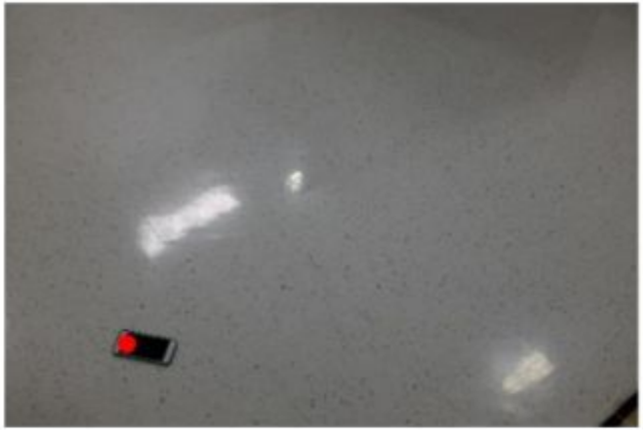
Assignment 1: Object Detection

Tahsin Reasat

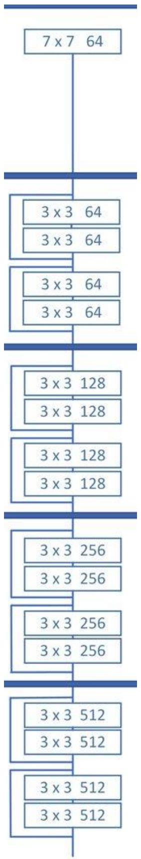
Task

Given an image of a mobile on a background, detect the location of the mobile.

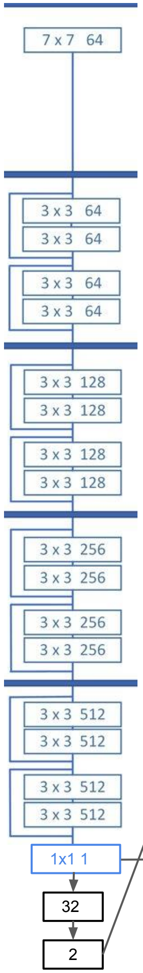


122.jpg

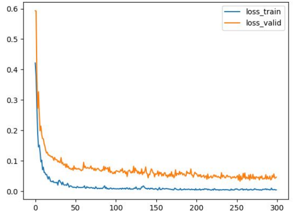
Network Architecture Rationale

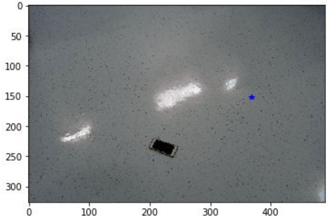


Backbone - ResNet18*

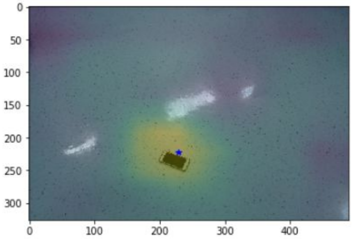


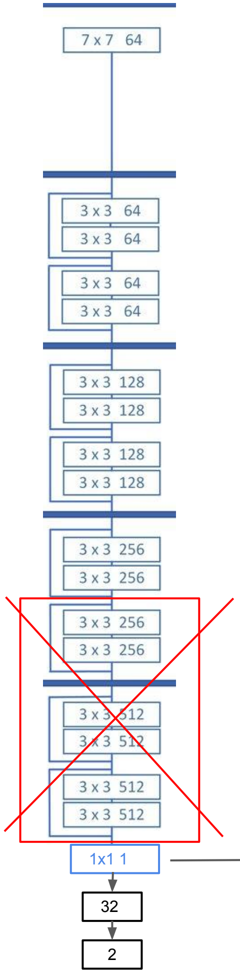
Initial Architecture





16 x 11 activation map





31 x 21 activation map

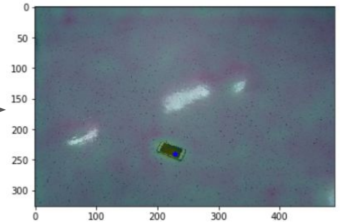


*Fig from Improved Selective Refinement Network for Face Detection. In reality, ResNet18 has 20 conv layers and 1 fc layer

Preprocessing

- Input RGB normalized according to ImageNet standards

  mean: (0.485, 0.456, 0.406),

  std: (0.229, 0.224, 0.225)

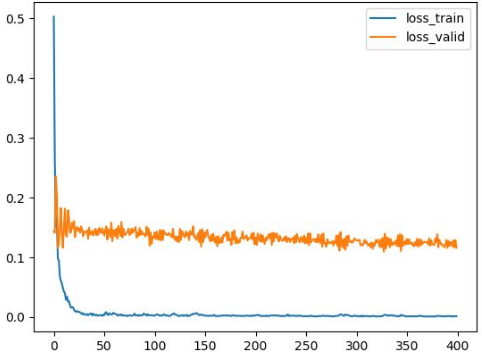
- Output x, y coordinates normalized to image dimension
- Augmentation: None

Parameters

- Optimizer: Adam
- Learning rate : 0.001
- Loss function: Mean Square Error
- Epoch Number: 400
- Batch Size: 32
- OS: Windows
- Specs: RTX 2070, AMD Ryzen 5 2600

Results: Loss Plots

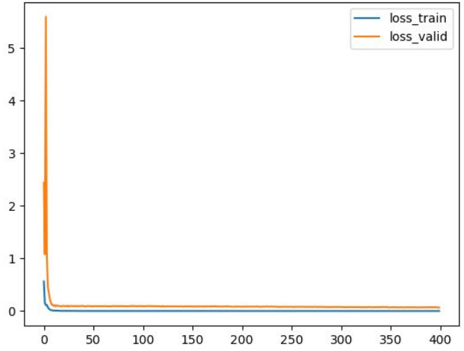
Loss



Epoch

Trained 400 epochs only updating the last conv+fc layers

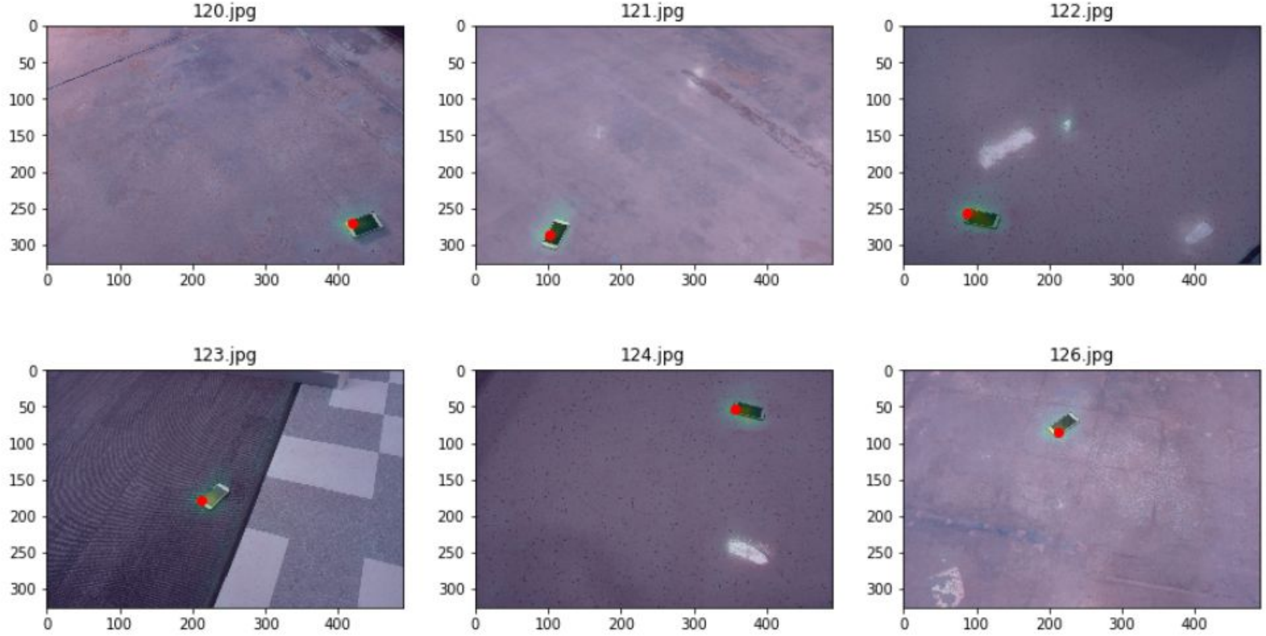
Loss



Epoch

Trained Another 400 epochs updating all the layers

Results: Test Output



Results: Test Coordinates

Name	Coordinate 1	Coordinate 2
120.jpg	0.831288	0.855102
121.jpg	0.880368	0.208163
122.jpg	0.785276	0.177551
123.jpg	0.546012	0.434694
124.jpg	0.165644	0.726531
126.jpg	0.260736	0.434694

Conclusion

The convolution to the fully connected layer mapping is not working properly.

Some sort of centerpooling layer might be better suited for this mapping.