HUNet: Homotopy Unfolding Network for Image Compressive Sensing

Feiyang Shen, Hongping Gan*
School of Software, Northwestern Polytechnical University, Xi'an, China
shenfeiyang@mail.nwpu.edu.cn; ganhongping@nwpu.edu.cn

1. Overview

In this Supplementary Material, we first introduce the technical details of HUNet in Sec. 2, including $\mathcal{S}(\cdot)$, $\widetilde{\mathcal{S}}(\cdot)$, $\mathcal{F}_B(\cdot)$ and $\mathcal{F}_B^{-1}(\cdot)$ in Sec. 2.1 and the mechanisms of PWA and PSWA in Sec. 2.2. Following that, Sec. 3 provides a detailed description of the experiments, covering Sec. 3.1 for detailed experimental settings, Sec. 3.2 for additional comparative experimental results, and Sec. 3.3 for experiments under various noise levels. Finally, Sec. 4 presents a feature visualization analysis, validating DFFM's role in HUNet.

2. Relevant Technical Details

2.1. Details of the Sampling Stage

Before sampling, a complete image with dimensions $l_h \times l_w$ is partitioned by $\mathcal{F}_B(\cdot)$ into a tensor of shape $\frac{l_h \times l_w}{H \times W} \times H \times W$. The inverse process, $\mathcal{F}_B^{-1}(\cdot)$, corresponds to reconstructing the tensor output from the reconstruction stage back into the complete image of size $l_h \times l_w$. To accommodate the sampling operation, the patch size $H \times W$ in HUNet is typically configured as $B \times B$.

The sampling operation can be abstracted as a forward pass using a convolution kernel of size $B \times B$ with a stride of B, which takes a single input channel and produces $\tau \times B \times B$ output channels. This operation is denoted as $\mathcal{S}(\cdot) : \mathbb{R}^{B\times B} \to \mathbb{R}^{\tau\times B\times B}$, where $\tau \times B \times B$ is rounded to the nearest integer, ensuring consistency in dimensions. In contrast, the initialization of $\mathbf{x}_0$ can be interpreted as a transposed convolution operation using the same convolutional kernel, denoted as $\widetilde{\mathcal{S}}(\cdot) : \mathbb{R}^{\tau\times B\times B} \to \mathbb{R}^{B\times B}$. For input images, zero-padding is applied as necessary to ensure that l_h and l_w are integer multiples of B.

2.2. Details of PWA and PSWA

PWA and PSWA receive the input feature map $\mathbf{Z} \in \mathbb{R}^{zw^2\times c}$, where w denotes the window size for segmentation, z denotes the number of windows, and c denotes the number of channels and perform attention operations based on the windows and shifted windows, respectively. Unlike conventional self-attention computations, when passing through the linear layer L_Q, L_K, L_V to get $\{\mathbf{Q}, \mathbf{K}, \mathbf{V}\}$, PWA and PSWA maintain $\mathbf{Q}$ with the same dimensions as $\mathbf{Z}$, while reducing the channel dimensions of $\mathbf{K}$ and $\mathbf{V}$ to c/r^2, resulting in $\mathbf{K}, \mathbf{V} \in \mathbb{R}^{zw^2\times c/r^2}$, expressed as:

$$\mathbf{Q}, \mathbf{K}, \mathbf{V} = L_Q(\mathbf{Z}), L_K(\mathbf{Z}), L_V(\mathbf{Z}). \quad (1)$$

Subsequently, spatial dimensions of $\mathbf{K}$ and $\mathbf{V}$ are reshaped into the channel dimension to get $\mathbf{K}_p$ and $\mathbf{V}_p$:

$$\mathbf{K} \in \mathbb{R}^{zw^2\times c/r^2} \to \mathbf{K}_p \in \mathbb{R}^{zw^2/r^2\times c}, \quad (2)$$

$$\mathbf{V} \in \mathbb{R}^{zw^2\times c/r^2} \to \mathbf{V}_p \in \mathbb{R}^{zw^2/r^2\times c}. \quad (3)$$

Thus, through reduction and reshaping operations, the window scope of $\mathbf{V}_p$ and $\mathbf{K}_p$ is reduced by a factor of r while maintaining consistency in channel dimensions with $\mathbf{Q}$, ensuring consistency in multi-channel information correspondence during attention map generation. Specifically, the window size w is always set as an integer multiple

Table A1. Detailed configurations of HUNet.

Configurations	Default
learning rate	1e-04
optimizer	AdamW
training epoch	200
learning rate schedule	[50,150,180]
learning rate decay	0.1
patch size B	64
batch size	48
phases count n	7
ISS count Θ	3
channels count C	48
window size w	8
scaling factor r	4
$\mathcal{S}(\cdot)$ / $\widetilde{\mathcal{S}}(\cdot)$ weight init	Gaussian random matrix
$\{\rho_k\}_{k=1}^n$ init	0.5
λ init	0.1
$\{\gamma_k\}_{k=1}^n$ init	0.1

*Corresponding author

Table A2. PSNR (dB)/SSIM comparisons between HUNet and other SOTA methods on OST300 [20] at various CS ratios.

Dataset	Methods	0.01	0.04	0.10	0.25	0.30	0.40	0.50
OST300	ISTA-Net$^+$ (CVPR 2018)	19.36/0.4208	22.06/0.5475	24.78/0.6896	28.53/0.8433	29.55/0.8722	31.34/0.9116	33.20/0.9396
	CSNet$^+$ (TIP 2020)	21.91/0.4983	24.33/0.6543	26.65/0.7875	29.86/0.8961	30.96/0.9178	33.19/0.9488	34.96/0.9649
	DPA-Net (TIP 2020)	18.66/0.4593	23.22/0.6186	25.08/0.7314	28.46/0.8562	29.17/0.8797	30.53/0.9140	31.98/0.9385
	OPINE-Net$^+$ (J-STSP 2020)	21.94/0.5089	24.76/0.6703	27.16/0.7941	30.76/0.9021	32.54/0.9311	34.72/0.9591	36.61/0.9727
	MADUN (ACM MM 2021)	-/-	-/-	26.30/0.7578	30.03/0.8807	31.05/0.9030	32.90/0.9345	34.86/0.9567
	AMP-Net-9BM (TIP 2021)	22.31/0.5288	24.92/0.6651	27.35/0.7859	31.06/0.9009	-/-	-/-	-/-
	DGUNet$^+$ (CVPR 2022)	22.36/0.5306	25.24/0.6973	27.84/0.8187	31.53/0.9170	32.44/0.9328	34.41/0.9558	36.44/0.9718
	CASNet (TIP 2022)	22.47/0.5338	25.15/0.6911	27.66/0.8124	31.35/0.9135	32.35/0.9303	34.28/0.9541	36.28/0.9700
	FSOINet (ICASSP 2022)	22.49/0.5335	25.25/0.6953	27.75/0.8159	31.55/0.9171	32.58/0.9338	34.57/0.9570	36.61/0.9723
	TransCS (TIP 2022)	21.67/0.4826	24.86/0.6756	27.31/0.8018	31.07/0.9096	31.87/0.9252	34.17/0.9534	36.24/0.9701
	OCTUF (CVPR 2023)	22.46/0.5298	25.19/0.6910	27.77/0.8148	31.60/0.9175	32.62/0.9339	34.61/0.9572	36.69/0.9726
	TCS-Net (TCI 2023)	22.28/0.5127	24.74/0.6728	27.04/0.8000	30.55/0.9084	30.81/0.9145	32.55/0.9400	34.52/0.9633
	CSformer (TIP 2023)	22.48/0.5299	25.19/0.6843	27.53/0.7950	31.05/0.9038	-/-	-/-	35.75/0.9657
	DPC-DUN (TIP 2023)	20.12/0.4645	23.61/0.6249	26.25/0.7561	29.93/0.8792	30.94/0.9013	32.81/0.9335	34.69/0.9554
	AutoBCS (TCYB 2023)	21.65/0.5176	24.51/0.6769	26.87/0.7991	30.52/0.9083	31.33/0.9230	33.13/0.9478	34.73/0.9640
	MTC-CSNet (TCYB 2024)	22.38/0.5217	24.92/0.6732	27.32/0.8053	31.06/0.9110	31.47/0.9278	32.98/0.9427	34.91/0.9532
	LTwIST (TCSVT 2024)	22.17/0.5105	24.86/0.6767	27.42/0.8064	31.16/0.9115	32.31/0.9273	34.10/0.9513	36.12/0.9643
	NesTD-Net (TIP 2024)	22.58/0.5313	25.17/0.6920	27.73/0.8156	31.60/0.9170	32.48/0.9322	34.57/0.9565	36.65/0.9720
	SCT$^+$(IJCV 2024)	-/-	-/-	25.27/0.7207	-/-	29.03/0.8656	-/-	31.38/0.9160
	UFC-Net (CVPR 2024)	22.40/0.5225	25.00/0.6830	27.53/0.8079	31.26/0.9108	32.25/0.9284	34.23/0.9529	36.31/0.9698
	CPP-Net (CVPR 2024)	22.76/0.5400	25.39/0.7001	27.93/0.8207	31.67/0.9185	32.71/0.9347	34.66/0.9573	36.68/0.9724
	HUNet (Our Method)	**22.78/0.5409**	**25.63/0.7103**	**28.20/0.8266**	**32.06/0.9222**	**33.09/0.9385**	**35.09/0.9600**	**37.29/0.9749**

of the scaling factor r. The self-attention computation, Attention$(\cdot)$, in PWA/PSWA is formulated as:

$$\text{Attention}(\mathbf{Q}, \mathbf{K}_p, \mathbf{V}_p) = \text{Softmax}(\mathbf{Q}\mathbf{K}_p^\top + \mathbf{B})\mathbf{V}_p. \quad (4)$$

Here, $\mathbf{B}$ represents alignment-relative positional embeddings, obtained through interpolation of the original embeddings [10]. Notably, by dividing the channels into multiple groups, the aforementioned equation can be seamlessly extended into a multi-head version.

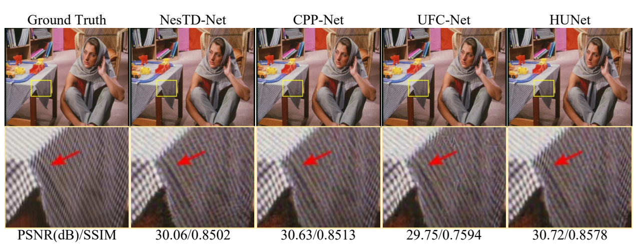

Figure A1. The visually examples of noise influence under Gaussian noise with $\sigma = 0.003$ on dataset Set14 [22] at sampling rate $\tau = 0.25$.

3. More Experiments

3.1. Experimental Settings

The training of HUNet is conducted using image patches of size 64×64, derived from 800 images in the DIV2K [1] dataset. The detailed parameter configurations used in HUNet are provided in Tab. A1.

Table A3. Comparison of the parameters, FLOPs, inference time and inference memory in the case of CS ratio $\tau = 0.1$.

Methods	Params. (M)	FLOPs (G)	Inference time (s)	Inference memory (MB)	PSNR (dB)
LTwIST	23.28	158.9	0.31346	552	27.42
NesTD-Net	5.36	372.58	0.23674	6140	27.73
CPP-Net	16.9	166.93	0.19615	2234	27.93
UFC-Net	1.65	112.42	0.21517	1506	27.53
HUNet	21.1	207.2	0.18203	1830	28.20

Table A4. Comparison of PSNR (dB)/SSIM under Gaussian noise intensities $\sigma \in \{0.001, 0.002, 0.004, 0.006\}$ on Urban100.

Methods	0.001	0.002	0.004	0.006
DGU-Net$^+$	31.81/0.8933	30.78/0.8626	29.39/0.8123	28.36/0.7716
OCTUF	32.00/0.8942	30.92/0.8633	29.45/0.8120	28.41/0.7707
DPC-DUN	30.33/0.8506	29.18/0.8105	27.67/0.7460	26.65/0.6963
NesTD-Net	32.08/0.8947	30.74/0.8634	29.48/0.8128	28.43/0.7727
CPP-Net	32.14/0.8949	31.05/0.8636	29.59/.8136	28.56/0.7732
UFC-Net	31.03/0.8881	30.22/0.8575	29.00/0.8072	28.05/0.7660
HUNet	**32.37/0.8987**	**31.18/0.8670**	**29.65/0.8162**	**28.58/0.7752**

3.2. More Comparison

In this section, we first perform a comprehensive evaluation of the top-performing algorithms discussed in the main text [2–9, 11–14, 16–19, 21]. To extend the analysis, we supplement these with additional methods: ISTA-Net$^+$ [23], MADUN [15], OPINE-Net$^+$ [24], AMP-Net-9BM [25], and AutoBCS [5]. All experimental results are consolidated in Tab. A2, where the best and second-best metrics are marked in **red** and blue, respectively. It can be observed that HUNet consistently outperforms the latest state-of-the-art methods, such as NesTD-

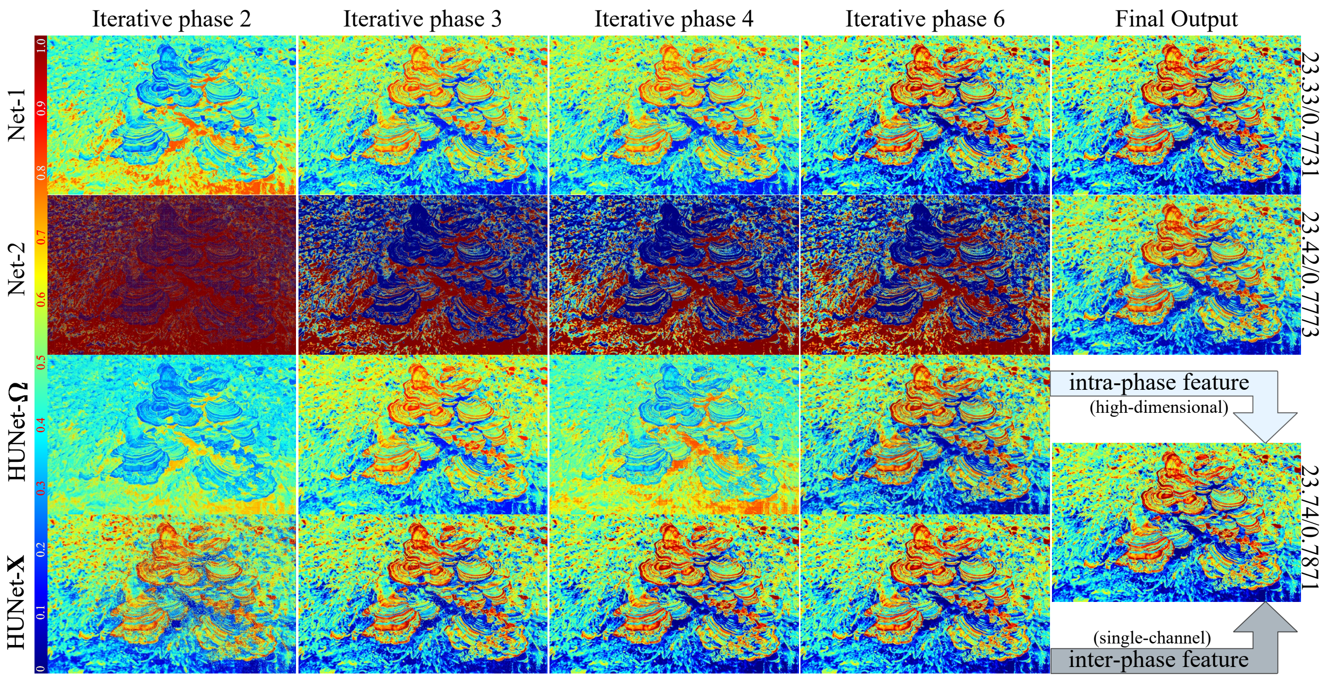

Figure A2. Visualization analysis of feature maps. The output feature maps from the 2nd, 3rd, 4th and 6th phases are displayed. The first and second rows present the feature visualization results of Net-1 and Net-2, respectively, while the third and fourth rows show $\{\mathbf{x}_k\}_{k=1}^n$ and $\{\boldsymbol{\omega}_k\}_{k=1}^n$ of HUNet.

Net and UFC-Net, across various sampling rates $\tau \in \{0.01, 0.04, 0.10, 0.25, 0.30, 0.40, 0.50\}$ in terms of PSNR and SSIM, highlighting its capability for superior image reconstruction. Additionally, Tab. A3 provides a comparison of HUNet with mainstream DUNs at a 0.1 sampling rate for reference. PSNR results from OST300 dataset, inference time and inference memory are the average of reconstructed 256×256 images. It can be observed that HUNet achieves the best reconstruction performance while maintaining optimal inference speed.

Table A5. Comparison of PSNR (dB)/SSIM under salt-and-pepper noise ratios $\delta \in \{0.01, 0.02, 0.04, 0.06\}$ on Set14.

Methods	0.01	0.02	0.04	0.06
DGU-Net$^+$	29.04/0.8212	27.13/0.7432	25.03/0.6440	23.56/0.5722
OCTUF	29.02/0.8219	27.16/0.7438	25.05/0.6443	23.60/0.5733
DPC-DUN	27.36/0.7649	25.64/0.6727	23.70/0.5541	22.35/0.4768
NesTD-Net	29.06/0.8221	27.17/0.7436	25.06/0.6462	23.70/0.5771
CPP-Net	29.03/0.8202	27.20/0.7463	25.11/0.6461	23.72/0.5782
UFC-Net	27.01/0.7926	25.30/0.7172	23.52/0.6220	22.41/0.5576
HUNet	**29.09/0.8222**	**27.21/0.7464**	**25.13/0.6465**	**23.74/0.5790**

3.3. More Comparison under Noises

We introduce varying levels of salt-and-pepper noise and different intensities of Gaussian noise to the Urban100 and Set14 datasets to evaluate HUNet's performance in handling noisy images within the context of compressed sensing. The results of this evaluation, presented in Tab. A4 and Tab. A5, compare HUNet's performance with other state-of-the-art methods under Gaussian and salt-and-pepper noises, respectively. It is evident that HUNet consistently outperforms all tested methods in reconstruction performance across different noise environments at a CS ratio $\tau = 0.25$. Moreover, to further highlight our model's remarkable performance, Fig. A1 presents several visual comparisons at a sampling rate $\tau = 0.25$ under Gaussian noise with $\sigma = 0.003$. The recovery images obtained by HUNet under noisy conditions exhibit details more faithful to the originals.

4. Visual Analysis

Furthermore, we visualize the inter-phase feature maps, $\{\mathbf{x}_k\}_{k=1}^n$, and intra-phase feature maps, $\{\boldsymbol{\omega}_k\}_{k=1}^n$ of HUNet. Specifically, for $\boldsymbol{\omega}_k \in \mathbb{R}^{H \times W \times C}$, we apply principal component analysis along the channel dimension to extract features, projecting them onto $\mathbb{R}^{H \times W \times 1}$ for easier observation. Given that existing DUNs, such as CPP-Net, typically only fuse information of type $\mathbf{X}$ obtained at each phase, we select the variant Net-2 to compare with HUNet and assess the impact of different fusion strategies. To better assess the impact of DFFM on model reconstruction performance, we uniformly set the number of training epochs to 30. As shown in Fig. A2, Net-1, which omits DFFM, performs worse than HUNet in phase-by-phase recovery, resulting in reconstruction PSNR and SSIM values that fall significantly below those of HUNet. Net-2, which solely fuses $\{\mathbf{x}_k\}_{k=1}^{n}$, lacks an explicit modeling of the reconstructed image through inter-phase feature maps, leading to suboptimal PSNR and SSIM values in the final reconstruction. In contrast, the $\{\mathbf{x}_k\}_{k=1}^{n}$ of HUNet exhibit phase-wise enhancement, with different phases of $\boldsymbol{\omega}_k$ focusing on varying aspects of the image, culminating in the most refined reconstructed image through final fusion and further validating the effectiveness of DFFM's dual-path feature fusion strategy.

References

[1] E irikur 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*, pages 126–135, 2017. 2

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

[3] Bin Chen, Xuanyu Zhang, Shuai Liu, Yongbing Zhang, and Jian Zhang. Self-supervised scalable deep compressed sensing. *International Journal of Computer Vision*, pages 1–36, 2024.

[4] Wenjun Chen, Chunling Yang, and Xin Yang. Fsoinet: feature-space optimization-inspired network for image compressive sensing. In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2460–2464. IEEE, 2022.

[5] Hongping Gan, Yang Gao, Chunyi Liu, Haiwei Chen, Tao Zhang, and Feng Liu. Autobcs: Block-based image compressive sensing with data-driven acquisition and noniterative reconstruction. *IEEE Transactions on Cybernetics*, 53 (4):2558–2571, 2023. 2

[6] Hongping Gan, Minghe Shen, Yi Hua, Chunyan Ma, and Tao Zhang. From patch to pixel: A transformer-based hierarchical framework for compressive image sensing. *IEEE Transactions on Computational Imaging*, 9:133–146, 2023.

[7] Hongping Gan, Zhen Guo, and Feng Liu. Nestd-net: Deep nesta-inspired unfolding network with dual-path deblocking structure for image compressive sensing. *IEEE Transactions on Image Processing*, 2024.

[8] Hongping Gan, Xiaoyang Wang, Lijun He, and Jie Liu. Learned two-step iterative shrinkage thresholding algorithm for deep compressive sensing. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(5):3943–3956, 2024.

[9] Zhen Guo and Hongping Gan. Cpp-net: Embracing multi-scale feature fusion into deep unfolding cp-ppa network for compressive sensing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 25086–25095, 2024. 2

[10] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 10012–10022, 2021. 2

[11] Chong Mou, Qian Wang, and Jian Zhang. Deep generalized unfolding networks for image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 17399–17410, 2022. 2

[12] 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*, 31:6991–7005, 2022.

[13] Minghe Shen, Hongping Gan, Chunyan Ma, Chao Ning, Hongqi Li, and Feng Liu. Mtc-csnet: Marrying transformer and convolution for image compressed sensing. *IEEE Transactions on Cybernetics*, 2024.

[14] Wuzhen Shi, Feng Jiang, Shaohui Liu, and Debin Zhao. Image compressed sensing using convolutional neural network. *IEEE Transactions on Image Processing*, 29:375–388, 2019. 2

[15] Jiechong Song, Bin Chen, and Jian Zhang. Memory-augmented deep unfolding network for compressive sensing. In *Proceedings of the 29th ACM international conference on multimedia*, pages 4249–4258, 2021. 2

[16] Jiechong Song, Bin Chen, and Jian Zhang. Dynamic path-controllable deep unfolding network for compressive sensing. *IEEE Transactions on Image Processing*, 32:2202–2214, 2023. 2

[17] Jiechong Song, Chong Mou, Shiqi Wang, Siwei Ma, and Jian Zhang. Optimization-inspired cross-attention transformer for compressive sensing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6174–6184, 2023.

[18] 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.

[19] Xiaoyang Wang and Hongping Gan. Ufc-net: Unrolling fixed-point continuous network for deep compressive sensing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 25149–25159, 2024. 2

[20] Xintao Wang, Ke Yu, Chao Dong, and Chen Change Loy. Recovering realistic texture in image super-resolution by deep spatial feature transform. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 606–615, 2018. 2

[21] Dongjie Ye, Zhangkai Ni, Hanli Wang, Jian Zhang, Shiqi Wang, and Sam Kwong. Csformer: Bridging convolution and transformer for compressive sensing. *IEEE Transactions on Image Processing*, 32:2827–2842, 2023. 2

[22] Roman Zeyde, Michael Elad, and Matan Protter. On single image scale-up using sparse-representations. In *Curves and Surfaces: 7th International Conference, Avignon, France, June 24-30, 2010, Revised Selected Papers 7*, pages 711–730. Springer, 2012. 2

[23] 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*, pages 1828–1837, 2018. 2

[24] Jian Zhang, Chen Zhao, and Wen Gao. Optimization-inspired compact deep compressive sensing. *IEEE Journal of Selected Topics in Signal Processing*, 14(4):765–774, 2020. 2

[25] 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, 2021. 2