APPENDIX
ADDITIONAL EXPERIMENTAL RESULTS

This appendix provides supplementary quantitative and qualitative comparisons to further validate the performance of HTDNet.

A. Supplementary Quantitative Analysis

Due to space constraints in the main manuscript, Table III provided only the average metrics. Table A-I details the per-image PSNR results on the Set12 dataset for three noise levels ($\sigma = 15, 25, 50$). HTDNet consistently achieves top-tier performance across varying textures and noise intensities.

To evaluate the effectiveness of our heterogeneous architecture against pure attention mechanisms, we conduct comparisons with DRANet [15], MLFAN [16], and GradNet [17]. Table A-II reports the PSNR results on the CBD68, Kodak24, and McMaster datasets. HTDNet consistently outperforms these methods across all noise levels, demonstrating that synergizing CNNs with Transformers yields superior robustness compared to relying solely on attention modules.

B. Additional Visual Comparisons

Fig. B presents an additional visual comparison on the Urban100 dataset. The accompanying heat map visualizes the salient features captured by our Transformer module, highlighting its ability to focus on high-frequency structural details.

TABLE A-I
AVERAGE PSNR (DB) RESULTS OF DIFFERENT METHODS ON 11 IMAGES FROM SET12 WITH NOISE LEVELS OF 15, 25 AND 50.

Images	C.man	House	Peppers	Starfish	Monarch	Airplane	Parrot	Barbara	Boat	Man	Couple	Average
Noise Level	$\sigma = 15$											
BM3D [1]	31.91	34.93	32.69	31.14	31.85	31.07	31.37	33.10	32.13	31.92	32.10	32.20
WNNM [2]	32.17	35.13	32.99	31.82	32.71	31.39	31.62	33.60	32.27	32.11	32.17	32.54
EPLL [3]	31.85	34.17	32.64	31.13	32.10	31.19	31.42	31.38	31.93	32.00	31.93	31.98
TNRD [4]	32.19	34.53	33.04	31.75	32.56	31.46	31.63	32.13	32.14	32.23	32.11	32.34
DnCNN [5]	32.61	34.97	33.30	32.20	33.09	31.70	31.83	32.64	32.42	32.46	32.47	32.70
IRCNN [6]	32.55	34.89	33.31	32.02	32.82	31.70	31.84	32.43	32.34	32.40	32.40	32.61
FFDNet [7]	32.43	35.07	33.25	31.99	32.66	31.57	31.81	32.54	32.38	32.41	32.46	32.60
ECNDNet [8]	32.56	34.97	33.25	32.17	33.11	31.70	31.82	32.41	32.37	32.39	32.39	32.65
SANet [9]	32.38	35.03	33.18	32.14	33.20	31.71	31.89	32.61	32.36	32.38	32.41	32.66
PSN-U [10]	32.04	35.03	33.21	31.94	32.93	31.61	31.62	32.49	32.41	32.37	32.43	32.55
RDDCNN [11]	32.61	35.01	33.31	32.13	33.13	31.67	31.93	32.62	32.42	32.38	32.46	32.70
DudeNet [12]	32.71	35.13	33.38	32.29	33.28	31.78	31.93	32.73	32.46	32.46	32.49	32.79
ADNet [13]	32.81	35.22	33.49	32.17	33.17	31.86	31.96	32.80	32.57	32.47	32.58	32.83
BRDNet [14]	32.80	35.27	33.47	32.24	33.35	31.85	32.00	32.93	32.55	32.50	32.62	32.87
HTDNet (Ours)	**32.90**	**35.74**	**33.66**	**32.63**	**33.55**	**32.06**	**32.18**	**33.62**	**32.72**	**32.62**	**32.75**	**33.13**
Noise Level	$\sigma = 25$											
BM3D	29.45	32.85	30.16	28.56	29.25	28.42	28.93	30.71	29.90	29.61	29.71	29.78
WNNM	29.64	33.22	30.42	29.03	29.84	28.69	29.15	31.24	30.03	29.76	29.82	30.08
EPLL	29.26	32.17	30.17	28.51	29.39	28.61	28.95	28.61	29.74	29.66	29.53	29.51
MLP	29.61	32.56	30.30	28.82	29.61	28.82	29.25	29.54	29.97	29.88	29.73	29.83
TNRD	29.72	32.53	30.57	29.02	29.85	28.88	29.18	29.41	29.91	29.87	29.71	29.88
DnCNN	30.18	33.06	30.87	29.41	30.28	29.13	29.43	30.00	30.21	30.10	30.12	30.25
IRCNN	30.08	33.06	30.88	29.27	30.09	29.12	29.47	29.92	30.17	30.04	30.08	30.20
FFDNet	30.10	33.28	30.93	29.32	30.08	29.04	29.44	30.01	30.25	30.11	30.20	30.25
ECNDNet	30.11	33.08	30.85	29.43	30.30	29.07	29.38	29.84	30.14	30.03	30.03	30.21
SANet	30.04	33.05	30.83	29.31	30.27	29.08	29.34	30.00	30.12	30.00	30.05	30.19
PSN-U	29.79	33.23	30.90	29.30	30.17	29.06	29.25	29.94	30.25	30.05	30.12	30.19
RDDCNN	30.20	33.13	30.82	29.38	30.36	29.05	29.53	30.03	30.19	30.05	30.10	30.26
DudeNet	30.23	33.24	30.98	29.53	30.44	29.14	29.48	30.15	30.24	30.08	30.15	30.33
ADNet	30.34	33.41	31.14	29.41	30.39	29.17	29.49	30.25	30.37	30.08	30.24	30.39
BRDNet	**31.39**	33.41	31.04	29.46	30.50	29.20	29.55	30.34	30.33	30.14	30.28	30.51
HTDNet (Ours)	30.56	**33.80**	**31.34**	**30.04**	**30.85**	**29.47**	**29.74**	**31.41**	**30.50**	**30.27**	**30.46**	**30.77**
Noise Level	$\sigma = 50$											
BM3D	26.13	29.69	26.68	25.04	25.82	25.10	25.90	27.22	26.78	26.81	26.46	26.51
WNNM	26.45	30.33	26.95	25.44	26.32	25.42	26.14	27.79	26.97	26.94	26.64	26.85
EPLL	26.10	29.12	26.80	25.12	25.94	25.31	25.95	24.83	26.74	26.79	26.30	26.27
MLP	26.37	29.64	26.68	25.43	26.26	25.56	26.12	25.24	27.03	27.06	26.67	26.55
TNRD	26.62	29.48	27.10	25.42	26.31	25.59	26.16	25.70	26.94	26.98	26.50	26.62
DnCNN	27.03	30.00	27.32	25.70	26.78	25.87	26.48	26.22	27.20	27.24	26.90	26.98
IRCNN	26.88	29.96	27.33	25.57	26.61	25.89	26.55	26.24	27.17	27.17	26.88	26.93
FFDNet	27.05	30.37	27.54	25.75	26.81	25.89	26.57	26.45	27.33	27.29	27.08	27.10
ECNDNet	27.07	30.12	27.30	25.72	26.82	25.79	26.32	26.26	27.16	27.11	26.84	26.96
SANet	26.92	29.93	27.27	25.52	26.64	25.71	26.18	26.37	27.20	27.11	26.80	26.88
PSN-U	27.21	30.21	27.53	25.63	26.93	25.89	26.62	26.56	27.27	27.23	27.04	27.10
RDDCNN	27.16	30.21	27.38	25.72	26.84	25.88	26.53	26.36	27.23	27.22	26.88	27.04
DudeNet	27.22	30.27	27.51	25.88	26.93	25.88	26.50	26.49	27.26	27.19	26.97	27.10
ADNet	27.31	30.59	27.69	25.70	26.90	25.88	26.56	26.64	27.35	27.17	27.07	27.17
BRDNet	**27.44**	30.53	27.67	25.77	26.97	25.93	26.66	26.85	27.38	27.27	27.17	27.24
HTDNet (Ours)	27.43	**30.99**	**27.75**	**26.67**	**27.30**	**26.18**	**26.84**	**27.96**	**27.50**	**27.39**	**27.38**	**27.58**

TABLE A-II

AVERAGE PSNR (DB) OF DIFFERENT METHODS BASED ON ATTENTIONS ON VARIOUS DATASETS WITH DIFFERENT NOISE LEVELS.

Methods	CBD68			Kodak24			McMaster		
Noise Levels	σ = 15	σ = 25	σ = 50	σ = 15	σ = 25	σ = 50	σ = 15	σ = 25	σ = 50
DRANet [15]	34.18	31.56	28.37	35.02	32.59	29.50	35.09	32.84	29.77
MLFAN [16]	34.06	31.37	28.11	34.85	32.33	29.15	35.08	32.68	29.47
GradNet [17]	34.07	31.39	28.12	34.85	32.35	29.23	34.81	32.45	29.39
MWDCNN [18]	34.18	31.45	28.13	34.91	32.40	29.26	/	/	/
HTDNet (Ours)	**34.36**	**31.70**	**28.46**	**35.26**	**32.70**	**29.64**	**35.49**	**33.15**	**29.99**

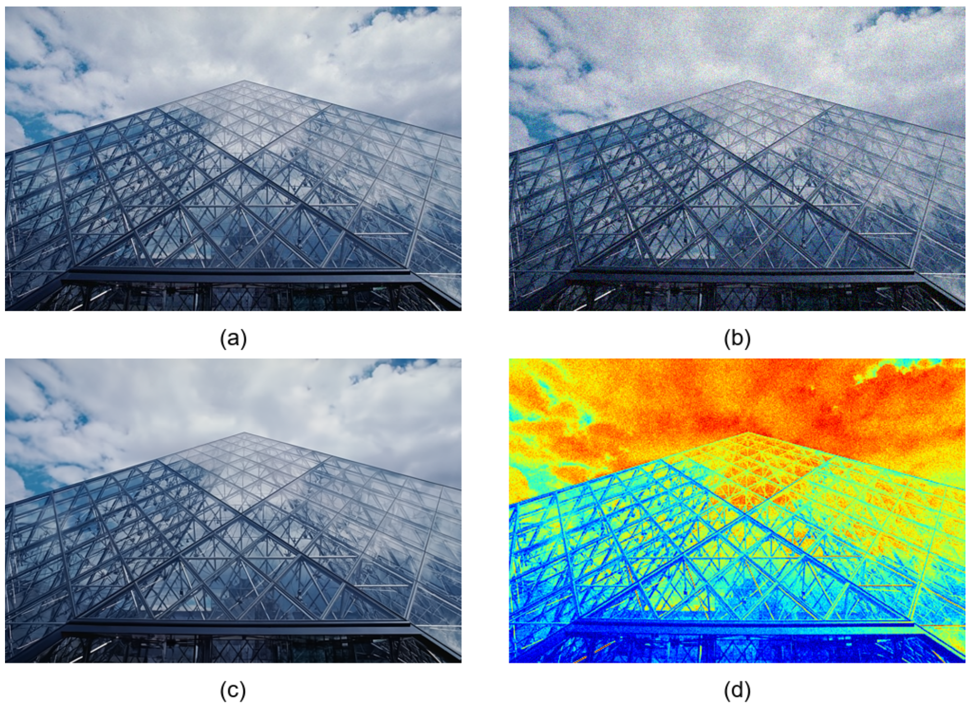

Fig. B. Visual comparison of one image from the Urban100 for a noise level of 15. (a) Clean Image, (b) Noisy Image, (c) Predicted clean image (Ours), and (d) Heat Map (Ours).

C. Enlarged Visual Comparisons from Main Manuscript

To address the concerns regarding the visibility of details in the main manuscript (due to page layout constraints), we provide enlarged, full-width versions of Figs. 2-4 in this section. These high-resolution images facilitate a clearer assessment of artifact suppression and fine texture preservation.

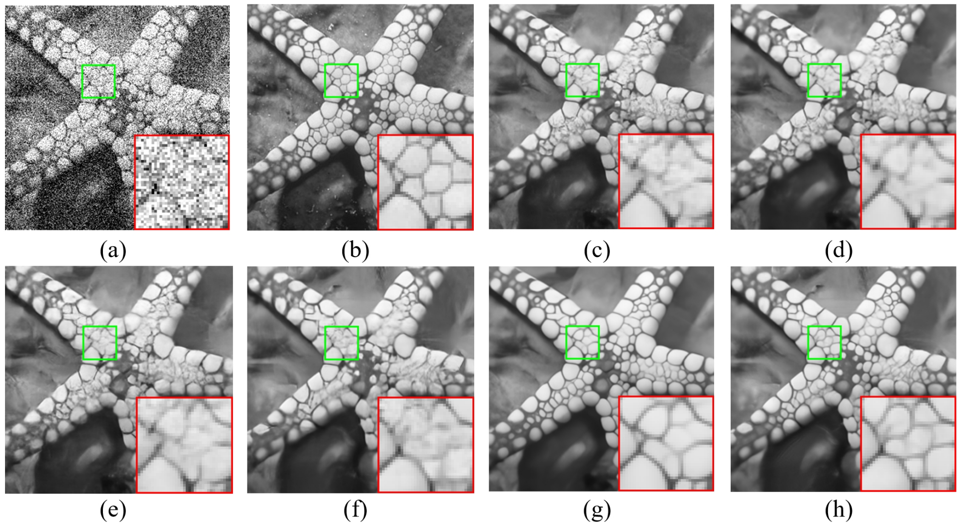

Fig. C-I. Visual denoising images of different methods on one image from Set12 for noise level of 50. (a) Noisy Image , (b) Clean Image, (c) DnCNN , (d) FFDNet , (e) IRCNN , (f) ADNet, (g) DRUNet and (h) HTDNet (Ours).

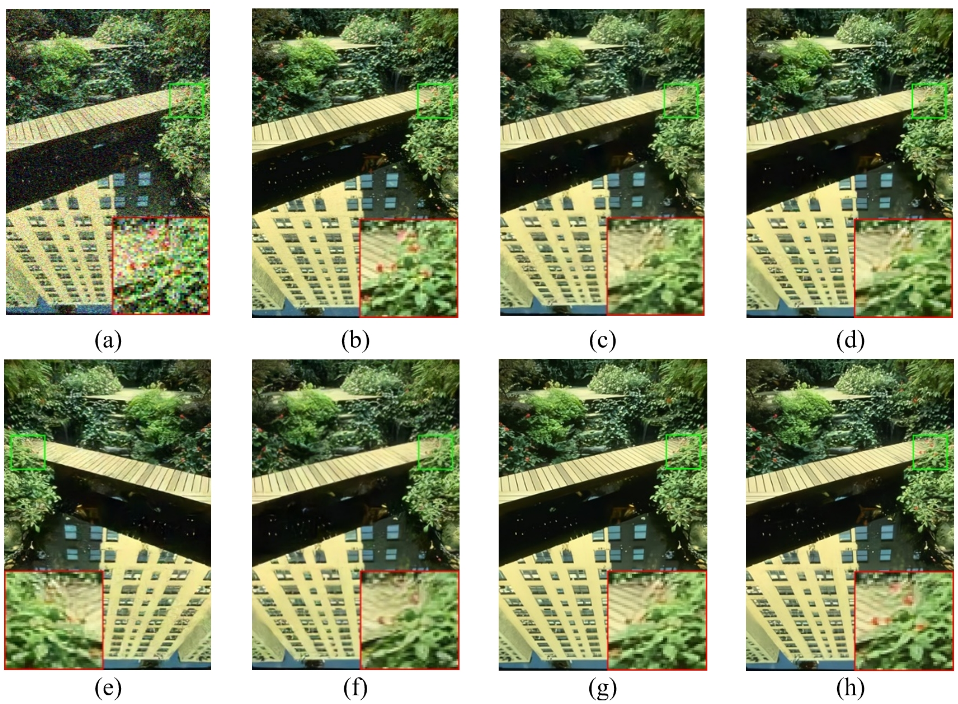

Fig. C-II. Visual denoising images of different methods on one image from CBDSD68 for noise level of 50. (a) Noisy Image, (b) Clean Image, (c) DnCNN, (d) FFDNet, (e) IRCNN, (f) ADNet, (g) DRUNet and (h) HTDNet (Ours).

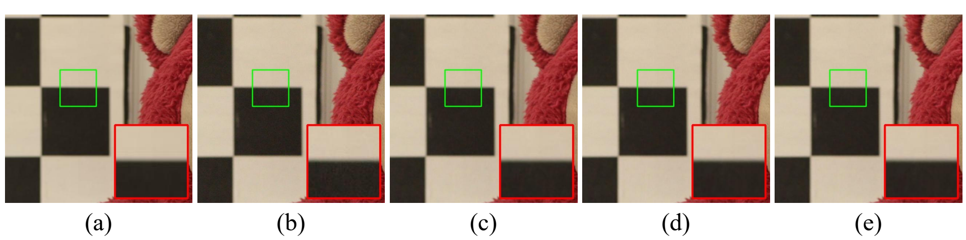

Fig. C-III. Visual denoising images of our method on three real-world noisy images from the CC dataset. (a) Noisy Image, (b) Clean Image, (c) ART, (d) Restormer and (e) HTDNet (Ours).

REFERENCES

[1] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising with block-matching and 3d filtering," in *Image processing: algorithms and systems, neural networks, and machine learning*, vol. 6064. SPIE, 2006, pp. 354–365.

[2] S. Gu, L. Zhang, W. Zuo, and X. Feng, "Weighted nuclear norm minimization with application to image denoising," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 2862–2869.

[3] D. Zoran and Y. Weiss, "From learning models of natural image patches to whole image restoration," in *2011 international conference on computer vision*. IEEE, 2011, pp. 479–486.

[4] Y. Chen and T. Pock, "Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 6, pp. 1256–1272, 2016.

[5] 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*, vol. 26, no. 7, pp. 3142–3155, 2017.

[6] K. Zhang, W. Zuo, S. Gu, and L. Zhang, "Learning deep cnn denoiser prior for image restoration," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 3929–3938.

[7] K. Zhang, W. Zuo, and L. Zhang, "Ffdnet: Toward a fast and flexible solution for cnn-based image denoising," *IEEE Transactions on Image Processing*, vol. 27, no. 9, pp. 4608–4622, 2018.

[8] C. Tian, Y. Xu, L. Fei, J. Wang, J. Wen, and N. Luo, "Enhanced cnn for image denoising," *CAAI Transactions on Intelligence Technology*, vol. 4, no. 1, pp. 17–23, 2019.

[9] L. Zhang, Y. Li, P. Wang, W. Wei, S. Xu, and Y. Zhang, "A separation–aggregation network for image denoising," *Applied Soft Computing*, vol. 83, p. 105603, 2019.

[10] R. Aljadaany, D. K. Pal, and M. Savvides, "Proximal splitting networks for image restoration," in *Image Analysis and Recognition: 16th International Conference, ICIAR 2019, Waterloo, ON, Canada, August 27–29, 2019, Proceedings, Part I 16*. Springer, 2019, pp. 3–17.

[11] Q. Zhang, J. Xiao, C. Tian, J. Chun-Wei Lin, and S. Zhang, "A robust deformed convolutional neural network (cnn) for image denoising," *CAAI Transactions on Intelligence Technology*, vol. 8, no. 2, pp. 331–342, 2023.

[12] C. Tian, Y. Xu, W. Zuo, B. Du, C.-W. Lin, and D. Zhang, "Designing and training of a dual cnn for image denoising," *Knowledge-Based Systems*, vol. 226, p. 106949, 2021.

[13] C. Tian, Y. Xu, Z. Li, W. Zuo, L. Fei, and H. Liu, "Attention-guided cnn for image denoising," *Neural Networks*, vol. 124, pp. 117–129, 2020.

[14] C. Tian, Y. Xu, and W. Zuo, "Image denoising using deep cnn with batch renormalization," *Neural Networks*, vol. 121, pp. 461–473, 2020.

[15] W. Wu, S. Liu, Y. Xia, and Y. Zhang, "Dual residual attention network for image denoising," *Pattern Recognition*, vol. 149, p. 110291, 2024.

[16] A. Ulu, G. Yildiz, and B. Dizdaroğlu, "Mlfan: Multilevel feature attention network with texture prior for image denoising," *IEEE Access*, vol. 11, pp. 34 260–34 273, 2023.

[17] Y. Liu, S. Anwar, L. Zheng, and Q. Tian, "Gradnet image denoising," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020, pp. 508–509.

[18] C. Tian, M. Zheng, W. Zuo, B. Zhang, Y. Zhang, and D. Zhang, "Multi-stage image denoising with the wavelet transform," *Pattern Recognition*, vol. 134, p. 109050, 2023.