A DEEP LEARNING APPROACH TO ANALYZING ACTIVE WORKING HOURS OF EMPLOYEES

Submitted By:

TANVIRUL ISLAM

191-35-2779

Department of Software Engineering

DAFFODIL INTERNATIONAL UNIVERSITY

Supervised By:

MD. SHOHEL ARMAN

Assistant Professor

Department of Software Engineering

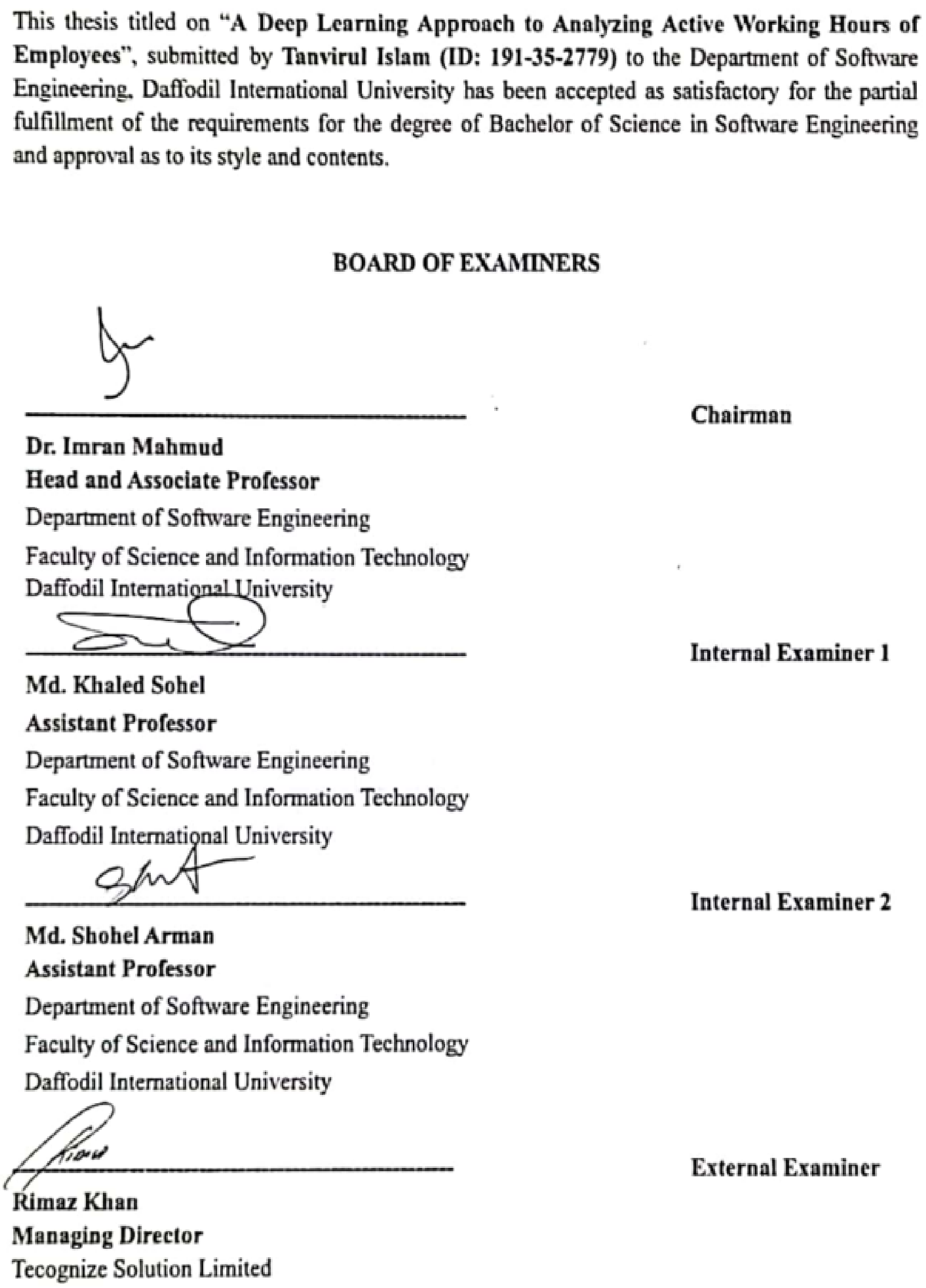
DAFFODIL INTERNATIONAL UNIVERSITY

This Thesis report has been submitted in fulfillment of the requirements

for the Degree of Bachelor of Science in Software Engineering

Fall-2022

APPROVAL



This thesis titled on "**A Deep Learning Approach to Analyzing Active Working Hours of Employees**", submitted by **Tanvirul Islam (ID: 191-35-2779)** to the Department of Software Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of Bachelor of Science in Software Engineering and approval as to its style and contents.

BOARD OF EXAMINERS

______________________________ **Chairman**

Dr. Imran Mahmud
Head and Associate Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

______________________________ **Internal Examiner 1**

Md. Khaled Sohel
Assistant Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

______________________________ **Internal Examiner 2**

Md. Shohel Arman
Assistant Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

______________________________ **External Examiner**

Rimaz Khan
Managing Director
Tecognize Solution Limited

DECLARATION

I hereby declare that I am submitting this research paper, the title **“A DEEP LEARNING APPROACH TO ANALYZING ACTIVE WORKING HOURS OF EMPLOYEES”** to **Mr. Md. Shohel Arman**, Assistant Professor, Daffodil International University's Department of Software Engineering. I thus certify that neither a bachelor's degree nor any other type of graduation was suggested for this work or any element of it.

Tanvirul Islam

Tanvirul Islam
ID: 191-35-2779
Batch: 28th
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

Approved By:

Mr. Md. Shohel Arman
Assistant professor
Department of Software Engineering
Faculty of Since and Information Technology
Daffodil International University

ACKNOWLEDGEMENT

I ought to commend Almighty Allah for His holiness and for enabling me to accomplish my undergraduate thesis.

In order to adequately praise and esteem my supervisor, **Md. Shohel Arman, Assistant Professor** in the Department of Software Engineering at Daffodil International University in Dhaka, Ideally I prefer to use the word retard. His extensive expertise and assistance in the "Deep Learning" portion really helped me to accomplish this full thesis work. It has been achieved by his unwavering compassion, scholarly literature, constant inspiration, routine and vigilant surveillance, insightful comments, useful advice, and reading several poor submissions and improving them on all levels.

I now want to symbolize my heartfelt indebtedness to **Dr. Imran Mahmud**, the Head of the Software Engineering Department within the Faculty of Science and Information Technology, as well as to the other professors, faculties, and administrators of the SWE Department at Daffodil International University for their thoughtful support in finalizing my work.

Finally, I would want to sincerely be grateful to my parents for their everlasting love and acceptance.

Table of Contents

List of Figures

List of Tables

List of Nomenclatures

YOLOv7 - “You Only Look Once version 7(Real-time Object detector.”

FLASK - “Python micro web framework”.

DRHN - “Deep Recurrent Hierarchical Network”.

LBPH - “Local Binary Pattern Histogram”.

CNN - “Convolutional Neural Network”.

KNN - “K-Nearest Neighbor”.

SVM - “Support Vector Machine”.

ANN - “Artificial Neural Network”.

RPN - “Region Proposal Network”.

LSTM - “Long Short-Term Memory”.

mAP - “Mean Average Precision”.

CCTV - “Closed-Circuit Television”.

ABSTRACT

It is conceivable to see how active each individual is at work by observing their various physical postures and gestures. With conventional systems, it is hard to constantly monitor every employee and make the best use of them. Our thesis' major goal is to determine the most effective strategy to utilize each employee's available work time. The challenge of object detection in visual data has been demonstrated to be almost fully resolved by deep learning, a method that mimics the information flow of the human mind. Allowing computers to distinguish not just things but also activities is one of the upcoming key issues in computer vision. In this work, the possibilities of deep learning are examined for the particular challenge of activity recognition in office settings. Data from CCTV footage of several offices during business hours was gathered in order to implement the 'YOLOv7 object detection' model. The research implemented a re-labeled dataset of distinct office worker motions to distinguish between employees' levels of activity. With the assistance of Flask, we can build a single-page website, deploy our model, and perform round-the-clock monitoring. After training, the model displays higher accuracy(95.7%), demonstrating how ideal it is for this situation. By employing this approach, it is possible to estimate an employee's productivity by observing their numerous motions while the workplace is in operation. Any official context can use our concept approach, and deployment services can provide ongoing monitoring.

CHAPTER 1

1. INTRODUCTION:

There are several workplaces across our region where people waste time working on unrelated tasks. As a result, office work moves forward at a relatively leisurely pace. In order to address this issue, we want to offer a model that makes it simple to identify office member activity utilizing various “Deep Learning” techniques or algorithms. To date, very little work has been done to measure active working hours. Notable among them - Karl Casserfelt employed a gesture recognition sensor to concentrate on the employee's physical movement in his 2018 paper, "A Deep Learning Approach to Video Processing for Scene Recognition in Smart Office Environments." Based on the camera shot, they divided the information into groupings that produced superior results. With the title "Work Engagement Recognition in Smart Office," Congcong Ma, Carman Ka Man Lee, Juan Du, Qimeng Li, and Raffaele Gravina published a paper in 2022. They identified head and body positions using “Machine Learning” techniques. However, their accuracy rating fell short of expectations. In their article titled "Detection of sitting posture using hierarchical image composition and deep learning" published in 2021, Audrius Kulikajevas, Rytis Maskeliunas, and Robertas Damaeviius stated the ‘CNN, ANN, and YOLOv3-based’ model to determine posture. Feng Yang, Xingle Zhang, and Bo Liu's proposal, "Video object tracking based on YOLOv7 and DeepSORT," employed ‘YOLOv7 and DeepSORT’ to identify objects.

We are looking for a technique that can work in any office setting on a shoestring budget. Developers are now taking notice of ‘YOLOv7’ due to its improved object identification efficiency. The context of the physical detector ‘YOLOv7’ is making “computer vision” object recognition more sophisticated than before. The utilization of a ‘YOLOv7’ model in conjunction with regular CCTV demonstrates excellent results in the identification of staff actions as opposed to installing sensors and heavy applications.

1.1. BACKGROUND:

How effectively an office operates is contingent on how efficient its employees are. Nevertheless, it is crucial to have a mindset of helping each other out and making efficient use of time if businesses are to accomplish their objectives. Utilizing every moment of the day is necessary to use the corporation's whole workforce. Every office worker must work energetically if any office job is to be finished on time. The company's annual production will rise greatly at the end of the year if all employees can utilize the office hours efficiently. However, a significant issue is that most individuals don't give their tasks their whole attention. Even when they have a lot to accomplish, they prefer to relax by doing nothing. That has a major impact on the business. The company's individuals are accountable for its reputation. If they spend a lot of time on private things during crucial business hours, it might be expensive for the corporation. This disregard for squandering time occasionally hurts the entire nation in addition to the office.

In response to such circumstances, we created a system based on ‘Deep Learning and Computer Vision’ methods that can identify persons engaging in any anomalous activity, such as standing, using their cellphones, or chatting, even if they are not present at their designated workstation. Consequently, we evaluated employee activity.

Systems and applications can use the field of computer vision to extract pertinent data from digital photos, videos, and other optical inputs and to make choices or provide options in light of that information. Installing sensors or other expensive components to detect posture or gesture is unnecessary and too pricey; we can achieve the same results with a few programs. Some methods can be used to resolve the problem even without a sensor. However, it also needs intensive upkeep and skilled labor. Some models are too sluggish to be effective over the long term. We looked for a system that could be executed with less complexity, with faster calculation times, and that was also cost-focused. That's why we suggested a technique that employed YOLOv7 (You Only Look Once-version 7) to identify and separate the applicants' proper and improper postures. Our implementation service allows us to keep a tight eye on all sorts of employee-related actions. We took into account every possible movement of the employee to detect their activeness.

Table 1.1: Classification of Employee's Activity

ACTIVITY TYPE	DETAILS	LABELS NAME
Actively Working	Engaged in Work	working
	Writing Something on Paper	writing
Idly Working	Eating/Drinking Something	eating_drinking
	Sleeping during Work-hour	sleeping
	Talking with Colleagues / Someone	gossiping
	Using Phone / Mobile	using_phone
	Standing without Working	standing
	Absent in the Desk	empty_desk
	Wasting Time as Leisure	not_working

1.2. MOTIVATION OF THE RESEARCH:

It becomes conventional to not be productive in the workplace during work time. Employees use a variety of tactics to avoid working during set hours, even if it means acting inappropriately. It is impossible to watch over everyone at once and constantly. Because of this, office hours cannot be used in certain situations in a productive or genuine way. The main target of this research project is to eliminate needless time loss by developing an integrated system that makes the workplace smarter. We computerized the whole area with the use of CCTV. A notion for a smart workplace environment was put out by several academics. Some of them built their models using real-world tools like cameras, motion detectors, and facial identification sensors. When installed extensively, it becomes quite costly. Some of them made use of primitive algorithms, which took a lot of time and occasionally had poor accuracy. We suggested the most recent object detection technique, which has quicker, more accurate computations and is cost-effective, to address all the problems.

1.3. PROBLEM STATEMENT:

Unused office time may cost a corporation a small fortune. That type of situation frequently occurs in nations like ours in the workplace. Because of this, office employees are unable to finish their tasks in a timely manner, which makes the work of all portions take longer. Organizations occasionally need to grow their workforce in order to resolve any concerns. The outcome is not as pleasant as it ought to be. It makes the company's financial losses worse. Even if it adds more work and stress to the employee's already heavy burden. Thus, the quality of businesses' yearly output is diminished. The lack of progress on that subject is turning into a serious worry for the growth prospects of the businesses as well as for the entire region.

1.4. RESEARCH QUESTIONS:

The inquiries for the study were:

- Q1: Can we use the “YOLOv7” deployment model to more accurately and quickly identify changes, motion, and activity?
- Q2: Can we retain the model for each user's financial means?

1.5. RESEARCH OBJECTIVE:

The main objective of this study is to recognize and track every individual. To broaden the applicability of our approach, we also sought a better result. The top priorities are:

- To increase the active working hours of the office.
- To keep track of every personnel.
- To improve the concentration, efficiency, and effectiveness of every individual.
- To decrease idle office time and non-work engagement events.

1.6. RESEARCH SCOPE:

We concentrated on the relevant perspectives when doing this analysis using ‘Deep Learning’ and ‘Computer Vision. These scopes are as follows:

- In conducting this study, we took into account people's movement patterns and activity inside an office setting.
- We're searching for an automated solution to monitor and manage the office.
- To construct the model, almost 2.5h real-life video footage was collected.

Along with this, some discrete videos were also processed.

- More use of training data will contribute to the result.

1.7. THESIS ORGANIZATION:

In the initial chapter, a portion on "ACTIVITY RECOGNITION OF EMPLOYEES" in particular is coated, along with its implementation and workflow, backstory context for the study, the key ‘Problem Statement’ of such research, a precise ‘Motivation for the Research’, a ‘Research Question’ on which it concentrates, a ‘Research Objective’, and a ‘Research Scope’. Our prosecution's additional components are as follows:

We shall analyze various researcher's studies that have already been conducted in this field in the following chapter, The Literature Review. We will look at their applied technique and any shortcomings. Afterward, we shall discuss our methodological approach. The methodology section will provide a description of the full data collection, data pre-processing, and work procedures and techniques. The ultimate outcomes will be examined in chapter 4. My attempts are fully reviewed in the concluding chapter, which also provides a list of my deficiencies.

1.8. SUMMARY:

There is always a percentage of employees who are not actually working throughout the scheduled hours. This model will have a robust activity detection approach that keeps track of every employee's movements and other activities.

CHAPTER 2

2. LITERATURE REVIEW:

2.1. INTRODUCTION:

We assessed prior work, research, conference papers, books, articles, etc. for moving forward with our research effort in the “Literature Review” segmentation. On the internet, there isn't a lot of relevant paper to be found. Therefore, we paid more attention to the areas that recognized activities. Most of them focused there in order to identify any personnel's motions in the various bodily sections. We thus took what we could from their earlier work and provided a broad summary of it. Afterward, we seek to combine such thoughts with ours.

2.2. PREVIOUS LITERATURE:

Many academics, both domestically and internationally, have not done a detailed analysis of the problem of office activity detection. Techniques that depend on employee conduct in the office environment approach based on whether or not something is actually operating and other techniques have been used to identify activity up to this point. We've discussed a few papers in this part that may be useful for our study.

Congcong Ma et al [5] use machine-learning to assess head and body motion (Decision tree, KNN). They mainly concentrated on the employee's seated position and were able to create digital photographs of the employee that were confidential. Although they took a very commendable strategy, their data analysis fell short, which had an influence on their final outcome.

Karl Casserfelt et al [1], in their research paper, proposed a smart office environment using a surveillance camera. He counted everyone's movements with the help of CNN and RCNN. He also used heat maps to recognize humans and their movements. He applied his proposed model for object detection. According to the camera shot, the results have improved since he divided the data into subgroups. But his dataset was not properly structured which makes this work questionable.

Audrius Kulikajevas et al [3] implemented some Deep-Learning approaches, like - ANN, CNN, DRHN, and YOLOv3 for detecting the sitting posture of every individual. Firstly they extracted frames and then removed unnecessary classes along with their frames. Then they performed their training for identifying the human states according to their sitting position. But here we saw, they used only some specific body sizes into their consideration. That means they follow specific body size indexes for their research paper. So, it may cause variations in the result when the model will apply to different shape's human bodies.

Keze Wang et al [5] focused on 3D human activity identification with the help of CNN. They researched raw RGB-D data for managing realistic obstacles to notify activity. Activity recognition is quite a complex problem for CNN to solve. In this era, there are a lot of efficient algorithms for solving this issue with higher accuracy and speed.

R. Meena et al [5] research has taken count the recognition of complicated human actions. After the extraction of features after cleaning the data, a histogram of oriented gradients(HOG) as a skeleton model was applied. They also used "CNN-LSTM and RNN". As a classifier KNN was implemented. Although they have employed many algorithms, their dataset incorporates semi-temporal information, and they employed a sophisticated architecture to address such problems. With genuine and precise datasets, they may concentrate on developing less complicated designs.

Feng Yang et al [3] claim object-detection with the help of a video camera and YOLOv7-Deep Sort algorithm. YOLOv7 works very well in object identification. So, its output is showing higher precision. But researchers can improve their proposed model if they can improve multiple object tracking at a time.

Sidrah Liaqat et al [5] created a hybrid model using Machine Learning and Deep Learning for posture detection. As a Deep-Learning approach, they used CNN and LSTM and for classification, they used “KNN, SVM, and Decision Trees”. Their model efficiency is quite impressive. But for implementing this model we need a highly configured pc.

Sudan Jha et al [4] with the help of a surveillance system generated a real-time object-detection or tracking system. They employed RCNN and YOLO as their training algorithm. In their case, they didn’t resize any frames. Instead of this, they received only fixed-sized images. Their model every time creates a new object-id. Thus they are not able to manage to re-enter. Their model can be improved if they focus on that issue.

Rithik Kapoor et al [3] produced a lightweight model that can recognize the seating position and during seating the inclination of the body. RCNN and YOLO were implemented as end-to-end posture detectors to identify bad posture. Their model has only two classes. Considering posture, they need to add more different body-shaped subjects for identification. Because various people have different sitting styles. It's difficult to distinguish between good and bad.

MCP Archana et al [3] proposed a biometric tool that operates in the online environment. Using LBPH, and CNN they mapped the faces. After that as a classifier, they used “Hear Cascaded”. Their accuracy depends on the camera quality. That means, they didn’t consider blur images or augmentation, when the dataset was trained.

2.3. SUMMARY:

We may infer from the experiments mentioned above that different methods or techniques were employed to recognize different types of human motions. After analyzing the data, they used the necessary algorithm, which is what they are concentrating on. In our study, we made every effort to address their weaknesses and incorporate their ideas into our own context.

CHAPTER 3

3. RESEARCH METHODOLOGY:

3.1. INTRODUCTION:

We employed YOLOv7 to identify individuals who had been engaged. It will serve as our learning procedure's object detection algorithm. The system deployment was then done via FLASK.

3.2. DATA COLLECTION:

We had to gather data from several real-world scenarios because we couldn't come across an appropriate and readily accessible dataset online. Data was gathered from CCTV recordings of various workplace settings. Each video is about three hours long. Then, in order to increase the precision of our model, we constructed a fictitious office atmosphere as well as some office space footage with a one-hour length.

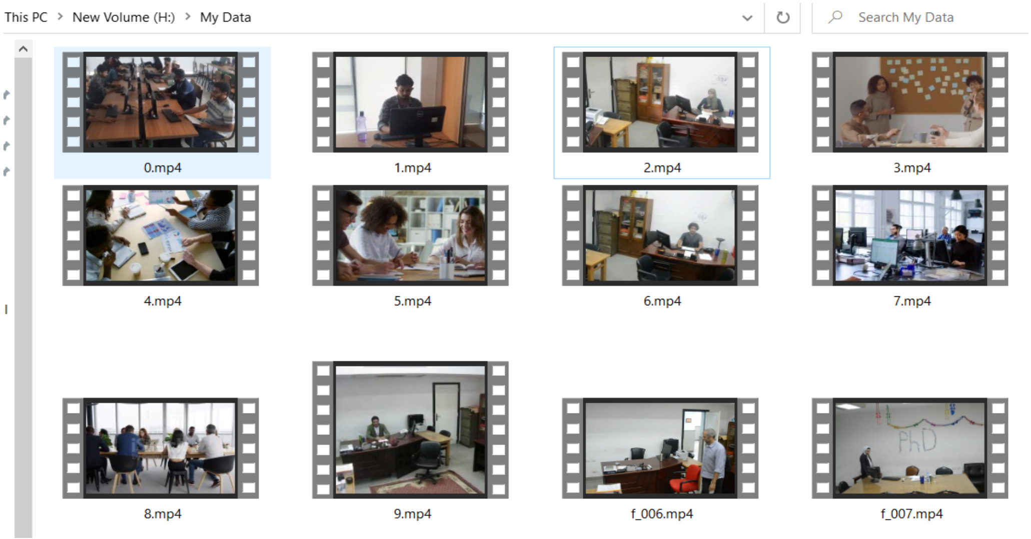


Figure 3.1: Uncut Video Clips

3.3. DATA PREPROCESSING:

After gathering footage from real-office circumstances, we extracted frames or images with a simple code, using OpenCV. We have achieved nearly 14466 pictures or frames after extraction. After removing some duplicate photographs, we were left with 9453 frames. Then, we conducted some augmenting and resizing of our images to make our training more efficient because, by this, our model can be trained at a different angle. Nine classes based on the office individuals' movement were created to learn our model.

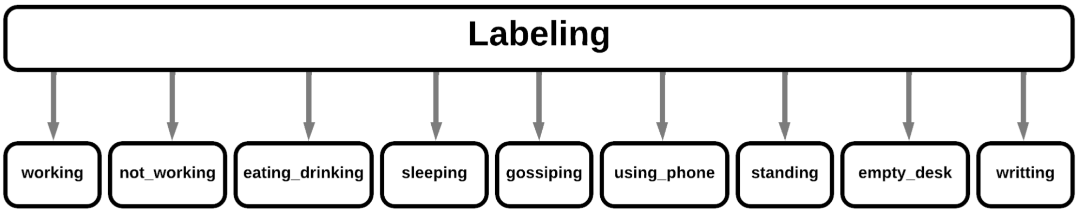


Figure 3.2: Annotation Labels

Applying the "labelMe" GitHub repo code and the 'YOLO format', we labeled those photographs, producing 9453 annotation text files.

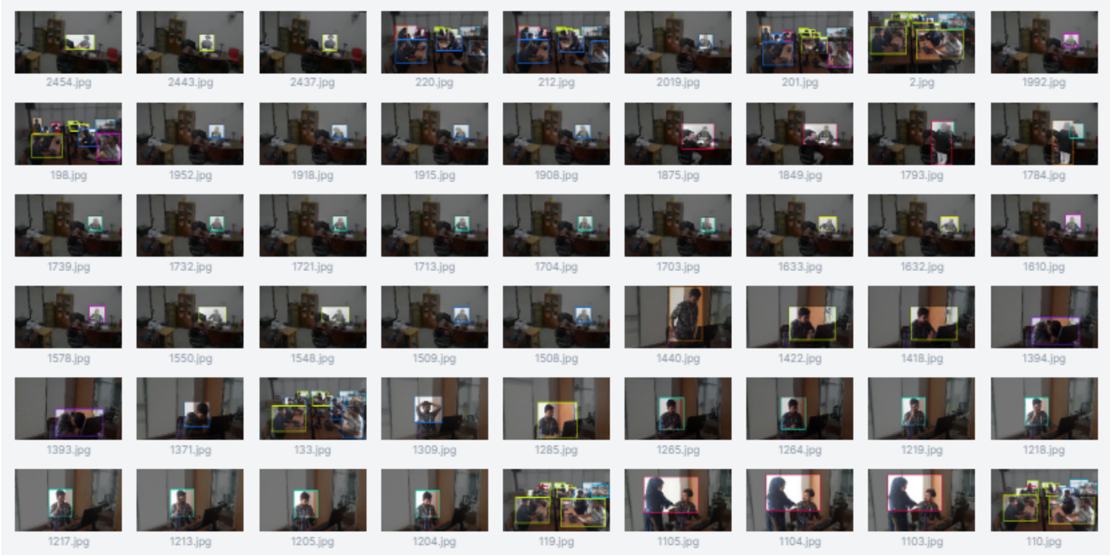
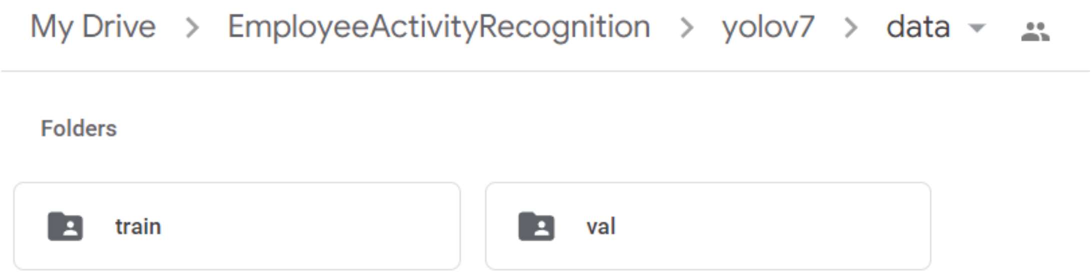


Figure 3.3: Annotated Image Data

Following that, we separated those tag files and pictures into several directories. Two directories, train, and validation were made. Eighty percent of the pictures with labels on them were saved in the collection for trains. The remaining photos were added to validation folders along with their appropriate labels.


My Drive > EmployeeActivityRecognition > yolov7 > data

Folders

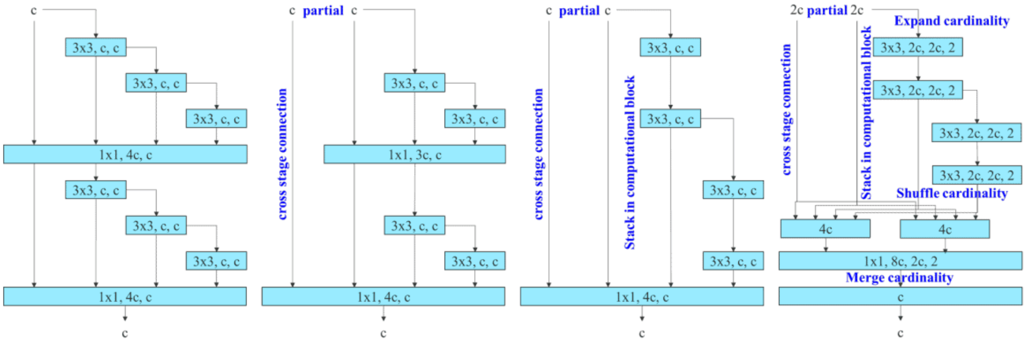
train

val

Figure 3.4: Splitting Image Directories

3.4. YOLOv7:

The authentic object identification model for computer vision programs that is quickest and most accurate is ‘YOLOv7’. Chien-Yao Wang, Alexey Bochkovskiy, and Hong-Yuan Mark Liao published the authorized YOLOv7 article titled "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors" in July 2022. The computer vision and machine learning community are buzzing about the “YOLOv7” algorithm. The most recent ‘YOLO’ algorithm outperforms all earlier object detection methods and YOLO iterations in terms of speed and precision. It can be taught significantly quicker on tiny datasets without any pre-learned weights than for other neural network models and requires technology that is several times less expensive. As a result, ‘YOLOv7’ is anticipated to overtake ‘YOLOv4’, the previous state-of-the-art for real-time applications, to become the accepted standard for object recognition in the near future. The “real-time object detection” performance is significantly increased by ‘YOLOv7’ without raising the inference expenses. ‘YOLOv7’ effectively outperforms other well-known object detectors by reducing about 40% of the parameters and 50% of the computation required for state-of-the-art real-time object detections. This allows it to perform inferences more quickly and with higher detection accuracy. In summary, ‘YOLOv7’ offers a quicker and more robust network architecture that offers a better feature integration approach, more precise object recognition performance, a more robust loss function, and an improved label assignment and model training efficiency. Because of this, ‘YOLOv7’ uses far less expensive computational hardware than other deep-learning models. It accepts the COCO dataset. The intended re-parameterized convolution design in ‘YOLOv7’ employs RepConv without identity connection (RepConvN). When re-parameterized convolution is used to replace a convolution process with residue or concatenated, the intention is to prohibit contact points from emerging.



(a) VoVNet [39] (b) CSPVoVNet [79] (c) ELAN [1] (d) E-ELAN

Figure 3.5: YOLOv7-Architecture

First, we modified the official “coco.yml” and “yolo.yml” files for ‘YOLOs’ in accordance with our classes or tags for our YOLO training. Then, we ran 8 batch sizes across 60 epochs. In GPU-based runtime, it operates really well.

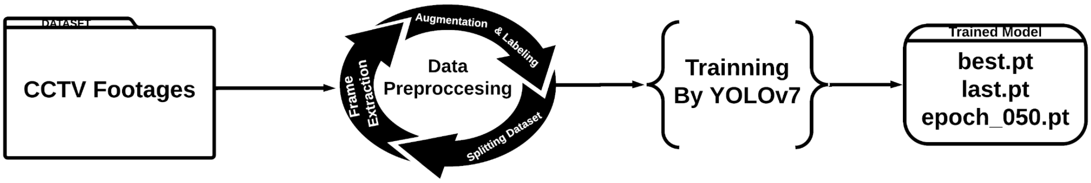


Figure 3.6: Custom Model Training Procedure

3.5. FLASK:

Python-based web application platform Flask offers modules to create compact apps. It is built using the Jinja2 template system and the WSGI tools. Considered a micro framework is Flask. Virtual server connector interface, sometimes known as WSGI, is standardized for creating web applications in Python. It is regarded as the standard for the common interface here between web applications and servers. Jinja2 is a web template generator that renders dynamic web pages by fusing a template with a specific data source. Python 2.7 or a later version must be present on the system in order to install flask. However, for the creation in the flask, we advise using Python 3.

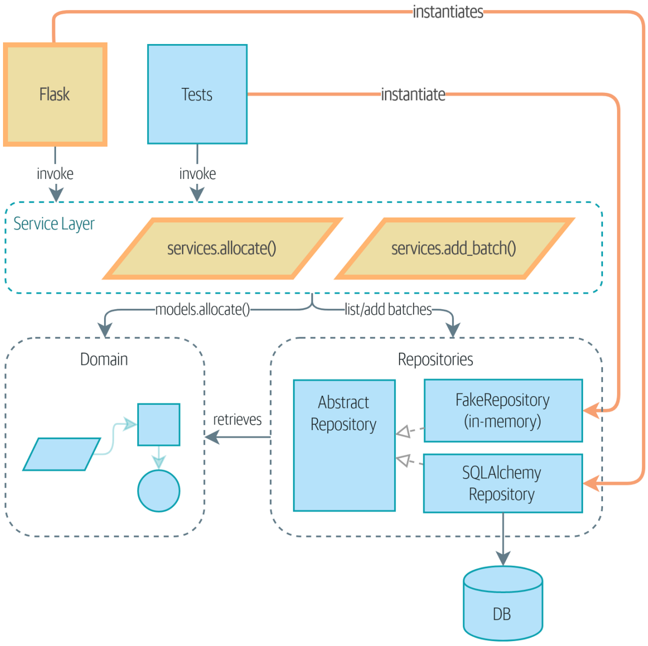


Figure 3.7: Flask-Architecture

Following the training, we developed a virtual platform for operating flasks. Then an app.py file was created. We then integrated our specially trained ‘YOLOv7 model’ (Weights file- "best.pt" or "last.pt") into that file after installing the flask library and inserting it into our app.py file. Then we launched our model into the server with the assistance of the index.html file.

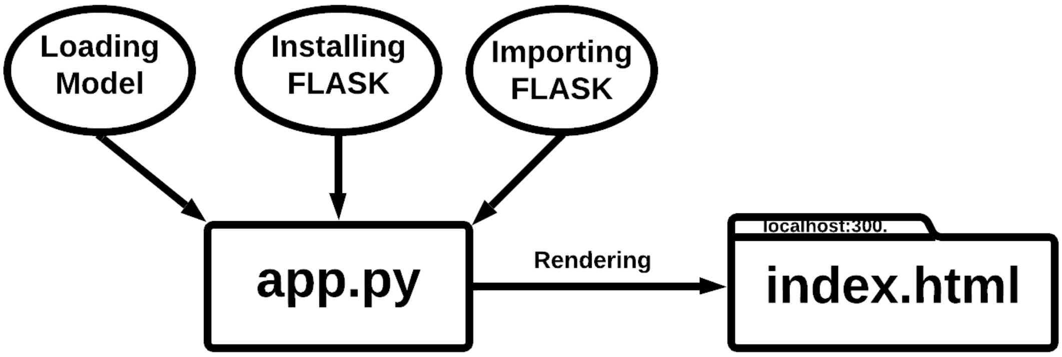


Figure 3.8: Process of Flask

3.6. TRANSFER LEARNING:

Transfer-learning is the practice of using streamlined hyperparameters that have already been developed. The neural-network process applies the model it has learned from solving one problem to others. Transfer learning techniques may be used to identify the general characteristics of pictures. It performed at the maximum level while preserving the efficiency of the system. The sole approach recommended in the research is deep-learning of vision algorithms. In our research work, our customized trained model produced some weights files that can be used for furthermore approaches.

3.7. EVALUATION METHODS:

The goal of model evaluation is to determine how well a model predicts the future.

Understanding some parameters is necessary for evaluating our model. Those are-

True_Positive(TP): A result where the parameter optimization predicted the positive class is referred to as a “True Positive”.

True_Negative(TN): True negative results are those for which the model accurately predicted the negative category.

False_Positive(FP): An outcome when the analysis prior to the positive class inaccurately is known as a false positive.

False_Negative(FN): False-negative results occur when the model predicts the negative category inaccurately.

3.7.1. Accuracy:

The accuracy of a machine's outcome prediction depends on how accurate the model is. When each class is equally important, something crucial has occurred. Every class is crucial to our area of work. As a result, precision is essential in establishing whether the model is adequate.

$$\text{Accuracy}= \frac{True\_Positive+True\_Negative}{True\_Positive+False\_Positive+True\_Negative+False\_Negative} \quad \text{…………..…..(3.1)}$$

3.7.2. Precision:

It is a tool to gauge how effectively a trained model works. Precision is obtained by dividing the true positive value by the overall positive value.

$$\text{Precision} = \frac{True\_Positive}{True\_Positive+False\_Positive} \quad \text{…………………....…………………(3.2)}$$

3.7.3. Recall:

Recall is an assessment of the exactly specified true positive. To estimate the recall, split the actual positive value by the overall number of related items that are extant.

$$\text{Recall}= \frac{True\_Positive}{True\_Positive+False\_Negative} \quad \text{………………..……....…………………(3.3)}$$

3.7.4. F1 Score:

Test accuracy is assessed by the F1 score. Both recall and precision are used to obtain the F1 value.

F1 Score = 2 * $\frac{Precision*Recall}{Precision+Recall}$ …………………………………..……………(3.4)

We have shown the validity of the model and mAP for this method, the YOLOv7 technique. In the predictive performance mAP algorithm, training consistency and efficiency are taken into account while calculating accuracy. The costs for the system are also considered in terms of training and testing productivity.

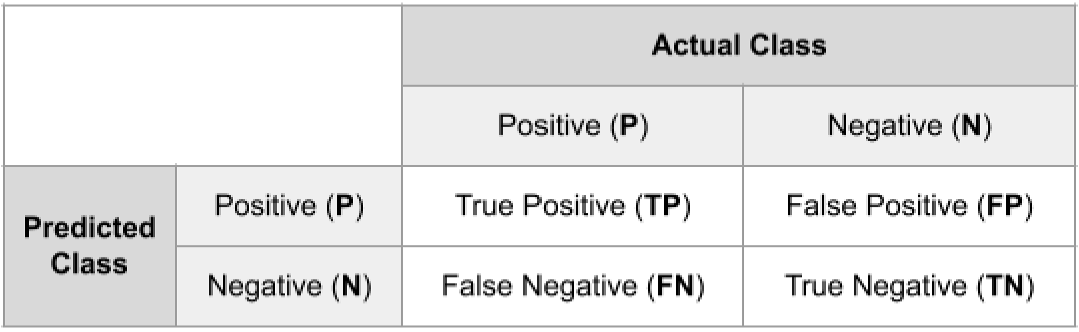
3.7.5. Mean Average Precision(mAP):

To evaluate algorithms like "Fast R- CNN, YOLO, Mask R-CNN," etc., "Mean Average Precision (mAP)" is employed. The "average accuracy (AP)" frequencies (mean) are calculated throughout recall levels scale from zero to 1.

mAP = $\frac{1}{n}\sum_{k=1}^{k=n} AP_k$ ……………………………………………………………(3.5)

3.7.6. Confusion Matrix:

The confusion Matrix essentially consists of the following indicators: “True_Positive (TP), True_Negative(TN), False_Positive(FP), and False_Negative(FN)”. And the following is a graphic illustration of it:



		Actual Class	
		Positive (**P**)	Negative (**N**)
Predicted Class	Positive (**P**)	True Positive (**TP**)	False Positive (**FP**)
	Negative (**N**)	False Negative (**FN**)	True Negative (**TN**)

Figure 3.9: Confusion Matrix Representation

3.8. SUMMARY:

We implemented YOLOv7 to detect motion after preprocessing the data, and FLASK to launch the web applications on the basis of these observations. An insight of algorithmic structure is also provided. We provide a few assessment strategies and their associated equations so that we can assess our system.

CHAPTER 4

4. RESULTS AND DISCUSSION:

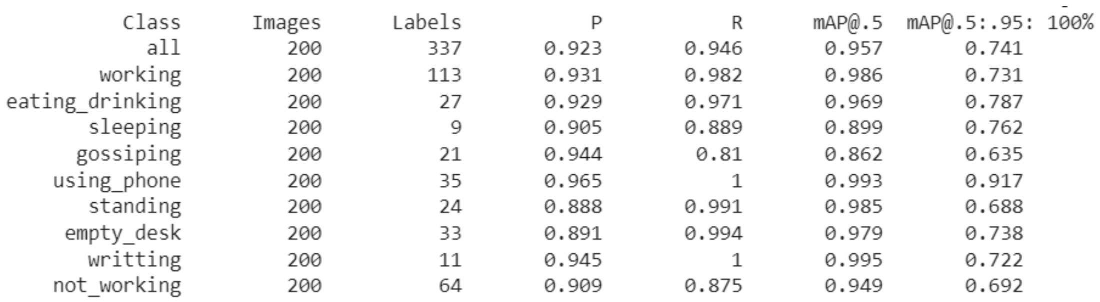
4.1. INTRODUCTION:

We defined the model's effective implementation after the data gathering and preparatory stage. Here, we'll discuss the model's final output following training.

4.2. RESULT:

Our custom-trained model has achieved a 95.7% mAP rate after accomplishing 60 epochs. After training that custom model, we have got a custom weight file(best.pt, last.pt, epoch_050.pt) which will help for further detection. For detecting purposes, We assign our confidence-level at 0.5 and image-size at 640. With the help of detect.py and custom-weighted files we always received the highest accuracy in our results.

Result with description at a glance-



```
          Class     Images     Labels          P          R     mAP@.5 mAP@.5:.95: 100%
            all        200        337      0.923      0.946      0.957      0.741
        working        200        113      0.931      0.982      0.986      0.731
eating_drinking        200         27      0.929      0.971      0.969      0.787
       sleeping        200          9      0.905      0.889      0.899      0.762
      gossiping        200         21      0.944       0.81      0.862      0.635
    using_phone        200         35      0.965          1      0.993      0.917
       standing        200         24      0.888      0.991      0.985      0.688
     empty_desk        200         33      0.891      0.994      0.979      0.738
       writting        200         11      0.945          1      0.995      0.722
    not_working        200         64      0.909      0.875      0.949      0.692
```

Figure 4.1: Result Description

4.2.1. YOLOv7 OUTPUT:

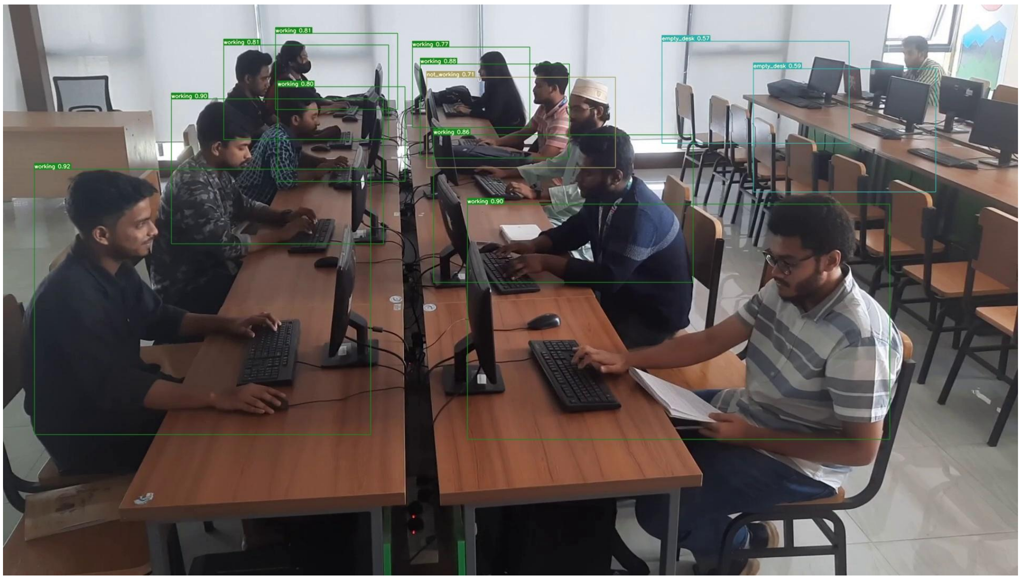


Figure 4.2: ‘YOLOv7 Object Detection’ Result-1

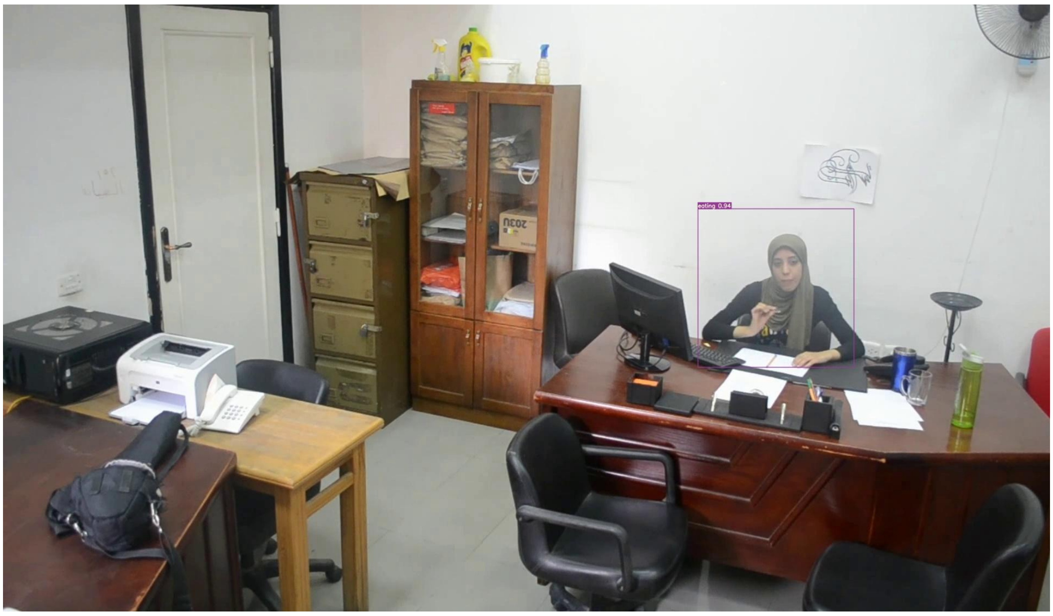


Figure 4.3: ‘YOLOv7 Object Detection’ Result-2

With the model's train and validation value associated, some plotting diagrams are produced.

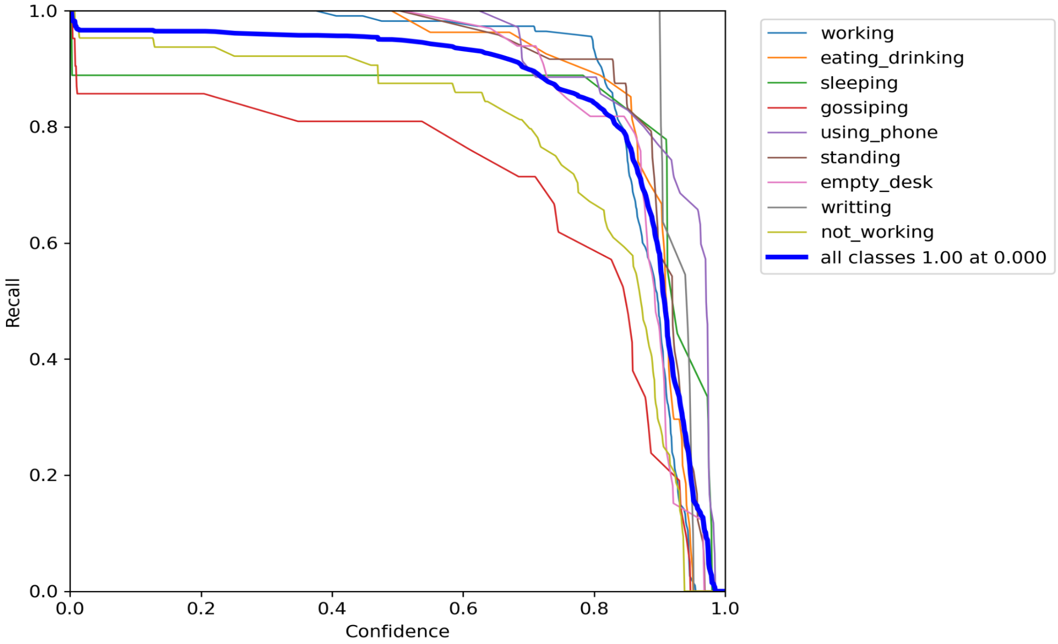


Figure 4.4: Recall_Curve

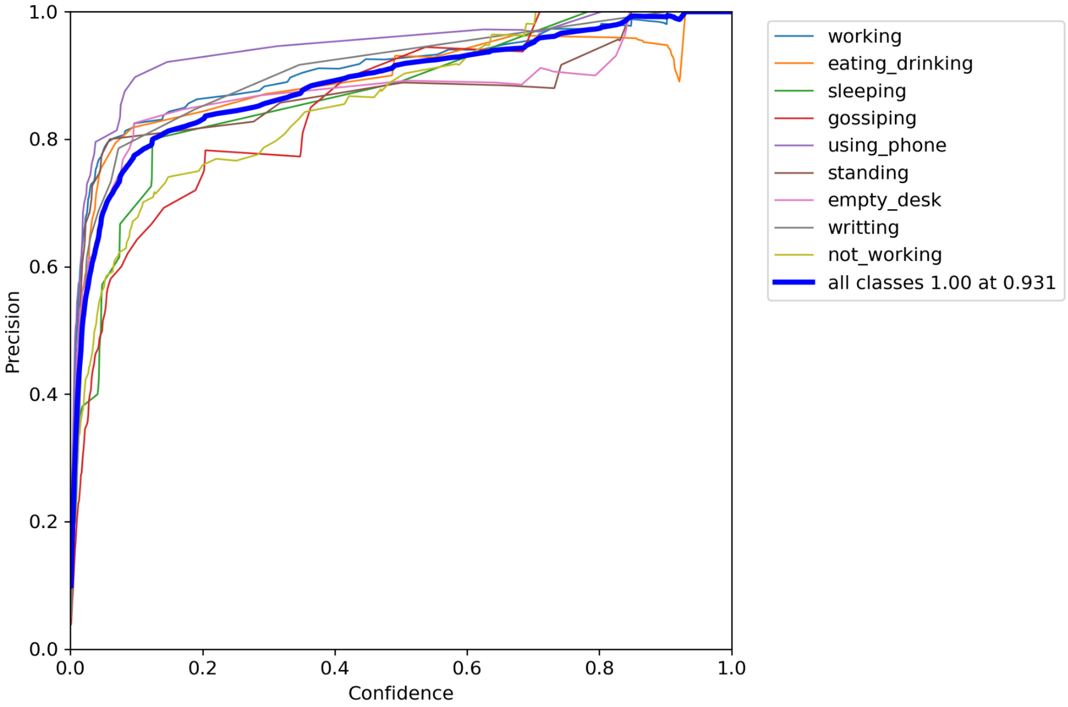


Figure 4.5: Precision_Curve

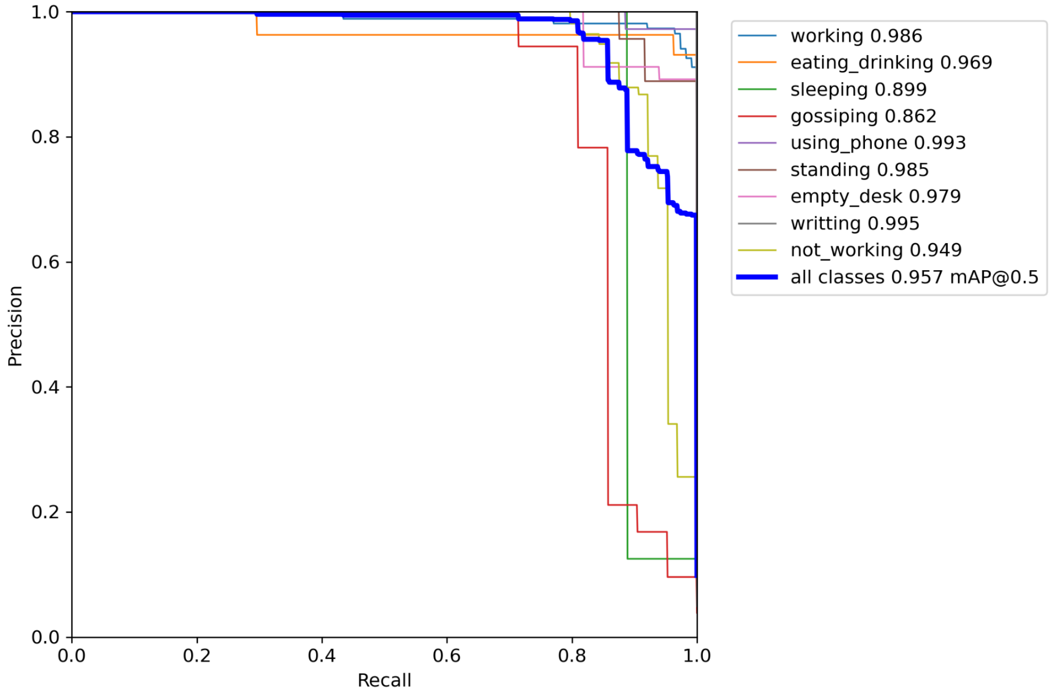


Figure 4.6: Precision-Recall_Curve

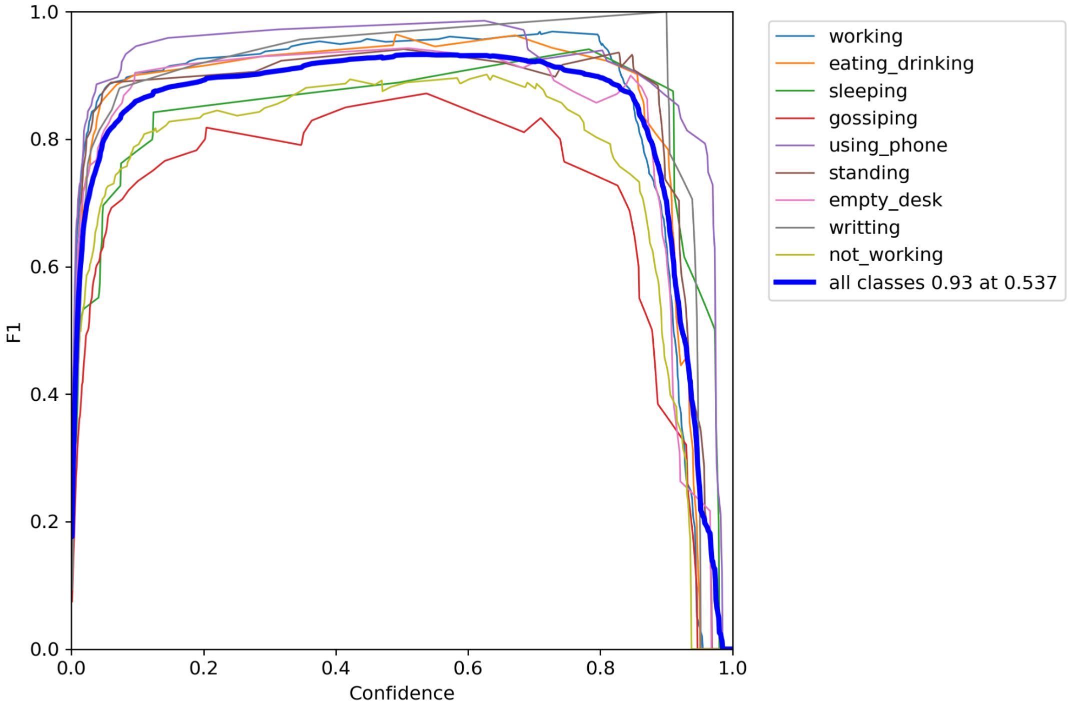


Figure 4.7: F1_Score

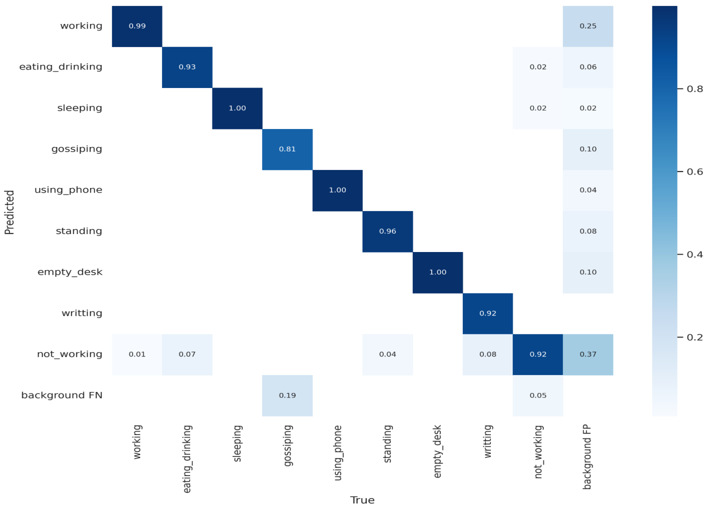


Figure 4.8: Confusion_Matrix

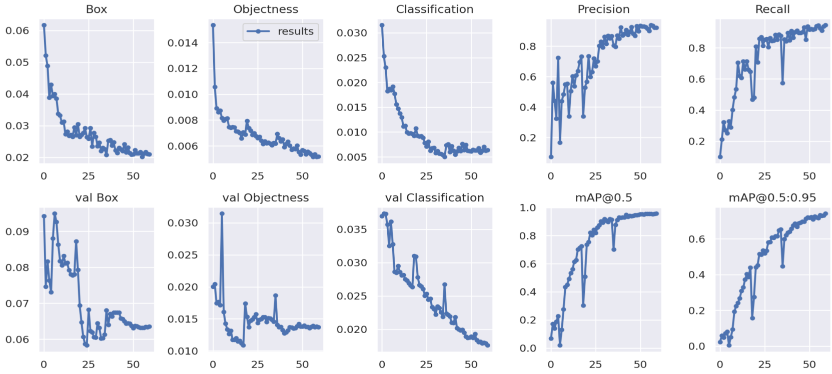


Figure 4.9: Result at a Glance

4.2.2. FLASK OUTPUT:

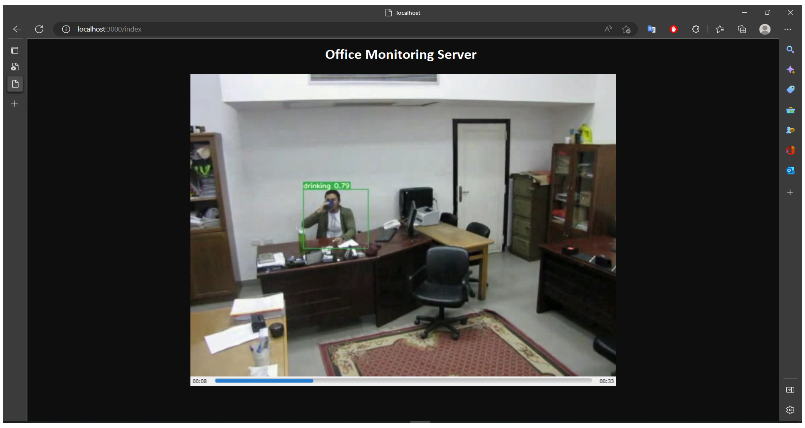


Figure 4.10: FLASK Hosting Output

4.3. DISCUSSION:

According to YOLOv7 documentation and their claims of greater object-detection accuracy, our specially trained ‘YOLOv7 model’ is displaying excellent outcomes in detecting different office members’ movements. By categorizing them as "working, not_working, eating_drinking, sleeping, gossiping, using_phone, standing, empty_desk, and writing," we were able to discriminate between different levels of physical activity. In comparison to our base article, our ‘YOLOv7 model’ appears to be significantly more effective. since sensors are not required. Additionally, we used FLASK to deploy the model into the server that our base paper couldn’t achieve.

4.4. SUMMARY:

We addressed the ultimate outcomes of our proposed approach in this part. The same position was captured on many recordings (CCTV video) that we gathered from different sources. This is why our model has been trained so well and has such high detection precision. In this study, FLASK is employed for deployment whereas ‘YOLOv7’ is applied to differentiate between postures.

CHAPTER 5

5. CONCLUSION AND RECOMMENDATION:

Unutilized office hours create difficulties to achieve the ultimate goal of the company at the end of the year. Additionally, it is impossible to use the company's workforce to its maximum potential.

In contrast to previous "object-detection" techniques, YOLOv7 offers more accurate results and more dependability for tracking purposes. The suggested model divides office employees' daily activities during work hours into nine categories using YOLOv7 as a predictor. Individuals who are focusing on their assigned tasks are referred to as proactively active members (working and writing), whereas those who are not functioning are referred to as inactive individuals (not_working, eating_drinking, sleeping, gossiping, using_phone, standing, empty_desk).

5.1. FINDINGS AND CONTRIBUTIONS:

In this recommended content, an intelligent monitoring system is utilized to spot employees working on their allocated tasks during business hours. On the Examination dataset, a classifier built that uses the YOLOv7 deep learning model is generated and evaluated. It has a 95.7% overall accuracy. The recommended model achieves better than the existing model because it can monitor all office employees' activities simultaneously and uses less processing power to provide the desired result than earlier versions.

5.2. RECOMMENDATIONS FOR FUTURE WORK:

In the office, this strategy is quite effective. But it also has certain shortcomings. It doesn't apply to any form of conference space. The resolution of the camera has a slight influence on reliability. Despite our usage of blurry images and augmentation, our object-detection technique is unable to demonstrate high precision in a blurry context. Moreover, our model cannot generate any kind of ID for creating any kind of tracking records Therefore, these are the area for improvement that we can work on soon.

CHAPTER 6

6. REFERENCES:

[1] Yang, F., Zhang, X., & Liu, B. (2022). Video object tracking based on YOLOv7 and DeepSORT. *arXiv preprint arXiv:2207.12202*.

[2] Wang, K., Wang, X., Lin, L., Wang, M., & Zuo, W. (2014, November). 3d human activity recognition with reconfigurable convolutional neural networks. In *Proceedings of the 22nd ACM international conference on Multimedia* (pp. 97-106).

[3] Casserfelt, K. (2018). A Deep Learning Approach to Video Processing for Scene Recognition in Smart Office Environments.

[4] Serpush, F., & Rezaei, M. (2020). Complex human action recognition in live videos using hybrid FR-DL method. *arXiv preprint arXiv:2007.02811*.

[5] Kapoor, R., Jaiswal, A., & Makedon, F. (2022, June). Light-Weight Seated Posture Guidance System with Machine Learning and Computer Vision. In *Proceedings of the 15th International Conference on PErvasive Technologies Related to Assistive Environments* (pp. 595-600).

[6] Ma, C., Lee, C. K. M., Du, J., Li, Q., & Gravina, R. (2022). Work Engagement Recognition in Smart Office. *Procedia Computer Science*, *200*, 451-460.

[7] Kulikajevas, A., Maskeliunas, R., & Damaševičius, R. (2021). Detection of sitting posture using hierarchical image composition and deep learning. *PeerJ computer science*, *7*, e442.

[8] Liaqat, S., Dashtipour, K., Arshad, K., Assaleh, K., & Ramzan, N. (2021). A hybrid posture detection framework: Integrating machine learning and deep neural networks. *IEEE Sensors Journal*, *21*(7), 9515-9522.

[9] Jha, S., Seo, C., Yang, E., & Joshi, G. P. (2021). Real time object detection and trackingsystem for video surveillance system. *Multimedia Tools and Applications*, *80*(3), 3981-3996.

[10] Bathija, A., & Sharma, G. (2019). Visual object detection and tracking using Yolo and sort. *International Journal of Engineering Research Technology*, *8*(11).

[11] Patalas-Maliszewska, J., & Halikowski, D. (2020). A deep learning-based model for the automated assessment of the activity of a single worker. *Sensors*, *20*(9), 2571.

[12] Archana, M. C. P., Nitish, C. K., & Harikumar, S. (2022). Real time face detection and optimal face mapping for online classes. In *Journal of Physics: Conference Series* (Vol. 2161, No. 1, p. 012063). IOP Publishing.

[13] Yan, X., & Su, X. (2009). *Linear regression analysis: theory and computing*. world scientific.

[14] Chen, Y., & Xue, Y. (2015, October). A deep learning approach to human activity recognition based on a single accelerometer. In *2015 IEEE international conference on systems, man, and cybernetics* (pp. 1488-1492). IEEE.

[15] Hammerla, N. Y., Halloran, S., & Plötz, T. (2016). Deep, convolutional, and recurrent models for human activity recognition using wearables. *arXiv preprint arXiv:1604.08880.*

[16] Bewley, A., Ge, Z., Ott, L., Ramos, F., & Upcroft, B. (2016, September). Simple online and realtime tracking. In *2016 IEEE international conference on image processing (ICIP)* (pp. 3464-3468). IEEE.

[17] Chen, L., Ai, H., Zhuang, Z., & Shang, C. (2018, July). Real-time multiple people tracking with deeply learned candidate selection and person re-identification. In *2018 IEEE international conference on multimedia and expo (ICME)* (pp. 1-6). IEEE.

[18] Bochkovsky, A., Wang, C. Y., & Liao, H. &. M. (2020). Yolov4: Optimal speed and accuracy of object detection. *arXiv preprint arXiv:2004.10934*.

[19] Ni, B., Pei, Y., Liang, Z., Lin, L., & Moulin, P. (2013, April). Integrating multi-stage depth-induced contextual information for human action recognition and localization. In *2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)* (pp. 1-8). IEEE.

[20] Chaquet, J. M., Carmona, E. J., & Fernández-Caballero, A. (2013). A survey of video datasets for human action and activity recognition. *Computer Vision and Image Understanding*, *117*(6), 633-659.

[21] Cheng, Z., Qin, L., Huang, Q., Jiang, S., Yan, S., & Tian, Q. (2011, November). Human group activity analysis with fusion of motion and appearance information. In *Proceedings of the 19th ACM international conference on Multimedia* (pp. 1401-1404).

[22] Brendel, W., & Todorovic, S. (2011, November). Learning spatiotemporal graphs of human activities. In *2011 International Conference on Computer Vision* (pp. 778-785). IEEE.

[23] Koppula, H., & Saxena, A. (2013, May). Learning spatio-temporal structure from rgb-d videos for human activity detection and anticipation. In *International conference on machine learning* (pp. 792-800). PMLR.

[24] Packer, B., Saenko, K., & Koller, D. (2012, June). A combined pose, object, and feature model for action understanding. In *2012 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1378-1385). IEEE.

[25] Sadanand, S., & Corso, J. J. (2012, June). Action bank: A high-level representation of activity in video. In *2012 IEEE Conference on computer vision and pattern recognition* (pp. 1234-1241). IEEE.

[26] Scovanner, P., Ali, S., & Shah, M. (2007, September). A 3-dimensional sift descriptor and its application to action recognition. In *Proceedings of the 15th ACM international conference on Multimedia* (pp. 357-360).

[27] Zhu, S. C., & Mumford, D. (2007). A stochastic grammar of images. *Foundations and Trends® in Computer Graphics and Vision*, *2*(4), 259-362.

[28] Yang, X., Zhang, C., & Tian, Y. (2012, October). Recognizing actions using depth motion maps-based histograms of oriented gradients. In *Proceedings of the 20th ACM international conference on Multimedia* (pp. 1057-1060).

[29] Wang, X., Lin, L., Huang, L., & Yan, S. (2013). Incorporating structural alternatives and sharing into hierarchy for multiclass object recognition and detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 3334-3341).

[30] Lin, L., Wu, T., Porway, J., & Xu, Z. (2009). A stochastic graph grammar for compositional object representation and recognition. *Pattern Recognition*, *42*(7), 1297-1307.

[31] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).