J. Vis. Commun. Image R. 65 (2019) 102648

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

journal homepage: www.elsevier.com/locate/jvci

Generative image deblurring based on multi-scaled residual adversary network driven by composed prior-posterior loss ☆

Meng Wang *, Shengyu Hou, Huafeng Li, Fan Li

Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, China
Yunnan Key Laboratory of Artificial Intelligence, Kunming University of Science and Technology, Kunming 650500, China

ARTICLE INFO

Article history:
Received 21 May 2019
Revised 20 July 2019
Accepted 14 September 2019
Available online 26 September 2019

Keywords:
Image deblurring
Generative adversarial network
Residual learning
Prior distribution
Histogram of gradients

ABSTRACT

Conditional Generative Adversarial Networks (CGANs) have been introduced to generate realistic images from extremely degraded inputs. However, these generative models without prior knowledge of spatial distributions has limited performance to deal with various complex scenes. In this paper, we proposed a image deblurring network based on CGANs to generate ideal images without any blurring assumption. To overcome adversarial insufficiency, an extended classifier with different attribute domains is formulated to replace the original discriminator of CGANs. Inspired by residual learning, a set of skip-connections are cohered to transfer multi-scaled spatial features to the following high-level operations. Furthermore, this adversary architecture is driven by a composite loss that integrates histogram of gradients (HoG) and geodesic distance. In experiments, an uniformed adversarial iteration is circularly applied to improve image degenerations. Extensive results show that the proposed deblurring approach significantly outperforms state-of-the-art methods on both qualitative and quantitative evaluations.

1. Introductions

Image deblurring aims to recover the spatial details from blurred images that are captured by defective imaging mechanisms [1,2]. The fine images recovered by deblurring procedures not only satisfy visually perception but also assists various high-level vision tasks, such as image classification [3] and object recognition [4]. At present, though numerous deblur approaches have been proposed, it is still challenge to precisely determine a reverse blur model for image recovering, since the solution is generally limited by ill-posed conditions [5–7]. The existing solutions can be categorized into two groups, i.e. blind deblur and non-blind deblur. Early work mostly focused on the non-blind deblurring which assumes the blur kernel to be utilized is known [8]. These works perform deconvolutions by Lucy-Richardson algorithm [9], Wiener or Tikhonov filter [10], however, the blur kernel is commonly unknown in reality, thus the blind methodology has been widely developed by simultaneously estimating ideal images and blur kernels.

An early landmark of the blind deblurring is achieved by Miskin and MacKay [11]. In their work, a mixture of Laplace distributions is utilized to restore cartoon images. Later, Likas and Galatsanos adopt a Gaussian prior distribution to enhance smoothness on the spatial reconstruction [12], and another similar work as in [13]. Fergus et al. suggest using a mixture-of-Gaussian (MOG) to ensure visual saliency [14]. Bishop et al. [15] review blind deblurring methods, including classical prior models such as conditional auto-regression (CAR) and simultaneous auto-regression (SAR) introduced by Molina et al. [16], or the total variation as suggested in Ref. [17] by imposing piecewise-smoothness. The above approaches mainly focused on image prior distributions. Furthermore, according to the mechanism of motion blurring, Whyte et al. develop a method for non-uniform blind deblurring by a parametrized geometric model with the rotational velocity of the camera during exposure [18]. Yan and Shao et al. parameterize blur kernels and then estimate them by classification algorithms [19]. Similarly, Gupta et al. estimate the blur kernels by simulating motion trajectories of the camera [20]. These mentioned approaches are usually based on a assumption that the blur kernel is partially uniform or locally linear, such as Gaussian blur models [21]. However, this assumption cannot match challenging cases with occluded regions or depth variations. In addition, these shallow models often yield erroneous kernel estimations and undesired distortions.

☆ This paper has been recommended for acceptance by Zicheng Liu.

* Corresponding author at: Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, China.

E-mail address: vicong68@qq.com (M. Wang).

https://doi.org/10.1016/j.jvcir.2019.102648

In recent years, due to the success of deep networks, deblurring approaches based on CNNs have attracted wide attention [5,6,22]. Sun et al. utilize CNNs to learn blur kernels [6], whereas Gong et al. develop a fully convolutional network for motion flow estimation [22]. In addition, Schuler et al. train a deep network to gain the blur kernel, and then adopt a conventional non-blind deconvolution algorithm to recover the latent sharp image [5]. With the development of deep generative networks, image restoration have achieved promising performance [7,23,24]. Xu et al. suggest to train CNNs for space-invariant non-blind deconvolution [7]. This network can handle complex blurring scenes, but the blur kernels has to be properly initialized. Moreover, Jain and Seung utilize a small CNN to remove additive white Gaussian noise with unknown energy [23]. A common objective function of these CNN-based models is the pixel-wise loss between the deblurred and ideal images, such as L2 and SSIM loss [25,26]. The pixel-wise losses lead to high structural evaluations, nevertheless, the outputs often tend to involve distortions in large scales as the lack of content priors. Recently, studies on conditional generative adversarial networks (CGANs) indicate that the discriminant features are expressive enough to generate realistic images. According to this, CGAN-based deblurring has been verified more powerful than the conventional CNNs [27,28]. However, though the original CGANs is effective at learning spatial features, the binary discriminator fails to capture more specific domain knowledge, such as the categorical information.

Overall, the above approaches that employ different generative architectures have three major limitations: (i) Due to the simple assumptions of blur kernels, the kernel estimations are practically inaccurate and limited to specific blurring models. Besides, the general deep architectures are difficult to be perfectly optimized since the over-redundant layered kernels. (ii) Although the evaluations can be improved through the posterior distance loss, the visual perception of detail contents is not satisfactory as lack of recovering prior. (iii) Existing deblurring networks based on CGANs are fail to capture the domain knowledge such as detailed semantics contents. In this case, the adversarial balance might be hard to reach during the training iterations.

To overcome these limitations, this paper focuses on an extended framework based on CGANs for high-quality image deblurring. The main idea is that an accurate recovery model should be able to represent salient contents from a global perspective with a progressive learning procedure. A widely-used prior assumption is that gradients of natural images have a heavy-tailed distribution [14]. Inspired by it, a gradient histogram is thus utilized to restore salient textures while the adversarial loss focuses on restoring general contents. To overcome deficiencies of the original CGANs, we formulate an extended classifier to instead of the binary discriminator, then enhance the residual architecture [29] to relieve the loss of high-frequency features in the generative procedures. Finally, state-of-the-art deblurring methods are evaluated on various scenes including ImageNet [3], CelebA [30], MNIST, GoPro [24] and Lai's dataset [31] to verify the effectiveness and adaptability of the proposed framework. The main contributions in this paper can be summarized as follows:

(1) A deep generative network is proposed to directly recover blurred images without any assumption to the deblurring kernels. Aim at this, a multi-scaled residual adversarial network (ResCGAN) is further applied, which can greatly relieve the loss of transferring high-frequency features in the generative network.
(2) Then, an extended classifier is formulated to replace the original discriminator of CGANs. This classifier can not only collaborate with the binary discrimination, but also learn attributive domain representations to generate more detail contents.
(3) In addition, a composite loss is introduced by integrating posterior geodesic distances and prior gradient histograms. The experimental results verify that perceptive evaluations of the proposed network are significant improved, and the adversarial balance is efficiently achieved.

The rest of this paper is organized as following. Section 2 briefly reviews related work as basic background, and the proposed approach is detailed in the next section. Section 4 illuminates the comparison results as well as their analysis. We finally conclude this paper in Section 5.

2. Related work

2.1. Deep generative formulas

CNNs have been widely applied in discriminative vision tasks, such as object detection [32], image classification [33], image fusion [34,35] and video deblurring [36,37]. On the basis, deep generative architectures that obtain images with desired contents recently achieve impressive performance [6,7,14,38–41]. In Ref. [7], a non-blind deblurring model with deep convolution architecture is proposed. These data-driven approaches bridge traditional optimization schemes and empirically determinative CNNs. Wong et al. consider model parameter estimation as a key issue in the overall optimization of image restorations [41], thus establish relationships between the mask and the regional parameters. Sun et al. estimate the probabilistic distribution of motion kernels by CNNs, then utilize a Markov random field to formulate a non-uniform deblurring model [6]. Cai et al. removes motion blurring from a single image by a blind prediction model, which simultaneously maximizes the sparsity of blur kernels [40]. Fergus et al. recover blur kernels by using natural image prior on pixel gradients in a variational Bayes framework [14]. Jia et al. use transparency maps to obtain motion blur kernels by performing blind-deconvolution [38]. Joshi et al. estimate a sharp image that should be consistent with an blurred image [39]. However, these deblurring approaches still involve explicit kernel estimation, and if the kernel prediction is inaccurate, the deblurred results often involve undesired artifacts. Furthermore, CNNs have shown its effectiveness in video deblurring. Zhang et al. [36] propose a CNN model called DBLRNet, which learns spatio-temporal features by 3D convolutions, and integrates the 3D deblurring network into a generative adversarial network to achieve photo-realistic results. Su et al. [37] train an end-to-end CNN for video deblurring, where the input is a stack of neighboring frames and the output is the deblurred central frame in the stack.

Image prior is an effective component to generative models, since it can relieve the ill-posed problem on image enhancing. Ren et al. [42] propose an enhanced prior for blind image deblurring by combining the low rank prior of similar patches from both the blurry image and its gradient map. In this model, the low rank prior is employed to generate the intermediate images by eliminating fine texture details and tiny edges while maintaining the dominant structures in blurry images. Cao et al. [43] propose a text image deblurring method using Text-specific Multi-scale Dictionaries and a natural scene dictionary, which constructs the dedicated priors for real world text and non-text fields. According to them, we further propose a composite loss by integrating posterior geodesic distances and prior gradient histograms. Our prior is designed from a simple criterion, that is the prior should be applied to discriminate clear details from blurred scenarios.

On the other hand, residual networks [27,29,44] exhibit excellent performance in computer vision problems aligning the