

University of Amsterdam and Euvision Technologies at ILSVRC2013

Koen van de Sande

Daniel Fontijne

Harro Stokman

Cees Snoek

Arnold Smeulders



UNIVERSITY OF AMSTERDAM



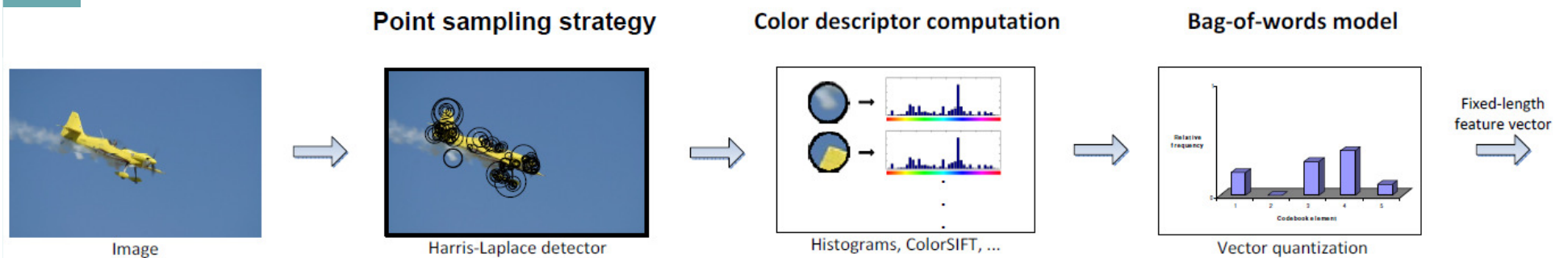
ILSVRC Workshop 2013 - December 7th 2013

About EU VISION technologies

- Spin-off from University of Amsterdam in 2010
- Brings University's concept detection software to the market
- We are hiring

<http://www.euvt.eu/>

Lessons from Pascal VOC, ILSVRC & TRECVID Classification



What works? [Zhang IJCV 2007, Song CVPR 2011]

- Ultra-dense sampling [Jurie ICCV 2005]
- Color descriptors [van de Sande TPAMI 2010]
- Fisher vectors [Sanchez IJCV 2013]

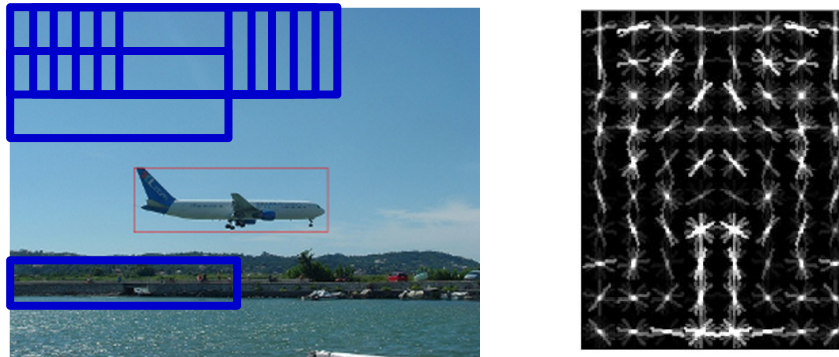
Bag-of-words proven effective for classification

Convolutional networks [Krizhevsky NIPS 2012] even better
(given enough data)

Software available for download at <http://www.colordescriptors.com/>

Lessons from Pascal VOC Detection

- Exhaustive search is great
 - Part-based [Felzenszwalb TPAMI 2010]
 - Improved by many [Zhang CVPR 2011] [Zhu TPAMI 2012]
 - Cheap features mandatory
 - Fast with accuracy loss [Dean CVPR 2013]



- Constrained search facilitates expensive features
 - Efficient subwindow search [Lampert TPAMI 2010]
 - Jumping Windows [Vedaldi TPAMI 2009]
 - Fine Spatial Pyramids [Russakovsky ECCV 2012]

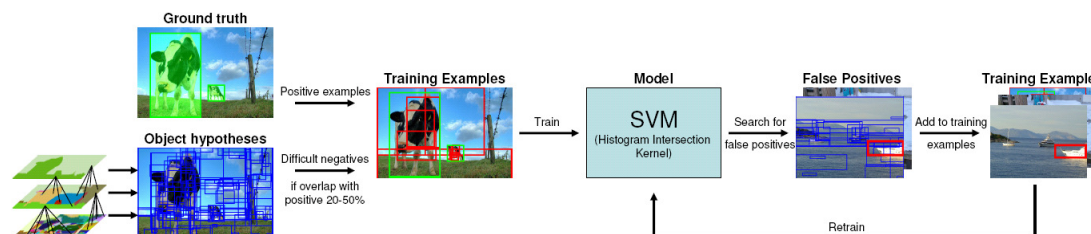
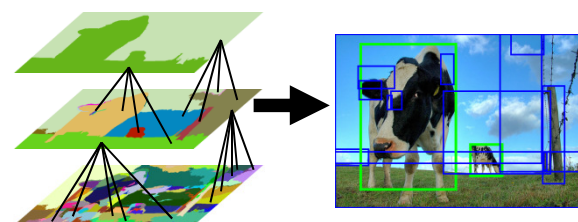
Our approach

- Classification priors

- Selective search

- Features

- Retraining





Features

- Use SIFT descriptors
- Novelty: New encoding method
- Faster & more accurate than bag-of-words
- **Submitted**

Selective Search

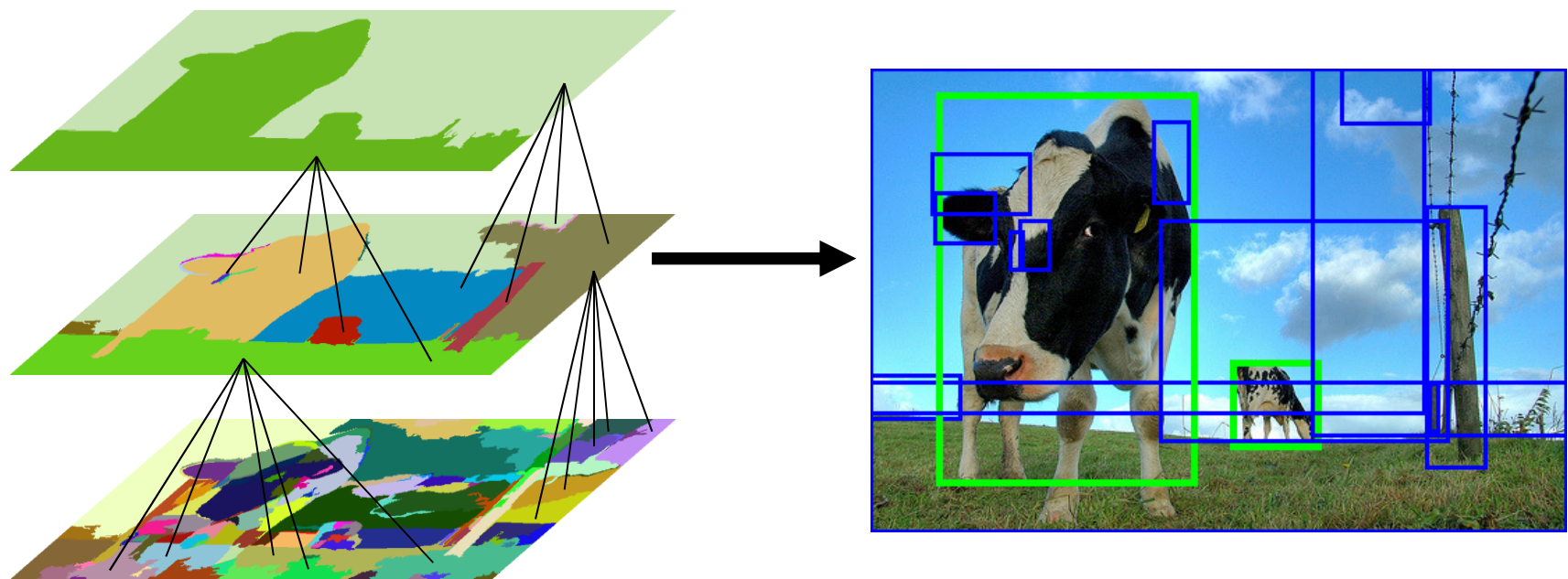
- Once discarded, an object will never be found again



- Image is intrinsically hierarchical
- Segmentation at a single scale won't find all objects

Selective Search: Approach

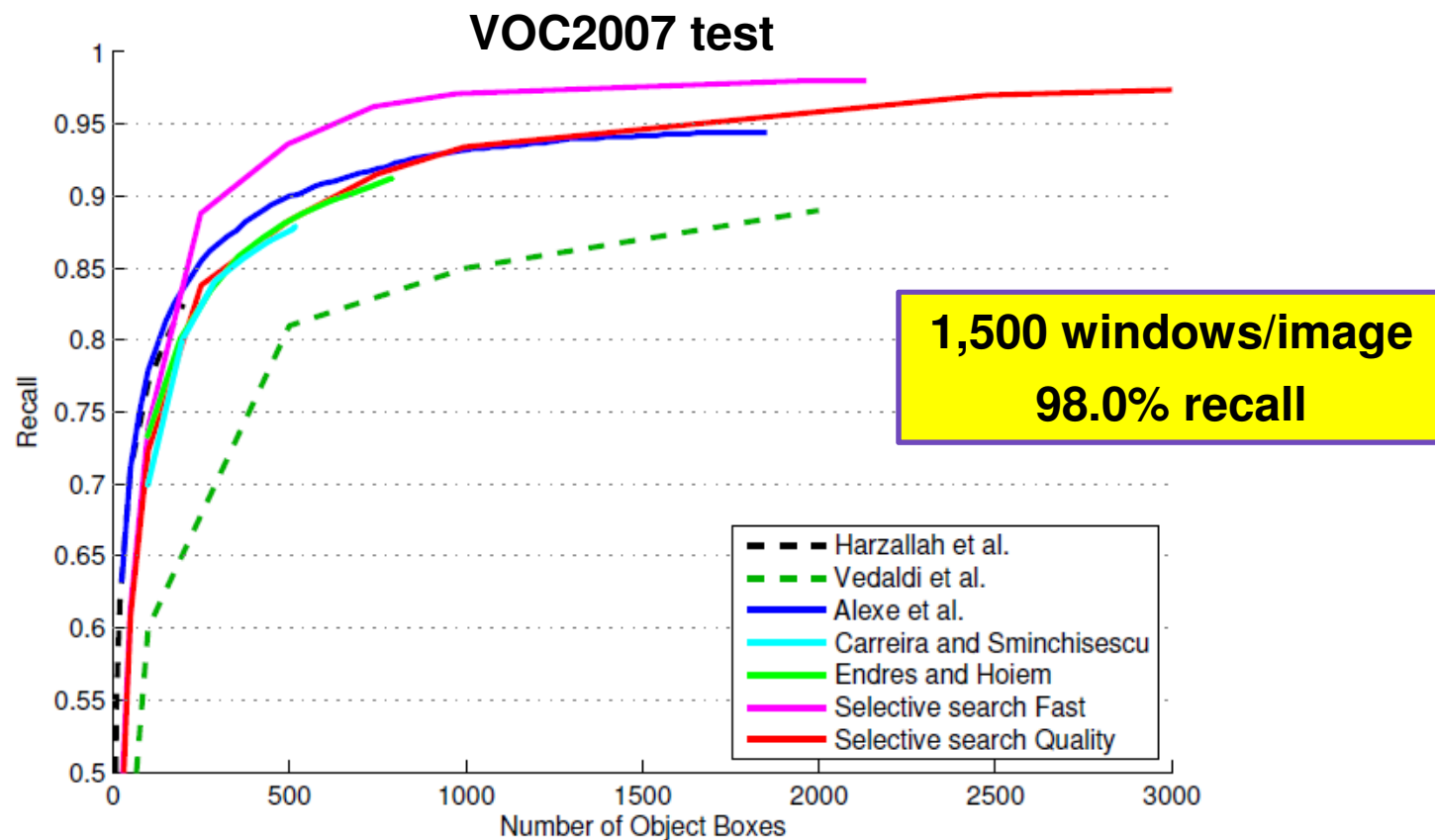
- Hypotheses based on hierarchical grouping



Group adjacent regions on color/texture cues

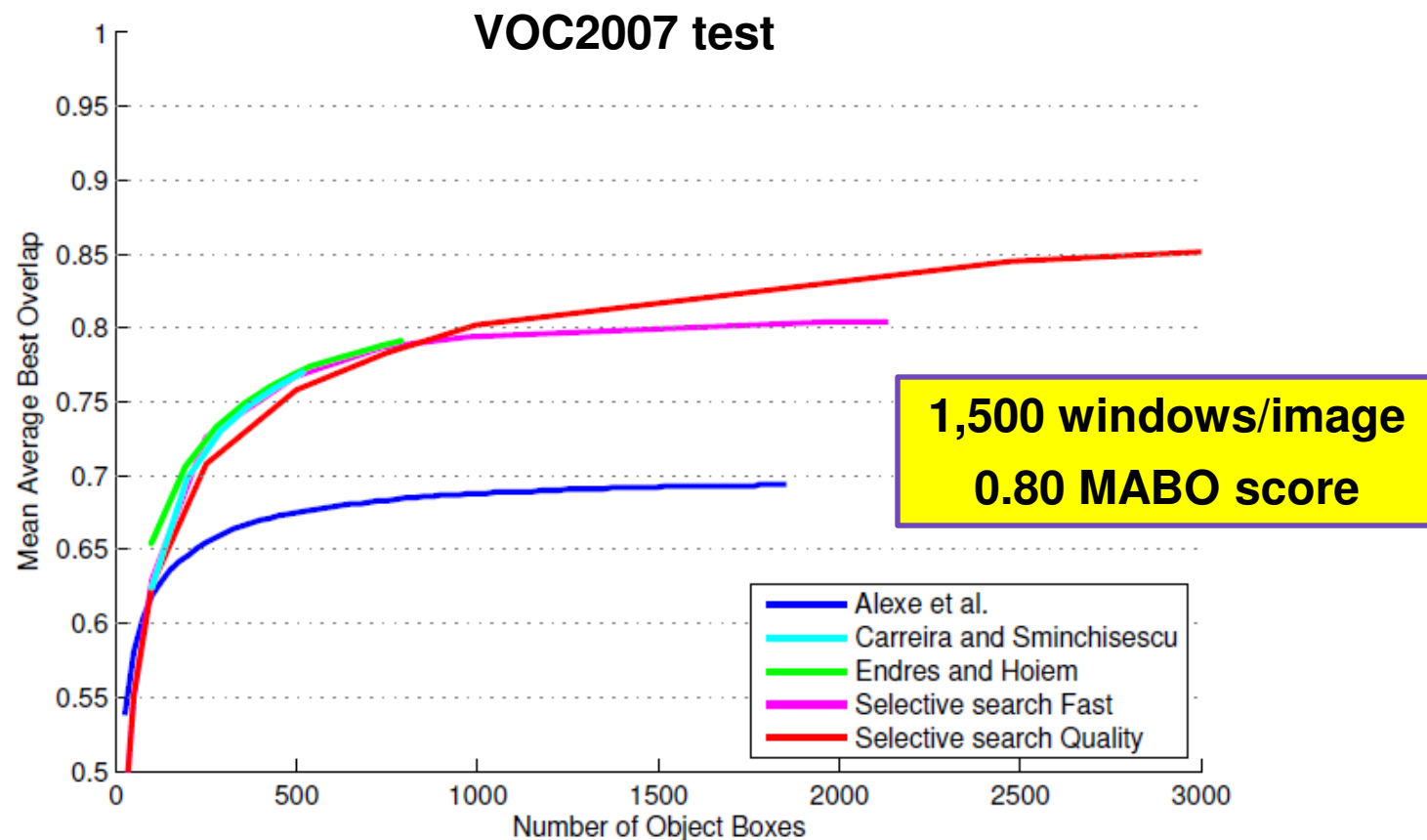
Selective Search

- Multiple complementary invariant color spaces
- Location hypotheses are class-independent



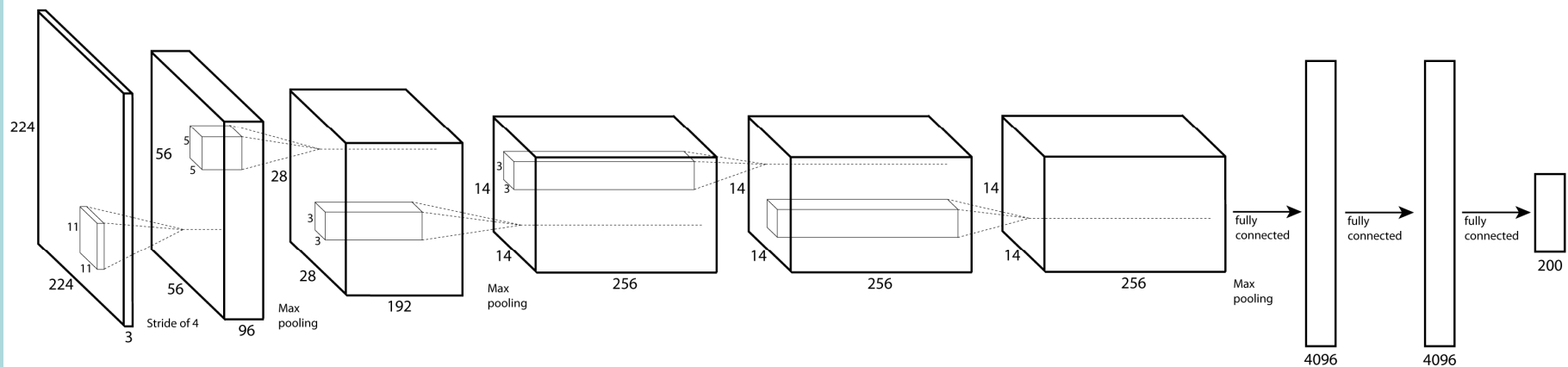
Selective Search

- Multiple complementary invariant color spaces
- Location hypotheses are class-independent



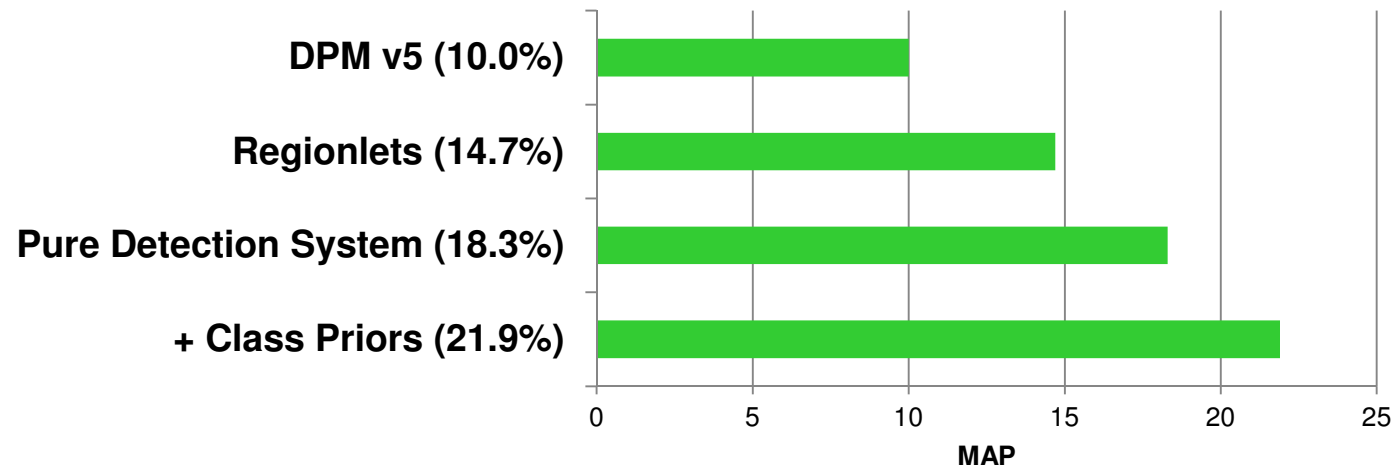
Classification priors

- Found in TRECVID localisation task:
 - CNN prior boosts even more than BoW prior
- Therefore trained multiple nets on DET 200 on GPUs
- High error rate found, due to limited dataset
- Scores used to rank images



DET quantitative results

ILSVRC 2013 DET Validation Set

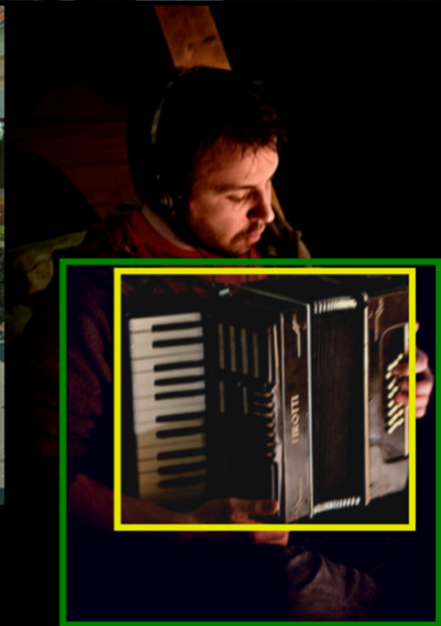
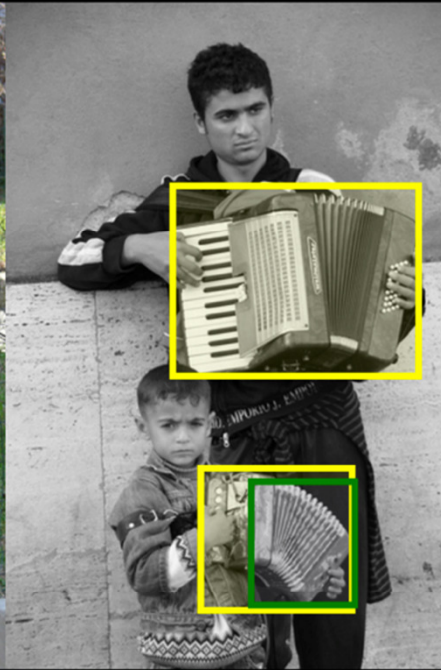
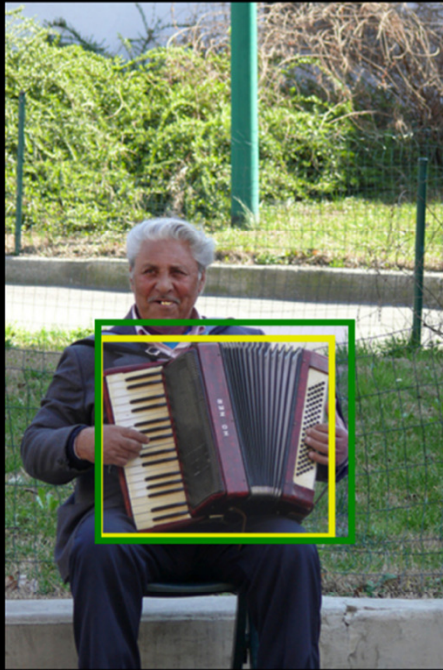


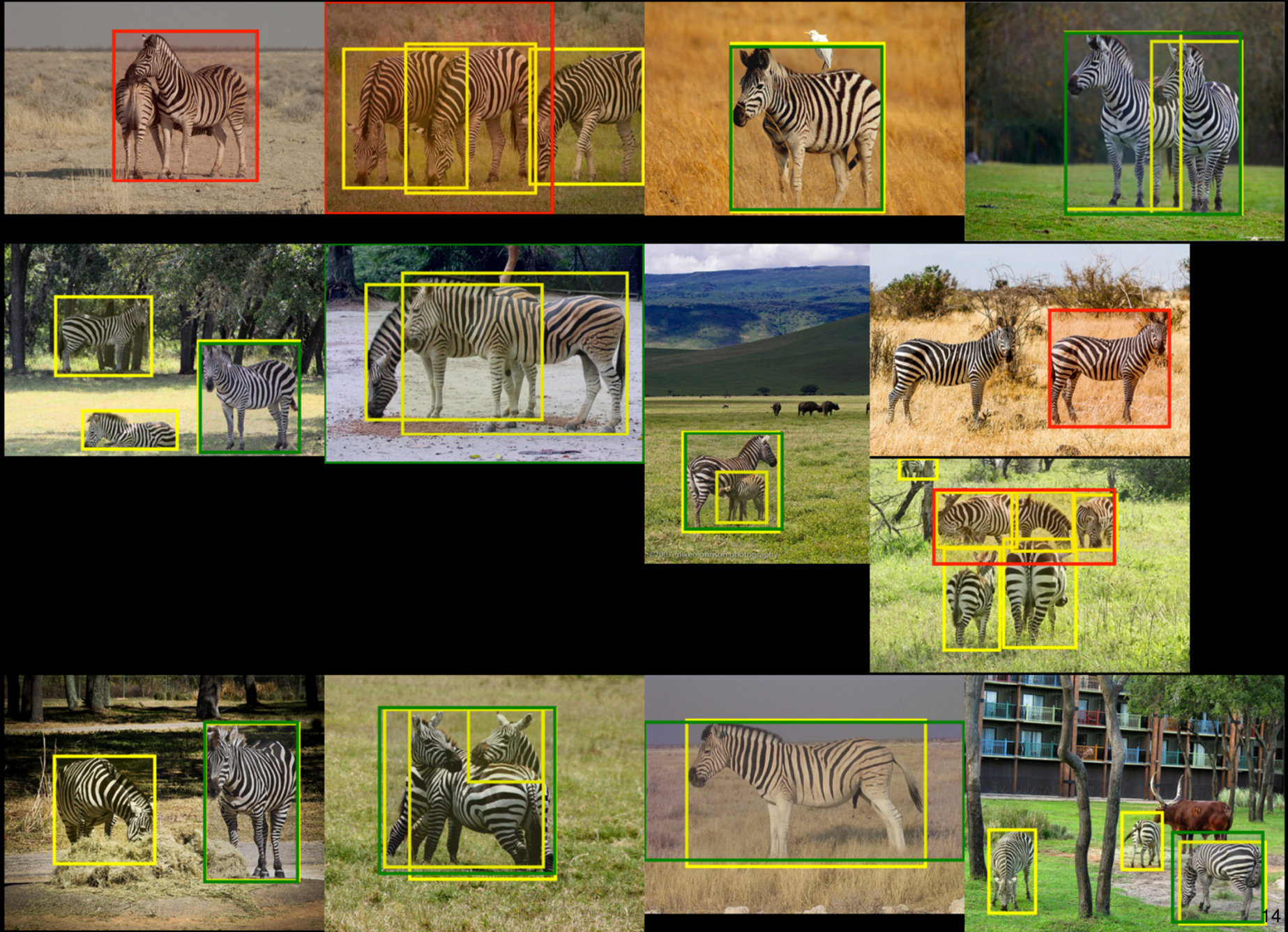
Test set:

- Pure Detection System: 19.2%
- Added Classification Priors: 22.6%

accordion (n02672831, MAP on val=40.9)

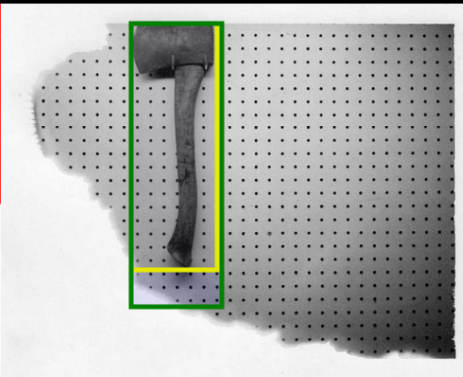
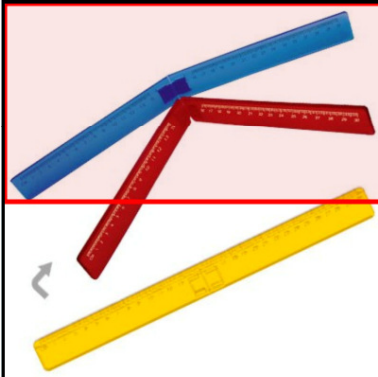
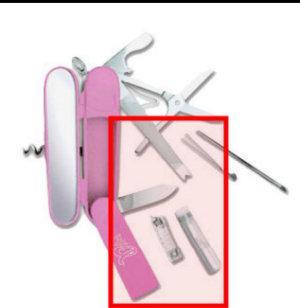
1 2 3 4 5 ... 17 18 Next »





axe (n02764044, MAP on val=1.1)

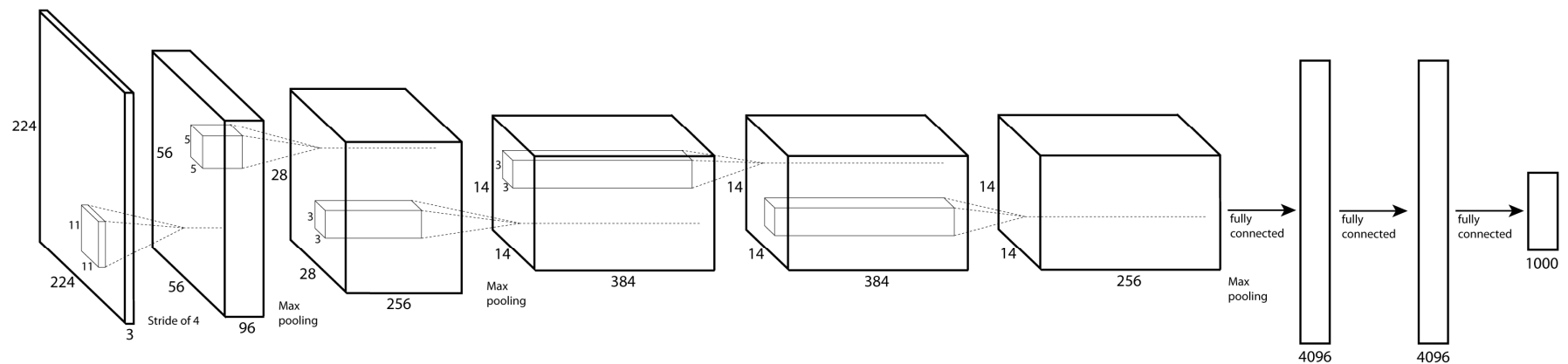
1 2 3 4 5 ... 104 105 Next





ImageNet 1000 classification task

- Trained multiple CNNs, achieved 14.3%
- Novelties:
 - Trained for 200+ epochs. Found that training for long times at high learning rate really improves
 - Employed larger convolutional layers
 - Used scaling as data augmentation



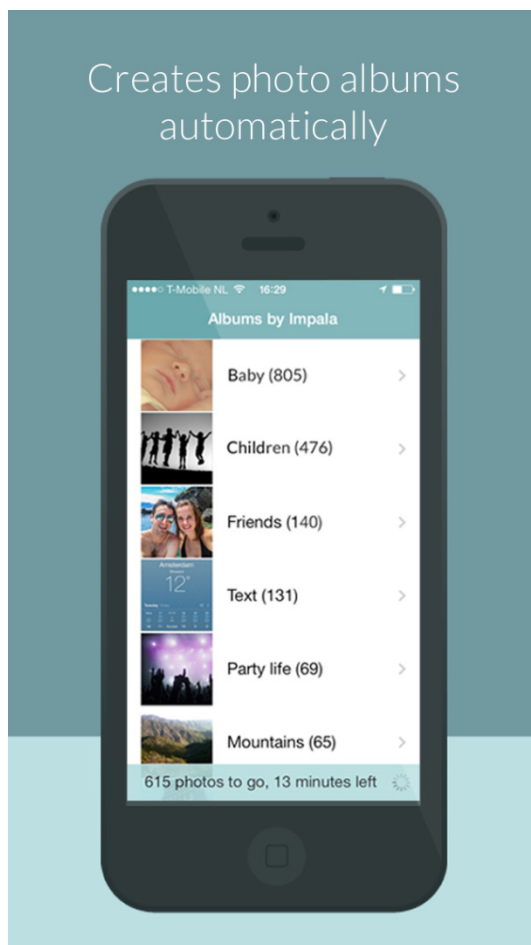
ImageNet 1000 on iPhone

- Our second run (16.6% top 5 error rate) was performed on our 'iPhone cluster'
- Evusion classification engine optimized for mobile
- 3 seconds per 8 images on iPhone 5s
- Available for free in App Store

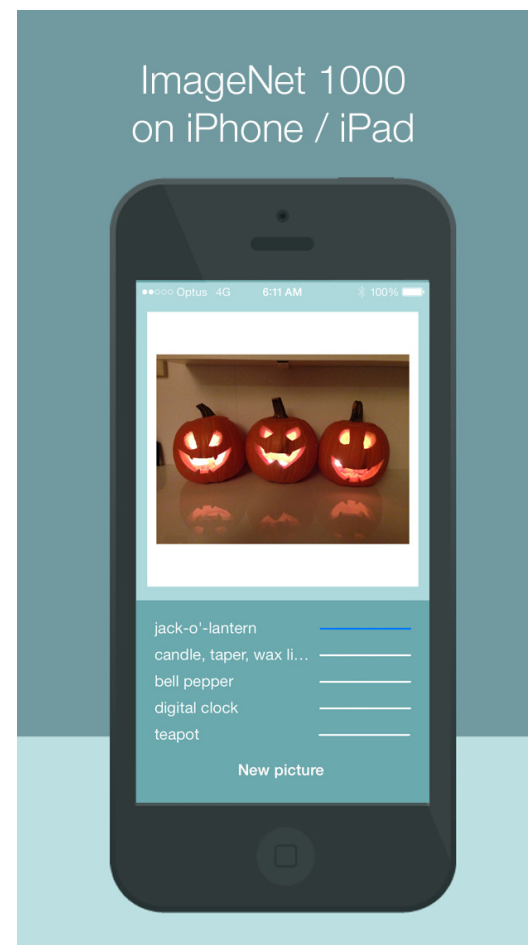
Demo . . .



Try it on your own photos



[Euvision Impala](#)



[UvA-Euvision ImageNet](#)



Conclusions

- New features (submitted)
- Selective search for few high quality object hypothesis
- Classification priors help
- ImageNet-scale classification on mobile