Supplementary Materials for "Self-supervised Blind Motion Deblurring with Deep Expectation Maximization"

Ji Li, Weixi Wang, Yuesong Nan, and Hui Ji
Department of Mathematics, National University of Singapore, 119076, Singapore
matliji@nus.edu.sg, wangweixi@u.nus.edu, nanyuesong@u.nus.edu, matjh@nus.edu.sg

The content of the supplementary file is listed as follows.

- Section 1. Introduction to Monte Carlo (MC) sampling via Langevin dynamics (LD)
- Section 2. Computational cost on processing blurred images
- Section 3. Visual comparison on Köhler *et al*.'s dataset
- Section 4. Visual comparison on the non-uniform dataset of Lai *et al*. [1]
- Section 5. Studies on model hyper-parameters
- Section 6. Experiments of uniform motion deblurring on Levin's dataset
- Section 7. Demonstration on processing the images of dynamic scenes from GOPRO dataset

1. Introduction to Monte Carlo (MC) sampling via Langevin dynamics (LD)

In the section, we give a brief introduction to MC sampling via LD, a Markov chain Monte Carlo (MCMC) sampling method, which is used in our work for sampling the network weights. More details can be found in most textbooks on MC sampling. LD is about sampling a given distribution $\pi(\theta) \propto \exp(-L(\theta))$. Suppose that $L(\theta) \in C^2(\mathbb{R}^n)$, we consider the following Langevin dynamics, which is governed by the continuous-time stochastic differential equation (SDE)

$$d\theta_t = -\nabla L(\theta_t)dt + \sqrt{2}dW_t, \tag{1}$$

where W_t is a Brownian motion and the equality $\nabla \log \pi(\theta) = -\nabla L(\theta)$ is used. By the dynamics, θ_t is a random variable for each time t, which is associated with a distribution $p_t(\theta)$. Let $t \to \infty$ and assume that $p_t(\theta) \to p_\infty(\theta)$, the limiting distribution $p_\infty(\theta)$ is defined as the *stationary distribution* of the SDE (1). The Fokker-Planck equation [2] is the tool to investigate the evolution of the distribution $p_t(\theta)$. The evolution dynamics of $p_t(\theta)$ is given by

$$\partial_t p_t(\theta) = \frac{\partial}{\partial \theta}[\nabla L(\theta) p_t(\theta)] + \frac{\partial^2}{\partial \theta^2}[p_t(\theta)]. \tag{2}$$

Let $t \to \infty$, then

$$\frac{\partial}{\partial \theta}[\nabla L(\theta) p_\infty(\theta)] + \frac{\partial^2}{\partial \theta^2}[p_\infty(\theta)] = 0, \tag{3}$$

holds, where $\partial_t p_t(\theta) \to 0$ when $t \to \infty$ is used. By solving the equation (3), we have the stationary distribution $p_\infty(\theta) \propto \exp(-L(\theta))$.

The discretization of the SDE (1) gives a Markov chain Monte Carlo (MCMC) sampling method:

$$\theta_{k+1} = \theta_k - \gamma_k \nabla L(\theta_k) + \sqrt{2\gamma_k}\boldsymbol{z}, \boldsymbol{z} \sim \mathcal{N}(0, 1). \tag{4}$$

The above scheme is closely related to the stochastic version of gradient descent method where the stochastic behavior comes from the injected random noise. Thus, the resultant sampling method is called **stochastic gradient Langevin dynamics** (SGLD) sampling method. The convergence of the discretization scheme (4) to the continuous-SDE (1) requires that the stepsize $\{\gamma_k\}$ satisfies the Robbins-Monro condition [3]. When k is large enough, the iterative sequence $\{\theta_k\}$ from (4) can be regarded as the samples from the stationary distribution $p_\infty(\theta) \propto \exp(-L(\theta))$.

2. Computational cost on processing blurred images

See Table 1 for the comparison of computational efficiency when processing a color 800×800 blurred image in a workstation with a single NVIDIA TITAN RTX GPU. It is noted that for non-learning iterative methods, *e.g.* [4–6], the time cost is calculated on CPU, as no GPU version of their codes is available online. It can be seen that the proposed self-supervised method has comparable time cost when running on GPU, in comparison to existing iterative methods for non-uniform blind motion deblurring running on CPU.

The only available self-supervised deep learning method for blind deblurring is Ren *et al.* [7], which is only applicable to uniform deblurring. Our method is also applicable to uniform motion deblurring by degrading the SVOLA model to uniform blurring model with $P = 1$ in (2). Under the same hardware, for a color 800×800 blurred image, the blur kernel size is set to 75×75, the comparison of time cost of two methods is: 844 (ours) vs 1357 ([7]). It can be seen that ours is faster than the existing method [7].

It is noted that there is no concept of "pre-trained" model for an self-supervised method, as it directly trains a NN when processing a blurred image. While such a scheme avoids the expensive time cost of training, the lack of a pre-trained model makes it more suitable for processing a small to modest size of dataset.

Table 1. Comparison of time cost when deblurring a color 800×800 image using nonuniform deblurring model.

	Non-learning methods (on CPU)			Self-supervised (on GPU)
Methods	Hirsch *et al.* [4]	Xu *et al.* [5]	Whyte *et al.* [6]	Ours
Time (s)	1567	1128	1002	1082

3. Visual comparison on Köhler *et al.*'s dataset

In this section, we visualize the results from Köhler *et al.*'s dataset [8]. See Figure 1–2 for visual inspection. It can be seen that our methods recover more sharp details compared with the existing non-uniform deblurring methods, ranging from modest to severe non-uniform blurring degree. The advantage of our methods over others on visual quality is consistent with the advantage of our method in quantitative comparison reported in Table 1 in the main manuscript.

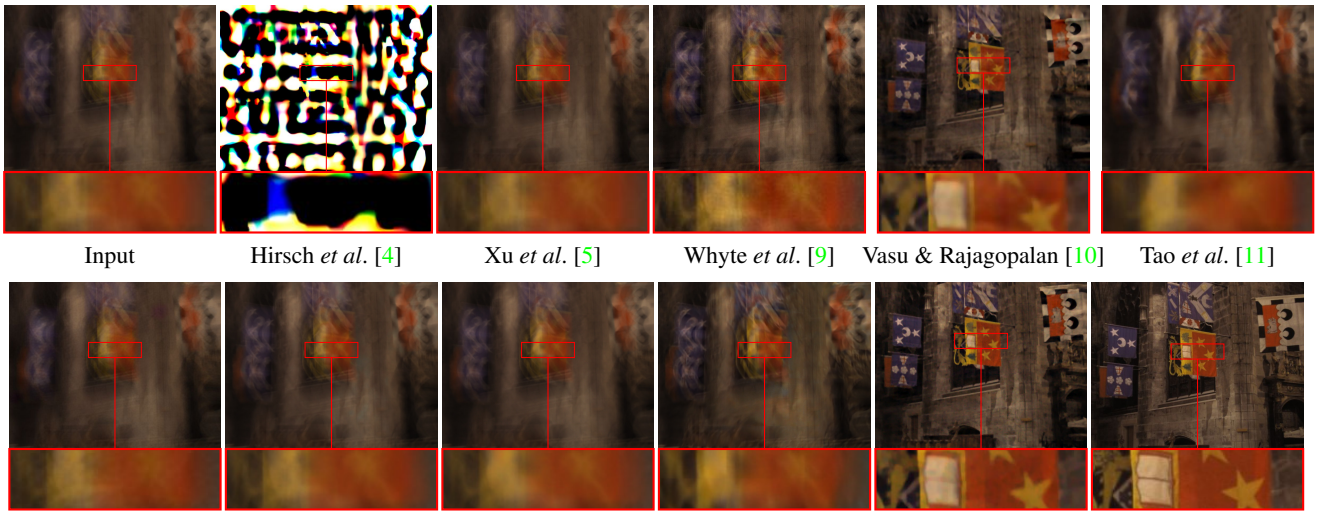


Figure 1. Visual comparison of the deblurred result of "Church" from the dataset of Köhler *et al.* [8]. Zoom-in for better inspection.

4. Visual comparison on the non-uniform dataset of Lai *et al.* [1]

In this section, we showed the results on Lai *et al.*'s [1] non-uniform dataset for visual inspection. See Figure 3–5 for the visual comparison of some results. It can be seen that our recovered results are consistently better than that from the

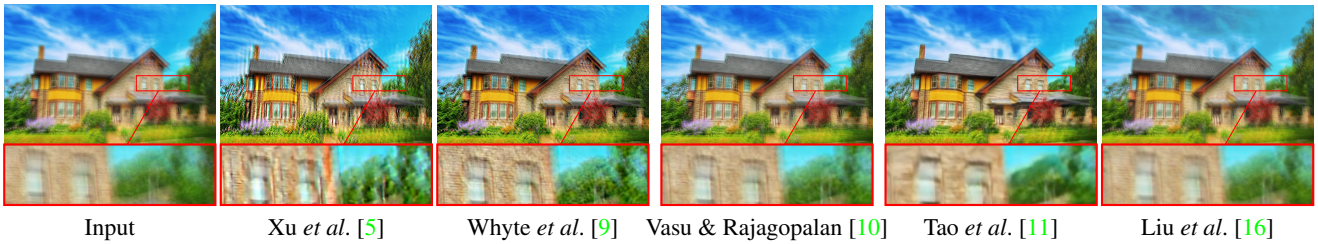
Input | Hirsch *et al*. [4] | Xu *et al*. [5] | Whyte *et al*. [9] | Vasu & Rajagopalan [10] | Tao *et al*. [11]

Kupyn *et al*. [12] | Zamir *et al*. [13] | Cho *et al*. [14] | Li *et al*. [15] | Ours | GroundTruth

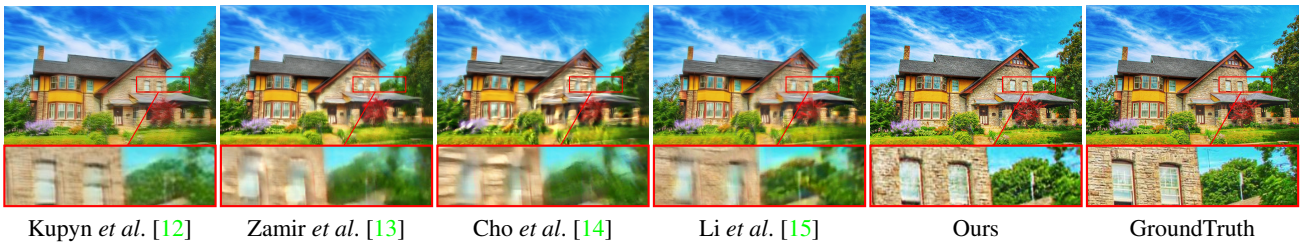
Figure 2. Visual comparison of the deblurred result of "Roof" from the dataset of Köhler *et al*. [8]. Zoom-in for better inspection.

compared methods in terms of visual quality, which is consistent with the quantitative evaluation shown in Table 2 in the main manuscript.

For the real-world dataset, we use the proposed method to solve the deblurring problem with the assumption of spatially-variant blurring. See 6-10 for the results on the images from the real-world dataset Lai *et al*. [1] without truth images available. It can be seen that our method consistently recovers more details with fewer artifacts compared with the existing methods, showing that our method is very competitive when being used for processing non-uniform-blurred images.



Input | Xu *et al*. [5] | Whyte *et al*. [9] | Vasu & Rajagopalan [10] | Tao *et al*. [11] | Liu *et al*. [16]



Kupyn *et al*. [12] | Zamir *et al*. [13] | Cho *et al*. [14] | Li *et al*. [15] | Ours | GroundTruth

Figure 3. Visual comparison of the deblurred results of "Manmade04" from non-uniform dataset of Lai *et al*. [1]. Zoom-in for better inspection.

5. Studies on model hyper-parameters

Recall that the SVOLA model we adopted for modeling non-uniform blurring has a hyper-parameter P:

$$\boldsymbol{g} = \boldsymbol{K} \circ \boldsymbol{f} + \boldsymbol{n} = \sum_{i=1}^{P} \boldsymbol{k}_i \otimes (\boldsymbol{w}(\cdot - c_i) \odot \mathcal{P}_i \boldsymbol{f}) + \boldsymbol{n}. \tag{5}$$

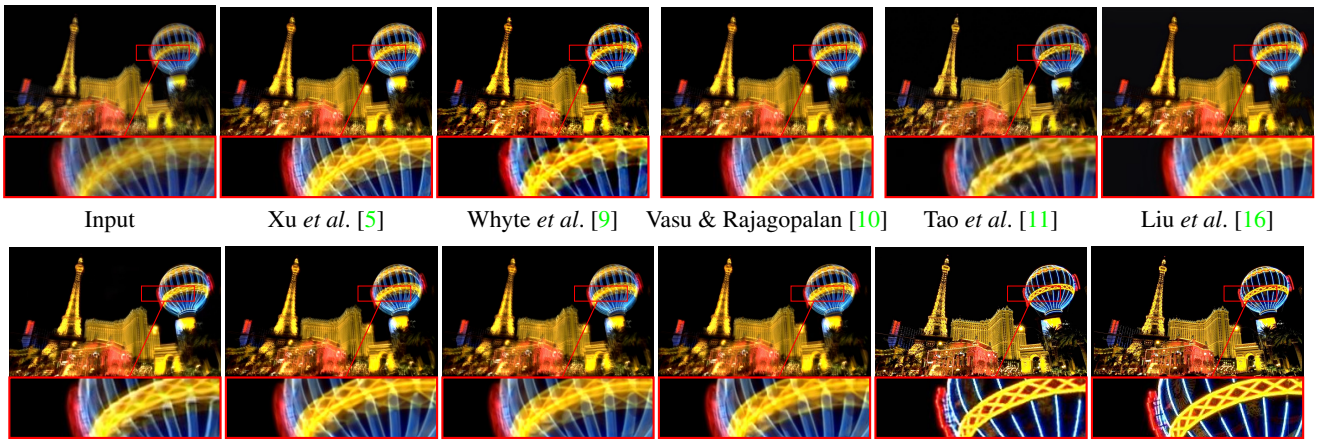


Figure 4. Visual comparison of the deblurred results of "Saturated01" from non-uniform dataset of Lai *et al*. [1]. Zoom-in for better inspection.

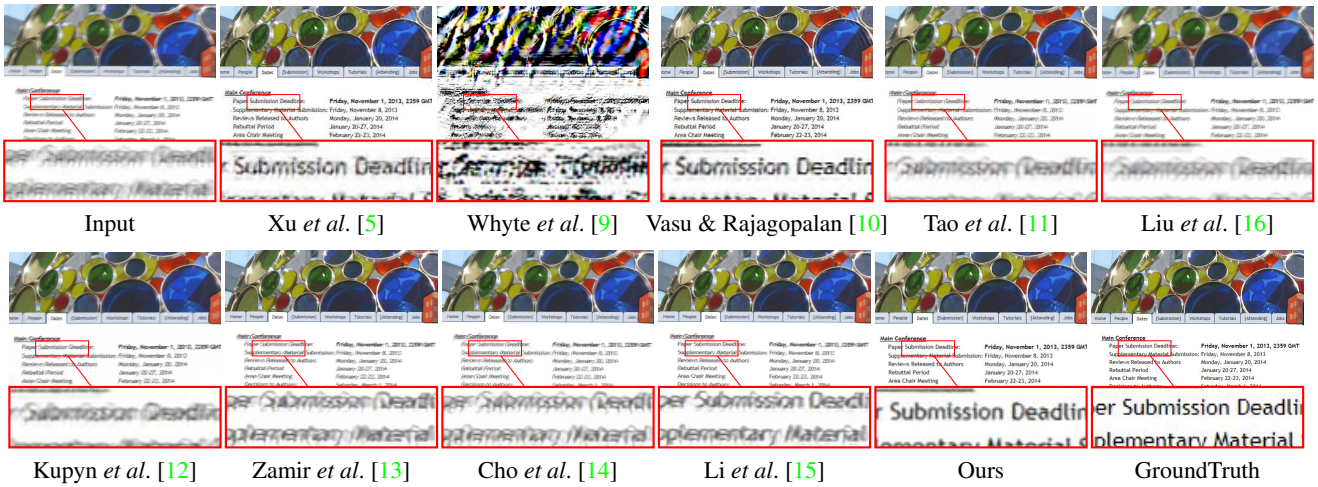
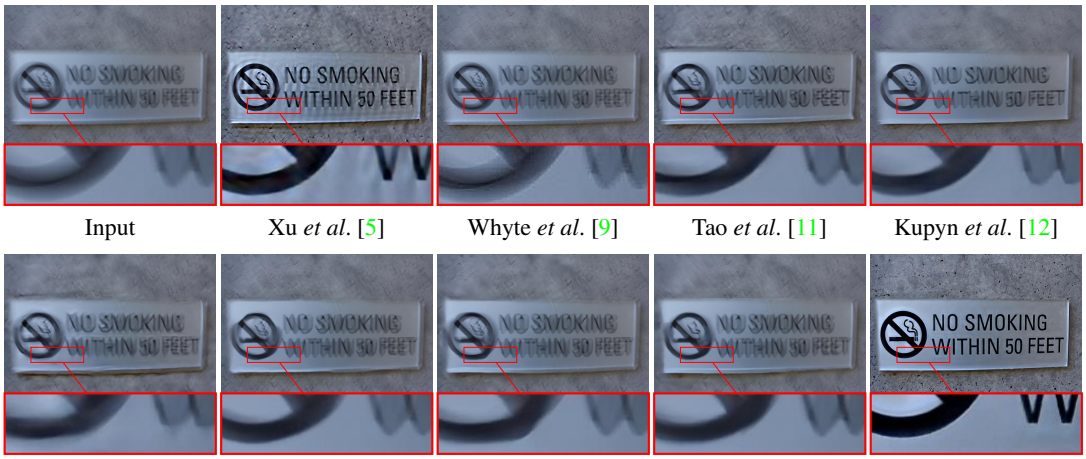


Figure 5. Visual comparison of the deblurred results of "Text02" from non-uniform dataset of Lai *et al*. [1]. Zoom-in for better inspection.

Such a hyper-parameter determines how fragmented the non-uniform blurring is. A larger value of P leads to a more accurate model of non-uniform blurring, but makes the problem harder to solve. In practice, there is a trade-off between model accuracy and computational feasibility. See Figure 11 for an illustration of how the visual quality of a non-uniform blurred image will be impacted by the choice of different values of P. It can be seen that larger P leads to a more accurate model and more sharp recovery in most regions, but will fail in certain regions. A smaller P leads to a more consistent recovery without very bad results in any region, but overall the result appears to remain a little blurry.

6. Experiments of uniform motion deblurring on Levin's dataset

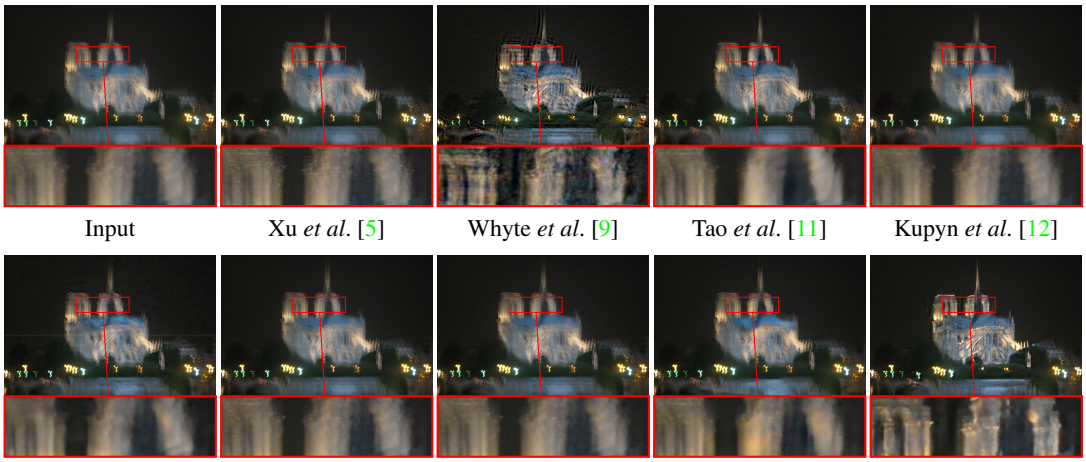
The performance of our method with simplification is evaluated on the synthesized uniform blurring dataset in [1] in the paper. We include the quantitative comparison of the performance on Levin *et al*. [18] in the supplementary file, see Table 2. It shows that the supervised learning methods perform poorly for the unseen blur effects. And our proposed self-supervised method is the best performer for the uniform motion deblurring task, our method outperforms the existing work [7]. For visual inspection, see Figure 12-13 for visual inspection of the proposed method and other compared methods on the images from the uniform dataset Lai *et al*. [1]. It can be seen that our method consistently recovers more details with fewer artifacts compared with the existing methods, showing that our method is very competitive when being used for processing uniform-



Input | Xu *et al*. [5] | Whyte *et al*. [9] | Tao *et al*. [11] | Kupyn *et al*. [12]

Zhang *et al*. [17] | Zamir *et al*. [13] | Cho *et al*. [14] | Li *et al*. [15] | Ours

Figure 6. Visual comparison of the deblurred result of "Text12" from Lai's [1]. Zoom-in for better inspection.



Input | Xu *et al*. [5] | Whyte *et al*. [9] | Tao *et al*. [11] | Kupyn *et al*. [12]

Zhang *et al*. [17] | Zamir *et al*. [13] | Cho *et al*. [14] | Li *et al*. [15] | Ours

Figure 7. Visual comparison of the deblurred result of "Notredame" from Lai's [1] . Zoom-in for better inspection.

Table 2. Average PSNR/SSIM of the results from different methods on the dataset Levin *et al*. [18].

	Non-learning methods			Supervised learning methods						Self-supervised	
Metric	Xu & Jia 2010' [19]	Yan *et al*. 2017' [20]	Yang & Ji 2019' [21]	Chakrabarti 2016' [22]	Pan *et al*. 2017' [23]	Zuo *et al*. 2016' [24]	Tao *et al*. 2018' [11]	Kupyn *et al*. 2019' [12]	Li *et al*. 2022' [15]	Ren *et al*. 2020' [7]	Ours
PSNR	31.64	31.28	32.04	25.21	30.42	32.66	26.12	25.70	25.28	<u>33.31</u>	34.34
SSIM	0.910	0.912	0.912	0.785	0.907	0.933	0.797	0.790	0.780	<u>0.943</u>	0.939

blurred images.

7. Demonstration on processing the images of dynamic scenes from GOPRO dataset

Recall that the proposed approach is designed to recover blurred images of static scenes with non-uniform blurring with static scenes. The "blending" effect often seen in the images of dynamic scenes is not considered in our method. As a result, the proposed method cannot process most images in GOPRO dataset [12], which are the ones of dynamic scenes.

For the purpose of illustration, we select some images from the GOPRO dataset, whose contents are dominated by static scenes. See Figure 14 for a demonstration. It can be seen that the proposed method works well. However, when being used for processing the images whose blurring are mainly caused by moving objects, the proposed method fails. See Figure 15

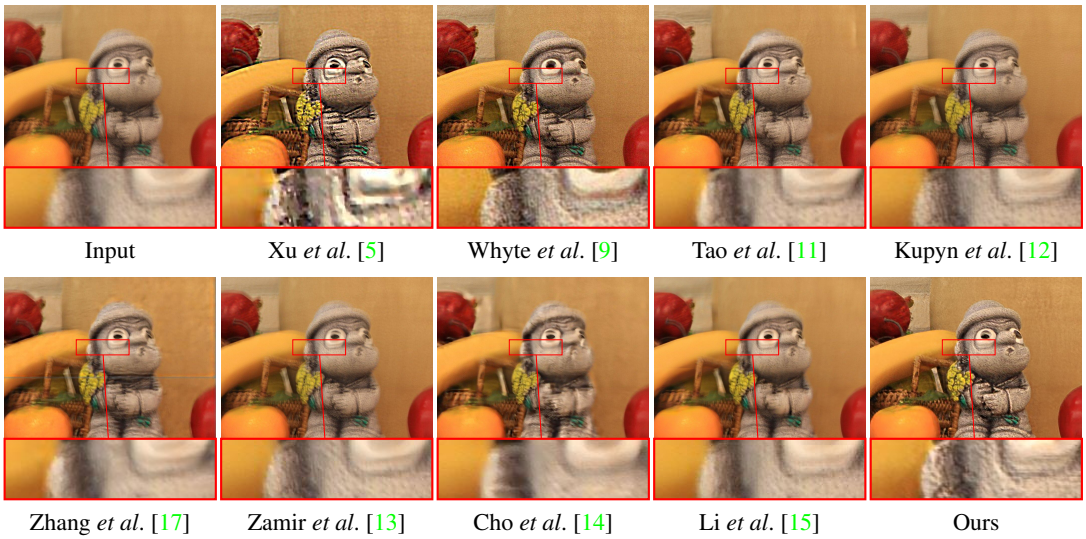


Figure 8. Visual comparison of the deblurred result of "Harubang" from Lai's [1]. Zoom-in for better inspection.

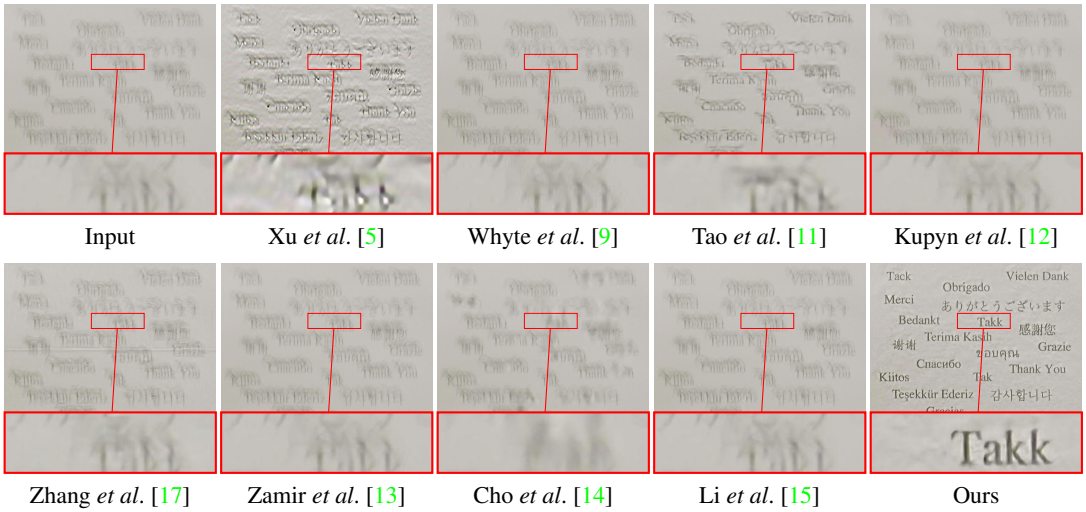
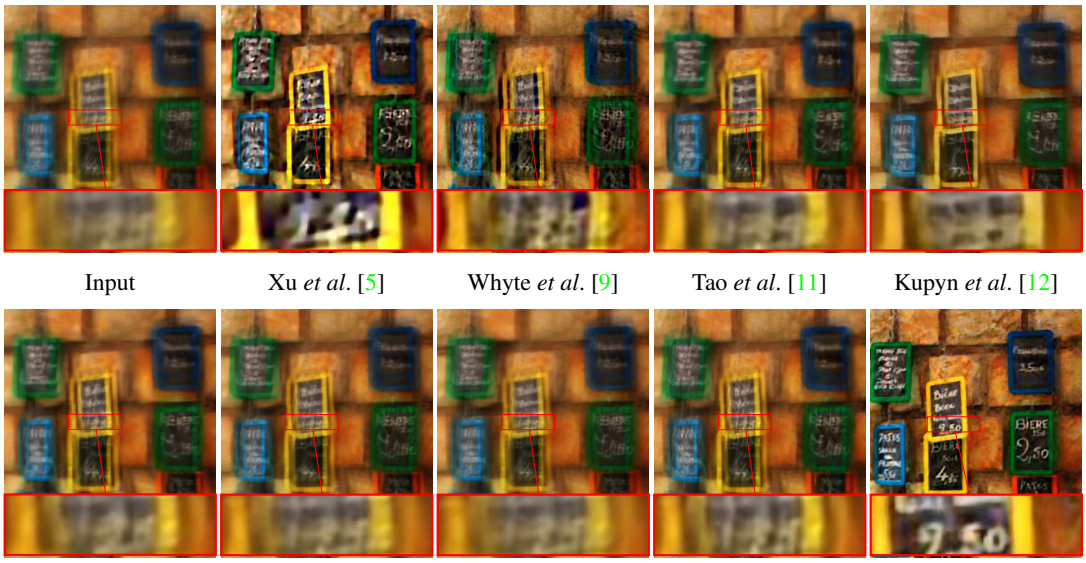


Figure 9. Visual comparison of the deblurred result of "Text3" from Lai's [1]. Zoom-in for better inspection.

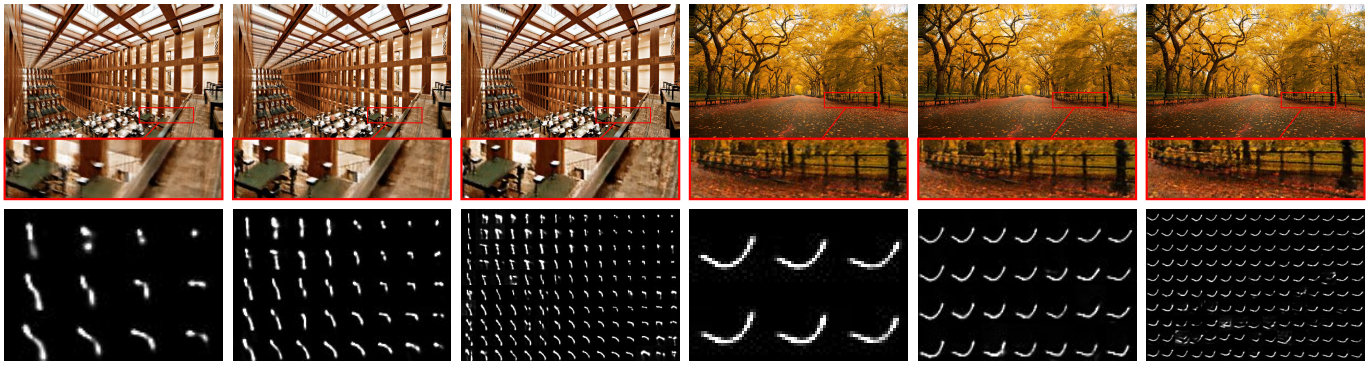
for an illustration. It would be our future work to extend the proposed method to process the images of dynamic scenes, *i.e.*, how to deal with the "blending" effect caused by moving objects.



Input | Xu *et al*. [5] | Whyte *et al*. [9] | Tao *et al*. [11] | Kupyn *et al*. [12]

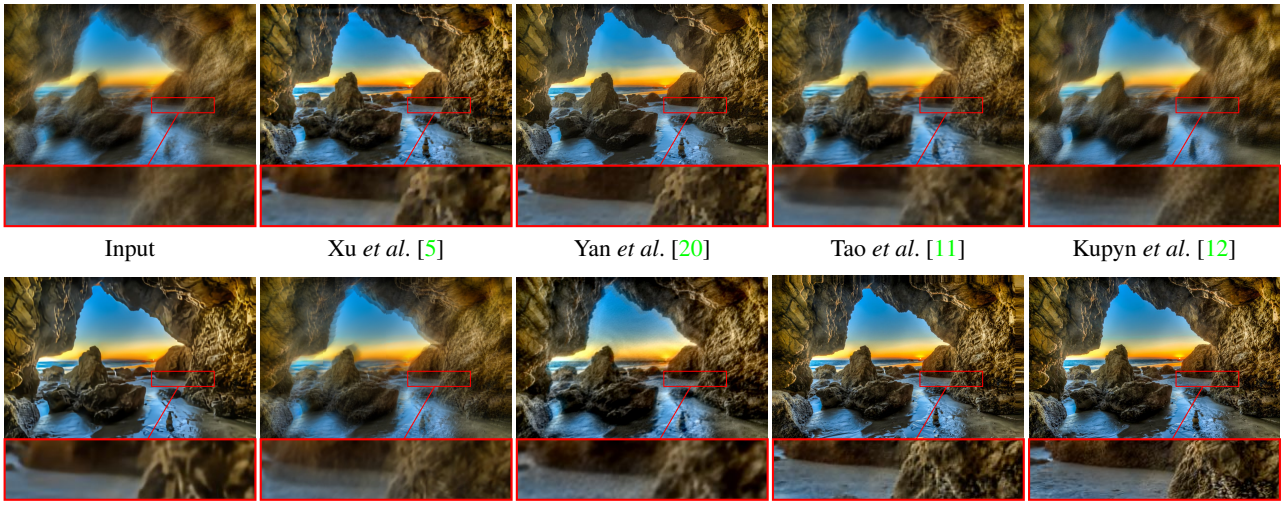
Zhang *et al*. [17] | Zamir *et al*. [13] | Cho *et al*. [14] | Li *et al*. [15] | Ours

Figure 10. Visual comparison of the deblurred result of "Wall" from Lai's [1]. Zoom-in for better inspection.



$P = 3 \times 4$ | $P = 5 \times 8$ | $P = 11 \times 16$ | $P = 3 \times 4$ | $P = 5 \times 8$ | $P = 11 \times 16$

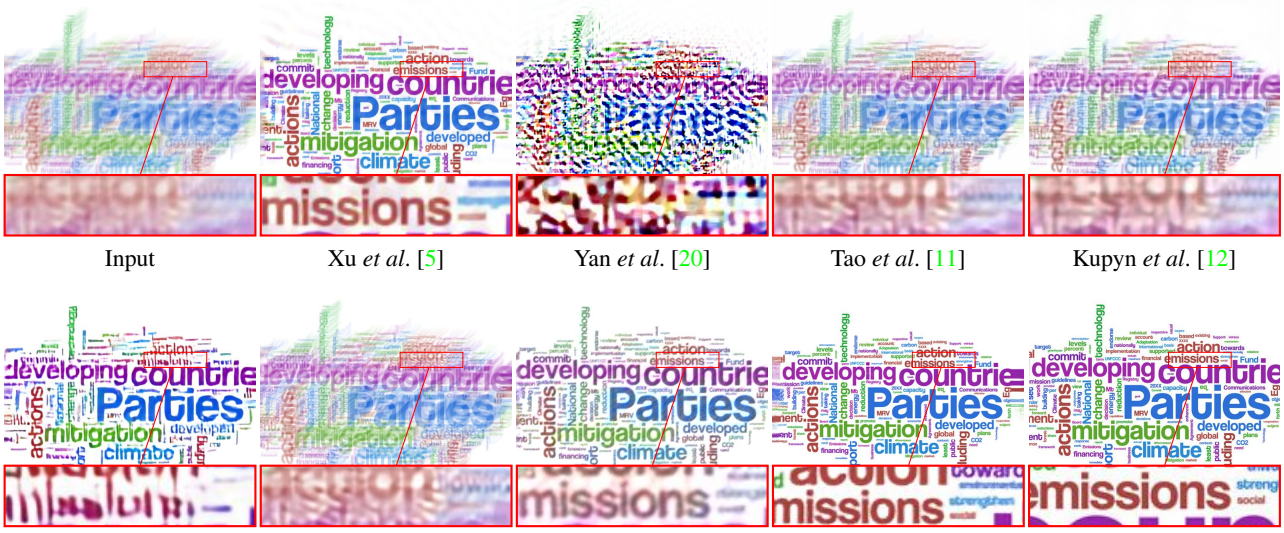
Figure 11. Visual comparison of the recovered results w.r.t the different choices of number of patch P. Noted that the kernel for different P has the same support. Zoom-in for better inspection.



Input | Xu *et al*. [5] | Yan *et al*. [20] | Tao *et al*. [11] | Kupyn *et al*. [12]

Kaufman *et al*. [25] | Li *et al*. [15] | Ren *et al*. [7] | Ours | GroundTruth

Figure 12. Visual comparison of the deblurred results of category "natural" from uniform dataset of Lai *et al*. [1]. Zoom-in for better inspection.



Input | Xu *et al*. [5] | Yan *et al*. [20] | Tao *et al*. [11] | Kupyn *et al*. [12]

Kaufman *et al*. [25] | Li *et al*. [15] | Ren *et al*. [7] | Ours | GroundTruth

Figure 13. Visual comparison of the deblurred results of "natural" from uniform dataset of Lai *et al*. [1]. Zoom-in for better inspection.

Input | Tao *et al*. [11] | Kupyn *et al*. [12] | Zhang *et al*. [17] | Li *et al*. [15] | Ours

Figure 14. Visual comparison of some deblurred results of blurring images dominated by static scene in Gopro dataset [12]. Zoom-in for better inspection.

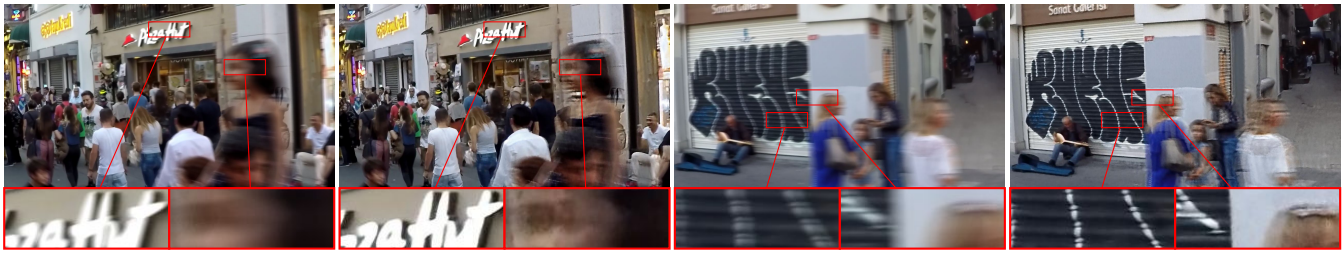


Figure 15. Illustration of our methods can not be adapted to dynamic scene deblurring. The dynamic deblurring contains moving objects, which needs explicitly/implicitly segmentation. Our methods can faithfully recover the static regions, but it can not account for the boundary issue between moving objects and the background.

References

[1] Wei-Sheng Lai, Jia-Bin Huang, Zhe Hu, Narendra Ahuja, and Ming-Hsuan Yang. A comparative study for single image blind deblurring. In *CVPR*, pages 1701–1709, 2016. 1, 2, 3, 4, 5, 6, 7, 8

[2] Crispin W Gardiner et al. *Handbook of stochastic methods*, volume 3. springer Berlin, 1985. 1

[3] Max Welling and Yee W Teh. Bayesian learning via stochastic gradient Langevin dynamics. In *Proceedings of the 28th international conference on machine learning (ICML-11)*, pages 681–688, 2011. 1

[4] Michael Hirsch, Christian J Schuler, Stefan Harmeling, and Bernhard Schölkopf. Fast removal of non-uniform camera shake. In *ICCV*, pages 463–470. IEEE, 2011. 2, 3

[5] Li Xu, Shicheng Zheng, and Jiaya Jia. Unnatural ℓ_0 sparse representation for natural image deblurring. In *CVPR*, pages 1107–1114, 2013. 2, 3, 4, 5, 6, 7, 8

[6] Oliver Whyte, Josef Sivic, and Andrew Zisserman. Deblurring shaken and partially saturated images. *IJCV*, 110(2):185–201, 2014. 2

[7] Dongwei Ren, Kai Zhang, Qilong Wang, Qinghua Hu, and Wangmeng Zuo. Neural blind deconvolution using deep priors. In *CVPR*, pages 3341–3350, 2020. 2, 4, 5, 8

[8] Rolf Köhler, Michael Hirsch, Betty Mohler, Bernhard Schölkopf, and Stefan Harmeling. Recording and playback of camera shake: Benchmarking blind deconvolution with a real-world database. In *ECCV*, pages 27–40. Springer, 2012. 2, 3

[9] Oliver Whyte, Josef Sivic, Andrew Zisserman, and Jean Ponce. Non-uniform deblurring for shaken images. *IJCV*, 98(2):168–186, 2012. 2, 3, 4, 5, 6, 7

[10] Subeesh Vasu and AN Rajagopalan. From local to global: Edge profiles to camera motion in blurred images. In *CVPR*, pages 4447–4456, 2017. 2, 3, 4

[11] Xin Tao, Hongyun Gao, Xiaoyong Shen, Jue Wang, and Jiaya Jia. Scale-recurrent network for deep image deblurring. In *CVPR*, 2018. 2, 3, 4, 5, 6, 7, 8, 9

[12] Orest Kupyn, Tetiana Martyniuk, Junru Wu, and Zhangyang Wang. Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better. In *ICCV*, pages 8878–8887, 2019. 2, 3, 4, 5, 6, 7, 8, 9

[13] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Multi-stage progressive image restoration. In *CVPR*, pages 14821–14831, 2021. 2, 3, 4, 5, 6, 7

[14] Sung-Jin Cho, Seo-Won Ji, Jun-Pyo Hong, Seung-Won Jung, and Sung-Jea Ko. Rethinking coarse-to-fine approach in single image deblurring. In *ICCV*, pages 4641–4650, 2021. 2, 3, 4, 5, 6, 7

[15] Dasong Li, Yi Zhang, Ka Chun Cheung, Xiaogang Wang, Hongwei Qin, and Hongsheng Li. Learning degradation representations for image deblurring. In *ECCV*, 2022. 2, 3, 4, 5, 6, 7, 8, 9

[16] Peidong Liu, Joel Janai, Marc Pollefeys, Torsten Sattler, and Andreas Geiger. Self-supervised linear motion deblurring. *IEEE Robot. Autom. Lett.*, 5(2):2475–2482, 2020. 3, 4

[17] Hongguang Zhang, Yuchao Dai, Hongdong Li, and Piotr Koniusz. Deep stacked hierarchical multi-patch network for image deblurring. In *CVPR*, pages 5978–5986, 2019. 5, 6, 7, 9

[18] Anat Levin, Yair Weiss, Fredo Durand, and William T Freeman. Understanding and evaluating blind deconvolution algorithms. In *CVPR*, pages 1964–1971, 2009. 4, 5

[19] Li Xu and Jiaya Jia. Two-phase kernel estimation for robust motion deblurring. In *ECCV*, pages 157–170, 2010. 5

[20] Yanyang Yan, Wenqi Ren, Yuanfang Guo, Rui Wang, and Xiaochun Cao. Image deblurring via extreme channels prior. In *CVPR*, pages 4003–4011, 2017. 5, 8

[21] Liuge Yang and Hui Ji. A variational EM framework with adaptive edge selection for blind motion deblurring. In *CVPR*, pages 10167–10176, 2019. 5

[22] Ayan Chakrabarti. A neural approach to blind motion deblurring. In *ECCV*, pages 221–235, 2016. 5

[23] Jinshan Pan, Jiangxin Dong, Yu-Wing Tai, Zhixun Su, and Ming-Hsuan Yang. Learning discriminative data fitting functions for blind image deblurring. In *ICCV*, pages 1068–1076, 2017. 5

[24] Wangmeng Zuo, Dongwei Ren, David Zhang, Shuhang Gu, and Lei Zhang. Learning iteration-wise generalized shrinkage–thresholding operators for blind deconvolution. *IEEE Trans. Image Processing*, 25(4):1751–1764, 2016. 5

[25] Dilip Krishnan and Rob Fergus. Fast image deconvolution using hyper-Laplacian priors. In *NIPS*, pages 1033–1041, 2009. 8