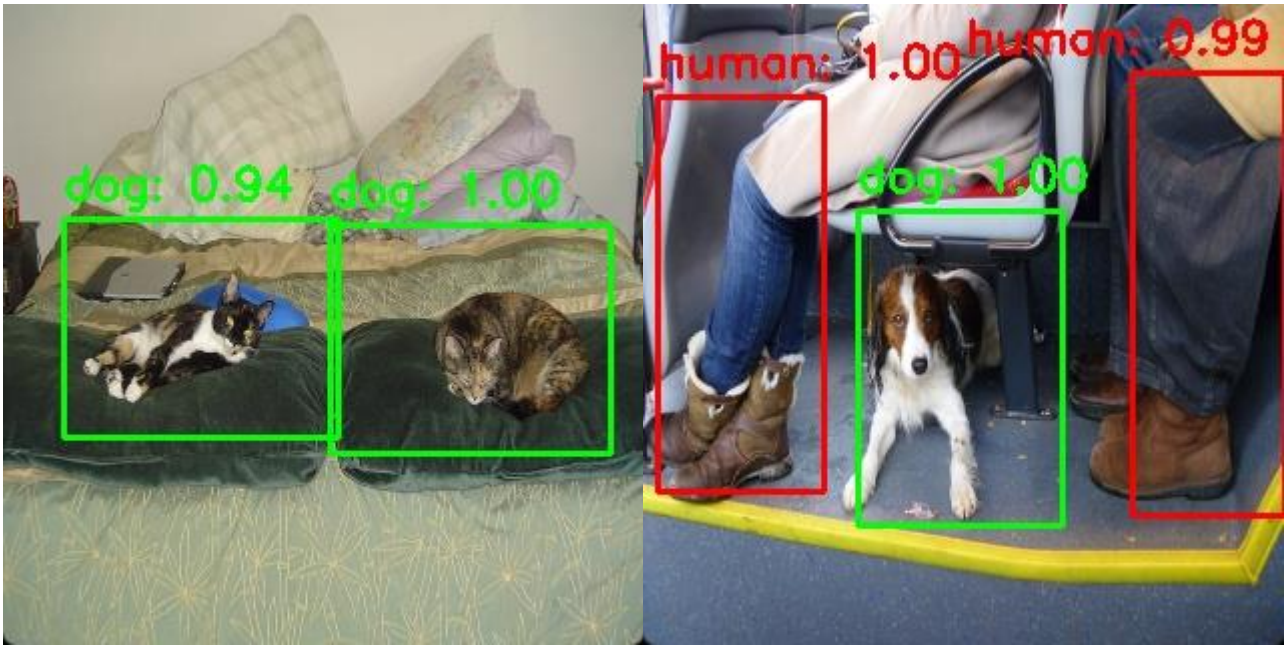
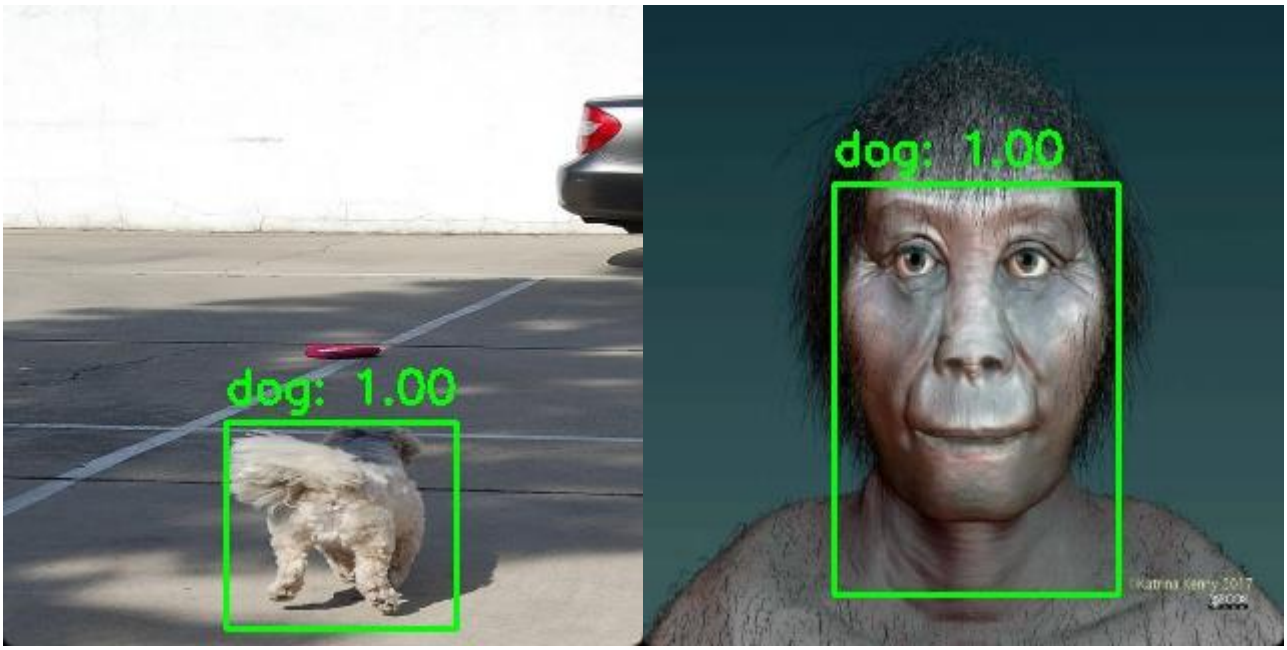
CMPT 742 Visual Computing I Assignment 3: SingleShotDetection Report

Name: Jiaqing Hu (Vincent)
SFU ID: 301368526
SFU Mail: jiaqingh@sfu.ca

Part A Test Images:


dog: 0.99
human: 1.00
human: 1.00


dog: 0.94
dog: 1.00
human: 1.00
human: 0.99
dog: 1.00


dog: 1.00
dog: 1.00

Part B Disadvantages of the current model with analysis:

Based on the results, the model has some disadvantages:

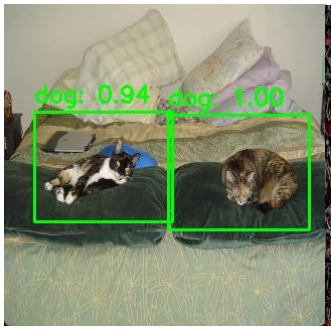


Figure 1

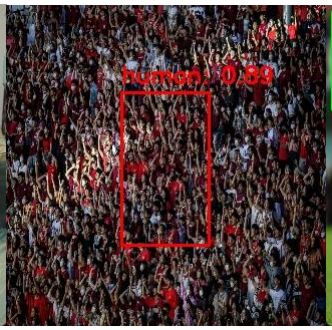


Figure 2

Figure 3

1. The model makes mistakes when identifying classes, in some cases it mislabels cats to dogs as shown in fig 1. This is a general problem with animals that have similar features, especially for cats and dogs which have similar body structures, furs, even behaviors. The model is also trained on imbalanced dataset (in 7000 images there are 5000 humans and 1000 each for cats and dogs), I tried to implement methods like WeightedRandomSampler and using weighted loss in the loss function to give more weights to imbalanced classes (cats and dogs). It didn't outperform too much compared to the original model.
2. The model doesn't perform well when the objects are too small and too many (fig 2). This is because of the default bounding box sizes were not dedicated to such tasks.
3. The model generates false positives when there are some objects in the background (fig 3). The model could also generate false positives when there exist many noises in the background. Intuitively the model could be overfitting the training dataset, which causes it to be sensitive to noises.

Part C Possible Improvements:

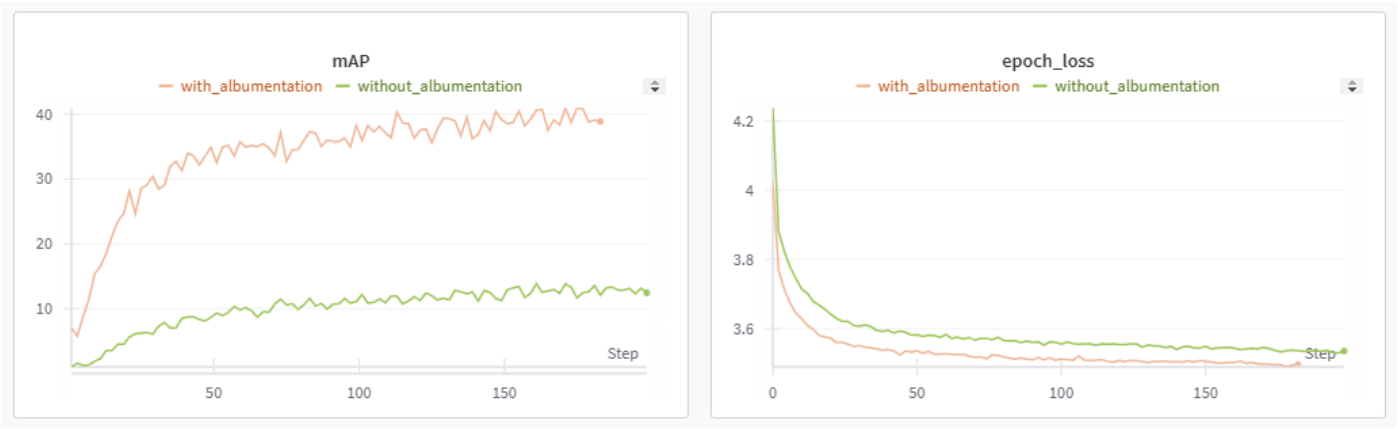
1. Rule of thumb is to get more data. The current dataset is too small. And if we train the model on a larger dataset, it will learn more features and should be more robust when classifying different objects.
2. Train the model on a more balanced dataset which has fair distribution across different classes.
3. Apply more image processing techniques during data augmentation stage to improve robustness of the model.

Part D Best confidence threshold:

I am choosing 0.9 as my threshold, 0.1 for overlapping in non-max suppression. Such combination eliminates redundant bounding boxes most efficiently and provides the best visualization results.

Part E Evalutaions:

mAP & loss



Precision-Recall curve

0: Cat

1: Dog

2: Human