The 2006 PASCAL Visual Object Classes Challenge

Mark Everingham
Luc Van Gool
Chris Williams
Andrew Zisserman

Challenge

- Ten object classes
  - bicycle, bus, car, cat, cow, dog, horse, motorbike, person, sheep
- Classification
  - Predict whether at least one object of a given class is present
- Detection
  - Predict bounding boxes of objects of a given class

Competitions

- Train on the supplied data
  - Which methods perform best given specified training data?
- Train on any (non-test) data
  - How well do state-of-the-art methods perform on these problems?
  - Which methods perform best?

Dataset

- Images taken from three sources
  - Personal photos contributed by Edinburgh/Oxford
  - Microsoft Research Cambridge images
  - Images taken from “flickr” photo-sharing website
- Annotation
  - Bounding box
  - Viewpoint: front, rear, left, right, unspecified
  - “Truncated” flag: Bounding box $\neq$ object extent
  - “Difficult” flag: Objects ignored in challenge

Examples

Bicycle

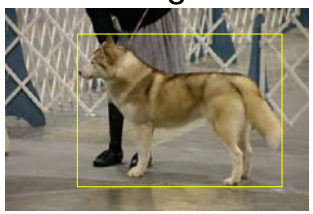
Bus

Car

Cat

Cow

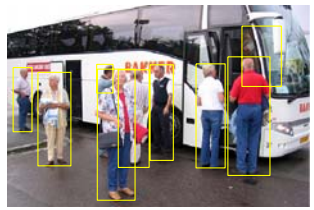
Dog

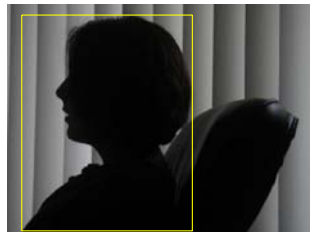


Horse

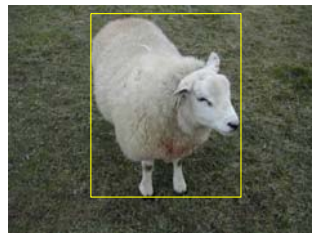
Motorbike

Person





Sheep



Annotation Procedure

- All annotation performed in a single session in a single location by seven annotators
- Detailed guidelines decided beforehand
  - What to label
    - Not excessive motion blur, poor illumination etc.
    - Object size, “recognisability”, level of occlusion
    - “Close-fitting occluders” e.g. snow/mud treated as object
    - Through glass, mirrors, pictures: label, reflections (=occlusion)
    - Non-photorealistic pictures: don’t label
  - Viewpoint
  - Bounding box e.g. don’t extend greatly for few pixels
  - Truncation: significant amount of object outside bounding box
- “Difficult” flag set afterwards by a single annotator examining individual objects in isolation

Dataset Statistics

	train		val		trainval		test	
	img	**obj**	**img**	**obj**	**img**	**obj**	**img**	**obj**
Bicycle	127	161	143	162	270	323	268	326
Bus	93	118	81	117	174	235	180	233
Car	271	427	282	427	553	854	544	854
Cat	192	214	194	215	386	429	388	429
Cow	102	156	104	157	206	313	197	315
Dog	189	211	176	211	365	422	370	423
Horse	129	164	118	162	247	326	254	324
Motorbike	118	138	117	137	235	275	234	274
Person	319	577	347	579	666	1156	675	1153
Sheep	119	211	132	210	251	421	238	422
Total	1277	2377	1341	2377	2618	4754	2686	4753

Participation

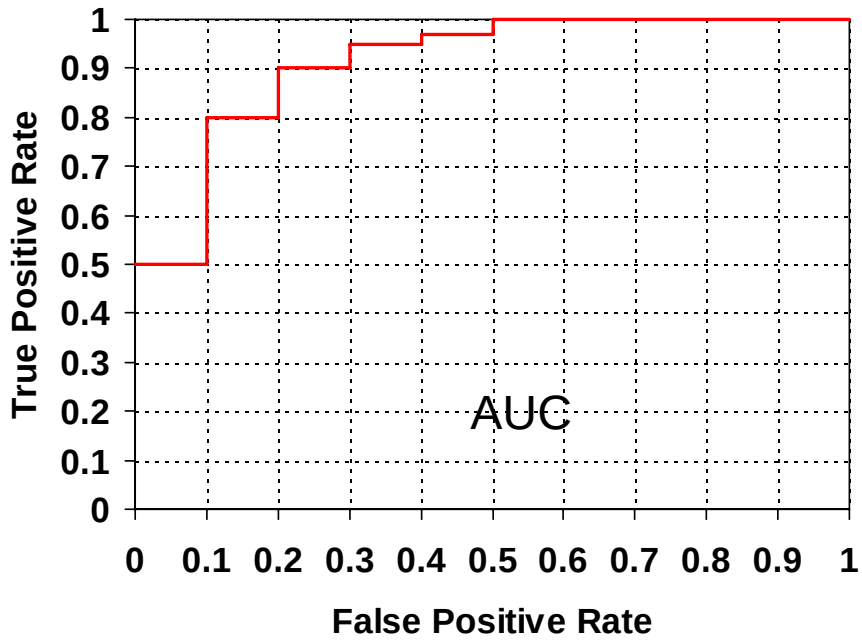
- 22 participants submitted results
  - 14 different institutions
- 28 different methods
  - 19 for classification task only
  - 4 for detection task only
  - 5 for classification and detection

1. Classification Task

Predict whether at least one object of a given class is present

Evaluation

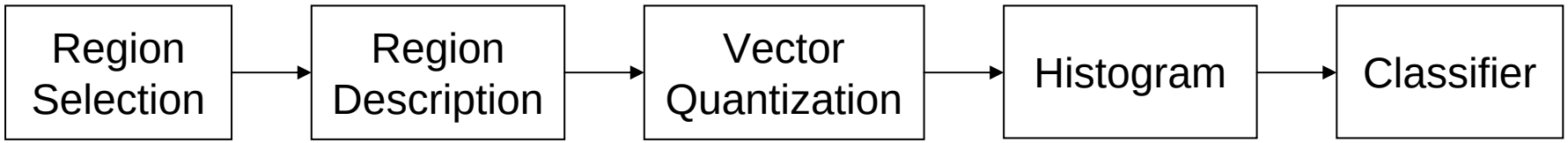
- Receiver Operating Characteristic (ROC)
  - Area Under Curve (AUC)



Methods

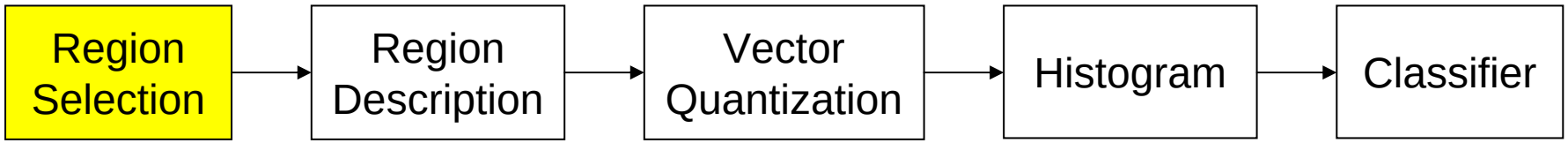
- **Bag of words: 15/20 (75%)**
- Correspondence-based
- Classification of individual patches/regions
- Local classification of “concepts”
- Graph neural network
- Classification by detection
  - Generalized Hough transform
  - “Star” constellation model
  - Sliding-window classifier

“Bag of words” Methods



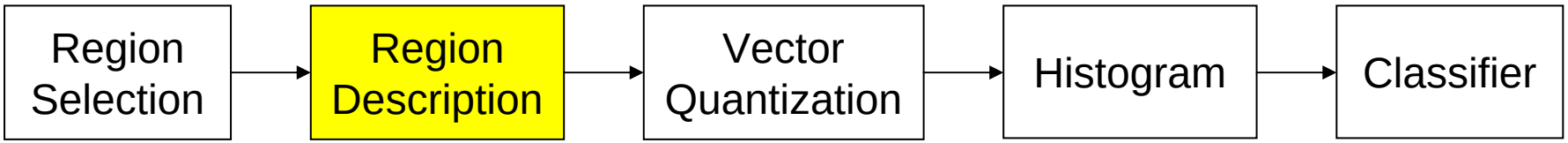
- Local regions are extracted from the image
- Region appearance is described by a descriptor
- Descriptors are quantized into “visual words”
- Image is represented as a histogram of visual words
- Classifier is trained to output class/non-class

Region Selection



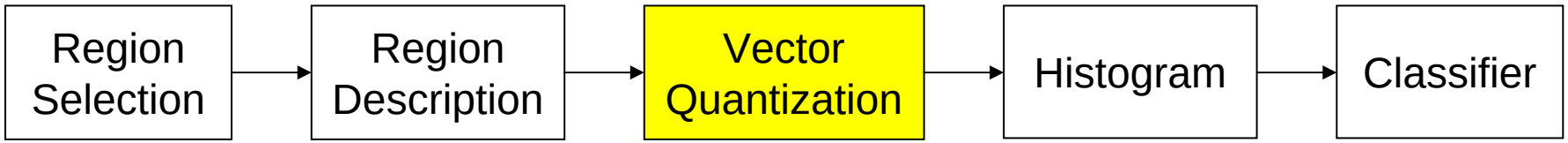
- "Sparse" methods based on interest points
  - Scale invariant: Harris-Laplace, Laplacian, DoG
  - Affine invariant: Hessian-Affine, MSER
  - Wavelets
- "Dense" methods
  - Multi-scale (overlapping) grid
- Other methods
  - Random position and scale patches with feedback from classifier
  - Segmented regions
- Combination of multiple methods

Region Description



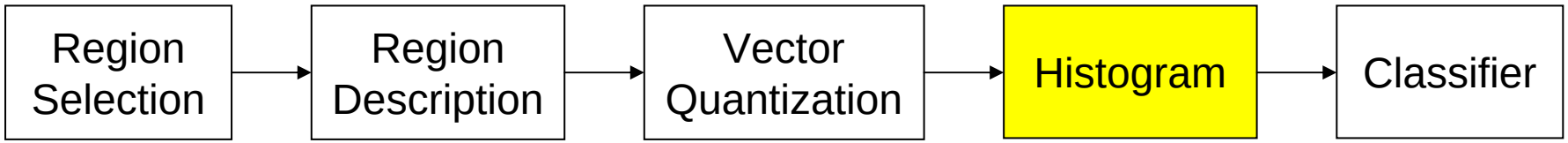
- SIFT
- PCA on vector of pixel values
- Haar wavelets
- Grey-level moments and invariants
- Colour and colour histograms
- Shape context
- Texture moments, texton histograms
- Position in spatial pyramid

Vector Quantization



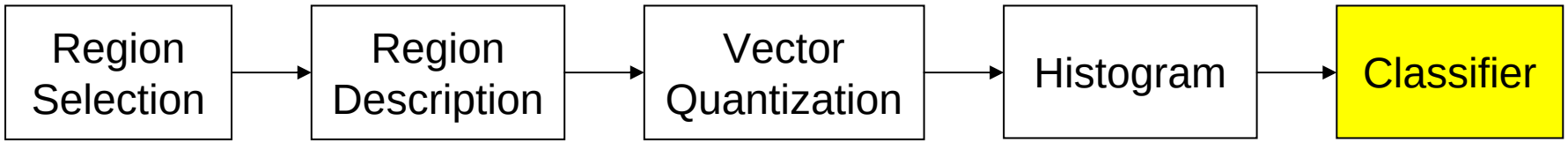
- Single codebook
- Multiple codebooks: per class, per region type, per descriptor type
- K-means, LBG clustering
- Supervised clustering
- Random cluster centres + selection by validation

Histogramming



- “Continuous valued”
  - Record frequency of each visual word
- Binary valued
  - Record only presence/absence of each visual word

Classifier



- Non-linear SVM: χ^2 kernel
  - Single classifier
  - Classifier per pyramid level
- Linear
  - Logistic regression/iterative scaling
  - Linear SVM
  - Least angle regression
- Other
  - Linear programming boosting

Other Methods

- Correspondence-based: Find nearest neighbour region in training images (with geometric context) and vote by class of training image
- Classification of individual patches/regions: Classify patches and accumulate class confidence over patches in the image
  - Nearest neighbour, boosting, self-organizing map
- Graph neural network: Segment image into a fixed number of regions and classify based on region descriptors and neighbour relations

Classification by Detection

- Detect objects of particular class in the image
  - Generalized Hough transform
  - “Star” Constellation model
  - Sliding-window classifier
- Assign maximum detection confidence as image classification confidence

- More in-line with human intuition: “There is a car *here* therefore the image contains a car”

Classification Results

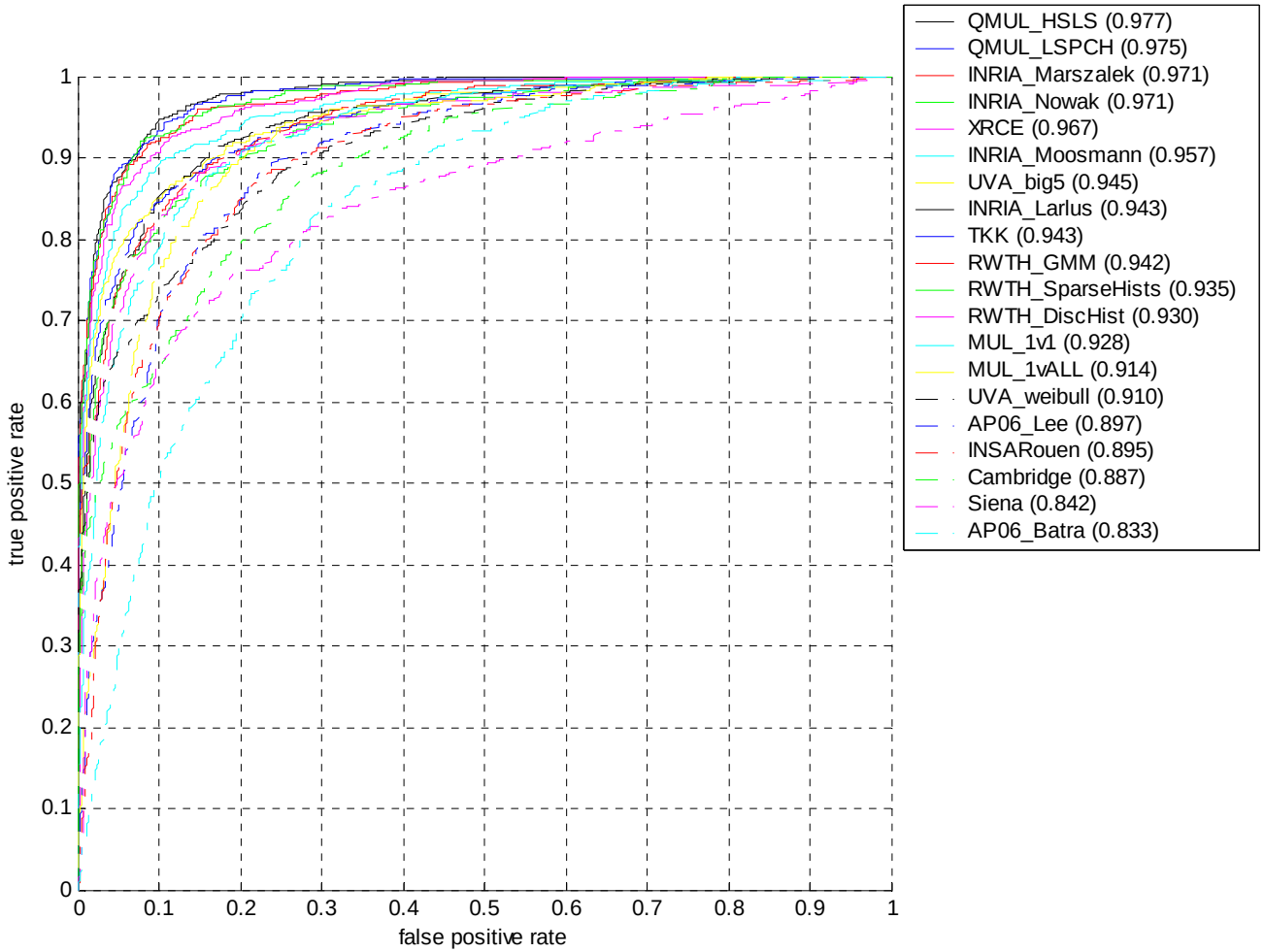
Competition 1: Train on VOC data

Participants

	bicycle	bus	car	cat	cow	dog	horse	motor bike	person	sheep
AP06_Batra	×	×	×	×	×	×	×	×	×	×
AP06_Lee	×	×	×	×	×	×	×	×	×	×
Cambridge	×	×	×	×	×	×	×	×	×	×
ENSMP	–	–	–	–	–	–	–	–	–	–
INRIA_Douze	–	–	–	–	–	–	–	–	–	–
INRIA_Laptev	–	–	–	–	–	–	–	–	–	–
INRIA_Larlus	×	×	×	×	×	×	×	×	×	×
INRIA_Marszalek	×	×	×	×	×	×	×	×	×	×
INRIA_Moosmann	×	×	×	×	×	×	×	–	×	×
INRIA_Nowak	×	×	×	×	×	×	×	×	×	×
INSARouen	–	–	×	–	–	×	–	–	–	×
KUL	–	–	–	–	–	–	–	–	–	–
MIT_Fergus	–	–	–	–	–	–	–	–	–	–
MIT_Torralba	–	–	–	–	–	–	–	–	–	–
MUL	×	×	×	×	×	×	×	×	×	×
QMUL	×	×	×	×	×	×	×	×	×	×
RWTH	×	×	×	×	×	×	×	×	×	×
Siena	×	×	×	×	×	×	×	×	×	×
TKK	×	×	×	×	×	×	×	×	×	×
TUD	–	–	–	–	–	–	–	–	–	–
UVA	×	×	×	×	×	×	×	×	×	×
XRCE	×	×	×	×	×	×	×	×	×	×

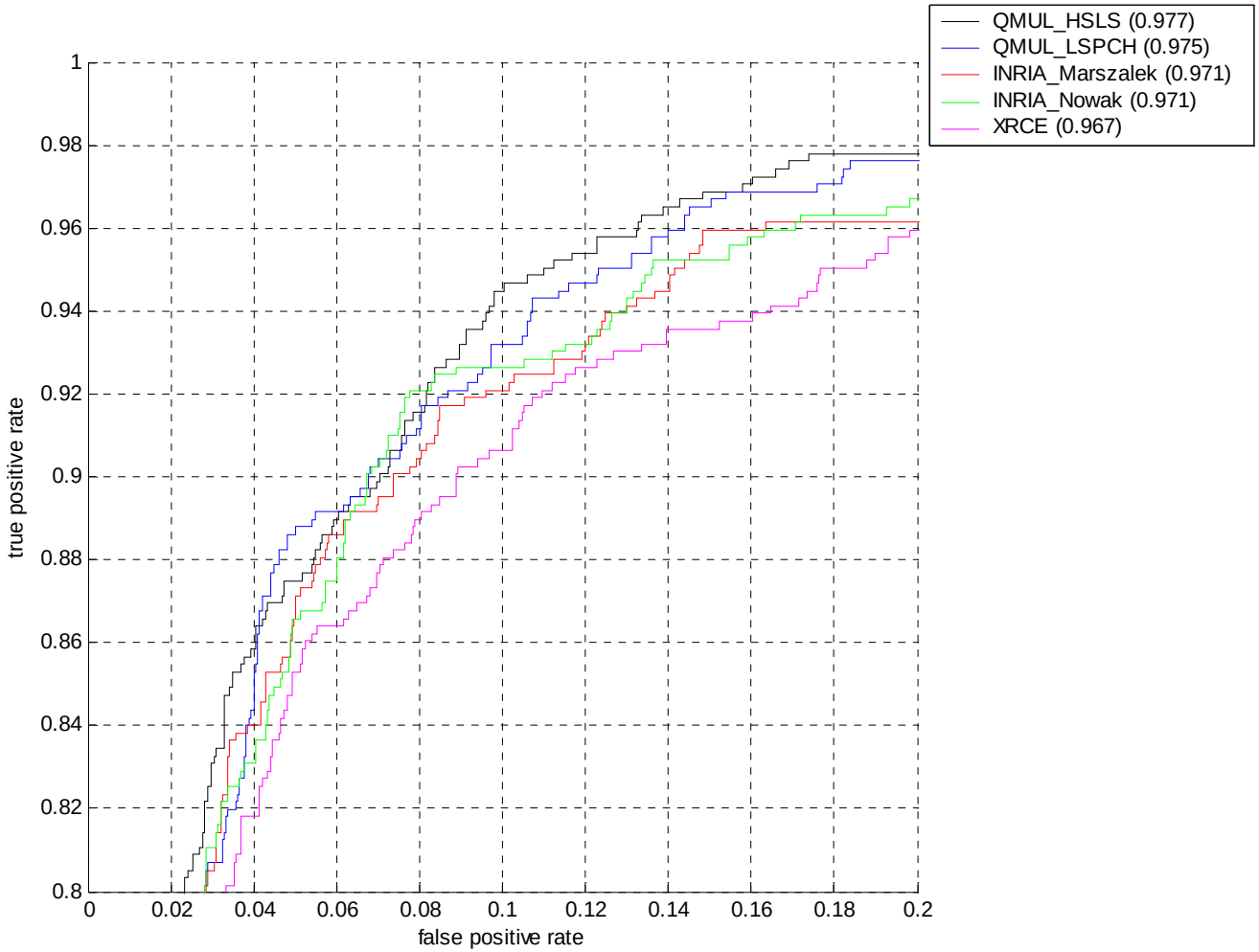
Competition 1: Car

- All methods



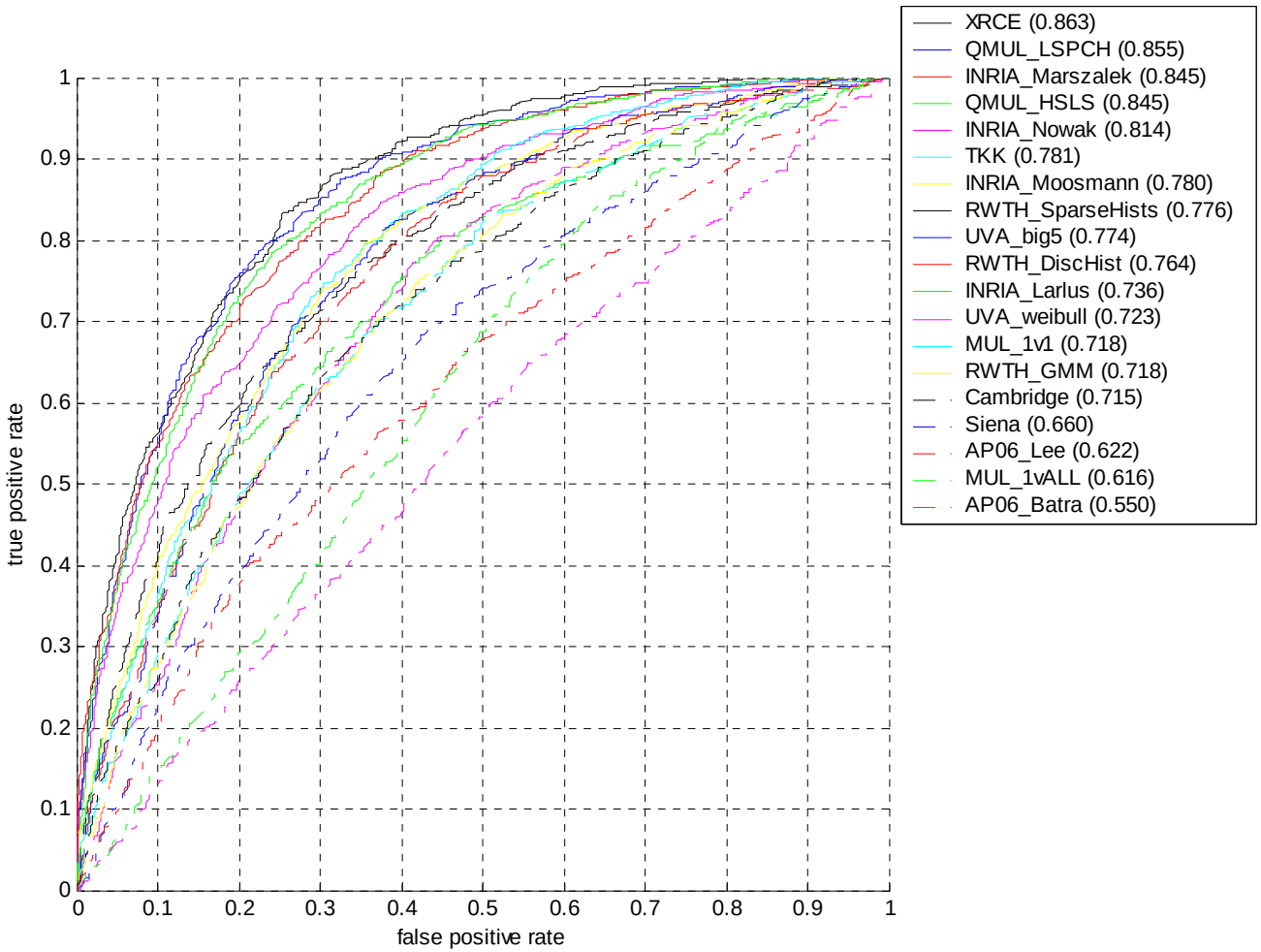
Competition 1: Car

- Top 5 methods by AUC



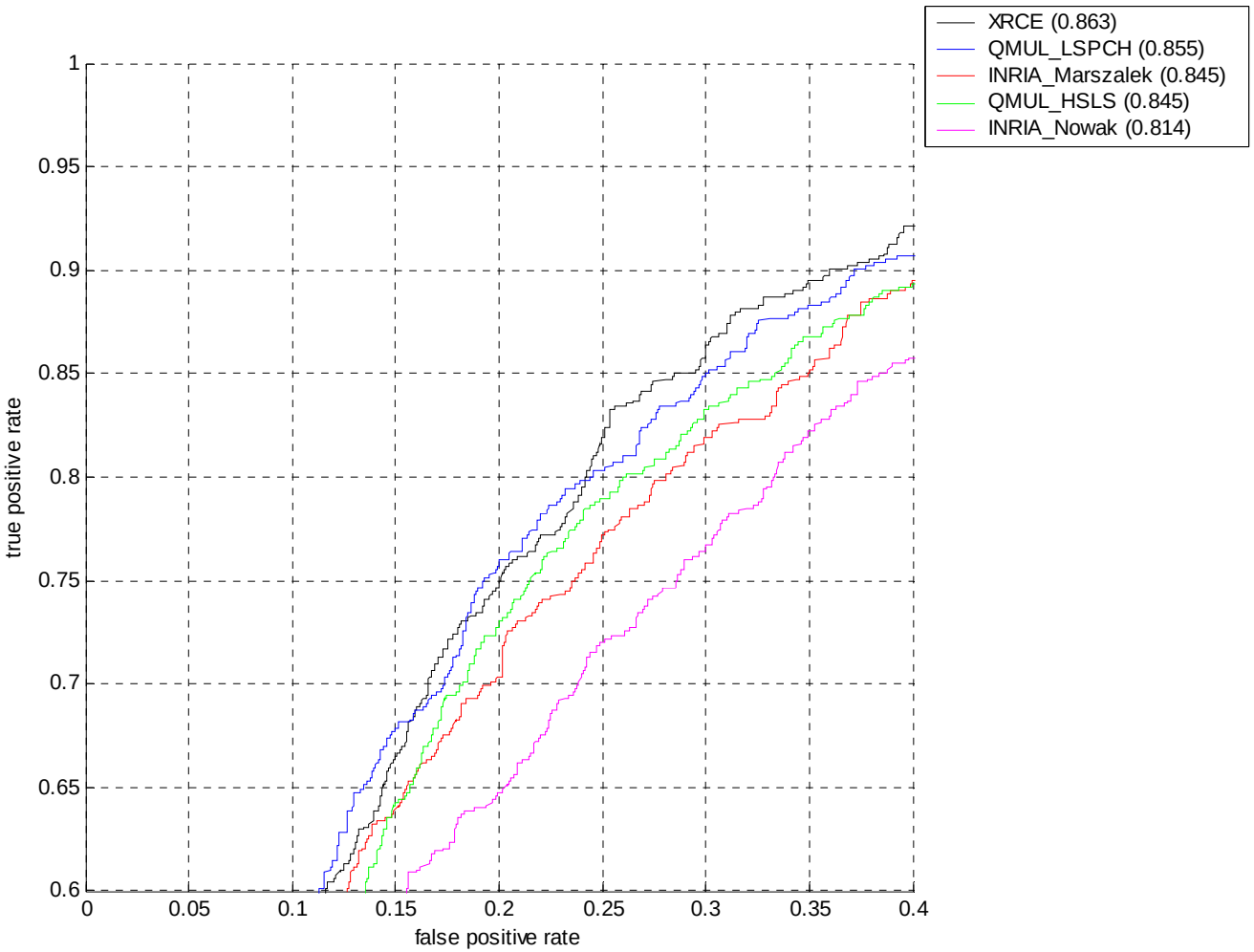
Competition 1: Person

- All methods



Competition 1: Person

- Top 5 methods by AUC



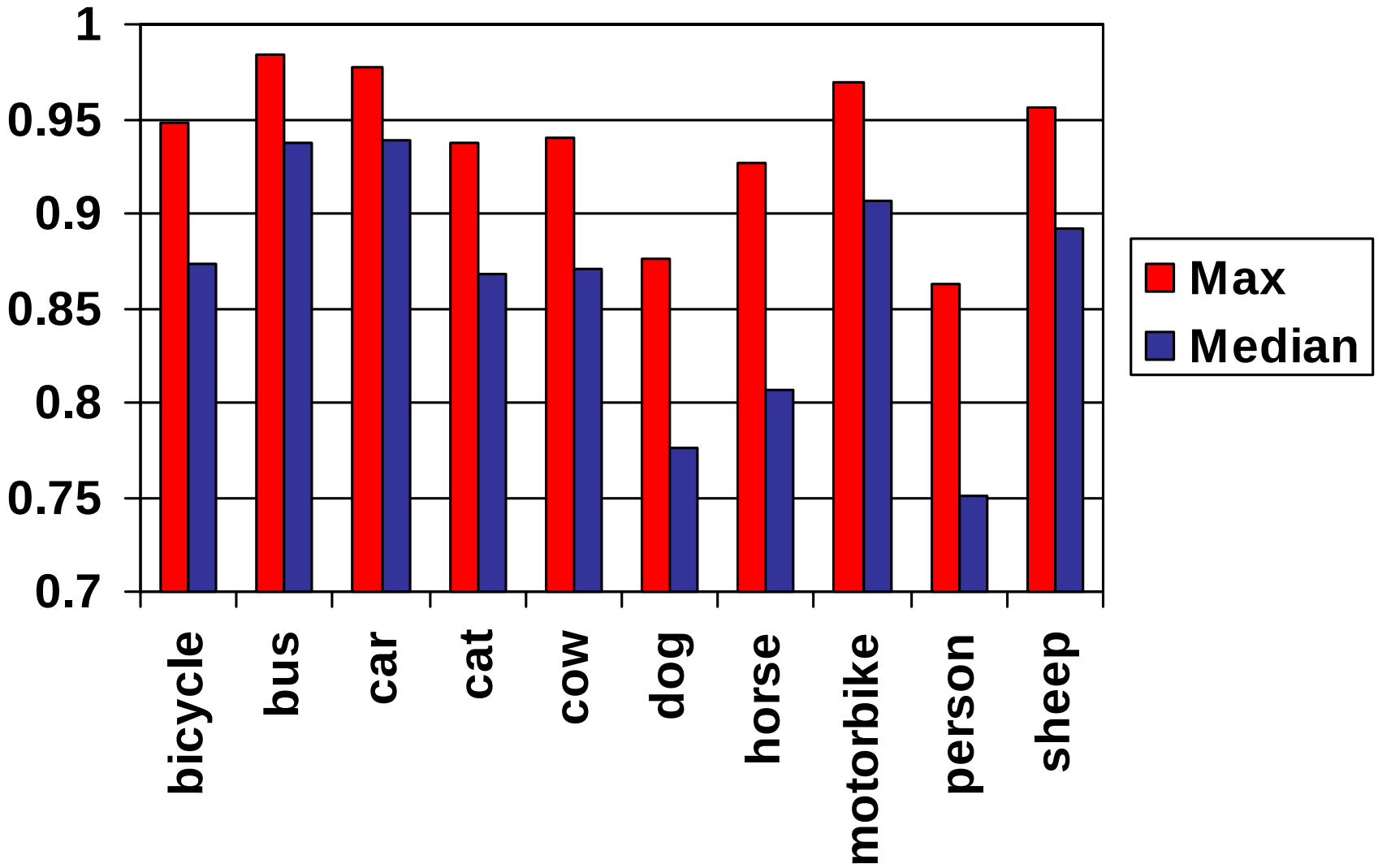
AUC by Method and Class

	bicycle	bus	car	cat	cow	dog	horse	motor bike	person	sheep
AP06_Batra	0.791	0.637	0.833	0.733	0.756	0.644	0.607	0.672	0.550	0.792
AP06_Lee	0.845	0.916	0.897	0.859	0.838	0.766	0.694	0.829	0.622	0.875
Cambridge	0.873	0.864	0.887	0.822	0.850	0.768	0.754	0.844	0.715	0.866
INRIA_Larlus	0.903	0.948	0.943	0.870	0.880	0.743	0.850	0.890	0.736	0.892
INRIA_Marszalek	0.929	0.984	0.971	0.922	0.938	0.856	0.908	0.964	0.845	0.944
INRIA_Moosmann	0.903	0.933	0.957	0.883	0.895	0.825	0.824	-	0.780	0.930
INRIA_Nowak	0.924	0.973	0.971	0.906	0.892	0.797	0.904	0.961	0.814	0.940
INSARouen	-	-	0.895	-	-	0.764	-	-	-	0.869
MUL_1vALL	0.857	0.852	0.914	0.562	0.632	0.584	0.525	0.831	0.616	0.758
MUL_1v1	0.864	0.945	0.928	0.826	0.789	0.764	0.733	0.906	0.718	0.872
QMUL_HSLS	0.944	0.984	0.977	0.936	0.936	0.874	0.922	0.966	0.845	0.946
QMUL_LSPCH	0.948	0.981	0.975	0.937	0.938	0.876	0.926	0.969	0.855	0.956
RWTH_DiscHist	0.874	0.955	0.930	0.879	0.910	0.799	0.854	0.938	0.764	0.906
RWTH_GMM	0.882	0.935	0.942	0.866	0.856	0.825	0.802	0.905	0.718	0.892
RWTH_SparseHists	0.863	0.941	0.935	0.883	0.883	0.704	0.844	0.858	0.776	0.907
Siena	0.671	0.749	0.842	0.696	0.774	0.677	0.644	0.701	0.660	0.768
TKK	0.857	0.928	0.943	0.871	0.892	0.811	0.806	0.908	0.781	0.900
UVA_big5	0.897	0.929	0.945	0.845	0.862	0.785	0.806	0.923	0.774	0.885
UVA_weibull	0.855	0.880	0.910	0.818	0.849	0.762	0.759	0.888	0.723	0.811
XRCE	0.943	0.978	0.967	0.933	0.940	0.866	0.925	0.957	0.863	0.951

Ranking by AUC per Class

	bicycle	bus	car	cat	cow	dog	horse	motor bike	person	sheep
AP06_Batra	18	19	20	17	18	19	18	18	19	18
AP06_Lee	17	14	16	12	15	12	16	16	17	13
Cambridge	11	16	18	15	13	11	14	14	15	16
INRIA_Larlus	6	7	8	10	10	16	7	11	11	10
INRIA_Marszalek	4	1	3	4	2	4	4	3	3	4
INRIA_Moosmann	7	11	6	6	6	5	9	-	7	6
INRIA_Nowak	5	5	4	5	7	9	5	4	5	5
INSARouen	-	-	17	-	-	13	-	-	-	15
MUL_1vALL	14	17	14	19	19	20	19	15	18	20
MUL_1v1	12	8	13	14	16	14	15	9	13	14
QMUL_HSLS	2	2	1	2	4	2	3	2	4	3
QMUL_LSPCH	1	3	2	1	3	1	1	1	2	1
RWTH_DiscHist	10	6	12	8	5	8	6	6	10	8
RWTH_GMM	9	10	10	11	12	6	12	10	14	11
RWTH_SparseHists	13	9	11	7	9	17	8	13	8	7
Siena	19	18	19	18	17	18	17	17	16	19
TKK	15	13	9	9	8	7	10	8	6	9
UVA_big5	8	12	7	13	11	10	11	7	9	12
UVA_weibull	16	15	15	16	14	15	13	12	12	17
XRCE	3	4	5	3	1	3	2	5	1	2

AUC by Class

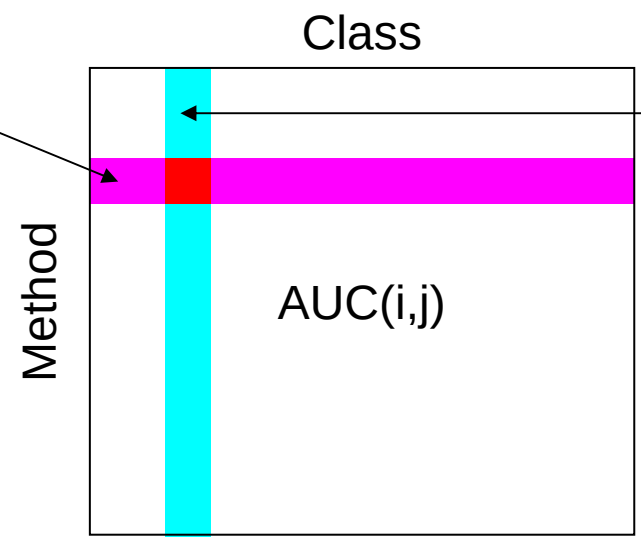

1
0.95
0.9
0.85
0.8
0.75
0.7
bicycle
bus
car
cat
cow
dog
horse
motorbike
person
sheep
Max
Median

“ANOVA” Analysis

- Explain AUC as a function of method *i* and class *j*:

$$AUC(i,j)=\alpha_i+\beta_j+\mu$$

Method	α
QMUL_LSPCH	0.074
QMUL_HSLS	0.071
XRCE	0.070
INRIA_Marszalek	0.064
INRIA_Nowak	0.046
RWTH_DiscHist	0.019
TKK	0.007
INRIA_Larlus	0.003
UVA_big5	0.003
RWTH_GMM	0.000
RWTH_SparseHists	-0.003
MUL_1v1	-0.028
UVA_weibull	-0.037
Cambridge	-0.038
AP06_Lee	-0.048
Siena	-0.144
MUL_1vALL	-0.149
AP06_Batra	-0.161



Class	β
car	0.047
bus	0.029
motorbike	0.003
sheep	0.000
bicycle	-0.008
cow	-0.025
cat	-0.039
horse	-0.089
dog	-0.109
person	-0.138

Median ranked images: Bicycle

- Highest ranked class images

- Highest ranked non-Microsoft class images

Median ranked images: Bicycle

- Lowest ranked class images










- Highest ranked non-class images










Median ranked images: Bus

- Highest ranked class images

Median ranked images: Bus

- Lowest ranked class images










- Highest ranked non-class images



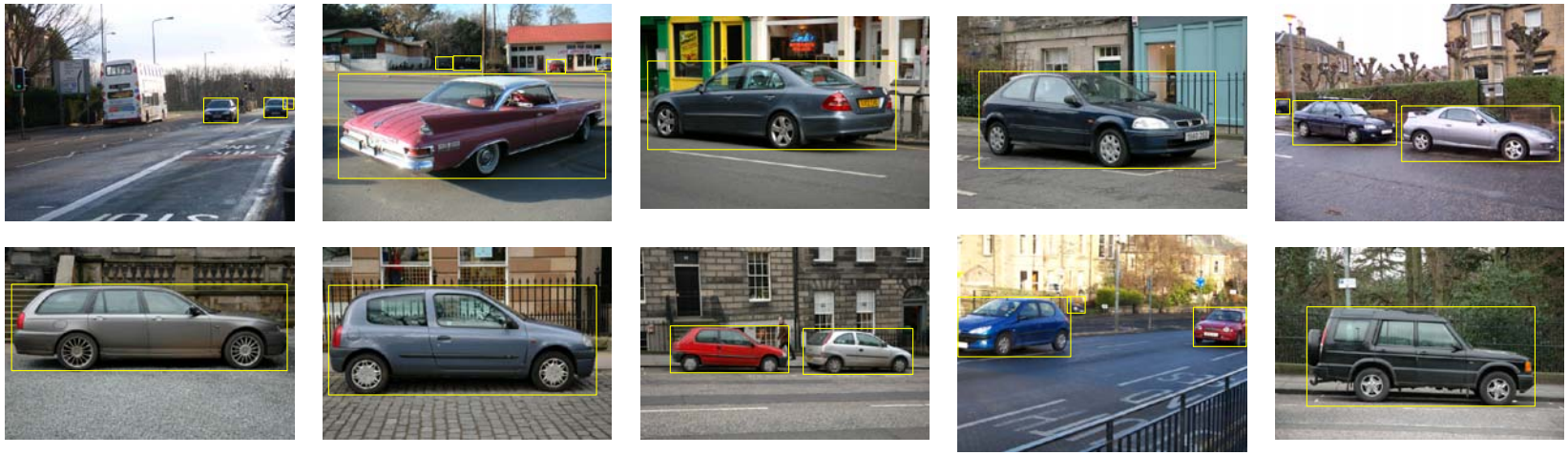





Median ranked images: Car

- Highest ranked class images

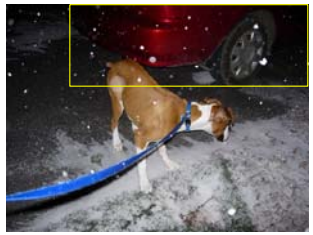
- Highest ranked non-Microsoft class images



Median ranked images: Car

- Lowest ranked class images










- Highest ranked non-class images



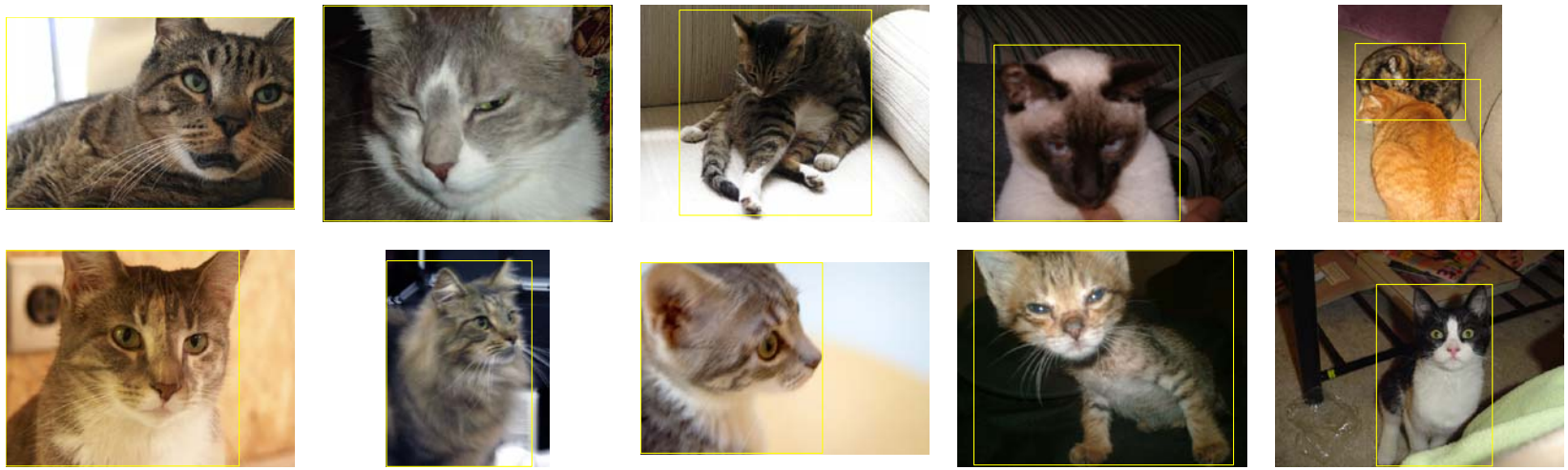






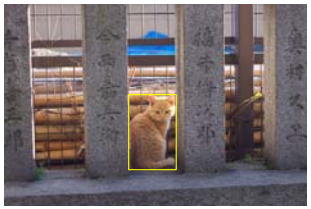
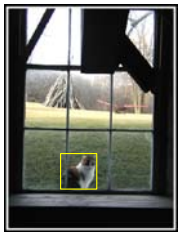
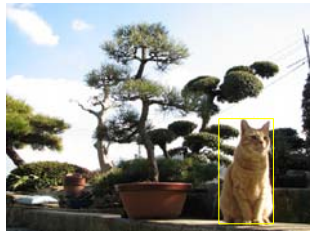

Median ranked images: Cat

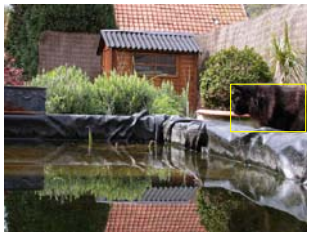
- Highest ranked class images

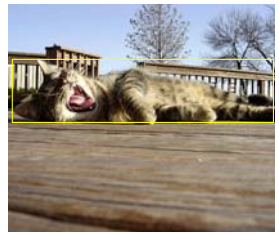


Median ranked images: Cat

- Lowest ranked class images

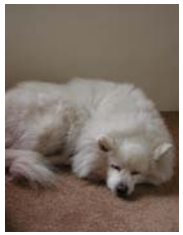




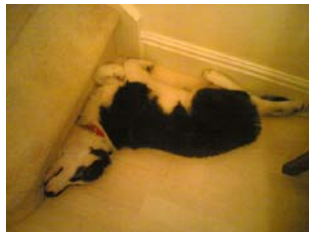
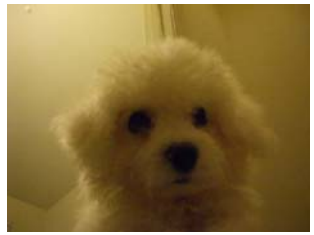




- Highest ranked non-class images







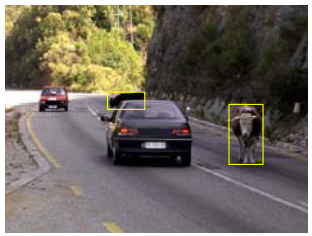

Median ranked images: Cow

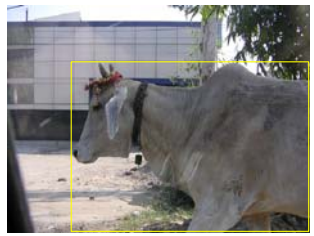
- Highest ranked class images

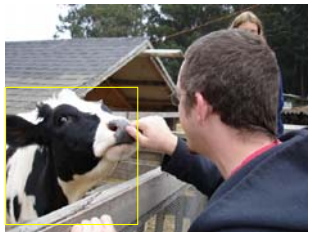
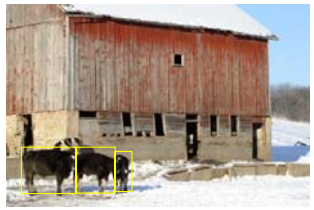
- Highest ranked non-Microsoft class images

Median ranked images: Cow

- Lowest ranked class images



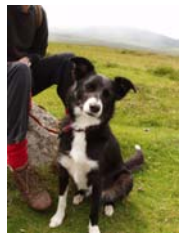




- Highest ranked non-class images



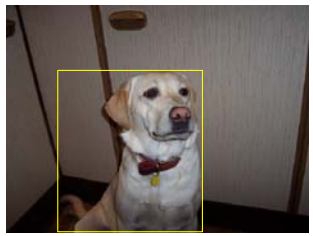
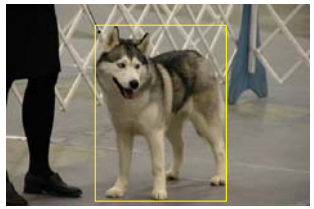
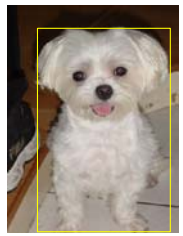


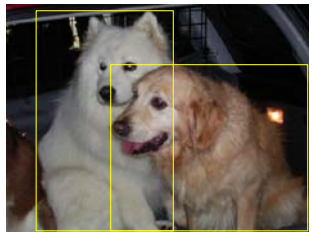
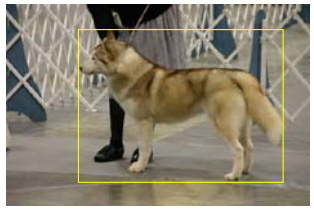
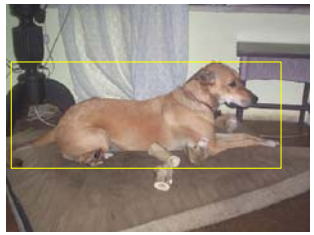


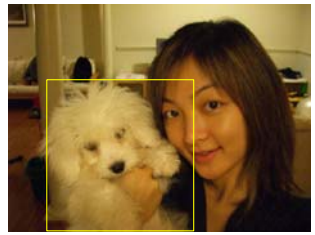



Median ranked images: Dog

- Highest ranked class images

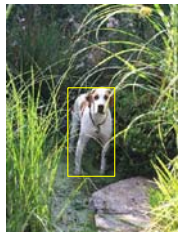






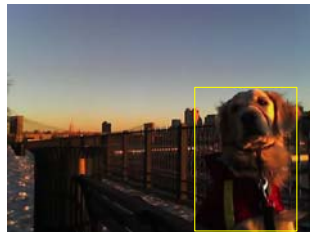
Median ranked images: Dog

- Lowest ranked class images

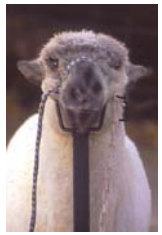



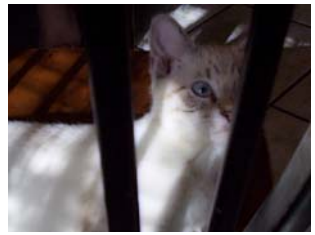




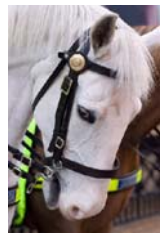

- Highest ranked non-class images



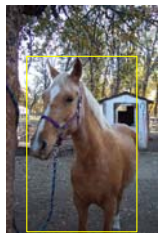
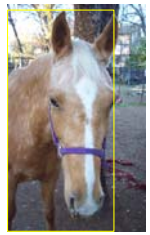


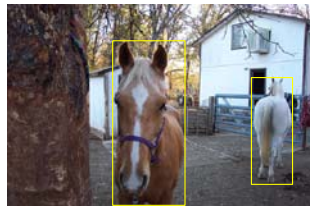
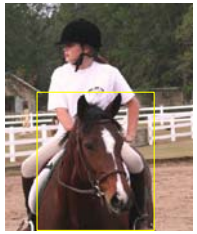


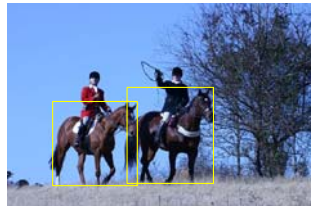


Median ranked images: Horse

- Highest ranked class images

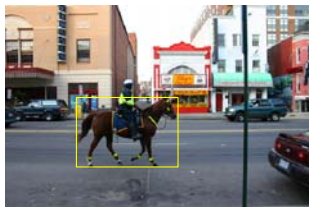



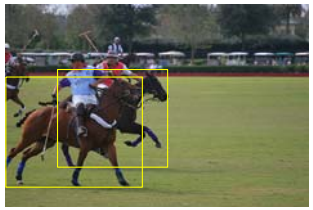

Median ranked images: Horse

- Lowest ranked class images








- Highest ranked non-class images



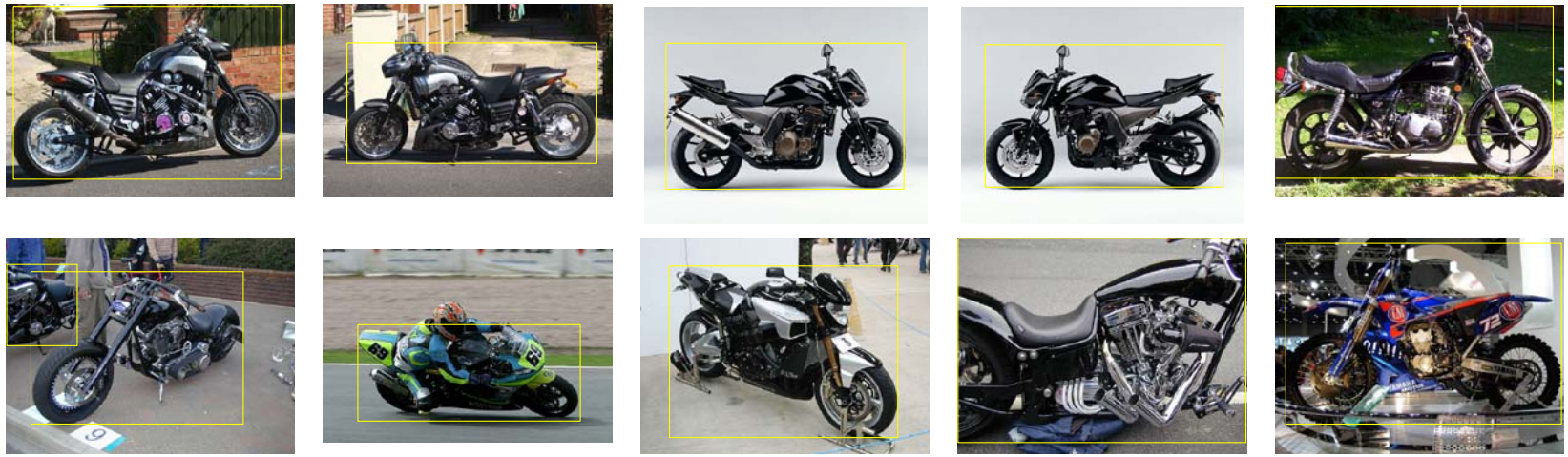







Median ranked images: Motorbike

- Highest ranked class images



Median ranked images: Motorbike

- Lowest ranked class images










- Highest ranked non-class images



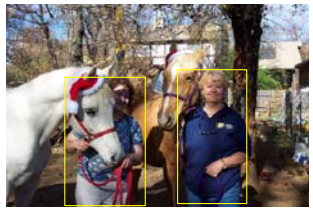


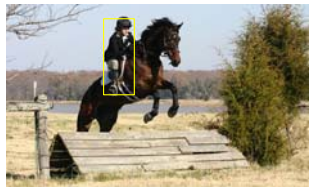
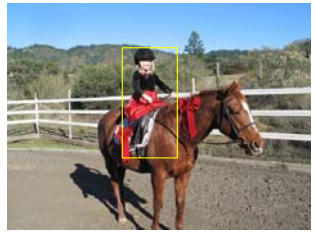


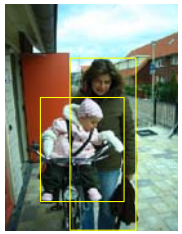
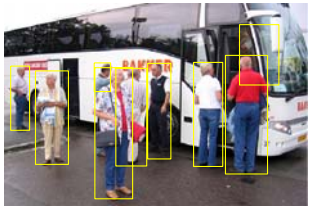
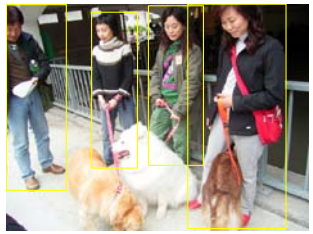
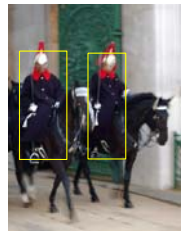
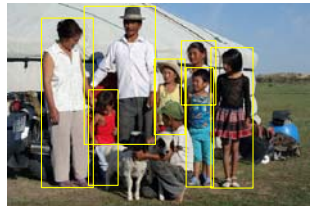


Median ranked images: Person

- Highest ranked class images



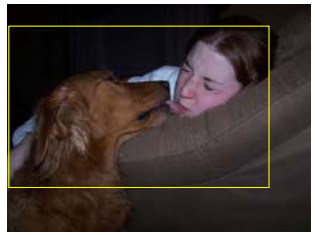



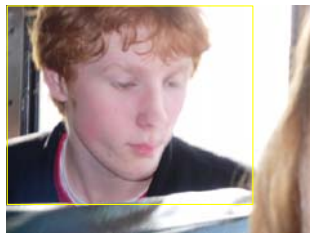
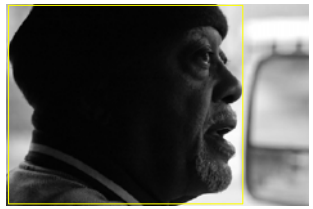






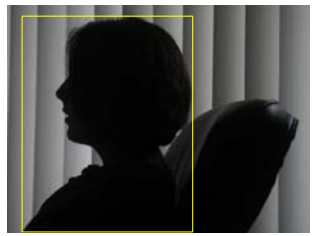
Median ranked images: Person

- Lowest ranked class images





- Highest ranked non-class images



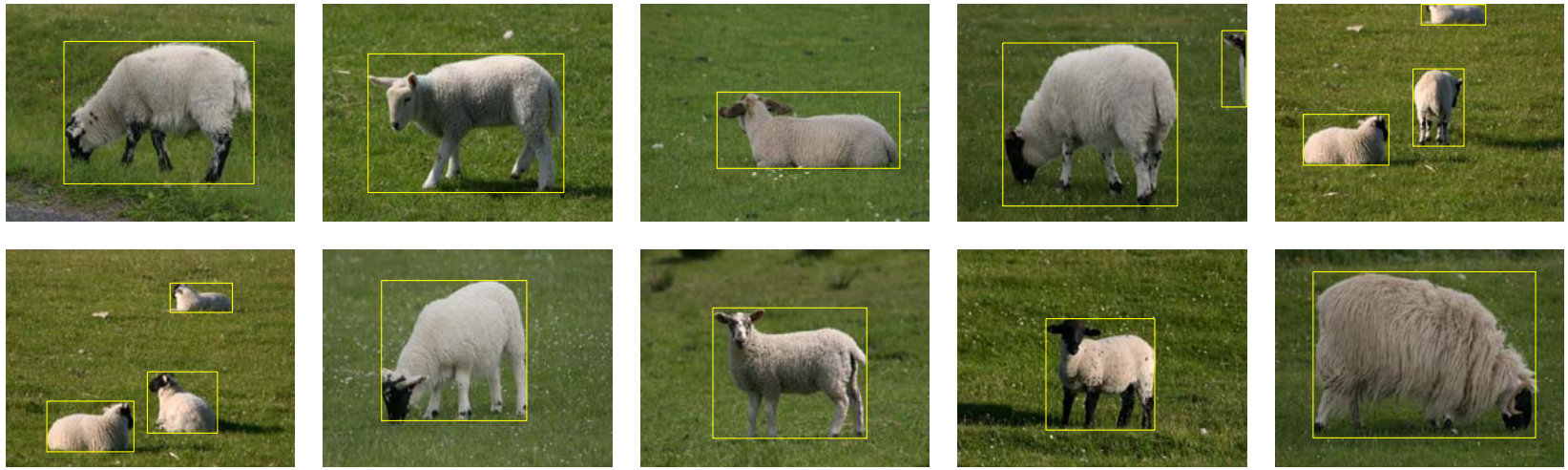




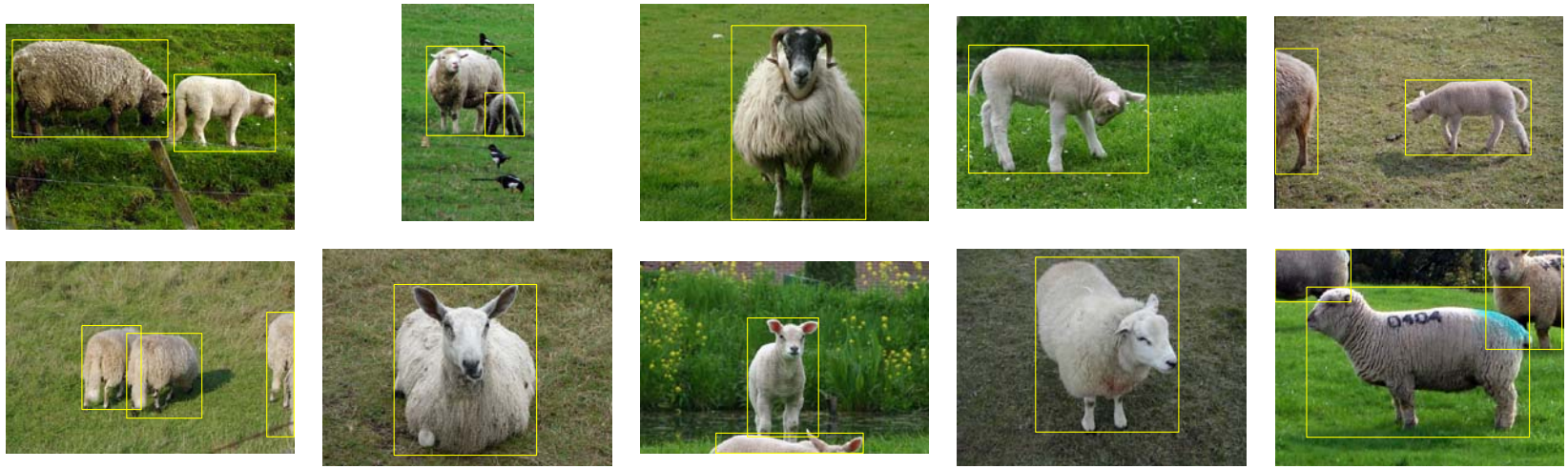


Median ranked images: Sheep

- Highest ranked class images

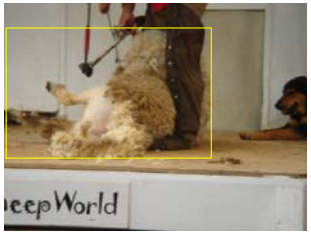


- Highest ranked non-Microsoft class images

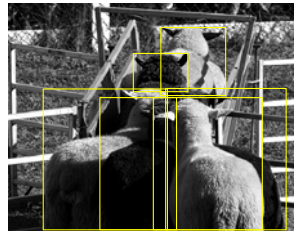
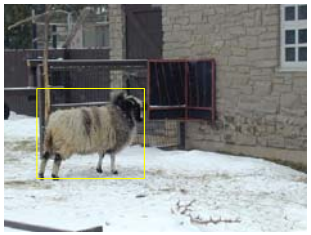


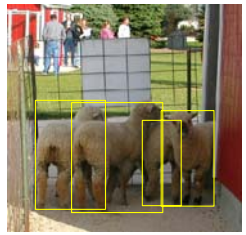
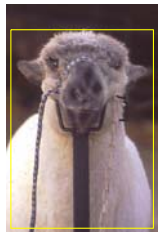
Median ranked images: Sheep

- Lowest ranked class images





- Highest ranked non-class images










Classification Results

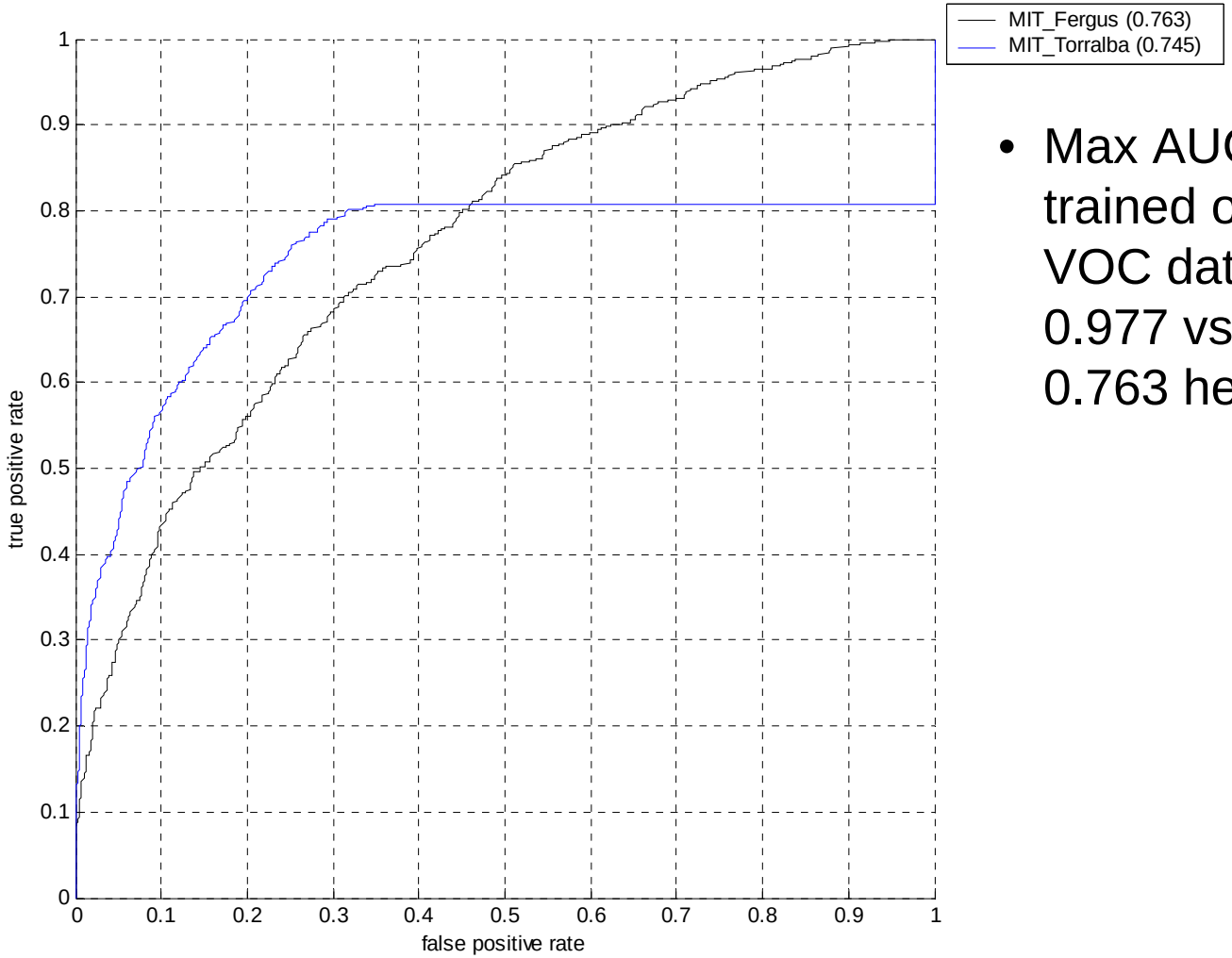
Competition 2: Train on own data

Participants

	bicycle	bus	car	cat	cow	dog	horse	motor bike	person	sheep
AP06_Batra	–	–	–	–	–	–	–	–	–	–
AP06_Lee	–	–	–	–	–	–	–	–	–	–
Cambridge	–	–	–	–	–	–	–	–	–	–
ENSMP	–	–	–	–	–	–	–	–	–	–
INRIA_Douze	–	–	–	–	–	–	–	–	–	–
INRIA_Laptev	–	–	–	–	–	–	–	–	–	–
INRIA_Larlus	–	–	–	–	–	–	–	–	–	–
INRIA_Marszalek	–	–	–	–	–	–	–	–	–	–
INRIA_Moosmann	–	–	–	–	–	–	–	–	–	–
INRIA_Nowak	–	–	–	–	–	–	–	–	–	–
INSARouen	–	–	–	–	–	–	–	–	–	–
KUL	–	–	–	–	–	–	–	×	–	–
MIT_Fergus	–	–	×	–	–	–	–	×	–	–
MIT_Torralba	–	–	×	–	–	–	–	–	–	–
MUL	–	–	–	–	–	–	–	–	–	–
QMUL	–	–	–	–	–	–	–	–	–	–
RWTH	–	–	–	–	–	–	–	–	–	–
Siena	–	–	–	–	–	–	–	–	–	–
TKK	–	–	–	–	–	–	–	–	–	–
TUD	–	–	–	–	–	–	–	–	–	–
UVA	–	–	–	–	–	–	–	–	–	–
XRCE	–	–	–	–	–	–	–	–	–	–

Competition 2: Train on own data

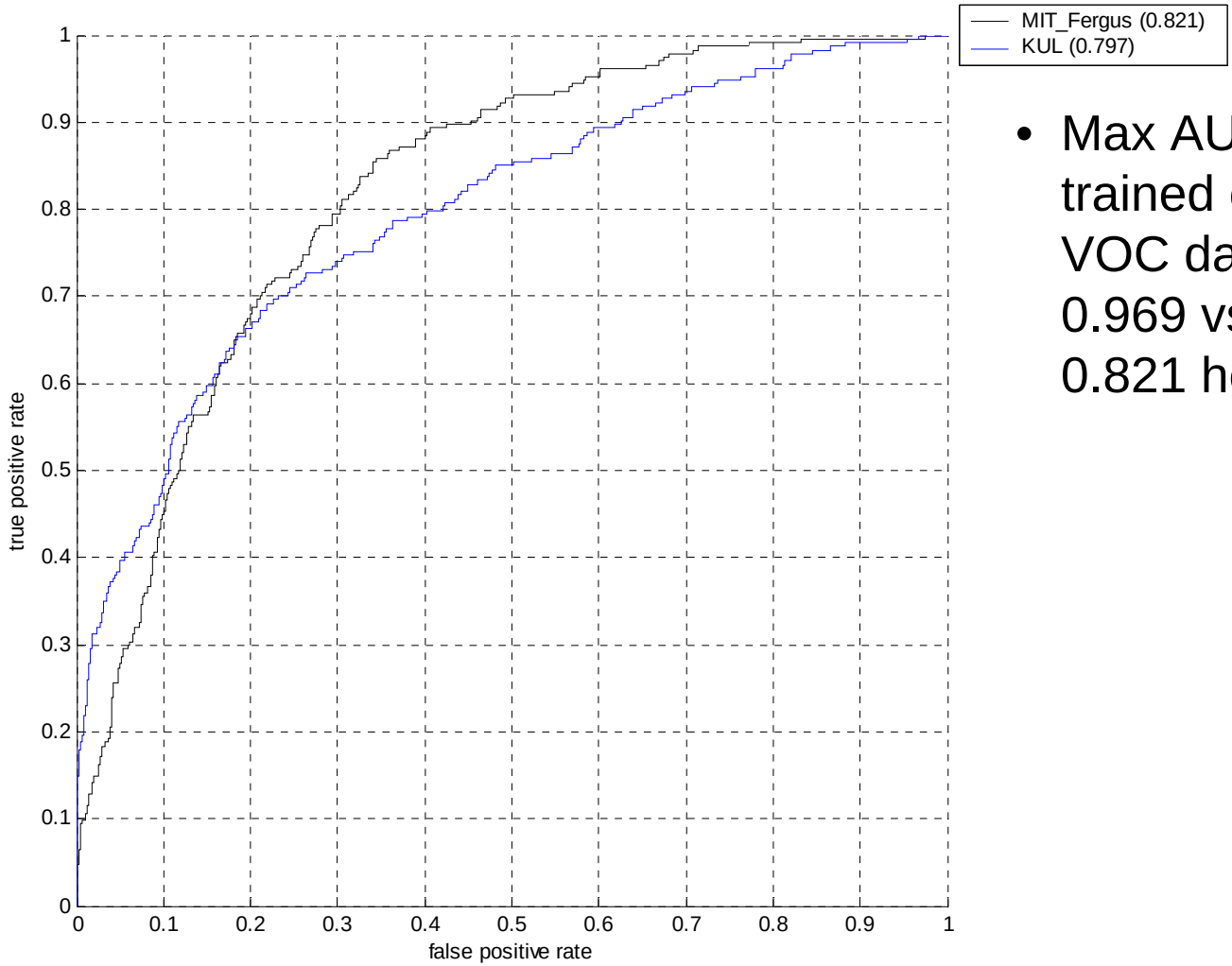
- Class “car”



- Max AUC trained on VOC data: 0.977 vs. 0.763 here

Competition 2: Train on own data

- Class “motorbike”



- Max AUC trained on VOC data: 0.969 vs. 0.821 here

Conclusions?

- Best results obtained by “bag of words” model
  - Number of small variations on basic bag of words model giving small differences in performance
  - Less diverse than VOC2005: χ^2 kernel
- Seemingly better results than VOC2005 test2
  - More balanced mix of close-up/distant images?
- Qualitative observations
  - Bias towards close-up views
  - Exploitation of context? bicycle/railings
  - Bias towards particular image composition? cats/dogs
  - Not always intuitive confusions? motorbike/bicycle

2. Detection Task

Predict bounding boxes of objects of a given class

Participants

	bicycle	bus	car	cat	cow	dog	horse	motor bike	person	sheep
AP06_Batra	–	–	–	–	–	–	–	–	–	–
AP06_Lee	–	–	–	–	–	–	–	–	–	–
Cambridge	×	×	×	×	×	×	×	×	×	×
ENSMP	–	–	×	–	×	–	–	–	–	–
INRIA_Douze	×	×	×	–	×	–	–	×	×	×
INRIA_Laptev	×	–	–	–	×	–	×	×	×	–
INRIA_Larlus	–	–	–	–	–	–	–	–	–	–
INRIA_Marszalek	–	–	–	–	–	–	–	–	–	–
INRIA_Moosmann	–	–	–	–	–	–	–	–	–	–
INRIA_Nowak	–	–	–	–	–	–	–	–	–	–
INSARouen	–	–	–	–	–	–	–	–	–	–
KUL	–	–	–	–	–	–	–	–	–	–
MIT_Fergus	–	–	–	–	–	–	–	–	–	–
MIT_Torralba	–	–	–	–	–	–	–	–	–	–
MUL	–	–	–	–	–	–	–	–	–	–
QMUL	–	–	–	–	–	–	–	–	–	–
RWTH	–	–	–	–	–	–	–	–	–	–
Siena	–	–	–	–	–	–	–	–	–	–
TKK	×	×	×	×	×	×	×	×	×	×
TUD	–	–	–	–	–	–	–	×	×	–
UVA	–	–	–	–	–	–	–	–	–	–
XRCE	–	–	–	–	–	–	–	–	–	–

Detection Results

Competition 3: Train on VOC data

AP by Method and Class

	bicycle	bus	car	cat	cow	dog	horse	motor bike	person	sheep
Cambridge	0.249	0.138	0.254	0.151	0.149	0.118	0.091	0.178	0.030	0.131
ENSMP	-	-	0.398	-	0.159	-	-	-	-	-
INRIA_Douze	0.414	0.117	0.444	-	0.212	-	-	0.390	0.164	0.251
INRIA_Laptev	0.440	-	-	-	0.224	-	0.140	0.318	0.114	-
TUD	-	-	-	-	-	-	-	0.153	0.074	-
TKK	0.303	0.169	0.222	0.160	0.252	0.113	0.137	0.265	0.039	0.227

Rank by AP per Class

	bicycle	bus	car	cat	cow	dog	horse	motor bike	person	sheep
Cambridge	4	2	3	2	5	1	3	4	5	3
ENSMP	-	-	2	-	4	-	-	-	-	-
INRIA_Douze	2	3	1	-	3	-	-	1	1	1
INRIA_Laptev	1	-	-	-	2	-	1	2	2	-
TUD	-	-	-	-	-	-	-	5	3	-
TKK	3	1	4	1	1	2	2	3	4	2

Programme

- 15:00 - 15:30 **Overview of the 2006 VOC Challenge.**
  Mark Everingham, University of Oxford.
- 15:30 - 15:55 **TextonBoost: Joint Appearance, Shape and Context Modeling for Multi-Class Object Recognition and Segmentation.**
  John Winn, Microsoft Research Cambridge.
- 15:55 - 16:20 **Object Detection using Histograms of Oriented Gradients.**
  Navneet Dalal, INRIA Rhones-Alpes.

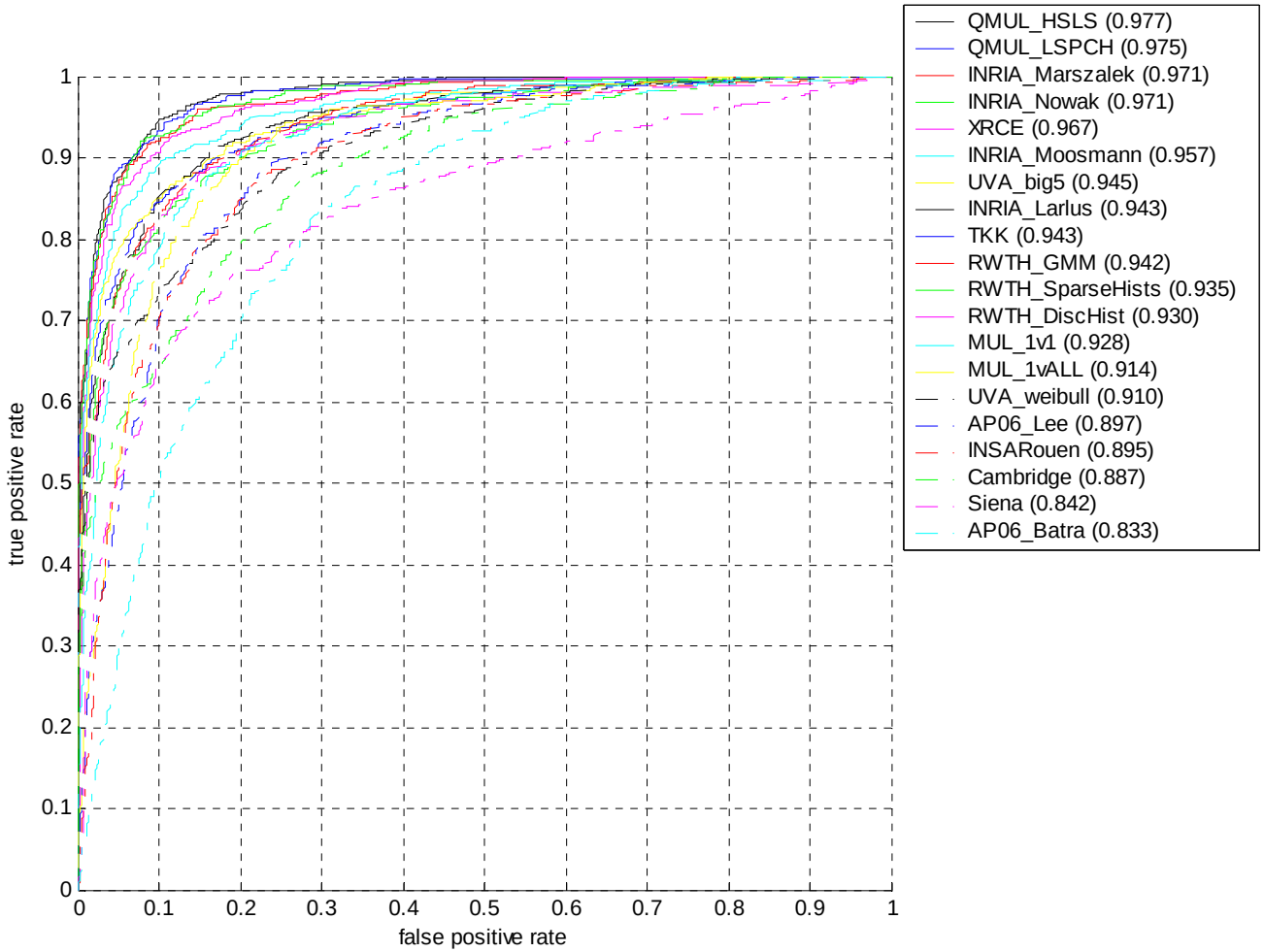
- 16:20 - 16:35 *Break.*

- 16:35 - 17:00 **Local Features and Kernels for Classification of Object Categories.**
  Jianguo Zhang, Queen Mary University of London.
- 17:00 - 17:30 **The MUSCLE / ImageCLEF Image Retrieval Evaluation Campaigns.**
  Allan Hanbury, Vienna University of Technology.
- 17:30 - 18:00 **Conclusions & Discussion.**

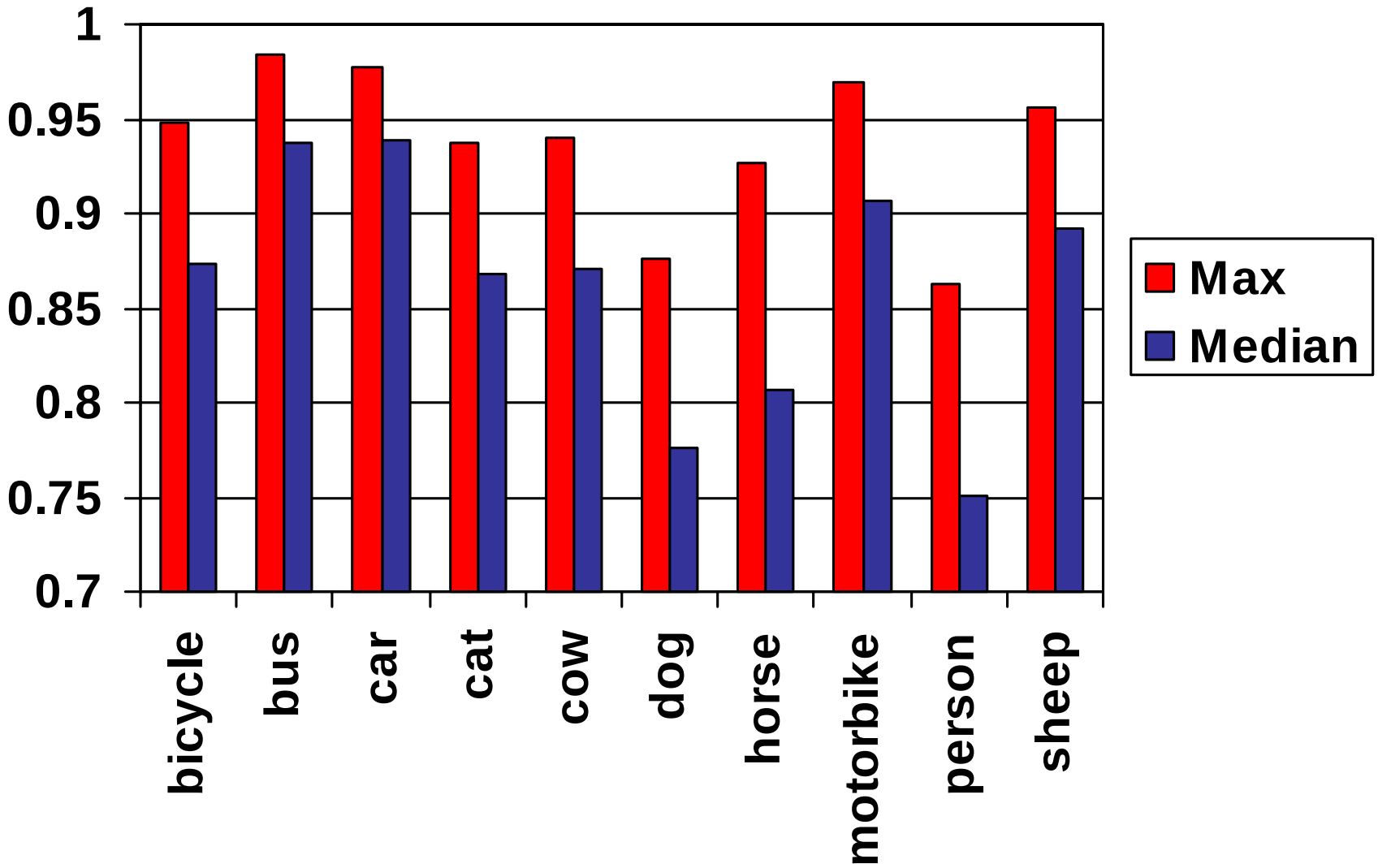
Recap: Classification Task

Competition 1: Car

- All methods



AUC by Class


1
0.95
0.9
0.85
0.8
0.75
0.7
bicycle
bus
car
cat
cow
dog
horse
motorbike
person
sheep
Max
Median

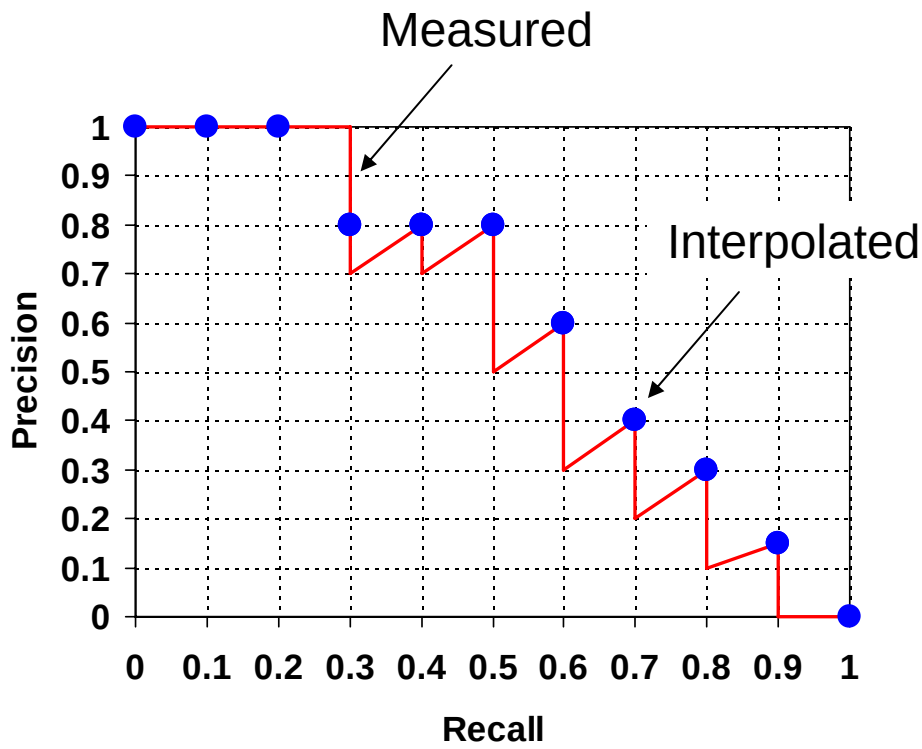
2. Detection Task

Predict bounding boxes of objects of a given class

Evaluation

- Correct detection: 50% overlap in bounding boxes
  - Multiple detections considered as (one true +) false positives

- Precision/Recall
  - Average Precision (AP) as defined by TREC
    - Mean precision interpolated at recall = 0,0.1,…,0.9,1



Methods

- Sliding-window classifiers
  - Assign confidence to windows over an image pyramid
  - Non-maximum suppression to obtain bounding boxes
  - Multi-view methods: one classifier per view
  - Classifiers and features
    - Linear SVM classifier with SIFT-like spatial/orientation histogram
    - Boosted classifier with spatial orientation histogram features
    - Boosted classifier with pixel-level features
    - Boosted classifier with template correlation features shared across views

Methods

- Generalized Hough transform
  - Vector-quantized regions around interest points “vote” for centre of object by a non-parametric distribution
  - Single-view and multi-view methods
    - Multi-view method reinforces votes by “overlapping” views
- Constellation model
  - Gaussian distribution with full covariance over position and appearance of small number of regions, detected at interest points

Methods

- Pixel-wise classification
  - Boosted classifier assigns class to each pixel based on spatial neighbourhood. Bounding boxes are derived from connected components of same class
- Classification of segmented regions
  - Regions from a segmentation algorithm classified and bounding boxes derived from region classification
  - Region confidence is combination of global (image) and local (region) confidence

Detection Results

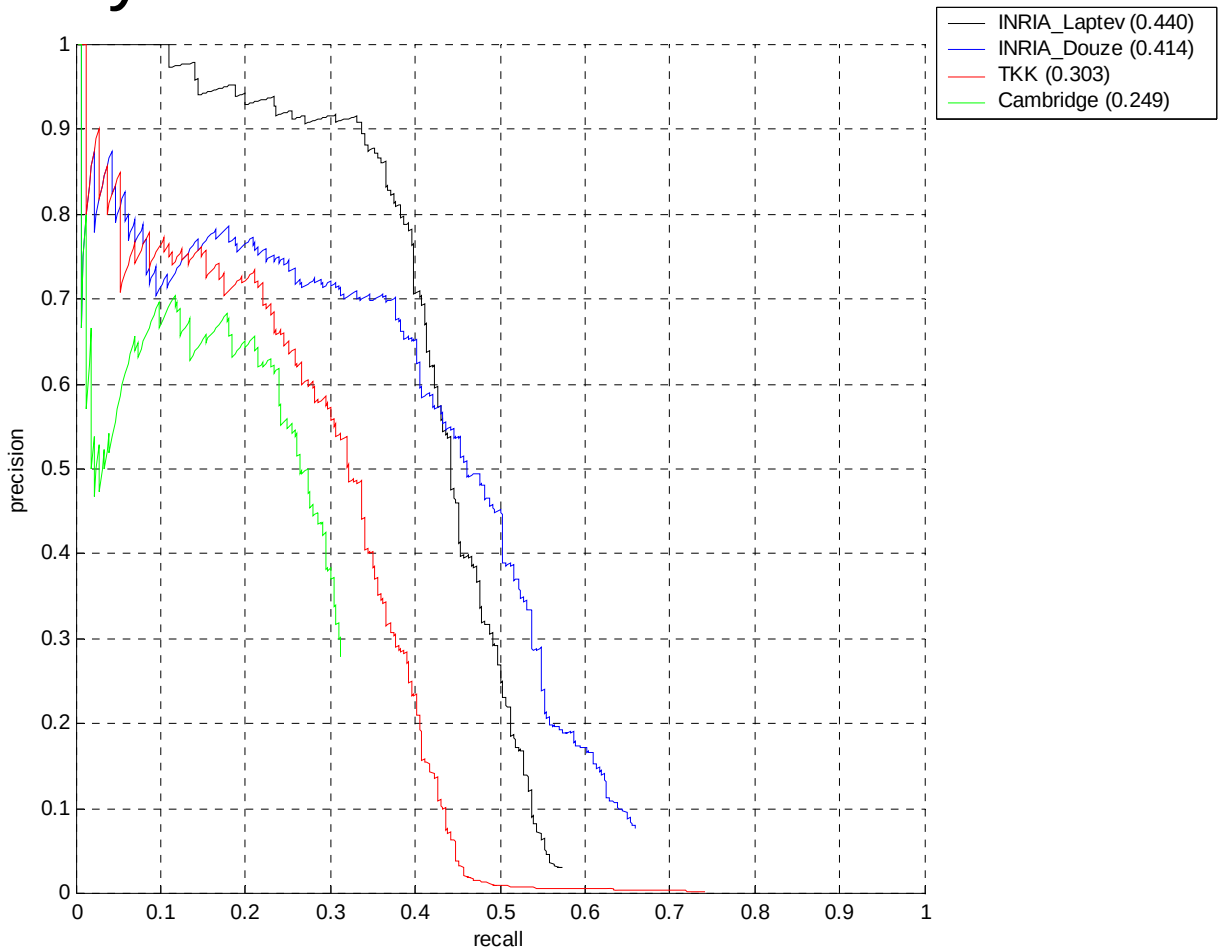
Competition 3: Train on VOC data

Participants

	bicycle	bus	car	cat	cow	dog	horse	motor bike	person	sheep
AP06_Batra	–	–	–	–	–	–	–	–	–	–
AP06_Lee	–	–	–	–	–	–	–	–	–	–
Cambridge	×	×	×	×	×	×	×	×	×	×
ENSMP	–	–	×	–	×	–	–	–	–	–
INRIA_Douze	×	×	×	–	×	–	–	×	×	×
INRIA_Laptev	×	–	–	–	×	–	×	×	×	–
INRIA_Larlus	–	–	–	–	–	–	–	–	–	–
INRIA_Marszalek	–	–	–	–	–	–	–	–	–	–
INRIA_Moosmann	–	–	–	–	–	–	–	–	–	–
INRIA_Nowak	–	–	–	–	–	–	–	–	–	–
INSARouen	–	–	–	–	–	–	–	–	–	–
KUL	–	–	–	–	–	–	–	–	–	–
MIT_Fergus	–	–	–	–	–	–	–	–	–	–
MIT_Torralba	–	–	–	–	–	–	–	–	–	–
MUL	–	–	–	–	–	–	–	–	–	–
QMUL	–	–	–	–	–	–	–	–	–	–
RWTH	–	–	–	–	–	–	–	–	–	–
Siena	–	–	–	–	–	–	–	–	–	–
TKK	×	×	×	×	×	×	×	×	×	×
TUD	–	–	–	–	–	–	–	×	×	–
UVA	–	–	–	–	–	–	–	–	–	–
XRCE	–	–	–	–	–	–	–	–	–	–

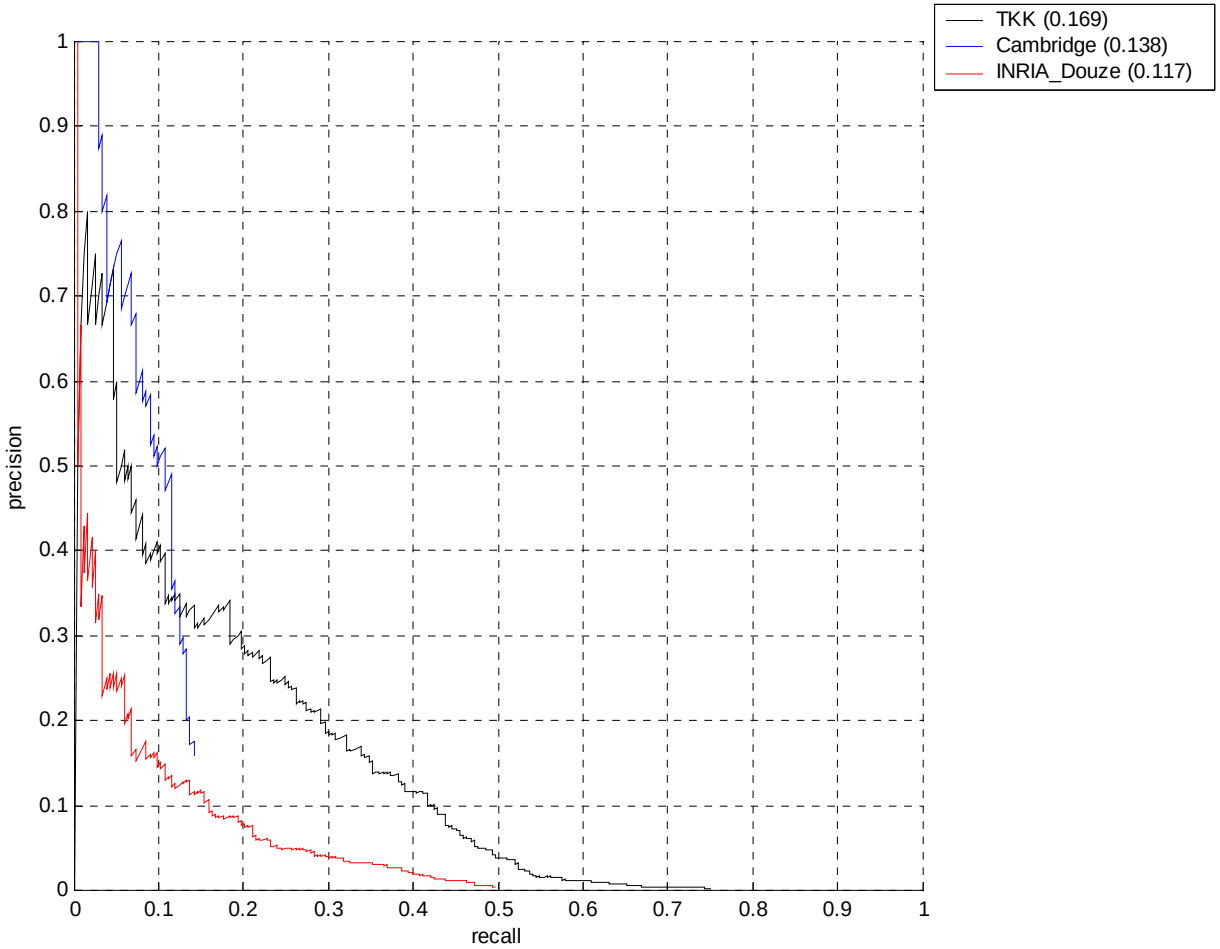
Competition 3: Train on VOC data

- Class “bicycle”



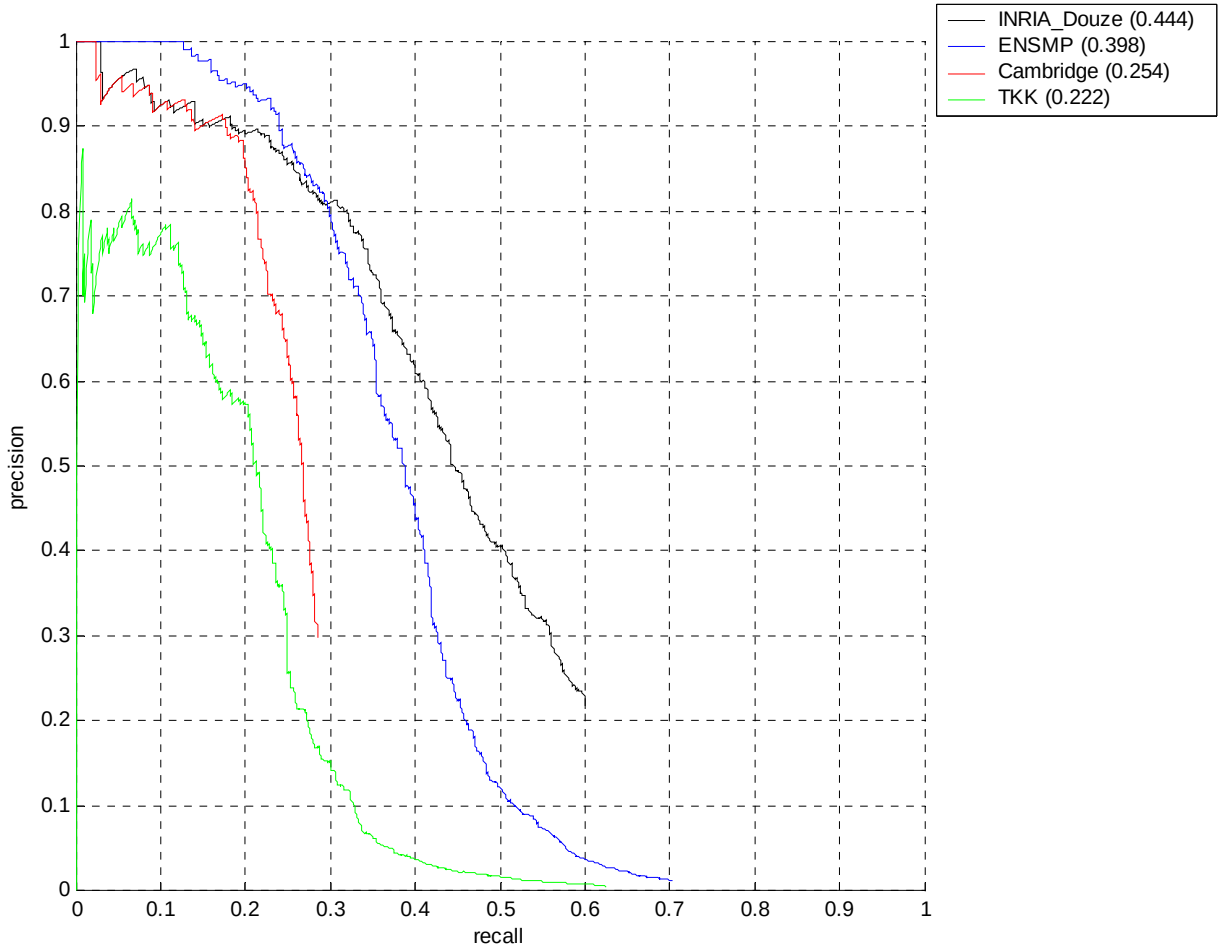
Competition 3: Train on VOC data

- Class “bus”



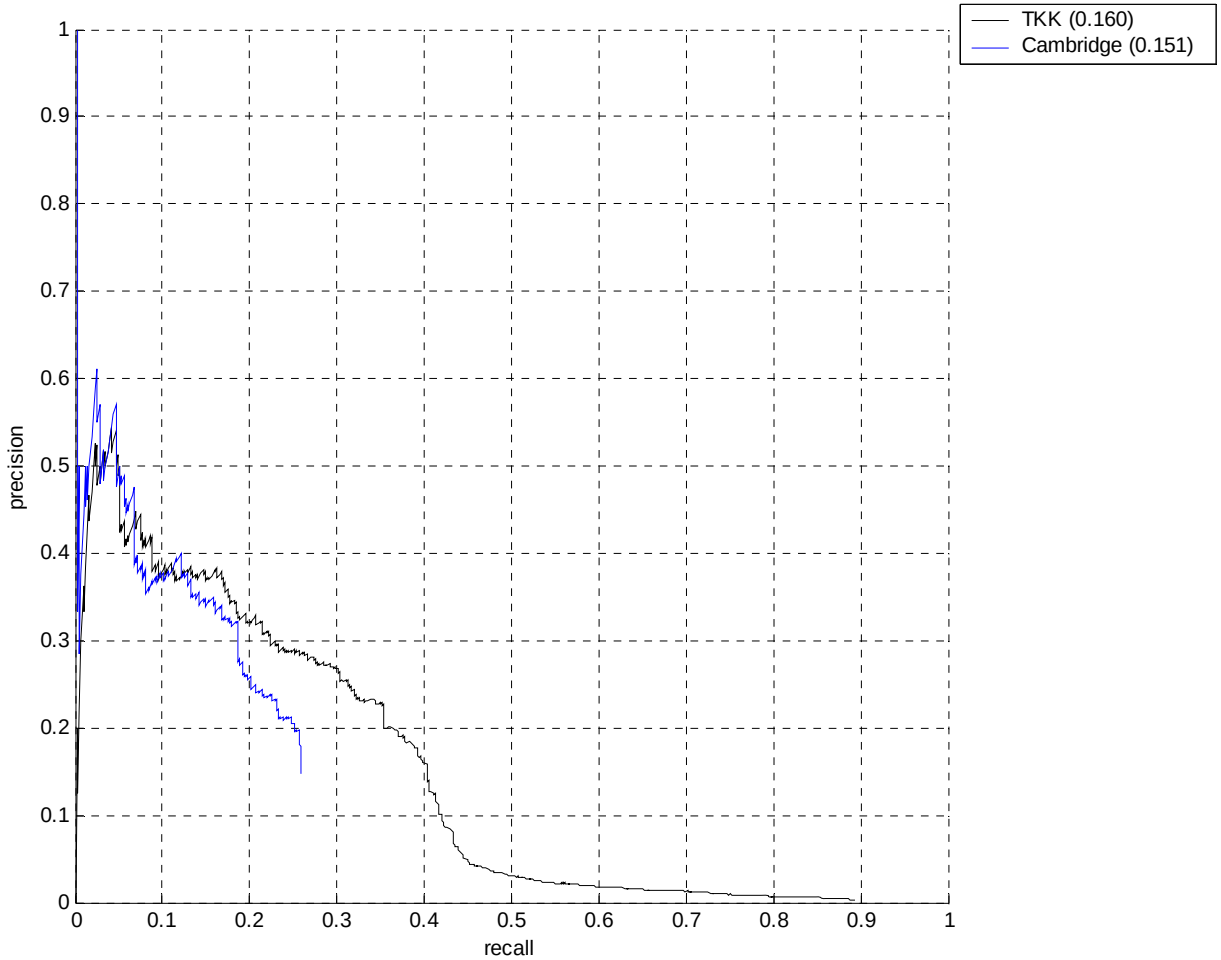
Competition 3: Train on VOC data

- Class “car”



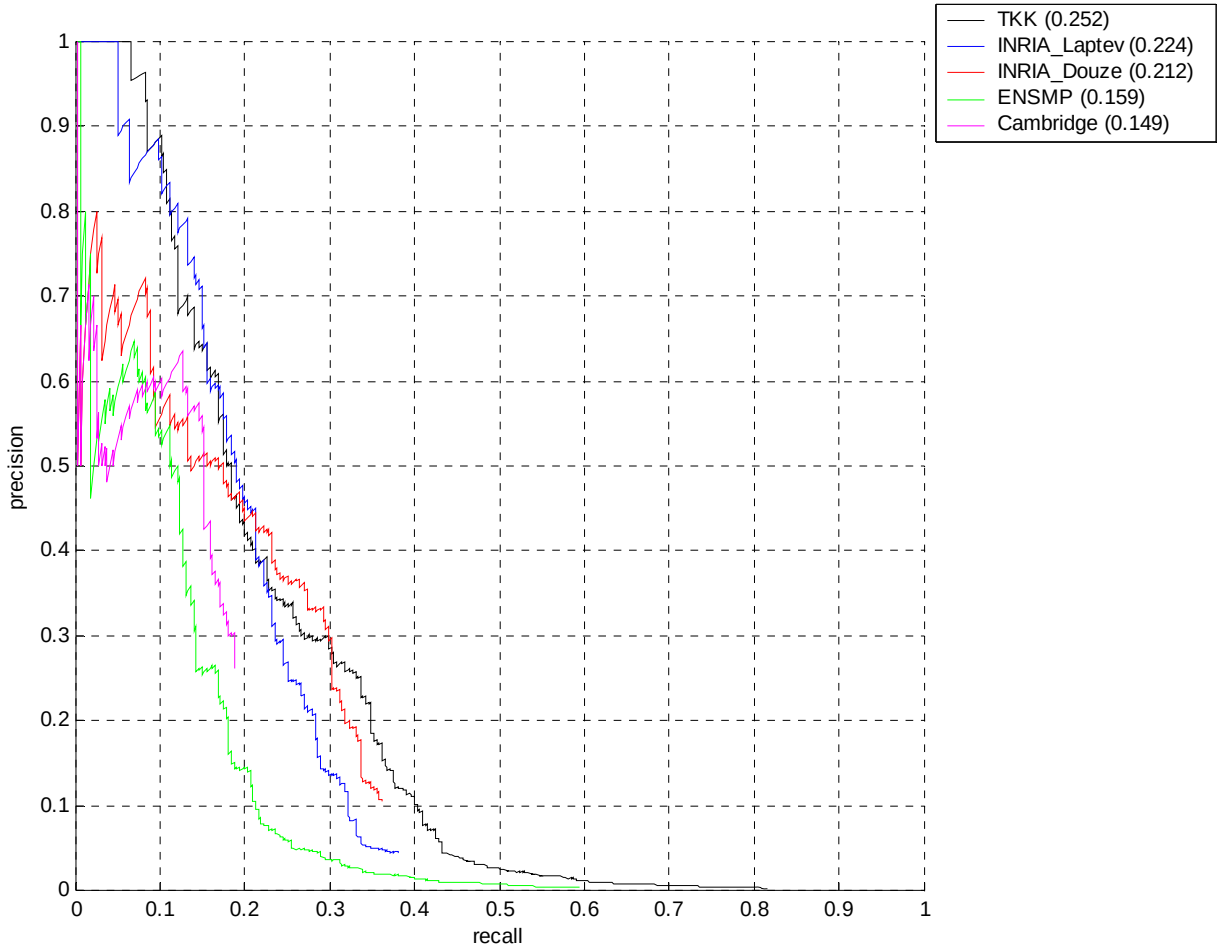
Competition 3: Train on VOC data

- Class “cat”



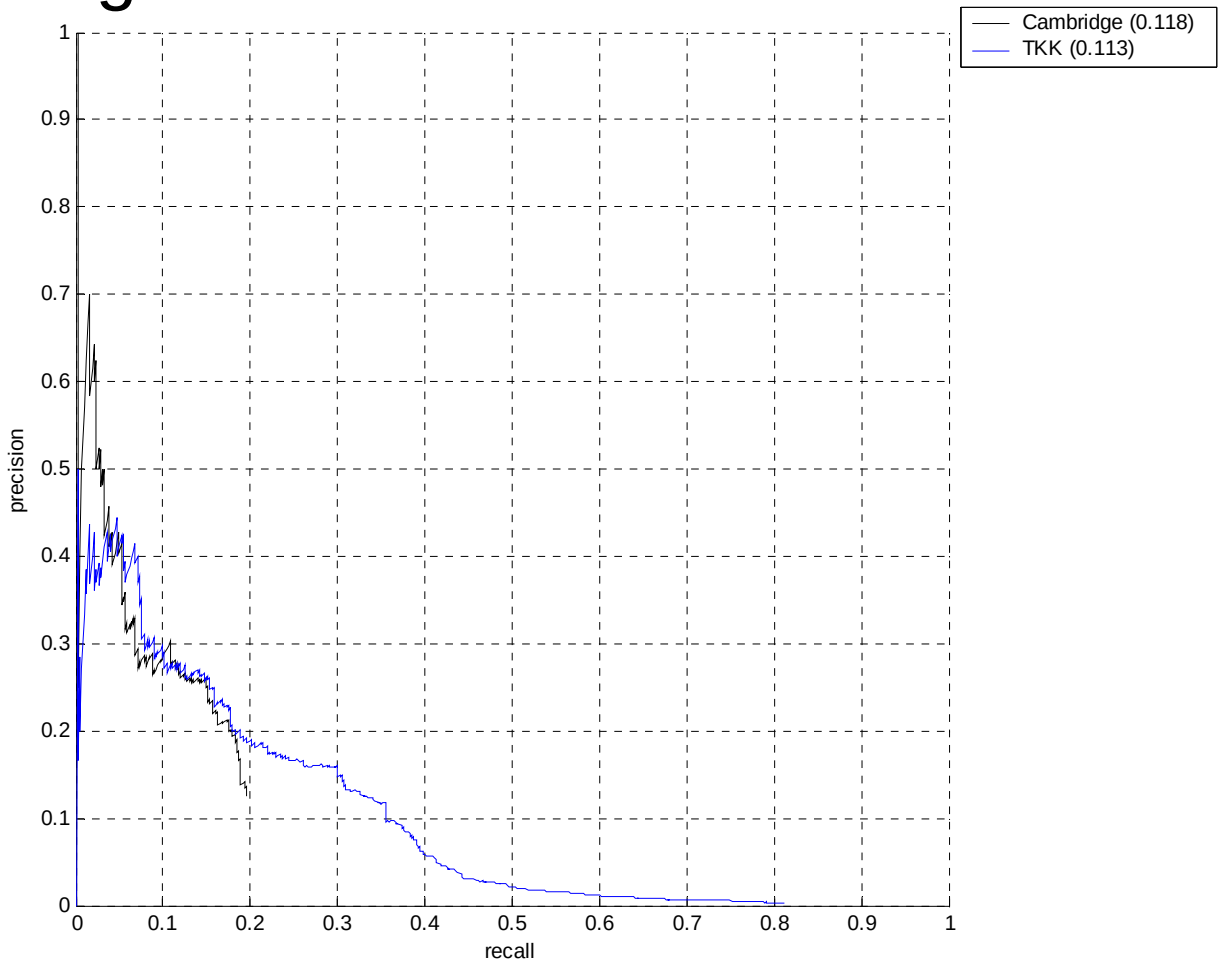
Competition 3: Train on VOC data

- Class “cow”



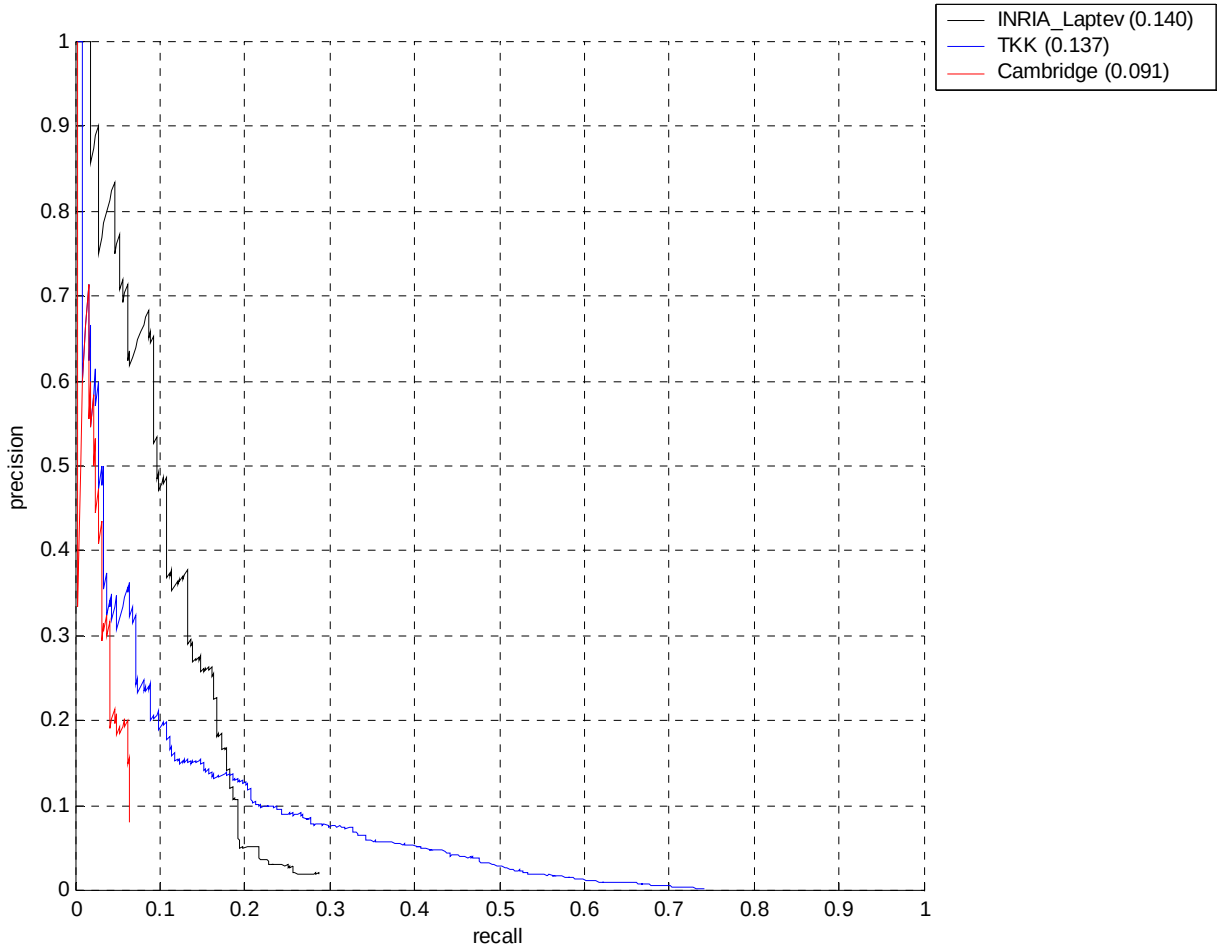
Competition 3: Train on VOC data

- Class “dog”



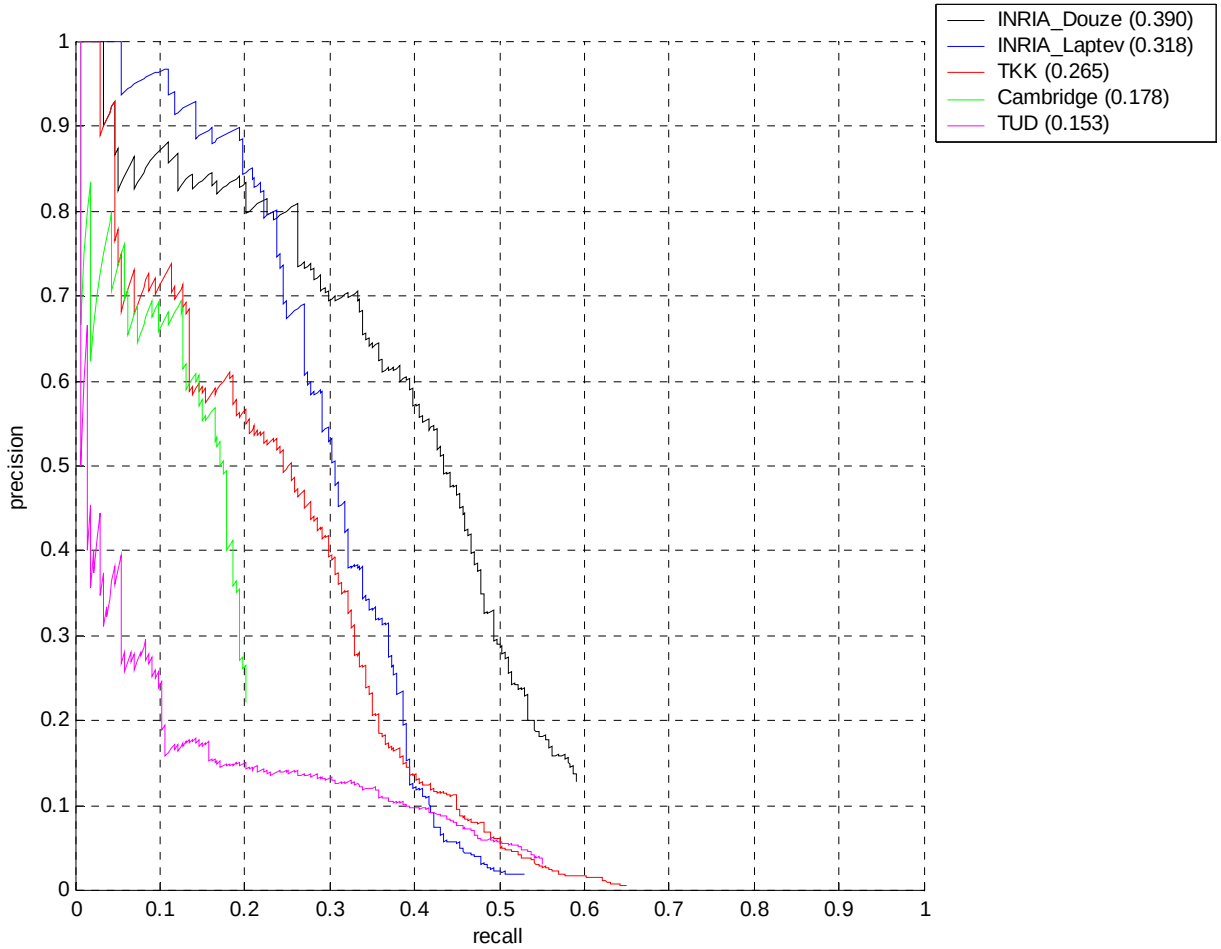
Competition 3: Train on VOC data

- Class “horse”



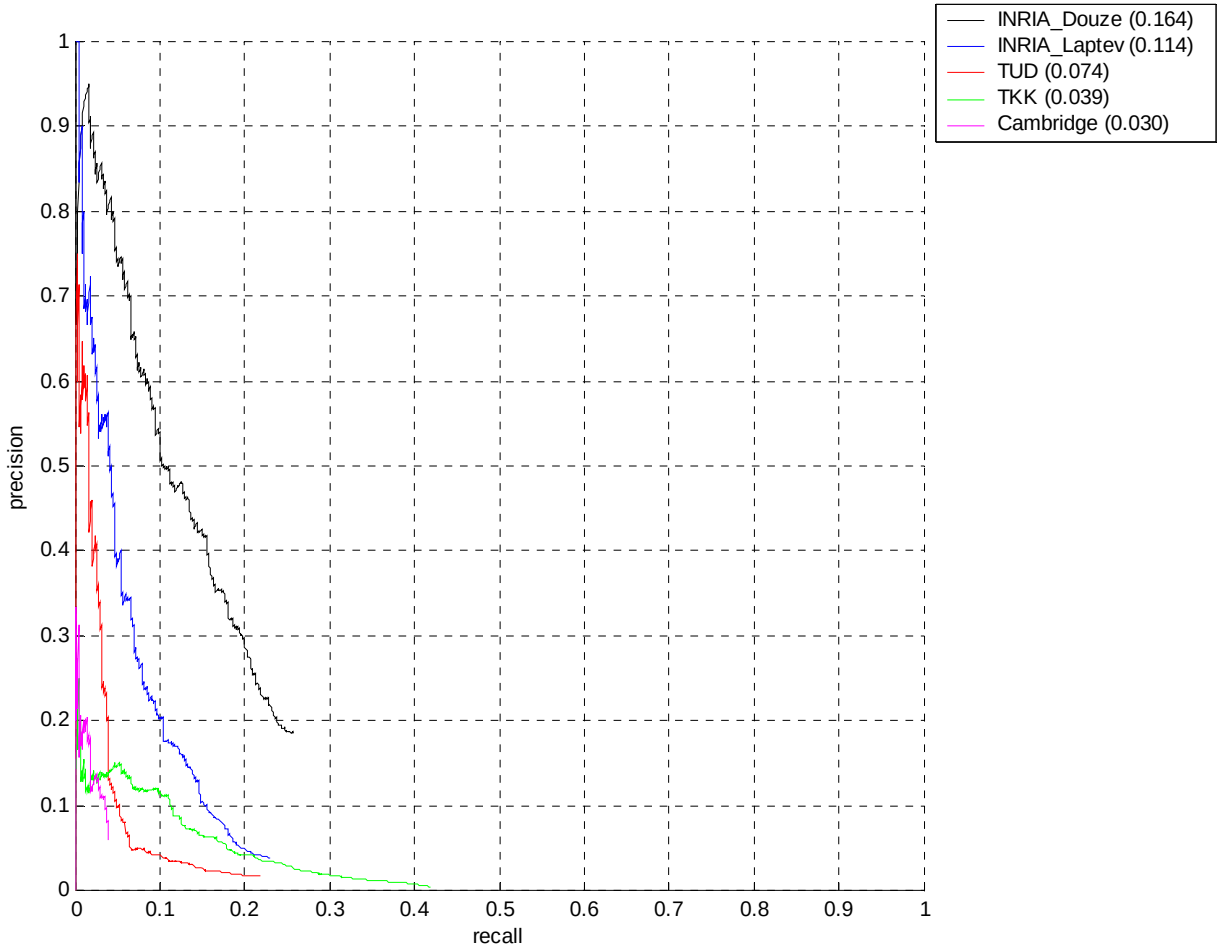
Competition 3: Train on VOC data

- Class “motorbike”



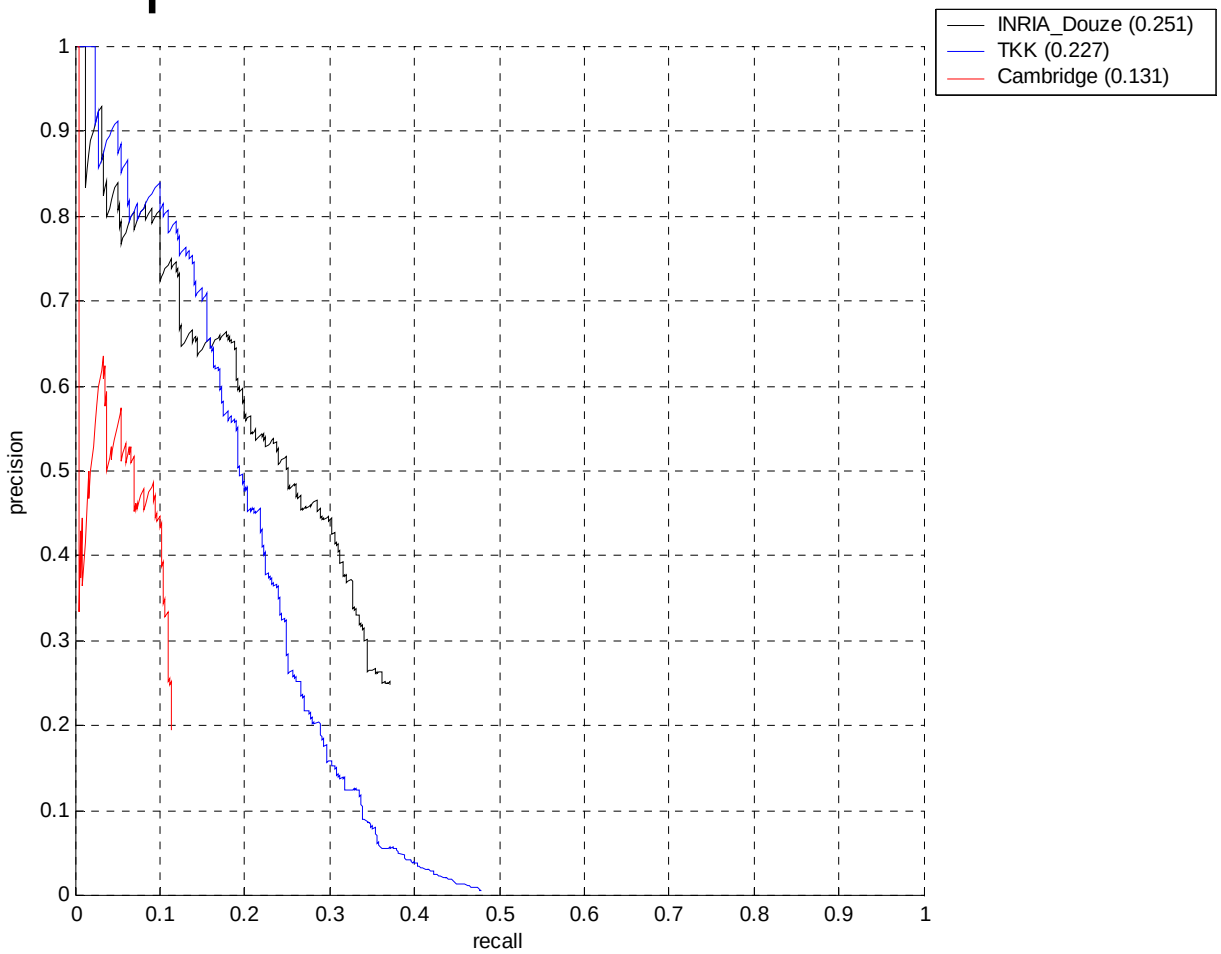
Competition 3: Train on VOC data

- Class “person”



Competition 3: Train on VOC data

- Class “sheep”



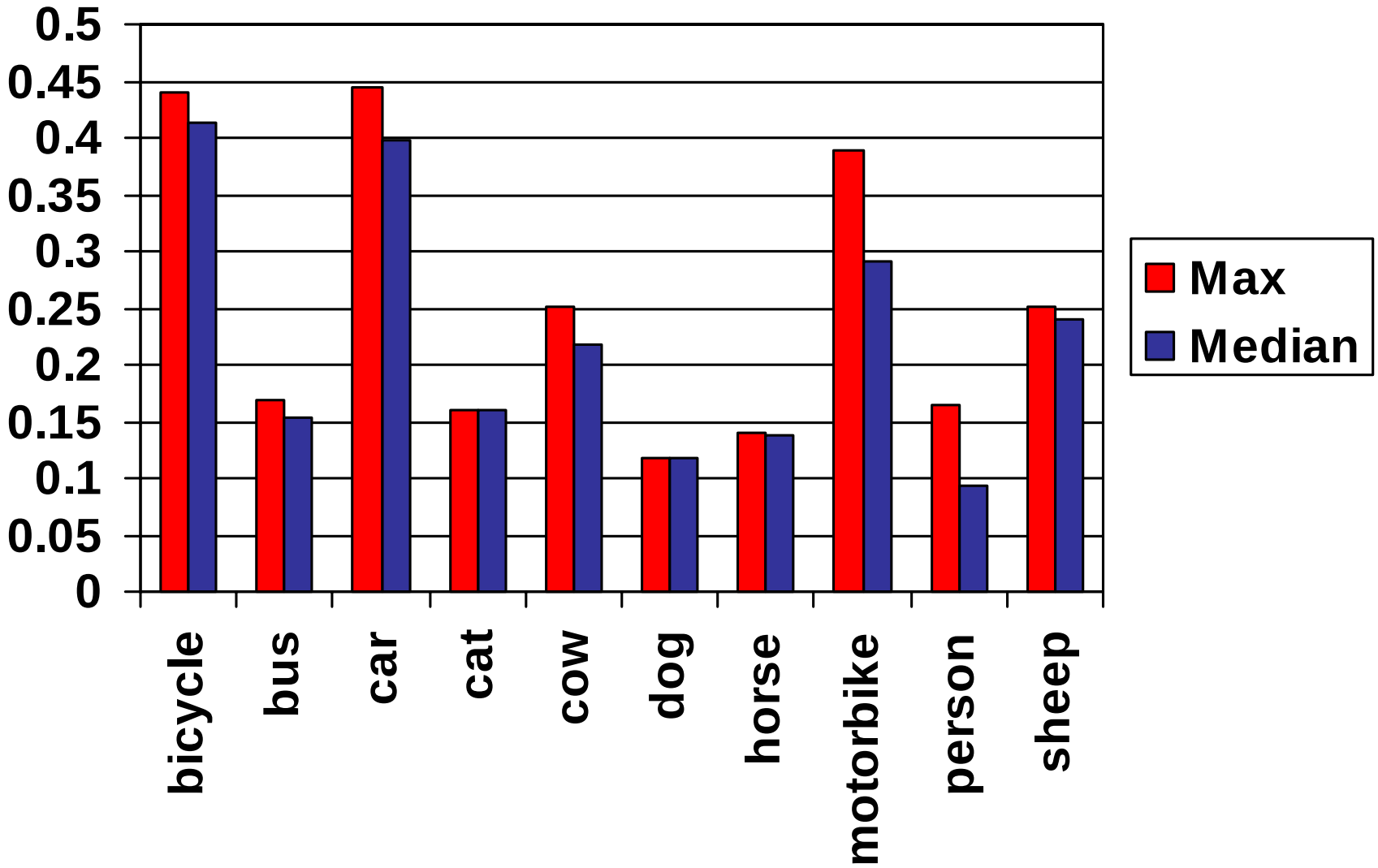
AP by Method and Class

	bicycle	bus	car	cat	cow	dog	horse	motor bike	person	sheep
Cambridge	0.249	0.138	0.254	0.151	0.149	0.118	0.091	0.178	0.030	0.131
ENSMP	-	-	0.398	-	0.159	-	-	-	-	-
INRIA_Douze	0.414	0.117	0.444	-	0.212	-	-	0.390	0.164	0.251
INRIA_Laptev	0.440	-	-	-	0.224	-	0.140	0.318	0.114	-
TUD	-	-	-	-	-	-	-	0.153	0.074	-
TKK	0.303	0.169	0.222	0.160	0.252	0.113	0.137	0.265	0.039	0.227

Rank by AP per Class

	bicycle	bus	car	cat	cow	dog	horse	motor bike	person	sheep
Cambridge	4	2	3	2	5	1	3	4	5	3
ENSMP	-	-	2	-	4	-	-	-	-	-
INRIA_Douze	2	3	1	-	3	-	-	1	1	1
INRIA_Laptev	1	-	-	-	2	-	1	2	2	-
TUD	-	-	-	-	-	-	-	5	3	-
TKK	3	1	4	1	1	2	2	3	4	2

AP by Class


0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
bicycle
bus
car
cat
cow
dog
horse
motorbike
person
sheep
Max
Median

Detection Results

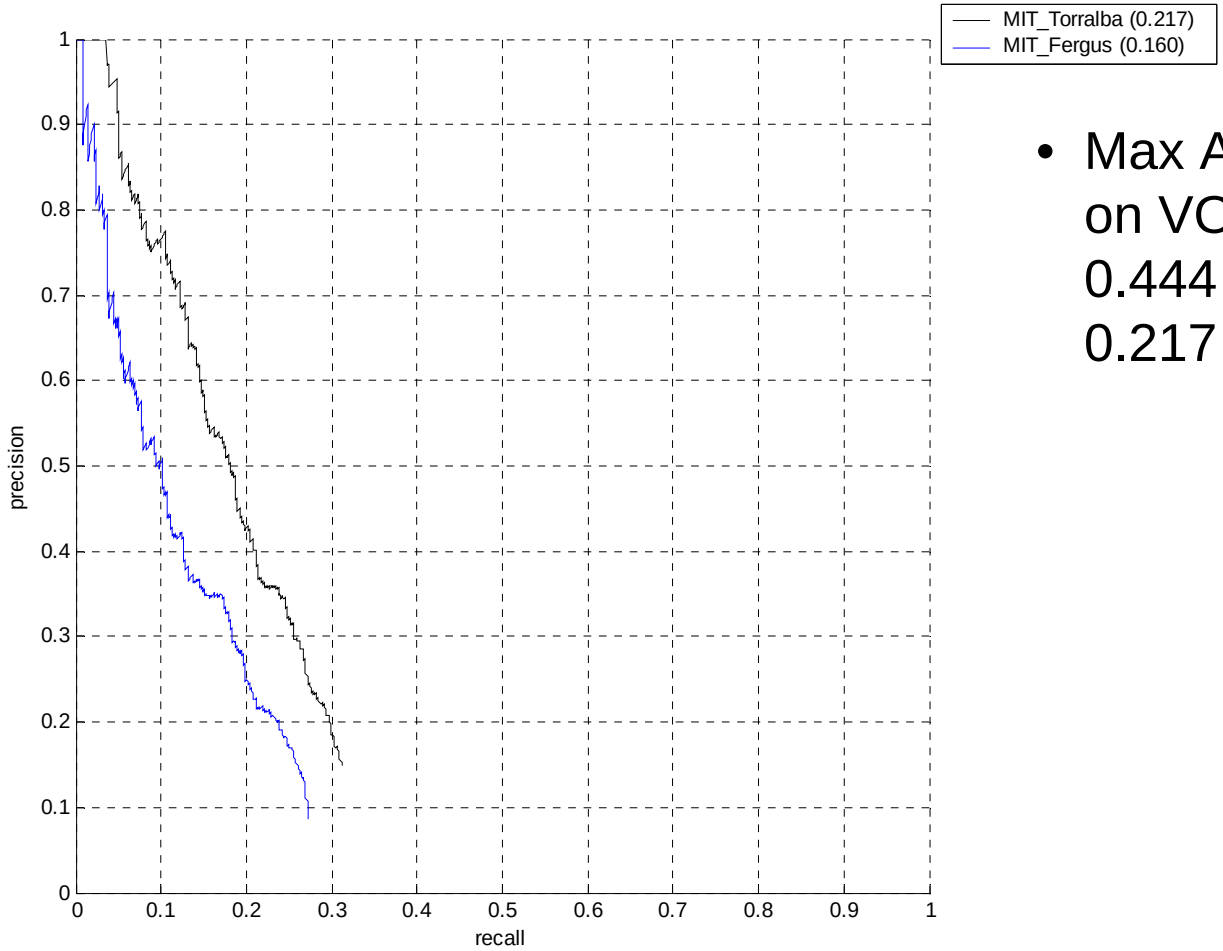
Competition 4: Train on own data

Participants

	bicycle	bus	car	cat	cow	dog	horse	motor bike	person	sheep
AP06_Batra	–	–	–	–	–	–	–	–	–	–
AP06_Lee	–	–	–	–	–	–	–	–	–	–
Cambridge	–	–	–	–	–	–	–	–	–	–
ENSMP	–	–	–	–	–	–	–	–	–	–
INRIA_Douze	–	–	–	–	–	–	–	–	×	–
INRIA_Laptev	–	–	–	–	–	–	–	–	–	–
INRIA_Larlus	–	–	–	–	–	–	–	–	–	–
INRIA_Marszalek	–	–	–	–	–	–	–	–	–	–
INRIA_Moosmann	–	–	–	–	–	–	–	–	–	–
INRIA_Nowak	–	–	–	–	–	–	–	–	–	–
INSARouen	–	–	–	–	–	–	–	–	–	–
KUL	–	–	–	–	–	–	–	×	–	–
MIT_Fergus	–	–	×	–	–	–	–	×	–	–
MIT_Torralba	–	–	×	–	–	–	–	–	–	–
MUL	–	–	–	–	–	–	–	–	–	–
QMUL	–	–	–	–	–	–	–	–	–	–
RWTH	–	–	–	–	–	–	–	–	–	–
Siena	–	–	–	–	–	–	–	–	–	–
TKK	–	–	–	–	–	–	–	–	–	–
TUD	–	–	–	–	–	–	–	–	–	–
UVA	–	–	–	–	–	–	–	–	–	–
XRCE	–	–	–	–	–	–	–	–	–	–

Competition 4: Train on own data

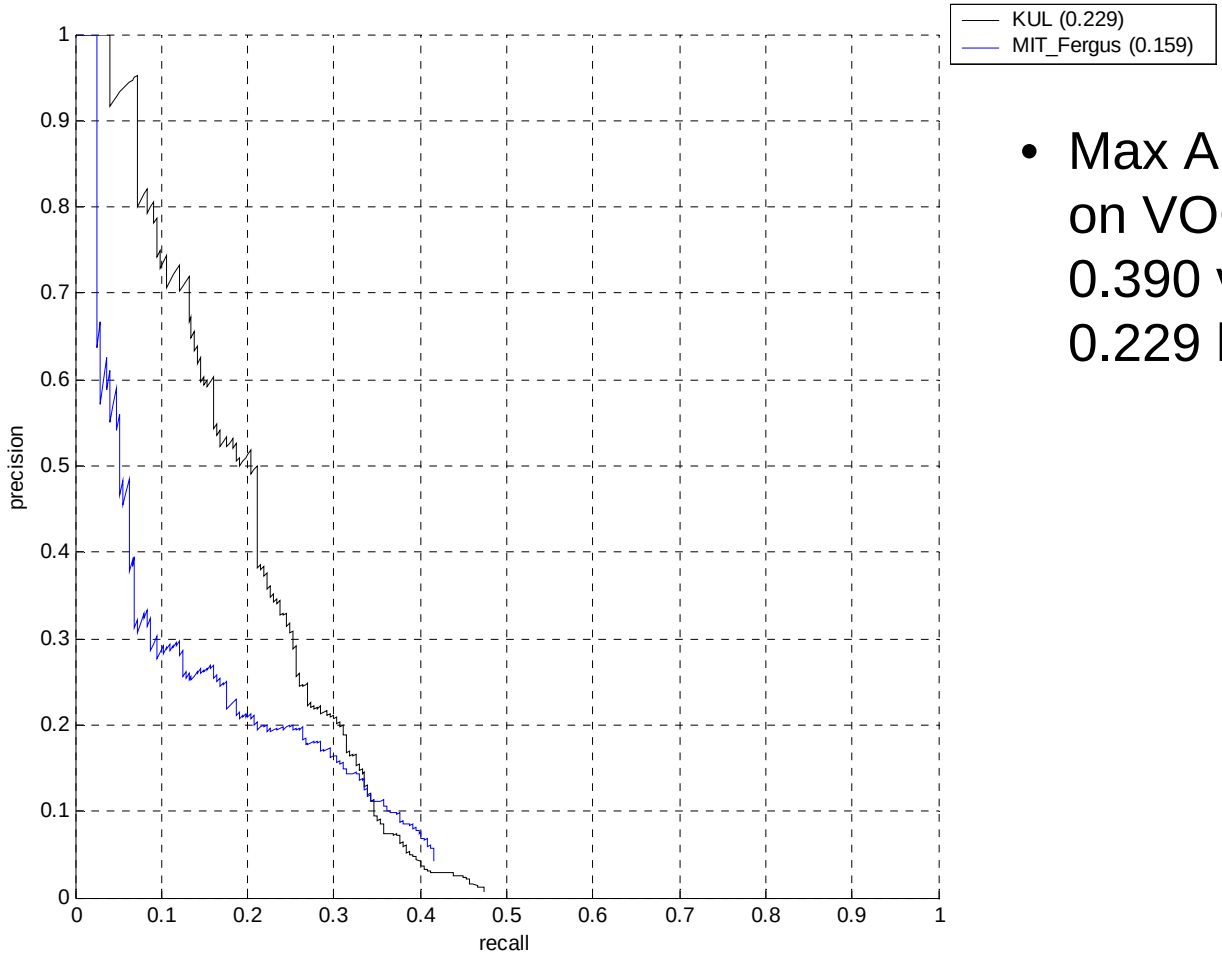
- Class “car”



- Max AP trained on VOC data: 0.444 vs. 0.217 here

Competition 4: Train on own data

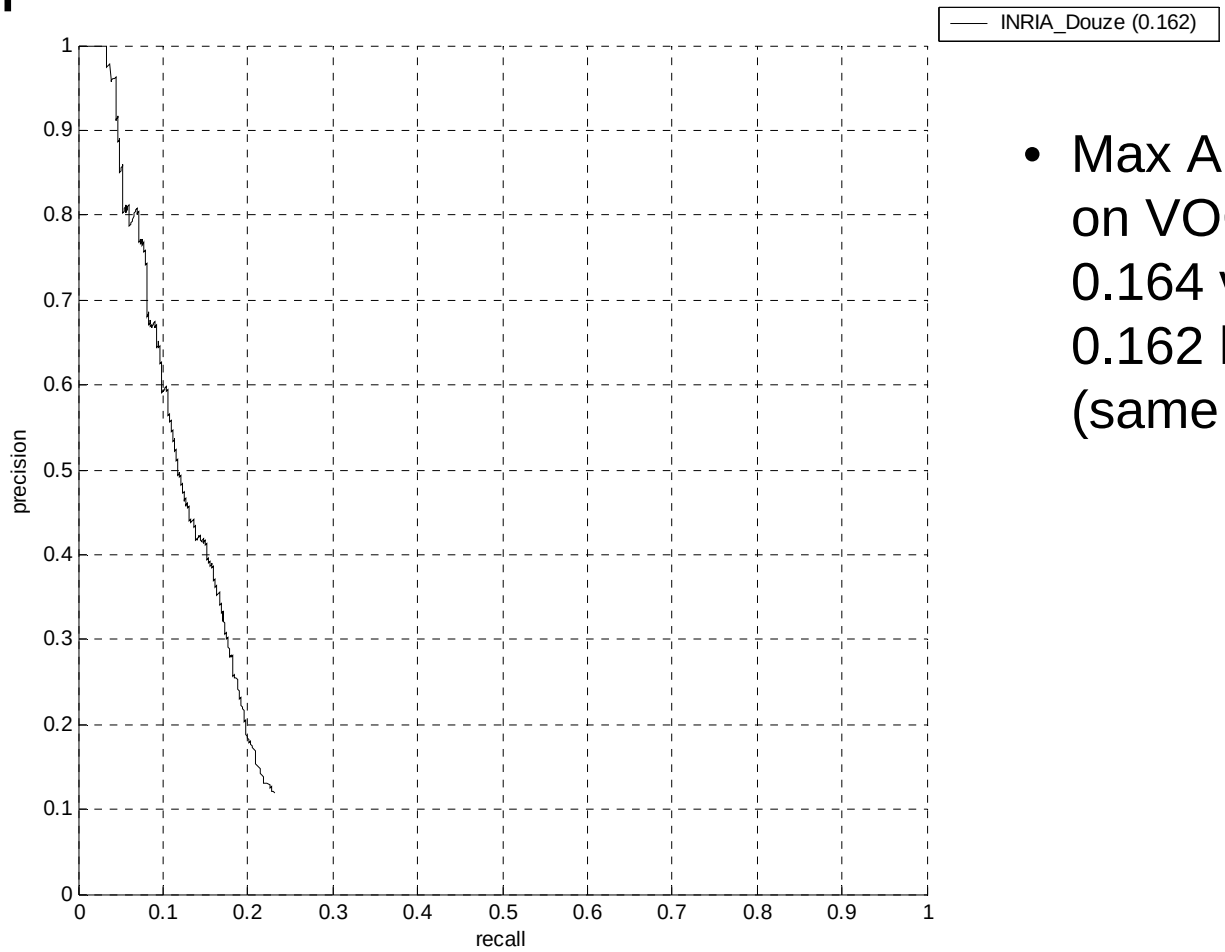
- Class “motorbike”



- Max AP trained on VOC data: 0.390 vs. 0.229 here

Competition 4: Train on own data

- Class “person”



- Max AP trained on VOC data: 0.164 vs. 0.162 here (same method)

Conclusions?

- Much more challenging than classification task
- No overall winner but sliding-window methods tended to give best results
- Generalized Hough transform method gave poor results compared to VOC2005
  - Greater viewpoint variation? Lack of SVM stage?
- For “person” class, use of own training data changed results little cf. VOC2005
  - Sufficiently large training set? Extremely difficult?

Overall Conclusions?

- **Classification**: Variety of methods with predominance of "bag of words"
  - Some re-introduction of spatial information
- Results on less rigid/non-manmade classes (dogs, people) worse then "traditional" cars, motorbikes
  - Bias towards classes with distinctive local appearance?
- Hard to distinguish between many classification methods
  - Usefulness of this task exhausted?
- **Detection**: Sliding-window methods gave better results than more explicit modelling of object "parts"
- Still much progress to be made
  - Unconstrained viewpoint etc. remain very challenging