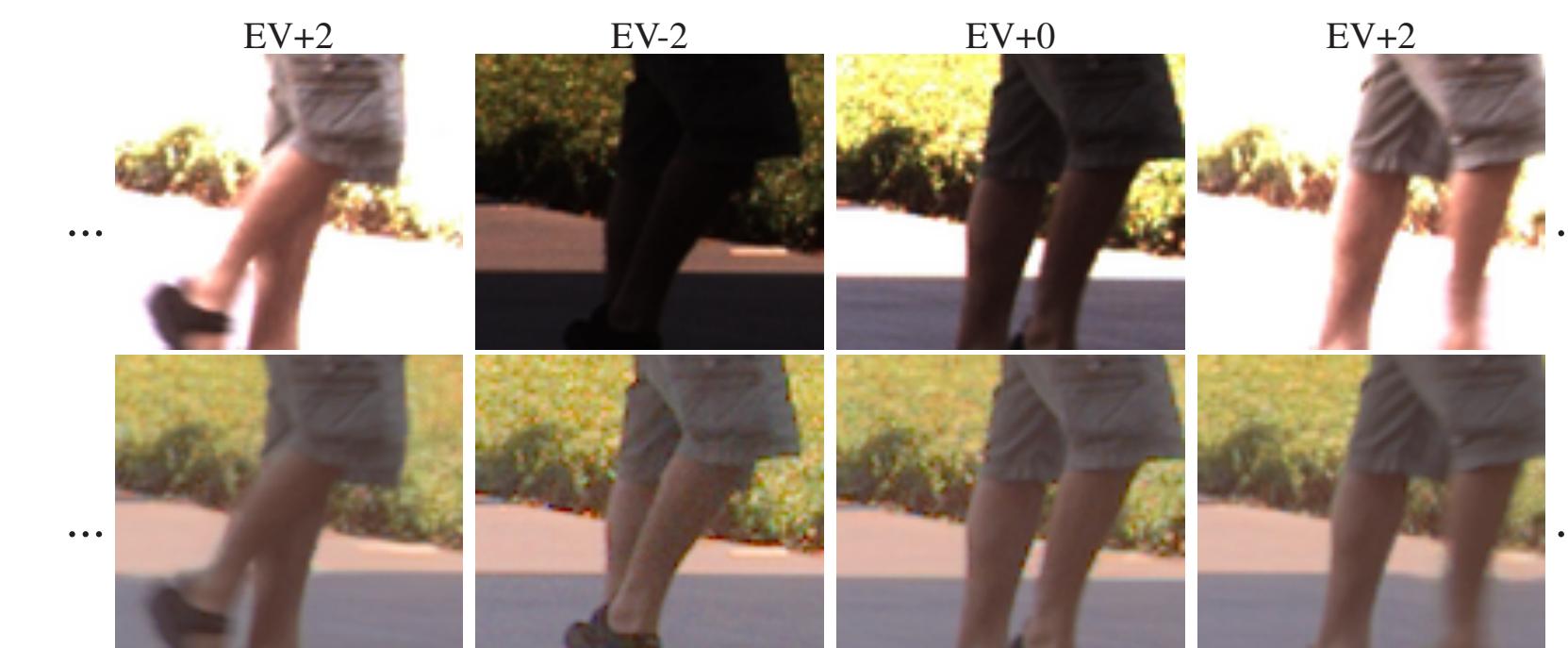




Problem and Contribution

Goal: Introducing a coarse-to-fine framework and a real-world video dataset for HDR video reconstruction from sequences with alternating exposures.



Contributions:

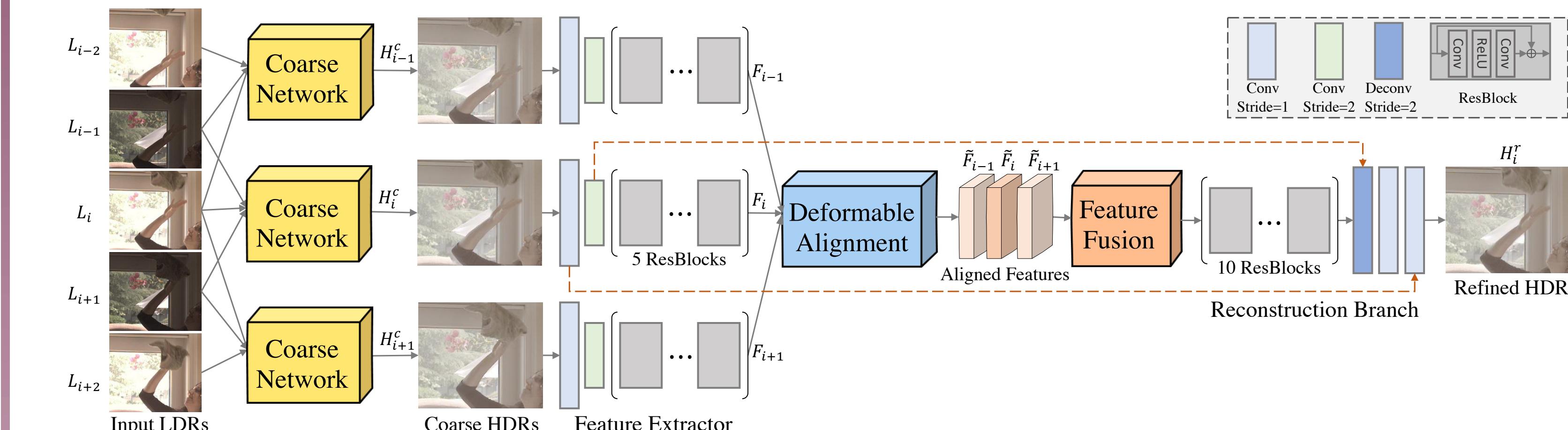
- We propose a two-stage framework, which first performs image alignment and HDR fusion in the image space and then in feature space, for HDR video reconstruction from sequences with alternating exposures.
- We create a real-world video dataset captured with alternating exposures as a benchmark to enable quantitative evaluation for this problem.

The Proposed Coarse-to-fine Network

Input Preprocessing

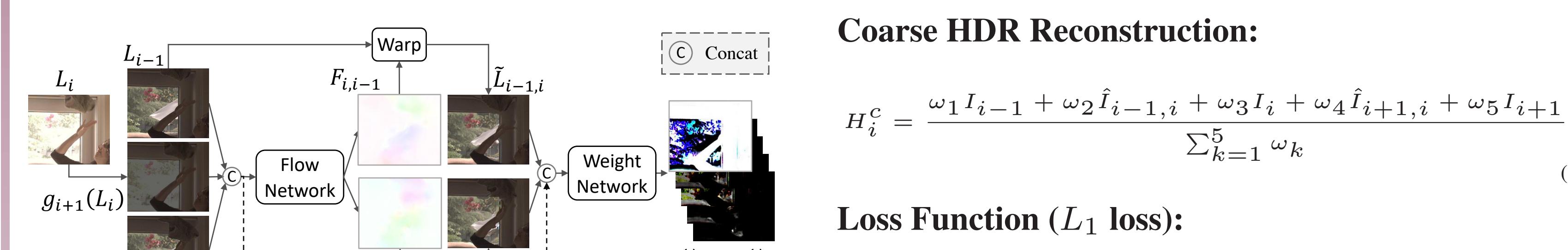
- We replace the CRF of the input images with a fixed gamma curve as $L_i = (\mathcal{F}^{-1}(\tilde{L}_i))^{1/\gamma}$, where $\gamma = 2.2$
- Global alignment is then performed using a similarity transformation to compensate camera motions among neighboring frames

Network Architecture for Two Alternating Exposures:



CoarseNet Performs Alignment and Fusion in Image Space:

- CoarseNet warps neighboring frames to the center frame using optical flows, and reconstructs the HDR image by pixel blending



Coarse HDR Reconstruction:

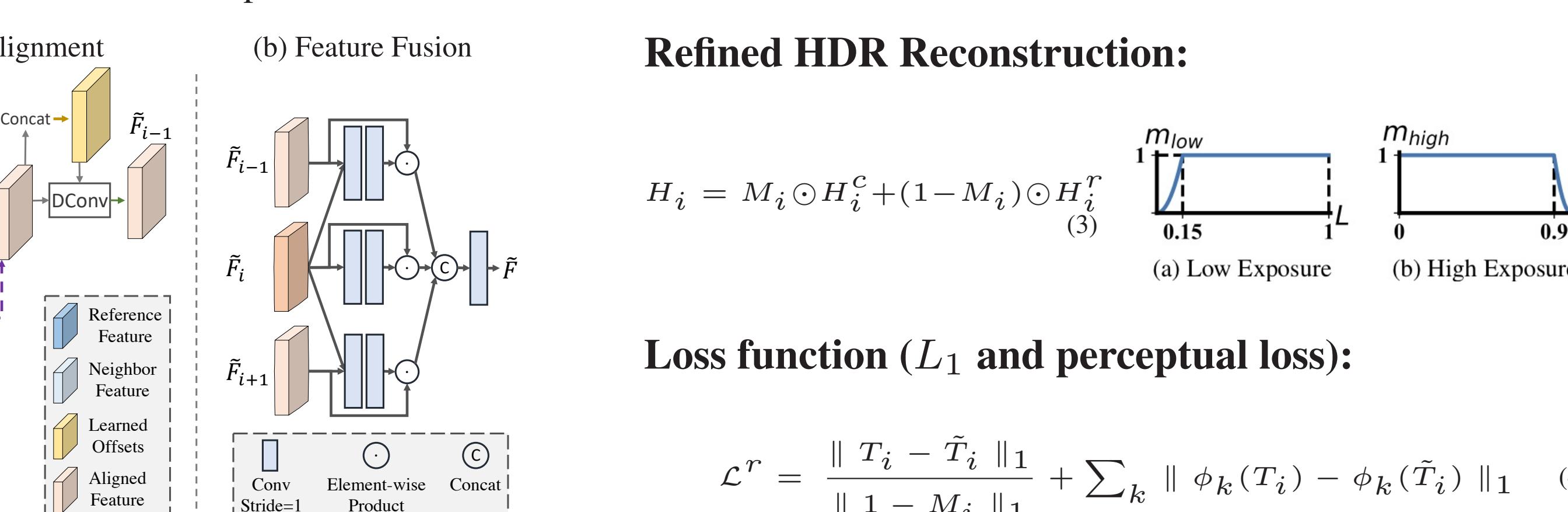
$$H_i^C = \frac{\omega_1 I_{i-1} + \omega_2 \hat{I}_{i-1,i} + \omega_3 I_i + \omega_4 \hat{I}_{i+1,i} + \omega_5 I_{i+1}}{\sum_{k=1}^5 \omega_k}.$$

Loss Function (L_1 loss):

$$\mathcal{L}^C = \|T_i^C - \bar{T}_i\|_1, \quad T_i^C = \frac{\log(1 + \mu H_i^C)}{\log(1 + \mu)}$$

RefineNet Performs Alignment and Fusion in Feature Space:

- Features of neighboring frames are aligned to the center frame using deformable alignment module
- The aligned feature is fused with a temporal attention for HDR reconstruction



Refined HDR Reconstruction:

$$H_i = M_i \odot H_i^C + (1 - M_i) \odot H_i^T$$

Loss function (L_1 and perceptual loss):

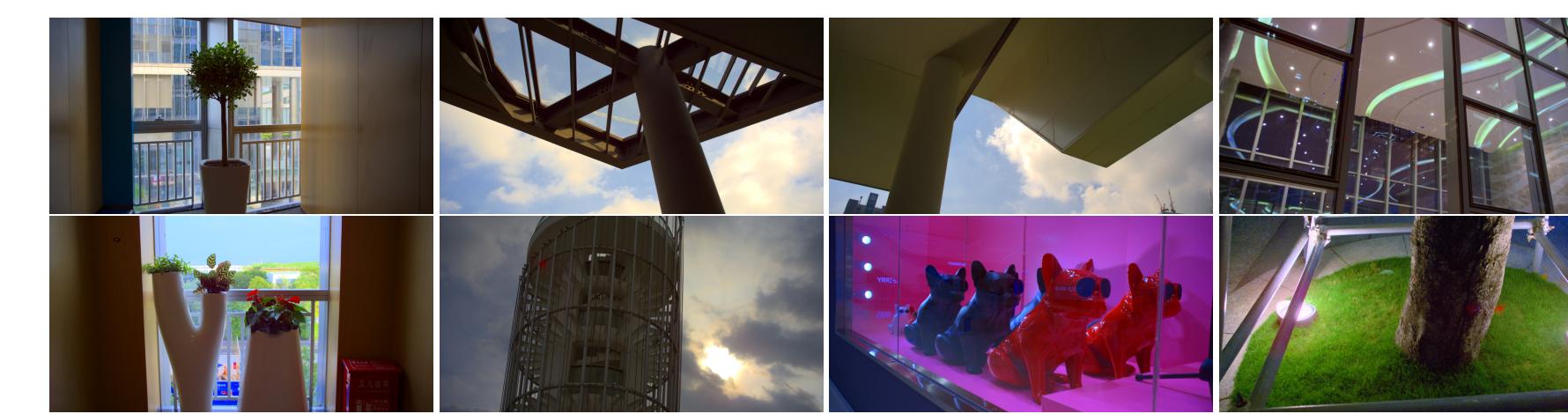
$$\mathcal{L}^r = \frac{\|T_i - \bar{T}_i\|_1}{\|1 - M_i\|_1} + \sum_k \|\phi_k(T_i) - \phi_k(\bar{T}_i)\|_1$$

The Introduced Real-world Video Dataset

Dataset Statistics

Data	Size	Static Scenes w/ GT		Dynamic Scenes w/ GT		Dynamic Scenes w/o GT	
		2-Exp	3-Exp	2-Exp	3-Exp	2-Exp	3-Exp
Kalantari13	1280 × 720	-	-	-	-	5	4
Ours	4096 × 2168	49	48	76	108	50	50

Samples in Static Scenes With GT



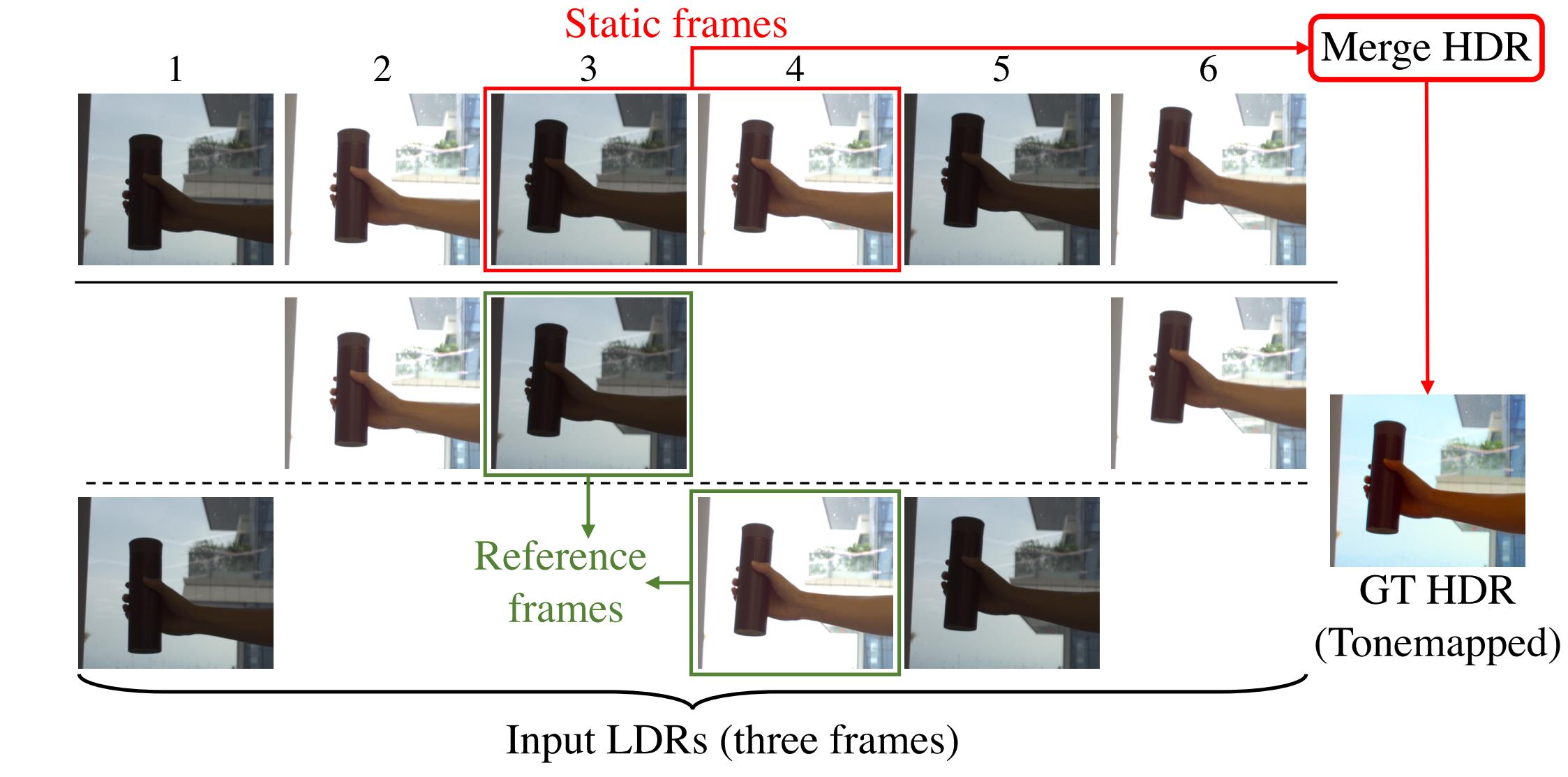
Samples in Dynamic Scenes Without GT



Generating the LDRs-HDR Pairs for Dynamic Scenes with GT

Row 1: the selected image sequence.

Rows 2 and 3: two sample pairs with low-exposure and high-exposure reference frames, respectively.



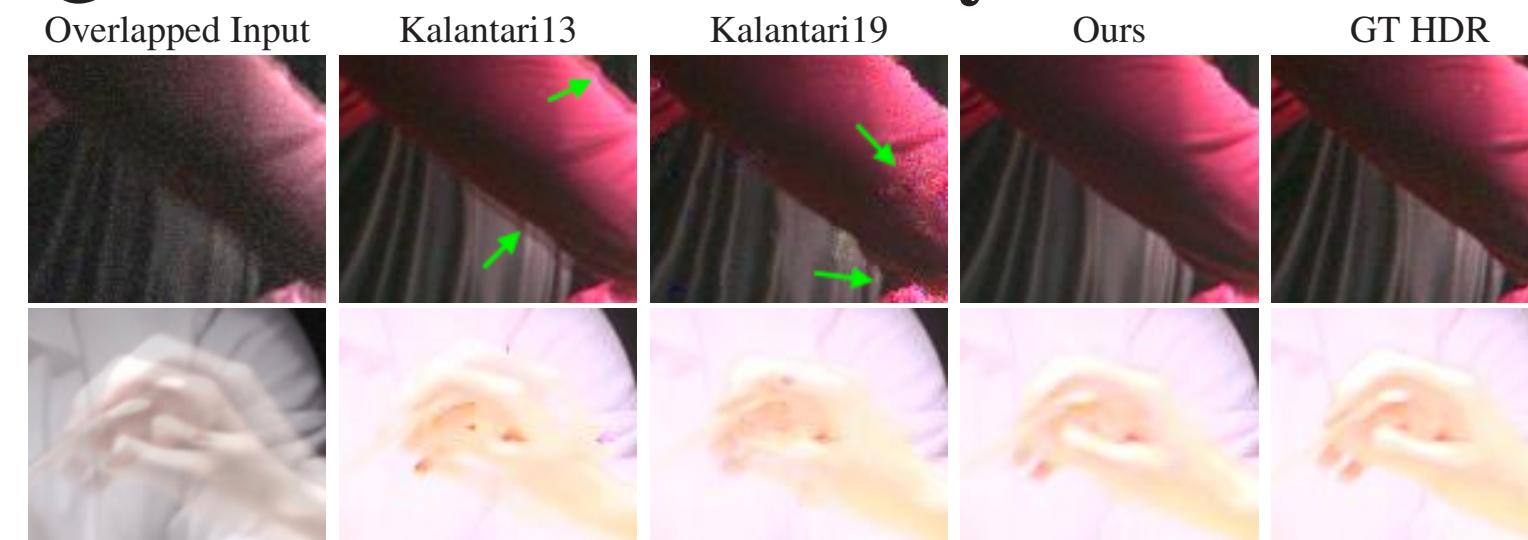
- The subject is asked to keep still for 2s, an HDR image is generated for this static frame
- The subject is asked to move back-and-forth (e.g., waving hands or walking)
- Select a sequence whose center frame was the static frame, and arrange it to be proper pairs

Experiments & Results

Quantitative Results on Synthetic Data

Method	2-Exposure			3-Exposure		
	PSNR	HDR-VDP2	HDR-VQM	PSNR	HDR-VDP2	HDR-VQM
Kalantari13	37.53	59.07	84.51	30.36	56.56	65.90
Yan19	39.05	70.61	71.27	36.28	65.47	72.20
Kalantari19	37.48	70.67	84.57	36.27	65.51	72.58
Ours	40.34	71.79	85.71	37.04	66.44	73.38

Qualitative Results on Synthetic Data



Quantitative Results on the Introduced Real-world Dataset

(a) Results on static scenes with GT (\mathcal{D}_s^{gt}) augmented with random global motion.

Method	2-Exposure			3-Exposure		
	Low-Exposure PSNR	High-Exposure HDR-VDP2	PSNR	Low-Exposure PSNR	Middle-Exposure HDR-VDP2	PSNR
Kalantari13	40.00	73.70	40.04	70.08	40.02	71.89
Yan19	34.54	80.22	39.25	65.96	36.90	73.09
Kalantari19	39.79	81.02	39.96	67.25	39.88	74.13
Ours	41.95	81.03	40.41	71.27	41.18	76.15
					39.67	73.24
					40.01	67.90
					39.77	70.37
					39.75	79.55

(b) Results on dynamic scenes with GT (\mathcal{D}_d^{gt}).

Method	2-Exposure			3-Exposure		
	Low-Exposure PSNR	High-Exposure HDR-VDP2	PSNR	Low-Exposure PSNR	Middle-Exposure HDR-VDP2	PSNR
Kalantari13	37.73	74.05	45.71	66.67	41.72	70.36
Yan19	36.41	85.68	49.89	69.90	43.15	77.79
Kalantari19	39.94	86.77	49.49	69.04	44.72	77.91
Ours	40.83	86.84	50.10	71.33	45.46	79.09
					38.70	78.44
					38.77	78.11
					41.47	68.49
					43.24	65.08
					41.16	70.56
					39.75	73.92
						82.87

Visual Comparison on THROWING TOWEL Scene from Kalantari13 Dataset



Model Parameter and Runtime:

Method	# Parameter	2-Exposure		3-Exposure	
		1280 × 720	1920 × 1080	1280 × 720	1920 × 1080
Kalantari13	-	125s	185s	300s	520s
Kalantari19	9.0M	0.35s	0.59s	0.42	0.64
Ours	6.1M	0.51s	0.97s	0.64</	