EfficientSCI: Densely Connected Network with Space-time Factorization for Large-scale Video Snapshot Compressive Imaging

Lishun Wang*, Miao Cao* and Xin Yuan†

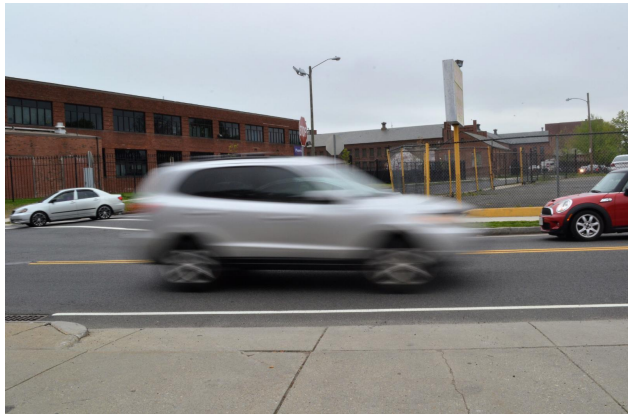
Outline

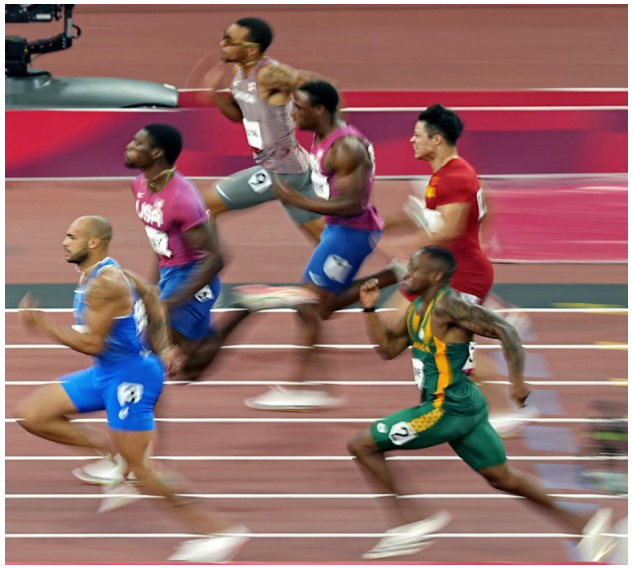
- Introduction
- The proposed network
- Experimental Results

Outline

- Introduction
- The proposed network
- Experimental Results

What is Video Snapshot Compressive Imaging (SCI)?





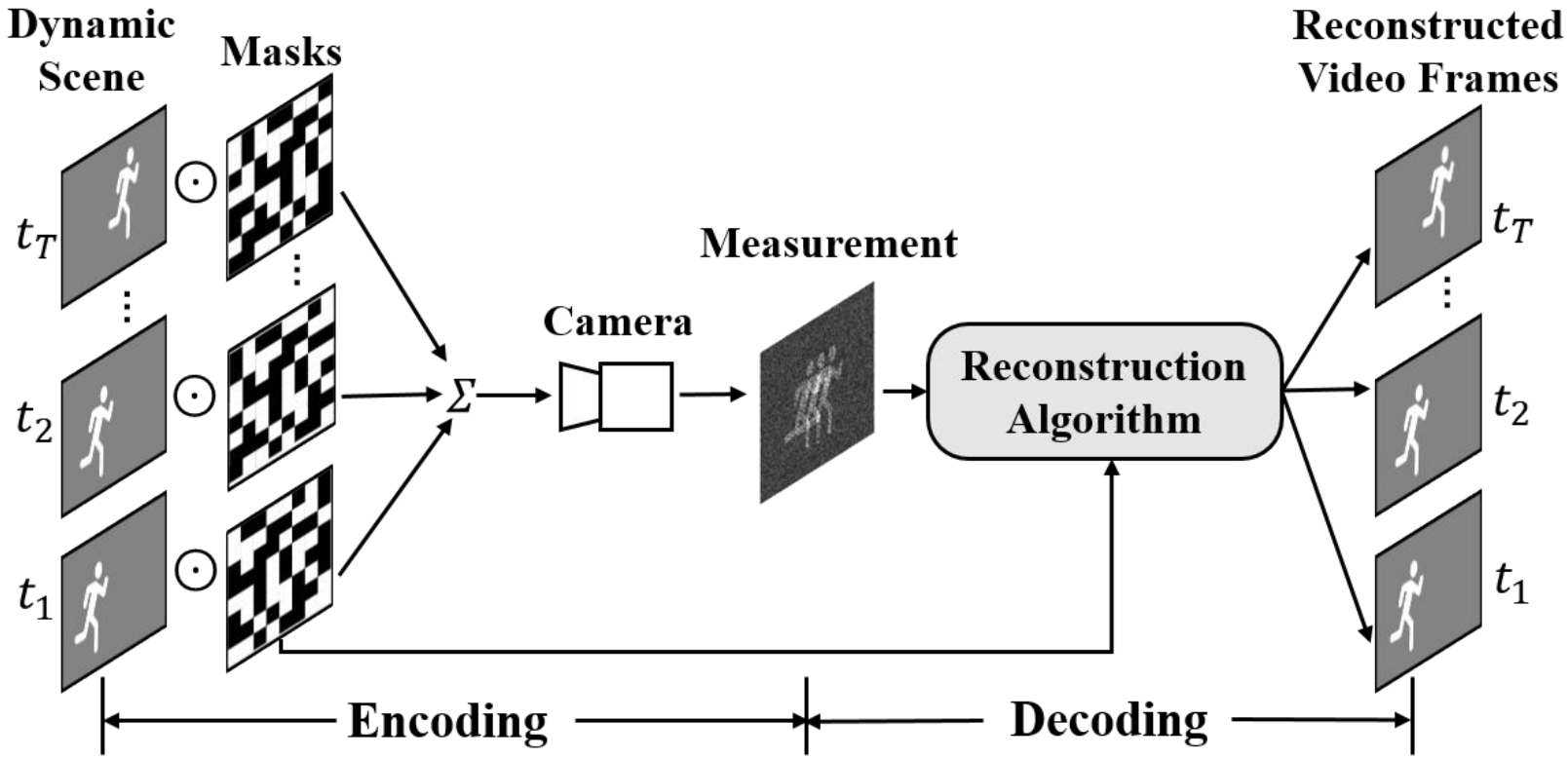
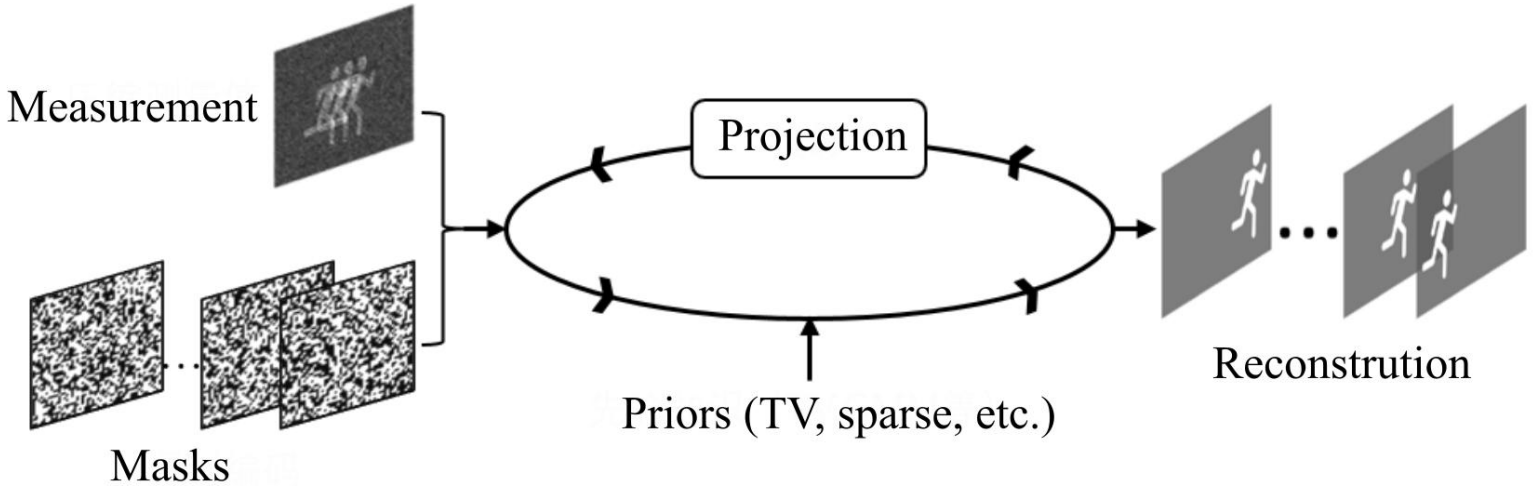


Figure1. Principle of the video SCI system

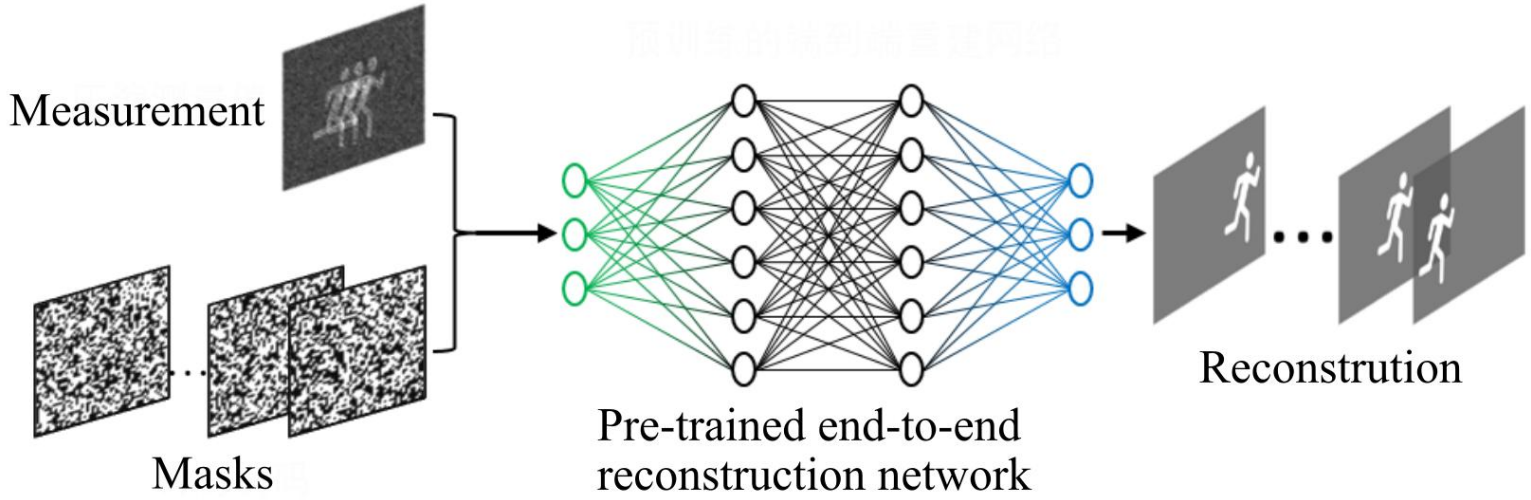
Current Challenges of Video SCI



Model-based methods:

long reconstruction time

poor reconstruction quality



Learning-based methods:

high model complexity

Outline

- Introduction
- The proposed network
- Experimental Results

The proposed network

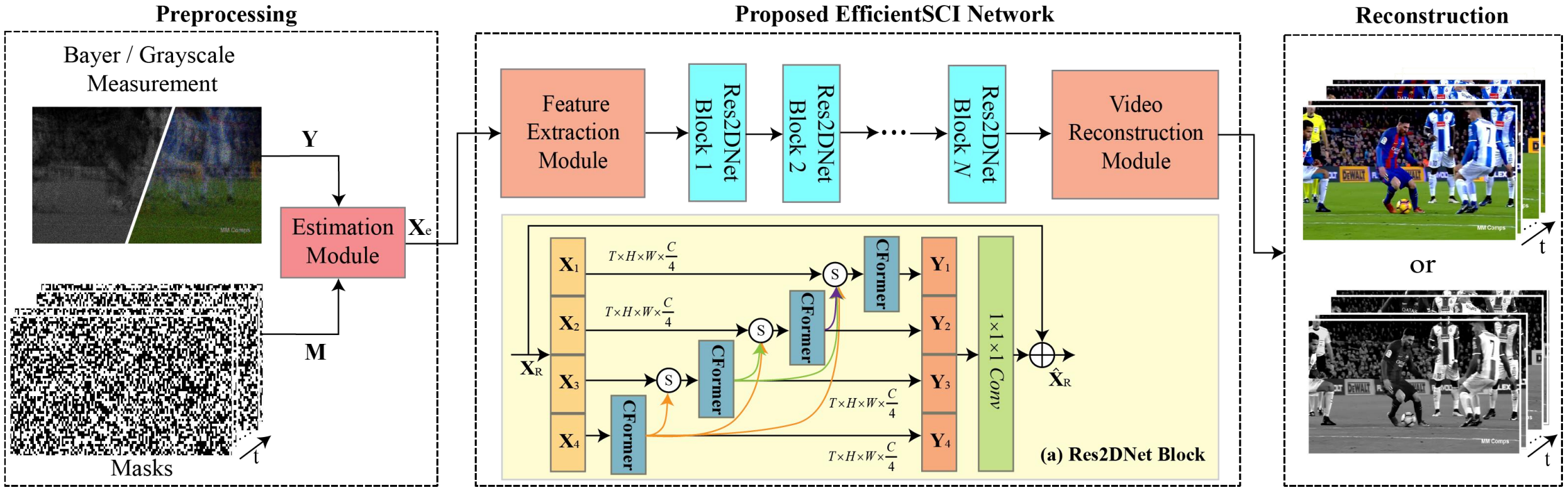


Figure 3. Architecture of the proposed EfficientSCI network

The proposed network

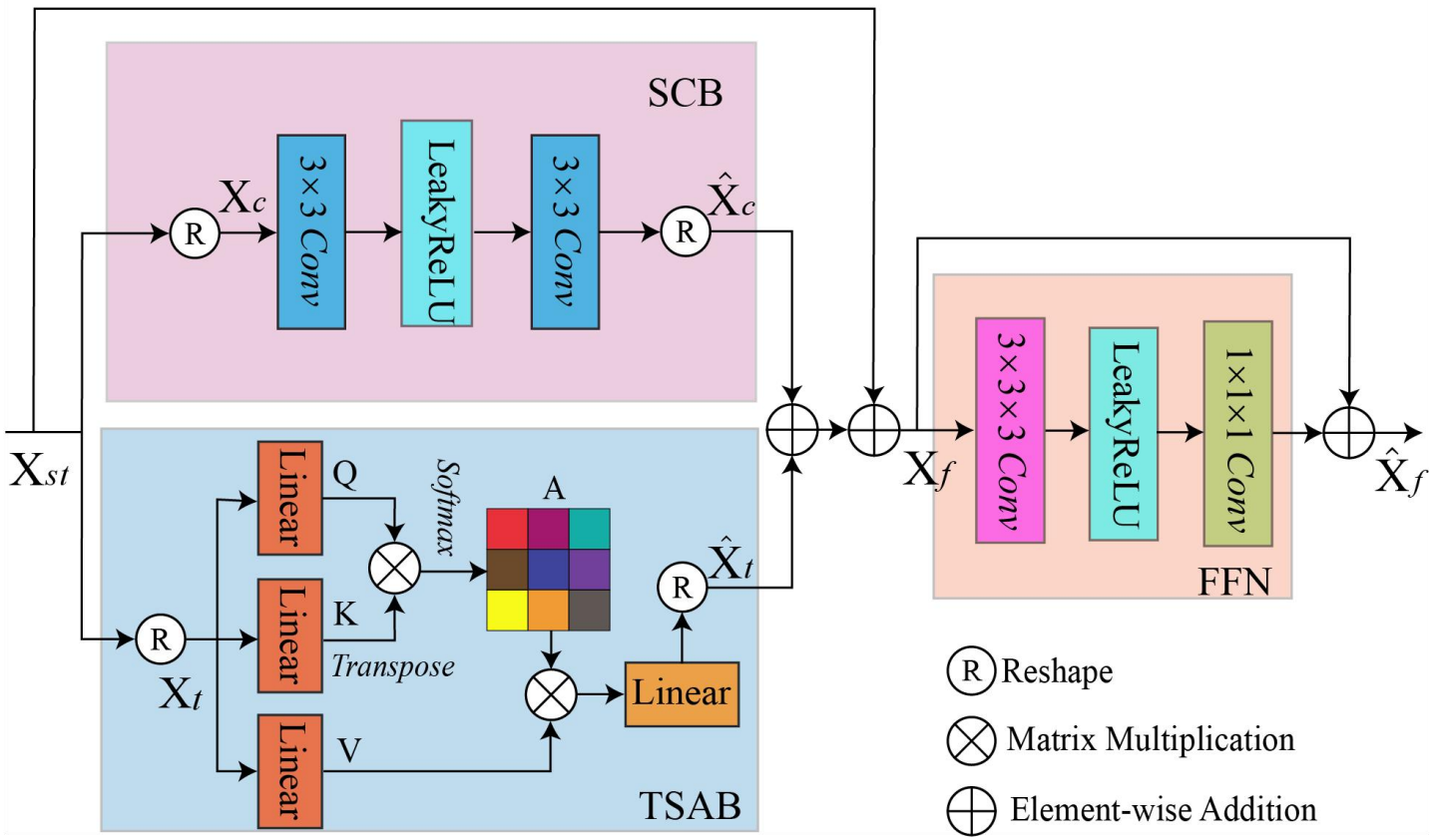


Figure 4. The CFormer block

Spatial Convolution Branch (SCB): SCB is used to extract spatial local information

Temporal Self-attention Branch (TSAB): TSAB is used to calculate temporal attention of the feature points at the same spatial position in each frame.

Feed Forward Network (FFN): FFN is used to further integrate spatial-temporal information.

The proposed network

Table 2. Reconstruction quality and test time (s) using EfficientSCI with different number of channels and blocks

Model	Channel	Block	PSNR	SSIM	Test time(s)
EfficientSCI-T	64	8	34.22	0.961	0.07
EfficientSCI-S	128	8	35.51	0.970	0.15
EfficientSCI-B	256	8	36.48	0.975	0.31
EfficientSCI-L	256	12	36.92	0.977	0.45

Table 3. Computational complexity and reconstruction quality of several SOTA algorithms

Method	Params (M)	FLOPs (G)	PSNR	SSIM
BIRNAT	4.13	390.56	33.31	0.951
RevSCI	5.66	766.95	33.92	0.956
DUN-3DUnet	61.91	3975.83	35.26	0.968
ELP-Unfolding	565.73	4634.94	35.41	0.969
EfficientSCI-T	0.95	142.18	34.22	0.961
EfficientSCI-S	3.78	563.87	35.51	0.970
EfficientSCI-B	8.82	1426.38	36.48	0.975
EfficientSCI-L	12.39	1893.72	36.92	0.977

Outline

- Introduction
- The proposed network
- Experimental Results

Results on Grayscale Simulation Datasets

Table 1. The average PSNR in dB (left entry) and SSIM (right entry) and running time per measurement of different algorithms on 6 benchmark grayscale datasets. The best results are shown in bold and the second-best results are underlined

Method	Kobe	Traffic	Runner	Drop	Crash	Aerial	Average	Test time(s)
GAP-TV	26.46, 0.885	20.89, 0.715	28.52, 0.909	34.63, 0.970	24.82, 0.838	25.05, 0.828	26.73, 0.858	4.2 (CPU)
PnP-FFDNet	30.50, 0.926	24.18, 0.828	32.15, 0.933	40.70, 0.989	25.42, 0.849	25.27, 0.829	29.70, 0.892	3.0 (GPU)
PnP-FastDVDnet	32.73, 0.947	27.95, 0.932	36.29, 0.962	41.82, 0.989	27.32, 0.925	27.98, 0.897	32.35, 0.942	6.0 (GPU)
DeSCI	33.25, 0.952	28.71, 0.925	38.48, 0.969	43.10, 0.993	27.04, 0.909	25.33, 0.860	32.65, 0.935	6180 (CPU)
BIRNAT	32.71, 0.950	29.33, 0.942	38.70, 0.976	42.28, 0.992	27.84, 0.927	28.99, 0.917	33.31, 0.951	0.10 (GPU)
RevSCI	33.72, 0.957	30.02, 0.949	39.40, 0.977	42.93, 0.992	28.12, 0.937	29.35, 0.924	33.92, 0.956	0.19 (GPU)
GAP-CCoT	32.58, 0.949	29.03, 0.938	39.12, 0.980	42.54, 0.992	28.52, 0.941	29.40, 0.923	33.53, 0.958	0.08 (GPU)
DUN-3DUnet	35.00, 0.969	31.76, 0.966	40.03, 0.980	44.96, 0.995	29.33, 0.956	30.46, 0.943	35.26, 0.968	0.58 (GPU)
ELP-Unfolding	34.41, 0.966	31.58, 0.962	41,16, 0.986	44.99, 0.995	29.65, 0.959	30.68, 0.944	35.41, 0.969	0.34 (GPU)
EfficientSCI-T	33.45, 0.960	29.20, 0.942	39.51, 0.981	43.56, 0.993	29.27, 0.954	30.32, 0.937	34.22, 0.961	**0.07 (GPU)**
EfficientSCI-S	34.79, 0.968	31.21, 0.961	41.34, 0.986	44.61, 0.994	30.34, 0.965	30.78, 0.945	35.51, 0.970	0.15 (GPU)
EfficientSCI-B	35.76, 0.974	32.30, 0.968	43.05, 0.988	45.18, 0.995	31.13, 0.971	31.50, 0.953	36.48, 0.975	0.31 (GPU)
EfficientSCI-L	**36.27**, **0.976**	**32.83**, **0.971**	**43.79**, **0.991**	**45.46**, **0.995**	**31.52**, **0.974**	**31.64**, **0.955**	**36.92**, **0.977**	0.45 (GPU)

Results on Grayscale Simulation Datasets

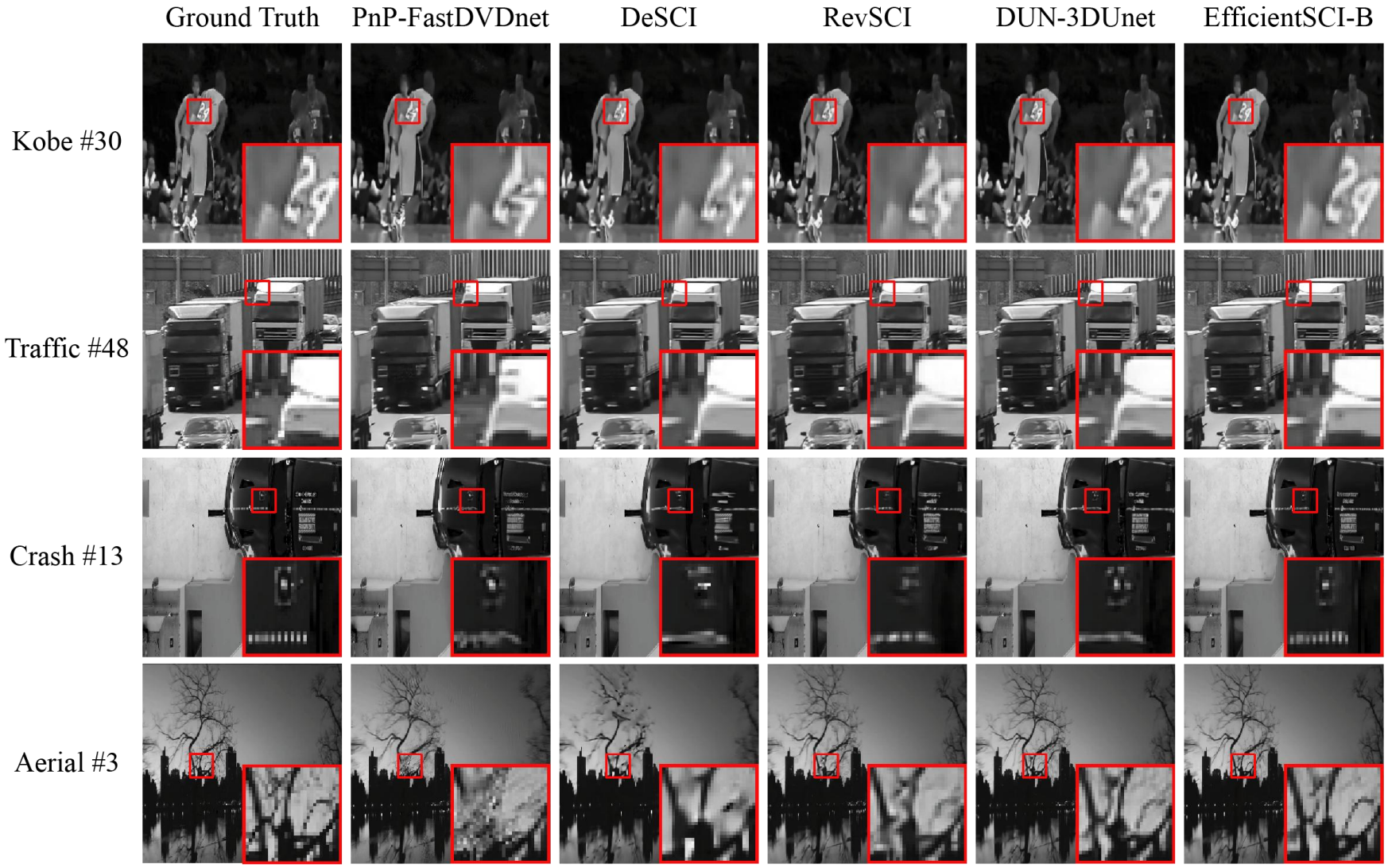


Figure 5. Selected reconstruction frames of simulated grayscale data.

Results on Large-scale Color Simulation Datasets

Table 4. The average PSNR in dB (left entry) and SSIM (middle entry) and test time (minutes) per measurement (right entry) of different algorithms on 4 benchmark large-scale datasets. Best results are in bold

Dataset	Pixel resolution	GAP-TV	PnP-FFDNet-color	PnP-FastDVDnet-color	EfficientSCI-S
Messi	$1080 \times 1920 \times 3 \times 8$	25.20, 0.874, 0.66	34.28, 0.968, 14.93	34.34, 0.970, 15.94	**34.41, 0.973, 0.09**
Hummingbird	$1080 \times 1920 \times 3 \times 30$	25.10, 0.750, 20.3	28.79, 0.665, 61.20	31.17, 0.916, 54.00	**35.56, 0.952, 0.39**
Swinger	$2160 \times 3840 \times 3 \times 15$	22.68, 0.769, 39.2	29.30, 0.934, 138.8	30.57, 0.949, 138.4	**31.05, 0.951, 0.62**
Football	$1644 \times 3840 \times 3 \times 40$	26.19, 0.858, 83.0	32.70, 0.951, 308.8	32.31, 0.947, 298.1	**34.81, 0.964, 1.52**

Results on Large-scale Color Simulation Datasets

Figure 7. Comparison of reconstruction results of different algorithms on several benchmark large-scale color video simulation datasets

Ablation Study

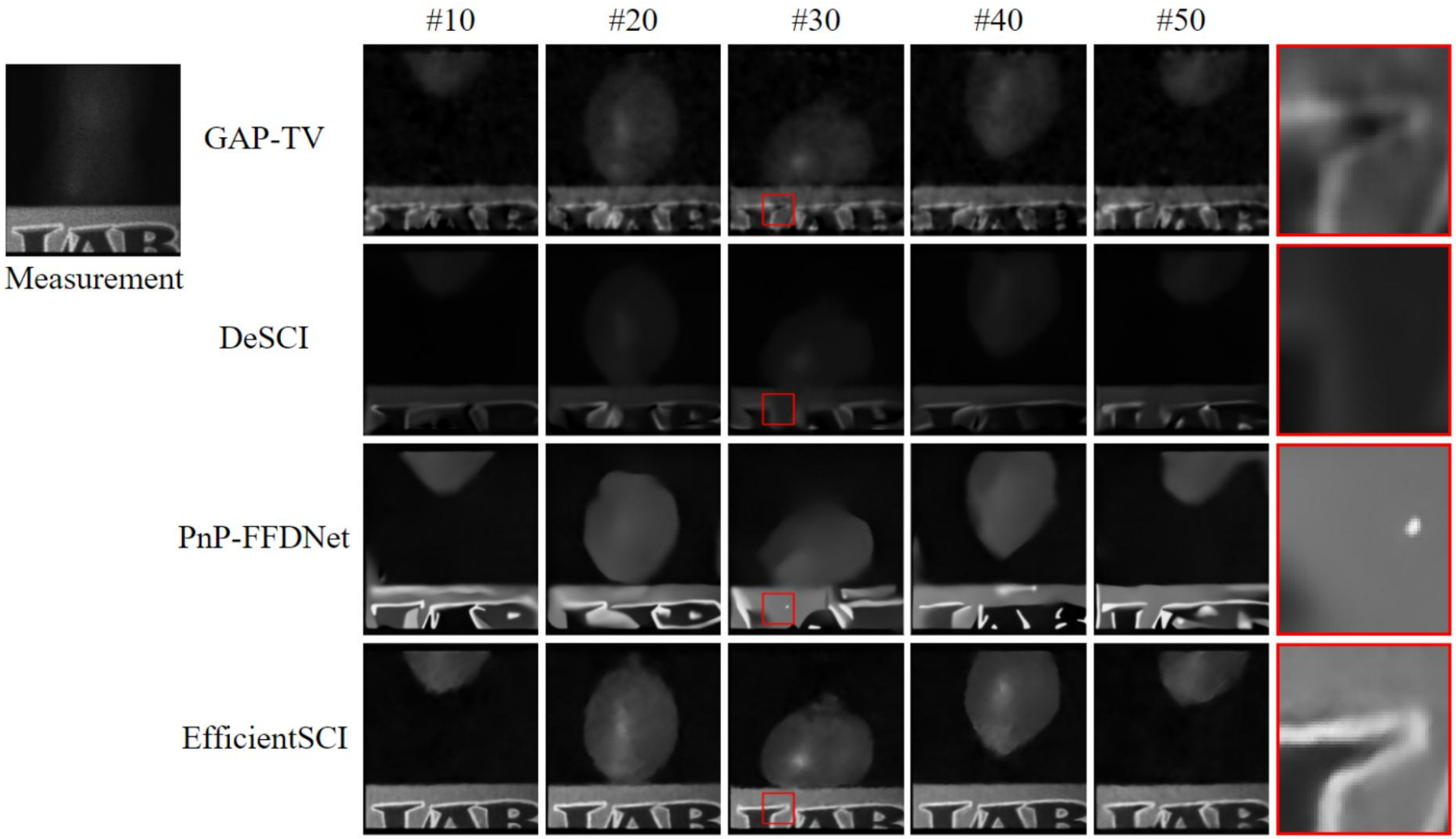
Table 5. Ablation study on the ResDNet block without dense connections (left entry) and with dense connections (right entry)

GN	Params	FLOPs	PSNR	SSIM
1	27.40, 27.40	3860.94, 3860.94	35.17, 35.17	0.967, 0.967
2	14.82, 15.08	2211.77, 2246.39	35.09, 36.02	0.966, 0.974
4	8.53, 8.82	1387.33, 1426.38	35.02, 36.48	0.966, 0.975
8	5.38, 5.65	975.39, 1013.45	34.31, 35.68	0.961, 0.971

Table 6. Ablation study on the CFormer block

SCB	TSAB	Swin	S2-3D	Params (M)	FLOPs (G)	PSNR	SSIM
	✓	✓		2.88	471.03	34.99	0.967
			✓	6.93	999.31	34.93	0.966
✓	✓			3.78	563.87	35.51	0.970

Results on Real Video SCI Data



Reconstruction results of different algorithms on real data with compression rate $B = 50$

Thank you ☺