Paper Tag: WED-PM-156

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

Highlights

- We present a self-supervised deep learning approach to restore motion-blurred images due to camera shake
- Contributions: The proposed approach is
  - The first dataset-free deep learning method for removing general motion blur (uniform and non-uniform) from images due to camera shake
  - The first approach that combines DNN-based re-parametrization and EM algorithm
  - A powerful method that significantly outperforms existing solutions for blind motion deblurring

Background

- Motion blur occurs when the camera shakes during the shutter time, resulting in a blurring effect
- To remove the uniform and non-uniform motion blur caused by camera shake from an image

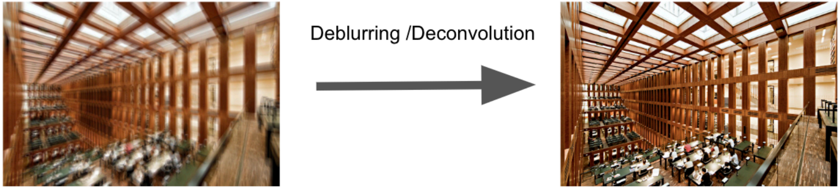


Figure 1: Blind image deblurring

Main idea

- DNN-based re-parametrization of the image and the kernel set
  - DIP for latent image
  - Multi-head NN for kernel set with embedded two priors
    - Implicit low-dimensional prior
    - Physical constraints prior with Softmax layer
- **Contribution**: Monte Carlo Expectation Maximization (MCEM) approach for NN weights inference of kernel network motivated from the Bayesian inference for blind deconvolution

Cont’d

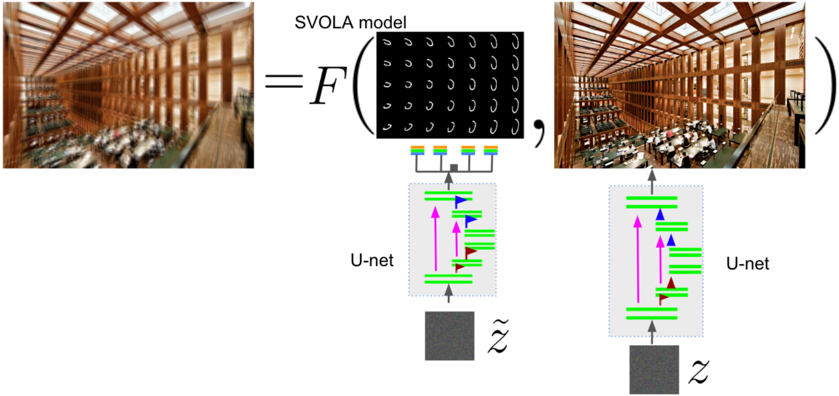


Figure 2: The diagram of our solution

Outline of MCEM algorithm

- Two unknowns are reparameterized by two networks
  - Image network $T_f(\boldsymbol{\theta}_f, \boldsymbol{z})$ and the kernel network $T_k(\boldsymbol{\theta}_k, \tilde{\boldsymbol{z}})$
- The setting of kernel inference problem in EM framework
  1. Observation data: the blurred image $\boldsymbol{g}$
  2. Latent variable: the weights $\boldsymbol{\theta}_f$ of image-relating network $T(\boldsymbol{\theta}_f, \boldsymbol{z})$
  3. Parameters: the weights $\boldsymbol{\theta}_K$ of kernel-relating network $T(\boldsymbol{\theta}_K, \tilde{\boldsymbol{z}})$

MCEM algorithm using Langevin dynamics

- E-step: Calculate the expectation of logarithm likelihood with respective to $p(\boldsymbol{\theta}_f|\boldsymbol{g};\boldsymbol{\theta}_K^{t-1})$:

$$Q(\boldsymbol{\theta}_K|\boldsymbol{\theta}_K^{t-1}) = \mathbb{E}_{\boldsymbol{\theta_f}\sim p(\boldsymbol{\theta}_f|\boldsymbol{g};\boldsymbol{\theta}_K^{t-1})}\left[\log p(\boldsymbol{g}|\boldsymbol{\theta}_f;\boldsymbol{\theta}_K)\right] \tag{1}$$

- M-step: Maximize the expectation of the likelihood:

$$\boldsymbol{\theta}_K^t = \arg\max_{\boldsymbol{\theta}_K} Q(\boldsymbol{\theta}_K|\boldsymbol{\theta}_K^{t-1}) \tag{2}$$

- Use Monte-Carlo to address the intractable expectation computation

$$Q(\boldsymbol{\theta}_K|\boldsymbol{\theta}_K^{t-1}) \approx \frac{1}{n_s}\sum_{i=1}^{n_s}\log p(\boldsymbol{g}|\boldsymbol{\theta}_f^i;\boldsymbol{\theta}_K), \boldsymbol{\theta}_f^i \sim p(\boldsymbol{\theta}_f|\boldsymbol{g};\boldsymbol{\theta}_K^{t-1}) \tag{3}$$

Langevin dynamics based sampling

- LD samples the distribution $p(\boldsymbol{\theta}_f|\boldsymbol{g};\boldsymbol{\theta}_K^{t-1})$ by the so-called *stochastic gradient Langevin dynamics* (SGLD): For $i = 1, 2, \ldots, n_s$

$$\boldsymbol{\theta}_f^i = \boldsymbol{\theta}_f^{i-1} + \alpha \nabla_{\boldsymbol{\theta}_f} \log p(\boldsymbol{\theta}_f^{i-1}|\boldsymbol{g};\boldsymbol{\theta}_K^{t-1}) + \sqrt{2\alpha}\boldsymbol{w}, \tag{4}$$

where $\boldsymbol{w} \sim \mathcal{N}(0, I)$

Training stage with warm-up trick

- Use the alternative EM algorithm to estimate the kernel set and the latent image
- A warm-up strategy is implemented to initialize the multi-head subnetwork for kernel set using the uniform deblurring

Experimental setting

- The image NN $T_{\boldsymbol{f}}(\boldsymbol{\theta_f})$ is implemented as 5-level U-Net with channel size 64. The kernel NN $T_{\boldsymbol{K}}(\boldsymbol{\theta_K})$ is implemented as U-Net with 4 levels whose channel size is $[32, 32, 64, 64]$
- The learning rate is set to be 0.01 for $T_{\boldsymbol{f}}$ and 0.0001 for M-step when optimizing $T_{\boldsymbol{K}}$

Quantitative evaluation

Table 1: Average PSNR comparison on the non-uniform dataset of Köhler

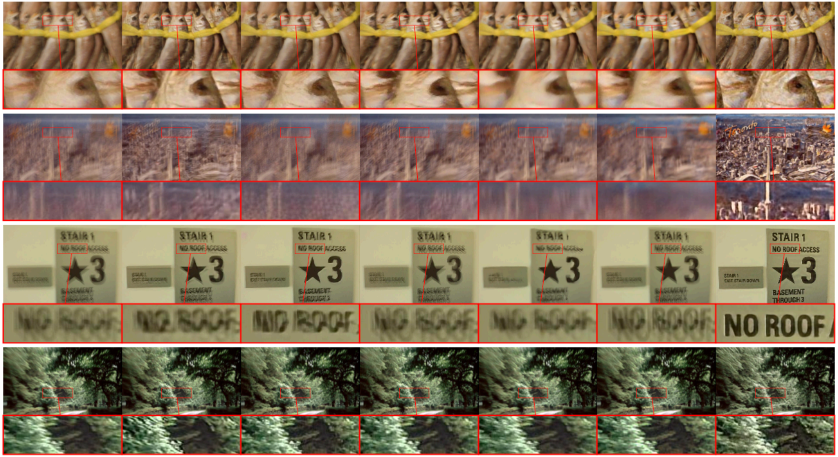
	Non-learning methods			Supervised learning methods					Self-supervised	
No.	Xu	Whyte	Vasu	Tao	Kupyn	Zamir	Cho	Li	Liu	Ours
1	29.19	29.77	32.44	29.14	28.99	29.86	27.66	29.82	27.21	<u>32.41</u>
2	24.43	24.27	<u>26.52</u>	23.10	23.78	22.57	21.69	22.56	21.19	26.84
3	29.97	30.73	<u>32.60</u>	29.96	30.00	28.04	28.09	30.21	28.20	33.18
4	25.76	26.60	<u>27.99</u>	25.22	25.09	24.78	23.91	24.95	23.49	28.60
Avg.	27.34	27.84	<u>29.89</u>	26.85	26.97	26.32	25.34	26.89	25.02	30.26

Cont'd

Table 2: Average PSNR comparison on the non-uniform dataset of Lai

	Non-learning methods			Supervised learning methods					Self-supervised	
	Xu	Whyte	Vasu	Tao	Kupyn	Zamir	Cho	Li	Liu	Ours
Manmade	17.90	17.33	17.93	18.45	<u>18.73</u>	17.42	16.78	17.28	17.39	19.17
Natural	21.99	21.04	21.94	<u>22.28</u>	22.24	20.76	19.88	20.59	20.90	22.70
People	25.42	23.92	25.63	<u>26.87</u>	26.71	23.95	23.64	24.23	24.76	26.90
Saturated	18.39	17.33	17.57	<u>20.10</u>	17.91	16.73	16.58	16.67	18.52	21.46
Text	18.97	13.22	<u>19.19</u>	18.66	19.11	15.63	17.17	17.45	17.42	21.91
Average	20.53	18.57	20.45	<u>21.27</u>	20.94	18.90	18.81	19.25	19.80	22.42

Visualization



Input Tao Kupyn Park Zamir Li Ours

Figure 3: Visual comparison of the results for samples images from real dataset of Lai and Sun

Acknowledgment

- People
  - Hui Ji (Prof. from NUS), who leads the project
  - Yuesong Nan (PhD from NUS), now works for Zoom. Inc.
  - Weixi Wang (PhD from NUS), now works for DBS, Singapore
- Grant: Singapore MOE Academic Research Fund (AcRF) Tier 1 with WBS number A-8000981-00-00