

Multi-Channel and Multi-Model-Based Autoencoding Prior for Grayscale Image Restoration

Sanqian Li, Binjie Qin ^{ID}, *Member, IEEE*, Jing Xiao ^{ID}, *Member, IEEE*, Qiegen Liu ^{ID}, *Member, IEEE*,
Yuhao Wang ^{ID}, *Senior Member, IEEE*, and Dong Liang ^{ID}, *Senior Member, IEEE*

Abstract—Image restoration (IR) is a long-standing challenging problem in low-level image processing. It is of utmost importance to learn good image priors for pursuing visually pleasing results. In this paper, we develop a multi-channel and multi-model-based denoising autoencoder network as image prior for solving IR problem. Specifically, the network that trained on RGB-channel images is used to construct a prior at first, and then the learned prior is incorporated into single-channel grayscale IR tasks. To achieve the goal, we employ the auxiliary variable technique to integrate the higher-dimensional network-driven prior information into the iterative restoration procedure. In addition, according to the weighted aggregation idea, a multi-model strategy is put forward to enhance the network stability that favors to avoid getting trapped in local optima. Extensive experiments on image deblurring and deblocking tasks show that the proposed algorithm is efficient, robust, and yields state-of-the-art restoration quality on grayscale images.

Index Terms—Grayscale image restoration, denoising autoencoder network, channel prior, auxiliary variable technique, multi-model, proximal gradient descent.

I. INTRODUCTION

THE task of image restoration (IR) is to recover a high-quality image from its degraded measurement, which is known to be an ill-posed inverse problem. The under-determined problem needs to be constrained by effective priors for ensuring acceptable solutions. Good priors can yield pleasing results and have been applied to different tasks including image deblurring [1]–[7], image deblocking [8]–[12],

etc. [13]–[19]. The mathematical formulation of IR can be generally modeled as:

$$f = Mu + e \quad (1)$$

where u is the unknown image to be estimated, f represents the degraded image and e is additive noise. M is the degrading operator: If M is a blurring operator or a down-sampling operator, IR becomes image deblurring or image super-resolution [5]; When M is the identity operator and e is the quantization noise, the task becomes image deblocking [8]. In this paper, we focus on grayscale image deblurring and deblocking applications.

In the past decade, there are many prior-guided IR approaches based on non-local patch modeling [4]–[9], [14]–[17], [20]–[22]. For example, Awate and Whitaker [20], [21] described an unsupervised, information-theoretic, adaptive filter (UINTA) that could automatically discover the statistical properties of the signal and thereby restore a wide spectrum of images. Kindermann *et al.* [1] investigated the regularization functionals with non-local correlation terms for image deblurring and defined a non-local variant of the well-known bounded variation regularization, which does not suffer from staircase effects. Dabov *et al.* [15] proposed the well-known block-matching and 3D filter (BM3D) method for image denoising based on sparse representation in transform domain and a specially developed collaborative Wiener filtering. Mairal *et al.* [4] explicitly exploited the self-similarities of patches in natural images and demonstrated that the self-similarities can result in successful non-local means approach. Zoran and Weiss [6] compared the likelihood of several patch models and showed that priors can give high likelihood to data. To improve the performance of sparse representation-based IR, the concept of sparse coding noise was introduced in [7], where the authors exploited the non-local self-similarity and obtained a good estimate for blurred images. Milanfa [22] presented a practical and accessible framework for arriving at new insights to understand some of the basic methods. In particular, they discussed several novel optimality properties of algorithms such as BM3D, and methods for their iterative improvement. Gu *et al.* [16] studied the weighted nuclear norm minimization (WNNM) problem by exploiting the non-local self-similarity, where the singular values were assigned different weights. Ren *et al.* proposed a novel image deblocking method via patch clustering

Manuscript received December 27, 2018; revised June 4, 2019 and July 11, 2019; accepted July 17, 2019. Date of publication July 31, 2019; date of current version September 12, 2019. This work was supported in part by the National Natural Science Foundation of China under Grant 61871206 and Grant 61661031 and in part by the Project of Innovative Special Funds for Graduate Students in Jiangxi Province under Grant YC2017-S108. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Lisimachos P. Kondi. (*Corresponding authors: Qiegen Liu; Dong Liang.*)

S. Li, Q. Liu, and Y. Wang are with the Department of Electronic Information Engineering, Nanchang University, Nanchang 330031, China (e-mail: lisanqian@email.ncu.edu.cn; liuqiegen@ncu.edu.cn; wangyuhao@ncu.edu.cn).

B. Qin is with the School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: bj Qin@sjtu.edu.cn).

J. Xiao is with the School of Computer Science, Wuhan University, Wuhan 430072, China (e-mail: jing@whu.edu.cn).

D. Liang is with the Paul C. Lauterbur Research Center for Biomedical Imaging, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China (e-mail: dong.liang@siat.ac.cn).

Digital Object Identifier 10.1109/TIP.2019.2931240

and low-rank minimization, which simultaneously exploits the local and non-local sparse representations in a unified framework [11]. Zhang *et al.* [12] also exploited the non-convex rank minimization on non-local blocks for reducing compression artifacts.

In the patch modeling procedure, patch matching/clustering and patch weighted/aggregation are two main procedures that contribute to the success of non-local based approaches [17]. The patch matching procedure enables image 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 the objects from pixel domain to patch domain and returning the restored results from patch domain to pixel domain, respectively.

As an advanced formulation of the patch-based approaches, several researchers studied some discriminative learning methods to learn image priors by means of convolutional features. For example, the shrinkage fields proposed by Schmidt and Roth [18] is a random field-based architecture, and it unifies the random field-based model and the half-quadratic optimization algorithm into a single learning framework. With the development of deep learning [23]–[26], convolutional neural networks (CNN) [19] was applied in IR due to its increased capacity. For example, Zhang *et al.* [19] proposed a denoising convolutional neural network (DnCNN) for image denoising and general IR tasks. DnCNN treats image denoising as a plain discriminative learning problem to embrace the progress in very deep architecture and regularization method. However, it works well only if the noise level is in the preset range due to the lack of network flexibility. In order to alleviate the drawback of DnCNN that lacks flexibility to deal with strong noise, Zhang *et al.* [27] further proposed a fast and flexible solution for CNN based image denoising, namely FFDNet. FFDNet works on downsampled subimages and achieves a good trade-off between inference speed and denoising performance. In [29], denoising autoencoders network [28] as priors (DAEP) was developed to address IR problems. A key advantage of DAEP 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 representation and output of network are unstable at the same conditions. Hence, there is a large room in enhancing network representation with better 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 IR tasks. More specifically, built on the observation that optimally trained DAEP can provide a good performance for image processing, we adopt a multi-channel and multi-model version of DAEP for the grayscale IR problems.

The contributions of this work are summarized as follows:

- Multi-channel strategy: An enhanced DAEP (EDAEP) is explored for recovering grayscale image. At the network training stage, the multi-channel prior information

is obtained via the relevance among R, G, B channel of color image training sets. Then, at the iterative IR stage for grayscale image, auxiliary variable technique is applied to embed the three-channel prior into the single-channel intermediate solution.

- Multi-model strategy: In order to enable the EDAEP to be generally applied for different tasks with enhanced network stability, multi-model weighted EDAEP (MEDAEP) is developed. According to the aggregation principle, the multi-model scheme prefers to avoid getting stuck in local minima and makes the iterative process to be more robust.
- After formulating the mathematical model for MEDAEP, we adopt the alternative optimization and proximal gradient method to tackle the non-convex grayscale IR minimization.

The remainder of the paper is organized as follows. Section II provides a brief description of preliminary work with regard to the basic DAE network and DAEP. Section III presents MEDAEP model and the corresponding iterative solver. Extensive experimental comparisons between the proposed MEDAEP and state-of-the-art methods are conducted in Section IV. Finally, concluding remarks and directions for future research are given in Section V.

II. PRELIMINARIES

Designing suitable priors is a classical strategy for successfully constraining IR problem. Rather than learning a discriminative model with an explicit image prior, deep learning has been successfully exhibited in various low-level IR tasks recently by treating image denoising as a plain discriminative learning problem [19]. However, it often trains an end-to-end network for a specific task. Subsequently, more time is spent on training various models for different IR tasks. Built on the key observation [26] that the output of an optimal denoising autoencoder is a local mean of the true natural image density for each input, DAE network was used to construct priors to address IR tasks and achieved computational efficiency [29]. It was not necessary to train separate networks at different circumstances.

A. DAE Network

The DAE architecture is motivated by residual network (ResNet) [26]. ResNet contains the following main characteristics: Rectified Linear Unit (ReLU) [26], Batch normalization (BN) [25] and Residual learning [23]. It is well known that ReLU layer [24] can introduce sparsity automatically into the networks and result in learning features faster. BN [25] is proposed to accelerate network learning and boost accuracy via normalizing the weights and parameters to avoid outliers in the intermediate feature maps. Finally, residual learning [26] explicitly lets the stacked layers fit a residual mapping, which is assumed to be easier for optimization. In summary, equipped with the above three operators, extremely deep CNN can be easily trained and achieves improved accuracy [26].

Specifically, the DAE network consists of 20 convolutional (Conv) layers followed by BN except for the first and last

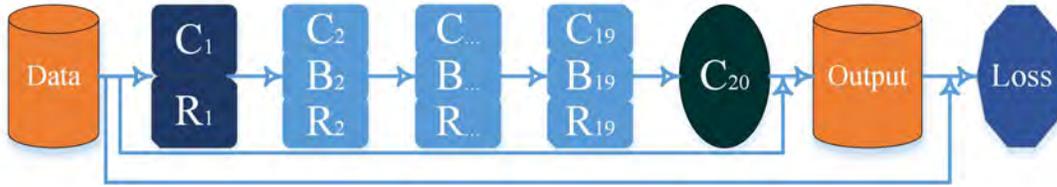


Fig. 1. The schematic flowchart of DAE network. The Conv layers, BN layers and ReLU layers are denoted as “C”, “B”, and “R”, respectively. The number at the subscript of word “C”, “B”, “R” stands for the layer index.

layers. Additionally, DAE uses ReLU activations except for the last convolutional layer. The size of kernels is 3×3 among all convolutional layers and there exist 64 filters (i.e. 64 channels) for the intermediate layers. In order to match color images and grayscale images, the channel number at the first and last layers are set to be 3 and 1, respectively. The schematic flowchart of DAE network model is shown in Fig. 1. The Conv, BN and ReLU layers are denoted as “C”, “B”, and “R”, respectively. For the first layer, 64 filters of size $3 \times 3 \times 3$ (color images with 3 channels, i.e. R, G, B) are used to generate 64 feature maps, and ReLU is utilized for nonlinearity. For 2-19 layers, 64 filters of size $3 \times 3 \times 64$ are used, and BN is used between convolution layer and ReLU layer. For the last layer, 3 filters of size $3 \times 3 \times 3$ are used to reconstruct the images.

It is worth noting that, in most of the existing CNN algorithms, the channel number of network input and output at the training phase should be the same as that at the testing phase. Take the basic DAE network for grayscale image as an example, DAE is first trained from grayscale image dataset and then the learned network is used for testing grayscale images again. Instead, the core idea in our work is to train the DAE network with RGB-channels variables at first and then employ it to the single-channel grayscale IR. The principle behind the new idea is to exploit the higher-dimensional structure information for enhancing the DAE priors, as will be introduced below.

B. DAE as Priors

For the sake of tackling the ill-posed IR problem, regularization term $R(u)$ has been used to incorporate the image prior information. Formally, the desired IR solution can be achieved by solving the following minimization:

$$\min_u \|Mu - f\|^2 + \lambda R(u) \quad (2)$$

where λ is a penalty parameter. M is a two-dimensional degradation operation.

The key idea of DAEP is leveraging a neural autoencoder to define a natural image prior. Specifically, denoting a DAE as A_{σ_η} and u as the input image, then the DAE output $A_{\sigma_\eta}(u)$ is trained by adding artificial Gaussian noise and using an expected quadratic loss:

$$L_{DAE} = E_{\eta, u} [\|u - A_{\sigma_\eta}(u + \eta)\|^2] \quad (3)$$

where the expectation is over all images u and Gaussian noise η with standard variance σ_η .

According to [28], it is revealed that the network output $A_{\sigma_\eta}(u)$ is related to the true data density $q(u)$ as follows:

$$\begin{aligned} A_{\sigma_\eta}(u) &= \frac{\int (u - \eta) g_{\sigma_\eta}(\eta) q(u - \eta) d\eta}{\int g_{\sigma_\eta}(\eta) q(u - \eta) d\eta} \\ &= u - \frac{\int g_{\sigma_\eta}(\eta) q(u - \eta) \eta d\eta}{\int g_{\sigma_\eta}(\eta) q(u - \eta) d\eta} \end{aligned} \quad (4)$$

where $g_{\sigma_\eta}(\eta)$ represents a local Gaussian kernel with standard variance σ_η . It can be observed that the output of an optimal DAE $A_{\sigma_\eta}(u)$ is a local mean of the true data density and the autoencoder error is a mean shift vector [30].

More importantly, following Eq. (4), it can be derived that the autoencoder error $A_{\sigma_\eta}(u) - u$ is proportional to the gradient of the log likelihood of the smoothed density, i.e.,

$$A_{\sigma_\eta}(u) - u = \sigma_\eta^2 \nabla \log[g_{\sigma_\eta} * q](u) \quad (5)$$

where the data distribution is $Probability(u) = \int q(u + \eta) d\eta$. The autoencoder error vanishes at stationary points, including local extrema, of the true density smoothed by the Gaussian kernel. Hence, DAEP utilizes the migratory characteristic of prior information $R(u)$ and uses the magnitude of this mean shift vector $\|A_{\sigma_\eta}(u) - u\|^2$ as the negative log likelihood of natural image prior. It is interesting that this regularization is perceptually similar to the well-known non-local means and non-local total variation [31], [32].

III. PROPOSED MEDAEP MODEL

In this section, we present the multi-channel and multi-model derived DAEP model in details. First, motivated by the idea of utilizing multi-channel prior information for single-channel image recovery, we train the DAE network with RGB-channel images as input-output pair samples and use the resulting higher-dimensional prior to solve grayscale IR problems. Second, according to the weighted aggregation idea, a multi-model scheme including employing different levels and different implementations of generating noises is put forward to enhance the iterative stability. Subsequently, it avoids getting stuck in locally optimal solutions. Finally, with regard to the resulting IR mathematical model, we adopt the alternative iterative strategy and proximal gradient method to address it.

A. Multi-Channel Enhanced DAEP (EDAEP)

The proposed EDAEP involves two characteristics: (i) learning some prior information in higher-dimensional

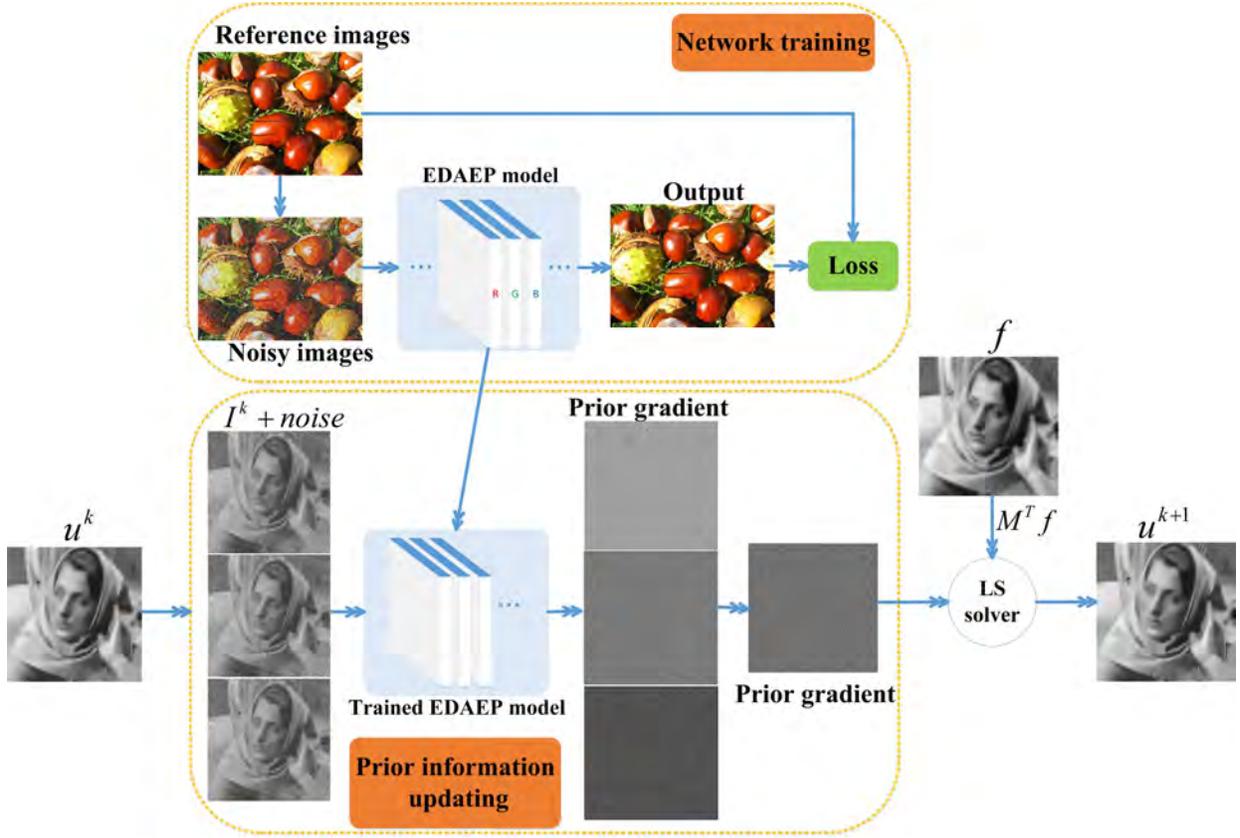


Fig. 2. Flowchart illustration of the proposed EDAEP. Top: the RGB three-channel network scheme at the training stage. Bottom: the auxiliary variable technique used for single-channel intermediate image at the iterative restoration phase. The auxiliary variable is $I^k = [u^k, u^k, u^k]$. Prior gradient stands for the gradient of the prior information.

space, rather than the original space; (ii) incorporating the higher-dimensional prior into the iterative restoration procedure to handle the original IR problem. Three important steps are as follows:

First, at the prior learning stage, we train a three-channel network from data pairs consisting of the original color image and its noisy version. Accordingly, the EDAEP prior is denoted as:

$$L_{EDAEP} = E_{\eta, I} [\|I - A_{\sigma_\eta}(I + \eta)\|^2] \quad (6)$$

where the training data is a set of color images $\{I | I = [I_r, I_g, I_b]\}$. As can be concluded, the autoencoder error $A_{\sigma_\eta}(I) - I$ of the EDAEP prior is

$$A_{\sigma_\eta}(I) - I = \sigma_\eta^2 \nabla \log [g_{\sigma_\eta} * q](I) \quad (7)$$

where the data distribution is $Probability(I) = \int q(I + \eta) d\eta$. Compared Eq. (7) with Eq. (5), it can be seen that, the effectiveness of prior (i.e., the autoencoder error) depends on the distribution of training data, and is proportional to the gradient of its log-density. As well known, the RGB-channel color image contains more structural information than that in the corresponding single-channel grayscale image, hence prior-based regularization can be fully exploited in color image. In fact, there already exist many works that utilize the channel priors in color images to enhance the restoration process [33]–[35]. The utmost

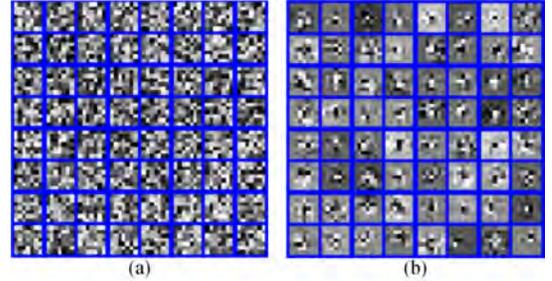


Fig. 3. Representative convolutional kernels of the final convolutional layer within the ResNet block with $\sigma_\eta = 25$. (a) DAEP and (b) EDAEP.

innovation here is that we learn prior information from color images, and then use it in grayscale IR tasks. The above theoretical analysis partly gives the rationality that the representation and recovery ability of EDAEP will be better than DAEP.

A flowchart for illustrating the EDAEP training stage is shown in Fig. 2, where the detailed network architecture is shown in Fig. 1. The visualized kernel comparison between DAEP learned by single-channel data and EDAEP learned by RGB-channels data is depicted in Fig. 3. The representative convolutional kernels of the final convolutional layer within the ResNet block in DAEP tend to be random distributional. By contrast, the convolutional kernels in EDAEP look more

regular and contain more structural information. This phenomenon indicates that, although the input noises in three channels are randomly given, the R, G, B channels in natural color images inherently involve some channel priors, the joint learning of the three-channel data exhibits some structural information.

Second, in the single-channel grayscale IR task, we apply auxiliary variable technique to utilize the three-channel network induced high-dimensional structural prior information. Specifically, after the network-driven prior learning procedure, the mathematical model for grayscale image recovery can be achieved by solving the following minimization:

$$\min_u \|Mu - f\|^2 + \lambda \|I(u) - A_{\sigma_\eta}(I(u))\|^2 \quad (8)$$

where the auxiliary variable $I(u) = [u, u, u]$ is a three-channel image that the element at each channel is copied by the desired solution u . In brief, the duplicated formulation $I(u)$ of u is dubbed as I . Obviously, the space of testing samples belongs to the space of learning samples, i.e., $\{I(u) = [u, u, u]\} \subset \{I | I = [I_r, I_g, I_b]\}$. The essence of Eq. (8) is to integrate the prior information learned from higher-dimensional space into the lower-dimensional IR problem. It is pointed out that the idea is partially motivated by our previous work in image decolorization and image fusion [36], [37], where auxiliary variable technique was applied to the grayscale image to obtain a new variable whose elements at three channels were the same. The flowchart of illustrating the EDEAP testing stage is shown in Fig. 2.

Last but not least, as will be shown in Section III.C, the solver of Eq. (8) can be decomposed to the alternative minimization procedure between data-fidelity updating and prior information updating. In the popular patch-based IR methods, they often involve a patch extraction stage and a patch weighting/average stage, such as exchanging between the pixel domain and patch domain. The proposed method also adopts similar strategy. Specifically, for constructing the variable I^k at the iterative procedure of the testing stage, we can copy and rearrange the single channel image u^k to be a multi-channel image formulation with same values, thus the prior information of single-channel grayscale image can be obtained by utilizing the trained multi-channel model. After obtaining the prior information I^{k+1} , we can average the three-channel output to generate the single-channel image u^{k+1} and regard it as the intermediate result at the next iteration.

B. Multi-Model Weighted EDAEP (MEDAEP)

Recent advances have shown that better recovery performance can be obtained by employing aggregation function to algorithms in various situations as it promotes more sparsity, while at the price of higher computational complexity. For example, Salmon [38] proposed an efficient modification of the central weight based on the Stein's Unbiased Risk estimate principle and investigated how to exploit this theoretical framework to design novel weighting schemes. Zhang *et al.* [39] proposed that exploit the intensity similarity and geometry closeness of the denoised patches, to reduce the unwanted artifacts in the synthesized denoised image.

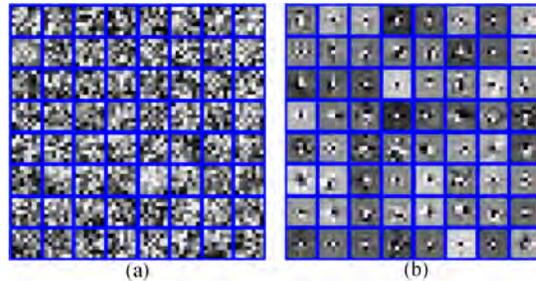


Fig. 4. Representative convolutional kernels of the final convolutional layer within the ResNet block with $\sigma_\eta = 15$. (a) DAEP and (b) EDAEP.

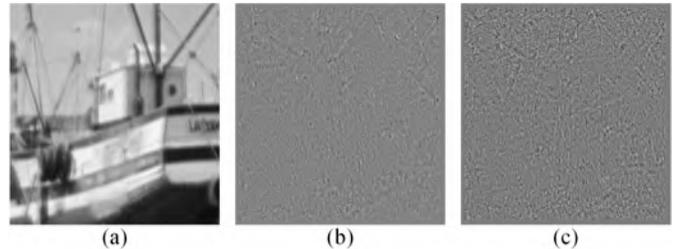


Fig. 5. Illustration of the gradient image at $\sigma_{\eta_1} = 15$ and $\sigma_{\eta_2} = 25$. (a) The corrupted image with Gaussian blur size 13×13 and noisy level $\delta_\epsilon = 2.55$. (b) Prior gradient image with $\sigma_{\eta_1} = 15$. (c) Prior gradient image with $\sigma_{\eta_2} = 25$.

It is well known that good representation learning should contain abundant features at different levels of granularity. Thus, the networks stability can be enhanced and local solutions may be avoided. Considering that EDAEP performance is influenced with the noise distribution, it is desirable to further exploit random Gaussian noise at different levels with standard variation σ_η . Specifically, when the standard variation σ_η is set to be relatively large, the EDAEP model removes high-frequency information so that it prefers to denoise better and produces more smoothed results. On the other hand, if the standard variation σ_η is set to be relatively smaller, it tends to retain more detail while containing more noisy results. In other words, at high noise level, the highly corrupted training data forces the network to learn more global and coarse-grained features. Meanwhile, at low noise level, the network is devoted to learning features for reconstructing fine details of the training data. Comparing to the trained networks with $\sigma_\eta = 25$ shown in Fig. 3, one visual demonstration of the networks with $\sigma_\eta = 15$ is shown in Fig. 4.

In order to illustrate the importance of the training data at different noise levels for restoration results, we provide two prior gradient images (i.e., the gradients of the prior) at $\sigma_{\eta_1} = 15$ and $\sigma_{\eta_2} = 25$ in Fig. 5, respectively. As can be observed, the gradient prior image with $\sigma_{\eta_2} = 25$ focuses on more large-scale structural information while gradient prior image with $\sigma_{\eta_1} = 15$ is devoted to more small-scale detail information. Inspired by the work in [38], [39] and for the sake of eliminating the influence of noise distribution in DAEP, we adopt the aggregation function on the prior term with different Gaussian noise, as shown in Fig. 6. For conveniences, by assuming that the estimates from different noise branches contribute to the final output equally, we formulate the prior

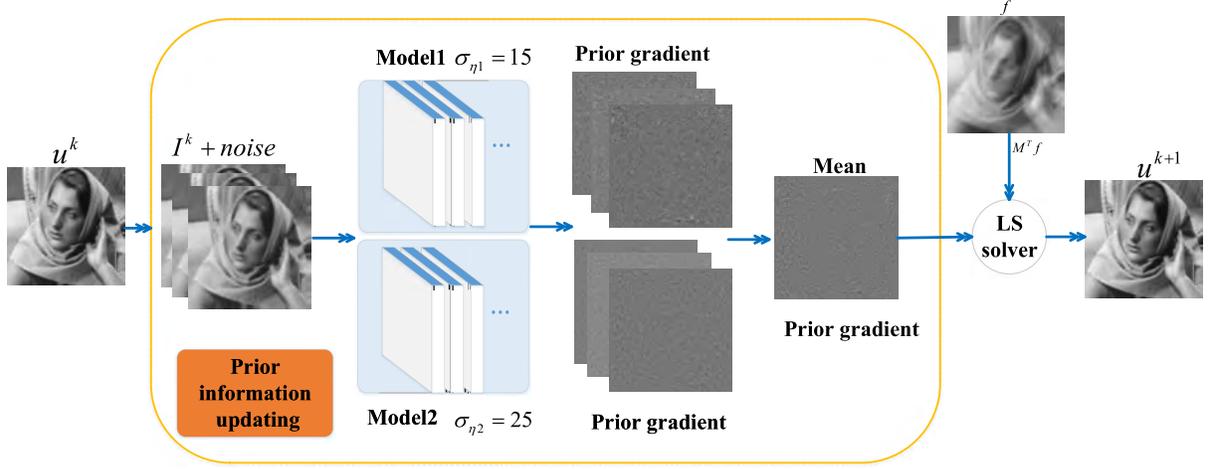


Fig. 6. Illustration of the multi-model MEDAEP with RGB three-channel network scheme at the iterative restoration phase. Prior gradient stands for the gradient of the prior information.

term as follows:

$$L_{MEDAE} = E_{\eta, I} \left[\frac{1}{N} \sum_{i=1}^N \left\| I - A_{\sigma_{\eta_i}}(I) \right\|^2 \right], \quad N = 2 \quad (9)$$

C. Iterative Solver for MEDAEP Model

Considering the multi-channel scheme in Section III. A and multi-model version in Section III. B, the general mathematical model for IR can be derived as follows:

$$\min_u \|Mu - f\|^2 + \frac{\lambda}{N} \sum_{i=1}^N \left\| I(u) - A_{\sigma_{\eta_i}}(I(u)) \right\|^2 \quad (10)$$

For the sake of convenience, the three-channel auxiliary variables $I(u) = [u, u, u]$ is simply termed as I . N stands for the number of EDAEP 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 [40] to tackle it. Subsequently, the model is approximated by a least square (LS) minimization:

$$\min_u \|Mu - f\|^2 + \frac{\lambda}{\beta N} \sum_{i=1}^N \left\| I - (I^k - \beta \nabla G_i(I^k)) \right\|^2 \quad (11)$$

where $G_i(I) = \left\| I - A_{\sigma_{\eta_i}}(I) \right\|^2$ and $\nabla G_i(I) = [1 - \nabla_I A_{\sigma_{\eta_i}}^T(I)][I - A_{\sigma_{\eta_i}}(I)]$.

The function $G(I)$ is $1/\beta$ -Lipschitz smooth, i.e., $\left\| \nabla G(I') - \nabla G(I'') \right\|_2 \leq \left\| I' - I'' \right\|_2 / \beta$. k denotes the index number of iterations. Here, we empirically set $\beta = 1$ and it works well in our experiments.

Given $\beta = 1$, Eq. (11) is a standard LS problem, which can be solved by calculating the gradient as follows:

$$M^T(Mu - f) + \lambda \left\{ I + \frac{1}{N} \sum_{i=1}^N [\nabla_I A_{\sigma_{\eta_i}}^T(I^k)(A_{\sigma_{\eta_i}}(I^k) - I^k) - A_{\sigma_{\eta_i}}(I^k)] \right\} = 0 \quad (12)$$

Algorithm 1 MEDAEP

Training Stage

Training images: Color image dataset $\{I | I = [I_r, I_g, I_b]\}$
Noisy levels: σ_{η_1} and σ_{η_2}
Network: Three-channel EDAE network
Outputs: Trained network $A_{\sigma_{\eta_1}}(\circ)$ and $A_{\sigma_{\eta_2}}(\circ)$

Testing Stage

Initialization: $u^0 = M^T f$; K ; $N = 2$

for $k = 1, 2, \dots, K$ **do**

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

Calculate the components in prior gradient:

$A_{\sigma_{\eta_i}}(I^k), \nabla_I A_{\sigma_{\eta_i}}^T(I^k)[A_{\sigma_{\eta_i}}(I^k) - I^k]; i = 1, 2, \dots, N$

Update the solution via solving LS problem:

$$u^{k+1} = \frac{M^T f + \frac{\lambda}{N} \sum_{i=1}^N R[\{A_{\sigma_{\eta_i}}(I^k) - \nabla_I A_{\sigma_{\eta_i}}^T(I^k)[A_{\sigma_{\eta_i}}(I^k) - I^k]\}}{(M^T M + \lambda)}$$

end for

and it yields:

$$u^{k+1} = \frac{M^T f + \frac{\lambda}{N} \sum_{i=1}^N R[\{A_{\sigma_{\eta_i}}(I^k) - \nabla_I A_{\sigma_{\eta_i}}^T(I^k)[A_{\sigma_{\eta_i}}(I^k) - I^k]\}}{(M^T M + \lambda)} \quad (13)$$

where R stands for the mean operator employed on the three-channel variable. It can be observed that the solution formulation contains $A_{\sigma_{\eta_i}}(I^k)$ and $\nabla_I A_{\sigma_{\eta_i}}^T(I^k)[A_{\sigma_{\eta_i}}(I^k) - I^k]$, where the parameters in $A_{\sigma_{\eta_i}}(\circ)$ are already learned at the network training stage. In particular, the $A_{\sigma_{\eta_i}}(I^k)$ is the forward output with network input $I^k + \sigma_{\eta_i}$. $\nabla_I A_{\sigma_{\eta_i}}^T(I^k)[A_{\sigma_{\eta_i}}(I^k) - I^k]$ is the network backward 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 $A_{\sigma_{\eta_i}}(I^k)$, $\nabla_I A_{\sigma_{\eta_i}}^T(I^k)$ and the LS solver until the value of solution u converges. In brief, the mathematical model is tackled by the proximal gradient and alternative optimization. In summary, the overall training phase and testing phase of MEDAEP algorithm are as follows:



Fig. 7. Testing grayscale images used in the experiments.

TABLE I
PSNR AND SSIM VALUES OF SIX TEST IMAGES DESTROYED BY VARIOUS BLUR SIZE WITH DIFFERENT LEVEL NOISE

Blur	Noisy		Barbara	Boats	Cameraman	Baboon	Peppers	Straw	Average
13	2.55	LevinSps	30.18/0.891	32.78/0.909	29.68/0.880	22.90/0.699	30.15/0.885	27.27/0.893	28.83/0.860
		EPLL	30.28/0.923	30.46/0.914	29.95/0.886	22.70/0.763	27.46/0.889	25.90/0.920	27.79/0.883
		DMSP	30.06/0.928	31.50/0.916	30.81/0.879	22.14/0.734	26.69/0.882	26.04/0.922	27.88/0.875
		DPE	32.25/0.916	33.20/0.902	31.48/0.861	24.57/0.804	32.05/0.876	29.02/0.938	30.43/0.883
		DAEP	29.42/0.827	30.41/0.825	28.05/0.728	23.46/0.731	29.40/0.789	27.12/0.898	27.98/0.800
		EDAEP	32.45/0.922	33.56/0.915	31.38/0.880	23.55/0.737	32.42/0.890	28.96/0.933	30.39/0.880
		MEDAEP	32.82/0.929	33.71/0.917	31.70/0.884	23.99/0.764	32.68/0.891	29.27/0.937	30.70/0.887
	7.65	LevinSps	23.51/0.491	24.04/0.451	23.46/0.310	20.67/0.492	23.56/0.414	22.40/0.732	22.94/0.482
		EPLL	25.58/0.726	28.07/0.830	26.90/0.812	20.81/0.553	26.16/0.830	23.25/0.786	25.13/0.756
		DMSP	26.84/0.821	28.65/0.829	27.92/0.797	20.69/0.580	25.45/0.806	23.62/0.830	25.53/0.777
		DPE	26.00/0.675	27.28/0.679	26.23/0.576	21.64/0.565	26.86/0.656	24.57/0.816	25.43/0.661
		DAEP	25.17/0.656	27.19/0.709	24.73/0.562	21.16/0.520	26.08/0.677	23.15/0.732	24.58/0.643
		EDAEP	26.76/0.779	30.09/0.855	27.96/0.826	21.22/0.526	29.65/0.840	24.99/0.818	26.78/0.774
		MEDAEP	27.53/0.816	30.32/0.862	28.18/0.831	21.43/0.557	29.67/0.843	25.18/0.833	27.05/0.790
19	2.55	LevinSps	30.14/0.885	31.34/0.884	29.55/0.861	23.11/0.701	30.84/0.863	26.02/0.860	28.50/0.842
		EPLL	28.21/0.904	28.83/0.890	29.58/0.867	22.57/0.754	26.22/0.867	24.50/0.884	26.65/0.861
		DMSP	28.78/0.897	27.83/0.884	30.22/0.854	22.39/0.749	24.59/0.848	24.43/0.883	26.37/0.853
		DPE	30.94/0.889	31.58/0.875	30.44/0.833	24.29/0.808	29.27/0.833	27.34/0.921	28.98/0.860
		DAEP	28.50/0.798	28.03/0.751	27.28/0.672	23.44/0.729	27.97/0.748	25.25/0.849	26.75/0.758
		EDAEP	31.83/0.910	32.46/0.898	30.76/0.860	23.71/0.745	31.73/0.874	27.94/0.920	29.74/0.868
		MEDAEP	31.88/0.912	32.74/0.903	31.11/0.865	24.42/0.786	31.92/0.877	28.20/0.925	30.05/0.878
	7.65	LevinSps	21.49/0.423	21.78/0.380	21.41/0.264	19.54/0.445	21.66/0.340	20.54/0.646	21.07/0.416
		EPLL	24.92/0.729	25.88/0.773	25.91/0.775	20.40/0.501	24.66/0.786	21.38/0.633	23.86/0.700
		DMSP	26.05/0.803	25.80/0.788	27.03/0.769	20.32/0.536	23.76/0.762	22.15/0.756	24.19/0.736
		DPE	25.14/0.625	25.55/0.596	25.23/0.521	21.03/0.529	25.20/0.579	23.10/0.753	24.21/0.601
		DAEP	24.87/0.627	24.68/0.599	23.60/0.506	20.75/0.483	24.53/0.639	21.55/0.615	23.33/0.578
		EDAEP	27.02/0.789	28.00/0.811	26.99/0.803	20.57/0.446	28.21/0.809	23.14/0.724	25.66/0.730
		MEDAEP	27.34/0.807	28.39/0.820	27.25/0.806	20.80/0.486	28.46/0.812	23.63/0.764	25.98/0.749
25	2.55	LevinSps	28.33/0.854	30.92/0.886	28.65/0.850	21.72/0.604	29.52/0.857	24.87/0.824	27.34/0.813
		EPLL	26.05/0.842	26.64/0.872	27.90/0.849	21.09/0.645	23.97/0.829	23.20/0.822	24.80/0.810
		DMSP	27.10/0.877	26.22/0.869	28.83/0.826	21.24/0.642	23.07/0.815	23.60/0.836	25.01/0.811
		DPE	28.95/0.884	29.30/0.872	28.47/0.829	22.38/0.729	26.35/0.797	25.00/0.883	26.74/0.832
		DAEP	26.93/0.779	27.63/0.768	26.48/0.665	21.58/0.625	26.38/0.731	24.38/0.828	25.56/0.733
		EDAEP	29.79/0.900	31.49/0.904	29.66/0.853	21.39/0.618	30.46/0.870	25.87/0.880	28.11/0.838
		MEDAEP	30.15/0.906	31.66/0.907	30.05/0.857	21.72/0.652	30.71/0.872	26.09/0.882	28.43/0.846
	7.65	LevinSps	21.24/0.393	21.58/0.373	21.53/0.260	19.39/0.412	21.45/0.339	19.98/0.589	20.86/0.394
		EPLL	23.46/0.670	25.04/0.771	25.28/0.780	19.81/0.441	23.16/0.765	21.50/0.652	23.04/0.680
		DMSP	25.05/0.777	25.02/0.798	26.47/0.779	19.95/0.480	22.69/0.760	21.87/0.727	23.51/0.720
		DPE	25.12/0.663	25.98/0.661	25.53/0.576	20.27/0.479	24.68/0.620	22.71/0.747	24.05/0.624
		DAEP	23.90/0.608	24.84/0.629	23.35/0.516	19.91/0.439	23.95/0.608	21.79/0.640	22.96/0.573
		EDAEP	25.61/0.767	28.04/0.825	26.87/0.804	19.95/0.438	27.30/0.808	22.99/0.731	25.13/0.729
		MEDAEP	25.91/0.775	28.11/0.829	27.15/0.809	20.02/0.458	27.33/0.811	23.23/0.754	25.29/0.739

IV. EXPERIMENTAL RESULTS

In this section, extensive experimental results are conducted to verify the performance of the proposed method MEDAEP for image deblurring and image deblocking. In the prior

learning stage, we follow ref. [41] to use 400 images of 180×180 and set the training patch size as 40×40 . As a result, 128×1600 patches are cropped to train the model. In the IR testing stage, we choose $N = 2$ and, $\sigma_{\eta_1} = 25$, $\sigma_{\eta_2} = 15$

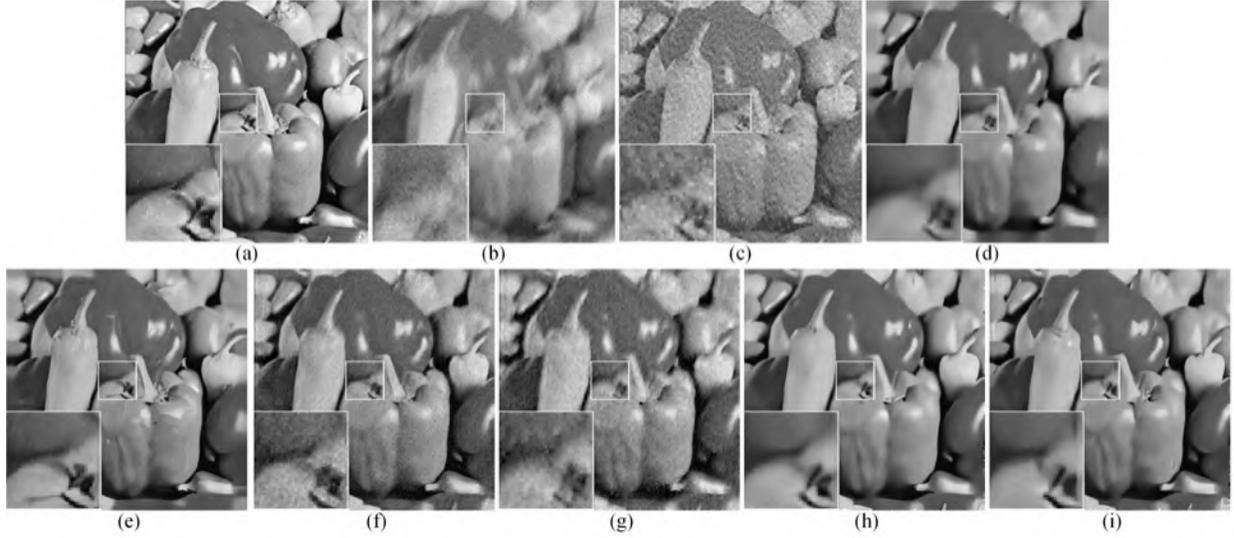


Fig. 8. Deblurring comparison on image “Peppers”. (a) Original image; (b) Noisy and blurred image (Gaussian kernel: 19×19 , $\delta_e = 7.65$); Results by (c) LevinSps (PSNR = 21.66dB; SSIM = 0.340), (d) EPLL (PSNR = 24.66dB; SSIM = 0.786), (e) DMSp (PSNR = 23.76dB; SSIM = 0.762), (f) DPE (PSNR = 25.20dB; SSIM = 0.579), (g) DAEP (PSNR = 24.53dB; SSIM = 0.639), (h) EDAEP (PSNR = 28.21dB; SSIM = 0.809), and (i) MEDAEP (PSNR = **28.46dB**; SSIM = **0.812**).

for image deblurring and $\sigma_{\eta 2} = 25$, $\sigma_{\eta 1} = 3$ for image deblocking. In the experiments, all the competing algorithms are implemented in MATLAB 2013 on a PC equipped with Intel(R) Core (TM) i7-7700 CPU @ 3.60GHz, 16G RAM and Windows 10 operating system. To evaluate the quality of the restoration image, in addition to PSNR (Peak Signal to Noise Ratio, dB), the powerful perceptual quality metric SSIM (Structural Similarity) are also calculated. The higher PSNR and SSIM values mean that retaining the more structures with better visual quality. All the experimental test images are given in Fig. 7. For the convenience of reproducible research, source code of MEDAEP is available at <https://github.com/yqx7150/MEDAEP>.

A. Image Deblurring

In the experiment of image deblurring, three blur kernels (i.e., 13×13 , 19×19 and 25×25) are selected from the dataset of Levin *et al.* [42] as the degrading operators. Additionally, two different levels of Gaussian noise with standard variations $\delta_e = 2.55$ and $\delta_e = 7.65$ are added into the intermediate blurred image to generate the final observation. Six grayscale images are tested to verify the performance of the proposed MEDAEP method with comparison to Levin *et al.* [42], EPLL framework [6], DMSp [43], DPE [44] and DAEP [29] with $\sigma_{\eta} = 25$. The regularization parameters in DAEP, EDAEP and MEDAEP are set as $\lambda = 3.875/\sigma_{\eta}^2$ and $\lambda = 3.875/\sigma_{\eta 1}^2$ respectively.

Visual quality comparison of image deblurring at varying blur kernel size and Gaussian noise level for grayscale images “Peppers”, “Barbara” and “Boats” with size 256×256 as shown in Figs. 8-10. It can be observed that LevinSps deblurs image better but retains a lot of noise and the DAEP method can remove noise better 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 recent DPE, DMSp and proposed EDAEP, MEDAEP can outperform the LevinSps, DAEP and EPLL largely. Particularly, DMSp, EDAEP and MEDAEP methods not only remove noise but also preserve the structure details. Moreover, the proposed MEDAEP produces cleaner and sharper image edges and textures than other competing methods.

The PSNR and SSIM values of six test images by various blur kernels with different level noise are shown in Table I. As can be seen, EPLL and DAEP produce very similar results and the proposed EDAEP method achieves highly competitive performance compared with other leading deblurring methods at various noise variances and blur kernel size. Although the PSNR/SSIM values of DPE and DMSp method gain over the proposed EDAEP in some case, the advanced MEDAEP method improves the result of image deblurring by aggregating the EDAEP, and it produces slightly higher average PSNR and SSIM than EDAEP.

B. Image Deblocking

In this subsection, we introduce the proposed MEDAEP to the JPEG-coded IR. According to ref. [45], the observed image can be modelled as the corrupted one by the quantization noise (i.e., $f = u + e$). Specifically, we adopt Gaussian model to characterize the noise e . Its variance δ_e^2 is adaptively adjusted depending on image content and quantization, and can be estimated as follows:

$$\delta_e^2 = 1.196 * (s)^{0.6394} + 0.9693, \quad s = \frac{1}{9} \sum_{i,j=1}^3 M_{[i,j]}^q \quad (14)$$

where M^q is the 8×8 quantization matrix with the quality factor (QF), and s is the mean value of the nine upper-left entries in M^q . With the Gaussian quantization noise model,



Fig. 9. Deblurring comparison on image “Barbara”. (a) Original image; (b) Noisy and blurred image (Gaussian kernel: 19×19 , $\delta_e = 2.25$); Results by (c) LevinSps (PSNR = 30.14dB; SSIM = 0.885), (d) EPLL (PSNR = 28.21dB; SSIM = 0.904), (e) DMSP (PSNR = 28.78dB; SSIM = 0.897), (f) DPE (PSNR = 30.94dB; SSIM = 0.889), (g) DAEP (PSNR = 28.50dB; SSIM = 0.798), (h) EDAEP (PSNR = 31.83dB; SSIM = 0.910), and (i) MEDAEP (PSNR = **31.88dB**; SSIM = **0.912**).

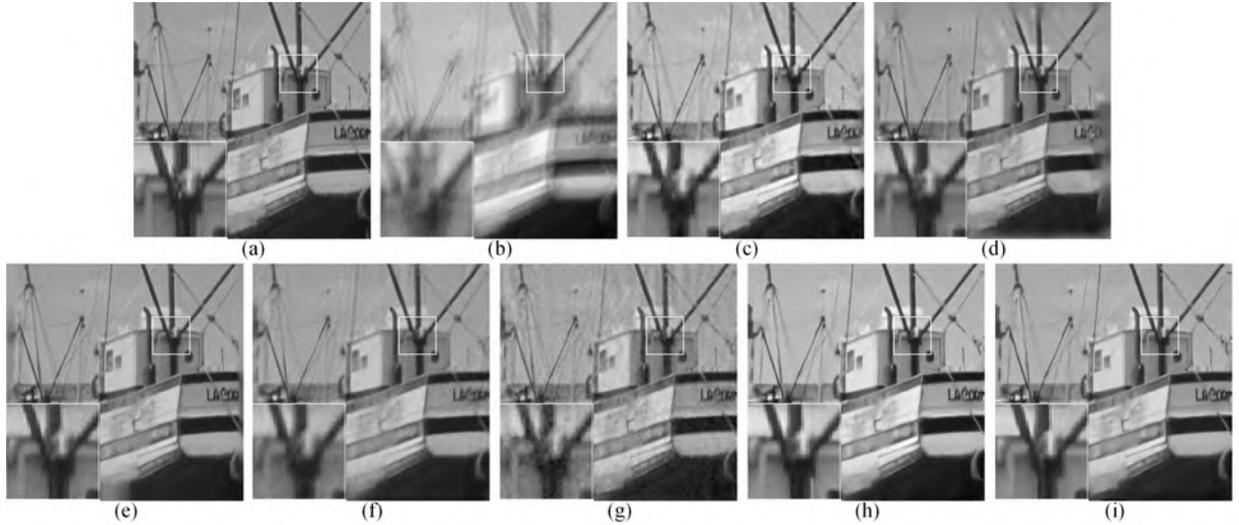


Fig. 10. Comparison on image “Boats”. (a) Original image; (b) Noisy and blurred image (Gaussian kernel: 25×25 , $\delta_e = 2.25$); (c) Results by LevinSps (PSNR = 30.92dB; SSIM = 0.886), (d) EPLL (PSNR = 26.64dB; SSIM = 0.869), (e) DMSP (PSNR = 26.22dB; SSIM = 0.869), (f) DPE (PSNR = 29.30dB; SSIM = 0.872), (g) DAEP (PSNR = 27.63dB; SSIM = 0.768), (h) EDAEP (PSNR = 31.49dB; SSIM = 0.904), and (i) MEDAEP (PSNR = **31.66dB**; SSIM = **0.907**).

the data-fidelity term can be formulated as $\|u - f\|_{2\delta_e^2}^2$. Hence, the image deblurring can be modeled as:

$$\min_u \frac{1}{2\delta_e^2} \|u - f\|^2 + \frac{\alpha}{2\delta_s^2} \left\| I - \frac{1}{N} \sum_{i=1}^N A_{\delta_{\eta_i}}(I) \right\|^2 \quad (15)$$

Particularly, it can be observed that we use adaptive parameter scheme as in ref. [12] (i.e. $\lambda = \alpha/2\delta_s^2$), and $\delta_s^2 = \delta\sqrt{\delta_e^2 - \|u - f\|^2}$ is adaptively adjusted depending on image content and quantization [16]. The parameter δ is a scaling factor that can control the variance estimation and we empirically set δ to be 0.2.

In the experiment for image deblurring, all test images are encoded by a JPEG coder [46] with QF = 10. The proposed MEDAEP is compared with six representative image deblurring methods, i.e. Sun’s [8], Foi’s [46], DicTV [9], Zhang’s [10], Ren’s [11], CONCOLOR [12], DnCNN-3 [19] and two representative image denoising methods, i.e. BM3D [15] and WNNM [16].

The PSNR and SSIM values for all grayscale test images in the case of QF = 10 are provided in Table II, where the best results are highlighted in bold. As shown, the proposed MEDAEP achieves the highest PSNR and SSIM values among almost test images. Besides, one can observe that DnCNN-3 and CONCOLOR can achieve a pleasing result

TABLE II
PSNR/SSIM VALUES WITH VARIOUS IMAGE DEBLOCKING METHODS AT SIX IMAGES

	Forem	House	Leaves	Lena	Parrots	Peppers	Average
size	288×352	256×256	256×256	512×512	256×256	512×512	
JPEG	29.11/0.808	30.56/0.818	25.40/0.861	30.41/0.818	28.96/0.834	30.14/0.784	29.10/0.821
Sun	30.31/0.860	32.00/0.850	26.60/0.914	31.72/0.859	30.04/0.878	31.62/0.832	30.38/0.866
Foi	30.94/0.865	32.09/0.849	27.30/0.928	31.84/0.859	30.15/0.878	31.69/0.832	30.67/0.869
Zhang	30.78/0.865	32.11/0.851	27.26/0.921	31.92/0.860	30.50/0.880	31.68/0.831	30.71/0.868
Ren	30.94/0.866	32.41/0.853	27.59/0.931	31.92/0.857	30.34/0.878	31.63/0.830	30.81/0.869
BM3D	30.72/0.862	32.07/0.849	26.98/0.921	31.77/0.855	30.05/0.875	31.42/0.825	30.50/0.865
DicTV	30.30/0.857	31.77/0.848	26.33/0.903	31.55/0.856	29.82/0.874	31.29/0.824	30.18/0.860
WNNM	30.94/0.866	32.42/0.853	27.66/0.933	31.93/0.857	30.33/0.878	31.64/0.830	30.82/0.870
CONCOLOR	31.46/0.875	33.04/0.861	28.20/0.941	32.19/0.866	30.66/0.884	31.94/0.836	31.25/0.877
DnCNN	31.26/0.873	32.87/0.858	29.15/0.946	32.34/0.866	31.26/0.887	32.17/0.838	31.51/0.878
DAEP	29.28/0.804	30.68/0.808	25.66/0.870	30.65/0.812	29.19/0.828	30.34/0.781	29.30/0.817
EDAEP	31.25/0.870	32.78/0.857	28.53/0.943	32.13/0.866	30.88/0.886	31.95/0.835	31.25/0.876
MEDAEP	31.50/0.876	32.90/0.861	28.94/0.951	32.34/0.872	31.10/0.895	32.22/0.848	31.50/0.884

TABLE III
AVERAGE PSNR (dB)/SSIM/PSNR-B (dB) SCORES OF DIFFERENT SOFT DECODING ALGORITHMS ON CLASSIC5

	10	20	30	40
JPEG	27.82/0.760/25.21	30.12/0.834/27.50	31.48/0.867/28.94	32.43/0.885/29.92
CONCOLOR	29.24/0.796/29.14	31.38/0.854/31.18	32.70/0.881/32.50	33.60/0.896/ 33.36
D2SD	29.21/0.796/28.87	31.47/0.855/31.15	32.79/0.881/32.40	33.66/0.896/33.20
ARCNN	29.04/0.811/28.75	31.16/0.852/30.60	32.52/0.881/32.00	33.33/0.895/32.81
TNRD	29.28/0.799/29.04	31.47/0.858/31.05	32.78/0.884/32.24	–
DnCNN-3	29.40/0.803/29.13	31.63/0.861/31.19	32.90/0.886/32.36	33.77/0.900/33.20
DPW-SDNet	29.74/0.812/29.37	31.95/0.866/31.42	33.22/0.890/32.51	34.07/0.904/33.24
MEDAEP	29.23/ 0.822 /28.96	31.49/ 0.876 /31.01	32.81/ 0.899 /32.37	33.71/ 0.913 /33.21

TABLE IV
AVERAGE PSNR (dB)/SSIM/PSNR-B (dB) SCORES OF DIFFERENT SOFT DECODING ALGORITHMS ON CLASSIC5

	10	20	30	40
JPEG	27.77/0.773/25.34	30.08/0.851/27.57	31.41/0.885/28.93	32.36/0.904/29.96
CONCOLOR	28.87/0.802/28.76	31.08/0.868/30.90	32.42/0.899/32.16	33.39/0.916/33.07
D2SD	28.83/0.802/25.34	31.08/0.869/30.80	32.41/0.899/32.10	33.37/0.916/33.06
ARCNN	29.04/0.808/28.77	31.31/0.873/30.79	32.73/0.904/32.22	33.63/0.920/33.14
TNRD	29.14/0.811/28.88	31.46/0.877/31.04	32.84/0.906/32.28	–
DnCNN-3	29.19/0.812/28.91	31.59/0.880/31.08	32.99/0.909/32.35	33.96/0.925/33.29
DPW-SDNet	29.53/0.821/29.13	31.90/0.885/31.27	33.31/0.913/32.52	34.30/0.928/33.44
MEDAEP	29.06/ 0.828 /28.78	31.41/ 0.892 /31.11	32.84/ 0.918 /32.43	33.78/ 0.932 /33.39

well. To further illustrate the difference among DnCNN-3, CONCOLOR method and proposed method, the deblocked images “Leaves” and “Parrots” obtained by proposed method, DnCNN-3 and the algorithm CONCOLOR are shown in Figs. 11-12. It can be observed that the blocking artifacts are obvious in the images that decoded directly by the standard JPEG. In particular, CONCOLOR produces results with good visual quality and preserves image details well, while DnCNN-3 can well reconstruct the piecewise smooth regions but often fails to recover fine image details. Furthermore, MEDAEP not only reduces most of the blocking artifacts significantly, but also provides better recovery on both edges and textures than other competing methods.

In order to further compare with state-of-the-art CNN-based methods, the MEDAEP is compared with six deblocking methods for JPEG-compressed images at the widely tested Classic5 and LIVE1 dataset, including two restoration-based

approaches (i.e., CONCOLOR [12] and D2SD [49]) and four deep learning-based algorithms (i.e., ARCNN [50], TNRD [41], DnCNN-3 [19], and DPW-SDNet [51]). We compare the performance of all algorithms in the cases of $QF = 10, 20, 30,$ and 40 . Table III and IV report the objective assessment scores achieved by all tested algorithms, including the PSNR, SSIM, and PSNR-B [52]. Note that the PSNR-B is a specifically developed assessment metric for blocky and deblocked images. It can be observed from Table III and IV that the DPW-SDNet consistently outperforms the six competitors with higher PSNR and PSNR-B improvements. While the proposed method preserves the detail structures with higher SSIM values. It is worth noting that the deep learning-based algorithms can achieve a better PSNR value and the difference values between PSNR and PSNR-B are bigger than the restoration-based approaches. The proposed method MEDAEP is an approach that combine the benefit of deep learning

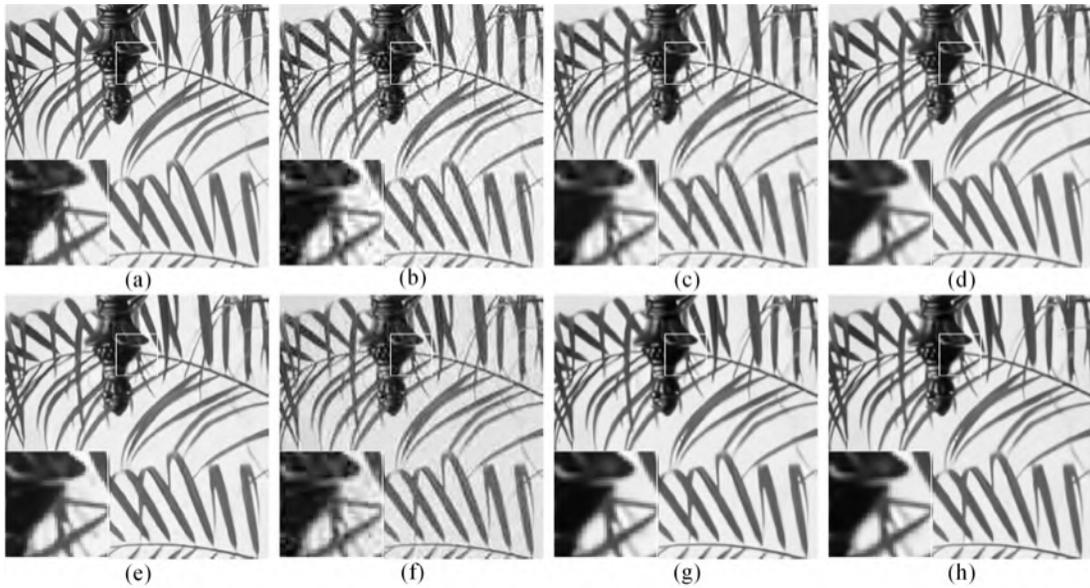


Fig. 11. Visual comparison of image deblocking for “Leaves” in the case of $QF = 10$. (a) Original image, (b) JPEG compressed image (PSNR = 25.40dB; SSIM = 0.861), (c) results by Sun’s (PSNR = 26.60dB; SSIM = 0.914), (d) CONCOLOR (PSNR = 28.20dB; SSIM = 0.941), (e) DnCNN-3 (PSNR = **29.15dB**; SSIM = 0.946), (f) DAEP (PSNR = 25.66dB; SSIM = 0.870), (g) EDAEP (PSNR = 28.53dB; SSIM = 0.943), and (h) MEDAEP (PSNR = 28.94dB; SSIM = **0.947**).

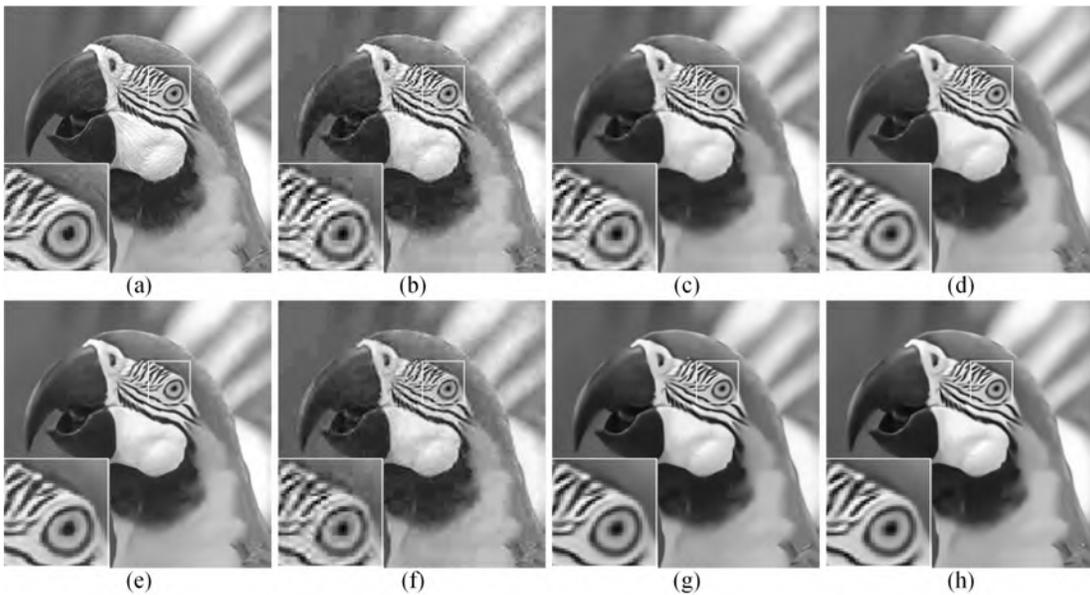


Fig. 12. Visual comparison of image deblocking for “Parrots” in the case of $QF = 10$. (a) Original image, (b) JPEG compressed image (PSNR = 28.96dB; SSIM = 0.834), (c) results of Sun’s (PSNR = 30.04dB; SSIM = 0.878), (d) CONCOLOR (PSNR = 30.66dB; SSIM = 0.884), (e) DnCNN-3 (PSNR = **31.26dB**; SSIM = 0.887), (f) DAEP (PSNR = 29.19dB; SSIM = 0.828), (g) EDAEP (PSNR = 30.88dB; SSIM = 0.886), and (h) MEDAEP (PSNR = 31.10dB; SSIM = **0.890**).

algorithm and restoration-based approaches. Figs. 13-14 illustrate the deblocked images “Womanhat” and “Lena” obtained by proposed method MEDAEP, CONCOLOR, ARCNN, and DnCNN-3. One can observe that the proposed MEDAEP can preserve the detailed structure well.

C. Variants of Network

In this subsection, some experimental investigations of the network parameter evaluations are conducted.

In the first test, we investigate the sensitivity of high-dimensional structural prior in the three-channel learned

TABLE V
PERFORMANCE OF MEDAEP WITH SOME SINGLE CHANNEL HAS BEEN EXCHANGED (r PERCENT)

r	0%	10%	30%
PSNR/SSIM	30.23/0.874	30.03/0.870	29.83/0.867

MEDAEP network. In the experiment, we deliberately exchange r percent training samples with only one changed channel (randomly choose R, G, or B). Therefore, the similarity in 3D space degrades and only remains in 2D space.



Fig. 13. Visual comparison of image deblocking for “Womanhat” in the case of QF = 10. (a) Original image, (b) JPEG compressed image (30.49, 28.18, 0.772), (c) CONCOLOR (**31.89**, 0.815, **31.89**), (d) ARCNN (31.71, 0.810, 31.59), (e) DnCNN-3 (31.49, 0.803, 31.49), and (f) MEDAEP (31.72, **0.817**, 31.67).

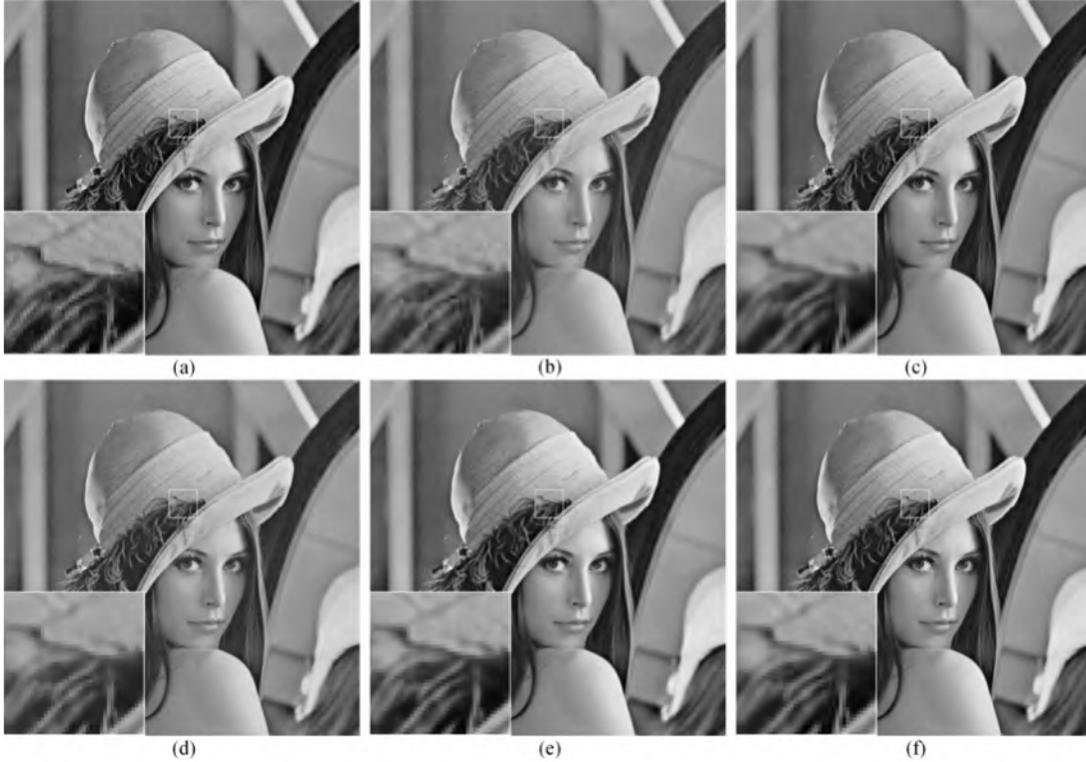


Fig. 14. Comparison of image deblocking for “Lena” in the case of QF = 40. (a) Original image, (b) JPEG compressed image (35.13, 0.921, 31.89), (c) CONCOLOR (36.21, 0.929, 35.90), (d) ARCNN (36.13, 0.930, 35.39), (e) DnCNN-3 (36.15, 0.928, **35.96**), and (f) MEDAEP (**36.35**, **0.931**, 35.65).

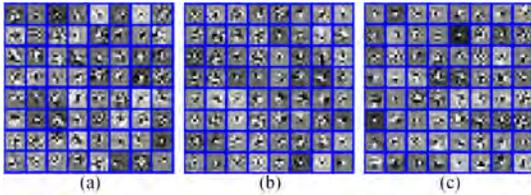


Fig. 15. Representative convolutional kernels of the final convolutional layer within the EDAEP with percent training samples have been exchanged. (a) $r = 0$, (b) $r = 10\%$, (c) $r = 30\%$. $\sigma_\eta = 25$.

As depicted in Fig. 15, when the r -value increases from 0% to 30%, the learned filters become flatter. The deblocking results with these three learned networks are listed in Table V, where the deblocking performance degrades as the r -value increases.

These phenomena reveal that the high-dimensional structural prior is very important for IR of lower-dimensional images.

In the second test, we examine the network sensitivity to filter number and filter size. The default filter number value of the DAEP, EDAEP and MEDAEP is set to be 64, and the filter size is set to be 32×32 . We vary one parameter at a time while keeping the others fixed at their nominal values. Besides, the network in EDAEP model is trained by 100000 iterations and the image deblurring process is stopped until 150 iterations. The PSNR and SSIM values for image deblurring with the blur size 13×13 and Gaussian noise level $\delta_e = 2.55$ on image “Barbara” are shown in Table VI and VII. One can observe a phenomenon that the performance improves if the filter size or number increase. As the filter number increases from 32 to 128 or the filter sizes increases from

TABLE VI
DEBLURRING PERFORMANCE WITH VARIOUS FILTER NUMBERS

Filter number	32	64	128
PSNR/SSIM	30.69/0.891	31.19/0.902	31.67/0.911
Time (s)	10.71/150	19.845/150	51.015/150

TABLE VII
DEBLURRING PERFORMANCE WITH VARIOUS FILTER SIZES

Filter size	3×3	5×5	7×7
PSNR/SSIM	31.19/0.902	31.65/0.910	31.76/0.912
Time (s)	19.845/150	54.36/150	102.315/150

TABLE VIII
PERFORMANCE OF MEDAEP FOR DEBLURRING “BOATS”
IMAGE WITH DIFFERENT NOISE MODELS

σ_η	PSNR/SSIM	σ_η	PSNR/SSIM
3,8,15	32.66/0.916	3,25	31.99/0.873
3,8,25	31.53/0.861	8,11	32.16/0.905
3,11,15	32.65/0.914	8,15	32.69/0.912
3,11,25	32.03/0.880	8,25	32.60/0.906
3,8,11	32.61/0.913	11,15	32.39/0.905
3,15,25	32.22/0.887	11,25	32.73/0.902
8,11,25	32.37/0.895	15,25	32.73/0.903
8,15,25	32.59/0.900	25	32.53/0.897
11,15,25	32.69/0.903	15	32.11/0.900
8,11,15	32.76/0.913	11	31.47/0.894
3,8	32.41/0.910	8	31.88/0.889
3,11	32.41/0.911	3	32.46/0.868
3,15	32.69/0.915	–	–

TABLE IX
PSNR PERFORMANCE OF MEDAEP IN IMAGE DEBLOCKING
WITH TWO DIFFERENT NOISE MODELS

$(\sigma_{\eta 1}, \sigma_{\eta 2})$	(3, 8)	(3, 11)	(3, 15)	(3, 25)	(8, 11)
Butterfly	25.69	25.67	25.68	28.23	27.84
Parrots	28.38	30.51	30.56	30.96	30.54
$(\sigma_{\eta 1}, \sigma_{\eta 2})$	(8, 15)	(8, 25)	(11, 15)	(11, 25)	(15, 25)
Butterfly	28.28	27.57	28.28	28.16	28.26
Parrots	30.57	30.85	30.55	30.89	30.55

3×3 to 7×7 , the PSNR and SSIM values become better while at the cost of running time. Thus, superior performance will be achieved if the current dataset supports a good training.

Finally, the performance of MEDAEP with regard to N -value in Eq. (10) is discussed. The deburred PSNR/SSIM value of image “Boats” under the blur size 19×19 and noise level is $\delta_e = 2.55$ recorded in Table VIII. As can be observed, when N -value increases, the PSNR/SSIM value increases accordingly. On the other hand, the computational cost also becomes larger. By considering the tradeoff between algorithm performance and complexity, $N = 2$ is selected in all experiments. Furthermore, we investigate the two-noise model rule for MEDAEP by varying the values of standard derivation $(\sigma_{\eta 1}, \sigma_{\eta 2})$ for images that encoded by a JPEG coder with QF = 10 in Table IX. As depicted, better PSNR value at the point (3, 25) is achieved both on images “Butterfly” and

“Parrots”. Thus, $(\sigma_{\eta 1}, \sigma_{\eta 2}) = (3, 25)$ is chosen for image deblocking.

V. CONCLUSION AND FUTURE WORKS

This work paved a new way to incorporate higher-dimensional prior information into the lower-dimensional recovery procedure. Specifically, a novel three-channel denoising autoencoder prior was presented for grayscale IR applications, which built on the assumption that an optimal denoising autoencoder is a local mean of the true data density. We further enhanced the network stability by adopting the aggregation function. In particular, auxiliary variable technique was applied to impose higher-dimensional structural information learned by color RGB-channels network. The formulated mathematic model was tackled by proximal gradient and alternative optimization. Both qualitative and quantitative experimental results on image deblurring and deblocking demonstrated that the proposed MEDAEP achieved significant performance improvements over many current state-of-the-art methods.

In the forthcoming future, it is precise to make the weighted aggregation function adaptively map the different multi-model based DAEP. For instance, we can explore some priori information as criteria to adaptively weight the multi-model such as to obtain desired IR effect. Besides, some researchers have already applied sparse representation or low-rank regularization in multi-filters domain [47, 48]. Therefore, it is natural to extend our MEDAEP methodology in these scenarios.

ACKNOWLEDGMENT

The authors sincerely thank the anonymous reviewers for their valuable comments and constructive suggestions that are very helpful in the improvement of this paper. The authors also would like to thank D. Zoran, A. Levin, D. Sun, G. K. Wallace, H. Chang, X. Zhang, J. Ren, J. Zhang, K. Dabov, S. Gu, R. Liu, S. A. Bigdeli, K. Zhang, X. Liu, C. Dong, Y. Chen and H. Chen for their source codes for comparisons.

REFERENCES

- [1] S. Kindermann, S. Osher, and P. W. Jones, “Deblurring and denoising of images by nonlocal functionals,” *Multiscale Model. Simul.*, vol. 4, no. 4, pp. 1091–1115, Jul. 2005.
- [2] J. A. Dobrosotskaya and A. L. Bertozzi, “A wavelet-laplace variational technique for image deconvolution and inpainting,” *IEEE Trans. Image Process.*, vol. 17, no. 5, pp. 657–663, May 2008.
- [3] A. Beck and M. Teboulle, “Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems,” *IEEE Trans. Image Process.*, vol. 18, no. 11, pp. 2419–2434, Nov. 2009.
- [4] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, “Non-local sparse models for image restoration,” in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Sep./Oct. 2009, pp. 2272–2279.
- [5] W. Dong, L. Zhang, G. Shi, and X. Wu, “Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization,” *IEEE Trans. Image Process.*, vol. 20, no. 7, pp. 1838–1857, Jul. 2011.
- [6] D. Zoran and Y. Weiss, “From learning models of natural image patches to whole image restoration,” in *Proc. IEEE Int. Conf. Comput. Vis.*, Nov. 2011, pp. 479–486.
- [7] W. Dong, L. Zhang, G. Shi, and X. Li, “Nonlocally centralized sparse representation for image restoration,” *IEEE Trans. Image Process.*, vol. 22, no. 4, pp. 1620–1630, Apr. 2013.
- [8] D. Sun and W.-K. Cham, “Postprocessing of low bit-rate block DCT coded images based on a fields of experts prior,” *IEEE Trans. Image Process.*, vol. 16, no. 11, pp. 2743–2751, Nov. 2007.

- [9] H. Chang, M. K. Ng, and T. Zeng, "Reducing artifacts in JPEG decomposition via a learned dictionary," *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 718–728, Feb. 2014.
- [10] X. Zhang, R. Xiong, X. Fan, S. Ma, and W. Gao, "Compression artifact reduction by overlapped-block transform coefficient estimation with block similarity," *IEEE Trans. Image Process.*, vol. 22, no. 12, pp. 4613–4626, Dec. 2013.
- [11] J. Ren, J. Liu, M. Li, W. Bai, and Z. Guo, "Image blocking artifacts reduction via patch clustering and low-rank minimization," in *Proc. IEEE Data Compress. Conf. (DCC)*, Mar. 2013, p. 516.
- [12] J. Zhang, R. Xiong, C. Zhao, Y. Zhang, S. Ma, and W. Gao, "CON-COLOR: Constrained non-convex low-rank model for image deblocking," *IEEE Trans. Image Process.*, vol. 25, no. 3, pp. 1246–1259, Mar. 2016.
- [13] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
- [14] Z. Xu and J. Sun, "Image inpainting by patch propagation using patch sparsity," *IEEE Trans. Image Process.*, vol. 19, no. 5, pp. 1153–1165, May 2010.
- [15] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.
- [16] S. Gu, L. Zhang, W. Zuo, and X. Feng, "Weighted nuclear norm minimization with application to image denoising," in *Proc. IEEE Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 2862–2869.
- [17] W. Zhao, Y. Lv, Q. Liu, and B. Qin, "Detail-preserving image denoising via adaptive clustering and progressive PCA thresholding," *IEEE Access*, vol. 6, pp. 6303–6315, 2018.
- [18] U. Schmidt and S. Roth, "Shrinkage fields for effective image restoration," in *Proc. IEEE Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 2774–2781.
- [19] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian Denoiser: Residual learning of deep CNN for image denoising," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142–3155, Jul. 2017.
- [20] S. P. Awate and R. T. Whitaker, "Higher-order image statistics for unsupervised, information-theoretic, adaptive, image filtering," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2, Jun. 2005, pp. 44–51.
- [21] S. P. Awate and R. T. Whitaker, "Unsupervised, information-theoretic, adaptive image filtering for image restoration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 3, pp. 364–376, Mar. 2006.
- [22] P. Milanfar, "A tour of modern image filtering: New insights and methods, both practical and theoretical," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 106–128, Jan. 2013.
- [23] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Learn. Represent.*, 2015, pp. 1–14.
- [24] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
- [25] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 448–456.
- [26] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.
- [27] K. Zhang, W. Zuo, and L. Zhang, "FFDNet: Toward a fast and flexible solution for CNN-based image denoising," *IEEE Trans. Image Process.*, vol. 27, no. 9, pp. 4608–4622, Sep. 2018.
- [28] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, "Extracting and composing robust features with denoising autoencoders," in *Proc. Int. Conf. Mach. Learn.*, 2008, pp. 1096–1103.
- [29] S. A. Bigdeli and M. Zwicker, "Image restoration using autoencoding priors," Mar. 2017, *arXiv:1703.09964*. [Online]. Available: <https://arxiv.org/abs/1703.09964>
- [30] G. Alain and Y. Bengio, "What regularized auto-encoders learn from the data-generating distribution," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 3743–3773, 2014.
- [31] X. Zhang, M. Burger, X. Bresson, and S. Osher, "Bregmanized non-local regularization for deconvolution and sparse reconstruction," *SIAM J. Imag. Sci.*, vol. 3, no. 3, pp. 253–276, 2010.
- [32] K. N. Chaudhury and A. Singer, "Non-local Euclidean medians," *IEEE Signal Process. Lett.*, vol. 19, no. 11, pp. 745–748, Nov. 2012.
- [33] N. Joshi, C. L. Zitnick, R. Szeliski, and D. J. Kriegman, "Image deblurring and denoising using color priors," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 1550–1557.
- [34] H. S. Mousavi and V. Monga, "Sparsity-based color image super resolution via exploiting cross channel constraints," *IEEE Trans. Image Process.*, vol. 26, no. 11, pp. 5094–5106, Nov. 2017.
- [35] S. Dai, M. Han, W. Xu, Y. Wu, Y. Gong, and A. K. Katsaggelos, "SoftCuts: A soft edge smoothness prior for color image super-resolution," *IEEE Trans. Image Process.*, vol. 18, no. 5, pp. 969–981, May 2009.
- [36] Q. Liu, P. X. Liu, W. Xie, Y. Wang, and D. Liang, "GcsDecolor: Gradient correlation similarity for efficient contrast preserving decolorization," *IEEE Trans. Image Process.*, vol. 24, no. 9, pp. 2889–2904, Sep. 2015.
- [37] Q. Liu and H. Leung, "Variable augmented neural network for decolorization and multi-exposure fusion," *Inf. Fusion*, vol. 46, pp. 114–127, Mar. 2018.
- [38] J. Salmon, "On two parameters for denoising with non-local means," *IEEE Signal Process. Lett.*, vol. 17, no. 3, pp. 269–272, Mar. 2010.
- [39] D. Zhang, S.-H. Lee, H. Kim, J.-W. Han, and S.-J. Ko, "An effective aggregation function for image denoising based on low rank matrix completion," in *Proc. 4th Int. Symp. Knowl. Acquisition Modeling*, 2011, pp. 219–221.
- [40] N. Parikh and S. Boyd, "Proximal algorithms," *Found. Trends Optim.*, vol. 1, no. 3, pp. 127–239, 2014.
- [41] Y. Chen and T. Pock, "Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1256–1272, Jun. 2017.
- [42] A. Levin, R. Fergus, F. Durand, and W. T. Freeman, "Image and depth from a conventional camera with a coded aperture," *ACM Trans. Graph.*, vol. 26, no. 3, p. 70, 2007.
- [43] S. A. Bigdeli, M. Zwicker, P. Favaro, and M. Jin, "Deep mean-shift priors for image restoration," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 1–10.
- [44] R. Liu, L. Ma, Y. Wang, and L. Zhang, "Learning converged propagations with deep prior ensemble for image enhancement," *IEEE Trans. Image Process.*, vol. 28, no. 3, pp. 1528–1543, Mar. 2019.
- [45] A. Foi, V. Katkovnik, and K. Egiazarian, "Pointwise shape-adaptive DCT for high-quality denoising and deblocking of grayscale and color images," *IEEE Trans. Image Process.*, vol. 16, no. 5, pp. 1395–1411, May 2007.
- [46] G. K. Wallace, "The JPEG still picture compression standard," *Commun. ACM*, vol. 34, no. 4, pp. 30–44, 1991.
- [47] B. Xiong, Q. Liu, J. Xiong, S. Li, S. Wang, and D. Liang, "Field-of-Experts filters guided tensor completion," *IEEE Trans. Multimedia*, vol. 20, no. 9, pp. 2316–2329, Sep. 2018.
- [48] H. Lu, S. Li, Q. Liu, and M. Zhang, "MF-LRTC: Multi-filters guided low-rank tensor coding for image restoration," *Neurocomputing*, vol. 303, pp. 88–102, Aug. 2018.
- [49] X. Liu, X. Wu, J. Zhou, and D. Zhao, "Data-driven soft decoding of compressed images in dual transform-pixel domain," *IEEE Trans. Image Process.*, vol. 25, no. 4, pp. 1649–1659, Apr. 2016.
- [50] C. Dong, Y. Deng, C. C. Loy, and X. Tang, "Compression artifacts reduction by a deep convolutional network," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 576–584.
- [51] H. Chen, X. He, L. Qing, S. Xiong, and T. Q. Nguyen, "DPW-SDNet: Dual pixel-wavelet domain deep CNNs for soft decoding of JPEG-compressed images," in *Proc. IEEE Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 711–720.
- [52] C. Yim and A. C. Bovik, "Quality assessment of deblocked images," *IEEE Trans. Image Process.*, vol. 20, no. 1, pp. 88–98, Jan. 2011.



Sanqian Li received the master's degree from the School of Information Engineering, Nanchang University. Her current research interests include convolutional sparse coding, low-rank tensor, and deep learning, especially the applications in image processing, computer vision, and MRI.



Binjie Qin (M'07) received the M.Sc. degree in measuring and testing technologies and instruments from the Nanjing University of Science and Technology, Nanjing, and the Ph.D. degree in biomedical engineering from Shanghai Jiao Tong University, Shanghai, China, in 1999 and 2002, respectively. He was a Lecturer and an Associate Professor with the Department of Biomedical Engineering, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, China. From August 2012 to August 2013, he was a Visiting

Professor with the Department of Computer Science, University College London, U.K. He is currently an Associate Professor with the School of Biomedical Engineering, Shanghai Jiao Tong University. His current research interests include biomedical imaging, image processing, machine learning, computer vision, and biomedical instrumentation.



Jing Xiao (M'15) received the B.S. and M.S. degrees from Wuhan University in 2006 and 2008, respectively, and the Ph.D. degree from the Institute of Geo-Information Science and Earth Observation, Twente University, The Netherlands, in 2013. She is currently an Associate Professor with the National Engineering Research Center for Multimedia Software, School of Computer Science, Wuhan University. She is also a Project Researcher with the National Informatics Institute, Japan. Her research interests include image/video processing and compression and analysis.



Qiegen Liu (M'16) received the B.S. degree in applied mathematics from Gannan Normal University, the M.S. degree in computation mathematics, and the Ph.D. degree in biomedical engineering from Shanghai Jiao Tong University (SJTU). Since 2012, he has been with the School of Information Engineering, Nanchang University, Nanchang, China, where he is currently an Associate Professor. From 2015 to 2017, he held a postdoctoral position with UIUC and University of Calgary. His current research interests include sparse representations,

deep learning and their applications in image processing, computer vision, and MRI reconstruction.



Yuhao Wang (SM'14) received the Ph.D. degree in space physics from Wuhan University, Wuhan, China, in 2006. He was a Visiting Professor with the Department of Electrical Communication Engineering, University of Calgary, Calgary, AB, Canada, in 2008, and the China National Mobile Communication Research Laboratory, Southeast University, Nanjing, China, from 2010 to 2011. He is currently a Professor with the Cognition Sensor Network Laboratory, School of Information Engineering, Nanchang University, Nanchang, China. His current

research interests include radio measurement and channel modeling, software-defined radio, complex environment sensing and monitoring, nonlinear signal processing, multimedia, sensor networks, and wireless communications.



Dong Liang (M'11–SM'14) received the B.S. degree in electrical engineering and the M.S. degree in signal and information processing from the Hefei University of Technology, China, in 1998 and 2002, respectively, and the Ph.D. degree in pattern recognition and intelligent system from Shanghai Jiao Tong University, China, in 2006. From 2006 to 2007, he was a Postdoctoral Researcher with The University of Hong Kong. From 2007 to 2011, he was a Postdoctoral Researcher and a Research Scientist with the University of Wisconsin–Milwaukee. Since

2011, he has been with the Paul C. Lauterbur Research Centre for Biomedical Imaging, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, China, where he is currently a Full Professor. His current research interests include compressed sensing, image reconstruction, biomedical imaging, and machine learning.