

### Interactive Adaptation of Real-Time Object Detectors



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

Daniel Goehring 02. June 2014



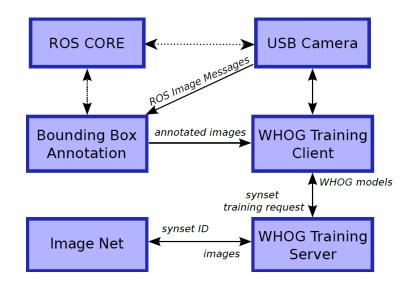
## Motivation

- 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 timeconsuming
- Models trained on large databases often perform poorly in real situations, efficient domain adaptation is required



## Contribution

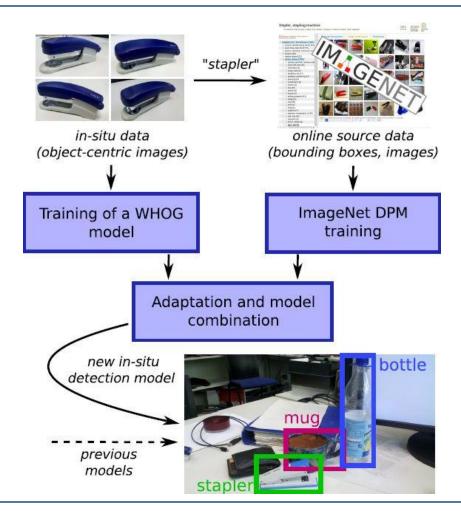
- Main components:
  - take advantage of large scale internet database, thousands of classes: ImageNet
  - fast training with whitened HOG
  - use in-situ images
  - fast model adaptation
  - realtime detection with 2d-FFT
  - ROS framework to execute on the PR2 robot platform







## **System Overview**



Daniel Goehring 02. June 2014



## ImageNet

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

A padded helmet worn by people ridir	g bicycles of motorcycles, pr	sects the field III case of a	accidento	pictures		We IDs
Numbers in brackets: (the number of nsets in the subtree ).	Treemap Visualization	Images of the Synset	Downloads			
ImageNet 2011 Fall Release (32326)	🖌 👌 ImageNet 2011 Fall Re	lease $\rangle$ / $\rangle$ ( $\rangle$ ( $\rangle$ ) $\rangle$ Helm	et $ angle$ Crash helmet			
- plant, flora, plant life (4486)	Skid					Ē
▶- geological formation, formation (1		💹 🥌 🎑 🧭	è 🔛 🔛 📲 🛲			
- natural object (1112)	🔊 🔊 🏐 🏶 🡝					
⊩- sport, athletics (176)						1
- artifact, artefact (10504)	- I 💭 🧠 🦛 👔	🗿 🚓 🙆 🖾 🍕	) 🗩 🕰 👁 🍘	len 🙈 🔛 .	🚗 🏔 🕋	Þ
👘 instrumentality, instrumentation				VOC	V VV	'
- structure, construction (1405)	💎 🎨 🧊 🧠 🗩	X 🔊 🗐 🦱 😤 🛍	😭 🥐 💏 🍗	D 🗇 🔫	SA 🐨 🗭	1
paving, pavement, paving mate						
▶- creation (650)				<b>100 100 100</b>	😥 🕅 🔧	9
⊩– sheet, flat solid (115)						-
k- layer, bed (13)		277200	T T T	A 19 19 19 19 19 19 19 19 19 19 19 19 19		P
⊨- facility (4)						ò
– lemon, stinker (0)	V 111 ** 111 **		PT TOT	ARA	***	1
🖡 fabric, cloth, material, textile (2)		A	1 🦛 🥎 🚄 19.0	🚮 🌊 🕋		4
* covering (1013)		P 4 🥙 P 🖪			4 00 00	2
- thumb (0)		ia: 🛖 💭 🗛 📷			Ð. 🗩 🗑	h
- imbrication, overlapping, lap					00	1
- finger (0)		SC 🖚 🗣 🖉 🖉	) <b>er 🧟 fr 4</b>	R 40 .	7 9 7	9
i⊧- folder (2)					0	r
- upholstery (0)	N 10 7 10 1		7 💎 🙉 💎 🗺	A 2 P	477 🔍 🧶	
- artificial skin (0)					a 🔿	
i⊷ mask (2)		1 T 🖉 🚺 🐬 📆	♥ ♥ ♥ ♥ ♥	28 🗶 TH	M 77 77	-
- cover plate (0)		***				
paddle box, paddle-box (0)	7 7 7 7 7	77 V V 2	7 🜌 🥌 🌌 🏹		$\Delta 4 \Delta$	1
chafing gear (0)		S A 😤 🧟 🗖			la 🖕 🙆	5
hood, exhaust hood (1)			i 🛃 🚍 🖡 💽 🦉	2 2 2		2
- cloak (0)	A A A A A A				CA . CAS	3
- canopy (0)					e 🔁 🔁	2



## **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) :

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



#### **Example: Bottle**



New **in-situ** training image and one-shot detection model without adaptation IM GENET detection model for synset bottle

Daniel Goehring 02. June 2014



## 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)

Daniel Goehring 02. June 2014

8



## Whitened HOG Features

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

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

- covariance matrix S<sub>o</sub> 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)

Daniel Goehring 02. June 2014



# **Interactive Learning Interface**

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





## **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

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

Interactive Adaptation of Real-Time Object Detectors

# 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



### **Detection Examples**



Daniel Goehring 02. June 2014



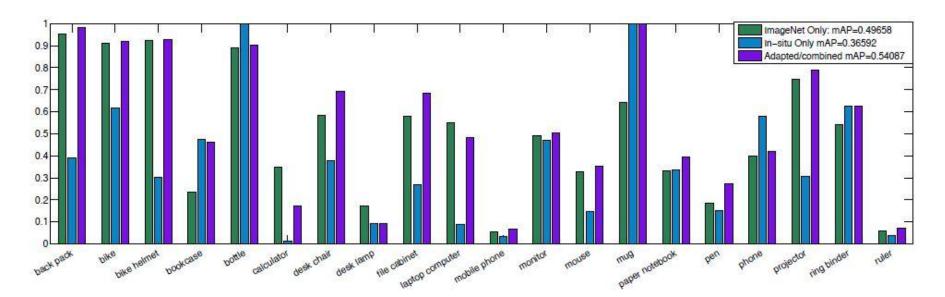
#### **Example: Office Data**

In-situ model only ImageNet model only Adapted model

Daniel Goehring 02. June 2014



### **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

Daniel Goehring 02. June 2014



## 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