The PASCAL Visual Object Classes Challenge 2009 (VOC2009)

Part 2 – Classification Task

Mark Everingham
Luc Van Gool
Chris Williams
John Winn
Andrew Zisserman

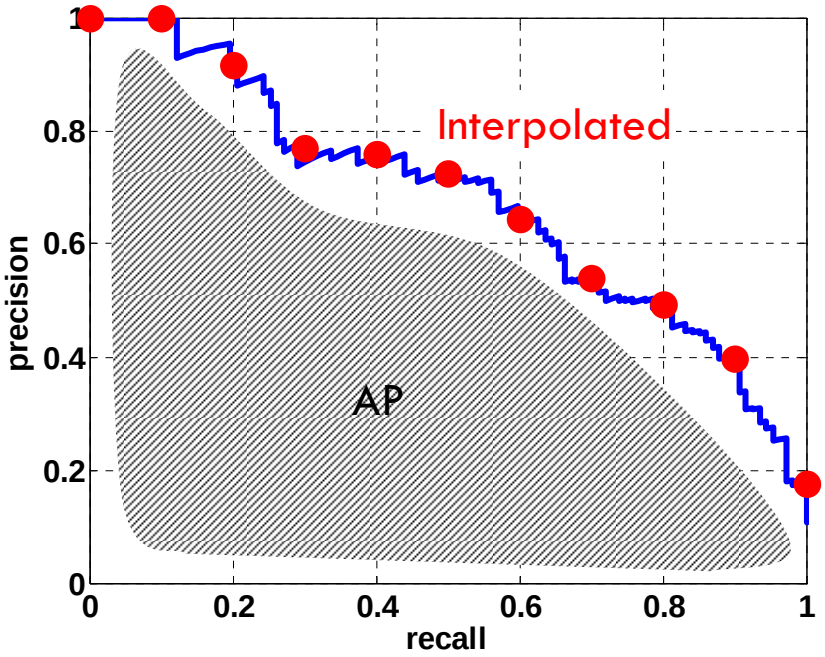
Classification Challenge

- Predict whether at least one object of a given class is present in an image

- Competition 1: Train on the supplied data
  - Which methods perform best given specified training data?
- Competition 2: Train on any (non-test) data
  - How well do state-of-the-art methods perform on these problems?

Evaluation

- **Average Precision [TREC]** averages precision over the entire range of recall
  - Curve interpolated to reduce influence of "outliers"



- A good score requires both high recall **and** high precision
- Application-independent

Participation

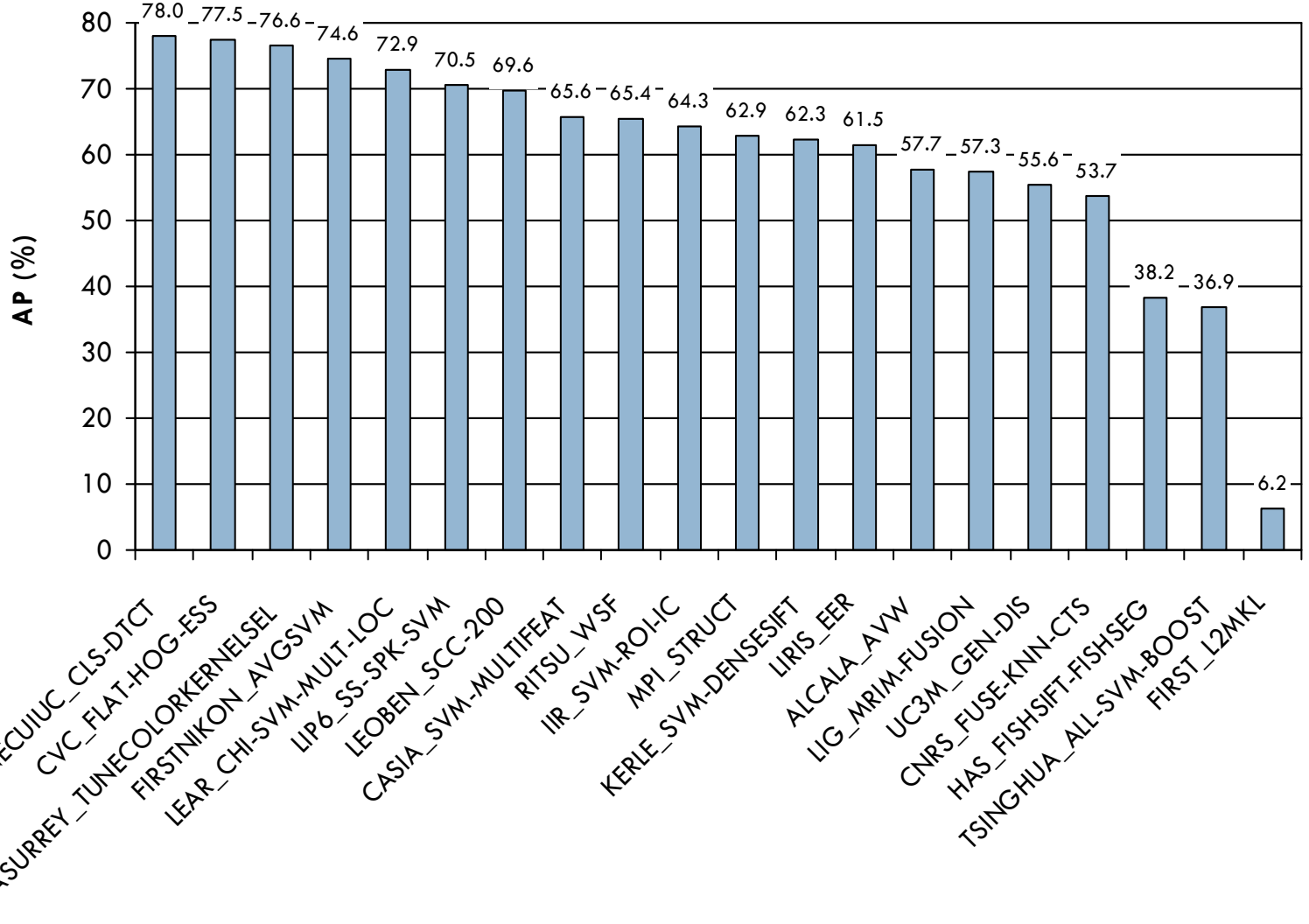
- 48 Methods, 20 Groups
- VOC2008: 21 Methods, 11 Groups
- Overwhelmingly "bag of visual words" methods with multiple features e.g. SIFT, color
- Multiple submissions of methods with small variations e.g. different features or classifier architectures

Results: AP by Method and Class

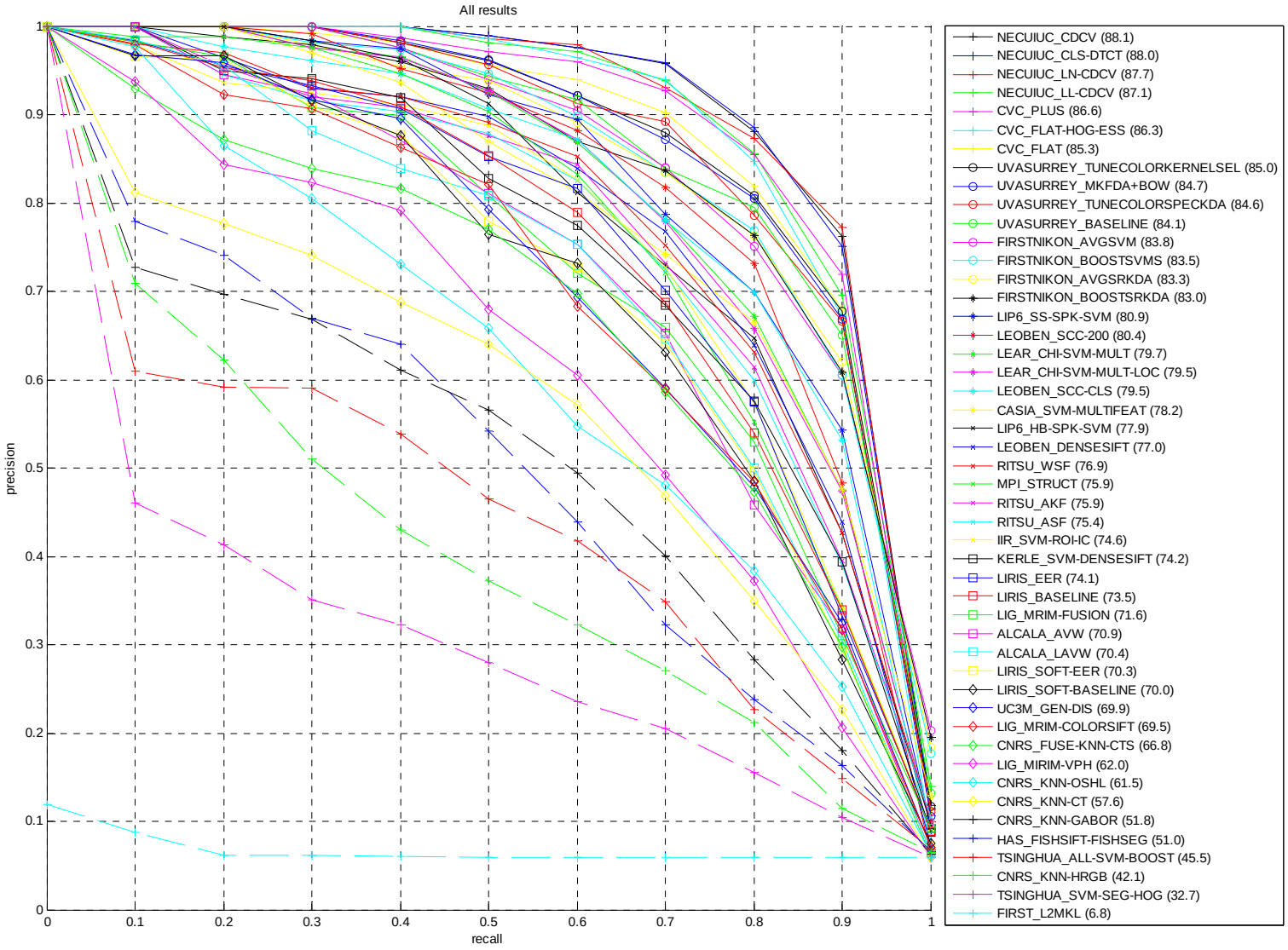
	aero plane	bicycle	bird	boat	bottle	bus	car	cat	chair	cow	dining table	dog	horse	motor bike	person	potted plant	sheep	sofa	train	tv/ monitor
CVC_FLAT	85.3	57.8	66.0	66.1	36.2	70.6	60.6	63.5	55.1	44.6	53.4	49.1	64.4	66.8	84.8	37.4	44.1	47.9	81.9	67.5
CVC_FLAT-HOG-ESS	86.3	60.7	66.4	65.3	41.0	71.7	64.7	63.9	55.5	40.1	51.3	45.9	65.2	68.9	85.0	40.8	49.0	49.1	81.8	68.6
CVC_PLUS	86.6	58.4	66.7	67.3	34.8	70.4	60.0	64.2	52.5	43.0	50.8	46.5	64.1	66.8	84.4	37.5	45.1	45.4	82.1	67.0
FIRSTNIKON_AVGSRKDA	83.3	59.3	62.7	65.3	30.2	71.6	58.2	62.2	54.3	40.7	49.2	50.0	66.6	62.9	83.3	34.2	48.2	46.1	83.4	65.5
FIRSTNIKON_AVGSVM	83.8	58.2	62.6	65.2	32.0	69.8	57.7	61.1	54.5	44.0	50.3	49.6	64.6	61.7	83.2	33.4	46.5	48.0	81.6	65.3
FIRSTNIKON_BOOSTSRKDA	83.0	59.2	61.4	64.6	33.2	71.1	57.5	61.0	54.8	40.7	48.3	50.0	65.5	63.4	82.8	32.8	47.0	47.1	83.3	64.6
FIRSTNIKON_BOOSTSVMS	83.5	56.8	61.8	65.5	33.2	69.7	57.3	60.5	54.6	43.1	48.3	50.3	64.3	62.4	82.3	32.9	46.9	48.4	82.0	64.2
LEAR_CHI-SVM-MULT-LOC	79.5	55.5	54.5	63.9	43.7	70.3	66.4	56.5	54.4	38.8	44.1	46.2	58.5	64.2	82.2	39.1	41.3	39.8	73.6	66.2
NECUIUC_CDCV	88.1	68.0	68.0	72.5	41.0	78.9	70.4	70.4	58.1	53.4	55.7	59.3	73.1	71.3	84.5	32.3	53.3	56.7	86.0	66.8
NECUIUC_CLS-DTCT	88.0	68.6	67.9	72.9	44.2	79.5	72.5	70.8	59.5	53.6	57.5	59.0	72.6	72.3	85.3	36.6	56.9	57.9	85.9	68.0
NECUIUC_LL-CDCV	87.1	67.4	65.8	72.3	40.9	78.3	69.7	69.7	58.5	50.1	55.1	56.3	71.8	70.8	84.1	31.4	51.5	55.1	84.7	65.2
NECUIUC_LN-CDCV	87.7	67.8	68.1	71.1	39.1	78.5	70.6	70.7	57.4	51.7	53.3	59.2	71.6	70.6	84.0	30.9	51.7	55.9	85.9	66.7
UVASURREY_BASELINE	84.1	59.2	62.7	65.4	35.7	70.6	59.8	61.3	56.7	45.3	52.4	50.6	66.1	66.6	83.7	34.8	47.2	47.7	80.8	65.9
UVASURREY_MKFDA+BOW	84.7	63.9	66.1	67.3	37.9	74.1	63.2	64.0	57.1	46.2	54.7	53.5	68.1	70.6	85.2	38.5	47.2	49.3	83.2	68.1
UVASURREY_TUNECOLORKERNELSEL	85.0	62.8	65.1	66.5	37.6	73.5	62.1	62.0	57.4	45.1	54.5	52.5	67.7	69.8	84.8	39.1	46.8	49.9	82.9	68.1
UVASURREY_TUNECOLORSPECKDA	84.6	62.4	65.6	67.2	39.4	74.0	63.4	62.8	56.7	43.8	54.7	52.7	67.3	70.6	85.0	38.8	46.9	50.0	82.2	66.2

- Only methods in 1st, 2nd or 3rd place by group shown
- Groups: CVC, FIRST/Nikon, NEC/UIUC, UVA/Surrey

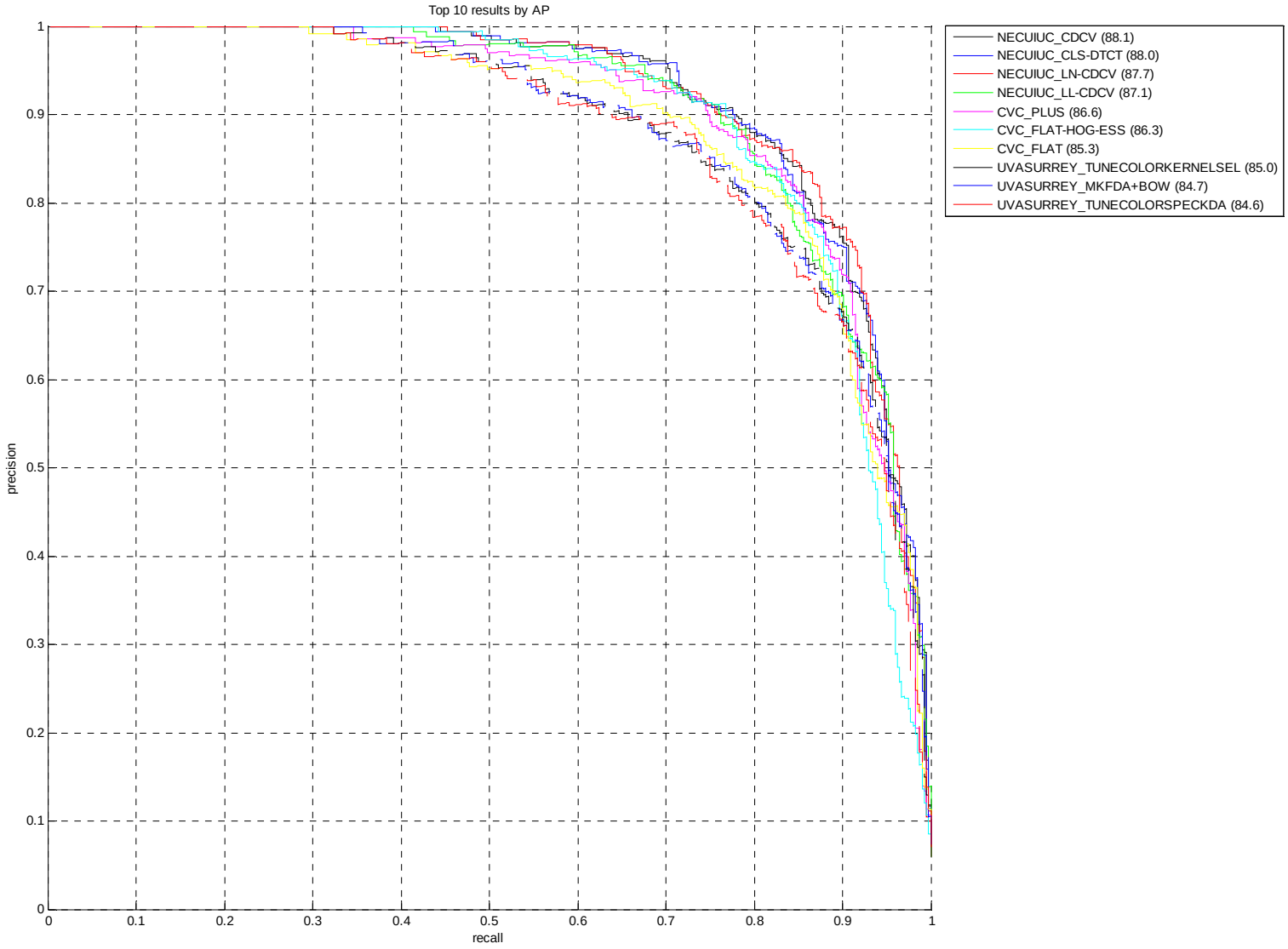
Median AP: Best Result by Group



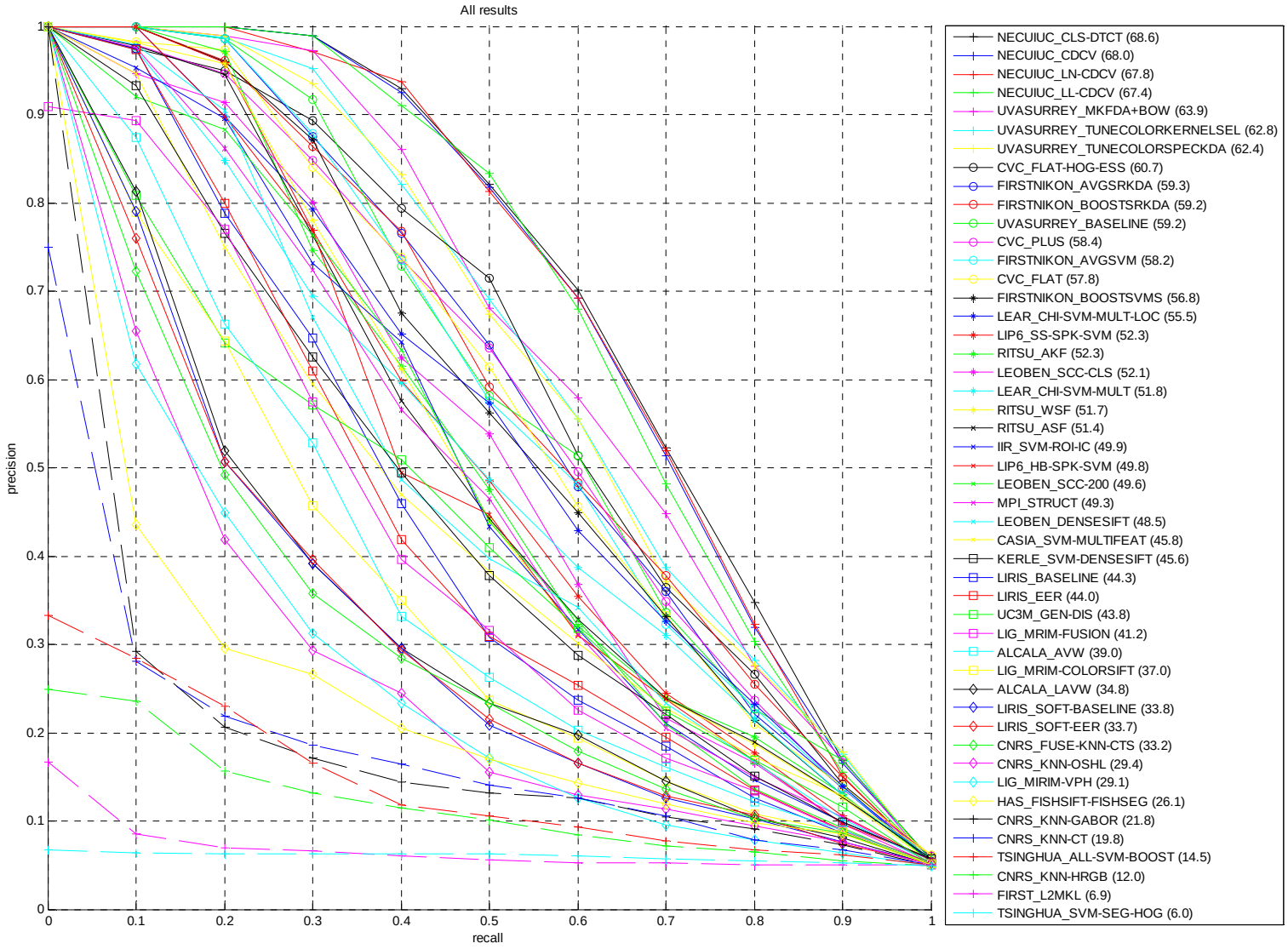
Precision/Recall: Aeroplane (All)



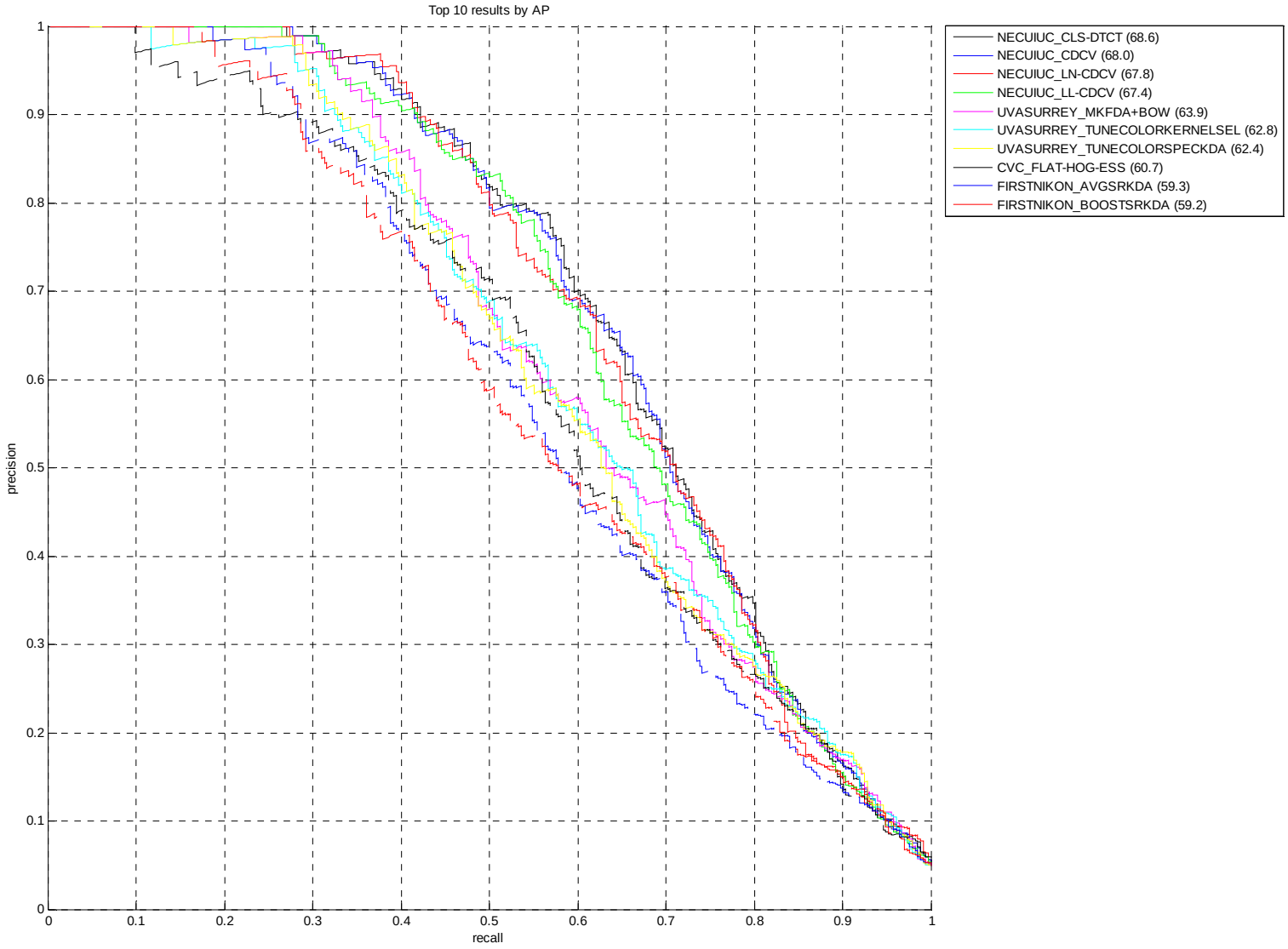
Precision/Recall: Aeroplane (Top 10 by AP)



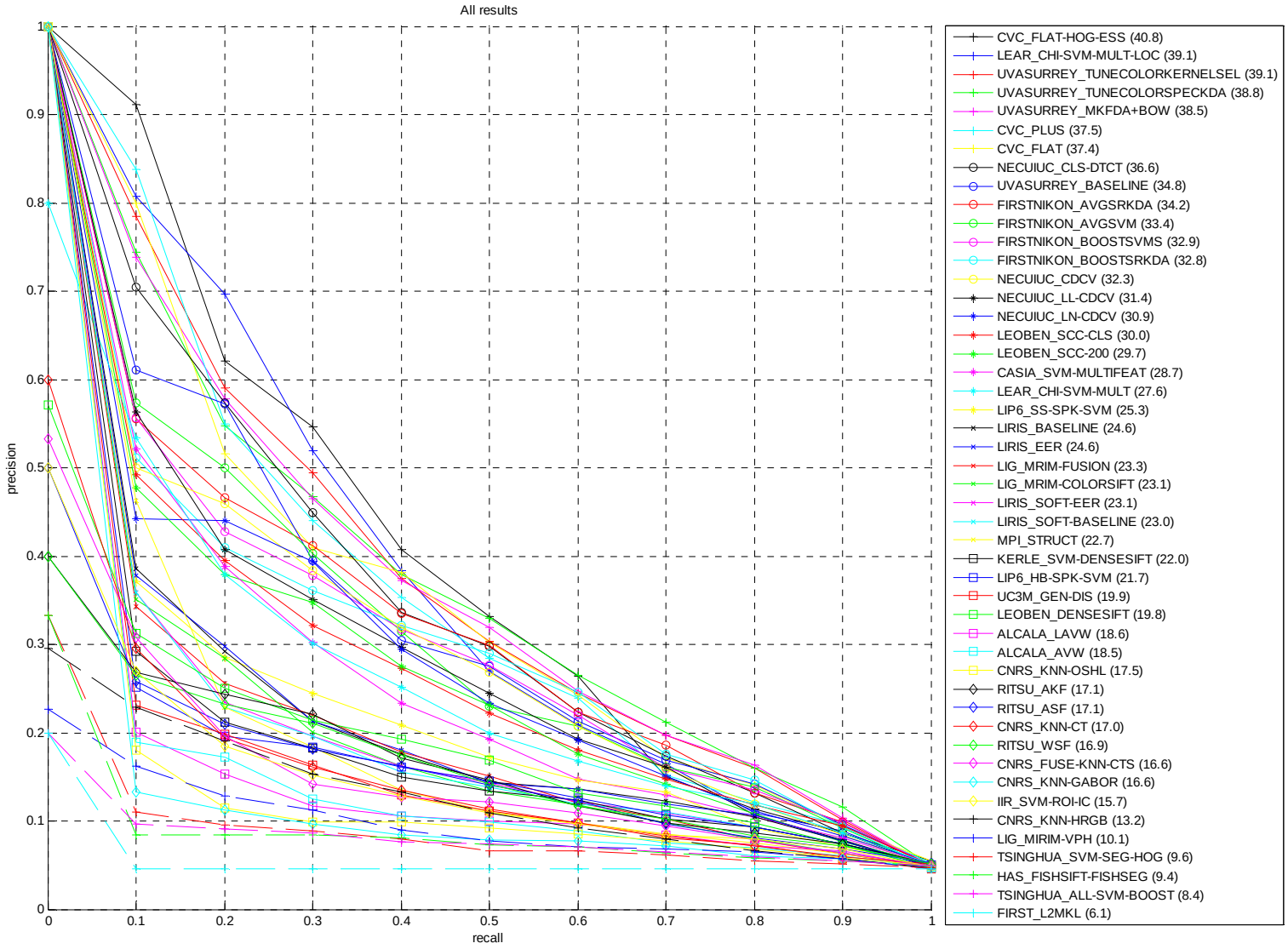
Precision/Recall: Bicycle (All)



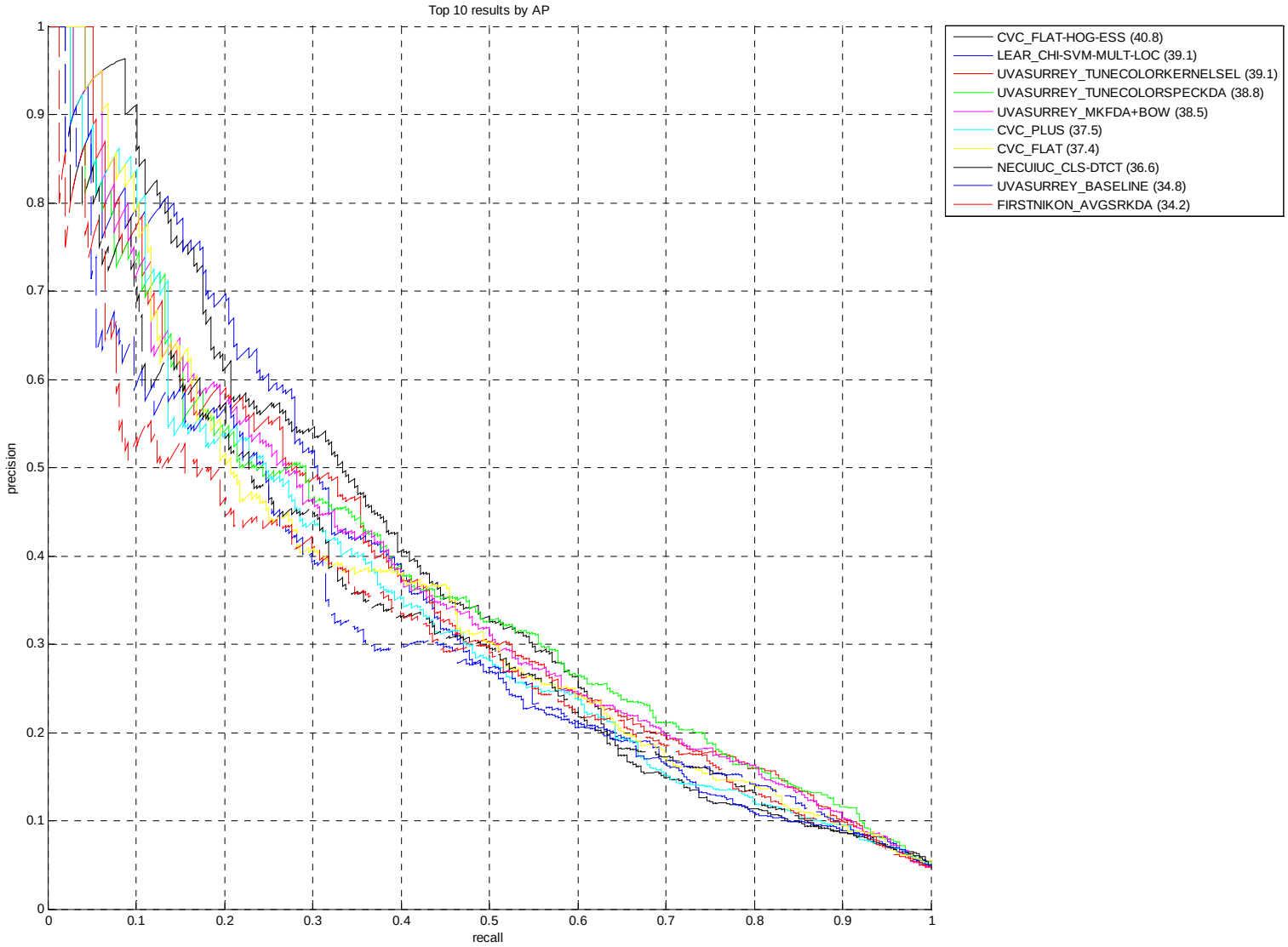
Precision/Recall: Bicycle (Top 10 by AP)



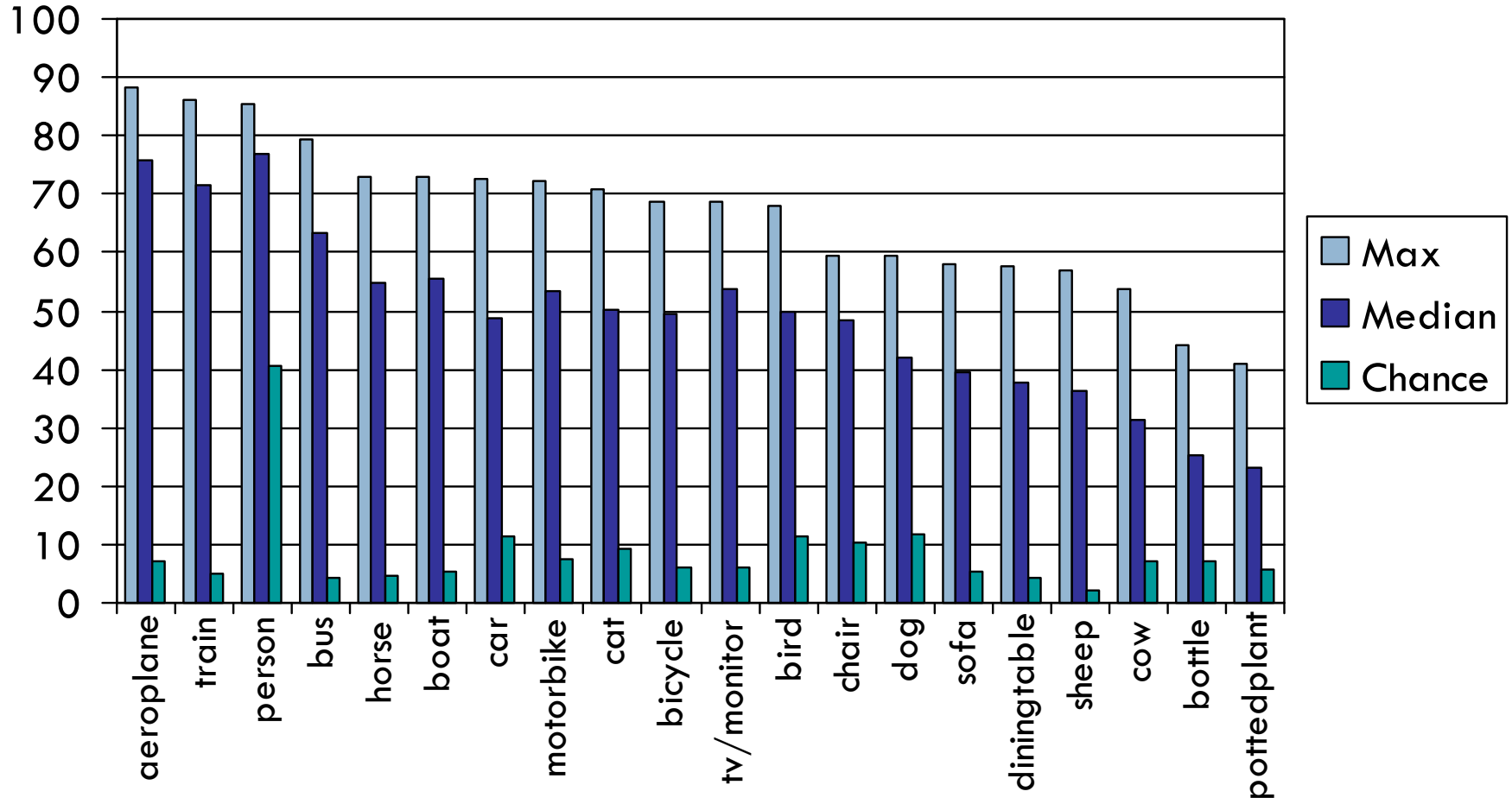
Precision/Recall: Potted plant (All)



Precision/Recall: Potted plant (Top 10 by AP)



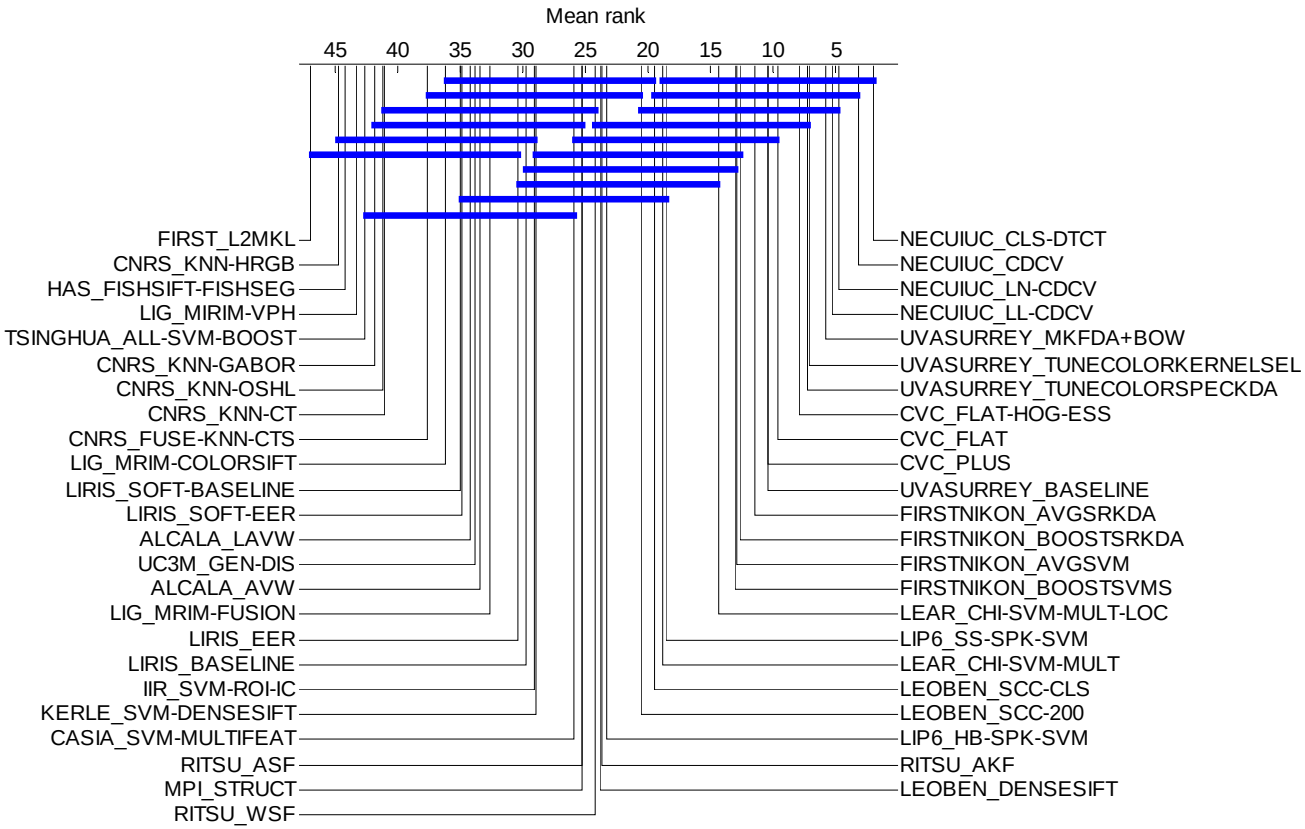
AP by Class



- Max AP: 88.1% (aeroplane) ... 40.8% (potted plant)

Statistical Significance

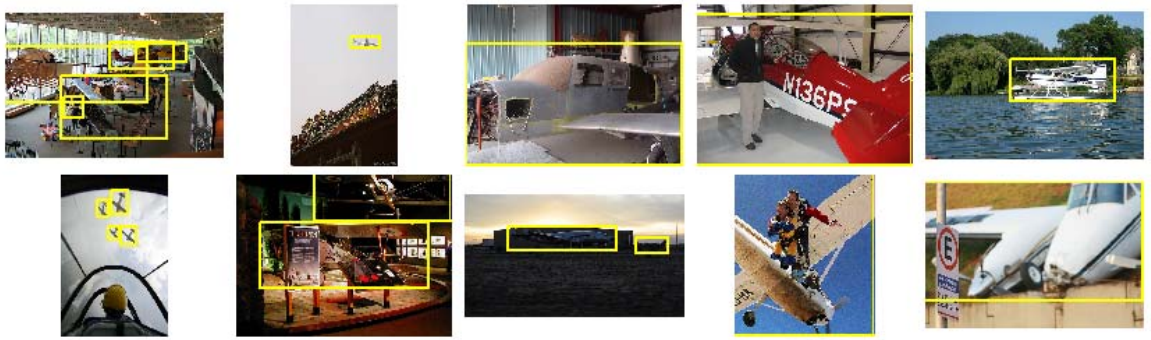
- Friedman/Nemenyi analysis
  - Compare differences in **mean rank** of methods over classes using non-parametric version of ANOVA
  - Mean rank must differ by at least 5.6 to be considered significant (p=0.05)



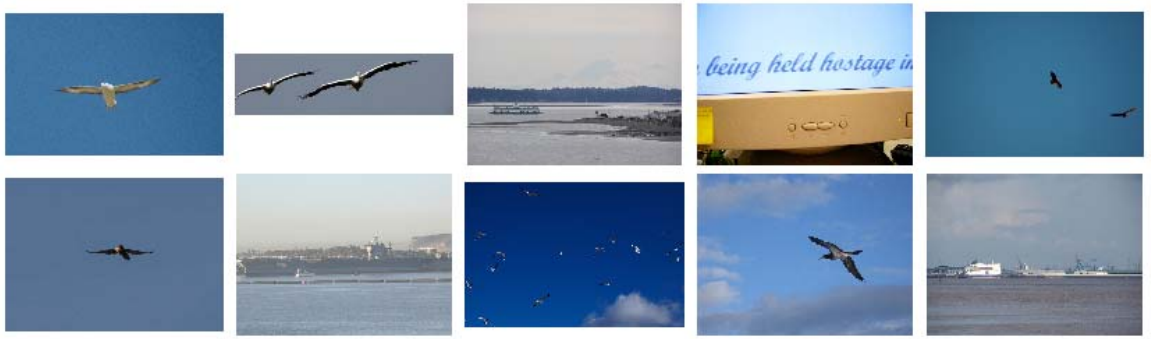
Ranked Images: Aeroplane

- Class images: Highest ranked

- Class images: Lowest ranked



- Non-class images: Highest ranked



- Context?

Ranked Images: Bicycle

- Class images: Highest ranked

- Class images: Lowest ranked

- Non-class images: Highest ranked

- "Texture"?

Ranked Images: Cat

- Class images: Highest ranked

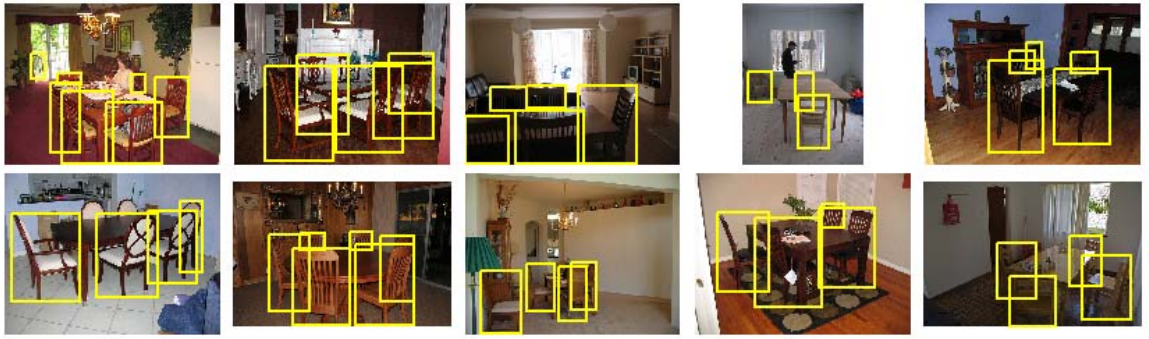
- Class images: Lowest ranked

- Non-class images: Highest ranked

- "Composition"?

Ranked Images: Chair

- Class images: Highest ranked



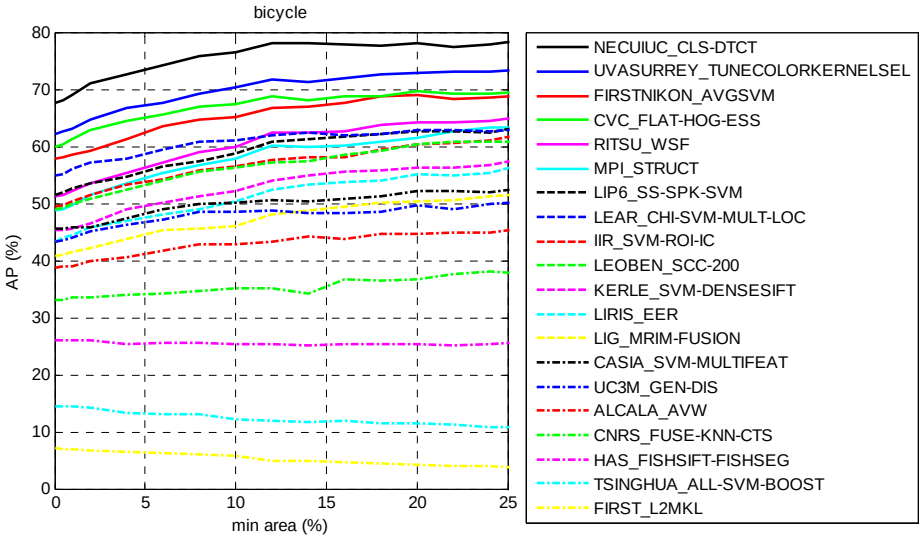
- Class images: Lowest ranked

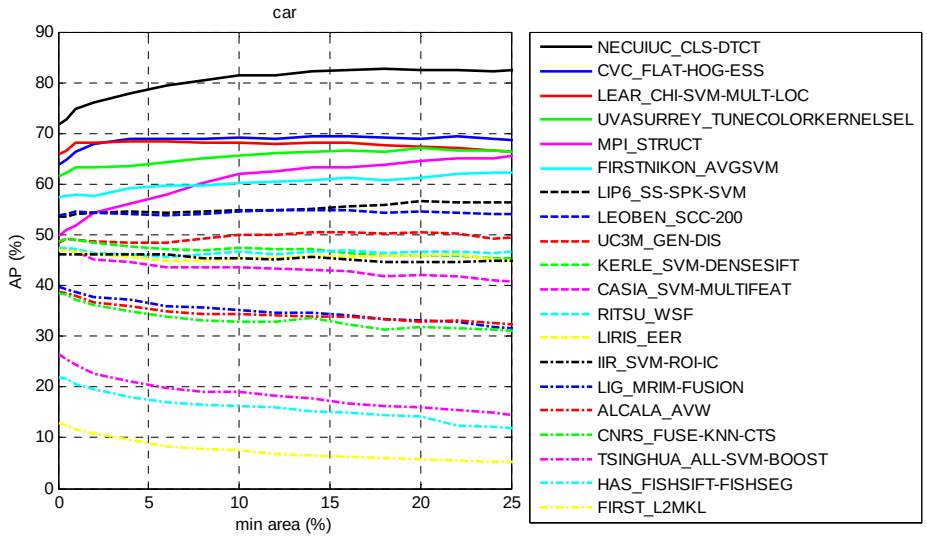
- Non-class images: Highest ranked

- Scene context?

AP vs. Object Class Area

- Do these methods have a bias toward larger objects?

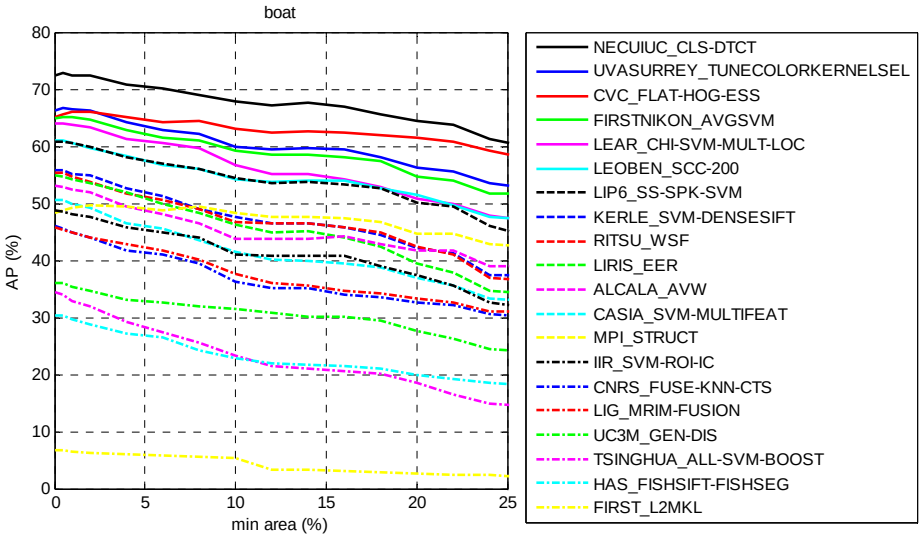


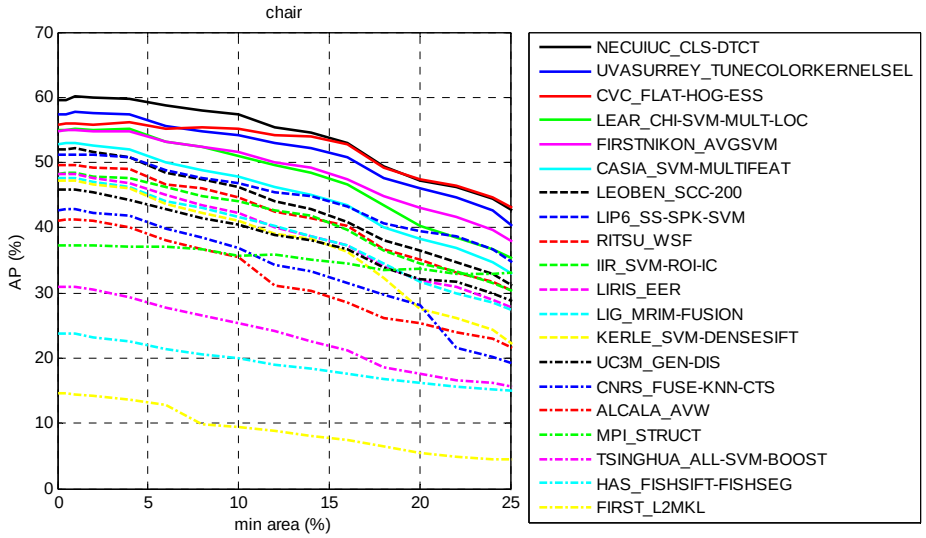


- Moderate evidence for some classes e.g. bicycle, car
- Accuracy tends to peak by 5% of image area

AP vs. Object Class Area

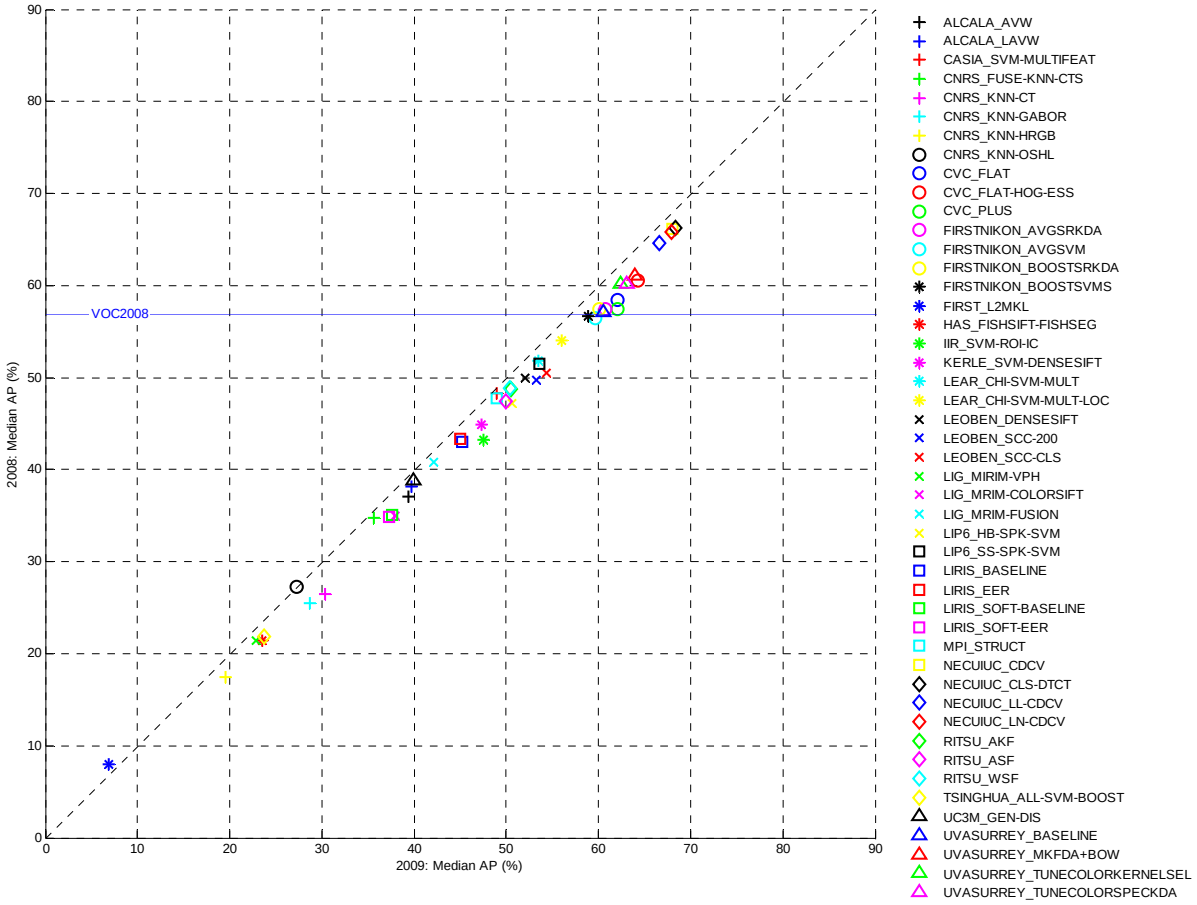
- For most classes, correlation with object class area is zero or negative





- Methods are learning more about context/scene appearance than object appearance?
- Possibility of occlusion negatively effects accuracy?

VOC2008 vs. VOC2009 Test Data



- High correlation, slightly better results on 2009 – "over-fitting"?
- Best methods are better than best 2008 result – better methods and/or advantage of more training data

Prizes

- Winner:
  - **NEC/UIUC**
    Yihong Gong, Fengjun Lv, Jingjun Wang, Chen Wu, Wei Xu, Jianchao Yang, Kai Yu, Xi Zhou, Thomas Huang
    NEC Laboratories America; University of Illinois at Urbana-Champaign
- Honourable mentions:
  - **UVA/SURREY**
    Koen van de Sande, Fei Yan, Atif Tahir, Jasper Uijlings, Mark Barnard, Hongping Cai, Piotr Koniusz, Theo Gevers, Arnold Smeulders, Krystian Mikolajczyk, Josef Kittler
    University of Amsterdam; University of Surrey
  - **CVC**
    Fahad Shahbaz Khan, Joost van de Weijer, Andrew Bagdanov, Noha Elfiky, David Rojas, Marco Pedersoli, Xavier Boix, Pep Gonfaus, Hany salahEldeen, Robert Benavente, Jordi Gonzalez, Maria Vanrell
    Computer Vision Centre Barcelona