



# Video Frame Synthesis using Deep Voxel Flow

Ziwei Liu<sup>1</sup>      Raymond A. Yeh<sup>2</sup>      Xiaou Tang<sup>1</sup>  
Yiming Liu<sup>3</sup>      Aseem Agarwala<sup>4</sup>

1. *The Chinese University of Hong Kong*
2. *University of Illinois at Urbana-Champaign*
3. *Pony AI*
4. *Google Inc.*

# Problem



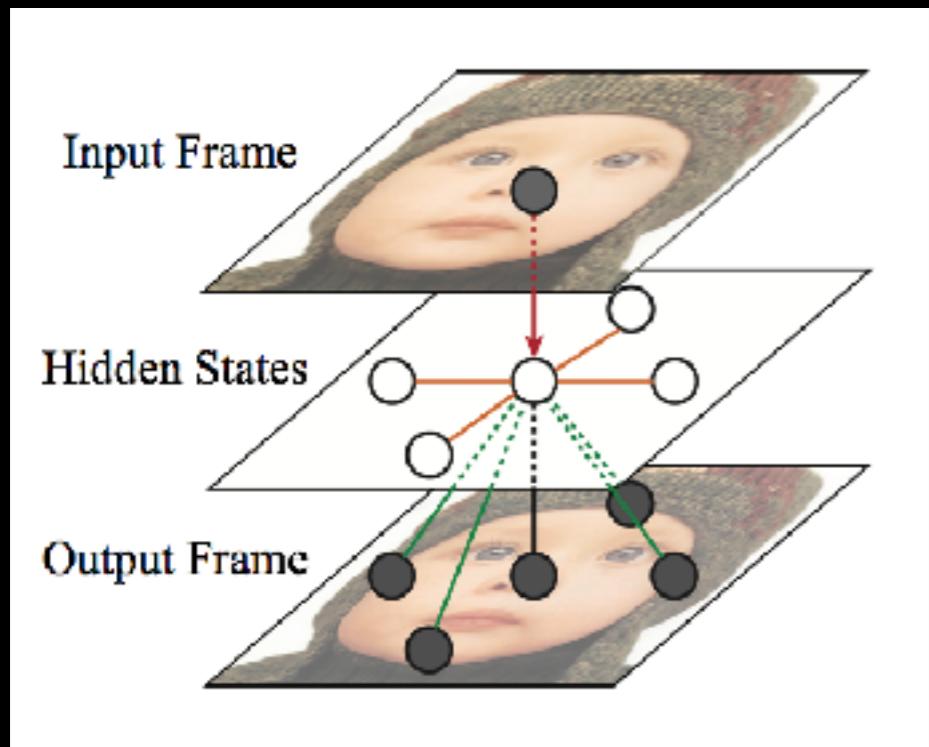
Video interpolation/extrapolation

# Challenges



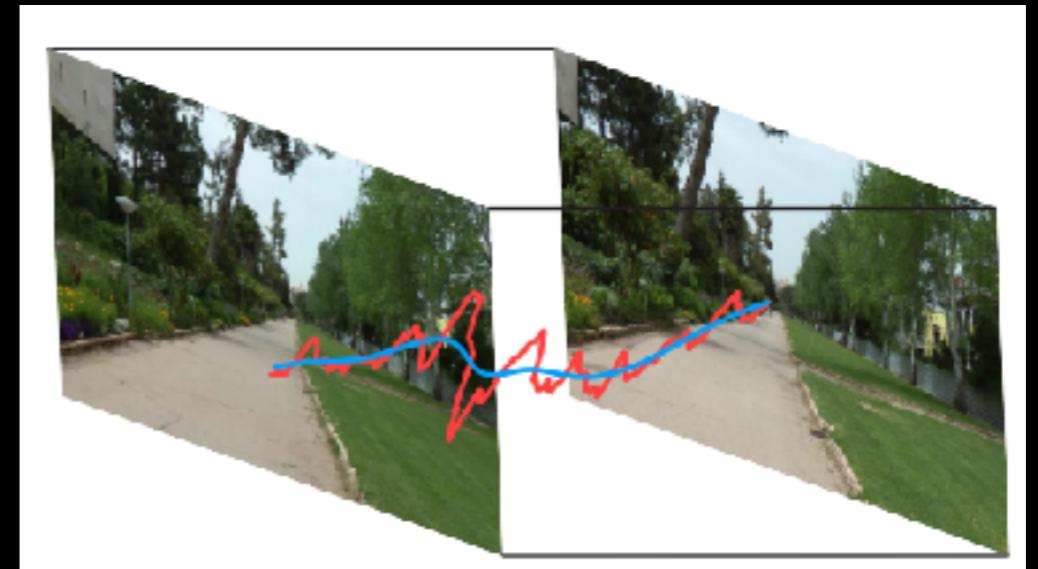
1. Complex motion (camera motion & scene motion)
2. High-res images (1280 \* 720)

# Previous Attempts



Hallucination-based Methods

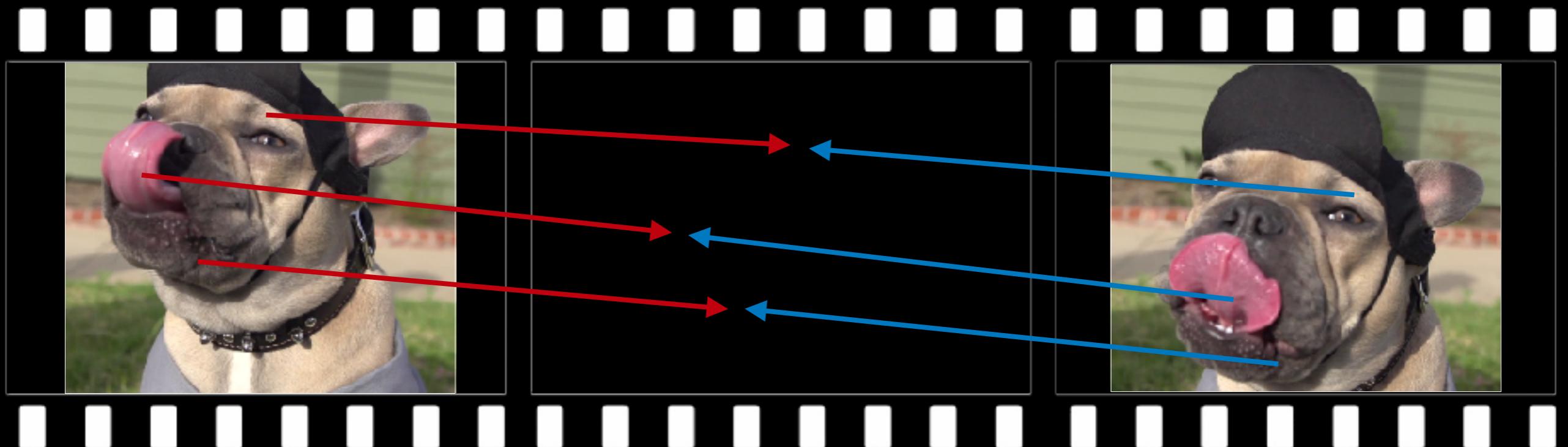
[Ranzato et al.], [Srivastava et al.], [Mathieu et al.]



Flow-based Methods

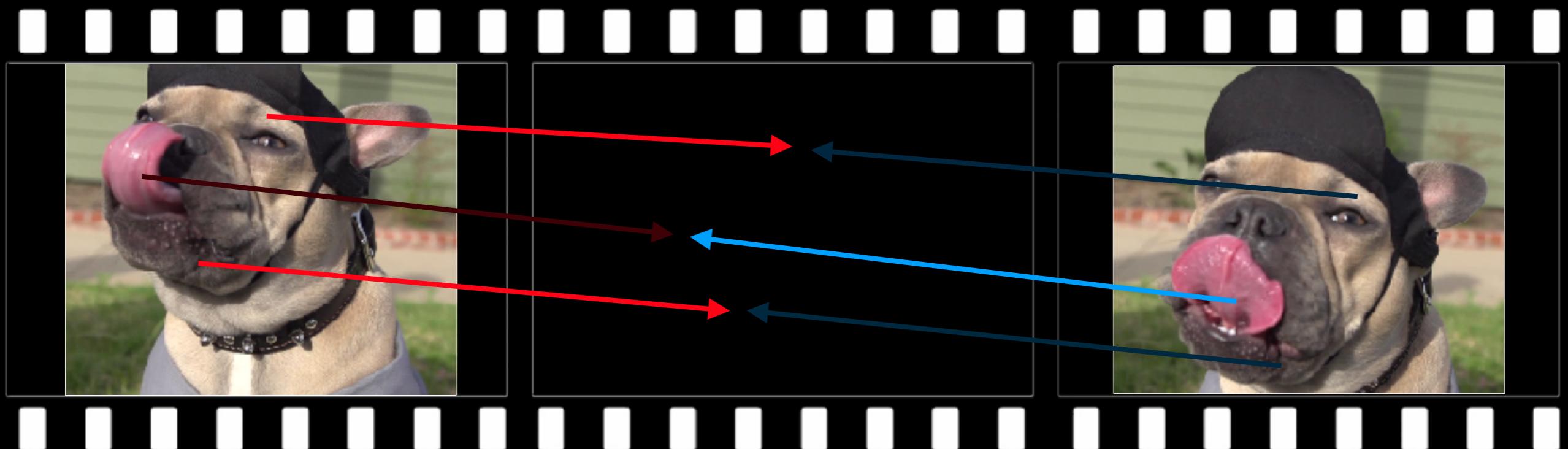
[Baker et al.], [Mahajan et al.], [Walker et al.]

# Voxel Flow



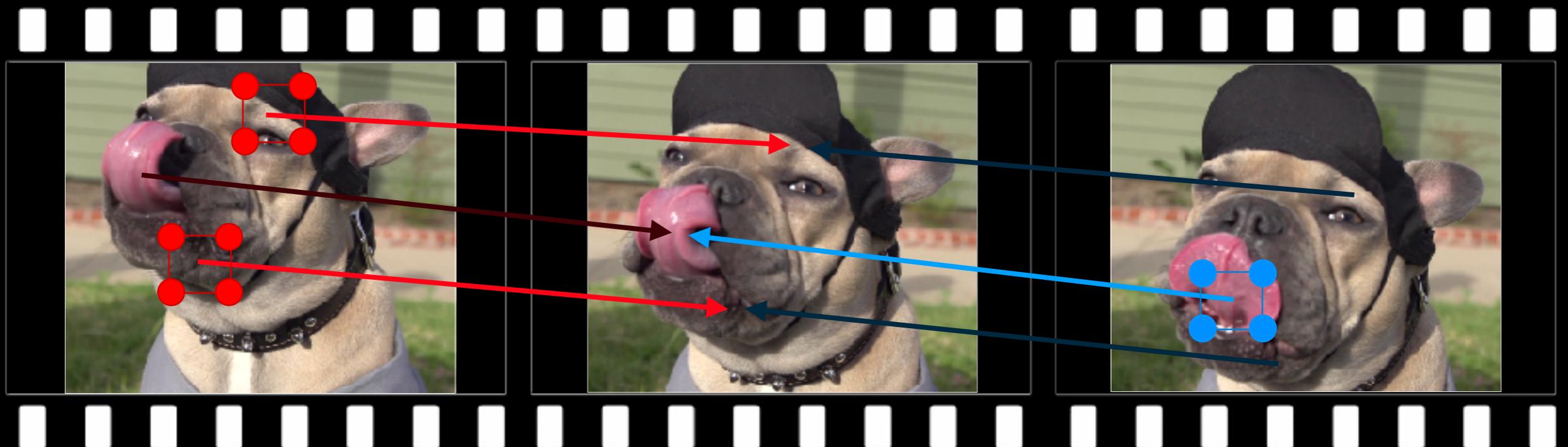
symmetric bi-directional flows

# Voxel Flow



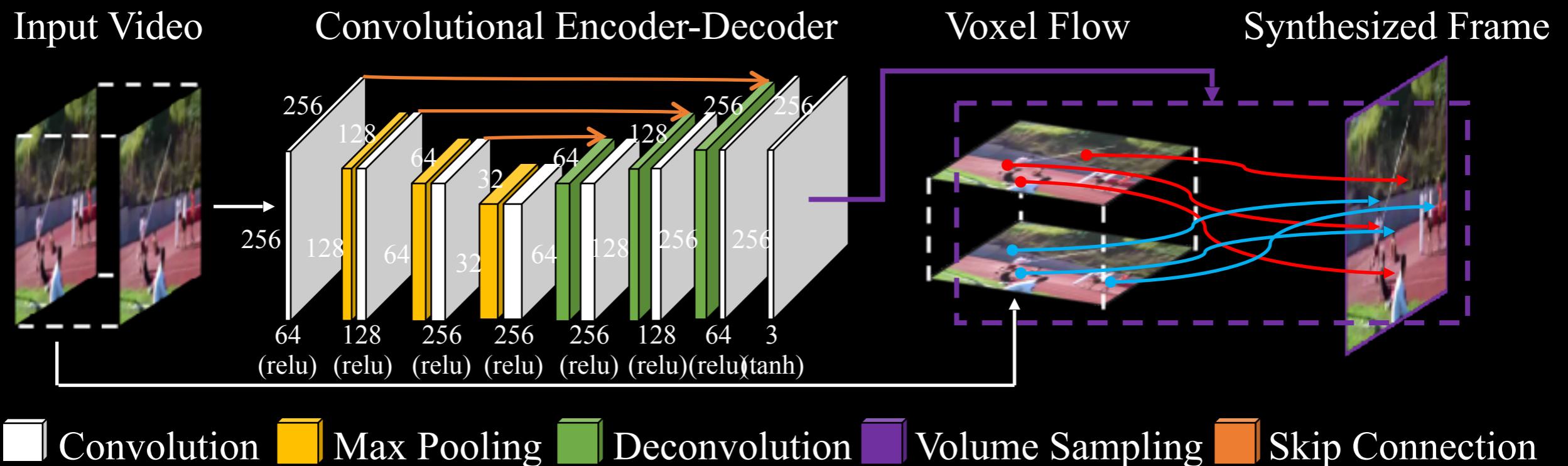
selection mask between frames

# Voxel Flow

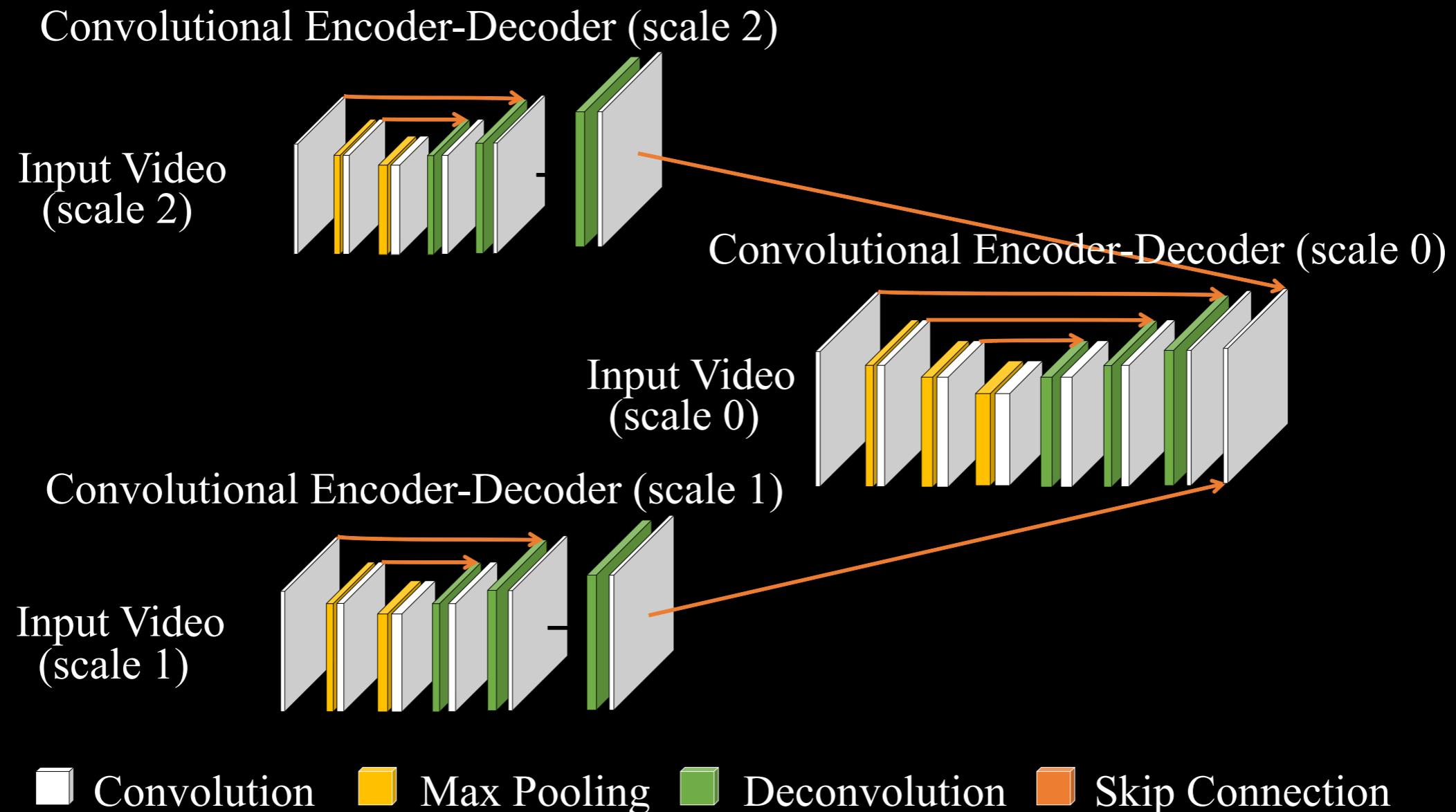


differentiable bilinear sampling

# Deep Voxel Flow



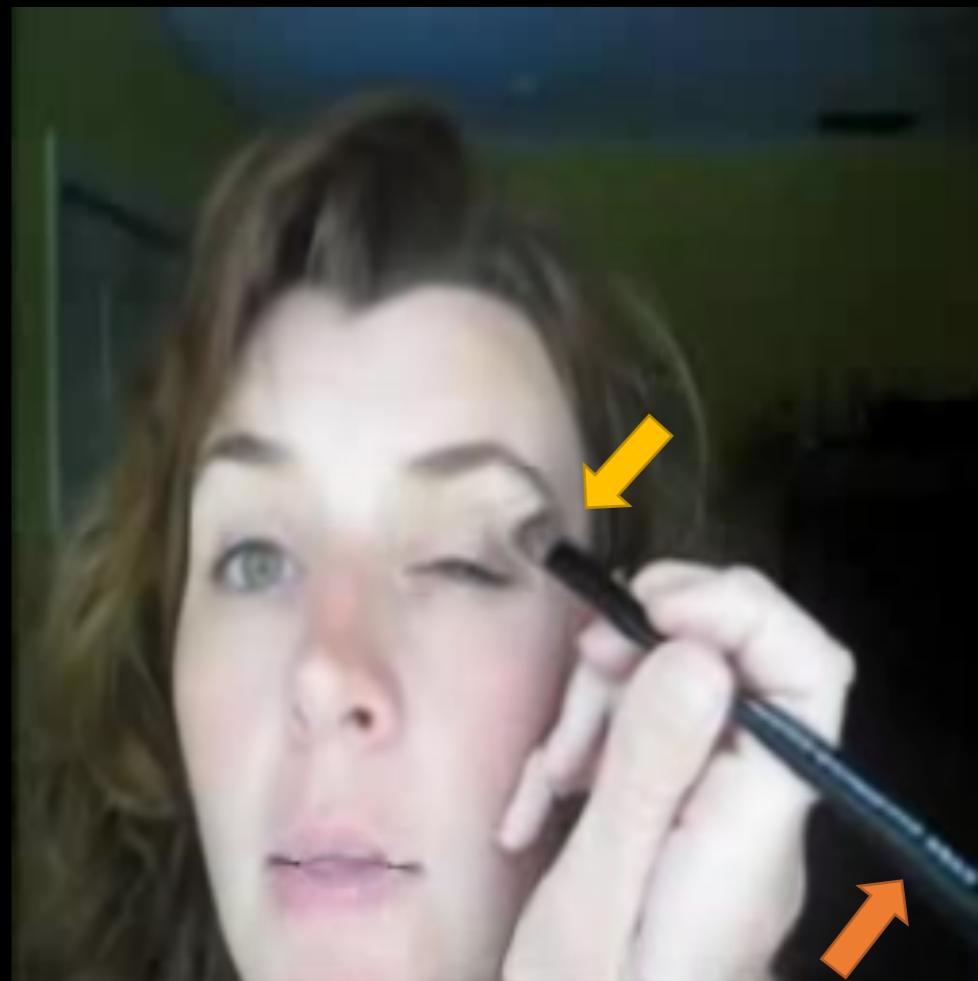
# Multi-scale Deep Voxel Flow



# Multi-scale Deep Voxel Flow



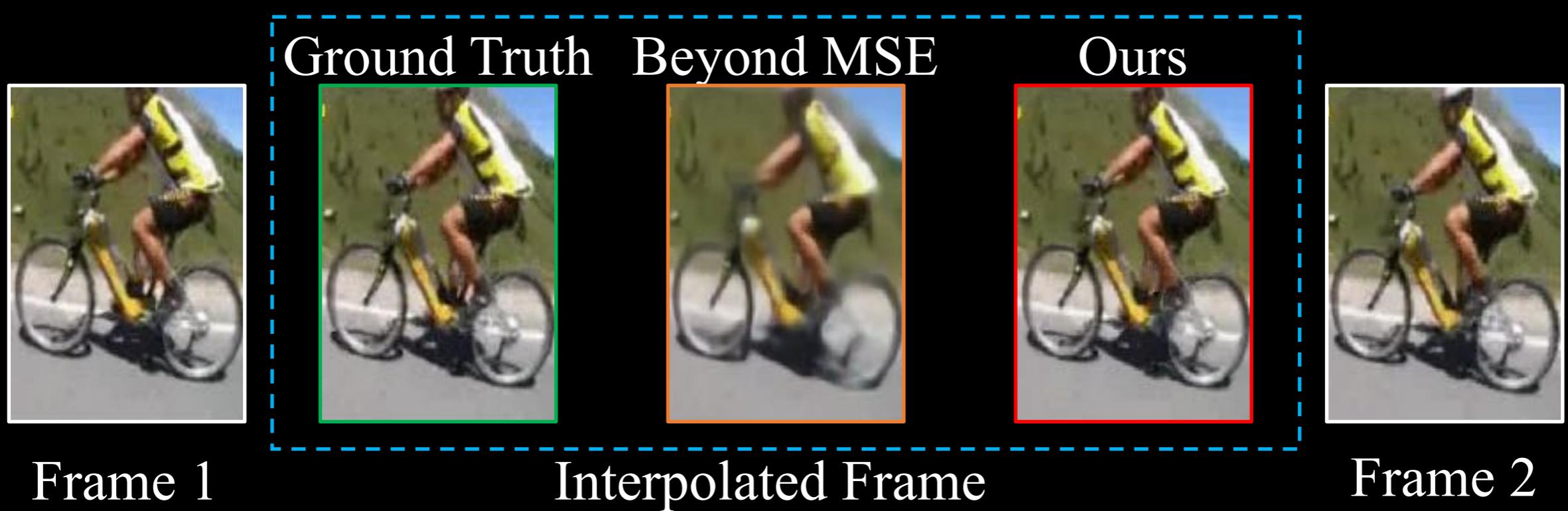
(a) Voxel Flow



(b) Multi-scale Voxel Flow

# Comparisons

- UCF-101



# Comparisons

- UCF-101



# Comparisons

- UCF-101



# Comparisons

- KITTI



Ground Truth

Appearance Flow

Ours

# Comparisons

- KITTI



# Comparisons

- KITTI



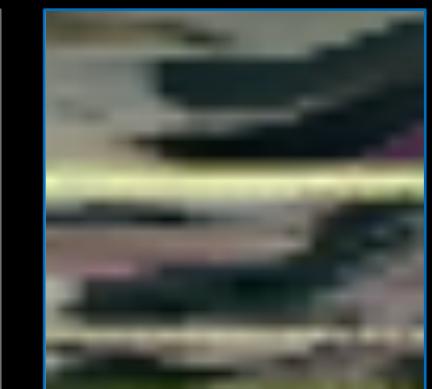
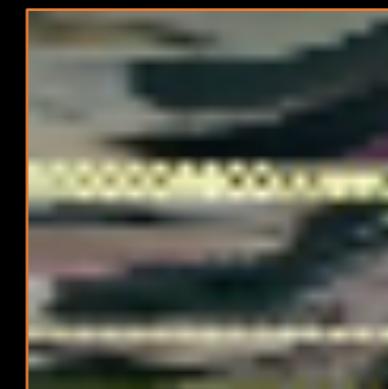
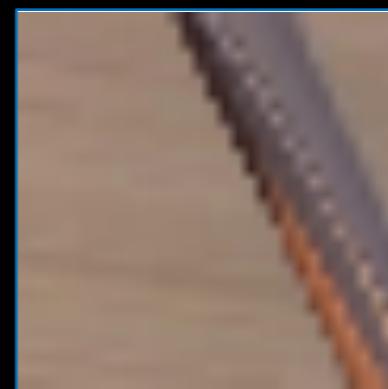
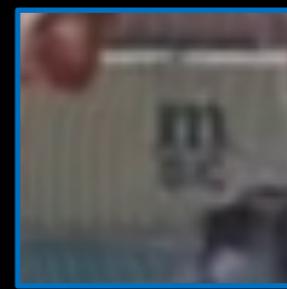
# Self-supervised Learning

Method	EPE	Method	Acc.
LD Flow [3]	12.4	Random	39.1
FlowNet [5]	9.1	Unsup. Video [30]	43.8
EpicFlow [22]	3.8	ImageNet [14]	63.3
Ours (w/o ft.)	14.6	Ours (w/o ft.)	48.7
Ours	<b>9.5</b>	Ours	<b>52.4</b>

Flow estimation

Action Recognition

# Spatio-temporal Coherence



EpicFlow

Ground Truth

Ours

EpicFlow

Ground Truth

Ours

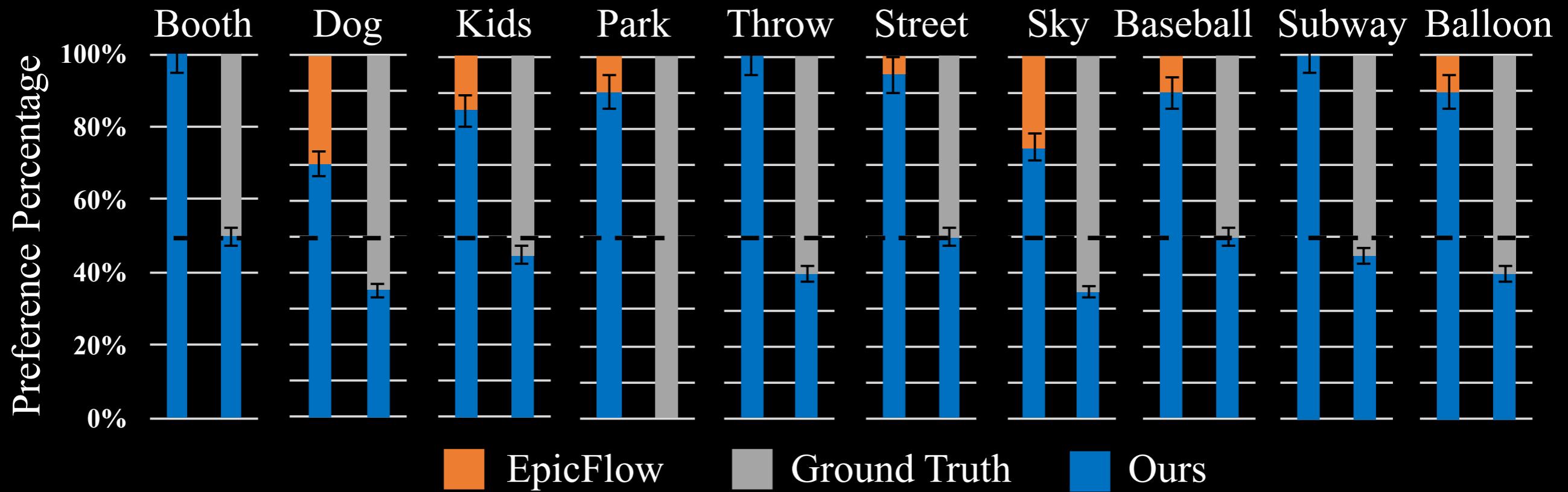
# User Study

Diagonal-split Comparison



Method 1 \ Method 2

# User Study



# Real-life Applications

## **Video Frame Synthesis using Deep Voxel Flow**

Ziwei Liu<sup>1</sup>, Raymond Yeh<sup>2</sup>, Xiaoou Tang<sup>1</sup>,  
Yiming Liu<sup>3</sup>, Aseem Agarwala<sup>3</sup>

<sup>1</sup>The Chinese University of Hong Kong

<sup>2</sup>University of Illinois at Urbana-Champaign

<sup>3</sup>Google

# Conclusions

- Unified motion representation
- End-to-end learning system
- Long-range dependencies in the future

# Thanks!

Project Page: <https://liuziwei7.github.io/projects/VoxelFlow>