R-CNN

TUNSA

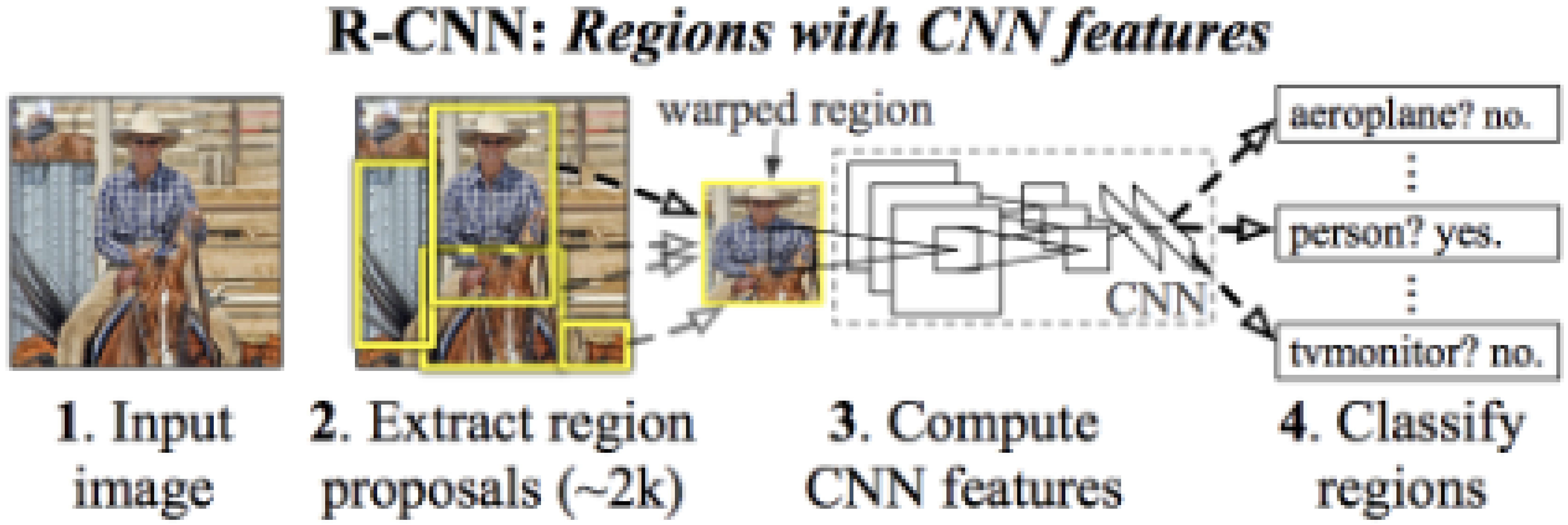
Tunisian Space Association

الجمعية التونسية للفضاء

Object detection / classification

THE DIFFERENCE BETWEEN OBJECT DETECTION ALGORITHMS AND CLASSIFICATION ALGORITHMS IS THAT IN DETECTION ALGORITHMS, WE TRY TO DRAW A BOUNDING BOX AROUND THE OBJECT OF INTEREST TO LOCATE IT WITHIN THE IMAGE

R-CNN


R-CNN: Regions with CNN features
warped region
aeroplane? no.
person? yes.
tvmonitor? no.
CNN
1. Input image
2. Extract region proposals (~2k)
3. Compute CNN features
4. Classify regions

R-CNN

SELECTIVE SEARCH

```
Selective Search:
1. Generate initial sub-segmentation, we generate many candidate
regions
2. Use greedy algorithm to recursively combine similar regions into
larger ones
3. Use the generated regions to produce the final candidate region
proposals
```

1

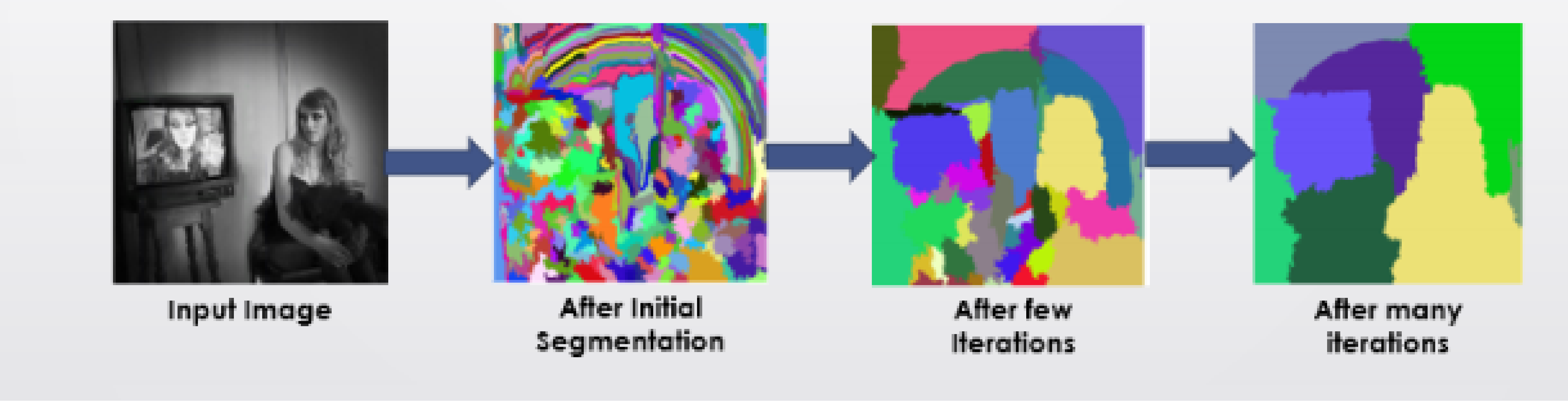
1. Generate initial sub-segmentation of input image using the method describe by *Felzenszwalb et al* in his paper "Efficient Graph-Based Image Segmentation ".

2

Color Similarity
Texture Similarity
Size Similarity
Fill Similarity

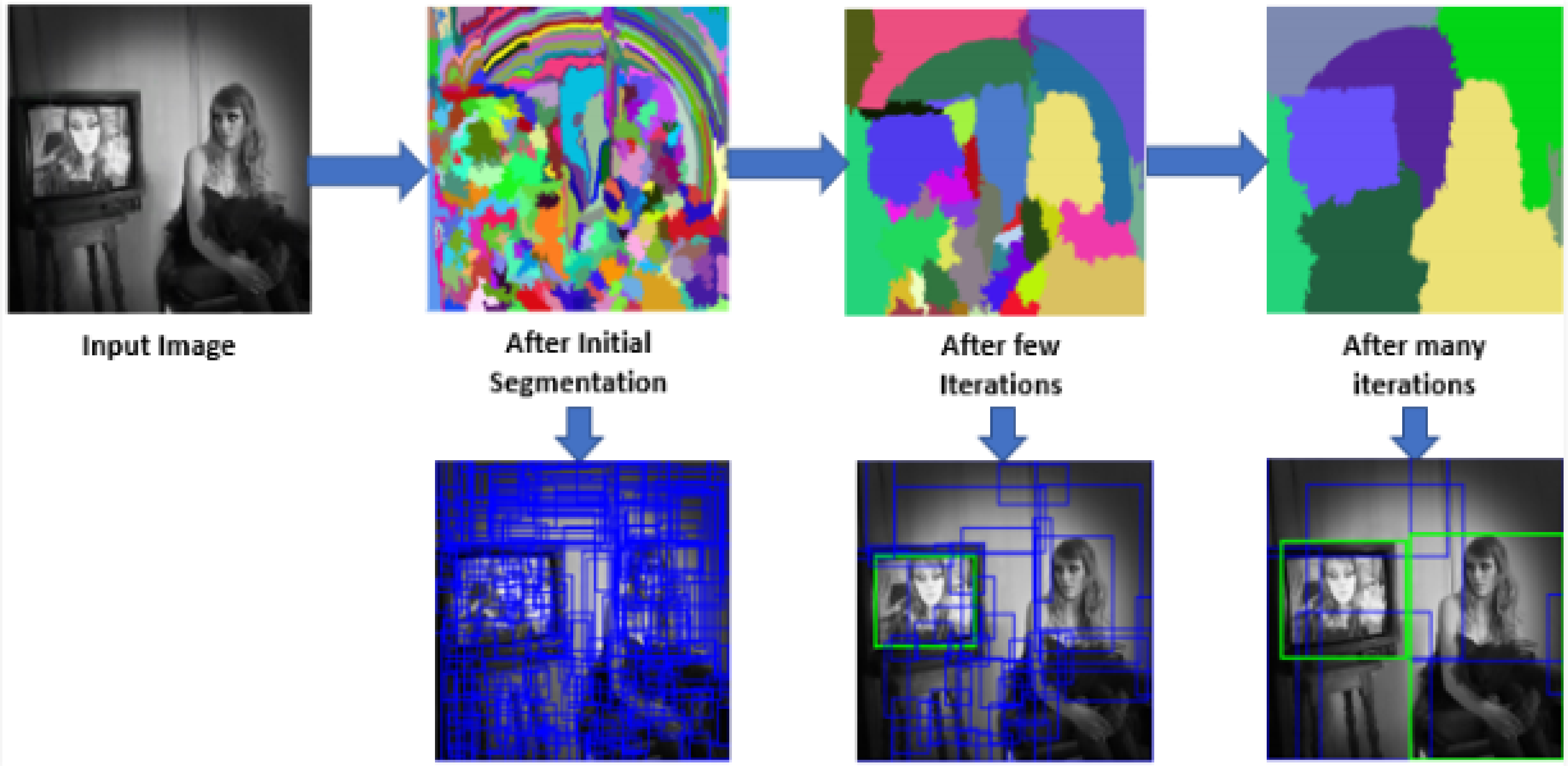
2. Recursively combine the smaller similar regions into larger ones. We use Greedy algorithm to combine similar regions to make larger regions. The algorithm is written below.

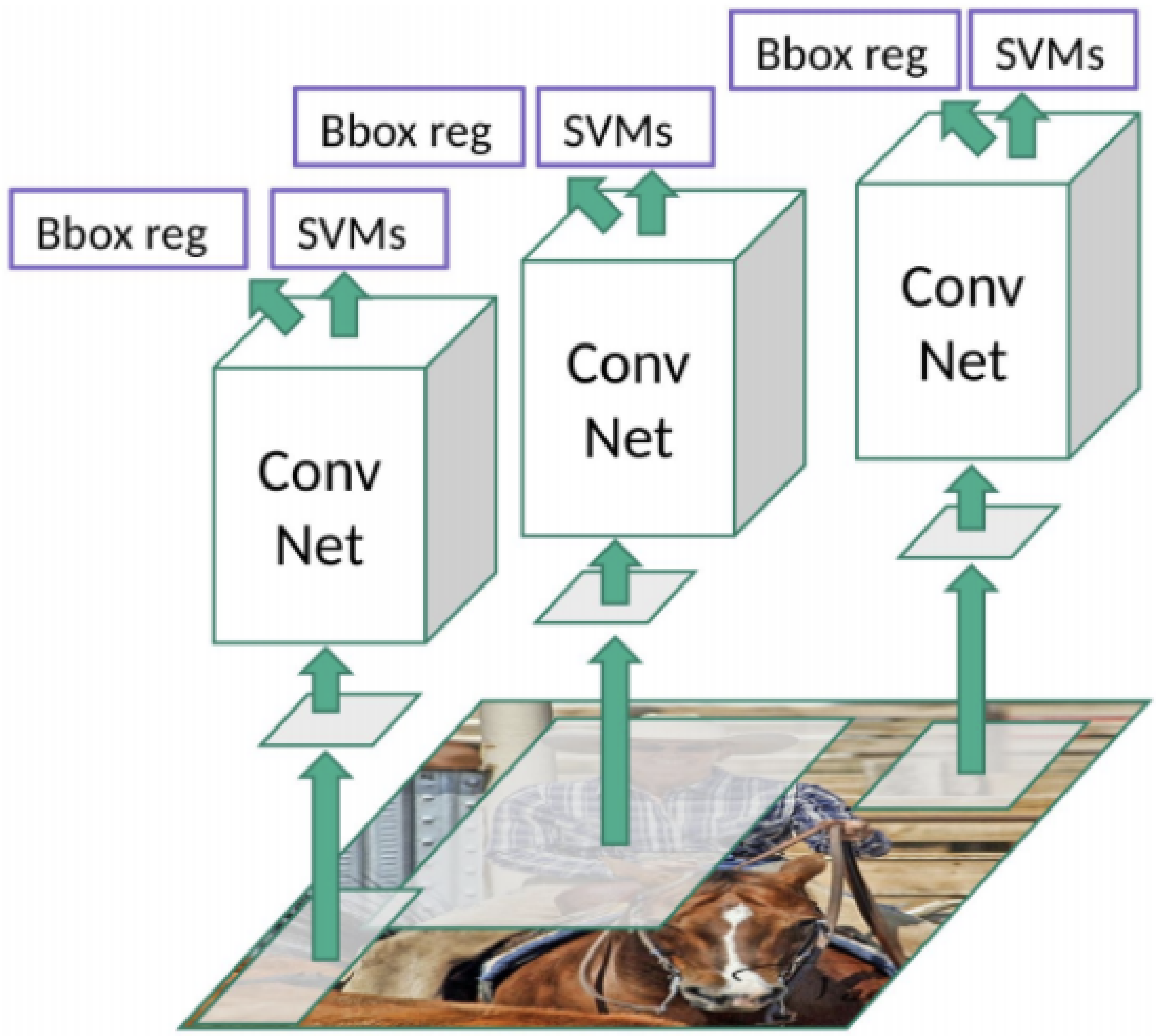
```
Greedy Algorithm :

1. From set of regions, choose two that are most similar.
2. Combine them into a single, larger region.
3. Repeat the above steps for multiple iterations.
```



3

3. Use the segmented region proposals to generate candidate object locations.




Bbox reg
SVMs
Bbox reg
SVMs
Bbox reg
SVMs
Conv
Net
Conv
Net
Conv
Net

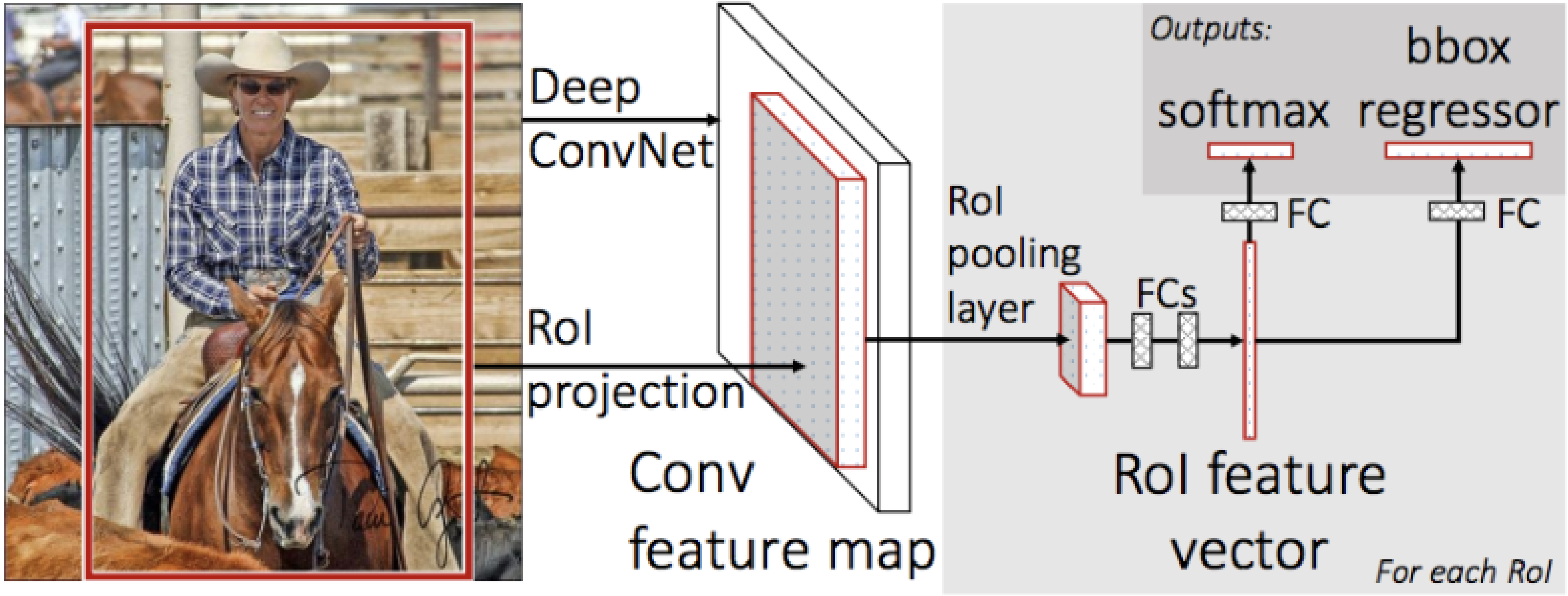
R-CNN

PROBLEMS WITH R-CNN

Problems with R-CNN

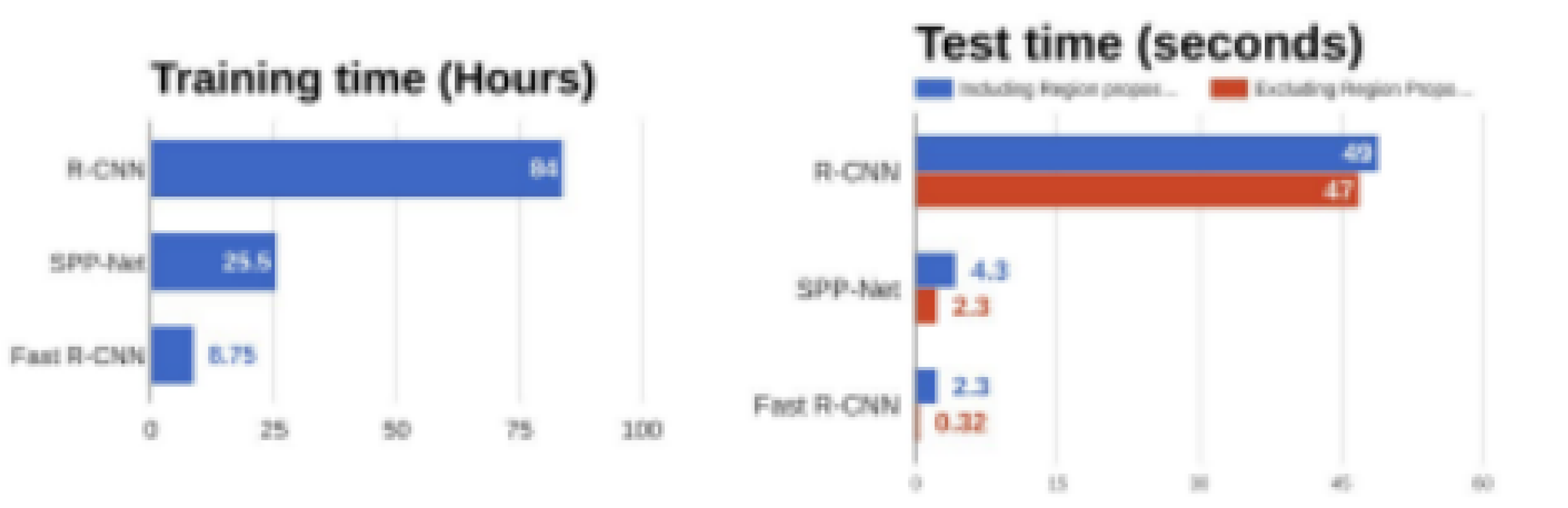
- It still takes a huge amount of time to train the network as you would have to classify 2000 region proposals per image.
- It cannot be implemented real time as it takes around 47 seconds for each test image.
- The selective search algorithm is a fixed algorithm. Therefore, no learning is happening at that stage. This could lead to the generation of bad candidate region proposals.

Fast R-CNN



Fast R-CNN

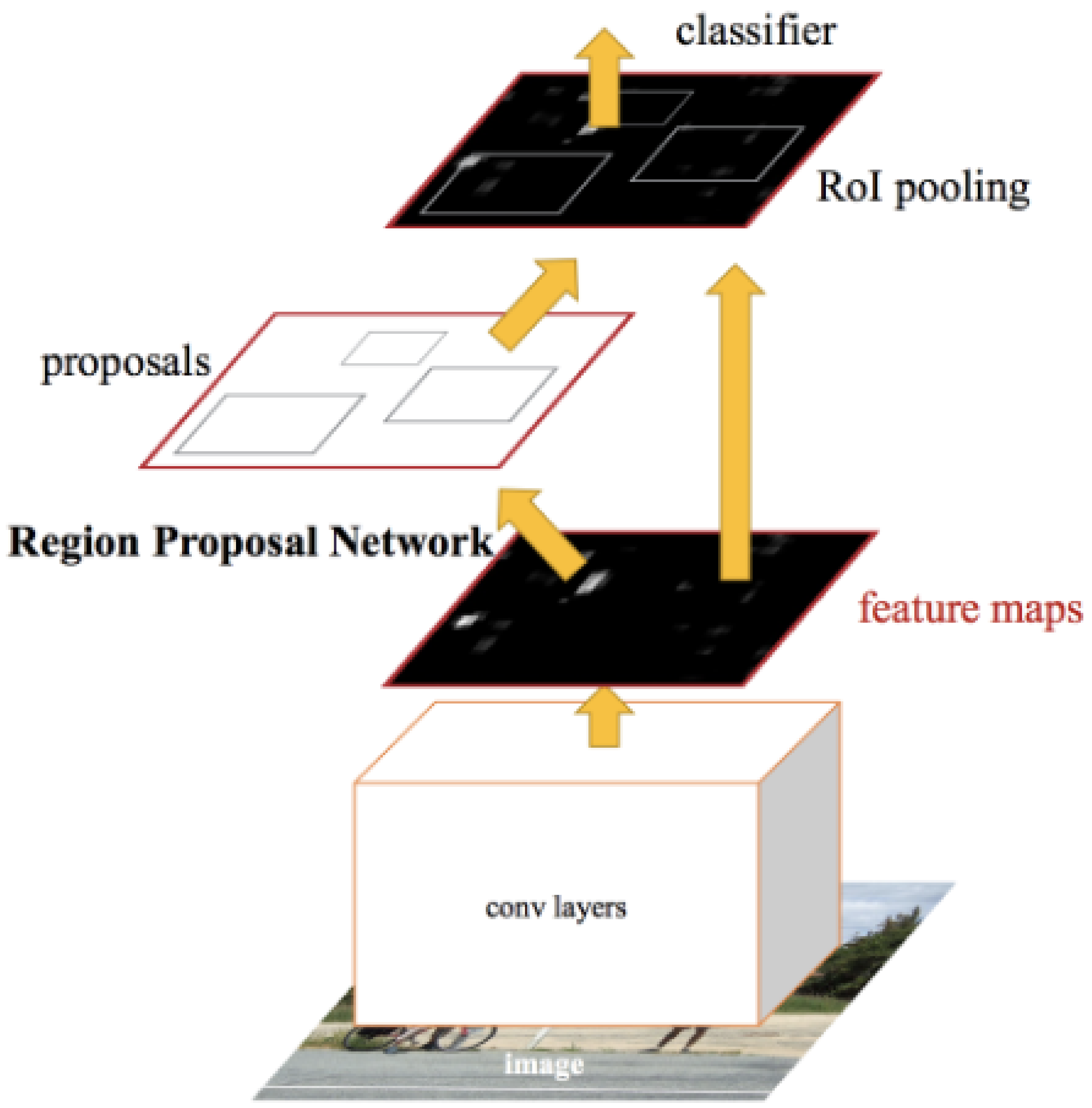
The reason "Fast R-CNN" is faster than R-CNN is because you don't have to feed 2000 region proposals to the convolutional neural network every time. Instead, the convolution operation is done only once per image and a feature map is generated from it.



Comparison of object detection algorithms

INSTEAD OF USING SELECTIVE SEARCH ALGORITHM ON THE FEATURE MAP TO IDENTIFY THE REGION PROPOSALS, A SEPARATE NETWORK IS USED TO PREDICT THE REGION PROPOSALS.

Faster R-CNN



Faster R-CNN

Mask R-CNN ?

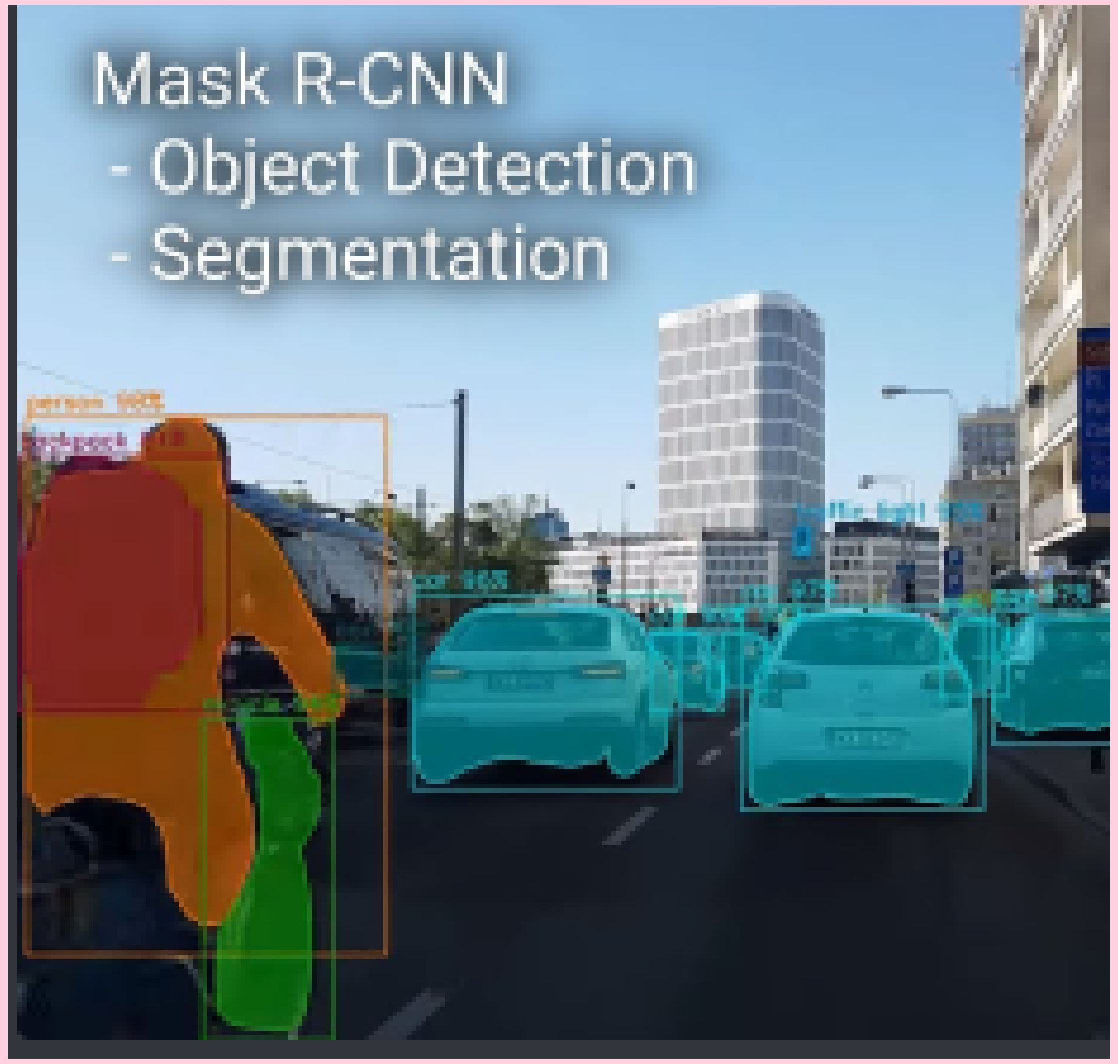
STAY TUNED

MASK RCNN

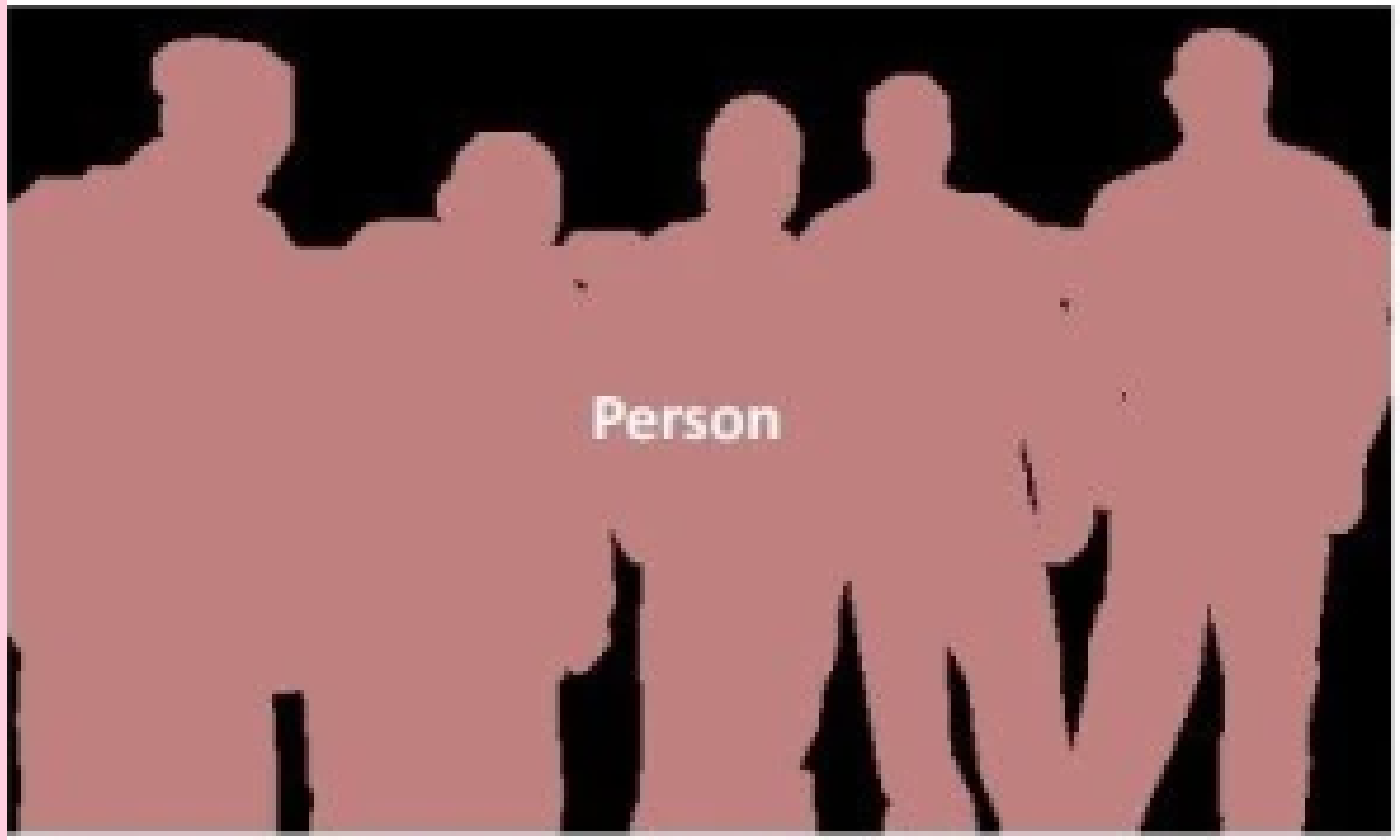
state-of-the-art in terms of image segmentation and instance segmentation. Mask R-CNN was developed on top of Faster R-CNN,

IMAGE SEGMENTATION

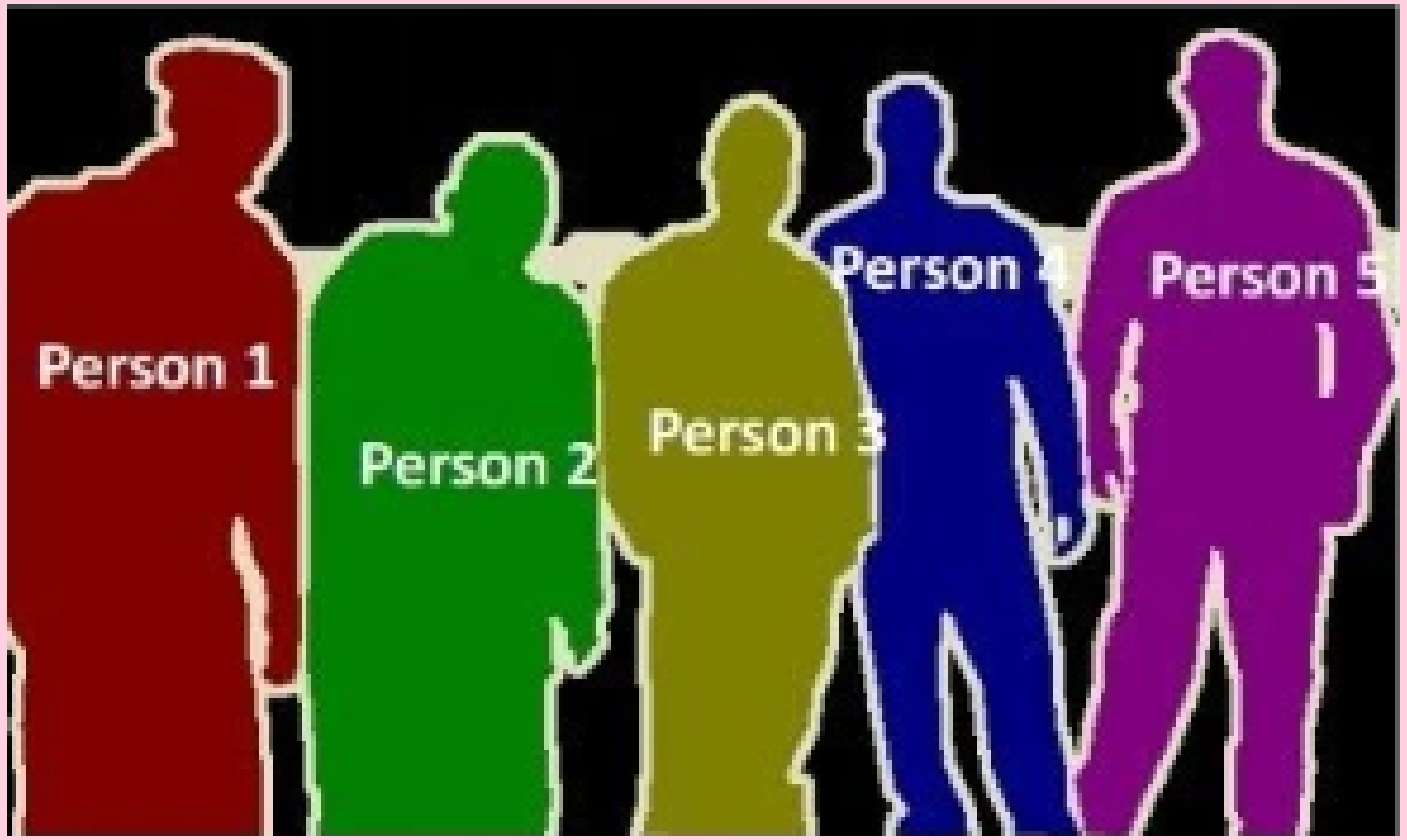
The computer vision task Image Segmentation is the process of partitioning a digital image into multiple segments (sets of pixels, also known as image objects). This segmentation is used to locate objects and boundaries (lines, curves, etc.).



SEMANTIC SEGMENTATION



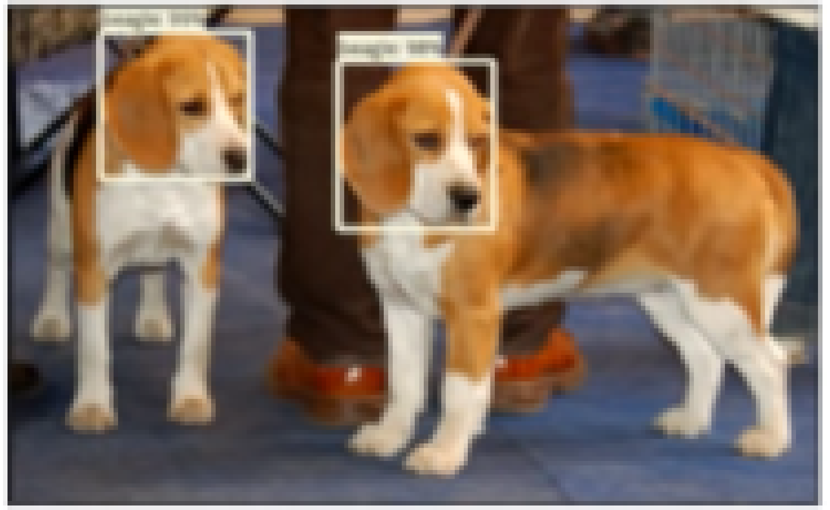
INSTANCE SEGMENTATION



Masked R-CNN

- approach to *Instance Segmentation*
- Two sub-problems:

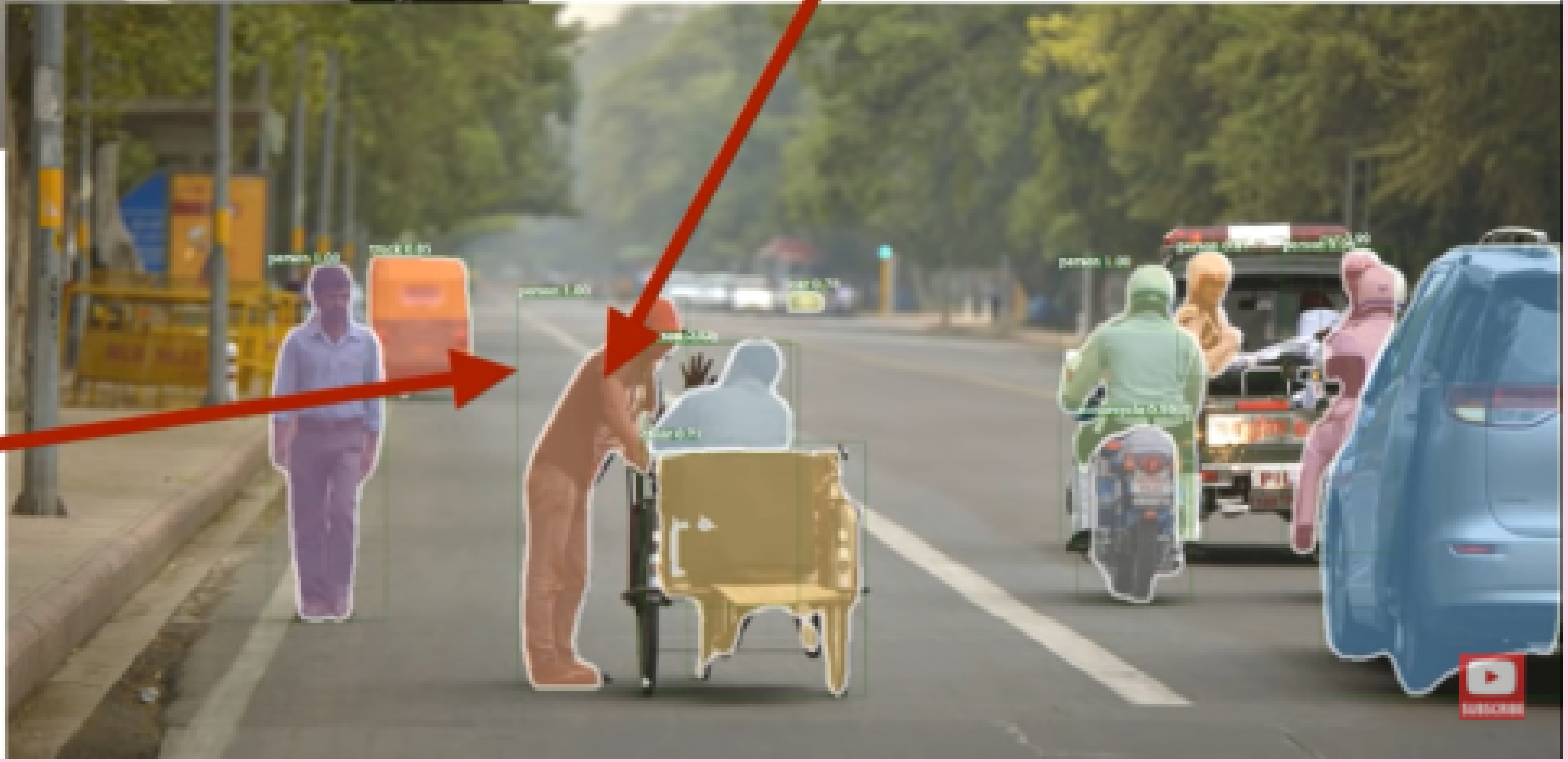
1. Object Detection



2. Semantic Segmentation

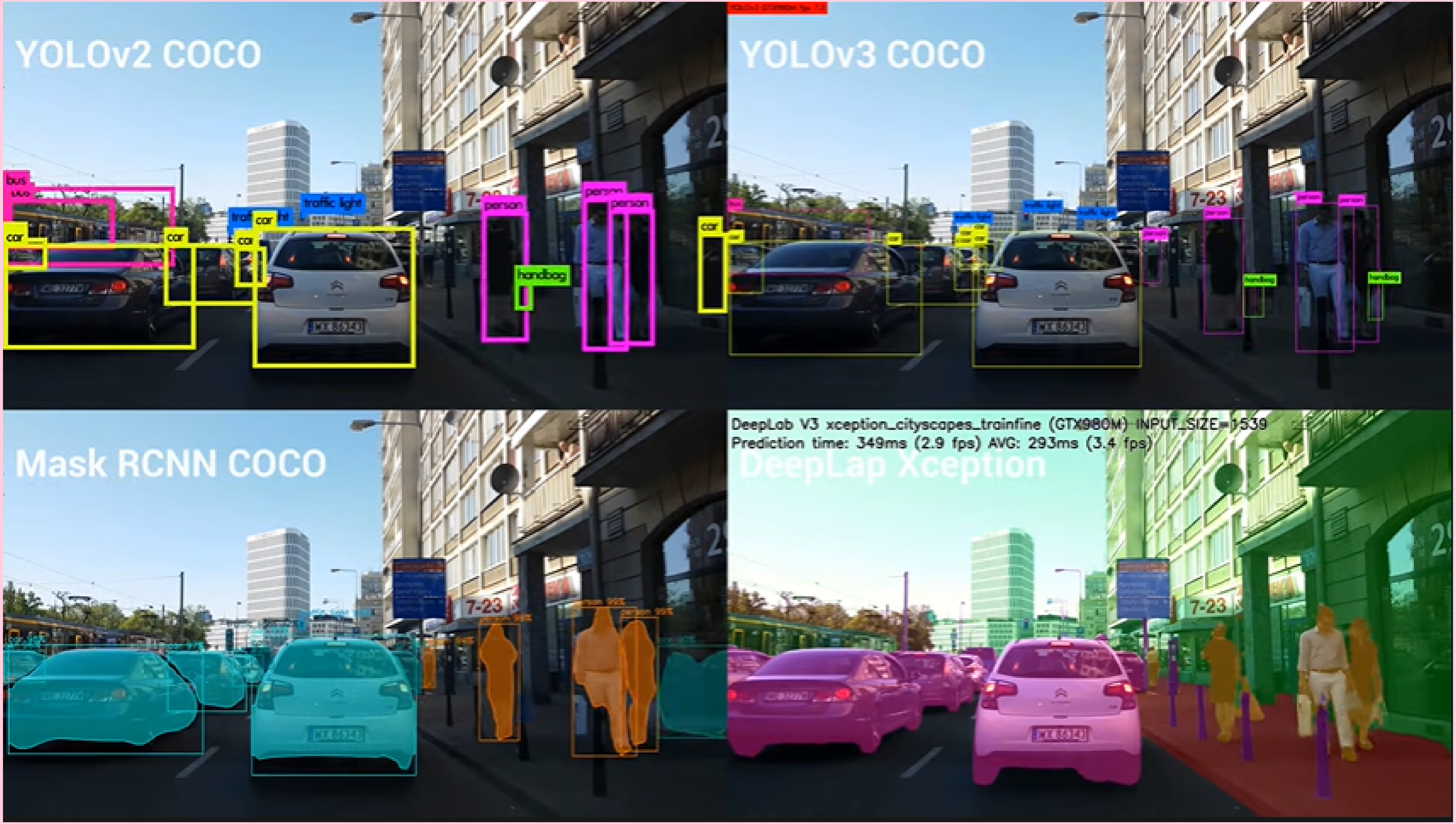
Colored Mask

Light green
bounding box



ADVANTAGES OF MASK R-CNN

- *Simplicity: Mask R-CNN is simple to train.*
- *Performance: Mask R-CNN outperforms all existing, single-model entries on every task.*
- *Efficiency: The method is very efficient and adds only a small overhead to Faster R-CNN.*
- *Flexibility: Mask R-CNN is easy to generalize to other tasks. For example, it is possible to use Mask R-CNN for human pose estimation in the same framework.*


YOLOv2 COCO
bus
car
traffic light
person
handbag
YOLOv3 COCO
Mask RCNN COCO
DeepLab V3 xception_cityscapes_trainfine (GTX980M) INPUT_SIZE=1539
Prediction time: 349ms (2.9 fps) AVG: 293ms (3.4 fps)
DeepLap Xception
7-23