Supplementary Material for
Comprehensive and Delicate: An Efficient Transformer for Image Restoration

Haiyu Zhao*, Yuanbiao Gou*, Boyun Li, Dezhong Peng, Jiancheng Lv, Xi Peng†
College of Computer Science, Sichuan University.
{haiyuzhao.gm, gouyuanbiao, liboyun.gm, pengx.gm}@gmail.com;
{pengdz, lvjiancheng}@scu.edu.cn

1. Introduction

In this material, we present more experiment results and qualitative comparisons to show the effectiveness of our CODE and discuss the broader impact. From the visual results, one could see that CODE recovers image structures better and image details finer, thus obtaining clearer restorations.

2. Ablation for Aggregation Factors

In addition to the analysis experiments presented in the main body of the paper, here we show more results about r_c and r_s. The experimental settings are the same as other analysis experiments, which are conducted on Set12 with noise level of 50. As shown in Tabs. 1 and 2, our method is not sensitive to r_c and r_s. Increasing r_c or decreasing r_s only leads to slight performance drop.

Table 1. Ablation experiment for the aggregation factor r_c.

r_c	#Params	FLOPs	PSNR	SSIM
2	18.67M	25.84G	27.92	0.8077
4	12.18M	22.44G	27.93	0.8083
8	8.94M	20.75G	27.90	0.8074

Table 2. Ablation experiment for the aggregation factor r_s.

r_s	#Params	FLOPs	PSNR	SSIM
$[16, 8, 4, 2]$	9.51M	22.10G	27.91	0.8072
$[32, 16, 8, 4]$	12.18M	22.44G	27.93	0.8083
$[64, 32, 16, 4]$	17.40M	23.11G	27.92	0.8079

3. Qualitative Results on Image Denoising

In addition to the qualitative results presented in the main body of the paper, we show more results on grayscale and color image denoising at noise level 15, 25 and 50. As shown in Figs. 1 to 5, CODE demonstrates better visual results on grayscale image denoising compared with the other methods. To be specific, DnCNN [11] and FFDNet [12] leave noises and artifacts in the images, DRUNet [10] and SwinIR [4] obtain oversmoothed results and lose the details in the door and window. However, our CODE fully recovers the structures of the door and the window while preserving the textures better. Through careful observation, one can find that our visual results are the closest ones to GTs. Figs. 8 to 10 and Figs. 11 to 13 respectively show the qualitative results of color image denoising on Kodak24 [1] and McMaster [13]. From the figures, one can observe that DnCNN and FFDNet have residual noises and distortions, DRUNet and SwinIR smooth out the image details, thus obtaining the blurry images. In contrast, our CODE adequately removes the noises without introducing artifacts, while achieving better restoration on details and textures.

4. Qualitative Results on Motion Deblurring

In this section, we show the qualitative results on motion deblurring. Fig. 14 and Fig. 15 show the visual results on GoPro [5] test set and HIDE [6] dataset, respectively. From the figures, one can observe that DGAN [2] and DGAN-v2 [3] have residual black streaks and artifacts, SRN [8], DMPHN [9], and SAPHNet [7] obtain blurry results with distortions. By contrast, our CODE recovers better structures and more details with much less residual blur in the restored images. Such an observation is more obvious in Fig. 14, where CODE obtains almost the same result as GT.

5. Qualitative Results on JPEG Compression Artifact Reduction

Figs. 16 to 19 and Figs. 20 to 23 show the qualitative results on JPEG compression artifact reduction at compression level 10, 20, 30, and 40. From the figures, one can observe that DnCNN has residual artifacts, DRUNet, SwinIR, and CODE all obtain visually pleasant results with fine textures and structures, in which our CODE is much more ef-

*indicates equal contribution.
†Corresponding author.

ficient than the other two methods. Specifically, CODE has $\sim$ 37% parameters and $\sim$31% FLOPs of DRUNet, and only $\sim$6% FLOPs of SwinIR. In other words, our CODE can provide similar results with much less computation and/or memory costs.

6. Broader Impact

In this section, we discuss the impact of our CODE in a broader vision. Generally, CODE is much more efficient than other competing image restoration Transformers, and thus has more opportunities to be deployed on mobile devices. However, CODE is a general neural network and might be trained with customized data and used for unauthorized purposes, such as watermark removal, which might prejudice the rights of others. Moreover, the training and testing of the model consume a lot of electricity, which causes carbon emissions.

References

[1] Rich Franzen. Kodak lossless true color image suite. *source: http://r0k. us/graphics/kodak*, 4(2), 1999. 1

[2] Orest Kupyn, Volodymyr Budzan, Mykola Mykhailych, Dmytro Mishkin, and Jiří Matas. Deblurgan: Blind motion deblurring using conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8183–8192, 2018. 1

[3] Orest Kupyn, Tetiana Martyniuk, Junru Wu, and Zhangyang Wang. Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8878–8887, 2019. 1

[4] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. *arXiv preprint arXiv:2108.10257*, 2021. 1

[5] Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3883–3891, 2017. 1

[6] Ziyi Shen, Wenguan Wang, Xiankai Lu, Jianbing Shen, Haibin Ling, Tingfa Xu, and Ling Shao. Human-aware motion deblurring. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5572–5581, 2019. 1

[7] Maitreya Suin, Kuldeep Purohit, and AN Rajagopalan. Spatially-attentive patch-hierarchical network for adaptive motion deblurring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3606–3615, 2020. 1

[8] Xin Tao, Hongyun Gao, Xiaoyong Shen, Jue Wang, and Jiaya Jia. Scale-recurrent network for deep image deblurring. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8174–8182, 2018. 1

[9] Hongguang Zhang, Yuchao Dai, Hongdong Li, and Piotr Koniusz. Deep stacked hierarchical multi-patch network for image deblurring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5978–5986, 2019. 1

[10] Kai Zhang, Yawei Li, Wangmeng Zuo, Lei Zhang, Luc Van Gool, and Radu Timofte. Plug-and-play image restoration with deep denoiser prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021. 1

[11] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE transactions on image processing*, 26(7):3142–3155, 2017. 1

[12] Kai Zhang, Wangmeng Zuo, and Lei Zhang. Ffdnet: Toward a fast and flexible solution for cnn-based image denoising. *IEEE Transactions on Image Processing*, 27(9):4608–4622, 2018. 1

[13] Lei Zhang, Xiaolin Wu, Antoni Buades, and Xin Li. Color demosaicking by local directional interpolation and nonlocal adaptive thresholding. *Journal of Electronic imaging*, 20(2):023016, 2011. 1

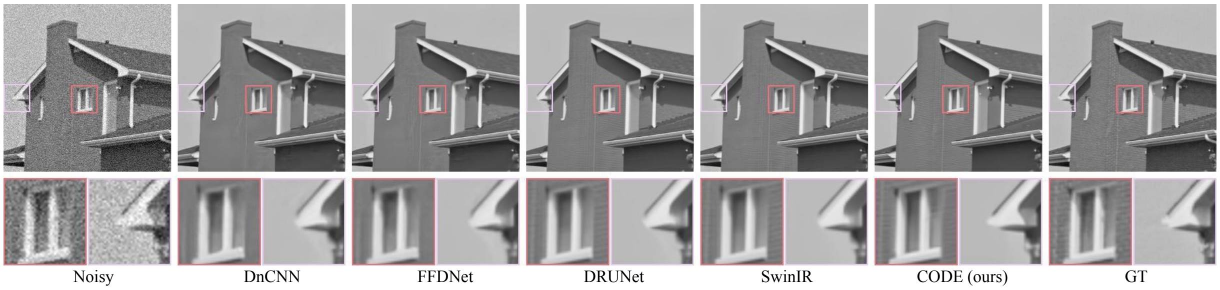

Figure 1. Qualitative comparisons of **grayscale image denoising** (noise level 15) on the Set12 dataset.

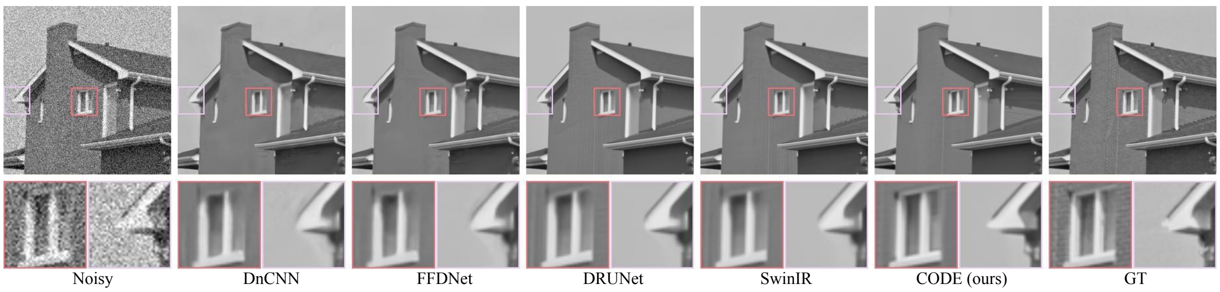

Figure 2. Qualitative comparisons of **grayscale image denoising** (noise level 25) on the Set12 dataset.

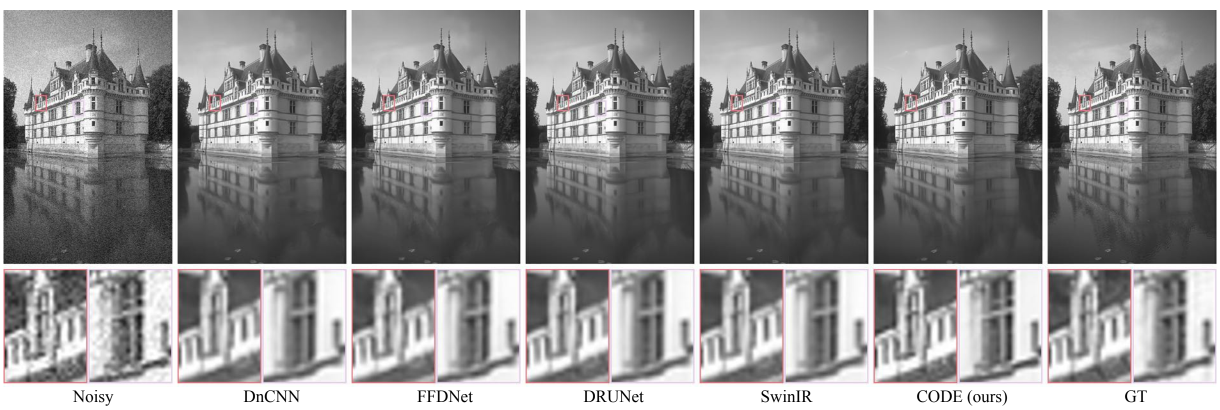

Figure 3. Qualitative comparisons of **grayscale image denoising** (noise level 15) on the BSD68 dataset.

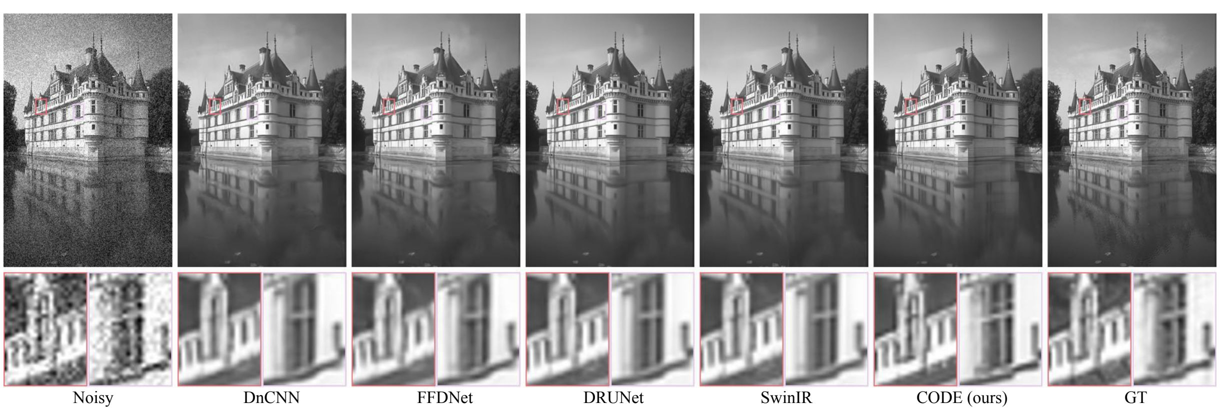

Figure 4. Qualitative comparisons of **grayscale image denoising** (noise level 25) on the BSD68 dataset.

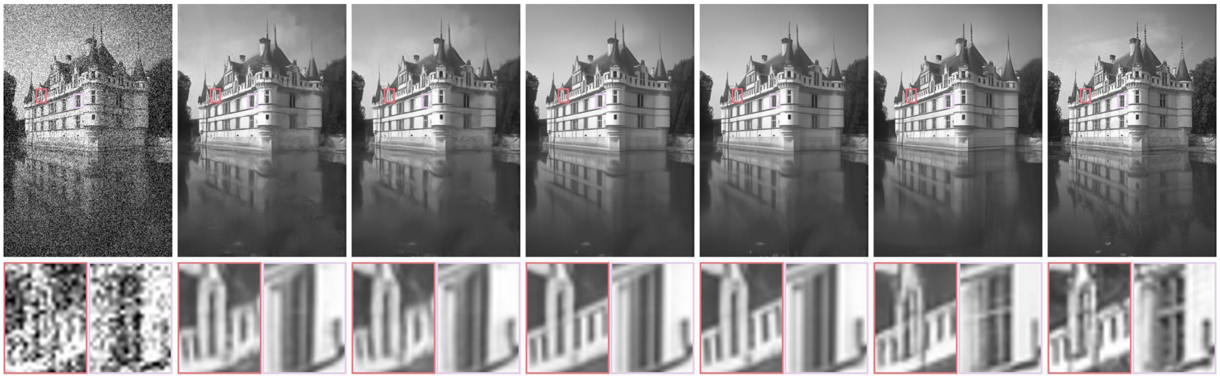

Figure 5. Qualitative comparisons of **grayscale image denoising** (noise level 50) on the BSD68 dataset.

Figure 6. Qualitative comparisons of **color image denoising** (noise level 15) on the CBSD68 dataset.

Figure 7. Qualitative comparisons of **color image denoising** (noise level 25) on the CBSD68 dataset.

Figure 8. Qualitative comparisons of **color image denoising** (noise level 15) on the Kodak24 dataset.

Figure 9. Qualitative comparisons of **color image denoising** (noise level 25) on the Kodak24 dataset.

Figure 10. Qualitative comparisons of **color image denoising** (noise level 50) on the Kodak24 dataset.

Figure 11. Qualitative comparisons of **color image denoising** (noise level 15) on the McMaster dataset.

Figure 12. Qualitative comparisons of **color image denoising** (noise level 25) on the McMaster dataset.

Figure 13. Qualitative comparisons of **color image denoising** (noise level 50) on the McMaster dataset.

Figure 14. Qualitative comparisons of **motion deblurring** on the GoPro test set.

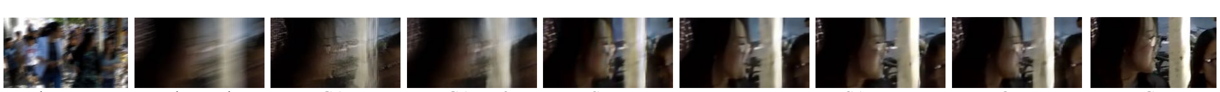

Figure 15. Qualitative comparisons of **motion deblurring** on the HIDE dataset.

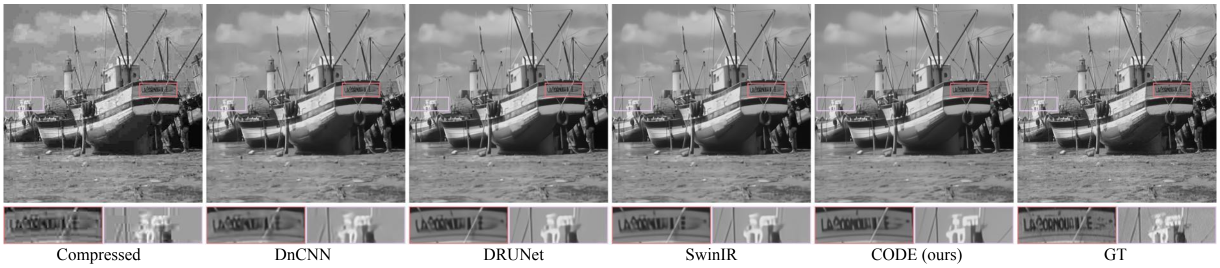

Figure 16. Qualitative comparisons of **JPEG compression artifact reduction** (compression level $q = 10$) on the Classic5 dataset.

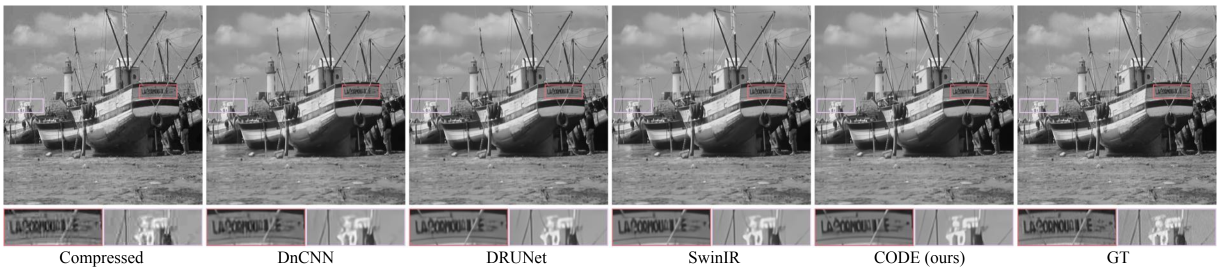

Figure 17. Qualitative comparisons of **JPEG compression artifact reduction** (compression level $q = 20$) on the Classic5 dataset.

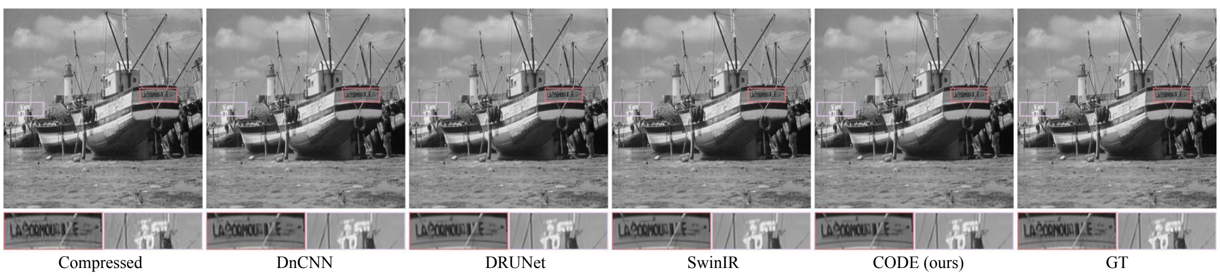

Figure 18. Qualitative comparisons of **JPEG compression artifact reduction** (compression level $q = 30$) on the Classic5 dataset.

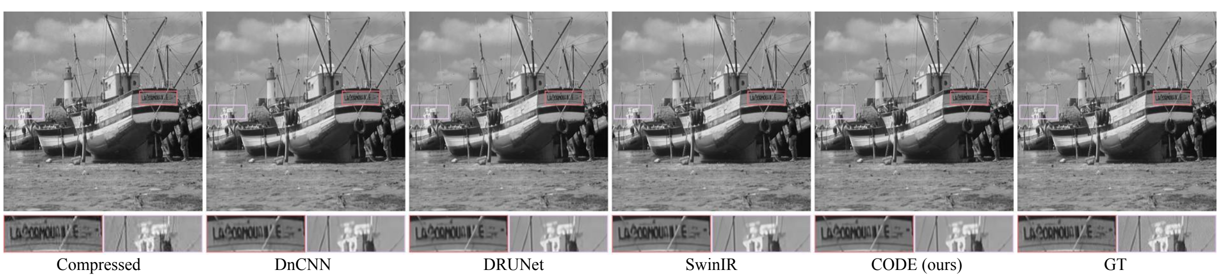

Figure 19. Qualitative comparisons of **JPEG compression artifact reduction** (compression level $q = 40$) on the Classic5 dataset.

Figure 20. Qualitative comparisons of **JPEG compression artifact reduction** (compression level $q = 10$) on the LIVE1 dataset.

Compressed DnCNN DRUNet SwinIR CODE (ours) GT

Figure 21. Qualitative comparisons of **JPEG compression artifact reduction** (compression level $q = 20$) on the LIVE1 dataset.

Compressed DnCNN DRUNet SwinIR CODE (ours) GT

Figure 22. Qualitative comparisons of **JPEG compression artifact reduction** (compression level $q = 30$) on the LIVE1 dataset.

Compressed DnCNN DRUNet SwinIR CODE (ours) GT

Figure 23. Qualitative comparisons of **JPEG compression artifact reduction** (compression level $q = 40$) on the LIVE1 dataset.