

LSD₂ – Joint Denoising and Deblurring of Short and Long Exposure Images with CNNs

Janne Mustaniemi¹
janne.mustaniemi@oulu.fi

Juho Kannala²
juho.kannala@aalto.fi

Jiri Matas³
matas@cmp.felk.cvut.cz

Simo Särkkä⁴
simo.sarkka@aalto.fi

Janne Heikkilä¹
janne.heikkila@oulu.fi

¹ Center for Machine Vision and
Signal Analysis, University of Oulu,
Finland

² Department of Computer Science,
Aalto University, Finland

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

⁴ Department of Electrical
Engineering and Automation,
Aalto University, Finland

Abstract

The paper addresses the problem of acquiring high-quality photographs with hand-held 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. Unlike existing methods, we take advantage of both images simultaneously and perform a joint denoising and deblurring using a convolutional neural network. A novel approach is introduced to generate realistic short-long exposure image pairs. The method produces good images in extremely challenging conditions and outperforms existing denoising and deblurring methods. It also enables exposure fusion in the presence of motion blur.

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. Besides the problems of noise and motion blur, the camera sensors have limited dynamic range. Details are typically lost either in dark shadows or bright highlights. These problems are most pronounced in smartphones, where the camera and optics need to be small, light-weight and cheap.

A satisfactory compromise between short and long exposure times does not always exist. To get rich colors, good brightness and low noise levels, one should choose long exposure with low sensor sensitivity setting. This will cause motion blur if the camera is moving (shaking), or if there is motion in the scene. A short exposure with high sensitivity setting will produce sharp but noisy images. Examples of such images are shown in Fig. 1. These

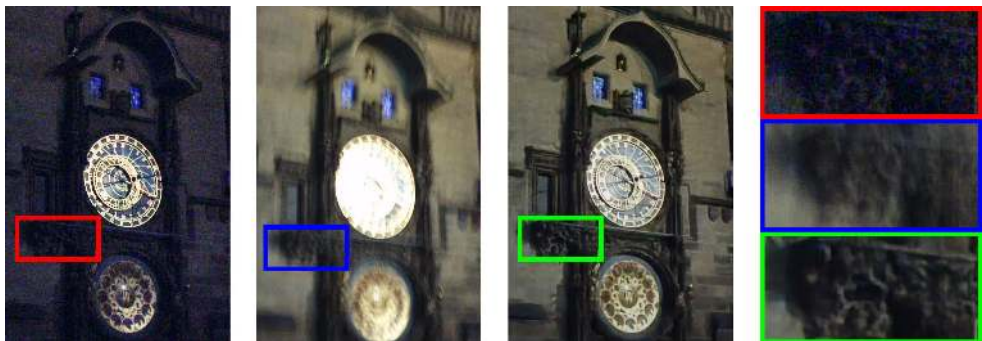


Figure 1: A pair of short and long exposure images captured by a hand-held tablet at night in $7+210=217$ milliseconds. A jointly deblurred and denoised image by the proposed LSD_2 method. The sharp and noise-free LSD_2 output has been tone mapped using an exposure fusion method. See Fig. 4 for dynamic scene results.

issues can be addressed using image denoising and deblurring. However, the conventional methods are limited by the information in a single image.

We propose a novel approach that addresses the aforementioned challenges by taking “the best of both worlds” via computational photography. Current mobile devices can often be programmed to capture bursts of images without extra hardware or notable delay. We capture pairs of short and long exposure images and fuse them into a single high-quality image. The method does not rely on existing denoising algorithms unlike previous methods that utilize short-long exposure image pairs [33, 37].

The proposed method, called LSD_2 ¹ performs joint image denoising and deblurring, exploiting information from both images, adapting their contributions to the conditions at hand. We train a deep convolutional neural network (CNN) that takes a pair of short and long exposure images as input and provides a single image as output. A novel approach is proposed for generating realistic short- and long-exposure images. LSD_2 is shown to outperform existing single-image and multi-image methods. Additionally, we train a second network for exposure fusion. Processing the LSD_2 output with the exposure fusion network improves the colors and brightness compared to a single-exposure smartphone image. The LSD_2 network, training data, and the Android software we developed for acquisition of the back-to-back short and long exposure images will be made public.

2 Related work

Single-image denoising has been addressed using approaches such as sparse representations [7], transform-domain collaborative filtering [5] or nuclear norm minimization [9]. Several deep learning based approaches have been proposed recently [2, 14, 17, 38]. The networks have been trained with pairs of clean and noisy images [2, 14, 38], and without clean targets [17]. Besides the end-to-end deep learning approaches, some utilize either conventional feed-forward networks [39] or recurrent networks [4] as learnable priors for denoising. Randomly initialized networks have been used as priors without pretraining [31]. Some methods

¹ LSD_2 stands for Long-Short Denoising and Deblurring.

utilize raw sensor data [3]. In contrast to our approach, the aforementioned methods focus on single image restoration and do not address multi-image denoising and deblurring.

Single-image deblurring is an ill-posed problem. Various priors have been proposed to regularize the solutions, *e.g.* the dark and bright channel priors [22, 36]. These methods assume spatially invariant blur which limits their practicality. Priors based on deep networks have also been proposed [39]. Some methods first estimate blur kernels and thereafter perform non-blind deconvolution [8, 29]. There are end-to-end approaches that directly produce a deblurred image [16, 20, 21]. Some methods aim to remove deconvolution artifacts [27, 32]. Others utilize additional information like inertial measurements [11, 19]. Despite recent progress, single-image deblurring methods often fail to produce satisfactory results. Unlike our approach, they cannot utilize a sharp but noisy image to guide the deblurring.

Several multi-image denoising [10, 18] and deblurring approaches [1, 6, 34, 35] have been proposed recently. They process a burst of images captured with a constant exposure time. Therefore, they address either denoising or deblurring, but not both problems jointly like we do. As the input images are not as complementary, they cannot get “the best of both worlds” but suffer the drawbacks of either case. Dynamic scenes are difficult to handle, especially when the capture time is long. Images may be severely misaligned and fast-moving objects might disappear from the view. With a constant exposure, the saturated regions can not be easily avoided and high dynamic range imaging is not achieved. On top of that, based on our observations and earlier studies [1, 18], it seems that due to the non-complementary nature of constant exposure, it is necessary to use more input frames than two. This may increase the consumption of memory, power, and processing time.

A similar problem setting as in our work is considered in [33, 37]. These methods utilize short-long exposure image pairs for image deblurring. They first estimate blur kernels for the blurry image and thereafter use the so-called residual deconvolution, proposed by [37], to iteratively estimate the residual image that is to be added to the denoised sharp image. A non-uniform blur model was introduced in [33] to improve the results. We note that both methods use [23] for denoising. A significant drawback of [37] and [33] is that they require a separate photometric and geometric registration stage, where the rotation is estimated manually [37]. Moreover, their model is not applicable to non-static scenes.

3 Joint Denoising and Deblurring

An overview of the proposed LSD_2 method is shown in Fig. 2. The goal is to recover the underlying sharp and noise-free image using a pair of long and short exposure images. A short exposure image is sharp but noisy as it is taken with a high sensitivity setting. A long exposure image is typically blurry due to camera or scene motion. Note that colors of the short exposure image are often distorted w.r.t. the long exposure image as shown in Fig. 1. Furthermore, the images are slightly misaligned even though they are captured immediately one after the other.

3.1 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

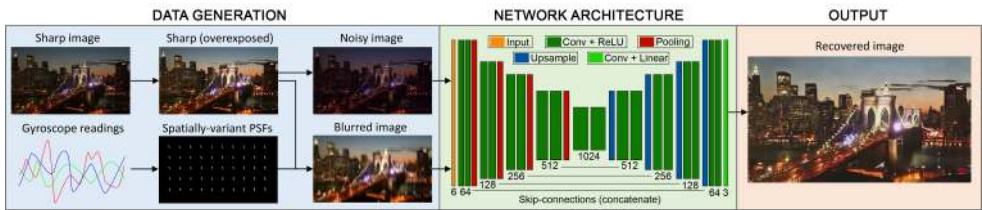


Figure 2: Overview of the LSD₂ joint denoising and deblurring method. To train the network, we generate pairs of short and long exposure images with realistic motion blur, spatial misalignment, image noise, color distortion and saturated regions.

can generate unlimited amount of training data with realistic blur while covering a wide range of different scene types. The LSD₂ network operates with images having intensity range $[0, 1]$ and hence we first scale the original RGB values to that range. We first invert the gamma correction of the input images (assuming $\gamma = 2.2$). Once the images have been generated, the gamma is re-applied.

3.1.1 Synthesis of Long Exposure Images

We take a regular high-quality RGB image I from Internet as the starting point of our simulation. At test time, the long exposure image should be slightly overexposed in order to enable high dynamic range and ensure sufficient illumination of darkest scene regions. Hence, we 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 $[1, 3]$. Then, by clipping the maximum intensity to value of 1, we get the sharp long exposure image, which will be the ground truth target. That is, we train the network to predict an output with similar exposure as the long exposure image.

3.1.2 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 $(aI + b)$ 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. The real image pairs are captured so that the short exposure time is always 1/30 of the long exposure time. Based on our observations, a constant exposure time ratio improves the performance. The absolute exposure time is allowed to vary based on auto-exposure algorithm.

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 is obtained by adding noise as described in Sec. 3.1.5.

3.1.3 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 [11, 26], 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 idea is to simulate a real life imaging situation with a shaking hand.

The starting time of the synthetic image exposure t_1 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 [30]. Assuming that the translation is zero (or that the scene is far away), the motion blur can be modelled using a planar homography

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

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 [28]. When computing the PSFs, the start time of the exposure is adjusted based on the y -coordinate of the point \mathbf{x} . Let t_r denote the camera readout 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

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

The blurred image is produced by performing a spatially-variant convolution between the sharp image and PSFs.

3.1.4 Spatial Misalignment

It is assumed that the blurry image is captured right after the noisy image. Nevertheless, the blurry image might be misaligned with respect to the noisy image due to camera or scene motion. Normally, the origin of the PSF 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. This approach also extends to cases when there is a known gap between the two exposures.

3.1.5 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 varies across different images since it depends on the sensitivity setting (ISO) 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. After this, 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. Later in Sec. 3.2.1, the network is fine-tuned with real examples of noisy images. This way the noise characteristics can be learned directly from the data.

3.2 Network and Training Details

The network is based on the popular U-Net architecture [25]. This type of network has been successfully used in many image-to-image translation problems [13]. It was chosen because of its simplicity and because it produced excellent results for this problem. The architecture of the network is shown in Fig. 2. The input of the network is a pair of blurry and noisy images (stacked). The images can be of arbitrary size since the network is fully convolutional. The number of feature maps is shown below the layers. 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.

The LSD₂ network was trained on 100k images taken from an online image collection [12]. The synthetically corrupted images have resolution of 270×480 pixels. We used the Adam [15] 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.

3.2.1 Fine-tuning

The method is targeted for images that have gone through an unknown image processing pipeline of the camera. To this end, we fine-tune the network with real images captured with the NVIDIA Shield tablet. This way, the network can learn the noise and color distortion models directly from the data. Real noise has a relatively coarse appearance as can be seen in Fig. 3. Our synthetic noise model assumes that the noise is independent for each pixel. This 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. The ratio of exposure times is fixed to 1/30. The ISO settings for the long and short exposure images are set to 200 and 800, respectively. The long exposure image is used as the ground truth sharp image and the short exposure image directly corresponds to the noisy image. The blurred image is generated from the sharp image as described in Sec. 3.1.3.

To increase the amount of training samples, we capture several image pairs at once while varying the long exposure between 30 - 330 milliseconds. Moreover, the original images are divided to four sub-images of size 480×960 pixels. The network was fine-tuned on 3500 images for 30 epochs. The rest of the details are the same as in the previous section.

4 Experiments

We capture pairs of noisy and blurry images in rapid succession with the NVIDIA Shield tablet and the Google Pixel 3 smartphone. The image acquisition setup is the same as in Sec. 3.2.1, except this time the camera and/or scene is moving. The resolution of the images is 800×800 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.

4.1 Single-Image Approaches

A comparison between LSD₂ and the state-of-the-art denoising methods BM3D [5] and FDNnCNN [38] is shown in Figs 3 and 4. The short exposure image (noisy) has been normalized so that its intensity matches the blurry image for visualization. The most apparent weakness of BM3D and FDNnCNN is that the color information is partly lost and cannot be recovered using a noisy image alone. Both methods tend to over-smooth some of the details even though their noise standard deviation parameters have been manually tuned to achieve a good overall balance between noise removal and detail preservation. A quantitative comparison in Table 1 shows that LSD₂ outperforms the other methods by a fair margin.

Fig. 4 show a comparison against the state-of-the-art deblurring method [16]. The results of DeblurGAN are unsatisfactory as it fails to remove most of the blur. Note that saturated image regions, such as light streaks, do not cause problems for LSD₂. Furthermore, LSD₂ performs surprisingly well on a dynamic scene even though it has not been trained for this type of situations. However, fine details such as the bike wheels remain blurry. See the supplementary material for more results.

We also tried retraining FDNnCNN [38] and DeblurGAN [16] using our dataset but with poor results. The FDNnCNN network has to learn to adjust the colors and brightness in addition to denoising. This seems to be too challenging task at least for the given network. Retraining did not improve DeblurGAN likely because the blurred and sharp images are misaligned in our dataset.

The importance of using real data for fine-tuning is demonstrated in Fig. 3. The fine-tuning clearly helps as the output is significantly less noisy and the colors are better. Furthermore, the fine-tuning does not make the network device specific.

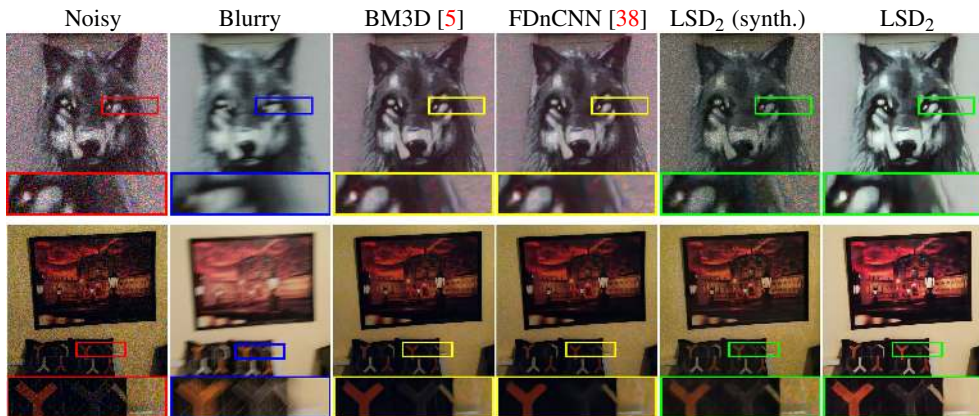


Figure 3: A comparison of LSD₂ and single-image denoising methods BM3D [5] and FDNnCNN [38]. The second column from the right shows the results without fine-tuning (synthetic data only). Note that the LSD₂ network was fine-tuned using NVIDIA Shield data but the input images on the second row were captured with Google Pixel 3.

4.2 Multi-Image Approches

Fig. 5 shows a comparison against Yuan *et al.* [37]. Input images were copied from the original paper (PDF file) as their inputs are not available. The LSD₂ output is significantly

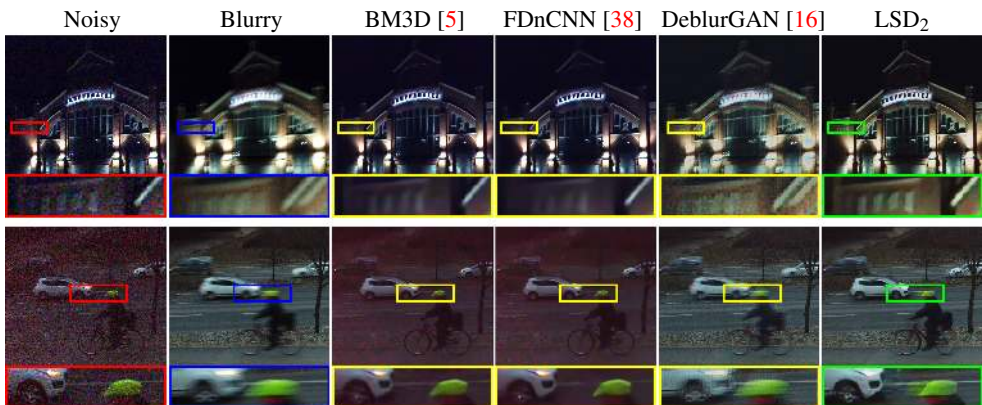


Figure 4: A low light performance in the presence of saturated pixels (top) and a dynamic scene performance (bottom).

	Noisy	Blurred	DeblurGAN [16]	BM3D [5]	FDnCNN [38]	LSD ₂
PSNR	16.43	16.88	15.78	23.48	23.83	25.67
SSIM	0.51	0.57	0.54	0.79	0.81	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). For fairness, the outputs of [5] and [38] have been adjusted so that the colors match the blurred images before computing the scores.

less noisy. The performance of [37] depends heavily on the separate denoising step. Furthermore, it produces major artifacts when inputs are misaligned (supplementary material). The parameters of [37] were kept default. Note that LSD₂ does not have any tunable parameters.

The implementation of Whyte *et al.* [33] is not publicly available. For comparison, we use the input images provided by the authors of [33]. Fig. 5 shows a comparison against the original result by [33]. LSD₂ produces equally good if not better results. The output of [33] shows a little bit of ringing and slightly less details. Keep in mind that [33] and [37] utilize an existing denoising algorithm. Moreover, the inputs need to be manually registered.

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. 6.

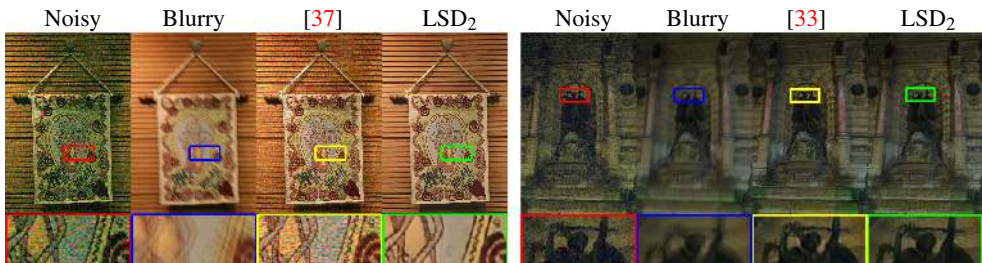


Figure 5: A comparison with Yuan *et al.* [37] and Whyte *et al.* [33]. They both require a separate denoising step and manual alignment.



Figure 6: A comparison with Aittala and Durand [1]. A burst of 6 blurry images was given to [1] as input (supplementary material).

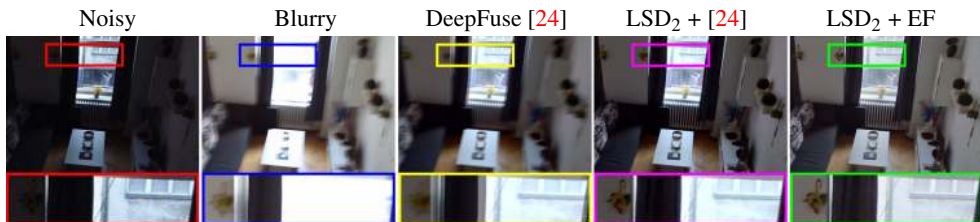


Figure 7: Exposure fusion without and with LSD₂. DeepFuse [24]: using a pair of noisy and blurry images as input. LSD₂ + DeepFuse [24]: the blurry image is replaced with the LSD₂ output. LSD₂ + EF: using our exposure fusion method instead of [24].

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 [33, 37], the input images need to be registered in advance.

4.3 Exposure Fusion

LSD₂ network produces a sharp version of the long exposure image that is aligned with the short exposure image. Thus, the short exposure image and the LSD₂ output would be suitable inputs for an exposure fusion method, such as DeepFuse [24], which assume that the input images are not blurry or misaligned. Fig. 7 shows the result of DeepFuse when using a pair of noisy and blurry images as input. The results are significantly improved when DeepFuse is used together with LSD₂.

We also trained a second network for exposure fusion since DeepFuse [24] does not consider the noise and color distortions of the short exposure image. Details of the network architecture and training are given in the supplementary material. The training was done using similar synthetic long and short exposure image pairs as described in Sections 3.1.1 and 3.1.2. This time the random number s was uniformly sampled from the interval $[1/3, 3]$ and the ground truth target is the original image, which has not been scaled by s and is presumably taken with "good exposure". This type of approach differs from existing methods, which often use hand-crafted features and assume that ground truth targets are not available.

Figs 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. Notice also the lack

of details in the dark areas of the DeepFuse output (see e.g. the curtains). The main purpose of this experiment is to demonstrate the suitability of LSD₂ approach for handheld high-dynamic range imaging with smartphones. A more comprehensive evaluation of different exposure fusion techniques is left for future work.

5 Conclusion

We proposed a CNN-based joint image denoising and deblurring method called LSD₂. 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, LSD₂ compares favorably with existing multi-image approaches. Unlike previous methods that utilize pairs of noisy and blurry images, LSD₂ 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 LSD₂ output makes exposure fusion possible even in the presence of motion blur and misalignment.

Acknowledgement

The authors would like to thank Business Finland for the financial support of this research project (grant no. 1848/31/2015). J. Matas was supported by project CZ.02.1.01/0.0/0.0/16019/0000765 Research Center for Informatics.

References

- [1] Miika Aittala and Frédo Durand. Burst image deblurring using permutation invariant convolutional neural networks. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 731–747, 2018.
- [2] Harold C Burger, Christian J Schuler, and Stefan Harmeling. Image denoising: Can plain neural networks compete with BM3D? In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2392–2399, 2012.
- [3] Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. Learning to see in the dark. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3291–3300, 2018.
- [4] Yunjin Chen and Thomas 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.
- [5] 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.
- [6] Mauricio Delbracio and Guillermo Sapiro. Removing camera shake via weighted fourier burst accumulation. *IEEE Transactions on Image Processing*, 24(11):3293–3307, 2015.

- [7] Michael Elad and Michal Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image processing*, 15(12): 3736–3745, 2006.
- [8] Dong Gong, Jie Yang, Lingqiao Liu, Yanning Zhang, Ian D. Reid, Chunhua Shen, Anton van den Hengel, and Qinfeng 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.
- [9] 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.
- [10] Samuel W Hasinoff, Dillon Sharlet, Ryan Geiss, Andrew Adams, Jonathan T Barron, Florian Kainz, Jiawen Chen, and Marc Levoy. Burst photography for high dynamic range and low-light imaging on mobile cameras. *ACM Transactions on Graphics (TOG)*, 35(6):192, 2016.
- [11] Sung Hee Park and Marc 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.
- [12] Mark J Huiskes, Bart Thomee, and Michael 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.
- [13] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. *CVPR*, 2017.
- [14] Viren Jain and Sebastian Seung. Natural image denoising with convolutional networks. In *Advances in Neural Information Processing Systems*, pages 769–776, 2009.
- [15] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [16] Orest Kupyn, Volodymyr Budzan, Mykola Mykhailych, Dmytro Mishkin, and Jiri Matas. DeblurGAN: Blind motion deblurring using conditional adversarial networks. *ArXiv e-prints*, 2017.
- [17] Jaakko Lehtinen, Jacob Munkberg, Jon Hasselgren, Samuli Laine, Tero Karras, Miika Aittala, and Timo Aila. Noise2Noise: Learning image restoration without clean data. *arXiv preprint arXiv:1803.04189*, 2018.
- [18] Ben Mildenhall, Jonathan T Barron, Jiawen Chen, Dillon Sharlet, Ren Ng, and Robert Carroll. Burst denoising with kernel prediction networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2502–2510, 2018.
- [19] Janne Mustaniemi, Juho Kannala, Simo Särkkä, Jiri Matas, and Janne Heikkilä. Inertial-aided motion deblurring with deep networks. *arXiv preprint arXiv:1810.00986*, 2018.

- [20] Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.
- [21] Thekke Madam Nimisha, Akash Kumar Singh, and Ambasamudram N Rajagopalan. Blur-invariant deep learning for blind-deblurring. In *ICCV*, pages 4762–4770, 2017.
- [22] Jinshan Pan, Deqing Sun, Hanspeter Pfister, and Ming-Hsuan Yang. Blind image deblurring using dark channel prior. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1628–1636, 2016.
- [23] Javier Portilla, Vasily Strela, Martin J. Wainwright, and Eero P. Simoncelli. Image denoising using scale mixtures of gaussians in the wavelet domain. *IEEE Transactions on Image Processing*, 12(11):1338–1351, 2003.
- [24] K Ram Prabhakar, V Sai Srikar, and R Venkatesh 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.
- [25] Olaf Ronneberger, Philipp Fischer, and Thomas 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.
- [26] Ondřej Šindelář and Filip Šroubek. Image deblurring in smartphone devices using built-in inertial measurement sensors. *Journal of Electronic Imaging*, 22(1):011003–011003, 2013.
- [27] Hyeongseok Son and Seungyong 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.
- [28] Shuochen Su and Wolfgang Heidrich. Rolling shutter motion deblurring. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1529–1537, 2015.
- [29] Jian Sun, Wenfei Cao, Zongben Xu, and Jean 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.
- [30] D. H. Titterton and J. L. Weston. *Strapdown Inertial Navigation Technology*. The Institution of Electrical Engineers, 2004.
- [31] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018.
- [32] Ruxin Wang and Dacheng Tao. Training very deep CNNs for general non-blind deconvolution. *IEEE Transactions on Image Processing*, 27(6):2897–2910, 2018.
- [33] Oliver Whyte, Josef Sivic, Andrew Zisserman, and Jean Ponce. Non-uniform deblurring for shaken images. *International journal of computer vision*, 98(2):168–186, 2012.

- [34] Patrick Wieschollek, Bernhard Schölkopf, Hendrik PA Lensch, and Michael Hirsch. End-to-end learning for image burst deblurring. In *Asian Conference on Computer Vision*, pages 35–51. Springer, 2016.
- [35] Patrick Wieschollek, Michael Hirsch, Bernhard Schölkopf, and Hendrik PA Lensch. Learning blind motion deblurring. In *ICCV*, pages 231–240, 2017.
- [36] Yanyang Yan, Wenqi Ren, Yuanfang Guo, Rui Wang, and Xiaochun Cao. Image deblurring via extreme channels prior. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 6978–6986, 2017.
- [37] Lu Yuan, Jian Sun, Long Quan, and Heung-Yeung Shum. Image deblurring with blurred/noisy image pairs. In *ACM Transactions on Graphics (TOG)*. ACM, 2007.
- [38] 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.
- [39] Kai Zhang, Wangmeng Zuo, Shuhang Gu, and Lei Zhang. Learning deep CNN denoiser prior for image restoration. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, 2017.