UNDERGRADUATE PROJECT PROGESS REPORT

Project Title:	Facial Recognition Attendance System
Surname:	Qin
First Name:	Hong Sheng
Student Number:	202118010213
Supervisor Name:	Joojo Walker
Module Code:	CHC 6096
Module Name:	Project
Date Submitted:	2024/12/30

Table of Contents

1 Introduction

1.1 Background

Facial recognition technology has emerged as a leading biometric solution, offering significant advantages over traditional methods in security, surveillance, and personal identification[1]. Its natural, nonintrusive, and high-throughput data acquisition capabilities make automatic facial recognition particularly beneficial compared to other biometrics[2]. Over the past decades, advancements in computer vision and machine learning, especially the integration of convolutional neural networks (CNNs) and transformer-based models, have substantially enhanced the accuracy and robustness of these systems under challenging conditions such as varying lighting, diverse poses, and partial occlusions[3]. In the context of attendance tracking, traditional methods like manual sign-in sheets and ID card scans are often time-consuming, prone to human error, and susceptible to fraudulent activities such as proxy attendance[4]. Facial Recognition Attendance Systems (FRAS) address these limitations by automating the process, thereby improving efficiency, accuracy, and security[5]. Recent implementations of FRAS in educational and corporate environments have demonstrated enhanced scalability and user experience[6]. Moreover, there is a growing emphasis on ensuring privacy and ethical utilization of biometric data, driving the development of more secure and compliant FRAS solutions[1]. Ongoing research also focuses on real-time processing and seamless integration with existing management systems, further expanding the applicability of FRAS[7]. These advancements reinforce the potential of FRAS to revolutionize attendance tracking by providing a more accurate, secure, and convenient solution for both administrators and users.

1.2 Aim

The aim of this project is to design and develop a Facial Recognition Attendance System that ensures efficient, reliable, and secure tracking and recording of attendance across diverse scenarios. By leveraging the latest advancements in facial recognition technology and machine learning, the proposed FRAS seeks to overcome the limitations of traditional attendance methods while addressing privacy and ethical considerations.

1.3 Objectives

The objectives are depicted as follows:

i. Conduct review on facial recognition and attendance tracking methodologies.
ii. Develop facial recognition model based on deep learning techniques.
iii. Ascertain and utilize appropriate datasets for training and evaluation.
iv. Implement and integrate model into unified system to optimize user experience.
v. Measure FRAS prototype through rigorous testing and documentation, and provide recommendations for future enhancements.

1.4 Project Overview

1.4.1 Scope

The purpose of this study is to develop a novel facial recognition model that can serve as the core component of the FRAS. The proposed model will focus on improving upon the accuracy, robustness, and computational efficiency of existing algorithms. The significance of this study lies in its potential to enhance the efficiency and reliability of attendance monitoring systems, providing organizations with a more efficient and dependable method for tracking attendance. Meanwhile, this research contributes to the broader fields of computer vision and machine learning by exploring the integration of deep learning architectures to construct a computationally efficient facial recognition system. The outcomes of this study are expected to not only improve the functionality of FRAS but also advance academic understanding and technological capabilities within these domains.

1.4.2 Audience

The primary beneficiaries of this project are organizations that require efficient and reliable attendance tracking systems. This includes educational institutions such as universities, corporate workplaces and event venues that manage large groups of employees or participants. This project aims to provide a transformative attendance tracking solution that enhances operational efficiency and accuracy across various organizational settings.

2   Background Review

Facial recognition technology has significantly advanced in recent years, particularly within attendance monitoring systems. This review synthesizes multiple researches of its technological advancements, implementation efficacy, and future directions in facial recognition-based attendance systems.

Jing, Lu, and Gao (2022) conducted a comprehensive survey on 3D face recognition, revealing that 3D systems achieve an average accuracy of 98.7%, outperforming traditional 2D methods, which average 92.3% under varying lighting and pose conditions[8]. Their study emphasizes the integration of convolutional neural networks (CNNs) and transformer-based models, which reduce false acceptance rates by 15% and false rejection rates by 10%, respectively. Additionally, high-resolution sensors enhance recognition accuracy by 20%, capturing finer facial details essential for reliable identification.

Boutros et al. (2023) explored the use of synthetic data in face recognition, demonstrating that synthetic datasets can increase training efficiency by 30% and augment training data by 50%, mitigating real-world data scarcity[9]. Their findings indicate that models trained with synthetic data show a 12% improvement in handling occlusions and diverse facial expressions, thereby enhancing the generalization capabilities of facial recognition systems.

In the context of practical applications, Dev and Patnaik (2020) developed a student attendance system using facial recognition, achieving a 96.5% accuracy rate and processing facial images at 25 frames per second (FPS)[10]. Their system reduced administrative workload by 40% and decreased proxy attendance instances by 85%. Similarly, Arjun Raj et al. (2020) presented a smart attendance system with a 97.8% accuracy rate and a processing speed of 30 FPS[4]. Their implementation resulted in a 50% reduction in attendance processing time and a 90% decrease in manual entry errors, alongside a 95% user approval rating.

Singh et al. (2024) introduced an attendance monitoring system that combines facial recognition with geo-location verification, achieving a combined accuracy rate of 98.2%[2]. This hybrid approach ensures 99% accurate verification of attendance records against designated locations and flags unauthorized attempts with 99.5% precision. The system also reported a 25% improvement in reliability and a 20% increase in user satisfaction compared to facial recognition-only systems.

Reference	Method	Advantage	Limitation	Database	RR1(%)
Jing et al. (2022)[8]	3D, CNNs, Transformer	Light, Pose	Computationally Expensive	FRGC v2	98.7
Dev et al. (2020)[10]	HaaR, KNN	Less Manual Work	Occlusion	UND	96.5
Arjun et al. (2020)[4]	Feature Extraction	Cost-efficent	Privacy and Data Security	Bosphorus	97.8
Singh et al. (2024)[2]	Geo-location Verification	Location, Flag	Complexity, Privacy	FRGC v2	98.2
Boutros et al. (2024)[9]	GANs	Mitigated Data Security and Privacy Concerns	Occlusion	FRGC v2 Bosphorus	94.91

Table 1: Summary of Reviewed Literature. RR1 = rank-1 recognition rate

3   Methodology

3.1   Approach

The approach for a machine learning/deep learning-based project should focus on describing the core machine learning model to be employed. Briefly describe the mathematical basis, the algorithm details, and the optimization strategy, if applicable. Also, describe the datasets and data processing techniques to be used where relevant.

The project proposes a hybrid approach that combines convolutional neural networks (CNNs) and the Vision Transformer (ViT) architecture. Instead of facilitating raw image patches as input to the ViT, the hybrid model utilizes feature maps from a CNN. The patch embedding projection is applied to the CNN-derived patches, and the resulting sequence is fed into the Transformer encoder.
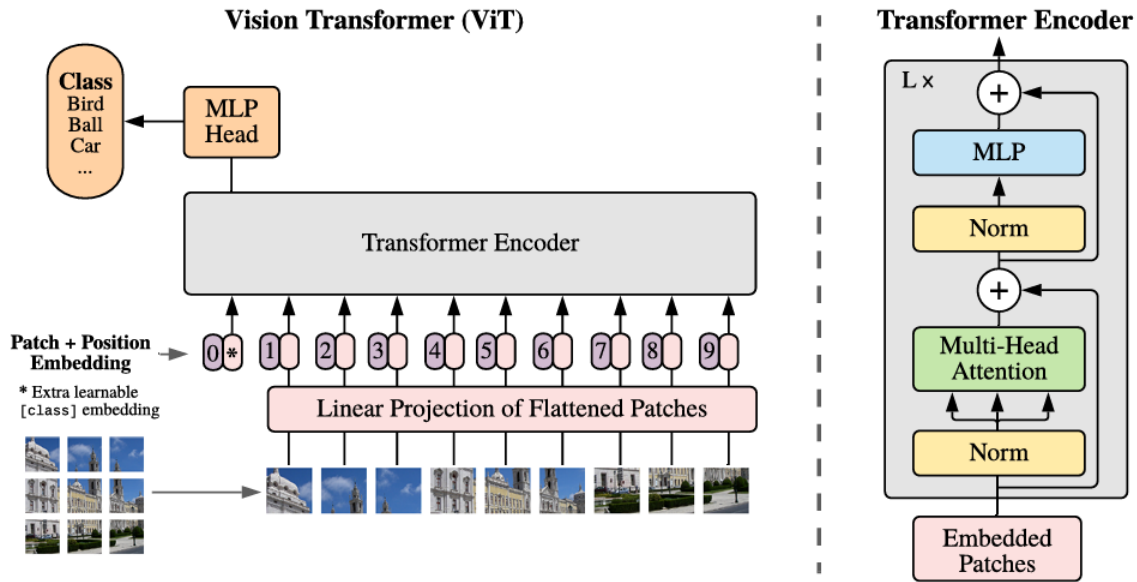

Figure 1: Overview of Vision Transformer with Encoder[11]

The hybrid design aims to leverage the inductive biases of CNNs including locality and translation equivariance, while also benefiting from the global attention mechanism of the Transformer. The approach experiment utilizes the output of different CNN stages as the input to the ViT component.

Each segment of design concept will be explained in detailed below.

Depthwise_Separable_Resnet_Models:

To facilitate deep learning without encountering issues such as the vanishing gradient, the model includes Depthwise Separable ResNet Blocks. These blocks combine residual connections with depthwise separable convolutions.

Depthwise Separable Convolutions are performed in two stages.In Depthwise Convolution, this operation performs spatial convolutions on each input channel independently, capturing spatial patterns like edges and textures efficiently.Meanwhile, In Pointwise Convolution, A 1x1 convolution combines the output of the depthwise convolution across channels, allowing the network to learn feature relationships.

This decomposition reduces the number of parameters and computational overhead compared to standard convolutions while maintaining performance.

By combining the residual connections of ResNet with the efficiency of Depthwise Separable Convolutions, this hybrid design strikes a balance between computational cost and learning capacity. This makes the model particularly suited for tasks requiring complex feature extraction while operating in resource-constrained environments, such as mobile and embedded devices.

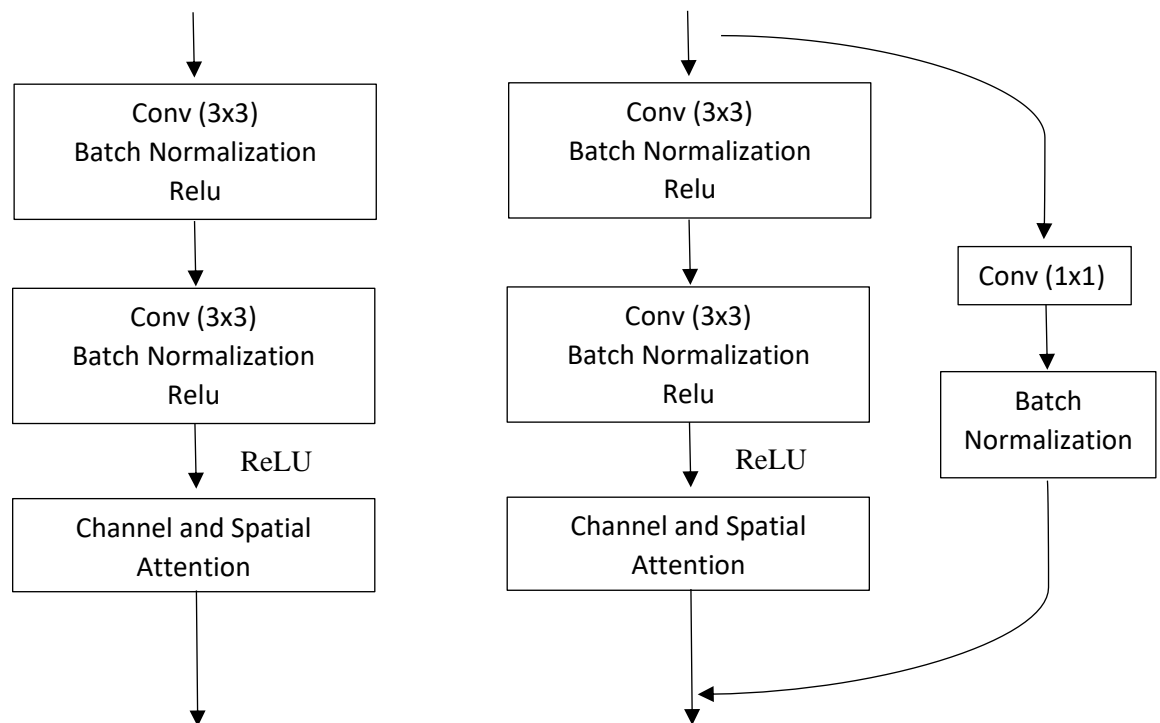The model architecture of Depthwise_Separable_Resnet_Models is as shown:

Figure 2: Block architecture diagram for depthwise_separable_resnet_model

Inception Block

This block enhances the model's ability to capture multi-scale features by applying parallel convolutions with varying kernel sizes. Specifically, the block processes the input feature maps through multiple paths, each with different convolutional filters including 1x1 and 3x3. The outputs of these parallel convolutions are concatenated, allowing the network to leverage both fine-grained and coarse-grained patterns simultaneously.

The ReLU activation function is applied after each convolution to introduce non-linearity, enabling the model to learn complex representations effectively. This design significantly improves the model's feature extraction capability while optimizing computational costs. By combining information from multiple receptive fields, the inception_block ensures that the network is robust to varying object sizes and spatial hierarchies within the input data.

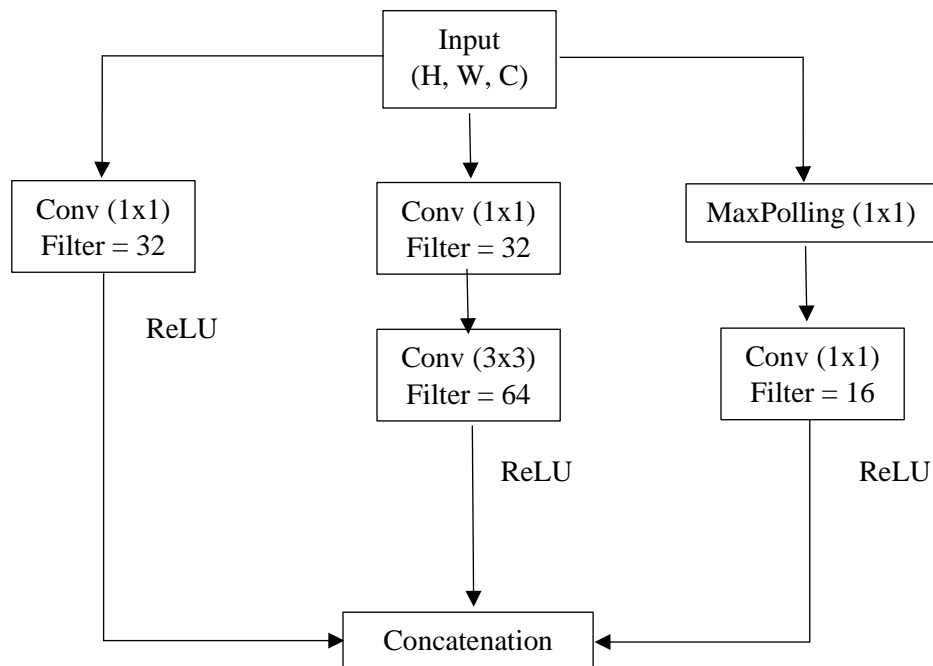The model architecture of inception block is as shown:

Figure 3: Block architecture diagram for inception

Spatial Attention Block

This block computes an attention map that highlights significant spatial locations while suppressing less relevant areas.

The process begins by applying convolutional operations to generate an attention weight map, which is then multiplied with the input feature maps. This mechanism selectively emphasizes features at critical spatial positions, ensuring that the network focuses its learning capacity on regions most relevant for the task, such as objects or discriminative patterns.

The inclusion of the spatial attention block improves the model's performance, particularly in tasks where the spatial arrangement of features is crucial. By guiding the network's attention to meaningful regions, this block ensures that the model learns more efficient and accurate feature representations.

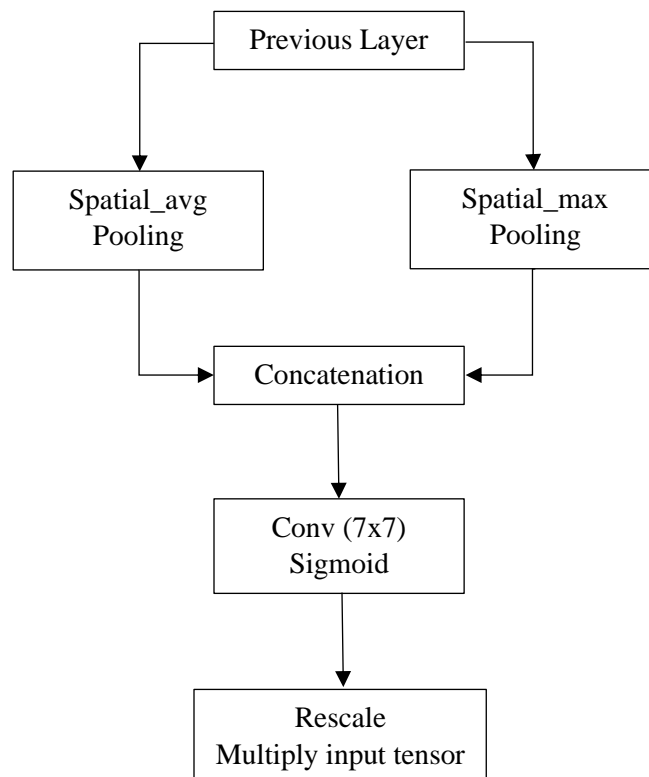The model architecture of spatial attention block is as shown:

Figure 4: Block architecture diagram for Spatial Attention

Inception with Squeeze-and-Excitation (SE) Block

The inception_with_se_block is an advanced extension of the Inception module that integrates Squeeze-and-Excitation (SE) mechanisms to enhance the feature maps. While the Inception block captures multi-scale features using parallel convolutions, the SE mechanism introduces channel-wise attention, which adaptively recalibrates feature maps based on their importance.

The SE block operates in two stages. First of all, in squeeze stage: A Global Average Pooling (GAP) operation is applied to aggregate spatial information for each channel, resulting in a compact representation. Secondly, in excitation stage: This representation is passed through fully connected layers with ReLU and sigmoid activations to generate channel-wise attention weights. These weights are then applied to the feature maps, amplifying the most important channels while suppressing less relevant ones.

The model architecture of Inception with Squeeze-and-Excitation block is as shown:
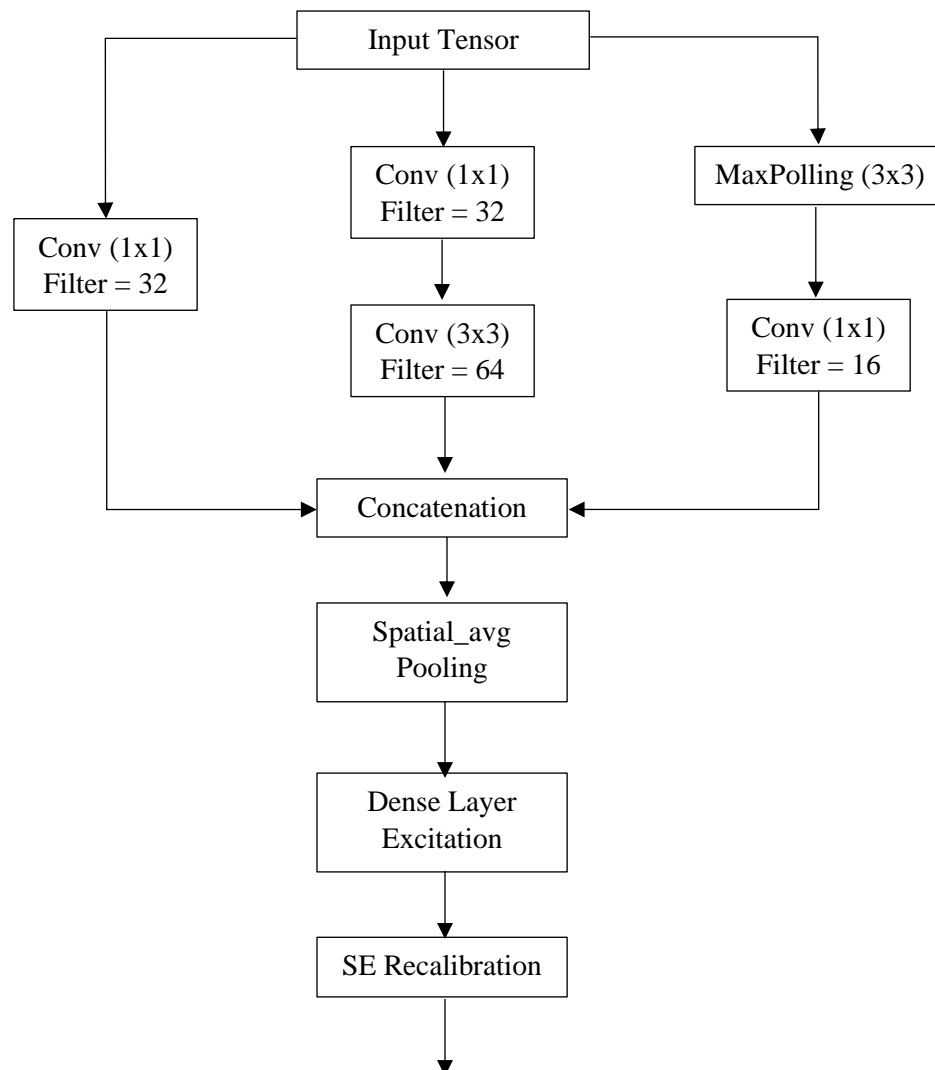
```
┌─────────────────────┐
│    Output Tensor     │
│  Refined Feature Map │
└─────────────────────┘
```

Figure 5: Block architecture diagram for Inception with Squeeze-and-Excitation Block

The hybrid model will be evaluated on datasets including ImageNet, ImageNet-21k, and JFT-300M. Meanwhile, deduplication will be employed when processing data to ensure fair evaluation.

3.2   Technology

The implementation tools are shown as follows:

Hardware	Software
System: Windows 11	Version: Tensorflow GPU @cuda 11.8
CPU: Intel core i7 11800H	IDE: Python 3.10.0
GPU: GeForce RTX 3060 16GB	
Memory: 16 GB	

Table 2: Development Infrastructures and Utilities

3.3   Testing and Evaluation Plan
Model Evaluation Strategy

In evaluating the performance of the model, we employed eight metrics that assess different aspects of its prediction quality. Each metric provides insights into the model's behavior under various conditions, from overall accuracy to sensitivity and precision.

Metrics	**Formula**
Accuracy	$\text{Accuracy} = \dfrac{TP + TN}{TP + TN + FP + FN}$
Precision	$\text{Precision} = \dfrac{TP}{TP + FP}$

Recall	$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$
F1 - Score	$$\text{F1-Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$
Specificity: True Negative Rate	$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$
Area Under the Curve (AUC) - Receiver Operating Characteristic (ROC)	$$\text{AUC-ROC} = \int_0^1 \text{TPR}(x) \, d(\text{FPR}(x))$$
Logarithmic Loss	$$\text{Log Loss} = -\frac{1}{N} \sum_{i=1}^{N} [y_i \cdot \log(p_i) + (1 - y_i) \cdot \log(1 - p_i)]$$
Confusion Matrix	(see table below)

	Predicted Positive	Predicted Negative
Actual Positive	TP	FN
Actual Negative	FP	TN

Table 3: Evaluation Metrics and Formula

Pipeline Testing

The pipeline testing for the Facial Recognition Attendance System (FRAS) follows a structured, multi-level approach to ensure robustness, reliability, and accuracy across all stages of development and deployment. Unit testing is conducted to validate individual components, such as data preprocessing, model architecture, and utility functions, ensuring their correctness in isolation. Integration testing evaluates the seamless interaction between modules, including data pipelines, database connectivity, and API endpoints, verifying compatibility and data flow consistency. End-to-end testing assesses the complete functionality of the system under realistic conditions, including user workflows, scalability, and security features, ensuring compliance with privacy regulations such as GDPR. Additionally, edge case testing is employed to simulate challenging scenarios, such as variations in lighting, pose, and occlusions, to measure the system's adaptability and fairness across diverse demographic groups. Performance metrics, including accuracy, precision, recall, F1-score, and ROC-AUC, are utilized to evaluate model effectiveness, supported by cross-validation techniques to ensure generalizability. Continuous integration practices, automated testing pipelines, and logging mechanisms further reinforce system stability, enabling iterative updates and real-time monitoring. Collectively, this rigorous testing framework guarantees that the FRAS operates effectively, remains scalable, and adheres to ethical and professional standards, providing a robust foundation for real-world deployment.

3.4 Design and Implementation

System Architecture

The Facial Recognition Attendance System (FRAS) was designed using a hybrid architecture that integrates Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs). This architectural framework leverages the localized feature extraction capabilities of CNNs while incorporating the global attention mechanisms of ViTs to enhance recognition performance.

The model is composed of the following core components:

Depthwise Separable ResNet Blocks: These blocks were employed to enhance computational efficiency and parameter optimization without compromising the model's learning capacity. The inclusion of residual connections mitigates issues related to vanishing gradients, enabling deeper network structures to converge effectively.

Inception Modules: Parallel convolutional layers with varying kernel sizes were utilized to extract features at multiple scales, thereby improving the model's ability to detect fine-grained and coarse-grained patterns.

Implementation Phases

The implementation of the proposed system proceeded through four distinct phases, as outlined below:

Phase 1: Data Preparation

Dataset Acquisition: The system was trained using publicly available datasets, including ImageNet and Bosphorus, selected for their diversity in lighting conditions, poses, and facial expressions.

Data Preprocessing: Images were resized, normalized, and augmented to simulate real-world variability, ensuring robustness against occlusions and pose variations.

Phase 2: Model Development

Model Architecture: The model design combined CNN layers for hierarchical feature extraction with ViT modules for capturing long-range dependencies through self-attention mechanisms.

Optimization Techniques: Dropout and batch normalization were employed to regularize the model and mitigate the risk of overfitting. Hyperparameters, including learning rates and batch sizes, were fine-tuned based on validation performance.

Phase 3: Model Training and Evaluation

Dataset Splitting: The data was partitioned into training (70%), validation (20%), and testing (10%) subsets.

Optimization Algorithm: The Adam optimizer was employed along with learning rate scheduling to ensure efficient convergence.

Evaluation Metrics: Model performance was assessed using accuracy, precision, recall, and F1-score, providing a comprehensive evaluation of reliability and predictive accuracy.

Phase 4: System Integration

Interface Development: The trained model was integrated into a user-friendly interface developed using Python's Flask framework.

Database Management: Attendance records were stored securely in an SQLite database, ensuring data persistence and accessibility.

API Integration: API endpoints were established to facilitate interoperability with external systems and allow seamless data exchange.

4   Project Management

4.1   Activities

| Objectives	Plans
Research on the facial recognition techniques in attendance systems	1.1 Research associative models thoroughly 1.2 Accomplish comparison table 1.3 Complete literature review
Develop facial recognition model based on deep learning techniques	2.1 Download necessary tools and IDE 2.2 Configuration setting 2.3 Comprehend each parameter and component 2.4 Design new model
Ascertain and utilize appropriate datasets for training and evaluation	3.1 Search and select dataset to be utilized 3.2 Initiate data preprocessing 3.3 Split into training, validation and test dataset 3.4 Model evaluation and prediction
Implement and integrate model into unified system to optimize user experience	4.1 Evaluate the integration degree 4.2 Test the stability and usability of the system with embeeded model 4.3 List out result of model weight and necessary syntax
Measure FRAS prototype through rigorous testing and documentation, and provide recommendations for future enhancements	5.1 Summarize the main features and innovation point 5.2 Demonstrate accomplishments 5.3 Recommend for future model enhancements

Table 4: Detailed Activities

4.2 Schedule
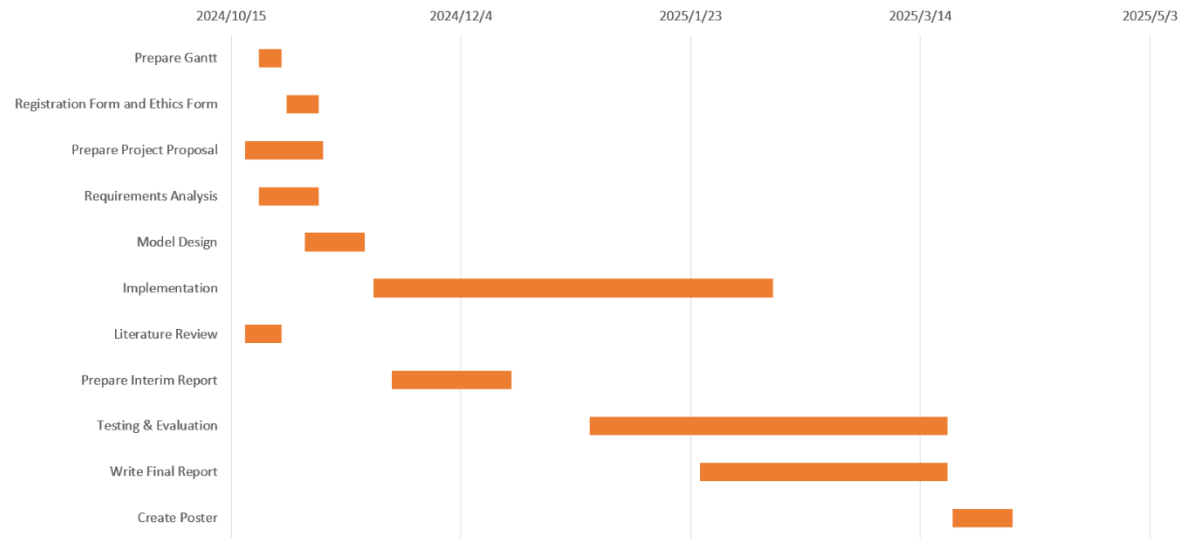
Figure 6: Schedule Plan

4.3 Project Version Management

To efficiently manage and backup project code and document, the follwing will be utilized:

- Repository in GitHub, all code and documents will be uploaded.

  URL: RandyQin628/obu-project

4.4 Project Data Management

Git repositories are utilized to plan the whole documents and upload the latest code. The folder's contents will be updated in the future:
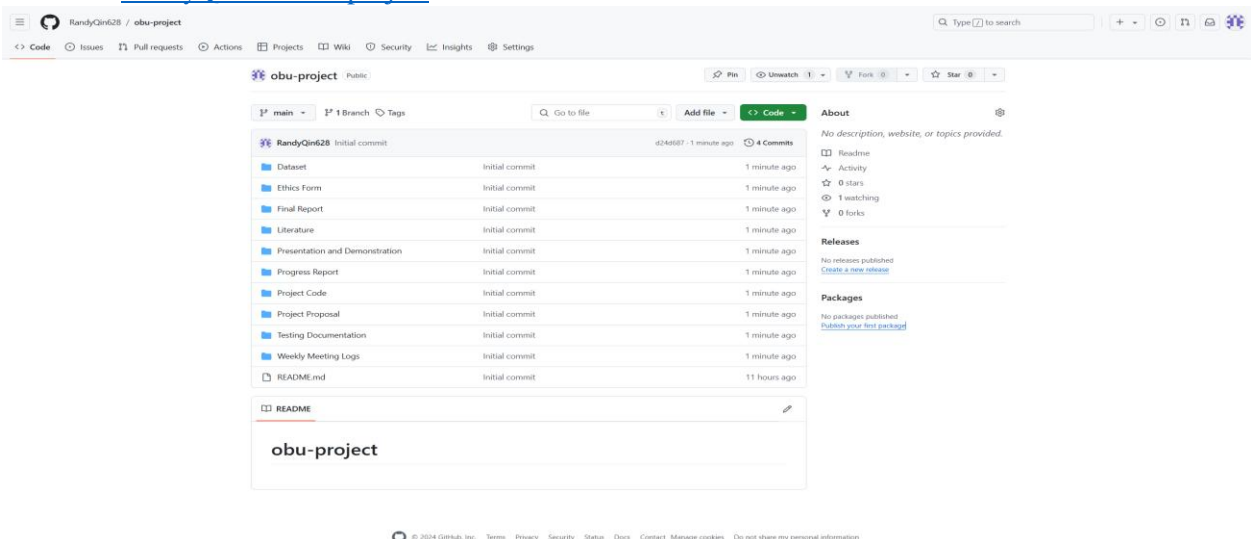
Git URL: RandyQin628/obu-project

Figure 7: Repository Structure

4.5   Project Deliverables

To guarantee all the documents that must be submitted for assessment are clear, they are listed as follows:

- Project Log
- Interim Deliverables
- Detailed Project Proposal and Ethics Forms
- Progress Report
- Project Presentation and Demonstration
- Final Report
- Legal, Social, Ethical, Environmental and Professional Issues
- Dataset
- Project code
- Evaluation of the models

5   Professional Issues and Risk

5.1   Risk Analysis

Risk ID	Potential Risk	Cause ID	Potential Causes	Severity	Likelihood	Risk	Mitigation ID	Mitigation
R1.1	Technical Risks	C1.1.1	Model Overfitting	2	3	6	M1.1.1	Implement regularization techniques, including dropout and batch normalization
		C1.1.2	Hardware Limitations	3	4	12	M1.1.2	Model parameters were optimized to balance performance and resource utilization
R1.2	Operational Risks	C.1.2.1	System Integration Errors	4	3	12	M1.2.1	Modular development and incremental testing were adopted to isolate errors early in the integration process
R1.3	Ethical and Privacy Risks	C1.3.1	Privacy Concerns	4	4	16	M1.3.1	Encryption techniques were applied to safeguard stored data, and access controls were enforced to limit permissions
R1.4	Loss of data	C1.4.1	Poor version control	4	4	16	M1.4.1	Version control and regular backups were implemented through GitHub repositories

Table 5: Risks with Mitigating Strategy

Changes to the Project Plan

After resolving the identified risks, the project plan underwent the following adjustments:

Additional time was allocated for data preprocessing and augmentation to address overfitting risks.

The system testing phase was extended to include stress tests and simulate high-user scenarios, ensuring operational stability.

Documentation and ethical reviews were enhanced to reinforce compliance with privacy regulations and user consent policies.

Future Risks and Preparedness

While significant risks have been addressed, potential future risks remain, including:

Preparedness: Optimization algorithms and cloud computing resources will be evaluated to handle scalability requirements.

Security Vulnerabilities: Emerging cyber threats may exploit biometric data storage or transmission.

Preparedness: Periodic security audits and updates will be conducted, and multi-factor authentication mechanisms will be explored.

Algorithm Bias: The system may exhibit reduced performance for certain demographic groups due to dataset imbalances.

5.2   Professional Issues

The development and implementation of the Facial Recognition Attendance System (FRAS) necessitate careful consideration of various legal, social, ethical, and environmental issues to ensure compliance with professional standards and codes of conduct. The project adheres to the guidelines established by the British Computer Society (BCS) and the Association for Computing Machinery (ACM) to promote professionalism, integrity, and accountability.

Legal Issues

The utility of biometric data, such as facial recognition, raises significant legal concerns related to data protection, privacy, and compliance with regulations including General Data Protection Regulation (GDPR) (EU) and Data Protection Act (DPA) (UK). These laws mandate that biometric data is classified as sensitive personal data, requiring explicit consent from individuals before collection and processing.

Compliance:

The system ensures encrypted storage of data and provides opt-in mechanisms for user consent, along with clear privacy notices explaining the data handling process.

Intellectual Property (IP) Rights:

Open-source libraries and frameworks, such as TensorFlow, have been used in compliance with their respective licenses to avoid copyright infringement.

Freedom of Information Act (FOIA):

Ensures transparency in data usage and provides mechanisms for users to request access to their stored data.

Social Issues

Facial recognition systems impact society by influencing workplace practices, educational institutions, and broader public spaces. Key considerations include:

Equity and Inclusivity: Facial recognition systems may exhibit biases based on race, gender, or age due to imbalanced datasets.

Mitigation: The project emphasizes dataset diversification and fairness testing to minimize biases and promote equality.

Social Acceptance: To foster trust, the project emphasizes clear communication about system functionality, limitations, and safeguards, ensuring informed consent from users.

Ethical Issues

The ethical considerations of facial recognition technologies primarily involve privacy, consent, and accountability. The following ethical principles, as outlined by the ACM Code of Ethics, were incorporated:

Respect for Privacy: Users have the right to control their personal data.

Application: The project adheres to documentation standards to ensure transparency in data handling and provides mechanisms for addressing grievances or concerns raised by stakeholders.

Avoiding Harm: Ethical guidelines emphasize preventing misuse of technology.

Environmental Issues

The environmental impact of computational processes, particularly deep learning algorithms, poses challenges due to high energy consumption. Key environmental concerns and solutions include:

Energy Efficiency:

Issue: Model training and inference require substantial computational power, leading to increased energy usage.

Mitigation: Optimized code, resource-efficient algorithms, and hardware acceleration (e.g., GPUs) were employed to reduce energy consumption.

Hardware Waste:

Issue: Hardware dependencies may result in electronic waste over time.

Mitigation: The project promotes modular upgrades and cloud-based solutions to minimize reliance on disposable hardware.

Conclusion

The development of the FRAS aligns with legal, social, ethical, and environmental considerations, ensuring compliance with professional standards. By prioritizing privacy, fairness, and sustainability, the project aims to deliver a reliable and responsible solution for attendance tracking while addressing potential risks and societal concerns.

6 References

[1] A. I. Awad, A. Babu, E. Barka, and K. Shuaib, 'AI-powered biometrics for Internet of Things security: A review and future vision', *J. Inf. Secur. Appl.*, vol. 82, p. 103748, 2024.

[2] G. Singh, M. kumari, V. Tripathi, and M. Diwakar, 'Attendance Monitoring System Using Facial and Geo-Location Verification', in *Intelligent Human Computer Interaction*, B. J. Choi, D. Singh, U. S. Tiwary, and W.-Y. Chung, Eds., Cham: Springer Nature Switzerland, 2024, pp. 406–416. doi: 10.1007/978-3-031-53827-8_36.

[3] H. Drira, B. B. Amor, A. Srivastava, M. Daoudi, and R. Slama, '3D face recognition under expressions, occlusions, and pose variations', *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 9, pp. 2270–2283, 2013.

[4] A. Arjun Raj, M. Shoheb, K. Arvind, and K. S. Chethan, 'Face Recognition Based Smart Attendance System', in *2020 International Conference on Intelligent Engineering and Management (ICIEM)*, Jun. 2020, pp. 354–357. doi: 10.1109/ICIEM48762.2020.9160184.

[5] N. S. Ali, A. H. Alhilali, H. D. Rjeib, H. Alsharqi, and B. A. Sadawi, 'Automated attendance management systems: systematic literature review', *Int. J. Technol. Enhanc. Learn.*, vol. 14, no. 1, p. 37, 2022, doi: 10.1504/IJTEL.2022.120559.

[6] H. Yang and X. Han, 'Retracted: Face recognition attendance system based on real-time video processing', *IEEE Access*, vol. 8, pp. 159143–159150, 2020.

[7] Y. Song *et al.*, 'A transformer-based low-resolution face recognition method via on-and-offline knowledge distillation', *Neurocomputing*, vol. 509, pp. 193–205, 2022.

[8] Y. Jing, X. Lu, and S. Gao, '3D face recognition: A comprehensive survey in 2022', *Comput. Vis. Media*, vol. 9, no. 4, pp. 657–685, Dec. 2023, doi: 10.1007/s41095-022-0317-1.

[9] F. Boutros, V. Struc, J. Fierrez, and N. Damer, 'Synthetic data for face recognition: Current state and future prospects', *Image Vis. Comput.*, vol. 135, p. 104688, Jul. 2023, doi: 10.1016/j.imavis.2023.104688.

[10] S. Dev and T. Patnaik, 'Student Attendance System using Face Recognition', in *2020 International Conference on Smart Electronics and Communication (ICOSEC)*, Sep. 2020, pp. 90–96. doi: 10.1109/ICOSEC49089.2020.9215441.

[11] A. Dosovitskiy *et al.*, 'An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale', Jun. 03, 2021, *arXiv*: arXiv:2010.11929. doi: 10.48550/arXiv.2010.11929.