The PASCAL Visual Object Classes Challenge 2011 (VOC2011)

Part 5 – Action Classification Taster Challenge

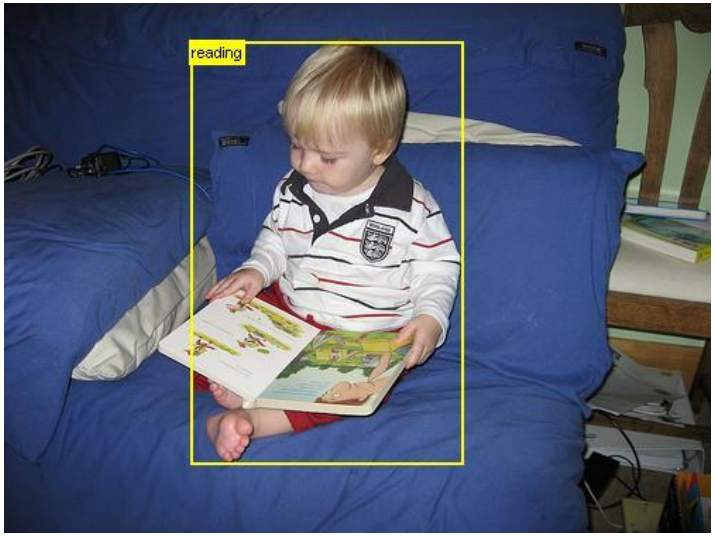
Mark Everingham
Luc Van Gool
Chris Williams
John Winn
Andrew Zisserman

PASCAL2
Pattern Analysis, Statistical Modelling and Computational Learning

Action Classification Taster Challenge

- Given the bounding box of a person, predict whether they are performing a given action

Playing Instrument?

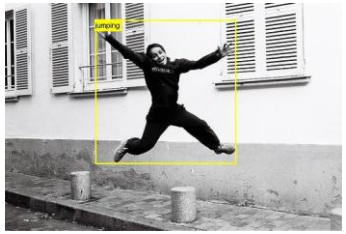


Reading?

- Encourage research on **still-image** activity recognition: more detailed understanding of image

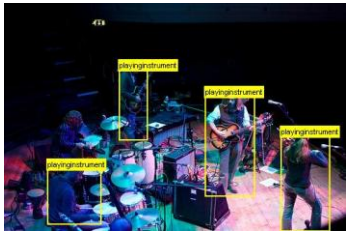
Ten Action Classes + **“Other”**

Jumping

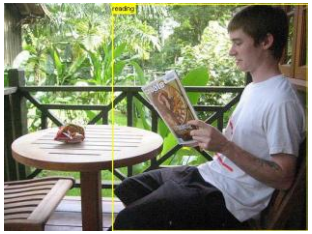


Phoning

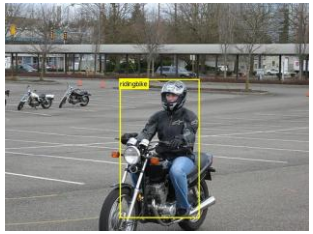
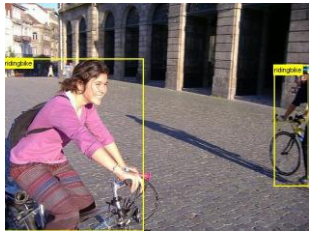
Playing Instrument



Reading



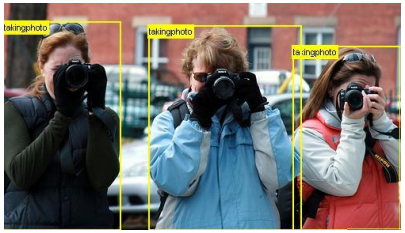
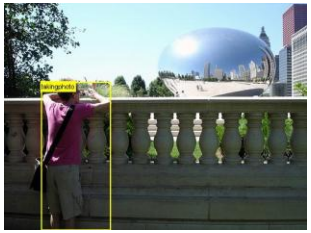
Riding Bike

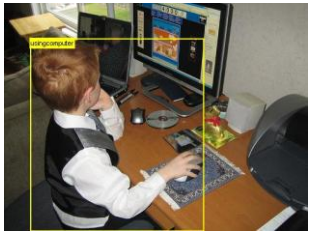
Riding Horse

Running

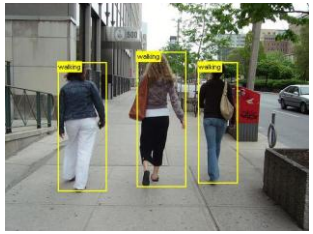
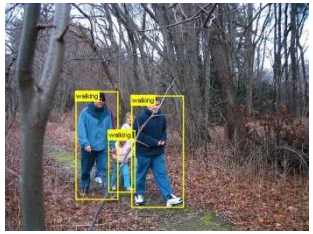
Taking Photo

Using Computer



Walking

Dataset Statistics

- ~5 times size of VOC2010 dataset

	Training		Testing	
Actions	2,772	(608)	2,788	(613)
"Other"	575	(0)	575	(0)

VOC2010 counts shown in brackets

- Minimum ~200 people per action category
- Only subset of people are annotated (bounding box + actions)
- Actions are not mutually-exclusive

Methods

- Train on VOC data: 7 Methods, 5 Groups
  - Image classification within bounding box
    - SVM, bag of words/spatial pyramid
    - Multiple features: SIFT, PHOG, color SIFT, etc.
    - Random Forest, dense spatial pyramid features
  - Multi-scale context (bounding box neighbourhood, image)
  - Discriminative Saliency Map weighting BoW features
  - Unsupervised learning of relevant object detector
- Train on own data: 3 Methods, 3 Groups
  - Poselets + image context
  - Attribute classifiers + object detectors (trained on ImageNet)
  - Poselets + foreground/background descriptors (segmentation)

AP by Class/Method

(1st, 2nd, 3rd place)

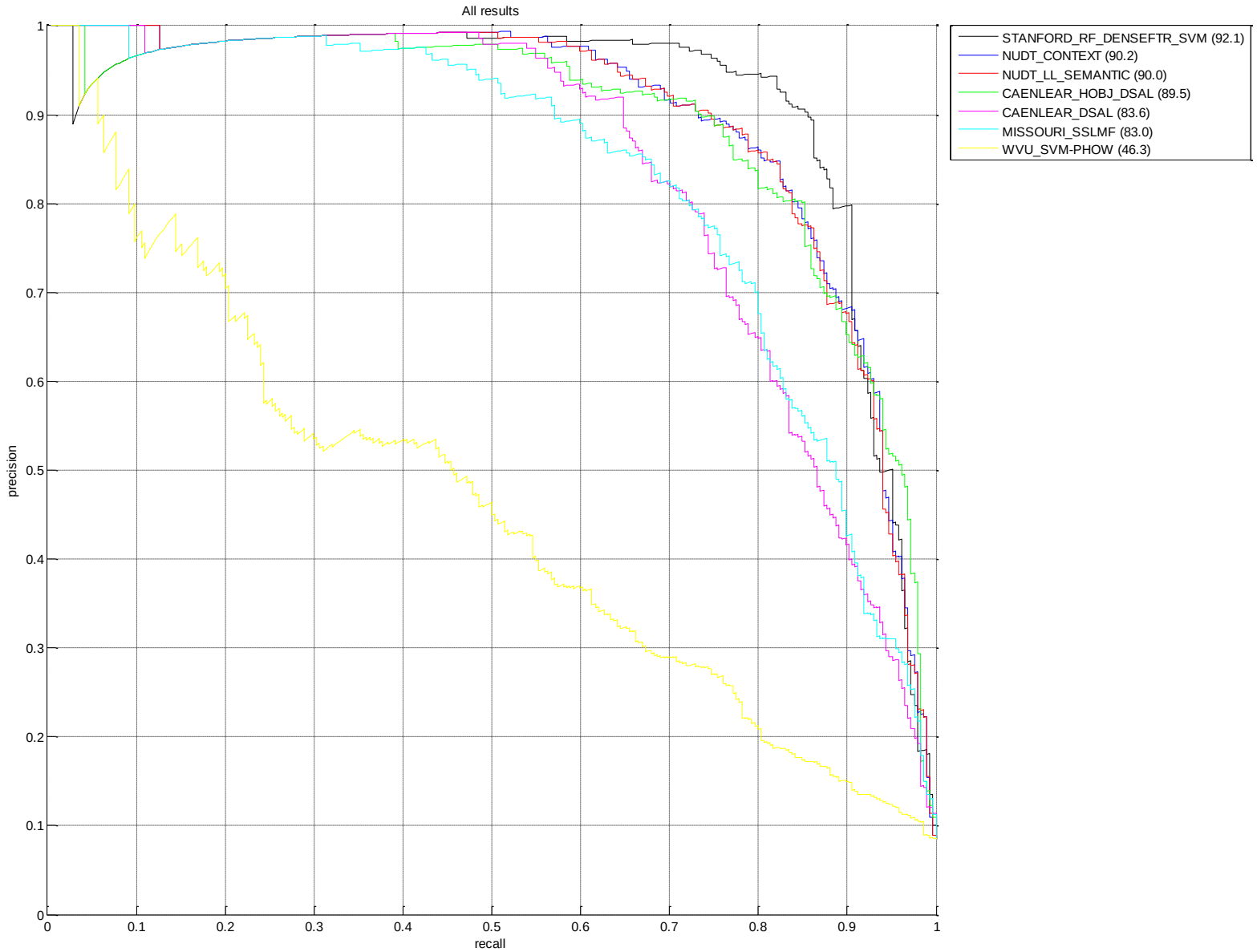
VOC data

	jumping	phoning	playing instrument	reading	riding bike	riding horse	running	taking photo	using computer	walking
CAENLEAR_DSAL	62.1	39.7	60.5	33.6	80.8	83.6	80.3	23.2	53.4	50.2
CAENLEAR_HOBJ_DSAL	71.6	50.7	77.5	37.8	86.5	89.5	83.8	25.1	58.9	59.2
MISSOURI_SSLMF	58.8	36.8	48.5	30.6	81.5	83.0	78.5	21.3	50.7	53.8
NUDT_CONTEXT	65.9	41.5	57.4	34.7	88.8	90.2	87.9	25.7	54.5	59.5
NUDT_LL_SEMANTIC	66.3	41.3	53.9	35.2	88.8	90.0	87.6	25.5	53.7	58.2
STANFORD_RF_DENSEFTR_SVM	66.0	41.0	60.0	41.5	90.0	92.1	86.6	28.8	62.0	65.9
WVU_SVM-PHOW	42.5	29.5	32.1	26.7	48.5	46.3	59.2	13.5	24.3	35.6

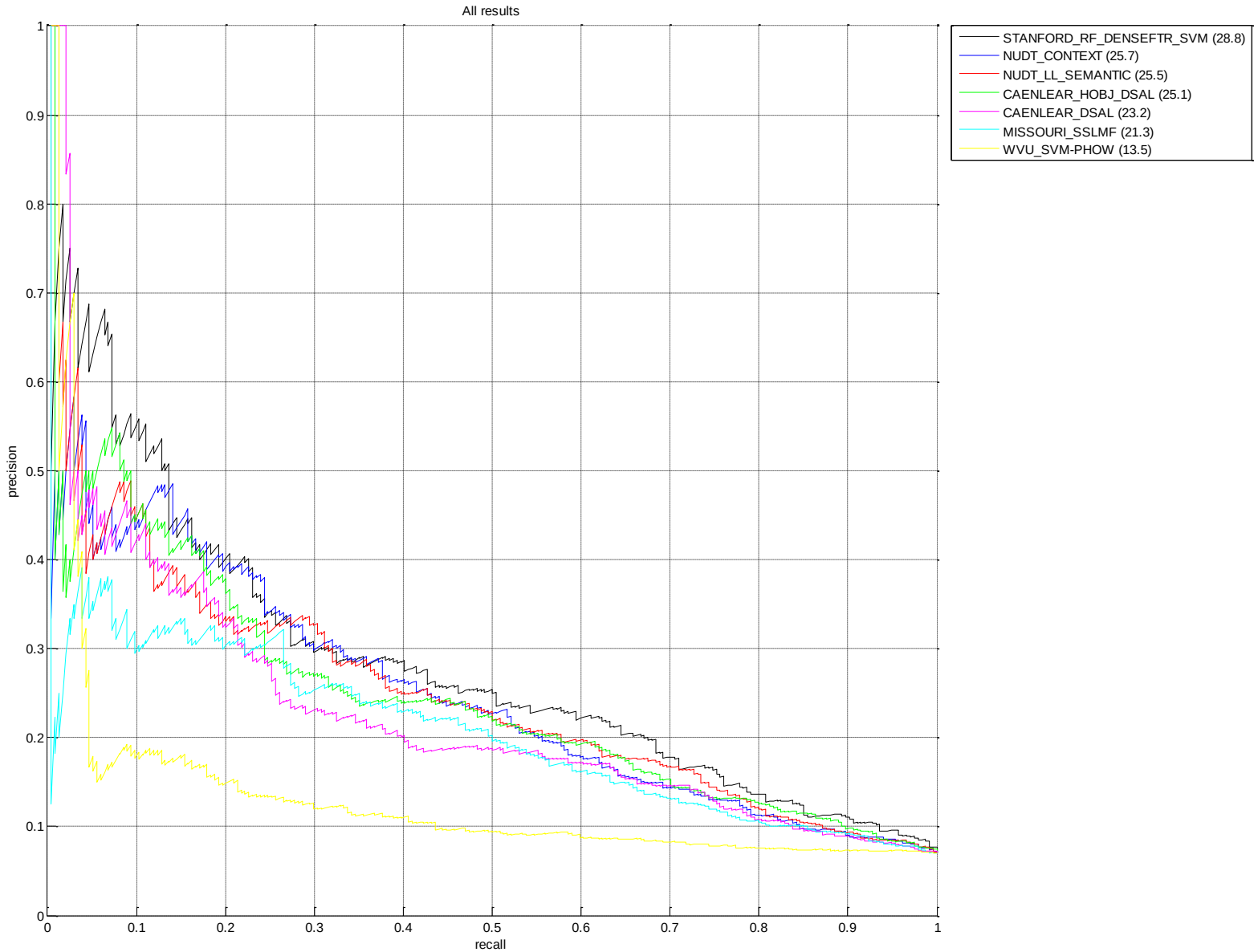
Own data

	jumping	phoning	playing instrument	reading	riding bike	riding horse	running	taking photo	using computer	walking
BERKELEY_ACTION_POSELETS	59.5	31.3	45.6	27.8	84.4	88.3	77.6	31	47.4	57.6
STANFORD_COMBINE_ATTR_PART	66.7	41.1	60.8	42.2	90.5	92.2	86.2	28.8	63.5	64.2
STANFORD_MAPSVM_POSELET	27	29.3	28.3	23.8	71.9	82.4	67.3	20.1	26	46.4

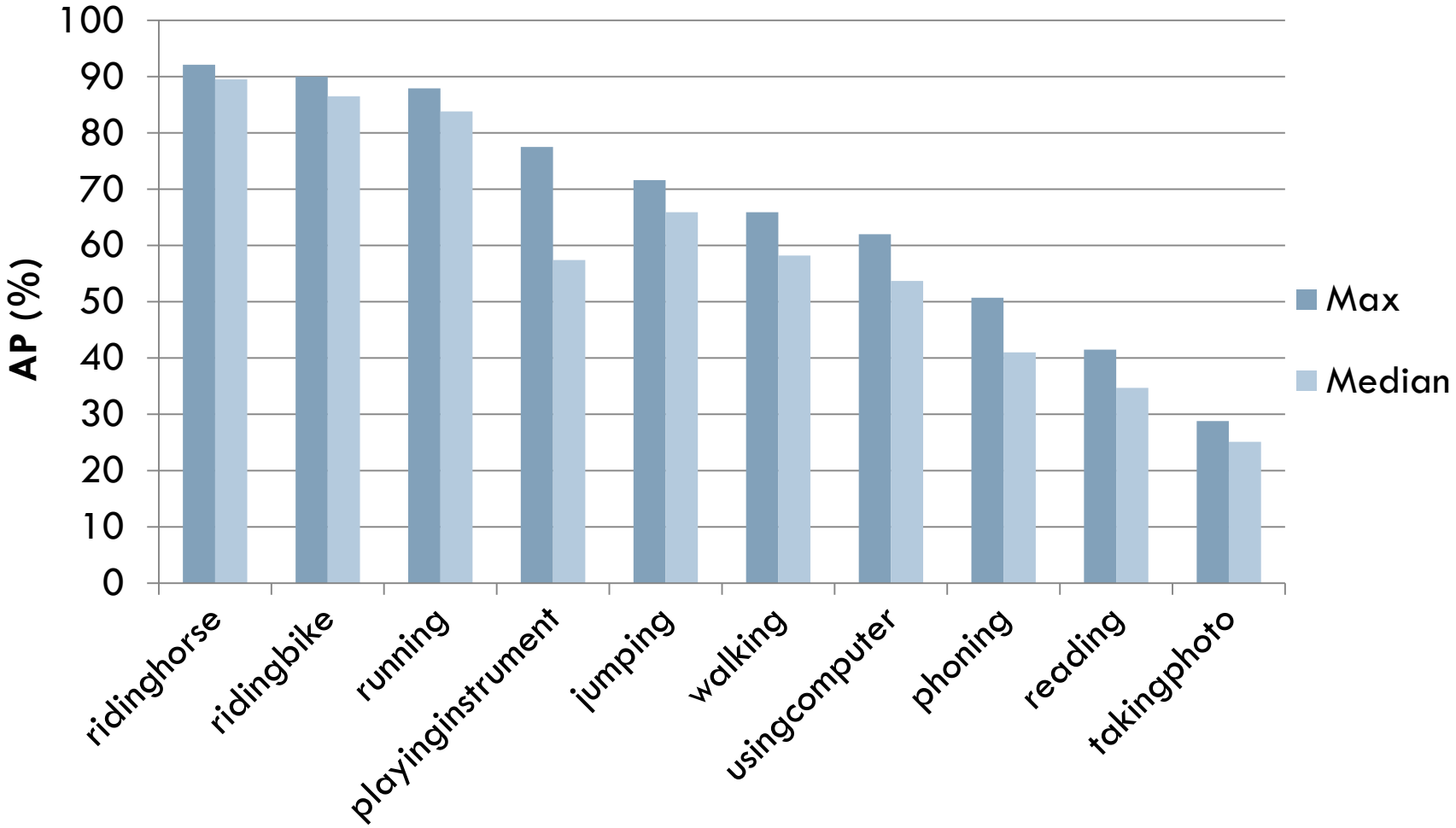
Precision/Recall – Riding Horse



Precision/Recall – Taking Photo

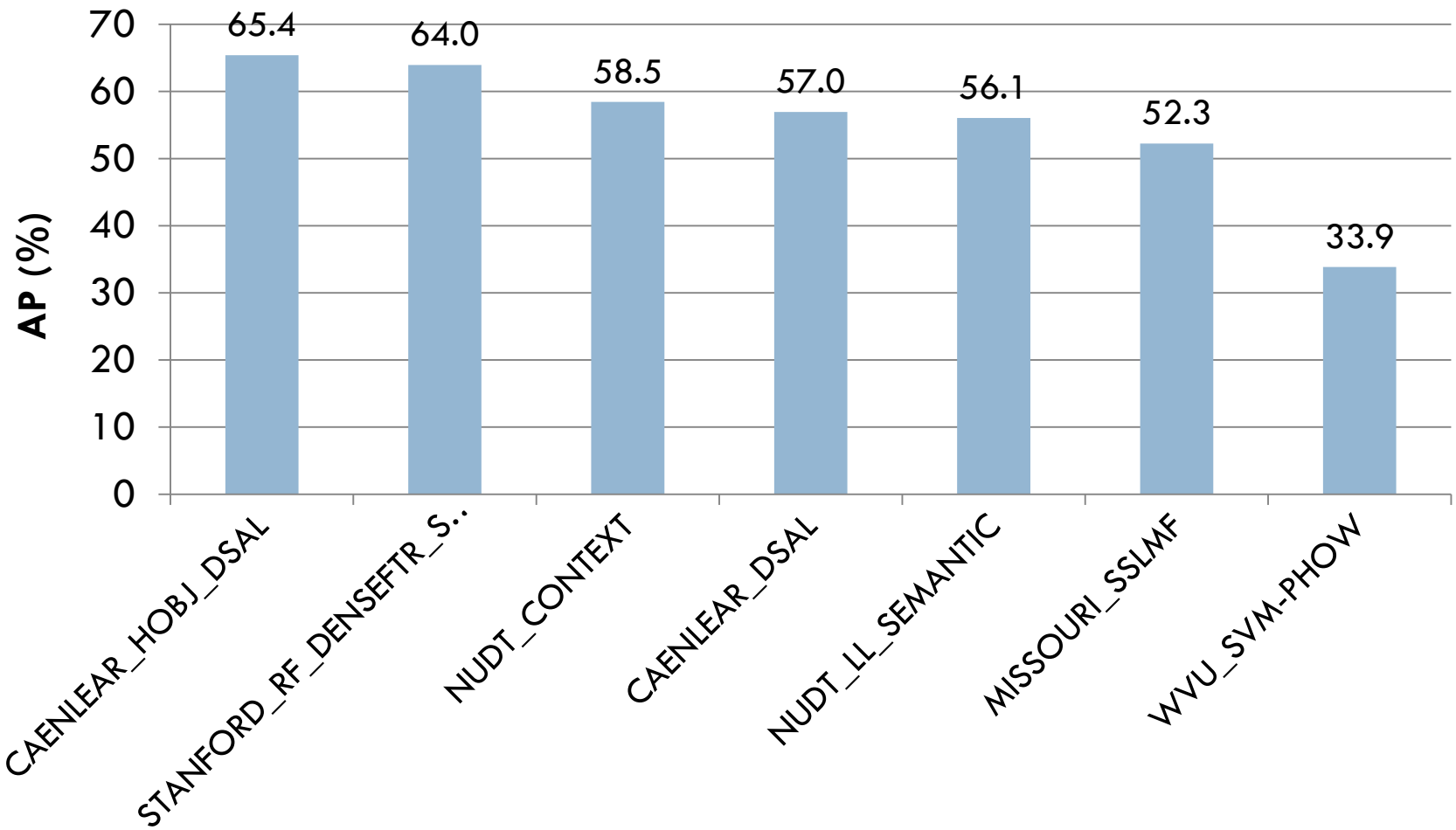


AP by Class



- Max AP: 92.1% (riding horse) … 28.8% (taking photo)

Median AP by Method



“True Positives” - Riding Horse

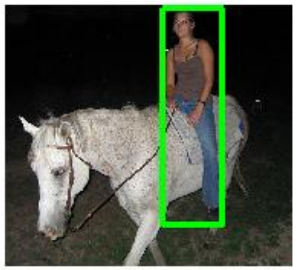
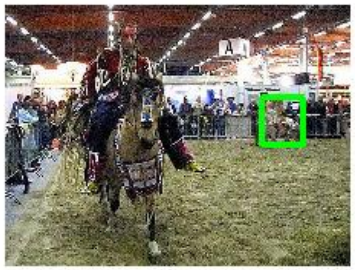
STANFORD_RF_DENSEFTR_SVM

NUDT_CONTEXT

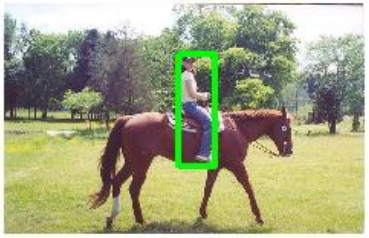
CAENLEAR_HOBJ_DSAL

“False Negatives” - Riding Horse

STANFORD_RF_DENSEFTR_SVM

NUDT_CONTEXT



CAENLEAR_HOBJ_DSAL

“False Positives” - Riding Horse

STANFORD_RF_DENSEFTR_SVM

NUDT_CONTEXT

CAENLEAR_HOBJ_DSAL

“True Positives” - Jumping

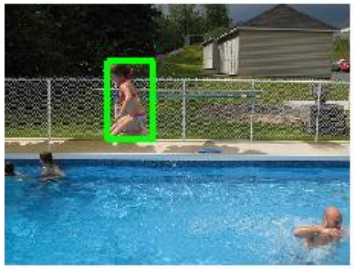
CAENLEAR_HOBJ_DSAL

NUDT_LL_SEMANTIC

STANFORD_RF_DENSEFTR_SVM

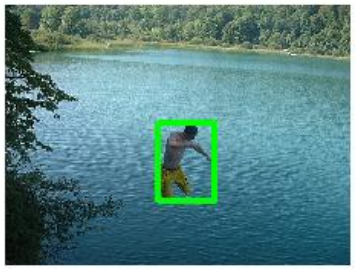
“False Negatives” - Jumping

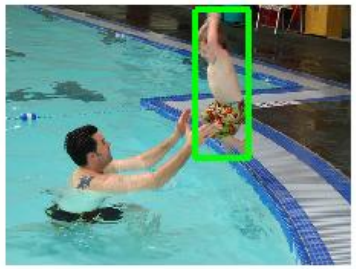
CAENLEAR_HOBJ_DSAL



NUDT_LL_SEMANTIC

STANFORD_RF_DENSEFTR_SVM





“False Positives” - Jumping

CAENLEAR_HOBJ_DSAL

NUDT_LL_SEMANTIC

STANFORD_RF_DENSEFTR_SVM

“True Positives” - Taking Photo

STANFORD_RF_DENSEFTR_SVM

NUDT_CONTEXT

CAENLEAR_HOBJ_DSAL

“False Negatives” - Taking Photo

STANFORD_RF_DENSEFTR_SVM

NUDT_CONTEXT

CAENLEAR_HOBJ_DSAL

“False Positives” - Taking Photo

STANFORD_RF_DENSEFTR_SVM

NUDT_CONTEXT

CAENLEAR_HOBJ_DSAL

Prizes

- Winner
  - **STANFORD_RF_DENSEFTR_SVM**
    Bangpeng Yao, Aditya Khosla, Li Fei-Fei
    Stanford University
- Runners-up
  - **CAENLEAR_HOBJ_DSAL**
    Gaurav Sharma, Alessandro Prest, Frederic Jurie, Vittorio Ferrari, Cordelia Schmid
    University of Caen, INRIA LEAR
  - **NUDT_CONTEXT/NUDT_LL_SEMANTIC**
    Li Zhou, Dewen Hu, Zongtan Zhou
    National University of Defense Technology, China