Supplementary Materials to "Blind Perceptual Quality Assessment for Single Image Motion Deblurring"

No Author Given

No Institute Given

Due to space limitation, many details and effects of the proposed DBQA-NET are not shown in the manuscript. In the supplementary file, we provide more details on the method, as well as more results of DBQA-NET.

1 More about DBQA-NET

We spend a lot of time on model design, training, verification, and choice of loss function, and have tried many other ways. Here we focus on the structure of the most important module.

Table S1: The network structure from f_1 to f_4

Layer	f_1	f_2	f_3	f_4
1	conv(3,32,5,2) gdn(32) pool(2)	conv(64,128,5,2) gdn(128) pool(2)	conv(128,256,5,2) gdn(256) pool(2)	conv(256,256,5,2) gdn(256) pool(2)
2	conv(32,64,5,2) gdn(64) pool(2)	conv(128,256,5,2) gdn(256) pool(2)	conv(256,256,5,2) gdn(256) pool(2)	conv(256,256,3,1) gdn(256) pool(2)
3	conv(64,128,5,2) gdn(128) pool(2)	conv(256,256,3,1) gdn(256) pool(2)	conv(256,256,3,1) gdn(256) pool(2)	conv(256,256,3,1) gdn(256) pool(2)
4	conv(128,256,3,1) gdn(256) pool(2)	conv(256,256,3,1) gdn(256) pool(2)		

The Multi-resolution Feature Calculation is the core module of DBQA-NET, and is responsible for extracting multi-scale features of distorted images for the regression of pMOSs in subsequent module. In order to find a suitable CNN, we search many literatures and conduct "pristine-distortion" classification experiments with the CNN structure proposed by them, and find that MEON[S1] has the best convergence and accuracy when having the same number of layers. But the input of MEON is fixed, so we improve its structure to generate feature vectors of the same dimension for input sizes of 256×256, 128×128, 64×64 and 32×32. Table S1 shows the structure, where conv(3,32,5,2) means the convolution input has 3 channels, the output has 32 channels, the kernel size is 5, and the step size is 2; gdn(32) means a 32-channels GDN layer;

pool(2) represents the maximum pooling layer whose step size and kernel size are both 2. Please read the code for more details.

2 More about Experiment

We have done many experiments to prove the superiority of the proposed method. Due to space limitations, only a small part is shown in the manuscript. Here we give more experiments, including visualization, more ranking results, hyperparameter search and optimization experiments.

2.1 Visualization

In order to more visually show that the model has learned the essential features of distortion rather than overfitting, we visualize the features learned in stage one through CAM[S2], which determine the quality classification and the regression in stage two. The class label is back-propagated to form a certain degree of response in the perception area of the image, which is displayed by a heat map, as shown in Fig. S1. Both cases are from the proposed DeblurQA.

It can be seen that the scales and positions of the distortion features detected by each branch are different. The branch f_1 tends to perceive global and large-scale distortions, while f_4 tends to detail distortion. f_2 and f_3 reflect fewer features, but pay more attention to the edges of the object. The ability of each resolution branch is complementary, which also explains why our model has excellent performance.

2.2 More results of quality ranking

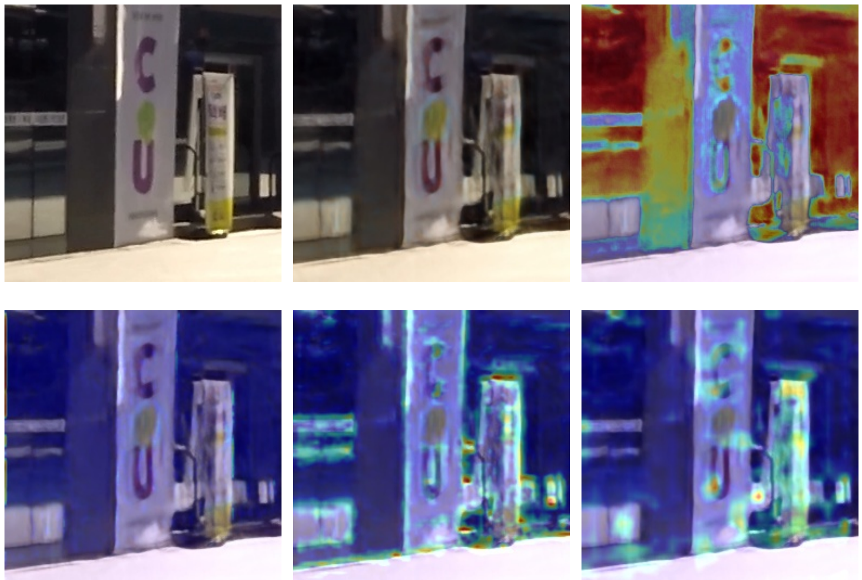
We manually screen out 6 images from low to high quality from each group, and use different BIQAs to evaluate them. Fig. S2 shows 4 image groups with continuous quality change, and Table S2 shows the assessment scores of each method. DBQA-NET also shows better consistency with subjective perception on these samples.

2.3 More results of automatic hyperparameter selection

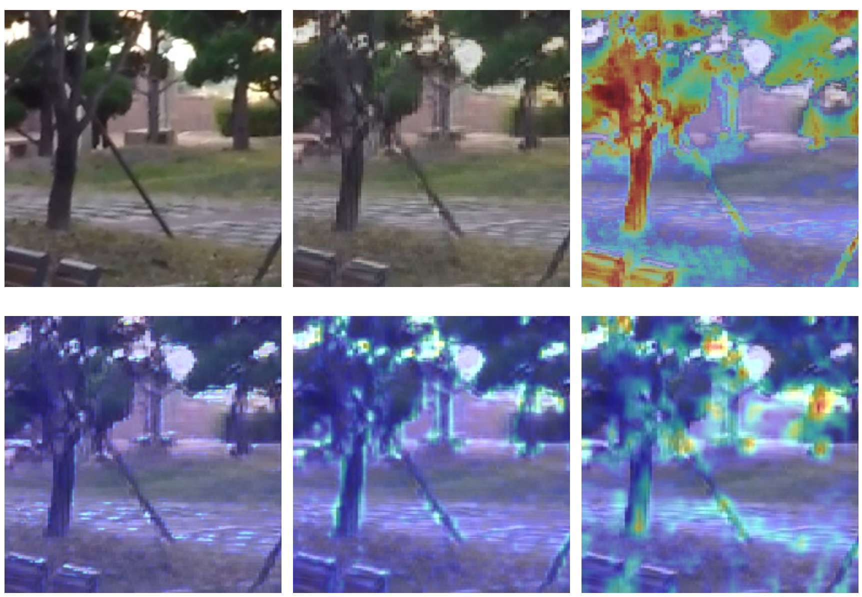
In Fig. S3, we present more results for Pan[1], and some examples for Pan[2] using DBQA-NET. In these examples, we use more actually shot images, which prove that DBQA-NET has good adaptability in more realistic scenes.

2.4 More results of automatic hyperparameter selection

Fig. S4 provide more results on optimization using Eq. 7. The images used are all from VOC. The comparison shows that DBQA-NET can not only eliminate the extra texture and ringing caused by optimization, but also enhance the details.

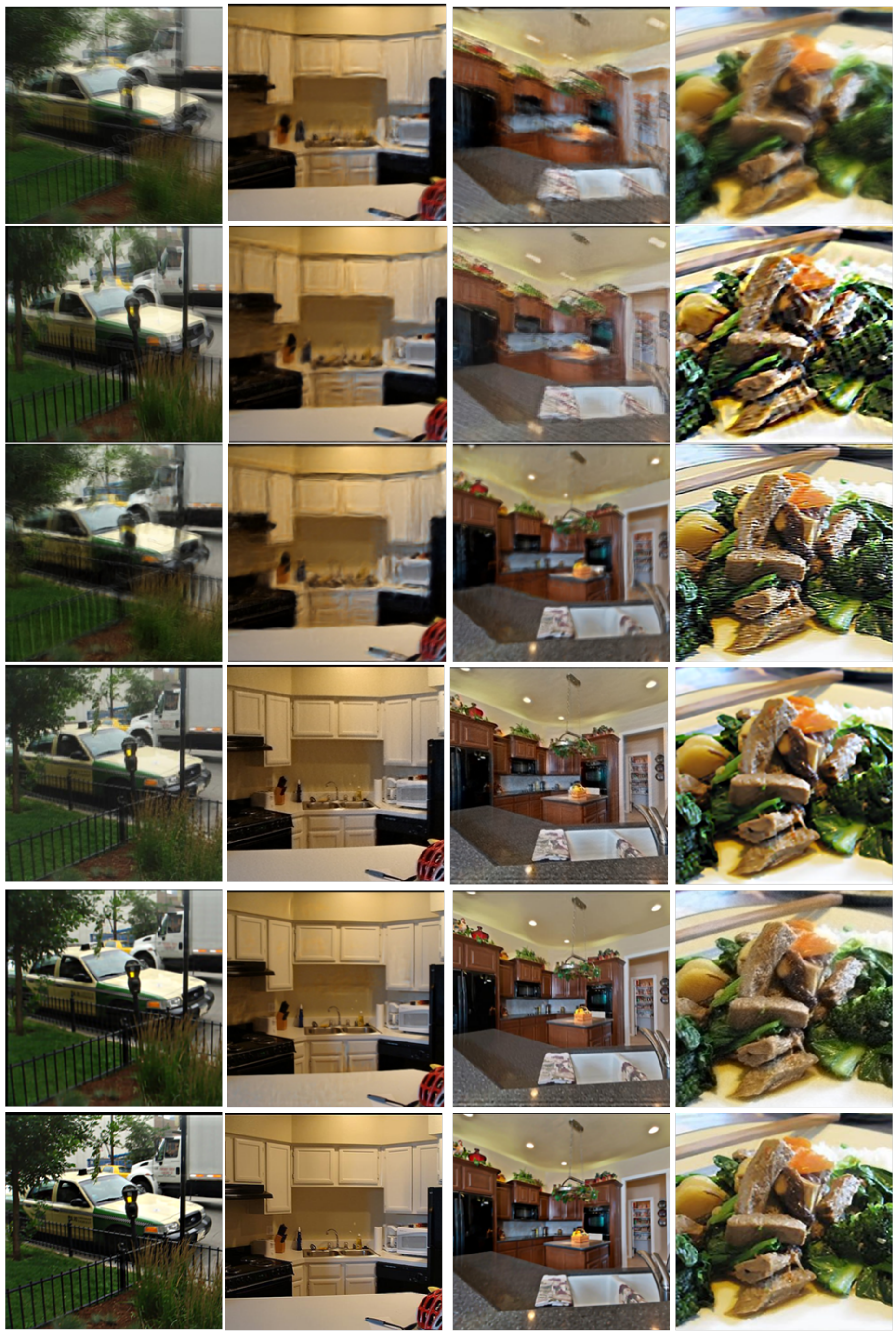


(a)The first example, from the upper left corner to the lower right corner are the pristine image, the deblurred and the distortion heat map from branch f_1 to f_4 respectively, the same below.



(b) The second example

Fig. S1: Two examples of feature visualization



(a) (b) (c) (d)

Fig. S2: Samples for ranking experiment, from low to high quality

Table S2: Evaluation of each method on Figure S2

(a)

ILNIQE	BIBLE	MT-A	MT-S	DBQA-NET
28.7349	3.2339	48.2592	53.9020	50.7296
29.8288	3.5248	50.6517	53.5153	51.1102
29.0991	2.3631	44.2273	51.1144	74.8749
25.3149	3.0076	52.1935	55.8374	76.9485
29.3008	3.3404	54.3278	56.8709	78.1981
26.1660	4.6017	54.1187	56.4937	81.7199

(b)

ILNIQE	BIBLE	MT-A	MT-S	DBQA-NET
44.0388	2.0126	45.2189	45.2189	55.9500
41.1549	1.8487	41.1431	41.1431	56.5297
49.3999	1.9568	41.7205	41.7205	57.8179
32.5475	4.7325	50.4162	50.4162	72.8241
31.6427	3.2054	48.8158	48.8158	79.1101
29.3394	4.8597	51.4954	51.4954	80.0560

(c)

ILNIQE	BIBLE	MT-A	MT-S	DBQA-NET
38.4171	2.6420	44.5336	45.2848	45.1663
32.0443	2.4634	45.7712	48.1741	56.4681
38.0456	1.7969	47.4016	48.9642	56.0179
20.7464	3.8582	54.6252	56.8537	61.5565
22.7777	4.0598	55.2665	56.9355	77.8930
31.6882	3.5995	54.8035	57.3914	79.1479

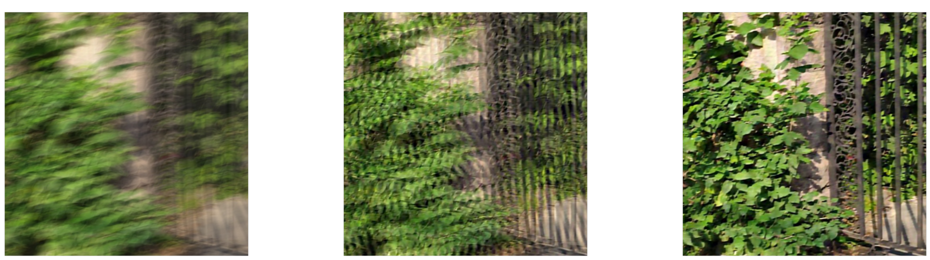
(d)

ILNIQE	BIBLE	MT-A	MT-S	DBQA-NET
55.3631	1.9591	40.1173	38.5677	40.0978
39.8988	4.1654	54.6502	52.1420	42.9655
48.2464	4.7913	52.5571	53.7649	50.3924
48.4018	3.1768	52.2411	49.4431	65.7094
35.7170	3.5696	49.6982	48.3981	70.9823
32.4989	4.7828	54.3200	53.1837	78.9030

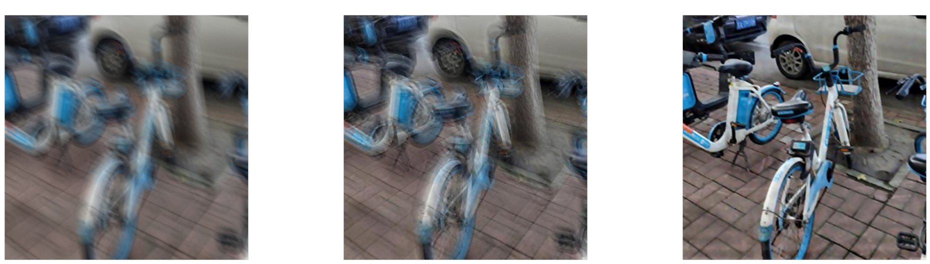
(a)Pristine image, blur image, the worst and best deblurring results of Pan[1]. The worst scores 47.1440(smooth=1×10^{-4}, Ksize=29), and the best 66.2538(smooth=1×10^{-5}, Ksize=42).

(b)Pristine image, blur image, the worst and best deblurring results of Pan[1]. The worst scores 54.4995(smooth=1×10^{-5}, Ksize=20), and the best 76.5108(smooth=1×10^{-4}, Ksize=46).

(c)Pristine image, blur image, the worst and best deblurring results of Pan[1]. The worst scores 49.3346(smooth=1×10^{-3}, Ksize=23), and the best 73.1362(smooth=1×10^{-5}, Ksize=31).

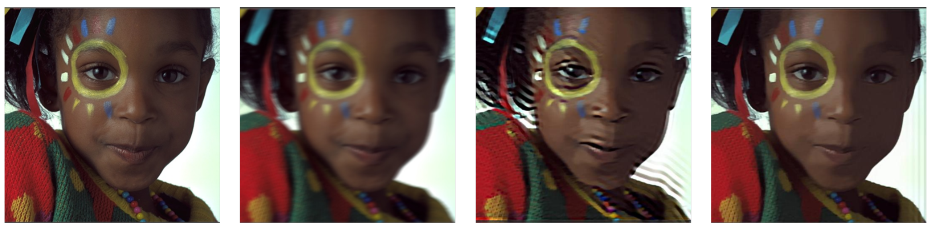


(d)Pristine image, blur image, the worst and best deblurring results of Pan[1]. The worst scores 44.9649(smooth=1×10^{-4}, Ksize=21), and the best 66.2502(smooth=1×10^{-5}, Ksize=57).

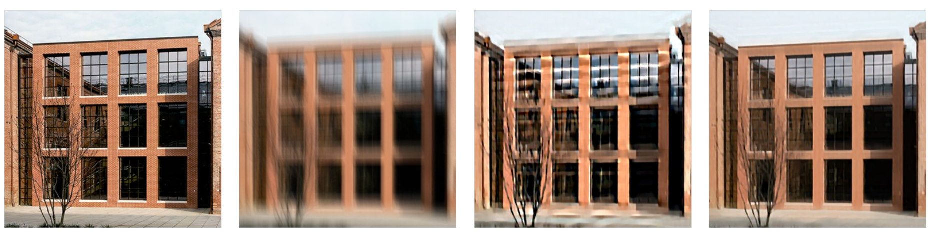


(e)Pristine image, blur image, the worst and best deblurring results of Pan[1]. The worst scores 36.8607(smooth=1×10^{-3}, Ksize=20), and the best 80.0392(smooth=1×10^{-5}, Ksize=51).

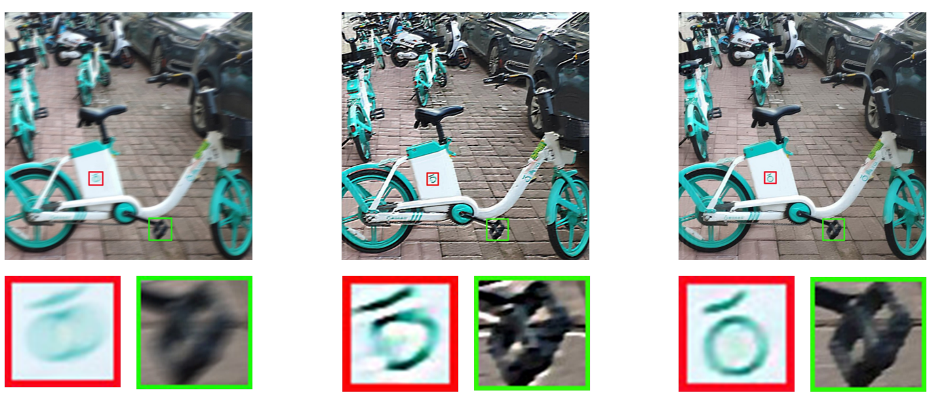
(f)Pristine image, blur image, the worst and best deblurring results of Pan[2]. The worst scores 2.3855(λ_{l0} =5 $\times$ 10^{-4}, Ksize=91), and the best 88.4276(λ_{l0}=1 $\times$ 10^{-3}, Ksize=53).



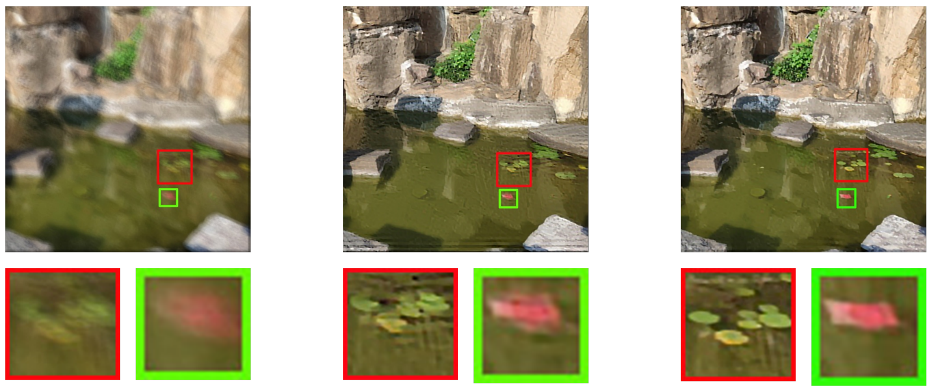
(g)Pristine image, blur image, the worst and best deblurring results of Pan[2]. The worst scores 49.6256(λ_{l0} =1 $\times$ 10^{-3}, Ksize=95), and the best 74.2920(λ_{l0} =2 $\times$ 10^{-3}, Ksize=55).



(h)Pristine image, blur image, the worst and best deblurring results of Pan[2]. The worst scores 51.8172(λ_{l0} =5 $\times$ 10^{-4}, Ksize=77 and the best 77.3829(λ_{l0} =2 $\times$ 10^{-3}, Ksize=49).



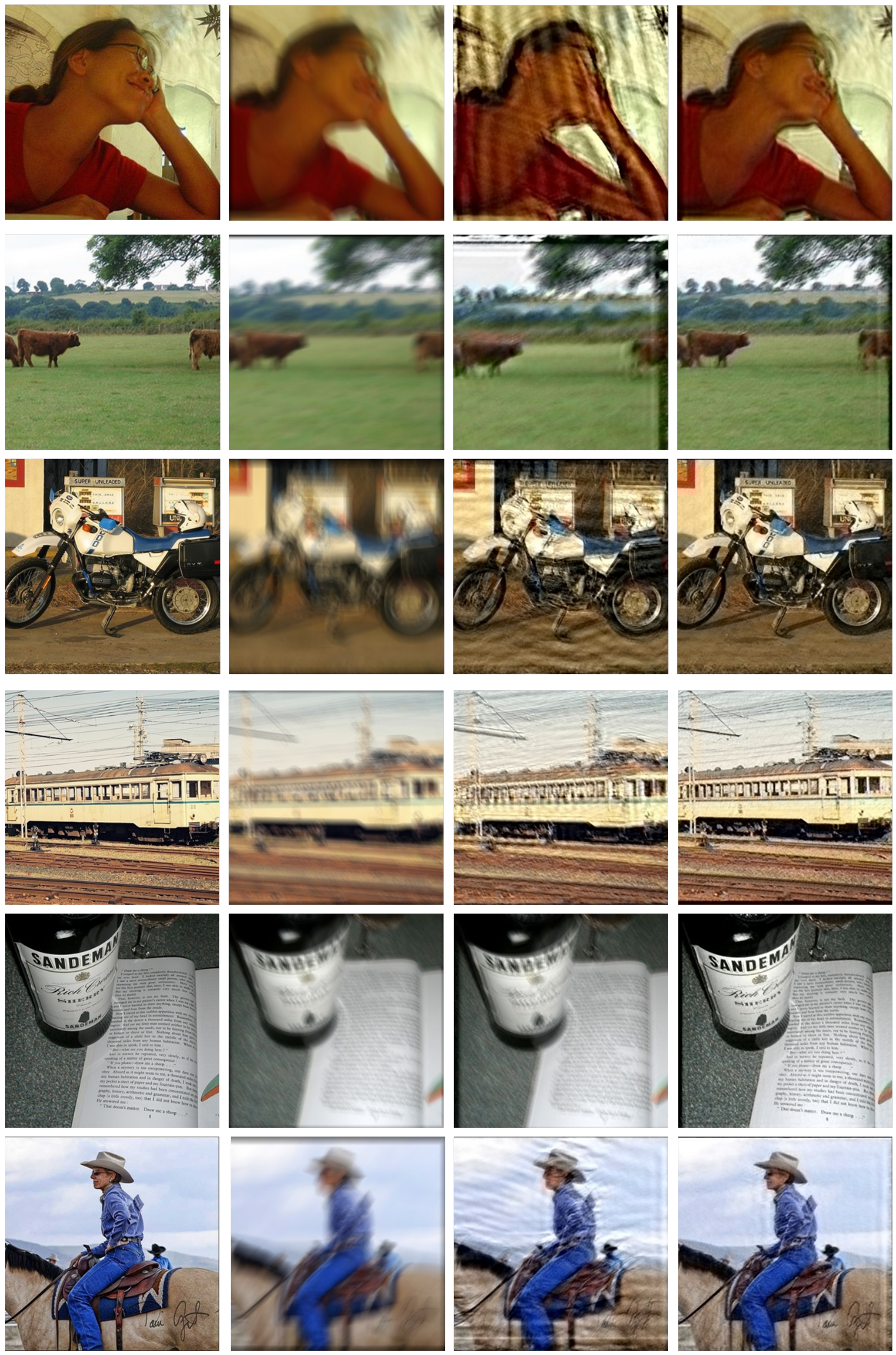
(i) Actually shot blur image, the worst and best deblurring results of Pan[2]. The worst scores 58.4930(λ_{l0} =1 $\times$ 10^{-3}, Ksize=95 and the best 79.0780(λ_{l0} =2 $\times$ 10^{-3}, Ksize=77).



(j)Actually shot blur image, the worst and best deblurring results of Pan[2]. The worst scores 61.2668(λ_{l0} =2 $\times$ 10^{-3}, Ksize=95 and the best 70.9147(λ_{l0} =5 $\times$ 10^{-4}, Ksize=43).

(k)Actually shot blur image, the worst and best deblurring results of Pan[2]. The worst scores 65.9769(λ_{l0} =2 $\times$ 10^{-3}, Ksize=95 and the best 83.8776(λ_{l0} =2 $\times$ 10^{-3}, Ksize=41).

Fig. S3: Automatically selection of hyperparameters using DBQA-NET.



(a) (b) (c) (d)

Fig. S4: The image quality improvement brought by Eq. 7. (a) Pristine images. (b) Blur images. (c) Results obtained without Eq. 7. (d) Results obtained with Eq. 7.

References

S1. Ma, K., Liu, W., Zhang, K., et al.: End-to-end blind image quality assessment using deep neural networks. IEEE Transactions on Image Processing (TIP), 27(3), 1202-1213 (2017)
S2. Rs R., Cogswell M., Das A., et al.: Grad-cam: Visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE international conference on computer vision, pp. 618-626. (2017)