IE534: Deep Learning – Course Project
Spring 2021
Faster R-CNN Implementation

Group 16:

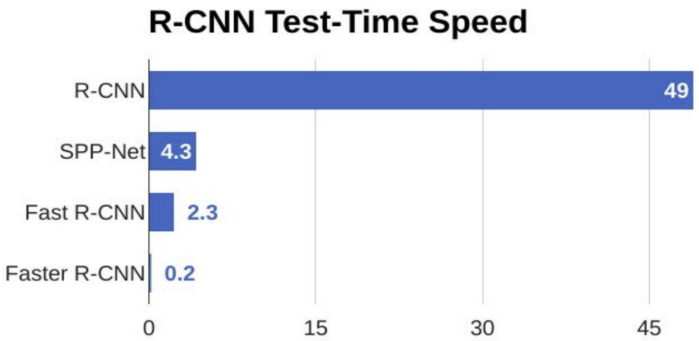
Refik Mert Cam (rcam2)

Sanyukta Deshpande (spd4)

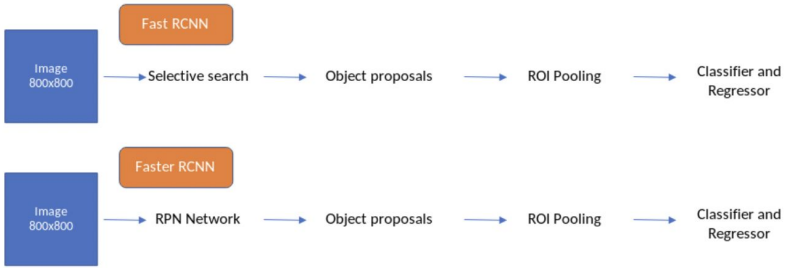
Abhinav Garg (garg19)

Background

- Aim: Real-time object detection within an image
- Belongs to RCNN family

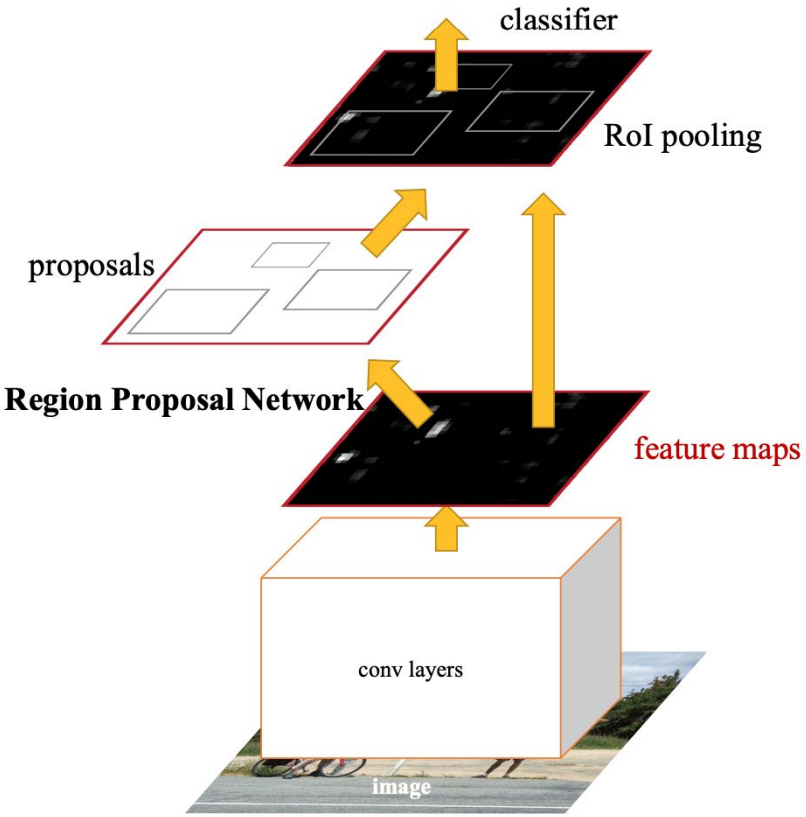


Comparison of test-time speed of object detection algorithms



Fast RCNN and Faster RCNN

Faster-RCNN Architecture

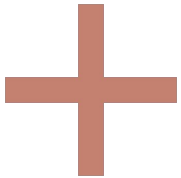


Algorithm

1. Feature extraction from the image
2. Creating anchor targets
3. Locations and objectness score prediction from the RPN network
4. Taking the top N locations and their objectness scores (proposal layer)
5. Passing these top N locations through Fast R-CNN network and generating locations and class predictions for each location suggested in 4
6. Generating proposal targets for each location suggested in 4
7. Use 2 and 3 to calculate **rpn_cls_loss** and **rpn_reg_loss**
8. Use 5 and 6 to calculate **roi_cls_loss** and **roi_reg_loss**

Our Work and Contributions

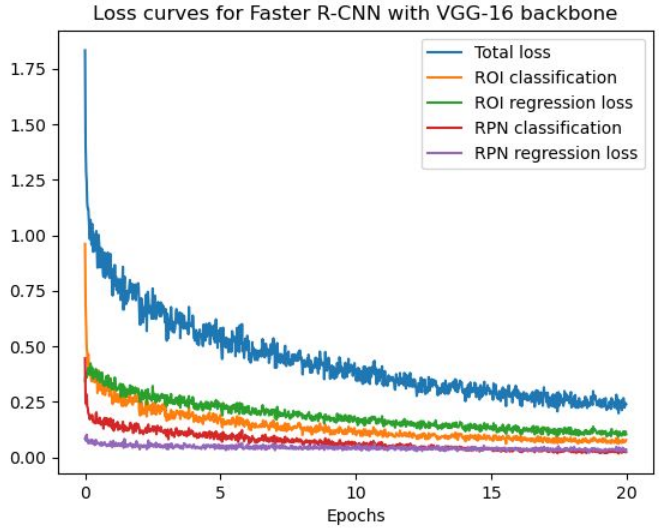
Implemented the Faster- RCNN architecture with approximate joint training approach

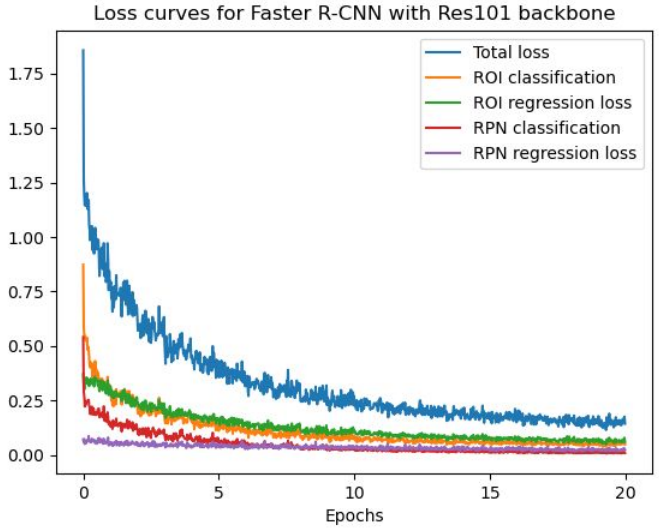


Used ResNet101 in addition to VGG16.

Used multiple anchor settings for training

Details and Results





- <u>Training strategy:</u>
  - Online learning
  - SGD with initial learning rate 1e-3 and learning rate decay at 8th epoch with 0.1
  - Data - 9,963 images with 24,640 objects over 20 classes (train and test split stratified by class)

Details and Results

Overall

Backbone network	mAP (%)	Rate (testing mode)
VGG16	67.6	13 fps
Res101	61.4	13 fps
Original Paper	69.9	17 fps

Per Class

	fly	bike	bird	boat	pin	bus	car	cat	chair	cow	table	dog	horse	moto	person	plant	sheep	sofa	train	tv
VGG16	69.7	78.8	65.4	51.8	53.3	77.1	78.5	79.5	46.2	71.4	66.2	75.7	78.9	74.0	75.2	36.0	65.3	61.5	73.2	73.8
Res101	69.0	69.8	55.6	48.5	34.7	68.0	70.4	78.1	40.1	65.6	55.1	75.8	78.1	66.6	68.8	33.1	59.3	60.0	74.1	54.2

Details and Results

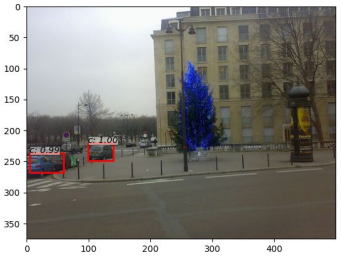
VGG16

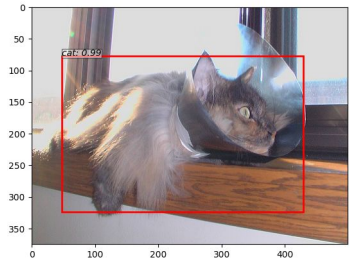
Anchor scales	Anchor ratios	mAP (%)
128	1:1	64.1
128	2:1, 1:1, 1:2	64.3
128, 256, 512	1:1	65.0
128, 256, 512	2:1, 1:1, 1:2	67.6

Res101

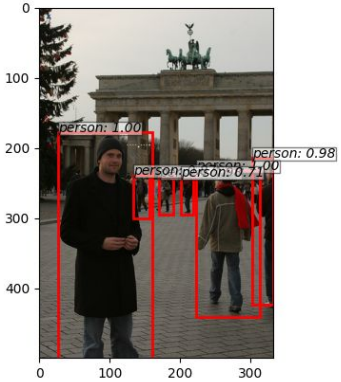
Anchor scales	Anchor ratios	mAP (%)
128	1:1	61.5
128	2:1, 1:1, 1:2	61.7
128, 256, 512	1:1	61.8
128, 256, 512	2:1, 1:1, 1:2	61.4

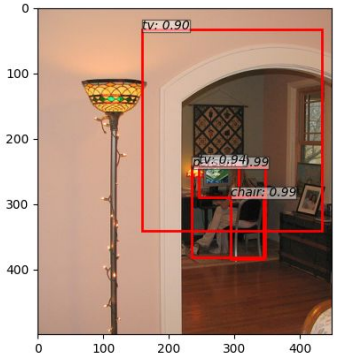
Selected Sample Results

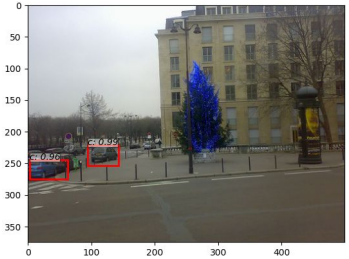


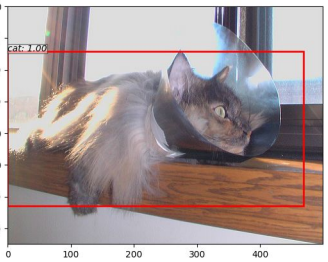


VGG-16 (Backbone)









Res101 (Backbone)

Conclusion

- Successfully implemented the Faster RCNN architecture
- Training time - 20 minutes per epoch
- Total GPU hours used - 53.3
- Test time rate - 13 fps
- VGG16 with 9 anchors performed best with a mAP value of 67.6%
- Further improvement can be made by tuning the hyperparameters, trying 2-step alternate training approach and training for longer periods