Compressive Image Recovery Using Recurrent Generative Model

Akshat Dave, Anil Kumar Vadathya, Kaushik Mitra
{ee12b073, ee15s055, kmitra}@ee.iitm.ac.in
Indian Institute of Technology, Madras

Abstract

Generative models are considered as the swiss knives for data modelling. In this paper we leverage the recently proposed recurrent generative model, RIDE, for applications like image inpainting and compressive image reconstruction. Recurrent networks can model long range dependencies in images and hence are suitable to handle global multiplexing in reconstruction from compressive imaging. We perform MAP inference with RIDE as prior using backpropagation to the inputs and projected gradient method. We propose an entropy thresholding based approach for preserving texture well. Our approach shows comparable results for image inpainting task. It shows superior results in compressive image reconstruction compared to traditional methods D-AMP and TVAL3 which uses global prior of minimizing TV norm.

1. Introduction

Recent advances in deep learning especially Convolutional Neural Nets (CNN) have led to a revolution in many computer vision tasks like object recognition [14] and detection [8]. Given the learning capabilities of CNNs, they have been used for image processing tasks like denoising [20], compressive signal recovery [15], [22] and image deconvolution [29] with good success. In all these recent approaches, the network is trained in a discriminative fashion, i.e., it is trained to learn a function mapping inputs to desired outputs provided through training data. Although this makes the network very efficient during run time, it results in a task specific network which is not appealing for many image processing problems. Since this approach would require us to learn a different network for each possible image degradations like noisy, compression artifacts and blur etc. [24] and with every changing random sensing matrix in compressive sensing application [15]. Given the ambiguity with the design aspects of CNNs like convolutional filter size and number of layers to stack (depth), it might be a tedious job to train again and again. We believe that deep generative image models can come handy here because once we learn a generative image prior, it can be used for solving any image degradation problem which can be framed as
noisy linear model and do Bayesian inference for image recovery. This avoids the need to relearn the network for each new task.

Advances in complex learning methods and training deeper architectures have rekindled the interest in generative models [13, 9]. Especially using recurrent models like Spatial Long Short Term Memory (SLSTM)[10] units there has been a good advancement both in terms of loglikelihood scores and visually appealing samples generated [26, 28]. Added to their expressiveness, these recurrent generative models being directed result in exact likelihood evaluation which is not possible with their counterpart undirected models like Restricted Boltzmann Machines (RBM)s [11]. Also this likelihood evaluation facilitates exact inference. More importantly, these models capture the long range dependencies in images very well as is demonstrated through samples generated containing crisp variation and being coherent [28]. In this work we propose to use one such model Recurrent Image Density Estimator (RIDE) [26] as an image prior for compressive image recovery. Our contributions are as follows:

- We utilize RIDE’s ability as an image prior to model long term dependencies for reconstructing compressively sensed images.
- We use backpropagation to inputs while doing gradient ascent for MAP inference.
- We hypothesize that the model’s uncertainty in prediction can be related to the entropy of component posterior probabilities. By thresholding the entropy, we enhance texture preserving ability of the model.

2. Prior Work

Role of Signal Priors: Image data priors have played a significant role for signal reconstruction from ill posed problems which are very common in image processing and computational photography. Initially such image priors were constructed through empirical observation of data statistics, for example TV norm minimization, sparse gradient prior [16] and sparsity of coefficients in wavelet domain [23]. On the other hand, many methods were proposed to learn the priors directly from data such as dictionary learning [1], mixture models like GMMs [30] and their variants GSMs [23], conditional models like Mixture of Conditional GSMs (MCGSM) [27], undirected models like Field of Experts (FoEs) [25]. In dictionary learning an overcomplete set of basis is learnt by representing natural image patches as sparse linear combination of these basis. It has been successfully applied for many image processing tasks [19, 1]. On the contrary, rest of the approaches explicitly model the data distribution by maximizing likelihood. GMMs are quite popular image patch priors and have been used for restoration tasks like image denoising and de-blurring [30] where it gives competitive results compared to state-of-the-art methods like BM3D [5] and KSVD [1]. FoEs [25] is another popular model which is a Product of Experts (PoEs) with the desirable property of translational invariance making it a whole image prior. It has been used for image inpainting and denoising.

Deep Nets for Image Processing: Many recent approaches have been proposed to use feed forward deep networks for image reconstruction problems. Burget et al. [3] used Multilayer perceptrons (MLP) for image patch denoising performing on par with BM3D. Mao et al. [20] used very deep convolutional encoder-decoder with skip connections for image denoising even handling different levels of Gaussian noise. It has surpassed BM3D’s performance. Xu et al. [29] have used convnets for image deconvolution. Kulkarni et al. [15] trained a convnet, termed as ReconNet, to recover image from compressed measurements of image patches with measurements being as low as 1%. Although these feed forward discriminative models are very fast at run time, their application is limited to the task they are trained for. Burget et al. [3] reported difficulty in generalizing a MLP network trained at a particular noise level for different levels of Gaussian noise. Mao et al. [20] handle this but at the cost of a huge network. ReconNet proposed for CS signal recovery requires the network to be trained again and again for each different sensing matrix and at each different measurement rate.

Deep Generative Models: Owing to the inherent problems posed by discriminative models, recently much effort has gone into building generative models such as, Generative Adversarial Nets (GAN) [9], Variational Auto Encoders (VAE) [13], Pixel Recurrent Neural Networks (PixelRNN) [28] and Recurrent Image Density Estimator (RIDE) [26]. GANs learn the ability to generate a plausible sample from the distribution of natural images. VAE provides a probabilistic framework for both encoding data to latent representation and decoding from it. Auto regressive models like RIDE model the current pixel distribution conditioned on the causal context where Spatial Long Short Term Memory (SLSTM) [10] units are used to obtain the contextual summary. PixelRNN is also an auto regressive model like RIDE but with much more complex architecture achieving the state-of-the-art performance in terms of loglikelihood scores. Apart from being expressive, RIDE and PixelRNN come with added advantages. Their directed nature facilitates the computation of exact likelihood. Also, these priors being auto regressive aren’t limited to patch size, as is the case with discriminative and even non deep generative models. This is very useful particularly in cases like single pixel camera where the reconstruction has to take account of global multiplexing and patch based methods can’t be used directly.
Among these deep generative models we find RIDE particularly suitable as low level image prior for our tasks involving Bayesian inference. GANs don’t model the data distribution and VAE doesn’t provide the exact likelihood. PixelRNN although models the distribution, it discretizes the distribution of a pixel to 256 intensity values resulting in optimization difficulties. In this work we extend RIDE as an image prior for reconstruction problems in compressive sensing and image inpainting.

Inpainting: Image inpainting has been previously attempted with image priors. FoEs were applied to remove scratches or unwanted effects like text from an image. This was extended to include a prior for image inpainting. Dictionary learning [18] has also been proposed for image inpainting although not ideal since it is patch based. A multiscale adaptive version of dictionary learning [19] is shown to perform well.

Single Pixel Camera: SPC [7] is a compressive sensing framework [4], where the goal is to reconstruct the image back from a very less number of random linear measurements. Typically this is an ill-posed problem and hence we need to use signal priors. Initially algorithms were proposed to minimize the $l_1$ norm assuming sparsity in the domain of wavelet coefficients, DCT coefficients or gradients [17]. Later class of algorithms known as approximate message passing (AMP) algorithms [6] [21] use off-the-shelf denoiser to iteratively refine their solution. ReconNet is another recent method using CNNs. But it can only handle local multiplexing since it is a patch based approach. Here we propose to do compressive image reconstruction with recurrent generative model RIDE as the image prior. Since it is not patch limited, we can handle global multiplexing.

3. Background

Let $x$ be a gray scale-image and $x_{ij}$ be the pixel intensity at location $ij$ then $x_{<ij}$ describes the causal context around that pixel containing all $x_{mn}$ such that $m \leq i$ and $j < n$. Now the joint distribution over the image can be factorized as follows:

$$p(x) = \prod_{ij} p(x_{ij}|x_{<ij}, \theta_{ij})$$

where $\theta_{ij}$ are distribution parameters at that location. By making the Markov assumption we can limit the extent of $x_{<ij}$ to a smaller neighbourhood. Another valid assumption is stationarity of the data which results in sharing the same parameters $\theta$ across all locations $ij$, thus achieving translational invariance.

Now each factor in the above equation can be modeled by a mixture of GSMs with shared parameters $\theta$ which makes it Mixture of Conditional Gaussian Scale Mixtures (MCGSM) as proposed by [27],

$$p(x_{ij}|x_{<ij}, \theta) = \sum_{c,s} p(c, s|x_{<ij}, \theta)p(x_{ij}|x_{<ij}, c, s, \theta),$$

Where the sum is over components and scales,

$$p(c, s|x_{<ij}) \propto \exp(\eta_{cs} - 0.5 * e^{\alpha cs}x_{<ij}^T K_c x_{<ij}),$$

$$p(x_{ij}|x_{<ij}, c, s) = \mathcal{N}(x_{ij}; \alpha_s x_{<ij}, e^{-\alpha cs})$$

In MCGSM, Markov assumption was made and the past context $x_{<ij}$ was actually limited to a small causal neighborhood. However natural images exhibit long range correlations and any smaller neighbourhood fails to capture them. On the other hand increasing the neighbourhood leads to dramatic increase in number of parameters. In order to take into account such dependencies [26] have proposed to use two dimensional Spatial Long Short Term Memory (LSTMs) [10] units for summarizing the causal context through their hidden representation $h_{ij}$ at location $ij$ as,

$$h_{ij} = f(x_{<ij}, h_{ij-1}, h_{i,j-1})$$

where $f$ is a complex non linear function with memory elements analogous to physical read, write and erase elements thus giving it the ability to model the long term dependencies in sequences. This formulation results in replacement
of the finite context $x_{<ij}$ in conditional modeling equation (2) with $h_{ij}$, thus bringing in the summary of entire causal context. Thus, the complete model is specified as follows:

$$p(x) = \prod_{ij} p(x_{ij}|h_{ij}, \theta) \quad (5)$$

$$p(x_{ij}|h_{ij}, \theta) = \sum_{c,s} p(c,s|h_{ij}, \theta)p(x_{ij}|c,s,h_{ij}, \theta) \quad (6)$$

Using Recurrent Image Density Estimator (RIDE) [26] have achieved one of the state-of-the-art results in terms of log-likelihood scores. For more details we recommend the reader to go through [26].

4. Compressive Image Recovery Using RIDE

Here we consider the problem of image restoration from linearly compressed measurements $y = Ax + n$, where the linear transformation $A$ is a $M \times N$ with $M < N$, $n$ is noise in the observation with known statistics.

4.1. MAP Inference via Backpropagation

Sequential sampling of the conditional factors has been used by RIDE to generate image samples from the joint distribution [26]. On similar lines, one method to do inference is to sample from the posterior distribution. But here sequential sampling is not possible and we have to resort to Markov Chain Monte Carlo methods such as Gibbs sampling which are computationally expensive even for smaller image sizes. Hence, we use Maximum-A-Posteriori principle to find the desired image $\hat{x}$,

$$\hat{x} = \arg \max_x p(x|y) = \arg \max_x p(x)p(y|x) \quad (7)$$

The prior term $p(x)$ is specified by the generative model (5),(6) and the likelihood is given by $p(y|x) \propto \exp(-||y - Ax||^2/\sigma^2)$ for the isotropic Gaussian noise case.

We apply gradient ascent to the net posterior distribution in order to obtain the reconstructed image. After log transforming the product in (7), the gradient with respect to the prior is given by:

$$\frac{\partial \log p(x)}{\partial x_{ij}} = \sum_{k \geq i, l \geq j} \frac{\partial \log p(x_{kl}|h_{kl}, \theta)}{\partial x_{ij}} \quad (8)$$

Due to the recurrent nature of the model, each pixel through its hidden representation can contribute to the likelihood of all the pixels that come after it in forward pass. In a similar fashion during backward pass the gradient from each pixel propagates to all the pixels prior to it in the sequence. Gradients with respect to log-likelihood are much easier to evaluate is given by:

$$\nabla_x \log p(y|x) \propto 2A^T(y - Ax) \quad (9)$$

4.2. Tricks used for inference

4.2.1 Four directions

Joint distribution (5) can be factorized in multiple ways, for example along each of the four diagonal directions of an image, i.e., top-right, top-left, bottom-right and bottom-left. Gradients from different factorizations are considered at each iteration of the inference, by flipping the image in the corresponding direction. This leads to faster convergence as compared to just considering one direction. While doing the inference on crops from randomly sampled BSDS test images, we observe that the convergence rate is roughly 2 times faster when considering four directions.

4.2.2 Entropy-based Thresholding

While solving the MAP optimization, we observed that we can recover the edges quite well but texture regions are blurred. This happens because the RIDE model may not have the right mixture component (see (6)) to explain the
Figure 4: Inpainting comparisons: We compare our approach with the multiscale dictionary learning approach [19]. Our method is able to recover the sharp edges better than the multiscale KSVD approach, as is evident in the zoomed region around zebra’s eye (top) and elephant’s head (bottom). This is because our method is a global prior as compared to the patch-based multiscale KSVD approach. The numbers mentioned below the figures are PSNR(left) and SSIM(right).

latent texture. In such cases, all the mixture components can be chosen with almost uniform probability, resulting in blurred texture. To detect such cases, in each iteration, we consider the posterior probability of scales and components in RIDE at each point as a metric to understand how confident the model is in modeling the distribution at that point. This is evaluated through posterior entropy given as,

$$H(i, j) = -\sum_{c,s} p(c,s|x_{<ij}, x_{ij}) \log(p(c,s|x_{<ij}, x_{ij}))$$

(11)

If the point lies on an edge, the posterior entropy is low as there are only certain selected components which can explain that edge. Whereas, if the point lies in a flat or textured patch, the posterior entropy is high and the point is equiprobable to come from different components and scales. Therefore, to reduce blurring we maintain a threshold on posterior entropy above which we clip the gradients to zero. Figure 3 shows the effect of entropy constraint on the texture reconstruction.

4.3. Compressive Image Recovery

To demonstrate the effectiveness of our method, we consider the problems of image inpainting and compressive sensing imaging [2]. In image inpainting our goal is to recover the missing pixels from a randomly masked image. We estimate the missing pixels by maximizing the prior over missing pixels, keeping the observed pixels constant. This is done by updating the gradients for only missing pixels. We have used the above mentioned entropy based gradient thresholding to avoid blurring the texture region. For SPC, we formulated the MAP inference as,

$$\hat{x} = \arg\max_{x} p(x) \text{ s.t. } y = \Phi x$$

(12)

To optimize the above we use projected gradients method, where after each gradient update solution is projected back on to the affine solution space for $y = \Phi x$. Every $k$-th iteration consists of the following two steps.

$$\hat{x}_k = x_{k-1} + \eta \nabla x_{k-1} p(x)$$

(13)

$$x_k = \hat{x}_k - (\Phi \Phi^T)^{-1} (\Phi \hat{x}_k - y)$$

(14)

In our experiments we consider row orthonormalized $\Phi$ and the term $(\Phi \Phi^T)^{-1}$ reduces to identity matrix.

5. Experiments

For training the RIDE model we have used publicly available Berkeley Segmentation dataset (BSDS300). Following the instincts from [26], we trained the model with increasing patch size in each epoch. Starting with 8x8 patch we go till 22x22 in steps of 2 for 8 epochs. We used the code provided by authors of RIDE in caffe available here\(^1\). We

\(^1\)https://github.com/lucastheis/ride/
start with a very low learning rate (0.0001) and decrease it to half the previous value after every epoch. We used Adam optimization [12] for training the model. We observe that models with more than one spatial LSTM layer, although gave slight improvement in loglikelihood score, don’t result in much of improvement for our tasks of interest. Also each additional layer results in lot of computational overhead given the sequential nature of the model. Hence we proceed with a single layer RIDE model for all the inference tasks mentioned in this paper. In order to accommodate for boundary issues we remove a two pixel neighbourhood around the image for PSNR and SSIM calculation also in the figures of SPC and inpainting experiments. For a fair comparison, we also do the same for the results of other methods.

5.1. Image Inpainting

For image inpainting, we randomly removed 70% of pixels and estimated them using aforementioned inference method. We compared our approach with the multiscale adaptive dictionary learning approach [19], which is an improvement over the KSVD algorithm, see Figure 4. It is clear from the figure that our approach is able to recover the sharp edges better than the multiscale KSVD approach. This is because our method is based on global image prior as compared to the patch-based multiscale KSVD approach.

5.2. Single Pixel Camera

In general, the SPC framework involves global multiplexing of the scene. But the recently proposed state-of-the-
art methods for signal reconstruction, like ReconNet, are designed for local spatial multiplexing and can’t handle the global multiplexing case directly. Our model, using Spatial LSTMs, can reason for long term dependencies in image sequences and is preferable for such kind of tasks. We show SPC reconstruction results on some randomly chosen images from the BSDS300 test set which were cropped to $160 \times 160$ size for computational feasibility, see Figure 5. We generate compressive measurements from them using random Gaussian measurement matrix with orthonormalized rows. We take measurements at four different rates 0.4, 0.3, 0.25 and 0.15. Using the projected gradient method, we perform gradient ascent for 200 iterations for 0.4 and 0.3 measurement rates. For lower measurement rates, we run gradient ascent for 400 iterations. Also, we follow the entropy thresholding procedure mentioned in section 4.2.2 with a threshold value of 3.5 which we empirically found to be good for preserving textures. In all the cases, we start with a random image uniformly sampled from $(0,1)$. Reconstruction results for five images are shown in Table 1 and Figure 6. We were able to show improvements both in terms of PSNR and SSIM values for different measurement rates. Even at low measurement rates, our method preserves the sharp and prominent structure in the image. D-AMP has the tendency to over-smooth the image, whereas TV AL3 adds blotches to even the smooth parts of the image. Figure 2 shows the reconstructions for a color image where the measurements are taken from individual color channels. We can notice that even at very low measurement rates the overall reconstruction is good, although fine details are lost.

6. Conclusions and Future Work

We demonstrate that deep recurrent generative image models such as RIDE can be used effectively for solving compressive image recovery problems. The main advantages of using such deep generative models is that they are global priors and hence can model long term image dependencies. Also using the proposed MAP formulation we can solve many other image restoration tasks such as image de-blurring, superresolution, demosaicing and computational photography problems such as coded aperture and exposure. Another direction of future work would be to adapt the trained generative model to the specific image that we are interested in restoring. We can use our entropy thresholding step to detect which part of the image is not modeled well by the generic model and then adapt the model accordingly.

References

[1] M. Aharon, M. Elad, and A. Bruckstein. K-svd: An algorithm for designing overcomplete dictionaries for sparse representation. IEEE Transactions on Signal Processing, 54(11):4311, 2006.

[2] R. G. Baraniuk. Compressive sensing. IEEE Signal Processing Magazine, 24(4), 2007.

[3] H. C. Burger, C. J. Schuler, and S. Harmeling. Image denoising: Can plain neural networks compete with bm3d? In Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, pages 2392–2399. IEEE, 2012.

[4] E. J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. IEEE Transactions on Information Theory, 52(2):489–509, 2006.

[5] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Bm3d image denoising with shape-adaptive principal component analysis. In SPARS’09-Signal Processing with Adaptive Sparse Structured Representations, 2009.

[6] D. L. Donoho, A. Maleki, and A. Montanari. Message-passing algorithms for compressed sensing. Proceedings of the National Academy of Sciences, 106(45):18914–18919, 2009.

[7] M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. E. Kelly, R. G. Baraniuk, et al. Single-pixel imaging via compressive sampling. IEEE Signal Processing Magazine, 25(2):83, 2008.

[8] R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 580–587, 2014.

[9] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In Advances in Neural Information Processing Systems, pages 2672–2680, 2014.

[10] A. Graves. Neural networks. In Supervised Sequence Labelling with Recurrent Neural Networks, pages 15–35. Springer, 2012.

[11] G. E. Hinton, T. J. Sejnowski, and D. H. Ackley. Boltzmann machines: Constraint satisfaction networks that learn. Carnegie-Mellon University, Department of Computer Science Pittsburgh, PA, 1984.

[12] D. Kingma and J. Ba. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980, 2014.

[13] D. P. Kingma and M. Welling. Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114, 2013.

[14] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems, pages 1097–1105, 2012.

[15] K. Kulkarni, S. Lohit, P. Turaga, R. Kerviche, and A. Ashok. Reconnet: Non-iterative reconstruction of images from compressively sensed measurements. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 449–458, 2016.

[16] A. Levin, R. Fergus, F. Durand, and W. T. Freeman. Deconvolution using natural image priors. Massachusetts Institute of Technology, Computer Science and Artificial Intelligence Laboratory, 2007.

[17] C. Li, W. Yin, H. Jiang, and Y. Zhang. An efficient augmented lagrangian method with applications to total varia-
Reconstruction with 30% measurement rate

| Original image | TVAL3     | DAMP      | Ours      |
|----------------|-----------|-----------|-----------|
|                | 26.2, 0.789| 31.66, 0.923| 33.02, 0.919 |
|                | 30.20, 0.849| 30.17, 0.858 | 34.70, 0.919 |

Reconstruction with 15% measurement rate

|                | 22.63, 0.638| 24.5, 0.757 | 25.32, 0.754 |
|                | 25.44, 0.719| 25.81, 0.70  | 25.88, 0.798 |

Figure 6: Images obtained by reconstruction from compressive measurements using D-AMP, TVAL3 and our method. Even at low measurement rates, our method preserves the sharp and prominent structures in the image. D-AMP has the tendency to over-smooth the image, whereas TVAL3 adds blotches to even the smooth parts of the image.

[18] J. Mairal, M. Elad, and G. Sapiro. Sparse representation for color image restoration. *IEEE Transactions on image processing*, 17(1):53–69, 2008.

[19] J. Mairal, G. Sapiro, and M. Elad. Learning multiscale sparse representations for image and video restoration. *Multiscale Modeling & Simulation*, 7(1):214–241, 2008.

[20] X.-J. Mao, C. Shen, and Y.-B. Yang. Image restoration using convolutional auto-encoders with symmetric skip conn-

[17] J. Mairal, M. Elad, and G. Sapiro. Computational Optimization and Applications, 56(3):507–530, 2013.

[19] J. Mairal, G. Sapiro, and M. Elad. Learning multiscale sparse representations for image and video restoration. *Multiscale Modeling & Simulation*, 7(1):214–241, 2008.
nections. arXiv preprint arXiv:1606.08921, 2016. 1, 2

[21] C. A. Metzler, A. Maleki, and R. G. Baraniuk. From denoising to compressed sensing. 2014. 3

[22] A. Mousavi, A. B. Patel, and R. G. Baraniuk. A deep learning approach to structured signal recovery. In 2015 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton), pages 1336–1343. IEEE, 2015. 1

[23] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli. Image denoising using scale mixtures of gaussians in the wavelet domain. IEEE Transactions on Image processing, 12(11):1338–1351, 2003. 2

[24] D. Rosenbaum and Y. Weiss. The return of the gating network: Combining generative models and discriminative training in natural image priors. In Advances in Neural Information Processing Systems, pages 2683–2691, 2015. 1

[25] S. Roth and M. J. Black. Fields of experts: A framework for learning image priors. In 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05), volume 2, pages 860–867. IEEE, 2005. 2

[26] L. Theis and M. Bethge. Generative image modeling using spatial lstms. In Advances in Neural Information Processing Systems, pages 1927–1935, 2015. 1, 2, 3, 4, 5

[27] L. Theis, R. Hosseini, and M. Bethge. Mixtures of conditional gaussian scale mixtures applied to multiscale image representations. PloS one, 7(7):e39857, 2012. 2, 3

[28] A. van den Oord, N. Kalchbrenner, and K. Kavukcuoglu. Pixel recurrent neural networks. arXiv preprint arXiv:1601.06759, 2016. 2

[29] L. Xu, J. S. Ren, C. Liu, and J. Jia. Deep convolutional neural network for image deconvolution. In Advances in Neural Information Processing Systems, pages 1790–1798, 2014. 1, 2

[30] D. Zoran and Y. Weiss. From learning models of natural image patches to whole image restoration. In 2011 International Conference on Computer Vision, pages 479–486. IEEE, 2011. 2