Supplementary Material for
"Attention-based Adaptive Selection of Operations for Image Restoration in the Presence of Unknown Combined Distortions"

Masanori Suganuma[1,2] Xing Liu[1] Takayuki Okatani[1,2]
[1]Graduate School of Information Sciences, Tohoku University [2]RIKEN Center for AIP
{suganuma,ryu,okatani}@vision.is.tohoku.ac.jp

This document provides additional experimental results.

A. Comparison with Existing Methods Dedicated to Single Types of Distortion

In this study, we consider restoration of images with combined distortion of multiple types. To further analyze effectiveness of the proposed method, we compare it with existing methods dedicated to single types of distortion on their target tasks, i.e., single-distortion image restoration. To be specific, for the four single-distortion tasks, i.e., removal of Gaussian noise, motion blur, JPEG artifacts, and raindrops, we compare the proposed method with existing methods that are designed for each individual task and trained on the corresponding (single-distortion) dataset. On the other hand, the proposed model (the same as the one explained in the main paper) is trained on the combined distortion dataset explained in Sec. 4.4.1. Then, they are tested on the test splits of the same single-distortion datasets.

As this setup is favorable for the dedicated methods, they are expected to yield better results. The purpose of this experiment is to understand how large the differences will be.

A.1. Noise Removal

Experimental Configuration We use the DIV2K dataset for the base image set. We first cropped 128×128 pixel patches from the training and validation sets of the DIV2K dataset. Then we added Gaussian noise to them, yielding a training set and a testing set which consist of $50,000$ and $1,000$ patches, respectively. The standard deviation of the Gaussian noise is randomly chosen from the range of $[10, 20]$. Using these datasets, we compare our method with DnCNN [9], FFDNet [10], and E-CAE [5], which are the state-of-the-art dedicated models for this task.

Results Table 4 shows the results. As expected, the proposed method is inferior to the state-of-the-art methods by a certain margin. Although the margin is quantitatively not small considering the differences among the state-of-the-art methods, qualitative differences are not so large; an example is shown in Fig. 9.

Table 4. Results on the Gaussian noise removal.

Method	PSNR	SSIM
DnCNN [9]	34.38	0.9289
FFDNet [10]	34.90	0.9355
E-CAE [5]	35.08	0.9365
Ours	31.49	0.8972
Ours*	35.20	0.9381

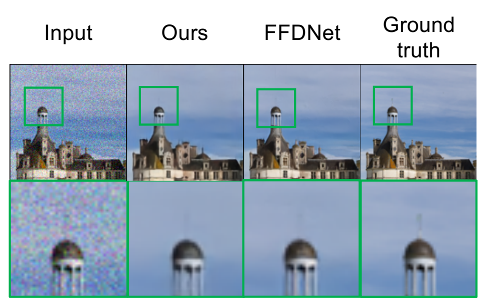

Figure 9. Examples of noise removal by our method and FFDNet [10].

To evaluate the potential of the proposed method, we also report the performance of our model trained on the single-distortion training data, which is referred to as 'Ours*' in the table. It is seen that it achieves similar or even better accuracy. This proves the potential of the proposed model as well as implies its usefulness in the scenario where there is only a single but *unidentified* type of distortion in images for which training data are available.

A.2. JPEG Artifacts Removal

Experimental Configuration We follow the same procedure as noise removal to construct a training set and a test set. The quality of the JPEG compression is randomly cho-

Table 5. Results of JPEG artifact removal.

Method	PSNR	SSIM
DnCNN [9]	31.24	0.8827
MemNet [7]	30.85	0.8785
Ours	29.87	0.8684
Ours*	31.64	0.8902

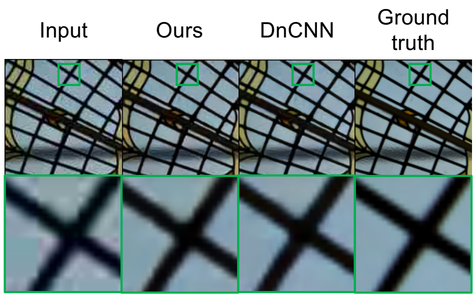

Figure 10. Examples of JPEG artifact removal by our method and DnCNN [9].

Table 6. Results of motion blur removal.

Method	PSNR	SSIM
Sun *et al.* [6]	24.6	0.842
Nah *et al.* [3]	29.1	0.916
Xu *et al.* [8]	25.1	0.890
DeblurGAN (*Synth*) [2]	23.6	0.884
DeblurGAN (*Wild*) [2]	27.2	0.954
DeblurGAN (*Comb*) [2]	28.7	0.958
Ours	25.4	0.793

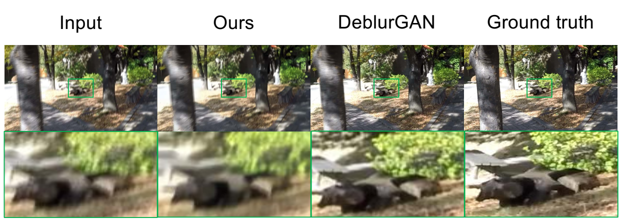

Figure 11. Examples of motion blur removal by our method and DeblurGAN (*Wild*) [2].

sen from the range of $[15, 35]$. We compare our method with two single distortion methods, DnCNN [9] and MemNet [7].

Results Table 5 shows the results. We can observe the same tendency as noise removal with smaller qualitative differences between our method and the dedicated methods. Their qualitative gap is further smaller; examples of restored images are shown in Fig.10. Our method trained on the same single distortion dataset ('Ours*') achieves better results with noticeable differences in this case.

Table 7. Results on the raindrop removal.

Method	PSNR	SSIM
Attentive GAN [4]	31.57	0.9023
Attentive GAN (*w/o D*) [4]	29.25	0.7853
Ours	28.57	0.8878

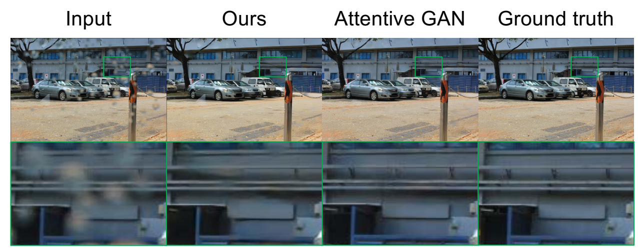

Figure 12. Examples of raindrop removal by our method and Attentive GAN [4].

A.3. Blur Removal

Experimental Configuration We used the GoPro dataset [3] for evaluation. It consists of 2,013 training and 1,111 test pairs of blurred and sharp images. We compare our model, which is trained on the combined distortion dataset, to the state-of-the-art method of [2]. In [2], the authors provide several versions of their method; DeblurGAN (*Synth*), DeblurGAN (*Wild*), and DeblurGAN (*Comb*). DeblurGAN (*Synth*) indicates a version of their model trained on a synthetic dataset generated by the method of [1], which is also employed in our experiments. Note that the dataset for training DeblurGAN (*Synth*) is generated from images of the MS COCO dataset, whereas the dataset for training our model is generated from the Raindrop dataset [4] and moreover it contains combined distortion. DeblurGAN (*Wild*) indicates a version of their model trained on random crops from the GoPro training set [3]. DeblurGAN (*Comb*) is a model trained on the both datasets.

Results Table 6 shows the results for the three variants of DeblurGAN along with three other existing methods, where their PSNR and SSIM values are copied from [2]. It is seen that our method is comparable to the existing ones, in particular earlier methods, in terms of PSNR but is inferior in terms of SSIM by a large margin. In fact, qualitative difference between our method and DeblurGAN (*Wild*) is large; see Fig.11.

However, we think that some of the gap can be explained by the difference in training data. DeblurGAN (*Wild*) is trained using the dataset lying in the same domain as the test data. On the other hand, our model is trained only on synthetic data, which must have different distribution from the GoPro test set. Thus, it may be fairer to make qualitative comparison with DeblurGAN (*Synth*), but we do not do so here, since its implementation is unavailable.

A.4. Raindrop Removal

Experimental Configuration We use the Raindrop dataset [4], which contains two test sets called TestA and TestB; the former is a subset of the latter. We use TestA for evaluation following [4]. We compare our model, which is trained on the combined distortion dataset as above, against the state-of-the-art method, Attentive GAN [4].

Results Table 7 shows the results. Attentive GAN (*w/o D*) is a variant that is trained without a discriminator and with only combined loss functions of the mean squared error and the perceptual loss etc. Our model achieves slightly lower accuracy than Attentive GAN (*w/o D*) in terms of PSNR and slightly lower accuracy than Attentive GAN in terms of SSIM. Figure 12 shows example images obtained by our method and Attentive GAN. Although there is noticeable difference between the images generated by the two methods, it is fair to say that our method yields reasonably good result, considering the fact that *our model can handle other types of distortion as well.*

B. More Results of Object Detection From Distorted Images

We have shown a few examples of object detection on PASCAL VOC in Fig.5 of the main paper. We provide more examples in Fig.13. They demonstrate that our method is able to remove combined distortions effectively, which contributes to the improvement of detection accuracy.

References

[1] G. Boracchi and A. Foi. Modeling the performance of image restoration from motion blur. *IEEE Transactions on Image Processing*, 21(8):3502–3517, 2012. 2

[2] O. Kupyn, V. Budzan, M. Mykhailych, D. Mishkin, and J. Matas. Deblurgan: Blind motion deblurring using conditional adversarial networks. In *CVPR*, 2018. 2

[3] S. Nah, T. H. Kim, and K. M. Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *CVPR*, 2017. 2

[4] R. Qian, R. T. Tan, W. Yang, J. Su, and J. Liu. Attentive generative adversarial network for raindrop removal from a single image. In *CVPR*, 2018. 2, 3

[5] M. Suganuma, M. Ozay, and T. Okatani. Exploiting the potential of standard convolutional autoencoders for image restoration by evolutionary search. In *ICML*, 2018. 1

[6] J. Sun, W. Cao, Z. Xu, and J. Ponce. Learning a convolutional neural network for non-uniform motion blur removal. In *CVPR*, 2015. 2

[7] Y. Tai, J. Yang, X. Liu, and C. Xu. Memnet: A persistent memory network for image restoration. In *CVPR*, 2017. 2

[8] L. Xu, S. Zheng, and J. Jia. Unnatural L0 sparse representation for natural image deblurring. In *CVPR*, 2013. 2

[9] K. Zhang., W. Zuo, Y. Chen, D. Meng, and L. Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Transactions on Image Processing*, 26(7):3142–3155, 2017. 1, 2

[10] K. Zhang, W. Zuo, and L. Zhang. Ffdnet: Toward a fast and flexible solution for cnn based image denoising. *IEEE Transactions on Image Processing*, 27(9):4608–4622, 2018. 1

Figure 13. Examples of object detection results on PASCAL VOC. The box colors indicate class categories.