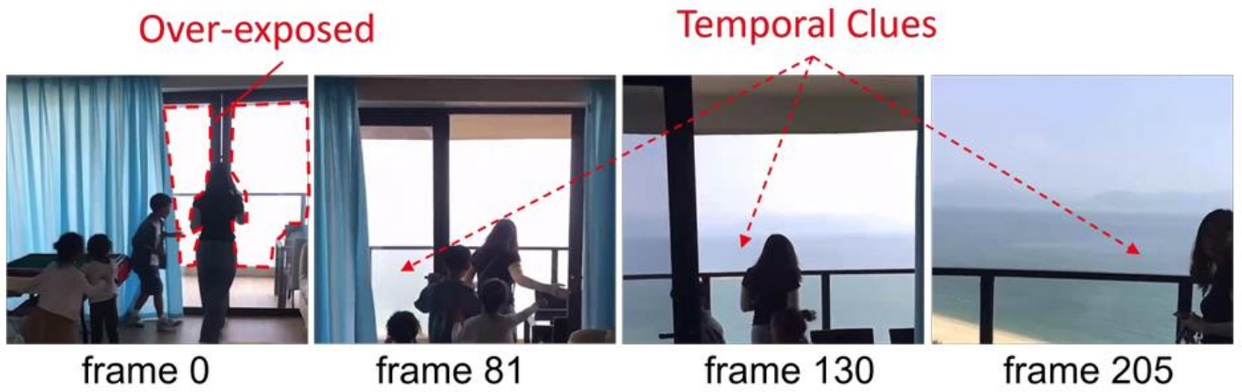
Deep Video Inverse Tone Mapping Based on Temporal Clues

Yuyao Ye[1],Ning Zhang[2], Yang Zhao[3], Hongbin Cao[1,4], Ronggang Wang[1]
[1]School of Electronic and Computer Engineering, Peking University [2]Baidu Netdisk
[3]School of Computer and Information, Hefei University of Technology [4]Bytedance Inc.

Background



LDR video shooting process:

1. Camera sets a proper exposure value based on lighting conditions and the object. -> exposure changes -> **temporal clues**
2. To prevent flickering, exposure changes are usually slow and smooth. -> **temporal clues far from over-exposed frames**

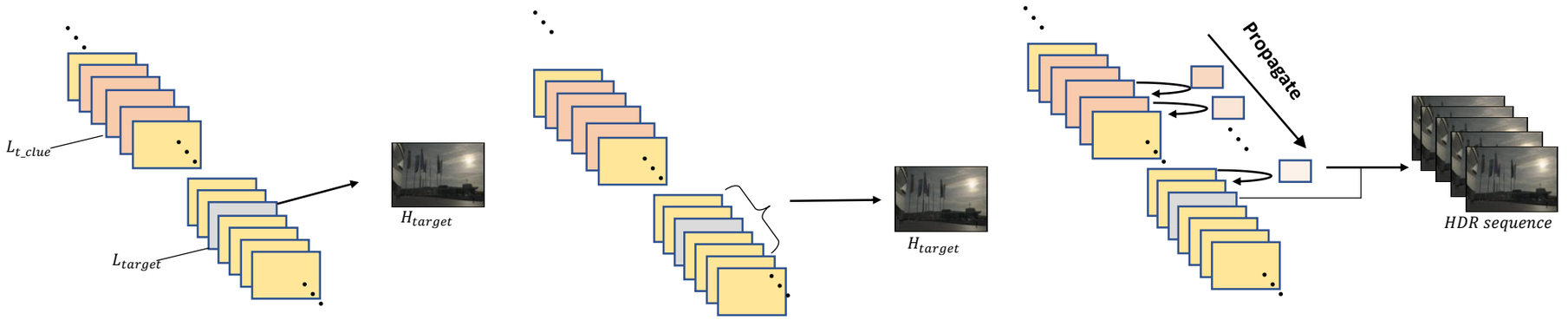
Temporal Clues:

Same content as over-exposed area, but normally exposed in other frames

Motivation:

Leverage temporal clues to efficiently convert real world LDR videos to HDR videos with fidelity and temporal consistency.

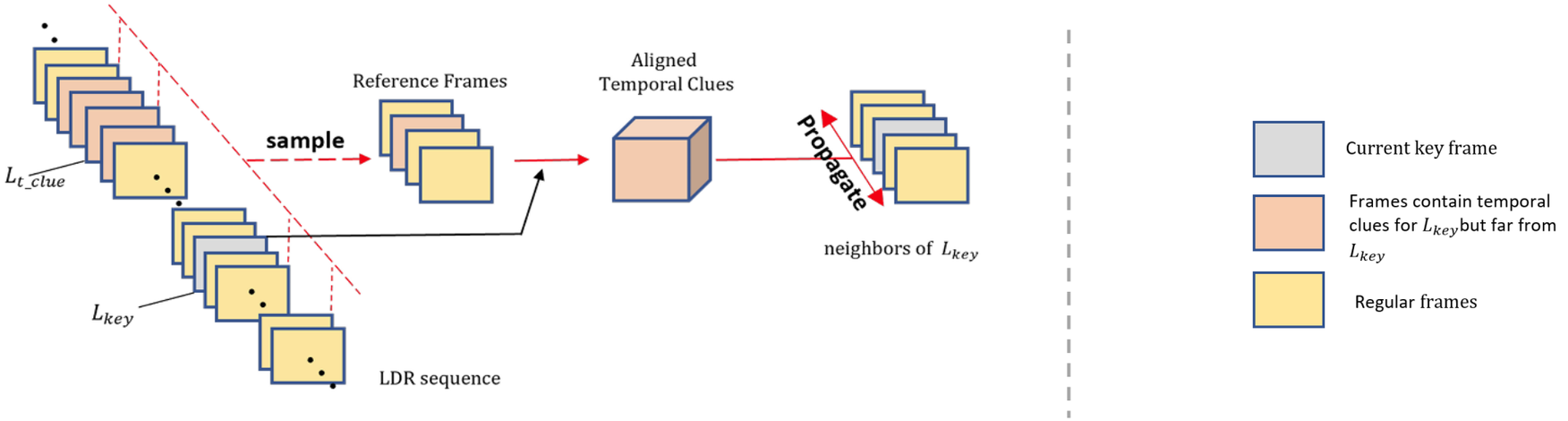
Challenges



(a) Image ITM (b) Sliding window based (c) propagation based

Existing temporal models failed(repeating image ITM, sliding window based, RNN based):
temporal inconsistency, window can’t reach temporal clues, temporal clues fade when propagation, high computational complexity.

Temporal Model



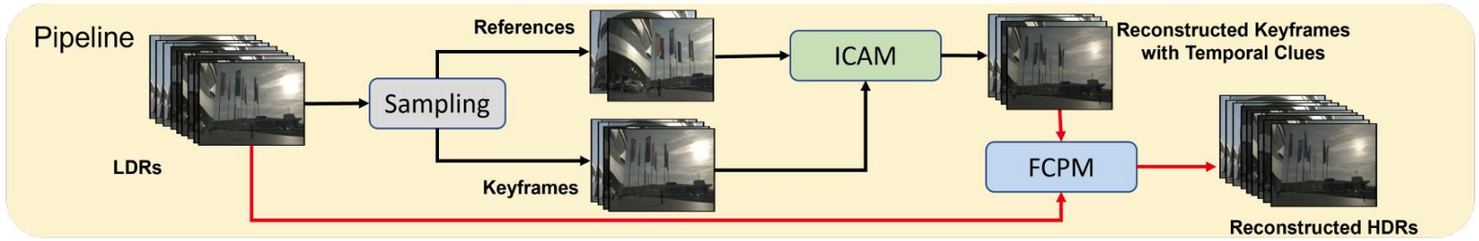
Global sampling and local propagation strategy.

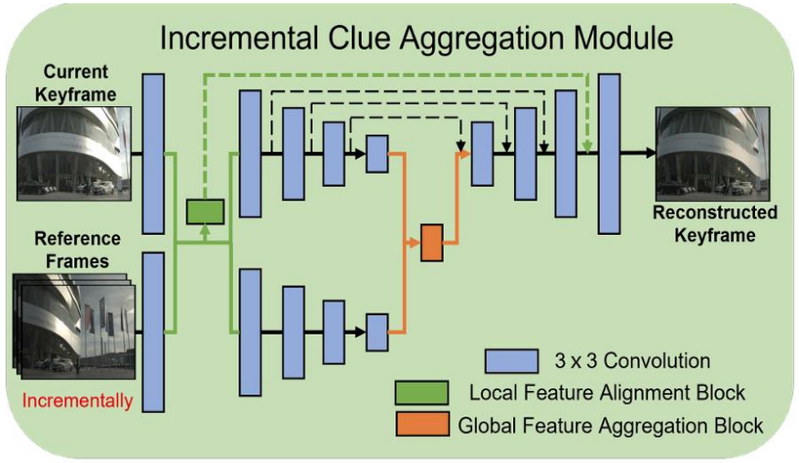
1. globally sample some reference frames with a large stride to reconstruct current key frame.

because exposure changes are slow, it is very likely that sampled reference frames contain valid temporal clues(orange frames in this figure).

2. propagate information in reconstructed key frames to their neighbors.

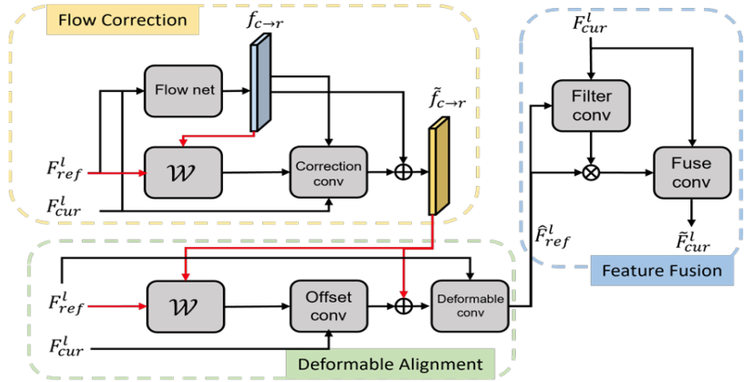
Two-stage video ITM pipeline





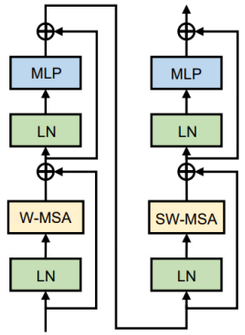
ICAM

- Reconstruct keyframes with references



Local Feature Alignment Block

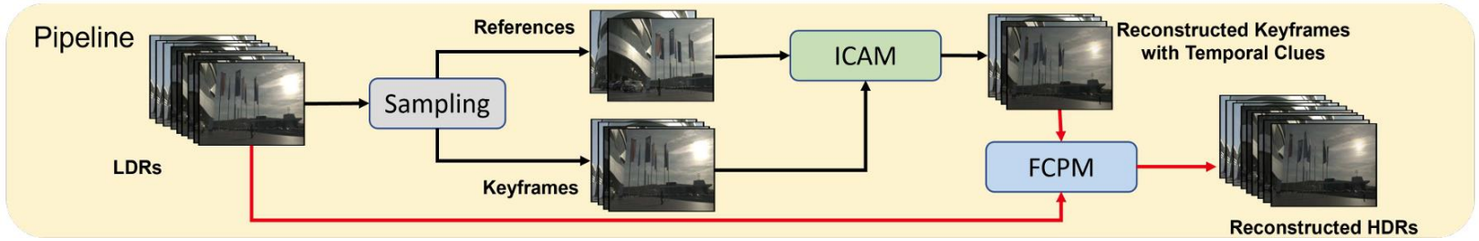
- alignment of overexposed and normal exposed areas
- large motion alignment due to global sampling
- efficient feature fusion

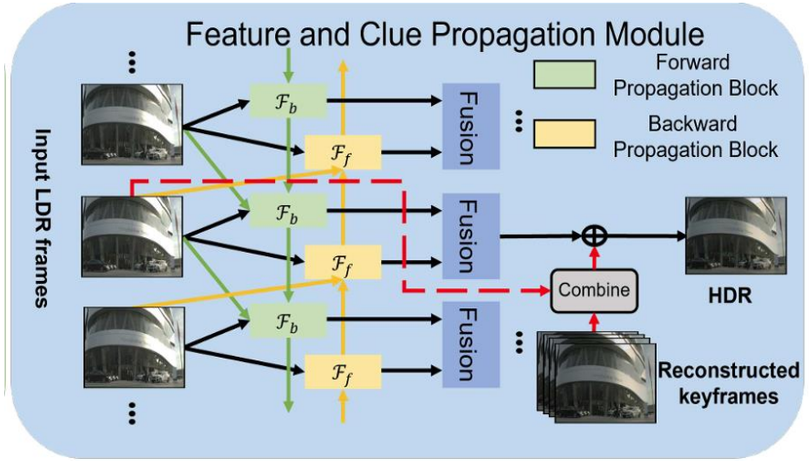


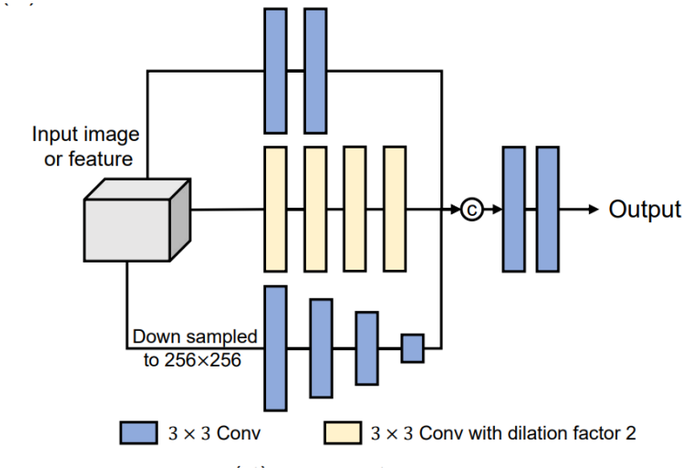
Global Feature Aggregation Block

- adjust global exposure

Two-stage video ITM pipeline



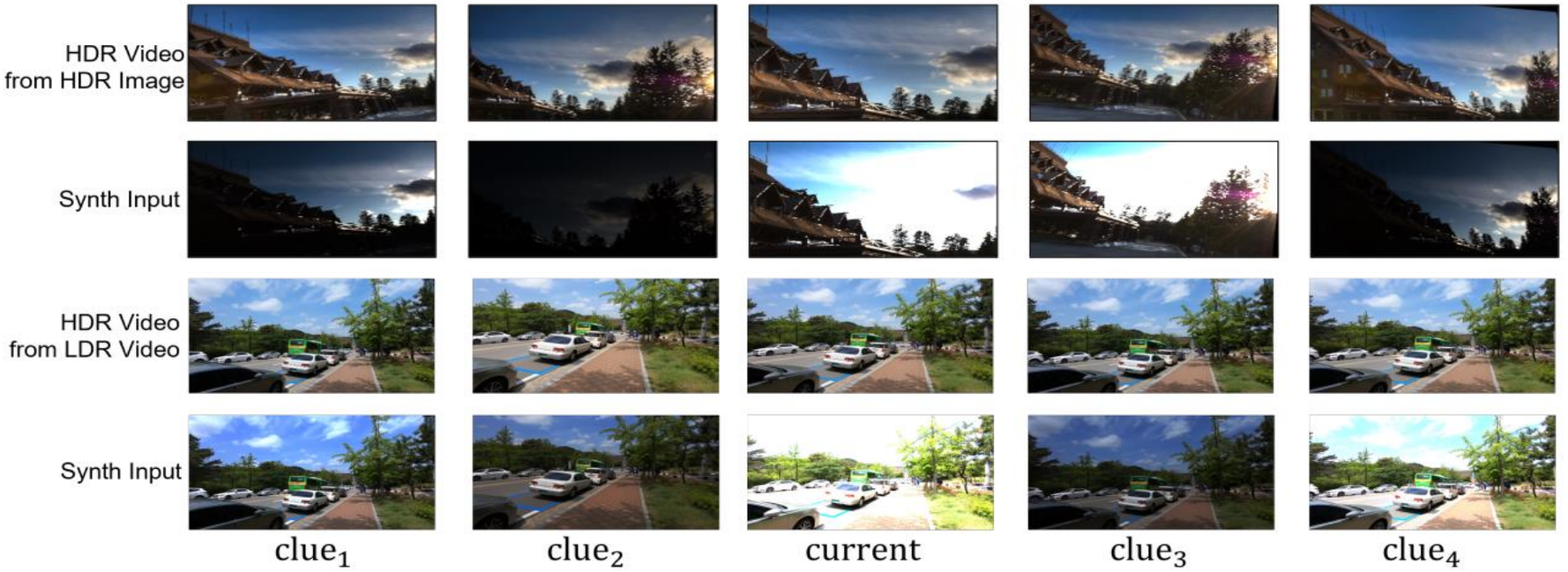




FCPM

- propagate the information in the reconstructed key frames to their local neighbor frames -> **efficiency**
- Regard each frame in several adjacent groups of pictures(taken care by several reconstructed keyframes) **-> temporal consistency**

Synthetic dataset



Low cost, diverse training dataset.

High quality LDR video datasets —Inverse tone mapping→

HDR image datasets —Random warp→

HDR video datasets —degrade→ LDR-HDR video datasets

Experiments and Results

Visual comparisons on the frame of “fishing longshot” sequence from HDM-HDRv (above) and “sb-tunnel-exr” sequence from MPI-HDRv(below).

Visual comparison on the frame of a real-world LDR video shot by iPhone 13.

Experiments and Results

Quantitative comparison on HDR videos with existing methods. The scores here are HDR-VDP-3/HDR-VQM, where a higher score of HDR-VDP-3 and a lower score of HDR-VQM mean better. Red text indicates the best and blue text indicates the second best result, respectively

	REDS-val [27]	HDM-HDR [9]	LiU-HDR [17]	MPI-HDRv [10]
HDRCNN [7]	6.774/0.452	5.822/0.467	7.884/0.545	6.733/0.053
Diff HDRI [14]	6.820/0.502	5.703/0.514	7.751/0.594	6.501/0.061
Single HDRI [22]	7.059/0.492	6.346/0.481	8.143/0.569	7.225/0.058
FMNet [36]	6.895/0.483	5.847/0.496	7.974/0.572	6.847/0.057
Deep VITM [37]	7.106/0.458	5.992/0.445	8.036/0.553	7.104/0.049
Bascivsr++ [3]	7.254/0.443	6.131/0.427	8.265/0.547	7.119/0.052
Proposed	**7.891**/**0.398**	**6.754**/**0.356**	**8.591**/**0.522**	**7.970**/**0.037**

Summary

- we analyze the temporal clues in LDR videos.
- A new temporal model with the Global Sampling and Local Propagation strategy and a two-stage video ITM pipeline
- Reconstruct over-exposed areas with fidelity, high efficiency, temporal consistency.
- Modules to align and fuse frames with exposure changes and large motions under ITM task. Modules to propagate the reconstructed key frames to their neighbors and generate temporally consistent results.
- A novel dataset synthesis method to obtain HDR video training dataset only using available HDR images and LDR video dataset.
- Outperform the SOTA methods both on quantitative and visual quality.