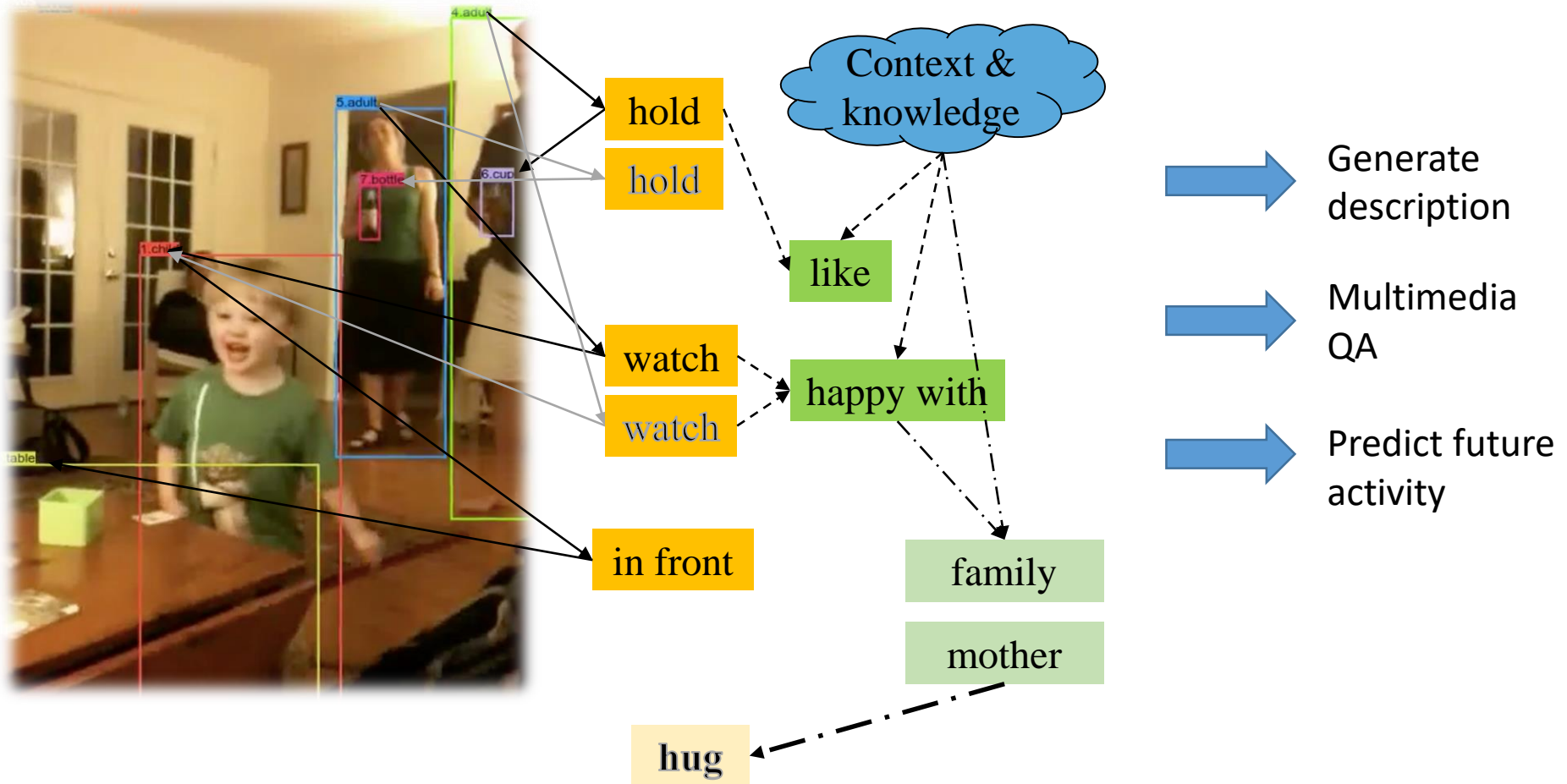




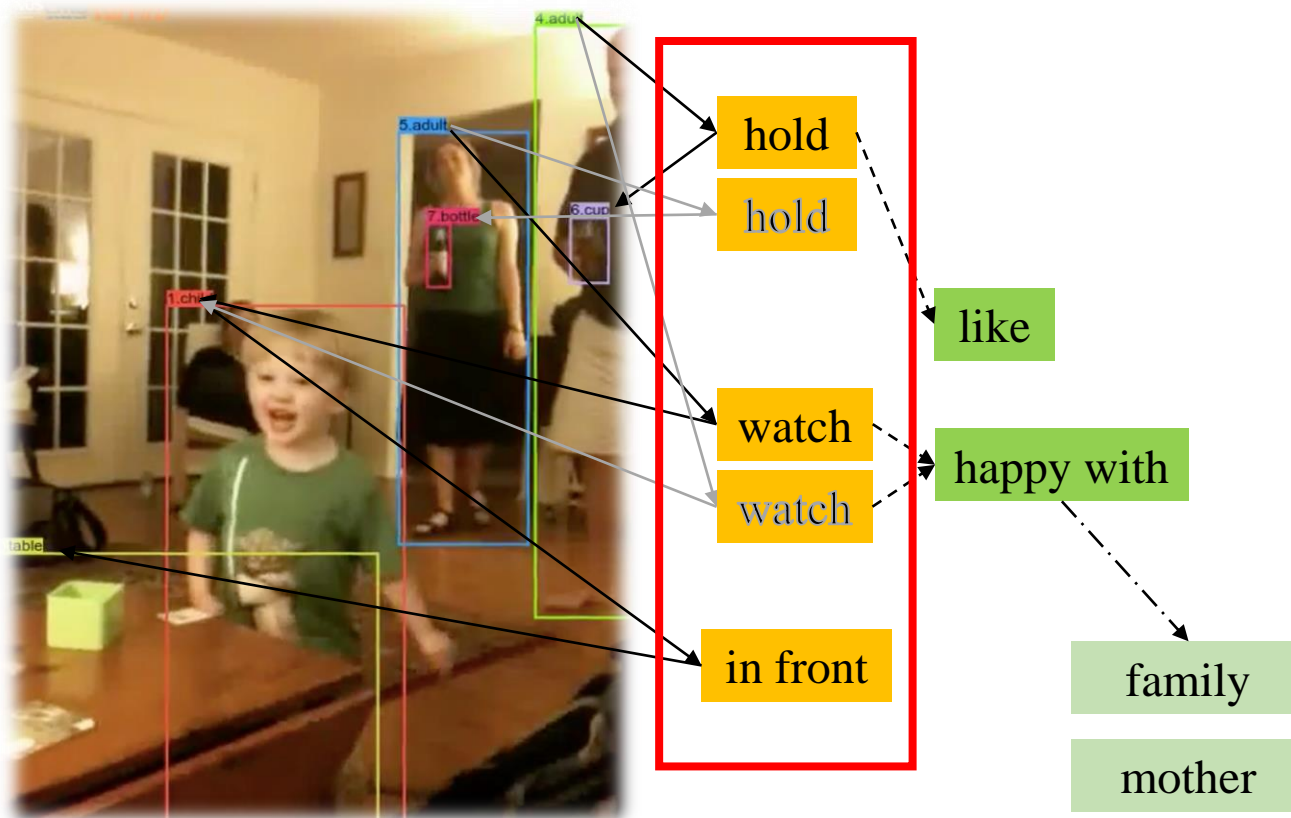
NUS-Tsinghua-Southampton
Centre for Extreme Search

Video Relation Inference

Xindi Shang, NUS



- Visual relation provides evidences for high-level semantic relation
- Visual Relation Detection (VRD)



Video Object Tracking

Image Object Detection

Video Feature Engineering

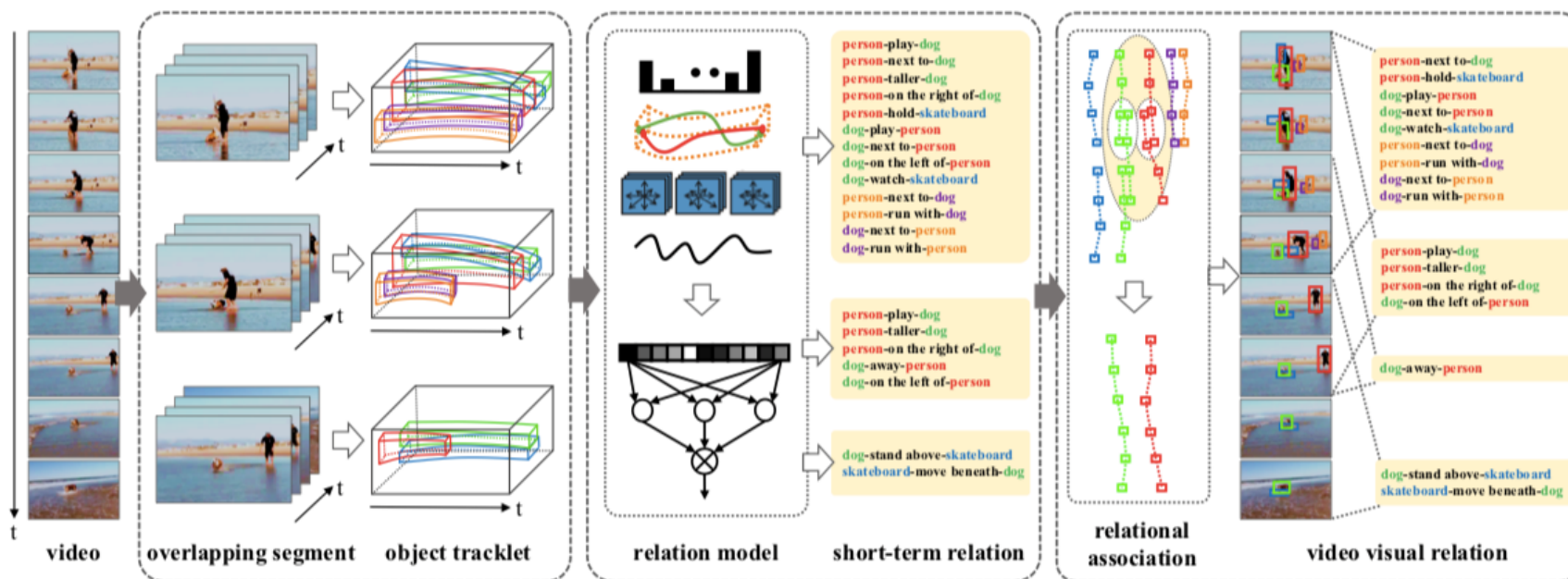
Action Recognition

Time-Series Modeling

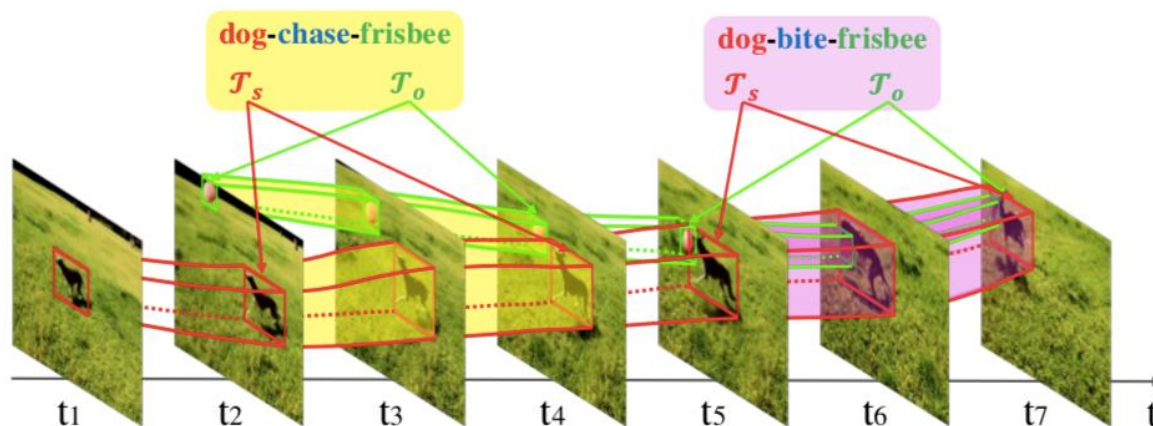
Video Object Detection

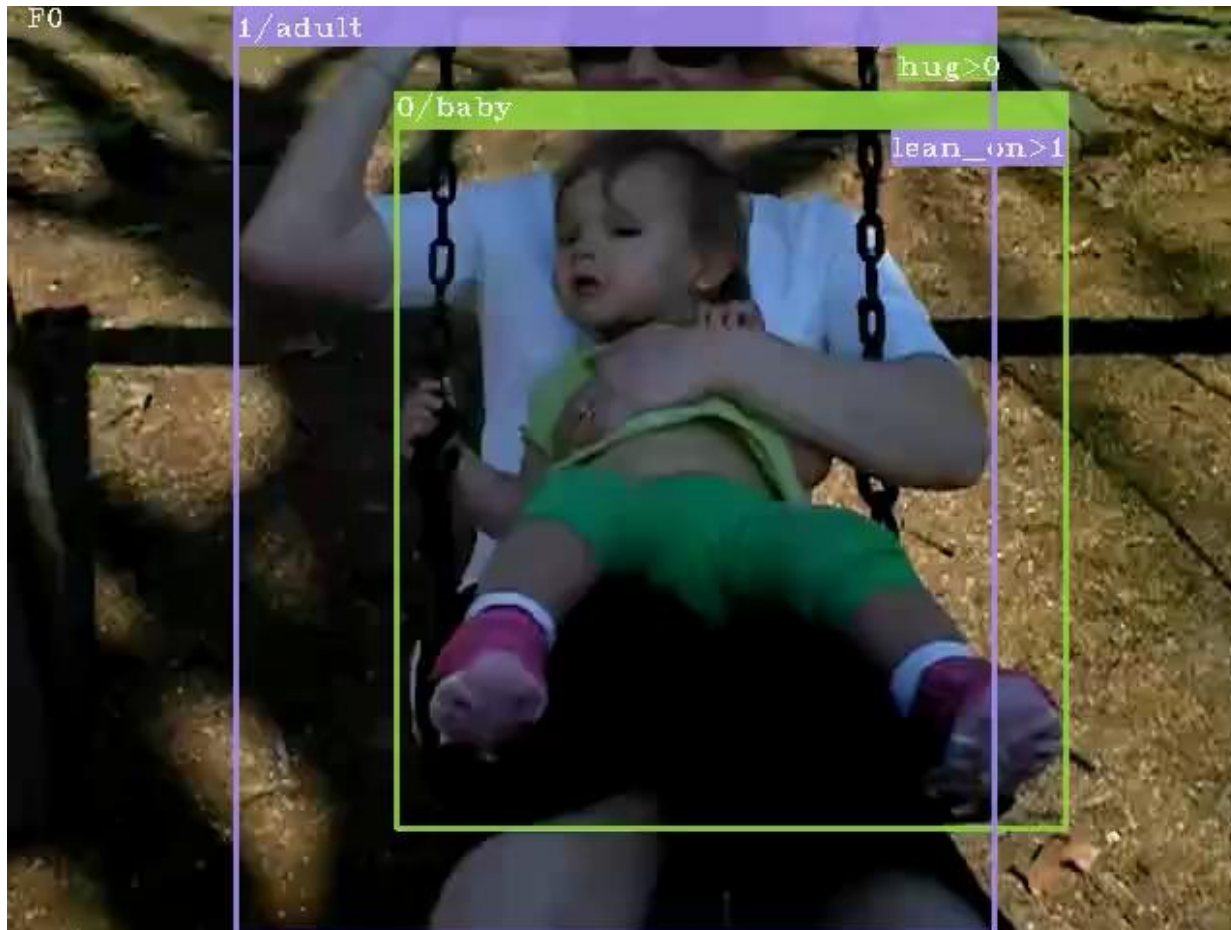
Relation Prediction

Temporal Localization

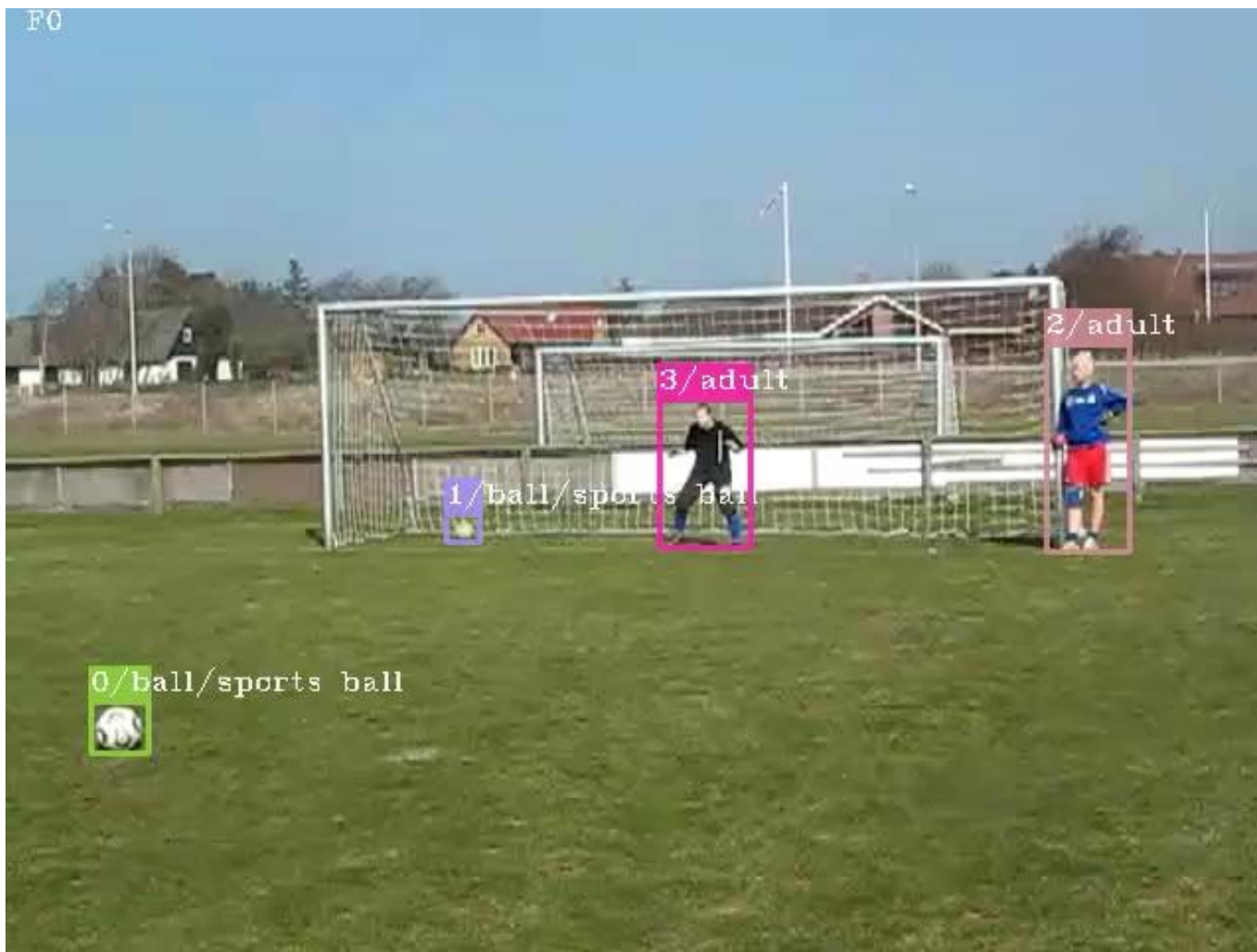


- 10K videos from social platform
 - 30s long in average
- Object entity annotation
 - 80 categories: adult, child, dog, chair, ball, laptop, etc.
 - Annotated 2,000,000 bounding boxes to localize the objects
- Relation annotation
 - 50 categories: watch, hold, lean_on, push, lift, kiss, grab, get_on, etc.
 - Annotated 100,000 relation instances



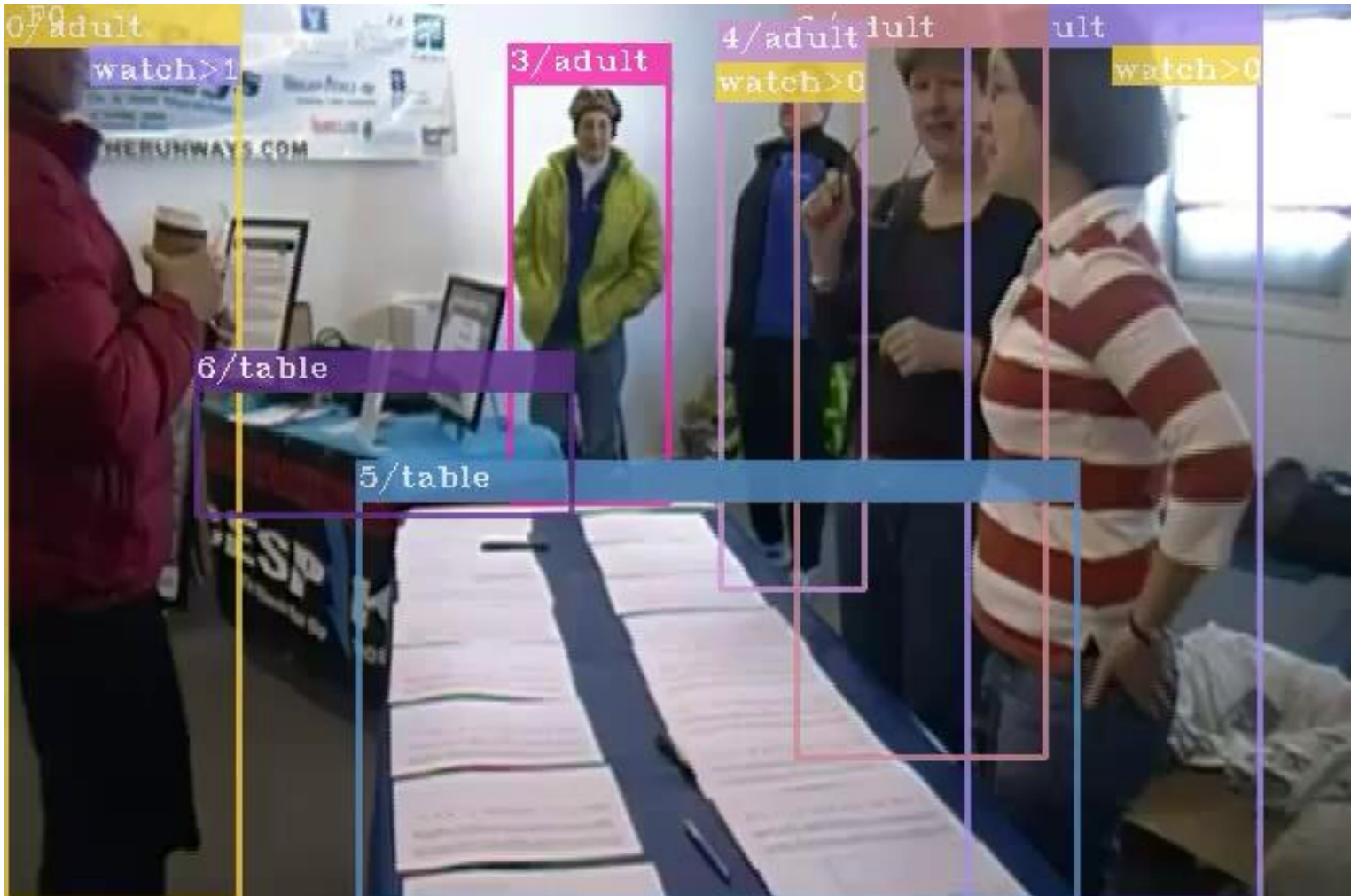


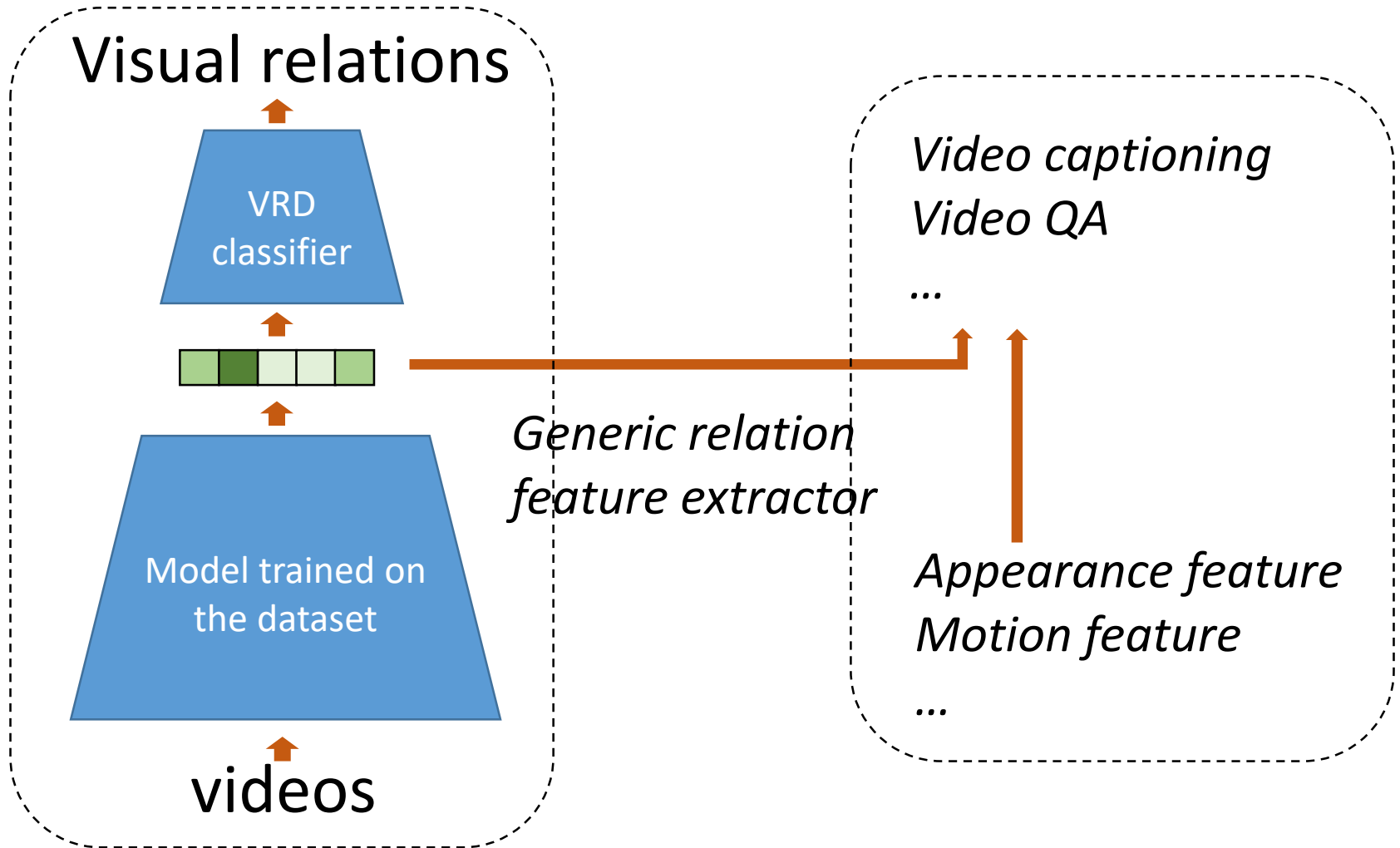




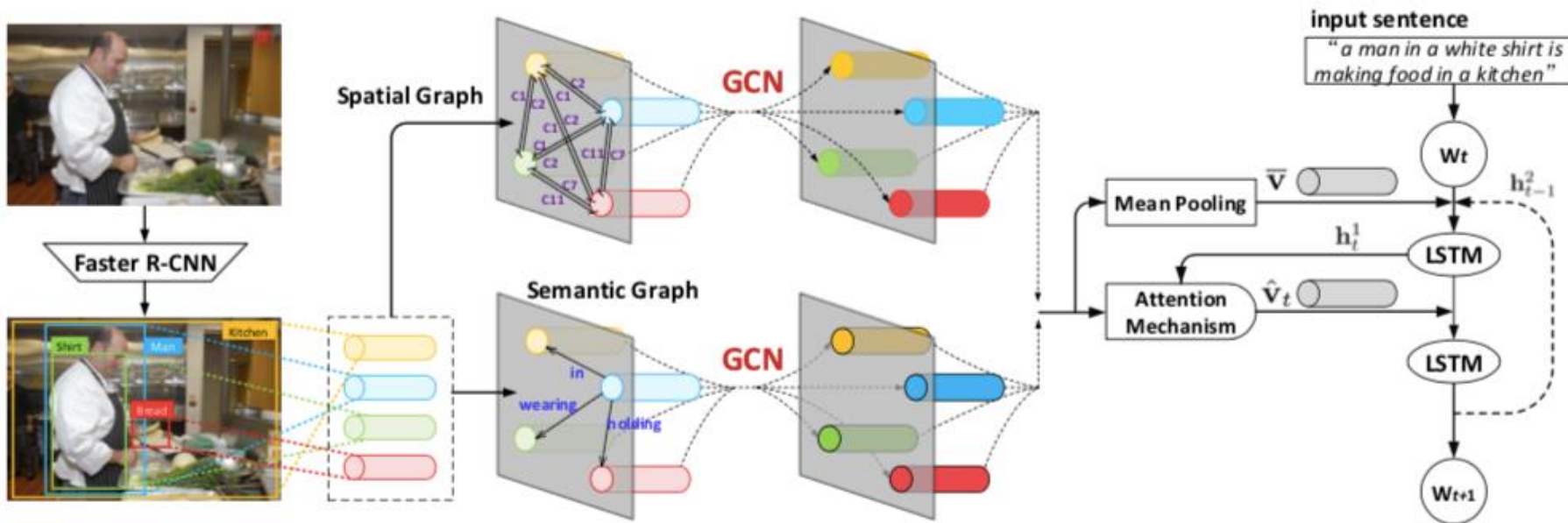


NEXT++ Dataset Samples





- Directly utilize relation triplets



“Exploring Visual Relationship for Image Captioning”, Yao Ting, Pan Yingwei , Li Yehao and Mei Tao, ECCV 2018



Summary

- Understanding relations is critical towards holistic video content understanding
- We made 10K video relation dataset to improve current VRD model
- In future work, use VRD to improve video captioning and QA
- <https://lms.comp.nus.edu.sg/research/VidVRD.html>

NExT++

National University of Singapore
13 Computing Drive
Singapore 117417



NUS-Tsinghua-Southampton
Centre for Extreme Search

THANK YOU

NExT++ research is supported by the National Research Foundation,
Prime Minister's Office, Singapore under its IRC@SG Funding Initiative.

shangxin@comp.nus.edu.sg