

Tell Me Where It is Still Blurry: Adversarial Blurred Region Mining and Refining

Jen-Chun Lin^{§‡}, Wen-Li Wei[§], Tyng-Luh Liu[§], C.-C. Jay Kuo[¶], Hong-Yuan Mark Liao[§]

[§]Academia Sinica, Taiwan, [‡]Yuan Ze University, Taiwan, [¶]University of Southern California, Los Angeles, USA
 {jenchunlin, lilijinjin}@gmail.com; liutyng@iis.sinica.edu.tw; cckuo@sipi.usc.edu; liao@iis.sinica.edu.tw



Figure 1: Deblurring results on challenging low-light concert images. Left column shows the blurry image and right column the respective deblurring result by our method.

ABSTRACT

Mobile devices such as smart phones are ubiquitously being used to take photos and videos, thus increasing the importance of image deblurring. This study introduces a novel deep learning approach that can automatically and progressively achieve the task via adversarial blurred region mining and refining (adversarial BRMR). Starting with a collaborative mechanism of two coupled conditional generative adversarial networks (CGANs), our method first learns the image-scale CGAN, denoted as iGAN, to globally generate a deblurred image and locally uncover its still blurred regions through an adversarial mining process. Then, we construct the patch-scale CGAN, denoted as pGAN, to further improve sharpness of the most blurred region in each iteration. Owing to such complementary designs, the adversarial BRMR indeed functions as a bridge between iGAN and pGAN, and yields the performance synergy in better solving blind image deblurring. The overall formulation is self-explanatory and effective to globally and locally restore an underlying sharp image. Experimental results on benchmark datasets demonstrate that the proposed method outperforms the current state-of-the-art technique for blind image deblurring both quantitatively and qualitatively.

CCS CONCEPTS

• **Computing methodologies** → **Artificial intelligence**; **Computer vision**.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

MM '19, October 21–25, 2019, Nice, France

© 2019 Association for Computing Machinery.

ACM ISBN 978-1-4503-6889-6/19/10...\$15.00

<https://doi.org/10.1145/3343031.3351038>

KEYWORDS

image deblurring, deep learning, generative adversarial networks

ACM Reference Format:

Jen-Chun Lin^{§‡}, Wen-Li Wei[§], Tyng-Luh Liu[§], C.-C. Jay Kuo[¶], Hong-Yuan Mark Liao[§]. 2019. Tell Me Where It is Still Blurry: Adversarial Blurred Region Mining and Refining. In *Proceedings of the 27th ACM International Conference on Multimedia (MM '19)*, October 21–25, 2019, Nice, France. ACM, New York, NY, USA, 9 pages. <https://doi.org/10.1145/3343031.3351038>

1 INTRODUCTION

Blind image deblurring is a classical problem in image processing and computer vision. It aims to estimate an unknown sharp image given a blurred version of the original. The increasing prevalence of mobile devices raises the importance of this problem as more photos and videos are taken with smart phones. For example, people can now easily film a live concert, and create video clips of a specific performance. However, as shown in Figure 1, these videos (or images) frequently give rise to poor viewing experience because they are blurred due to issues such as camera shake, out of focus and fast object motion. As reproducing captured moments is generally difficult, to ensure pleasant viewing experience, it is of great interest to enhance sharpness for a higher-quality image or video.

Let \otimes denote the convolution operator. The formation of a blurred image B can be conveniently expressed by

$$B = K \otimes S + n, \quad (1)$$

where K , S , and n correspond to an unknown blur kernel, latent sharp image, and noise, respectively. As only B is available, blind image deblurring methods aim to estimate the latent sharp image S and blur kernel K simultaneously. This problem is ill-posed, since different pairings of S and K may result in the same B . To make blind image deblurring well-posed, early research relies on heuristics, image statistics and assumptions on the sources of the blur, and mainly focuses on removing uniform and spatially invariant

blur caused by simple translational or rotational camera motions, called uniform blur [11, 18]. For uniform blur, \mathbf{K} is the unknown blur kernel that acts globally on \mathbf{S} to generate a blurry image. In view of the success of the iterative approach [5], most methods improve the estimation of the blur kernel and latent sharp image for each iteration by using either the parametric prior model or the assumptions of local linearity of the blur function [2, 5, 21, 31]. However, run time, stop criteria and accuracy estimation remain problems of these algorithms. Although some researchers started to apply deep-net models, such as convolutional neural networks (CNN), to improve the estimation of blur kernels and/or latent sharp images, these models are mostly designed to handle uniform blurs in early stages [3, 23]. Blurring artifacts are generally not uniformly distributed in a captured image. The observation prompts recent studies [11, 18, 22, 27] to explore the non-uniform blur problem, and develop end-to-end deblurring approaches to deal with the non-uniform blur effect caused by depth variation, camera shake and object motion in dynamic scenes. For non-uniform blur, \mathbf{K} typically represents a large sparse matrix where each row contains a local unknown blur kernel acting on \mathbf{S} to generate a corresponding blurred image. It has been shown in [18, 27] that the blur kernel-free end-to-end multi-scale deep-net models are capable of dealing with non-uniform blind image deblurring. Kupyn et al. [11] and Ramakrishnan et al. [22] further construct the conditional generative adversarial network-based (CGAN-based) approaches that solve the non-uniform deblurring issue and achieve impressive performance improvements. Although the introduction of adversarial loss enables learning a deep-net model to generate a sharper image distribution, it still has the difficulty in satisfactorily characterizing local details of image distribution. Such insufficiency in modeling will eventually lead to poor representations on challenging foreground/background areas and degrade the final deblurring effect. Therefore, a new and effective method, that can additionally refine the local areas/details into a sharp image, is preferable.

To address the aforementioned issue, we propose a novel adversarial Blurred Region Mining and Refining (adversarial BRMR) approach that can drive CGANs to learn real sharp image distribution globally and locally for blind image deblurring. The proposed adversarial BRMR approach can be viewed as establishing a couple of collaborators, trying to challenge the full image-scale generator network to discover evidence of still blurred regions and use the patch-scale generator network to refine such regions until no supportable evidence is left. To be more specific, we first train an image-scale CGAN, denoted as iGAN, and a patch-scale CGAN, denoted as pGAN. The former learns from pairs of blurred images and ground truth sharp images, while the latter learns from pairs of blurred patches (i.e., mined from the discriminator of iGAN) and sharp patches. For testing, the discriminator of iGAN is applied to localize the most blurred region within a deblurred image generated by the generator for inferring the blur class. The localized blur region is then refined by using the generator of pGAN to improve the performance of the image-scale generator network. To continuously enhance the local details, the discriminator of iGAN is driven to discover another blurred region for better generating a local patch. With such repetitive adversarial BRMR operations, our network can mine and refine blur regions progressively until no further improvement, i.e., using the discriminator of iGAN to check

whether the confidence score of the updated deblurred image is increased in the sharp class. The process is illustrated in Figure 2, in which a blue patch represents the localized most blurred region within a deblurred image generated from the image-scale generator network. After accepting the updated deblurred image with the newly refined region (green patch) by using the patch-scale generator network and the discriminator of iGAN, another blur region (refer to as the red patch) is mined. Repeating such adversarial BRMR approach gradually improves the quality of the deblurred image. We summarize our main contributions as follows:

- This study proposes a novel adversarial BRMR approach that couples two complementary CGANs, enabling the progressive mining and refining of the blurred regions to improve image sharpness.
- We collect a new concert-related dataset and would make it publicly available. The related research communities can access the data and develop learning techniques for sharpening blurry images under different shot types defined in the language of film [1, 17].
- Our method achieves state-of-the-art performance on both the challenging low-light concert-related dataset and non-uniform dynamic scene blurry benchmark [18].

2 RELATED WORK

Image deblurring has been studied extensively in the literature of image processing and computer vision, e.g., [5, 11, 18, 23]. The following discussion covers blind image deblurring approaches from uniform blur to non-uniform blur and recent deep learning techniques.

Since blind image deblurring is an ill-posed problem, early methods require certain assumptions or prior knowledge to constrain the solution space, and focus on dealing with uniform blur. For example, Fergus et al. [5] showed that photographs of natural scenes typically follow a heavy-tailed distribution of image gradients, and used a mixture-of-Gaussians model to learn the statistical prior for deblurring. Similarly, Shan et al. [24] adopted a parametric model to approximate the heavy-tailed prior for natural images. Levin et al. [13] pointed out the limitation of the maximum *a posteriori* deblurring method with a sparse prior, and presented a variational Bayesian approach to avoid such limitation. As the variational Bayesian approach [13] is computationally expensive, an efficient algorithm for approximation of marginal likelihood has been developed for image deblurring [14]. Pan et al. [19] developed a new natural image prior based on the dark channel prior [6] for blur kernel estimation. To restore images regularized by the dark channel prior, they developed an effective optimization algorithm based on a half-quadratic splitting strategy and look-up tables. Following the success of deep learning, researchers have adopted deep-net models such as CNN to improve the accuracy of both blur kernel and latent sharp image estimation over the past few years. Since no pairs of real blurry image and ground truth sharp image are available for learning, they generally adopted blurry images generated by convolving synthetic uniform blur kernels. For example, Schuler et al. [23] incorporated a sharpening CNN into an iterative blind deconvolution method to estimate a blur kernel. To obtain pairs of sharp and blurry images for network training, they generate

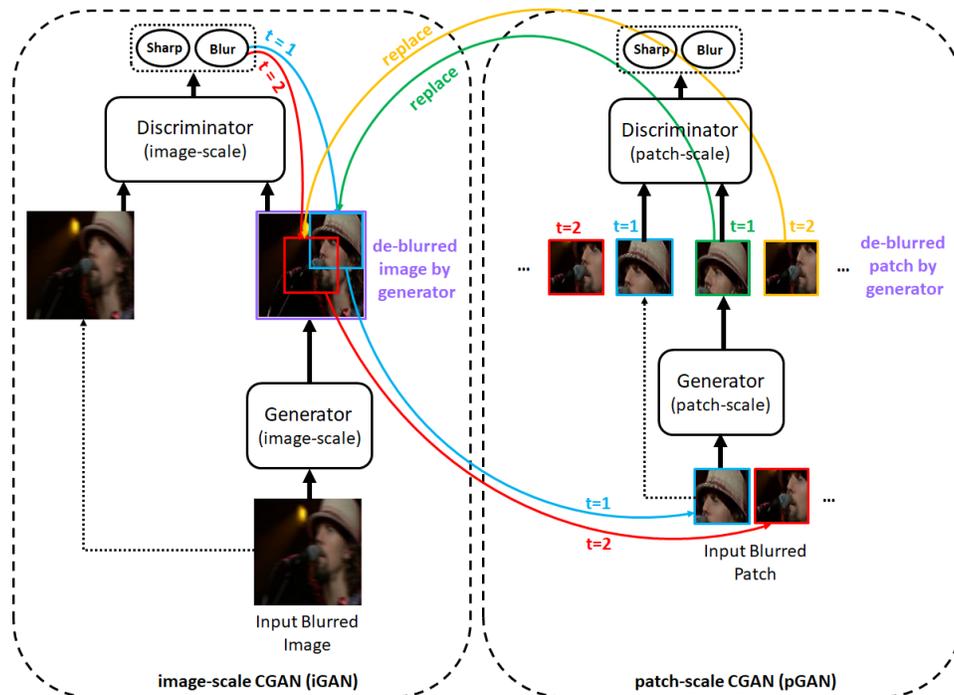


Figure 2: The proposed adversarial blurred region mining and refining (adversarial BRMR) functions as a bridge between the two (collaborative) conditional GANs: iGAN and pGAN. Here t indexes the t th local patch refinement under the adversarial BRMR iterative operation.

uniform blur kernels using a Gaussian process, and synthesize a great deal of blurry images by convolving them to the sharp images collected from the ImageNet dataset [4]. Through a similar data generation process, Chakrabarti [3] learned neural network to predict complex Fourier coefficients of motion kernel for executing deblurring in Fourier space. Note that all these methods require a blur kernel estimation step to restoring the latent sharp image. However, these blur kernel approximations may be inaccurate, i.e., blur is far more complex than using the simplified assumptions/models in blur kernel estimation, causing difficulties in constraining the solution space.

For non-uniform blind image deblurring, the research trend has recently shifted from the traditional iterative approach [19, 25] to a kernel-free deblurring approach based on end-to-end trainable deep-net models [11, 18, 22, 27]. For example, Nah et al. [18] proposed a multi-scale CNN that restores latent sharp images from a coarse-to-fine architecture in an end-to-end manner. Such multi-scale CNN learns directly from paired sharp images and non-uniform blurred images (i.e., generated by averaging sharp images in dynamic scenes) without collaborating any blur kernel estimation steps. Tao et al. [27] proposed a scale-recurrent network based on coarse-to-fine scheme, which gradually restores the latent sharp image at different resolutions in an end-to-end manner. Ramakrishnan et al. [22] used the CGAN architecture integrated with the global skip connection and the densely connected CNN [7] to learn a kernel-free deblurring network in an end-to-end manner. Similarly, Kupyn et al. [11] proposed a DeblurGAN that combines adversarial loss and content loss into the CGAN architecture to perform blind image deblurring. Although multi-scale deep-net models [18, 27]

and CGAN-based approaches [11, 22] have performed well recently, their deblurring frameworks still struggle to correctly match some local details of the sharp image distribution, causing the local foreground/background area to have poor quality.

3 ADVERSARIAL BRMR TO IMAGE DEBLURRING

To address the insufficiency that current deblurring frameworks lack the ability to satisfactorily recover local sharp image distributions, we propose an adversarial BRMR approach for automatically and progressively localizing and sharpening the blurred regions. The adversarial BRMR involves two processes, namely blur region mining with the image-scale CGAN (iGAN), and blur region refinement with the patch-scale CGAN (pGAN).

iGAN The iGAN model for blur region mining mainly involves two operations: generating a full-scale deblurred image, and localizing the regions that are still blurred (as shown in Figure 3). Motivated by pix2pix [8] and *classification activation maps* (CAM) [32], we construct iGAN based on the pix2pix architecture with the following modifications in the discriminator D_i . Specifically, the CAM concept is incorporated into D_i by considering the task as two-class image classification for sharp and blur. Now let the feature maps yielded at *conv4* of D_i be $\{f_k\}$ and the weights of the succeeding fully-connected (FC) layer be $\mathbf{w}^s = \{w_k^s\}$ and $\mathbf{w}^b = \{w_k^b\}$. Analogous to [32], the global average pooling (GAP) is applied to each *conv4* feature map f_k for resulting the representation F_k , where F_k is $\sum_{x,y} f_k(x,y)$, and (x,y) indicates the spatial location of f_k . Then, through forward propagation, the representation $\{F_k\}$ is passed

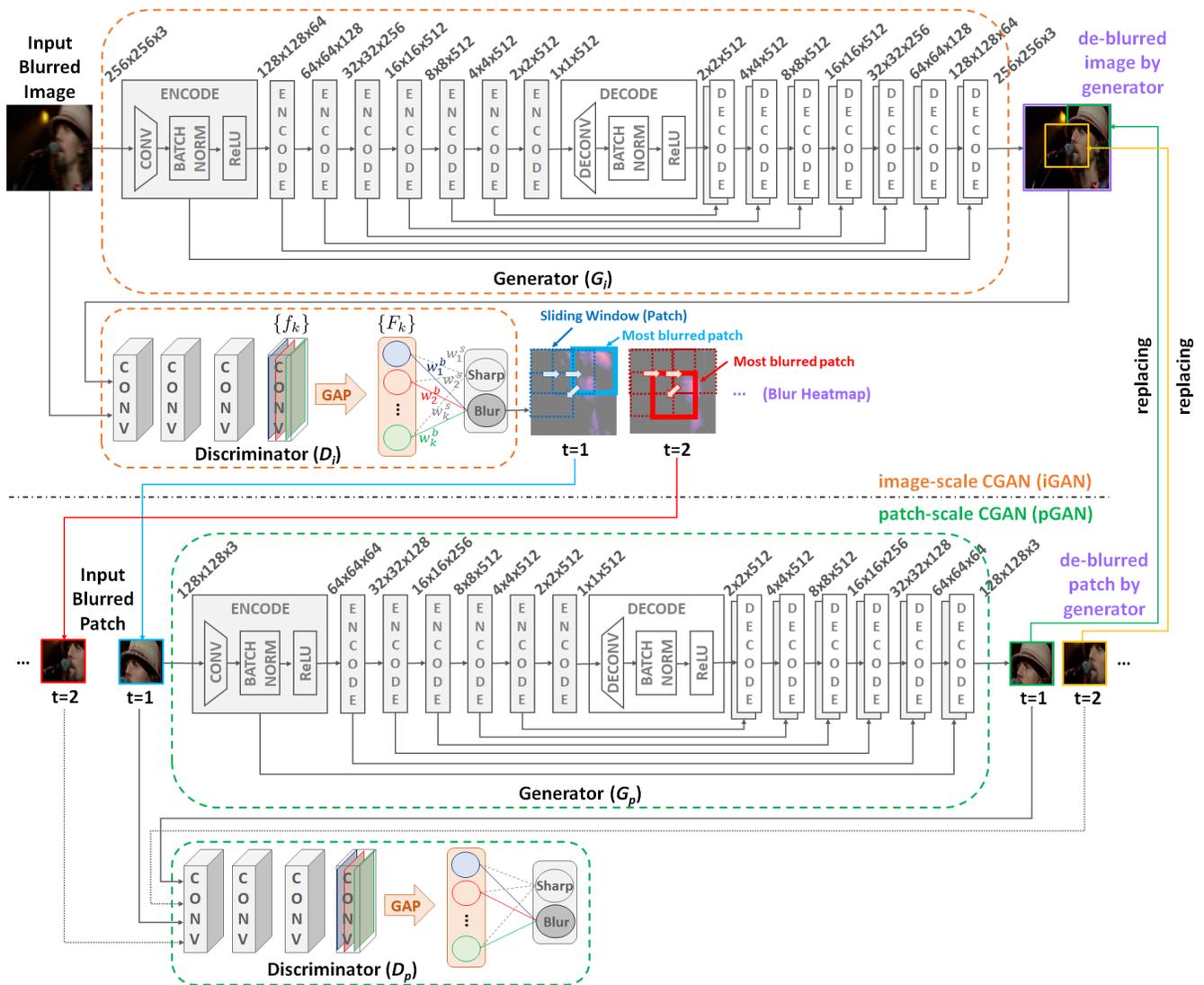


Figure 3: Architecture of the coupled CGANs with the proposed adversarial BRMR. We first train the image-scale CGAN (iGAN) to generate a deblurred image and a couple of response heatmaps for blur and sharp, respectively. Applying the sliding window (so-called patch) to the blur heatmap is used to detect the most blurred region. The patch-scale CGAN (pGAN) is then trained to refine and replace the mined blur region. After refining, the discriminator of iGAN is driven to discover another blurred region for enhancing sharpness progressively. The refined regions from such an iterative processes together constitute the final deblurred image as output.

through the FC layer (w^s, w^b) to predict the confidence of the sharp and blur classes by implementing a two-class classification. Essentially, w_k^s and w_k^b indicate the importance of F_k for predicting sharp and blur classes, respectively. Accordingly, by multiplying the weights w_k^s and w_k^b back to the $f_k(x, y)$, the activation at spatial grid (x, y) to sharp and blur classes can be revealed by the two corresponding heatmaps $H^s(x, y)$ and $H^b(x, y)$ where

$$\begin{aligned}
 H^s(x, y) &= \sum_k w_k^s \cdot f_k(x, y) \\
 \text{and } H^b(x, y) &= \sum_k w_k^b \cdot f_k(x, y).
 \end{aligned}
 \tag{2}$$

By simply upsampling the heatmap to the size of the input image, we can identify the image regions most relevant to the particular (sharp or blur) category [32]. Given the generated deblurred output by iGAN, the regions that are still blurred can be readily revealed in the heatmap H^b and thus further improvements can be made via the heatmap CGAN, i.e., pGAN. We now summarize the two operations of iGAN. In the first operation, the generator G_i of iGAN, trained by using a combination of adversarial and $L1$ losses (refer to Subsection 3.1), is applied to globally perform image deblurring. In the second operation, the discriminator D_i is used to output the heatmap H^b for the blur class. The most blurred

Algorithm 1: Image Deblurring with adversarial BRMR**** Training:**

Blurred images $\mathbf{B} = \{B_j\}_{j=1}^N$.
 Sharp images (ground truth) $\mathbf{S} = \{S_j\}_{j=1}^N$.
 Blurred patches $\mathbf{p}^b = \{p_j^b\}_{j=1}^M$.
 Sharp patches (ground truth) $\mathbf{p}^s = \{p_j^s\}_{j=1}^M$.
 (G_i, D_i): The generator and discriminator of iGAN.
 (G_p, D_p): The generator and discriminator of pGAN.
 Train iGAN (G_i, D_i) with \mathbf{B} and \mathbf{S} .
 Mine blurred patches \mathbf{p}^b using D_i through inside testing.
 Extract the corresponding patches \mathbf{p}^s from \mathbf{S} .
 Train pGAN (G_p, D_p) with \mathbf{p}^b and \mathbf{p}^s .

**** Testing:**

Input : A testing blurred image B .
Initialize: Deblurred image $D = \emptyset$,
 $t = 1$, indexing the t th local patch refinement,
 current confidence for sharp class $\delta = 0$,
 updated confidence for sharp class $\tilde{\delta} = 0$.

Output : A deblurred image D .

- 1 Generate a deblurred image D from $G_i(B)$.
 - 2 Compute blur heatmap H_{t-1}^b and δ via $D_i(B, D)$.
 - 3 **repeat**
 - 4 Mine the most blurred patch p_t from H_{t-1}^b .
 - 5 Generate a deblurred patch \tilde{p}_t via $G_p(p_t)$.
 - 6 $\tilde{D} \leftarrow$ Replace p_t in D with \tilde{p}_t .
 - 7 Compute blur heatmap H_t^b and $\tilde{\delta}$ via $D_i(B, \tilde{D})$.
 - 8 **if** $\tilde{\delta} > \delta$ **then**
 - 9 $D \leftarrow \tilde{D}$, $\delta \leftarrow \tilde{\delta}$, $t = t + 1$.
- until** patch refinement is not good and $\tilde{\delta} \leq \delta$

region is mined by comparing the heatmap response of the sliding window (so-called patch), where patch size is set to 128×128 (achieved best performance) and shifted 64 pixels away to overlap 50%, sequentially from left to right and from top to bottom.

pGAN For blur region refinement, the pGAN is trained to locally restore the sharpness of a particular blur region singled out in the mining operation. Similar to the iGAN architecture, pGAN is established based on the pix2pix architecture and trained from pairs of blurred image patches and corresponding sharp patches. Those blurred patches are mined by driving the discriminator D_i of iGAN through inside testing during the training stage. In performing blur region refinement for the globally-generated deblurred image by iGAN, pixel values within the mined blur region are replaced with those generated from the patch-scale generator G_p of pGAN. The updated deblurred image with the newly refined region is then fed back to the discriminator D_i of iGAN to verify if the refinement results in increase in the confidence of predicting the updated version as sharp. Once a blurred region is sharpened and the verification evidence is supported, the discriminator D_i is driven to discover another blurred region to be enhanced. This process is repeated until

no more supportable evidence can be detected. This collaborative mechanism allows the adversarial BRMR approach to drive the two coupled CGANs to restore a latent sharp image distribution globally and locally for blind image deblurring. Algorithm 1 summarizes the procedure of the adversarial BRMR approach.

3.1 Objective

The objective of both iGAN and pGAN can be expressed as

$$\mathcal{L}_{CGAN}(G, D) = \mathbb{E}_{b,s} [\log D(b, s)] + \mathbb{E}_{b,z} [\log(1 - D(b, G(b, z)))], \quad (3)$$

where G and D represent the generator and the discriminator (classifier) in both iGAN and pGAN. b , s , and z denote the blurred image (or patch), sharp image (or patch), and random noise, respectively. To avoid learning deterministic mapping from b to s , similar to the pix2pix architecture [8], z is provided in the form of dropout, which is applied to several layers of the G in the training and testing phases. During training, G tries to minimize this objective against an adversarial D that tries to maximize it, i.e., $G^* = \arg \min_G \max_D \mathcal{L}_{CGAN}(G, D)$.

In addition, previous researches have demonstrated that it is beneficial to mix GAN objective (adversarial loss) with a traditional loss, such as $L1$ or $L2$ distance [8, 20]. The discriminator's job remains unchanged, but the generator is tasked not only to fool the discriminator but also to be near the ground truth output in an $L1$ or $L2$ sense. Accordingly, we also consider this option, using $L1$ loss rather than $L2$ as $L1$ encourages less blurring:

$$\mathcal{L}_{L1}(G) = \mathbb{E}_{b,s,z} [\|s - G(b, z)\|_1]. \quad (4)$$

Thus, our final objective is

$$G^* = \arg \min_G \max_D \mathcal{L}_{CGAN}(G, D) + \lambda \mathcal{L}_{L1}(G), \quad (5)$$

where λ is set to 100 for both iGAN and pGAN in all experiments.

4 EXPERIMENTS

To demonstrate the effectiveness of the adversarial BRMR approach, experiments were conducted on two challenging datasets, namely the proposed concert-related dataset and the GOPRO dataset [18].

4.1 Concert-Related Dataset

Although researchers have devoted considerable effort to image deblurring [3, 10, 12, 13, 18, 23, 26], the collected datasets do not consider the shot types defined in film terminology [1, 17], thus ignoring the details of deblurring under different shots. In the language of film [1, 17], a type of shot is defined as how much a target subject and its surrounding area can be seen. Totally, six types of shots are defined, as described in Table 1. Take, for example, that the two images in Figure 4 are obtained with the medium close-up shot and extreme long shot, respectively. One can easily see that the left image emphasizes the singer Bruno Mars from shoulders to the top of head, and the right image instead highlights the small body of singer in vista. Such examples suggest that the method of creating an image deblurring should take into account the properties of the shot. For example, when performing image deblurring on close up and medium close-up shots, the model should focus on maintaining the details (e.g., detail texture) of the target subject,

Table 1: The definition of six types of shots [1, 17].

Types of Shots	Description	Example
Close-Up (CU)	A Close-Up is used to show emotion on the subject’s face. That is, the face occupies most of the screen (image) .	
Medium Close-Up (MCU)	A Medium Close-Up contains a subject’s head and shoulders completely.	
Medium Shot (MS)	A Medium Shot contains a subject from the waist to the top of the head .	
Medium Long Shot (MLS)	A Medium Long Shot would contain a subject from his/her knees to the top of the head .	
Long Shot (LS)	A Long Shot would contain a subject’s entire body from the top of the head to the bottom of the feet .	
Extreme Long Shot (XLS)	An Extreme Long Shot covers a large area or landscape . It would be hard to see any reactions/emotion from people in the shot since they are too far away.	



Figure 4: Two images from an official concert video of the song “When I was Your Man” by Bruno Mars live at BBC Radio 2013. The left image is yielded by a medium close-up shot, and the right by an extreme long shot.

while long and extreme long shots should instead strive to maintain the main structure of the target because the size of the target is too small. Therefore, learning how to sharpen recordings under different shots is crucial to guaranteeing a high-quality visual storytelling process [1, 16, 17, 29, 30]. Besides the six shot types listed in Table 1, this study also considers two additional variants that focus on either the audience shot (ADS) or musical instrument shot (MIS) to enrich the shot variations in a concert video.

To this end, a dataset was created from a set of official concert videos downloaded from YouTube. The dataset comprises 37 official video clips collected from 30 live concerts, with a total of 440 annotated sharp images. Each image was specified with an appropriate film shot type [1, 17]. To obtain pairs of sharp and blurry images, seven uniform blur kernels, namely horizontal, vertical, averaging, Gaussian, median, bilateral and resizing, were first generated using

Table 2: Quantitative deblurring performance comparison on the concert-related dataset.

Metric	multi-scale CNN [18]	DeblurGAN [11]	iGAN	adversarial BRMR
Avg. PSNR	29.90 dB	28.57 dB	30.77 dB	31.93 dB

the OpenCV Python Program library¹. The blurry images were then synthesized by convolving the blur kernels to the sharp images collected from our dataset. Finally, the dataset was composed of 3,080 pairs of blurry and sharp images at 256×256 resolution. Among the 3,080 pairs, 2,800 were adopted for training, and the remaining 280 pairs were adopted for testing. The dataset links and the detailed information such as blur kernel setting are provided in <https://sites.google.com/site/adversarialbrmr/blur-kernel-script>.

The performance of the proposed adversarial BRMR approach was compared with two state-of-the-art approaches for blind image deblurring, namely multi-scale CNN [18] and DeblurGAN [11]. For training, both iGAN and pGAN in the adversarial BRMR approach were trained by using mini-batch stochastic gradient descent (SGD) and applying the Adam solver [9]. The batch size and learning rate were set to 1 and 2×10^{-4} , respectively. Multi-scale CNN [18] was also trained by using the Adam solver. The batch size is set to 4, and the learning rate is initialized with 5×10^{-5} and updated with a linear decay. Regarding DeblurGAN, it was trained with the Adam solver, initialized with a learning rate of 1×10^{-4} and updated with a linear decay [11]. The batch size was set to 1. Techniques such as batch normalization, data augmentation, and dropout were used for the mentioned approaches to alleviate overfitting. When the batch size is set to 1, batch normalization has been termed “instance normalization” and has proven to be effective in image generation tasks [8, 28]. In the experiments, the standard peak signal-to-noise ratio (PSNR) metric was used for performance evaluation. The average PSNR over the testing set was reported.

The experimental results in Table 2 and Figure 5 demonstrate that the proposed adversarial BRMR approach not only shows superior results in terms of PSNR than multi-scale CNN and DeblurGAN, but also provides better visual perception. We believe that it is because multi-scale CNN and DeblurGAN have difficulty in correctly matching some local details of sharp image distribution, resulting in poor local foreground/background area quality. Further, DeblurGAN [11] has integrated the PatchGAN architecture [8, 15] in the discriminator network for alleviating the effects due to mismatched local details, but the improvements are moderate. Using the collaborative mechanism of the two CGANs, the proposed adversarial BRMR approach can indeed learn the real sharp image distribution globally and locally. Overall, the adversarial BRMR approach increased the PSNR by approximately 2.03dB and 3.4dB, respectively, compared to multi-scale CNN and DeblurGAN. In addition, the results also indicate that the adversarial BRMR approach is superior to the iGAN in both qualitative and quantitative ways. As shown in Figure 5, the results demonstrate that without collaborate with pGAN by using adversarial BRMR approach, the iGAN has difficulty in resulting the sharp local details. Again, it confirms the importance of modeling the local details of sharp image distribution.

¹https://docs.opencv.org/3.1.0/d4/d13/tutorial_py_filtering.html

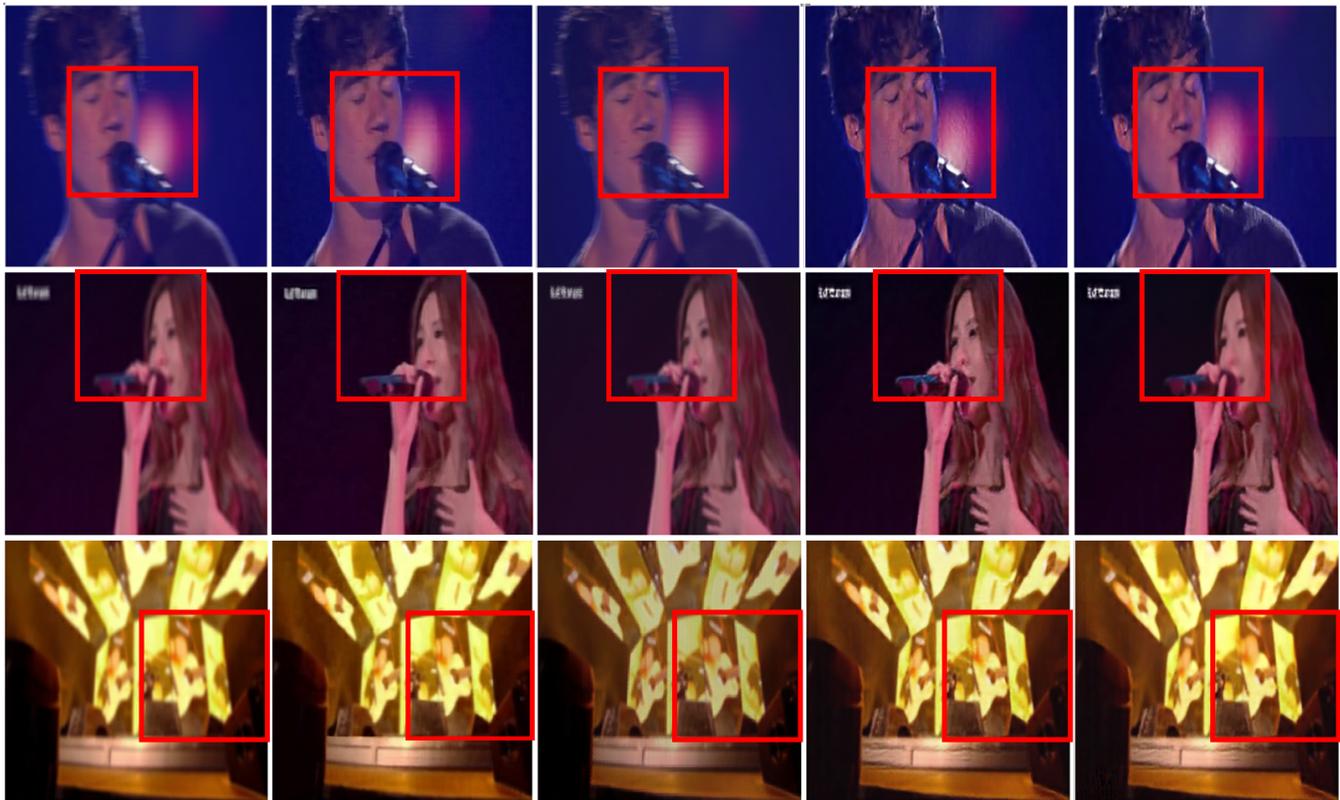


Figure 5: Test results on the concert-related dataset. From left to right: blurry image, result of multi-scale CNN [18], result of DeblurGAN [11], result of iGAN, and result of our adversarial BRMR approach. From top to down: close-up with horizontal blur, and medium close-up and extreme long shot with Gaussian blur. Red bounding box highlights the deblurring result.

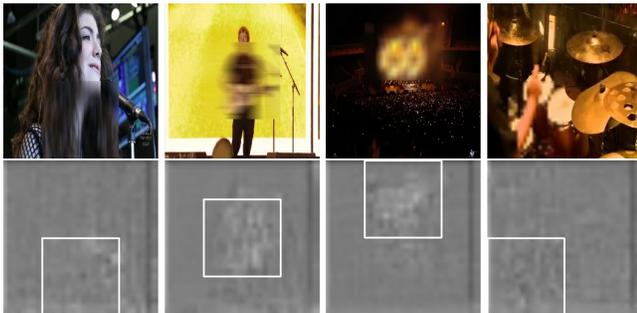


Figure 6: The results of blurred region detection. From top to down: partially blurred images and the corresponding blur region detection results. From left to right represent different types of blur kernel: vertical, horizontal, averaging, and resizing that respect to medium close up shot, long shot, extreme long shot, and musical instrument shot.

To demonstrate whether the iGAN discriminator D_i in the adversarial BRMR approach can effectively localize a blurred region, we further carry out the experiment of the blurred region detection. For experimental setup, a test image was generated by convolving a blur kernel to a single patch of the sharp image. The patch size is set to 128×128 and the sharp image comes from above mentioned 280 pairs of testing data. Both the location of the patch and the type of

the blur kernel (from mentioned seven types) are randomly chosen. Finally, a total of 280 partially blurred images were generated for testing. In the experiments, the D_i was used to detect the blurred patch of the partially blurred image. The intersection over union (IoU) metric was used for performance evaluation. The IoU score, which is defined as $\frac{|P \cap Q|}{|P \cup Q|}$, where P denotes the detected blurred patch and Q denotes the ground-truth blurred patch. The mean IoU over the testing set was reported.

The quantitative result first indicates that our D_i reached a mean IoU score of 0.676. Such a result confirms that the blurred patch detected by D_i is highly overlapping with the blurred patch of the ground-truth. Some qualitative results are further shown in Figure 6. The results show that the D_i can effectively localize a blurred region regardless of the location of the patch and the type of the blur kernel. Both quantitative and qualitative results support the D_i , which is positive for adversarial BRMR approach to achieve the blur region mining process. For more experiments, we also provide a video demo at <https://sites.google.com/site/adversarialbrmr/demo>.

4.2 GOPRO Dataset

To further verify the effectiveness of the proposed adversarial BRMR approach, a challenging benchmark dataset GOPRO [18] was adopted for performance evaluation. The GOPRO dataset is composed of 3,214 pairs of non-uniform blurry images and sharp



Figure 7: Test results on the GOPRO dataset. From left to right: blurry image, result of DeblurGAN [11], result of our adversarial BRMR approach, and ground truth sharp image. Red bounding box highlights the deburring result.

Table 3: Quantitative deburring performance comparison on the GOPRO dataset. Run time indicates the average execution time of deburring a single image.

Metric	multi-scale CNN [18]	DeblurGAN [11]	adversarial BRMR
Avg. PSNR	29.08 dB	27.94 dB	30.27 dB
Run Time	4.22 s	0.91 s	2.16 s

images. Among the 3,214 pairs, 2,103 pairs were set for training, and the remaining 1,111 pairs were set for the test [18]. Similar to the experimental setup in the concert-related dataset, the parameter settings of multi-scale CNN, DeblurGAN, and our adversarial BRMR approach were optimized in the training stage. In the experiments, quantitative (average PSNR and run time) and qualitative results of the testing set were reported.

The results in Table 3 first indicate that even in the challenging GOPRO dataset, the proposed adversarial BRMR approach has a better average PSNR performance than multi-scale CNN and DeblurGAN. The qualitative results shown in Figure 7 further demonstrate that compared with DeblurGAN [11], the latent sharp image generated by the adversarial BRMR approach is notably close to that of the ground truth sharp image. The findings are in accordance with the analytical results of the concert-related dataset. That is, modeling the local details of sharp image distribution can indeed help sharpen the local foreground/background area, thereby enhancing subjects’ viewing experiences. Despite the challenging GOPRO dataset, the adversarial BRMR approach still increased the average PSNR by 1.19dB and 2.33dB compared to multi-scale CNN and DeblurGAN, and achieved an average PSNR of 30dB. We further

show the run time of each approach on a single NVIDIA GTX 1080 GPU. The results demonstrate that even though the adversarial BRMR couples two CGANs, it achieves the run time comparable to the other techniques. This is because for deburring a single image, our approach only needs to perform an average of 2.3 iterations of adversarial BRMR, so it does not consume a lot of run time. Compared with the state-of-the-art approaches for image deburring, the proposed adversarial BRMR technique has shown promising results both quantitatively and qualitatively.

5 CONCLUSION

This study introduces a novel adversarial blurred region mining and refining (adversarial BRMR) approach that effectively integrates two conditional generative adversarial networks (CGANs) to automatically and progressively localize and sharpen the blurred regions for blind image deburring. Experiments on both uniform and non-uniform blurry datasets demonstrate that the adversarial BRMR approach outperforms the state-of-the-art approaches, and generates satisfactory blind image deburring results. Based on these promising outcomes, our future work in this field will focus on expanding the adversarial BRMR approach from image deburring to video deburring, which would require incorporating temporal information into the CAM-motivated collaborative mechanism of the two coupled CGANs.

ACKNOWLEDGMENTS

This work was supported in part by MOST grants 108-2634-F-001-003, 108-2634-F-001-007 and 107-2218-E-155-011-MY2.

REFERENCES

- [1] Dale Andrews. 2011. *Digital overdrive: Communications & multimedia technology*. Digital Overdrive.
- [2] S. Derin Babacan, Rafael Molina, Minh N. Do, and Aggelos K. Katsaggelos. 2012. Bayesian blind deconvolution with general sparse image priors. In *Proc. ECCV*. 341–355.
- [3] Ayan Chakrabarti. 2016. A neural approach to blind motion deblurring. In *Proc. ECCV*. 221–235.
- [4] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. ImageNet: A large-scale hierarchical image database. In *Proc. CVPR*. 248–255.
- [5] Rob Fergus, Barun Singh, Aaron Hertzmann, Sam T. Roweis, and William T. Freeman. 2006. Removing camera shake from a single photograph. *ACM Transactions on Graphics* 25, 3 (2006), 787–794.
- [6] Kaiming He, Jian Sun, and Xiaoou Tang. 2009. Single image haze removal using dark channel prior. In *Proc. CVPR*. 1956–1963.
- [7] Gao Huang, Zhuang Liu, Laurens van der Maaten, and Kilian Q. Weinberger. 2017. Densely connected convolutional networks. In *Proc. CVPR*. 2261–2269.
- [8] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. 2017. Image-to-image translation with conditional adversarial networks. In *Proc. CVPR*. 5967–5976.
- [9] Diederik P. Kingma and Jimmy Lei Ba. 2015. Adam: A method for stochastic optimization. In *Proc. ICLR*.
- [10] Rolf Köhler, Michael Hirsch, Betty Mohler, Bernhard Schölkopf, and Stefan Harmeling. 2012. Recording and playback of camera shake: Benchmarking blind deconvolution with a real-world database. In *Proc. ECCV*. 27–40.
- [11] Orest Kupyn, Volodymyr Budzan, Mykola Mykhailych, Dmytro Mishkin, and Jiri Matas. 2018. DeblurGAN: Blind motion deblurring using conditional adversarial networks. In *Proc. CVPR*.
- [12] Wei-Sheng Lai, Jia-Bin Huang, Zhe Hu, Narendra Ahuja, and Ming-Hsuan Yang. 2016. A comparative study for single image blind deblurring. In *Proc. CVPR*. 1701–1709.
- [13] Anat Levin, Yair Weiss, Fredo Durand, and William T. Freeman. 2009. Understanding and evaluating blind deconvolution algorithms. In *Proc. CVPR*. 1964–1971.
- [14] Anat Levin, Yair Weiss, Fredo Durand, and William T. Freeman. 2011. Efficient marginal likelihood optimization in blind deconvolution. In *Proc. CVPR*. 2657–2664.
- [15] Chuan Li and Michael Wand. 2016. Precomputed real-time texture synthesis with Markovian generative adversarial networks. In *Proc. ECCV*. 702–716.
- [16] Jen-Chun Lin, Wen-Li Wei, Tyng-Luh Liu, Yi-Hsuan Yang, Hsin-Min Wang, Hsiao-Rong Tyan, and Hong-Yuan Mark Liao. 2018. Coherent deep-net fusion to classify shots in concert videos. *IEEE Transactions on Multimedia* 20, 11 (2018), 3123–3136.
- [17] Gustavo Mercado. 2010. *The filmmaker's eye: Learning (and breaking) the rules of cinematic composition*. Taylor & Francis.
- [18] Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. 2017. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *Proc. CVPR*. 3883–3891.
- [19] Jinshan Pan, Deqing Sun, Hanspeter Pfister, and Ming-Hsuan Yang. 2016. Blind image deblurring using dark channel prior. In *Proc. CVPR*. 1628–1636.
- [20] Deepak Pathak, Philipp Krähenbühl, Jeff Donahue, Trevor Darrell, and Alexei A. Efros. 2016. Context encoders: Feature learning by inpainting. In *Proc. CVPR*. 2536–2544.
- [21] Daniele Perrone and Paolo Favaro. 2014. Total variation blind deconvolution: The devil is in the details. In *Proc. CVPR*. 2909–2916.
- [22] Sainandan Ramakrishnan, Shubham Pachori, Aalok Gangopadhyay, and Shanmuganathan Raman. 2017. Deep generative filter for motion deblurring. In *Proc. ICCVW*. 2993–3000.
- [23] Christian J. Schuler, Michael Hirsch, Stefan Harmeling, and Bernhard Schölkopf. 2016. Learning to deblur. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 38, 7 (2016), 1439–1451.
- [24] Qi Shan, Jiaya Jia, and Aseem Agarwala. 2008. High-quality motion deblurring from a single image. *ACM Transactions on Graphics* 27, 3 (2008), 73:1–10.
- [25] Jian Sun, Wenfei Cao, Zongben Xu, and Jean Ponce. 2015. Learning a convolutional neural network for non-uniform motion blur removal. In *Proc. CVPR*. 769–777.
- [26] Libin Sun, Sunghyun Cho, Jue Wang, and James Hays. 2013. Edge-based blur kernel estimation using patch priors. In *Proc. ICCP*. 1–8.
- [27] Xin Tao, Hongyun Gao, Xiaoyong Shen, Jue Wang, and Jiaya Jia. 2018. Scale-recurrent network for deep image deblurring. In *Proc. CVPR*. 8174–8182.
- [28] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. 2016. Instance normalization: The missing ingredient for fast stylization. *arXiv* (2016).
- [29] Wen-Li Wei, Jen-Chun Lin, Tyng-Luh Liu, Yi-Hsuan Yang, Hsin-Min Wang, Hsiao-Rong Tyan, and Hong-Yuan Mark Liao. 2017. Deep-net fusion to classify shots in concert videos. In *Proc. ICASSP*. 1383–1387.
- [30] Wen-Li Wei, Jen-Chun Lin, Tyng-Luh Liu, Yi-Hsuan Yang, Hsin-Min Wang, Hsiao-Rong Tyan, and Hong-Yuan Mark Liao. 2018. Seethevoice: Learning from music to visual storytelling of shots. In *Proc. ICME*.
- [31] Li Xu, Shicheng Zheng, and Jiaya Jia. 2013. Unnatural L0 sparse representation for natural image deblurring. In *Proc. CVPR*. 1107–1114.
- [32] Bolei Zhou, Aditya Khosla, Agata Lapedriza, Aude Oliva, and Antonio Torralba. 2016. Learning deep features for discriminative localization. In *Proc. CVPR*. 2921–2929.