Supplementary Materials for
MetaSCI: Scalable and Adaptive Reconstruction for Video Compressive Sensing

Zhengjue Wang†, Hao Zhang†, Ziheng Cheng, Bo Chen*
National Laboratory of Radar Signal Processing, Xidian University, Xian, China
{zhengjuewang,zhanghao_xidian,zhchengxd}@163.com,bchen@mail.xidian.edu.cn

Xin Yuan*
Bell Labs, NJ USA
xyuan@bell-labs.com

1. Network Structure

In this section, we provide the detailed network structure of the proposed fully convolutional network backbone in Section 4.2 of the main paper, with the illustrations shown in Fig. 1 and Table 1.

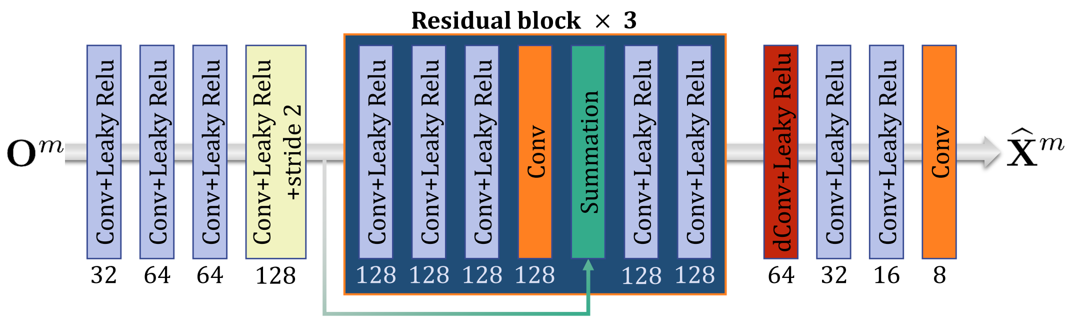

Figure 1: The brief structure of our proposed fully CNN backbone for SCI reconstruction, where the numbers denote the number of kernels at each layer.

2. Simulated Large-scale Benchmark

Since limited researches provide large-scale data for SCI, we create a large-scale benchmark, including $512 \times 512 \times 8$, $1024 \times 1024 \times 8$, and $2048 \times 2048 \times 8$ videos cropped from the Ultra Video Group (UVG) dataset [1]. Each scale have five different video sequences, and the compression ratio is $B = 8$. We download the original UVG dataset from http://ultravideo.cs.tut.fi/#testsequences, with the parameters listed in Table 2. Since the motion speed varies among different videos, for each video, we sequentially choose 64 frames at an interval of '*Step*' frame(s). For example, for the video '*Beauty*' of spatial size 512×512, we extract 1 frame from every 5 frames in the original video, and totally obtain 64 frames.

3. Discussion on the Overlapping Size in Large-scale Video Reconstruction

As discussed in Section 4.5 of the main paper, we spatially decompose a large-scale video into multiple sub-videos, which can be realized by non-overlapping or overlapping cropping. In this section, we discuss the effect of overlapping size in recovering large-scale videos.

In Fig. 2, we can see that cropping without overlapping brings some boundary effect, while overlapping cropping leads to better reconstruction. The corresponding PSNR comparisons is shown in Table 3. Generally speaking, with the increase of the overlapping size, the performance gets better. We observe that as overlapping size is 128, the performance is good enough

† Equal contributions. * Corresponding authors.

Table 1: Network structure of the proposed fully CNN backbone for SCI reconstruction.

	Module	Stride	Kernel Size	Output Size
$\times 1$	Conv. + Leaky Relu	1	$5 \times 5 \times (B+1) \times 32$	$256 \times 256 \times 32$
	Conv. + Leaky Relu	1	$3 \times 3 \times 32 \times 64$	$256 \times 256 \times 64$
	Conv. + Leaky Relu	1	$1 \times 1 \times 64 \times 64$	$256 \times 256 \times 64$
	Conv. + Leaky Relu	2	$3 \times 3 \times 64 \times 128$	$128 \times 128 \times 128$
$\times 3$	Conv. + Leaky Relu	1	$3 \times 3 \times 128 \times 128$	$128 \times 128 \times 128$
	Conv. + Leaky Relu	1	$1 \times 1 \times 128 \times 128$	$128 \times 128 \times 128$
	Conv. + Leaky Relu	1	$3 \times 3 \times 128 \times 128$	$128 \times 128 \times 128$
	Conv. + Leaky Relu	1	$3 \times 3 \times 128 \times 128$	$128 \times 128 \times 128$
	Summation	–	–	$128 \times 128 \times 128$
	Conv. + Leaky Relu	1	$3 \times 3 \times 128 \times 128$	$128 \times 128 \times 128$
	Conv. + Leaky Relu	1	$1 \times 1 \times 128 \times 128$	$128 \times 128 \times 128$
$\times 1$	dConv. + Leaky Relu	1	$3 \times 3 \times 64 \times 128$	$256 \times 256 \times 64$
	Conv. + Leaky Relu	1	$3 \times 3 \times 64 \times 32$	$256 \times 256 \times 32$
	Conv. + Leaky Relu	1	$1 \times 1 \times 32 \times 16$	$256 \times 256 \times 16$
	Conv. + Leaky Relu	1	$3 \times 3 \times 16 \times B$	$256 \times 256 \times B$

Table 2: Details of the Simulated Large-scale Benchmark.

Spatial Size	Name	Original data				Processing		
		Resolution	Bit depth	Format	Container	Step	Num. of frames	Num. of measurements
512×512	Beauty	1920×1080	8	YUV	RAW	5	64	8
	Bosphorus	1920×1080	8	YUV	RAW	5	64	8
	HoneyBee	1920×1080	8	YUV	RAW	1	64	8
	Jockey	1920×1080	8	YUV	RAW	1	64	8
	ShakeNDry	1920×1080	8	YUV	RAW	3	64	8
1024×1024	Beauty	1920×1080	8	YUV	RAW	5	64	8
	Jockey	1920×1080	8	YUV	RAW	5	64	8
	ReadyGo	1920×1080	8	YUV	RAW	5	64	8
	ShakeNDry	1920×1080	8	YUV	RAW	3	64	8
	YachtRide	1920×1080	8	YUV	RAW	9	64	8
2048×2048	City	3840×2160	8	YUV	RAW	8	64	8
	Kids	3840×2160	8	YUV	RAW	5	64	8
	Lips	3840×2160	8	YUV	RAW	5	64	8
	Night	3840×2160	8	YUV	RAW	5	64	8
	RiverBank	3840×2160	8	YUV	RAW	5	64	8

to obtain high-quality reconstructions. To balance the computational complexity and performance, we set overlapping size as 128 to evaluate MetaSCI with other methods in the main paper.

Table 3: PSNR on large-scale video reconstruction with different overlapping sizes.

Overlapping size	Scale		
	512×512	1024×1024	2048×2048
0	34.81	34.18	31.87
64	35.25	34.66	32.28
128	35.47	34.89	32.49
192	35.50	34.87	32.51

Figure 2: The reconstructed example frames based on different overlapping size. Clearly, with overlapping size as 0, there is some boundary effect. With the increase of overlapping size, the reconstruction becomes better.

References

[1] Alexandre Mercat, Marko Viitanen, and Jarno Vanne. Uvg dataset: 50/120fps 4k sequences for video codec analysis and development. In *Proceedings of the 11th ACM Multimedia Systems Conference*, pages 297–302, 2020. 1