YOLOv8-Enhanced Facial Recognition for One-Shot Learning Attendance System

Krishna P Hasaraddi
KLE Technological University
Hubballi,Karnataka,India
krishnahasaraddi@gmail.com

Gauri Thambkar
KLE Technological University
Hubballi,Karanataka,India
gaurithambkar@gmail.com

Anupama P Bidargaddi
KLE Technological University
Hubballi,Karnataka,India
anupamapb@gmail.com

Aishwarya P Patil
KLE Technological University
Hubballi,Karnataka,India
aishu14032003@gmail.com

Gayatri Betageri
KLE Technological University
Hubballi,Karnataka,India
betagerigayatri@gmail.com

Meena S M
KLE Technological University
Hubballi,Karnataka,India
msm@kletech.ac.in

Uday Kulkarni
KLE Technological University
Hubballi,Karnataka,India
udaykulkarni.3@gmail.com

***Abstract*—Assessing student attendance is crucial in classrooms because it influences how well students perform. The integration of advanced image processing technology has streamlined the critical functions of our classroom attendance monitoring system. We proposed a method for student attendance control using face recognition with advanced one-shot learning to track student attendance efficiently. Our approach included employing YOLOv8, OpenCV, and face recognition frameworks for face detection and storage optimization by storing a single image per person. Our work achieved an accuracy of 97% for 50 student's datasets using a webcam. This innovative method surpasses traditional training-testing methodologies, enhancing effectiveness while effectively addressing authentication and logistical challenges within attendance management systems.**

***Index Terms*—face recognition, OneShot literacy, image processing, face discovery, Siamese Networks, YOLOv8**

I. Introduction

Attendance plays a pivotal role in academic and professional settings serving as a fundamental factor for both individual and collective success in educational institutions [1]. Regular attendance ensures that students not only receive essential information in classrooms but also actively participate in the learning process, fostering discipline, responsibility, and commitment to future success. Consistent attendance enables educators to evaluate student progress and adjust teaching methods for a more personalized learning experience in the professional sphere. Regular attendance reflects employee dedication contributing to team cohesion and overall productivity. Beyond mere punctuality attendance signifies a commitment to growth learning and contributing to shared goals. Our study outlines the link between regular attendance in educational institutions and improved academic performance, underscoring the need for effective student participation in classrooms. It points out the common use of attendance management systems in educational settings globally while raising concerns about their efficiency.

The integration of networking smart devices and advanced technologies such as machine learning has led to the digitalization of attendance systems with high-processing computing facilities. Manual attendance tracking is streamlined and smart devices facilitate seamless data collection. Machine learning supports sophisticated attendance systems capable of handling complex tasks.

The text introduces the concept of using face recognition technology for attendance systems as an emerging solution. These systems employ biometric methods to confirm identities by analyzing distinct facial features and converting these features into numerical data for database comparison. Digital imaging devices such as webcams and CCTV cameras are crucial in this process, capturing images to establish a unique facial pattern for each individual [8]. Facial recognition algorithms, which have advanced to adapt to various environmental and physical changes, are at the core of this system.

We propose a research study on an attendance system that integrates machine learning with facial recognition. It uses the One-Shot Learning method for effective face detection attendance recording, culminating in the compilation of attendance data in an Excel format for academic staff to assess.

Our proposed crowd-sourced framework for attendance recognition is utilizing YOLOv8 which aims to surpass other projects by presenting a robust classification and query-based retrieval model. This innovative approach represents a significant advancement in attendance management systems underscoring our commitment to leveraging cutting-edge technologies for enhanced performance and reliability.

The sections are arranged as follows: Section II delves into the background study for our proposed system. Section III elaborates on the implementation details of OneShot learning and YOLOv8 architecture. Section IV covers the results derived from the conducted experiments, with the conclusion presented in Section V.

II. Background Study

The usual method for tracking attendance involves instructors manually marking attendance during lectures. This can be time-consuming, particularly with a sizable student population, leading to difficulties in identifying those skipping class and potential inaccuracies in recording attendance for others [14].

To improve these in universities today scholars' attendance is monitored and evaluated through colorful systems despite the fact these technologies are beneficial as was already said their limitations and practicality cause problems in the process the systems include. One such system is where students' identity cards are equipped with RFID cards, which are swiped through RFID readers upon entering the classroom [5]. A notable drawback of this approach is the risk of unauthorized individuals using someone else's identity card to falsely record attendance [2]. A Bluetooth-grounded attendance system is also one of the ways where data for attendance is gathered by using the scholar's phone Bluetooth signals, considering that about 95% of council scholars always have their phones on them. This strategy works veritably well to ameliorate delicacy. The system can apply deputy removal ways. Its usability is oppressively limited, however, as a Bluetooth-based device can only establish connections with over 8 other biases. As a result, this system can only be used in classrooms with fewer than ten scholars [11].

After more study, it was discovered that this attendance control system is flawed from the ground up. 'Proxy attendances' can be avoided by using facial recognition and detection for attendance, meaning that only students who are present are graded appropriately. The necessary corridors are affordable because all classrooms have laptops and webcams.

In an attempt to provide a better approach in this area, we have implemented a system using YOLOv8 and the OneShot learning approach.

A. Face Detection

Initiating a recognition pipeline, face detection within an image stands as the initial phase, crucial for removing extraneous information. This process involves identifying and extracting one or more faces if the algorithm recognizes their presence in the original image, facilitating separate analysis [12]. Authors of [6] described a Real-Time Multiple Face Recognition system that uses Face Tracking and Convolutional Neural Network (CNN) for Face detection on an Embedded GPU system. The training of these algorithms involves utilizing diverse images, some containing faces and others without. Despite the apparent simplicity of the binary classification problem, the effectiveness of various face detection algorithms hinges on comprehensive training to yield optimal results [13]. The approach applied in this work is YOLOv8.

In the dynamic domain of object detection, the YOLO (You Only Look Once) model, particularly its latest version, YOLOv8, stands out for its real-time detection capabilities. YOLOv8 expands upon the fundamental principles set by its forerunners, introducing the concept of framing object detection in a streamlined manner. The model's structure entails partitioning, the image given as input is partitioned into a grid, where each cell is tasked with predicting individual bounding boxes and the probabilities associated with different classes. YOLOv8 introduces enhancements over earlier versions, leveraging both the Path Aggregation Network (PAN) and the Feature Pyramid Network (FPN) introducing altogether a fresh architecture for neural networks.

B. Face Recognition

Facial recognition involves capturing and extracting the facial features from an image, which are then input into a mathematical model. The objective of this model is to ascertain whether the detected face corresponds to any face previously recorded in a database [14]. This task yields two possible results:

• Verification: The model is involved in determining the validity of a specific user or assesses and denies the legitimacy of an individual.

• Identification: The model undertakes the task of categorizing (one-to-many), establishing the connection between the scrutinized face and all others stored in the database.

In the biometric attendance system, facial contours are employed for comparison, but this proves to be a time-consuming and tedious process [10]. The authors have chosen to implement Principal Component Analysis (PCA), which outperforms other algorithms by exhibiting a superior recognition rate and lower false positive rate. It is noteworthy that the developed system is designed to recognize faces within a 30-degree angle variation.

The main method is to match current face encodings with those in the database, marking attendance when a match is found. The authors of [7] employed an Enhanced Real-Time Face Recognition system based on the Original Binary Pattern Histogram (LBPH) for live face recognition across diverse image scripts, employing a similar methodology. A Computerized attendance recording system incorporating both face recognition and radio frequency identification (RFID) to identify and count scholars entering and exiting the classroom was proposed by authors of [5]. For face recognition in attendance systems, authors of [2] used a mongrel algorithm combining Eigenface star element analysis, Linear Discriminant Analysis (LDA), and Principal Component Analysis (PCA). Utilizing Eigenface, Haar Cascade Classifier, and PCA algorithms, authors of [4] created a face recognition attendance system that tested with a database of 70 scholars and achieved 97 chance accuracy. An attendance system grounded on the LBPH algorithm and the Haar Cascade Classifier was proposed by authors of [3]. It updates attendance based on group images taken during class hours.

In our work, we have integrate One-Shot Learning here to reduce the database size. The one-shot learning model operates efficiently with a minimal database, enabling the identification of similarities among objects. Siamese networks are a kind of neural network architecture widely applied here, that focuses on learning a similarity metric for pairs of inputs. The Siamese network involves multiple identical sub-networks that share weights and architecture. These sub-networks receive two distinct images as input and their outputs are compared

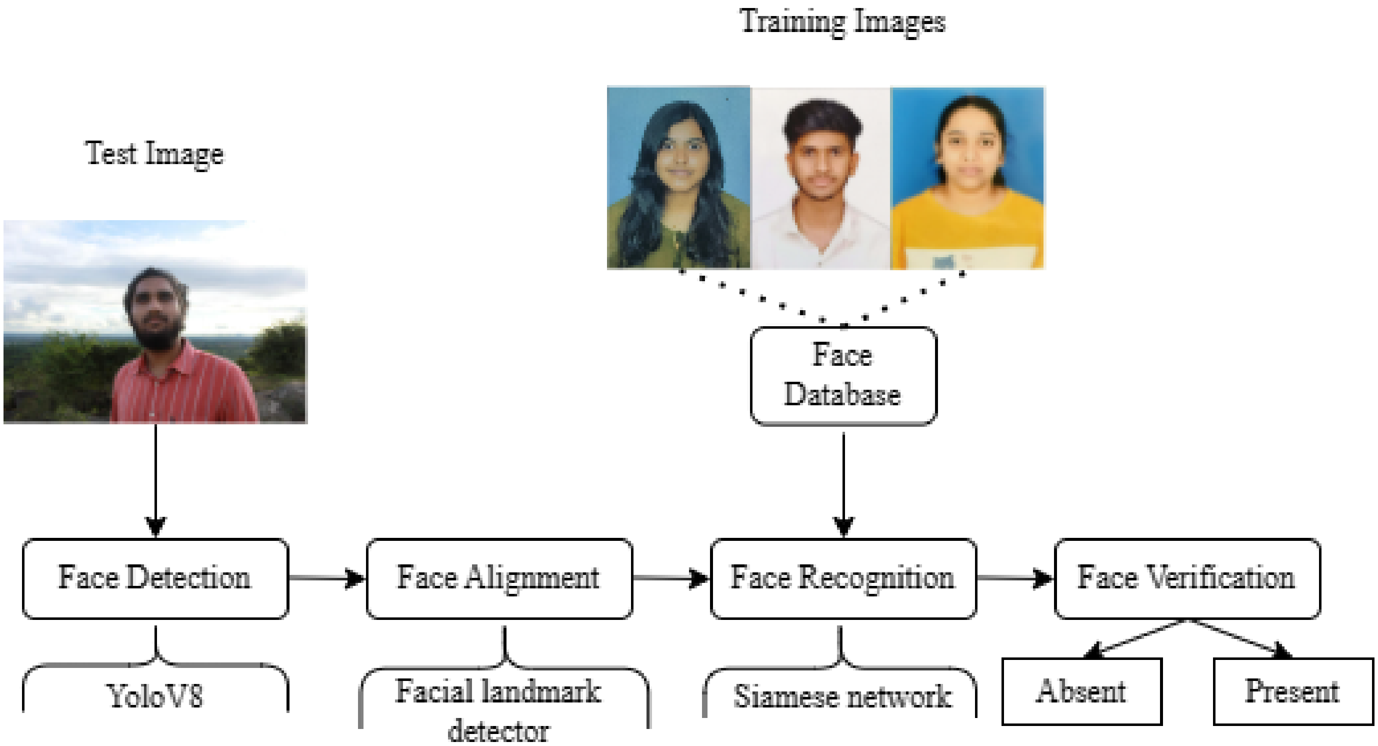

Fig. 1. Face Recognition Pipeline

employing a distance metric, for instance, Euclidean distance or cosine similarity. The distance metric is then used to classify whether the two images belong to the same class or different classes.

III. PROPOSED WORK

In our approach, we harness the capabilities of a webcam to acquire a clear image of each student's face. Subsequently, this image undergoes a multi-step pipeline, outlined in Figure 1, designed to achieve the following objectives:

Face Detection: The initial phase entails identifying the presence of a face within the captured image. This fundamental step ensures that the subsequent processes are applied exclusively to valid facial data.

Face Alignment: In order to enhance accuracy, we proceed to align the detected faces to a standardized position. This alignment helps mitigate variations caused by differences in head angles or positioning during image capture.

Student Recognition: Following successful alignment, our system employs facial recognition technology to match the aligned face with known student profiles. This step enables us to accurately identify and verify student attendance

A. Face Detection

We've made the YOLOv8 detector a core component of our methodology to enable quick and reliable face detection. This option allows us to decrease the possibility of false negatives and good performance even under difficult situations.YOLOv8 stands out for its real-time detection capabilities. This model offers a generalized approach for detecting flying objects [16]. YOLOv8 is being regarded as the new state-of-the-art [17], achieving a balance between inference speed and model accuracy (mAP). The YOLOv8 architecture was chosen for its high performance on the COCO dataset, assuming it would provide the highest probability of success [18].

It divides the picture into SxS gris and predicts bounding box B and the confidence score for the boxes for each grid. We define trust as Pr(Object) IOUtruth pred. If there is no item in that cell, the confidence rating should be 0. Otherwise, we want the confidence score to be equal to the intersection of the forecast box and the ground truth. Each bounding box is made up of five predictions: x, y, w, h, and confidence [19]. The (x, y) coordinates denote the center of the box about the grid cell limits.

$$P(\text{Class}_i|\text{Object}) \times P(\text{Object}) \times \text{IOU}^{\text{pred}}_{\text{truth}} = P(\text{Class}_i) \times \text{IOU}^{\text{pred}}_{\text{truth}} \tag{1}$$

Which gives us class-specific confidence scores for each box.

We have trained it without a face dataset, We then converted the model to perform detection. Ren et al. show that adding both convolutional and connected layers to pre-trained networks can improve performance [20]. During training we optimize the following, multi-part loss function

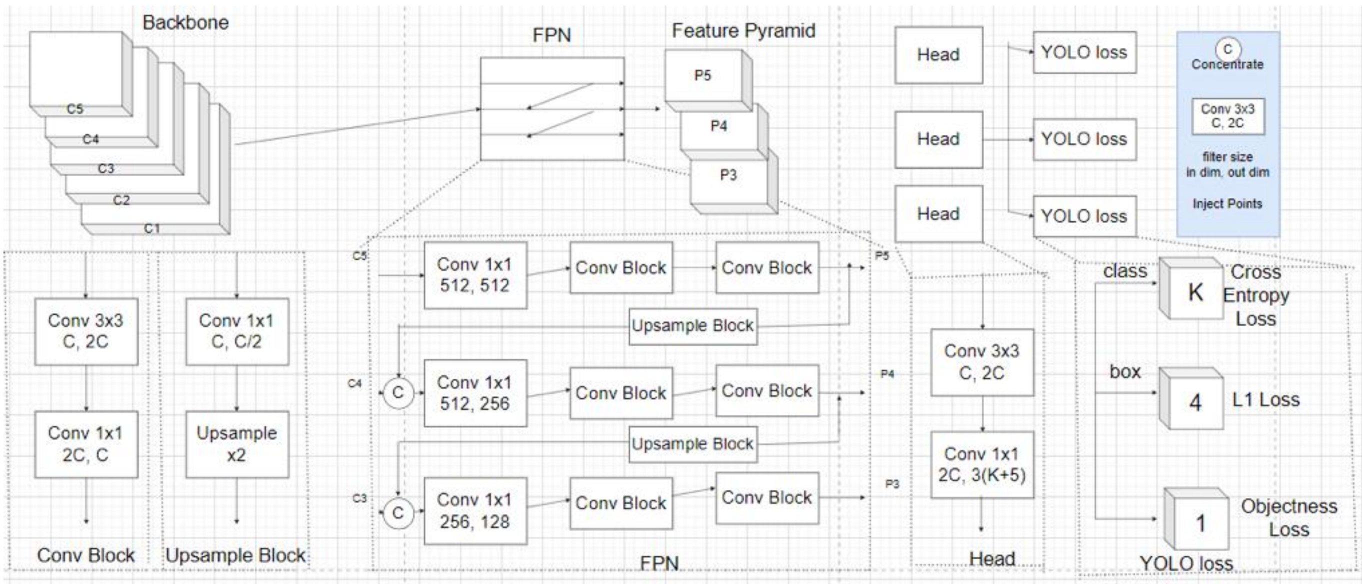

Fig. 2. YOLOV8 Architecture

$$
\begin{aligned}
&\lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right] \\
&\quad + \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \Big[(\sqrt{w_i} - \sqrt{\hat{w}_i})^2 \\
&\quad + (\sqrt{h_i} - \sqrt{\hat{h}_i})^2 \Big] \\
&\quad + \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} (C_i - \hat{C}_i)^2 \\
&\quad + \lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{noobj}} (C_i - \hat{C}_i)^2 \\
&\quad + \sum_{i=0}^{S^2} \mathbb{1}_{i}^{\text{obj}} \sum_{c \in \text{classes}} (p_i(c) - \hat{p}_i(c))^2
\end{aligned} \tag{2}
$$

The indicator $\mathbb{1}_{i}^{\text{obj}}$ signals the presence of an object in the i-th cell, while $\mathbb{1}_{i}^{\text{obj}}$ specifies that the j-th bounding box predictor within cell i is accountable for the prediction. It is important to note that the loss function is designed to target classification errors only in the scenario where an object is actually located within the grid cell, which aligns with the previously mentioned class probability aspect. Furthermore, the loss function will only consider errors in the bounding box coordinates if the specific predictor is deemed accountable for the correct ground truth box, meaning it has achieved the highest Intersection Over Union (IOU) compared to other predictors within the same grid cell.

The prediction mechanism is mathematically represented as follows

$$\text{Box} = (\sigma(x), \sigma(y), e^w, e^h) \tag{3}$$

$$\text{Confidence Score} = \Pr(\text{Object}) \times \text{IOU}_{\text{pred}}^{\text{truth}} \tag{4}$$

$$q_{xy} = \frac{\beta_{xy} \cap \hat{\beta}_{xy}}{\beta_{xy} \cup \hat{\beta}_{xy}} \tag{5}$$

$$\nu_{xy} = \frac{4}{\pi^2} \left(\arctan \frac{w_{xy}}{h_{xy}} - \arctan \frac{\hat{w}_{xy}}{\hat{h}_{xy}} \right)^2 \tag{6}$$

$$\alpha_{xy} = 1 - q_{xy} \tag{7}$$

$$\hat{y}_c = \sigma(\cdot), \quad \hat{q}_{xy} = \text{softmax}(\cdot) \tag{8}$$

These formulas represent the IoU and other components essential for accurate bounding box prediction in YOLOv8, significantly improving the model's detection capabilities [24] [25].

In Figure 2 YOLOv8 introduces enhancements over earlier versions, including a new neural network architecture that utilizes both Feature Pyramid Network (FPN) and Path Aggregation Network (PAN) [27].

B. Face Alignment

For our facial recognition initiative, we've adopted the 'face-alignment' library, engineered atop the PyTorch ecosystem. Following facial detection with YOLOv8, this library is deployed to precisely map out facial reference points eye locations, nasal apex, and mouth boundaries. It executes pivotal transformations rotational, scaling, and translational to orchestrate these points into alignment with a normative facial blueprint. It's a critical factor in enhancing the accuracy of identity verification and recognition

C. Face recognition

1) Face Feature Extraction: The foundation of our system is built upon the One-Shot Learning Model, a strategy that relies on using a single sample to handle subsequent samples. In face recognition systems that employ the One-Shot Learning Model, we acquire a rich low- dimensional feature representation. This feature representation greatly enhances the efficiency of comparisons for verification and identification tasks [21].

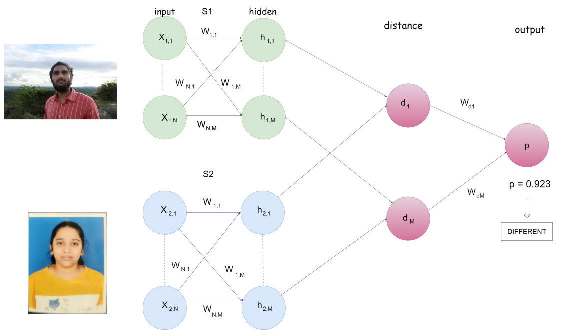

Fig. 3. Siamese Network for Face Recognition

Let's delve into the scenario of face identification within an attendance management system. Traditional Convolutional Neural Networks (CNNs) demand a substantial Dataset for effective training and often require iterative retraining even for minor changes in the Dataset. For example, the process of updating the Dataset due to student departures or new admissions can be resource-intensive and time-consuming. To overcome these challenges, we've adopted the One-Shot Learning Model, which significantly reduces the training time since it only requires a single image per student. Siamese networks are widely utilized for implementing the One-Shot Learning Model. Fig 1 illustrates the facial recognition process based on Siamese networks. These networks rely on a similarity function and consist of two parallel networks, each receiving distinct inputs. Their outputs are combined to make predictions. In the context of face recognition, Siamese networks learn a function denoted as f(d), which takes two input images: one from an external source (referred to as the candidate image) and the other from the Dataset (referred to as the actual image). The output reflects the similarity between these images. When the distance between the encodings of the two images is low, it indicates a match within the Dataset [21].

Siamese networks are widely utilized for implementing the One-Shot Learning Model. Fig 2 illustrates the facial recognition process based on Siamese networks. These networks rely on a similarity function and consist of two parallel networks, each receiving distinct inputs. Their outputs are combined to make predictions. In the context of face recognition, Siamese networks learn a function denoted as f(d), which takes two input images: one from an external source (referred to as the candidate image) and the other from the Dataset (referred to as the actual image). The output reflects the similarity between

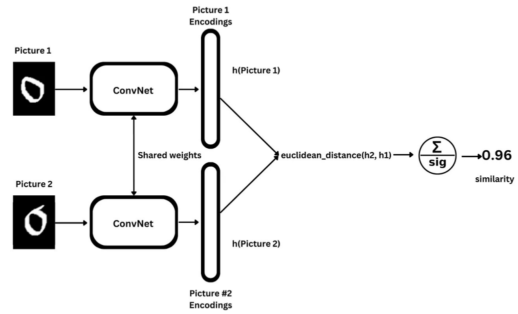

Fig. 4. Feature vector extraction in sister networks

these images. When the distance between the encodings of the two images is low, it indicates a match within the Dataset [22]. The essential objective is to train the model to minimize the

Euclidean distance of embeddings for identical photos while maximizing the distance for different photos by employing a triplet loss function. The framework of this approach is outlined in Equation.

$$\|y_i^a - y_i^b\|_2^2 + \beta \leq \|f(y_i^a) - f(y_i^c)\|_2^2 \quad \forall (y_i^a, y_i^b, y_i^c) \in \mathcal{D} \quad (9)$$

The loss to be minimized is then described in Equation 4

$$\text{Loss} = \sum_i \left[\|f(y_i^a) - f(y_i^b)\|_2^2 - \|f(y_i^a) - f(y_i^c)\|_2^2 + \beta \right] \quad (10)$$

- β: A margin or threshold that distinguishes the distance between positive and negative pairs.
- y_i^a: The anchor embedding, representing the reference facial image.
- y_i^b: The positive embedding, corresponds to an image with the same identity as the anchor.
- y_i^c: The negative embedding, represents an image with a different identity from the anchor.
- $\mathcal{D}$: The collection of all possible triplets within the dataset.

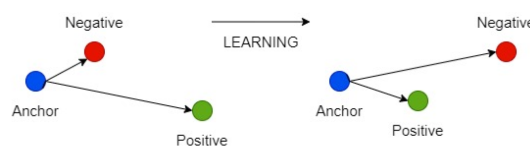

Fig. 5. Triplet Loss

2)Face Verification: A technique is utilized that relies on the measurement of Euclidean distances between facial embeddings to confirm identity. A predetermined threshold is critical in this evaluation—falling below it implies an identity correspondence. When confronted with instances of several students having distances beneath this threshold for a singular identity, the individual whose features are quantitatively nearest to the threshold, verging on zero, is deemed the accurate identity

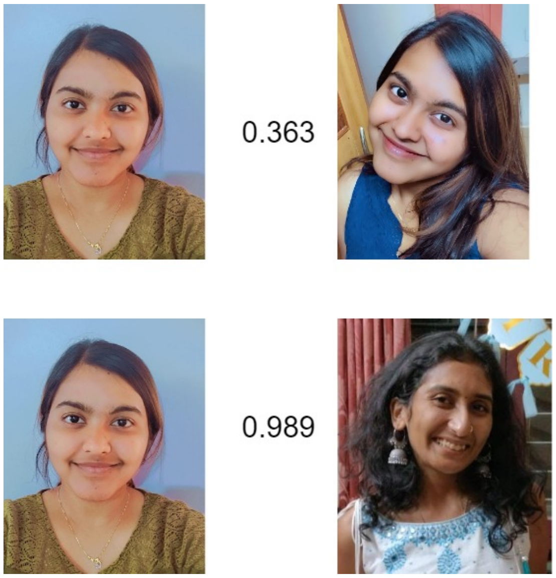

Fig. 6. Example of embedding distances

Algorithm 1: d Algorithm for Face Detection and Recognition

Result: List of recognized faces with identities

Procedure `MainProcedure`(*Image_Set, Model_Weights_YOLO, Model_Config_YOLO, Model_Weights_Siamese, Architecture_Config_Siamese, Database*) **:**
- Initialize YOLO and Siamese Network with given models and configurations
- **for** *each image I in Image_Set* **do**
  - Preprocess and detect faces using YOLO
  - Filter detections and crop faces
  - Generate embeddings and recognize faces from the database
  - Update the list of recognized faces
- **end**
- **return** *Recognized_Faces*

return

IV. RESULT AND ANALYSIS

A. *YOLO and Siamese Network performance*

YOLO Version	Accuracy (%)
YoloV8	97
YoloV7	94
YoloV6	80.3

Fig. 7. YOLO Comparision

Fig 7 provides a comparative overview of different versions of the YOLO (You Only Look Once) object detection algorithm, highlighting their accuracy percentages for our dataset. [26]. YoloV8, the latest iteration, achieves the highest accuracy at 97%, indicating a significant advancement in the algorithm's ability to correctly identify and classify objects within images. YoloV7 shows a slightly lower accuracy of 94%, while YoloV6, an earlier version, demonstrates 80.3% accuracy. The progression in accuracy from YoloV6 through YoloV8 suggests continuous improvements in the model's architecture, feature extraction, and overall object detection capabilities. This evolution underscores the technological strides in deep learning-based computer vision, with each version building upon the previous to achieve more precise and reliable detection.

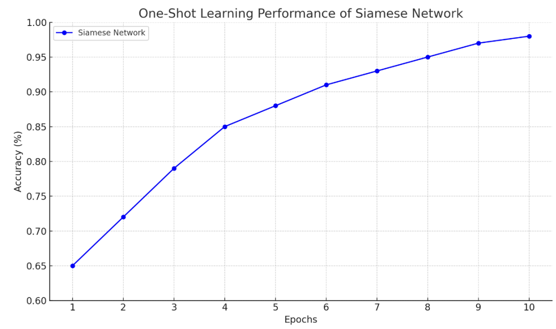

Fig. 8. Performance of a Siamese Network

In Fig 8 starting at an accuracy of 65 the network demonstrates consistent improvement with each epoch reaching an accuracy of 98 by the tenth epoch this trend illustrates the network's ability to effectively learn from limited data. It compares input pairs and learns from the differences and similarities is particularly suited for tasks

B. *Face Recognition Results*

In our work, the match indicator represents the minimum similarity required between the facial encodings from the live image of a student and those stored in our database. A demonstration of how this system operates in real-time image capture is shown in the figure. We have set a default confidence threshold for the match index at 0.6 [23]. When the match index falls below this threshold, the system effectively recognizes and confirms the identity of the individual. Subsequently, attendance is recorded in a CSV file. This file includes the names of scholars whose faces were successfully scanned and identified by the system. The generated CSV file is compatible with various applications that support the CSV format, such as Excel, Google Sheets, and Numbers, facilitating easy access and assessment. Our work had an accuracy of 97% for 50 students' datasets using a webcam.

Following the identification procedure the recognized faces are indicated as present on a CSV file that may be made and evaluated in Excel in soft copy format

V. CONCLUSION

In this period of rapid-fire technological progression and advancements in Biometrics, the facial recognition system

Fig. 9. Live webcam capture of the Students with Identification.

S . NO	NAME	Clock-in time
1	Krishna	19:25:27
2	Gauri	19:43:22
3	Aishwarya	19:57:34
4	Gayatri	20:12:45

Fig. 10. Attendance List

we've finagled is a significant advance. This system adeptly tackles challenges like cover attendance and the unreliability constantly set up in conventional attendance styles with their jacked security features. Biometric technologies are gaining traction in various sectors including seminaries businesses and universities despite the limitations of certain Biometric systems that render them less doable for real-world operations. Our facial recognition technology stands out as both reliable and economically doable. Our innovative system is compatible with ordinary PCs or laptops taking just a camera to serve its capability to operate independently of internet connectivity further making it an attractive budget-friendly choice for managing attendance.

This perfection-focused system, still, might affect in slower processing in surroundings with a high number of scholars. In addition, the recent relinquishment of mask-wearing in educational institutions as part of health safety measures presents a new challenge for our system, which was not originally set up for mask-covered facial recognition. Addressing this, we're laboriously developing advancements for our system that will enable it to effectively identify and corroborate scholars' faces, indeed when obscured by masks.

REFERENCES

[1] Fadelelmoula, T., & Almaarefa Colleges. (2018). The impact of class attendance on student performance. *International Research Journal of Medicine and Medical Sciences*, 47–49.

[2] Sawhney, S., Kacker, K., Jain, S., Singh, S. N., & Garg, R. (2019). Real-time smart attendance system using face recognition techniques. In *2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, pp. 522–525.

[3] D'Souza, J. W. S., Jothi, S., & Chandrasekar, A. (2019). Automated attendance marking and management system by facial recognition using histogram. In *2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS)*, pp. 66–69.

[4] Kiran, T. A., Reddy, N. D. K., Ninan, A. I., Krishnan, P., Aravindhar, D. J., & Geetha, A. (2020). PCA based Facial Recognition for Attendance System. In *2020 International Conference on Smart Electronics and Communication (ICOSEC)*, pp. 248–252.

[5] Akbar, M. S., Sarker, P., Mansoor, A. T., Al Ashray, A. M., & Uddin, J. (2018). Face recognition and RFID verified attendance system. In *2018 International Conference on Computing, Electronics & Communications Engineering (iCCECE)*, pp. 168–172.

[6] Saypadith, S., & Aramvith, S. (2018). Real-time multiple face recognition using deep learning on embedded GPU system. In *2018 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, pp. 1318–1324.

[7] Deeba, F., Memon, H., Ali, F., Ahmed, A., & Ghaffar, A. (2019). LBPH-based enhanced real-time face recognition. *International Journal of Advanced Computer Science and Applications*, *10*(5).

[8] Electronic ID. (2021). Face Recognition. Retrieved from https://www.electronicid.eu/en/blog/post/face-recognition/en

[9] Software Suggest. (2018). Manual Vs Automated Attendance System: Pros & Cons. Retrieved from https://www.softwaresuggest.com/blog/manual-vs-automated-attendance-system-pros-cons/ (Accessed: 2023-12-18).

[10] Jadhav, A., Jadhav, A., Ladhe, T. and Yeolekar, K., 2017. "Automated attendance system using face recognition." *International Research Journal of Engineering and Technology (IRJET), 4(1), pp.1467-1471*.

[11] Baharin, S. K., Zulkifli, Z., & Ahmad, S. B. (2020). Student absenteeism monitoring system using Bluetooth smart location-based technique. In *2020 International Conference on Computational Intelligence (ICCI)*, pp. 109–114.

[12] Muttu, Y. and Virani, H.G., 2015, December. Effective face detection, feature extraction & neural network based approaches for facial expression recognition.*In 2015 International conference on information processing (ICIP) (pp. 102-107). IEEE.*

[13] Zhang, K., Zhang, Z., Li, Z. and Qiao, Y., 2016. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE signal processing letters, 23(10), pp.1499-1503.*

[14] Crosswhite, N., Byrne, J., Stauffer, C., Parkhi, O., Cao, Q. and Zisserman, A., 2018. Template adaptation for face verification and identification. *Image and Vision Computing, 79, pp.35-48.*

[15] Menezes, A.G., Sá, J.M.D.C., Llapa, E. and Estombelo-Montesco, C.A., 2020, July. Automatic attendance management system based on deep one-shot learning. *In 2020 International Conference on Systems, Signals and Image Processing (IWSSIP) (pp. 137-142). IEEE.*

[16] Pierre Sermanet, David Eigen, Xiang Zhang, Michaël Mathieu, Rob Fergus, and Yann LeCun. *Overfeat: Integrated recognition, localization and detection using convolutional networks*. arXiv preprint arXiv:1312.6229, 2013.

[17] Jeff Donahue, Yangqing Jia, Oriol Vinyals, Judy Hoffman, Ning Zhang, Eric Tzeng, and Trevor Darrell. *Decaf: A deep convolutional activation feature for generic visual recognition*. In Proceedings of the International Conference on Machine Learning, pages 647–655, 2014. PMLR.

[18] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. *Going deeper with convolutions*. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 1–9, 2015.

[19] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. *You only look once: Unified, real-time object detection*. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 779–788, 2016.

[20] Shaoqing Ren, Kaiming He, Ross Girshick, Xiangyu Zhang, and Jian Sun. *Object detection networks on convolutional feature maps*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(7):1476–1481, 2016. IEEE.

[21] Jason Brownlee. *One-Shot Learning with Siamese Networks, Contrastive and Triplet Loss for Face Recognition*. Available online:https://machinelearningmastery.com/one-shot-learning-with-siamese-networks-contrastive-and-triplet-loss-for-face-recogni 2020. Accessed: 2023-12-25.

[22] Fritz AI. *One-Shot Learning (Part 2/2): Facial Recognition Using a Siamese Network*. Available online: https://fritz.ai/one-shot-learning/, 2023. Accessed: 2023-12-25.

[23] Gary B Huang, Marwan Mattar, Tamara Berg, and Eric Learned-Miller. *Labeled faces in the wild: A database for studying face recognition in*

unconstrained environments. Workshop on Faces in'Real-Life'Images: Detection, Alignment, and Recognition, 2008.

[24] P. Viola and M. J. Jones, "Robust real-time face detection," *International Journal of Computer Vision*, vol. 57, pp. 137–154, 2004, Springer.

[25] J. Yan, Z. Lei, L. Wen, and S. Z. Li, "The fastest deformable part model for object detection," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2497–2504, 2014.

[26] Mohamed Traore. *Face Detection Dataset*. Available at: https://universe.roboflow.com/mohamed-traore-2ekkp/face-detection-mik1i/dataset/7. Accessed: 2024-01-04.

[27] Dillon Reis, Jordan Kupec, Jacqueline Hong, and Ahmad Daoudi. Real-Time Flying Object Detection with YOLOv8. *arXiv preprint arXiv:2305.09972*, 2023.