Name: 李華健 Dep.: 電機三 Student ID:B05901119

Reference:

Github – Pytorch YoLov1

https://github.com/xiongzihua/pytorch-YOLO-v1/blob/master/yoloLoss.py

1. (5%) Print the network architecture of your YoloV1-vgg16bn model and describe your training config. (optimizer, batch size….and so on)

Optimizer: SGD
Batch Size: 16
Momentum: 0.9
Weight Decay: 1e-4
Learning Rate:

(1-20 Epoches)	1e-3
(21-40 Epoches)	1e-4
(41-55 Epoches)	1e-5
(56+ Epoches)	1e-6

NMS setting:
Keep bounding box: 0.05 (Keep when P(class) > 0.05)
IOU threshold: 0.5 (Remove when IOU>0.5)

Training Set augmentation:

- Random Horizon Flip

- Random Zoom In (x1.1) and crop

Final Result: Stopped at epoch 47.

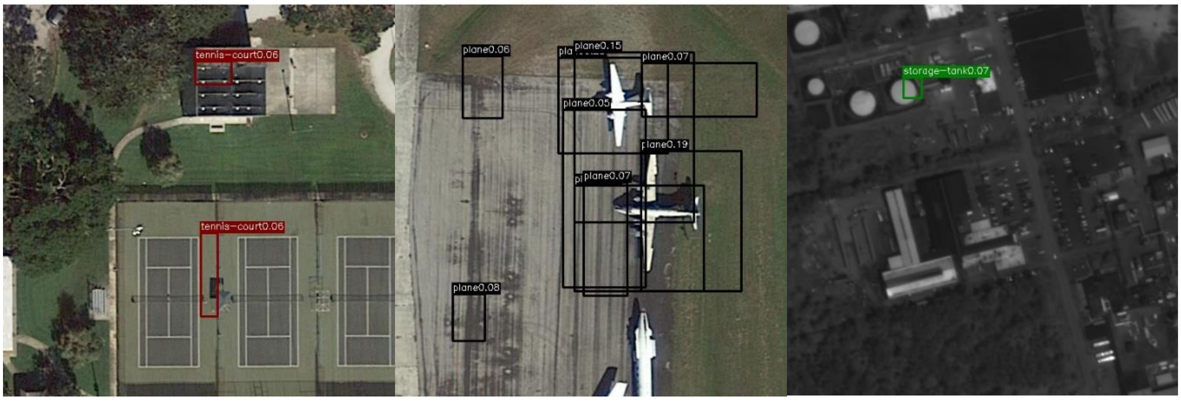
Input size: (3, 448, 448)

Layer: (Layer Name: Output size)

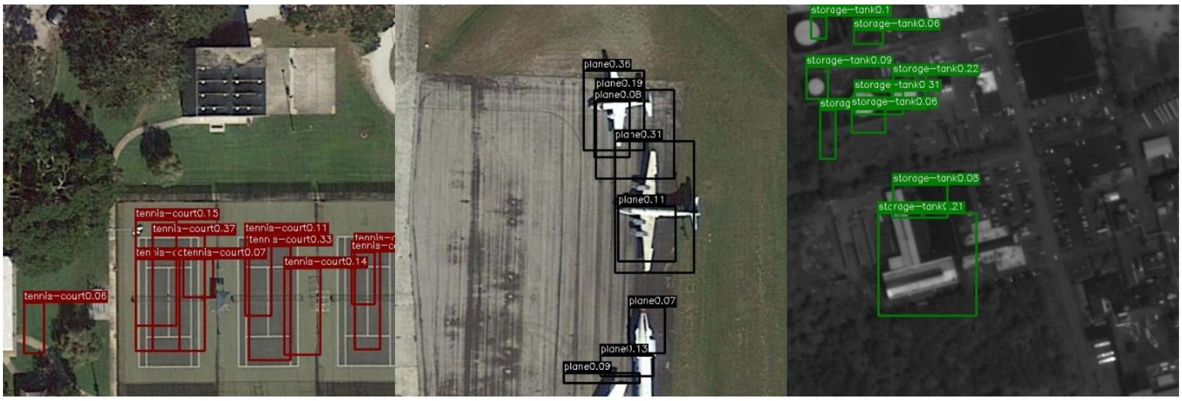
VGG16-bn:	(512, 7, 7)
Flatten-layer:	(25088)
Fully Connected:	(4096)
LeakyReLU(0.02):	(4096)
Dropout(0.5):	(4096)
Fully Connected:	(1274)
Sigmoid:	(1274)
Reshape:	(7, 7, 26)

2. (10%) Show the predicted bbox image of "val1500/0076.jpg", "val1500/0086.jpg", "val1500/0907.jpg" during the early, middle, and the final stage during the training stage. (For example, results of 1st, 10th, 20th epoch)

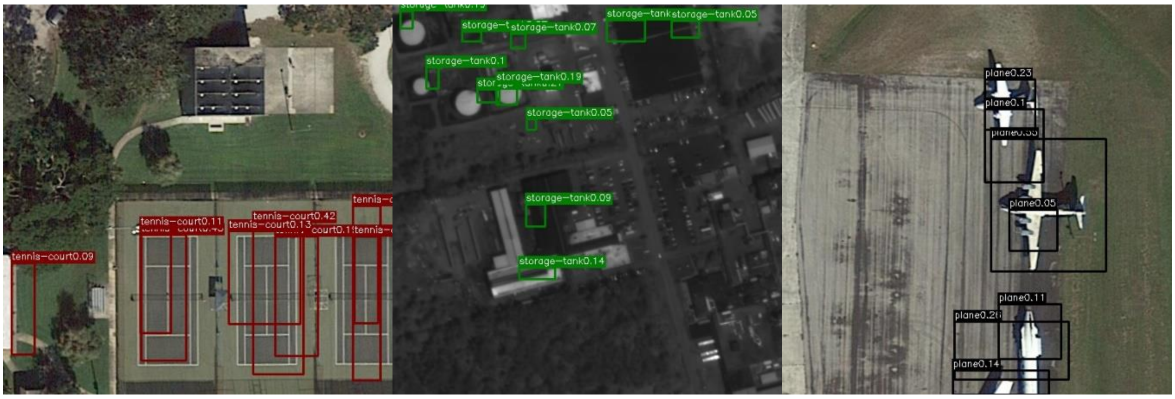
Early State (Epoch 25):



Middle State (Epoch 35):



Final Result (Epoch 47):



3. (10%) Implement an improved model which performs better than your baseline model. Print the network architecture of this model and describe it.

   Optimizer: SGD
   Batch Size: 16
   Momentum: 0.9
   Weight Decay: 1e-4
   Learning Rate:
   (1-20 Epoches) 1e-3
   (21-40 Epoches) 1e-4
   (41-55 Epoches) 1e-5
   (66-70 Epoches) 1e-6

   NMS setting:
   Keep bounding box: 0.05 (Keep when P(class) > 0.05)
   IOU threshold: 0.5 (Remove when IOU>0.5)

   Trainset augment:

   - Random Horizon Flip / Vertical Flip

   Final Stopped at 5 Epochs

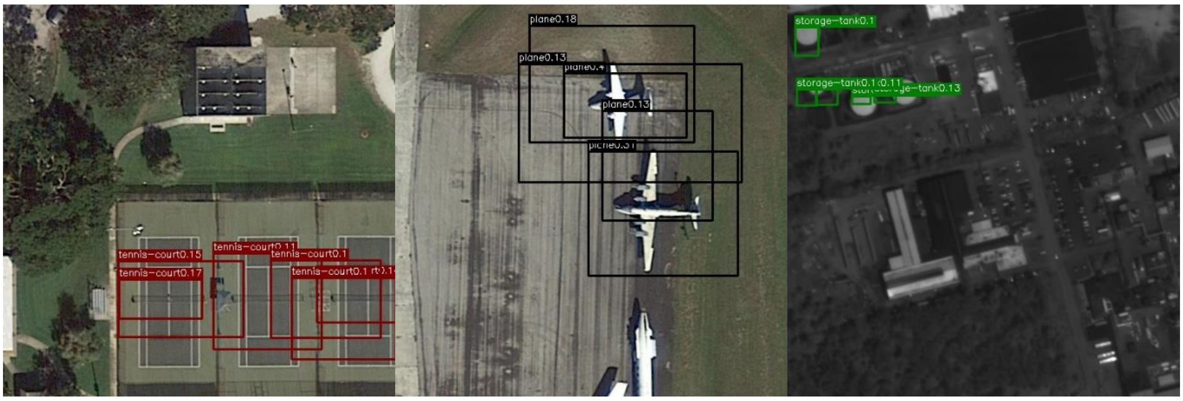
Input size: (3, 448, 448)

Layer: (Layer Name: Output size)

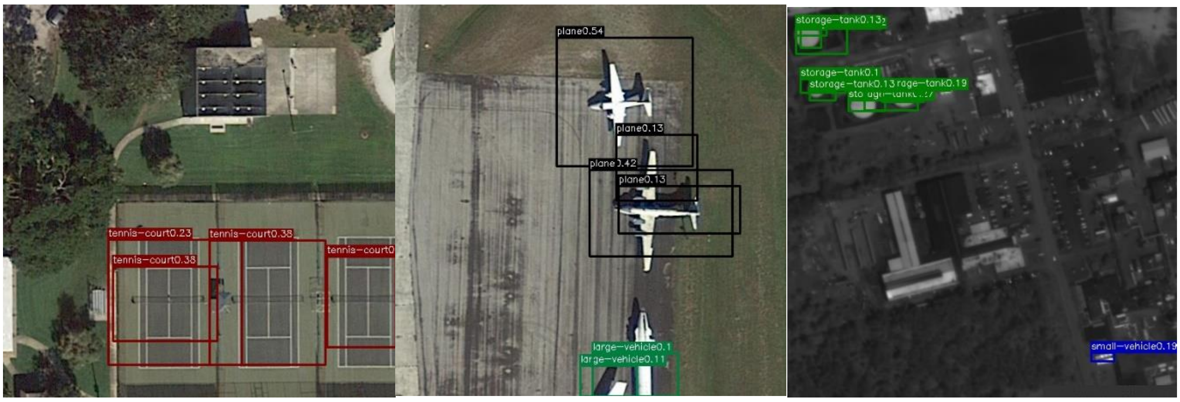
VGG16-without Maxpooling:	(512, 14, 14)
Conv2D(size=1)	(26, 14, 14)
Sigmoid:	(5096)
Reshape:	(14, 14, 26)

4. (10%) Show the predicted bbox image of “val1500/0076.jpg”, “val1500/0086.jpg”, “val1500/0907.jpg” during the early, middle, and the final stage during the training process of this improved model.

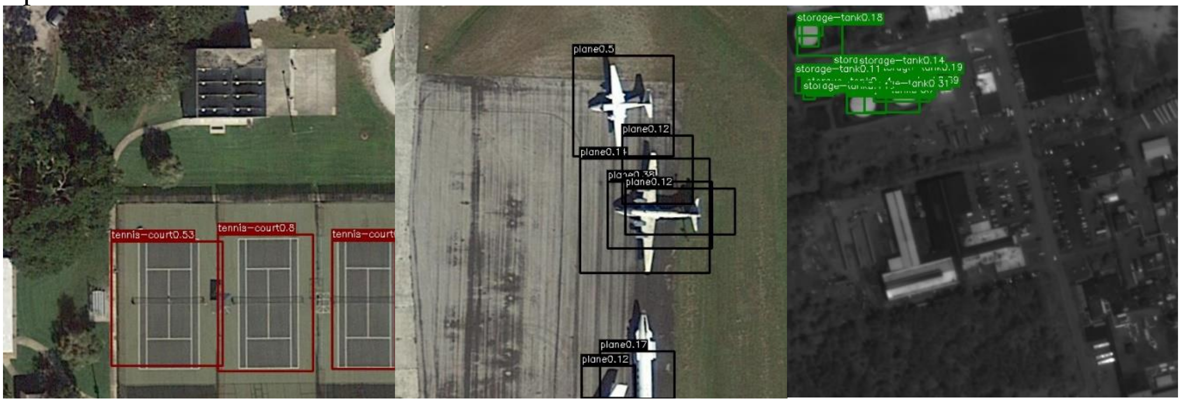
Epoch 1:



Epoch 3:



Epoch 5:



5. (15%) Report mAP score of both models on the validation set. Discuss the reason why the improved model performs better than the baseline one. You may conduct some experiments and show some evidences to support your reasoning.

   mAPs of basic model: 10.55% (at IoU 0.5, minimum prob 0.05)

   mAPs of improve model: 11.76% (at IoU 0.5, minimum prob 0.05)

6. **bonus (5%)** Which classes prediction perform worse than others? Why? You should describe and analyze it.

Statistics: AP of each class in Basic Model with the best mAP performance result.

	Final AP (Base)	Final AP (Improve)	Number in train15000 (Base)	Number in train15000 (Improve)
Plane	21.01%	35.75%	11407	14451
Baseball-diamond	16.67%	15.58%	1412	1836
Bridge	1.14%	0.56%	2619	2898
Ground-track-field	9.09%	0.0%	1879	2357
Small-vehicle	0.38%	9.09%	63229	98381
Large-vehicle	9.09%	18.26%	30684	47664
Ship	2.7%	14.27%	25012	38773
Tennis-Court	36.48%	52.36%	5663	6963
Basketball-court	14.54%	18.18%	1620	1969
Storage-tank	9.09%	10.08%	4575	6551
Soccer-ball-field	32.32%	0.0%	1872	2380
Roundabout	0.0%	1.13%	2172	3129
Harbor	7.20%	2.28%	21380	32986
Swimming-pool	9.09%	10.56%	6334	8366
Helicopter	0.0%	0.0%	637	896
Container-crane	0.0%	0.0%	188	245

Issue 0: Encoding problem.

First, when I encoded the ground truth to tensors, there’s only a class in one grid. Random choice is applied when multiple bounding boxes appear in a class.

Issue 1: The number of each class is unbalanced.

The table above counts that the class appears in my ground truth tensors. I found that some class seldom appears, such as “Helicopter” and “Container-crane”. Maybe this is a reason that the model can’t identify those class with a high confidence.

Issue 2: Average size of each class.

When we use the 7x7 grids, only 0.38% of small vehicle is detected. But a large improve occurred when using 14x14 grids. The reason is the size of a small vehicle is very small, it’s very difficult to learn by the MSE loss function and sigmoid activation.

Issue 3: The improved model hadn’t fine tune enough.

Only 5 epochs were applied to the improve model, maybe the parameter have a large range to fine tune. Therefore, there is some uncertainly of why some class detects well in the basic model then improve model.