Galgotias University

Plot No. 2, Sector 17-A, Yamuna Expressway
Greater Noida, Gautam Buddha Nagar,
Uttar Pradesh -201310

SHOOL OF COMPUTER SCIENCE AND ENGINEERING

PERSON TRACKING IN LIVE VIDEO FEEDS

Project Report submitted in partial

Fulfillment for the award of degree of

Bachelors of Technology

Group No.:BT8283

S.No	Enrollment Number	Admission Number	Student Name	Degree/ Branch	Sem
1.	1713101234	17SCSE1077022	ANSHIKA SONI	B.tech / CSE(BA)	8th
2.	1713101441	17SCSE101459	HARSH PRATAP	B.tech / CSE	8th
3.	1713101173	17SCSE101183	RACHIT SRIVASTAVA	B.tech / CSE	8th

Under the Supervision of

Dr. S. Annamalai

School of Computing Science and Engineering

Greater Noida,Uttar Pradesh

Fall 2020 – 2021

GALGOTIAS
UNIVERSITY

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING

BONAFIDE CERTIFICATE

Certified that this project report “**PERSON TRACKING IN LIVE VIDEO FEEDS”** is the bonafide word of **ANSHIKA SONI, HARSH PRATAP, RACHIT SRIVASTAVA** who carried out the project work under my supervision.

SIGNATURE OF DEAN SIGNATURE OF
SUPERVISOR

Approval Sheet

This thesis/dissertation/report entitled **PERSON TRACKING IN LIVE VIDEO FEEDS** by **Anshika Soni, Harsh Pratap, Rachit Srivastava** is approved for the degree of BACHELORS OF TECHNOLOGY.

Examiners

____

____

Supervisor

____

____

Date:______________

Place:_______________

Statement of Project Report Preparation

1. Thesis title: Person Tracking in Live Video Feeds
2. Degree for which the report is submitted: B.tech
3. Project Supervisor was referred to for preparing the report.
4. Specifications regarding thesis format have been closely followed.
5. The contents of the thesis have been organized based on the guidelines.
6. The report has been prepared without resorting to plagiarism.
7. All sources used have been cited appropriately
8. The report has not been submitted elsewhere for a degree.

(1st student)

Anshika Soni - 1713101234

(2nd Student)

Harsh Pratap - 1713101441

(3rd Student)

Rachit Srivastava – 1713101173

Acknowledgement

We sincerely express indebtedness to esteemed and revered guide **Mr. S.Annamalai** Sir of Department of Computer Science and Engineering for his invaluable guidance, supervision, and encouragement throughout the work. Without his kind patronage and guidance, the synopsis would not have taken shape. We take this opportunity to express a deep sense of gratitude to our guide for his encouragement and kind approval. In addition, we thank him for providing the facility. We would like to express our sincere regards to him for advice and counseling from time to time. We owe sincere thanks to all the faculties in the Department of Computer Science and Engineering for their advice and counseling from time to time.

Table of Contents

Abstract

The idea is to utilize the combination of Deep Learning Object Detection, using YOLO (You Only Look Once) and Deep SORT (Simple Online and Realtime Tracking). We are looking into first classifying the person inside the video frames by passing the frames into the neural networks and based on the output co-ordinates we track the person using SORT. The idea is to track the person even when his/her face is not visible.

There will be utilization of feature extraction and mapping along with frames history. The analysis is fully done in videos after decomposition of videos. The feature like clothes color, location coordinates will be utilized for this purpose. The system will fully be capable of tracking multiple suspects in real-time. The system need not to be fed with facial data, though it will be a useful and implementable for improved accuracy. The system requires manual feeding of target, which can be done by manual targeting or a presumed database with suspicious person data, where descriptors of the person is gives. The system utilizes Kalman Filter of original SORT algorithm for this purpose and the outcome of this system is a warning or analysis on a person.

Literature Review/ Comparative Study

Video surveillance system market in India is at its emerging pace and is increasing day by day. Video surveillance is a system that uses video cameras to monitor a person or place, especially in camera; the recording composed by such surveillance. India has the world's second largest population, an enlarging middle class and as India is a developing country so for sure it's a huge market for international investors. Some of the world's largest corporations have offices in India.

The market of video surveillance in India has emerge continuously in last 20 years and it has seen multiple changes in its market dynamics especially after 2008 Mumbai terrorist attacks, the Government of India has taken several initiatives to install intelligent security systems such as video surveillance systems. The video surveillance system is now used as security weapons across societies, government & transportation, etc.

In India, video surveillance camera market had seen the entry of Canon, a major player in the Digital Imaging industry. The company has taken steps into the surveillance domain with the new range of IP cameras featuring intelligent video analytics. The major factor that propels the growth of India video surveillance market is initiatives taken by government to install video surveillance because of increasing number of crimes in India. The major restraint that can hamper the growth of India video surveillance market is the alternatives for video surveillance rapidly changing technological environment and increase in the flux of private labels tends to impact global vendors of this market.

Problem Formulation

- We have data in hand but we are not able to utilize it accordingly.
- Manual handling of tracking of people are done which requires heavy manpower.
- There are multiple rooms as well as cases of human error, where our program would have proved to be more efficient.
- Police narrows down the location of suspect based on information from unreliable sources, which can be done automatically.
- Currently there is a requirement of police officers to install blockade based on unreliable information.
- All cases of suspect of different crimes are handled manually.

Police have difficulty in tracking a certain person or suspect in video feeds from CCTV cameras. They require a manual searching of person to track a person in video feeds.

Current Scenario: The videos are manually scanned for suspects by police officers or another person. This process requires a consistent concentration and there is a large room for human error, sometimes there is insufficient data for human to analyze. There is a massive number of videos for analysis and it is a very slow process of analysis. Expected Scenario: The application after deployment, it won’t require manual handling of analysis for videos the application will be fully capable of analyzing huge videos without manual handling and room for human error will be decreased greatly.

Tools Required for Implementation

❖ Software Requirements:

- Custom YoloV4
- Google Open Image Dataset
- TensorFlow or Darknet for deployment
- OpenCV for Python
- Google Cloud or Other Cloud Platform
- MongoDB for Database
- DeepSort trained on person re-id dataset

❖ Hardware Requirements for Training:

- I5 Processor Based Computer or higher
- Memory: 4 GB RAM or Higher
- Hard Drive: 50 GB
- NVIDIA GTX1050ti or Higher

Description of Algorithms Used

YOLOv4

YOLO, as stated, stands for You Only Look Once, it is an object detection system in real-time that recognizes various objects in a single enclosure. Moreover, it identifies objects more rapidly and more precisely than other recognition systems.

As completely based on Convolutional Neural Network(CNN), it isolates a particular image into regions and envisioned the confined-edge box and probabilities of every region. Concurrently, it also anticipates various confined-edge boxes and probabilities of these classes.

YOLOv4 outruns the existing methods significantly in both terms “detection performance” and “superior speed”. In reference to a paper published, the research team mentions it as a “speedily operating” object detector that can be trained smoothly and used in production systems.

The main objective was “to optimize neural networks detector for parallel computations”, the team also introduces various different architectures and architectural selections after attentively analyzing the effects on the performance of numerous detector, features suggested in the previous YOLO models.

YOLOv4 consists of;

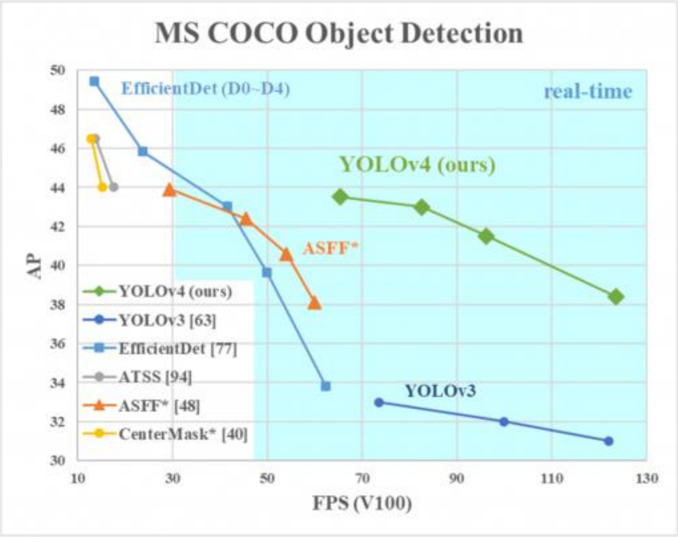
Backbone: CSPDarknet53,

Neck: Spatial Pyramid Pooling additional module, PANet path-aggregation, and

Head: YOLOv3

CSPDarknet53 is a unique backbone that augments the learning capacity of CNN, the spatial pyramid pooling section is attached overhead CSPDarknet53 for improving the receptive field and distinguish the highly important context features.

The PANet is deployed in terms of the method for parameter aggregation for distinctive detector levels rather than Feature pyramid networks (FPN) for object detection applied in YOLOv3.



Speed(AP) vs. Performance(FPS) of YOLOv4 and different model

Feasibility Analysis

At recent time, there are CCTV cameras at every corner of roads, highways, commercial areas and everywhere and they are properly surveilled. Those cameras have enough quality input for machines or humans to detect and identify the person. We have the data in hand, what we require is a processing part for those videos to classify and localize classes of interest and run analysis on those classes of interest. If given a system/hardware with enough computing power to handle those to handle those frames it is totally feasible to implement this project. We require a system/server with multiple and sufficient GPUs with CUDA support and high bandwidth network connection installed for transmission of video data to the server and the whole analysis will take place on the server and outputs will be sent to the standalone machine in JSON format.

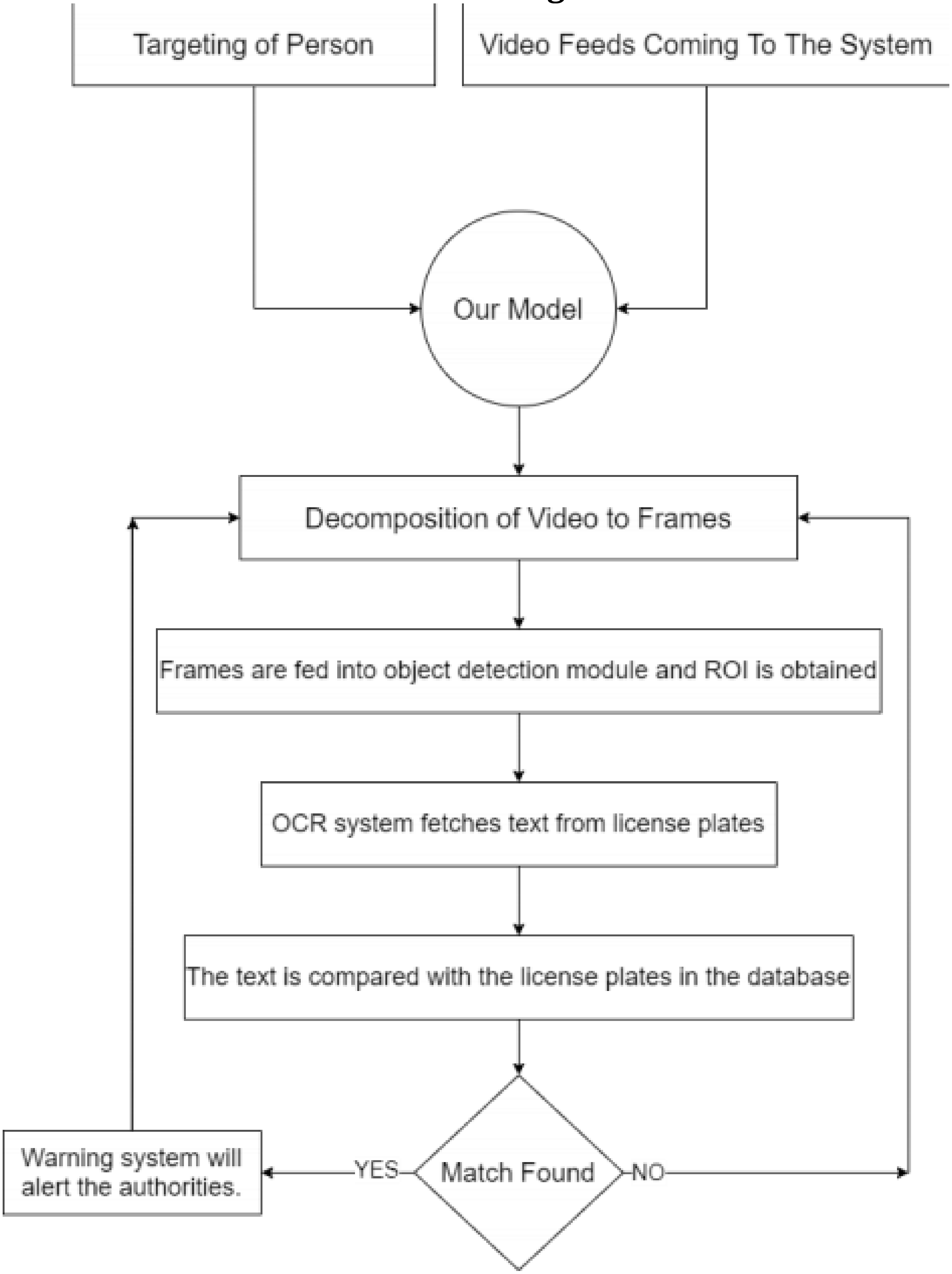
Steps Involved

- We train our Custom YoloV4 on Cloud with custom (person) dataset from Open Image Dataset with converted annotation
- We train deepsort with person Re-ID dataset
- We deploy YoloV4 + DeepSort on TensorFlow 2.1
- The deployment will be done on flask-based server on cloud
- Videos will be sent to the server and timestamp analysis will come as output.

Use Cases

- Any type of suspect tracking
- Evidence Gathering
- Many more use cases concerning police activities

Data Flow Diagram


Targeting of Person
Video Feeds Coming To The System
Our Model
Decomposition of Video to Frames
Frames are fed into object detection module and ROI is obtained
OCR system fetches text from license plates
The text is compared with the license plates in the database
Match Found
YES
NO
Warning system will alert the authorities.

Modules

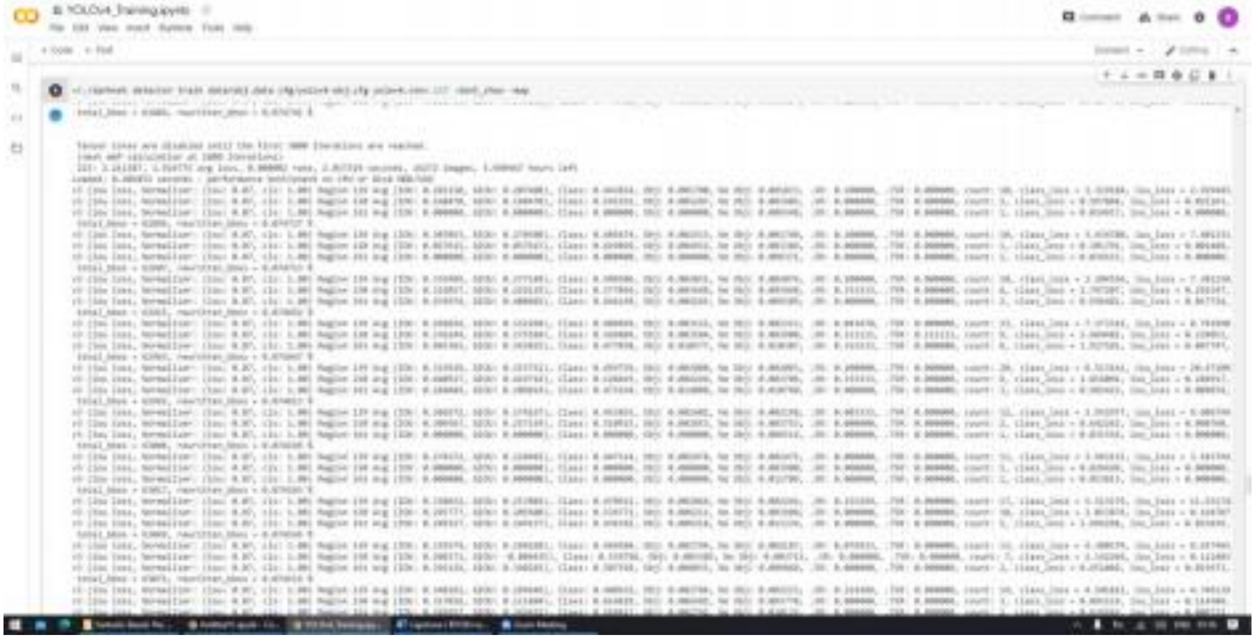
1. Data Collection

- Images

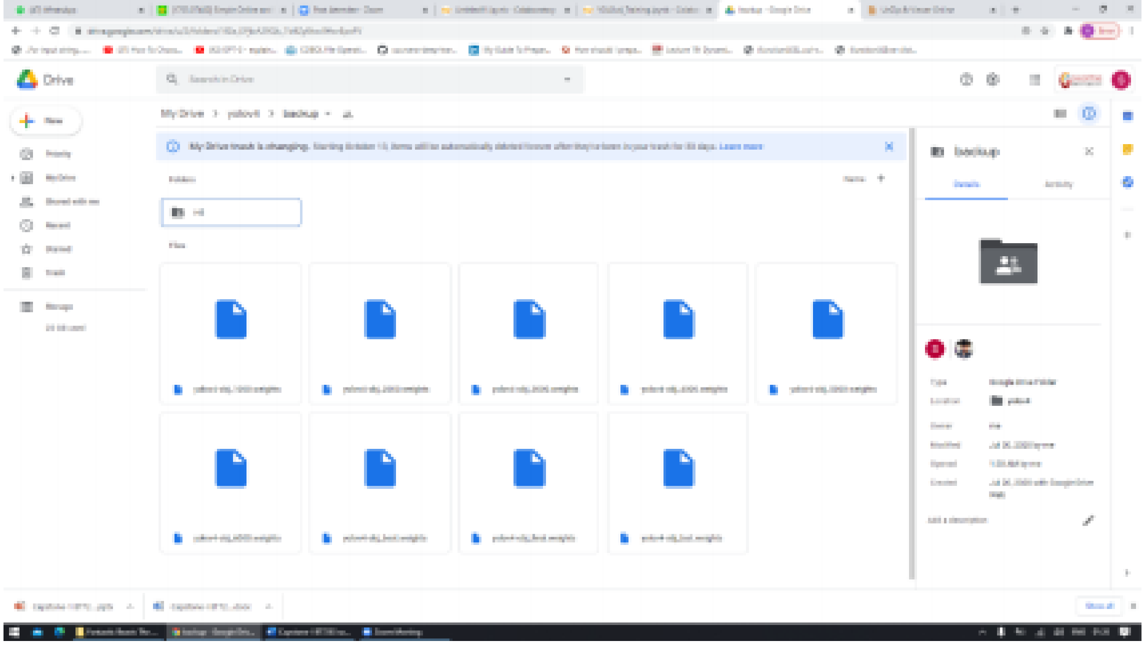
- Annotations

2. Object Model Detection

Model has been trained on Google Colab



The weights files are generated



References

- YOLOv4: Optimal Speed and Accuracy of Object Detection

  https://arxiv.org/abs/2004.10934
- Simple Online and Realtime Tracking

  https://arxiv.org/abs/1602.00763