LSD_2 – Joint Denoising and Deblurring of Short and Long Exposure Images with Convolutional Neural Networks

[1]Janne Mustaniemi [2]Juho Kannala [3]Jiri Matas [2]Simo Särkkä [1]Janne Heikkilä

[1]Center for Machine Vision and Signal Analysis, University of Oulu, Finland [2]Aalto University, Finland

[3]Center for Machine Perception, Faculty of Electrical Engineering, Czech Technical University in Prague, Czech Republic

janne.mustaniemi@oulu.fi

Abstract

This paper addresses the challenging problem of acquiring high-quality photographs with handheld smartphone cameras in low-light imaging conditions. We propose an approach based on capturing pairs of short and long exposure images in rapid succession and fusing them into a single high-quality photograph using a convolutional neural network. The network input consists of a pair of images, where the short exposure image is typically noisy and has poor colors due to low lighting and the long exposure image is susceptible to motion blur when the camera or scene objects are moving. The network is trained using a combination of real and simulated data and we propose a novel approach for generating realistic synthetic short-long exposure image pairs. Our approach is the first one to address the joint denoising and deblurring problem using deep networks. It outperforms the existing denoising and deblurring methods in this task and allows to produce good images in extremely challenging conditions. Our source code, pretrained models and data will be made publicly available to facilitate future research.

1. Introduction

Capturing high-quality images in difficult acquisition conditions is a formidable challenge. Such conditions, which are not uncommon, include low lighting levels and dynamic scenes with significant motion or high dynamic range, e.g. in the presence of both dark shadows and bright highlights. The problems related to low-light imaging affect all cameras but they are most pronounced in smartphones, the currently most commonly used acquisition device, where the camera and optics need to be small, lightweight and cheap.

The situation is particularly challenging if the device is handheld or the scene is dynamic as no satisfactory compromise between short and long exposure times exists. To get rich colors and good brightness with low noise, one should choose long exposure with low sensor sensitivity setting (ISO number). However, this will cause strong motion blur when the camera is moving (shaking) or if there is motion in the scene. On the other hand, a short exposure and high sensitivity setting will produce sharp but noisy images of moving objects. Examples of such short and long exposure images are shown in Fig. 1.

Figure 1. A noisy, short exposure and a blurry, long exposure image (top) captured by a hand-held tablet at night in 7+210=217 milliseconds. Jointly deblurred and denoised image by the proposed LSD_2 method (bottom left). The real-valued LSD_2 CNN output, approximating the long exposure intensity, may be tone-mapped by an exposure fusion method guided by the short exposure image (bottom right). See Fig. 5 for dynamic scene results.

We propose a novel approach that addresses the aforementioned challenges by taking "the best of both worlds"

via computational photography, avoiding the unsatisfactory trade-off between the short and long exposure settings. The method captures pairs of short and long exposure images in almost instantaneous succession and fuses them into a single high-quality image using a convolutional neural network (CNN). The overall capture time is only fractionally longer than the long exposure time.

Our CNN-based method, called LSD_2[1] performs joint image denoising and deblurring, exploiting information from both images, adapting their contributions to the conditions at hand. Experiments show that LSD_2 handles well situations traditionally falling under the rubric of either blind deblurring or denoising methods. Moreover, the real-valued output of the convolutional neural net may be treated as a high dynamic range image.

Many current mobile devices can be programmed to capture sequences of images with different exposure times in rapid bursts without extra hardware or notable delay. The proposed approach thus brings significant practical benefits in comparison to conventional denoising methods, which are limited by the information in a single image and solve only one of the problems, not covering situations when both blur and noise have to be addressed.

Besides the problems of noise and blur, mobile imaging suffers from the limited dynamic range of camera sensors, which is often more severe in smartphone cameras than in digital single-lens reflex cameras. Even if the user were able to keep the camera perfectly still, the camera might not be able to capture the full dynamic range of the scene with a single exposure. Thus, details are typically lost either in dark shadows or bright highlights. Our approach provides a solution to this problem and produces more faithful colors and brightness values than in single-exposure input images. We note that previous exposure fusion algorithms such as [23] assume that input images are neither blurry nor misaligned.

There are only few papers addressing both the denoising and deblurring problems jointly in a similar setup [36, 32] and, to the best of our knowledge, our work is the first one utilizing deep neural networks for this task.

Our approach has the following key ingredients. We train a U-net-shaped deep convolutional neural network that takes a pair of short-long exposure images as input and provides a single high-quality image as output. The network is trained using both simulated and real data. Large volume of simulated data is generated from regular high-quality photographs by synthesizing both under- and over-exposed images and a realistic blur to the latter. Real training data are acquired by capturing image pairs of static scenes with varying exposure times using a tripod. The long exposure image in each real pair is the ground truth target for the network and the blurred input is obtained by adding synthetic blur to it. Additionally, we train a second U-net for exposure fusion, which takes the short-exposure image and the output of the LSD_2 network as input and produces a tone-mapped result as shown in Fig. 1 (bottom right).

The main contributions of the paper are the following:

- We present LSD_2, the first joint denoising and deblurring approach based on convolutional neural networks, and show results superior to the state-of-the art. The network will be made public.
- We propose a novel approach for generating realistic training and evaluation data. The data will be published to facilitate future research.
- We show that processing the output of the LSD_2 network with an exposure fusion network achieves better reproduction of colors and brightness than a single-exposure smartphone image.
- We will publish the Android software we developed for acquisition of the back-to-back short and long exposure images, enabling reproducibility of our results and further exploitation of multi-exposure imagery.

2. Related work

Single-image denoising is a classical problem, which has been addressed using various approaches such as sparse representations [6], transform-domain collaborative filtering [4] or nuclear norm minimization [8]. In addition, several deep learning based approaches have been proposed recently [13, 2, 37, 16]. Typically the deep networks are trained with pairs of clean and noisy images [13, 2, 37], but it has been shown that training is possible without clean targets [16]. Besides the end-to-end deep learning approaches there are methods that utilize either conventional feed-forward networks [38] or recurrent networks [3] as learnable priors for denoising. Randomly initialized networks have been used as priors without pretraining [30]. Many of the recent methods can be applied to other restoration tasks, such as inpainting [16, 30] and single-image super-resolution [37, 3]. Nevertheless, in contrast to our approach, the aforementioned methods focus on single image restoration and do not address multi-image denoising and deblurring, which is essential in our case.

Single-image deblurring is an ill-posed problem and various kind of priors have been utilized to regularize the solutions. For example, the so called dark and bright channel priors [21, 35] have been used with promising results. However, these methods assume spatially invariant blur which limits their practicality. Priors based on deep networks have also been proposed [38]. There are end-to-end approaches, where a neural network takes the blurry image as input and directly outputs a deblurred result [20, 19, 15]. Some methods utilize inertial sensor data in addition to images [18, 10].

[1] LSD_2 stands for Long-Short Denoising and Deblurring.

Other methods first estimate blur kernels and thereafter perform non-blind deconvolution [28, 7], and some approaches utilize deep networks for removing the deconvolution artifacts [26, 31]. Despite recent progress, single-image deblurring methods often fail to produce satisfactory results since the problem is very challenging and ill-posed. That is, unlike our approach, the aforementioned methods can not utilize a sharp but noisy image to guide the deblurring.

Recently, several multi-image denoising [9, 17] or deblurring approaches [5, 34, 33, 1] have been proposed that are based on processing a burst of input images that are captured consecutively. However, unlike our approach, these methods do not vary the exposure time of the images but use either short or long exposure bursts and, hence, they address either denoising or deblurring, but not both problems jointly like we do. Moreover, since the characteristics of their input images are not as complementary as in our case, they can not get "the best of both worlds" but suffer the drawbacks of either case. For example, a burst of short exposure images may suffer from too low light and low signal to noise ratio in the darkest scene regions, although alignment and weighted averaging of multiple frames can alleviate the problem to some extent [9, 17]. On the other hand, using only relatively long exposure has problems with dynamic scenes as there may be severe spatial misalignment between the images, and the capture time is longer so that fast-moving objects may disappear from the view. On top of that, based on our own observations and earlier studies [17, 1], it seems that due to the non-complementary nature of constant exposure images it is necessary to use more input frames than two and this may increase the consumption of memory and power besides time. Moreover, with a constant exposure the saturated bright regions can not be easily avoided and high dynamic range imaging is not achieved.

A similar problem setting as in our work is considered in [36, 32] but without utilizing CNNs. Both [36] and [32] first estimate blur kernels for the blurry image and thereafter use so-called residual deconvolution, proposed by [36], to iteratively estimate the residual image that is to be added to the denoised sharp image. Both methods use [22] for denoising, and [32] estimates spatially varying blur kernels whereas [36] assumes uniform blur. One limitation of [32] is that their model is not applicable to non-static scenes and it assumes that the motion of the camera during exposure is limited to rotations about its optical center, whereas our approach can generalize to more varied motions. Another drawback of [36] and [32] is that they require a separate photometric and geometric registration stage, where the rotation is estimated manually [36]. We compared our approach to [32] using their images (static scene, pure rotation) and observed that our results are better or comparable despite the fact that the images have unknown exposure times and they are captured with another camera having different noise characteristics than our camera used for training our model (see Fig. 6).

3. Method Overview

The short and long exposure images can be captured with a modern mobile device that supports per-frame camera control. An example is shown in Fig. 1. The short exposure image is sharp but noisy as it is taken with a high sensitivity setting of ISO equal to 800. Notice that the colors are also distorted compared to the long exposure image with ISO equal to 200, which is blurry due to camera motion. Furthermore, the images are slightly misaligned even though they are captured immediately one after the other.

Fig. 2 shows an overview of the proposed LSD_2 method. The goal is to recover the underlying sharp and noise-free image using a pair of blurry and noisy images. The input images are jointly denoised and deblurred by a convolutional neural network similar to U-net [24]. The architecture of the network and training details are covered in Sec. 5.

Capturing real pairs of noisy and blurry images together with the ground truth sharp images is a major challenge. To train the network, we propose a data generation framework that produces realistic training data with the help of real gyroscope readings recorded from handheld movements. Details of the data generation framework are given in the next section. To further improve the performance, the network is fine-tuned with real short and long exposure images captured with a mobile device as described in Sec. 5.3.

4. Data Generation

In order to train the network, we need pairs of noisy and blurry images together with the corresponding sharp images. Since there is no easy way to capture such real-world data, we propose a data generation framework that synthesizes realistic pairs of short and long exposure images. By utilizing images taken from the Internet and gyroscope readings, we can generate unlimited amount of training data with realistic blur while covering a wide range of different scene types.

In the following subsections, we describe the different stages of our data generation pipeline: synthesis of long and short exposure image pairs, addition of noise and realistic blur, and simulation of spatial misalignment. The LSD_2 network operates with images having intensity range $[0, 1]$ and hence we first scale the original RGB values to that range. Since the aforementioned imaging effects occur in linear color space, we invert the gamma correction of the input images. As we do not know the real value of the gamma, it is assumed that $\gamma = 2.2$. Once the images have been generated, the gamma is re-applied.

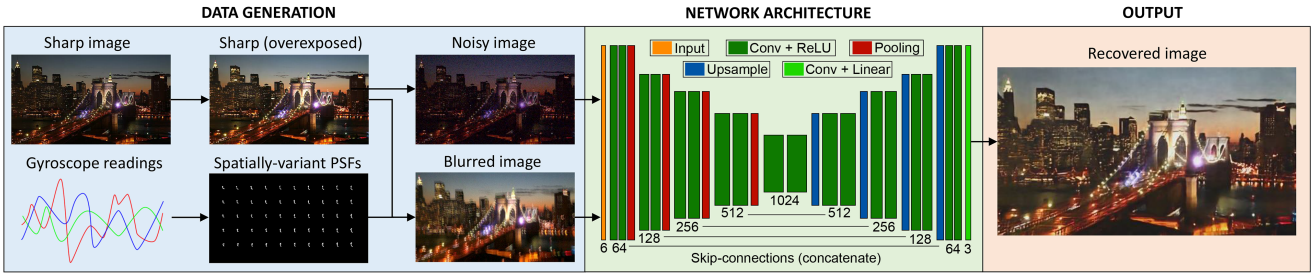

Figure 2. Overview of the proposed joint denoising and deblurring method. To train the network, we generate pairs of short and long exposure images with realistic motion blur, spatial misalignment, image noise, color distortion and saturated regions. The blurred image (misaligned) is generated with the help of gyroscope readings recorded with a mobile device. The output of the network is a sharp and noise-free image, which can be further used for exposure fusion as demonstrated in Fig. 1.

4.1. Synthesis of Long Exposure Images

We take a regular high-quality RGB image I from Internet as the starting point of our simulation. We avoid overexposed or underexposed photographs. However, at test time our long exposure input image should be slightly overexposed in order to enable high dynamic range and ensure sufficient illumination of darkest scene regions. Hence, we need to simulate the saturation of intensities due to overexposure. We do that by first multiplying the intensity values with a random number s uniformly sampled from the interval $[1, 3]$. The short exposure image is generated from this intensity-scaled version sI, as described in the next subsection. Then, by clipping the maximum intensity to value of 1, we get the sharp long exposure image, which will be the ground truth target for network training. That is, we train the network to predict an output with similar exposure as the long exposure image. This enables us to use the real long exposure images captured with a tripod as targets when fine-tuning with real data (Sec. 5.3). In practical use, the degree of overexposure can be controlled by utilizing an auto-exposure algorithm to determine the long exposure time. Further, the performance can be improved by selecting the ratio between the short and long exposure time to be always constant even if the absolute time varies, e.g. based on brightness of the scene. Thus, we record the real image pairs so that the short exposure time is always 1/30 of the long exposure time.

4.2. Underexposure and Color Distortion

The underexposed short exposure image is synthesized from the aforementioned long exposure image sI, where intensities can exceed 1, by applying affine intensity change $(asI + b)$ with random coefficients (a, b) sampled from uniform distributions, whose parameters are determined by analyzing the intensity distributions of real short and long exposure pairs, captured with a constant exposure time ratio (1/30).

Our analysis of real image pairs showed that the colors are often distorted in the noisy short exposure image as show in Fig. 1. Hence, in order to simulate the distortion, we randomly sample different affine transformation parameters (a_i, b_i) for each color channel i. Moreover, the parameters of the uniform distributions for a_i and b_i are determined independently for each color channel and they are such that $a_i < 0.3$ and $b_i < 0.01$ always. By introducing random color distortions, we encourage the network to learn the colors and brightness mainly from the (blurry) long exposure image.

The final short exposure image for network training is obtained by adding noise after the affine intensity change. An example of synthetic short exposure image is shown in Fig. 3 and details of added noise are described in Sec. 4.5.

4.3. Motion Blur

The motion blur is simulated only to the long exposure image sI. Synthetically blurred images are generated with help of gyroscope measurements. Similar to prior work [10, 25], we assume that motion blur is mainly caused by the rotation of the camera. We start by recording a long sequence of gyroscope readings with a mobile device. The device is kept more or less steady during the recording to simulate a real life imaging situation with a shaking hand.

Let t_1 denote the starting time of the synthetic image exposure. It is randomly selected to make each of the blur fields different. The level of motion blur is controlled by the exposure time parameter t_e, which defines the end time of the exposure $t_2 = t_1 + t_e$. The rotation of the camera $\mathbf{R}(t)$ is obtained by solving the quaternion differential equation driven by the angular velocities and computing the corresponding direction cosine matrices [29]. Assuming that the translation is zero (or that the scene is far away), the motion blur can be modelled using a planar homography

$$\mathbf{H}(t) = \mathbf{K}\mathbf{R}(t)\mathbf{K}^{-1}, \tag{1}$$

where $\mathbf{K}$ is the intrinsic camera matrix. Let $\mathbf{x} = (x, y, 1)^\top$

be a projection of the 3D point in homogeneous coordinates. The point-spread-function (PSF) of the blur at the given location can be computed by $\mathbf{x}' = \mathbf{H}(t)\mathbf{x}$.

Since mobile devices are commonly equipped with a rolling shutter camera, each row of pixels is exposed at slightly different time. This is another cause of spatially-variant blur [27]. When computing the PSFs, the start time of the exposure needs to be adjusted based on the y-coordinate of the point $\mathbf{x}$. Let t_r denote the camera read-out time, i.e. the time difference between the first and last row exposure. The exposure of the y:th row starts at $t_1(y) = t_f + t_r \frac{y}{N}$, where t_f corresponds to the starting time of the first row exposure and N is the number of pixel rows. To take this into account, we modify Eq. 1 so that

$$\mathbf{H}(t) = \mathbf{K}\mathbf{R}(t)\mathbf{R}^\top(t_1)\mathbf{K}^{-1}. \quad (2)$$

An example of computed PSFs is shown in Fig. 2. The blurred image is produced by performing a spatially-variant convolution between the sharp image and the blur kernels (PSFs). To speed-up the convolution, we only store and process the nonzero elements of each blur kernel.

4.4. Spatial Misalignment

It is assumed that the blurry image is captured right after the noisy image. Still, the blurry image might be misaligned with respect to the noisy image due to camera or scene motion. Let us consider a horizontal blur kernel with the length of 5 pixels $(1/5) * [11111]$. Normally, the origin would be at the center of the kernel (middle of the exposure). To introduce the effect of spatial misalignment, we set the origin of each PSF kernel to be at the beginning of the exposure. In the previous example, that would correspond to the first or last position of the kernel depending on the motion direction. The effect of misalignment is visualized in Fig. 3. Although we assumed that the images can be taken immediately one after the other, this approach also extends to cases when there is a known gap between the two exposures.

4.5. Realistic Noise

As a final step, we add shot noise to both generated images. The shot noise is considered to be the dominant source of noise in photographs, modeled by a Poisson process. The noise magnitude is varied across different images since it depends on the (ISO) sensitivity setting of the camera. In general, the noise will be significantly more apparent in the short exposure image, and we model this by setting the noise magnitude for the short exposure image larger by a constant factor of 4. Later in Sec. 5.3, the network is fine-tuned with real examples of noisy images. This way the noise characteristics can be learned directly from the data.

Finally, after adding the noise, we ensure that the maximum intensity of the blurry long exposure image does not exceed the maximum brightness value of 1. That is, we clip larger values at 1.

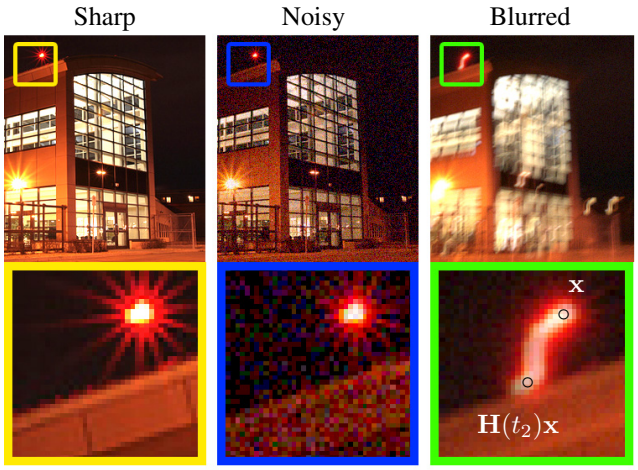

Figure 3. Noisy and blurred images generated from synthetically overexposed sharp image. This will produce realistic light-streaks as the pixel values of the sharp image exceed the range. We also model the misalignment between the noisy and blurred image as they are captured one after the other.

5. Network and Training Details

5.1. Architecture

The network is based on the popular U-Net architecture [24]. This type of network has been successfully used in many image-to-image translation problems [12]. In our case, the input of the network is a pair of blurry and noisy images (stacked). Since the network is fully convolutional, the images can be of arbitrary size. The architecture of the network is shown in Fig. 2. First, the input goes through a series of convolutional and downsampling layers. Once the bottleneck, i.e. the lowest resolution is reached, this process is reversed. The upsampling layers expand the low-resolution image back into a full resolution image. The feature maps from the encoder are concatenated with equally sized feature maps of the decoder. The number of feature maps is shown below the layers in Fig. 2. All convolutional layers use a 3x3 window, except the last layer, which is a 1x1 convolution. Downsampling layers are 2x2 max-pooling operations with a stride of 2.

5.2. Training

The LSD_2 network was trained on 100k images taken from an online image collection [11]. The synthetically corrupted images have resolution of 270 $\times$ 480 pixels. We used the Adam [14] optimizer with the L2 loss function. The learning rate was initially set to 0.00005 and it was halved after every 10th epoch. The network was trained for 50 epochs.

5.3. Fine-tuning

The method is targeted for real-world images that have gone through unknown image processing pipeline of the camera. To this end, we fine-tune the network with real images captured with the NVIDIA Shield tablet, the same device that will be used in testing. This way, the network can learn the noise and color distortion models directly from the data. Examples of real noise are shown in Fig. 4. Notice the relatively coarse appearance of the noise. Our synthetic noise model assumes that the noise is independent for each pixel. This is clearly does not hold because of the camera's internal processing (demosaicing, etc.).

We capture pairs of short and long exposure images while the camera is on a tripod. In this case, the long exposure image is used as the ground truth sharp image. It is also used to generate the blurred image as described in Sec. 4.3. The short exposure image directly corresponds to noisy image. To increase the amount of training samples, we capture several image pairs at once while varying the long exposure between 30 - 330 milliseconds. The ratio of exposure times remains fixed so that the short exposure is always 1/30 of the long exposure. The ISO settings for the long and short exposure images are set to 200 and 800, respectively. The original images are divided to four sub-images to further increase the training data. The network was fine-tuned on 3500 images (480 x 960 pixels) for 30 epochs. The rest of the details are the same as in Sec. 5.2.

6. Experiments

We capture pairs of noisy and blurry images in rapid succession with the NVIDIA Shield tablet. The image acquisition setup is the same as in Sec. 5.3, except this time the camera and/or scene is moving. The resolution of the images is 800 $\times$ 800 pixels (cropped from the original images). For the quantitative comparison, we use synthetically blurred and noisy image pairs taken from the validation set. An example of such pair is shown in Fig. 2.

6.1. Single-Image Approaches

The proposed approach is first compared against the state-of-the-art deblurring and denoising methods DeblurGAN [15] and BM3D [4]. The noise standard deviation parameter of BM3D has been manually tuned to achieve a good overall balance between noise removal and detail preservation.

Fig. 4 show the results on static scenes. The short exposure image (noisy) has been normalized so that its intensity matches the blurry image (for visualization). The most apparent weakness of the BM3D is that the color information is partly lost and cannot be recovered using the noisy image alone. LSD_2 does a good job at extracting the colors from the blurry image. Saturated image regions, such as the light streaks, do not cause problems for LSD_2. There is significantly less noise compared to BM3D, which also tends to over-smooth some of the details. The results of DeblurGAN [15] are unsatisfactory as it fails to remove most of the blur.

Method	PSNR	SSIM
Noisy	16.14	0.51
Blurred	16.88	0.57
DeblurGAN [15]	15.78	0.54
BM3D [4]	23.48	0.79
LSD_2	**25.67**	**0.89**

Table 1. The average peak-signal-to-noise ratio (PSNR) and structural similarity (SSIM) computed for 30 synthetically corrupted image pairs (shown in the supplementary material).

Fig. 5 shows the performance on a dynamic scene. Although LSD_2 has not been trained for this type of situations, the results are surprisingly good. However, fine details such as the bike wheels remain blurry. A quantitative comparison of the methods is presented in Table 1. LSD_2 outperforms the other methods by a fair margin. DeblurGAN [15] generates a "grid-like" pattern over the blurry images, which partly explains the poor results. See the supplementary material for more results.

6.2. Multi-Image Approches

The implementations of Whyte *et al*. [32] or Yuan *et al*. [36] are not publicly available. To compare the methods, we use a pair of blurry and noisy images provided by the authors of [32]. As the exposure and ISO settings are different, we skip the fine-tuning of LSD_2. A comparison against the original result by [32] is shown in Fig. 6. Even though the setup is not ideal for LSD_2, it produces equally good if not better results. The output of [32] shows a little bit of ringing and slightly less details. Note that Whyte *et al*. [32] and Yuan *et al*. [36] perform a separate denoising step and their inputs are registered (manually).

A recent burst deblurring method by Aittala and Durand [1] takes an arbitrary number of blurry images as input. Using their implementation, we compare the methods in Fig. 7. Their result clearly improves as more images are added. Nevertheless, the final result appears less sharp compared to ours, which is obtained with only two images (blurry and noisy). Furthermore, the saturated regions such as the overexposed windows, cannot be recovered using the long exposure images alone. We also tried feeding a pair of noisy and blurry images to [1] but the results were poor. This is not surprising as their method is designed for blurry images only. Similar to [32, 36], the input images need to be registered in advance.

6.3. Exposure Fusion

As described in previous sections, LSD_2 network performs joint denoising and deblurring and outputs a sharp

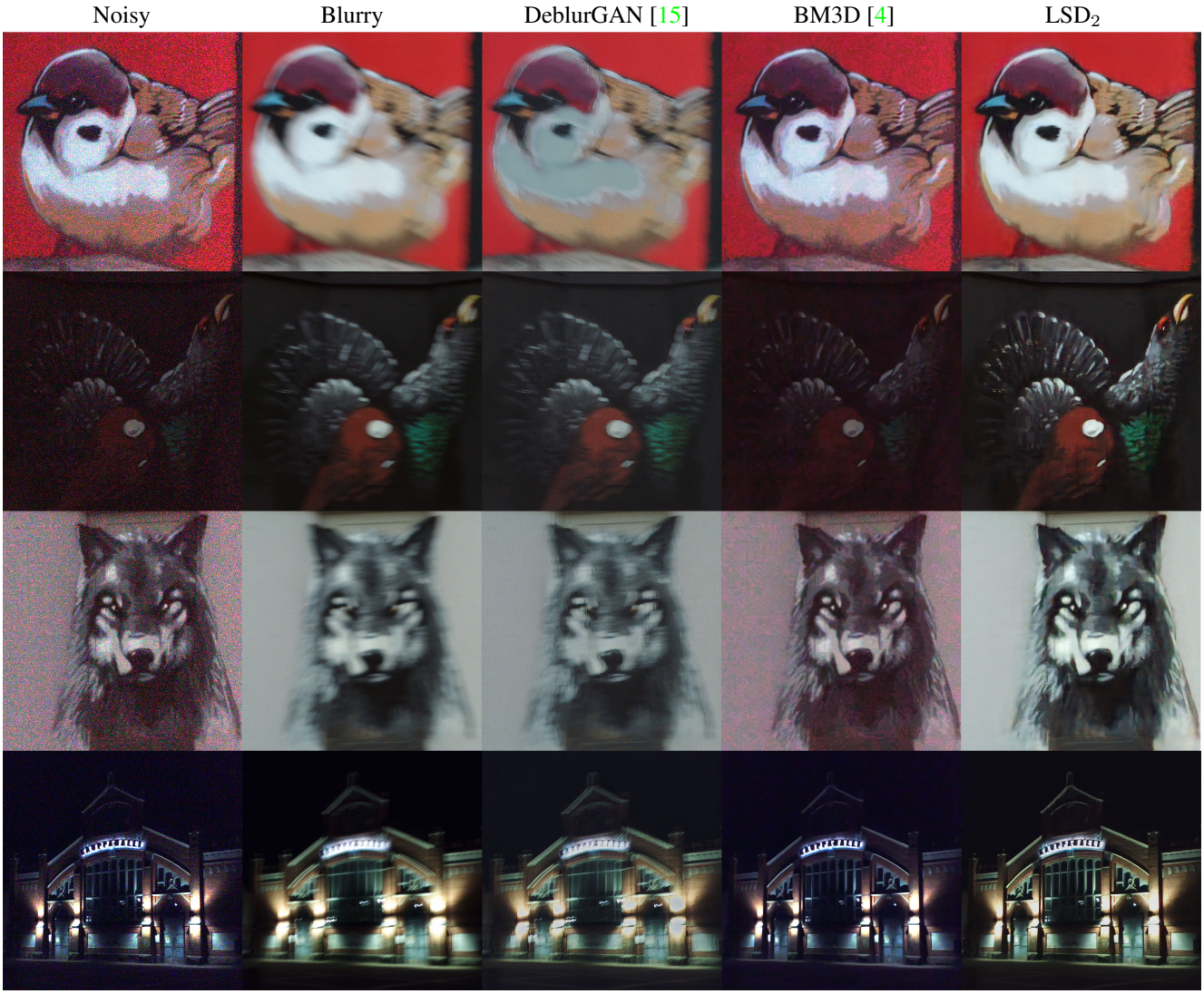

Figure 4. Static scene performance: LSD_2, DeblurGAN [15] and BM3D [4].

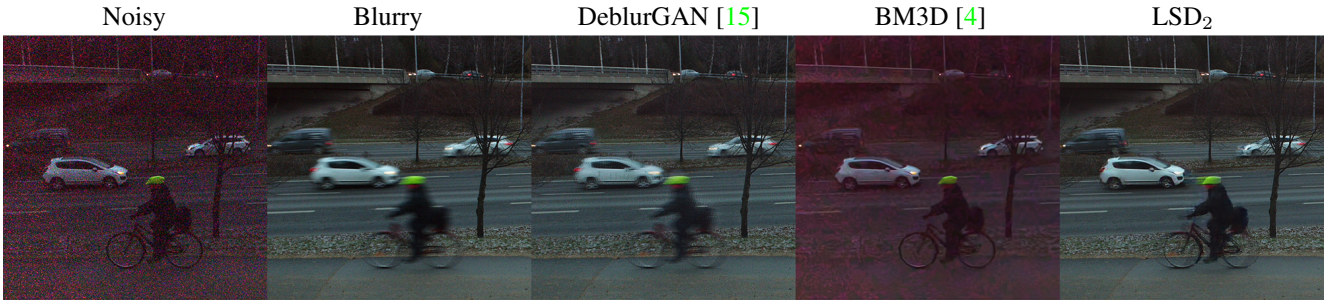

Figure 5. Dynamic scene performance: LSD_2, DeblurGAN [15] and BM3D [4].

version of the long exposure image that is aligned with the short exposure image. Thus, the short exposure image and the output of LSD_2 network would be suitable inputs to exposure fusion methods, such as [23], which assume that the input images are not blurry or misaligned. However, instead of utilizing existing methods, we simply train a second U-net for exposure fusion by using similar synthetic long and short exposure image pairs as described in Sections 4.1 and 4.2. This time the random number s was uniformly sampled from the interval $[1/3, 3]$ and the ground truth target is

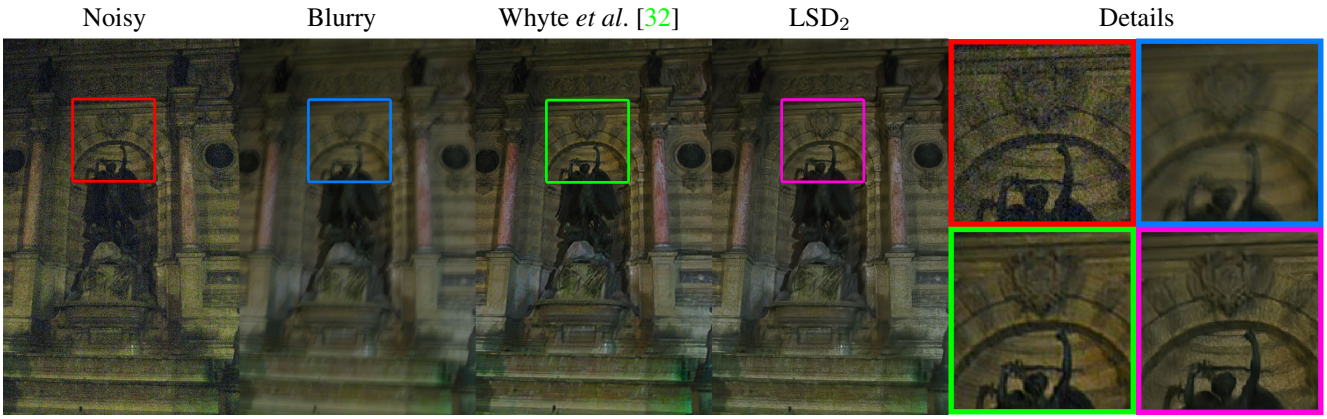

Figure 6. A comparison of LSD$_2$ and Whyte *et al*. [32]. Note that [32] requires manual alignment and a separate denoising step.

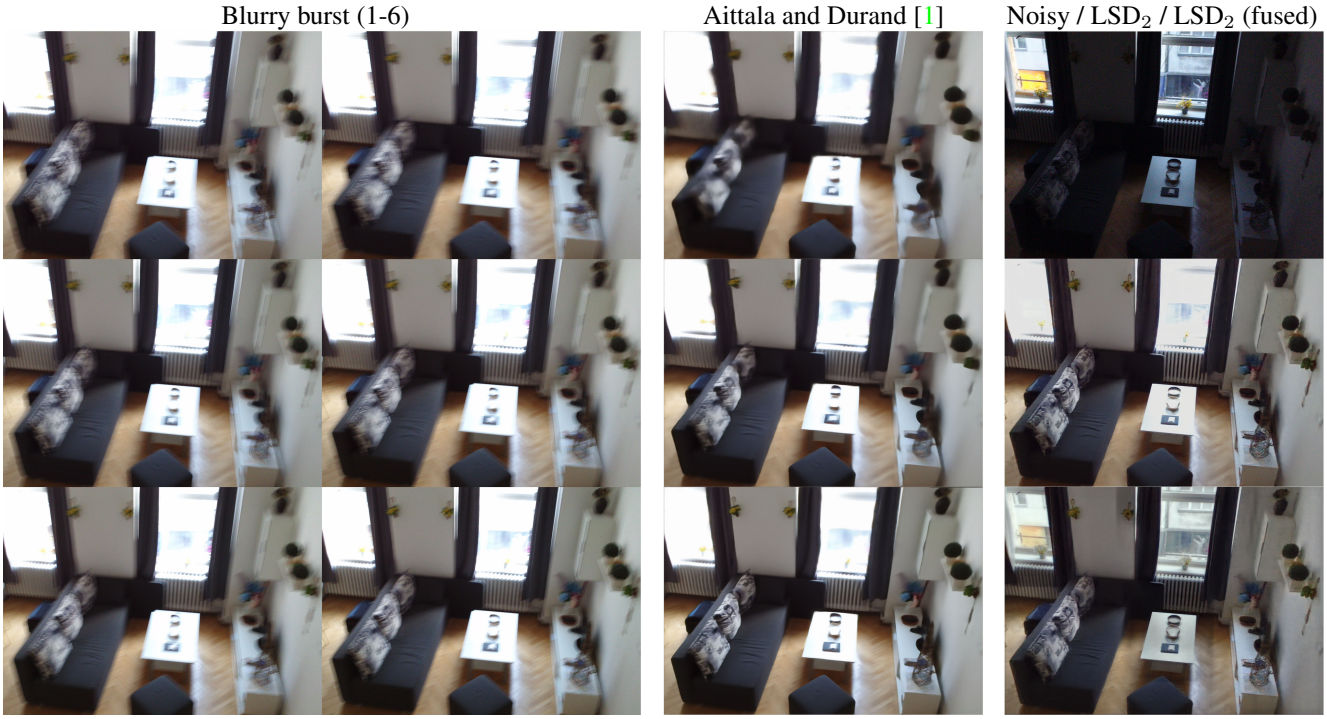

Figure 7. A comparison with Aittala and Durand [1]. A burst of blurry images (left). The results of Aittala and Durand [1] obtained using a growing number of input images: 1, 3 and 6 (middle). The noisy input image and our results without and with the exposure fusion (right).

the original image, which has not been scaled by s and is presumably taken with "good exposure".

In order to demonstrate high-dynamic range imaging, we then process the short exposure image and the output of the LSD$_2$ network with our exposure fusion U-net. The results in Figures 1 and 7 show that we get higher dynamic range and better reproduction of colors and brightness than in either one of the single-exposure input images.

The main purpose of this experiment is to demonstrate the suitability of LSD$_2$ approach for handheld high-dynamic range imaging with smartphones. Since exposure fusion is not the main focus in this paper, a more comprehensive evaluation of different approaches is left for future work.

7. Conclusion

We proposed a CNN-based joint image denoising and deblurring method called LSD$_2$. It recovers a sharp and noise-free image given a pair of short and long exposure images. Its performance exceeds the conventional single-image denoising and deblurring methods on both static and dynamic scenes. Furthermore, LSD$_2$ compares favorably with existing multi-image approaches. Unlike previous methods that utilize pairs of noisy and blurry images, LSD$_2$ does not rely on any existing denoising algorithm. Moreover, it does not expect the input images to be pre-aligned. Finally, we demonstrated that the LSD$_2$ output makes exposure fusion possible even in the presence of motion blur and misalignment.

References

[1] M. Aittala and F. Durand. Burst image deblurring using permutation invariant convolutional neural networks. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 731–747, 2018. 3, 6, 8

[2] H. C. Burger, C. J. Schuler, and S. Harmeling. Image denoising: Can plain neural networks compete with BM3D? In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2392–2399, 2012. 2

[3] Y. Chen and T. Pock. Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6):1256–1272, 2017. 2

[4] 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, 6, 7

[5] M. Delbracio and G. Sapiro. Removing camera shake via weighted fourier burst accumulation. *IEEE Transactions on Image Processing*, 24(11):3293–3307, 2015. 3

[6] M. Elad and M. Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image processing*, 15(12):3736–3745, 2006. 2

[7] D. Gong, J. Yang, L. Liu, Y. Zhang, I. D. Reid, C. Shen, A. van den Hengel, and Q. Shi. From motion blur to motion flow: A deep learning solution for removing heterogeneous motion blur. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3806–3815, 2017. 3

[8] S. Gu, L. Zhang, W. Zuo, and X. 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. 2

[9] S. W. Hasinoff, D. Sharlet, R. Geiss, A. Adams, J. T. Barron, F. Kainz, J. Chen, and M. Levoy. Burst photography for high dynamic range and low-light imaging on mobile cameras. *ACM Transactions on Graphics (TOG)*, 35(6):192, 2016. 3

[10] S. Hee Park and M. Levoy. Gyro-based multi-image deconvolution for removing handshake blur. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3366–3373, 2014. 2, 4

[11] M. J. Huiskes, B. Thomee, and M. S. Lew. New trends and ideas in visual concept detection: the MIR flickr retrieval evaluation initiative. In *Proceedings of the international conference on Multimedia information retrieval*, pages 527–536. ACM, 2010. 5

[12] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. *CVPR*, 2017. 5

[13] V. Jain and S. Seung. Natural image denoising with convolutional networks. In *Advances in Neural Information Processing Systems*, pages 769–776, 2009. 2

[14] D. P. Kingma and L. Ba. J. adam: a method for stochastic optimization. In *International Conference on Learning Representations*, 2015. 5

[15] O. Kupyn, V. Budzan, M. Mykhailych, D. Mishkin, and J. Matas. Deblurgan: Blind motion deblurring using conditional adversarial networks. *ArXiv e-prints*, 2017. 2, 6, 7

[16] J. Lehtinen, J. Munkberg, J. Hasselgren, S. Laine, T. Karras, M. Aittala, and T. Aila. Noise2noise: Learning image restoration without clean data. *arXiv preprint arXiv:1803.04189*, 2018. 2

[17] B. Mildenhall, J. T. Barron, J. Chen, D. Sharlet, R. Ng, and R. Carroll. Burst denoising with kernel prediction networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2502–2510, 2018. 3

[18] J. Mustaniemi, J. Kannala, S. Särkkä, J. Matas, and J. Heikkilä. Inertial-aided motion deblurring with deep networks. *arXiv preprint arXiv:1810.00986*, 2018. 2

[19] S. Nah, T. H. Kim, and K. M. Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017. 2

[20] T. M. Nimisha, A. K. Singh, and A. N. Rajagopalan. Blur-invariant deep learning for blind-deblurring. In *ICCV*, pages 4762–4770, 2017. 2

[21] J. Pan, D. Sun, H. Pfister, and M.-H. Yang. Blind image deblurring using dark channel prior. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1628–1636, 2016. 2

[22] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli. Image denoising using scale mixtures of gaussians in the wavelet domain. *IEEE Transactions on Image Processing*, 12(11):1338–1351, 2003. 3

[23] K. R. Prabhakar, V. S. Srikar, and R. V. Babu. Deepfuse: A deep unsupervised approach for exposure fusion with extreme exposure image pairs. In *IEEE International Conference on Computer Vision (ICCV)*, pages 4724–4732, 2017. 2, 7

[24] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015. 3, 5

[25] O. Šindelář and F. Šroubek. Image deblurring in smartphone devices using built-in inertial measurement sensors. *Journal of Electronic Imaging*, 22(1):011003–011003, 2013. 4

[26] H. Son and S. Lee. Fast non-blind deconvolution via regularized residual networks with long/short skip-connections. In *IEEE International Conference on Computational Photography (ICCP)*, pages 1–10, 2017. 3

[27] S. Su and W. Heidrich. Rolling shutter motion deblurring. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1529–1537, 2015. 5

[28] J. Sun, W. Cao, Z. Xu, and J. Ponce. Learning a convolutional neural network for non-uniform motion blur removal. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 769–777, 2015. 3

[29] D. H. Titterton and J. L. Weston. *Strapdown Inertial Navigation Technology*. The Institution of Electrical Engineers, 2004. 4

[30] D. Ulyanov, A. Vedaldi, and V. Lempitsky. Deep image prior. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018. 2

[31] R. Wang and D. Tao. Training very deep CNNs for general non-blind deconvolution. *IEEE Transactions on Image Processing*, 27(6):2897–2910, 2018. 3

[32] O. Whyte, J. Sivic, A. Zisserman, and J. Ponce. Non-uniform deblurring for shaken images. *International journal of computer vision*, 98(2):168–186, 2012. 2, 3, 6, 8

[33] P. Wieschollek, M. Hirsch, B. Schölkopf, and H. P. Lensch. Learning blind motion deblurring. In *ICCV*, pages 231–240, 2017. 3

[34] P. Wieschollek, B. Schölkopf, H. P. Lensch, and M. Hirsch. End-to-end learning for image burst deblurring. In *Asian Conference on Computer Vision*, pages 35–51. Springer, 2016. 3

[35] Y. Yan, W. Ren, Y. Guo, R. Wang, and X. Cao. Image deblurring via extreme channels prior. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 6978–6986, 2017. 2

[36] L. Yuan, J. Sun, L. Quan, and H.-Y. Shum. Image deblurring with blurred/noisy image pairs. In *ACM Transactions on Graphics (TOG)*. ACM, 2007. 2, 3, 6

[37] 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*, 26(7):3142–3155, 2017. 2

[38] K. Zhang, W. Zuo, S. Gu, and L. Zhang. Learning deep cnn denoiser prior for image restoration. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, 2017. 2