

复现论文题目：Zero-Shot Noise2Noise: Efficient Image Denoising without any Data

1. 背景介绍

图像去噪是图像处理的一个经典问题，其目标是从受噪声干扰的退化图像中尽可能的恢复原始的真实图像，是图像进行后续处理的关键性一步。

目前大部分的图像去噪技术依赖于大量无噪声和，噪声的图像对，并且通常由经过训练的神经网络组成，将噪声图像映射到非噪声图像。这些基于数据集方法的存在着明显的缺点，数据集的采集时昂贵和耗时的，并且对于来自不同分布的图像，在数据集上训练的网络性能会受到影响。这些缺点就激发对无数据集方法的研究。

该文章提出基于 Noise2Noise 和 Neighbour2Neighbour 的方法，在去噪质量和计算资源之间做到了一个很好的权衡。Noise2Noise 是对同一静态场景的两幅噪声图像进行训练，训练一个网络将噪声图像映射到同一场景的另一个噪声图像。由于一般很难获取同一静态场景的噪声图像对（例如，被捕获的物体可能是非静态的，或者光照条件变化很快），其可用性受到了限制。Neighbour2Neighbour 拓展了 Noise2Noise，通过对单张噪声图像进行下采样来创建噪声图像对，但其在去噪真实世界的相机噪声时，性能会明显下降。

2. 主体方法框架

整个算法框架可以分为两个部分：下采样器和轻量级网络。其大致思路是首先将图像分解为一对下采样图像，然后用正则化训练一个轻量级网络，将一个下采样图像映射到另一个下采样图像。利用经过训练的网络对噪声图像去噪。

下采样器是通过大小为 $W \times H \times C$ 的输入图像生成两张大小为 $W/2 \times H/2 \times C$ 的图像。它先将图像划分为大小为 2×2 的不重叠的 patches，再对每个 patch 的对角像素取均值分配给第一张低分辨率的图像，然后将反对角线的像素取平均值分配给另一张低分辨率的图像，如下图所示。

干净图像附近的像素具有很高的相关性，噪声像素是非结构化，并且是独立的。所以，下采样得到的噪声图像对有相似的信号，但独立的噪声。

对于给定要去噪的噪声图像 y ，首先拟合一个小的图像到图像的神经网络 f_{θ} ，通过最小化下面的损失将第一个下采样的图像 $D_1(y)$ 映射到另一张下采样的图像 $D_2(y)$ 。

$$\mathcal{L}(\theta) = \|f_{\theta}(D_1(y)) - D_2(y)\|_2^2$$

当我们拟合了这个网络，可以将这个网络应用于检测原始的噪声 $\hat{x} = f_{\theta}(y)$ 。考虑到残差学习、对称损失以及一个额外的一致性加强项能够很好的提升网络的性能。所以，最终我们要最小化的损失函数是

$$\mathcal{L}(\theta) = \mathcal{L}_{\text{res.}}(\theta) + \mathcal{L}_{\text{cons.}}(\theta)。$$

$$\mathcal{L}_{\text{res.}}(\theta) = \frac{1}{2} \left(\|D_1(\mathbf{y}) - f_{\theta}(D_1(\mathbf{y})) - D_2(\mathbf{y})\|_2^2 + \|D_2(\mathbf{y}) - f_{\theta}(D_2(\mathbf{y})) - D_1(\mathbf{y})\|_2^2 \right).$$

$$\mathcal{L}_{\text{cons.}}(\theta) = \frac{1}{2} \left(\|D_1(\mathbf{y}) - f_{\theta}(D_1(\mathbf{y})) - D_1(\mathbf{y} - f_{\theta}(\mathbf{y}))\|_2^2 + \|D_2(\mathbf{y}) - f_{\theta}(D_2(\mathbf{y})) - D_2(\mathbf{y} - f_{\theta}(\mathbf{y}))\|_2^2 \right)$$

这篇文章使用的是一个轻量级网络，只有两层图像到图像的网络，由两个大小为 3×3 的卷积核和一个大小为 1×1 的卷积核组成。没有归一化和池化层，只有约 20k 的参数。

3. 实验结果

下图是添加随机的高斯噪声，去噪的效果图。

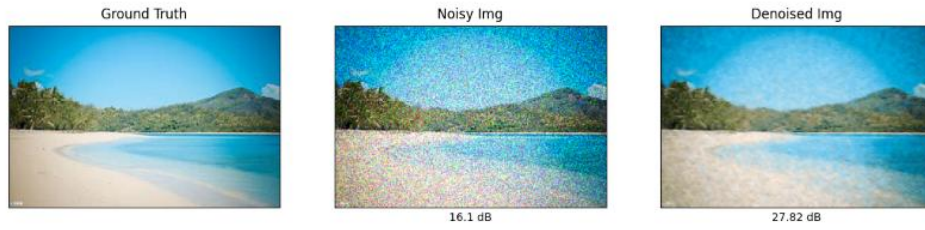


图 1 对添加随机噪声的图像去噪效果图

下图是对真实世界的相机噪声图像的去噪效果图

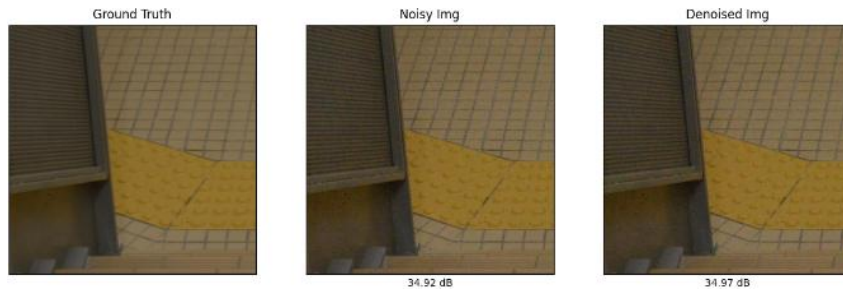


图 2 对真实世界噪声图像的去噪效果图

下图是对显微镜噪声图像的去噪效果图。

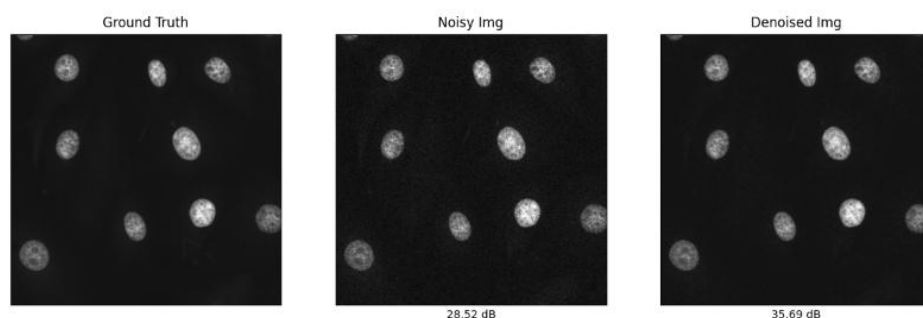


图 3 对显微镜噪声图像的去噪效果图

4. 结论以及改进思路

该篇文章提出了一种新的零镜头图像去噪算法，它不需要任何训练样例或噪声模型或水平的知识。我们的工作使用一个简单的两层网络，即使在没有 GPU 的情况下执行，也可以在相对较短的时间内去噪。

该方法可以很好地处理模拟噪声以及现实世界中的相机和显微镜噪声，并且与现有的无数据集方法相比，在泛化、去噪质量和计算资源之间取得了很好的平衡。

后续可以考虑将神经网络架构改成 unet 的结构。使用一个简单的 unet 网络来提取噪声特征，提高模型的去噪效果。也可以将 LeakyReLU 激活函数修改为 Smooth Maximum Unit (SMU) 激活函数。

参考文献：

[1] Mansour Y, Heckel R. Zero-Shot Noise2Noise: Efficient Image Denoising without any Data[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 14018-14027.