Motion Blur Decomposition with Cross-shutter Guidance

Xiang Ji Haiyang Jiang Yinqiang Zheng†

The University of Tokyo, Japan

{jixiang,jiang-haiyang777}@g.ecc.u-tokyo.ac.jp,

yqzheng@ai.u-tokyo.ac.jp

CVPR 2024

Presenter: Hao Wang

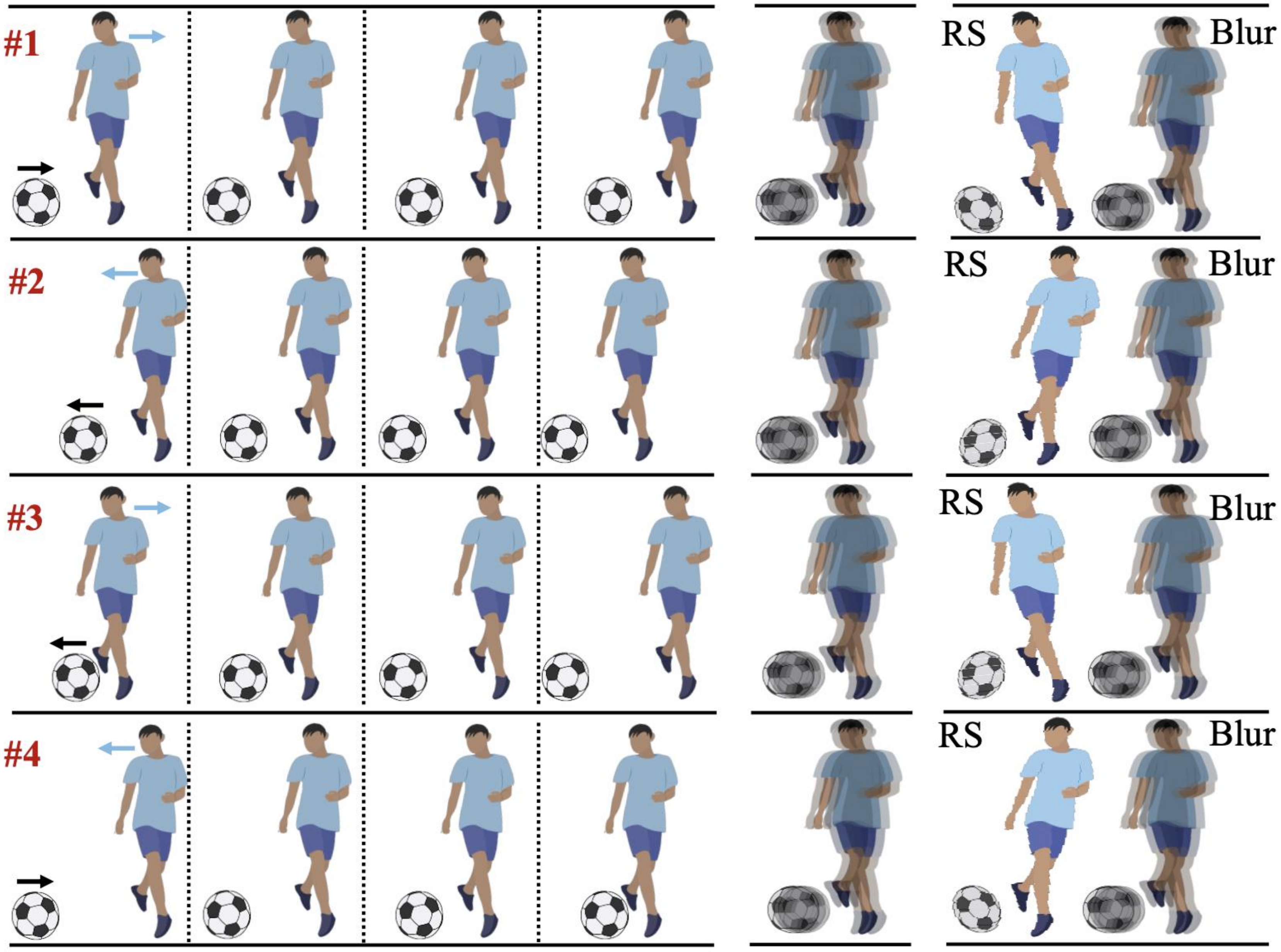
Advisor: Prof. Chia-Wen Lin

Outline

- Introduction
- Framework
- Method
- Experiment
- Conclusion

Introduction

- Decomposing a blurry image into **multiple sharp images**
- **Motion ambiguity**
- Dual
  - global shutter (**GS**)
  - rolling shutter (**RS**)



(a) Possible state of movement (b) Blur view (c) Dual view

Introduction

- New setting of **dual Blur-RS combination** to address the motion ambiguity of blur decomposition.
  - ➢ RS branch offers **local details** and **disambiguates motion directions**.
  - ➢ Blur (GS) counterpart with **full global context** will elevate the accuracy of **motion magnitude**.

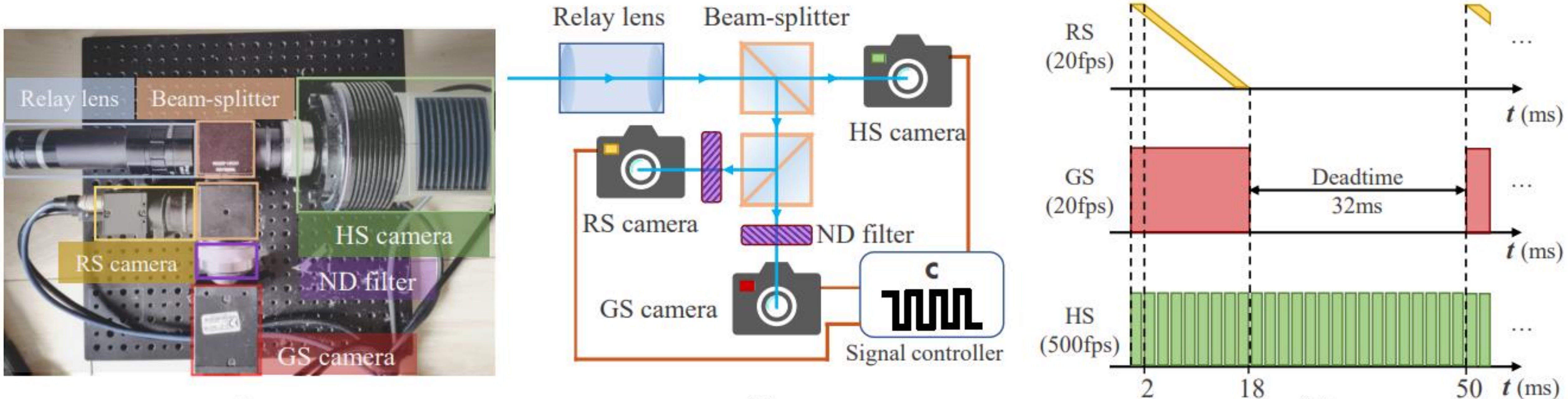
- **Triaxial imaging system** that simultaneously captures **Blur-RS** pairs along with high-speed **ground truth**, and collect a real dataset named **RealBR**.

- Novel neural network architecture.

Outline

- Introduction
- Framework
- Method
- Experiment
- Conclusion

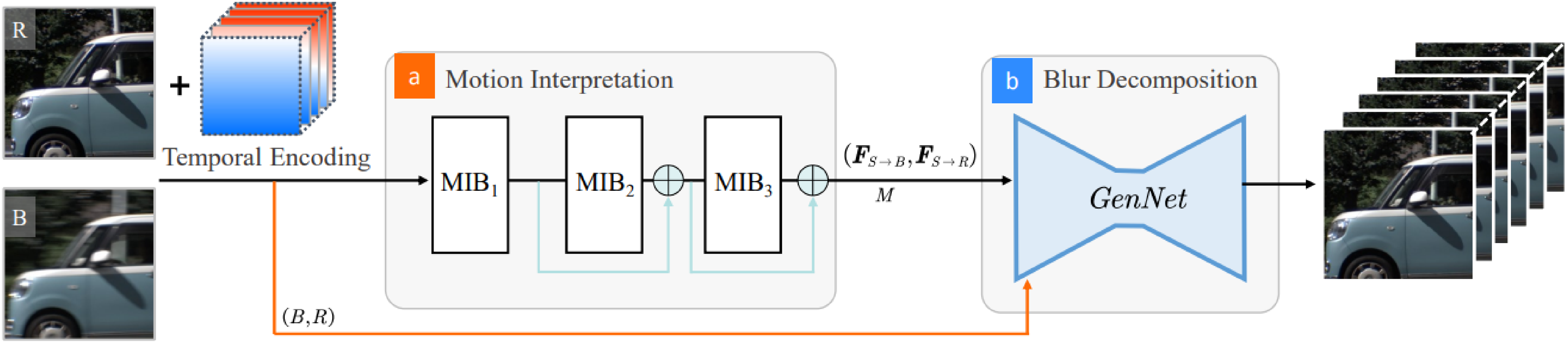
Dataset



(a) (b) (c)

- **54** distinct street **scenes**
    - containing objects, like vehicles and pedestrians, and various camera motions.
- In **each scene**, we have **56 pairs of consecutive RS and GS blur** frames, and 1400 corresponding sharp HS images.

Framework

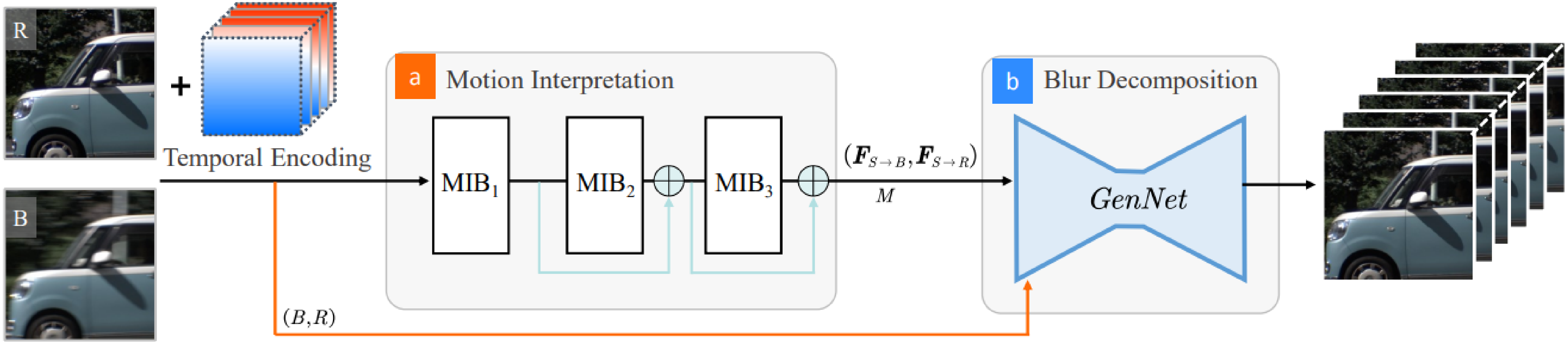


(a) Overall architecture of our proposed model

Outline

- Introduction
- Framework
- Method
- Experiment
- Conclusion

Architecture

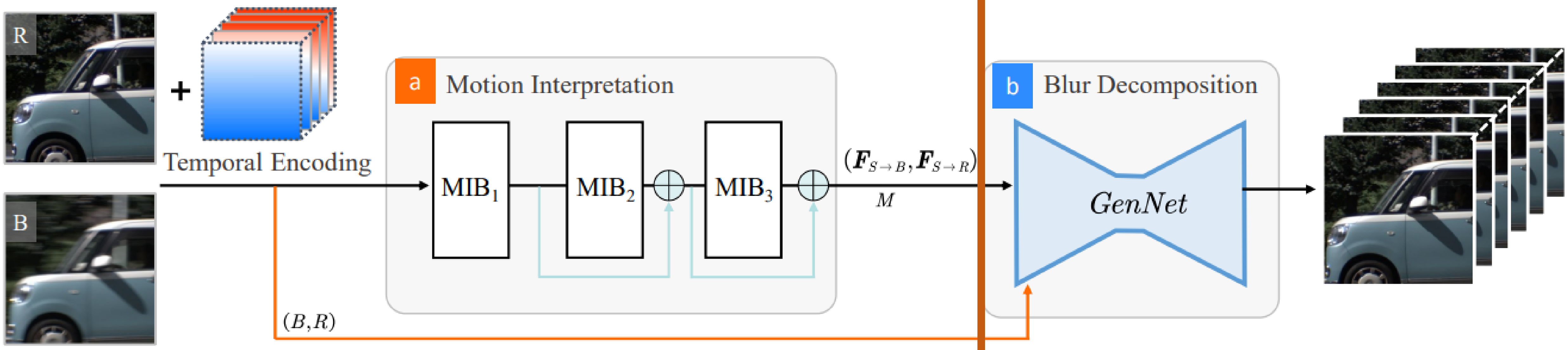


(a) Overall architecture of our proposed model

$$\mathbf{S} = \mathcal{F}\left(B, R\right) \qquad (2)$$

$$\mathbf{S} = \{S^t, t \in 0, \cdots, N-1\}$$

Architecture

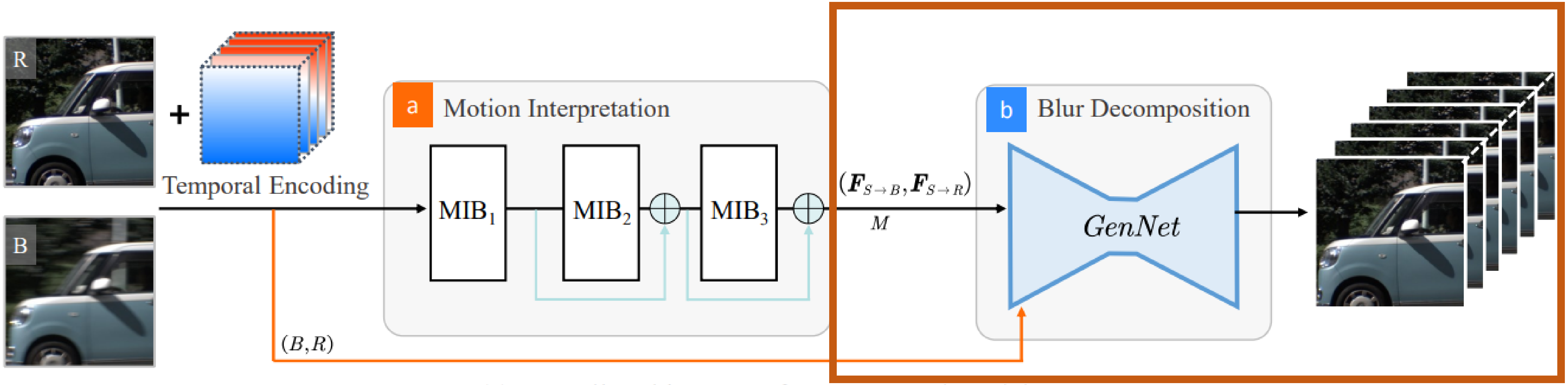


(a) Overall architecture of our proposed model

$$(\mathbf{F}_{S\to B}, \mathbf{F}_{S\to R}, M) = \mathcal{MI}(B, R) \tag{3}$$

$$\mathbf{F}_{S\to B} = \{F_{S^t\to B}, t \in 1, \cdots, N\}$$

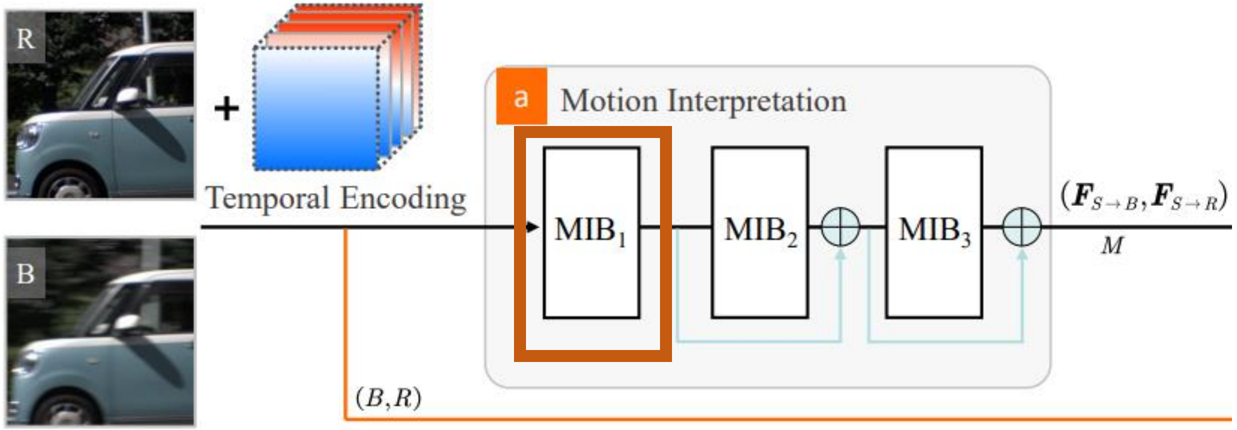
Architecture

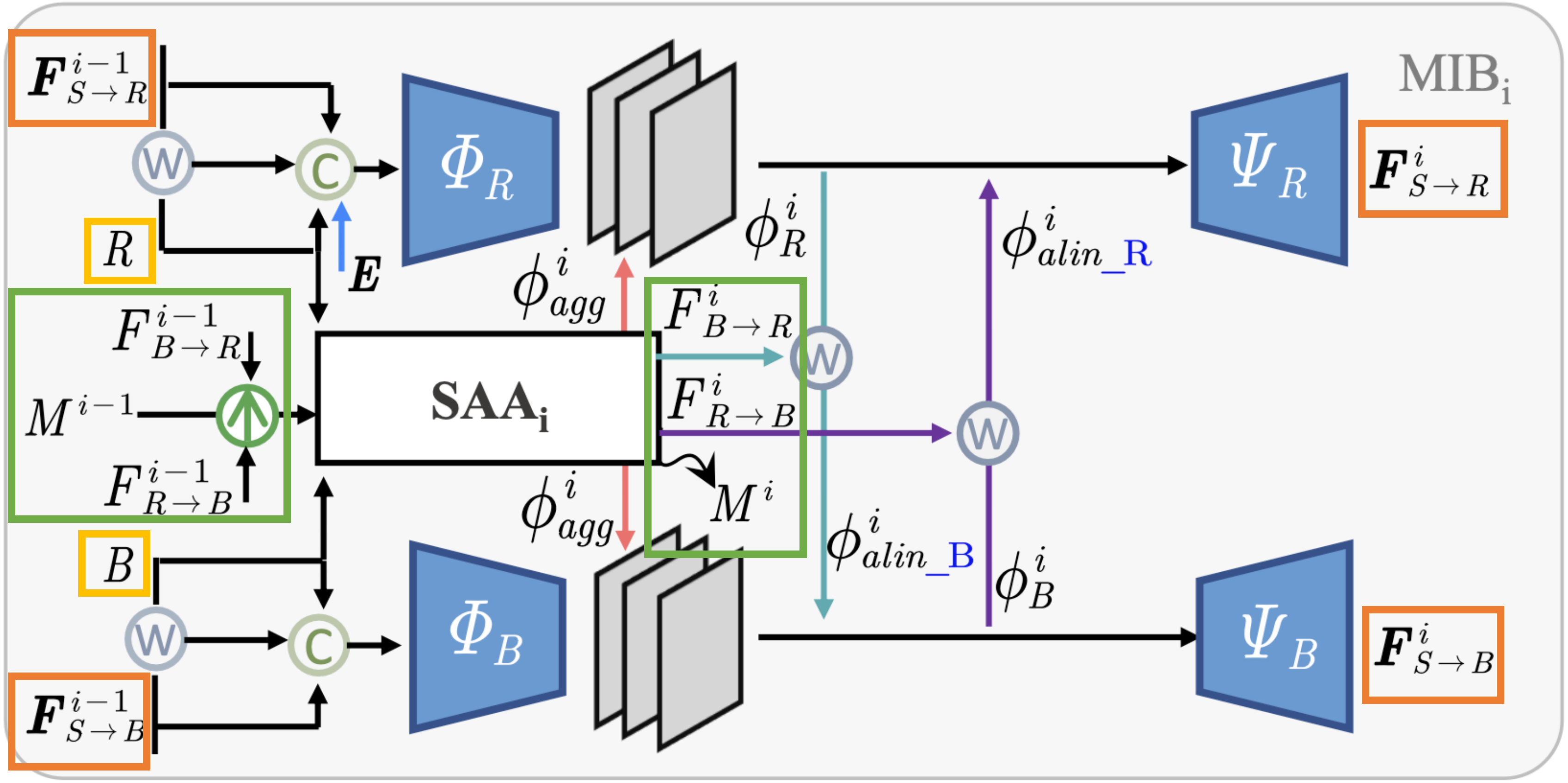


(a) Overall architecture of our proposed model

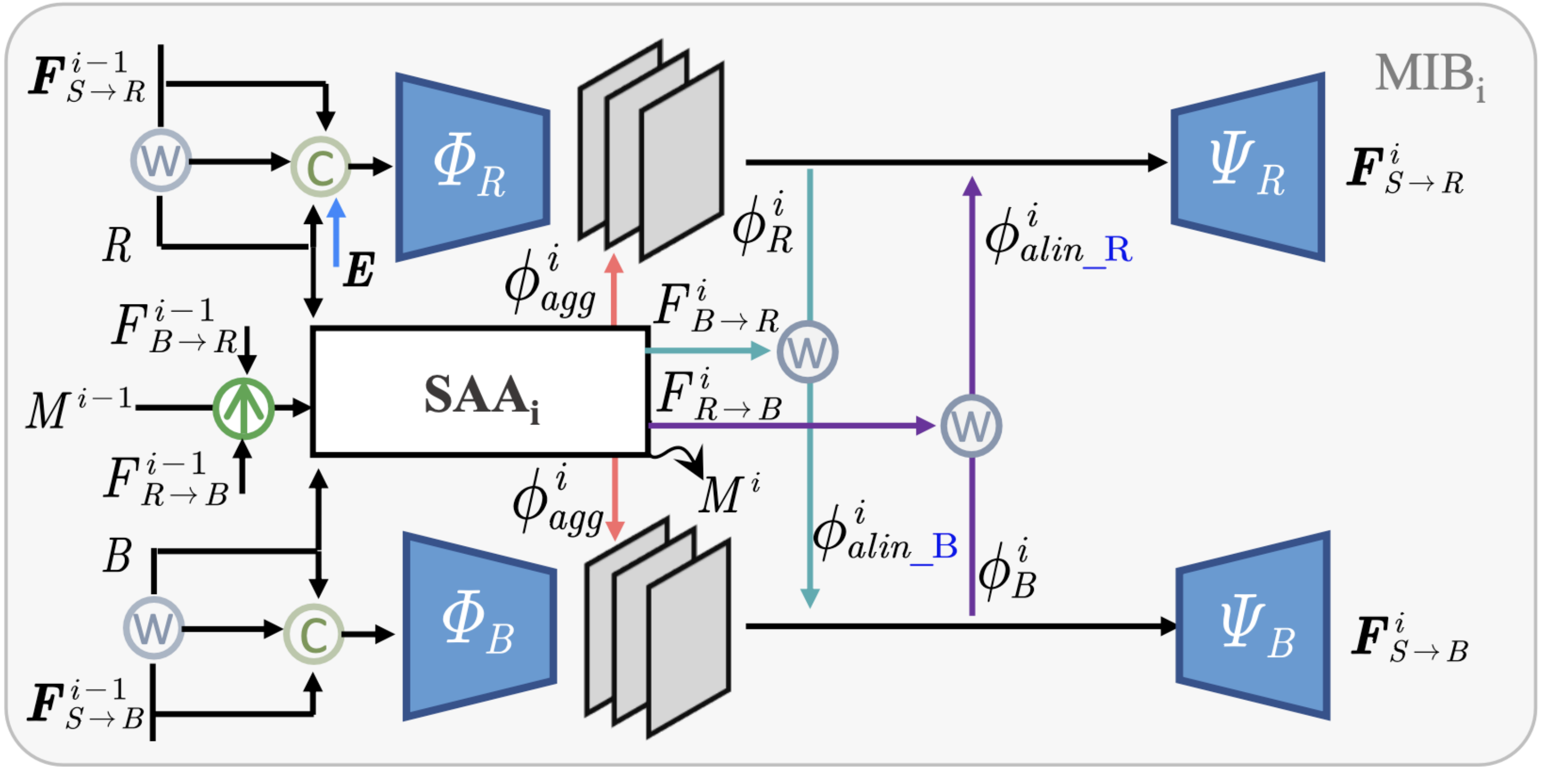
$$\mathbf{S} = GenNet\left(B, R, \mathbf{F}_{S\rightarrow B}, \mathbf{F}_{S\rightarrow R}, M\right) \qquad (4)$$

Motion interpretation block





Motion interpretation block



$$\phi^i_{alin\_R} = \mathcal{W}\left(\phi^i_B, F^i_{R\to B}\right)$$

$$\phi^i_{alin\_B} = \mathcal{W}\left(\phi^i_R, F^i_{B\to R}\right)$$

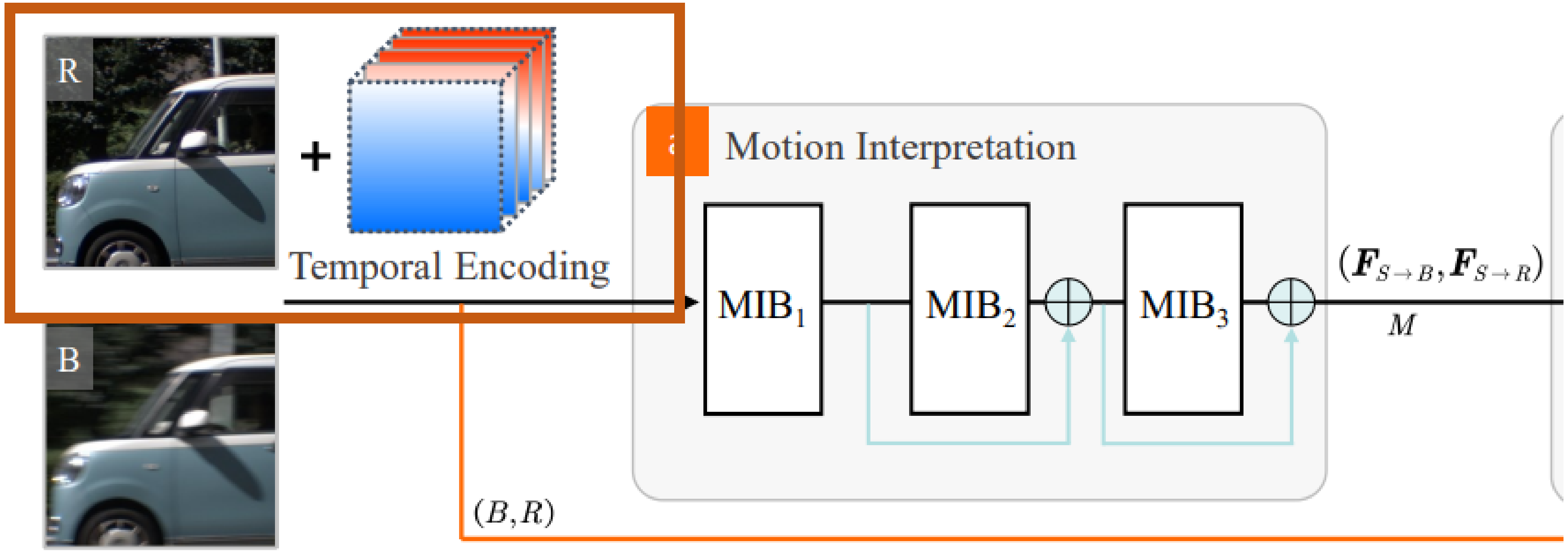
$$\phi^i_B = \Phi_B\left(B, \mathbf{F}^{i-1}_{S\to B}\right)$$

$$\mathbf{F}^i_{S\to B} = \Psi_B\left(\left[\phi^i_B, \phi^i_{alin\_B}, \phi^i_{agg}\right]\right)$$

$$\mathbf{F}^i, F^i, M^i = \mathcal{MIB}_i\left(B, R, \mathbf{E}, \mathbf{F}^{i-1}, F^{i-1}, M^{i-1}\right)$$

Temporal Positional Encoding

- Further enhance model's ability to **disambiguate motion direction** of latent frames

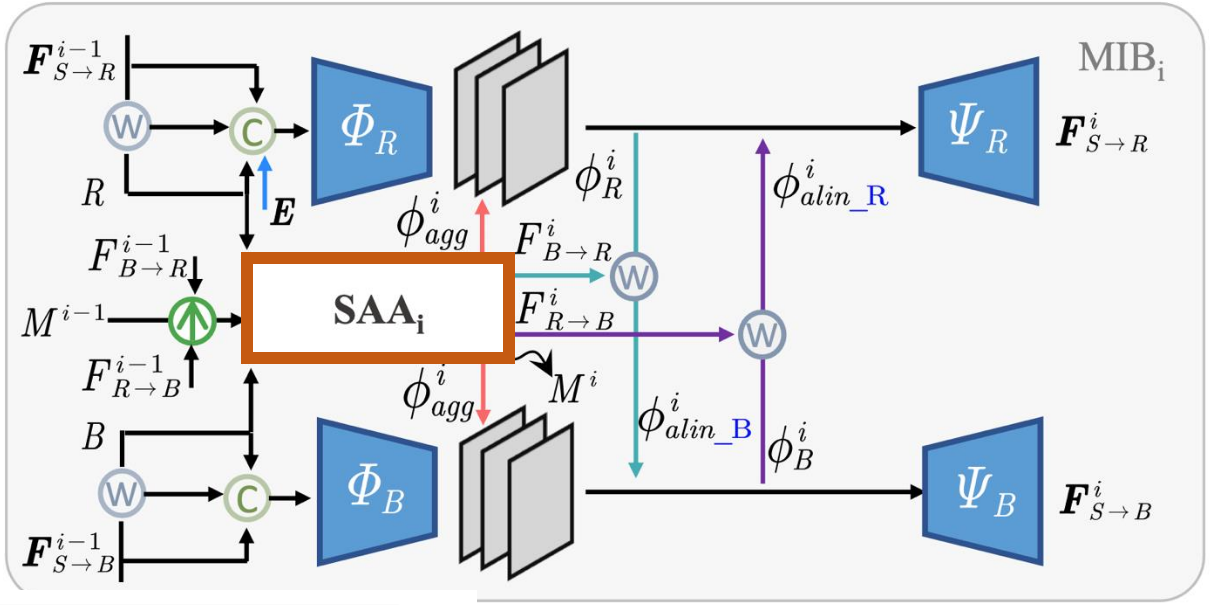


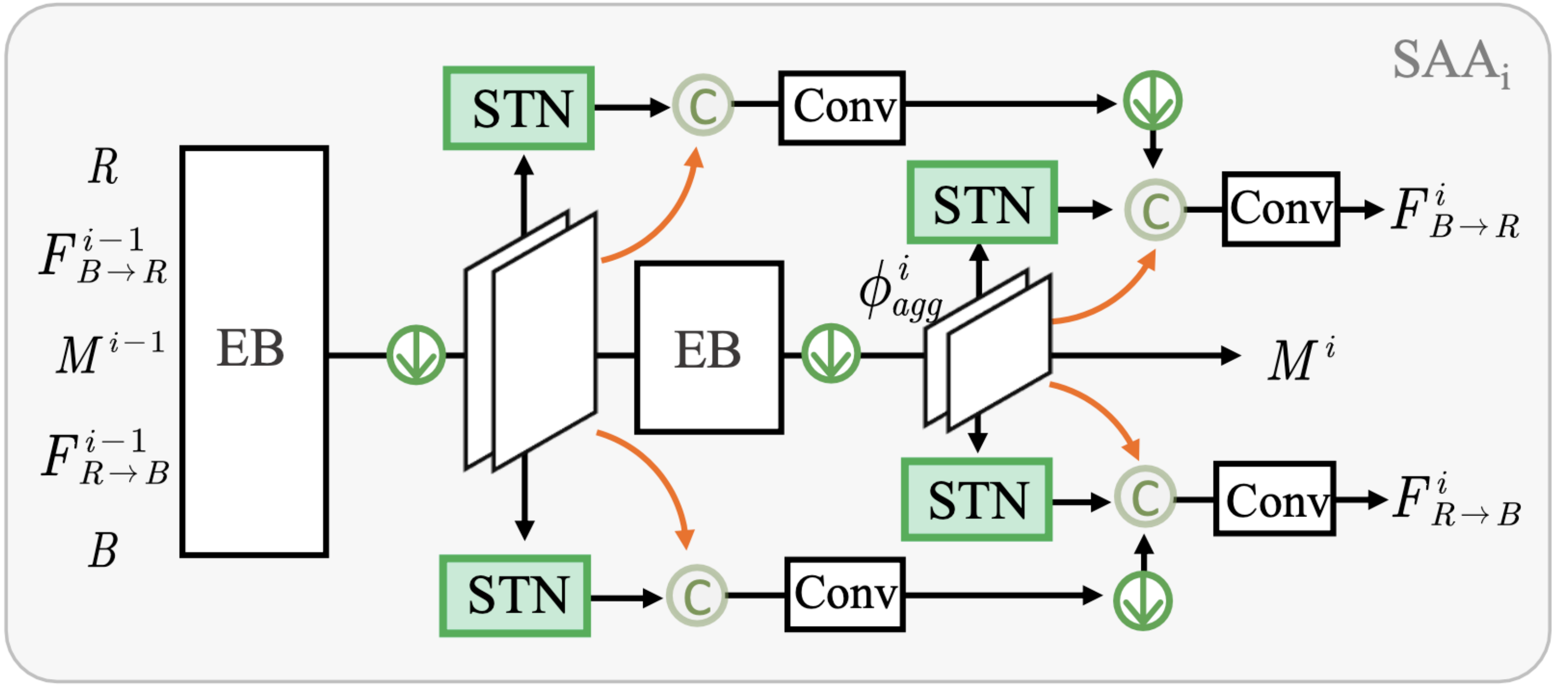
$$[E_R]_k = k, k = 0, 1, \cdots, N-1$$

$$E_{S^t} = \frac{H-1}{N-1} t \cdot \mathbb{1}$$

$$\mathbf{E} = \{(E_R - E_{S^t}), t = 0, 1, \cdots, N-1\}$$

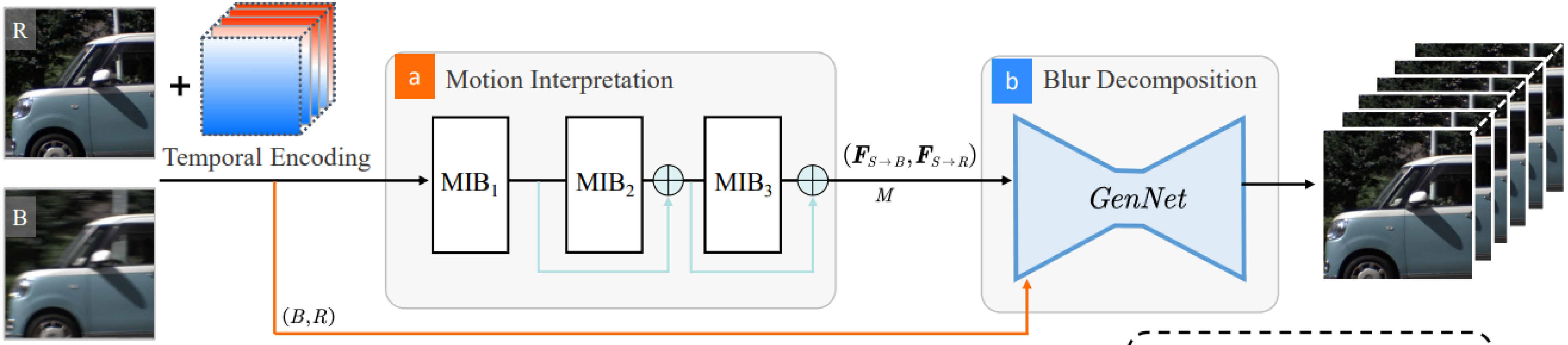
Shutter Alignment and Aggregation (SAA)



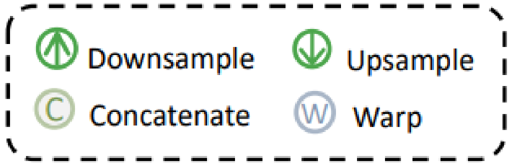


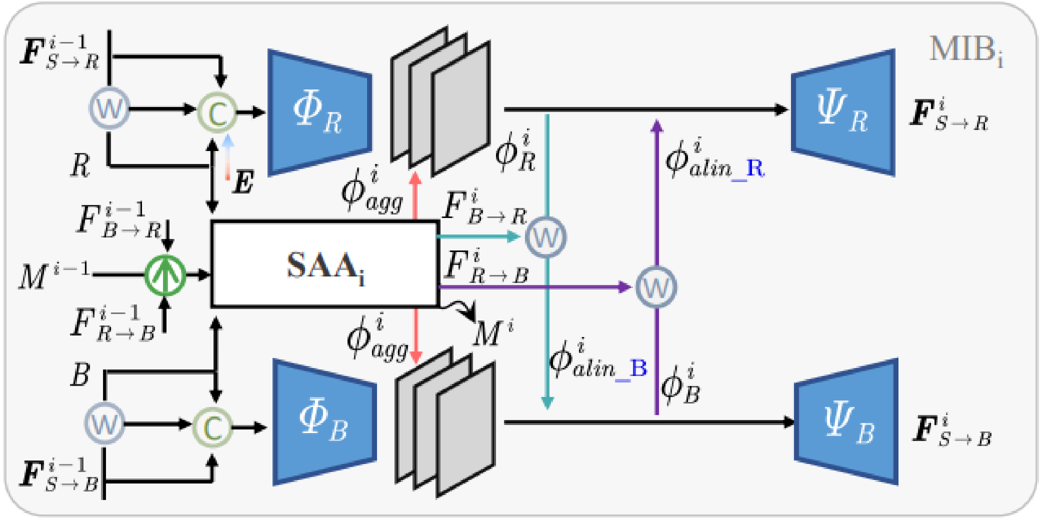
$$F^i, M^i = \mathcal{SAA}_i\left(B, R, F^{i-1}, M^{i-1}\right)$$

Architecture

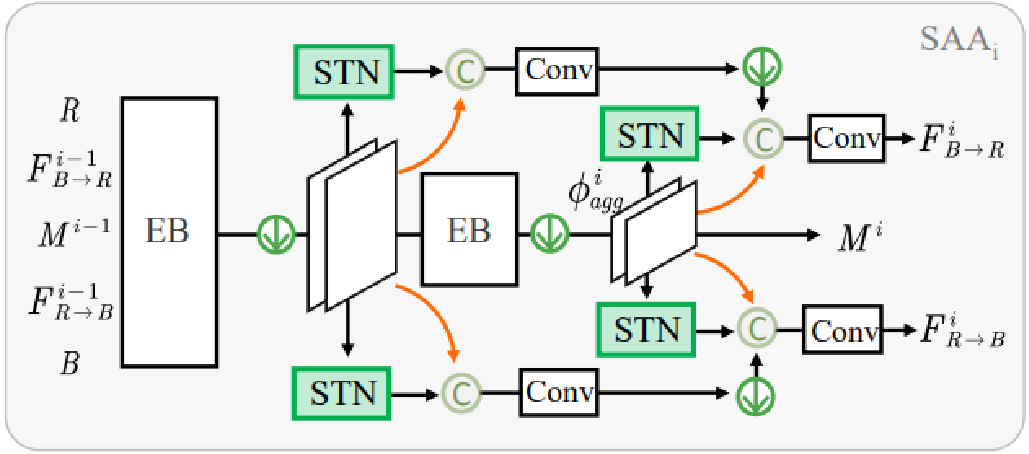


(a) Overall architecture of our proposed model





(b) Motion Interpretation Block (MIB)



(c) Shutter Alignment and Aggregation (SAA)

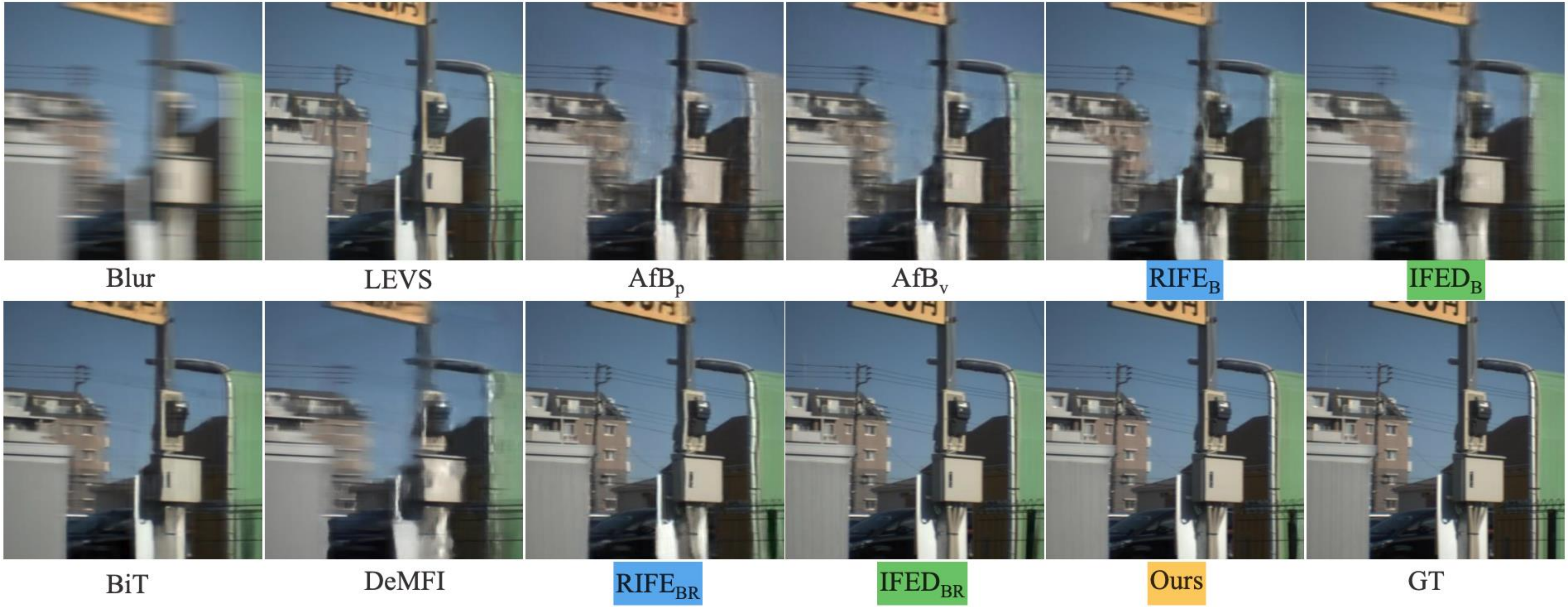
Outline

- Introduction
- Framework
- Method
- Experiment
- Conclusion

Quantitative comparisons

Method	Input	×3 PSNR	×3 SSIM	×3 LPIPS	×5 PSNR	×5 SSIM	×5 LPIPS	×9 PSNR	×9 SSIM	×9 LPIPS	Time (s)	Params (M)	FLOPs (G)
LEVS [11]	*1·B*	21.77	0.7042	0.2886	21.62	0.7153	0.2683	21.83	0.7277	0.2535	1.47	15.9	304
AfB_p [36]	*2·B*	21.50	0.7596	0.4102	21.65	0.7648	0.4055	21.82	0.7686	0.4017	0.15	190	839
AfB_v [36]	*2·B*	22.83	0.7877	0.3904	22.96	0.7903	0.3883	23.10	0.7924	0.3860	0.22	129	793
$RIFE_B$ [8]	*2·B*	24.60	0.8172	0.2254	24.73	0.8199	0.2268	24.83	0.8219	0.2268	1.33	54.8	71.1
$IFED_B$ [35]	*2·B*	24.45	0.8105	0.1817	24.62	0.8141	0.1811	24.74	0.8164	0.1798	1.33	10.8	29.5
BiT [34]	*3·B*	21.90	0.7664	0.2583	21.88	0.7694	0.2574	22.02	0.7729	0.2546	0.11	11.3	57.4
DeMFI [19]	*4·B*	25.55	0.8485	0.2247	25.26	0.8466	0.2275	26.20	0.8577	0.2165	4.86	7.41	420
$RIFE_{BR}$ [8]	*B-R*	30.26	0.8983	0.1071	30.53	0.9030	0.1046	30.67	0.9053	0.1042	1.33	54.8	71.1
$IFED_{BR}$ [35]	*B-R*	30.46	0.9030	0.0467	30.70	0.9064	0.0445	30.84	0.9084	0.0434	1.33	10.8	29.5
Ours	*B-R*	30.87	0.9073	0.0696	31.05	0.9103	0.0684	31.15	0.9120	0.0678	1.30	105	183

Qualitative comparison



Quantitative comparisons

- **more competitive settings** under task of blur decomposition and RS **temporal super resolution** based on synthetic data

Method	Input	×3			×7		
		PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
LEVS [11]	*1·B*	17.27	0.6063	0.3410	16.64	0.58	0.3811
AfB_p [36]	*2·B*	23.38	0.7411	0.2271	23.41	0.7517	0.2183
AfB_v [36]		28.10	0.8760	0.1496	28.39	0.8815	0.1461
$RIFE_B$ [8]		31.26	0.9410	0.0896	31.49	0.9430	0.0892
$IFED_B$ [35]		29.46	0.9193	0.0897	29.75	0.9225	0.0874
BiT [34]	*3·B*	32.31	0.9234	0.0708	32.56	0.9266	0.0691
DeMFI [19]	*4·B*	27.57	0.9002	0.1332	27.44	0.8984	0.1304
PMB [23]	*B-SL*	35.48	0.9723	0.0349	35.11	0.9715	0.0324
EBFI [30]	*B-Event*	33.21	0.9568	0.0703	33.51	0.9591	0.0685
RSSR [4]	*2·R*	22.73	0.8116	0.1039	22.65	0.8090	0.1154
CVR [6]		23.50	0.8342	0.0818	23.47	0.8332	0.0815
$RIFE_R$ [8]		24.16	0.8318	0.1697	24.32	0.8365	0.1618
$IFED_R$ [35]		28.30	0.9122	0.0475	28.63	0.9181	0.0446
IFED [35]	*R-iR*	30.89	0.9417	0.0372	31.96	0.9530	0.0307
EvUnroll [37]	*R-Event*	33.06	0.9558	0.0737	33.48	0.9587	0.0699
$RIFE_{BR}$ [8]	*B-R*	34.49	0.9701	0.0398	35.02	0.9733	0.0366
$IFED_{BR}$ [35]		33.03	0.9627	0.0332	33.72	0.9675	0.0304
Ours		34.92	0.9732	0.0310	35.51	0.9764	0.0305


Event
RS
Inputs
EvUnroll
$I_{3/7}$
$I_{6/7}$
Short
Long
Inputs
PMB
$I_{3/7}$
$I_{6/7}$
RS
iRS
Inputs
IFED
$I_{3/7}$
$I_{6/7}$
Blur1
Blur2
Inputs
BiT
$I_{3/7}$
$I_{6/7}$
Event
Blur
Inputs
EBFI
$I_{3/7}$
$I_{6/7}$
RS
Blur
Inputs
Ours
$I_{3/7}$
$I_{6/7}$

Outline

- Introduction
- Framework
- Method
- Experiment
- Conclusion

Conclusion

- **Novel cross-shutter setting** for motion decomposition of a single blurry image, inspired by the complementary exposure characteristics of **GS** and **RS cameras.**

- Proposed a **novel network** architecture that actively addresses the **contextual characterization** and **temporal abstraction** in a mutual incentive manner.

- Experiments on real dataset and synthetic data have verified the effectiveness of algorithm and global-shutter/rolling-shutter dual imaging setting.