JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Blur Interpolation Transformer for Real-World Motion from Blur

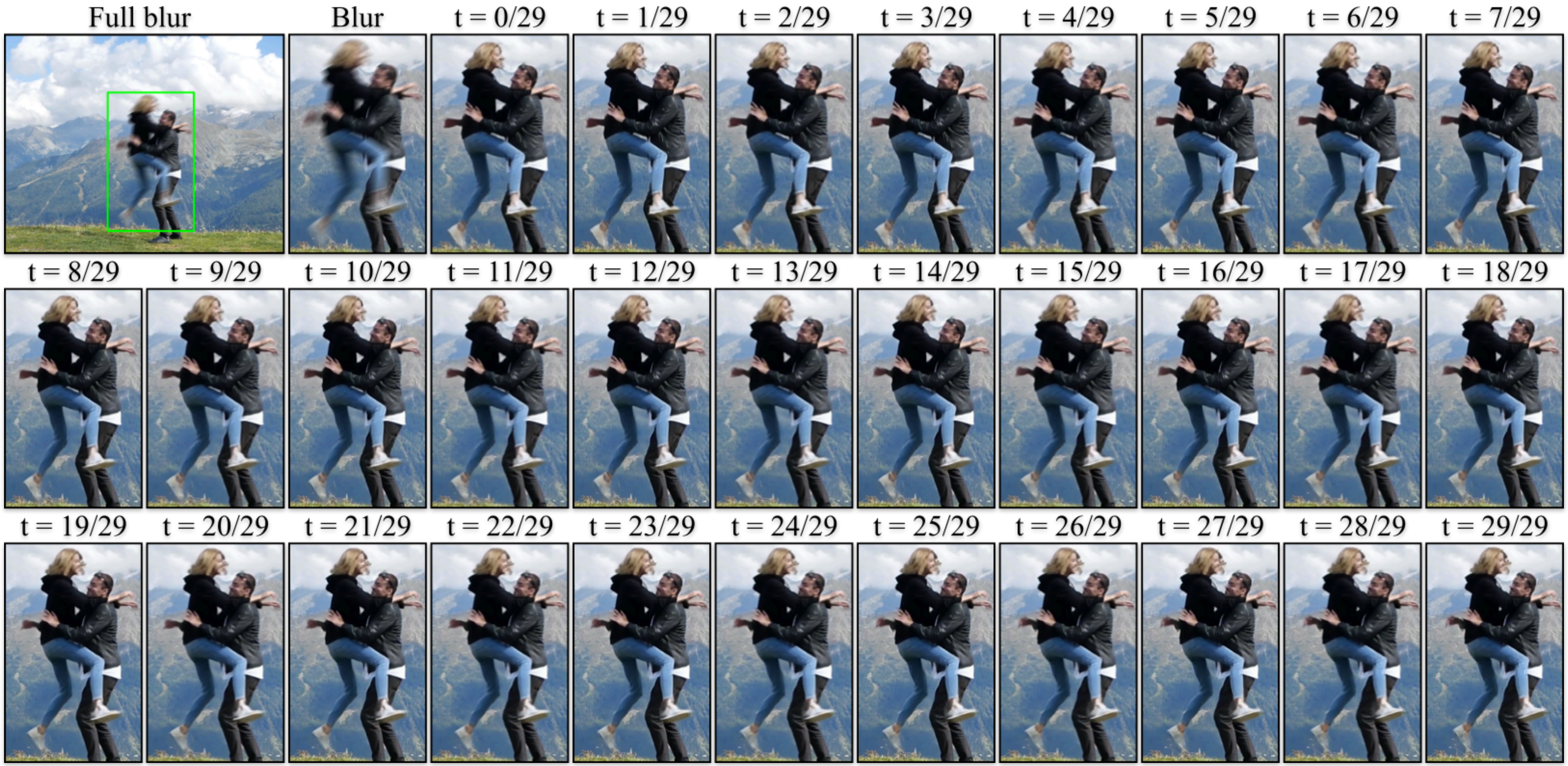
Zhihang Zhong, Mingdeng Cao, Xiang Ji,
Yinqiang Zheng, and Imari Sato

The University of Tokyo
National Institute of Informatics

https://zzh-tech.github.io/BiT/

Arbitrary time blur interpolation

Transforming a motion blurred image into a sharp video clip



Arbitrary time blur interpolation

Arbitrary temporal super-resolution via motion blur

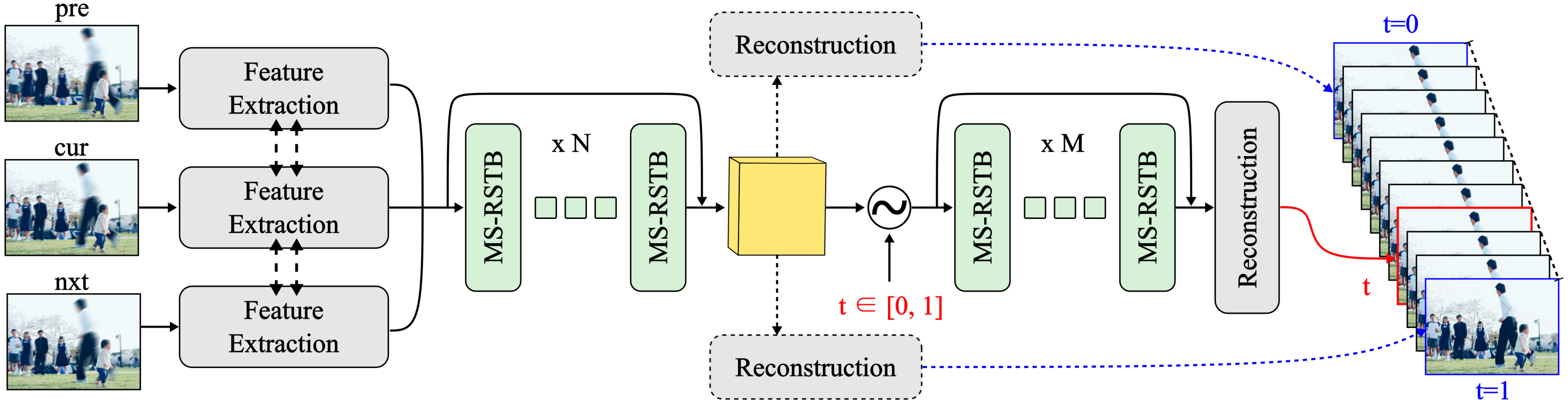
Existing challenges

Performance and generalization

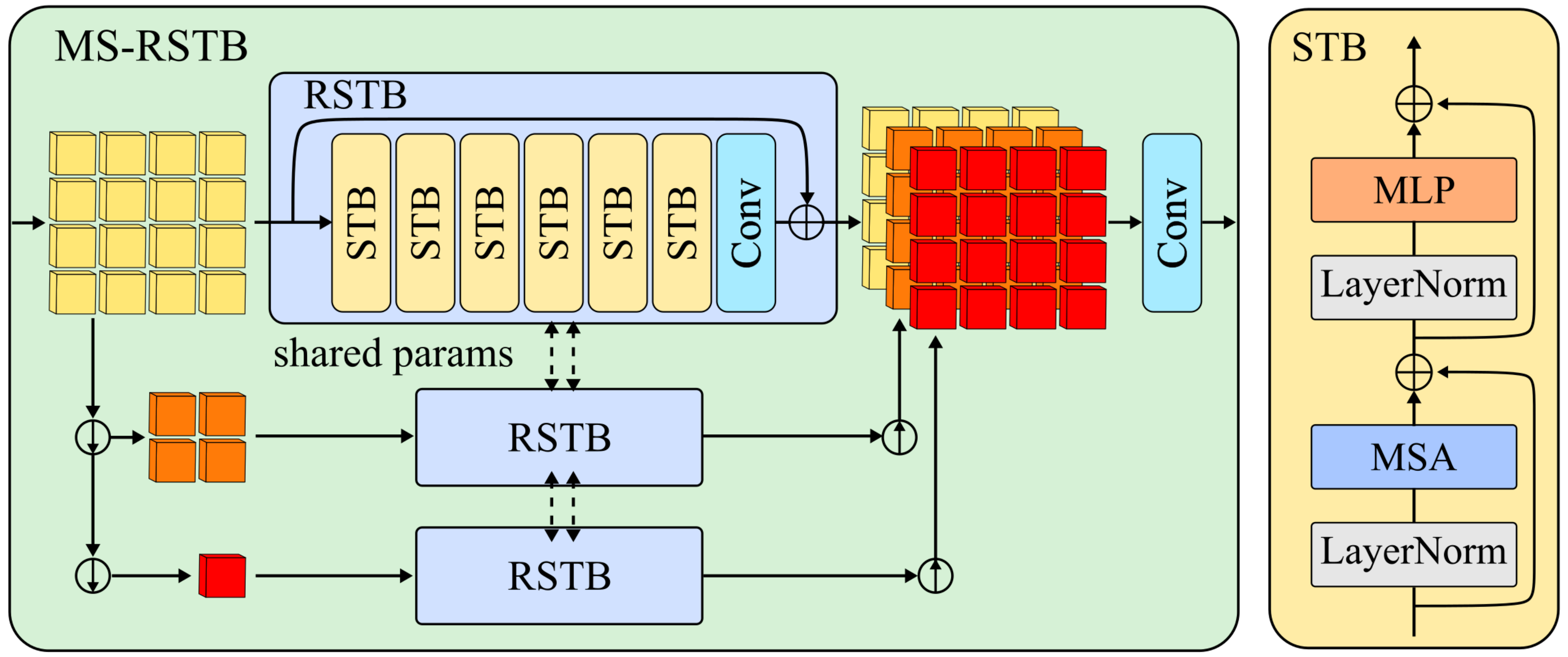
- The current methods still leave considerable room for improvement in terms of visual quality even with synthetic datasets
  - Blur interpolation Transformer (BiT)
- Poor generalization to real-world data
  - Real-world Blur Interpolation dataset (RBI)

Blur interpolation Transformer

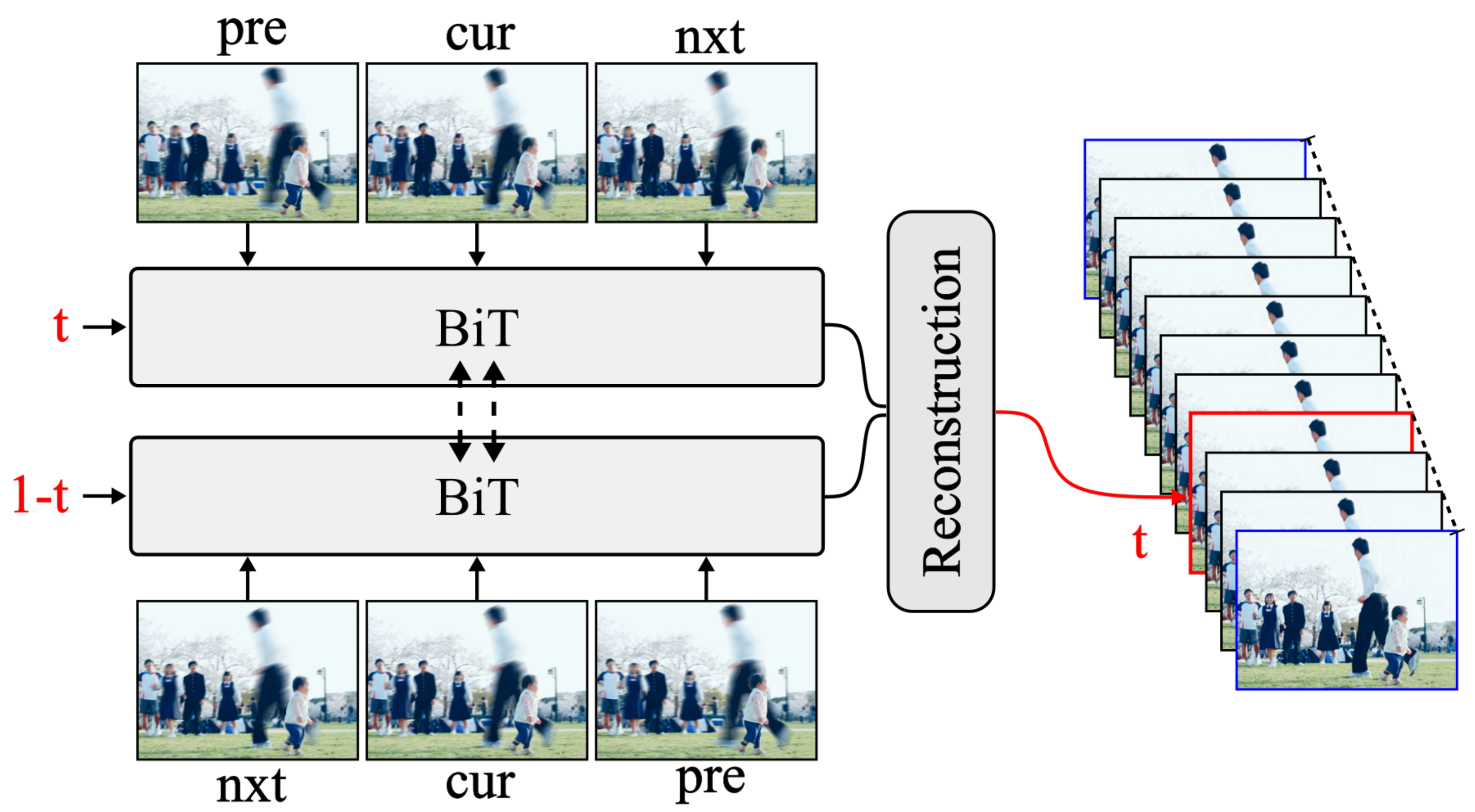
Efficient blur interpolation network boosted by temporal strategies



(a) Overview of blur intra-interpolation transformer



(b) Multi-scale residual Swin transformer block

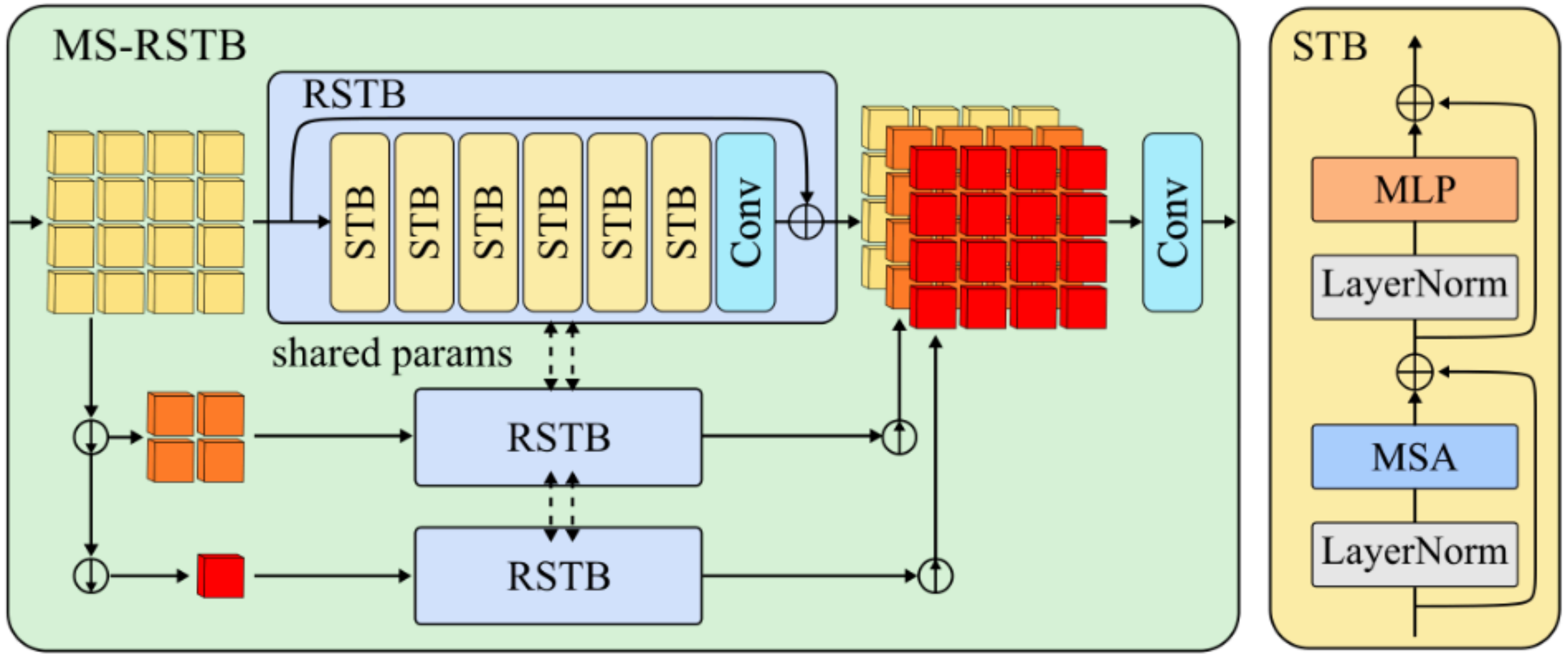


(c) Temporally symmetric ensembling

Blur interpolation Transformer

Multi-scale residual Swin transformer block (MS-RSTB)

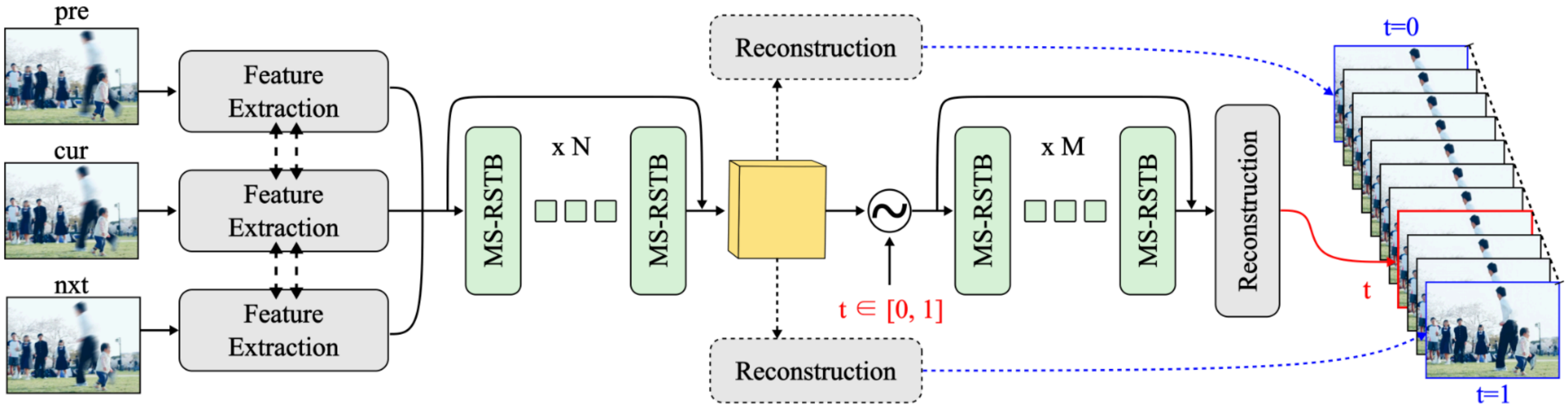
- Multi-scale residual Swin transformer block to efficiently tackle different blur-scales and merge information from nearby frames in a coarse-to-fine manner



Blur interpolation Transformer

Dual-ended temporal supervision strategy (DTS)

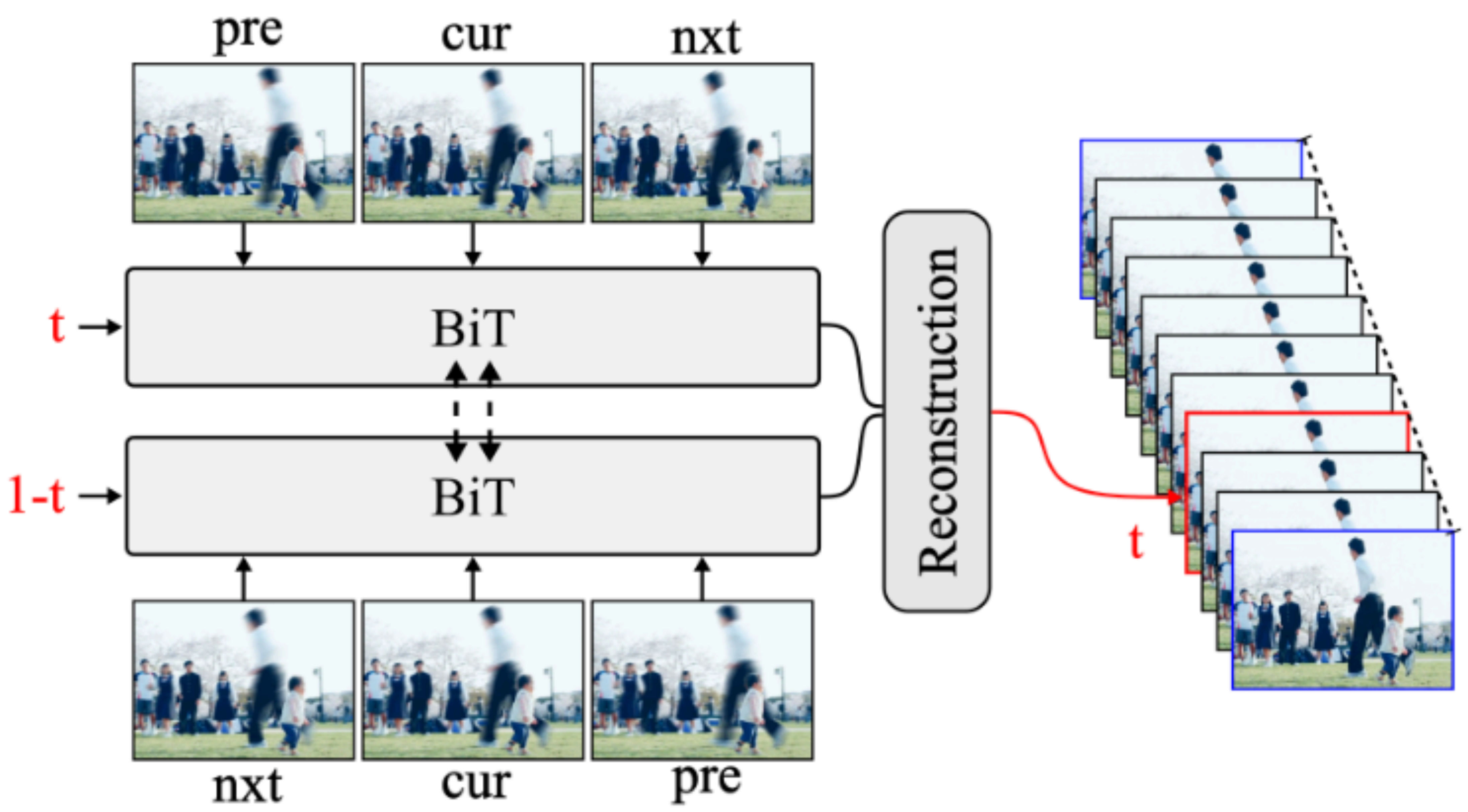
- Dual-ended temporal supervision is used to “underpin and spread” the shared intermediate features, making them more conducive to the reconstruction of the latent sharp frame at arbitrary time instance



Blur interpolation Transformer

Temporal symmetric ensembling strategy (TSE)

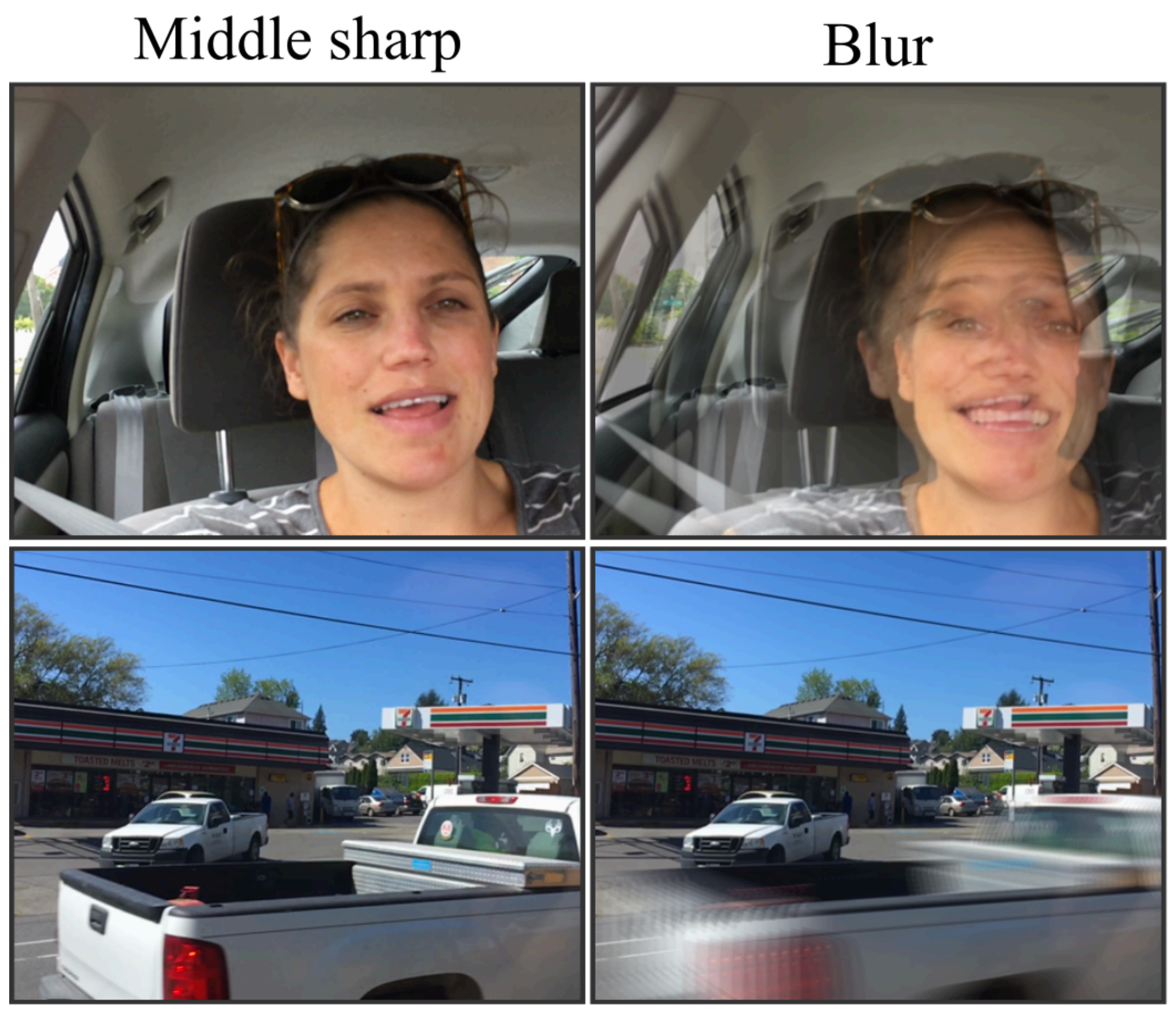
- Using temporal symmetric properties to further enhance the features for reconstruction



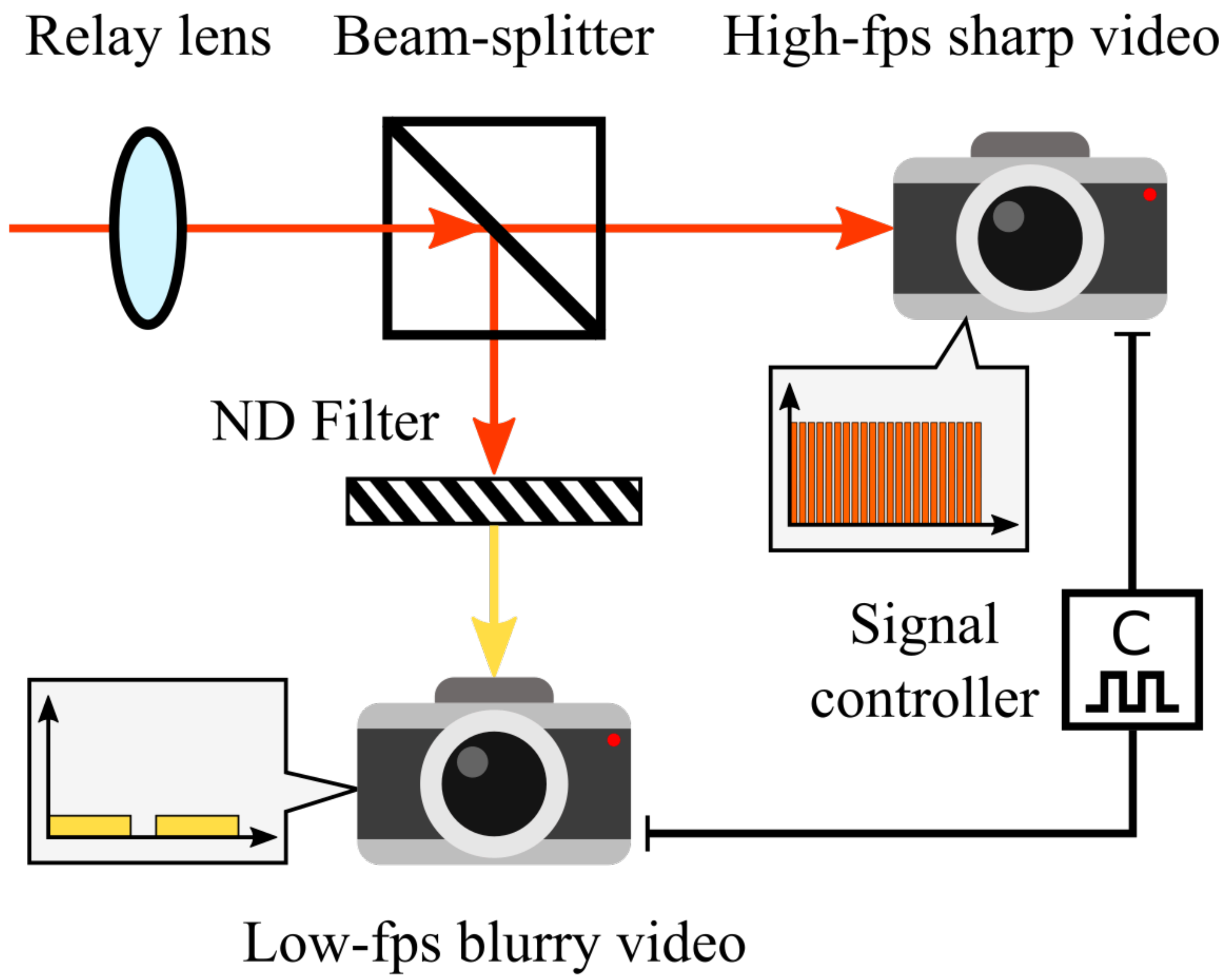
Blur interpolation dataset

Beam-splitter-based co-axis camera system

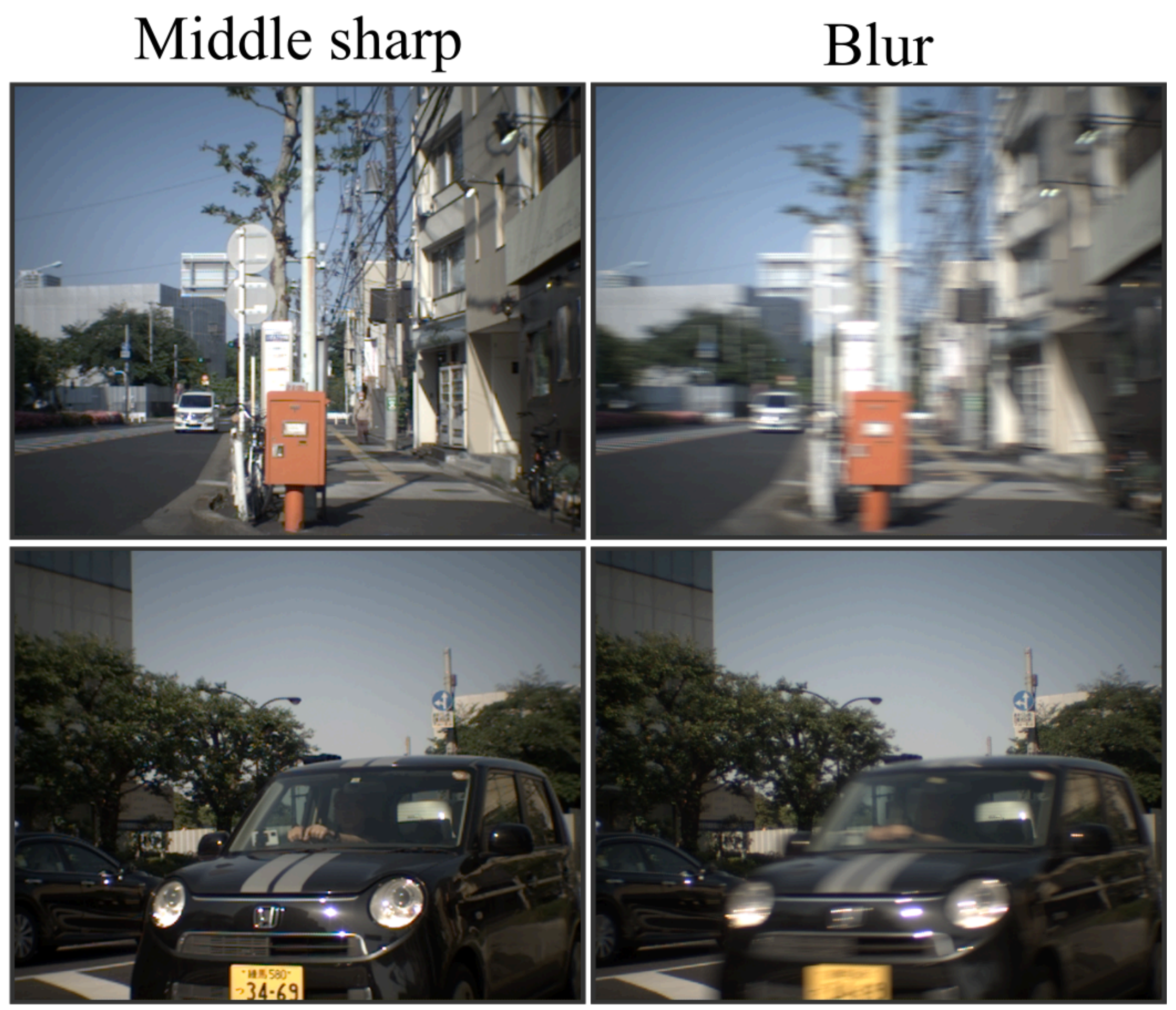
- Using discrete consecutive frames to synthesize the blur will cause significant discontinuities, and models trained on such data have poor generalization on real-world data



(a) Samples of Adobe240



(b) Hybrid-camera system



(c) Samples of RBI

Experiments

Quantitative comparison

- Our method surpasses the prior arts with a significant margin while also being much faster (BiT: w/o TSE; BiT++: w/ TSE)

Table 1. **Comparison with the state-of-the-arts on synthetic dataset Adobe240 and our real-world dataset RBI. Red** denotes the best performance, and blue denotes the second best performance. Runtime is calculated uniformly using images from the Adobe240 dataset with size of 640 × 352 on a single RTX2080 Ti GPU.

	Adobe240		RBI		Runtime		
	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	1x [s] ↓	60x [s] ↓	Params [M] ↓
EDVR [49]+XVFI [43]	33.19	0.934	28.17	0.847	0.294	17.64	29.2
Jin *et al.* [9]	32.47	0.924	27.73	0.853	0.250	15.00	10.8
RPF_4 [40]	33.32	0.935	28.55	0.872	0.746	44.76	11.4
DeMFI [28]	34.34	0.945	29.03	0.895	0.513	30.78	**7.41**
BiT	34.34	0.948	29.90	0.900	**0.203**	**5.76**	11.3
BiT++	**34.97**	**0.954**	**30.45**	**0.908**	0.395	11.64	11.3

Experiments

Ablation studies

- Table 2 verifies the effectiveness of the proposed module and strategies
- Table 3 indicates large N makes inference faster with slight performance loss

Table 2. **Ablation studies.** BiT w/o MS denotes BiT using single-scale RSTB module. BiT w/o DTS denotes BiT without dual-end temporal supervision. BiT+ denotes BiT that has the same training epochs as BiT++.

	Adobe240					RBI				
	BiT w/o MS	BiT w/o DTS	BiT	BiT+	BiT++	BiT w/o MS	BiT w/o DTS	BiT	BiT+	BiT++
PSNR ↑	33.96	34.10	34.34	34.52	**34.97**	29.40	29.44	29.90	29.99	**30.45**
SSIM ↑	0.944	0.946	0.948	0.946	**0.954**	0.893	0.894	0.900	0.901	**0.908**

Table 3. **Effect of # of MS-RSTB.** The performance is evaluated on Adobe240 using BiT.

	$N=0$ $M=6$	$N=1$ $M=5$	$N=2$ $M=4$	$N=3$ $M=3$	$N=4$ $M=2$	$N=5$ $M=1$	$N=6$ $M=0$
PSNR ↑	34.08	34.09	34.18	**34.34**	34.30	34.05	27.13
SSIM ↑	0.947	0.942	0.943	**0.948**	**0.948**	0.944	0.832
60x Runtime [s] ↓	11.34	9.36	7.98	5.76	4.02	2.16	**0.36**

Experiments

Qualitative comparison

(a) Comparison on Adobe240 dataset

(b) Comparison on RBI dataset

Experiments

Cross-validation between synthetic and real-world datasets

- In figure 5 (a), testing on real-world samples from RBI, we can observe severe artifacts in the results of the model trained on Adobe240
- Conversely, in figure 5 (b), the model trained on RBI does not introduce artifacts to synthetic samples from Adobe240

(a) Test samples from RBI (b) Test samples from Adobe240