Table 1. Quantitative evaluation of **our method** on BasicLFSR test dataset. Four different models w.r.t. τ and ablation models are compared. Reported scores are PSNR/SSIM; larger is better for both. † indicates that model was trained with second loss term.

Configuration	train τ	test τ	EPFL	HCI (new)	HCI (old)	INRIA	Stanford	ALL
Fixed-τ-low	0.075	0.075	34.76/0.9534	32.04/0.8522	39.53/0.9612	36.11/0.9454	29.80/0.8973	34.45/0.9219
Fixed-τ-mid	0.15	0.15	34.50/0.9503	31.22/0.8337	38.86/0.9548	35.81/0.9437	28.78/0.8762	33.83/0.9118
Fixed-τ-high	0.3	0.3	34.04/0.9474	30.90/0.8287	38.41/0.9531	35.48/0.9421	28.13/0.8684	33.39/0.9080
Flexible-τ	Random	0.075	34.01/0.9459	31.38/0.8434	38.41/0.9531	35.19/0.9358	28.80/0.8783	33.56/0.9113
		0.15	34.30/0.9502	31.26/0.8424	38.95/0.9570	35.68/0.9442	28.75/0.8806	33.79/0.9149
		0.3	32.98/0.9348	29.60/0.7839	35.35/0.9013	34.47/0.9358	26.87/0.8213	31.85/0.8754
Image only	–	–	27.35/0.8639	25.43/0.7016	31.13/0.8462	29.08/0.8854	21.28/0.6841	26.85/0.7962
Events only†	Random	0.15	16.29/0.5747	12.98/0.4669	18.99/0.5835	14.92/0.5997	10.88/0.4720	14.81/0.5394
Events only	Random	0.15	28.40/0.8447	27.17/0.7035	32.26/0.8300	29.43/0.8611	24.50/0.7403	28.35/0.7963

Table 2. Quantitative evaluation of **other imaging methods** on BasicLFSR test dataset. Second column ("#") shows number of acquired images. CA captures one or more images, while full 4-D, JAEC, and LA capture single image.

Method	#	EPFL	HCI (new)	HCI (old)	INRIA	Stanford	ALL
CA ($N = 4$) + RecNet	4	35.52/0.9556	33.06/0.8796	40.10/0.9654	36.89/0.9471	31.39/0.9254	35.39/0.9346
CA ($N = 2$) + RecNet	2	34.06/0.9455	31.98/0.8599	38.82/0.9546	35.84/0.9414	29.76/0.9037	34.09/0.9210
CA ($N = 1$) + RecNet	1	27.78/0.8654	26.61/0.7251	31.31/0.8352	29.40/0.8915	22.99/0.7522	27.62/0.8139
Full-4D + RecNet	1	32.91/0.9336	31.26/0.8371	37.90/0.9434	34.88/0.9345	29.16/0.8895	33.22/0.9076
JAEC ($N = 4$) + RecNet	1	31.84/0.9195	30.05/0.8078	36.28/0.9213	33.36/0.9256	27.80/0.8569	31.97/0.8862
JAEC ($N = 4$) [28]	1	30.38/0.9263	28.87/0.7732	34.96/0.8293	33.17/0.9685	26.14/0.8473	30.70/0.8689
LA + RecNet	1	24.26/0.6843	26.17/0.6714	30.81/0.7978	25.85/0.7628	24.03/0.6926	26.22/0.7218
LA (naive)	1	22.25/0.5820	24.75/0.6050	28.65/0.7157	23.71/0.6829	22.42/0.5934	24.36/0.6358

Full-4D [42] is an idealized hypothetical imaging model without physical hardware implementations, which enables arbitrary 4-D coding while capturing a single image:

$$I_{x,y} = \sum_{u,v} m_{x,y,u,v} L_{x,y,u,v} \tag{14}$$

where $m_{x,y,u,v} \in [0, 1]$. With these three methods, we appended the same amount of noise as our method to the observed images ($\sigma = 0.005$ w.r.t. the intensity range $[0, 1]$). As indicated by "+RecNet", each method was combined with a light-field reconstruction network that had the same architecture as RecNet.[4] The network was trained for each imaging method from scratch; the coding patterns and weights in RecNet were jointly optimized on the same dataset as ours.[5] We also used the software of JAEC provided by Mizuno et al. [28], which had 12 times the parameters of our RecNet, and was retrained on the same dataset as ours. Finally, to simulate **lens-array based imaging (LA)** [1, 2, 31, 32], which can take a light field in a single shot, we down-sampled each view of a light field into the 1/8×1/8 spatial resolution, and up-sampled it into the original resolution using bicubic interpolation ("naive"). We also used RecNet to enhance the quality of the up-sampled light field ("+RecNet").[6]

CA ($N = 4$) can be regarded as the upper-bound reference for our method since our method can obtain a set of data that is *quasi-equivalent* to the four coded-aperture images (See 3.2). Aligned with this theory, the quantitative scores of our method (fixed-τ-low) were close to those of CA ($N = 4$). Our method with a moderate configuration (flexible-τ, test $\tau = 0.15$) still performed comparably to CA ($N = 2$) and outperformed CA ($N = 1$), full-4D, JAEC, and LA. Our method (flexible-τ) was also stable over a wide range of τ. As shown in Fig. 5. our method (flexible-τ) consistently ($\tau \in [0.075, 0.275]$) outperformed the other imaging methods that can complete the measurement in a single exposure (CA ($N = 1$), JAEC, and LA).

Several visual results are presented in Fig. 6. Our method obtained a visually-convincing result comparable to that of the reference (CA with $N = 4$). The image-only model reconstructed the overall appearance but lost some details and the consistency among the viewpoints. The result obtained with the events-only† model seems somewhat consistent among the viewpoints but lacking correct intensity. Please refer to the supplementary video for more results with better visualization.

[4] With CA, N observed images were stacked along the channel dimension and fed to RecNet. With JAEC and full4D, we lifted the observed image to a sparse tensor with 64 channels before feeding to RecNet, as was done in a previous study [42].

[5] With JAEC, we fixed the coding patterns for the imaging plane, $p^1, \ldots, p^4$, to those provided by Mizuno et al.' [28] to ensure the compatibility with the hardware constraints.

[6] The up-sampled 64 views were stacked along the channel dimension and fed to RecNet, and RecNet was trained on the same dataset as ours.