

计算机视觉 课程报告

IEEE Transactions on Image Processing, 2018

《FFDNet: Toward a Fast and Flexible Solution for CNN based Image Denoising》论文解读

—— 一种高速并且灵活的基于卷积神经网络的图像去噪算法

计算机科学与技术学院 21821152 朱策

简介

论文提出了一种快速并且灵活的用于图像去噪的卷积神经网络，将此命名为 FFDNet。该算法的网络以可调节的去噪等级图作为输入，该网络包括了几种特点：1.对于多种噪音的等级，该网络都可以有较为有效的去噪效果。2.通过指定一种不均匀的噪音等级图，网络即可以进行空间域的不同噪音的去除。3.即使在 CPU 上运行，算法也拥有不输于 BM3D 算法[1]的速度，并且不需要牺牲图像去噪的性能。

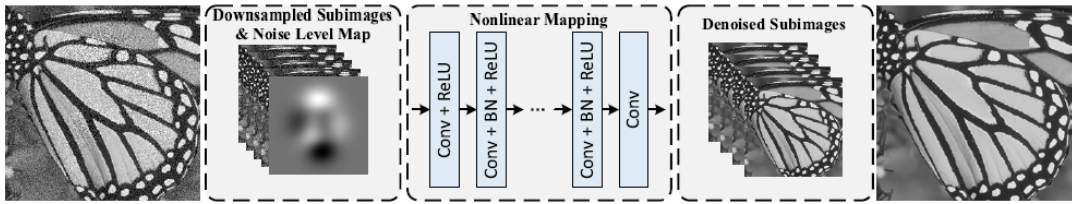
算法背景

对低层视觉的图像去噪的重要性，对于人们的日常生活来说不言而喻。噪声经常出现在图像的感知过程中，因为噪声的影响，经常会是的所需要的图像的视觉效果以及图像质量大打折扣。因此在图像的处理以及计算机视觉的领域，去除观察图像的噪声是一个十分关键的步骤。

尤其是对于一些图像复原的问题来说，其实可以将整个复原问题分为一系列的图像去噪的子问题来看待。因此我们说图像去噪算法在各种应用的领域中都有着不可或缺的作用。

在本算法中将假设图像的噪音为白色的高斯噪声，并且噪声的等级会被事先人工定义，以作为图像去噪算法的依据之一。

网络结构



组图 1 网络结构

如图 1，首先网络使用一种可逆的下采样操作将带噪声的图像 y 分为四张下采样后的子图像。并且我们将一个可调节的噪音等级图 M 与下采样后的子图像进行结合，组成一个尺寸为 $\frac{W}{2} \times \frac{H}{2} \times (4C + 1)$ 的张量 \tilde{y} ，作为卷积神经网络的输入。

以张量 \tilde{y} 作为输入，网络包含了一系列卷积核为 3×3 的卷积层组成。对于每一个卷积层，包含三种操作：卷积（Conv），线性整流函数（ReLU）^[2]以及批标准化（Batch Normalization）^[3]。其中激励函数将不运用在最后一层，在首尾层将不运用批标准化。在卷积层后，论文使用上采样的操作来进行与下采样操作相反的过程来生成去噪后的尺寸为 $W \times H \times C$ 图像 \hat{x} 。

综合考虑算法的复杂度以及表现后，论文选择对于灰度图，采用 15 层卷积层，提取 64 通道的特征。而对于色彩图，采用 12 层卷积层提取 96 通道的特征。因为对于 R,G,B 三通道的信息有着高度的依赖性，因此使用较小数目的卷积层从而激励模型学到通道间的关系，并且对于色彩图像有着更多的通道作为输入，因此需要更多的特征。

算法理解

对于论文中提出的算法以及噪音等级图的定义，我的理解如下

对于大多数的基于模型的去噪方法总是试图解决式 1 的问题

$$\hat{x} = arg \min_x \frac{1}{2\sigma^2} \|y - x\|^2 + \lambda\theta(x)$$

式 1

$\frac{1}{2\sigma^2} \|y - x\|^2$ 指数据的在噪声等级为 σ 下的保真项， $\varphi(x)$ 指图像的正则项，参数 λ 控制了两者的平衡。因为 λ 可以被整合进入 σ 所以可以有隐函数式 2

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

式 2

在这种表达下，设置噪声的等级 σ 同时扮演着控制去噪以及特征保留的平衡。因此通常来讲使用卷积神经网络来学习式 2 的映射，使用带噪声的图像以及噪声等级作为输入是一种自然的解决方案。但是因为输入 y 与 σ 在维度上差异过大，直接将它们输入卷积神经网络并不容易。因此为了解决维度不匹配的问题，将噪声等级 σ 伸展至噪声等级图 M 。在矩阵 M 中，所有的元素都是 σ ，因此最终可以得到隐函数式 3

$$\hat{x} = F(y, M, \theta)$$

式 3

实验结果

当前用于去噪评估的标准数据集如 Kodak24, BSD68, CBSD68 等多为清晰无噪声的图像集。在评估过程中，人为地在图像上添加指定类型和参数的噪声，然后通过去噪算法进行处理，再将原始图像与去噪图像进行各种比较。通常，加入的噪声为均匀方差的加性高斯白噪声。

本人在一些真实的图像数据上进行评估，这些数据集中包含自然出现的真实噪声。由于本文的网络模型是非盲的去噪算法，对于每张噪声未知的真实图像，我们用[3]中基于 PCA 的方法来估计整个图像的高斯噪声水平。

图像评估结果

本人将本文的模型和近年的一篇论文 noise2noise^[4]进行比较，测试了用 openCV 自带的库函数实现双边滤波器(bilateral filter)和非局部均值去噪(fast non-local means)的结果。其中需要设置的方差和窗口大小根据不同图像的去噪效果进行一定的调整。

下图所示的图像结果中，第一张为原图（或加入噪声后的原图），第二张为双边滤波的结果，第三张为非局部均值的结果，第四张为 noise2noise 模型的结果，最后一张为 FFDNet 模型的结果。

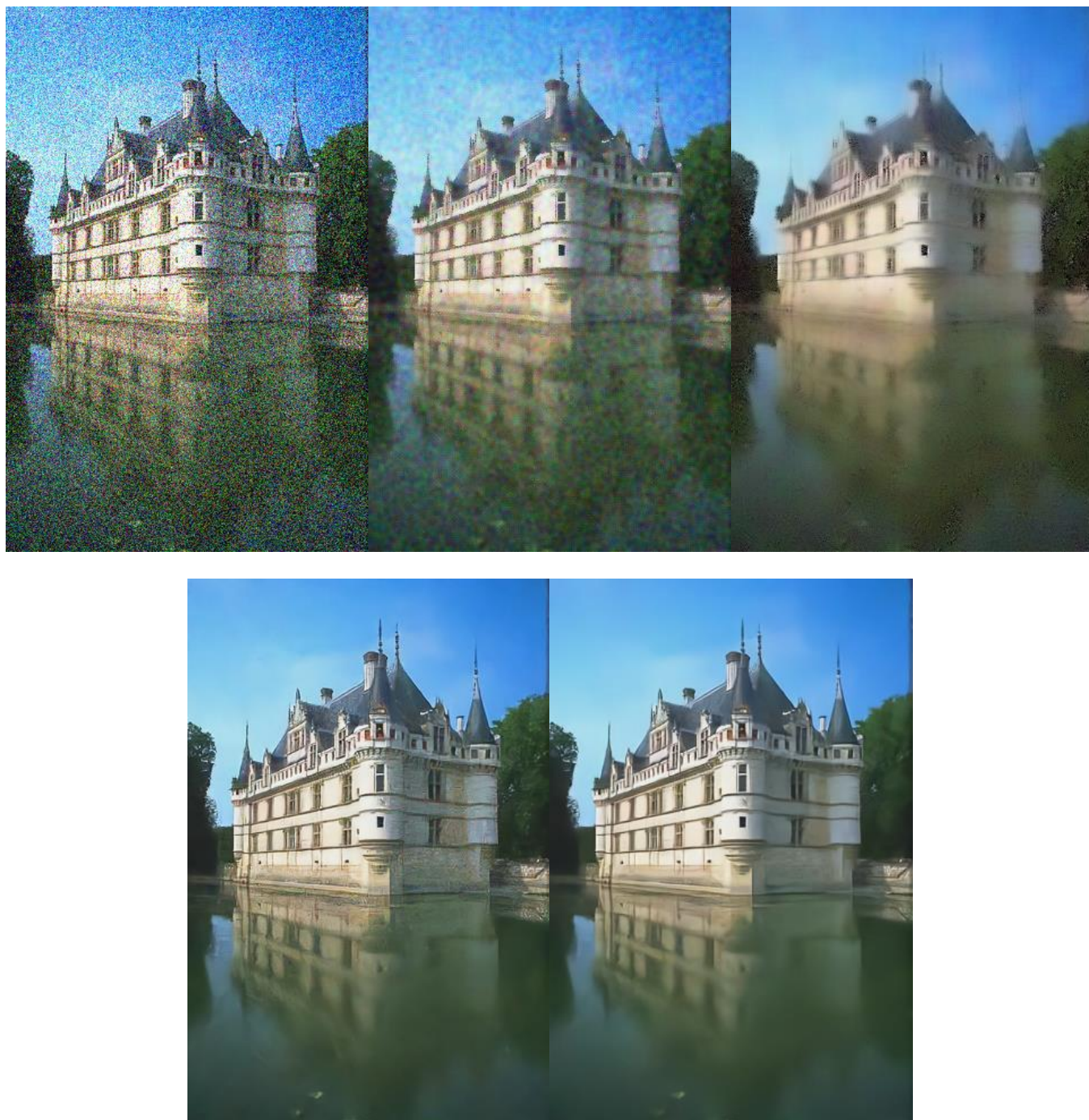


图 1 原图为清晰图像，来自 CBSD68 数据集，加入 $\sigma = 45$ 的高斯噪声

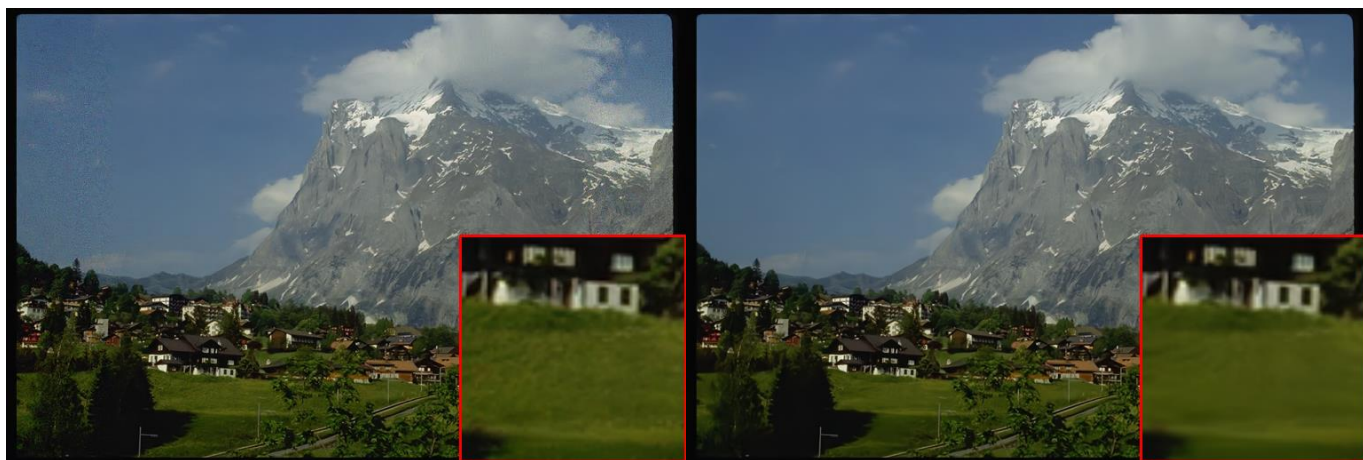
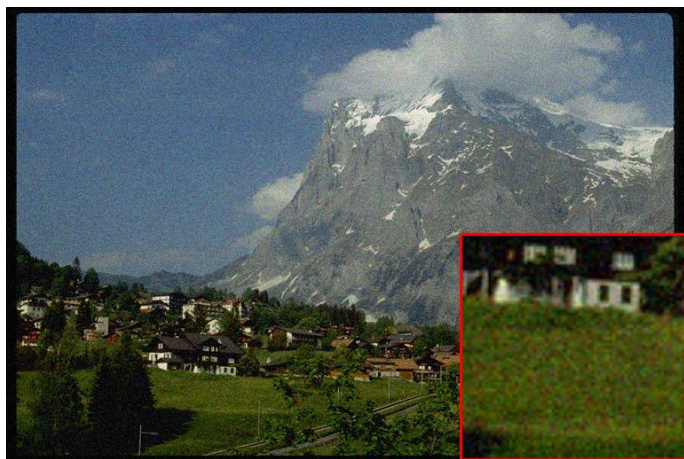


图 2 原图来自 RNI15 数据集，噪声为真实噪声

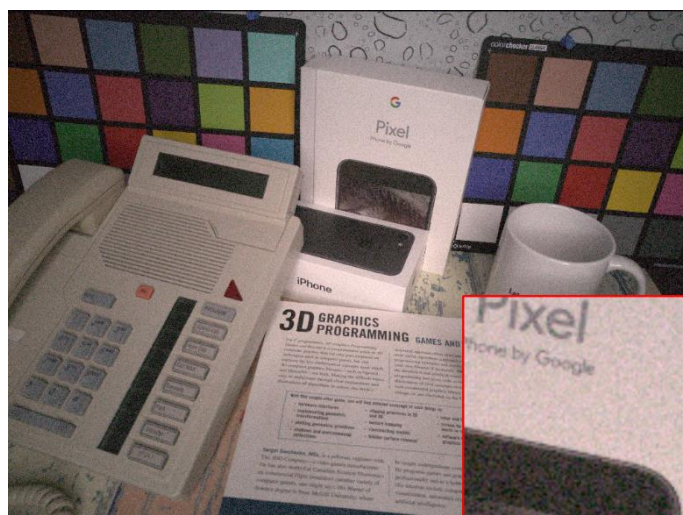


图 3 原图来自 Smartphone Image Denoising Dataset, 为手机拍摄的真实图像

数值评估结果

我们在整个 CBSD68 数据集（68 张彩色图像）上进行测试，分别添加 $\sigma = 15, 25, 35, 45$ 四种不同标准差的高斯噪声，然后依据去噪图像和原始图像计算峰值信噪比（PSNR），然后对所有图像的 PSNR 求平均值。简单绘制 PSNR 与 σ 的关系曲线如下：

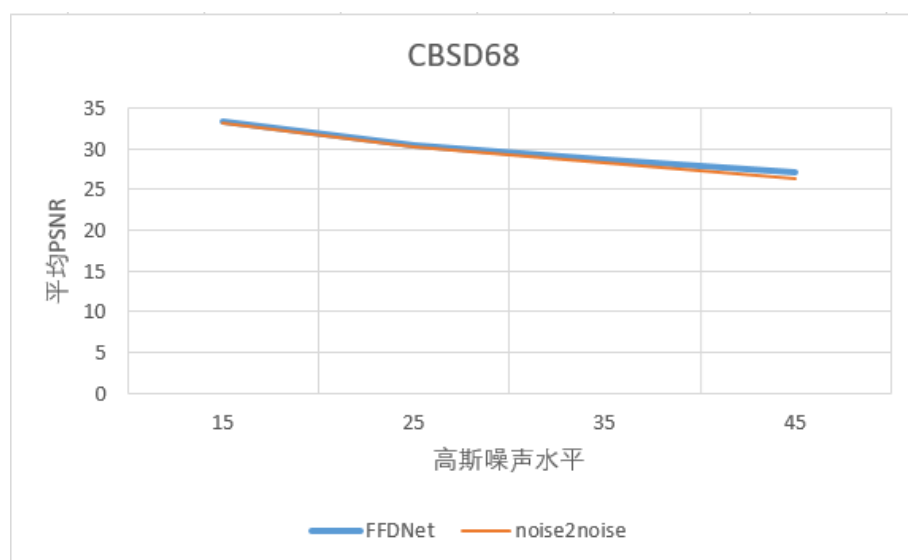


图 4 两个模型的 PSNR- σ 曲线

算法运行速度

我们测试了网络模型的在每张图像上的平均运行时间，测试环境为 GTX 1080Ti，彩色图像的运行时间约为 0.055s，而灰度图像的运行时间约为 0.033s.

失败的例子



图 5 从上到下分别为原图、 $\sigma=3$ 和 $\sigma=20$ 的 FFDNet 去噪结果、noise2noise 模型的结果

我们可以看到 $\sigma=3$ 的去噪结果基本没有消除背景的噪声，而 $\sigma=20$ 的去噪结果已经导致杯中泡沫纹理丢失了。其中噪声水平估计算法给出的是 2.8，导致该现象出现的原因是整张测试图像的噪声水平可能分布不均匀，以及图像中包含的噪声并非完全是高斯噪声，也可能是椒盐噪声或泊松噪声。

参考文献

- [1] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," IEEE Transactions on Image Processing, vol. 16, no. 8, pp. 2080–2095, 2007.
- [2] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in Advances in neural information processing systems, 2012, pp. 1097–1105.
- [3] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in Advances in neural information processing systems, 2012, pp. 1097–1105.
- [4] Liu X, Tanaka M, Okutomi M. Single-image noise level estimation for blind denoising[J]. IEEE transactions on image processing, 2013, 22(12): 5226-5237.
- [5] Lehtinen J, Munkberg J, Hasselgren J, et al. Noise2noise: Learning image restoration without clean data[J]. arXiv preprint arXiv:1803.04189, 2018.