杂性的多头 SA。

Transformer 的主要计算负担来自自注意力层。在常规自注意力中, key-query 的点乘操作会随分辨率以  $O(W^2H^2)$  增长。因此, 基本不可能将自注意力层用于图像复原任务 (图像复原的输入分辨率一般都比较大)。

为解决该问题, 模型利用具有线性复杂度的 Multi-Dconv Head Transposed Attention(MDTA) 模块实现了: 在通道维度 (而非空间维度) 执行自注意力计算跨通道的交叉协方差以生成关于全局上下文的隐式注意力特征图。作为 MDTA 的另一个重要成分, 在计算特征协方差生成全局特征图之前, 模型引入了深度卷积以突出局部上下文。

对于输入  $Y \in R^{\hat{H} \times \hat{W} \times \hat{C}}$ , MDTA 的处理流程如下:

- 首先, 生成  $Q$ 、 $K$  与  $V$  (采用  $1 \times 1 + 3 \times 3$  深度卷积实现, 注: bias-free)
- 然后, 对 query 与 key 进行 reshape 并进行点乘生成转置注意力图  $A \in R^{\hat{C} \times \hat{C}}$

MDTA 的计算公式可以表示为:

$$\hat{X} = W_p \text{Attention}(\hat{Q}, \hat{K}, \hat{V}) + X \quad (1)$$

$$\text{Attention}(\hat{Q}, \hat{K}, \hat{V}) = \hat{V} \cdot \text{Softmax}(\hat{K} \cdot \hat{Q} / \alpha) \quad (2)$$

注:  $\alpha$  是一个可学习缩放参数, 用于控制点积的幅值。

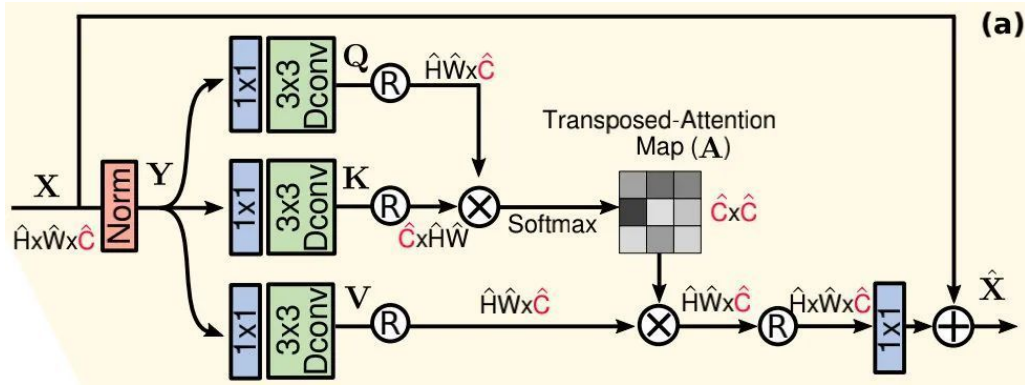


图 2: Multi-Dconv Head Transposed Attention

#### 4.2.2 Gated-Dconv Feed-forward Network

模型还提出了一种新的 gated-Dconv feed-forward network (GDFN), 用于执行受控特征转换, 抑制信息量较小的特征。GDFN 模块由两个全连接层组成, 层间具有非线性, 模块使用门控机制重新制定常规 FN 的第一个线性变换层, 以改善通过网络的信息流, 从而保证了只允许有用的信息进一步通过网络层次结构为进行特征变换, 常规 FFN (由两个  $1 \times 1$  卷积构成) 对每个像素进行独立处理。本文则对齐进行了两个改动以提升表达能力 (见下图):

- Gating 机制: 它表现为两个并行分支的元素乘操作;
- 深度卷积: 它表现为  $3 \times 3$  深度卷积以学习局部结构信息。

综上所述, GDFN 可以控制信息流动, 进而使得每层聚焦于不同的细节信息。由于 GDFN 具有比 FFN 更多的计算量, 我们降低扩展因子以控制参数数量与计算量。

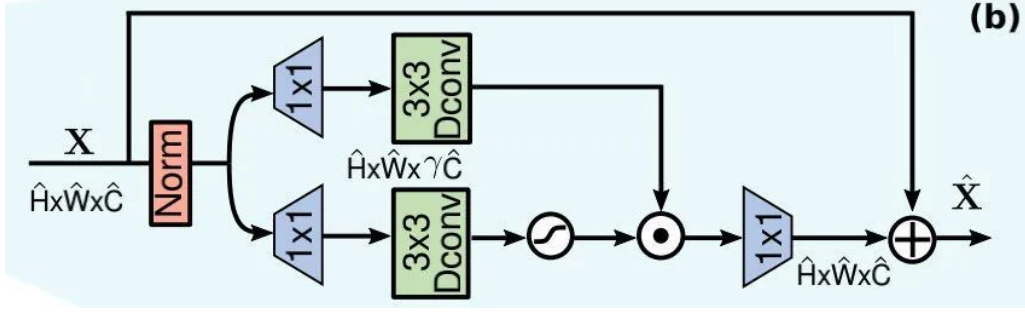


图 3: Gated-Dconv Feed-forward Network

#### 4.2.3 Progressive Learning

由于 CNN 的复原模型通常通过采用固定尺寸图像块进行训练。然而，Transformer 模型在较小块上训练可能无法进行全局统计信息编码，进而导致全分辨率测试时的次优性能。

针对该问题，模型提出了 Progressive Learning 机制：在训练的初期，本模型在较小图像块上进行训练；在训练的后期，模型采用更大图像块进行训练。由于更大的图像块会导致更长的计算耗时，模型随图像块提升而降低 batch 参数以保持与固定 batch 训练相当的耗时。总的来说，模型通过混合尺寸图像块训练的模型具有更优的性能。采取这种训练策略有助于 Restormer 从大型图像中学习上下文，并在测试时提供质量性能改善。

## 5 测试结果

我们在多个雨滴数据集上对 Restormer 模型和行业内其他解决方案进行了对比评测，证明了我们采用的模型具有行业内最佳的实践效果，并满足了任务功能性要求。为了验证模型的效果，我们还在单一图像运动去模糊，离焦去模糊，和图像去噪等数据集上验证了 Restormer 模型的性能。

实验过程中，我们针对不同的图像修复任务训练不同的模型。在所有的实验中，我们均使用以下训练参数。实验过程中的 Restormer 采用了一个 4 级编码器-解码器。从第 1 级到第 4 级，Transformer 块的数量为 [4, 6, 6, 8]，MDTA 的注意力头为 [1, 2, 4, 8]，通道的数量为 [48, 96, 192, 384]。细化阶段包含 4 个块。GDFN 中的通道扩展系数为  $\times 2.66$ 。我们用 AdamW 优化器 ( $\beta_1=0.9$ ,  $\beta_2=0.999$ , 权重衰减  $1e-4$ ) 和 L1 损失训练模型，迭代 30 万次，初始学习率  $3e-4$ ，用余弦退火法逐渐降低到  $1e-6$ [51]。对于渐进式学习，我们以  $128 \times 128$  的斑块大小和 64 的批次大小开始训练。在迭代 [92K, 156K, 204K, 240K, 276K] 时，补丁大小和批次大小对被更新为 [(1602, 40), (1922, 32), (2562, 16), (3202, 8), (3842, 8)]。

### 5.1 雨滴数据集结果

我们使用 YCbCr 色彩空间中的 Y 通道计算 PSNR/SSIM 分数。如图中所示，在所有所有雨滴数据集上，Restormer 模型比现有其他的方法取得了显著的性能提升。与此前的最佳方法 SPAIR 相比，Restormer 在所有数据集上平均实现了 1.05dB 的改进。在个别数据集上，增益可以达到 2.06dB，例如 Rain100L。在实验中我们还发现 Restormer 可以生成更好的无雨图像，且可以有效的保持结构内容。

| 模型        | Rain100L PSNR | Rain100L SSIM | 所有数据集 PSNR | 所有数据集 SSIM |
|-----------|---------------|---------------|------------|------------|
| DerainNet | 27.03         | 0.884         | 22.48      | 0.796      |
| SEMI      | 25.03         | 0.842         | 22.88      | 0.744      |
| DIDMDN    | 25.23         | 0.741         | 24.58      | 0.770      |
| UMRL      | 29.18         | 0.923         | 28.02      | 0.880      |
| RESCAN    | 29.80         | 0.881         | 28.59      | 0.857      |
| PreNet    | 32.44         | 0.950         | 29.42      | 0.897      |
| MSPFN     | 32.40         | 0.933         | 30.75      | 0.903      |
| MPRNet    | 36.40         | 0.965         | 32.73      | 0.921      |
| SPAIR     | 36.93         | 0.969         | 32.91      | 0.926      |
| Restormer | 38.99         | 0.978         | 33.96      | 0.935      |

表 1: 相比此前行业效果最佳的 SPAIR, Restormer 模型在数据集上取得了平均 1.05dB 指标提升

## 5.2 其他场景结果

Restormer 模型在单一图像运动去模糊, 离焦去模糊, 和图像去噪等数据集上的测试结果见下图。

**Table 2. Single-image motion deblurring results.** Our Restormer is trained only on the GoPro dataset [56] and directly applied to the HIDE [67] and RealBlur [65] benchmark datasets.

| Method                   | GoPro [56]   |              | HIDE [67]    |              | RealBlur-R [65] |              | RealBlur-J [65] |              |
|--------------------------|--------------|--------------|--------------|--------------|-----------------|--------------|-----------------|--------------|
|                          | PSNR         | SSIM         | PSNR         | SSIM         | PSNR            | SSIM         | PSNR            | SSIM         |
| Xu <i>et al.</i> [84]    | 21.00        | 0.741        | -            | -            | 34.46           | 0.937        | 27.14           | 0.830        |
| DeblurGAN [38]           | 28.70        | 0.858        | 24.51        | 0.871        | 33.79           | 0.903        | 27.97           | 0.834        |
| Nah <i>et al.</i> [56]   | 29.08        | 0.914        | 25.73        | 0.874        | 32.51           | 0.841        | 27.87           | 0.827        |
| Zhang <i>et al.</i> [98] | 29.19        | 0.931        | -            | -            | 35.48           | 0.947        | 27.80           | 0.847        |
| DeblurGAN-v2 [39]        | 29.55        | 0.934        | 26.61        | 0.875        | 35.26           | 0.944        | 28.70           | 0.866        |
| SRN [72]                 | 30.26        | 0.934        | 28.36        | 0.915        | 35.66           | 0.947        | 28.56           | 0.867        |
| Shen <i>et al.</i> [67]  | -            | -            | 28.89        | 0.930        | -               | -            | -               | -            |
| Gao <i>et al.</i> [23]   | 30.90        | 0.935        | 29.11        | 0.913        | -               | -            | -               | -            |
| DBGAN [100]              | 31.10        | 0.942        | 28.94        | 0.915        | 33.78           | 0.909        | 24.93           | 0.745        |
| MT-RNN [58]              | 31.15        | 0.945        | 29.15        | 0.918        | 35.79           | 0.951        | 28.44           | 0.862        |
| DMPHN [94]               | 31.20        | 0.940        | 29.09        | 0.924        | 35.70           | 0.948        | 28.42           | 0.860        |
| Suin <i>et al.</i> [71]  | 31.85        | 0.948        | 29.98        | 0.930        | -               | -            | -               | -            |
| SPAIR [61]               | 32.06        | 0.953        | 30.29        | 0.931        | -               | -            | <u>28.81</u>    | <u>0.875</u> |
| MIMO-UNet+ [14]          | 32.45        | 0.957        | 29.99        | 0.930        | 35.54           | 0.947        | 27.63           | 0.837        |
| IPT [13]                 | 32.52        | -            | -            | -            | -               | -            | -               | -            |
| MPRNet [93]              | <u>32.66</u> | <u>0.959</u> | <u>30.96</u> | <u>0.939</u> | <u>35.99</u>    | <u>0.952</u> | 28.70           | 0.873        |
| <b>Restormer</b>         | <b>32.92</b> | <b>0.961</b> | <b>31.22</b> | <b>0.942</b> | <b>36.19</b>    | <b>0.957</b> | <b>28.96</b>    | <b>0.879</b> |

图 4: 单一图像运动去模糊



Table 3. **Defocus deblurring** comparisons on the DPDD testset [3] (containing 37 indoor and 39 outdoor scenes). **S**: single-image defocus deblurring. **D**: dual-pixel defocus deblurring. Restormer sets new state-of-the-art for both single-image and dual pixel defocus deblurring.

| Method                       | Indoor Scenes   |                 |                  |                    | Outdoor Scenes  |                 |                  |                    | Combined        |                 |                  |                    |
|------------------------------|-----------------|-----------------|------------------|--------------------|-----------------|-----------------|------------------|--------------------|-----------------|-----------------|------------------|--------------------|
|                              | PSNR $\uparrow$ | SSIM $\uparrow$ | MAE $\downarrow$ | LPIPS $\downarrow$ | PSNR $\uparrow$ | SSIM $\uparrow$ | MAE $\downarrow$ | LPIPS $\downarrow$ | PSNR $\uparrow$ | SSIM $\uparrow$ | MAE $\downarrow$ | LPIPS $\downarrow$ |
| EBDB <sub>S</sub> [33]       | 25.77           | 0.772           | 0.040            | 0.297              | 21.25           | 0.599           | 0.058            | 0.373              | 23.45           | 0.683           | 0.049            | 0.336              |
| DMENet <sub>S</sub> [40]     | 25.50           | 0.788           | 0.038            | 0.298              | 21.43           | 0.644           | 0.063            | 0.397              | 23.41           | 0.714           | 0.051            | 0.349              |
| JNB <sub>S</sub> [68]        | 26.73           | 0.828           | 0.031            | 0.273              | 21.10           | 0.608           | 0.064            | 0.355              | 23.84           | 0.715           | 0.048            | 0.315              |
| DPDNet <sub>S</sub> [3]      | 26.54           | 0.816           | 0.031            | 0.239              | 22.25           | 0.682           | 0.056            | 0.313              | 24.34           | 0.747           | 0.044            | 0.277              |
| KPAC <sub>S</sub> [70]       | 27.97           | 0.852           | 0.026            | 0.182              | 22.62           | 0.701           | 0.053            | 0.269              | 25.22           | 0.774           | 0.040            | 0.227              |
| IFAN <sub>S</sub> [41]       | 28.11           | 0.861           | 0.026            | 0.179              | 22.76           | 0.720           | 0.052            | 0.254              | 25.37           | 0.789           | 0.039            | 0.217              |
| <b>Restormer<sub>S</sub></b> | <b>28.87</b>    | <b>0.882</b>    | <b>0.025</b>     | <b>0.145</b>       | <b>23.24</b>    | <b>0.743</b>    | <b>0.050</b>     | <b>0.209</b>       | <b>25.98</b>    | <b>0.811</b>    | <b>0.038</b>     | <b>0.178</b>       |
| DPDNet <sub>D</sub> [3]      | 27.48           | 0.849           | 0.029            | 0.189              | 22.90           | 0.726           | 0.052            | 0.255              | 25.13           | 0.786           | 0.041            | 0.223              |
| RDPD <sub>D</sub> [4]        | 28.10           | 0.843           | 0.027            | 0.210              | 22.82           | 0.704           | 0.053            | 0.298              | 25.39           | 0.772           | 0.040            | 0.255              |
| Uformer <sub>D</sub> [80]    | 28.23           | 0.860           | 0.026            | 0.199              | 23.10           | 0.728           | 0.051            | 0.285              | 25.65           | 0.795           | 0.039            | 0.243              |
| IFAN <sub>D</sub> [41]       | 28.66           | 0.868           | 0.025            | 0.172              | 23.46           | 0.743           | 0.049            | 0.240              | 25.99           | 0.804           | 0.037            | 0.207              |
| <b>Restormer<sub>D</sub></b> | <b>29.48</b>    | <b>0.895</b>    | <b>0.023</b>     | <b>0.134</b>       | <b>23.97</b>    | <b>0.773</b>    | <b>0.047</b>     | <b>0.175</b>       | <b>26.66</b>    | <b>0.833</b>    | <b>0.035</b>     | <b>0.155</b>       |

图 5: 离焦去模糊

Table 6. **Real image denoising** on SIDD [1] and DND [60] datasets. \* denotes methods using additional training data. Our Restormer is trained only on the SIDD images and directly tested on DND. Among competing approaches, only Restormer surpasses 40 dB PSNR.

| Dataset            | Method          | DnCNN | BM3D  | CBDNet* | RIDNet* | AINDNet* | VDN   | SADNet* | DANet+* | CycleISP* | MIRNet | DeamNet* | MPRNet | DAGL  | Uformer | Restormer    |
|--------------------|-----------------|-------|-------|---------|---------|----------|-------|---------|---------|-----------|--------|----------|--------|-------|---------|--------------|
|                    |                 | [101] | [15]  | [25]    | [6]     | [35]     | [89]  | [12]    | [90]    | [91]      | [92]   | [63]     | [93]   | [55]  | [80]    | (Ours)       |
| <b>SIDD</b><br>[1] | PSNR $\uparrow$ | 23.66 | 25.65 | 30.78   | 38.71   | 39.08    | 39.28 | 39.46   | 39.47   | 39.52     | 39.72  | 39.47    | 39.71  | 38.94 | 39.77   | <b>40.02</b> |
|                    | SSIM $\uparrow$ | 0.583 | 0.685 | 0.801   | 0.951   | 0.954    | 0.956 | 0.957   | 0.957   | 0.957     | 0.959  | 0.957    | 0.958  | 0.953 | 0.959   | <b>0.960</b> |
| <b>DND</b><br>[60] | PSNR $\uparrow$ | 32.43 | 34.51 | 38.06   | 39.26   | 39.37    | 39.38 | 39.59   | 39.58   | 39.56     | 39.88  | 39.63    | 39.80  | 39.77 | 39.96   | <b>40.03</b> |
|                    | SSIM $\uparrow$ | 0.790 | 0.851 | 0.942   | 0.953   | 0.951    | 0.952 | 0.952   | 0.955   | 0.956     | 0.956  | 0.953    | 0.954  | 0.956 | 0.956   | <b>0.956</b> |

图 6: 图像去噪

## 6 数据集及程序指令说明

详见 doc 文件夹下：数据集及程序指令说明文档.md

## 参考文献

- [1] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration, 2021.