A Nonparametric Bayesian Model for Image Restoration

Jinke Li
Department of Applied Mathematics and Statistics
Johns Hopkins University
jli300@jh.edu

Abstract

Image restoration aims to recover the underlying clean image from corrupted observations. This paper extends the popular expected patch log-likelihood (EPLL) method, and proposes a novel hierarchical Bayesian model which can captures the similarity of images. As the number of components for Gaussian mixture is unknown practically, this work aims to allow the model's complexity to grow when additional images are observed. Furthermore, we derive a variational inference algorithm that is easily applicable to massive data. To test the proposed method, this study conducts extensive experiments on standard benchmarks. The performance is comparable to state-of-the-art, demonstrating the effectiveness of this model.

1 Introduction

Image restoration is an important problem in the are of image processing and computer vision. It is not only useful to simply provide us a high quality image, but also an essential pre-processing step for many vision tasks including object recognition, image segmentation. Due to the complex structures of images captured by modern cameras, patch-based methods form a very popular and successful class of image restoration [10] [17]. These methods process an image patch-by-patch where patch is a small sub-image ($e.g.$, of 4×4 pixels). Among these various patch-based methods, the *block matching and 3D filtering* (BM3D, [1]) employs a non-local filtering of images by collecting similar patches in 3D arrays. This method becomes a reliable benchmark in most of experiments over the past 10 years [16].

Unlike the filtering methods, learning good image priors is also of utmost importance for the study of image restoration [17]. We can apply the image priors learned from external images (patches) to be regularization for the prediction. The state-of-the-art *expected patch log-likelihood* (EPLL, [17]) models the image patches in Gaussian mixtures, and maximizes the log-likelihood of patches using priors learned from uncorrupted images. Moreover, this approach concludes that patch priors with high likelihoods yield better performance on the whold image.

To extend the EPLL framework, this paper adapt the *hierarchical Dirichlet process* (HDP, [13]) to design a nonparametric generative model for non-overlapping image patches. While the Gaussian Mixture model requires a fixed number of components, this model can be used to model the mixed-membership data with a potential infinite number of components [14]. Unlike the *Dirichlet Process Mixtures*, hierarchical model is more applicable for capturing the self-similarities that are common in nature images [6]. While Gibbs sampling remains an important method of inference in such models, variational inference has certain advantages such as easy assessment of convergence, easy optimization [12]. With the scalable variational algorithm, we further conduct extensive experiments to test our method on real and synthesized noisy images with different noise levels. Compared with other state-of-the-art models, our method achieves superior performance consistently.

32nd Conference on Neural Information Processing Systems (NeurIPS 2018), Montréal, Canada.

2 Learning Image Priors

We begin by briefly reviewing the EPLL framework as described in Zoran [17]. Given an image x, the *expected patch log likelihood* with some prior p is defined as

$$EPLL_p(\mathbf{x}) = \sum_i \log p\left(\mathbf{P}_i \mathbf{x}\right), \tag{1}$$

where $\mathbf{P}_i$ is a binary matrix that extracts the i^{th} patch in image x and form it as a vector. For example, if our patch size is (8×8), $\mathbf{P}_i x \in \mathbb{R}^{64}$. Now, assume we are given a corrupted image y, the cost function for reconstructing the clean image x is given by:

$$\min_x \frac{\lambda}{2}\|x - y\|^2 - EPLL_p(x), \tag{2}$$

the first item is a general corruption model , it aims to find clean image x that is close to y. The second item is a regularization, aiming to find image x with higher log likelihood. Zoran and Weistrass learns a finite Gaussian mixture model, the log likelihood of a given patch is

$$\log p(\mathbf{P}_i \mathbf{x}) = \log \left(\sum_{k=1}^{K} \pi_k N\left(\mathbf{P}_i \mathbf{x} \mid \mu_k, \Sigma_k^{-1}\right) \right) \tag{3}$$

Direct optimization of the cost function in Equation (2) may be very hard, an alternative optimization method called *Half Quadratic Splitting* (Geman [5]) is used here. We introduce an auxiliary variable $\bar{v}_i$ for each patch i, the cost function can be reformulated as follows:

$$\min_{x,\bar{v}} \frac{\lambda}{2}\|x - y\|^2 + \sum_i \frac{\delta}{2}\left\|P_i x - \bar{v}_i\right\|^2 - \log p\left(\bar{v}_i\right) \tag{4}$$

With this objective, the closed-form coordinate descent updates can be presented in Algorithm 1.

Algorithm 1: Coordinate Descent Algorithm

Result: Clean image x
Initialization $y, \{\mu_k\}_{k=1}^K, \{\Sigma_k\}_{k=1}^K, x^0$;
while *convergence is reached* **do**
  instructions;
  for *patch i in image y* **do**
    cluster index $I_i = \arg\max_k \mu_k \, \mathrm{N}\left(P_i x \mid 0, \Sigma_k^{-1} + \delta I\right)$;
    solving $\bar{v}_i = \left(I + \delta^{-1}\Sigma_{z_i}\right)^{-1} P_i x$
  end
  solving $x = \left(\lambda I + \delta \sum_i P_i^T P_i\right)^{-1} \left(\lambda y + \delta \sum_i P_i^T \bar{v}_i\right)$;
end

In EPLL framework, patch models which give high likelihood values for patches sampled from natural images perform better in patch and image restoration tasks. However, the GMM model in clustering the complex image structure is extremely naive [16]. Our study addresses this question by extending the GMM to a nonparametric Bayesian generative model.

3 Model

In this section we develop the HDP model on the image patches prior, which creates new components of patches automatically, and captures the self-similarity within and across the images.

3.1 Hierarchical Dirichlet Process Mixtures

The *Hierarchical Dirichlet Process* [13] addresses the problem in which observations are organized into groups, and where the observations are assumed to be exchangeable within the group. Applied

(a) (b) (c) (d)

Figure 1: Images of animals from *Berkeley Segmentation Dataset*. Patches in different images can share the same atoms, but with different weights in each image.

to document collections, the HDP provides a topic model where documents are viewed as groups of observed words, mixture components (called topics) are distributions over terms, and each document exhibits the topics with different proportions. In our image restoration task, we model the images as "document groups", and patches as the observations within each group. The HDP model allows the patches in different image shares the same set of atoms, but with different set of weights. Example of these patch similarities is illustrated in Fig.1, the animals' fur may share the same atoms of parameters, but different images have different proportions of such patches.

In our HDP model for image restoration, we consider a model where each image in a dataset is modelled as a mixture over topics, and each topic is a distribution that's drawn from a Dirichlet process. Let there be $\mathbf{M}$ images, and $\mathbf{N}$ patches in each image. We use a stick breaking construction to represent the model, the corpus-level frequency vector $\pi = [\pi_0, \pi_1, ...\pi_k, ...]$ is given by:

$$\beta_k \sim Beta(1, \gamma), \pi_k = \beta_k \prod_{\ell=1}^{k-1} (1 - \beta_\ell), \tag{5}$$

For each image m, the HDP model assigns a frequency vector $\pi_m = [\pi_{m0}, \pi_{m1}, ..., \pi_{mk}, ...]$, it is defined as:

$$\pi_m \mid \alpha, \pi \sim \mathbf{DP}(\alpha\pi) \tag{6}$$

Let z_{mn} be an indicator variable such that $\theta_k = \theta_{z_{mn}}$, where $\theta_k = (\Sigma_k, \mu_k) \sim H$. According to Teh [13], we have:

$$z_{mn} \mid \pi_m \sim \mathbf{Cat}(\pi_m) \tag{7}$$

$$\Sigma_k, \mu_k | H, z_{mn} = \Sigma_{z_{mn}}, \mu_{z_{mn}} \tag{8}$$

Once we have the indicator vector for each image m, the patch v_{mn} is generated by

$$v_{mn} \mid z_{mn}, (\Sigma_k)_{k=1}^{\infty}, (\mu_k)_{k=1}^{\infty} \sim N\left(\mu_{z_{mn}}, \Sigma_{z_{mn}}^{-1}\right) \tag{9}$$

Thus Fig.2 gives an equivalent representation of the hierarchical Dirichlet process mixture, with conditional distributions summarized here:

$$\pi|\gamma \sim GEM(\gamma), \quad \pi_m|\alpha, \pi \sim \mathbf{DP}(\alpha\pi)$$

$$z_{mn}|\pi_m \sim \mathbf{Cat}(\pi_m)$$

$$\Sigma_k, \mu_k|H, z_{mn} = \Sigma_{z_{mn}}, \mu_{z_{mn}}$$

$$v_{mn} \mid z_{mn}, (\Sigma_k)_{k=1}^{\infty}, (\mu_k)_{k=1}^{\infty} \sim N\left(\mu_{z_{mn}}, \Sigma_{z_{mn}}^{-1}\right)$$

3.2 Remarks on HDP

Learning priors from images is a very effective method, and the very popular deep learning method can also be regarded as a subcategory of this approach. Although deep learning has achieved good results in some image processing experiments, the Bayesian models are more reasonable and safer [2]. Marinescu [8] combines Bayesian method with generative network and achieved amazing results in image reconstruction tasks. Since HDP can handle very complex data, it is very suitable for the complex structure of high-dimensional images. Zoran [17] shows that that learning a better patch prior is equivalent with learning a better prior for the whole image. HDP can better captures the similarities between patches than Gaussian Mixture model, so it can help achieve better performance on the entire image.

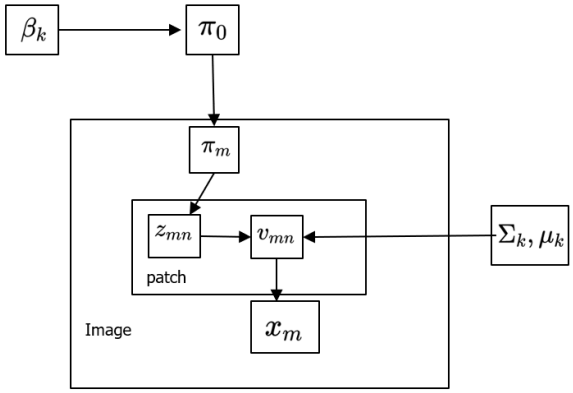

Figure 2: A alternative representation of a hierarchical Dirichlet process mixture model in terms of the stick-breaking construction.

3.3 Variational Inference

Let Θ be all the parameters in the above HDP model, suppose we are given the traininig set $\mathbf{D}$ containing image $x_1, x_2, ...x_M$. The posterior distribution $p(\Theta|\mathbf{x})$ is very complex and computationally intractable. Thus, we use the variational inference algorithm to obtain a tight approximation of the posterior. The variational inference methods seeks a distribution $q(\Theta)$ to minimize the KL divergence from $p(\Theta|\mathbf{x})$.

We use a fully factorized variational distribution and mean-field variational inference like Wang [14]. Here we use the Gaussian distribution for the observation model, which is different from Wang. The latent variables with the HDP model are the topic-level proportions $\pi = \{\pi_k\}_{k=1}^{\infty}$, the image-level proportions $\pi_m = \{\pi_{mt}\}_{t=1}^{\infty}$, and the vector of indicators $c_m = \{c_{mt}\}_{t=1}^{\infty}$ for each G_m. Also, the atoms $\{\Sigma_k\}$,$\{\mu_k\}$; and the index of each patch $\{z_{mn}\}$. Note that we use Sethuraman's Stick-breaking construction [4] rather than Teh's construction [12]. Our variational distribution is given by:

$$q(\pi, \pi', c, z, \phi) = q(\pi)\, q(\pi')\, q(c) q(z) q(\phi) \tag{10}$$

where $\pi'_j = (\pi_{j1}, \pi_{j2}, ...)$, $\phi = (\Sigma, \mu)$. Then we have

$$\begin{aligned} q(c) &= \textstyle\prod_j \prod_t q(c_{jt} \mid \varphi_{jt}) \\ q(z) &= \textstyle\prod_j \prod_n q(z_{jn} \mid \zeta_{jn}) \\ q(\Sigma) &= \textstyle\prod_k q(\Sigma_k \mid a_k, \mathbf{B}_k) \\ q(\mu) &= \textstyle\prod_k q(\mu_k \mid v_{\mu_k}, \mathbf{I}) \end{aligned} \tag{11}$$

where $q(c) = \mathbf{Mult}(\varphi_{jt})$, $q(z) = \mathbf{Mult}(\zeta_{jn})$, $q(\Sigma) = \mathbf{Wish}(a_k, \mathbf{B}_k)$,$q(\mu) = \mathbf{N}(v_{\mu_k}, \mathbf{I})$. For the proportion π, π', we have:

$$\begin{aligned} q(\pi) &= \textstyle\prod_{k=1}^{K-1} q(\pi_k \mid u_k, v_k) \\ q(\pi') &= \textstyle\prod_j \prod_{t=1}^{T-1} q(\pi'_{jt} \mid a_{jt}, b_{jt}) \end{aligned} \tag{12}$$

where $q(\pi) = \mathbf{Beta}(u_k, v_k), q(\pi') = \mathbf{Beta}(a_k, b_k)$ By the ELBO bound given in Wang [14], we derive the follwing coordinate ascent updates.

"Document" Level Updates At this level, we update the parameters in every image.

$$a_{jt} = 1 + \sum_n \zeta_{jnt} \tag{13}$$

$$b_{jt} = \alpha + \sum_n \sum_{s=t+1}^{T} \zeta_{jns} \tag{14}$$

$$\varphi_{jtk} \propto \exp\left(\sum_n \zeta_{jnt} \mathbb{E}_q \left[\log p(v_{jn} \mid \phi_k) \right] + \mathbb{E}_q \left[\log \pi_k \right] \right) \tag{15}$$

$$\zeta_{jnt} \propto \exp\left(\sum_{k=1}^{K} \varphi_{jtk} \mathbb{E}_q\left[\log p\left(v_{jn} \mid \phi_k\right)\right] + \mathbb{E}_q\left[\log \pi_{jt}\right]\right) \tag{16}$$

"Corpus" Level Updates

$$u_k = 1 + \sum_j \sum_{t=1}^{T} \varphi_{jtk} \tag{17}$$

$$v_k = \gamma + \sum_j \sum_{t=1}^{T} \sum_{l=k+1}^{K} \varphi_{jtl} \tag{18}$$

$$a_k = 2 + D + \sum_i \nu_{z_{i,k}} \tag{19}$$

$$\mathbf{B}_k = \left(\left(\sum_{m,n} \nu_{z_{m,n}} + 1\right)\mathbf{I} + \sum_{m,n} \nu_{z_{m,n}} \left(v_{mn} - \nu_{\mu_k}\right)\left(v_{mn} - \nu_{\mu_k}\right)^T\right)^{-1} \tag{20}$$

where $\Sigma_k \sim \mathbf{Wish}(D, I)$, v_{mn} denote the patches.

4 Experiments

Dataset The HDP model and other state-of-art methods are evaluated on two image datasets:The Berkeley Segmentation Dataset (BSDS, [9]), Tampere Image Database (TID 2008, [11]). Our patch size is 8×8, which is suitable for the most of images with 512 × 512 pixels in the dataset. To evaluate the various methods, we use 200 images in BSDS as training images, while the remaining 100 images as test images. Since the number of images in TID2008 is quite large, we randomly select 200 images, 100 images to be training set, test set, respectively.

Compared Methods K-SVD [3], LSSC [7], EPLL [17], and BM3D [1] are chosen to as the compared methods. For the filtering method BM3D, we use the same test images to evaluate its performance. To better illustrative our model's effectiveness, we train the HDP model and use the number of components as the hyperparameter of K-SVD.

Evaluation Metrics The image restoration performance of these models is evaluated by the peak signal-to-noise ratio (PSNR), a logarithmic transform of the mean squared error (MSE) between images with normalized intensities,

$$\mathrm{PSNR} = -20\log_{10}\mathrm{MSE}$$

We also evaluate the structural similarity index (SSIM, [15]), which can be viewed as a quality measure of one of the images being compared, provided the other image is regarded as of perfect quality.

Synthesis Noisy Images Although the best method to evaluate the denoising performance is using real noisy images, the "real" images with ground truth are very hard to collect. On the other hand, Gaussian noise can be treated as the combination of other noise, so the synthetic white noise is generally used in denoising tasks.

4.1 Why "Hierarchical"

HDP models the patches by different weights, but allows the patches share the same atoms. We conduct the experiment on the two figures using both Dirichlet Process and Hierarchical Dirichlet Process. For DP model, we just put all the patches together and do clustering. In the training dataset, patches from different images can have some similarities, but DP give the same proportions in different images, which leads to some inaccuracy. As shown in Fig.3, HDP model can reduces the artifacts in the images, which captures the similarities of patches better than non-hierarchical model.

(a) DP (b) HDP (c) DP (d) HDP

Figure 3: The Hierarchical Dirichlet Process can better capture the patches similarities, which reduces the artifacts in the image. In Dirichlet Process mixture model, the algorithm tends to give lower log-likelihood than HDP.

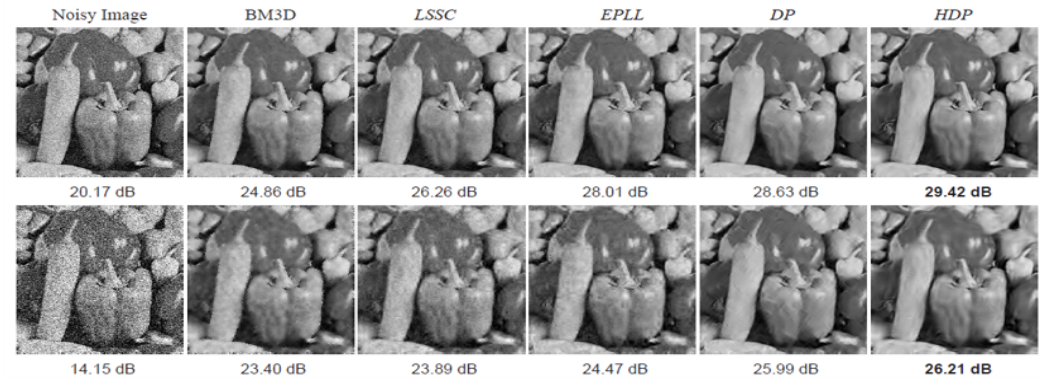

Figure 4: Denoising results of six methods; the first row's noise level is $\sigma = 25$, the second row's noise level is $\sigma = 50$

4.2 Overall Performance

We first evaluate our method on noisy images contaminated by homogeneous white Gaussian noise. We compare our HDP method to other four classical methods on images contaminated by noise with intensity σ from 10 to 100. (The true standard deviation of the Gaussian noise is $\sigma/255.0$) Table 1 shows the average quantitative performance of the methods on the two datasets. It can be seen that our HDP model has a better performance when the noise level is high, this illustrates the efficiency of our method. Fig. 4 shows the denoising results on one noisy image with $\sigma = 25, 50$. It can be observed that our method can well eliminate the noise and better preserve the details. The overall performance shows that our method can well cope with images contaminated by Gaussian white noise with different intensity.

4.3 Complex Noise

Although Gaussian white noise is widely used in image processing, there are many much more complex noise model. In this part, we compare different models' performance on images with: (1) Laplacian noise with $\sigma = 15, 30, 45$; (2) Combined noise: We divide the image into two parts, one using Laplacian noise with $\sigma = 45$, another using Gaussian noise with $\sigma = 50$. Fig.5 shows the result of our experiments, the proposed model perform better when the noises are complex, which illustrates our model's advantage.

5 Conclusions

We propose a novel image restoration framework which can efficiently recover noisy images even the noise model is complex and unavailable. We assume the clean image patches are complex and we model them using the nonparametric technique. We show that the HDP mixture models of can grow in complexity as additional images are observed and capture the similarity of natural images. To test

(a) Combined Noise PSNR:21.9742 PSNR:28.3826

(b) Laplacian Noise PSNR:28.5219 PSNR:29.4927

Figure 5: Comparison with complex noise; left:noise image; middle: EPLL; right: HDP

	σ	20	40	60	80	100
Dataset	Method	PSNR				
BSDS	K-SVD	30.46	27.89	24.23	23.80	22.56
	LSSC	30.52	28.19	23.41	23.94	23.01
	BM3D	31.49	27.42	24.17	23.64	23.24
	EPLL	**31.75**	29.94	24.88	23.84	23.11
	HDP	31.42	**29.97**	**25.03**	**23.99**	**23.32**
TID2008	K-SVD	36.25	30.51	27.52	23.33	20.15
	LSSC	**36.92**	30.29	27.94	23.14	21.24
	BM3D	36.77	30.62	27.68	23.75	21.85
	EPLL	36.82	30.82	28.00	23.77	21.70
	HDP	36.79	**30.97**	**28.03**	23.92	**21.88**
		SSIM				
BSDS	K-SVD	0.942	0.930	0.817	0.775	0.744
	LSSC	0.921	0.935	0.823	0.774	0.748
	BM3D	**0.988**	**0.938**	0.818	0.800	0.765
	EPLL	0.952	0.934	0.825	0.794	0.762
	HDP	0.972	0.931	**0.829**	**0.803**	**0.769**
TID2008	K-SVD	0.938	0.925	0.822	0.766	0.754
	LSSC	0.953	0.926	0.827	0.774	0.725
	BM3D	0.954	0.931	0.835	0.792	0.741
	EPLL	0.979	**0.937**	0.834	0.794	0.765
	HDP	**0.981**	0.927	**0.838**	**0.805**	**0.775**

Table 1: Performance of various approaches on noisy images with white Gaussian noise.

our method, we conduct extensive experiments on synthesis and real images. Our HDP model are competitive with the state-of-the-art, and it is easy to do posterior sampling.

References

[1] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Image restoration by sparse 3d transform-domain collaborative filtering. In *Image Processing: Algorithms and Systems VI*, volume 6812, page 681207. International Society for Optics and Photonics, 2008.

[2] Wei Deng, Guang Lin, and Faming Liang. A contour stochastic gradient langevin dynamics algorithm for simulations of multi-modal distributions. *arXiv preprint arXiv:2010.09800*, 2020.

[3] 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.

[4] Emily B Fox, Erik B Sudderth, Michael I Jordan, and Alan S Willsky. An hdp-hmm for systems with state persistence. In *Proceedings of the 25th international conference on Machine learning*, pages 312–319, 2008.

[5] Donald Geman and Chengda Yang. Nonlinear image recovery with half-quadratic regularization. *IEEE transactions on Image Processing*, 4(7):932–946, 1995.

[6] Hervé Jégou, Matthijs Douze, and Cordelia Schmid. On the burstiness of visual elements. In *2009 IEEE conference on computer vision and pattern recognition*, pages 1169–1176. IEEE, 2009.

[7] Julien Mairal, Francis Bach, Jean Ponce, Guillermo Sapiro, and Andrew Zisserman. Non-local sparse models for image restoration. In *2009 IEEE 12th international conference on computer vision*, pages 2272–2279. IEEE, 2009.

[8] Razvan V Marinescu, Daniel Moyer, and Polina Golland. Bayesian image reconstruction using deep generative models. *arXiv preprint arXiv:2012.04567*, 2020.

[9] David Martin, Charless Fowlkes, Doron Tal, and Jitendra Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, volume 2, pages 416–423. IEEE, 2001.

[10] Shibin Parameswaran, Charles-Alban Deledalle, Loïc Denis, and Truong Q Nguyen. Accelerating gmm-based patch priors for image restoration: Three ingredients for a 100 times speed-up. *IEEE Transactions on Image Processing*, 28(2):687–698, 2018.

[11] Nikolay Ponomarenko, Vladimir Lukin, Alexander Zelensky, Karen Egiazarian, Marco Carli, and Federica Battisti. Tid2008-a database for evaluation of full-reference visual quality assessment metrics. *Advances of Modern Radioelectronics*, 10(4):30–45, 2009.

[12] Yee W Teh, Kenichi Kurihara, and Max Welling. Collapsed variational inference for hdp. In *Advances in neural information processing systems*, pages 1481–1488, 2008.

[13] Yee Whye Teh, Michael I Jordan, Matthew J Beal, and David M Blei. Hierarchical dirichlet processes. *Journal of the american statistical association*, 101(476):1566–1581, 2006.

[14] Chong Wang, John Paisley, and David Blei. Online variational inference for the hierarchical dirichlet process. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, pages 752–760. JMLR Workshop and Conference Proceedings, 2011.

[15] Zhou Wang, Eero P Simoncelli, and Alan C Bovik. Multiscale structural similarity for image quality assessment. In *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*, volume 2, pages 1398–1402. Ieee, 2003.

[16] Fengyuan Zhu, Guangyong Chen, and Pheng-Ann Heng. From noise modeling to blind image denoising. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 420–429, 2016.

[17] Daniel Zoran and Yair Weiss. From learning models of natural image patches to whole image restoration. In *2011 International Conference on Computer Vision*, pages 479–486. IEEE, 2011.