



## **Project Outline and Plan**

### **Object Detection Website Based on Deep Learning**

**Yunjia Feng**

**<BSc Computer Science>**

The candidate confirms that the work submitted is their own and the appropriate credit has been given where reference has been made to the work of others.

I understand that failure to attribute material which is obtained from another source may be considered as plagiarism.

(Name of student) <Yunjia Feng>

## **1. Introduction**

Object detection has always been of great significance in the field of computer vision whose main objective is to enable the computer to accurately classify the objects in a given picture or video along with their positions. Recently, there is an enormously growing interest in this field and a large number of excellent object detection neural network algorithms were developed by many international computer vision research institutions, including Faster R-CNN, SSD, YOLO. From both the perspectives of research and application, the object detection field has indicated the significance and prospect, since it is not only the basis of many other high-level tasks (image classification, face recognition, target tracking, pedestrian re-recognition) but also a foundation among a wide range of practical usages: face detection technology, vehicle detection applied in aided driving, automatic driving.

However, the lack of direct interaction between user and object detection application leads to the unfamiliarity of this promising technology. Therefore, the motivation of this project is to select appropriate object detection algorithms with proper data sets and train the deep learning model, then use the form of the website to allow the user to experience object detection tasks easily.

## **2 Aims, Objectives and Deliverables**

### **2.1 Aims**

The primary aims of this project are to take research into the object detection algorithms including their implementations, which dataset to choose, hyperparameters to set in the training process, which acquire the model to perform high standard in both precision and speed. Moreover, the representation of the application should be built in consideration of a user-friendly interface that allows people to experience object detection technology conveniently. At last, a final report should be constructed in a clear logic and objectivity in describing and evaluating the whole process.

### **2.2 Objectives**

- (1) The final application should allow the user to upload images (or possibly videos) for object detection purposes in the interaction of a website form.
- (2) The trained model should perform a relatively high standard in precision according to its specific algorithm benchmark.
- (3) The trained model should achieve a relatively high standard in fast responding time according to its specific algorithm benchmark.
- (4) The application should be built in a user-friendly standard for interfaces and interactions in order to provide convenience.

(5) The final report should be conducted in an academic way with clear logic, which should show the details for the building and evaluating process.

## 2.3 Deliverables

(1) A model should be provided for the outcome of multi-classes object detection tasks training, along with its precision, processing speed, and other important indicators.

(2) The code includes the integration of both the object detection module and representation by the website (front end and back end). Multiple history versions should be included according to the code version control tool (for instance, Github).

(3) The final report writing in an academic way, by demonstrating details for the implementation and evaluating process in a logical and structural format.

## 3. Project Plan

9.20 - 10.8 Complete the project outline and plan, discuss the topic and laboratory usage in the project with the supervisor.

10.9 - 10.17 Object Detection background research:

(1) Basic background research about the general deep learning field on its significance, current application, bottleneck, and future development.

(2) Background research on object detection and related tasks: classification, location, instance segmentation, and read relevant papers.

(3) Research on object detection algorithms: YOLOv5, Faster R-CNN (refer to R-CNN, Fast R-CNN as the basis), and SSD, then summarize the advantages and disadvantages of each algorithm.

(4) Research on object detection dataset: VOC, COCO, ImageNet, and select one dataset for training.

(5) Investigate basic indicators and benchmarks related to object detection, read relevant papers, and acquainted with the process for subsequent model evaluation.

10.18 - 10.25 Prepare for the Thesis Proposal Defense.

10.25 - 11.30 Determine the project structure and technology stack, built the development environment, and complete the basic demo:

(1) The basic construction of front and back end modules and deep learning framework (technology stack: Vue + Flask + Pytorch + YOLOv5)

(2) Build Anaconda deep learning development environment and install related dependent libraries.

(3) Implement data set format conversion code.

(4) Complete the demo based on the pre-training model, test the object detection accuracy of pictures, videos, and cameras.

12.1 - 12.31 Start training on multiple pre-training weights:

(1) Record and visualize loss, AP, recall, speed and other important indicators of each training.

(2) Evaluate the training results, use test sets to evaluate the training results, and formulate improvement plans according to the situation.

(3) Initial the setup of the front-end page (using Node.js and Vue framework).

(4) Write Project Outline and Plan.

1.1 - 1.31 Continue the remaining development tasks:

(1) Develop and beautify the web page, add status hints and other information to improve user experience.

(2) Complete the interface for accessing web pages at the back end, including uploading and downloading pictures.

(3) Complete the deep learning Yolo-v5 image object detection module and encapsulate the image object detection processing interface.

2.1 - 2.28 Continue remaining development tasks and thesis writing:

(1) Complete the front-end and back-end of selecting weight for object detection.

(2) Add sign up/log functionality in for users, allow user to favorite their uploaded images and results by using MySQL database.

(3) Design charts based on previous work (training and test samples, comparison with benchmark)

(4) Begin to write the abstract, introduction, and literature review of the thesis.

3.1 - 3.15 Continue to write the thesis and prepare for the mid-term examination.

3.16 - 4.20 Prepare for Thesis Defense

## 3.1 Timeline



## 4. Risk Mitigation

The potential risk for this project may attribute to the necessary hardware for the deep learning training process for graphics, which acquires high-performance computer equipment, including large

memory capacity (more than 16G), Nvidia GPU (least for GTX1060, 8G memory for display card).

However, the possible solution is listed which cloud perfectly avoided the risk:

- (1) Use my own device which cloud offers the least standard for the training process (though may lead to some precision loss).
- (2) Proposal of using the equipment in school laboratory by communicating with the supervisor.
- (3) Use the online solution for training (for instance Google Colaboratory), or rent specific cloud services for deep learning purposes.

## 5. Ethics

There are no ethical issues for this project.

## List of References

- [1] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in Advances in neural information processing systems, 2012, pp. 1097–1105.
- [2] Vishwakarma S, Agrawal A. A survey on activity recognition and behavior understanding in video surveillance [J]. The Visual Computer, 2012: 1-27.
- [3] Jiang, B., Luo, R., Mao, J., Xiao, T., & Jiang, Y. (2018). Acquisition of localization confidence for accurate object detection. In ECCV (pp. 784–799).
- [4] Hei Law and Jia Deng. Cornernet: Detecting objects as paired keypoints. In Proc. Eur. Conf. Comp. Vis., pages 734–750, 2018.
- [5] Liu, Wei, et al. "SSD: Single Shot MultiBox Detector." European Conference on Computer Vision Springer International Publishing, 2016:21-37.S
- [6] C.-Y. Fu, W. Liu, A. Ranga, A. Tyagi, and A. C. Berg. DSSD: Deconvolutional single shot detector. arXiv:1701.06659, 2016. 1, 2, 8
- [7] Z. Li and F. Zhou. Fssd: Feature fusion single shot multibox detector. arXiv preprint arXiv:1712.00960, 2017.
- [8] R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In CVPR, 2014.
- [9] S. Ren, K. He, R. Girshick, and J. Sun. Faster R-CNN: Towards real-time object detection with region proposal networks. In NIPS, 2015.
- [10] Cai, Z., Vasconcelos, N.: Cascade r-cnn: Delving into high quality object detection. arXiv preprint arXiv:1712.00726 (2017)
- [11] Redmon, Joseph, et al. "You Only Look Once: Unified, Real-Time Object Detection." (2015):779-788.