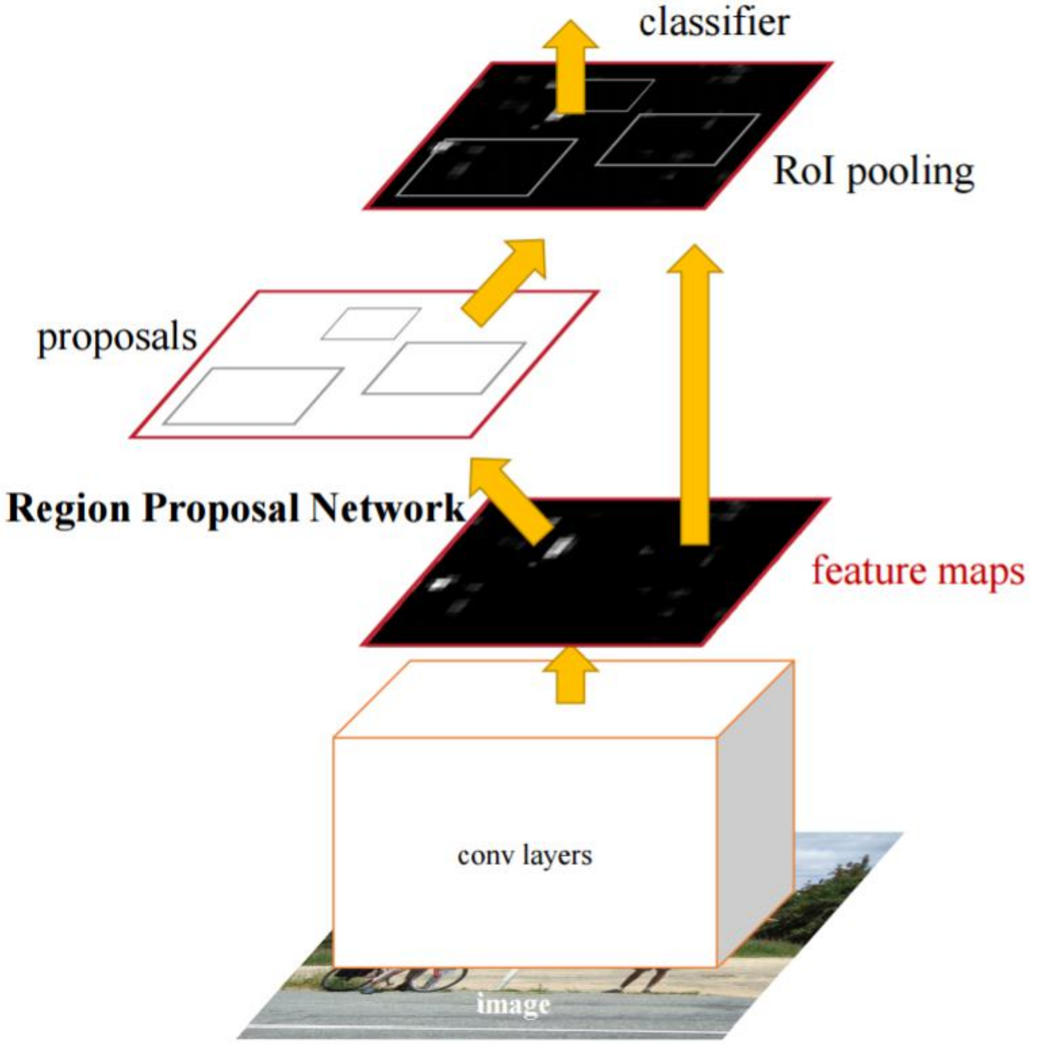
I lost my Faster R-CNN's report and YOLO's report when I reinstalled the Ubuntu,so this is a short vision I rewrote.Sorry for that.

# INTRODUCTION

Faster R-CNN, is composed of two modules. The first module is a deep fully convolutional network that proposes regions, and the second module is the Fast R-CNN detector that uses the proposed regions.Using the recently popular terminology of neural networks with 'attention' mechanisms, the RPN module tells the Fast R-CNN module where to look.



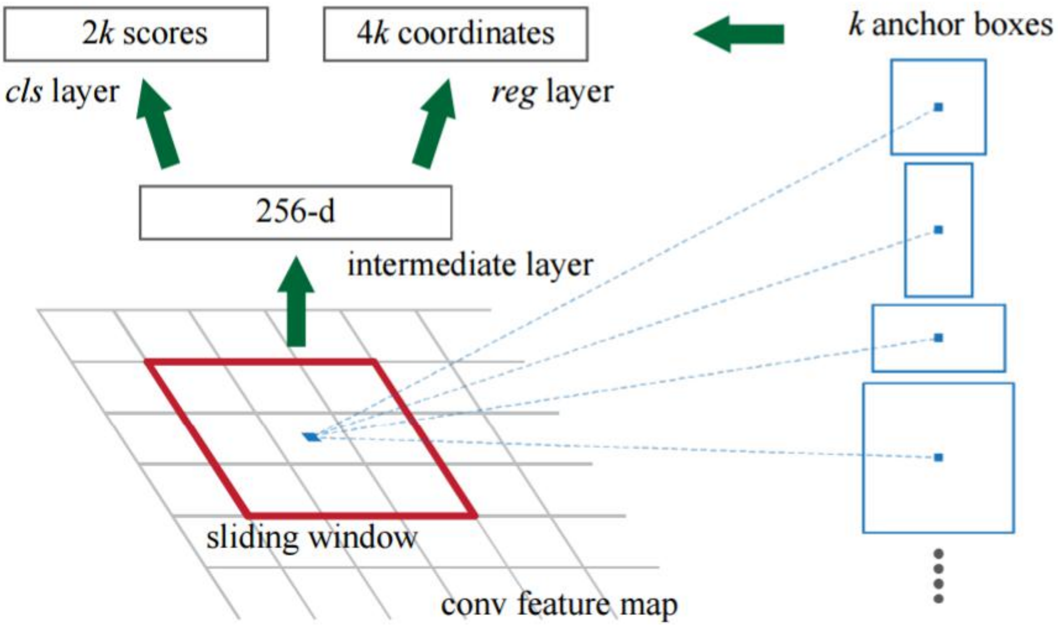
# Implementing Details

## Region Proposal Networks

A Region Proposal Network (RPN) takes an image (of any size) as input and outputs a set of rectangular object proposals, each with an objectness score. This small network takes as input an n × n spatial window of the input convolutional feature map. Each sliding window is mapped to a lower-dimensional feature. This feature is fed into two sibling fullyconnected layers—a box-regression layer (reg) and a box-classification layer (cls). We use n = 3 in this paper.

## Anchors

They choose some anchors by combining 3 scales and 3 aspect ratios.The sacles and aspect ratios are picked up manually.YOLO have provided a more stable way to picked.Their design of anchors provided a more cost-efficient way .



## Loss Function

$$L(\{p_i\},\{t_i\}) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) \\ + \lambda \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}(t_i, t_i^*).$$

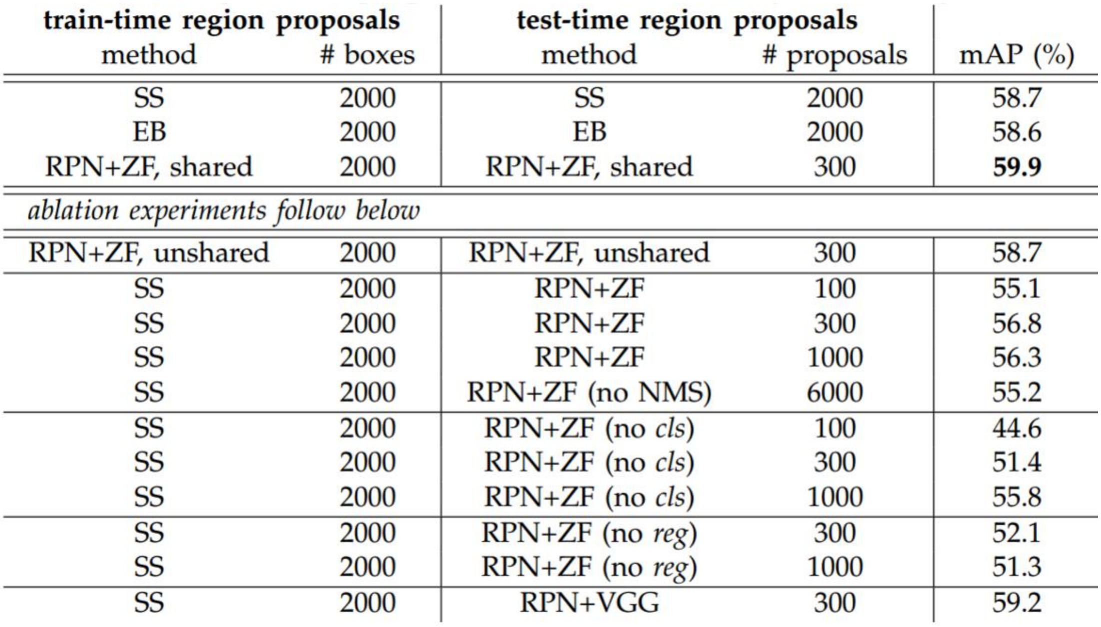
$$\begin{aligned} t_{\mathrm{x}} &= (x - x_{\mathrm{a}})/w_{\mathrm{a}}, \quad t_{\mathrm{y}} = (y - y_{\mathrm{a}})/h_{\mathrm{a}}, \\ t_{\mathrm{w}} &= \log(w/w_{\mathrm{a}}), \quad t_{\mathrm{h}} = \log(h/h_{\mathrm{a}}), \\ t_{\mathrm{x}}^* &= (x^* - x_{\mathrm{a}})/w_{\mathrm{a}}, \quad t_{\mathrm{y}}^* = (y^* - y_{\mathrm{a}})/h_{\mathrm{a}}, \\ t_{\mathrm{w}}^* &= \log(w^*/w_{\mathrm{a}}), \quad t_{\mathrm{h}}^* = \log(h^*/h_{\mathrm{a}}), \end{aligned}$$

## 4-Step Alternating Training

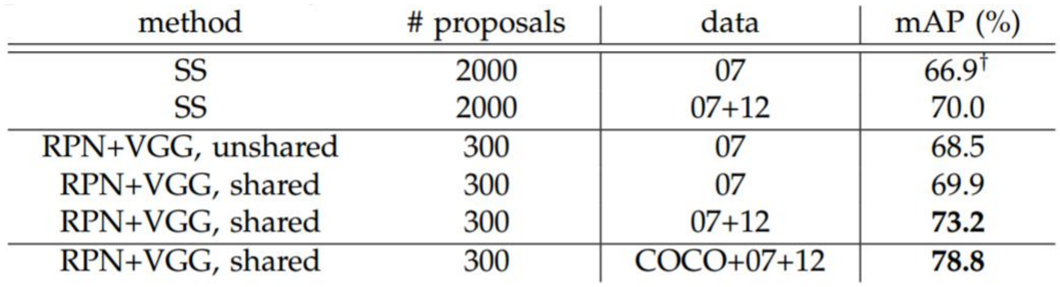
The RPN can be trained end-to-end by backpropagation and stochastic gradient descent (SGD).

1. Train the RPN.
2. Train a separate detection network by Fast R-CNN
3. Use the detector network to initialize RPN training, but fix the shared convolutional layers and only fine-tune the layers unique to RPN.
4. Keeping the shared convolutional layers fixed, fine-tune the unique layers of Fast R-CNN.

# Experiments

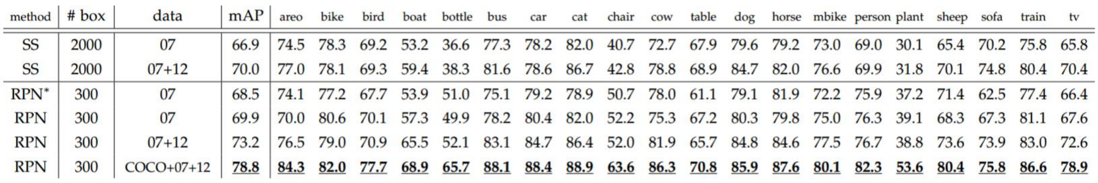


| train-time region proposals | | test-time region proposals | | |
|---|---|---|---|---|
| method | # boxes | method | # proposals | mAP (%) |
| SS | 2000 | SS | 2000 | 58.7 |
| EB | 2000 | EB | 2000 | 58.6 |
| RPN+ZF, shared | 2000 | RPN+ZF, shared | 300 | **59.9** |
| *ablation experiments follow below* | | | | |
| RPN+ZF, unshared | 2000 | RPN+ZF, unshared | 300 | 58.7 |
| SS | 2000 | RPN+ZF | 100 | 55.1 |
| SS | 2000 | RPN+ZF | 300 | 56.8 |
| SS | 2000 | RPN+ZF | 1000 | 56.3 |
| SS | 2000 | RPN+ZF (no NMS) | 6000 | 55.2 |
| SS | 2000 | RPN+ZF (no *cls*) | 100 | 44.6 |
| SS | 2000 | RPN+ZF (no *cls*) | 300 | 51.4 |
| SS | 2000 | RPN+ZF (no *cls*) | 1000 | 55.8 |
| SS | 2000 | RPN+ZF (no *reg*) | 300 | 52.1 |
| SS | 2000 | RPN+ZF (no *reg*) | 1000 | 51.3 |
| SS | 2000 | RPN+VGG | 300 | 59.2 |

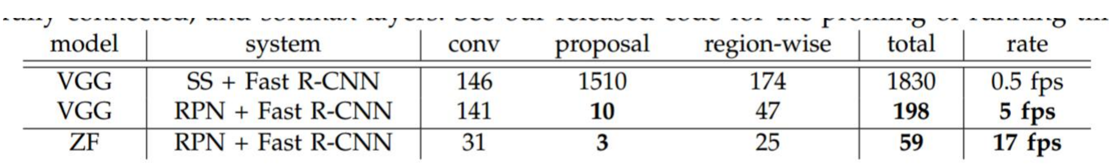


| method | # proposals | data | mAP (%) |
|---|---|---|---|
| SS | 2000 | 07 | $66.9^{\dagger}$ |
| SS | 2000 | 07+12 | 70.0 |
| RPN+VGG, unshared | 300 | 07 | 68.5 |
| RPN+VGG, shared | 300 | 07 | 69.9 |
| RPN+VGG, shared | 300 | 07+12 | **73.2** |
| RPN+VGG, shared | 300 | COCO+07+12 | **78.8** |

## This display the effect of sharing convulition layers.

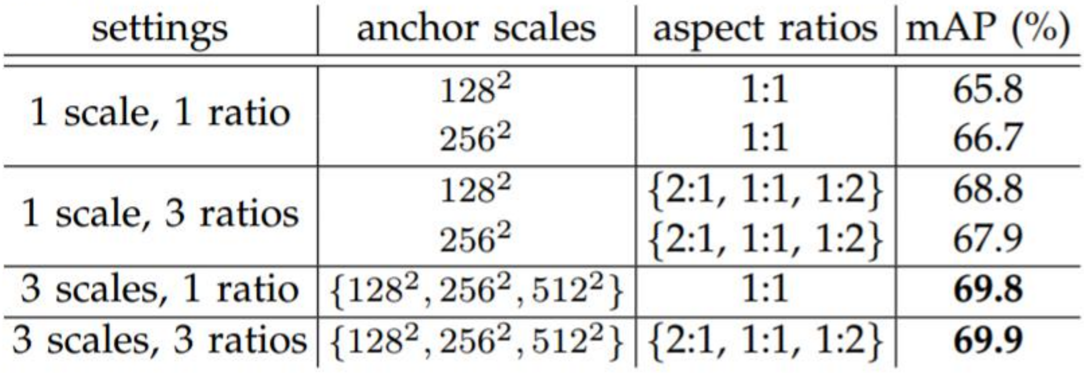


| method | # box | data | mAP | areo | bike | bird | boat | bottle | bus | car | cat | chair | cow | table | dog | horse | mbike | person | plant | sheep | sofa | train | tv |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| SS | 2000 | 07 | 66.9 | 74.5 | 78.3 | 69.2 | 53.2 | 36.6 | 77.3 | 78.2 | 82.0 | 40.7 | 72.7 | 67.9 | 79.6 | 79.2 | 73.0 | 69.0 | 30.1 | 65.4 | 70.2 | 75.8 | 65.8 |
| SS | 2000 | 07+12 | 70.0 | 77.0 | 78.1 | 69.3 | 59.4 | 38.3 | 81.6 | 78.6 | 86.7 | 42.8 | 78.8 | 68.9 | 84.7 | 82.0 | 76.6 | 69.9 | 31.8 | 70.1 | 74.8 | 80.4 | 70.4 |
| RPN* | 300 | 07 | 68.5 | 74.1 | 77.2 | 67.7 | 53.9 | 51.0 | 75.1 | 79.2 | 78.9 | 50.7 | 78.0 | 61.1 | 79.1 | 81.9 | 72.2 | 75.9 | 37.2 | 71.4 | 62.5 | 77.4 | 66.4 |
| RPN | 300 | 07 | 69.9 | 70.0 | 80.6 | 70.1 | 57.3 | 49.9 | 78.2 | 80.4 | 82.0 | 52.2 | 75.3 | 67.2 | 80.3 | 79.8 | 75.0 | 76.3 | 39.1 | 68.3 | 67.3 | 81.1 | 67.6 |
| RPN | 300 | 07+12 | 73.2 | 76.5 | 79.0 | 70.9 | 65.5 | 52.1 | 83.1 | 84.7 | 86.4 | 52.0 | 81.9 | 65.7 | 84.8 | 84.6 | 77.5 | 76.7 | 38.8 | 73.6 | 73.9 | 83.0 | 72.6 |
| RPN | 300 | COCO+07+12 | **78.8** | **84.3** | **82.0** | **77.7** | **68.9** | **65.7** | **88.1** | **88.4** | **88.9** | **63.6** | **86.3** | **70.8** | **85.9** | **87.6** | **80.1** | **82.3** | **53.6** | **80.4** | **75.8** | **86.6** | **78.9** |

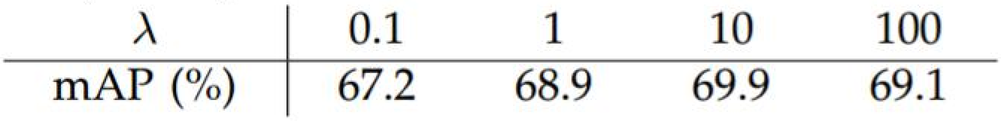


| model | system | conv | proposal | region-wise | total | rate |
|---|---|---|---|---|---|---|
| VGG | SS + Fast R-CNN | 146 | 1510 | 174 | 1830 | 0.5 fps |
| VGG | RPN + Fast R-CNN | 141 | **10** | 47 | **198** | **5 fps** |
| ZF | RPN + Fast R-CNN | 31 | **3** | 25 | **59** | **17 fps** |

in Table 8.

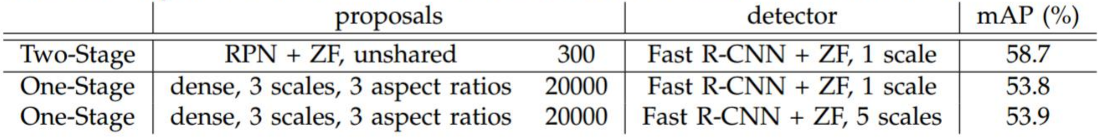


| settings | anchor scales | aspect ratios | mAP (%) |
|---|---|---|---|
| 1 scale, 1 ratio | $128^2$ | 1:1 | 65.8 |
| | $256^2$ | 1:1 | 66.7 |
| 1 scale, 3 ratios | $128^2$ | {2:1, 1:1, 1:2} | 68.8 |
| | $256^2$ | {2:1, 1:1, 1:2} | 67.9 |
| 3 scales, 1 ratio | $\{128^2, 256^2, 512^2\}$ | 1:1 | **69.8** |
| 3 scales, 3 ratios | $\{128^2, 256^2, 512^2\}$ | {2:1, 1:1, 1:2} | **69.9** |



| $\lambda$ | 0.1 | 1 | 10 | 100 |
|---|---|---|---|---|
| mAP (%) | 67.2 | 68.9 | 69.9 | 69.1 |

test set using the ZF model and Fast R-CNN. RPN uses unshared features.



| | proposals | | detector | mAP (%) |
|---|---|---|---|---|
| Two-Stage | RPN + ZF, unshared | 300 | Fast R-CNN + ZF, 1 scale | 58.7 |
| One-Stage | dense, 3 scales, 3 aspect ratios | 20000 | Fast R-CNN + ZF, 1 scale | 53.8 |
| One-Stage | dense, 3 scales, 3 aspect ratios | 20000 | Fast R-CNN + ZF, 5 scales | 53.9 |