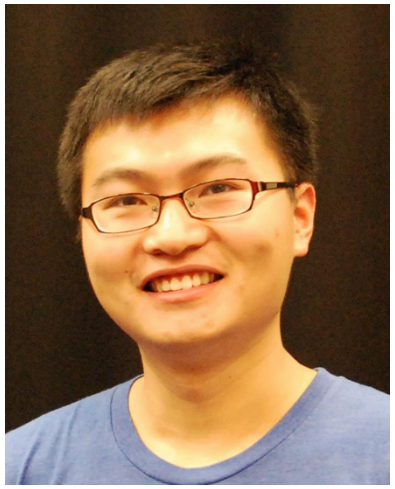
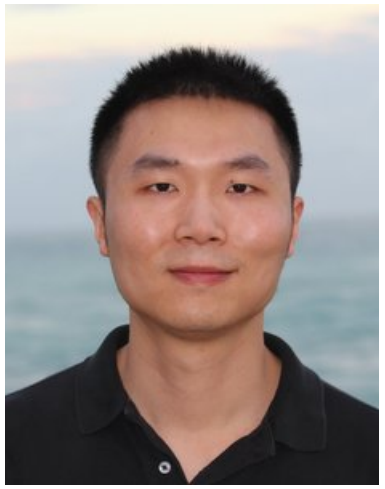
IMAGENET Large Scale Visual Recognition Challenge (ILSVRC) 2016

Object Detection from Video (VID)

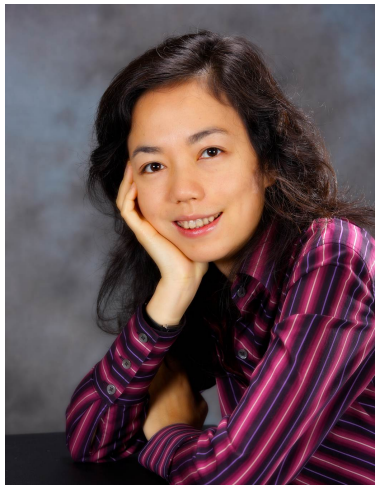


Wei Liu
UNC Chapel Hill

Olga Russakovsky
CMU



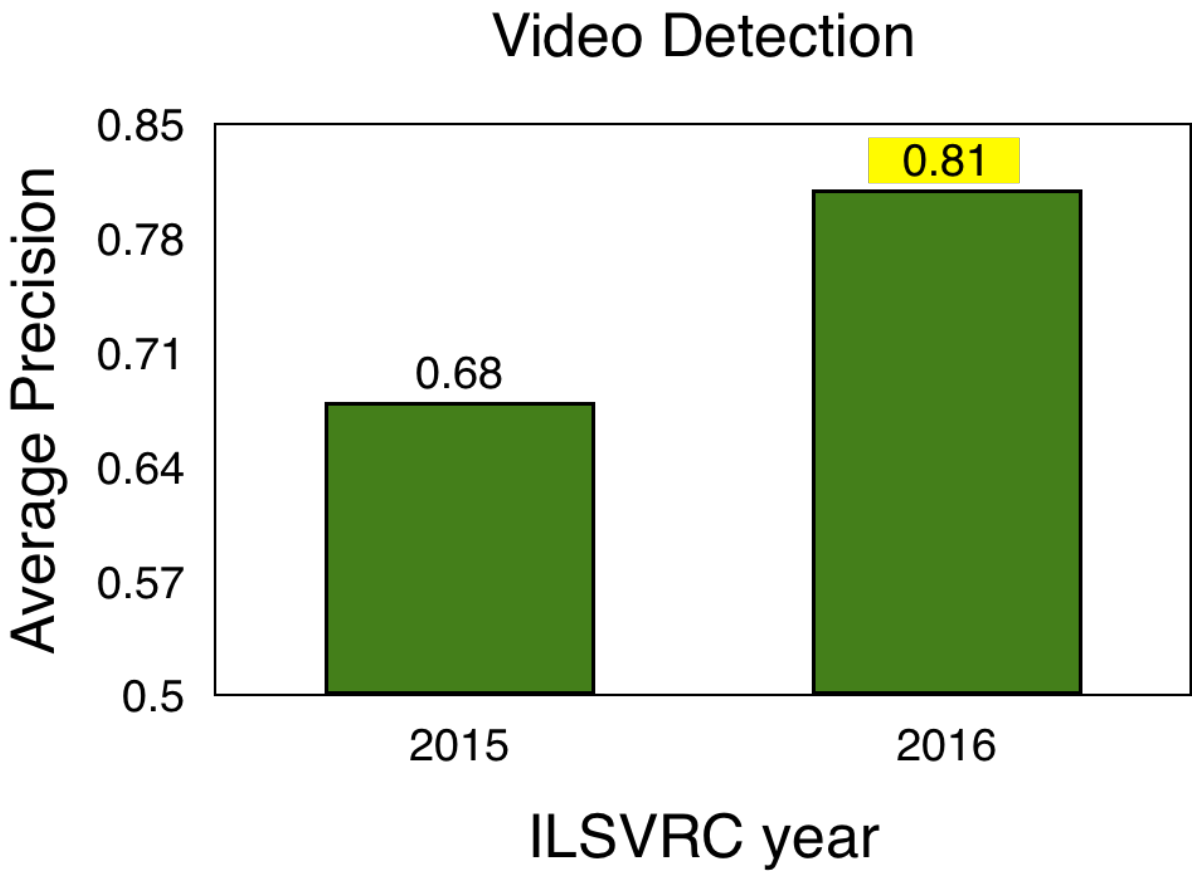
Jia Deng
Univ. of Michigan



Fei-Fei Li
Stanford

Alex Berg
UNC Chapel Hill

Result in ILSVRC VID over the years



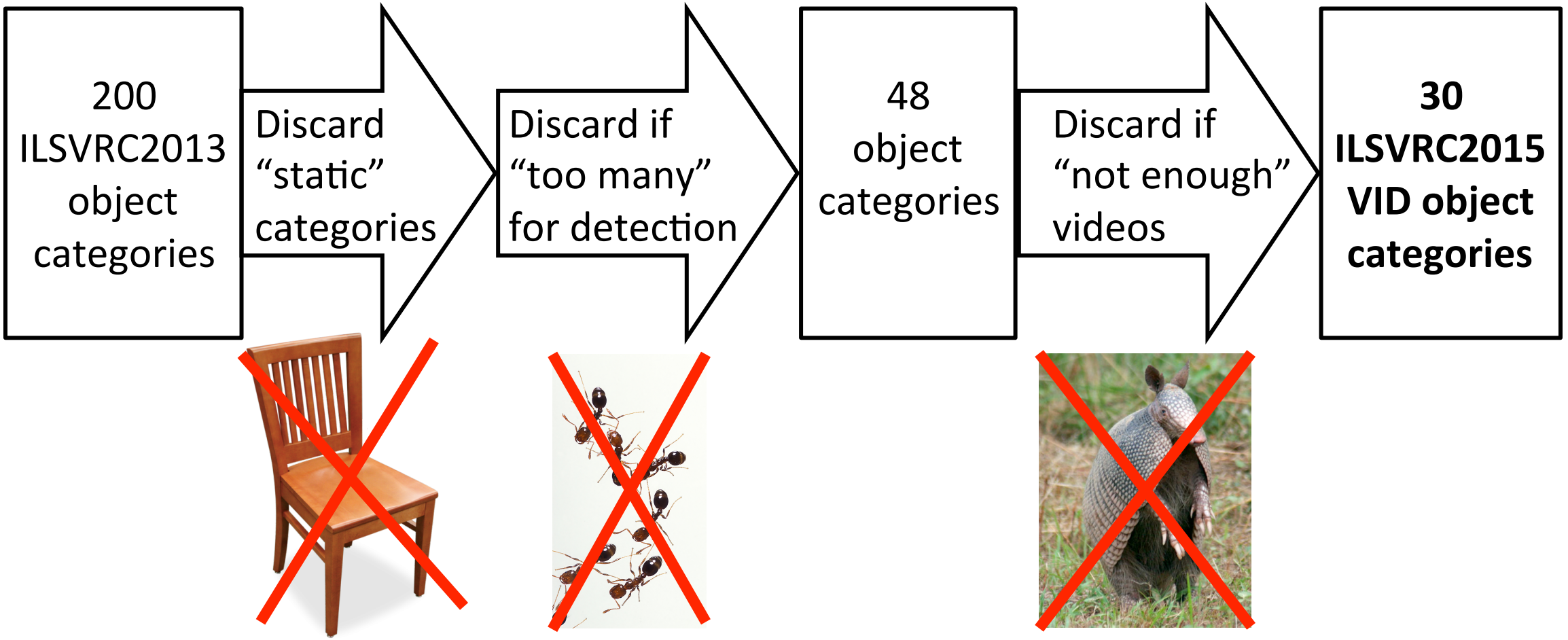
Object detection from video (VID)

Fully annotated 30 object classes across 6,278 snippets (train+test)

Allows evaluation of generic object detection
in cluttered videos at scale

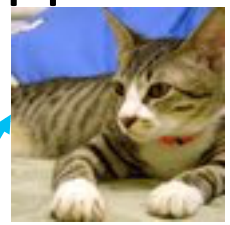
video data collection

- Step 1: Define object categories



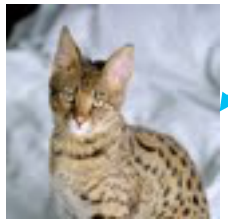
video data collection

- Step 1: Define object categories
- Step 2: Collect snippets



Tabby cat

Cat → Freebase → /m/0g4cd0 … /m/068m4y → YouTube Data API → Cat videos → Manually cleanup → Segment video to snippets → Manually cleanup



Savannah cat

Creative Common License

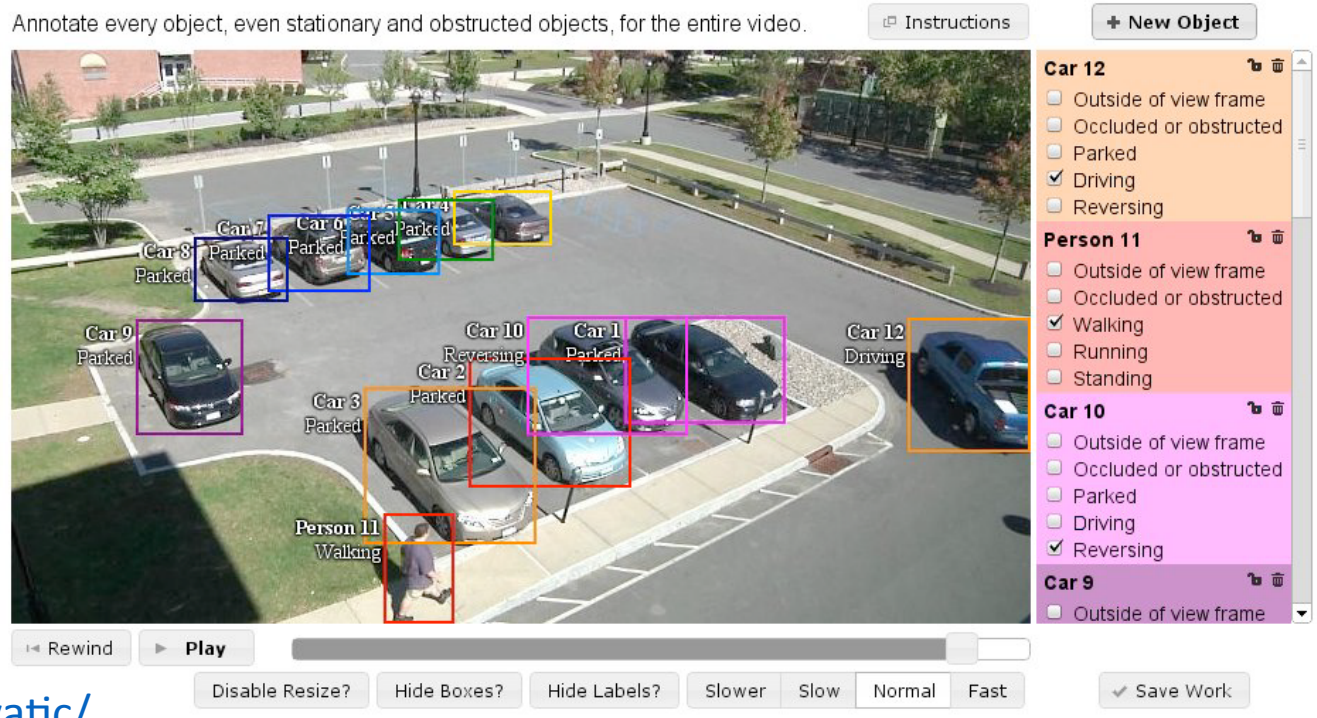
Has cat in video
Fix missing class
No cartoon video
No music video
…

Shot detection
Manually segment

Not too crowded
Not too simple
…

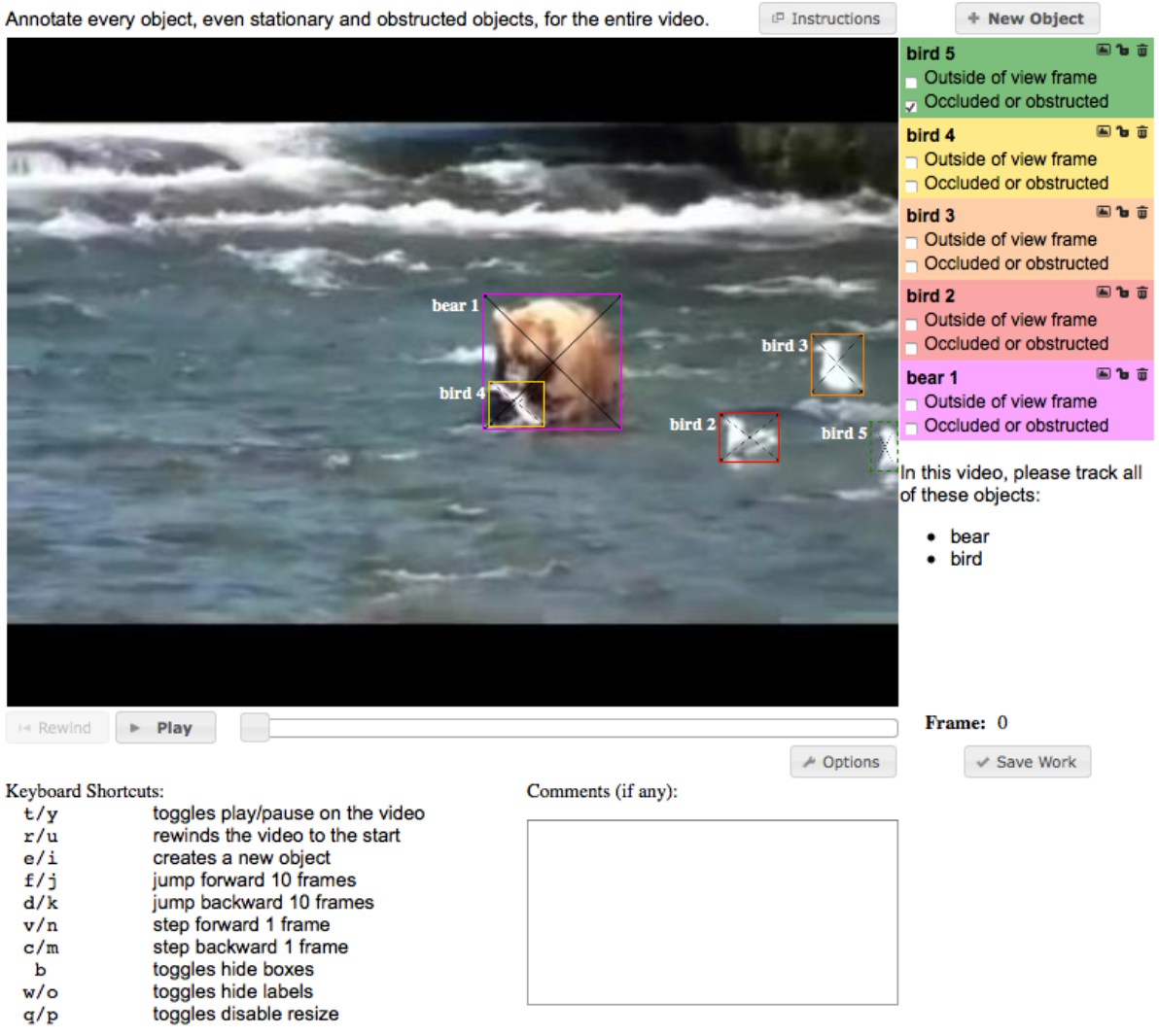
video data collection

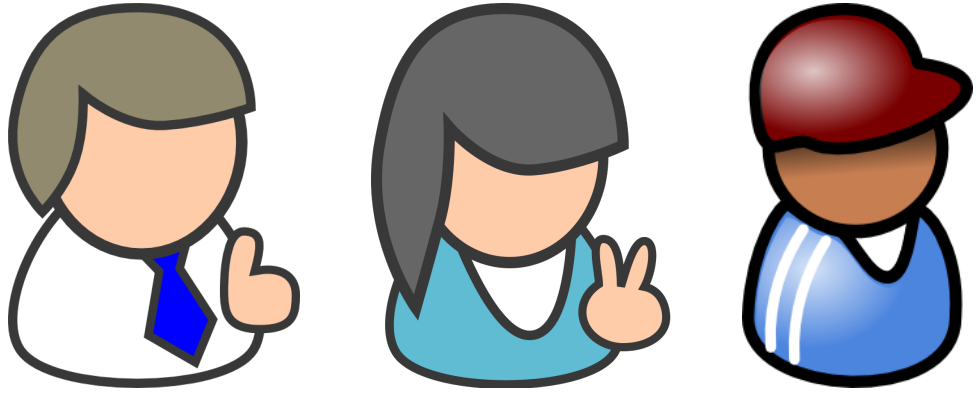
- Step 1: Define object categories
- Step 2: Collect videos
- Step 3: Annotate bounding boxes completely for all categories



http://web.mit.edu/vondrick/vatic/

video data collection





http://github.com/weiliu89/vatic/tree/vid

ILSVRC object detection from video (VID)

Evaluation modeled after PASCAL VOC:

- Algorithm outputs a list of bounding box detections with confidences
- A detection is considered correct if intersection over union (IoU) overlap with ground truth > threshold (0.5)
- Evaluated by average precision per object class
- Winners of challenge is the team that wins the most object categories

Everingham, Van Gool, Williams, Winn and Zisserman.
The PASCAL Visual Object Classes (VOC) Challenge. IJCV 2010.

ILSVRC object detection from video (VID)

New metric: Take tracking into account

- Algorithm outputs a list of bounding box detections with confidences and tracklet ID.
- Tracklets are sorted by the mean confidence.
- A tracklet is considered correct if intersection over union (IoU) overlap with ground truth tracklet > threshold (0.25, 0.5, 0.75).
- Evaluated by average precision per class. Final score is an average over different thresholds.
- Winners of challenge is the team that has the highest score.

New metric: Take tracking into account

panda1 ground truth

panda2 ground truth

Exemplar video with ground truth

New metric: Take tracking into account

Exemplar video with detection results

Step 1: Compute mean score for tracklet

Exemplar video with detection results

Step 2: Sort tracklets based on mean score

Exemplar video with detection results

Step 3: Evaluate each tracklet

✔ ✔ ✔

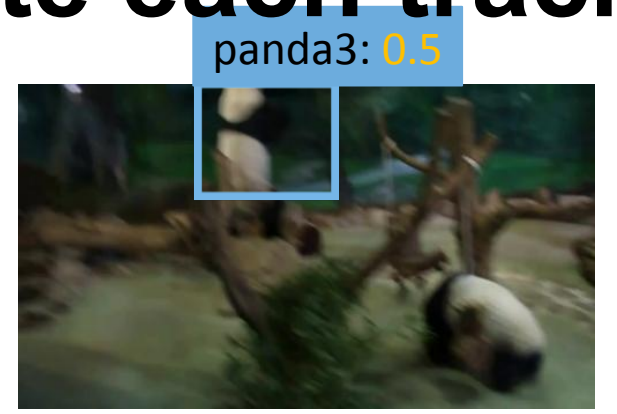
positive tracklet

1. Every frame uses bbox IoU overlap 0.5 to determine if it is a true positive detection or not.
2. Then, compute the IoU overlap between detected tracklet and ground truth tracklet.

$$\text{tracklet IoU} = \frac{\text{\# of detected frames}}{\text{\# of union frames}} = 3 / 4 = 0.75 \quad > 0.5 \text{ (a threshold)}$$

Exemplar video with detection results

Step 3: Evaluate each tracklet



✔ ✔ ✔

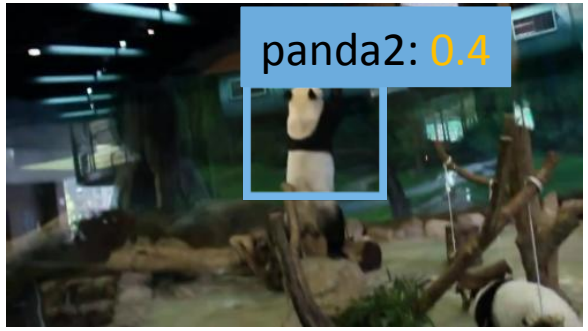
positive tracklet

$$\text{tracklet IoU} = \frac{\text{\# of detected frames}}{\text{\# of union frames}} = 2 / 3 = 0.66$$

Exemplar video with detection results

Step 3: Evaluate each tracklet



✔

negative tracklet

$$\text{tracklet IoU} = \frac{\text{\# of detected frames}}{\text{\# of union frames}} = 1 / 3 = 0.33$$

Note: Duplicate detected tracklets are considered as negative as well

Exemplar video with detection results

ILSVRC2016 VID results – with "provided" data

Team Name	Number of categories won	Mean Average Precision(%)
NUIST	10	80.8
CUVideo	9	76.8
Trimps-Soushen	1	71.0
MCG-ICT-CAS	0	73.3
KAIST-SLSP	0	64.3

NUIST:
Jing Yang, Hui Shuai, Zhengbo Yu, Rongrong Fan, Qiang Ma, Qingshan Liu, Jiankang Deng

CUvideo:
Hongsheng Li*, Kai Kang*, Wanli Ouyang, Junjie Yan, Tong Xiao, Xingyu Zeng, Kun Wang, Xihui Liu, Qi Chu, Junming Fan, Yucong Zhou, Yu Liu, Ruohui Wang, Shengen Yan, Dahua Lin, Xiaogang Wang

(* indicates equal contribution)
The Chinese University of Hong Kong, SenseTime Group Limited

ILSVRC2016 VID results – with "external" data

Team Name	Number of categories won	Mean Average Precision(%)
NUIST	17	79.6
Trimps-Soushen	5	72.1
ITLab-Inha	3	73.1
DPAI Vision	0	61.5
TEAM1	0	21.8

NUIST:
Jing Yang, Hui Shuai, Zhengbo Yu, Rongrong Fan, Qiang Ma, Qingshan Liu, Jiankang Deng

Trimps-Soushen:
Jie Shao, Xiaoteng Zhang, Zhengyan Ding, Yixin Zhao, Yanjun Chen, Jianying Zhou, Wenfei Wang, Lin Mei, Chuanping Hu
The Third Research Institute of the Ministry of Public Security, P.R. China.

ILSVRC2016 VID tracking results – with "provided" data

Team Name	Mean Average Precision(%)
CUVideo	55.9
NUIST	54.9
MCG-ICT-CAS	48.9
KAIST-SLSP	32.7
CIGIT_Media	23.0

CUvideo:
Hongsheng Li*, Kai Kang*, Wanli Ouyang, Junjie Yan, Tong Xiao, Xingyu Zeng, Kun Wang, Xihui Liu, Qi Chu, Junming Fan, Yucong Zhou, Yu Liu, Ruohui Wang, Shengen Yan, Dahua Lin, Xiaogang Wang

(* indicates equal contribution)
The Chinese University of Hong Kong, SenseTime Group Limited

NUIST:
Jing Yang, Hui Shuai, Zhengbo Yu, Rongrong Fan, Qiang Ma, Qingshan Liu, Jiankang Deng

ILSVRC2016 VID tracking results – with "external" data

Team Name	Mean Average Precision(%)
NUIST	58.4
ITLab-Inha	49.1

NUIST:
Jing Yang, Hui Shuai, Zhengbo Yu, Rongrong Fan, Qiang Ma, Qingshan Liu, Jiankang Deng

ITLab-Inha:
Byungjae Lee[1], Songguo Jin[1], Enkhbayar Erdenee[1], Mi Young Nam[2], Young Giu Jung[2], Phill Kyu Rhee[1]

1. Inha University
2. NaeulTech

Object detection from video (VID)

Winner with “provided” and “external” data

NUIST:
Jing Yang, Hui Shuai,
Zhengbo Yu, Rongrong
Fan, Qiang Ma, Qingshan
Liu, Jiankang Deng

Object detection from video with tracking

Winner with "provided" data

CUvideo:
Hongsheng Li*, Kai Kang*, Wanli Ouyang, Junjie Yan, Tong Xiao, Xingyu Zeng, Kun Wang, Xihui Liu, Qi Chu, Junming Fan, Yucong Zhou, Yu Liu, Ruohui Wang, Shengen Yan, Dahua Lin, Xiaogang Wang
(* indicates equal contribution)
The Chinese University of Hong Kong, SenseTime Group Limited

Winner with "external" data

NUIST:
Jing Yang, Hui Shuai, Zhengbo Yu, Rongrong Fan, Qiang Ma, Qingshan Liu, Jiankang Deng