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

Submitted by:-
Deepankar Sharma
Amit Sati
Pawan Chandra

Under the Guidance of:-
Ms. Richa Pandey
Faculty of GEHU, Haldwani

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:

Ms. Richa Pandey
Faculty of GEHU, Haldwani

Signature of the Student:

Deepankar Sharma(2092014)
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

1. 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.

1.1 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.

1.2 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.

2. System Requirement Analysis

2.1 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.

2.2 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.

3. 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.

4. 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.*

5. 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.

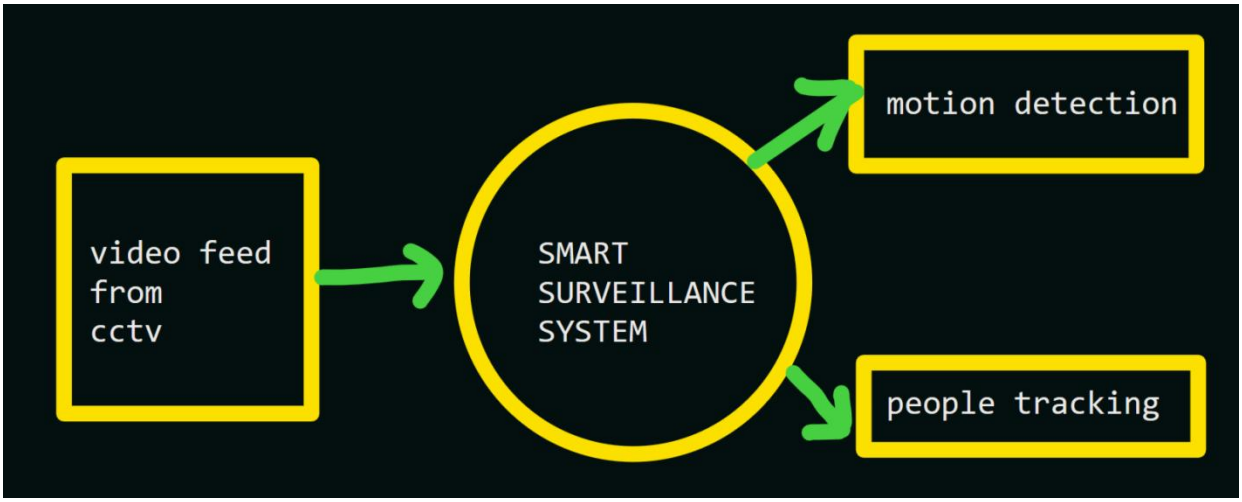
Integration with other systems: The system can be integrated with other security systems, such as access control systems or intrusion detection systems, to provide a more comprehensive security solution.

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.

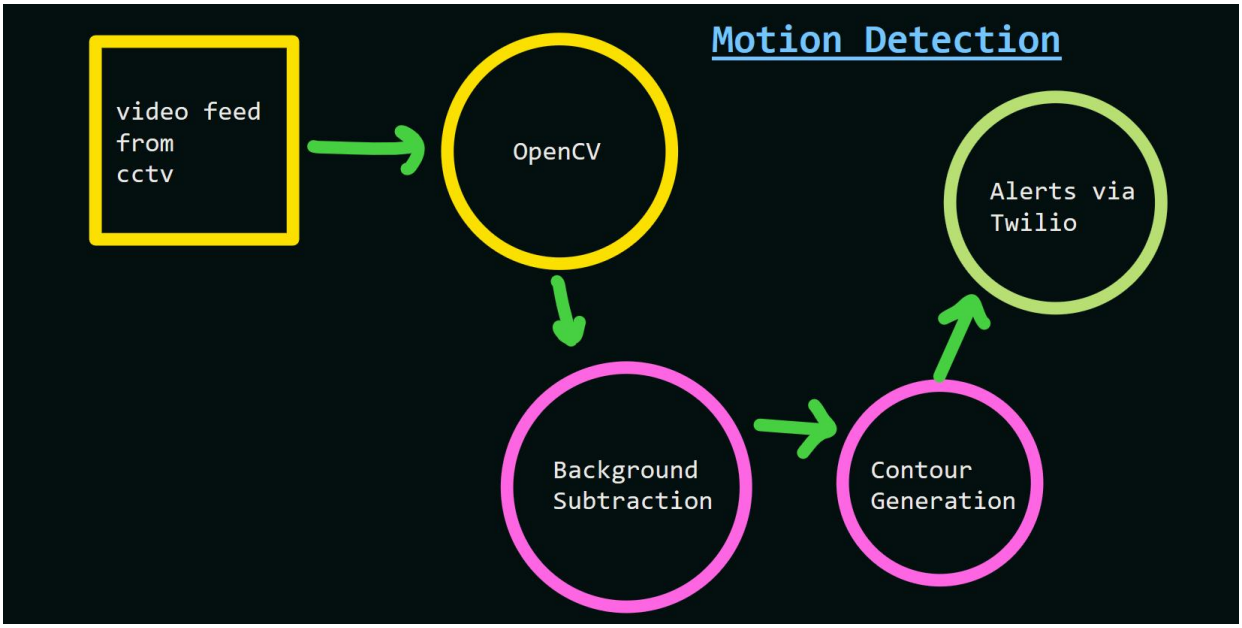
6. Appendices

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

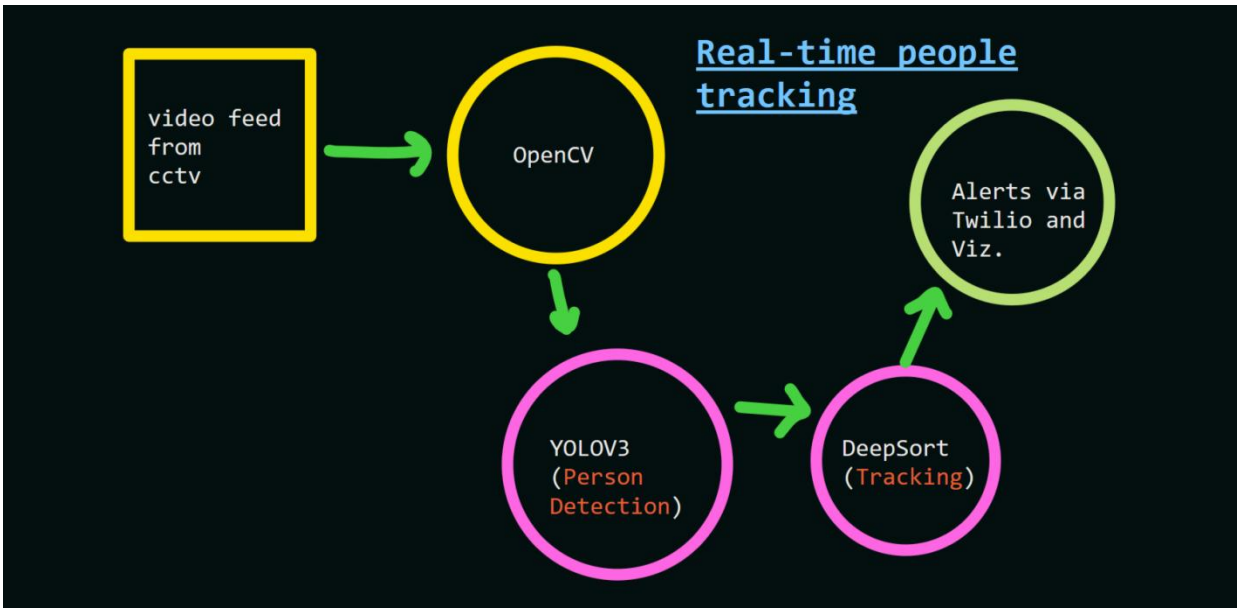
- **Level 0 DFD:** The system has a single input, which is the video feed from the surveillance camera.The system has two main outputs: motion detection alerts and real-time people tracking.



- **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:

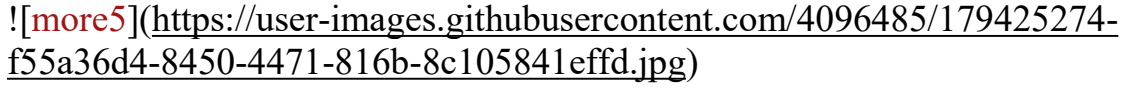
* **paper** - YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors: https://arxiv.org/abs/2207.02696

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

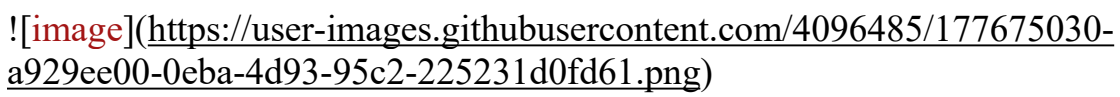
Official YOLOv7 is more accurate and faster than YOLOv5 by **120%** FPS, than YOLOX by **180%** FPS, than Dual-Swin-T by **1200%** FPS, than ConvNext by **550%** FPS, than SWIN-L by **500%** FPS.

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.

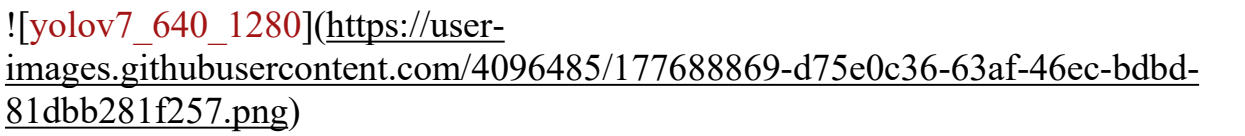
* YOLOv7-e6 (55.9% AP, 56 FPS V100 b=1) by `+500%` FPS faster than SWIN-L Cascade-Mask R-CNN (53.9% AP, 9.2 FPS A100 b=1)
* YOLOv7-e6 (55.9% AP, 56 FPS V100 b=1) by `+550%` FPS faster than ConvNeXt-XL C-M-RCNN (55.2% AP, 8.6 FPS A100 b=1)
* YOLOv7-w6 (54.6% AP, 84 FPS V100 b=1) by `+120%` FPS faster 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)



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



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



(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

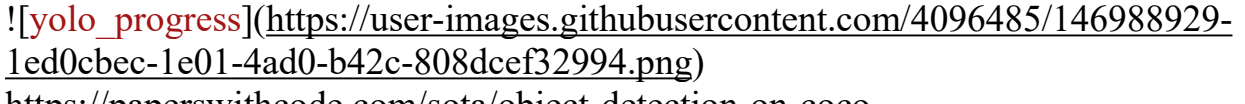
* **source code:** https://github.com/AlexeyAB/darknet

* **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> Expand </summary>



(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>
```

```c
#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)
```

CX

```
    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()
```

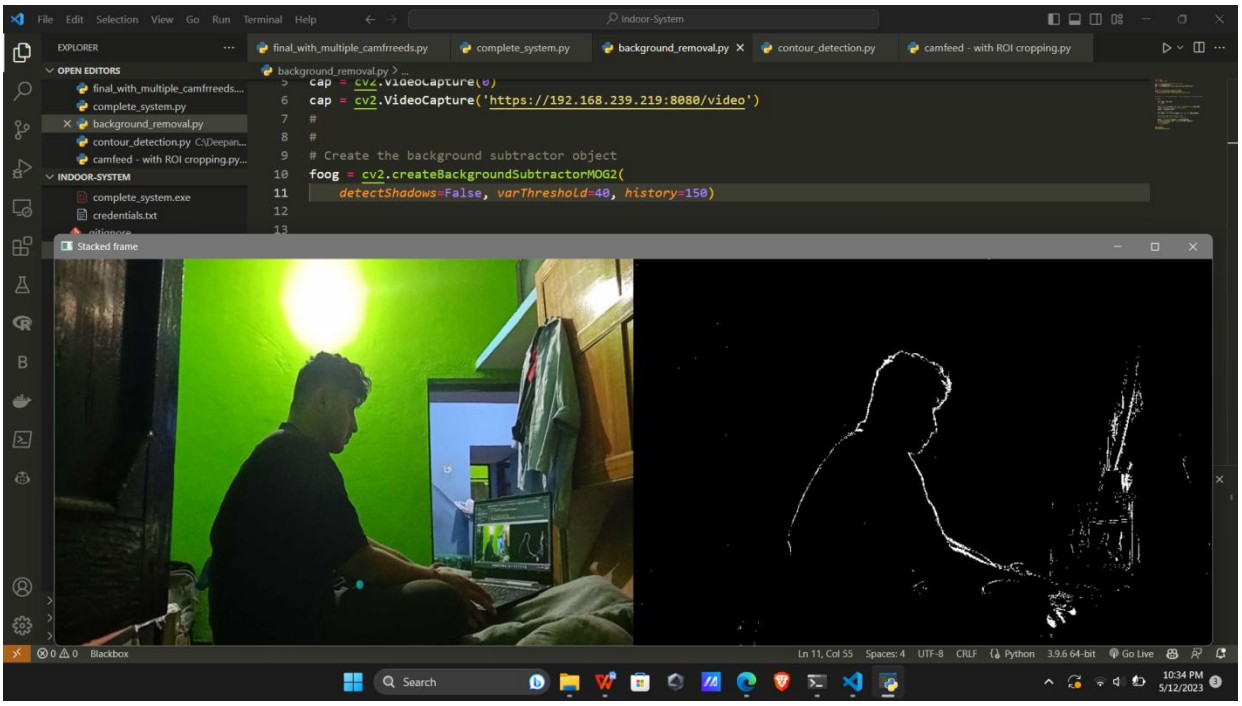
```
        if YOLO_FRAMEWORK == "tf":
            pred_bbox = Yolo.predict(image_data)
        elif YOLO_FRAMEWORK == "trt":
            batched_input = tf.constant(image_data)
            result = Yolo(batched_input)
            pred_bbox = []
            for key, value in result.items():
                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()
```

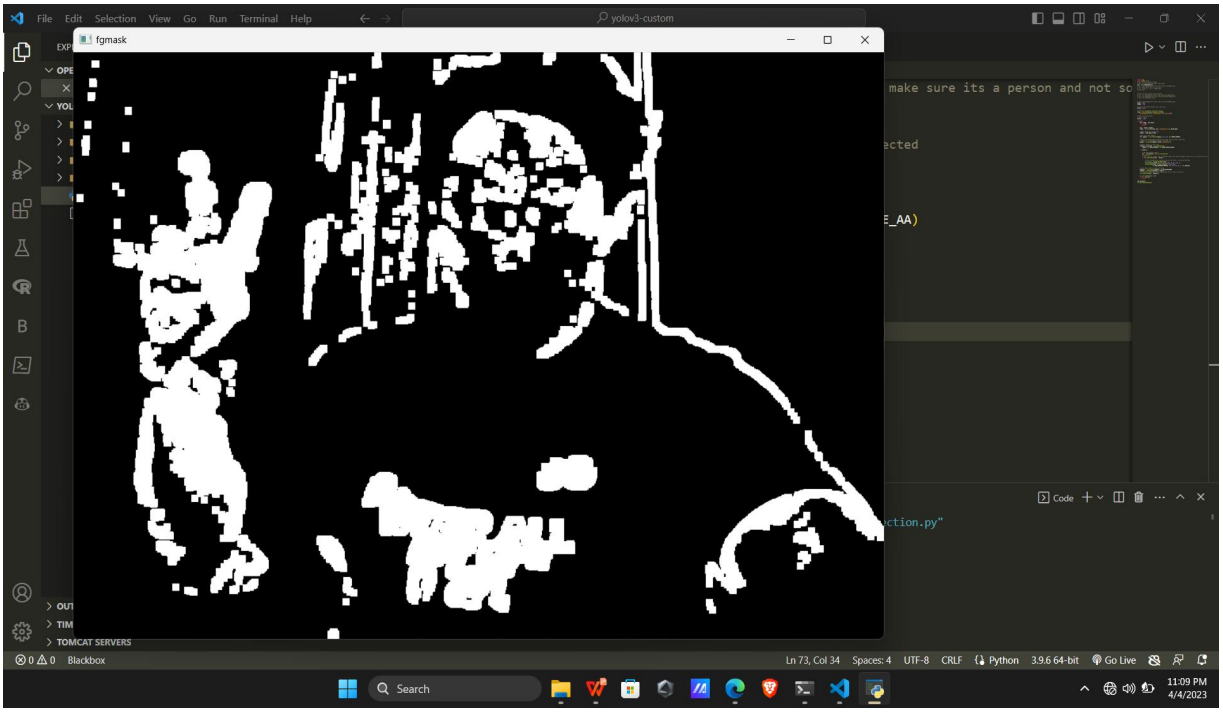
CLI

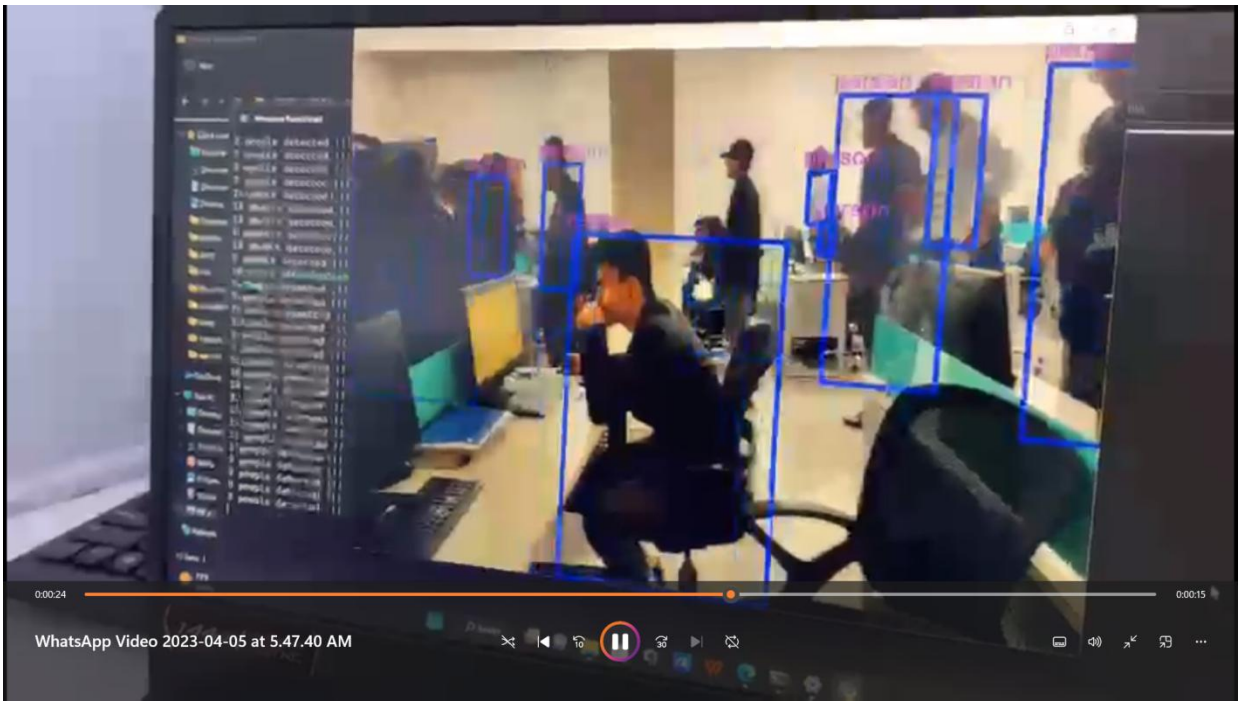
```
        tracker.update(detections)

        # 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)
```

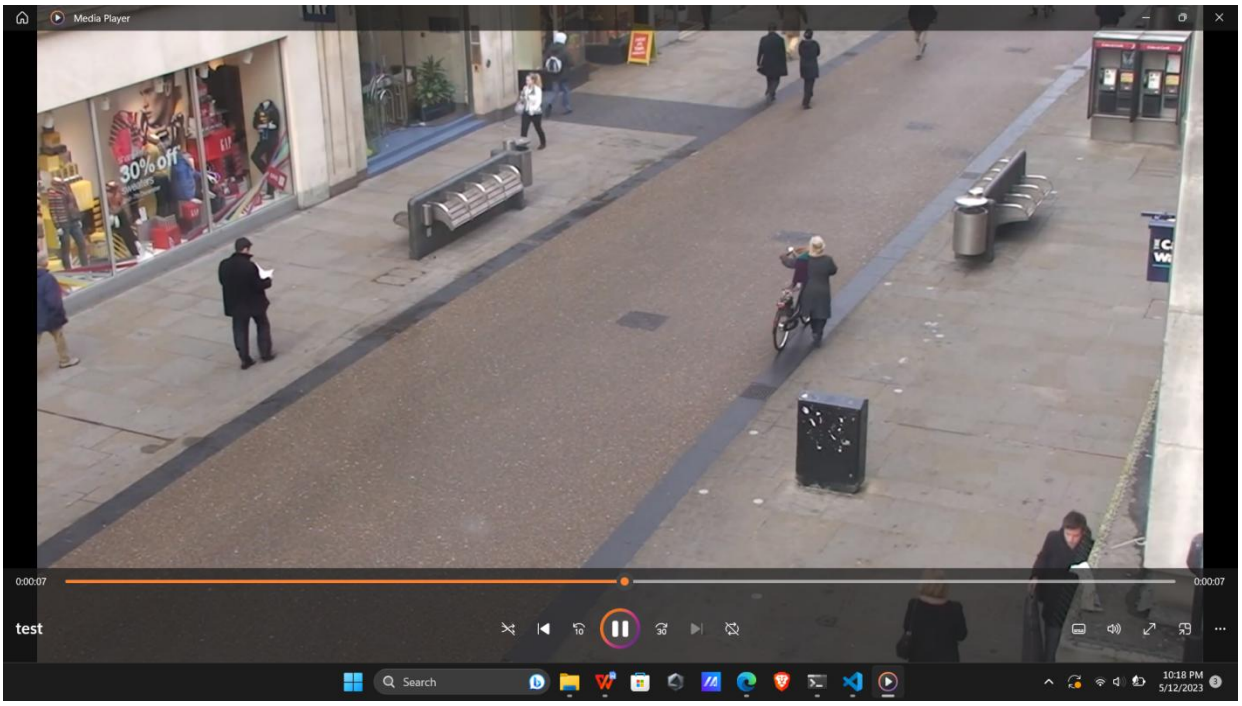
```
            # draw original yolo detection
            # image = draw_bbox(image, bboxes, CLASSES=CLASSES,
show_label=False, rectangle_colors=rectangle_colors, tracking=True)

            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"])
```

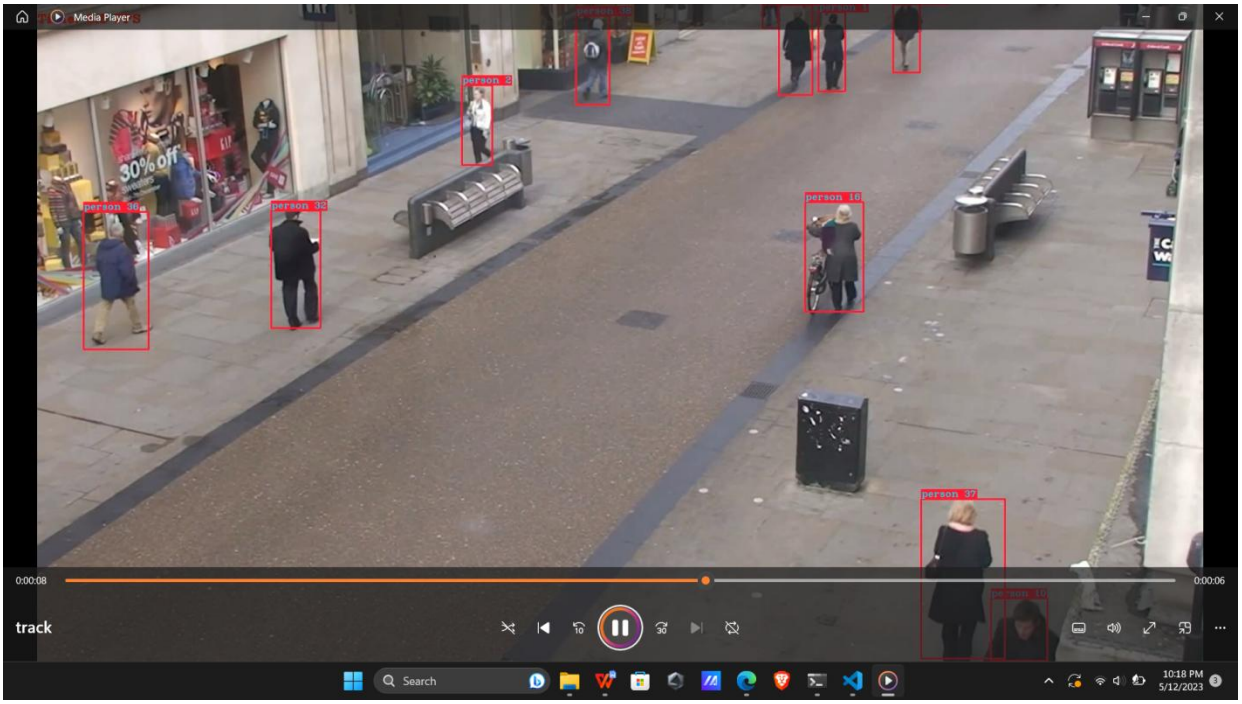
6.3 Test Data: A set of sample videos and images used to test the system's object detection and tracking capabilities.

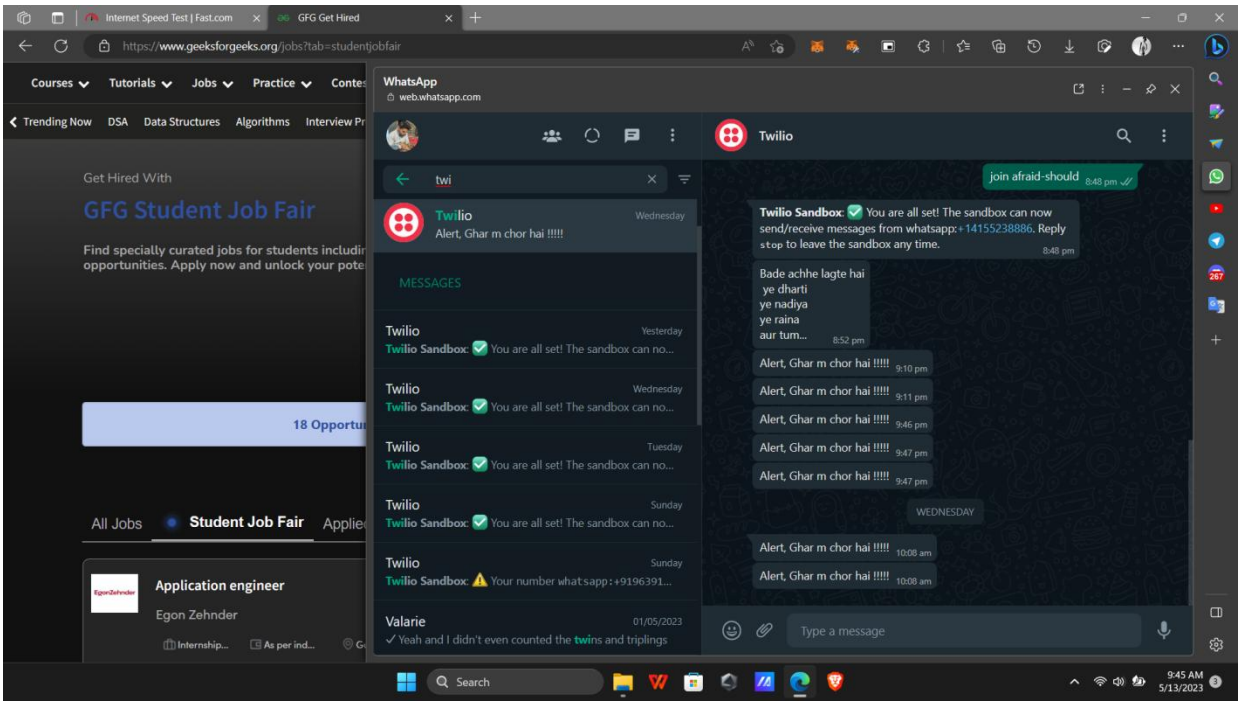


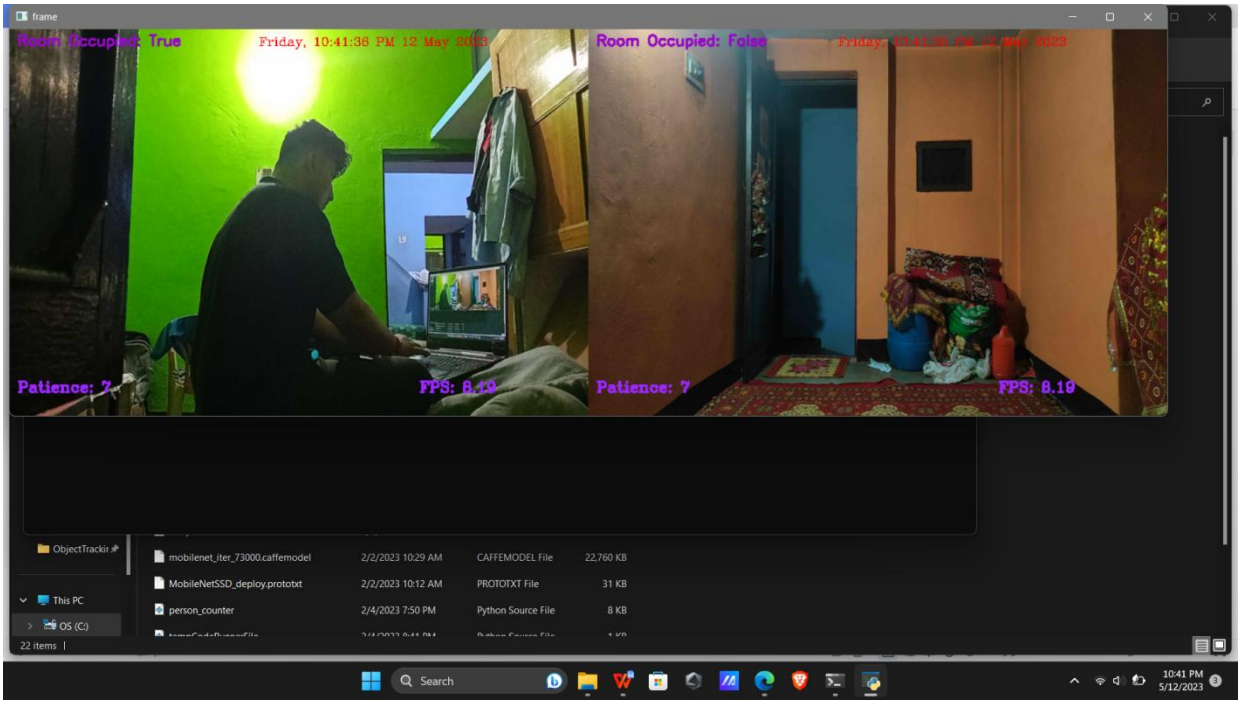



0:00:24
0:00:15
WhatsApp Video 2023-04-05 at 5.47.40 AM


Media Player
30% off
0:00:07
0:00:07
test
Search
10:18 PM
5/12/2023


Media Player
person 3
person 36
person 32
person 16
person 37
0:00:08
0:00:06
track
Search
10:18 PM
5/12/2023


Internet Speed Test | Fast.com
GFG Get Hired
https://www.geeksforgeeks.org/jobs?tab=studentjobfair
Courses
Tutorials
Jobs
Practice
Trending Now
DSA
Data Structures
Algorithms
Get Hired With
GFG Student Job Fair
Find specially curated jobs for students includir
opportunities. Apply now and unlock your pote
All Jobs
Student Job Fair
Application engineer
Egon Zehnder
WhatsApp
web.whatsapp.com
twi
Twilio
Alert, Ghar m chor hai !!!!!
MESSAGES
Twilio Sandbox: You are all set! The sandbox can no...
Twilio Sandbox: You are all set! The sandbox can now send/receive messages from whatsapp:+14155238886. Reply stop to leave the sandbox any time.
join afraid-should
Bade achhe lagte hai
ye dharti
ye nadiya
ye raina
aur tum...
Alert, Ghar m chor hai !!!!!
WEDNESDAY
Type a message
Search
9:45 AM
5/13/2023



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.

7. 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.

8. 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.