IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received March 23, 2020, accepted April 6, 2020, date of publication April 9, 2020, date of current version April 21, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2986827

Deep Orthogonal Transform Feature for Image Denoising

YOON-HO SHIN, MIN-JE PARK, OH-YOUNG LEE, AND JONG-OK KIM, (Member, IEEE)
School of Electrical Engineering, Korea University, Seoul 02841, South Korea

Corresponding author: Jong-Ok Kim (jokim@korea.ac.kr)

This work was supported in part by the National Research Foundation of Korea (NRF), Ministry of Science and ICT (MSIT), South Korea, funded by the Korea Government, under Grant 2019R1A2C1005834, and in part by the Information Technology Research Center (ITRC) Support Program, supervised by the Institute of Information and Communications Technology Planning and Evaluation (IITP), under Grant IITP-2020-2018-0-01421.

ABSTRACT Recently, CNN-based image denoising has been investigated and shows better performance than conventional vision based techniques. However, there are still a couple of limits that are weak partly in restoring image details like textured regions or produce other artifacts. In this paper, we introduce noise-separable orthogonal transform features into a neural denoising framework. We specifically choose wavelet and PCA as an orthogonal transform, which achieved a good denoising performance conventionally. In addition to spatial image signals, the orthogonal transform features (OTFs) are fed into a denoising network. For the guide of the denoising process, we also concatenate OTFs from the image denoised by the existing method. This can play a role of prior for learning a denoising process. It has been confirmed that our proposed multi-input network can achieve better denoising performance than other single-input networks.

INDEX TERMS Image denoising, deep learning for image denoising, orthogonal transform, multi-input network, PCA, wavelet transform.

I. INTRODUCTION

Image denoising is a classical method required for various vision fields, which restores a noisy image to its original version with the least loss. For this goal, there have been proposed many methods of image denoising [19], [20]. Recently, CNN-based methods have been widely investigated and shows better performance than conventional vision based techniques. However, there are still a couple of limits in that they are weak partly in restoring image details like textured regions or produce other artifacts through the noise reduction process.

Most of those previous image restoration methods based on deep learning primarily work on spatial domain. Some research works [1], [2], [21] proposed denoising algorithms running on various domains such as Fourier [22], wavelet [23], edge, and etc. Those domains have been already used popularly as a tool of solving the problem of image restoration [13]–[17]. And there has been also proposed a method that decomposes image signals into cartoon and texture components and they are fused after independent processing for multi-modality image restoration [42].

In particular, orthogonal transform such as wavelet, DCT [25] and PCA [24] is beneficial in that it can divide an input noisy image into its original and additive noise roughly. Its separable ability enables noise to be removed easily. For example, if a noisy image is transformed into PCA domain, noise can be separated from the original information to some extent, and it can be reduced by simply thresholding PCA coefficients.

Motivated by this observation, we introduce orthogonal transforms into a neural denoising framework. Noise removal behaviors are learned on noise-separable transform domain as well as spatial one. Unlike the existing methods on either spatial or transform domain, both domains are simultaneously used. We specifically choose wavelet and PCA as an orthogonal transform, which achieved a good denoising performance conventionally. In addition to spatial image signals, the orthogonal transform features (OTFs) are fed into a denoising network for learning noise characteristics more accurately. There are two sorts of OTFs to be put into the network. One is the feature extracted from a noisy input image

The associate editor coordinating the review of this manuscript and approving it for publication was Gangyi Jiang.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

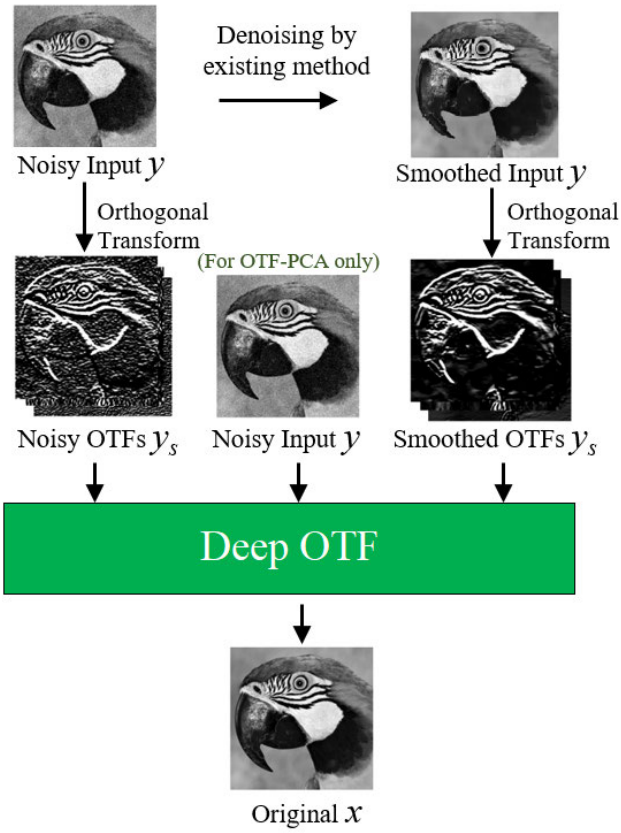

FIGURE 1. The proposed multiple inputs to deep network for image denoising. (Deep OTF means our proposed two networks, OTF-WT and OTF-PCA.)

itself, and it still includes some important information like image details as well as separated noises. For the guide of the denoising process, we also concatenate OTFs from the image denoised by the existing method. Using the smoothed features, the previous knowledge of image denoising is reflected to the neural network as prior information [10]. As illustrated in Fig. 1, the existing denoising network is extended from a single spatial input to include two additional OTFs. In other words, the proposed denoising network includes multiple inputs, and it is designed to efficiently learn to remove noise on the noise-separable orthogonal transform domain. It has been confirmed that our proposed multi-input network can achieve better denoising performance than other single-input networks throughout numerous experiments.

Multiple inputs to the deep network have been recently studied in literature for further performance improvement [6], [18]. However, the problem is which direction a single input is extended to. Namely, what information should be put into the input of the network additionally? In this paper, the extension is made in two ways as shown in Fig. 1. One is transform domain and the other is the existing denoising method to provide an example of denoising. In summary, the main contributions of the paper are summarized as

- The existing denoising network is extended from a single spatial input to multiple inputs. We propose a new multi-input denoising network which includes the orthogonal transform features of both an input noisy image and its denoised version.
- The proposed architecture paired with OTF adopts depth-wise convolutional layers [9] before feature fusion for independent learning among multiple inputs. This is because two OTF inputs are orthogonal to each other.
- Any existing denoising method can be flexibly used to obtain the smoothed OTF, and its denoising behaviors can be easily reflected to the learning process for superior denoising performance. This can play a role of prior for learning denoising.
- The proposed network architecture can be easily extended to other image restoration tasks such as image super-resolution [27], [31] and deblurring [26]. It is a general framework for neural image restoration.

The rest of the paper is organized as follows. We proceed by explaining related works, followed by describing how OTFs are generated using principal component analysis (PCA) and wavelet transform. Then, we describe a new network architecture including depth-wise convolutional blocks. We highlight better performance for both PCA and wavelet features as OTFs. Finally, the paper is concluded.

II. RELATED WORK

Image denoising has been popularly studied on orthogonal domain [2], [38], [39]. Because orthogonal transform can separate noises from a noisy image to some extent. Noises commonly tend to be distributed in high frequency regions on orthogonal domain, and they can be removed by carefully thresholding high frequency components (e.g., HH band in wavelet transform and low-priority coefficients in PCA). Previous works have concentrated on how to threshold transform coefficients on orthogonal domain in the fields of image processing and computer vision. In this paper, the denoising principle to exploit orthogonal transform has been reflected to the deep learning approach.

A. DENOISING ON ORTHOGONAL DOMAIN

Wavelet transform is an orthogonal transformation which integrates frequency and spatial information. It decomposes input data into four frequency bands (i.e., LL, LH, HL, and HH). It has been popularly used for image denosing due to its separable capability of noises. By soft-thresholding wavelet coefficients of high frequency bands, image noises can be removed [2]. In [6], it was demonstrated that Haar wavelet [28] is excellent in image denoising and a single image super-resolution, in particular in preserving image details such as texture and edge. Motivated by the superiority of wavelet transform in previous works, the proposed method chooses wavelet domain features for deep denoising network.

PCA (Principal Components Analysis) is also an orthogonal transformation which uses correlation matrix of input data, and it divides input data into a couple of orthogonal principal axes. Similar to wavelet transform, it is also good at separating image noises from original information on

transform domain. It has been popularly exploited for image denosing by thresholding PCA coefficients with less priority, which are probably expected to contain noises.

One of the representative PCA denoising methods is LPGPCA [1], which applies PCA to a local patch group and denoising is done by updating the noise level iteratively. It accomplished decent denoising performance by spatially adaptive denoising. But it has limitation in the points that the noise level must be known in advance and that overfitting can occur due to the limited number of samples resulting from the selection of similar patches in a local region [40]. Another representative PCA denoising method is PGPCA, which thresholds the PCA coefficients in a global manner [40]. But it has limitation on the performance because less similarity among patches. With the growth of deep learning, features through PCA filters were used to help neural networks to learn classification works in PCANet [11]. So, the features of PCA filters were proved to be effective for deep neural networks.

In this paper, we design a method using CNN that does not require the manual thresholding of PCA coefficients and does not have a potential risk of overfitting due to the selection of similar patches in a local region because the importance of each PCA coefficient is determined using the multiple OTFs (Noisy and smoothed OTFs).

B. DEEP LEARNING BASED DENOISING

Recently, there have been so many methods to handle a noise reduction problem, based on deep neural networks. Deep learning based denoising methods have been popularly proposed in literature, and they have some advantages, compared to the traditional vision based approach as follows. (1) They do not need optimization methods for the test phase, (2) Less need for configuring parameters manually, (3) Flexibility for architectures [36]. However, most of the state-of-the-art deep learning based methods only take a single spatial image as an input, and thus, there exists limitation in the feature extraction on high frequency domain. As a result, high frequency components like textures and edges in the output image can be damaged.

The work in [4] tried to solve a denoising problem by exploiting an MLP (multi-layer perceptron) [29] architecture. It was the first work to use an artificial neural network for image denoising. After that, with the growth of CNN, [5] proposed a CNN based residual learning architecture (so called DnCNN). Instead of generating a denoised image at the output, it tries to estimate noise signals. Motivated by DnCNN, [6] also proposed residual CNN based denoising on wavelet domain instead of spatial. It showed that topologically simple manifold of data improves the learning process of deep neural network. The work in [7] added a generative model [30] to blind denoising problems. Using the generative adversarial network, it attempts to estimate noise levels from observed noisy images. It accomplished high PSNR [33] results comparable to DnCNN [5].

Unlike the previous works mentioned above, we feed a pair of the OTFs to the denoising network. This is meaningful from the following perspectives. The smoothed OTF is roughly noise-separable, and is compared to the noisy one by concatenating them. From their comparison, we can learn the transition process from the noisy OTF to the clean.

III. ORTHOGONAL TRANSFORM FEATURE

Orthogonal transform has been popularly studied for image denoising due to its separable capability between the original information and noise. The conventional approach in image processing and computer vision typically thresholds high-frequency components on orthogonal domain. In this paper, we propose a deep learning method which works on orthogonal transform domain unlike the conventional spatial one. The proposed network can learn the complex denoising process easily by comparing noisy signals with their smoothed version on frequency domain.

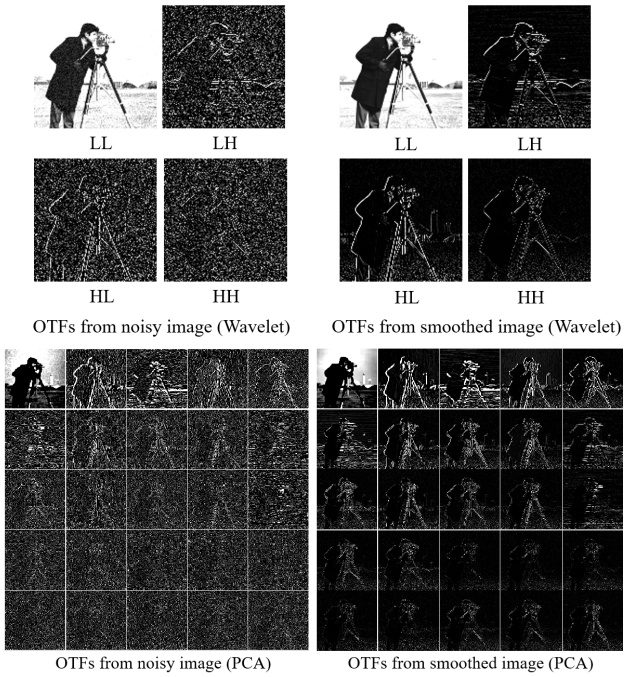

FIGURE 2. Orthogonal features of Haar wavelet transform and PCA (Multiplied by 5 for visibility except for the LL band of Wavelet domain and 1st component of PCA).

A. WAVELET TRANSFORM FEATURE

Wavelet transform (WT) has shown that it is an effective tool to reduce Gaussian noise from an image. It decomposes spatial signals into multi-levels of different time-frequency components. Especially, Haar filters are used to decompose a noisy image on discrete wavelet transform domain, and as a consequence, we can obtain four wavelet transform features (WTFs) such as LL, LH, HL, HH (L and H indicate low and high frequency components, respectively and the first letter represents horizontal component while the second does vertical) as illustrated in Fig. 2. The smoothed image

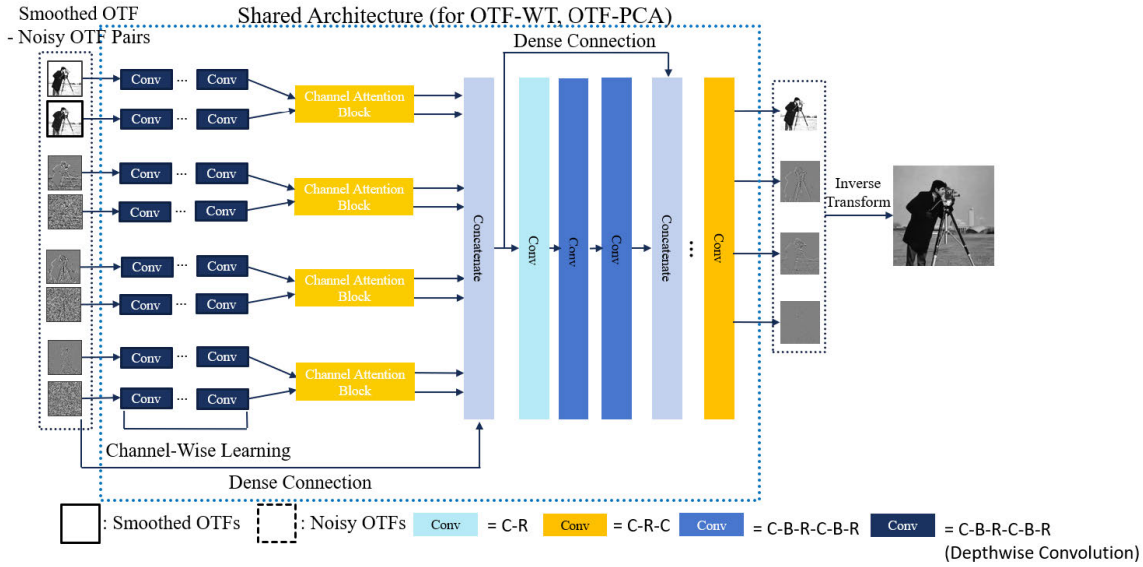

FIGURE 3. The architecture of OTF-WT for image denoising. From a noisy image y, the noisy OTFs and smoothed OTFs are extracted and trained independently. After the channel-wise learning, the trained features are fused and restored with the aim of estimating the original OTFs. Finally, the estimated features are inverse-transformed into the estimated image on spatial domain.

which is noise-reduced by the existing denoising method is also partitioned into four wavelet bands in an identical way.

If a noisy image is compared with its smoothed version on wavelet domain, it can be observed that the LL band includes less noise surely, and relatively preserves most of important structure information. On the other hand, we can see the outstanding difference between the noisy and smoothed HH bands, which contain high frequency components on both horizontal and vertical directions. The strong noise in HH can prevent the neural network learning the denoising process correctly. This is the reason that the smoothed WTFs are put into the network additionally. The smoothed features provide the network with the overall important information about the original image. Also, they are meaningful in introducing a good example of denoising, inducing the network to learn correct denoising behaviors.

B. PCA FEATURE

PCA has been an effective technique to decompose input signals into orthogonal components, and it is typically used in pattern recognition and dimensionality reduction. By transforming a noisy image into PCA domain and preserving only the most significant principal components, the noise can be removed easily from a noisy image. Similar to wavelet transform feature in previous subsection, the multiple PCA coefficients can be alternatively used as an OTF as shown in Fig. 2. By applying PCA to a noisy image, we obtain 25 PCA coefficients and eigenvectors every a 5×5 target image patch. The coefficients form the orthogonal PCA feature. The size of the patch is appropriately decided through experiments. Also, the smoothed PCA coefficients are calculated, and they are concatenated with the noisy PCA features, identical to WTF. This can help the network to converge more optimally.

C. COMPARISON BETWEEN BOTH FEATURES

When the WTF is compared with the PCA feature, both OTFs are common in that noise is partitioned to some extent from the original image information. However, there is quite distinction between two orthogonal transforms. For decomposition, wavelet transform uses pre-defined filters, which are uniformly applied to both a noisy image and its smoothed version. For PCA, basis vectors (or PCA eigenvectors) for decomposition are derived adaptively to a target image patch. Because the original image is not available at the task of denoising, the basis vectors are commonly calculated from a noisy image, and they are also used for the smoothed one. The eigenvectors of the original image may be different from those of its noisy and smoothed ones, depending on the amount of noise. Therefore, the proposed network with PCA features is designed to directly generate a denoised image at its output due to unavailable eigenvectors for reconstruction as shown in Fig. 4. Meanwhile, for WTFs the network outputs wavelet transformed signals as shown in Fig. 3, and they are inverse-transformed with wavelet filters as a post-processing. This is a key difference between WTFs and PCA features for the proposed network. The core network architecture is shared for both WTFs and PCA features except for input and output.

IV. THE PROPOSED METHOD

In previous section, we have described how to generate two types of OTFs using wavelet transform and PCA. In this section, we show how to construct a CNN-based network to perform deep sub-band learning on orthogonal domain.

A. ADDITIONAL OTF INPUTS

A noisy image y can be separated into an original clean image x and noise signal v, and is given by $y = x + v$, where the

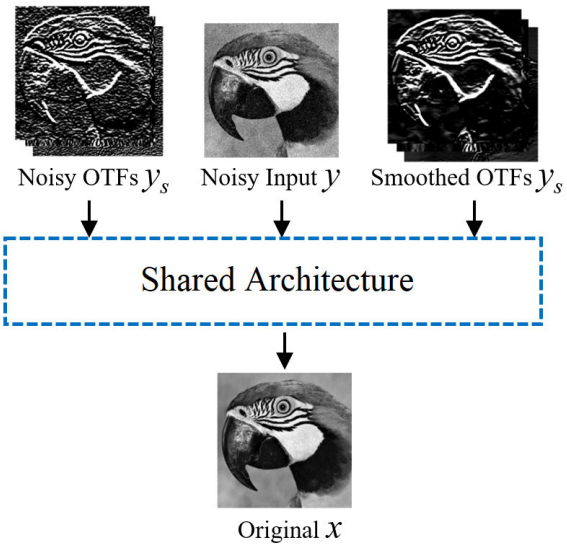

FIGURE 4. The architecture of OTF-PCA. The network adds a noisy image to the input of the network and targets at the original image on spatial domain instead of orthogonal domain of OTF- WT.

noise signal v is additive white Gaussian noise (AWGN) with standard deviation. To further improve the existing network, we propose to additionally feed OTFs from a noisy image y to the deep convolutional network. In addition to noisy OTFs, the smoothed features from the denoised image are also concatenated with the noisy ones to regularize the recovery ability. It is worth noting that the denoising is performed with conventional competitive methods. This means that the denoising behaviors of the existing method become a prior for deep learning, and they guide the network to learn denoising process correctly. Compared with a noisy image on spatial domain, two types of orthogonal transform features such as WT and PCA are more informative about subtle details of the original image such as texture and edge. Also, noises to be distributed globally on spatial domain are intensively gathered on high frequency bands of OTFs. Therefore, we develop a deep orthogonal transform feature network which learns on OTFs as well as a spatial noisy image. For wavelet transform, OTFs are formed by concatenating 4 noisy features with the associated smoothed ones, leading to 8 inputs to the network. In a similar way, input features consist of 25 noisy and the corresponding smoothed features for the case of PCA, leading to 51 inputs in total (including a spatial noisy image). Then we train the network with those multiple inputs to predict the full details of the original image.

B. CHANNEL-WISE LEARNING

We train deep orthogonal transform feature networks for image denoising. The network is constructed based on the two important assumptions. First, as described in previous section, orthogonal transform features benefit the recovery of the original image details. Second, those orthogonal features need to be trained in an independent way as the individual feature in OTFs is orthogonal to each other. It is worth noting that the OTFs are generated from either wavelet transform or PCA, which corresponds to orthogonal decomposition. At the first stage of the deep network, it would be efficient to train the features independently, and thus, this is realized by depth-wise convolutional blocks as shown in Fig. 3. Using these depth-wise convolutions, each channel is trained separately so that each orthogonal feature should not be mixed with the other ones. For channel-wise learning, a depth-wise convolution block is constructed by the combination of a depth-wise convolutional layer, a batch normalization layer and a residual layer [12]. After the channel-wise learning processes, their output features are combined with the input ones by dense connections [8]. With residual blocks and two dense connections (one concatenates the initial inputs to the feature maps after the channel-wise learning and another concatenates the feature maps after the channel-wise learning to the 3th block of the channel fusion and restoration), the deep network which includes 33 layers in total can be trained without any degradation or vanishing gradient problems. Particularly, channel-wise learning process contains 10 convolution blocks that consist of C-B-R-C-B-, and the number of feature maps in these blocks are set to 8 and 51 for OTF-WT and OTF-PCA, respectively. (C: Convolution, B: Batch-normalization, R: Relu). In Fig. 3, the difference between the light-blue and the yellow blocks is simply the number of convolution layers. Also, the dark-blue block adopts depth-wise convolution while the blue one uses normal convolution.

C. BAND-WISE CHANNEL ATTENTION

After learning orthogonal sub-band images, we consider the inter-dependencies between the noisy and the smoothed images on each sub-band. Channel attention is conducted to acquire their inter-dependent relationship [37]. It is implemented by the squeeze-and-excitation operation which consists of a squeeze operation that summarizes global information of each feature map and an excitation operation that scales each feature map's importance.

First, the global information of each feature map is extracted through the global average pooling. The inter-dependent relationship between feature maps are learned through the fully connected layers that use the global information. Finally, the outputs of the fully connected layers go through the sigmoid function and each value from the sigmoid function is multiplied to each feature map element-wisely. Through the channel attention, we can get the weights that model the inter-dependencies between the noisy and the smoothed images. The weights are multiplied to the feature maps, which are transferred to the next layer with an initial input concatenated. Fig. 5 shows the process of the channel attention.

D. CHANNEL FUSION AND RESTORATION

In sub-band based orthogonal decomposition such as wavelet transform, low-frequency and high-frequency components are extracted at different bands in a hierarchical way.

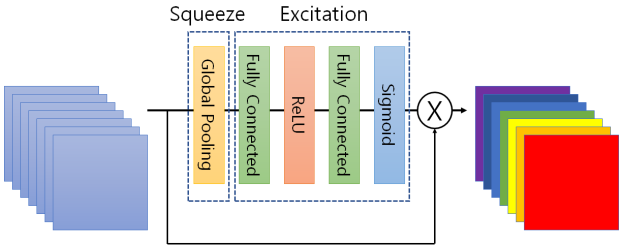

FIGURE 5. Process of channel attention. The weights acquired from the squeeze and excitation operation are multiplied to each feature map element-wisely.

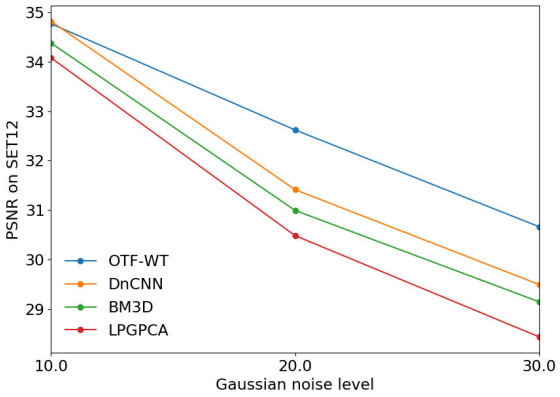

FIGURE 6. PSNR comparisons according to gaussian noise level for various denoising methods.

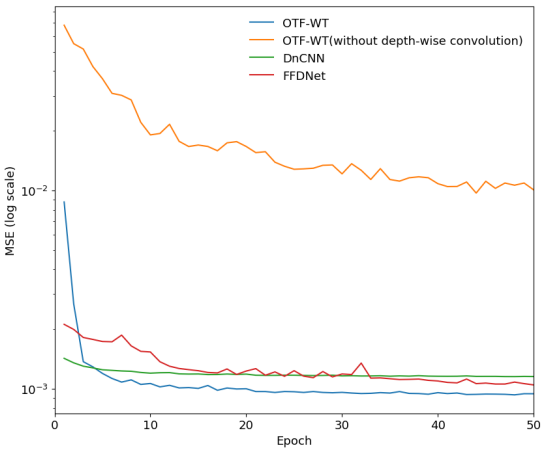

FIGURE 7. Comparison of the convergence speed for OTF-WT, OTF-WT without depth-wise convolution, FFDNet and DnCNN. The proposed OTF-WT is converged more quickly and optimally than OTF-WT without depth-wise convolution and other methods.

For PCA, orthogonal eigenvectors are determined in a descending eigenvalue order. To train the input OTFs in a channel-independent way, the first stage of the denoising network is intentionally designed to be channel-wise using depth-wise convolutional blocks as shown in Fig. 3. After the channel-wise learning, the denoised features are fused together at the second stage of the network, and a noisy input image is finally restored to the original ground truth.

Particularly, channel fusion and restoration process contain 7 convolution blocks that consist of 13 convolution layers as shown in Fig. 3. For wavelet transform, the OTFs can be perfectly inverse-transformed into the original image on spatial domain without any data loss. Thus, the denoising network is designed to generate the wavelet OTFs of the denoised image at its output. On the other hand, for PCA, the OTFs correspond to PCA coefficients, which require the corresponding eigenvectors in order to be perfectly restored into the original spatial image. However, the eigenvectors are unknown because the original clean image is not available. They can be alternatively calculated from a noisy image, but may be quite different from the original eigenvectors, depending on the amount of noise. Therefore, as shown in Fig. 4, the network is designed to directly generate a target image on spatial domain, not PCA OTFs unlike wavelet transform.

E. TRAINING

Let G(•) represent the trained network to restore the original clean image x from the noisy featuresy_N and the smoothed features y_S. Let us define m as the number of total OTFs, i.e., 4 in wavelet transform and 25 in PCA. (For PCA, the 2D feature described in Section III is aligned to a vector form fed as the network's input.) All OTFs including both the noisy and smoothed ones are represented by

$$\Theta = \{y_{N,1}, .., y_{N,m}, y_{S,1}, .., y_{S,m}\} \tag{1}$$

Given n pairs of noisy and clean images for training, a mean squared error (MSE) cost function is formulated to train the network and it is differently defined according to the type of the OTF. For wavelet transform and PCA, it is given by

$$L_{WT}(\Theta) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m}\left\|G_j(\Theta_i) - x_{F,i,j}\right\|^2 \tag{2}$$

and

$$L_{PCA}(\Theta) = \frac{1}{n}\sum_{i=1}^{n}\|G(\Theta) - x_i\|^2, \tag{3}$$

respectively.

OTF-WT is trained to reduce the MSE (mean square error) between the ground truth image and the corresponding output on wavelet domain, and OTF-PCA is trained to reduce the MSE between the ground truth image and the corresponding output on spatial domain. Indeed, the overall structures of OTF-WT and OTF-PCA are kept identical, except for two points, the number of inputs and outputs. Note that for OTF-WT, the outputs are four wavelet domain images while for OTF-PCA, the output is only one spatial image.

V. EXPERIMENTAL RESULTS

A. EXPERIMENTAL SETTING

After we conducted a couple of careful experiments, we used the same dataset and parameters for patch extraction with that of DnCNN, which we used as our network's initial denoising method.

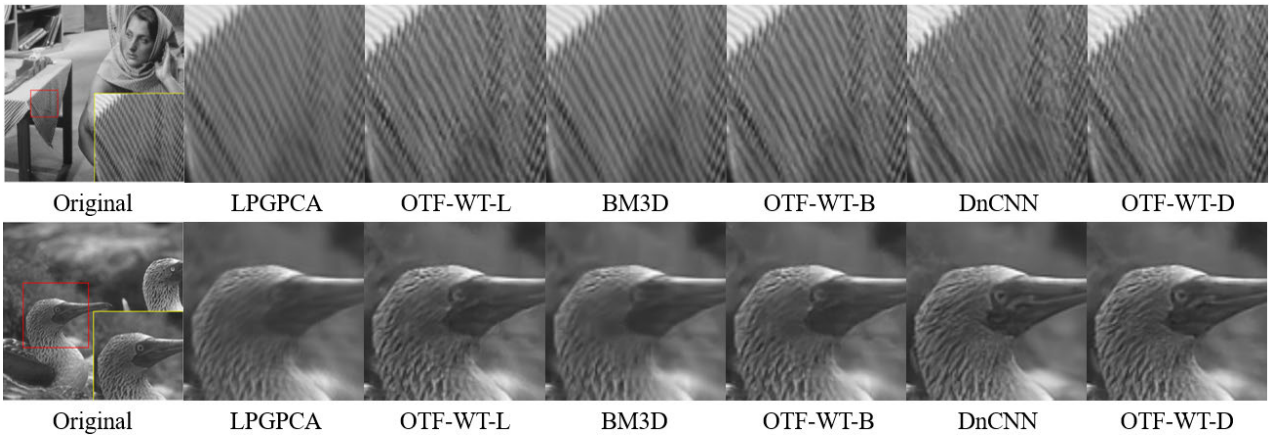

FIGURE 8. Comparison among initial denoising methods, DnCNN (OTF-WT-D), BM3D (OTF-WT-B), LPGPCA (OTF-WT-L).

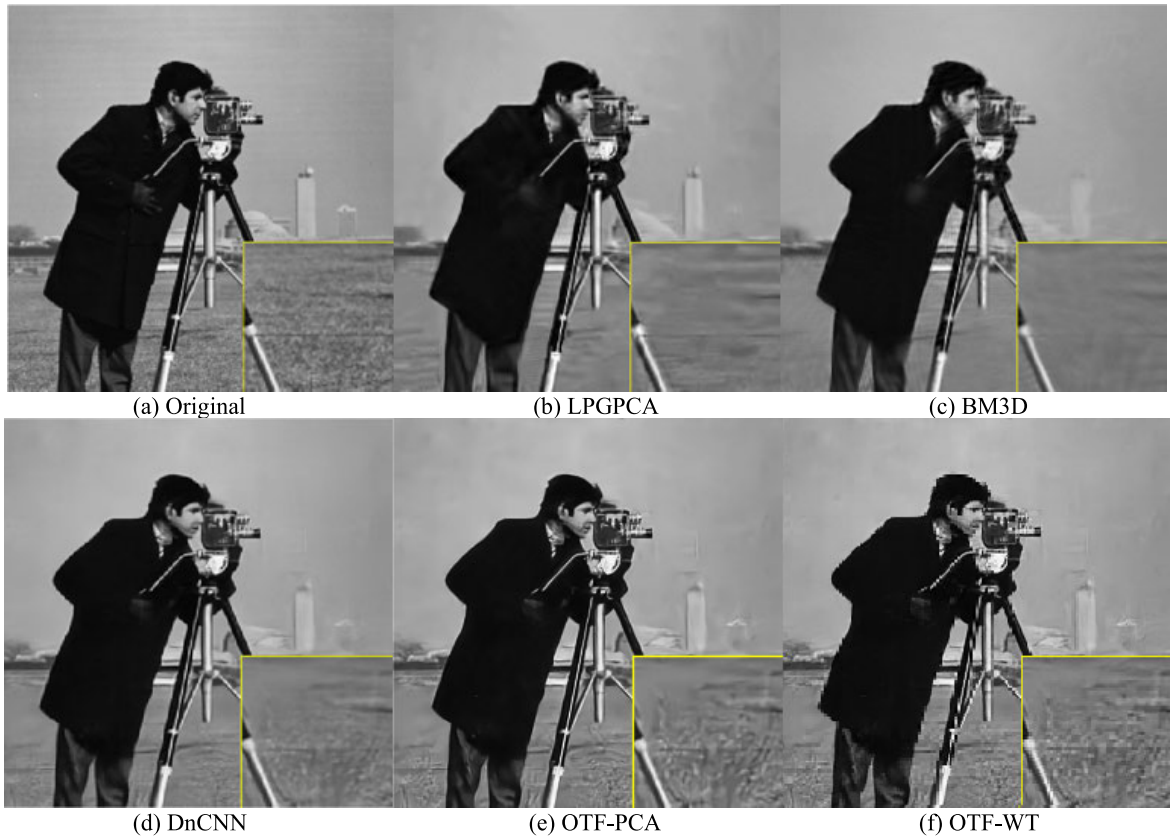

FIGURE 9. Denoising performance comparison in Cameraman. Gaussian noises with $\sigma = 30$ are added.

To train the network, we use 400 images of size 180×180 and crop them into 40×40 patches. Using stride 10 and four types of scale (1, 0.9, 0.8 and 0.7), the patches are extracted from the train images, leading to the total number of the patches, 128×1600. We consider two noise levels (i.e., Gaussian standard deviation is set to 20 and 30). For the test, we used two datasets. One is a Set68 (Berkeley segmentation dataset) [41], and the other is Set12 which is used widely in the field of image restoration. The learning rate is set to 0.0001 and decayed by 0.9 every 10 epochs. The batch size is 64 and the patch size of a noisy input is fixed to 40×40 during training. For each batch, one of eight data augmentation modes (rotation / flip / mix) is applied for training pairs. The OTF network is trained with 300 epochs. It was implemented with Pytorch. We used an Nvidia GTX 1080ti graphic processor and i7-8700 CPU. DnCNN is used to obtain the smoothed features of a noisy input as initial denoising.

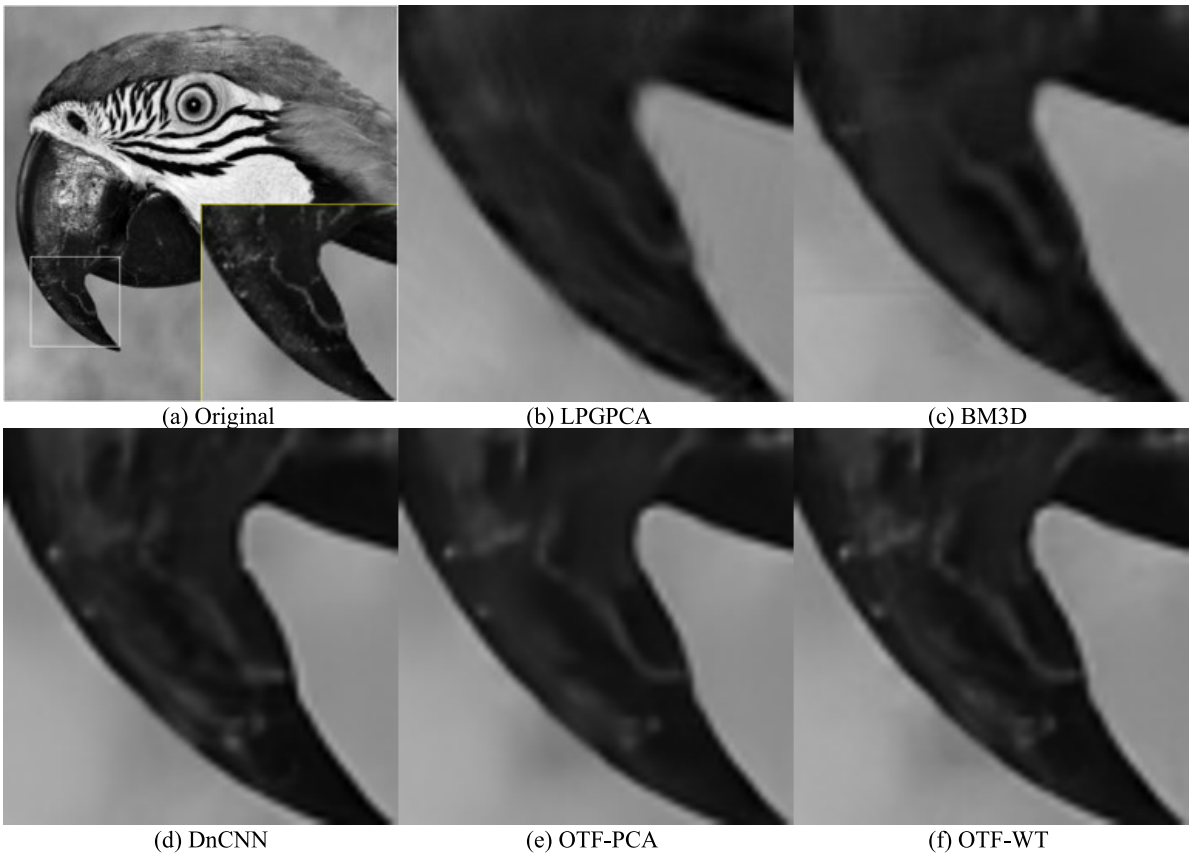

FIGURE 10. Denoising performance comparison in Parrot. Gaussian noises with $\sigma = 30$ are added.

TABLE 1. Performance comparisons in terms of PSNR/SSIM for Set12/ Set68. Note that OTF-PCA and OTF-WT represent the proposed method with PCA and WT as OTF, respectively.

Dataset		Algorithm						
		LPGPCA[1]	BM3D[3]	TWSC[32]	DnCNN[5]	FFDNet[35]	OTF-PCA	OTF-WT
σ=10	Set12	34.08/0.9179	34.38/0.9230	34.58/0.9249	**34.82/0.9276**	34.59/0.9275	34.53/0.9227	34.77/0.9271
	Set68	33.05/0.9113	33.26/0/9156	33.41/0.9195	33.85/**0.9275**	33.68/0.9264	33.64/0.9214	**33.87**/0.9270
σ=20	Set12	30.48/0.8605	30.99/0.8714	31.21/0.8737	31.41/0.8801	31.44/0.8829	32.45/0.8943	**32.62/0.8981**
	Set68	29.03/0.8188	29.60/0.8338	29.63/0.8360	30.23/0.8570	30.20/0.8580	31.49/0.8803	**31.77/0.8895**
σ=30	Set12	28.43/0.8147	29.14/0.8318	29.23/0.8351	29.49/0.8428	29.64/0.8478	30.54/0.8610	**30.66/0.8644**
	Set68	27.27/0.7562	27.76/0.7731	27.55/0.7717	28.35/0.7991	28.38/0.8041	29.49/0.8297	**29.69/0.8375**

TABLE 2. Performance comparisons in terms of PSNR/SSIM for Set12/ Set68 on $\sigma = 20$. Note that OTF-WT-L, B, D means the network using the each initial denoising method LPGPCA, BM3D and DnCNN.

Dataset	Algorithm					
	LPGPCA[1]	OTF-WT-L	BM3D[3]	OTF-WT-B	DnCNN[5]	OTF-WT-D
Set12	30.48/0.8605	31.21/0.8768	30.99/0.8714	31.28/0.8769	31.41/0.8801	**32.62/0.8981**
Set68	29.03/0.8188	29.97/0.8512	29.60/0.8338	30.02/0.8497	30.23/0.8570	**31.77/0.8895**

B. VISUAL COMPARISON

For the quantitative comparison of the denoising performance, we used the objective measures such as the peak signal to noise ratio (PSNR) [33] and SSIM [34]. Table 1 shows that the proposed networks, i.e., OTF-PCA and OTF-WT, outperform the state-of-the-art denoising methods in terms of PSNR and SSIM for datasets Set12, Set68, except for the case of $\sigma = 10$, in which we achieved almost same with DnCNN for OTF-WT and slightly lower for OTF-PCA. But when the denoised image is compared to that of the DnCNN, the proposed method accomplished the better visual quality removing the artifacts that exist in the denoised image of

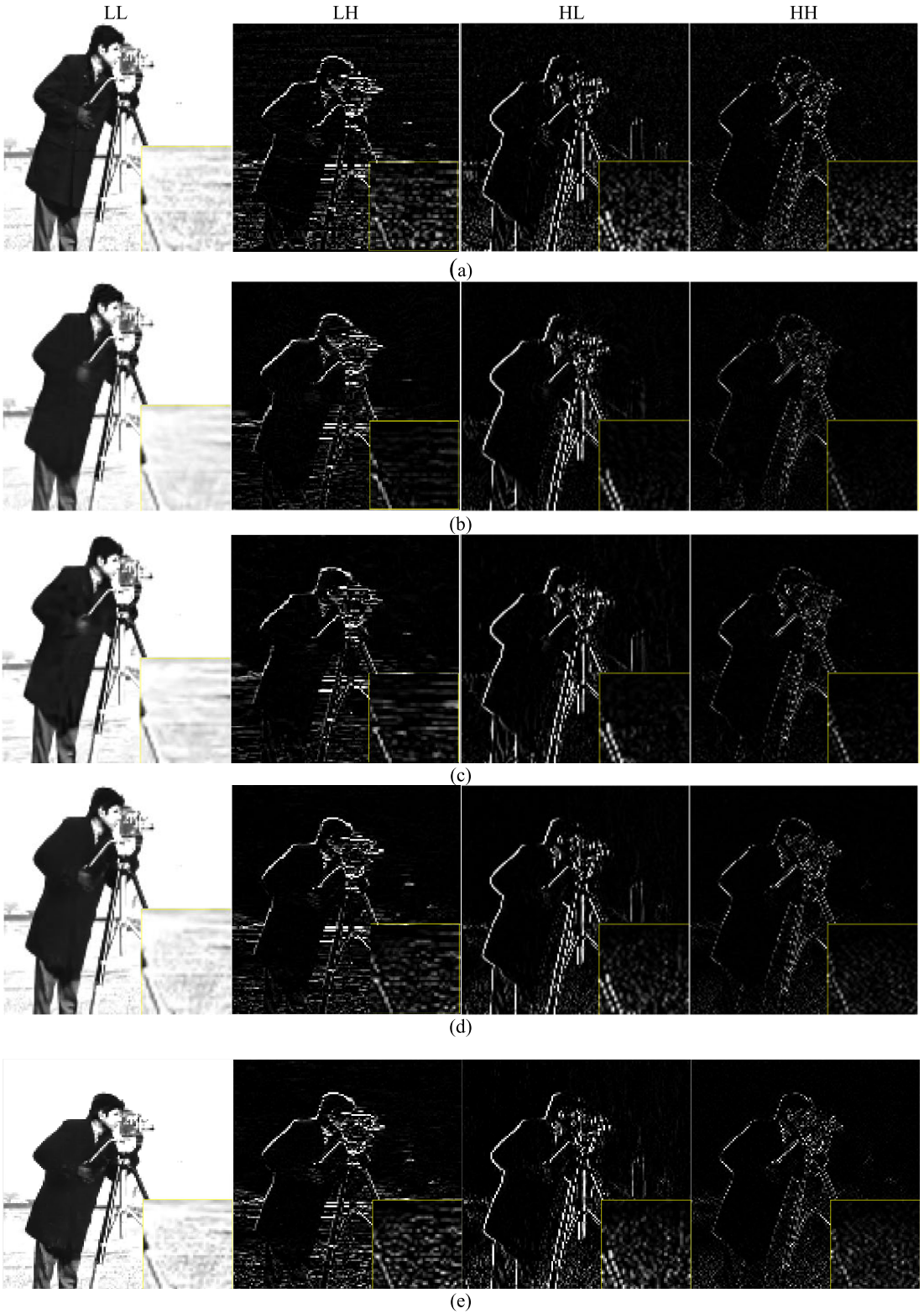

FIGURE 11. Wavelet domain comparison in Cameraman. Gaussian noises with $\sigma = 20$ are added and each image's pixel values except for LL band are multiplied by 5 for visibility. (a) Original, (b) LPGPCA, (c) BM3D, (d) DnCNN, (e) OTF-WT.

DnCNN as shown in Fig. 12. Especially, for the image with regular patterns, we accomplished better performance than the other denoising methods. Fig. 9 and Fig. 10 show the denoising examples of C*amera man* and *Parrot*. The proposed method accomplishes the best visual quality subjectively as well as objectively, particularly in the textured regions.

FIGURE 12. Visual quality comparison between DnCNN (Smoothed input for our network) and OTF-WT. Gaussian noises with $\sigma = 10$ are added.

The results of *Camera man* are also compared on wavelet domain in Fig. 11. We can see that the proposed OTF-WT restores high frequency components better than the existing methods in HL and HH bands. How the noise level (standard deviation of gaussian noises) affects the denoising performance is shown in Fig. 6.

C. EFFECT OF SMOOTHED FEATURE

Additional experiments are performed in order to find the effect of the smoothed features among OTFs on the overall learning performance. For the case of wavelet transform, the network is trained for three combinations of inputs (i.e., the original OTF-WT, OTF-WT without the smoothed features and OTF-WT without the noisy features) with noise standard deviation. Without the smoothed features, the network can be trained but the output image quality of the network is very poor as listed in Table 3. On the other hand, the network without the noisy features is not trained well and new artifacts are observed in the output image. We can see from the observation of those experiments that the smoothed features play an important role by correctly

TABLE 3. PSNR comparisons between OTF-WT and OTF-WT without the smoothed features.

Dataset		Algorithms	
		OTF-WT	OTF-WT w/o smoothed features
$\sigma = 20$	Set12	**32.62**	31.02
	Set68	**31.77**	29.92
$\sigma = 30$	Set12	**30.66**	28.65
	Set68	**29.69**	27.60

guiding denoising, and consequently, they enable the network to be learned stably. Finally, by concatenating the smoothed features additionally, the learning for denoising is boosted significantly.

D. EFFECT OF CHANNEL-WISE LEARNING

To evaluate the effect of depth-wise convolutions, a network is also constructed using only convolutional blocks. In Fig. 7, the convergence behaviors of the original OTF-WT and the OTF-WT with only convolutional blocks are compared. Even though the number of parameters in OTF WT is much less than the convolutional version, we can find that the learning ability of OTF-WT is much faster than OTF-WT with only convolutional blocks. Using depth-wise convolutions enables the network to achieve superior computational efficiency and provides better network architecture for independent OTF feature learning.

E. COMPUTATIONAL EFFICIENCY

Our OTF-WT network is better than other existing methods in terms of the computational efficiency as well as the performance. We evaluated the computational efficiency in terms of two criteria, the number of the network parameters and the convergence speed. Fig. 7 and Table 4 show the results of the comparisons. We can see that the proposed method is better than other methods in both criteria.

TABLE 4. Comparison of the number of the network parameters for DnCNN, FFDNet and OTF-WT.

Network	DnCNN	FFDNet	OTF-WT
# of parameters	556,096	486,080	430,228
% relative to DnCNN	100%	87.41%	77.37%

F. CO-EXISTENCE WITH EXISTING METHODS

In this paper, we attempt to develop the efficient learning of complex denoising process by referring to the denoising capability of existing excellent methods. Thus, our method can friendly work with any other denoising methods. Fig. 8 compares the results of initial denoising with LPGPCA, BM3D and DnCNN. OTF-WT-L, OTF-WT-B, OTF-WT-D means using the each initial denoising method LPGPCA, BM3D and DnCNN. OTF-WT based on BM3D (OTF-WT- B) generates more clean results at grid-shaped patterns than OTF-D and OTF-WT-L. In comparison, OTF-WT based on DnCNN (OTF-WT-D) shows better restoration in complex texture regions as shown in Fig. 8.

VI. CONCLUSION

In this paper, we proposed a novel orthogonal transform feature such as wavelet transform and PCA for deep image denoising. Wavelet transform and PCA have been popularly investigated for image denoising due to their capability to separate noises from a noisy image. Inspired by this observation, we add orthogonal transform features to the image denoising network in addition to a spatial noisy image input. We also guide the network to correctly learn a denoising process by providing a pair of denoising examples using a competitive existing method. In order to reflect the prior experience of the existing denoising method to the learning process, we feed a pair of noisy orthogonal features and its smoothed version to the network together. It would be easier to identify noises on orthogonal transform domain rather than spatial one. It is expected that this could help the network to learn the denoising process optimally and stably. The proposed orthogonal transform features have been thoroughly evaluated with a variety of test images. Experimental results show that they can accomplish superior visual quality to the existing CNN and non-CNN denoising methods.

As a future work, our network will be extended to other image restoration applications such as image super-resolution and deblurring. Also, we try to find a more optimal network architecture to train OTFs. We may construct a fully orthogonal network without the original convolutional layers which prevent the orthogonality of input data.

ACKNOWLEDGMENT

(Yoon-Ho Shin and Min-Je Park contributed equally to this work.)

REFERENCES

[1] L. Zhang, W. Dong, D. Zhang, and G. Shi, "Two-stage image denoising by principal component analysis with local pixel grouping," *Pattern Recognit.*, vol. 43, no. 4, pp. 1531–1549, Apr. 2010.
[2] S. G. Chang, B. Yu, and M. Vetterli, "Adaptive wavelet thresholding for image denoising and compression," *IEEE Trans. Image Process.*, vol. 9, no. 9, pp. 1532–1546, Sep. 2000.
[3] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image restoration by sparse 3D transform-domain collaborative filtering," *Proc. SPIE Electron. Imag.*, vol. 6812, pp. 681207-1–681207-12, Mar. 2008.
[4] H. C. Burger, C. J. Schuler, and S. Harmeling, "Image denoising: Can plain neural networks compete with BM3D?" in *Proc. CVPR*, Jun. 2012, pp. 2392–2399.
[5] 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.
[6] W. Bae, J. Yoo, and J. C. Ye, "Beyond deep residual learning for image restoration: Persistent homology-guided manifold simplification," in *Proc. CVPR*, Jul. 2017, pp. 2392–2399.
[7] J. Chen, J. Chen, H. Chao, and M. Yang, "Image blind denoising with generative adversarial network based noise modeling," in *Proc. CVPR*, Jun. 2018, pp. 3155–3164.
[8] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. CVPR*, Jul. 2017, pp. 4700–4708.
[9] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. CVPR*, Jul. 2017, pp. 1251–1258.
[10] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," Mar. 2015, *arXiv:1503.02531*. [Online]. Available: http://arxiv.org/abs/1503.02531
[11] T.-H. Chan, K. Jia, S. Gao, J. Lu, Z. Zeng, and Y. Ma, "PCANet: A simple deep learning baseline for image classification?" *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5017–5032, Dec. 2015.
[12] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. CVPR*, Jun. 2016, pp. 770–778.

[13] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, Dec. 2006.
[14] 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.
[15] M. D. Robinson, C. A. Toth, J. Y. Lo, and S. Farsiu, "Efficient Fourier-wavelet super-resolution," *IEEE Trans. Image Process.*, vol. 19, no. 10, pp. 2669–2681, Oct. 2010.
[16] S. Zhao, H. Han, and S. Peng, "Wavelet-domain HMT-based image super-resolution," in *Proc. ICIP*, 2003, p. II-953.
[17] W. Zhang and W.-K. Cham, "Learning-based face hallucination in DCT domain," in *Proc. CVPR*, Jun. 2008, pp. 1–8.
[18] W. Yang, J. Feng, J. Yang, F. Zhao, J. Liu, Z. Guo, and S. Yan, "Deep edge guided recurrent residual learning for image super-resolution," *IEEE Trans. Image Process.*, vol. 26, no. 12, pp. 5895–5907, Dec. 2017.
[19] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Non-local sparse models for image restoration," in *Proc. CVPR*, Sep. 2009, pp. 2272–2279.
[20] D. Zoran and Y. Weiss, "From learning models of natural image patches to whole image restoration," in *Proc. ICCV*, Nov. 2011, pp. 479–486.
[21] S. G. Chang, B. Yu, and M. Vetterli, "Spatially adaptive wavelet thresholding with context modeling for image denoising," *IEEE Trans. Image Process.*, vol. 9, no. 9, pp. 1522–1531, Sep. 2000.
[22] R. N. Bracewell, *The Fourier Transform and Its Applications*, 3rd ed. New York, NY, USA: McGraw-Hill, 1999.
[23] I. Daubechies, "The wavelet transform, time-frequency localization and signal analysis," *IEEE Trans. Inf. Theory*, vol. 36, no. 5, pp. 961–1005, Sep. 1990.
[24] I. Jolliffe, *Principal Component Analysis*. New York, NY, USA: Springer-Verlag, 1986.
[25] N. Ahmed, T. Natarajan, and K. R. Rao, "Discrete cosine transform," *IEEE Trans. Comput.*, vol. C-23, no. 1, pp. 90–93, Jan. 1974.
[26] 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.
[27] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2861–2873, Nov. 2010.
[28] R. S. Stankovic and B. J. Falkowski, "The Haar wavelet transform: Its status and achievements," *Comput. Electr. Eng.*, vol. 29, no. 1, pp. 25–44, Jan. 2003.
[29] M. W. Gardner and S. R. Dorling, "Artificial neural networks (the multilayer perceptron)—A review of applications in the atmospheric sciences," *Atmos. Environ.*, vol. 32, nos. 14–15, pp. 2627–2636, Aug. 1998.
[30] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. NIPS*, 2014, pp. 2672–2680.
[31] O.-Y. Lee, Y.-H. Shin, and J.-O. Kim, "Multi-perspective discriminators-based generative adversarial network for image super resolution," *IEEE Access*, vol. 7, pp. 136496–136510, 2019.
[32] J. Xu, L. Zhang, and D. Zhang, "A trilateral weighted sparse coding scheme for real-world image denoising," in *Proc. ECCV*, Sep. 2018, pp. 20–36.
[33] Y. Fisher, *Fractal Image Compression: Theory and Application*. New York, NY, USA: Springer, 2012.
[34] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[35] 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.
[36] C. Tian, L. Fei, W. Zheng, Y. Xu, W. Zuo, and C.-W. Lin, "Deep learning on image denoising: An overview," 2019, *arXiv:1912.13171*. [Online]. Available: http://arxiv.org/abs/1912.13171
[37] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. CVPR*, Jul. 2018, pp. 7132–7141.
[38] 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.
[39] L. Kaur, S. Gupta, and R. C. Chauhan, "Image denoising using wavelet thresholding," in *Proc. ICVGIP*, vol. 2, 2002, pp. 16–18.
[40] C.-A. Deledalle, J. Salmon, and A. Dalalyan, "Image denoising with patch based PCA: Local versus global," in *Proc. BMVC*, vol. 81, 2011, pp. 425–455.
[41] S. Roth and M. J. Black, "Fields of experts," *Int. J. Comput. Vis.*, vol. 82, no. 2, pp. 205–229, Apr. 2009.
[42] Z. Zhu, H. Yin, Y. Chai, Y. Li, and G. Qi, "A novel multi-modality image fusion method based on image decomposition and sparse representation," *Inf. Sci.*, vol. 432, pp. 516–529, Mar. 2018.

YOON-HO SHIN received the B.S. degree from the School of Electrical Engineering, Korea University, Seoul, South Korea, in 2017, and the M.S. degree in electrical engineering with a specialization in signal processing and multimedia from Korea University. His research interests are image denoising, image super-resolution, and deep learning.

MIN-JE PARK received the B.S. degree from the School of Electrical Engineering, Korea University, Seoul, South Korea, in 2019, where he is currently pursuing the master's degree in electrical engineering specializing in signal processing and multimedia. His research interests are image denoising, hyperspectral imaging, and deep learning.

OH-YOUNG LEE received the B.S. degree from the School of Electrical Engineering, Korea University, Seoul, South Korea, in 2011, and the master's and Ph.D. degrees in electrical engineering with a specialization in signal processing and multimedia from Korea University. His research interests are image/video processing, super resolution, noise reduction, and deep learning.

JONG-OK KIM (Member, IEEE) received the B.S. and M.S. degrees in electronics engineering from Korea University, Seoul, South Korea, in 1994 and 2000, respectively, and the Ph.D. degree in information networking from Osaka University, Osaka, Japan, in 2006. From 1995 to 1998, he served as an Officer of Korea Air Force. From 2000 to 2003, he was with SK Telecom Research and Development Center and Mcubeworks, Inc., South Korea, where he was involved in research and development on mobile multimedia systems. From 2006 to 2009, he was a Researcher with the Advanced Telecommunication Research Institute International (ATR), Kyoto, Japan. He joined Korea University, in 2009, where he is currently a Professor. His current research interests are in the areas of image processing, computer vision, and intelligent media systems. He was a recipient of the Japanese Government Scholarship, from 2003 to 2006.