

# Real-world Noisy Image Denoising: A New Benchmark

Jun Xu<sup>1</sup>, Hui Li<sup>1</sup>, Zhetong Liang<sup>1</sup>, David Zhang<sup>1,2</sup>, *Fellow, IEEE*, and Lei Zhang<sup>1,\*</sup>, *Fellow, IEEE*

<sup>1</sup>Department of Computing, The Hong Kong Polytechnic University, Hong Kong SAR, China

<sup>2</sup>School of Science and Engineering, The Chinese University of Hong Kong (Shenzhen), Shenzhen, China

**Abstract:** Most of previous image denoising methods focus on additive white Gaussian noise (AWGN). However, the real-world noisy image denoising problem with the advancing of the computer vision techniques. In order to promote the study on this problem while implementing the concurrent real-world image denoising datasets, we construct a new benchmark dataset which contains comprehensive real-world noisy images of different natural scenes. These images are captured by different cameras under different camera settings. We evaluate the different denoising methods on our new dataset as well as previous datasets. Extensive experimental results demonstrate that the recently proposed methods designed specifically for realistic noise removal based on sparse or low rank theories achieve better denoising performance and are more robust than other competing methods, and the newly proposed dataset is more challenging. The constructed dataset of real photographs is publicly available at <https://github.com/csjunxu/PolyUDataset> for researchers to investigate new real-world image denoising methods. We will add more analysis on the noise statistics in the real photographs of our new dataset in the next version of this article.

**Index Terms**—Image denoising, Real-world noisy images, benchmark datasets

## I. INTRODUCTION

DURING the past decades, the statistical property of real-world noise has been studied for CCD and CMOS image sensors [1]–[5]. There are five major sources for real-world noise, including photon shot noise, fixed pattern noise, dark current, readout noise, and quantization noise, etc. The shot noise is one inevitable source of noise, which is induced by the stochastic arrival process of photons to the sensor. The arrival of photons can be modeled by a Poisson process in which the number of photons arriving the sensor follows a Poisson distribution. This type of noise is proportional to the mean intensity of the specific pixel and is not stationary across the whole image. The fixed pattern noise include pixel response non-uniformity (PRNU) noise and dark current non-uniformity (DCNU) noise. In PRNU noise, each pixel will have a slightly different output level or response for a fixed light level. The major cause of the PRNU noise is the loss of light and color mixture in the neighboring pixels. The DCNU noise comes from the electronics within the sensor chip, and it is generated due to thermal agitation, even there is no light reaching the camera sensor. The readout noise and quantization noise come from the discretization of measured signals. The readout noise is generated during the process of charge-to-voltage conversion, which is inherently not accurate. The quantization noise is generated when the readout values are quantized to integers. The final pixel values are discretizations of the original raw pixel values. Other noise include CCD specific sources such as transfer efficiency, and CMOS specific sources such as column noise.

Different from additive white Gaussian noise (AWGN), the real-world noise is signal dependent, and cannot be modeled by an explicit distribution. It becomes much more complex after being processed in the camera imaging pipelines. Hence, removing noise from real-world noisy images is a more challenging task than its synthetic AWGN counterpart. Another issue about real-world image denoising is how to evaluate the

quality of the denoised images. The image quality assessment by subjective evaluation would be time-consuming, since it needs huge number of subjects to take part in the evaluation experiments. An alternative is to resort to the objective evaluation. However, since the real-world noisy images have no corresponding “ground truth” images, the objective evaluation on the quality of denoised images is very hard. Another choice is to resort to some blind image quality assessment (BIQA) methods [6], [7]. However, these BIQA methods are mostly developed based on the commonly used datasets such as TID dataset [8] and LIVE IQA dataset [9], whose images have very different properties from real-world noisy images.

Recently, several works have been done to address the issue of missing corresponding “ground truth” image of the captured real-world noisy image. In [4], a dataset containing 11 scenes is constructed for analyzing the properties of real-world noise produced in the camera imaging pipeline. However, this dataset is limited in several aspects. It contains only printed pictures on the package of several products, having few real objects. Other problems include that the intensity transform does not model heteroscedastic noise, and low-frequency bias is not removed, etc. In [3] and [5], the corresponding “ground truth” image of the captured real-world noisy image is captured with low ISO values (e.g., ISO=100), with other post-processing steps such as linear intensity changes, spatial misalignment, and low-frequency residual correction, etc. However, the “ground truth” images with low ISO values may have slightly different illuminations from the corresponding real-world noisy images captured under high ISO values (e.g., ISO=6,400). Besides, the post-processing steps may introduce human bias into the “ground truth” images. The work of [10] proposes a less tedious capture protocol similar to [5], where multiple exposures of a static scene are used to aggregate the measurements at every pixel site temporally. The works of [11], [12] propose to illuminate the sensor with approximately constant irradiation and subsequently aggregates intensity measurements spatially.

This is repeated for different irradiation levels to capture the intensity dependence of the noise. In contrast, in [5] the employed Tobit regression allows to estimate the parameters of the noise process by having access to just two images.

In this work, we construct a large dataset of real-world noisy images with reasonably obtained corresponding “ground truth” images. The basic idea is to capture the same and unchanged scene for many (e.g., 500) times and compute their mean image, which can be roughly taken as the “ground truth” image for the real-world noisy images. The rational of this strategy is that for each pixel, the noise is generated randomly larger or smaller than 0. Sampling the same pixel many times and computing the average value will approximate the truth pixel value and alleviate significantly the noise.

## II. EXISTING DATASETS

Currently, there are some datasets available aiming at benchmarking the denoising methods on real-world noisy images [3]–[5].

As far as we know, the RENOIR dataset [3] is the first dataset on real-world noisy images with “ground truth” noise-free images. The cameras used in this dataset are Canon Rebel T3i, Canon S90, and Xiaomi T3i. The authors took photos of a static scene with different ISO values. However, the post-processing is less refined. Image pairs appear to exhibit spatial misalignment, the intensity transform does not model heteroscedastic noise, and low-frequency bias is not removed. In [3], experiments have been conducted to validate that ignoring these factors makes the dataset less useful. It is often useful to measure the noise characteristics of a sensor at a certain ISO level. It was proposed [3] to illuminate the sensor with approximately constant irradiation and subsequently aggregate intensity measurements spatially. This is repeated for different irradiation levels to capture the intensity dependence of the noise. A less tedious capture protocol was also proposed in [3], where multiple exposures of a static scene are used to aggregate the measurements at every pixel site temporally. The detailed description of this dataset is listed in Table I.

The second work along this direction is reported in [4], which involves 11 static scenes. The real-world noisy images and the corresponding “ground truth” images are collected. For each scene, 500 JPEG images are captured and the mean image of the 500 images is roughly taken as the “ground truth” image. Utilizing the mean of temporal images as “ground truth” image has also been employed in [13], [14], but the authors did not build a benchmark dataset. In the dataset of [4], the images are mostly with resolution of  $7630 \times 4912$  and captured by Nikon D800 (ISO=1,600, 3,200, and 6,400), Nikon D600 (ISO=3,200), and Canon 5D Mark III (ISO=3,200). There are totally 15 cropped regions of size  $512 \times 512$  provided for evaluating different denoising methods. The major problems of this dataset is that the captured images are almost printed scenes, which share similar noise statistical property. The camera settings such as the ISO values are also somewhat limited. The detailed description of this dataset is listed in Table II.

One recent benchmark is reported in [5]. Different from the previous two datasets in [3] and [4], [5] employs the Tobit

regression to estimate the parameters of the noise process by accessing just two images. In order to obviate the unrealistic setting by developing a methodology for benchmarking denoising techniques on real photographs, the authors of [5] captured 50 different pairs of images with different ISO settings and shutter speeds. The image captured with high ISO and faster shutter speed is taken as the real-world noisy image, while the image captured with low ISO and slower shutter speed is roughly taken as the “ground truth” image. To derive better “ground truth”, careful post-processing is designed in [5]. The authors corrected spatial misalignment, coped with the inaccurate exposure parameters through a linear intensity transform based on a novel heteroscedastic Tobit regression model, and removed residual low-frequency bias that stems from minor illumination changes, etc. The proposed dataset is called the Darmstadt Noise Dataset (DND) [5], in which the cameras used for capturing the dataset include Sony A7R, Olympus E-M10, Sony RX100 IV, and Huawei Nexus 6P. The authors extracted the linear raw intensities from the captured images using the free software *Dcraw*, and then normalized the image intensities to the range of  $[0, 1]$  by scaling the black and white levels. One interesting finding is that various denoising techniques that perform well on synthetic noisy images are clearly outperformed by BM3D [15] on realistic photographs. This benchmark delineates realistic evaluation scenarios that deviate strongly from those commonly used in the scientific literature. The detailed description of this dataset is listed in Table III.

## III. THE PROPOSED DATASET

### A. Motivation

As discussed previously, existing real-world noisy image datasets [3]–[5] have several limitations in evaluating existing and future image denoising methods. These limitations include camera brands, camera settings, and captured scenes, etc.

**Camera Brands:** In the RENOIR dataset [3], the authors used two different camera brands, Canon (T3i and S90) and Xiaomi (Mi3), for image collection. In the dataset [4], the authors also used two camera brands, i.e., the Canon (5D) and Nikon (D600 and D800), for image collection. In the DND dataset [5], the authors used three different cameras including Sony (A7R and RX100 IV), Olympus (E-M10), and Huawei (Nexus 6P).

**Camera Settings:** In the RENOIR dataset [3], the “ground truth” images are all captured by setting the ISO as 100. The ISO in noisy images are set as follows: for Xiaomi Mi3, the ISO is set as 1,600 or 3,200; for Canon S90, the ISO is set as 640 or 1,000; for Canon T3i, the ISO is set as 3,200 or 6,400. For all the cases except for the reference image of Canon S90, the shutter speed is set as automatic. For Canon S90, the shutter speed is set as 3.2 seconds. In the dataset [4], three different ISOs (e.g., 1,600, 3,200, and 6,400) are employed when capturing images with Nikon D800, while ISO=3,200 is utilized for Canon 5D and Nikon D600. The DND dataset [5], the ranges of ISO are 100 ~ 25,600 for Sony A7R, 200 ~ 25,600 for Olympus E-M10, 125 ~ 8,000 for Sony RX100 IV, and 100 ~ 6,400 for Huawei Nexus 6P, respectively.

TABLE I: Cameras and camera settings used in the dataset [3].

Camera	Sensor Size (mm)	# of Scenes	“Ground Truth”		Noisy Image	
			ISO	Time (s)	ISO	Time (s)
Canon S90	$7.4 \times 5.6$	40	100	3.2	640, 1k	Auto
Canon T3i	$22.3 \times 14.9$	40	100	Auto	3.2k, 6.4k	Auto
Xiaomi Mi3	$4.69 \times 3.52$	40	100	Auto	1.6k, 3.2k	Auto

TABLE II: The detailed information of the cropped regions from the dataset [4].

Camera	ISO	# of Images	JPEG	Image Size
Canon 5D Mark III	3.2k	3	Fine	$512 \times 512$
Nikon D600	3.2k	3	Normal	$512 \times 512$
Nikon D800	1.6k, 3.2k, 6.4k	9	Normal	$512 \times 512$

TABLE III: Cameras and camera settings used in the dataset [5].

Camera	# of Scenes	Sensor Size (mm)	ISO
Sony A7R	13	$36 \times 24$	100-25.6k
Olympus E-M10	13	$17.3 \times 13$	200-25.6k
Sony RX100 IV	12	$13.2 \times 8.8$	125-8k
Huawei Nexus 6P	12	$6.17 \times 4.55$	100-6.4k

**Captured Scenes:** The RENOIR dataset [3] captures 40 scenes for each camera brand, and overall 120 scenes are included in the dataset. However, the noisy images and corresponding “ground truth” images in this dataset have distinct color difference, which is largely caused by inconsistent lighting conditions. The dataset [4] contains only 11 indoor scenes, which are overlapped by similar contents and objects. Though containing 50 different scenes, the “ground truth” images in the DND dataset [5] are not accessible yet. This limits the evaluation of the proposed denoising methods on visual quality.

**Discussion.** Among the above mentioned factors, the camera settings are very important when we capture the real-world noisy images, while the camera brands and captured scenes are relatively easy to improve. The camera settings include mainly ISO value, the shutter speed, and the aperture, etc. In general, the faster the shutter speed, the darker the captured images when we fix the other settings, and vice versa. Similarly, the smaller the ISO value (or aperture), the darker the captured images when we fix the other settings, and vice versa.

In order to make the image less affected by the change of environment (e.g., object motion, change of illumination, camera shake, etc.), the shutter speed should be set as faster as possible. For example, the shutter speed of the Sony A7II camera is between 1/80,000 second and 30 seconds. Given suitable aperture and ISO, it is possible to capture images with normal illuminations when we set the shutter speed between 1/100 second and 1 second. The aperture could be set as any value as long as it is in the reasonable range. The aperture of the Sony camera is between F3.5 and F22. Setting the aperture between F3.5 and F15 can allow us to obtain images with normal illumination under the fixed ranges of shutter and ISO. In our capturing process, we fixed the shutter and aperture in a suitable range, and tuned the ISO values according to the

given camera. In general, the noise level would be higher when the ISO is higher. We set the ISO values from a low value to a high value with fixed gap to more comprehensively evaluate the denoising methods.

To analyze how ISO, shutter speed, and aperture influence the contents and illumination of the captured images, we perform some heuristic experiments with different camera settings. In Figure 1, we show some images captured with different camera settings. Comparing Figures 1(a) and 1(b) (or 1(g) and 1(i)), we can find that higher aperture results in darker illumination. Comparing Figures 1(b) and 1(c) (or 1(g) and 1(h)), we can find that slower shutter results in brighter illumination. Comparing Figures 1(b), 1(d), 1(e), 1(f), and 1(g), we can find that the illumination becomes brighter when the ISO is higher. Besides, given fixed aperture and shutter, the images captured by the camera can avoid the over-exposure or under-exposure when the ISO is set between 400 and 3,200. When ISO=200, the captured images would have the problem of under-exposure, while when ISO=6,400, the captured images would have the problem of over-exposure. However, this can be alleviated by changing the aperture and shutter speed and finally we can obtain images with normal illuminations.

Given fixed shutter speed and aperture, enhancing the camera sensitivity will generate stronger noise than lowering the shutter speed. In the construction of our dataset, we only changed the ISO values while fixing shutter speed and aperture with suitable values to ensure that the images will not suffer over-exposure or under-exposure. It is commonly accepted that the noise in images will become stronger when the scene is under darker light conditions.



Fig. 1: Captured images with the Sony A7 II camera under different (ISO, Shutter speed, Aperture) settings.

### B. The Dataset Construction Process

To alleviate the limitations of the previous datasets [3]–[5], we propose to construct a new dataset which could: 1) contain more camera brands; 2) contain **more carefully designed camera settings**; 3) capture more real-world scenes with realistic objects; 4) capture both the raw data and sRGB data for comparison analysis. The captured images are stored in raw data and JPEG format without compression. For each scene, we capture it for 500 times. Figure 2 shows how we capture images of a static scene in indoor environment. The camera is fixed by a tripod. The data collection is automatically done with shutter release after the button is pressed by a person. Hence, the misalignment problem can be nearly avoided in the acquisition process of 500 images for one scene. We capture images with different camera settings. The cameras are set based on the following rules. First, the shutter speed should be faster than the blink of the fluorescent lights, otherwise the flickering of the light will make the global luminances of the captured images very different. Second, we **set the shutter speed, the aperture, and the ISO value to ensure that the scenes are in a naturally lighting condition**. Besides, since the digital single-lens reflex cameras (DSLRs) use mechanical shutter, the shutter speed of each shot is a little different. This small difference results in **slightly different brightness of different shots**. However, we **ignore this small difference in our dataset**, as that in [4].

**More Camera Brands:** In our dataset, we use 5 different cameras of three camera brands, including Canon (Mark 5D, 80D, 600D), Nikon (D800), and Sony (A7 II), to capture

real-world noisy images. According to a recent survey [16], the three camera brands occupy 48 of 50 most commonly used camera-lens combinations. Hence, our dataset is more comprehensive than the previous datasets on camera brands.

**More Camera Settings:** In our new dataset, each scene is captured **with 6 different ISO settings**, e.g., 800, 1,600, 3,200, 6,400, 12,800, and 25,600. For each ISO setting, we carefully adjust the shutter speed and aperture, and choose other suitable camera settings to make the captured scene neither under-exposed nor over-exposed. With the increase of ISO, the luminance of the captured images will also increase, and we tune the shutter speed and aperture accordingly to capture an image with normal luminance. **For example, to make the images captured with ISO=25,600 be normally exposed, we set the shutter speed to 1/320 second and aperture to F10.0.**

**More Captured Scenes:** We capture the images with indoor normal lighting condition, dark lighting condition, and outdoor normal lighting condition. The scenes we captured are also versatile (including the buildings, classrooms, coffee rooms, and outdoor scenes, etc.). The objects in the scenes include books, pens, bottles, boxes, and joys, etc. In summary, we capture totally 40 different scenes by using 5 different cameras in different camera settings, including 12 scenes captured by Canon 5D Mark II, 5 scenes captured by Canon 80D, 3 scenes captured by Canon 600D, 13 scenes captured by Nikon D800, 7 scenes captured by Sony A7 II. Since the images are of large size ( $3000 \times 3000$ ), we crop some regions from these images and obtain 100 regions of size  $512 \times 512$ .

**Removing Outlier Images:** The outlier images are those



Fig. 2: The static scene is captured with a camera fixed by tripod. The data collection is automatically done with shutter release after the button is pressed by a person.

images which have misalignment or different illuminance from the base image (we usually choose the first image of the 500 shots as the base image). In the dataset of [4], the authors did not remove the images with misalignment or different luminances. In the DND dataset [5], the authors corrected the misalignment of each image. However, this operation largely depends on the misalignment detection method and the correction method, which may make the corrected images less natural. Besides, the DND dataset [5] takes the image captured with low ISO as “ground truth”, and linearly transfer the noisy image captured with high ISO to the scale of the “ground truth” image. This step, in our opinion, is problematic since the image pixels are not linearly dependent on the ISO values. In our dataset, we browse the captured images and remove the outlier images with clear misalignment or different luminances. For each scene, three volunteers are invited to do the screening successively, and the remaining images are used to compute the “ground truth” image.

**Generating “Ground Truth” Image:** The “ground truth” images of the RENOIR dataset [3] are generated when the camera is set with ISO=100, while the other settings are fixed the same as those for the noisy images. The “ground truth” images of the dataset [4] are generated by averaging the static images captured on the same scene under the same camera setting. The “ground truth” images of the DND dataset [5] is generated mainly by using low ISO values (e.g., ISO=100), and other post-processing steps include linear intensity changes, spatial misalignment, and low-frequency residual correction, etc. In our dataset, we employ the same strategy as that method of [4] due to its simplicity. We capture images of the same static scenes for many (500 ~ 1000) times and average the captured images to obtain the “ground truth” image.

We first remove the images with misalignment by careful subjective evaluation. The images with several pixels displacement will be deleted. After this stage, we will remove the images with inconsistent luminance. The luminance is affected by two factors. One is the lighting conditions of the environment. The other is that the camera will automatically make up the illumination when the scene is in a relatively dark lighting condition. Since we shot the scene for many times, some shots may have different illumination, though captured under the same lighting condition. To remove the images with outlier luminance, we first sample 10,000 pixels uniformly (the pixels are on the 100 equidistantly sampled rows and 100 equidistantly sampled columns) from each image, and then compute the mean luminance of the 10,000 pixels. Each of the captured images will have one value representing its mean luminance. We sort these values in a descending order. The images with the lowest or highest mean luminances will be referred as outlier images. We remove these images until the lowest and highest mean luminances are close enough to the “center” of the mean luminances. Here, “center” means the median of the sorted mean luminances. In this way, the images which are much darker or much brighter than the “center” image with the “center” luminance will be deleted, and the remaining images are very close to each other in luminance. The remaining images will be averaged to obtain the mean image, which will be used as the “ground truth” image of each scene.

### C. Summary of the Dataset

In our constructed dataset, we captured images from 40 different scenes with different contents and objects. Figure 3 shows some samples of the real-world noisy images in

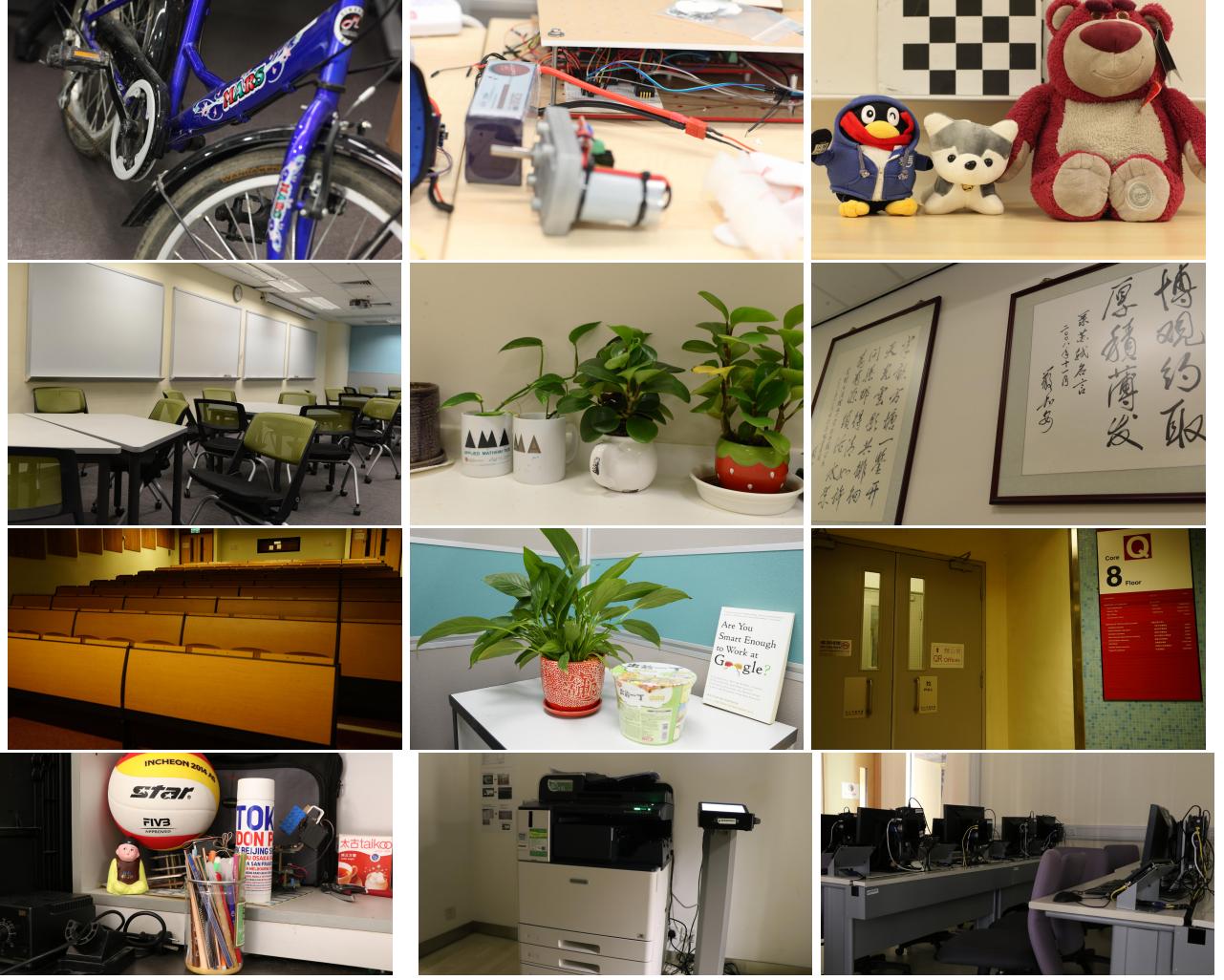


Fig. 3: Some sample images in our newly constructed dataset.

our dataset. The images cover from different types of indoor scenes and versatile objects, etc.

Since the images we captured are very large in size, we **crop 100 regions of size  $512 \times 512$  from the 40 scenes** to evaluate the existing image denoising methods. Some examples of the cropped regions and their corresponding “ground truth” images are listed in Figure 4. Besides, one can see that the “ground truth” image contains much less noise than the noisy image and has much better visual quality. Hence, this dataset provides us a good platform for evaluating the image denoising methods. The detailed description on cameras and camera settings is listed in Table IV. One can see that in our dataset, the ISO values are more comprehensive than the previous datasets [3]–[5].

#### IV. EXPERIMENTS

##### A. Benchmark Datasets

To better evaluate the effectiveness of existing image denoising methods, we **apply the competing methods on existing datasets [4], [5] and our constructed new dataset**. Since the captured real-world noisy images in [3] have clear color

differences with the corresponding “ground truth” images, we do not evaluate image denoising methods on this dataset.

**Dataset 1** is provided in [4], which includes noisy images of 11 static scenes captured by Canon 5D Mark 3, Nikon D600, and Nikon D800 cameras. 15 regions of size  $512 \times 512$  were cropped to evaluate different denoising methods.

**Dataset 2** is called the Darmstadt Noise Dataset (DND) [5], which includes 50 different pairs of images of the same scenes captured by Sony A7R, Olympus E-M10, Sony RX100 IV, and Huawei Nexus 6P. The authors cropped 20 bounding boxes of  $512 \times 512$  pixels from each image in the dataset, yielding 1,000 testing crops in total. However, the “ground truth” images are not open access, and we can only submit the denoising results to the authors’ [Project Website](#) and get the PSNR and SSIM [17] results.

**Dataset 3** is our constructed dataset, which includes noisy images of 40 static scenes captured by Canon 5D Mark II, Canon 80D, Canon 600D, Nikon D800, and Sony A7 II cameras. 100 regions of size  $512 \times 512$  were cropped to evaluate different denoising methods.

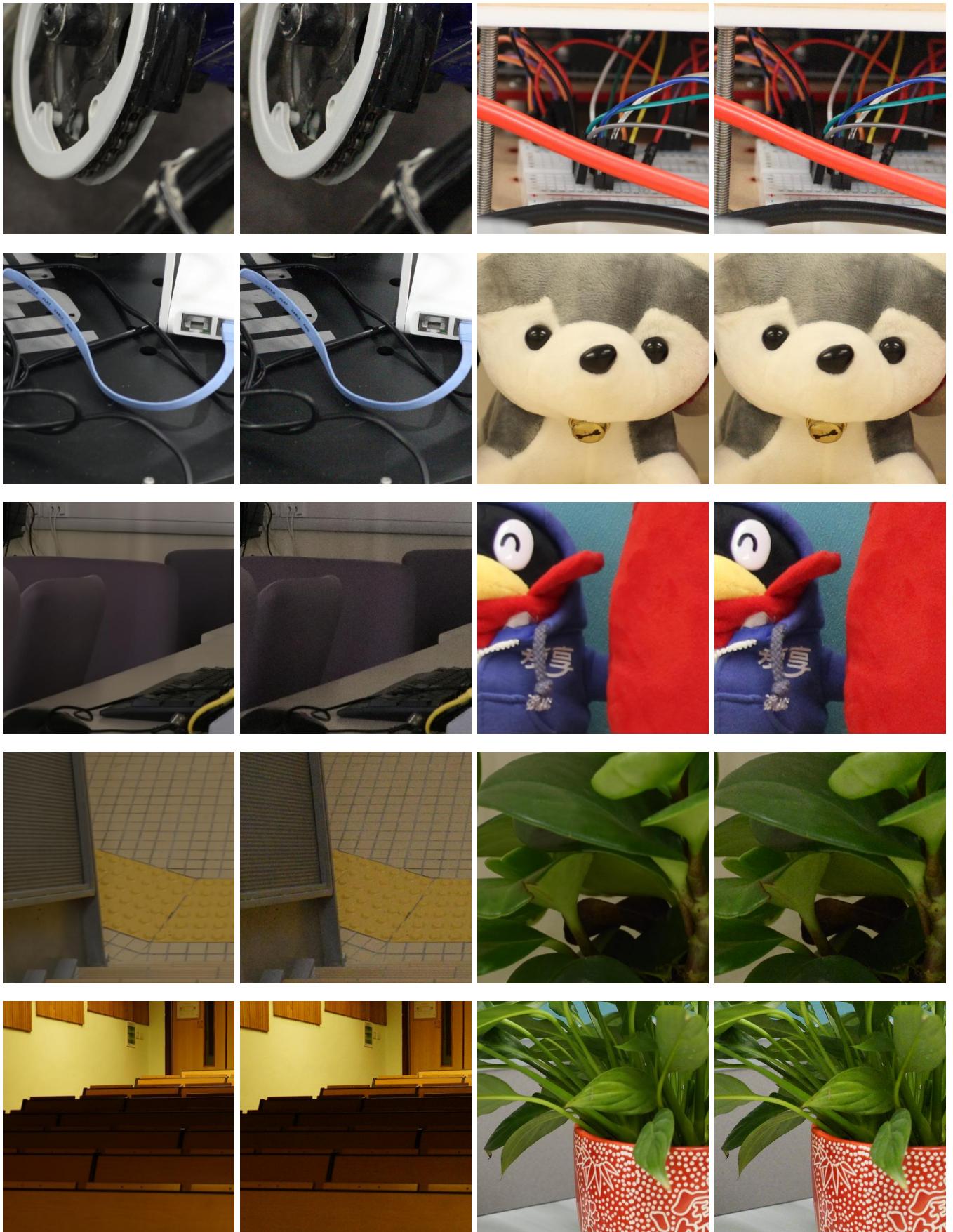


Fig. 4: Some cropped regions of the “ground truth” images (left) and their corresponding noisy images (right) in our constructed dataset.

TABLE IV: Cameras and camera settings used in our new dataset.

Camera	# of Scenes	Sensor Size (mm)	# of Cropped Regions	ISO
Canon 5D	10	36 × 24	29	3.2k,6.4k
Canon 80D	6	22.5 × 15	15	800,1.6k,3.2k,6.4k,12.8k
Canon 600D	5	22.3 × 14.9	11	1.6k,3.2k
Nikon D800	12	35.9 × 24	33	1.6k,1.8k,3.2k,5k,6.4k
Sony A7II	7	35.8 × 23.9	12	1.6k,3.2k,6.4k

### B. Comparison Methods

With the proposed dataset, we make a comprehensive evaluation on the state-of-the-art image denoising methods, including CBM3D [18], Expected Patch Log Likelihood (EPLL) [19], Patch Group Prior based Denoising (PGPD) [29], Nonlocally Centralized Sparse Representations (NCSR) [21], Weighted Nuclear Norm Minimization (WNNM) [23], multi-layer perception (MLP) [20], Cascades of Shrinkage Fileds (CSF) [22], Trainable Nonlinear Reactive Diffusion (TNRD) [24], the residual network based method DnCNN [25], the “Noise Clinic” method [26], [27], the commercial software Neat Image [28], and the recently proposed methods external prior guided internal prior learning for image denoising (Guided) [30], Multi-channel Weighted Nuclear Norm Minimization (MCWNNM) [31], the Trilateral Weighted Sparse Coding (TWSC) [32]. CBM3D is a state-of-the-art color image denoising method, which assumes that the noise is AWGN. The EPLL, PGPD, NCSR, WNNM, MLP, CSF, TNRD, DnCNN are state-of-the-art methods for AWGN noise removal on greyscale images, and we apply these methods on each channel of the realistic color images. The “Noise Clinic” (NC) is a blind image denoising method while Neat Image (NI) is a set of commercial software for image denoising, which has been embedded into Photoshop and Corel Paint Shop. Besides, the method of DnCNN [25] can also deal with real-world noisy images. the recently proposed methods includes Guided, MCWNNM, and TWSC, which are proposed for real-world noisy image denoising.

**Noise level estimation for comparison methods.** For the CBM3D method, the standard deviation of noise on color images should be given as a parameter. For methods of EPLL, PGPD, NCSR, WNNM, MLP, CSF, and TNRD, the noise level in each color channel should be input. For the DnCNN method, it is trained to deal with noise in a range of levels 0 ~ 55. We retrain the models of discriminative denoising methods MLP, CSF, and TNRD (using the released codes by the authors) at different noise levels from  $\sigma = 5$  to  $\sigma = 50$  with a gap of 5. The denoising is performed by processing each channel with the model trained at the same (or nearest) noise level. **The noise levels ( $\sigma_r, \sigma_g, \sigma_b$ ) in R, G, B channels are estimated via some noise estimation methods** [33], [34]. In this chapter, we employ the method [33] to estimate the noise level for each channel of the input color image.

### C. Results and Discussion

**Results on Dataset 1.** The average PSNR and SSIM results on the 15 cropped images by competing methods are listed in Table V. One can see that the recently proposed Guided,

MCWNNM, and TWSC methods perform much better than the other competing methods. Figure 5 shows the denoised images of a scene captured by Canon 5D Mark 3 at ISO = 3200. One can see that TWSC results in not only higher PSNR and SSIM measurements, but also much better visual quality than other methods.

**Results on Dataset 2.** In Table VI, we list the average PSNR and SSIM results of the competing methods on the 1,000 cropped images in the DND dataset [5]. We can see that on this dataset the Guided, MCWNNM, and TWSC methods achieve much better performance than the other competing methods. Note that the “ground truth” images of this dataset have not been released, but one can submit the denoised images to the project website and get the PSNR and SSIM results. Figure 6 shows the denoised images of a scene captured by a Nexus 6P camera. One can see that the TWSC method results in better visual quality than the other denoising methods.

**Results on Dataset 3.** The PSNR and SSIM [17] results on 100 images of the cropped regions are listed in Table VII. We can see that the traditional methods proposed for AWGN are no longer effective enough for the real-world noisy images. The discriminative methods achieve slightly better performance than the traditional methods, while still being inferior to the methods designed for real-world noisy images. The recently proposed Guided, MCWNNM, and TWSC methods work much better than previous methods. Some visual comparisons are given in Figure 7, from which one can see that the TWSC method removes most of the noise while maintaining the details.

The realistic noise is not AWGN, and this point can be valid in our new dataset. In Figure 8, we compute the mean noise levels ( $\sigma$ ) with respect to different ISO values (left) and Red, Green, and Blue channels (right) on our real-world noisy image dataset. One can see that, with the increasing of the ISO values, the noise levels will be increased. This trend is also true for each channel of the color images. One possible correlation is that the noise level in Green channel would be lower than the other two channels, i.e.,  $\sigma_b, \sigma_r \geq \sigma_g$ .

**Discussion.** The experimental results on the three datasets demonstrate that:

- The methods CBM3D, EPLL, PGPD, NCSR, WNNM, MLP, CSF, TNRD, and DnCNN which are designed for AWGN achieve lower PSNR and SSIM compared with the methods developed for real-world noisy image denoising;
- The denoising methods EPLL, PGPD, NCSR, and WNNM which are designed for grey scale images would generate much artifacts since they process each channel of the RGB image individually [35]. They cannot deal

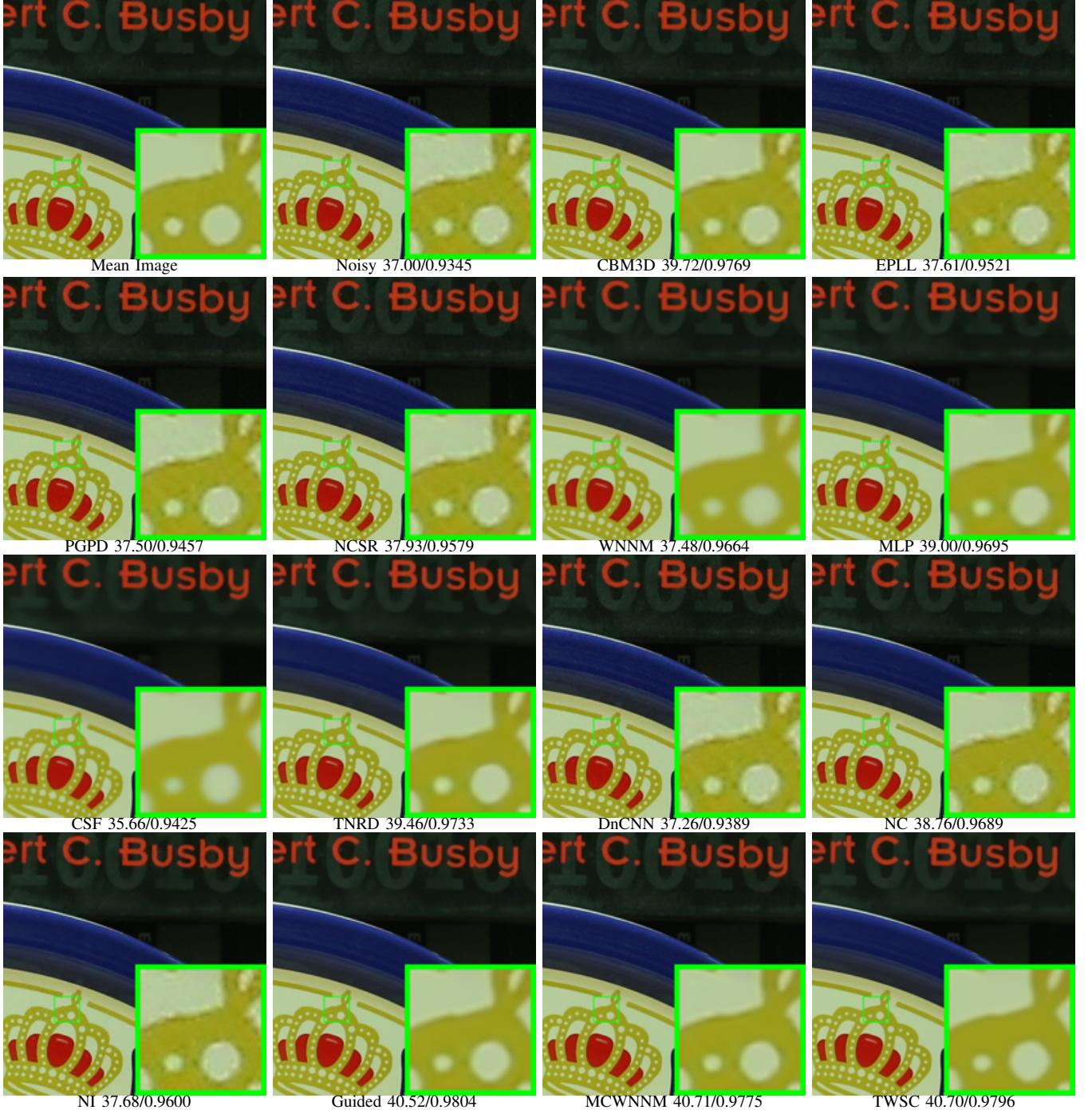


Fig. 5: Denoised images and PSNR (dB)/SSIM results of the real-world noisy image *Canon 5D Mark 3 ISO 3200 1* [4] by different methods. The images are better to be zoomed in on screen.

TABLE V: Average results on PSNR(dB) and SSIM of different denoising algorithms on the 15 cropped images in **Dataset 1** [4].

Metric	CBM3D	EPLL	PGPD	NCSR	WNNM	MLP	CSF
PSNR	35.19	33.66	33.69	33.46	35.77	36.46	35.33
SSIM	0.8580	0.8591	0.8591	0.8512	0.9381	0.9436	0.9250
Metric	TNRD	DnCNN	NC	NI	Guided	MCWNNM	TWSC
PSNR	36.61	33.86	36.43	35.49	37.15	37.71	<b>37.81</b>
SSIM	0.9463	0.8635	0.9364	0.9126	0.9504	0.9542	<b>0.9586</b>

TABLE VI: Average results on PSNR(dB) and SSIM of different denoising algorithms on the 1,000 cropped images in **Dataset 2** [5].

Metric	CBM3D	EPLL	PGPD	NCSR	WNNM	MLP	CSF
PSNR	32.14	32.65	33.12	32.81	33.28	34.02	33.87
SSIM	0.7773	0.7889	0.8002	0.7912	0.8012	0.8201	0.8128
Metric	TNRD	DnCNN	NC	NI	Guided	MCWNNM	TWSC
PSNR	34.15	32.41	36.07	35.11	36.41	37.38	<b>37.94</b>
SSIM	0.8271	0.7897	0.9013	0.8778	0.9101	0.9294	<b>0.9403</b>

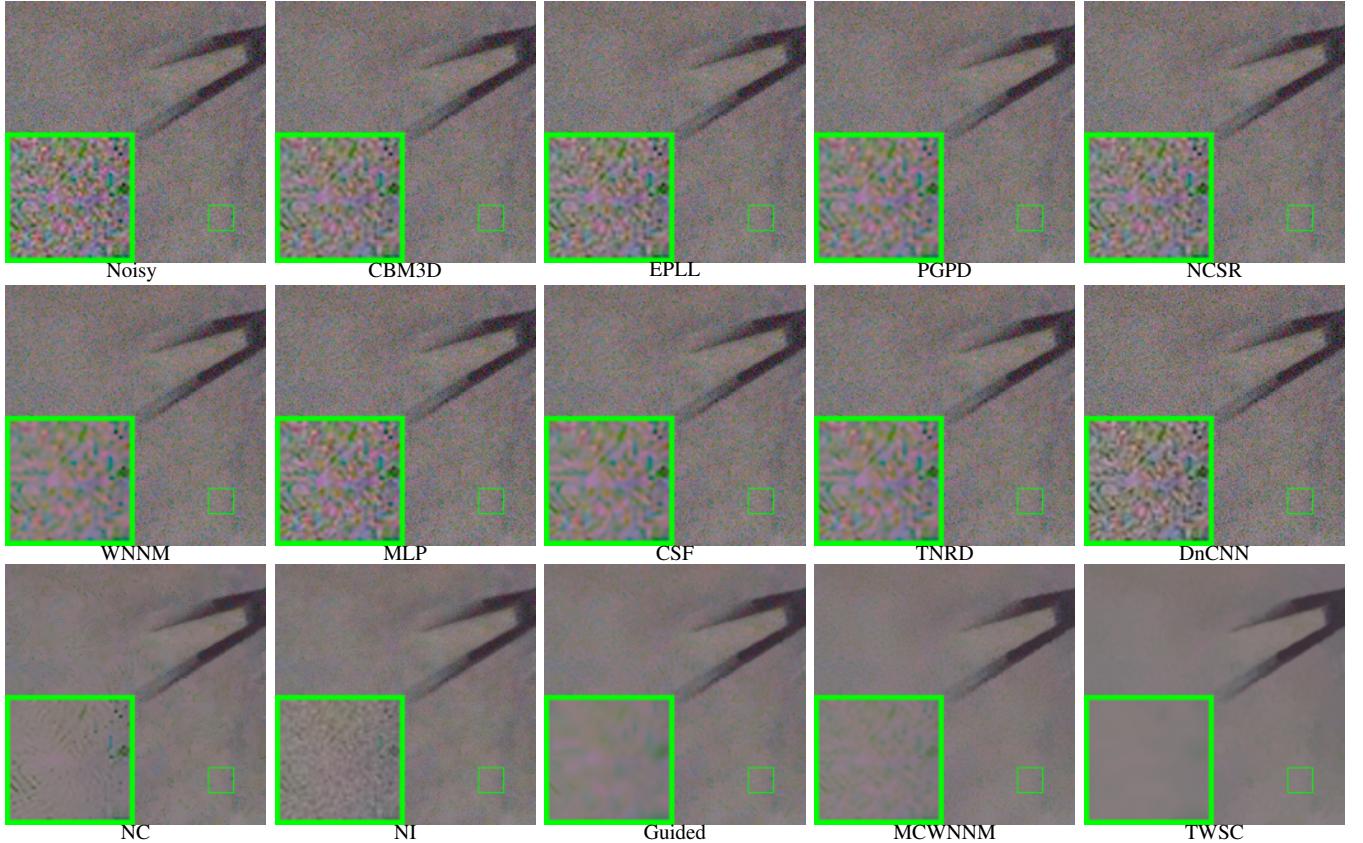


Fig. 6: Denoised images of the real-world noisy image 0001\_1 [5] by different methods. The images are better to be zoomed in on screen.

TABLE VII: Average results on PSNR(dB) and SSIM of different denoising algorithms on the 100 cropped images in our new dataset (**Dataset 3**).

Metric	CBM3D	EPLL	PGPD	NCSR	WNNM	MLP	CSF
PSNR	37.40	36.17	36.18	36.40	36.59	38.07	37.71
SSIM	0.9526	0.9216	0.9206	0.9290	0.9247	0.9615	0.9571
Metric	TNRD	DnCNN	NC	NI	Guided	MCWNNM	TWSC
PSNR	38.17	36.08	36.92	37.77	38.35	38.51	<b>38.60</b>
SSIM	0.9640	0.9161	0.9449	0.9570	0.9669	0.9671	<b>0.9685</b>

with the images which have different noise statistics in different channels as well as different local patches. Hence, these methods may fail to process the real-world noisy images which have complex noise statistics;

- The discriminative learning based methods MLP, CSF, TNRD, and DnCNN are trained on paired clean and

noisy images. These methods largely depends on the training dataset, and would achieve inferior performance on images whose noise has different statistics from those in the training images. Besides, discriminative methods will also be sensitive to the resolution (DPI) of the images in the training set;

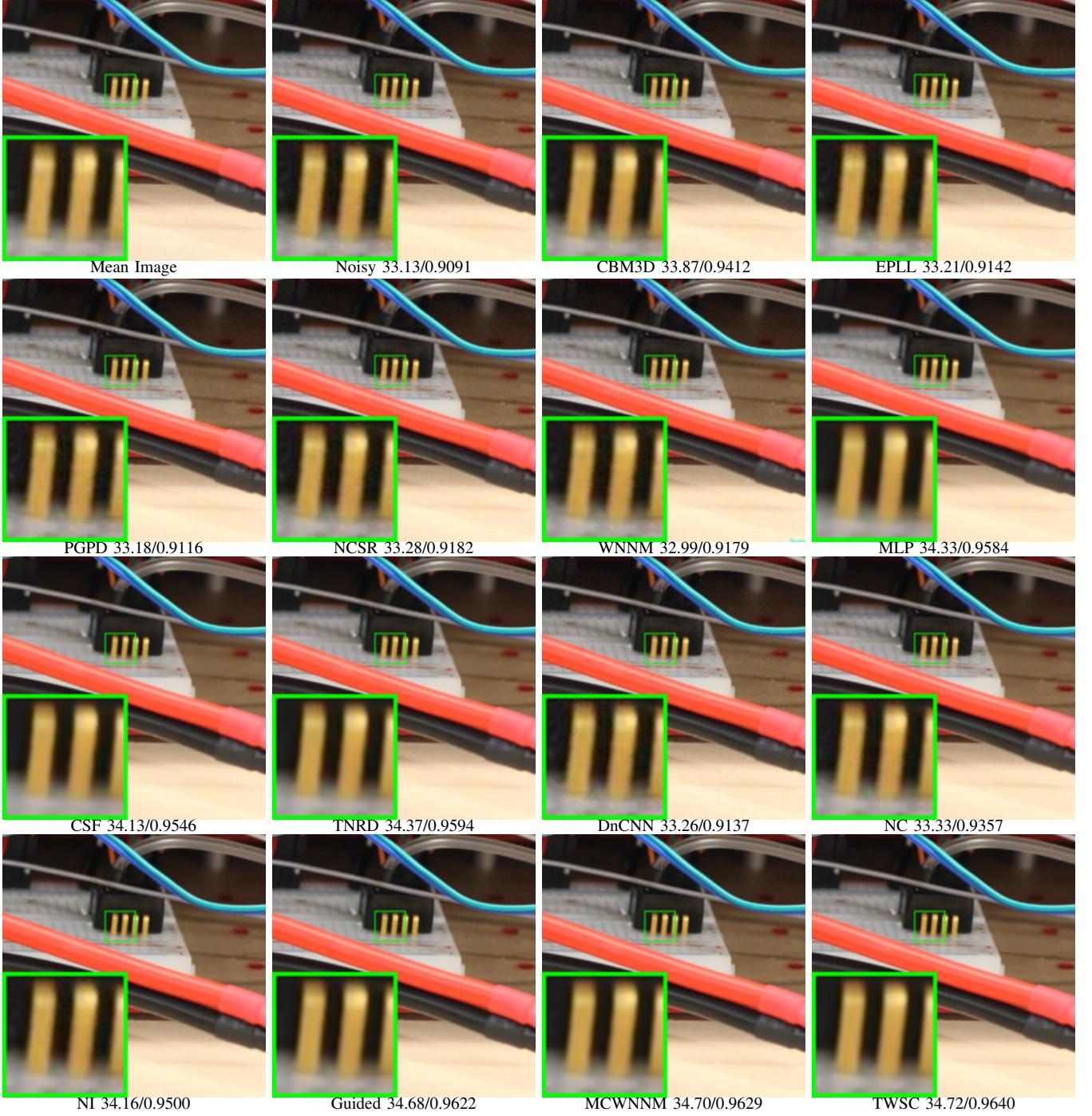


Fig. 7: Denoised images and PSNR (dB)/SSIM results of the real-world noisy image *Canon5D\_2.5\_160\_6400\_circuit\_3* in our new dataset by different methods. The images are better to be zoomed in on screen.

- The performance of recently proposed denoising methods, i.e., Guided, MCWNNM, and TWSC, on datasets [4], [5] are much better than those of previous denoising methods. As we can see from Tables V and VI, the highest PSNR of these methods (TWSC) and the highest PSNR of the methods proposed for AWGN removal (TNRD on Dataset 1 and NC on Dataset 2) have a difference of over 1.2dB. However, **on our new dataset**, as we can see from Tables VII, the **highest PSNR of these methods (TWSC)** and the **highest PSNR of previous**

**methods (TNRD)** only have a difference of around 0.4dB. This indicates that on our new dataset, the **recently proposed methods such as Guided, MCWNNM, and TWSC do not show significant advantages over the previous methods** such as TNRD. This is because **our dataset is more comprehensive** in the scene contents and have more camera settings. This also shows that our dataset is more challenging than previous datasets, and new real-world image denoising methods are needed.

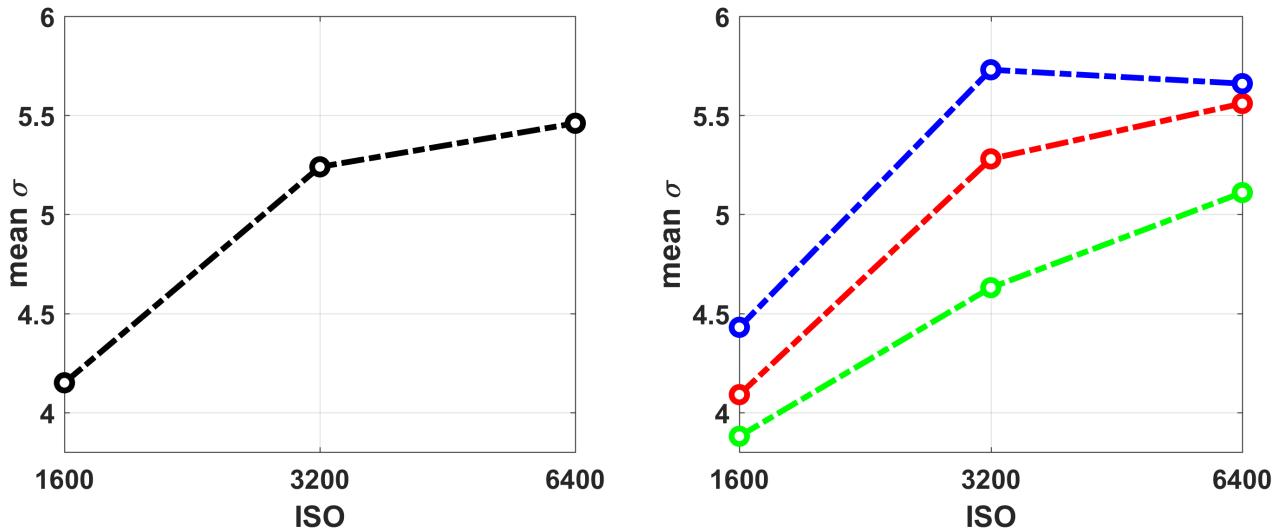


Fig. 8: The mean noise levels ( $\sigma$ ) with respect to different ISO values (left) and Red, Green, and Blue channels (right) on our real-world noisy image dataset.

## V. CONCLUSION

To evaluate the existing denoising methods on real photographs and promote new methods for removing real-world noise, we constructed a novel dataset which contains comprehensive real-world noisy images of different natural scenes. These images were captured by different cameras under different camera settings. Each scene was shot 500 times in a short time. We first selected the images without misalignment, then deleted the images which do not have consistant lumiance with the baseline image. Since the captured images are very large in size, we cropped smaller regions of size  $512 \times 512$  to evaluate the existing image denoising methods and the methods we proposed in previous sections. We evaluated the different denoising methods on the new dataset and previous datasets. The results demonstrated that the proposed methods are more robust than other competing methods, and the newly proposed dataset is more challenging. We will make the constructed dataset of real photographs publicly available for researchers to investigate new real-world image denoising methods.

## REFERENCES

- [1] G. E Healey and R. Kondepudy. Radiometric CCD camera calibration and noise estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(3):267–276, 1994. 1
- [2] Y. Tsai, Visvanathan Ramesh, and Takeo Kanade. Statistical calibration of ccd imaging process. *IEEE International Conference on Computer Vision (ICCV)*, 1:480–487, 2001. 1
- [3] Josue Anaya and Adrian Barbu. Renoir - a dataset for real low-light noise image reduction. *arXiv preprint arXiv:1409.8230*, 2014. 1, 2, 3, 4, 5, 6
- [4] S. Nam, Y. Hwang, Y. Matsushita, and S. J. Kim. A holistic approach to cross-channel image noise modeling and its application to image denoising. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1683–1691, 2016. 1, 2, 3, 4, 5, 6, 9, 11
- [5] Tobias Plötz and Stefan Roth. Benchmarking denoising algorithms with real photographs. *CVPR*, 2017. 1, 2, 3, 4, 5, 6, 8, 10, 11
- [6] Michele A Saad, Alan C Bovik, and Christophe Charrier. Blind image quality assessment: A natural scene statistics approach in the dct domain. *IEEE transactions on Image Processing*, 21(8):3339–3352, 2012. 1
- [7] Anush Krishna Moorthy and Alan Conrad Bovik. A two-step framework for constructing blind image quality indices. *IEEE Signal processing letters*, 17(5):513–516, 2010. 1
- [8] Nikolay Ponomarenko, Vladimir Lukin, Alexander Zelensky, and Karen Egiazarian. Tid2008-a database for evaluation of full-reference visual quality assessment metrics. 10(4):30–45, 2009. 1
- [9] H. R. Sheikh, M. F. Sabir, and A. C. Bovik. A statistical evaluation of recent full reference image quality assessment algorithms. *IEEE Transactions on Image Processing*, 15(11):3440–3451, Nov 2006. 1
- [10] European Machine Vision Association. Emva standard 1288: Standard for characterization of image sensors and cameras. 2012. 1
- [11] A. Foi, S. Alenius, V. Katkovnik, and K. Egiazarian. Noise measurement for raw-data of digital imaging sensors by automatic segmentation of nonuniform targets. *IEEE Sensors Journal*, 7(10):1456–1461, 2007. 1
- [12] Teodor Mihai Moldovan, Stefan Roth, and Michael J Black. Denoising archival films using a learned bayesian model. In *Image Processing, 2006 IEEE International Conference on*, pages 2641–2644. IEEE, 2006. 1
- [13] C. Liu, R. Szeliski, S. Bing Kang, C. L. Zitnick, and W. T. Freeman. Automatic estimation and removal of noise from a single image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(2):299–314, 2008. 2
- [14] X. Liu, M. Tanaka, and M. Okutomi. Practical signal-dependent noise parameter estimation from a single noisy image. *IEEE Transactions on Image Processing*, 23(10):4361–4371, 2014. 2
- [15] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE Transactions on Image Processing*, 16(8):2080–2095, 2007. 2
- [16] The most commonly used camera-lens combinations. <https://explorecams.com/stats/top/models>. Accessed: 2017-09-11. 4
- [17] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004. 6, 8
- [18] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Color image denoising via sparse 3D collaborative filtering with grouping constraint in luminance-chrominance space. *IEEE International Conference on Image Processing (ICIP)*, pages 313–316, 2007. 8
- [19] D. Zoran and Y. Weiss. From learning models of natural image patches to whole image restoration. *IEEE International Conference on Computer Vision (ICCV)*, pages 479–486, 2011. 8
- [20] H. C. Burger, C. J. Schuler, and S. Harmeling. Image denoising: Can plain neural networks compete with BM3D? *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2392–2399, 2012. 8
- [21] W. Dong, L. Zhang, G. Shi, and X. Li. Nonlocally centralized sparse representation for image restoration. *IEEE Transactions on Image Processing*, 22(4):1620–1630, 2013. 8
- [22] U. Schmidt and S. Roth. Shrinkage fields for effective image restoration. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2774–2781, June 2014. 8
- [23] S. Gu, L. Zhang, W. Zuo, and X. Feng. Weighted nuclear norm minimization with application to image denoising. *IEEE Conference on*

- Computer Vision and Pattern Recognition (CVPR)*, pages 2862–2869, 2014. 8
- [24] Y. Chen, W. Yu, and T. Pock. On learning optimized reaction diffusion processes for effective image restoration. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5261–5269, 2015. 8
- [25] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Transactions on Image Processing*, 2017. 8
- [26] M. Lebrun, M. Colom, and J.-M. Morel. Multiscale image blind denoising. *IEEE Transactions on Image Processing*, 24(10):3149–3161, 2015. 8
- [27] M. Lebrun, M. Colom, and J. M. Morel. The noise clinic: a blind image denoising algorithm. <http://www.ipol.im/pub/art/2015/125/>. Accessed 01 28, 2015. 8
- [28] Neatlab ABSoft. Neat Image. <https://ni.neatvideo.com/home>. 8
- [29] Jun Xu, Lei Zhang, Wangmeng Zuo, David Zhang, and Xiangchu Feng. Patch group based nonlocal self-similarity prior learning for image denoising. In *ICCV*, 2015. 8
- [30] Jun Xu, Lei Zhang, and David Zhang. External prior guided internal prior learning for real noisy image denoising. <https://arxiv.org/abs/1705.04505>, 2017. 8
- [31] Jun Xu, Lei Zhang, David Zhang, and Xiangchu Feng. Multi-channel weighted nuclear norm minimization for real color image denoising. In *ICCV*, 2017. 8
- [32] Jun Xu, Lei Zhang, and David Zhang. A trilateral weight sparse coding scheme for robust image denoising. *Unpublished Draft*, 2017. 8
- [33] X. Liu, M. Tanaka, and M. Okutomi. Single-image noise level estimation for blind denoising. *IEEE Transactions on Image Processing*, 22(12):5226–5237, 2013. 8
- [34] G. Chen, F. Zhu, and A. H. Pheng. An efficient statistical method for image noise level estimation. *IEEE International Conference on Computer Vision (ICCV)*, December 2015. 8
- [35] J. Mairal, M. Elad, and G. Sapiro. Sparse representation for color image restoration. *IEEE Transactions on Image Processing*, 17(1):53–69, 2008. 8