DL Assignment 1: Object Detection
Iram Kamdar

A February 8, 2022

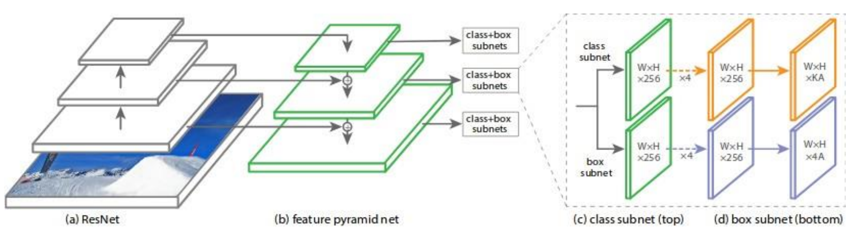
1 Baseline

For the purpose of object detection on the given data set, the baseline has been motivated by the implementation (available open-source) of a recent paper ***Focal Loss for Dense Object Detection***.

Retina Net consists of 3 composite networks : z

1. Backbone Network : This consists of a feature pyramid network built on ResNet50 or ResNet101.
2. Classification Subnet : This network focuses on predicting the presence of an object in the input image.
3. Regression Box Subnet : This network aims to output the location of the object detected in the image.

This implementation uses a pre-trained ResNet model of depth 50 along with Adam optimizer. The classification threshold is kept 0.5. If in case, prediction for none of the classes crosses the threshold, it does not output any threshold.



(a) Retina Net

1.1 Data Analysis and Cleaning

The data set contains a total of 12,617 annotation files. Since an image may contain more than one bounding box, it contributes to as many distinct annotation files as the number of bounding boxes in it. After cleaning the data, we

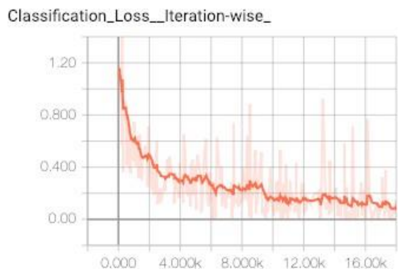
get the total number of images to be equal to 12,099.
Below is the data set statistics as per the initial annotation files.

Weighted Loss: We observed that the accuracy in bounding box predictionby our baseline model is satisfactory while there is still scope for improvementin prediction of the right class id after we have drawn the bounding boxes. Therefore, we weighted the classification loss twice the magnitude of regression loss.

Category	Sample Percentage	maP with baseline
Bobcat	684	0.754
Opossum	2514	0.932
Coyote	1371	0.745
Raccoon	1030	0.706
Bird	560	0.245
Dog	769	0.595
Cat	1170	0.756
Squirrel	1037	0.808
Rabbit	2278	0.817
Skunk	214	0.817
Rodent	264	0.804
Badger	9	0
Deer	44	0.4
Car	668	0.999
Fox	5	0

Table 1: Data Set Statistics on baseline method

The training loss plots are as follows:



(a) Classification Loss

(b) Regression Loss

(c) Running Loss

2 Significant Changes

2.1 Innovation 1

2.1.1 Motivation

Since the data provided is biased and not equally divided among classes, a few classes were highly deficient with respect to the number of data samples belong- ing to them. As evident from the maP, the precision is extremely low specificallyfor these classes, thereby decreasing the overall performance of the model as well.

To improve upon this, the idea is to use Weighted Random Sampling. In order to compensate the scarcity of data samples for some classes, they are given more weightage so to affect the loss term more.

2.1.2 Improvement

The following are the training loss plots after adding weighted random sampling to our model:

(a) Classification Loss

(b) Regression Loss

(c) Running Loss

Category	maP
Bobcat	0.948
Opossum	0.950
Coyote	0.886
Raccoon	0.874
Bird	0.673
Dog	0.838
Cat	0.957
Squirrel	0.954
Rabbit	0.958
Skunk	0.956
Rodent	0.964
Badger	0
Deer	1.0
Car	0.998
Fox	1.0

Table 2: maP after adding weighted random sampling

2.2 Innovation 2

2.2.1 Motivation

As we have substantially changed the number of samples of classes that are originally very low in number, there is a significant chance that our model overfits. We need to prevent this overfitting by using a combination of various models trained on different sets of data.

2.2.2 Improvement

In order to prevent such overfitting, we make use ensemble approach i.e. bagging between different models. In this approach what we basically do it that we parallely train a number a strong learners and then combine all strong learners to smooth out their prediction. We pass the test image through all our models, average out bounding box coordinates to output generalised bounding boxes coordinates. The predicted class is the one that has the maximum confidence score.

Category	maP
Bobcat	0.746
Opossum	0.942
Coyote	0.825
Raccoon	0.808
Bird	0.250
Dog	0.599
Cat	0.728
Squirrel	0.841
Rabbit	0.882
Skunk	0.802
Rodent	0.884
Badger	0
Deer	0.414
Car	0.998
Fox	0

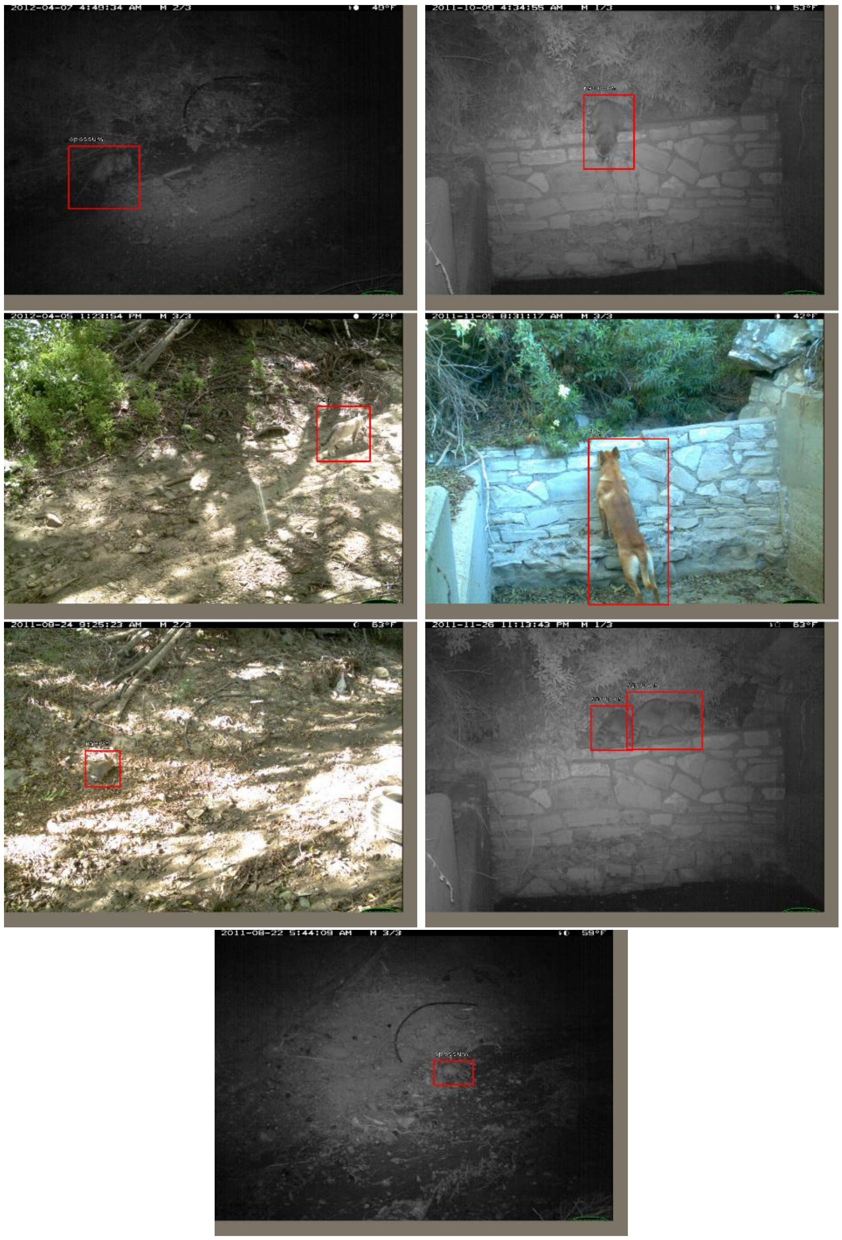
Table 3: maP after adding ensemble with baseline models

2.3 Ensemble with weighted random sampling

Category	maP
Bobcat	0.880
Opossum	0.920
Coyote	0.732
Raccoon	0.850
Bird	0.686
Dog	0.780
Cat	0.940
Squirrel	0.955
Rabbit	0.878
Skunk	0.956
Rodent	0.964
Badger	0
Deer	1.0
Car	0.999
Fox	1.0

Table 4: maP after adding ensemble with weighted random sampling models

3 Some results



4 References

- Code borrowed from pytorch-retinanet by yhenon
- Lin, Tsung-Yi et al. "Focal Loss for Dense Object Detection." 2017 IEEE International Conference on Computer Vision (ICCV) (2017): 2999-3007.
- Ensemble approach motivated from ensemble-objdet by ahrnbom