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 important 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 solutions of this problem and compared them with the proposed method. Extensive experiments carried out on the 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 weighted 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 well studied in the past decades. With the recent advancement of convolutional neural networks (CNNs), performances of image restoration methods have significantly improved and several CNN-based image restoration methods achieved the state-of-the-art (SOTA) performance [17, 39, 35, 44].

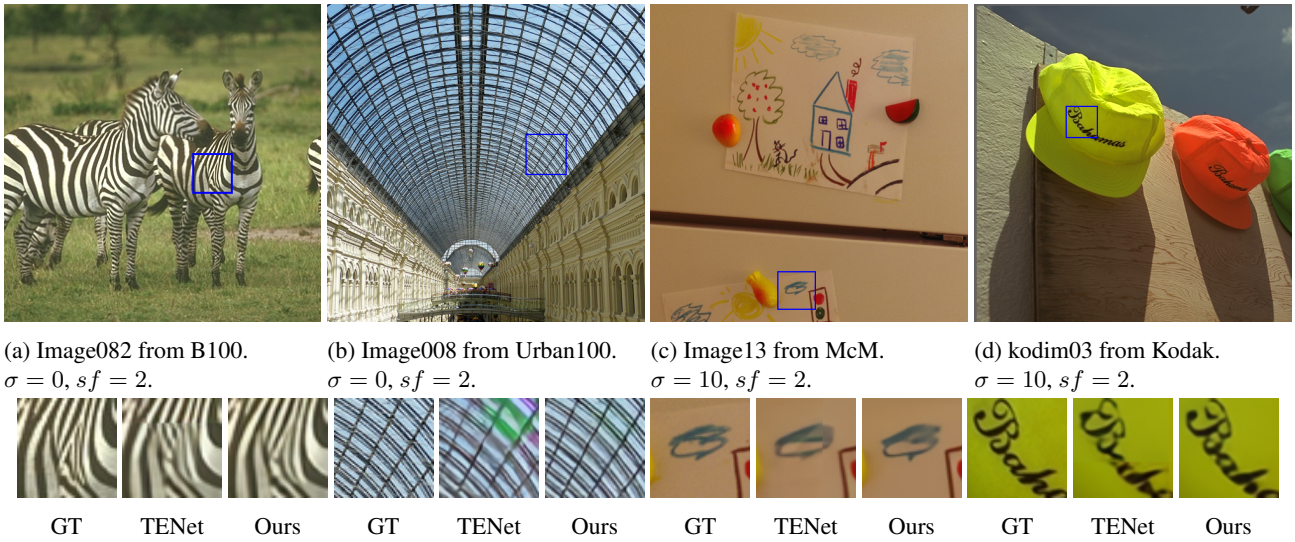
Since in many practical applications an acquired image is a subject of multiple degradations, the above mentioned image restoration problems must to be solved simultaneously. A most natural choice for a combined solution is to apply SOTA methods of individual restoration tasks in the sequential order. Existing SOTA, however, are not perfect. Addressing a problem of image denoising, most of the existing algorithms tend to smooth out high-frequency content, such as image details and texture, while eliminating noise in the flat areas of the image. Image demosaicing and super-resolution algorithms may generate color artifacts especially in the high-frequency texture regions and around edges. Image reconstruction errors of individual reconstruction algorithms are accumulated even if one is using SOTA algorithms in the sequential order.

To tackle this problem, joint solutions for the combined problems have been proposed in the literature [3, 6, 7, 8, 16, 19, 23, 34, 41, 46]. However, the problem of finding a joint solution for a triplet of problems of image demosaicing, denoising and SR has received much less attention [23, 27].

In this paper, we comprehensively study various solutions to this combined problem. First we adjust the execution order, and then start to investigate possible joint solutions under this execution order. Then we propose an end-to-end learning for this joint problem by designing a very deep network JD_ND_MSR. We carried out numerous experiments to show that the proposed solution outperforms other joint solutions quantitatively and qualitatively. To further optimize the proposed solution, different loss functions are used in the proposed scheme, and the comparative analysis of the resulting solutions is demonstrated in Section. 5.1. A comparison with the state-of-the-art method TENet [27] and the ablation study (see Fig. 1) are presented in Section. 5.2 and Section. 5.3, respectively.

2. Related work

Denoising. The existing image denoising methods can be roughly classified into two categories, model-based methods and deep learning methods. Among the model-based methods, BM3D [4] is often regarded as a denoising benchmark. In 2017, Zhang et al. have applied a deep convolutional neural network (CNN) to the problem of denoising images corrupted by Gaussian noise. This method, called DnCNN [39], adopts residual learning and batch normalization on CNN for denoising and attains top performance. Later on, a variety of machine-learning based methods [2, 14, 36, 39, 40] have been successfully used in denoising.



(a) Image082 from B100. $\sigma = 0$, $sf = 2$.
(b) Image008 from Urban100. $\sigma = 0$, $sf = 2$.
(c) Image13 from McM. $\sigma = 10$, $sf = 2$.
(d) kodim03 from Kodak. $\sigma = 10$, $sf = 2$.

GT TENet Ours GT TENet Ours GT TENet Ours GT TENet Ours

Figure 1: Qualitative comparison between the SOTA model TENet and the proposed $JD_ND_MSR^+$.

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 [10, 11, 24, 29, 42], which recover images based on mathematical models and image priors in the spatial-spectral domain; and learning-based methods [10, 30], based on process mapping learned from abundant training data. The deep learning methods [9, 15, 31] of image demosaicing attain 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 3 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 [17]. To reduce the occupation of memory and accelerate the speed of computation, Shi et al. have introduced a sub-pixel CNN ESPCN [28] to upscale feature maps to the desired solution. In 2017, Ledig et al. [20] have applied ResNet architecture in SR and proposed a SRResNet scheme. EDSR [21] 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 [44], which introduces the channel attention into residual blocks (RCAB). Wang et al. [35] have proposed a perceptual-driven method ESRGAN based on the proposed Residual-in-Residual Dense Block (RRDB). In 2020, Liu et al. [22] 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 SR [3, 6, 8, 16, 19], joint demosaicing and SR [7, 34, 46], and joint denoising and SR [41]. A research on the triplet of denoising, demosaicing and SR is still lacking a special attention. In 2019, Qian *et al*. [27] proposed a trinity network to jointly solve this composite problem. In 2020, Liu *et al*. proposed the SGNet [23] for joint image demosaicing and super-resolution, which also can handle the mixture problem of denosing, 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 subproblems, 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, all joint solutions of this execution order are considered. 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 [27], 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: $Dn \rightarrow Dm \rightarrow SR$, the first solution is to sequentially utilize three targeted methods to solve the corresponding image restoration problems one by one:

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

where M denotes image restoration method, Dn, Dm, 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:

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

and

$$\widehat{I_{HR}} = M_{JDmSR}(M_{Dn}(I_{LR_M^N})). \tag{3}$$

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

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

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

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 Figure 2, consisting 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 or is properly estimated in advance, thus, one can parametrize a network with this. One way is to add a noise level input σ, and replicate it spatially and concatenate it with the packed mosaic vector. Every layer downstream depends on it, which effectively parametrizes the learned filters. The color extraction step includes one convolutional layer with a big filter (256), and one deconvolutional layer to upscale the feature maps to the prime resolution. The feature extraction stage learns the residual features by several basic blocks and a convolutional layer. The basic block can be any effective block applied in SOTA methods, such as residual block (RB) [21], residual-in-residual dense block (RRDB) [35], and residual group (RG) with residual channel attention block (RCAB) [44]. Through the ablation study of the basic blocks (see Section 5.3), we utilize 4 residual groups in the JD_ND_MSR network structure and each RG includes 20 RCABs. In the reconstruction part, the deconvolutional 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 Chapter 4 will demonstrate that the proposed JD_ND_MSR and JD_MSR achieve notable performance improvement in comparison with the other joint solutions and the SOTA joint solution.

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 testing experiments run on a Linux desktop computer, with 3.4 GHz Intel i7-3770 CPU, 32 GB of RAM, and Nvidia GTX 1050Ti GPU with 4GB of memory.

Dataset. For training and validation of the network, we used publicly available dataset DIV2K [1] which contains 900 2K resolution images (800 for training, 100 for validation) for image restoration tasks. We compared different joint solutions on two public benchmark datasets, McMaster [43] and Kodak, which are often used for benchmarking in demosaicing. These two datasets are widely used in other image restoration papers as well [8, 17, 27, 33, 37, 39].

Data preprocessing. For data preprocessing of denoising, noisy input image is 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 [18]

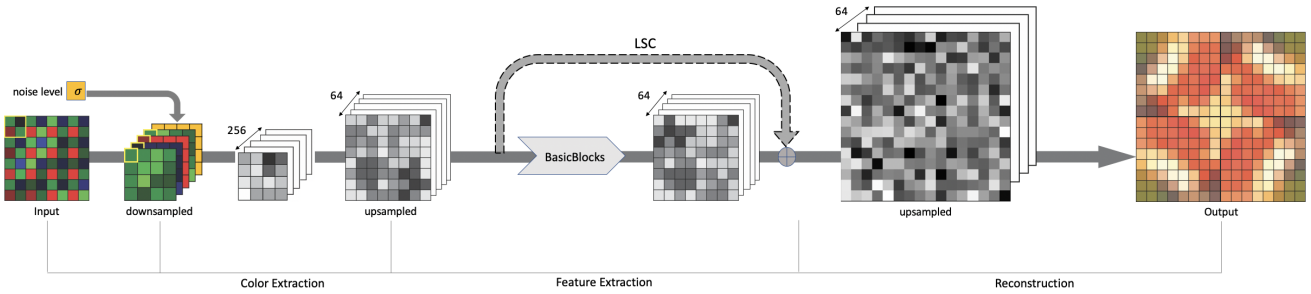


Figure 2: Illustration of the proposed deep joint denoising, demosaicing and super-resolution network JD_ND_MSR.

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 the 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 their high-resolution counterparts, the goal of training JD_ND_MSR is to minimize the loss function:

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

where Θ denotes the parameter set of JD_ND_MSR. The models in this part are trained with mean squared error (MSE). We also further optimize our network 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 all joint solutions (Eqn. 1-4) of the mixture problem of image demosaicing, denoising and super-resolution. Except the M_{JD_mSR} in Eqn. 3 solved by the proposed JD_MSR, other methods in Eqn. 1-3 are replaced by the state-of-the-art image restoration networks. DnCNN [39], DJDD [8], and VDSR [17] are selected methods for denoising, demosaicing and super-resolution, respectively. It should be noted that there are two versions of DJDD [8] (for noisy and noise-free inputs). The noisy version model is used in Eqn. 2 for joint demosaicing and denoising (M_{JDnDm}). 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 SOTA results. Therefore, we chose some simple but effective methods instead of computationally more demanding ones with the better performances. Similar to [38], we also introduce transfer-learning strategy to further improve JD_ND_MSR (we name the transfer-learning method as $JD_ND_MSR^+$). $JD_ND_MSR^+$ transfers the learned parameters from a trained model of JD_MSR (see our supplementary material) for joint $\times 2$ super-resolution and image demosaicing. The ablation study of transfer learning is also included in Section 5.3.

Table 1: Quantitative comparison of different solutions on the mixture problem of joint denoising, demosaicing and super-resolution using datasets Kodak and McMaster [43]. 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.

Solution type	Pipeline	McMaster cPSNR	McMaster SSIM	Kodak cPSNR	Kodak SSIM
Eqn. 1	DnCNN→DJDD→VDSR	25.99	0.8522	26.18	0.7868
	DnCNN* →DJDD* →VDSR*	29.14	0.9248	28.53	0.8913
Eqn. 2	DJDD→VDSR	28.40	0.9248	28.13	0.8812
	DJDD* →VDSR*	28.88	0.9212	28.43	0.8887
Eqn. 3	DnCNN→ JD_MSR	25.91	0.8522	26.11	0.7846
	DnCNN* → JD_MSR*	29.51	0.9293	28.75	0.8948
Eqn. 4	JD_ND_MSR	29.34	0.9274	28.80	0.8942
	$JD_ND_MSR^+$	29.56	0.9296	28.80	0.8965

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 comparisons of all solutions for joint image demosaicing, denoising and super-resolution. We fix the noise level to 10 and scale factor to 2. The loss function used in this comparison is mean square error (MSE). Since CNN models are sensitive to the input data, all models in the first three types of solutions (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. These retrained models are marked

by $*$. When compared with other solutions, our combined solution $JD_ND_MSR^+$ performs the best on both datasets. Even without transfer-learning, our next combined solution JD_ND_MSR also outperforms most of the other 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 $JD_ND_MSR^+$ not only eliminates the noise but also recovers more details in high frequency region.

Effects of combined solution. In Table 1, we can observe that our JD_ND_MSR 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_ND_MSR can achieve a comparable performance by the additional noise level estimation input. Our $JD_ND_MSR^+$ demonstrates a superior performance. This observation indicates that the combined solution can avoid an accumulation of errors. Considering the storage of models and weights, the combined solution is a good trade-off between performance and operational complexity.

Figure 3: 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 ground truth images are shown in the left part, and the resulting images of different joint solutions are shown in the right part. The sequence of resulting images corresponds to the illustration in the lower right corner.

5. Optimization

5.1. Comparison on cost functions

In Section 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. Inspired by [45], we train the network with six different cost functions: MSE, MAE, SSIM, MS-SSIM, Mix1 and Mix2. The Mix cost functions are combinations of MS-SSIM with l_1 or l_2 loss. These six models are compared on four evaluation metrics: cPSNR, SSIM, MS-SSIM and CSSIM [26]. 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 0.21dB on two datasets. The patch size during training was 32×32, and the number of epochs is 100.

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 [27] 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 [32] datasets with $\times 2$ scale factor and the noise level randomly sampled from [0, 20], called the resulting models as TENet-df2k and $JD_ND_MSR^+$-df2k, respectively. In addition to McMaster [43] and Kodak datasets, we also test them on B100 [25] and Urban100 [13] 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 ((Fig. 1 (a,c))) without color artifacts (Fig. 1 (b)) and

Table 2: Average value of different image quality metrics of $JD_ND_MSR^+$ trained with different cost functions. The three values in a single cell are tested on McMaster, Kodak and the mix of these two datasets, respectively. The noise level is 10 and the sale factor is 2. For cPSNR, PSNRw, SSIM, MS-SSIM, and cSSIM the value reported here has been obtained as an average of the three color channels. Best results are shown in bold.

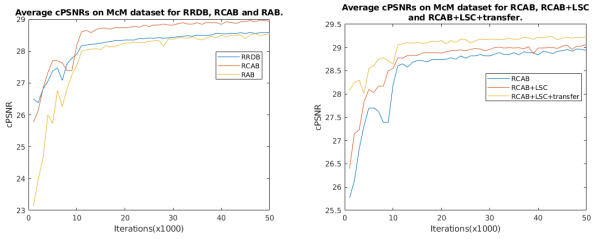
Image quality metric	Training cost function						
	MSE	MSEw	MAE	SSIM	MS-SSIM	Mix1	Mix2
cPSNR	29.05/28.45/28.71	29.16/28.56/28.81	**29.32/28.61/28.92**	27.74/27.48/27.59	27.52/27.25/27.36	27.61/27.30/27.43	27.48/27.20/27.32
PSNRw	23.71/23.23/23.44	23.80/23.30/23.51	**23.99/23.38/23.64**	22.10/22.00/22.04	21.91/21.81/21.85	22.02/21.89/21.94	21.84/21.74/21.78
SSIM	0.9224/0.8889/0.9033	0.9232/0.8914/0.9050	**0.9268/0.8933/0.9077**	0.8948/0.8600/0.8749	0.8895/0.8549/0.8697	0.8936/0.8565/0.8724	0.8858/0.8473/0.8638
MS-SSIM	0.9539/0.9442/0.9483	0.9552/0.9456/0.9497	**0.9572/0.9464/0.9511**	0.9409/0.9326/0.9361	0.9400/0.9329/0.9360	0.9399/0.9316/0.9352	0.9398/0.9329/0.9359
CSSIM	0.9775/0.9707/0.9736	0.9783/0.9717/0.9745	**0.9790/0.9723/0.9752**	0.9696/0.9617/0.9651	0.9700/0.9625/0.9657	0.9700/0.9618/0.9653	0.9701/0.9622/0.9656

blur (Fig. 1 (d)).

5.3. Ablation Study

In order to study the effects of each components in the proposed model $JD_ND_MSR^+$, we gradually modify the baseline $JD_ND_MSR^+$ model and compare their differences. In this comparison, all models are trained for 50 epochs with 1000 training iterations per epoch, 100 validation iterations per epoch, and 32×32 patch size. During the training, the scale factor was 2 and the noise level was randomly sampled from [0, 20]. The evaluation is done by calculating cPSNR values on McM dataset with noise level 10.

We start our investigation from a selection of the basic module. We compare three types of residuals in residual blocks, RRDB [35], RCAB [12] and RAB [44]. For a fair comparison, we tuned the number of three basic modules to keep all networks to have similar parameters. The comparison is shown in Fig. 4 (a). With a similar model size, the network with RCAB blocks performs better than those with the other two basic modules. Based on the comparison result of the basic modules, we have decided to add the Long Skip Connection (LSC) to our baseline. Fig. 4 (b) demonstrates that the additional LSC improves the performance of the network. In addition, we exploit the transfer learning, which transfers the well-learned parameters from the pre-trained noise-free model JD_MSR. The curves (yellow and red lines) in Fig. 4 (b) prove that this kind of easy-to-hard transfer learning strategy not only improves the performance of network, but also supports a better starting point (at least 1.5 dB higher cPSNR).



(a) Comparison of basic modules.

(b) Long skip connection and transfer learning influence.

Figure 4: Comparisons to show the effects of each component in $JD_ND_MSR^+$.

6. Conclusion

We have systematically and comprehensively compared all joint solutions of the mixture problem of image demosaicing, denoising and super-resolution, under the fixed execution order. Extensive experiments have demonstrated that the proposed combined version solution, $JD_ND_MSR^+$ surpasses other joint solutions, both quantitatively and qualitatively. Besides, the performance of $JD_ND_MSR^+$ is improved by training with mean absolute error cost function, instead of mean square error. The performance of this optimized model surpassed the state-of-the-art method TENet over four benchmark datasets on both noisy and noise-free data. 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.

[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.

[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.

[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.

[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.

Table 3: The average cPSNR/SSIM results of TENet-df2k and $JD_ND_MSR^+$-MAE-df2k on different datasets. The scale factor is 2 and the noise levels are 0, 10, 20 and 30.

Noise level	McMaster		Kodak		B100		Urban100	
	TENet-df2k	$JD_ND_MSR^+$-df2k	TENet-df2k	$JD_ND_MSR^+$-df2k	TENet-df2k	$JD_ND_MSR^+$-df2k	TENet-df2k	$JD_ND_MSR^+$-df2k
0	31.40/0.9579	**32.59/0.9652**	30.82/0.9394	**31.49/0.9456**	29.13/0.9189	**29.87/0.9283**	28.40/0.9268	**28.99/0.9331**
10	29.05/0.9247	**29.66/0.9315**	28.63/0.8950	**28.85/0.8982**	27.09/0.8670	**27.37/0.8725**	26.72/0.8904	**26.89/0.8922**
20	26.97/0.8877	**27.54/0.9000**	26.90/0.8518	**27.13/0.8595**	25.37/0.8142	**25.67/0.8250**	25.06/0.8480	**25.22/0.8524**
30	25.49/0.8535	**26.11/0.8724**	25.73/0.8173	**26.03/0.8309**	24.20/0.7725	**24.57/0.7924**	23.77/0.8082	**24.01/0.8156**

[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.

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

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

[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.

[10] 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.

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

[12] 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.

[13] 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.

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

[15] 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.

[16] 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.

[17] 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.

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

[19] 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.

[20] 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.

[21] 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.

[22] 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.

[23] 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.

[24] 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.

[25] 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.

[26] Mykola Ponomarenko, Karen Egiazarian, Vladimir Lukin, and Victoriya Abramova. Structural similarity index with predictability of image blocks. In *2018 IEEE 17th International Conference on Mathematical Methods in Electromagnetic Theory (MMET)*, pages 115–118. IEEE, 2018.

[27] 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.

[28] 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.

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

[30] 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.

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

[32] 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.

[33] 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.

[34] 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.

[35] 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.

[36] 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.

[37] 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.

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

[39] 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.

[40] 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.

[41] 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.

[42] 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.

[43] 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.

[44] 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.

[45] 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.

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