Asistencia - A Deep Learning-Based Face Recognition System for Automated Attendance Monitoring

Ahmed Gomaa, Ahmed Nagah, Jana Elsalhy, Kareem Mazrou, Shiref Ashraf, Tasbih Neamatalla
Egypt-Japan University of Science and Technology, Alexandria, Egypt
{ahmed.gomaa, ahmed.nagah, jana.elsalhy, kareem.mazrou, shiref.ashraf, tasbih.neamatalla}@ejust.edu.eg

***Abstract*—Traditionally, taking attendance during university classes consumes noticeable time, and most of the time, it is inappropriate for the professor to do this task himself. In this paper, an alternative – Asistencia – is introduced as a solution for a real-time facial recognition system installed in existing CCTV infrastructure, saving the time and cost of installing new monitoring systems. We have utilized various models for this task: ResNet18, ResNet50, YOLOv8, and YOLOv11. Each model was trained on the SCface (a surveillance dataset). The proposed strategy works in two phases: detection and recognition. As the students enter the classroom, the model detects faces and recognizes them. However, to avoid any inaccuracies, an interface was developed for an administrator who monitors camera feeds with the capability of manually adding students to the attendance sheet and generating reports and statistics about specific classrooms to collect insights. Experimental results have shown the models' accuracy and speed, achieving 99.5% accuracy with 127 frames per second for the YOLOv11, which we have utilized in our system.**

***Index Terms*—Computer vision, Attendance System, Facial Recognition, Automated attendance, Student Monitoring**

I. Introduction

Facial recognition research is a trendy topic in the computer vision field. It can be used with surveillance and monitoring applications. For educational purposes, the automated attendance system is being explored. Automated systems are much more efficient than manual, error-prone, and time-consuming approaches. One of these automatic ways is face recognition. It offers a contactless, efficient, and accurate alternative, especially in environments like universities where scalability is essential [1]. Despite its potential, real-world deployment of facial recognition in classrooms remains limited due to technical and practical challenges.

The main challenge is that classroom surveillance camera setups often capture low-resolution images, uncontrolled lighting, and various poses and angles, which makes it difficult for the model to identify the subject correctly. [2] states that when face size decreases to 32×32 pixels, the accuracy declines significantly. On the SCface dataset, accuracy drops to 36% at a camera distance of 4.2 meters, highlighting the sensitivity of recognition systems to image quality.

Available recognition systems are typically trained on high-quality, controlled datasets like VGGFace2 or MS-Celeb-1M. These models fail to generalize when tested on low-quality surveillance images. [3] evaluated popular deep learning models, including FaceNet and SphereFace on QMUL-SurvFace, and the accuracy dropped to 32%. Even though image super-resolution has been tested, the improvements were negligible or had adverse effects.

To address these issues, **Asistencia** is our solution to this challenge: an attendance system that uses facial recognition to record student presence automatically. It leverages state-of-the-art facial recognition under real-world surveillance conditions. By integrating YOLOv11-based face detectors with existing CCTV infrastructure, **Asistencia** automatically identifies students as they enter the classroom and marks their attendance. The records are saved in the system, and instructors can manually edit or add students in case of any unexpected errors. It also lets instructors view or update students' attendance with a click.

This paper is organized as follows: Section II reviews related work on attendance management systems and surveillance-based facial recognition. Section III describes the dataset, preprocessing steps, and model architectures. It also presents the system design and implementation of Asistencia. Section IV discusses experimental results and performance evaluations. Finally, Section V concludes the paper and outlines directions for future work.

II. Related Work

Face recognition is a trending topic in computer vision in this era. One challenge facing the recognition models is the resolution of datasets. Since our project is dependent on implementing surveillance cameras, which often have low quality or the faces take a small portion of images, we must overcome this challenge to build a dependable system. Moreover, they pose differences and illuminations. We have tried our models on the SCface Dataset, which covers all these issues. There are previous efforts that have been attempted in [4]–[6].

Aghdam et al. [4] focus on finding a better model for low-quality image detection. They tried models trained on MS-Celeb-1M and fine-tuned on the VGGFace2 dataset. Several pre-trained CNNs have been used, which are SENet-50 and LResNet50E-IR. These models have been tested on SCface. Their Model has resulted in significant outcomes since they

achieved and Rank-1 reported Rank-1 recognition rates ranging from 78.5% to nearly 100% depending on the camera distance, demonstrating the effectiveness of resolution matching. While this approach is simple, it potentially sacrifices fine discriminative details due to down-sampling.

For better feature extraction in the low-resolution dataset, the multiscale parallel deep CNN (mpdCNN) was proposed by Mishra et al. [5]. The mpdCNN architecture consists of parallel pooling layers and fusion layers, so it captures facial features at multiple spatial scales simultaneously, which is suited for low-resolution images. It achieved an accuracy of 88.6% on SCface; the mpdCNN effectively balances feature representation and generalization. Nevertheless, this increased capability comes at the cost of higher computational requirements.

Tuvskog does one more implementation [6] evaluated the performance of pretrained FaceNet models on SCface and other datasets. A decrease in accuracy is noticed when models are trained on high-quality datasets rather than low-quality ones. The work highlights how crucial it is to incorporate low-quality photos into training to increase model accuracy for those applications. Furthermore, the evaluations' actual efficacy is limited because they rely on models without undergoing fine-tuning on the target surveillance datasets.

Both Aghdam et al. [4] and Mishra et al. [5] noted the limited utility of super-resolution techniques in this domain. It often introduces artifacts that confuse recognition models and may not improve accuracy. These observations suggest that approaches addressing resolution mismatch and multiscale feature extraction are more effective in the real world.

III. Methodology

A. System Overview

This system introduces an intelligent, vision-based attendance monitoring solution for university classrooms. By leveraging facial recognition powered by the YOLOv11 model and integrating it with existing CCTV camera infrastructure, the system automatically identifies students and records their attendance in real-time.

The core functionality is managed through an admin interface that enables login authentication, classroom selection, live camera feed monitoring, and real-time attendance marking. When students are recognized, they are automatically recorded in the attendance sheet. The admin can manually add unrecognized students and generate statistical reports on student attendance performance.

B. Datasets and Processing

1) Dataset Description: This study uses the SCface (Surveillance Cameras Face) database to evaluate face recognition algorithms in real-world surveillance applications [7]. The SCface dataset contains 4,160 images of 130 people captured using five commercially available surveillance cameras of varying quality and resolution. The images were collected in uncontrolled indoor environments with natural lighting and at three distinct distances (1.0 m, 2.6 m, and 4.2 m) to simulate typical surveillance conditions.

While SCface includes several types of images, such as high-quality frontal mug shots, infrared (IR) images, and various pose images, our study uses only the visible spectrum surveillance images captured by the surveillance cameras shown in Figure 1. These images are the ones that are relevant to our system, which is designed for face recognition under typical surveillance camera conditions. We excluded the mug shot images as they represent a controlled condition. Similarly, we ignored the IR images for preprocessing and training, as the infrared modality differs significantly from visible light imagery and is not the focus of our system.

Figure 1: Different Images from the Same Subject

2) Preprocessing: The dataset was preprocessed for consistency and to simulate real-world data issues. A significant problem in the SCface dataset is the gender imbalance, where the male subjects vastly outnumber females (114 males vs. 16 females).

To address this imbalance and avoid bias during model training, we applied oversampling techniques to increase the female samples, ensuring more balanced gender representation in the training data. Exploratory data analysis (EDA) visualizations, such as gender distribution histograms, confirm the initial imbalance and show the result of oversampling in the dataset in Figure 2.

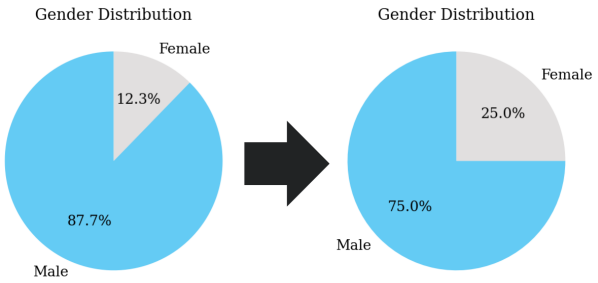

Figure 2: Gender Distribution before and after oversampling

Additionally, we analyzed the age distribution of subjects in SCface in Figure 4, which mainly covers young adults, a demographic consistent with university populations. This

makes SCface suited for applications such as university campuses. The dataset also includes data on facial hair presence, providing further demographic variety in Figure 3. All images were resized to a shape of 416×416 pixels. The dataset was divided into training, validation, and test sets with proportions of 70%, 15%, and 15%, respectively.

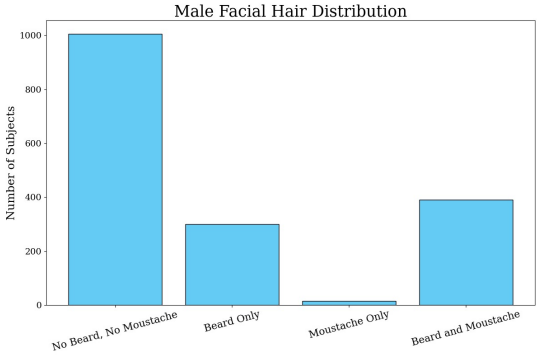

Figure 3: Males Facial Hair

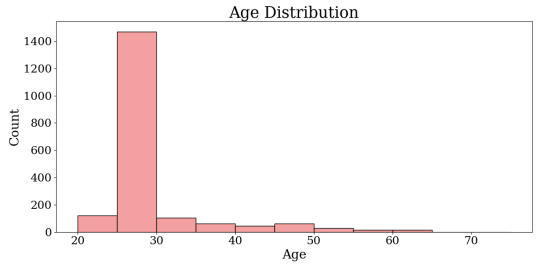

Figure 4: Age Distribution in SCface

The SCface dataset's advantages are that its images are captured under realistic indoor conditions, with surveillance cameras of varying quality, and at different distances. This setup simulates the variability in images exposed in practical scenarios. By focusing on visible spectrum surveillance The dataset directly reflects the challenging conditions that a surveillance system would face.

One major issue is the gender imbalance, with males outnumbering females, which can introduce bias and affect the fairness of the model. Though oversampling female images can solve this to an extent, the relatively small number of female subjects remains a constraint. Moreover, the dataset consists of Caucasian subjects only, limiting ethnic diversity.

C. Face Recognition Models

As we explored the models that provide both speed and accuracy, we came across multiple models that offer robust performance regarding face detection and recognition, starting from that basic CNN model and moving to more specific architectures such as YOLO.

First, we explored **ResNet50** [8], a deep network of 50 conventional layers that employ bottleneck blocks to learn more complex representations efficiently. ResNet50 was trained using the Adam optimizer with a learning rate 0.001 and weighted cross-entropy loss.

On the other hand, we explored two versions of the **YOLO family, 8 and 11**. Version **8** [9] benefits from data augmentation techniques like mosaic and mixup, which improve generalization by exposing the network to more varied training examples. However, version **11** [10] features enhanced convolutional blocks and a task-aligned anchor-free detection head which improves the results of the model regarding both accuracy and speed. YOLO is trained with SGD and momentum, using the CIoU loss to optimize classification and localization simultaneously. To sum up, each model has its strengths that we tend to explore more by training it on the SCface dataset and through a comprehensive comparison of all models' accuracy and frames per second.

D. Software Design

The Asistancia system performs tracking of students by integration of facial recognition technology into a web-based management system. Built on a scalable client-server architecture, it is build in three steps: a frontend (user interface), a backend (server logic), and a face recognition AI model. The frontend, made on web technology, has a user-friendly interface so the administrators and teachers can manage classes, register students, and monitor the real-time attendance also contains manually addition capabilities. The Node.js and MySQL-powered backend serves for the system, processing all business logic, database queries, and API endpoints to secure user authentication and data management. The core of the system is the Face Recognition Service, which uses Python, OpenCV, and YOLOv11 to process images automatically and recognize students and feed attendance directly to the backend. The next process is made by a solid MySQL database schema which supports the recordings of the attendance.

IV. Results

We calculated and compared the proposed models' accuracy based on accuracy as a first filtration. ResNet50 (98.83%) and YOLOv11 (99.5%) outperformed ResNet18 (95.43%) and YOLOv8 (92.75%). Then, we compared ResNet50 and YOLOv11 based on frames per second. ResNet50 detected 125 frames per second, while YOLOv11 outperformed it by detecting 127 frames per second. Figures 5 and 6 show the results of both YOLOv11 and ResNet50 during the training process.

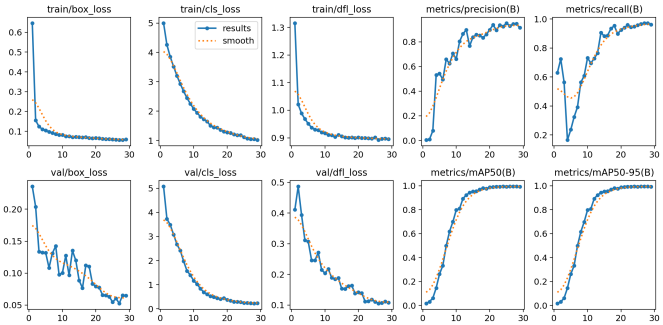

Figure 5: Results of YOLOV11

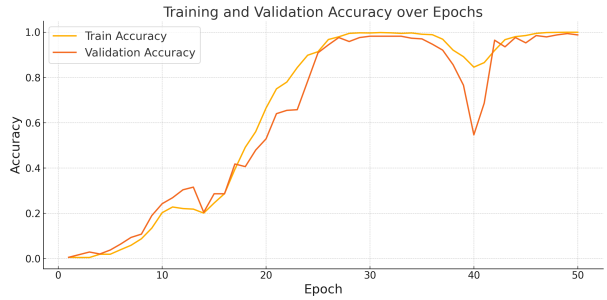

Figure 6: Results of ResNet50

Based on these results, we applied YOLOv11 to our system. However, we compared our models with previous related works': Aghdam et al. [4], Mishra et al. [5], and Tuvskog [6] who worked on the same dataset applying different models: SENet-50, LResNet50E-IR, mpdCNN, 512-VGG, 512-CASIA. Our model - YOLOV11 - outperformed all previous models, as shown in Table I and Figure 7.

Table I: Accuracy comparison across face recognition models.

Paper / Model	**Accuracy (%)**	**Notes**
Aghdam et al. [4]	SENet-50: 97.23 LResNet50E-IR: 98.15	Highest accuracy reported. IR data pretraining.
Mishra et al. [5]	mpdCNN: 88.6	Multiscale CNN architecture.
Tuvskog [6]	512-VGG: 99.49 512-CASIA: 97.61	Pretrained on VGGFace2. Pretrained on CASIA-WebFace.
YOLOV11	99.50	127 FPS, face alignment, lighting normalization.
ResNet50	98.83	125 FPS

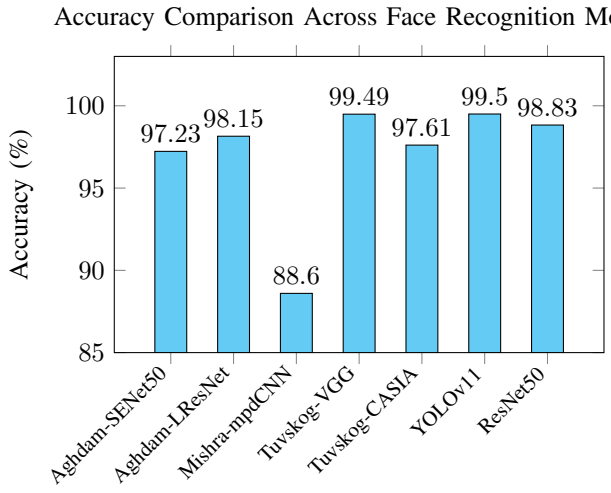

Figure 7: Accuracy comparison across face recognition models.

The model (YOLO-v11)achieved an accuracy of **99.5%**, outperforming several state-of-the-art models referenced, including SENet-50, LResNet50E-IR, and FaceNet-based architectures.

In addition to its high recognition accuracy, our model reached an average of **127 frames per second (FPS)**. This makes it highly suitable for live classroom environments where low latency is critical for practical deployment.

Preprocessing steps included face alignment and lighting normalization to improve robustness under varying classroom conditions. As shown in Table I, our approach outperformed comparable models in terms of both accuracy and runtime efficiency, validating its effectiveness for attendance automation in academic settings.

V. Conclusion

Asistencia performs better than state-of-the-art facial recognition algorithms coupled with existing CCTV infrastructure as it provides a contactless, automated solution. The system addresses real-world challenges commonly found in surveillance-based environments, such as low-resolution imagery, variable lighting, and diverse facial orientations.

Through the use of the SCface dataset and robust preprocessing techniques—including face alignment and gender balancing—Asistencia was trained to operate reliably under unconstrained classroom conditions. Our evaluation of multiple models showed that YOLOv11 achieved superior performance, with an accuracy of 99.5% and real-time operation at 127 frames per second, making it highly suitable for live deployment.

Furthermore, Asistencia accommodates market needs for scalable and equitable facial recognition systems in educational settings. It demonstrates that with perfect training data and architectural design, real-world constraints can be effectively implemented. In the long term, we intend to scale Asistencia to multi-camera systems, study cross-domain generalization, and enhance privacy defenses for deployment.

References

[1] N. Ali, A. Alhilali, H. Rjeib, B. Al-Sadawi, and H. Alsharqi, "Automated attendance management systems: systematic literature review," *International Journal of Technology Enhanced Learning*, vol. 14, p. 37, 01 2022.

[2] P. Li, L. Prieto, D. Mery, and P. Flynn, "Face recognition in low quality images: A survey," *arXiv preprint arXiv:1805.11519*, 2018.

[3] Z. Cheng, X. Zhu, and S. Gong, "Surveillance face recognition challenge," 04 2018.

[4] O. A. Aghdam, B. Bozorgtabar, H. K. Ekenel, and J.-P. Thiran, "Exploring factors for improving low resolution face recognition," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2019, pp. 2363–2370.

[5] N. K. Mishra, M. Dutta, and S. K. Singh, "Multiscale parallel deep cnn (mpdcnn) architecture for the real low-resolution face recognition for surveillance," *Image and Vision Computing*, vol. 115, p. 104290, 2021.

[6] J. Tuvskog, "Evaluation of face recognition accuracy in surveillance video," 2020.

[7] M. Grgic, K. Delac, and S. Grgic, "Scface - surveillance cameras face database," *Multimedia Tools Appl.*, vol. 51, pp. 863–879, 02 2011.

[8] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2016, pp. 770–778.

[9] Ultralytics, "Ultralytics yolov8," https://docs.ultralytics.com/models/yolov8/, 2023, accessed: 2025-05-27.

[10] R. Khanam and M. Hussain, "Yolov11: An overview of the key architectural enhancements," *arXiv preprint arXiv:2410.17725*, 2024. [Online]. Available: https://arxiv.org/abs/2410.17725