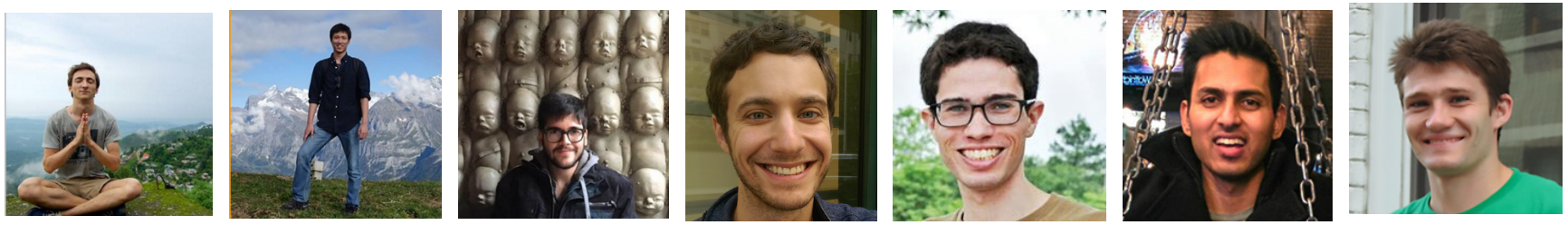
FAIR COCO Object Detection

Sergey Zagoruyko*, Tsung-Yi Lin*,
Pedro Pinheiro*, Adam Lerer, Sam
Gross, Soumith Chintala, Piotr Dollár



(*equal contribution)

Results

	AP bbox	AP small	AP medium	AP large	AR max=100	AP segm
MSRA	0.373	0.183	0.419	0.524	0.491	0.282
FAIRCNN	0.335	0.139	0.378	0.477	0.485	0.251
ION	0.310	0.123	0.332	0.447	0.457	
FastRCNN	0.197	0.035	0.188	0.346	0.298	

66% improvement over FastRCNN baseline

Overview

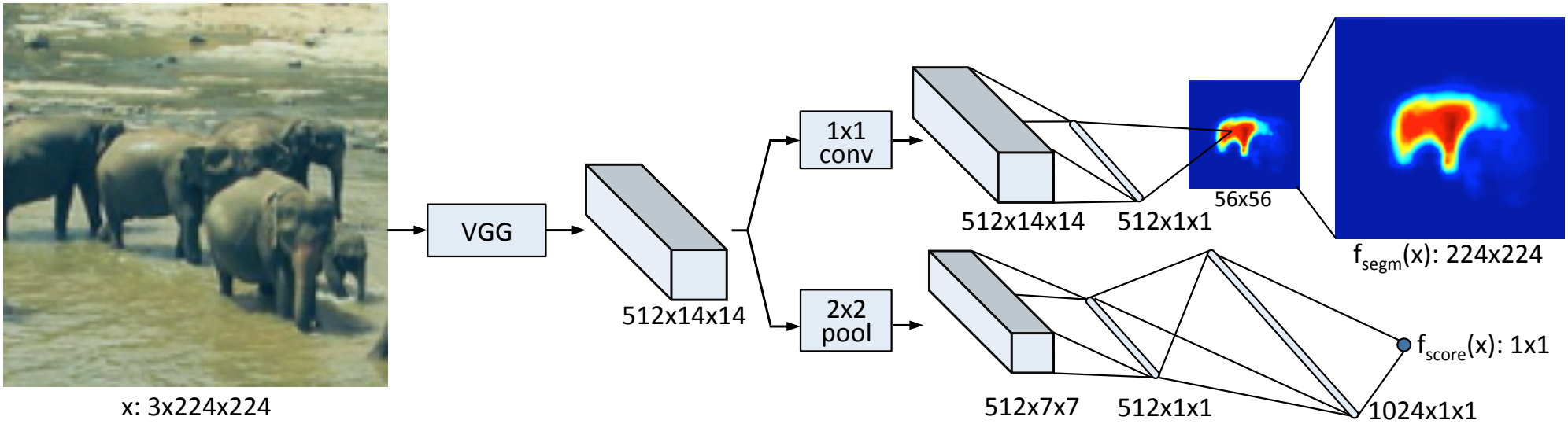
I. DeepMask segmentation proposals [Pinheiro NIPS 15]
   + iterative localization
   + top-down refinement

II. Fast R-CNN object detector [Girshick ICCV 15]
   + foveal context regions
   + modified loss function
   + skip connections
   + ensembling

I. DEEP MASK OBJECT PROPOSALS

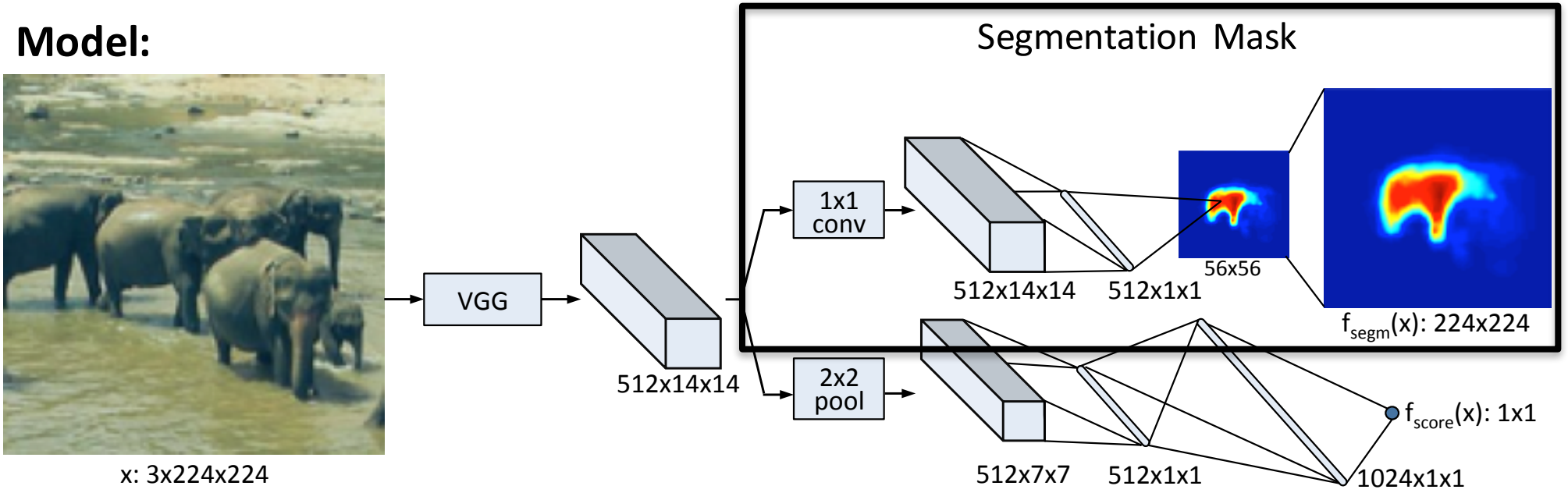
DeepMask Framework

Model:



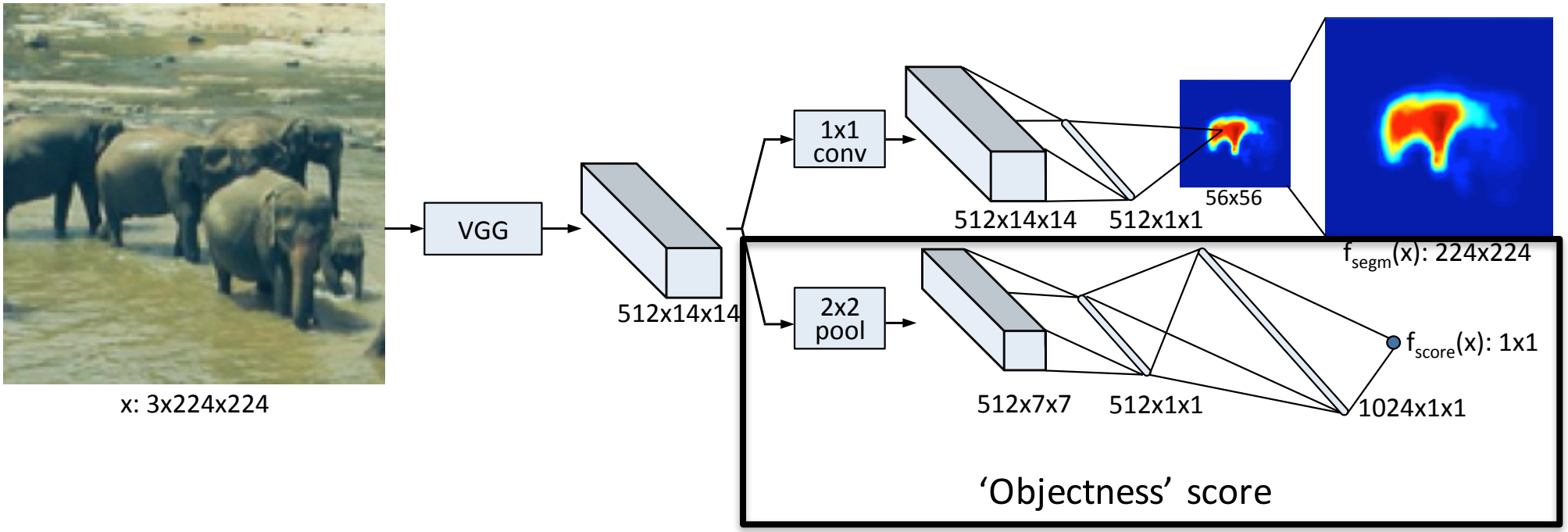
DeepMask Framework

Model:



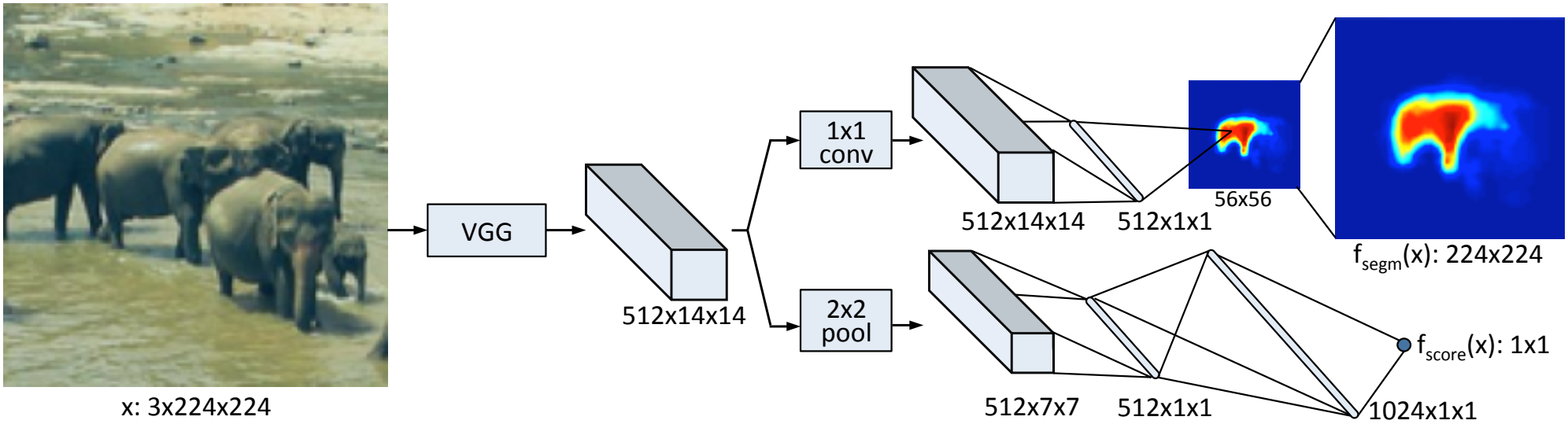
DeepMask Framework

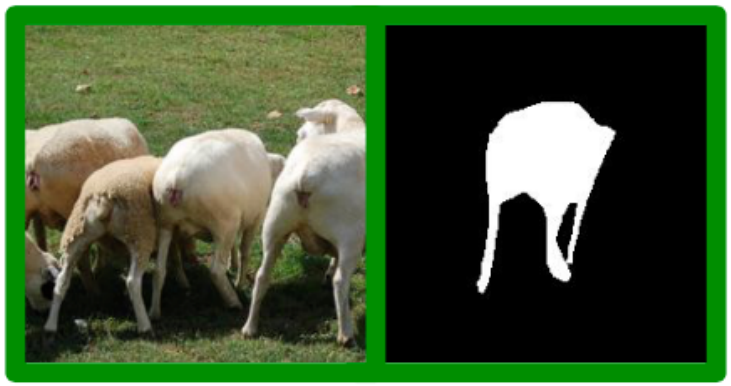
Model:



DeepMask Framework

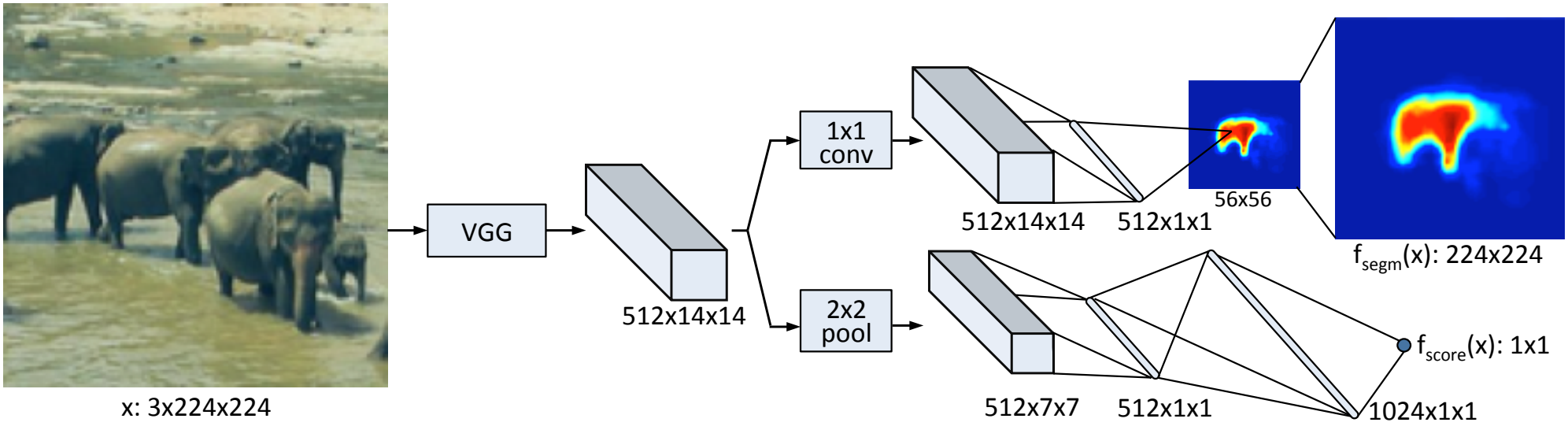
Model:





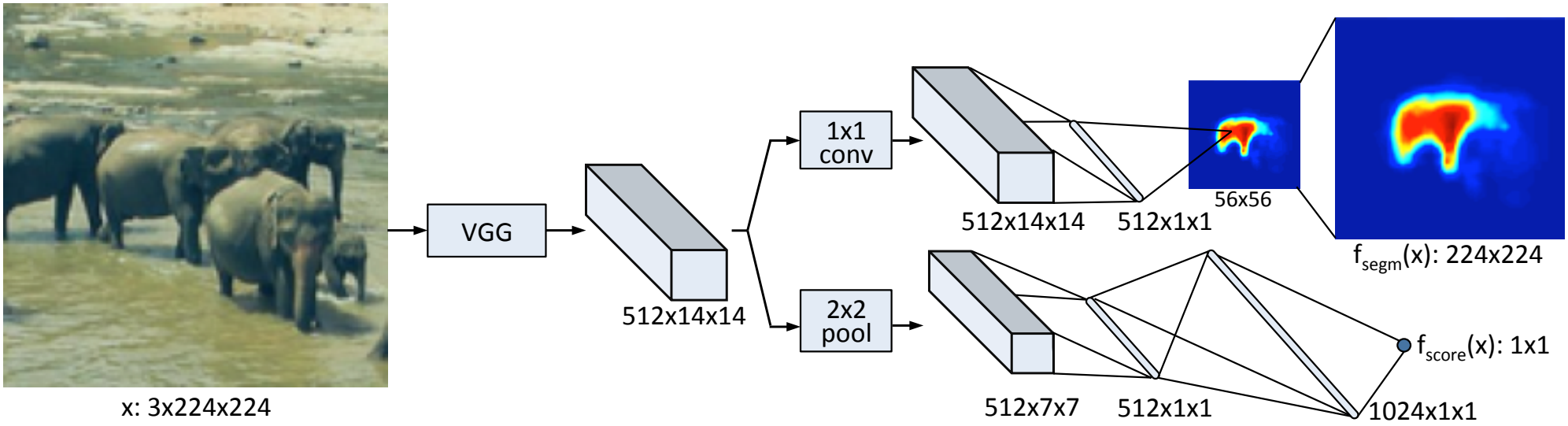
DeepMask Framework

Model:



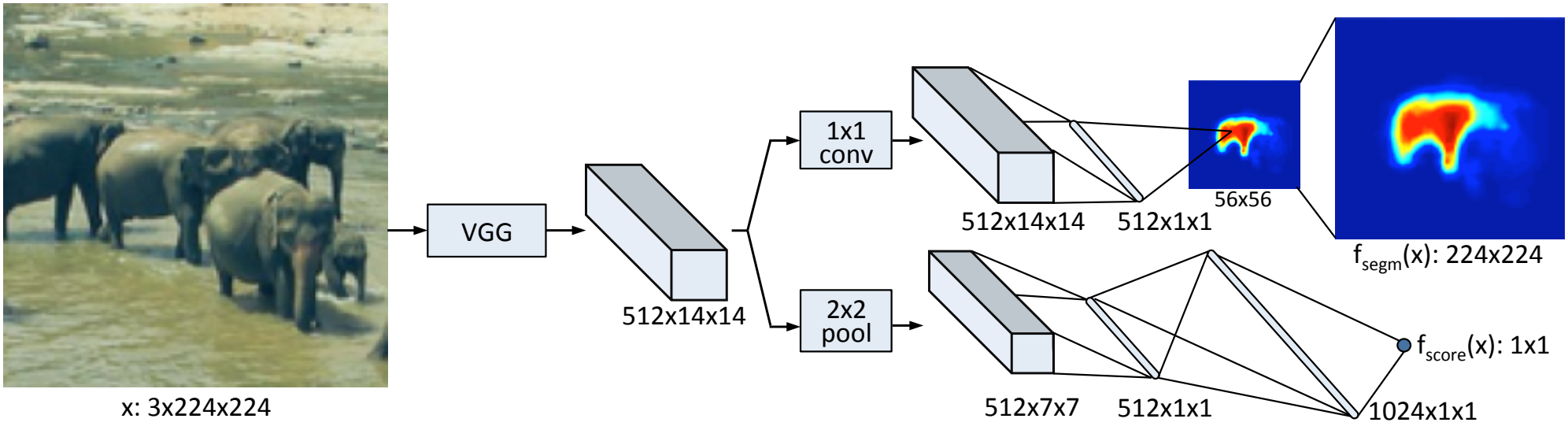
DeepMask Framework

Model:



DeepMask Framework

Model:

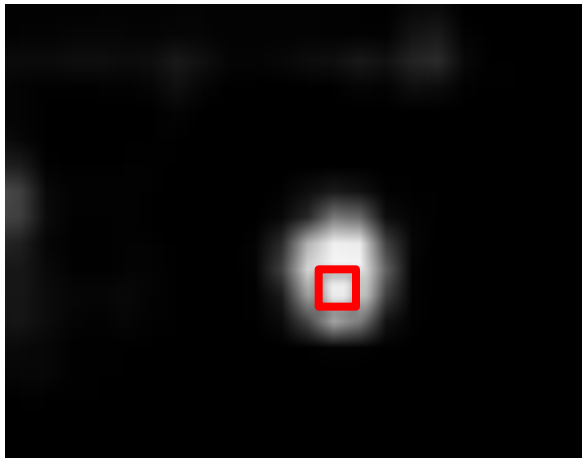


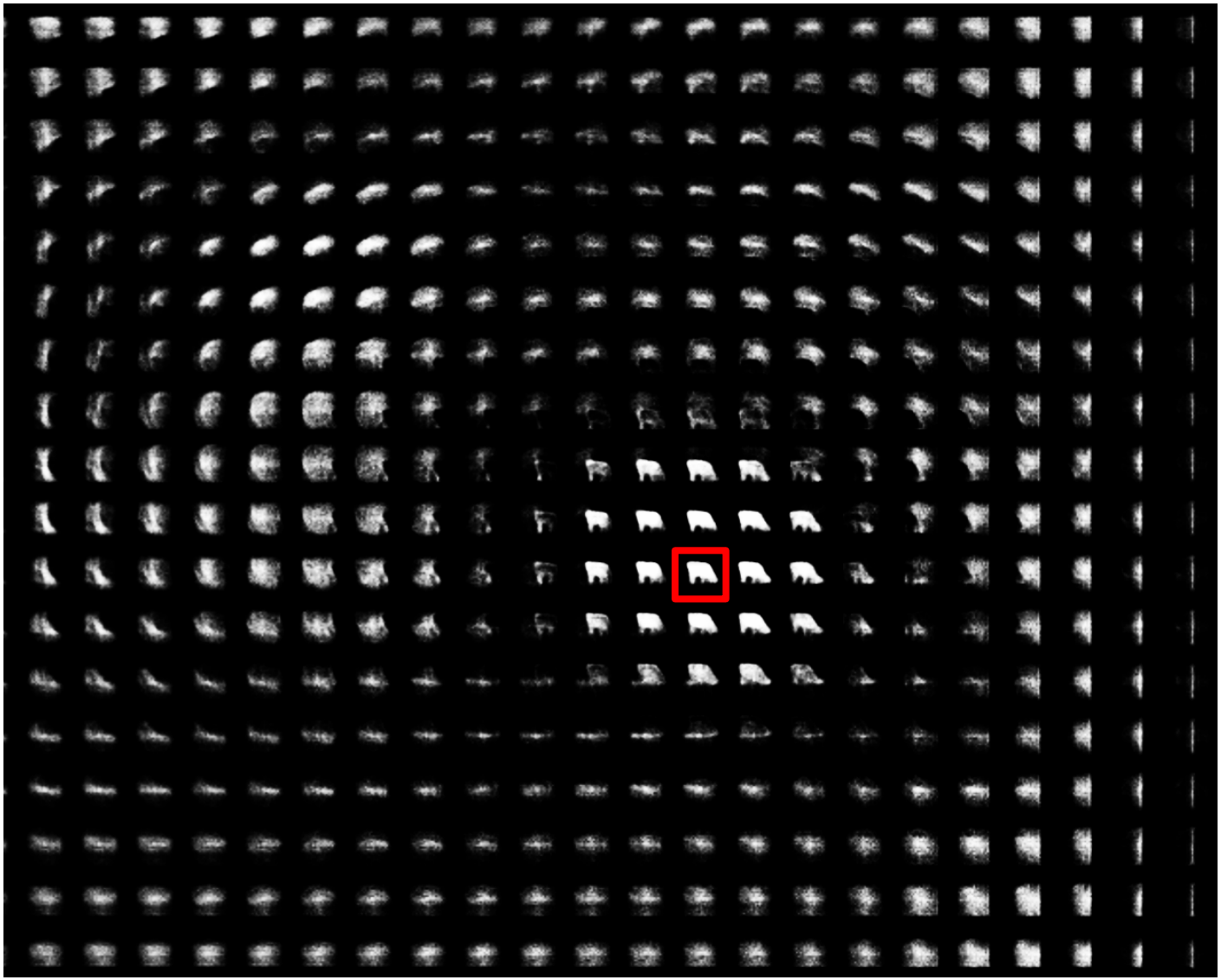
Single Scale Inference

masks

image

scores

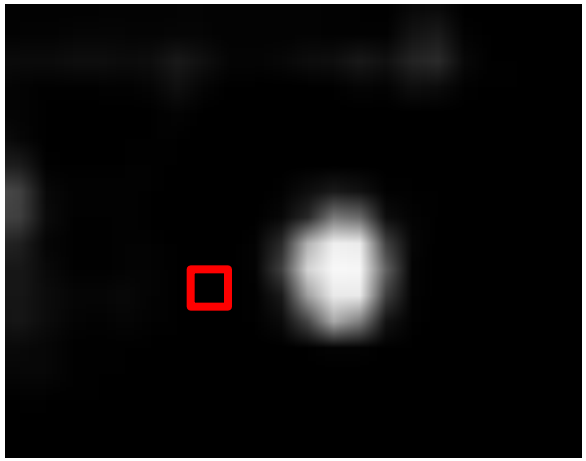




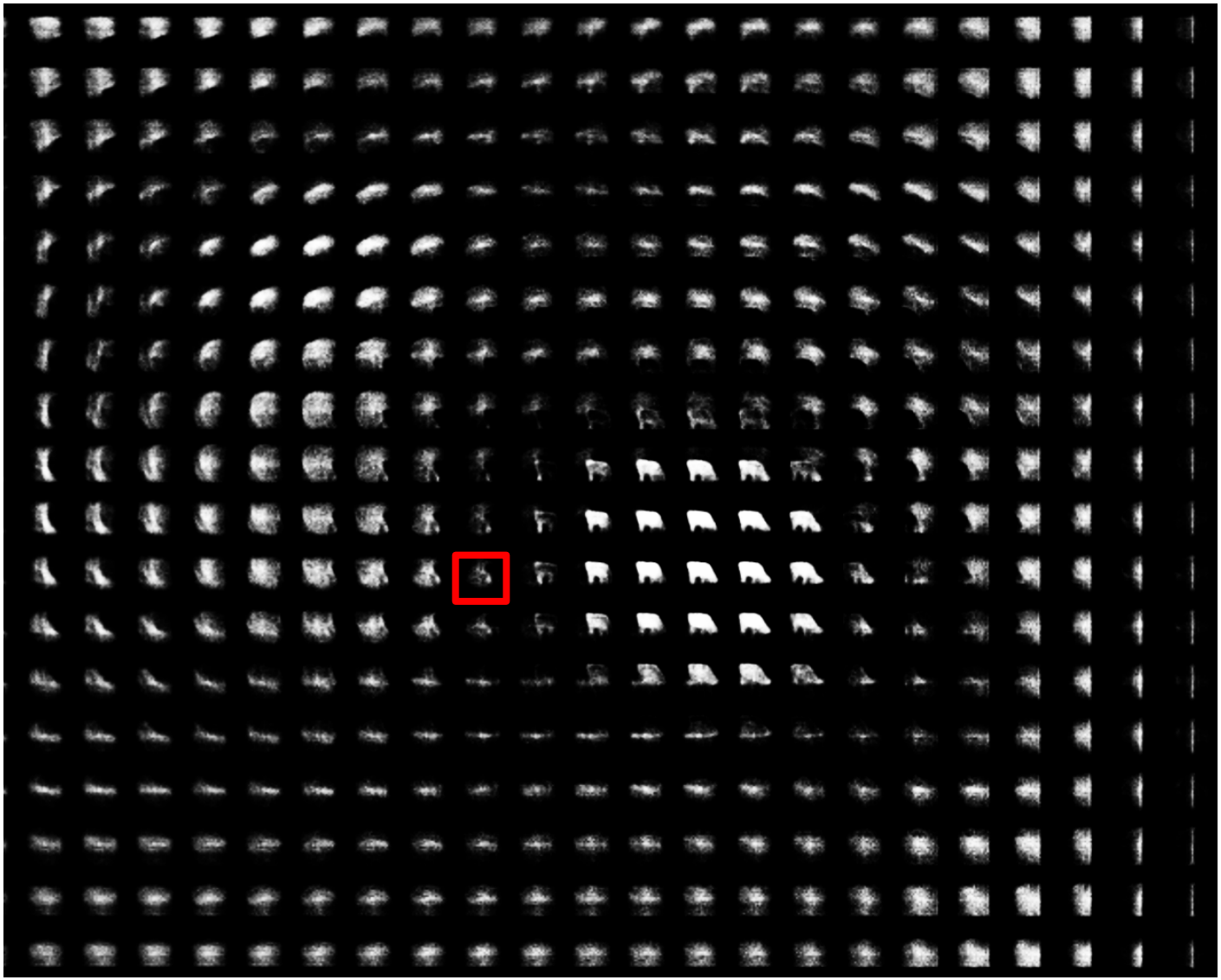
Single Scale Inference

image

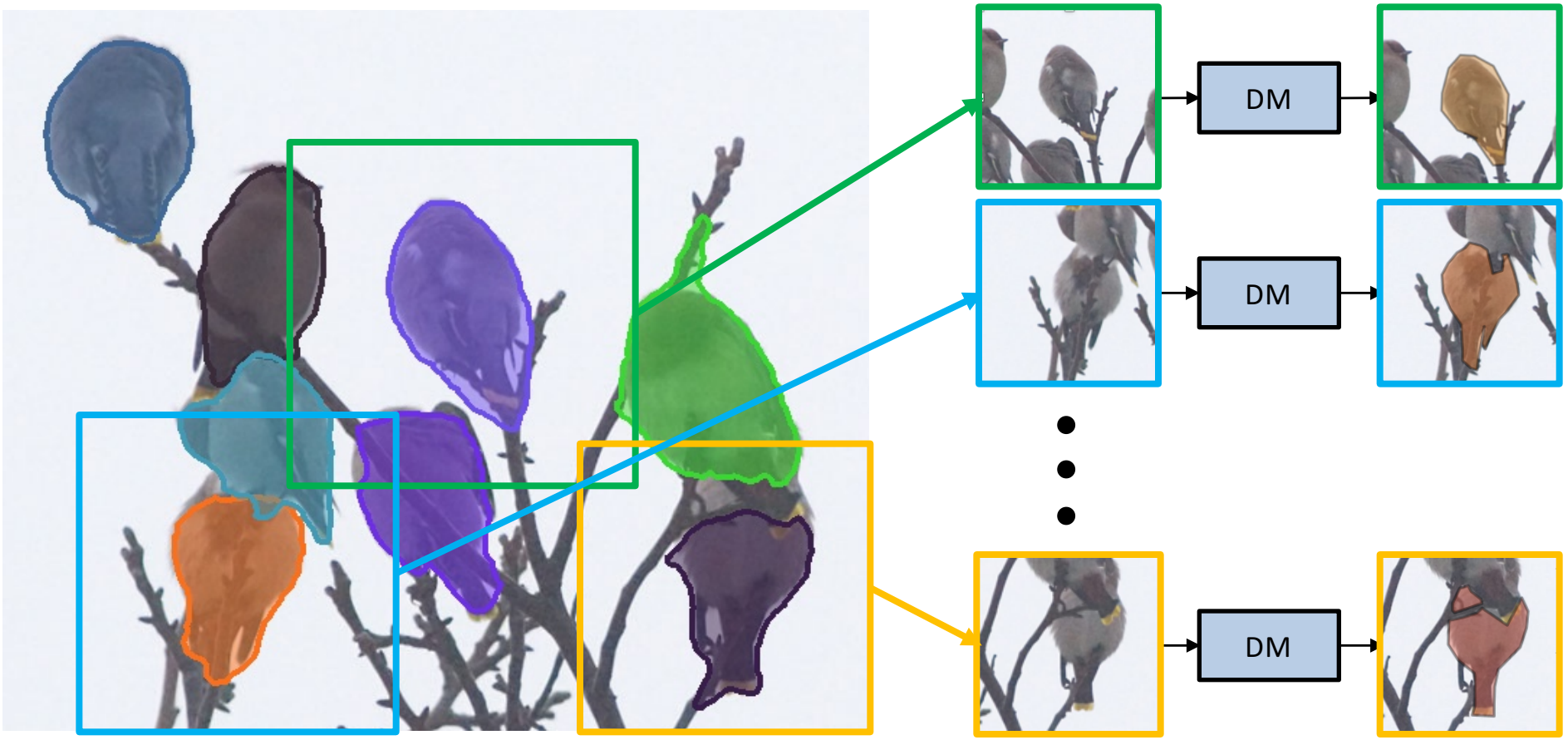
scores



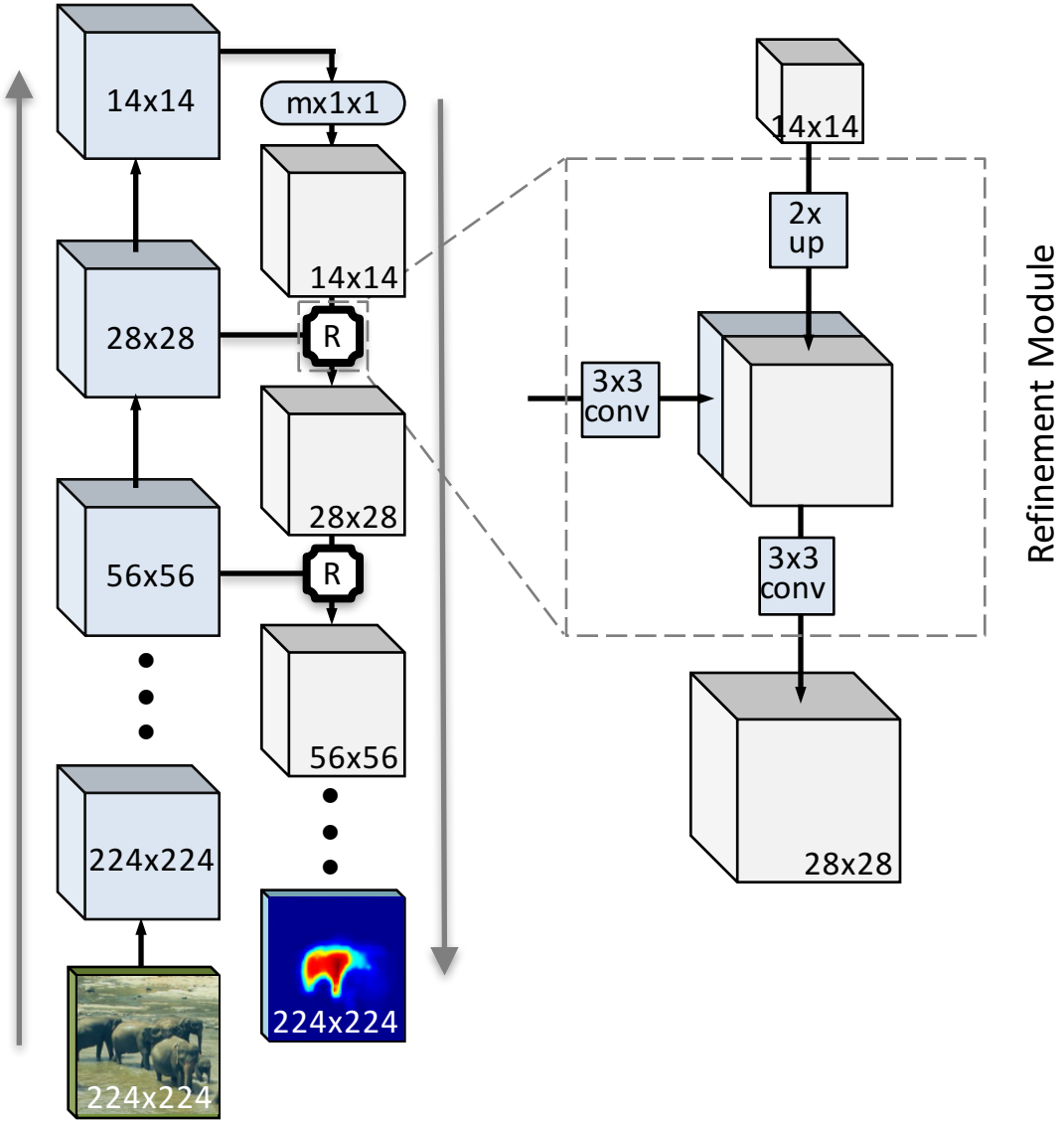
masks



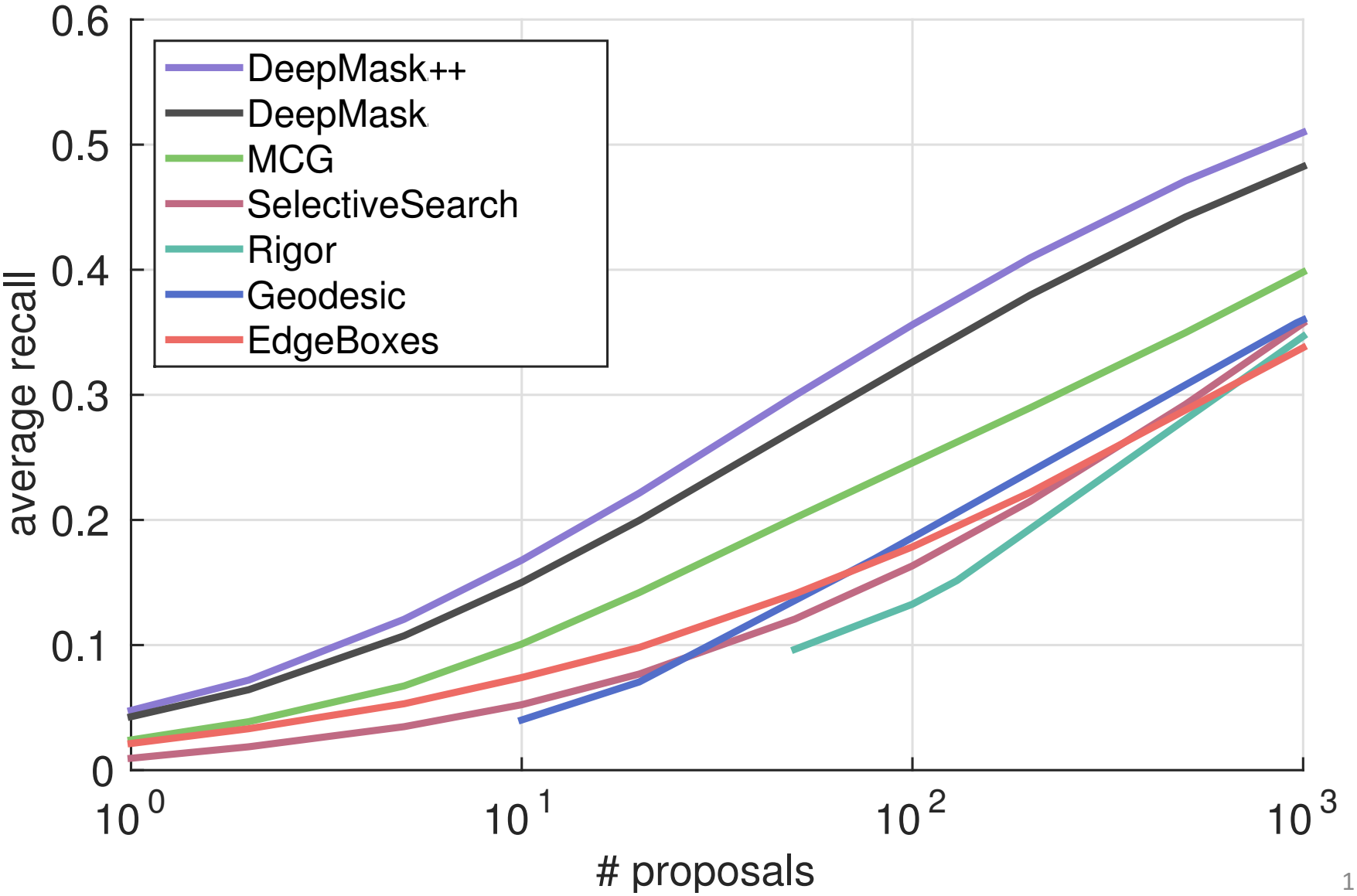
New: Iterative Localization (+1.0 AP)



New: Top-Down Refinement (+0.7 AP)



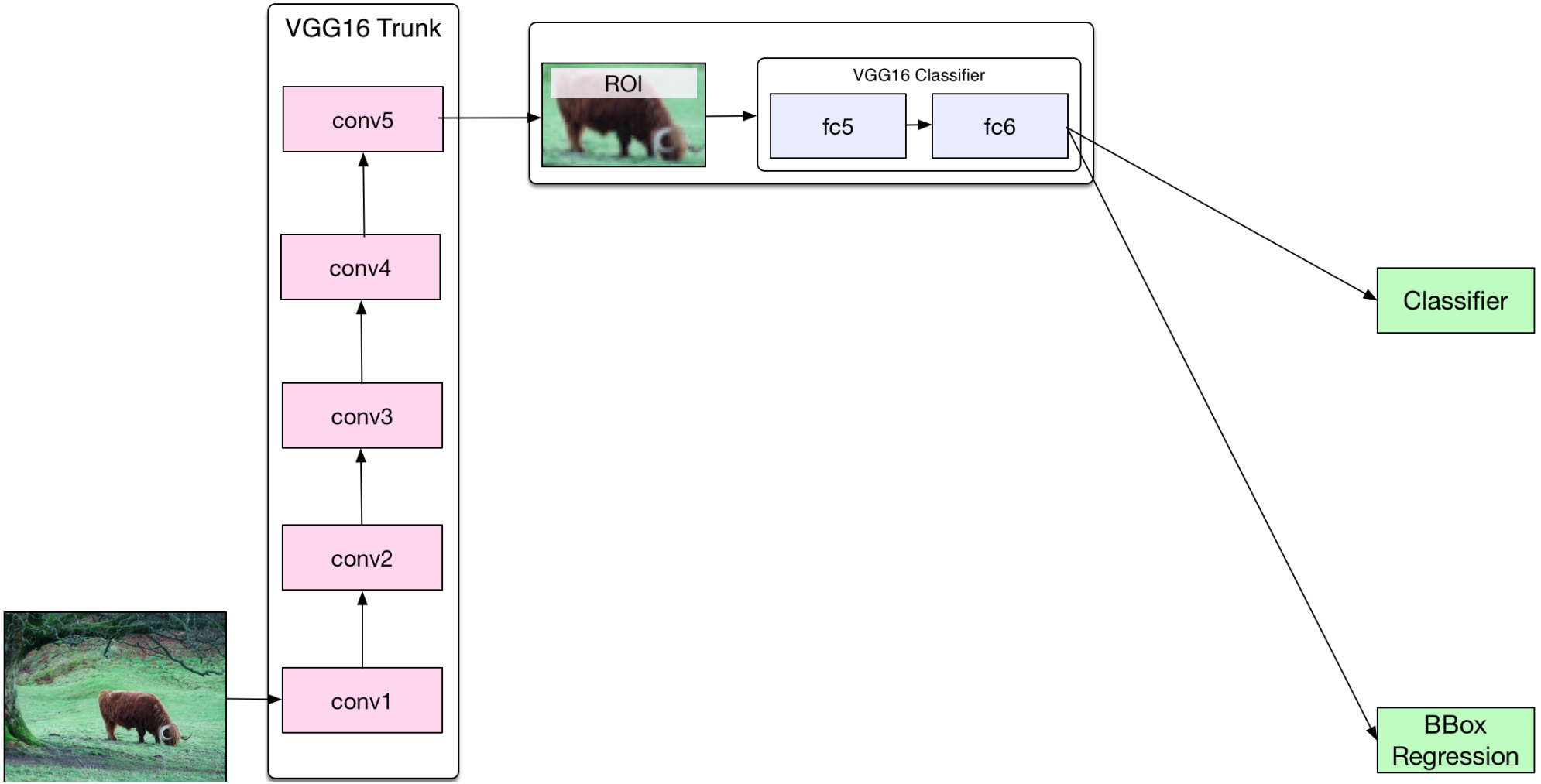
Proposal Quality (boxes)



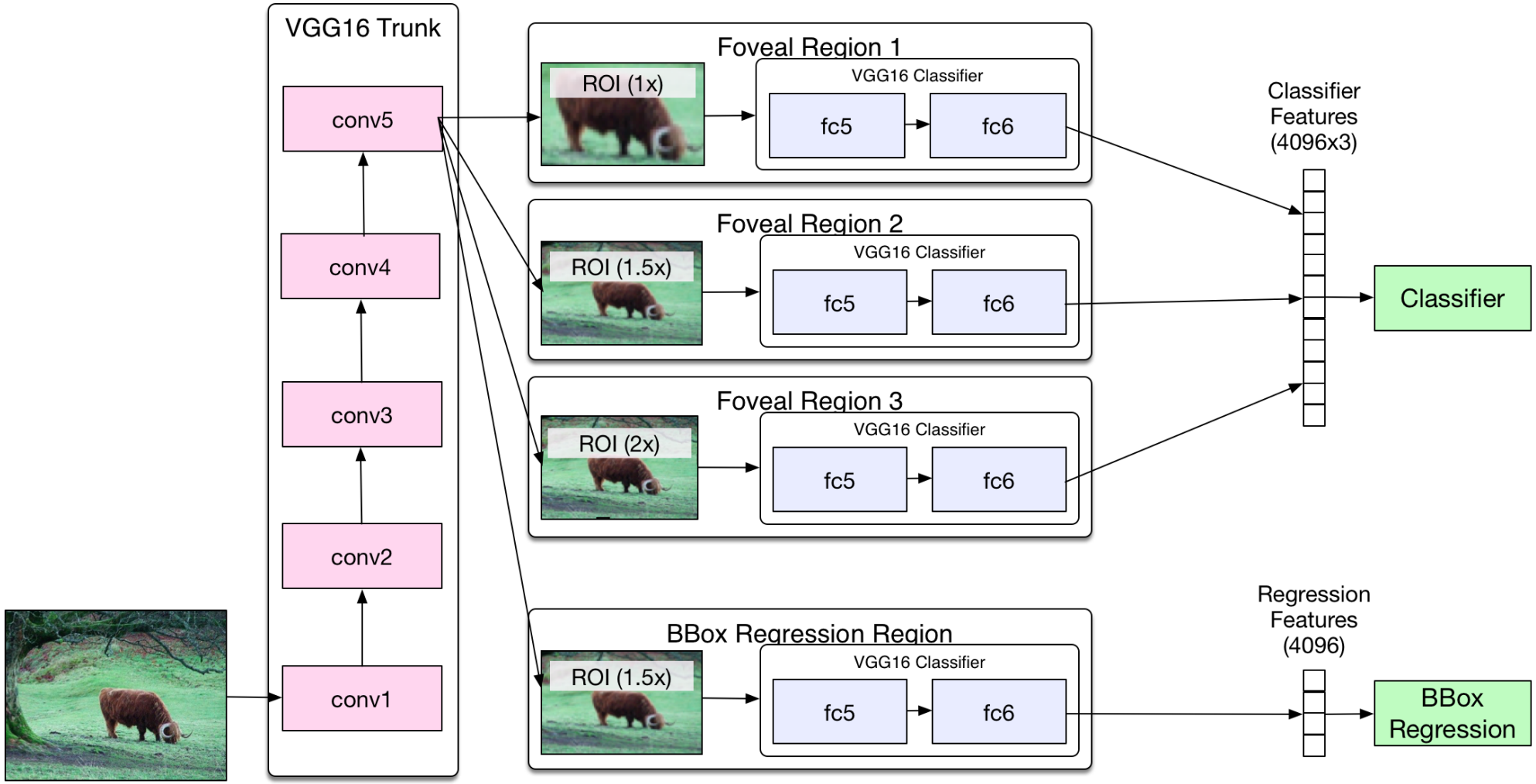
DeepMask Object Proposals

II. CLASSIFICATION FRAMEWORK

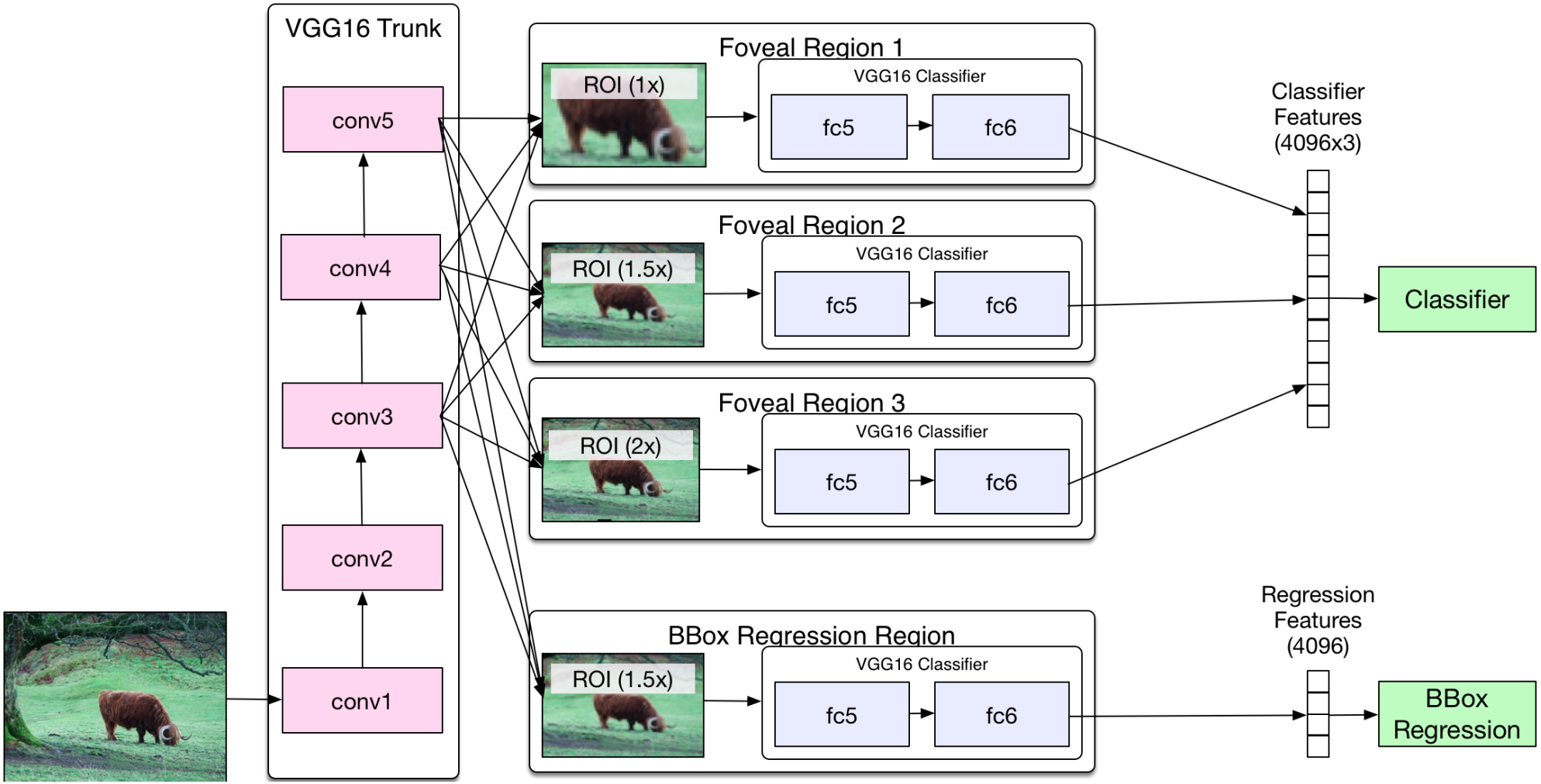
- Fast R-CNN setup [Girshick, ICCV15]



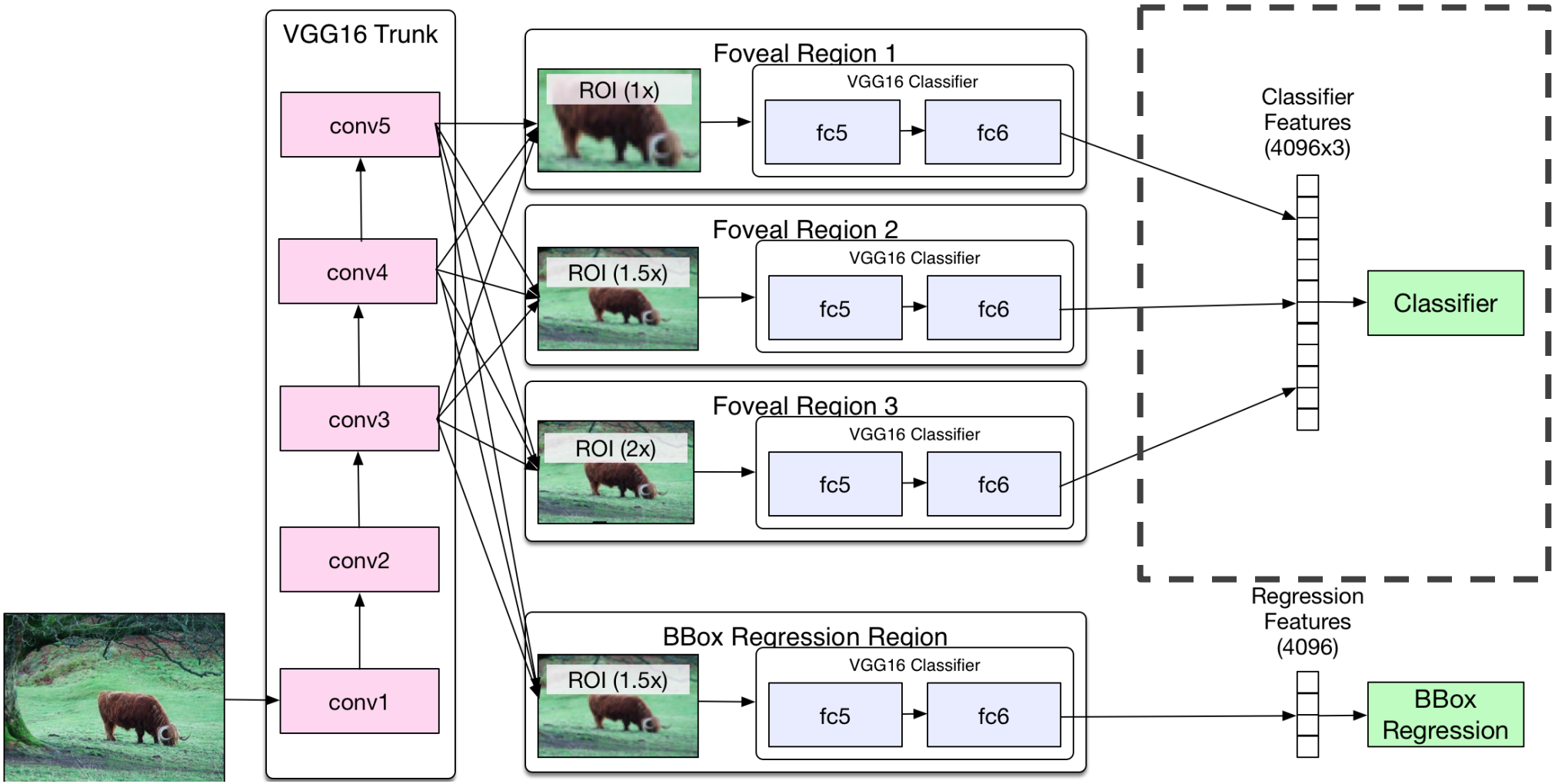
- Fast R-CNN setup [Girshick, ICCV15]
- Foveal structure [inspired by Gidaris & Komodakis, ICCV15] (+2 AP)



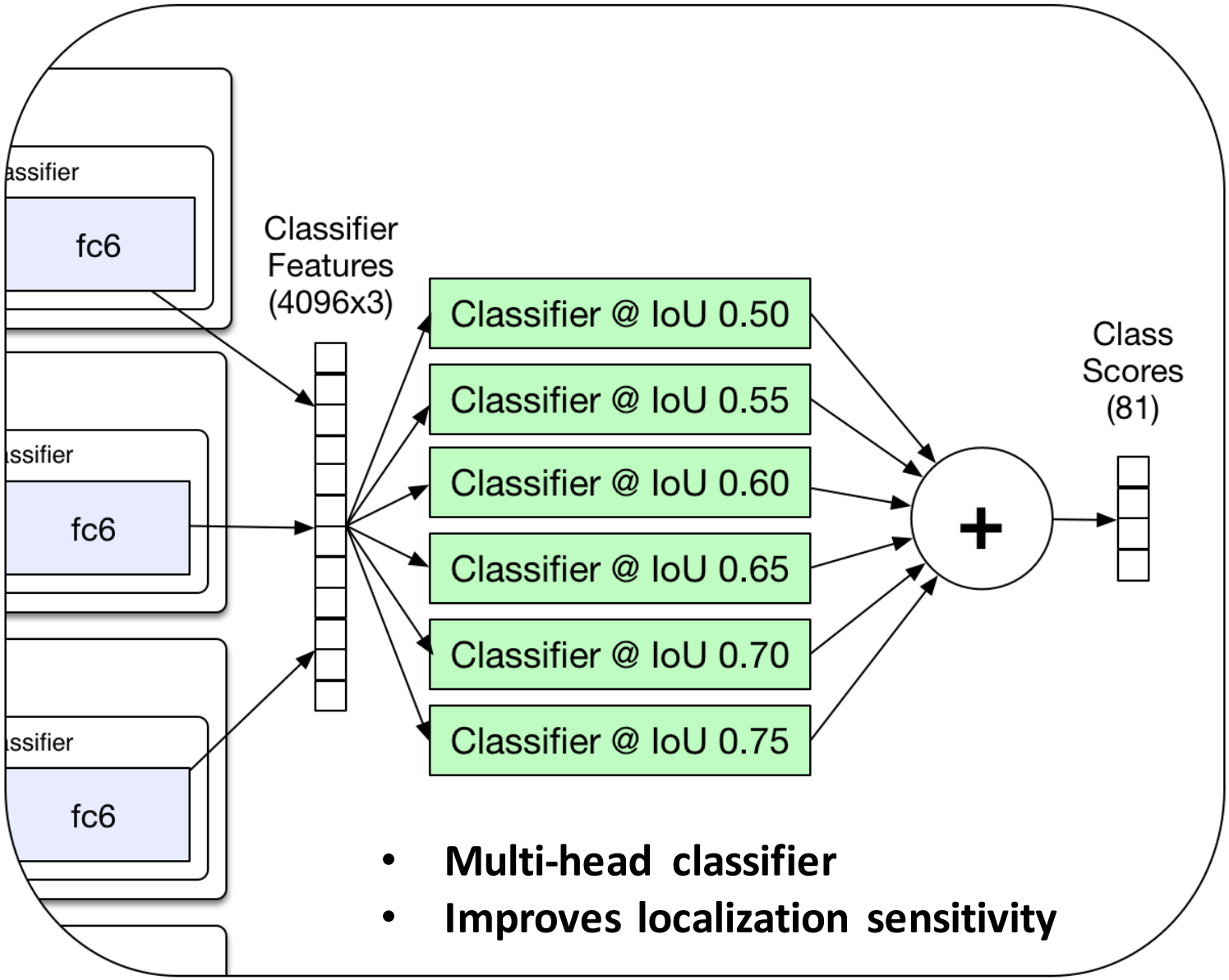
- Fast R-CNN setup [Girshick, ICCV15]
- Foveal structure [inspired by Gidaris & Komodakis, ICCV15] (+2 AP)
- Skip connections (+1 AP)



- Fast R-CNN setup [Girshick, ICCV15]
- Foveal structure [inspired by Gidaris & Komodakis, ICCV15] (+2 AP)
- Skip connections (+1 AP)



Multi-threshold Loss (+1.5 AP)



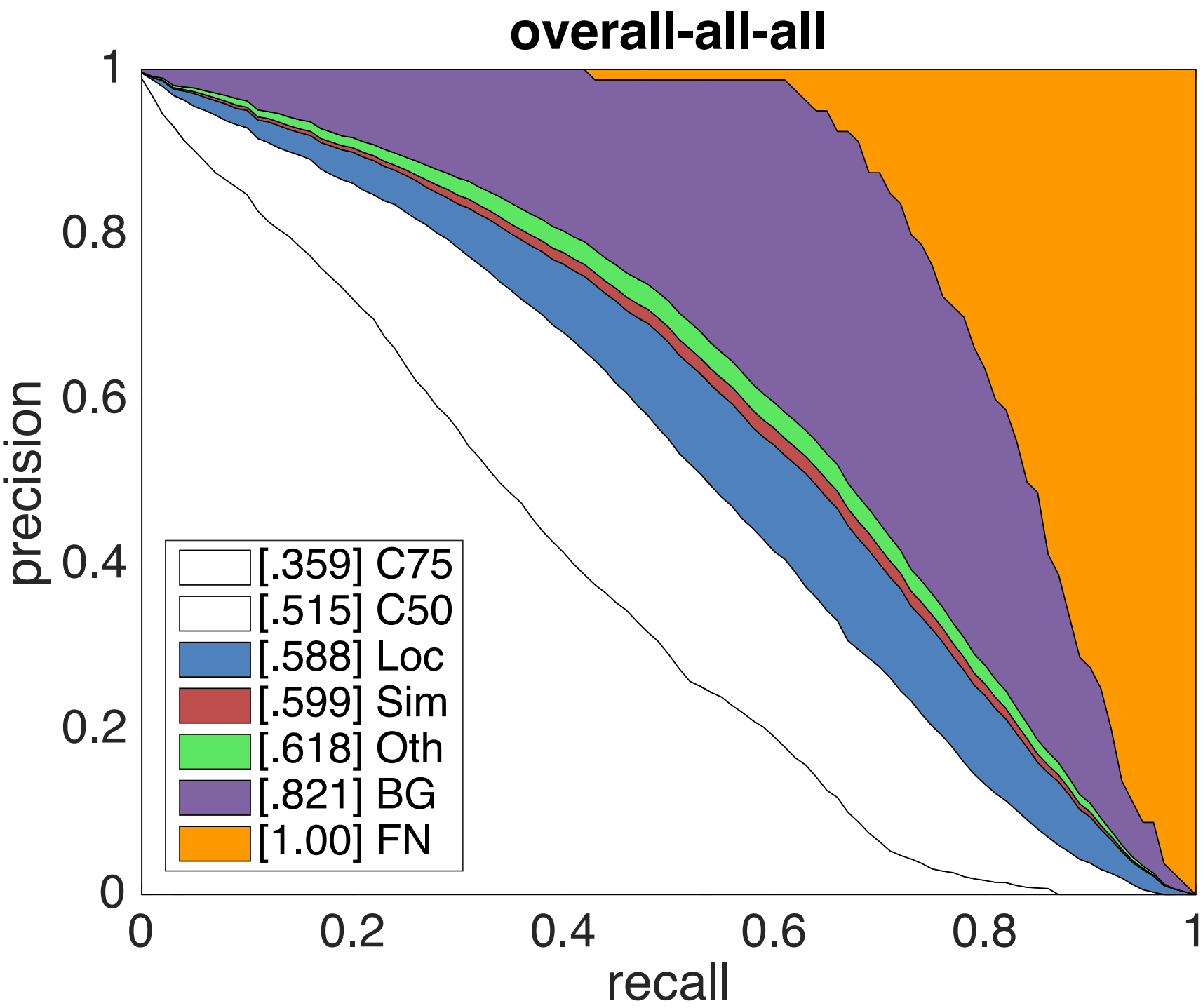
Training

Two strategies:

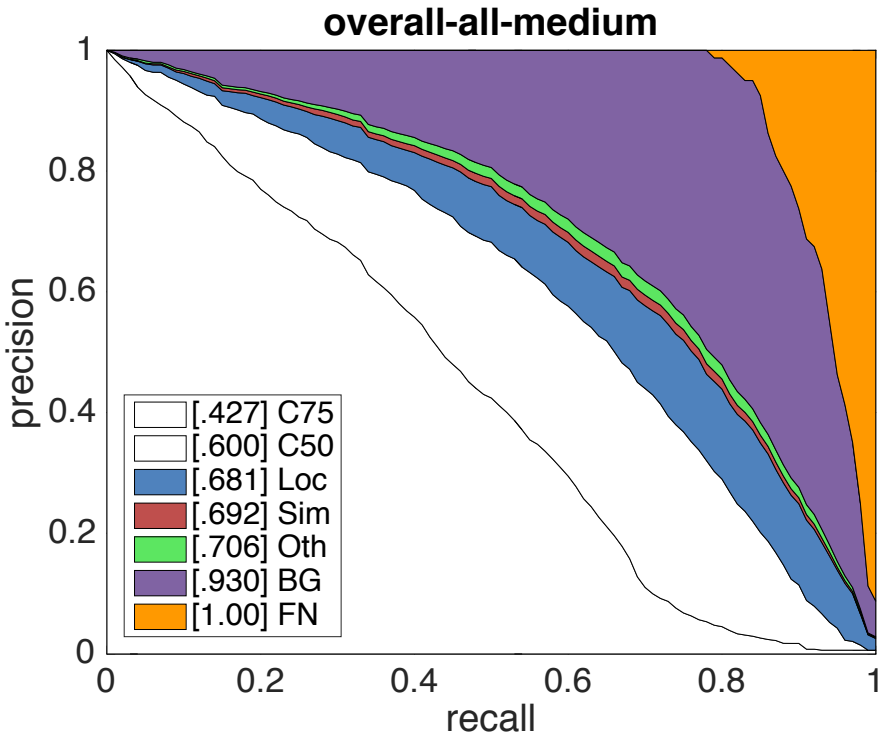
- 4 days on 4 Maxwell GPUs [Big Sur]
- 2.5 days on 8x4 Kepler with Elastic Averaging SGD [Zhang, NIPS 15]

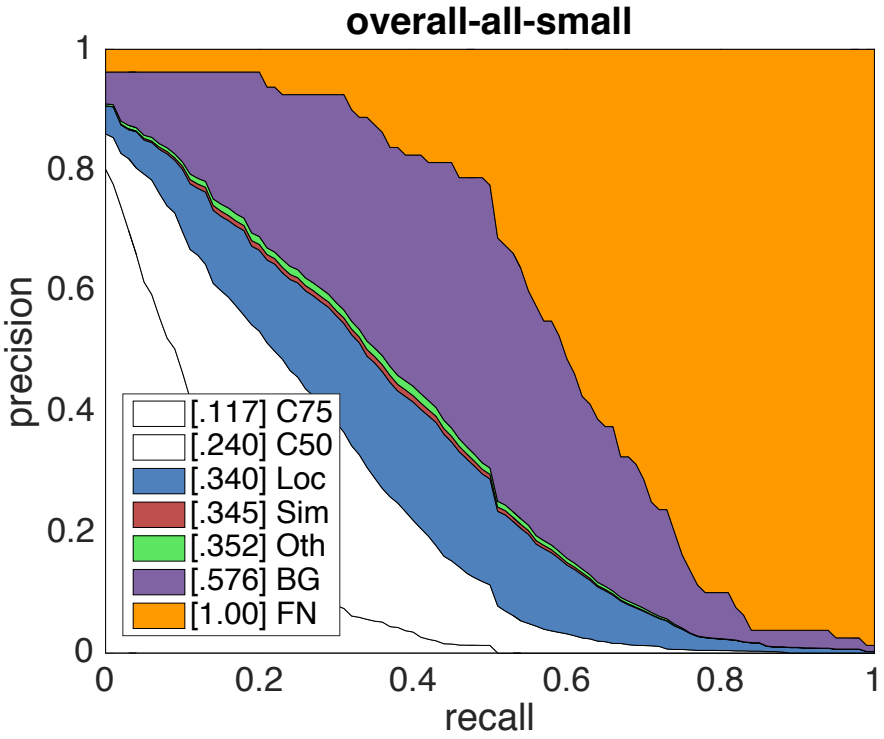
Inference

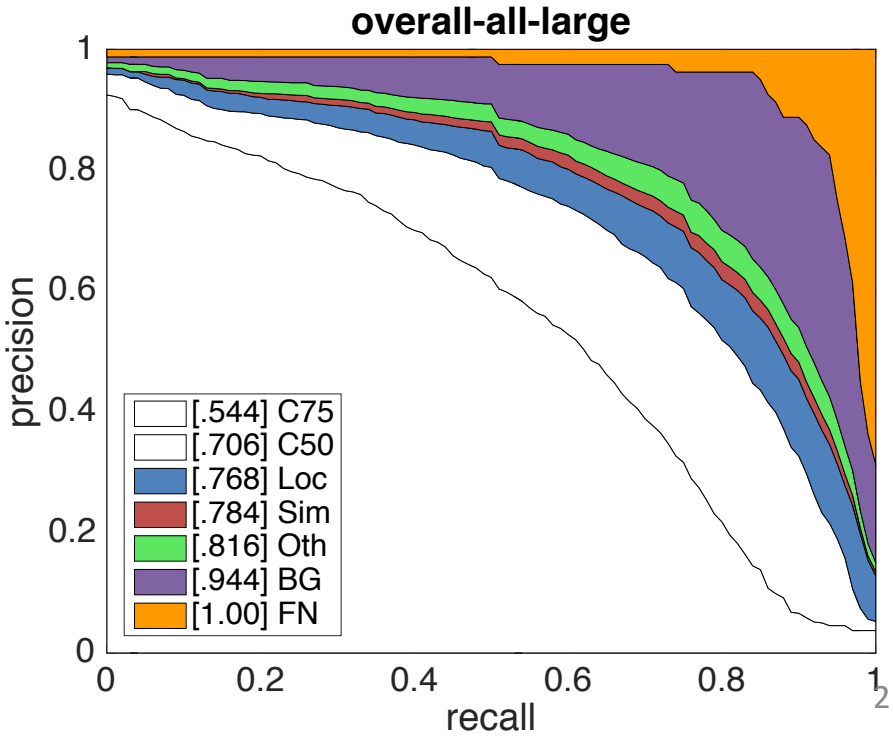
Base Model	**30.1 AP**
+ horizontal flip	**31.1 AP**
+ ROI Pooling ‘2 crop’	**32.1 AP**
+ 7-model Ensemble	**33.5 AP**


overall-all-all
precision
recall
0
0.2
0.4
0.6
0.8
1
[.359] C75
[.515] C50
[.588] Loc
[.599] Sim
[.618] Oth
[.821] BG
[1.00] FN

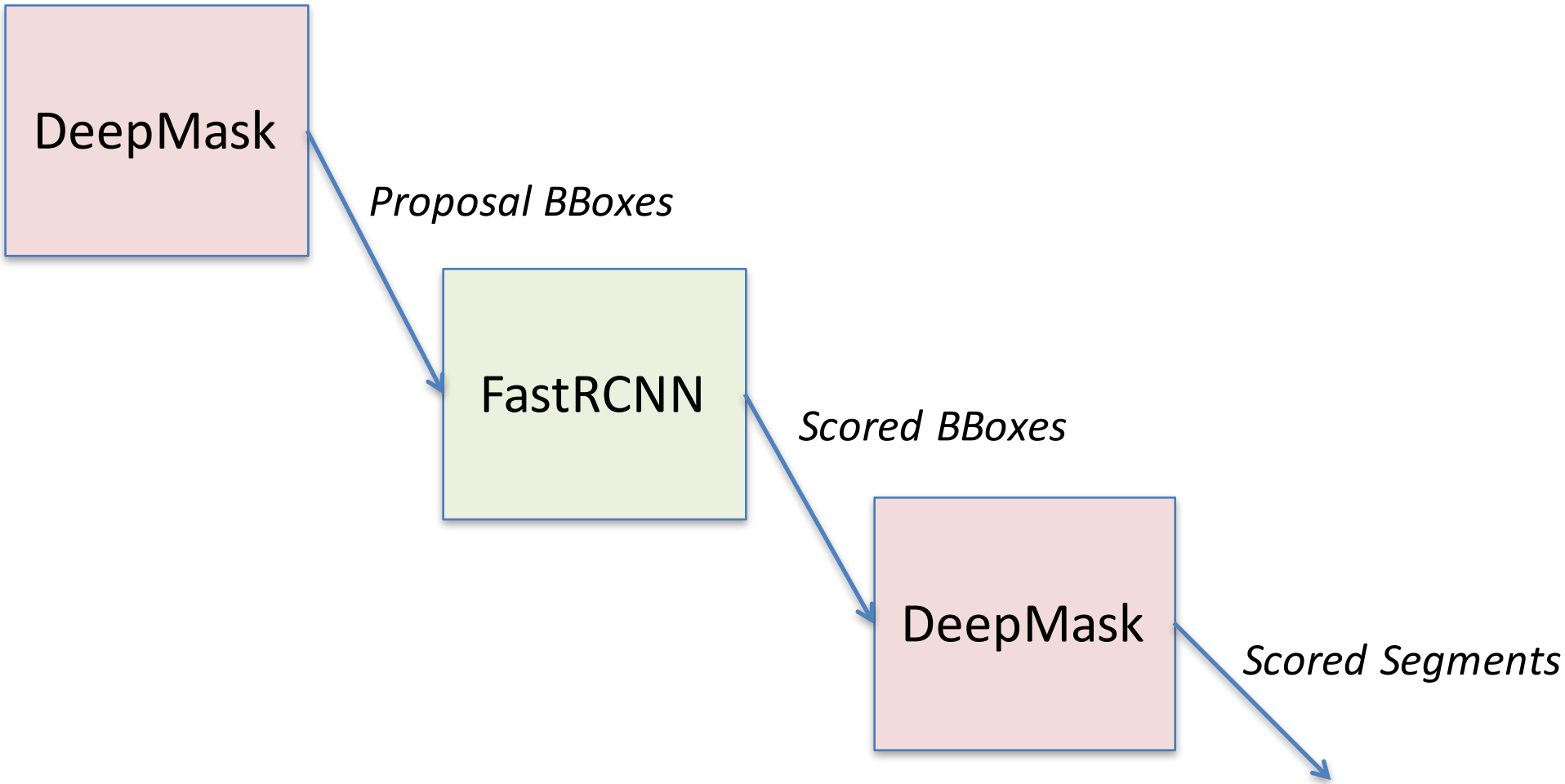
	AP
Small	0.139
Medium	0.378
Large	0.477

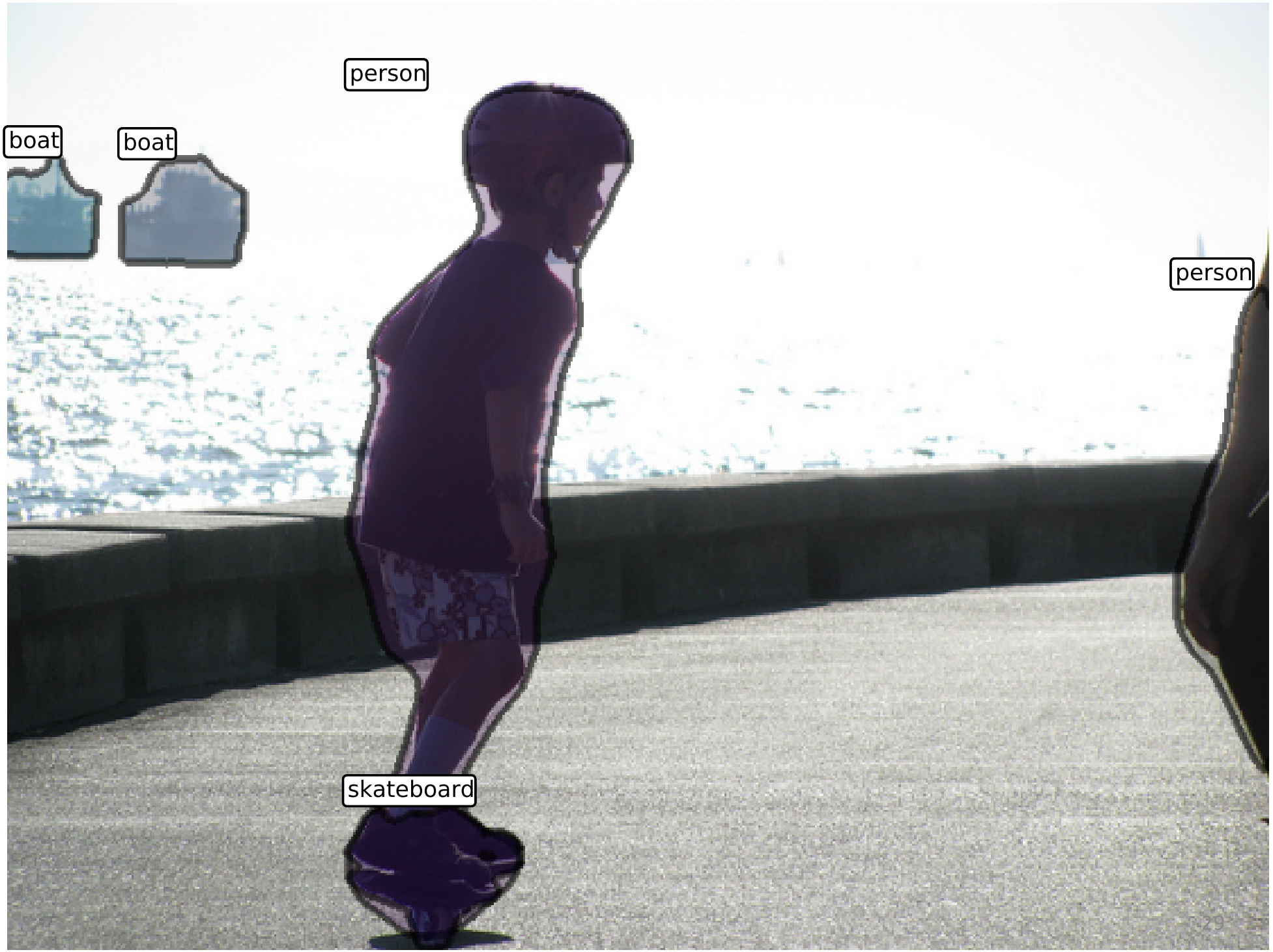


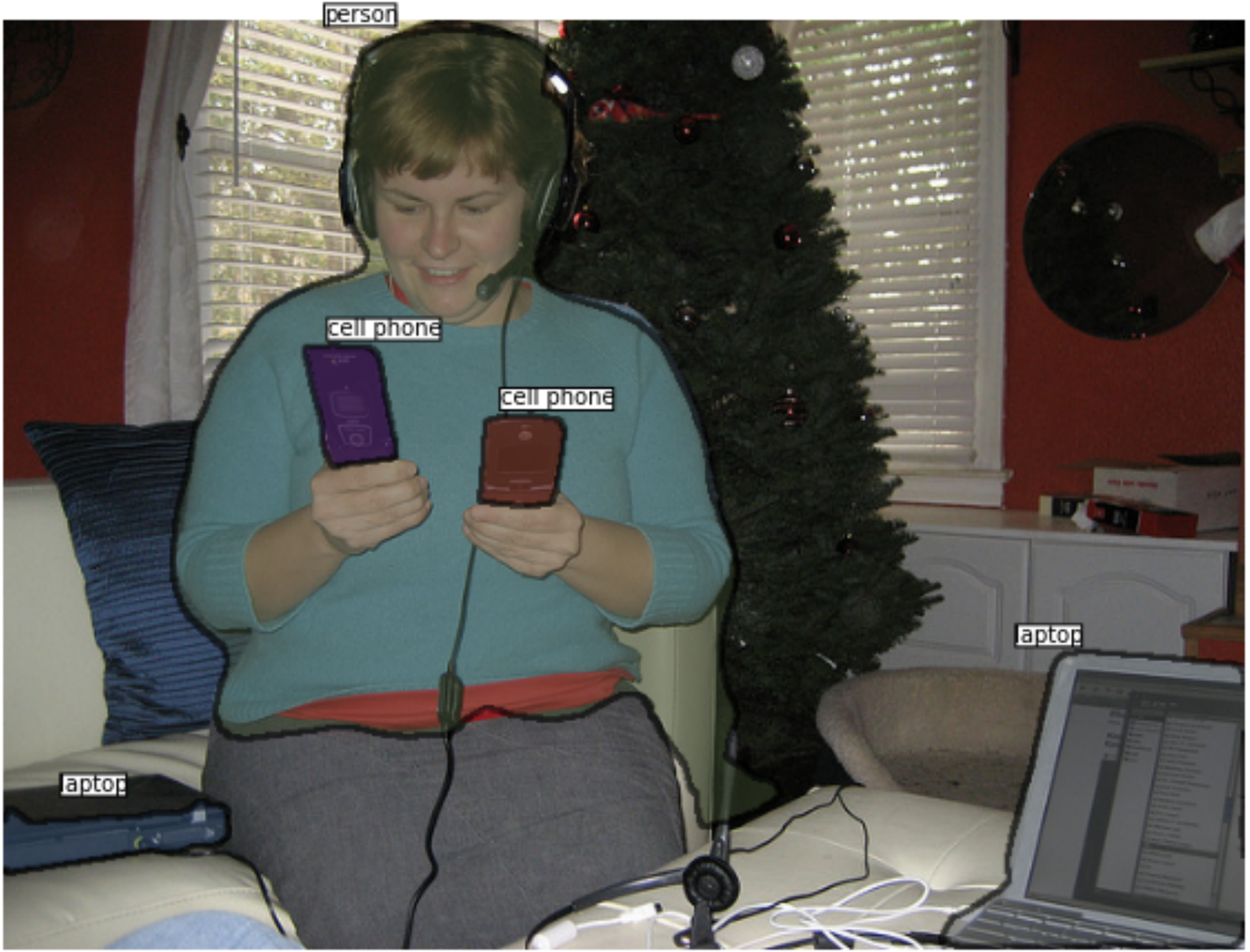


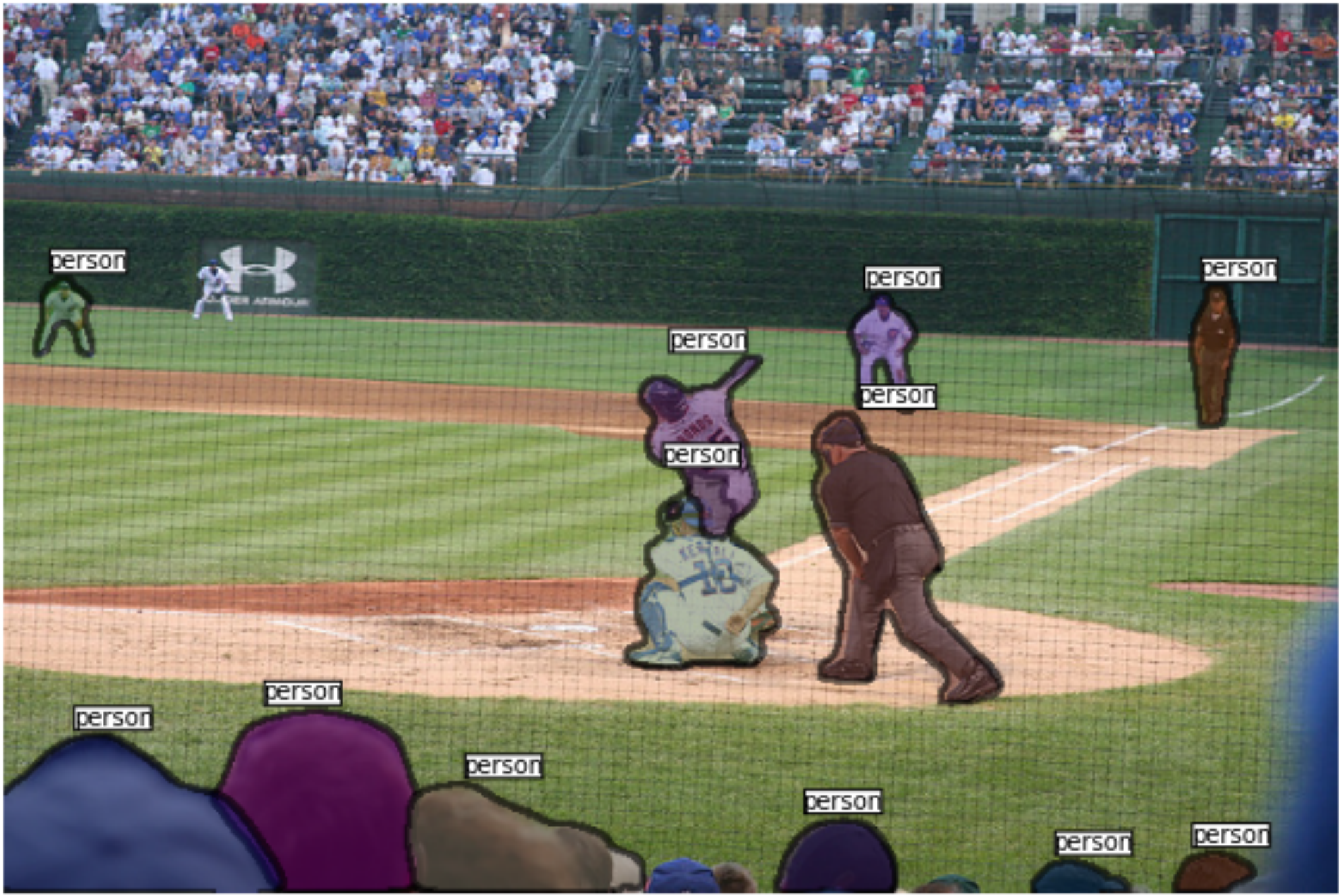


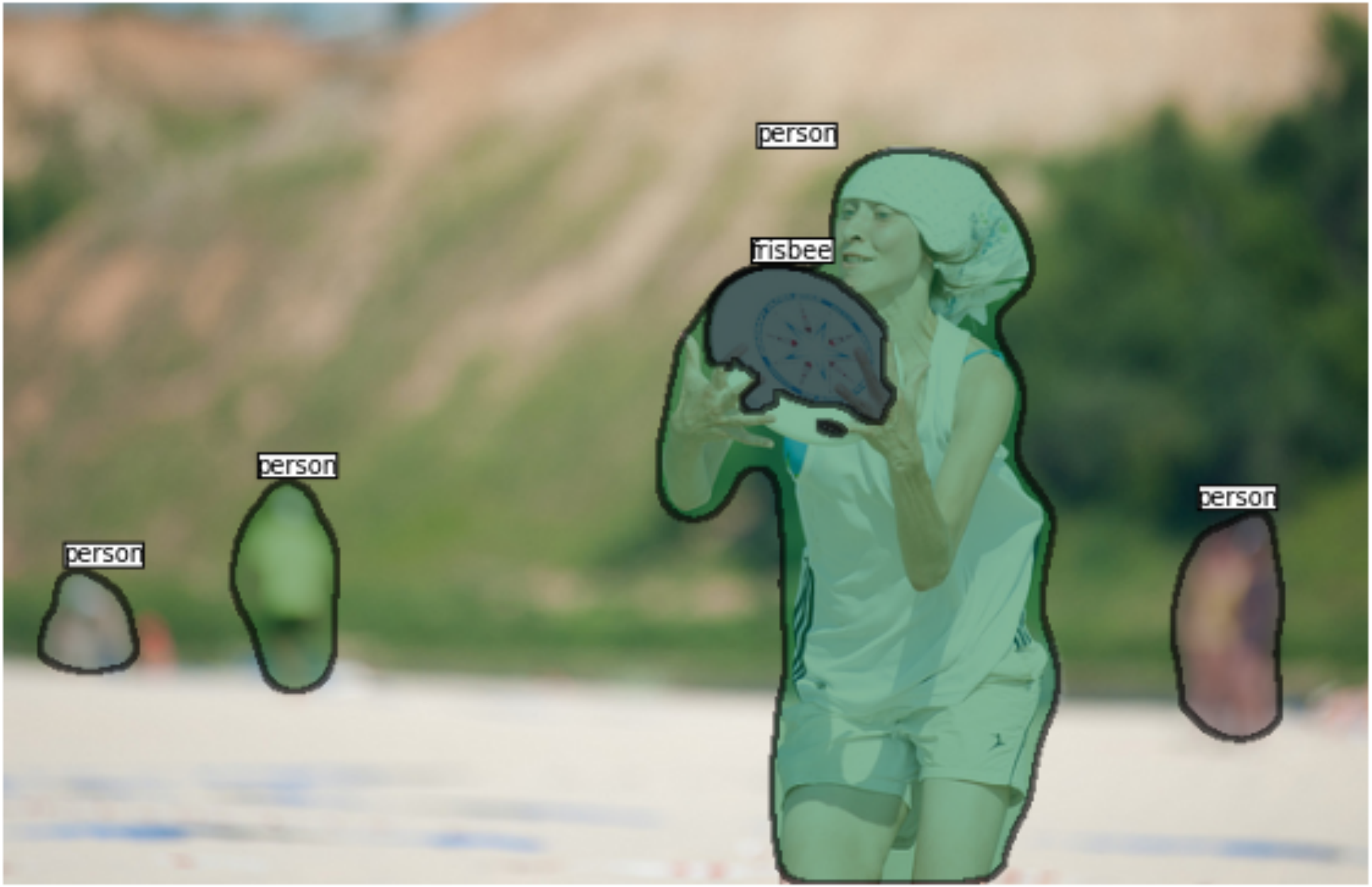
Segmentation Examples

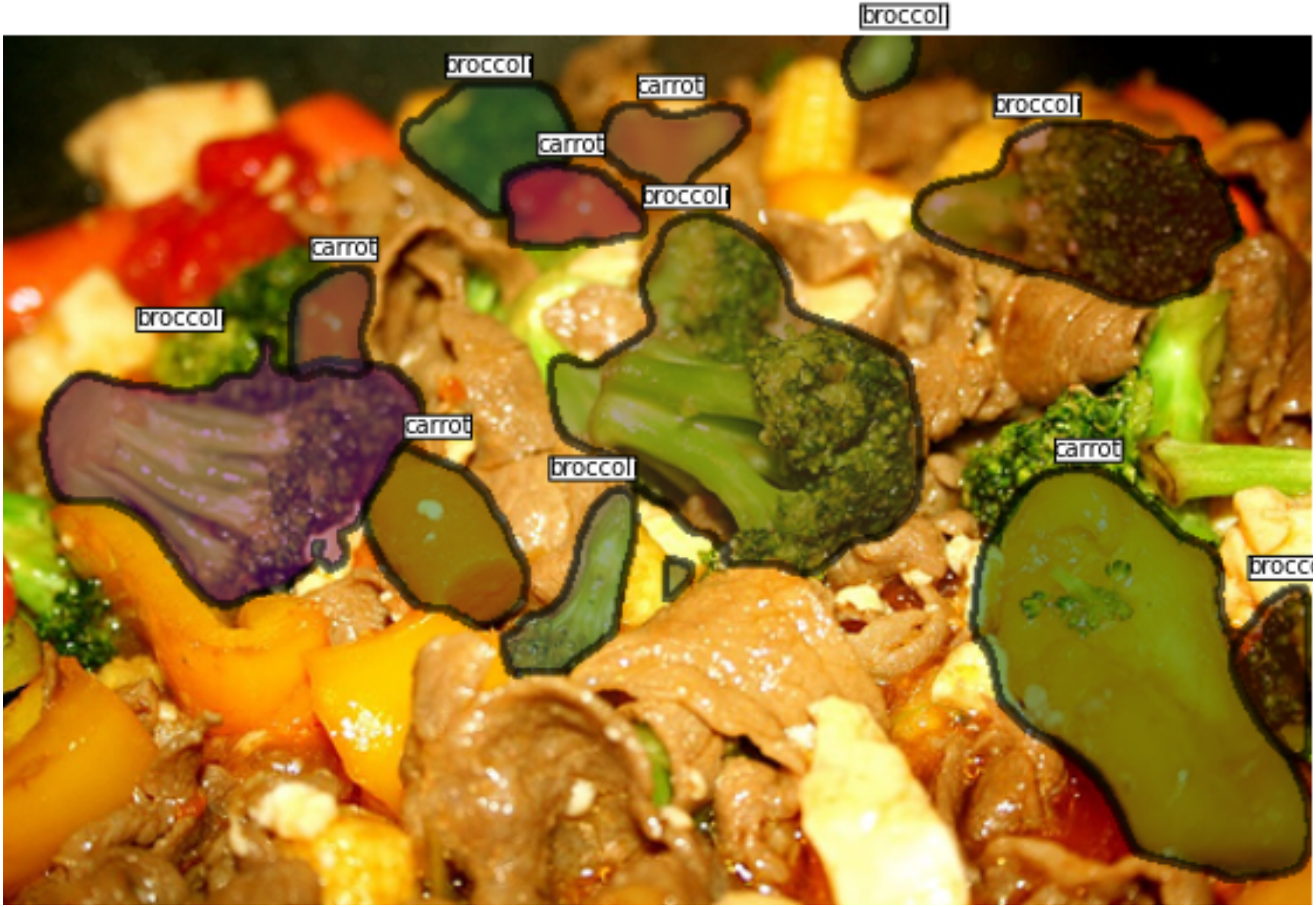



person
boat
boat
person
skateboard

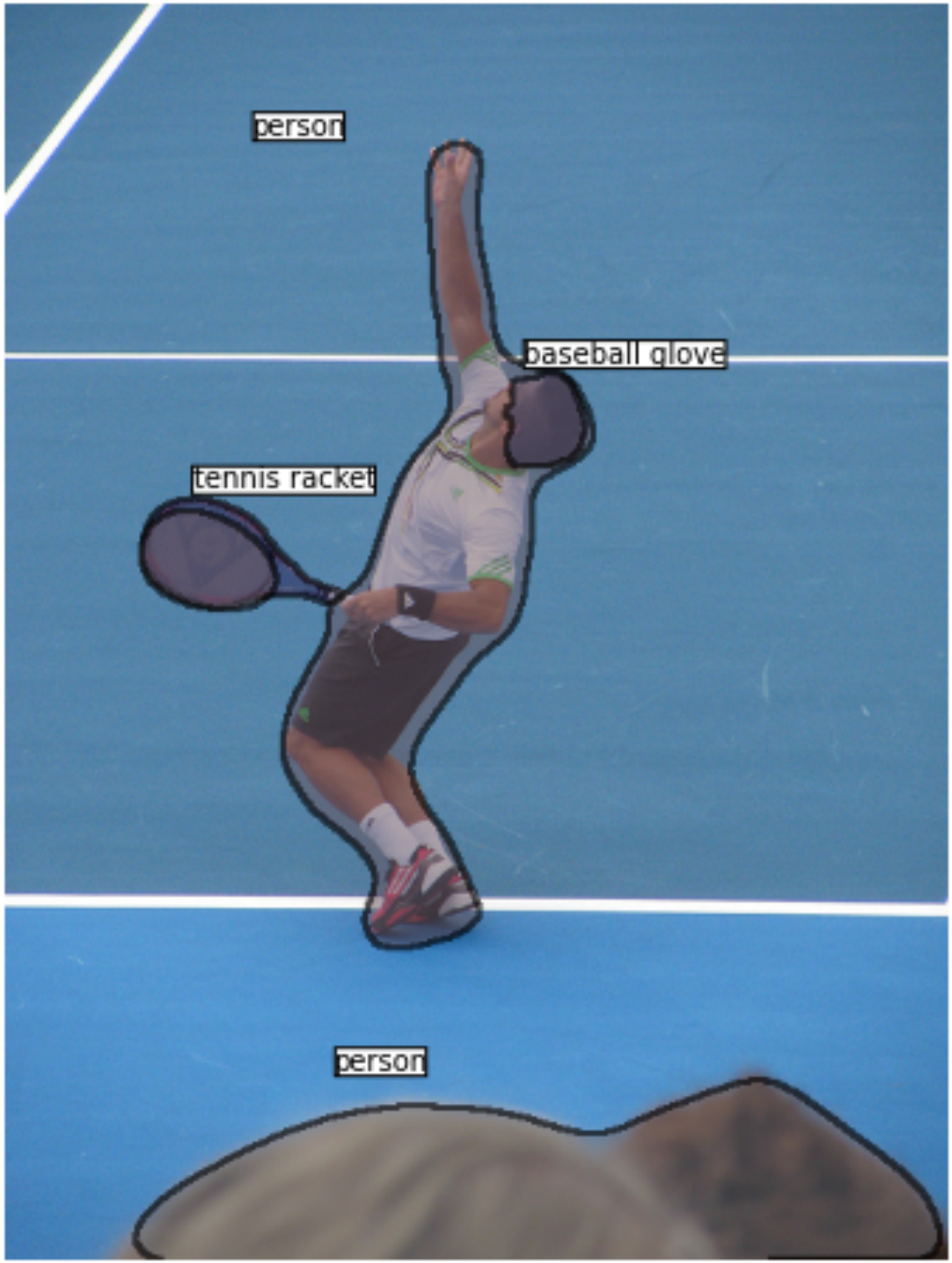

person
cell phone
cell phone
aptop
aptop


person
person
person
person
person
person
person
person
person
person
person
person


person
frisbee
person
person
person


broccol
broccol
carrot
broccol
carrot
broccol
carrot
broccol
carrot
broccol
carrot
broccol


traffic light
traffic light
car
car
car
person
car


person
baseball glove
tennis racket
person


bear
frisbee
frisbee


sink
sink


mouse
refrigerator
refrigerator
sink
sink


sheep
sheep
sheep
sheep
sheep
sheep
sheep
sheep
sheep
sheep


sheep
sheep
sheep
sheep
sheep
sheep
sheep
sheep
sheep
sheep
sheep
sheep
sheep
sheep
sheep

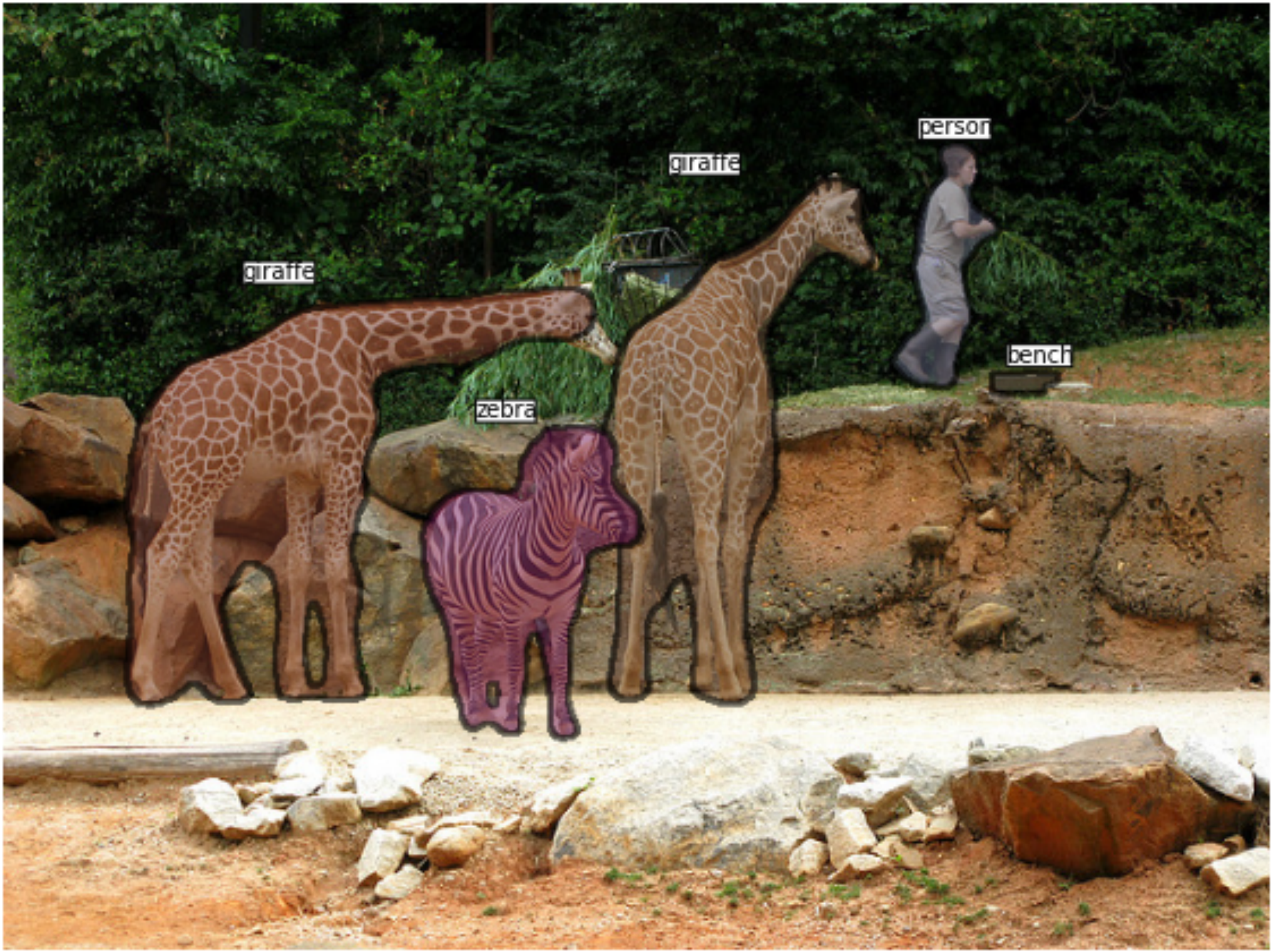

vase
vase
vase
vase
vase
vase
vase
vase
vase
vase
vase
vase
vase
vase
vase
vase
vase
vase
vase
vase
vase
vase
vase
vase

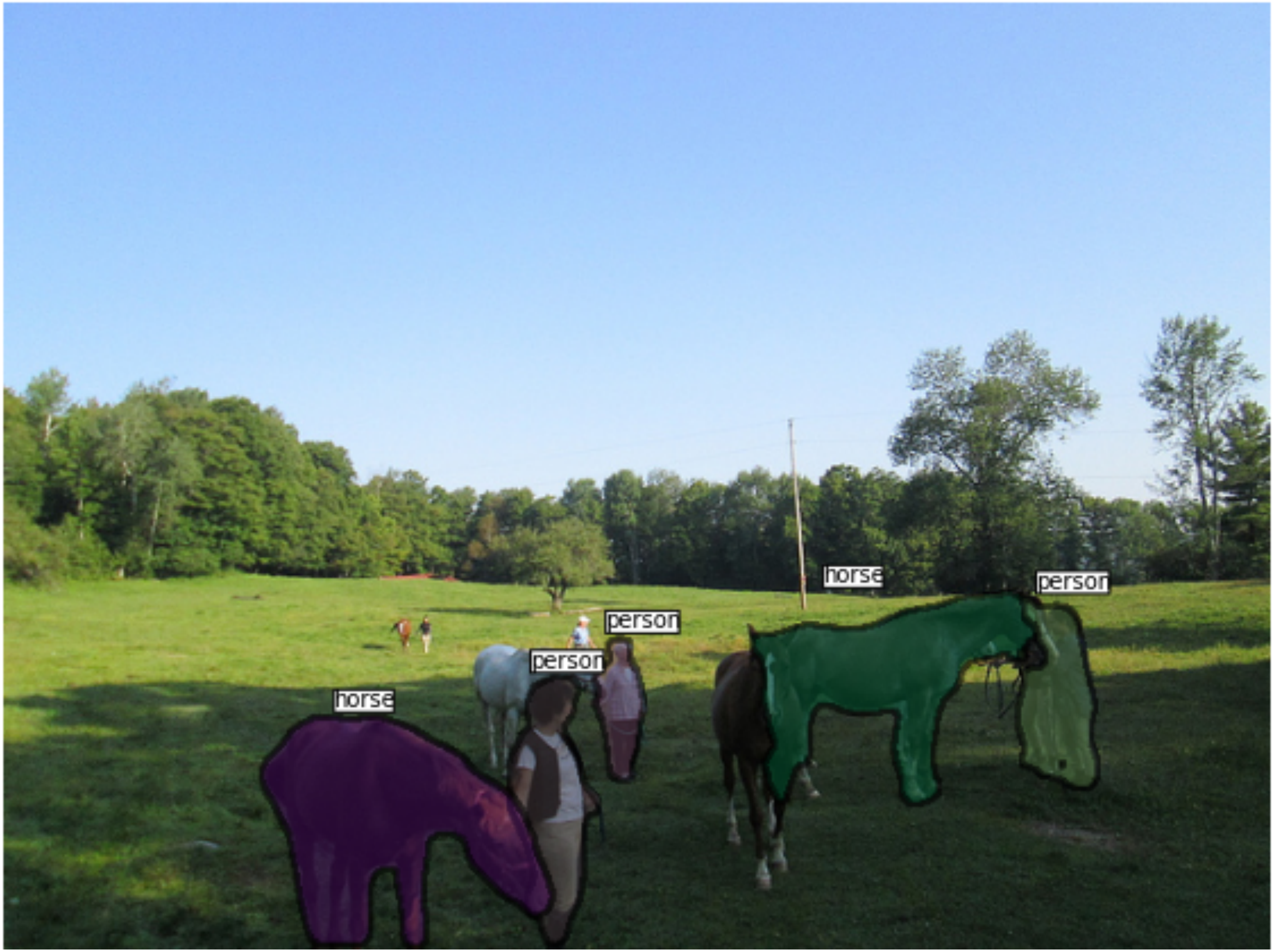

sandwich
donut
donut
donut
donut
donut
donut
donut
donut
donut
donut
donut
donut
donut
donut
donut
donut
donut
donut
donut
donut
donut
donut
donut
donut
donut
donut
donut


zebra
zebra

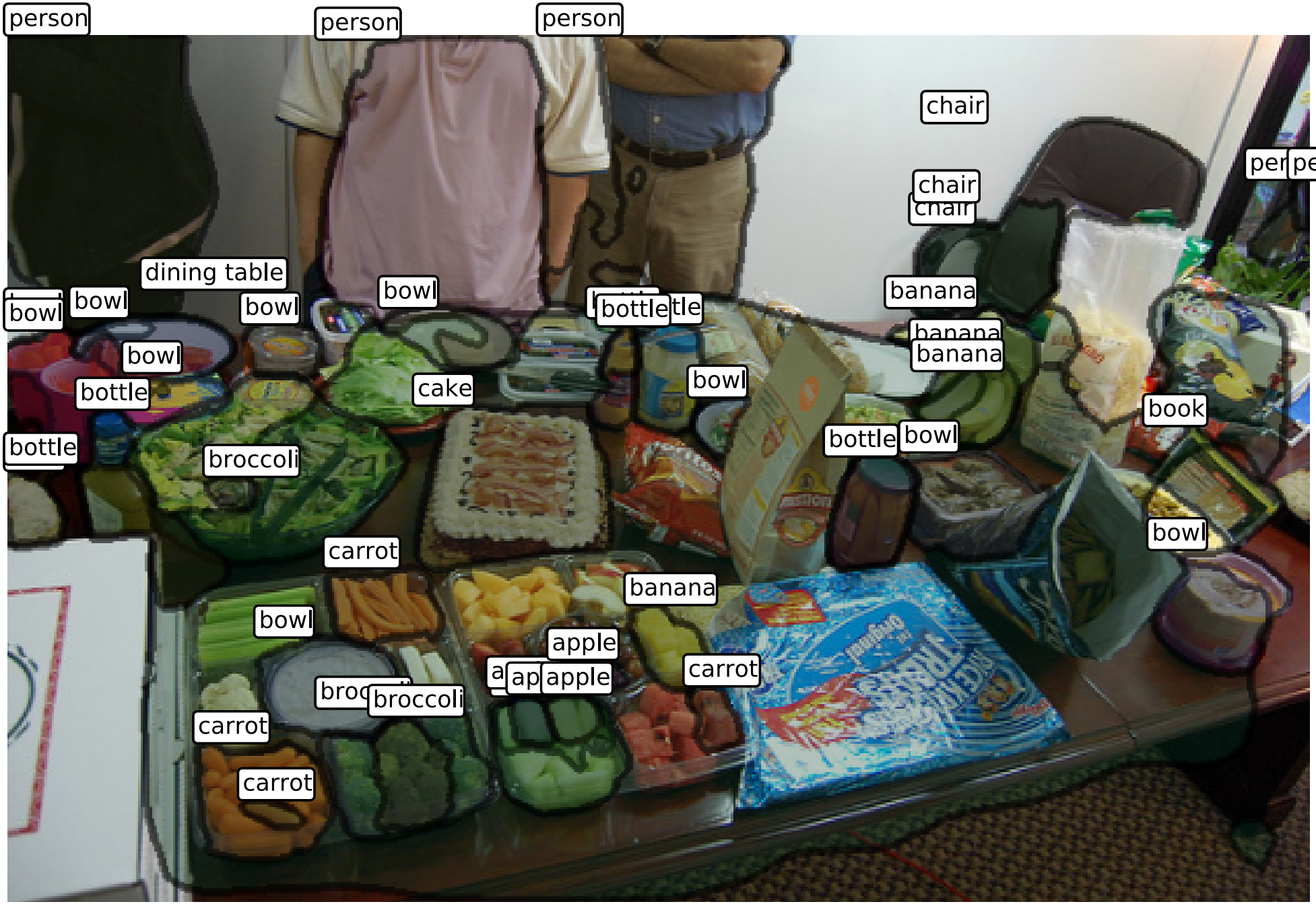
tv
book
book
book
book
book
book
book
book
bird
FRANK
A Visual
MUSEUM OF FINE ARTS
50 FAV
ORITE ROOMS BY FRANK LLOYD WRIGHT


bottle
toilet
toilet


person
giraffe
giraffe
bench
zebra


horse
horse
persor
persor
persor


potted plant
teddy bear
teddy bear
chair
teddy bear
cake
cake
cake
cake
cake
cake
cake
cake


person
person
person
chair
chair
dining table
bowl
bowl
bowl
bowl
bowl
bottle
banana
banana
banana
bowl
cake
bottle
bottle
bowl
book
bottle
broccoli
carrot
bowl
banana
bowl
apple
carrot
apple
broccoli
carrot
carrot

Future Directions

- most room for improvement:
  - background confusion (FP/FN)
  - small objects
- more effective use of context
- fast / proposal-free detection


cup
bowl
clock
person
bowl
clock
person
clock
vase
vase
bottle
bowl
Questions?
clock
vase
vase
vase
cup
cup
cup
vase
vase
vase
vase
vase
cup
voyageAnatolia.blogspot.com