Supplementary Material for "Ensemble Learning Priors Driven Deep Unfolding for Scalable Snapshot Compressive Imaging"

Chengshuai Yang[0000−0003−2840−5344], Shiyu Zhang[0000−0001−7111−3895], and Xin Yuan[0000−0002−8311−7524]

School of Engineering, Westlake University, Hangzhou, Zhejiang 310030, China
integrityyang@gmail.com {zhangshiyu, xyuan}@westlake.edu.cn

1 Denoising network structure

In this section, we provide the detailed network structure of the proposed fully convolutional network backbone. The first prior network structure is shown in Fig. S1 and the non-first in Fig. S2. Here, $\boldsymbol{u}^0 = \boldsymbol{x}^0 - \frac{\boldsymbol{\lambda}_2^0}{\gamma_2^1}$ and $\boldsymbol{u}^{i-1} = \boldsymbol{x}^{i-1} - \frac{\boldsymbol{\lambda}_2^{i-1}}{\gamma_2^i}$.

2 Scalable comparison

In this section, we test four different size datasets: 256×256×24, 512×512 ×10, 1024 ×1024×18 and 1536×1536×12. These four datasets have different sizes not only in spatial dimension but also in *temporal dimension*. Here, we use GAP as the PnP framework [2]. Thus, we can denote the GAP-FFDNet (GAP-FastDVDnet) as PnP-FFDNet (PnP-FastDVDnet). To help compare, the detailed results are presented, including tables, images and videos. For videos, please refer to the folder 'video/ scalable'(https://github.com/integritynoble/ELP-Unfolding/tree/master/video).

2.1 Data size 256 × 256 × 24, (B=24).

In this case, the dataset is the same with the benchmark but the compression ratio B is now set to 24. Because of high compressed ratio, which is three times of the benchmark, the reconstruction accuracy is reduced but our proposed ELP-Unfolding but we can still obtain the acceptable results (31.53 dB for PSNR), as show in Table S1 and Fig. S3. In this case, one measurement is used to test. As shown in Fig. S3, FFDNet just gets the deformed shapes because of high compressed ratio. But our proposed ELP-Unfolding achieves clearer images than other three algorithms.

2.2 Data size 512 × 512 × 10, (B=10).

In this case, the spatial size is 512 × 512 and temporal size is 10. Here, three measurements are used to test. The dataset is cropped from the Ultra Video

Group (UVG) dataset [1]. The results are shown in Table S2 and Fig. S4. Because of low compressed ratio and large spatial size, high reconstruction can easily be achieved. But our proposed ELP-unfolding still get the best result at the shortest testing time.

2.3 Data size $1024 \times 1024 \times 18$, ($B$=18).

In this case, the spatial size is 1024×1024 and temporal size is 18. Here, four measurements are used to test. The dataset is also cropped from the Ultra Video Group (UVG) dataset [1]. The results are shown in Table S3 and Fig. S5. Because of high compressed ratio, the reconstruction accuracy of other algorithms gets degraded, especially GAP-TV. By contrast, our proposed ELP-unfolding still keeps high reconstruction accuracy (36.00 dB for PSNR).

2.4 Data size is $1536 \times 1536 \times 12$, ($B$=12).

In this case, the spatial size is 1536×1536 and temporal size is 12. Here, one measurement is used to test. The dataset is also cropped from the Ultra Video Group (UVG) dataset [1]. The results are shown in Table S4 and Fig. S6. Because of large spatial size, it is not difficult to get good reconstruction results for all algorithms. However, as shown in the zoomed area, our proposed ELP-unfolding gets more details.

Therefore, these four datasets indicate that our proposed ELP-unfolding method has achieved excellent performance for SCI reconstruction not only in model's scalability but also in accuracy and speed.

3 Real data

What's more, all the real dynamic scenes are made into videos. Please refer to the folder 'video/ real'(https://github.com/integritynoble/ELP-Unfolding/tree/master/video).

Table S1: Scalability: Data size is $256 \times 256 \times 24$, (B=24). PSNR (left in dB) and SSIM (right) are shown

Algorithm	Kobe	Runner	Drop	Crash	Aerial	Average	Run time (s)
GAP-TV [2]	23.27, 0.680	25.06, 0.796	29.41, 0.907	22.60, 0.729	22.87, 0.728	24.64, 0.768	2.78(CPU)
PnP-FFDNet [3]	20.31, 0.606	22.60, 0.747	25.76, 0.846	19.91, 0.679	19.97, 0.646	21.71, 0.705	5.96(GPU)
PnP-FastDVDnet [4]	23.34, 0.695	27.83, 0.867	31.75, 0.952	24.60, 0.801	23.82, 0.756	26.27, 0.814	17.84(GPU)
ELP-Unfolding (Ours)	**28.68**, **0.879**	**34.71**, **0.950**	**40.01**, **0.986**	**27.50**, **0.910**	**26.75**, **0.864**	**31.53**, **0.918**	**0.257(GPU)**

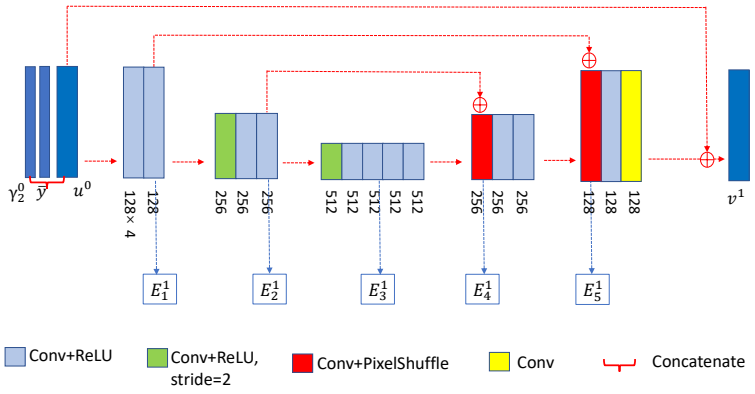

Fig. S1: The first denoising prior network structure

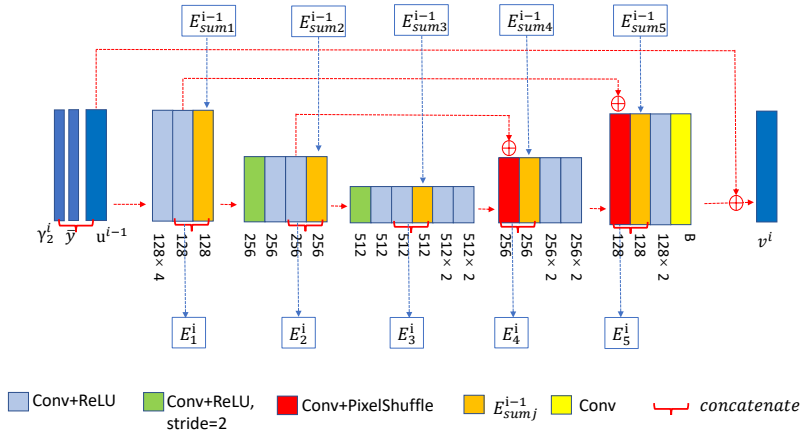

Fig. S2: The non-fisrt denoising prior network structure

Table S2: Scalability: Data size is $512 \times 512 \times 10$, (B=10). PSNR (left in dB) and SSIM (right) are shown

Algorithm	Beauty512	Bosphorus512	HoneyBee512	Jockey512	ShakeNDry512	Average	Run time (s)
GAP-TV [2]	36.37, 0.910	33.63, 0.893	37.62, 0.947	32.68, 0.899	32.66, 0.836	34.59, 0.897	4.99(CPU)
PnP-FFDNet [3]	38.78, 0.922	37.23, 0.932	37.68, 0.945	37.60, 0.937	32.59, 0.832	36.78, 0.914	8.23(GPU)
PnP-FastDVDnet [4]	37.98, 0.915	35.49, 0.899	40.44, 0.957	36.74, 0.921	31.84, 0.808	36.50, 0.900	27.36(GPU)
ELP-Unfolding (Ours)	**40.58**, **0.940**	**40.85**, **0.960**	**43.05**, **0.970**	**40.26**, **0.956**	**34.52**, **0.876**	**39.85**, **0.940**	**0.896(GPU)**

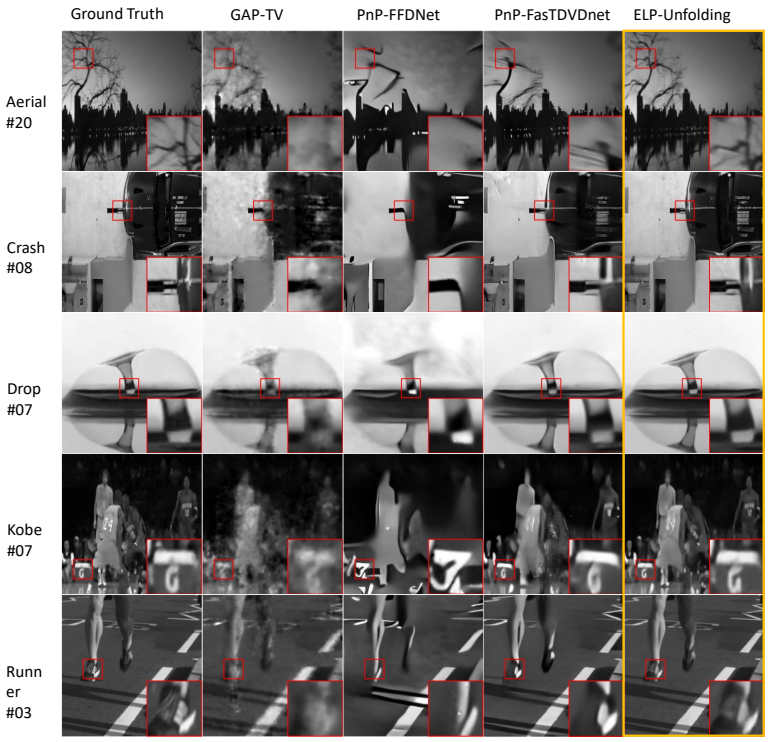

Fig. S3: Scalability: Data size is $256 \times 256 \times 24$.

Table S3: Scalability: Data size is $1024 \times 1024 \times 18$, ($B$=18). PSNR (left in dB) and SSIM (right) are shown

Algorithm	Beauty1024	Jockey1024	ReadySteadyGo1024	ShakeNDry1024	YachtRide1024	Average	Run time (s)
GAP-TV [2]	34.50, 0.873	28.13, 0.815	25.78, 0.755	32.19, 0.861	26.32, 0.761	29.38, 0.813	71.83(CPU)
PnP-FFDNet [3]	37.15, 0.898	31.42, 0.896	28.78, 0.858	29.81, 0.793	25.99, 0.787	30.63, 0.846	90.49(GPU)
PnP-FastDVDnet [4]	35.54, 0.889	33.52, 0.927	32.28, 0.911	32.65, 0.849	30.51, 0.864	32.90, 0.888	249.41(GPU)
ELP-Unfolding (Ours)	**38.95**, **0.916**	**38.44**, **0.948**	**34.07**, **0.924**	**35.42**, **0.902**	**33.10**, **0.909**	**36.00**, **0.920**	**3.575(GPU)**

Table S4: Scalability: Data size is $1536 \times 1536 \times 12$, B=12. PSNR (left in dB) and SSIM (right) are shown

Algorithm	CityAlley1536	FlowerKids1536	Lips1536	RaceNight1536	RiverBank1536	Average	Run time (s)
GAP-TV [2]	31.88, 0.860	32.89, 0.873	33.23, 0.754	30.63, 0.799	28.43, 0.754	31.41, 0.808	100.82(CPU)
PnP-FFDNet [3]	34.67, 0.894	37.62, 0.934	33.43, 0.738	34.24, 0.827	31.15, 0.826	34.22, 0.844	114.26(GPU)
PnP-FastDVDnet [4]	35.40, 0.883	35.58, 0.912	33.23, 0.728	33.57, 0.818	31.74, 0.837	33.90, 0.836	368.54(GPU)
ELP-Unfolding (Ours)	**37.79**, **0.924**	**38.57**, **0.940**	**34.82**, **0.780**	**34.99**, **0.846**	**34.32**, **0.891**	**36.10**, **0.876**	**8.206(GPU)**

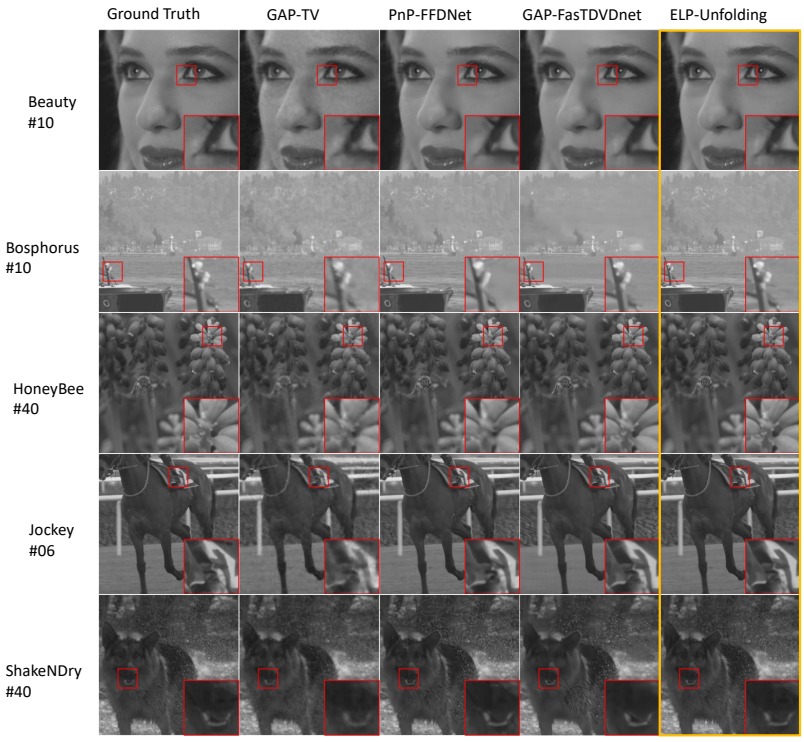

Fig. S4: Scalability: Data size is $512 \times 512 \times 10$.

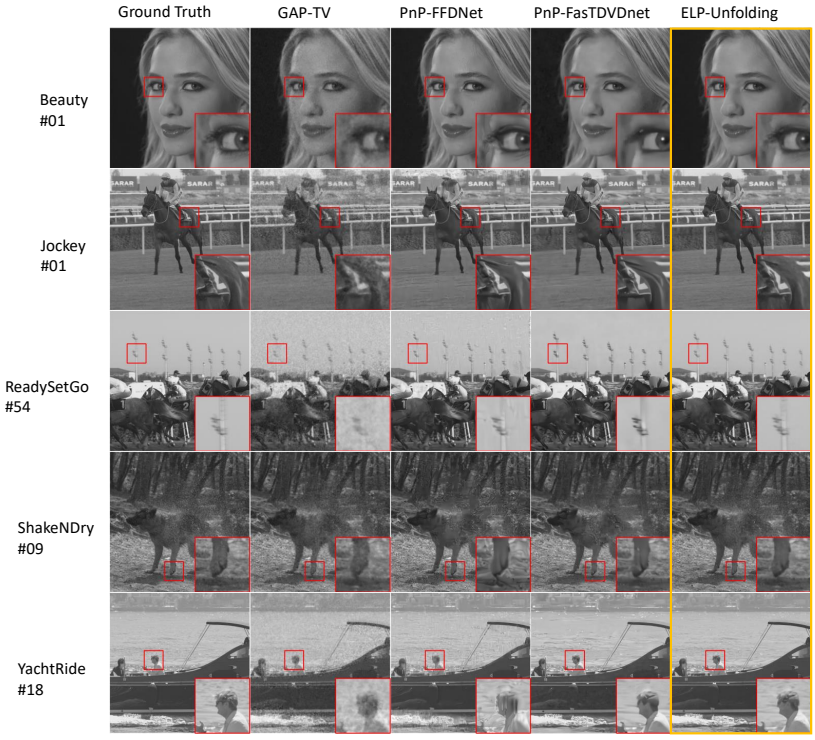

Fig. S5: Scalability: Data size is $1024 \times 1024 \times 18$.

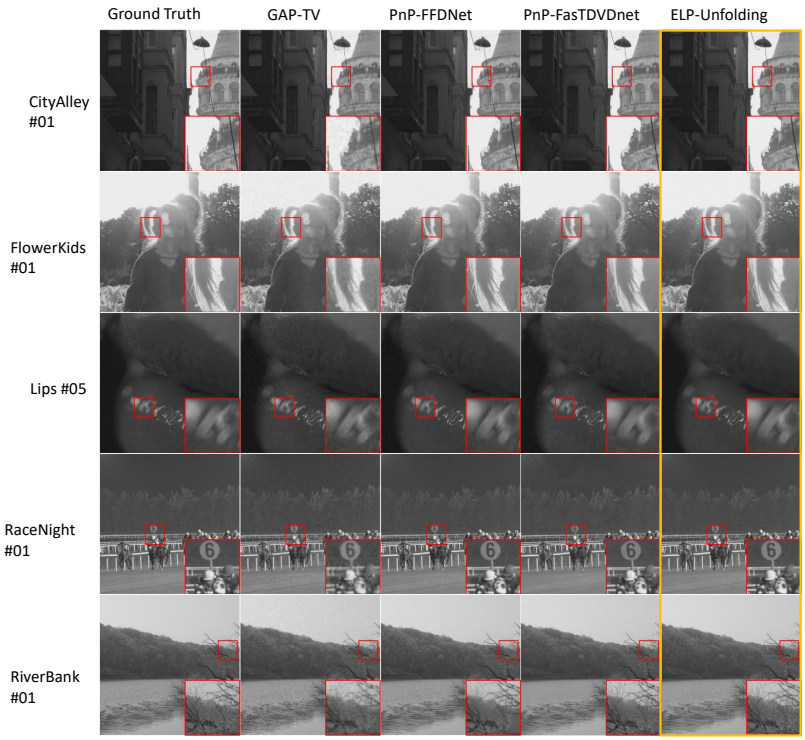

Fig. S6: Scalability: Data size is $1536 \times 1536 \times 12$.

References

1. Mercat, A., Viitanen, M., Vanne, J.: Uvg dataset: 50/120fps 4k sequences for video codec analysis and development. In: Proceedings of the 11th ACM Multimedia Systems Conference. pp. 297–302 (2020) 2
2. Yuan, X.: Generalized alternating projection based total variation minimization for compressive sensing. In: 2016 IEEE International Conference on Image Processing (ICIP). pp. 2539–2543 (September 2016) 1, 2, 3, 4
3. Yuan, X., Liu, Y., Suo, J., Dai, Q.: Plug-and-play algorithms for large-scale snapshot compressive imaging. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1447–1457 (June 2020) 2, 3, 4
4. Yuan, X., Liu, Y., Suo, J., Durand, F., Dai, Q.: Plug-and-play algorithms for video snapshot compressive imaging. arXiv: 2101.04822 (January 2021) 2, 3, 4