Trinity of Pixel Enhancement: a Joint Solution for Demosaicking, Denoising and Super-Resolution

Guocheng Qian[1*], Jinjin Gu[12*], Jimmy S. Ren[1], Chao Dong[3], Furong Zhao[1], Juan Lin[1]
[1]SenseTime Research [2]The Chinese University of Hong Kong, Shenzhen
[3]Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences
{gujinjin, qianguocheng, rensijie, zhaofurong, linjuan}@sensetime.com, chao.dong@siat.ac.cn

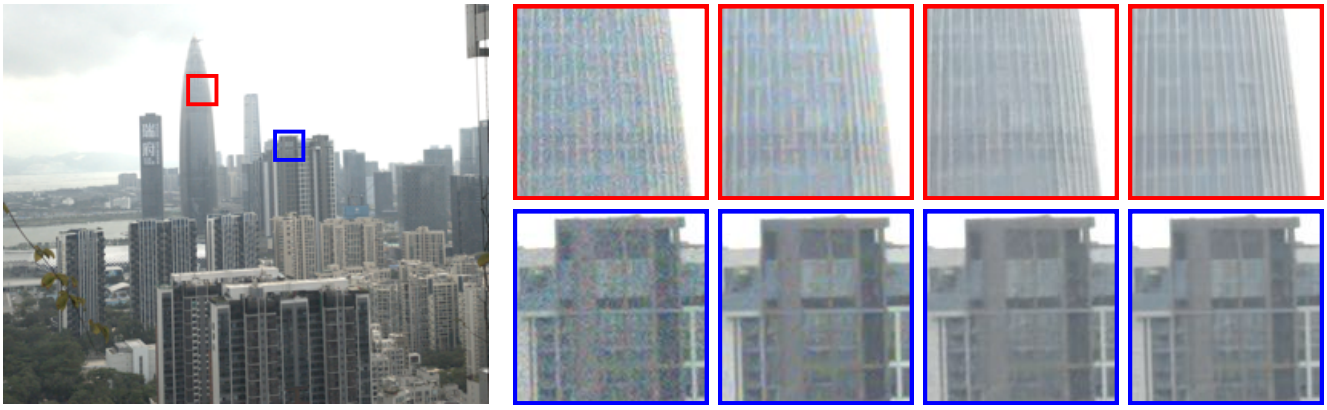

Figure 1: Our model TENet achieves better result on the mixture problem of demosaicing, denoising and SR on the real raw sensor test image captured by iPhone X. We conduct comparison with the most popular commercial software (Camera Raw) and the state-of-the-art demosaicing method [13] and SR method [3]. Our output is artifact-free and preserves detail even for challenging regions. Here, *CARN is fine tuned from CARN [3] using pixel averaging downsampling for fair comparison.

Abstract

Demosaicing, denoising and super-resolution (SR) are of practical importance in digital image processing and have been studied independently in the passed decades. Despite the recent improvement of learning-based image processing methods in image quality, there lacks enough analysis into their interactions and characteristics under a realistic setting of the mixture problem of demosaicing, denoising and SR. In existing solutions, these tasks are simply combined to obtain a high-resolution image from a low-resolution raw mosaic image, resulting in a performance drop of the final image quality. In this paper, we first rethink the mixture problem from a holistic perspective and then propose the Trinity Enhancement Network (TENet), a specially designed learning-based method for the mixture problem, which adopts a novel image processing pipeline order and a joint learning strategy. In order to obtain the correct color sampling for training, we also contribute a new dataset namely PixelShift200, which consists of high-quality full color sampled real-world images using the advanced pixel shift technique. Experiments demonstrate that our TENet is superior to existing solutions in both quantitative and qualitative perspective. Our experiments also show the necessity of the proposed PixelShift200 dataset.

1. Introduction

In computational photography, obtaining high-quality high-resolution even super-resolution images has attracted increasingly attention in research community and commercial industry. However, obtaining such images is of practical difficulty under limited hardware conditions (small prime lens and compact sensors, etc.), especially for mobile devices. The limitations mainly come from three aspects. First, most digital cameras contain sensor arrays covered by color filter arrays (CFAs, e.g. Bayer pattern), resulting in incomplete color sampling of images and loss in resolution. Second, the images captured directly by the image sensor are usually noisy. Especially when the pixel density of sensor becomes larger, the noise becomes more obvious.

* J. Gu and G. Qian contributed equally to this work. This work was done when they were interns at SenseTime.

Third, most the lenses used in mobile devices have fixed and short focal length, which not only causes difficulties to the imaging of distant objects, but also limits the resolution of the images. In order to break through the above hardware limitations, some post-processing methods are introduced to enhance the images. Demosaicing, denoising and super-resolution (SR) are three fundamental processing tasks, respectively. In the past decades, these three tasks have been well studied separately, and all have made breakthrough progress recently with the help of deep learning.

However, the problems encountered in practical applications are more complicated than any single problem – it is usually a mixture problem of noise and resolution limitation (color mosaic and insufficient resolution). Although perform well when applied separately to solve a single problem, it will bring in new problems when those tasks are simply combined to solve mixture problem (e.g., unexpected artifacts and blurry), which are caused by the interactions between tasks. Such a mixture problem has received lower attention in research field. In this paper, we rethink the mixture problem from a holistic perspective. By thoroughly analyzing the characteristics of each task and the behaviors of their interactions, we propose a new method, namely Trinity Enhancement Network (TENet), to solve the mixture problem. Experiments demonstrate the superiority of the proposed TENet method under realistic settings.

The motivation behind this work is three-fold. Firstly, we adjust the order of demosaicing and SR in image processing pipeline. Although formulated differently, both demosaicing and SR are meant to overcome the sampling limitation of imaging. In the existing solutions, the image is first demosaiced to obtain a full color image. Then SR is performed to further enhance the resolution. However, demosaicing will introduce artifacts when the resolution is limited (such as color aliases, zippering and moiré artifacts), and these artifacts will be magnified by the followed SR process. To address this problem, we propose to super-resolve the raw mosaic image before demosaicing. In the new pipeline, not only the artifacts of demosaicing is reduced, but SR also helps demosaicing to break the resolution limit. Secondly, simply combining two or more tasks usually causes severe performance drop, e.g. new artifacts and blurry. An important reason for this drop is that there is no appropriate model or algorithm can perfectly handle the *middle state*, which refers to the intermediate result after one or two steps of processing [46]. These middle states usually involve task related complicated defects. With the advent of deep learning based methods, we are able to address complicated multi-task image processing problems in an end-to-end manner, which is also known as "joint solution". When jointly performed, if one task produces the result that is difficult to process directly, the followed task will compensate for the middle state, and provide better final results. Thus, we propose to perform demosaicing, denoising and SR in such a joint scheme for the mixture problem. Thirdly, we contribute a real-world dataset with the advanced pixel shift technique namely *PixelShift200* for this mixture problem. By further diving into the training data, we find that the existing datasets have limitations in training demosaicing related tasks. As the images in those datasets are demosaiced from raw mosaic images, so they contains potential artifacts. The proposed *PixelShift200* consists of 200 high-quality 4k resolution full color sampled real-world images. By training with the above dataset, our TENet can reconstruct high-quality high-resolution images with less artifacts.

We summarize our contributions as follows: (1) We are the first to analyze the mixture problem of demosaicing, denoising and SR and propose the Trinity Enhancement Network (TENet) to solve the mixture problem. (2) We propose to super-resolve mosaic image before demosaicing. We show the superior performance of the proposed pipeline with experiments. (3) We contribute a new real-world dataset namely *PixelShift200* for demosaicing and SR with novel pixel shift technique. Experiments show the necessity of proposed dataset in training demosaicing related tasks.

2. Related Work

We aim to solve the mixture problem of demosaicing, denoising and SR for a single Bayer image. All the above tasks are well-studied separately. Since we are the first to address the mixture problem with joint solution, in this section, we first briefly present the previous work and existing problems for the above tasks, and then review the literature of joint solutions.

2.1. Demosaicking

Image demosaicing is an ill-posed problem of interpolating full-resolution color images from the color mosaic images (e.g. Bayer mosaic images), and is usually preformed in the beginning of image processing pipeline. Existing approaches can be mainly classified into two categories: model-based and learning-based methods. Model-based approaches [31, 50, 19, 37, 42, 17] focus on the construction of mathematical models and image priors in the spatial-spectral domain facilitating the recovery of missing data. Learning-based approaches [17, 38] build the process mapping by learning from abundant training data. Recently, deep learning has also used successfully for image demosaicing and achieved competitive performance [21, 14, 13, 39]. Michaël *et al.* [13] train a deep convolutional neural network (CNN) on millions of carefully selected image patches and achieve the state-of-the-art performance of demosaicking.

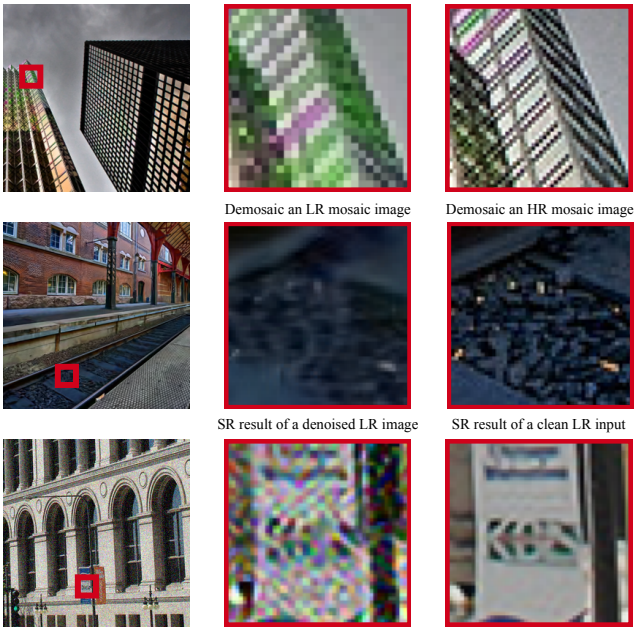

Figure 2: The interactions between different tasks. As shown in the first row, the image demosaiced from an LR image contains severe color distortion. The image demosaiced from an HR image provides better result. The second row indicates that the denoising task tends to smooth the high frequency details. The last row shows the serious artifacts of super-resolving a noisy input.

In general, demosaicing algorithms perform well in flat regions of the image. However, it leads to conspicuous artifacts in the high-frequency texture regions and strong edges. Due to the input resolution limitation, serious artifacts such as zippering, color moiré and loss of detail are prone to occur in this area. This kind of problem is related to resolution limitation of the input Bayer image [54], and will be alleviated when the image resolution is increased, as shown in the first row of Figure 2 When the input low-resolution mosaic raw image contains noise, the demosaicking is further difficult. This leads to unpleasant artifacts, as the estimation of edge orientation is less reliable.

2.2. Denoising

Image noise is inevitable during imaging and it may heavily degrade the visual quality. In past decades, plenty of methods have been proposed for denoising not only for color images but also mosaic images. Early methods such as anisotropic diffusion [33], total variation denoising [34] and wavelet coring [36] use hand-craft features and algorithms to recover a clean signal from noisy input. However these parametric methods have limited capacity and expressiveness. Advanced methods usually exploit effective image priors such as self-similarity [8, 16, 4] and sparse representation [2]. With the increasing of interests to learning-based methods, in recent years, most successful denoising algorithms are entirely data-driven, consisting of CNNs trained to recover from noisy images to noise-free images [5, 47, 48, 35, 45, 13].

Same as demosaicing, denoising algorithms work well in flat regions in the image, they eliminates high-frequency noise to make image smooth and clean. Unfortunately, most denoising algorithms not only eliminate noise, but also smooth the high-frequency detail and texture in the image. If we further conduct post-process on the denoised image such as SR, the blur effect will be magnify and affect image quality, as shown in the second row of Figure 2. Note that when the noise of the input image is complicated, denoising algorithms will hardly remove this kind of defects. Thus, the denoising algorithms have limited performance when removing artifacts left by other algorithms.

2.3. Super-Resolution

SR aims to recover the high-resolution (HR) image from its low-resolution (LR) version. Since the seminal work of employing CNN for SR [10], various deep learning based methods with different network architectures [11, 23, 53, 3, 24, 30, 15] and training strategies [29, 44] have been proposed to continuously improve the SR performance. However, problems occur when apply such algorithm in real-world applications. When SR algorithms enhance the image details and texture, the unexpected noise, blurry and artifacts are also magnified. If the input image is noisy or blurry, the problems that were not serious will be magnified, especially for artifacts and noise caused by previous processing. It may lead to unsatisfactory results when apply SR separately after demosaicking or denoising. An example is shown in Figure 2.

2.4. Mixture Problem of Image Processing

In practical applications, in addition to the above well defined problems, more common is the mixture problem of multiple image defects. For example, the mixture problem of SR and denoising [49], demosaicing and denoising [6, 22, 26], and the problem of SR and demosaicing [12, 43, 54]. For the mixture problem of multiple tasks, the difficulty of solving is greatly increased. Yu *et al.* [46] study the order of execution of tasks in the mixture problem and use reinforcement learning to learn the order of execution of the task. More relevant to this work, Michaël *et al.* [13] train a CNN to jointly perform these tasks and achieve the state of art performance. Zhang *et al.* [49] propose a SR network to jointly perform SR and denoising, as the denoising pre-precessing step tends to lose detail information and would deteriorate the subsequent SR performance. Zhou *et al.* [54] introduce deep residual network for joint demosaicking and super-resolution. However, the mixture

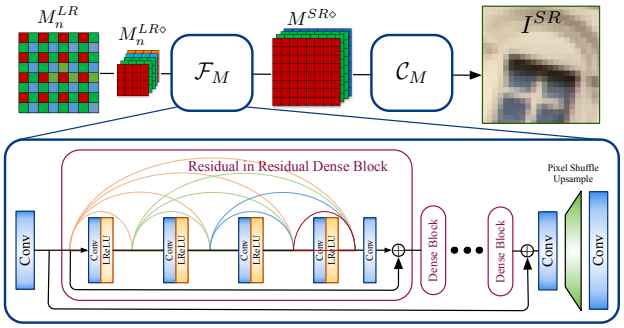

Figure 3: Our proposed Trinity Enhancement Network (TENet).

problem of demosaicking, denoising and SR has not witnessed the usage of jointly perform strategy to the best of our knowledge.

3. Method

Our main aim is to improve the overall image quality for the mixture problem of demosaicing, denoising and SR. In this section, we first discuss the improvements from the proposed new pipeline. And then we describe the proposed joint objective function. At last we present the network design.

3.1. Pipeline

As mentioned above, different tasks will interact with each other. When multiple processing tasks are executed in sequence, the defects generated by the previous task will affect the subsequent tasks and then cause performance drop. In our approach, we carefully adjust the execution order of denoising, SR and demosaicing to minimize the effects caused by task interaction.

Firstly, we suggest to denoise before other tasks to minimize the effects of noise. If the denoising operation is performed after SR or demosaicing, the noise will impact the processing of SR and demosaicing and cause severe artifacts that is difficult to remove. Secondly, different from the previous popular image processing pipeline which first demosaic the raw image into a full color image and then perform SR, we propose to super-resolve the raw image to a higher resolution and then perform demosaicing to get the SR color image. There are at least two advantages: (1) The artifacts caused by super-resolving the defects of the demosaiced images can be avoided. (2) SR can help demosaicing task to break the limitation of resolution. Demosaic a higher resolution raw images will obtain better results.

In our pipeline, for a given noisy LR raw mosaic image M_n^{LR}, its corresponding HR color image I^{HR} can be written as a composite function:

$$I^{HR} = \mathcal{C}(\mathcal{S}_M(\mathcal{D}_M(M_n^{LR}))), \tag{1}$$

where $\mathcal{C}$ is the demosaicing mapping, $\mathcal{S}_M$[1] is the SR mapping for mosaic images and $\mathcal{D}_M$ denotes the denoising mapping for mosaic images. The denoising is first performed to obtain noise-free mosaic LR image $M^{LR} = \mathcal{D}_M(M_n^{LR})$. We then use an SR mapping to super-resolve the LR mosaic image in order to obtain HR mosaic image $M^{HR} = \mathcal{S}_M(M^{LR})$. At last, we perform demosaicing to convert HR mosaic image into full color image $I^{HR} = \mathcal{C}(M^{HR})$. In our approach, we employ deep convolutional neural networks to implement the above mappings.

3.2. Joint Objective Function

With a carefully designed image processing pipeline, we can avoid the serious performance drop caused by the interaction between different tasks to a certain extent. However, we still cannot totally solve the problem caused by the middle state. In the proposed pipeline, although the denoising is performed at first to eliminate serious artifacts, it will lose high-frequency textures and image details, which still causes difficulties for subsequent tasks – no SR or demosaicing method is designed to compensate the lost high-frequency details. The distribution of the super-resolved mosaic image is also different from the real-world mosaic image. Directly performing existing demosaicing method cannot achieve the satisfactory processing effect. To address this problem, we propose to joint perform denoising, SR and demosaicing in an end-to-end manner. In our approach, we calculate the l_2-norm loss on the final result I^{HR} directly:

$$\mathcal{L}_{joint} = \|\mathcal{C}(\mathcal{S}_M(\mathcal{D}_M(M_n^{LR}))) - I_{gt}^{HR}\|_2^2, \tag{2}$$

where I_{gt}^{HR} represents the ground-truth HR color image of M_n^{LR}. In order to provide more information to the network during training, we also calculate the SR loss on raw image to optimize the functionality of the network.

$$\mathcal{L}_{SR} = \|\mathcal{S}_M(\mathcal{D}_M(M_n^{LR})) - M_{gt}^{HR}\|_2^2, \tag{3}$$

where M_{gt}^{HR} represents the corresponding HR noise-free mosaic image of M_n^{LR}. The SR loss term makes the first half of the network focus on denoising and SR and the second half of the network focuses on demosaicing the super-resolved mosaic image. Although the joint perform strategy mainly focuses on the final results, providing the supervision information of intermediate state can also optimize network performance during joint processing. The final objective function in our approach is

$$\mathcal{L} = \mathcal{L}_{joint} + \lambda \mathcal{L}_{SR}, \tag{4}$$

where λ is the trade-off parameter.

[1]The subscript $_M$ stands for 'mosaic'.

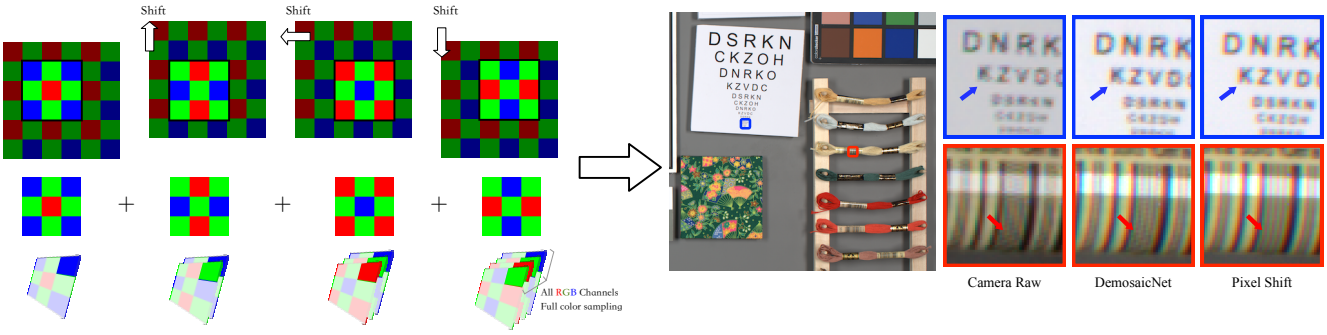

Figure 4: The pixel shift technique used to contribute dataset *PixelShift200* and samples of qualitative comparison among dcraw, Camera Raw and pixel shift results.

3.3. Network Design

As mentioned above, our approach can be divided into two parts. The first part is the mapping of joint denoising and SR, denoted with $\mathcal{F}_M$, which actually jointly implements $\mathcal{D}_M$ and $\mathcal{S}_M$. The second part is mapping $\mathcal{C}_M$, which converts the SR mosaic image into a full color image. The mapping $\mathcal{F}_M$ and $\mathcal{C}_M$ can be trained and performed jointly. At first, the Bayer mosaic image M_n^{LR} is extracted into four color maps $M_n^{LR\diamond}$, so that the spatial information of the same color is more easily extracted with convolution operation. Mapping $\mathcal{F}_M$ maps these four color maps to a SR noise-free mosaic color maps $M^{SR\diamond}$, which is then mapped to a SR three-channel color image by the mapping $\mathcal{C}_M$. We employ the deep network of ESRGAN [44] to implement these two mappings, which uses a specially designed Residual in Residual Dense Block (RRDB) to increase the stability of the training. The network structure is illustrated in Figure 3. For the network $\mathcal{F}_M$, the input has five channels (including a noise map, stretched with noise level to indicate the sigma of the Gaussian noise) and the output has four channels. For network $\mathcal{C}_M$, the input is a four-channel SR mosaic image and the output is a three-channel RGB image. In order to balance the number of parameters and running time, the number of RRDBs for both $\mathcal{F}_M$ and $\mathcal{C}_M$ is set to 6.

4. Data Collection

Although there are many high-resolution image datasets available, we find that existing datasets are difficult to meet the requirements related to the training of demosaicing tasks. When it comes to demosaicing tasks, we need the full color sampled ground truth images. However, since most of the high-quality images are obtained by demosaicing the mosaic raw images, training using such data will introduce the artifacts generated by the existing demosaicing algorithm. In previous work, the high-quality images are first preprocessed to eliminate the effect of demosaicing artifacts. Zhou *et al.* [54] perform bicubic downsampling operation to the original high-resolution images to eliminate artifacts that have potentially been introduced by the demosaicing algorithm as well as by other factors in the camera processing pipeline (like sensor noise). Michaël *et al.* [13] use ImageNet [9] dataset and propose a novel training data selection method to select the 'hard case' of training data. Note that the ImageNet images can also be viewed as downsampled images. Although the downsampled image no longer contains obvious artifacts, the downsampled image is somewhat different from the natural image distribution. We need the real-world high-resolution images with full sampling of the color to train demosaicing related tasks.

In this paper, we contribute a novel dataset *PixelShift200* and a new testset *PixelShiftTest*. We employ advanced pixel shift technology to perform a full color sampling of the image. Pixel shift technology takes four samples of the same image, and physically controls the camera sensor to move one pixel horizontally or vertically at each sampling to capture all color information at each pixel (see Figure 4). The pixel shift technology ensures that the sampled images follow the distribution of natural images sampled by the camera, and the full information of the color is completely obtained. In this way, the collected images are artifacts-free, which could lead to better training results for demosaicing related tasks.

During data collection, we use a Sony A7R3 digital camera with pixel shift technology. In order to control the quality of the data, most of the images were taken on the real scene and the finely printed pictures in the darkroom, the light intensity and color temperature is predefined and fixed. In order to avoid motion parallax when moving the sensor, We control the depth of field of the scene to a small range. We use a lens with fixed focal length and aperture, and use a low photosensibility (ISO 100 or less) to avoid possible serious noise. We divided the data into 10 test images, namely *PixelShiftTest* and 200 4k resolution training pictures namely *PixelShift200*. Some examples are shown in Figure 4, one can see that the pixel shift results is artifacts-free and thus provide better ground truth for training.

Table 1: Quantitative comparison of the performance of different approaches on the demosaicing and SR mixture problem using dataset Kodak, McM [51], BSD100 [32] and Urban100 [20]. The SR factor is 2. Note the DemosaicNet [13] used in this comparison is the noise-free version.

Method	Kodak		McMaster [51]		BSD100 [32]		Urban100 [20]	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Malvar *et al.* [31] + *CARN [3]	28.40	0.8421	28.51	0.8442	27.00	0.8201	24.49	0.8143
NLM [52] + *CARN [3]	29.29	0.8477	29.70	0.8731	27.72	0.8230	25.95	0.8440
NAT [52] + *CARN [3]	29.20	0.8551	29.64	0.8733	27.60	0.8293	25.89	0.8451
DemosaicNet [13] + *CARN [3]	30.82	0.8864	31.60	0.9052	28.99	0.8644	28.14	0.8886
†DemosaicNet [13] + *CARN [3]	30.29	0.8886	31.75	0.9073	29.22	0.8675	28.44	0.8942
TENet (noise-free)	31.39	0.8965	32.40	0.9163	29.39	0.8736	29.37	0.9061

Table 2: Quantitative comparison of different approaches on the mixture problem of demosaicing, denoising and SR using dataset Kodak, McM [51], BSD100 [32] and Urban100 [20]. The SR factor is 2 and the noise level is set to 10, 20 and 50. The DemosaicNet [13] do not provide the model for noise level more than 20.

Method	Noise level	Kodak		McMaster [51]		BSD100 [32]		Urban100 [20]	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
ADMM [40] + *CARN [3]	10	26.71	0.7310	27.53	0.7793	25.80	0.6992	24.10	0.7414
Condak [7] + *CARN [3]		27.21	0.7654	26.43	0.7717	25.90	0.7382	24.64	0.7823
FlexISP [18] + *CARN [3]		25.29	0.6362	25.26	0.6601	24.30	0.6153	23.50	0.6784
DemosaicNet [13] + *CARN [3]		27.82	0.7830	28.75	0.8153	26.82	0.7601	25.58	0.7960
†DemosaicNet [13] + *CARN [3]		27.96	0.7874	28.87	0.8202	26.92	0.7640	25.87	0.8055
TENet (ours)		28.60	0.8067	29.56	0.8423	27.32	0.7783	27.18	0.8470
ADMM [40] + *CARN [3]	20	25.76	0.6893	26.03	0.7101	24.72	0.6480	23.45	0.7029
Condak [7] + *CARN [3]		25.74	0.6920	24.81	0.6893	24.40	0.6462	23.54	0.7211
FlexISP [18] + *CARN [3]		23.03	0.4573	22.77	0.4822	22.30	0.4613	21.30	0.5176
DemosaicNet [13] + *CARN [3]		26.15	0.6989	26.53	0.7239	25.09	0.6644	23.89	0.7142
†DemosaicNet [13] + *CARN [3]		26.22	0.7029	26.61	0.7308	25.15	0.6677	24.04	0.7218
TENet (ours)		26.99	0.7388	27.51	0.7799	25.68	0.6973	25.50	0.7932
ADMM [40] + *CARN [3]	50	23.06	0.5629	22.24	0.5461	22.02	0.5152	20.85	0.5723
Condak [7] + *CARN [3]		22.92	0.5743	21.36	0.5301	21.66	0.5003	20.52	0.5683
FlexISP [18] + *CARN [3]		18.47	0.2071	18.17	0.2292	18.07	0.2257	17.27	0.2789
TENet (ours)		23.93	0.5867	23.79	0.6010	22.81	0.5483	22.16	0.6513

5. Experiments

5.1. Data Preprocessing and Network Training

Since downsampling operation is performed on the mosaic raw image, we propose to employ pixel averaging as the downsampling method. In pervious work, camera hardware binning was used to implement the downsampling operation on a monochromatic sensor directly [28, 27]. However, due to the existence of CFA in a color sensor, it is difficult to adopt such hardware binning technique on the color mosaic images. In our experiments, we simulate the hardware binning downsampling by performing a pixel averaging downsampling on the full color sampled images obtained by the pixel shift technique. We employ white Gaussian noise for the noisy input synthesis. We conduct the comparison on both existing high-quality image datasets and real-world dataset. For the high-quality data, we use the DIV2K dataset [1], which contains 800 2K resolution images for image restoration tasks. Beyond the training set of DIV2K, we further use the Flickr2K dataset [41] consisting of 2650 2K resolution images to enrich our training set. For the real-world data set, we use the proposed *PixelShift200* contains 200 4K resolution images as the training set. For the optimization of network parameters, we use Adam [25] with $\beta_1 = 0.9, \beta_2 = 0.999$ and the learning rate is 1×10^{-4}. The mini-batch size is set to 16. The spatial size of cropped HR patch of color images is 256×256. We implement our models with the PyTorch framework and train them using NVIDIA Titan Xp GPUs. The entire training process takes about two days.

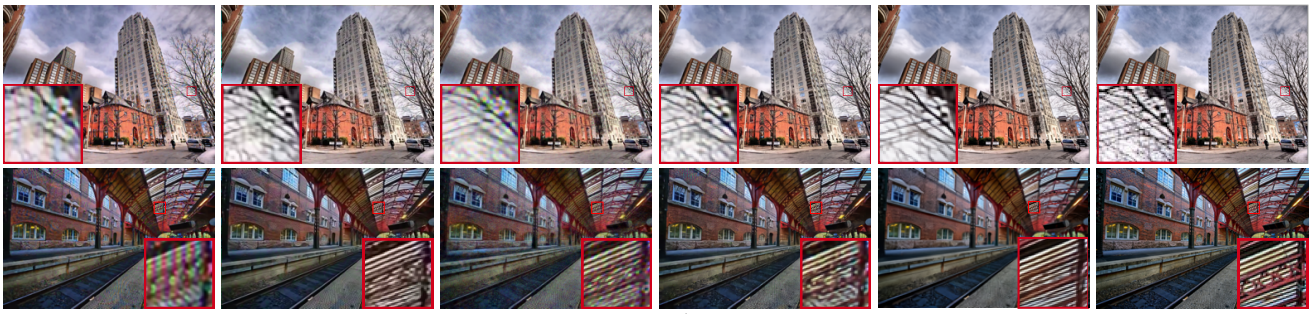

Figure 5: Comparison of our approach with ADMM [40], Condat [6], FlexISP [18] and fine-tuned DemosaickNet [13] on the noisy synthetic test images. The noise level of Gaussian noise is 10 and the SR factor is 2.

Table 3: Quantitative comparison of different pipelines on the demosaicing, denoising and SR mixture problem. In this experiment, the tasks are perform step by step with different order with fixed network for each task.

Method	Kodak		Urban100 [20]	
	PSNR	SSIM	PNSR	SSIM
DM $\rightarrow$ SR $\rightarrow$ DN	26.40	0.6495	24.98	0.7029
SR $\rightarrow$ DM $\rightarrow$ DN	26.86	0.6796	25.42	0.7311
SR $\rightarrow$ DN $\rightarrow$ DM	27.28	0.7089	25.86	0.7589
DM $\rightarrow$ DN $\rightarrow$ SR	26.97	0.6991	25.38	0.7491
DN $\rightarrow$ DM $\rightarrow$ SR	28.40	0.8028	26.55	0.8355
DN $\rightarrow$ SR $\rightarrow$ DM	28.45	0.8038	26.75	0.8395

Table 4: Quantitative comparison of different joint solutions on the demosaicing, denoising and SR mixture problem. In this experiment, the tasks are joint or partially joint performed (denoted by $+$) in different orders.

Method	Kodak		Urban100 [20]	
	PSNR	SSIM	PNSR	SSIM
SR $\rightarrow$ DN $+$ DM	27.27	0.7062	25.89	0.7579
DN $\rightarrow$ DM $+$ SR	28.47	0.8041	26.76	0.8396
DM $\rightarrow$ DN $+$ SR	27.07	0.6869	25.86	0.7468
DM $+$ DN $\rightarrow$ SR	28.43	0.8039	26.59	0.8365
DM $+$ SR $\rightarrow$ DN	26.67	0.6618	25.26	0.7149
DN $+$ SR $\rightarrow$ DM	28.54	0.8048	26.96	0.8437
SR $+$ DN $+$ DM	28.56	0.8050	27.10	0.8451
SR $+$ DN $+$ DM, $w/\mathcal{L}_{SR}$	28.60	0.8051	27.14	0.8458

5.2. Experiments on Synthesis Test Images

We compare our method on several public benchmark datasets under both noise-free and noisy settings. Note that it lacks research for the mixture problem of demosaicing and SR, we implement such comparison with the combination of demosaicing methods and the state-of-the-art SR method CARN [3]. For fair comparison, the CARN model used is fine tuned using pixel averaging downsampling, denoted with *CARN. We also provide the comparison with the joint trained DemosaicNet and *CARN (denoted by †DemosaicNet + *CARN).

For the noise-free setting, we compare our noise-free version model with the combination of *CARN and the state-of-the-art demosaicing methods. The quantitative comparison result is shown in Table 1. One can see that the joint fine tuned †DemosaicNet and *CARN achieves better result compared to the original model, which demonstrates the effectiveness of joint strategy. Also, our TENet outperform the all the existed solutions on the mixture problem of SR and demosaicing.

For the noisy input setting, we compare our final model with the combination of *CARN and the state-of-the-art demosaicing and denoising methods. Table 2 shows the quantitative comparison result. One can see that our TENet outperform the all the existed solutions on the such mixture problem. Some examples are shown in Figure 5, and more comparison results are shown in supplementary material. As can be seen, for ADMM and Condak, the demosaicing are affect by the noise, resulting in over-smooth results and color aliasing artifacts. The subsequent SR task further magnify this image distortion. For FlexISP, the demosaiced image contains serious artifacts caused by noise, which is a damage to the final visual effect. †DemosaicNet causes artifacts and also fails in the recovery of high frequency details. The proposed TENet is able to provide clean image with rich and accurate details.

5.3. Experiments on Real-World Test Images

We test the proposed TENet using a randomly selected raw image shot by IPhone X mobile phone. We compare our method with the joint fine-tuned DemosaicNet and CARN, dcraw and a popular commercial photography software Camera Raw. Figure 1 shows the visual effect comparison. As one can see, the proposed TNEet provides clean processing result with rich details. Figure 6 shows the results on the proposed *PixelShiftTest* testset. Our proposed successfully reconstruct the high frequency texture without generating any artifacts and color aliasing.

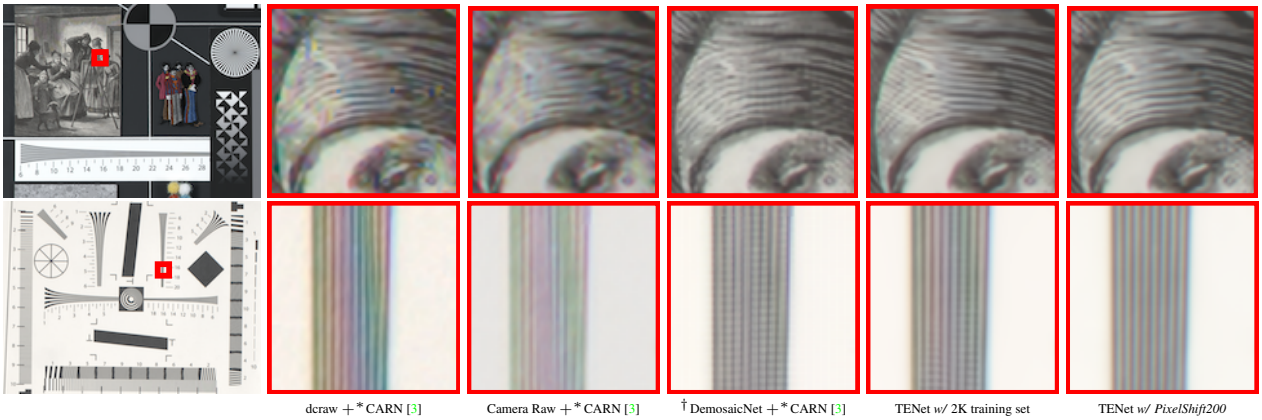

Figure 6: Comparison of our approach with dcraw, Camera Raw, DemosaicNet and two TENet trained using 2K image dataset and the propoed *PixelShift200* using real-world test raw images.

6. Ablation Study

In order to study the effects of each component in the proposed, we gradually modify the baseline model and compare their differences. In ablation study, we denote DM as Demosaicing, DN as Denoising and SR as Super-Resolution. We use the same network architecture with 6 RRDBs to implement SR, denoising and demosaicing tasks, respectively. The models are all well trained using DIV2K and Flickr2K dataset separately. For denoising and SR tasks, we prepare the models for both mosaic images and color images.

6.1. Comparison of Pipelines

In this section, we study the performance of different pipeline orders. Based on the above models, we implement different image processing pipelines and test with SR factor equals 2 and noise level equals to 10. The quantitative comparison results are shown in Table 3. As one can see, when DN is not performed at first, the numerical performance will decline sharply due to the artifacts which is difficult to remove. When DN is fixed as the first task, exchange DM and SR will improve the performance. The proposed pipeline order out perform the others.

6.2. Effects of Joint Solution

In this section, we study the performance of different joint solutions. We perform joint strategy to two or more tasks mentioned above, and then perform with other task wit different order. The quantitative comparison results are shown in Table 4. As can be seen, with the similar pipeline orders, the performance of the joint solution is generally better than the solution without joint strategy. Same as the above experiment, when DN is not performed at first, the numerical performance will decline sharply. According to the Table 4 row 4 and row 6, the solution that joint perform DN with other tasks at first achieves relatively good performance. The solution with SR at first outperform the traditional solution of DM + DN $\rightarrow$ SR, which also demonstrate the effectiveness of the proposed pipeline order. As revealed in the last two rows, joint performing DN, SR and DM using a large network outperforms other partial joint solutions. In particular, we are able to further improve the performance with the employing of the additional SR loss $\mathcal{L}_{SR}$.

6.3. Comparison of Different Datasets

In this section, we compare the TENet trained with different training datasets on real-world images. Figure 6 shows the qualitative comparison of several popular or state-of-the-art solutions and TENet. We can further observe that the results of TENet trained on 2K images perform well in most regions while fail under extreme conditions. One of the reason is that it lacks the good sampling of such difficult case in the training set. Our results demonstrate the effectiveness of the proposed *PixelShift200* dataset.

7. Conclusion

In this paper, we conduct thorough analysis of interactions of denoising, demosaicing and SR. We propose TENet for jointly solving these three tasks in a specific pipeline order. Our quantitive and qualitative experiments results demonstrate the effectiveness of the proposed new pipeline order and joint strategy. We also contribute a fully color-sampled datasets namely *PixelShift200* for training demosaicking related tasks. The qualitative result on real mosaic raw images shows the our model trained on *PixelShift200* outperforms the combination of the state-of-the-art demosaicking methods and SR methods. Our work shows the potentiality of conduct complex processing raw images.

References

[1] Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 126–135, 2017. 6

[2] Michal Aharon, Michael Elad, Alfred Bruckstein, et al. K-svd: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on signal processing*, 54(11):4311, 2006. 3

[3] Namhyuk Ahn, Byungkon Kang, and Kyung-Ah Sohn. Fast, accurate, and lightweight super-resolution with cascading residual network. pages 252–268, 2018. 1, 3, 6, 7, 8

[4] Antoni Buades, Bartomeu Coll, and J-M Morel. A non-local algorithm for image denoising. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 2, pages 60–65. IEEE, 2005. 3

[5] 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

[6] 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. 3, 7

[7] 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. 6

[8] 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. 3

[9] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009. 5

[10] 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, 2016. 3

[11] Chao Dong, Chen Change Loy, and Xiaoou Tang. Accelerating the super-resolution convolutional neural network. In *European Conference on Computer Vision*, pages 391–407. Springer, 2016. 3

[12] Sina Farsiu, Michael Elad, and Peyman Milanfar. Multiframe demosaicing and super-resolution from undersampled color images. In *Computational Imaging II*, volume 5299, pages 222–234. International Society for Optics and Photonics, 2004. 3

[13] Michaël Gharbi, Gaurav Chaurasia, Sylvain Paris, and Frédo Durand. Deep joint demosaicking and denoising. *ACM Transactions on Graphics (TOG)*, 35(6):191, 2016. 1, 2, 3, 5, 6, 7

[14] 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. 2

[15] Jinjin Gu, Hannan Lu, Wangmeng Zuo, and Chao Dong. Blind super-resolution with iterative kernel correction. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2019. 3

[16] Shuhang Gu, Lei Zhang, Wangmeng Zuo, and Xiangchu Feng. Weighted nuclear norm minimization with application to image denoising. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2862–2869, 2014. 3

[17] 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. 2

[18] Felix Heide, Markus Steinberger, Yun-Ta Tsai, Mushfiqur Rouf, Dawid Pajak, Dikpal Reddy, Orazio Gallo, Jing Liu, Wolfgang Heidrich, Karen Egiazarian, et al. Flexisp: A flexible camera image processing framework. *ACM Transactions on Graphics (TOG)*, 33(6):231, 2014. 6, 7

[19] Keigo Hirakawa and Thomas W Parks. Adaptive homogeneity-directed demosaicing algorithm. *IEEE Transactions on Image Processing*, 14(3):360–369, 2005. 2

[20] 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, 7

[21] 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 112–121. International Society for Optics and Photonics, 2000. 2

[22] 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. 3

[23] 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. 3

[24] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Deeply-recursive convolutional network for image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1637–1645, 2016. 3

[25] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 6

[26] 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. 3

[27] Thomas Köhler, Michel Bätz, Farzad Naderi, André Kaup, Andreas Maier, and Christian Riess. Bridging the simulated-to-real gap: Benchmarking super-resolution on real data. *arXiv preprint arXiv:1809.06420*, 2018. 6

[28] Thomas Köhler, Michel Bätz, Farzad Naderi, André Kaup, Andreas K Maier, and Christian Riess. Benchmarking

super-resolution algorithms on real data. *arXiv preprint arXiv:1709.04881*, 2017. 6

[29] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew P Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *CVPR*, volume 2, page 4, 2017. 3

[30] 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

[31] Henrique S Malvar, Li-wei He, and Ross Cutler. High-quality linear interpolation for demosaicing of bayer-patterned color images. In *Acoustics, Speech, and Signal Processing, 2004. Proceedings.(ICASSP'04). IEEE International Conference on*, volume 3, pages iii–485. IEEE, 2004. 2, 6

[32] David Martin, Charless Fowlkes, Doron Tal, and Jitendra Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, volume 2, pages 416–423. IEEE, 2001. 6

[33] Pietro Perona and Jitendra Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on pattern analysis and machine intelligence*, 12(7):629–639, 1990. 3

[34] Leonid I Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: nonlinear phenomena*, 60(1-4):259–268, 1992. 3

[35] Uwe Schmidt and Stefan Roth. Shrinkage fields for effective image restoration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2774–2781, 2014. 3

[36] Eero P Simoncelli and Edward H Adelson. Noise removal via bayesian wavelet coring. In *Proceedings of 3rd IEEE International Conference on Image Processing*, volume 1, pages 379–382. IEEE, 1996. 3

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

[38] 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, 2013. 2

[39] Nai-Sheng Syu, Yu-Sheng Chen, and Yung-Yu Chuang. Learning deep convolutional networks for demosaicing. *arXiv preprint arXiv:1802.03769*, 2018. 2

[40] Hanlin Tan, Xiangrong Zeng, Shiming Lai, Yu Liu, and Maojun Zhang. Joint demosaicing and denoising of noisy bayer images with admm. In *2017 IEEE International Conference on Image Processing (ICIP)*, pages 2951–2955. IEEE, 2017. 6, 7

[41] 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

[42] Chi-Yi Tsai and Kai-Tai Song. A new edge-adaptive demosaicing algorithm for color filter arrays. *Image and Vision Computing*, 25(9):1495–1508, 2007. 2

[43] 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. 3

[44] 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 *European Conference on Computer Vision*, pages 63–79. Springer, 2018. 3, 5

[45] 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. 3

[46] Ke Yu, Chao Dong, Liang Lin, and Chen Change Loy. Crafting a toolchain for image restoration by deep reinforcement learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2443–2452, 2018. 2, 3

[47] 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. 3

[48] 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. 3

[49] 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. 3

[50] 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. 2

[51] 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. 6

[52] 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. 6

[53] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. Residual dense network for image super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2472–2481, 2018. 3

[54] Ruofan Zhou, Radhakrishna Achanta, and Sabine Süsstrunk. Deep residual network for joint demosaicing and super-resolution. In *Color and Imaging Conference*, volume 2018, pages 75–80. Society for Imaging Science and Technology, 2018. 3, 5