Neural Networks 124 (2020) 117–129

Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

Attention-guided CNN for image denoising

Chunwei Tian [a], Yong Xu [a,b,*], Zuoyong Li [c,**], Wangmeng Zuo [d], Lunke Fei [e], Hong Liu [f]

[a] Bio-Computing Research Center, Harbin Institute of Technology, Shenzhen, 518055, Guangdong, China
[b] Peng Cheng Laboratory, Shenzhen, 518055, Guangdong, China
[c] Fujian Provincial Key Laboratory of Information Processing and Intelligent Control, Minjiang University, Fuzhou, 350121, Fujian, China
[d] School of Computer Science and Technology, Harbin Institute of Technology, Harbin, 150001, Heilongjiang, China
[e] School of Computer Science and Technology, Guangdong University of Technology, Guangzhou, 510006, Guangdong, China
[f] Engineering Lab on Intelligent Perception for Internet of Things, Shenzhen Graduate School, Peking University, Shenzhen, 518055, Guangdong, China

ARTICLE INFO

Article history:
Received 13 August 2019
Received in revised form 15 November 2019
Accepted 23 December 2019
Available online 7 January 2020

Keywords:
Image denoising
CNN
Sparse block
Feature enhancement block
Attention block

ABSTRACT

Deep convolutional neural networks (CNNs) have attracted considerable interest in low-level computer vision. Researches are usually devoted to improving the performance via very deep CNNs. However, as the depth increases, influences of the shallow layers on deep layers are weakened. Inspired by the fact, we propose an attention-guided denoising convolutional neural network (ADNet), mainly including a sparse block (SB), a feature enhancement block (FEB), an attention block (AB) and a reconstruction block (RB) for image denoising. Specifically, the SB makes a tradeoff between performance and efficiency by using dilated and common convolutions to remove the noise. The FEB integrates global and local features information via a long path to enhance the expressive ability of the denoising model. The AB is used to finely extract the noise information hidden in the complex background, which is very effective for complex noisy images, especially real noisy images and bind denoising. Also, the FEB is integrated with the AB to improve the efficiency and reduce the complexity for training a denoising model. Finally, a RB aims to construct the clean image through the obtained noise mapping and the given noisy image. Additionally, comprehensive experiments show that the proposed ADNet performs very well in three tasks (i.e. synthetic and real noisy images, and blind denoising) in terms of both quantitative and qualitative evaluations. The code of ADNet is accessible at https://github.com/hellloxiaotian/ADNet.

1. Introduction

Image denoising is a typical problem for low-level vision applications in the real world (Xu, Li, Liang, Zhang, & Zhang, 2018). Since image denoising has ill-posed nature and important realistic significance, it has become a hot topic in the field of image processing and computer vision (Xu, Zhang, & Zhang, 2018). Specifically, the typical image denoising methods (Liu, Zhang, Zhang, Lin, & Zuo, 2018) usually apply the model $y = x + \upsilon$ to recover the latent clean image, x, where y and υ represent the given noisy image and additive white Gaussian noise with standard deviation, σ, respectively. From Bayesian of view, it is known that the image prior with a given likelihood is key to overcome the ill-posed denoising problem. Inspired by that fact, a lot of methods based on image degradation model have been developed to suppress the noise from noisy images in the past 20 years (Dabov, Foi, Katkovnik, & Egiazarian, 2007). Specifically, sparse methods were utilized to improve the performance and reduce the computational cost (Zha, et al., 2017). Dong et al. used nonlocal self-similarity to compute the sparse coefficients, then, centralized these obtained coefficients to recover the clean image (Dong, Zhang, Shi, & Li, 2012). Taking the efficiency into account, the dictionary learning embedded sparse method was employed to remove the noise from the given image (Mairal, Bach, Ponce, Sapiro, & Zisserman, 2009). To deal with the practical application, Gu, Zhang, Zuo, and Feng (2014) extracted more information under different conditions as weights to address the nuclear norm minimization problem for image denoising. Also, using image detailed information to obtain the latent clean image was also a good tool in image denoising (Barbu, 2009). Based on this idea, a Markov random field (MRF) consolidated wavelet transform to smooth the noisy image and filter the noise (Malfait & Roose, 1997). Additionally, there were other excellent methods such as total-variation (TV) (Chambolle, 2004) for image denoising.

Although the methods above have obtained competitive performance in image denoising, most of these methods are faced with three major problems: (1) manually selected parameters, (2) complex optimized algorithms, and (3) certain noise task.

* Corresponding author at: Bio-Computing Research Center, Harbin Institute of Technology, Shenzhen, 518055, Guangdong, China.
** Corresponding author.
E-mail addresses: yongxu@ymail.com (Y. Xu), fzulzytdq@126.com (Z. Li).

https://doi.org/10.1016/j.neunet.2019.12.024
0893-6080/

Due to strong representation ability, deep learning methods have become the dominant techniques to solve these drawbacks. Specifically, deep CNNs with flexible modular architectures are also very popular for image applications, especially image denoising (Guo, Yan, Zhang, Zuo, & Zhang, 2019; Tian, et al., 2019). Zhang, Zuo, Chen, Meng, and Zhang (2017) first proposed a denoising convolutional neural network (DnCNN), comprising residual learning (RL) (He, Zhang, Ren, & Sun, 2016) and batch normalization (BN) (Ioffe & Szegedy, 2015) to recover the corrupted image. It is noted that the proposed DnCNN can use a model to deal with multi denoising tasks, such as Gaussian noisy, JPEG and low-resolution images. However, there is a challenge that with growth of the depth, deep networks may result in performance degradation. For resolving the problem, iterative and skip connection operations in CNNs were developed for image restoration (Zhang, Tian, Kong, Zhong, & Fu, 2018). For example, a deep recursive residual network (DRRN) utilized global and local RL techniques to enhance the representation ability of the trained model in image restoration (Tai, Yang, & Liu, 2017a). Similarly, a deeply-recursive convolution network (DRCN) fused information of each layer into the final layer by utilizing the RL technique to improve the denoising performance (Kim, Kwon Lee, & Mu Lee, 2016). Also, increasing the width of the network is useful to mine more information in image denoising (Liu & Fang, 2017). Additionally, combining the prior and CNN is also very beneficial to accelerate the training in image restoration (Liu, Sun, Xu, & Kamilov, 2019). Although the proposed methods above have visual effects on image restoration, they still suffer from the following drawbacks: (1) some of the methods such as residual dense network (Zhang et al., 2018) have high computational cost and memory consumption. (2) Some of these deep networks do not make full use of the effects from shallow layers on the deep layers. (3) Most of these methods neglect that complex background can hide some key features.

In this paper, we propose a denoising network as well as ADNet, which is composed of a SB, a FEB, an AB and a RB. Specifically, the SB based on dilated and common convolutions is used to extract useful features from the given noisy image. That can improve both the denoising performance and efficiency. Then, the FEB fuses global and local features via a long path to enhance the expressive ability of denoising model. It is noted that complex background from the given image or video is easier to hide the features, which increases the difficulty of training (Li, et al., 2019). Thus, we use an attention mechanism to guide a CNN for image denoising. That is, the AB can finely extract noise information hidden in the complex background to deal with complex noisy images, i.e. real noisy image and blind denoising. Actually, the FEB is also utilized to consolidate AB for improving the efficiency and reducing the complexity of the denoising model. Finally, the RB is used to construct the clean image. Additionally, extensive experiments illustrate that our ADNet outperforms state-of-the-art denoising methods such as block-matching and 3-D filtering (BM3D) (Dabov et al., 2007) and DnCNN in terms of both quantitative and qualitative analysis. Besides this, the model is small, which is suitable to practical applications, i.e. mobile phones and cameras.

The main contributions of this work can be summarized as follows:

(1) The SB comprising dilated and common convolutions is proposed to improve the denoising performance and the efficiency. Also, it is used to reduce the depth.

(2) The FEB uses a long path to fuse information from shallow and deep layers for enhancing the expressive ability of the denoising model.

(3) The AB is used to deeply mine noise information hidden in the complex background from the given noisy images, which is very useful to handle complex noisy images, such as real noisy image and blind denoising.

(4) The FEB is integrated with AB to improve the efficiency and reduce the complexity for training a denoising model.

(5) The ADNet is very superior to the state-of-the-arts on six benchmark datasets for synthetic and real noisy images, and blind denoising.

The remainder of this paper is organized as follows. Section 2 offers the related work of deep CNN on image denoising, dilated convolution and attention mechanism for image applications. Section 3 provides the proposed method. Section 4 shows the extensive experiments and results of the proposed method for image denoising. Section 5 presents the conclusion.

2. Related work

2.1. Deep CNNs for image denoising

Flexible plug-in components in CNNs have strong abilities to treat with different applications, e.g. object detection (Wang, Wang, Gao, Li, & Zuo, 2018b) and image restoration (Chen, Xiong, Tian, & Wu, 2018; Ren, Zuo, Hu, Zhu, & Meng, 2019). Specifically, most of deep CNNs are designed in terms of improving denoising performance and efficiency.

For promoting denoising performance, enlarging the receptive fields of CNNs is the most popular method. A dilated dense fusion network (DDFN) (Chen et al., 2018) combined dilated convolutions and a feed-forward way to facilitate more useful features for filtering the noise. Further, increasing the number of training samples is also a good choice to suppress the noise. For example, a generative adversarial network (GAN) (Tripathi, Lipton, & Nguyen, 2018) comprising a generative network and a discriminative network was used to generate virtual samples, according to the training samples in image denoising. The two sub-networks ruled the game theory. That is, the generative network was in charge of increasing training samples. The other was utilized to verify the reliability of the generated samples. Additionally, using the skip connection operation in CNNs to enhance the expressive ability of the denoising model was effective. For example, a very deep persistent memory network (MemNet) (Tai, Yang, Liu, & Xu, 2017b) was proposed to enhance the effects of shallow layers on deep layers through recursive and gate units. A very deep residual encoder–decoder network30 (RED30) (Mao, Shen, & Yang, 2016) utilized the skip connections for between convolutions and deconvolutions to eliminate the noise or corruption. Further, the combination of CNN and signal processing technique was beneficial to image denoising. A multi-level wavelet CNN (MWCNN) (Liu et al., 2018) fused the wavelet transform and a U-Net to extract detailed information of the corrupted image.

In improving the efficiency of denoising task, deep CNNs can be regarded as a modular part to plug into some classical optimized methods for recovering the latent clean image, which was very effective to cope with the noisy image (Zhang, Zuo, Gu, & Zhang, 2017). An image restoration CNN (IRCNN) (Zhang et al., 2017) integrated a CNN and half quadratic splitting (HQS) to accelerate the training speed for image denoising. Actually, reducing the number of parameters was very popular to improve the efficiency from the training of the denoising model. Additionally, small filter sizes have been adopted to reduce the computational cost and memory consumption. An information distillation network (IDN) (Hui, Wang, & Gao, 2018) including feature extraction, stacked information distillation and reconstruction blocks employed partial filters of size 1×1 to distill more useful information for image super-resolution. That is also very suitable to image denoising. Also, combining the nature of the given task and CNN can ease the difficulty of training.

For example, a fast and flexible denoising convolutional neural network (FFDNet) (Zhang, Zuo, & Zhang, 2018) used the noisy image patches and the noise mapping patches as input of the network to efficiently tackle blind denoising. Further, separating the convolutions can improve the denoising efficiency (Cho & Kang, 2018). From these descriptions, we can see that deep CNNs are very competing to both performance and efficiency in image denoising. Motivated by the fact, we use a deep CNN for image denoising.

2.2. Dilated convolution

It is known that the context information is important to reconstruct the corrupted pixel point for image denoising (Yu & Koltun, 2015). Specifically, enlarging the receptive field size is the common way to capture more context information in CNNs. In general, the CNNs have two ways to enlarge the receptive field, i.e. increasing the depth and width of the deep networks. However, the first way with deeper CNNs suffer from difficulty of training. The second way may be involved into more parameters and increase the complexity of the denoising model. Inspired by the fact, dilated convolutions (Yu & Koltun, 2015) are developed. Here we use an example to show the principle of dilated convolutions as follows. A dilated convolution with dilated factor of 2 has the receptive field size of $(4n+1)\times(4n+1)$, where n denotes the depth of a given deep CNN. We assume that n is 10, the receptive field size of the given CNN is 41×41. Thus, it can map context information of 41×41. Moreover, it has the same effect as 20-layer CNN with standard convolutions of 3×3, dilated factor of 1. Thus, dilated convolutions make a tradeoff between increasing the depth and width of deep CNNs. Taking these reasons into account, some scholars used dilated convolutions in CNN for image processing. To resolve the difficulty of training, Peng, et al. (2019) proposed to use the symmetric skip connections and dilated convolutions rather than batch normalization (Ioffe & Szegedy, 2015) to improve the denoising speed and performance. To reduce the complexity, Wang, Sun, and Hu (2017) combined dilated convolutions and the BN technique to reduce the computational cost in image denoising. To solve the artifact removal, Zhang, Yang, Hu, and Liu (2018) utilized multi-scale loss and dilated convolutions to eliminate the effect of artifact on JPEG compression task. These methods verify the effectiveness of dilated convolutions in image applications. Thus, we propose a sparse mechanism based on dilated convolutions to reduce the complexity and improve the denoising performance.

2.3. Attention mechanism

Extracting and choosing suitable features are very important for image processing applications (Du, Wei, & Liu, 2019; Li, Lu, Zhang, You, & Zhang, 2019; Liang, Zhang, Lu, Guo, & Luo, 2019; Zhang, et al., 2018; Lu, et al., 2015). However, the given image with complex background is challenge to extract features (Li, et al., 2019). An attention mechanism based on this reason is proposed (Zhu, Wu, Zou, & Yan, 2018). There are usually two kinds of attention mechanisms: attentions from different sub-networks and different branches in the same network.

The first method emphasizes the effect of different perspectives to obtain the features of background. For example, Zhu et al. (2018) extracted the saliency features via different sub-networks for video tracking.

The second method focuses on the effects of different branches in the same network to guide previous stage for image application. Wang et al. (2018b) used higher-order statistics to direct the deep network for improving the performance and efficiency of object detection. Specifically, the proposed CNN utilized latter stage to offer complementary information for the previous network. The two methods can quickly find the object. However, the first method referred to multi sub-networks, it may increase the computational cost of training. Moreover, there is little research to study the attention mechanism on image denoising, especially the real noisy image. Inspired by these reasons, we integrate the second method into CNN for image denoising in this paper.

3. Proposed method

In this section, we introduce the proposed denoising network, ADNet, which is composed of a SB, a FEB, an AB and a RB. Specifically, the design of network architecture of ADNet follows between the performance and efficiency for image denoising. For improving the performance, we use three blocks (i.e. a SB, a FEB and an AB) to remove the noise from different perspectives. The SB uses the dilated and standard convolutions to enlarge the receptive field size for improving denoising performance. The FEB integrates the global and local features of ADNet via a long path to enhance the expressive ability in image denoising. The AB can quickly capture the key noisy features hidden in the complex background for complex noisy tasks, such as real noisy image and blind denoising. For improving the efficiency, there are three phases: firstly, the SB facilitates the ADNet to obtain shallow network architecture. Secondly, the FEB compresses the output from the sixteenth layer as c. Thirdly, the AB uses a convolutional filter of 1×1 to reduce the number of parameters. Further, we will introduce these techniques in later sub-sections.

3.1. Network architecture

The proposed 17-layer ADNet consists of four blocks, a SB, a FEB, an AB and a RB as shown in Fig. 1. The 12-layer sparse block is used to enhance the performance and efficiency in image denoising. We assume that I_N and I_R denote the input noisy image and the predicted residual image (also referred to as noise mapping image), respectively. The sparse block is shown as

$$O_{SB} = f_{SB}(I_N), \tag{1}$$

where f_{SB} represents the function of the SB. O_{SB} is the output of the SB and it serves the FEB. The 4-layer FEB makes full use of global and local features of ADNet to enhance the expressive ability in image denoising, where the global features are the input noisy image, I_N. O_{SB} is regarded as local features. The implementations of this block can be transformed as the following formula:

$$O_{FEB} = f_{FEB}(I_N, O_{SB}), \tag{2}$$

where f_{FEB} and O_{FEB} denote the function and output of the FEB, respectively. The O_{FEB} is applied on the AB. It is noted that complex background from the given image or video might be more easier to hide the features, which increases the difficulty of extracting key features in the training process (Li, et al., 2019). To overcome this problem, the 1-layer AB is proposed to predict the noise. The AB can be expressed as

$$I_R = f_{AB}(O_{FEB}), \tag{3}$$

where f_{AB} and I_R stand for the function and output of the AB, respectively. Specifically, I_R is used as the input of the RB. The RB is utilized to reconstruct the clean image via a RL technique. This process is illustrated as Eq. (4).

$$\begin{aligned} I_{LC} &= I_N - I_R \\ &= I_N - f_{AB}(f_{FEB}(I_N, f_{SB}(I_N))), \\ &= I_N - f_{ADNet}(I_N) \end{aligned} \tag{4}$$

where f_{ADNet} is the function of ADNet to predict the residual image and I_{LC} is the latent clean image. Further, the ADNet is optimized by applying the loss function as illustrated in Section 3.2.

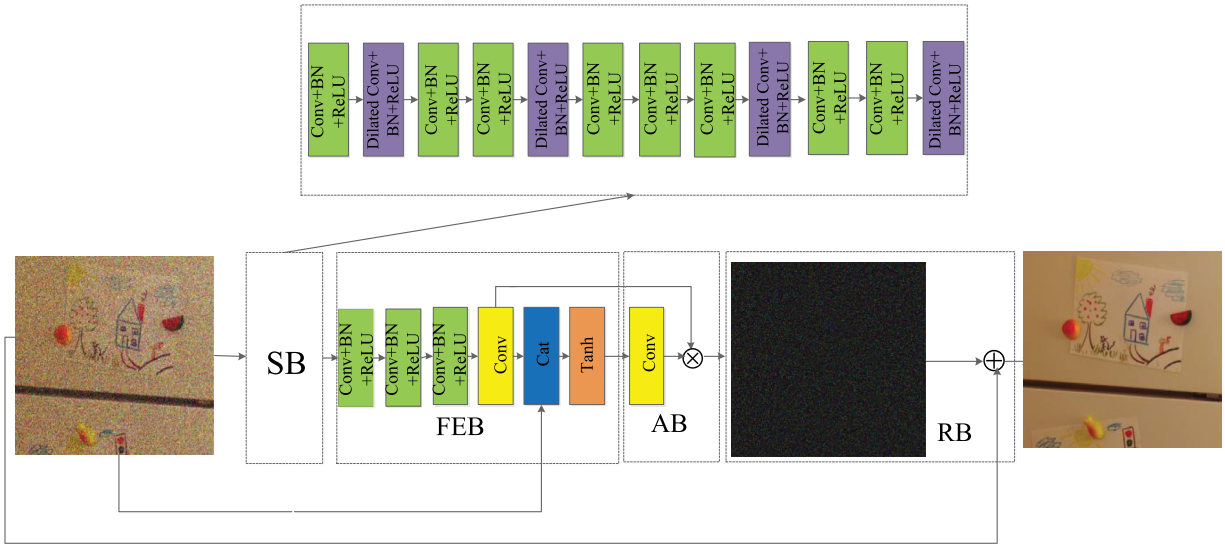

Fig. 1. Network architecture of the proposed ADNet.

3.2. Loss function

The proposed ADNet is trained by the degradation equation $y = x + \upsilon$. It is known that the ADNet is used to predict the residual image, υ via $\upsilon = y - x$. Thus, we use the given pair $\{I_C^i, I_N^i\}_{i=1}^N$ and the mean square error (MSE) (Ephraim & Malah, 1984) to train a denoising model, where I_C^i and I_N^i denote the i-th given clean and noisy images, respectively. The implementations of this process can be formulated as

$$l(\theta) = \frac{1}{2N}\sum_{i=1}^{N}\left\|f_{ADNet}(I_N^i) - (I_N^i - I_C^i)\right\|^2, \tag{5}$$

where θ stands for parameters in training the denoising model.

3.3. Sparse block

It is known that sparsity is effective for image application (Tian, Zhang, Sun, Song, & Li, 2018). Motivated by that, we propose a sparse block (also named SB) to improve the denoising performance and efficiency. Also, it can reduce the depth of the denoising network, which is very beneficial to cut down the computational cost and memory consumption. Specifically, the proposed sparse block based on dilated and standard convolutions in CNN is different from the common sparse mechanism. That is, the 12-layer SB includes two types: dilated Conv+BN+ReLU and Conv+BN+ReLU. The dilated Conv+BN+ReLU denotes that a dilated convolution with dilated factor of 2, BN (Ioffe & Szegedy, 2015) and activation function, ReLU (Krizhevsky, Sutskever, & Hinton, 2012) are connected. The other is that a convolution, BN and ReLU are connected. The dilated Conv+BN+ReLU is placed at the second, fifth, ninth and twelfth layers in ADNet. It is noted that dilated convolutions can map more context (Yu & Koltun, 2015). Based on this idea, these layers can be regarded as high energy points. The Conv+BN+ReLU is set at the first, third, fourth, sixth, seventh, eighth, tenth and eleventh layers in ADNet, which can be treated as low energy points. Specifically, the convolution filter sizes of 1–12 layers are 3×3. The input of the first layer is c, which is the number of channels from the input noisy image. It is noted that if the given noisy image is color, c is 3; Otherwise, c is 1. The input and output of 2–12 layers are 64. The combination of several high energy points and some low energy points can be considered as the sparsity. Additionally, the sparse block uses less high-energy points rather than many high-energy points to capture more useful information, which can not only improve the denoising performance and efficiency of training, but also reduce the complexity. That can be proved in Section 4.3. The implementations of the sparse block are converted as a formula. First, we define some symbols as follows. D stands for the function of a dilated convolution. R and B represent the function of ReLU and BN, respectively. CBR is the function of Conv+BN+ReLU. Then, according to the previous descriptions, we use the following equation to show the SB.

$$O_{SB} = R(B(D(CBR(CBR(R(B(D(CBR(CBR(R(B(D(CBR(CBR(R(B(D(CBR(I_N)))))))))))))))))). \tag{6}$$

3.4. Feature enhancement block

It is known that very deep network might suffer from weaken influences from the shallow layers on the deep layers as the growth of depth (Tai et al., 2017b). For resolving this problem, a feature enhancement block (FEB) is proposed in ADNet for image denoising. The FEB fully utilizes the global and local features through a long path to mine more robust features, which is complementary with SB in handling the given noisy image. The 4-layer FEB consists of three types: Conv+BN+ReLU, Conv and Tanh, where the Tanh is an activate function (Malfliet & Hereman, 1996). The Conv+BN+ReLU is fitted at the 13–15 layers of the ADNet and their filter sizes are $64 \times 3 \times 3 \times 64$. The Conv is used for the sixteenth layer in ADNet and its filter size is $64 \times 3 \times 3 \times c$, where the setting of c is the same as Section 3.3. Finally, we use a concatenation operation to fuse the input noisy image and the output of the sixteenth layer to enhance the representation ability of the denoising model. Thus, the final output size is $64 \times 3 \times 3 \times 2c$. Additionally, a Tanh is used to convert the features obtained into nonlinearity. The descriptions of this process are explained as

$$O_{FEB} = T(Cat(C(CBR(CBR(CBR(O_{SB})))), I_N)), \tag{7}$$

where C, Cat and T are functions of a convolution, concatenation and Tanh, respectively. Cat is used to denote the function of concatenation in Fig. 1. Further, O_{FEB} is used for the AB.

Fig. 2. 12 images from Set12 dataset.

3.5. Attention block

As we all know, the complex background can easily hide the features for image and video applications, which can increase the difficulty of training (Li, et al., 2019). In this paper, we apply an AB to guide the CNN for training a denoising model. The AB uses the current stage to guide the previous stage for learning the noise information, which is very useful for unknown noisy images, i.e. blind denoising and real noisy images. The 1-layer AB only includes a Conv and its size is $2c \times 1 \times 1 \times c$, where c is the number of channels from the given corrupted image and its value is the same as Section 3.3. The AB exploits the following two steps to implement the attention mechanism. The first step uses a convolution of size 1×1 from the seventeenth layer to compress the obtained features into a vector as the weights for adjusting the previous stage, which can also improve the denoising efficiency. The second step utilizes obtained weights to multiply the output of the sixteenth layer for extracting more prominent noise features. The implemental procedure can be transformed as the following formulates.

$$O_t = C(O_{FEB}), \tag{8}$$

$$I_R = O_t \times O_{FEB}, \tag{9}$$

where O_t is the output of convolution from the seventeenth layer in ADNet. Specifically, $\otimes$ is used to express the multiplication operator in Fig. 1 rather than '$\times$' in Eq. (9).

3.6. Reconstruction block

As shown from Eq. (5), the ADNet is employed to predict the residual image. Thus, we use the RL technique in the reconstruction block (also treated as RB) to reconstruct the clean image. This process can be explained as Eq. (4).

4. Experiments

4.1. Datasets

4.1.1. Training datasets

We use 400 images with size of 180×180 from the Berkeley Segmentation Dataset (BSD) (Martin, Fowlkes, Tal, Malik, et al., 2001) and 3,859 images from the waterloo exploration database (Ma, et al., 2016) to train the Gaussian synthetic denoising model. It is noticeable that different areas of an image contain different detailed information (Zoran & Weiss, 2011). Inspired by that fact, we divide the training noisy images into 1,348,480 patches of size 50×50. The patch is useful to facilitate more robust features and improve the efficiency of training a denoising model. A side attraction is that noise is varying and complex in the real world. Based on this reason, we use 100 real noisy images with size of 512×512 from the benchmark dataset (Xu, Li, Liang, Zhang, & Zhang, 2018) to train a real-noisy denoising model. This dataset is captured by five cameras, such as Canon 5D Mark II, Cannon 80D, Cannon 600D, Nikon D800 and Sony A7 II with different sensor sizes (e.g. 1, 600, 3,200 and 6, 400). To accelerate the speed of training, the 100 real-noisy images are also divided into 211,600 patches of size 50×50. Additionally, each training image above is randomly rotated by one way from eight ways: original image, 90°, 180°, 270°, original image with flopped itself horizontally, 90° with flopped itself horizontally, 180° with flopped itself horizontally and 270° with flopped itself horizontally.

4.1.2. Test datasets

We evaluate the denosing performance of ADNet via six datasets, i.e. BSD68 (Roth & Black, 2005), Set12, CBSD68 (Roth & Black, 2005), Kodak24 (Franzen, 1999), McMaster (Zhang, Wu, Buades, & Li, 2011) and cc (Nam, Hwang, Matsushita, & Joo Kim, 2016), which consist of 68, 12, 68, 24, 18 and 15 images, respectively. Specifically, the BSD68 and Set12 are gray images. The Set12 is shown in Fig. 2. The other datasets are color images. Also, the BSD68 and CBSD68 have the same scenes. The real-noisy cc dataset is captured from three different cameras (i.e. Nikon D800, Nikon D600 and Canon 5D Mark III) with different ISO (1,600, 3,200 and 6,400). The size of each real-noisy image is 512×512. The nine real-noisy images from the cc are shown in Fig. 3.

4.2. Implementation details

The depth of ADNet is the same as DnCNN. The initial parameters are learning rate of 1e−3, epsilon of 1e−8, beta_1 of 0.9, beta_2 of 0.999 and Ref. He, Zhang, Ren, and Sun (2015) for training a denoising model. Specifically, the epsilon, beta_1 and beta_2 are parameters of the BN. Also, the batch size and the number of epochs are 128 and 70, respectively. The learning rates vary from 1e−3 to 1e−5 for the 70 epochs. That is, the learning rates of the 1st–30th epochs are 1e−3. The learning rates of the 31st–60th epochs are 1e−4. The learning rates of the final ten epochs are 1e−5. Further, the Adam (Kingma & Ba, 2014) is exploited to optimize the loss function.

We apply Pytorch 1.01 (Paszke, Gross, Chintala, & Chanan, 2017) and Python 2.7 to train and test the proposed ADNet in image denoising. Specifically, all the experiments are conducted on the Ubuntu 14.04 from a PC: an Intel Core i7-6700 CPU, a RAM 16G and an Nvidia GeForce GTX 1080 Ti GPU. Finally, the Nvidia CUDA of 8.0 and cuDNN of 7.5 are employed to accelerate the training speed of GPU.

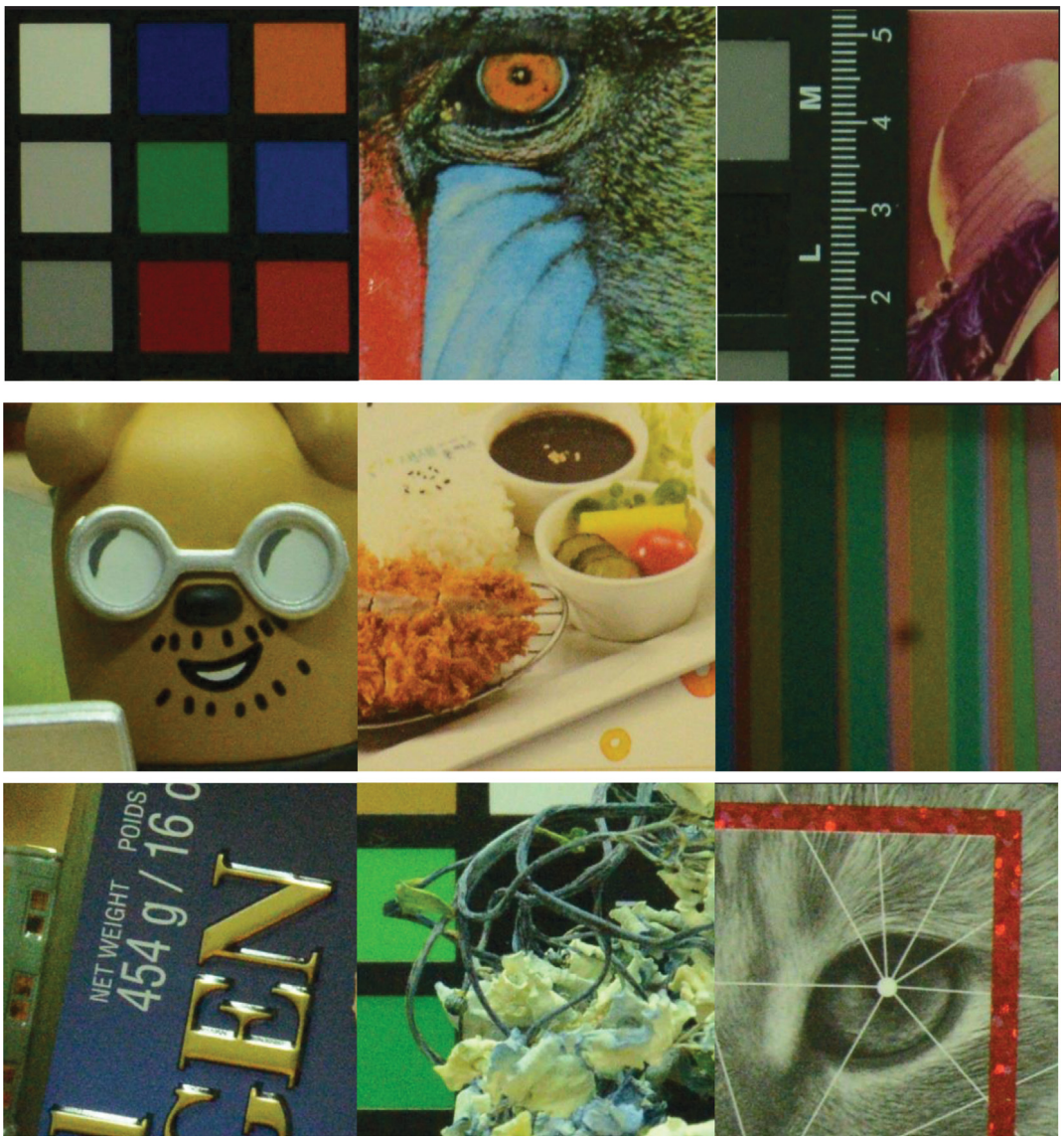

Fig. 3. 9 real-noisy images from cc dataset.

4.3. Network analysis

In this subsection, we will introduce and verify the rationalities of several key techniques: sparse block, feature enhancement block and attention block.

Sparse block: As shown in Xu, Zhang, Lu, and Yang (2016), it is known that the sparse mechanism meets the following requirements in general: (1) less high-energy points and more low-energy points. (2) Irregular high-energy points. Motivated by these ideas, we propose a novel sparse mechanism in CNN for image denoising. Specifically, due to mapping more context information from dilated convolutions, the second, fifth, ninth and twelfth layers are regarded as high-energy points. Other layers in the SB are treated as low-energy points. The combination of several high-energy and many low-energy points can be considered as sparsity. However, how to choose the high-energy points is important. Here we choose the high-energy points from the design of network architecture.

It is noted that with the growth of depth, the shallow layers have poorer effect on deep layers in CNN (Tai et al., 2017b). If successive dilated convolutions are placed at the shallow layers, the deep layers cannot fully map the information of shallow layers. Additionally, if the dilated convolution is set at the first layer, the first layer would be padded into zero. That might cut down the denoising performance. The fact is tested by Table 1, where 'SB with extended layers' has higher peak signal to noise ratio (PSNR) value (Tai et al., 2017a) than that of 'SHE with extended layers'. Additionally, the 'SHE' is successive high-energy points, which consists of dilated convolutions with dilated factor 2 from the second, third, fourth and fifth layers. $PSNR = 10 * \log_{10}(MAX)^2/MSE$, where MAX is the maximum pixel value of each image. The MSE denotes the error between a real clean image and a latent clean image (Tian, Xu, & Zuo, 2020). It is noticed that the depth of the ADNet is 17. To keep the same depth as the ADNet, we add the same 3-layer Conv+BN+ReLU and 2-layer convolutions in the final five layers to conduct the comparative experiments as shown in Tables 1 and 3, respectively. The added five layers are called extended layers.

It is known that diversity of the network is bigger, its representation ability is stronger (Zhang et al., 2018). Inspired by the idea, we do not apply equidistant points as high-energy points. That is proved through 'SB with extended layers' and 'EHE with extended layers' as shown in Table 1, where 'EHE' denotes the equidistant high-energy points. Also, 'EHE' is that dilated

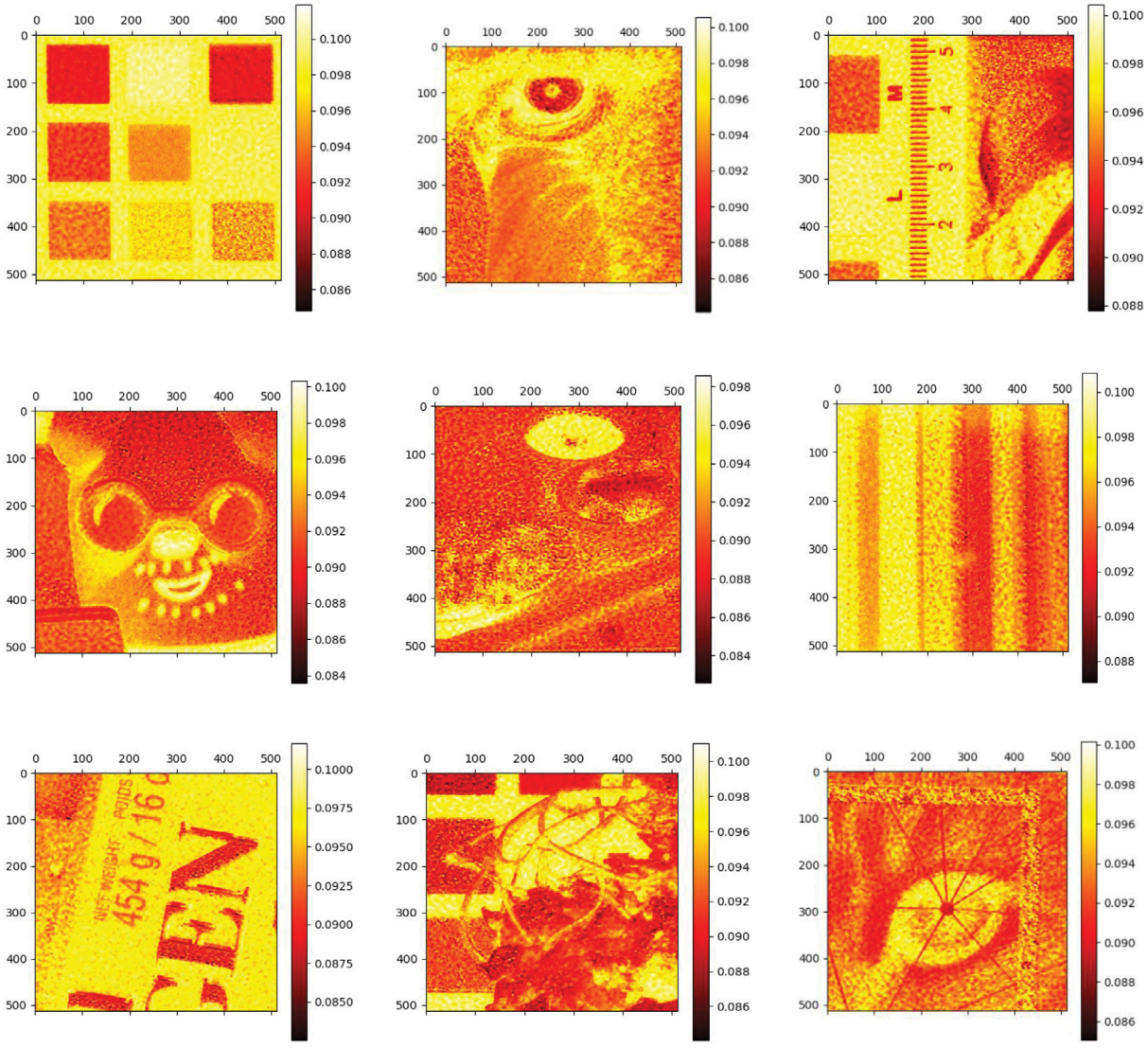

Fig. 4. 9 thermodynamic images from the proposed AB.

Table 1
PSNR (dB) results of several networks from BSD68 with $\sigma = 25$.

Methods	PSNR (dB)
SB with extended layers	29.016
EHE with extended layers	28.886
SHE with extended layers	28.999
ADC with extended layers	28.842
CB	28.891

Table 2
PSNR (dB) results of different methods from BSD68 on noise level of 25.

Methods	PSNR (dB)
ADNet	29.119
FEB + AB	29.088
FEB	29.045
RL + BN	28.988
RL	20.477

convolutions with dilated factor of 2 are used at the second, fifth, eighteenth and eleventh layers, respectively. Combining the illustrations above, the second, fifth, ninth and twelfth layers with dilated convolutions as high-energy points are reasonable. Additionally, the SB is also designed, according to both the denoising performance and efficiency.

To improve the denoising performance, 12-layer SB has the receptive field of size 33 × 33, which has the same effect as a 16-layer network. That reduces the depth of the network. Also, irregular high-energy points can promote the diversity of the network architecture, which is useful to improve the performance in low-level task (Zhang et al., 2018). Further, the fact above is verified by Table 2, where 'ADNet' achieves greater PSNR value (Tai et al., 2017a) than that of 'FEB+AB'. Specifically, 'FEB+AB' is ADNet without SB. Additionally, the 'SB with extended layers' outperforms the conventional block ('CB') as illustrated in Table 1, where the 'CB' is 15-layer Conv+BN+ReLU and two convolutions with size 3 × 3.

Table 3
Running time of different networks the noisy images of sizes 256×256, 512×512 and 1024 × 1024 with $\sigma = 25$.

Methods	Device	256 × 256	512 × 512	1024 × 1024
SB with extended layers	GPU	0.0461	0.0791	0.2061
ADC with extended layers	GPU	0.0534	0.1036	0.3094

Table 4
Complexity of different denoising networks.

Methods	Parameters	Flops
ADNet	0.52 M	1.29 G
Uncompressed ADNet	0.56 M	1.39 G
DnCNN	0.55 M	1.39 G
RED30	4.13 M	10.33 G

To improve the efficiency, the SB only uses four high-energy points and eight low-energy points to reduce the running time for dealing with each noisy image, which is very competing with a lot of high-energy points. The low-efficiency reason from a lot of high-energy points is that dilated convolution of each layer needs to pad some things to capture more context information. Thus, the 'SB with extended layers' is faster than the 'ADC with extended layers' as shown in Table 3. The ADC is dilated convolutions for the whole twelve layers. Additionally, the receptive field size of 58 × 58 from 'ADC with extended layers' is greater than the given patch size of 50 × 50, which results in patch size cannot full map the 'ADC with extended layers'. Thus, the proposed SB has better denoising performance than that of ADC as shown in Table 1. In summary, our proposed SB is reasonable and advantageous from the network architecture, performance and efficiency.

Feature enhancement block: influence of the network is very important to mine features for image processing (Tai et al., 2017b). However, with the growth of the network depth, the effects from the shallow layers on deep layers are weakened. For resolving the problem, Zhang et al. (2018) combined the local and global features to enhance the performance for image restoration. Inspired by that, the feature enhancement block concatenates the input noisy image (regarded as global features) and the output of the sixteenth layer (treated as local features) to enhance the effect of the shallow layer, where concatenation operation is called a long path operation in this paper. Actually, the input noisy image includes strong noise information. Thus, we choose the original input to be complementary with deep layer for the denoising task. Finally, performance of the FEB is proved as shown in Table 2, where 'FEB' has higher PSNR than that of 'RL+BN', where the 'RL+BN' is the combination of RL and BN techniques. Additionally, the output of the sixteenth layer in ADNet is c, where is useful to improve the denoising efficiency.

Attention block: environment (i.e. dark light) may increase the noise of captured image by the camera. Thus, how to extract useful information from the noisy image with complex

Table 5
Average PSNR (dB) of different methods on BSD68 with different noise levels of 15, 25 and 50.

Methods	BM3D	WNNM	EPLL	MLP	CSF	TNRD	DnCNN	IRCNN	ECNDNet	ADNet	ADNet-B
$\sigma = 15$	31.07	31.37	31.21	–	31.24	31.42	31.72	31.63	31.71	31.74	31.56
$\sigma = 25$	28.57	28.83	28.68	28.96	28.74	28.92	29.23	29.15	29.22	29.25	29.14
$\sigma = 50$	25.62	25.87	25.67	26.03	–	25.97	26.23	26.19	26.23	26.29	26.24

Table 6
Average PSNR (dB) results of different methods on Set12 with noise levels of 15, 25 and 50.

Images	C.man	House	Peppers	Starfish	Monarch	Airplane	Parrot	Lena	Barbara	Boat	Man	Couple	Average
Noise level	$\sigma = 15$												
BM3D (Dabov et al., 2007)	31.91	34.93	32.69	31.14	31.85	31.07	31.37	34.26	33.10	32.13	31.92	32.10	32.37
WNNM (Gu et al., 2014)	32.17	35.13	32.99	31.82	32.71	31.39	31.62	34.27	33.60	32.27	32.11	32.17	32.70
EPLL (Zoran & Weiss, 2011)	31.85	34.17	32.64	31.13	32.10	31.19	31.42	33.92	31.38	31.93	32.00	31.93	32.14
CSF (Schmidt & Roth, 2014)	31.95	34.39	32.85	31.55	32.33	31.33	31.37	34.06	31.92	32.01	32.08	31.98	32.32
TNRD (Chen & Pock, 2016)	32.19	34.53	33.04	31.75	32.56	31.46	31.63	34.24	32.13	32.14	32.23	32.11	32.50
DnCNN (Zhang, Zuo, Chen, et al., 2017)	32.61	34.97	33.30	32.20	33.09	31.70	31.83	34.62	32.64	32.42	32.46	32.47	32.86
IRCNN (Zhang et al., 2017)	32.55	34.89	33.31	32.02	32.82	31.70	31.84	34.53	32.43	32.34	32.40	32.40	32.77
FFDNet (Zhang, Zuo, Zhang, 2018)	32.43	35.07	33.25	31.99	32.66	31.57	31.81	34.62	32.54	32.38	32.41	32.46	32.77
ECNDNet (Tian, et al., 2019)	32.56	34.97	33.25	32.17	33.11	31.70	31.82	34.52	32.41	32.37	32.39	32.39	32.81
ADNet	32.81	35.22	33.49	32.17	33.17	31.86	31.96	34.71	32.80	32.57	32.47	32.58	32.98
ADNet-B	31.98	35.12	33.34	32.01	33.01	31.63	31.74	34.62	32.55	32.48	32.34	32.43	32.77
Noise level	$\sigma = 25$												
BM3D (Dabov et al., 2007)	29.45	32.85	30.16	28.56	29.25	28.42	28.93	32.07	30.71	29.90	29.61	29.71	29.97
WNNM (Gu et al., 2014)	29.64	33.22	30.42	29.03	29.84	28.69	29.15	32.24	31.24	30.03	29.76	29.82	30.26
EPLL (Zoran & Weiss, 2011)	29.26	32.17	30.17	28.51	29.39	28.61	28.95	31.73	28.61	29.74	29.66	29.53	29.69
MLP (Burger, Schuler, & Harmeling, 2012)	29.61	32.56	30.30	28.82	29.61	28.82	29.25	32.25	29.54	29.97	29.88	29.73	30.03
CSF (Schmidt & Roth, 2014)	29.48	32.39	30.32	28.80	29.62	28.72	28.90	31.79	29.03	29.76	29.71	29.53	29.84
TNRD (Chen & Pock, 2016)	29.72	32.53	30.57	29.02	29.85	28.88	29.18	32.00	29.41	29.91	29.87	29.71	30.06
DnCNN (Zhang, Zuo, Chen, et al., 2017)	30.18	33.06	30.87	29.41	30.28	29.13	29.43	32.44	30.00	30.21	30.10	30.12	30.43
IRCNN (Zhang et al., 2017)	30.08	33.06	30.88	29.27	30.09	29.12	29.47	32.43	29.92	30.17	30.04	30.08	30.38
FFDNet (Zhang, Zuo, Zhang, 2018)	30.10	33.28	30.93	29.32	30.08	29.04	29.44	32.57	30.01	30.25	30.11	30.20	30.44
ECNDNet (Tian, et al., 2019)	30.11	33.08	30.85	29.43	30.30	29.07	29.38	32.38	29.84	30.14	30.03	30.03	30.39
ADNet	30.34	33.41	31.14	29.41	30.39	29.17	29.49	32.61	30.25	30.37	30.08	30.24	30.58
ADNet-B	29.94	33.38	30.99	29.22	30.38	29.16	29.41	32.59	30.05	30.28	30.01	30.15	30.46
Noise level	$\sigma = 50$												
BM3D (Dabov et al., 2007)	26.13	29.69	26.68	25.04	25.82	25.10	25.90	29.05	27.22	26.78	26.81	26.46	26.72
WNNM (Gu et al., 2014)	26.45	30.33	26.95	25.44	26.32	25.42	26.14	29.25	27.79	26.97	26.94	26.64	27.05
EPLL (Zoran & Weiss, 2011)	26.10	29.12	26.80	25.12	25.94	25.31	25.95	28.68	24.83	26.74	26.79	26.30	26.47
MLP (Burger et al., 2012)	26.37	29.64	26.68	25.43	26.26	25.56	26.12	29.32	25.24	27.03	27.06	26.67	26.78
TNRD (Chen & Pock, 2016)	26.62	29.48	27.10	25.42	26.31	25.59	26.16	28.93	25.70	26.94	26.98	26.50	26.81
DnCNN (Zhang, Zuo, Chen, et al., 2017)	27.03	30.00	27.32	25.70	26.78	25.87	26.48	29.39	26.22	27.20	27.24	26.90	27.18
IRCNN (Zhang et al., 2017)	26.88	29.96	27.33	25.57	26.61	25.89	26.55	29.40	26.24	27.17	27.17	26.88	27.14
FFDNet (Zhang, Zuo, Zhang, 2018)	27.05	30.37	27.54	25.75	26.81	25.89	26.57	29.66	26.45	27.33	27.29	27.08	27.32
ECNDNet (Tian, et al., 2019)	27.07	30.12	27.30	25.72	26.82	25.79	26.32	29.29	26.26	27.16	27.11	26.84	27.15
ADNet	27.31	30.59	27.69	25.70	26.90	25.88	26.56	29.59	26.64	27.35	27.17	27.07	27.37
ADNet-B	27.22	30.43	27.70	25.63	26.92	26.03	26.56	29.53	26.51	27.22	27.19	27.05	27.33

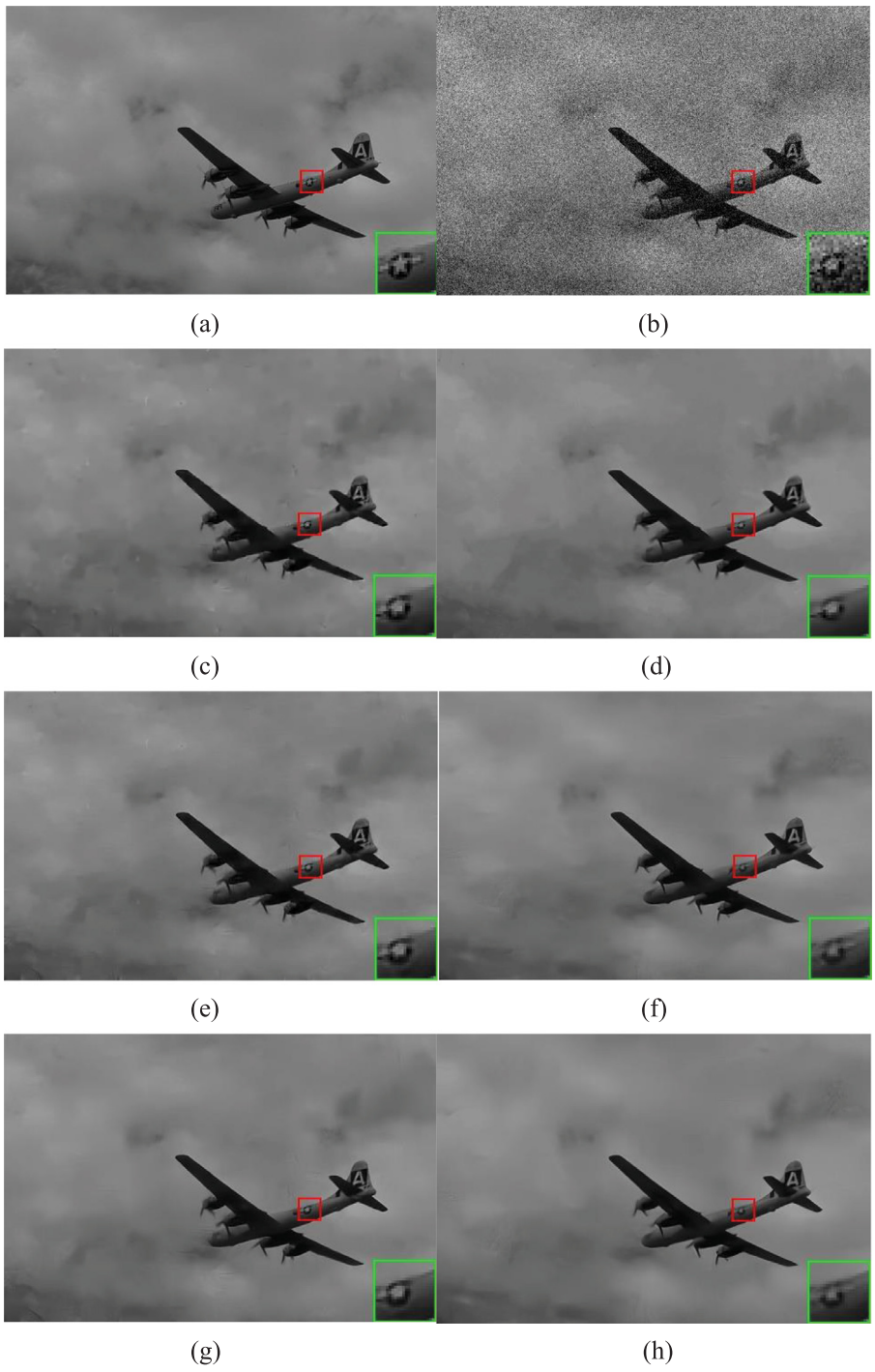

Fig. 5. Denoising results of different methods on one image from BSD68 with $\sigma = 25$. (a) Original image (b) Noisy image/20.23dB (c) EPLL/37.21 dB (d) WNNM/37.26 dB (e) TNRD /37.83 dB (f) ECNDNet/38.31 dB (g) DnCNN/38.35 dB (h) ADNet/38.48 dB.

background is very important. For addressing this problem, the attentive idea is developed to extract saliency features for image applications (Wang, Girshick, Gupta, & He, 2018a). Specifically, the attentive idea uses the current stage to guide the previous stage and obtain high-frequency features. From this point of view, we propose an attention mechanism to extract the latent noise hidden in the complex background. Further, the detailed information of AB is illustrated in Section 3.5. Additionally, we can see that the noise in Fig. 4 is more obvious than that of Fig. 3, where redder area denotes the higher weight from the attention mechanism (also regarded as smaller value in Fig. 4). Moreover, the 'FEB+AB' is also higher than 'FEB' in PSNR as illustrated in Table 2. These prove that the proposed AB is very effective for complex noisy images. Also, the convolution size of 1×1 in AB is combined with the c output channels from the sixteenth layer in FEB to reduce the complexity of denoising model as shown in Table 4. The 'ADNet' has smaller parameters and flops than that of 'Uncompressed ADNet', where 'Uncompressed ADNet' is the 64 output channels from the sixteenth layer and the filter size of the seventeenth layer is 3×3. Thus, taking into account between task characteristic and complexity, we think that the proposed ADNet is competing in image denoising, especially complex noisy images. Also, the BN is useful to promote the denoising performance in ADNet as shown in Table 2.

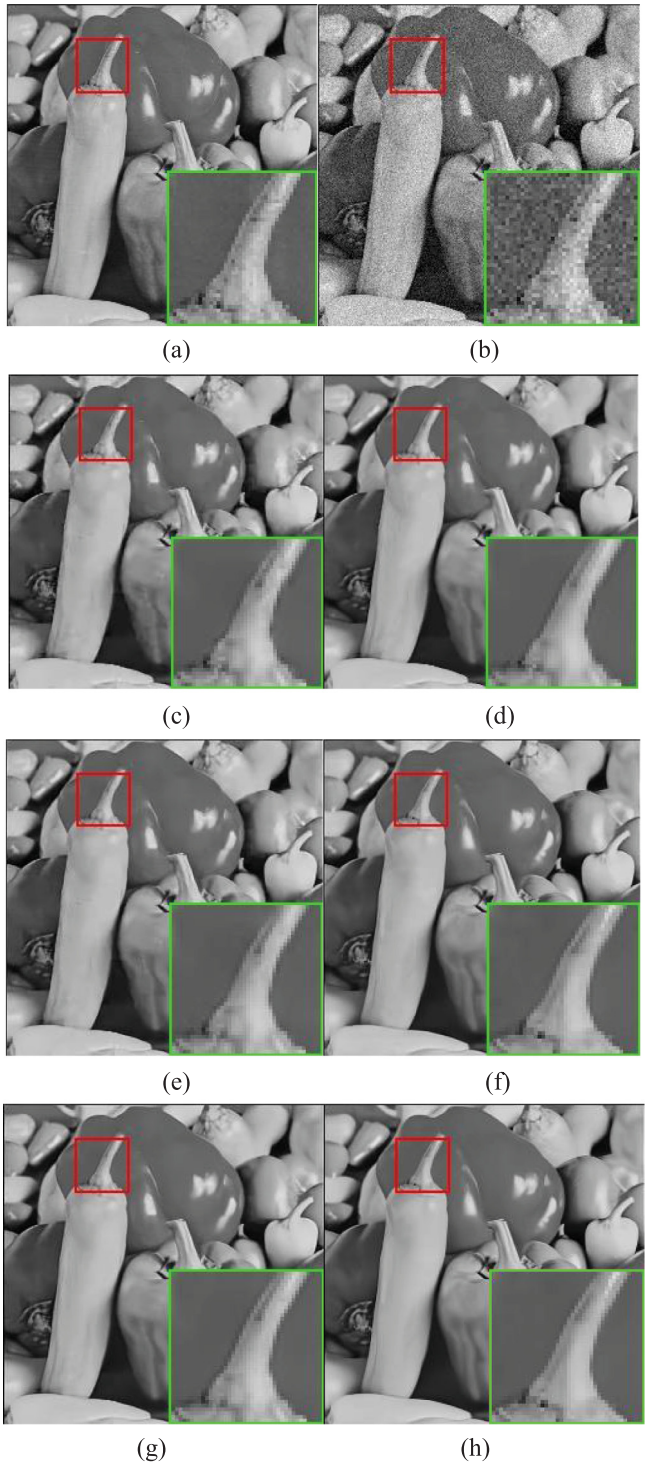

Fig. 6. Denoising results of different methods on one image from Set12 when $\sigma = 15$. (a) Original image (b) Noisy image/24.64 dB (c) EPLL/32.64 dB (d) WNNM/32.99 dB (e) TNRD/33.04 dB (f) ECNDNet/33.25 dB (g) DnCNN/30.30 dB (h) ADNet/33.49 dB.

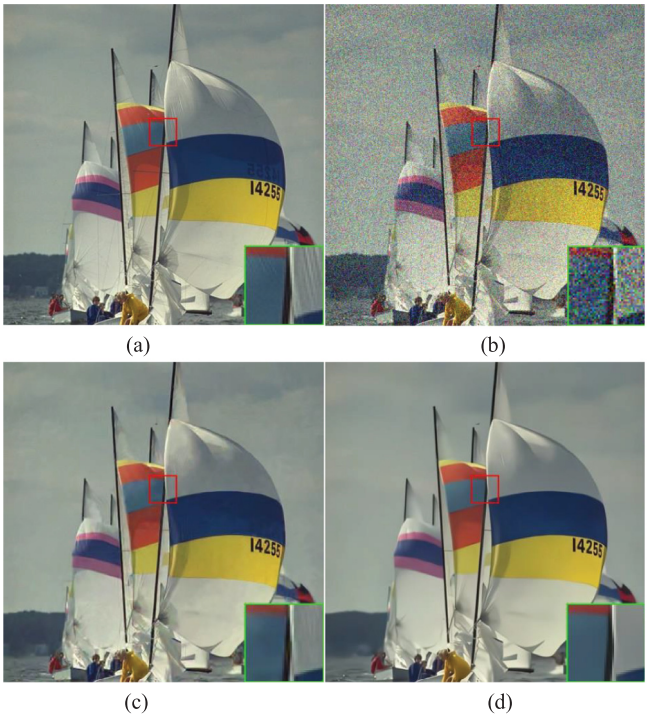

Fig. 7. Denoising results of one color Gaussian noisy image from Kodak 24 with $\sigma = 35$. (a) Original image (b) Noisy image/17.40 dB (c) CBM3D/32.56 dB (d) ADNet/33.71 dB.

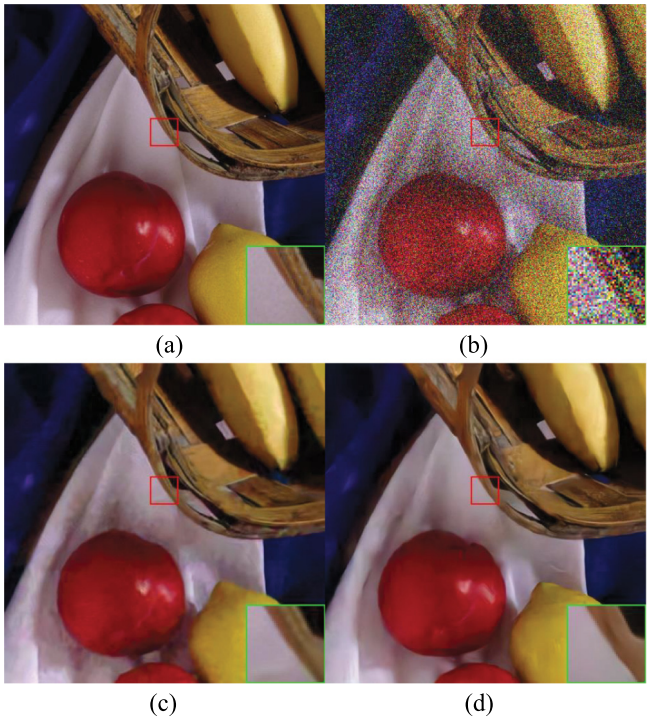

Fig. 8. Denoising results of one color Gaussian noisy image from McMaster with $\sigma = 75$. (a) Original image (b) Noisy image/12.34 dB (c) CBM3D/29.64 dB (d) ADNet/30.44 dB.

4.4. Comparisons with state-of-the-art denoising methods

In this section, we test the denoising performance of ADNet in terms of both quantitative and qualitative analysis. The quantitative analysis is that compares the PSNR, running time and complexity of ADNet with competing denoising methods, such as BM3D (Dabov et al., 2007), weighted nuclear norm minimization method (WNNM) (Gu et al., 2014), Multi-Layer Perception (MLP) (Burger et al., 2012), trainable nonlinear reaction diffusion (TNRD) (Chen & Pock, 2016), expected patch log likelihood (EPLL) (Zoran & Weiss, 2011), cascade of shrinkage

Table 7
PSNR (dB) results of different methods on CBSD68, Kodak24 and McMaster datasets with noise levels of 15, 25, 35, 50 and 75.

Datasets	Methods	$\sigma = 15$	$\sigma = 25$	$\sigma = 35$	$\sigma = 50$	$\sigma = 75$
CBSD68	CBM3D (Dabov et al., 2007)	33.52	30.71	28.89	27.38	25.74
	FFDNet (Zhang, Zuo, Zhang, 2018)	33.80	31.18	29.57	27.96	26.24
	DnCNN (Zhang, Zuo, Chen, et al., 2017)	33.98	31.31	29.65	28.01	–
	IRCNN (Zhang et al., 2017)	33.86	31.16	29.50	27.86	–
	ADNet	33.99	31.31	29.66	28.04	26.33
	ADNet-B	33.79	31.12	29.48	27.83	–
Kodak24	CBM3D (Dabov et al., 2007)	34.28	31.68	29.90	28.46	26.82
	FFDNet (Zhang, Zuo, Zhang, 2018)	34.55	32.11	30.56	28.99	27.25
	DnCNN (Zhang, Zuo, Chen, et al., 2017)	34.73	32.23	30.64	29.02	–
	IRCNN (Zhang et al., 2017)	34.56	32.03	30.43	28.81	–
	ADNet	34.76	32.26	30.68	29.10	27.40
	ADNet-B	34.53	32.03	30.44	28.81	–
McMaster	CBM3D (Dabov et al., 2007)	34.06	31.66	29.92	28.51	26.79
	FFDNet (Zhang, Zuo, Zhang, 2018)	34.47	32.25	30.76	29.14	27.29
	DnCNN (Zhang, Zuo, Chen, et al., 2017)	34.80	32.47	30.91	29.21	–
	IRCNN (Zhang et al., 2017)	34.58	32.18	30.59	28.91	–
	ADNet	34.93	32.56	31.00	29.36	27.53
	ADNet-B	34.60	32.28	30.72	29.03	–

fields (CSF) (Schmidt & Roth, 2014), DnCNN (Zhang, Zuo, Chen, et al., 2017), IRCNN (Zhang et al., 2017), FFDNet (Zhang, Zuo, Zhang, 2018), enhanced CNN denoising network (ECNDNet) (Tian, et al., 2019), CBM3D (Dabov et al., 2007), neat image (NI) (ABSoft, 2017), RED30 (Mao et al., 2016), MemNet (Tai et al., 2017b) and MWCNN (Liu et al., 2018) for synthetic images and real noisy images. Specifically, the synthetic noisy images include two kinds: gray and color Gaussian synthetic noisy images. The gray and color Gaussian noisy images include noisy images of certain noise level and varying noise levels from 0 to 55. The noisy image with varying noise levels is called the image with blind noise. For the first noisy image, we design the experiments on BSD68 and Set12, respectively. As shown in Table 5, the ADNet achieves the best performance on noise levels of 15, 25 and 50, respectively. Also, the ADNet for blind denoising (ADNet-B) outperforms the state-of-the-art denosing method, such as DnCNN for noise level of 50. Table 6 is used to show the denoising result of each image from Set12. From that, we can see that the ADNet is superior to DnCNN, IRCNN and FFDNet for different noise levels (i.e. 15, 25 and 50) in image denosing. Moreover, the ADNet-B obtains the second performance for $\sigma = 25$ and $\sigma = 50$. That shows that our proposed ADNet is very robust for the certain noisy image and blind denoising. Specifically, red and blue lines are used to denote the best and second denoising results in Tables 5 and 6, respectively. For the second noisy image, we choose several popular denoising, e.g. CBM3D, FFDNet, IRCNN, ADNet and ADNet-B on three benchmark datasets (i.e. CBSD68, Kodak24 and McMaster) with different noise levels of 15, 25, 35, 50 and 75 for image denoising, respectively. As shown in Table 7, we can see that the ADNet achieves excellent results on color Gaussian synthetic noisy image. Additionally, our proposed ADNet is very outstanding on real noisy images. As shown in Table 8, the ADNet obtains the improvement of 1.83dB than that of DnCNN in PSNR, where the popular denoising methods such as CBM3D, DnCNN, NI and CSF are used as comparative experiments to remove the noise from the real noisy image. That proves that the proposed AB is useful for complex noisy images, such as real noisy images. Specifically, the red and blue lines are marked the highest and second denoising performance in Tables 7–9, respectively.

Table 9
Running time of 11 popular denoising methods for the noisy images of sizes 256×256, 512×512 and 1024×1024.

Methods	Device	256×256	512×512	1024×1024
BM3D (Dabov et al., 2007)	CPU	0.59	2.52	10.77
WNNM (Gu et al., 2014)	CPU	203.1	773.2	2536.4
EPLL (Zoran & Weiss, 2011)	CPU	25.4	45.5	422.1
MLP (Burger et al., 2012)	CPU	1.42	5.51	19.4
TNRD (Chen & Pock, 2016)	CPU	0.45	1.33	4.61
CSF (Schmidt & Roth, 2014)	CPU	–	0.92	1.72
DnCNN (Zhang, Zuo, Chen, et al., 2017)	GPU	0.0344	0.0681	0.1556
RED30 (Mao et al., 2016)	GPU	1.362	4.702	15.77
MemNet (Tai et al., 2017b)	GPU	0.8775	3.606	14.69
MWCNN (Liu et al., 2018)	GPU	0.0586	0.0907	0.3575
ADNet	GPU	0.0467	0.0798	0.2077

For running time, we choose 11 state-of-the-art denoising methods to conduct the experiments for testing the denoising running time of a noisy image from different sizes (i.e. 256×256, 512×512 and 1024×1024) as illustrated in Table 9. From that, we can see that although the proposed ADNet obtains the second result, its running time is very competing in contrast to other popular methods. Further, the ADNet has the smaller

Table 8
PSNR (dB) results of different methods on real noisy images.

Camera settings	CBM3D (Dabov et al., 2007)	DnCNN (Zhang, Zuo, Chen, et al., 2017)	CSF (Schmidt & Roth, 2014)	NI (ABSoft, 2017)	ADNet
Canon 5D ISO = 3200	39.76	37.26	35.68	37.68	35.96
	36.40	34.13	34.03	34.87	36.11
	36.37	34.09	32.63	34.77	34.49
Nikon D600 ISO = 3200	34.18	33.62	31.78	34.12	33.94
	35.07	34.48	35.16	35.36	34.33
	37.13	35.41	39.98	38.68	38.87
Nikon D800 ISO = 1600	36.81	37.95	34.84	37.34	37.61
	37.76	36.08	38.42	38.57	38.24
	37.51	35.48	35.79	37.87	36.89
Nikon D800 ISO = 3200	35.05	34.08	38.36	36.95	37.20
	34.07	33.70	35.53	35.09	35.67
	34.42	33.31	40.05	36.91	38.09
Nikon D800 ISO = 6400	31.13	29.83	34.08	31.28	32.24
	31.22	30.55	32.13	31.38	32.59
	30.97	30.09	31.52	31.40	33.14
Average	35.19	33.86	35.33	35.49	35.69

complexity than that of state-of-the-arts, such as DnCNN and RED30 as shown in Table 4. Thus, the ADNet is very effective for image denoising in terms of quantitative analysis.

For qualitative analysis, we use some visual figures from BSD68, Set12, Kodak24 and McMaster to show the denoising performance of different methods. Specifically, the visual figure is obtained by amplifying the observed area. The amplified area is clearer, the corresponding denosing method is more robust. Figs. 5 and 6 show the visual effects from the gray noisy image denoising model. Figs. 7 and 8 describe the visual effects from the color noisy image denoising model. From these figures, we can see that the ADNet is clearer than state-of-the-art denoising methods, such as DnCNN, ECNDNet and CBM3D. Also, red and blue lines are used to express the best and second denoising results in these figures, respectively. The fact shows that our proposed ADNet is effective for qualitative analysis. According to the presentations above, we conclude that the ADNet is very suitable to image denoising.

5. Conclusion

In this paper, we propose an attention-guided denoising CNN as well as ADNet for image denoising. Main components of ADNet play the following roles. The SB is based on dilated and common convolutions and can make a tradeoff between denoising performance and efficiency. The FEB is utilized to enhance the representation capability on noisy images of the model by virtue of cooperation of the global and local information. The AB can extract the latent noise information hidden in the complex background, which is very effective for complex noisy images, such as images with blind noise and real noisy images. After the components above are carried out, the RB is implemented to obtain the resultant clean image. Since FEB is consolidated by AB, the efficiency and complexity for training a denoising model can be improved. The experimental results show that our proposed ADNet is very effective for image denoising in terms of both quantitative and qualitative evaluations.

Acknowledgments

This paper is supported in part by the National Nature Science Foundation of China under Grant No. 61972187, in part by the Shenzhen Municipal Science and Technology Innovation Council under Grant No. JCYJ20170811155725434 and in part by the Shenzhen Municipal Science and Technology Innovation Council under Grant No. GJHZ20180419190732022.

References

ABSoft, N. (2017). Neat image.

Barbu, A. (2009). Learning real-time MRF inference for image denoising. In *2009 IEEE conference on computer vision and pattern recognition* (pp. 1574–1581). IEEE.

Burger, H. C., Schuler, C. J., & Harmeling, S. (2012). Image denoising: Can plain neural networks compete with BM3D? In *2012 IEEE conference on computer vision and pattern recognition* (pp. 2392–2399). IEEE.

Chambolle, A. (2004). An algorithm for total variation minimization and applications. *Journal of Mathematical imaging and vision*, *20*(1–2), 89–97.

Chen, Y., & Pock, T. (2016). Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration. *IEEE transactions on pattern analysis and machine intelligence*, *39*(6), 1256–1272.

Chen, C., Xiong, Z., Tian, X., & Wu, F. (2018). Deep boosting for image denoising. In *Proceedings of the european conference on computer vision* (pp. 3–18).

Cho, S. I., & Kang, S.-J. (2018). Gradient prior-aided CNN denoiser with separable convolution-based optimization of feature dimension. *IEEE Transactions on Multimedia*, *21*(2), 484–493.

Dabov, K., Foi, A., Katkovnik, V., & Egiazarian, K. (2007). Image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE Transactions on Image Processing*, *16*(8), 2080–2095.

Dong, W., Zhang, L., Shi, G., & Li, X. (2012). Nonlocally centralized sparse representation for image restoration. *IEEE transactions on Image Processing*, *22*(4), 1620–1630.

Du, B., Wei, Q., & Liu, R. (2019). An improved quantum-behaved particle swarm optimization for endmember extraction. *IEEE Transactions on Geoscience and Remote Sensing*.

Ephraim, Y., & Malah, D. (1984). Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator. *IEEE Transactions on acoustics, speech, and signal processing*, *32*(6), 1109–1121.

Franzen, R. (1999). Kodak lossless true color image suite. source: http://r0k.us/graphics/kodak 4.

Gu, S., Zhang, L., Zuo, W., & Feng, X. (2014). Weighted nuclear norm minimization with application to image denoising. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2862–2869).

Guo, S., Yan, Z., Zhang, K., Zuo, W., & Zhang, L. (2019). Toward convolutional blind denoising of real photographs. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1712–1722).

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision* (pp. 1026–1034).

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770–778).

Hui, Z., Wang, X., & Gao, X. (2018). Fast and accurate single image super-resolution via information distillation network. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 723–731).

Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:1502.03167.

Kim, J., Kwon Lee, J., & Mu Lee, K. (2016). Deeply-recursive convolutional network for image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1637–1645).

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097–1105).

Li, Y., Chen, X., Zhu, Z., Xie, L., Huang, G., Du, D., & Wang, X. (2019). Attention-guided unified network for panoptic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7026–7035).

Li, J., Lu, G., Zhang, B., You, J., & Zhang, D. (2019). Shared linear encoder-based multikernel Gaussian process latent variable model for visual classification. *IEEE transactions on cybernetics*.

Liang, X., Zhang, D., Lu, G., Guo, Z., & Luo, N. (2019). A novel multicamera system for high-speed touchless palm recognition. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*.

Liu, P., & Fang, R. (2017). Learning pixel-distribution prior with wider convolution for image denoising. arXiv preprint arXiv:1707.09135.

Liu, J., Sun, Y., Xu, X., & Kamilov, U. S. (2019). Image restoration using total variation regularized deep image prior. In *ICASSP 2019-2019 IEEE international conference on acoustics, speech and signal processing* (pp. 7715–7719). IEEE.

Liu, P., Zhang, H., Zhang, K., Lin, L., & Zuo, W. (2018). Multi-level wavelet-CNN for image restoration. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops* (pp. 773–782).

Lu, Y., Lai, Z., Li, X., Wong, W. K., Yuan, C., & Zhang, D. (2018). Low-rank 2-D neighborhood preserving projection for enhanced robust image representation. *IEEE transactions on cybernetics*, *49*(5), 1859–1872.

Lu, Y., Lai, Z., Xu, Y., Li, X., Zhang, D., & Yuan, C. (2015). Low-rank preserving projections. *IEEE transactions on cybernetics*, *46*(8), 1900–1913.

Ma, K., Duanmu, Z., Wu, Q., Wang, Z., Yong, H., Li, H., et al. (2016). Waterloo exploration database: New challenges for image quality assessment models. *IEEE Transactions on Image Processing*, *26*(2), 1004–1016.

Mairal, J., Bach, F. R., Ponce, J., Sapiro, G., & Zisserman, A. (2009). Non-local sparse models for image restoration. In *ICCV, Vol. 29* (pp. 54–62). Citeseer.

Malfait, M., & Roose, D. (1997). Wavelet-based image denoising using a Markov random field a priori model. *IEEE Transactions on image processing*, *6*(4), 549–565.

Malfliet, W., & Hereman, W. (1996). The tanh method: I. exact solutions of nonlinear evolution and wave equations. *Physica Scripta*, *54*(6), 563.

Mao, X., Shen, C., & Yang, Y.-B. (2016). Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections. In *Advances in neural information processing systems* (pp. 2802–2810).

Martin, D., Fowlkes, C., Tal, D., Malik, J., et al. (2001). *A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics*. Iccv Vancouver:.

Nam, S., Hwang, Y., Matsushita, Y., & Joo Kim, S. (2016). A holistic approach to cross-channel image noise modeling and its application to image denoising. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1683–1691).

Paszke, A., Gross, S., Chintala, S., & Chanan, G. (2017). Pytorch. computer software. vers. 0.3 1.

Peng, Y., Zhang, L., Liu, S., Wu, X., Zhang, Y., & Wang, X. (2019). Dilated residual networks with symmetric skip connection for image denoising. *Neurocomputing*, *345*, 67–76.

Ren, D., Zuo, W., Hu, Q., Zhu, P., & Meng, D. (2019). Progressive image deraining networks: a better and simpler baseline. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3937–3946).

Roth, S., & Black, M. J. (2005). Fields of experts: A framework for learning image priors. In *2005 IEEE computer society conference on computer vision and pattern recognition, Vol. 2* (pp. 860–867). Citeseer.

Schmidt, U., & Roth, S. (2014). Shrinkage fields for effective image restoration. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2774–2781).

Tai, Y., Yang, J., & Liu, X. (2017). Image super-resolution via deep recursive residual network. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3147–3155).

Tai, Y., Yang, J., Liu, X., & Xu, C. (2017). Memnet: A persistent memory network for image restoration. In *Proceedings of the IEEE international conference on computer vision* (pp. 4539–4547).

Tian, C., Xu, Y., Fei, L., Wang, J., Wen, J., & Luo, N. (2019). Enhanced CNN for image denoising. *CAAI Transactions on Intelligence Technology*, *4*(1), 17–23.

Tian, C., Xu, Y., & Zuo, W. (2020). Image denoising using deep CNN with batch renormalization. *Neural Networks*, *121*, 461–473.

Tian, C., Zhang, Q., Sun, G., Song, Z., & Li, S. (2018). FFT consolidated sparse and collaborative representation for image classification. *Arabian Journal for Science and Engineering*, *43*(2), 741–758.

Tripathi, S., Lipton, Z. C., & Nguyen, T. Q. (2018). Correction by projection: Denoising images with generative adversarial networks. arXiv preprint arXiv:1803.04477.

Wang, X., Girshick, R., Gupta, A., & He, K. (2018). Non-local neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7794–7803).

Wang, T., Sun, M., & Hu, K. (2017). Dilated deep residual network for image denoising. In *2017 IEEE 29th international conference on tools with artificial intelligence* (pp. 1272–1279). IEEE.

Wang, H., Wang, Q., Gao, M., Li, P., & Zuo, W. (2018). Multi-scale location-aware kernel representation for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1248–1257).

Xu, J., Li, H., Liang, Z., Zhang, D., & Zhang, L. (2018). Real-world noisy image denoising: A new benchmark. arXiv preprint arXiv:1804.02603.

Xu, Y., Zhang, Z., Lu, G., & Yang, J. (2016). Approximately symmetrical face images for image preprocessing in face recognition and sparse representation based classification. *Pattern Recognition*, *54*, 68–82.

Xu, J., Zhang, L., & Zhang, D. (2018). A trilateral weighted sparse coding scheme for real-world image denoising. In *Proceedings of the european conference on computer vision* (pp. 20–36).

Yu, F., & Koltun, V. (2015). Multi-scale context aggregation by dilated convolutions. arXiv preprint arXiv:1511.07122.

Zha, Z., Liu, X., Huang, X., Shi, H., Xu, Y., Wang, Q., et al. (2017). Analyzing the group sparsity based on the rank minimization methods. In *2017 IEEE international conference on multimedia and expo* (pp. 883–888). IEEE.

Zhang, Y., Tian, Y., Kong, Y., Zhong, B., & Fu, Y. (2018). Residual dense network for image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2472–2481).

Zhang, L., Wu, X., Buades, A., & Li, X. (2011). Color demosaicking by local directional interpolation and nonlocal adaptive thresholding. *Journal of Electronic imaging*, *20*(2), 023016.

Zhang, X., Yang, W., Hu, Y., & Liu, J. (2018). DMCNN: Dual-domain multi-scale convolutional neural network for compression artifacts removal. In *2018 25th IEEE international conference on image processing* (pp. 390–394). IEEE.

Zhang, K., Zuo, W., Chen, Y., Meng, D., & Zhang, L. (2017). Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Transactions on Image Processing*, *26*(7), 3142–3155.

Zhang, K., Zuo, W., Gu, S., & Zhang, L. (2017). Learning deep CNN denoiser prior for image restoration. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3929–3938).

Zhang, K., Zuo, W., & Zhang, L. (2018). FFDNet: Toward a fast and flexible solution for CNN-based image denoising. *IEEE Transactions on Image Processing*, *27*(9), 4608–4622.

Zhu, Z., Wu, W., Zou, W., & Yan, J. (2018). End-to-end flow correlation tracking with spatial-temporal attention. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 548–557).

Zoran, D., & Weiss, Y. (2011). From learning models of natural image patches to whole image restoration. In *2011 international conference on computer vision* (pp. 479–486). IEEE.