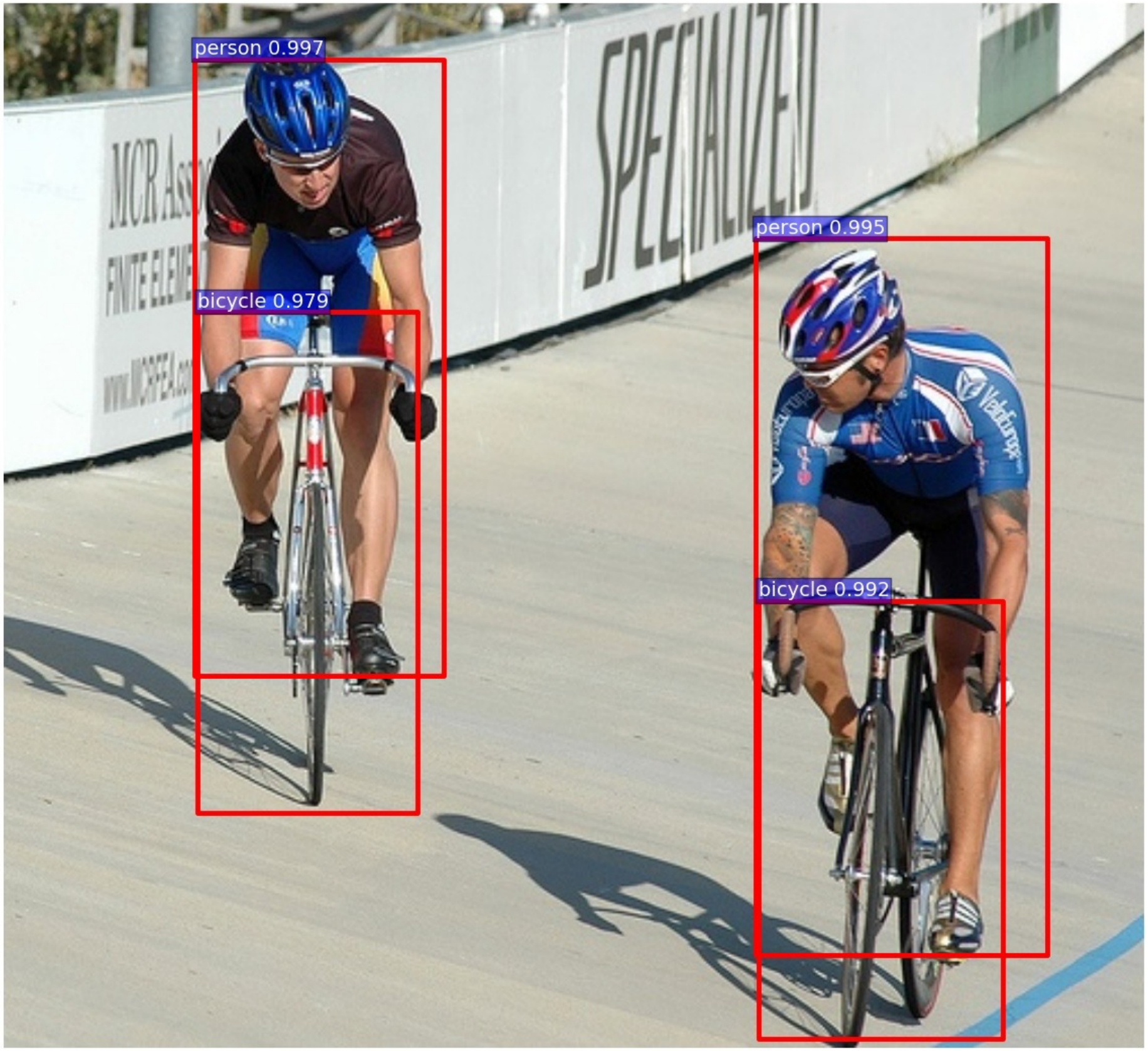
Object Detection using R-CNN Experiments

CS381V: Visual Recognition, Spring 2016

William Xie

Feb. 24, 2016


person 0.997
bicycle 0.979
person 0.995
bicycle 0.992
SPECIALIZED

Fast R-CNN

- R-CNN: Girshick et al., CVPR 2013
- **Fast R-CNN: Girshick, ICCV 2015**
- Faster R-CNN: Ren et al., NIPS 2015

Fast R-CNN

Ross Girshick
Microsoft Research
rbg@microsoft.com

Abstract

while achieving top accuracy on PASCAL VOC 2012 [7]

Fast R-CNN

- Implemented in modified Caffe, requires Matlab

- With VGG16
  Train: 9x faster than traditional R-CNN
  Test: 200x faster than R-CNN *

*https://github.com/rbgirshick/fast-rcnn

Fast R-CNN

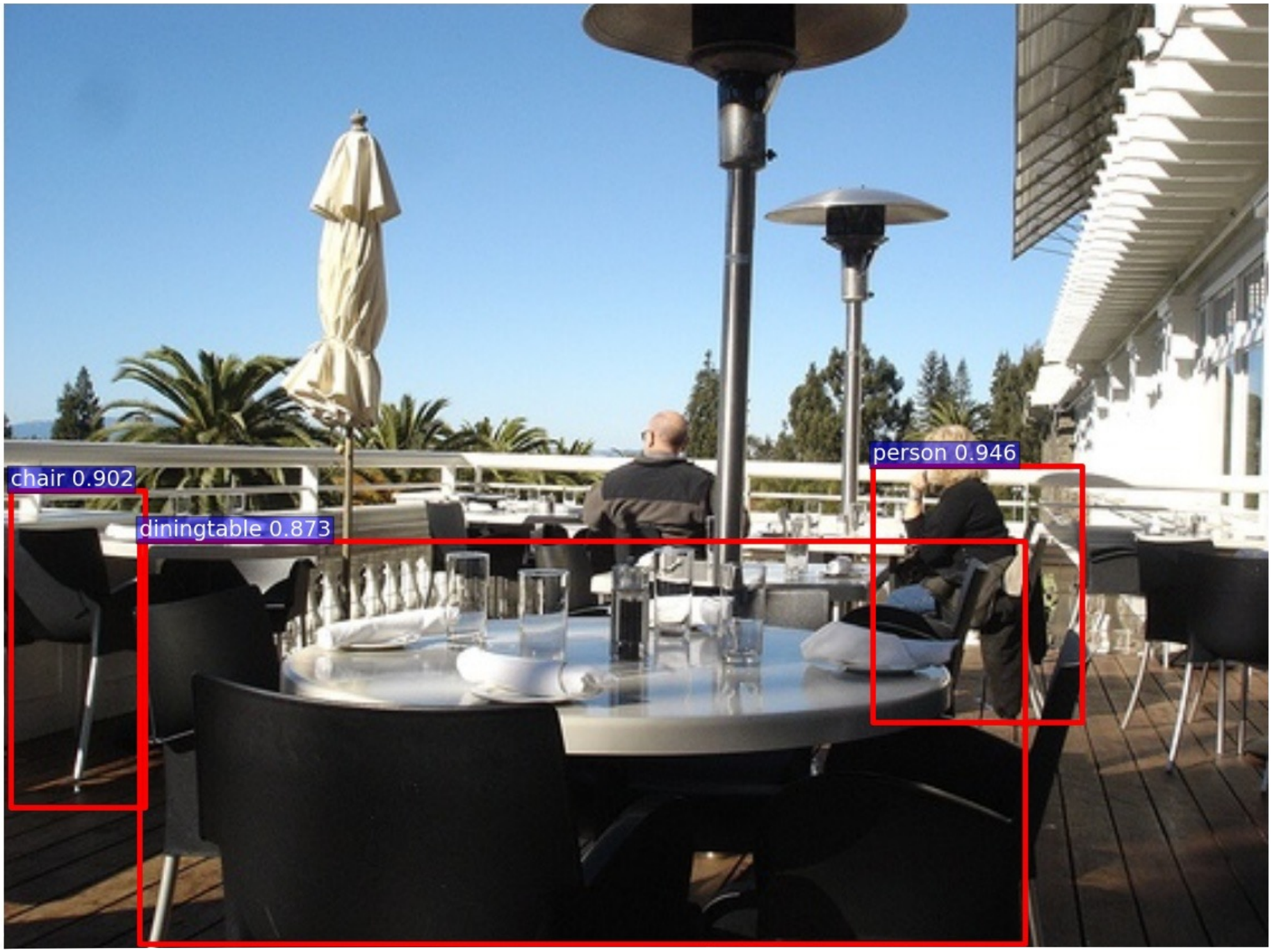
- Available models: CaffeNet, VGG16, VGG_M_1024
- Trained with ImageNet (ILSVRC 2012), fine-tuned on PASCAL VOC 2007

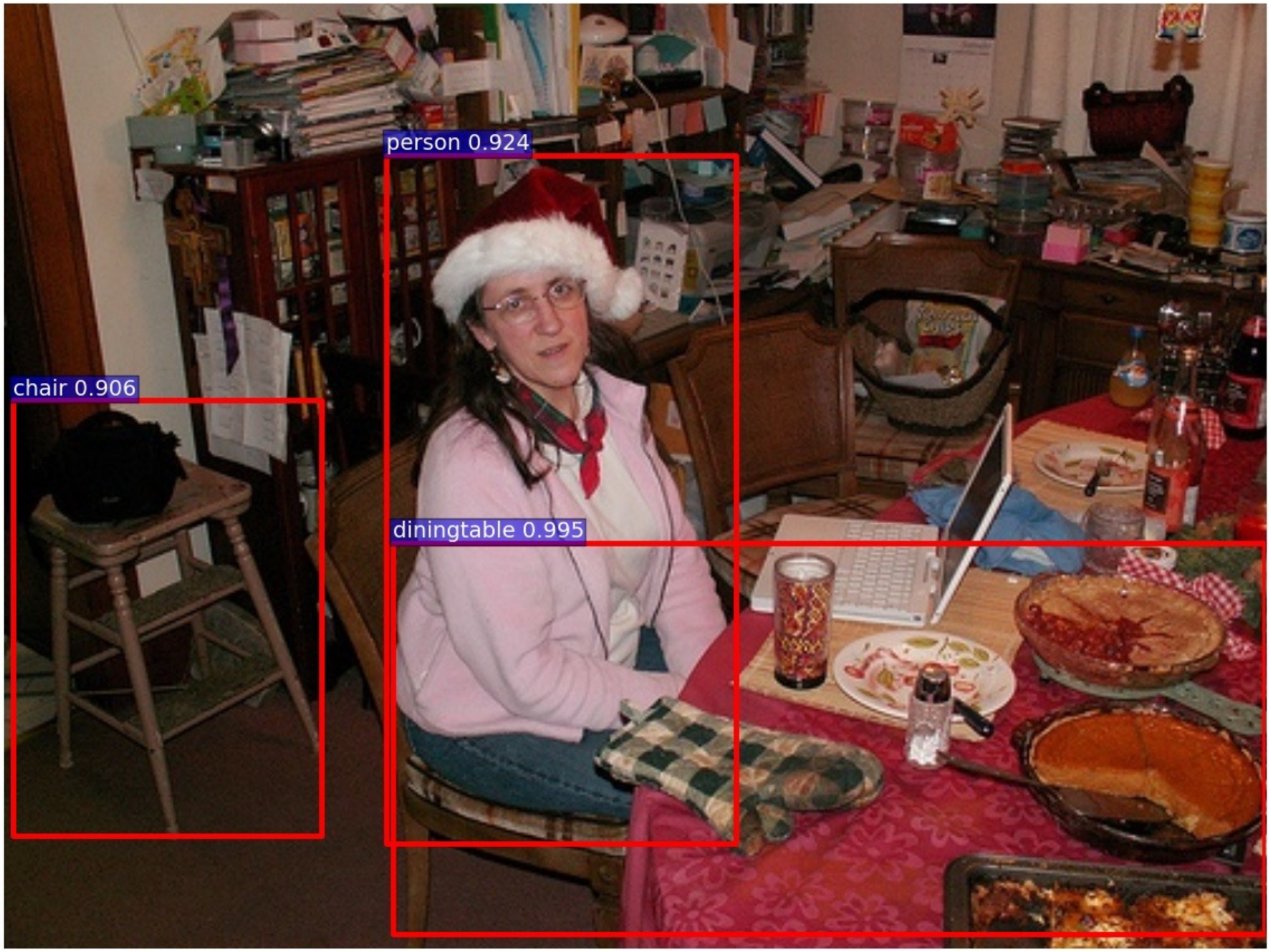
PASCAL VOC

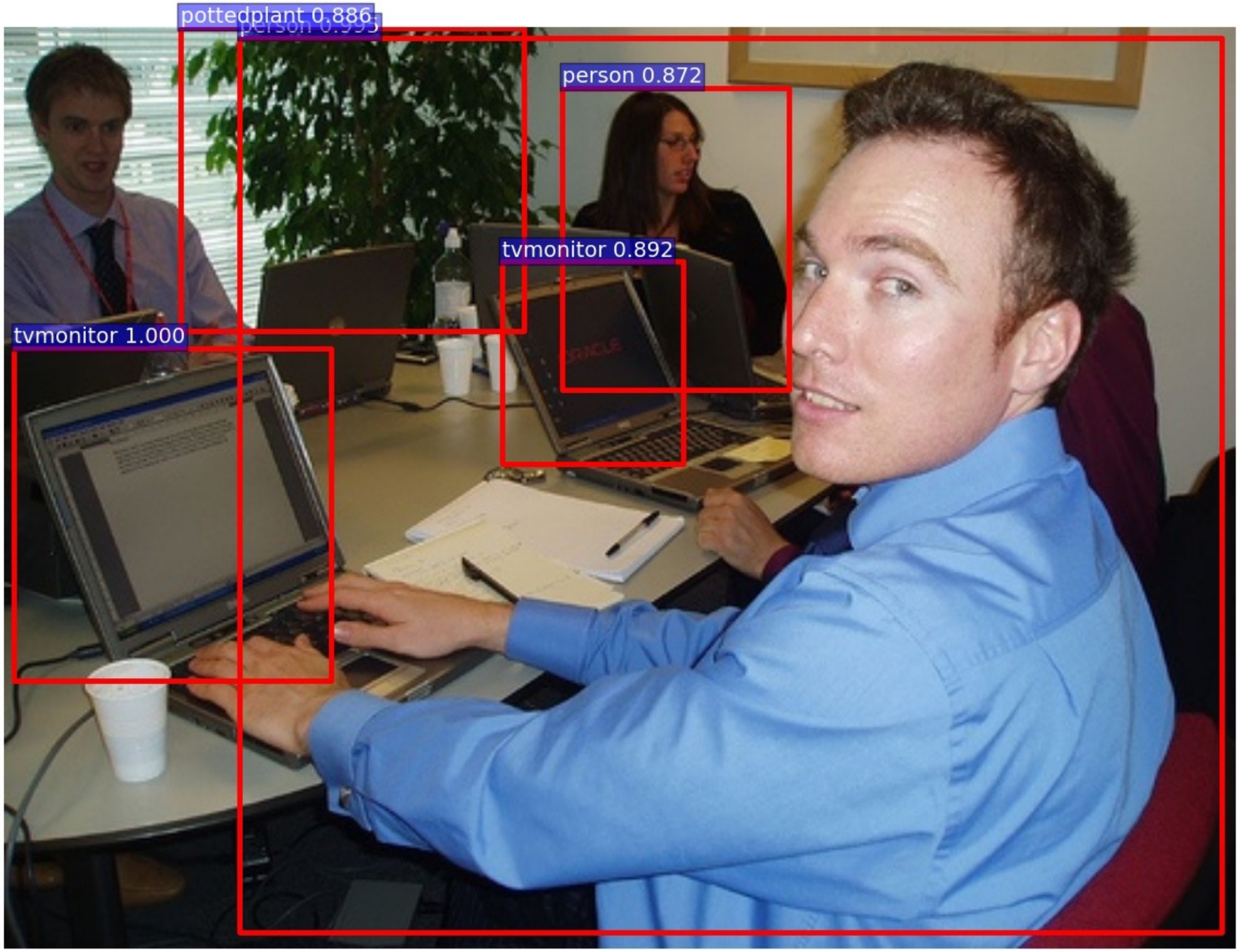
- 20 classes + background

```
CLASSES = ('__background__',
    'aeroplane', 'bicycle', 'bird', 'boat',
     'bottle', 'bus', 'car', 'cat', 'chair',
    'cow', 'diningtable', 'dog', 'horse',
    'motorbike', 'person', 'pottedplant',
     'sheep', 'sofa', 'train', 'tvmonitor')
```

Positive examples

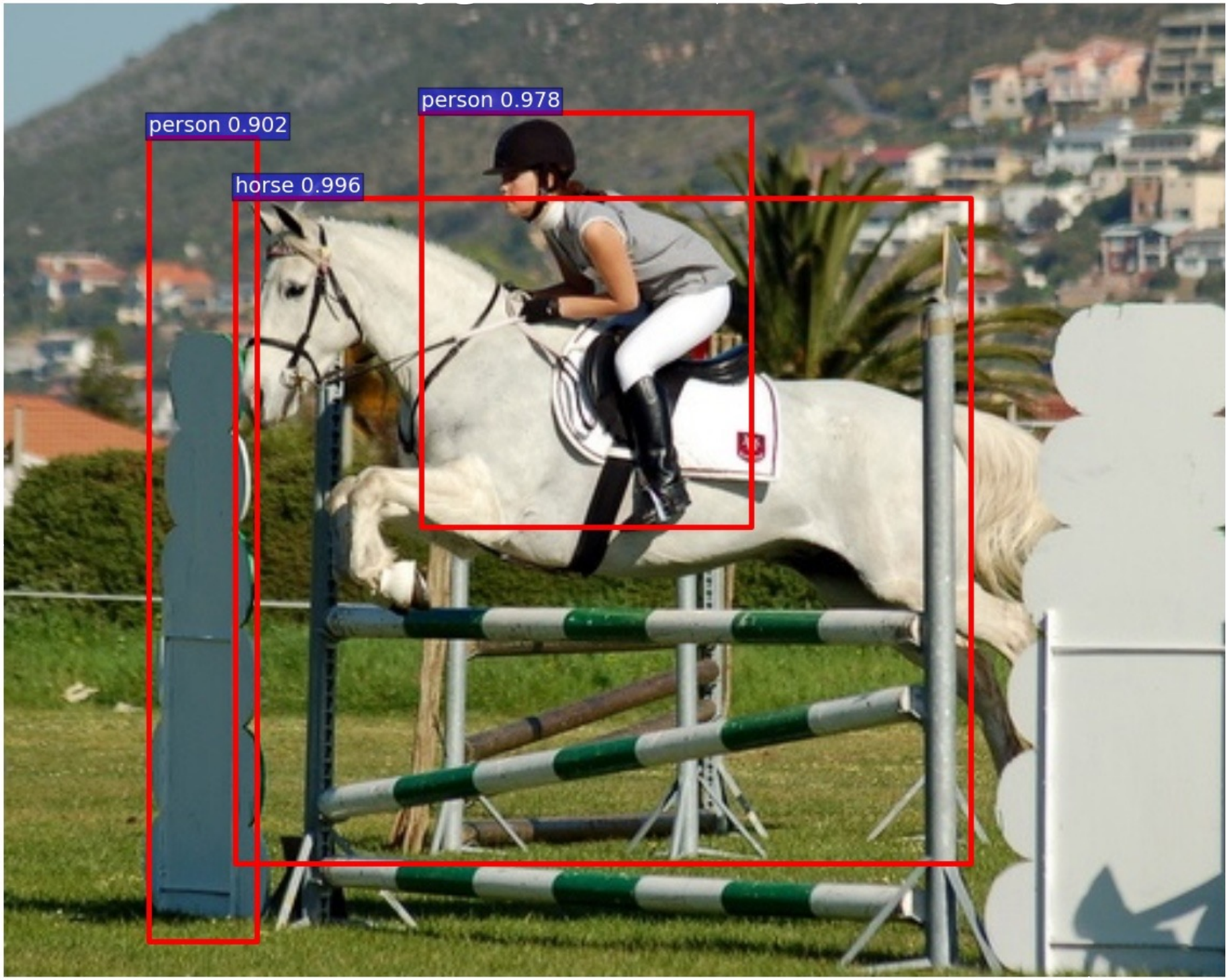



person 0.924
chair 0.906
diningtable 0.995

- 
pottedplant 0.886
person 0.993
person 0.872
tvmonitor 0.892
tvmonitor 1.000

person 0.982
person 0.892
person 0.998
person 0.919
person 0.944

Negative examples



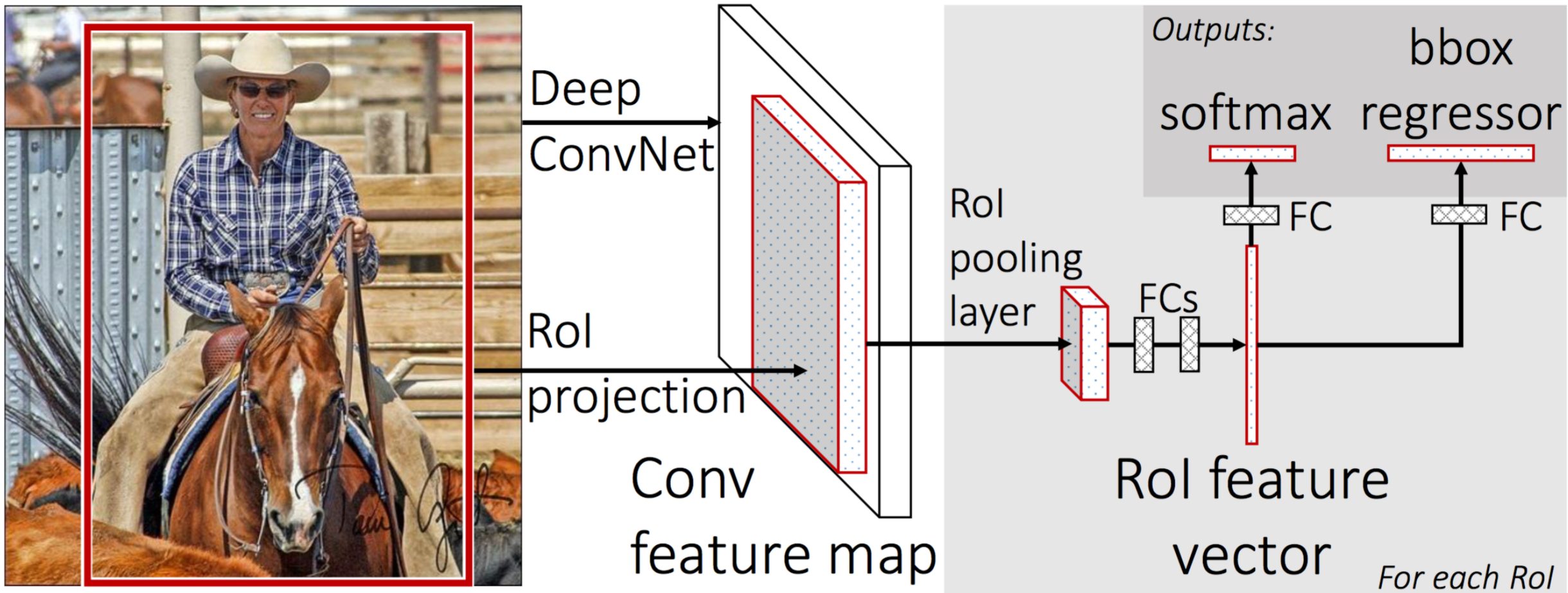
bird 0.838

cat 0.822
dog 0.932

person 0.986
person 0.857
cat 0.876

tvmonitor 0.910

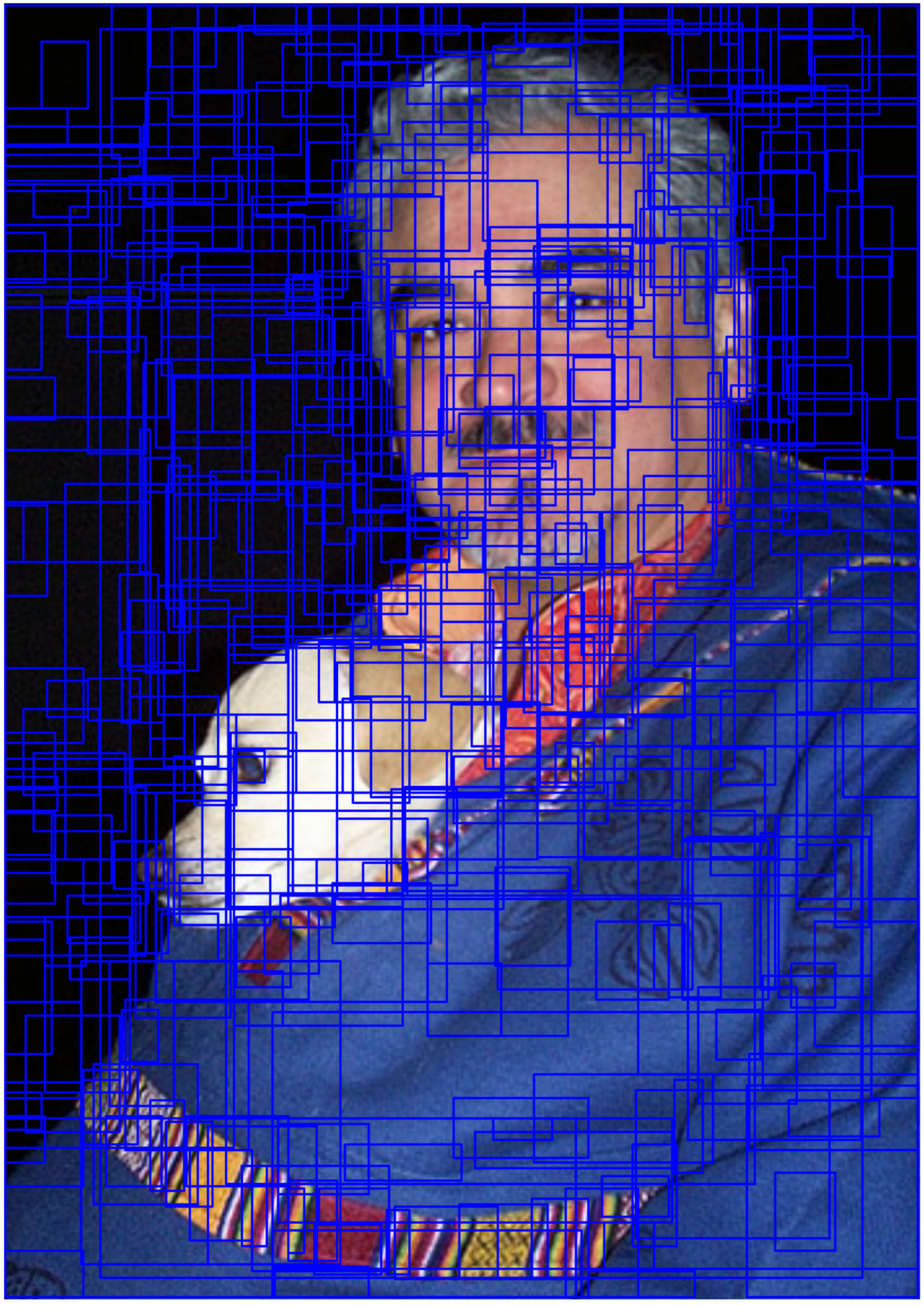
aeroplane 0.887
person 0.829



- Each region of interest -> 21 scores, 21 boxes
- Non-maximum suppression and probability threshold

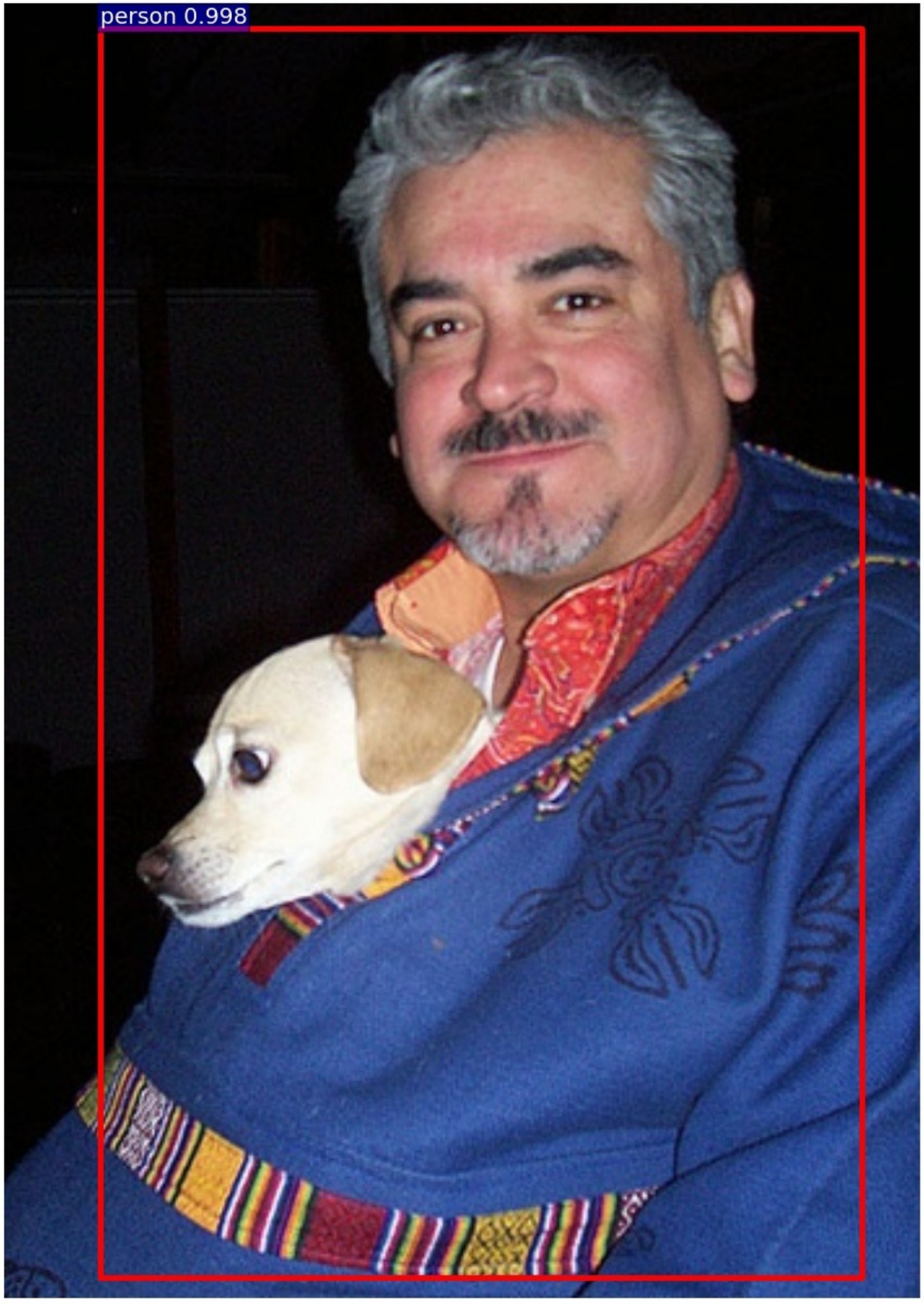
image: Girshick, Ross. "Fast r-cnn." Proceedings of the IEEE International Conference on Computer Vision. 2015.

Input image



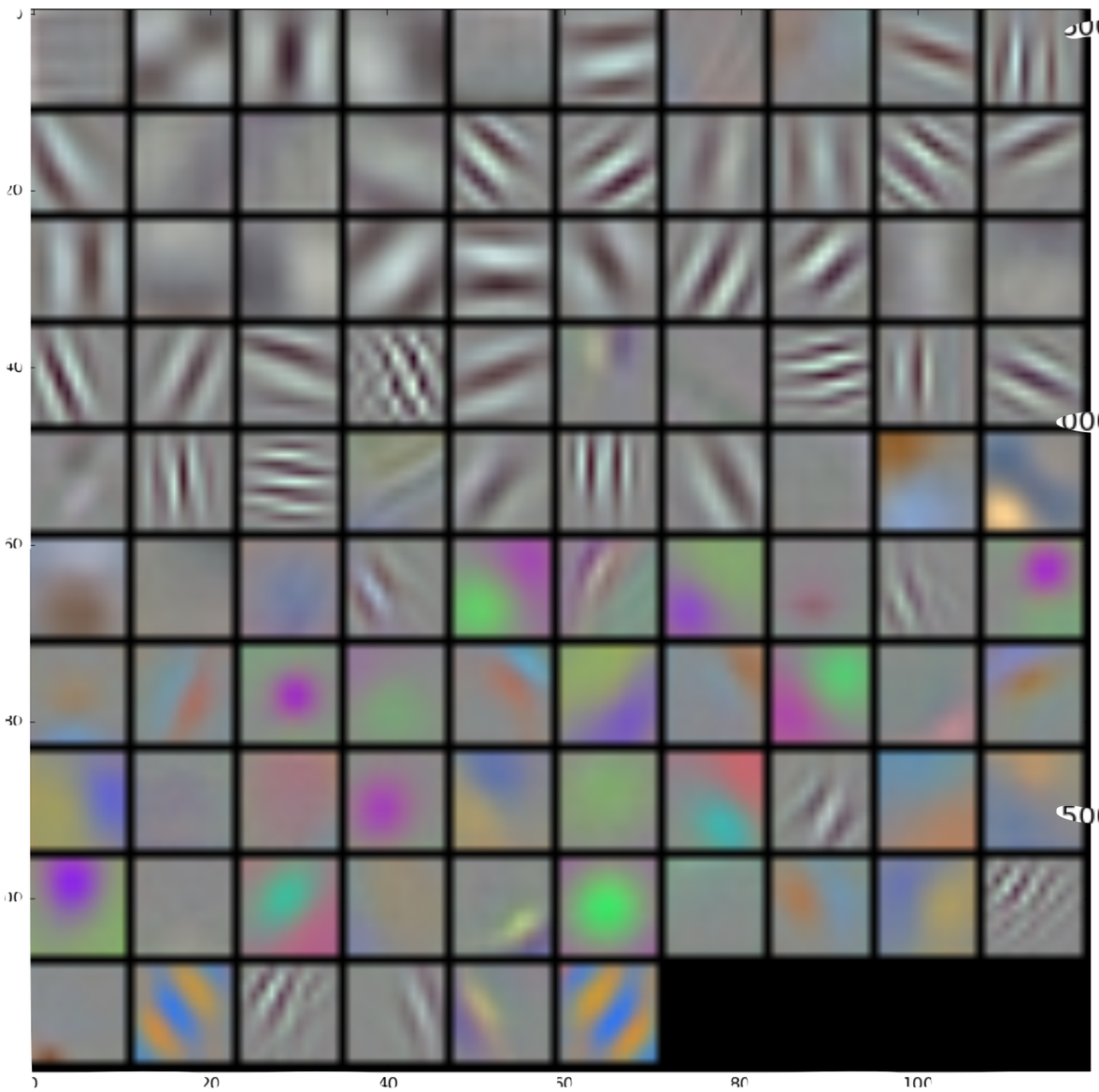
Region proposal

~2000 per image
(Selective search)

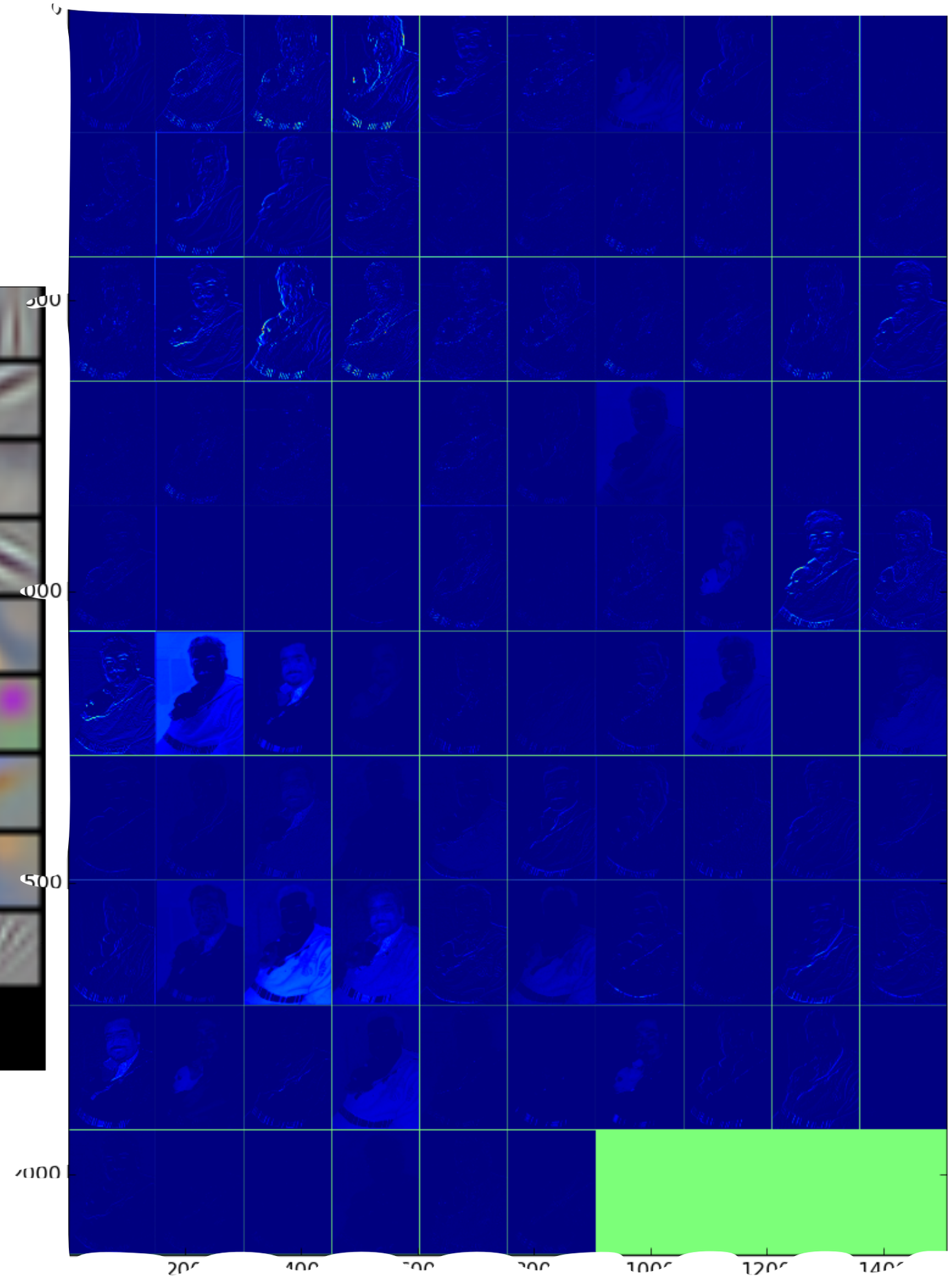


Detection and classification

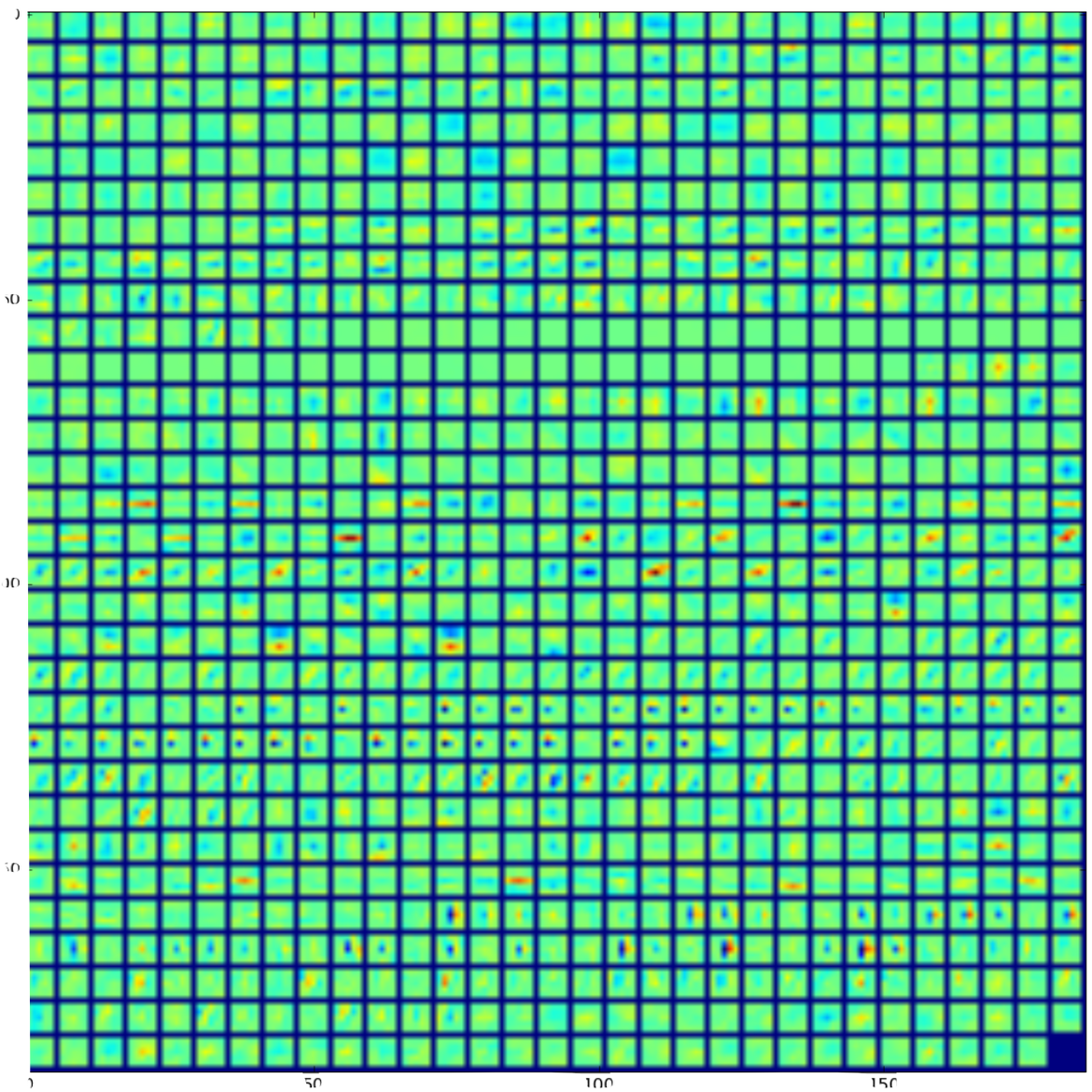
Conv1



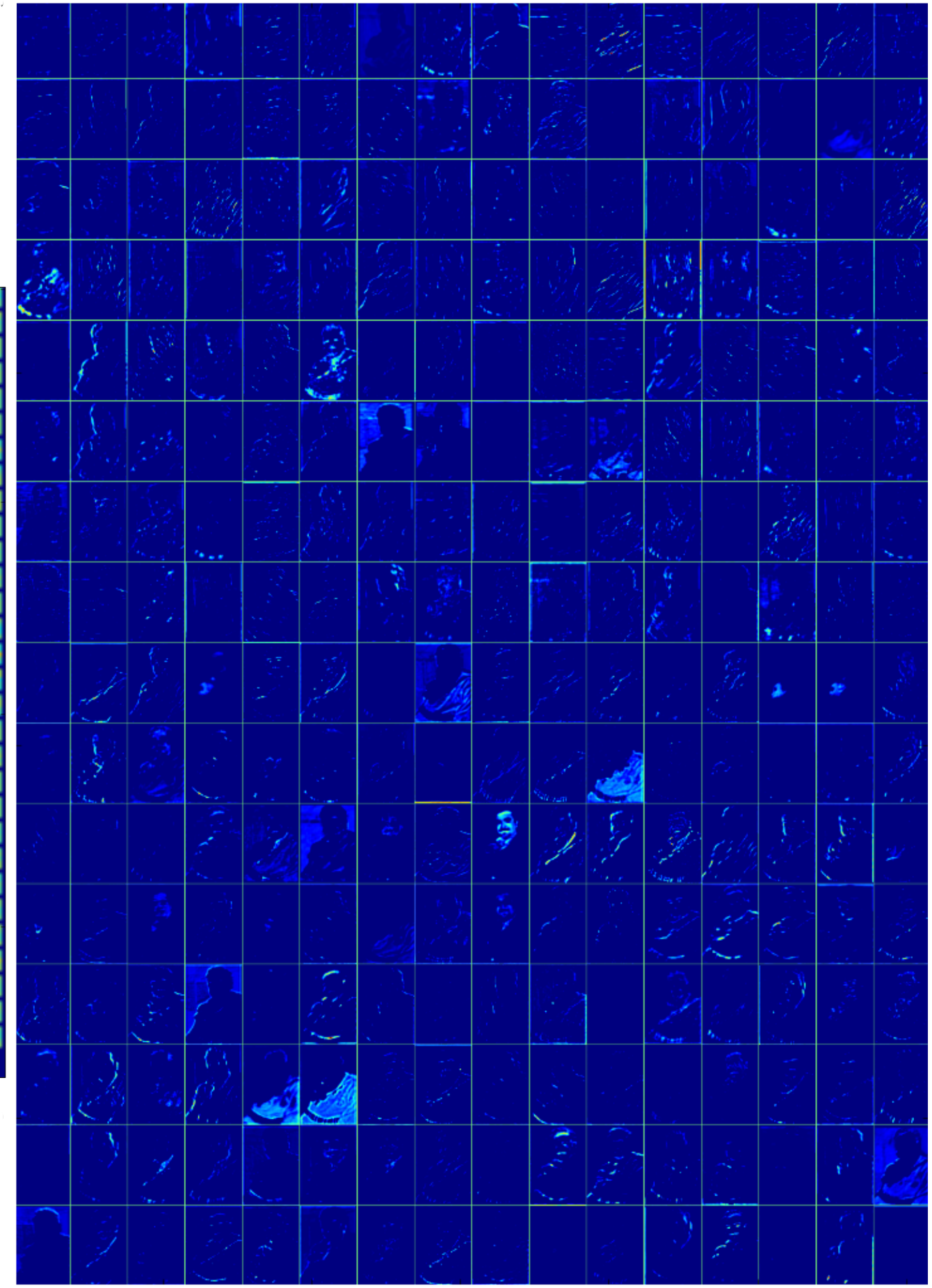
11 x 11



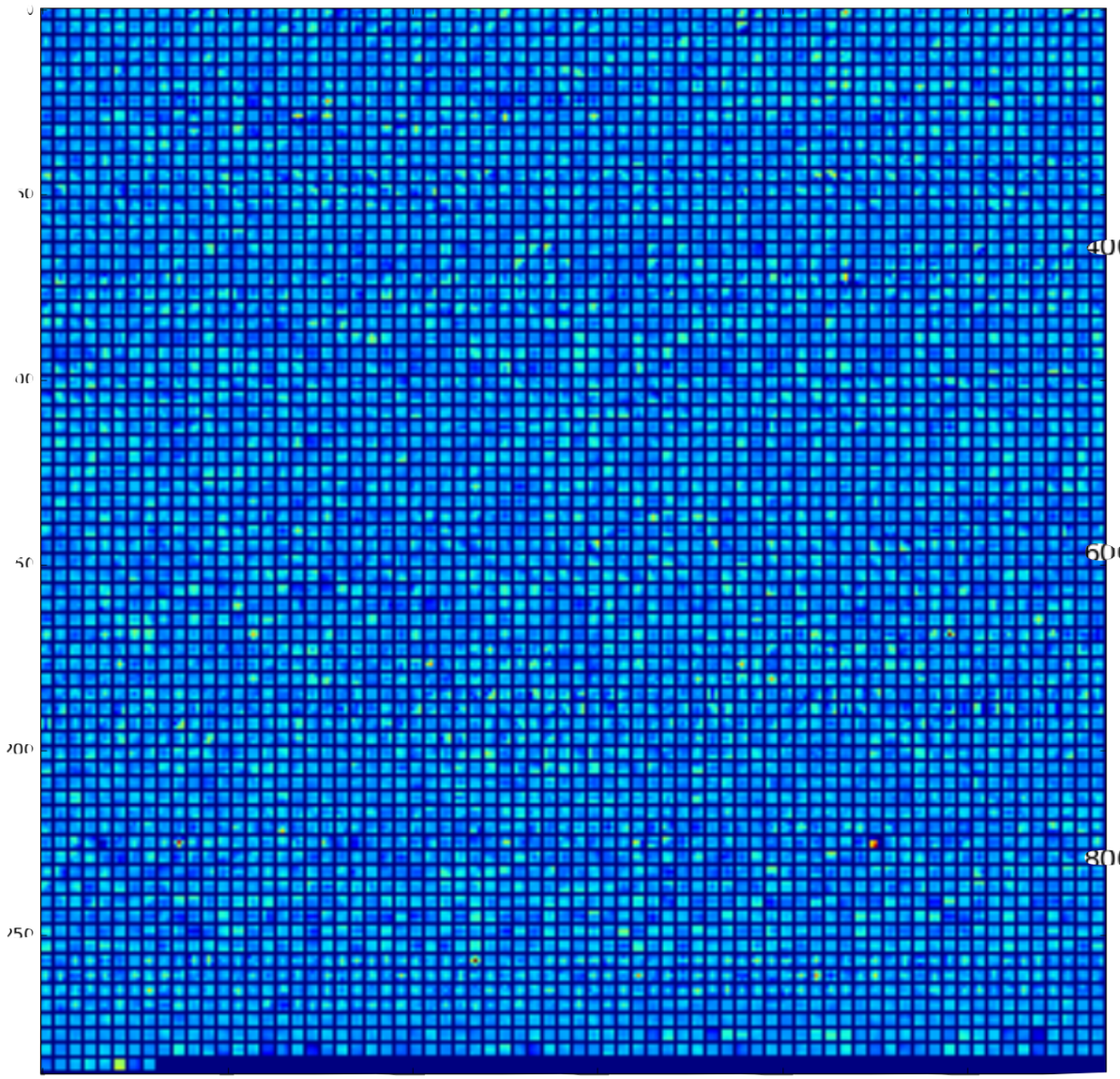
Conv2



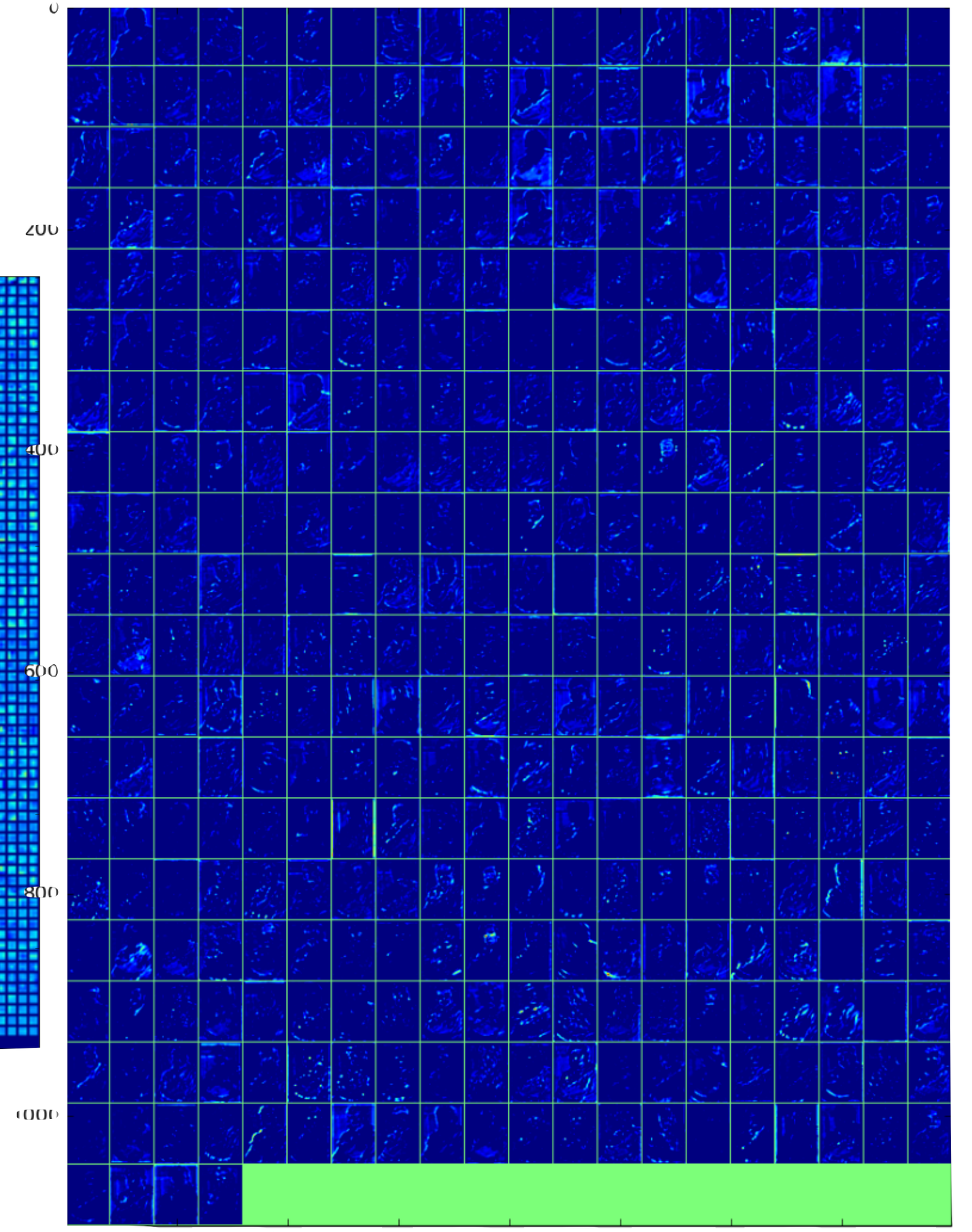
5 x 5



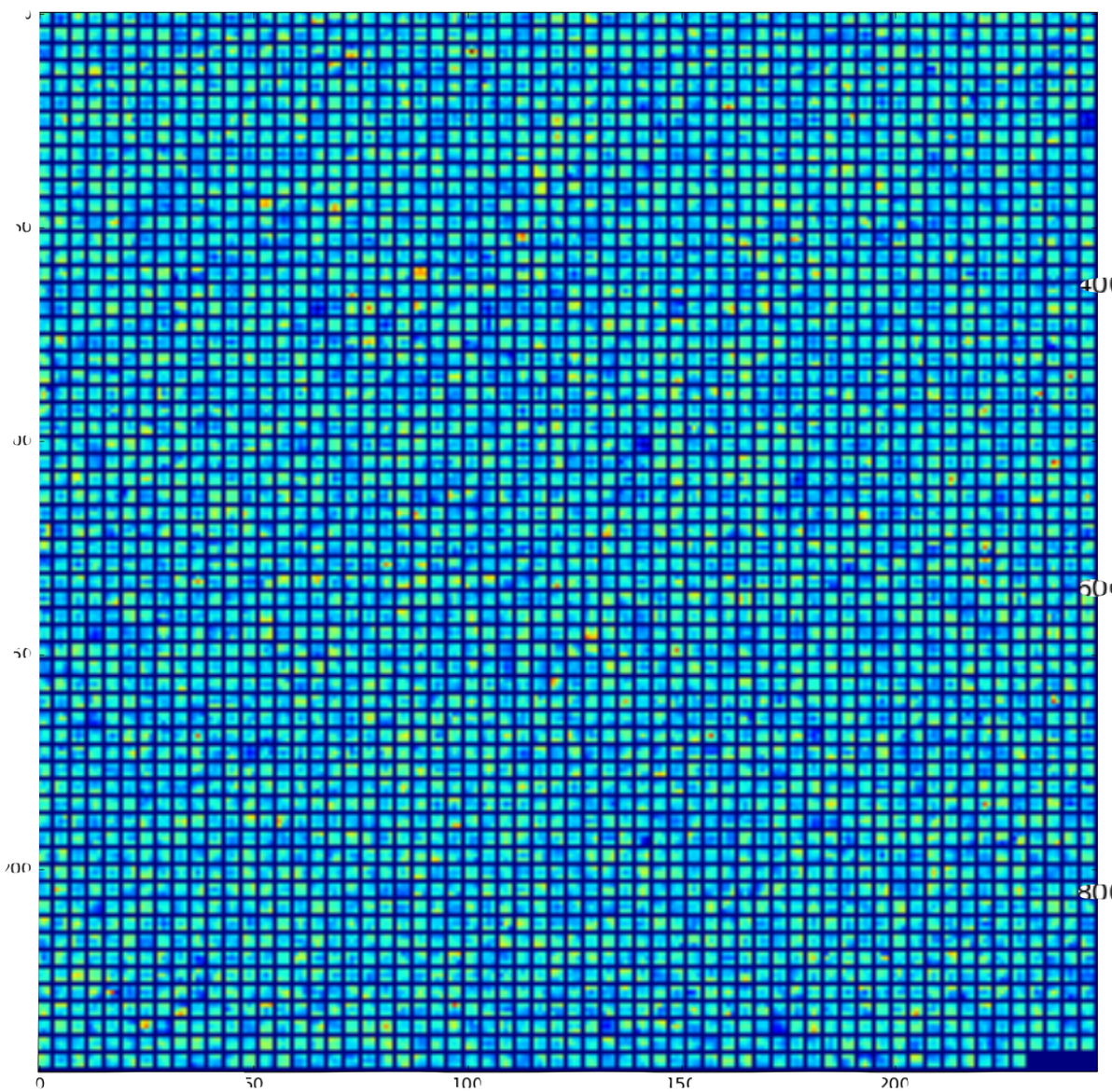
Conv3



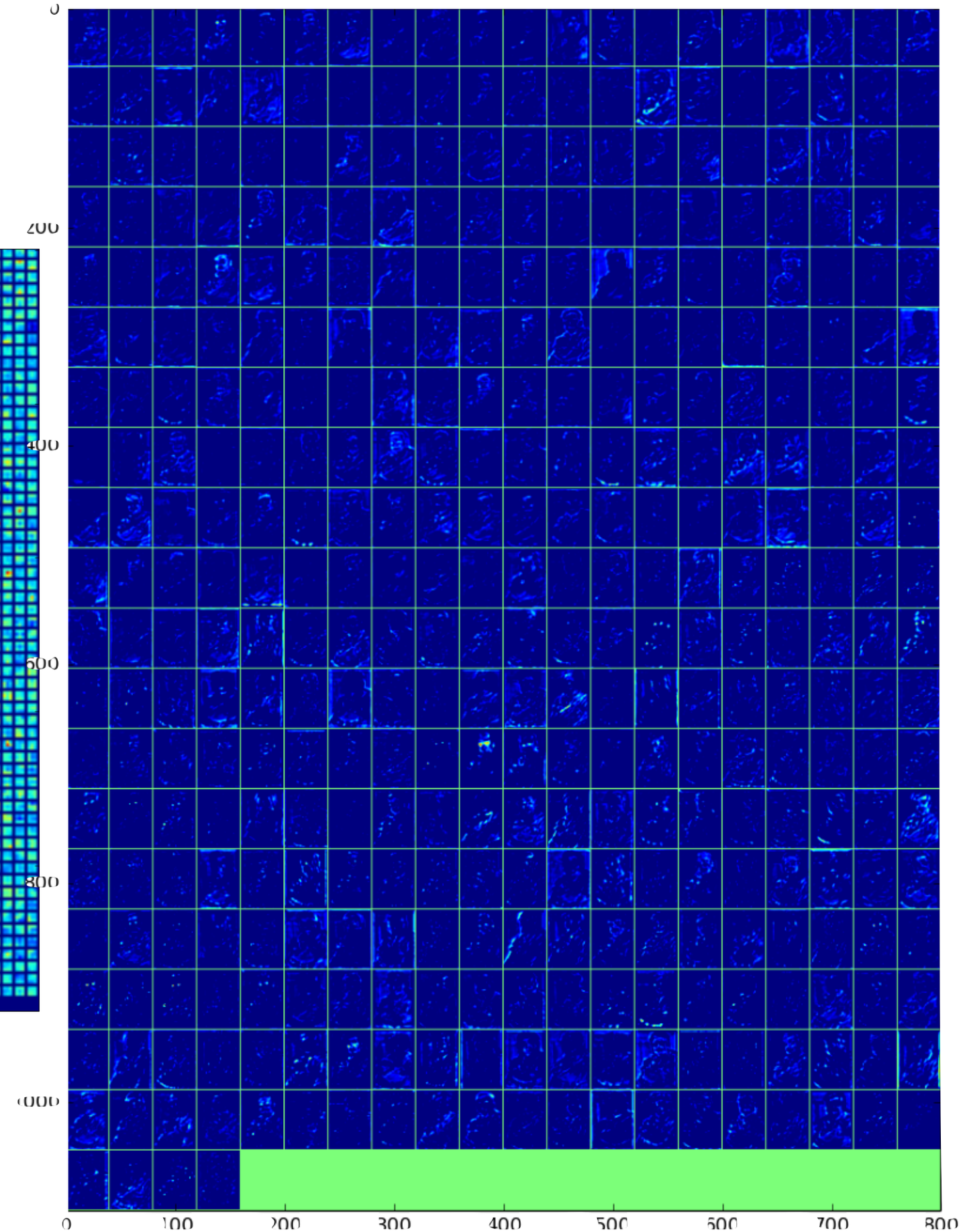
3 x 3



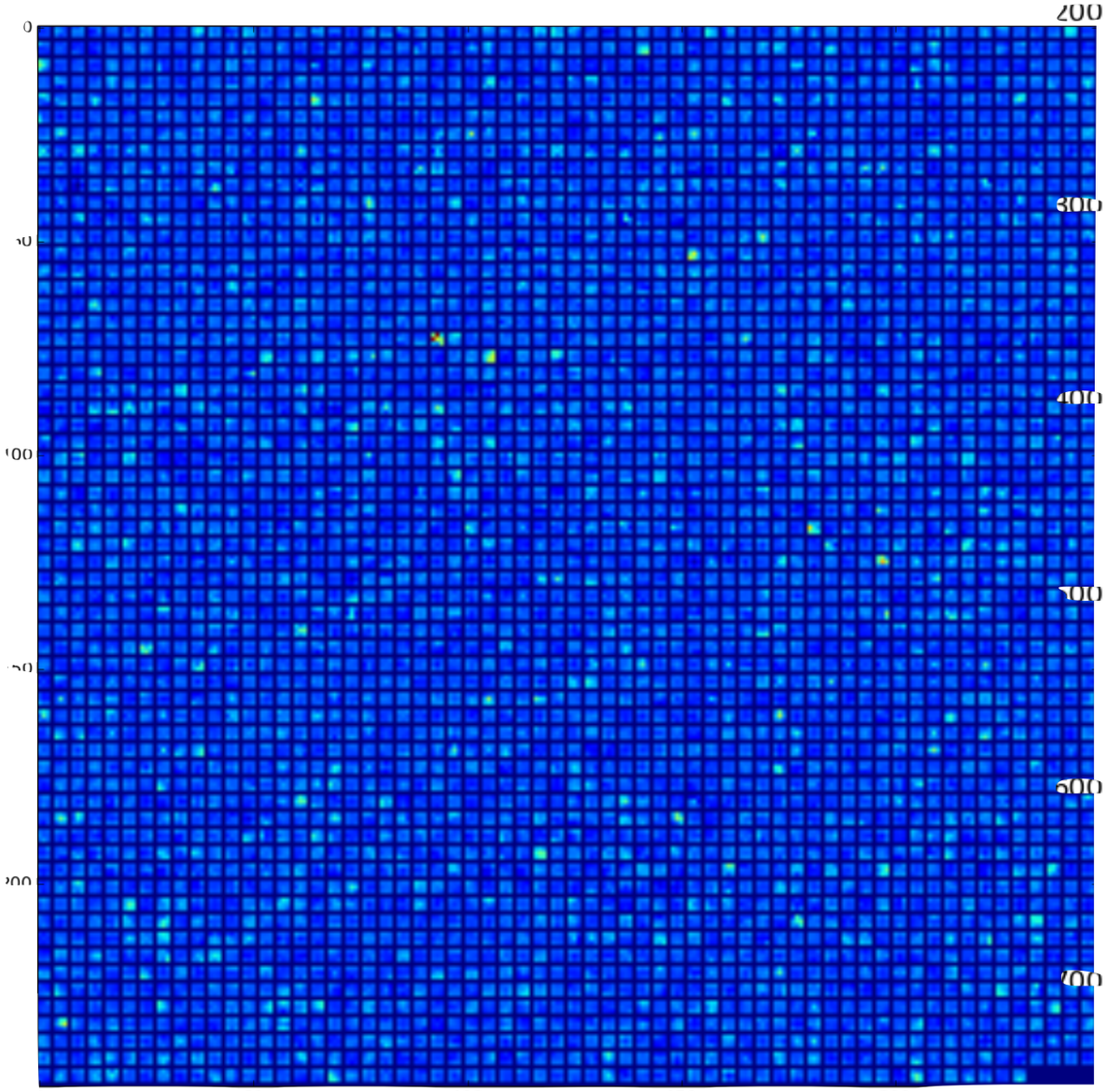
Conv4


0
50
100
150
200

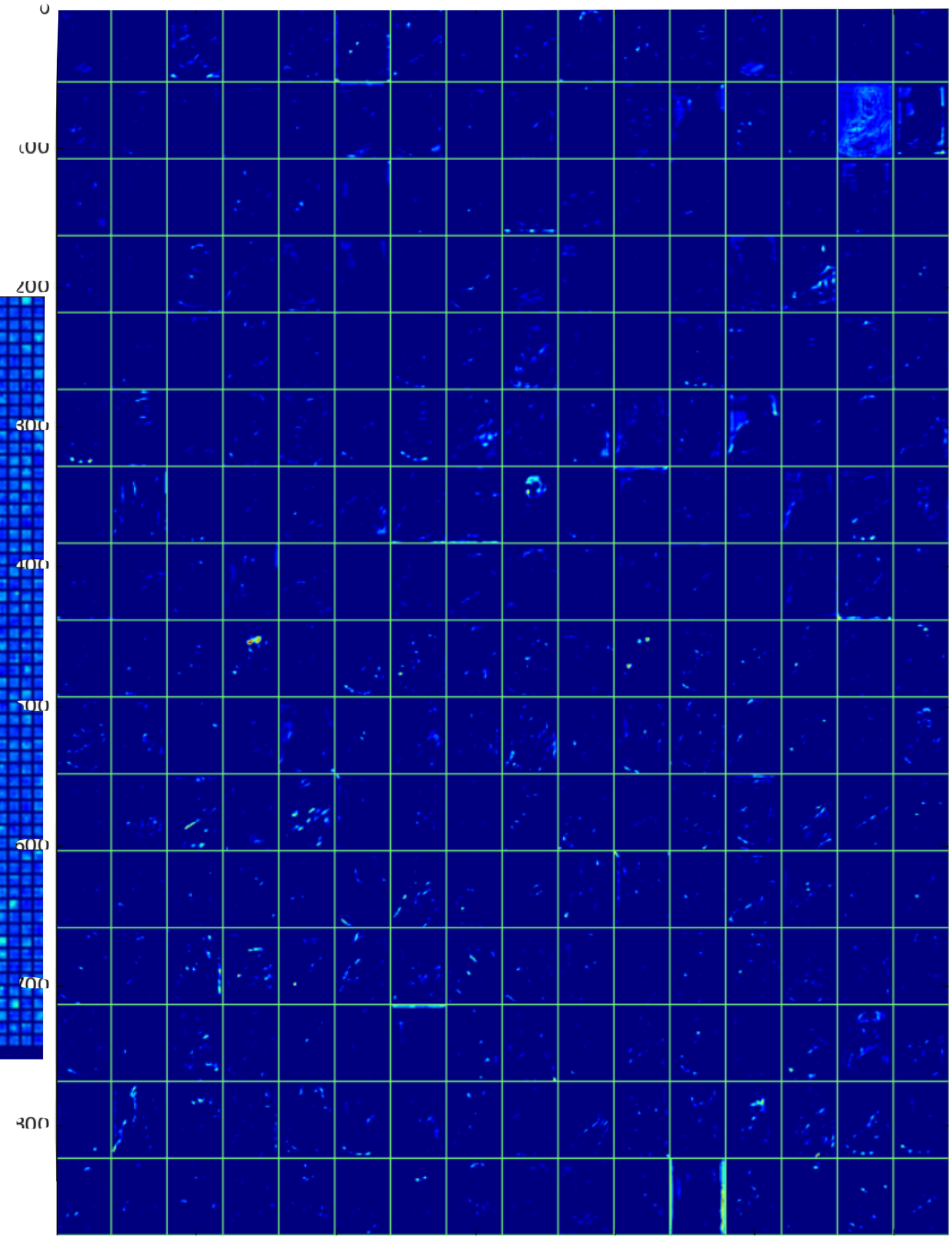
3 x 3


0
200
400
600
800
1000
0
100
200
300
400
500
600
700
800

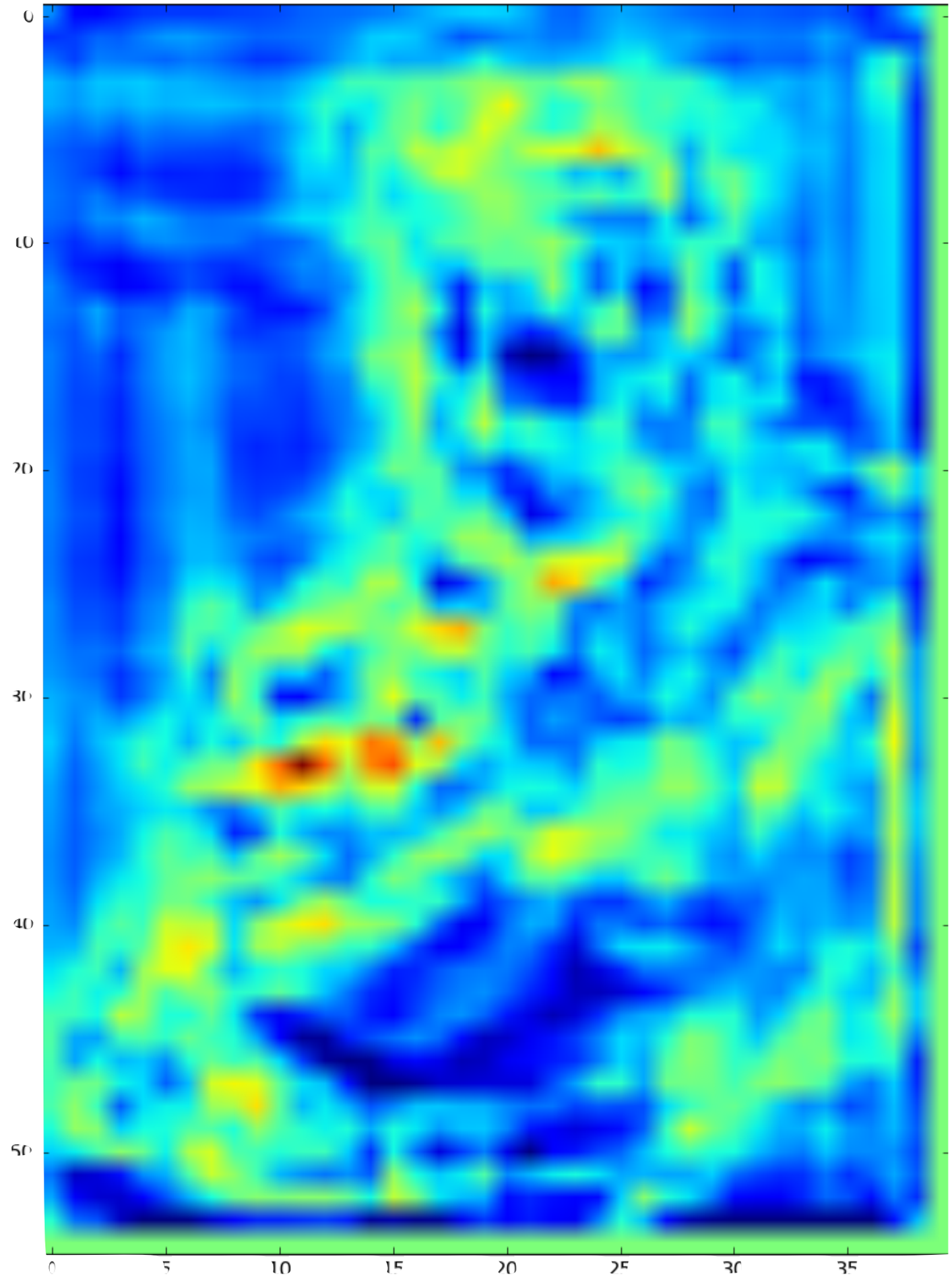
Conv5



3 x 3



Conv5

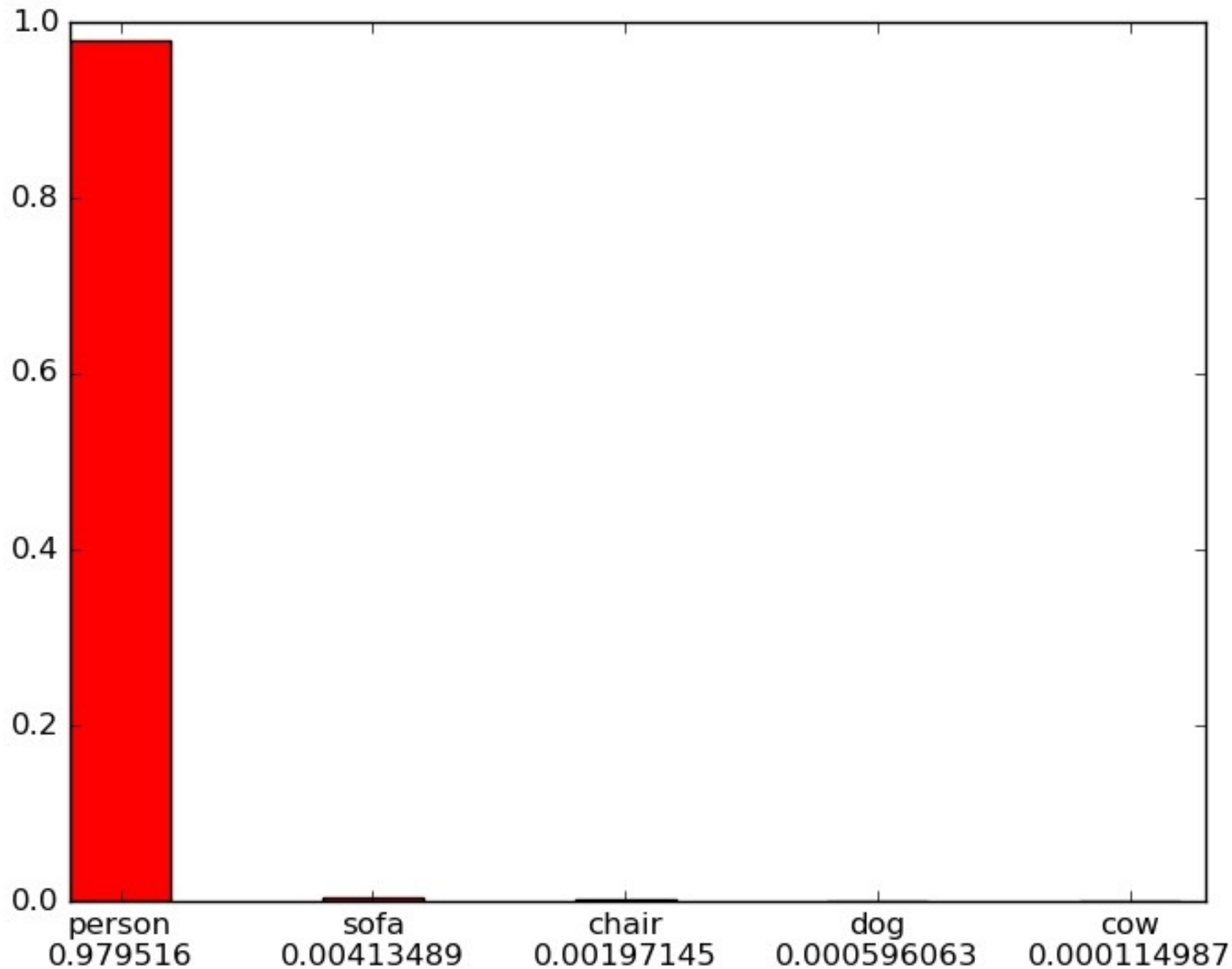


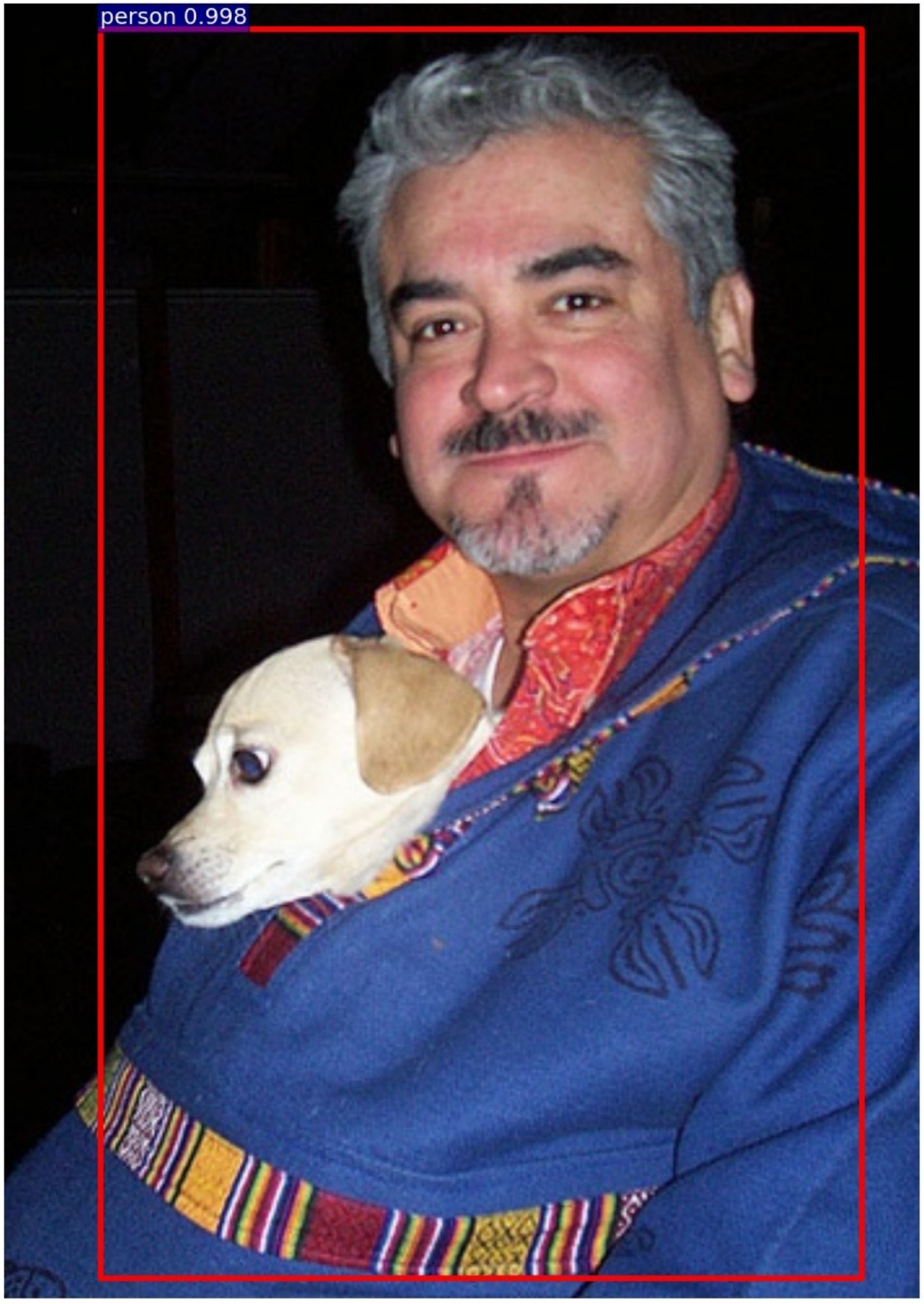
Running time

- CPU mode
- Intel Core i7-3770 @ 3.40 GHz (4 cores)
- CaffeNet
  - Pre-computed bounding boxes: ~8s / image
  - Single image level bounding box: ~1s / image
- VGG16 pre-computed: ~35s / image

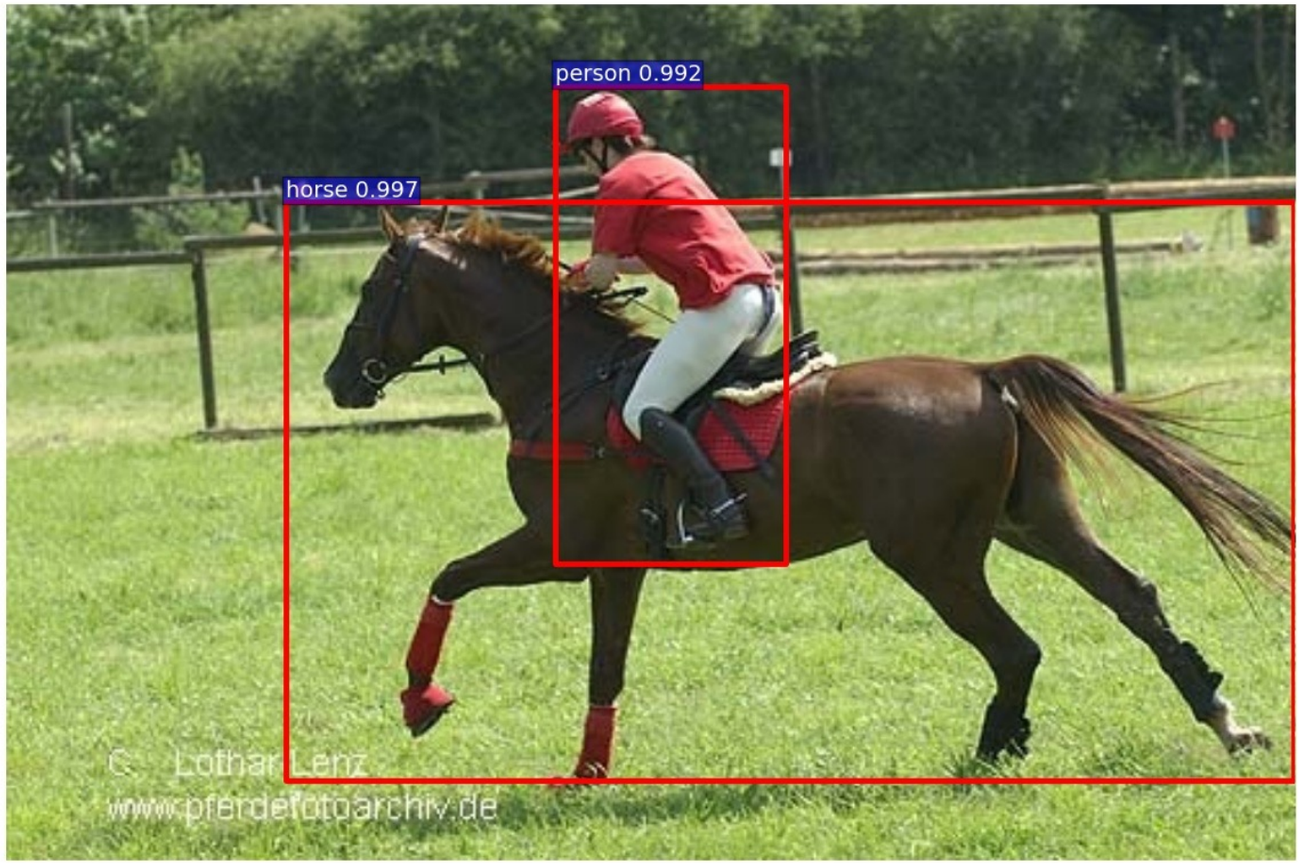
Image level detection and classification

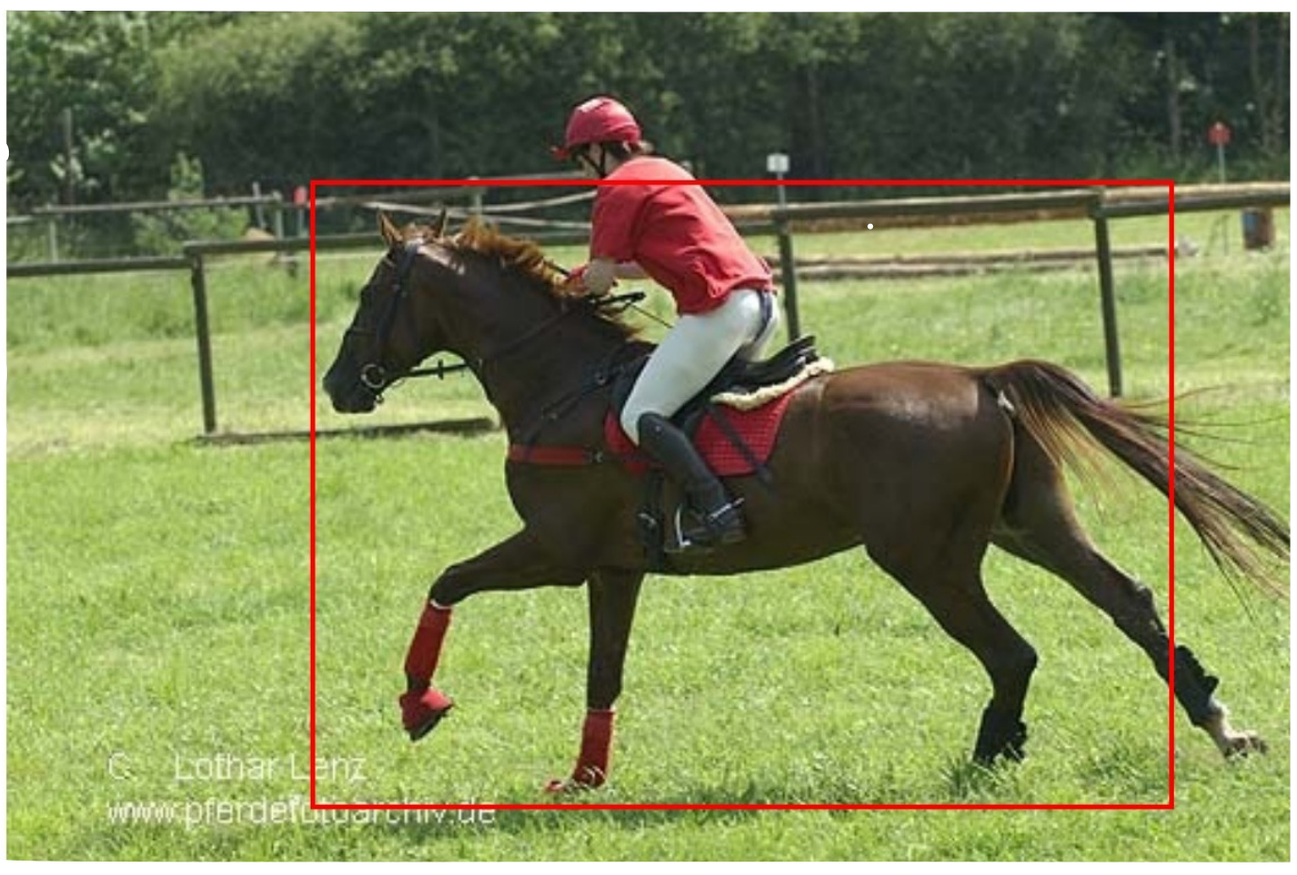
- No region proposals
- Input: 1 bounding box of the entire image

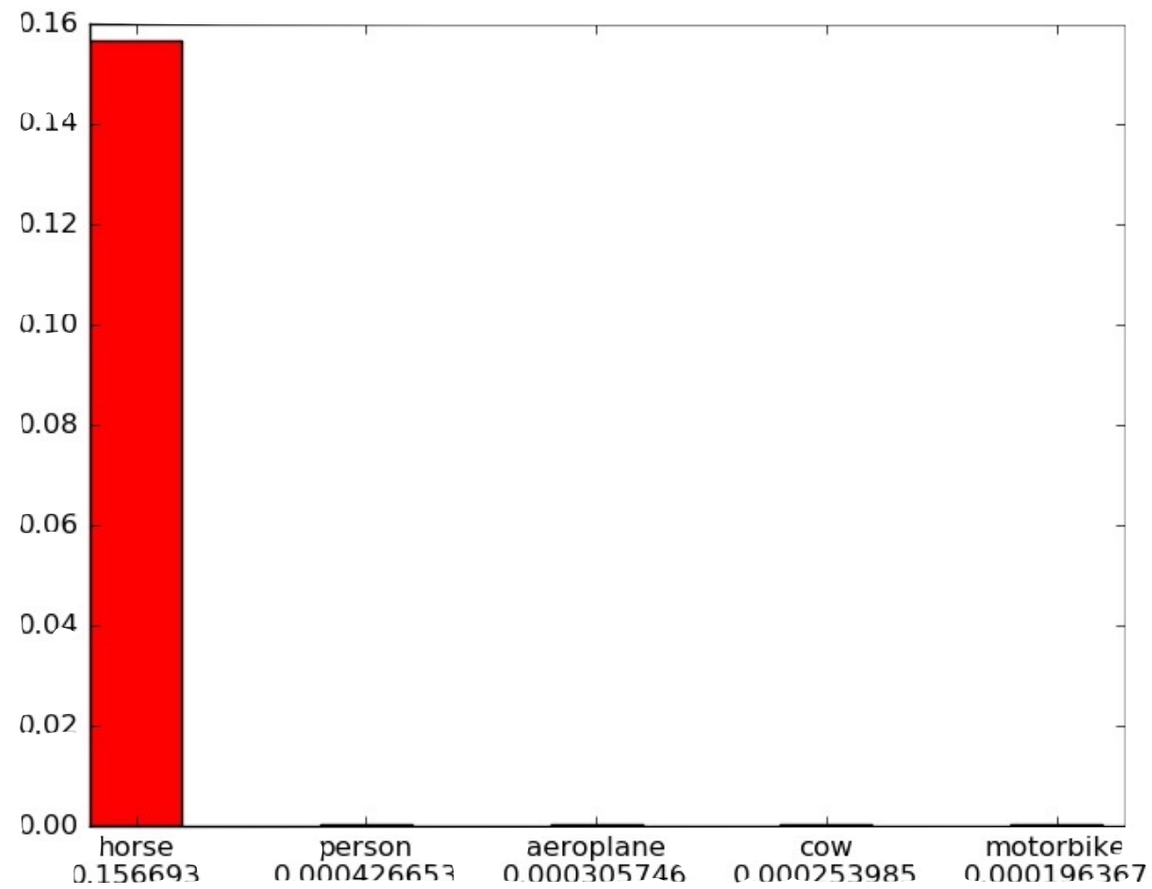

1.0
0.8
0.6
0.4
0.2
0.0
person
0.979516
sofa
0.00413489
chair
0.00197145
dog
0.000596063
cow
0.000114987


person 0.998

PASCAL

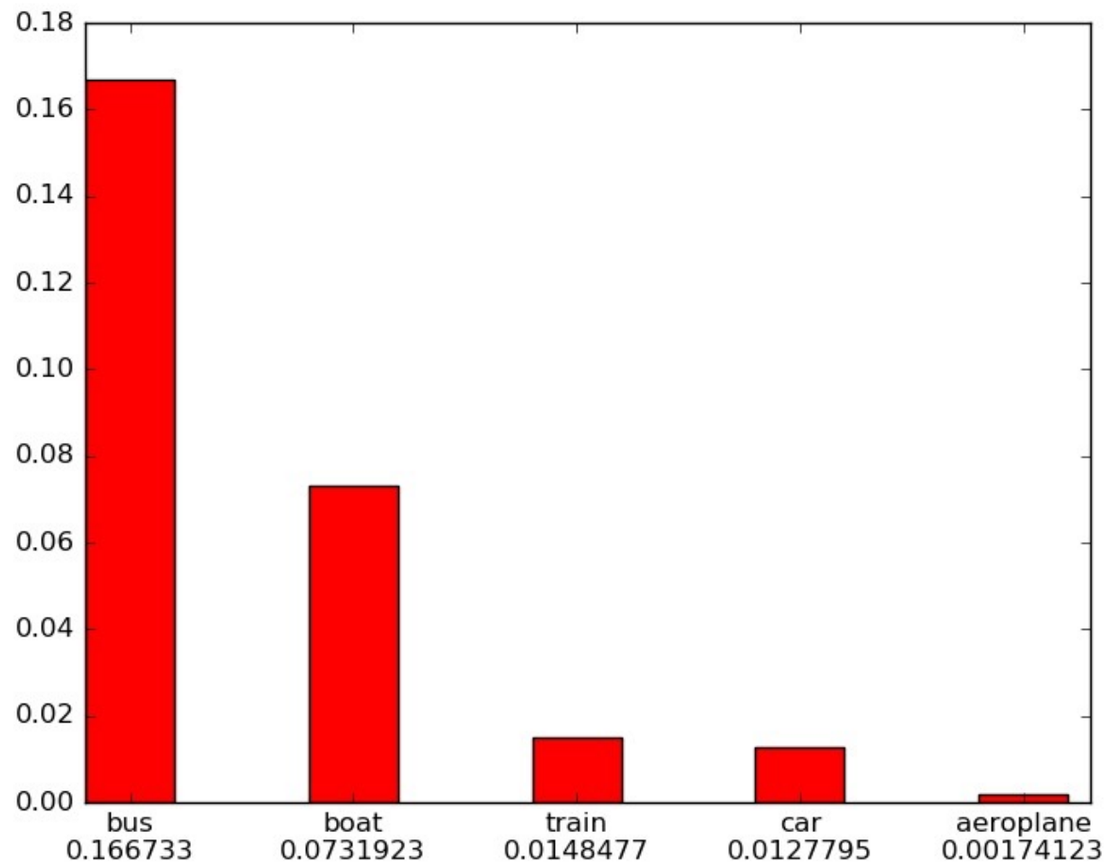

person 0.992
horse 0.997
© Lothar Lenz
www.pferdefotoarchiv.de


© Lothar Lenz
www.pferdefotoarchiv.de

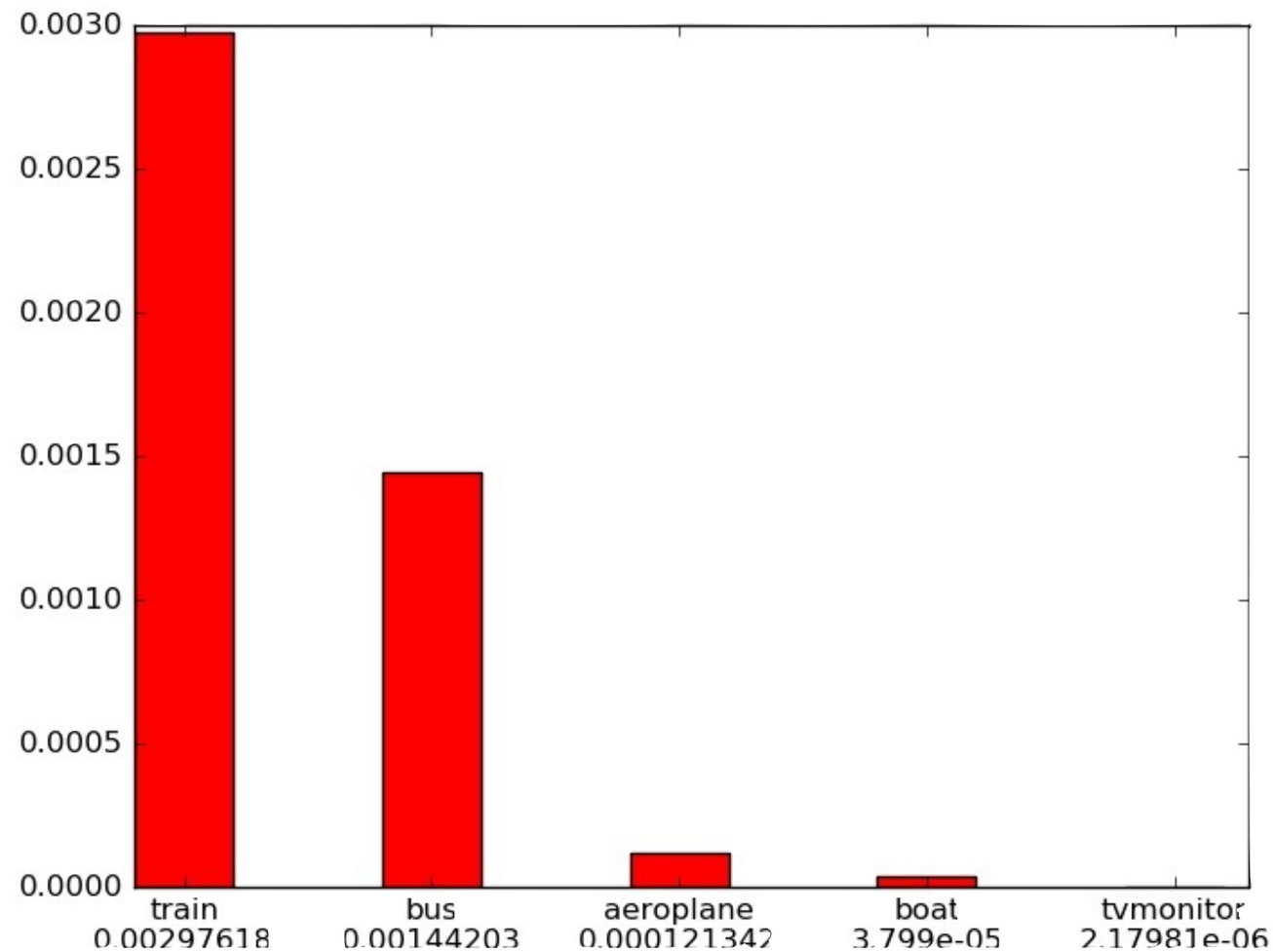

0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
horse
0.156693
person
0.000426653
aeroplane
0.000305746
cow
0.000253985
motorbike
0.000196367

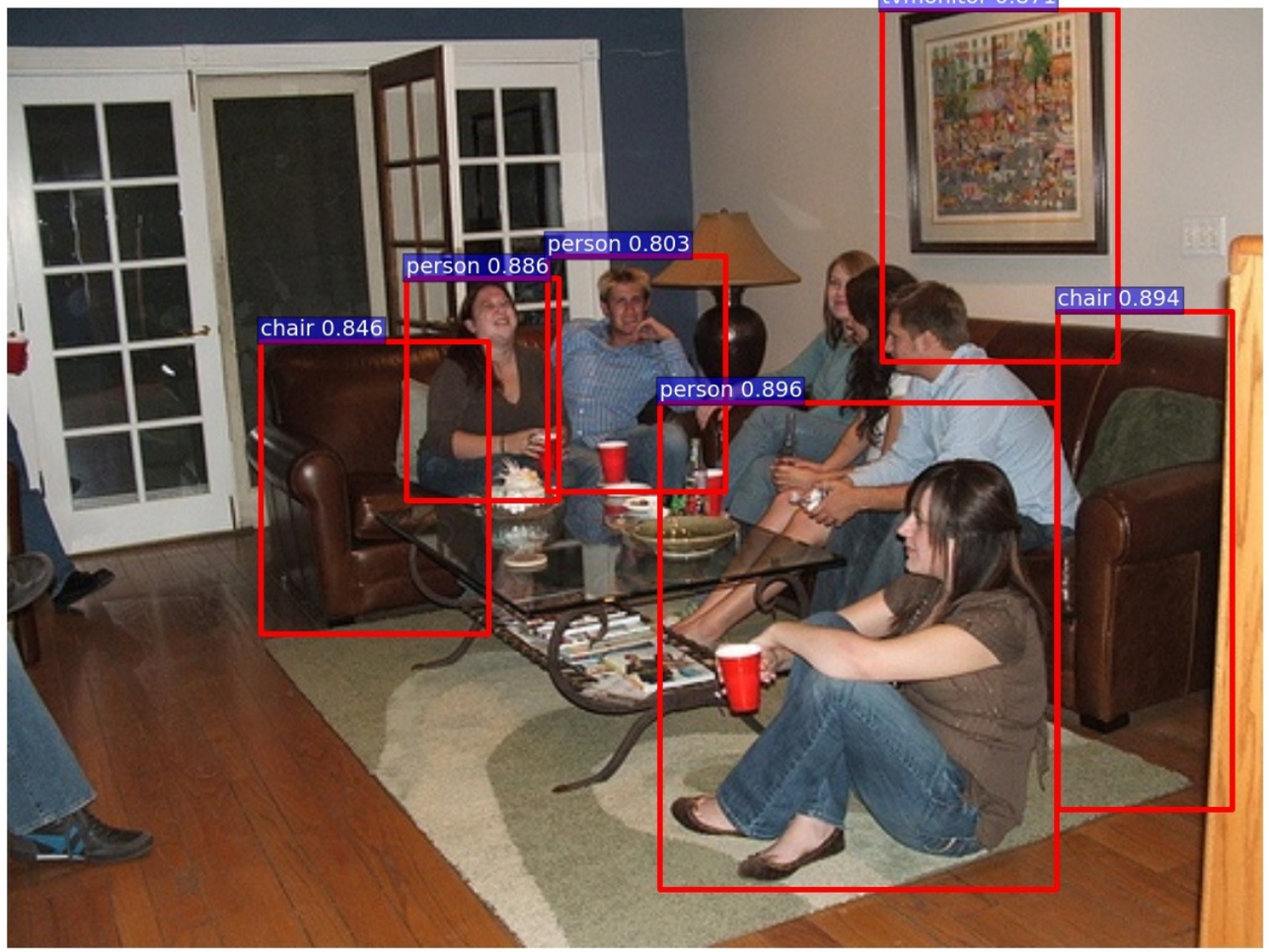
bus 0.992
CAUTION
E40 SBO

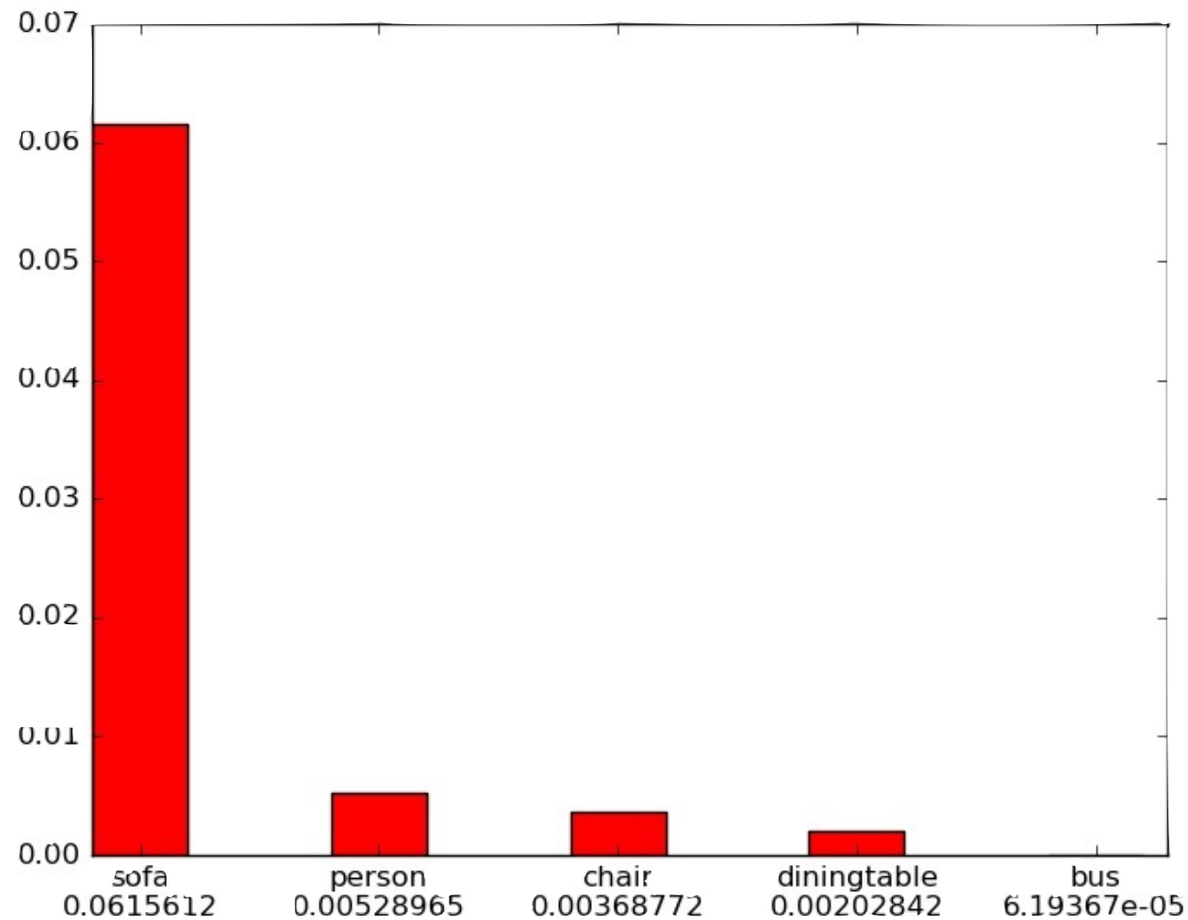
CAUTION
E40 SBO


0.18
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
bus
0.166733
boat
0.0731923
train
0.0148477
car
0.0127795
aeroplane
0.00174123

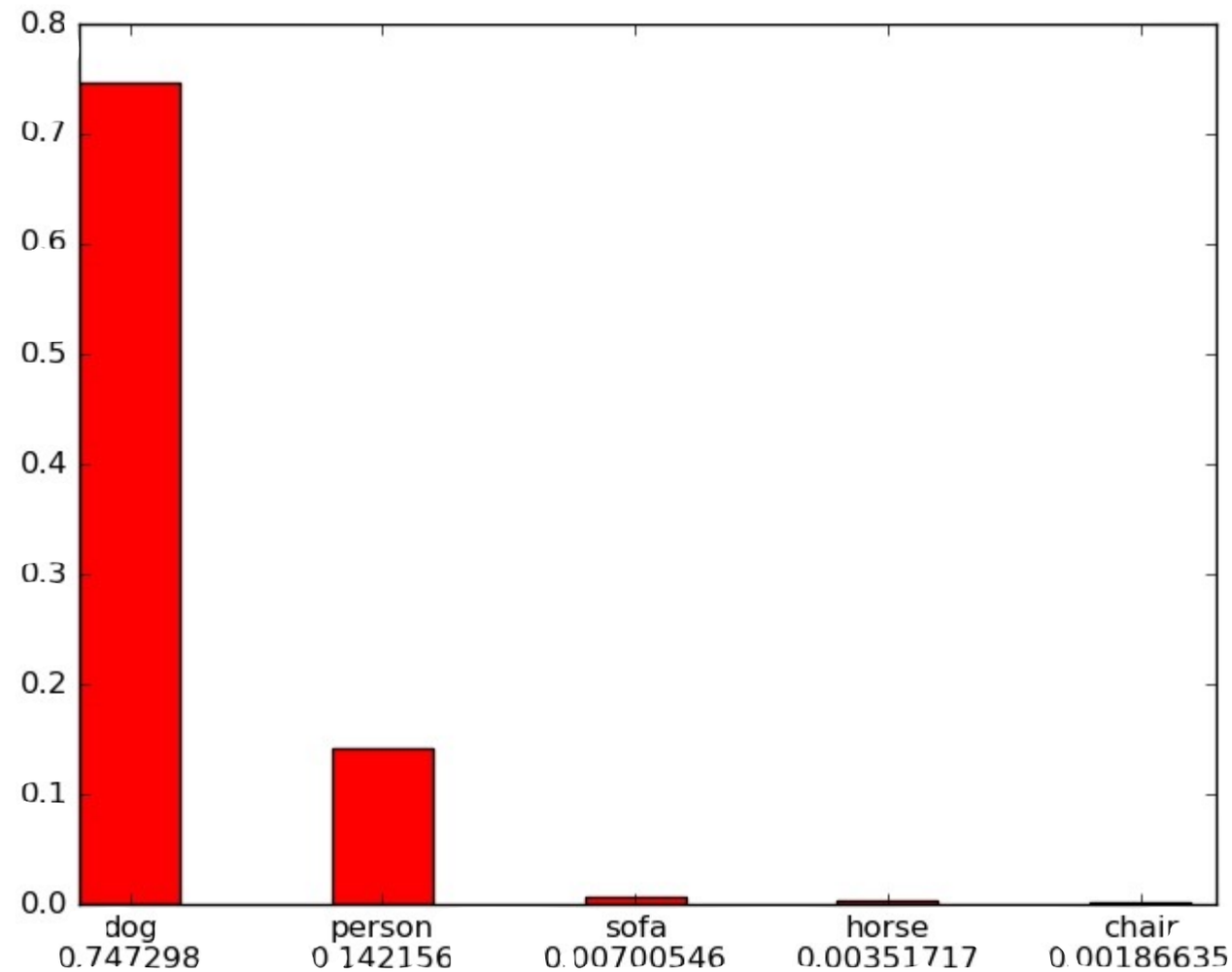
car 0.988
car 0.929
car 0.885
car 0.994
car 0.985

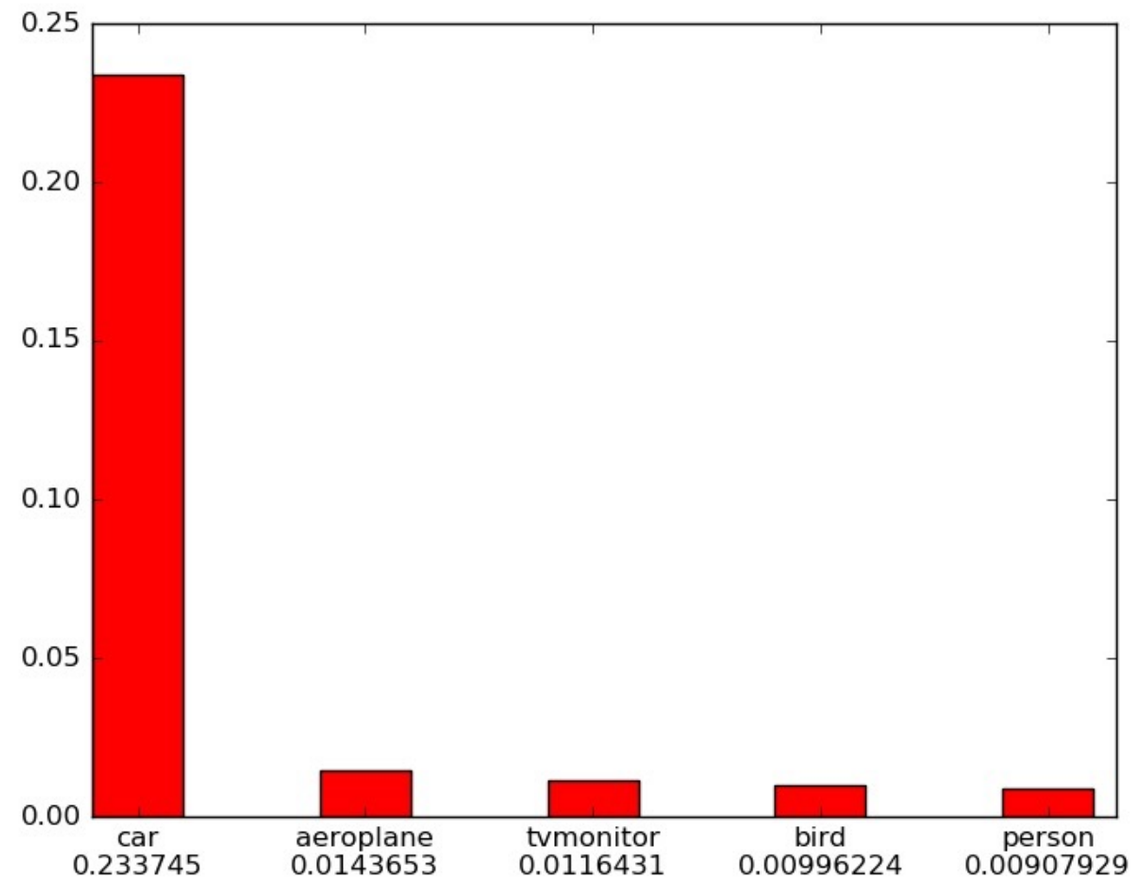

0.0030
0.0025
0.0020
0.0015
0.0010
0.0005
0.0000
train
0.00297618
bus
0.00144203
aeroplane
0.000121342
boat
3.799e-05
tvmonitor
2.17981e-06

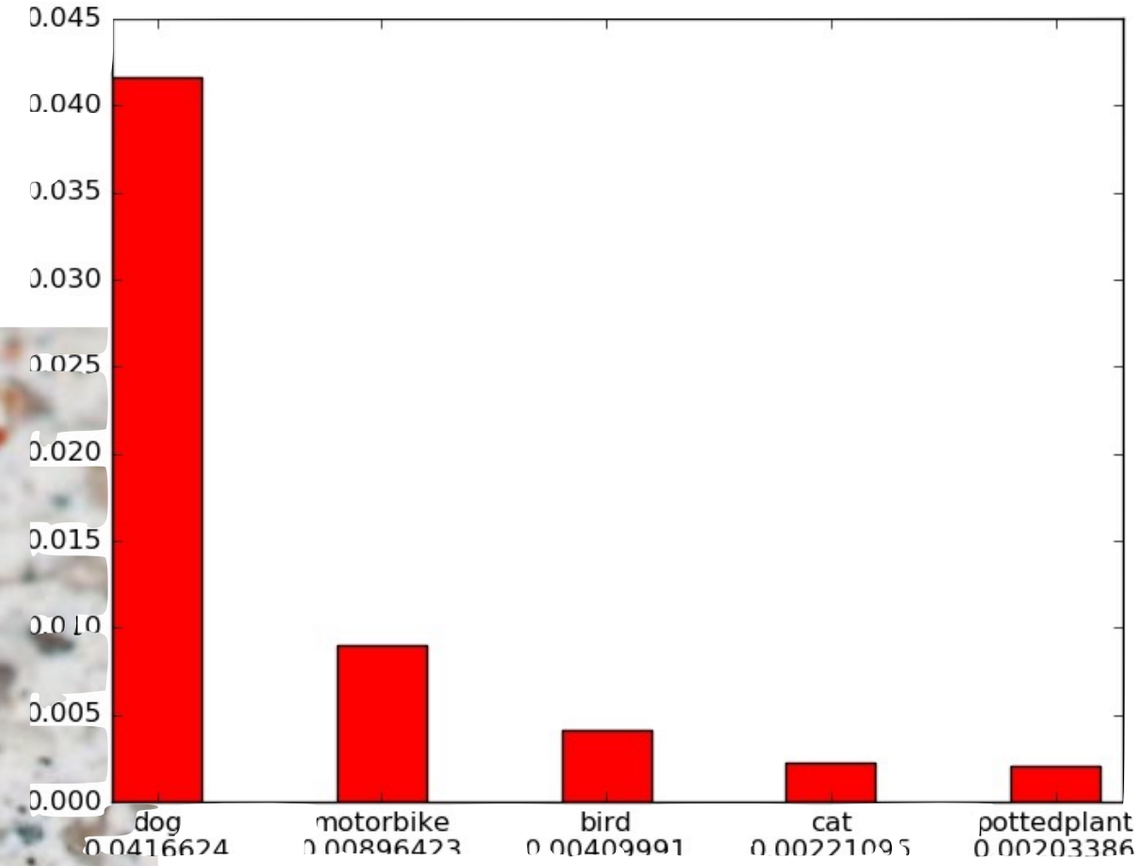

person 0.803
person 0.886
chair 0.846
chair 0.894
person 0.896

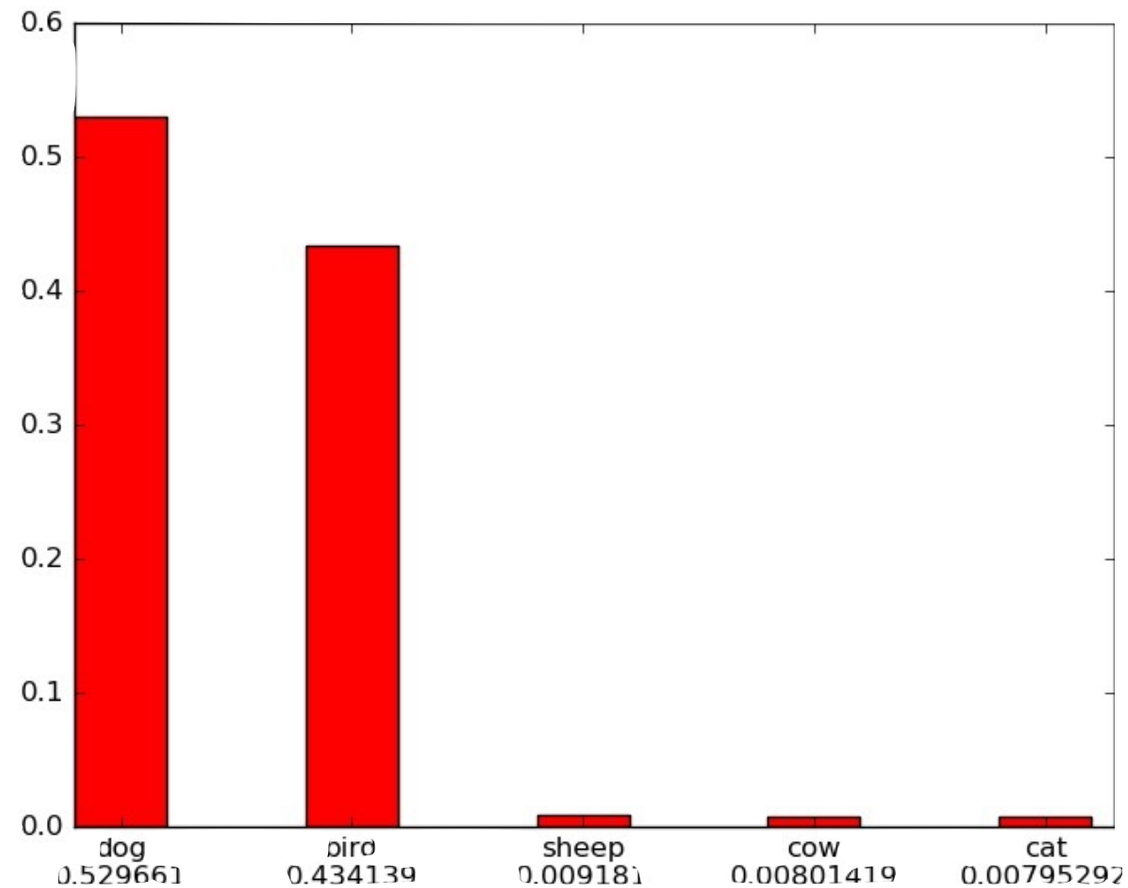

0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
sofa
0.0615612
person
0.00528965
chair
0.00368772
diningtable
0.00202842
bus
6.19367e-05

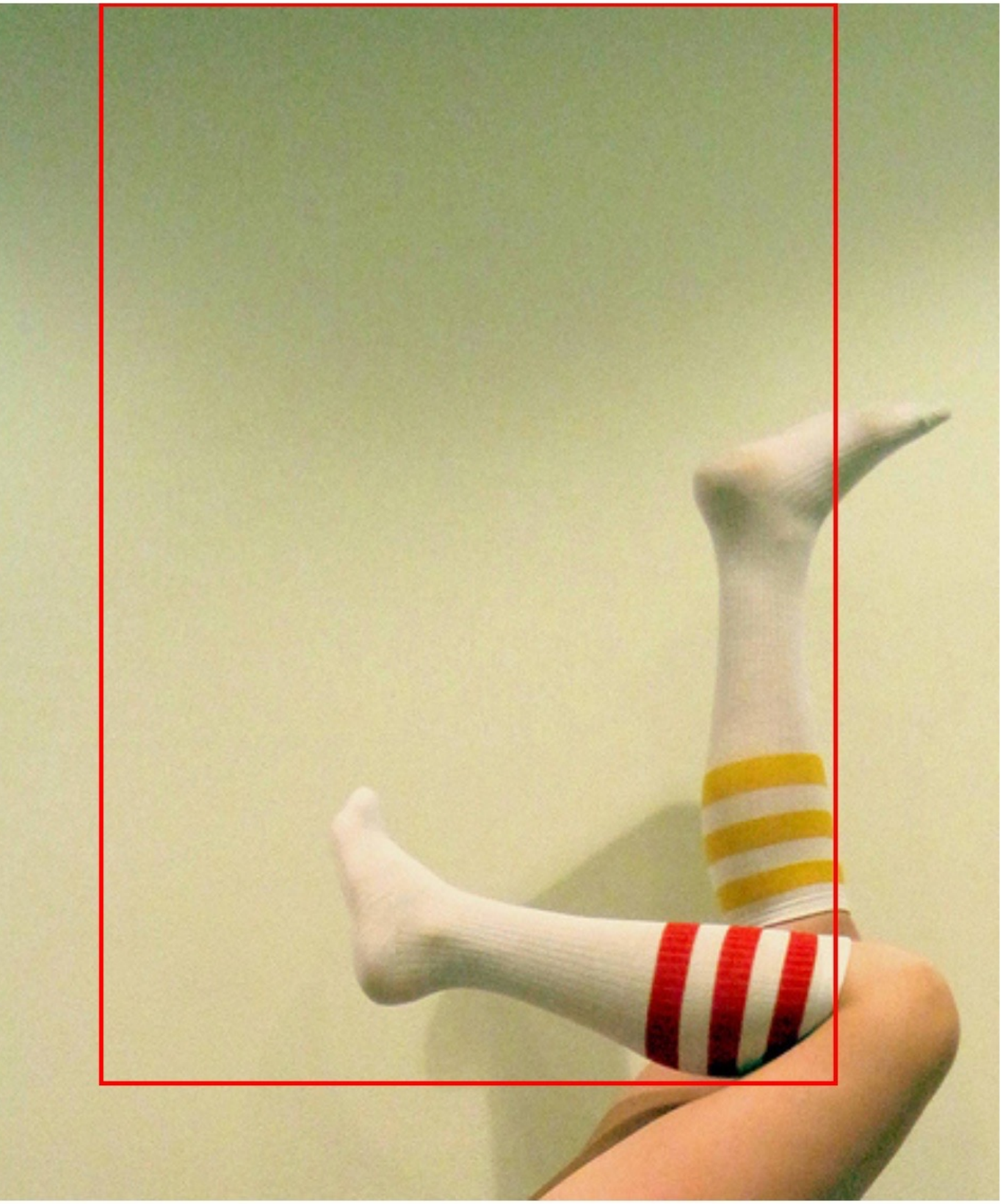
Imagenet


0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
dog
0.747298
person
0.142156
sofa
0.00700546
horse
0.00351717
chair
0.00186635


0.25
0.20
0.15
0.10
0.05
0.00
car
0.233745
aeroplane
0.0143653
tvmonitor
0.0116431
bird
0.00996224
person
0.00907929


0.045
0.040
0.035
0.030
0.025
0.020
0.015
0.010
0.005
0.000
dog
0.0416624
motorbike
0.00896423
bird
0.00409991
cat
0.00221095
pottedplant
0.00203386


0.6
0.5
0.4
0.3
0.2
0.1
0.0
dog
0.529661
bird
0.434139
sheep
0.009181
cow
0.00801419
cat
0.00795292



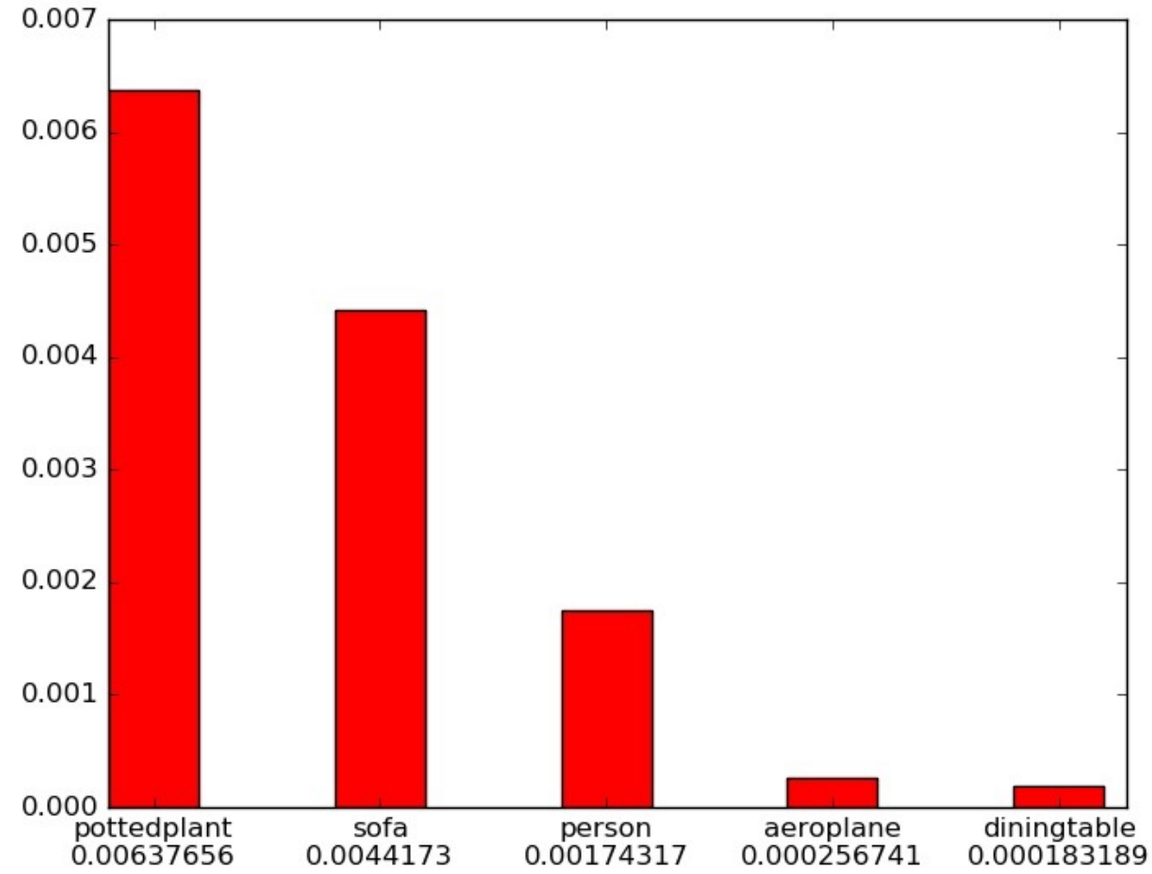

0.007
0.006
0.005
0.004
0.003
0.002
0.001
0.000
pottedplant
0.00637656
sofa
0.0044173
person
0.00174317
aeroplane
0.000256741
diningtable
0.000183189

Image classification accuracy

- Imagenet data, 100 images per class

	car	bottle	chair	tv	plant	person	cat
Sample data accuracy	87	45	19	87	76	72	69
VOC 07 with detection AP	74.2	36.5	34.4	64.8	33.4	58.7	67.6

Takeaway

- Works for image level classification
- Detection works without region proposal
- Class independent detection
- Detection is only as good as the classification

Questions?