CSC2529 Project Proposal: Weakly-supervising the Deep Priors for Blind Deconvolution

Zhixiang Chi
University of Toronto, Canada
zhixiang.chi@mail.utoronto.ca

Abstract

In this project, we are going to investigate the deep image prior method for blind image deblurring. There are several drawbacks observed for the existing work, 1) lack of direct supervision; 2) non-adaptive and sub-optimal early stopping policy; 3) limitation on the dynamic scene / spatially non-uniform deblurring. Several potential improvements are proposed using a pre-trained deblurring network as a weak supervisor to provide direct supervision and a early stopping strategy.

1. Motivation and Related works

In this work, we mainly focus on the task of blind image deconvolution where the blur kernel is unknown. We start from the image degradation model:

$$\mathbf{y} = \mathbf{x} * \mathbf{H} + \mathbf{N}, \tag{1}$$

where $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{N}$ are the clean image, degraded image and noise respectively. $\mathbf{H}$ and $*$ are the task-dependent degradation function and operator. For example, for blury image formation, $\mathbf{H}$ is the blur kernel and $*$ is the convolution [3]. while for image in-painting, $\mathbf{H}$ can be a mask image and $*$ is the element-wise multiplication [7].

To restore a degraded image, we aim to develop a method $f(\cdot)$. When given a degraded image $\mathbf{y}$, it generates a restored image $\hat{\mathbf{x}} = f(\mathbf{y})$ that is closer to the ground truth clean image $\mathbf{x}$. For deep learning-based method, especially the end-to-end ones, we denote the mapping function parameterized by θ as $f_\theta(\cdot)$. For learning-based methods, $f_\theta(\cdot)$ is learned from large-scale degraded/clean image pairs, so that the restoration prior or knowledge is encoded in the weights θ [5].

As an alternative research direction, Deep Image Prior (DIP) [6] finds that an untrained deep model is also capable to capture some of the low-level statistics of natural images. Given each degraded image, the optimization is formulated as:

$$\theta^* = \arg\min_\theta E(f_\theta(z); \mathbf{y}),\ \mathbf{y}^* = f_{\theta^*}(z). \tag{2}$$

$E(\cdot)$ is the data fidelity term, which normally aims to minimize the distance between two inputs, therefore, making the network outputs closer to the given data. z is the noise that follows Gaussian Distribution.

Through iterative optimization of Eq. 2. The network $f_\theta(z)$ aims to reproduce the degraded image. However, as discovered in [6], such parametrization offers high impedance to noise but low impedance to signals. In other words, the optimization process produces natural clean images first before fitting to the degraded image. Thus, an early stopping for such degradation reconstruction optimization leads to a image restoration solution. It has been shown the effectiveness on restoration tasks, such as image denoising, super-resolution, in-painting.

However, for the task of blind image deblurring, DIP is degenerated as it has the limitation on capturing the prior of blur kernels [4]. Thus, two separate generative networks $\mathcal{G}_x$ and $\mathcal{G}_k$ are proposed to replace f_θ. $\mathcal{G}_x$ aims to capture the image prior, while $\mathcal{G}_k$ is expected to model the blur kernel k. The optimization is then becomes:

$$\min_{(\mathcal{G}_x, \mathcal{G}_k)} = \left\|\mathcal{G}_k(z_k) \otimes \mathcal{G}_x(z_x) - \mathbf{y}\right\|^2, \tag{3}$$

where z_k and z_x are the input noise for each generator and $\otimes$ is the convolution operator. The goal of Eq. 3 is to generate a clean image and the corresponding blur kernel so that they can be transformed into the input blurry image via blur degradation model. Note, such supervision is self-supervised without any ground-truth image.

There are several drawbacks observed:

- The clean image is the intermediate output of the whole system, but the supervision is applied to the final reconstructed blurry image. It leads to indirect optimization of the expected output, and the training objective and evaluate protocol do not match.
- Different input blurry images may require different number of iterations. The current method has a fixed

iteration number for all images, which is sub-optimal. Although the intermediate output is constrained by the reconstruction loss, its convergence cannot be guaranteed.

- The current method is designed for spatially uniform blurs. An investigation is needed to assess its limitations on dynamic scenes, which involves non-uniform blurs (e.g., moving objects).

2. Project overview

In this project, we are going to explore if providing a supervision directly at the expected output is beneficial. Concretely, the direct supervision on the expected intermediate output might enhance the prior extracting process. However, the only available information is the degraded input image $\mathbf{y}$, which reduces the flexibility of such method.

To alleviate such problem, we follow the self-training method [2], which use the trained model to label the unlabeled data. Specifically, we employ a deep model g that is trained on a large-scale dataset for the deblurring task:

$$\bar{\mathbf{x}} = g(\mathbf{y}). \tag{4}$$

We name $\bar{\mathbf{x}}$ as the surrogate ground truth which is the deblurred version of $\mathbf{y}$ using g. Although $\bar{\mathbf{x}}$ highly depends on the architecture of g and the training dataset, and may contain defected results, it has learned through abundant data samples to extract features. $\bar{\mathbf{x}}$ can be used as a weak supervision to guide $\mathcal{G}_x(z_x)$. Therefore, we propose to add the following loss:

$$\mathcal{L}_{weak} = E(\mathcal{G}_x(z_x), \bar{\mathbf{x}}). \tag{5}$$

E measures the distance between $\mathcal{G}_x(z_x)$ or $\bar{\mathbf{x}}$ which can be $L1/L2$ norm, or other similarity metrics.

Note, $\bar{\mathbf{x}}$ also represents the knowledge of the trained model g. Thus, using Eq. 5 is analogy to the knowledge distillation technique [1]. Specifically, the knowledge of a well-trained model is transferred to a model with lower learning capability.

Another constraint can be added at the final degraded output. We expect that passing the generated degraded image $\hat{x}$ to g will have similar deblur results (either comparing the final predictions or the intermediate features) as passing $\bar{x}$. Therefore, another loss term can be added as:

$$\mathcal{L}_{rec} = E(g(\hat{x}), g(\bar{x})). \tag{6}$$

To determine the adaptive early stopping strategy, g can also be utilized. For example, a stopping policy can be determined by choosing a threshold t:

$$E(g(\hat{x}), g(\bar{x})) \leq t. \tag{7}$$

We can also set a rule so that the training can be stopped if the criteria of Eq. 7 can last for certain consecutive iterations to eliminate the effect of unstable training iterations.

3. Milestones, timeline & goals

- Week 1 (Nov.17 - Nov.23): get familiar with the codebase of SelfDeblur [4]. Investigate the proper methods for g. Evaluation of g is needed for the performance on different datasets.

- Week 2 (Nov.24 - Nov.30): conduct the experiments of the proposed methods and record the results quantitatively and qualitatively.

- Week 3 (Dec.1 - Dec.8) continue the experiments and investigate potential extension to dynamic scene deblurring. Report write-up, prepare the posters.

References

[1] Geoffrey Hinton, Oriol Vinyals, Jeff Dean, et al. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*, 2(7), 2015. 2

[2] Xinzhe Li, Qianru Sun, Yaoyao Liu, Qin Zhou, Shibao Zheng, Tat-Seng Chua, and Bernt Schiele. Learning to self-train for semi-supervised few-shot classification. *Advances in Neural Information Processing Systems*, 2019. 2

[3] Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017. 1

[4] Dongwei Ren, Kai Zhang, Qilong Wang, Qinghua Hu, and Wangmeng Zuo. Neural blind deconvolution using deep priors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020. 1, 2

[5] Ying Tai, Jian Yang, and Xiaoming Liu. Image super-resolution via deep recursive residual network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017. 1

[6] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018. 1

[7] Junyuan Xie, Linli Xu, and Enhong Chen. Image denoising and inpainting with deep neural networks. *Advances in neural information processing systems*, 2012. 1