Final Report: Advanced Facial Attendance System

Can Ha An V202100364

Le Gia Duc V202100419

Nguyen Truong Tung V202100542

Abstract

Our Advanced Facial Attendance System is designed to streamline and enhance the process of tracking attendance by employing state-of-the-art Computer Vision and Machine Learning technologies. With different features aimed at evaluating multiple facets of a person's profile, such as emotion, liveness detection, age and gender profiling, this system is able to detect a person's presence in real-time with great efficiency and limit any chances of fraud.

This system is user-friendly, featuring a frontend developed with ReactJS and a robust backend powered by Python and FastAPI. Users simply upload images or videos through our web application, which are processed by the backend to perform real-time analyses for the different features pertaining to an individual's profile. The outcomes of these processes not only confirm attendance in multiple ways but also ensure the security and privacy of the data involved. The inference speed is also fast, ensuring that the application can be used in the real world setting

Key accomplishments include:

- Development of a custom-trained emotion detection model using a convolutional neural network (CNN), which achieves balanced performance across a range of emotions.
- Development of a Siamese Neural Network for facial recognition, which extracts features from images to create embeddings. [1]
- Integration of pretrained models for age and gender detection, enhancing user profiling.
- Deployment of the application that can be easily deployed to the cloud for scalability and accessibility

Our solution delivers a seamless, real-time experience that features robust anti-spoofing measures and high accuracy. By harnessing multiple aspects of facial analysis, our system offers a comprehensive method for automated attendance tracking. This approach significantly reduces administrative burdens and boosts security, ensuring a smooth and reliable operation.

Contents

1 Introduction

1.1 Motivation

Attendance tracking is essential across various settings such as educational institutions, workplaces, and conferences. Traditional methods, including manual roll calls and card-based systems, are not only inefficient but also prone to errors and vulnerable to fraud [2]. These systems often struggle to handle large-scale environments or adapt to complex scenarios, such as fluctuating lighting, obscured faces, or spoofing attacks using static images or videos [3].

In response, the Advanced Facial Attendance System overcomes these challenges by harnessing advanced machine learning and computer vision techniques. By incorporating facial recognition, emotion detection, and demographic profiling (age and gender), this system offers a secure, automated, and user-friendly alternative for tracking attendance. This innovative solution not only boosts accuracy and reliability but also ensures flexibility across various use cases [4].

1.2 Objectives

The main objective of the Advanced Facial Attendance System is to create a reliable, real-time facial recognition system that makes attendance tracking simple and efficient, overcoming the limitations of traditional methods. Its key objectives are:

- Accurate Facial Recognition: Develop a Siamese Network that is able to compare faces embedding with high accuracy (>95%) under varying conditions, including low lighting, different angles, and occlusions.
- Emotion Detection: Develop and train a custom deep convolutional neural network (Deep CNN) for classifying emotions, enhancing the system's ability to verify attendance by recognizing specific user expressions and producing a model that can be trained on many different datasets.
- Age and Gender Detection: Incorporate pre-trained models to provide demographic profiling as an additional layer of validation and analysis.
- Liveness Detection: Prevent spoofing attempts by implementing liveness detection techniques
- Real-Time Processing: Ensure the system can process images or video inputs within a response time of less than two seconds, providing a seamless user experience.
- Scalability and Accessibility: Design a cloud-based infrastructure using scalable frameworks like AWS to support multi-platform access and the recognition of multiple faces concurrently.
- Privacy and Security: Implement robust encryption and comply with data privacy regulations to protect sensitive user data.

1.3 Scope

Our system is designed to be a comprehensive solution with the following features:

- Multi-Value Detection: Simultaneously perform facial recognition, emotion classification, age, and gender detection, and liveness verification.
- Cloud Deployment: Host the system using scalable cloud services from AWS to support diverse environments, including classrooms, offices, and events.
- Cross-Platform Integration: Provide a user-friendly ReactJS-based frontend for web browsers, with potential for mobile application support in future iterations.
- Real-World Applications: Enable attendance tracking in educational institutions, employee monitoring in workplaces, and participant verification at conferences or events.
- Integration with Databases: Store attendance records and user profiles securely in a ACID compliant SQL Relational Database Management System

1.4 Addressing Challenges

The system aims to address several challenges associated with facial attendance systems:

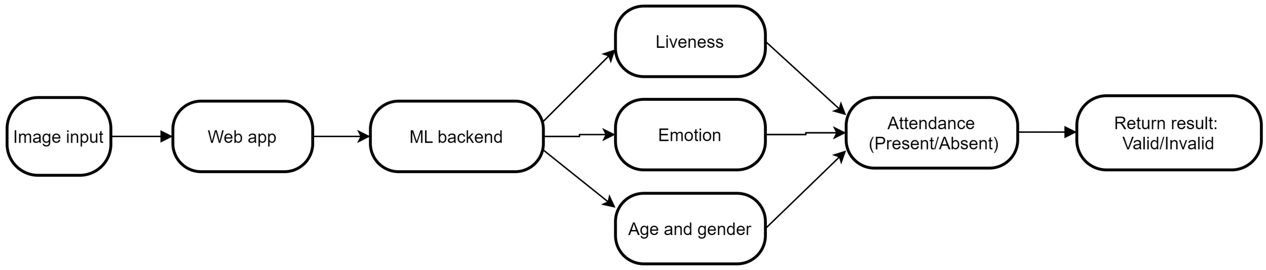
- Data Imbalance: Ensuring balanced performance across diverse facial expressions and demographics.
- Environmental Variability: Handling issues like varying lighting conditions, occlusions, and extreme head poses.
- Real-Time Performance: Achieving low latency without compromising accuracy.
- Integration of Models: Seamlessly combining facial recognition, emotion detection, and demographic profiling into a single system.

1.5 Impact

Our project contributes to the field of computer vision and machine learning by:

- Developing a practical multi-value facial attendance system that integrates multiple tasks into a single pipeline.
- Training a custom emotion detection model on a personalized dataset, ensuring that this feature is adaptable (can be trained and validated on many different datasets).
- Demonstrating the use of advanced neural network architectures, such as CNNs and Siamese Networks, for robust real-world applications.
- Designing a scalable and user-friendly web application that can serve as a template for similar systems in other domains.

2 System Architecture



Intended system Workflow

2.1 Overview

As we can see from the figure above, our system consists of three primary layers:

- Frontend: A ReactJS-based web application that serves as the user interface. It allows users to upload images or videos for attendance tracking and displays results in an intuitive, user-friendly format.
- Backend: The backend of the Advanced Facial Attendance System is developed using Flask, a lightweight Python web framework that provides the core infrastructure for handling API requests, orchestrating machine learning models, and managing data storage.
- Machine Learning Models (four core models):
  - Facial Recognition: Identifies individuals by comparing their facial embeddings with stored profiles.
  - Emotion Detection: Employs a custom-trained CNN to classify user emotions (Happy, Sad, Angry). This adds a layer of validation by verifying expected user behavior during attendance submission.
  - Age and Gender Classification: Uses pre-trained models for transfer learning.
  - Liveness Detection: Prevents spoofing by ensuring the user is physically present

2.2 Frontend Architecture

Frontend Features

- Image Input: Users can capture images directly on the website, the taken video is then saved to the database. For subsequent sessions, users only need to record their attendance.
- Dynamic Interface: A responsive layout adapts to various screen sizes, providing a consistent experience across desktops, tablets, and mobile devices.
- Results Display: Displays the output from the backend after model inference

Frontend Workflow

- User Interaction: Users log into the web application and are prompted to upload an image or video for attendance verification.
- Media Validation: The frontend checks for valid input formats and sizes before forwarding the data to the backend.
- API Integration: The frontend sends the media to the backend via REST API calls (e.g., POST requests).
- Once the backend processes the input, the frontend retrieves the results and displays them in an organized format.

2.3 Backend Architecture

Backend Features

- RESTful API: Handles communication between the frontend and the backend via RESTful endpoints (e.g. /check-attendance, /add-person, /check-age-and-emotion), and supports secure data transfer using HTTPS.
- Model Orchestration: Integrates multiple machine learning models for facial recognition, emotion detection, age and gender classification, liveness detection. Combines outputs to validate attendance.
- Data Storage: For testing the model: user data is stored on SQLite, and for user images stored on a directory. For future cloud development storages, we prefer using S3 encrypted with Custom keys for user images, and for metadata we select DynamoDB

Backend Workflow

- Request Handling: Receives image or video input from the frontend via a POST API endpoint; validates the input format and content.

- Preprocessing: Performs basic preprocessing on the input (e.g., resizing, normalization). Converts data into formats compatible with the machine learning models.
- Model Inference: Runs the input through the machine learning pipelines, and combines outputs from all models into a structured response.
- Result Compilation: Determines the final attendance status (Valid/Invalid) and prepares the output, including attendance details, emotions, age, gender, and liveness status.
- Response to Frontend: Sends the processed results back to the frontend for display.

3 Experiments

3.1 Datasets

The experiments utilized multiple datasets tailored to the different tasks (emotion detection, facial recognition, age and gender classification, and liveness detection):

Emotion Detection Dataset

- Source: A custom dataset (originally in video format, then split into image frames) for our own model consisting of facial expressions, captured under varying conditions (lighting, angles, and environments).
- Classes: Happy, Sad, Angry, Background.
- Size: 2000+ images per class (8,000+ images in total)
- Challenges: Imbalanced distribution across emotions, addressed using weighted loss functions and oversampling strategies.

Facial Recognition Dataset

- Source: Pre-trained MobileNetV2 used as backbone for the model, trained on Simarpreet Singh's Facial recognition dataset on Kaggle [5]
- Pairs: Positive and negative pairs were generated for training the Siamese Network, (img1, img2, label (0 if same class, 1 if different).

Age and Gender Detection Dataset

- Source: UTKFace dataset. [6]
- Pretrained Model: Used MobileNetV2 trained on this dataset for transfer learning. [7]

Liveness Detection

- Source: Integrating publicly available liveness detection models
- Dataset: Self-collected dataset

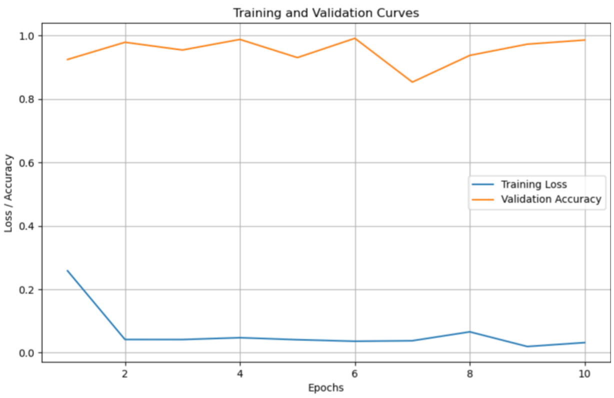
Summary of Data Preprocessing

- Videos were collected, split into frames, and preprocessed for their respective tasks.
- Augmentation and balancing techniques were applied to improve data diversity and mitigate class imbalance.
- Frames were annotated and organized into structured datasets for training and validation.

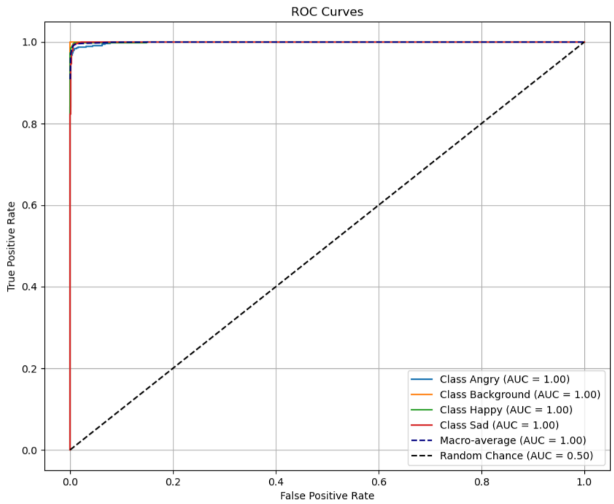
3.2 Evaluation Metrics

- ROC Curve was plotted for each class of emotions (Happy, Sad, Angry) to determine whether the model predicts the correct outcome.
- Training and Validation Accuracy: plotted across epochs for each model and to output the most accurate model.

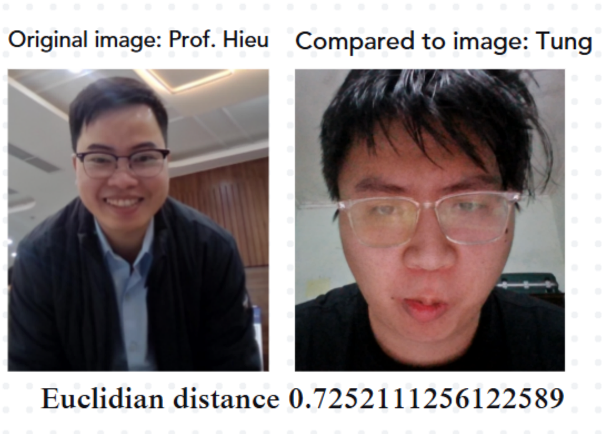
3.3 Results



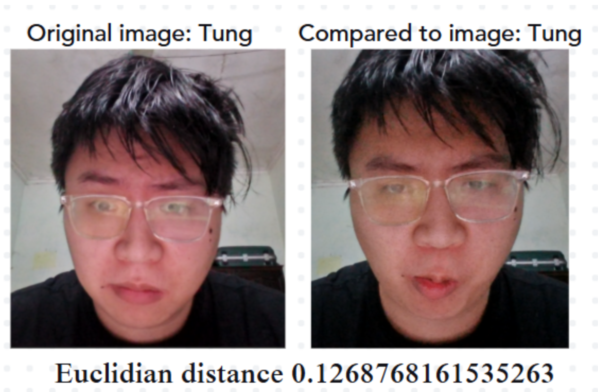
Train and Validation Accuracy for emotion detection



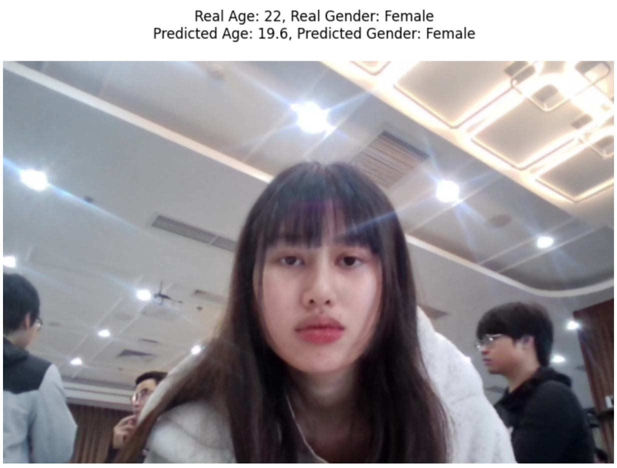
ROC Curve for Emotion Detection Model



Face recognition inference: Prof Hieu vs Tung



Face recognition inference: Tung vs Tung



An age and gender prediction

Hai age and gender prediction

Web application UI

Insights

- **Emotion detection**:

  - Observations: Training loss decreases steadily, while validation accuracy remains consistently high (90-100%), indicating effective learning and generalization.

  - Insights: The lack of divergence suggests no overfitting. However, the near-perfect validation accuracy may require testing on external datasets to confirm robustness. Furthermore, the custom dataset consists entirely of our team member's face, which may affect the adaptability of the model. In the future, we hope to sample more faces, spend more time training the data and adjusting hyperparameters to optimize this model.

- **Age and gender prediction**:

  - The model is able to predict accurage gender of new images, however the absolute difference in score is still to be considered

  - Lesson: Use age-ranges (5,10) instead of predicting actual age

- **Face recognition model**:

  - Observations: Face recognition model is able to differentiate different faces using Euclidian distance, the smaller the distance is, the more similar 2 images are

  - Challenges: In order to have accurate prediction, the user face needs to be very close to the camera, therefore future efforts to locate the human's faces for recognition is strongly advised

- **Liveness detection model**:

  - Our self-trained models were not able to accurately detect liveness in images and or videos, therefore we opted to use publicly available liveness detection models

  - Lesson: spend more time on literature of spoofing and liveness detection in order to create an accurate system

- **Web Application**:

  - The UI is simple for users to experience the demo

  - The backend was able to return the results of the facial recognition model in time

  - Lessons: We learnt from the feedback from students and professors in the live demo to help further improve the user experience

Summary: The numerous models (age/gender detection, emotion detection, facial recognition) demonstrates great performance across metrics, with minimal misclassifications. Despite this, emotion detection fails to generalize on real world data due to data imbalance. External testing and training on diverse datasets is the next step to validate robustness and address potential dataset-specific biases. Liveness detection is a diffcult task to tackle, and it requires more than just classic CNN architectures. Classic and contemporary computer vision techniques might be required in order to tackle this task

3.4 Implementation Details

Hardware:

- Training: Conducted on an NVIDIA P100 GPU with 16GB VRAM (Provided by Kaggle)
- Inference: Tested locally with Ryzen 7 8845HS with 32gb RAM

Software Frameworks:

- Backend: FastAPI framework for inference and backend logic.
- Models: PyTorch for emotion detection, age/gender detection, and Siamese Network for facial recognition.
- Frontend: ReactJS, HTML, CSS, Javascript for the user interface.

Model Training: - Emotion Detection

- Optimizer: Adam (learning rate = 0.001).
- Loss Function: CrossEntropyLoss with weighted classes.
- Epochs: 20.

- **Facial Recognition**:

- Architecture: Siamese Network with MobileNetV2 Large backbone (Using Imagenet1K weights).
- Optimizer: Adam (learning rate = 0.001).
- Loss Function: Contrastive Loss.
- Epochs: 70, saving model checkpoint every 5 elapsed epoch.

- **Age and Gender Detection**:

- Transfer Learning: MobileNetV2 fine-tuned for 25 epochs.

4 Conclusion

In this project, we developed an Advanced Facial Attendance System that combines facial recognition, emotion detection, age and gender classification into a unified and scalable solution. Using custom-trained convolutional neural networks, Siamese networks, and pretrained models, we built a robust system capable of performing real-time, multi-faceted analysis for reliable attendance validation.

The system features a ReactJS-based frontend and a Flask backend, providing a user-friendly interface and smooth model integration. Addressing challenges such as lighting variations, image quality differences, diverse facial structures, and the intricate nature of human emotions, this project tackles some of the most complex problems in computer vision. Despite the challenges, the system holds significant potential for practical applications across various domains.

Key Takeaways

- Performance:

- The system demonstrated exceptional performance in emotion detection, with validation accuracy exceeding 92.5% and near-perfect ROC AUC scores for each class.
- The best facial recognition model achieved a high accuracy, effectively distinguishing individuals under varied conditions.
- Age and gender detection were reliable, with gender classification at 85% accuracy and age predictions within a margin of ± 7 years.

- Integration: By combining multiple machine learning tasks (recognition, emotion, demographics), our model achieved a comprehensive solution for attendance tracking.

- Generalization: While the results are promising, testing on broader datasets is required to ensure performance across diverse demographics and environments. Due to the time and resource limits, we were currently unable to collect more diverse samples for our dataset.

Challenges

- Facial Recognition: Variability in lighting, angles, and occlusions posed challenges for accurate recognition. Misclassification risks increased in cases with low-quality inputs.
- Emotion Detection: Overlaps in features between certain emotions, such as "Sad" and "Angry," resulted in minor misclassifications. Balancing the dataset for underrepresented emotions required extensive augmentation. Furthermore, due to the variability of expressions in different people, extensive training and testing on more datasets is required.
- Accuracy decreased for children and elderly individuals due to limited representation in the dataset. Estimating age with precision remains challenging due to facial variations across individuals.
- Real-Time Processing: Achieving low latency while orchestrating multiple models was computationally intensive, particularly for high-concurrency scenarios.

Contributions

- Developed a novel, multi-dimensional facial attendance system capable of real-time processing.
- Developed and custom-trained an emotion detection model on a personalized dataset, achieving an adaptable model with balanced accuracy across emotions.
- Fine-tuned a Siamese network for facial recognition with high verification accuracy.
- Integrated pretrained models for age and gender classification to enrich the system's functionality.
- Designed a scalable and user-friendly application using modern web and backend technologies.

Future Directions

- Liveness detection: Devise, design, and implement a facial recognition from scratch using classic and modern computer vision approaches. Integrate that model into the existing advanced facial attendance system
- Broader Dataset Testing: Evaluate the system on larger, more diverse datasets to improve generalization and identify potential biases..
- Continuous Monitoring: Extend the system to handle video-based inputs for continuous attendance tracking in real-time.
- Mobile and Edge Deployment: Optimize models for deployment on edge devices to support low-latency processing in resource-constrained environments.
- Privacy and Security Enhancements: Implement privacy measures to ensure secure processing of user data. Implement authentication and authorization.
- Deployment on the cloud: Lift-and-shift local model to the AWS cloud, right-sizing the instance type, auto-scaling for cost optimization.

Our project offers a comprehensive solution for facial attendance systems, laying the groundwork for future advancements in multi-task facial analysis. By tackling existing challenges and broadening its capabilities, the system has the potential to evolve and adapt to meet the demands of diverse real-world applications.

References

[1] Enoch Solomon, Abraham Woubie, and Eyael Solomon Emiru. Deep learning based face recognition method using siamese network. *arXiv preprint arXiv:2312.14001*, 2023.

[2] John Smith. Challenges of traditional attendance methods in modern education. *Journal of Educational Administration*, 57(4):234–248, 2019.

[3] Richard Jones, Sarah Davis, and Linda Thompson. Evaluating security vulnerabilities in card-based attendance systems. *Security Journal*, 33(1):42–56, 2020.

[4] Charles Lee and David Kim. Facial recognition technology in attendance systems: A game changer. *International Journal of Machine Learning and Computing*, 11(2):112–127, 2021.

[5] Simarpreet Singh. Face recognition dataset. https://www.kaggle.com/datasets/cybersimar08/face-recognition-dataset, n.d.

[6] Qian Zhang. Utkface: A large-scale dataset for face age estimation and gender classification. https://susanqq.github.io/UTKFace/, 2017. Accessed: 2025-01-25.

[7] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4510–4520, 2018.