基于残差学习的图像去噪

杨昌林
21821139
计算机科学与技术学院
21821139@zju.com

摘要

由于生成模型在解决图像去噪任务上效率低下问题，近些年来判别模型逐渐成为去噪任务的主流方法。在本文中，我们尝试通过改造基本的*CNN* 架构，引入残差学习的思想，利用批标准化等机制，使得深度*CNN* 网络的训练得以加速。通过进一步堆叠残差模块单元，提升我们的深度*CNN* 网络的性能表现。我们的实验表明，相较于其他主流去噪方法，*CNN*架构在图像细节保留，边缘平滑处理，伪影问题解决等方面取得了长足的进步，更为重要的是，*CNN*架构能够处理未知噪声水平的盲去噪任务。

1. 背景介绍

图像恢复(Image Restoration，以下简称IR) 是图像视觉领域一个长期存在的问题。IR 的目标是尝试从一个被污染的图片y中恢复原始清洁图片x，一般x, y满足$y = x + v(\theta)$，$v(\theta)$表示噪声分布。通常假设v是服从一个尺度参数为σ的高斯噪声(Gaussian Noise)分布,如果能找寻参数θ 的建模方法，就可以建立一个从原始图像到噪声图像的映射，从而解决图像去噪问题。早期图像去噪就是基于这个朴素的想法，包括非局部自相似性模型(NSS Models) [1, 2, 5, 16, 23]，稀疏建模(Sparse Models) [6, 7, 24]，梯度模型[18, 17, 21]，马尔科夫随机场模型(MRF)[12, 14] 等在内。其中，NSS 模型的表现突出，并从中衍生出了BM3D[5]，LSSC[16]，NCSR[6] 等变形。

此类基于统计建模的方法，通常会遇到一些问题，如，

- 一方面，这种建模模型参数的方法往往依赖于特定的优化算法(凸优化) 求解，通常来说，这类优化算法的效率并不高，某种意义上会造成计算效率低下。
- 另一方面，有些模型甚至可能是非凸的，这意味传统的凸优化算法无法求解这类问题。通

常我们需要加入额外的先验，但这会牺牲去噪的效果。

针对这种建模参数的模型(生成模型) 的弊端，一些基于图像先验判别模型被逐渐提出。相较生成模型，判别模型可以摆脱测试阶段的迭代优化过程，弥合计算效率和恢复质量之间的差距，是一种较为折中的做法。CSF[19] 和TNRD[4] 是此类的方法的代表，但此类方法也存在下列问题，

- 一方面，采用阶段式地贪婪训练以及后续阶段的微调来学习参数，这一过程中通常会涉及到很多的手工调参。
- 另一方面，这种方法往往会受限于某种特定的噪声类型，对于未知噪声类型的去噪情况表现不佳。

事实上，我们可以完全摒弃图像先验，尝试直接建立一个从污染噪声图像到原始清洁图像的端到端去噪映射。因此考虑引入CNN 架构，它的优势主要体现在以下几点，

- 更加全面灵活地捕捉图片特征，这是深度CNN 网络的天然优势。
- Batch Normalization[10]，Residual Learning[9] 等方法的提出，使得深度CNN 网络可训练。

本文基于深度CNN 架构，引入残差学习的思想，尝试设计不同的网络架构以解决图像去噪问题。

2. 相关工作

2.1. 深度学习图像去噪

[3] 一文中，多层感知机(MLP) 首次被利用在图像去噪任务。

[22] 一文中，针对高斯噪声去噪，提出了堆叠的稀疏自编码器结构。

[11] 一文中，利用浅层CNN 网络解决图像去噪问题，性能表现上略微超过了传统MRF模型。

[15]一文中，提出了multi-level wavelet CNN (MWCNN)模型就是为了更好的在感受野大小和计算效率之间取一个权衡。

事实上，在图像去噪任务，某种特定的深度学习框架(比如AutoEncoder) 大多只能解决某类特定的噪声问题。而相较之下，得益于CNN 在提取图片特征上的优势，基于CNN 的网络架构最有可能被改造以解决一般性的噪声问题。

2.2. 残差学习

CNN 架构通常遭遇这样一个问题：随着深度的不断增加，容易梯度消失，这导致靠前的网络层很难得到有效的梯度更新。残差学习基于这样的背景提出。

其核心在于引入了恒等快捷连接，跳过一个或多个网络层。一个直观的想法是，如果一个比较浅层的网络已经达到了较高的准确率，那么深层网络直接采取$y = x$的映射也不会使整体网络的准确率下降。基于这个想法，假设某段神经网络的输入是x，理想学习目标是$H(x)$，我们直接把x 传到这段神经网络的输出作为初始值，那我们的学习目标就变为了$H(x) - x$，这就是所谓的残差学习单元。这个快捷连接并不会增加网络整体的参数和计算量，但是却可以提高网络的训练速度以及最终的训练效果。实际上残差学习蕴含了集成学习的思想。

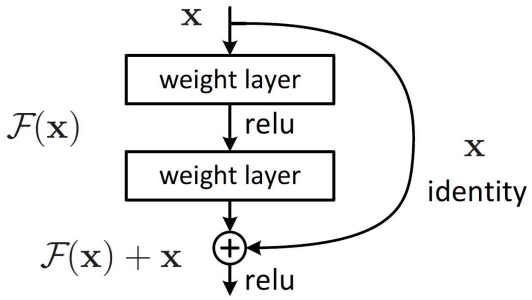

图 1. 残差模块[9]

2.3. 批标准化

批标准化(Batch Normalization)[10] 基于梯度更新时常出现的消失或爆炸问题提出，其关键在于原始架构中经由网络单元处理后，数据的分布将会发生较大的变化。假设非线性激活单元是如sigmoid 之类的函数，其特点是在极大或极小时梯度值很小。如果数据恰巧落在了sigmoid激活单元的两端，那么其梯度信息将得不到足够的更新。因此，Batch Normalization 尝试归一化输出分布，将数据根据其均值和方差重新分布。同时，为了规避改变分布而造成的非线性表达能力的下降，常常会引入缩放和偏置参数。即,

$$x = \frac{x - \mu}{\sigma}, x = scale \times x + shift$$

3. 网络架构

3.1. 网络深度

[20] 一文里，对于CNN 的网络设计进行了考量，基于文中提供的设计原则，网络架构中的卷积核大小被设置为3×3，并且移除了所有的池化层。

基于这样的设定，如果我们的网络深度为d，那么对应的原始感受野为$(2d+1)\times(2d+1)$。显然，增加深度伴随一个更大的感受野，更能充分利用较大图片区域内的特征信息，然后增加深度也会便随着计算效率的损失。因此，设置一个合适的网络深度，在网络表现和网络效率之间进行折中选择，是尤为重要的。

显然，越高的噪声水平通常需要更大的有效噪声补丁来捕获更多的上下文信息以进行恢复[13]。而有效噪声补丁的大小其实与感受野的大小密切相关[3, 11]。因此关于网络中卷积野大小的选择问题，就转化成噪声补丁大小的选择。

关于噪声补丁的选择问题，我们尝试通过以往论文的实验来解决。固定噪声水平，$\sigma = 25$，下表是几种主流去噪方法中的有效补丁大小选择。在BM3D[5]中，非局部相似补丁在大小为25×25的局部区域中自适应搜索两次，因此最终有效贴片大小为49×49。与BM3D类似，WNNM[8]使用更大的搜索窗口并迭代地执行非局部搜索，从而产生相当大的有效补丁大小，361×361。MLP[3]首先使用尺寸为39×39的补丁生成预测补丁，然后采用尺寸为9×9的滤波器对输出补丁进行平均，因此其有效补丁尺寸为47。CSF[19]和TNRD[4] 总共有五个阶段十个卷积层，过滤器尺寸为7×7，它们的有效补丁尺寸为61×61。

以EPLL具有最小的有效补丁尺寸。而CNN架构相较于这些传统方法，应具有更大的特征提取能力，这意味着如果网络得以生效，达到与之前工作相同的性能(如EPL等)，应具有更小的卷积野大小$(<=36)$。因此不妨将卷积野大小设定为35×35，根据之前的卷积野大小-深度公式$(2d+1)\times(2d+1)$，将网络深度设置为至多17层。

值得注意的一点，上述结论只在我们的设定(噪声水平$\sigma = 25$)。对于其他更一般的任务，我们可以基于这个结论进行推广，即增加网络的深度，通常来说，当网络深度>20 时，盲去噪任务亦能解决。

3.2. 架构设计

基于形式化定义，$y = x + v(\theta)$，其中y为污染图像，x为清洁图像。直接学习图像残差的思想，即直接学习污染部分v，$R(y)\approx v$，基于这个学习到的残差，我们可以获得原始清洁图像$x = y - R$。

在这个学习思路下，网络输出为残差$R(y_i;\Theta)$，Ground Truth 为$y_i - x_i$，因此均方损失函数(MSE)在此架构上的定义为，

$$\ell(\Theta) = \frac{1}{2N}\sum_{i=1}^{N}\|R(y_i;\Theta) - (y_i - x_i)\|_F^2$$

输入层，网络输入可能面临黑白图像(channel 为1) 和彩色图像(channel 为3) 的区别，因此对于输入我们需要采取一个$3\times3\times c$ 的卷积核，以转化不同情况的图像输入，紧接着经过一个ReLU 单元进行非线性转化，输入层的卷积核个数共64个。

隐藏层，前一层的输入首先经过64个3×3×64的卷积核处理，在施加非线性激活函数ReLU之前，首先将输出经由BN层进行归一化处理。

输出层，需要重新输出对应真实图像的表征，因此我们需要采取c个$3\times3\times64$的卷积核，这样输出形状对应的channel 仍为输入时的值。

对应上述分析的网络结构示意图如下所示，

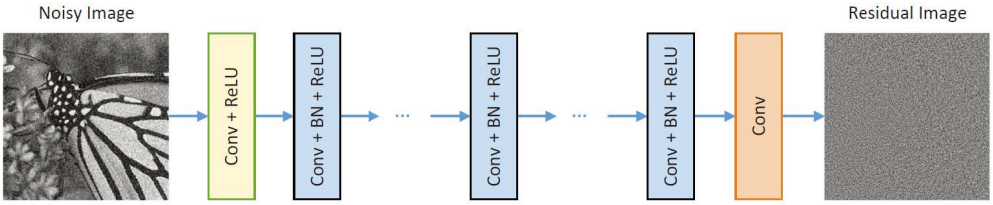

图 2. 网络架构：直接学习图像残差[25]

另外，边界伪影(Boundary Artifacts)[3]是一个在低维视觉处理领域经常遭遇的问题，造成的原因可能是图像尺寸大小和原始尺寸大小不同。为了尝试解决这个问题，通过直接的零填充策略[19]使得中间层特征映射大小和原始输入层大小相同，实验结果证明这样简单的策略确实可以减少边界伪影情况的出现。

直接的网络设计中，我们只将残差单元应用在输出层，即建立一个从输入端到输出端的快捷连接，这相较传统CNN 架构，引入了集成思想，的确加快了收敛与训练效果。

然而残差单元其实可以应用在更多的隐藏层，应用在更多的卷积单元之间。从直观来说，更密集的残差连接会进一步加快网络的收敛速度。一个可行的模型架构如下图。

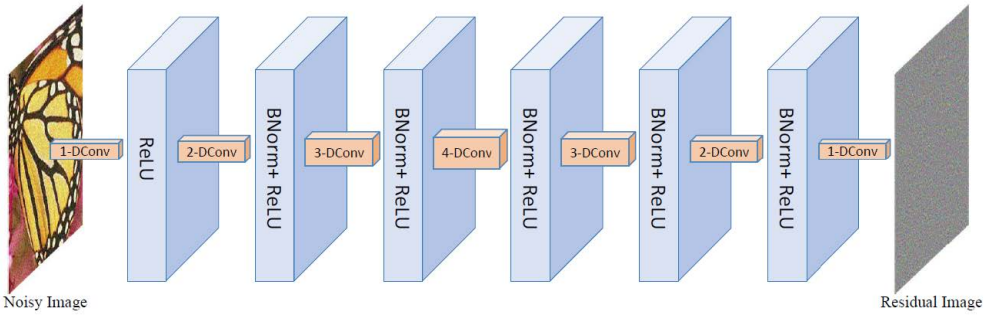

图 3. 网络架构：堆叠残差单元学习[26]

具体来说，我们以步长为2，跨度为2建立快捷连接。即如第6层的输入，来自于第5层的输出和第4层的输入(第3层的输出)。

4. 实验结果

针对之前的网络设计，我们首先实现了基本的CNN 架构，之后建立输入端到输出端的快

捷连接，直接学习图像残差的网络结构。并在此基础上，堆叠残差单元进行学习。实验环境如下，

- OS：Ubuntu 14.04.5 LTS
- CPU：Intel(R) Xeon(R) CPU E5-2630 v4 @ 2.20GHz
- GPU：NVIDIA TITAN X (Pascal) CUDA 9.0.176 Cudnn7.0.5
- 编译环境: Python3.6 TensorFlow1.8

实验数据集选取为CIFAR-10 数据集，该数据集共有60000 张彩色图像，这些图像大小均是32×32，数据集本身是带有分类标签的，但在去噪任务中并未使用。对于数据集里的任一图片，我们可以认为是该图片是清洁图像。基于预先设定的规则，比如($\sigma = 25$ 的高斯噪声分布)，我们在清洁图像上施加噪声，作为污染后的图像。这里的σ是一个超参数。

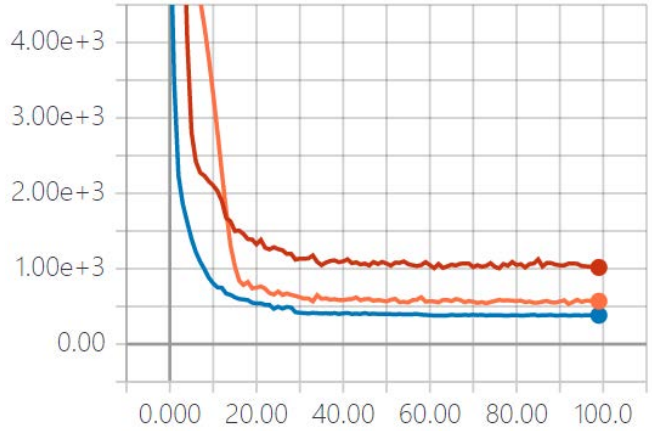

图 4. 训练误差

基于这样的处理方法，我们对数据集进行随机分组，按照$8:1:1$的比例产生训练集，验证集和测试集。在初始学习率为$1e-3$的情况下，迭代100 个Epoch，学习率每隔30个Epoch衰减一半。

上图描述了基本CNN 框架，基于残差学习的CNN架构以及堆叠残差单元的CNN架构三个网络，在训练时的表现。可以预见的是，随着残差单元的引入，收敛加速，并且最终的达到一个更低的收敛程度。随着残差单元的引入，最终同样深度的网络结构收敛到一个更小的极小值。

另外，我们比较了以上架构和其他主流方法的PSNR 值，如下表。其中D-CNN 表示直接学习图像残差的CNN 架构，I-CNN 表示堆叠残差单元的CNN 架构。

Method	$\sigma = 15$	$\sigma = 25$	$\sigma = 50$
BM3D	31.07	28.57	25.62
WNNM	31.37	28.837	25.87
EPLL	31.21	28.68	25.67
MLP	-	28.96	26.03
CSF	31.24	28.74	-
TNRD	31.42	28.92	25.97
D-CNN	32.73	29.91	26.49
I-CNN	33.01	30.45	26.92

表 1. PSNR 结果对比

从上表可以得到和测试集上类似的结构，随着残差单元的堆叠，CNN 架构更容易得到收敛，能达到一个更小的局部最小值，有更好的性能表现。

另外，我们从图像细节方面对生成效果进行比较。可以发现，传统方法BM3D，WNNM，EPLL和MLP等倾向于产生光滑的边缘和纹理。在保留锐边和精细细节的同时，TNRD很有可能会在平滑区域产生伪影。相比之下，CNN架构不仅可以较好地恢复锋利的边缘和细节，并且不会产生视觉伪影的问题，在平滑区域上表现良好。对于彩色图像去噪，基准的BM3D 会在某些区域产生假色伪影，而D-CNN可以恢复具有自然色彩的图像。此外，D-CNN相比BM3D，具有更多的图像细节和更为锐利的图像边缘。

5. 结论

在本文中，基于残差学习思想，我们提出了一个深度CNN 残差网络，通过学习图像残差，实现图像去噪。另外，我们尝试堆叠更多的残差单元，在加速收敛和性能提高等方面都取得了进步。基于残差学习的CNN 架构，相较于普通的CNN 架构有更好的性能表现和更快的训练速度，相较于其他主流方法在图像细节(锐利边缘，伪影问题等) 上有更好的表现。并且，深度CNN 架构在处理未知噪声水平的盲去噪任务上，也达到了极高的去噪水平。

参考文献

[1] A. Buades, B. Coll, and J.-M. Morel. A non-local algorithm for image denoising. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 2, pages 60–65. IEEE, 2005.

[2] A. Buades, B. Coll, and J.-M. Morel. Nonlocal image and movie denoising. *International journal of computer vision*, 76(2):123–139, 2008.

[3] H. C. Burger, C. J. Schuler, and S. Harmeling. Image denoising: Can plain neural networks compete with bm3d? In *2012 IEEE conference on computer vision and pattern recognition*, pages 2392–2399. IEEE,

2012.

[4] Y. Chen and T. Pock. Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration. *IEEE transactions on pattern analysis and machine intelligence*, 39(6):1256–1272, 2016.

[5] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. image processing, ieee transactions on 16 (8), pp. 2080-2095. 2007.

[6] W. Dong, L. Zhang, G. Shi, and X. Li. Nonlocally centralized sparse representation for image restoration. *IEEE transactions on Image Processing*, 22(4):1620–1630, 2012.

[7] M. Elad and M. Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image processing*, 15(12):3736–3745, 2006.

[8] S. Gu, L. Zhang, W. Zuo, and X. Feng. Weighted nuclear norm minimization with application to image denoising. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2862–2869, 2014.

[9] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[10] S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*, 2015.

[11] V. Jain and S. Seung. Natural image denoising with convolutional networks. In *Advances in neural information processing systems*, pages 769–776, 2009.

[12] X. Lan, S. Roth, D. Huttenlocher, and M. J. Black. Efficient belief propagation with learned higher-order markov random fields. In *European conference on computer vision*, pages 269–282. Springer, 2006.

[13] A. Levin and B. Nadler. Natural image denoising: Optimality and inherent bounds. In *CVPR 2011*, pages 2833–2840. IEEE, 2011.

[14] S. Z. Li. *Markov random field modeling in image analysis*. Springer Science & Business Media, 2009.

[15] P. Liu, H. Zhang, K. Zhang, L. Lin, and W. Zuo. Multi-level wavelet-cnn for image restoration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 773–782, 2018.

[16] J. Mairal, F. R. Bach, J. Ponce, G. Sapiro, and A. Zisserman. Non-local sparse models for image restoration. In *ICCV*, volume 29, pages 54–62. Citeseer, 2009.

[17] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin. An iterative regularization method for total variation-based image restoration. *Multiscale Modeling & Simulation*, 4(2):460–489, 2005.

[18] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: nonlinear phenomena*, 60(1-4):259–268, 1992.

[19] U. Schmidt and S. Roth. Shrinkage fields for effective image restoration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2774–2781, 2014.

[20] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[21] Y. Weiss and W. T. Freeman. What makes a good model of natural images? In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, 2007.

[22] J. Xie, L. Xu, and E. Chen. Image denoising and inpainting with deep neural networks. In *Advances in neural information processing systems*, pages 341–349, 2012.

[23] J. Xu, L. Zhang, W. Zuo, D. Zhang, and X. Feng. Patch group based nonlocal self-similarity prior learning for image denoising. In *Proceedings of the IEEE international conference on computer vision*, pages 244–252, 2015.

[24] Z. Zha, X. Liu, X. Huang, H. Shi, Y. Xu, Q. Wang, L. Tang, and X. Zhang. Analyzing the group sparsity based on the rank minimization methods. In *2017 IEEE International Conference on Multimedia and Expo (ICME)*, pages 883–888. IEEE, 2017.

[25] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Transactions on Image Processing*, 26(7):3142–3155, 2017.

[26] K. Zhang, W. Zuo, S. Gu, and L. Zhang. Learning deep cnn denoiser prior for image restoration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3929–3938, 2017.