LEARNED PATCH-BASED REGULARIZATION FOR INVERSE PROBLEMS IN IMAGING

Davis Gilton, Greg Ongie, Rebecca Willett

ABSTRACT

Many modern approaches to image reconstruction are based on learning a regularizer that implicitly encodes a prior over the space of images. For large-scale images common in imaging domains like remote sensing, medical imaging, astronomy, and others, learning the entire image prior requires an often-impractical amount of training data. This work describes a deep image patch-based regularization approach that can be incorporated into a variety of modern algorithms. Learning a regularizer amounts to learning the a prior for image patches, greatly reducing the dimension of the space to be learned and hence the sample complexity. Demonstrations in a remote sensing application illustrates that learning patch-based regularizers produces high-quality reconstructions and even permits learning from a single ground-truth image.

Index Terms— Patch-based methods, deep learning, inverse problems, deblurring, remote sensing

1. INTRODUCTION

Linear image reconstruction is the process of estimating an image $\boldsymbol{x}$ from observed noisy projections of the form

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{\epsilon},$$

where $\boldsymbol{A}$ is a linear model of a measurement system and $\boldsymbol{\epsilon}$ is a noise vector. Example settings include deblurring in remote sensing, k-space measurements in MRI, radar returns, missing pixels, low-resolution measurements, and more. Estimating $\boldsymbol{x}$ from $\boldsymbol{y}$ with knowledge of $\boldsymbol{A}$, which may be underdetermined or ill-posed, has been well-studied in the computational imaging literature. A standard approach is to find an image $\hat{\boldsymbol{x}} \in \mathbb{R}^p$ which satisfies

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \frac{1}{2}||\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}||_2^2 + r(\boldsymbol{x}), \tag{1}$$

where $r(\cdot)$ is a *regularizer* that encourages the solution $\hat{\boldsymbol{x}}$ to have particular desirable properties.

D. Gilton is with the Department of Electrical and Computer Engineering, University of Wisconsin, Madison, WI, 53706 USA (e-mail: gilton@wisc.edu). G. Ongie is with the Department of Statistics, University of Chicago, Chicago, IL, 60637 USA. R. Willett is with the Department of Statistics and Computer Science, University of Chicago, Chicago, IL, 60637 USA. This material is based upon work supported by the National Science Foundation under Grant Nos. 1740707, 1447449, 0353079, and DMS-1925101, by the US Department of Energy under Award No. DE-AC02-06CH11357, as well as AFOSR FA9550-18-1-0166.

Classical reconstruction methods specify a choice of regularizer to promote smoothness, sparsity in some dictionary or basis, or other geometric properties. However, an emerging body of research explores the idea that training data can be used to learn a regularizer, either explicitly or implicitly, using neural networks. Such learned regularizers can be "decoupled", so they are learned in a way which is agnostic to the forward operator [1, 2], or may be learned "end-to-end", where an existing optimization algorithm is unrolled and trained for specific reconstruction tasks [3, 4].

Most past work in this regime has focused on learning a regularizer for whole images, capturing global image geometry. Essentially, this approach learns the *natural image manifold* that we expect $\boldsymbol{x}$ to lie upon [5]. This framework is effective when the images of interest are small, as in standard benchmark training sets. However, for larger images common in remote sensing, medical imaging, astronomy, microscopy, and other applications, we face two key challenges: (1) the underlying image manifold is much higher dimensional because the images themselves lie in a higher dimensional space with more complex geometry, and (2) there is a paucity of large-scale training image datasets.

To address these challenges, this paper proposes learning a patch-based regularizer. Exploiting the geometry of image patches is leveraged in a myriad of image processing tools, including nonlocal means [6], dictionary learning [7, 8], BM3D [9, 10], Gaussian mixture model priors [11, 12], and more. However, state-of-the-art methods for neural network-based image reconstruction do not leverage patch geometry. This work systematically explores patch-based reconstruction and its impact on sample complexity, *i.e.,* how reconstruction error scales with the number of training images.

2. PREVIOUS WORK

Much of the past work in learned regularizers for image reconstruction has origins in optimization theory. Specifically, imagine we had a fixed regularizer $r(\cdot)$ and set $\widehat{\boldsymbol{x}}$ to be the solution to (1). A proximal gradient algorithm [13] starts with an initial estimate $\boldsymbol{x}^{(0)}$ and step size $\eta > 0$ and then computes for $k = 0, 1, 2, \ldots$

$$\overline{\boldsymbol{x}}^{(k)} = \boldsymbol{x}^{(k)} + \eta \boldsymbol{A}^\top(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}^{(k)}) \tag{2}$$

$$\boldsymbol{x}^{(k+1)} = \arg\min_{\boldsymbol{x}} \frac{1}{2}\|\overline{\boldsymbol{x}}^{(k)} - \boldsymbol{x}\|_2^2 + r(\boldsymbol{x}). \tag{3}$$

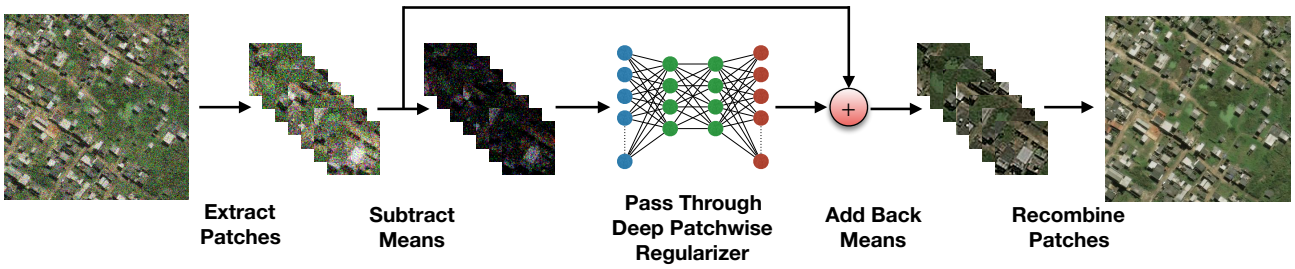


Fig. 1. Visual representation of patchwise regularization using a deep network. The network separately processes N patches of dimension $p \times p \times 3$, where N is the number of (potentially overlapping) patches of size $p \times p$ that are used to represent the original image. By operating on small $p \times p$ patches, the regularizer learns from $N \times n_{samples}$ unique patches during training.

Essentially, this algorithm alternates between (2), a gradient descent step that pushes the current estimate towards a better fit to the data, and (3), a *proximal operator* that finds an estimate in the proximity of $\overline{\boldsymbol{x}}^{(k)}$ that is well-regularized (as measured by $r(\cdot)$). This second step is often thought of as a denoising step. One approach to learning to solve inverse problems is to implicitly learn $r(\cdot)$ by explicitly learning a proximal operator mapping (3) [14, 15].

In a similar vein, Plug-and-play priors [16] and Regularization by Denoising [2] demonstrate that many reconstruction problems can be solved by using denoisers as regularization terms in ADMM, fixed-point, or steepest descent algorithms. These regularizers, in particular, may be learned offline, permitting deep learning methods like TNRD [17], DnCNN [18], or FFDNet [19] to be used.

These and related methods have been explored primarily in the context of small images for which we have large collections of training images. For instance, [18] trains on 3,000 50×50 images. To train for larger image sizes (224×224) and more complex scenes like ImageNet, a state-of-the-art superresolution method [20] requires an even larger dataset of 350,000 images. Such large datasets are not available in many application domains. For instance, [21] trains on 52,850 2-D slices of MRI acquisitions from 350 patients, and [22] trains on 4,096 images in a hyperspectral superresolution setting.

3. DEEP PATCH-BASED LEARNING

The aforementioned learned approaches to image reconstruction require learning a regularization term that can correspond to a prior over the space of images of interest. Unfortunately, learning such priors is plagued by the curse of dimensionality: even for moderately-sized images the number of samples required to learn a full prior is unreasonable.

In contrast, our focus is on learning a prior for image *patches* – small blocks of pixels, such as 8×8 or 16×16. Some prior work has explored leveraging patch geometry. For instance, [23, 24] both train entirely on patches but on special forward models where the operator $\boldsymbol{A}$ can be decomposed across patches; they split the corrupted image into multiple patches and reconstruct those independently. In contrast, this paper describes a framework in which $\boldsymbol{A}$ may be an global operator, such as k-space measurements in MRI or radar returns, but in which the *regularizer* may be decomposed across patches. This is similar in spirit to the patch-based decomposition of regularizers in [25], but while that prior work assumed patches lie along a low-dimensional subspace, here we learn the patch geometry from training images using a neural network.

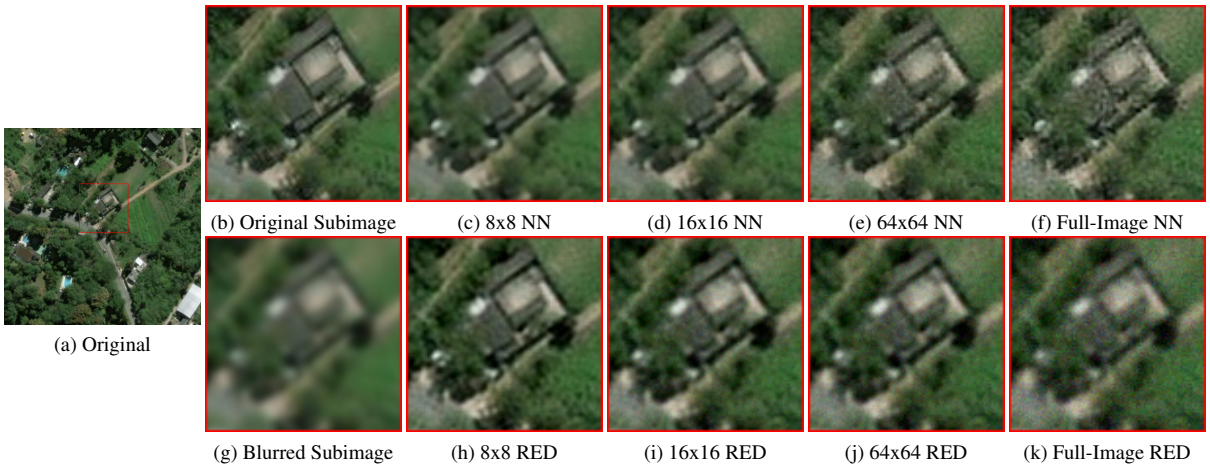
We utilize a flexible procedure in which a deep regularizer is applied in a patchwise manner. Specifically, the regularizer operates by dividing the input image into overlapping patches, subtracting the mean from each patch (a standard preprocessing technique in patch-based methods [26]), and passing each mean-subtracted patch through the learned component (*e.g.,* neural network). The means from the original patches are added to the outputs of the patches, which are then recombined. Figure 1 is a graphical representation of the described procedure. Mathematically, this can be represented by the following expression:

$$R(\boldsymbol{x}) = \mathrm{P}^{-1}\left(\left\{\tilde{R}\left(\mathrm{P}(\boldsymbol{x}, i) - \overline{\mathrm{P}(\boldsymbol{x}, i)}\right) + \overline{\mathrm{P}(\boldsymbol{x}, i)}\right\}_{i=0}^{n}\right). \tag{4}$$

Here, $\mathrm{P}(\cdot, i)$ is a patch extraction operator which outputs the ith patch of an image, and $\mathrm{P}^{-1}(\cdot)$ recombines a set of potentially overlapping patches, where overlapping pixels are averaged. In this case, $\tilde{R}(\cdot)$ is a deep network operating on each patch individually. $\overline{\mathrm{P}(\boldsymbol{x}, i)}$ denotes the average of an individual patch.

In the above regularizer, all processing steps other than patch extraction and aggregation are local to a single patch, and hence are trivially parallizable. The ability to explicitly distribute regularization permits larger minibatch sizes in iterative methods with learned components. Several state-of-the-art learnable iterative methods are trained using minibatch sizes of 1 or 2 because of memory constraints [23, 27, 4, 21], which is much smaller than those used by non-iterative approaches [28, 20].

Patch-based regularizers may be incorporated into and trained as part of an end-to-end image reconstruction tech-



(a) Original (b) Original Subimage (c) 8x8 NN (d) 16x16 NN (e) 64x64 NN (f) Full-Image NN (g) Blurred Subimage (h) 8x8 RED (i) 16x16 RED (j) 64x64 RED (k) Full-Image RED

Fig. 2. Zoomed-in visual comparison of reconstruction quality for a variety of regularization patch sizes for both the Neumann Network (NN) and Regularization by Denoising (RED). Larger patch sizes result in visual artifacts and reduced PSNR. Sub-images are 100×100 pixels, representing a 50 m $\times$ 50 m area on the ground.

nique, e.g. [4], or may be trained alone for use in an iterative reconstruction algorithm like [2]. The former case is straightforward, as a single whole-image network can be replaced by the regularizer in Eq. 4. Training is simple, as gradients can flow along the patch-combination and patch-extraction operators. Training an offline patchwise regularizer typically requires learning a denoising neural network, which is trained using image patches instead of full images.

4. EXPERIMENTS

In this section, we illustrate the effects of learned patch-based regularization on sample complexity, and follow these results by demonstrating that systems with local regularization are able to generalize well even when trained with a single image.

4.1. Training Data and Methods

In all experiments we train and test using subsets of the SpaceNet dataset [29]. We use the AOI1 subset of images, which consists of aerial 3-band imagery of a 2544 square kilometer region of Rio de Janeiro captured by a WorldView-2 satellite system. Training images have a resolution of 50 cm, and are 416x416 pixels in size. The forward model in this case is a Gaussian blur with kernel size 9×9 and variance parameter $\sigma^2 = 2$, where noise is added after blur with standard deviation $\sigma = 0.03$.

Some fraction of the dataset contains images with missing regions. We removed all such images by hand, and selected the training set from the remaining images. All numerical results shown below use a single held-out test set.

For illustration purposes, we use as reconstruction methods the Neumann Network (NN) [4] and Regularization by Denoising (RED) [2]. The Neumann network is an end-to-end trainable method which permits preconditioning and arbitrary learned components. The Neumann network estimator is of the form:

$$\hat{\boldsymbol{x}}_{\text{NN}} = \sum_{i=0}^{n} (\boldsymbol{I} - \eta \boldsymbol{A}^\top \boldsymbol{A} - R(\cdot))^i \left(\eta \boldsymbol{A}^\top \boldsymbol{x}_{\text{init}}\right)$$

where $R(\cdot)$ is a learned regularizer and η a scaling constant. The Neumann network is attractive in the patch-based setting because of its guarantees for data drawn from a union-of-subspaces model, which has been demonstrated to be a good model for small image patches [30, 31].

RED, by contrast, leverages a pretrained learned denoiser, which we denote $\hat{R}(\cdot)$. The RED estimator minimizes a particular energy functional:

$$\hat{\boldsymbol{x}}_{\text{RED}} = \arg\min_{\boldsymbol{x}} ||\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}||_2^2 + \frac{\lambda}{2} \boldsymbol{x}^\top (\boldsymbol{x} - \hat{R}(\boldsymbol{x})) \qquad (5)$$

In our experiments, we use ADMM to minimize (5).

We use FFDNet [19] as a denoiser for RED. FFDNet is a denoising convolutional network that is designed to work with spatially-variable noise levels. We train the FFDNet by learning to denoise image patches, whose size depends on the experiment. At test time, the whole image is regularized by combining patchwise estimates in the manner illustrated by Fig. 1. The learned element of the Neumann network is a two-block residual network with identical structure to the residual learned component used in [21]. The original Neumann network [4] utilized a network based on a residual autoencoder that resembled the U-Net [28], but memory restrictions for the full-image regularizers necessitated a more lightweight network. We chose the learned component of [21] because of

its empirical performance as well as prior use in an unrolled iterative algorithm. The learned component of the Neumann network is trained in an end-to-end fashion, with the regularizer as proposed in the original work replaced with the scheme illustrated in Figure 1.s

We use both methods to demonstrate that the improvements seen below are not specific to end-to-end or pre-trained iterative methods, but instead a result of the improved sample complexity permitted by regularizing locally.

4.2. Sample Complexity

In this section, we explore our previous hypothesis that a deep regularizer that operates on small patches will have better performance in the small-sample case than a comparable regularizer that operates on larger patch sizes.

		8×8	16×16	64×64	Full Image
RED	100	30.01 ± 1.14	29.46 ± 1.16	29.33 ± 1.07	29.32 ± 1.18
	500	31.33 ± 1.22	30.19 ± 1.10	30.96 ± 1.27	29.51 ± 1.29
	1000	33.47 ± 1.26	31.47 ± 1.26	31.12 ± 1.21	30.03 ± 1.32
NN	100	32.64 ± 1.26	31.22 ± 1.35	28.88 ± 1.14	27.08 ± 1.00
	500	32.87 ± 1.24	31.93 ± 1.46	31.09 ± 1.34	27.81 ± 1.01
	1000	32.90 ± 1.25	32.20 ± 1.44	32.19 ± 1.44	29.85 ± 1.17

Table 1. PSNR (dB) comparison for deblurring across several training set sizes (*i.e.,* number of training images) with different regularizer patch sizes. Results are the mean plus and minus the standard deviation of PSNR over a test set of size 64.

In Table 1, the empirical results support out sample complexity hypothesis. While both NN and RED enjoy higher PSNR with increasing training set sizes, decreasing the size of the patches that are input to the regularizer has a dramatic effect on PSNR. **For the Neumann network, shrinking the patch size from 64×64 to 8×8 has comparable effect on PSNR as a $10\times$ increase in training set size.** A qualitative, visual comparison of detail in a test image reconstructed by the networks trained on 1000 images is given in Figure 2.

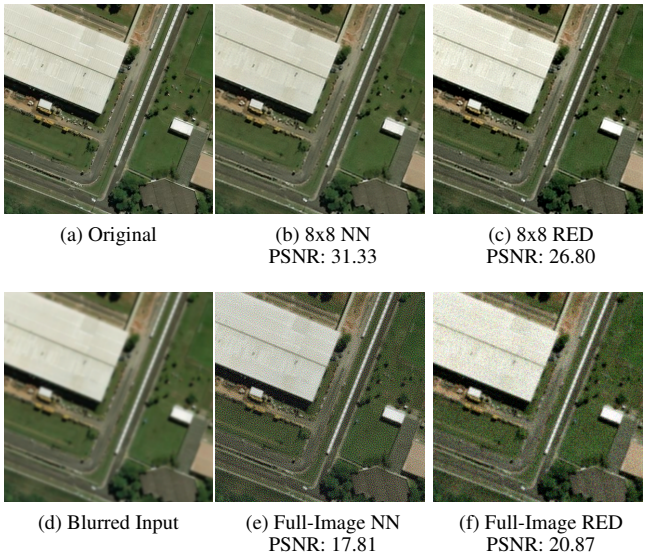
4.3. Single-Image Training

We demonstrate that enforcing locality in a learned regularizer permits training on a single ground truth image while still enjoying competitive reconstruction accuracy. In this experiment, the training set consists only of a single clean image. Training was performed with identical parameter settings as prior experiments. The number of training steps was 1000. The held-out test set was the same as previous experiments.

Table 2 contains test PSNR results. The results shown here are for a single instance of this experiment: while any training image may be chosen, we found empirically it is beneficial to train on an image with a variety of visual features. Fig. 3 contains some sample reconstructions of an image from the test set. We find using patchwise regularizers produces high-quality reconstructions, while regularizers operating on the full image result in noisy artifacts that are likely a result of overfitting to the single training image.

	8×8	Full Image
RED	26.60 ± 1.54	21.02 ± 0.57
NN	31.90 ± 1.42	18.34 ± 1.31

Table 2. PSNR (dB) comparison for single-training-image reconstruction. When there is just one training image, local regularization does not overfit, unlike full-image regularization. Results are the mean plus and minus the standard deviation over test set of size 64.



(a) Original (b) 8x8 NN PSNR: 31.33 (c) 8x8 RED PSNR: 26.80

(d) Blurred Input (e) Full-Image NN PSNR: 17.81 (f) Full-Image RED PSNR: 20.87

Fig. 3. Single-image Training Reconstruction. Local regularization enables competitive reconstruction quality with a single training image, while unrestricted training on a single image results in a poor, noisy reconstruction. The inverse problem is Gaussian deblurring with kernel size 9×9, variance $\sigma^2 = 2$, and noise level 0.03.

5. CONCLUSION

In this paper, we propose a deep patchwise regularizer for use in iterative solutions to inverse problems. By regularizing in a patchwise manner, fewer samples are needed to learn the lower-dimensional prior distribution over the space of image patches, which is demonstrated through experiment. Restricting a regularizer to operate in a patchwise manner sacrifices the ability to learn long-range correlations, but empirically improves performance in the small-sample setting. Patchwise regularizers are not limited to end-to-end-trained methods or pretrained regularizers, and in both cases permit learning from even a single training image.

6. REFERENCES

[1] J. Sun, H. Li, and Z. Xu, "Deep ADMM-Net for compressive sensing MRI," in *Advances in Neural Information Processing Systems (NeurIPS)*, pp. 10–18, 2016.

[2] Y. Romano, M. Elad, and P. Milanfar, "The little engine that could: Regularization by denoising (red)," *SIAM Journal on Imaging Sciences*, vol. 10, no. 4, pp. 1804–1844, 2017.

[3] J. Adler and O. Öktem, "Learned primal-dual reconstruction," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1322–1332, 2018.

[4] D. Gilton, G. Ongie, and R. Willett, "Neumann networks for inverse problems in imaging," *arXiv preprint arXiv:1901.03707*, 2019.

[5] A. Bora, A. Jalal, E. Price, and A. G. Dimakis, "Compressed sensing using generative models," in *International Conference on Machine Learning (ICML)*, pp. 537–546, 2017.

[6] 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)*, vol. 2, pp. 60–65, IEEE, 2005.

[7] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online dictionary learning for sparse coding," in *International Conference on Machine Learning (ICML)*, pp. 689–696, ACM, 2009.

[8] M. Aharon, M. Elad, and A. Bruckstein, "K-svd: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on signal processing*, vol. 54, no. 11, pp. 4311–4322, 2006.

[9] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising with block-matching and 3d filtering," in *Image Processing: Algorithms and Systems, Neural Networks, and Machine Learning*, vol. 6064, p. 606414, International Society for Optics and Photonics, 2006.

[10] W. Marais and R. Willett, "Proximal-gradient methods for Poisson image reconstruction with BM3D-based regularization," in *IEEE 7th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, pp. 1–5, 2017.

[11] D. Zoran and Y. Weiss, "From learning models of natural image patches to whole image restoration," in *2011 International Conference on Computer Vision*, pp. 479–486, IEEE, 2011.

[12] A. M. Teodoro, J. M. Bioucas-Dias, and M. A. Figueiredo, "A convergent image fusion algorithm using scene-adapted gaussian-mixture-based denoising," *IEEE Transactions on Image Processing*, vol. 28, no. 1, pp. 451–463, 2018.

[13] N. Parikh and S. Boyd, "Proximal algorithms," *Foundations and Trends® in Optimization*, vol. 1, no. 3, pp. 127–239, 2014.

[14] J. Rick Chang, C.-L. Li, B. Poczos, B. Vijaya Kumar, and A. C. Sankaranarayanan, "One network to solve them all–solving linear inverse problems using deep projection models," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 5888–5897, 2017.

[15] J. R. Chang, C.-L. Li, B. Póczos, and B. V. Kumar, "One network to solve them all — Solving linear inverse problems using deep projection models," in *IEEE International Conference on Computer Vision (ICCV)*, pp. 5889–5898, 2017.

[16] S. V. Venkatakrishnan, C. A. Bouman, and B. Wohlberg, "Plug-and-play priors for model based reconstruction," in *2013 IEEE Global Conference on Signal and Information Processing*, pp. 945–948, IEEE, 2013.

[17] 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*, vol. 39, no. 6, pp. 1256–1272, 2016.

[18] 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*, vol. 26, no. 7, pp. 3142–3155, 2017.

[19] K. Zhang, W. Zuo, and L. Zhang, "FFDNet: Toward a fast and flexible solution for CNN-based image denoising," *IEEE Transactions on Image Processing*, vol. 27, no. 9, pp. 4608–4622, 2018.

[20] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. P. Aitken, A. Tejani, J. Totz, and Z. Wang, "Photo-realistic single image super-resolution using a generative adversarial network.," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4681–4690, 2017.

[21] M. Mardani, Q. Sun, S. Vasawanala, V. Papyan, H. Monajemi, J. Pauly, and D. Donoho, "Neural proximal gradient descent for compressive imaging," *arXiv preprint arXiv:1806.03963*, 2018.

[22] L. Liebel and M. Körner, "Single-image super resolution for multispectral remote sensing data using convolutional neural networks," *ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 41, pp. 883–890, 2016.

[23] C. Qin, J. Schlemper, J. Caballero, A. N. Price, J. V. Hajnal, and D. Rueckert, "Convolutional recurrent neural networks for dynamic mr image reconstruction," *IEEE transactions on medical imaging*, vol. 38, no. 1, pp. 280–290, 2018.

[24] E. Bostan, U. S. Kamilov, and L. Waller, "Learning-based image reconstruction via parallel proximal algorithm," *IEEE Signal Processing Letters*, vol. 25, no. 7, pp. 989–993, 2018.

[25] W. Dong, G. Shi, X. Li, Y. Ma, and F. Huang, "Compressive sensing via nonlocal low-rank regularization," *IEEE Transactions on Image Processing*, vol. 23, no. 8, pp. 3618–3632, 2014.

[26] M. Lebrun, M. Colom, A. Buades, and J.-M. Morel, "Secrets of image denoising cuisine," *Acta Numerica*, vol. 21, pp. 475–576, 2012.

[27] H. K. Aggarwal, M. P. Mani, and M. Jacob, "MoDL: Model based deep learning architecture for inverse problems," *IEEE Transactions on Medical Imaging*, 2018.

[28] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 234–241, Springer, 2015.

[29] A. Van Etten, D. Lindenbaum, and T. M. Bacastow, "Spacenet: A remote sensing dataset and challenge series," *arXiv preprint arXiv:1807.01232*, 2018.

[30] J. Silva, M. Chen, Y. C. Eldar, G. Sapiro, and L. Carin, "Blind compressed sensing over a structured union of subspaces," *arXiv preprint arXiv:1103.2469*, 2011.

[31] J. Salmon, Z. Harmany, C.-A. Deledalle, and R. Willett, "Poisson noise reduction with non-local PCA," *Journal of mathematical imaging and vision*, vol. 48, no. 2, pp. 279–294, 2014.