HIGH-DIMENSIONAL EMBEDDING DENOISING AUTOENCODING PRIOR FOR COLOR IMAGE RESTORATION

Yuan Yuan[1], *Jinjie Zhou*[1], *Zhuonan He*[1], *Shanshan Wang*[2], *Biao Xiong*[3*], *Qiegen Liu*[1*]
[1]Department of Electronic Information Engineering, Nanchang University, Nanchang, 330031, China
[2]Paul C. Lauterbur Research Center for Biomedical Imaging, SIAT, CAS, Shenzhen 518055, China
[3]School of Computer Science and Technology, Wuhan University of Technology, Wuhan, 430070, China

ABSTRACT

This work exploits the basic denoising autoencoding (DAE) as enhanced priori for color image restoration (IR). The proposed method consists of two steps: enhanced DAE network learning and iterative restoration. To be special, at the training phase, a denoising network taking 6-dimensional variable as input is trained. Then, the network-driven high-dimensional prior information embedded DAE priori is utilized in the iterative restoration procedure. We first map the intermediate color image to be 6-dimensional and employ the higher-dimensional network to handle its corrupted version. The average operator is used to turn it back to the 3-channel image. The higher-dimensional prior alleviates the issue of the basic DAE that getting trapped in local optimal solution and effectively overcomes the instability. Experimental results on single image super-restoration (SISR) and deblurring demonstrate that the proposed algorithm can achieve good performance and prime visual inspection.

***Index Terms*—Color image restoration, denoising autoencoding, higher-dimensional embedding, SISR, deblurring**

1. INTRODUCTION

Image restoration (IR), which aims to recover high quality images from its degraded measurement, has always been a research hot spot in image processing [1-18]. The mathematical definition of IR can be formulated as:

$$f = Hu + n \tag{1}$$

where f denotes a corrupted image, u denotes a clean image, H represents a degradation matrix associated with the imaging system, and n is additive white Gaussian noise with standard deviation σ. For different degradation matrices, we can obtain different IR tasks accordingly. Two traditional IR tasks would be single image super-resolution (SISR) [1] when H is a composite operator of blurring and down-sampling, image deblurring [8] when H is a blurring matrix/operator. Because of the ill-posed essence of IR, the regularization-based techniques have been extensively studied by regularizing the solution spaces [1, 2, 4]. From a Bayesian viewpoint, the solution $\hat{u}$ is obtained by dealing with a Maximum A Posteriori (MAP) issue,

$$\hat{u} = \arg\min_{u} \| Hu - f \|^2 + \lambda\varphi(u) \tag{2}$$

*Corresponding author: b.xiong@whut.edu.cn, liuqiegen@ncu.edu.cn
This work was supported in part by the National NSFC under 61871206, 61661031, 81830056.

where $\varphi(u)$ is the regularizer connected with prior knowledge. The fidelity term $\| Hu - f \|^2$ ensures that the solution conforms to the degradation process. λ is an adjusted parameter.

In general, the solution to tackle Eq. (2) can be divided into two categories, namely model-based optimization method [1-18, 19-21] and discriminative learning method [22-27]. In the circumstances of the color IR, the classic algorithm is the kind of total variation (TV). For instance, Blomgr *et al.* [2] proposed total variation norm for vector valued functions applied to restore color image. A significant improvement has been made in which adaptive dictionary learning and sparse representation are applied in nonlocally extracted image patches of color images [4-11]. Dabov *et al.* [11] proposed the well-known block-matching and 3-D filter (BM3D) method for image denoising based on sparse representation in transform domain and a specially developed collaborative Wiener filtering. As an advanced formulation of the sparse representation, the structural sparse representation like low-rank prior is also exploited to color IR [12-16]. Gu *et al.* [13] studied the weighted nuclear norm minimization (WNNM) problem by exploiting the image non-local self-similarity, where the singular values were assigned different weights. Zhao *et al*. [15] proposed an adaptive clustering method and an adaptive rule for thresholding in PCA domain.

Nowadays, many deep learning-based SISR methods have been proposed, which aim to directly learn mappings from low-resolution (LR) patches to high resolution (HR) [22-27]. The SRCNN proposed by Dong *et al.* [22] used a three-layer network to learn the end-to-end mapping between LR images and HR images and has achieved real-time application speed. Wang *et al.* [23] proposed to combine the key ingredients of deep learning with the conventional sparse coding, and then they designed a neural network model for SISR. The VDSR method proposed by Kim *et al*. [25] adopted the residual-learning to explore contextual information over the large image regions.

In [19], denoising autoencoders network [28] as priors (DAEP) was developed to address various IR problems. A key advantage is that it just trains a single network for different IR tasks including deblurring with different kernels and super-resolution at different magnification factors. Although DAEP trained for a denoising instance can be used for removing different degradations, the output of the network is unstable at the same conditions and is prone to be trapped in locally optimal solutions. There is a large room in enhancing network representation with priors for performance improvement.

In this work, we explore the central idea of similarity samples clustering and aggregation strategy applied in the context of convolutional features learning and testing for color IR tasks. More specifically, built on the observation that optimally trained DAEP can provide good performance for image processing, we adopt a multi-channels and multi-models version of DAEP, termed M^{2}DAEP.

2. PROPOSED M^{2}DAEP

In this section, we present the 6D and multi-models derived M^{2}DAEP in details. First, we train the DAE network with input-output pair sample to be 6-channel image, consisting of the original color image and its copy version. Second, by means of auxiliary variable augmentation and average techniques, we use the resulting network-driven higher-dimensional prior to solve the color IR problem. Additionally, like in [19], multi-models scheme including employing different levels and different implementations of generating noises is put forward. Finally, with regard to the resulting IR mathematical model, proximal gradient method is adopted.

2.1. Motivation

In many non-local patch-based approaches, there usually consist of three steps: patch matching/clustering, sparsity or low-rank modeling as a regularizer, and patch weighted/aggregation modeling [4-15]. As shown in Fig. 1(a), the patch matching procedure enables multi-patches with similar structural patterns to be found and grouped. Meanwhile, patch aggregation strategy applied on the clustered patches can achieve better restoration. These two procedures play the role of converting pixel domain to patch domain and returning the restored results in patch domain to pixel domain, respectively.

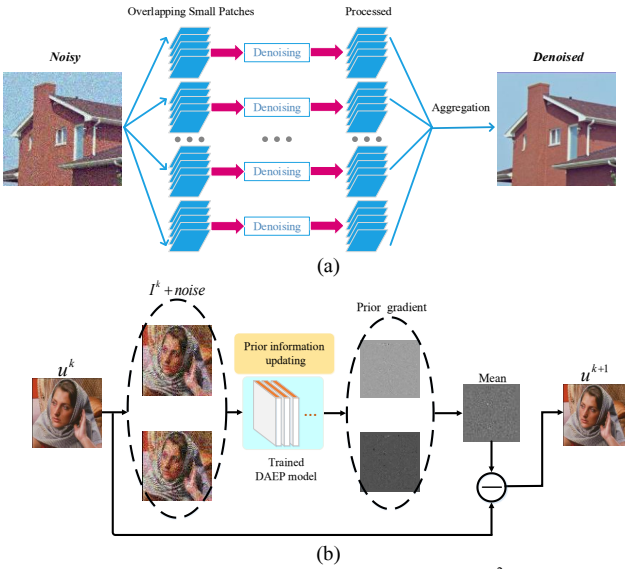

Fig. 1. Visual illustrations of patch-based approach and M^{2}DEAP. (a) Matrix operator in clustered patches for image denoising. (b) Convolution operator in multi-channel image features at iterative procedure of M^{2}EDAP.

Inspired by the central idea existed in the patch-based models, we adopt a 6-dimensional and multi-models version of DAEP for the color IR. Visual illustration of employing high-dimensional prior at iterative procedure in M^{2}DAEP is shown in Fig. 1(b). By mapping 3-channel image to be 6-channel via copy operator, we train a network taking 6-channel as input. After higher-dimensional denoising procedure, we use average operator to attain the solution.

In summary, in this work we use 6 channel image information as input and output samples to train DAE network, and obtain high dimensional priori information to solve the color IR tasks. More details of the training and testing phases of our M^{2}DAEP will be elaborated in the next two subsections. The proposed M^{2}DAEP involves two characteristics: (i) learns some prior information in higher-dimension space rather than the original space, and (ii) incorporates the higher-dimensional prior into the iterative restoration procedure for handling the original IR problem.

2.2. Proposed Training Model

At the prior learning stage, we train a 6-channel network from data pairs consisting of the 6-dimensional image (i.e., the original and its copy) and the noisy version. Accordingly, the M^{2}DAEP prior containing DAE $A_{\sigma_\eta}(I)$ is denoted as:

$$L_{M^2EDA} = E_{\eta,I}[\left\|I - A_{\sigma_\eta}(I)\right\|^2] \tag{3}$$

where the training data is a set of 6-dimensional images $\{I \mid I(u) = [u, u_1]\}$. $A_{\sigma_\eta}(I) = A(I+\eta)$ and η with standard deviation σ_η is the artificial noise. The copy u_1 of u is an auxiliary variable that accounts to map the 3-channel color image to be 6-channel. The auxiliary variable augmentation technique was also adopted in deep learning-based color-to-gray conversion [29].

The network architecture used for learning a DAE in this work is the residual encoder-decoder network (RED-Net) [27], as shown in Fig. 2. The RED-Net network consists of 20 layers, including 10 convolutional and 10 deconvolutional layers symmetrically arranged. Shortcuts connect matching convolutional and deconvolutional layers. Each layer is followed by its rectified linear units (ReLU). The convolution kernels are of size 3×3. The channel number at input and output layers is 6, and at the rest of the layers is 64.

The R, G, B channels in natural color images inherently involve some channel priors, the joint learning of the 6-channel data exhibits some structural information. In color IR task, we apply auxiliary variables technique to obtain the six-channel network induced high-dimensional structural prior information. This partly gives the rationality that the representation ability of M^{2}DAEP will be better than DAEP. In fact, there already exist many color IR works that utilize the channel priors in color images to enhance the restoration process [5, 17]. The utmost innovation here is that we learn prior information from 6D images and use it in color IR tasks.

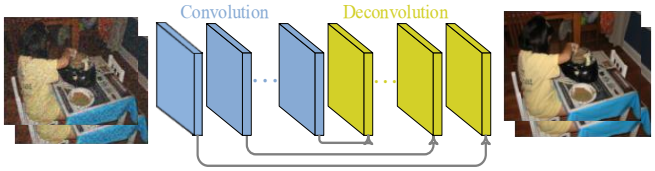

Fig. 2. The flowchart of training the 6-channel network in M^{2}EDAP.

2.3. Proposed IR Solver

Considering the 6D and multi-models ($N = 2$), the general mathematical model for color IR can be derived as follows:

$$\min_u \|Hu-f\|^2+\frac{\lambda}{N}\sum_{i=1}^{N}\left\|I(u)-A_{\sigma_{\eta i}}(I(u))\right\|^2 \quad (4)$$

For the sake of convenience, the 6-channel auxiliary variables $I(u)=[u,u_1]$ are simply termed as I . N stands for the number of M^{2}DAEP model. The first term is the data-fidelity term, and the second term consists of the network-driven prior information. Due to the nonlinearity of the model, we apply the proximal gradient method to tackle it [30]. The model is approximated by standard least square (LS) minimization:

$$\min_u \|Hu-f\|^2+\frac{\lambda}{\beta N}\sum_{i=1}^{N}\left\|I-(I^k-\beta\nabla G_i(I^k))\right\|^2 \quad (5)$$

where $G_i(I)=\left\|I-A_{\sigma_{\eta i}}(I)\right\|^2$ and $\nabla G_i(I)=[1-\nabla_I A^T_{\sigma_{\eta i}}(I)][I-A_{\sigma_{\eta i}}(I)]$. The function $G(I)$ is $1/\beta$ -Lipschitz smooth, i.e., $\left\|\nabla G(I')-\nabla G(I'')\right\|_2\le\left\|I'-I''\right\|_2/\beta$. k denotes the index number of iterations. Here, we empirically set β=1 and it works well in our experiments.

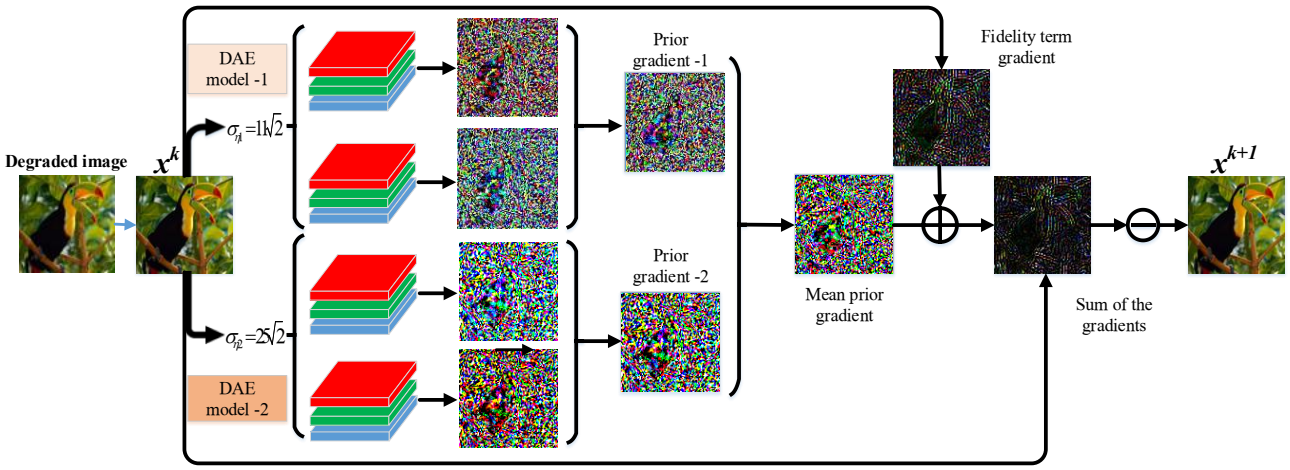

Fig. 3. The flowchart of employing the learned M^{2}EDAP to SISR application. Notice that multi-channels and multi-models is ultilized. 6-channal image is obtained by copying the original color image. Additionally, 6-channel networks implemeted at noise level $11\sqrt{2}$ and $25\sqrt{2}$ are simulateously used.

Given $\beta=1$, Eq. (5) is a standard LS problem, which can be solved by calculating the gradient as follows:

$$H^T(Hu-f)+\lambda\{I+\frac{1}{N}\sum_{i=1}^{N}[\nabla_I A^T_{\sigma_{\eta i}}(I^k)(A_{\sigma_{\eta i}}(I^k)-I^k)-A_{\sigma_{\eta i}}(I^k)]\}=0 \quad (6)$$

and it yields:

$$u^{k+1}=\frac{H^Tf+\frac{\lambda}{N}\sum_{i=1}^{N}R[\{A_{\sigma_{\eta i}}(I^k)-\nabla_I A^T_{\sigma_{\eta i}}(I^k)[A_{\sigma_{\eta i}}(I^k)-I^k]\}]}{(H^TH+\lambda)} \quad (7)$$

where R stands for the mean operator employed on the six channels. It can be observed that the solution formulation contains $A_{\sigma_{\eta i}}(I^k)$ and $\nabla_I A^T_{\sigma_{\eta i}}(I^k)[A_{\sigma_{\eta i}}(I^k)-I^k]$, where the parameters $A_{\sigma_{\eta i}}(\circ)$ are already learned at the network training stage. In particular, the $A_{\sigma_{\eta i}}(I^k)$ is the forward output with the input $I^k+\sigma_\eta$. $\nabla_I A^T_{\sigma_{\eta i}}(I^k)[A_{\sigma_{\eta i}}(I^k)-I^k]$ is the backward network output with the input $A_{\sigma_{\eta i}}(I^k)-I^k$. Additionally, we update the solution u^k by alternately updating the network estimation $A_{\sigma_{\eta i}}(I^k)$, $\nabla_I A^T_{\sigma_{\eta i}}(I^k)$ and the LS solver until the u value convergences. In brief, the mathematical model is tackled by the proximal gradient and alternative optimization.

The flowchart for illustrating the M^{2}DAEP training stage is shown in Fig. 3. In summary, the overall training phase and testing phase of M^{2}DAEP algorithm are as follows:

Algorithm: M^{2}DEAP

Training stage

Training images: 6-dimensional dataset $\{I\,|\,I(u)=[u,u_1]\}$

Network: 6-channel DAE network

Outputs: Trained network $A_{\sigma_{\eta 1}}(\circ)$ and $A_{\sigma_{\eta 2}}(\circ)$

Testing stage

Initialization: $u^0=H^Tf$; K ; $N=2$

For $k=1,\ 2,\ \cdots,\ K$

Update the auxiliary variable: $I^k=[u^k,u_1^k]$

Calculate the prior gradient components:

$A_{\sigma_{\eta i}}(I^k)$, $\nabla_I A^T_{\sigma_{\eta i}}(I^k)[A_{\sigma_{\eta i}}(I^k)-I^k]$; $i=1,2,\cdots,N$

Update the solution via solving the LS problem:

$$u^{k+1}=\frac{H^Tf+\frac{\lambda}{N}\sum_{i=1}^{N}R[\{A_{\sigma_{\eta i}}(I^k)-\nabla_I A^T_{\sigma_{\eta i}}(I^k)[A_{\sigma_{\eta i}}(I^k)-I^k]\}]}{(H^TH+\lambda)}$$

End

3. EXPERIMENTAL RESULTS

In this section, the performance of M^{2}DAEP is evaluated for SISR and image deblurring. We implemented the proposed architecture in MATLAB on a PC equipped with Inter(R) Core (TM) i7-7700 CPU and GeForce Titan XP. Our training set use 400 images of 180×180 and set the training patch size as 40×40 . Image patches with different noise levels were used to train models for learning a mapping from noisy images to denoising results. To evaluate the quality of the restoration image,

PSNR (Peak Signal to Noise Ratio, dB) is calculated. The higher PSNR values mean that retaining the more structures with better visual quality. For the convenience of reproducible research, codes source code of M^{2}DAEP is available at: https://github.com/yqx7150/M2DAEP.

3.1. Single image Super-Resolution (SISR)

We compared our method with the recent techniques, including SRCNN [25], TNRD [31], DnCNN-3 [24], IRCNN [32], DMSP [20], SRMD [26] and DAEP [19], where the networks of SRCNN, TNRD, and SRMD were trained separately for each scale factor, however, the DnCNN-3 and IRCNN models were trained jointly on ×2, 3 and 4. Specifically, for the super-resolution under scale ×5, we used the SRCNN and TNRD models trained on ×4, and the DnCNN-3 and IRCNN models trained jointly on ×2, ×3 and ×4.

Table 1 shows the performance of above different methods on Set 5 for varying scale factors. It can be seen that M^{2}DAEP achieves higher PSNR results, except for SRMD. More importantly, M^{2}DAEP gains higher PSNR values than our previous DAEP [19], implying that the quality of DAEP is significantly improved by embedding higher-dimensional structure. Fig. 4 depicts some visual inspections.

Fig. 4. Comparison of super-resolution on image "Butterfly" in Set 5 for scale ×3 with the corresponding PSNR (dB) values. From top to bottom and left to right: SRCNN, DnCNN-3, IRCNN, DMSP, SRMD, DAEP and M^{2}DAEP.

Table 1. Average PSNR (dB) of different methods on Set 5 for the different scale factors.

Scale	Dataset	Bicubic	SRCNN	TNRD	DnCNN-3	IRCNN	DMSP	SRMD	DAEP	MDAEP	M^{2}DAEP
×2	Set 5	31.80	34.50	34.62	35.20	35.07	35.16	35.28	35.23	35.41	**35.48**
×3		28.67	30.84	31.08	31.58	31.25	31.38	**31.84**	31.44	31.68	31.75
×4		26.73	28.60	28.83	29.30	29.00	29.14	**29.64**	29.01	29.25	29.44
×5		25.32	26.12	26.88	26.30	27.13	27.35	-	27.19	27.40	**27.59**

3.2. Image Deblurring

In deblurring experiment, a blur kernels with 35×25 size is selected as the degrading operator. Four color images in Fig. 5 were tested to verify the performance of the proposed M^{2}DAEP method with comparison EPLL framework [12], DMSP [20] and DAEP [19]. The PSNR values recorded in Table 2 indicate that M^{2}DAEP achieves the highest values for almost images.

Visual quality comparison of image deblurring at varying Gaussian noise level for the color images "Lena" with size of 256×256 is shown in Fig. 6. It can be observed that the DAEP method can remove noise very well while the image is still blurry. In additional, the EPLL method can well reconstruct the piecewise smooth regions but often fails to recover fine image details. Finally, the DMSP and DAEP can outperform the EPLL largely. Particularly, DMSP and M^{2}DAEP methods not only can remove noise but also preserve the structure details. Moreover, the proposed M^{2}DAEP produces cleaner and sharper image edges and textures than other competing methods.

Fig. 5. The photographic images. *Barbara, Butterfly, House and Lena*.

Table 2. The deblurring performance (PSNR) on four images.

Noisy	image	EPLL	DMSP	DAEP	M^{2}DAEP
7.65	Barbara	20.10	**21.23**	20.66	21.10
	Butterfly	19.03	20.18	25.45	**26.97**
	House	24.05	25.45	25.47	**29.01**
	Lena	23.56	26.97	27.37	**28.68**
	Average	21.69	23.45	24.47	**26.07**
12.75	Barbara	19.59	**20.21**	19.33	19.54
	Butterfly	18.81	19.52	22.68	**23.47**
	House	22.57	26.25	24.74	**28.54**
	Lena	24.91	24.25	25.94	**27.45**
	Average	21.47	22.56	23.17	**24.75**

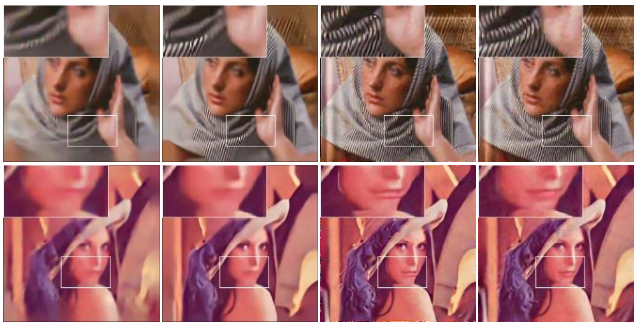

Fig. 6. Visual comparison of image deblurring. Top line: image "Barbara" under noise level of 7.65. Bottom line: image "Lena" under noise level of 12.75. Results obtained by EPLL, DMSP, DAEP and M^{2}DAEP.

4. CONCLUSIONS

This work paved a new way to incorporate higher-dimensional prior information into color IR applications. Specifically, a 6-channel denoising autoencoder prior was presented, which built on the assumption that an optimal denoising autoencoder is a local mean of the correct data density. In particular, auxiliary variables technique was applied to integrate higher-dimensional structural information. The formulated mathematic model was tackled by proximal gradient and alternative optimization. Both qualitative and quantitative results on SISR and image deblurring demonstrated that M^{2}DAEP achieves performance improvements over state-of-the-art methods.

5. REFERENCES

[1] A. Marquina and S. J. Osher, "Image super-resolution by TV regularization and bregman iteration," *Journal of Scientific Computing*, vol. 37, no. 3, pp. 367-382, 2008.

[2] P. Blomgren, T. F. Chan, "Color TV: total variation methods for restoration of vector-valued images." *IEEE Transactions on Image Processing*, vol. 7, no. 3, pp. 304-309, 1998.

[3] Y. Dong, M. Hintermüller, M. M. Rinconcamacho, "A multi-scale vectorial L, τ, -TV framework for color image restoration," *Kluwer Academic Publishers,* 2011.

[4] J. Mairal, M. Elad, G. Sapiro, "Sparse representation for color image restoration," *IEEE Transactions on Image Processing*, vol. 17, no. 1, pp. 53-69, 2007.

[5] H. Mousavi, V. Monga, "Sparsity-based color image super resolution via exploiting cross channel constraints," *IEEE Transactions on Image Processing*, vol. 26, no. 11, pp. 5094-5106, 2017.

[6] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution as sparse representation of raw image patches," in *Proc. of the IEEE CVPR*, pp. 1-8, 2008.

[7] X. Gao, K. Zhang, D. Tao, and X. Li, "Image super-resolution with sparse neighbor embedding," *IEEE Transactions on Image Processing*, vol. 21, no. 7, pp. 3194-3205, 2012.

[8] W. Dong, L. Zhang, G. Shi, and X. Wu, "Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization," *IEEE Transactions on image processing*, vol. 20, no. 7, pp. 1838-1857, 2011.

[9] A. Danielyan, V. Katkovnik, and K. Egiazarian, "Bm3d frames and variational image deblurring," *IEEE Transactions on image processing*, vol. 21, no. 4, pp. 1715-1728, 2012.

[10] W. Dong, G. Shi, Y. Ma, and X. Li, "Image restoration via simultaneous sparse coding: Where structured sparsity meets gaussian scale mixture," *International Journal of Computer Vision*, pp. 1-16, 2015.

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

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

[13] S. Gu, L. Zhang, W. Zuo , X. Feng, "Weighted nuclear norm minimization with application to image denoising," *IEEE Conference on Computer Vision & Pattern Recognition*, 2014.

[14] G. Yu, G. Sapiro, and S. Mallat, "Solving inverse problems with piecewise linear estimators: From gaussian mixture models to structured sparsity," *IEEE Transactions on Image Processing*, vol. 21, no. 5, pp. 2481-2499, 2012.

[15] W. Zhao, Y. Lv, Q. Liu, B. Qin, "Detail-preserving image denoising via adaptive clustering and progressive PCA thresholding," *IEEE Access*, vol. 6, 6303-6315, 2018.

[16] B. Xiong, Q. Liu, J. Xiong, S. Li, S Wang, D Liang, "Field-of-Experts filters guided tensor completion," *IEEE Transactions on Multimedia*, vol. 20, no. 9, pp. 2316-2329, 2018.

[17] N. Joshi, C L.Zitnick, R. Szeliski, D. Kriegman, "Image deblurring and denoising using color priors," *IEEE Conference on Computer Vision & Pattern Recognition,* 2009.

[18] R. Timofte, V. De Smet, and L. Van Gool, "A+: Adjusted anchored neighborhood regression for fast super-resolution," in Asian Conference on *Computer Vision. Springer*, pp. 111-126, 2014.

[19] S.A. Bigdeli and M. Zwicker, "Image restoration using autoencoding priors," arXiv preprint arXiv: 1703.09964, 2017.

[20] S. A. Bigdeli, M. Zwicker, "Deep mean-shift priors for image restoration", *In Proceedings of the 31st Conference on Neural Information Processing Systems*, 2017.

[21] Y Wang, Q Liu, H Zhou, Y Wang, "Learning multi-denoising autoencoding priors for image super-resolution," *Journal of Visual Communication and Image Representation*, vol. 57, pp. 152-162, 2018.

[22] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *European Conference on Computer Vision*. Springer, pp. 184-199, 2014.

[23] Z. Wang, D. Liu, J. Yang, W. Han, and T. Huang, "Deep networks for image super-resolution with sparse prior," in Proc. of the *IEEE ICCV*, pp. 370-378, 2015.

[24] 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.

[25] J. Kim, J. K. Lee, K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," *IEEE Conference on Computer Vision & Pattern Recognition*, pp. 1646-1654, 2016.

[26] K. Zhang, W. Zuo, L. Zhang, "Learning a single convolutional super-resolution network for multiple degradations," *IEEE Conference on Computer Vision & Pattern Recognition*, 2018.

[27] X.J. Mao, C. Shen, and Y.B. Yang, "Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections," in *Proc. NIPS*, 2016.

[28] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, "Extracting and composing robust features with denoising autoencoders," in *International Conference on Machine Learning,* pp. 1096-1103, 2008.

[29] Q. Liu, H. Leung, "Variable augmented neural network for decolorization and multi-exposure fusion," *Information Fusion*, vol. 46, 114-127, 2019.

[30] N. Parikh and S. Boyd, "Proximal algorithms," *Found. Trends Optim.*, vol. 1, no. 3, pp. 127–239, 2014.

[31] Y. Chen, 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

[32] K. Zhang, W. Zuo, Y. Chen, S. Gu, L. Zhang, "Learning deep CNN denoiser prior for image restoration", IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, 2017.