End-to-End Learning for Joint Image Demosaicing, Denoising and Super-Resolution

Wenzhu Xing and Karen Egiazarian
Computational Image Group, Tampere University, Finland
{wenzhu.xing, karen.eguiazarian}@tuni.fi

Abstract

Image denoising, demosaicing and super-resolution are key problems of image restoration well studied in the recent decades. Often, in practice, one has to solve these problems simultaneously. A problem of finding a joint solution of the multiple image restoration tasks just begun to attract an increased attention of researchers. In this paper, we propose an end-to-end solution for the joint demosaicing, denoising and super-resolution based on a specially designed deep convolutional neural network (CNN). We systematically study different methods to solve this problem and compared them with the proposed method. Extensive experiments carried out on large image datasets demonstrate that our method outperforms the state-of-the-art both quantitatively and qualitatively. Finally, we have applied various loss functions in the proposed scheme and demonstrate that by using the mean absolute error as a loss function, we can obtain superior results in comparison to other cases.

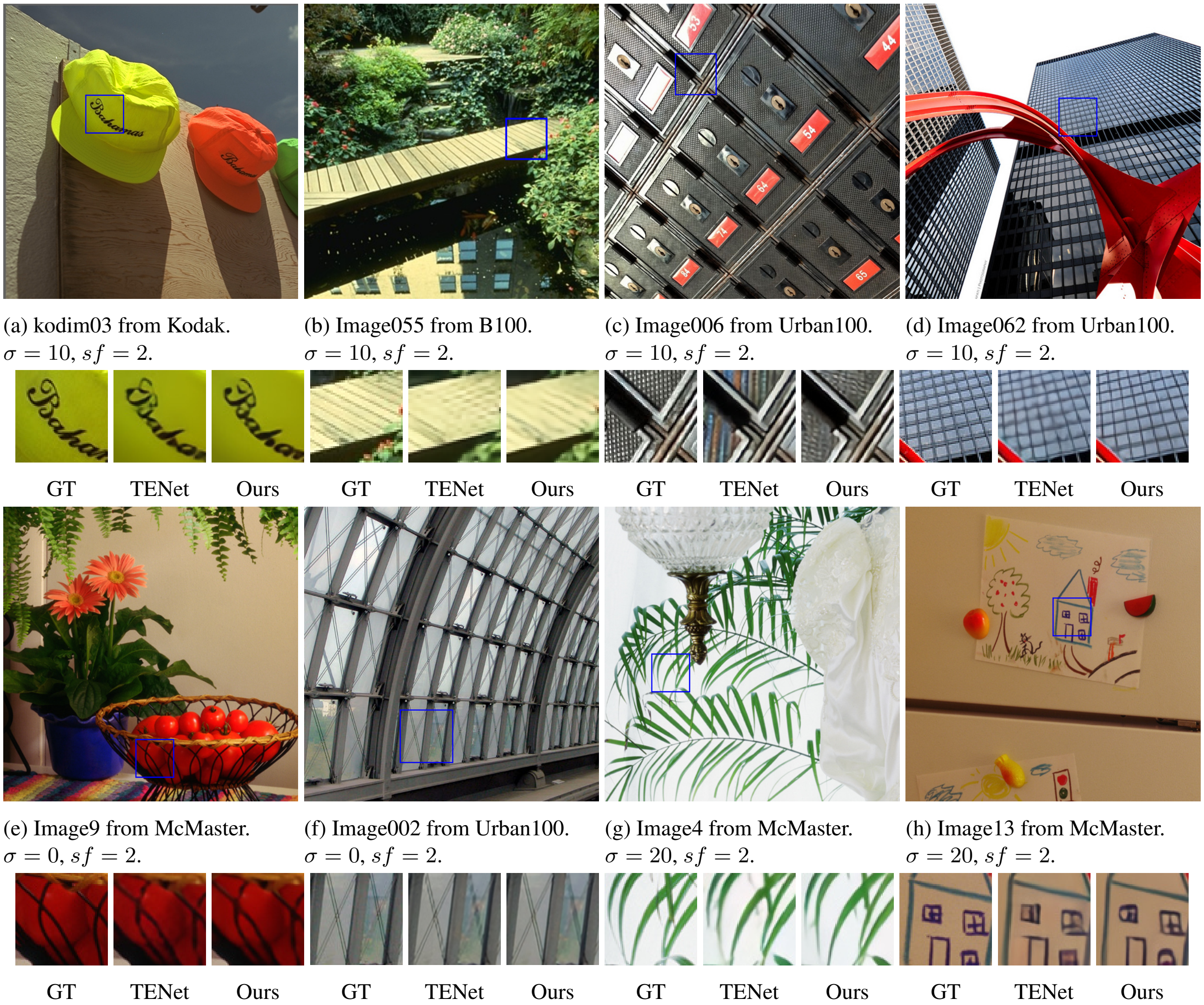
1. Introduction

Image demosaicing, denoising, and super-resolution (SR) are classical image restoration problems. With the recent advancement of deep convolutional neural networks (CNNs) and their application in image restoration, several deep learning-based methods achieve the state-of-the-art (SOTA) performance [19, 42, 37, 47].

In many practical applications an acquired image is distorted by multiple degradations, thus the above mentioned individual image restoration problems have to be solved simultaneously. A most natural choice is to apply best methods of individual image restoration tasks in a sequence. However, the existing solutions are not ideal. Addressing a problem of image denoising, most of the algorithms smooth out high-frequency content, such as image details and texture, while eliminating noise in flat areas. Image demosaicing and super-resolution algorithms often introduce color artifacts especially in the texture regions and around image edges. Thus, a sequential application of the individual image restoration methods will result in an accumulation of errors produced by the individual methods. Another drawback of the sequential methods is an increased complexity of a solution (considering both speed and memory).

As an alternative to this, joint solutions for the combined problems have been proposed in the literature [3, 6, 7, 8, 18, 21, 26, 36, 44, 49]. However, the problem of finding a joint solution for a triplet of problems of image demosaicing, denoising and SR has received much less attention [26, 29]. In 2019, Qian *et al*. [29] proposed a trinity network (TENet) to jointly solve this composite problem. Although the TENet is an end-to-end network, the execution order of different tasks is fixed. To this end, Qian *et al*. have divided the network into two modules and calculated the middle loss to supervise the functionality of the first module and optimize the network. Recently, Liu *et al*. proposed another solution to the joint problem, SGNet [26]. In order to improve the performance of demosaicing, SGNet introduces two self-guidance methods, the green channel guidance and the density map guidance.

In this paper, we comprehensively study various solutions of this combined problem. First, in subsection 3.1, we adjust the execution order, and investigate possible joint solutions under this execution order. Then, in subsection 3.2, we propose an end-to-end learning for the combined problem by designing a very deep convolutional neural network JD_ND_MSR. Differently from TENet and SGNet, our network uses the residual channel connection block (RCAB) [47] instead of residual-in-residual dense block (RRDB) [37] as the basic block (see subsection 6.1). We have carried out numerous experiments and demonstrate that the proposed method outperforms other joint solutions both quantitatively and qualitatively (Section 4). To further optimize the proposed network, different loss functions are utilized, and the comparative analysis of the resulting solutions is demonstrated in subsection 5.1. A comparison with the state-of-the-art method TENet [29] and the ablation study (see Fig. 1) are presented in subsection 5.2 and Section 6, respectively.



(a) kodim03 from Kodak. $\sigma = 10$, $sf = 2$.

(b) Image055 from B100. $\sigma = 10$, $sf = 2$.

(c) Image006 from Urban100. $\sigma = 10$, $sf = 2$.

(d) Image062 from Urban100. $\sigma = 10$, $sf = 2$.

(e) Image9 from McMaster. $\sigma = 0$, $sf = 2$.

(f) Image002 from Urban100. $\sigma = 0$, $sf = 2$.

(g) Image4 from McMaster. $\sigma = 20$, $sf = 2$.

(h) Image13 from McMaster. $\sigma = 20$, $sf = 2$.

Figure 1: Qualitative comparison between the SOTA model TENet and the proposed $JD_ND_MSR^+$. sf means the scale factor. The noise levels (σ) of (e-f), (a-d), (g-h) are 0, 10 and 20, respectively.

The main contributions of this paper are listed below.

1. We propose an end-to-end network ($JD_ND_MSR^+$) based on residual channel attention blocks for joint image demosaicing, denoising and super-resolution. This network is universal: one can turn off denoising and/or super-resolution operations of the network by setting the noise level parameter to 0 and the scale factor parameter to 1.

2. We systematically compare our $JD_ND_MSR^+$ with diverse solutions to the joint problem. The quantitative and qualitative experimental results on the benchmark datasets show that the proposed method not only surpasses other solutions, but also outperforms the state-of-the-art, including cases when denoising or super-resolution operations are not performed.

2. Related work

Denoising. The advanced image denoising methods can be classified into two main categories: model-based and deep learning-based methods. BM3D [4], often regarded as a denoising benchmark, belongs to the first category. In 2017, Zhang *et al*. applied a deep convolutional neural network (CNN) , called DnCNN [42]. DnCNN adopts residual learning and batch normalization on CNN for blind Gaussian denoising and attains top performance. Later on, many other machine-learning based methods of image denoising have appeared [2, 15, 38, 42, 43].

Demosaicing. To reduce manufacturing costs, most digital camera sensors capture only one color (red, green and blue) at each pixel. The camera sensor is covered by the color filter arrays (CFAs). Image demosaicing is the process of interpolating full-resolution color image from incomplete color samples output by an image sensor. Most demosaicing methods have been specifically designed for the Bayer CFA which is the most popular CFA. Existing algorithms can be also classified into two categories: model-based methods [12, 27, 31, 45], which recover images based on mathematical models and image priors in the spatial-spectral domain; and learning-based methods [11, 32], based on process mapping learned from abundant training data. The deep learning methods [9, 17, 33] of image demosaicing also attain the state-of-the-art performance.

Single image super-resolution. Single image super-resolution aims at recovering a high-resolution (HR) image from its corresponding low-resolution (LR) image. The emergence of convolutional neural network has made the performance of super-resolution methods advance by leaps and bounds. In 2015, Dong *et al.* proposed SRCNN [5], which utilizes a three-layers CNN in a single image super-resolution task. Inspired by VGG-net, Kim *et al.* have presented a very deep residual learning super-resolution network, VDSR [19]. To reduce the occupation of memory and accelerate the speed of computation, Shi *et al.* have introduced a sub-pixel CNN ESPCN [30] to upscale feature maps to the desired solution. In 2017, Ledig *et al.* [23] have applied ResNet architecture in SR and proposed a SRResNet scheme. EDSR [24] further ameliorate the residual block and develop a very deep and wide CNN to enhance the performance of SR. In 2018, Zhang *et al.* have presented RDN, which is a residual dense network for SR. They have also proposed an attention-based network, RCAN [47], which introduces the channel attention into residual blocks (RCAB). Wang *et al.* [37] have proposed a perceptual-driven method ESRGAN based on the proposed Residual-in-Residual Dense Block (RRDB). In 2020, Liu *et al.* [25] proposed the RFANet by improving the chain of residual modules and adding an enhanced spatial attention (ESA) block at the end of each residual block.

Joint solutions. In practical applications, multiple image restoration problems appear simultaneously, resulting in the combined problems that one needs to solve. Recently, the combined solutions to the mixture problem of multiple image distortions replace traditional sequential solutions. Examples are joint denoising and demosaicing [3, 6, 8, 10, 16, 18, 21], joint demosaicing and SR [7, 36, 39, 49], and joint denoising and SR [44, 50]. However, a research on the triplet of denoising, demosaicing and SR is still lacking a special attention. In 2019, Qian *et al.* [29] proposed a trinity network to jointly solve this composite problem. In 2020, Liu *et al.* proposed the SGNet [26] for joint image demosaicing and super-resolution, which also can handle the mixture problem of denoising, demosaicing and SR.

In this paper, we propose the end-to-end solution of demosaicing, denoising and SR, JD_ND_MSR, and compared it with the sequential application of SOTA methods for each of these sub-problems, as well as with the state-of-the-art method to solve this mixture problem.

3. Proposed method

In what follows, we first study the execution order of image demosaicing, denoising and super-resolution. Then, solutions of this execution order are analysed. Later, we propose a deep CNN for the mixture problem. Note that we only consider the CNN-based methods in this paper.

3.1. Joint solutions

For the mixture problem of image demosaicing, denoising and super-resolution, a clean high-resolution color image I_{HR} should be estimated from its noisy low-resolution raw image $I_{LR_M^N}$. For the execution order, demosaicing should follow denoising, like in [29], to avoid complications in filtering correlated noise after demosaicing. In addition, the demosaicing should be performed before super-resolution because the correlation across color channels can be exploited when super-resolving color image. Besides this reason, performing super-resolution on raw image will destroy the original mosaic pattern, which increases the difficulty of demosaicing. Therefore, for the fixed execution order: $D_N \rightarrow D_M \rightarrow SR$, the first solution is to sequentially utilize three targeted methods to solve the corresponding image restoration problems one by one:

$$\widehat{I_{HR}} = M_{SR}(M_{D_M}(M_{D_N}(I_{LR_M^N}))). \tag{1}$$

where M denotes image restoration method, D_N, D_M, SR denote denoising, demosaicing and super-resolution, respectively, and $\widehat{I_{HR}}$ is the estimation of high-resolution image I_{HR}.

Naturally, another approach is to combine two image restoration tasks and then execute the remaining one:

$$\widehat{I_{HR}} = M_{SR}(M_{JD_ND_M}(I_{LR_M^N})), \tag{2}$$

and

$$\widehat{I_{HR}} = M_{JD_MSR}(M_{D_N}(I_{LR_M^N})), \tag{3}$$

where J indicates joint processing. Similarly, the third solution is a fully combined end-to-end solution:

$$\widehat{I_{HR}} = M_{JD_ND_MSR}(I_{LR_M^N}). \tag{4}$$

A comparison of the solutions (Eqn. 1-4) is presented in Section 4.2.

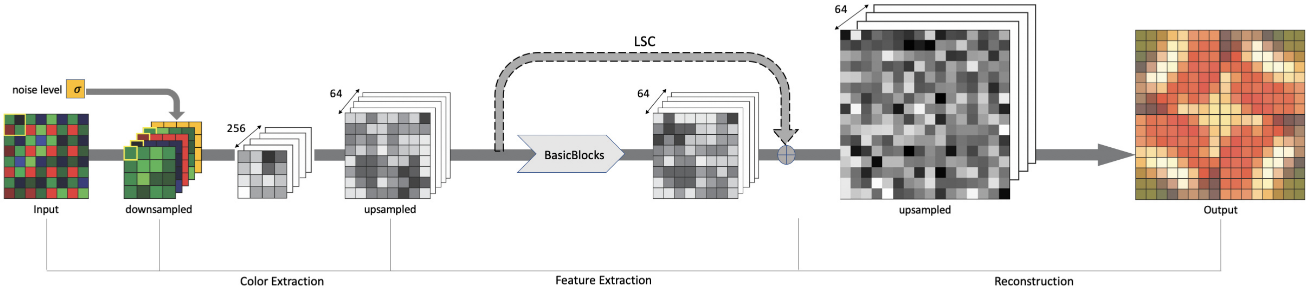


Figure 2: The featured visualization of the proposed deep joint denoising, demosaicing and super-resolution network JD_ND_MSR.

3.2. Network architecture

The proposed end-to-end solution of the mixture problem is based on the deep CNN-based network, JD_ND_MSR shown in Fig. 2, and consists of three parts: color extraction, feature extraction and reconstruction. Inspired by the method presented in [8], the Bayer input is first reshaped to a quarter-resolution multi-channel image, which is concatenated with the noise level estimation input. In this paper, we assume that a noise level is known in advance or is properly estimated, thus, one can parametrize a network with it. One way is to add a noise level input σ, and replicate it spatially, concatenating with the packed mosaic vector. Every layer downstream depends on it, which effectively parametrizes the learned filters. The color extraction step includes a convolutional layer with a large filter (256), and a transposed convolution layer to upscale the feature maps to the prime resolution. Upsampling the features before the next module improves the performance of the network (Section 6.2). The feature extraction module is composed of several basic blocks and a Long Skip Connection (LSC). The basic block can be any effective block applied in SOTA, such as the residual block (RB) [24], the residual-in-residual dense block (RRDB) [37], or the residual group (RG) with residual channel attention block (RCAB) [47]. Through the ablation study of the basic blocks (see Section 6.1), we have chosen the RCAB to be our basic block of feature extraction module. We utilize 4 residual groups in the JD_ND_MSR network structure, each group including 20 RCABs. In the reconstruction part, the transposed convolution layer is used again to convert the extracted features into full resolution features. The following is the final convolutional layer to generate the desired resolution color image. The proposed JD_ND_MSR can be changed to a noise-free version JD_MSR by removing the noise level input σ ($\sigma = 0$). The experiments presented in Section 5 will demonstrate that the proposed JD_ND_MSR and JD_MSR achieve notable performance improvement in comparison with the other solutions including the state-of-the-art.

4. Experiments

4.1. Settings

For the training, we have applied Nvidia Tesla P100 GPU with 16 GB memory from the Tampere University TCSC Narvi computing cluster. All experiments run on a Linux computer with 3.4 GHz Intel i7-3770 CPU, 32 GB of RAM, and Nvidia GTX 1050Ti GPU with 4GB of memory.

Dataset. For network training and validation, we used publicly available dataset DIV2K [1] consisting of 900 2K resolution images (800 for training, 100 for validation). We compared different joint solutions on two public datasets, McMaster [46] and Kodak, widely used in the papers on image restoration [8, 19, 29, 35, 40, 42].

Data preprocessing. For data preprocessing of denoising, noisy input images are generated by adding Gaussian noise with the noise levels (σ) 10, 20 and 30. For data preprocessing of demosaicing, we mosaic the color image to a single-channel image in the Bayer CFA pattern. For data preprocessing of super-resolution, the HR image is BICUBIC down scaled with the scale factors (SF) 2.

Training details. Data augmentation is performed on images, which are randomly rotated by 90°, 180°, 270° and flipped horizontally. For each training epoch, the mini-batch size is 16, and the patch size is 64×64. All models are implemented in Python with the platform Keras. For the optimization of network parameters, we use Adam [20] with $\beta_1 = 0.9, \beta_2 = 0.999$ and the learning rate is initialized to 0.001. All training continue 100 epochs. There are 2000 training steps and 200 validation steps in each epoch. For the first 10 epochs, the learning rate is constant, then the learning rate is decreased by 10 times for the remaining 90 epochs. Only a model with the smallest validation loss is saved.

Loss function. The proposed JD_ND_MSR is optimized with different loss functions. Given a training set $\{I^i_{LR^N_M}, I^i_{HR}\}^N_{i=1}$, which contains N low-resolution inputs and corresponding high-resolution counterparts, the goal of

Table 1: Quantitative comparison of different solutions on the mixture problem of joint denoising, demosaicing and super-resolution using datasets Kodak and McMaster [46]. $*$ represents the model is re-trained by our. The noise level is 10 and the scale factor is set to 2. The best, second and third best results are highlighted with red, blue and green, respectively. The efficiency is computed as an average time to process an image.

Solution type	Pipeline	McMaster		Kodak		Parameters	Efficiency
		cPSNR	SSIM	cPSNR	SSIM		
$D_N \rightarrow D_M \rightarrow SR$	DnCNN→DJDD→VDSR	25.99	0.8522	26.18	0.7868	21.1MB	0.36s
	DnCNN* →DJDD* →VDSR*	29.14	0.9248	28.53	0.8913		
Joint $D_N D_M \rightarrow SR$	DJDD→VDSR	28.40	0.9248	28.13	0.8812	14.2MB	0.23s
	DJDD* →VDSR*	28.88	0.9212	28.43	0.8887		
D_N →Joint $D_M SR$	DnCNN→ $JD_M SR$	25.91	0.8522	26.11	0.7846	85.1MB	0.77s
	DnCNN* → $JD_M SR^*$	29.51	0.9293	28.75	0.8948		
Joint $D_N D_M SR$	$JD_N D_M SR$	29.34	0.9274	28.80	0.8942	78.2MB	0.64s
	$JD_N D_M SR^+$	29.56	0.9296	28.80	0.8965		

training $JD_N D_M SR$ is to minimize the loss function:

$$\mathcal{L}(\Theta) = \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}(JD_N D_M SR(I^i_{LR^N_M}), I^i_{HR}). \qquad (5)$$

where Θ denotes the parameter set of $JD_N D_M SR$. The models in this part are trained with mean squared error (MSE). We also further optimize our network by training it with different error criteria and comparing the results by different loss functions (see Section 5.1).

4.2. Comparison of solutions

In this section, we compare the joint solutions, presented in Eqns. (1)-(4), of the mixture problem of image demosaicing, denoising and super-resolution. Except $M_{JD_M SR}$ in Eqn. (3), solved by the proposed $JD_M SR$, other methods in Eqn. (1)-(3) are replaced by the state-of-the-art image restoration networks, DnCNN [42], DJDD [8], and VDSR [19], used for denoising, demosaicing and super-resolution, respectively. It should be noted that there are two versions of DJDD (for noisy and noise-free inputs). The noisy version model is used in Eqn. (2) for joint demosaicing and denoising ($M_{JD_N D_M}$). In contrast, the noise-free version is adopted in Eqn. (1) for M_{D_N}. Here we mainly focus on the comparison of various joint solutions, rather than aiming at obtaining state-of-the-art results. Therefore, we chose simple yet effective methods instead of computationally more demanding ones with better performances. Similar to [41], we introduce a transfer-learning strategy to further improve $JD_N D_M SR$ (we name the transfer-learning method as $JD_N D_M SR^+$). $JD_N D_M SR^+$ transfers the learned parameters from the trained model of $JD_M SR$ for joint $\times 2$ super-resolution and image demosaicing. The details of the pre-trained $JD_M SR$ model and the ablation study of transfer learning are provided in the supplementary material.

Quantitative results. Quantitative analysis was performed with cPSNR and SSIM metrics, by calculating them on full RGB image. The results are averaged over whole dataset. For super-resolved image, the borders of the image are shaved off, with the scaling factor as the width of the shaved border.

Table 1 shows the quantitative comparison of all solutions for joint image demosaicing, denoising and super-resolution. We fix a noise level to 10 and a scale factor to 2. The loss function used in this comparison is MSE. Since CNN models are sensitive to the input data, all models in (Eqn. (1)-(3)) are re-trained with the specific input and output pairs. In order to reduce the interaction among different tasks, a model should input the results of the previous model and try to correct the errors produced by the previous processing at the same time. Comparing to other solutions, our combined solution $JD_N D_M SR^+$ performs better on both datasets. Even without transfer-learning, our closest combined solution $JD_N D_M SR$ also outperforms most of the compared solutions. On the other hand, the re-trained models can obtain better performance than the directly exploiting trained models. We also presented the qualitative results in Fig. 3. Our $JD_N D_M SR^+$ not only eliminates the noise but also recovers more details in high frequency region.

Effects of combined solution. In Table 1, one can observe that our $JD_N D_M SR$ is the third best joint solution. In contrast, the specific re-trained models of solution in Eqn. (3) achieves the second best performance. In addition, the fourth best solution is the retrained version of Eqn. (1). These two solutions both begin from the specific re-trained DnCNN model. Therefore, a specific trained DnCNN model can support a good start for joint denoising, demosaicing and SR.

However, our $JD_N D_M SR$ can achieve a comparable performance by the additional noise level estimation input. Our $JD_N D_M SR^+$ demonstrates a superior performance.

This observation indicates that the combined solution can avoid an accumulation of errors. According to Table 1, the combined solution, $JD_ND_MSR^+$, outperforms other solutions in consideration of performance, storage, and computation efficiency.

5. Optimization

5.1. Comparison on cost functions

In subsection 4.2, the proposed $JD_ND_MSR^+$ surpasses other joint solutions both quantitatively and qualitatively. In order to further optimize $JD_ND_MSR^+$, we train several models with different cost functions besides MSE, including MAE, SSIM, MS-SSIM, Mix1. Inspired by [48], the Mix1 cost function is defined as $\alpha\mathcal{L}_{MS-SSIM}+(1-\alpha)\mathcal{L}_1$[1]. These five models are compared on three evaluation metrics: cPSNR, SSIM, and MS-SSIM. The results of their comparison on McMaster and Kodak datasets are shown in Table 2. As one can see, the model trained with MAE (mean absolute error) cost function attains the best performance for all image quality metrics and on both datasets. Compared with the model trained with MSE, the cPSNR values of MAE version is improved by 0.25dB on two datasets.

Table 2: Quantitative comparison of different cost functions. The results are averaged both on McMaster and Kodak. The noise level is 10 and the scale factor is 2. For cPSNR, SSIM, MS-SSIM, the value reported here has been obtained as an average of the three color channels. Best results are shown in bold.

Metric	Training cost function				
	MSE	MAE	SSIM	MS-SSIM	Mix1
cPSNR	28.48	**28.73**	26.64	26.64	27.36
SSIM	0.8991	**0.9041**	0.8513	0.8531	0.8733
MS-SSIM	0.9452	**0.9487**	0.9312	0.9326	0.9297

5.2. Comparison with State-of-the-Art

In this section, we compare the proposed $JD_ND_MSR^+$-MAE with the state-of-the-art method TENet [29] on four datasets with four noise levels (see Table 3). For a fair comparison, we re-trained the TENet network and our $JD_ND_MSR^+$-MAE on both DIV2K and Flickr2K [34] datasets with ×2 scale factor and the noise level randomly sampled from [0, 20]. In addition to McMaster and Kodak datasets, we also test them on B100 [28] and Urban100 [14] datasets, which are often applied in the comparison of different super-resolution methods. The dataset B100 contains 100 human segmented natural images, and the dataset Urban100 contains 100 urban images with many similar structures. For the pre-processing of the test images, the scale factor is set to 2 and the noise levels to 0, 10, 20 and 30. We use cPSNR and SSIM metrics for the quantitative evaluation. As shown in Table 3, our model outperforms the TENet over all noise levels on all datasets. We also present the visual comparison both on noisy and noise-free versions in Fig. 1. In comparison with the resulting images of TENet, our $JD_ND_MSR^+$-df2k enables to reconstruct the high resolution images more accurately with less blur and less color artifacts. Although $JD_ND_MSR^+$-df2k can handle higher noise ($\sigma > 20$), more details are eliminated along with the noise (see our supplementary material).

Joint denoising and demosaicing. As it was mentioned above, our $JD_ND_MSR^+$ can switch off denoising by setting the parameter σ to 0. In addition, the super-resolution can also be turned off by setting the scale factor to 1. Based on this idea, we train our JD_ND_MSR network with scale factor 1 on DIV2K dataset, named as JD_ND_M. We compare the JD_ND_M with three state-of-the-art methods: DJDD [8], Kokkinos [22], and SGNet [26]. The comparison on three datasets with four noise levels is shown in Table 4[2]. This table demonstrates that the performance of our JD_ND_M surpasses the state-of-the-art joint denoising and demosaicing methods on both noisy and noise-free data. When the noise level of JD_ND_M is 0, the model works as demosaicing only, *i.e.* denoising and super-resolution are turned off. Therefore, our JD_ND_MSR network can be adjusted according to different requirements, including switching on/off super-resolution, switching on/off denoise, and setting scale factor and noise level. Meanwhile, our JD_ND_MSR attains favorable performance on different mixture problems, such as denoising and demosaicing (Table 4), demosaicing and super-resolution (Table 3), and denoising, demosaicing and super-resolution (Table 3).

6. Ablation study

6.1. Basic blocks of feature extraction module

In order to study the effects of each component in the proposed model $JD_ND_MSR^+$, we gradually modify the baseline $JD_ND_MSR^+$ model and compare their differences. The investigation starts from the selection of the basic blocks of feature extraction module. We compare three types of residuals in the residual blocks: RRDB [37], RCAB [13] and RAB [47]. For a fair comparison, we tuned the number of three basic blocks to keep all networks to have similar parameters (Table 5). The performance curves of different basic blocks is shown in Fig. 4. With a similar model size, the network with RCAB blocks performs bet-

[1] We tested a few different values for α, and set $\alpha = 0.1$

[2] For fair comparison, the models we tested for noise-free data are the noisy version supported by the authors. The max noise level of Kokkinos is 10. Since the public pre-trained model are not available, the values of SGNet are from the corresponding paper.

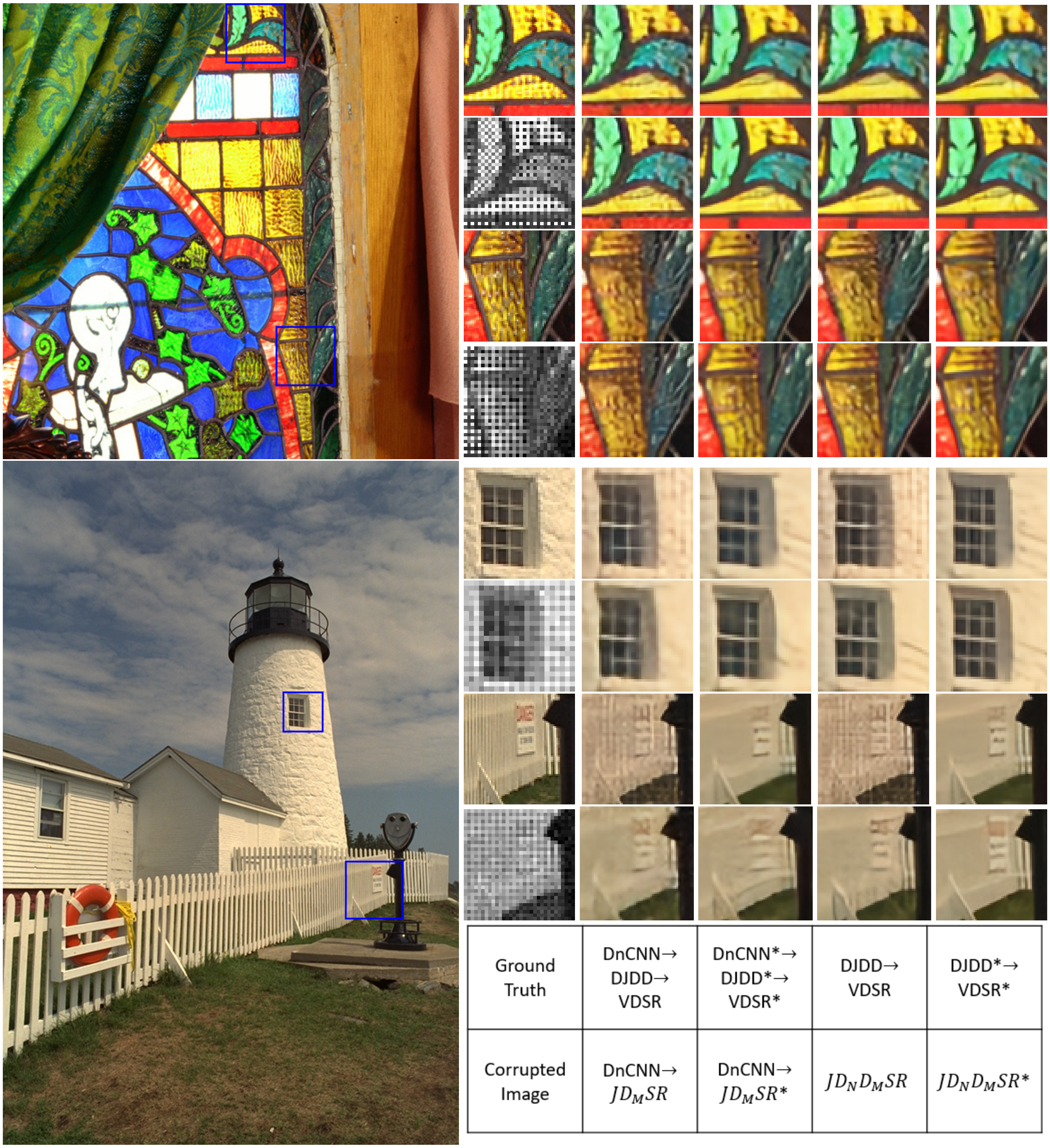


Figure 3: Visual comparison of the joint solutions of denoising, demosaicing and super-resolution. The scale factor is 2 and noise level is 10. The upper half part is the Image01 from McMaster dataset. The lower half part is the kodim19 from Kodak dataset. The sequence of ground truth image, corrupted image and resulting images corresponds to the illustration in the lower right corner.

ter than those with the other two basic blocks. Besides the selection of the basic blocks, we find that LSC and transfer learning can also improve the performance of the network (See our supplementary material).

Table 3: Quantitative comparison for joint denoising, demosaicing and super-resolution. The evaluation metrics are cPSNR and SSIM. The best values are shown in bold. The scale factor is 2 and the noise levels are 0, 10, 20 and 30.

Noise level	McMaster		Kodak		B100		Urban100	
	TENet	$JD_ND_MSR^+$	TENet	$JD_ND_MSR^+$	TENet	$JD_ND_MSR^+$	TENet	$JD_ND_MSR^+$
0	31.48/0.9574	**32.59/0.9652**	30.80/0.9386	**31.49/0.9456**	29.24/0.9200	**29.87/0.9283**	28.05/0.9225	**28.99/0.9331**
10	29.28/0.9269	**29.66/0.9315**	28.70/0.8963	**28.85/0.8982**	27.25/0.8711	**27.37/0.8725**	26.53/0.8872	**26.89/0.8922**
20	27.29/0.8943	**27.54/0.9000**	27.04/0.8558	**27.13/0.8595**	25.60/0.8230	**25.67/0.8250**	24.98/0.8464	**25.22/0.8524**
30	25.88/0.8636	**26.11/0.8724**	25.90/0.8226	**26.03/0.8309**	24.50/0.7854	**24.57/0.7924**	23.75/0.8072	**24.01/0.8156**

Table 4: Quantitative comparison for joint denoising and demosaicing. The best values are shown in bold.

Method	Noise level	McMaster cPSNR	McMaster SSIM	Kodak cPSNR	Kodak SSIM	Urban100 cPSNR	Urban100 SSIM
DJDD[8]	5	35.48	0.9775	36.21	0.9749	34.04	0.9728
kokkinos[22]		32.41	0.9601	34.65	0.9665	33.09	0.9654
SGNet[26]		–	–	–	–	34.54	0.9533
JD_ND_M		**36.05**	**0.9805**	**36.87**	**0.9782**	**35.07**	**0.9767**
DJDD[8]	10	33.14	0.9629	33.22	0.9537	31.80	0.9547
kokkinos[22]		29.30	0.9253	30.70	0.9215	30.02	0.9246
SGNet[26]		–	–	–	–	32.14	0.9229
JD_ND_M		**33.74**	**0.9677**	**33.90**	**0.9599**	**32.83**	**0.9619**
DJDD[8]	15	31.49	0.9478	31.43	0.9323	30.14	0.9356
kokkinos[22]		25.98	0.8517	27.17	0.8295	26.74	0.8492
SGNet[26]		–	–	–	–	30.37	0.8923
JD_ND_M		**32.11**	**0.9550**	**32.05**	**0.9420**	**31.25**	**0.9477**
DJDD[8]	0	37.90	0.9880	40.33	0.9918	36.47	0.9858
kokkinos[22]		33.82	0.9655	37.64	0.9815	33.94	0.9570
JD_ND_M		**38.85**	**0.9904**	**42.23**	**0.9947**	**38.34**	**0.9895**

Table 5: Performance comparison of different basic blocks. The performance is the best cPSNR value on McMaster dataset.

Basic block	Amounts	Total parameters	Performance
RAB	7	3,247,063	28.55
RRDB	6	3,299,031	28.59
RCAB	40	3,504,855	28.97

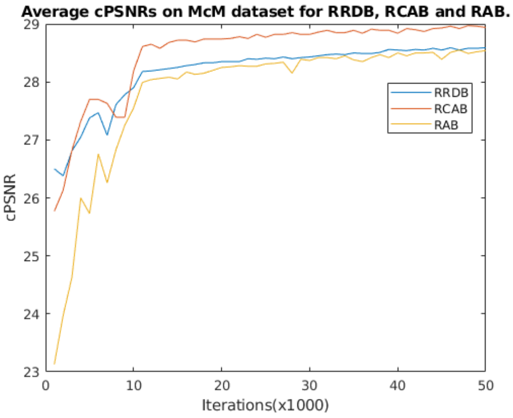


Figure 4: The comparison of convergence curves of different basic modules.

6.2. Color extraction module

Fig. 2 shows that the color extraction module of our network is composed of two layers: a convolutional layer with a large filter (256) for color extraction (CE) and a deconvolutional layer for upsampling (UP1). In this part, we prove the importance of this module. Table 6 displays the effect of the CE layer and the position of the deconvolutional layer. When the features are upsampled before feature extraction, performance of the network improves by 0.19 dB compared to the case when the features are upsampled after feature extraction. The CE layer can also provide a small performance improvement.

Table 6: Investigation of color extraction module. The models are tested on McMaster dataset. The scale factor is set to 2 and the noise sigma is 10.

CE?	✗	✓	✗	✓
UP1?	After	After	Before	Before
cPSNR on McMaster	28.86	28.87	29.05	**29.07**

7. Conclusion

We have systematically compared possible solutions of the joint problem of image demosaicing, denoising and super-resolution, under fixed execution order. Extensive experiments have demonstrated that the proposed end-to-end learning-based solution, $JD_ND_MSR^+$ surpasses others, both quantitatively and qualitatively. Besides, the performance of $JD_ND_MSR^+$ is improved by training with the mean absolute error cost function used instead of mean square error. The performance of this optimized model surpassed the state-of-the-art method TENet on four benchmark datasets for noisy and noise-free data. In addition, the denoising operation and the super-resolution operation of the proposed network can be turned off (by setting the noise level to 0 and the scale factor to 1). When the super-resolution operation is switched off, our JD_ND_M model for joint denoising and demosaicing outperforms the state-of-the-art methods. In the future, we will explore more prior information to further improve the performance of joint image demosaicing, denoising and super-resolution.

References

[1] Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, July 2017. 4

[2] Harold C Burger, Christian J Schuler, and Stefan Harmeling. Image denoising: Can plain neural networks compete with bm3d? In *2012 IEEE conference on computer vision and pattern recognition*, pages 2392–2399. IEEE, 2012. 2

[3] Laurent Condat and Saleh Mosaddegh. Joint demosaicking and denoising by total variation minimization. In *2012 19th IEEE International Conference on Image Processing*, pages 2781–2784. IEEE, 2012. 1, 3

[4] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Transactions on image processing*, 16(8):2080–2095, 2007. 2

[5] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, 38(2):295–307, 2015. 3

[6] Thibaud Ehret, Axel Davy, Pablo Arias, and Gabriele Facciolo. Joint demosaicing and denoising by overfitting of bursts of raw images. *arXiv preprint arXiv:1905.05092*, 2019. 1, 3

[7] Sina Farsiu, Michael Elad, and Peyman Milanfar. Multi-frame demosaicing and super-resolution from undersampled color images. In *Computational Imaging II*, volume 5299, pages 222–233. International Society for Optics and Photonics, 2004. 1, 3

[8] Michaël Gharbi, Gaurav Chaurasia, Sylvain Paris, and Frédo Durand. Deep joint demosaicking and denoising. *ACM Transactions on Graphics (TOG)*, 35(6):191, 2016. 1, 3, 4, 5, 6, 8

[9] Jinwook Go, Kwanghoon Sohn, and Chulhee Lee. Interpolation using neural networks for digital still cameras. *IEEE Transactions on Consumer Electronics*, 46(3):610–616, 2000. 3

[10] Yu Guo, Qiyu Jin, Gabriele Facciolo, Tieyong Zeng, and Jean-Michel Morel. Residual learning for effective joint demosaicing-denoising. *arXiv preprint arXiv:2009.06205*, 2020. 3

[11] Fang-Lin He, Yu-Chiang Frank Wang, and Kai-Lung Hua. Self-learning approach to color demosaicking via support vector regression. In *2012 19th IEEE International Conference on Image Processing*, pages 2765–2768. IEEE, 2012. 3

[12] Keigo Hirakawa and Thomas W Parks. Adaptive homogeneity-directed demosaicing algorithm. *IEEE Transactions on Image Processing*, 14(3):360–369, 2005. 3

[13] Guanqun Hou, Yujiu Yang, and Jing-Hao Xue. Residual dilated network with attention for image blind denoising. In *2019 IEEE International Conference on Multimedia and Expo (ICME)*, pages 248–253. IEEE, 2019. 6

[14] Jia-Bin Huang, Abhishek Singh, and Narendra Ahuja. Single image super-resolution from transformed self-exemplars. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5197–5206, 2015. 6

[15] Viren Jain and Sebastian Seung. Natural image denoising with convolutional networks. In *Advances in neural information processing systems*, pages 769–776, 2009. 2

[16] Qiyu Jin, Gabriele Facciolo, and Jean-Michel Morel. A review of an old dilemma: Demosaicking first, or denoising first? In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 514–515, 2020. 3

[17] Oren Kapah and Hagit Zabrodsky Hel-Or. Demosaicking using artificial neural networks. In *Applications of Artificial Neural Networks in Image Processing V*, volume 3962, pages 112–120. International Society for Optics and Photonics, 2000. 3

[18] Daniel Khashabi, Sebastian Nowozin, Jeremy Jancsary, and Andrew W Fitzgibbon. Joint demosaicing and denoising via learned nonparametric random fields. *IEEE Transactions on Image Processing*, 23(12):4968–4981, 2014. 1, 3

[19] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1646–1654, 2016. 1, 3, 4, 5

[20] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 4

[21] Teresa Klatzer, Kerstin Hammernik, Patrick Knobelreiter, and Thomas Pock. Learning joint demosaicing and denoising based on sequential energy minimization. In *2016 IEEE International Conference on Computational Photography (ICCP)*, pages 1–11. IEEE, 2016. 1, 3

[22] Filippos Kokkinos and Stamatios Lefkimmiatis. Deep image demosaicking using a cascade of convolutional residual denoising networks. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 303–319, 2018. 6, 8

[23] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4681–4690, 2017. 3

[24] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 136–144, 2017. 3, 4

[25] Jie Liu, Wenjie Zhang, Yuting Tang, Jie Tang, and Gangshan Wu. Residual feature aggregation network for image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2359–2368, 2020. 3

[26] Lin Liu, Xu Jia, Jianzhuang Liu, and Qi Tian. Joint demosaicing and denoising with self guidance. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2240–2249, 2020. 1, 3, 6, 8

[27] Henrique S Malvar, Li-wei He, and Ross Cutler. High-quality linear interpolation for demosaicing of bayer-

patterned color images. In *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 3, pages iii–485. IEEE, 2004. 3

[28] David Martin, Charless Fowlkes, Doron Tal, Jitendra Malik, et al. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. Iccv Vancouver:, 2001. 6

[29] Guocheng Qian, Jinjin Gu, Jimmy S Ren, Chao Dong, Furong Zhao, and Juan Lin. Trinity of pixel enhancement: a joint solution for demosaicking, denoising and super-resolution. *arXiv preprint arXiv:1905.02538*, 2019. 1, 3, 4, 6

[30] Wenzhe Shi, Jose Caballero, Ferenc Huszár, Johannes Totz, Andrew P Aitken, Rob Bishop, Daniel Rueckert, and Zehan Wang. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1874–1883, 2016. 3

[31] Chung-Yen Su. Highly effective iterative demosaicing using weighted-edge and color-difference interpolations. *IEEE Transactions on Consumer Electronics*, 52(2):639–645, 2006. 3

[32] Jian Sun and Marshall F Tappen. Separable markov random field model and its applications in low level vision. *IEEE transactions on image processing*, 22(1):402–407, 2012. 3

[33] Nai-Sheng Syu, Yu-Sheng Chen, and Yung-Yu Chuang. Learning deep convolutional networks for demosaicing. *arXiv preprint arXiv:1802.03769*, 2018. 3

[34] Radu Timofte, Eirikur Agustsson, Luc Van Gool, Ming-Hsuan Yang, and Lei Zhang. Ntire 2017 challenge on single image super-resolution: Methods and results. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 114–125, 2017. 6

[35] Radu Timofte, Vincent De Smet, and Luc Van Gool. A+: Adjusted anchored neighborhood regression for fast super-resolution. In *Asian conference on computer vision*, pages 111–126. Springer, 2014. 4

[36] Patrick Vandewalle, Karim Krichane, David Alleysson, and Sabine Süsstrunk. Joint demosaicing and super-resolution imaging from a set of unregistered aliased images. In *Digital Photography III*, volume 6502, page 65020A. International Society for Optics and Photonics, 2007. 1, 3

[37] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 0–0, 2018. 1, 3, 4, 6

[38] Jun Xu, Lei Zhang, and David Zhang. A trilateral weighted sparse coding scheme for real-world image denoising. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 20–36, 2018. 2

[39] Xuan Xu, Yanfang Ye, and Xin Li. Joint demosaicing and super-resolution (jdsr): network design and perceptual optimization. *IEEE Transactions on Computational Imaging*, 2020. 3

[40] Chih-Yuan Yang and Ming-Hsuan Yang. Fast direct super-resolution by simple functions. In *Proceedings of the IEEE international conference on computer vision*, pages 561–568, 2013. 4

[41] Ke Yu, Chao Dong, Chen Change Loy, and Xiaoou Tang. Deep convolution networks for compression artifacts reduction. *arXiv preprint arXiv:1608.02778*, 2016. 5

[42] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Transactions on Image Processing*, 26(7):3142–3155, 2017. 1, 2, 4, 5

[43] Kai Zhang, Wangmeng Zuo, and Lei Zhang. Ffdnet: Toward a fast and flexible solution for cnn-based image denoising. *IEEE Transactions on Image Processing*, 27(9):4608–4622, 2018. 2

[44] Kai Zhang, Wangmeng Zuo, and Lei Zhang. Learning a single convolutional super-resolution network for multiple degradations. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3262–3271, 2018. 1, 3

[45] Lei Zhang and Xiaolin Wu. Color demosaicking via directional linear minimum mean square-error estimation. *IEEE Transactions on Image Processing*, 14(12):2167–2178, 2005. 3

[46] Lei Zhang, Xiaolin Wu, Antoni Buades, and Xin Li. Color demosaicking by local directional interpolation and nonlocal adaptive thresholding. *Journal of Electronic imaging*, 20(2):023016, 2011. 4, 5

[47] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 286–301, 2018. 1, 3, 4, 6

[48] Hang Zhao, Orazio Gallo, Iuri Frosio, and Jan Kautz. Loss functions for image restoration with neural networks. *IEEE Transactions on computational imaging*, 3(1):47–57, 2016. 6

[49] Ruofan Zhou, Radhakrishna Achanta, and Sabine Süsstrunk. Deep residual network for joint demosaicing and super-resolution. *arXiv preprint arXiv:1802.06573*, 2018. 1, 3

[50] Ruofan Zhou, Majed El Helou, Daniel Sage, Thierry Laroche, Arne Seitz, and Sabine Süsstrunk. W2s: a joint denoising and super-resolution dataset. *arXiv preprint arXiv:2003.05961*, 2020. 3