

[深度学习] 图像反卷积的深度卷积神经网络 Deep Convolutional Neural Network for Image Deconvolution

摘要

许多基本的图像处理问题涉及反卷积计算。由于相机噪声，饱和度，图像压缩等原因，实际模糊降级很少符合理想的线性卷积模型。不同于传统的通过设计完美的模型来解决这一问题，我们要设计了一种深度卷积神经网络来捕捉图像退化的特征，这是相当具有挑战性的。我们注意到直接应用现有的深度神经网络不会产生合理的结果。我们的解决方案是建立传统的基于优化的方案与神经网络架构之间的联系，其中引入了一种新颖的、可分离结构的卷积结构来进行反卷积。我们的网络包含两个子模块，这两个子模块都以监督学习的方式进行了适当的初始化训练。与先前基于生成模型的方法相比，它们在非盲图像（non-blind image）去卷积上有不错的效果。

引言

许多图像和视频降级过程可以建模为一种平移不变卷积。为了恢复这些视觉数据，逆过程即反卷积（去卷积）成为去模糊[1,2,3,4]，超分辨率[5,6]和扩展景深[7]处理过程中的重要工具。

在涉及由相机捕获的图像的应用中，诸如饱和度，有限图像边界，噪声或压缩伪像的异常值是不可避免的。先前的研究表明，不正确地处理这些问题可能会产生与图像内容相关的大量伪像，这些伪像很难被删除。因此，有一些工作致力于在非盲反卷积中对每种特定类型的伪像进行建模和寻址，以抑制振铃效应[8]，去除噪声[9]，并处理饱和区域[9,10]。这些方法可以通过结合补丁级统计[11]或其他方案[4]进一步完善。因为每种方法都有自己的特点和局限性，所以还没有解决所有这些问题的解决方案。图1中示出了一个示例 - 具有压缩误差的部分饱和的模糊图像可能已经使许多现有方法失败。

去除这些伪影的一种方案是通过生成模型。然而，这些模型通常基于强有力的假设，例如相同且独立分布的噪声，其可能不适用于真实图像。这解释了当图像模糊特性稍微改变时，即使是高级算法也会受到影响。

在本文中，我们启动自然图像反卷积的过程，而不是基于它们的物理或数学特征。另外，我们展示了使用图像样本构建数据驱动系统的新方向，这些图像样本可以从相机轻松生成或在线收集。

我们使用卷积神经网络（CNN）来学习反卷积操作，不需要知道人为视觉效果产生的原因，与之前的基于学习的图像去模糊方法不同，它不依赖任何预处理。本文的工作是在反卷积的伪逆背景下，我们利用生成模型来弥补经验决定的卷积神经网络与现有方法之间的差距。我们产生一个实用的系统，提供了有效的策略来初始化网络的权重值，否则在卷积随机初始化训练过程中很难得到，实验证明，当输入的模糊图像是部分饱和的，我们的系统比之前的方法效果都要好。

模糊退化

图像模糊往往由多种因素导致，如：剪强度（饱和度），相机噪声，压缩等，定义如下模糊模型：

$$\hat{y} = \psi_b[\phi(\alpha x * k + n)],$$

αx 表示潜在的清晰图像，符号 $\alpha \geq 1$ 表示 αx 可能值超过相机传感器的动态范围，因此被修剪。 k 为卷积核，通常被称为一个点扩展函数（PSF）， n 为模型添加的相机噪声， $\phi(\cdot)$ 是饱和限幅函数模型，定义 $\phi(z) = \min(z, z_{\max})$ ， z_{\max} 是限幅阈值范围， $\psi_b[\cdot]$ 是非线性（如JPEG）压缩操作。

即使我们知道 y 和 k ，想要复原 αx 还是比较困难，因为在剪裁的时候会有信息损失，所以我们的目标是复原剪裁过的输入 \hat{x} ，其中 $\hat{x} = \phi(\alpha x)$ 。

尽管求解 \hat{x} 和一个复杂的能量函数涉及到上面的公式很困难，从输入图像 x 得到模糊图像比较简单，根据卷积模型将各种可能的图像退化转化为生成来合成图像，这激发了反卷积的学习过程，训练图像对 $\{\hat{x}_i, \hat{y}_i\}$ 。

分析

我们的目标是训练一个网络结构 $f(\cdot)$ ，使得下面式子最小：

$$\frac{1}{2|N|} \sum_{i \in N} \|f(\hat{y}_i) - \hat{x}_i\|^2,$$

$|N|$ 为样本集中图像对的个数。

我们已经使用了最近的两种深度神经网络来解决这个问题，但都不成功。其中一个为Stacked Sparse Denoise Autoencoder (SSDAE)，另一个为卷积神经网络（CNN），这两种方法都是为了图像去噪设计的，对于SSDAE，我们选择patch大小为17*17（patch表示某一点关联的一片区域，而点和点之间的权要用到patch和patch之间的距离，然后把搜索区域内的点加权平均）。在训练中，我们收集了两百万个清晰的样例及对应的模糊版本。其中一个例子如下：



Figure 2: Existing stacked denoise autoencoder and convolutional neural network structures cannot solve the deconvolution problem.

https://blog.csdn.net/weixin_41028208

从图上看，SSDAE的结果仍然比较模糊，CNN的效果稍微好一些，但它仍有模糊的边缘和强烈的重影痕迹，这是因为这些网络结构是为了去噪，而没有考虑到反卷积性质。

Pseudo Inverse Kernels(伪逆内核)

假设线性模糊模型为： $y = x * k$

空间卷积可以转化为频域乘法： $F(y) = F(x) \cdot F(k)$.

其中 $F(\cdot)$ 表示离散傅里叶变换 (DFT) , 操作符 \cdot 是元素相乘。在傅里叶域中, x 可以表示为:

$$x = \mathcal{F}^{-1}(\mathcal{F}(y)/\mathcal{F}(k)) = \mathcal{F}^{-1}(1/\mathcal{F}(k)) * y,$$

其中, \mathcal{F}^{-1} 为离散傅里叶逆变换。 x 的解写成一种空间卷积, 核为 $\mathcal{F}^{-1}(1/\mathcal{F}(k))$, 内核实际上是一个跨越整个空间域的重复信号。当噪声出现时, 通常使用正则化项来避免频域中除数为零, 使得伪逆 (Pseudo Inverse) 在空间域中快速衰减。

经典的维纳反卷积相当于使用Tikhonov正则化矩阵。维纳反卷积如下:

$$x = \mathcal{F}^{-1}\left(\frac{1}{\mathcal{F}(k)} \left\{ \frac{|\mathcal{F}(k)|^2}{|\mathcal{F}(k)|^2 + \frac{1}{SNR}} \right\}\right) * y = k^{\dagger} * y,$$

SNR 为信噪比, k^{\dagger} 为伪逆核, 噪声越大, $1/SNR$ 就越大。下图 (a) 显示了一个半径为7的磁盘模糊核 (disk blur kernel) , 通常用于模拟焦点模糊。下图 (b) 是伪逆核 k^{\dagger} 且 $SNR=1E-4$, 一张使用此模糊

核的图像如图 (c) , k^+ 结合反卷积的结果如图 (d) 。这种方法可以去掉一部分图像中的模糊, 但是噪声和饱和会引起视觉伪影, 这与我们对维纳反卷积的理解是一致的。总结来说, 使用深度卷积网络来做图像反卷积其实并不简单, 增加卷积核来简单地修改网络结构会导致训练难度增加, 我们采用一种新的结构来改进网络, 结果如图 (e) 。



Figure 3: Pseudo inverse kernel and deconvolution examples.

https://blog.csdn.net/weixin_41028208

虽然Wiener方法不是最先进的, 但是它的逆副核虽然有限但能够提供足够大的空间支持, 在我们的神经网络系统中变得非常有用, 这表明反卷积可以通过足够空间卷积很好地近似。这解释了中SSDA和CNN直接应用于反卷积的不成功。

- SSDA无法很好地捕捉其完全连接结构的卷积性质。
- CNN表现更好, 因为如上所述, 解卷积可以通过大内核卷积来近似。
- 以前的CNN使用小卷积内核。然而, 在我们的反卷积问题中, 它不是一个合适的配置。

因此可以总结出, 使用深度神经网络进行反卷积绝不是那么简单的。然而, 简单地通过使用大卷积核来修改网络将导致更高的训练难度。我们提出了一种新的结构来更新网络。

网络结构

基于核可分性定理, 将反卷积的伪逆核转化为卷积网络。它使网络具有更强的表现力, 映射到更高的维度以适应非线性特征。

核可分性定理

核可分性是通过奇异值分解 (SVD) 来实现的, 给定伪逆核 k^+ , 可以分解为

$$K^+ = USV^T$$

用 u_j, v_j 分别表示 U 和 V 的第 j 列, s_j 为第 j 个奇异值, 原始的伪反卷积核表示为:

$$k^{\dagger} * y = \sum_j s_j \cdot u_j * (v_j^T * y),$$

这样，二维卷积可以看成是一维滤波器的加权和。在实验中，近似地将 k^+ 看成一个可分的滤波器，将核与接近0或者非常小的 s_j 相关联，我们尝试使用真正的模糊核来忽略小于0.01的奇异值，得到的可分离核的平均值30左右。使用的SNR越小，逆核的空间支持就越小，我们还发现，长度为100的逆核通常足以产生看似合理的反卷积结果，这一点对我们设计网络的结构非常重要。

图像反卷积神经网络 (DCNN)

这个网络可以表示为：

$$h_3 = W_3 * h_2; \quad h_l = \sigma(W_l * h_{l-1} + b_{l-1}), \quad l \in \{1, 2\}; \quad h_0 = \hat{y},$$

其中， W_l 是第 $(l-1)$ 层到第 l 层的权值映射， b_{l-1} 为偏差。 $\sigma(\cdot)$ 为非线性激活函数，可以是sigmoid或者tanh。

我们的网络包含两个隐藏层，类似于可分离核的反转设置。第一个隐藏层 h_1 使用38个大规模一维核，大小为1211。第二个隐藏层 h_2 使用38个1121的卷积核对应着 h_1 中的38个映射。为了产生结果，使用一个11138的核，类似于使用奇异值 s_j 的线性组合。

这样设计的优点：

- 它聚集了可分离核反转进行反卷积，因此保证是最优的；
- 非线性特征和高维度结构使得网络比传统的伪逆结构更具有表现力。

训练DCNN

我们都自然图像采取两种策略进行实验，添加高斯噪声（AWG）和JPEG压缩技术，分为两类：一类具有强烈的色相饱和度，一类没有。饱和度影响了许多现有的反卷积算法。PSNR（峰值信噪比）如

图：

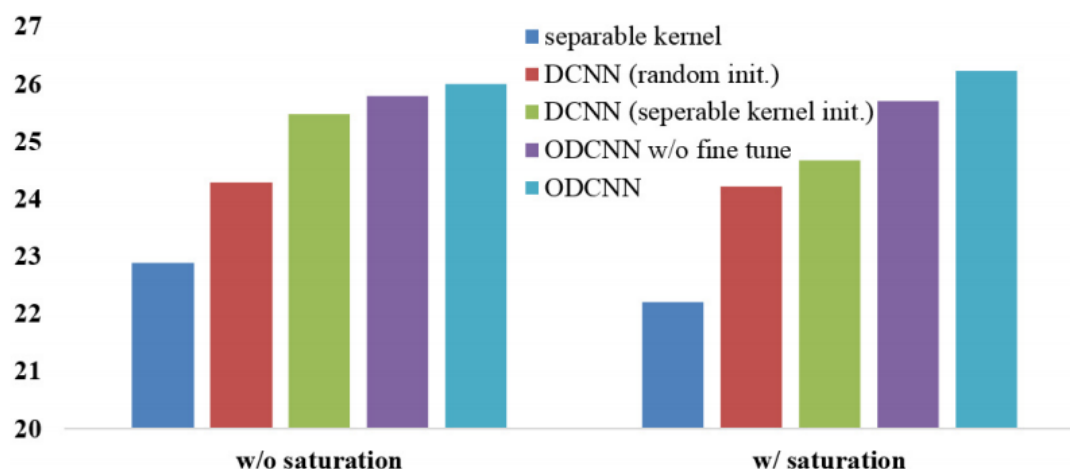


Figure 4: PSNRs produced in different stages of our convolutional neural network architecture.

https://blog.csdn.net/weixin_41028208

从图中可以得出两点：

- 训练后的网络比简单地执行可分离核反演具有优势，无论是随机初始化还是使用伪逆初始化。而且高维度映射和非线性特征是我们的网络比简单的可分离核反演更具表现力。
- 使用可分离核反演初始化的PSNR值比随机初始化后的更高，表明初始值会对网络产生影响，因此要进行调优。

其视觉比较如下：

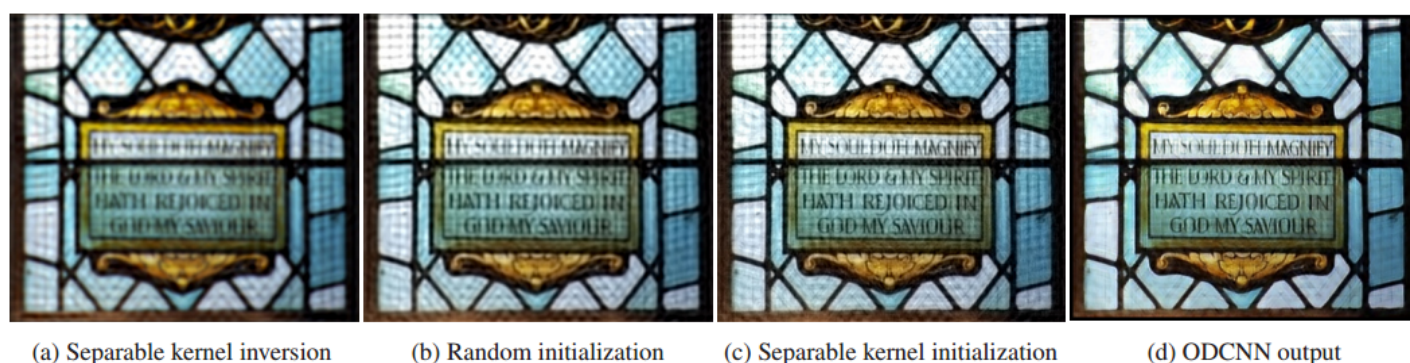


Figure 5: Results comparisons in different stages of our deconvolution CNN.

https://blog.csdn.net/weixin_41028208

其中，(a) - (c) 是可分离核反演，使用随机初始化权重和可分离反转初始化，图 (c) 明显包含更多的边缘和细节特征，最后的DCNN并不等同于任何现有的反核函数，即使有各种正则化，因为它涉及到具有非线性的高维映射。

在彩色饱和的图像中，CNN的反卷积性能降低，由于噪声和压缩，视觉伪影也可能产生，接下来，我们将通过合并一个消除噪声的CNN模块，转向更深层次的结构来解决这些遗留的问题。

Outlier-rejection Deconvolution CNN (ODCNN)

完整的网络结构是将反卷积CNN与去噪CNN结合，如下：

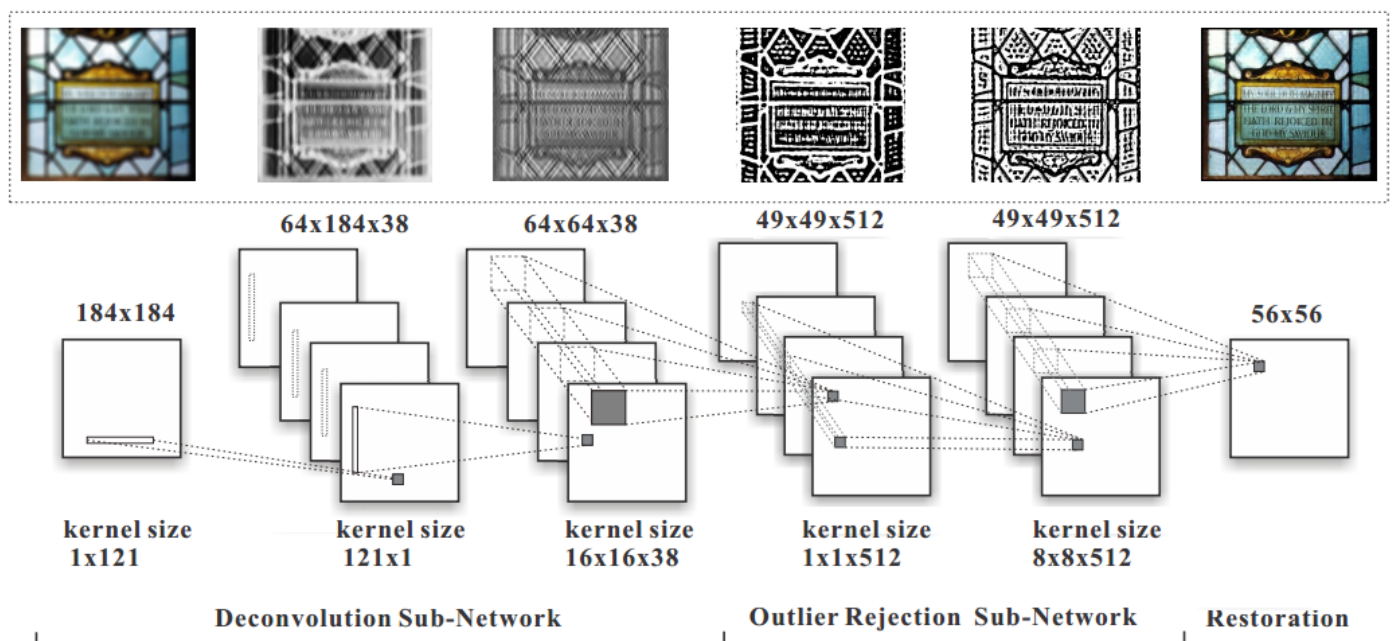


Figure 6: Our complete network architecture for deep deconvolution.

https://blog.csdn.net/weixin_41028208

去噪CNN具有两个隐藏层，512个特征映射，输入图像卷积512内核的大1616送入隐层中。将反卷积CNN的最后一层与去噪CNN的第一层网络相连接，合并1136的核及512个1616的核产生512个大小为161636的核，在结合两个模块的时候没有非线性特征，合并后权值的数量增加，我们将对其进行微调。

训练ODCNN

我们为了训练而模糊自然图像，使用2500张自然图像，随机采样得到两百万patch。我们独立地训练子网络，反卷积CNN使用前面说到的可分离初始化，其输出作为去噪CNN的输入。使用十万张184*184的小块到整个网络中进行微调，训练样本包括有噪声、饱和度、有压缩的小块，微调后的ODCNN的PSNR值增加了2dB，特别是对那些饱和区域。

实验和结论

本文中，我们呈现了几种反卷积结果，使用磁盘和运动内核来比较性能，平均PSNR如下表，视觉效果对比如下图：

kernel type	Krishnan [3]	Levin [7]	Cho [9]	Whyte [10]	Schuler [13]	Schmidt [4]	Ours
disk sat.	24.05dB	24.44dB	25.35dB	24.47dB	23.14dB	24.01dB	26.23dB
disk	25.94dB	24.54dB	23.97dB	22.84dB	24.67dB	24.71dB	26.01dB
motion sat.	24.07dB	23.58dB	25.65 dB	25.54dB	24.92dB	25.33dB	27.76dB
motion	25.07dB	24.47 dB	24.29dB	23.65dB	25.27dB	25.49dB	27.92dB

Table 1: Quantitative comparison on the evaluation image set. https://blog.csdn.net/weixin_41028208

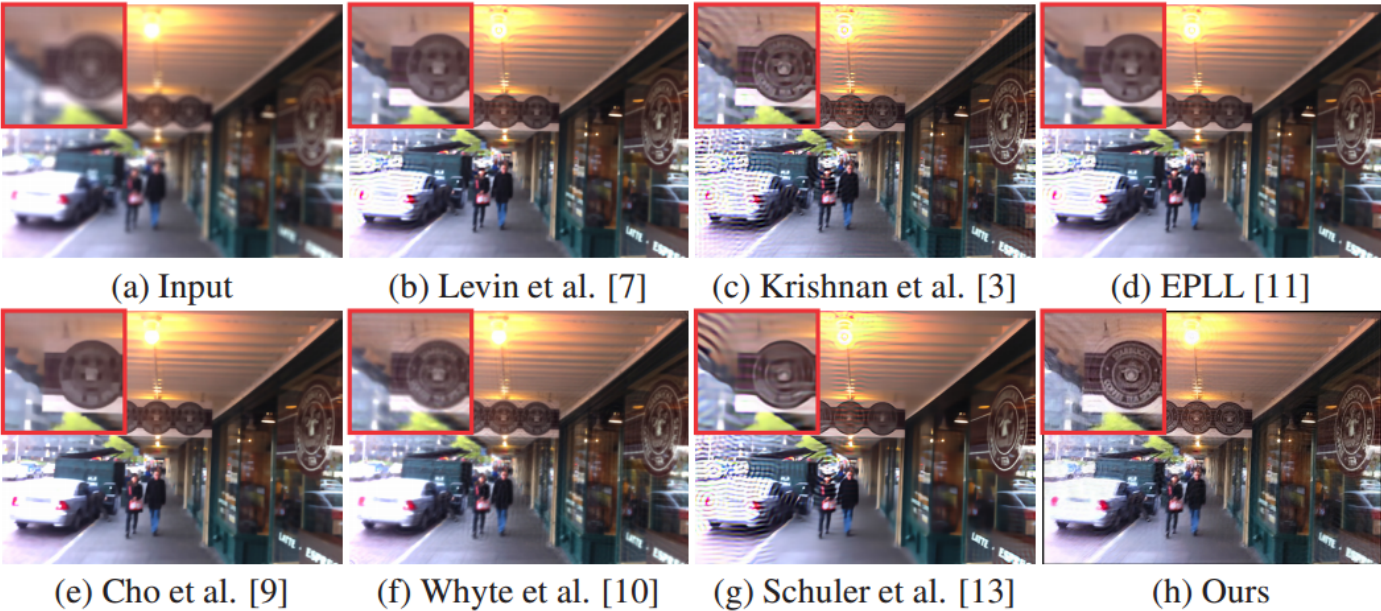


Figure 8: Visual comparison of deconvolution results. https://blog.csdn.net/weixin_41028208

我们的方法取得了极好的质量和视觉效果。

总的来说，我们提出了一种深度卷积网络结构来完成具有挑战性的反卷积任务，我们的主要贡献是使得传统的反卷积方法通过一系列的卷积步骤来指导神经网络的设计和近似的反卷积，我们的系统新颖地使用了两个模块对应的反卷积和伪影去除。由于网络很难训练，采用监督的预训练方法来初始化子网络，高质量的反卷积结果证明了该方法的有效性。