Student Face Detection and Identification from CCTV Footage

Submitted By

Darji Akshatkumar Hiteshbhai

23MCD001

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

INSTITUTE OF TECHNOLOGY

NIRMA UNIVERSITY

AHMEDABAD-382481

May 2024

Student Face Detection and Identification from CCTV Footage

Minor Project - I

Submitted in partial fulfillment of the requirements

for the degree of

Master of Technology in Computer Science and Engineering (Data Science)

Submitted By

Darji Akshatkumar Hiteshbhai

(23MCD001)

Guided By

Dr. Usha Patel

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

INSTITUTE OF TECHNOLOGY

NIRMA UNIVERSITY

AHMEDABAD-382481

May 2024

Certificate

This is to certify that the minor project entitled **"Student Face Detection and Identification from CCTV Footage"** submitted by **Darji Akshatkumar (Roll No: 22MCD001)**, towards the partial fulfillment of the requirements for the award of degree of Master of Technology in Computer Science and Engineering (Data Science) of Nirma University, Ahmedabad, is the record of work carried out by him under my supervision and guidance. In my opinion, the submitted work has reached a level required for being accepted for examination. The results embodied in this major project part-I, to the best of my knowledge, haven't been submitted to any other university or institution for award of any degree or diploma.

Dr. Usha Patel
Guide & Associate Professor,
CSE Department,
Institute of Technology,
Nirma University, Ahmedabad.

Dr. Swati Jain
Associate Professor,
Coordinator M.Tech - CSE (Specialization)
Institute of Technology,
Nirma University, Ahmedabad

Statement of Originality

I, **Darji Akshatkumar**, **23MCD001**, give undertaking that the Minor Project entitled "**Student Face Detection and Identification from CCTV Footage**" submitted by me, towards the partial fulfillment of the requirements for the degree of Master of Technology in **Computer Science & Engineering (Data Science)** of Institute of Technology, Nirma University, Ahmedabad, contains no material that has been awarded for any degree or diploma in any university or school in any territory to the best of my knowledge. It is the original work carried out by me and I give assurance that no attempt of plagiarism has been made.It contains no material that is previously published or written, except where reference has been made. I understand that in the event of any similarity found subsequently with any published work or any dissertation work elsewhere; it will result in severe disciplinary action.

_____

Signature of Student

Date:

Place:

Endorsed by

Dr. Usha Patel

(Signature of Guide)

Acknowledgements

It gives me immense pleasure in expressing thanks and profound gratitude to **Dr. Usha Patel**, Associate Professor, Computer Engineering Department, Institute of Technology, Nirma University, Ahmedabad for his valuable guidance and continual encouragement throughout this work. And a special thanks to **Dr. Swati Jain**, Coordinator of Data-Science for all kinds of permission to access the CCTV footage and the support for this project.

It gives me an immense pleasure to thank **Dr. Madhuri Bhavsar**, Head of Computer Science and Engineering Department, Institute of Technology, Nirma University, Ahmedabad for his kind support and providing basic infrastructure and healthy research environment.

A special thank you is expressed wholeheartedly to **Dr. R. N. Patel**, Hon'ble Director, Institute of Technology, Nirma University, Ahmedabad for the unmentionable motivation he has extended throughout course of this work.

I would also like to thank the Institution, all faculty members of Computer Engineering Department, Nirma University, Ahmedabad for their special attention and suggestions towards the project work.

Akshat Darji
23MCD001

Abstract

The title "Student Face Detection and Identification from CCTV Footage" itself suggests that the project involves the applications of Deep Learning and Computer Vision for the purpose of Face Detection and Face Verification tasks. In this study, the major focus is on Detection and Verification of the detected face from CCTV footage. During this study, various challenges are faced in the face detection and verification part such as occlusion, low-resolution image that is detected from the CCTV footage, Face Detection from CCTV footage for the Overlapped section, Preprocessing of the detected image to improve the quality of the detected image. To overcome the challenges multiple Deep learning and computer vision applications are used. For the Face Detection task, YOLOv8 is used and it provides better accuracy than other object detection algorithms, For the Face Verification task Siamese Network is used which is the combination of the Embedding Layer, Disance layer, and Classification layer. For the Embedding Layer in the Siamese network, VGG16 pre-trained model is used. The Siamese network is followed by the Recognization part to fetch the student information for each successful verification by the Siamese network. Experimental results show that for the face detection and verification task YOLOv8 and Siamese Network work better than any other application of deep learning and computer vision. The proposed methodology gives 87% accuracy for fetching the information of detected faces from the Student Database.

Abbreviations

ML	Machine Learning.
DL	Deep Learning.
CNN	Convolution Nerual Network.
VGG	Visual Geometry Group
CV	Computer Vision
YOLO	You Only Look Once
MTCNN	Multi-Task Cascaded Convolutional Neural Networks

Contents

List of Figures

Chapter 1

Introduction

1.1 Importance of student face detection and identification in educational institutes

"Student face detection and identification from CCTV footage" plays a crucial role in detecting and identifying the students by their faces. The main objective of this project is to detect students' faces from CCTV footage and match them with the faces in the student database. Once a match is made, the system displays the respective student's information, making it easy and convenient to access student information from the CCTV footage. There are some key points in which this project is useful and plays a crucial role i.e. campus security and preventing academic fraud.

1.2 Knowledge Discovery Process

"Student Face Detection and Identification from the CCTV footage" includes various steps such as problem definition and understanding, Literature Review, Data Collection and Integration, Data Prepossessing, and Model Training. These are the steps to create an integrated system which is a combination of YOLOv8(For face detection) and Siamese Network(For face verification) followed by face recognition.

Knowledge Discovery Process Steps:-

- **Problem Definition and Understanding**

  In this project, there are two main concepts are used computer vision and Deep

learning techniques to implement student face detection and Identification from CCTV footage. The title "Student face detection and identification from CCTV footage" itself indicates that the problem is divided into two stages first stage is to detect the face from the CCTV footage and the second stage is Face verification that is followed by face recognition. Here for the Face Detection YOLO algorithm is used and for the face verification task there is one network called the Siamese network is used which is also a combination of various deep learning techniques.

- **Literature Review**

  Conducted a literature review on the current methodologies, related technologies, and existing models that help us to develop an integrated system "Student face detection and identification from CCTV footage" that can detect the faces from the CCTV footage and on the successful verification of that detected face it fetches the respective student information from the database.

- **Data Collection and Integration**

  For training the model data collection and integration is the key part of any project. Here mainly two kinds of databases are created. As mentioned in the Problem definition part this project is the integration of Face Detection and Identification. Here for Face Detection YOLO algorithm is used, we use the YOLO(version8) pre-trained model for face detection so there is no need for a dataset to train the YOLO algorithm. But for the face verification part, we use a Siamese network that is trained on Positive and Negative pairs of data so here we require a dataset. For the Recognition part as well we need the dataset to fetch the details it contains the Student image and with that, there is one text file that contains the student information i.e. student's name, and roll number.

  Figures 1.1 and 1.2 show the Dataset structure for the Siamese network and the Face recognition part respectively.

  As shown in **Fig 1.1** Siamese network is trained on Positive and negative pairs so it can validate the similarity between two images that is used in the Face verification part for this project. Here Positive images means all the images which is in

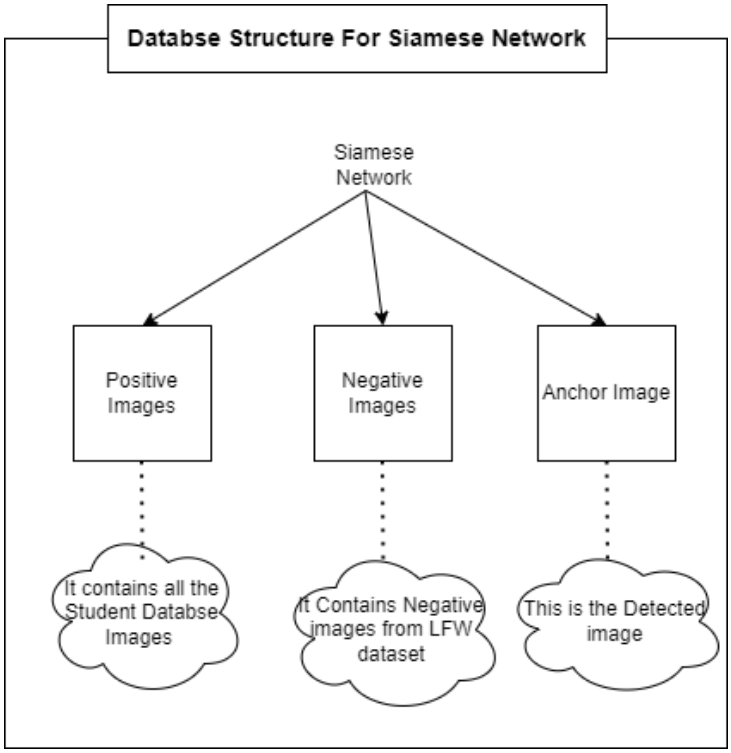

Figure 1.1: Dataset structure for Siamese Network(For Training the Siamese Network)

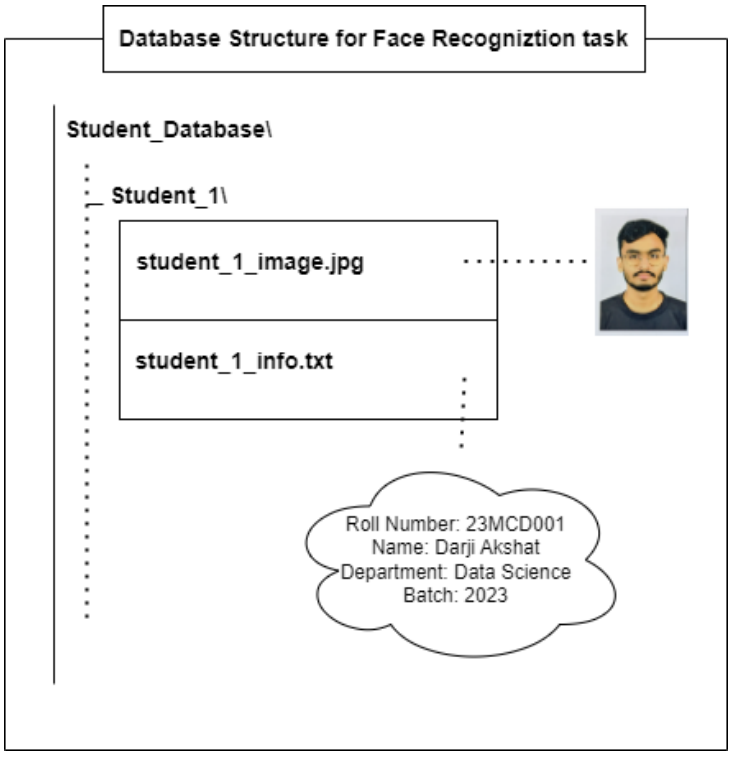

Figure 1.2: Dataset structure for Recognition part(Used to fetch the Information)

our student database are authorized images and Negative images means the images that are not in our student database are unauthorized images. An anchor image is the image that is detected from the CCTV footage. Positive images and Nega-

tive images are then used to create the pairs of (positive, positive), and (positive, Negative) which are feed to the Siamese Network for training purposes.

As shown in **Fig 1.2** the Recognition part is used to fetch the details of the student on each successful identification (using Siamese Network). In `Student_Database` Folder contains each student Folder i.e. `Student_1, Student_2, etc`, and each individual student folder contains `student_1_image.jpg and student_1_info.txt` as shown in the above figure.

- **Data Prepossessing**

  As mentioned in the problem Definition part the detected faces are preprocessed and faded to the Siamese network. Now the detection part is done using the YOLO algorithm and the quality of detected images is very low, so for better accuracy, preprocessing of detected images is required. The dimension of the detected images from CCTV footage is 100x100, so we use the interpolation technique to increase the quality of the detected image. The basic method that we use for the interpolation technique is nearest neighbor where the output pixel value is estimated by considering the nearest pixel's value to the specific input coordinates.

  Siamese network are trained on the Positive and negative pairs so for that task from the positive and negative images that we collected we made the pairs of images and converted them into 100x100 images for a better understanding.

1.3 Aim Of the Project

Student Face Detection and Identification from CCTV Footage is a topic that involves the application of Deep learning and computer vision to address the challenges associated with face detection and identification followed by face recognition. Detecting the faces from the CCTV footage is a critical aspect of many applications, including surveillance. However, it's challenging because it requires understanding and research on the latest models introduced in the field of face detection and identification in deep learning. Deep learning and computer vision such as YOLO and Siamese network shown remarkable success in such tasks. This project aims to localize and recognize student faces from CCTV footage, Displaying bounding boxes with corresponding roll numbers upon successful identification.

1.4 Project Scope

Developing the student face detection and identification system which utilizes the YOLO (For detecting the face) and Siamese Network(for face verification) is the scope of this project. The main objective is to retrieve the student information from the Student Database on each successful identification and showcase the student information on the bounding box is the major focus. real-time face detection using YOLO, Face verification which utilizes the Siamese network to match the detected faces with the images in the database, and face recognition for identifying and retrieving the student information, these are the main features provided by the Student face detection and identification from CCTV footage system.

1.5 Methods

Various computer vision such as YOLO, and HaarCascade are used to detect the face features from any video frame, and deep learning applications such as convolution neural network(CNN) and transfer Learning are used to create the siamese network for calculating the similarity between detected image and student database image. Below are the methods that are used in "Student Face Detection and Identification System from CCTV footage".

1.5.1 HaarCascade Algorithm

A Haar Cascade is an object detection method used for object detection in images. A Haar Cascade classifier is a machine learning-based approach proposed by Paul Viola and Michael Jones in their 2001 paper [1] that employs a cascade of classifiers to identify objects based on Haar-like features. Haar-like features are simple rectangular patterns that can be used to detect changes in intensity in a nearby image region. By combining several Haar-like features, a classifier can identify things of interest efficiently. Haar Cascade classifiers can be trained on both positive and negative examples of the object to be identified. As a result, the classifier learns what distinguishes an image from the background from many training examples.

1.5.2 YOLO Algorithm

YOLO, which stands for You Only Look Once, is a widely popular object detection algorithm. It works by dividing the slide of an input image into a grid as it predicts bounding boxes and class probabilities for the grid cells. Contrary to most object detection methods, this model does not use region proposal algorithms to locate objects, followed by feeding them to a classifier. On the contrary, YOLO approaches the problem in its entirety, delivering bounding boxes and class predictions simultaneously. As a result, with the single pass that the algorithm takes through the network to predict and locate the objects from an image, YOLO does this exceptionally fast and thus is suitable for real-time object detection systems such as automation vehicles, and robotics.

The YOLO algorithm has different versions such as YOLOv1, YOLOv2, etc. YOLOv3 and YOLOv5 are widely used versions and give the state of art results but YOLOv8 gives a more convenient result than the previous version of it.

1.5.3 Siamese Network

Siamese Network is used for the Face Verification task to compare the similarity between two images Siamese network is more convenient than other techniques. Siamese Network is a combination of the Embedding layer and Distance Layer that is followed by the classifier. The Embedding Layer is a simple convolution neural network that is used to extract the features of input images and the distance layer is used to find the distance between those features. A combination of the Embedding Layer and Distance layer followed by the classifier makes the Siamese network for the face verification task.

Chapter 2

Literature Survey

Koch et al. [2] discussed the problem of learning good features for machine learning in the face of a lack of training data, such as the case of one-shot learning. It tackles this by proposing siamese neural networks, which inherently identify similarities between inputs, thus allowing for good generalization to novel classes. With the use of convolutional architectures, the implementation demonstrated strong performance in one-shot classification, outperforming other deep learning approaches.

Stalin et al. [3] presented a novel solution for real-world face recognition applications, addressing common issues such as long-distance identification or CCTV footage inference. Additionally, by utilizing one-shot learning, the model is designed to ignore natural constraints that are currently impossible to avoid, like face masks during the COVID-19 outbreak. Based on Siamese network design, trained with triplet loss function, the solution offered increased adaptability and accuracy in identifying people from limited training sets, representing a valuable contribution to the implemented face recognition solutions.

Celine et al. [4] reiterate the significance of Closed-Circuit Television systems in enhancing security in different institutions. In addition to the observation settings, the paper has also focused on the underlying needs and issues that make forensic analysis using CCTV challenging in actual situations, including bank robberies. The distance measurements and the camera viewing angles and their effects on the identification process were some of the considered determining factors. The facial recognition model recom-

mended involves two solid stages: the first involves acquiring a detailed database of faces from different subjects; the second involves obtaining and enhancing target faces from the available video records to use in comparing them to a detailed face database.

Zhang et al. [5] propose an algorithm for small-sample face recognition. The algorithm is based on a new Siamese network, which is modeled into SiameseFacel. Face pairs are used to map each of the face images to the target space, and the norm distance reflects the semantic distance of the new input space. Moreover, a lighter model named SiameseFace2 was designed to reduce the number of network parameters while maintaining model accuracy. In addition, the augmentation of training samples was artificially used to improve the accuracy of recognition by increasing the number of training samples. The experimental results of AR and LFW data sets show that the new Siamese network model combined with our contrastive loss function can effectively improve the accuracy of face recognition.

Satyagama et al. [6] developed a system that can detect faces even in low-quality photos, such as those obtained from security cameras or older webcams. It employs specialized networks known as Siamese networks to aid with facial recognition, picture quality improvement, and determining who is who. The researchers had three main objectives: create a full facial recognition system for low-quality photographs, test several approaches to discover what works best, and then include the best-performing ones into the final system. So, what are the results? They achieved a staggering 93% accuracy for 36x36 photos, 84% for 24x24, and 56.2% for 12x12 images, all in just 0.086 seconds.

Powale et al. [7] used deep learning to address the difficulty of detecting humans in low-quality CCTV or webcam footage. Their proposed approach, a convolutional neural network (CNN), has excellent performance. The CNN is trained using 6667 face photos from 62 participants and achieves exceptional accuracies: 89.99% for training, 88.45% for validation, and 86.03% for testing over 1599 images. Furthermore, when evaluated on the TinyFace dataset, which is designed for low-resolution facial recognition, it scores an impressive 84.55% accuracy, indicating its potential for real-world applications.

Viola et al. [1] describes a machine-learning approach for speedy and accurate visual object recognition. It introduces the "Integral Image" model, which allows the detector's features to be computed quickly. It effectively identifies key visual elements using an AdaBoost-based learning algorithm. Furthermore, it uses a cascade approach to quickly remove background regions while concentrating computation on probable object-like areas, similar to an object-specific focus-of-attention mechanism. Notably, it provides face identification rates equivalent to leading systems, running at a rapid 15 frames per second in real-time applications without depending on picture differencing or skin color detection.

Heidari et al. [8] performed two primary functions: validating if two photographs depict the same person and recognizing a specific face within a database. Despite its use, face recognition confronts limitations such as illumination and insufficient data. This research addresses these concerns by transfer learning in a Siamese network with two comparable CNNs. By evaluating pairs of face photos, the network determines if they belong to the same individual, attaining an astonishing 85.62% accuracy on the LFW dataset.

Tariyal et al. [9] examined four common face detection methods—Viola-Jones, MTCNN, SSD, and YOLO—according to how well they balance speed and accuracy. Viola-Jones is accurate, but not as fast as later algorithms. MTCNN is accurate, particularly with facial landmarks, but may not keep up in real-time. SSD strikes a fair compromise between speed and accuracy, but YOLO excels in real-time performance, which is ideal for fast-paced applications with limited resources. This analysis assists practitioners and researchers in selecting the best approach for their needs by demonstrating how face detection technology is always advancing for faster and more accurate results.

Detecting faces within computer vision is not only a hot topic but a cornerstone regarding security systems, video surveillance, and human-computer interaction. To solve this problem, researchers have already developed various methods such as Viola-Jones, RCNN, and SSD. Aung et al. [10] developed an improved face detection system by combining the popular YOLO algorithm with a pre-trained VGG16 convolutional neural network. The obtained results include an accuracy of over 85% .

Author	Year	Objective	Findings
Tariyal et al. [9]	2024	Comparative analysis of MTCNN, Viola-Jones, SSD, and YOLO face detection algorithm, and evaluating accuracy using MAP, F1-Score, Precision, and Recall.	After evaluating the accuracy using MAP, F1-Score, Precision, and Recall finding of this paper is that Viola-jones is fast but sacrifices accuracy compared to YOLO. MTCNN is very accurate but real-time detecting speed is very low. And YOLO is fast, provides good speed and good accuracy.
Stalin et al. [3]	2022	To achieve Face recognition from unique CCTV footage using the Siamese network. Use One-shot Classification to train the model for the minimal images.	Identified the method for long-distance CCTV footage scenarios. The proposed system can identify the face even the person with a face-mask, and training using the triplet loss function enhances the performance.
Aung et al. [10]	2021	To Develop an integrated system to improve the face detection system with YOLO and VGG16 CNN, proposed a system which is a combination of pre-trained VGG16 network with YOLOv2 algorithm for face detection for obtaining the higher accuracy.	Real-time face detection using YOLO algorithm and Achieved 95% precision on real-time live video and detection speed is also high but it requires GPU to proceed.
Heidari et al. [8]	2020	Face Recognition using a Siamese network with the concept of transfer learning in small datasets. For the dataset with few samples, increase the accuracy of face recognition. Using the CNN for the feature extraction in the Siamese Network	Proposed model improves the face recognition accuracy on small sample datasets, and achieves the accuracy of 85.62% on the LFW dataset.
Satyagama et al. [6]	2020	Developed a complete low-resolution face recognition system with higher accuracy using MTCNN.	Achieved 83% accuracy for 36x36 Resolution images for the face recognition task.
Powale et al. [7]	2020	Identify a person in low-resolution CCTV footage using Deep learning techniques such as CNN, use of pre-trained models such as FaceNet, VGG16, etc.	Achieved 80% testing accuracy with 1599 test images. And Tested on the Tinyface dataset with 82.55% classification accuracy.

Table 2.1: Literature Survay

Author	Year	Objective	Findings
Zhang et al. [5]	2018	Face recognition using siamese network with the CNN on a small dataset and improve the accuracy for the face recognition task on the dataset with the few samples	Improves the recognization task accuracy on AR and LFW dataset. Archives 85% accuracy on AR and 82.34% on LFW dataset.
Koch et al. [2]	2015	Address the challenges of learning good features for machine learning applications in cases with limited data availability, and the focus is on one-shot image recognition using deep convolution siamese networks for verification tasks.	For the One-Shot Classification proposed method achieved an accuracy of 90% for the limited data of any one category of MNIST dataset. The siamese neural network performs well on existing classifiers on one-shot classification tasks.
Viola et al. [1]	2001	machine-learning approach for speedy and accurate visual object recognition. introduces the "Integral Image" model, which allows the detector's features to be computed quickly.	effectively identifies key visual elements using an AdaBoost-based learning algorithm. Furthermore, it uses a cascade approach to quickly remove background regions while concentrating computation on probable object-like areas, similar to an object-specific focus-of-attention mechanism.

Table 2.2: Literature Survay

Chapter 3

Proposed Methodology

The proposed methodology involves deep learning and a computer vision approach for Student face detection and identification from CCTV footage. Conducted a Literature survey in Table2.2 and the two of YOLOv8 and Siamese Network with deep neural architecture is chosen to enhance the accuracy and efficiency in face detection and recognition system. YOLOv8 is used for the Face detection task from the CCTV footage. In this proposed methodology YOLOv8 is used because it is faster than other versions of YOLO and other object detection algorithms like Haarcascade [1]. So compared to all the object detection algorithms YOLOv8 gives better accuracy than other versions of YOLO and detection algorithms. Siamese Network is good for face verification tasks for the smaller dataset [2] and gives better accuracy for the verification task.

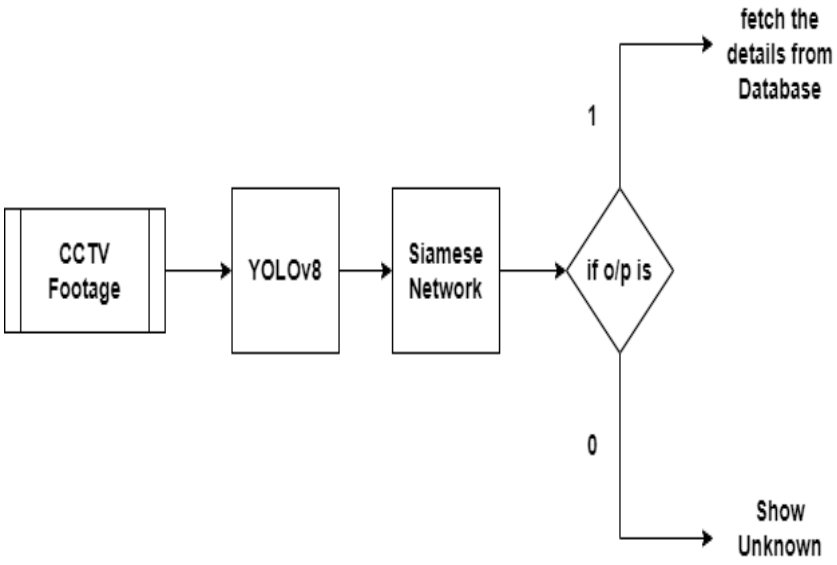

Figure 3.1: Basic Flow Chart of the Proposed Method

As shown in figure 3.9 the working of the developed system. The proposed system starts with the Face detection step which is followed by the Face. CCTV footage is given to the YOLOv8 algorithm which detects all the faces from the CCTV footage. In between YOLOv8 and Siamese Network preprocessing of Detected images is done. The detected faces are stored and preprocessed to the specific format which is required by the Siamese network. After detecting the face the detected face is given as input to the Siamese Network. Siamese Network compares that detected image and the Database images and gives the output as classifier means in 0 and 1. If the output of the Siamese Network is 0 then There is no similarity between those two images and if 1 then there is a similarity between those two images. If the output of Siamese Network is 1 then the recognition part takes place that retrieves the respective student information from the Student Database. And for each unsuccessful verification, it shows the unknown on the bounding box.

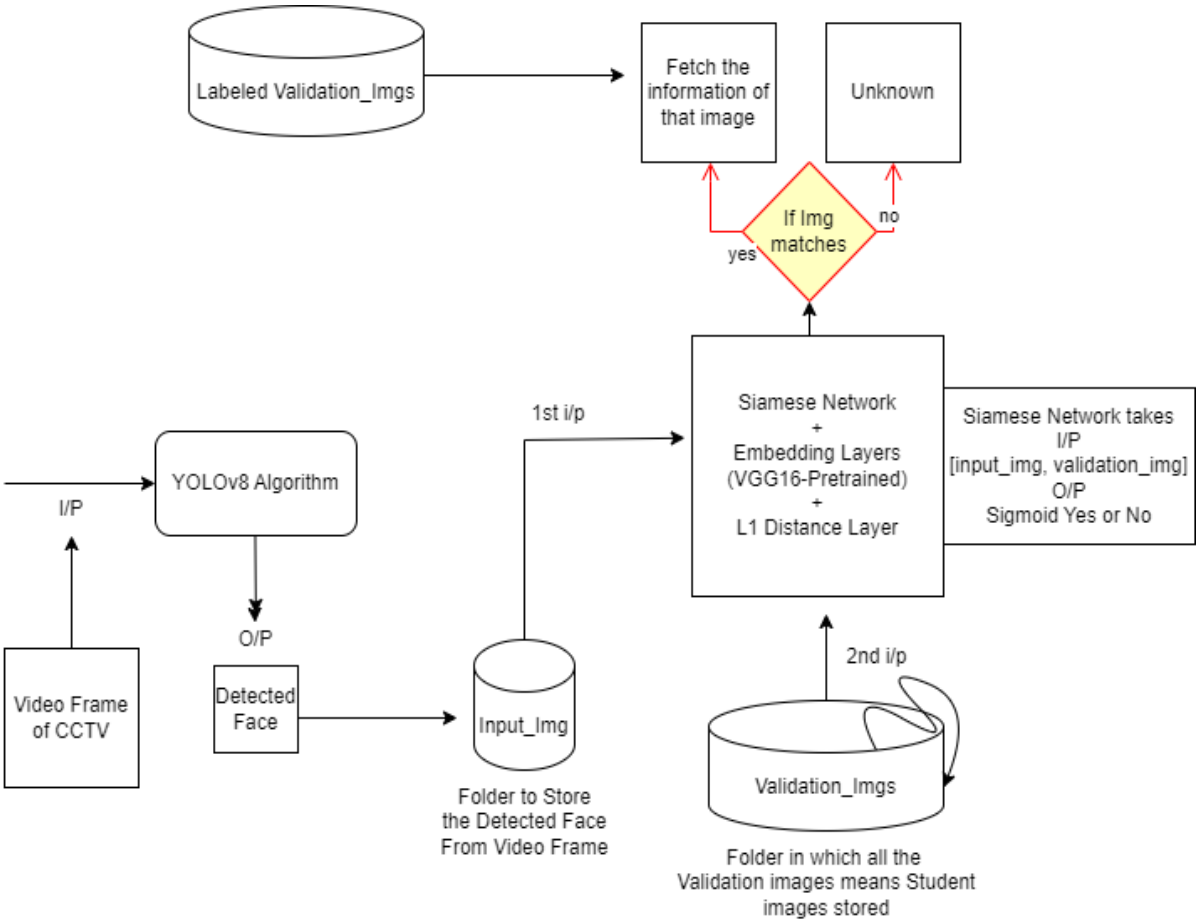

Figure 3.2: Proposed Method (Detailed)

3.1 Stages of Proposed Methodology

As shown in figure 3.9 "Student Face Detection and Identification from CCTV footage" is divided into two main stages Face Detection and Face Verification which is followed by face recognition.

3.1.1 Face Detection

For the Face Detection task, the YOLOv8 pre-trained model is used. YOLOv8 is a state-of-the-art object detection algorithm that is known for its higher accuracy. Other YOLO versions are dependent on the sliding window approach and region proposal networks, YOLOv8 uses a single neural network architecture that divides the input image into grids and predicts the bounding boxes and the class probabilities directly. So YOLOv8 detects the faces in the single pass through the neural network so it improves the speed of detection. In this project pre-trained model of YOLOv8 is used for the face detection task. YOLOv8 model can accurately detect the faces in unseen CCTV footage by predicting the bounding box around the detected face and also shows the confidence score for each face detection that is decided by the Iou(Intersection over Union). Iou is the measure of the overlap between two bounding boxes.

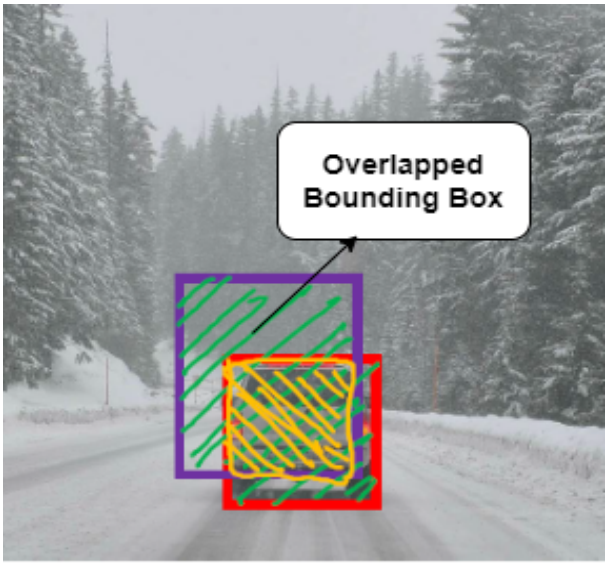

Figure 3.3: Intersection Over Union (IOU)

If the IOU is greater than 0.5 then only it detects the faces from the CCTV frames for each overlapped section.

3.1.2 Face Verification

For the Face Verification task, the Siamese network is used. Siamese Network is the efficient choice when we have fewer dataset [2]. Siamese Network is a combination of three components As shown in figure 3.2 Embedding Layer, Distance Layer, and Classification part. In the Embedding Layer VGG16 is used for the feature extraction, Distance Layer is used for finding the distance vector which is the input of the classification part which classifies that whether both the images are Similar or Not.

Here main use of the Siamese network in the project is to find the similarity between the detected image which is named input_image and the student image that is stored in the student's database which is named validation_image, Siamese Network finds the similarity between these two inputs and gives the output (classifier – either 1 or 0). So this is the work of the Siamese network in this project and then using the input of the Siamese network to fetch the labels of that image.

Expected Input and Output of Siamese Network

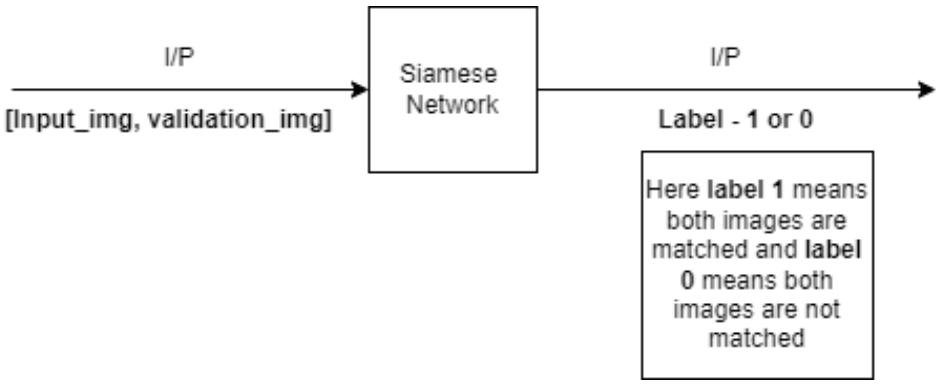

Figure 3.4: Expected Input & Output of Siamese Network

In Siamese Network data is divided into three parts Anchor Image, Positive Image, and Negative Image. Anchor images are the images that are detected in the frame. The positive image is the image that contains all the authorized images of the student, and Negative images are the images that are not authorized means not part of the student database here used the LFW dataset for the Negative Images. Siamese Network is trained on pairs of images. there are mainly two combinations of pairs on which Siamese network is trained which are (Anchor image + Positive image) and (Anchor image + Negative Image)

CASE-I:- Anchor Image + Positive Image

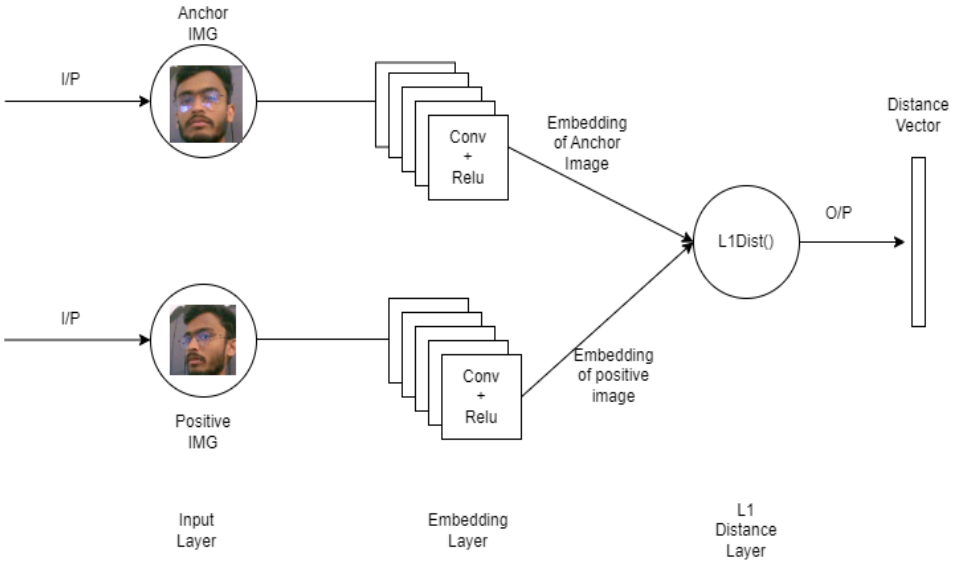

Figure 3.5: pair of Anchor image ad Positive image

Case 1 suggests that if there is a pair of Anchor image and Positive image (means image from Student Database) which means that both the images are same and the classifier gives me **label 1**.

CASE-II:- Anchor Image + Positive Image

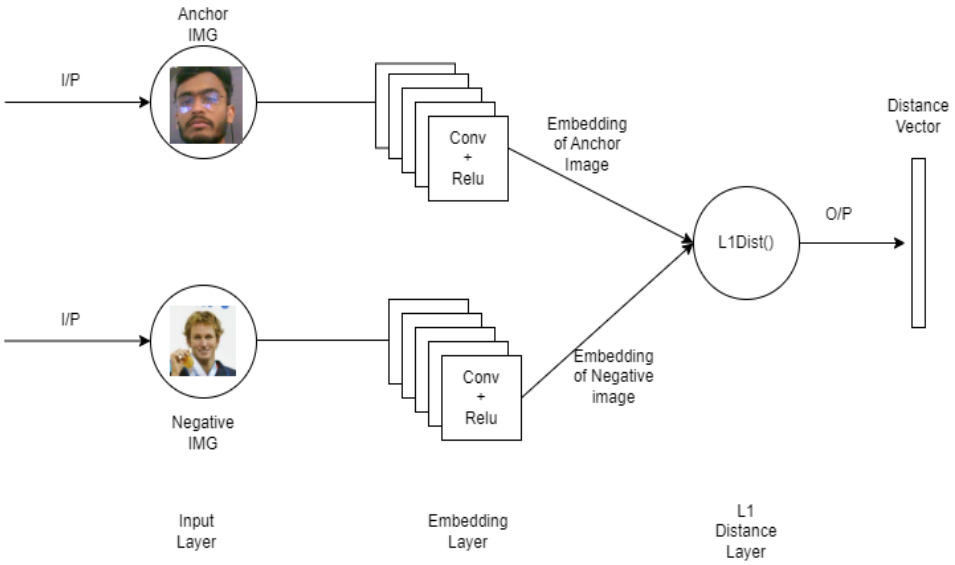

Figure 3.6: pair of Anchor image ad Negative image

Case 2 suggests that if there is a pair of Anchor image and Negative image (means

image from LFW dataset - Unauthorized image) which means that both the images are not same and the classifier gives me **label 0**.

3.2 Components of Siamese Network

As discussed there are three main components of Siamese network Embedding Layer, Distance Layer, and Classifier Part.

3.2.1 Embedding Layer

Embedding Layer is used to extract the features of input images [input_image, validation_image]. In the embedding Layer VGG16 model is used for the feature extraction task. VGG16 has 16 layers, in which 13 layers are convolution layers and 3 are fully connected layer as shown in Fig 3.7. The Input of the Embedding layer is 100x100 RGB images[input_image, validation_image]. The output of the Embedding layer is a Feature vector of size 4096 which is input to the next layer of Siamese network which is Distance Layer.

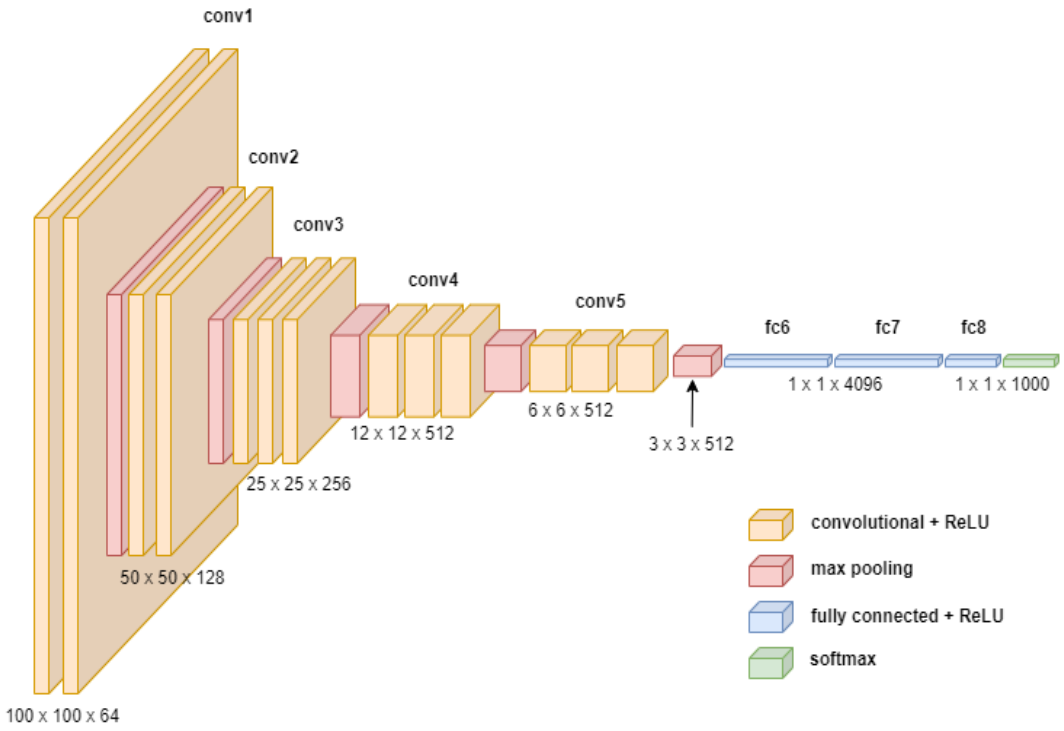

Figure 3.7: Fine tuned VGG16 Architecture (Embedding layer)

3.2.2 Distance Layer

The output of the Embedding layer is the feature vector of size 4096. so two feature vectors are generated one is of input_image and another is of vaidation_image two feature vectors of size 4096 are the input of the Distance layer. It is the function that finds the similarity between embedding generated by the Embedding Layer of input_image and validation_image and gives the distance vector that input to the Classification part as shown in Fig 3.8. Distance Layer finds the Euclidian distance between input_image embedding and validation_image embedding.

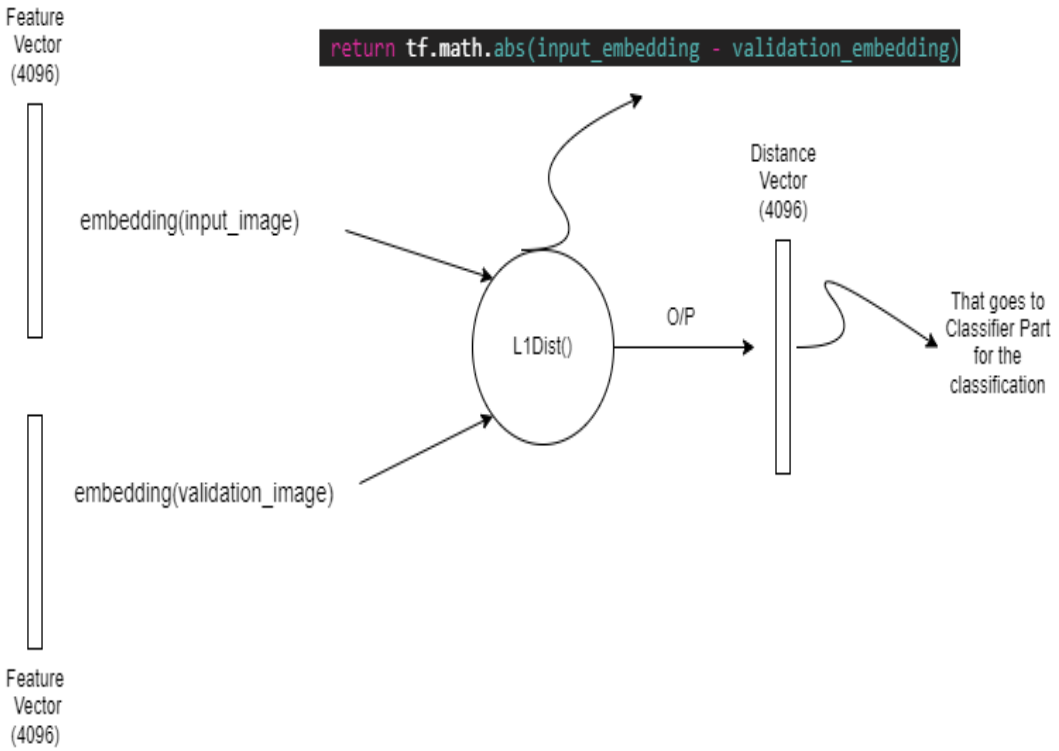

Figure 3.8: Distance Layer

3.2.3 Classifier Part

The main objective of the classification part in the Siamese network is to classify the similarity in 0 and 1. The output of the Distance layer is the Distance Vector of size 4096 which is input to the classifier and gives the output whether both the images are same or not as shown in Fig 3.9. If both the images are the same then it should predict Label-1 and if both the images are not the same then it should predict label-0. The sigmoid activation function is used in the Distance Layer because at the end we need the output in Binary format i.e. in YES or NO format.

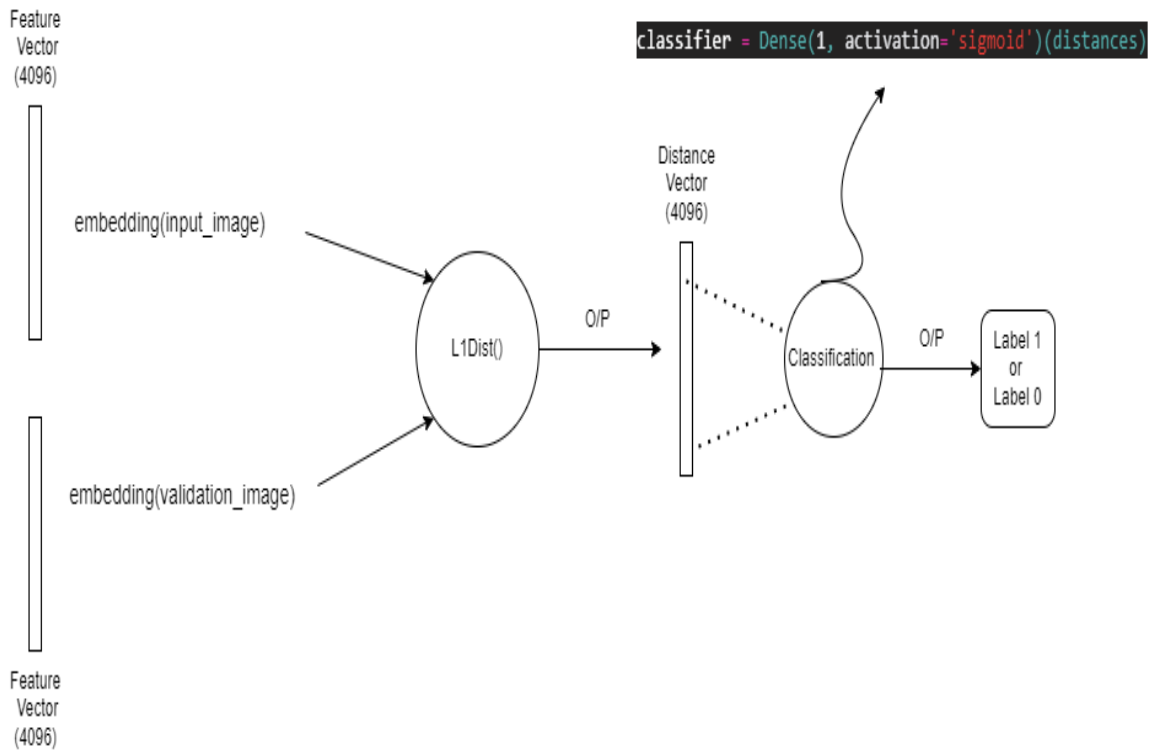

Figure 3.9: Classification Layer

Using above three main components Siamese Network is created. First input_image and validation_image are gone into the Embedding layer which generates the feature vector of size 4096, which is provided to the Distance Layer which finds the Euclidian Distance between input_image Embedding and validation_image Embedding which generates the Distance Vector of size 4096 which is input to the Classification part which classifies the result in YES and NO means whether both images are same or not same. **Below table 3.1 shows the summary of Siamese Network.**

Layer (type)	Output Shape	Param	Connected to
input_image (InputLayer)	[(None,100,100,3)]	0	[]
validation_image (InputLayer)	[(None,100,100,3)]	0	[]
embedding (Functional)	[(None,4096)]	38960448	['input_image[0][0]', 'validation_image[0][0]']
l1_dist_1 (L1Dist)	[(None,4096)]	38960448	['embedding[0][0]', 'embedding[1][0]']
dense_2 (Dense)	[(None,1)]	4097	['l1_dist_1[0][0]']

Table 3.1: Literature Survay

Total parameters in siamese network is **38964545 (148ḋ4 MB)**, Total trainable parameters is **38964545 (148ḋ4 MB)**.

3.2.4 Preprocessing Algorithm

Algorithm 1 Preprocessing algorithm

Require: Pairs of images P_i with corresponding labels
Ensure: Preprocessed dataset P' for model training
1: $P' \leftarrow$ Null
2: **for all** I in P **do**
3: **Resize:** $I \leftarrow \text{Resize}(I, 100, 100)$
4: **Convert to GrayScale:** $I \leftarrow \text{GrayScale}(I)$
5: **Interpolation Method:** $I \leftarrow \text{Interpolation}(I)$
6: $P' \leftarrow P' \cup \{I\}$
7: **end for**

3.2.5 Real-Time Verification Flow Chart using Siamese Network

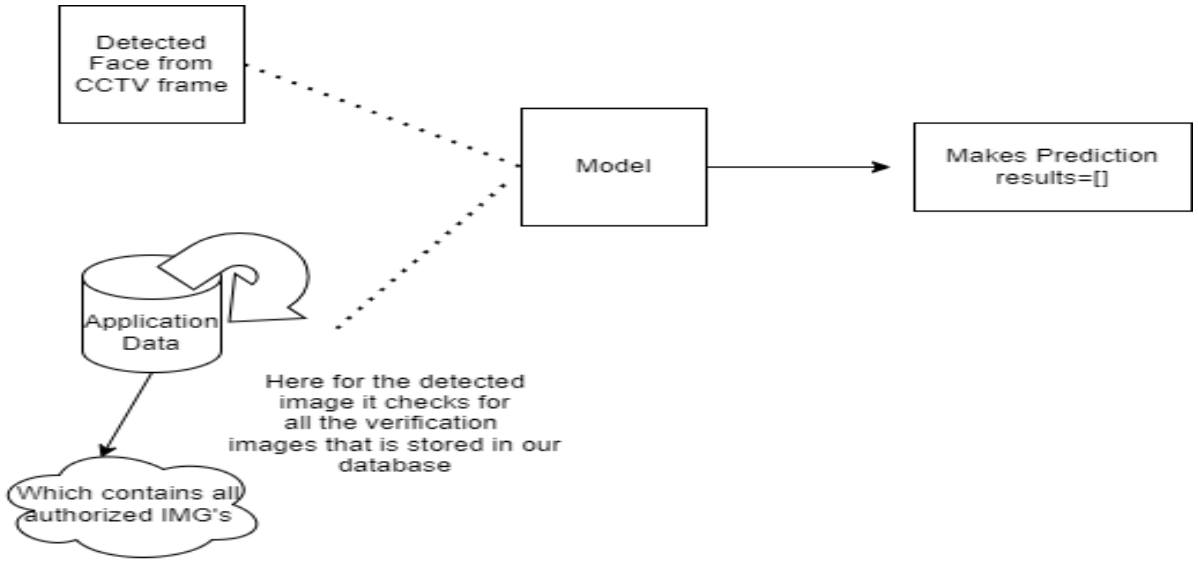

Figure 3.10: Workflow of Siamese Network for Real-time Verification

Chapter 4

Result Analysis

4.1 Results of Face Detection Stage

Fig 4.1. and Fig 4.2. shows the results of Detected Faces using YOLOv8 and HaarCascade algorithm. As the figure shows YOLOv8 is able to detect more faces then the HaarCascade Algorithm, in Addition YOLOv8 is the optimal choice for this system because it is time convenient. YOLOv8 is faster and gives better accuracy than the HaarCascade and other versions of YOLO algorithms.

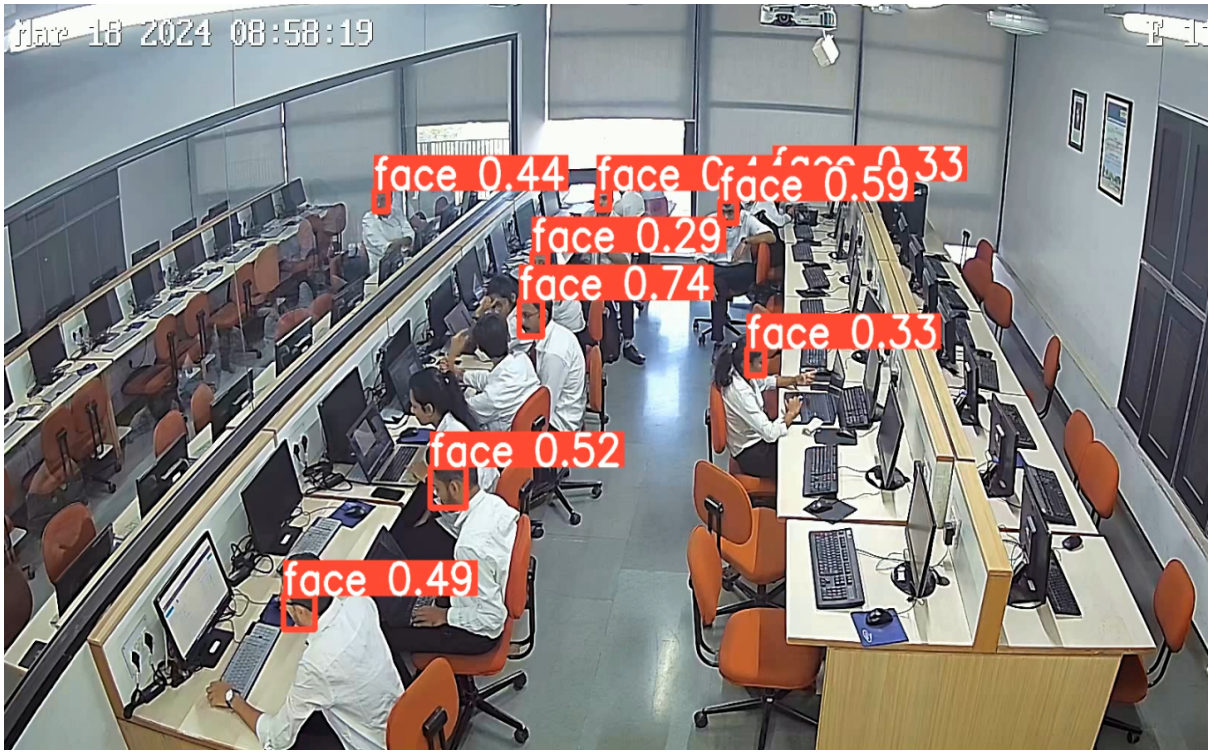

Figure 4.1: Detected Faces From CCTV Footage using YOLOv8

For Evaluating the YOLO algorithm CCTV footage of one class is taken from different two angles. In the CCTV frame, there are a total of 19 student faces on different frame of CCTV. YOLOv8 can detect more faces than the Haar cascade. In YOLOv8 the predicted bounding boxes are correct because it has better significant overlap means IOU(Intersection over Union) is greater than the ground truth bounding box.

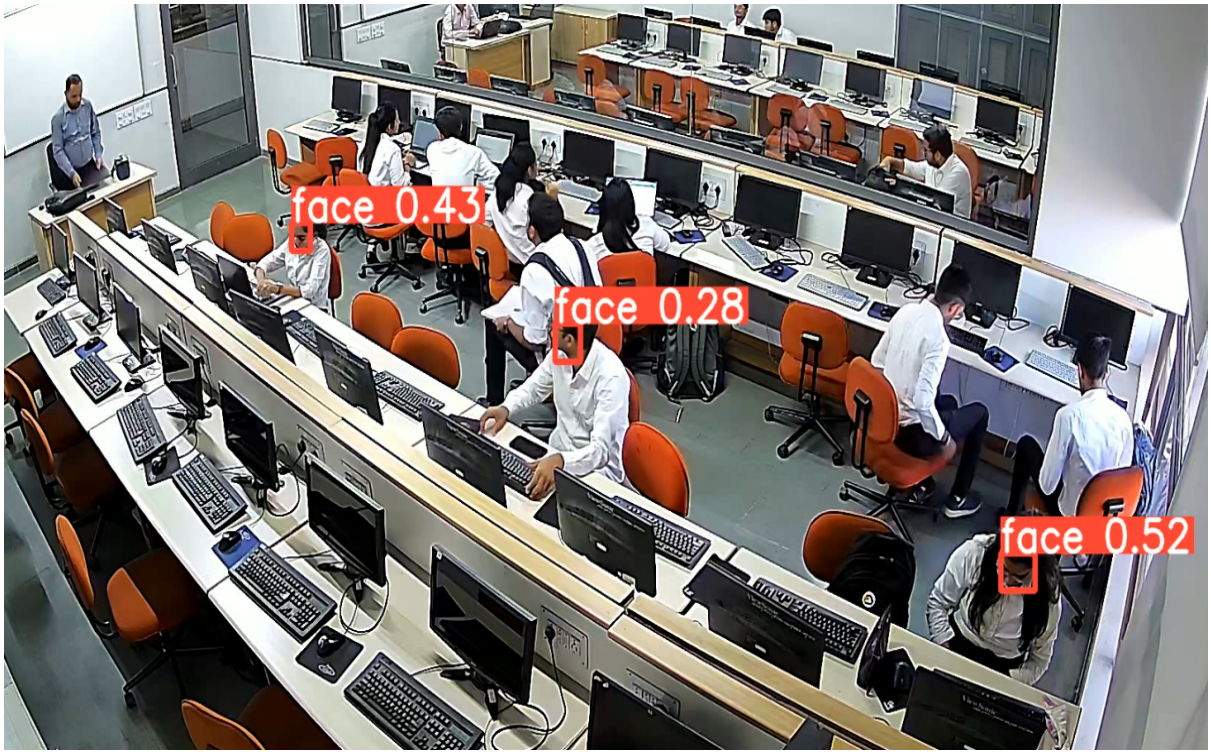

Figure 4.2: Detected Faces From CCTV Footage using HaarCascade

Formula to Calculate Accuracy

$$\text{Accuracy} = \frac{\text{correctly detected faces}}{\text{Total number of faces}} \times 100\%$$

4.2 Results of Face Verification Stage

As discussed in the Proposed methodology section for the Face verification task Siamese Network is used which is the combination of three components Embedding Layer for which we use the VGG16 pretrained model, Distance Layer, and Classifier part. For the evaluation purpose take a batch of Data in which there are a total of 16 pairs of images with a shape of 100x100x3. For the 100 EPOCHS, the classifier achieves the 1.0 Accuracy. it verifies each predicted face correctly. Below is the classification report of the Siamese

Network on One batch of data which contains 16 pairs of images. That is detected from the CCTV footage.

Classification Report:

	precision	recall	f1-score	support
0.0	1.00	1.00	1.00	10
1.0	1.00	1.00	1.00	6
accuracy			1.00	16
macro avg	1.00	1.00	1.00	16
weighted avg	1.00	1.00	1.00	16

Figure 4.3: Classification Report Of Siamese Network for a Single Batch

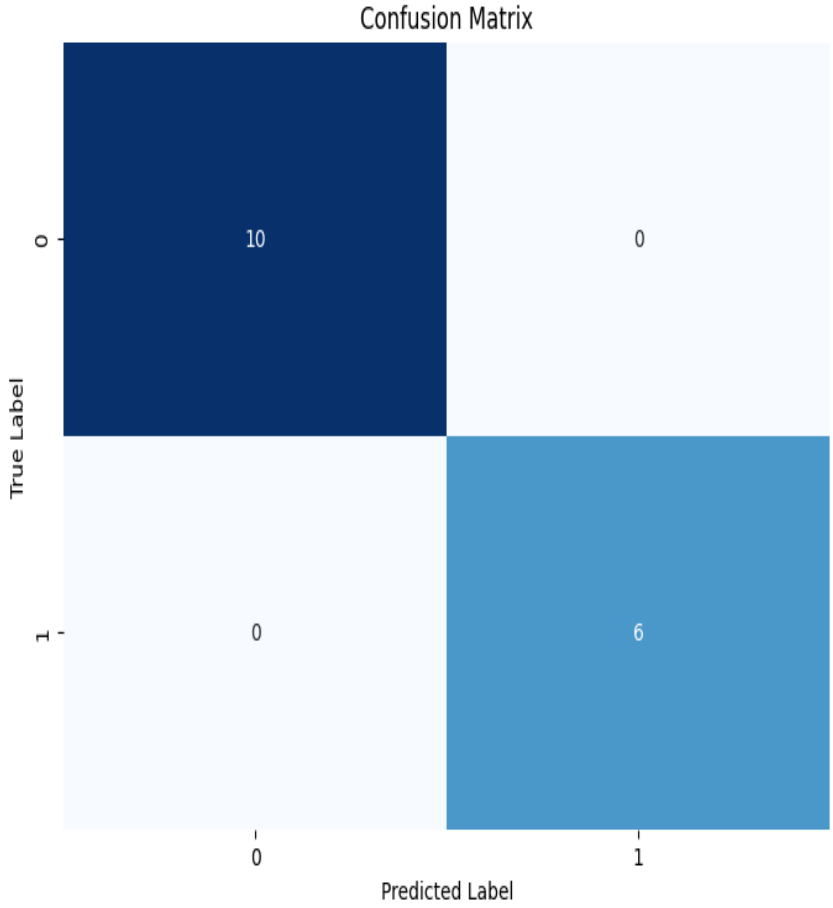

Figure 4.4: Confusion Metrix Of Siamese Network for a Single Batch

For a Single batch of Data, the Siamese Network gives an Accuracy of 1.0 and an F1-Score of 0.999. it concludes that the Siamese network works well for fewer datasets. It can accurately find the similarity between two images and gives accurate results i.e. whether both the images are similar or not, which helps in the recognition part of retrieving the information from the Student Database. For each successful Verification, it fetches that student's details from the Student Database.

4.3 Results of Combined Pipeline (Fetching the details from the Database)

As shown in Fig 1.1. and Fig 1.2. The dataset structure of The methodology is described. For each successful verification of the detected face, fetch the details of that detected face information from the Student Database. Fig 4.5. shows the stored information of Student_1. If a detected image and student_1 image (As shown in Fig 4.5.) are verified, the similarity of both images is good, then it fetches the student roll number from the student_1.info text file associated with the Student_1. Display the Roll number of that detected student on the bounding box on that CCTV frame. The evaluation is based on one branch which contains 19 unseen validation_images with information.

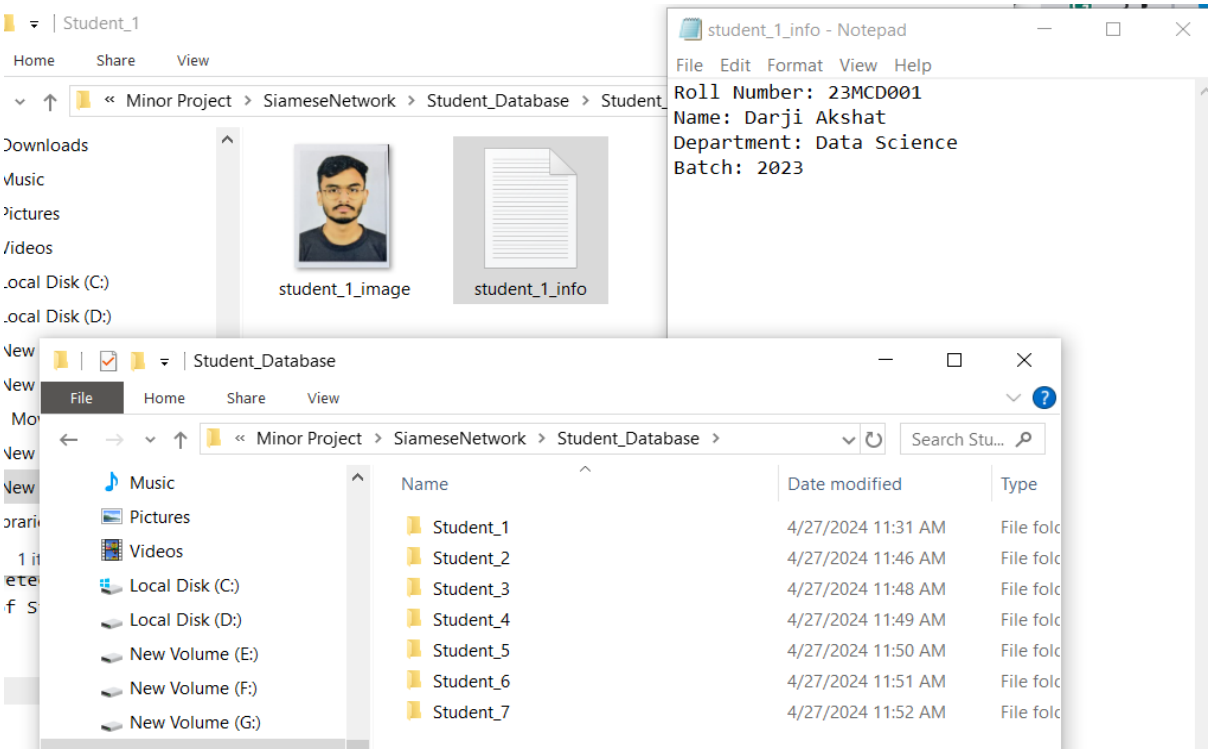

Figure 4.5: Stored Information of Student in Student_Database

Figure 4.6: Results of retrieval of information from Student_Database for threshold 0.5

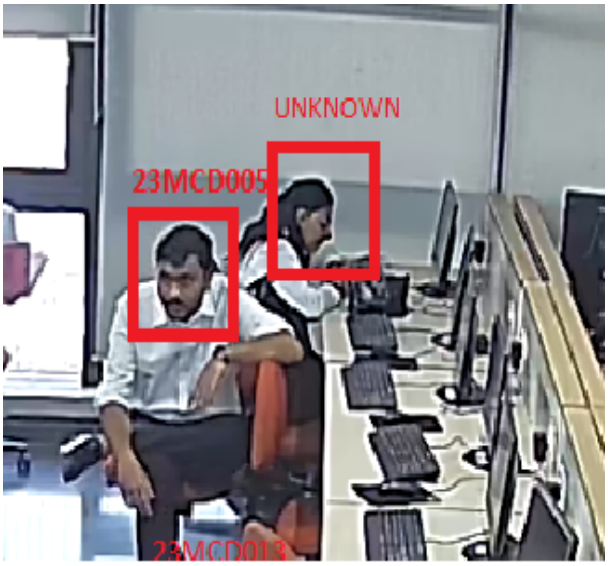

Figure 4.7: Results of Face Which are Not Detected

Fig 4.6. Shows the final results of retrieval of information on each successful verification by the Siamese network. Achieved 87% accuracy of retrieval of information with the threshold of 0.5 (**means if verification_result of the Siamese network is**

greater than threshold 0.5 then only we fetch the student_id from the Student_Database). Getting 81% accuracy for threshold=0.6, 77.89% accuracy for threshold=0.7. Increasing the threshold value reduces the accuracy of verification and fetching the information from the Student_Database. **To run this project, a higher-quality GPU and processor are necessary.**

Chapter 5

Conclusion and Future Plan

In this study, the main objective is to detect the student faces from the CCTV footage and recognize the detected face means fetch the student information i.e. roll number. There are two main phases of this project first is Face Detection and second is Face Verification which is followed by the face recognition task. For the Face Detection task, YOLOv8 gives better accuracy than the Haar cascade and other versions of the YOLO algorithm. For the Face Verification task Siamese network gives good accuracy and for 0.5 threshold total of 87% accuracy is achieved.

5.1 Future Work

During this study, lots of challenges occurred especially for the face detection task from the CCTV footage. And making the integrated system which is the combination of all these main steps of the proposed methodology. Enhancing the speed of this methodology is the real challenge so optimizing the proposed methodology, Optimization for Real-Time detection and fetching, Dataset Expansion using all branch unseen footage for the validation of the proposed methodology, A Desktop Application for this methodology, Apply Data Augmentation and preprocessing for the better results is the major future work for this study.

Bibliography

[1] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001*, vol. 1, pp. I–I, Ieee, 2001.

[2] G. Koch, R. Zemel, R. Salakhutdinov, *et al.*, "Siamese neural networks for one-shot image recognition," in *ICML deep learning workshop*, vol. 2, Lille, 2015.

[3] A. Stalin, A. Sha, A. S. Kumar, S. Nandakumar, and G. Gopakumar, "Face recognition at varying angles from distant cctv footage using siamese architecture," in *2022 3rd International Conference For Emerging Technology (INCET)*, pp. 1–6, IEEE, 2022.

[4] J. Celine *et al.*, "Face recognition in cctv systems," in *2019 International Conference on Smart Systems and Inventive Technology (ICSSIT)*, pp. 111–116, IEEE, 2019.

[5] J. Zhang, X. Jin, Y. Liu, A. K. Sangaiah, and J. Wang, "Small sample face recognition algorithm based on novel siamese network," *Journal of Information Processing Systems*, vol. 14, no. 6, pp. 1464–1479, 2018.

[6] P. Satyagama and D. H. Widyantoro, "Low-resolution face recognition system using siamese network," in *2020 7th International Conference on Advance Informatics: Concepts, Theory and Applications (ICAICTA)*, pp. 1–6, IEEE, 2020.

[7] S. Powale, A. Dhanawade, S. Bagwe, S. Kawale, N. L. Chutke, and S. Chavan, "Person identification in low resolution cctv footage using deep learning," in *2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN)*, pp. 236–240, IEEE, 2020.

[8] M. Heidari and K. Fouladi-Ghaleh, "Using siamese networks with transfer learning for face recognition on small-samples datasets," in *2020 international conference on machine vision and image processing (MVIP)*, pp. 1–4, IEEE, 2020.

[9] S. Tariyal, R. Chauhan, Y. Bijalwan, R. Rawat, and R. Gupta, "A comparitive study of mtcnn, viola-jones, ssd and yolo face detection algorithms," in *2024 International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE)*, pp. 1–7, IEEE, 2024.

[10] H. Aung, A. V. Bobkov, and N. L. Tun, "Face detection in real time live video using yolo algorithm based on vgg16 convolutional neural network," in *2021 International conference on industrial engineering, applications and manufacturing (ICIEAM)*, pp. 697–702, IEEE, 2021.

minor project

ORIGINALITY REPORT

6%	2%	4%	2%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Loughborough University Student Paper	1%
2	Li Yang, Ying Li, Jin Wang, Neal N. Xiong. "FSLM: An Intelligent Few-Shot Learning Model Based on Siamese Networks for IoT Technology", IEEE Internet of Things Journal, 2021 Publication	<1%
3	mail.easychair.org Internet Source	<1%
4	dokumen.pub Internet Source	<1%
5	Submitted to University of Greenwich Student Paper	<1%
6	dergipark.org.tr Internet Source	<1%
7	Ge Wen, Yi Mao, Deng Cai, Xiaofei He. "Split-Net: Improving Face Recognition In One Forwarding Operation", Neurocomputing, 2018	<1%