Deep Semantic Face Deblurring
Supplementary Material

Ziyi Shen[1] Wei-Sheng Lai[2] Tingfa Xu[1] Jan Kautz[3] Ming-Hsuan Yang[2,4]
[1]Beijing Institute of Technology [2]University of California, Merced
[3]Nvidia [4]Google Cloud
https://sites.google.com/site/ziyishenmi/cvpr18_face_deblur

1. Overview

In this supplementary document, we present additional results to complement the paper. First, we provide the network parameters and configuration of our face parsing network, face deblurring network and discriminator for the adversarial training. Second, we present additional analysis on the multi-scale deblurring network, perceptual loss and adversarial loss. Finally, we show more qualitative comparisons with state-of-the-art algorithms on both synthetic datasets and real blurred face images.

2. Network Architectures

The proposed semantic face deblurring network consists of two sub-networks: 1) a semantic face parsing network that generates the semantic labels of an input blurred image and 2) a multi-scale deblurring network that restores a clear face image in a coarse-to-fine manner. We also impose a perceptual loss based on the VGG-Face network [5] and an adversarial loss from a discriminator. Table 1 and 2 list the detailed configuration and parameters of our face parsing and deblurring networks, respectively. Table 3 shows the architecture of our discriminator for the adversarial training.

Table 1. **Detailed architecture of our face deblurring network.** Our network consists of two scales. The input of the first scale is the concatenation of a blurred image and 11-channel semantic labels. The input of the second scale is the concatenation of the output from the first scale, the blurred image and the corresponding 11-channel semantic labels. We use the transposed convolution to upsample the output image of the first scale by $2\times$.

Layer	Kernel size	#input channels	#output channels	Stride	Output Size
conv1-1	11×11	14	64	1	64×64
conv1-2	5×5	64	64	1	64×64
conv1-3	5×5	64	64	1	64×64
$6\times$ ResBlock	5×5	64	64	1	64×64
conv1-4	5×5	64	64	1	64×64
conv1-5	5×5	64	64	1	64×64
conv1-6	5×5	64	3	1	64×64
upsampling	4×4	3	3	1/2	128×128
conv2-1	11×11	17	64	1	128×128
conv2-2	5×5	64	64	1	128×128
conv2-3	5×5	64	64	1	128×128
$6\times$ ResBlock	5×5	64	64	1	128×128
conv2-4	5×5	64	64	1	128×128
conv2-5	5×5	64	64	1	128×128
conv2-6	5×5	64	3	1	128×128

Table 2. **Detailed architecture of our face parsing network.** We use the encoder-decoder architecture with skip connections from the encoder to decoder. The output is the 11-channel semantic labels.

Input	Output	Kernel size	#input channels	#output channels	Stride	Output Size
RGB	conv1	3×3	3	16	1	128×128
conv1	pool1	2×2	16	16	2	64×64
pool1	conv2	3×3	16	32	1	64×64
conv2	pool2	2×2	32	32	2	32×32
pool2	conv3	5×5	32	64	1	32×32
conv3	pool3	2×2	64	64	2	16×16
pool3	conv4	3×3	64	128	1	16×16
conv4	pool4	2×2	128	128	2	8×8
conv4	conv5	3×3	128	256	1	8×8
conv5	pool5	2×2	256	256	2	4×4
pool5	conv6	3×3	256	512	1	4×4
conv6	deconv6	4×4	512	512	1/2	8×8
deconv6	conv7	3×3	512	256	1	8×8
conv7 + conv5	deconv7	4×4	256	256	1/2	16×16
deconv7	conv8	3×3	256	128	1	16×16
conv8 + conv4	deconv8	4×4	128	128	1/2	32×32
deconv8	conv9	3×3	128	64	1	32×32
conv9 + conv3	deconv9	4×4	64	64	1/2	64×64
deconv9	conv10	3×3	64	32	1	64×64
conv10 + conv2	deconv10	4×4	32	32	1/2	128×128
deconv10	conv11	3×3	32	16	1	128×128
conv11	labels	3×3	16	11	1	128×128

Table 3. **Detailed architecture of the discriminator network for adversarial training.** The input is a 128×128 RGB image. The output is a single scalar.

Layer	Kernel size	#input channels	#output channels	Stride	Output Size
conv1	5×5	3	64	2	64×64
conv2	5×5	64	128	2	32×32
conv3	5×5	128	256	2	16×16
conv4	5×5	256	512	2	8×8
conv5	5×5	612	1024	2	4×4
conv6	4×4	1024	1	1	1×1

3. Additional Experimental Results

In this section, we provide additional analysis on the multi-scale deblurring network and the effect of perceptual and adversarial losses on rendering photo-realistic results. We also conduct an experiment of face recognition on the CelebA dataset to demonstrate the performance of the proposed semantic face deblurring approach.

3.1. Single-scale and multi-scale network

We use a variant of multi-scale network [3] as our face deblurring network. Here we compare the performance of the deblurring network using a single scale and multiple scales (i.e., two image scales). Both networks have the semantic global priors as input and are optimized with the content and local structural losses. We compare the PSNR and SSIM on our Random-kernel dataset in Figure 1 and show visual comparisons in Figure 2. The multi-scale network produces sharper and clearer results than the single-scale network.

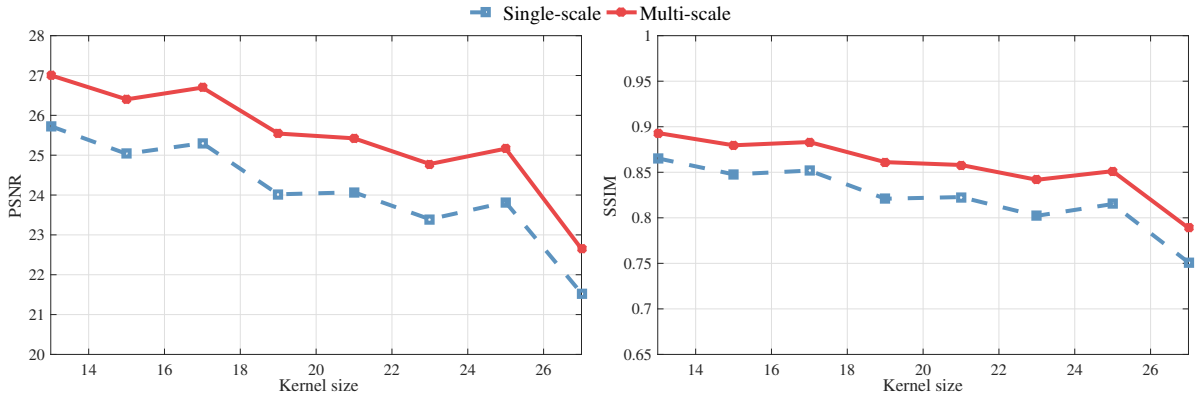
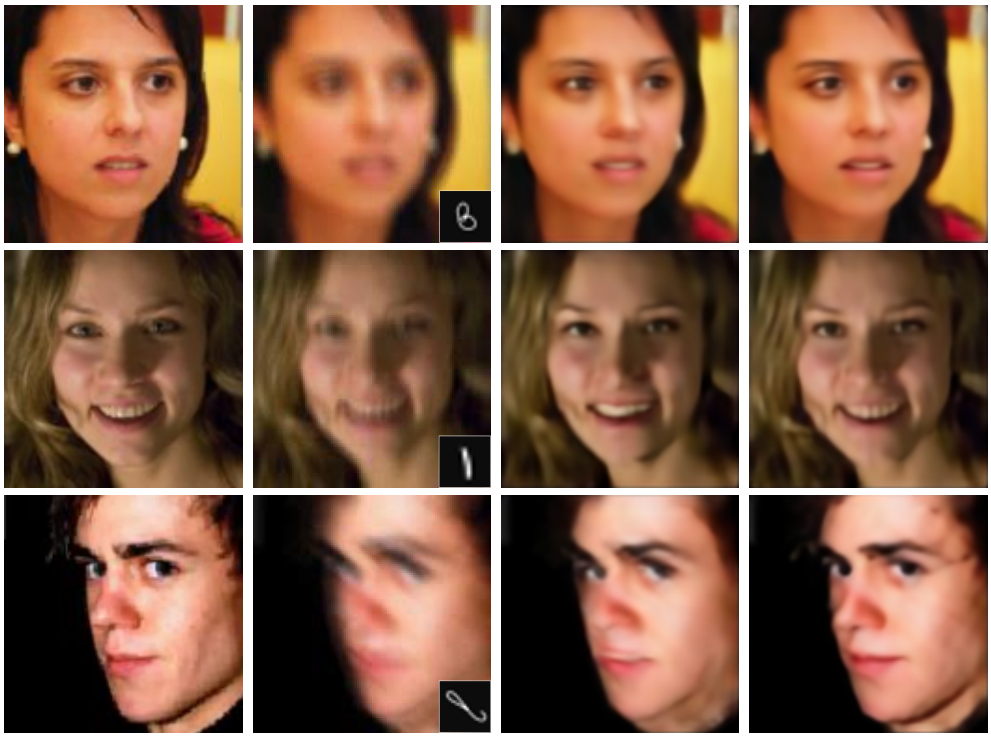


Figure 1. **Quantitative comparison of single-scale and multi-scale deblurring networks.** We evaluate the PSNR and SSIM on the Random-Kernel dataset.



(a) Clear images (b) Blurred images (c) Single-scale network (d) Multi-scale network

Figure 2. **Visual comparison of single-scale and multi-scale deblurring networks.** The multi-scale network produces sharper results.

3.2. Perceptual and adversarial losses

We compare the deblurring results with and without perceptual and adversarial losses in Figure 3. The perceptual loss encourages the images to match the high-level activations of the VGG-Face network and thus makes the output look photo-realistic. The adversarial loss further introduces more details on hairs and beard. Overall, the perceptual loss improves the average PSNR (on the Random-Kernel dataset) from 25.46 to 25.91. Although including the adversarial loss provides more fine details, the average PSNR is slightly dropped to 25.48. We give a quantitative comparison on perceptual loss and adversarial loss in Figure 3.

(a) Clear images (b) Blurred images (c) $\mathcal{L}_c + \mathcal{L}_s$ (d) $\mathcal{L}_c + \mathcal{L}_s + \mathcal{L}_p$ (e) $\mathcal{L}_c + \mathcal{L}_s + \mathcal{L}_p + \mathcal{L}_{\text{adv}}$

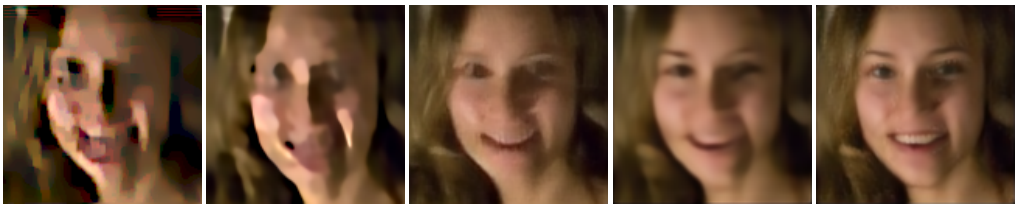
Figure 3. **Comparison between perceptual and adversarial losses.** We train our face deblurring network using the content loss $\mathcal{L}_c$, local structural losses $\mathcal{L}_s$, perceptual loss $\mathcal{L}_p$ and adversarial loss $\mathcal{L}_{\text{adv}}$.

4. Qualitative Comparisons

In this section, we provide more visual comparisons with state-of-the-art methods [1, 2, 3, 4, 6, 7, 8] on our Helen and CelebA test sets as well as real blurred images. Finally, we show some failure cases of the proposed algorithm.

4.1. Helen test set

(a) Clear image (b) Blurred image (c) Krishnan et al. [2] (d) Pan et al. [4] (e) Shan et al. [6] (f) Xu et al. [7]



(g) Cho and Lee [1] (h) Zhong et al. [8] (i) Nah et al. [3] (j) Ours w/o semantics (k) Ours w/ semantics

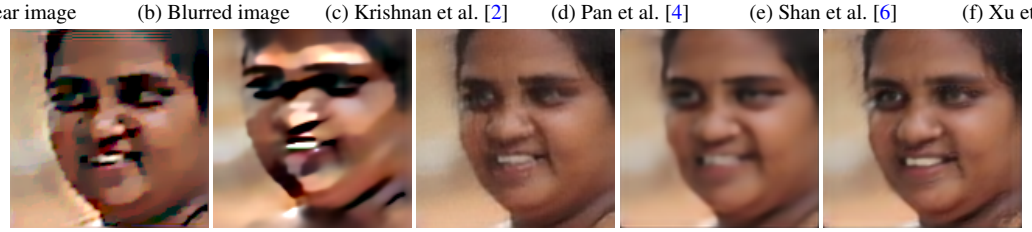
Figure 4. **Visual comparisons on the Helen test set.**

(a) Clear image (b) Blurred image (c) Krishnan et al. [2] (d) Pan et al. [4] (e) Shan et al. [6] (f) Xu et al. [7]

(g) Cho and Lee [1] (h) Zhong et al. [8] (i) Nah et al. [3] (j) Ours w/o semantics (k) Ours w/ semantics

Figure 5. **Visual comparisons on the Helen test set.**

(a) Clear image (b) Blurred image (c) Krishnan et al. [2] (d) Pan et al. [4] (e) Shan et al. [6] (f) Xu et al. [7]



(g) Cho and Lee [1] (h) Zhong et al. [8] (i) Nah et al. [3] (j) Ours w/o semantics (k) Ours w/ semantics

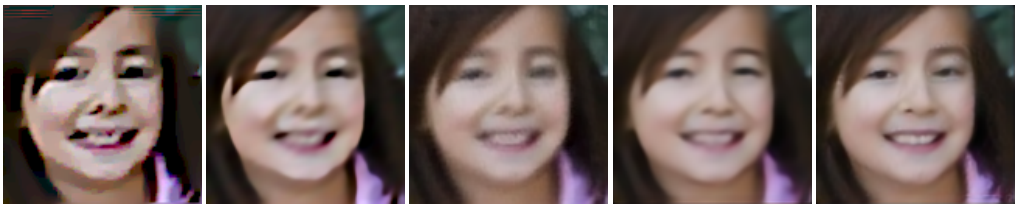
Figure 6. **Visual comparisons on the Helen test set.**

(a) Clear image (b) Blurred image (c) Krishnan et al. [2] (d) Pan et al. [4] (e) Shan et al. [6] (f) Xu et al. [7]

(g) Cho and Lee [1] (h) Zhong et al. [8] (i) Nah et al. [3] (j) Ours w/o semantics (k) Ours w/ semantics

Figure 7. **Visual comparisons on the Helen test set.**

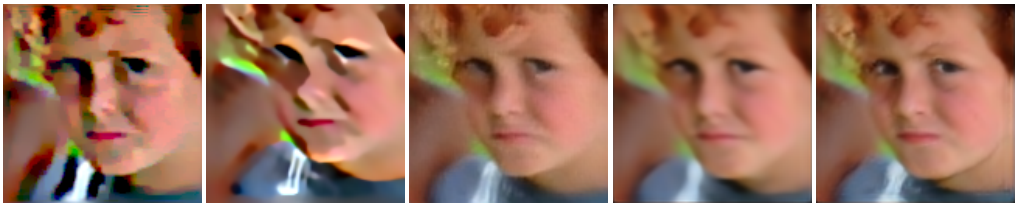
(a) Clear image (b) Blurred image (c) Krishnan et al. [2] (d) Pan et al. [4] (e) Shan et al. [6] (f) Xu et al. [7]



(g) Cho and Lee [1] (h) Zhong et al. [8] (i) Nah et al. [3] (j) Ours w/o semantics (k) Ours w/ semantics

Figure 8. **Visual comparisons on the Helen test set.**

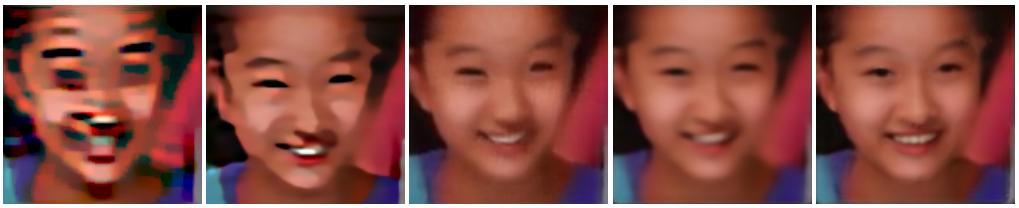
(a) Clear image (b) Blurred image (c) Krishnan et al. [2] (d) Pan et al. [4] (e) Shan et al. [6] (f) Xu et al. [7]



(g) Cho and Lee [1] (h) Zhong et al. [8] (i) Nah et al. [3] (j) Ours w/o semantics (k) Ours w/ semantics

Figure 9. **Visual comparisons on the Helen test set.**

(a) Clear image (b) Blurred image (c) Krishnan et al. [2] (d) Pan et al. [4] (e) Shan et al. [6] (f) Xu et al. [7]



(g) Cho and Lee [1] (h) Zhong et al. [8] (i) Nah et al. [3] (j) Ours w/o semantics (k) Ours w/ semantics

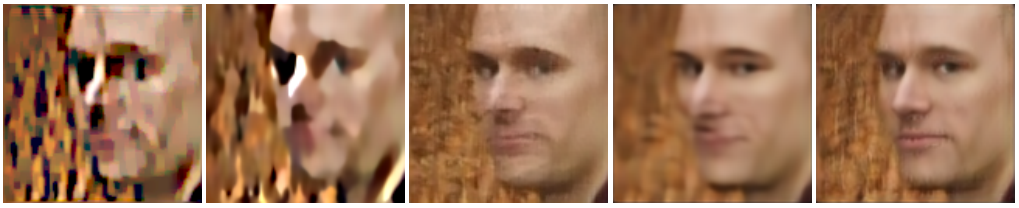
Figure 10. **Visual comparisons on the Helen test set.**

(a) Clear image (b) Blurred image (c) Krishnan et al. [2] (d) Pan et al. [4] (e) Shan et al. [6] (f) Xu et al. [7]

(g) Cho and Lee [1] (h) Zhong et al. [8] (i) Nah et al. [3] (j) Ours w/o semantics (k) Ours w/ semantics

Figure 11. **Visual comparisons on the Helen test set.**

(a) Clear image (b) Blurred image (c) Krishnan et al. [2] (d) Pan et al. [4] (e) Shan et al. [6] (f) Xu et al. [7]



(g) Cho and Lee [1] (h) Zhong et al. [8] (i) Nah et al. [3] (j) Ours w/o semantics (k) Ours w/ semantics

Figure 12. **Visual comparisons on the Helen test set.**

4.2. CelebA test set

(a) Clear image (b) Blurred image (c) Krishnan et al. [2] (d) Pan et al. [4] (e) Shan et al. [6] (f) Xu et al. [7]

(g) Cho and Lee [1] (h) Zhong et al. [8] (i) Nah et al. [3] (j) Ours w/o semantics (k) Ours w/ semantics

Figure 13. **Visual comparisons on the CelebA test set.**

(a) Clear image (b) Blurred image (c) Krishnan et al. [2] (d) Pan et al. [4] (e) Shan et al. [6] (f) Xu et al. [7]

(g) Cho and Lee [1] (h) Zhong et al. [8] (i) Nah et al. [3] (j) Ours w/o semantics (k) Ours w/ semantics

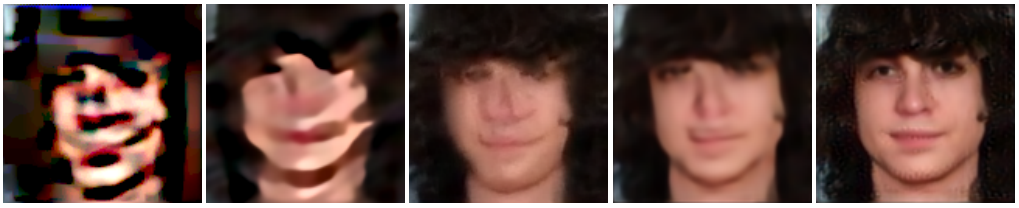
Figure 14. **Visual comparisons on the CelebA test set.**

(a) Clear image (b) Blurred image (c) Krishnan et al. [2] (d) Pan et al. [4] (e) Shan et al. [6] (f) Xu et al. [7]

(g) Cho and Lee [1] (h) Zhong et al. [8] (i) Nah et al. [3] (j) Ours w/o semantics (k) Ours w/ semantics

Figure 15. **Visual comparisons on the CelebA test set.**

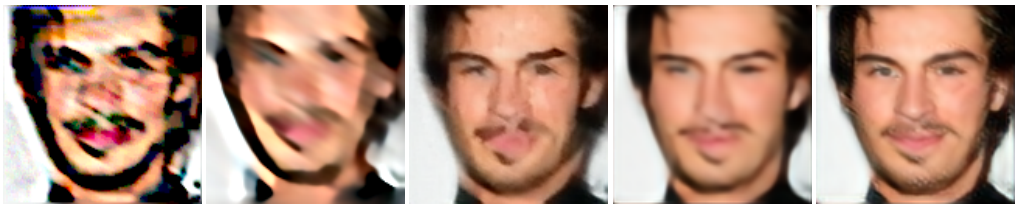
(a) Clear image (b) Blurred image (c) Krishnan et al. [2] (d) Pan et al. [4] (e) Shan et al. [6] (f) Xu et al. [7]



(g) Cho and Lee [1] (h) Zhong et al. [8] (i) Nah et al. [3] (j) Ours w/o semantics (k) Ours w/ semantics

Figure 16. **Visual comparisons on the CelebA test set.**

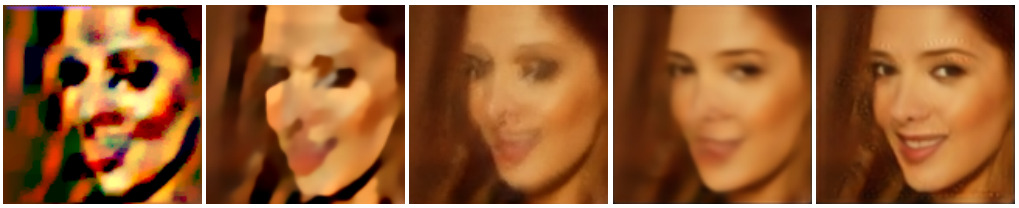
(a) Clear image (b) Blurred image (c) Krishnan et al. [2] (d) Pan et al. [4] (e) Shan et al. [6] (f) Xu et al. [7]



(g) Cho and Lee [1] (h) Zhong et al. [8] (i) Nah et al. [3] (j) Ours w/o semantics (k) Ours w/ semantics

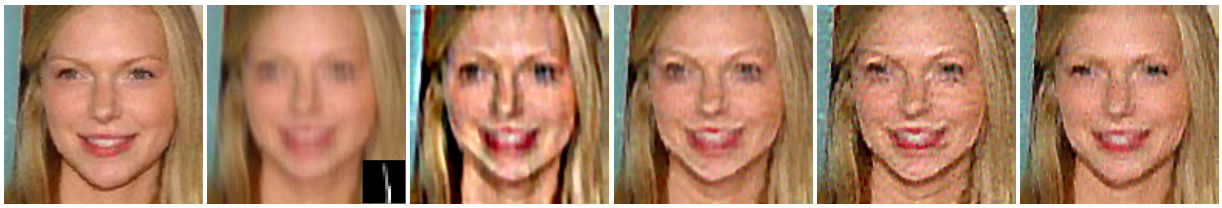
Figure 17. **Visual comparisons on the CelebA test set.**

(a) Clear image (b) Blurred image (c) Krishnan et al. [2] (d) Pan et al. [4] (e) Shan et al. [6] (f) Xu et al. [7]

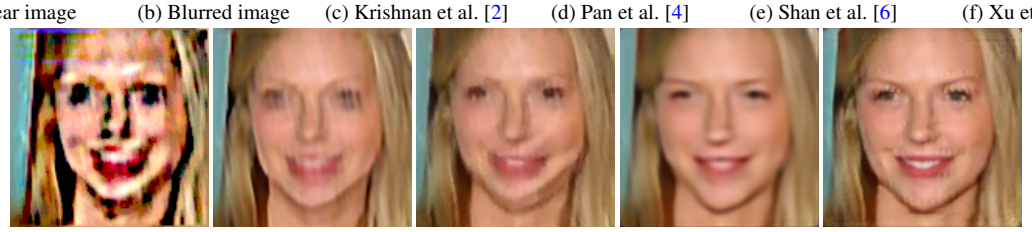


(g) Cho and Lee [1] (h) Zhong et al. [8] (i) Nah et al. [3] (j) Ours w/o semantics (k) Ours w/ semantics

Figure 18. **Visual comparisons on the CelebA test set.**



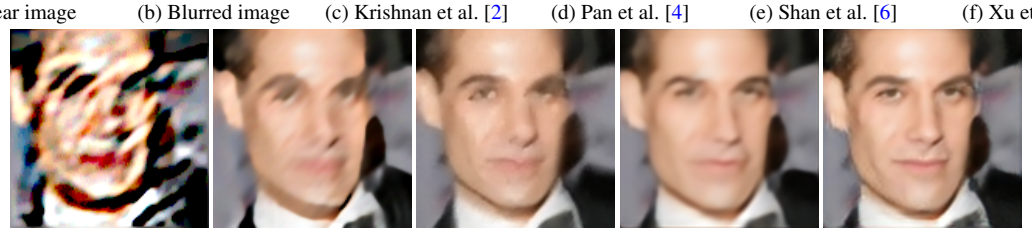
(a) Clear image (b) Blurred image (c) Krishnan et al. [2] (d) Pan et al. [4] (e) Shan et al. [6] (f) Xu et al. [7]



(g) Cho and Lee [1] (h) Zhong et al. [8] (i) Nah et al. [3] (j) Ours w/o semantics (k) Ours w/ semantics

Figure 19. **Visual comparisons on the CelebA test set.**

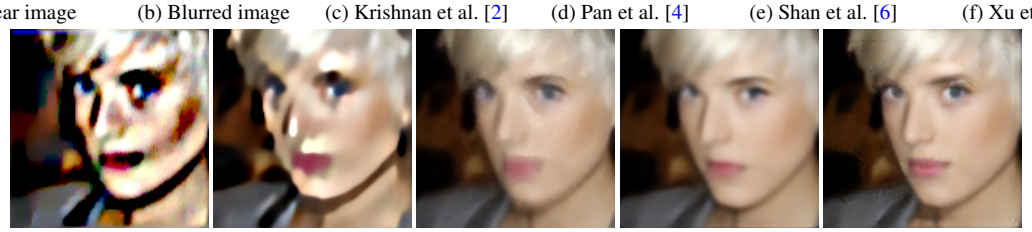
(a) Clear image (b) Blurred image (c) Krishnan et al. [2] (d) Pan et al. [4] (e) Shan et al. [6] (f) Xu et al. [7]



(g) Cho and Lee [1] (h) Zhong et al. [8] (i) Nah et al. [3] (j) Ours w/o semantics (k) Ours w/ semantics

Figure 20. **Visual comparisons on the CelebA test set.**

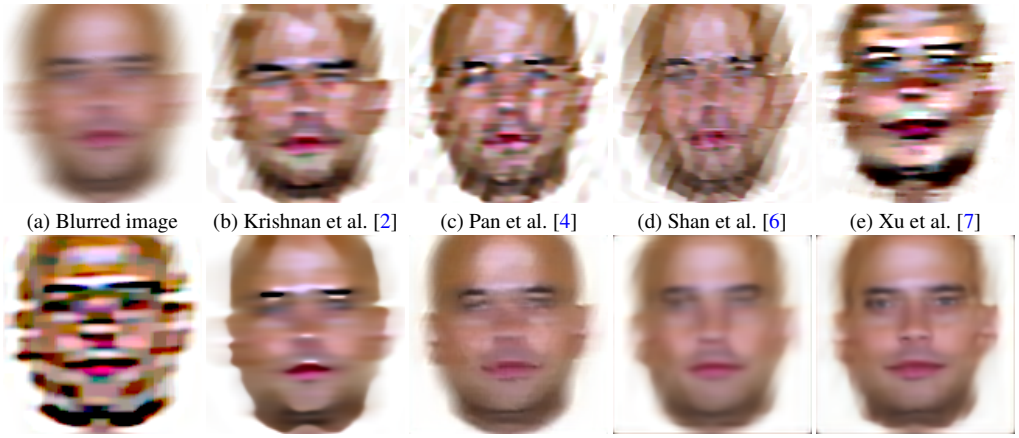
(a) Clear image (b) Blurred image (c) Krishnan et al. [2] (d) Pan et al. [4] (e) Shan et al. [6] (f) Xu et al. [7]



(g) Cho and Lee [1] (h) Zhong et al. [8] (i) Nah et al. [3] (j) Ours w/o semantics (k) Ours w/ semantics

Figure 21. **Visual comparisons on the CelebA test set.**

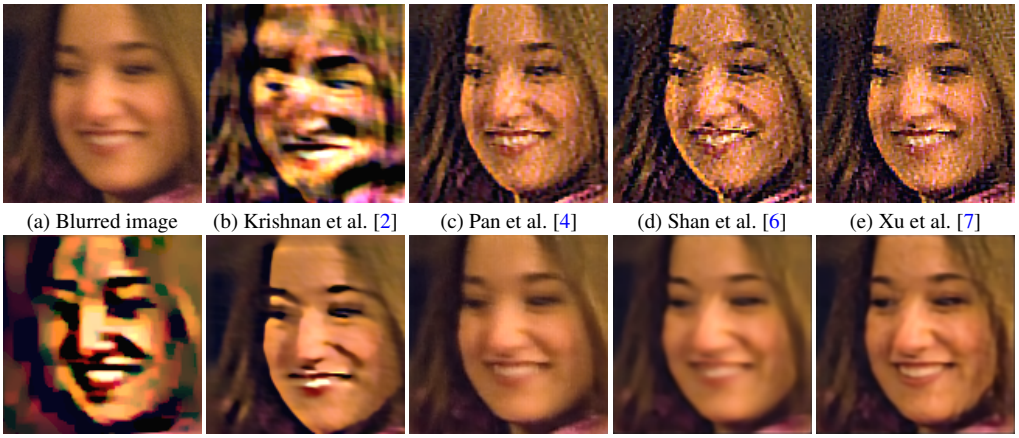
4.3. Real-world blurred images



(a) Blurred image (b) Krishnan et al. [2] (c) Pan et al. [4] (d) Shan et al. [6] (e) Xu et al. [7]

(f) Cho and Lee [1] (g) Zhong et al. [8] (h) Nah et al. [3] (i) Ours w/o semantics (j) Ours w/ semantics

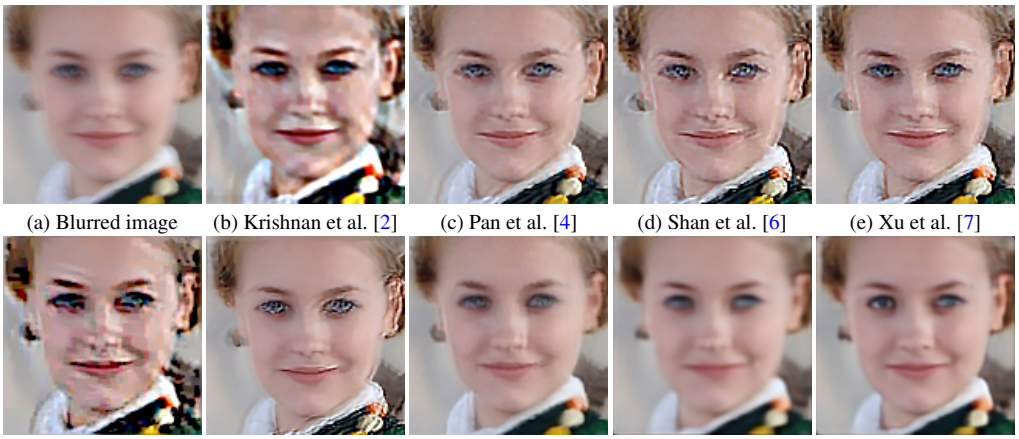
Figure 22. **Visual comparison of real-world face images.**



(a) Blurred image (b) Krishnan et al. [2] (c) Pan et al. [4] (d) Shan et al. [6] (e) Xu et al. [7]

(f) Cho and Lee [1] (g) Zhong et al. [8] (h) Nah et al. [3] (i) Ours w/o semantics (j) Ours w/ semantics

Figure 23. **Visual comparison of real-world face images.**



(a) Blurred image (b) Krishnan et al. [2] (c) Pan et al. [4] (d) Shan et al. [6] (e) Xu et al. [7]

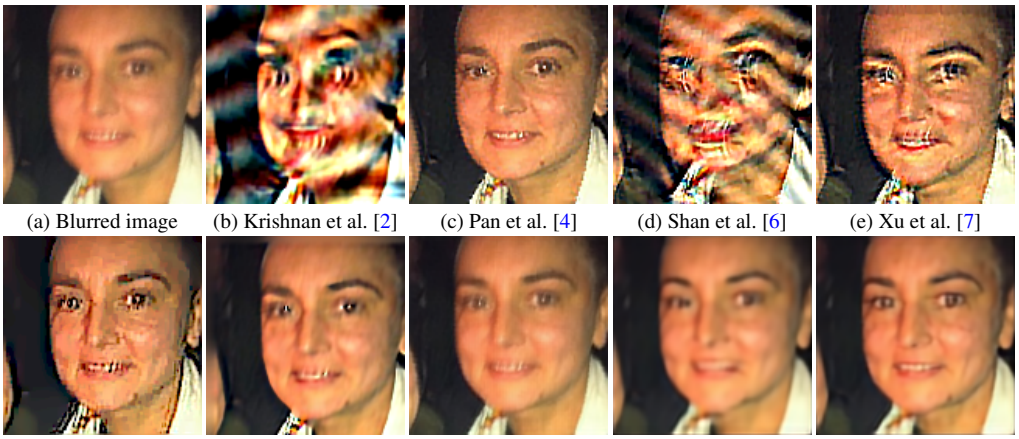
(f) Cho and Lee [1] (g) Zhong et al. [8] (h) Nah et al. [3] (i) Ours w/o semantics (j) Ours w/ semantics

Figure 24. **Visual comparison of real-world face images.**

(a) Blurred image (b) Krishnan et al. [2] (c) Pan et al. [4] (d) Shan et al. [6] (e) Xu et al. [7]

(f) Cho and Lee [1] (g) Zhong et al. [8] (h) Nah et al. [3] (i) Ours w/o semantics (j) Ours w/ semantics

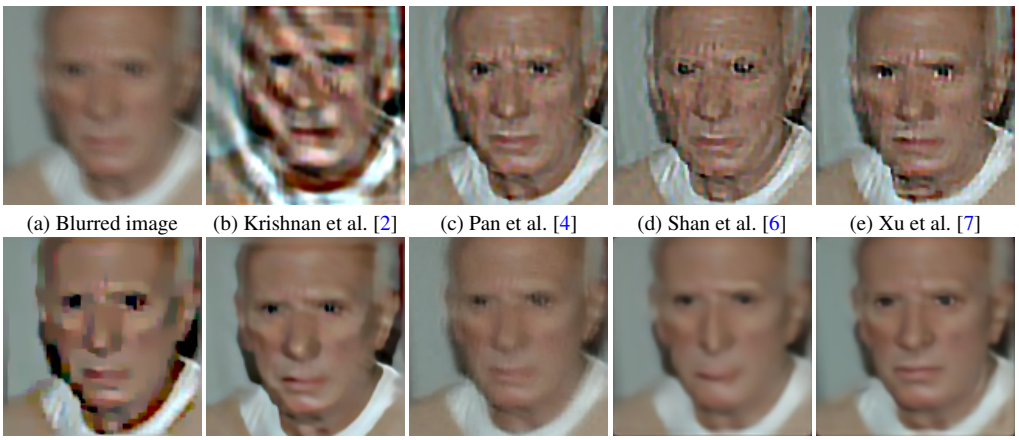
Figure 25. **Visual comparison of real-world face images.**



(a) Blurred image (b) Krishnan et al. [2] (c) Pan et al. [4] (d) Shan et al. [6] (e) Xu et al. [7]

(f) Cho and Lee [1] (g) Zhong et al. [8] (h) Nah et al. [3] (i) Ours w/o semantics (j) Ours w/ semantics

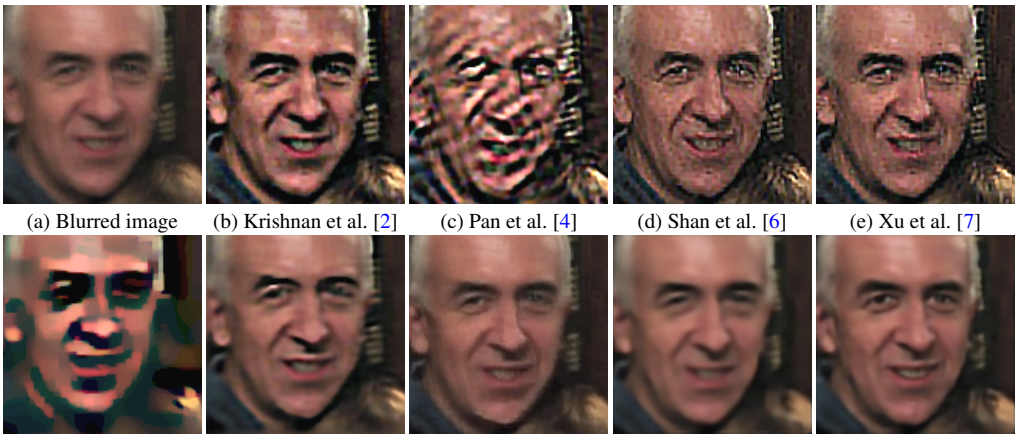
Figure 26. **Visual comparison of real-world face images.**



(a) Blurred image (b) Krishnan et al. [2] (c) Pan et al. [4] (d) Shan et al. [6] (e) Xu et al. [7]

(f) Cho and Lee [1] (g) Zhong et al. [8] (h) Nah et al. [3] (i) Ours w/o semantics (j) Ours w/ semantics

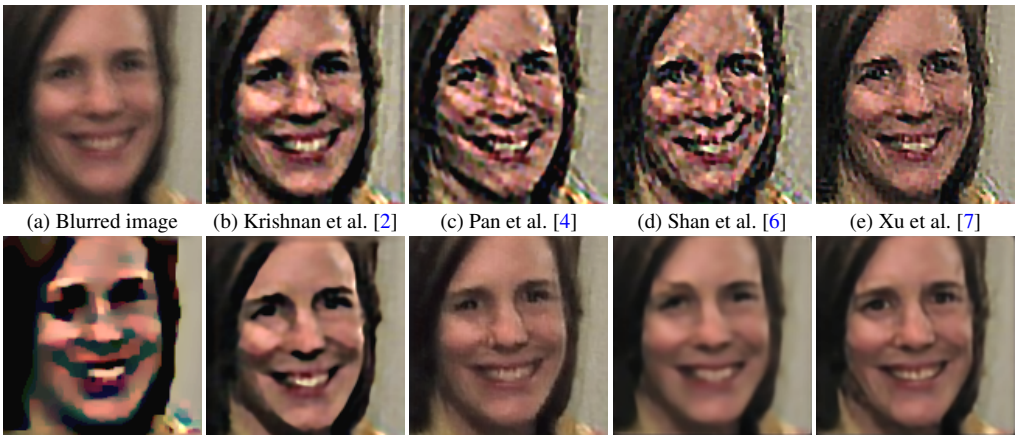
Figure 27. **Visual comparison of real-world face images.**



(a) Blurred image (b) Krishnan et al. [2] (c) Pan et al. [4] (d) Shan et al. [6] (e) Xu et al. [7]

(f) Cho and Lee [1] (g) Zhong et al. [8] (h) Nah et al. [3] (i) Ours w/o semantics (j) Ours w/ semantics

Figure 28. **Visual comparison of real-world face images.**



(a) Blurred image (b) Krishnan et al. [2] (c) Pan et al. [4] (d) Shan et al. [6] (e) Xu et al. [7]

(f) Cho and Lee [1] (g) Zhong et al. [8] (h) Nah et al. [3] (i) Ours w/o semantics (j) Ours w/ semantics

Figure 29. **Visual comparison of real-world face images.**

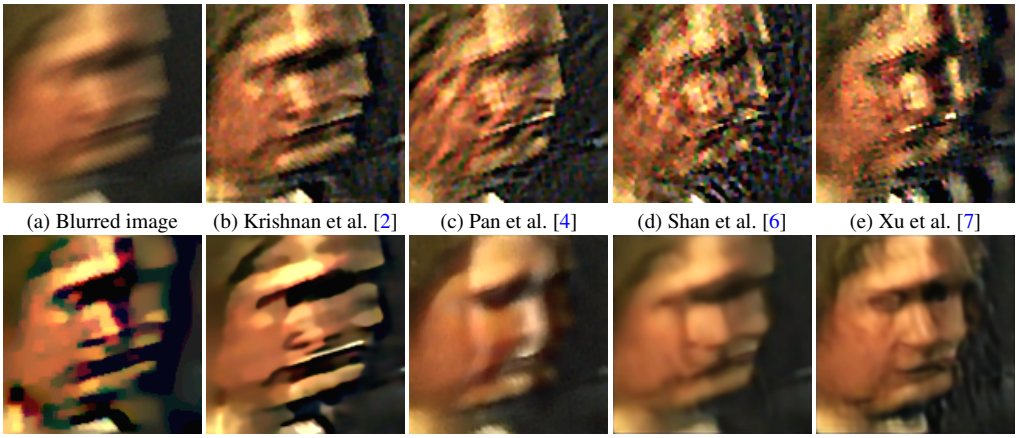
(a) Blurred image (b) Krishnan et al. [2] (c) Pan et al. [4] (d) Shan et al. [6] (e) Xu et al. [7]

(f) Cho and Lee [1] (g) Zhong et al. [8] (h) Nah et al. [3] (i) Ours w/o semantics (j) Ours w/ semantics

Figure 30. **Visual comparison of real-world face images.**

4.4. Failure cases

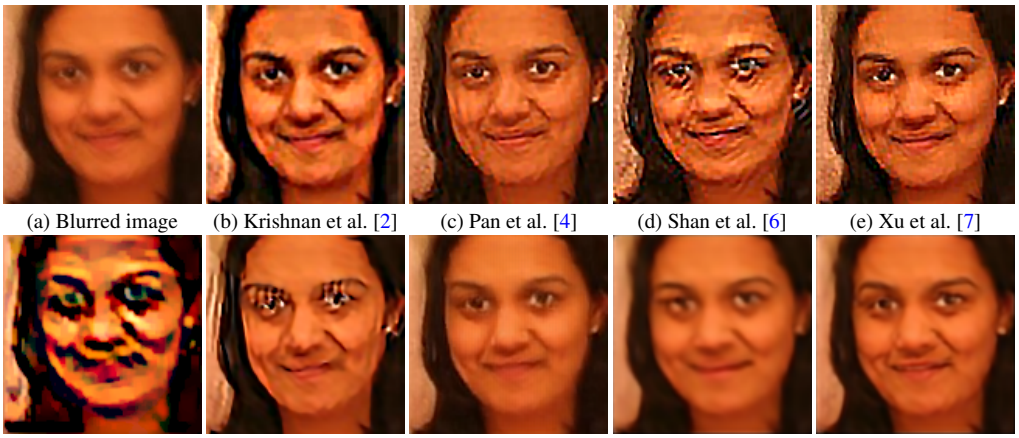
The proposed face deblurring method may not be robust to extremely large motion blur or side faces, as shown in Figure 31. In some cases, our method may hallucinate non-existing details, e.g., teeth in Figure 32, due to the strong prior of face parsing. However, our result in Figure 32 still looks visually pleasing and has less ringing artifacts.



(a) Blurred image (b) Krishnan et al. [2] (c) Pan et al. [4] (d) Shan et al. [6] (e) Xu et al. [7]

(f) Cho and Lee [1] (g) Zhong et al. [8] (h) Nah et al. [3] (i) Ours w/o semantics (j) Ours w/ semantics

Figure 31. **A failure case.** Our method may not be robust to large motion blur or side faces.



(a) Blurred image (b) Krishnan et al. [2] (c) Pan et al. [4] (d) Shan et al. [6] (e) Xu et al. [7]

(f) Cho and Lee [1] (g) Zhong et al. [8] (h) Nah et al. [3] (i) Ours w/o semantics (j) Ours w/ semantics

Figure 32. **A failure case.** Our method may hallucinate wrong details, e.g., teeth, that does not exist in the input image. However, our result still look photo-realistic with less ringing artifacts.

References

[1] S. Cho and S. Lee. Fast motion deblurring. *ACM TOG (Proceedings of SIGGRAPH Asia)*, 28(5):145:1–145:8, 2009. 5, 6, 7, 8, 9, 10, 11, 12, 13, 14

[2] D. Krishnan, T. Tay, and R. Fergus. Blind deconvolution using a normalized sparsity measure. In *CVPR*, 2011. 5, 6, 7, 8, 9, 10, 11, 12, 13, 14

[3] S. Nah, T. Hyun Kim, and K. Mu Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *CVPR*, 2017. 3, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14

[4] J. Pan, Z. Hu, Z. Su, and M. Yang. Deblurring face images with exemplars. In *ECCV*, 2014. 5, 6, 7, 8, 9, 10, 11, 12, 13, 14

[5] O. M. Parkhi, A. Vedaldi, and A. Zisserman. Deep face recognition. In *BMVC*, 2015. 1

[6] Q. Shan, J. Jia, and A. Agarwala. High-quality motion deblurring from a single image. *ACM TOG (Proceedings of SIGGRAPH)*, 27(3):73:1–73:10, 2008. 5, 6, 7, 8, 9, 10, 11, 12, 13, 14

[7] L. Xu, S. Zheng, and J. Jia. Unnatural L0 sparse representation for natural image deblurring. In *CVPR*, 2013. 5, 6, 7, 8, 9, 10, 11, 12, 13, 14

[8] L. Zhong, S. Cho, D. N. Metaxas, S. Paris, and J. Wang. Handling noise in single image deblurring using directional filters. In *CVPR*, 2013. 5, 6, 7, 8, 9, 10, 11, 12, 13, 14