Towards a unified view of unsupervised non-local methods for image denoising: the NL-Ridge approach

Sébastien Herbreteau, Charles Kervrann

▶ **To cite this version:**

Sébastien Herbreteau, Charles Kervrann. Towards a unified view of unsupervised non-local methods for image denoising: the NL-Ridge approach. ICIP 2022 - 29th IEEE International Conference on Image Processing, Oct 2022, Bordeaux, France. pp.3376-3380, 10.1109/ICIP46576.2022.9897992 . hal-03926888

HAL Id: hal-03926888
https://hal.science/hal-03926888v1

Submitted on 6 Jan 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

TOWARDS A UNIFIED VIEW OF UNSUPERVISED NON-LOCAL METHODS FOR IMAGE DENOISING: THE NL-RIDGE APPROACH

Sébastien Herbreteau, Charles Kervrann
Inria Rennes - Bretagne Atlantique and UMR144-CNRS Institut Curie PSL, Paris, France
{sebastien.herbreteau, charles.kervrann}@inria.fr

ABSTRACT

We propose a unified view of unsupervised non-local methods for image denoising that linearly combine noisy image patches. The best methods, established in different modeling and estimation frameworks, are two-step algorithms. Leveraging Stein's unbiased risk estimate (SURE) for the first step and the "internal adaptation", a concept borrowed from deep learning theory, for the second one, we show that our NL-Ridge approach enables to reconcile several patch aggregation methods for image denoising. In the second step, our closed-form aggregation weights are computed through multivariate Ridge regressions. Experiments on artificially noisy images demonstrate that NL-Ridge may outperform well established state-of-the-art unsupervised denoisers such as BM3D and NL-Bayes, as well as recent unsupervised deep learning methods, while being simpler conceptually.

***K*eywords** Patch-based image denoising · non-local method · Stein's unbiased risk estimation · statistical aggregation · ridge regression

1 Introduction

From anisotropic diffusion to recently deep learning-based methods, image denoising is probably one of the most important topic in image processing. Among them, the N(on)-L(ocal) means algorithm [1] that exploits the self-similarity assumption and information redundancy may be considered as a milestone in image denoising. It is assumed that, in a natural image, a patch rarely appears alone: almost perfect copies can be found in its surroundings. The NL-means algorithm [1] then amounts to computing, for each pixel, an average of its neighboring noisy pixels, weighted by the degree of similarity of patches they belong to. This inspired a lot of methods afterward that manage several groups of similar noisy patches [2], [3], [4], [5]. In the final step, patches are re-positioned to their initial locations and a final averaging is performed at a given position as each pixel has been denoised multiple times. To exploit patches more collaboratively, frequency-based methods [2] or low-rank assumptions [4], [5] were proposed and led to state-of-the-art performances for more than a decade.

In this paper, our contribution is twofold. First, we propose a unified view, well grounded in Stein's unbiased risk estimation theory, to reinterpret and reconcile previous state-of-the-art non-local methods. In our estimation framework, no prior model for the distribution on patches is required. Second, the resulting two-step NL-Ridge algorithm, based on closed-form expressions, may outperform BM3D [2] and NL-Bayes [3], as well as several unsupervised deep-learning methods [6] [7] [8] on artificially noisy images with additive white Gaussian noise.

2 NL-Ridge for image denoising

2.1 A two-step approach applied to patches

Leveraging the self-similarity assumption, our objective is to design a local denoiser f_Θ, where Θ is a set of parameters, that operates on groups of similar noisy patches extracted from a noisy image and combines them collaboratively. Formally, let $X \in \mathbb{R}^{n\times m}$ be a matrix gathering m clean similar patches of size $\sqrt{n}\times\sqrt{n}$ and $Y \in \mathbb{R}^{n\times m}$ its noisy version. The matrices X and Y are generally referred to similarity matrices in the literature. One wants to estimate the

matrix X from $Y = X + W$, where $W \in \mathbb{R}^{n\times m}$ is a matrix whose elements $W_{i,j} \sim \mathcal{N}(0, \sigma^2)$ (Gaussian noise) are assumed to be independent along each row and each column, respectively. Our local denoiser is of the form:

$$f_\Theta : Y \mapsto Y\Theta \tag{1}$$

where $\Theta \in \mathbb{R}^{m\times m}$. The function f_Θ aims here at denoising the m noisy similar patches contained in Y all at once. To this end, each noisy patch is processed by a linear combination of its most similar patches whose weights are gathered in the matrix Θ. Aggregating similar patches via a linear combination has already been exploited in the past [1] [9] [10]. However, the originality of our method lies in the way of computing the weights Θ, which significantly boosts performance.

The optimal local denoiser f_Θ is found by minimizing the quadratic risk defined as:

$$R_\Theta(X) = \mathbb{E}\|f_\Theta(Y) - X\|_F^2 \tag{2}$$

where $\|\cdot\|_F^2$ denotes the Frobenius norm. In other words, we look for the Minimum-Mean-Squared-Error (MMSE) estimator among the family of functions $(f_\Theta)_{\Theta\in\mathbb{R}^{m\times m}}$. The optimal estimator f_{Θ^*} minimizes the risk, *i.e.*

$$\Theta^* = \arg\min_\Theta \; R_\Theta(X). \tag{3}$$

Unfortunately, Θ^* requires the knowledge of X which is unknown. The good news is that the risk $R_\Theta(X)$ can be approximated through the following two-step algorithm:

- In the first step, an approximation of Θ^* is computed for each group of similar patches, through the use of an unbiased estimate of $R_\Theta(X)$, Stein's unbiased risk estimate (SURE) [11]. After reprojection [12] of all denoised patches, a first denoised image $\hat{I}_1$ is obtained.
- In the second step, $\hat{I}_1$ is improved with a second estimation of Θ^* which is found thanks to the technique of "internal adaptation" described in [13].

2.2 Step 1: Stein's unbiased risk estimate

Preliminarily, let us consider the two following propositions.

Proposition 1. *Let $Y = X + W$ where $Y, X, W \in \mathbb{R}^{n\times m}$ and $W_{i,j} \sim \mathcal{N}(0, \sigma^2)$ are independent along each row. An unbiased estimate of the risk $R_\Theta(X) = \mathbb{E}\|f_\Theta(Y) - X\|_F^2$ is Stein's unbiased risk estimate (SURE):*

$$\mathrm{SURE}_\Theta(Y) = \|Y\Theta - Y\|_F^2 + 2n\sigma^2 \,\mathrm{tr}(\Theta) - nm\sigma^2$$

where tr *denotes the trace operator.*

Proposition 2. *Let $Y \in \mathbb{R}^{n\times m}$ and $\Theta \in \mathbb{R}^{m\times m}$.*

$$\hat{\Theta}_1 = \arg\min_\Theta \; \mathrm{SURE}_\Theta(Y) = I_m - n\sigma^2(Y^\top Y)^{-1}$$

where I_m is the $m \times m$ identity matrix and $\top$ denotes the transpose operator.

Our objective is to get an approximation of Θ^* from (3). Proposition 1 gives an unbiased estimate of the risk $R_\Theta(X)$ that does not depend on X, but only on the observations Y. A common idea that has been previously exploited in image denoising (*e.g.* see [14], [15], [16], [17], [18]) is to use this estimate as a surrogate for minimizing the risk $R_\Theta(X)$ in (3) which is inaccessible. Based on Proposition 2, we get a first estimation:

$$\hat{\Theta}_1 = I_m - n\sigma^2(Y^\top Y)^{-1}. \tag{4}$$

Note that $\hat{\Theta}_1$ is close to Θ^* as long as the variance of SURE is low. A rule of thumbs used in [16] states that the number of parameters must not be "too large" compared to the number of data in order for the variance of SURE to remain small. In our case, this suggests that $m < n$.

2.3 Step 2: Internal adaptation

Proposition 3. *The quadratic risk $R_\Theta(X)$ defined in (2) is:*

$$\mathbb{E}\|f_\Theta(X + W) - X\|_F^2 = \|X\Theta - X\|_F^2 + n\sigma^2\|\Theta\|_F^2$$

which is minimal for:

$$\Theta^* = (X^\top X + n\sigma^2 I_m)^{-1} X^\top X$$

(solution of the multivariate Ridge regression).

At the end of the first step, we get a first denoised image $\hat{I}_1$ that will serve as a pilot in the second step. Once again, we focus on the solution of (3) to denoise locally similar patches. As X and $\hat{X}_1$, the corresponding group of similar patches in $\hat{I}_1$, are supposed to be close, the "internal adaptation" procedure [13] consists in solving (3) by substituting $\hat{X}_1$ for X. From Proposition 3, we get the following closed-form solution:

$$\hat{\Theta}_2 = (\hat{X}_1^\top \hat{X}_1 + n\sigma^2 I_m)^{-1} \hat{X}_1^\top \hat{X}_1 \tag{5}$$

which is nothing else than the solution of a multivariate Ridge regression. Interestingly, in practice, this second estimate of Θ^* produces a significant boost in terms of denoising performance compared to $\hat{\Theta}_1$. The second step can be iterated but we did not notice improvements in our experiments.

2.4 Weighted average reprojection

After the denoising of a group of similar patches, each denoised patch is repositioned at its right location in the image. As several pixels are denoised multiple times, a final step of aggregation, or reprojection [12], is necessary to produce a final denoised image $\hat{I}_1$ or $\hat{I}_2$. With inspiration from [12], each pixel belonging to column j of Y is assigned, after denoising, the weight $w_j = 1/(\|\Theta_{\cdot,j}\|_2^2)$. Those weights are at the end pixel-wise normalized such that the sum of all weights associated to a same pixel equals one.

3 A unified view of non-local denoisers

In NL-Ridge, the local denoiser f_Θ is arbitrarily of the form given by (1) involving the weighted aggregation of similar patches with closed-form weights given in (4) and (5). In this section, we show that NL-Ridge can serve to interpret two popular state-of-the-art non-local methods (NL-Bayes [3] and BM3D [2]), which were originally designed with two very different modeling and estimation frameworks. It amounts actually to considering two particular families (f_Θ) of local denoisers.

3.1 Analysis of NL-Bayes algorithm

The NL-Bayes [3] algorithm has been established in the Bayesian setting and the resulting maximum a posteriori estimator is computed with a two-step procedure as NL-Ridge. As starting point, let us consider the following local denoiser:

$$f_{\Theta,\beta} : Y \mapsto \Theta Y + \beta u^\top \tag{6}$$

where $\Theta \in \mathbb{R}^{n\times n}$, $\beta \in \mathbb{R}^n$, and u denotes a m-dimensional vector with all entries equal to one.

- **Step 1:** Stein's unbiased estimate of the quadratic risk $R_{\Theta,\beta}(X)$ reaches its minimum for:

$$\hat{\Theta}_1 = (C_Y - \sigma^2 I_n) C_Y^{-1} \quad \text{and} \quad \hat{\beta}_1 = (I_n - \hat{\Theta}_1)\mu_Y \tag{7}$$

where μ and C denote the empirical mean and covariance matrix of a group of patches. Interestingly, $f_{\hat{\Theta}_1,\hat{\beta}_1}(Y)$ is the expression given in [3] (first step), which is actually derived from the prior distribution of patches assumed to be Gaussian. Furthermore, our framework provides guidance on the choice of the parameters n and m. Indeed, SURE is helpful provided that its variance remains small which is achieved if $n < m$ (the number of parameters must not be "too large" compared to the number of data). This result suggests that NL-Bayes is expected to be efficient if small patches are used, as confirmed in the experiments in [19].

- **Step 2:** Interpreting the second step in [3] as an "internal adaptation" step, the updated parameters become:

$$\hat{\Theta}_2 = C_{\hat{X}_1}(C_{\hat{X}_1} + \sigma^2 I_n)^{-1} \quad \text{and} \quad \hat{\beta}_2 = (I_n - \hat{\Theta}_2)\mu_{\hat{X}_1} \tag{8}$$

and $f_{\hat{\Theta}_2,\hat{\beta}_2}(Y)$ corresponds to the original second-step expression in [3].

3.2 Analysis of BM3D algorithm

BM3D [2] is probably the most popular non-local method for image denoising. It assumes a locally sparse representation of images in a transform domain. A two step algorithm was described in [2] to achieve state-of-the-art results for several years. By using the generic NL-Ridge formulation, we consider the following family:

$$f_\Theta : Y \mapsto P^{-1}(\Theta \odot (PYQ))Q^{-1} \tag{9}$$

where $\Theta \in \mathbb{R}^{n\times m}$ and $\odot$ denotes the Hadamard product. $P \in \mathbb{R}^{n\times n}$ and $Q \in \mathbb{R}^{m\times m}$ are two orthogonal matrices that model a separable 3D-transform (typically a 2D and 1D *Discrete Cosine Transform*, respectively).

- **Step 1:** The minimization of SURE yields:

$$\hat{\Theta}_{1a} = U - \frac{\sigma^2}{(PYQ)^2} \tag{10}$$

where $U \in \mathbb{R}^{n\times m}$ whose elements equal to one and where division and square are element-wise operators. Unfortunately, $f_{\hat{\Theta}_{1a}}(Y)$ does not provide very satisfying denoising results. This result is actually expected as the number of parameters equals the size of data (nm), making SURE weakly efficient. To overcome this difficulty, we can force the elements of Θ to be either 0 or 1. Minimizing SURE with this constraint yields to:

$$\hat{\Theta}_{1b} = \mathbb{1}(|PYQ| \geq \sqrt{2}\sigma). \tag{11}$$

where $\mathbb{1}(x \geq \tau) = 1$ if $x \geq \tau$ and 0 otherwise. Our denoiser f_Θ is then a hard thresholding estimator as in BM3D: the coefficients of the transform domain (*i.e.* the elements of the matrix PYQ) below $\sqrt{2}\sigma$, in absolute value, are canceled before applying the inverse 3D-transform. This result suggests that the threshold should be linearly dependent on σ but also independent on orthogonal transforms P and Q. In [2], a threshold of 2.7σ was carefully chosen in Step 1, which is approximately twice the SURE-prescribed threshold.

- **Step 2:** the "internal adaptation" yields the same expression as the Wiener filtering step in BM3D:

$$\hat{\Theta}_2 = \frac{(P\hat{X}Q)^2}{\sigma^2 + (P\hat{X}Q)^2} \tag{12}$$

In conclusion, we have shown that BM3D [2] and NL-Bayes [3] can be interpreted within NL-Ridge framework, enabling to set the size of the patches and to potentially relax the need to specify the prior distribution of patches.

4 Experimental results

In this section, we compare the performance of our NL-Ridge method with state-of-the-art methods, including related learning-based methods [20] [21] [13] [6] [7] [8] applied to standard gray images artificially corrupted with additive white Gaussian noise with zero mean and variance σ^2. Performances of NL-Ridge and other methods are assessed in terms of PSNR values. We want to emphasize that NL-Ridge is as fast as BM3D [2] and NL-Bayes [3] as they share the same paradigm. It has been implemented in Python with Pytorch, enabling it to run on GPU unlike its traditional counterparts. The code can be downloaded at: https://github.com/sherbret/NL-Ridge/.

4.1 Setting of algorithm parameters

For the sake of computational efficiency, the search for groups of similar patches across the image is restricted to a small local window of size $w \times w$ (in our experiments $w = 45$). More precisely, for a reference patch y_j, one looks for its $m-1$ more similar patches of size $\sqrt{n} \times \sqrt{n}$ in the local window in order to form the similarity matrix Y of size $n \times m$. We used the conventional L_2 distance to form groups of similar patches. Considering iteratively each overlapping patch of the image as reference patch is also computationally demanding. To speed up the algorithm, NL-Ridge is performed on a subsampled position grid. We used a moving step $\delta = 4$ from one reference patch to its neighbors both horizontally and vertically.

Table 1: Recommended patch size n and patch number m for the Step 1 and Step 2 versus noise standard deviation σ.

σ	n_1	n_2	m_1	m_2
$0 < \sigma \leq 15$	7×7	7×7	18	55
$15 < \sigma \leq 35$	9×9	9×9	18	90
$35 < \sigma \leq 50$	11×11	9×9	20	120

Finally, the choice of the parameters n and m depend on the noise level. Experimentally, bigger patches have to be considered for higher noise levels as well as a higher quantity of patches for the second step. An empirical analysis leads to choose the parameters reported in Table 1.

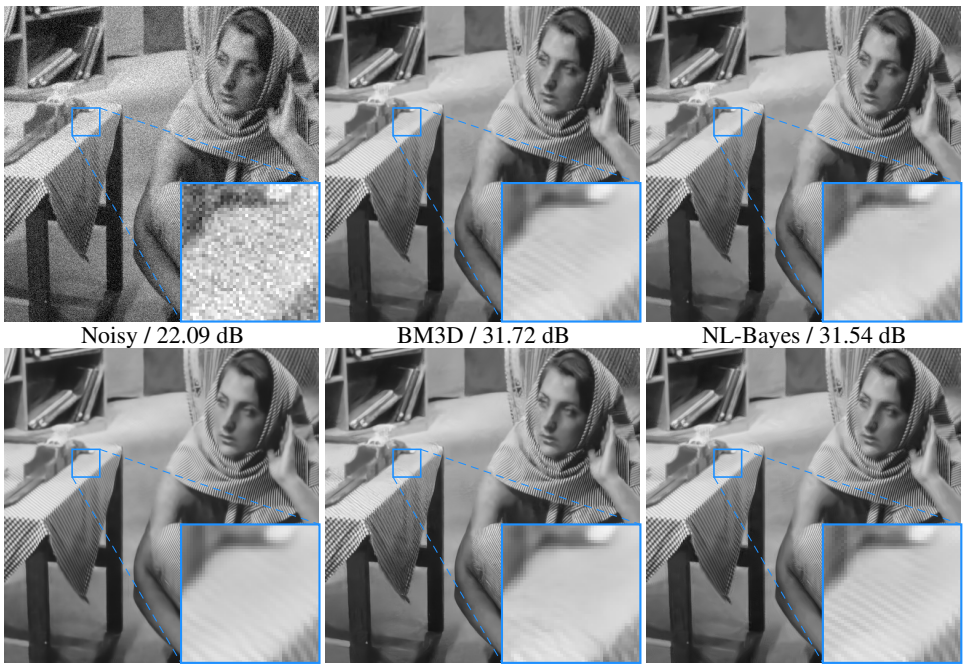

Figure 1: Denoising results (in PSNR) on *Barbara* corrupted with additive white Gaussian noise ($\sigma = 20$).

Table 2: The average PSNR (dB) results of different methods on various datasets corrupted with Gaussian noise ($\sigma = 15$ and 25). Best performance among each category is in bold.

	Methods	Set12	BSD68	Urban100
	Noisy	24.61 / 20.17	24.61 / 20.17	24.61 / 20.17
Unsupervised	BM3D	32.37 / 29.97	31.07 / 28.57	32.35 / 29.70
Trad.	NL-Bayes	32.25 / 29.88	31.16 / **28.70**	31.96 / 29.34
Trad.	**NL-Ridge**	**32.46** / 30.00	**31.20** / 28.67	**32.53** / **29.90**
DL	DIP	30.12 / 27.54	28.83 / 26.59	- / -
DL	Noise2Self	31.01 / 28.64	29.46 / 27.72	- / -
DL	Self2Self	32.07 / **30.02**	30.62 / 28.60	- / -
Super-vised	DnCNN	**32.86** / **30.44**	**31.73** / **29.23**	32.68 / 29.97
Super-vised	FFDnet	32.75 / 30.43	31.63 / 29.19	32.43 / 29.92
Super-vised	LIDIA	32.85 / 30.41	31.62 / 29.11	**32.80** / **30.12**

4.2 Results on test datasets

We tested the denoising performance of our method on three well-known datasets: Set12, BSD68 [22] and Urban100 [23]. A comparison with state-of-the-art algorithms is reported in Table 2. For a fair comparison, algorithms are divided into two categories: unsupervised methods (either traditional or deep learning-based) and supervised deep learning-based ones that require a training phase beforehand on an external dataset. We used the implementations provided by the authors for all algorithms. As for Noise2Self [7], only the single-image extension was considered.

NL-Ridge, exclusively based on weighted aggregation of noisy patches, performs surprisingly at least as well as its traditional counterparts [2] [3]. It is particularly efficient on Urban100 dataset which contains abundant structural patterns and textures, achieving comparable performances with FFDnet [21], a popular supervised network composed of hundreds of thousands of parameters.

Figure 1 illustrates the visual results of different methods. NL-Ridge is very competitive with respect to well-established methods such as BM3D [2]. The self-similarity assumption is particularly useful to recover subtle details such as the stripes on the *Barbara* image that are better reconstructed than DnCNN [20].

5 Conclusion

We presented a unified view to reconcile state-of-the-art unsupervised non-local denoisers through the minimization of a risk from a family of estimators, exploiting Stein's unbiased risk estimate on the one hand and the "internal adaptation" on the other. We derive NL-Ridge algorithm, which leverages local linear combinations of noisy similar patches. Our experimental results show that NL-Ridge compares favorably with its state-of-the-art counterparts, including recent unsupervised deep learning methods which are much more computationally demanding.

References

[1] A. Buades, B. Coll, and J.-M. Morel, "A review of image denoising algorithms, with a new one," *SIAM Journal on Multiscale Modeling and Simulation*, vol. 4, no. 2, pp. 490–530, 2005.

[2] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3d transform-domain collaborative filtering," *IEEE Transactions on Image Processing*, vol. 16, no. 8, pp. 2080–2095, 2007.

[3] A. Buades, M. Lebrun, and J.-M. Morel, "A non-local bayesian image denoising algorithm," *SIAM Journal on Imaging Science*, vol. 6, no. 3, pp. 1665–1688, 2013.

[4] S. Gu, L. Zhang, W. Zuo, and X. Feng, "Weighted nuclear norm minimization with application to image denoising," *IEEE Conference on Computer Vision and Pattern Recognition*, p. 2862–2869, 2014.

[5] H. Hu, J. Froment, and Q. Liu, "A note on patch-based low-rank minimization for fast image denoising," *Journal of Visual Communication and Image Representation*, vol. 50, pp. 100–110, 2018.

[6] Y. Quan, M. Chen, T. Pang, and Ji H., "Self2self with dropout: Learning self-supervised denoising from single image," *CVPR*, 2020.

[7] J. Batson and L. Royer, "Noise2self: Blind denoising by self-supervision," *Proceedings of the 36th International Conference on Machine Learning*, vol. 97, pp. 524–533, 2019.

[8] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Deep image prior," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, p. 9446–9454, 2018.

[9] C. Kervrann, "Pewa: Patch-based exponentially weighted aggregation for image denoising," *Neural Information Processing Systems*, pp. 2150–2158, Montreal, Canada, 2014.

[10] Q. Jin, I. Grama, C. Kervrann, and Q. Liu, "Non-local means and optimal weights for noise removal," *SIAM Journal on Imaging Sciences*, vol. 10, no. 4, pp. 1878–1920, 2017.

[11] C. Stein, "Estimation of the mean of a multivariate normal distribution," *Annals of Statistics*, vol. 9, no. 6, pp. 1135–1151, 1981.

[12] J. Salmon and Y. Strozecki, "From patches to pixels in non-local methods: Weighted-average reprojection," *2010 IEEE International Conference on Image Processing*, pp. 1929–1932, 2010.

[13] G. Vaksman, M. Elad, and P. Milanfar, "Lidia: Lightweight learned image denoising with instance adaptation," *Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2020.

[14] A. Benazza-Benyahia and J.-C. Pesquet, "An extended sure approach for multicomponent image denoising," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 2, pp. ii–945, 2004.

[15] F. Luisier, T. Blu, and M. Unser, "A new sure approach to image denoising: Interscale orthonormal wavelet thresholding," *IEEE Transactions on Image Processing*, vol. 16, no. 3, pp. 593–606, 2007.

[16] T. Blu and F. Luisier, "The sure-let approach to image denoising," *IEEE Transactions on Image Processing*, vol. 16, no. 11, pp. 2778–2786, 2007.

[17] D. Van De Ville and M. Kocher, "Sure based non-local means," *IEEE Signal Processing Letters*, vol. 16, no. 11, pp. 973–976, 2009.

[18] Y.-Q. Wang and J.-M. Morel, "Sure guided gaussian mixture image denoising," *SIAM Journal on Imaging Sciences*, vol. 6, no. 2, pp. 999–1034, 2013.

[19] M. Lebrun, A. Buades, and J.-M. Morel, "Implementation of the "non-local bayes" (nl-bayes) image denoising algorithm," *Image Processing On Line*, vol. 3, pp. 1–42, 2013.

[20] 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.

[21] 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.

[22] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," *International Conference on Computer Vision*, vol. 2, pp. 416–423, 2001.

[23] J.-B. Huang, A. Singh, and N. Ahuja, "Single image super-resolution from transformed self-exemplars," *Conference on Computer Vision and Pattern Recognition (CVPR)*, p. 5197–5206, 2015.