Image Restoration Using Total Variation Regularized Deep Image Prior

Jiaming Liu
Washington University in St. Louis
St. Louis, MO 63105, USA
jiaming.liu@wustl.edu

Yu Sun
Washington University in St. Louis
St. Louis, MO 63105, USA
sun.yu@wustl.edu

Xiaojian Xu*
Washington University in St. Louis
St. Louis, MO 63105, USA
xiaojianxu@wustl.edu

Ulugbek S. Kamilov
Washington University in St. Louis
St. Louis, MO 63105, USA
kamilov@wustl.edu

Abstract

- In the past decade, sparsity-driven regularization has led to significant improvements in image reconstruction. Traditional regularizers, such as total variation (TV), rely on analytical models of sparsity. However, increasingly the field is moving towards trainable models, inspired from deep learning. Deep image prior (DIP) is a recent regularization framework that uses a convolutional neural network (CNN) architecture without data-driven training. This paper extends the DIP framework by combining it with the traditional TV regularization. We show that the inclusion of TV leads to considerable performance gains when tested on several traditional restoration tasks such as image denoising and deblurring.

Background

- Consider the restoration as a linear inverse problem

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{e},$$

where the goal is to reconstruct an unknown image $\mathbf{x} \in \mathbb{R}^N$ from the noisy measurements $\mathbf{y} \in \mathbb{R}^M$. Here, $\mathbf{H} \in \mathbb{R}^{M \times N}$ is a degradation matrix and $\mathbf{e} \in \mathbb{R}^M$ corresponds to the measurement noise, which is assumed to be additive white Gaussian (AWGN) of variance σ^2.

- As practical inverse problems are often ill-posed, it is common to regularize the task by constraining the solution according some prior knowledge.In practice, the reconstruction often relies on the regularized least-squares formulation

$$\mathbf{x}^* = \arg\min_{\mathbf{x}} \left\{ \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_{\ell_2}^2 + \lambda \rho(\mathbf{x}) \right\},$$

where the data-fidelity term ensures the consistency with measurements, and regularizer ρ constrains the solution to the desired image class. The parameter $\lambda > 0$ controls the strength of regularization.

- Total variation (TV) is one of the most widely used image priors that promotes sparsity in image in image gradients [1]. It has been shown to be effective in a number of applications. The ℓ_1-based anisotropic TV is given by

$$\rho_{TV}(\mathbf{x}) \triangleq \sum_{i=1}^{N} |[\mathbf{D_1}\mathbf{x}]_n| + |[\mathbf{D_2}\mathbf{x}]_n|,$$

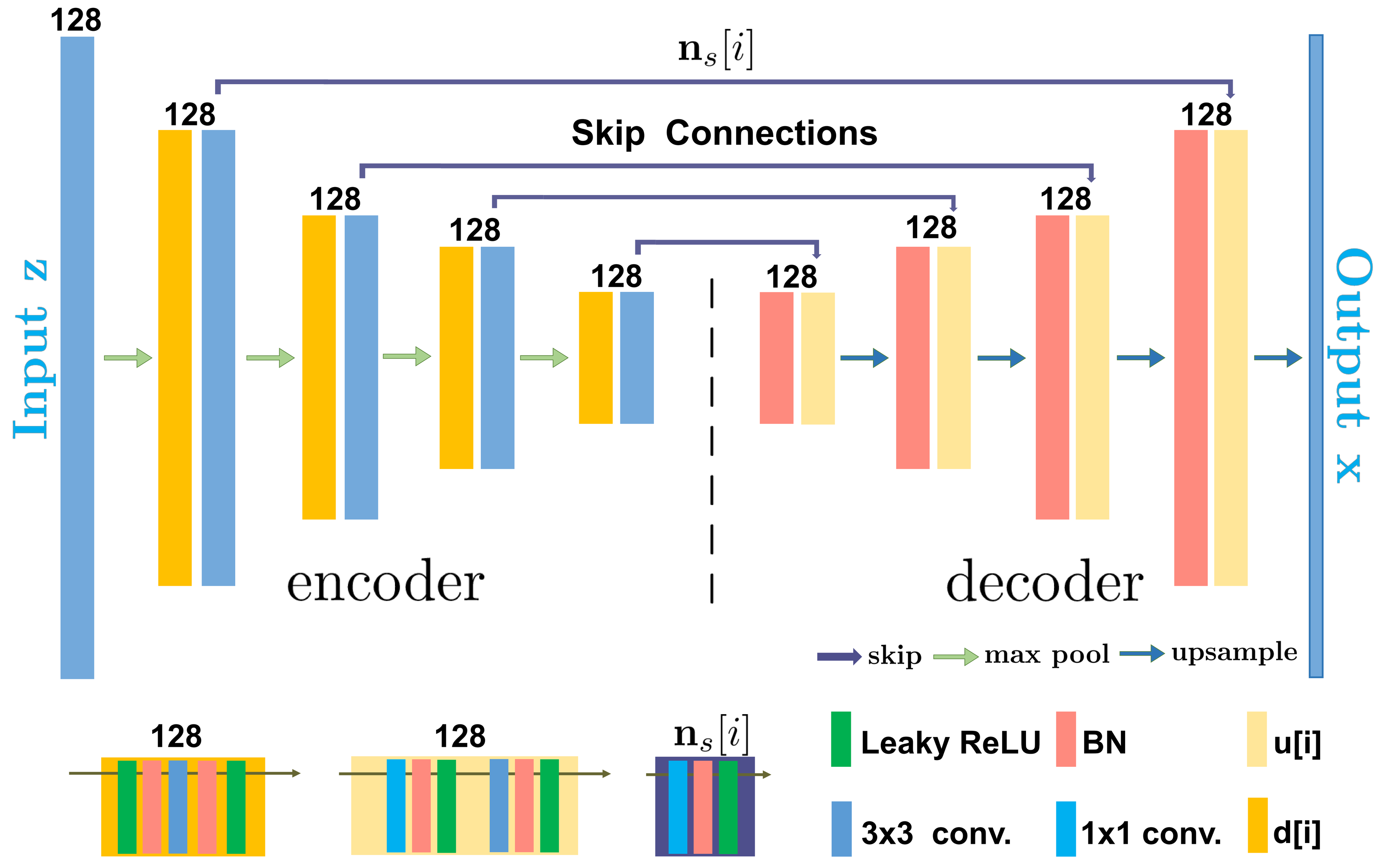
where $\mathbf{D_1}$ and $\mathbf{D_2}$ denote the finite difference operation along the first and second dimension of a two-dimensional (2D) image with appropriate boundary conditions.

- Currently, deep learning achieves the state-of-the-art performance for different image restoration problems. The core idea is to train a CNN via the following optimization

$$\mathbf{\Theta}^* = \arg\min_{\mathbf{\Theta}} \mathcal{L}(f_{\mathbf{\Theta}}(\mathbf{y}), \mathbf{x}),$$

where $f_{\mathbf{\Theta}}(\cdot)$ represents the CNN parametrized by $\mathbf{\Theta}$. $\mathcal{L}$ denotes the loss function.

CNN Architecture [2]



- Recently, Ulyanov *et al.* [2] proposed to use CNN-based methods in an alternative way. They discovered that the architecture of deep CNN models is well-suited for representing natural images, but not random noise. With a random input vector, CNN can reproduce the clear image without supervised training on a large dataset. In the context of image restoration, the associated optimization for DIP can be formulated as

$$\mathbf{\Theta}^* = \arg\min_{\mathbf{\Theta}} \|\mathbf{y} - \mathbf{H} f_{\mathbf{\Theta}}(\mathbf{z})\|_{\ell_2}^2, \text{ such that } \mathbf{x}^* = f_{\mathbf{\Theta}^*}(\mathbf{z}).$$

where $\mathbf{z} \in \mathbb{R}^N$ denotes the random input vector. The CNN generator is initialized with random variables $\mathbf{\Theta}$, and these variables are iteratively optimized so that the output of the network is as close to the target measurement as possible.

Proposed Method

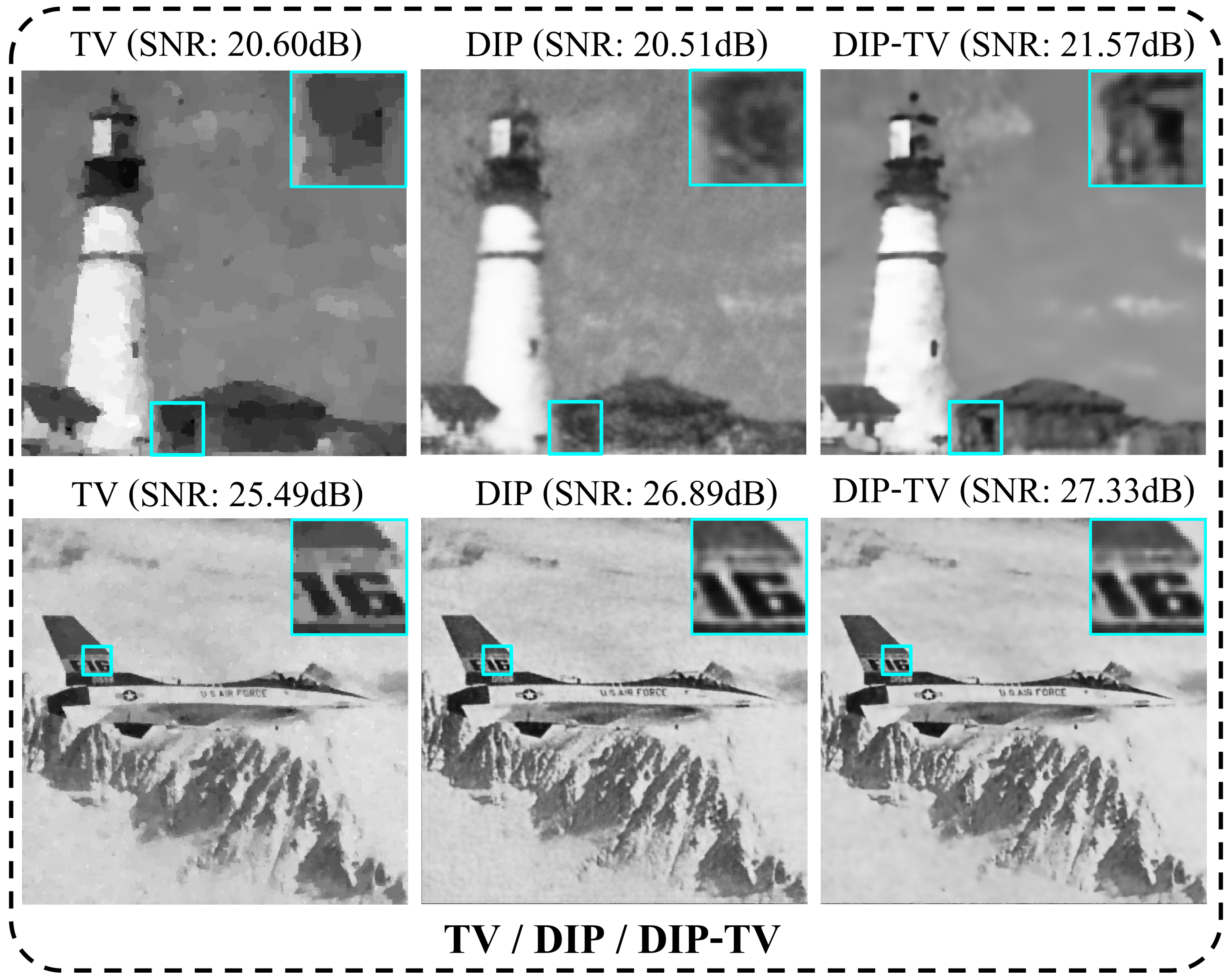
- The goal of DIP-TV is to use the TV regularization to improve the basic DIP approach. We first consider the optimization problem and the objective function of DIP. One can find that the $\|\mathbf{y} - \mathbf{H} f_{\mathbf{\Theta}}(\mathbf{z})\|_{\ell_2}^2$ term corresponds to the data-fidelity term in by replacing $f_\Theta(\mathbf{z})$ with an unknown image output. Thus, we can consider replacing with an optimization problem

$$\Theta^* = \arg\min_{\Theta} \left\{ \|\mathbf{y} - \mathbf{H} f_\Theta(\mathbf{z})\|_{\ell_2}^2 + \lambda \rho_{\mathrm{TV}}(f_\Theta(\mathbf{z})) \right\},$$

$$\text{such that } \mathbf{x}^* = f_{\Theta^*}(\mathbf{z}).$$

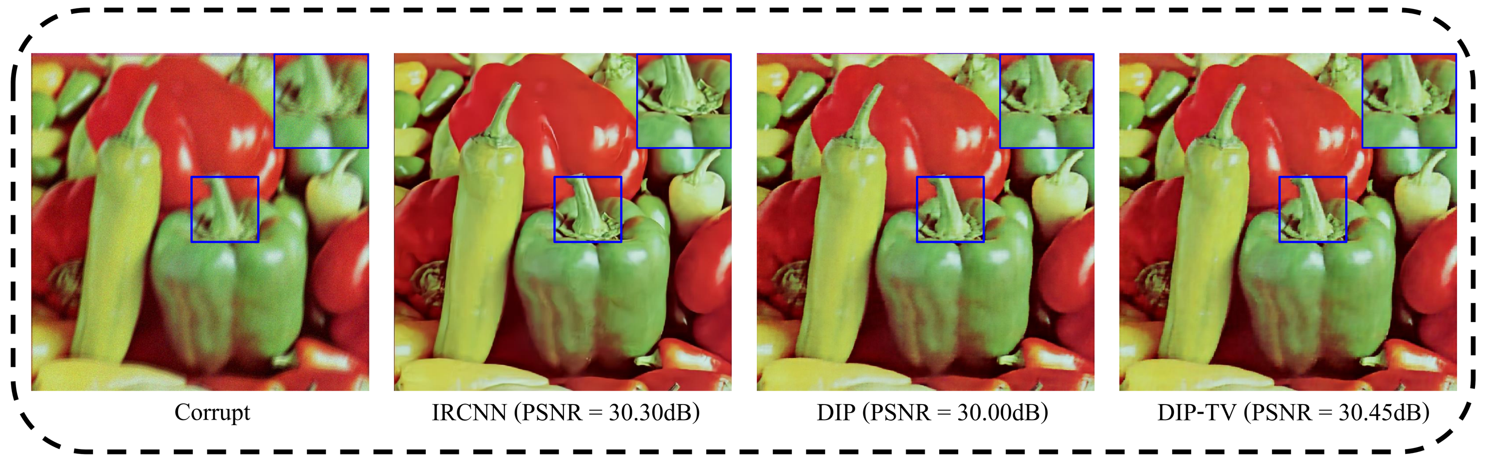
It is similar to training of a CNN and one can rely on any standard optimization algorithms.

Experiments on Image Denoising



TV / DIP / DIP-TV

Experiments on Image Debluring



Conclusion

- This work has presented a simple method, namely DIP-TV, to improve the deep image prior framework, leading to promising performance, equivalent to and sometimes surpassing recently published leading alternatives, such as BM3D and IRCNN. The proposed method is based on the recent idea that a CNN model itself can act as a prior on images and improve sparsity promoting priors via the ℓ_1-norm penalty on the image gradient. The results on images denoising and deblurring demonstrate that TV regularization can further improve on DIP and provides high-quality results.

References

[1] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: nonlinear phenomena*, 1992.

[2] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Deep image prior," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2018.

Funding: This material is based upon work supported by the National Science Foundation (NSF) under Grant No. 1813910.