

LSDA: Large Scale Detection through Adaptation

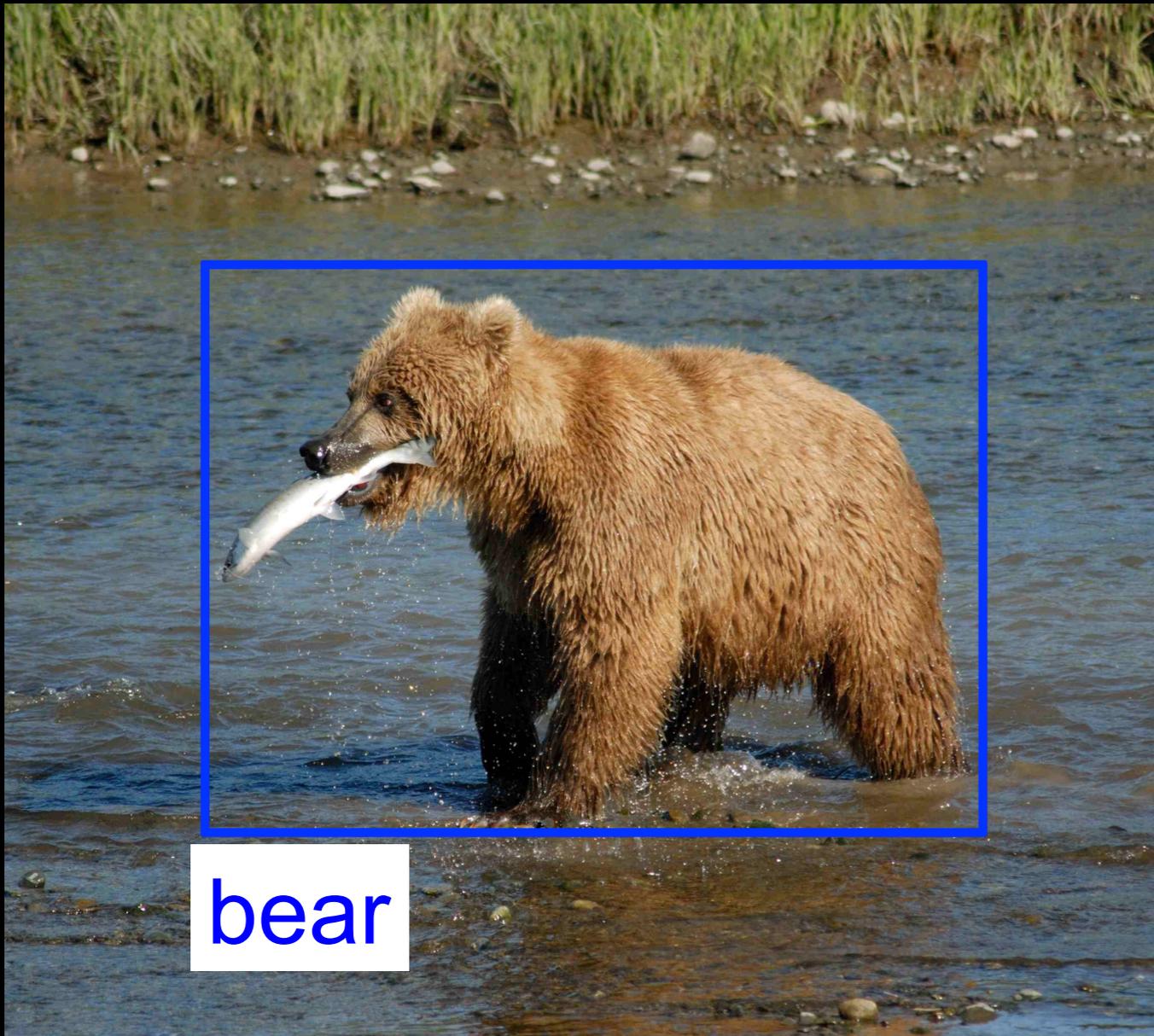
Judy Hoffman*, Sergio Guadarrama*, Eric Tzeng*,
Ronghang Hu*, Jeff Donahue*, Ross Girshick*,
Trevor Darrell*, Kate Saenko^o
*UC Berkeley, ^oUMass Lowell

ImageNet

Dataset: Millions of images, >10K classes



State of the art: 200 class detection



bear



bear

What we want



Grizzly bear



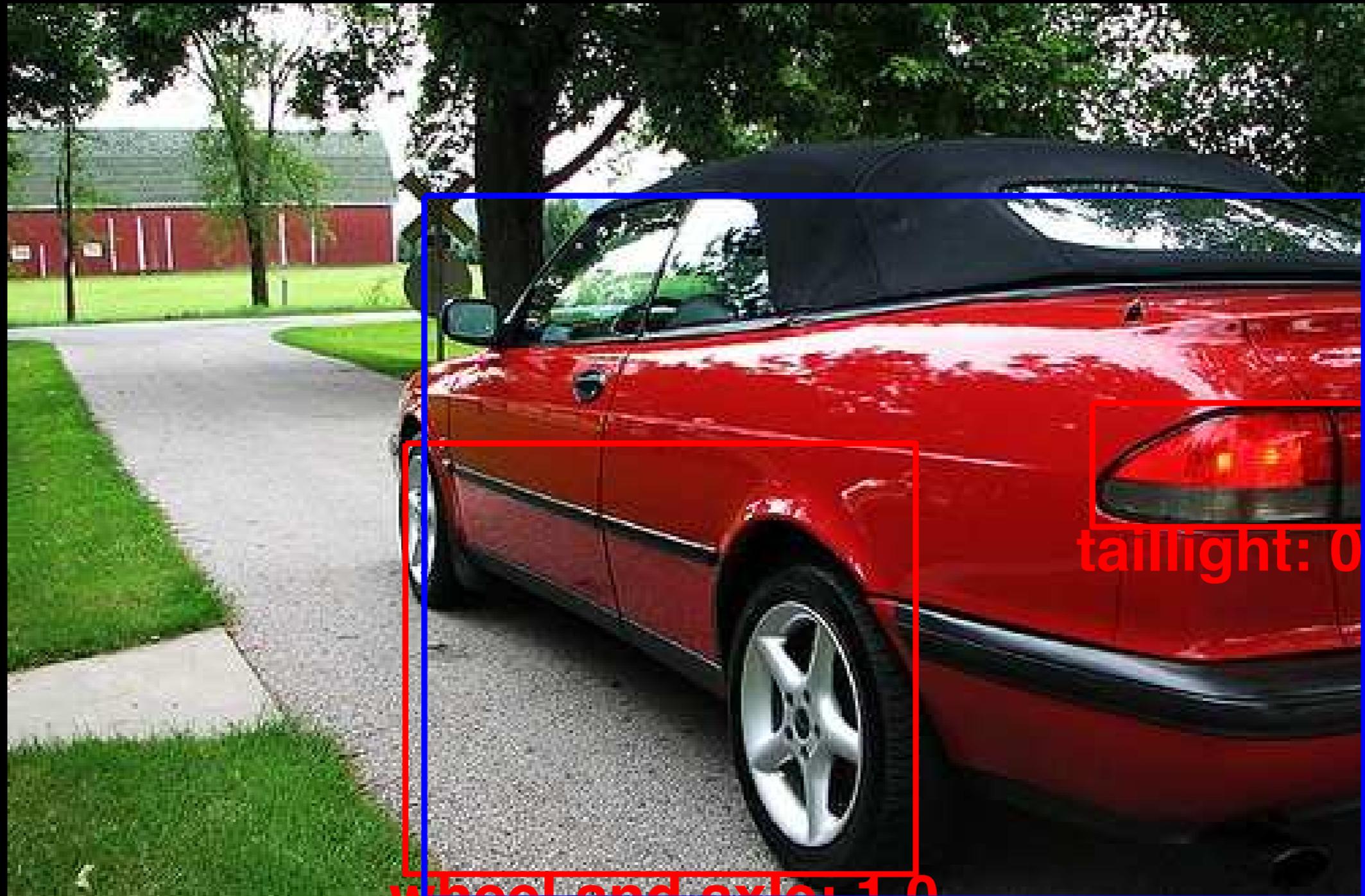
Teddy bear

State of the art: 200 class detection



car

Our Model: 7.5K classes



car: 6.0
wheel and axle: 1.0

taillight: 0.9

Object classification

{ airplane, bird, motorbike, person, sofa }



Input

person
motorbike

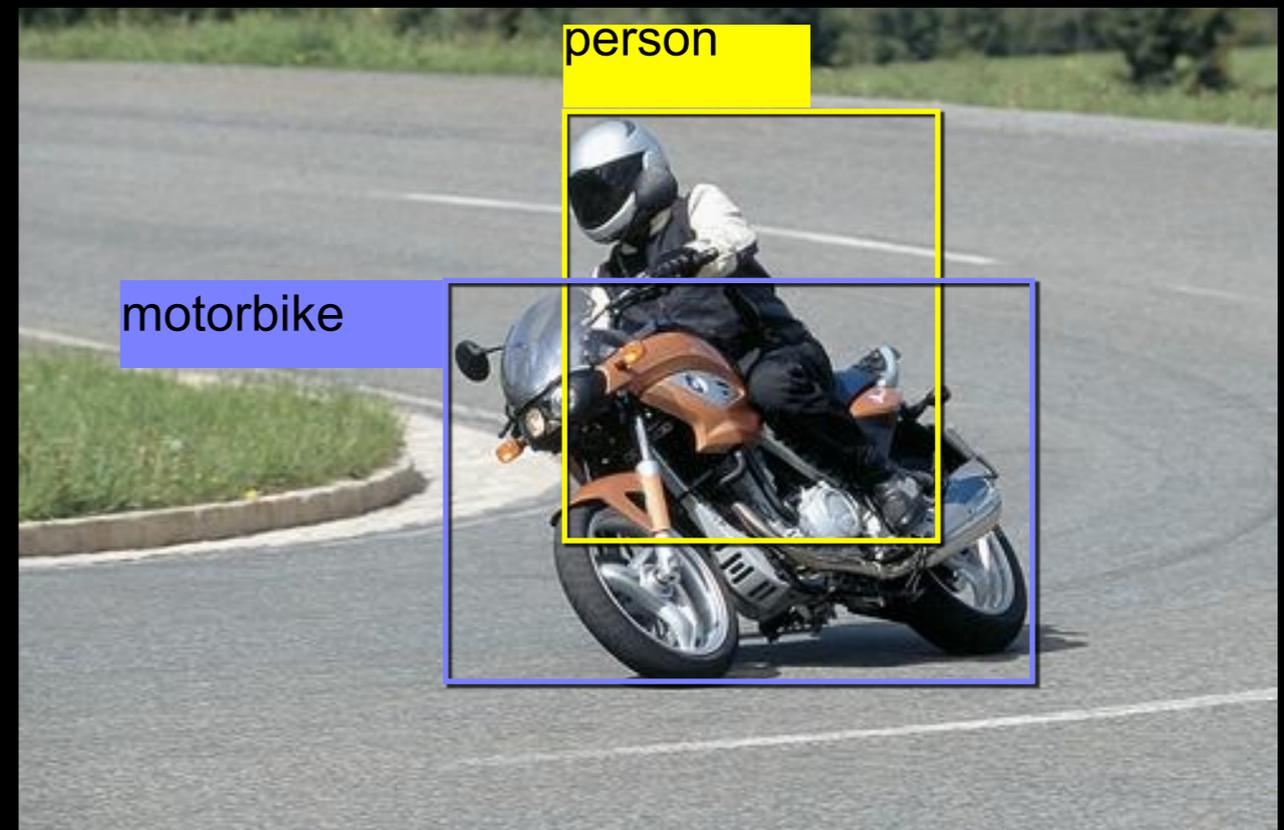
Desired output

Object detection

{ airplane, bird, motorbike, person, sofa }



Input



Desired output

ImageNet-LSVRC Classification

- 1000 classes
- 1.2 million training images
- Image classification

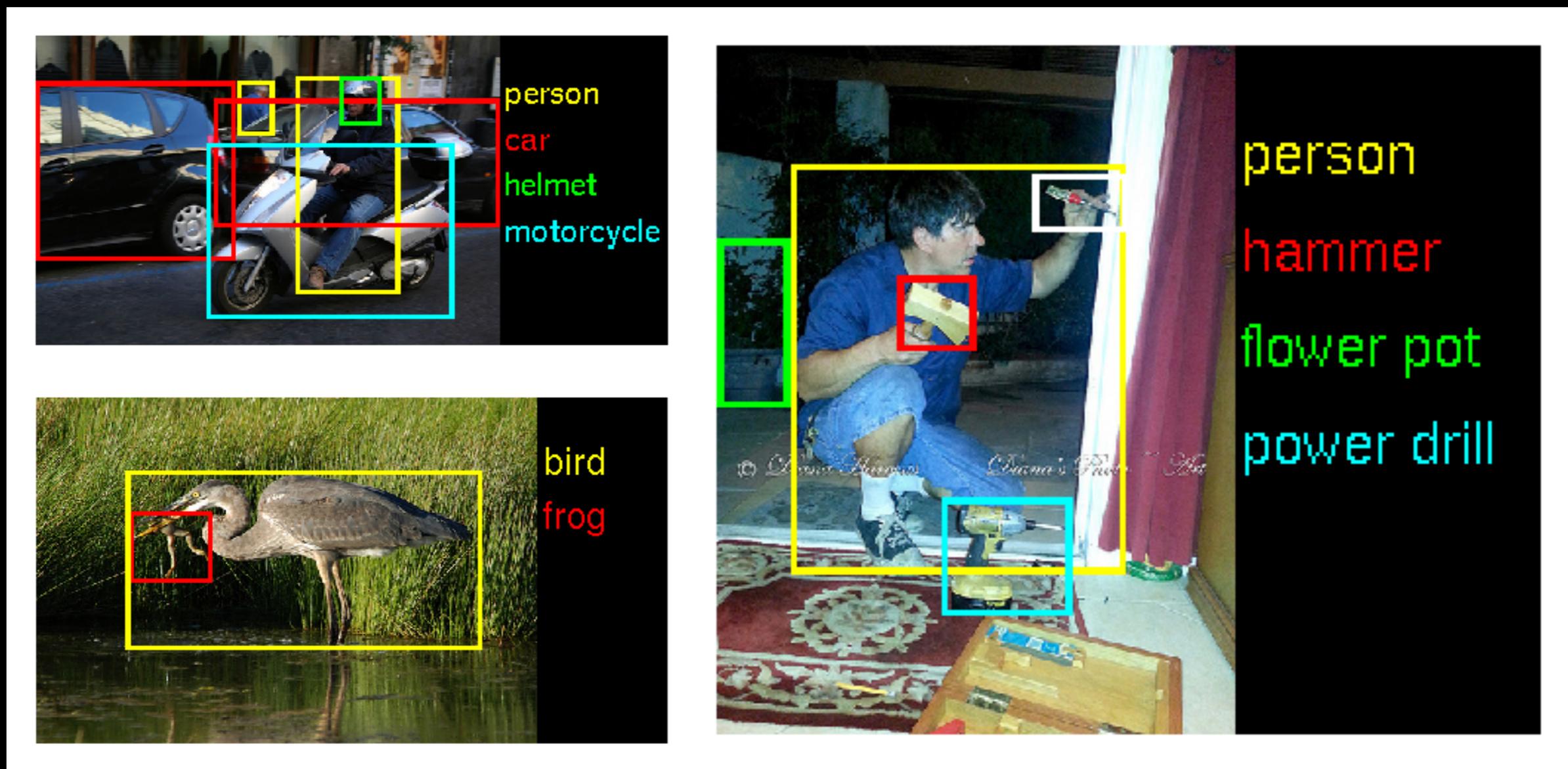


bus anywhere?

[Deng et al. CVPR'09]

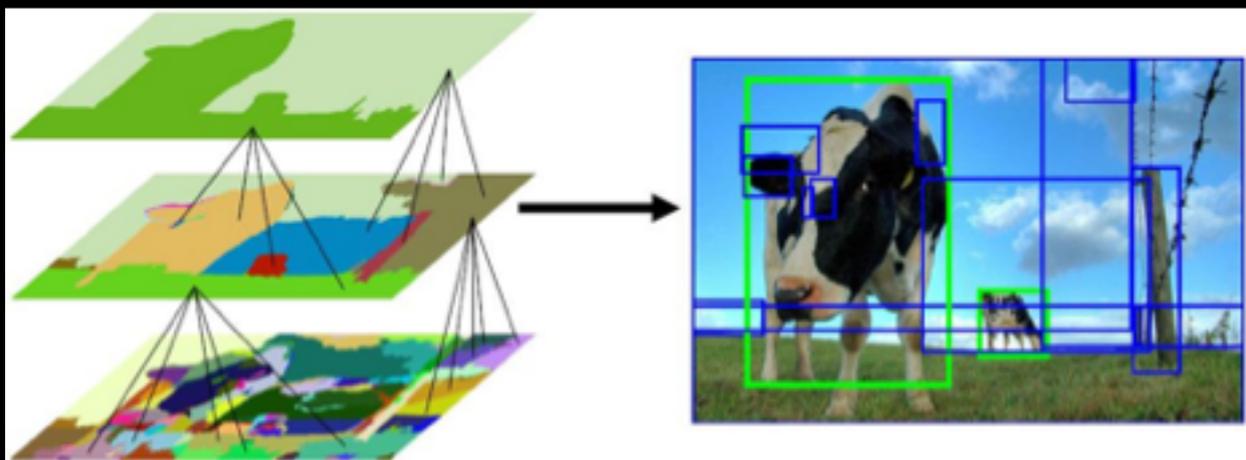
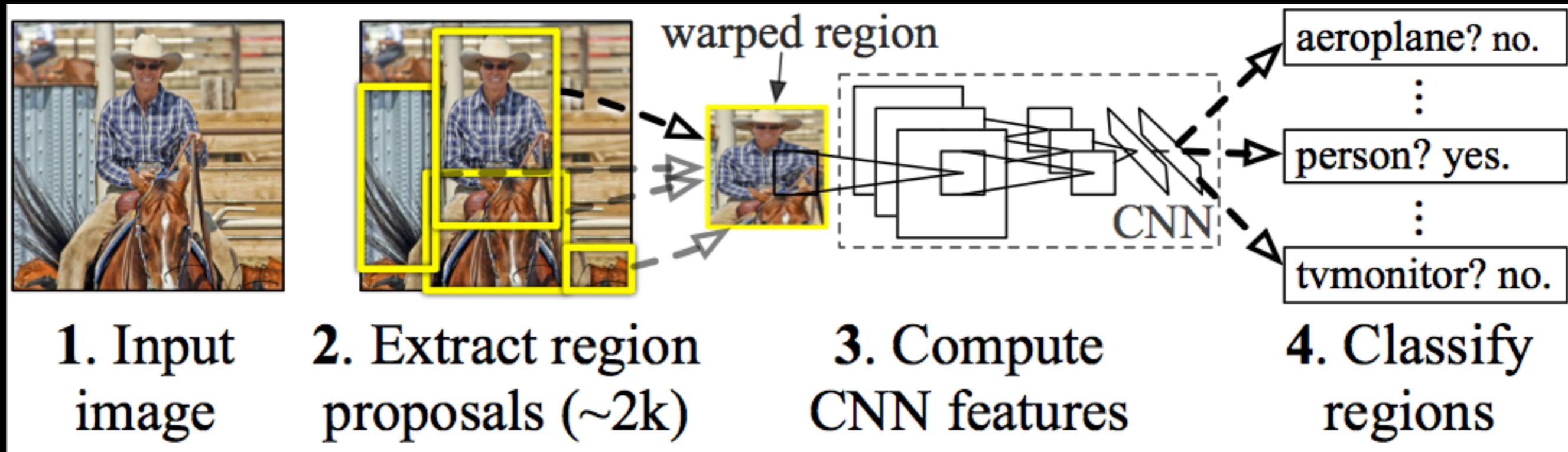
ImageNet-LSVRC Detection

Dataset: 400k images, 350k objects, 200 classes



Detect: people, horses, sofas, bicycles, pizza, ...

R-CNN: “Regions with CNN features”



[Girshick et al. CVPR'14]

Alternative approach: “overfeat”
[Sermanet et al. ICLR'14]

“selective search” [van de Sande et al. 2011]

ImageNet

Dataset: Millions of images, >10K classes



Can we produce detectors for all of ImageNet?
And use *all* available labeled image data?

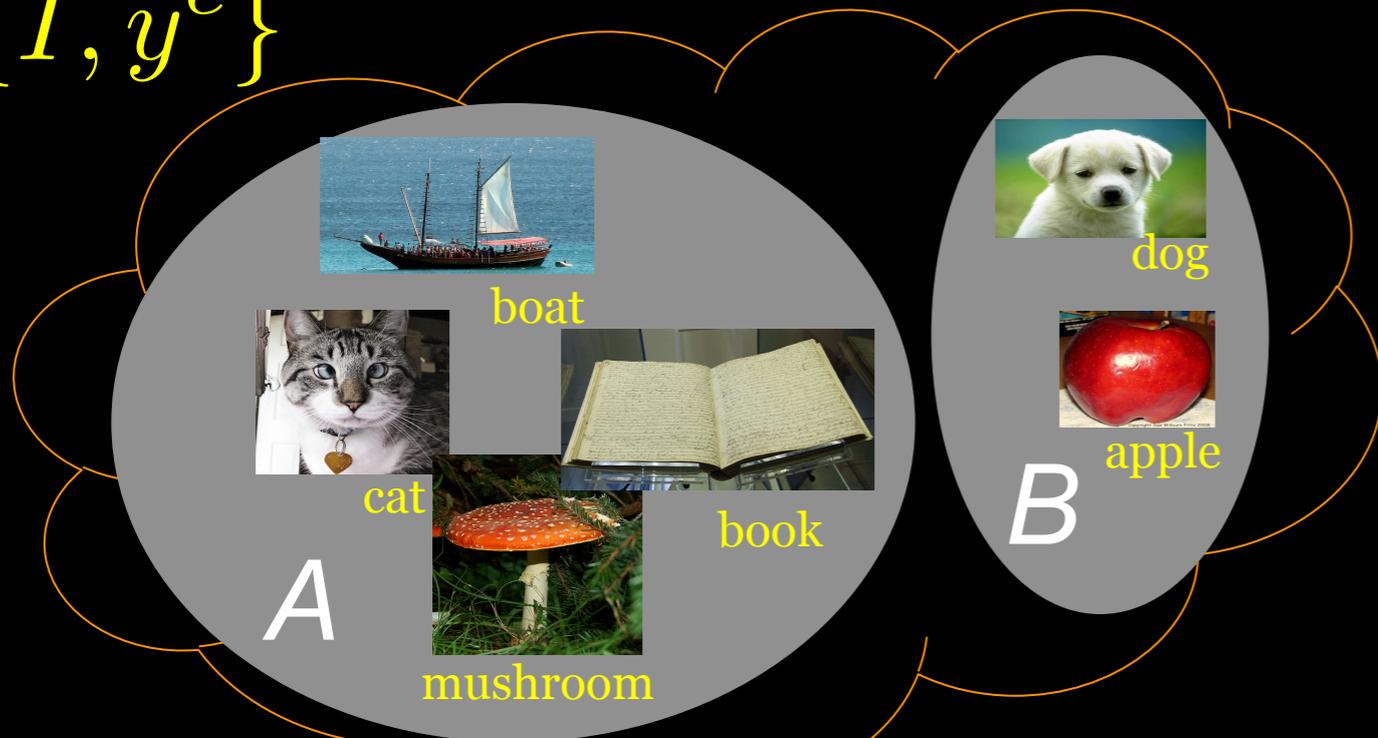
Bounding boxes for only 200 classes
Nearly all models only use 1K class images

Adaptation Paradigm

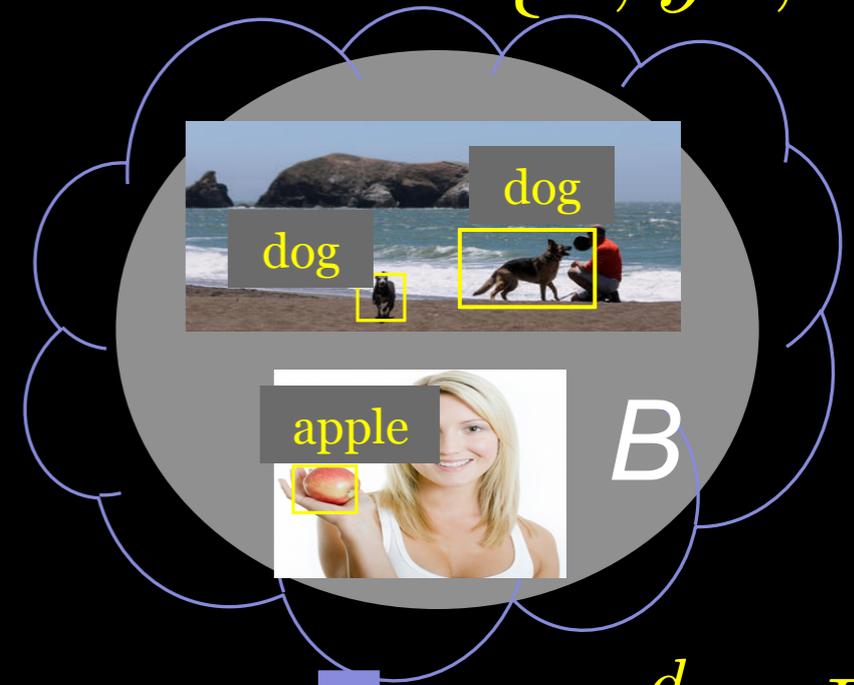
- Weak-label learning commonly considered as a MIL problem
 - requires inference per category.
- Consider a *domain adaptation* paradigm
 - Is there something common to “detection”?

Transform Classifiers into Detectors

$\{I, y^c\}$



$\{I, y^d, b\}$



$y^c \in \{A \cup B\}$

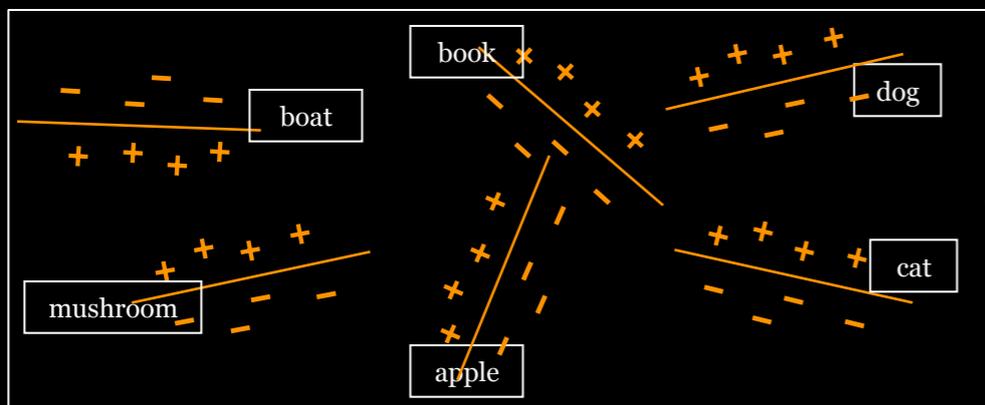
$y^d \in B$

$\phi^c(I)$

?

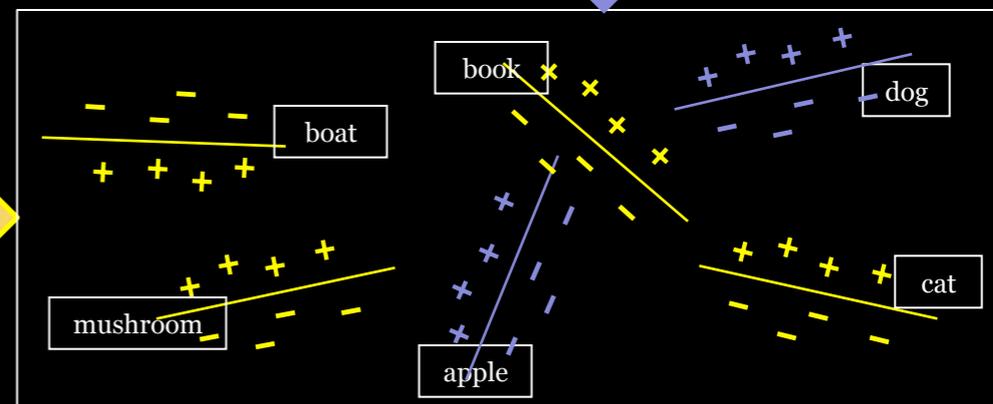
$\phi^d(I(b))$

W^c

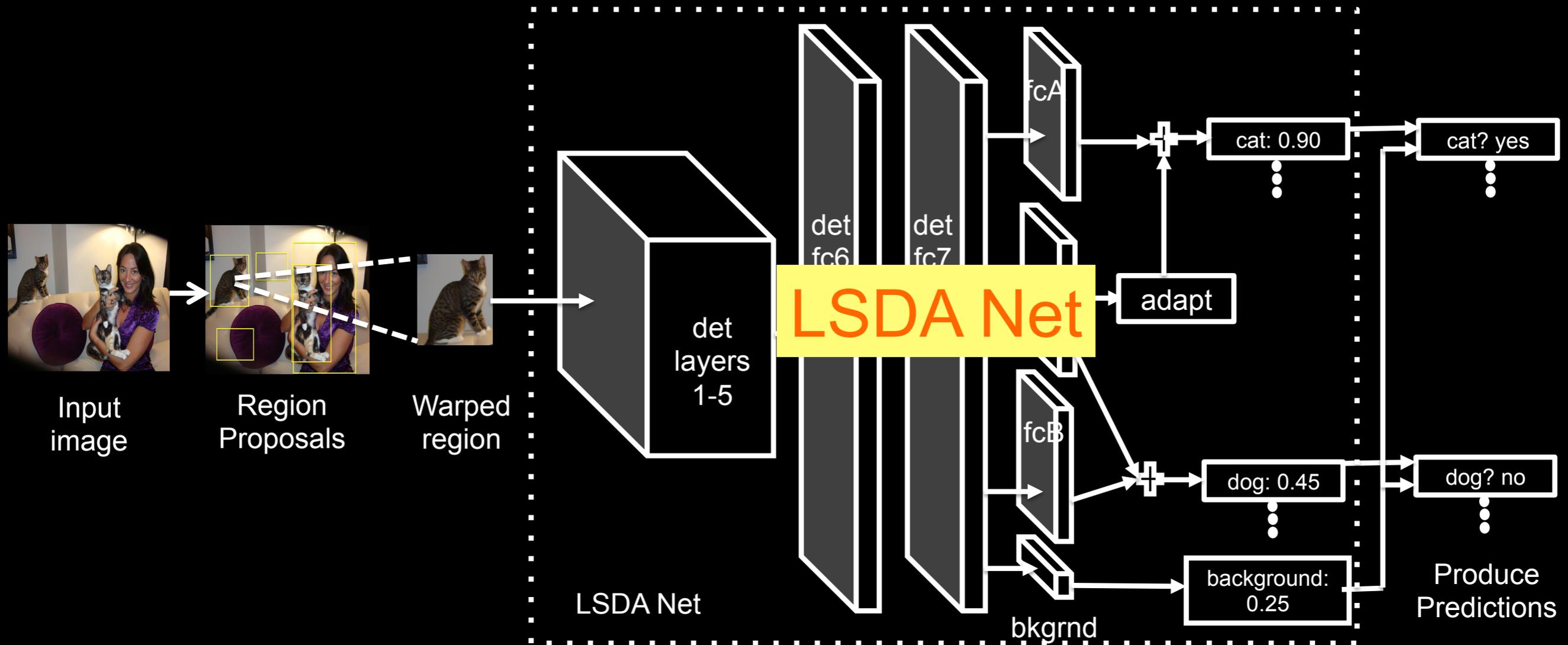


?

W^d



LSDA Overview



LSDA Overview

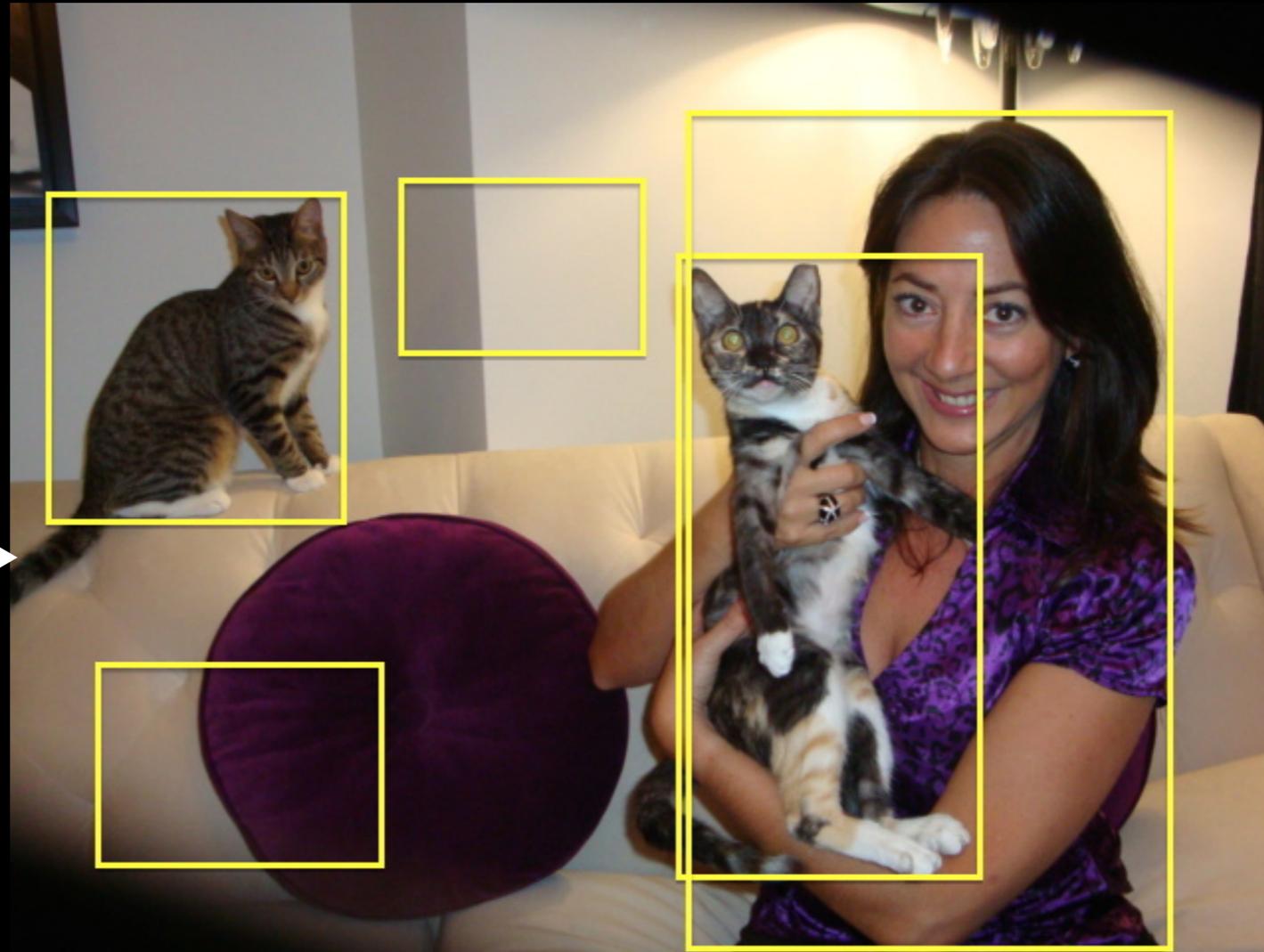


Input Image

LSDA Overview

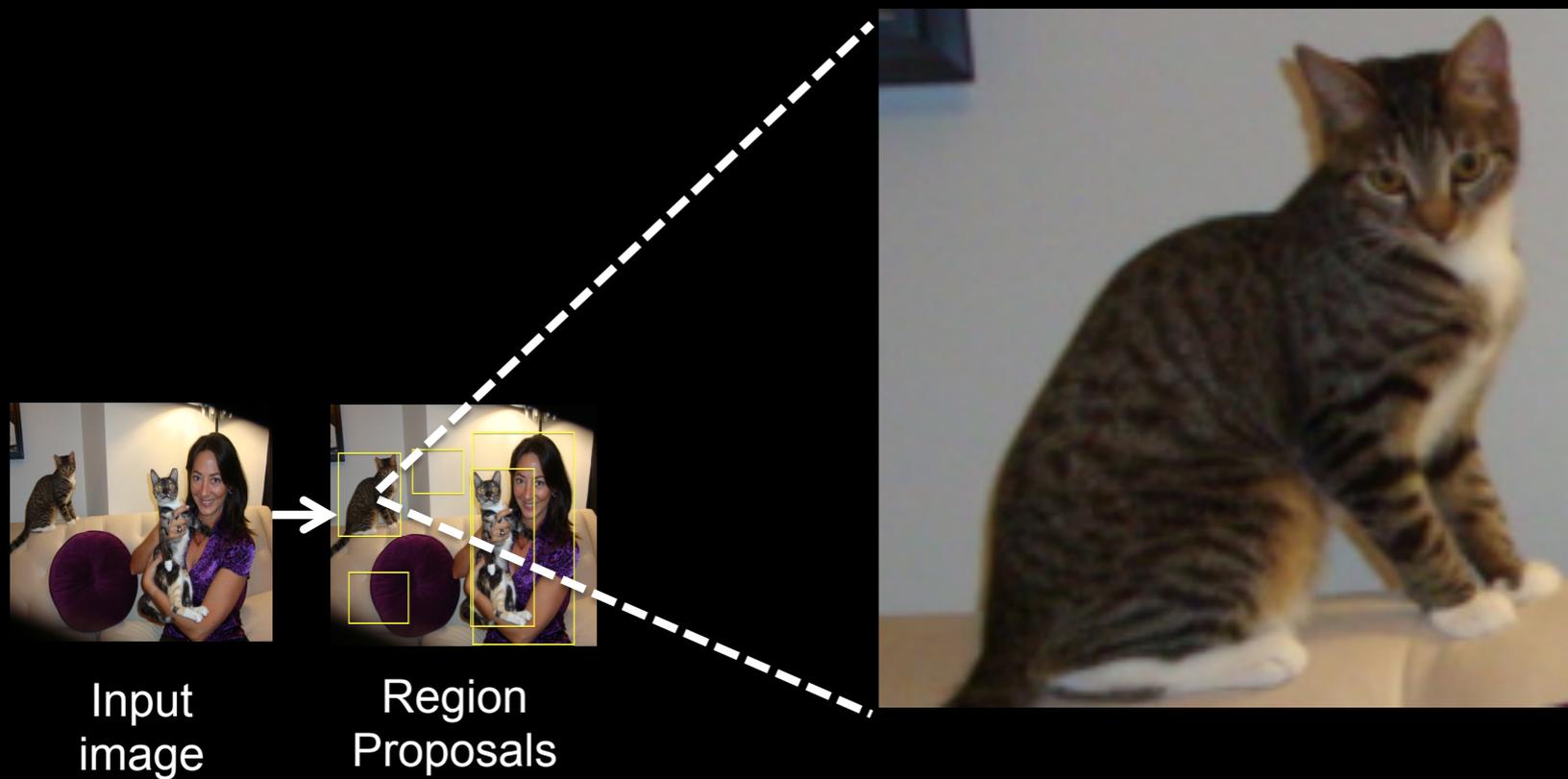


Input image



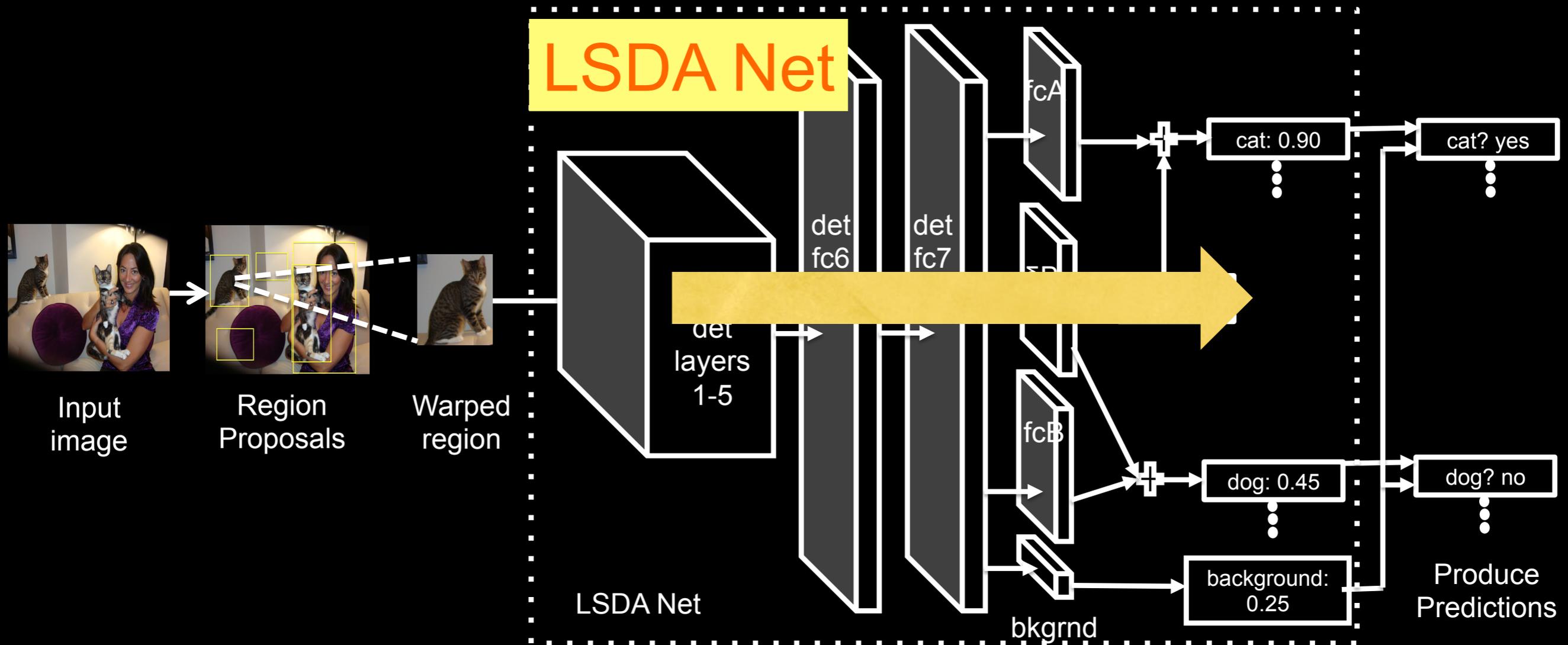
Region Proposals

LSDA Overview



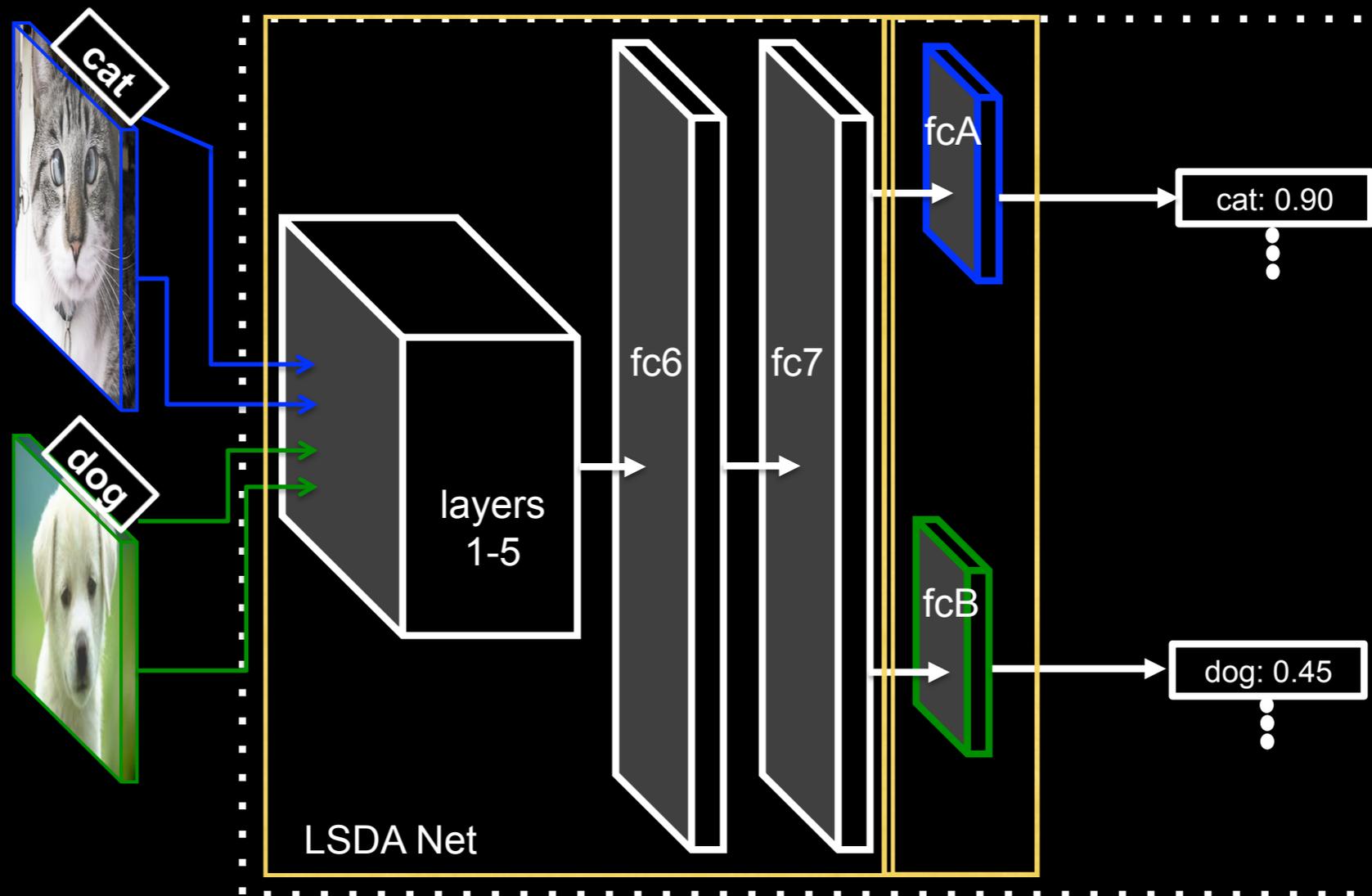
Select and Warp a Region

LSDA Overview



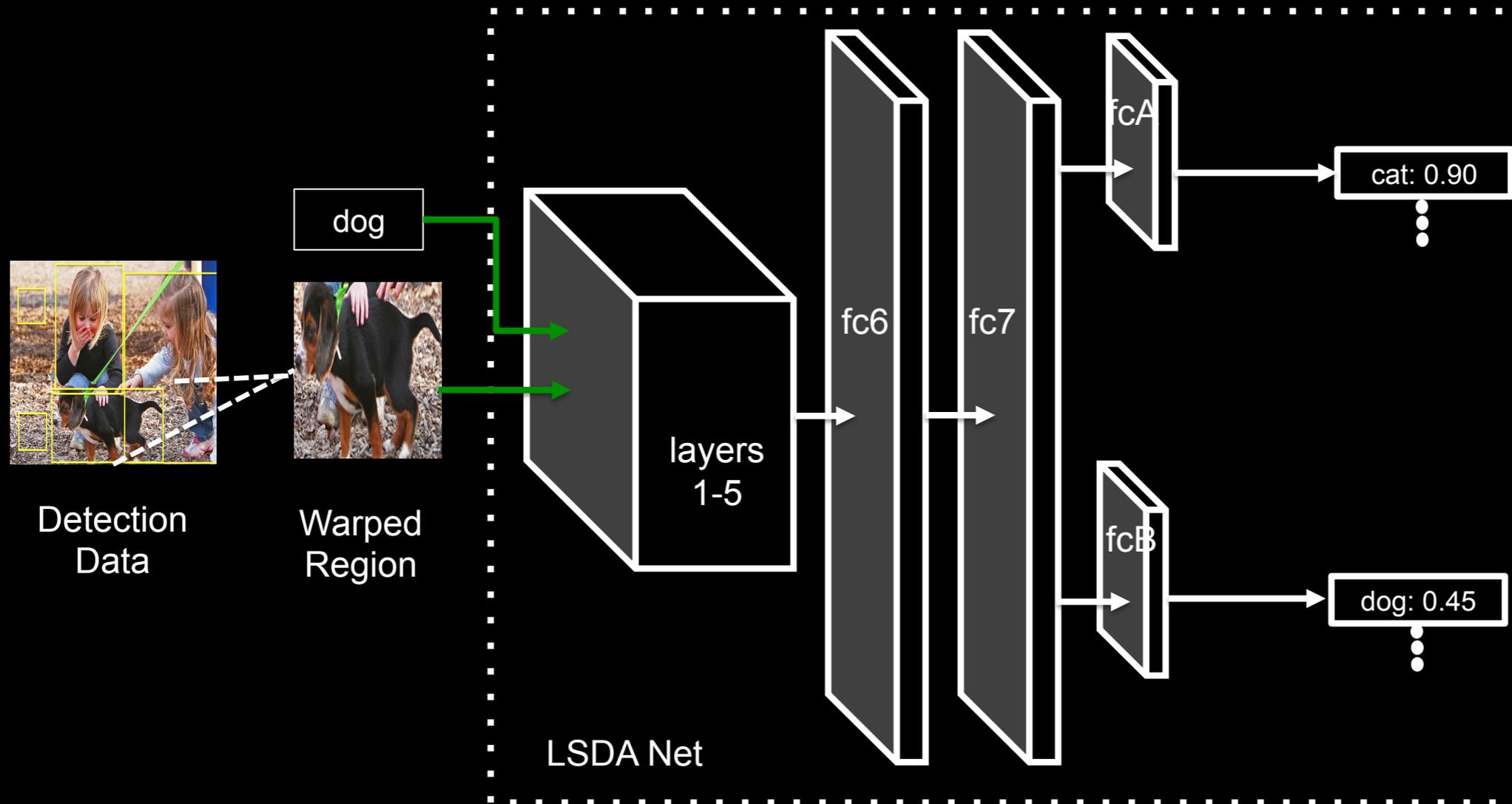
Compute Region Scores

Training LSDA



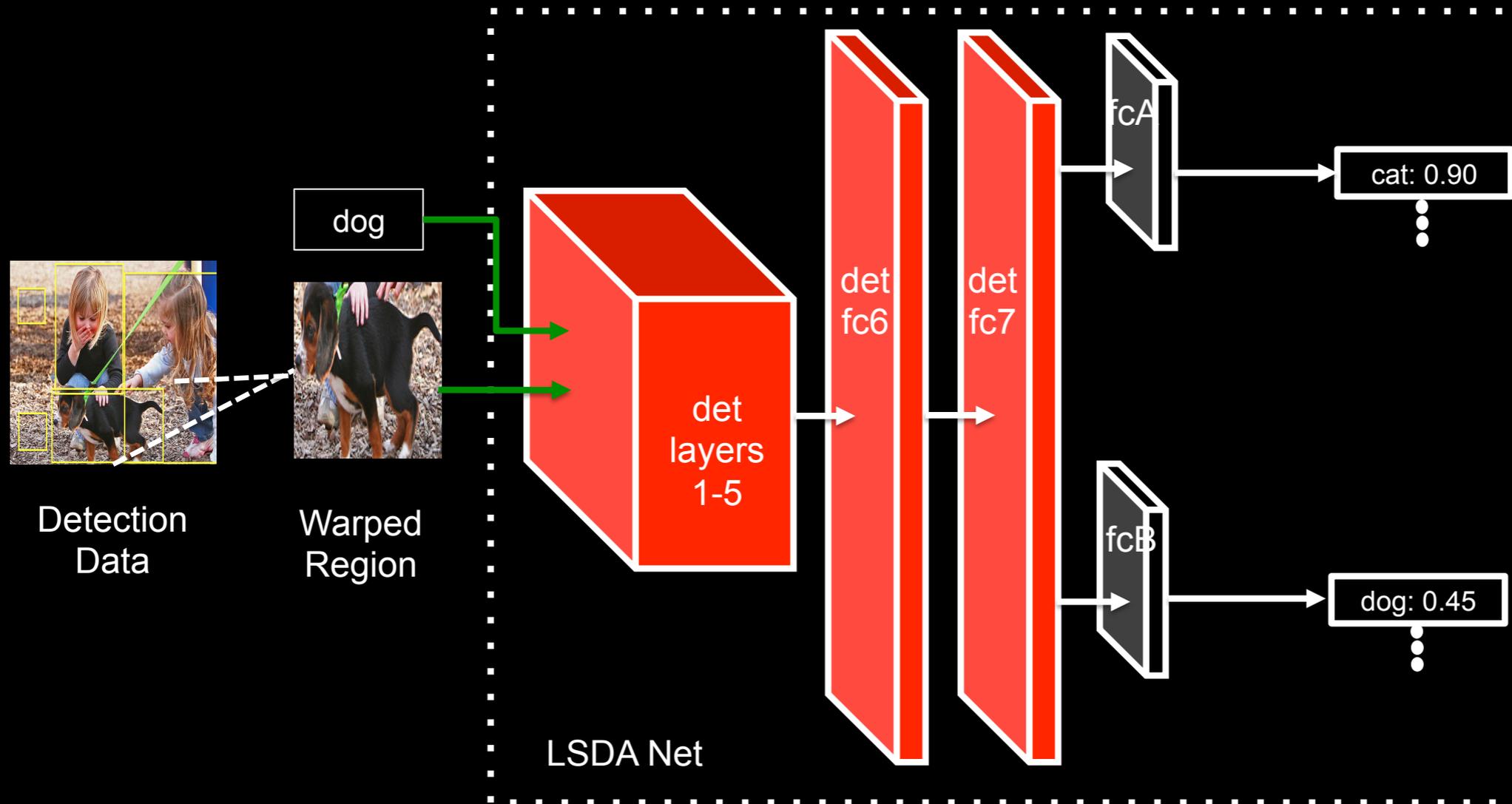
Classification Net

Training LSDA



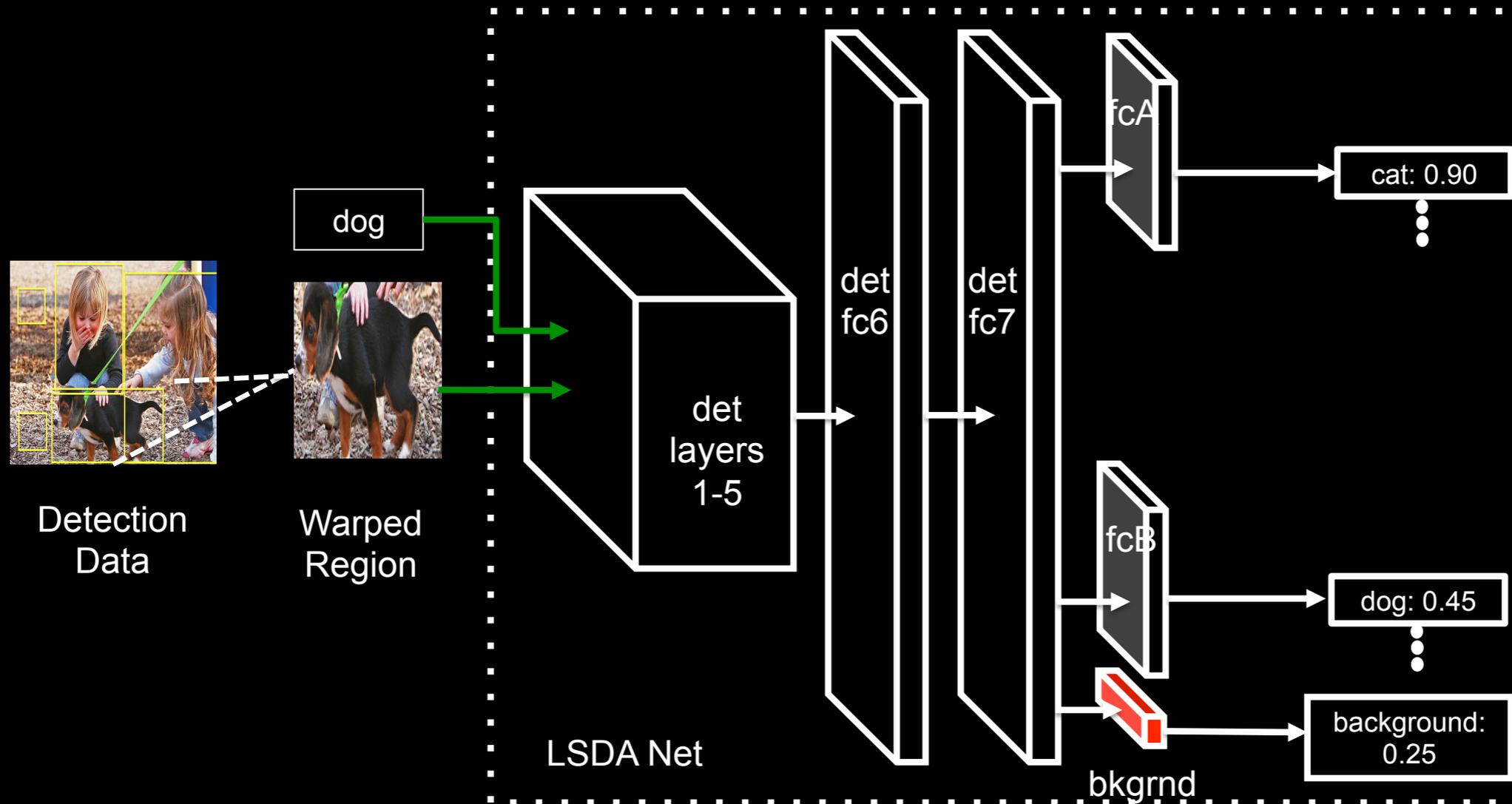
Fine-tune Representation with Detection Data

Training LSDA



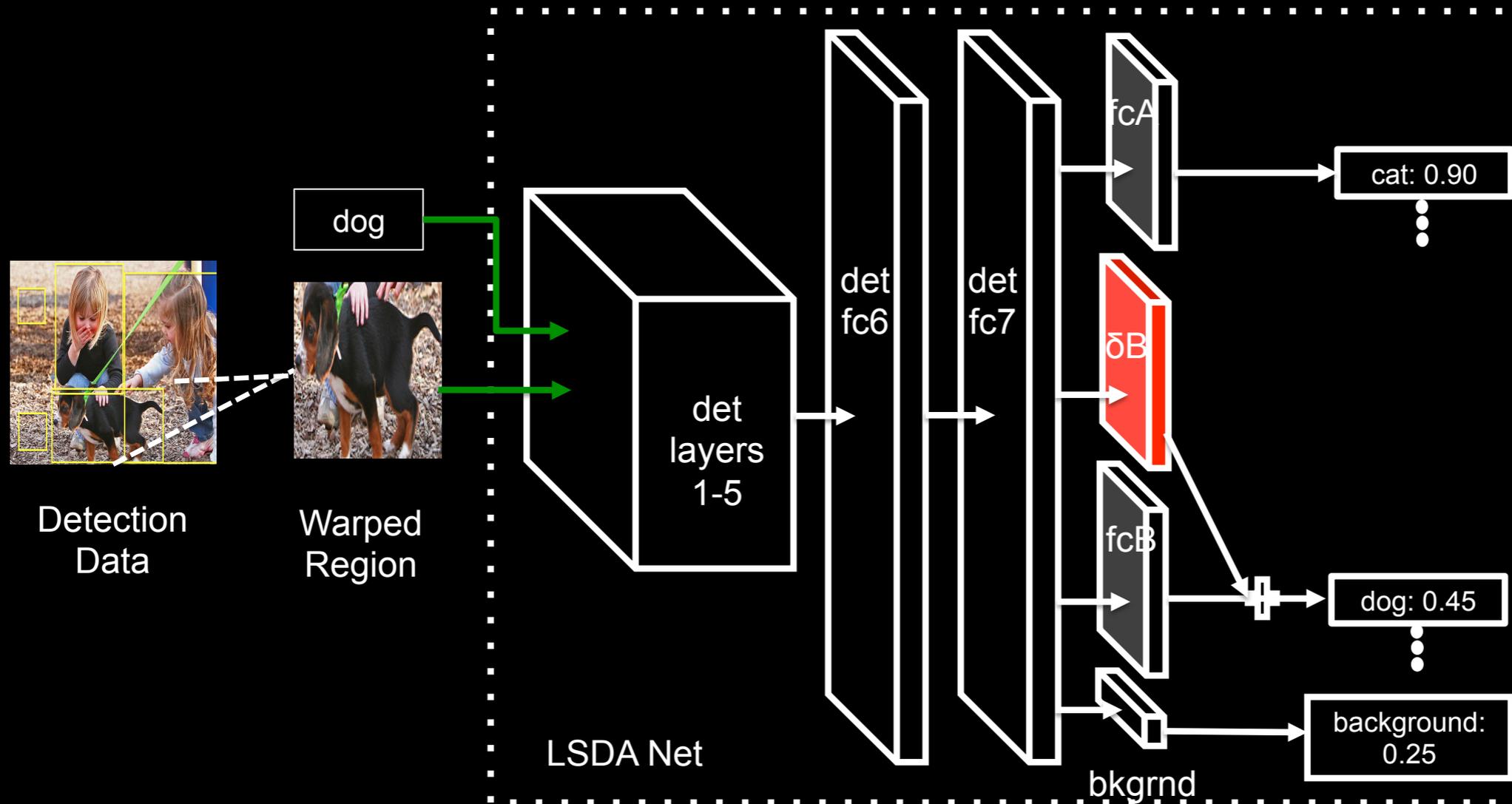
Fine-tune Representation with Detection Data

Training LSDA



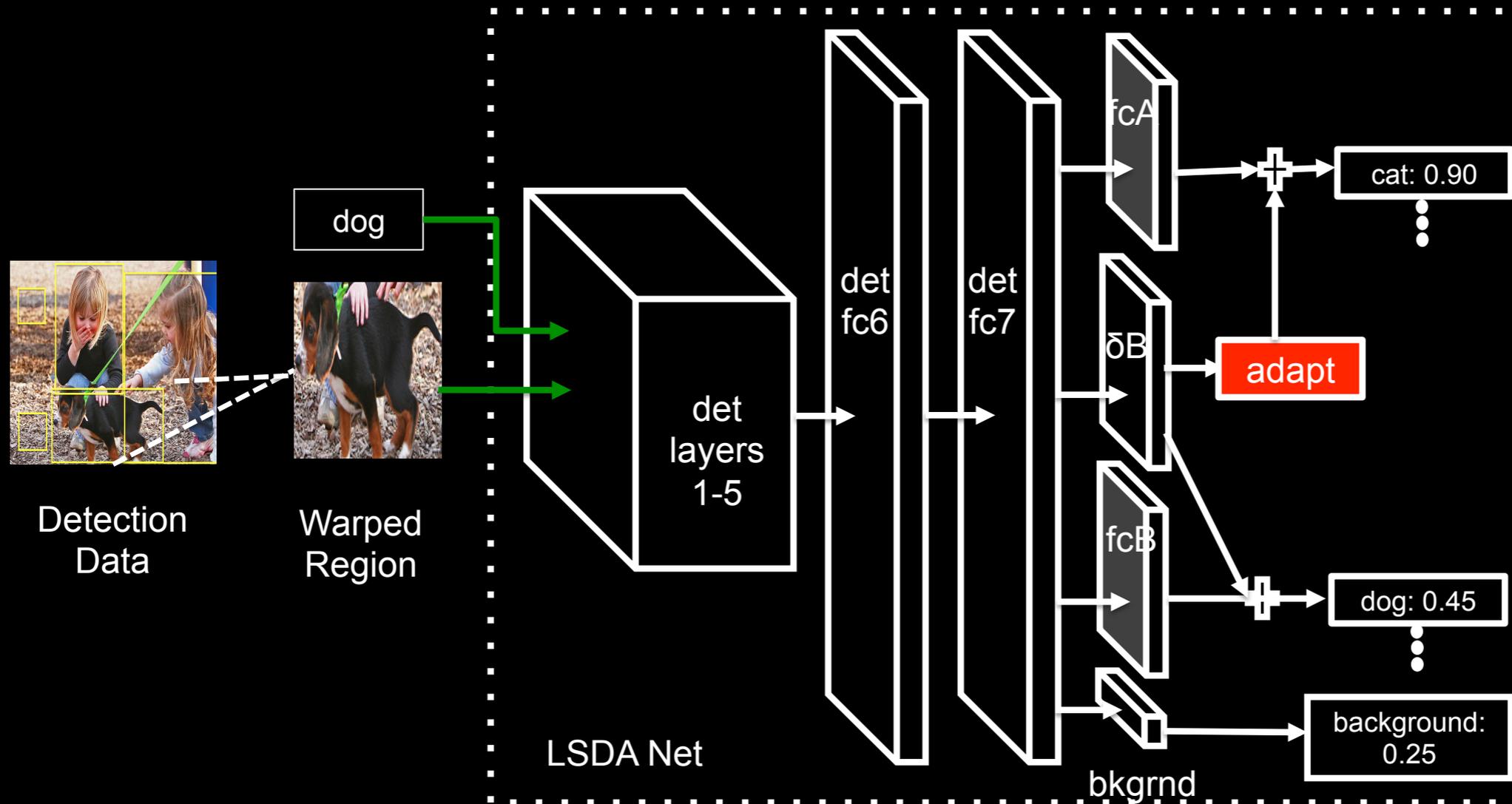
Learn a background class detector

Training LSDA



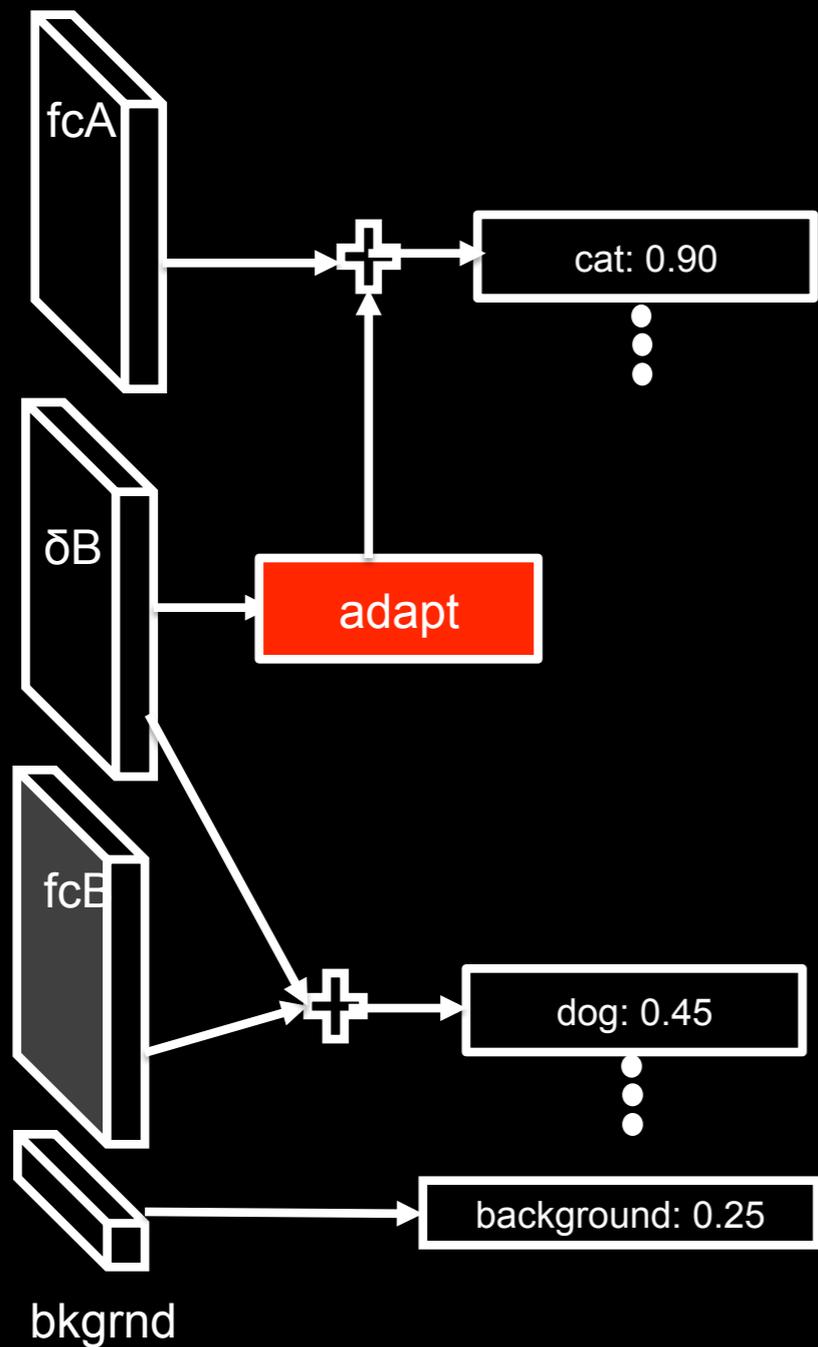
Fine-tune Category Specific Parameters

Training LSDA



Adapt output layer for held-out classes

LSDA Model: Adapt



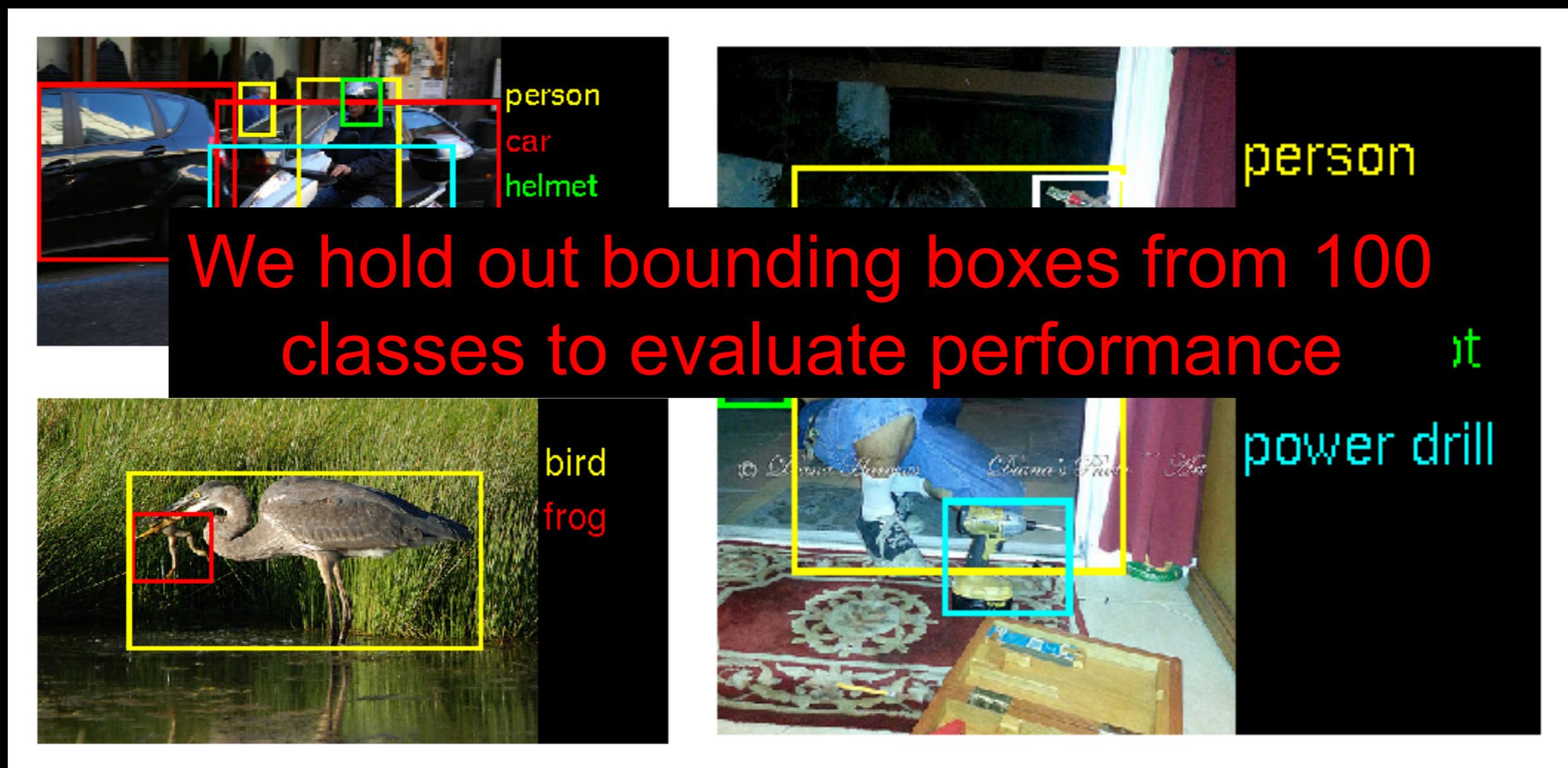
For category i in set A find j^{th} nearest neighbor in set B :

adapt

$$\delta A_i = \frac{1}{K} \sum_{j=1}^K \delta B_{N_B(A_i, j)}$$

ImageNet-LSVRC Detection

Dataset: 400k images, 350k objects, 200 classes



Detect: people, horses, sofas, bicycles, pizza, ...

Evaluating a detector



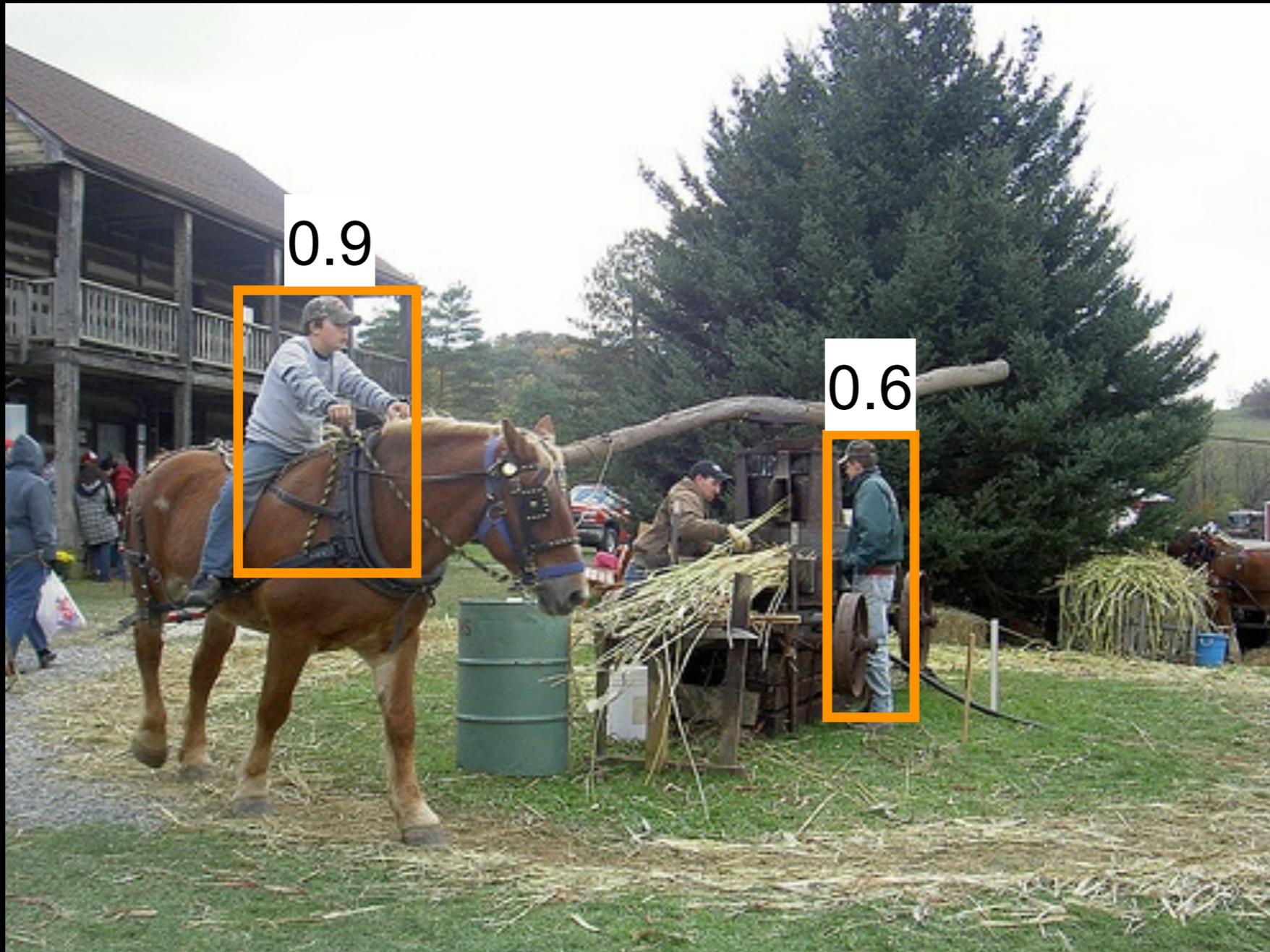
Test image (previously unseen)

First detection ...



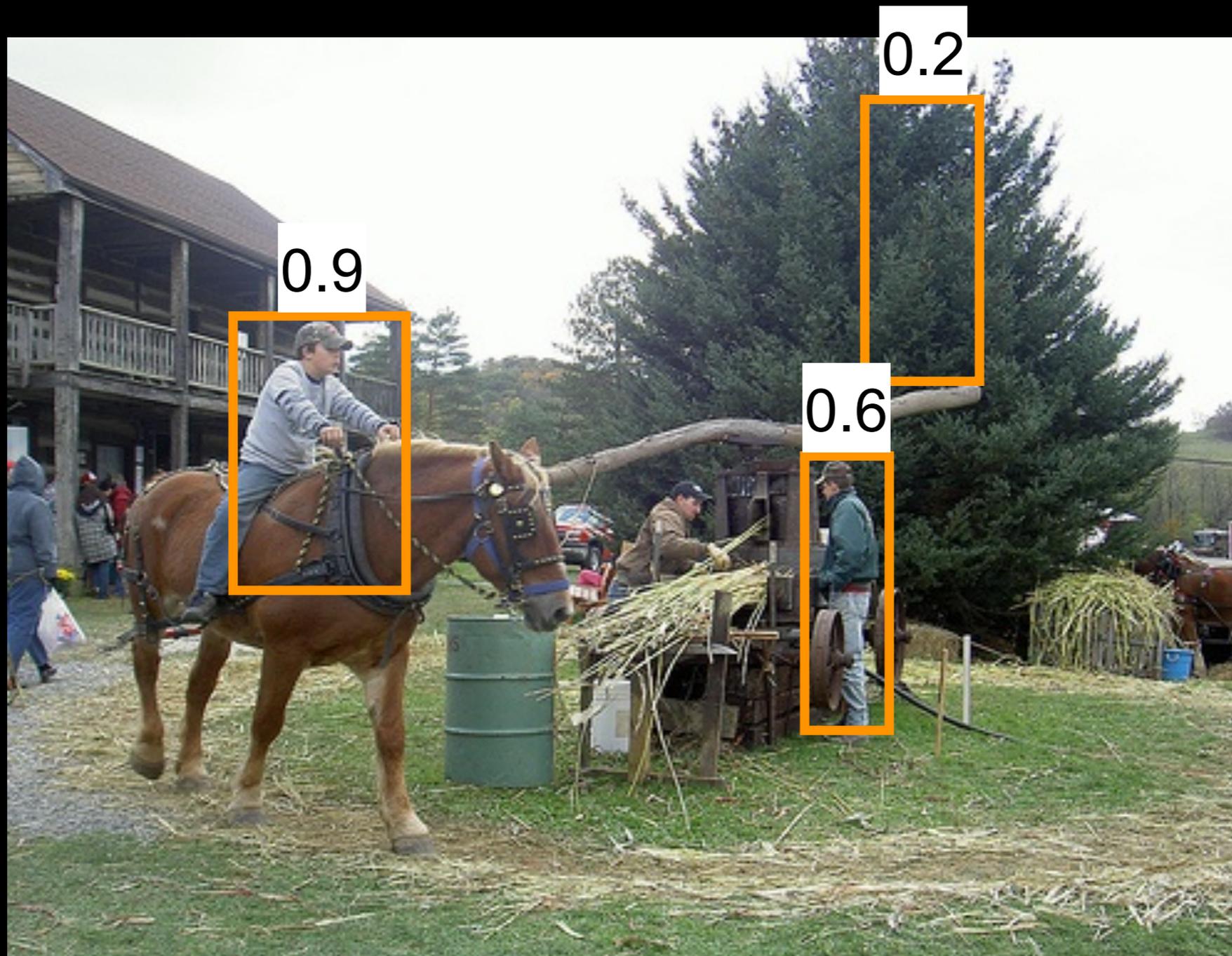
□ 'person' detector predictions

Second detection ...



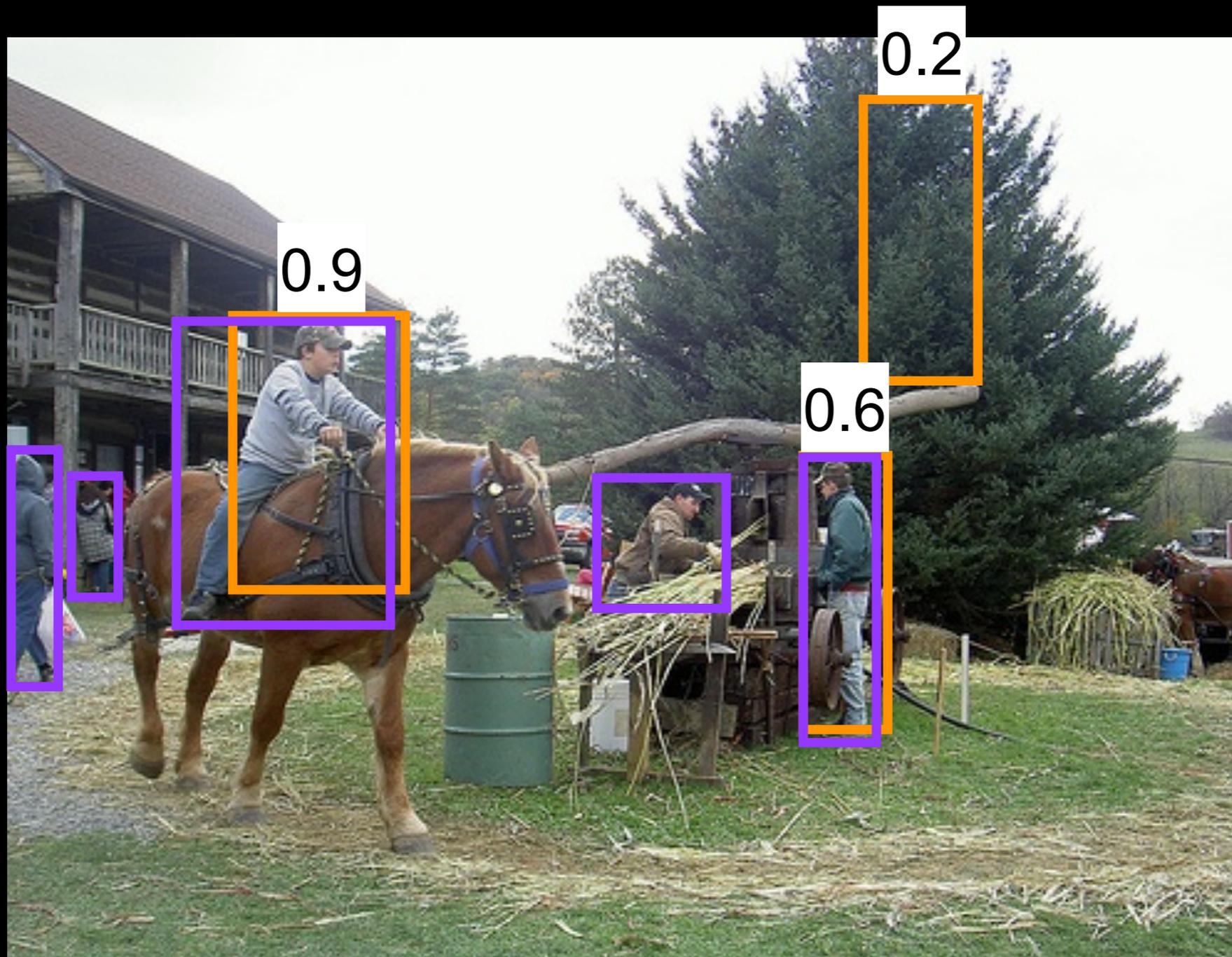
 'person' detector predictions

Third detection ...



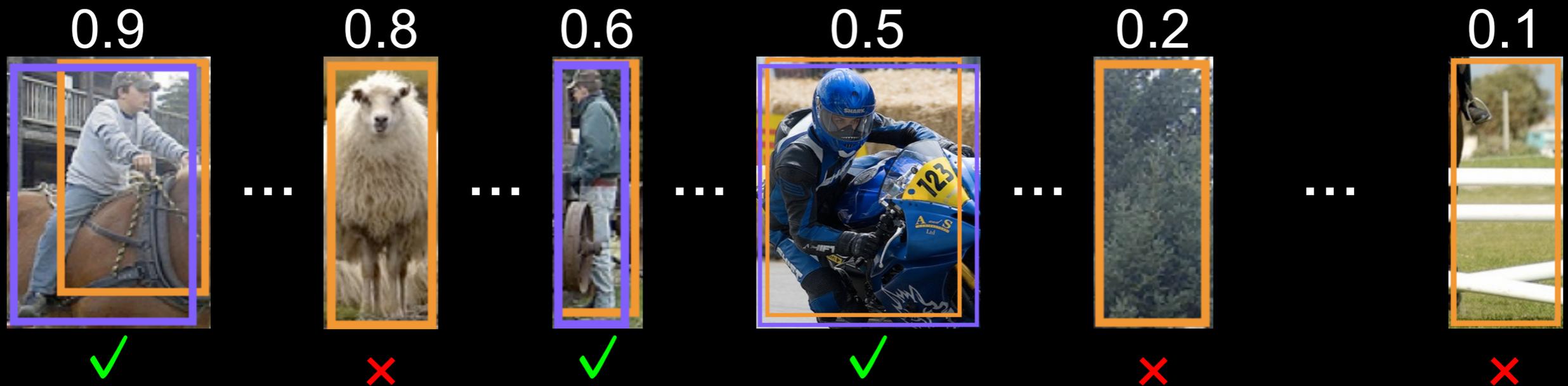
□ 'person' detector predictions

Compare to ground truth



-  'person' detector predictions
-  ground truth 'person' boxes

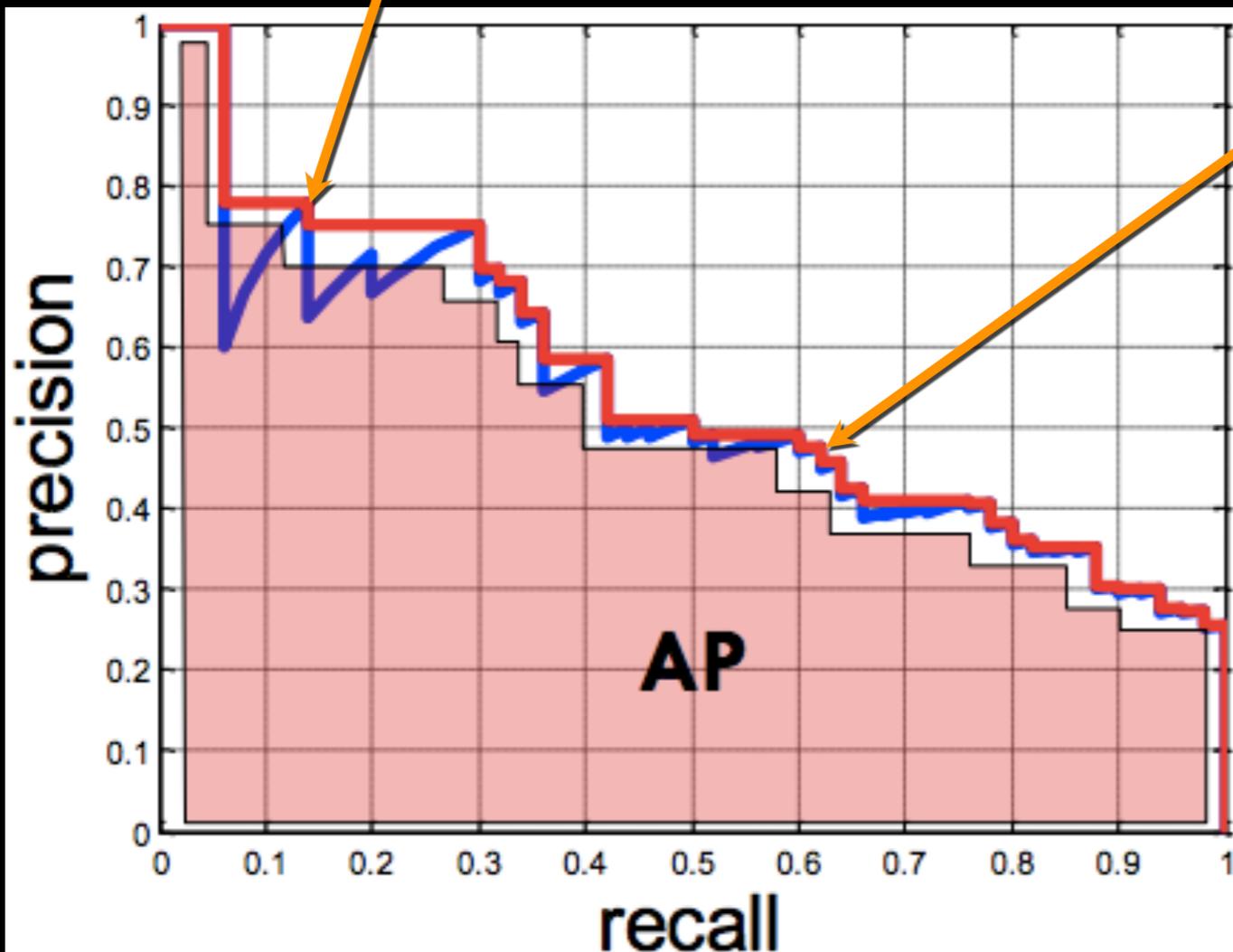
Sort by confidence



true
positive
(high overlap)

false
positive
(low overlap or
duplicate)

Evaluation metric



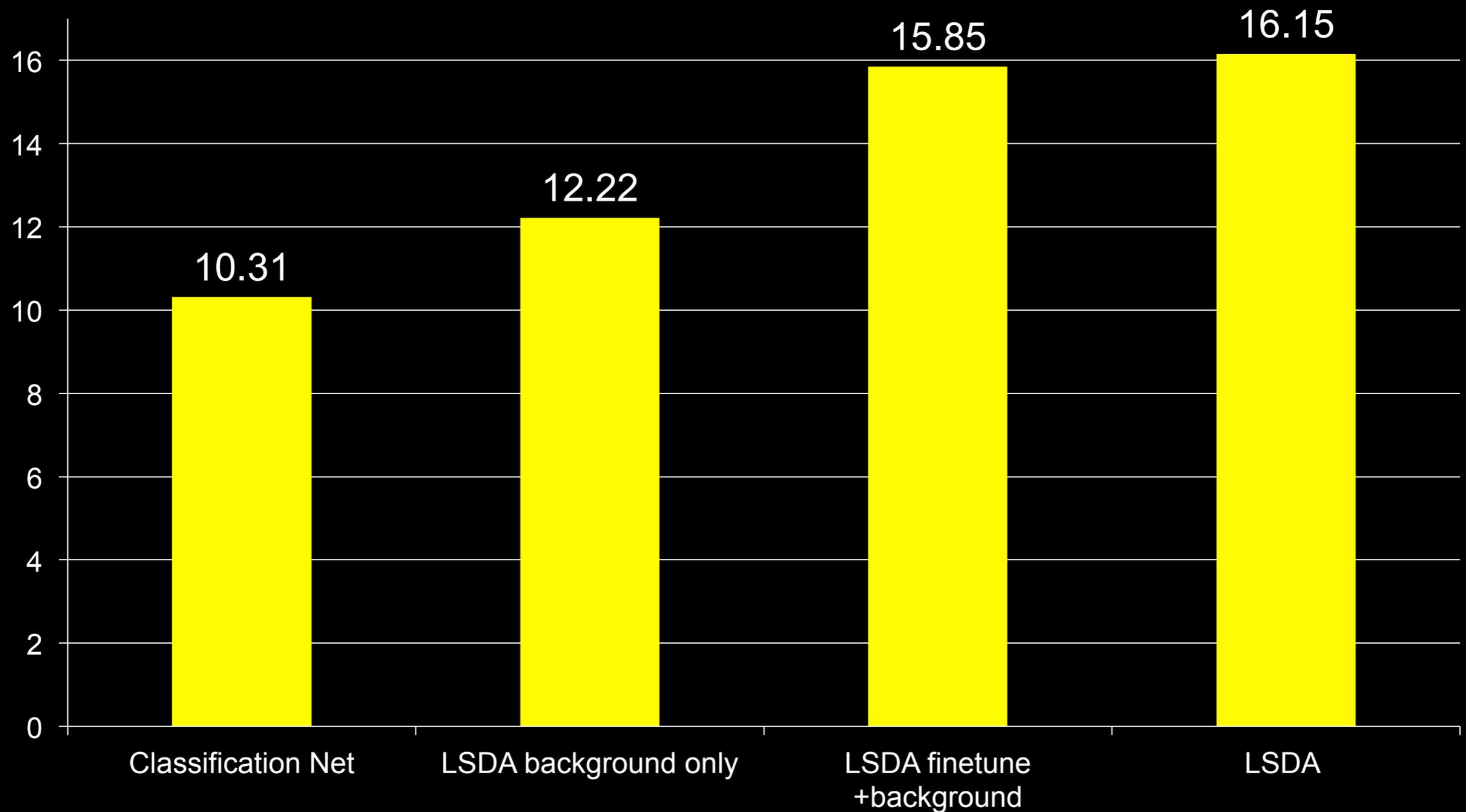
Average Precision (AP)

0% is worst

100% is best

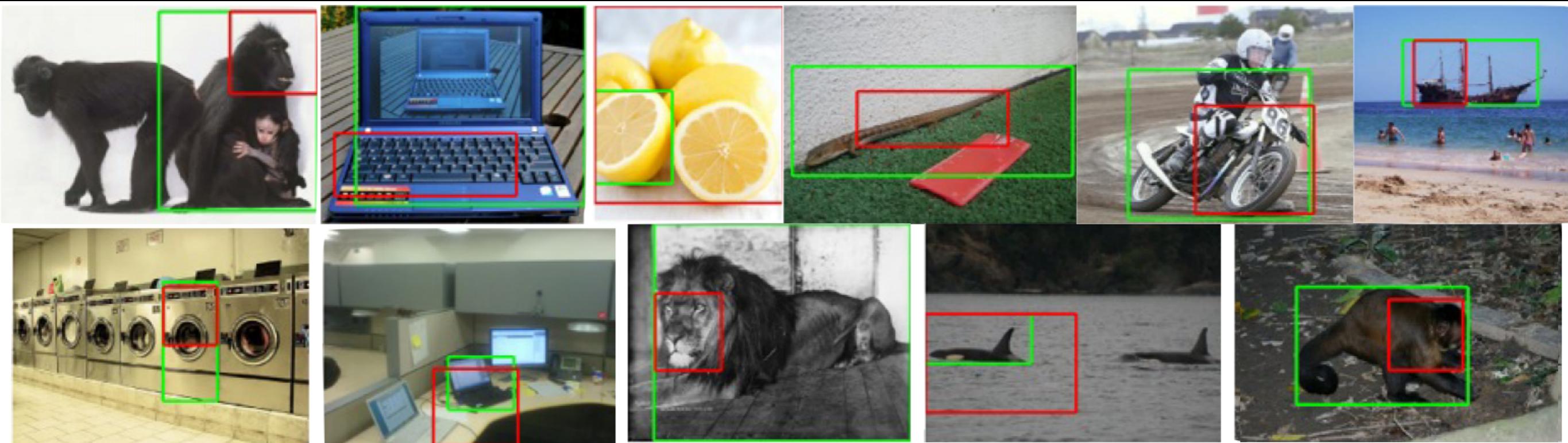
mean AP over classes
(mAP)

mAP (%) Held-out 100 Categories

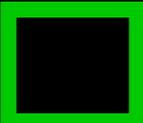


Oracle performance on this set is 26.25

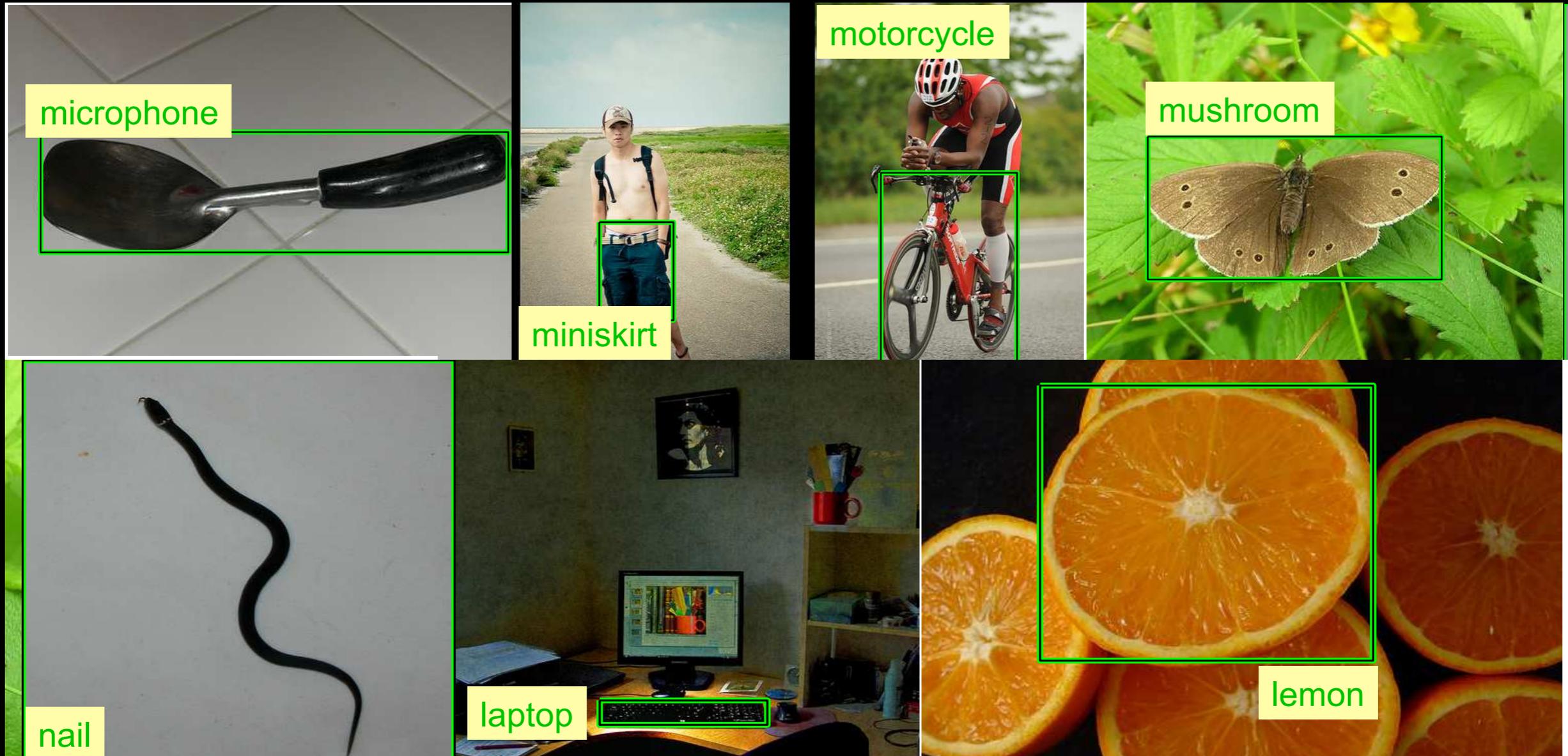
Improved Localization



 Classification Network

 LSDA Network

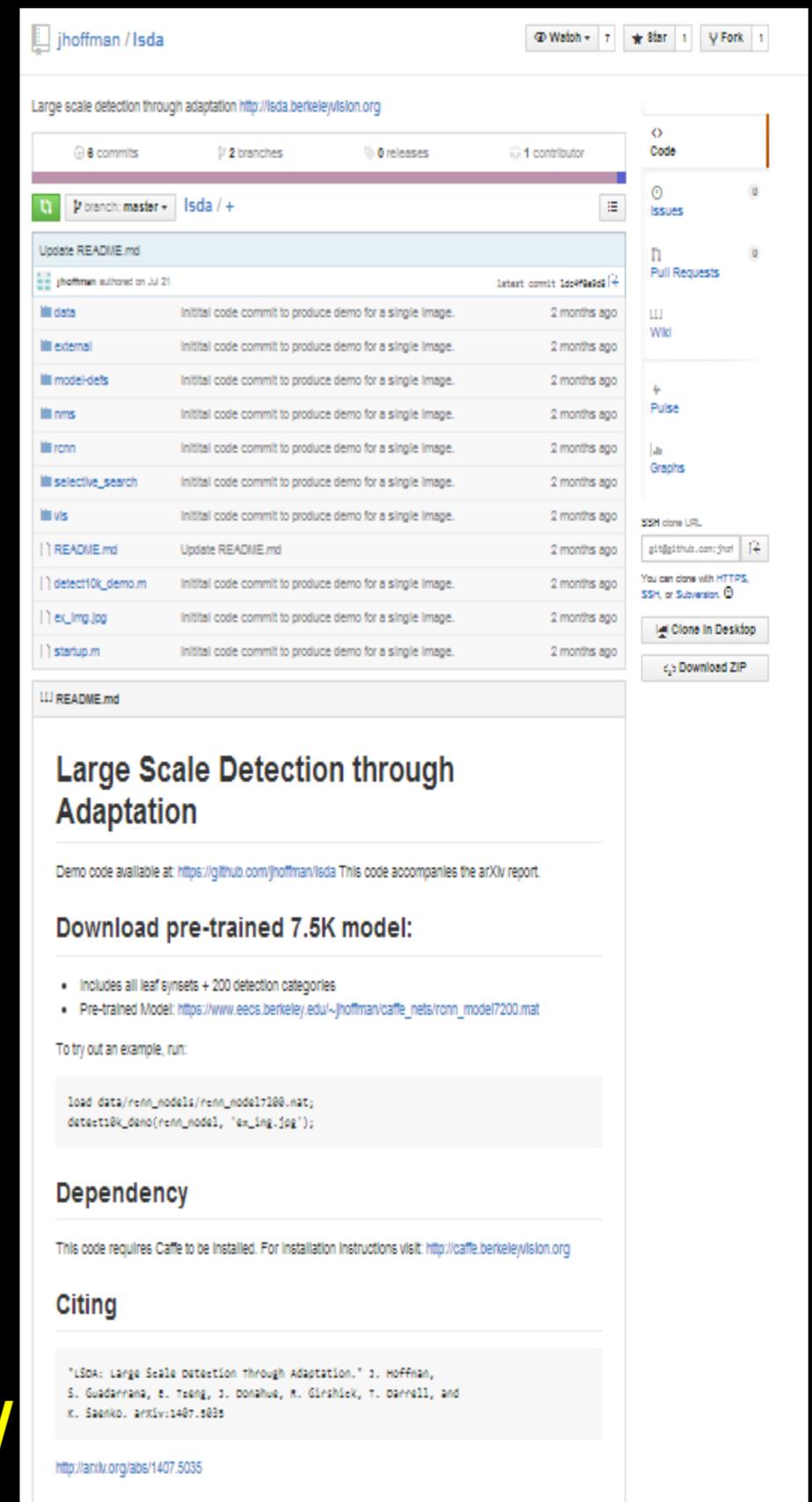
False Positives



7K class detector!

- Public release of a 7604 category detector trained using this method
- 200 ILSVRC2013 classes trained with bounding box data
- 7404 ImageNet leaf nodes trained with adaptation

<http://lsda.berkeleyvision.org/>



The screenshot shows the GitHub repository page for 'lsda' by jhoffman. The repository is titled 'Large scale detection through adaptation' and is linked to the website <http://lsda.berkeleyvision.org>. The repository has 8 commits, 2 branches, 0 releases, and 1 contributor. The main branch is 'master'. The repository contains several files and folders, including 'data', 'external', 'model-dets', 'nms', 'ronn', 'selective_search', 'vis', 'README.md', 'detect10k_demo.m', 'ex_img.jpg', and 'startup.m'. The README.md file is expanded, showing the title 'Large Scale Detection through Adaptation' and a description: 'Demo code available at: <https://github.com/jhoffman/lsda> This code accompanies the arXiv report.' The README also includes a section for 'Download pre-trained 7.5K model:' which lists the following details:

- Includes all leaf synsets + 200 detection categories
- Pre-trained Model: https://www.eecs.berkeley.edu/~jhoffman/caffe_nets/ronn_model7200.mat

To try out an example, run:

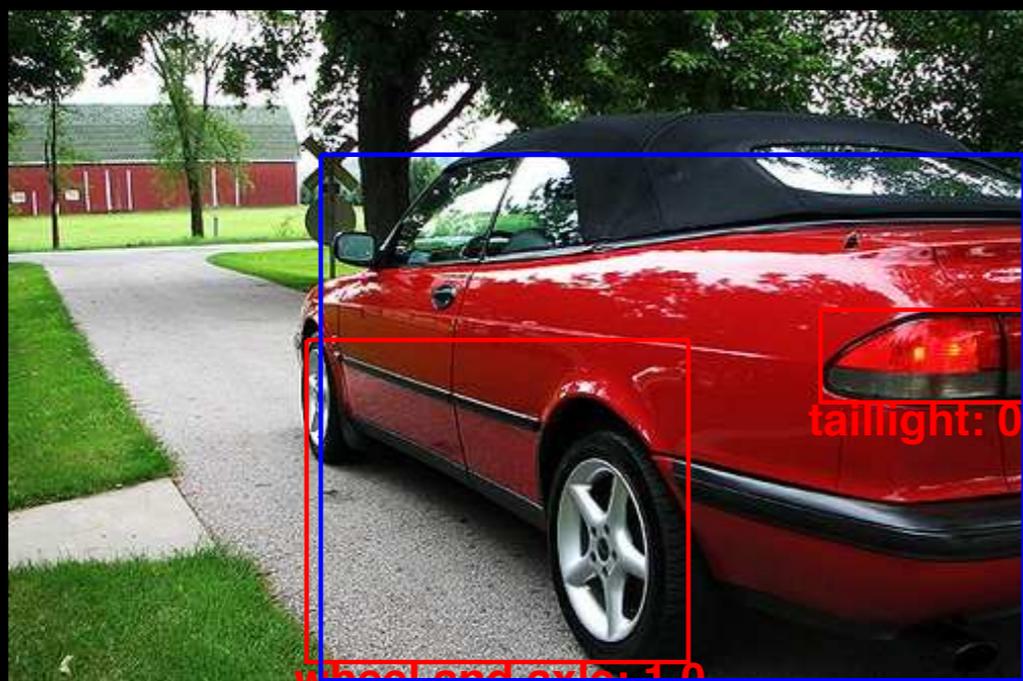
```
load data/ronn_models/ronn_model7200.mat;
detect10k_demo(ronn_model, 'ex_img.jpg');
```

The README also includes a 'Dependency' section stating that Caffe must be installed, with a link to the Caffe installation instructions. Finally, there is a 'Citing' section with the following citation:

"LSDA: Large Scale Detection through Adaptation." J. Hoffman, S. Guadarrama, E. Tseng, J. Donahue, K. Girshick, T. Darrell, and K. Saenko. arXiv:1407.5035

<http://arxiv.org/abs/1407.5035>

Detection with 7K Classes



car: 6.0
wheel and axle: 1.0

taillight: 0.9



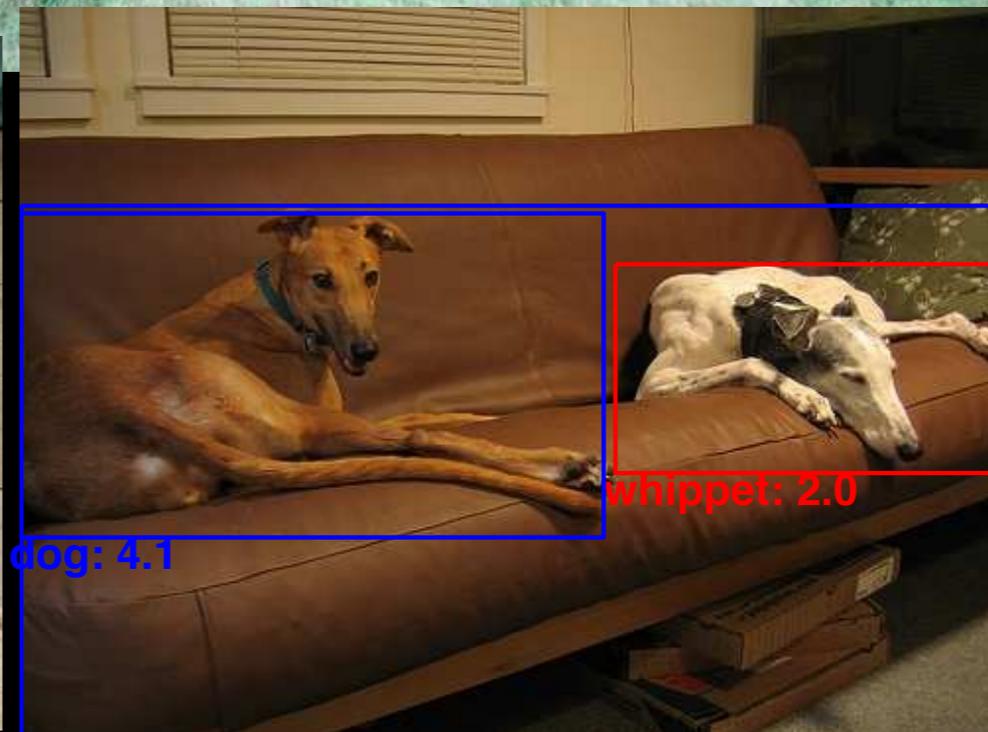
American bison: 7.0



sunglasses: 5.7

person: 3.0

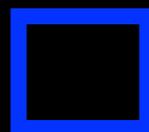
person: 3.1



dog: 4.1

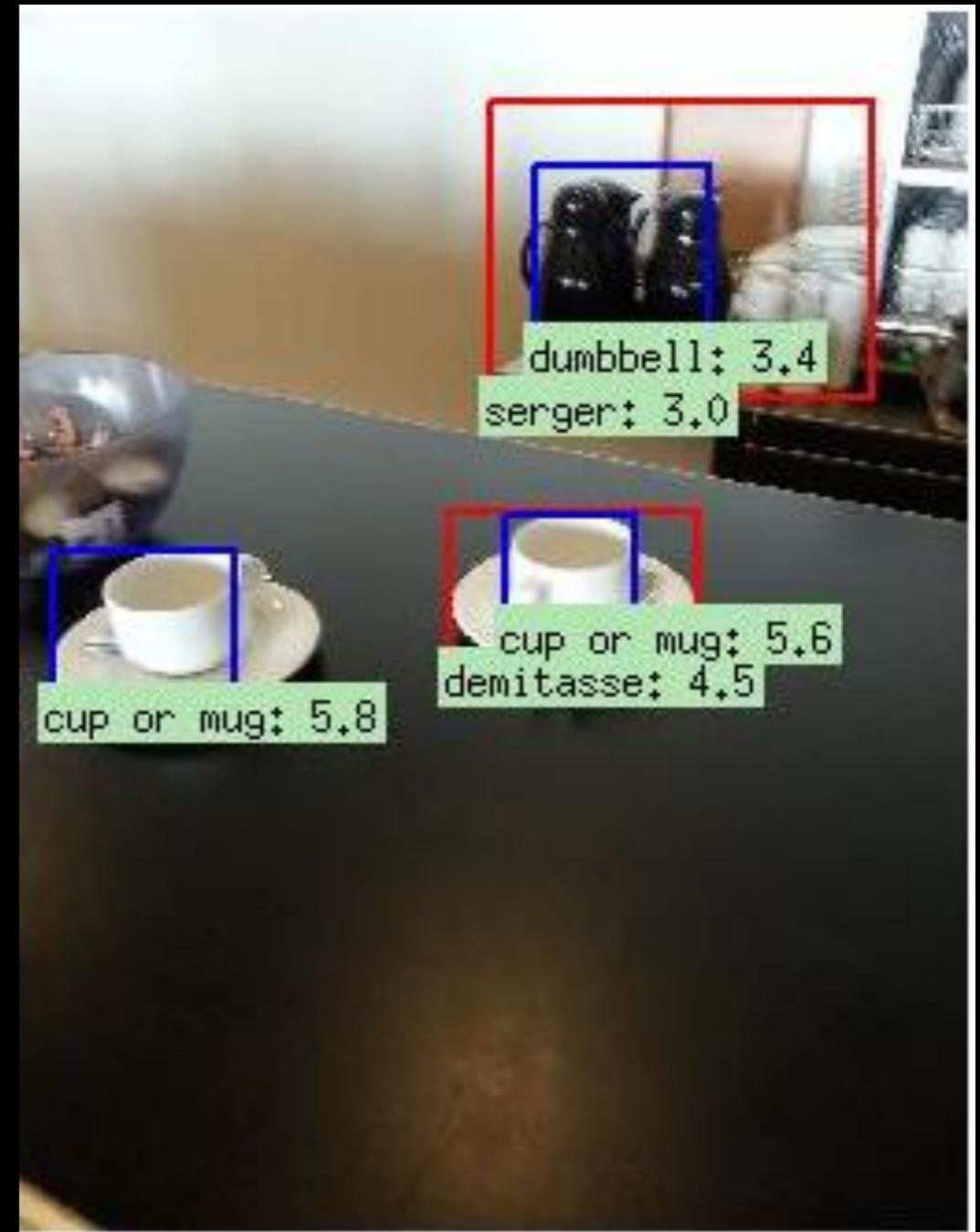
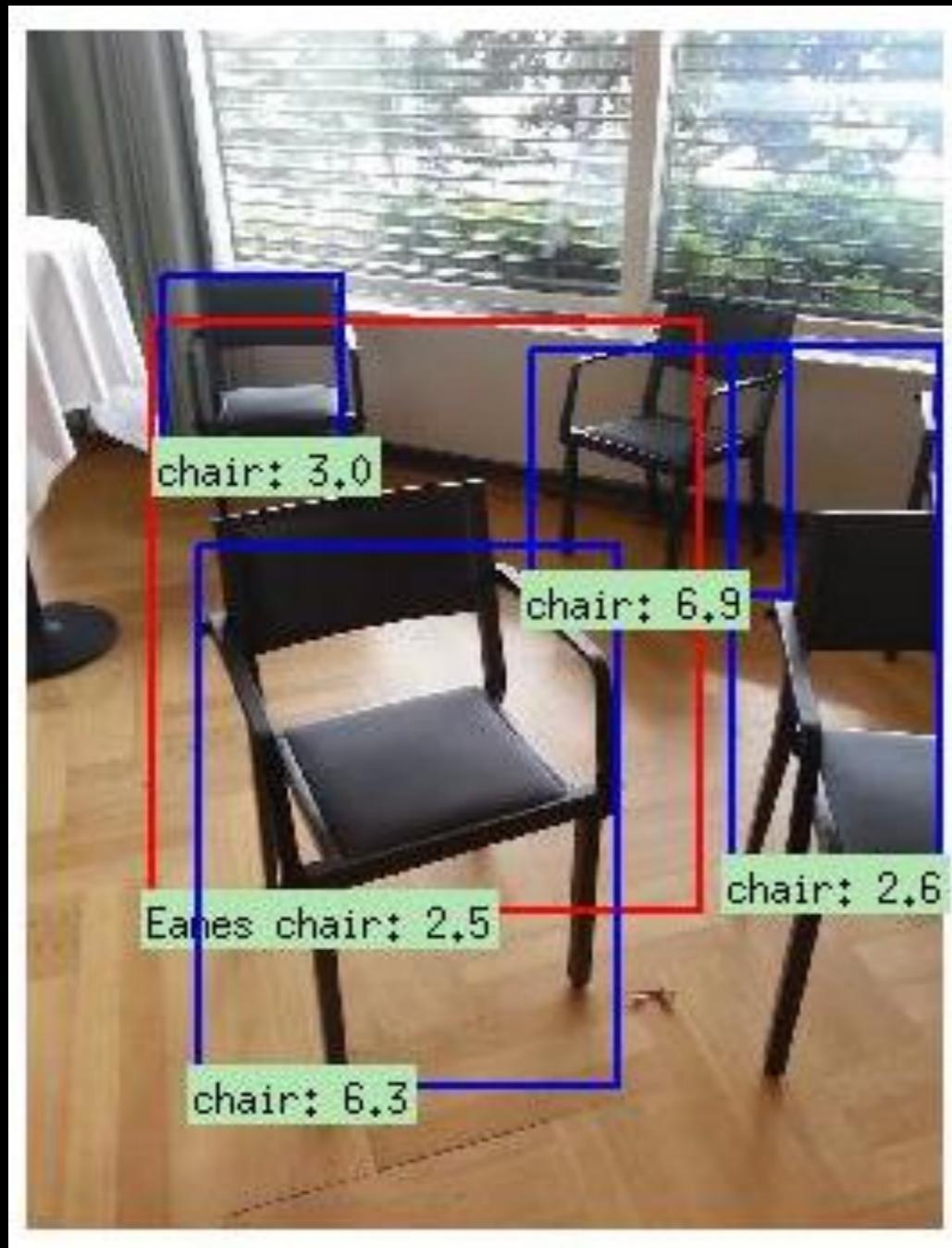
whippet: 2.0

sofa: 8.0

 200 trained with bbox

 7K trained without bbox

ECCV 2014 Demo...



ECCV 2014 Demo



Funny wagon

An ambulance used to transport patients to a mental hospital

240 pictures

53.64% Popularity Percentile

Wordnet IDs

- armored vehicle, carrier (0)
- forklift (0)
- locomotive, engine
- motor vehicle, auto
 - amphibian, amphibious
 - bloodmobile (0)
 - car, auto, automobile
 - ambulance
 - funny wagon**
 - beach wagon
 - bus, jalopy
 - cab, hack, taxi
 - compact, compact car
 - convertible
 - coupe (0)
 - cruiser, police
 - electric, electric car
 - gas guzzler
 - hardtop (0)
 - hatchback
 - horseless carriage
 - hot rod, hot rod car
 - jeep, landrover
 - limousine, limo
 - loaner (0)

Treemap Visualization

Images of the Synset

Downloads



Conclusions

- Domain adaptation approach to weak-label detector learning
- “Detection” can transfer across categories
- 7.4K SPP R-CNN Detector runs $\sim < 1$ fps
- Model and code available at lsda.berkeleyvision.org
- Future work: context, hierarchical backoff

Conclusion

- Domain adaptation approach to weak-label detector learning
- 7.4K SPP R-CNN
Detector runs $\sim < 1$ fps
- Model and code available at lsda.berkeleyvision.org
- Future work: context, hierarchical backoff

