INFORMATION TECHNOLOGY LABORATORY

ActEV18: Activities in Extended Video

PIs: Afzal Godil, **Jonathan Fiscus**
Yooyoung Lee, David Joy, Andrew Delgado

TRECVID 2018 Workshop

November 13-15, 2018

IARPA BE THE FUTURE

NIST National Institute of Standards and Technology U.S. Department of Commerce

DIVA

Disclaimer

Certain commercial equipment, instruments, software, or materials are identified in this paper to specify the experimental procedure adequately. Such identification is not intended to imply recommendation or endorsement by NIST, nor necessarily the best available for the purpose.

The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, NIST, or the U.S. Government.

Outline

- ActEV Overview
- Evaluation Framework
- Tasks and Measures
- ActEV18 Evaluations
- ActEV18 Dataset
- ActEV18 Results and Analyses
- Next Steps

ActEV Overview

What is ActEV?

- ActEV (Activities in Extended Video) is an extension of TRECVID Surveillance Event Detection (SED) evaluations
- Goal
  - To advance video analytics technology that can automatically detect a target activity and identify and track objects associated with the activity.
- A series of challenges are also designed for:
  - Activity detection in a multi-camera environment
  - Temporal (and spatio-temporal) localization of the activity for reasoning

What's New? (SED -> ActEV)

- New activity-annotated and unannotated data for 4 years!
  - DARPA Video and Image Retrieval and Analysis Tool (VIRAT) data (16, 28 hrs)
  - Newly-collected DIVA data (Rough est. ~200 hrs, ~20K hrs)
- New evaluation tasks
  - Activity Detection (AD) : similar to the retrospective SED task
  - Activity and Object Detection (AOD): activity + object detection
  - Activity and Object Detection and Tracking (AODT): activity + object detection + tracking
- A series of evaluations rather than one per year
  - Blind: participants deliver system output (typical TRECVID)
  - Leader board: participants deliver many system output
  - Independent: participants deliver working systems for NIST to test on sequestered data

NIST, IARPA, and Kitware

- NIST developed the ActEV evaluation series to support the metrology needs of the Intelligence Advanced Research Projects Activity (IARPA) Deep Intermodal Video Analytics (DIVA) Program
- The ActEV's datasets collected and annotated by Kitware, Inc.

NIST
National Institute of Standards and Technology
U.S. Department of Commerce

Evaluation Framework

Evaluation Framework

- Target applications
  - Retrospective analysis of archives (e.g., forensic analytics)
  - Real-time analysis of live video streams (e.g., alerting)
- Evaluation Type
  - Self-reported evaluation
  - Independent (& sequestered) evaluation
- Evaluation conditions
  - Activity-level (1.A phase evaluation)
  - Reference temporal segmentation
  - Leaderboard

Tasks and Measures
(AD, AOD, AODT)

Evaluation Tasks (AD)

- Activity Detection (AD)
  - Given a target activity, a system automatically 1) detects its presence and then temporally localizes all instances of the activity in video sequences
  - The system output includes:
    - Start and end frames indicating the temporal location of the target activity
    - A presence confidence score that indicates how likely the activity occurred

Evaluation Tasks (AOD)

- Activity and Object Detection (AOD)
  - A system not only 1) detects/localizes the target activity, but also 2) detects the presence of required objects and spatially localizes the objects that are associated with the activity
  - The system output includes:
    - Start and end frames indicating the temporal location of the target activity
    - A presence confidence score that indicates how likely the activity occurred
    - Coordinates of object bounding boxes and object presence confidence scores
  - Scoring protocol: AOD_AD and AOD_AOD.

Evaluation Tasks (AODT)

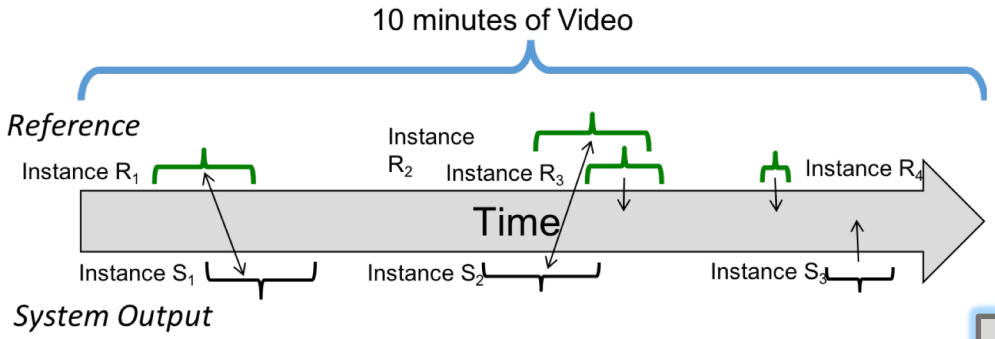
- Activity Object Detection/Tracking (AODT)
  - A system 1) correctly detects/localizes the target activity, 2) correctly detects/localizes the required objects in that activity, and 3) correctly tracks those objects over time.

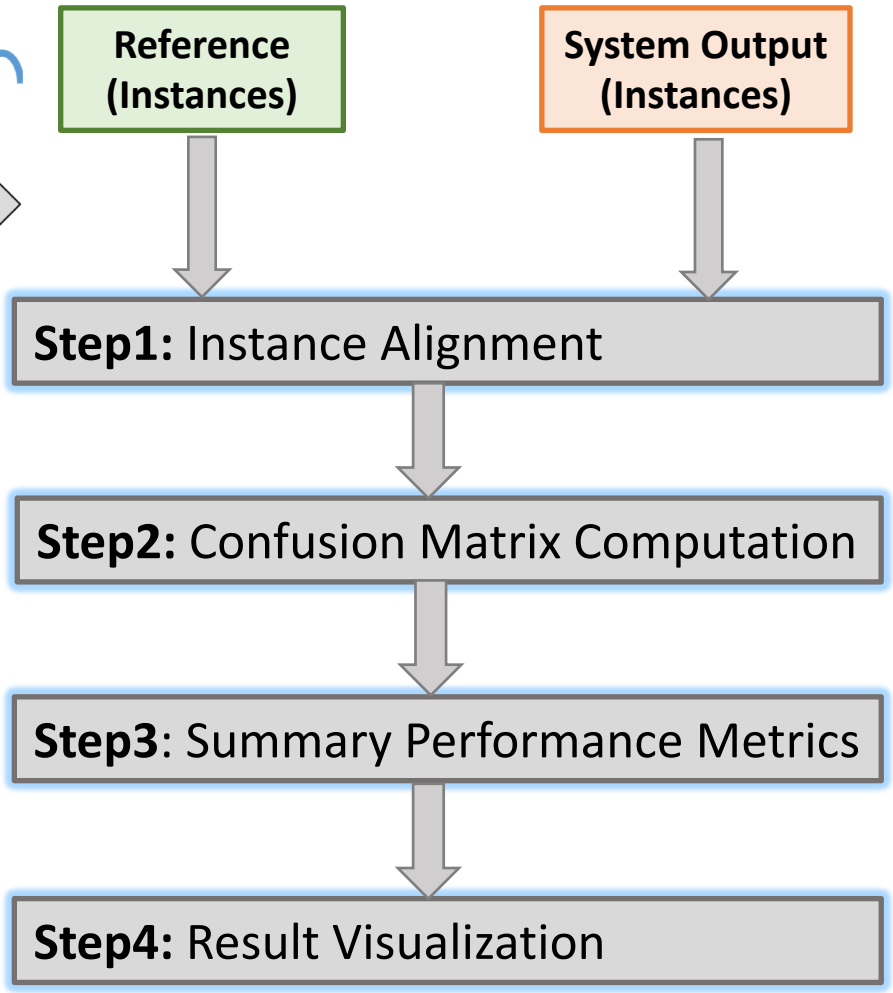
  - The AODT task is NOT addressed in ActEV18 evaluations

Performance Measures (AD)

- Primary metrics
  - J. Fiscus, “TRECVID Surveillance Event Detection Evaluation.” https://www.nist.gov/itl/iad/mig/trecvid-2017-evaluation-surveillance-event-detection
- Secondary metrics
  - K. Bernardin and R. Stiefelhagen, “Evaluating Multiple Object Tracking Performance: The CLEAR MOT Metrics,” *EURASIP J. Image Video Process.*, vol. 2008

Primary: Activity Occurrence Detection



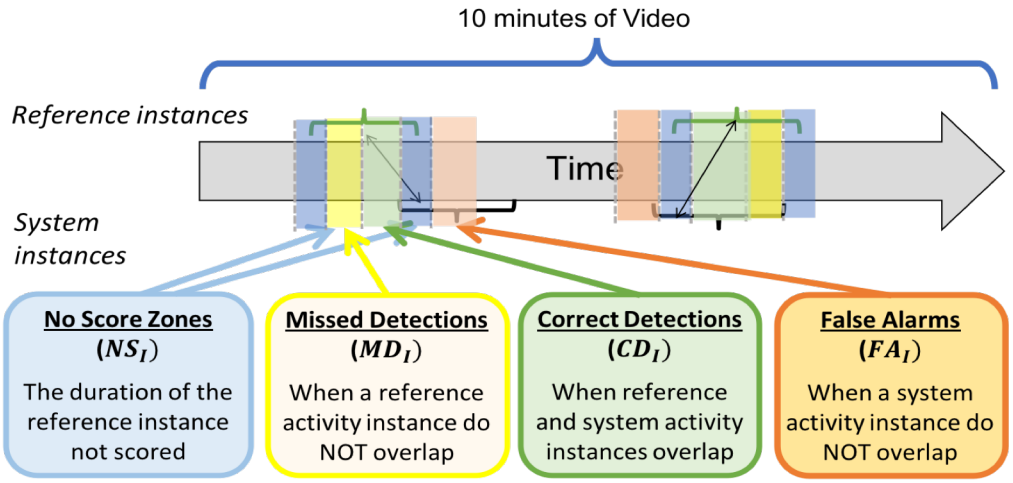


$$P_{miss}(\tau) = \frac{N_{MD}(\tau)}{N_{TrueInstance}}$$

$$R_{FA}(\tau) = \frac{N_{FA}(\tau)}{VideoDurInMinutes}$$

- P_{miss} at $R_{FA} = 0.15$
- Further details in "ActEV 2018 Evaluation Plan", https://actev.nist.gov/

Secondary: Temporal Localization



- N_MIDE (Normalized Multiple Instance Detection Error)

$$N_{MIDE} = \sum_{I=1}^{N_{mapped}} \frac{(C_{MD} * \frac{MD_I}{MD_I + CD_I} + C_{FA} * \frac{FA_I}{Dur_V - (MD_I + CD_I + NS_I)})}{N_{mapped}}$$

Further detail in "ActEV 2018 Evaluation Plan", https://actev.nist.gov/

Performance Measures (AOD)

- Primary
  - Similar to AD, however, instance alignment step uses an additional term for the object detection congruence
- Secondary
  - N_MODE (Normalized Multiple Object Detection Error)

$$N_{MODE(\tau)} = \sum_{t=1}^{N_{frames}} \frac{\left(C_{MD} * MD_t(\tau) + C_{FA} * FA_t(\tau)\right)}{\sum_{t=1}^{N_{frames}} N_R^t}$$

  - The minimum N_MODE value (minMODE) is calculated for object detection performance
  - 1-minMODE is used for the object detection congruence term

Performance Measures (AODT)

- Primary
  - Similar to AD, however, instance alignment step uses an additional term for the object tracking congruence
- Secondary
  - MOTE (Multiple Object Tracking Error)

$$MOTE(\tau) = \sum_{t=1}^{N_{frames}} \frac{(C_{MD} * MD_t(\tau) + C_{FA} * FA_t(\tau) + C_{ID} * \boldsymbol{IDSwitchs_t}(\tau)}{\sum_{t=1}^{N_{frames}} N_R^t}$$

  - The minimum MOTE value (minMOTE) is calculated for object tracking performance
  - 1-minMOTE is used for the tracking congruence term

ActEV18 Evaluations

ActEV18 Evaluations are focusing on

- The AD and AOD tasks only
- Retrospective analysis applications in mind
- The single camera view and at the activity observation level
- Self-reported evaluation only
- A series of the evaluations:
  - Activity-level
  - Reference temporal segmentation (RefSeg)
  - Leaderboard

ActEV18 Dataset

Activities and Number of Instances

VIRAT V1 dataset

12 activities for activity-level/RefSeg

Activity Type	Train	Validation
Closing	126	132
Closing_trunk	31	21
Entering	70	71
Exiting	72	65
Loading	38	37
Open_Trunk	35	22
Opening	125	127
Transport_HeavyCarry	45	31
Unloading	44	32
Vehicle_turning_left	152	133
Vehicle_turning_right	165	137
Vehicle_u_turn	13	8

Additional 7 activities for leaderboard

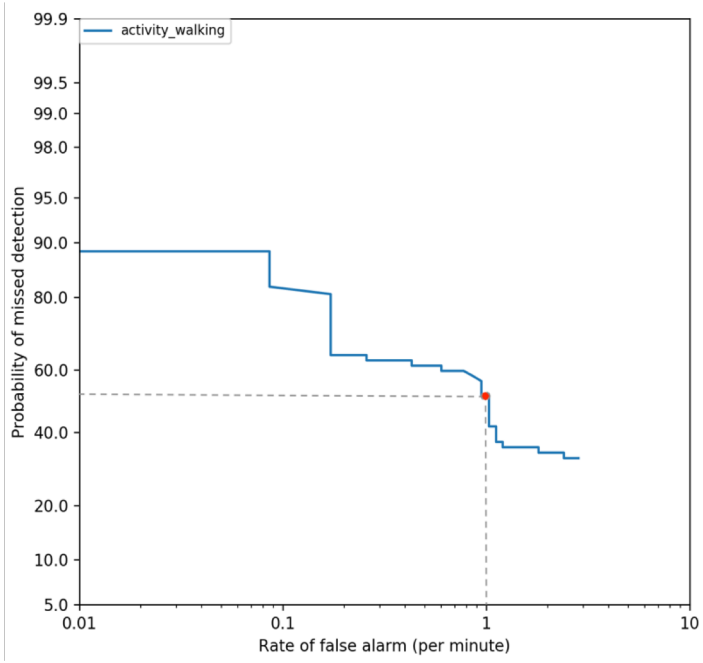
Activity Type	Train	Validation
Interacts	88	101
Pull	21	22
Riding	21	22
Talking	67	41
Activity_carrying	364	237
Specialized_talking_phone	16	17
Specialized_texting_phone	20	5

Due to ongoing evaluations, the test sets are not included in the table

ActEV18
Results and Analyses

ActEV18 Activity-Level Evaluation

- 15 Participants from the academic and industrial sectors
- AD
  - 20 systems from 13 teams (including baseline)
  - Activity Detection (Primary):
    - P_{miss} at $R_{FA} = 0.15$, P_{miss} at $R_{FA} = 1$
  - Temporal Localization (Secondary):
    - N_{MIDE} at $R_{FA} = 0.15$, N_{MIDE} at $R_{FA} = 1$
- AOD
  - 16 systems from 11 teams
  - Two scoring protocols
    - AOD_AD: the same with the AD task
    - AOD_AOD: In addition to the AD metrics, $\mu ObjectP_{miss}$ at $R_{FA} = 0.5$ is used for object detection



Detection Error Tradeoff (DET) curve

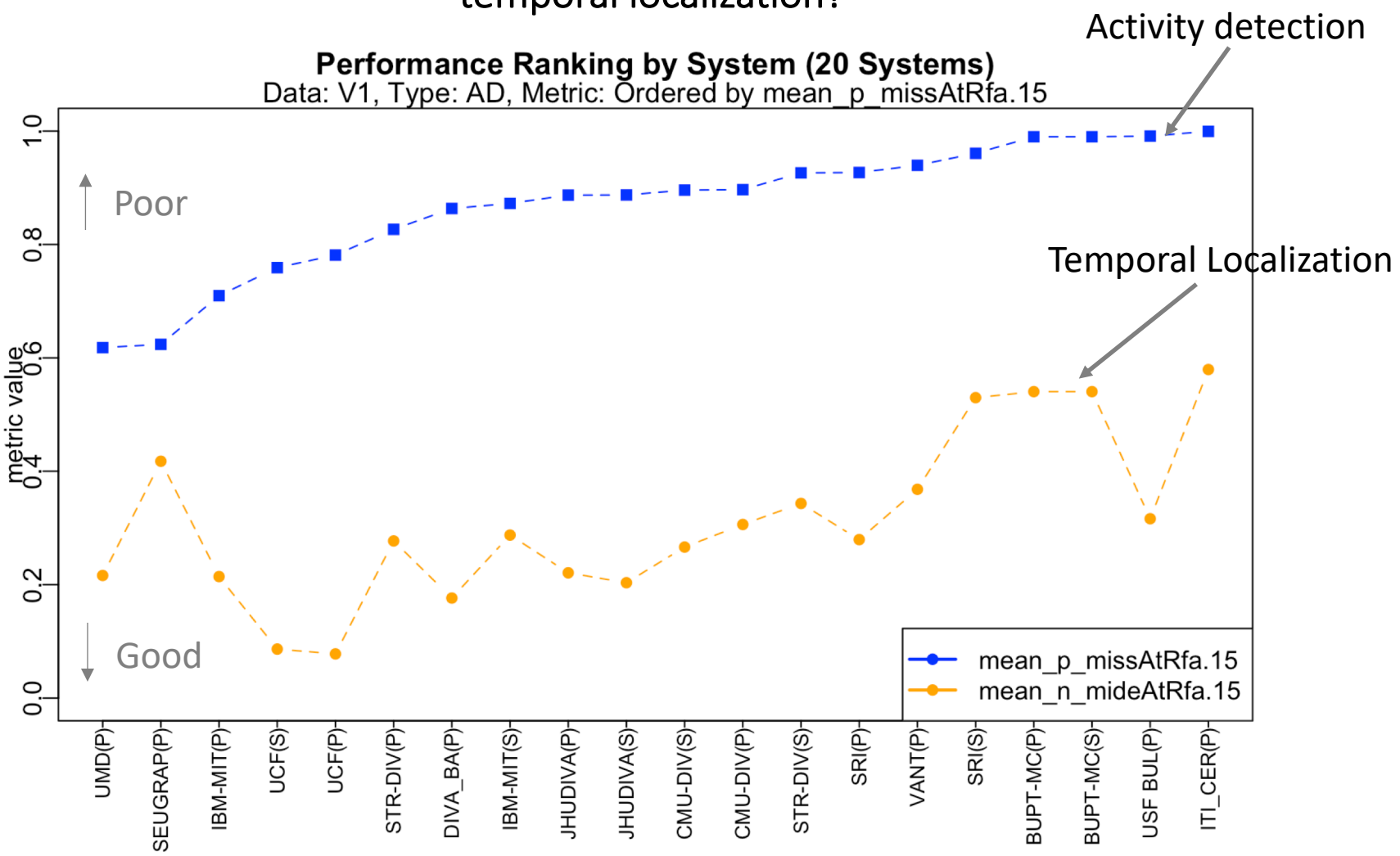
ActEV18 activity-level evaluation results

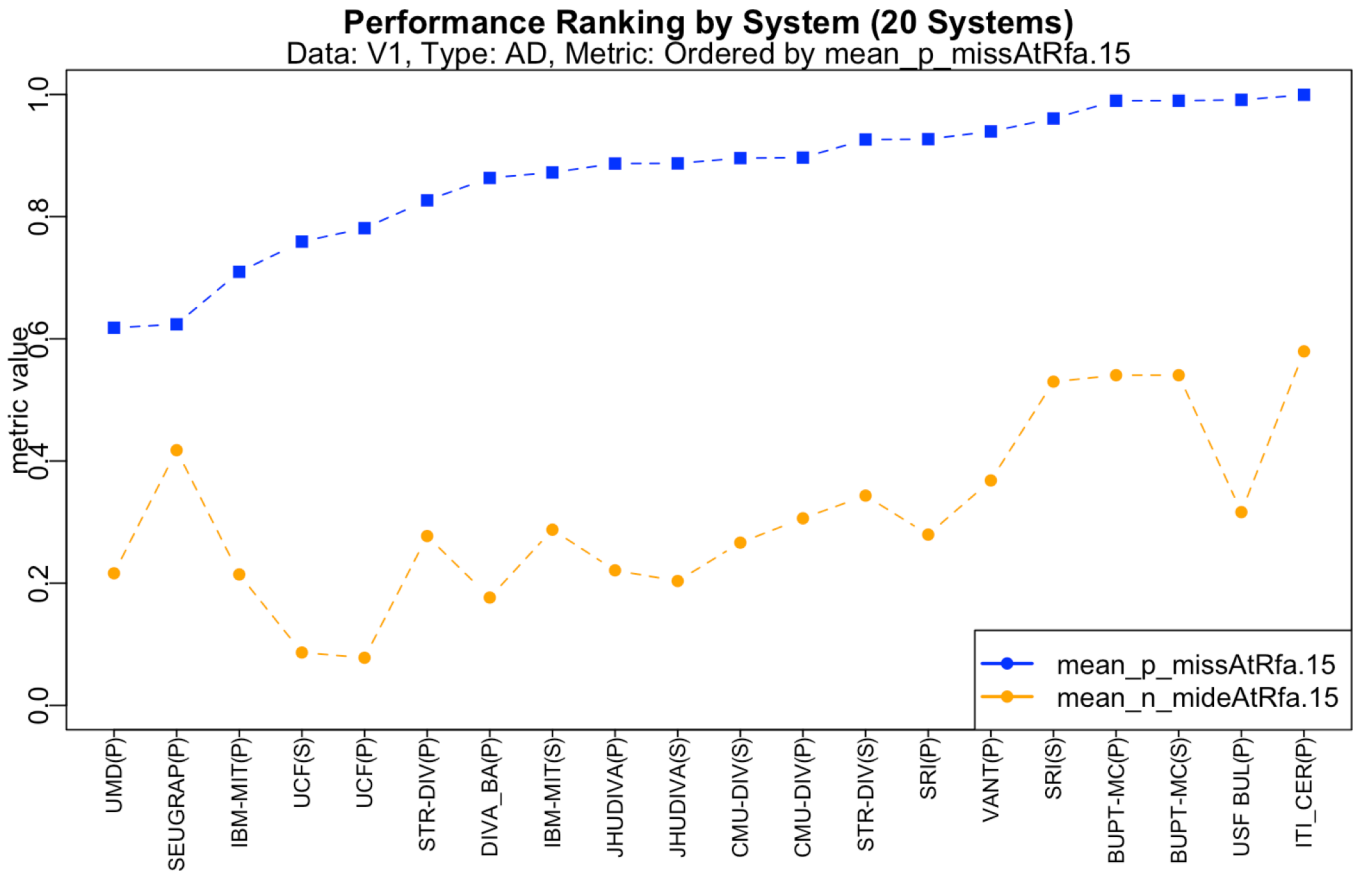
P: Primary, S: Secondary, PR.15: μP_{miss} at $R_{FA} = 0.15$, NR.15: μN_{MIDE} at $R_{FA} = 0.15$,
PR1: μP_{miss} at $R_{FA} = 1$, NR1: μN_{MIDE} at $R_{FA} = 1$, OPR.5: $\mu ObjectP_{miss}$ at $R_{FA} = 0.5$

System and Version		AD				AOD		
						AOD_AD	AOD_AOD	
		PR.15↓	PR1↓	NR.15↓	NR1↓	PR.15↓	PR.15↓	OPR.5↓
UMD	P	0.618	0.441	0.216	0.223	0.618	0.680	0.306
SeuGraph	P	0.624	0.621	0.418	0.416	0.624	0.664	0.362
IBM-MIT-Purdue	P	0.710	0.603	0.214	0.230	0.710	0.726	0.110
UCF	S	0.759	0.624	0.086	0.129	n/a	n/a	n/a
UCF	P	0.781	0.654	0.078	0.112	n/a	n/a	n/a
STR-DIVA Team	P	0.827	0.722	0.277	0.321	0.827	0.838	0.443
DIVA_Baseline	P	0.863	0.720	0.176	0.196	n/a	n/a	n/a
IBM-MIT-Purdue	S	0.872	0.704	0.288	0.282	0.872	0.878	0.329
JHUDIVATeam	P	0.887	0.829	0.221	0.219	0.887	0.933	0.266
JHUDIVATeam	S	0.887	0.813	0.203	0.240	0.887	0.926	0.332
CMU-DIVA	S	0.896	0.831	0.266	0.317	0.896	0.904	0.421
CMU-DIVA	P	0.897	0.766	0.306	0.349	0.897	0.908	0.244
STR-DIVA Team	S	0.926	0.905	0.343	0.355	n/a	n/a	n/a
SRI	P	0.927	0.856	0.279	0.282	0.927	0.936	0.406
VANT	P	0.940	0.918	0.368	0.385	0.940	0.945	0.837
SRI	S	0.961	0.885	0.530	0.490	0.961	0.963	0.446
BUPT-MCPRL	P	0.990	0.839	0.540	0.248	0.990	1.000	0.669
BUPT-MCPRL	S	0.990	0.839	0.540	0.248	0.990	1.000	0.669
USF Bulls	P	0.991	0.949	0.316	0.375	n/a	n/a	n/a
ITI_CERTH	P	0.999	0.998	0.579	0.667	0.999	0.999	0.955
HSMW_TUC	P	n/a	n/a	n/a	n/a	0.961	0.968	0.502

Performance Ranking (AD)

What is the general trend on performance between activity detection and temporal localization?

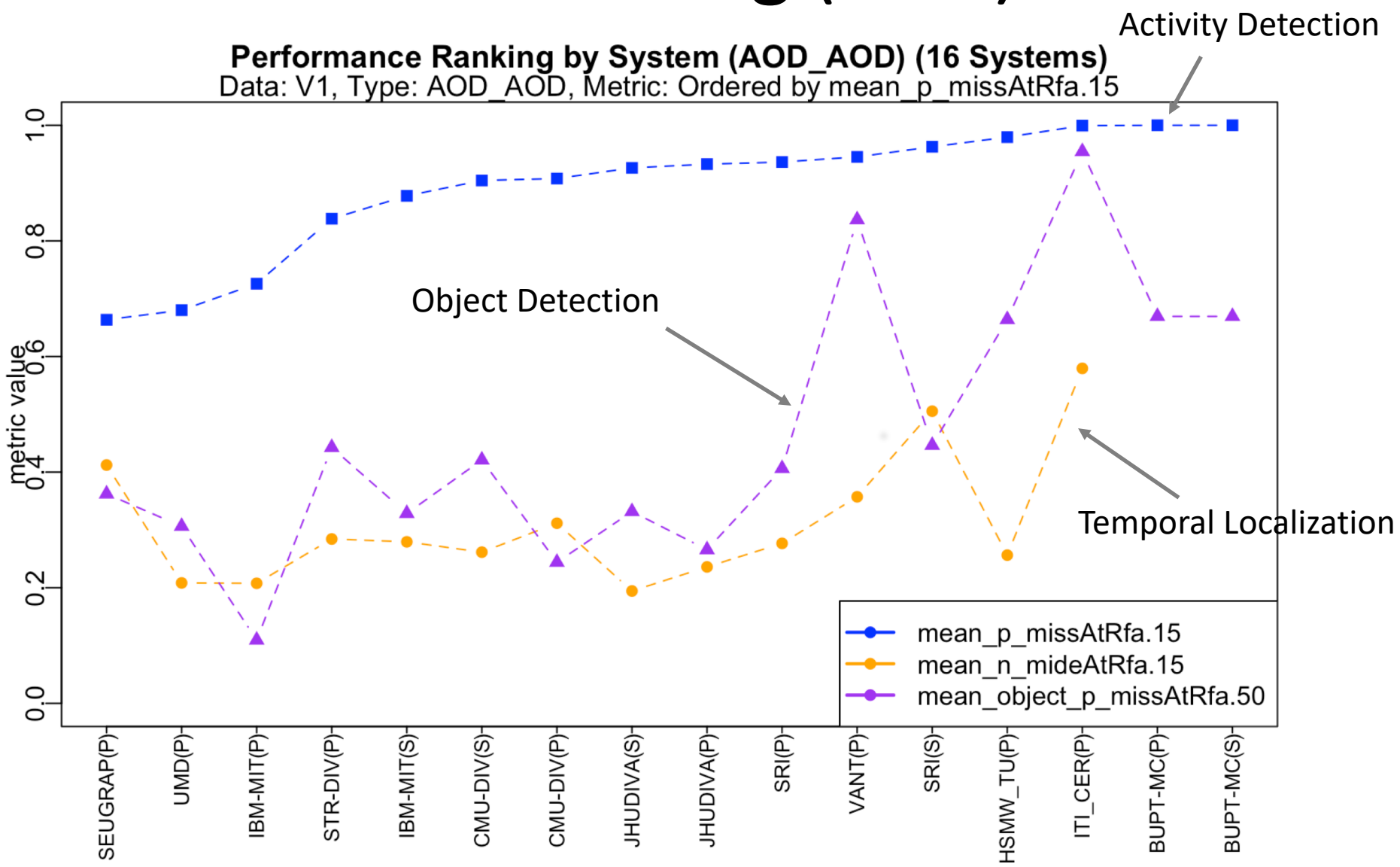


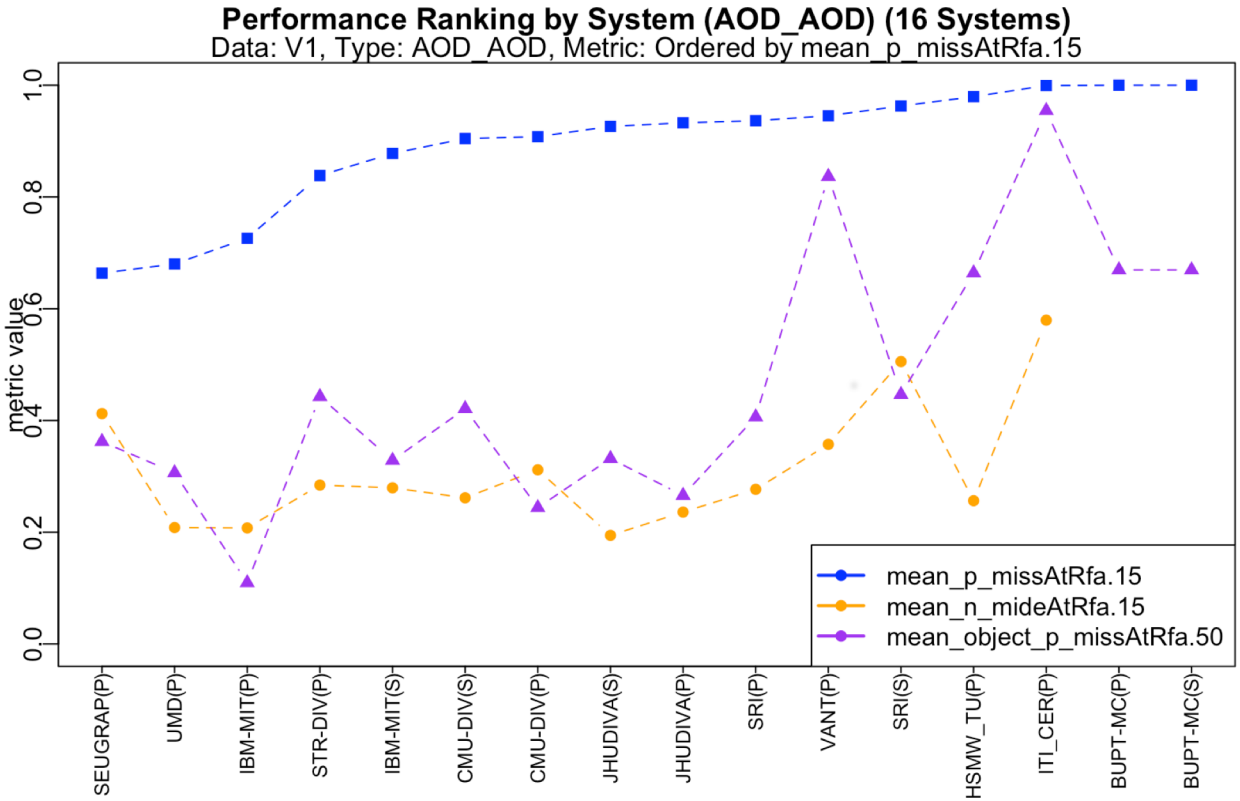


Observation

- Highest performance on activity detection:
  - UMD (PR.15: 61.8%) followed by SeuGraph (PR.15: 62.4%)
- Highest performance on temporal localization
  - UCF (NR.15: 7.8%)
- Different trend between activity detection and temporal localization

Performance Ranking (AOD)

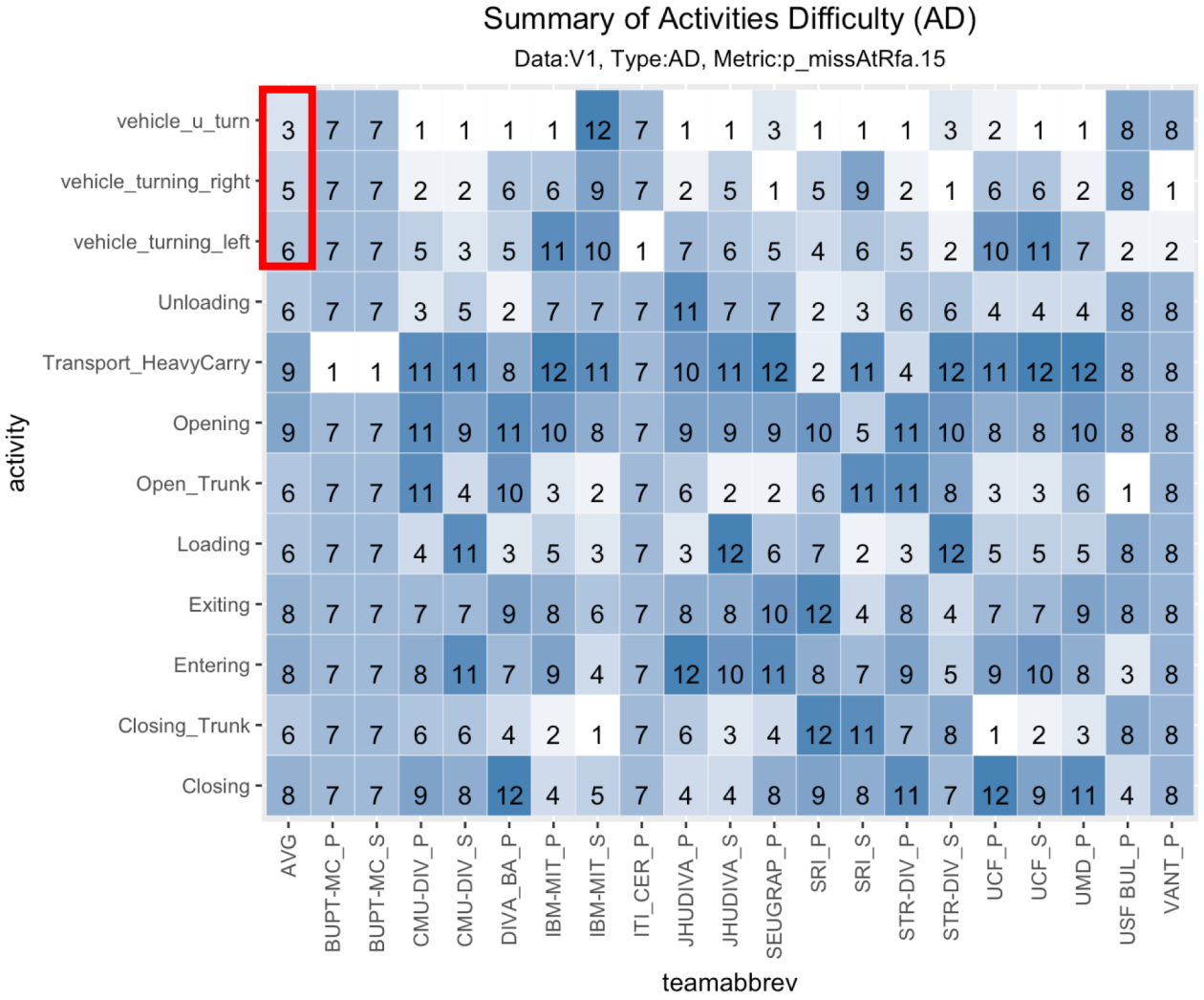




Observation

- Highest performance on activity detection:
  - SeuGraph (PR.15: 66.4%), UMD (PR.15: 68%)
- Highest performance on temporal localization
  - JHU (NR.15: 19.4%), IBM_MIT_PURDUE (20.7%) , UMD (20.8%)
- Highest performance on object detection
  - IBM_MIT_PURDUE (OPR.5: 11%)
- Different trend among activity detection, temporal localization, and object detection

Which activities are easier or more difficult to detect?



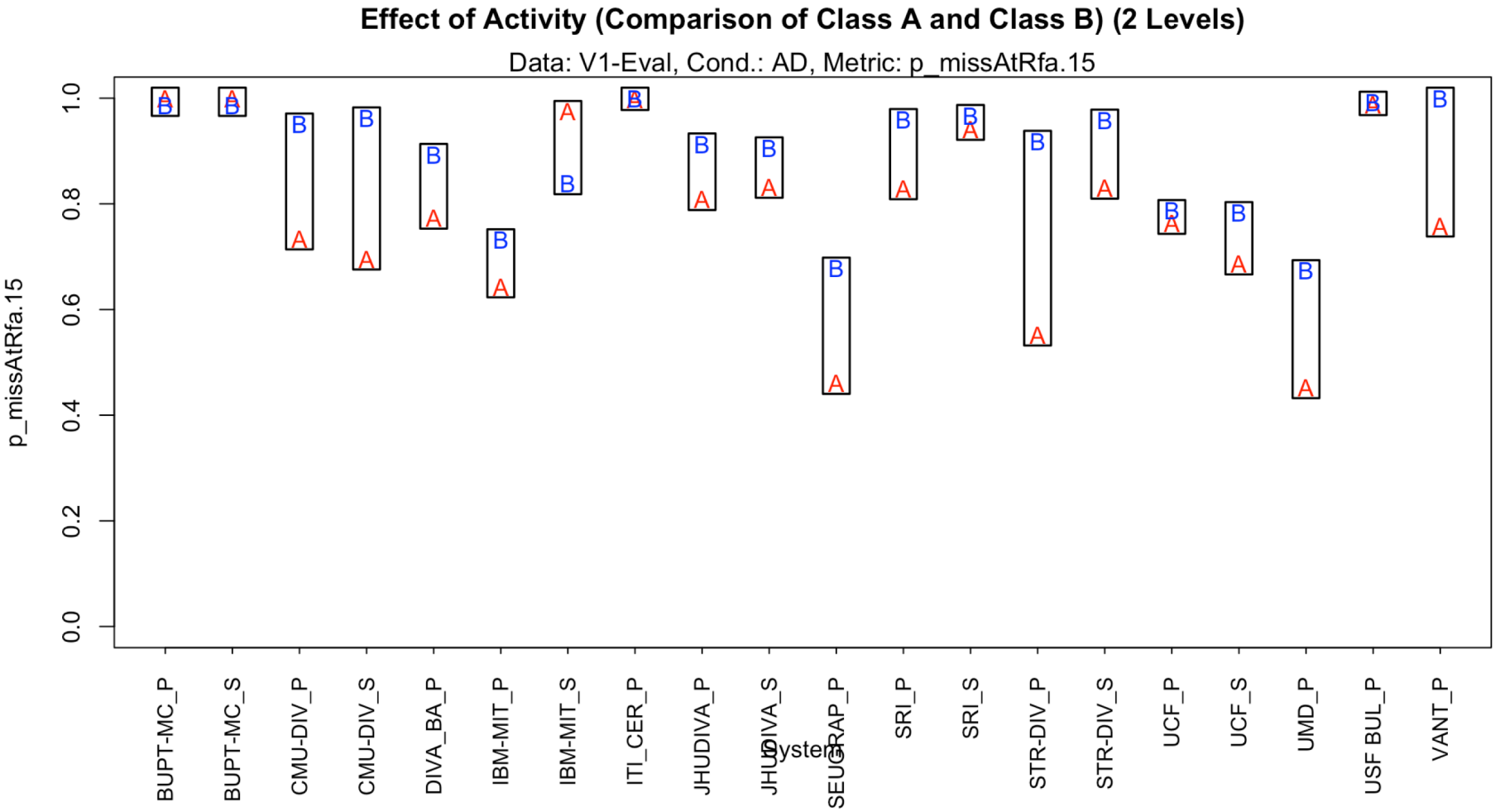
- X-axis: systems ordered by name
- Y-axis:12 activities and average activity ranking (AVG)
- Numbers in the matrix: the ranking of 12 activities per system

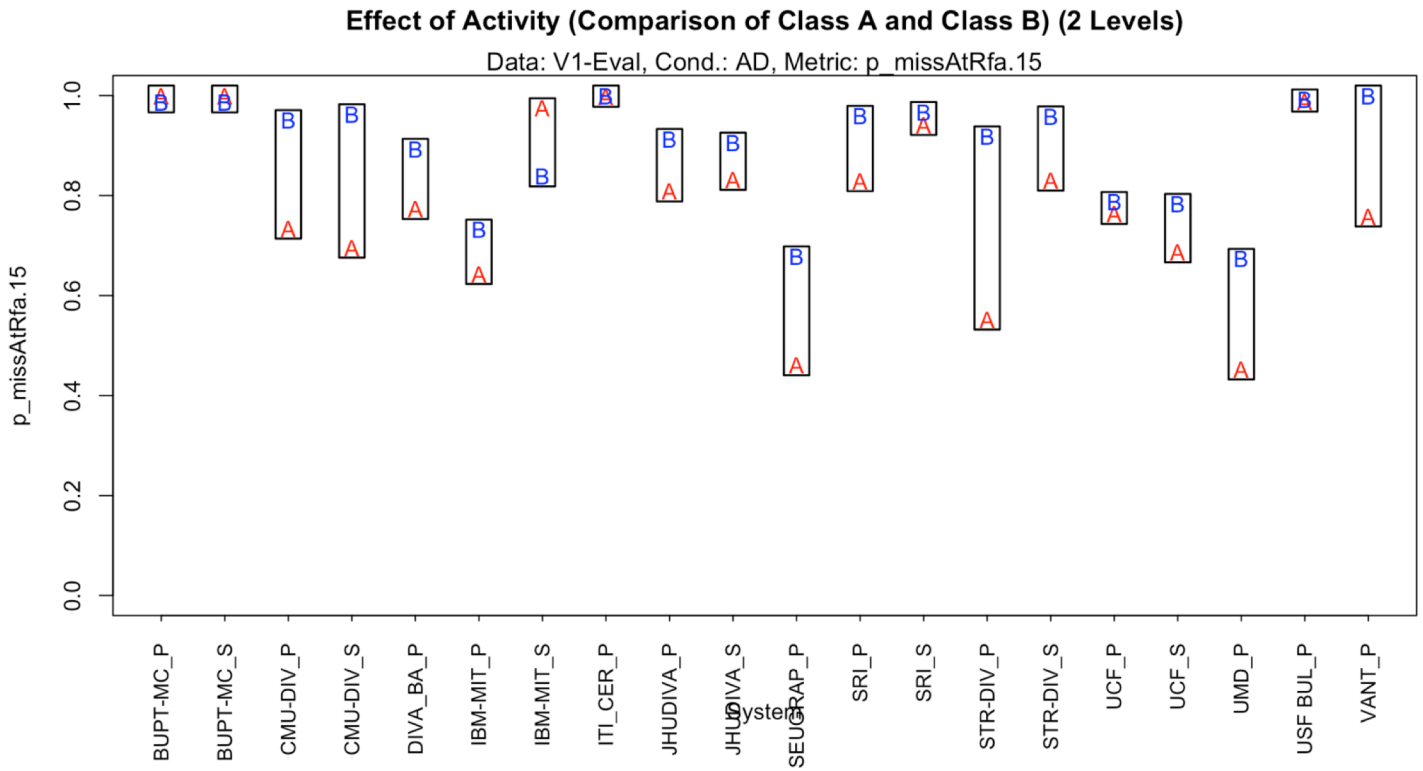
The activity class was characterized by systems and baseline performance

Observation: the vehicle-turn related activities are easier to detect compared to the rest of the other activities

How does the activity class behave per system?

Class A: Vehicle-turn related activities, **Class B**: the rest of the other activities





Observation

1. How does the activity class behave per system?

- In general, the class A activities are easier to detect

2. Robustness?

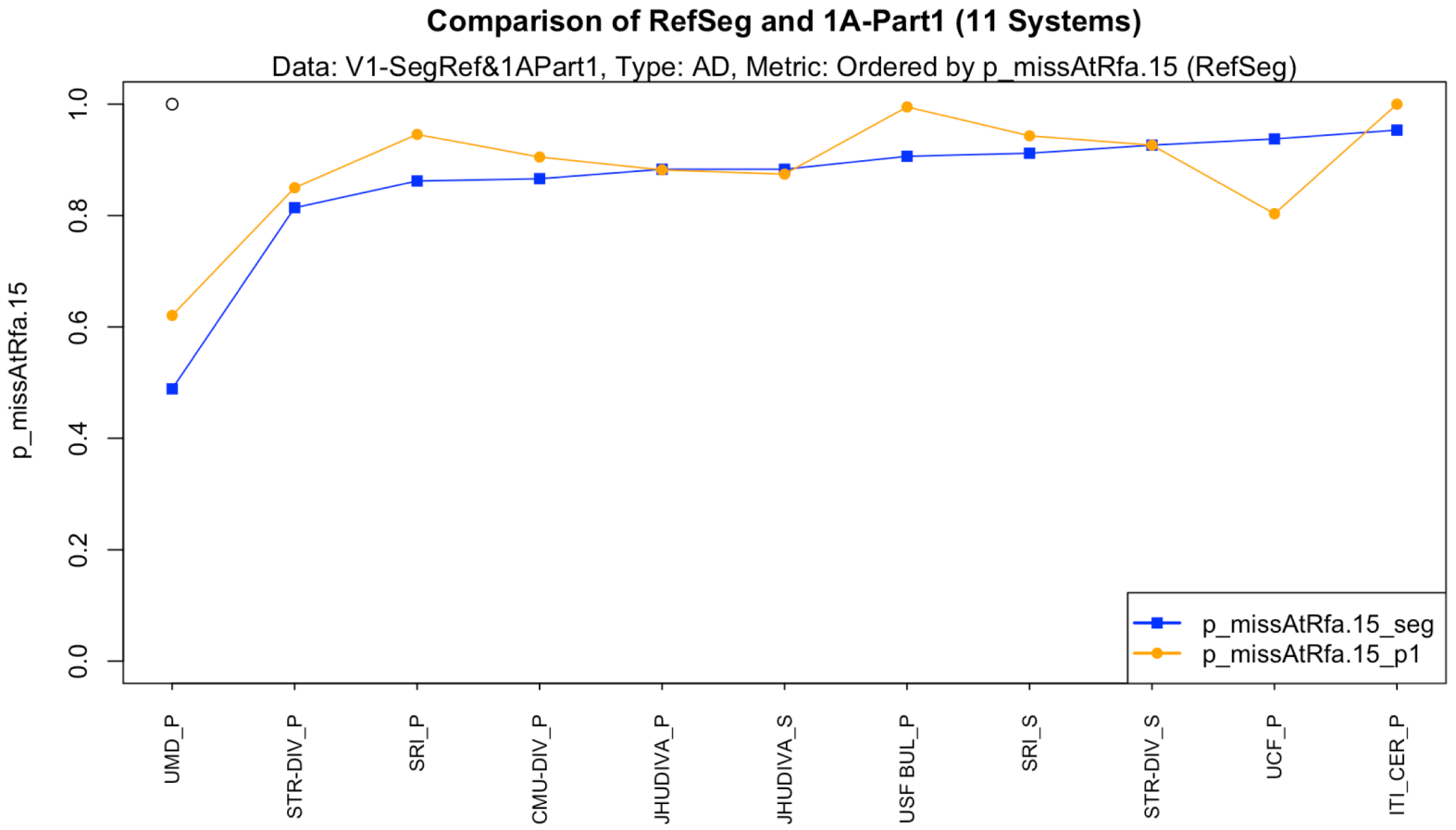
- The conclusion is consistent across systems with a few exception (e.g., IBM_MIT_Purdue)

3. Effect comparison?

- STR and CMU have larger effect on the activity class

Comparison of RefSeg and EvalPart1 (AD)

RefSeg: the systems were scored on the reference temporal segment test set
EvalPart1: the systems submitted for the activity-level evaluation were scored on the same test set

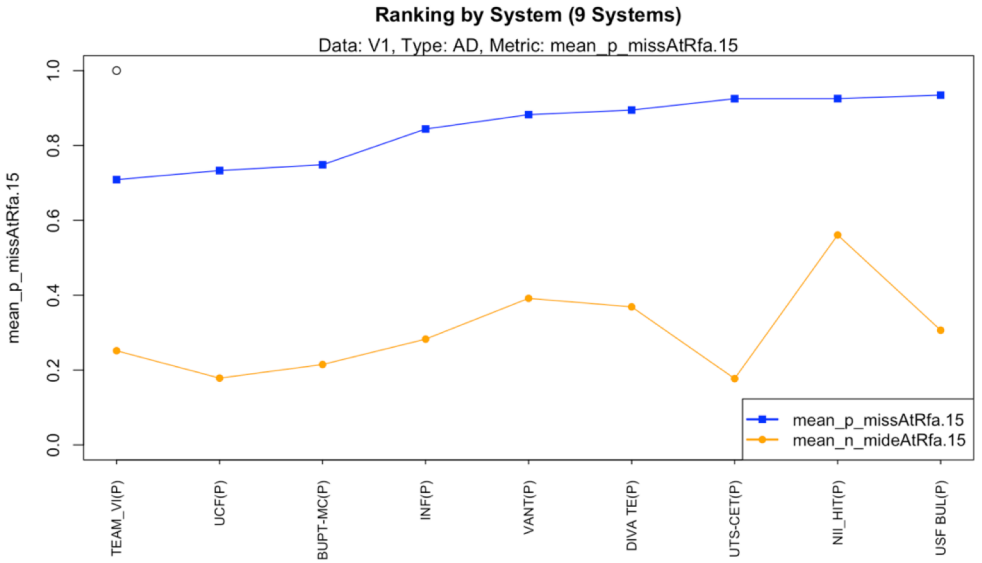


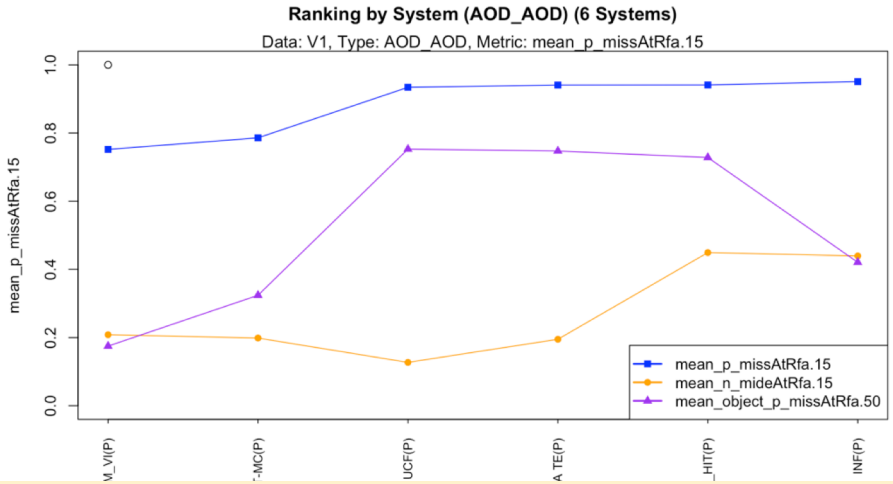
Observation: with a few exceptions, system performance with reference segment info is better than system performance without

Leaderboard (as of 11/08/18)

Teams	AD	
	PR.15	NR.15
Team_Vision	0.709	0.252
UCF	0.733	0.179
BUPT-MCPRL	0.749	0.215
INF	0.844	0.283
VANT	0.882	0.392
DIVA Baseline	0.895	0.369
UTS-CETC	0.925	0.177
NII_Hitachi_UIT	0.925	0.561
USF Bulls	0.934	0.306

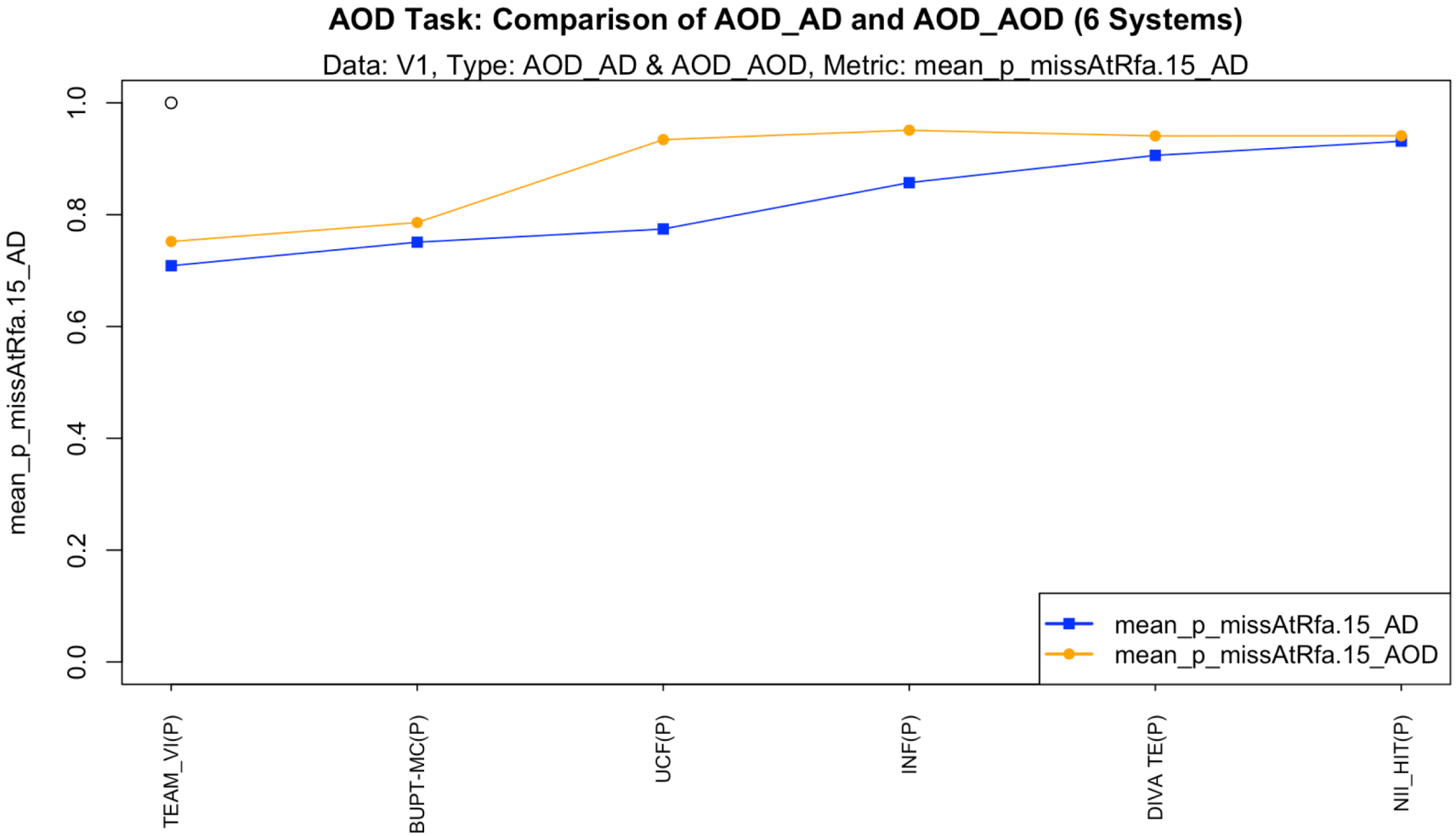
Teams	AOD		
	AOD_AD	AOD_AOD	
	PR.15	PR.15	OPR.5
Team_Vision	0.709	0.752	0.175
BUPT-MCPRL	0.751	0.786	0.324
UCF	0.774	0.934	0.753
DIVA Baseline	0.906	0.941	0.747
NII_Hitachi_UIT	0.931	0.941	0.728
INF	0.857	0.951	0.421





Observation: Team-Vision (IBM-MIT-Purdue) team achieved the highest performance on AD and AOD

How does activity detection behave when object detection was taken into account?



Observation: when the object detection was taken into account, the AOD_AOD performance under-performs compared to AOD_AD

Next Steps

Next Steps

- ActEV18 next phase evaluation incudes AODT (on VIRAT V1/V2 dataset)—ongoing
- 50K ActEV-PC (IARPA Activity in Extended Videos Prize Challenge)--ongoing https://actev.nist.gov/prizechallenge
- ActivityNet workshop under CVPR19
- New datasets (M1/M2) are coming soon

Questions?

https://actev.nist.gov/

Contact: actev-nist@nist.gov