Learning Event-guided Exposure-agnostic Video Frame Interpolation via Adaptive Feature Blending

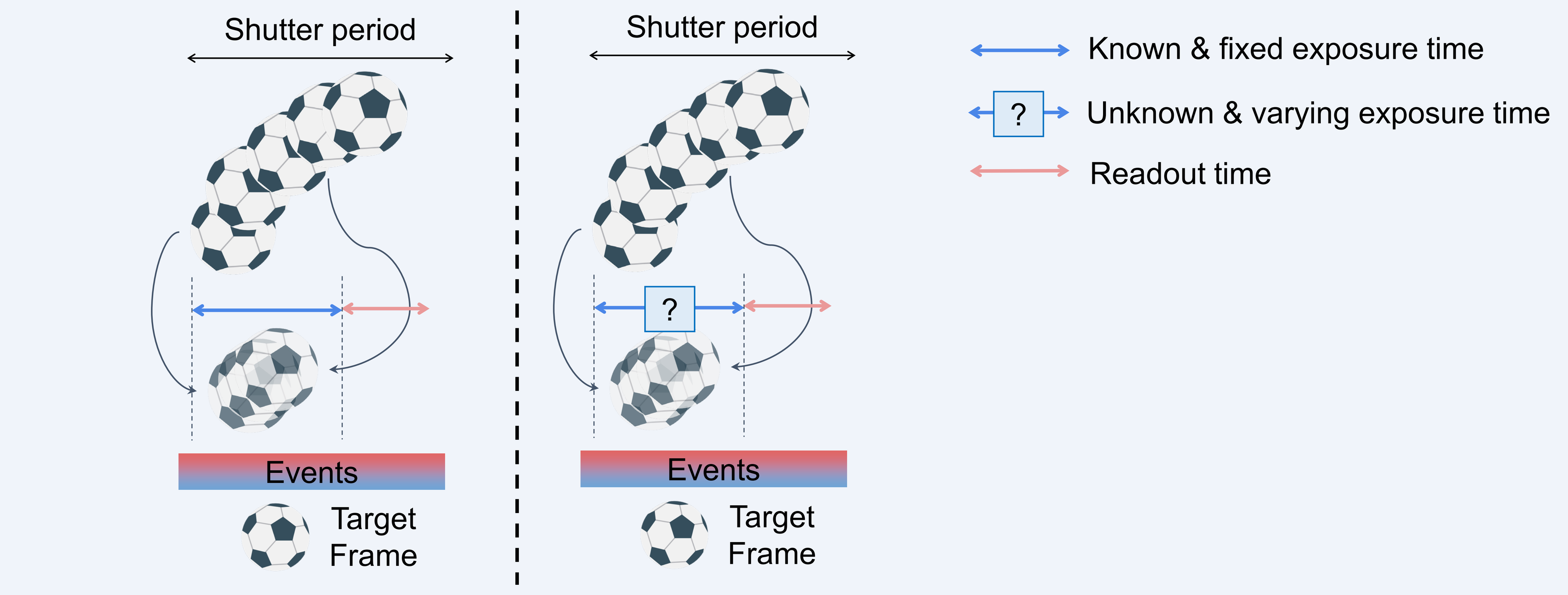
Junsik Jung[1], Yoonki Cho[1], Woo Jae Kim[1], Lin Wang[2], Sung−Eui Yoon[1]

Korea Advanced Institute of Science and Technology[1], Nanyang Technological University[2]

Background

Exposure-agnostic Event-guided Video Frame Interpolation

- Restoring a sharp frame from captured frames **without knowing the exposure time**, using **events as precise motion cues** to restore details lost due to blur and low frame rate.

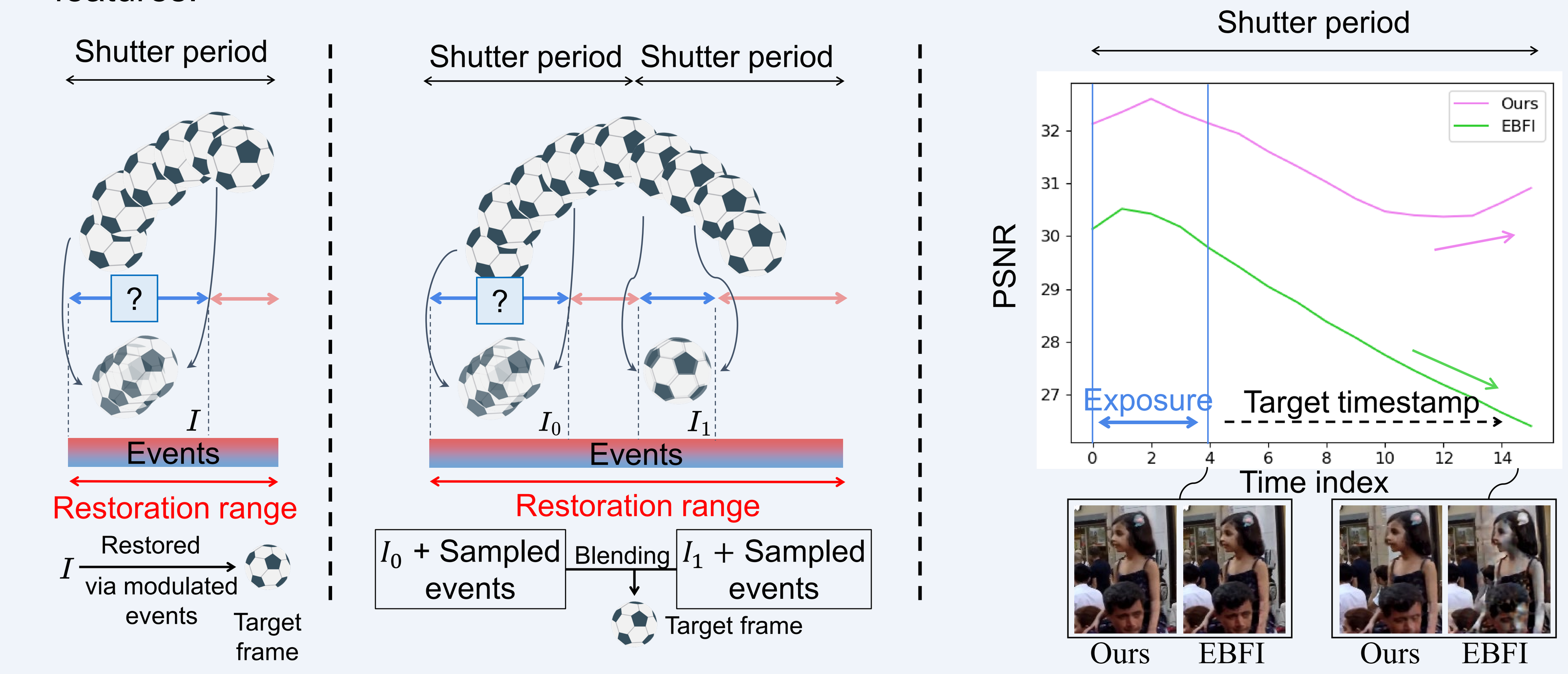


Motivation

- Existing methods do not utilize **temporal constraints** between consecutive frames and events
- As a result, when the target timestamp is far from the exposure time, the model lacks reliable temporal reference, leading to **noticeable degradation in sharpness and consistency**.

Proposed Solution

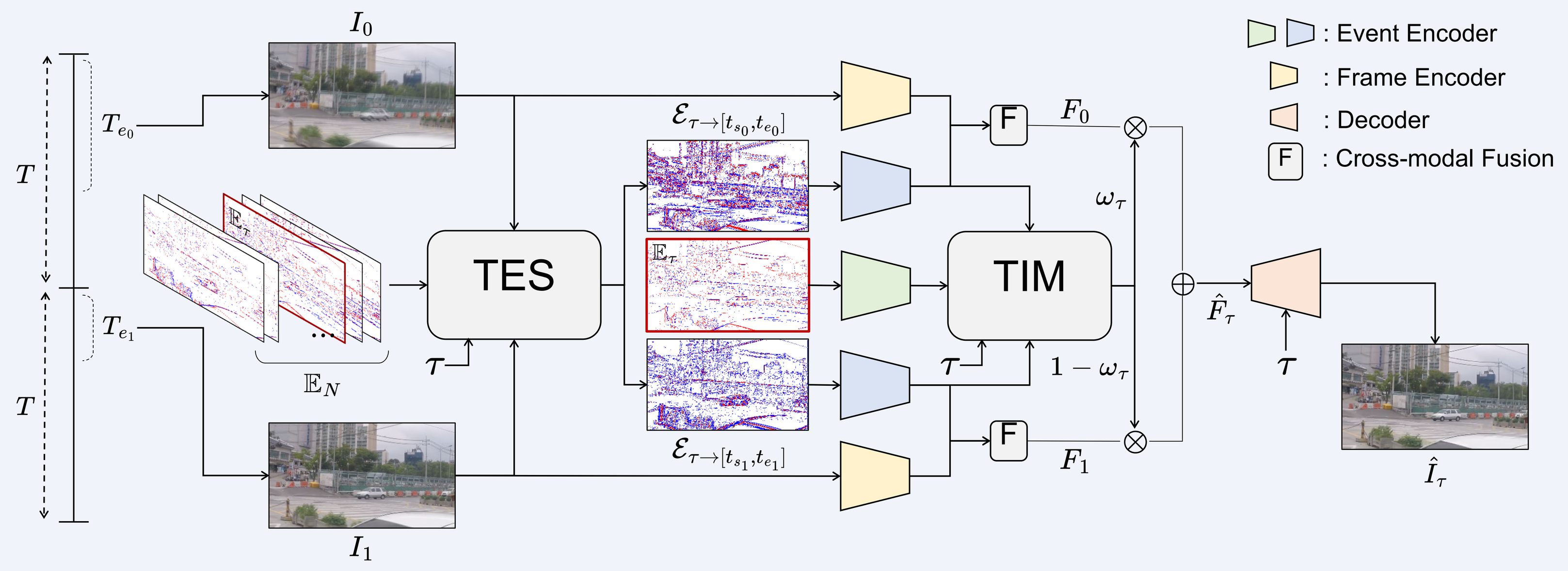
- We propose a **temporal-adaptive event-guided VFI framework** that addresses the limitations of prior blind-exposure methods by explicitly modeling **temporal relevance** between features.



Method

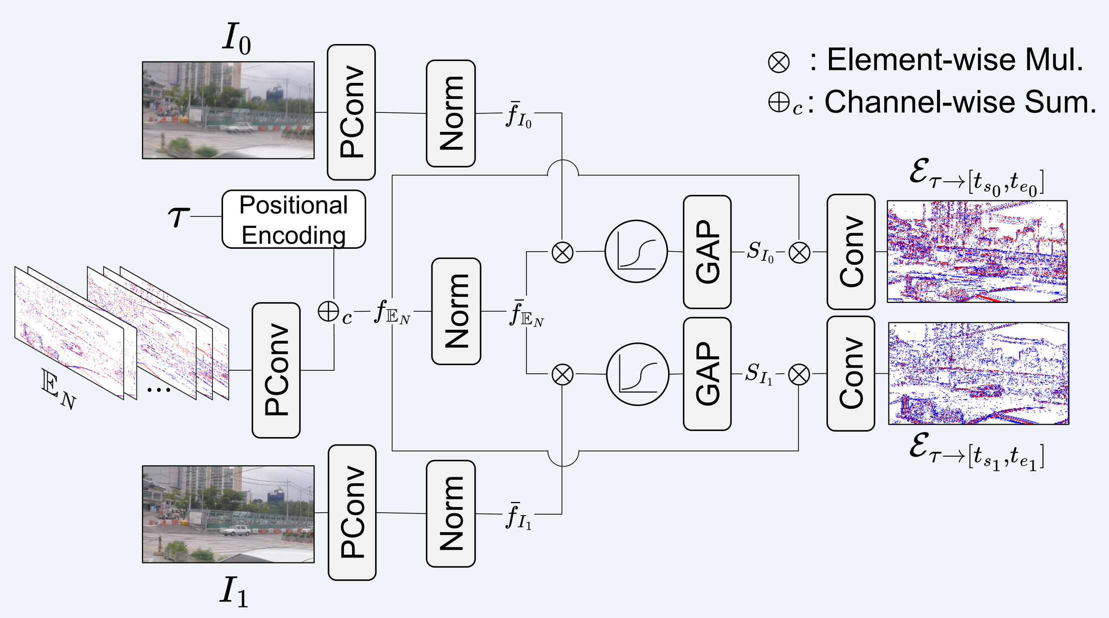
Overview

- Our framework takes two blurry frames (I_0, I_1) and the corresponding event stream, along with a target timestamp (τ), and synthesizes a sharp frame at that time $(\hat{I}_\tau)$



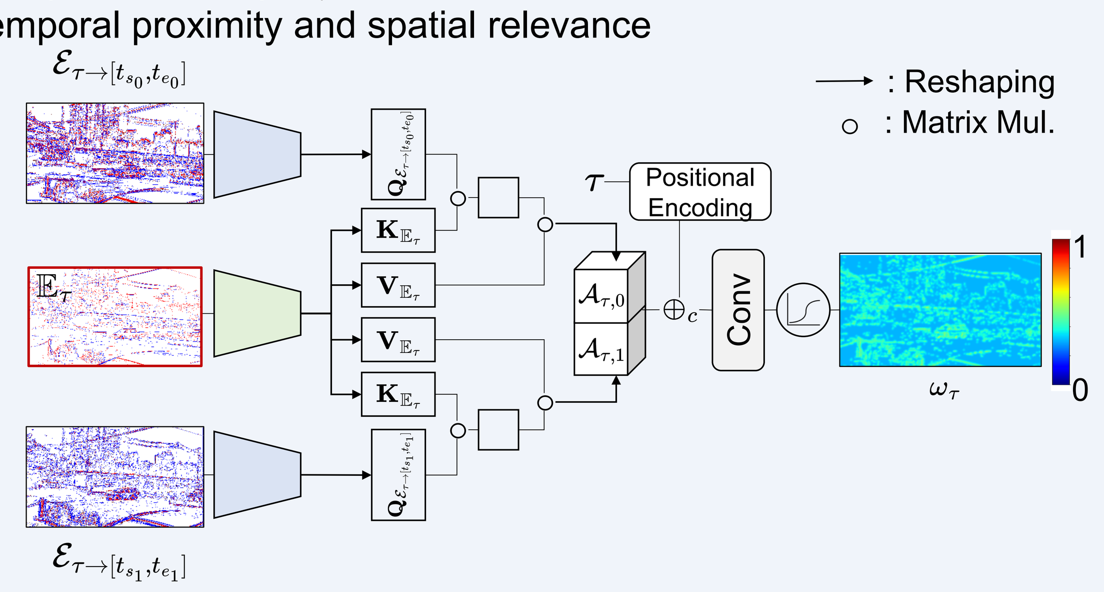
Target-adaptive Event Sampling Module (TES)

- Selects events that are temporally aligned with the target timestamp and the (unknown) exposure time
- Produces event-conditioned feature representations for each frame



Target-adaptive Importance Mapping Module (TIM)

- Predicts an importance map ω_τ that indicates how much each feature should contribute, based on temporal proximity and spatial relevance



Adaptive Feature Blending

- The final target feature is obtained by: $\hat{F}_\tau = \omega_\tau F_0 + (1 - \omega_\tau) F_1$
- A decoder then reconstructs the target sharp frame $(\hat{I}_\tau)$

Evaluation

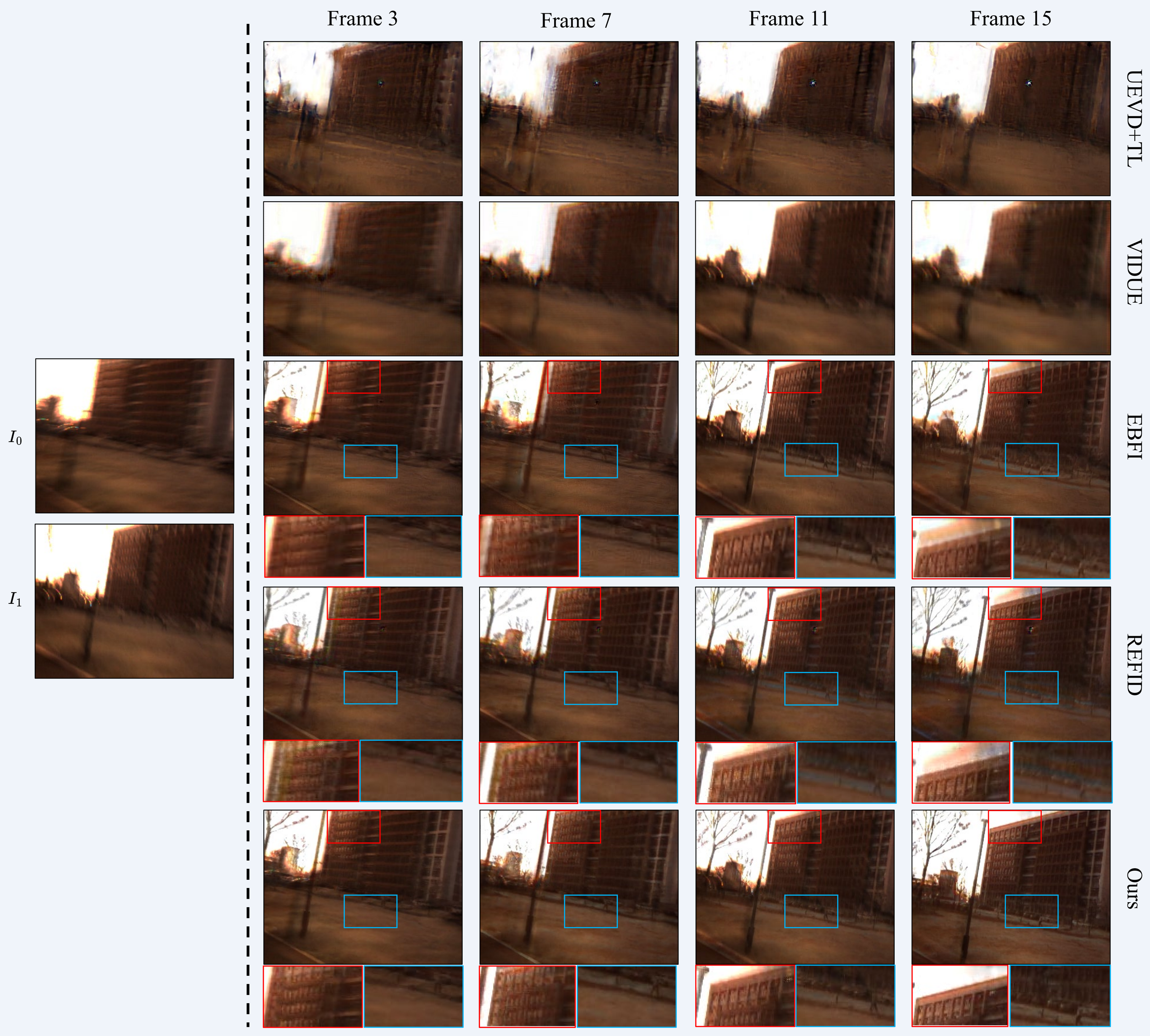
Quantitative comparison

- Synthetic datasets

Method	PSNR ↑ / SSIM ↑ / LPIPS ↓			
GoPro-10⇓	**9+1**	**5+5**	**1+9**	**RandEx**
NAFnet [4]+RIFE [11]	23.15 / 0.736 / 0.164	21.82 / 0.665 / 0.188	18.78 / 0.520 / 0.282	21.05 / 0.633 / 0.219
UTI [49]	23.84 / 0.805 / 0.205	23.25 / 0.771 / 0.166	20.95 / 0.640 /0.170	22.17 / 0.714 / 0.189
VIDUE [34]	25.37 / 0.818 / 0.181	26.24 / 0.835 / 0.145	24.76 / 0.764 / 0.146	25.86 / 0.819 / 0.152
UEVD [16]+TL [41]	23.40 / 0.720 / 0.194	23.31 / 0.755 / 0.161	26.33 / 0.866 / 0.103	23.51 / 0.757 / 0.164
EVDI [47]	28.18 / 0.898 / <u>0.052</u>	27.73 / 0.884 / <u>0.067</u>	26.07 / 0.834 / 0.084	27.51 / 0.877 / <u>0.071</u>
REFID [40]	<u>31.08</u> / 0.939 / 0.077	<u>31.28</u> / <u>0.941</u> / 0.071	30.27 / <u>0.927</u> / <u>0.078</u>	<u>31.03</u> / <u>0.938</u> / 0.074
EBFI [44]	31.06 / <u>0.942</u> / 0.072	31.08 / 0.940 / 0.071	<u>30.32</u> / 0.921 / 0.087	30.89 / 0.936 / 0.075
Ours	**33.22 / 0.960 / 0.050**	**33.61 / 0.963 / 0.042**	**32.87 / 0.954 / 0.048**	**33.39 / 0.961 / 0.045**
HighREV-10⇓				
NAFnet [4]+RIFE [11]	25.09 / 0.805 / 0.446	24.90 / 0.874 / 0.401	26.07 / 0.897 / 0.361	25.20 / 0.844 / 0.402
UTI [49]	26.22 / 0.834 / 0.371	26.15 / 0.821 / 0.395	26.51 / 0.851 / 0.395	26.19 / 0.831 / 0.398
VIDUE [34]	26.65 / 0.847 / 0.375	27.43 / 0.860 / 0.367	25.31 / 0.841 / 0.387	26.76 / 0.850 / 0.373
UEVD [16]+TL [41]	26.02 /0.844 / 0.399	27.41 / 0.854 / 0.378	27.77 / 0.861 / 0.337	27.43 / 0.854 / 0.373
EVDI [47]	30.26 / 0.910 / <u>0.278</u>	29.75 / 0.896 / <u>0.277</u>	27.01 / 0.839 / 0.292	29.25 / 0.887 / <u>0.273</u>
REFID [40]	<u>33.60</u> / <u>0.937</u> / 0.306	<u>34.18</u> / <u>0.938</u> / 0.296	<u>33.38</u> / <u>0.928</u> / <u>0.260</u>	<u>33.91</u> / <u>0.936</u> / 0.292
EBFI [44]	28.23 / 0.898 / 0.362	27.84 / 0.887 / 0.342	26.25 / 0.855 / 0.279	27.55 / 0.882 / 0.333
Ours	**35.73 / 0.948 / 0.263**	**36.27 / 0.949 / 0.250**	**35.45 / 0.942 / 0.248**	**36.02 / 0.947 / 0.253**

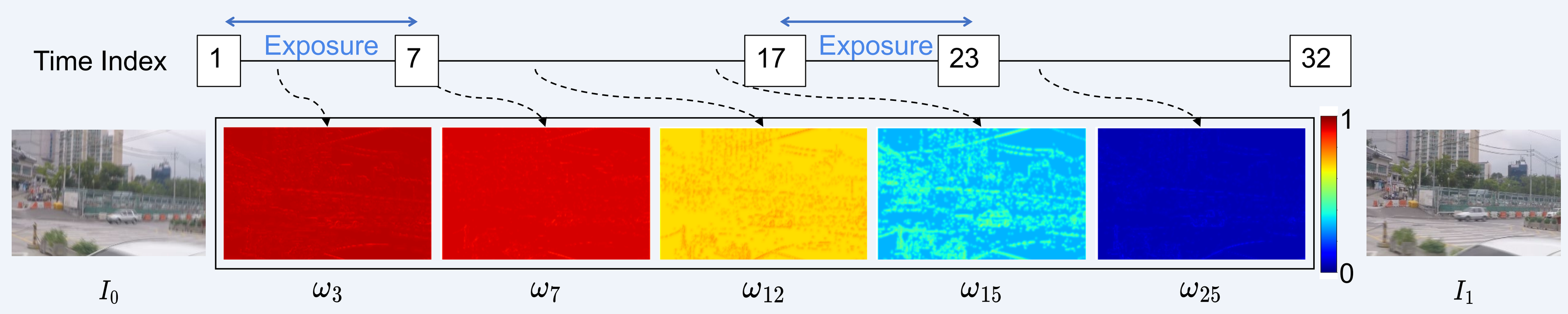
Qualitative comparison

- Real-world dataset



Analysis

- As the target timestamp τ moves closer to either exposure window, the importance map ω_τ shifts to assign greater weight to the corresponding frames' features
- This demonstrates that our framework adaptively emphasizes temporally closer and spatially relevant information, ensuring stable interpolation across varying τ



Takeaways

- We propose a learnable framework that effectively incorporates temporal constrains for event-guided VFI under blind exposure
- Experiments show that temporal relevance is crucial for robust interpolation in real-world settings