A YOLO-based Innovative Approach for Attendance Tracking Using Facial Recognition

Omkar Mestry[1], Mayoor Moolya[2], Sarita Mohapatra[3], Rohan Mohite[4], N.R.Khaire[5]
[1,2,3,4]Student, Dept. of Information Technology, Datta Meghe College of Engineering, Maharashtra
[5]Asst. Professor, Dept. of Information Technology, Datta Meghe College of Engineering, Maharashtra

Abstract - The conventional method of recording attendance in educational settings is manual, time-consuming, and error-prone. This research introduces a novel attendance management system (AMS) that employs facial recognition technology to automate the attendance procedure. The system utilizes the You Only Look Once (YOLO) algorithm for facial detection, is well recognized as it is highly efficient and accurate at the same time in real-time applications. The AMS comprises three primary components: Authentication and Authorization, Detection Portal, and Attendance Logging. It captures high-resolution images of students using advanced cameras, detects their faces using the YOLO model, and records their attendance in real-time. The attendance data is stored in a centralized database for easy access and management. The YOLOv8n model was trained using 700 augmented images to enhance detection accuracy in various conditions. Training was conducted on a computing platform consisting of an Intel i5-12400F CPU, 16GB RAM, and NVIDIA GTX 1650 GPU, with specific parameter settings for optimization. This proposed AMS simplifies attendance management by eliminating the need for manual data entry, thereby reducing administrative workload. By automating attendance tracking, educators can allocate more time to teaching tasks, ultimately improving productivity and operational efficiency in educational institutions.

Index Terms - Machine Learning (ML), YOLO, Recurrent Neural Network (RNN), Region-Based Convolutional Neural Network (RCNN), Attendance Management System(AMS)

I. Introduction

The Attendance System is important in the functioning of an institution. This attendance system mostly used in the institutes and schools are still the same since attendance was made mandatory in many places to check the regularity of a student or a particular person. This system is basically calling out the names of the person or a number assigned to a particular person and then marking present or absent based on the response. This method is time as well as energy consuming. The scope for error in this method was so much as people found out new ways to cheat the system (pretending someone else) and many more. This system also was modified to take the signatures before the name and roll number of the person and then to mark the attendance manually. As this system was also compromised by fake proxies (Marking in place of someone else) and many more. Furthermore, the respective in-charge had to manually store the attendance and then convert it into a digital copy. Students also raise concerns about the response such as loud and clear voice of "Present", absence of attention while the roll call etc. This important function had so many problems from both the in-charge as well as the attendant side. Considering these problems, a need to formulate and create a new attendance system felt necessary as we all go through the rough phase of the same old school technique with so many flaws. Face recognition is one of the answers to these problems. Numerous fields, including robotic navigation, security, medical, and surveillance systems, use face recognition technology. Face recognition has advanced significantly in the modern era because of deep learning. Facial recognition becomes even more crucial during pandemics for a variety of purposes, including social distancing, health care, attendance tracking, and people monitoring. There are several facial recognition algorithms on the market now. However, these algorithms' performance is lacking in several areas, which makes them perform poorly in real-time applications. R-CNN, Fast R-CNN, and other facial recognition algorithms are the most widely used ones.

One of the most basic applications of deep learning forfacial recognition is the automated labeling of publicly accessible Facebook images. It's crucial to remember that recognition occurs following the detection phase. One time detection and Two way detection are the two categories of detection techniques used in deep learning. Single shot detection, as the name implies, is a technique that allows for face identification in a particular one step. Face recognition in Two way detection involves two steps: extracting the features and classifying them. The first phase is the extraction of features. Real-time operations frequently involve face recognition. However, the algorithm used for recognition has a major role in the quality of facial recognition [1]

Joseph Redmon et al. introduced YOLO to the computer vision community in 2015 with a paper edition under the title "You only look once: unified, real-time object detection". The research framed object classification essentially by stating it as one-way reverting issue that starts at the pixels of the image and progresses with annotation and label higher occurrence chances. The suggested idea based on the above given paper allowed detection of various labels at same time with the addition of fast and correct results. The YOLO family has continued to develop rapidly since its establishment in 2016 until this year (2023). Although YOLO-v3 was cancelled by the original author (Joseph Redmon) for further work in computer vision, the performance, and capabilities of a "unified and number 039" core; several authors have developed this concept further and the most recent. The new version of the YOLO came with YOLOv8 [2]. The dataset that is used to train the Yolo-v8 model in this case will be a custom dataset where the data is in the form of images containing faces that are to be detected by the model.Yolo-v8 dataset has a specific format where a bounding box is added to the part of the image that has to be recognized and the model will then use that particular part of image in the training process. Training will be carried out with a number of epochs (iteration unit) that will be determined using trial and error method. This trial-and-error method will give us the best possible outcomes that will be selected for the actual project. This together will give us an attendance system that uses face recognition with a proper User Interface and to store the results at backend that will be powered by Node.js. The results stored then can be downloaded on the local machine in the excel sheet format without manually entering attendance of each student, thus saving extra time and effort of both students and the in-charge.

II. LITERATURE REVIEW

Attendance monitoring methods could be revolutionized by facial recognition systems, which have garnered significant attention in recent years. These systems use advanced algorithms to accurately recognize faces in different conditions, with the YOLOv8 pretrained model being a prominent one to be used for object detection. YOLO, a convolutional neural network-based approach, has demonstrated promising results in object recognition tasks, with Garg et al. (2018) proving its effectiveness in accurately recognizing faces in images. The YOLO algorithm's ability to process images live makes it subjectively suitable for various purpose which involve fast and efficient object detection (Face detection in our case) [10]. However, its performance can be further enhanced by increasing ROI strategies. Gunawan et al. (2023) suggested a long-range facial detection model with respect to the ROI-YOLOv8 that optimizes computing resources by focusing on specific regions of interest, resulting in high detection accuracy[10]. Advances in YOLO-based facial recognition systems have paved the way for robust and reliable attendance monitoring solutions in various environments.

Apart from YOLO-based approaches, previous studies have explored hybrid methods that combine different face detection and recognition techniques. Goel and Agarwal (2012) presented a hybrid approach that combines Haar cascade classifiers with facial geometric features to classify gender under different lighting conditions. This hybrid approach offers opportunities to combine different techniques to get a better and more dependable and strong face detecting model [5]. Furthermore, integrating facial recognition with other authentication mechanisms such as RFID enhances the security and reliability of attendance monitoring systems. Akbar et al. (2018) suggested a face detecting model that makes use of RFID-controlled attendance taking model that combines facial recognition technology with RFID authentication to ensure accurate and secure attendance tracking [7]. Integrated systems offer versatile authentication methods that ensure accurate attendance tracking while minimizing potential security risks. Overall, the literature demonstrates significant progress in face recognition systems, particularly using the YOLO algorithm for effective face recognition. Integrating ROI strategies and hybrid methods further enhances the accuracy and reliability of the face detecting models. Further invention could target on enhancing the respective algorithms and identifying additional integration opportunities to develop even more efficient search solutions.

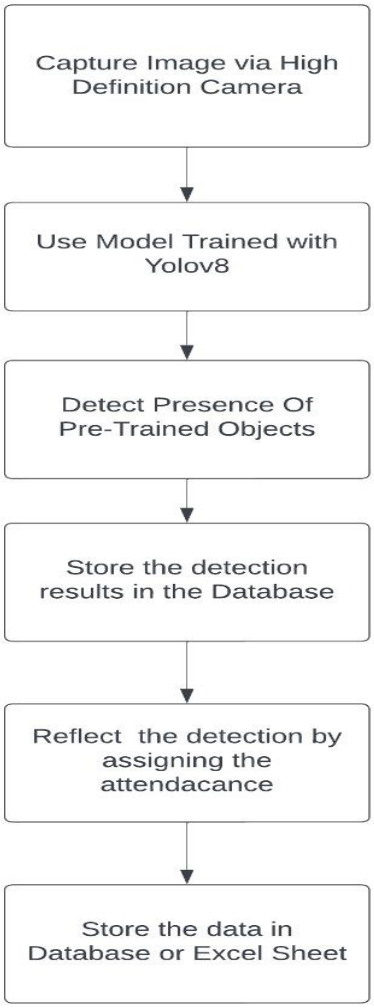


Fig. 1. Flow diagram illustrating the attendance tracking process.

III. Proposed Methodology

This study presents a new method to ease manual attendance system and achieve least human intervention. Our approach is carefully designed to handle the by breaking it down into manageable steps.

A. Proposed System Structure

The streamlined process begins with the capture of high-definition images of designated students using advanced cameras situated within the classroom. These images are then subjected to facial detection using a YOLO (You Only Look Once) model, meticulously trained to identify faces within images. Upon successful detection, the system promptly logs the attendance of individual students in real-time, leveraging the facial detection data as shown in Fig. 1. These attendance records are efficiently stored in a centralized database, ensuring easy access and revision as necessary. By automating the attendance tracking process through facial detection technology, the system eliminates the need for manual data entry or traditional paper-based systems, thereby significantly reducing the time required for attendance management. This streamlined approach not only enhances the accuracy of attendance records but also allows educators to allocate more time to instructional tasks, ultimately optimizing productivity and operational efficiency within educational environments.

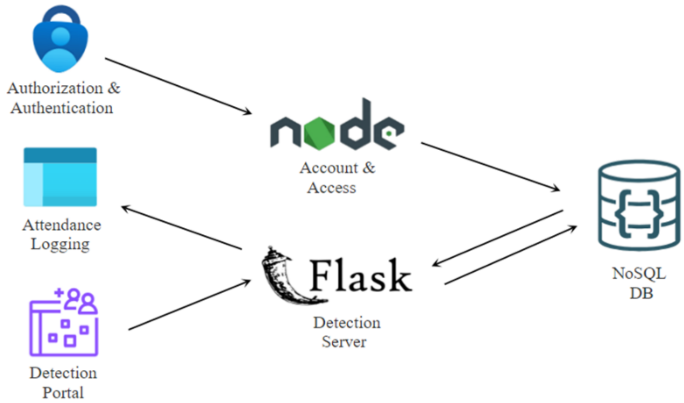


Fig. 2. Components of AMS

B. Components of AMS

The model underwent training, where YOLOv8n was employed to process both the images and their respective labels. This comprehensive approach ensured that the model was effectively trained to accurately detect and classify objects within the images. This model is further used in Attendance Management System (AMS). The AMS consists of 3 major features providing a seamless and consistent interface for management and analysis. According to the Fig. 2 , the system consists of 1) Authentication and Authorization 2) Detection Portal 3) Attendance Logging . The AMS provides an environment for teachers to collectively manage the attendance of each subject . This involves access to respective subject ,real time logging and storage and analysis of the attendance logs The authorization makes sure about the data is secured and access control is maintained .The system's subsequent component enables the identification of the targeted students by giving us access to an input image portal that further employs the model to identify their presence and correspondingly store data. The data is stored in a NoSQL database which is MongoDB. The authentication and authorization is handled by Nodejs Server and Detection Portal employes Flask. Flask Python works with the model giving out predictions and sending data to the common NoSQL database. This data is pulled out and processed to be displayed on the attendance log.

C. Training Parameter and Methods.

To initiate our project, we've commenced by constructing a robust facial detection model, harnessing the capabilities of YOLOv8n. The training images targeted two human(s).A total of 36 images which were further up sampled to 700 images using augmentation. The augmented data will represent a more comprehensive set of possible data points, thus minimizing the distance between the training and validation set, as well as any future testing sets.[1]. ROIs isolate regions within an image that contain relevant information, such as distinct facial features. This focus on ROIs mitigates challenges posed by varying lighting, pose variation, and reduced resolution[2]. The script utilized libraries such as OpenCV and NumPy for image manipulation and processing tasks. Augmentation techniques included rotations, flips, scaling, and adjustments in brightness and contrast. These simulated variations in the training data aimed to improve the model's robustness and generalization. By augmenting the dataset in this manner, the model became more adept at recognizing patterns across different orientations and lighting conditions

D. The training was done by computer:

(i) CPU: Intel i5-12400F, (ii) RAM: 16GB, (iii) GPU:NVIDIA GTX 1650. All training has the same parameter setting: (i) Batch: 16; (ii) Cache: True; (iii) Optimizer: SGD; (iv)Epoch: 300, (v) Epoch patience: 50. Furthermore, all trained models use the same pre-trained weight, the decay training rate of 0.0005. All the parameters of the two models which had a significant change in improvement are mentioned in the table below :

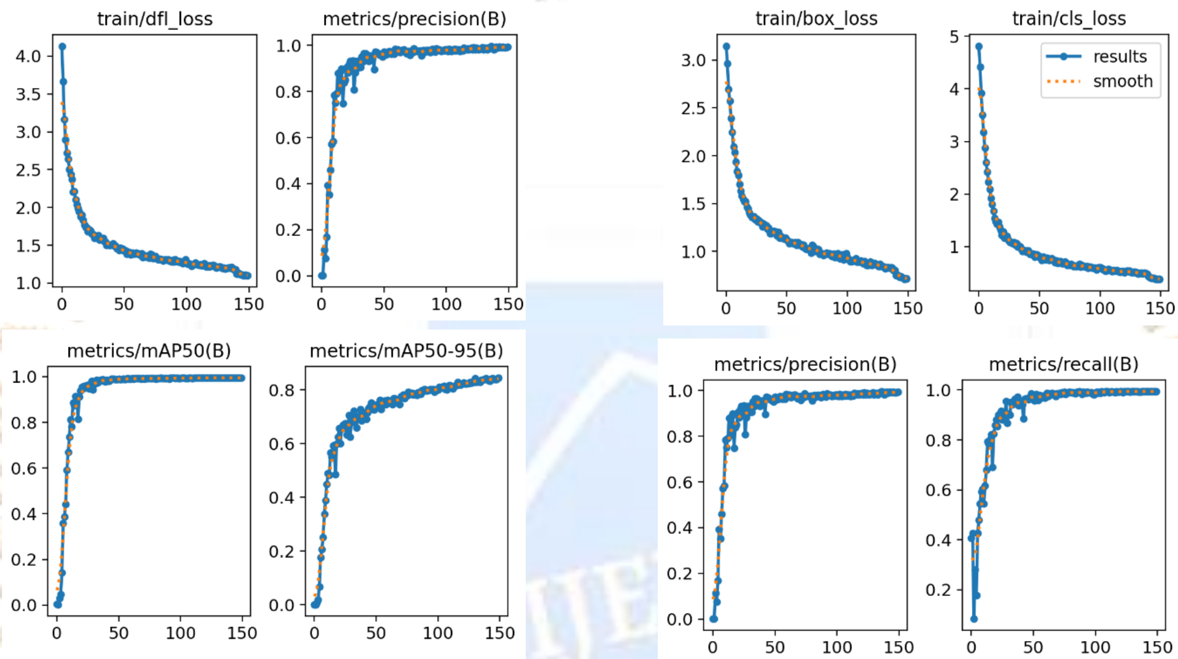
Model	Total Images	Images per class	Original Images per class	Epocs
1	300	150	18	200
2	700	350	18	300

E. Comparisons with previous models.

By using better parameters and an expanded dataset, the second model outperformed the first model by a wide margin. It showed improved performance with 700 photos total, 350 images per class, and trained over 300 epochs. Richer training data was ensured by this augmentation, which expanded the dataset from its initial 18 photos per class. Additionally, the 300 epoch training period enabled more thorough learning, which improved classification accuracy and resilience. Together, these enhancements produced a more sophisticated and trustworthy model that was better equipped to manage the intricacies of the categorization assignment. The second model's substantial advancement highlighted the transformative impact of refined parameters and expanded datasets on overall model efficacy.

IV. RESULT AND DISCUSSION

The images went through rigorous training to develop multiple models , out of which two models showed efficient detection results. Both the models were trained on the same set of images which consisted a total of 18 images per class with students at 5-6 meters. The training images had the dimensions of 640x640. The best model trained used a total of 700 images and 300 epochs. These 700 images were created using augmentation of the 18 images for each class . An epoch represents a full pass over the entire dataset. Adjusting this value can affect training duration and model performance. The below tables showcases all the training parameters of the model which the most accurate results :



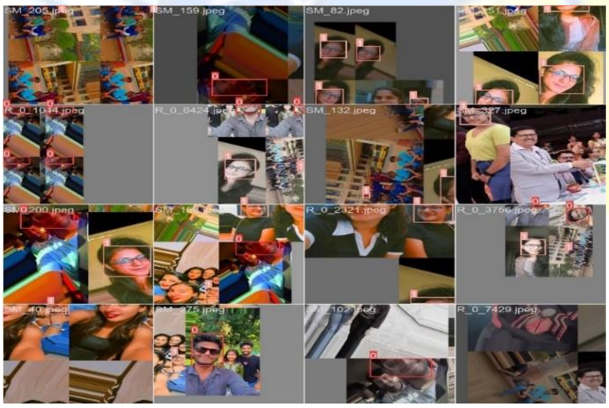


Fig.3 contains some of the training images that are used in the batch.

Fig. 4. Model 1

1) Model 1:

Model 1: A dataset of 300 photos, or 150 images on average each class, was used to train Model 1. As with Model 2, the original dataset was enhanced to improve the training data. In contrast, Model 1 was trained for a shorter amount of time—200 epochs were allotted for model convergence. Even with a reduced training period, the model was still able to go through a significant number of learning iterations. In order to attain sufficient performance within a limited time frame, this technique proposes striking a compromise between training time and computational resources. Furthermore, the fewer epochs may have reduced the likelihood of overfitting, enhancing the model's capacity for generalisation. Furthermore, Model 1's effective use of computing power demonstrates a pragmatic approach to model creation and deployment.

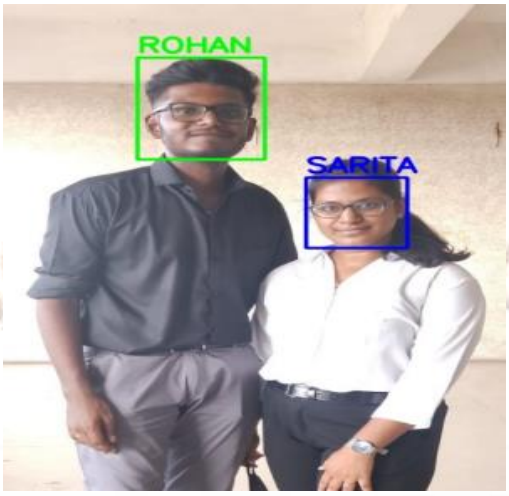


Fig. 5. Model 2

2) Model 2:

700 photos in all, or 350 images each class, were used to train Model 2. In order to increase the initial number of photos per class, the dataset was augmented. After 300 epochs of intensive training, the model underwent a great deal of learning iterations that allowed it to recognize complex patterns in the data. The act of augmentation probably added value to the dataset by offering a variety of viewpoints for the model to absorb, which could improve its capacity to extrapolate to previously unobserved data. Using a significant number of epochs indicates that the dataset was thoroughly explored, which helped the model converge on ideal parameters and improve its prediction power. The training plan of Model 2 seems strong, suggesting a thorough method for reaching high performance.

Comparisons with previous models:

By using better parameters and an expanded dataset, the second model outperformed the first model by a wide margin. It showed improved performance with 700 photos total, 350 images per class, and trained over 300 epochs. Richer training data was ensured by this augmentation, which expanded the dataset from its initial 18 photos per class. Additionally, the 300 epoch training period enabled for more thorough learning, which improved classification accuracy and resilience. Together, these enhancements produced a more sophisticated and trustworthy model that was better equipped to manage the intricacies of the categorization assignment. The second model's substantial advancement highlighted the transformative impact of refined parameters and expanded datasets on overall model efficacy.

V. CONCLUSION

Our suggested attendance management system is unique because it uses facial recognition technology, and this marks a shift from the way things were done previously. Our system utilizes high quality cameras and YOLOv8n facial detection model that makes sure that we have real-time and accurate attendance tracking. Clear images of students are taken into consideration, these images are subjected to proper analysis for the purpose of attendance logging in time. These records are put together in a secured centralized database by means of using authentication through Node.js and detection portal via Flask.

VI. REFERENCES

[1] S. M. M, A. George, A. N and J. James, "Custom Face Recognition Using YOLO.V3," 2021 3rd International Conference on Signal Processing and Communication (ICPSC), Coimbatore, India, 2021, pp. 454-458, doi: 10.1109/ICSPC51351.2021.9451684.

[2] Hussain, M. (2023, July 1). YOLO-v1 to YOLO-v8, the Rise of YOLO and Its Complementary Nature toward Digital Manufacturing and Industrial Defect Detection. Machines. Multidisciplinary Digital Publishing Institute (MDPI).

[3] D. Garg, P. Goel, S. Pandya, A. Ganatra and K. Kotecha, "A Deep Learning Approach for Face Detection using YOLO," 2018 IEEE Punecon, Pune, India, 2018, pp. 1-4, doi: 10.1109/PUNECON.2018.8745376. keywords: Face Detection;YOLO;Neural Network;object detection;Convolutional Neural Network,

[4] Goel, P., Agarwal, S. (2012). Hybrid Approach of Haar Cascade Classifiers and Geometrical Properties of Facial Features Applied to Illumination Invariant Gender Classification System. 2012 International Conference on Computing Sciences.

[5] Akbar, Md Sajid, et al. "Face Recognition and RFID Verified Attendance System." 2018 International Conference on Computing, Electronics Communications Engineering (iCCECE). IEEE, 2018.

[6] Shorten, C., Khoshgoftaar, T.M. A survey on Image Data Augmentation for Deep Learning. J Big Data 6, 60 (2019). https://doi.org/10.1186/s40537-019-0197-0

[7] Gunawan, Felix Hwang, Chih-Lyang Cheng, ZihEn. (2023). ROI-YOLOv8-Based Far-Distance FaceRecognition. 1-6. 10.1109/ARIS59192.2023.10268512.

[8] .Hapani, Smit, et al. ”Automated Attendance System Using Image Processing.” 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA). IEEE, 2018.

[9] Okokpujie, Kennedy O., et al. ”Design and implementation of a student attendance system using iris biometric recognition.” 2017 International Conference on Computational Science and Computational Intelligence (CSCI). IEEE, 2017

[10] Li, L., Mu, X., Li, S., and Peng, H. (2020). A Review of Face Recognition Technology. IEEE Access, 8, 139110–139120.

[11] R. Asyrofi and Y. A. Winata, ”The Improvement Impact Performance of Face Detection Using YOLO Algorithm,” 2020 7th International Conference on Electrical Engineering, Computer Sciences and Informatics (EECSI), Yogyakarta, Indonesia, 2020, pp. 177-180, doi: 10.23919/EECSI50503.2020.9251905.

[12] Viola, P., Jones, M.J. Robust Real-Time Face Detection. International Journal of Computer Vision 57, 137–154 (2004).

[13] K. Zhang, Z. Zhang, Z. Li and Y. Qiao, ”Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks,” in IEEE Signal Processing Letters, vol. 23, no. 10, pp. 1499-1503, Oct. 2016, doi: 10.1109/LSP.2016.2603342.

[14] W. Wang, X. Wang, W. Yang and J. Liu, ”Unsupervised Face Detection in the Dark,” in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 45, no. 1, pp. 1250-1266, 1 Jan. 2023, doi: 10.1109/TPAMI.2022.3152562.