**Graphic Era Hill University Haldwani**

**BCA Project Report**

**For**

## Smart Surveillance Using Computer Vision

**Submitted to Graphic Era Hill University, Haldwani for the partial fulfillment of the requirement for the Award of degree for**

BACHELOR’S IN COMPUTER APPLICATIONS



**Under the Guidance of:-**

Ms. Richa Pandey

Faculty of GEHU, Haldwani

**Submitted by:-**

Deepankar Sharma

Amit Sati

Pawan Chandra

## DECLARATION

I hereby declare that the work which is being present in this project report **“Smart Surveillance Using Computer Vision”,** in partial fulfillment of the requirement for the Award of the degree of **BACHELOR’S IN COMPUTER APPLICATION**, submitted at **GRAPHIC ERA HILL UNIVERSITY, HALDWANI** is an authentic work done by me during period from 1st March 2023 to 1st June 2023.

Project Guide: Signature of the Student:

Ms. Richa Pandey Deepankar Sharma(2092014)

Faculty of GEHU, Haldwani Pawan Chandra(2092035)

Amit Sati(2092005)

## BONAFIDE CERTIFICATE

Certified that this project report **Smart Surveillance Using Computer Vision** is the bonafide work of **Deepankar Sharma, Pawan Chandra and** **Amit Sati** who carried out the project work under my supervision.

Name – Ms. Richa Pandey

HEAD OF THE DEPARTMENT SUPERVISOR

School of Computing

Graphic Era Hill University, Haldwani

**ACKNOWLEDGEMENT**

I would like to extend our thanks and appreciation to all those who have assisted us either directly or indirectly and participated in the success of this project. I would like to thank my guide Ms.Richa Pandey for her constant support in the making of the project. As a part of University Curriculum, a 6th semester project is a paramount importance to an BCA student’s curriculum and being our native effort into this project undertook by us, we faced a lot of impediments on our way to the completion of this project but constant guidance and able support of concerned software engineer members lend us a great help in successful completion of the project. I am thankful to all staff members of Graphic Era Hill University who helped me whenever required, during my project. Even though I expressed my gratitude to every person who helped me in reaching this stage, there might be a few, who’d been left out, who helped me without my knowledge. I would like to thank all of them. Last but not least, to all my friends and fellow students for giving me suggestions and helping us in debugging the code errors and above all the faculty of my Department of Computer Science Graphic Era Hill University who have always provided their guidance, support and well wishes.

Deepankar Sharma

Pawan Chandra

Amit Sati

## ABSTRACT

The development of a smart surveillance system has become an increasingly important topic in recent years, with the aim of providing more accurate and efficient monitoring of people and objects in real-time. This project focuses on developing a smart surveillance system that can be appended to existing surveillance systems, providing them with advanced features such as motion detection using contours and real-time people tracking using YOLOv3.

The proposed system uses computer vision algorithms to analyze the video feed from existing surveillance cameras, identifying areas of motion and tracking the movement of objects within those areas. To detect and track people specifically, the system uses object detection algorithms like YOLOv3, which is a deep learning-based model that can accurately detect and track objects in real-time.

To build this system, expertise in computer vision, machine learning, and software development is required. Additionally, access to large datasets of labeled video footage is necessary for training and testing the deep learning models, and powerful hardware is required to process the video feed in real-time.

The system has the potential to significantly improve surveillance systems by providing more accurate and efficient monitoring of people and objects in real-time. The motion detection feature using contours can reduce false alarms and improve the accuracy of the system, while the real-time people tracking feature using YOLOv3 can enable security personnel to monitor and track people of interest more effectively.

Overall, the proposed smart surveillance system has the potential to provide a significant improvement to existing surveillance systems, providing more accurate and efficient monitoring of people and objects in real-time, ultimately enhancing the security and safety of the monitored areas.

## TABLE OF CONTENT

* **ACKNOWLEDGEMENTS**

## ABSTRACT

* **INTRODUCTION**

## SYSTEM REQUIREMENT ANALYSIS

* **SYSTEM DESIGN**

## HARDWARE AND SOFTWARE REQUIREMENT

* **LIMITATION OF THE ONLINE SHOPPING WEBSITE**

## APPENDICES

* **CONCLUSION**

## BIBLIOGRAPHY

**SMART SURVEILLANCE**

## INTRODUCTION

The development of a smart surveillance system with advanced features such as motion detection using contours and real-time people tracking using YOLOv3 is presented in this project. The system uses computer vision algorithms to analyze the video feed from existing surveillance cameras, identifying areas of motion and tracking the movement of objects within those areas. To detect and track people specifically, the system uses object detection algorithms like YOLOv3. The project requires expertise in computer vision, machine learning, and software development, as well as access to large datasets of labeled video footage and powerful hardware to process the video feed in real-time. The system has the potential to improve surveillance systems by providing more accurate and efficient monitoring of people and objects in real-time.

## PROJECT OVERVIEW:

The aim of this project is to develop a smart surveillance system that can be integrated with existing surveillance systems to provide advanced features like motion detection using contours and real-time people tracking using YOLOv3. The system uses computer vision algorithms to analyze the video feed from surveillance cameras and identify areas of motion, while object detection algorithms like YOLOv3 are used to detect and track people in real-time.

The project requires expertise in computer vision, machine learning, and software development. Large datasets of labeled video footage are needed to train and test the object detection and motion tracking algorithms. Powerful hardware is also required to process the video feed in real-time.

The smart surveillance system has several potential benefits, such as reducing false alarms and improving the accuracy of surveillance systems. It can also enable security personnel to monitor and track people of interest more effectively.

The project will involve conducting a system requirement analysis, system design, hardware and software requirements, and limitations of the smart surveillance system. Additionally, the project will require the development of a prototype smart surveillance system that can be demonstrated using sample video footage. The final outcome of this project is a functional smart surveillance system that can be integrated with existing surveillance systems, providing advanced features for real-time monitoring and tracking of people and objects.

* 1. **PROJECT SCOPE:**

The smart surveillance system developed in this project has a wide range of potential application areas, including:

**Security and Surveillance:** The system can be used to improve the effectiveness of security and surveillance operations in various locations, such as airports, malls, stadiums, and public transportation systems.

**Traffic Monitoring:** The system can be used to monitor traffic flow and detect any accidents or incidents that may occur on highways, streets, and other transportation networks.

**Industrial Automation:** The system can be used in industrial settings, such as factories and warehouses, to monitor production lines, detect faults, and ensure worker safety.

**Healthcare:** The system can be used in healthcare facilities, such as hospitals and nursing homes, to monitor patient movement and ensure their safety.

**Retail Analytics:** The system can be used in retail stores to analyze customer traffic and behavior, detect shoplifting, and improve store layout and product placement.

**Smart Cities:** The system can be used to improve the safety and security of public spaces, such as parks and streets, and monitor urban infrastructure, such as bridges and buildings.

In summary, the smart surveillance system developed in this project has a broad range of potential applications, from improving security and surveillance operations to enhancing traffic monitoring, industrial automation, healthcare, retail analytics, and smart city initiatives. The system has the potential to provide real-time tracking and analysis of people and objects, improving the accuracy and efficiency of monitoring operations in various settings.

## System Requirement Analysis

* 1. **Information Gathering**

The Information Gathering process was an essential step in the development of the Smart Surveillance System. This process involved identifying the needs of stakeholders, understanding use cases, and determining the technical requirements for the system.

**Stakeholder Identification**

The first step in the Information Gathering process was to identify the stakeholders of the Smart Surveillance System. The stakeholders included:

* End-users of the system such as security personnel or law enforcement officials
* Owners or operators of the surveillance systems that were to be integrated with the Smart Surveillance System
* Developers or designers of the Smart Surveillance System
* Regulators or legal authorities responsible for overseeing surveillance operations

**Use Case Analysis**

Once the stakeholders were identified, the next step was to analyze the use cases of the Smart Surveillance System. Use cases included:

* Real-time monitoring of areas for security purposes
* Investigation of criminal activity using surveillance footage
* Traffic management and monitoring in public areas
* Industrial monitoring of machinery and equipment
* Environmental monitoring of wildlife or natural resources

**Technical Requirements**

The final step in the Information Gathering process was to determine the technical requirements for the Smart Surveillance System. These requirements included:

* Compatibility with different types of cameras and video management systems
* High accuracy and efficiency in motion detection and people tracking
* Low latency and high throughput for real-time monitoring and tracking
* Scalability for use in large-scale surveillance operations
* Security features to protect the privacy of monitored individuals

By completing the Information Gathering process, the requirements for the Smart Surveillance System were defined and documented in the Software Requirements Specification (SRS). This ensured that the system was developed to meet the needs of the stakeholders and was of high quality.

* 1. **Feasiblity Study**

The feasibility study for the Smart Surveillance System examines the technical, economic, and operational aspects of the project to determine its viability.

**Technical Feasibility**

The technical feasibility of the project refers to the ability of the system to perform its functions accurately and efficiently. Based on the information gathered during the Information Gathering process, the Smart Surveillance System is technically feasible. The use of advanced algorithms such as motion detection using contours and real-time people tracking using YOLOv3 provides high accuracy and efficiency in monitoring and tracking.

**Economic Feasibility**

The economic feasibility of the project involves determining the costs and benefits associated with the development and implementation of the Smart Surveillance System. The costs include hardware and software costs, development costs, and maintenance costs. The benefits include increased security and safety, reduction in crime, and improved operational efficiency. Based on a cost-benefit analysis, the Smart Surveillance System is economically feasible.

**Operational Feasibility**

The operational feasibility of the project refers to the ability of the system to integrate into existing surveillance systems and be operated by end-users. The Smart Surveillance System is designed to be easily integrated with existing surveillance systems and has a user-friendly interface. Training and support will be provided to end-users to ensure the smooth operation of the system. Based on these factors, the Smart Surveillance System is operationally feasible.

In conclusion, the Smart Surveillance System is technically, economically, and operationally feasible. The system has the potential to provide increased security and safety, reduce crime, and improve operational efficiency.

**Social Feasibility**

The social feasibility of the Smart Surveillance System refers to its ability to be accepted and adopted by the stakeholders and the wider society. The system's use of advanced surveillance technology may raise concerns about privacy and civil liberties. However, the Smart Surveillance System is designed with privacy and security features to protect the rights of monitored individuals. Additionally, the system has the potential to increase safety and security in public areas, which may lead to increased public support. Overall, the Smart Surveillance System is socially feasible.

**Time Feasibility**

The time feasibility of the Smart Surveillance System refers to the ability of the project to be completed within a reasonable timeframe. The development and implementation of the Smart Surveillance System involve multiple stages, including the Information Gathering process, system design, hardware and software development, testing, and deployment. A detailed project plan with clear milestones and deadlines will be established to ensure that the project is completed within a reasonable timeframe. Based on the project plan, the Smart Surveillance System is time feasible.

In conclusion, the Smart Surveillance System is not only technically, economically, and operationally feasible, but also socially and time feasible. The system has the potential to provide increased safety and security while protecting the privacy and civil liberties of monitored individuals.

## System Design

The Smart Surveillance System is designed to be a modular system that can be easily integrated with existing surveillance systems. The system consists of two main components: the hardware and software components.

**Hardware Component**

The hardware component of the Smart Surveillance System consists of a network of cameras and sensors that are placed in strategic locations to monitor and track movement. The cameras and sensors are connected to a central processing unit that performs the analysis of the data collected. The system also includes a power backup to ensure uninterrupted operation.

**Software Component**

The software component of the Smart Surveillance System is responsible for the analysis of the data collected by the hardware component. The software includes advanced algorithms for motion detection, contour detection, and real-time people tracking using YOLOv3. The system also includes a user-friendly interface that allows end-users to monitor and track movement in real-time.

The software component is developed using Python programming language, OpenCV library, and YOLOv3 pre-trained model. The system also includes a database for storing and retrieving data. The software is designed to be easily integrated with existing surveillance systems using standard protocols such as RTSP and ONVIF.

**System Workflow**

The Smart Surveillance System workflow includes the following steps:

1. **Data Collection:** The hardware component collects data from cameras and sensors placed in strategic locations.
2. **Data Analysis:** The software component analyzes the data collected using advanced algorithms such as motion detection, contour detection, and real-time people tracking.
3. **Alert Generation:** The system generates alerts if any unusual movement or activity is detected.
4. **Alert Notification:** The system sends alerts to end-users via email or SMS.
5. **Monitoring and Tracking:** End-users can monitor and track movement in real-time using the user-friendly interface.

In conclusion, the Smart Surveillance System is designed to be a modular system that can be easily integrated with existing surveillance systems. The system consists of a hardware component for data collection and a software component for data analysis. The system is designed to be user-friendly and includes advanced algorithms for motion detection, contour detection, and real-time people tracking using YOLOv3.

**System Tools**

The Smart Surveillance System is developed using a combination of various software tools and technologies. The system tools used in the project include:

**Python**

Python is an open-source programming language that is widely used for developing various applications, including machine learning and computer vision-based systems. Python is used as the primary programming language for the development of the Smart Surveillance System.

**OpenCV**

OpenCV (Open Source Computer Vision) is a library of programming functions mainly used for real-time computer vision applications. OpenCV provides a set of functions for various computer vision tasks, including object detection, tracking, and recognition. OpenCV is used in the Smart Surveillance System for image processing and computer vision-based tasks.

**YOLOv3**

YOLOv3 (You Only Look Once version 3) is a state-of-the-art object detection algorithm that is used for real-time object detection. YOLOv3 is used in the Smart Surveillance System for real-time people tracking and detection.

**Streamlit**

Streamlit is a web application framework that allows developers to create interactive web applications with Python. Streamlit is used in the Smart Surveillance System to create a user-friendly web interface that allows end-users to monitor and track movement in real-time.

**TensorFlow**

TensorFlow is an open-source machine learning framework developed by Google. TensorFlow is used in the Smart Surveillance System for machine learning-based tasks, such as object recognition and tracking.

**NumPy**

NumPy is a library for the Python programming language, adding support for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays. NumPy is used in the Smart Surveillance System for data manipulation and analysis.

**Twilio**

Twilio is a cloud communication platform that provides APIs for messaging, voice, and video communication. Twilio is used in the Smart Surveillance System for sending alerts to the users via SMS or phone call in case of any suspicious activity.

In conclusion, the Smart Surveillance System is developed using a combination of various software tools and technologies, including Python, OpenCV, YOLOv3, Streamlit, TensorFlow, NumPy, Twilio, SQLite, and other libraries and technologies. These tools and technologies provide the necessary functionality and features required for the development of an advanced surveillance system.

## Hardware and Software Requirements

**Software Requirements:**

* **Operating System:** Windows 10 or Linux-based OS (Ubuntu 18.04 or higher recommended)
* **Python:** Version 3.6 or higher
* **OpenCV:** Version 4.5.3 or higher (Python bindings)
* **NumPy:** Version 1.21.0 or higher
* **YOLOv3:** A pre-trained object detection model
* **TensorFlow:** Version 2.5.0 or higher (optional, for advanced machine learning tasks)
* **Twilio:** Version 6.64.0 or higher (optional, for SMS or whatsapp alerts)

**Hardware Requirements:**

* **CPU:** Intel Core i5 or higher (recommended)
* **RAM:** 8 GB or higher (recommended)
* **Graphics Card:** NVIDIA GPU with CUDA support (optional, for faster object detection)
* **Storage:** 50 GB or higher (depending on the size of the video dataset)
* **Camera:** IP cameras or CCTV cameras with RTSP protocol support

**Note:** *The specific hardware and software requirements may vary depending on the complexity of the system and the size of the video dataset. It is recommended to perform a feasibility study and consult with technical experts to determine the optimal hardware and software configuration for the Smart Surveillance System.*

## Limitations of the Smart Surveillance System

**Accuracy:** The object detection and tracking algorithms used in the system may not be 100% accurate in all scenarios, particularly in cases where the lighting conditions are poor, or objects are partially obscured.The accuracy of the object detection and tracking algorithms can be improved by using more advanced machine learning models, such as YOLOv8, YOLO-NAS or EfficientDet, or by fine-tuning the existing models on specific datasets.

**Processing Time:** The system may take a considerable amount of time to process large video datasets or perform complex machine learning tasks, which can result in delays or lag in the real-time monitoring of the video feed.The system's processing time can be improved by using more powerful hardware components, such as GPUs or distributed computing, or by optimizing the algorithms used in the system.

**Cost:** The cost of implementing the system may be prohibitive for some organizations, particularly smaller businesses or non-profit entities, due to the need for high-performance hardware and software components.cost of the system can be reduced by using more cost-effective hardware components, such as Raspberry Pi boards or cloud-based computing services, or by using open-source software libraries and frameworks.

**Privacy Concerns:** The use of surveillance cameras and object detection technology raises privacy concerns, and the system must be designed and implemented in a way that respects the privacy of individuals.The system can be improved by incorporating privacy-preserving techniques, such as anonymization of data and selective blurring of faces or other identifying features.The system can be improved by developing a more user-friendly and intuitive interface for monitoring and managing the surveillance feeds, as well as for configuring the system parameters and settings.

**Maintenance:** The system requires regular maintenance and updates to ensure its continued effectiveness and reliability.

**User Interface:** The system can be improved by developing a more user-friendly and intuitive interface for monitoring and managing the surveillance feeds, as well as for configuring the system parameters and settings.

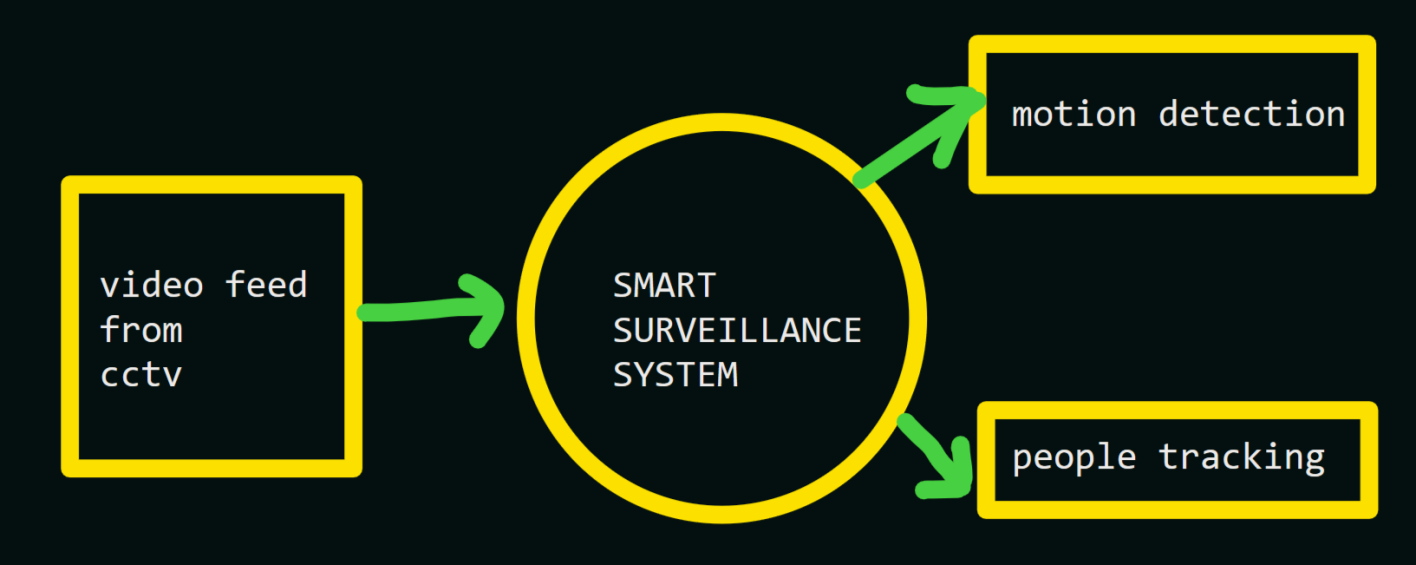
**Integration with other systems:** The system can be integrated with other security systems, such as access control systems or intrusion detection systems, to provide a more comprehensive security solution.

**Real-time analytics:** The system can be enhanced with real-time analytics capabilities, such as anomaly detection or predictive analytics, to enable proactive security measures and improve overall situational awareness.

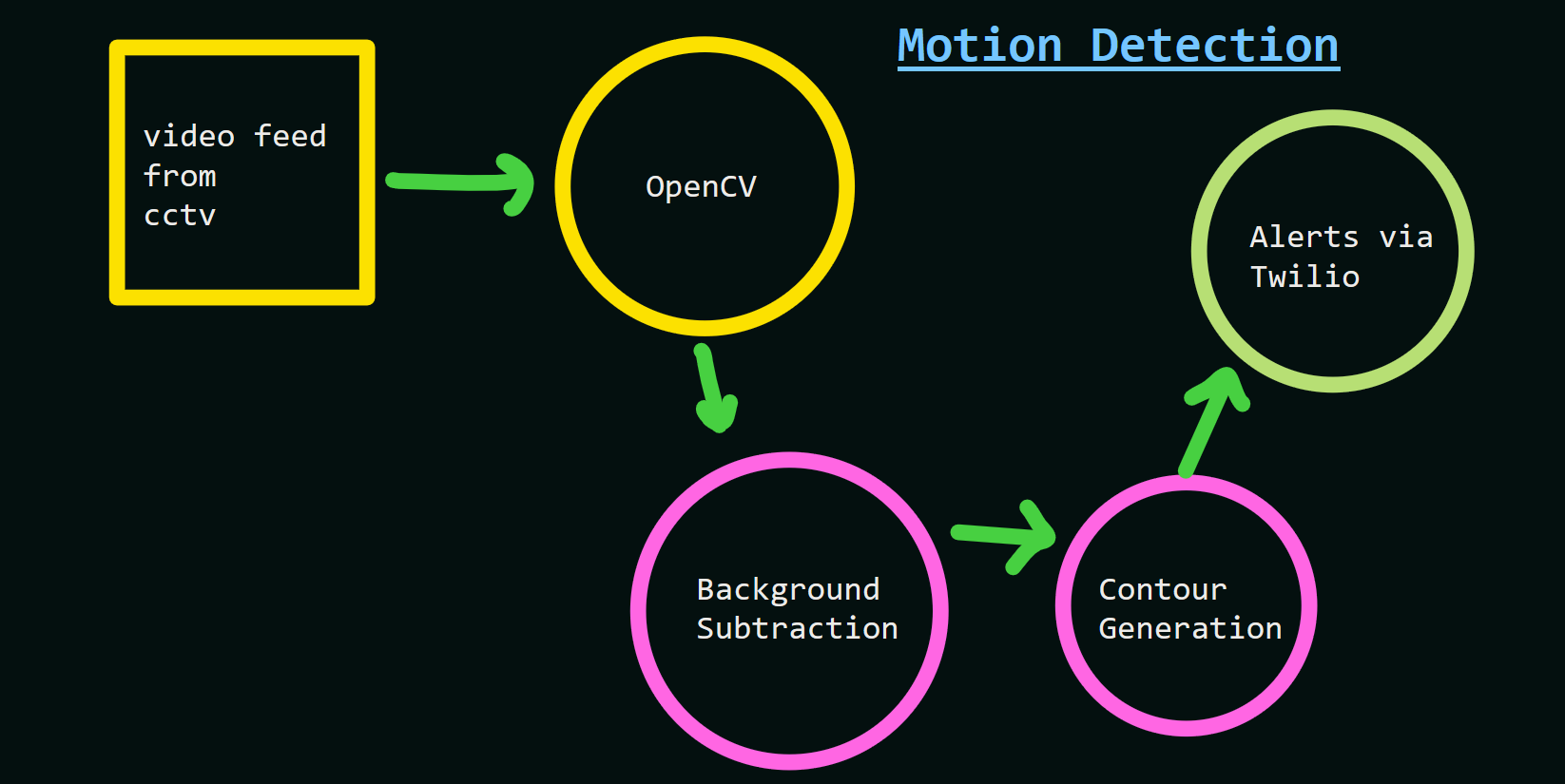
## Appendices

**6.1 Technical Documentation:** A detailed technical document that describes the system architecture, algorithms, data flow, and system components.

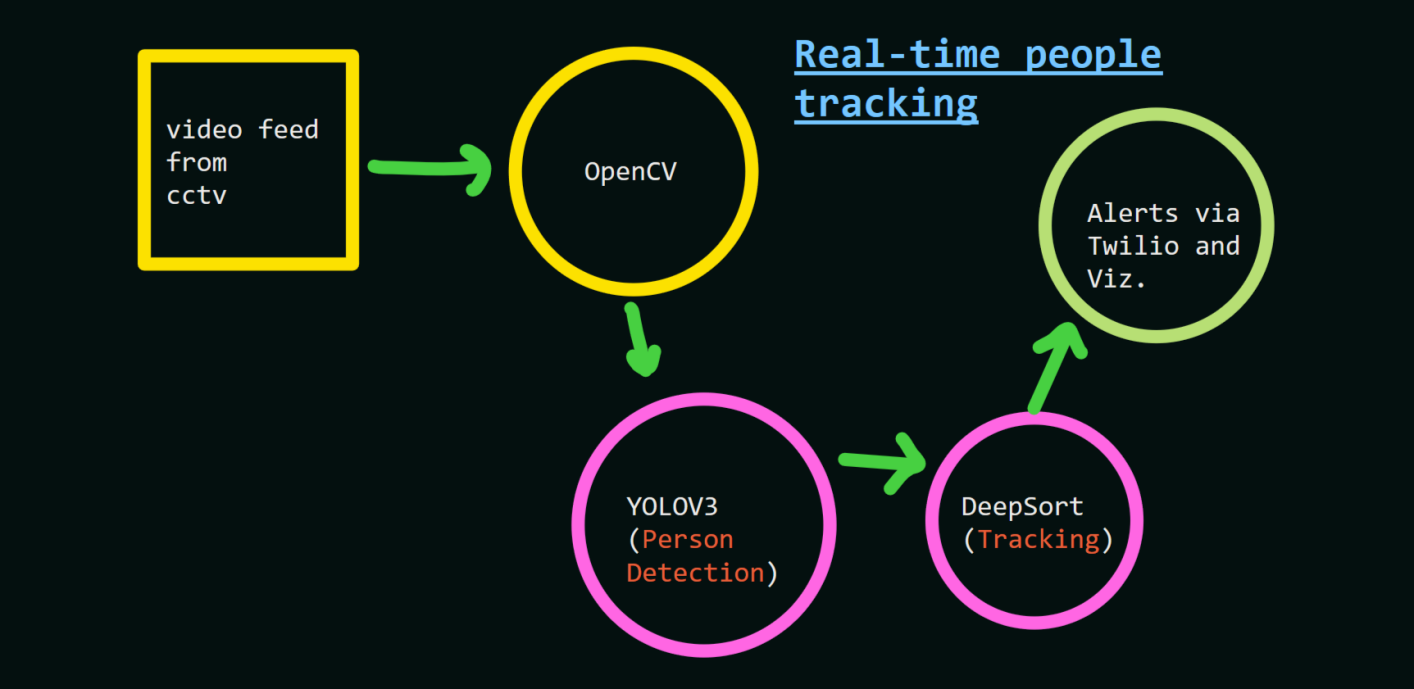
* **Level 0 DFD:** The system has a single input, which is the video feed from the surveillance camera.The system has two main outputs: motion detection alerts and real-time people tracking.



* **Level 1 DFDs:** 
  + **Motion detection alerts:** The video input is analyzed using OpenCV and motion detection algorithms. If motion is detected, the system sends an alert via Twilio to a designated recipient.



* + **Real-time people tracking:** The video input is analyzed using YOLOv3 and machine learning algorithms. If people are detected, their movements are tracked and visualized using Streamlit and alerts are generated accordingly.



**6.2 Source Code:** A copy of the source code used to develop the Smart Surveillance System.

camfeed.py

# Import the required libraries

import numpy as np

import cv2

import time

import datetime

from collections import deque

# Set Window normal so we can resize it

# cv2.namedWindow('frame', cv2.WINDOW\_NORMAL)

# Note the starting time

start\_time = time.time()

# Initialize these variables for calculating FPS

fps = 0

frame\_counter = 0

# Read the video stream from the camera

# cap = cv2.VideoCapture('http://192.168.18.4:8080/video')

# cap = cv2.VideoCapture('https://10.137.131.218:8080/video')

cap= cv2.VideoCapture(0)

while(True):

    ret, frame = cap.read()

    if not ret:

        break

    # Calculate the Average FPS

    frame\_counter += 1

    fps = (frame\_counter / (time.time() - start\_time))

    # Display the FPS

    cv2.putText(frame, 'FPS: {:.2f}'.format(fps), (20, 20), cv2.FONT\_HERSHEY\_SIMPLEX, 0.6, (0, 0, 255),1)

    # Show the Frame

    cv2.imshow('frame',frame)

    # Exit if q is pressed.

    if cv2.waitKey(1) == ord('q'):

        break

# Release Capture and destroy windows

cap.release()

cv2.destroyAllWindows()

countourdetection.py

import cv2

import numpy as np

# initlize video capture object

# cap = cv2.VideoCapture('sample\_video.mp4')

cap = cv2.VideoCapture(0)

# cap = cv2.VideoCapture('http://192.168.137.114:8080/video')

#

# cap = cv2.VideoCapture('https://10.137.131.218:8080/video')

width = 1024

height = 720

# you can set custom kernel size if you want

kernel = None

# initilize background subtractor object

foog = cv2.createBackgroundSubtractorMOG2(

    detectShadows=True, varThreshold= 50, history=200)

# Noise filter threshold

# thresh = 1100

thresh = 1100

while(1):

    ret, frame = cap.read()

    if not ret:

        break

    dim = (width, height)

    frame = cv2.resize(frame, dim, interpolation=cv2.INTER\_AREA)

    # Apply background subtraction

    fgmask = foog.apply(frame)

    # Get rid of the shadows

    ret, fgmask = cv2.threshold(fgmask, 250, 255, cv2.THRESH\_BINARY)

    # Apply some morphological operations to make sure you have a good mask

    # fgmask = cv2.erode(fgmask,kernel,iterations = 1)

    fgmask = cv2.dilate(fgmask, kernel, iterations=4)

    # Detect contours in the frame

    contours, hierarchy = cv2.findContours(

        fgmask, cv2.RETR\_EXTERNAL, cv2.CHAIN\_APPROX\_SIMPLE)

    if contours:

        # Get the maximum contour

        cnt = max(contours, key=cv2.contourArea)

        # print(cnt)

        # make sure the contour area is somewhat hihger than some threshold to make sure its a person and not some noise.

        if cv2.contourArea(cnt) > thresh:

            # Draw a bounding box around the person and label it as person detected

            x, y, w, h = cv2.boundingRect(cnt)

            cv2.rectangle(frame, (x, y), (x+w, y+h), (0, 0, 255), 2)

            cv2.putText(frame, 'Person Detected', (x, y-10),

                        cv2.FONT\_HERSHEY\_SIMPLEX, 0.3, (0, 255, 0), 1, cv2.LINE\_AA)

    # Stack both frames and show the image

    fgmask\_3 = cv2.cvtColor(fgmask, cv2.COLOR\_GRAY2BGR)

    stacked = np.hstack((fgmask\_3, frame))

    cv2.imshow('Combined', cv2.resize(stacked, None, fx=0.65, fy=0.65))

    k = cv2.waitKey(40) & 0xff

    if k == ord('q'):

        break

cap.release()

cv2.destroyAllWindows()

multiplecamfeeds.py

import cv2

import numpy as np

# initlize video capture object

# capture video from webcam

cap1 = cv2.VideoCapture(0)

# capture video from file

cap2 = cv2.VideoCapture('https://10.143.38.102:8080/video')

cap3 = cv2.VideoCapture('https://192.168.137.66:8080/video')

# cap2 = cv2.VideoCapture(0)

# cap3 = cv2.VideoCapture(0)

# you can set custom kernel size if you want

kernel = None

# initilize background subtractor object

# foog = cv2.createBackgroundSubtractorMOG2(

#     detectShadows=True, varThreshold=50, history=500)

foog = cv2.createBackgroundSubtractorMOG2(

    detectShadows=True, varThreshold=50, history=350)

# Noise filter threshold

# thresh = 1100

thresh = 1100

while(1):

    # read frames from both sources

    ret1, frame1 = cap1.read()

    ret2, frame2 = cap2.read()

    ret3, frame3 = cap3.read()

    dim = (480, 720)

    frame1 = cv2.resize(frame1, dim, interpolation=cv2.INTER\_AREA)

    frame2 = cv2.resize(frame2, dim, interpolation=cv2.INTER\_AREA)

    frame3 = cv2.resize(frame3, dim, interpolation=cv2.INTER\_AREA)

    # Apply background subtraction

    fgmask\_f1 = foog.apply(frame1)

    fgmask\_f2 = foog.apply(frame2)

    fgmask\_f3 = foog.apply(frame3)

    # Get rid of the shadows

    ret, fgmask\_f1 = cv2.threshold(fgmask\_f1, 250, 255, cv2.THRESH\_BINARY)

    ret, fgmask\_f2 = cv2.threshold(fgmask\_f2, 250, 255, cv2.THRESH\_BINARY)

    ret, fgmask\_f3 = cv2.threshold(fgmask\_f3, 250, 255, cv2.THRESH\_BINARY)

    # Apply some morphological operations to make sure you have a good mask

    # fgmask = cv2.erode(fgmask,kernel,iterations = 1)

    fgmask\_f1 = cv2.dilate(fgmask\_f1, kernel, iterations=4)

    fgmask\_f2 = cv2.dilate(fgmask\_f2, kernel, iterations=4)

    fgmask\_f3 = cv2.dilate(fgmask\_f3, kernel, iterations=4)

    # Detect contours in the frame

    contours\_f1, hierarchy\_f1 = cv2.findContours(

        fgmask\_f1, cv2.RETR\_EXTERNAL, cv2.CHAIN\_APPROX\_SIMPLE)

    contours\_f2, hierarchy\_f2 = cv2.findContours(

        fgmask\_f2, cv2.RETR\_EXTERNAL, cv2.CHAIN\_APPROX\_SIMPLE)

    contours\_f3, hierarchy\_f3 = cv2.findContours(

        fgmask\_f3, cv2.RETR\_EXTERNAL, cv2.CHAIN\_APPROX\_SIMPLE)

    if contours\_f1:

        # Get the maximum contour

        cnt = max(contours\_f1, key=cv2.contourArea)

        # print(cnt)

        # make sure the contour area is somewhat hihger than some threshold to make sure its a person and not some noise.

        if cv2.contourArea(cnt) > thresh:

            # Draw a bounding box around the person and label it as person detected

            x, y, w, h = cv2.boundingRect(cnt)

            cv2.rectangle(frame1, (x, y), (x+w, y+h), (0, 0, 255), 2)

            cv2.putText(frame1, 'Person Detected', (x, y-10),

                        cv2.FONT\_HERSHEY\_SIMPLEX, 0.3, (0, 255, 0), 1, cv2.LINE\_AA)

    if contours\_f2:

        # Get the maximum contour

        cnt = max(contours\_f2, key=cv2.contourArea)

        # print(cnt)

        # make sure the contour area is somewhat hihger than some threshold to make sure its a person and not some noise.

        if cv2.contourArea(cnt) > thresh:

            # Draw a bounding box around the person and label it as person detected

            x, y, w, h = cv2.boundingRect(cnt)

            cv2.rectangle(frame2, (x, y), (x+w, y+h), (0, 0, 255), 2)

            cv2.putText(frame2, 'Person Detected', (x, y-10),

                        cv2.FONT\_HERSHEY\_SIMPLEX, 0.3, (0, 255, 0), 1, cv2.LINE\_AA)

    if contours\_f3:

        # Get the maximum contour

        cnt = max(contours\_f3, key=cv2.contourArea)

        # print(cnt)

        # make sure the contour area is somewhat hihger than some threshold to make sure its a person and not some noise.

        if cv2.contourArea(cnt) > thresh:

            # Draw a bounding box around the person and label it as person detected

            x, y, w, h = cv2.boundingRect(cnt)

            cv2.rectangle(frame3, (x, y), (x+w, y+h), (0, 0, 255), 2)

            cv2.putText(frame3, 'Person Detected', (x, y-10),

                        cv2.FONT\_HERSHEY\_SIMPLEX, 0.3, (0, 255, 0), 1, cv2.LINE\_AA)

    # Stack both frames and show the image

    fgmask\_3\_f1 = cv2.cvtColor(fgmask\_f1, cv2.COLOR\_GRAY2BGR)

    fgmask\_3\_f2 = cv2.cvtColor(fgmask\_f2, cv2.COLOR\_GRAY2BGR)

    fgmask\_3\_f3 = cv2.cvtColor(fgmask\_f3, cv2.COLOR\_GRAY2BGR)

    stacked\_frame = np.hstack((frame1, frame2, frame3))

    stacked\_countours = np.hstack((fgmask\_3\_f1, fgmask\_3\_f2, fgmask\_3\_f3))

    stacked= np.vstack((stacked\_frame, stacked\_countours))

    cv2.imshow('Combined', cv2.resize(stacked, None, fx=0.90, fy=0.50))

    k = cv2.waitKey(40) & 0xff

    if k == ord('q'):

        break

# release video capture objects and close windows

cap1.release()

cap2.release()

cap3.release()

cv2.destroyAllWindows()

IntruderDetector.py

import cv2

import imutils

import numpy as np

import argparse

# model

HOGCV = cv2.HOGDescriptor()

HOGCV.setSVMDetector(cv2.HOGDescriptor\_getDefaultPeopleDetector())

def detect(frame):

    bounding\_box\_cordinates, weights = HOGCV.detectMultiScale(

        frame, winStride=(4, 4), padding=(8, 8), scale=1.03)

    person = 1

    for x, y, w, h in bounding\_box\_cordinates:

        cv2.rectangle(frame, (x, y), (x+w, y+h), (0, 255, 0), 2)

        cv2.putText(frame, f'person {person}', (x, y),

                    cv2.FONT\_HERSHEY\_SIMPLEX, 0.5, (0, 0, 255), 1)

        person += 1

    cv2.putText(frame, 'Status : Detecting ', (40, 40),

                cv2.FONT\_HERSHEY\_DUPLEX, 0.8, (255, 0, 0), 2)

    cv2.putText(frame, f'Total Persons : {person-1}',

                (40, 70), cv2.FONT\_HERSHEY\_DUPLEX, 0.8, (255, 0, 0), 2)

    cv2.imshow('output', frame)

    return frame

# to detect human

def humanDetector(args):

    image\_path = args["image"]

    video\_path = args['video']

    if str(args["camera"]) == 'true':

        camera = True

    else:

        camera = False

    writer = None

    if args['output'] is not None and image\_path is None:

        writer = cv2.VideoWriter(

            args['output'], cv2.VideoWriter\_fourcc(\*'MJPG'), 10, (600, 600))

    print('[INFO] Opening Web Cam.')

    detectByCamera('outputs-cv/feed.mp4', writer)

def detectByCamera(path, writer):

    video = cv2.VideoCapture(0)

    print('Detecting people...')

    while True:

        check, frame = video.read()

        frame = detect(frame)

        if writer is not None:

            writer.write(frame)

        key = cv2.waitKey(1)

        if key == ord('q'):

            break

    video.release()

    cv2.destroyAllWindows()

def argsParser():

    arg\_parse = argparse.ArgumentParser()

    arg\_parse.add\_argument("-v", "--video", default=None, help="path to Video File ")

    arg\_parse.add\_argument("-i", "--image", default=None, help="path to Image File ")

    arg\_parse.add\_argument("-c", "--camera", default=False, help="Set true if you want to use the camera.")

    arg\_parse.add\_argument("-o", "--output", type=str, help="path to optional output video file")

    args = vars(arg\_parse.parse\_args())

    return args

if \_\_name\_\_ == "\_\_main\_\_":

    HOGCV = cv2.HOGDescriptor()

    HOGCV.setSVMDetector(cv2.HOGDescriptor\_getDefaultPeopleDetector())

    args = argsParser()

    humanDetector(args)

PersonCounter.py

from imutils.video import VideoStream

from imutils.video import FPS

import argparse

import imutils

import time

import cv2

from datetime import datetime, time

import numpy as np

import time as time2

ap = argparse.ArgumentParser()

ap.add\_argument("-v", "--video", help="path to the video file")

ap.add\_argument("-a", "--min-area", type=int, default=500, help="minimum area size")

ap.add\_argument("-t", "--tracker", type=str, default="csrt", help="OpenCV object tracker type")

args = vars(ap.parse\_args())

# extract the OpenCV version info

(major, minor) = cv2.\_\_version\_\_.split(".")[:2]

# if we are using OpenCV 3.2 or an earlier version, we can use a special factory

# function to create the entity that tracks objects

if int(major) == 3 and int(minor) < 3:

    tracker = cv2.Tracker\_create(args["tracker"].upper())

    #tracker = cv2.TrackerGOTURN\_create()

# otherwise, for OpenCV 3.3 or newer,

# we need to explicity call the respective constructor that contains the tracker object:

else:

    # initialize a dictionary that maps strings to their corresponding

    # OpenCV object tracker implementations

    OPENCV\_OBJECT\_TRACKERS = {

        "csrt": cv2.TrackerCSRT\_create,

        "kcf": cv2.TrackerKCF\_create,

        "boosting": cv2.legacy.TrackerBoosting\_create,

        "mil": cv2.TrackerMIL\_create,

        "tld": cv2.legacy.TrackerTLD\_create,

        "medianflow": cv2.legacy.TrackerMedianFlow\_create,

        "mosse": cv2.legacy.TrackerMOSSE\_create

    }

# grab the appropriate object tracker using our dictionary of

# OpenCV object tracker objects

    tracker = OPENCV\_OBJECT\_TRACKERS[args["tracker"]]()

    #tracker = cv2.TrackerGOTURN\_create()

# if the video argument is None, then the code will read from webcam (work in progress)

if args.get("video", None) is None:

    vs = VideoStream(src=0).start()

    time2.sleep(2.0)

# otherwise, we are reading from a video file

else:

    vs = cv2.VideoCapture(args["video"])

# loop over the frames of the video, and store corresponding information from each frame

firstFrame = None

initBB2 = None

fps = None

differ = None

now = ''

framecounter = 0

trackeron = 0

while True:

    frame = vs.read()

    frame = frame if args.get("video", None) is None else frame[1]

    # if the frame can not be grabbed, then we have reached the end of the video

    if frame is None:

            break

    # resize the frame to 500

    frame = imutils.resize(frame, width=500)

    framecounter = framecounter+1

    if framecounter > 1:

        (H, W) = frame.shape[:2]

        gray = cv2.cvtColor(frame, cv2.COLOR\_BGR2GRAY)

        gray = cv2.GaussianBlur(gray, (21, 21), 0)

        # if the first frame is None, initialize it

        if firstFrame is None:

            firstFrame = gray

            continue

        # compute the absolute difference between the current frame and first frame

        frameDelta = cv2.absdiff(firstFrame, gray)

        thresh = cv2.threshold(frameDelta, 25, 255, cv2.THRESH\_BINARY)[1]

        # dilate the thresholded image to fill in holes, then find contours on thresholded image

        thresh = cv2.dilate(thresh, None, iterations=2)

        # cnts = cv2.findContours(thresh.copy(), cv2.RETR\_EXTERNAL,cv2.CHAIN\_APPROX\_SIMPLE)

        # cnts = cnts[0] if imutils.is\_cv2() else cnts[1]

        contours, heirarchy = cv2.findContours(thresh.copy(), cv2.RETR\_EXTERNAL,cv2.CHAIN\_APPROX\_SIMPLE)

        cnts = max(contours, key=cv2.contourArea)

        # loop over the contours identified

        contourcount = 0

        for c in cnts:

            contourcount =  contourcount + 1

           # if the contour is too small, ignore it

            if cv2.contourArea(c) < args["min\_area"]:

                continue

            # compute the bounding box for the contour, draw it on the frame,

            (x, y, w, h) = cv2.boundingRect(c)

            initBB2 =(x,y,w,h)

            prott1 = r'MobileNetSSD\_deploy.prototxt'

            prott2 = r'mobilenet\_iter\_73000.caffemodel'

            net = cv2.dnn.readNetFromCaffe(prott1, prott2)

            CLASSES = ["person"]

            COLORS = np.random.uniform(0, 255, size=(len(CLASSES), 3))

            trackbox = frame[y:y+h, x:x+w]

            trackbox = cv2.resize(trackbox, (224, 224))

            cv2.imshow('image',trackbox)

            blob = cv2.dnn.blobFromImage(cv2.resize(trackbox, (300, 300)),0.007843, (300, 300), 127.5)

            net.setInput(blob)

            detections = net.forward()

            for i in np.arange(0, detections.shape[2]):

                confidence = detections[0, 0, i, 2]

                confidence\_level = 0.7

                if confidence > confidence\_level:

                    # extract the index of the class label from the `detections`, then compute the (x, y)-coordinates of

                    # the bounding box for the object

                    idx = int(detections[0, 0, i, 1])

                    box = detections[0, 0, i, 3:7] \* np.array([w, h, w, h])

                    (startX, startY, endX, endY) = box.astype("int")

                    # draw the prediction on the frame

                    label = "{}: {:.2f}%".format(CLASSES[idx],

                                                 confidence \* 100)

                    cv2.rectangle(frame, (startX, startY), (endX, endY),

                                  COLORS[idx], 2)

                    y = startY - 15 if startY - 15 > 15 else startY + 15

                    cv2.putText(frame, label, (startX, y),

                                cv2.FONT\_HERSHEY\_SIMPLEX, 0.5, COLORS[idx], 2)

            cv2.rectangle(frame, (x, y), (x + w, y + h), (255, 255, 0), 2)

            # Start tracker

            now = datetime.now()

            if differ == None or differ > 9:

                tracker.init(frame, initBB2)

                fps = FPS().start()

    # check to see if we are currently tracking an object, if so, ignore other boxes

    # this code is relevant if we want to identify particular persons (section 2 of this tutorial)

    if initBB2 is not None:

        # grab the new bounding box coordinates of the object

        (success, box) = tracker.update(frame)

        # check to see if the tracking was a success

        differ = 10

        if success:

            (x, y, w, h) = [int(v) for v in box]

            cv2.rectangle(frame, (x, y), (x + w, y + h), (0, 255, 0), 2)

            differ = abs(initBB2[0]-box[0]) + abs(initBB2[1]-box[1])

            i = tracker.update(lastframe)

            if i[0] != True:

                time2.sleep(4000)

        else:

            trackeron = 1

        # update the FPS counter

        fps.update()

        fps.stop()

        # initialize the set of information we'll be displaying on

        # the frame

        info = [

            ("Success", "Yes" if success else "No"),

            ("FPS", "{:.2f}".format(fps.fps())),

        ]

        # loop over the info tuples and draw them on our frame

        for (i, (k, v)) in enumerate(info):

            text = "{}: {}".format(k, v)

            cv2.putText(frame, text, (10, H - ((i \* 20) + 20)),

                        cv2.FONT\_HERSHEY\_SIMPLEX, 0.6, (0, 0, 255), 2)

        # draw the text and timestamp on the frame

        now2 = datetime.now()

        time\_passed\_seconds = str((now2-now).seconds)

        cv2.putText(frame, 'Detecting persons', (10, 20),

                    cv2.FONT\_HERSHEY\_SIMPLEX, 0.5, (0, 0, 255), 2)

    # show the frame and record if the user presses a key

    cv2.imshow("Video stream", frame)

    key = cv2.waitKey(1) & 0xFF

    # if the `q` key is pressed, break from the lop

    if key == ord("q"):

        break

    if key == ord("d"):

        firstFrame = None

    lastframe = frame

# finally, stop the camera/stream and close any open windows

vs.stop() if args.get("video", None) is None else vs.release()

cv2.destroyAllWindows()

Twilio\_api.py

from twilio.rest import Client

# Lucky@1234567891011

# deepankarsharma2003@gmail.com

# Your Account SID from twilio.com/console

# account\_sid = "AC0436791453c88f23bb818240cbd471a2"

# Your Auth Token from twilio.com/console

# auth\_token  = "33dfa5eaaf0c90be139db142df619323"

# Read text from the credentials file and store in data variable

with open('credentials.txt', 'r') as myfile:

  data = myfile.read()

# Convert data variable into dictionary

info\_dict = eval(data)

account\_sid = 'AC0436791453c88f23bb818240cbd471a2'

# auth\_token = '[Redacted]'

# Your Auth Token from twilio.com/console

auth\_token = info\_dict['auth\_token']

client = Client(account\_sid, auth\_token)

message = client.messages.create(

    from\_='whatsapp:+14155238886',

    body='Bade achhe lagte hai\n ye dharti\nye nadiya\nye raina\naur tum...',

    to='whatsapp:+919639102301'

)

print(message.sid)

BackgroundRemoval.py

import cv2

import numpy as np

# cap = cv2.VideoCapture('sample\_video.mp4')

cap = cv2.VideoCapture(0)

# cap = cv2.VideoCapture('https://10.137.131.218:8080/video')

#

#

# Create the background subtractor object

foog = cv2.createBackgroundSubtractorMOG2(

    detectShadows=False, varThreshold=40, history=150)

# history ----> jitni jyada history , utna slow motions k liye robust hoga

while(1):

    ret, frame = cap.read()

    if not ret:

        break

    # Apply the background object on each frame

    fgmask = foog.apply(frame)

    # Get rid of the shadows

    ret, fgmask = cv2.threshold(fgmask, 250, 255, cv2.THRESH\_BINARY)

    # Show the background subtraction frame.

    # cv2.imshow('All three', fgmask) # original

    fgmask = cv2.cvtColor(fgmask, cv2.COLOR\_GRAY2BGR)

    cv2.imshow('Stacked frame', np.hstack((frame, fgmask)))

    k = cv2.waitKey(10)

    if k == ord('q'):

        break

cap.release()

cv2.destroyAllWindows()

CompleteSystem.py

# Import the required libraries

import numpy as np

import cv2

import time

import datetime

from collections import deque

from twilio.rest import Client

def is\_person\_present(frame, thresh=1100):

    global foog

    # Apply background subtraction

    fgmask = foog.apply(frame)

    # Get rid of the shadows

    ret, fgmask = cv2.threshold(fgmask, 250, 255, cv2.THRESH\_BINARY)

    # Apply some morphological operations to make sure you have a good mask

    fgmask = cv2.dilate(fgmask, kernel=None, iterations=4)

    # Detect contours in the frame

    contours, hierarchy = cv2.findContours(

        fgmask, cv2.RETR\_EXTERNAL, cv2.CHAIN\_APPROX\_SIMPLE)

    # Check if there was a contour and the area is somewhat higher than some threshold so we know its a person and not noise

    if contours and cv2.contourArea(max(contours, key=cv2.contourArea)) > thresh:

        # Get the max contour

        cnt = max(contours, key=cv2.contourArea)

        # Draw a bounding box around the person and label it as person detected

        # x, y, w, h = cv2.boundingRect(cnt)

        # cv2.rectangle(frame, (x, y), (x+w, y+h), (0, 0, 255), 2)

        # cv2.putText(frame, 'Person Detected', (x, y-10),

        #             cv2.FONT\_HERSHEY\_SIMPLEX, 0.3, (0, 255, 0), 1, cv2.LINE\_AA)

        return True, frame

    # Otherwise report there was no one present

    else:

        return False, frame

def send\_message(body, info\_dict):

    # Your Account SID from twilio.com/console

    account\_sid = 'AC0436791453c88f23bb818240cbd471a2'

    # Your Auth Token from twilio.com/console

    auth\_token = info\_dict['auth\_token']

    client = Client(account\_sid, auth\_token)

    message = client.messages.create(

        from\_='whatsapp:+14155238886',

        body='Alert, Ghar m chor hai !!!!!',

        to='whatsapp:+919639102301'

    )

    print(message)

#time.sleep(15)

# Set Window normal so we can resize it

cv2.namedWindow('frame', cv2.WINDOW\_AUTOSIZE)

# This is a test video

# cap = cv2.VideoCapture('sample\_video.mp4')

ip1 = input('Enter the ip of the cam1: ')

ip1 = 'https://' + ip1 + ':8080/video'

print('ip1: ', ip1)

ip2 = input('Enter the ip of the cam2: ')

ip2 = 'https://' + ip2 + ':8080/video'

print('ip2: ', ip2)

cap = cv2.VideoCapture(ip1)

cap2 = cv2.VideoCapture(ip2)

# Read the video steram from the camera

#cap = cv2.VideoCapture('http://192.168.18.4:8080/video')

# Get width and height of the frame

# width = int(cap.get(3))

# height = int(cap.get(4))

width = 720

height = 480

# Read and store the credentials information in a dict

with open('credentials.txt', 'r') as myfile:

  data = myfile.read()

info\_dict = eval(data)

# Initialize the background Subtractor

foog = cv2.createBackgroundSubtractorMOG2(

    detectShadows=True, varThreshold=100, history=2000)

# Status is True when person is present and False when the person is not present.

status = False

status2 = False

# After the person disapears from view, wait atleast 7 seconds before making the status False

patience = 7

patience2 = 7

# We don't consider an initial detection unless its detected 15 times, this gets rid of false positives

detection\_thresh = 15

# Initial time for calculating if patience time is up

initial\_time = None

initial\_time2 = None

# We are creating a deque object of length detection\_thresh and will store individual detection statuses here

de = deque([False] \* detection\_thresh, maxlen=detection\_thresh)

de2 = deque([False] \* detection\_thresh, maxlen=detection\_thresh)

# Initialize these variables for calculating FPS

fps = 0

fps2 = 0

frame\_counter = 0

frame\_counter2 = 0

start\_time = time.time()

start\_time2 = time.time()

while(True):

    ret, frame = cap.read()

    ret2, frame2 = cap2.read()

    if not ret or not ret2:

        break

    # This function will return a boolean variable telling if someone was present or not, it will also draw boxes if it

    # finds someone

    dim = (width, height)

    frame = cv2.resize(frame, dim, interpolation=cv2.INTER\_AREA)

    frame2 = cv2.resize(frame2, dim, interpolation=cv2.INTER\_AREA)

    # frame= np.hstack((frame, frame2))

    detected, annotated\_image = is\_person\_present(frame)

    detected2, annotated\_image2 = is\_person\_present(frame2)

    # Register the current detection status on our deque object

    de.appendleft(detected)

    de2.appendleft(detected2)

    # If we have consectutively detected a person 15 times then we are sure that soemone is present

    # We also make this is the first time that this person has been detected so we only initialize the videowriter once

    if sum(de) == detection\_thresh and not status:

        status = True

        entry\_time = datetime.datetime.now().strftime("%A, %I-%M-%S %p %d %B %Y")

        # out = cv2.VideoWriter('outputs/{}.mp4'.format(entry\_time),

        #                       cv2.VideoWriter\_fourcc(\*'XVID'), 15.0, (width, height))

    if sum(de2) == detection\_thresh and not status2:

        status2 = True

        entry\_time2 = datetime.datetime.now().strftime("%A, %I-%M-%S %p %d %B %Y")

        # out = cv2.VideoWriter('outputs/{}.mp4'.format(entry\_time),

        #                       cv2.VideoWriter\_fourcc(\*'XVID'), 15.0, (width, height))

    # If status is True but the person is not in the current frame

    if status and not detected:

        # Restart the patience timer only if the person has not been detected for a few frames so we are sure it was'nt a

        # False positive

        if sum(de) > (detection\_thresh/2):

            if initial\_time is None:

                initial\_time = time.time()

        elif initial\_time is not None:

            # If the patience has run out and the person is still not detected then set the status to False

            # Also save the video by releasing the video writer and send a text message.

            if time.time() - initial\_time >= patience:

                status = False

                exit\_time = datetime.datetime.now().strftime("%A, %I:%M:%S %p %d %B %Y")

                # out.release()

                initial\_time = None

                body = "Alert: n A Person Entered the Room at {} n Left the room at {}".format(

                    entry\_time, exit\_time)

                print(body)

                send\_message(body, info\_dict)

    # If significant amount of detections (more than half of detection\_thresh) has occured then we reset the Initial Time.

    elif status and sum(de) > (detection\_thresh/2):

        initial\_time = None

    # Get the current time in the required format

    current\_time = datetime.datetime.now().strftime("%A, %I:%M:%S %p %d %B %Y")

    # Display the FPS

    cv2.putText(annotated\_image, 'FPS: {:.2f}'.format(

        fps), (510, 450), cv2.FONT\_HERSHEY\_COMPLEX, 0.6, (255, 40, 155), 2)

    # Display Time

    cv2.putText(annotated\_image, current\_time, (310, 20),

                cv2.FONT\_HERSHEY\_COMPLEX, 0.5, (0, 0, 255), 1)

    # Display the Room Status

    cv2.putText(annotated\_image, 'Room Occupied: {}'.format(str(status)), (10, 20), cv2.FONT\_HERSHEY\_SIMPLEX, 0.6,

                (200, 10, 150), 2)

    # Show the patience Value

    if initial\_time is None:

        text = 'Patience: {}'.format(patience)

    else:

        text = 'Patience: {:.2f}'.format(

            max(0, patience - (time.time() - initial\_time)))

    cv2.putText(annotated\_image, text, (10, 450),

                cv2.FONT\_HERSHEY\_COMPLEX, 0.6, (255, 40, 155), 2)

    # If status is true save the frame

    # if status:

    #     out.write(annotated\_image)

    # If status is True but the person is not in the current frame

    if status2 and not detected2:

        # Restart the patience timer only if the person has not been detected for a few frames so we are sure it was'nt a

        # False positive

        if sum(de2) > (detection\_thresh/2):

            if initial\_time2 is None:

                initial\_time2 = time.time()

        elif initial\_time2 is not None:

            # If the patience has run out and the person is still not detected then set the status to False

            # Also save the video by releasing the video writer and send a text message.

            if time.time() - initial\_time2 >= patience2:

                status2 = False

                exit\_time2 = datetime.datetime.now().strftime("%A, %I:%M:%S %p %d %B %Y")

                # out.release()

                initial\_time2 = None

                body2 = "Alert: n A Person Entered the Room at {} n Left the room2 at {}".format(

                    entry\_time2, exit\_time2)

                print(body2)

                send\_message(body2, info\_dict)

    # If significant amount of detections (more than half of detection\_thresh) has occured then we reset the Initial Time.

    elif status2 and sum(de2) > (detection\_thresh/2):

        initial\_time2 = None

    # Get the current time in the required format

    current\_time2 = datetime.datetime.now().strftime("%A, %I:%M:%S %p %d %B %Y")

    # Display the FPS

    cv2.putText(annotated\_image2, 'FPS: {:.2f}'.format(

        fps2), (510, 450), cv2.FONT\_HERSHEY\_COMPLEX, 0.6, (255, 40, 155), 2)

    # Display Time

    cv2.putText(annotated\_image2, current\_time2, (310, 20),

                cv2.FONT\_HERSHEY\_COMPLEX, 0.5, (0, 0, 255), 1)

    # Display the Room Status

    cv2.putText(annotated\_image2, 'Room Occupied: {}'.format(str(status2)), (10, 20), cv2.FONT\_HERSHEY\_SIMPLEX, 0.6,

                (200, 10, 150), 2)

    # Show the patience Value

    if initial\_time2 is None:

        text = 'Patience: {}'.format(patience2)

    else:

        text = 'Patience: {:.2f}'.format(

            max(0, patience2 - (time.time() - initial\_time2)))

    cv2.putText(annotated\_image2, text, (10, 450),

                cv2.FONT\_HERSHEY\_COMPLEX, 0.6, (255, 40, 155), 2)

    # If status is true save the frame

    # if status:

    #     out.write(annotated\_image)

    frame= np.hstack((frame, frame2))

    # Show the Frame

    cv2.imshow('frame', frame)

    # Calculate the Average FPS

    frame\_counter += 1

    fps = (frame\_counter / (time.time() - start\_time))

    frame\_counter2 += 1

    fps2 = (frame\_counter2 / (time.time() - start\_time2))

    # Exit if q is pressed.

    if cv2.waitKey(30) == ord('q'):

        break

# Release Capture and destroy windows

cap.release()

cap2.release()

cv2.destroyAllWindows()

# out.release()

Using\_mobilenet.py

import cv2

import numpy as np

# Initialize the HOG descriptor/person detector

hog = cv2.HOGDescriptor()

hog.setSVMDetector(cv2.HOGDescriptor\_getDefaultPeopleDetector())

width = 1024

height = 720

img = cv2.imread('firstframe.jpg')

dim = (width, height)

img= cv2.resize(img, dim, interpolation=cv2.INTER\_AREA)

# firstFrame= img.copy()

gray = cv2.cvtColor(img, cv2.COLOR\_BGR2GRAY)

firstFrame = cv2.GaussianBlur(gray, (21, 21), 0)

# initlize video capture object

# cap = cv2.VideoCapture('sample\_video.mp4')

# cap = cv2.VideoCapture(0)

cap = cv2.VideoCapture('http://192.168.137.3:8080/video')

# cap = cv2.VideoCapture('https://10.137.131.218:8080/video')

# you can set custom kernel size if you want

kernel = None

# initilize background subtractor object

foog = cv2.createBackgroundSubtractorMOG2(

    detectShadows=False, varThreshold=200, history=500)

# Noise filter threshold

# thresh = 1100

thresh = 3000

while(1):

    ret, frame = cap.read()

    if not ret:

        break

    dim = (width, height)

    frame = cv2.resize(frame, dim, interpolation=cv2.INTER\_AREA)

    gray = cv2.cvtColor(frame, cv2.COLOR\_BGR2GRAY)

    gray = cv2.GaussianBlur(gray, (21, 21), 0)

    # Apply background subtraction

    # fgmask = foog.apply(frame)

    # fgmask = foog.apply(gray)

    # firstFrame = foog.apply(firstFrame)

    # Get rid of the shadows

    # ret, fgmask = cv2.threshold(fgmask, 250, 255, cv2.THRESH\_BINARY)

    # ret, firstFrame = cv2.threshold(firstFrame, 250, 255, cv2.THRESH\_BINARY)

    # Apply some morphological operations to make sure you have a good mask

    # fgmask = cv2.erode(fgmask,kernel,iterations = 1)

    # fgmask = cv2.dilate(fgmask, kernel, iterations=4)

    # fgmask = cv2.dilate(fgmask, kernel, iterations=4)

    # firstFrame = cv2.dilate(firstFrame, kernel, iterations=4)

    # print(frame.shape==img.shape)

    frameDelta = cv2.absdiff(firstFrame, gray)

    thresh = cv2.threshold(frameDelta, 25, 255, cv2.THRESH\_BINARY)[1]

    thresh = cv2.dilate(thresh, None, iterations=2)

    # Detect people in the frame

    boxes, weights = hog.detectMultiScale(thresh, winStride=(8, 8))

    # Draw bounding boxes around the people

    for (x, y, w, h) in boxes:

        cv2.rectangle(frame, (x, y), (x + w, y + h), (0, 0, 255), 2)

    '''

    # Detect contours in the frame

    contours, hierarchy = cv2.findContours(

        thresh, cv2.RETR\_EXTERNAL, cv2.CHAIN\_APPROX\_SIMPLE)

    if contours:

    # for cnt in contours:

        # Get the maximum contour

        cnt = max(contours, key=cv2.contourArea)

        # make sure the contour area is somewhat higher than some threshold to make sure its a person and not some noise.

        # if cv2.contourArea(cnt[0][0]) > thresh:

        if 1:

            print(cnt)

            # Draw a bounding box around the person and label it as person detected

            x, y, w, h = cv2.boundingRect(cnt)

            cv2.rectangle(frame, (x, y), (x+w, y+h), (0, 0, 255), 2)

            cv2.putText(frame, 'Person Detected', (x, y-10),

                        cv2.FONT\_HERSHEY\_SIMPLEX, 0.3, (0, 255, 0), 1, cv2.LINE\_AA)

    '''

    # Stack both frames and show the image

    fgmask\_3 = cv2.cvtColor(thresh, cv2.COLOR\_GRAY2BGR)

    stacked = np.hstack((fgmask\_3, frame))

    cv2.imshow('Combined', cv2.resize(stacked, None, fx=0.65, fy=0.65))

    k = cv2.waitKey(40) & 0xff

    if k == ord('q'):

        break

cap.release()

cv2.destroyAllWindows()

Camfeed with ROI cropping.py

# Import the required libraries

import numpy as np

import cv2

import time

import datetime

from collections import deque

# Set Window normal so we can resize it

# cv2.namedWindow('frame', cv2.WINDOW\_NORMAL)

# Note the starting time

start\_time = time.time()

# Initialize these variables for calculating FPS

fps = 0

frame\_counter = 0

classes = None

with open('coco.names', 'r') as f:

    classes = [line.strip() for line in f.readlines()]

net = cv2.dnn.readNet('yolov3-tiny.weights', 'yolov3-tiny.cfg')

layer\_names = net.getLayerNames()

output\_layers = [layer\_names[i-1] for i in net.getUnconnectedOutLayers()]

ip= input('Enter the ip of the cam: ')

ip = 'https://' + ip + ':8080/video'

print(ip)

# Read the video stream from the camera

# cap = cv2.VideoCapture('http://192.168.46.101:8080/video')

# cap = cv2.VideoCapture('https://192.168.205.234:8080/video')

cap = cv2.VideoCapture(ip)

# cap = cv2.VideoCapture('https://10.133.173.57:8080/video')

# cap = cv2.VideoCapture(0)

skip= 1

while(True):

    if skip==1:

        skip=2

    elif skip == 2:

        skip = 3

        continue

    elif skip==3:

        skip=4

        continue

    elif skip==4:

        skip=5

        continue

    elif skip==5:

        skip=6

        continue

    elif skip==6:

        skip=7

        continue

    else:

        skip=1

        continue

    ret, frame = cap.read()

    if not ret:

        break

    # dim = (1024, 720)

    dim = (720, 480)

    frame = cv2.resize(frame, dim, interpolation=cv2.INTER\_AREA)

    # vertices = np.array(

    #     [[(0, 0), (0, 200), (200, 200), (200, 0)]], dtype=np.int32)

    # vertices = np.array(

    #     [[(50, 50), (50, 50+300), (50+300, 50+300), (50+300, 50)]], dtype=np.int32)

    vertices = np.array(

        [[(250, 50), (250, 50+300), (250+300, 50+300), (250+300, 50)]], dtype=np.int32)

    mask = np.zeros\_like(frame)

    # # cv2.fillPoly(mask, vertices, (255, 255, 255))

    cv2.fillPoly(mask, vertices, (255, 255, 255))  # BGR

    # cv2.imshow('mask', mask)

    masked\_frame= frame.copy()

    frame = cv2.bitwise\_and(frame, mask)

    # frame = cv2.resize(frame, dim, interpolation=cv2.INTER\_AREA)

    frame= cv2.cvtColor(frame, cv2.COLOR\_BGR2RGB)

    frame= cv2.cvtColor(frame, cv2.COLOR\_BGR2RGB)

    # frame= cv2.cvtColor(frame, cv2.COLOR\_BGR2GRAY)

    # frame= cv2.cvtColor(frame, cv2.COLOR\_GRAY2RGB)

    # Calculate the Average FPS

    frame\_counter += 1

    fps = (frame\_counter / (time.time() - start\_time))

    # Display the FPS

    cv2.putText(frame, 'FPS: {:.2f}'.format(

        fps), (20, 20), cv2.FONT\_HERSHEY\_SIMPLEX, 0.6, (0, 0, 255), 1)

    image= frame

    net.setInput(cv2.dnn.blobFromImage(image, 0.00392,

                (416, 416), (0, 0, 0), True, crop=False))

    outs = net.forward(output\_layers)

    class\_ids = []

    confidences = []

    boxes = []

    Width = image.shape[1]

    Height = image.shape[0]

    for out in outs:

        for detection in out:

            scores = detection[5:]

            class\_id = np.argmax(scores)

            confidence = scores[class\_id]

            # if confidence > 0.1:

            if confidence > 0.15:

                center\_x = int(detection[0] \* Width)

                center\_y = int(detection[1] \* Height)

                w = int(detection[2] \* Width)

                h = int(detection[3] \* Height)

                x = center\_x - w / 2

                y = center\_y - h / 2

                class\_ids.append(class\_id)

                confidences.append(float(confidence))

                boxes.append([x, y, w, h])

    indices = cv2.dnn.NMSBoxes(boxes, confidences, 0.1, 0.1)

    #check if is people detection

    count= 0

    for i in indices:

        # i = i[0]

        box = boxes[i]

        # if class\_ids[i] == 0 or class\_ids[i]==56:

        if class\_ids[i] == 0:

            count+=1

            # label = str(classes[class\_id])

            label = str(classes[class\_ids[i]])

            cv2.rectangle(image, (round(box[0]), round(box[1])), (round(

                box[0]+box[2]), round(box[1]+box[3])), (200, 10, 10), 5)

            cv2.putText(image, label, (round(

                box[0])-10, round(box[1])-10), cv2.FONT\_HERSHEY\_SIMPLEX, 0.9, (200, 10, 150), 2)

    print(f'{count} people detected !!!')

    # Show the Frame

    cv2.imshow('frame', image)

    # Exit if q is pressed.

    if cv2.waitKey(1) == ord('q'):

        break

# Release Capture and destroy windows

cap.release()

cv2.destroyAllWindows()

Yolo/camfeed.py

# Import the required libraries

import numpy as np

import cv2

import time

import datetime

from collections import deque

# Set Window normal so we can resize it

# cv2.namedWindow('frame', cv2.WINDOW\_NORMAL)

# Note the starting time

start\_time = time.time()

# Initialize these variables for calculating FPS

fps = 0

frame\_counter = 0

classes = None

with open('coco.names', 'r') as f:

    classes = [line.strip() for line in f.readlines()]

net = cv2.dnn.readNet('yolov3-tiny.weights', 'yolov3-tiny.cfg')

layer\_names = net.getLayerNames()

output\_layers = [layer\_names[i-1] for i in net.getUnconnectedOutLayers()]

# Read the video stream from the camera

# cap = cv2.VideoCapture('http://192.168.46.101:8080/video')

# cap = cv2.VideoCapture('https://10.145.108.7:8080/video')

cap= cv2.VideoCapture(0)

# cap = cv2.VideoCapture('https://10.133.173.57:8080/video')

# cap = cv2.VideoCapture(0)

skip= 1

while(True):

    if skip==1:

        skip=2

    elif skip == 2:

        skip = 3

        continue

    elif skip==3:

        skip=4

        continue

    elif skip==4:

        skip=5

        continue

    elif skip==5:

        skip=6

        continue

    elif skip==6:

        skip=7

        continue

    else:

        skip=1

        continue

    ret, frame = cap.read()

    if not ret:

        break

    dim = (1024, 720)

    dim = (720, 480)

    frame = cv2.resize(frame, dim, interpolation=cv2.INTER\_AREA)

    # vertices = np.array(

    #     [[(0, 0), (0, 200), (200, 200), (200, 0)]], dtype=np.int32)

    # vertices = np.array(

    #     [[(50, 50), (50, 50+300), (50+300, 50+300), (50+300, 50)]], dtype=np.int32)

    vertices = np.array(

        [[(250, 50), (250, 50+300), (250+300, 50+300), (250+300, 50)]], dtype=np.int32)

    mask = np.zeros\_like(frame)

    # # cv2.fillPoly(mask, vertices, (255, 255, 255))

    cv2.fillPoly(mask, vertices, (255, 255, 255))  # BGR

    # cv2.imshow('mask', mask)

    unmasked\_frame= frame.copy()

    frame = cv2.bitwise\_and(frame, mask)

    # frame = cv2.resize(frame, dim, interpolation=cv2.INTER\_AREA)

    frame= cv2.cvtColor(frame, cv2.COLOR\_BGR2RGB)

    frame= cv2.cvtColor(frame, cv2.COLOR\_BGR2RGB)

    # frame= cv2.cvtColor(frame, cv2.COLOR\_BGR2GRAY)

    # frame= cv2.cvtColor(frame, cv2.COLOR\_GRAY2RGB)

    # Calculate the Average FPS

    frame\_counter += 1

    fps = (frame\_counter / (time.time() - start\_time))

    # Display the FPS

    cv2.putText(frame, 'FPS: {:.2f}'.format(

        fps), (20, 20), cv2.FONT\_HERSHEY\_SIMPLEX, 0.6, (0, 0, 255), 1)

    image= frame

    net.setInput(cv2.dnn.blobFromImage(image, 0.00392,

                (416, 416), (0, 0, 0), True, crop=False))

    outs = net.forward(output\_layers)

    class\_ids = []

    confidences = []

    boxes = []

    Width = image.shape[1]

    Height = image.shape[0]

    for out in outs:

        for detection in out:

            scores = detection[5:]

            class\_id = np.argmax(scores)

            confidence = scores[class\_id]

            # if confidence > 0.1:

            if confidence > 0.15:

                center\_x = int(detection[0] \* Width)

                center\_y = int(detection[1] \* Height)

                w = int(detection[2] \* Width)

                h = int(detection[3] \* Height)

                x = center\_x - w / 2

                y = center\_y - h / 2

                class\_ids.append(class\_id)

                confidences.append(float(confidence))

                boxes.append([x, y, w, h])

    indices = cv2.dnn.NMSBoxes(boxes, confidences, 0.1, 0.1)

    #check if is people detection

    count= 0

    for i in indices:

        # i = i[0]

        box = boxes[i]

        # if class\_ids[i] == 0 or class\_ids[i]==56:

        if class\_ids[i] == 0:

            count+=1

            # label = str(classes[class\_id])

            label = str(classes[class\_ids[i]])

            cv2.rectangle(image, (round(box[0]), round(box[1])), (round(

                box[0]+box[2]), round(box[1]+box[3])), (200, 10, 10), 5)

            cv2.putText(image, label, (round(

                box[0])-10, round(box[1])-10), cv2.FONT\_HERSHEY\_SIMPLEX, 0.9, (200, 10, 150), 2)

    print(f'{count} people detected !!!')

    stacked= np.hstack((unmasked\_frame, frame))

    # Show the Frame

    cv2.imshow('stacked', stacked)

    # Exit if q is pressed.

    if cv2.waitKey(1) == ord('q'):

        break

# Release Capture and destroy windows

cap.release()

cv2.destroyAllWindows()

Yolo-layer.c

#include "yolo\_layer.h"

#include "activations.h"

#include "blas.h"

#include "box.h"

#include "cuda.h"

#include "utils.h"

#include <stdio.h>

#include <assert.h>

#include <string.h>

#include <stdlib.h>

layer make\_yolo\_layer(int batch, int w, int h, int n, int total, int \*mask, int classes)

{

    int i;

    layer l = {0};

    l.type = YOLO;

    l.n = n;

    l.total = total;

    l.batch = batch;

    l.h = h;

    l.w = w;

    l.c = n\*(classes + 4 + 1);

    l.out\_w = l.w;

    l.out\_h = l.h;

    l.out\_c = l.c;

    l.classes = classes;

    l.cost = calloc(1, sizeof(float));

    l.biases = calloc(total\*2, sizeof(float));

    if(mask) l.mask = mask;

    else{

        l.mask = calloc(n, sizeof(int));

        for(i = 0; i < n; ++i){

            l.mask[i] = i;

        }

    }

    l.bias\_updates = calloc(n\*2, sizeof(float));

    l.outputs = h\*w\*n\*(classes + 4 + 1);

    l.inputs = l.outputs;

    l.truths = 90\*(4 + 1);

    l.delta = calloc(batch\*l.outputs, sizeof(float));

    l.output = calloc(batch\*l.outputs, sizeof(float));

    for(i = 0; i < total\*2; ++i){

        l.biases[i] = .5;

    }

    l.forward = forward\_yolo\_layer;

    l.backward = backward\_yolo\_layer;

#ifdef GPU

    l.forward\_gpu = forward\_yolo\_layer\_gpu;

    l.backward\_gpu = backward\_yolo\_layer\_gpu;

    l.output\_gpu = cuda\_make\_array(l.output, batch\*l.outputs);

    l.delta\_gpu = cuda\_make\_array(l.delta, batch\*l.outputs);

#endif

    fprintf(stderr, "yolo\n");

    srand(0);

    return l;

}

void resize\_yolo\_layer(layer \*l, int w, int h)

{

    l->w = w;

    l->h = h;

    l->outputs = h\*w\*l->n\*(l->classes + 4 + 1);

    l->inputs = l->outputs;

    l->output = realloc(l->output, l->batch\*l->outputs\*sizeof(float));

    l->delta = realloc(l->delta, l->batch\*l->outputs\*sizeof(float));

#ifdef GPU

    cuda\_free(l->delta\_gpu);

    cuda\_free(l->output\_gpu);

    l->delta\_gpu =     cuda\_make\_array(l->delta, l->batch\*l->outputs);

    l->output\_gpu =    cuda\_make\_array(l->output, l->batch\*l->outputs);

#endif

}

box get\_yolo\_box(float \*x, float \*biases, int n, int index, int i, int j, int lw, int lh, int w, int h, int stride)

{

    box b;

    b.x = (i + x[index + 0\*stride]) / lw;

    b.y = (j + x[index + 1\*stride]) / lh;

    b.w = exp(x[index + 2\*stride]) \* biases[2\*n]   / w;

    b.h = exp(x[index + 3\*stride]) \* biases[2\*n+1] / h;

    return b;

}

float delta\_yolo\_box(box truth, float \*x, float \*biases, int n, int index, int i, int j, int lw, int lh, int w, int h, float \*delta, float scale, int stride)

{

    box pred = get\_yolo\_box(x, biases, n, index, i, j, lw, lh, w, h, stride);

    float iou = box\_iou(pred, truth);

    float tx = (truth.x\*lw - i);

    float ty = (truth.y\*lh - j);

    float tw = log(truth.w\*w / biases[2\*n]);

    float th = log(truth.h\*h / biases[2\*n + 1]);

    delta[index + 0\*stride] = scale \* (tx - x[index + 0\*stride]);

    delta[index + 1\*stride] = scale \* (ty - x[index + 1\*stride]);

    delta[index + 2\*stride] = scale \* (tw - x[index + 2\*stride]);

    delta[index + 3\*stride] = scale \* (th - x[index + 3\*stride]);

    return iou;

}

void delta\_yolo\_class(float \*output, float \*delta, int index, int class, int classes, int stride, float \*avg\_cat)

{

    int n;

    if (delta[index]){

        delta[index + stride\*class] = 1 - output[index + stride\*class];

        if(avg\_cat) \*avg\_cat += output[index + stride\*class];

        return;

    }

    for(n = 0; n < classes; ++n){

        delta[index + stride\*n] = ((n == class)?1 : 0) - output[index + stride\*n];

        if(n == class && avg\_cat) \*avg\_cat += output[index + stride\*n];

    }

}

static int entry\_index(layer l, int batch, int location, int entry)

{

    int n =   location / (l.w\*l.h);

    int loc = location % (l.w\*l.h);

    return batch\*l.outputs + n\*l.w\*l.h\*(4+l.classes+1) + entry\*l.w\*l.h + loc;

}

void forward\_yolo\_layer(const layer l, network net)

{

    int i,j,b,t,n;

    memcpy(l.output, net.input, l.outputs\*l.batch\*sizeof(float));

#ifndef GPU

    for (b = 0; b < l.batch; ++b){

        for(n = 0; n < l.n; ++n){

            int index = entry\_index(l, b, n\*l.w\*l.h, 0);

            activate\_array(l.output + index, 2\*l.w\*l.h, LOGISTIC);

            index = entry\_index(l, b, n\*l.w\*l.h, 4);

            activate\_array(l.output + index, (1+l.classes)\*l.w\*l.h, LOGISTIC);

        }

    }

#endif

    memset(l.delta, 0, l.outputs \* l.batch \* sizeof(float));

    if(!net.train) return;

    float avg\_iou = 0;

    float recall = 0;

    float recall75 = 0;

    float avg\_cat = 0;

    float avg\_obj = 0;

    float avg\_anyobj = 0;

    int count = 0;

    int class\_count = 0;

    \*(l.cost) = 0;

    for (b = 0; b < l.batch; ++b) {

        for (j = 0; j < l.h; ++j) {

            for (i = 0; i < l.w; ++i) {

                for (n = 0; n < l.n; ++n) {

                    int box\_index = entry\_index(l, b, n\*l.w\*l.h + j\*l.w + i, 0);

                    box pred = get\_yolo\_box(l.output, l.biases, l.mask[n], box\_index, i, j, l.w, l.h, net.w, net.h, l.w\*l.h);

                    float best\_iou = 0;

                    int best\_t = 0;

                    for(t = 0; t < l.max\_boxes; ++t){

                        box truth = float\_to\_box(net.truth + t\*(4 + 1) + b\*l.truths, 1);

                        if(!truth.x) break;

                        float iou = box\_iou(pred, truth);

                        if (iou > best\_iou) {

                            best\_iou = iou;

                            best\_t = t;

                        }

                    }

                    int obj\_index = entry\_index(l, b, n\*l.w\*l.h + j\*l.w + i, 4);

                    avg\_anyobj += l.output[obj\_index];

                    l.delta[obj\_index] = 0 - l.output[obj\_index];

                    if (best\_iou > l.ignore\_thresh) {

                        l.delta[obj\_index] = 0;

                    }

                    if (best\_iou > l.truth\_thresh) {

                        l.delta[obj\_index] = 1 - l.output[obj\_index];

                        int class = net.truth[best\_t\*(4 + 1) + b\*l.truths + 4];

                        if (l.map) class = l.map[class];

                        int class\_index = entry\_index(l, b, n\*l.w\*l.h + j\*l.w + i, 4 + 1);

                        delta\_yolo\_class(l.output, l.delta, class\_index, class, l.classes, l.w\*l.h, 0);

                        box truth = float\_to\_box(net.truth + best\_t\*(4 + 1) + b\*l.truths, 1);

                        delta\_yolo\_box(truth, l.output, l.biases, l.mask[n], box\_index, i, j, l.w, l.h, net.w, net.h, l.delta, (2-truth.w\*truth.h), l.w\*l.h);

                    }

                }

            }

        }

        for(t = 0; t < l.max\_boxes; ++t){

            box truth = float\_to\_box(net.truth + t\*(4 + 1) + b\*l.truths, 1);

            if(!truth.x) break;

            float best\_iou = 0;

            int best\_n = 0;

            i = (truth.x \* l.w);

            j = (truth.y \* l.h);

            box truth\_shift = truth;

            truth\_shift.x = truth\_shift.y = 0;

            for(n = 0; n < l.total; ++n){

                box pred = {0};

                pred.w = l.biases[2\*n]/net.w;

                pred.h = l.biases[2\*n+1]/net.h;

                float iou = box\_iou(pred, truth\_shift);

                if (iou > best\_iou){

                    best\_iou = iou;

                    best\_n = n;

                }

            }

            int mask\_n = int\_index(l.mask, best\_n, l.n);

            if(mask\_n >= 0){

                int box\_index = entry\_index(l, b, mask\_n\*l.w\*l.h + j\*l.w + i, 0);

                float iou = delta\_yolo\_box(truth, l.output, l.biases, best\_n, box\_index, i, j, l.w, l.h, net.w, net.h, l.delta, (2-truth.w\*truth.h), l.w\*l.h);

                int obj\_index = entry\_index(l, b, mask\_n\*l.w\*l.h + j\*l.w + i, 4);

                avg\_obj += l.output[obj\_index];

                l.delta[obj\_index] = 1 - l.output[obj\_index];

                int class = net.truth[t\*(4 + 1) + b\*l.truths + 4];

                if (l.map) class = l.map[class];

                int class\_index = entry\_index(l, b, mask\_n\*l.w\*l.h + j\*l.w + i, 4 + 1);

                delta\_yolo\_class(l.output, l.delta, class\_index, class, l.classes, l.w\*l.h, &avg\_cat);

                ++count;

                ++class\_count;

                if(iou > .5) recall += 1;

                if(iou > .75) recall75 += 1;

                avg\_iou += iou;

            }

        }

    }

    \*(l.cost) = pow(mag\_array(l.delta, l.outputs \* l.batch), 2);

    printf("Region %d Avg IOU: %f, Class: %f, Obj: %f, No Obj: %f, .5R: %f, .75R: %f,  count: %d\n", net.index, avg\_iou/count, avg\_cat/class\_count, avg\_obj/count, avg\_anyobj/(l.w\*l.h\*l.n\*l.batch), recall/count, recall75/count, count);

}

void backward\_yolo\_layer(const layer l, network net)

{

   axpy\_cpu(l.batch\*l.inputs, 1, l.delta, 1, net.delta, 1);

}

void correct\_yolo\_boxes(detection \*dets, int n, int w, int h, int netw, int neth, int relative)

{

    int i;

    int new\_w=0;

    int new\_h=0;

    if (((float)netw/w) < ((float)neth/h)) {

        new\_w = netw;

        new\_h = (h \* netw)/w;

    } else {

        new\_h = neth;

        new\_w = (w \* neth)/h;

    }

    for (i = 0; i < n; ++i){

        box b = dets[i].bbox;

        b.x =  (b.x - (netw - new\_w)/2./netw) / ((float)new\_w/netw);

        b.y =  (b.y - (neth - new\_h)/2./neth) / ((float)new\_h/neth);

        b.w \*= (float)netw/new\_w;

        b.h \*= (float)neth/new\_h;

        if(!relative){

            b.x \*= w;

            b.w \*= w;

            b.y \*= h;

            b.h \*= h;

        }

        dets[i].bbox = b;

    }

}

int yolo\_num\_detections(layer l, float thresh)

{

    int i, n;

    int count = 0;

    for (i = 0; i < l.w\*l.h; ++i){

        for(n = 0; n < l.n; ++n){

            int obj\_index  = entry\_index(l, 0, n\*l.w\*l.h + i, 4);

            if(l.output[obj\_index] > thresh){

                ++count;

            }

        }

    }

    return count;

}

void avg\_flipped\_yolo(layer l)

{

    int i,j,n,z;

    float \*flip = l.output + l.outputs;

    for (j = 0; j < l.h; ++j) {

        for (i = 0; i < l.w/2; ++i) {

            for (n = 0; n < l.n; ++n) {

                for(z = 0; z < l.classes + 4 + 1; ++z){

                    int i1 = z\*l.w\*l.h\*l.n + n\*l.w\*l.h + j\*l.w + i;

                    int i2 = z\*l.w\*l.h\*l.n + n\*l.w\*l.h + j\*l.w + (l.w - i - 1);

                    float swap = flip[i1];

                    flip[i1] = flip[i2];

                    flip[i2] = swap;

                    if(z == 0){

                        flip[i1] = -flip[i1];

                        flip[i2] = -flip[i2];

                    }

                }

            }

        }

    }

    for(i = 0; i < l.outputs; ++i){

        l.output[i] = (l.output[i] + flip[i])/2.;

    }

}

int get\_yolo\_detections(layer l, int w, int h, int netw, int neth, float thresh, int \*map, int relative, detection \*dets)

{

    int i,j,n;

    float \*predictions = l.output;

    if (l.batch == 2) avg\_flipped\_yolo(l);

    int count = 0;

    for (i = 0; i < l.w\*l.h; ++i){

        int row = i / l.w;

        int col = i % l.w;

        for(n = 0; n < l.n; ++n){

            int obj\_index  = entry\_index(l, 0, n\*l.w\*l.h + i, 4);

            float objectness = predictions[obj\_index];

            if(objectness <= thresh) continue;

            int box\_index  = entry\_index(l, 0, n\*l.w\*l.h + i, 0);

            dets[count].bbox = get\_yolo\_box(predictions, l.biases, l.mask[n], box\_index, col, row, l.w, l.h, netw, neth, l.w\*l.h);

            dets[count].objectness = objectness;

            dets[count].classes = l.classes;

            for(j = 0; j < l.classes; ++j){

                int class\_index = entry\_index(l, 0, n\*l.w\*l.h + i, 4 + 1 + j);

                float prob = objectness\*predictions[class\_index];

                dets[count].prob[j] = (prob > thresh) ? prob : 0;

            }

            ++count;

        }

    }

    correct\_yolo\_boxes(dets, count, w, h, netw, neth, relative);

    return count;

}

#ifdef GPU

void forward\_yolo\_layer\_gpu(const layer l, network net)

{

    copy\_gpu(l.batch\*l.inputs, net.input\_gpu, 1, l.output\_gpu, 1);

    int b, n;

    for (b = 0; b < l.batch; ++b){

        for(n = 0; n < l.n; ++n){

            int index = entry\_index(l, b, n\*l.w\*l.h, 0);

            activate\_array\_gpu(l.output\_gpu + index, 2\*l.w\*l.h, LOGISTIC);

            index = entry\_index(l, b, n\*l.w\*l.h, 4);

            activate\_array\_gpu(l.output\_gpu + index, (1+l.classes)\*l.w\*l.h, LOGISTIC);

        }

    }

    if(!net.train || l.onlyforward){

        cuda\_pull\_array(l.output\_gpu, l.output, l.batch\*l.outputs);

        return;

    }

    cuda\_pull\_array(l.output\_gpu, net.input, l.batch\*l.inputs);

    forward\_yolo\_layer(l, net);

    cuda\_push\_array(l.delta\_gpu, l.delta, l.batch\*l.outputs);

}

void backward\_yolo\_layer\_gpu(const layer l, network net)

{

    axpy\_gpu(l.batch\*l.inputs, 1, l.delta\_gpu, 1, net.delta\_gpu, 1);

}

#endif

Yolo-layer.h

#ifndef YOLO\_LAYER\_H

#define YOLO\_LAYER\_H

#include "darknet.h"

#include "layer.h"

#include "network.h"

layer make\_yolo\_layer(int batch, int w, int h, int n, int total, int \*mask, int classes);

void forward\_yolo\_layer(const layer l, network net);

void backward\_yolo\_layer(const layer l, network net);

void resize\_yolo\_layer(layer \*l, int w, int h);

int yolo\_num\_detections(layer l, float thresh);

#ifdef GPU

void forward\_yolo\_layer\_gpu(const layer l, network net);

void backward\_yolo\_layer\_gpu(layer l, network net);

#endif

#endif

Darknet/readme.md

![Darknet Logo](http://pjreddie.com/media/files/darknet-black-small.png)

**# Darknet #**

Darknet is an open source neural network framework written in C and CUDA. It is fast, easy to install, and supports CPU and GPU computation.

**\*\*Discord\*\*** invite link for for communication and questions: https://discord.gg/zSq8rtW

**## YOLOv7:**

\* **\*\*paper\*\*** - YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors: https://arxiv.org/abs/2207.02696

\* **\*\*source code - Pytorch (use to reproduce results):\*\*** https://github.com/WongKinYiu/yolov7

----

Official YOLOv7 is more accurate and faster than YOLOv5 by **\*\*120%\*\*** FPS, than YOLOX by **\*\*180%\*\*** FPS, than Dual-Swin-T by **\*\*1200%\*\*** FPS, than ConvNext by **\*\*550%\*\*** FPS, than SWIN-L by **\*\*500%\*\*** FPS.

YOLOv7 surpasses all known object detectors in both speed and accuracy in the range from 5 FPS to 160 FPS and has the highest accuracy 56.8% AP among all known real-time object detectors with 30 FPS or higher on GPU V100, batch=1.

\* YOLOv7-e6 (55.9% AP, 56 FPS V100 b=1) by `+500%` FPS faster than SWIN-L Cascade-Mask R-CNN (53.9% AP, 9.2 FPS A100 b=1)

\* YOLOv7-e6 (55.9% AP, 56 FPS V100 b=1) by `+550%` FPS faster than ConvNeXt-XL C-M-RCNN (55.2% AP, 8.6 FPS A100 b=1)

\* YOLOv7-w6 (54.6% AP, 84 FPS V100 b=1) by `+120%` FPS faster than YOLOv5-X6-r6.1 (55.0% AP, 38 FPS V100 b=1)

\* YOLOv7-w6 (54.6% AP, 84 FPS V100 b=1) by `+1200%` FPS faster than Dual-Swin-T C-M-RCNN (53.6% AP, 6.5 FPS V100 b=1)

\* YOLOv7x (52.9% AP, 114 FPS V100 b=1) by `+150%` FPS faster than PPYOLOE-X (51.9% AP, 45 FPS V100 b=1)

\* YOLOv7 (51.2% AP, 161 FPS V100 b=1) by `+180%` FPS faster than YOLOX-X (51.1% AP, 58 FPS V100 b=1)

----

![more5](https://user-images.githubusercontent.com/4096485/179425274-f55a36d4-8450-4471-816b-8c105841effd.jpg)

----

![image](https://user-images.githubusercontent.com/4096485/177675030-a929ee00-0eba-4d93-95c2-225231d0fd61.png)

----

![yolov7\_640\_1280](https://user-images.githubusercontent.com/4096485/177688869-d75e0c36-63af-46ec-bdbd-81dbb281f257.png)

----

**## Scaled-YOLOv4:**

\* **\*\*paper (CVPR 2021)\*\***: https://openaccess.thecvf.com/content/CVPR2021/html/Wang\_Scaled-YOLOv4\_Scaling\_Cross\_Stage\_Partial\_Network\_CVPR\_2021\_paper.html

\* **\*\*source code - Pytorch (use to reproduce results):\*\*** https://github.com/WongKinYiu/ScaledYOLOv4

\* **\*\*source code - Darknet:\*\*** https://github.com/AlexeyAB/darknet

\* **\*\*Medium:\*\*** https://alexeyab84.medium.com/scaled-yolo-v4-is-the-best-neural-network-for-object-detection-on-ms-coco-dataset-39dfa22fa982?source=friends\_link&sk=c8553bfed861b1a7932f739d26f487c8

**## YOLOv4:**

\* **\*\*paper:\*\*** https://arxiv.org/abs/2004.10934

\* **\*\*source code:\*\*** https://github.com/AlexeyAB/darknet

\* **\*\*Wiki:\*\*** https://github.com/AlexeyAB/darknet/wiki

\* **\*\*useful links:\*\*** https://medium.com/@alexeyab84/yolov4-the-most-accurate-real-time-neural-network-on-ms-coco-dataset-73adfd3602fe?source=friends\_link&sk=6039748846bbcf1d960c3061542591d7

For more information see the [Darknet project website](http://pjreddie.com/darknet).

<details><summary> <b>Expand</b> </summary>

![yolo\_progress](https://user-images.githubusercontent.com/4096485/146988929-1ed0cbec-1e01-4ad0-b42c-808dcef32994.png) https://paperswithcode.com/sota/object-detection-on-coco

----

![scaled\_yolov4](https://user-images.githubusercontent.com/4096485/112776361-281d8380-9048-11eb-8083-8728b12dcd55.png) AP50:95 - FPS (Tesla V100) Paper: https://arxiv.org/abs/2011.08036

----

![YOLOv4Tiny](https://user-images.githubusercontent.com/4096485/101363015-e5c21200-38b1-11eb-986f-b3e516e05977.png)

----

![YOLOv4](https://user-images.githubusercontent.com/4096485/90338826-06114c80-dff5-11ea-9ba2-8eb63a7409b3.png)

</details>

----

![OpenCV\_TRT](https://user-images.githubusercontent.com/4096485/90338805-e5e18d80-dff4-11ea-8a68-5710956256ff.png)

**## Citation**

```

@misc{https://doi.org/10.48550/arxiv.2207.02696,

  doi = {10.48550/ARXIV.2207.02696},

  url = {https://arxiv.org/abs/2207.02696},

  author = {Wang, Chien-Yao and Bochkovskiy, Alexey and Liao, Hong-Yuan Mark},

  keywords = {Computer Vision and Pattern Recognition (cs.CV), FOS: Computer and information sciences, FOS: Computer and information sciences},

  title = {YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors},

  publisher = {arXiv},

  year = {2022},

  copyright = {arXiv.org perpetual, non-exclusive license}

}

```

```

@misc{bochkovskiy2020yolov4,

      title={YOLOv4: Optimal Speed and Accuracy of Object Detection},

      author={Alexey Bochkovskiy and Chien-Yao Wang and Hong-Yuan Mark Liao},

      year={2020},

      eprint={2004.10934},

      archivePrefix={arXiv},

      primaryClass={cs.CV}

}

```

```

@InProceedings{Wang\_2021\_CVPR,

    author    = {Wang, Chien-Yao and Bochkovskiy, Alexey and Liao, Hong-Yuan Mark},

    title     = {{Scaled-YOLOv4}: Scaling Cross Stage Partial Network},

    booktitle = {Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)},

    month     = {June},

    year      = {2021},

    pages     = {13029-13038}

}

```

Coco.names

person

bicycle

car

motorbike

aeroplane

bus

train

truck

boat

traffic-light

fire-hydrant

stop-sign

parking-meter

bench

bird

cat

dog

horse

sheep

cow

elephant

bear

zebra

giraffe

backpack

umbrella

handbag

tie

suitcase

frisbee

skis

snowboard

sports-ball

kite

baseball-bat

baseball-glove

skateboard

surfboard

tennis-racket

bottle

wine-glass

cup

fork

knife

spoon

bowl

banana

apple

sandwich

orange

broccoli

carrot

hot-dog

pizza

donut

cake

chair

sofa

pottedplant

bed

diningtable

toilet

tvmonitor

laptop

mouse

remote

keyboard

cell-phone

microwave

oven

toaster

sink

refrigerator

book

clock

vase

scissors

teddy-bear

hair-drier

toothbrush

Yolov3-tiny.cfg

[net]

# Testing

batch=1

subdivisions=1

# Training

# batch=64

# subdivisions=2

width=416

height=416

channels=3

momentum=0.9

decay=0.0005

angle=0

saturation = 1.5

exposure = 1.5

hue=.1

learning\_rate=0.001

burn\_in=1000

max\_batches = 500200

policy=steps

steps=400000,450000

scales=.1,.1

[convolutional]

batch\_normalize=1

filters=16

size=3

stride=1

pad=1

activation=leaky

[maxpool]

size=2

stride=2

[convolutional]

batch\_normalize=1

filters=32

size=3

stride=1

pad=1

activation=leaky

[maxpool]

size=2

stride=2

[convolutional]

batch\_normalize=1

filters=64

size=3

stride=1

pad=1

activation=leaky

[maxpool]

size=2

stride=2

[convolutional]

batch\_normalize=1

filters=128

size=3

stride=1

pad=1

activation=leaky

[maxpool]

size=2

stride=2

[convolutional]

batch\_normalize=1

filters=256

size=3

stride=1

pad=1

activation=leaky

[maxpool]

size=2

stride=2

[convolutional]

batch\_normalize=1

filters=512

size=3

stride=1

pad=1

activation=leaky

[maxpool]

size=2

stride=1

[convolutional]

batch\_normalize=1

filters=1024

size=3

stride=1

pad=1

activation=leaky

###########

[convolutional]

batch\_normalize=1

filters=256

size=1

stride=1

pad=1

activation=leaky

[convolutional]

batch\_normalize=1

filters=512

size=3

stride=1

pad=1

activation=leaky

[convolutional]

size=1

stride=1

pad=1

filters=255

activation=linear

[yolo]

mask = 3,4,5

anchors = 10,14,  23,27,  37,58,  81,82,  135,169,  344,319

classes=80

num=6

jitter=.3

ignore\_thresh = .7

truth\_thresh = 1

random=1

[route]

layers = -4

[convolutional]

batch\_normalize=1

filters=128

size=1

stride=1

pad=1

activation=leaky

[upsample]

stride=2

[route]

layers = -1, 8

[convolutional]

batch\_normalize=1

filters=256

size=3

stride=1

pad=1

activation=leaky

[convolutional]

size=1

stride=1

pad=1

filters=255

activation=linear

[yolo]

mask = 0,1,2

anchors = 10,14,  23,27,  37,58,  81,82,  135,169,  344,319

classes=80

num=6

jitter=.3

ignore\_thresh = .7

truth\_thresh = 1

random=1

Yolov3.cfg

[net]

# Testing

# batch=1

# subdivisions=1

# Training

batch=64

subdivisions=16

width=608

height=608

channels=3

momentum=0.9

decay=0.0005

angle=0

saturation = 1.5

exposure = 1.5

hue=.1

learning\_rate=0.001

burn\_in=1000

max\_batches = 500200

policy=steps

steps=400000,450000

scales=.1,.1

[convolutional]

batch\_normalize=1

filters=32

size=3

stride=1

pad=1

activation=leaky

# Downsample

[convolutional]

batch\_normalize=1

filters=64

size=3

stride=2

pad=1

activation=leaky

[convolutional]

batch\_normalize=1

filters=32

size=1

stride=1

pad=1

activation=leaky

[convolutional]

batch\_normalize=1

filters=64

size=3

stride=1

pad=1

activation=leaky

[shortcut]

from=-3

activation=linear

# Downsample

[convolutional]

batch\_normalize=1

filters=128

size=3

stride=2

pad=1

activation=leaky

[convolutional]

batch\_normalize=1

filters=64

size=1

stride=1

pad=1

activation=leaky

[convolutional]

batch\_normalize=1

filters=128

size=3

stride=1

pad=1

activation=leaky

[shortcut]

from=-3

activation=linear

[convolutional]

batch\_normalize=1

filters=64

size=1

stride=1

pad=1

activation=leaky

[convolutional]

batch\_normalize=1

filters=128

size=3

stride=1

pad=1

activation=leaky

[shortcut]

from=-3

activation=linear

# Downsample

[convolutional]

batch\_normalize=1

filters=256

size=3

stride=2

pad=1

activation=leaky

[convolutional]

batch\_normalize=1

filters=128

size=1

stride=1

pad=1

activation=leaky

[convolutional]

batch\_normalize=1

filters=256

size=3

stride=1

pad=1

activation=leaky

[shortcut]

from=-3

activation=linear

[convolutional]

batch\_normalize=1

filters=128

size=1

stride=1

pad=1

activation=leaky

[convolutional]

batch\_normalize=1

filters=256

size=3

stride=1

pad=1

activation=leaky

[shortcut]

from=-3

activation=linear

[convolutional]

batch\_normalize=1

filters=128

size=1

stride=1

pad=1

activation=leaky

[convolutional]

batch\_normalize=1

filters=256

size=3

stride=1

pad=1

activation=leaky

[shortcut]

from=-3

activation=linear

[convolutional]

batch\_normalize=1

filters=128

size=1

stride=1

pad=1

activation=leaky

[convolutional]

batch\_normalize=1

filters=256

size=3

stride=1

pad=1

activation=leaky

[shortcut]

from=-3

activation=linear

[convolutional]

batch\_normalize=1

filters=128

size=1

stride=1

pad=1

activation=leaky

[convolutional]

batch\_normalize=1

filters=256

size=3

stride=1

pad=1

activation=leaky

[shortcut]

from=-3

activation=linear

[convolutional]

batch\_normalize=1

filters=128

size=1

stride=1

pad=1

activation=leaky

[convolutional]

batch\_normalize=1

filters=256

size=3

stride=1

pad=1

activation=leaky

[shortcut]

from=-3

activation=linear

[convolutional]

batch\_normalize=1

filters=128

size=1

stride=1

pad=1

activation=leaky

[convolutional]

batch\_normalize=1

filters=256

size=3

stride=1

pad=1

activation=leaky

[shortcut]

from=-3

activation=linear

[convolutional]

batch\_normalize=1

filters=128

size=1

stride=1

pad=1

activation=leaky

[convolutional]

batch\_normalize=1

filters=256

size=3

stride=1

pad=1

activation=leaky

[shortcut]

from=-3

activation=linear

# Downsample

[convolutional]

batch\_normalize=1

filters=512

size=3

stride=2

pad=1

activation=leaky

[convolutional]

batch\_normalize=1

filters=256

size=1

stride=1

pad=1

activation=leaky

[convolutional]

batch\_normalize=1

filters=512

size=3

stride=1

pad=1

activation=leaky

[shortcut]

from=-3

activation=linear

[convolutional]

batch\_normalize=1

filters=256

size=1

stride=1

pad=1

activation=leaky

[convolutional]

batch\_normalize=1

filters=512

size=3

stride=1

pad=1

activation=leaky

[shortcut]

from=-3

activation=linear

[convolutional]

batch\_normalize=1

filters=256

size=1

stride=1

pad=1

activation=leaky

[convolutional]

batch\_normalize=1

filters=512

size=3

stride=1

pad=1

activation=leaky

[shortcut]

from=-3

activation=linear

[convolutional]

batch\_normalize=1

filters=256

size=1

stride=1

pad=1

activation=leaky

[convolutional]

batch\_normalize=1

filters=512

size=3

stride=1

pad=1

activation=leaky

[shortcut]

from=-3

activation=linear

[convolutional]

batch\_normalize=1

filters=256

size=1

stride=1

pad=1

activation=leaky

[convolutional]

batch\_normalize=1

filters=512

size=3

stride=1

pad=1

activation=leaky

[shortcut]

from=-3

activation=linear

[convolutional]

batch\_normalize=1

filters=256

size=1

stride=1

pad=1

activation=leaky

[convolutional]

batch\_normalize=1

filters=512

size=3

stride=1

pad=1

activation=leaky

[shortcut]

from=-3

activation=linear

[convolutional]

batch\_normalize=1

filters=256

size=1

stride=1

pad=1

activation=leaky

[convolutional]

batch\_normalize=1

filters=512

size=3

stride=1

pad=1

activation=leaky

[shortcut]

from=-3

activation=linear

[convolutional]

batch\_normalize=1

filters=256

size=1

stride=1

pad=1

activation=leaky

[convolutional]

batch\_normalize=1

filters=512

size=3

stride=1

pad=1

activation=leaky

[shortcut]

from=-3

activation=linear

# Downsample

[convolutional]

batch\_normalize=1

filters=1024

size=3

stride=2

pad=1

activation=leaky

[convolutional]

batch\_normalize=1

filters=512

size=1

stride=1

pad=1

activation=leaky

[convolutional]

batch\_normalize=1

filters=1024

size=3

stride=1

pad=1

activation=leaky

[shortcut]

from=-3

activation=linear

[convolutional]

batch\_normalize=1

filters=512

size=1

stride=1

pad=1

activation=leaky

[convolutional]

batch\_normalize=1

filters=1024

size=3

stride=1

pad=1

activation=leaky

[shortcut]

from=-3

activation=linear

[convolutional]

batch\_normalize=1

filters=512

size=1

stride=1

pad=1

activation=leaky

[convolutional]

batch\_normalize=1

filters=1024

size=3

stride=1

pad=1

activation=leaky

[shortcut]

from=-3

activation=linear

[convolutional]

batch\_normalize=1

filters=512

size=1

stride=1

pad=1

activation=leaky

[convolutional]

batch\_normalize=1

filters=1024

size=3

stride=1

pad=1

activation=leaky

[shortcut]

from=-3

activation=linear

######################

[convolutional]

batch\_normalize=1

filters=512

size=1

stride=1

pad=1

activation=leaky

[convolutional]

batch\_normalize=1

size=3

stride=1

pad=1

filters=1024

activation=leaky

[convolutional]

batch\_normalize=1

filters=512

size=1

stride=1

pad=1

activation=leaky

[convolutional]

batch\_normalize=1

size=3

stride=1

pad=1

filters=1024

activation=leaky

[convolutional]

batch\_normalize=1

filters=512

size=1

stride=1

pad=1

activation=leaky

[convolutional]

batch\_normalize=1

size=3

stride=1

pad=1

filters=1024

activation=leaky

[convolutional]

size=1

stride=1

pad=1

filters=255

activation=linear

[yolo]

mask = 6,7,8

anchors = 10,13,  16,30,  33,23,  30,61,  62,45,  59,119,  116,90,  156,198,  373,326

classes=80

num=9

jitter=.3

ignore\_thresh = .7

truth\_thresh = 1

random=1

[route]

layers = -4

[convolutional]

batch\_normalize=1

filters=256

size=1

stride=1

pad=1

activation=leaky

[upsample]

stride=2

[route]

layers = -1, 61

[convolutional]

batch\_normalize=1

filters=256

size=1

stride=1

pad=1

activation=leaky

[convolutional]

batch\_normalize=1

size=3

stride=1

pad=1

filters=512

activation=leaky

[convolutional]

batch\_normalize=1

filters=256

size=1

stride=1

pad=1

activation=leaky

[convolutional]

batch\_normalize=1

size=3

stride=1

pad=1

filters=512

activation=leaky

[convolutional]

batch\_normalize=1

filters=256

size=1

stride=1

pad=1

activation=leaky

[convolutional]

batch\_normalize=1

size=3

stride=1

pad=1

filters=512

activation=leaky

[convolutional]

size=1

stride=1

pad=1

filters=255

activation=linear

[yolo]

mask = 3,4,5

anchors = 10,13,  16,30,  33,23,  30,61,  62,45,  59,119,  116,90,  156,198,  373,326

classes=80

num=9

jitter=.3

ignore\_thresh = .7

truth\_thresh = 1

random=1

[route]

layers = -4

[convolutional]

batch\_normalize=1

filters=128

size=1

stride=1

pad=1

activation=leaky

[upsample]

stride=2

[route]

layers = -1, 36

[convolutional]

batch\_normalize=1

filters=128

size=1

stride=1

pad=1

activation=leaky

[convolutional]

batch\_normalize=1

size=3

stride=1

pad=1

filters=256

activation=leaky

[convolutional]

batch\_normalize=1

filters=128

size=1

stride=1

pad=1

activation=leaky

[convolutional]

batch\_normalize=1

size=3

stride=1

pad=1

filters=256

activation=leaky

[convolutional]

batch\_normalize=1

filters=128

size=1

stride=1

pad=1

activation=leaky

[convolutional]

batch\_normalize=1

size=3

stride=1

pad=1

filters=256

activation=leaky

[convolutional]

size=1

stride=1

pad=1

filters=255

activation=linear

[yolo]

mask = 0,1,2

anchors = 10,13,  16,30,  33,23,  30,61,  62,45,  59,119,  116,90,  156,198,  373,326

classes=80

num=9

jitter=.3

ignore\_thresh = .7

truth\_thresh = 1

random=1

Utils.h

#ifndef UTILS\_H

#define UTILS\_H

#include <stdio.h>

#include <time.h>

#include "darknet.h"

#include "list.h"

#define TIME(a) \

    do { \

    double start = what\_time\_is\_it\_now(); \

    a; \

    printf("%s took: %f seconds\n", #a, what\_time\_is\_it\_now() - start); \

    } while (0)

#define TWO\_PI 6.2831853071795864769252866f

double what\_time\_is\_it\_now();

void shuffle(void \*arr, size\_t n, size\_t size);

void sorta\_shuffle(void \*arr, size\_t n, size\_t size, size\_t sections);

void free\_ptrs(void \*\*ptrs, int n);

int alphanum\_to\_int(char c);

char int\_to\_alphanum(int i);

int read\_int(int fd);

void write\_int(int fd, int n);

void read\_all(int fd, char \*buffer, size\_t bytes);

void write\_all(int fd, char \*buffer, size\_t bytes);

int read\_all\_fail(int fd, char \*buffer, size\_t bytes);

int write\_all\_fail(int fd, char \*buffer, size\_t bytes);

void find\_replace(char \*str, char \*orig, char \*rep, char \*output);

void malloc\_error();

void file\_error(char \*s);

void strip(char \*s);

void strip\_char(char \*s, char bad);

list \*split\_str(char \*s, char delim);

char \*fgetl(FILE \*fp);

list \*parse\_csv\_line(char \*line);

char \*copy\_string(char \*s);

int count\_fields(char \*line);

float \*parse\_fields(char \*line, int n);

void translate\_array(float \*a, int n, float s);

float constrain(float min, float max, float a);

int constrain\_int(int a, int min, int max);

float rand\_scale(float s);

int rand\_int(int min, int max);

void mean\_arrays(float \*\*a, int n, int els, float \*avg);

float dist\_array(float \*a, float \*b, int n, int sub);

float \*\*one\_hot\_encode(float \*a, int n, int k);

float sec(clock\_t clocks);

void print\_statistics(float \*a, int n);

int int\_index(int \*a, int val, int n);

#endif

Utils.c

#include <stdio.h>

#include <stdlib.h>

#include <string.h>

#include <math.h>

#include <assert.h>

#include <unistd.h>

#include <float.h>

#include <limits.h>

#include <time.h>

#include <sys/time.h>

#include "utils.h"

/\*

// old timing. is it better? who knows!!

double get\_wall\_time()

{

    struct timeval time;

    if (gettimeofday(&time,NULL)){

        return 0;

    }

    return (double)time.tv\_sec + (double)time.tv\_usec \* .000001;

}

\*/

double what\_time\_is\_it\_now()

{

    struct timeval time;

    if (gettimeofday(&time,NULL)){

        return 0;

    }

    return (double)time.tv\_sec + (double)time.tv\_usec \* .000001;

}

int \*read\_intlist(char \*gpu\_list, int \*ngpus, int d)

{

    int \*gpus = 0;

    if(gpu\_list){

        int len = strlen(gpu\_list);

        \*ngpus = 1;

        int i;

        for(i = 0; i < len; ++i){

            if (gpu\_list[i] == ',') ++\*ngpus;

        }

        gpus = calloc(\*ngpus, sizeof(int));

        for(i = 0; i < \*ngpus; ++i){

            gpus[i] = atoi(gpu\_list);

            gpu\_list = strchr(gpu\_list, ',')+1;

        }

    } else {

        gpus = calloc(1, sizeof(float));

        \*gpus = d;

        \*ngpus = 1;

    }

    return gpus;

}

int \*read\_map(char \*filename)

{

    int n = 0;

    int \*map = 0;

    char \*str;

    FILE \*file = fopen(filename, "r");

    if(!file) file\_error(filename);

    while((str=fgetl(file))){

        ++n;

        map = realloc(map, n\*sizeof(int));

        map[n-1] = atoi(str);

    }

    return map;

}

void sorta\_shuffle(void \*arr, size\_t n, size\_t size, size\_t sections)

{

    size\_t i;

    for(i = 0; i < sections; ++i){

        size\_t start = n\*i/sections;

        size\_t end = n\*(i+1)/sections;

        size\_t num = end-start;

        shuffle(arr+(start\*size), num, size);

    }

}

void shuffle(void \*arr, size\_t n, size\_t size)

{

    size\_t i;

    void \*swp = calloc(1, size);

    for(i = 0; i < n-1; ++i){

        size\_t j = i + rand()/(RAND\_MAX / (n-i)+1);

        memcpy(swp,          arr+(j\*size), size);

        memcpy(arr+(j\*size), arr+(i\*size), size);

        memcpy(arr+(i\*size), swp,          size);

    }

}

int \*random\_index\_order(int min, int max)

{

    int \*inds = calloc(max-min, sizeof(int));

    int i;

    for(i = min; i < max; ++i){

        inds[i] = i;

    }

    for(i = min; i < max-1; ++i){

        int swap = inds[i];

        int index = i + rand()%(max-i);

        inds[i] = inds[index];

        inds[index] = swap;

    }

    return inds;

}

void del\_arg(int argc, char \*\*argv, int index)

{

    int i;

    for(i = index; i < argc-1; ++i) argv[i] = argv[i+1];

    argv[i] = 0;

}

int find\_arg(int argc, char\* argv[], char \*arg)

{

    int i;

    for(i = 0; i < argc; ++i) {

        if(!argv[i]) continue;

        if(0==strcmp(argv[i], arg)) {

            del\_arg(argc, argv, i);

            return 1;

        }

    }

    return 0;

}

int find\_int\_arg(int argc, char \*\*argv, char \*arg, int def)

{

    int i;

    for(i = 0; i < argc-1; ++i){

        if(!argv[i]) continue;

        if(0==strcmp(argv[i], arg)){

            def = atoi(argv[i+1]);

            del\_arg(argc, argv, i);

            del\_arg(argc, argv, i);

            break;

        }

    }

    return def;

}

float find\_float\_arg(int argc, char \*\*argv, char \*arg, float def)

{

    int i;

    for(i = 0; i < argc-1; ++i){

        if(!argv[i]) continue;

        if(0==strcmp(argv[i], arg)){

            def = atof(argv[i+1]);

            del\_arg(argc, argv, i);

            del\_arg(argc, argv, i);

            break;

        }

    }

    return def;

}

char \*find\_char\_arg(int argc, char \*\*argv, char \*arg, char \*def)

{

    int i;

    for(i = 0; i < argc-1; ++i){

        if(!argv[i]) continue;

        if(0==strcmp(argv[i], arg)){

            def = argv[i+1];

            del\_arg(argc, argv, i);

            del\_arg(argc, argv, i);

            break;

        }

    }

    return def;

}

char \*basecfg(char \*cfgfile)

{

    char \*c = cfgfile;

    char \*next;

    while((next = strchr(c, '/')))

    {

        c = next+1;

    }

    c = copy\_string(c);

    next = strchr(c, '.');

    if (next) \*next = 0;

    return c;

}

int alphanum\_to\_int(char c)

{

    return (c < 58) ? c - 48 : c-87;

}

char int\_to\_alphanum(int i)

{

    if (i == 36) return '.';

    return (i < 10) ? i + 48 : i + 87;

}

void pm(int M, int N, float \*A)

{

    int i,j;

    for(i =0 ; i < M; ++i){

        printf("%d ", i+1);

        for(j = 0; j < N; ++j){

            printf("%2.4f, ", A[i\*N+j]);

        }

        printf("\n");

    }

    printf("\n");

}

void find\_replace(char \*str, char \*orig, char \*rep, char \*output)

{

    char buffer[4096] = {0};

    char \*p;

    sprintf(buffer, "%s", str);

    if(!(p = strstr(buffer, orig))){  // Is 'orig' even in 'str'?

        sprintf(output, "%s", str);

        return;

    }

    \*p = '\0';

    sprintf(output, "%s%s%s", buffer, rep, p+strlen(orig));

}

float sec(clock\_t clocks)

{

    return (float)clocks/CLOCKS\_PER\_SEC;

}

void top\_k(float \*a, int n, int k, int \*index)

{

    int i,j;

    for(j = 0; j < k; ++j) index[j] = -1;

    for(i = 0; i < n; ++i){

        int curr = i;

        for(j = 0; j < k; ++j){

            if((index[j] < 0) || a[curr] > a[index[j]]){

                int swap = curr;

                curr = index[j];

                index[j] = swap;

            }

        }

    }

}

void error(const char \*s)

{

    perror(s);

    assert(0);

    exit(-1);

}

unsigned char \*read\_file(char \*filename)

{

    FILE \*fp = fopen(filename, "rb");

    size\_t size;

    fseek(fp, 0, SEEK\_END);

    size = ftell(fp);

    fseek(fp, 0, SEEK\_SET);

    unsigned char \*text = calloc(size+1, sizeof(char));

    fread(text, 1, size, fp);

    fclose(fp);

    return text;

}

void malloc\_error()

{

    fprintf(stderr, "Malloc error\n");

    exit(-1);

}

void file\_error(char \*s)

{

    fprintf(stderr, "Couldn't open file: %s\n", s);

    exit(0);

}

list \*split\_str(char \*s, char delim)

{

    size\_t i;

    size\_t len = strlen(s);

    list \*l = make\_list();

    list\_insert(l, s);

    for(i = 0; i < len; ++i){

        if(s[i] == delim){

            s[i] = '\0';

            list\_insert(l, &(s[i+1]));

        }

    }

    return l;

}

void strip(char \*s)

{

    size\_t i;

    size\_t len = strlen(s);

    size\_t offset = 0;

    for(i = 0; i < len; ++i){

        char c = s[i];

        if(c==' '||c=='\t'||c=='\n') ++offset;

        else s[i-offset] = c;

    }

    s[len-offset] = '\0';

}

void strip\_char(char \*s, char bad)

{

    size\_t i;

    size\_t len = strlen(s);

    size\_t offset = 0;

    for(i = 0; i < len; ++i){

        char c = s[i];

        if(c==bad) ++offset;

        else s[i-offset] = c;

    }

    s[len-offset] = '\0';

}

void free\_ptrs(void \*\*ptrs, int n)

{

    int i;

    for(i = 0; i < n; ++i) free(ptrs[i]);

    free(ptrs);

}

char \*fgetl(FILE \*fp)

{

    if(feof(fp)) return 0;

    size\_t size = 512;

    char \*line = malloc(size\*sizeof(char));

    if(!fgets(line, size, fp)){

        free(line);

        return 0;

    }

    size\_t curr = strlen(line);

    while((line[curr-1] != '\n') && !feof(fp)){

        if(curr == size-1){

            size \*= 2;

            line = realloc(line, size\*sizeof(char));

            if(!line) {

                printf("%ld\n", size);

                malloc\_error();

            }

        }

        size\_t readsize = size-curr;

        if(readsize > INT\_MAX) readsize = INT\_MAX-1;

        fgets(&line[curr], readsize, fp);

        curr = strlen(line);

    }

    if(line[curr-1] == '\n') line[curr-1] = '\0';

    return line;

}

int read\_int(int fd)

{

    int n = 0;

    int next = read(fd, &n, sizeof(int));

    if(next <= 0) return -1;

    return n;

}

void write\_int(int fd, int n)

{

    int next = write(fd, &n, sizeof(int));

    if(next <= 0) error("read failed");

}

int read\_all\_fail(int fd, char \*buffer, size\_t bytes)

{

    size\_t n = 0;

    while(n < bytes){

        int next = read(fd, buffer + n, bytes-n);

        if(next <= 0) return 1;

        n += next;

    }

    return 0;

}

int write\_all\_fail(int fd, char \*buffer, size\_t bytes)

{

    size\_t n = 0;

    while(n < bytes){

        size\_t next = write(fd, buffer + n, bytes-n);

        if(next <= 0) return 1;

        n += next;

    }

    return 0;

}

void read\_all(int fd, char \*buffer, size\_t bytes)

{

    size\_t n = 0;

    while(n < bytes){

        int next = read(fd, buffer + n, bytes-n);

        if(next <= 0) error("read failed");

        n += next;

    }

}

void write\_all(int fd, char \*buffer, size\_t bytes)

{

    size\_t n = 0;

    while(n < bytes){

        size\_t next = write(fd, buffer + n, bytes-n);

        if(next <= 0) error("write failed");

        n += next;

    }

}

char \*copy\_string(char \*s)

{

    char \*copy = malloc(strlen(s)+1);

    strncpy(copy, s, strlen(s)+1);

    return copy;

}

list \*parse\_csv\_line(char \*line)

{

    list \*l = make\_list();

    char \*c, \*p;

    int in = 0;

    for(c = line, p = line; \*c != '\0'; ++c){

        if(\*c == '"') in = !in;

        else if(\*c == ',' && !in){

            \*c = '\0';

            list\_insert(l, copy\_string(p));

            p = c+1;

        }

    }

    list\_insert(l, copy\_string(p));

    return l;

}

int count\_fields(char \*line)

{

    int count = 0;

    int done = 0;

    char \*c;

    for(c = line; !done; ++c){

        done = (\*c == '\0');

        if(\*c == ',' || done) ++count;

    }

    return count;

}

float \*parse\_fields(char \*line, int n)

{

    float \*field = calloc(n, sizeof(float));

    char \*c, \*p, \*end;

    int count = 0;

    int done = 0;

    for(c = line, p = line; !done; ++c){

        done = (\*c == '\0');

        if(\*c == ',' || done){

            \*c = '\0';

            field[count] = strtod(p, &end);

            if(p == c) field[count] = nan("");

            if(end != c && (end != c-1 || \*end != '\r')) field[count] = nan(""); //DOS file formats!

            p = c+1;

            ++count;

        }

    }

    return field;

}

float sum\_array(float \*a, int n)

{

    int i;

    float sum = 0;

    for(i = 0; i < n; ++i) sum += a[i];

    return sum;

}

float mean\_array(float \*a, int n)

{

    return sum\_array(a,n)/n;

}

void mean\_arrays(float \*\*a, int n, int els, float \*avg)

{

    int i;

    int j;

    memset(avg, 0, els\*sizeof(float));

    for(j = 0; j < n; ++j){

        for(i = 0; i < els; ++i){

            avg[i] += a[j][i];

        }

    }

    for(i = 0; i < els; ++i){

        avg[i] /= n;

    }

}

void print\_statistics(float \*a, int n)

{

    float m = mean\_array(a, n);

    float v = variance\_array(a, n);

    printf("MSE: %.6f, Mean: %.6f, Variance: %.6f\n", mse\_array(a, n), m, v);

}

float variance\_array(float \*a, int n)

{

    int i;

    float sum = 0;

    float mean = mean\_array(a, n);

    for(i = 0; i < n; ++i) sum += (a[i] - mean)\*(a[i]-mean);

    float variance = sum/n;

    return variance;

}

int constrain\_int(int a, int min, int max)

{

    if (a < min) return min;

    if (a > max) return max;

    return a;

}

float constrain(float min, float max, float a)

{

    if (a < min) return min;

    if (a > max) return max;

    return a;

}

float dist\_array(float \*a, float \*b, int n, int sub)

{

    int i;

    float sum = 0;

    for(i = 0; i < n; i += sub) sum += pow(a[i]-b[i], 2);

    return sqrt(sum);

}

float mse\_array(float \*a, int n)

{

    int i;

    float sum = 0;

    for(i = 0; i < n; ++i) sum += a[i]\*a[i];

    return sqrt(sum/n);

}

void normalize\_array(float \*a, int n)

{

    int i;

    float mu = mean\_array(a,n);

    float sigma = sqrt(variance\_array(a,n));

    for(i = 0; i < n; ++i){

        a[i] = (a[i] - mu)/sigma;

    }

    mu = mean\_array(a,n);

    sigma = sqrt(variance\_array(a,n));

}

void translate\_array(float \*a, int n, float s)

{

    int i;

    for(i = 0; i < n; ++i){

        a[i] += s;

    }

}

float mag\_array(float \*a, int n)

{

    int i;

    float sum = 0;

    for(i = 0; i < n; ++i){

        sum += a[i]\*a[i];

    }

    return sqrt(sum);

}

void scale\_array(float \*a, int n, float s)

{

    int i;

    for(i = 0; i < n; ++i){

        a[i] \*= s;

    }

}

int sample\_array(float \*a, int n)

{

    float sum = sum\_array(a, n);

    scale\_array(a, n, 1./sum);

    float r = rand\_uniform(0, 1);

    int i;

    for(i = 0; i < n; ++i){

        r = r - a[i];

        if (r <= 0) return i;

    }

    return n-1;

}

int max\_int\_index(int \*a, int n)

{

    if(n <= 0) return -1;

    int i, max\_i = 0;

    int max = a[0];

    for(i = 1; i < n; ++i){

        if(a[i] > max){

            max = a[i];

            max\_i = i;

        }

    }

    return max\_i;

}

int max\_index(float \*a, int n)

{

    if(n <= 0) return -1;

    int i, max\_i = 0;

    float max = a[0];

    for(i = 1; i < n; ++i){

        if(a[i] > max){

            max = a[i];

            max\_i = i;

        }

    }

    return max\_i;

}

int int\_index(int \*a, int val, int n)

{

    int i;

    for(i = 0; i < n; ++i){

        if(a[i] == val) return i;

    }

    return -1;

}

int rand\_int(int min, int max)

{

    if (max < min){

        int s = min;

        min = max;

        max = s;

    }

    int r = (rand()%(max - min + 1)) + min;

    return r;

}

// From http://en.wikipedia.org/wiki/Box%E2%80%93Muller\_transform

float rand\_normal()

{

    static int haveSpare = 0;

    static double rand1, rand2;

    if(haveSpare)

    {

        haveSpare = 0;

        return sqrt(rand1) \* sin(rand2);

    }

    haveSpare = 1;

    rand1 = rand() / ((double) RAND\_MAX);

    if(rand1 < 1e-100) rand1 = 1e-100;

    rand1 = -2 \* log(rand1);

    rand2 = (rand() / ((double) RAND\_MAX)) \* TWO\_PI;

    return sqrt(rand1) \* cos(rand2);

}

/\*

   float rand\_normal()

   {

   int n = 12;

   int i;

   float sum= 0;

   for(i = 0; i < n; ++i) sum += (float)rand()/RAND\_MAX;

   return sum-n/2.;

   }

 \*/

size\_t rand\_size\_t()

{

    return  ((size\_t)(rand()&0xff) << 56) |

        ((size\_t)(rand()&0xff) << 48) |

        ((size\_t)(rand()&0xff) << 40) |

        ((size\_t)(rand()&0xff) << 32) |

        ((size\_t)(rand()&0xff) << 24) |

        ((size\_t)(rand()&0xff) << 16) |

        ((size\_t)(rand()&0xff) << 8) |

        ((size\_t)(rand()&0xff) << 0);

}

float rand\_uniform(float min, float max)

{

    if(max < min){

        float swap = min;

        min = max;

        max = swap;

    }

    return ((float)rand()/RAND\_MAX \* (max - min)) + min;

}

float rand\_scale(float s)

{

    float scale = rand\_uniform(1, s);

    if(rand()%2) return scale;

    return 1./scale;

}

float \*\*one\_hot\_encode(float \*a, int n, int k)

{

    int i;

    float \*\*t = calloc(n, sizeof(float\*));

    for(i = 0; i < n; ++i){

        t[i] = calloc(k, sizeof(float));

        int index = (int)a[i];

        t[i][index] = 1;

    }

    return t;

}

Darknet.py

from ctypes import \*

import math

import random

def sample(probs):

    s = sum(probs)

    probs = [a/s for a in probs]

    r = random.uniform(0, 1)

    for i in range(len(probs)):

        r = r - probs[i]

        if r <= 0:

            return i

    return len(probs)-1

def c\_array(ctype, values):

    arr = (ctype\*len(values))()

    arr[:] = values

    return arr

class BOX(Structure):

    \_fields\_ = [("x", c\_float),

                ("y", c\_float),

                ("w", c\_float),

                ("h", c\_float)]

class DETECTION(Structure):

    \_fields\_ = [("bbox", BOX),

                ("classes", c\_int),

                ("prob", POINTER(c\_float)),

                ("mask", POINTER(c\_float)),

                ("objectness", c\_float),

                ("sort\_class", c\_int)]

class IMAGE(Structure):

    \_fields\_ = [("w", c\_int),

                ("h", c\_int),

                ("c", c\_int),

                ("data", POINTER(c\_float))]

class METADATA(Structure):

    \_fields\_ = [("classes", c\_int),

                ("names", POINTER(c\_char\_p))]

#lib = CDLL("/home/pjreddie/documents/darknet/libdarknet.so", RTLD\_GLOBAL)

lib = CDLL("libdarknet.so", RTLD\_GLOBAL)

lib.network\_width.argtypes = [c\_void\_p]

lib.network\_width.restype = c\_int

lib.network\_height.argtypes = [c\_void\_p]

lib.network\_height.restype = c\_int

predict = lib.network\_predict

predict.argtypes = [c\_void\_p, POINTER(c\_float)]

predict.restype = POINTER(c\_float)

set\_gpu = lib.cuda\_set\_device

set\_gpu.argtypes = [c\_int]

make\_image = lib.make\_image

make\_image.argtypes = [c\_int, c\_int, c\_int]

make\_image.restype = IMAGE

get\_network\_boxes = lib.get\_network\_boxes

get\_network\_boxes.argtypes = [c\_void\_p, c\_int, c\_int, c\_float, c\_float, POINTER(c\_int), c\_int, POINTER(c\_int)]

get\_network\_boxes.restype = POINTER(DETECTION)

make\_network\_boxes = lib.make\_network\_boxes

make\_network\_boxes.argtypes = [c\_void\_p]

make\_network\_boxes.restype = POINTER(DETECTION)

free\_detections = lib.free\_detections

free\_detections.argtypes = [POINTER(DETECTION), c\_int]

free\_ptrs = lib.free\_ptrs

free\_ptrs.argtypes = [POINTER(c\_void\_p), c\_int]

network\_predict = lib.network\_predict

network\_predict.argtypes = [c\_void\_p, POINTER(c\_float)]

reset\_rnn = lib.reset\_rnn

reset\_rnn.argtypes = [c\_void\_p]

load\_net = lib.load\_network

load\_net.argtypes = [c\_char\_p, c\_char\_p, c\_int]

load\_net.restype = c\_void\_p

do\_nms\_obj = lib.do\_nms\_obj

do\_nms\_obj.argtypes = [POINTER(DETECTION), c\_int, c\_int, c\_float]

do\_nms\_sort = lib.do\_nms\_sort

do\_nms\_sort.argtypes = [POINTER(DETECTION), c\_int, c\_int, c\_float]

free\_image = lib.free\_image

free\_image.argtypes = [IMAGE]

letterbox\_image = lib.letterbox\_image

letterbox\_image.argtypes = [IMAGE, c\_int, c\_int]

letterbox\_image.restype = IMAGE

load\_meta = lib.get\_metadata

lib.get\_metadata.argtypes = [c\_char\_p]

lib.get\_metadata.restype = METADATA

load\_image = lib.load\_image\_color

load\_image.argtypes = [c\_char\_p, c\_int, c\_int]

load\_image.restype = IMAGE

rgbgr\_image = lib.rgbgr\_image

rgbgr\_image.argtypes = [IMAGE]

predict\_image = lib.network\_predict\_image

predict\_image.argtypes = [c\_void\_p, IMAGE]

predict\_image.restype = POINTER(c\_float)

def classify(net, meta, im):

    out = predict\_image(net, im)

    res = []

    for i in range(meta.classes):

        res.append((meta.names[i], out[i]))

    res = sorted(res, key=lambda x: -x[1])

    return res

def detect(net, meta, image, thresh=.5, hier\_thresh=.5, nms=.45):

    im = load\_image(image, 0, 0)

    num = c\_int(0)

    pnum = pointer(num)

    predict\_image(net, im)

    dets = get\_network\_boxes(net, im.w, im.h, thresh, hier\_thresh, None, 0, pnum)

    num = pnum[0]

    if (nms): do\_nms\_obj(dets, num, meta.classes, nms);

    res = []

    for j in range(num):

        for i in range(meta.classes):

            if dets[j].prob[i] > 0:

                b = dets[j].bbox

                res.append((meta.names[i], dets[j].prob[i], (b.x, b.y, b.w, b.h)))

    res = sorted(res, key=lambda x: -x[1])

    free\_image(im)

    free\_detections(dets, num)

    return res

if \_\_name\_\_ == "\_\_main\_\_":

    #net = load\_net("cfg/densenet201.cfg", "/home/pjreddie/trained/densenet201.weights", 0)

    #im = load\_image("data/wolf.jpg", 0, 0)

    #meta = load\_meta("cfg/imagenet1k.data")

    #r = classify(net, meta, im)

    #print r[:10]

    net = load\_net("cfg/tiny-yolo.cfg", "tiny-yolo.weights", 0)

    meta = load\_meta("cfg/coco.data")

    r = detect(net, meta, "data/dog.jpg")

    print(r)

Activation-kernels.cu

#include "cuda\_runtime.h"

#include "curand.h"

#include "cublas\_v2.h"

extern "C" {

#include "activations.h"

#include "cuda.h"

}

\_\_device\_\_ float lhtan\_activate\_kernel(float x)

{

    if(x < 0) return .001f\*x;

    if(x > 1) return .001f\*(x-1.f) + 1.f;

    return x;

}

\_\_device\_\_ float lhtan\_gradient\_kernel(float x)

{

    if(x > 0 && x < 1) return 1;

    return .001;

}

\_\_device\_\_ float hardtan\_activate\_kernel(float x)

{

    if (x < -1) return -1;

    if (x > 1) return 1;

    return x;

}

\_\_device\_\_ float linear\_activate\_kernel(float x){return x;}

\_\_device\_\_ float logistic\_activate\_kernel(float x){return 1.f/(1.f + expf(-x));}

\_\_device\_\_ float loggy\_activate\_kernel(float x){return 2.f/(1.f + expf(-x)) - 1;}

\_\_device\_\_ float relu\_activate\_kernel(float x){return x\*(x>0);}

\_\_device\_\_ float elu\_activate\_kernel(float x){return (x >= 0)\*x + (x < 0)\*(expf(x)-1);}

\_\_device\_\_ float selu\_activate\_kernel(float x){return (x >= 0)\*1.0507f\*x + (x < 0)\*1.0507f\*1.6732f\*(expf(x)-1);}

\_\_device\_\_ float relie\_activate\_kernel(float x){return (x>0) ? x : .01f\*x;}

\_\_device\_\_ float ramp\_activate\_kernel(float x){return x\*(x>0)+.1f\*x;}

\_\_device\_\_ float leaky\_activate\_kernel(float x){return (x>0) ? x : .1f\*x;}

\_\_device\_\_ float tanh\_activate\_kernel(float x){return (2.f/(1 + expf(-2\*x)) - 1);}

\_\_device\_\_ float plse\_activate\_kernel(float x)

{

    if(x < -4) return .01f \* (x + 4);

    if(x > 4)  return .01f \* (x - 4) + 1;

    return .125f\*x + .5f;

}

\_\_device\_\_ float stair\_activate\_kernel(float x)

{

    int n = floorf(x);

    if (n%2 == 0) return floorf(x/2);

    else return (x - n) + floorf(x/2);

}

\_\_device\_\_ float hardtan\_gradient\_kernel(float x)

{

    if (x > -1 && x < 1) return 1;

    return 0;

}

\_\_device\_\_ float linear\_gradient\_kernel(float x){return 1;}

\_\_device\_\_ float logistic\_gradient\_kernel(float x){return (1-x)\*x;}

\_\_device\_\_ float loggy\_gradient\_kernel(float x)

{

    float y = (x+1)/2;

    return 2\*(1-y)\*y;

}

\_\_device\_\_ float relu\_gradient\_kernel(float x){return (x>0);}

\_\_device\_\_ float elu\_gradient\_kernel(float x){return (x >= 0) + (x < 0)\*(x + 1);}

\_\_device\_\_ float selu\_gradient\_kernel(float x){return (x >= 0)\*1.0507 + (x < 0)\*(x + 1.0507\*1.6732);}

\_\_device\_\_ float relie\_gradient\_kernel(float x){return (x>0) ? 1 : .01f;}

\_\_device\_\_ float ramp\_gradient\_kernel(float x){return (x>0)+.1f;}

\_\_device\_\_ float leaky\_gradient\_kernel(float x){return (x>0) ? 1 : .1f;}

\_\_device\_\_ float tanh\_gradient\_kernel(float x){return 1-x\*x;}

\_\_device\_\_ float plse\_gradient\_kernel(float x){return (x < 0 || x > 1) ? .01f : .125f;}

\_\_device\_\_ float stair\_gradient\_kernel(float x)

{

    if (floorf(x) == x) return 0;

    return 1;

}

\_\_device\_\_ float activate\_kernel(float x, ACTIVATION a)

{

    switch(a){

        case LINEAR:

            return linear\_activate\_kernel(x);

        case LOGISTIC:

            return logistic\_activate\_kernel(x);

        case LOGGY:

            return loggy\_activate\_kernel(x);

        case RELU:

            return relu\_activate\_kernel(x);

        case ELU:

            return elu\_activate\_kernel(x);

        case SELU:

            return selu\_activate\_kernel(x);

        case RELIE:

            return relie\_activate\_kernel(x);

        case RAMP:

            return ramp\_activate\_kernel(x);

        case LEAKY:

            return leaky\_activate\_kernel(x);

        case TANH:

            return tanh\_activate\_kernel(x);

        case PLSE:

            return plse\_activate\_kernel(x);

        case STAIR:

            return stair\_activate\_kernel(x);

        case HARDTAN:

            return hardtan\_activate\_kernel(x);

        case LHTAN:

            return lhtan\_activate\_kernel(x);

    }

    return 0;

}

\_\_device\_\_ float gradient\_kernel(float x, ACTIVATION a)

{

    switch(a){

        case LINEAR:

            return linear\_gradient\_kernel(x);

        case LOGISTIC:

            return logistic\_gradient\_kernel(x);

        case LOGGY:

            return loggy\_gradient\_kernel(x);

        case RELU:

            return relu\_gradient\_kernel(x);

        case ELU:

            return elu\_gradient\_kernel(x);

        case SELU:

            return selu\_gradient\_kernel(x);

        case RELIE:

            return relie\_gradient\_kernel(x);

        case RAMP:

            return ramp\_gradient\_kernel(x);

        case LEAKY:

            return leaky\_gradient\_kernel(x);

        case TANH:

            return tanh\_gradient\_kernel(x);

        case PLSE:

            return plse\_gradient\_kernel(x);

        case STAIR:

            return stair\_gradient\_kernel(x);

        case HARDTAN:

            return hardtan\_gradient\_kernel(x);

        case LHTAN:

            return lhtan\_gradient\_kernel(x);

    }

    return 0;

}

\_\_global\_\_ void binary\_gradient\_array\_kernel(float \*x, float \*dy, int n, int s, BINARY\_ACTIVATION a, float \*dx)

{

    int id = (blockIdx.x + blockIdx.y\*gridDim.x) \* blockDim.x + threadIdx.x;

    int i = id % s;

    int b = id / s;

    float x1 = x[b\*s + i];

    float x2 = x[b\*s + s/2 + i];

    if(id < n) {

        float de = dy[id];

        dx[b\*s + i] = x2\*de;

        dx[b\*s + s/2 + i] = x1\*de;

    }

}

extern "C" void binary\_gradient\_array\_gpu(float \*x, float \*dx, int n, int size, BINARY\_ACTIVATION a, float \*y)

{

    binary\_gradient\_array\_kernel<<<cuda\_gridsize(n/2), BLOCK>>>(x, dx, n/2, size, a, y);

    check\_error(cudaPeekAtLastError());

}

\_\_global\_\_ void binary\_activate\_array\_kernel(float \*x, int n, int s, BINARY\_ACTIVATION a, float \*y)

{

    int id = (blockIdx.x + blockIdx.y\*gridDim.x) \* blockDim.x + threadIdx.x;

    int i = id % s;

    int b = id / s;

    float x1 = x[b\*s + i];

    float x2 = x[b\*s + s/2 + i];

    if(id < n) y[id] = x1\*x2;

}

extern "C" void binary\_activate\_array\_gpu(float \*x, int n, int size, BINARY\_ACTIVATION a, float \*y)

{

    binary\_activate\_array\_kernel<<<cuda\_gridsize(n/2), BLOCK>>>(x, n/2, size, a, y);

    check\_error(cudaPeekAtLastError());

}

\_\_global\_\_ void activate\_array\_kernel(float \*x, int n, ACTIVATION a)

{

    int i = (blockIdx.x + blockIdx.y\*gridDim.x) \* blockDim.x + threadIdx.x;

    if(i < n) x[i] = activate\_kernel(x[i], a);

}

\_\_global\_\_ void gradient\_array\_kernel(float \*x, int n, ACTIVATION a, float \*delta)

{

    int i = (blockIdx.x + blockIdx.y\*gridDim.x) \* blockDim.x + threadIdx.x;

    if(i < n) delta[i] \*= gradient\_kernel(x[i], a);

}

extern "C" void activate\_array\_gpu(float \*x, int n, ACTIVATION a)

{

    activate\_array\_kernel<<<cuda\_gridsize(n), BLOCK>>>(x, n, a);

    check\_error(cudaPeekAtLastError());

}

extern "C" void gradient\_array\_gpu(float \*x, int n, ACTIVATION a, float \*delta)

{

    gradient\_array\_kernel<<<cuda\_gridsize(n), BLOCK>>>(x, n, a, delta);

    check\_error(cudaPeekAtLastError());

}

darknet.h

#ifndef DARKNET\_API

#define DARKNET\_API

#include <stdlib.h>

#include <stdio.h>

#include <string.h>

#include <pthread.h>

#ifdef GPU

    #define BLOCK 512

    #include "cuda\_runtime.h"

    #include "curand.h"

    #include "cublas\_v2.h"

    #ifdef CUDNN

    #include "cudnn.h"

    #endif

#endif

#ifdef \_\_cplusplus

extern "C" {

#endif

#define SECRET\_NUM -1234

extern int gpu\_index;

typedef struct{

    int classes;

    char \*\*names;

} metadata;

metadata get\_metadata(char \*file);

typedef struct{

    int \*leaf;

    int n;

    int \*parent;

    int \*child;

    int \*group;

    char \*\*name;

    int groups;

    int \*group\_size;

    int \*group\_offset;

} tree;

tree \*read\_tree(char \*filename);

typedef enum{

    LOGISTIC, RELU, RELIE, LINEAR, RAMP, TANH, PLSE, LEAKY, ELU, LOGGY, STAIR, HARDTAN, LHTAN, SELU

} ACTIVATION;

typedef enum{

    PNG, BMP, TGA, JPG

} IMTYPE;

typedef enum{

    MULT, ADD, SUB, DIV

} BINARY\_ACTIVATION;

typedef enum {

    CONVOLUTIONAL,

    DECONVOLUTIONAL,

    CONNECTED,

    MAXPOOL,

    SOFTMAX,

    DETECTION,

    DROPOUT,

    CROP,

    ROUTE,

    COST,

    NORMALIZATION,

    AVGPOOL,

    LOCAL,

    SHORTCUT,

    ACTIVE,

    RNN,

    GRU,

    LSTM,

    CRNN,

    BATCHNORM,

    NETWORK,

    XNOR,

    REGION,

    YOLO,

    ISEG,

    REORG,

    UPSAMPLE,

    LOGXENT,

    L2NORM,

    BLANK

} LAYER\_TYPE;

typedef enum{

    SSE, MASKED, L1, SEG, SMOOTH,WGAN

} COST\_TYPE;

typedef struct{

    int batch;

    float learning\_rate;

    float momentum;

    float decay;

    int adam;

    float B1;

    float B2;

    float eps;

    int t;

} update\_args;

struct network;

typedef struct network network;

struct layer;

typedef struct layer layer;

struct layer{

    LAYER\_TYPE type;

    ACTIVATION activation;

    COST\_TYPE cost\_type;

    void (\*forward)   (struct layer, struct network);

    void (\*backward)  (struct layer, struct network);

    void (\*update)    (struct layer, update\_args);

    void (\*forward\_gpu)   (struct layer, struct network);

    void (\*backward\_gpu)  (struct layer, struct network);

    void (\*update\_gpu)    (struct layer, update\_args);

    int batch\_normalize;

    int shortcut;

    int batch;

    int forced;

    int flipped;

    int inputs;

    int outputs;

    int nweights;

    int nbiases;

    int extra;

    int truths;

    int h,w,c;

    int out\_h, out\_w, out\_c;

    int n;

    int max\_boxes;

    int groups;

    int size;

    int side;

    int stride;

    int reverse;

    int flatten;

    int spatial;

    int pad;

    int sqrt;

    int flip;

    int index;

    int binary;

    int xnor;

    int steps;

    int hidden;

    int truth;

    float smooth;

    float dot;

    float angle;

    float jitter;

    float saturation;

    float exposure;

    float shift;

    float ratio;

    float learning\_rate\_scale;

    float clip;

    int noloss;

    int softmax;

    int classes;

    int coords;

    int background;

    int rescore;

    int objectness;

    int joint;

    int noadjust;

    int reorg;

    int log;

    int tanh;

    int \*mask;

    int total;

    float alpha;

    float beta;

    float kappa;

    float coord\_scale;

    float object\_scale;

    float noobject\_scale;

    float mask\_scale;

    float class\_scale;

    int bias\_match;

    int random;

    float ignore\_thresh;

    float truth\_thresh;

    float thresh;

    float focus;

    int classfix;

    int absolute;

    int onlyforward;

    int stopbackward;

    int dontload;

    int dontsave;

    int dontloadscales;

    int numload;

    float temperature;

    float probability;

    float scale;

    char  \* cweights;

    int   \* indexes;

    int   \* input\_layers;

    int   \* input\_sizes;

    int   \* map;

    int   \* counts;

    float \*\* sums;

    float \* rand;

    float \* cost;

    float \* state;

    float \* prev\_state;

    float \* forgot\_state;

    float \* forgot\_delta;

    float \* state\_delta;

    float \* combine\_cpu;

    float \* combine\_delta\_cpu;

    float \* concat;

    float \* concat\_delta;

    float \* binary\_weights;

    float \* biases;

    float \* bias\_updates;

    float \* scales;

    float \* scale\_updates;

    float \* weights;

    float \* weight\_updates;

    float \* delta;

    float \* output;

    float \* loss;

    float \* squared;

    float \* norms;

    float \* spatial\_mean;

    float \* mean;

    float \* variance;

    float \* mean\_delta;

    float \* variance\_delta;

    float \* rolling\_mean;

    float \* rolling\_variance;

    float \* x;

    float \* x\_norm;

    float \* m;

    float \* v;

    float \* bias\_m;

    float \* bias\_v;

    float \* scale\_m;

    float \* scale\_v;

    float \*z\_cpu;

    float \*r\_cpu;

    float \*h\_cpu;

    float \* prev\_state\_cpu;

    float \*temp\_cpu;

    float \*temp2\_cpu;

    float \*temp3\_cpu;

    float \*dh\_cpu;

    float \*hh\_cpu;

    float \*prev\_cell\_cpu;

    float \*cell\_cpu;

    float \*f\_cpu;

    float \*i\_cpu;

    float \*g\_cpu;

    float \*o\_cpu;

    float \*c\_cpu;

    float \*dc\_cpu;

    float \* binary\_input;

    struct layer \*input\_layer;

    struct layer \*self\_layer;

    struct layer \*output\_layer;

    struct layer \*reset\_layer;

    struct layer \*update\_layer;

    struct layer \*state\_layer;

    struct layer \*input\_gate\_layer;

    struct layer \*state\_gate\_layer;

    struct layer \*input\_save\_layer;

    struct layer \*state\_save\_layer;

    struct layer \*input\_state\_layer;

    struct layer \*state\_state\_layer;

    struct layer \*input\_z\_layer;

    struct layer \*state\_z\_layer;

    struct layer \*input\_r\_layer;

    struct layer \*state\_r\_layer;

    struct layer \*input\_h\_layer;

    struct layer \*state\_h\_layer;

    struct layer \*wz;

    struct layer \*uz;

    struct layer \*wr;

    struct layer \*ur;

    struct layer \*wh;

    struct layer \*uh;

    struct layer \*uo;

    struct layer \*wo;

    struct layer \*uf;

    struct layer \*wf;

    struct layer \*ui;

    struct layer \*wi;

    struct layer \*ug;

    struct layer \*wg;

    tree \*softmax\_tree;

    size\_t workspace\_size;

#ifdef GPU

    int \*indexes\_gpu;

    float \*z\_gpu;

    float \*r\_gpu;

    float \*h\_gpu;

    float \*temp\_gpu;

    float \*temp2\_gpu;

    float \*temp3\_gpu;

    float \*dh\_gpu;

    float \*hh\_gpu;

    float \*prev\_cell\_gpu;

    float \*cell\_gpu;

    float \*f\_gpu;

    float \*i\_gpu;

    float \*g\_gpu;

    float \*o\_gpu;

    float \*c\_gpu;

    float \*dc\_gpu;

    float \*m\_gpu;

    float \*v\_gpu;

    float \*bias\_m\_gpu;

    float \*scale\_m\_gpu;

    float \*bias\_v\_gpu;

    float \*scale\_v\_gpu;

    float \* combine\_gpu;

    float \* combine\_delta\_gpu;

    float \* prev\_state\_gpu;

    float \* forgot\_state\_gpu;

    float \* forgot\_delta\_gpu;

    float \* state\_gpu;

    float \* state\_delta\_gpu;

    float \* gate\_gpu;

    float \* gate\_delta\_gpu;

    float \* save\_gpu;

    float \* save\_delta\_gpu;

    float \* concat\_gpu;

    float \* concat\_delta\_gpu;

    float \* binary\_input\_gpu;

    float \* binary\_weights\_gpu;

    float \* mean\_gpu;

    float \* variance\_gpu;

    float \* rolling\_mean\_gpu;

    float \* rolling\_variance\_gpu;

    float \* variance\_delta\_gpu;

    float \* mean\_delta\_gpu;

    float \* x\_gpu;

    float \* x\_norm\_gpu;

    float \* weights\_gpu;

    float \* weight\_updates\_gpu;

    float \* weight\_change\_gpu;

    float \* biases\_gpu;

    float \* bias\_updates\_gpu;

    float \* bias\_change\_gpu;

    float \* scales\_gpu;

    float \* scale\_updates\_gpu;

    float \* scale\_change\_gpu;

    float \* output\_gpu;

    float \* loss\_gpu;

    float \* delta\_gpu;

    float \* rand\_gpu;

    float \* squared\_gpu;

    float \* norms\_gpu;

#ifdef CUDNN

    cudnnTensorDescriptor\_t srcTensorDesc, dstTensorDesc;

    cudnnTensorDescriptor\_t dsrcTensorDesc, ddstTensorDesc;

    cudnnTensorDescriptor\_t normTensorDesc;

    cudnnFilterDescriptor\_t weightDesc;

    cudnnFilterDescriptor\_t dweightDesc;

    cudnnConvolutionDescriptor\_t convDesc;

    cudnnConvolutionFwdAlgo\_t fw\_algo;

    cudnnConvolutionBwdDataAlgo\_t bd\_algo;

    cudnnConvolutionBwdFilterAlgo\_t bf\_algo;

#endif

#endif

};

void free\_layer(layer);

typedef enum {

    CONSTANT, STEP, EXP, POLY, STEPS, SIG, RANDOM

} learning\_rate\_policy;

typedef struct network{

    int n;

    int batch;

    size\_t \*seen;

    int \*t;

    float epoch;

    int subdivisions;

    layer \*layers;

    float \*output;

    learning\_rate\_policy policy;

    float learning\_rate;

    float momentum;

    float decay;

    float gamma;

    float scale;

    float power;

    int time\_steps;

    int step;

    int max\_batches;

    float \*scales;

    int   \*steps;

    int num\_steps;

    int burn\_in;

    int adam;

    float B1;

    float B2;

    float eps;

    int inputs;

    int outputs;

    int truths;

    int notruth;

    int h, w, c;

    int max\_crop;

    int min\_crop;

    float max\_ratio;

    float min\_ratio;

    int center;

    float angle;

    float aspect;

    float exposure;

    float saturation;

    float hue;

    int random;

    int gpu\_index;

    tree \*hierarchy;

    float \*input;

    float \*truth;

    float \*delta;

    float \*workspace;

    int train;

    int index;

    float \*cost;

    float clip;

#ifdef GPU

    float \*input\_gpu;

    float \*truth\_gpu;

    float \*delta\_gpu;

    float \*output\_gpu;

#endif

} network;

typedef struct {

    int w;

    int h;

    float scale;

    float rad;

    float dx;

    float dy;

    float aspect;

} augment\_args;

typedef struct {

    int w;

    int h;

    int c;

    float \*data;

} image;

typedef struct{

    float x, y, w, h;

} box;

typedef struct detection{

    box bbox;

    int classes;

    float \*prob;

    float \*mask;

    float objectness;

    int sort\_class;

} detection;

typedef struct matrix{

    int rows, cols;

    float \*\*vals;

} matrix;

typedef struct{

    int w, h;

    matrix X;

    matrix y;

    int shallow;

    int \*num\_boxes;

    box \*\*boxes;

} data;

typedef enum {

    CLASSIFICATION\_DATA, DETECTION\_DATA, CAPTCHA\_DATA, REGION\_DATA, IMAGE\_DATA, COMPARE\_DATA, WRITING\_DATA, SWAG\_DATA, TAG\_DATA, OLD\_CLASSIFICATION\_DATA, STUDY\_DATA, DET\_DATA, SUPER\_DATA, LETTERBOX\_DATA, REGRESSION\_DATA, SEGMENTATION\_DATA, INSTANCE\_DATA, ISEG\_DATA

} data\_type;

typedef struct load\_args{

    int threads;

    char \*\*paths;

    char \*path;

    int n;

    int m;

    char \*\*labels;

    int h;

    int w;

    int out\_w;

    int out\_h;

    int nh;

    int nw;

    int num\_boxes;

    int min, max, size;

    int classes;

    int background;

    int scale;

    int center;

    int coords;

    float jitter;

    float angle;

    float aspect;

    float saturation;

    float exposure;

    float hue;

    data \*d;

    image \*im;

    image \*resized;

    data\_type type;

    tree \*hierarchy;

} load\_args;

typedef struct{

    int id;

    float x,y,w,h;

    float left, right, top, bottom;

} box\_label;

network \*load\_network(char \*cfg, char \*weights, int clear);

load\_args get\_base\_args(network \*net);

void free\_data(data d);

typedef struct node{

    void \*val;

    struct node \*next;

    struct node \*prev;

} node;

typedef struct list{

    int size;

    node \*front;

    node \*back;

} list;

pthread\_t load\_data(load\_args args);

list \*read\_data\_cfg(char \*filename);

list \*read\_cfg(char \*filename);

unsigned char \*read\_file(char \*filename);

data resize\_data(data orig, int w, int h);

data \*tile\_data(data orig, int divs, int size);

data select\_data(data \*orig, int \*inds);

void forward\_network(network \*net);

void backward\_network(network \*net);

void update\_network(network \*net);

float dot\_cpu(int N, float \*X, int INCX, float \*Y, int INCY);

void axpy\_cpu(int N, float ALPHA, float \*X, int INCX, float \*Y, int INCY);

void copy\_cpu(int N, float \*X, int INCX, float \*Y, int INCY);

void scal\_cpu(int N, float ALPHA, float \*X, int INCX);

void fill\_cpu(int N, float ALPHA, float \* X, int INCX);

void normalize\_cpu(float \*x, float \*mean, float \*variance, int batch, int filters, int spatial);

void softmax(float \*input, int n, float temp, int stride, float \*output);

int best\_3d\_shift\_r(image a, image b, int min, int max);

#ifdef GPU

void axpy\_gpu(int N, float ALPHA, float \* X, int INCX, float \* Y, int INCY);

void fill\_gpu(int N, float ALPHA, float \* X, int INCX);

void scal\_gpu(int N, float ALPHA, float \* X, int INCX);

void copy\_gpu(int N, float \* X, int INCX, float \* Y, int INCY);

void cuda\_set\_device(int n);

void cuda\_free(float \*x\_gpu);

float \*cuda\_make\_array(float \*x, size\_t n);

void cuda\_pull\_array(float \*x\_gpu, float \*x, size\_t n);

float cuda\_mag\_array(float \*x\_gpu, size\_t n);

void cuda\_push\_array(float \*x\_gpu, float \*x, size\_t n);

void forward\_network\_gpu(network \*net);

void backward\_network\_gpu(network \*net);

void update\_network\_gpu(network \*net);

float train\_networks(network \*\*nets, int n, data d, int interval);

void sync\_nets(network \*\*nets, int n, int interval);

void harmless\_update\_network\_gpu(network \*net);

#endif

image get\_label(image \*\*characters, char \*string, int size);

void draw\_label(image a, int r, int c, image label, const float \*rgb);

void save\_image(image im, const char \*name);

void save\_image\_options(image im, const char \*name, IMTYPE f, int quality);

void get\_next\_batch(data d, int n, int offset, float \*X, float \*y);

void grayscale\_image\_3c(image im);

void normalize\_image(image p);

void matrix\_to\_csv(matrix m);

float train\_network\_sgd(network \*net, data d, int n);

void rgbgr\_image(image im);

data copy\_data(data d);

data concat\_data(data d1, data d2);

data load\_cifar10\_data(char \*filename);

float matrix\_topk\_accuracy(matrix truth, matrix guess, int k);

void matrix\_add\_matrix(matrix from, matrix to);

void scale\_matrix(matrix m, float scale);

matrix csv\_to\_matrix(char \*filename);

float \*network\_accuracies(network \*net, data d, int n);

float train\_network\_datum(network \*net);

image make\_random\_image(int w, int h, int c);

void denormalize\_connected\_layer(layer l);

void denormalize\_convolutional\_layer(layer l);

void statistics\_connected\_layer(layer l);

void rescale\_weights(layer l, float scale, float trans);

void rgbgr\_weights(layer l);

image \*get\_weights(layer l);

void demo(char \*cfgfile, char \*weightfile, float thresh, int cam\_index, const char \*filename, char \*\*names, int classes, int frame\_skip, char \*prefix, int avg, float hier\_thresh, int w, int h, int fps, int fullscreen);

void get\_detection\_detections(layer l, int w, int h, float thresh, detection \*dets);

char \*option\_find\_str(list \*l, char \*key, char \*def);

int option\_find\_int(list \*l, char \*key, int def);

int option\_find\_int\_quiet(list \*l, char \*key, int def);

network \*parse\_network\_cfg(char \*filename);

void save\_weights(network \*net, char \*filename);

void load\_weights(network \*net, char \*filename);

void save\_weights\_upto(network \*net, char \*filename, int cutoff);

void load\_weights\_upto(network \*net, char \*filename, int start, int cutoff);

void zero\_objectness(layer l);

void get\_region\_detections(layer l, int w, int h, int netw, int neth, float thresh, int \*map, float tree\_thresh, int relative, detection \*dets);

int get\_yolo\_detections(layer l, int w, int h, int netw, int neth, float thresh, int \*map, int relative, detection \*dets);

void free\_network(network \*net);

void set\_batch\_network(network \*net, int b);

void set\_temp\_network(network \*net, float t);

image load\_image(char \*filename, int w, int h, int c);

image load\_image\_color(char \*filename, int w, int h);

image make\_image(int w, int h, int c);

image resize\_image(image im, int w, int h);

void censor\_image(image im, int dx, int dy, int w, int h);

image letterbox\_image(image im, int w, int h);

image crop\_image(image im, int dx, int dy, int w, int h);

image center\_crop\_image(image im, int w, int h);

image resize\_min(image im, int min);

image resize\_max(image im, int max);

image threshold\_image(image im, float thresh);

image mask\_to\_rgb(image mask);

int resize\_network(network \*net, int w, int h);

void free\_matrix(matrix m);

void test\_resize(char \*filename);

int show\_image(image p, const char \*name, int ms);

image copy\_image(image p);

void draw\_box\_width(image a, int x1, int y1, int x2, int y2, int w, float r, float g, float b);

float get\_current\_rate(network \*net);

void composite\_3d(char \*f1, char \*f2, char \*out, int delta);

data load\_data\_old(char \*\*paths, int n, int m, char \*\*labels, int k, int w, int h);

size\_t get\_current\_batch(network \*net);

void constrain\_image(image im);

image get\_network\_image\_layer(network \*net, int i);

layer get\_network\_output\_layer(network \*net);

void top\_predictions(network \*net, int n, int \*index);

void flip\_image(image a);

image float\_to\_image(int w, int h, int c, float \*data);

void ghost\_image(image source, image dest, int dx, int dy);

float network\_accuracy(network \*net, data d);

void random\_distort\_image(image im, float hue, float saturation, float exposure);

void fill\_image(image m, float s);

image grayscale\_image(image im);

void rotate\_image\_cw(image im, int times);

double what\_time\_is\_it\_now();

image rotate\_image(image m, float rad);

void visualize\_network(network \*net);

float box\_iou(box a, box b);

data load\_all\_cifar10();

box\_label \*read\_boxes(char \*filename, int \*n);

box float\_to\_box(float \*f, int stride);

void draw\_detections(image im, detection \*dets, int num, float thresh, char \*\*names, image \*\*alphabet, int classes);

matrix network\_predict\_data(network \*net, data test);

image \*\*load\_alphabet();

image get\_network\_image(network \*net);

float \*network\_predict(network \*net, float \*input);

int network\_width(network \*net);

int network\_height(network \*net);

float \*network\_predict\_image(network \*net, image im);

void network\_detect(network \*net, image im, float thresh, float hier\_thresh, float nms, detection \*dets);

detection \*get\_network\_boxes(network \*net, int w, int h, float thresh, float hier, int \*map, int relative, int \*num);

void free\_detections(detection \*dets, int n);

void reset\_network\_state(network \*net, int b);

char \*\*get\_labels(char \*filename);

void do\_nms\_obj(detection \*dets, int total, int classes, float thresh);

void do\_nms\_sort(detection \*dets, int total, int classes, float thresh);

matrix make\_matrix(int rows, int cols);

#ifdef OPENCV

void \*open\_video\_stream(const char \*f, int c, int w, int h, int fps);

image get\_image\_from\_stream(void \*p);

void make\_window(char \*name, int w, int h, int fullscreen);

#endif

void free\_image(image m);

float train\_network(network \*net, data d);

pthread\_t load\_data\_in\_thread(load\_args args);

void load\_data\_blocking(load\_args args);

list \*get\_paths(char \*filename);

void hierarchy\_predictions(float \*predictions, int n, tree \*hier, int only\_leaves, int stride);

void change\_leaves(tree \*t, char \*leaf\_list);

int find\_int\_arg(int argc, char \*\*argv, char \*arg, int def);

float find\_float\_arg(int argc, char \*\*argv, char \*arg, float def);

int find\_arg(int argc, char\* argv[], char \*arg);

char \*find\_char\_arg(int argc, char \*\*argv, char \*arg, char \*def);

char \*basecfg(char \*cfgfile);

void find\_replace(char \*str, char \*orig, char \*rep, char \*output);

void free\_ptrs(void \*\*ptrs, int n);

char \*fgetl(FILE \*fp);

void strip(char \*s);

float sec(clock\_t clocks);

void \*\*list\_to\_array(list \*l);

void top\_k(float \*a, int n, int k, int \*index);

int \*read\_map(char \*filename);

void error(const char \*s);

int max\_index(float \*a, int n);

int max\_int\_index(int \*a, int n);

int sample\_array(float \*a, int n);

int \*random\_index\_order(int min, int max);

void free\_list(list \*l);

float mse\_array(float \*a, int n);

float variance\_array(float \*a, int n);

float mag\_array(float \*a, int n);

void scale\_array(float \*a, int n, float s);

float mean\_array(float \*a, int n);

float sum\_array(float \*a, int n);

void normalize\_array(float \*a, int n);

int \*read\_intlist(char \*s, int \*n, int d);

size\_t rand\_size\_t();

float rand\_normal();

float rand\_uniform(float min, float max);

#ifdef \_\_cplusplus

}

#endif

#endif

Yolov3/utils.py

from multiprocessing import Process, Queue, Pipe

import cv2

import time

import random

import colorsys

import numpy as np

import tensorflow as tf

from yolov3.configs import \*

from yolov3.yolov3 import \*

from tensorflow.python.saved\_model import tag\_constants

def load\_yolo\_weights(model, weights\_file):

    tf.keras.backend.clear\_session() # used to reset layer names

    # load Darknet original weights to TensorFlow model

    if YOLO\_TYPE == "yolov3":

        range1 = 75 if not TRAIN\_YOLO\_TINY else 13

        range2 = [58, 66, 74] if not TRAIN\_YOLO\_TINY else [9, 12]

    with open(weights\_file, 'rb') as wf:

        major, minor, revision, seen, \_ = np.fromfile(wf, dtype=np.int32, count=5)

        j = 0

        for i in range(range1):

            if i > 0:

                conv\_layer\_name = 'conv2d\_%d' %i

            else:

                conv\_layer\_name = 'conv2d'

            if j > 0:

                bn\_layer\_name = 'batch\_normalization\_%d' %j

            else:

                bn\_layer\_name = 'batch\_normalization'

            conv\_layer = model.get\_layer(conv\_layer\_name)

            filters = conv\_layer.filters

            k\_size = conv\_layer.kernel\_size[0]

            in\_dim = conv\_layer.input\_shape[-1]

            if i not in range2:

                # darknet weights: [beta, gamma, mean, variance]

                bn\_weights = np.fromfile(wf, dtype=np.float32, count=4 \* filters)

                # tf weights: [gamma, beta, mean, variance]

                bn\_weights = bn\_weights.reshape((4, filters))[[1, 0, 2, 3]]

                bn\_layer = model.get\_layer(bn\_layer\_name)

                j += 1

            else:

                conv\_bias = np.fromfile(wf, dtype=np.float32, count=filters)

            # darknet shape (out\_dim, in\_dim, height, width)

            conv\_shape = (filters, in\_dim, k\_size, k\_size)

            conv\_weights = np.fromfile(wf, dtype=np.float32, count=np.product(conv\_shape))

            # tf shape (height, width, in\_dim, out\_dim)

            conv\_weights = conv\_weights.reshape(conv\_shape).transpose([2, 3, 1, 0])

            if i not in range2:

                conv\_layer.set\_weights([conv\_weights])

                bn\_layer.set\_weights(bn\_weights)

            else:

                conv\_layer.set\_weights([conv\_weights, conv\_bias])

        assert len(wf.read()) == 0, 'failed to read all data'

def Load\_Yolo\_model():

    gpus = tf.config.experimental.list\_physical\_devices('GPU')

    if len(gpus) > 0:

        print(f'GPUs {gpus}')

        try: tf.config.experimental.set\_memory\_growth(gpus[0], True)

        except RuntimeError: pass

    if YOLO\_FRAMEWORK == "tf": # TensorFlow detection

        if YOLO\_TYPE == "yolov3":

            Darknet\_weights = YOLO\_V3\_TINY\_WEIGHTS if TRAIN\_YOLO\_TINY else YOLO\_V3\_WEIGHTS

        if YOLO\_CUSTOM\_WEIGHTS == False:

            yolo = Create\_Yolo(input\_size=YOLO\_INPUT\_SIZE, CLASSES=YOLO\_COCO\_CLASSES)

            load\_yolo\_weights(yolo, Darknet\_weights) # use Darknet weights

        else:

            yolo = Create\_Yolo(input\_size=YOLO\_INPUT\_SIZE, CLASSES=TRAIN\_CLASSES)

            yolo.load\_weights(YOLO\_CUSTOM\_WEIGHTS) # use custom weights

    elif YOLO\_FRAMEWORK == "trt": # TensorRT detection

        saved\_model\_loaded = tf.saved\_model.load(YOLO\_CUSTOM\_WEIGHTS, tags=[tag\_constants.SERVING])

        signature\_keys = list(saved\_model\_loaded.signatures.keys())

        yolo = saved\_model\_loaded.signatures['serving\_default']

    return yolo

def image\_preprocess(image, target\_size, gt\_boxes=None):

    ih, iw    = target\_size

    h,  w, \_  = image.shape

    scale = min(iw/w, ih/h)

    nw, nh  = int(scale \* w), int(scale \* h)

    image\_resized = cv2.resize(image, (nw, nh))

    image\_paded = np.full(shape=[ih, iw, 3], fill\_value=128.0)

    dw, dh = (iw - nw) // 2, (ih-nh) // 2

    image\_paded[dh:nh+dh, dw:nw+dw, :] = image\_resized

    image\_paded = image\_paded / 255.

    if gt\_boxes is None:

        return image\_paded

    else:

        gt\_boxes[:, [0, 2]] = gt\_boxes[:, [0, 2]] \* scale + dw

        gt\_boxes[:, [1, 3]] = gt\_boxes[:, [1, 3]] \* scale + dh

        return image\_paded, gt\_boxes

def draw\_bbox(image, bboxes, CLASSES=YOLO\_COCO\_CLASSES, show\_label=True, show\_confidence = True, Text\_colors=(255,255,0), rectangle\_colors='', tracking=False):

    NUM\_CLASS = read\_class\_names(CLASSES)

    num\_classes = len(NUM\_CLASS)

    image\_h, image\_w, \_ = image.shape

    hsv\_tuples = [(1.0 \* x / num\_classes, 1., 1.) for x in range(num\_classes)]

    #print("hsv\_tuples", hsv\_tuples)

    colors = list(map(lambda x: colorsys.hsv\_to\_rgb(\*x), hsv\_tuples))

    colors = list(map(lambda x: (int(x[0] \* 255), int(x[1] \* 255), int(x[2] \* 255)), colors))

    random.seed(0)

    random.shuffle(colors)

    random.seed(None)

    for i, bbox in enumerate(bboxes):

        coor = np.array(bbox[:4], dtype=np.int32)

        score = bbox[4]

        class\_ind = int(bbox[5])

        bbox\_color = rectangle\_colors if rectangle\_colors != '' else colors[class\_ind]

        bbox\_thick = int(0.6 \* (image\_h + image\_w) / 1000)

        if bbox\_thick < 1: bbox\_thick = 1

        fontScale = 0.75 \* bbox\_thick

        (x1, y1), (x2, y2) = (coor[0], coor[1]), (coor[2], coor[3])

        # put object rectangle

        cv2.rectangle(image, (x1, y1), (x2, y2), bbox\_color, bbox\_thick\*2)

        if show\_label:

            # get text label

            score\_str = " {:.2f}".format(score) if show\_confidence else ""

            if tracking: score\_str = " "+str(score)

            label = "{}".format(NUM\_CLASS[class\_ind]) + score\_str

            # get text size

            (text\_width, text\_height), baseline = cv2.getTextSize(label, cv2.FONT\_HERSHEY\_COMPLEX\_SMALL,

                                                                  fontScale, thickness=bbox\_thick)

            # put filled text rectangle

            cv2.rectangle(image, (x1, y1), (x1 + text\_width, y1 - text\_height - baseline), bbox\_color, thickness=cv2.FILLED)

            # put text above rectangle

            cv2.putText(image, label, (x1, y1-4), cv2.FONT\_HERSHEY\_COMPLEX\_SMALL,

                        fontScale, Text\_colors, bbox\_thick, lineType=cv2.LINE\_AA)

    return image

def bboxes\_iou(boxes1, boxes2):

    boxes1 = np.array(boxes1)

    boxes2 = np.array(boxes2)

    boxes1\_area = (boxes1[..., 2] - boxes1[..., 0]) \* (boxes1[..., 3] - boxes1[..., 1])

    boxes2\_area = (boxes2[..., 2] - boxes2[..., 0]) \* (boxes2[..., 3] - boxes2[..., 1])

    left\_up       = np.maximum(boxes1[..., :2], boxes2[..., :2])

    right\_down    = np.minimum(boxes1[..., 2:], boxes2[..., 2:])

    inter\_section = np.maximum(right\_down - left\_up, 0.0)

    inter\_area    = inter\_section[..., 0] \* inter\_section[..., 1]

    union\_area    = boxes1\_area + boxes2\_area - inter\_area

    ious          = np.maximum(1.0 \* inter\_area / union\_area, np.finfo(np.float32).eps)

    return ious

def nms(bboxes, iou\_threshold, sigma=0.3, method='nms'):

    """

    :param bboxes: (xmin, ymin, xmax, ymax, score, class)

    Note: soft-nms, https://arxiv.org/pdf/1704.04503.pdf

          https://github.com/bharatsingh430/soft-nms

    """

    classes\_in\_img = list(set(bboxes[:, 5]))

    best\_bboxes = []

    for cls in classes\_in\_img:

        cls\_mask = (bboxes[:, 5] == cls)

        cls\_bboxes = bboxes[cls\_mask]

        # Process 1: Determine whether the number of bounding boxes is greater than 0

        while len(cls\_bboxes) > 0:

            # Process 2: Select the bounding box with the highest score according to socre order A

            max\_ind = np.argmax(cls\_bboxes[:, 4])

            best\_bbox = cls\_bboxes[max\_ind]

            best\_bboxes.append(best\_bbox)

            cls\_bboxes = np.concatenate([cls\_bboxes[: max\_ind], cls\_bboxes[max\_ind + 1:]])

            # Process 3: Calculate this bounding box A and

            # Remain all iou of the bounding box and remove those bounding boxes whose iou value is higher than the threshold

            iou = bboxes\_iou(best\_bbox[np.newaxis, :4], cls\_bboxes[:, :4])

            weight = np.ones((len(iou),), dtype=np.float32)

            assert method in ['nms', 'soft-nms']

            if method == 'nms':

                iou\_mask = iou > iou\_threshold

                weight[iou\_mask] = 0.0

            if method == 'soft-nms':

                weight = np.exp(-(1.0 \* iou \*\* 2 / sigma))

            cls\_bboxes[:, 4] = cls\_bboxes[:, 4] \* weight

            score\_mask = cls\_bboxes[:, 4] > 0.

            cls\_bboxes = cls\_bboxes[score\_mask]

    return best\_bboxes

def postprocess\_boxes(pred\_bbox, original\_image, input\_size, score\_threshold):

    valid\_scale=[0, np.inf]

    pred\_bbox = np.array(pred\_bbox)

    pred\_xywh = pred\_bbox[:, 0:4]

    pred\_conf = pred\_bbox[:, 4]

    pred\_prob = pred\_bbox[:, 5:]

    # 1. (x, y, w, h) --> (xmin, ymin, xmax, ymax)

    pred\_coor = np.concatenate([pred\_xywh[:, :2] - pred\_xywh[:, 2:] \* 0.5,

                                pred\_xywh[:, :2] + pred\_xywh[:, 2:] \* 0.5], axis=-1)

    # 2. (xmin, ymin, xmax, ymax) -> (xmin\_org, ymin\_org, xmax\_org, ymax\_org)

    org\_h, org\_w = original\_image.shape[:2]

    resize\_ratio = min(input\_size / org\_w, input\_size / org\_h)

    dw = (input\_size - resize\_ratio \* org\_w) / 2

    dh = (input\_size - resize\_ratio \* org\_h) / 2

    pred\_coor[:, 0::2] = 1.0 \* (pred\_coor[:, 0::2] - dw) / resize\_ratio

    pred\_coor[:, 1::2] = 1.0 \* (pred\_coor[:, 1::2] - dh) / resize\_ratio

    # 3. clip some boxes those are out of range

    pred\_coor = np.concatenate([np.maximum(pred\_coor[:, :2], [0, 0]),

                                np.minimum(pred\_coor[:, 2:], [org\_w - 1, org\_h - 1])], axis=-1)

    invalid\_mask = np.logical\_or((pred\_coor[:, 0] > pred\_coor[:, 2]), (pred\_coor[:, 1] > pred\_coor[:, 3]))

    pred\_coor[invalid\_mask] = 0

    # 4. discard some invalid boxes

    bboxes\_scale = np.sqrt(np.multiply.reduce(pred\_coor[:, 2:4] - pred\_coor[:, 0:2], axis=-1))

    scale\_mask = np.logical\_and((valid\_scale[0] < bboxes\_scale), (bboxes\_scale < valid\_scale[1]))

    # 5. discard boxes with low scores

    classes = np.argmax(pred\_prob, axis=-1)

    scores = pred\_conf \* pred\_prob[np.arange(len(pred\_coor)), classes]

    score\_mask = scores > score\_threshold

    mask = np.logical\_and(scale\_mask, score\_mask)

    coors, scores, classes = pred\_coor[mask], scores[mask], classes[mask]

    return np.concatenate([coors, scores[:, np.newaxis], classes[:, np.newaxis]], axis=-1)

def detect\_image(Yolo, image\_path, output\_path, input\_size=416, show=False, CLASSES=YOLO\_COCO\_CLASSES, score\_threshold=0.3, iou\_threshold=0.45, rectangle\_colors=''):

    original\_image      = cv2.imread(image\_path)

    original\_image      = cv2.cvtColor(original\_image, cv2.COLOR\_BGR2RGB)

    original\_image      = cv2.cvtColor(original\_image, cv2.COLOR\_BGR2RGB)

    image\_data = image\_preprocess(np.copy(original\_image), [input\_size, input\_size])

    image\_data = image\_data[np.newaxis, ...].astype(np.float32)

    if YOLO\_FRAMEWORK == "tf":

        pred\_bbox = Yolo.predict(image\_data)

    elif YOLO\_FRAMEWORK == "trt":

        batched\_input = tf.constant(image\_data)

        result = Yolo(batched\_input)

        pred\_bbox = []

        for key, value in result.items():

            value = value.numpy()

            pred\_bbox.append(value)

    pred\_bbox = [tf.reshape(x, (-1, tf.shape(x)[-1])) for x in pred\_bbox]

    pred\_bbox = tf.concat(pred\_bbox, axis=0)

    bboxes = postprocess\_boxes(pred\_bbox, original\_image, input\_size, score\_threshold)

    bboxes = nms(bboxes, iou\_threshold, method='nms')

    image = draw\_bbox(original\_image, bboxes, CLASSES=CLASSES, rectangle\_colors=rectangle\_colors)

    # CreateXMLfile("XML\_Detections", str(int(time.time())), original\_image, bboxes, read\_class\_names(CLASSES))

    if output\_path != '': cv2.imwrite(output\_path, image)

    if show:

        # Show the image

        cv2.imshow("predicted image", image)

        # Load and hold the image

        cv2.waitKey(0)

        # To close the window after the required kill value was provided

        cv2.destroyAllWindows()

    return image

def Predict\_bbox\_mp(Frames\_data, Predicted\_data, Processing\_times):

    gpus = tf.config.experimental.list\_physical\_devices('GPU')

    if len(gpus) > 0:

        try: tf.config.experimental.set\_memory\_growth(gpus[0], True)

        except RuntimeError: print("RuntimeError in tf.config.experimental.list\_physical\_devices('GPU')")

    Yolo = Load\_Yolo\_model()

    times = []

    while True:

        if Frames\_data.qsize()>0:

            image\_data = Frames\_data.get()

            t1 = time.time()

            Processing\_times.put(time.time())

            if YOLO\_FRAMEWORK == "tf":

                pred\_bbox = Yolo.predict(image\_data)

            elif YOLO\_FRAMEWORK == "trt":

                batched\_input = tf.constant(image\_data)

                result = Yolo(batched\_input)

                pred\_bbox = []

                for key, value in result.items():

                    value = value.numpy()

                    pred\_bbox.append(value)

            pred\_bbox = [tf.reshape(x, (-1, tf.shape(x)[-1])) for x in pred\_bbox]

            pred\_bbox = tf.concat(pred\_bbox, axis=0)

            Predicted\_data.put(pred\_bbox)

def postprocess\_mp(Predicted\_data, original\_frames, Processed\_frames, Processing\_times, input\_size, CLASSES, score\_threshold, iou\_threshold, rectangle\_colors, realtime):

    times = []

    while True:

        if Predicted\_data.qsize()>0:

            pred\_bbox = Predicted\_data.get()

            if realtime:

                while original\_frames.qsize() > 1:

                    original\_image = original\_frames.get()

            else:

                original\_image = original\_frames.get()

            bboxes = postprocess\_boxes(pred\_bbox, original\_image, input\_size, score\_threshold)

            bboxes = nms(bboxes, iou\_threshold, method='nms')

            image = draw\_bbox(original\_image, bboxes, CLASSES=CLASSES, rectangle\_colors=rectangle\_colors)

            times.append(time.time()-Processing\_times.get())

            times = times[-20:]

            ms = sum(times)/len(times)\*1000

            fps = 1000 / ms

            image = cv2.putText(image, "Time: {:.1f}FPS".format(fps), (0, 30), cv2.FONT\_HERSHEY\_COMPLEX\_SMALL, 1, (0, 0, 255), 2)

            #print("Time: {:.2f}ms, Final FPS: {:.1f}".format(ms, fps))

            Processed\_frames.put(image)

def Show\_Image\_mp(Processed\_frames, show, Final\_frames):

    while True:

        if Processed\_frames.qsize()>0:

            image = Processed\_frames.get()

            Final\_frames.put(image)

            if show:

                cv2.imshow('output', image)

                if cv2.waitKey(25) & 0xFF == ord("q"):

                    cv2.destroyAllWindows()

                    break

# detect from webcam

def detect\_video\_realtime\_mp(video\_path, output\_path, input\_size=416, show=False, CLASSES=YOLO\_COCO\_CLASSES, score\_threshold=0.3, iou\_threshold=0.45, rectangle\_colors='', realtime=False):

    if realtime:

        vid = cv2.VideoCapture(0)

    else:

        vid = cv2.VideoCapture(video\_path)

    # by default VideoCapture returns float instead of int

    width = int(vid.get(cv2.CAP\_PROP\_FRAME\_WIDTH))

    height = int(vid.get(cv2.CAP\_PROP\_FRAME\_HEIGHT))

    fps = int(vid.get(cv2.CAP\_PROP\_FPS))

    codec = cv2.VideoWriter\_fourcc(\*'XVID')

    out = cv2.VideoWriter(output\_path, codec, fps, (width, height)) # output\_path must be .mp4

    no\_of\_frames = int(vid.get(cv2.CAP\_PROP\_FRAME\_COUNT))

    original\_frames = Queue()

    Frames\_data = Queue()

    Predicted\_data = Queue()

    Processed\_frames = Queue()

    Processing\_times = Queue()

    Final\_frames = Queue()

    p1 = Process(target=Predict\_bbox\_mp, args=(Frames\_data, Predicted\_data, Processing\_times))

    p2 = Process(target=postprocess\_mp, args=(Predicted\_data, original\_frames, Processed\_frames, Processing\_times, input\_size, CLASSES, score\_threshold, iou\_threshold, rectangle\_colors, realtime))

    p3 = Process(target=Show\_Image\_mp, args=(Processed\_frames, show, Final\_frames))

    p1.start()

    p2.start()

    p3.start()

    while True:

        ret, img = vid.read()

        if not ret:

            break

        original\_image = cv2.cvtColor(img, cv2.COLOR\_BGR2RGB)

        original\_image = cv2.cvtColor(original\_image, cv2.COLOR\_BGR2RGB)

        original\_frames.put(original\_image)

        image\_data = image\_preprocess(np.copy(original\_image), [input\_size, input\_size])

        image\_data = image\_data[np.newaxis, ...].astype(np.float32)

        Frames\_data.put(image\_data)

    while True:

        if original\_frames.qsize() == 0 and Frames\_data.qsize() == 0 and Predicted\_data.qsize() == 0  and Processed\_frames.qsize() == 0  and Processing\_times.qsize() == 0 and Final\_frames.qsize() == 0:

            p1.terminate()

            p2.terminate()

            p3.terminate()

            break

        elif Final\_frames.qsize()>0:

            image = Final\_frames.get()

            if output\_path != '': out.write(image)

    cv2.destroyAllWindows()

def detect\_video(Yolo, video\_path, output\_path, input\_size=416, show=False, CLASSES=YOLO\_COCO\_CLASSES, score\_threshold=0.3, iou\_threshold=0.45, rectangle\_colors=''):

    times, times\_2 = [], []

    vid = cv2.VideoCapture(video\_path)

    # by default VideoCapture returns float instead of int

    width = int(vid.get(cv2.CAP\_PROP\_FRAME\_WIDTH))

    height = int(vid.get(cv2.CAP\_PROP\_FRAME\_HEIGHT))

    fps = int(vid.get(cv2.CAP\_PROP\_FPS))

    codec = cv2.VideoWriter\_fourcc(\*'XVID')

    out = cv2.VideoWriter(output\_path, codec, fps, (width, height)) # output\_path must be .mp4

    while True:

        \_, img = vid.read()

        try:

            original\_image = cv2.cvtColor(img, cv2.COLOR\_BGR2RGB)

            original\_image = cv2.cvtColor(original\_image, cv2.COLOR\_BGR2RGB)

        except:

            break

        image\_data = image\_preprocess(np.copy(original\_image), [input\_size, input\_size])

        image\_data = image\_data[np.newaxis, ...].astype(np.float32)

        t1 = time.time()

        if YOLO\_FRAMEWORK == "tf":

            pred\_bbox = Yolo.predict(image\_data)

        elif YOLO\_FRAMEWORK == "trt":

            batched\_input = tf.constant(image\_data)

            result = Yolo(batched\_input)

            pred\_bbox = []

            for key, value in result.items():

                value = value.numpy()

                pred\_bbox.append(value)

        t2 = time.time()

        pred\_bbox = [tf.reshape(x, (-1, tf.shape(x)[-1])) for x in pred\_bbox]

        pred\_bbox = tf.concat(pred\_bbox, axis=0)

        bboxes = postprocess\_boxes(pred\_bbox, original\_image, input\_size, score\_threshold)

        bboxes = nms(bboxes, iou\_threshold, method='nms')

        image = draw\_bbox(original\_image, bboxes, CLASSES=CLASSES, rectangle\_colors=rectangle\_colors)

        t3 = time.time()

        times.append(t2-t1)

        times\_2.append(t3-t1)

        times = times[-20:]

        times\_2 = times\_2[-20:]

        ms = sum(times)/len(times)\*1000

        fps = 1000 / ms

        fps2 = 1000 / (sum(times\_2)/len(times\_2)\*1000)

        image = cv2.putText(image, "Time: {:.1f}FPS".format(fps), (0, 30), cv2.FONT\_HERSHEY\_COMPLEX\_SMALL, 1, (0, 0, 255), 2)

        # CreateXMLfile("XML\_Detections", str(int(time.time())), original\_image, bboxes, read\_class\_names(CLASSES))

        print("Time: {:.2f}ms, Detection FPS: {:.1f}, total FPS: {:.1f}".format(ms, fps, fps2))

        if output\_path != '': out.write(image)

        if show:

            cv2.imshow('output', image)

            if cv2.waitKey(25) & 0xFF == ord("q"):

                cv2.destroyAllWindows()

                break

    cv2.destroyAllWindows()

# detect from webcam

def detect\_realtime(Yolo, output\_path, input\_size=416, show=False, CLASSES=YOLO\_COCO\_CLASSES, score\_threshold=0.3, iou\_threshold=0.45, rectangle\_colors=''):

    times = []

    vid = cv2.VideoCapture(0)

    # by default VideoCapture returns float instead of int

    width = int(vid.get(cv2.CAP\_PROP\_FRAME\_WIDTH))

    height = int(vid.get(cv2.CAP\_PROP\_FRAME\_HEIGHT))

    fps = int(vid.get(cv2.CAP\_PROP\_FPS))

    codec = cv2.VideoWriter\_fourcc(\*'XVID')

    out = cv2.VideoWriter(output\_path, codec, fps, (width, height)) # output\_path must be .mp4

    while True:

        \_, frame = vid.read()

        try:

            original\_frame = cv2.cvtColor(frame, cv2.COLOR\_BGR2RGB)

            original\_frame = cv2.cvtColor(original\_frame, cv2.COLOR\_BGR2RGB)

        except:

            break

        image\_data = image\_preprocess(np.copy(original\_frame), [input\_size, input\_size])

        image\_data = image\_data[np.newaxis, ...].astype(np.float32)

        t1 = time.time()

        if YOLO\_FRAMEWORK == "tf":

            pred\_bbox = Yolo.predict(image\_data)

        elif YOLO\_FRAMEWORK == "trt":

            batched\_input = tf.constant(image\_data)

            result = Yolo(batched\_input)

            pred\_bbox = []

            for key, value in result.items():

                value = value.numpy()

                pred\_bbox.append(value)

        t2 = time.time()

        pred\_bbox = [tf.reshape(x, (-1, tf.shape(x)[-1])) for x in pred\_bbox]

        pred\_bbox = tf.concat(pred\_bbox, axis=0)

        bboxes = postprocess\_boxes(pred\_bbox, original\_frame, input\_size, score\_threshold)

        bboxes = nms(bboxes, iou\_threshold, method='nms')

        times.append(t2-t1)

        times = times[-20:]

        ms = sum(times)/len(times)\*1000

        fps = 1000 / ms

        print("Time: {:.2f}ms, {:.1f} FPS".format(ms, fps))

        frame = draw\_bbox(original\_frame, bboxes, CLASSES=CLASSES, rectangle\_colors=rectangle\_colors)

        # CreateXMLfile("XML\_Detections", str(int(time.time())), original\_frame, bboxes, read\_class\_names(CLASSES))

        image = cv2.putText(frame, "Time: {:.1f}FPS".format(fps), (0, 30),

                          cv2.FONT\_HERSHEY\_COMPLEX\_SMALL, 1, (0, 0, 255), 2)

        if output\_path != '': out.write(frame)

        if show:

            cv2.imshow('output', frame)

            if cv2.waitKey(25) & 0xFF == ord("q"):

                cv2.destroyAllWindows()

                break

    cv2.destroyAllWindows()

Yolov3/configs.py

# YOLO options

YOLO\_TYPE                   = "yolov3"

YOLO\_FRAMEWORK              = "tf" # "tf" or "trt"

YOLO\_V3\_WEIGHTS             = "model\_data/yolov3.weights"

YOLO\_V3\_TINY\_WEIGHTS        = "model\_data/yolov3-tiny.weights"

YOLO\_TRT\_QUANTIZE\_MODE      = "INT8" # INT8, FP16, FP32

YOLO\_CUSTOM\_WEIGHTS         = False # "checkpoints/yolov3\_custom" # used in evaluate\_mAP.py and custom model detection, if not using leave False

                            # YOLO\_CUSTOM\_WEIGHTS also used with TensorRT and custom model detection

YOLO\_COCO\_CLASSES           = "model\_data/coco/coco.names"

YOLO\_STRIDES                = [8, 16, 32]

YOLO\_IOU\_LOSS\_THRESH        = 0.5

YOLO\_ANCHOR\_PER\_SCALE       = 3

YOLO\_MAX\_BBOX\_PER\_SCALE     = 100

YOLO\_INPUT\_SIZE             = 416

YOLO\_ANCHORS            = [[[10,  13], [16,   30], [33,   23]],

                           [[30,  61], [62,   45], [59,  119]],

                           [[116, 90], [156, 198], [373, 326]]]

# Train options

# TRAIN\_YOLO\_TINY             = False

TRAIN\_YOLO\_TINY             = True # \*

TRAIN\_SAVE\_BEST\_ONLY        = True # saves only best model according validation loss (True recommended)

TRAIN\_SAVE\_CHECKPOINT       = False # saves all best validated checkpoints in training process (may require a lot disk space) (False recommended)

TRAIN\_CLASSES               = "model\_data/custom\_data.names"

TRAIN\_ANNOT\_PATH            = "model\_data/custom\_data\_train.txt"

TRAIN\_LOGDIR                = "log"

TRAIN\_CHECKPOINTS\_FOLDER    = "checkpoints"

TRAIN\_MODEL\_NAME            = f"{YOLO\_TYPE}\_custom"

TRAIN\_LOAD\_IMAGES\_TO\_RAM    = True # With True faster training, but need more RAM

TRAIN\_BATCH\_SIZE            = 4

TRAIN\_INPUT\_SIZE            = 416

TRAIN\_DATA\_AUG              = True

TRAIN\_TRANSFER              = True

TRAIN\_FROM\_CHECKPOINT       = False # "checkpoints/yolov3\_custom"

# TRAIN\_FROM\_CHECKPOINT       = True # "checkpoints/yolov3\_custom"\*

TRAIN\_LR\_INIT               = 1e-4

TRAIN\_LR\_END                = 1e-6

TRAIN\_WARMUP\_EPOCHS         = 2

TRAIN\_EPOCHS                = 100

# TEST options

TEST\_ANNOT\_PATH             = "model\_data/custom\_data\_test.txt"

TEST\_BATCH\_SIZE             = 4

TEST\_INPUT\_SIZE             = 416

TEST\_DATA\_AUG               = False

TEST\_DECTECTED\_IMAGE\_PATH   = ""

TEST\_SCORE\_THRESHOLD        = 0.3

# TEST\_IOU\_THRESHOLD          = 0.45

TEST\_IOU\_THRESHOLD          = 0.45 # \*

#YOLOv3-TINY WORKAROUND

if TRAIN\_YOLO\_TINY:

    YOLO\_STRIDES            = [16, 32, 64]

    YOLO\_ANCHORS            = [[[10,  14], [23,   27], [37,   58]],

                               [[81,  82], [135, 169], [344, 319]],

                               [[0,    0], [0,     0], [0,     0]]]

Yolov3/dataset.py

import os

import cv2

import random

import numpy as np

import tensorflow as tf

from yolov3.utils import read\_class\_names, image\_preprocess

from yolov3.yolov3 import bbox\_iou

from yolov3.configs import \*

class Dataset(object):

    # Dataset preprocess implementation

    def \_\_init\_\_(self, dataset\_type, TEST\_INPUT\_SIZE=TEST\_INPUT\_SIZE):

        self.annot\_path  = TRAIN\_ANNOT\_PATH if dataset\_type == 'train' else TEST\_ANNOT\_PATH

        self.input\_sizes = TRAIN\_INPUT\_SIZE if dataset\_type == 'train' else TEST\_INPUT\_SIZE

        self.batch\_size  = TRAIN\_BATCH\_SIZE if dataset\_type == 'train' else TEST\_BATCH\_SIZE

        self.data\_aug    = TRAIN\_DATA\_AUG   if dataset\_type == 'train' else TEST\_DATA\_AUG

        self.train\_input\_sizes = TRAIN\_INPUT\_SIZE

        self.strides = np.array(YOLO\_STRIDES)

        self.classes = read\_class\_names(TRAIN\_CLASSES)

        self.num\_classes = len(self.classes)

        self.anchors = (np.array(YOLO\_ANCHORS).T/self.strides).T

        self.anchor\_per\_scale = YOLO\_ANCHOR\_PER\_SCALE

        self.max\_bbox\_per\_scale = YOLO\_MAX\_BBOX\_PER\_SCALE

        self.annotations = self.load\_annotations(dataset\_type)

        self.num\_samples = len(self.annotations)

        self.num\_batchs = int(np.ceil(self.num\_samples / self.batch\_size))

        self.batch\_count = 0

    def load\_annotations(self, dataset\_type):

        final\_annotations = []

        with open(self.annot\_path, 'r') as f:

            txt = f.readlines()

            annotations = [line.strip() for line in txt if len(line.strip().split()[1:]) != 0]

        np.random.shuffle(annotations)

        for annotation in annotations:

            # fully parse annotations

            line = annotation.split()

            image\_path, index = "", 1

            for i, one\_line in enumerate(line):

                if not one\_line.replace(",","").isnumeric():

                    if image\_path != "": image\_path += " "

                    image\_path += one\_line

                else:

                    index = i

                    break

            if not os.path.exists(image\_path):

                raise KeyError("%s does not exist ... " %image\_path)

            if TRAIN\_LOAD\_IMAGES\_TO\_RAM:

                image = cv2.imread(image\_path)

            else:

                image = ''

            final\_annotations.append([image\_path, line[index:], image])

        return final\_annotations

    def \_\_iter\_\_(self):

        return self

    def Delete\_bad\_annotation(self, bad\_annotation):

        print(f'Deleting {bad\_annotation} annotation line')

        bad\_image\_path = bad\_annotation[0]

        bad\_image\_name = bad\_annotation[0].split('/')[-1] # can be used to delete bad image

        bad\_xml\_path = bad\_annotation[0][:-3]+'xml' # can be used to delete bad xml file

        # remove bad annotation line from annotation file

        with open(self.annot\_path, "r+") as f:

            d = f.readlines()

            f.seek(0)

            for i in d:

                if bad\_image\_name not in i:

                    f.write(i)

            f.truncate()

    def \_\_next\_\_(self):

        with tf.device('/cpu:0'):

            self.train\_input\_size = random.choice([self.train\_input\_sizes])

            self.train\_output\_sizes = self.train\_input\_size // self.strides

            batch\_image = np.zeros((self.batch\_size, self.train\_input\_size, self.train\_input\_size, 3), dtype=np.float32)

            batch\_label\_sbbox = np.zeros((self.batch\_size, self.train\_output\_sizes[0], self.train\_output\_sizes[0],

                                          self.anchor\_per\_scale, 5 + self.num\_classes), dtype=np.float32)

            batch\_label\_mbbox = np.zeros((self.batch\_size, self.train\_output\_sizes[1], self.train\_output\_sizes[1],

                                          self.anchor\_per\_scale, 5 + self.num\_classes), dtype=np.float32)

            batch\_label\_lbbox = np.zeros((self.batch\_size, self.train\_output\_sizes[2], self.train\_output\_sizes[2],

                                          self.anchor\_per\_scale, 5 + self.num\_classes), dtype=np.float32)

            batch\_sbboxes = np.zeros((self.batch\_size, self.max\_bbox\_per\_scale, 4), dtype=np.float32)

            batch\_mbboxes = np.zeros((self.batch\_size, self.max\_bbox\_per\_scale, 4), dtype=np.float32)

            batch\_lbboxes = np.zeros((self.batch\_size, self.max\_bbox\_per\_scale, 4), dtype=np.float32)

            exceptions = False

            num = 0

            if self.batch\_count < self.num\_batchs:

                while num < self.batch\_size:

                    index = self.batch\_count \* self.batch\_size + num

                    if index >= self.num\_samples: index -= self.num\_samples

                    annotation = self.annotations[index]

                    image, bboxes = self.parse\_annotation(annotation)

                    try:

                        label\_sbbox, label\_mbbox, label\_lbbox, sbboxes, mbboxes, lbboxes = self.preprocess\_true\_boxes(bboxes)

                    except IndexError:

                        exceptions = True

                        self.Delete\_bad\_annotation(annotation)

                        print("IndexError, something wrong with", annotation[0], "removed this line from annotation file")

                    batch\_image[num, :, :, :] = image

                    batch\_label\_sbbox[num, :, :, :, :] = label\_sbbox

                    batch\_label\_mbbox[num, :, :, :, :] = label\_mbbox

                    batch\_label\_lbbox[num, :, :, :, :] = label\_lbbox

                    batch\_sbboxes[num, :, :] = sbboxes

                    batch\_mbboxes[num, :, :] = mbboxes

                    batch\_lbboxes[num, :, :] = lbboxes

                    num += 1

                if exceptions:

                    print('\n')

                    raise Exception("There were problems with dataset, I fixed them, now restart the training process.")

                self.batch\_count += 1

                batch\_smaller\_target = batch\_label\_sbbox, batch\_sbboxes

                batch\_medium\_target  = batch\_label\_mbbox, batch\_mbboxes

                batch\_larger\_target  = batch\_label\_lbbox, batch\_lbboxes

                return batch\_image, (batch\_smaller\_target, batch\_medium\_target, batch\_larger\_target)

            else:

                self.batch\_count = 0

                np.random.shuffle(self.annotations)

                raise StopIteration

    def random\_horizontal\_flip(self, image, bboxes):

        if random.random() < 0.5:

            \_, w, \_ = image.shape

            image = image[:, ::-1, :]

            bboxes[:, [0,2]] = w - bboxes[:, [2,0]]

        return image, bboxes

    def random\_crop(self, image, bboxes):

        if random.random() < 0.5:

            h, w, \_ = image.shape

            max\_bbox = np.concatenate([np.min(bboxes[:, 0:2], axis=0), np.max(bboxes[:, 2:4], axis=0)], axis=-1)

            max\_l\_trans = max\_bbox[0]

            max\_u\_trans = max\_bbox[1]

            max\_r\_trans = w - max\_bbox[2]

            max\_d\_trans = h - max\_bbox[3]

            crop\_xmin = max(0, int(max\_bbox[0] - random.uniform(0, max\_l\_trans)))

            crop\_ymin = max(0, int(max\_bbox[1] - random.uniform(0, max\_u\_trans)))

            crop\_xmax = max(w, int(max\_bbox[2] + random.uniform(0, max\_r\_trans)))

            crop\_ymax = max(h, int(max\_bbox[3] + random.uniform(0, max\_d\_trans)))

            image = image[crop\_ymin : crop\_ymax, crop\_xmin : crop\_xmax]

            bboxes[:, [0, 2]] = bboxes[:, [0, 2]] - crop\_xmin

            bboxes[:, [1, 3]] = bboxes[:, [1, 3]] - crop\_ymin

        return image, bboxes

    def random\_translate(self, image, bboxes):

        if random.random() < 0.5:

            h, w, \_ = image.shape

            max\_bbox = np.concatenate([np.min(bboxes[:, 0:2], axis=0), np.max(bboxes[:, 2:4], axis=0)], axis=-1)

            max\_l\_trans = max\_bbox[0]

            max\_u\_trans = max\_bbox[1]

            max\_r\_trans = w - max\_bbox[2]

            max\_d\_trans = h - max\_bbox[3]

            tx = random.uniform(-(max\_l\_trans - 1), (max\_r\_trans - 1))

            ty = random.uniform(-(max\_u\_trans - 1), (max\_d\_trans - 1))

            M = np.array([[1, 0, tx], [0, 1, ty]])

            image = cv2.warpAffine(image, M, (w, h))

            bboxes[:, [0, 2]] = bboxes[:, [0, 2]] + tx

            bboxes[:, [1, 3]] = bboxes[:, [1, 3]] + ty

        return image, bboxes

    def parse\_annotation(self, annotation, mAP = 'False'):

        if TRAIN\_LOAD\_IMAGES\_TO\_RAM:

            image\_path = annotation[0]

            image = annotation[2]

        else:

            image\_path = annotation[0]

            image = cv2.imread(image\_path)

        bboxes = np.array([list(map(int, box.split(','))) for box in annotation[1]])

        if self.data\_aug:

            image, bboxes = self.random\_horizontal\_flip(np.copy(image), np.copy(bboxes))

            image, bboxes = self.random\_crop(np.copy(image), np.copy(bboxes))

            image, bboxes = self.random\_translate(np.copy(image), np.copy(bboxes))

        #image = cv2.cvtColor(image, cv2.COLOR\_BGR2RGB)

        if mAP == True:

            return image, bboxes

        image, bboxes = image\_preprocess(np.copy(image), [self.input\_sizes, self.input\_sizes], np.copy(bboxes))

        return image, bboxes

    def preprocess\_true\_boxes(self, bboxes):

        label = [np.zeros((self.train\_output\_sizes[i], self.train\_output\_sizes[i], self.anchor\_per\_scale,

                           5 + self.num\_classes)) for i in range(3)]

        bboxes\_xywh = [np.zeros((self.max\_bbox\_per\_scale, 4)) for \_ in range(3)]

        bbox\_count = np.zeros((3,))

        for bbox in bboxes:

            bbox\_coor = bbox[:4]

            bbox\_class\_ind = bbox[4]

            onehot = np.zeros(self.num\_classes, dtype=np.float)

            onehot[bbox\_class\_ind] = 1.0

            uniform\_distribution = np.full(self.num\_classes, 1.0 / self.num\_classes)

            deta = 0.01

            smooth\_onehot = onehot \* (1 - deta) + deta \* uniform\_distribution

            bbox\_xywh = np.concatenate([(bbox\_coor[2:] + bbox\_coor[:2]) \* 0.5, bbox\_coor[2:] - bbox\_coor[:2]], axis=-1)

            bbox\_xywh\_scaled = 1.0 \* bbox\_xywh[np.newaxis, :] / self.strides[:, np.newaxis]

            iou = []

            exist\_positive = False

            for i in range(3):

                anchors\_xywh = np.zeros((self.anchor\_per\_scale, 4))

                anchors\_xywh[:, 0:2] = np.floor(bbox\_xywh\_scaled[i, 0:2]).astype(np.int32) + 0.5

                anchors\_xywh[:, 2:4] = self.anchors[i]

                iou\_scale = bbox\_iou(bbox\_xywh\_scaled[i][np.newaxis, :], anchors\_xywh)

                iou.append(iou\_scale)

                iou\_mask = iou\_scale > 0.3

                if np.any(iou\_mask):

                    xind, yind = np.floor(bbox\_xywh\_scaled[i, 0:2]).astype(np.int32)

                    label[i][yind, xind, iou\_mask, :] = 0

                    label[i][yind, xind, iou\_mask, 0:4] = bbox\_xywh

                    label[i][yind, xind, iou\_mask, 4:5] = 1.0

                    label[i][yind, xind, iou\_mask, 5:] = smooth\_onehot

                    bbox\_ind = int(bbox\_count[i] % self.max\_bbox\_per\_scale)

                    bboxes\_xywh[i][bbox\_ind, :4] = bbox\_xywh

                    bbox\_count[i] += 1

                    exist\_positive = True

            if not exist\_positive:

                best\_anchor\_ind = np.argmax(np.array(iou).reshape(-1), axis=-1)

                best\_detect = int(best\_anchor\_ind / self.anchor\_per\_scale)

                best\_anchor = int(best\_anchor\_ind % self.anchor\_per\_scale)

                xind, yind = np.floor(bbox\_xywh\_scaled[best\_detect, 0:2]).astype(np.int32)

                label[best\_detect][yind, xind, best\_anchor, :] = 0

                label[best\_detect][yind, xind, best\_anchor, 0:4] = bbox\_xywh

                label[best\_detect][yind, xind, best\_anchor, 4:5] = 1.0

                label[best\_detect][yind, xind, best\_anchor, 5:] = smooth\_onehot

                bbox\_ind = int(bbox\_count[best\_detect] % self.max\_bbox\_per\_scale)

                bboxes\_xywh[best\_detect][bbox\_ind, :4] = bbox\_xywh

                bbox\_count[best\_detect] += 1

        label\_sbbox, label\_mbbox, label\_lbbox = label

        sbboxes, mbboxes, lbboxes = bboxes\_xywh

        return label\_sbbox, label\_mbbox, label\_lbbox, sbboxes, mbboxes, lbboxes

    def \_\_len\_\_(self):

        return self.num\_batchs

Realtime\_object\_tracking.py

from deep\_sort import generate\_detections as gdet

from deep\_sort.tracker import Tracker

from deep\_sort.detection import Detection

from deep\_sort import nn\_matching

import time

from yolov3.configs import \*

from yolov3.utils import Load\_Yolo\_model, image\_preprocess, postprocess\_boxes, nms, draw\_bbox, read\_class\_names

import tensorflow as tf

import numpy as np

import cv2

import os

os.environ['CUDA\_VISIBLE\_DEVICES'] = '0'

def Object\_tracking(Yolo, input\_size=416, show=False, CLASSES=YOLO\_COCO\_CLASSES, score\_threshold=0.3, iou\_threshold=0.45, rectangle\_colors='', Track\_only=[]):

    # Definition of the parameters

    max\_cosine\_distance = 0.7

    nn\_budget = None

    #initialize deep sort object

    model\_filename = 'model\_data/mars-small128.pb'

    encoder = gdet.create\_box\_encoder(model\_filename, batch\_size=1)

    metric = nn\_matching.NearestNeighborDistanceMetric(

        "cosine", max\_cosine\_distance, nn\_budget)

    tracker = Tracker(metric)

    times, times\_2 = [], []

    # vid = cv2.VideoCapture(0)  # detect from webcam

    vid = cv2.VideoCapture('https://192.168.137.3:8080/video') # detect from mobile feed

    # vid = cv2.VideoCapture('https://10.137.131.218:8080/video')

    # by default VideoCapture returns float instead of int

    width = int(vid.get(cv2.CAP\_PROP\_FRAME\_WIDTH))

    height = int(vid.get(cv2.CAP\_PROP\_FRAME\_HEIGHT))

    fps = int(vid.get(cv2.CAP\_PROP\_FPS))

    # codec = cv2.VideoWriter\_fourcc(\*'XVID')

    # # output\_path must be .mp4

    # out = cv2.VideoWriter(output\_path, codec, fps, (width, height))

    NUM\_CLASS = read\_class\_names(CLASSES)

    key\_list = list(NUM\_CLASS.keys())

    val\_list = list(NUM\_CLASS.values())

    skip=1

    while True:

        \_, frame = vid.read()

        if skip==1:

            skip=2

        elif skip==2:

            skip=3

            continue

        elif skip==3:

            skip=4

            continue

        elif skip==4:

            skip=5

            continue

        elif skip==5:

            skip=6

            continue

        elif skip==6:

            skip=7

            continue

        else:

            skip=1

            continue

        try:

            original\_frame = cv2.cvtColor(frame, cv2.COLOR\_BGR2RGB)

            # original\_frame = cv2.cvtColor(frame, cv2.COLOR\_BGR2GRAY)

            original\_frame = cv2.cvtColor(original\_frame, cv2.COLOR\_BGR2RGB)

            width = 416

            height = 416

            dim = (width, height)

            original\_frame = cv2.resize(original\_frame, dim, interpolation=cv2.INTER\_AREA)

        except:

            break

        image\_data = image\_preprocess(np.copy(original\_frame), [

                                      input\_size, input\_size])

        #image\_data = tf.expand\_dims(image\_data, 0)

        image\_data = image\_data[np.newaxis, ...].astype(np.float32)

        t1 = time.time()

        if YOLO\_FRAMEWORK == "tf":

            pred\_bbox = Yolo.predict(image\_data)

        elif YOLO\_FRAMEWORK == "trt":

            batched\_input = tf.constant(image\_data)

            result = Yolo(batched\_input)

            pred\_bbox = []

            for key, value in result.items():

                value = value.numpy()

                pred\_bbox.append(value)

        # t1 = time.time()

        # pred\_bbox = Yolo.predict(image\_data)

        t2 = time.time()

        pred\_bbox = [tf.reshape(x, (-1, tf.shape(x)[-1])) for x in pred\_bbox]

        pred\_bbox = tf.concat(pred\_bbox, axis=0)

        bboxes = postprocess\_boxes(

            pred\_bbox, original\_frame, input\_size, score\_threshold)

        bboxes = nms(bboxes, iou\_threshold, method='nms')

        # extract bboxes to boxes (x, y, width, height), scores and names

        boxes, scores, names = [], [], []

        for bbox in bboxes:

            if len(Track\_only) != 0 and NUM\_CLASS[int(bbox[5])] in Track\_only or len(Track\_only) == 0:

                boxes.append([bbox[0].astype(int), bbox[1].astype(int), bbox[2].astype(

                    int)-bbox[0].astype(int), bbox[3].astype(int)-bbox[1].astype(int)])

                scores.append(bbox[4])

                names.append(NUM\_CLASS[int(bbox[5])])

        # Obtain all the detections for the given frame.

        boxes = np.array(boxes)

        names = np.array(names)

        scores = np.array(scores)

        features = np.array(encoder(original\_frame, boxes))

        detections = [Detection(bbox, score, class\_name, feature) for bbox,

                      score, class\_name, feature in zip(boxes, scores, names, features)]

        # Pass detections to the deepsort object and obtain the track information.

        tracker.predict()

        tracker.update(detections)

        # Obtain info from the tracks

        tracked\_bboxes = []

        for track in tracker.tracks:

            if not track.is\_confirmed() or track.time\_since\_update > 5:

                continue

            bbox = track.to\_tlbr()  # Get the corrected/predicted bounding box

            class\_name = track.get\_class()  # Get the class name of particular object

            tracking\_id = track.track\_id  # Get the ID for the particular track

            # Get predicted object index by object name

            index = key\_list[val\_list.index(class\_name)]

            # Structure data, that we could use it with our draw\_bbox function

            tracked\_bboxes.append(bbox.tolist() + [tracking\_id, index])

        # draw detection on frame

        image = draw\_bbox(original\_frame, tracked\_bboxes,

                          CLASSES=CLASSES, tracking=True)

        t3 = time.time()

        times.append(t2-t1)

        times\_2.append(t3-t1)

        times = times[-20:]

        times\_2 = times\_2[-20:]

        ms = sum(times)/len(times)\*1000

        fps = 1000 / ms

        fps2 = 1000 / (sum(times\_2)/len(times\_2)\*1000)

        image = cv2.putText(image, "Time: {:.1f} FPS".format(

            fps), (0, 30), cv2.FONT\_HERSHEY\_COMPLEX\_SMALL, 1, (0, 0, 255), 2)

        # draw original yolo detection

        # image = draw\_bbox(image, bboxes, CLASSES=CLASSES, show\_label=False, rectangle\_colors=rectangle\_colors, tracking=True)

        print("Time: {:.2f}ms, Detection FPS: {:.1f}, total FPS: {:.1f}".format(

            ms, fps, fps2))

        # if output\_path != '':

        #     out.write(image)

        cv2.imshow('output', image)

        if cv2.waitKey(25) & 0xFF == ord("q"):

            cv2.destroyAllWindows()

            break

    cv2.destroyAllWindows()

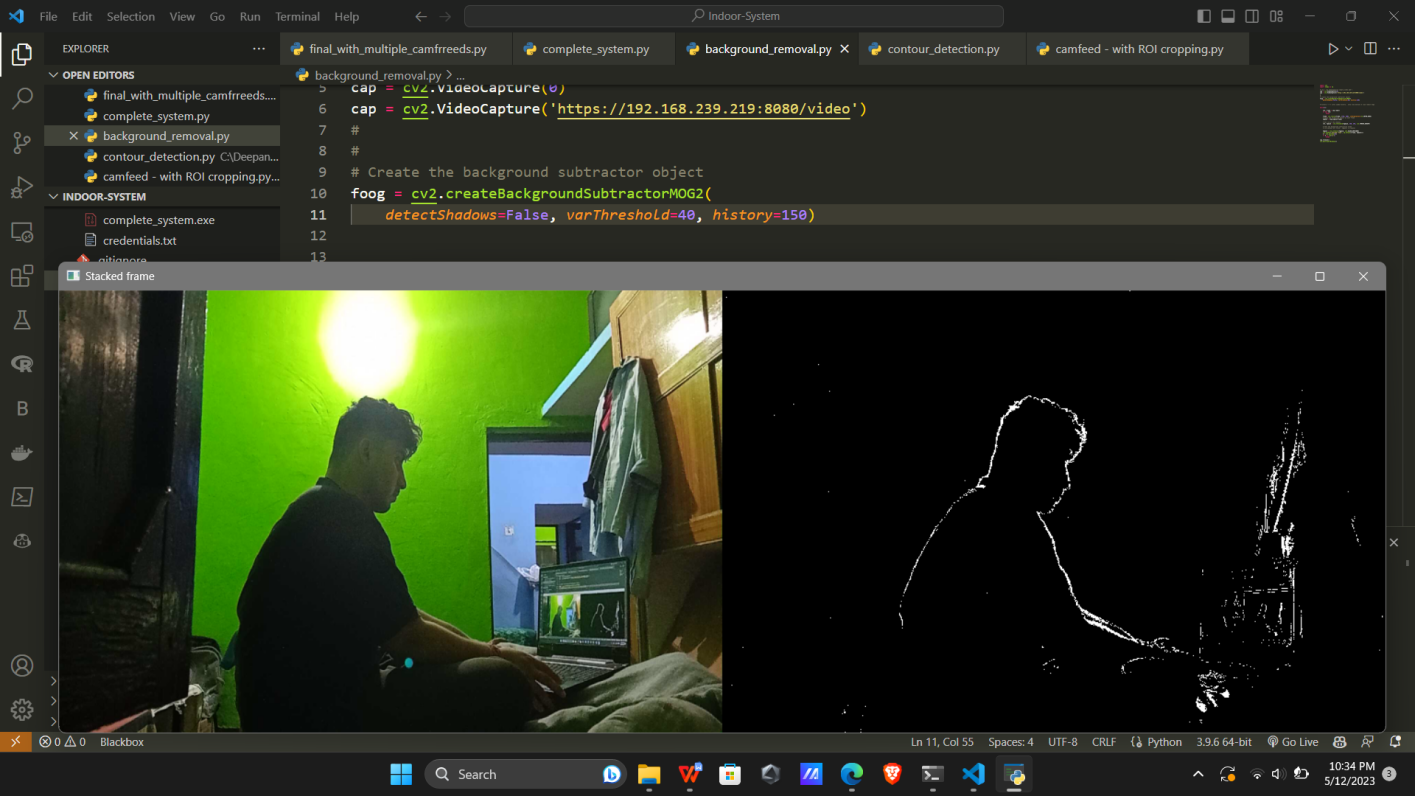
yolo = Load\_Yolo\_model()

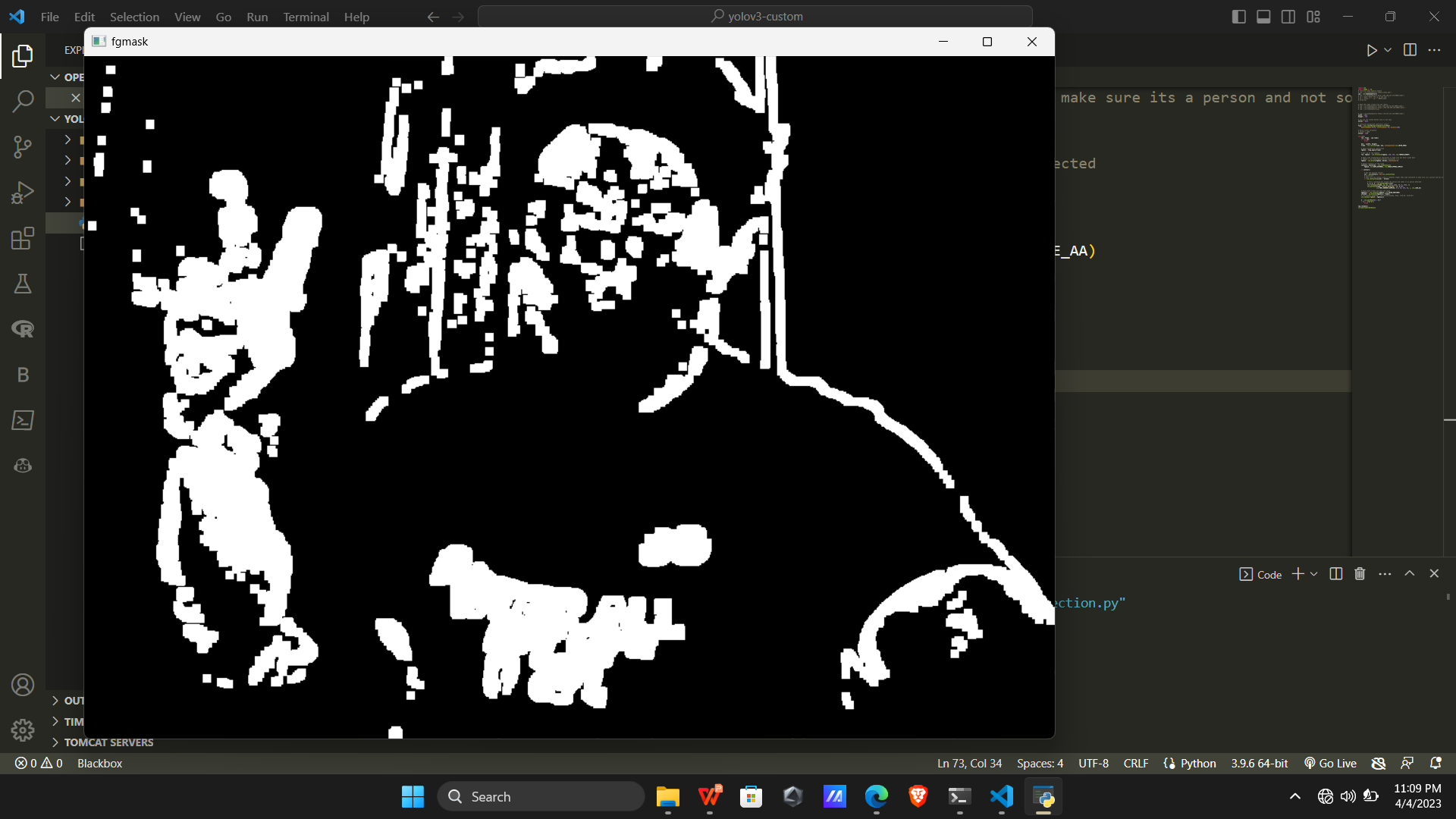
# Object\_tracking(yolo, video\_path, "track.mp4", input\_size=YOLO\_INPUT\_SIZE, show=False, iou\_threshold=0.1, rectangle\_colors=(255,0,0), Track\_only = ["person"])

Object\_tracking(yolo, input\_size=YOLO\_INPUT\_SIZE, show=True,

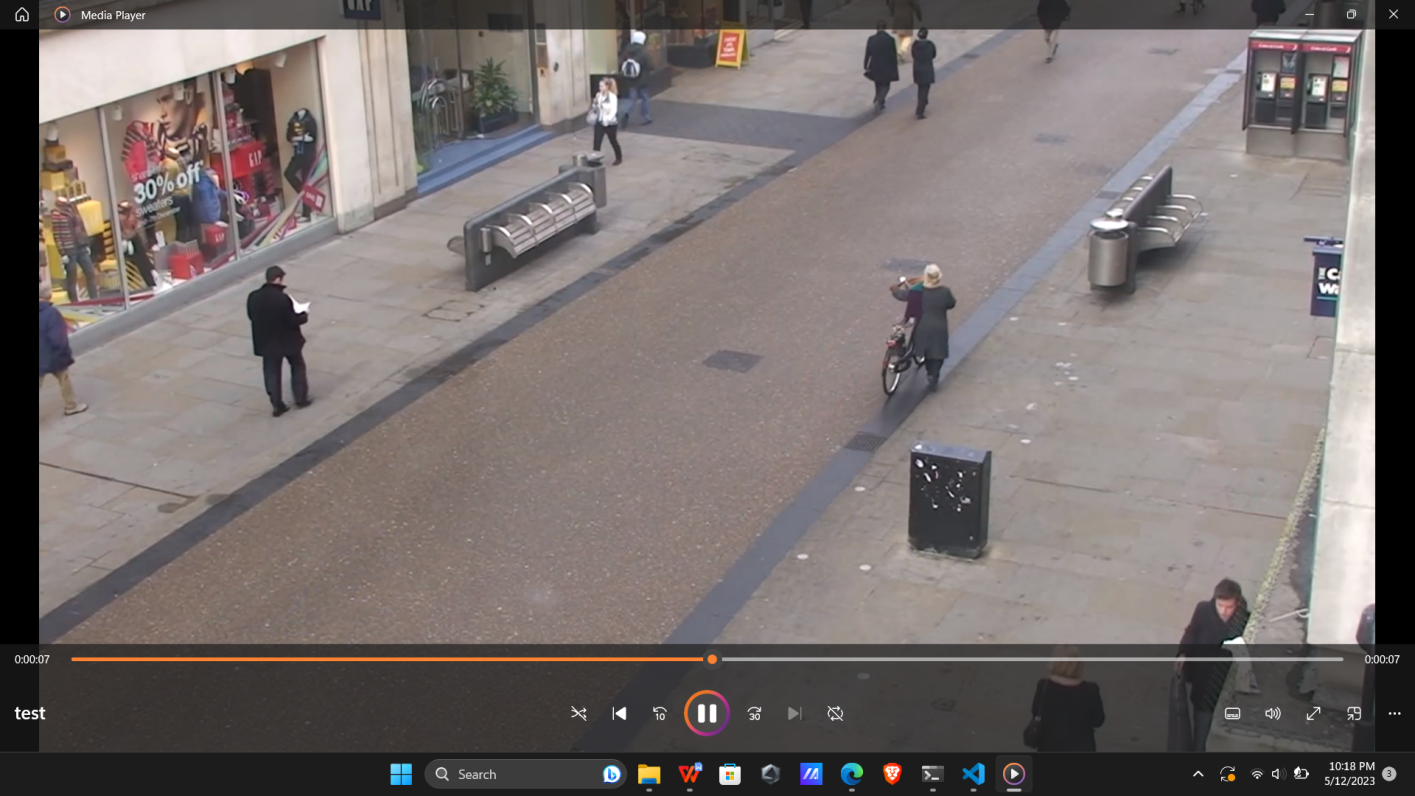
                iou\_threshold=0.1, rectangle\_colors=(255, 0, 0), Track\_only=["person", "chair"])

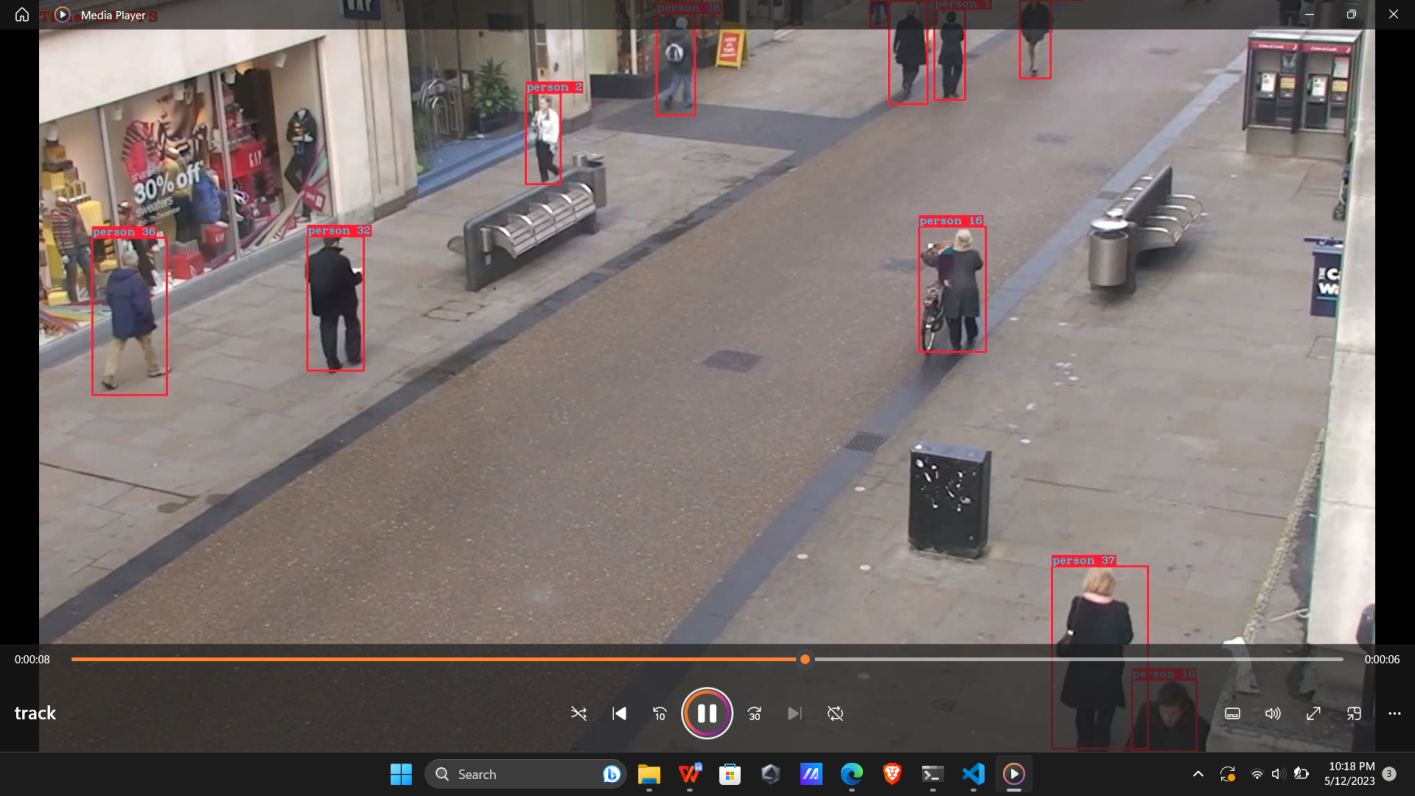
**6.3 Test Data:** A set of sample videos and images used to test the system's object detection and tracking capabilities.

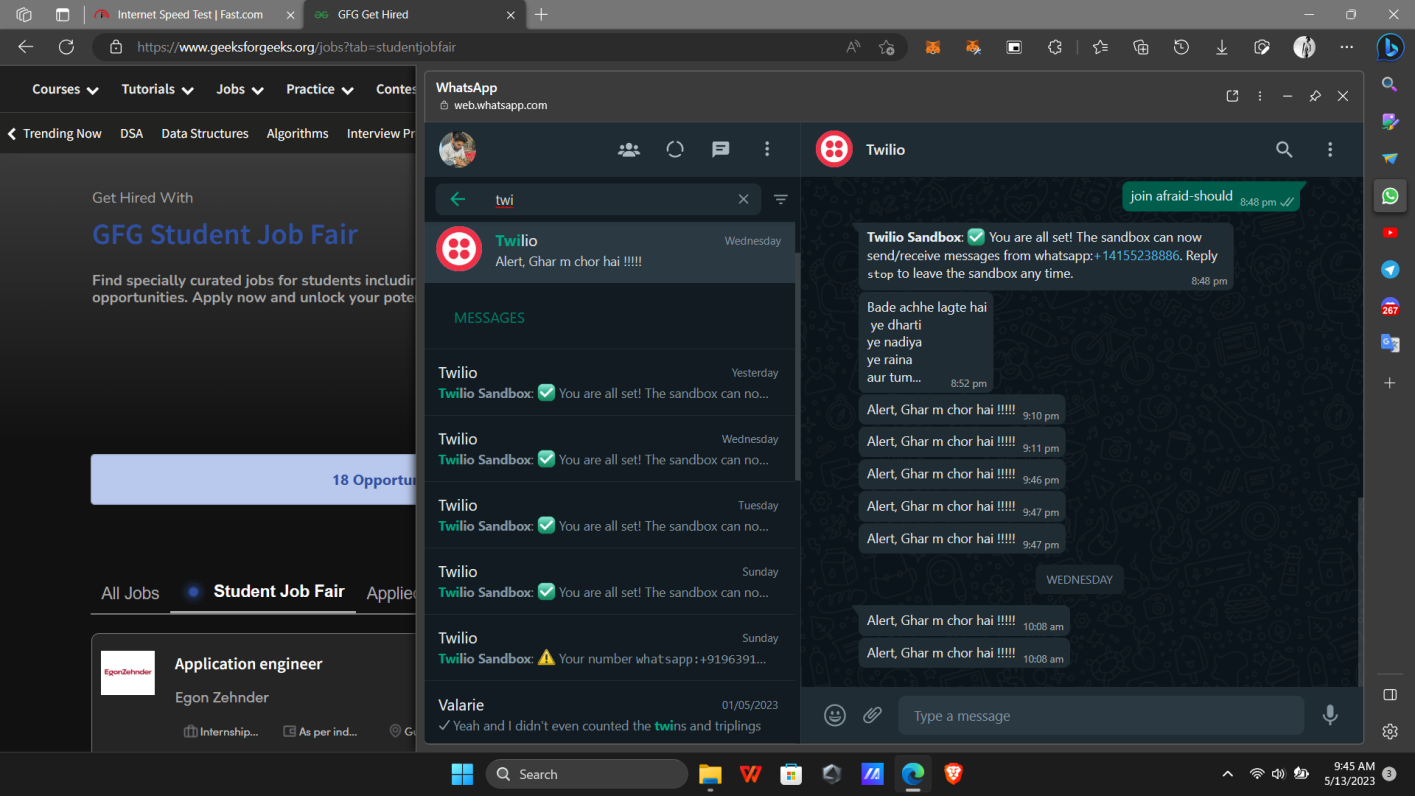


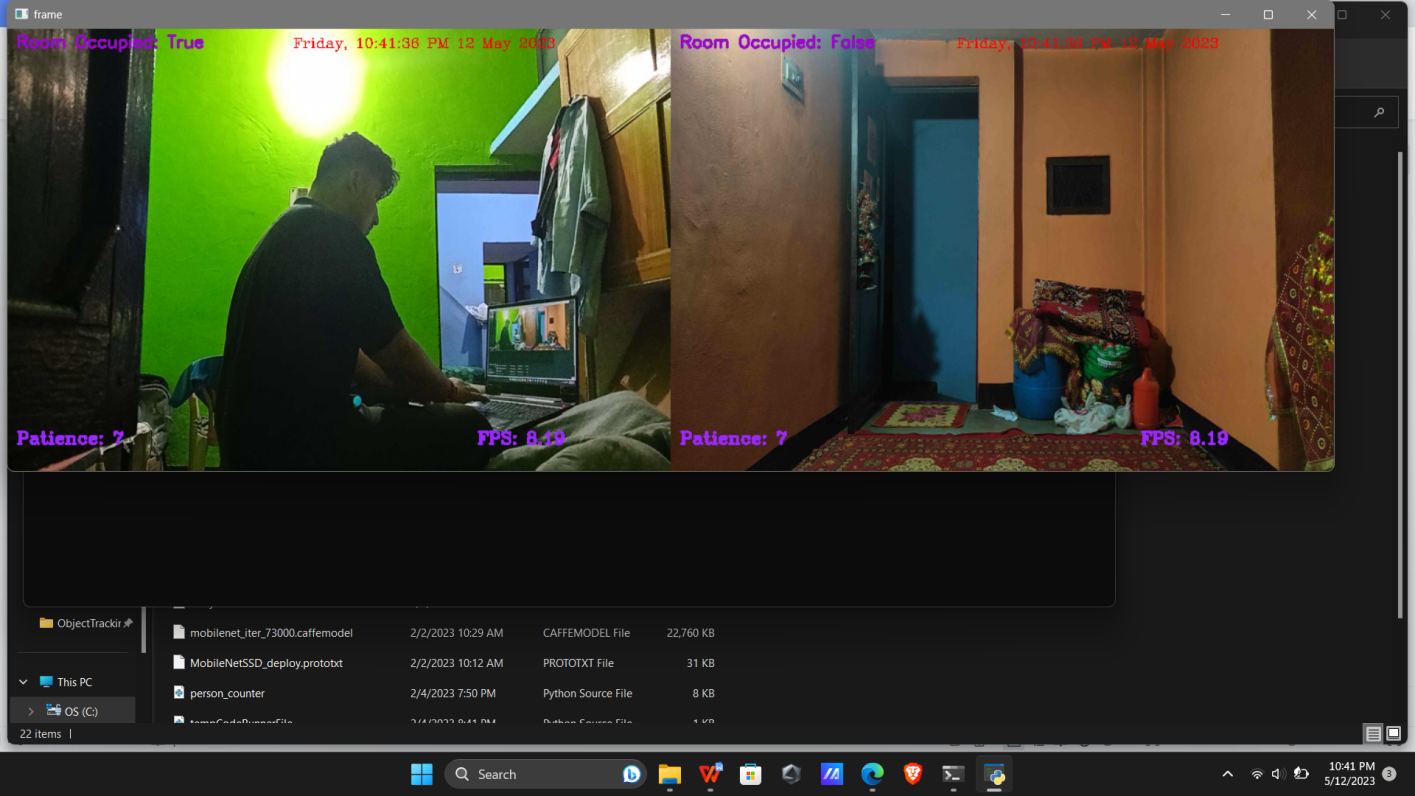












**6.4 Risk Assessment:** A report on the potential risks and vulnerabilities associated with the Smart Surveillance System, along with mitigation strategies.

* **Data privacy and security:** The system collects and stores sensitive personal information, which could be vulnerable to unauthorized access or breach. To mitigate this risk, the system should incorporate strong encryption and access controls to protect the data, and adhere to relevant data privacy regulations.
* **False positives and false negatives:** The system could generate false positives (i.e. detecting activity as a security threat when it is not) or false negatives (i.e. failing to detect a real security threat). To mitigate this risk, the system should incorporate robust machine learning algorithms and be continually trained on real-world data to improve accuracy.
* **System downtime:** The system could experience downtime due to hardware failures, network outages, or software bugs. To mitigate this risk, the system should be designed with redundancy and failover mechanisms to ensure continuity of surveillance operations.
* **Integration with existing systems:** The system may face challenges in integrating with existing security systems or infrastructure. To mitigate this risk, the system should be designed with flexibility and modularity, allowing for easy integration and customization.

**6.5 Future Work**: A section detailing future work and potential enhancements to the system, such as incorporating new algorithms or integrating with other security systems.

* **Integration with other sensors:** The system can be expanded to incorporate other types of sensors, such as audio or environmental sensors, to provide a more comprehensive view of the surveillance environment.
* **Multi-camera tracking:** The system can be enhanced to track individuals across multiple camera feeds, enabling more accurate and comprehensive monitoring of their movements and behavior.
* **Automated threat assessment:** The system can be improved with machine learning algorithms to automatically assess the threat level of an observed activity, enabling more rapid and targeted responses to potential security threats.
* **Facial recognition:** The system can be enhanced with facial recognition technology to identify individuals and track their movements, providing more granular control over access to secure areas.
* **Cloud-based storage and processing:** The system can be adapted to store and process surveillance data in the cloud, enabling more scalable and cost-effective deployments for larger or more complex security environments.
* **Edge computing:** The system can be enhanced with edge computing capabilities, enabling the processing and analysis of surveillance data to be performed closer to the source, reducing latency and improving overall performance.
* **Improved user interface:** The system can be improved with a more intuitive and user-friendly interface for security personnel, enabling them to more effectively monitor and respond to potential security threats.

**6.6 Glossary:** A list of technical terms and acronyms used in the project and their definitions.

* **Object Detection:** The process of identifying and locating objects within an image or video feed.
* **Motion Detection:** The process of detecting movement within an image or video feed, typically used to trigger alerts or notifications.
* **YOLOv3:** A deep learning algorithm used for object detection, which stands for "You Only Look Once" version 3.
* **OpenCV:** An open-source computer vision library used for image and video processing.
* **Streamlit:** A Python library used for building interactive web applications for data science and machine learning.
* **Twilio:** A cloud communications platform used for sending SMS messages and making voice calls.
* **TensorFlow:** An open-source machine learning library developed by Google, used for building and training deep learning models.
* **GPU:** A Graphics Processing Unit, a specialized hardware component used for accelerating the processing of large amounts of data, commonly used in machine learning and other computationally intensive tasks.
* **Raspberry Pi:** A small, low-cost computer used for building hardware prototypes and small-scale projects.
* **Real-time Analytics:** The process of analyzing data as it is generated in real-time, typically used for detecting anomalies, predicting future events, or making decisions based on the current situation.

## Conclusion

The Smart Surveillance System presented in this project makes use of advanced technologies such as YOLOv3, OpenCV, NumPy, and TensorFlow to detect and track objects in real-time. The system provides a reliable and efficient solution for enhancing the security of various premises.

The system's core functionality includes real-time object detection and tracking, motion detection, and generating alerts and notifications. The integration of Twilio APIs enables the system to send SMS and email notifications and even Whats-app alerts to the relevant authorities, improving the system's overall responsiveness and reliability.

Further advancements to the system could include the integration of additional machine learning algorithms, such as deep neural networks, to improve the accuracy and precision of object detection. The system could also incorporate edge computing capabilities to process video feeds on local devices, improving the system's response time and reducing network bandwidth requirements.

Additionally, the system could integrate with existing surveillance infrastructure, such as CCTV cameras and access control systems, to provide a comprehensive security solution. The system could also incorporate advanced analytics and reporting capabilities, providing valuable insights into the security and surveillance of the premises.

Overall, the Smart Surveillance System presented in this project is a highly effective and efficient solution for enhancing the security of various premises. The system's versatility and flexibility make it a valuable asset for organizations seeking to improve their security and surveillance capabilities.

## Bibliography

* YOLO: Real-Time Object Detection (pjreddie.com)
* Redmon, J., & Farhadi, A. (2018). YOLOv3: An Incremental Improvement. arXiv preprint arXiv:1804.02767.
* Bradski, G. (2000). The OpenCV Library. Dr. Dobb's Journal of Software Tools, 25-33.
* Rehman, M. Z., & Saleem, A. (2018). Streamlit: Build custom web applications for machine learning and data science.
* Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., ... & Zheng, X. (2016). TensorFlow: Large-scale machine learning on heterogeneous systems. arXiv preprint arXiv:1603.04467.
* Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., ... & Oliphant, T. E. (2020). Array programming with NumPy. Nature, 585(7825), 357-362.
* Twilio API Documentation. https://www.twilio.com/docs/api. Accessed on 7th April 2023.
* SQLite Documentation. https://www.sqlite.org/docs.html. Accessed on 7th April 2023.
* Matplotlib Documentation. https://matplotlib.org/stable/contents.html. Accessed on 7th April 2023.
* Flask Documentation. https://flask.palletsprojects.com/en/2.1.x/. Accessed on 7th April 2023.
* Docker Documentation. https://docs.docker.com/. Accessed on 7th April 2023.

These references were used as a guide to implement various components of the Smart Surveillance System and provided insights into the best practices for completing the project successfully.