Multi-Task learning

Mateusz Biegański
Kamil Bladoszewski
Piotr Gierda
Paweł Opiela

Who are we?

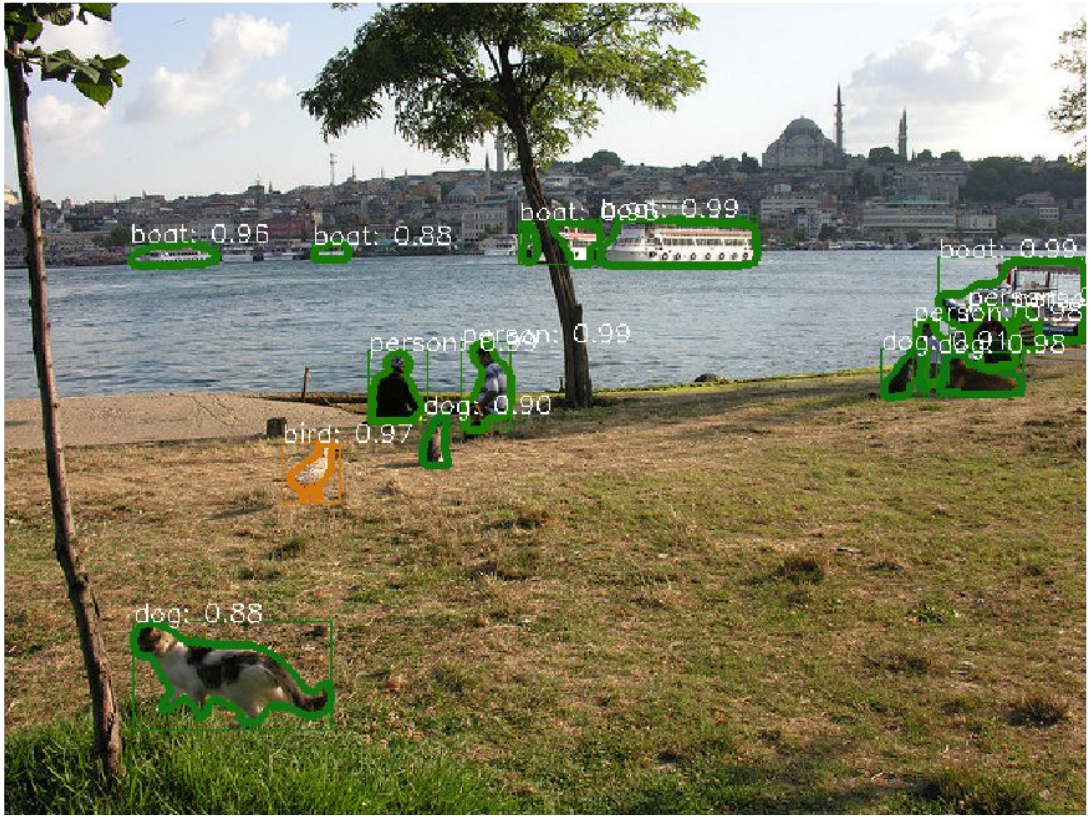
- Mateusz Biegański
- Kamil Bladoszewski
- Piotr Gierda
- Paweł Opiela

What's the fuss all about?

Multi-Task learning!

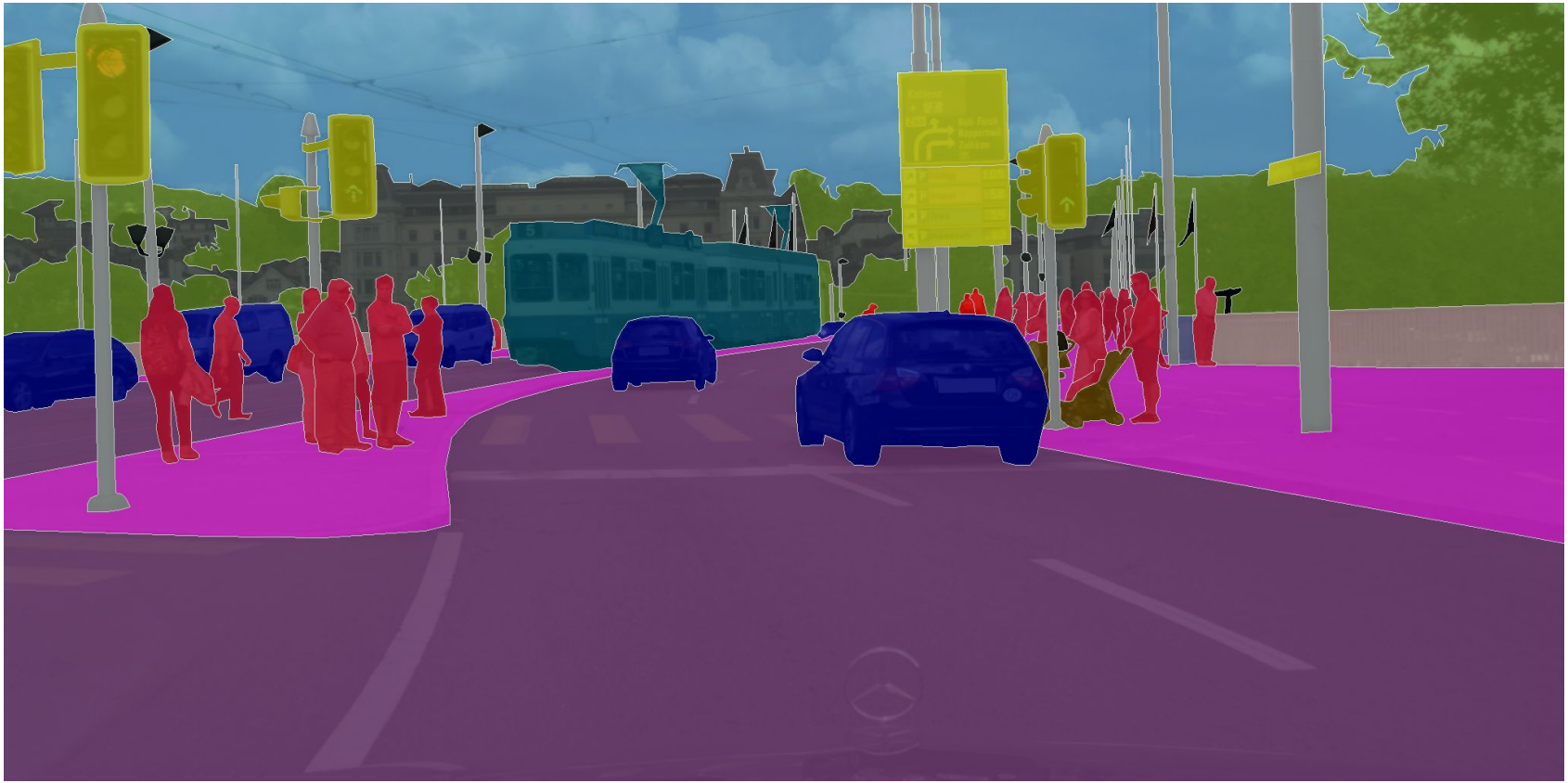
- One neural network performing multiple tasks
- Trained once, inference on whichever trained task
- Exploiting commonalities across tasks

Detection

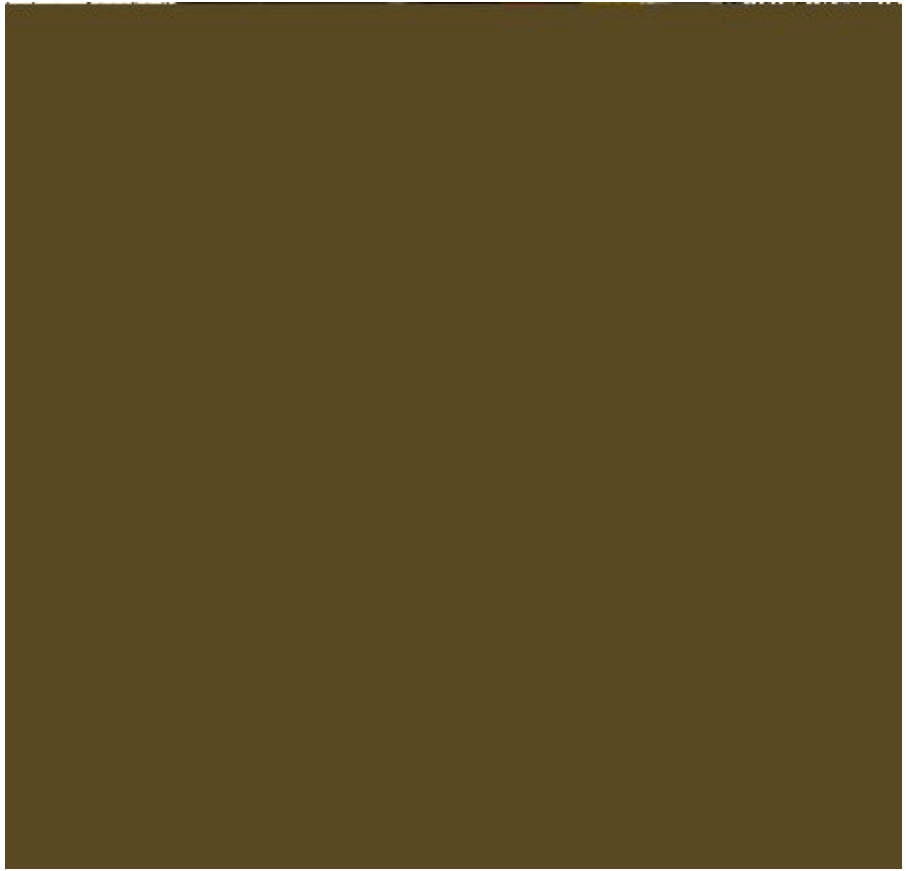

boat: 0.96
boat: 0.88
boat: 0.99
boat: 0.99
person: 0.99
person: 0.99
dog: 0.90
bird: 0.97
dog: 0.88

Human Keypoints Estimation

Semantic Segmentation



Instance Segmentation



Why multitask is important?

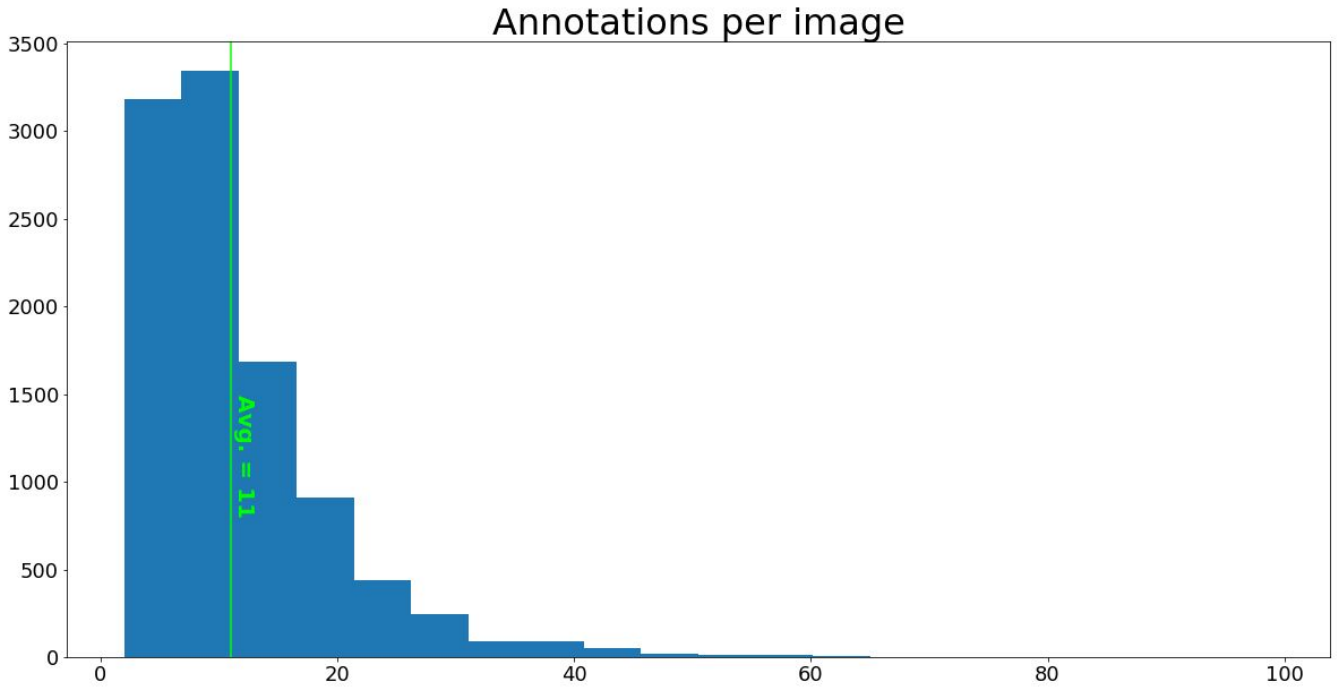
Source: https://nomagic.ai/

Pascal in Detail dataset

- Popular for multitask research
- Released for Pascal in Detail Challenge in July 2017
- Contains wide range of annotations:

  - detection
  - image classification
  - instance segmentation
  - semantic segmentation
  - human parts segmentation
  - human keypoints estimation
  - occlusion recognition
  - boundary detection

114425 Annotations



Guess the number of annotations!

Guess the number of annotations!

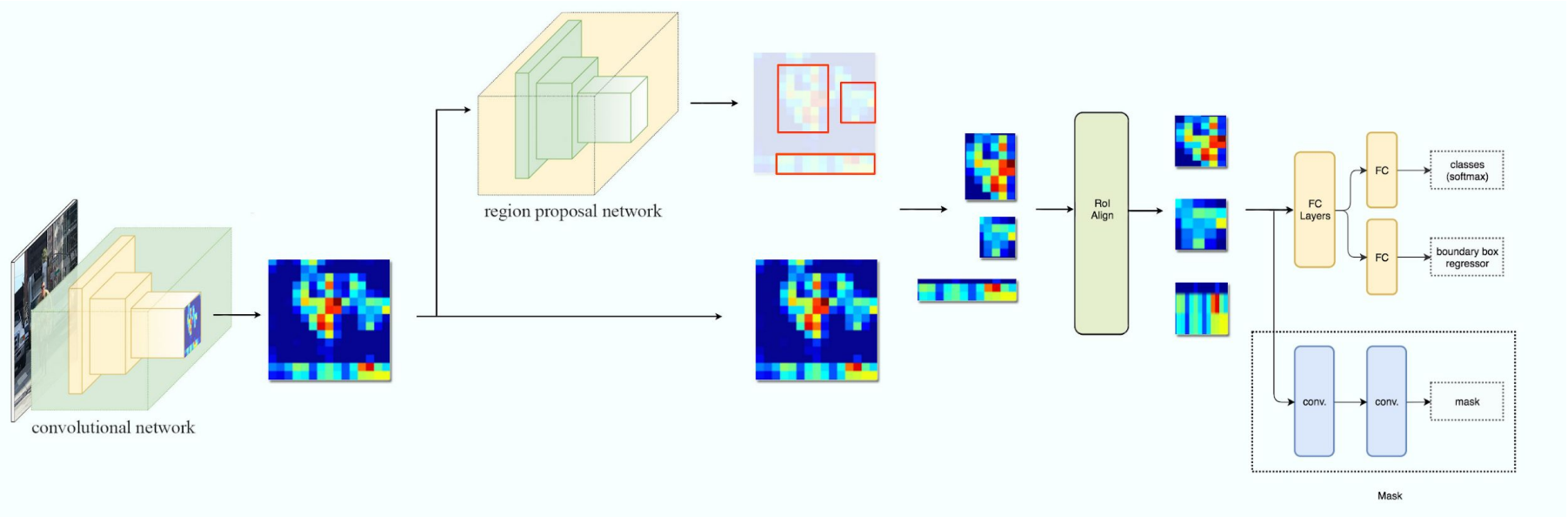
Architecture

Mask R-CNN

- Paper “Mask R-CNN” from Kaiming He et al. from March 2017
- Computer vision network (image processing)
- Facebook Research Pytorch 1.0 implementation
- Support for COCO dataset

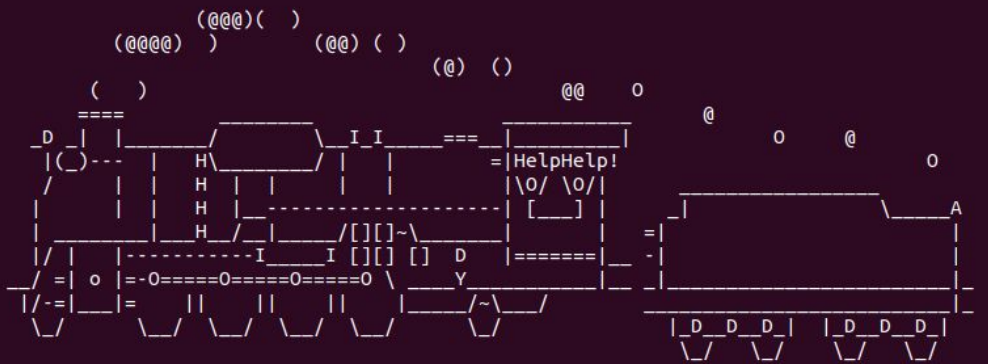
COCO dataset

- Annotations for detection, instance segmentation and keypoint estimation
- That’s not enough for us! (or so it seemed to us…)
- Thus we need another, better labeled dataset


region proposal network
convolutional network
RoI Align
FC Layers
FC
classes (softmax)
FC
boundary box regressor
conv.
conv.
mask
Mask

Deploy

```
kb392558@sylvester:~/maskrcnn-benchmark$ make train
```




```
zppmt@bozena:~/maskrcnn-benchmark$ make train
```

```
zppmt@kasia:~/maskrcnn-benchmark$ make train
```

Training

Our battlefield

- Mask encoding differences (polygons, RLE)
- Mask R-CNN does not support multitask itself
- Great effort to filter annotations
- Evaluation adjustment struggle

- detection
- image classification
- instance segmentation
- semantic segmentation
- human keypoints estimation
- occlusion recognition
- boundary detection

```
DONE (t=1.08s).
 Average Precision  (AP) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.068
 Average Precision  (AP) @[ IoU=0.50      | area=   all | maxDets=100 ] = 0.125
 Average Precision  (AP) @[ IoU=0.75      | area=   all | maxDets=100 ] = 0.068
 Average Precision  (AP) @[ IoU=0.50:0.95 | area= small | maxDets=100 ] = 0.094
 Average Precision  (AP) @[ IoU=0.50:0.95 | area=medium | maxDets=100 ] = 0.150
 Average Precision  (AP) @[ IoU=0.50:0.95 | area= large | maxDets=100 ] = 0.009
 Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=  1 ] = 0.070
 Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets= 10 ] = 0.107
 Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.107
 Average Recall     (AR) @[ IoU=0.50:0.95 | area= small | maxDets=100 ] = 0.128
 Average Recall     (AR) @[ IoU=0.50:0.95 | area=medium | maxDets=100 ] = 0.239
 Average Recall     (AR) @[ IoU=0.50:0.95 | area= large | maxDets=100 ] = 0.022
loading annotations into memory...
Done (t=3.42s)
creating index...
index created!
Loading and preparing results...
DONE (t=0.23s)
creating index...
index created!
Running per image evaluation...
Evaluate annotation type *segm*
DONE (t=7.73s).
Accumulating evaluation results...
DONE (t=1.10s).
 Average Precision  (AP) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.056
 Average Precision  (AP) @[ IoU=0.50      | area=   all | maxDets=100 ] = 0.107
 Average Precision  (AP) @[ IoU=0.75      | area=   all | maxDets=100 ] = 0.054
 Average Precision  (AP) @[ IoU=0.50:0.95 | area= small | maxDets=100 ] = 0.068
 Average Precision  (AP) @[ IoU=0.50:0.95 | area=medium | maxDets=100 ] = 0.116
 Average Precision  (AP) @[ IoU=0.50:0.95 | area= large | maxDets=100 ] = 0.006
 Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=  1 ] = 0.056
 Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets= 10 ] = 0.087
 Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.088
 Average Recall     (AR) @[ IoU=0.50:0.95 | area= small | maxDets=100 ] = 0.117
 Average Recall     (AR) @[ IoU=0.50:0.95 | area=medium | maxDets=100 ] = 0.195
 Average Recall     (AR) @[ IoU=0.50:0.95 | area= large | maxDets=100 ] = 0.004
loading annotations into memory...
```


boat: 0.97
person: 0.98
cow: 0.76

What went wrong?

- Working with code without documentation was very time consuming
- Really lots of problems, caveats and hacking
- Project was hard itself

And what was great?

- Project was demanding, and computer vision was totally foreign for us
- Research (lots frustration and sometimes gratification)
- Problems with communication

And what was great?

- Project was demanding, and computer vision was totally foreign for us
- Research (lots frustration and sometimes gratification)
- Problems with communication

It was fun!

Funny stuff

Funny stuff


TENSORFLIGHT


python™


PyTorch

Thank you for your attention!