Lab2
Object Detection & Mask

Chen Houshuang

chenhoushuang@sjtu.edu.cn

contents

- R-cnn
- Fast Rcnn
- Faster Rcnn
- Mask Rcnn
- Yolo

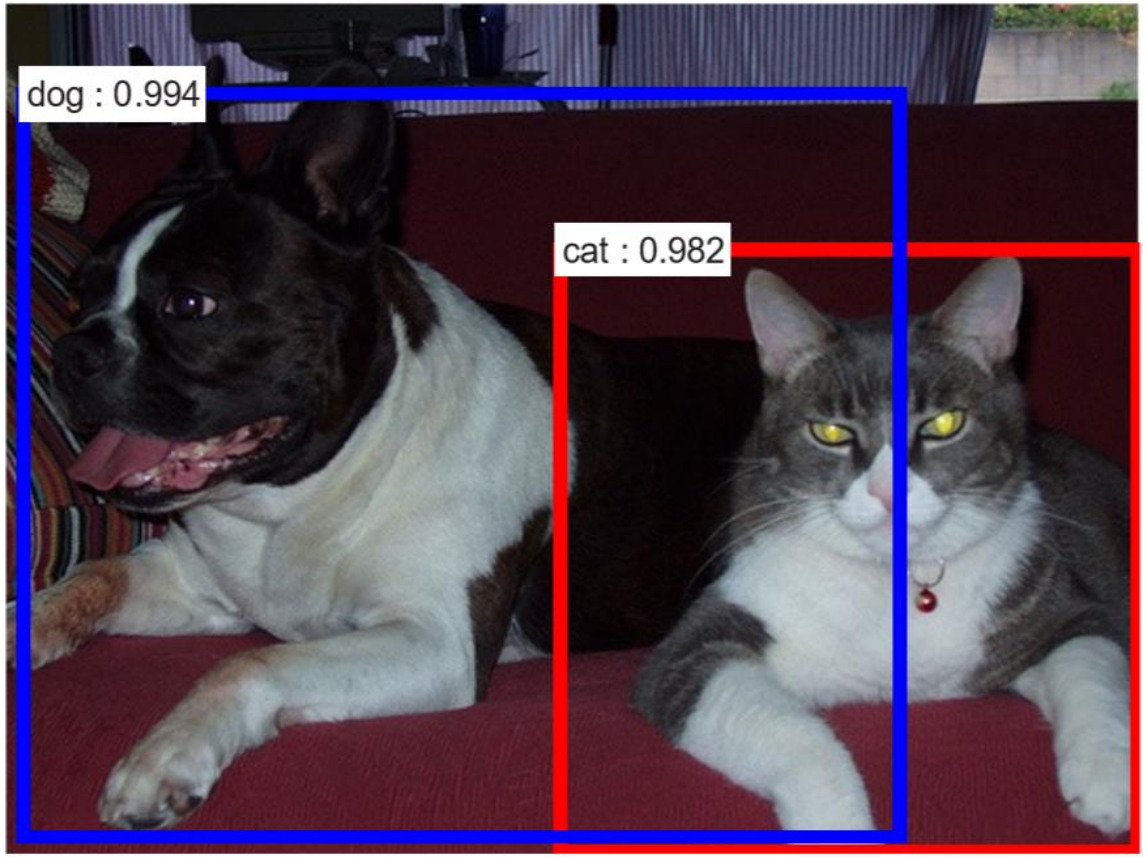
Different tasks

- Image classification

Different tasks

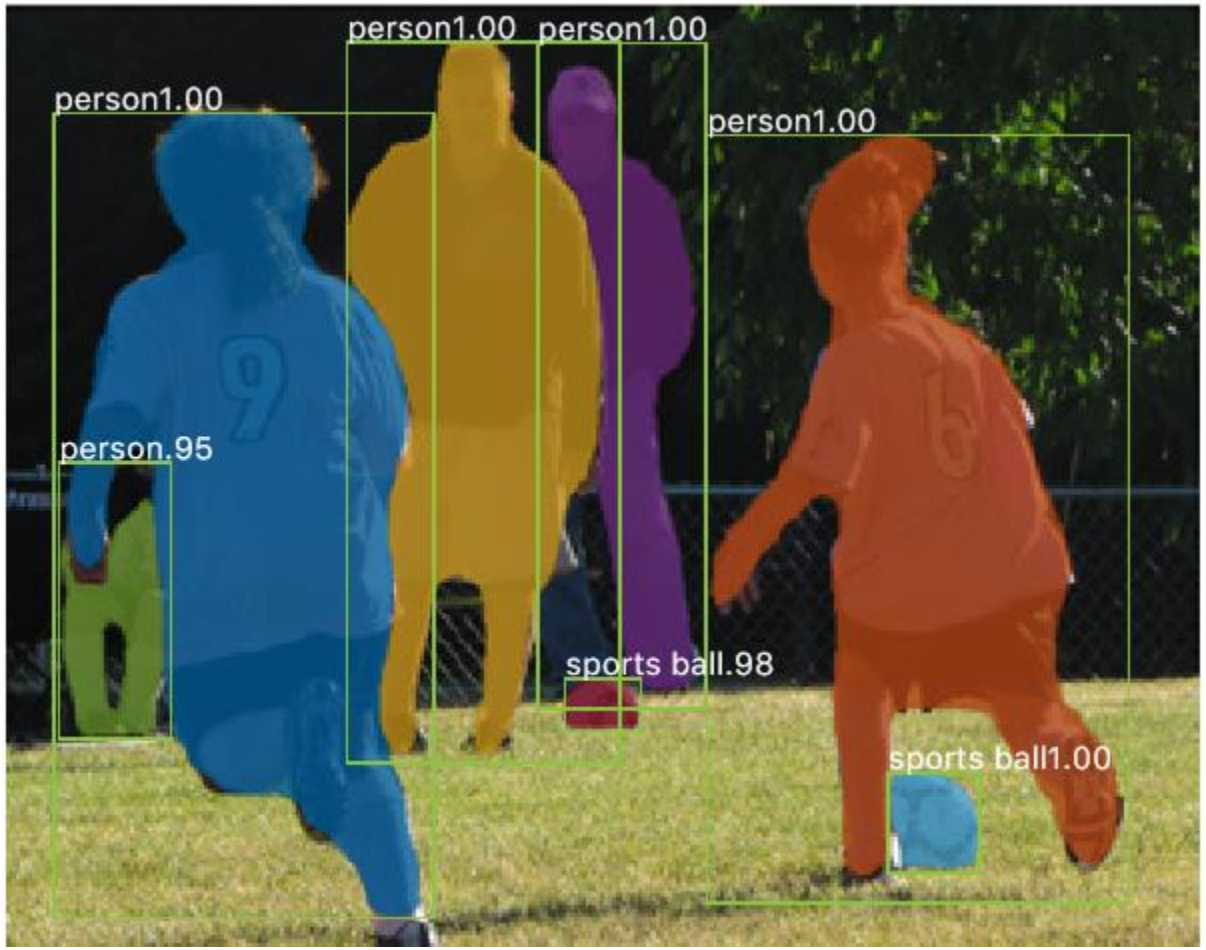
- Image classification

- Object detection



Different tasks

- Image classification
- Object detection
- Instance segmentation(mask)

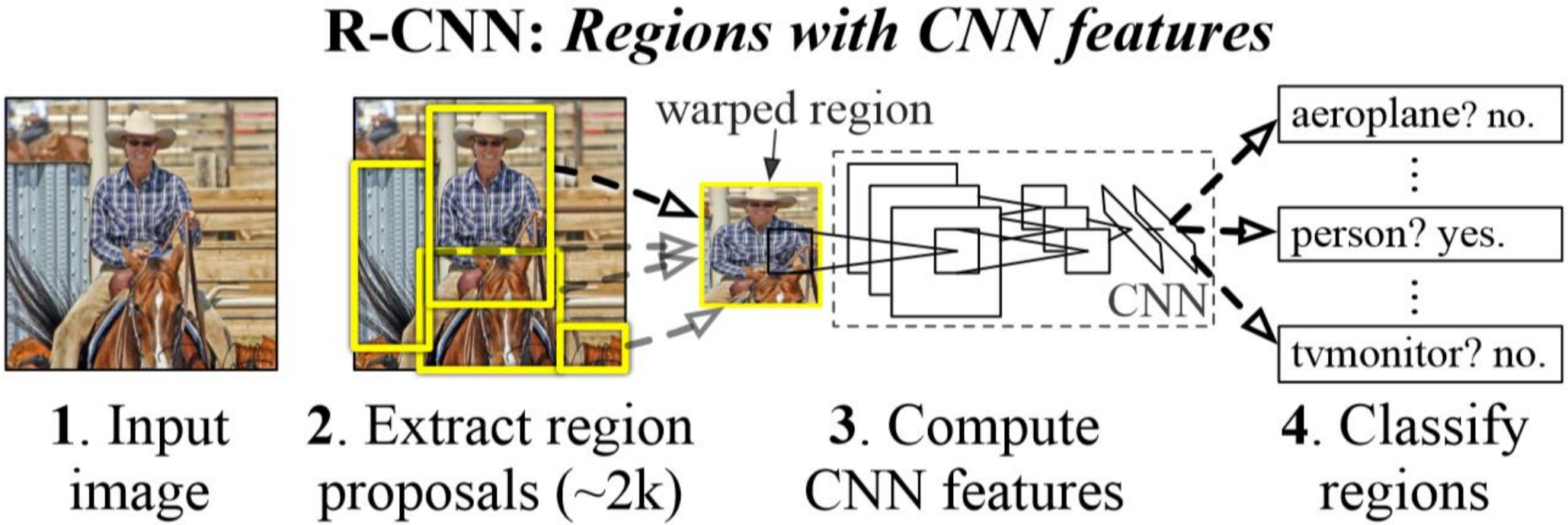


Different tasks

- Image classification
- Object detection
- Instance segmentation(mask)
- Keypoint detection

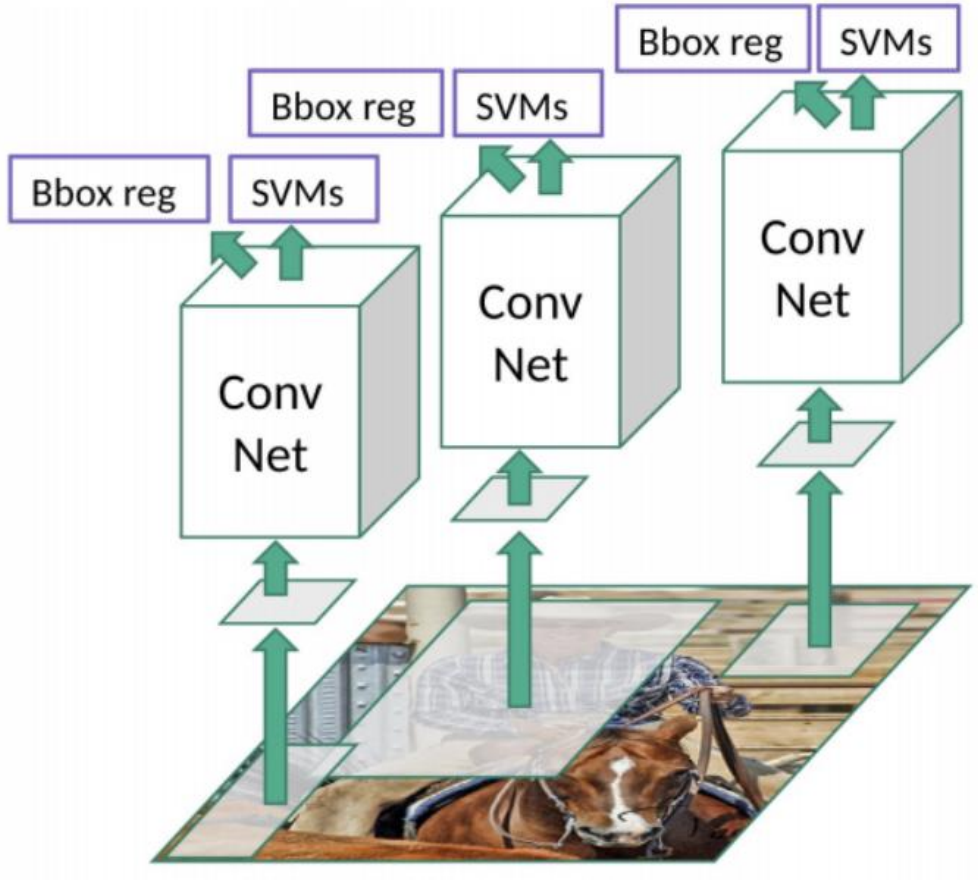
R-cnn

- **Inputs**: image
- **Outputs**: Bounding boxes(Bbox) + labels for each object in the image



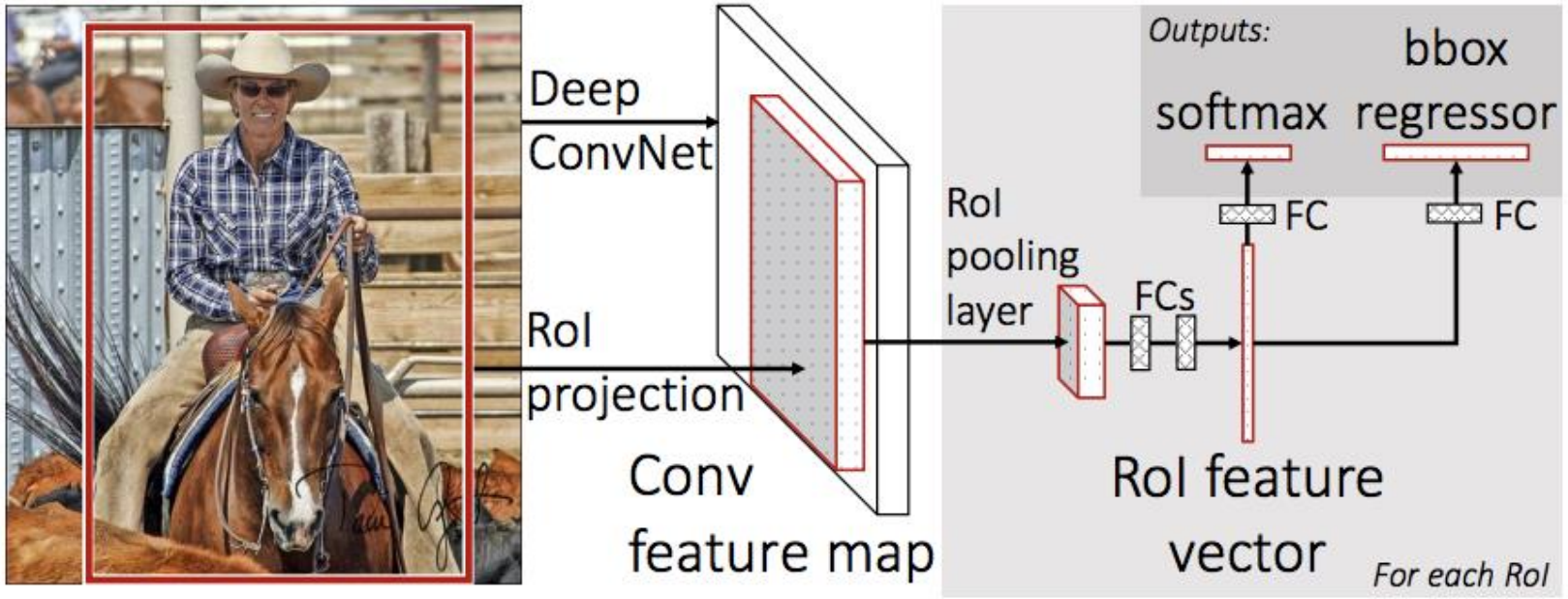
R-cnn Problem

- Use extra(traditional) algorithm to propose Bbox
  - Can’t learn and may generate bad proposal Bboxes
- Time-consuming
  - Selective search
  - Cnn for each Bbox



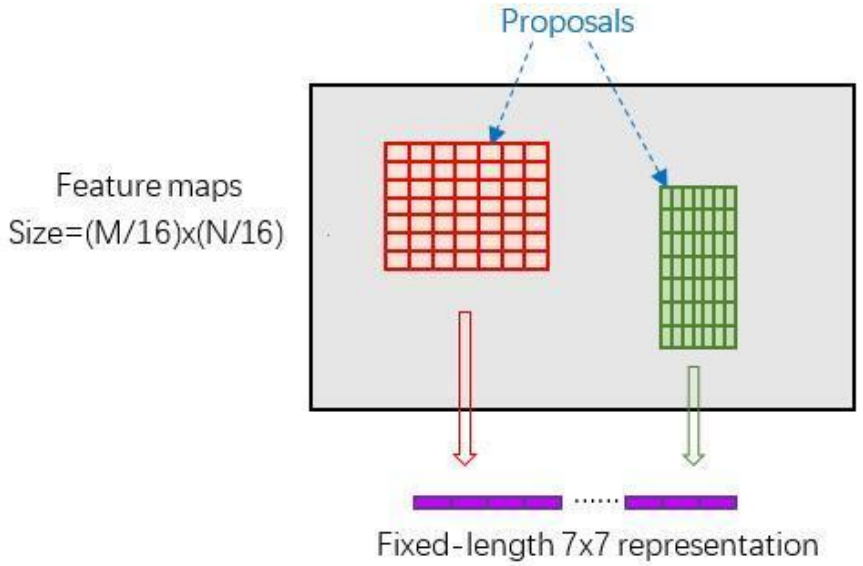
Fast Rcnn

- Feature proposal on feature map
  - Use RoI pooling
  - Use softmax to classify
  - Still use selective search



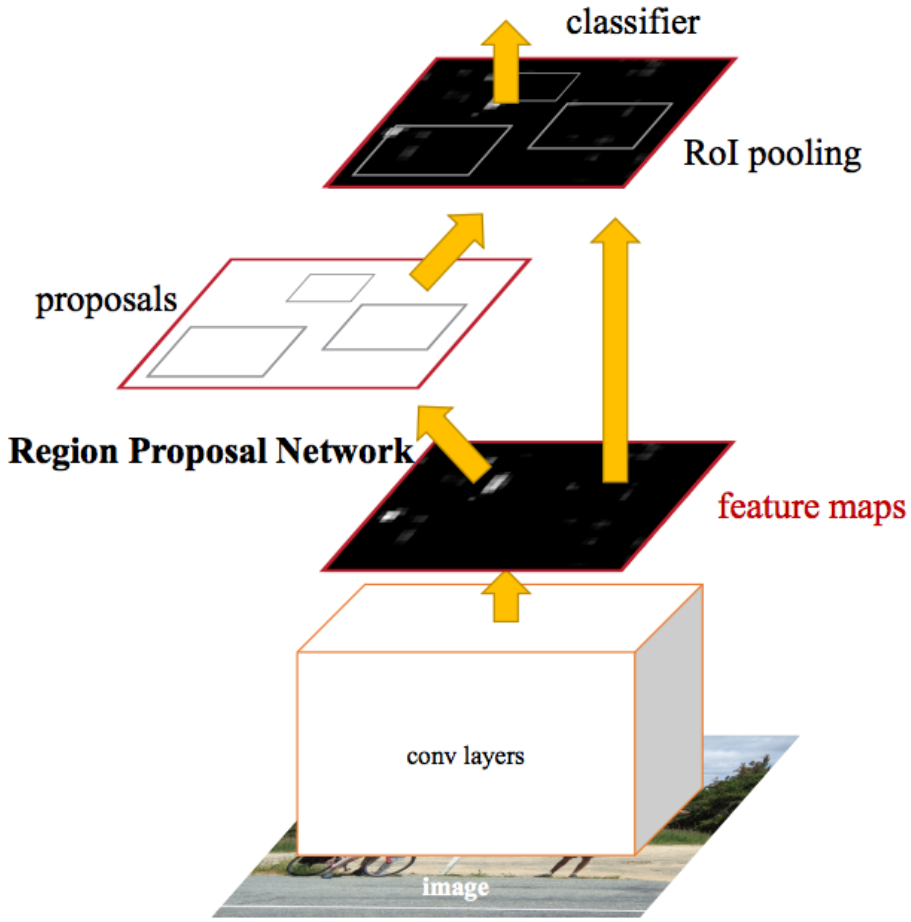
Fast Rcnn

- RoI pooling
  - divide the h x w RoI as H x W sub-windows
  - max-pool each sub-window to get H x W map to represent the ROI. (The max-pool kernel is [h/H], [w/W] respectively).
- RoI Align
  - the value of the four regularly sampled locations are computed directly through bilinear interpolation

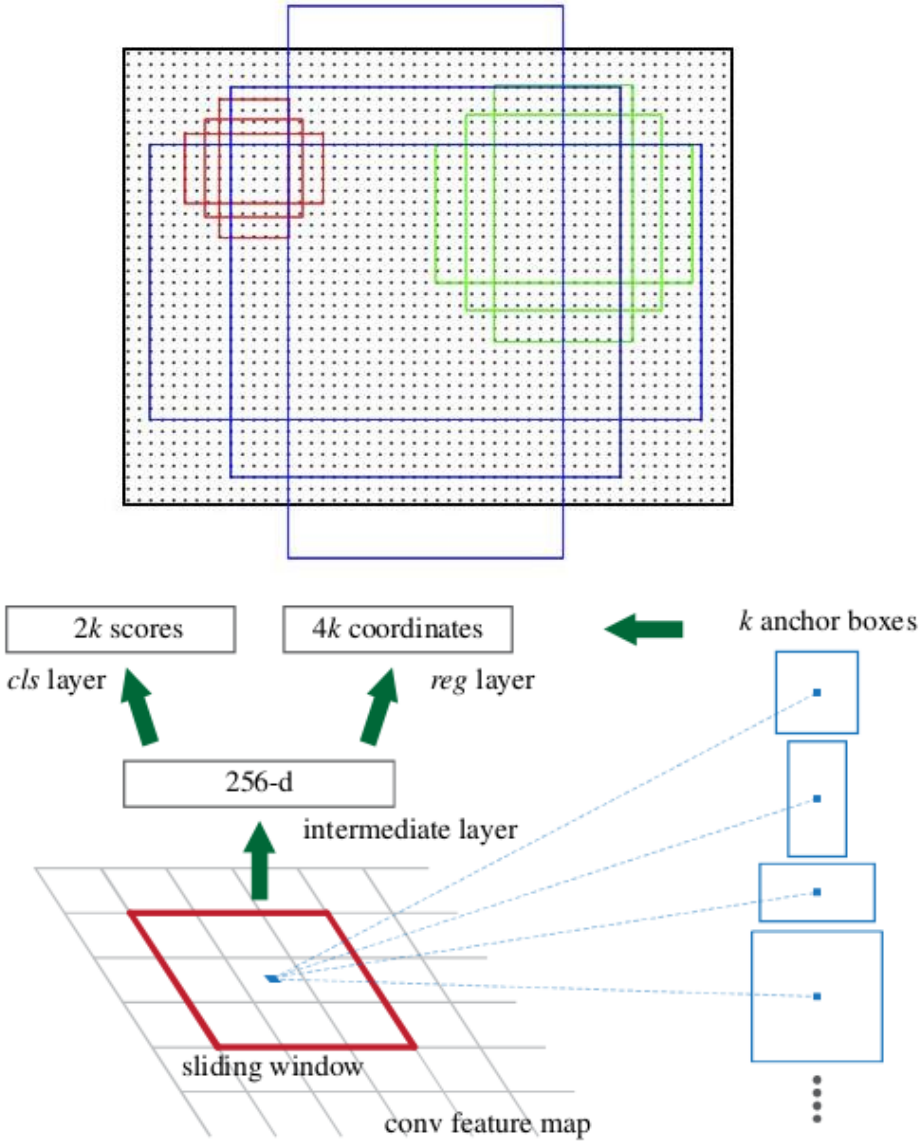


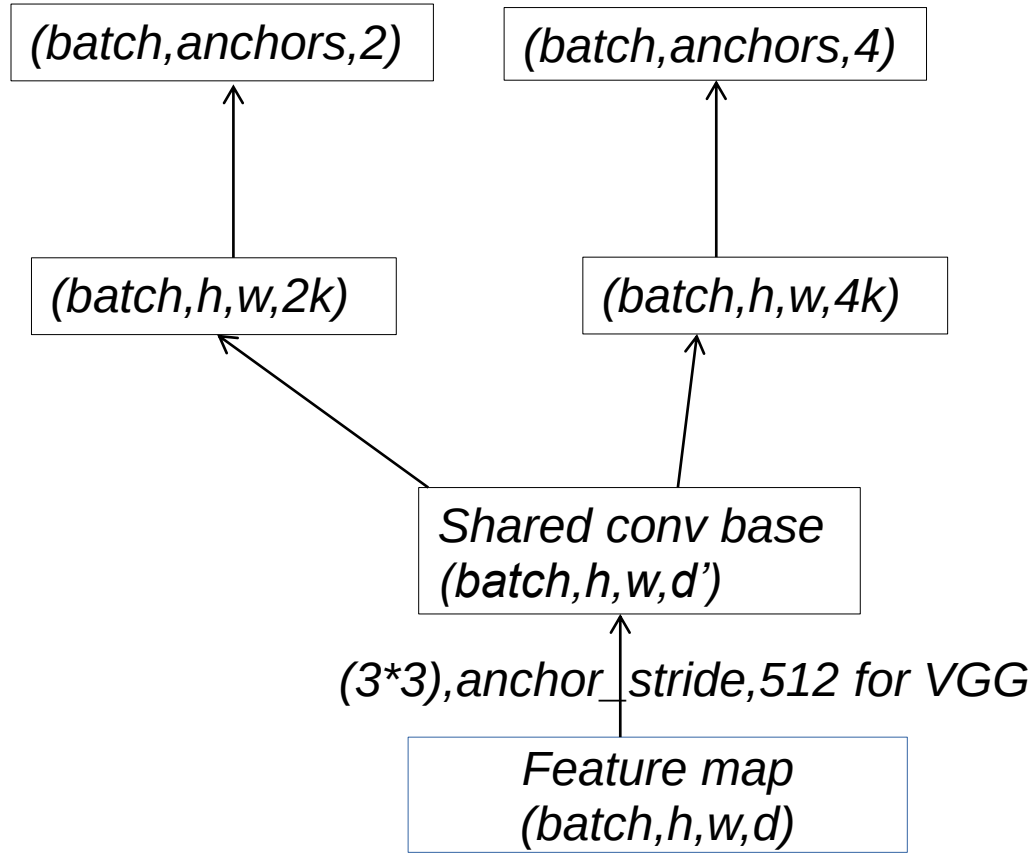
Faster Rcnn

- Use network to propose
  - Reuse the feature map



Faster Rcnn: RPN





Faster Rcnn: RPN

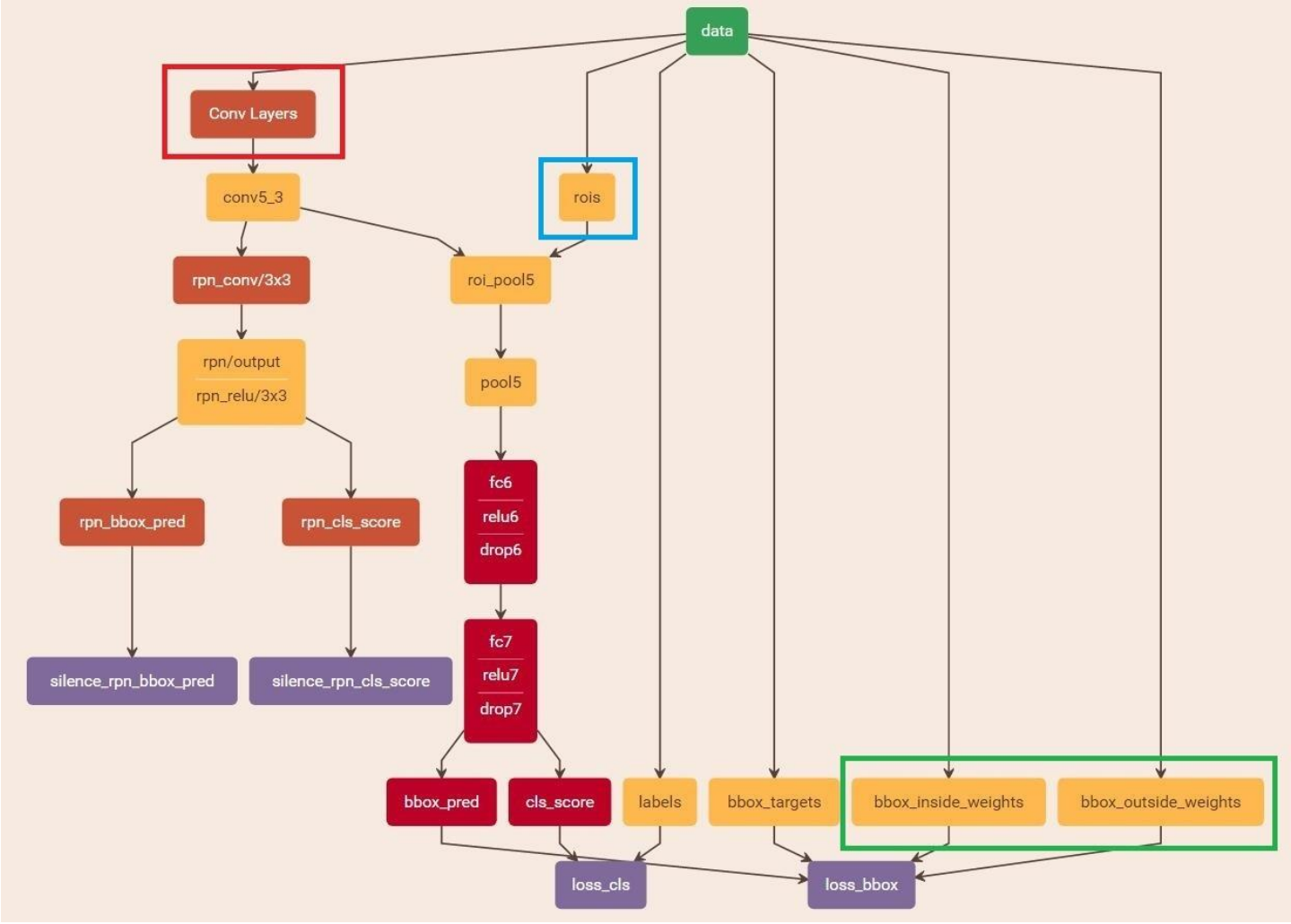
- Label
  - Intersection over Union(IoU)
  - positive
    i. the anchor/anchors with the highest IoU overlapwithaground-truthbox
    ii. an anchor that has an IoU overlap higher than 0.7 with any ground-truth box
  - negative
    - an anchor that has an IoU overlap lower than 0.3 with all ground-truth box

- Test
  - Non-maximum-suppression based on *cls* scores

Faster R-cnn: training

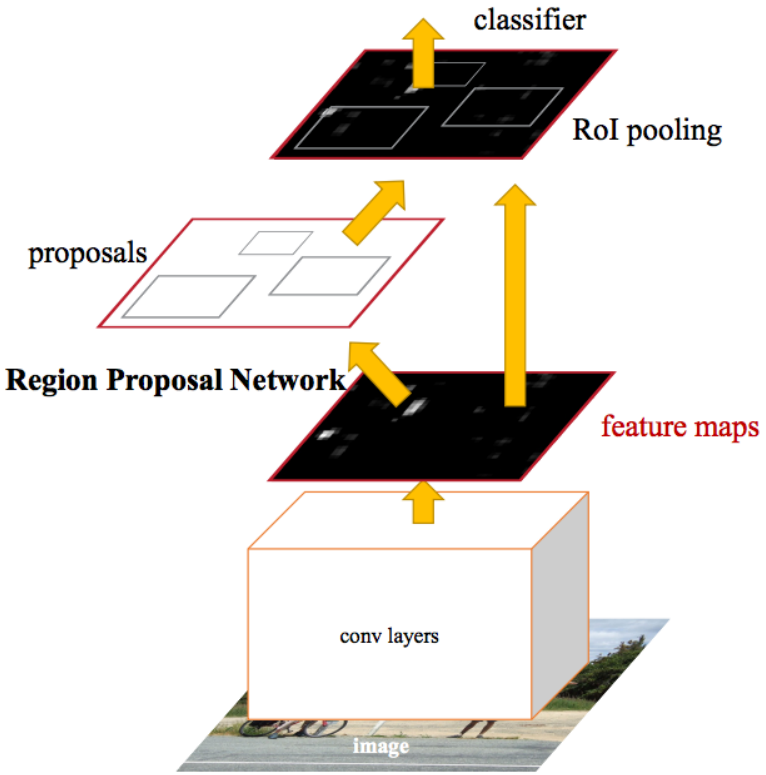
- Step1:(model1)
  - Train RPN network initialized by imageNet-pre-trained model weights
- Step2:(model2)
  - Train fast rcnn(initialized by imageNet-pre-trained model) with RPN network(model1)
- Step3:(model3)
  - Fine-tune RPN with fixed cnn initialized by model2’s cnn weights
- Step4:
  - Fine-tune model2 with model3’s region proposals and fix cnn weights

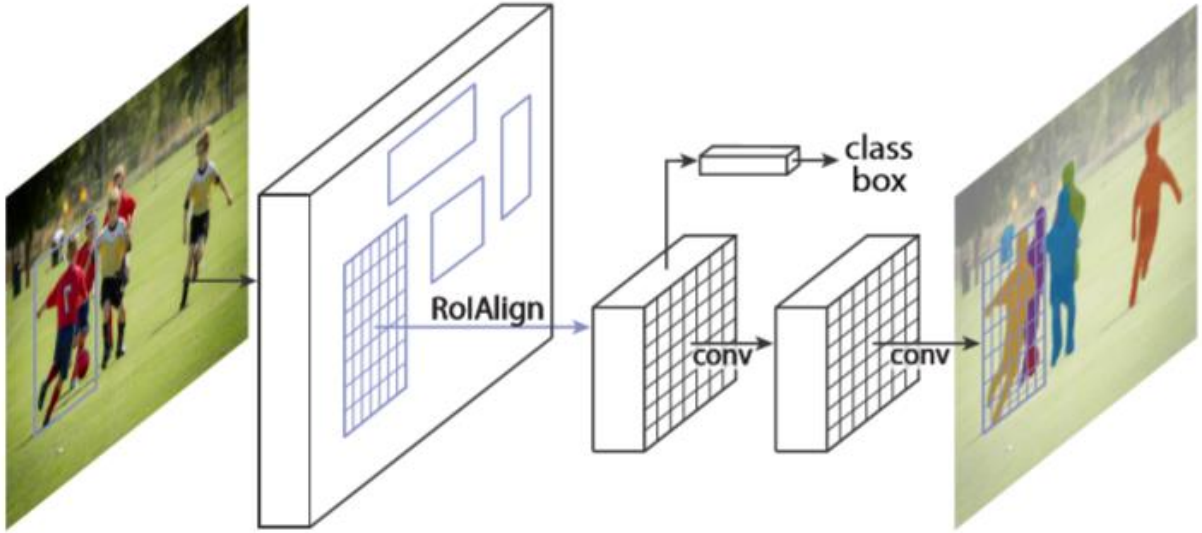
Faster R-cnn

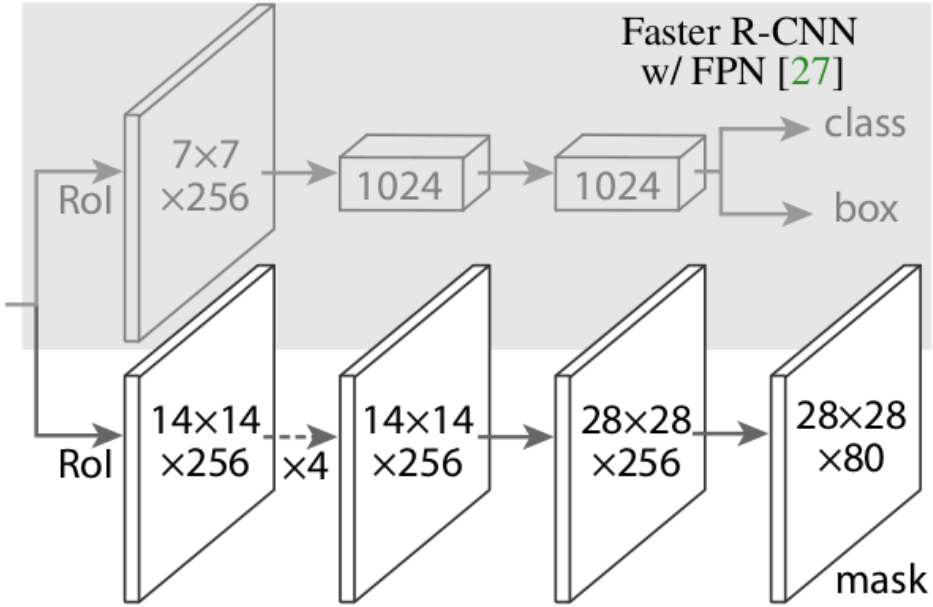

data
Conv Layers
conv5_3
rois
rpn_conv/3x3
roi_pool5
rpn/output
rpn_relu/3x3
pool5
fc6
relu6
drop6
rpn_bbox_pred
rpn_cls_score
fc7
relu7
drop7
silence_rpn_bbox_pred
silence_rpn_cls_score
bbox_pred
cls_score
labels
bbox_targets
bbox_inside_weights
bbox_outside_weights
loss_cls
loss_bbox

Mask R-Cnn

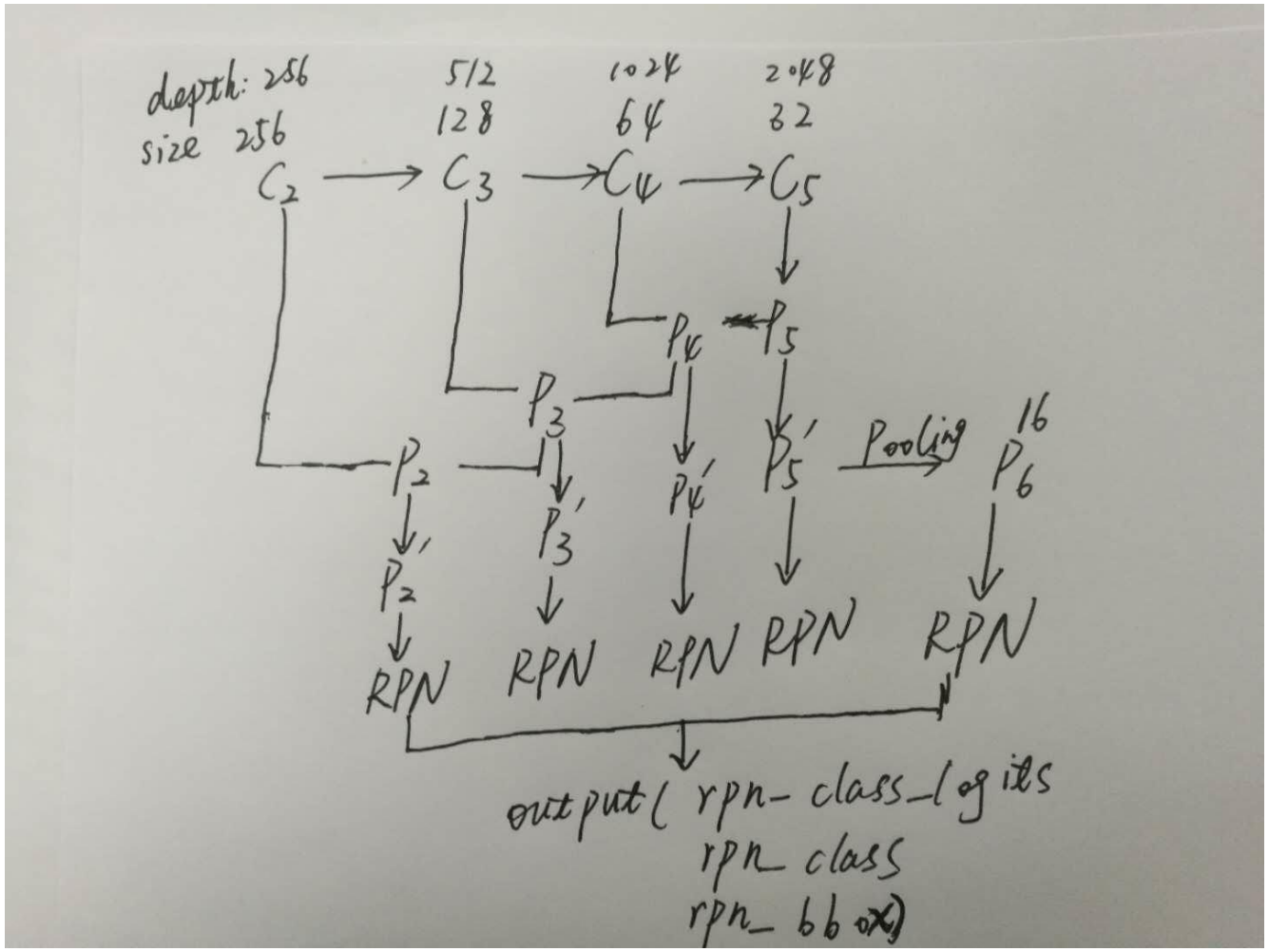
- Extending Faster R-CNN for Pixel Level Segmentation
- Use RoI Align







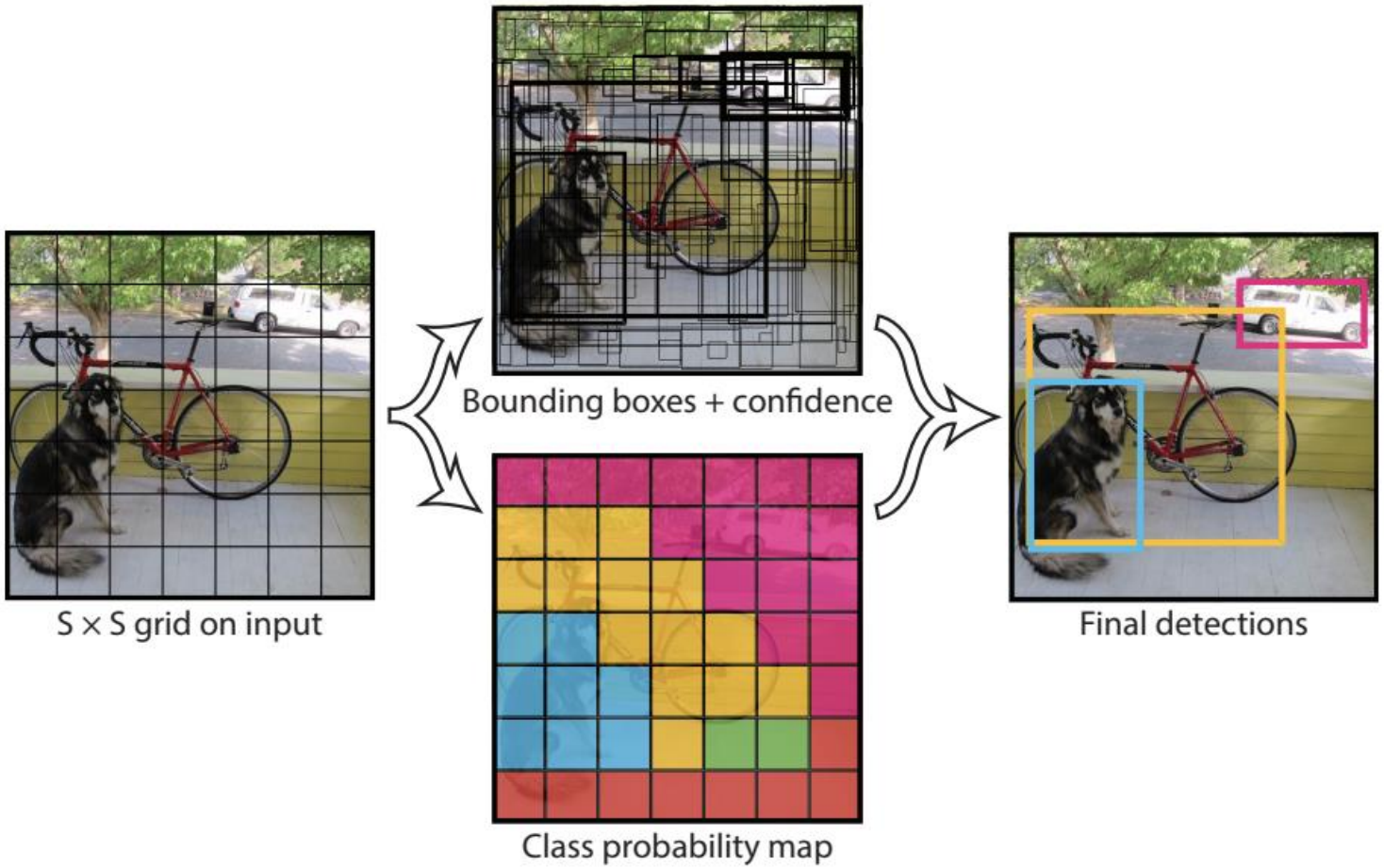
FPN(feature pyramid network)



roi_align_classifier: PyramidROIAlign	input:	[(None, 200, 4), (None, 256, 256, 256), (None, 128, 128, 256), (None, 64, 64, 256), (None, 32, 32, 256)]
	output:	(None, 200, 7, 7, 256)

You Only Look Once

- Pro
  - Simple and fast
- Con:
  - Lower accuracy than state-of-the art
  - Difficult to detect the small object



reference

- R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Rich feature hierarchies for accurate object detection and semantic segmentation,” in CVPR, 2014.
- K. He, X. Zhang, S. Ren, and J. Sun, “Spatial pyramid pooling in deep convolutional networks for visual recognition,” in European Conference on Computer Vision (ECCV), 2014.
- R. Girshick. Fast R-CNN. In ICCV, 2015.
- S. Ren, K. He, R. Girshick, and J. Sun. Faster R-CNN: Towards real-time object detection with region proposal networks. In NIPS, 2015
- K. He et al. “Mask R-CNN.” 2017 IEEE International Conference on Computer Vision (ICCV) (2017): 2980-2988.
- Lin, Tsung-Yi et al. “Feature Pyramid Networks for Object Detection.” 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016): 936-944.
- J. Redmon et al. “You Only Look Once: Unified, Real-Time Object Detection.” 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2015): 779-788.
- https://zhuanlan.zhihu.com/p/31426458
- https://github.com/matterport/Mask_RCNN