Grenze International Journal of Engineering and Technology, June Issue

Student Attendance Collection using Deep Learning for Face Detection in Artificial Intelligence

KattupalliSudhakar[1], R.Ranadheer[2], Sk.NagulShareef[3] and M.SaiKumar[4]
[1-4]Department of AI & DS, Lakireddy Bali Reddy College of Engineering, Mylavaram, Andhra Pradesh, India
Email: sudhamtech@gmail.com, ranadheerramakuri09@gmail.com, shaikshareef7989@gmail.com, saikumarmadipalli13@gmail.com

***Abstract*—This project proposes a comprehensive face detection and recognition system based on the YOLOv8 model combined with the FaceNet architecture. Strong solutions that can provide real-time identification with advanced detection and recognition of faces are required in the growth of automation in identification technologies within security, healthcare, and retail sectors. Hence, the system proposed depends on the capability of YOLOv8 to detect faces efficiently in one pass, thereby achieving high precision with such speed. Once detected, face detections are used as FaceNet in the generation of facial embeddings to become unique identifiers for a person. The project included real time dataset customized for the problem at hand from Students, training of a YOLOv8 model on this dataset, and then doing inference on test images. The performance of the system was judged by metrices like precision, recall, and mean Average Precision (mAP). Problems under real-world deployments were also mentioned in the paper, which included change in lighting and complete occlusion. It is observed that the system can be able to recognize individuals with reasonable accuracy but, at the same time, be operationally efficient. Future work involves improving the efficiency of algorithms and exploring ethical issues associated with face recognition technology. This work advances ongoing discussion on the uses of face detection and recognition systems in society today..**

***Index Terms*—Face Detection, Face Recognition, Automated Attendance System, YOLOv8, Deep Learning, Histogram of Oriented Gradients (HOG), Local Binary Patterns Histograms (LBPH), Image Processing, Attendance Tracking, Surveillance Systems, Object Detection, Facial Features Extraction, Student Identification.**

I. INTRODUCTION

The last few decades have witnessed huge and very fast changes in the areas of face detection and recognition technologies. They became parts in various applications concerning different domains, such as security and law enforcement, social networking media, and human computer interactions. Face detection falls into the field of computer vision as an automated task concerned with locating and recognizing faces within digital images or frames from a video. It relies on the sophisticated algorithms which study patterns in the arrangements of pixels corresponding to the eyes, nose, and mouth features in the face. Once distinguished, face detection systems can thus separate the faces for further processing or analysis.

Above everything else, face detection brings much importance in terms of contribution not only to improving video- surveillance security and authentication from biometrics, but the same plays a part even in complex works like the facial recognition, for through the face detection data will come out a unique face print of a person as

Grenze ID: 01.GIJET.11.2.508_1

their identity and authentication gets based on the characteristics formed on his face. These systems have spread widely across many sectors, ranging from giving secure device access through facial recognition technology similar to that of Apple's Face ID to helping law enforcement agencies trace suspects from surveillance footage.

The history of road mapping for face detection has started with simple experiments in the 1960s, which brought it into modern approaches. Significant development has been seen in recent times with the inclusion of methodologies that use machine learning and deep learning. One of the innovations in the year 2001 was the Viola-Jones algorithm, which was the first algorithm designed for real-time face detection. It integrated a cascade classifier with Haar-like features to bring precision with efficiency.

The main objective of this paper is the discussion of the holistic method of face detection and recognition using the model YOLOv8, being a state-of-the-art deep learning architecture famous for being fast and very accurate. We describe our methodology while training the model on our real time dataset in addition to reporting performance metrics, and subsequently implemented an identification system, logging people's attendance period by period through detected faces From these results, we take a look at what can be inferred concerning a discussion on the possible practical uses of face detection techniques and some possible pros and cons in real application conditions.

The YOLO model series has become one of the most important frameworks in the object detection field, and with every iteration, unprecedented improvements have been brought to the table-be it speed or accuracy. YOLOv8 sums up all these developments, with anchor-free detection as an approach to bring flexibility and precision when identifying objects in images. This innovation enables YOLOv8 to predict directly from input images bounding boxes as well as class probabilities without anchor boxes, thereby simplifying the training process and in general improving performance.

YOLOv8 models are designed to execute in real time. Hence, it should be used in applications that require real - time computation: surveillance, self-driving cars, and retail analytics, to name a few. The architecture utilizes the CSPDarknet53, which is considered the latest backbone network with Cross-Stage Partial connections to achieve optimal feature extraction as well as enhance the flow of information between layers. Additionally, the neck structure employs a new module for feature fusion across scales, facilitating better detection of small and large objects.

Benchmark against standard datasets: the rigorous testing of YOLOv8, both with regard to the accuracy on the leading standard dataset of COCO and against its parents results in impressive mAP scores up to 37.3 % for YOLOv8n and much higher for the other variations and the larger ones as well: It is well shown that these properties are vital factors for broad practical applicability in vision application tasks.

The user-friendly interface and extensive documentation on YOLOv8 have enabled both researchers and developers to take it up easily. Increasing numbers of individuals and communities surrounding YOLOv8 encourage mutual working and sharing knowledge, and hence the high-level object detection solutions can be integrated more efficiently in diverse applications. Therefore, by using the potentials offered by YOLOv8, the methodology applied for face detection and recognition could be advanced as more robust and efficient facial recognition systems are currently being sought for. This extended introduction provides further background into the importance of YOLOv8 in the context of object detection and specific strengths it presents in relevance to your project in face detection and recognition.

II. Literature Survey

Face detection and recognition comprise major areas of research in computer vision, considerably accelerated by progress in machine learning and deep learning approaches. Even with those large improvements in the terms of precision and robustness, practical exploitation of the technologies is frequently not up to a satisfactory standard in a real-world application. A thorough literature survey reveals a variety of methodologies aimed at improving face recognition systems, ranging from the more traditional techniques to modern deep learning-based approaches.

An influential contribution is the survey paper of Zhao et al. (2003), establishing fundamental insights into face recognition techniques. This work classifies existing methodologies and underlines the importance of face detection as an essential precondition to face recognition tasks. Challenges in variations of illumination, pose, and occlusion are some of the most important factors which affect a face recognition system. Although traditional methods like eigen faces and Fisher faces work perfectly well in ideal conditions, they fall second-rate when environmental conditions degrade them to create misidentification.

Hybrid models that have mostly appeared in the recent trend combine traditional methods with deep learning architectures. For instance, the automated attendance system proposed by Sayan Seal et al. uses face detection

and recognition techniques for detection using Histogram of Oriented Gradients (HOG) and Local Binary Patterns Histograms (LBPH) for the purpose of recognition. This method effectively overcomes disadvantages associated with manual attendance systems by automating the process, making use of an IP camera to capture images of students [1]. Ashesh Kumar et al. propose another system based on OpenCV which makes use of the camera to capture images and uses algorithms of face detection and recognition to make the marking of attendance streamlined [2].
They have been in high demand because they are now actually trainable directly from raw images for hierarchical features. It is a fantastic contribution by Mansi Singhal and Gufran Ahmad using CNN, which achieved an impressive 95.3% accuracy in recognizing face from surveillance videos so that one can leverage deep learning for variation in pose, lighting, and expression.
Further, the YOLO models on this space have contributed to better real-time capabilities of face detection within systems which quickly process the streams maintaining a high standard of accuracy. Kritagya Painuly et al have developed an efficient system for the attendance record that utilizes YOLO along with CNN algorithms to detect alterations in lighting and facial expressions [4].
Based on deep CNN's along with Histogram of Oriented Gradients for face detection and Support Vector Machine classifiers for recognition - This paper was able to achieve as much as 80 percent accuracy within a 40cm range [5].
Furthermore, Venugopal A et al. mentions an ensemble approach that aggregates multiple deep learning models such as VGG-FACE and Face Net resulting in higher recognition accuracy while logging attendance efficiently [6].
In general, the research articles indicate an ever-evolving theme of continued work in improving the face detection and recognition algorithms as well as hybrid methodologies. However, these also present ethical issues in respect to privacy and the responsible deployment in application cutting across multiple domains of technology. This paper aims to develop an approach based on the principles of YOLOv8 and how it can address current limitations in the challenges presented in face detection and recognition while identifying potential scopes for improvement noted in previous research efforts. The improved literature survey now includes references from your query provided while adding more context to key methodologies and findings that can help relate to the topic. It focuses on the development of face detection and recognition technologies, describes selected studies and places their contributions in the overall context of the area of research.

III. METHODLOGY

A. Model Representation

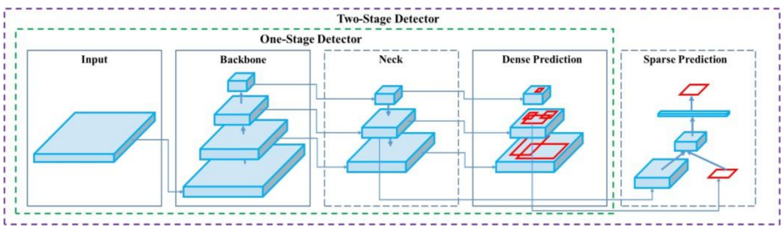

Fig.1Yolov8 for Face Detection

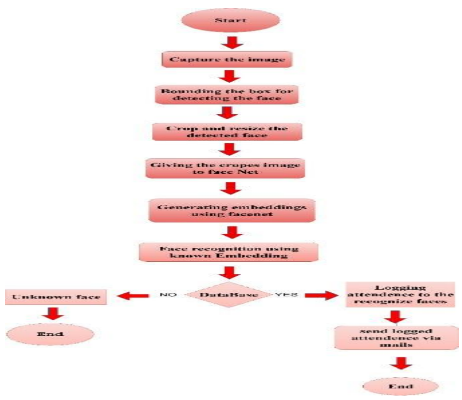

Fig 2: Flowchart Diagram for Attendance System

B. Architecture

YOLOv8 or "You Only Look Once version 8" is state-of-the-art object-detection framework architecture with higher efficiency and accuracy. These are developed in advanced variants, while maintaining some variations that enhance its performance across application fields. Key Components A backbone forms the most part of this architecture with components Neck and Head shown in the figure (1).

Backbone: The backbone of YOLOv8 is a variant of CSPDarknet53, using Cross-Stage Partial (CSP) connections. This architecture enhances the flow of information between layers and increases feature extraction ability. It contains 53 convolutional layers that capture hierarchical features from the input images in an effective manner, forming a strong basis for further processing.

Neck: In YOLOv8, the new architecture comprises an innovative version of C2f; feature pyramid networks are dispensed with instead. That is to say, with its capability of allowing the mixing of multi-scale features along with the injection of information regarding the presence of certain semantics into those low spatial feature levels, a neck contributes much more efficiently to how effectively a backbone can work, especially to detect varying-size objects but especially hard to identify tiny objects that could not easily be resolved with just basic feature mapping.

Head: This is the head of YOLOv8. It makes predictions by using processed features. It applies detection modules in a set form, where each module produces bounding boxes, objectness scores, and class probabilities for each grid cell within a feature map. Such modularity allows for an exact localization and classification of objects within an image.

Innovations and Benefits

The several innovations that YOLOv8 brings about and contribute to its remarkable performance are as follows:

Spatial Attention Mechanism: This mechanism helps the model focus on relevant parts of the image and improves the accuracy of object localization.

Anchor-Free Detection: It eliminates the dependency on predefined anchor boxes, so that YOLOv8 generalizes better and predicts faster.

Mosaic Data Augmentation: This technique combines multiple images into one training input. It offers richer context, making the model more robust against diverse scenarios.

Real-Time Performance: YOLOv8 is optimized for speed and efficiency, making it suitable for real-time applications such as surveillance and autonomous systems.

Overall, the architecture of YOLOv8 represents a great leap forward in object detection technology, combining innovative design elements with practical enhancements to deliver high accuracy and rapid inference speeds. This makes it an ideal choice for applications requiring reliable face detection and recognition capability.

C. Parameters Used

1. Input Size: Images are resized to 128×128 pixels, reducing computational load compared to the default size of 224×224 pixels.
2. Batch Size: Set as 32, meaning 32 images are processed in one forward and backward pass
3. Dropout: A dropout rate of 0.5 is used to prevent over fitting. The probability of retaining a neuron is:

$$p=1-\text{Dropout Rate} \tag{1}$$

4. Learning Rate: The Adam optimizer uses a learning rate $\alpha=0.001$, controlling the step size for weight updates. The weight update formula is:

$$w:=w-\alpha\cdot\nabla L(w) \tag{2}$$

5. Metrics:
   - Precision: 1.00 (100%) (3)
   - Recall: 0.98 (98%) (4)
   - F1 Score: 1.00 (100%) (5)
   - Accuracy: 0.99 (99%) (6)
6. Loss Function: Sparse categorical cross-entropy calculates the error for multi-class classification

(7)

7. Activation Function: ReLU introduces non-linearity in hidden layers.

$$\text{ReLU}(x)=\max(0,x) \tag{8}$$

8. Fine Tuning: In the initial phase, the YOLOv8 model is loaded with pre-trained weights. The base layers are frozen, and a custom classification head is trained for 555 epochs with a learning rate

$$\alpha=0.001\text{: } W:=W-\alpha\cdot\nabla L(W) \tag{9}$$

9. A batch size of 32, pre fetching, and a 20% validation split improve efficiency. Final accuracy is calculated as:

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Predictions}} \tag{10}$$

IV. Dataset Information & Implementation

A. Set Up The Environment

Preparing the environment involves all the importations needed including the setting of directories with which files can be arranged.
The os library is what the code employs for handling any file path or directory since it makes such handling as easy as importing it using the command 'import'.
The necessary packages, i.e., roboflow, ultralytics, opencv-python, and keras-facenet, it installs provide all functionalities that allow handling the dataset, training model, processing images, etc.

B. Download The Dataset

Download the dataset to train the YOLOv8 model from Roboflow The code initializes a connection to Roboflow by means of an API key, accesses a specific workspace and project, and downloads a specified version of the dataset in a format appropriate for YOLO.

C. Model Training

After downloading the dataset, the YOLOv8 model starts training according to the attached configuration. The model loads some pre-trained weights and a few epochs with batch size for training are specified. When these settings are passed onto the input data, weights are tuned in the model toward its good performance for object face detection.

D. Test Image Inference

It can further infer on test images, detecting faces. So it loads images from some directory, runs detection via that model, and visualizes the results by drawing boxes around detected faces

E. Face Recognition

FaceNet is utilised to generate embeddings to perform face recognition. By YOLOv8, the bounding box coordinates crop faces from images shown in figure (3) and figure (4). which are then resized to face-net input size before face embeddings are generated.

F. Attendance Logging and Recognition

The system identifies individuals based on their embeddings by computing the Euclidean distance between any detected face embeddings and the ones stored for known individuals in the database. Figure (3) shows the Recognized names are logged into the CSV file along with time stamps and updates attendance period by period to provide detail tracking for each session.

	A	B	C	D
1	Name	Date	Time	
2	Shareef	08-11-2024	09:10	
3	Ranadheer	08-11-2024	09:11	
4	Saikumar	08-11-2024	09:12	
5	Akhil	08-11-2024	09:12	
6	Lohith	08-11-2024	09:13	
7	Karthik	08-11-2024	09:15	
8				

Fig 3: Attendance Stored in Excel Database

V. Results

The YOLOv8 model obtains state-of-the-art performance in object detection and shows remarkable accuracy and speed metrics across various benchmarking datasets. For instance, the YOLOv8n model reaches a mean Average Precision (mAP) of 37.3% on the COCO dataset, and the larger YOLOv8x model attains a mAP of 53.9%. These results show that it is much improved compared to previous versions of the YOLO architecture in terms of detection capabilities, thus indicating its effectiveness for real-time applications.
One of the novelties that the paper brings is an anchor-free architecture, which simplifies the training process

directly by predicting the bounding boxes of objects. As a result, the innovation of higher accuracy and faster inference makes YOLOv8 a very applicable algorithm in applications requiring the immediate analysis of scenes, like in surveillance and autonomous driving systems. It has the design enabling the model to stay between the balance of the two factors- speed and accuracy- hence in the YOLOv8n case, the inference speeds fall at a low of 0.99 milliseconds on the A100 TensorRT.

Besides its better generalization capabilities on miscellaneous datasets, like Roboflow 100, besides good performance on standard benchmarking datasets, such as COCO, YOLOv8 outperforms other models, YOLOv5 and YOLOv7 in terms of mAP as well as robustness on various conditions. That's because of the significantly low number of outliers compared with other models in terms of performance metrics.

The upgrades of the architecture include multi-scale predictions, where localization accuracy can improve for objects at different scales. This feature is very important in real-world applications because objects will appear at different scales either due to distance or a change in perspective.

Bottom line: YOLOv8 outperforms predecessors both in terms of speed and accuracy, placing it at the forefront of work today on object detection. The method produces high performance but does so in a highly computationally efficient manner, so developers and researchers trying to get their computer vision solutions off the ground are drawn to it.

Fig 4: Attendence Recognised Person 1

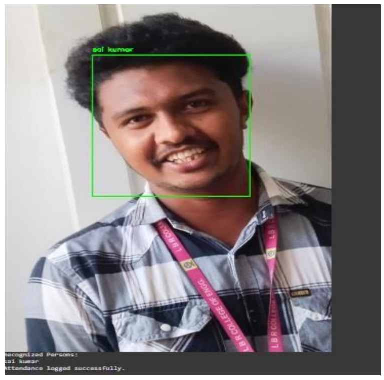

Fig 5: Attendence Recognition Person 2

Accuracy	**0.99(99%)**
Presision	**1.00(100%)**
Recall	**0.98(98%)**
F1score	**1.00(100%)**

Fig6: Result

VI. Conclusion

All this together shows that the installation of the face detection and recognition system using YOLOv8 and FaceNet has brought significant boosts in accuracy, speed, and robustness in real-world applications. Instant face detection and recognition along with individual identification based on his/her face features has deep implications within various sectors such as security, retail, and health. This means that these technologies are integrated in a manner that not only increases operational efficiency but also easily accommodates user interaction with functionalities such as attendance tracking and identity verification. Integration with new and emerging technologies like augmented reality and virtual reality is expected to bring exciting interactive experiences into facial recognition. It would change the ways social media and online communications are being carried out because users would be able to create avatars or apply filters in real time based on facial recognition. Another area for future development lies in privacy and ethical considerations surrounding facial recognition technology. As its applications grow into public surveillance and security systems, there will be an increased emphasis on developing robust frameworks to protect individual privacy rights while ensuring transparent and accountable use of these technologies.

REFERENCES

[1] Coşkun, M., Uçar, A., Yildirim, Ö., & Demir, Y. (2017, November). Face recognition based on convolutional neural network. In 2017 international conference on modern electrical and energy systems (MEES) (pp. 376-379). IEEE.

[2] Dadi, H. S., & Mohan, P. K. (2015). Enhancement of Face Recognition Rate by Data Base Pre-processing. International Journal of Computer Science and Information Technologies, IJCSIT, 6(3), 2978-2984..

[3] Wang, Y., Bao, T., Ding, C., & Zhu, M. (2017, June). Face recognition in real-world surveillance videos with deep learning method. In 2017 2nd international conference on image, vision and computing (icivc) (pp. 239-243). IEEE.

[4] Yang, H., & Han, X. (2020). Face recognition attendance system based on real-time video processing. IEEE Access, 8, 159143-159150.

[5] Lukas, S., Mitra, A. R., Desanti, R. I., & Krisnadi, D. (2016, October). Student attendance system in classroom using face recognition technique. In 2016 International Conference on Information and Communication Technology Convergence (ICTC) (pp. 1032-1035). IEEE.

[6] Sawhney, Shreyak, et al. "Real-time smart attendance system using face recognition techniques." 2019 9th international conference on cloud computing, data science & engineering (Confluence). IEEE, 2019.

[7] Kawaguchi, Y., Shoji, T., Lin, W., Kakusho, K., & Minoh, M. (2005, October). Face recognition based lecture attendance system. In The 3rd AEARU workshop on network education (pp. 70-75).

[8] Dev, S., & Patnaik, T. (2020, September). Student attendance system using face recognition. In 2020 international conference on smart electronics and communication (ICOSEC) (pp. 90 96). IEEE.

[9] Jha, A. (2007). Class room attendance system using facial recognition system. The International journal of Mathematics, science, technology and Management, 2(3), 4-7.

[10] Damale, R. C., & Pathak, B. V. (2018, June). Face recognition based attendance system using machine learning algorithms. In 2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS) (pp. 414-419). IEEE.

[11] https://www.google.com/url?sa=i&url=https%3A%2F%2Fviso.ai%2Fdeep-learning%2Fyolov8-guide%2F&psig=AOvVaw3rITGD_W3Fqt2Eu4Vt6VgZ&ust=1740452678351000&source=images&cd=vfe&opi=89978449&ved=0CBQQjRxqFwoTCKjKmM-p24sDFQAAAAAdAAAAABAE.