Frequency Attention Network: Blind Noise Removal for Real Images

Hongcheng Mo[1,2], Jianfei Jiang[1], Qin Wang[1(✉)], Dong Yin[2], Pengyu Dong[2], and Jingjun Tian[2]

[1] Shanghai Jiao Tong University, Shanghai 200240, China
{momo1689,qinqinwang}@sjtu.edu.cn
[2] Fullhan, Shanghai, China

Abstract. With outstanding feature extraction capabilities, deep convolutional neural networks(CNNs) have achieved extraordinary improvements in image denoising tasks. However, because of the difference of statistical characteristics of signal-dependent noise and signal-independent noise, it is hard to model real noise for training and blind real image denoising is still an important challenge problem. In this work we propose a method for blind image denoising that combines frequency domain analysis and attention mechanism, named frequency attention network (FAN). We adopt wavelet transform to convert images from spatial domain to frequency domain with more sparse features to utilize spectral information and structure information. For the denoising task, the objective of the neural network is to estimate the optimal solution of the wavelet coefficients of the clean image by nonlinear characteristics, which makes FAN possess good interpretability. Meanwhile, spatial and channel mechanisms are employed to enhance feature maps at different scales for capturing contextual information. Extensive experiments on the synthetic noise dataset and two real-world noise benchmarks indicate the superiority of our method over other competing methods at different noise type cases in blind image denoising.

1 Introduction

Image denoising is a very critical low-level task in computer vision, and the quality of the image has a significant impact on high-level tasks – image classification, semantic segmentation, object localization, instance segmentation. Image denoising is an ill-posed problem like a transcendental equation which is difficult to find a unique solution by reversing it but only by optimizing it. For a real-world noisy image, noise usually results from the interaction of photons of the image and electrons thermal movement when the sensor obtains the image signal, such as shot noise, dark current noise, quantification noise, etc [1] [2]. During ISP pipeline, some nonlinear operations like demosaicking and gamma correction also change the noise distribution which makes real-world noise more sophisticated.

For decades, methods of image denoising can be divided into two categories, model-driven and data-driven. One exact scheme of the first category is to find

similar blocks by exploiting the correlations that exist in the image information itself and using these similar blocks to estimate the clean image [3] [4] [5]. Another direction is to exploit prior information based on the transformed domain by converting image signal to another domain for shrinkage like frequency domain. The noise is usually contained in the medium and high frequency information of the image due to the randomness and sparsity of the noise which causes the gradient of the image increases so that reasonable reconstruction of the medium and high frequency information can effectively remove noise [6] [7]. Recently, data-driven deep convolutional neural networks(CNNs) are increasingly applied on the image denoising task in that CNN can extract high-dimensional features of images and utilize them to restore clean images.

Deep CNN denoisers significantly improve image denoising performance on synthetic noise model [8] [9] [10] [11] [12] [13]but tend to be over-fitted to the realistic noise model like additive white Gaussian noise(AWGN). When they are applied on the real-world denoising task, they are lack of the ability to eliminate both signal-dependent noise and signal-independent noise. Thus, real-world denoising is still a challenging task since noise distributions vary a lot in different scenes.

In this article, we tackle this issue by developing a frequency attention network (FAN) which combine frequency domain analysis and a deep CNN model for the image denoising task. From the perspective of Fourier, image transformation methods that can enhance the performance of the network are essentially to change the frequency domain information of the image to enhance the feature [14]. Motivated by this perspective and sensitivity of frequency components in human visual system(HVS) [15], we employ wavelet transform to extract the spectral information and structure information of images as the prior information of the network. The wavelet transform can preserve the structural information of the image and facilitate high-dimensional features for image restoration.

Flexibility and robustness are still significant problems for most denoising methods. [9] [16] introduce the noise map to train denoising networks jointly which can provide more information to help extract image features and we adopt it to accelerate our network training. We combine spatial attention mechanism and channel attention mechanism to enhance the feature map for improving the feature extraction ability of the network and removing signal-dependent noise. At last, we explore the influence of different individual components of the network and different wavelet basis functions.

To sum up, the contributions of this paper are following:

- We proposed frequency attention networks(FAN) combining traditional signal processing methods and deep learning, which makes the method based on neural network with more interpretability from the perspective of frequency domain.
- We introduced the *Spatial-Channel Attention Block* which combines the spatial attention and channel attention mechanisms to enhance feature maps and help to better extract the main features of the image.

- We evaluate the effect of different wavelet basis functions on denoising performance and experiments show Haar wavelet with symmetry, orthogonality and compactly supported characteristics can acheive the best result.
- Experiments on synthetic noise datasets and real noise datasets respectively prove the superiority of our model compared with competing methods and can achieve state-of-the-art results.

2 Related Work

With the rapid development of convolutional neural networks in the field of computer vision, many researchers have proposed algorithms based on deep learning to solve low-level and high-level computer vision problems. Some combination of traditional methods and deep learning provides more comprehensive interpretability for neural networks. Next, we briefly describe representative methods of image denoising.

Traditional Denoising Methods: Noise images can be seen as the sum of clean signal and noise signal with their relationship usually expressed as $\boldsymbol{Y} = \boldsymbol{X} + \boldsymbol{n}$. Towards employing high frequency characteristic of noise, image denoising methods on the different transform domains were widely proposed especially DCT [7] [17] and wavelet transform [6] [18]. Wavelet transform is a widely application method in signal processing, which uses the wavelet basis function as a filter to extract the frequency information of the image and then solve sub-band reconstruction coefficients in accordance with the desired expectation to obtain a denoised image. Based on images self-similar patches, it is effective to make full use of the structural information of the images themselves as prior information to approximate the optimal estimate clean patches through their statistical relationships involving CBM3D [19], NLM [3], NL-Bayes [5]. Another type of solution is to convert the image denoising task to a mathematical optimization problem based on the noise model and use decomposition or dictionary-based to solve it, such as low-rank model [20] [21], sparse representation [22] [23] [24]. However, the main drawbacks of these algorithms are computationally expensive and time-consuming in that they need to be re-implemented for new coming images so that they are difficult in gaining wider access.

Deep CNN denoiser: With the widespread use of deep convolutional networks in computer vision, deep CNN has also led to great performance on image denoising. DnCNN [8] is the first to introduce a residual network for denoising, allowing the network to learn the distribution of noise and then remove the noise to get the result. FFDNet [9] adopts downsampling, introducing noise map and orthogonalization to improve the speed, flexibility and robustness of the denoising network. CBDNet [16] adopts two-stage denoising strategy including estimating noise map and denoising, using asymmetric loss function for training. MWCNN [25] replaces the downsampling of UNet with wavelet transform to retain more information using orthogonalization. VDN [26] introduces variable inference to predict the noise distribution for denoising. CBDNet, MWCNN and VDN all adopt UNet network [27] as their backbone which includes downsam-

pling and can be operated for pixel-level tasks, with excellent results for image restoration task that require attention to pixel points.

An improvement for image denoising is to apply attention mechanisms for adapting to different regions. RIDNet [28] introduces enhancement attention modules to select essential features and use a residual on residual structure to build networks. KPN [29] utilizes the idea of non-local mean to train out filter windows and then use these filter cores for image reconstruction. [30] considers the deformable convolution to predict the distribution of filter kernels for obtaining good results. However, these networks are learned in spatial domain which are costly to learn and underutilize the frequency domain characteristics of the noise.

More recently, frequency domain learning has shown its potential to improve model efficiency and feature extraction capabilities. Both [31] and [32] employ DCT transformations to convert training data into the frequency domain while the front applied to image classification and the latter applied to image segmentation. For the image denoising task, the spectral information can reflect the relationship between noise and clean image signal in the frequency domain. Wavelet transform is commonly used to convert images to the frequency domain which preserves both the spatial structure information and the spectral information of the image at the same time. Thus, we propose our method based on the characteristic of wavelet transform and deep learning for blind image denoising.

3 Frequency Attention Network

This section presents our FAN consisting of data pre-processing, networks architecture and attention design. To begin with, we show FAN architecture including *Est-Net* and *De-net*. Then, we analysis the effects of wavelet transform characteristics on network performance. Finally, we introduce *Spatial-Channel Attention Block* used for feature map enhancement.

3.1 Network Architecture

Inspired by the observation that human visual system(HVS) is more sensitive to the spatial resolution of the luminance signal than that of the chrominance signal [33] and HVS has the varying sensitivity to different frequency components [15], we convert the image from the RGB color space to YCbCr color space for denoising.

As shown in figure 1, the network we designed contains two subnetworks including the noise estimation network and the denoising network. *Est-Net* takes the noisy image as input and estimates different noise level map for each channel. *Est-Net* is composed of five full-convolutional layers, each consisting of only the Conv and PReLU layers excluding the batch normalization layer and the pooling layer. The filter size is 3×3 and the feature map is set as 64. The noise map which can improve network flexibility and generalizability for different noise levels is also conductive to increase the convergence speed of the network because

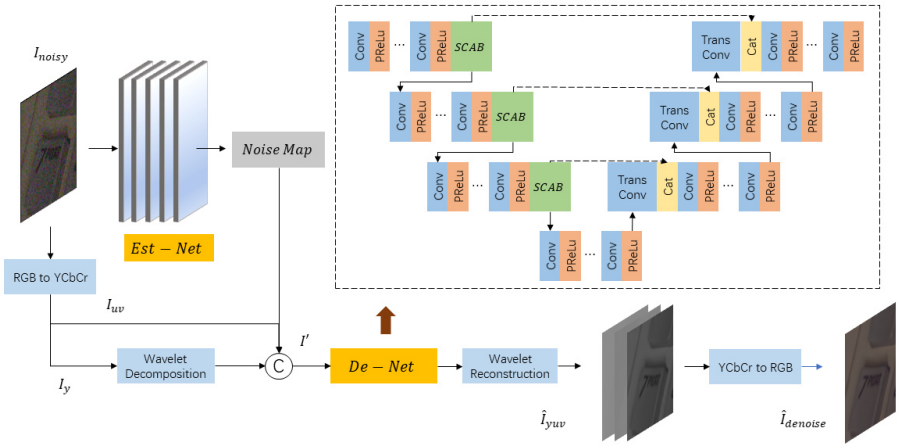


Fig. 1: Frequency Attention Network Architecture

its redundancy can help to better extract image features. Given a training set $\{I_{noisy}, I_{gt}, \sigma_{gt}\}_{i=1}^{N}$, loss of the estimated noise map $\hat{\sigma}(I_{noisy})$ is defined as

$$\mathcal{L}_{map} = \frac{1}{N}\sum_{i=1}^{N} ||\hat{\sigma}(I_{noisy}) - \sigma_{gt}||_1 \tag{1}$$

where $||\cdot||_1$ denotes l_1 norm.

In the denoising network, the input RGB images are transformed to the YCbCr color space including luminance I_y and chroma I_{uv}. Considering the sensitivity of the human visual system to luminance, I_y is converted to the frequency domain by wavelet transform for reserving spectral information and structure information while I_{uv} stays original. We concatenate the processed data and noise map to get $I^{'}$ as input to the *De-Net* network which contains 4 encoder blocks and 3 decoder blocks and there is a skip connection to concatenate two blocks under the same scale.

The U-Net structure can use feature fusion at different resolutions to obtain better contextualized representations, but it will cause the loss of image details at high resolution when the network depth at each scale is consistent. Aim at retaining more details, we adopted variable depth design for *De-Net* where the numbers of residual convolutional blocks at different scales increase with the resolution to ensure that our network can obtain a stronger expressive ability. Besides, we add the SCAB module after each encoder for feature enhancement. For the denoised image $\hat{I}(I_{noisy})$, we define image loss of *De-Net* as

$$\mathcal{L}_{img} = \frac{1}{N}\sum_{i=1}^{N} ||\hat{I}(I_{noisy}) - I_{gt}||_1 \tag{2}$$

Finally, we obtain the complete denoised image by wavelet reconstruction and convert it to RGB color space. Given a training set $\{I_{noisy}, I_{gt}\}_{i=1}^{N}$, loss

function of our networks is defined as

$$\mathcal{L}(\theta) = \mathcal{L}_{img} + \gamma \mathcal{L}_{map} \tag{3}$$

3.2 Wavelet Transform

Wavelet is an important tool for the analysis of unstable signals while the image as a 2D plane unstable signal is well suited for study with the wavelet. Wavelet transformation of one image decomposes the image into different sub-bands based on the frequency information and the processing of the medium and the high frequency sub-bands can result in noise removal. Two-dimensional wavelet decomposition of level j can be described as

$$f(x, y) = \sum_{j,m,n \in Z} a_{j,m,n}(k) \psi_{j,m,n}(x, y) \tag{4}$$

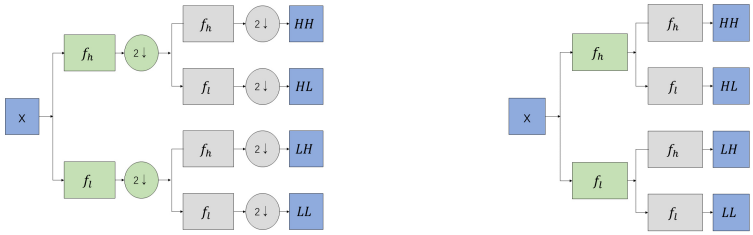
where $f(x)$ can be expanded into a linear combination of wavelet basis functions with $a_{j,m,n}(k)$ as the expansion factor and $\psi_{j,m,n}$ as the wavelet basis function. For FAN, reconstruction denoising result can be defined as following,

$$\hat{I}(I_{noisy}) = \sum_{k \in Z} W(a_k) \psi_k(I_{noisy}) \tag{5}$$

where $W(\cdot)$ refers to network output. It is difficult to deduce the expression relationship between $W(a_k)$ and a_k through mathematical theory but the optimal estimation of a_k can be obtained with the help of the nonlinear characteristics of the neural network and auxiliary noise map by deep learning to restore the image as close as possible to ground truth. Therefore, the learning objective of our neural network can be abstracted as the optimal solution to the wavelet coefficients.

Separable wavelets which is widely used for two-dimension wavelet transform generally use orthogonal wavelets and for discrete signal discrete orthogonal wavelets have completeness to retain all the energy of the image signal in the transformation process. Meanwhile, orthogonal wavelets can reduce the data correlation between different sub-bands. Therefore, using wavelet to transform the image into the frequency domain will not lose any information. Instead, it can use the frequency characteristics of the image signal in the frequency domain to help the deep convolutional network extract its nonlinear characteristics.

The wavelet transform includes discrete wavelet transform(DWT), stationary wavelet transform(SWT) and continuous wavelet transform(CWT) where CWT is the analysis of continuous signals, DWT and SWT are the analysis of discrete signals. The decompose transformation process of DWT and SWT is shown in the Figure 2. SWT is also called unsampled wavelet transform, which can calculate the wavelet transform value point by point. The biggest difference between SWT and DWT is that SWT has translation invariance because DWT performs downsampling operation during the calculation process which will cause the pseudo-Gibbs phenomena of reconstruction images. Meanwhile, SWT can



(a) DWT (b) SWT

Fig. 2: Discrete Wavelet Transform and Stationary Wavelet Tranform

make the decomposition result keep the same size as the original image to retain more prior information for network training and avoid information loss caused by the downsampling and upsampling operation of the chroma layer for ensuring the same size of the input and the output when we perform DWT on the luminance layer instead of the chroma layer.

3.3 Spatial-Channel Attention Block(SCAB)

Signal-independent noise can be easily filtered out from the wavelet sub-band through neural network learning, but signal-dependent noise is not easy to remove because of the high correlation between high-frequency signal and noise. In order to make full use of the inter-channel and inter-spatial relationships of the image, we used a *Spatial-Channel Attention Block* to extract the features in the convolutional stream. The schematic of SCAB is shown in Figure 3. We extracted the distribution of noise levels through *Est-Net* and also characterized the structure information of noise which we can use spatial attention mechanism to refine features map of I_y. Meanwhile, we apply channel attention mechanism [34] on I_y to achieve the feature recalibration.

Spatial attention is used to extract the inter-spatial relationship of images. Non-local block [35] [12] generates attention map through point-by-point calculation of the feature map, but this method has limitations on the size of the image and the calculation amount is large when the image size is too large. CBAM [36] uses GAP and GMP to utilize the full channel information which saves the amount of calculation while missing some feature information. We use a progressively expanding multi-layer convolution operation to obtain an effective tradeoff between model complexity and performance. Dilated convolution and expanding filter kernel size are adopted to increase the receptive field, and gradually decreasing channels reduce the computational complexity. At the same time, 1×1 convolution layers distributed between convolutional layers work to gather feature information in receptive fields of different ranges so as to calculate the dependency relationship on the feature map space.

Channel attention utilizes the squeeze and excitation operation [34] to enhance the main features of the feature map based on the inter-channel rela-

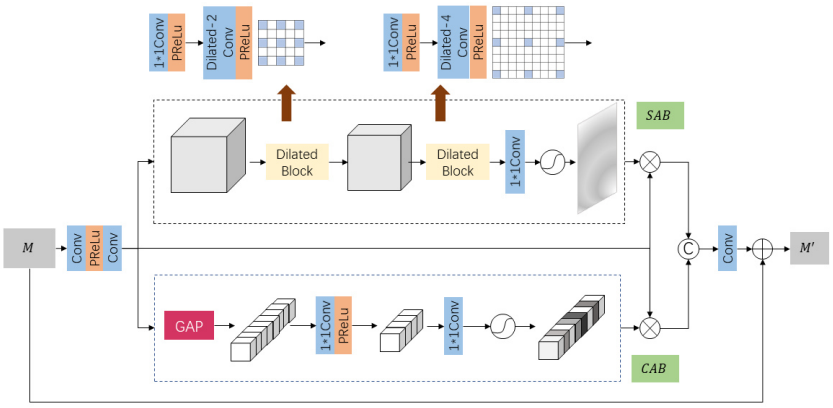


Fig. 3: Spatial-Channel Attention Block

tionship. For an input feature map $\boldsymbol{M} \in \mathbb{R}^{1\times1\times c}$,firstly generate the channel-by-channel statistics $\boldsymbol{d} \in \mathbb{R}^{1\times1\times c}$ through global pooling as the squeeze operation. This statistic expresses the entire image under this type of feature extraction convolution kernel global description. The excitation operation is used for fully capturing the dependencies between channels through two convolutional layers with the sigmoid gating and obtains activations $\hat{\boldsymbol{d}} \in \mathbb{R}^{1\times1\times c}$ as follow

$$\hat{d} = \sigma(W_2\delta(W_1\mathrm{GAP}(f_{in}))) \tag{6}$$

where δ refers to PReLU function and σ is sigmoid gating.

In the image restoration, the attention mechanism can be regarded as an extension of the classical method ideas on the neural network. Similar to the bilateral filter [37] which adopts the difference between the spatial domain and the value domain to calculate the weight of different locations for the central area, spatial attention mechanism can re-weight the feature map according to the location of the features and help the network learn where to be paid attention. Channel attention is the overall enhancement of different types of features, which emphasizes the features corresponding to edge information in order to achieve the effect of better retaining edges. The fusion of spatial and channel attention mechanisms can enhance the feature maps in the high dimension, which is conducive to smoothing the flat region and recovering the details of the texture region.

4 Experiments

In this section, we design various ablation study to demonstrate effectiveness of our strategy and evaluate performance by our method on synthetic and real noise datasets compard with previous outstanding methods.

Table 1: Impact of each individual components(Test on SIDD validation dataset)

Wavelet		✓		✓	✓	✓	✓
SCAB			✓		✓	✓	✓
Noise Map	✓	✓	✓	✓		✓	✓
Variable Depth			✓	✓	✓		✓
PSNR(dB)	39.04	39.13	39.29	39.16	39.24	39.37	39.45
SSIM	0.9124	0.9139	0.9157	0.9144	0.9158	0.9172	0.9184

Table 2: Impact of Different Types of Wavelet on Real-World Noise(SIDD validation dataset)

Wavelet	Haar	Daubechies				Biorthogonal		
		db2	db3	db4	db5	bior2.2	bior3.3	bior4.4
PSNR	39.45	39.36	39.25	39.17	39.10	39.39	39.35	39.26
SSIM	0.9184	0.9175	0.9163	0.9160	0.9157	0.9176	0.9173	0.9161

4.1 Implement Details

We employ SIDD real noise dataset and synthetic noise with as our training dataset, respectively. Each image is cropped to a size of 128*128*3 as input, each epoch trained 96000 images and 50 epochs trained each time. We adopt Adam [38] as the optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.999$ while we set initial learning rate as 2e-4 and adopt the cosine annealing strategy [39] with the final learning rate as 5e-10. The hyper-parameter γ is set as 0.2 both for real-noise and synthetic noise training.

4.2 Ablation Study

Table 1 shows our ablation study on the impact of different architectural components including wavelet transform, SCAB, noise map and variable depth when testing on the SIDD validation dataset.

Compared with the spatial domain image as input, the wavelet transform can simultaneously assemble frequency domain information and spatial structure information for learning and improves the network performance by 0.16dB. Wavelet decomposition essentially regards the wavelet basis function as a filter to decompose the image into different frequency bands. This operation can also be learned by the neural network without wavelet transformation. However, the Haar wavelet with orthogonality, compactly supported and symmetry provides a certain prior constraint for the image to help the network pay attention to the frequency domain information of the image during training process.

In order to choose a better wavelet basis function for wavelet transform, we compare the performance of Haar, Daubechies and Biorthogonal wavelets on

(a) SIDD (b) DND (c) LIVE1

Fig. 4: The noise map predicted by our FAN on SIDD, DND and LIVE1 dataset. The noisy image and noise map of one typical image of SIDD validation dataset, one of DND dataset and one of LIVE1 dataset with $\sigma = 50$.

our proposed FAN and the result is shown on Table 2. Daubechies wavelet is a continuous orthogonal compactly supported wavelet and suitable for whitening wavelet coefficients [40]. However, Daubechies wavelet is an asymmetric wavelet which will cause the phase distortion of images during wavelet reconstruction and the performance of Daubechies wavelet gets worse as the approximation order increases. In order to construct "linear phase" filters, biorthogonal wavelets are proposed to construct compactly supported symmetric wavelets [41]. Note that the biorthogonal wavelet does not perform better on the denoising task compared with Haar, which indicates that the strictly orthogonal wavelet basis function can decompose the image into an orthogonal space and is effective in eliminating the correlation of the image signal.

We also experiment on applying the wavelet decomposition of the chroma layer to concatenate that of the luminance layer as input and introducing the multi-resolution decomposition of wavelet for wavelet transform respectively and we observe these operations perform slightly worse than not using them. We conclude that chrominance noise is eventually greater than luminance noise and distributed in lower frequency band which results in insufficient seperation of chromiannce noise. In the case of the same number of convolution kernels, increasing the width of the network input will affect the ability of the network to extract features and limit the denoising performance.

Furthermore, considering that the distribution of luminance noise and chrminance noise is inconsistent, we also train a FAN_{dual} with two *De-Net*s for the lumianance layer and the chroma layer respectively. However, the final observation is that FAN_{dual} and FAN achieve close denoising results on the test dataset

Table 3: The PSNR(dB) results about AWGN removal of three datasets

Sigma	Datasets	CBM3D [19]	WNNM [20]	NCSR [23]	MWCNN [25]	DnCNN [8]	MemNet [10]	FFDNet [9]	VDN [26]	FAN (Ours)
σ=15	LIVE1	33.08	31.70	31.46	32.33	33.72	33.84	33.96	33.94	**34.16**
	Set5	33.90	32.92	32.57	33.84	34.04	34.18	34.31	34.34	**35.01**
	CBSD68	32.89	31.27	30.84	31.86	33.87	33.76	33.85	33.90	**34.08**
σ=25	LIVE1	30.39	29.15	29.05	31.56	31.23	31.26	31.37	31.50	**31.66**
	Set5	31.34	30.61	30.33	29.84	31.88	31.98	32.10	32.24	**32.59**
	CBSD68	30.13	28.62	28.35	29.41	31.22	31.17	31.21	31.35	**31.46**
σ=50	LIVE1	27.13	26.07	26.06	26.86	27.95	27.99	28.10	28.36	**28.44**
	Set5	28.25	27.58	27.20	28.61	28.95	29.10	29.25	29.47	**29.51**
	CBSD68	26.94	25.86	25.75	26.54	27.91	27.91	27.95	28.19	**28.24**

while parameters of FAN_{dual} is almost twice that of FAN so that we still adopt FAN for testing.

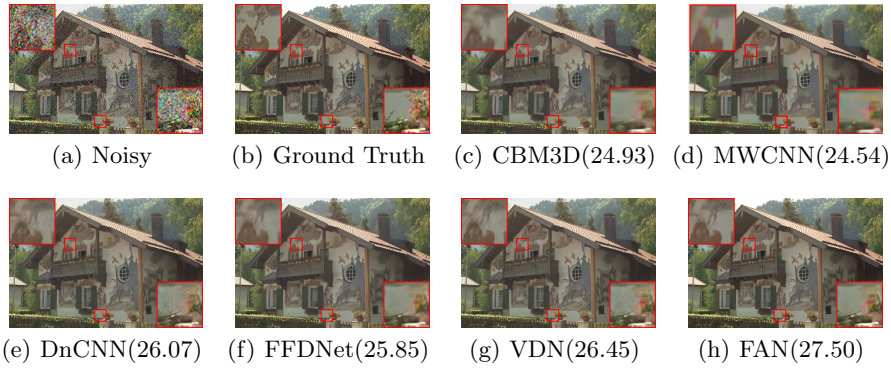
We develop SCAB to make the network pay more attention to the main features of the image that are more relevant to the surroundings and to some extent SCAB module aims at enhancing the self-similarity of the image. Experiments also show that the attention module has achieve 0.32dB improvement on the overall network.

For blind denoising tasks, the estimation of the noise level map can make the network adaptive to different scenes with various noise distribution and help the network adjust the ability of extract features to remove noise. Compared with no noise map, FAN with noise map can reach a faster convergence when training and better visual results. Figure 4 shows noise level map of real-world noise and synthetic noise predicted by our proposed FAN including a typical image of SIDD validation dataset, another of DND dataset and an image of LIVE1 with $\sigma = 50$. It can be seen that the noise intensity information in the predicted map of real-world noise is more hierarchical because of signal-dependent noise and signal-independent noise while noise map of synthetic noise has a smooth distribution in each channel.

4.3 Experiments on Synthetic Noise

We collect 4744 pictures from Waterloo Exploration Database [42] and cropped them into $N = 20 * 4744$ pictures with the size of 128*128*3 for training. We adopt three common image restoration datasets as test datasets to evaluate the performance of different competing methods. There is still another challenge to make noise contribution of synthetic noise as close as real-noise and it is unfair to adopt one noise model for training when other methods train with another. Considering that most denoising algorithms use additive white Gaussian noise(AWGN) for the assumption of synthetic noise, in order to compare the different methods more fairly, our noise model is defined as following,

$$Y = X + n, n \sim \mathcal{N}(0, \sigma^2) \tag{7}$$



(a) Noisy (b) Ground Truth (c) CBM3D(24.93) (d) MWCNN(24.54)

(e) DnCNN(26.07) (f) FFDNet(25.85) (g) VDN(26.45) (h) FAN(27.50)

Fig. 5: Denoised results on one typical image in LIVE1 dataset with $\sigma = 50$ by different methods

Table 3 lists the average PSNR results of different competing methods on three testing datasets and Fig5 shows some denoising results of different method on one typical images of CBSD68 dataset when $\sigma = 50$. From Table 3 and Fig 5, it can be easily observed that: 1) Although the CBM3D based on self-similarity model is a stable traditional method, there are color artifact and left noise in the denoising results while most methods based on neural network performs better at this problem. 2) Some outstanding CNN denoisers like FFDNet easily over-smooth the images and are unable to preserve edge information while our proposed FAN can retain more details by variable depth design. 3) For noise in the flat regions, our model can also deal with it well because SWT can avoid the pseudo-Gibbs effect and make the image look more natural and real.

4.4 Experiments on Real-World Noise

We select two real noise datasets DND [43] and SIDD [44] to evaluate the denoising performance of FAN. DND collected 50 images from 50 scenes captured by four consumer cameras. The carefully post-processing of low-ISO images results in clean images, but it does not provide ground truth while PSNR / SSIM results of denoised images through the online server. SIDD is another real noise dataset which contains 320 image pairs of 10 scenes taken by 5 smartphones. Clean Images are obtained by a series of ISP pipeline processing performing on multiple images of the same scene. SIDD uses some unpublished image pairs as a test set for verifying the performance of the denoising algorithms online.

Table 4 lists the denoising performance of different competing methods shown on the SIDD benchmark. It can be seen that our FAN has obvious advantages compared of model-driven traditional denoising algorithm and data-driven neural network denoising algorithm. In view of the fact that the noise type of CBDNet training data is inconsistent with SIDD, for fairly we also compared the results of CBDNet training on the SIDD training set [26] and over 0.64dB. Specially, FAN achieves 0.62dB higher than RIDNet and 0.06dB higher than VDN

Table 4: The comparison results of other methods on SIDD benchmark [44]

Method	DnCNN [8]	TNRD [45]	CBM3D [19]	NLM [3]	WNNM [20]	KSVD [22]	CBDNet [16]	CBDNet* [16]	RIDNet [28]	VDN [26]	FAN (Ours)
PSNR	23.66	24.73	25.65	26.75	25.78	26.88	33.28	38.68	38.71	39.26	**39.33**
SSIM	0.583	0.643	0.685	0.699	0.809	0.842	0.868	0.901	0.914	0.955	**0.956**

(a) Noisy (b) Ground Truth (c) Noisy (d) Ground Truth

(e) CBM3D(29.53) (f) CBDNet(25.47) (g) CBM3D(27.14) (h) CBDNet(21.13)

(i) VDN(34.46) (j) FAN(34.67) (k) VDN(35.58) (l) FAN(36.29)

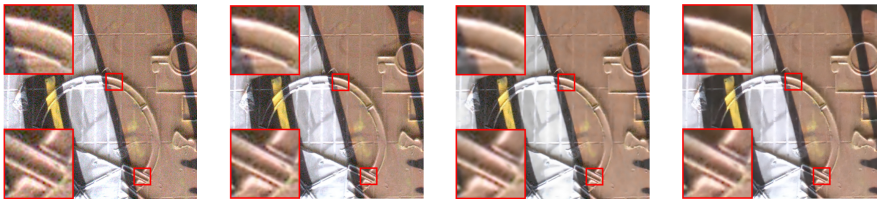
Fig. 6: Denoised results(PSNR: dB) on two typical images in SIDD validation dataset by different methods

on the SIDD test dataset. As shown in Fig 6, we compare our results with other competing algorithms. In the first example, our proposed FAN performs well in the smooth region and the color boundary distinction while avoiding speckled structures and chroma artifacts. For another example, the image denoised by VDN has lattice-like artifact on the upper and left sides while the denoising result of our FAN can maintain the spatial smoothness of the homogeneous regions and keep fine textural details.

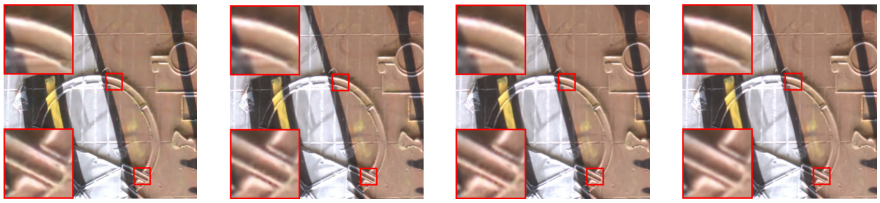
Table 5 summarizes the quantitative comparison of different methods on DND benchmark. It is easy to be seen that our proposed FAN surpasses other

Table 5: The comparison results of other methods on DND benchmark [43]

Method	CBM3D [19]	WNNM [20]	KSVD [22]	MCWNNM [25]	FFDNet [9]	DnCNN+ [8]	TWSC [24]	CBDNet [16]	RIDNet [28]	VDN [26]	FAN (Ours)
PSNR	34.51	34.67	36.49	37.38	37.61	37.90	37.96	38.06	39.26	39.38	**39.41**
SSIM	0.8507	0.8507	0.8978	0.9294	0.9415	0.9430	0.9416	0.9421	**0.9528**	0.9518	0.9507



(a) Noisy (b) CBM3D(30.91) (c) KSVD(31.73) (d) DnCNN(33.29)



(e) CBDNet(33.62) (f) RIDNet(34.09) (g) VDN(33.89) (h) FAN(34.23)

Fig. 7: Denoised results(PSNR: dB) on one typical image in DND benchmark by different methods

methods, especially has a performance gain of 1.35dB and 0.15dB compared to CBDNet and RIDNet respectively. Fig 7 shows some visualizing results of the comparison between FAN and other competitive algorithms on DND benchmark. We can see that our FAN can make the image smoother and retain perceptually-pleasing texture details.

5 Conclusion

In this paper, we propose frequency attention network for blind real noise removal which exploits spectral information and structural information of images and employ the attention mechanism to enhance the feature maps. Abundant ablation experiments indicate that Haar wavelet basis function which satisfies symmetry, orthogonality and compactly supported characteristics at the same time can achieve the best performance on our proposed FAN. Comprehensive evaluations on different noise distribution cases demonstrate the superiority and effectiveness of our method for image restoration tasks. Our method can also be implemented on other low-level tasks including super-resolution and deblurring.

References

1. Liu, C., Szeliski, R., Bing Kang, S., Zitnick, C.L., Freeman, W.T.: Automatic estimation and removal of noise from a single image. IEEE Transactions on Pattern Analysis Machine Intelligence **30** (2007) 299–314
2. Wang, W., Chen, X., Yang, C., Li, X., Hu, X., Yue, T.: Enhancing low light videos by exploring high sensitivity camera noise. In: Proceedings of the IEEE International Conference on Computer Vision. (2019) 4111–4119
3. Buades, A., Coll, B., Morel, J.M.: A non-local algorithm for image denoising. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). Volume 2., IEEE (2005) 60–65
4. Dabov, K., Foi, A., Katkovnik, V., Egiazarian, K.: Image denoising by sparse 3-d transform-domain collaborative filtering. IEEE Transactions on image processing **16** (2007) 2080–2095
5. Lebrun, M., Buades, A., Morel, J.M.: A nonlocal bayesian image denoising algorithm. SIAM Journal on Imaging Sciences **6** (2013) 1665–1688
6. Donoho, D.L., Johnstone, I.M.: Adapting to unknown smoothness via wavelet shrinkage. Journal of the american statistical association **90** (1995) 1200–1224
7. Yu, G., Sapiro, G.: Dct image denoising: a simple and effective image denoising algorithm. Image Processing On Line **1** (2011) 292–296
8. Zhang, K., Zuo, W., Chen, Y., Meng, D., Zhang, L.: Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. IEEE Transactions on Image Processing **26** (2017) 3142–3155
9. Zhang, K., Zuo, W., Zhang, L.: Ffdnet: Toward a fast and flexible solution for cnn-based image denoising. IEEE Transactions on Image Processing **27** (2018) 4608–4622
10. Tai, Y., Yang, J., Liu, X., Xu, C.: Memnet: A persistent memory network for image restoration. In: Proceedings of the IEEE international conference on computer vision. (2017) 4539–4547
11. Lehtinen, J., Munkberg, J., Hasselgren, J., Laine, S., Karras, T., Aittala, M., Aila, T.: Noise2noise: Learning image restoration without clean data. In: International Conference on Machine Learning. (2018) 2965–2974
12. Zhang, Y., Li, K., Li, K., Zhong, B., Fu, Y.: Residual non-local attention networks for image restoration. In: International Conference on Learning Representations. (2019)
13. Jia, X., Liu, S., Feng, X., Zhang, L.: Focnet: A fractional optimal control network for image denoising. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2019) 6054–6063
14. Yin, D., Lopes, R.G., Shlens, J., Cubuk, E.D., Gilmer, J.: A fourier perspective on model robustness in computer vision. In: Advances in Neural Information Processing Systems. (2019) 13255–13265
15. Kim, J., Lee, S.: Deep learning of human visual sensitivity in image quality assessment framework. In: Proceedings of the IEEE conference on computer vision and pattern recognition. (2017) 1676–1684
16. Guo, S., Yan, Z., Zhang, K., Zuo, W., Zhang, L.: Toward convolutional blind denoising of real photographs. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2019) 1712–1722
17. Foi, A., Katkovnik, V., Egiazarian, K.: Pointwise shape-adaptive dct for high-quality denoising and deblocking of grayscale and color images. IEEE transactions on image processing **16** (2007) 1395–1411

18. Chang, S.G., Yu, B., Vetterli, M.: Adaptive wavelet thresholding for image denoising and compression. IEEE transactions on image processing **9** (2000) 1532–1546
19. Dabov, K., Foi, A., Katkovnik, V., Egiazarian, K.: Color image denoising via sparse 3d collaborative filtering with grouping constraint in luminance-chrominance space. In: 2007 IEEE International Conference on Image Processing. Volume 1., IEEE (2007) I–313
20. Gu, S., Zhang, L., Zuo, W., Feng, X.: Weighted nuclear norm minimization with application to image denoising. In: Proceedings of the IEEE conference on computer vision and pattern recognition. (2014) 2862–2869
21. Xu, J., Zhang, L., Zhang, D., Feng, X.: Multi-channel weighted nuclear norm minimization for real color image denoising. In: Proceedings of the IEEE International Conference on Computer Vision. (2017) 1096–1104
22. Aharon, M., Elad, M., Bruckstein, A.: K-svd: An algorithm for designing overcomplete dictionaries for sparse representation. IEEE Transactions on signal processing **54** (2006) 4311–4322
23. Dong, W., Zhang, L., Shi, G., Li, X.: Nonlocally centralized sparse representation for image restoration. IEEE transactions on Image Processing **22** (2012) 1620–1630
24. Xu, J., Zhang, L., Zhang, D.: A trilateral weighted sparse coding scheme for real-world image denoising. In: Proceedings of the European Conference on Computer Vision (ECCV). (2018) 20–36
25. Liu, P., Zhang, H., Zhang, K., Lin, L., Zuo, W.: Multi-level wavelet-cnn for image restoration. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. (2018) 773–782
26. Yue, Z., Yong, H., Zhao, Q., Meng, D., Zhang, L.: Variational denoising network: Toward blind noise modeling and removal. In: Advances in Neural Information Processing Systems. (2019) 1688–1699
27. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention, Springer (2015) 234–241
28. Anwar, S., Barnes, N.: Real image denoising with feature attention. In: Proceedings of the IEEE International Conference on Computer Vision. (2019) 3155–3164
29. Mildenhall, B., Barron, J.T., Chen, J., Sharlet, D., Ng, R., Carroll, R.: Burst denoising with kernel prediction networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2018) 2502–2510
30. Xu, X., Li, M., Sun, W.: Learning deformable kernels for image and video denoising. ArXiv **abs/1904.06903** (2019)
31. Gueguen, L., Sergeev, A., Kadlec, B., Liu, R., Yosinski, J.: Faster neural networks straight from jpeg. In: Advances in Neural Information Processing Systems. (2018) 3933–3944
32. Kai Xu, Minghai Qin, F.S.Y.W.Y.K.C.F.R.: Learning in the frequency domain. In: 2020 IEEE Computer Conference on Computer Vision and Pattern Recognition (CVPR'20), IEEE (2020)
33. Chou, C.H., Li, Y.C.: A perceptually tuned subband image coder based on the measure of just-noticeable-distortion profile. IEEE Transactions on circuits and systems for video technology **5** (1995) 467–476
34. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. (2018) 7132–7141
35. Liu, D., Wen, B., Fan, Y., Loy, C.C., Huang, T.S.: Non-local recurrent network for image restoration. In: Advances in Neural Information Processing Systems. (2018) 1673–1682

36. Woo, S., Park, J., Lee, J.Y., So Kweon, I.: Cbam: Convolutional block attention module. In: Proceedings of the European Conference on Computer Vision (ECCV). (2018) 3–19
37. Tomasi, C., Manduchi, R.: Bilateral filtering for gray and color images. In: Sixth international conference on computer vision (IEEE Cat. No. 98CH36271), IEEE (1998) 839–846
38. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. CoRR **abs/1412.6980** (2015)
39. Loshchilov, I., Hutter, F.: Sgdr: Stochastic gradient descent with warm restarts. In: ICLR. (2017)
40. Po, D.Y., Do, M.N.: Directional multiscale modeling of images using the contourlet transform. IEEE Transactions on image processing **15** (2006) 1610–1620
41. Cohen, A., Daubechies, I., Feauveau, J.C.: Biorthogonal bases of compactly supported wavelets. Communications on pure and applied mathematics **45** (1992) 485–560
42. Ma, K., Duanmu, Z., Wu, Q., Wang, Z., Yong, H., Li, H., Zhang, L.: Waterloo exploration database: New challenges for image quality assessment models. IEEE Transactions on Image Processing **26** (2016) 1004–1016
43. Plotz, T., Roth, S.: Benchmarking denoising algorithms with real photographs. In: Proceedings of the IEEE conference on computer vision and pattern recognition. (2017) 1586–1595
44. Abdelhamed, A., Lin, S., Brown, M.S.: A high-quality denoising dataset for smartphone cameras. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2018) 1692–1700
45. Chen, Y., Pock, T.: Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration. IEEE transactions on pattern analysis and machine intelligence **39** (2016) 1256–1272