AI in Built Environment DCP4300

Lec10: Computer Vision

Object detection/ Segmentation

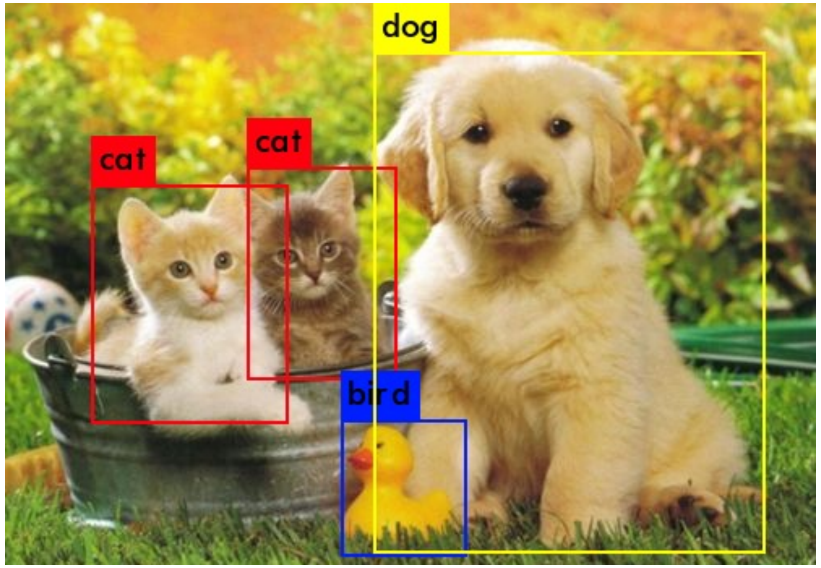
Dr. Chaofeng Wang
Jianhao Gao (TA)

University of Florida
College of Design Construction and Planning

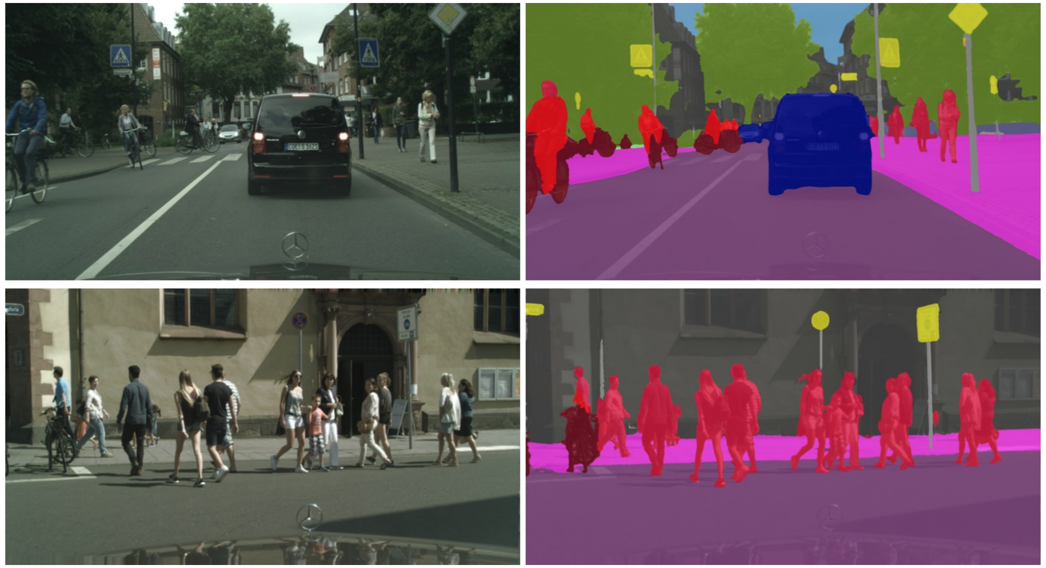
Major Tasks of Computer Vision

Classification

Object detection



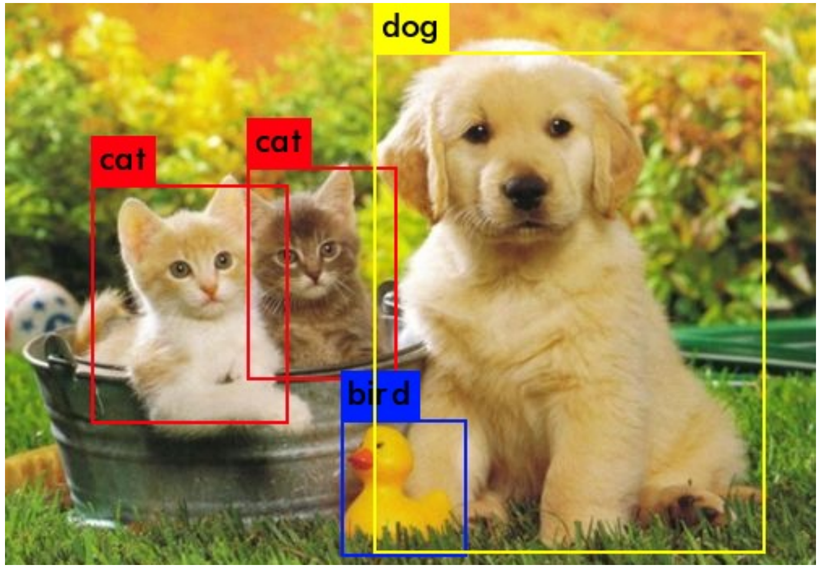
Segmentation



Major Tasks of Computer Vision

Classification

Object detection



Localization + Classification (sub-image)

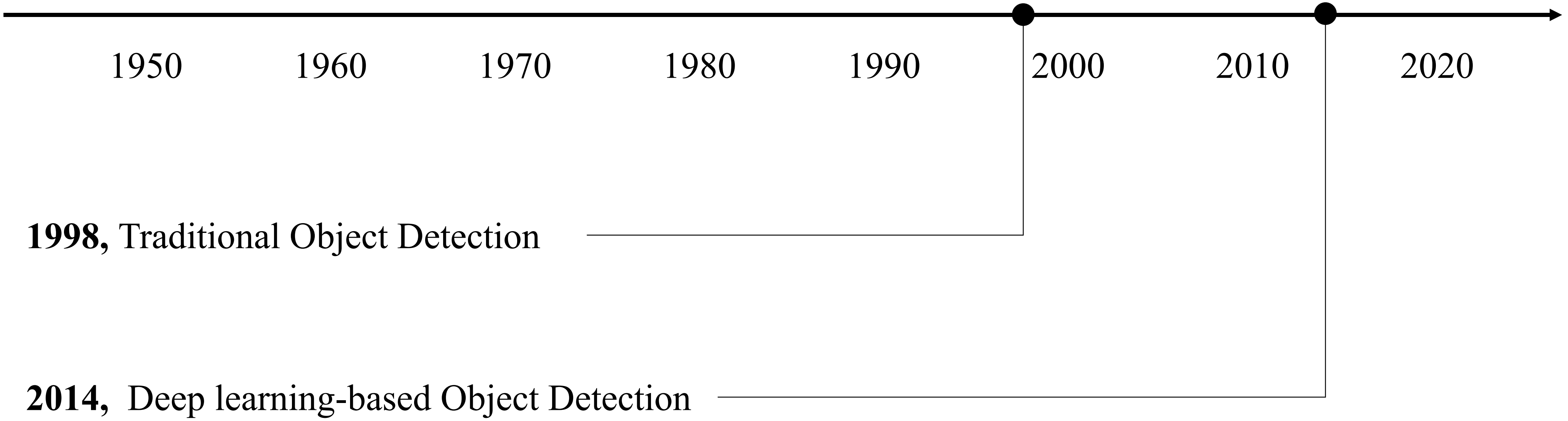
Scenarios to use object detection

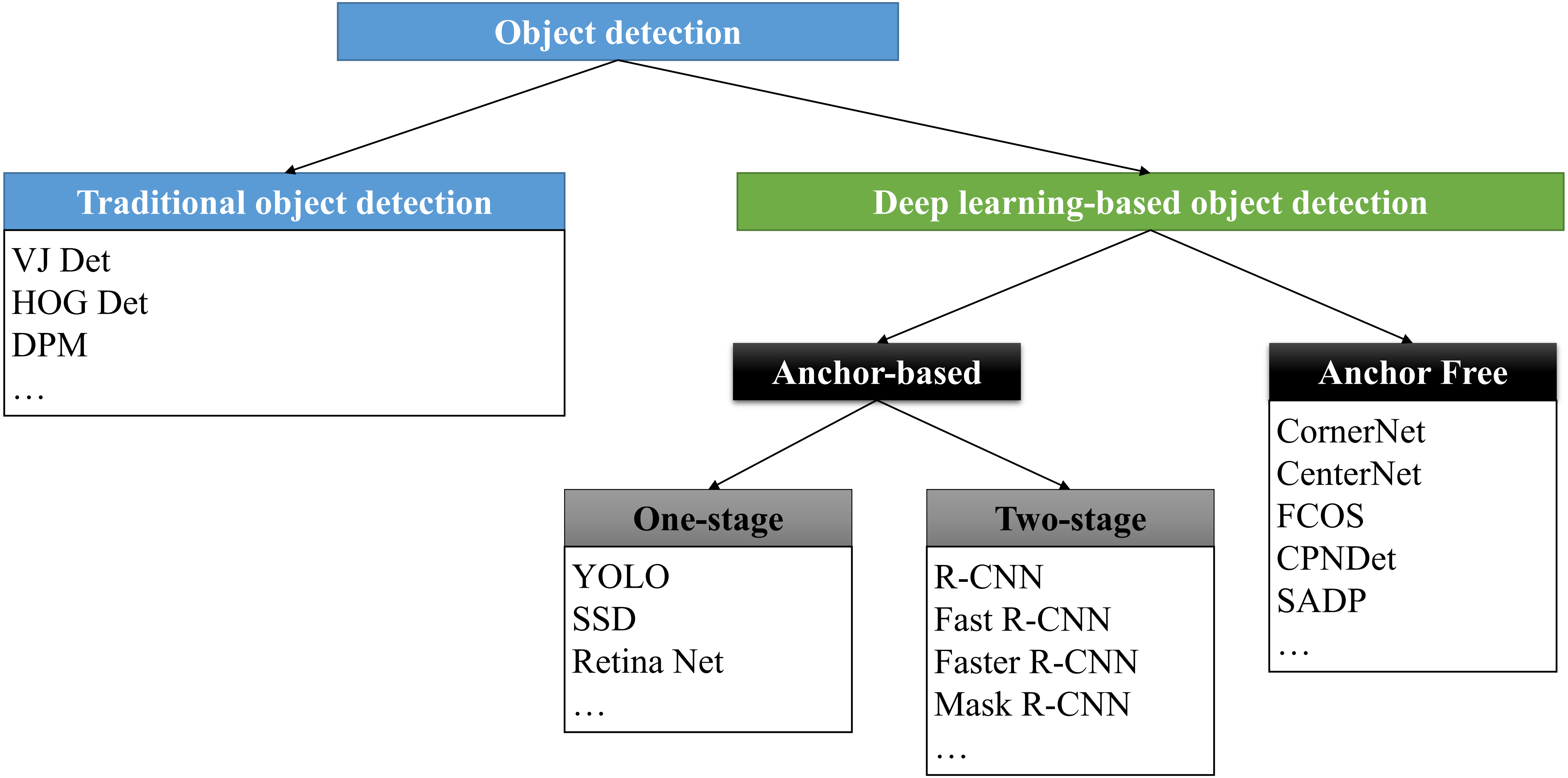
Multiple objects of interest in the image

Need to know the location of objects

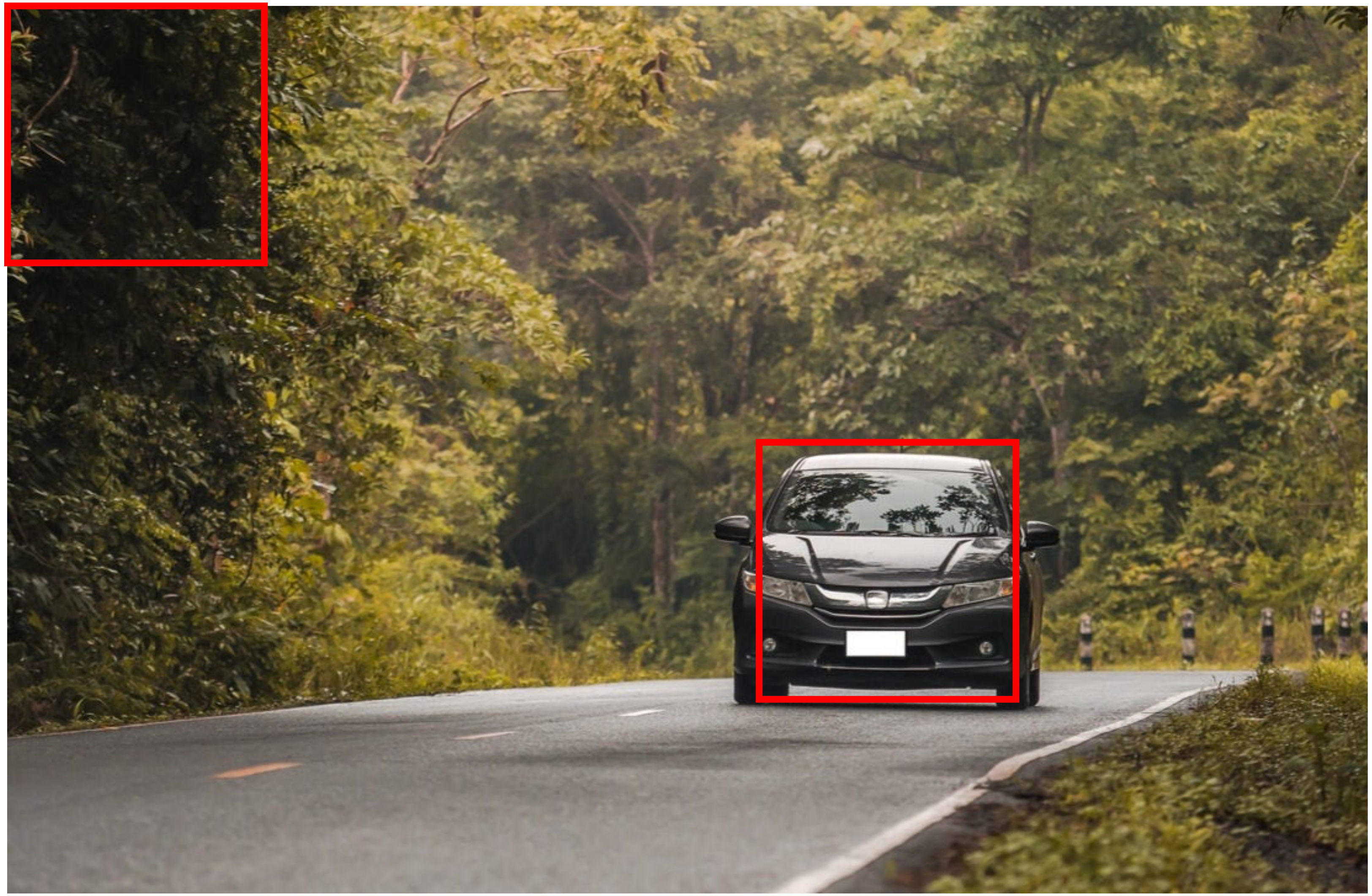
The objects of interest are not predominant

Key points in the history of object detection

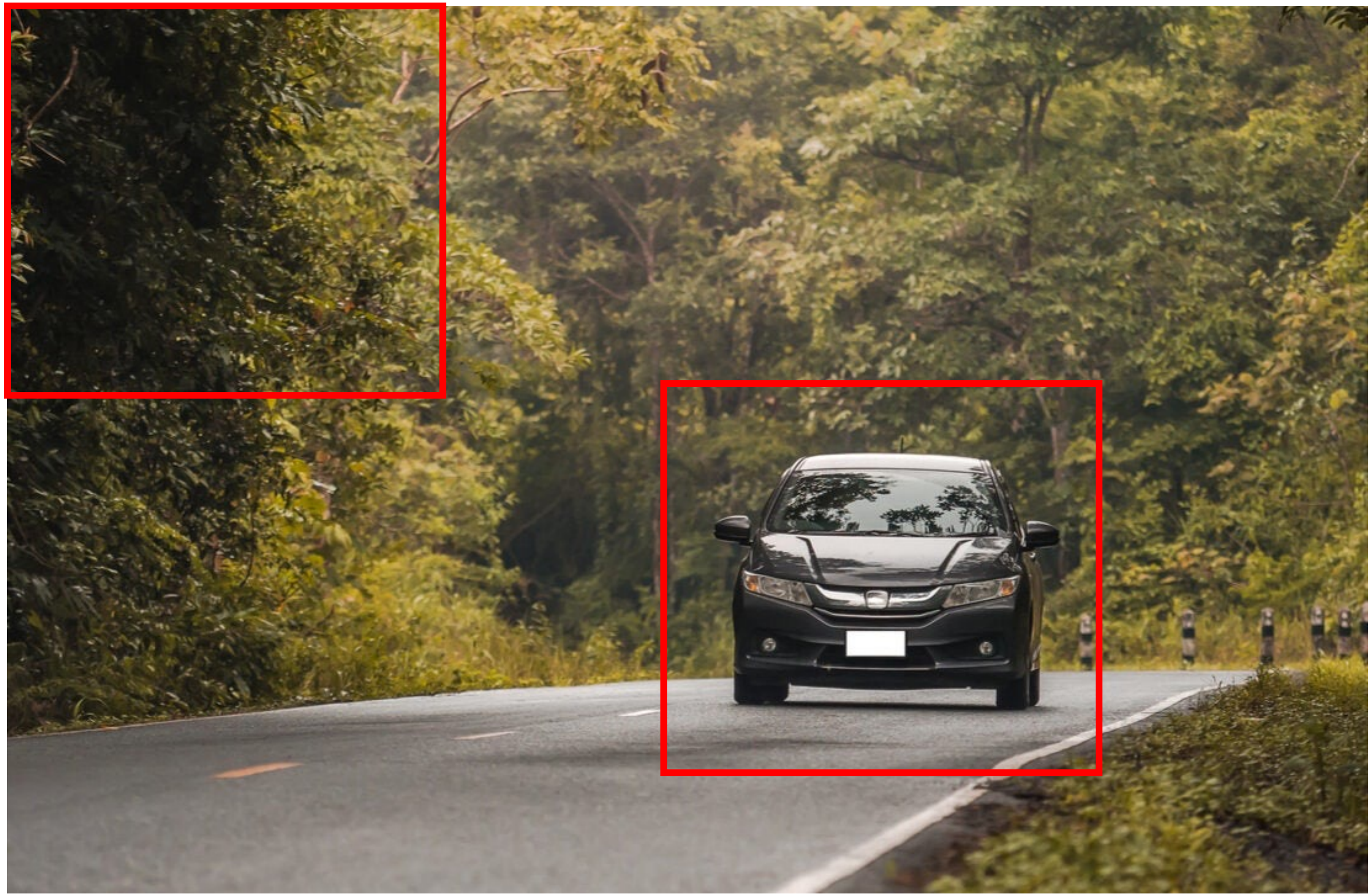



Object detection
Traditional object detection
VJ Det
HOG Det
DPM
…
Deep learning-based object detection
Anchor-based
Anchor Free
One-stage
YOLO
SSD
Retina Net
…
Two-stage
R-CNN
Fast R-CNN
Faster R-CNN
Mask R-CNN
…
CornerNet
CenterNet
FCOS
CPNDet
SADP
…

Sliding windows



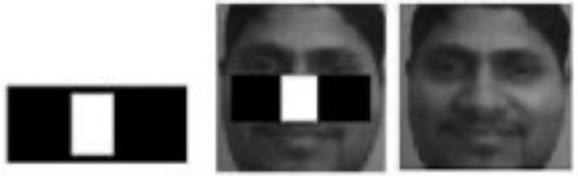
Sliding windows



Change window size

Sliding windows

Early stage: Hand-engineered features + Simple classifiers



Feature that looks similar to the bridge of the nose is applied onto the face

Feature that looks similar to the eye region which is darker than the upper cheeks is applied onto a face

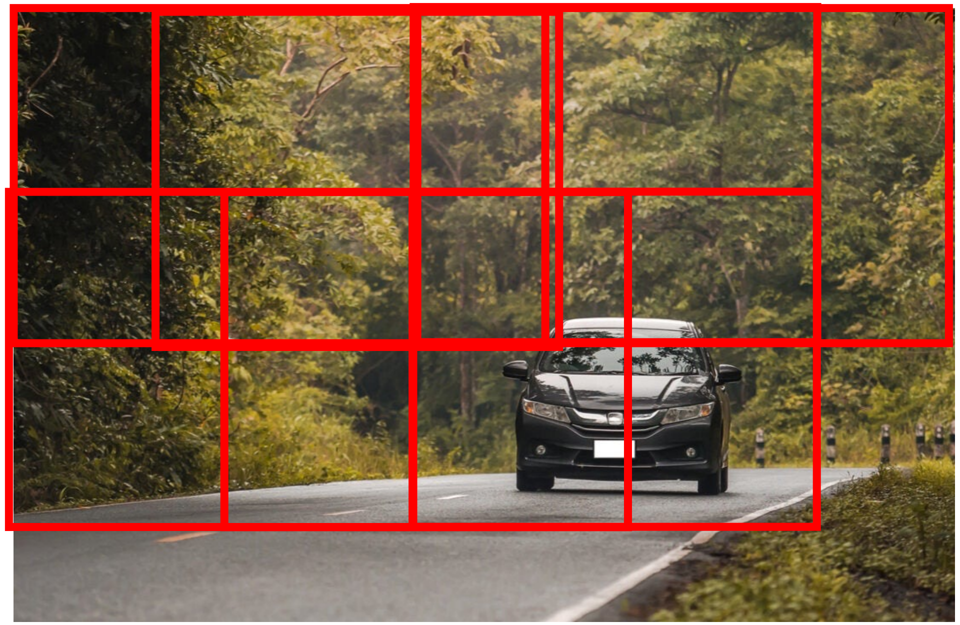
Viola–Jones method

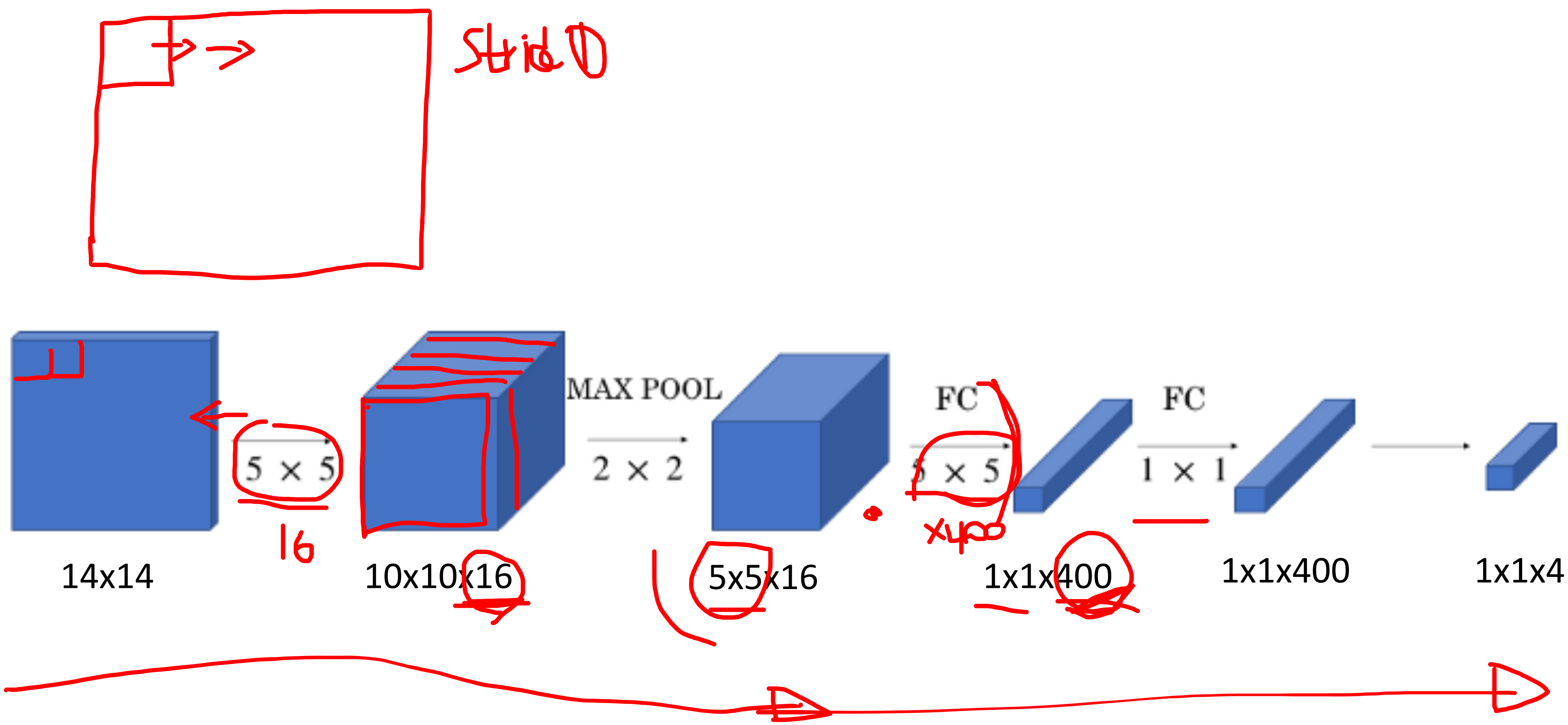
Viola, P., & Jones, M. (2001). Robust real-time object detection. International journal of computer vision, 4(34-47), 4.

Sliding windows

Sliding windows + Simple linear classifiers — Speed is ok, but performance is not good

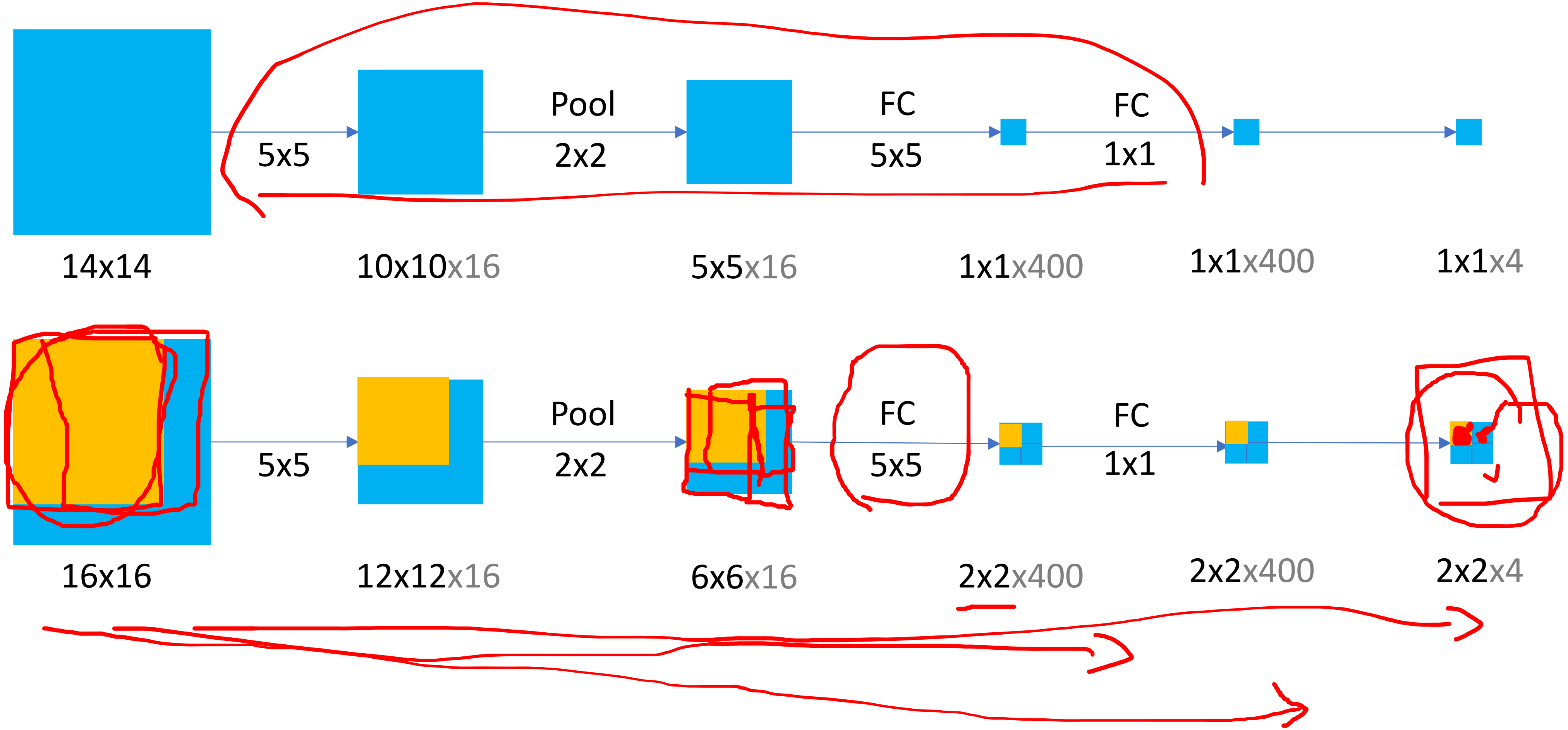
Sliding windows + Neural networks — Performance is higher but slower





Example of a convolutional neural network classifier

Convolutional Implementation of Sliding Windows



Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., Fergus, R., & LeCun, Y. (2013). Overfeat: Integrated recognition, localization and detection using convolutional networks. arXiv preprint arXiv:1312.6229. https://arxiv.org/abs/1312.6229

Convolutional Implementation of Sliding Windows

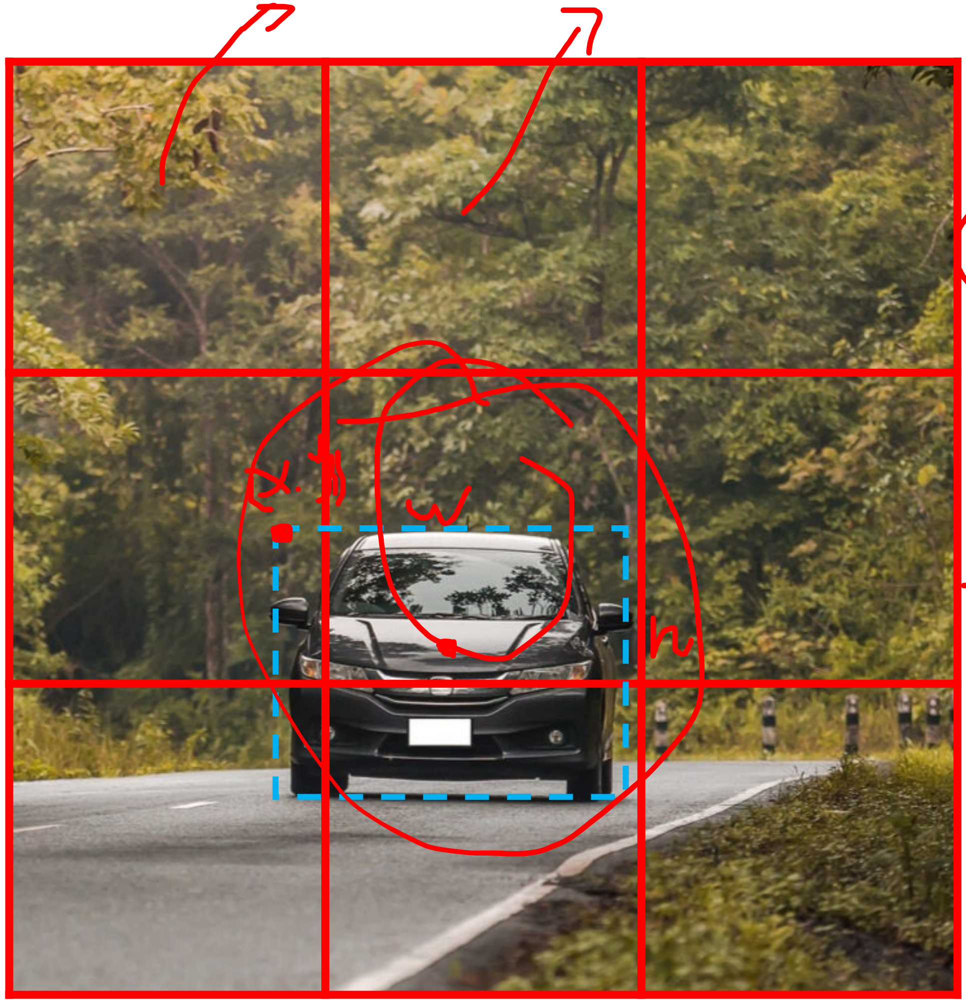
Disadvantages:

1. Computationally expensive, slow
2. Bounding box prediction not accurate

Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., Fergus, R., & LeCun, Y. (2013). Overfeat: Integrated recognition, localization and detection using convolutional networks. arXiv preprint arXiv:1312.6229. https://arxiv.org/abs/1312.6229

Anchor-based One-stage algorithm

YOLO



X

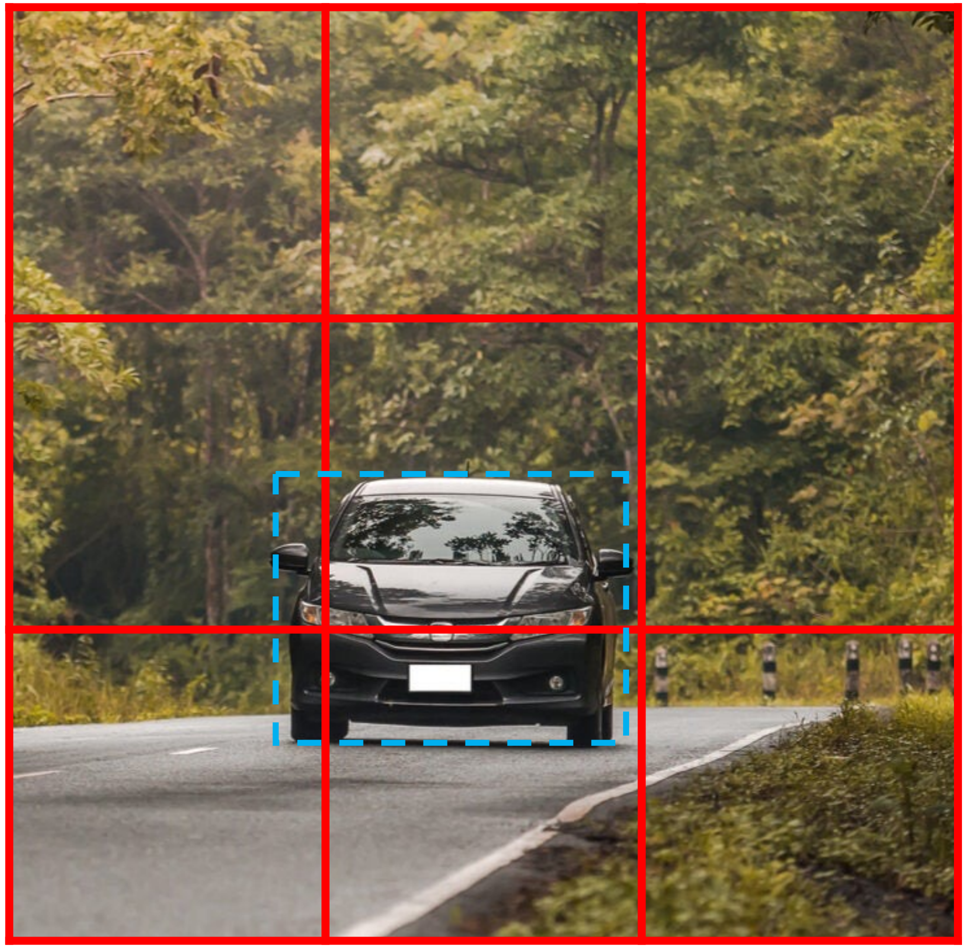
$$Y = 3 \times 3 \times 8$$

$$y = \begin{bmatrix} p \\ x \\ y \\ w \\ h \\ c_1 = 1 \\ c_2 = 0 \end{bmatrix} \qquad y = \begin{bmatrix} p \\ \vdots \end{bmatrix}$$

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 779-788). https://www.cv-foundation.org/openaccess/content_cvpr_2016/papers/Redmon_You_Only_Look_CVPR_2016_paper.pdf

Anchor-based One-stage algorithm

YOLO

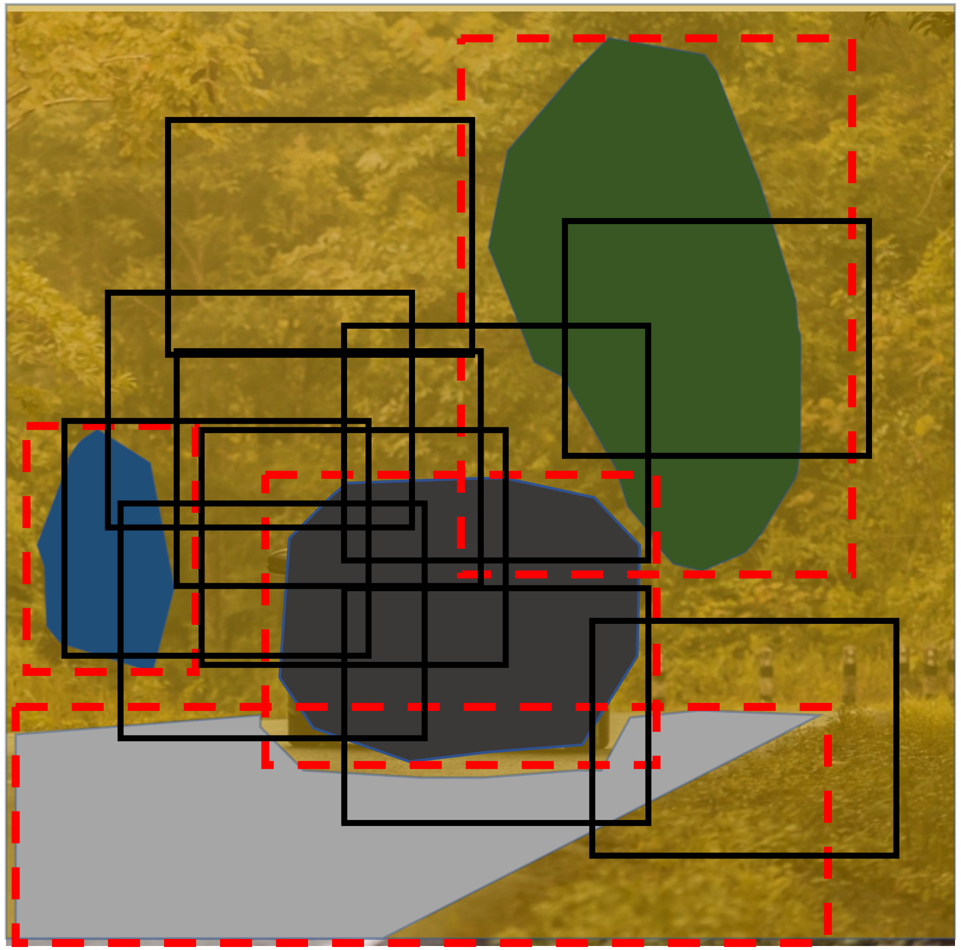


Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 779-788). https://www.cv-foundation.org/openaccess/content_cvpr_2016/papers/Redmon_You_Only_Look_CVPR_2016_paper.pdf

Traditional CV: Segmentation

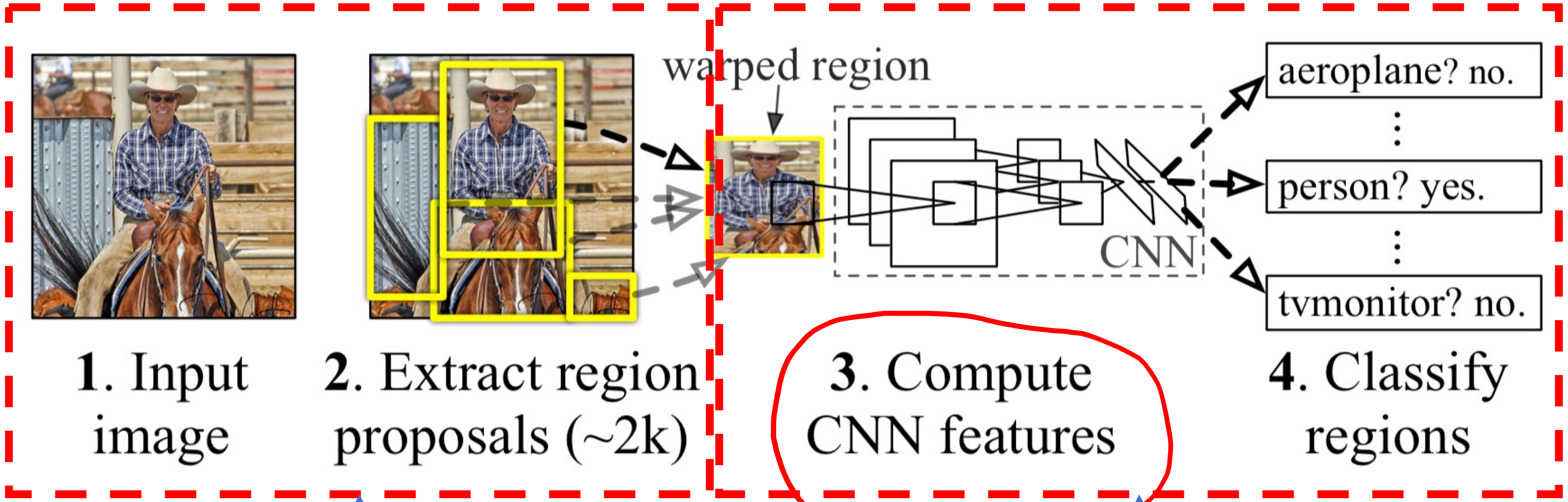
Anchor-based Two-stage algorithm

Region Proposals (R-CNN)



Stage-1: Generate region proposals from image
Stage-2: Predict bounding boxes from region proposals

R-CNN



Generate region proposals: method agnostic
The author chose the selective search method:
A greedy algorithm to recursively combine similar regions into larger ones

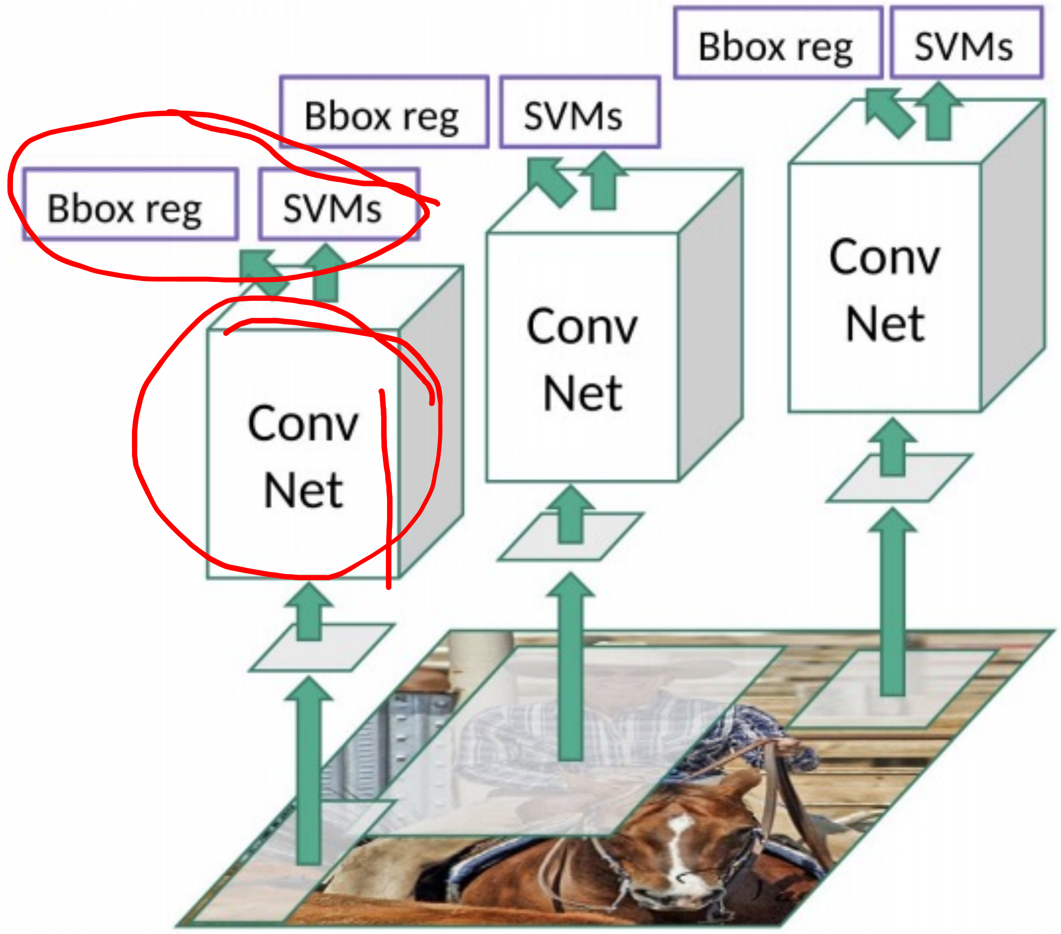
A CNN-based classifier

Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 580-587).
https://openaccess.thecvf.com/content_cvpr_2014/papers/Girshick_Rich_Feature_Hierarchies_2014_CVPR_paper.pdf

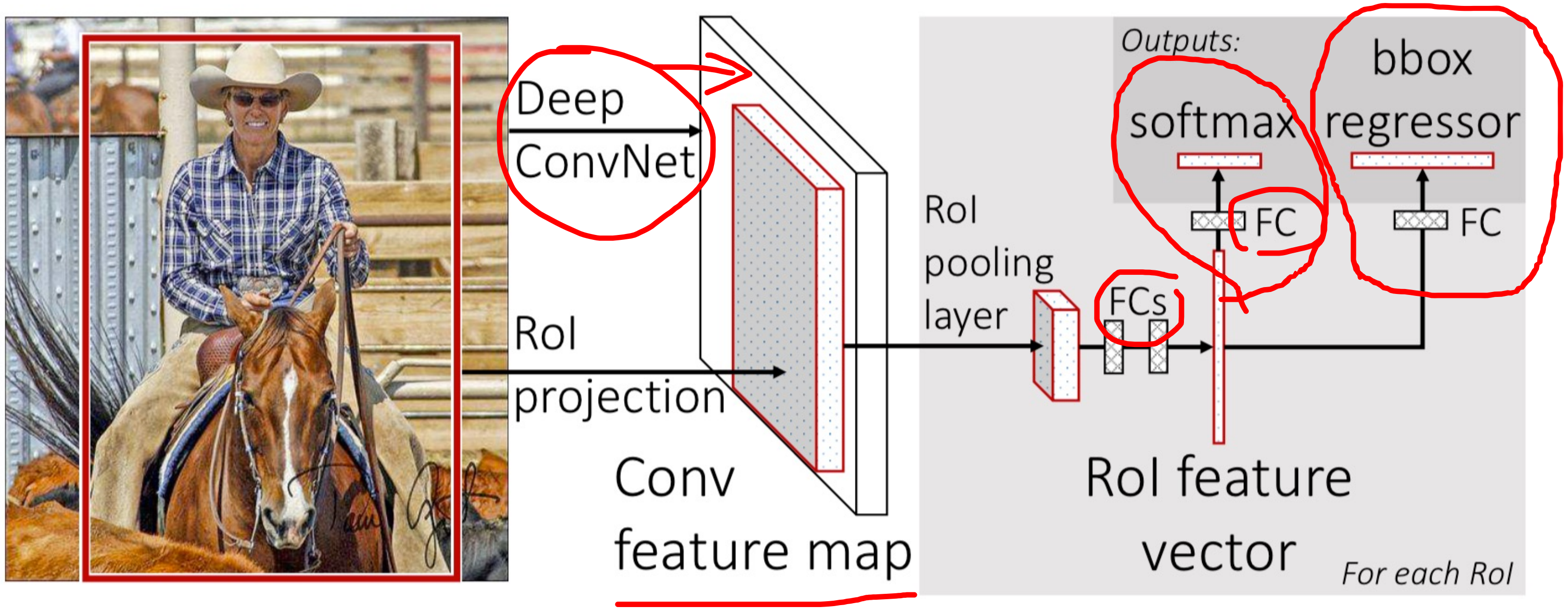
Selective search: J.Uijlings,K.vandeSande,T.Gevers,andA.Smeulders.Selective search for object recognition. *IJCV*, 2013.
https://staff.fnwi.uva.nl/th.gevers/pub/GeversIJCV2013.pdf

R-CNN

It’s very slow:
Too much proposals;
Each needs to be classified, independently.



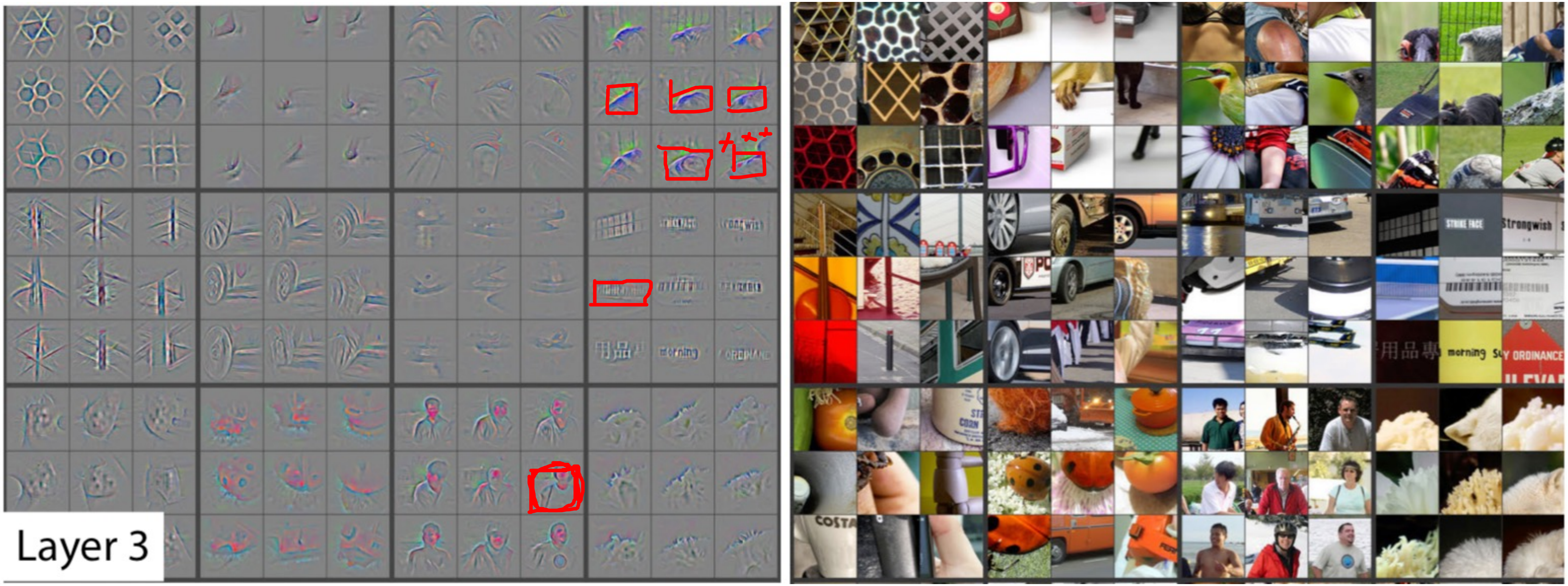
Fast R-CNN



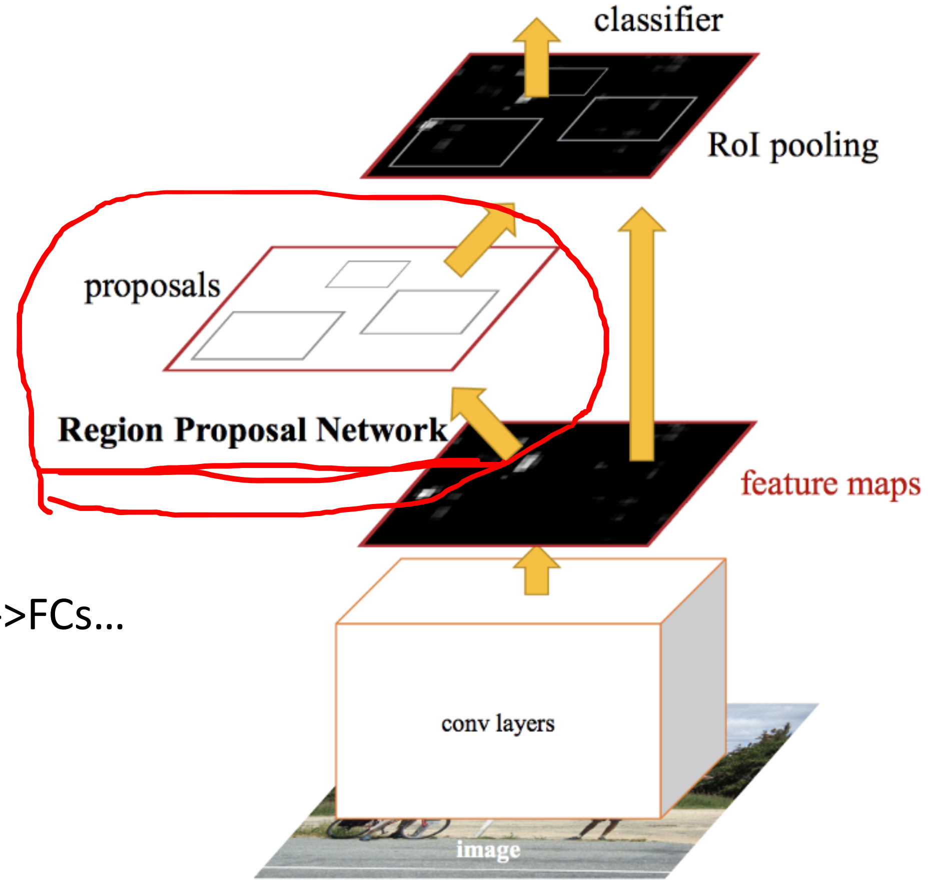
Input image-> ConvNet->Conv feature map->proposals->FCs…

Girshick, R. (2015). Fast r-cnn. In Proceedings of the IEEE international conference on computer vision (pp. 1440-1448).https://openaccess.thecvf.com/content_iccv_2015/papers/Girshick_Fast_R-CNN_ICCV_2015_paper.pdf

Conv feature map



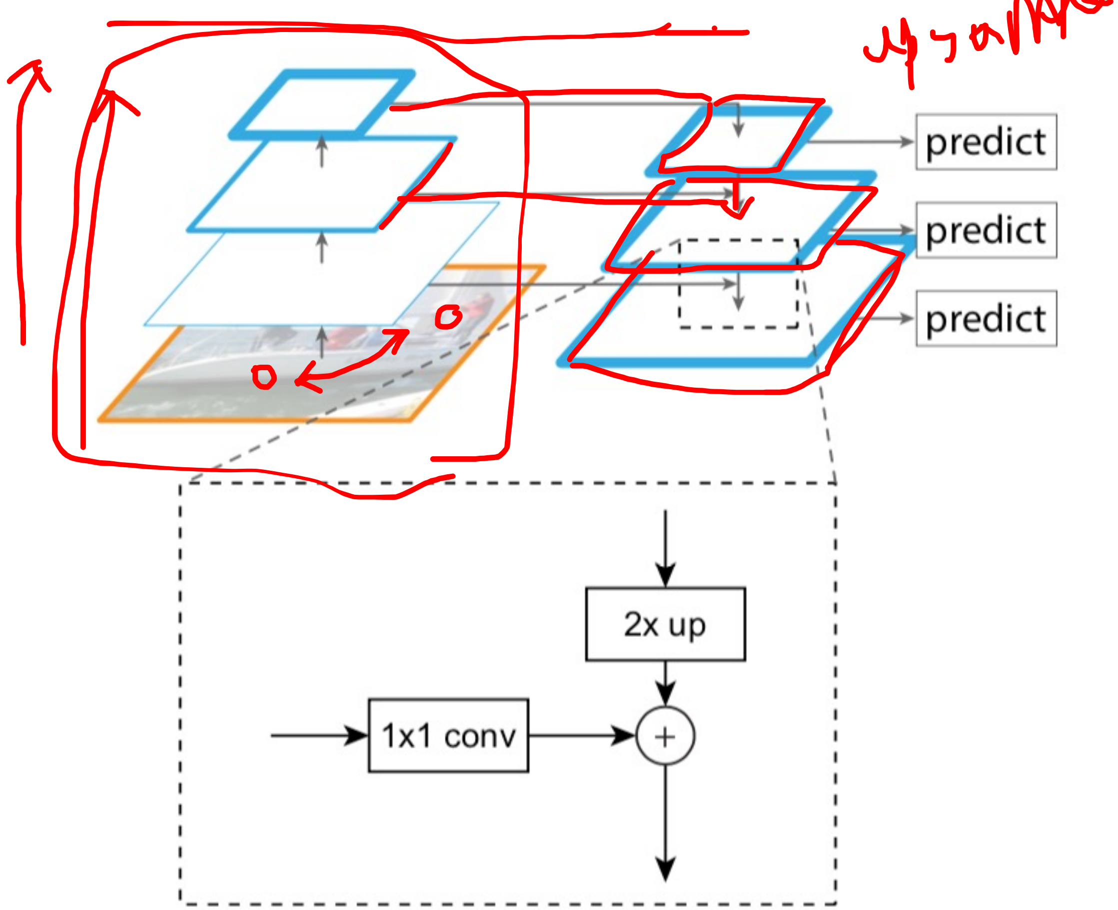
Faster R-CNN



Input image-> ConvNet->Conv feature map **by a network**->proposals->FCs…

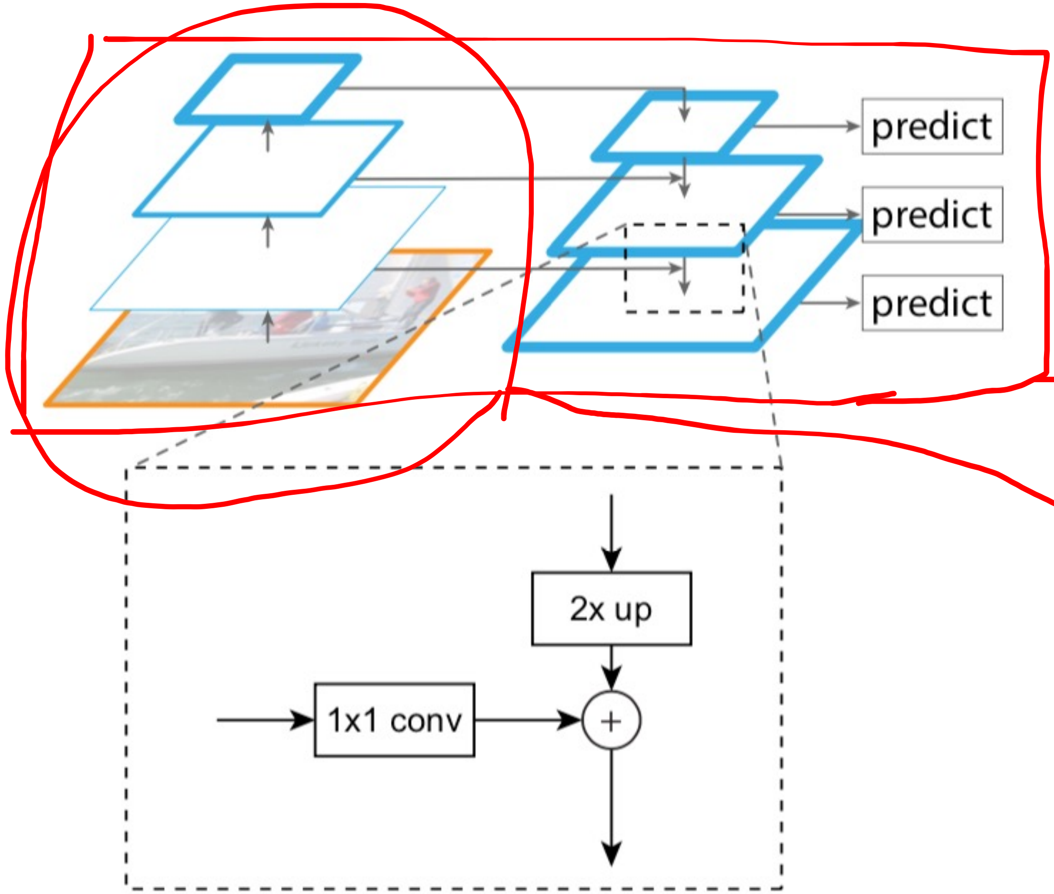
Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. Advances in neural information processing systems, 28, 91-99. https://proceedings.neurips.cc/paper/5638-faster-r-cnn-towards-real-time-object-detection-with-region-proposal-networks.pdf

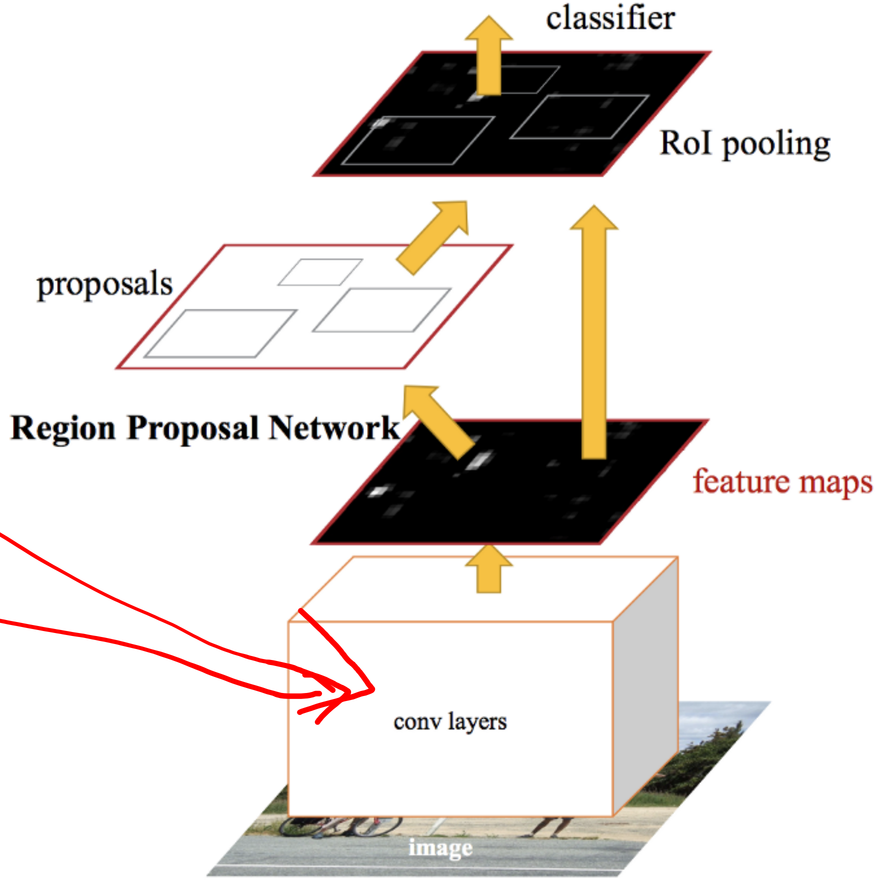
Feature Pyramid Networks (FPN)



Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). Feature pyramid networks for object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2117-2125). http://openaccess.thecvf.com/content_cvpr_2017/papers/Lin_Feature_Pyramid_Networks_CVPR_2017_paper.pdf

Feature Pyramid Networks (FPN) in Faster R-CNN





Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). Feature pyramid networks for object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2117-2125). http://openaccess.thecvf.com/content_cvpr_2017/papers/Lin_Feature_Pyramid_Networks_CVPR_2017_paper.pdf

One-stage vs. Two-stage

One-stage detectors:

Computational demand is relatively low

Generally faster than two-stage methods

Suitable for real-time detections

Not good at recognizing irregularly shaped objects or a group of small objects.

Popular one-stage detectors include the YOLO, SSD, and RetinaNet.

Two-stage detectors:

Demand more computational resources

Generally slower than one-stage methods

Two-stage methods achieve the highest detection accuracy

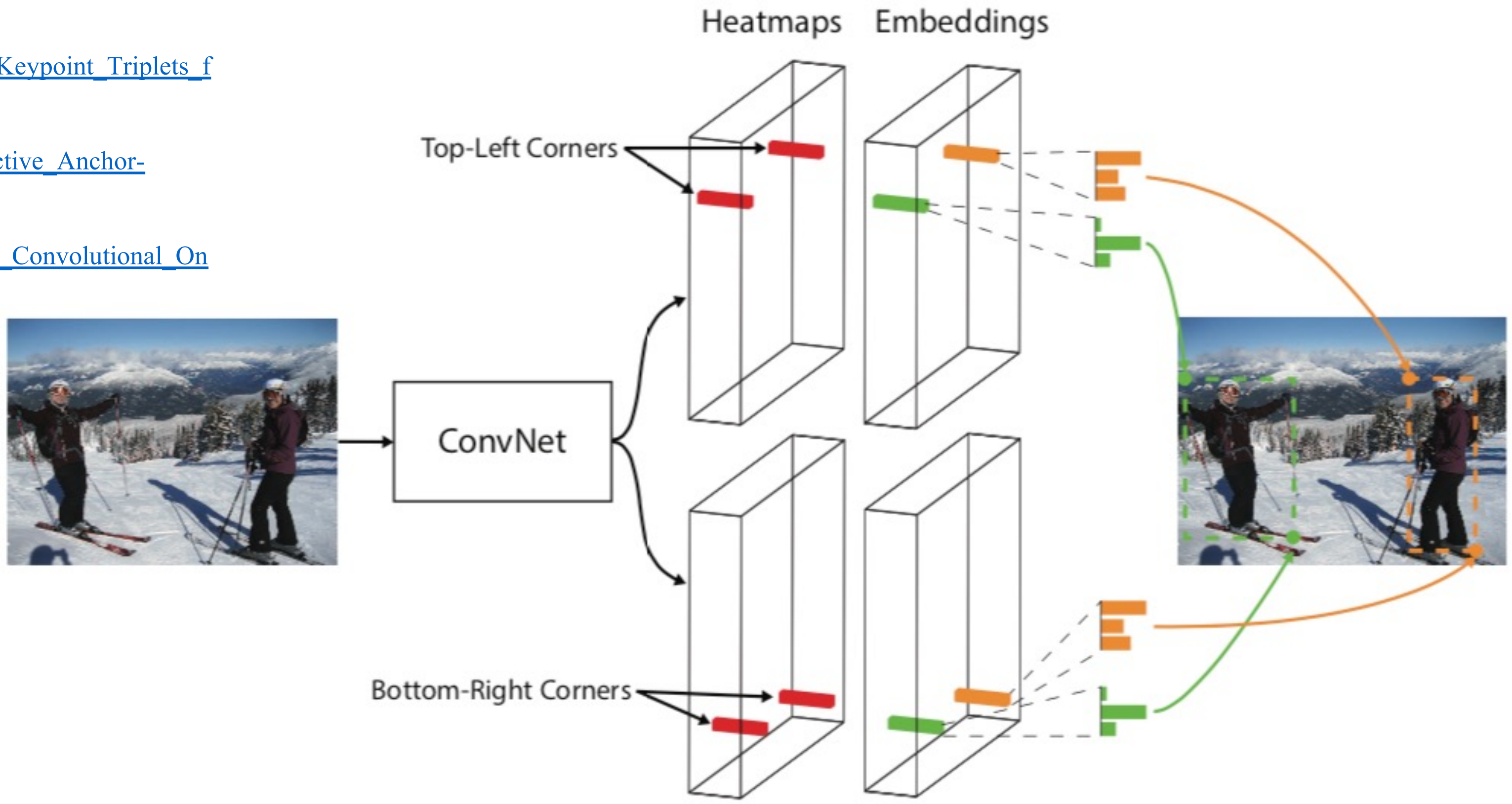
Various two-stage detectors include region convolutional neural network (RCNN), with evolutions Faster R-CNN or Mask R-CNN. The latest evolution is the granulated RCNN (G-RCNN).

Two-stage object detectors first find a region of interest and use this cropped region for classification. However, such multi-stage detectors are usually not end-to-end trainable because cropping is a non-differentiable operation.

Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. Advances in neural information processing systems, 28, 91-99. https://proceedings.neurips.cc/paper/5638-faster-r-cnn-towards-real-time-object-detection-with-region-proposal-networks.pdf

Anchor free methods

CornerNet
http://openaccess.thecvf.com/content_ECCV_2018/papers/Hei_Law_CornerNet_Detecting_Objects_ECCV_2018_paper.pdf
CenterNet
https://openaccess.thecvf.com/content_ICCV_2019/papers/Duan_CenterNet_Keypoint_Triplets_for_Object_Detection_ICCV_2019_paper.pdf
FSAF
http://openaccess.thecvf.com/content_CVPR_2019/papers/Zhu_Feature_Selective_Anchor-Free_Module_for_Single-Shot_Object_Detection_CVPR_2019_paper.pdf
FCOS
https://openaccess.thecvf.com/content_ICCV_2019/papers/Tian_FCOS_Fully_Convolutional_One-Stage_Object_Detection_ICCV_2019_paper.pdf
SAPD
https://arxiv.org/pdf/1911.12448

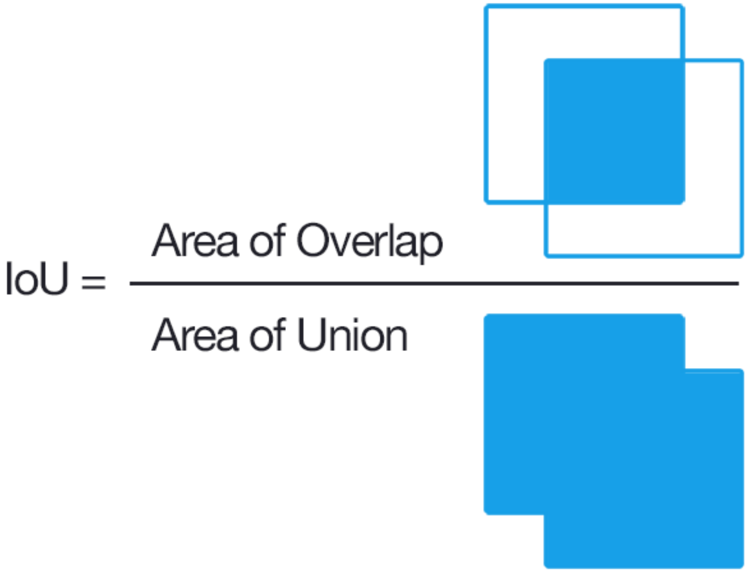


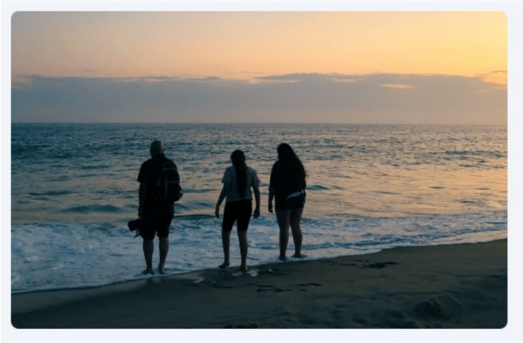
CornerNet

Proceedings of the European conference on computer vision (ECCV) (pp. 734-750).http://openaccess.thecvf.com/content_ECCV_2018/papers/Hei_Law_CornerNet_Detecting_Objects_ECCV_2018_paper.pdf

Metrics

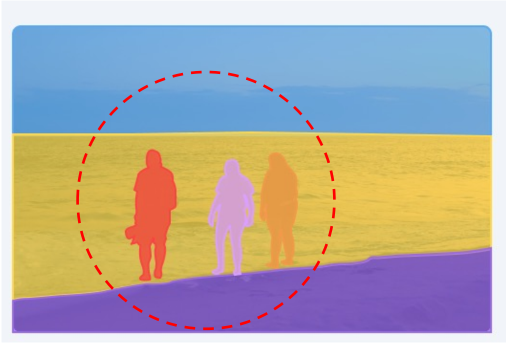
Intersection Over Union (IOU):
The ratio of intersection of ground truth and predicted bounding box or segmentation outputs over their union.





Image

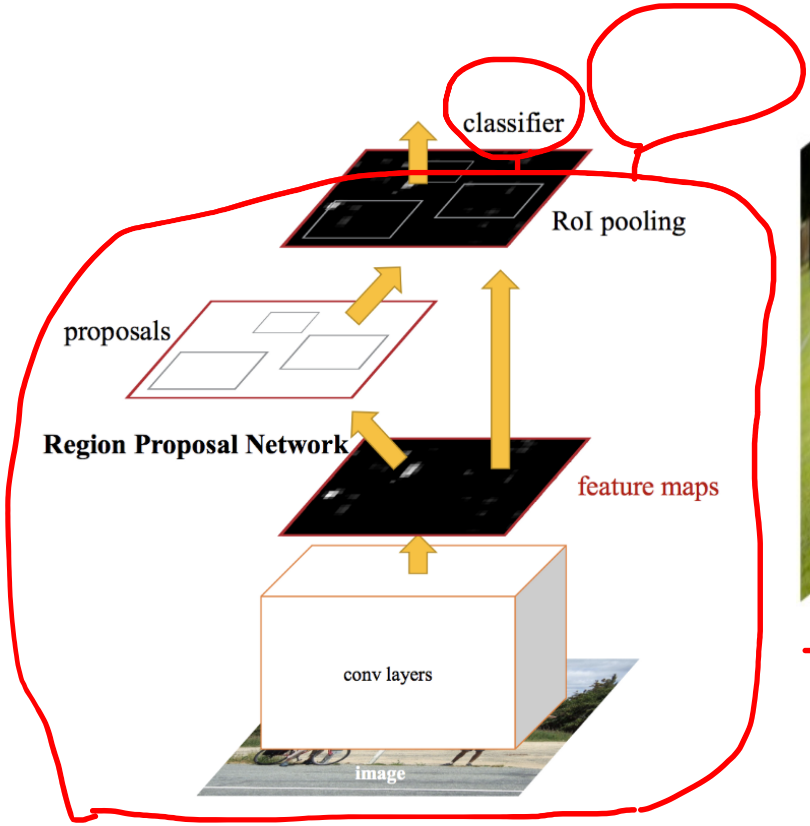
Semantic Segmentation



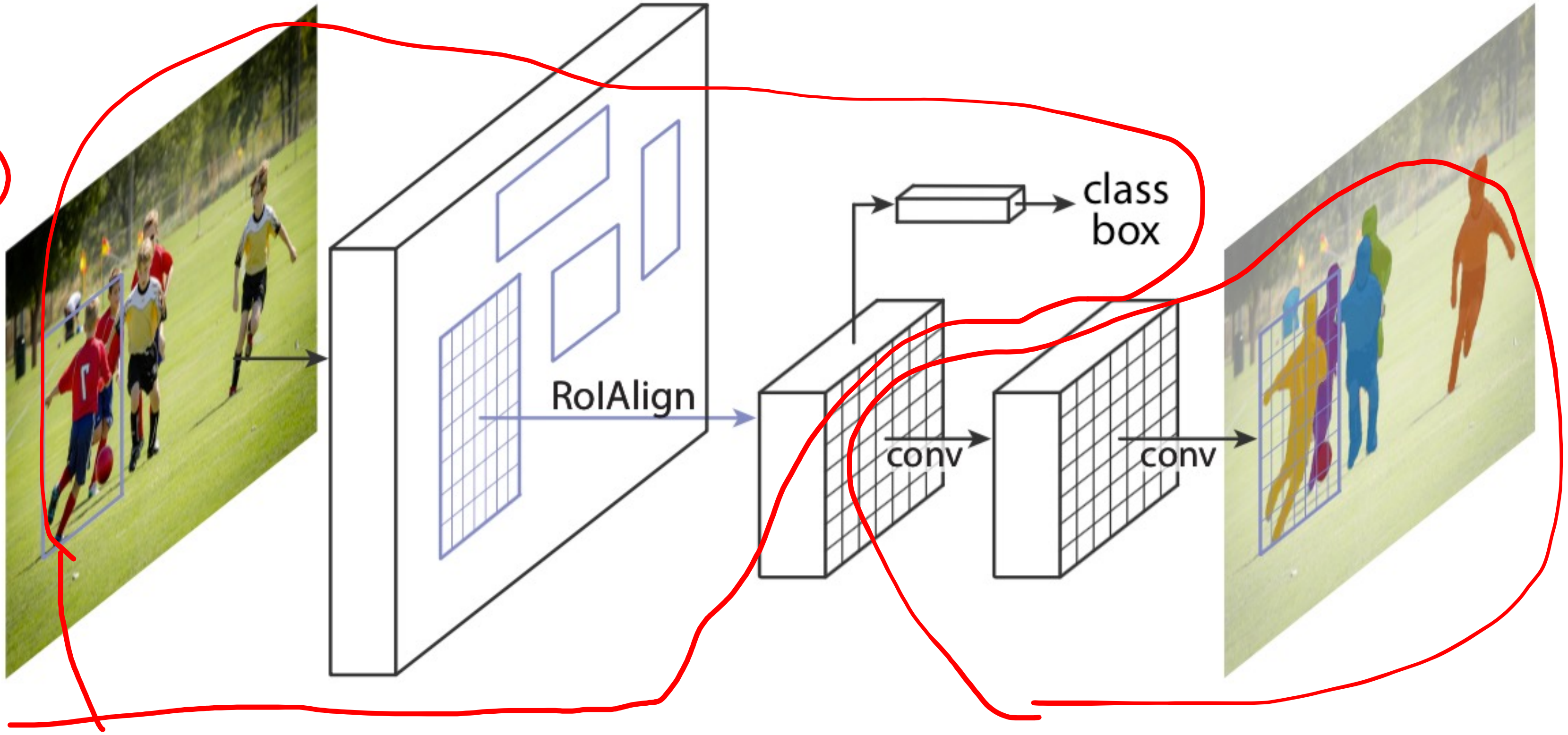
Instance Segmentation

Popular algorithms:
U-Net
Fast Fully Convolutional Network (FastFCN)
DeepLab
Mask R-CNN
…

Mask R-CNN For Detection + Segmentation



Faster R-CNN



Extends Faster R-CNN: adding a branch in the head for predicting segmentation masks on each Region of Interest

Mask R-CNN

He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision* (pp. 2961-2969). https://arxiv.org/pdf/1703.06870

Demo: Object detection and instance segmentation

Will run this demo in a Jupyter notebook on Google Colab:
https://colab.research.google.com/drive/1_FT8lzry_7uYRQ-jXF0Rg75H5vKB_oy2?usp=sharing