ICSI INTERNATIONAL COMPUTER SCIENCE INSTITUTE

Interactive Adaptation of Real-Time Object Detectors

Daniel Goehring, Judy Hoffman, Erik Rodner,
Kate Saenko, Trevor Darrell

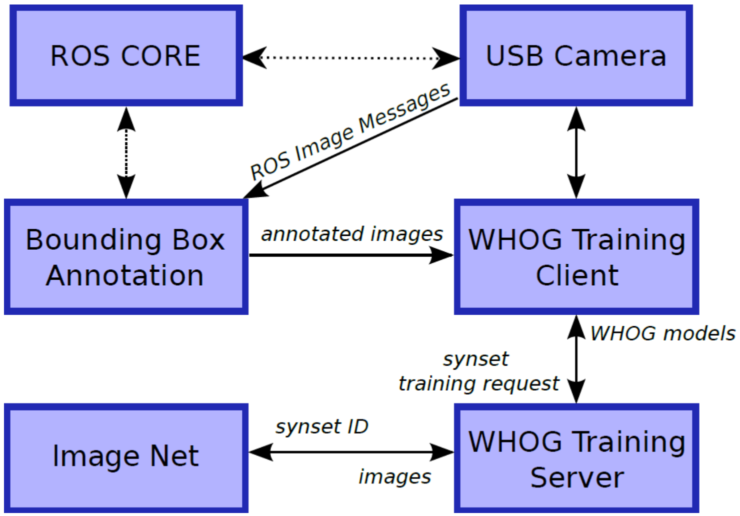
Motivation

- Fast object detection using 2d-images: What is possible with state-of-the-art vision techniques and databases?
- Training of HOG-feature based classifiers can be time-consuming
- Models trained on large databases often perform poorly in real situations, efficient domain adaptation is required

INTERNATIONAL COMPUTER SCIENCE INSTITUTE

Contribution

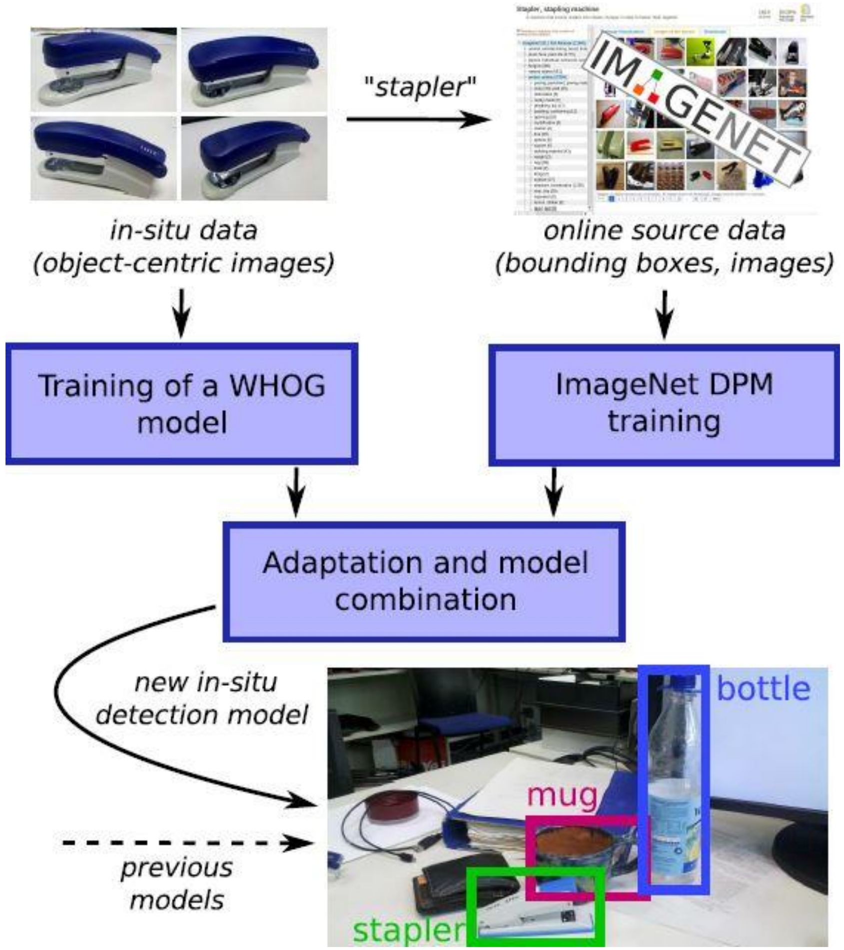
- Main components:
  - take advantage of large scale internet database, thousands of classes: ImageNet
  - fast training with whitened HOG
  - use in-situ images
  - fast model adaptation
  - realtime detection with 2d-FFT
  - ROS framework to execute on the PR2 robot platform



Daniel Goehring
02. June 2014

International Computer Science Institute, Berkeley, CA, U.S.A.

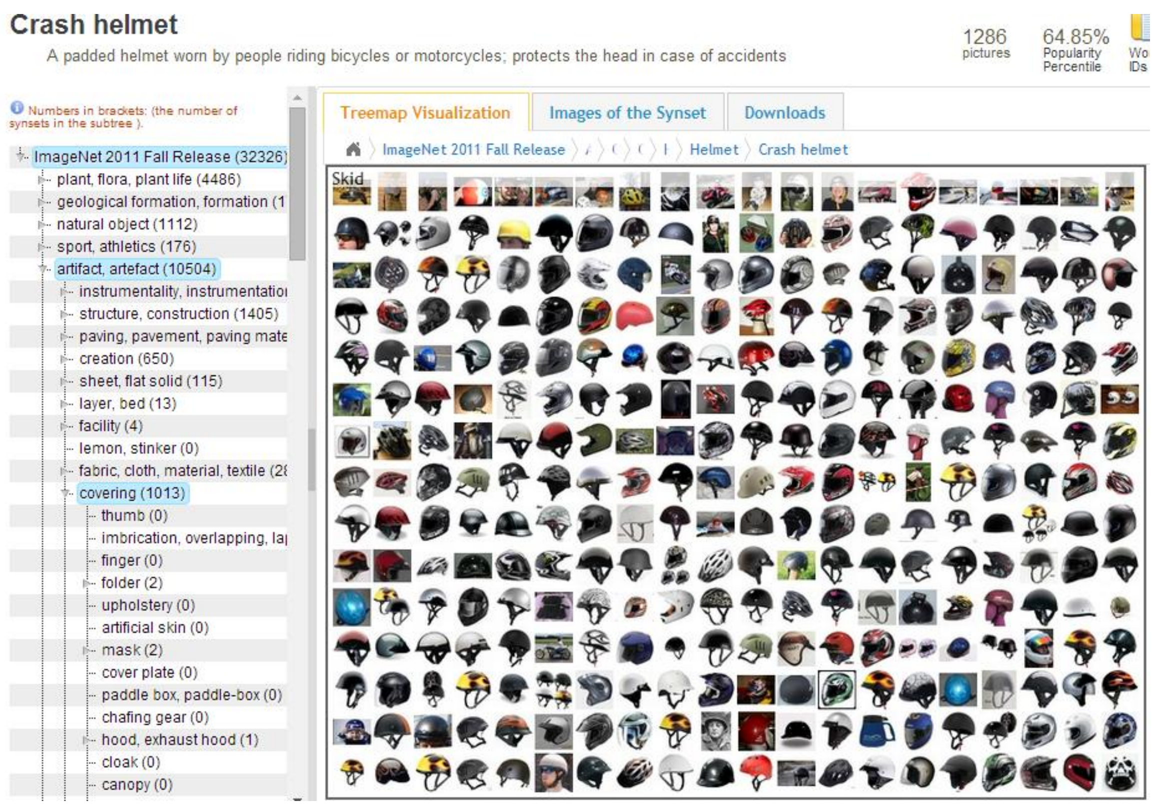
System Overview



INTERNATIONAL COMPUTER SCIENCE INSTITUTE

ImageNet

- 14 million images
- 21k semantic concepts
- many with annotated bounding boxes
- synset names and descriptions



INTERNATIONAL COMPUTER SCIENCE INSTITUTE

HOG Features

- Linear sliding window detectors, using HOG features are robust to illumination changes and small shifts

- detection score calculated with feature representation of an image *φ(I)*, with filter vector *w* on position *x* (and a certain scale) :

$$\text{argmax}_{\mathbf{x}} \, f_{\mathbf{w}}(\mathbf{x}) = \text{argmax}_{\mathbf{x}} \left[\mathbf{w} * \phi(I) \right](\mathbf{x})$$

Example: Bottle

*New **in-situ** training image and one-shot detection model without adaptation*

IMAGENET *detection model for synset bottle*

INTERNATIONAL COMPUTER SCIENCE INSTITUTE

Fast Learning of Detection Models

- Object detection over the past few years converged on using linear SVM over HOG features
- Linear SVM training of positive and negative examples is expensive, in particular for training of thousands of categories, suppression of false positives
- Whitened HOG features (WHO) based on Linear Discriminant Analysis: significant decrease of training time *(Hariharan, Malik, et. al. ECCV ‘12)*

INTERNATIONAL COMPUTER SCIENCE INSTITUTE

Whitened HOG Features

- Assumption: positive and negative training example are Gaussian distributed, which leads to an optimal hyperplane separating positive and negative sets:

$$\mathbf{w} = \mathbf{S}_0^{-1}(\mu_0 - \mu_1)$$

- covariance matrix S_0 can be estimated from unlabeled data and reused for all categories to whiten and implicitely decorrelate HOG features
- linear descriptor computes the difference of average positive and negative features in a whitened space *(Hariharan, Malik, et. al. ECCV ‘12)*

Interactive Learning Interface

- user inputs search term
- matching of terms with ImageNet synsets
- visual feedback for in-situ training
- bounding box visualization
- 5 images

INTERNATIONAL COMPUTER SCIENCE INSTITUTE

Adaptation of Model Mixtures

- Incorporate the models learned on in-situ images with max-fusion
- add in-situ model as additional component in the detection mixture model

$$f_{\mathcal{M}}(\mathbf{x}) = \max_{M \in \mathcal{M}_I \cup \mathcal{M}_O} f_M(\mathbf{x})$$

Fast inference with Fourier Transformation

- bottleneck during detection is the convolution of learned filters with HOG feature map of the image
- speedup by taking advantage of the convolution theorem
- detection speed: 2 Hz for 20 models on a 2.5 GHz machine with 320x240 pixel images

Click to play
PR2 demo

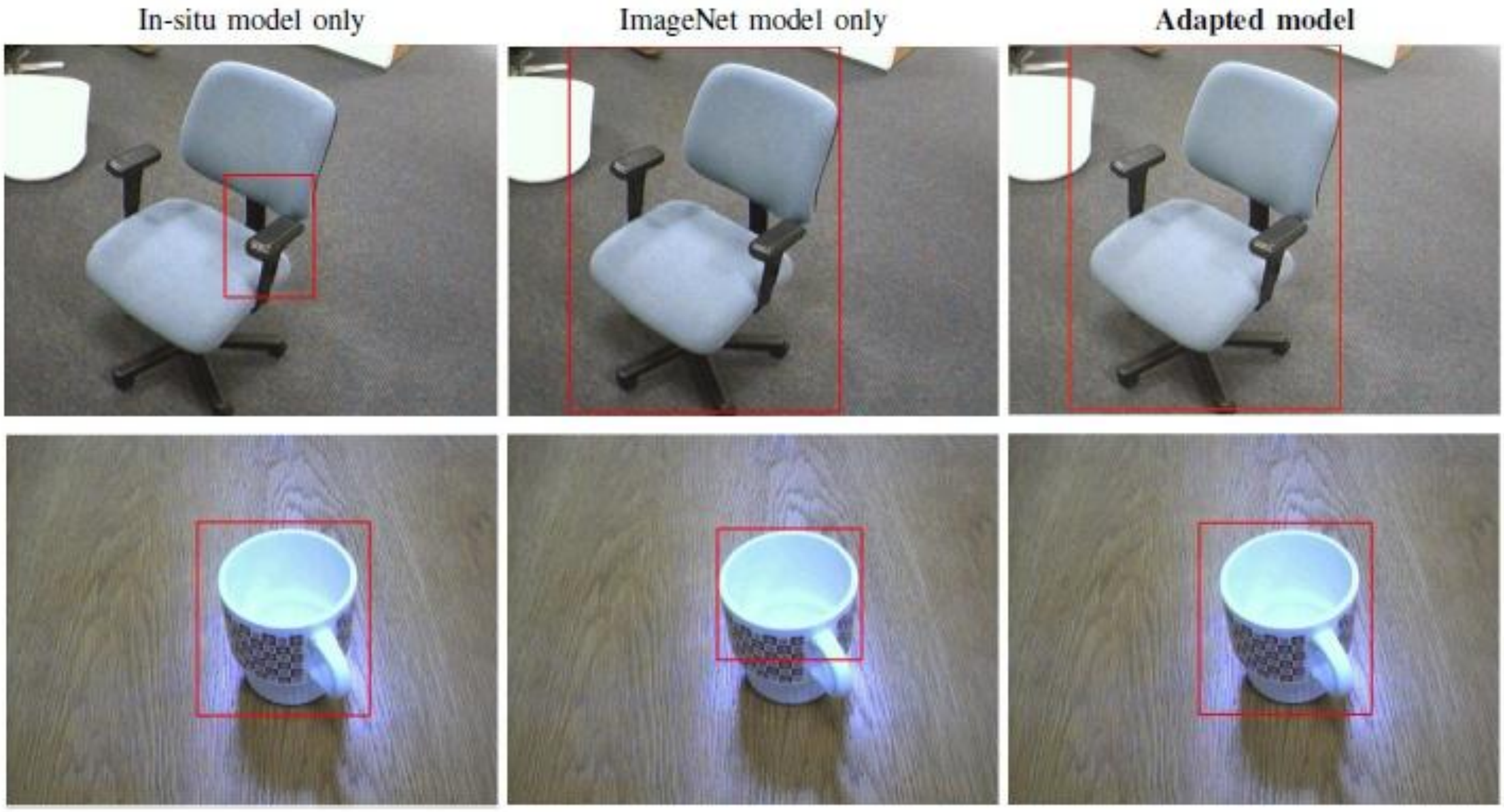
INTERNATIONAL COMPUTER SCIENCE INSTITUTE

Detection Examples

Daniel Goehring
02. June 2014

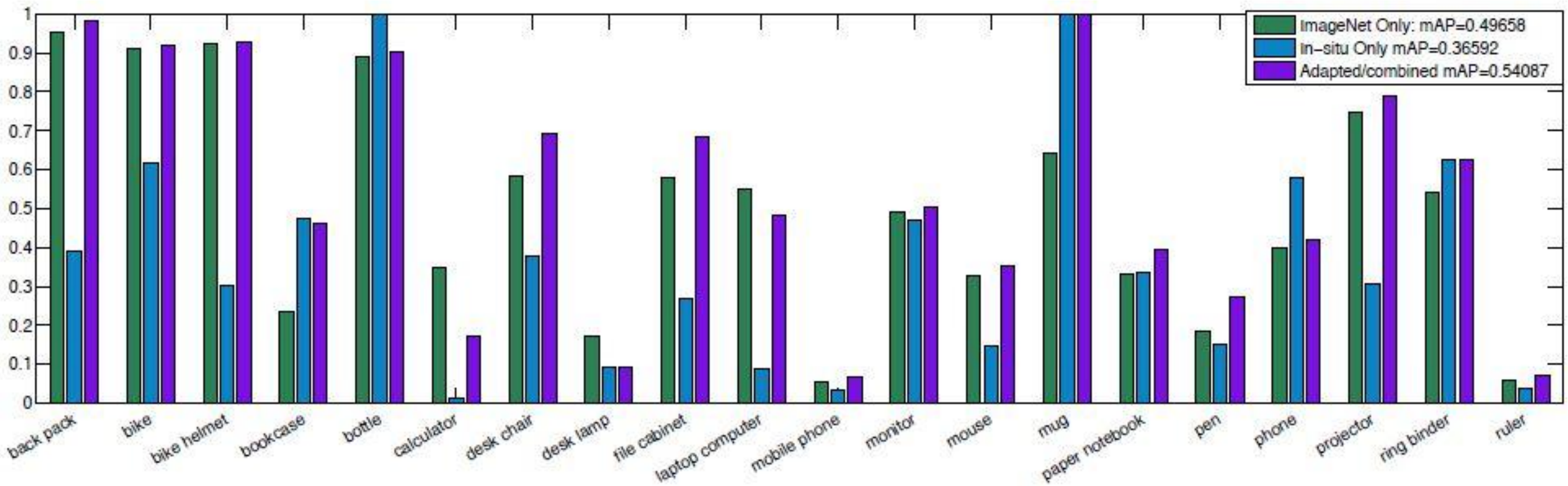
International Computer Science
Institute, Berkeley, CA, U.S.A.

Example: Office Data



INTERNATIONAL COMPUTER SCIENCE INSTITUTE

Experimental Results



- average precision (AP) for a category calculated as the integral of the precision-recall curve
- detection was correct when the detected bounding box overlapped at least 50% with the trained one

Conclusion

- we presented an approach to learning detection models on the fly
- combined training data from large-scale databases with few in-situ images
- adaptation of models learned from internet sources to the target environment led to better detection results
- simple adaptation scheme and fast training in less than 1 minute (including downloading bounding boxes)
- fast detection using 2d-FFT
- http://raptor.berkeleyvision.org

Future Work

- improve the detector by adding some rotational invariance to our models
- proposing object hypotheses to the user
- active learning techniques to guide the acquisition step during learning to examples with a significant impact on the classification model