**Project Document**

**CSCE 5390 (Summer 2022)**

**Project Title – Video Objection Detection**

Group Number – 13

Group Members

1. TARUNENDRA MALEPATI – 1399679
2. ROJA KAMBLE – 11454258
3. ABDUL RAZZAQ RAHIM – 11521098
4. PRUDHVI RAJ RAVI- 11546774
5. CHAITANYA GALLA -11554881

**Table of Contents**

**1.Introduction**

1.1 Purpose

1.2 Project Scope

**2. Project Overview**

2.1 Abstract

2.2 Introduction of YOLO

**3. Project Implementation**

3.1 Software tools

3.2 Libraries

3.3 Steps/Commands to implement

3.4 Code snippets

3.5 Output screenshots

**4. Future Scope**

**5. Conclusion**

**1.Introduction**

* 1. **Purpose**

The main idea of our project is to detect the objects from the sample videos that are uploaded in the system.

**Goals:**

* Implementing a system that detects the object in the video
* Searching the videos content based on the keywords
* Custom object searches in a video

**1.2 Project Scope**

**Scope:**

* Video files could be of any formats. For example, .mp4, .avi, .mov, .mkv, etc.
* Detection of objects
* New video can be uploaded, and it has to undergo backend processing
* Custom object detection
* Video length of short to medium length is preferable

**2. Project Overview**

**2.1 Abstract**

This 21St Century is data-driven, the information has achieved great   
value, and abstraction of information and analysis has been a vital role.   
The Cameras have the most information of all, and the techniques of   
information extraction from images and videos are increasing daily.   
We have planned to build an AI-based project, which has neural   
networks and deep learning frameworks. The idea is to build a system   
that has a video content search engine that helps to sort and search the   
contents of the video. To build this first we need to identify the objects   
in the video and then we need to apply machine learning techniques to   
categorize the contents and objects detected in the video. With this we   
can label each object that has been detected in the video and then when   
we build the search engine, we can search for the object of our desire   
and as we have labels for the video, we can get the redefined videos that   
have the contents of the searched objects. We need to first convert the   
video into frames and then each frame must be passed through an   
algorithm to get the object in the respective frames. The frames which   
have objects should be labeled accordingly and then the frames need to   
be concatenated. The whole system must have a search engine this helps   
to search the videos that have objects and the contents that we desire.   
Our video recording contains many objects like different persons,   
landscapes, buildings, and various objects.

**2.2 Introduction of YOLO**

The state-of-the-art, real-time object detection technology is called You Only Look Once (YOLO). To train your AI models to recognize objects, a Limited Number of Effective Object Detection Tools for Training Your AI Models.

* Classifications
* **Boxes with Bounds**
* Polygons
* Spans and Lines
* Key points in the Facial Recognition Point Cloud

In comparison to other detectors, YOLOv3 is incredibly quick and precise. YOLOv3 is comparable to Focal Loss in mAP measured at.5 IOU, but it is nearly four times faster. Additionally, you can easily compromise between accuracy and speed by just altering the model's size; no retraining is necessary!

**Comments:** mAP (mean Average Precision)

The accuracy of object detectors is frequently assessed using the statistic known as AP (Average precision). For recall values greater than or equal to 0, average precision calculates the average precision value.

The **precision** of your predictions is measured. i.e., what proportion of your predictions were accurate. Whereas **recall** gauges how well you remember all the advantages. For instance, our top K forecasts contain 80% of the potential positive cases.

**What has Version 3 added?**

YOLOv3 employs several techniques, such as multi-scale predictions, a better backbone classifier, and others, to enhance training and boost performance.

**How YOLO Works?**

Classifiers or localizers are repurposed for detection in earlier detection systems. At various scales and locations, they apply the model to an image. Regions of the image with high scores are referred to as detections.

We adopt a completely new methodology. We process the entire image using a single neural network. By dividing the image into regions, this network can forecast bounding boxes and probabilities for each region. These bounding boxes are weighted based on the probability that were projected.

**3. Project Implementation**

The project has 3 elements to search the video having a object, to search the objects in the videos and to search the specific objects in the video

3.1 Software tools

i) PyCharm

ii) Python v – 3.7

iii) Packages needed:

tensorflow

keras

opencv

3.2 Libraries

tensorflow==2.4.0

keras==2.4.3

numpy==1.19.3

pillow==7.0.0

scipy==1.4.1

h5py==2.10.0

matplotlib==3.3.2

opencv-python

keras-resnet==0.2.0

3.3 Steps/Commands to implement

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

3.4 Code snippetsText

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

Text

Description automatically generated

3.5 Output screenshots

A picture containing text, outdoor

Description automatically generated

Graphical user interface, application, table

Description automatically generated

**4.Future Scope**

The scope of this project if implemented in a large has wide application we can sort the videos according to the content like in google photos we can search only specific content in the videos. We can add the time frame of the video and then we can sort it accordingly.

**5.Conclusions**

The main aim of the project has been achieved. We are successful in building an artificial intelligence system that can detect objects of 200 kinds. We have developed a user interface that makes the system perform 3 different tasks of searching the videos has the content, video object detection this will intern help to train the system and we have also developed a custom detection system. The possibilities of the system can be even more enhanced with more processing and training of the system.

**Credits**

* [**https://pjreddie.com/media/files/papers/YOLOv3.pdf**](https://pjreddie.com/media/files/papers/YOLOv3.pdf)
* [**https://docs.nvidia.com/metropolis/TLT/tlt-user-guide/text/object\_detection/yolo\_v3.html**](https://docs.nvidia.com/metropolis/TLT/tlt-user-guide/text/object_detection/yolo_v3.html)
* [**https://github.com/rbgirshick/rcnn**](https://github.com/rbgirshick/rcnn)
* [**https://github.com/rbgirshick/fast-rcnn**](https://github.com/rbgirshick/fast-rcnn)