Supplementary Material-Optimization-Inspired Cross-Attention Transformer for Compressive Sensing

Jiechong Song[1,4], Chong Mou[1], Shiqi Wang[2], Siwei Ma[3,4], Jian Zhang[1,4*]
[1]Peking University Shenzhen Graduate School, Shenzhen, China
[2]Department of Computer Science, City University of Hong Kong, China
[3]School of Computer Science, Peking University, Beijing, China
[4]Peng Cheng Laboratory, Shenzhen, China
{songjiechong, swma, zhangjian.sz} @pku.edu.cn eechongm@stu.pku.edu.cn shiqwang@cityu.edu.hk

In this supplementary material, we provide the following additional details to facilitate the understanding of our paper:

(1) We analyze the influence of channel numbers in multi-channel transmission and prove that our multi-channel inertial term maximizes its advantages.

(2) We analyze the effect of parameter sharing for the corresponding parts in different iterations, which shows that further compression in OCTUF parameters is possible with limited impact on reconstruction performance.

(3) We present the objective results of our methods and some recent methods on the larger DIV2K dataset [1] and provide more subjective results of our proposed methods to present our high performance compared with other models.

1. Comparison of Channel Number

To investigate the effect of the channel number of input and output in each iteration, we do experiments with the channel dimension $C \in \{8, 16, 32, 64\}$ in Tab. 1. It can be seen that the performance grows obviously when $C \leqslant 32$ and then maintains stability, which indicates that the multi-channel inertial term is affected by the number of channels. Therefore, we select $C = 32$ as our setting, considering the tradeoff between performance and complexity.

Table 1. Comparison of channel number on Set11 dataset [6] in the case of CS ratio $= 50\%$.

Cases	Channels	PSNR(dB)	SSIM	Parameters
(a)	8	40.68	0.9825	0.55 M
(b)	16	41.03	0.9831	0.61 M
(c)	32	41.34	0.9838	0.82 M
(d)	64	41.35	0.9838	1.61 M

2. Comparison of Weight-sharing

To study the difference in sharing the model parameters of each iteration, we evaluate the recovered images in different cases, as shown in Tab. 2. One iteration consists of a Dual-CA sub-module and an FFN sub-module assigned the shared strategy separately. As seen from Tab. 2, Case (a) denotes that the parameters in each iteration are shared, and Cases (b)(c) represent that only FFN or Dual-CA sub-module is shared, respectively. Our OCTUF (the iteration number $K = 10$) without the shared strategy, *i.e.*, Case (d), has the best reconstruction performance. Obviously, the parameter numbers of the reconstruction network in Case (a) are much fewer than in our setting (Case (d)), which indicates that further compression in OCTUF parameters is possible, with limited effect on reconstruction performance.

3. More Comparison Results

We summarize the average PSNR(dB)/SSIM performances on DIV2K [1] dataset in Tab. 3. DIV2K dataset contains 100 high-resolution images and consists of various pictures such as characters, scenery, buildings, animals, people, etc. From

Table 2. Comparison of weight-sharing on Set11 dataset [6]. The best performance is labeled in **bold**.

Cases	Sharing Part		Ratio= 25%		Ratio= 40%		Parameters (M)	
	Dual-CA	FFN	PSNR(dB)	SSIM	PSNR(dB)	SSIM	Sampling Matrix	Reconstruction Network
(a)	√	√	35.20	0.9550	38.56	0.9744	0.34	0.03
(b)	-	√	35.49	0.9572	39.06	0.9762	0.34	0.13
(c)	√	-	35.75	0.9588	39.22	0.9768	0.34	0.20
(d)	-	-	**36.10**	**0.9604**	**39.41**	**0.9773**	0.34	0.30

Tab. 3, we observe that our proposed OCTUFs achieve superior performance against the existing deep network-based CS schemes. The visual comparisons are shown in Fig. 1, from which we can see that OCTUFs can recover more texture information compared to the other deep network-based CS methods.

Table 3. Average PSNR(dB)/SSIM performance comparisons of recent deep network-based CS methods on DIV2K dataset [1] with different CS ratios. The best result is labeled in **bold**.

Dataset	Methods	CS Ratio				
		25%	30%	40%	50%	Average
DIV2K	COAST (TIP 2021) [10]	33.45/0.9178	34.49/0.9323	36.41/0.9530	38.22/0.9668	35.64/0.9425
	MADUN (ACM MM 2021) [8]	35.63/0.9499	36.82/0.9596	38.96/0.9727	40.76/0.9810	38.04/0.9658
	TransCS (TIP 2022) [7]	35.20/0.9450	36.01/0.9535	38.54/0.9702	40.65/0.9800	37.60/0.9622
	FSOINet (ICASSP 2022) [4]	35.75/0.9495	36.92/0.9593	39.09/0.9725	41.18/0.9808	38.24/0.9655
	OCTUF (Ours)	35.82/0.9499	37.01/**0.9597**	39.17/**0.9731**	41.32/0.9819	38.33/**0.9662**
	OCTUF$^+$ (Ours)	**35.86/0.9500**	**37.04/0.9597**	**39.21/0.9731**	**41.33/0.9820**	**38.36/0.9662**

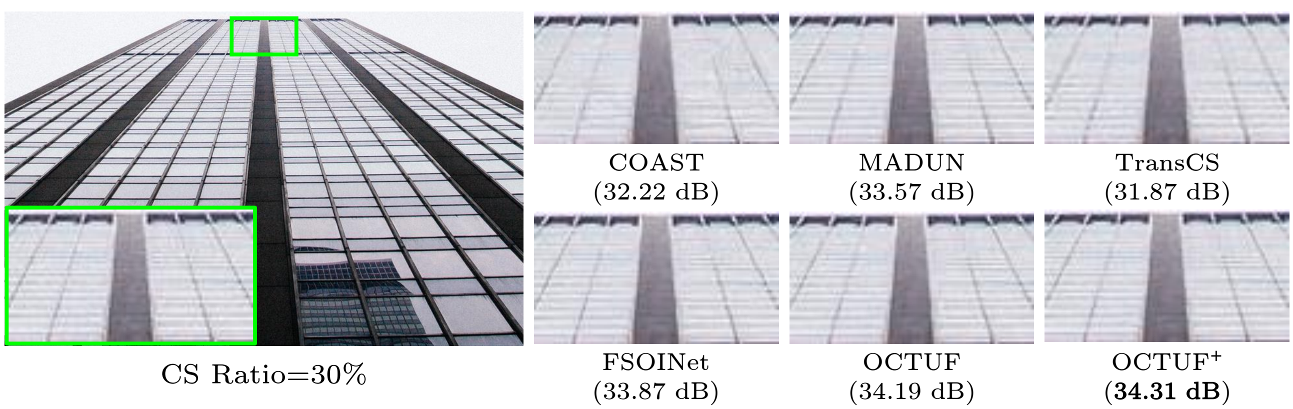

Figure 1. Comparisons on recovering an image from DIV2K dataset [1] in the case of CS ratio = 30%.

We furthermore provide more visual results of different competing approaches to prove the superiority of the proposed method, including ISTA-Net$^+$ [11], DPA-Net [9], AMP-Net [12], MAC-Net [3], COAST [10], MADUN [8], CASNet [2], TransCS [7] and FSOINet [4]. In Figs. 2 to 4, more results on Set11 [6] and Urban100 [5] datasets are provided for various ratios. We can observe from these figures that image edges and details are reconstructed well. In contrast, the other competing methods may lead to over-smooth results or generate results with higher remaining image noise than ours. These observations further verify the effectiveness of our methods for natural images CS both objectively and subjectively.

References

[1] Eirikur Agustsson and Radu Timofte. NTIRE 2017 challenge on single image super-resolution: Dataset and study. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2017. 1, 2

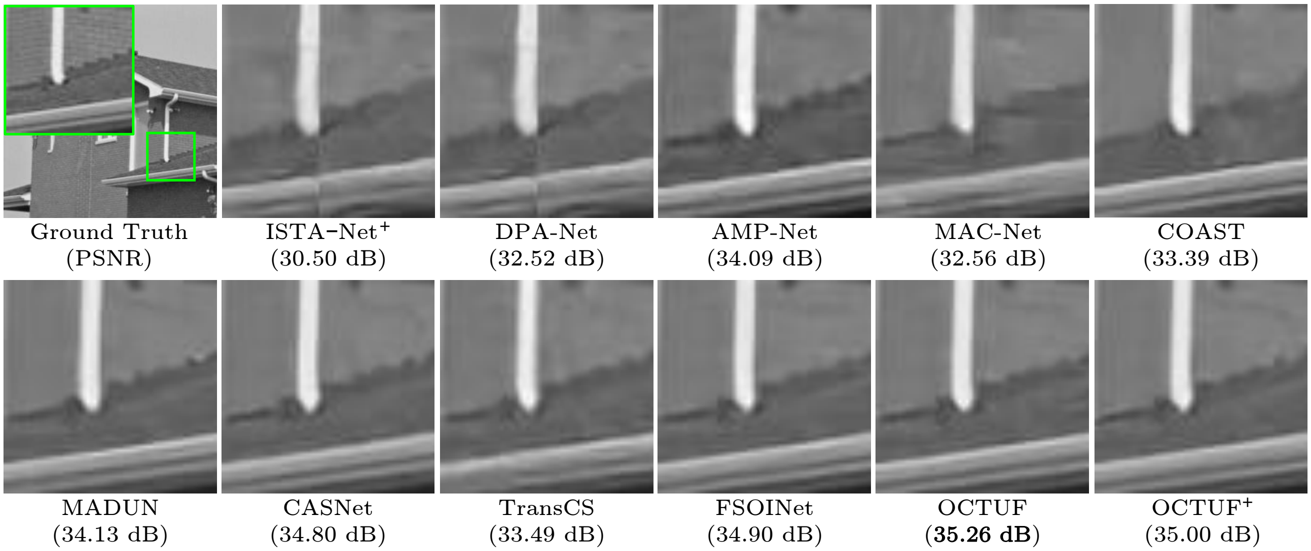

Figure 2. Comparisons on recovering an image from Set11 dataset [6] in the case of CS ratio $= 10\%$.

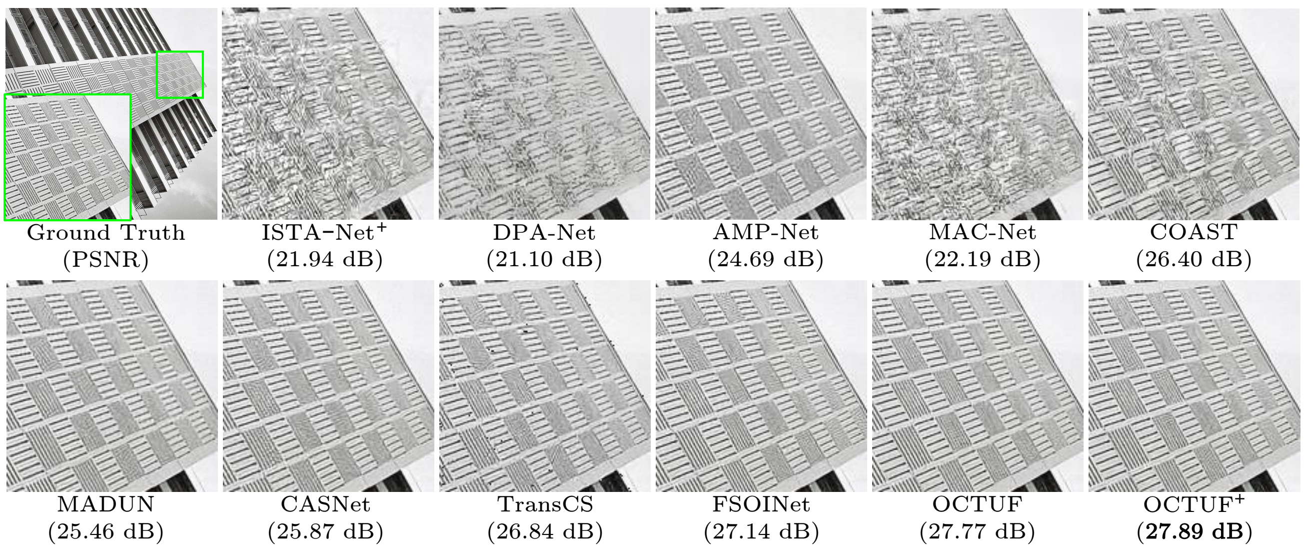

Figure 3. Comparisons on recovering an image from Urban100 dataset [5] in the case of CS ratio $= 40\%$.

[2] Bin Chen and Jian Zhang. Content-aware scalable deep compressed sensing. *IEEE Transactions on Image Processing*, 31:5412–5426, 2022. 2

[3] Jiwei Chen, Yubao Sun, Qingshan Liu, and Rui Huang. Learning memory augmented cascading network for compressed sensing of images. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2020. 2

[4] Wenjun Chen, Chunling Yang, and Xin Yang. FSOINET: feature-space optimization-inspired network for image compressive sensing. In *Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2022. 2

[5] Weisheng Dong, Peiyao Wang, Wotao Yin, Guangming Shi, Fangfang Wu, and Xiaotong Lu. Denoising prior driven deep neural network for image restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(10):2305–2318, 2018. 2, 3, 4

[6] Kuldeep Kulkarni, Suhas Lohit, Pavan Turaga, Ronan Kerviche, and Amit Ashok. ReconNet: Non-iterative reconstruction of images from compressively sensed measurements. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. 1, 2, 3

[7] Minghe Shen, Hongping Gan, Chao Ning, Yi Hua, and Tao Zhang. TransCS: A Transformer-based hybrid architecture for image compressed sensing. *IEEE Transactions on Image Processing*, 2022. 2

[8] Jiechong Song, Bin Chen, and Jian Zhang. Memory-augmented deep unfolding network for compressive sensing. In *Proceedings of the ACM International Conference on Multimedia (ACM MM)*, 2021. 2

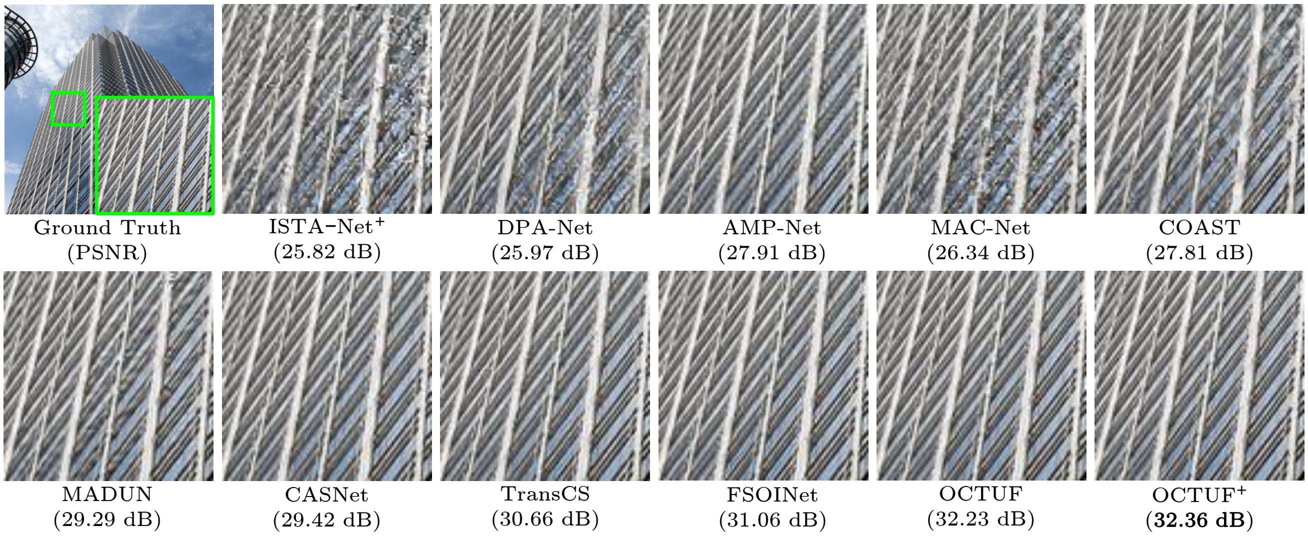

Figure 4. Comparisons on recovering an image from Urban100 dataset [5] in the case of CS ratio = 50%.

[9] Yubao Sun, Jiwei Chen, Qingshan Liu, Bo Liu, and Guodong Guo. Dual-path attention network for compressed sensing image reconstruction. *IEEE Transactions on Image Processing*, 29:9482–9495, 2020. 2

[10] Di You, Jian Zhang, Jingfen Xie, Bin Chen, and Siwei Ma. COAST: Controllable arbitrary-sampling network for compressive sensing. *IEEE Transactions on Image Processing*, 30:6066–6080, 2021. 2

[11] Jian Zhang and Bernard Ghanem. ISTA-Net: Interpretable optimization-inspired deep network for image compressive sensing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. 2

[12] Zhonghao Zhang, Yipeng Liu, Jiani Liu, Fei Wen, and Ce Zhu. AMP-Net: Denoising-based deep unfolding for compressive image sensing. *IEEE Transactions on Image Processing*, 30:1487–1500, 2020. 2