

# Artificial Intelligence Term Project Report

Mahsun ALTIN - 110510047

## Abstract

Computer Vision is one of the most important field in AI; and there are still some unsolved problems on it. CNN is a type of neural network that is generally used to classify images in AI. In this project, Microsoft Coco API is examined and analysed; also trained a gestures classifier neural network via using Tensorflow in Google Colab platform.

## **Contents**

<b>Abstract</b>	<b>0</b>
<b>Contents</b>	<b>1</b>
<b>Artificial Intelligence Term Project Report</b>	<b>2</b>
<b>1. Introduction</b>	<b>2</b>
<b>2. Related Work</b>	<b>2</b>
<b>2. 1. Coco API</b>	<b>3</b>
<b>2. 2. Hand Gesture Classifier</b>	<b>4</b>
<b>2. 2. a. Dataset</b>	<b>4</b>
<b>2. 2. b. Methods</b>	<b>4</b>
<b>2. 2. c. Results</b>	<b>4</b>
<b>3. Conclusion</b>	<b>4</b>
<b>4. References</b>	<b>4</b>



## Artificial Intelligence Term Project Report

### 1. Introduction

In today's world, most of technological developments happened and continue to happen. One of the technological field is Artificial Intelligence no doubt. If AI is examined, it can be easily seen that there are also many different type of fields in AI. One of these fields is Computer Vision. When viewed from sources of today, it will be seen that there are still many unsolved problems in some fields of CV such as Drones, Robots, Traffic Cameras, Autonomous Cars etc. Mainly, most of today's car companies have investment in that field. They can detect the objects very well. However, when thinking the moving humans' behaviours classification, these do not have very good results actually. As an example of that, matterport, openPose by CMU, Microsoft has really useful works

that is Microsoft Coco API for object detection.

In this project, firstly, one of the leading companies' project, that is Microsoft Coco API, will be examined and tried to use, secondly, what problems and outcomes are seen when processing the data, will be tried to understand through using Tensorflow. All of these will try on Google Colab platform, which is recommended for research projects. This platform is another useful platform to try and to understand basic projects freely. This project will be mainly on hand gestures and will try to classify their meanings.

### 2. Related Work

Currently, many projects and methods available to classify hand gesture activities especially human activity recognition via using by CNN.

Convolutional Neural Network is a kind of neural network generally to classify images.

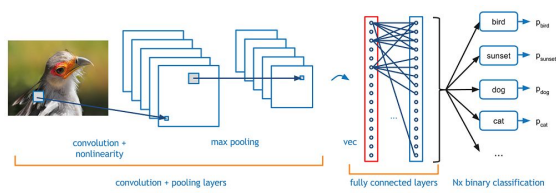


Figure 1

However, videos are cause of bigger datas, they need another different method and that is RNN. With this method, they produce another different Neural Network which is C3D.

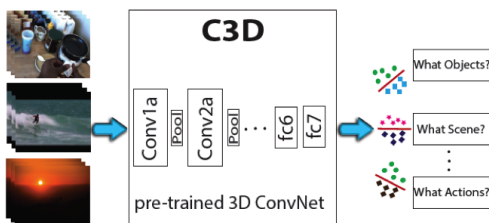


Figure 2

In this project, firstly Coco API will be analyzed, secondly, a hand gestures recogniser will be trained via Tensorflow.

## 2. 1. Coco API

Coco API is published by Microsoft to detect objects and to mask or

find key points etc. Mainly, this API can detect almost perfect percentages. You can see an example of results in figure 3 and figure 4.

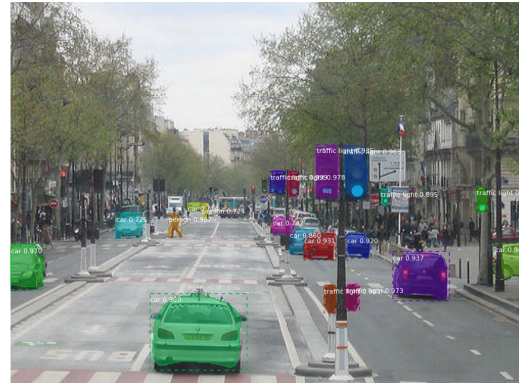


Figure 3



Figure 4

I worked on Google Colab platform to detect objects and mask with Coco API. Figure 4 is gave by me to Coco API; and as it can be seen that I got 100% percentage from Coco API. Figure 3 is taken from main project of Coco API.

## 2. 2. Hand Gesture Classifier

### 2. 2. a. Dataset

Currently, there are many kind of available datasets for human activity classification as videos. These are mainly MPII Human Pose, ActivityNet, UCF101, Sports 1M, KTH datasets. However, there is just some example to recognise hand gestures in [github](#). Also, in this project, our small dataset, which is produced by me, will be used and; labeled via [LabelImg](#). These dataset will be just to understand the logic, therefore it will be small dataset as can be seen in Figure 6.

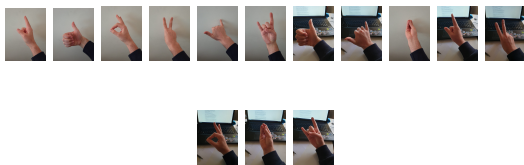


Figure 6

### 2. 2. b. Methods

In this project, number of training steps is defined as 1000 and [ssd\\_mobilenet\\_v2](#) model is used, which is

type of Tensorflow model. You can also examine the code from [Colab](#).

### 2. 2. c. Results

As a result, with this small dataset, these results such as between %95 and %99 percentage of prediction could be taken easily. However, if a hand gesture from different perspective will be given, our model cannot predict very well. Therefore, it should be used bigger dataset such as in 2.2.a Datasets or; there should be more labeled image which labeled by humans, but this is another heavy work.

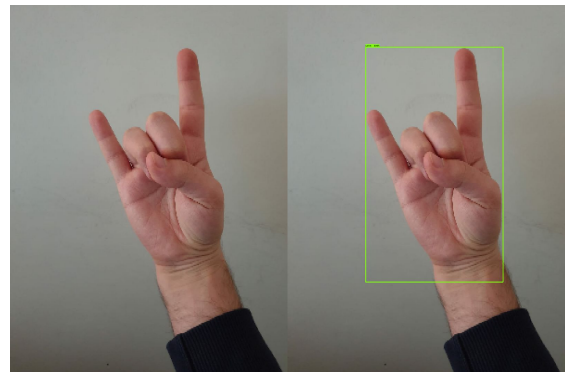


Figure 7

## 3. Conclusion

As a conclusion, this project open a door for me to understand what is Coco API and how neural network can be



trained. Also, If bigger datasets can be found, these results will increase and will be large scaled. As a conclusion, in today's technology even I can get some results and examine these with Google Colab freely. Surely, in the future like after 5 or 10 years, we will get very good results in whole fields of our life like autonomous cars, medical, traffic lights and security cameras etc..

#### 4. References

- <https://github.com/tensorflow/models/>
- <https://www.dlology.com/blog/how-to-train-an-object-detection-model-easy-for-free/>
- <https://github.com/cocodataset/cocoapi>