

**National University of Sciences and Technology
School of Electrical Engineering and Computer Science
Department of Computing**

**CS893: Advanced Computer Vision
Spring 2020**

Assignment 02

Object Detection using Deep Learning

Announcement Date: 14th Feb, 2020

Due Date: 8th March 2020 at 11:55 pm

(on Github and Dota Servers)

Instructor: Dr. Muhammad Moazam Fraz

Background Knowledge:

Object Detection lecture notes and shared references (Lectures 04 – 07)

Expected Outcomes:

The goals of this assignment are:

- To learn about the object detection pipeline
- Implementation of end to end object detection (localization + classification) on a real world dataset

Dataset:

DOTA: A Large-scale Dataset for Object Detection in Aerial Images

The images of **DOTA-v1.0** dataset are mainly collected from the Google Earth, satellite JL-1 and satellite GF-2 of the China Centre for Resources Satellite Data and Application. (available for research purpose only)

Default object categories/ classes of DOTA-v1.0 include: plane, ship, storage tank, baseball diamond, tennis court, basketball court, ground track field, harbor, bridge, large vehicle, small vehicle, helicopter, roundabout, soccer ball field and swimming pool.

In the dataset, each instance's location is annotated by a quadrilateral bounding boxes, which can be denoted as " $x_1, y_1, x_2, y_2, x_3, y_3, x_4, y_4$ " where (x_i, y_i) denotes the positions of the oriented bounding boxes' vertices in the image.

The ground truth consists of annotation of location, category label and a difficult/ not difficult label (which indicates whether the instance is difficult to be detected, 1 for difficult, 0 for not difficult). Annotations for an image are saved in a text file with the same file name. The annotation format is:

$x_1, y_1, x_2, y_2, x_3, y_3, x_4, y_4, \text{category}, \text{difficult}$

Dataset can be downloaded from:

<https://captain-whu.github.io/DOTA/dataset.html>

Task:

- Implement end to end object detection and classification pipeline using CNN backbone of your choice (i.e. resnet, resnext, VGG etc) for all classes of dataset i.e. DOTA-v1.0.
- Evaluate and visualize the results using your trained model.

Submission instruction:

Create your GitHub account and create repository for this assignment. The code should be properly documented and organized. The ReadMe.md (the main page) should also be nicely

created, with background information of the problem, your results, pre-trained model (if any) and a narrative that how to replicate your reported results. The main page may contain following.

- Network summary or diagram
- Training settings
- Confusion matrix
- Training and Validation Performance Graphs of accuracy and loss per epoch.
- Some Visual results of your trained model.
- Pretrained model
- How to setup the environment and replicate your result
- ...

You may like to look at other GitHub repositories for an idea that how to present your information more effectively.

Moreover, create your accounts on DOTA evaluation servers, using your CMS ID as user name and submit your results in their instructed format. I would like to see your results in the top ranks at the leader board.

You can use Amazon AWS / Microsoft Azure / Google Colab for this assignment. Following links may be helpful.

- [Getting Started With Google Colab](#)
- [Google Colab Free GPU Tutorial](#)