PatchDIP: Exploiting Patch Redundancy in Deep Image Prior for Denoising

Muhammad Asim, Fahad Shamshad, and Ali Ahmed
Department of Electrical Engineering, Information Technology University, Lahore, Pakistan
{msee16001, fahad.shamshad, ali.ahmed}@itu.edu.pk

Abstract

The structure of a deep convolutional neural network initialized with random weights is able to sufficiently capture the patterns in a natural image. This finding motivates using deep neural network as an effective prior for natural images. In this work, we show that this strong prior, enforced by the structure of a ConvNet, can be augmented with the information that recurs in different patches of a natural image to boost the performance. We demonstrate that the self-similarity in the image patches can be exploited alongside deep image prior by optimizing the network weights to fit patches extracted from a single noisy image. Our results indicate that employing deep image prior on noisy patches provides an additional disincentive for the network to fit noise, and is encouraged to exploit redundancies among the patches yielding better denoising performance.

1 Introduction

In this paper, we consider the problem of recovering an unknown signal $\boldsymbol{x} \in \mathbb{R}^n$ from its noisy measurements $\boldsymbol{y} \in \mathbb{R}^n$ of the form

$$\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{\eta}, \tag{1}$$

where $\boldsymbol{\eta} \in \mathbb{R}^n$ denotes additive Gaussian noise. Since denoising is an ill-posed and under-constrained inverse problem, some prior knowledge is often used to restrict the solution space.

Natural image statistics, such as sparsity [1], smoothness [2], and patch recurrence [3] have guided the development of priors for image recovery from noisy measurements. These hand-crafted priors have proven to be effective. However, a large majority of unnatural signals also satisfy the constraints specified by these hand-designed priors due to which they suffer. To overcome this limitation, several deep learning based methods have been developed to learn image prior models for reconstructing an estimate of the true image from the noisy observation. Specifically, these deep learning based approaches invert the forward acquisition model of denoising via end-to-end training of deep neural networks in a supervised manner [4]. These techniques require a large number of clean-noisy image pairs for network training that are inconvenient to obtain.

To recall classical engineered priors, a leap of improvement in denoising was obtained by exploiting the patch recurrence in natural images that has led to several advanced denoising algorithms, such as BM3D. The key idea is to leverage the self similarity of natural images i.e. patches extracted from a natural image tend to recur much more frequently (densely) inside the same image, than in any random collection of images.

Surprisingly, it was recently shown in [5] that the structure of a deep convolutional neural network (CNN) can regularize image reconstruction without any prior training. Deep Image Prior (DIP) is a new strategy for handling the regularization task in inverse problems. Rather than taking the supervised avenue, as most earlier methods did, DIP suggests to benefit from the structure of deep network itself assuming that convolutional neural network has lower impedance in fitting image

33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada.

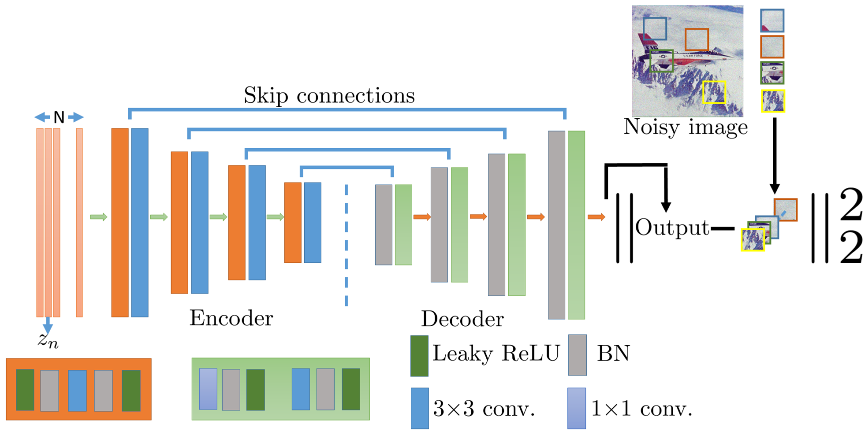


Figure 1: Block diagram of proposed approach.

signal as compared to noise signal. DIP fits the network weights to the corrupted image adapting it for each image to be treated.

It has been shown in [6] that implicit internal image statistics like self similarity are equally rewarding as explicit external image statistics that are learned from a collection of images. Neural network architectures that explicitly make use of these internal natural properties of images should able to produce even better reconstructions. In this work, we make a similar attempt to show that, by augmenting DIP framework with a network that is encouraged to exploit patch recurrence property in natural images significantly improves its reconstruction ability.

2 Methodology

From a Bayesian perspective, the solution $\hat{\boldsymbol{x}}$ can be obtained by solving the following Maximum A Posteriori (MAP) problem,

$$\hat{\boldsymbol{x}} = \underset{\boldsymbol{x}\in\mathbb{R}^n}{\arg\max} \quad \log p(\boldsymbol{y} \mid \boldsymbol{x}) + \log p(\boldsymbol{x}) \tag{2}$$

where $\log p(\boldsymbol{y} \mid \boldsymbol{x})$ represents the log-likelihood of observations $\boldsymbol{y}$, $\log p(\boldsymbol{x})$ is the prior knowledge about the true signal $\boldsymbol{x}$. (2) can be formulated as:

$$\hat{\boldsymbol{x}} = \underset{\boldsymbol{x}\in\mathbb{R}^n}{\arg\min} \quad \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{x}\|_2^2 + \lambda\phi(\boldsymbol{x}) \tag{3}$$

where the solution minimizes an energy function comprising of a fidelity term $\frac{1}{2}\|\boldsymbol{y} - \boldsymbol{x}\|_2^2$, a regularization term $\phi(\boldsymbol{x})$ and a trade-off parameter λ. The fidelity term guarantees that the solution accords with the degradation process, while the regularization term enforces desired property of the output.

In the context of image denoising, the associated optimization for DIP is formulated as:

$$\hat{\boldsymbol{\Theta}} = \underset{\boldsymbol{\Theta}}{\arg\min} \quad \|\boldsymbol{y} - f_{\boldsymbol{\Theta}}(\boldsymbol{z})\|_2^2, \qquad \hat{\boldsymbol{x}} = f_{\hat{\boldsymbol{\Theta}}}(\boldsymbol{z}) \tag{4}$$

where $\boldsymbol{z} \in \mathbb{R}^m$ is input to the network and $f_{\boldsymbol{\Theta}}(\cdot)$ is the deep convolutional neural network parametrized by weights $\boldsymbol{\Theta}$. The regularization here is the network structure itself and early stopping. Indeed, the fact that DIP operates well and recovers high quality images could be perceived as a manifestation of the "correctness" of the chosen architecture to represent image synthesis.

Our aim is to supplement the regularization capability of DIP with the classical property of patch recurrence in natural images. We do this by enforcing the network to operate on patches of a single image rather than the entire image; see Figure 1. Specifically, we aim to minimize the following loss function.

$$\hat{\mathbf{\Theta}} = \arg\min_{\mathbf{\Theta}} \sum_{n=1}^{N} \|\mathcal{P}_n(\boldsymbol{y}) - f_{\mathbf{\Theta}}(\boldsymbol{z}_n)\|_2^2, \qquad \hat{\boldsymbol{x}}_n = f_{\hat{\mathbf{\Theta}}}(\boldsymbol{z}_n) \tag{5}$$

where operator $\mathcal{P}_n(\cdot)$ extracts the n^{th} patch from the input image and $\hat{\boldsymbol{x}}_n$ is the denoised n^{th} patch. Final estimate of the image is reconstructed by combining all the denoised patches, $\hat{\boldsymbol{x}} = \tilde{\mathcal{P}}(\hat{\boldsymbol{x}}_1, \hat{\boldsymbol{x}}_2, ..., \hat{\boldsymbol{x}}_n)$, where $\tilde{\mathcal{P}}(\cdot)$ is the function that reconstructs image back from its patches. Minimizing (5) instead of (4) over all the noisy patches extracted from a single image provides two additional incentives for denoising: (1) it provides regularization in the fitting process as now instead of a single image, network weights are being optimized for multiple smaller image patches which in turn leads to higher impedance to noise (2) this aforementioned regularization encourages the network weights to leverage redundant information from the image patches in the reconstruction task. Henceforth, this patch based approach will be referred to as PatchDIP. Through experiments we show that PatchDIP shows better performance over DIP and offers higher impedance to noise.

3 Experiments

In this section, we evaluate the performance of the proposed approach, PatchDIP, against baseline methods, both qualitatively and quantitatively. For quantitative comparison, we use peak signal to noise ratio (PSNR) and structural similarity index measure (SSIM). All simulations are performed on core-i7 computer (3.40 GHz and 16GB RAM) equipped with Nvidia Titan X GPU using Pytorch framework.

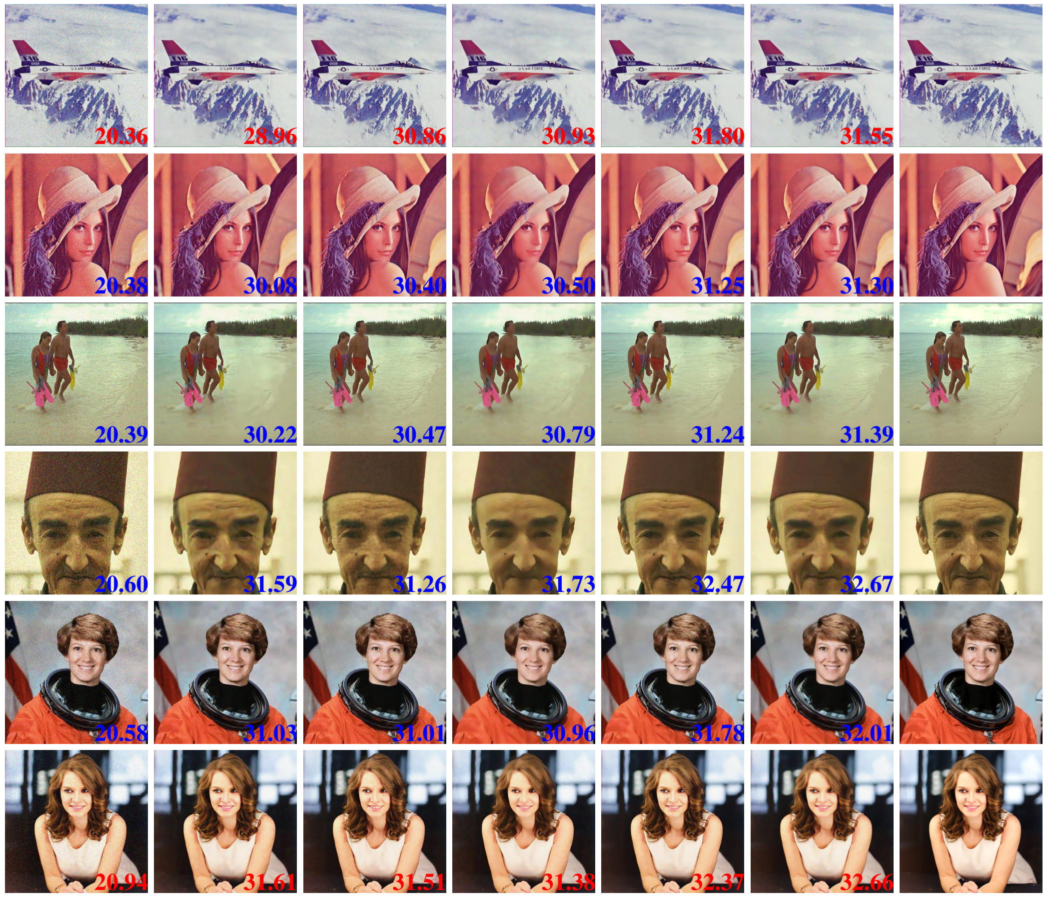
Deep Decoder (DD) [7], Deep Image Prior (DIP), Non-Local Means (NLM), [8] and BM3D [9] are used as baseline methods. We use default algorithmic parameters of all baseline methods unless stated otherwise. As for PatchDIP, we extract overlapping patches of size 64×64 with stride 32×32 from 512×512 sized images. We ran $10,000$ gradient iterations for DD, DIP and PatchDIP and extracted the solution with the best PSNR from the iterates. We also do not employ an exponential running average on the iterates as in DIP paper.

Both qualitative and quantitative results are shown in Figure 2. It can be seen that denoised images via NLM exhibit smoothness due to averaging effect of multiple patches. DIP reconstructions, although sharp as compared to non-local means, still contain noise residuals. On the other hand, the performance of PatchDIP exceeds DD, DIP and NLM, but slightly under performs as compared BM3D.

To provide further insights into PatchDIP, two additional experiments are performed in Figure 3. In the left pane, we show that PatchDIP provides additional impedance to noise over DIP. We plot mean squared error (MSE) versus gradient iterations for DIP and PatchDIP when fitted to random uniform noise. During the first 1000 iterations, PatchDIP shows high impedance to noise. Even though, MSE starts decreasing after 1000 iterations, the convergence is still slower as compared to DIP, indicating that PatchDIP in general fits noise even more reluctantly as compared to DIP. In the right pane MSE is plotted against gradient iterations of PatchDIP for patches extracted from a single image, and patches extracted randomly from a collection of images. Patches extracted from a single image are vividly similar as compared to randomly selected patches and thus, PatchDIP shows quicker and better convergence when patches belong to the same image. This is owing to the fact patches belonging to the same image share redundant information that is leveraged by the deep neural network during the fitting process.

4 Conclusion

To conclude, we propose to leverage the patch recurrence property of natural images alongside DIP. We show that optimizing network weights over patches of images has an additional regularization effect in the fitting process that impedes noisy signals. This regularization coupled with the observation that network fits patches from a single image better over random patches makes a strong case that the network implicitly learns to exploit redundancies among patches, thus leading to a better denoising performance.



(a) Noisy (b) DD (c) DIP (d) NLM (e) PatchDIP (f) BM3D (g) GT

Figure 2: Denosing at $\sigma = 25$.

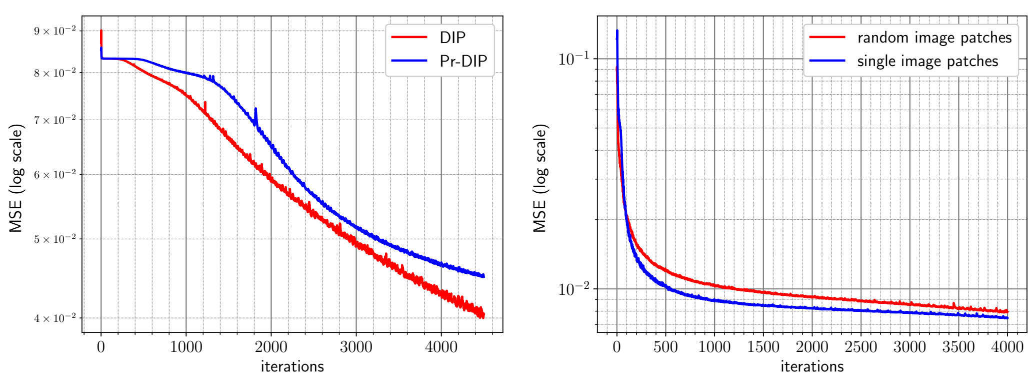


Figure 3: In the left pane, DIP and PatchDIP were fitted to noise sampled from $\mathcal{U}(0, 1)$. It can be seen that PatchDIP exhibits additional resistance over DIP in fitting noise. In the right pane, PatchDIP was fitted on noisy patches from single image and from random images. Better and faster convergence of MSE for patches from the same image provides insight that PatchDIP leverages redundant information in patches.

Acknowledgement

We gratefully acknowledge the support of the NVIDIA Corporation for the donation of NVIDIA TITAN Xp GPU for this research.

References

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

[2] Maurits Malfait and Dirk Roose. Wavelet-based image denoising using a markov random field a priori model. *IEEE Transactions on image processing*, 6(4):549–565, 1997.

[3] Maria Zontak, Inbar Mosseri, and Michal Irani. Separating signal from noise using patch recurrence across scales. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1195–1202, 2013.

[4] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436, 2015.

[5] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 9446–9454, 2018.

[6] Maria Zontak and Michal Irani. Internal statistics of a single natural image. In *CVPR 2011*, pages 977–984. IEEE, 2011.

[7] Reinhard Heckel and Paul Hand. Deep decoder: Concise image representations from untrained non-convolutional networks. *arXiv preprint arXiv:1810.03982*, 2018.

[8] Antoni Buades, Bartomeu 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.

[9] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Transactions on image processing*, 16(8):2080–2095, 2007.