



Face Image Deblurring: A Data-Driven Strategy

Abdelwahed Nahli[†], Yuanzhouhan Cao[§], Shugong Xu^{*}
Shanghai University, Shanghai, China

^{*}E-mail: shugong@shu.edu.cn, [§]E-mail: yuanzhouhan.cao@gmail.com

[†]E-mail: nahliabdelwahed@gmail.com Tel: +86-15-604302529.

Abstract

Nowadays remarkable progress has been observed in facial detection as a core part of computer vision. Nevertheless, motion blur still presents substantial challenges in face detection. The most recent face image deblurring methods make oversimplifying presumption and fail to restore the highly structured face shape/identity information. Therefore, we propose a data-driven based face image deblurring approach that foster facial detection and identity preservation. The proposed model includes two sequential data streams: Out of any supervision the first has been trained on real unlabeled clear/blurred data to generate a close realistic blurred image data during its inference. On the other hand, the generated labeled data has been exploited with by a second supervised learning-based data steam to learn the mapping function from blur domain to the clear one. We utilize the restored data to conduct an experimentation on face detection task. The experimental evaluation demonstrates the outperformance of our results and supports our system design and training strategy.

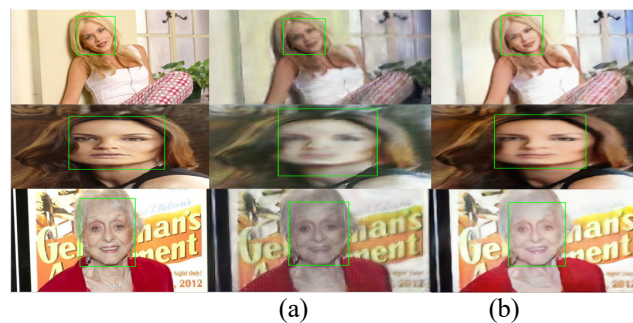


Figure 1. FaceBoxes [16] Face detection approach performance on our face deblurring results a, b and c respectively show the ground truth, blurred and restored version.

1. INTRODUCTION

Image deblurring task aims to restore a single clear image from a single blurred image. Classical methods model the blur process as the convolution operation between a latent clear image and blur kernels with an additive noise, and utilize the maximum a posteriori (MAP) framework to formulate the image blurring phenomena. As inverse problems are often ill-posed, in order to constrain the solution space, the state-of-the-art approaches adopt on natural image priors with some presumptions (e.g., dark channel prior [1] and L0 gradient [2]). The existing image priors are effective for deblurring some limited range of natural images, but the underlying assumptions may not hold for images from specific categories, for instance, text and face. Therefore, several approaches use task-specific priors, such as light streaks [3] for extremely low-light images and L0 intensity [4] for text images. Motivated by the fact that face images typically have fewer edges and textures for estimating blur kernels, another face image database searching technic was proposed by Pan et al. [5] to find a match similar face exemplar and extract the contour as reference edges. However, this approach is limited by the diversity of face images in the wild as a similar reference image might be not always exist in the used face image databases. Furthermore, the MAP framework-based methods typically require heavy computational power due to the reiterated estimations of blur kernels and latent images. In this work, we focus on face image deblurring task.

We propose an efficient and effective deep learning-based approach using an end-to-end cooperative learning strategy between CycleGAN [6] as unsupervised learning-based model and SRN [7] as supervised learning guided model. Unlike the state-of-the-art MAP-based approaches, our method is end-to-end without any blur kernel estimation or post-processing and with a much shorter execution time. To address natural image blurring phenomena, as the blur kernels are unknown the existing methods usually synthesize blur kernels by modeling the camera trajectories [8, 9] and convolve them with clear image data to build a paired training dataset. Instead we built our own face image deblurring generated dataset by strictly different and more reasonable and realistic manner. In order to translate realistically sharp face image data to its blurred counterpart and pair them to build a labeled data. We trained a CycleGAN [6] model on real unpaired clear/blurred face image data collected from the wild with a variety of phone device cameras. We then utilize the trained model to generate a large labeled paired clear/blurred training dataset during its inference. The figure 1. Illustrate our face image deblurring results impact on the face detection accuracy. Even the blurred faces are detectable but within this work we will demonstrate that is not always the case.

Our main contributions in this work are:

- Rather than using CycleGAN [6] just for data augmentation as it is communally used in some recent works. Instead we present a use case where we can exploit this algorithm for data labeling moreover than just data augmentation.
- To better address the face image blurring problematic, we present a sequential learning strategy in a learning chain consists of an unsupervised learning based-algorithm in charge of data labeling and a supervised learning-guided algorithm taking charge of face image recovery.
- And finally, we investigated the face image deblurring impact on the face detection accuracy.
- Our results are reproducible by following the detailed instructions attached with our project open source code [17].

2. RELATED WORK

Image deblurring is a classical problem in image processing. We can divide image deblurring learning-based methods into two main classes, supervised learning-based and unsupervised learning-based approaches.

Most of the recent image deblurring approaches are supervised learning-based. In which deep learning models have been trained on a labeled data to learn the mapping function from the blurred domain back to the sharp domain. Whoever this kind of methods are still always constrained by the lack of the appropriate labeled training data. The state-of-the-arts attempts to tackle the image blurring phenomena rely on GoPro dataset [12]. Which is a benchmarked labeled image deblurring dataset that was built by averaging high frame rate videos to one single image. Even though this data generation method still cannot emulate and cover all the sides of the phenomena under the scope. Moreover GoPro dataset [12] suffer from several shortcomings in term of its size which do not exceeds more than 2000 blur/clear labeled pairs and less than 10 different sceneries are covered by its contain. The thing that limit the supervised learning-based deblurring approaches capability to map accurately from the blurred domain to the clear domain.

To overcome the supervised learning-based methods limitations other works attention was attracted by the unsupervised learning concept as a labeled data-free method. inspired from [6] that propose a CycleGAN model for image translation that can be trained on unpaired data and assuming that image deblurring task is no more than a special case of image to image translation task, Yuan et al [13] trained CycleGAN model on a shuffled version of GoPro dataset[12] to tackle the image blurring phenomena.

From an opposite perspective, benefitting from unsupervised learning and supervised learning strengths, we present a method with two learning stages in one package to tackle image blurring effect. The first is in charge of generating labeled dataset starting from real clear/blurred unlabeled dataset. While the second is assigned to restore blurred face image data with fine details by using the first stage generated labeled data as training data.

3. PROPOSED METHOD

In this section, we describe the design methodology of the proposed face deblurring approach. We exploit the CycleGAN model [6] as bidirectional image translator to learn how to translate clear image data to blurred one by training the model with adversarial and cycle-consistent losses on real blurred/clear unpaired data. We then used the trained model to build a massive dataset for image face deblurring to serve the dependence of the supervised learning-based deblurring algorithms to labeled training data.

Finally, a Recurrent Neural Network as supervised learning-based model architecture has been trained on the built generated labeled dataset using the Euclidean loss to produce realistic deblurred results.

3.1. CycleGAN model Loss functions.

The full training objective of a CycleGAN model [6] consists of two kinds of loss functions: adversarial losses [11] for matching the distribution of generated images to the data distribution in the target domain, and cycle consistency losses to prevent the learned mappings G and F from contradicting each other.

Under the control of the both losses the training aim is to learn the mapping functions between two domains X and Y given training samples $\{x_i\}_{i=1}^N$ where $x_i \in X$ and $\{y_j\}_{j=1}^M$ where $y_j \in Y$. While $x \sim P_{data}(x)$ and $y \sim P_{data}(y)$ denote the data distribution. As illustrated in Figure 2 (a), the model includes two mappings $G: X \rightarrow Y$ and $F: Y \rightarrow X$. In addition, two adversarial discriminators D_X and D_Y , where D_X aims to distinguish between images $\{x\}$ and translated images $\{F(y)\}$; in the same way, D_Y aims to discriminate between $\{y\}$ and $\{G(x)\}$.

Adversarial Loss: it was applied to both mapping functions. For the mapping function $G: X \rightarrow Y$ and its discriminator D_Y , the Adversarial Loss function can be expressed as:

$$\mathcal{L}_{GAN}(G, D_Y, X, Y) = E_{y \sim p_{data}(y)}[\log D_Y(y)] + E_{x \sim p_{data}(x)}[\log(1 - D_Y(G(x)))] \quad (1)$$

Where G tries to generate images $G(x)$ that look similar to images from domain Y , while D_Y aims to distinguish between translated samples $G(x)$ and real samples y .

Cycle Consistency Loss: The mapping functions should be cycle-consistent, as shown in Figure 2 (b), for each image x from domain X , the image translation cycle should be able to bring x back to the original image, i.e., $x \rightarrow G(x) \rightarrow F(G(x)) \approx x$. We call this forward cycle consistency. Similarly, as illustrated in Figure 2 (c), for each image y from domain Y , G and F should also satisfy backward cycle

$$\mathcal{L}_{cyc}(G, F) = E_{x \sim p_{data}(x)}[\|F(G(x)) - x\|_1] + E_{y \sim p_{data}(y)}[\|G(F(y)) - y\|_1]. \quad (2)$$

consistency: $y \rightarrow F(y) \rightarrow G(F(y)) \approx y$. This behavior can be formulated as cycle consistency loss function:

3.2. SRN model loss function.

During SRN model training we used Euclidean loss between network output and the ground truth learning of the face image deblurring function which can be expressed as:

$$\mathcal{L} = \sum_{i=1}^n \frac{k_i}{N_i} \|I^i - I_*^i\|_2^2 \quad (3)$$

Where I^i and I_*^i are our network output and ground truth respectively in the i -th scale, $\{k_i\}$ are the trainable weights for each network scale.

4. MODEL ARCHITECTURE

As it is illustrated in figures 2 and 3. Our model architecture is a two training stages approach which includes two sub-models CycleGAN Figure 2. That is similar to that from [6] and a recurrent neural network Figure 3. Which was borrowed directly from [7] but with few changes. Our choice of these two models was motivated by the fact that each of them have its own limitations when it comes to

perform image deblurring task but they can complete each other if we jointly involve them to solve face image blurring challenge. Our model inference structure is shown in figure 4.

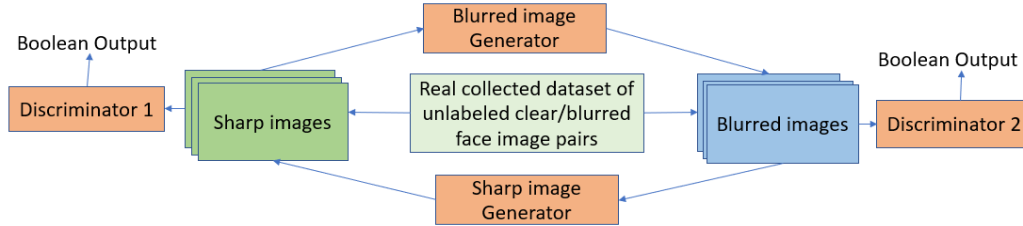


Figure 2. The first training phase: Training CycleGAN [6] on real data for clear to blurred image translation.

The CycleGAN [6] model weakness resides in its incapability to preserve accurately the semantic features and texture details in the restored face images. However, it is still very powerful in term of modeling highly complex distributions and learning from unlabeled data. On the other hand, as any other supervised learning-based algorithms, SRN [7] model performance is always depending on labeled training data with considerable amounts which is not available or hard to get. Even though it is still can restore blurred image data accurately if it is trained on the right and proper training data. Therefore, we use CycleGAN model as unsupervised learning-based model to generate a close realistic labeled face image deblurring dataset to overcome the limitations of SRN [7] in term of training data.

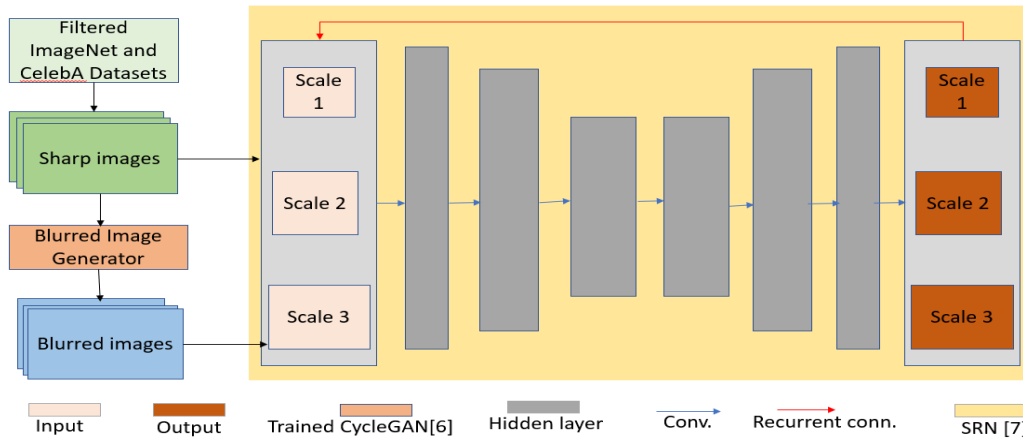


Figure 3. The second training phase: While CycleGAN model [6] is running on inference, SRN [7] is on training. The trained CycleGAN model input and output data are the input training data for SRN[7]

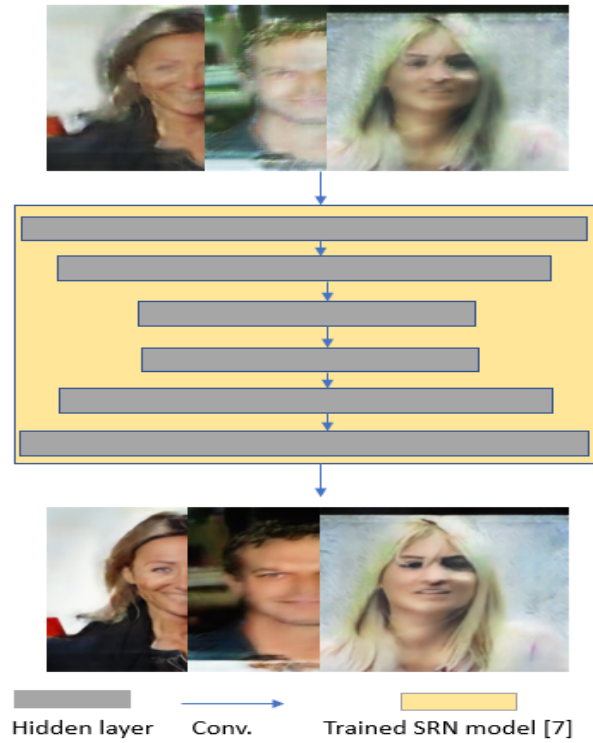


Figure 4. Blurred face Image restoration: Running our trained model on inference

5. IMPLEMENTATION DETAILS

Our experiments are conducted on a workstation with a NVIDIA Tesla T4 GPU. We implement our framework on TensorFlow platform [14]. Our evaluation is comprehensive to verify different network structures, as well as various network parameters. For fairness reason, all experiments are conducted on the same testing dataset with the same inference configuration of our baseline’s origin open sourced works.

We trained two CycleGAN networks [6] under the same training parameters. The first was trained with a batch size of 4, giving 70 epochs over a real wild unlabeled collected face image data for the aim of building our paired labeled dataset.

While the second was trained on the built dataset and included to our baselines list as well. The reconstructed images are regularized with the adversarial and cycle consistency loss functions respectively mentioned in (equation 1 & 2). SRN model [7] was trained differently on a batch size of 16 for 150 epochs on our built labeled face image dataset. The training process was supervised by the Euclidian loss function (Equation 3).

6. EXPERIMENTAL RESULTS

In this section, we first describe the training and testing datasets used in our experiments, we then analyze and evaluate our found results in term of their effect on face detection accuracy, and finally compare with the state-of-the-art deblurring algorithms.

6.1. Datasets

To create a large training dataset, the early learning-based methods [8, 15, 10] synthesize blurred images by convolving sharp images with real or generated uniform/non-uniform blur kernels. Due to the simplified image formation models, the synthetic data is still different from real ones that are captured by cameras.

In order to build more close realistic labeled face image deblurring dataset to serve the supervised learning-guided algorithms demand. A unique process was pursued to build a massive labeled dataset of 9000 blurred/clear image pairs. Starting from 1500 blurred/clear medium-sized unlabeled face image dataset taken by a variety of phone devise cameras, some of its samples are shown in the Figure. 5. 9000 Blurred face images have been generated taking 9000 clear face images filtered from CelebA and ImageNet datasets as an input during the inference of a CycleGAN [6] model that was trained on the mentioned collected dataset. By pairing its input with its generated output, we built a labeled dataset of 9000 blurred/clear pairs of face images .80% of the built dataset has been used as training dataset for SRN [7] and we kept 20% as a testing dataset.

For fair comparison with respect to our baselines networks structures and training strategy and parameters, we exploit their official pretrained model published on GitHub that was trained using the GoPro dataset [12].

While our proposed model was trained on our own built dataset. all experiments were conducted on the same testing dataset and on the same real-world deblurring challenges and scenarios.

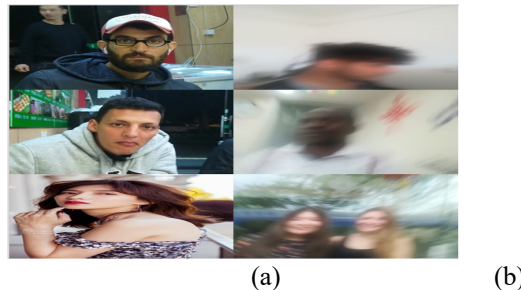


Figure 5. Some samples of the collected dataset. Blurred (b) and sharp (a) face images taken by phone

6.2. Comparisons against baselines

We compared our method with previous state-of-the-arts image deblurring approaches on both evaluation datasets and real images. Since our model deals with general camera shake and object motion, it is unfair to compare with traditional uniform deblurring methods. Instead we compare with more recent work of Yuan et al [38], which demonstrated very good results. Tao et al. [7] recovered the sharp

images by training a recurrent neural network model on GoPro dataset. Zhu et al. [6] which was included to our baselines list as well by training its proposed model on our built dataset.

We used the official implementation from the authors with default parameters. The quantitative results on our testing dataset are listed in Table 1. Visual comparison is shown in Figures. 6 and 7.

The second row of Figure.6 contains images from our generated testing dataset, which suffer from complex blur due to camera shake and object motion. The most of recent state-of-the-art methods used GoPro as a training dataset. These methods can produce good quality results but with remaining a few blurry structure and artifacts. In construct of video frames averaging based synthetic training sets. Our dataset complex blur patterns learnt from the real blurred face image data are the closest to accurately imitate the real blurring phenomena. The thing that puts these methods face to the real challenge and fail to predict reliable kernels on most cases. The returned results are only slightly sharpened or even worse than input Figure. 6(c), (d), and (e). Thanks to the designed framework and modules, our method produces superior results with sharper structures and clear details. According to our experiments, even on extreme cases, where motion is too large, our method can still produce sharp and smooth results. Quantitative results also validate our observation, where our framework outperforms others by a large margin.

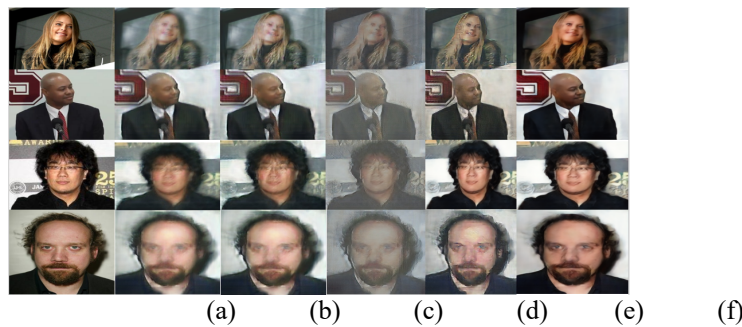


Figure 6. Visual comparisons on our testing dataset. from left to right: Ground truth, Blurred input, Tao et al [7], Yuan et al [13], Zhu et al [6], Ours.

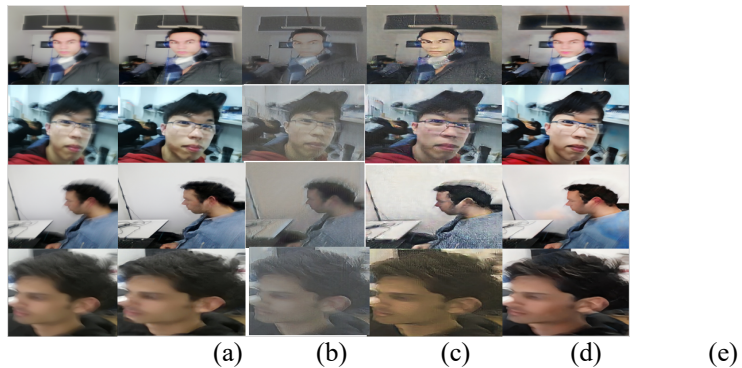


Figure 7. Visual comparisons on real blurred face images. From left to right: blurred input, Tao et al [7], Yuan et al [13], Zhu et al [6], Ours.

For more fairness we put our method and the baselines approaches face to real image deblurring challenges. We show our results on real-captured blurred images in Figure.7. As shown in Figure.7(e). compared to the baselines our model generalizes well on these images.

Table 1. Quantitative results on our and GOPRO testing datasets referring to PSNR and SSIM metrics.

Methods	Our testing set		GOPRO testing set [12]	
	PSNR	SSIM	PSNR	SSIM
Tao et al [7]	11.94	0.7614	30.26	0.9367
Yuan et al [13]	8.11	0.6749	26.55	0.9101
Zhu et al [6]	12.03	0.7345	28.81	0.9235
Ours	17.14	0.8509	32.62	0.9502

6.3 Deblurring Effect on Face Detection Accuracy.

Comparing our found results in term of PSNR and SSIR were necessary but not sufficient evaluation to prove their high-quality performance on real world applications. therefore, we decided to analyze face image deblurring effect on face detection accuracy within a further experiment.

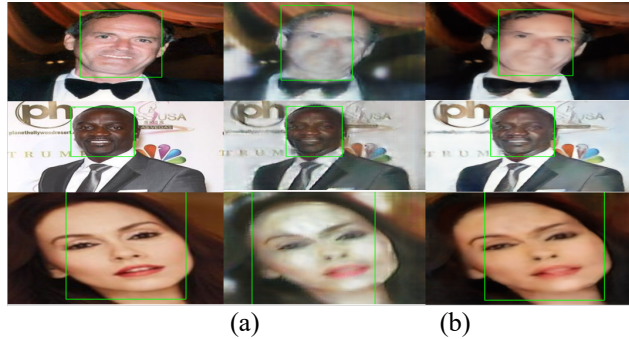


Figure 8. FaceBoxes [16] Face detection approach performance on our face deblurring results a, b and c respectively show the ground truth, blurred and restored version. Even the blurred faces are detectable but not always the case.

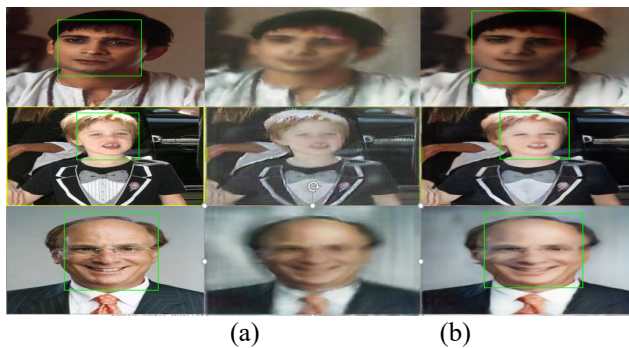


Figure 9. FaceBoxes [16] Face detection performance on our face deblurring results a, b and c respectively show the ground truth, blurred and restored version. In this case [16] failed to detect blurred faces in (b) while they were successfully detected on their recovered version (c) by our proposed face deblurring approach.

Table 2. The success rate of face detection on blurred, ground-truth and recovered images of our testing dataset and dataset of synthetically blurred images.

	Face Detection Accuracy		
	Ground Truth	Blurred	Restored
Testing dataset	0.87	0.32	0.61
Synthetic dataset	0.87	0.63	0.76

We used FaceBoxes [16] a face detection deep learning-based algorithm to detect the face for each image in our testing set and its recovered version. However, due to the motion blur and the ringing artifacts, as shown in Figure.8. Figure.9. faces in some of the blurred and deblurred images cannot be always detectable. We then compute the face detection accuracy on the restored version of our testing set. We show the face detection success rate for blurred, ground truth and recovered images in Table 2. The proposed method produces fewer artifacts and thus achieves a higher success rate on the restored images compared to blurred ones which proves that our face deblurring approach can dramatically increase the face detection deep learning-based algorithms accuracy on real wild blurred face images.

7. CONCLUSIONS

In this paper, we have explained what is the proper way to build a realistic labeled face image deblurring training dataset. We present a cooperative sequential learning strategy in a learning chain where supervised learning and an unsupervised are jointly involved to better address the face image blurring phenomena. We further investigate the face image deblurring impact on face detection accuracy. Our face image deblurring results supports our system design and training strategy and obviously outperforms our baselines, both qualitatively and quantitatively. We believe that this approach could be applied to other image processing tasks and other domain-specific or generic image deblurring tasks, and we will explore them in a future work.

8. REFERENCES

- [1] J. Pan, D. Sun, H. Pfister, and M. Yang. Blind image deblurring using dark channel prior. In CVPR, 2016. 1, 2
- [2] L. Xu, S. Zheng, and J. Jia. Unnatural L0 sparse representation for natural image deblurring. In CVPR, 2013. 1, 2, 6, 7, 8
- [3] Z. Hu, S. Cho, J. Wang, and M.-H. Yang. Deblurring lowlight images with light streaks. In CVPR, 2014. 1, 2
- [4] J. Pan, Z. Hu, Z. Su, and M. Yang. L0-regularized intensity and gradient prior for deblurring text images and beyond. TPAMI, 39(2):342–355, 2017. 1, 2
- [5] J. Pan, Z. Hu, Z. Su, and M. Yang. Deblurring face images with exemplars. In ECCV, 2014. 1, 2, 6, 7, 8
- [6] JY Zhu, T Park, P Isola, AA Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. Proceedings of the IEEE international conference on ..., 2017
- [7] Tao, Xin & Gao, Hongyun & Shen, Xiaoyong & Wang, Jue & Jia, Jiaya. (2018). Scale-Recurrent Network for Deep Image Deblurring. 8174-8182. 10.1109/CVPR.2018.00853.
- [8] A. Chakrabarti. A neural approach to blind motion deblurring. In ECCV, 2016. 2
- [9] M. Hradířs, J. Kotera, P. Zemečık, and F. Sroubek. Convolutional neural networks for direct text deblurring. In BMVC, 2015. 2

- [10] L. Sun, S. Cho, J. Wang, and J. Hays. Good image priors for non-blind deconvolution. In ECCV, 2014. 2
- [11] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., WardeFarley, D., Ozair, S., Courville, A., Bengio, Y. Generative adversarial nets. In: Advances in neural information processing systems, pp. 2672–2680 (2014).
- [12] S. Nah, T. H. Kim, and K. M. Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. pages 3883–3891, 2017.
- [13] Yuan, Quan & Li, Junxia & Zhang, Lingwei & Wu, Zhefu & Liu, Guangyu. (2019). Blind Motion Deblurring with Cycle Generative Adversarial Networks.
- [14] M. A. et. al. TensorFlow: large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.
- [15] J. Sun, W. Cao, Z. Xu, and J. Ponce. Learning a convolutional neural network for non-uniform motion blur removal. In CVPR, pages 769–777, 2015
- [16] Zhang, Shifeng & Zhu, Xiangyu & Lei, Zhen & Shi, Hailin & Wang, Xiaobo & Li, Stan. (2017). FaceBoxes: A CPU Real-time Face Detector with High Accuracy. 10.1109/BTAS.2017.8272675.
- [17] <https://github.com/nahliabdelwahed/Face-Image-Deblurring>