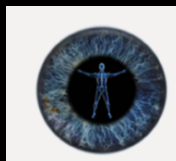




VISION SCIENCE
UNIVERSITY OF CALIFORNIA, BERKELEY

Learning Video Parsing, Tracking and Synthesis in the Wild

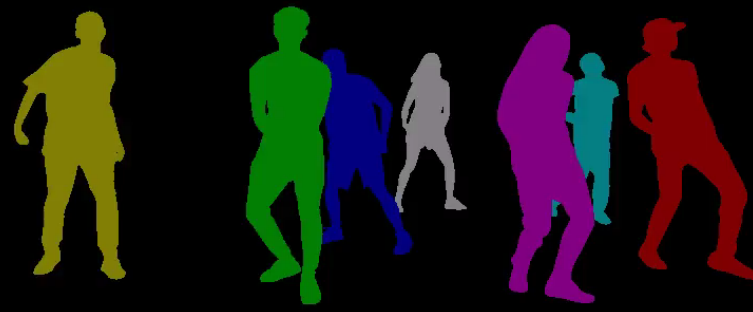
Ziwei Liu

International Computer Science Institute, UC Berkeley

Let computer understand videos

Video Parsing

Let computer manipulate videos



Video Tracking

Let computer create videos

Video Synthesis

Video Parsing

"Not All Pixels are Equal: Difficulty-aware Semantic Segmentation via Deep Layer Cascade", *CVPR 2017 (spotlight)*

Problem

Input Video

State-of-the-art Method (4 FPS)

Deep Layer Cascade (17 FPS)

Motivation: Not All Pixels are Equal

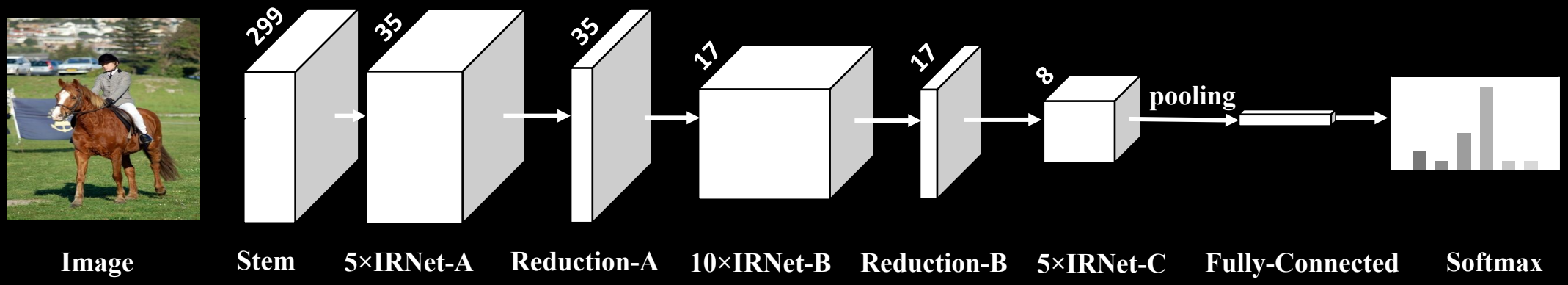
Image

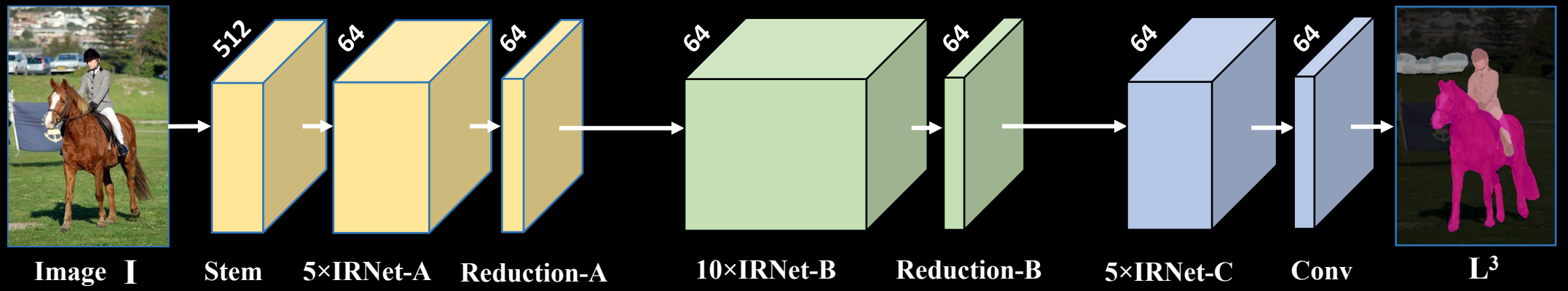
Easy Region

Moderate Region

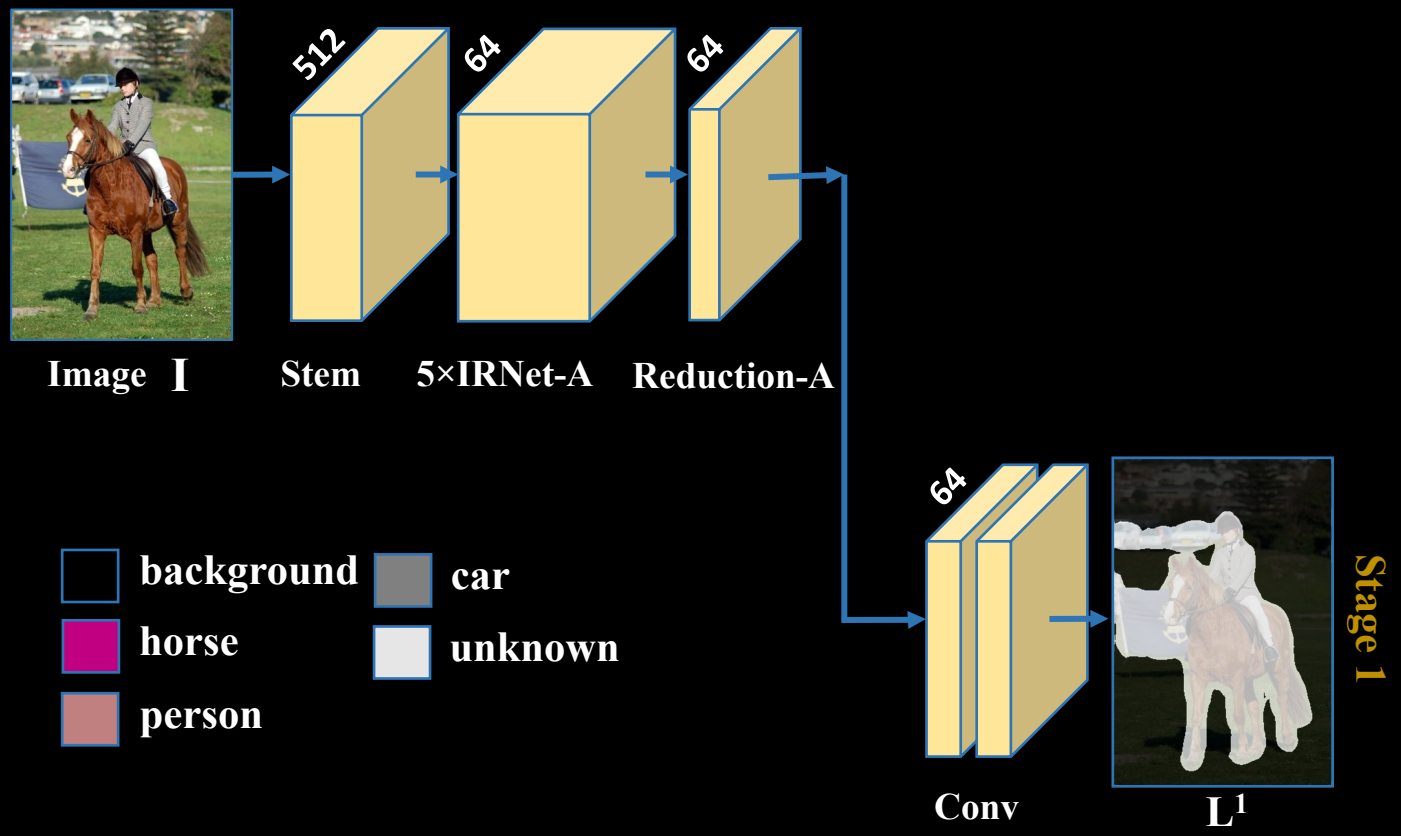
Hard Region

Contemporary Model

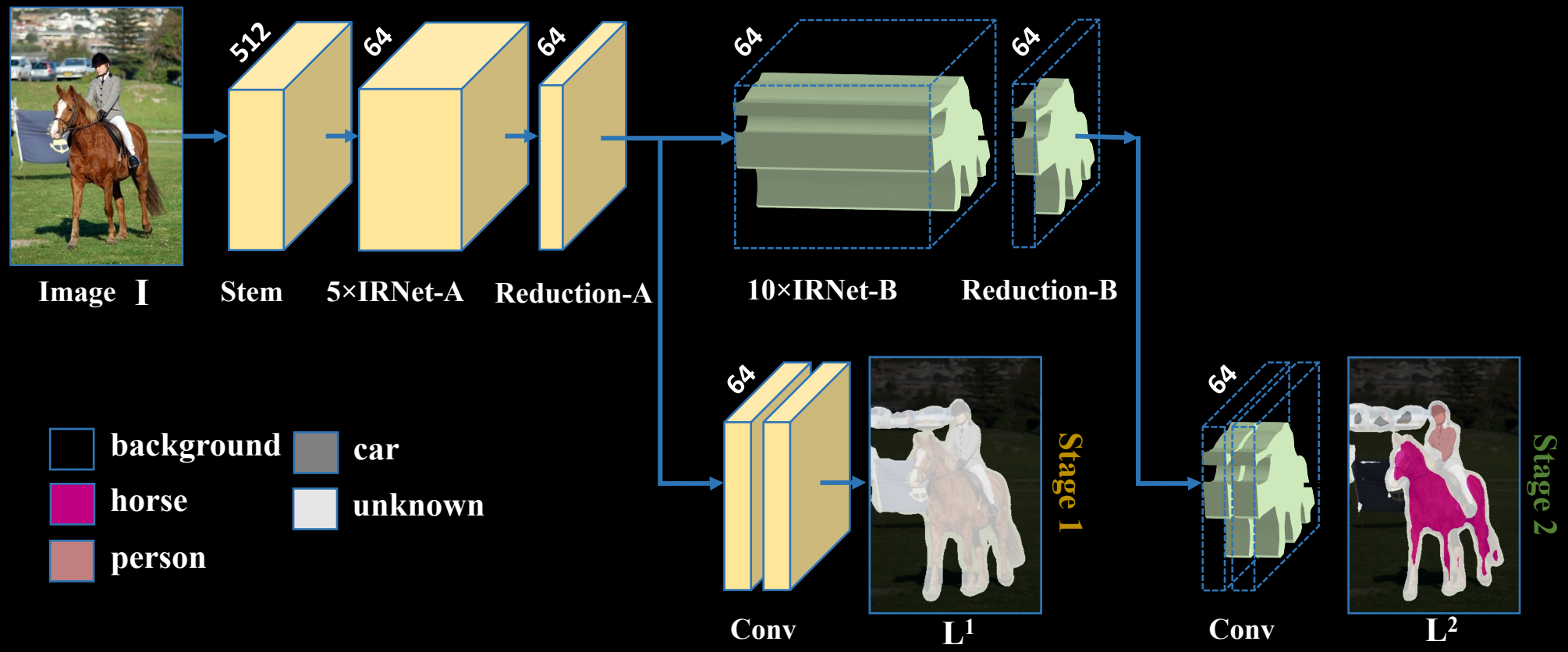




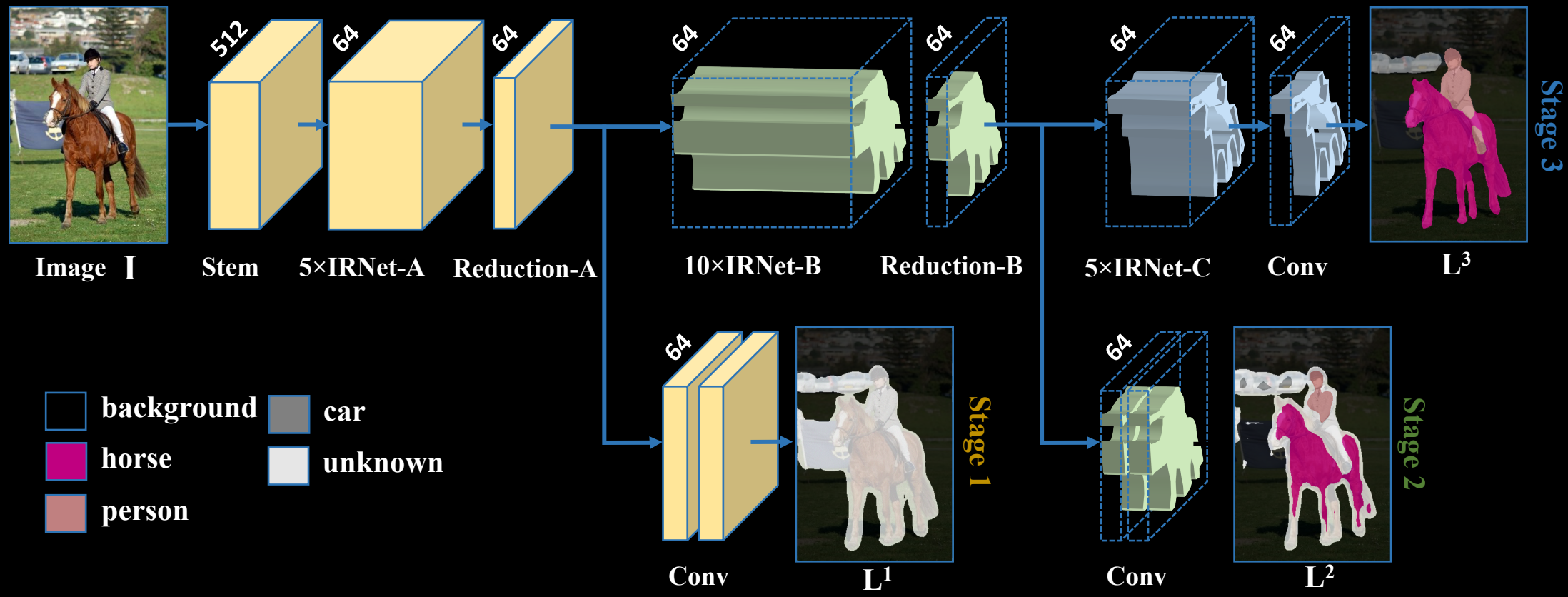
Deep Layer Cascade

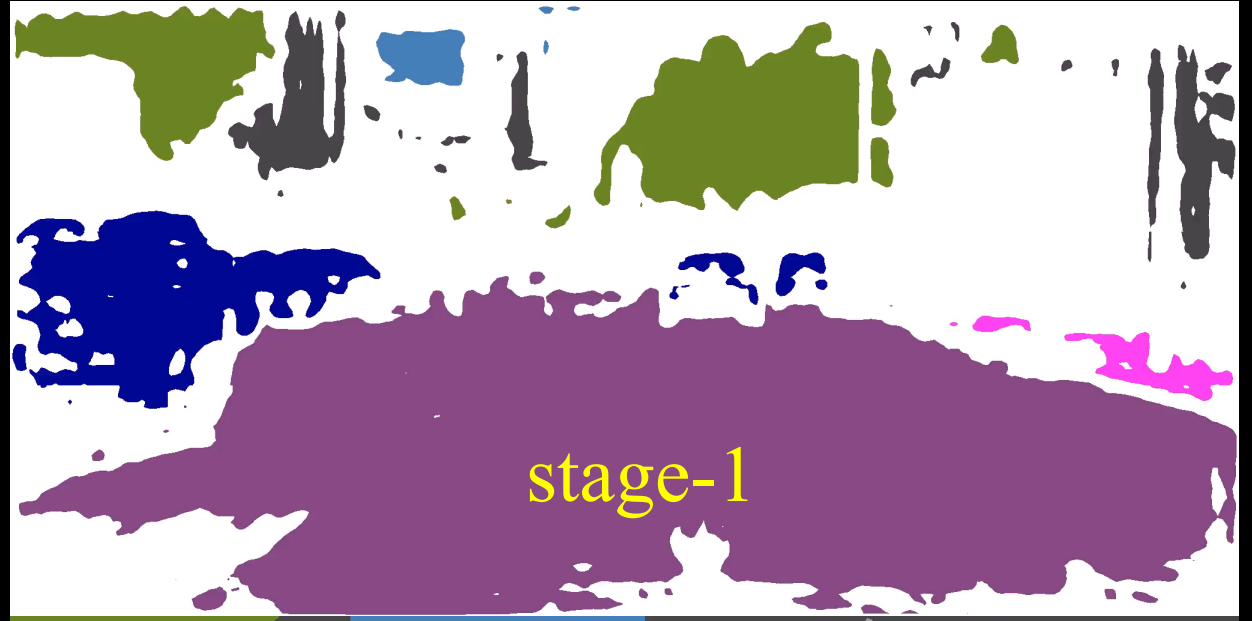


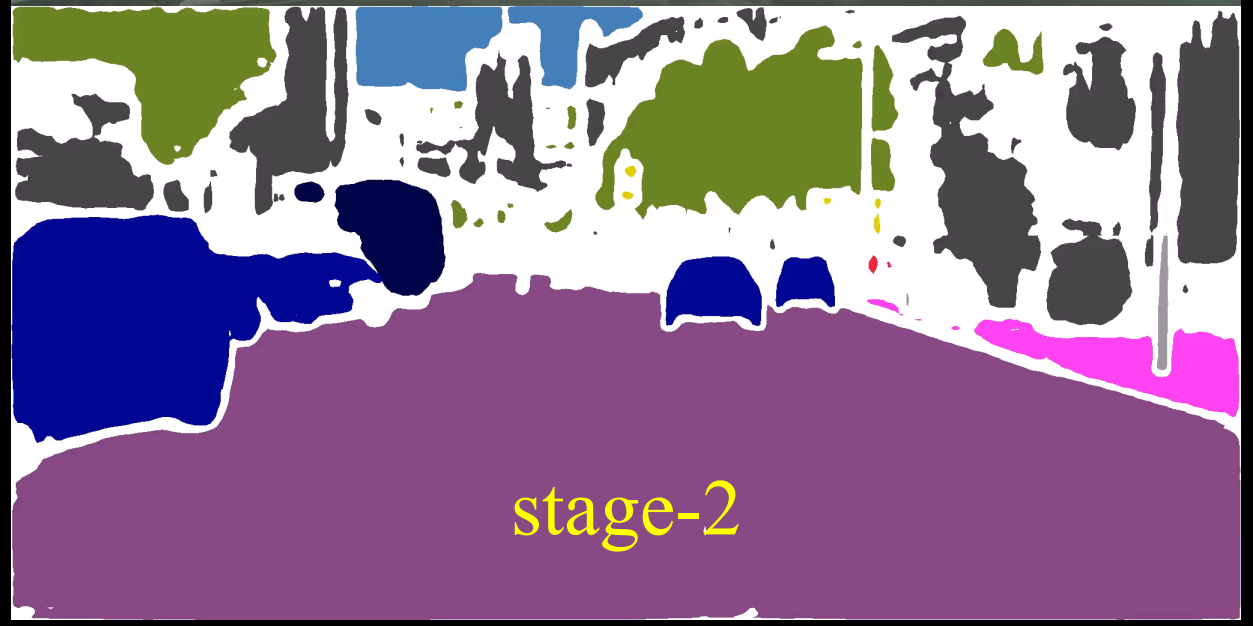
Deep Layer Cascade

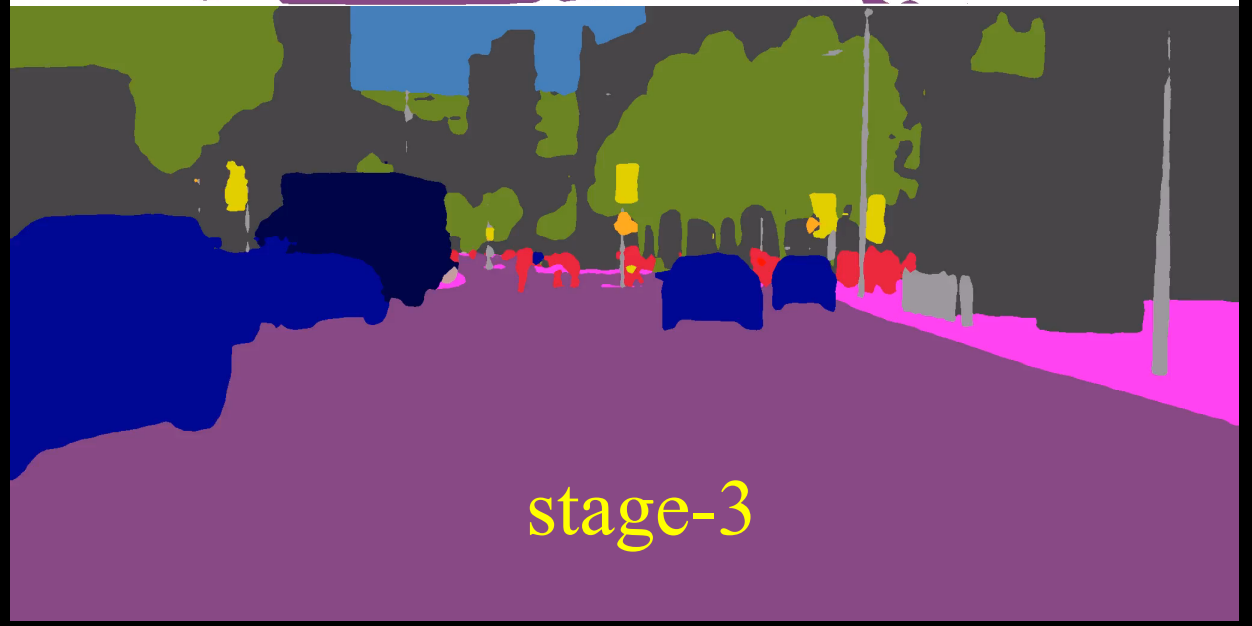


Deep Layer Cascade









- Difficulty-Aware Learning Paradigm
- Region Convolution → Real-Time
- End-To-End Trainable Framework

Video Tracking

"Video Object Segmentation with Re-identification", *CVPR 2017 Workshop (winning entry)*

Problem

- Input :Video sequence, ground-truth label of the first frame

- Output : Masks of all instances

Challenge

- Instance Segmentation
  - Small objects and fine structures
  - Scale & pose-variations
- Tracking
  - Frequent occlusions

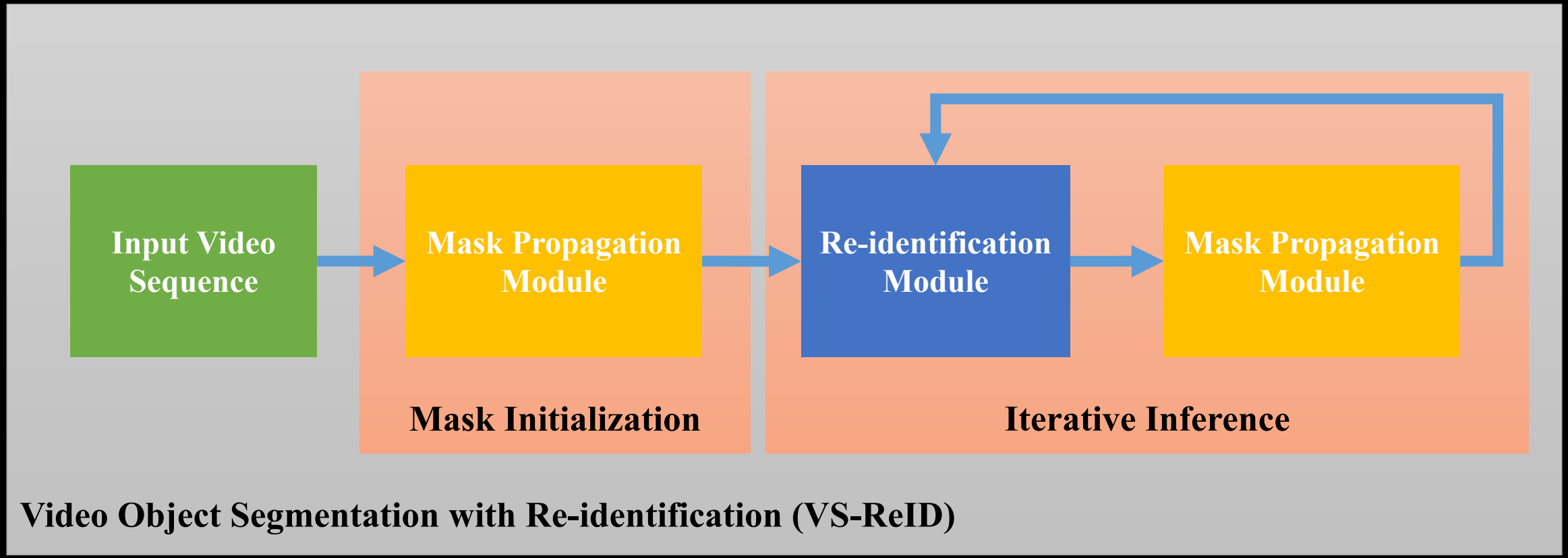
Challenge

- Instance Segmentation
  - Small objects and fine structures
  - Scale & pose-variations
- Tracking
  - Frequent occlusions

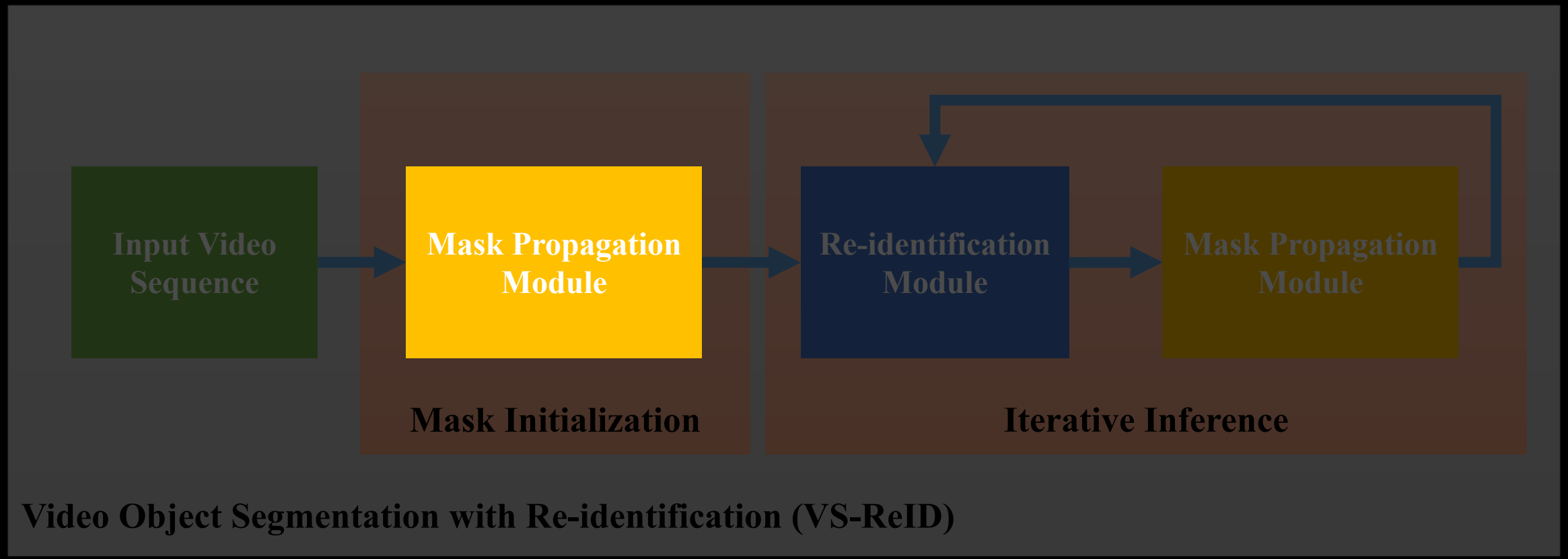
Mask Propagation Module — Short Term

Re-identification Module — Long Term

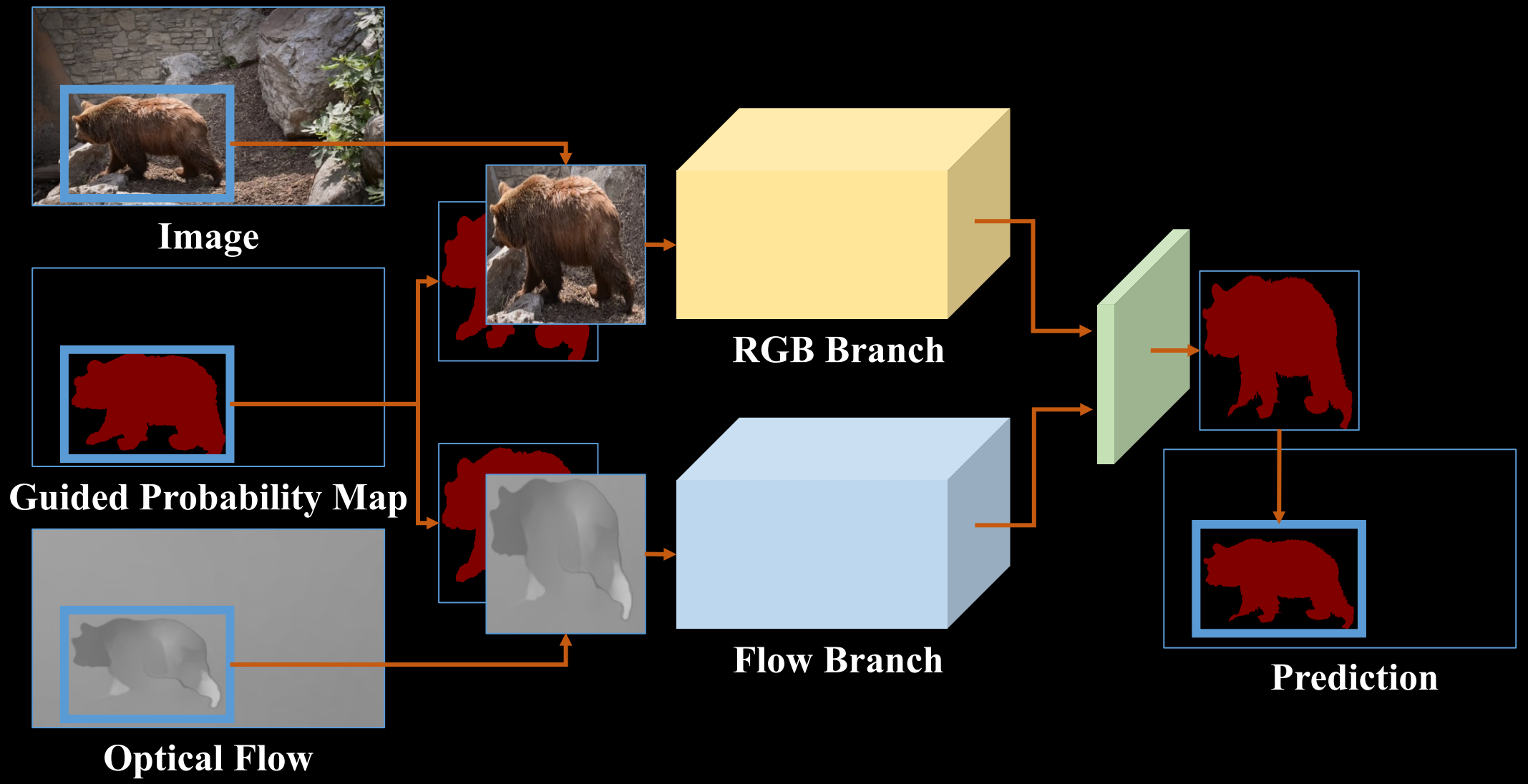
Proposed Framework



Mask Propagation Module



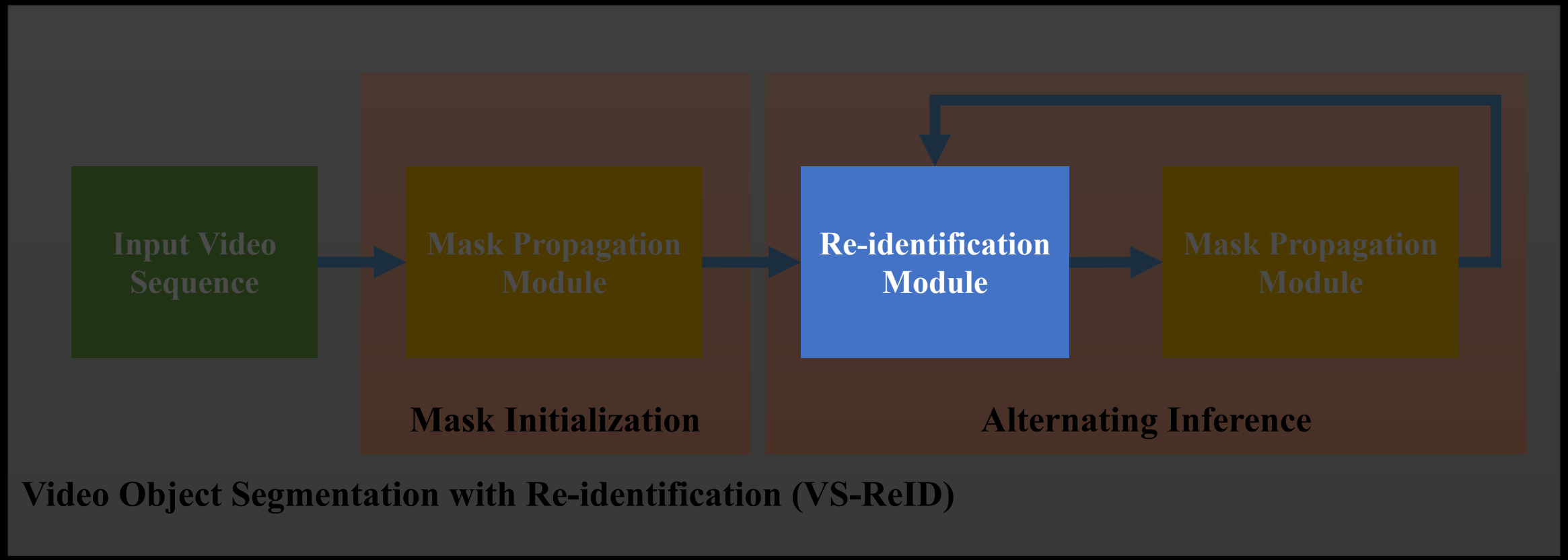
Mask Propagation Module



Video Frame
Guided Probability Map
Prediction
Warping
Mask Propagation Module

Mask Propagation Module

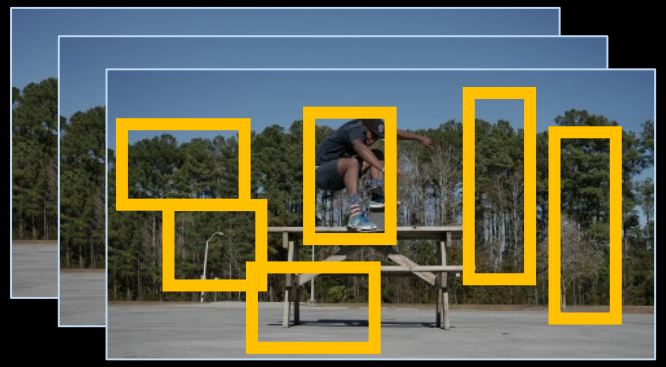
Proposed Framework



Re-identification Module

- Detection and re-identification

First Frame



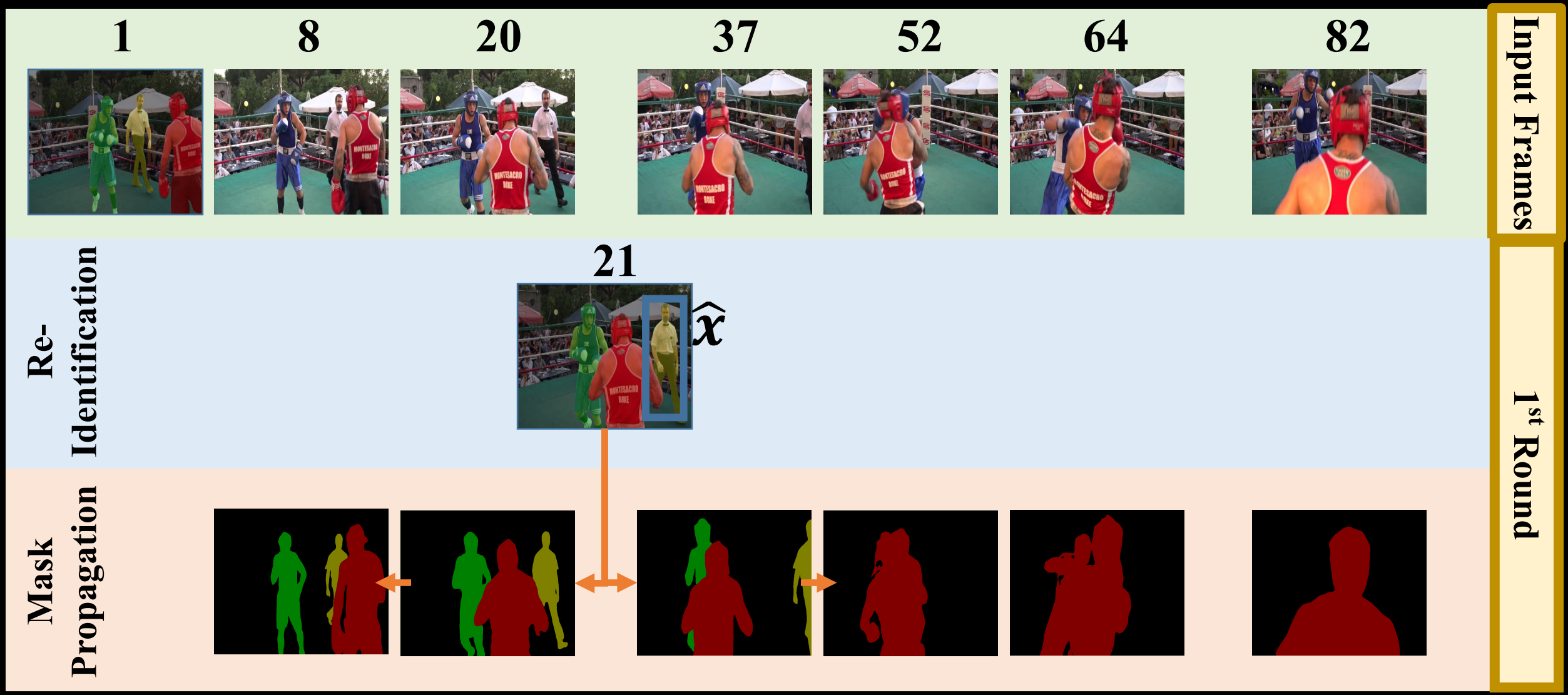
Rest Frames

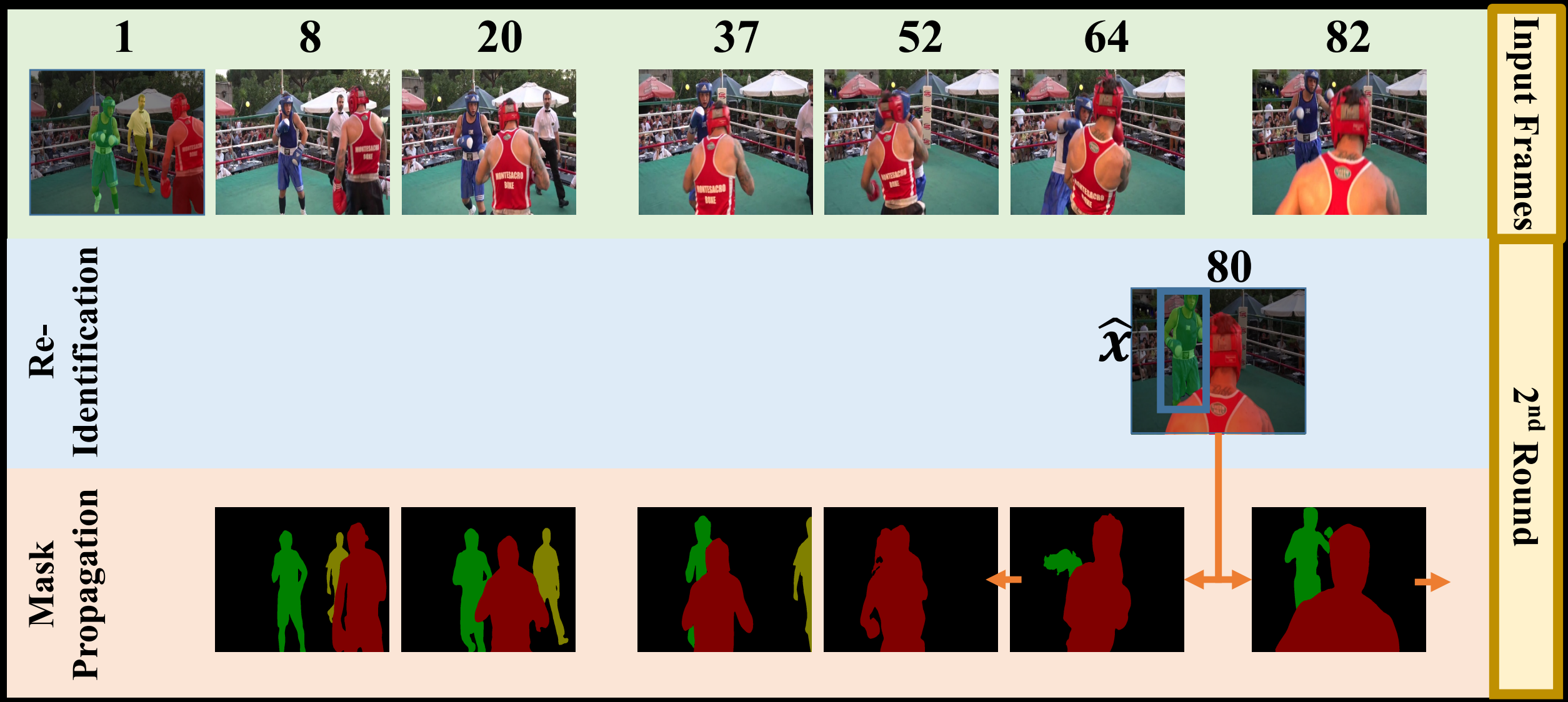
Template

Re-identification

Candidate Bounding Boxes

Most Confident Candidate


1
8
20
37
52
64
82
21
$\hat{x}$
Input Frames
1st Round
Re-Identification
Mask Propagation


1
8
20
37
52
64
82
80
$\hat{x}$
Input Frames
Re-Identification
Mask Propagation
2nd Round

Results

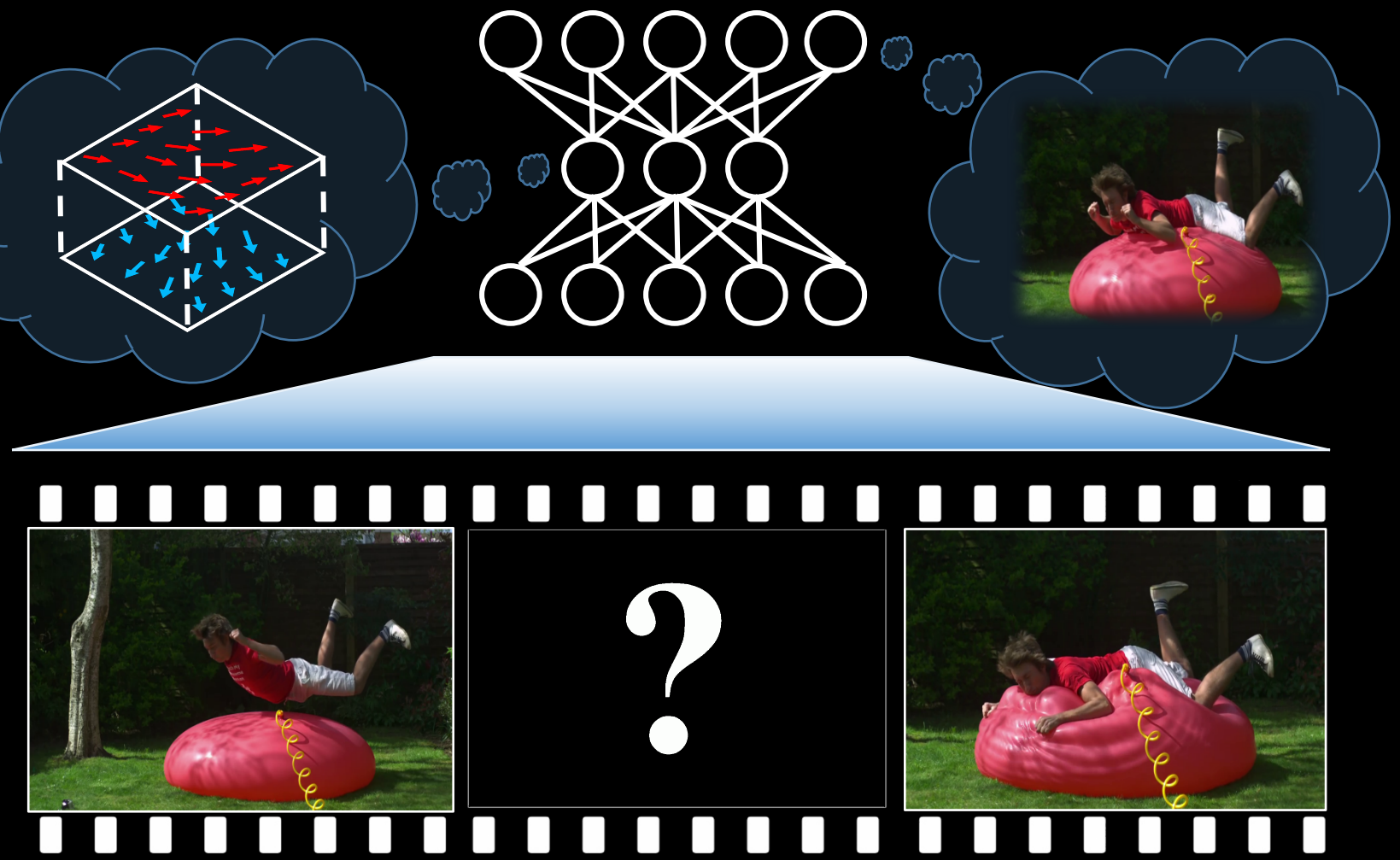
Video Synthesis

"Video Frame Synthesis using Deep Voxel Flow", *ICCV 2017 (oral)*

Video Frame Synthesis

- Problem

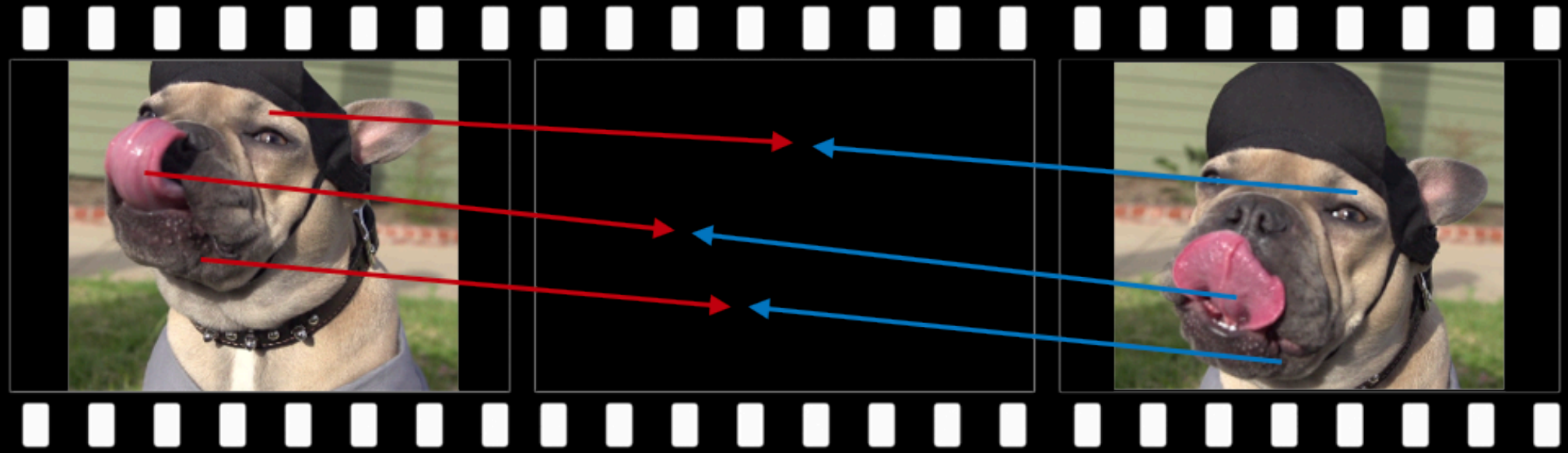
Video interpolation/ extrapolation



Video Frame Synthesis

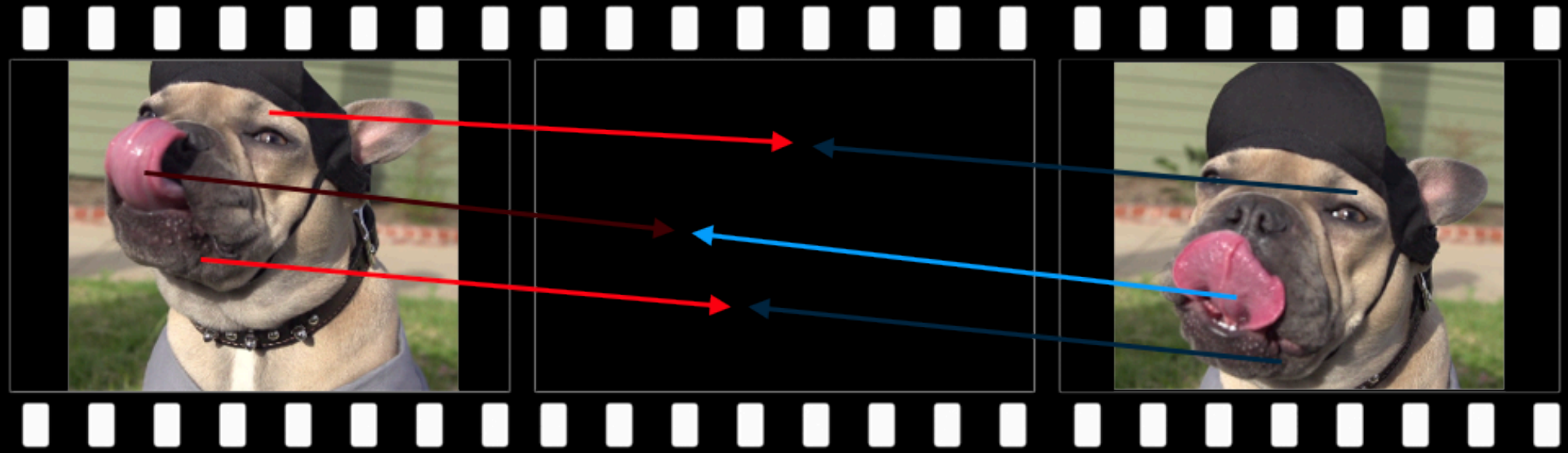
- Challenge
  1. Complex motion (camera motion & scene motion)
  2. High-res images (1280 * 720)

Voxel Flow



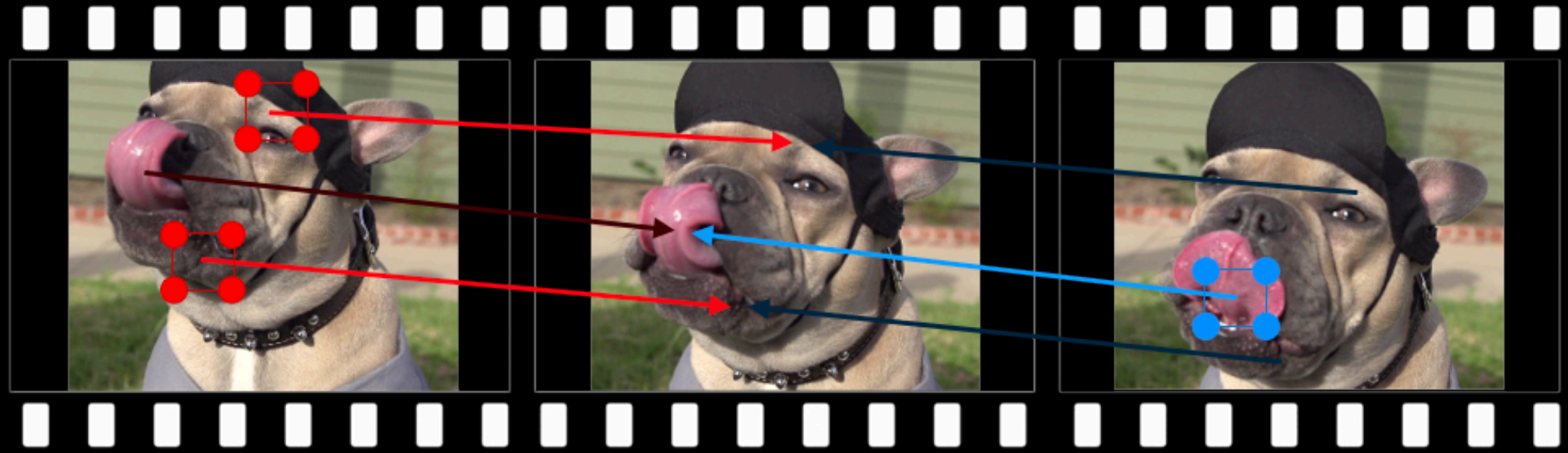
symmetric bi-directional flows

Voxel Flow



selection mask between frames

Voxel Flow

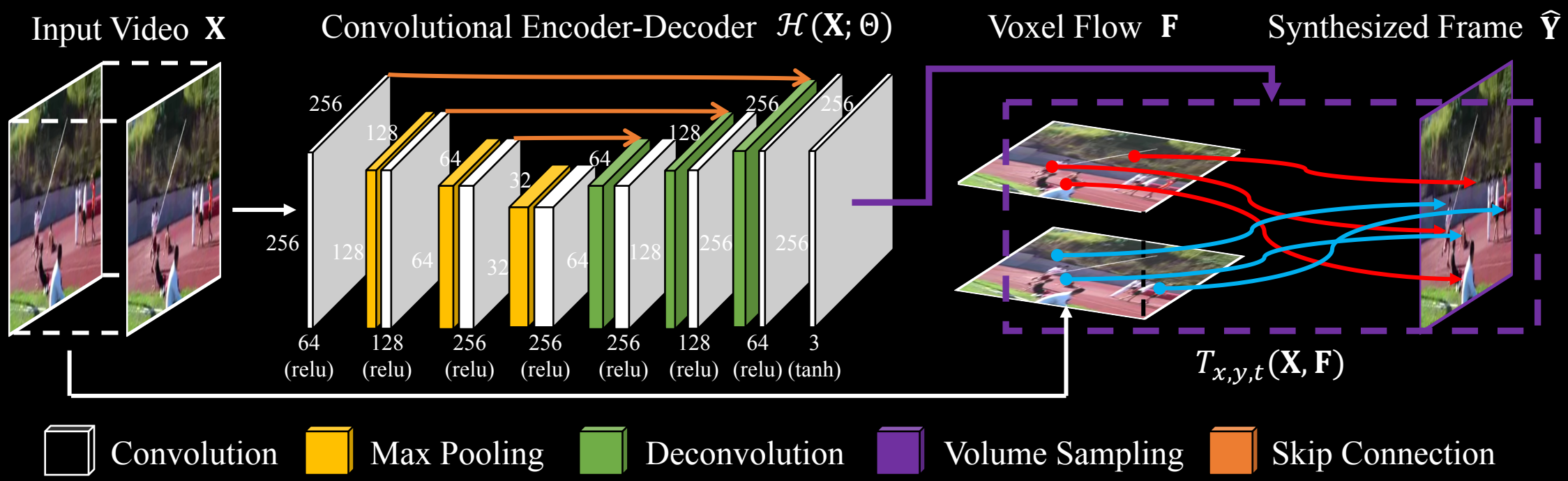


differentiable bilinear sampling

Deep Voxel Flow

- Motivation

Combining the strength of flow-based and NN-based methods



Visualization

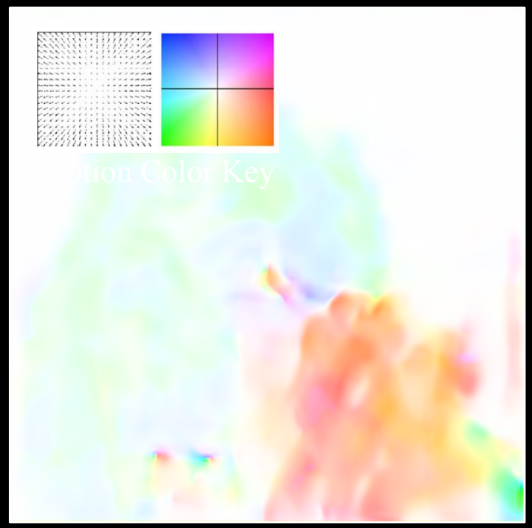
- Advantages

(a) 2D Flow + Mask

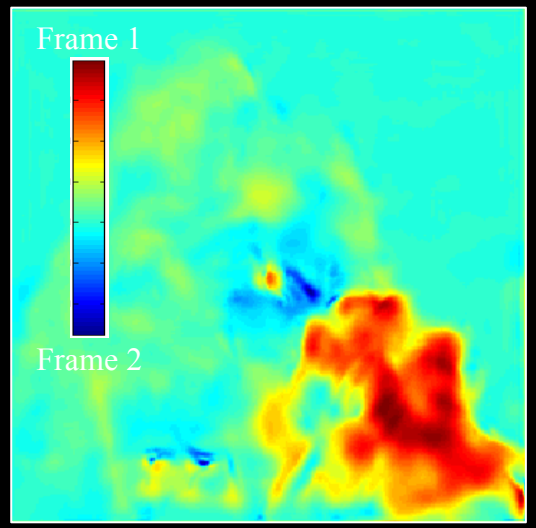
(b) Voxel Flow

(c) Multi-scale Voxel Flow

(d) Difference Image



(e) Projected Motion Field



(f) Projected Selection Mask

Comparisons

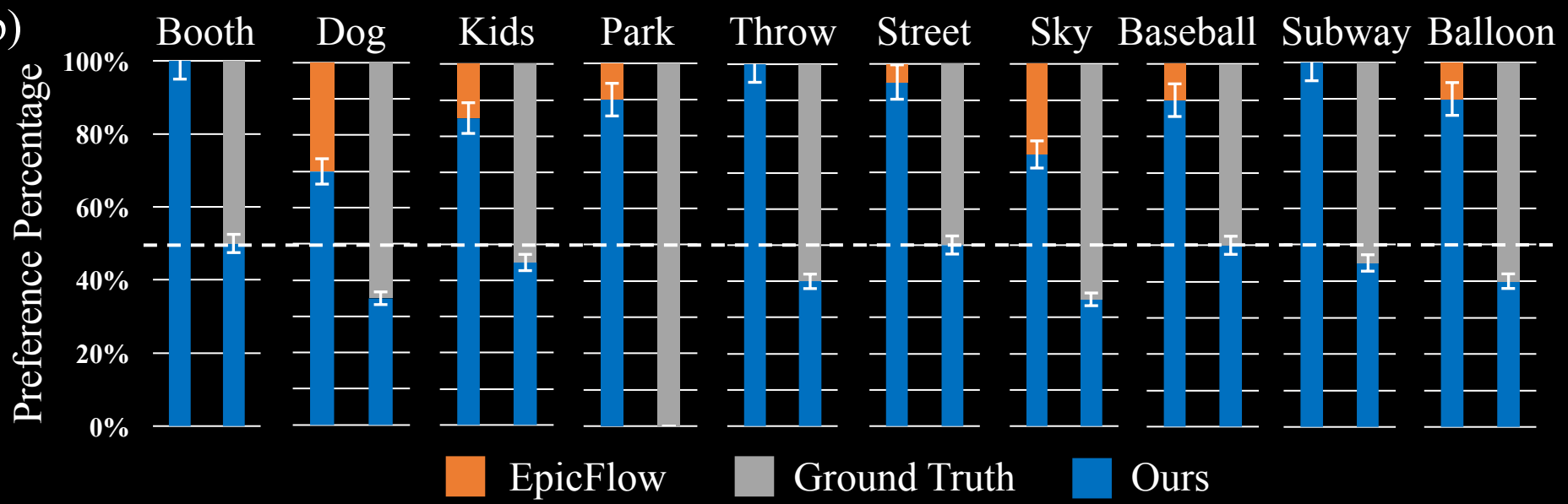
- User Study

(a)

Diagonal-split Comparison

Method 1 \ Method 2

(b)



Results

- UCF-101

Results

- KITTI

We can understand videos

We can manipulate videos

We can create videos

Video Frame Synthesis using Deep Voxel Flow

Ziwei Liu[1], Raymond Yeh[2], Xiaoou Tang[1],
Yiming Liu[3], Aseem Agarwala[3]

[1]The Chinese University of Hong Kong
[2]University of Illinois at Urbana-Champaign
[3]Google

Product Transfer

Google Clips

Thanks!