### HO CHI MINH UNIVERSITY OF TECHNOLOGY AND EDUCATION

**FACULTY FOR HIGH QUALITY TRAINING**



SENIOR PROJECT 2

**TRAFFIC SIGN AND LIGHT RECOGNITION USING YOLOV8**

#### COMPUTER ENGINEERING TECHNOLOGY

Students: **NGUYEN TRONG HAI** MSSV: 19119028 **NGUYEN MINH KHANH**

MSSV: 19119186

Advisor: Ph.D. **TRUONG NGOC SON**

### HO CHI MINH UNIVERSITY OF TECHNOLOGY AND EDUCATION

**FACULTY FOR HIGH QUALITY TRAINING**



SENIOR PROJECT 2

**TRAFFIC SIGN AND LIGHT RECOGNITION USING YOLOV8**

#### COMPUTER ENGINEERING TECHNOLOGY

Students: **PHAM DUC PHUONG**

MSSV: 19119054

Advisor: Ph.D. TRUONG NGOC SON

##### ACKNOWLEDGMENTS

The project of "Traffic Sign And Light Recognition Using YOLOv8" project was carried out in accordance with the requirements with the efforts of the individual au- thors, with the dedicated help of the teacher in the process of implementing the project. We would like to express our gratitude and thanks to our guide Mr. Truong Ngoc Son enthusiastically guided and facilitated us to complete the project. Although the authors have tried to complete the assigned tasks and on time, but due to limited knowledge and experience, errors and omissions cannot be avoided. We hope you understand and leave comments for us to absorb and edit

Project Exececutors

*(Sign with full name)*

##### DECLARATION

We hereby declare that the project with the topic "Traffic Sign And Light Recogni- tion Using YOLOv8" is our own research project under the guidance of Mr. Truong Son Nguyen. The opinions expressed in the project are the result of our serious research and are based on the research of published documents and scientific translations. The project has been implemented and ensures objectivity, honesty and science at the highest level.

Project Exececutors

*(Sign with full name)*

##### ABSTRACT

In this project, we propose to explore the YOLOv8 Architecture and its comity to break this problem. The ideal is locating and bracket of business signs in natural road scenes. The crucial challenge to be addressed in this problem is recognition of nanosecond targets in an extended and complex image background. The main downsides with similar styles is their speed- they fail to be real time. therefore, the provocation behind exploring YOLO for this task is the speed. In this work, we also propose a new modified loss function for the YOLO model to perform better for business sign discovery. Autonomous driving is one of the intriguing exploration areas of ultramodern times and business sign discovery is a veritably important and pivotal

problem in this exploration.

**Contents**

1. [OVERVIEW](#_bookmark0) 1
   1. [Introduction](#_bookmark1) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 1
   2. [Objection](#_bookmark2) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 1
   3. [Research Situation](#_bookmark3) . . . . . . . . . . . . . . . . . . . . . . . . . . . . 2
   4. [Research Method](#_bookmark4) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 2
   5. [Content Structure](#_bookmark5) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 3
2. [LITERATURE REVIEW](#_bookmark6) 4
   1. [Related Work](#_bookmark7) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 4
   2. [Outcome of Literature Review](#_bookmark8) . . . . . . . . . . . . . . . . . . . . . . 5
   3. [Problem Statement](#_bookmark9) . . . . . . . . . . . . . . . . . . . . . . . . . . . . 5
   4. [Research Objectives](#_bookmark10) . . . . . . . . . . . . . . . . . . . . . . . . . . . . 5
3. [METHODOLOGY AND FRAMEWORK](#_bookmark11) 6
   1. [Architecture](#_bookmark12) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 6
   2. [Loss Function](#_bookmark15) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 8
   3. [Output](#_bookmark16) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 8
   4. [YOLOv8 Algorithm](#_bookmark18) . . . . . . . . . . . . . . . . . . . . . . . . . . . . 9
   5. [Data Set](#_bookmark19) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 9
4. [CODE AND RESULTS](#_bookmark24) 12
   1. [Code](#_bookmark25) 12
   2. [Results](#_bookmark33) 15
5. [CONCLUSION AND RECOMMENDATION](#_bookmark35) 17
   1. [Conclusion](#_bookmark36) 17
      1. [Advantages](#_bookmark37) 17
      2. [Disadvantages](#_bookmark38) 17
   2. [Recommendations](#_bookmark39) 18

**List of Figures**

* 1. [YOLO’s Architecture](#_bookmark13) . . . . . . . . . . . . . . . . . . . . . . . . . . . 6
  2. [Layers in the Darknet-53 network](#_bookmark14) . . . . . . . . . . . . . . . . . . . . 7
  3. [Algorithm is applied at output](#_bookmark17) . . . . . . . . . . . . . . . . . . . . . . 9
  4. [New Project](#_bookmark20) 10
  5. [Upload Images](#_bookmark21) 10
  6. [Cut Frame Class And Label](#_bookmark22) 11
  7. [File Train](#_bookmark23) 11
  8. [Link with google drive](#_bookmark26) 12
  9. [Extract dataset](#_bookmark27) 13
  10. [Download version YOLOv8](#_bookmark28) 13
  11. [Download Ultralytics](#_bookmark29) 14
  12. [Create YAML](#_bookmark30) 14
  13. [Train with file best.pt](#_bookmark31) 15
  14. [Check Acc and Bounding box](#_bookmark32) 15
  15. [Results](#_bookmark34) 16

**List of abbreviations**

Below is the list of abbreviations used in the project.

##### Abbreviations Definition

DL Deep Learning

YOLO You Only Look Once

AI Artificial Intelligence

HSV Hue, Saturation, Value

RGB Red, Green, Blue

R-CNN Region Based Convolutional Neural Networks

**Chapter 1**

**OVERVIEW**

# Introduction

YOLOv8 stands out as the forefront iteration of YOLO models, catering to tasks encompassing object detection and image classification. Its development is credited to Ultralytics, the very team behind the prior successful YOLOv5 model. Noteworthy architectural refinements and enhancements define YOLOv8’s evolution. In the context of autonomous vehicles, visual acuity holds paramount importance, encompassing the capacity to discern and identify road-based traffic signs. Addressing the complication of simultaneously detecting multiple signs within a single frame poses an additional challenge.

# Objection

The core aim of this project revolves around the reimplementation and comprehensive analysis of YOLOv8. In the realm of real-time traffic sign detection, a pivotal concern is the latency during testing. Initially, CNNs (Convolutional Neural Networks) were deemed impractical for real-time traffic sign detection due to their intricate computa- tional demands. Nevertheless, the advancement of GPUs (Graphics Processing Units) has cleared a pathway for utilizing CNNs in this context, capitalizing on their robust computing capabilities. This development aligns ideally with our specific challenge,

demanding both high speed and accuracy.

# Research Situation

Our necessity entails a model capable of real-time sign detection and classification. Within this study, we delve into the YOLOv8 architecture, renowned for its ability to achieve object detection and classification in real time, averaging around forty-five frames per second. The objective of this endeavor encompasses conducting thorough experiments to evaluate the effectiveness of YOLOv8 and explore potential avenues for its enhancement. By doing so, we aim to augment the current body of knowledge con- cerning object detection methodologies.

# Research Method

To achieve the goal of the project, we use the methods of data collection, experimen- tation and analysis, summarizing experience.

* + - Data collection methodology:
      * Using the observation method: We conducted direct observation on building our own dataset which for implementing.
    - Experimental approach:
      * Design and system implementation: We carried out the design and implementa- tion on both Google Colab.
    - Method of experience analysis and summarization:
      * Evaluation: We conducted an evaluation of the effectiveness based on the average precision of the test results.
      * Data analysis: We analyzed the data which is mainly the time of testing process of YOLOv8.
      * Summary: Based on the analysis results, we summarized the experience working with other projects.

# Content Structure

Our report for the senior project 2 will consist of 5 chapters:

* + - Chapter 1: Overview
    - Chapter 2: Literature review
    - Chapter 3: Methodology and framework
    - Chapter 4: Code and results
    - Chapter 5: Conclusion and recommendation

**Chapter 2**

**LITERATURE REVIEW**

# Related Work

Traffic sign and light recognition methods can be classified into two categories. The classical object detection models which rely on the characteristics of the traffic signs and the class of techniques which include the recent deep learning based approaches which learn the general features instead of using the manually selected features as in the first approach.

The classical detection methods have mainly relied on feature algorithms. Features like color and shape are used for both classification and detection. Images are generally transformed into HSV to overcome RGB color space for various light condition presents a color probability model that computes maps based on Ohta space. The general shapes of the traffic signs include circle, triangle, rectangle or any others.

Although manually selected features have achieved higher precision for traffic signs, traditional detection approaches are very specific towards changing scenes and its as- sociated complexities. Thus DL approaches have become popular for object detection problems in recent years. The deep learning approaches can be further classified into region based techniques and regression based techniques.

Thus we can conclude that the two-stage methods do better in terms of the localization and classification accuracies, however at increased cost of resources and latencies. The end-to-end one stage methods are useful in producing faster results at a reduced preci-

sion. Here, the latter would be more appropriate as the traffic sign detection must be real time.

# Outcome of Literature Review

As explained above, different models for traffic sign detection have been explored and their pros and cons are understood. It is concluded that the YOLO model may be the most suitable for the application because of the speed and low latency of testing time compared to other models. Although the accuracy of the YOLO model is worse than that of other models as discussed above, YOLO’s real-time object detection and classification outweighs this drawback. The accuracy of R-CNN for object detection and classification is very good but would not be suitable for this application due to its high latency.

# Problem Statement

To develop a YOLOv8 based deep CNN model for Traffic Sign And Light Recogni- tion, we have to build a wide-range Data set. The objective is to obtain a model which can detect and classify traffic signs and light simultaneously in real time.

# Research Objectives

The general objective is to obtain a YOLOv8 based deep CNN model for Traffic Sign and Light Recognition on our Data set. The specific objectives in order to obtain above objective are:

* + - To analyse the existing YOLOv8 framework on data set.
    - To train and test YOLOv8 on data set.
    - To estimate modifications in the YOLOv8 framework that might suit better for data set.
    - To compare the performance of our model with that of the existing works.

**Chapter 3**

**METHODOLOGY AND FRAMEWORK**

Our model is built with YOLOv8 Architecture as base focusing on fine-tuning the existing model based on the traffic signs and light.

# Architecture

The YOLO architecture includes: base networks which is convolution networks that perform feature extraction. The second part is the extra layers applied to detect objects on the feature map of the base network. YOLO’s base network uses mainly convolu- tional layers and fully connected layers as shown in Figure [3.1.](#_bookmark13) YOLO architectures are also quite diverse and can be customized into versions for many different input shapes.

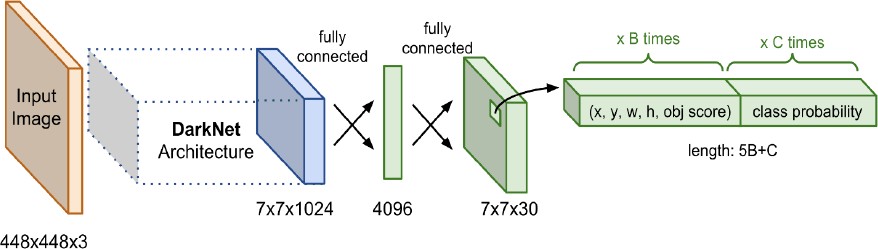


Figure 3.1: YOLO’s Architecture

In YOLO, apply a network feature extractor darknet-53 as shown in Figure [3.2.](#_bookmark14) This network consists of 53 consecutively connected convolutional layers, each of which is followed by a batch normalization and a Leaky Relu activation. To reduce the size of the output after each convolution layer, down sample is used with filters of size 2. This trick has the effect of minimizing the number of parameters for the model.

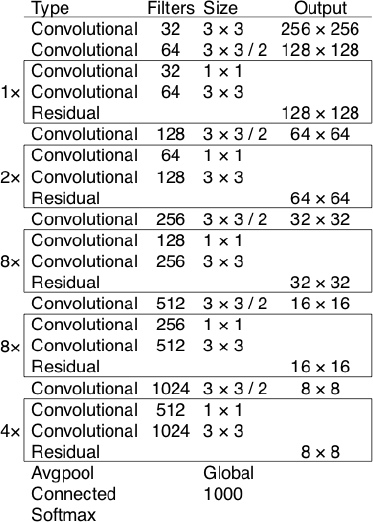


Figure 3.2: Layers in the Darknet-53 network

The images, when fed into the model, will be scaled to the same size as the input shape of the model and then collected into a batch for training.

Currently YOLO is supporting 2 main inputs, 416x416 and 608x608. Each input will have its own design of layers that match the shape of the input. After going through the convolutional layers, the shape decreases exponentially by 2. Finally, we get a feature map of relatively small size to predict the feature on each tile of the feature map.

The size of the feature map will depend on the input. For input 416x416, the feature map has dimensions of 13x13, 26x26 and 52x52. The input is 608x608 will generate 19x19, 38x38, 72x72 feature map.

# Loss Function

During this process, loss function is calculated. For the object recognition, it is nec- essary to calculate 3 loss functions.

* + - Classification Loss Function
    - Localization Loss Function which is used to calculate the error between predicted bounding boxes and anchor boxes, improved Localization Loss will make object detection more accurate.
    - Confidence Loss Function shows the error between the prediction of the bounding box and the actual label.

# Output

At the output, each grid will do the prediction of the 2 bounding boxes with the highest pc. Remove all low pc bounding boxes from the whole image. And finally, Non- max suppression algorithm will be implemented to keep the most accurate bounding box for each object as shown in Figure [3.3.](#_bookmark17)

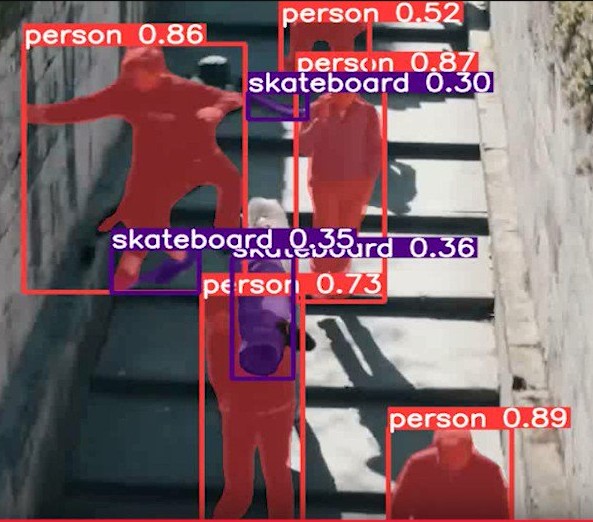


Figure 3.3: Algorithm is applied at output

# YOLOv8 Algorithm

YOLOv8 is a native extension of YOLOv5 PyTorch. These improvements were orig- inally called YOLOv5 but due to the recent release of YOLOv5 in the framework, to avoid version conflicts, it was renamed YOLOv8.

The YOLOv8 algorithm basically also inherits the basic methods of the YOLOs, how- ever, YOLOv8 applies some fast object detection algorithms, optimizing parallel opera- tions to increase recognition speed and reduce optimal training time.

# Data Set

Creating a dataset involves the following steps:

1. Initiate a fresh project.

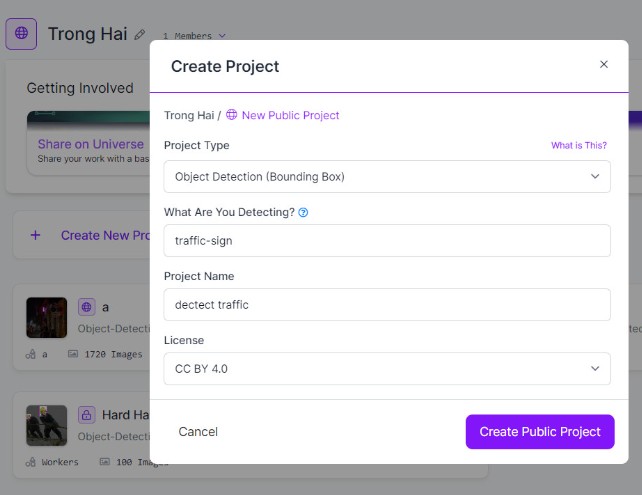


Figure 3.4: New Project

1. Import current traffic lights and signs into Roboflow after .

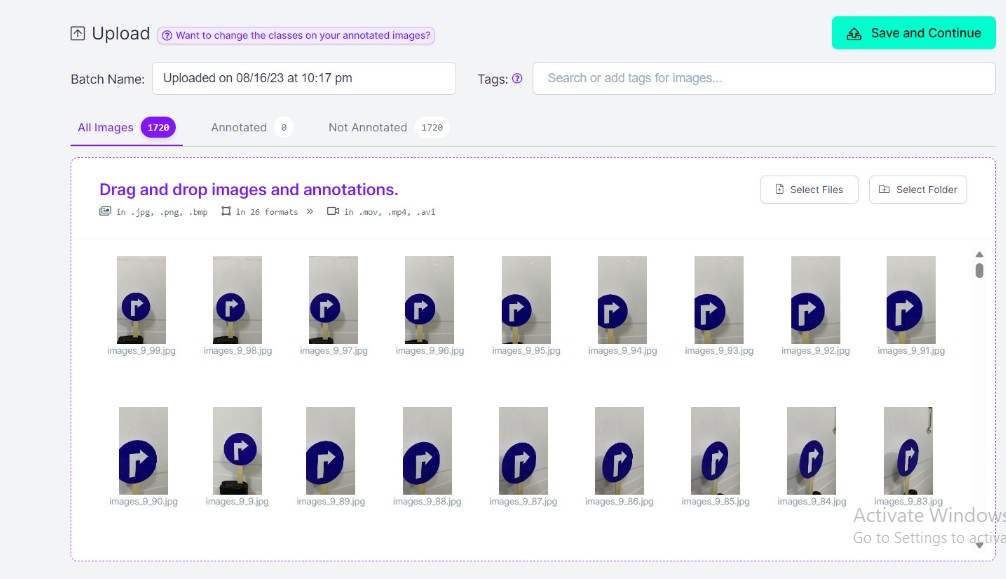


Figure 3.5: Upload Images

1. Proceed to segment frames, classify, and label each image to establish a training dataset file.

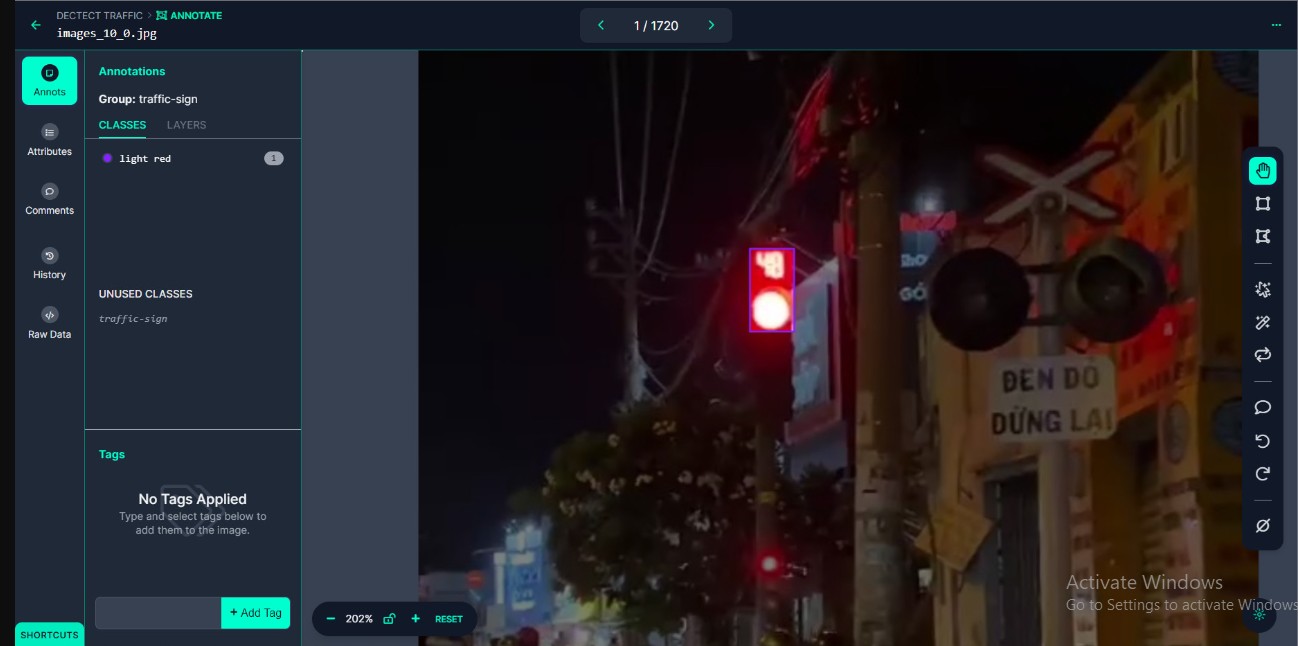


Figure 3.6: Cut Frame Class And Label

1. Ultimately, generate the training file comprising image and label components in two sections.

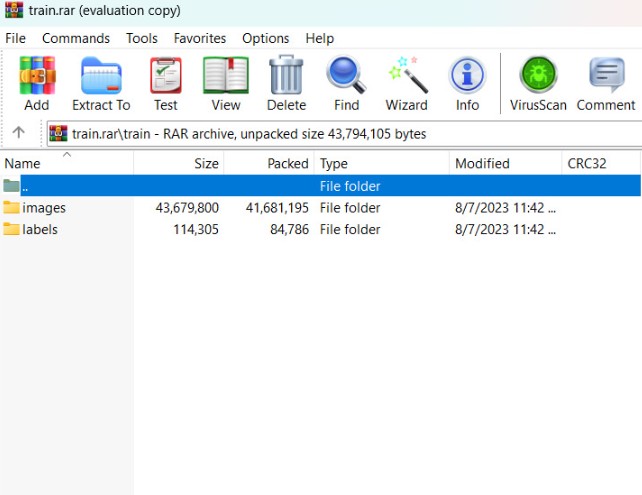


Figure 3.7: File Train

**Chapter 4**

**CODE AND RESULTS**

# Code

Initially, we establish the pathway to Google Drive. Subsequently, we navigate to the YOLOv8 GitHub repository to generate a ’requirements.txt’ directory tailored for our training data.

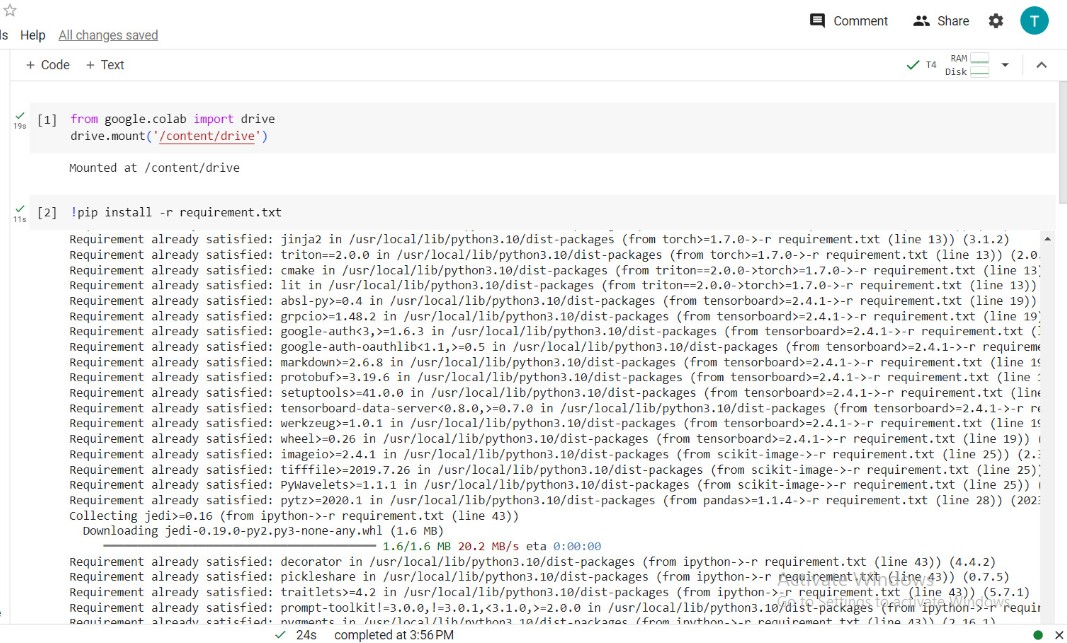


Figure 4.1: Link with google drive

Following that, we will employ the dataset wherein frames and corresponding labels have been segregated, and place them into the designated folder within Google Drive.

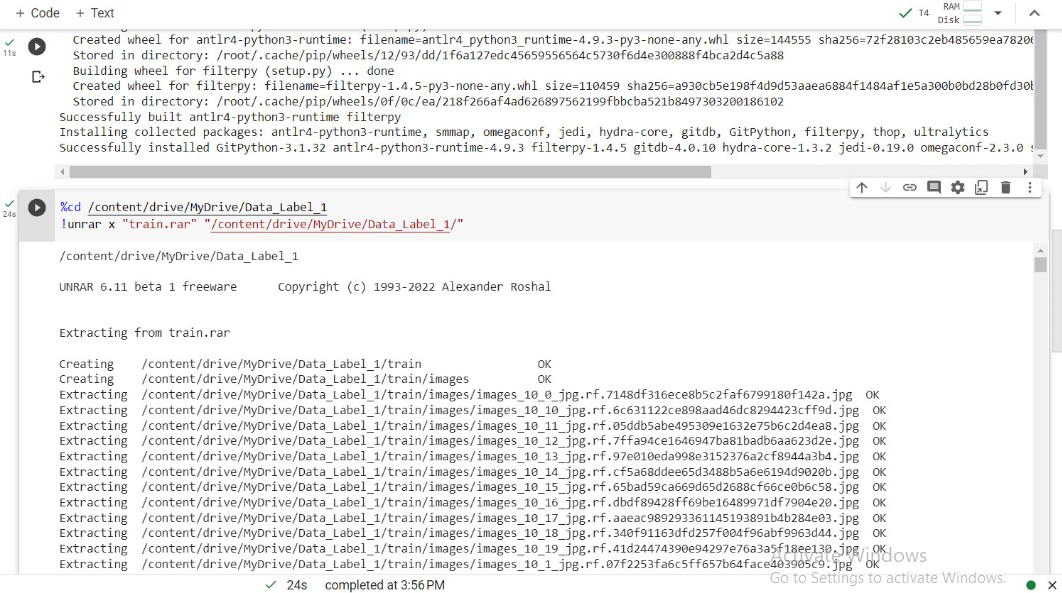


Figure 4.2: Extract dataset

In the subsequent phase, we proceed to download the desired versions of YOLOv8 for training the data.

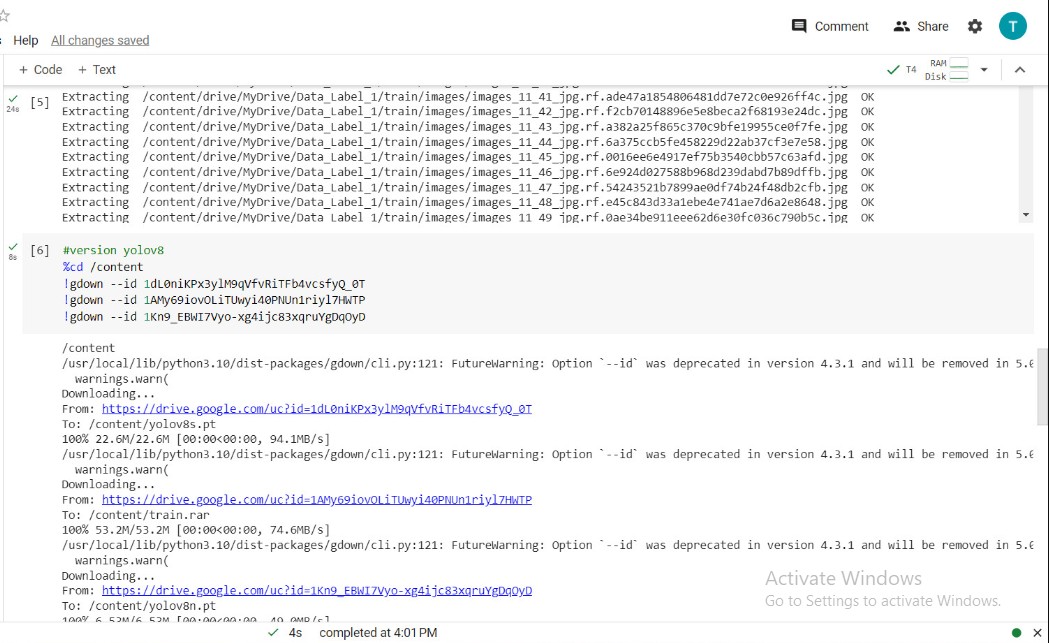


Figure 4.3: Download version YOLOv8

Subsequently, we acquire the Ultralytics environment through downloading.

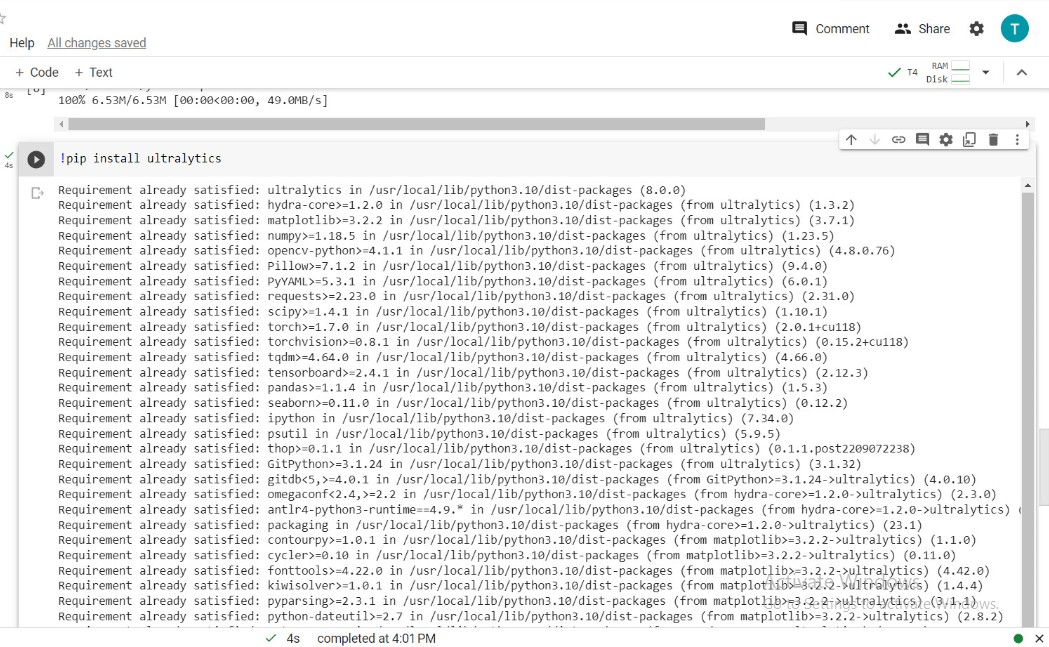


Figure 4.4: Download Ultralytics

Moving forward, we initiate a YAML file to represent the data in textual format.

Upon completing training for 200 epochs, a weight data file will be generated, consisting of two components: "best.pt" and "last.pt".

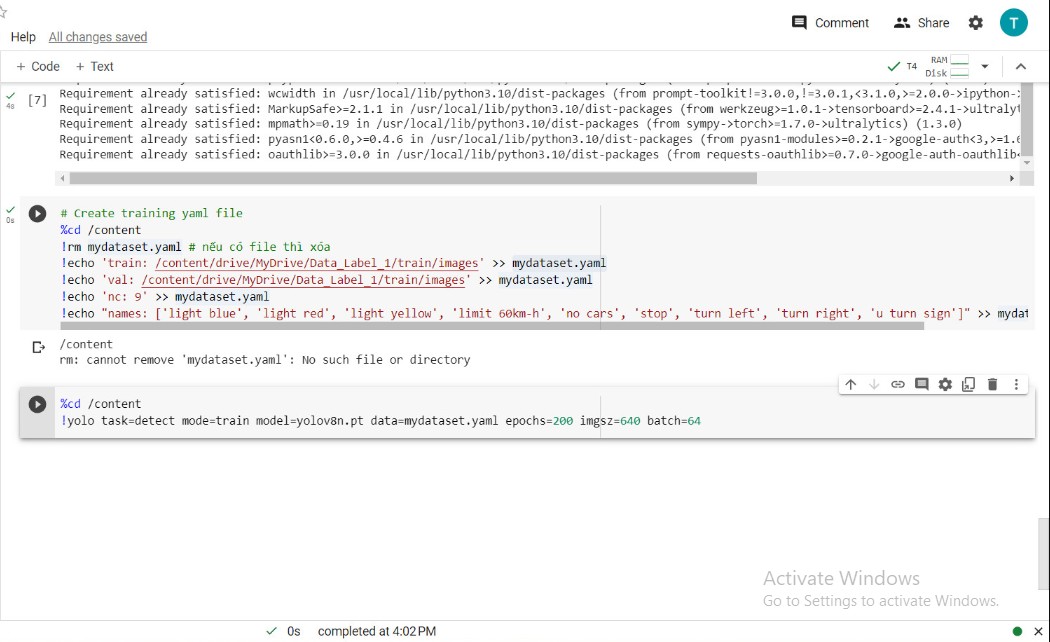


Figure 4.5: Create YAML

With this groundwork laid, we initiate testing to ascertain the model’s ability to rec- ognize objects from images.

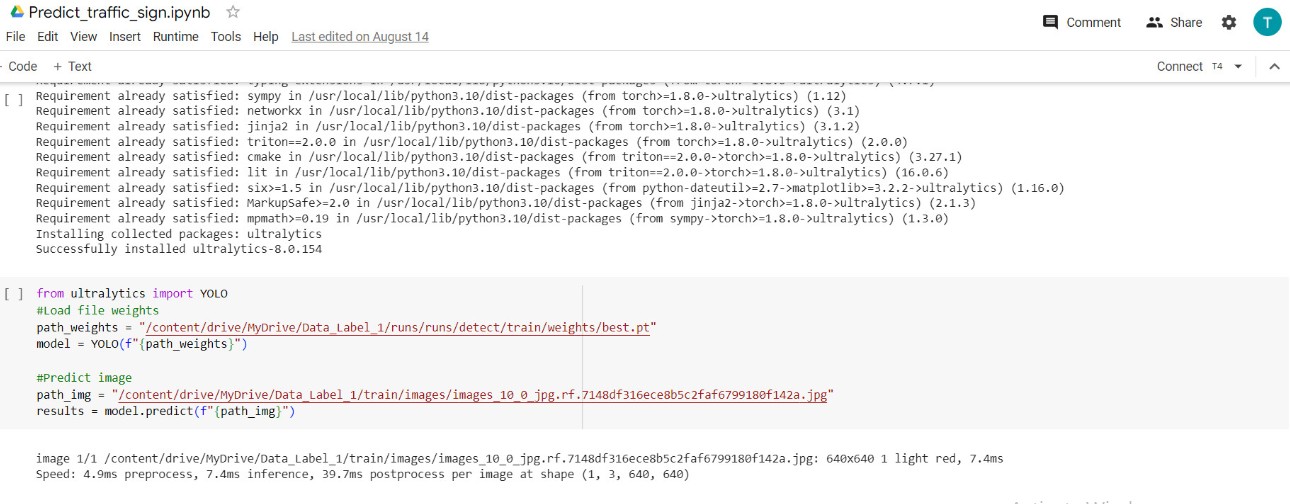


Figure 4.6: Train with file best.pt

Verify the accuracy and exhibit borders utilizing specified parameters.

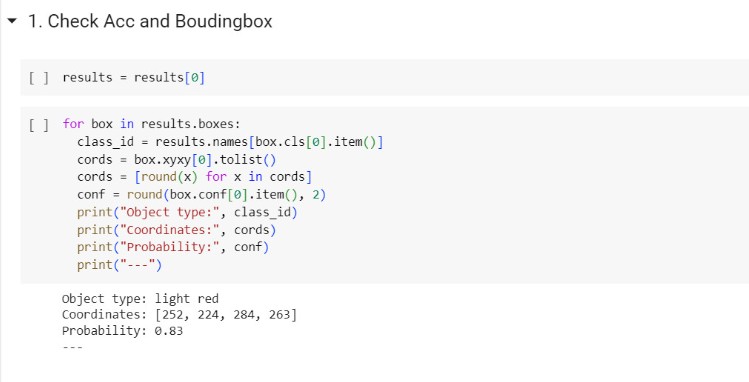
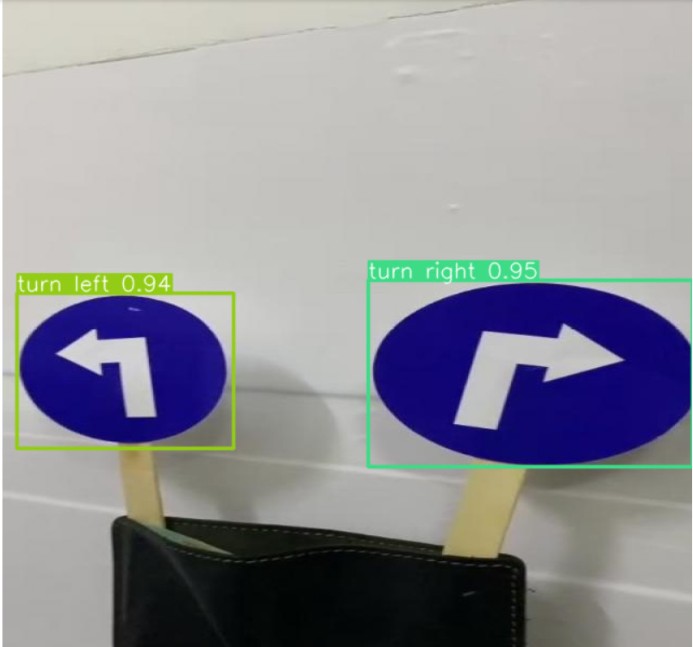


Figure 4.7: Check Acc and Bounding box

# Results

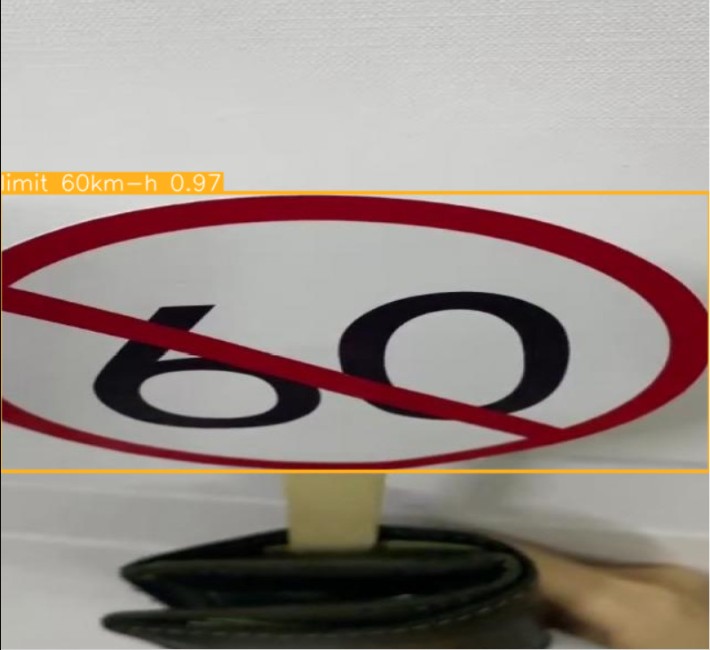
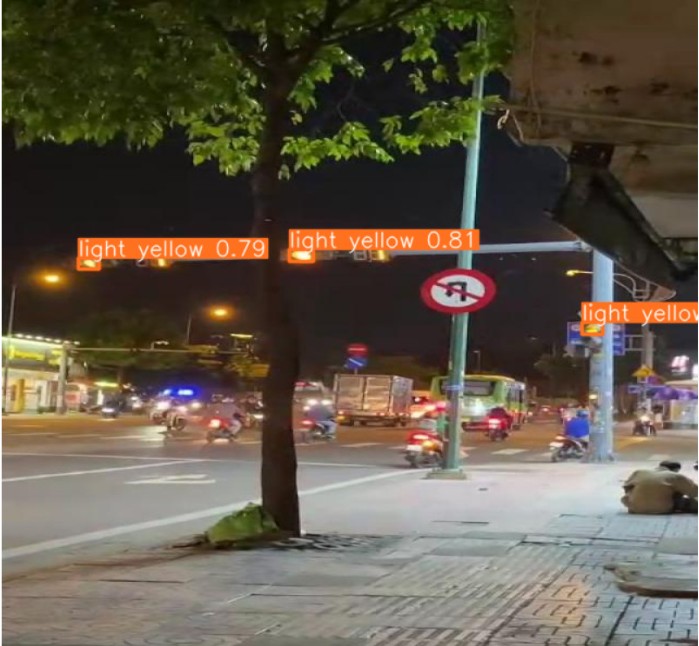


Figure 4.8: Results of all processes

**Chapter 5**

**CONCLUSION AND RECOMMENDATION**

# Conclusion

## Advantages

At the outset, a training model was constructed to identify and classify signs, em- ploying YOLOv8 network algorithms for effective sign detection and recognition. The robust Google Colab platform was utilized for building and training machine learning and deep learning models, leveraging Google’s GPU capabilities to optimize time effi- ciency.

The image classification predictive model’s identification and validation outcomes for traffic notifications exhibit a notable degree of precision and a reasonably consistent performance.

## Disadvantages

Because of the learning and system development time constraints, along with the constraints in knowledge, the system has undergone testing only with 9 distinct sign types. Moreover, given the extensive nature of the research domain, the operational

phase of the system has remained confined to the realm of research. Thus, it cannot be implemented practically and lacks full functionality for end-user utilization.

# Recommendations

* Expand the traffic sign database.
* Enhance the approach to address scenarios involving damaged or obscured signs.
* Upgrade and refine the system’s capabilities to transform it into a comprehensive recognition and real-time alert system for traffic participants within a complete pro- gram.

**Bibliography**

1. B. Gašparovic´, G. Mauša, J. Rukavina, and J. Lerga, “Evaluating yolov5, yolov6, yolov7, and yolov8 in underwater environment: Is there real improvement?” in *2023 8th International Conference on Smart and Sustainable Technologies (SpliTech)*, 2023, pp. 1–4.
2. Fatma M. Talaat, Hanaa ZainEldin, “Road traffic sign detection and recognition using adaptive color segmentation and deep learning,” in *2021 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*, 2021, pp. 134–139.
3. <https://www.mdpi.com/1424-8220/23/16/7190>
4. https://www.scirp.org/journal/paperinformation.aspx?paperid=126758