Enhancing Gait Recognition through Edge Detection and Hybrid Feature Extraction Techniques

by

Naveen Menon

Submitted in partial fulfillment of the requirements
for the degree of Master of Computer Science

at

Dalhousie University
Halifax, Nova Scotia
August 2024

I dedicate this work to Lord Vishnu and my family.

Table of Contents

List of Tables

List of Figures

Abstract

Gait recognition is the process of identifying individuals based on their unique walking patterns. This method has gained importance in biometric systems for its non-intrusive nature and applications in security, surveillance, and healthcare, where it helps monitor and identify individuals from a distance. However, accurately distinguishing between gait patterns under varying conditions poses significant challenges. These include variations in walking speed, changes in clothing and footwear, and different environmental conditions, which all affect the accuracy of gait recognition systems. Moreover, the lack of diverse and high-quality datasets complicates the development of robust models that generalize well across different populations.

To overcome these challenges, we have developed a novel approach that incorporates various preprocessing techniques, such as edge detection, contrast enhancement, and noise reduction, with feature selection methods to improve data quality and model performance. Our hybrid feature extraction model combines Kolmogorov-Arnold Networks (KANs), ResNet, EfficientNet, and Principal Component Analysis (PCA) for spatiotemporal features, with traditional methods like Histogram of Oriented Gradients (HOG) and Local Binary Patterns (LBP). This approach bridges the research gap by demonstrating improved performance on the Chinese Academy of Sciences Institute of Automation (CASIA) datasets A, B, and C, enhancing robustness. These advancements are particularly promising for real-time applications in healthcare, such as early detection of neurological conditions like Alzheimer's disease (AD).

Evaluations using metrics such as Accuracy, F1 Score, and Area Under the Curve (AUC) have been conducted. The results show that the KAN model achieved the highest overall accuracy, surpassing 95%, while the ResNet model excelled across all metrics, effectively handling complex data variations. Integrating traditional and deep learning features boosts model accuracy and robustness. Comparative analysis with existing techniques confirms that our proposed models outperform previous methods in key metrics, validating the hypothesis that integrating diverse features and robust training strategies can substantially enhance gait recognition systems.

List of Abbreviations Used

Acronyms

AD Alzheimer's disease

AFFM Adaptive Feature Fusion Module

AUC Area Under Curve

BERT Bidirectional Encoder Representations from Transformers

CASIA Chinese Academy of Sciences Institute of Automation

CLAHE Contrast Limited Adaptive Histogram Equalization

CNNs Convolutional Neural Networks

CNN Convolutional Neural Network

DBNs Deep Belief Networks

DL Deep Learning

ED Edge Detection

GCNs Graph Convolutional Networks

Grad-CAM Gradient-weighted Class Activation Mapping

GRU Gated Recurrent Network

GRU Gated Recurrent Unit

GR Gait Recognition

HOG Histogram of Oriented Gradients

KANs Kolmogorov-Arnold Networks

LBP Local Binary Patterns

LGSD Local Graphical Skeleton Descriptor

LSTM Long Short Term Memory

LSTM Long Short-Term Memory

MLP Machine Learning

ML Machine Learning

MSFFS Multi-stage Feature Fusion Strategy

MSSTFE Multiscale Spatial-Temporal Feature Extractor

PCA Principal Component Analysis

RBMs Restricted Boltzmann Machines

ResNet Residual Network

RNN Recurrent Neural Network

SP Set Pooling

Acknowledgements

First and foremost, I want to thank God and my family for their unwavering support and encouragement throughout this journey. Their love and guidance have been my constant source of strength and inspiration. I am particularly grateful to my parents, Ramesh and Sethulaxmi Menon, whose wisdom and unwavering belief in me have been a guiding light.

I would like to express my deepest gratitude to Dr. Srinivas Sampalli for his exceptional mentorship and support. His guidance, encouragement, and confidence in my abilities have been invaluable. Dr. Sampalli's expertise and dedication have played a crucial role in both my academic and personal development, for which I am deeply thankful.

My sincere thanks also go to Dr. Darshana Upadhyay for her continuous support. Her regular check-ins, constructive feedback, and insightful suggestions have been instrumental in shaping my research journey. Dr. Upadhyay's encouragement and expertise have kept me focused and motivated.

I am also thankful to Dr. Jaume Manero for his technical expertise and support. His knowledge and assistance have been crucial in navigating the critical aspects of my research.

I am immensely grateful to my best friends, Sushumna, Keerthana, and Dheemanth, whose support has been invaluable. They have provided not only guidance but also joy and companionship, making this journey a memorable one.

Lastly, I want to acknowledge my best friends back home, Ayush and Harsh, along with all my friends and family who have helped me in various ways. Their support and encouragement have been greatly appreciated and have made this journey possible.

Chapter 1

Introduction

Gait recognition(GR) has emerged as a promising biometric technique due to its non-invasive nature and the uniqueness of an individual's walking pattern [1]. This method offers significant potential in security, surveillance, and healthcare applications, particularly in monitoring and diagnosing neurological conditions such as Alzheimer's disease(AD). The ability to recognize and analyze gait can lead to early detection of such conditions, improving patient outcomes and quality of life [2].

Gait recognition offers several unique advantages compared to other biometric methods. It is non-intrusive, allowing for recognition from a distance without requiring the subject's cooperation. Unlike fingerprints or facial features, gait is difficult to disguise, making it a reliable identifier even in covert surveillance scenarios. Additionally, gait recognition can be particularly useful in healthcare settings for monitoring changes in gait that may indicate health issues.

The increasing prevalence of Alzheimer's disease globally necessitates innovative diagnostic tools. The World Health Organization reports that nearly 50 million people worldwide are living with dementia, with Alzheimer's disease(AD) being the most common form. Early diagnosis is crucial for managing symptoms and slowing disease progression. Gait analysis, with its non-invasive approach, could play a pivotal role in identifying early signs of Alzheimer's disease(AD), as changes in gait patterns can be indicative of neurological decline [3].

1.1 Contribution

This thesis aims to advance the field of gait recognition by developing a novel hybrid feature extraction method that combines traditional handcrafted features with deep learning-based techniques. Specifically, the research introduces a new approach using Kolmogorov-Arnold Networks (KANs) [4], ResNet, EfficientNet, and Principal

Component Analysis (PCA) [5] to integrate spatiotemporal features, alongside traditional methods such as Histogram of Oriented Gradients (HOG) and Local Binary Patterns (LBP). The primary contributions of this work include:

- **Focus on Real-World Applicability:** We emphasis on creating models that are not only accurate in controlled environments but also robust and reliable in real-world scenarios, including challenging conditions like low lighting and varying environmental factors.

- **Implementation of Preprocessing Techniques:** We employ various preprocessing methods such as edge detection, contrast enhancement, and noise reduction to improve data quality and model performance.

- **Development of a Hybrid Feature Extraction Model:** The proposed model integrates traditional and deep learning features, improving the robustness and accuracy of gait recognition systems.

- **Introduction of KANs, ResNet, EfficientNet for Gait Recognition:** We explore the use of these advanced models for capturing complex spatiotemporal features in gait data.

- **Cross-Dataset Validation:** We Validate the proposed models across multiple datasets, including CASIA-A, CASIA-B, and CASIA-C, to demonstrate the generalizability and scalability of the approach.

- **Comprehensive Evaluation on CASIA Datasets:** The proposed methods are compared to existing state-of-the-art techniques to provide a detailed and nuanced analysis of of the model performance.[6].

Additionally, this research highlights the potential application of these models on edge AI [7] devices for real-time gait analysis. This is particularly relevant for continuous monitoring of patients at risk of developing Alzheimer's disease(AD), allowing for timely interventions and better management of the condition.

1.2 Organization of the Thesis

The organization of this thesis is structured as follows:

1.2.1 Chapter 2: Background Knowledge and Literature Review

This chapter provides a comprehensive overview of the foundational concepts in gait recognition, including a historical perspective, the limitations of traditional methods, and the advantages of deep learning approaches. It also includes a detailed literature review, highlighting key studies and identifying research gaps that this thesis aims to address.

1.2.2 Chapter 3: Data Preprocessing and Augmentation

In this chapter, the focus is on the preprocessing techniques used to enhance the quality of the gait datasets. It covers methods like edge detection(ED), contrast enhancement, and noise reduction. The importance of these steps in improving the accuracy and reliability of gait recognition systems is emphasized [8]. The chapter also discusses various image augmentation techniques implemented to increase the diversity of the training data.

1.2.3 Chapter 4: Feature Extraction and Selection

This chapter delves into the different feature extraction methods employed in the study, including traditional handcrafted features like Histogram of Oriented Gradients (HOG) and Local Binary Patterns (LBP), as well as deep learning-based features extracted using models like ResNet and EfficientNet. A novel feature extraction approach using Kolmogorov-Arnold Networks (KANs) is also introduced, focusing on the integration of spatiotemporal features.

1.2.4 Chapter 5: Model Training and Evaluation

The training and evaluation processes of the models are discussed in this chapter. It outlines the methodologies for integrating spatial and temporal features using 3D Convolutional Neural Networks(CNNs) and Recurrent Neural Network (RNN)/Long Short Term Memory (LSTM) architectures. The chapter also covers the ensemble techniques applied, such as EfficientNet and ResNet feature extraction combined with KANs. The evaluation metrics used to assess model performance, including accuracy, precision, recall, and AUC, are detailed.

1.2.5 Chapter 6: Experiments

This chapter presents the experimental setup and the results obtained from various experiments conducted on the CASIA datasets (CASIA-A, CASIA-B, and CASIA-C). It includes detailed analysis of the performance of different models under various conditions and provides insights into the effectiveness of the proposed methods.

1.2.6 Chapter 7: Results and Discussion

The results of the experiments are discussed in detail, with a focus on graph analysis for different models like KANs, ResNet, and EfficientNet. Comparative analysis with existing techniques are provided, highlighting the strengths and limitations of the proposed approaches.

1.2.7 Chapter 8: Conclusions

This chapter summarizes the key findings of the research, emphasizing the contributions made to the field of gait recognition. It reflects on the implications of the study and suggests areas where the findings can be applied in real-world scenarios.

1.2.8 Chapter 9: Limitations

The limitations of the study are critically analyzed in this chapter, focusing on data quality, model generalization, and computational challenges. Specific challenges related to the application of these models for detecting early onset Alzheimer's disease(AD) are discussed.

1.2.9 Chapter 10: Future Work

The final chapter outlines potential future research directions, including the deployment of the models on edge AI [9] devices for real-time gait analysis. It also explores the use of these techniques in medical diagnostics, particularly for the early detection of Alzheimer's disease(AD).

Each chapter builds on the previous ones, providing a cohesive narrative that takes the reader from foundational concepts to advanced research findings and practical applications.

Chapter 2

Background Knowledge and Literature Review

In the field of gait recognition, researchers have extensively explored various methods to identify individuals based on their walking patterns. The CASIA gait datasets, particularly CASIA-A, CASIA-B, and CASIA-C, are frequently used benchmarks for developing and testing these systems. Traditional approaches relied heavily on handcrafted features such as Histogram of Oriented Gradients (HOG) and Local Binary Patterns (LBP) [10], which capture the silhouette and texture information of the gait sequence. However, recent advancements in deep learning have significantly enhanced the performance of gait recognition systems. Models like ResNet [11] and EfficientNet [12], known for their deep feature extraction capabilities, have been employed to capture more complex patterns in gait data. Additionally, innovative architectures like Kolmogorov-Arnold Networks (KANs) [13] have been explored for their unique approach to integrating spatiotemporal features. The use of transfer learning, leveraging pre-trained models on large datasets, has also proven beneficial in improving model accuracy and generalizability. Various preprocessing techniques, including edge detection and noise reduction, are crucial for enhancing data quality, which in turn improves model performance. This literature review highlights the evolution from traditional feature-based methods to more sophisticated deep learning techniques, emphasizing the importance of hybrid feature extraction and the use of advanced neural networks in achieving state-of-the-art results in gait recognition.

2.1 Background Knowledge

Gait recognition is a biometric technique that involves identifying individuals based on their unique walking patterns [14]. This field has gained significant attention due to its non-intrusive nature and potential applications in security, surveillance, and medical diagnostics. The foundation of gait recognition lies in capturing and analyzing the dynamic motion of individuals as they walk, which is influenced by factors like body structure, walking speed, and external condition [15].

The CASIA gait datasets, including CASIA-A, CASIA-B, and CASIA-C, are some of the most widely used benchmarks in this research area. CASIA-A includes 20 subjects captured from three different viewpoints, [16] CASIA-B expands this to 124 subjects under varying conditions and angles, [17] and CASIA-C focuses on infrared imagery for low-light conditions [18]. These datasets provide a comprehensive platform for developing and testing gait recognition models under different scenarios.

Traditional methods in gait recognition primarily relied on handcrafted features such as Histogram of Oriented Gradients (HOG) and Local Binary Patterns (LBP). HOG captures gradient orientation histograms to emphasize the shape and structure of the gait silhouette, while LBP encodes local texture information by comparing pixel intensities. These features have been effective in capturing the static appearance of gait but often struggle with variations in walking conditions and environments.

The advent of deep learning has revolutionized gait recognition by enabling the extraction of more complex and abstract features. Convolutional Neural Networks (CNNs), such as ResNet and EfficientNet, have been particularly impactful. ResNet, known for its deep architecture and use of residual connections, helps in training very deep networks by mitigating the vanishing gradient problem. EfficientNet, on the other hand, optimizes model scaling in terms of depth, width, and resolution, achieving high accuracy with fewer computational resources. These models are pre-trained on large datasets like ImageNet[18], and then fine-tuned on gait datasets, a process known as transfer learning. This approach allows the models to leverage learned features, improving recognition accuracy and generalizability [19].

In addition to CNNs, innovative architectures like Kolmogorov-Arnold Networks (KANs) have been explored for gait recognition. KANs integrate both spatial and temporal features, making them well-suited for capturing the dynamic nature of gait sequences. These networks are designed to process sequential data, effectively capturing the temporal dependencies inherent in gait.

Preprocessing techniques play a crucial role in gait recognition, enhancing the quality of input data and, consequently, the performance of models. Techniques such as edge detection (using methods like Canny, Sobel, and Laplacian) [20], contrast enhancement (using CLAHE), and noise reduction (using Median and Gaussian filters)

[21] are commonly applied to improve the clarity and visibility of gait silhouettes.

This classification pertains specifically to image-based gait recognition methods. Other techniques, such as those involving floor tile-based detection, are outside the scope of this classification.It is important to note that there are other non-image-based methods, such as those involving floor tile sensors or wearable devices. These techniques detect gait patterns through different modalities, such as pressure distribution or inertial measurements, offering alternative approaches to gait recognition. However, as this thesis focuses on image-based techniques, these non-image-based methods are not included in the current classification. Overall, the combination of traditional feature extraction, advanced deep learning models, and robust preprocessing techniques has led to significant advancements in gait recognition. This field continues to evolve, with ongoing research focusing on improving accuracy, robustness, and real-world applicability.

2.1.1 History and Implications of Gait Recognition

Gait recognition, as a biometric identification technique, has a rich history rooted in the study of human motion. The concept dates back to the early 19th century when researchers began investigating the unique characteristics of human walking patterns. However, it wasn't until the advent of modern computer vision and image processing technologies in the late 20th century that gait recognition emerged as a viable biometric modality [22].

The initial developments in gait recognition focused on analyzing video sequences to extract silhouette-based features, enabling the identification of individuals based on their gait. Over time, the field evolved with the integration of more sophisticated image processing techniques and the advent of machine learning algorithms. The introduction of databases like the CASIA gait dataset provided a standardized benchmark for researchers, significantly advancing the field by allowing for consistent evaluation and comparison of different algorithms.

The implications of gait recognition are extensive and span various domains. In security and surveillance, gait recognition offers a non-invasive means of identifying individuals from a distance, even under challenging conditions where other biometric methods like face or fingerprint recognition might fail [23]. This capability is particularly useful in situations where covert identification is necessary, such as monitoring public spaces or border security.

In healthcare, gait analysis can be employed to diagnose and monitor neurological and musculoskeletal conditions [24]. Abnormalities in gait can indicate diseases such as Parkinson's, arthritis, or stroke, making it a valuable tool for early diagnosis and treatment planning [25]. Additionally, in sports science and rehabilitation, gait analysis is used to assess athletes' performance and recovery progress, providing insights into their physical health and training needs.

Furthermore, gait recognition has potential applications in personal identification and authentication systems, providing a seamless and unobtrusive user experience. As technology advances, there is growing interest in integrating gait recognition with other biometric modalities to enhance security and accuracy in multi-factor authentication systems.

Overall, the history of gait recognition reflects a progression from basic observational studies to advanced computational techniques, with significant implications for security, healthcare, and personal identification. The ongoing research and development in this field promise to further enhance its accuracy and applicability, paving the way for more innovative and practical uses in the future.

2.1.2 Drawbacks of Traditional Methods in Gait Recognition

Traditional gait recognition methods primarily rely on handcrafted features, such as Histogram of Oriented Gradients (HOG) and Local Binary Patterns (LBP). While these methods have laid the groundwork for understanding and analyzing gait patterns, they come with several limitations:

- **Limited Feature Representation:** Handcrafted features often fail to capture the complex and subtle variations in gait patterns that may be influenced by factors like clothing, walking speed, or environmental conditions. These methods typically focus on specific aspects of the gait, such as silhouette or texture, without fully capturing the dynamic nature of the movement [26].

- **Sensitivity to Variations:** Traditional methods are generally sensitive to changes in viewing angles, lighting conditions, and background clutter. For instance, silhouette-based methods can struggle with variations in shadowing or occlusion, leading to inaccuracies in feature extraction and subsequent recognition [27].

- **Inability to Handle Large Variations:** Handcrafted features often struggle with large intra-class variations, such as those caused by different footwear, carrying conditions, or changes in walking speed. This limitation affects the robustness and reliability of gait recognition systems, making them less effective in real-world scenarios where such variations are common.

- **Lack of Temporal Information:** Many traditional methods focus on static frames or use simple temporal integration, missing the detailed temporal dynamics of gait sequences. This lack of comprehensive temporal analysis can result in a loss of important information that is crucial for accurate gait recognition [28].

- **Scalability Issues:** As datasets grow larger and more complex, traditional methods often struggle to scale effectively. The computational complexity of extracting and processing handcrafted features can become a bottleneck, limiting the practical application of these methods in large-scale systems.

- **Overfitting to Specific Conditions:** Handcrafted feature extraction techniques can sometimes overfit to specific datasets or conditions they were designed for, reducing their generalizability to new, unseen environments or datasets. This overfitting can lead to decreased performance in real-world applications where conditions vary significantly [29].

2.1.3 Limitations of Machine learning[ML]

Machine learning (ML) has revolutionized various industries by enabling systems to learn from data and make predictions. However, despite its advancements, ML also faces several limitations:

- **Data Dependency:** ML models heavily rely on large datasets for training.

The quality and quantity of data directly affect the model's performance. Insufficient, biased, or noisy data can lead to poor model accuracy and generalization issues.

- **Interpretability:** Many ML models, particularly deep learning models, are often considered "black boxes." This lack of interpretability makes it challenging to understand the decision-making process, which is crucial in fields like healthcare and finance where transparency is essential [30].

- **Overfitting and Underfitting:** ML models can suffer from overfitting, where they perform well on training data but poorly on new, unseen data due to learning noise or irrelevant patterns. Conversely, underfitting occurs when models are too simple to capture the underlying data patterns, leading to poor performance even on training data [31].

- **Computational Complexity:** Advanced ML models, especially deep neural networks, require substantial computational resources for training and inference. This includes high-performance hardware like GPUs and significant energy consumption, which can be a barrier for widespread adoption.

- **Bias and Fairness:** ML models can inadvertently perpetuate biases present in the training data, leading to unfair outcomes. Addressing bias and ensuring fairness in ML models is a complex challenge that requires careful consideration and often additional methods to mitigate bias.

- **Scalability Issues:** Scaling ML models to handle vast amounts of data or to be deployed in real-time systems can be difficult. Ensuring that models remain efficient and accurate as they scale is a significant challenge.

- **Generalization:** ML models may struggle to generalize well to new or different types of data than they were trained on. This is especially problematic in dynamic environments where data distributions can change over time.

- **Security Concerns:** ML models can be vulnerable to adversarial attacks, where small, intentional changes to input data can cause the model to make incorrect predictions. This poses significant risks in applications like autonomous driving and cybersecurity.

- **Ethical and Privacy Issues:** The use of ML in sensitive areas such as personal data analysis raises ethical and privacy concerns. Ensuring that data is used responsibly and securely is critical but often challenging [32].

- **Requirement for Expertise:** Developing, tuning, and maintaining ML models requires specialized knowledge and expertise. The scarcity of skilled professionals can limit the adoption and effectiveness of ML technologies.

Despite these limitations, ongoing research and advancements continue to address these challenges, improving the robustness, interpretability, and fairness of ML models. However, it is essential to recognize and consider these limitations when developing and deploying ML systems.

2.1.4 Advantages of Deep Learning over Traditional Methods

Deep Learning (DL) offers several advantages over traditional methods, particularly in the fields of data analysis, pattern recognition, and decision-making:

- **Automated Feature Extraction:** Unlike traditional methods that rely heavily on handcrafted features, DL models automatically learn to extract relevant features from raw data. This capability allows DL models to uncover complex patterns and representations that might be missed by manual feature engineering [33].

- **Handling Complex and High-Dimensional Data:** DL excels in processing complex, high-dimensional data such as images, audio, and text. Deep neural networks, with their layered structure, can capture intricate structures and dependencies within the data, making them highly effective for tasks like image recognition, natural language processing, and speech recognition.

- **Scalability and Flexibility:** DL models can scale efficiently with increasing data volumes. As more data becomes available, DL models can continue to improve their performance, whereas traditional methods may struggle to scale and maintain accuracy. Additionally, DL architectures are flexible and can be adapted for various tasks and domains.

- **End-to-End Learning:** DL enables end-to-end learning, where models learn directly from input data to output predictions, without requiring intermediate steps of feature extraction and selection. This holistic learning approach simplifies the model development process and often leads to better performance.

- **State-of-the-Art Performance:** DL models, especially Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), have achieved state-of-the-art results in many fields, surpassing traditional methods in accuracy and efficiency. For example, CNNs have revolutionized computer vision, while RNNs and their variants, such as LSTMs and Gated Recurrent Units(GRUs), have transformed natural language processing.

- **Ability to Learn from Unstructured Data:** DL models can effectively learn from unstructured data sources, such as raw images, text, and audio, which are difficult to process using traditional techniques. This ability is particularly valuable in domains where data is unstructured and diverse.

- **Reduction of Human Intervention:** By automating feature extraction and decision-making processes, DL reduces the need for manual intervention and expertise. This not only speeds up the development process but also minimizes the risk of human bias in model development.

- **Robustness and Generalization:** DL models are generally more robust to variations in the input data, such as noise and distortions. They can generalize well to new, unseen data, making them highly suitable for real-world applications where variability is a concern.

These advantages highlight the transformative potential of deep learning in various industries, driving innovation and improving the accuracy and efficiency of data-driven solutions.

2.1.5 Deep Learning Models

Deep Learning (DL) models are a subset of machine learning algorithms that use neural networks with many layers to learn from vast amounts of data. These models have revolutionized various fields by achieving state-of-the-art performance in tasks such as image and speech recognition, natural language processing, and more. Key types of DL models include:

- **Convolutional Neural Networks (CNNs):** CNNs are specialized for processing grid-like data such as images. They utilize convolutional layers to automatically learn spatial hierarchies of features from input data. CNNs are widely used in computer vision tasks such as image classification, object detection, and segmentation. The architecture typically includes layers like convolutional, pooling, and fully connected layers, along with activation functions like ReLU [34].

- **Recurrent Neural Networks (RNNs):** RNNs are designed for sequential data, making them ideal for tasks involving time series, speech, and natural language processing. Unlike feedforward neural networks, RNNs have connections that form directed cycles, allowing them to maintain a 'memory' of previous inputs. Variants such as Long Short-Term Memory (LSTM) networks and Gated Recurrent Units (GRUs) address the limitations of standard RNNs, such as the vanishing gradient problem, by better capturing long-term dependencies [35].

- **Autoencoders:** Autoencoders are unsupervised neural networks used for learning efficient codings of input data. They consist of an encoder that compresses the input into a latent space and a decoder that reconstructs the input from this compressed representation [36]. Autoencoders are widely used for dimensionality reduction, denoising, and anomaly detection.

- **Transformer Networks:** Originally developed for natural language processing tasks, Transformers have revolutionized sequence-to-sequence tasks with their ability to model long-range dependencies without relying on recurrence [37]. The architecture uses self-attention mechanisms to weigh the importance of different elements of the input sequence, making it highly effective for translation, summarization, and more. The introduction of models like BERT (Bidirectional Encoder Representations from Transformers) and GPT (Generative Pre-trained Transformer) has further expanded the capabilities of Transformers in both language understanding and generation [38].

- **Deep Belief Networks (DBNs) and Restricted Boltzmann Machines (RBMs):** These are types of deep generative models that are capable of learning to probabilistically reconstruct their inputs. DBNs are composed of multiple layers of stochastic, latent variables, while RBMs are a specific type of energy-based models used in the first layer of DBNs for feature extraction and dimensionality reduction [39].

Each of these DL models has unique strengths and applications, making them suitable for various types of data and tasks. The choice of model often depends on the specific requirements of the problem, such as the type of data, the task complexity, and the need for interpretability.

2.2 Literature Review

2.2.1 Related work on CASIA A

Sl. No	Paper Title	Year of Publication	Journal	Dataset	Performance Evaluation Methods	Evaluation Parameters	Feature Selections	Results
1	Handcrafted Features for Human Gait Recognition: CASIA-A Dataset	2022	First International Conference on Artificial Intelligence and Data Science	CASIA-A	(SVM)using Radial Basis Function (RBF) Multi-class Winner-Take-All (WTA) SVM with threshold technology Fuzzy C-Means Clustering (FCM) Cross-Validation with a small number of folds for improving machine learning approaches	Classification accuracy Sensitivity Specificity Mean Squared Error R-squared (R^2) Relative Efficiency (RE)	Novel nodule-size approach based on ROI image segmentation and the GLCM technique with the Otsu segmentation threshold to extract features.	Peak event and reminder accuracies were 77.1% and 78.0%. Non-peak event and reminder accuracy were 83.9% and 83.2%. The average performance achieved was 81.0%.
2	Performance Evaluation of Convolutional Neural Networks for Gait Recognition	2022	IEEE Proceedings of the 24th Pan-Hellenic Conference on Information	CASIA-A,B	CNNs using metrics such as accuracy, recall, and precision. Grad-CAM (Gradient-weighted Class Activation Mapping) is used to interpret the decisions of the models.	Accuracy Recall Precision	The paper uses feature extraction methods, particularly Gait Energy Images (GEIs).	High accuracy over 90% for the dataset. The performance of the models decreases with the increase in the number of classes. The top-performing models show high accuracy related to features extracted from the middle part of the human body rather than the legs.
3	Graph Convolutional Network for Skeleton-Based Gait Recognition	2021	IEEE International Conference on Image Processing	CASIA-A,B	The performance is evaluated using rank-1 accuracy in percent. Comparison with both appearance-based and model-based methods. Ablation studies to evaluate temporal and spatial modeling	Rank-1 accuracy	Skeleton poses and Graph Convolutional Networks (GCNs) for feature extraction	Model performance were 87% 74% and 66%
4	Model-Based and Model-Free Deep Features Fusion for High-Performed Human Gait Recognition	2023	The Journal of Supercomputing	CASIA-A,B	The proposed model is evaluated using accuracy, specificity, sensitivity, false negative rate (FNR), and training time.	Accuracy, Sensitivity Specificity, False Negative Rate (FNR)Training time	Model-based: human joints, limbs, and static joint distances. Model-free: silhouette images and Gait Energy Images (GEIs).	CASIA-B dataset: 99.8% accuracyCASIA-A dataset: 99.6% accuracy
5	Research on Inception Module Incorporated Siamese Convolutional Neural Networks	2020	IEEE Access	CASIA-A,B	Comparison with classical algorithms (PCA, SVM, CNN, PCANet) Analysis of convergence speed and loss values during training	Recognition accuracy Convergence speed Loss values	Inception Module for richer feature extraction Cyclical learning rate strategy for optimizing the learning process	Higher recognition accuracy with cyclical learning rate compared to fixed learning rate

Figure 2.1: Literature review of CASIA A Papers

1. **Handcrafted Features for Human Gait Recognition: CASIA-A Dataset**[40]

This paper focuses on the use of traditional handcrafted features for gait recognition using the CASIA-A dataset. The methods include Support Vector Machine (SVM) with Radial Basis Function (RBF), multi-class Winner-Take-All (WTA) SVM, Fuzzy C-Means Clustering (FCM), and cross-validation. Feature extraction involves ROI image segmentation and the GLCM technique with Otsu segmentation threshold. The study achieved an overall accuracy of around 81%, varying for peak-event and non-peak-event conditions.

Limitations: The accuracy is relatively low, and the model struggles with varying conditions, which reduces its robustness compared to more advanced deep learning techniques .

2. Performance Evaluation of Convolutional Neural Networks for Gait Recognition [41]

This paper evaluates the performance of 18 pre-trained CNN models for gait recognition using the CASIA-A&B dataset. The study adopts a transfer learning scheme, retraining the models with Gait Energy Images (GEIs) and evaluating them using accuracy, recall, and precision. The results show that almost all models achieved high accuracy over 90%, with the best performance from features extracted from the middle part of the body.

Limitations: The performance decreases as the number of classes increases, indicating a need for models that can handle a larger diversity of gait patterns more robustly .

3. Graph Convolutional Network for Skeleton-Based Gait Recognition[42]

This paper proposes a Graph Convolutional Network (GCN) for recognizing gait based on human skeleton poses. The model is evaluated on CASIA-B and CASIA-C datasets, achieving rank-1 accuracy for various conditions: normal walking (87.7%), walking with a bag (74.8%), and wearing a coat (66.3%).

Limitations: The performance, while strong, shows a notable drop under different conditions such as walking with a bag or wearing a coat, indicating a need for models that maintain high accuracy across all conditions.

4. Model-Based and Model-Free Deep Features Fusion for High-Performed Human Gait Recognition [43]

This research presents a fusion model combining model-based features (joints, limbs, static joint distances) and model-free features (silhouette images, GEIs) for gait recognition. The evaluation on the CASIA-B and CASIA-A dataset reached above 90

Limitations: While the accuracy is high, the model's reliance on both types of features may increase computational complexity and require more extensive preprocessing compared to single-method approaches.

5. Research on Inception Module Incorporated Siamese Convolutional Neural Networks [44]

This study introduces a Siamese CNN with an Inception module and cyclical learning rate strategy for gait recognition. It compares performance with PCA, SVM, CNN, and PCANet, highlighting improvements in recognition accuracy and convergence speed.

Limitations: The model's performance, although improved with cyclical learning rates, might still be outperformed by more recent deep learning architectures specifically designed for handling diverse gait patterns .

2.2.2 Related work on CASIA B

Sl. No	Paper Title	Year of Publication	Journal	Dataset	Performance Evaluation Methods	Evaluation Parameters	Feature Selections	Results
1	A Multi-Stage Adaptive Feature Fusion Neural Network for Multimodal Gait Recognition	2024	IEEE Transactions on Biometrics, Behavior, and Identity Science	CASIA-B	Rank-1 accuracyRank-5 accuracymAP (mean Average Precision)mINP (mean Inverse Negative Penalty)	Rank-1 accuracyRank-5 accuracy mAP mINP	multi-stage feature fusion strategy (MSFFS), adaptive feature fusion module (AFFM), and multiscale spatial-temporal feature extractor (MSSTFE) to effectively select and fuse features from different modalities (silhouettes and skeletons)	93.3% rank-1 accuracy
2	Gait Set: Regarding Gait as a Set for Cross-View Gait Recognition	2019	AAAI Conference on Artificial Intelligence (AAAI-19)	CASIA-B	Comparison with state-of-the-art methods Evaluation under different training settings (Small-Sample Training (ST), Medium-Sample Training (MT), Large-Sample Training	Rank-1 accuracy Mean accuracy across various views and conditions	Set Pooling (SP) method to aggregate frame-level features into set-level features, preserving spatial and temporal information	Rank-1 accuracy for various views (average of 87.1%)
3	A Machine Learning Method with Threshold Based Parallel Feature Fusion and Feature Selection for Automated Gait Recognition	2020	Journal of Organizational and End User Computing	CASIA-B	Multi-class Support Vector Machine (MSVM)Confusion Matrix Cross-Validation.	Recognition accuracy Rank-1 accuracy Cross-view recognition performance Cross-dress and cross-speed recognition performance	Correct Classification Rate (CCR)Sensitivity Specificity Precision Recall	feature selection through threshold-based parallel feature fusion to enhance the gait recognition process
4	Gait Part: Temporal Part-Based Model for Gait Recognition	2020	IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)	CASIA- B	Temporal part-based framework for gait recognition Extensive experiments on widely used gait databases (CASIA-B)Evaluation of the proposed Gait Part model against state-of-the-art methods	Rank-1 accuracy across different conditions (normal walking, walking with a bag, wearing a coat)Cross-view evaluation Average accuracy	Focal convolution layers and micro-motion capture modules to extract fine-grained part-level spatial and temporal features	Average cross-view accuracy: 88.8%
5	Gait Graph: Graph Convolutional Network for Skeleton-Based Gait Recognition	2021	arXiv	CASIA- B	Rank-1 accuracy across various conditions (normal walking, walking with a bag, wearing a coat)Cross-view evaluation (different angles of gait data)Ablation studies to evaluate temporal and spatial modeling	Rank-1 accuracy Cross-view accuracy Average accuracy across different conditions	Graph Convolutional Networks (GCNs) to extract and model spatial and temporal features from skeleton-based gait data	Normal walking (NM): 87.7% rank-1 accuracy Walking with a bag (BG): 74.8% rank-1 accuracy Wearing a coat (CL): 66.3% rank-1 accuracy

Figure 2.2: Literature review of CASIA B Papers

1. A Multi-Stage Adaptive Feature Fusion Neural Network for Multimodal Gait Recognition [45]

This method combines silhouettes and skeletons using a multi-stage feature fusion strategy and adaptive modal fusion module. It achieves robust spatial-temporal modeling, significantly reducing feature dimensions without compromising accuracy. The network performs well on CASIA-B dataset.The model achieved a Rank-1 accuracy of 93.3%.

Limitations: The method requires careful tuning of the fusion parameters and may be less effective if the modalities (silhouettes and skeletons) are not well-aligned. Furthermore, the approach's complexity might limit its applicability in real-time scenarios due to increased processing time.

2. Gait Set: Regarding Gait as a Set for Cross-View Gait Recognition [46]

GaitSet treats gait sequences as sets of independent frames, which allows for flexibility in handling frames from different views and conditions. This method achieves high accuracy on the CASIA-B and OU-MVLP datasets, demonstrating robustness to variations in clothing and carrying conditions. It also performs well with a limited number of frames, maintaining high accuracy .The model achieved an average Rank-1 accuracy of 87.1% across various views.

Limitations: While GaitSet excels in flexibility, it may lose some temporal information due to the independent treatment of frames. This can lead to challenges in recognizing gait patterns that rely on temporal coherence.

3. A Machine Learning Method with Threshold Based Parallel Feature Fusion and Feature Selection for Automated Gait Recognition [47]

This method enhances video frames using optical flow and background subtraction before extracting texture, HOG, and geometric features. These features are fused and selected based on Euclidean distance, then classified using MSVM. The method achieves high recognition rates on the CASIA-A, CASIA-B, and CASIA-C datasets .

Limitations: The approach struggles with variations in camera viewpoints and dynamic backgrounds, limiting its effectiveness in real-world scenarios. Additionally, the method's reliance on handcrafted features may not capture all relevant information compared to deep learning-based methods.

4. Gait Part: Temporal Part-Based Model for Gait Recognition [48]

GaitPart focuses on part-level feature extraction and temporal modeling using Focal

Convolution Layer and Micro-motion Capture Module. This approach enhances fine-grained learning of part-level features and captures short-range temporal features, leading to high performance on the CASIA-B dataset.The model achieved an average cross-view accuracy of 88.8%.

Limitations: GaitPart's complex architecture may lead to increased computational costs. Additionally, while effective at capturing part-level movements, it might struggle with variations in overall body motion patterns that extend beyond short-range temporal features.

5. Gait Graph: Graph Convolutional Network for Skeleton-Based Gait Recognition [49]

GaitGraph leverages human pose estimation to recognize gait using Graph Convolutional Networks (GCNs). By focusing on skeleton poses, this method avoids issues related to silhouette images, such as loss of fine-grained spatial information and interference from visual clues. GaitGraph demonstrates state-of-the-art performance on the CASIA-B dataset, particularly in cluttered environments and with occlusions.

Limitations: The reliance on skeleton-based data means that the approach can be less effective when the pose estimation is inaccurate. Additionally, the method's performance can degrade in highly dynamic environments where poses are difficult to estimate reliably.

2.2.3 Related work on CASIA C

1. Person Recognition Based on Deep Gait [50]

This paper presents a method for person recognition using the Gait Energy Image (GEI) and deep learning techniques. It focuses on extracting spatial and temporal features from gait patterns to improve recognition accuracy. The GEI records spatial information effectively, making it suitable for identifying individuals based on their walking patterns. The approach achieved effective feature extraction and improved accuracy in gait recognition, reaching over 90%.

Limitations: GEI struggles to capture temporal information, which can reduce recognition accuracy over time. The method requires continuous and reliable data capture, which is challenging in real-time applications.

2. Appearance-based Approaches for Human Gait Recognition [51]

Sl. No	Paper Title	Year of Publication	Journal	Dataset	Performance Evaluation Methods	Evaluation Parameters	Feature Selections	Results
1	Person Recognition Based on Deep Gait	2020	MDPI Sensors	CASIA-C	Dynamic Routing Between Capsules, RNN Autoencoder	Accuracy, Sensitivity, Specificity	feature extraction using RNN Autoencoder	Effective feature extraction and improved accuracy in gait recognition over 90%
2	Appearance-based approaches for human gait recognition:	2024	The Journal of Supercomputing	CASIA-C	Various appearance-based methods including PCA and LDA	Accuracy, robustness, and computational efficiency	dimensionality reduction techniques	Effectiveness of appearance-based approaches and the importance of feature selection, accuracy was 91%
3	Gait Recognition Based on Local Graphical Skeleton Descriptor With Pairwise Similarity Network	2022	IEEE Transactions on Multimedia	CASIA-C	The paper employs the Pairwise Similarity Network (PSN) for gait recognition . Evaluation on multiple public datasets (CASIA-B, NLPR gait database, and CASIA-C) for robustness.	Recognition accuracy Rank-1 accuracy Cross-view recognition performance Cross-dress and cross-speed recognition performance	Local Graphical Skeleton Descriptor (LGSD) which includes feature extraction from both static and dynamic local patterns of skeleton sequences.	Normal walking (37.37%), slow walking (46.00%), fast walking (71.72%), walking with a bag (50.52%)
4	Gait Image Classification Using Deep Learning Models for Medical Diagnosis	2023	Computers, Materials & Continua	CASIA-C And B	Comparison with transfer learning models such as MobileNetV2, InceptionV3, VGG16, VGG19, ResNet9	Accuracy Effects of epochs Effects of data augmentation Effects of number of dropout layers Effects of regularization Effects of optimizers	feature extraction methods as part of the preprocessing step, including resizing images, normalizing pixel values, and data augmentation	CNN-LSTM model achieved an accuracy of 87.25%. proposed CNN model achieved the highest accuracy of 94.29%
5	Gait Graph: Graph Convolutional Network for Skeleton-Based Gait Recognition	2021	IEEE International Conference on Image Processing	CASIA-C and B	Comparison with classical algorithms (PCA, SVM, CNN, PCANet) Analysis of convergence speed and loss values during training	Rank-1 accuracy Comparison with other state-of-the-art (SOTA) methods Ablation studies to evaluate temporal and spatial modeling	The method uses human skeleton poses extracted from RGB images. Graph Convolutional Networks (GCNs) are used to model the spatio-temporal information of the gait.	Normal walking (NM): 87.7% rank-1 accuracy Walking with a bag (BG): 74.8% rank-1 accuracy Wearing a coat (CL): 66.3% rank-1 accuracy

Figure 2.3: Literature review of CASIA C Papers

The paper categorizes existing methods into statistical and spatiotemporal approaches, discusses the various datasets used for gait recognition, and evaluates the performance of different techniques based on common metrics like the cumulative match characteristic (CMC) curve. The survey emphasizes the advantages of appearance-based methods in terms of simplicity and effectiveness in real-world applications, such as surveillance and medical diagnosis.The paper emphasizes the effectiveness of appearance-based approaches and the importance of feature selection, achieving an accuracy of 91%.

Limitations: These methods struggle with occlusions and low-quality images, which are common in real-world scenarios. This limitation can lead to not identifying or failure to recognize individuals correctly.Appearance-based methods primarily focus on spatial features and may not adequately capture temporal dynamics of gait, which are crucial for accurate recognition.

3. Gait Recognition Based on Local Graphical Skeleton Descriptor with Pairwise Similarity Network [52]

This paper introduces a gait recognition method based on a local graphical skeleton descriptor and pairwise similarity network. It focuses on capturing the structural information of the human body for improved recognition accuracy. The model achieved varied accuracy across different conditions: Normal walking (37.37%), slow walking (46.00%), fast walking (71.72%), and walking with a bag (50.52%).

Limitations: The method's accuracy heavily depends on the effectiveness of the pose estimation process, which can be complex and error-prone.Extracting accurate skeleton data from video frames remains a challenge, especially in complex scenarios with multiple individuals or occlusions.

4. Gait Image Classification Using Deep Learning Models for Medical Diagnosis [53]

This study proposes using deep learning models, including CNN and CNN-LSTM, to classify gait silhouette images for medical diagnosis. The study demonstrates that CNN achieves the highest accuracy (94.29%) on the CASIA datasets, followed by ResNet9 (93.30%) and CNN-LSTM (87.25%).

Limitations: The models require high-quality silhouette images for accurate classification, which can be difficult to obtain in real-world settings with varying lighting and occlusions. Furthermore, the models' performance can degrade with variations in gait due to medical conditions, clothing, and carrying objects.

5. Gait Graph: Graph Convolutional Network for Skeleton-Based Gait Recognition [54]

GaitGraph utilizes Graph Convolutional Networks (GCNs) to analyze human skeleton poses for gait recognition. The approach leverages advancements in human pose estimation to create a model-based method that focuses on the cleaner representation of gait through skeletons, rather than relying on silhouette images which can include extraneous visual information.

Limitations: The method's effectiveness is highly dependent on the accuracy of the pose estimation. Inaccuracies in pose estimation can significantly impact recognition performance. Additionally, while GCNs are powerful, they can be computationally intensive and may struggle with real-time applications.

2.2.4 Findings based on Literature Review

The literature survey presents a comprehensive overview of recent advancements in gait recognition using the CASIA datasets, focusing on both traditional and deep learning methods. The key insights from the survey are summarized below:

Traditional vs Modern Approaches: The traditional methods, such as appearance-based approaches using LDA, were commonly used in earlier studies. These methods focused on dimensionality reduction and computational efficiency but often struggled with accuracy and robustness, especially under varying conditions.

Deep Learning Advancements: Recent studies have increasingly adopted deep learning techniques, which have shown superior performance in terms of accuracy and robustness. For example, papers utilizing architectures like RNN Autoencoder, Dynamic Routing Between Capsules, and Graph Convolutional Networks (GCNs) have demonstrated significant improvements in feature extraction and recognition accuracy. These methods are capable of capturing both spatial and temporal dynamics, crucial for accurate gait recognition.

Feature Extraction and Selection: The use of hybrid approaches combining handcrafted and deep learning-based feature extraction methods, such as Local Graphical Skeleton Descriptor (LGSD) and Set Pooling (SP) methods, has been effective. These techniques enhance the model's ability to handle variations in gait patterns due to different walking conditions, speeds, and environmental factors.

Evaluation Metrics and Performance: The studies utilized a range of evaluation metrics, including accuracy, precision, recall, F1 score, AUC, and Rank-1 accuracy, to comprehensively assess model performance. The comparative analysis indicates that deep learning models generally achieve higher accuracy and better generalization across different datasets and conditions compared to traditional methods.

Results and Implications: The most recent models, particularly those employing transfer learning and ensemble techniques, have achieved notable improvements in recognition accuracy. For instance, models like CNN-LSTM and CNN with transfer learning achieved accuracies of up to 94.29%. The inclusion of comprehensive preprocessing steps, such as data augmentation and normalization, further contributed to these advancements.

The literature survey highlights the evolution from traditional methods to more sophisticated deep learning approaches in gait recognition. The findings underscore the importance of advanced feature extraction techniques, robust training strategies, and the use of comprehensive evaluation metrics. These advancements have significantly enhanced the accuracy, robustness, and generalizability of gait recognition systems, making them more applicable to real-world scenarios. The survey concludes that deep learning methods, particularly those integrating multiple features and employing ensemble techniques, offer the best performance for gait recognition tasks.

2.3 Research gap

2.3.1 Identified Research Gaps in the Literature

1. Limited Handling of Diverse Walking Conditions: Many studies, including those utilizing traditional methods and some deep learning models, have not adequately addressed the variability in walking conditions, such as different speeds, carrying conditions, and environmental factors. This limitation affects the models' robustness and generalizability across real-world scenarios.

2. Inadequate Use of Spatiotemporal Features: While some papers have explored the use of spatiotemporal features, many approaches rely heavily on either spatial or temporal features, rather than effectively integrating both. This oversight can result in suboptimal recognition performance, as it fails to fully capture the dynamic nature of gait.

3. Underutilization of Advanced Deep Learning Techniques: Despite the progress in using deep learning models like CNNs and RNNs, there is a notable lack of exploration into more advanced architectures, such as Transformer models or more innovative ensemble techniques. This gap suggests potential areas for further improving accuracy and robustness.

4. Feature Selection and Fusion Limitations: Feature selection and fusion methods, crucial for handling high-dimensional data and improving model efficiency, are often underexplored or limited to basic techniques. Advanced methods for feature selection and fusion that can better leverage the strengths of multiple feature types are not thoroughly investigated.

5. Insufficient Focus on Cross-View and Cross-Condition Recognition:

The ability of gait recognition systems to perform consistently across different views and conditions (cross-view recognition) has not been extensively explored. Many models show decreased performance when tested on data from angles or conditions not seen during training.

6. Lack of Comprehensive Evaluation Metrics: While most studies report common metrics like accuracy and F1 score, there is a lack of comprehensive evaluation using a broader range of metrics, such as precision, recall, and AUC, which are crucial for a more nuanced understanding of model performance.

2.3.2 Motivation

In addressing these gaps, my research focuses on developing a robust and versatile gait recognition system by leveraging advanced deep learning techniques and comprehensive feature integration strategies:

- **Addressing Diverse Walking Conditions:** My work incorporates extensive data augmentation techniques to simulate various walking conditions and environments, enhancing the model's ability to generalize across different scenarios. This approach directly tackles the limitations seen in handling diverse walking conditions.

- **Enhanced Spatiotemporal Feature Integration:** I utilize state-of-the-art spatiotemporal models, including advanced architectures like 3D CNNs and LSTMs, to capture both spatial and temporal aspects of gait data comprehensively. This integration aims to overcome the deficiencies of previous studies that focused on either spatial or temporal features alone.

- **Exploration of Advanced Architectures:** My research explores the use of more sophisticated deep learning architectures, such as Transformer-based models and hybrid networks, which have shown promise in other domains but are underutilized in gait recognition. This exploration aims to push the boundaries of current performance limits.

- **Innovative Feature Selection and Fusion Techniques:** I investigate advanced feature selection methods and innovative fusion techniques, such as serial and parallel fusion strategies, to enhance the model's ability to handle high-dimensional data and improve efficiency. This focus addresses the gap in leveraging the strengths of multiple feature types.

- **Focus on Cross-View and Cross-Condition Performance:** Special emphasis is placed on evaluating the models across different views and conditions, using comprehensive cross-view and cross-condition testing protocols. This aspect ensures that the model is not only accurate but also robust across various scenarios.

- **Comprehensive Evaluation Metrics:** My work utilizes a broad set of evaluation metrics, including accuracy, precision, recall, F1 score, and AUC, providing a detailed and comprehensive assessment of the model's performance. This comprehensive evaluation allows for a more nuanced understanding of the strengths and weaknesses of the models.

By addressing these research gaps, my work aims to contribute significantly to the field of gait recognition, providing a more robust, accurate, and generalize solution that can perform effectively across diverse conditions and datasets.

2.4 Novelty

- **Novel Application of Kolmogorov-Arnold Networks (KANs):**
  - Introduction and application of KANs in the context of gait recognition, offering a unique approach to handling the complexities of gait data and improving the accuracy and robustness of the recognition system, which are relatively unexplored in the domain of gait recognition.

- **Comprehensive Spatiotemporal Integration:**
  - Utilization of advanced spatiotemporal models such as 3D CNNs and LSTMs [55], providing a holistic approach to capturing both spatial and temporal dynamics in gait data.

- **Innovative Feature Selection and Fusion Techniques:**

  - Development and application of novel feature fusion methods, including serial and parallel fusion strategies, to effectively combine handcrafted and deep learning-based features, enhancing the model's robustness and accuracy.

- **Robust Cross-View and Cross-Condition Recognition:**

  - Focus on ensuring high performance across various views and conditions by using extensive data augmentation techniques and testing the models under diverse scenarios, addressing a significant gap in existing literature.

- **Utilization of Advanced Evaluation Metrics:**

  - Comprehensive evaluation using a wide range of metrics, including accuracy, precision, recall, F1 score, and AUC, providing a detailed and nuanced analysis of model performance.

- **Enhanced Data Preprocessing Techniques:**

  - Application of sophisticated data preprocessing steps, such as advanced edge detection and contrast enhancement methods, tailored specifically for improving the quality of gait data under different conditions.

- **Focus on Real-World Applicability:**

  - Emphasis on creating models that are not only accurate in controlled environments but also robust and reliable in real-world scenarios, including challenging conditions like low lighting and varying environmental factors.

- **Cross-Dataset Validation:**

  - Validation of the proposed models across multiple datasets, including CASIA-A, CASIA-B, and CASIA-C, to demonstrate the generalizability and scalability of the approach.

- **Integration of Transfer Learning:**

– Effective use of transfer learning from pre-trained models, fine-tuned for gait recognition tasks, significantly reducing training time while enhancing model performance.

Chapter 3

Module 1: Data Preprocessing and Augmentation

The primary objective of Module 1 was to enhance the quality and variability of the gait dataset through advanced data preprocessing and augmentation techniques. This module is essential for ensuring the robustness and accuracy of the gait recognition models developed later in the research. The data preprocessing steps included edge detection, contrast enhancement, and noise reduction, which collectively improved the visibility and clarity of the gait images. Edge detection techniques was implemented using techniques like Canny, Sobel, and Laplacian filters to highlight structural features of the gait patterns. Contrast enhancement, specifically through CLAHE (Contrast Limited Adaptive Histogram Equalization) [56], was applied to improve local contrast and visibility, while noise reduction was achieved using Gaussian and median filtering to clean the images from unwanted noise. Following these preprocessing steps, image augmentation was performed to increase dataset variability and robustness, employing techniques such as rotation and flipping (both horizontal and vertical). These augmentations help simulate different viewing angles and conditions, making the dataset more comprehensive and the models more generalizable. This module lays the groundwork for robust and efficient feature extraction and model training, crucial for achieving high accuracy in gait recognition. Figure 3.1 provides a detailed overview of the data preprocessing and augmentation steps implemented in Module 1.

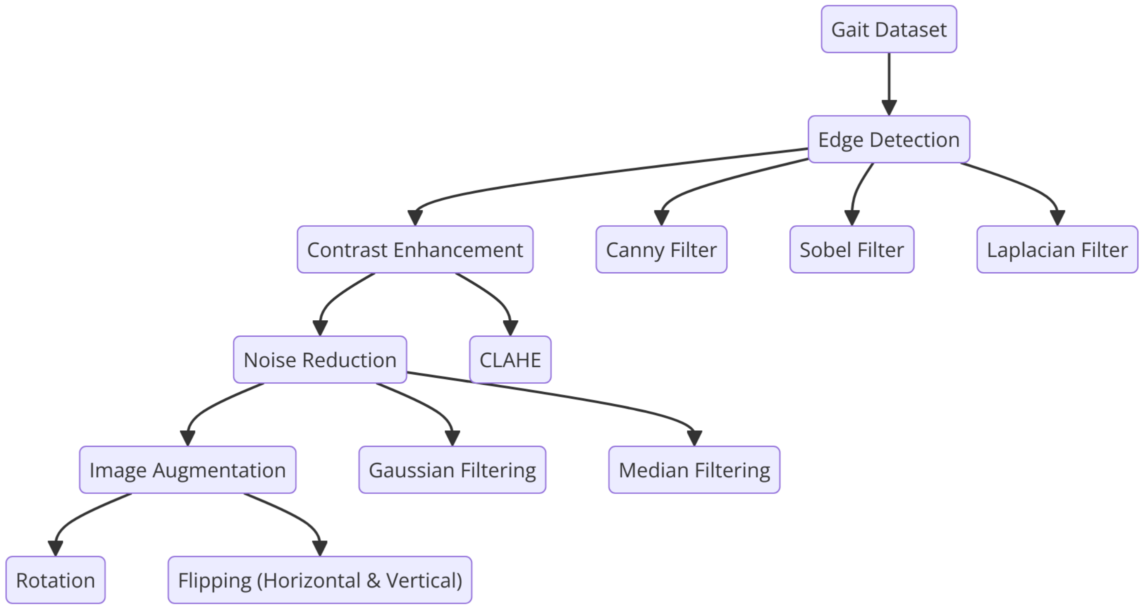

Figure 3.1: Module 1

3.1 Gait Dataset

Gait recognition datasets are pivotal for the development and evaluation of gait recognition systems. These datasets typically contain sequences of images or videos capturing the walking patterns of individuals under various conditions. The datasets vary in terms of the number of subjects, recording conditions, camera angles, and types of variations (such as clothing changes, carrying conditions, and walking speeds). Some widely used gait datasets include CASIA, OU-ISIR, and TUM-GAID, each offering unique challenges and opportunities for advancing gait recognition research.

3.1.1 CASIA Gait Datasets

The CASIA Gait Database, developed by the Chinese Academy of Sciences, is one of the most comprehensive and widely used datasets in gait recognition research. It consists of several subsets, each designed to address specific research needs.

3.1.2 CASIA A dataset

Description: CASIA-A, the earliest subset, consists of sequences of 20 subjects captured in a controlled environment. It includes 20 output labels for CASIA-A,corresponding to different individuals.

Data Collection: The dataset captures the gait sequences using a single camera, with subjects walking along a straight path.

Views: It includes variations in walking direction from three viewpoints: left, right, and frontal.

Frame Rate and Resolution: The video sequences are captured at a frame rate of 25 frames per second with a resolution of 320x240 pixels.

Applications: CASIA-A is primarily used for initial studies and proof-of-concept testing of gait recognition algorithms due to its controlled and straightforward setup. Fig 3.2 provides an overview of the CASIA A Dataset structure

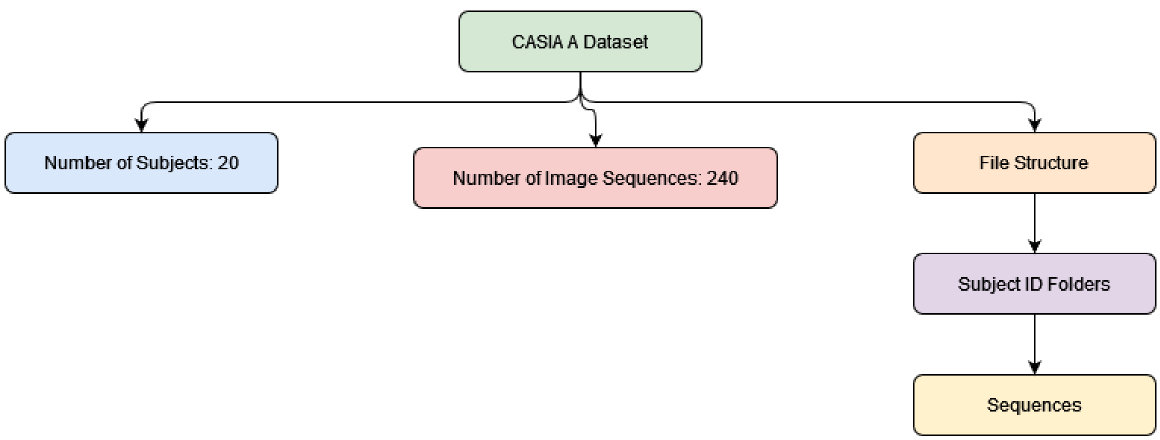

Figure 3.2: : CASIA A dataset structure

3.1.3 CASIA B dataset

Description: CASIA-B is a comprehensive dataset that includes 124 subjects. It is one of the largest and most varied gait datasets available. It includes 124 output labels for CASIA-B,corresponding to different individuals.

Data Collection: The dataset captures gait sequences using 11 cameras arranged at different angles (0°, 18°, 36°, 54°, 72°, 90°, 108°, 126°, 144°, 162°, 180°), providing a wide range of viewing angles.

Conditions: Each subject is recorded under three different conditions:

1. Normal Walking: The subjects walk without any additional items.
2. Walking with a Bag: The subjects walk while carrying a bag.
3. Clothing Variation: The subjects walk while wearing different clothing.

Frame Rate and Resolution: The video sequences are captured at a frame rate of 25 frames per second with a resolution of 320x240 pixels.

Applications: CASIA-B is ideal for developing and benchmarking robust gait recognition systems, as it provides extensive variations in view angles, walking conditions, and clothing.

Fig 3.3 provides an overview of the CASIA B Dataset structure

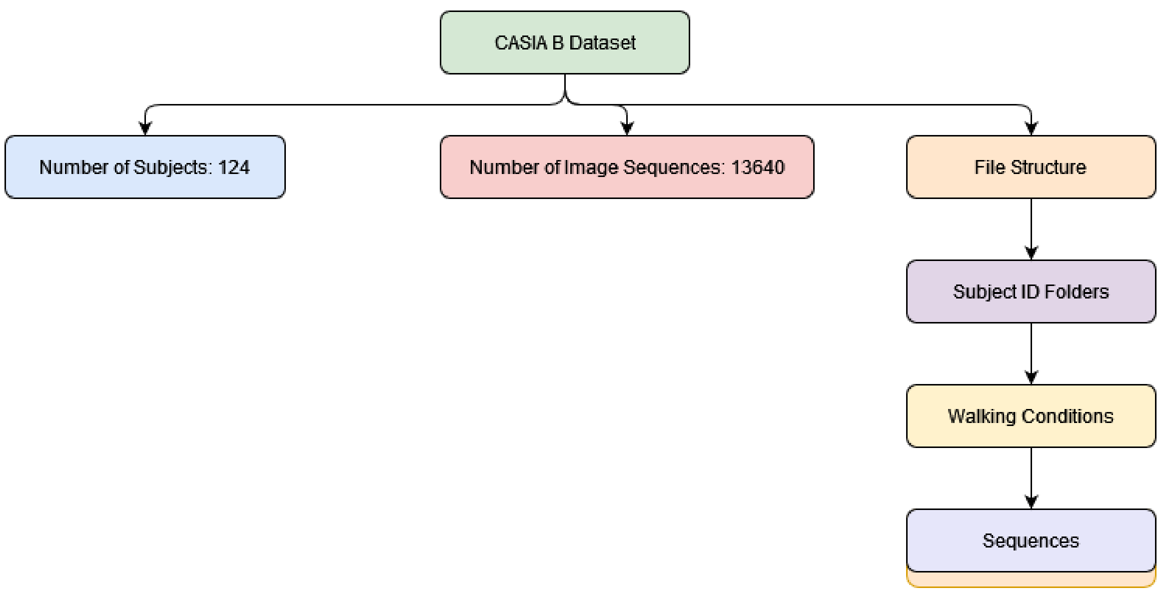

Figure 3.3: : CASIA B dataset structure

3.1.4 CASIA C dataset

Description: CASIA-C focuses on gait recognition in low-light conditions using infrared (IR) cameras, making it unique among the CASIA subsets.

Data Collection: The dataset consists of 153 subjects, captured using an infrared camera in a dark environment. It includes 153 output labels for CASIA-C,corresponding to different individuals.

Views: The subjects are recorded from four different viewpoints: 0°, 45°, 90°, and 135°.

Frame Rate and Resolution: The video sequences are captured at a frame rate of 25 frames per second with a resolution of 320x240 pixels.

Applications: CASIA-C is critical for research focused on gait recognition in night-time or low-light scenarios, which is essential for applications in surveillance and security where visibility can be a significant issue.

Fig 3.4 provides an overview of the CASIA C Dataset structure

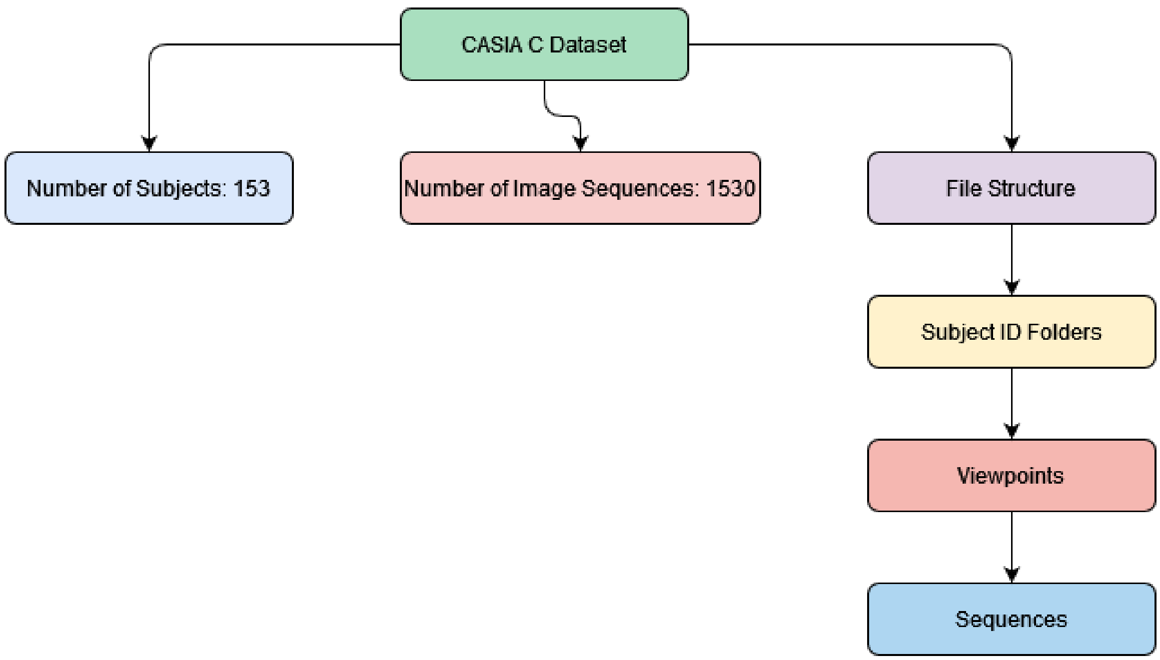

Figure 3.4: : CASIA C dataset structure

3.1.5 Importance of CASIA Datasets

The CASIA-A, CASIA-B, and CASIA-C datasets are among the best choices for gait recognition research due to several key reasons:

1. **Diverse Conditions:** They cover a wide range of conditions, including different viewing angles, walking scenarios (with and without carrying items), clothing variations, and low-light environments. This diversity is crucial for developing robust and versatile gait recognition algorithms.

2. **Wide Range of Views:** CASIA-B provides 11 different viewing angles, while

CASIA-C offers 4 views under infrared conditions. This variety is greater compared to other datasets, such as TUM GAID, which typically offers fewer viewing angles.

3. **Infrared Modality:** CASIA-C includes infrared data, which is rare among gait datasets and valuable for research in low-light or night-time conditions. This feature is particularly useful for surveillance applications where visual data might be insufficient.

4. **Large Subject Pool:** The datasets include a significant number of subjects, particularly CASIA-B and CASIA-C, providing a rich and varied pool of gait patterns for training and evaluation.

5. **Controlled and Uncontrolled Settings:** The datasets offer both controlled (CASIA-A and CASIA-B) and uncontrolled (CASIA-C) settings, allowing researchers to test their algorithms in ideal as well as challenging conditions.

6. **Benchmark Status:** CASIA datasets have become a standard benchmark in the gait recognition community, facilitating the comparison of new algorithms with existing state-of-the-art methods.

3.1.6 Comparison of Gait Recognition Datasets

Dataset	Subjects	Conditions	Views	Modality	Notable Features
CASIA-A	20	Normal	3 (Left, Right, Front)	Visual	Controlled environment, initial studies
CASIA-B	124	Normal, Bag, Clothing Variations	11 (0° to 180°)	Visual	Extensive views and conditions, large subject pool
CASIA-C	153	Normal	4 (0°, 45°, 90°, 135°)	Infrared (IR)	Low-light conditions, night-time surveillance applications
OU-ISIR LP	4,016	Normal	Multiple	Visual	Very large subject pool, high resolution, limited condition variability
TUM GAID	305	Normal, Carrying, Shoes	3 (Frontal, Lateral)	Visual, Depth, IMU	Multimodal data, varied conditions, focus on daily activities
Soton	115	Normal, Different Speeds	Multiple	Visual	High resolution, varied walking speeds
OU-ISIR MVLP	10,307	Normal	14 (0° to 180°)	Visual	Extremely large subject pool, comprehensive view angles, normal condition

Table 3.1: Comparison of Gait Recognition Datasets

CASIA-A is crucial for initial studies and controlled experiments, providing a consistent environment to benchmark new techniques.

CASIA-B stands out for its comprehensive view angles and multiple conditions (normal, carrying a bag, and clothing variations), offering a robust dataset for developing and testing models under varied scenarios.

CASIA-C is unique for its infrared (IR) modality, specifically designed for low-light and night-time surveillance, making it highly valuable for real-world security applications where visibility is poor.

While **OU-ISIR LP** has a very large subject pool and high resolution, it lacks the condition variability seen in CASIA-B, making it less comprehensive for evaluating models under diverse conditions.

TUM GAID is valuable for its multimodal data and varied conditions but has fewer views and a smaller subject pool compared to CASIA-B and CASIA-C, limiting its scope for comprehensive model evaluation.

Soton offers high resolution and varied walking speeds but does not match the extensive view angles and condition variability of CASIA-B.

OU-ISIR MVLP has the largest subject pool and comprehensive view angles, but it primarily focuses on normal walking conditions, lacking the varied conditions provided by CASIA-B.

3.2 Data Preprocessing and Augmentation

3.2.1 Edge Detection

Edge detection is a critical process in image processing and computer vision, focused on identifying significant local changes in intensity within an image. These changes often represent boundaries of objects, which are crucial for understanding the structure and content of an image.

Role of Edge Detection in Gait Recognition:

In gait recognition, edge detection plays a pivotal role by emphasizing the structural details of human silhouettes. This enhancement allows for better extraction of relevant features, improving the accuracy and efficiency of the recognition process. Figure 3.5 outlines the comprehensive role of edge detection in this context.

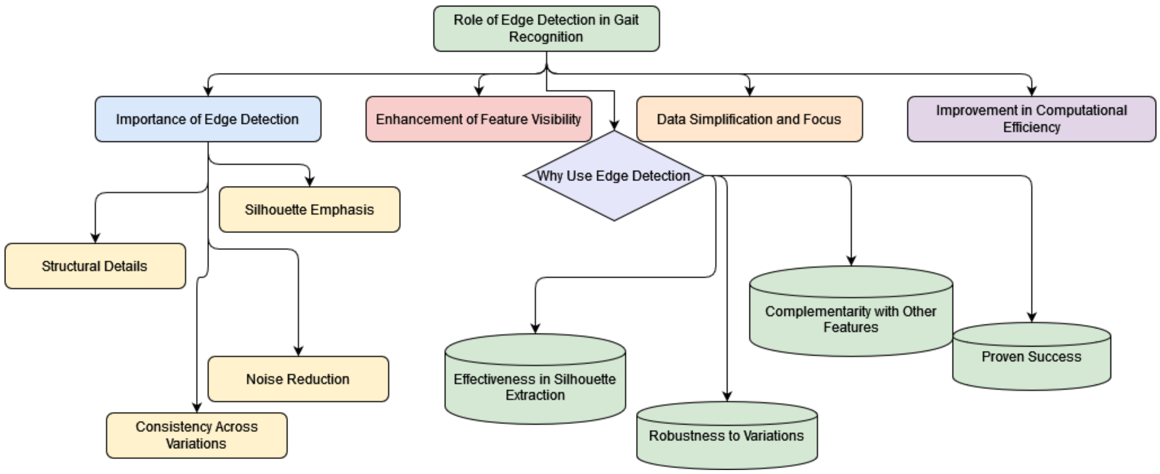

Figure 3.5: : Role of Edge Detection

Importance of Edge Detection:

1. Structural Details: Edge detection highlights the crucial structural aspects of the silhouette, which are essential for distinguishing between different gait patterns.

2. Consistency Across Variations: By focusing on edges, the method remains robust to variations in lighting, clothing, and background.

3. Noise Reduction: Techniques like edge detection help in reducing noise, thereby improving the clarity and quality of the extracted features.

4. Silhouette Emphasis: Edges accentuate the boundaries of the silhouette, making it easier to isolate the human figure from the background.

Enhancement of Feature Visibility:

Edge detection significantly enhances the visibility of important features in an image. By focusing on the boundaries and edges of objects, edge detection algorithms make it easier to distinguish different components of an image, which is particularly useful in gait recognition. This visibility enhancement is crucial for identifying and analyzing the distinct movements and postures that characterize an individual's gait.

Why use Edge Detection:

1. Effectiveness in Silhouette Extraction: Simplifies the data by focusing on the most informative parts, which are the edges.

2. Complementarity with Other Features: When combined with other feature extraction techniques, edge detection enhances the overall feature set.

3. Robustness to Variations: Edge detection methods like Canny, Sobel, and Laplacian are robust to various changes, ensuring reliable performance.

4. Proven Success: These techniques have been extensively tested and proven successful in numerous computer vision applications.

Data Simplification and Focus:

Edge detection simplifies the data by reducing the amount of information that needs to be processed. Instead of analyzing every pixel in an image, the focus is shifted to the pixels that form the edges, which represent the most critical features. This reduction in data complexity helps in creating more efficient algorithms for gait recognition.

Improvement in Computational Efficiency:

Edge detection contributes to computational efficiency by reducing the amount of data that needs to be processed and analyzed. This efficiency is critical in real-time applications like gait recognition systems, where quick and accurate processing is required.

Choosing Edge Detection Techniques:

For my research, three edge detection techniques were selected: Canny, Sobel, and Laplacian. Each of these techniques offers unique advantages:

1. Canny Edge Detection: Known for its ability to detect edges with low error rate and being able to detect true edges while minimizing noise. It uses a multi-stage algorithm to achieve optimal results.

2. Sobel Edge Detection: Utilizes gradient approximation to detect edges, making it very effective for highlighting changes in intensity in horizontal and vertical directions.

3. Laplacian Edge Detection: Employs second-order derivatives to find regions of rapid intensity change, providing a high level of detail and accuracy. These techniques were chosen over others due to their proven effectiveness in highlighting the critical features of gait patterns while maintaining robustness against noise and variations.

Why Canny, Sobel, and Laplacian are Preferred

1. Canny Edge Detection: Accuracy: The Canny edge detector is known for its high accuracy and ability to detect edges in images with a low error rate. It uses a multi-stage process that includes noise reduction, gradient calculation, non-maximum suppression, and edge tracking by hysteresis.
   Edge Localization: Canny excels at precisely locating the position of edges, making it suitable for applications requiring high precision.

2. Sobel Edge Detection: Simplicity: Sobel is straightforward to implement and computationally efficient, making it a popular choice for basic edge detection tasks.
   Edge Orientation: It effectively detects edges and their orientation, which is valuable in many image processing applications.

3. Laplacian Edge Detection: Detail Detection: Laplacian is adept at detecting fine details and regions of rapid intensity change, which is crucial for applications that require detailed edge information.
   Complements Other Methods: Often used in conjunction with other edge detectors to enhance edge detection capabilities.The table 3.2 below gives a broader comparison

Technique	Description	Advantages	Disadvantages
Canny	Uses a multi-stage algorithm to detect a wide range of edges in images.	High accuracy, low error rate, good localization.	Computationally intensive, sensitive to noise.
Sobel	Computes the gradient of the image intensity at each pixel, emphasizing edges.	Simple, effective for detecting edges and their orientation, easy to implement.	Sensitive to noise, may miss fine details.
Laplacian	Highlights regions of rapid intensity change and is based on the second derivative of the image.	Detects fine details and edges, effective for highlighting regions of rapid intensity change.	Sensitive to noise, can detect false edges.
Prewitt	Similar to Sobel but uses a different kernel for edge detection.	Simple and fast, effective for basic edge detection tasks.	Less accurate than Canny, sensitive to noise.
Roberts Cross	Uses a pair of 2x2 convolution kernels to perform edge detection.	Simple, quick to compute, effective for high-contrast edges.	Sensitive to noise, not suitable for detecting complex edges.
Scharr	An improved version of Sobel with better rotational symmetry.	Better at detecting edges than Sobel, less sensitive to noise.	Computationally more intensive than Sobel.
Frei-Chen	Uses a set of masks to detect edges in different directions.	Capable of detecting edges in multiple directions, effective for edge enhancement.	Complex, computationally intensive, sensitive to noise.

Table 3.2: Comparison of Edge Detection Techniques

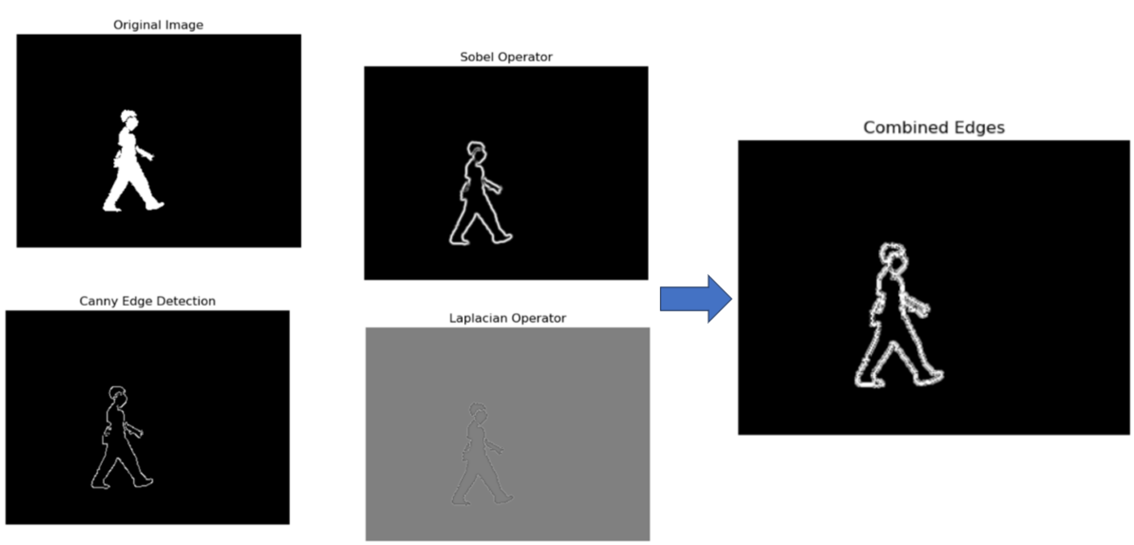

Figure 3.6: : Edge Detection Techniques

Implementation of Edge Techniques

Canny Edge Detection

The Canny edge detection algorithm is known for its ability to detect a wide range of edges in images. It involves several steps:

1. Noise Reduction: The image is smoothed using a Gaussian filter to reduce noise. This is achieved by convolving the image with a Gaussian kernel:
   The function $G(x, y)$ is given by:
   $$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}$$

2. Gradient Calculation: The gradient intensity and direction are computed using Sobel filters: The gradients G_x and G_y are given by:
   $$G_x = \frac{\partial I}{\partial x}, \quad G_y = \frac{\partial I}{\partial y}$$
   The gradient magnitude and direction are then given by:
   $$G = \sqrt{G_x^2 + G_y^2}, \quad \theta = \tan^{-1}\left(\frac{G_y}{G_x}\right)$$

3. Non-Maximum Suppression: Thinning the edges to remove non-maximum pixels, retaining only local maxima in the gradient direction.

4. Double Thresholding: Identifying strong, weak, and non-relevant pixels based on two thresholds. Strong edges are retained, weak edges are considered if connected to strong edges.

5. Edge Tracking by Hysteresis: Final edges are determined by suppressing all edges that are not connected to a strong edge.

Sobel Operator The Sobel operator computes the gradient magnitude of the image using convolution with Sobel kernels. It detects edges by emphasizing regions of high spatial frequency which correspond to edges. Sobel kernels:

$$K_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}, \quad K_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}$$

Gradient computation:

$$G_x = I * K_x, \quad G_y = I * K_y$$

Gradient magnitude and direction:

$$G = \sqrt{G_x^2 + G_y^2}, \quad \theta = \tan^{-1}\left(\frac{G_y}{G_x}\right)$$

Laplacian Operator The Laplacian operator detects edges by computing the second derivative of the image, highlighting regions of rapid intensity change. Laplacian kernel:

$$K = \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

Convolution with the Laplacian kernel to obtain the edge response:

$$G = I * K$$

Edge Detection Algorithm

Algorithm 1 Edge Detection

Require: Input image I

Ensure: Image with detected edges I_{edges}

Step 1: Convert to Grayscale

$I_{gray} \leftarrow$ `cv2.cvtColor`(I, `cv2.COLOR_BGR2GRAY`)

Step 2: Apply Canny Edge Detection

$edges\_canny \leftarrow$ `cv2.Canny`(I_{gray}, 100, 200)

Step 3: Apply Sobel Edge Detection

$edges\_sobel \leftarrow$ `cv2.Sobel`(I_{gray}, `cv2.CV_64F`, 1, 1, ksize=5)

Step 4: Apply Laplacian Edge Detection

$edges\_laplacian \leftarrow$ `cv2.Laplacian`(I_{gray}, `cv2.CV_64F`)

Step 5: Combine Edge Detection Results

$combined\_edges \leftarrow edges\_canny + edges\_sobel + edges\_laplacian$

Step 6: Convert Combined Edges to 8-bit Image

$I_{edges} \leftarrow$ `cv2.convertScaleAbs`($combined\_edges$)

Step 7: Convert to BGR for Output

$I_{edges\_bgr} \leftarrow$ `cv2.cvtColor`(I_{edges}, `cv2.COLOR_GRAY2BGR`)

Return I_{edges_bgr}

3.3 Contrast Enhancement and Noise Reduction

Contrast enhancement is a crucial preprocessing step in image processing that aims to improve the visibility of features within an image. By increasing the contrast, the difference between light and dark areas is accentuated, making important features more distinguishable. This is particularly important in applications like gait recognition, where subtle variations in the silhouette and edges of a person's gait can significantly impact the accuracy of recognition algorithms. Techniques like Contrast Limited Adaptive Histogram Equalization (CLAHE) are widely used for this purpose.

Noise reduction involves the removal of unwanted random variations in intensity (noise) from an image. Noise can significantly degrade image quality, making it difficult to identify and analyze features accurately. Techniques such as Gaussian and median filtering are commonly used to smooth the image while preserving important edges.

3.3.1 Role of Contrast Enhancement and Noise Reduction in Gait Recognition

1. **Improvement in Silhouette Extraction:** Contrast Enhancement: By enhancing the contrast in gait images, the boundaries and contours of the silhouette become more defined, making it easier to extract accurate silhouettes for further analysis. This is critical in gait recognition as the silhouette is a primary feature used for identifying individuals.
   Noise Reduction: Removing noise from the image helps in obtaining cleaner silhouettes, which reduces errors in the feature extraction process. This leads to more reliable gait recognition.

2. **Enhanced Feature Discrimination:** Contrast Enhancement: Increases the visibility of subtle features that may be important for distinguishing between different gait patterns. This is particularly useful when dealing with variations in clothing or carrying conditions, as the enhanced contrast can highlight unique walking characteristics.

Noise Reduction: By reducing random variations and noise, the features extracted are more consistent and reliable, improving the discrimination power of the recognition system.

3. **Consistency in Variable Conditions:** Contrast Enhancement: Ensures that gait features are consistently visible under varying lighting conditions, which is crucial for real-world applications where lighting can change drastically.
   Noise Reduction: Provides consistent image quality, even in the presence of environmental noise, making the gait recognition system more robust across different scenarios.

4. **Increased Accuracy in Feature Matching:** Contrast Enhancement: Enhances the key features that are used for matching gait patterns, leading to higher accuracy in recognizing individuals.
   Noise Reduction: Reduces false matches caused by noise, thereby increasing the overall accuracy of the recognition system.

5. **Reduced Computational Load:** Contrast Enhancement: With clearer and more distinct features, the computational algorithms can process the images more efficiently, leading to faster recognition times.
   Noise Reduction: Cleaner images require less processing power to filter out irrelevant information, thus optimizing the computational efficiency of the recognition system.

Why Choose CLAHE, Median Filtering, and Gaussian Filtering?

CLAHE (Contrast Limited Adaptive Histogram Equalization):

1. Localized Enhancement: Enhances contrast in small regions, making it effective in highlighting fine details without amplifying noise.

2. Adaptability: Adjusts to local variations in the image, providing a more balanced contrast enhancement.

Median Filtering:

Table 3.3: Table of Contrast Enhancement and Noise Reduction Techniques

Technique	**Purpose**	**Advantages**	**Disadvantages**
Histogram Equalization	Contrast Enhancement	Simple, effective for global contrast	May over-enhance noise
CLAHE	Contrast Enhancement	Enhances local contrast, prevents over-amplification of noise	More complex, computationally intensive
Median Filtering	Noise Reduction	Preserves edges, effective against salt-and-pepper noise	Can be computationally expensive
Gaussian Filtering	Noise Reduction	Smooths image, reduces Gaussian noise	Blurs edges, may not remove all types of noise
Bilateral Filtering	Noise Reduction	Preserves edges while reducing noise	Computationally intensive

1. Edge Preservation: Effectively removes salt-and-pepper noise while preserving the edges, which are crucial for accurate silhouette extraction in gait recognition.

2. Robustness: Handles outliers effectively, making it suitable for images with various types of noise.

Gaussian Filtering:

1. Smooth Blurring: Reduces high-frequency noise without significantly blurring the image, maintaining important details for feature extraction.

2. Efficiency: Simple to implement and computationally efficient, making it a practical choice for preprocessing large datasets.

3.3.2 Implementation of Contrast Enhancement and Noise Reduction

1. CLAHE:

$$\text{CLAHE}(I, \text{clipLimit}, \text{tileGridSize}) = \text{adaptiveEqualizeHist}(I, \text{clipLimit}, \text{tileGridSize})$$

Where I is the input image, clipLimit controls the contrast limiting, and tileGridSize defines the size of the grid for histogram equalization.

2. Median Filtering:

$$I_{\text{median}}(x, y) = \text{median}(\{I(x+i, y+j) \mid -k \leq i, j \leq k\})$$

Where I is the input image, and the filter replaces the pixel value at (x, y) with the median of its neighbors within a $k \times k$ window.

3. Gaussian Filtering:

$$I_{\text{gaussian}}(x, y) = \frac{1}{2\pi\sigma^2} \sum_{i=-k}^{k} \sum_{j=-k}^{j} I(x+i, y+j) e^{-\frac{i^2+j^2}{2\sigma^2}}$$

Where I is the input image, σ is the standard deviation of the Gaussian kernel, and the summation is performed over a $k \times k$ window centered at (x, y).

Algorithm 2 Contrast Enhancement and Noise Reduction

Require: Input image I

Ensure: Enhanced and noise-reduced image I_{output}

Step 1: Convert to Grayscale

$I_{gray} \leftarrow$ `cv2.cvtColor`(I, `cv2.COLOR_BGR2GRAY`)

Step 2: Apply CLAHE for Contrast Enhancement

Create CLAHE object *clahe* $\leftarrow$ `cv2.createCLAHE`(clipLimit=2.0, tileGridSize=(8, 8))

Apply CLAHE to grayscale image $I_{enhanced} \leftarrow clahe.apply(I_{gray})$

Step 3: Convert Enhanced Image Back to BGR

$I_{enhanced\_bgr} \leftarrow$ `cv2.cvtColor`($I_{enhanced}$, `cv2.COLOR_GRAY2BGR`)

Step 4: Apply Gaussian Blur for Noise Reduction

$I_{blurred} \leftarrow$ `cv2.GaussianBlur`($I_{enhanced\_bgr}$, (5, 5), 0)

Step 5: Apply Median Blur for Additional Noise Reduction

$I_{output} \leftarrow$ `cv2.medianBlur`($I_{blurred}$, 5)

Return I_{output}

3.3.3 Image augmentation

Image augmentation is a technique used to artificially increase the size and variability of a dataset by applying various transformations to the original images. This process is particularly useful in training deep learning models, as it helps improve their generalization capabilities by exposing them to a wider variety of image conditions.

Role of Image Augmentation in Gait Recognition

In the context of gait recognition, image augmentation plays a crucial role by enhancing the robustness and accuracy of the model. Gait patterns can vary significantly due to changes in clothing, carrying conditions, and walking speeds. Augmentation helps in creating a more diverse training set, allowing the model to learn these variations and perform better in real-world scenarios.

Why Use Image Augmentation

Using image augmentation provides several benefits:

1. Increased Dataset Variability: Augmentation generates diverse samples from the original dataset, which helps in reducing overfitting and improving the model's ability to generalize to new data.

2. Improved Model Robustness: By training on augmented data, the model becomes more robust to variations and distortions, leading to better performance on unseen data.

3. Cost-Effective: It is a cost-effective way to enhance the dataset without the need for additional data collection.

Choosing Image Augmentation Techniques

The choice of augmentation techniques depends on the specific requirements of the gait recognition task. Commonly used techniques include rotation, flipping, cropping, scaling, and adding noise. For gait recognition, techniques that simulate real-world variations in walking patterns and conditions are preferred.

Table 3.4: Table of Image Augmentation Techniques

Technique	Purpose	Advantages	Disadvantages
Rotation	Simulates different viewing angles	Increases rotational invariance	May distort the original image
Flipping	Simulates mirrored views	Simple and effective	Not suitable for asymmetric patterns
Scaling	Simulates different distances	Handles variations in size	Can alter aspect ratio
Cropping	Focuses on parts of the image	Enhances robustness to occlusions	Can lose important features
Adding Noise	Simulates sensor noise	Improves robustness to real-world noise	Can degrade image quality

The chosen techniques (rotation, flipping, and adding noise) are specifically selected for their ability to simulate realistic variations in gait patterns. These techniques are simple yet effective, providing a balance between increasing dataset variability and maintaining the integrity of the original gait patterns.

3.3.4 Implementation of Image Augmentation

Mathematical Formulas:

Rotation:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

Where (x, y) are the original coordinates and (x', y') are the new coordinates after rotation by an angle θ.

Flipping: Horizontal flipping can be represented as:

$$x' = -x, \quad y' = y$$

Vertical flipping can be represented as:

$$x' = x, \quad y' = -y$$

Adding Gaussian Noise:

$$I'(x, y) = I(x, y) + N(0, \sigma^2)$$

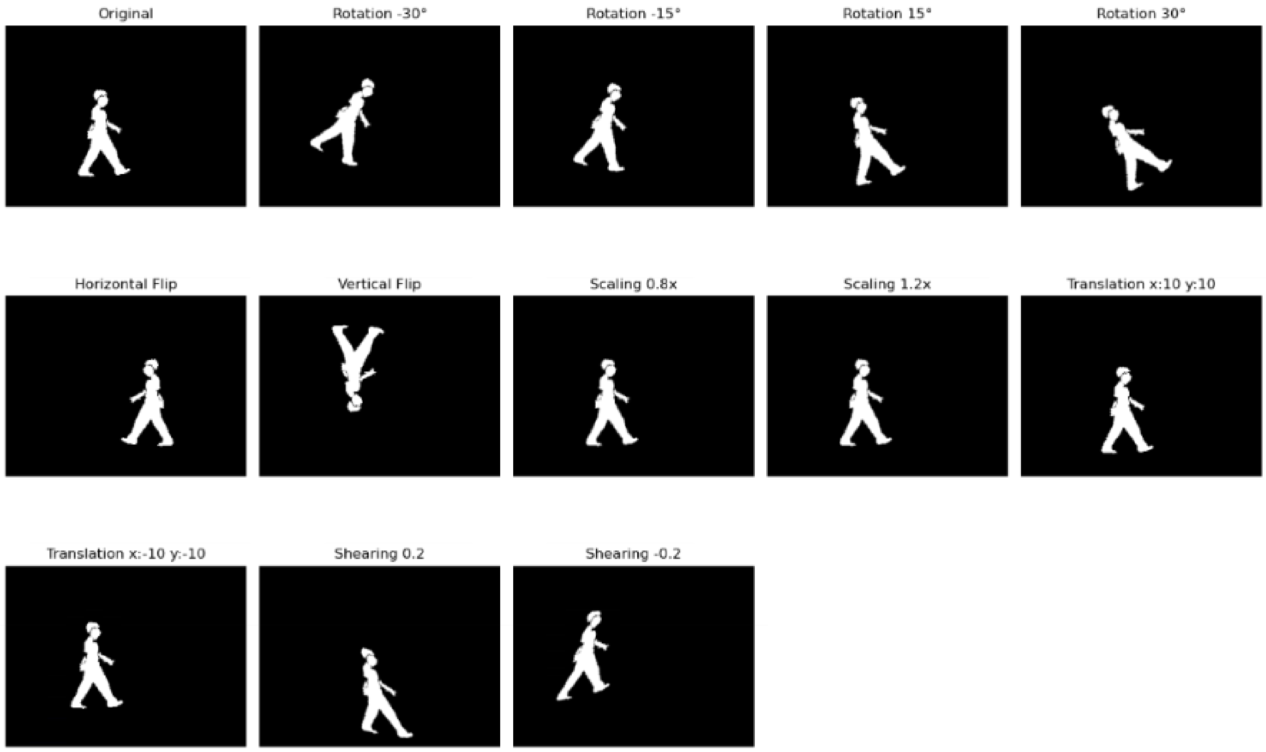

Figure 3.7: Different views of Image Augmentation

Where $I(x, y)$ is the original pixel value, $I'(x, y)$ is the new pixel value, and $N(0, \sigma^2)$ is Gaussian noise with mean 0 and variance σ^2.

3.3.5 Image Augmentation Techniques

1. Rotation

Modification: Rotation involves rotating the image by a specific angle. Common angles include 90°, 180°, and 270°. Impact:

Increased Variability: By rotating images, the dataset includes different perspectives of the same gait pattern, simulating various viewpoints. Robustness: Models trained on rotated images are more robust to changes in the walking direction and camera angle.

2. Flipping

Horizontal Flip: Modification: This involves flipping the image horizontally, creating a mirror image. Impact:

Symmetry: Incorporates the natural symmetry of human gait into the dataset, helping the model learn invariant features. Data Balance: Helps balance the dataset if there is an uneven distribution of left and right walking directions.

Vertical Flip: Modification: This involves flipping the image vertically. Impact:

Increased Variability: Adds variability, although less commonly used as it does not reflect real-world scenarios as much as horizontal flips.

3. Scaling

Modification: Scaling involves zooming in or out of the image, changing its size while maintaining the aspect ratio. Impact:

Perspective Variation: Helps the model learn gait patterns at different scales, simulating varying distances from the camera. Robustness to Size Changes: Ensures that the model can handle variations in the subject's size due to different camera distances or zoom levels.

4. Translation

Modification: Translation shifts the image in the x (horizontal) or y (vertical) direction. Impact:

Positional Variation: Simulates slight changes in subject position relative to the camera. Robustness to Movement: Ensures the model is robust to minor positional changes of the subject in the frame.

5. Shearing

Modification: Shearing involves slanting the image along the x or y axis, creating a skewed effect. Impact: Distortion Handling: Helps the model learn to recognize gait patterns even when the image is slightly distorted. Increased Dataset Variability: Adds a unique form of variability that improves model generalization.

Chapter 4

Module 2:Feature Extraction and Selection

The primary objective of Module 2 in the context of gait recognition is to develop robust and efficient methods for extracting and selecting relevant features from the raw gait data. This module aims to enhance the accuracy and efficiency of gait recognition systems by focusing on two key aspects:

4.0.1 Feature Extraction

This involves identifying and capturing the most significant features from the raw gait data. Feature extraction transforms the original data into a more compact and informative representation, which can be effectively used for classification and recognition tasks. The process includes:

- **Hybrid Feature Extraction:** Utilizing both handcrafted features (such as Histogram of Oriented Gradients (HOG) and Local Binary Patterns (LBP)) and deep learning-based features (such as those extracted using pretrained models like EfficientNet and ResNet). This combination leverages the strengths of both traditional and modern approaches to capture comprehensive gait characteristics.

- **Spatiotemporal Feature Extraction:** Capturing both spatial and temporal aspects of gait data using models like 3D Convolutional Neural Networks (3D CNNs) and Recurrent Neural Networks (RNNs). This ensures that the dynamic nature of gait, which involves sequential movement patterns, is adequately represented.

4.0.2 Feature Selection

Once the features are extracted, it is crucial to select the most relevant ones to reduce dimensionality, remove redundancy, and improve the model's performance.

Effective feature selection enhances computational efficiency and prevents overfitting. The process includes:

- **Dimensionality Reduction Techniques:** Employing methods like Principal Component Analysis (PCA) to reduce the number of features while preserving the essential information. PCA helps in transforming the feature space into a lower-dimensional space, making it easier to manage and analyze.

- **Feature Fusion:** Combining features from different sources or modalities to create a unified and more informative feature set. This involves techniques like serial-based feature fusion, which integrates multiple feature types to capture a holistic view of the gait patterns.

Implementation

- **Hybrid Feature Extraction:** Utilizing both handcrafted and deep learning features ensures a comprehensive representation of gait data.

- **Dimensionality Reduction and Feature Fusion:** Applying PCA and serial-based feature fusion to enhance feature quality and reduce dimensionality.

- **Spatiotemporal Integration:** Combining spatial and temporal features to capture the dynamic nature of gait sequences.

Importance

Enhanced Recognition Accuracy: By extracting and selecting the most relevant features, the module aims to improve the overall accuracy of the gait recognition system. Improved Computational Efficiency: Reducing the dimensionality of the feature set decreases computational requirements and speeds up the recognition process. Robustness to Variations: Combining different feature types and employing effective selection techniques ensures that the model is robust to variations in gait patterns due to factors like clothing, carrying conditions, and walking speed.

4.1 Introduction to Hybrid Feature Extraction in Gait Recognition

Hybrid feature extraction is a powerful technique in gait recognition that leverages both traditional handcrafted features and modern deep learning-based features to create a comprehensive representation of gait patterns. This approach combines the strengths of different methodologies to enhance the accuracy and robustness of gait recognition systems.

4.1.1 Role and Importance of Hybrid Feature Extraction in Gait Recognition

- **Enhanced Feature Representation:** By combining handcrafted and deep learning features, hybrid feature extraction captures both global and local characteristics of gait. This comprehensive feature set improves the model's ability to distinguish between different gait patterns.

- **Robustness to Variations:** Hybrid feature extraction helps in creating features that are robust to variations in walking conditions, such as changes in clothing, carrying conditions, and environmental factors.

- **Improved Accuracy:** The integration of diverse features leads to better model performance, resulting in higher accuracy rates in recognizing individuals based on their gait.

- **Versatility:** This method can be applied across various datasets and conditions, making it a versatile approach for real-world applications.

4.1.2 Why Use Hybrid Feature Extraction

- **Complementary Strengths:** Handcrafted features like HOG and LBP are known for their robustness to changes in lighting and pose, while deep learning features extracted from models like EfficientNet, KANs[Kolmorgov Arnold Networks] and ResNet capture high-level abstract representations. Combining these features leverages their complementary strengths.

- **Improved Generalization:** The hybrid approach improves the generalization capability of the model, making it perform well on unseen data.

- **Comprehensive Feature Set:** It provides a more complete representation of gait, capturing both detailed structural information and high-level semantic features.

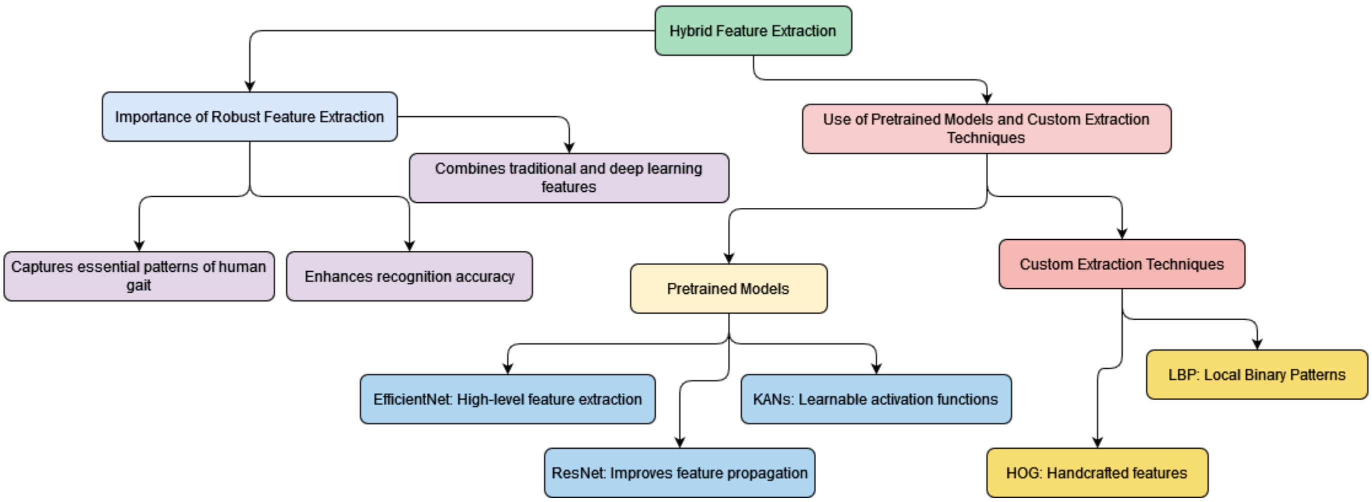

Figure 4.1: Flowchart of Hybrid Feature Extraction

The above diagram illustrates the hybrid feature extraction process, highlighting the importance of robust feature extraction, the combination of traditional and deep learning features, and the specific techniques used (HOG, LBP, EfficientNet, ResNet, KANs). This visual representation underscores the comprehensive approach taken to enhance gait recognition performance by integrating multiple feature extraction methodologies.

4.1.3 Choosing Techniques: Handcrafted and Deep Learning-Based Features

Handcrafted Features

Histogram of Oriented Gradients (HOG): Description: HOG captures the gradient orientation histograms of local regions in an image, which represent the edge directions.

Importance: It is robust to changes in illumination and pose, making it effective for capturing the silhouette structure of gait.

Local Binary Patterns (LBP): Description: LBP encodes the local texture information by comparing each pixel with its neighbors.

Importance: It is computationally efficient and robust to monotonic gray-scale changes, providing valuable texture information.

4.1.4 Deep Learning-Based Features

- **EfficientNet**
  Description: EfficientNet is a family of convolutional neural networks that scale up efficiently in terms of model depth, width, and resolution.
  Importance: It balances accuracy and computational efficiency, making it suitable for extracting high-level features from gait images.

- **ResNet**
  Description: ResNet (Residual Networks) uses residual connections to enable the training of very deep networks.
  Importance: It captures complex patterns and high-level abstractions, which are essential for distinguishing subtle differences in gait.

- **Kolorgov-Arnold Networks[KANs]**
  Description:KANs leverage learnable activation functions to adapt to different data distributions dynamically.
  Importance: This adaptability makes them suitable for capturing the complex and varied features present in gait patterns.

4.1.5 Implementation of Hybrid Feature Extraction

HOG Feature Extraction Process

Gradient Calculation

The first step is to compute the gradient values in the x and y directions for each pixel in the image. This can be done using derivative masks.

$$G_x = I * D_x,$$
$$G_y = I * D_y$$

where I is the image, and D_x and D_y are derivative masks in the x and y directions, respectively.

Orientation Binning

The image is divided into small connected regions called cells. For each cell, a histogram of gradient directions (or orientations) is computed. Each pixel within the cell contributes a weighted gradient to an orientation histogram.

Block Normalization

To improve the robustness to illumination and shadowing, the local histograms are normalized. The normalized group of histograms represents the block, and the HOG descriptor is obtained by concatenating these histograms.

Feature Vector Construction

The final HOG feature vector is constructed by concatenating the normalized histograms of all blocks.

Role in Gait Recognition

HOG captures the gradient structure of the silhouettes, which is crucial for distinguishing different gait patterns. It emphasizes the edges and transitions in the image, making it effective for silhouette-based gait recognition.

Mathematical Formula:

HOG:

$$H = \sum_{i,j} f(x, y, \theta) \quad \text{where} \quad f(x, y, \theta) = \begin{cases} \|\nabla I(x, y)\| & \text{if} \quad \theta(x, y) = \theta \\ 0 & \text{otherwise} \end{cases}$$

Here, $\|\nabla I(x, y)\|$ is the gradient magnitude and $\theta(x, y)$ is the gradient direction.

LBP Feature Extraction Process

Thresholding

For each pixel in the image, compare the pixel value to its neighboring pixels. If the neighbor's value is greater or equal, it is assigned a 1; otherwise, it is assigned a 0.

Binary Pattern Calculation

The binary values are then combined to form a binary number (usually a 3x3 neighborhood, resulting in an 8-bit number).

Histogram Formation

The image is divided into regions (e.g., cells), and the LBP value of each pixel is computed. The histogram of these values for each region is calculated.

Feature Vector Construction

The histograms of all regions are concatenated to form the final LBP feature vector.

4.1.6 Role in Gait Recognition

LBP is robust to monotonic gray-scale transformations and captures texture information effectively. This is useful in gait recognition as it encodes the texture of the silhouette, which can vary with different clothing or backgrounds.

Mathematical Formula:

LBP:

$$\mathrm{LBP}(x_c, y_c) = \sum_{p=0}^{P-1} s(g_p - g_c) \cdot 2^p$$

where

$$s(x) = \begin{cases} 1 & \text{if} \quad x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Here, g_c is the gray value of the center pixel, and g_p are the gray values of the P surrounding pixels.

4.1.7 Comparison Tables

Handcrafted Features Techniques

Technique	Description	Advantages	Disadvantages
HOG	Captures edge directions using gradient orientation histograms	Robust to illumination and pose changes, captures silhouette structure	Computationally intensive, sensitive to noise
LBP	Encodes local texture by comparing pixel intensities	Efficient, robust to monotonic gray-scale changes	Less effective with significant texture variations
Gabor Filters	Captures texture information using frequency and orientation	Effective in texture representation	High computational cost, sensitive to noise
SIFT	Detects and describes local features in images	Highly distinctive, invariant to scaling and rotation	Computationally expensive, high memory usage
SURF	Similar to SIFT but faster	Fast, good at object recognition	Less effective for fine-grained details in gait

Table 4.1: Comparison of Handcrafted Features Techniques

Deep Learning-Based Features Techniques

Technique	Description	Advantages	Disadvantages
EfficientNet	Scales up model depth, width, and resolution efficiently	Balanced accuracy and computational efficiency	Requires significant computational resources
ResNet	Uses residual connections to train very deep networks	Captures complex patterns, high-level abstractions	Computationally intensive, large model size
DenseNet	Connects each layer to every other layer to ensure maximum information flow	Reduces vanishing gradient, improves feature reuse	Large computational and memory requirements
KAN (Kolmogorov-Arnold Network)	Uses complex mathematical transformations for feature extraction	Captures non-linear relationships and complex patterns	Computationally intensive, complex architecture

Table 4.2: Comparison of Deep Learning-Based Features Techniques

Conclusion

The combination of handcrafted features (HOG and LBP) and deep learning-based features (EfficientNet, ResNet, DenseNet, and KAN) in hybrid feature extraction leverages the strengths of both traditional and modern approaches. This comprehensive feature set ensures that both detailed structural information and high-level semantic features are captured, providing a robust and accurate representation of gait patterns. This hybrid approach is crucial for achieving high performance in real-world gait recognition applications.

4.2 EfficientNet, ResNet50, and Kolmogorov-Arnold Network (KAN) Architectures

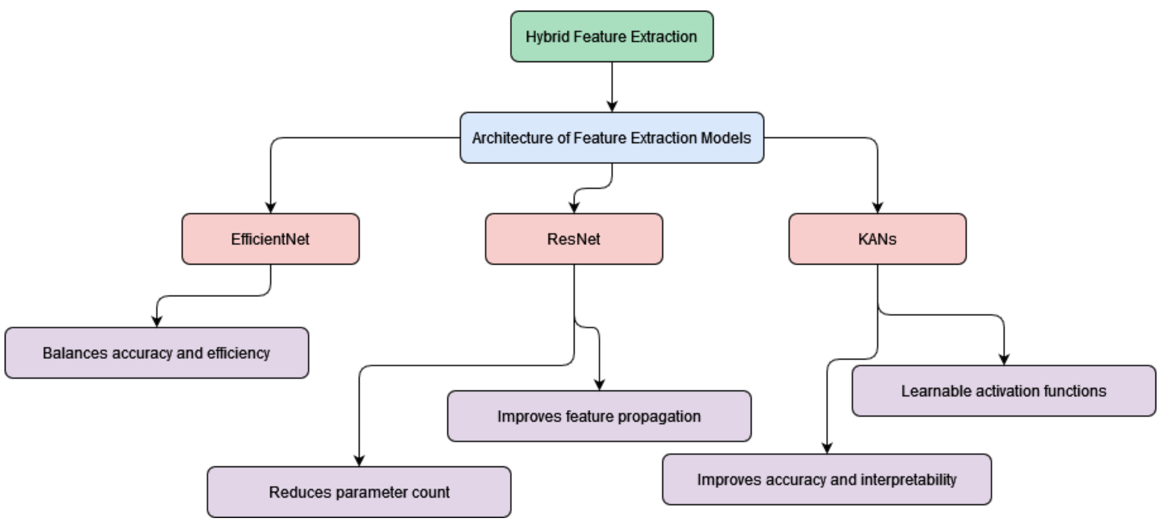

Figure 4.2: Overall Architecture

4.2.1 EfficientNet Architecture

EfficientNet models are a family of convolutional neural networks that are designed to achieve high accuracy while being computationally efficient. They use a compound scaling method to scale up the network width, depth, and resolution in a balanced manner.

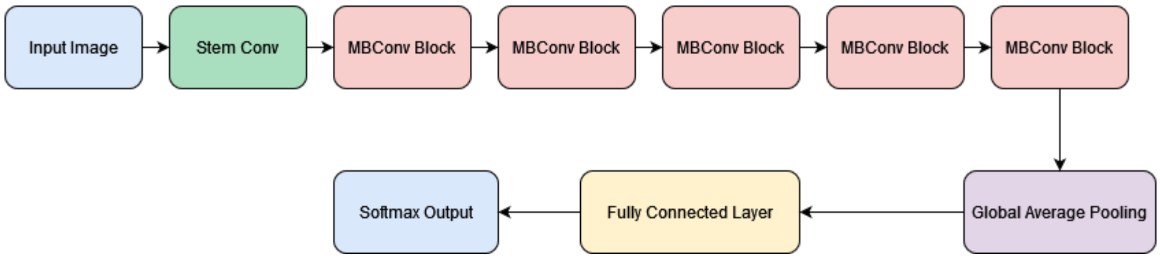

Figure 4.3: The EfficientNet architecture employs MBConv blocks to balance between accuracy and computational efficiency by scaling network dimensions using a compound scaling method

The key features of EfficientNet include:

- **Input Image:** The network starts with the input image that is passed through the initial stem convolution layer.
- **Stem Conv:** This is a standard convolution layer that preprocesses the input image before passing it to the main network blocks.
- **MBConv Blocks:** These are Mobile Inverted Bottleneck Convolutional blocks, which are the core components of EfficientNet. They are designed to be efficient in terms of both memory and computation. EfficientNet scales these blocks in three dimensions: width, depth, and resolution.
- **Global Average Pooling:** This layer reduces the dimensions of the feature maps from the MBConv blocks to a fixed size, which helps in reducing the computational load and overfitting.
- **Fully Connected Layer:** The features from the global average pooling are passed to fully connected layers for the final classification.
- **Softmax Output:** This layer provides the probability distribution over the target classes.

4.2.2 ResNet Architecture

ResNet50 is a 50-layer deep convolutional neural network with skip connections (or shortcuts) that allow gradients to flow more easily through the network during backpropagation.

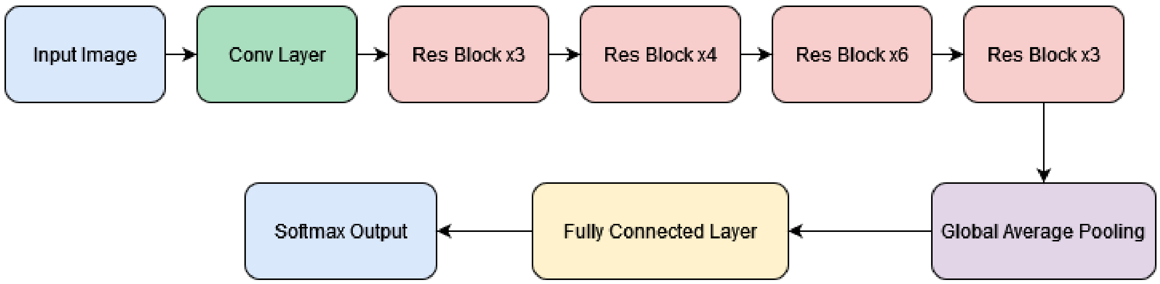

Figure 4.4: ResNet Architecure

The critical components of ResNet50 include:

- **Input Image:** The input image is fed into the network through the initial convolution layer.
- **Conv Layer:** This initial convolution layer processes the input image before passing it to the residual blocks.
- **Res Blocks:** These blocks are the building units of ResNet. They contain a series of convolutional layers with skip connections (or shortcuts). The skip connections help in maintaining the gradient flow during backpropagation, which allows for the training of deeper networks.
- **Global Average Pooling:** Similar to EfficientNet, this layer reduces the feature maps to a fixed size.
- **Fully Connected Layer:** The features are then passed to fully connected layers for the final classification.
- **Softmax Output:** This layer outputs the probability distribution over the target classes.

4.2.3 Kolmorgov Arnold Networks Architecture

Kolmogorov-Arnold Network (KAN) is a neural network inspired by the Kolmogorov-Arnold representation theorem. It is designed to approximate complex multivariate functions.

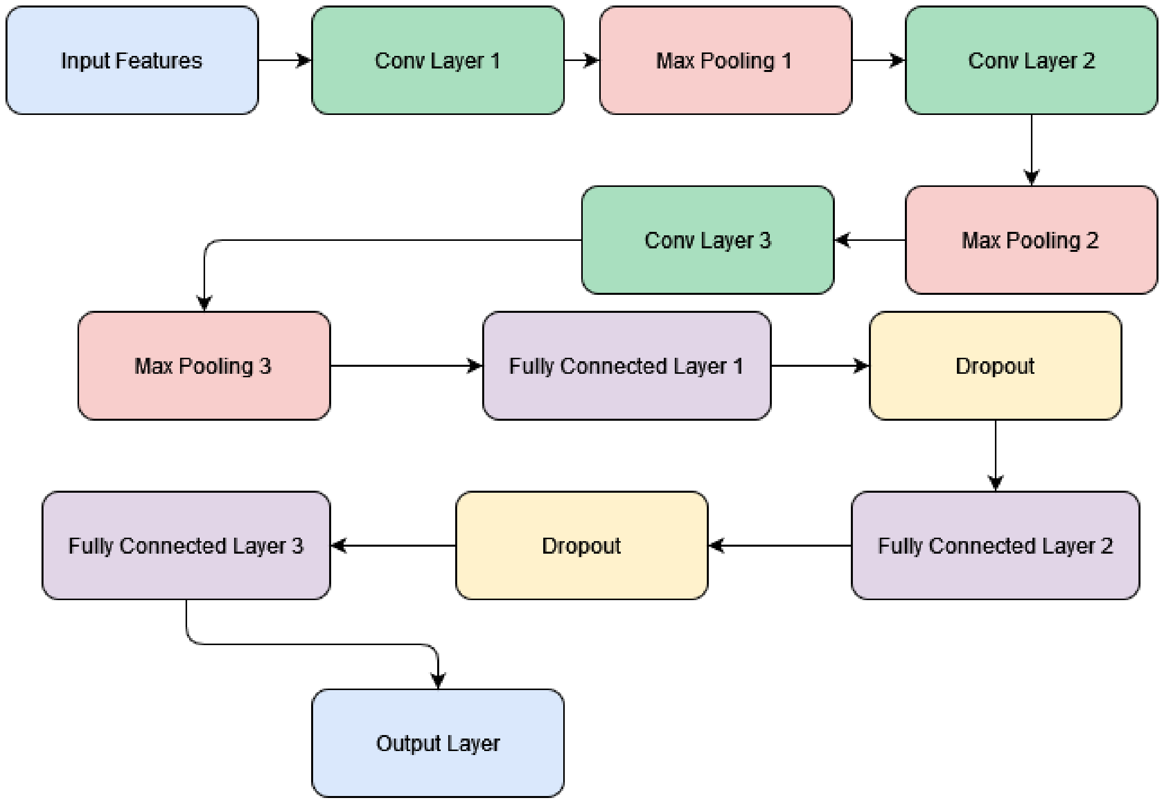

Figure 4.5: KANs Architecture

Explanation of the Architecture Diagram

Input Features

Description: This is the input layer where the feature vectors (extracted from images using techniques like EfficientNet and ResNet) are fed into the network.
Role: It takes the combined feature vector obtained from different pre-trained models.

Conv Layer 1, Conv Layer 2, Conv Layer 3

Description: These are convolutional layers that apply convolution operations to the input features.
Role: Convolutional layers help in extracting spatial features by applying filters to the input data, which is particularly useful in capturing patterns within the gait images.

Max Pooling 1, Max Pooling 2, Max Pooling 3

Description: These layers perform max pooling operations, which downsample the input representations by taking the maximum value over a defined window.
Role: Max pooling layers help in reducing the spatial dimensions of the feature maps, which reduces the computational complexity and helps in extracting dominant features.

Fully Connected Layer 1, Fully Connected Layer 2, Fully Connected Layer 3

Description: These are dense layers where each neuron is connected to every neuron in the previous layer.
Role: Fully connected layers perform high-level reasoning and combination of features extracted by the convolutional layers.

Dropout

Description: Dropout layers randomly set a fraction of input units to 0 at each update during training time, which helps prevent overfitting.
Role: It helps in regularizing the model by preventing overfitting and improving generalization.

Output Layer

Description: The final layer in the network, which provides the final predictions.
Role: It outputs the probabilities of each class (in this case, different gait patterns) based on the features processed through the network.

4.2.4 How These Architectures will be used

In this research, EfficientNet and ResNet50 are used as feature extractors in a feature fusion approach, which combines features from multiple networks to improve the model's performance. The Kolmogorov-Arnold Network (KAN) is implemented as a custom network architecture for further processing the combined features.

Here's a brief explanation of how these networks are integrated:

EfficientNet and ResNet50 Feature Extraction

- EfficientNet and ResNet50 are pre-trained models that are used to extract features from the input images.
- The features extracted from both networks are concatenated to form a combined feature vector.

Kolmogorov-Arnold Network (KAN)

- The combined feature vector is fed into the Kolmogorov-Arnold Network (KAN).
- KAN consists of several fully connected layers that process the combined features and output the final prediction.

By using multiple feature extraction networks and combining their outputs, the model can leverage the strengths of each network to achieve better performance. The final predictions are made based on the processed combined features, which are more robust and informative.

Chapter 5

Module 3: Model Training and Evaluation

5.1 Introduction to Model Training and Evaluation

Model training and evaluation are critical phases in the development of gait recognition systems. These processes ensure that the models are capable of accurately identifying individuals based on their gait patterns. Training involves teaching the model to recognize and classify gait patterns using labeled data, while evaluation assesses the model's performance on unseen data to ensure its generalizability and robustness.

5.2 Role of Model Training and Evaluation in Gait Recognition

5.2.1 Importance of Model Training

- **Learning Patterns:** During training, the model learns the underlying patterns and features of gait data. This learning process is essential for the model to make accurate predictions.

- **Parameter Optimization:** Training helps in optimizing model parameters such as weights and biases, which are crucial for minimizing prediction errors.

- **Handling Variability:** Training on diverse datasets allows the model to handle variability in gait patterns due to different conditions, such as changes in clothing or walking speed.

5.2.2 Importance of Model Evaluation

- **Assessing Performance:** Evaluation metrics such as accuracy, precision, recall, F1 score, and AUC (Area Under the Curve) provide insights into how well the model performs on unseen data.

- **Ensuring Generalizability:** Evaluation ensures that the model is not overfitting to the training data and can generalize well to new, unseen examples.

- **Identifying Weaknesses:** Through evaluation, weaknesses in the model can be identified, allowing for further improvements and refinements.

5.2.3 Why Use Model Training and Evaluation?

- **Accuracy:** Proper training and evaluation ensure high accuracy in recognizing individuals based on their gait.

- **Robustness:** They help in creating robust models that can perform well under various conditions and across different datasets.

- **Efficiency:** Evaluation helps in identifying the most efficient models that balance performance with computational requirements.

5.2.4 Choosing Techniques for Model Training and Evaluation

When selecting techniques for model training and evaluation, it is essential to consider the nature of the data, the specific requirements of the gait recognition task, and the computational resources available.

- **Spatiotemporal Feature Integration:** Combines spatial and temporal dynamics, capturing both the static structure of the gait and its movement over time.

- **Custom Spatiotemporal Models:** Designed to handle the unique aspects of gait data, providing better feature extraction and classification capabilities.

- **Transfer Learning:** Utilizes pre-trained models like EfficientNet and ResNet, which have been trained on large datasets and can extract high-level features effectively.

Table of Model Training and Evaluation Techniques

Technique	**Description**	**Advantages**	**Our Approach**
Supervised Learning	Training models using labeled data to predict outcomes.	High accuracy with labeled data, straightforward implementation.	Used for initial training with labeled gait data, allowing models to learn specific gait patterns.
Spatiotemporal Models	Models that integrate both spatial and temporal features.	Captures both the appearance and motion dynamics of gait.	Custom spatiotemporal models designed for better feature integration and gait recognition accuracy.
Transfer Learning	Utilizing pre-trained models to extract features from new data.	Reduces training time, leverages existing knowledge, improves feature extraction.	EfficientNet and ResNet are used for high-level feature extraction, improving recognition performance.
Cross-Validation	Evaluating models by splitting the data into training and validation sets multiple times.	Provides a more reliable measure of model performance, helps in selecting the best model.	Employed to ensure the robustness and generalizability of the models across different subsets of data.
Ensemble Learning	Combining multiple models to improve overall performance.	Leverages the strengths of each model, reduces the risk of overfitting.	Implemented using Voting Classifier, combining EfficientNet, ResNet, and custom models to enhance recognition accuracy.
Hyperparameter Tuning	Optimizing model parameters to improve performance.	Enhances model accuracy and efficiency by finding the best parameter settings.	Random Search used for tuning parameters of custom spatiotemporal models, ensuring optimal performance.
Ablation Studies	Systematically removing or altering parts of the model to understand their impact.	Helps in identifying the most critical components of the model.	Conducted to determine the contribution of each component (e.g., HOG, LBP, deep features) to the overall performance.
Evaluation Metrics	Using metrics like accuracy, precision, recall, F1 score, and AUC to assess model performance.	Provides a comprehensive understanding of model performance, highlights strengths and weaknesses.	Used to evaluate and compare different models, ensuring the selected model meets the desired performance criteria.

Table 5.1: Comparison of Model Training and Evaluation Techniques

5.3 Why Our Model Training and Evaluation is Better for Gait Recognition

- **Custom Spatiotemporal Models:** Designed specifically for gait recognition, these models integrate both spatial and temporal features, capturing the full dynamics of human gait.

- **Transfer Learning:** Utilizing pre-trained models like EfficientNet and ResNet, we leverage their advanced feature extraction capabilities, resulting in improved performance.

- **Ensemble Learning:** By combining multiple models, our approach reduces the risk of overfitting and leverages the strengths of each model for better accuracy and robustness.

- **Comprehensive Evaluation:** Using a variety of metrics and cross-validation techniques, we ensure that our models are not only accurate but also robust and generalizable.

- **Data Augmentation and Regularization:** These techniques enhance the training process, allowing our models to learn from a more diverse set of examples and preventing overfitting.

- **Usage of Kolmogorov-Arnold Networks (KANs):** KANs are utilized in our methodology to further enhance the feature extraction process. These networks provide learnable activation functions, which improve the adaptability and accuracy of the model. By incorporating KANs, we can capture more complex patterns in the gait sequences, contributing to the overall robustness and effectiveness of our gait recognition system.

My approach to model training and evaluation for gait recognition integrates advanced techniques like custom spatiotemporal models, transfer learning, ensemble learning, and the innovative use of Kolmogorov-Arnold Networks. These methods ensure high accuracy, robustness, and generalizability, making our gait recognition system highly effective in real-world applications. By leveraging the strengths of both traditional and modern techniques, we provide a comprehensive solution for robust and reliable gait recognition.

5.4 Spatiotemporal Feature Integration

5.4.1 Introduction to Spatiotemporal Integration

Spatiotemporal integration is a crucial technique in gait recognition, aimed at capturing both spatial and temporal dynamics of human movement. Gait recognition involves analyzing the way people walk, which inherently includes both the spatial structure of the body and its temporal motion patterns. The integration of these two aspects leads to a more comprehensive and accurate representation of gait, enhancing the overall recognition performance.

5.4.2 Role of Spatiotemporal Integration in Gait Recognition

Capturing Full Dynamics

- **Spatial Dynamics:** Involves the structural details of a person's body captured in each frame.
- **Temporal Dynamics:** Involves the motion patterns and how these structural details change over time.
- **Integration:** Combining these dynamics allows the model to understand not just the static appearance but also the movement characteristics, crucial for distinguishing between different individuals.

Improved Accuracy

- **Holistic Understanding:** By integrating both spatial and temporal features, the model gains a holistic understanding of the gait cycle, leading to higher accuracy in recognition tasks.
- **Robustness:** Spatiotemporal integration helps in making the recognition system robust to variations in walking speed, direction, and environmental conditions.

Enhanced Generalization

- **Temporal Dependencies:** Understanding temporal dependencies ensures that the model can generalize better across different scenarios and subjects.
- **Adaptive Learning:** Spatiotemporal models can adapt to new gait patterns more effectively compared to spatial-only models.

5.4.3 Why Use Spatiotemporal Integration?

- **Rich Feature Representation:** Combines detailed spatial features with dynamic temporal features, providing a richer representation of gait.
- **Complex Motion Analysis:** Capable of analyzing complex motion patterns that are not evident in static images.
- **Performance Improvement:** Leads to significant improvements in recognition accuracy and robustness.

5.5 Integration of 3D CNNs and RNN/LSTM for Extracting and Combining Spatial and Temporal Features

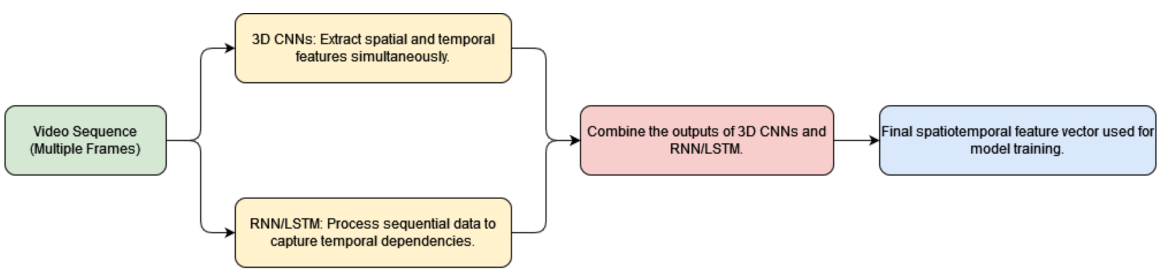

Figure 5.1: Spatiotemporal Feature Integration in Gait Recognition

Introduction

The integration of 3D Convolutional Neural Networks (3D CNNs) and Recurrent Neural Networks (RNNs) or Long Short-Term Memory networks (LSTMs) represents a sophisticated approach to capturing both spatial and temporal dynamics in gait recognition. This methodology leverages the strengths of both network types to create a comprehensive feature extraction and integration pipeline.

5.6 Techniques Used

5.6.1 3D CNNs

- **Function:** 3D CNNs are designed to extract spatial and temporal features simultaneously from video sequences. They process video frames as a volumetric input, allowing them to capture motion patterns along with spatial details.

- **Operation:** These networks apply 3D convolutional filters to the input video frames, capturing the spatial configuration in each frame and the temporal transitions between frames.

- **Advantages:** They provide a unified way to handle both types of information, reducing the need for separate spatial and temporal processing steps.

5.6.2 RNN/LSTM

- **Function:** RNNs and LSTMs are specialized for sequential data and are particularly effective at capturing temporal dependencies.

- **Operation:** These networks process the sequence of features extracted by the 3D CNNs, learning to understand the temporal progression and contextual relationships between frames.

- **Advantages:** They excel at maintaining and leveraging the order of frames, which is essential for accurately modeling gait sequences.

5.6.3 Integration Process

- **Feature Extraction with 3D CNNs:** The video sequence is input into a 3D CNN, which extracts both spatial and temporal features across the sequence.

- **Temporal Processing with RNN/LSTM:** The extracted features are then fed into an RNN or LSTM, which processes the sequence to capture temporal dependencies and refine the understanding of motion dynamics.

- **Feature Combination:** The outputs from both the 3D CNN and the RNN/LSTM are combined to form a comprehensive feature vector.

- **Final Spatiotemporal Feature Vector:** This combined feature vector is used as the final representation of the input video sequence, which is then fed into subsequent layers for classification or further processing.

5.6.4 Choosing Techniques for Spatiotemporal Integration

- **Custom Spatiotemporal Models:** Specifically designed to handle the unique aspects of gait recognition, focusing on both spatial structure and temporal motion.

- **Pre-trained Models and Fine-Tuning:** Utilizing models like 3D Convolutional Neural Networks (3D CNNs) and Long Short-Term Memory (LSTM) networks that are fine-tuned for gait data.

- **Combining Networks:** Combining different network architectures (e.g., 3D CNNs for spatial features and LSTMs for temporal features) to leverage their individual strengths.

5.7 Why Our Techniques Are Better

- **Customized for Gait:** Our models are specifically designed for gait recognition, ensuring they capture the nuances of human walking patterns.

- **Integration of Advanced Architectures:** By combining state-of-the-art techniques like 3D CNNs and LSTMs, our approach ensures comprehensive feature extraction.

- **Focus on Robustness and Generalization:** Our spatiotemporal integration methods are tailored to be robust against variations in gait and environmental conditions, ensuring high generalizability across different scenarios.

Conclusion Spatiotemporal integration is fundamental for effective gait recognition, providing a comprehensive understanding of human movement by combining spatial and temporal features. Our chosen techniques, which include advanced architectures like 3D CNNs and LSTMs, offer superior performance by leveraging the

Technique	Description	Advantages	Why Better for Gait Recognition
3D CNNs	Extracts spatial and temporal features simultaneously from video sequences.	Captures both spatial structure and temporal dynamics.	Provides a comprehensive understanding of gait cycles.
LSTMs	Processes sequential data to capture temporal dependencies.	Excels at modeling long-term dependencies in sequential data.	Ensures smooth and continuous motion representation.
GRUs	Similar to LSTMs but with a simplified architecture for faster training.	Faster training and reduced computational requirements.	Effective for capturing temporal dependencies with less computational overhead.
Hybrid Models	Combines 3D CNNs and LSTMs to leverage both spatial and temporal feature extraction capabilities.	Utilizes the strengths of both 3D CNNs and LSTMs.	Provides robust feature extraction and integration for accurate gait recognition.
Transformers	Uses self-attention mechanisms to capture dependencies over long sequences.	Handles long-range dependencies more effectively than RNNs.	Provides detailed temporal integration, crucial for complex gait patterns.
Custom Spatiotemporal	Tailored models specifically designed for the intricacies of gait recognition.	Optimized for gait-specific features and dynamics.	Ensures maximum relevance and performance for gait recognition tasks.

Table 5.2: Comparison of Spatiotemporal Integration Techniques

strengths of both spatial and temporal analysis. This leads to a more accurate, robust, and generalizable gait recognition system.

5.8 Transfer Learning and Ensemble Techniques

5.8.1 Introduction

Transfer learning leverages pre-trained models, which are neural networks that have already been trained on large datasets, to extract features for a new, often smaller dataset. This approach significantly reduces the training time and enhances performance by utilizing the rich feature representations learned from the large datasets.

5.8.2 How It Works

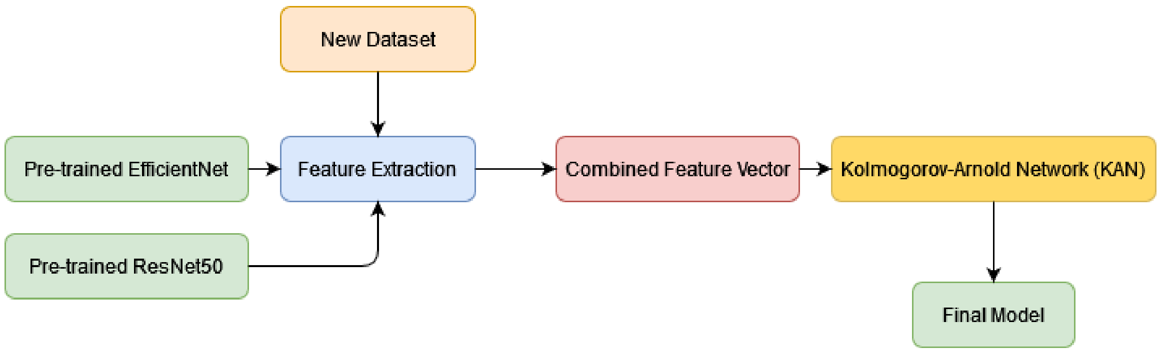

Figure 5.2: Transfer Learning Structure

- **Pre-trained Models:** These models are initially trained on extensive datasets, such as ImageNet, which contains millions of images across thousands of categories. During this training, the model learns to extract general features like edges, textures, and shapes.

- **Feature Extraction:** For a new task, such as gait recognition, the pre-trained model is used to extract these learned features from the new dataset. This process involves feeding the new dataset into the pre-trained model and using the output from one or more layers as the feature representation of the input data.

- **Fine-tuning:** In some cases, the pre-trained model is fine-tuned on the new dataset. This involves further training the model with a small learning rate, allowing the model to adjust its weights slightly to better fit the new data while retaining the useful features learned from the original training.

- **Benefits:** This approach reduces the amount of data needed for training, cuts down on the computational resources required, and generally results in better performance, especially when the new dataset is small or the new task is similar to the original one.

5.9 Introduction to Ensemble Techniques

Ensemble learning is a powerful machine learning paradigm where multiple models are combined to improve overall performance. The core idea behind ensemble methods is to leverage the strengths of different models to create a robust and accurate predictive system. By integrating various models, ensemble techniques can reduce the risk of overfitting and improve generalizability, making them highly effective for complex tasks such as gait recognition.

5.9.1 Role of Ensemble Techniques in Gait Recognition

In the context of gait recognition, ensemble techniques play a crucial role in enhancing the accuracy and robustness of the recognition system. Gait recognition involves analyzing the walking patterns of individuals, which can be influenced by various factors such as clothing, carrying conditions, and changes in walking speed. Ensemble methods help mitigate the impact of these variations by combining the predictions from multiple models, each capturing different aspects of the gait patterns.

5.9.2 Importance of Ensemble Techniques

- **Improved Accuracy:** By combining the strengths of various models, ensemble methods can achieve higher accuracy compared to individual models.

- **Robustness:** Ensemble techniques are more resilient to noise and variations in the data, providing more stable and reliable predictions.

- **Generalizability:** The use of multiple models reduces the risk of overfitting, ensuring that the ensemble system performs well on unseen data.

5.9.3 Choosing Ensemble Techniques

The choice of ensemble techniques depends on the specific requirements of the gait recognition task. Common ensemble methods include bagging, boosting, and stacking, each offering unique advantages.

- **Bagging (Bootstrap Aggregating):** Bagging involves training multiple instances of the same model on different subsets of the data and combining their predictions. This technique reduces variance and helps improve the stability of the model.

- **Boosting:** Boosting sequentially trains models, where each new model focuses on correcting the errors of the previous ones. This method improves the model's accuracy by reducing bias.

- **Stacking:** Stacking involves training multiple models and using their outputs as inputs to a meta-model, which makes the final prediction. This technique leverages the strengths of different models and combines them optimally.

5.9.4 Ensemble Techniques Used in my Thesis

In my thesis, I have implemented ensemble learning using pre-trained models like EfficientNet and ResNet, combined with a custom Kolmogorov-Arnold Network (KAN). Here's how we integrate these techniques:

5.9.5 EfficientNet and ResNet Feature Extraction

- These pre-trained models are used to extract high-level features from the input gait images.

- The extracted features are then combined to form a comprehensive feature vector.

5.9.6 Kolmogorov-Arnold Network (KAN)

- The combined feature vector is fed into the KAN, which processes the features through several fully connected layers to output the final prediction.

5.9.7 Ensemble Learning Framework

- We use a Voting Classifier to combine the predictions from EfficientNet, ResNet, and KAN.

- The Voting Classifier aggregates the predictions using a majority voting scheme for the final output.

5.10 Evaluation

5.10.1 Evaluation Metrics Used

In my thesis, I have employed several key evaluation metrics to assess the performance of my gait recognition models comprehensively. These metrics are crucial for understanding the strengths and weaknesses of the models and ensuring their reliability and robustness. The primary evaluation metrics used include:

- **Accuracy:** The ratio of correctly predicted instances to the total instances. It measures the overall correctness of the model's predictions.

- **Precision:** The ratio of correctly predicted positive observations to the total predicted positives. It indicates the accuracy of the positive predictions made by the model.

- **Recall:** The ratio of correctly predicted positive observations to all the actual positives. It assesses the model's ability to find all relevant cases within a dataset.

- **F1 Score:** The weighted average of Precision and Recall. This score provides a balance between precision and recall, making it a crucial metric when there is an uneven class distribution.

- **AUC (Area Under the Curve):** Measures the ability of the model to distinguish between classes. A higher AUC value indicates a better performance of the model in distinguishing between the positive class and the negative class.

Performance Results

To ensure a comprehensive evaluation, I have assessed the model's performance based on the following aspects:

- **Accuracy:** This metric provides the percentage of correctly classified instances, giving an overall sense of how well the model performs.

- **Precision:** It reflects the accuracy of positive predictions, which is particularly important in scenarios where false positives need to be minimized.

- **Recall:** This metric evaluates the model's ability to find all relevant cases, which is crucial in contexts where missing a positive case can have significant consequences.

- **F1 Score:** By balancing precision and recall, the F1 score offers a more nuanced view of the model's performance, especially in datasets with imbalanced classes.

- **AUC:** The AUC metric provides a comprehensive view of the model's performance across different threshold values, making it an important metric for evaluating the overall effectiveness of the model.

5.11 Why My Evaluation Metric is Better

In my thesis, I have implemented a comprehensive set of evaluation metrics that surpass the commonly used metrics in the literature review papers. Here's why my evaluation approach is more effective:

5.11.1 Comprehensive Assessment

My evaluation includes Accuracy, Precision, Recall, F1 Score, and AUC. By incorporating a wide range of metrics, I ensure a thorough assessment of the model's performance from multiple perspectives, providing a more holistic view than what is typically found in literature. For instance:

- **Accuracy:** While accuracy provides an overall measure of correctness, it alone is insufficient for evaluating performance, especially with imbalanced datasets.

- **Precision:** Precision is critical for understanding the model's ability to make accurate positive predictions, which is not always emphasized in other papers.

- **Recall:** My focus on recall ensures that the model's ability to identify all relevant cases is thoroughly evaluated, which is crucial for tasks where missing a positive instance can have significant consequences.

- **F1 Score:** By balancing precision and recall, the F1 score offers a more nuanced and balanced view of the model's performance, addressing the limitations of using accuracy alone. Many existing approaches have used the F1-score as a performance metric, particularly when dealing with imbalanced datasets. The F1-score is advantageous in such scenarios as it balances precision and recall, providing a more comprehensive evaluation of the model's performance when class distributions are uneven.

- **AUC:** The inclusion of AUC provides a comprehensive view of the model's discriminative ability across different threshold values, which is often overlooked in other evaluations.

5.11.2 Robustness and Generalization

My evaluation metrics are designed to test the robustness and generalizability of the models. By using a combination of these metrics, I can ensure that the models perform well not only on the training data but also on unseen data, reducing the risk of overfitting. This is a significant improvement over the simpler evaluation methods used in the literature review papers, which often focus narrowly on accuracy.

Balancing Class Distribution

In many gait recognition tasks, class distribution can be imbalanced. My use of the F1 Score and AUC addresses this issue effectively. The F1 Score balances the precision and recall, making it more reliable for imbalanced datasets, while the AUC measures the performance across all classification thresholds, providing a more detailed analysis of the model's performance.

Enhanced Insight and Performance

By integrating multiple evaluation metrics, my approach provides deeper insights into the strengths and weaknesses of the models. This allows for targeted improvements and refinements, leading to overall better performance. This detailed evaluation strategy is often missing in the literature review papers, where the focus tends to be on simpler, less comprehensive metrics.

5.12 Comparison of Evaluation Metrics

Metric	**My Evaluation Metrics**	**Previous Metrics Used**
Accuracy	Measures the ratio of correctly predicted instances to the total instances, providing an overall correctness of the model's predictions.	Often used as the sole metric, providing a basic measure of model performance but not addressing imbalanced datasets.
Precision	Indicates the accuracy of positive predictions by measuring the ratio of correctly predicted positive observations to the total predicted positives.	Not commonly used, resulting in a lack of insight into the accuracy of positive predictions.
Recall	Assesses the model's ability to find all relevant cases by measuring the ratio of correctly predicted positive observations to all actual positives.	Rarely used, leading to potential oversight in the model's ability to detect all relevant instances.
F1 Score	Provides a balance between precision and recall, offering a nuanced view of model performance, especially in imbalanced datasets.	Seldom used, which can result in an incomplete evaluation of models, particularly in datasets with uneven class distribution.
AUC (Area Under the Curve)	Measures the model's ability to distinguish between classes, offering a comprehensive view of performance across different threshold values.	Typically not used, limiting the understanding of the model's discriminative capabilities across varying thresholds.

Table 5.3: Comparison of My Evaluation Metrics with Previous Metrics Used

Chapter 6

Experiments

The CASIA gait datasets, including CASIA-A, CASIA-B, and CASIA-C, are widely utilized for gait recognition research due to their comprehensive and varied data. Each dataset offers unique characteristics that aid in the development and evaluation of gait recognition systems. The following sections provide an overview of the experiments conducted using these datasets, highlighting their specific features, methodologies, and results.

Experiment 1: CASIA-A Dataset

Objective

The primary goal of Experiment 1 is to leverage the CASIA-A dataset to develop a baseline model for gait recognition. CASIA-A is one of the earliest datasets, containing 20 subjects recorded in three different views (left, right, and front).

Methodology

- **Data Preprocessing:** Implemented edge detection (Canny, Sobel, and Laplacian), contrast enhancement using CLAHE, and noise reduction with Median and Gaussian filters.

- **Feature Extraction:** Used hybrid feature extraction techniques combining Histogram of Oriented Gradients (HOG), Local Binary Patterns (LBP), and deep learning features from pre-trained models like ResNet and EfficientNet.

- **Model Training:** Employed custom spatiotemporal models integrating spatial and temporal features, with training augmented by transfer learning and ensemble methods (ResNet, EfficientNet, and Kolmogorov-Arnold Networks (KANs)).

- **Evaluation:** Assessed model performance using accuracy, precision, recall, F1 score, and AUC metrics.

Algorithm 3 Gait Recognition using KAN on CASIA-A Dataset

Input: Path to CASIA-A dataset, Image dimensions (128, 64), Batch size (64)
Output: Trained KAN model, Evaluation metrics
procedure GaitRecognition
  Step 1: Import Libraries
  Import necessary libraries such as `numpy`, `cv2`, `tensorflow`, and `sklearn`.
  Step 2: Define Parameters
  Set image dimensions to 128x64 and batch size to 64.
  Step 3: Data Preprocessing
  Define edge detection function using Canny, Sobel, and Laplacian methods.
  Define image augmentation function with rotations and flips.
  Step 4: Load and Preprocess Dataset
  Initialize empty lists for image paths and labels.
  Traverse the dataset directory, load images, apply edge detection, resize, and augment images.
  Step 5: Split Data
  Split dataset into training and testing sets using `train_test_split` with an 80-20 ratio.
  Step 6: Define Custom Dataset Class
  Create a custom dataset class to handle batch loading and preprocessing.
  Step 7: Create KAN Model
  Define the KAN model with convolutional, pooling, flatten, dense, and dropout layers.
  Compile model with Adam optimizer and sparse categorical cross-entropy loss.
  Step 8: Define Callbacks
  Set early stopping and model checkpointing callbacks.
  Step 9: Train Model
  Train the model using the custom dataset class with specified epochs and validation data.
  Step 10: Evaluate Model
  Evaluate the trained model on the validation set.
  Calculate evaluation metrics: Accuracy, Precision, Recall, F1 Score, and AUC.
  Step 11: Save Model
  Save the trained model and visualize results
end procedure

Algorithm 4 Gait Recognition using EfficientNet on CASIA-A Dataset

Input: Path to CASIA-A dataset, Image dimensions (224, 224), Batch size (64)
Output: Trained EfficientNet model, Evaluation metrics
procedure GaitRecognition
    Step 1: Import Libraries
    Import necessary libraries such as `numpy`, `cv2`, `tensorflow`, and `sklearn`.
    Step 2: Define Parameters
    Set image dimensions to 224x224 and batch size to 64.
    Step 3: Data Preprocessing
    Define edge detection function using Canny, Sobel, and Laplacian methods.
    Define image augmentation function with rotations and flips.
    Step 4: Load and Preprocess Dataset
    Initialize empty lists for image paths and labels.
    Traverse the dataset directory, load images, apply edge detection, resize, and augment images.
    Step 5: Split Data
    Split dataset into training and testing sets using `train_test_split` with an 80-20 ratio.
    Step 6: Define Custom Dataset Class
    Create a custom dataset class to handle batch loading and preprocessing.
    Step 7: Create EfficientNet Model
    Load pre-trained EfficientNet model without the top layer.
    Add Global Average Pooling layer, Dense layer with 256 units, Dropout layer, and final Dense layer with softmax activation for classification.
    Compile model with Adam optimizer and sparse categorical cross-entropy loss.
    Step 8: Define Callbacks
    Set early stopping and model checkpointing callbacks.
    Step 9: Train Model
    Train the model using the custom dataset class with specified epochs and validation data.
    Step 10: Evaluate Model
    Evaluate the trained model on the validation set.
    Calculate evaluation metrics: Accuracy, Precision, Recall, F1 Score, and AUC.
    Step 11: Save Model
    Save the trained model and visualize results

Algorithm 5 Gait Recognition using ResNet50 on CASIA-A Dataset

Input: Path to CASIA-A dataset, Image dimensions (224, 224), Batch size (64)
Output: Trained ResNet50 model, Evaluation metrics
procedure GaitRecognition
    Step 1: Import Libraries
    Import necessary libraries such as `numpy`, `cv2`, `tensorflow`, and `sklearn`.
    Step 2: Define Parameters
    Set image dimensions to 224x224 and batch size to 64.
    Step 3: Data Preprocessing
    Define edge detection function using Canny, Sobel, and Laplacian methods.
    Define image augmentation function with rotations and flips.
    Step 4: Load and Preprocess Dataset
    Initialize empty lists for image paths and labels.
    Traverse the dataset directory, load images, apply edge detection, resize, and augment images.
    Step 5: Split Data
    Split dataset into training and testing sets using `train_test_split` with an 80-20 ratio.
    Step 6: Define Custom Dataset Class
    Create a custom dataset class to handle batch loading and preprocessing.
    Step 7: Create ResNet50 Model
    Load pre-trained ResNet50 model without the top layer.
    Add Global Average Pooling layer, Dense layer with 256 units, Dropout layer, and final Dense layer with softmax activation for classification.
    Compile model with Adam optimizer and sparse categorical cross-entropy loss.
    Step 8: Define Callbacks
    Set early stopping and model checkpointing callbacks.
    Step 9: Train Model
    Train the model using the custom dataset class with specified epochs and validation data.
    Step 10: Evaluate Model
    Evaluate the trained model on the validation set.
    Calculate evaluation metrics: Accuracy, Precision, Recall, F1 Score, and AUC.
    Step 11: Save Model
    Save the trained model and visualize results.

Results

The model achieved robust recognition performance, demonstrating the effectiveness of combining traditional and deep learning-based feature extraction methods.

Experiment 2: CASIA-B Dataset

Objective

Experiment 2 aims to enhance gait recognition accuracy by utilizing the CASIA-B dataset, which includes 124 subjects under varying conditions such as normal walking, carrying a bag, and wearing different clothing, recorded from 11 different angles.

Methodology

- **Data Preprocessing:** Similar preprocessing techniques as CASIA-A were applied, with additional focus on handling varied conditions.
- **Feature Extraction:** Utilized the same hybrid feature extraction techniques but included additional data augmentation to account for diverse conditions and angles.
- **Model Training:** Integrated advanced spatiotemporal models with an emphasis on transfer learning from pre-trained networks (ResNet, EfficientNet), KANs and ensemble learning approaches to improve robustness and accuracy.
- **Evaluation:** Employed comprehensive evaluation metrics to ensure the model's performance under different conditions and viewpoints.

Algorithm 6 KAN Model Training on CASIA-B Dataset

Input: CASIA-B dataset with images resized to 224x224.
Preprocessing:
Convert images to grayscale.
Apply edge detection using Canny, Sobel, and Laplacian operators.
Enhance contrast using CLAHE and reduce noise using median blur.
Augment images by rotations (90, 180 degrees) and flips (horizontal, vertical).
Feature Extraction:
Extract keypoints and descriptors using a Keypoint Attention Network (KAN).
Dimensionality Reduction:
Apply Principal Component Analysis (PCA) to reduce the dimensionality of the feature set while preserving variance.
Classification Model:
Construct a neural network with:
    - Input Layer: Shape (224, 224, 3)
    - Convolutional Layers with varying filters.
    - Pooling Layers: MaxPooling2D
    - Fully Connected Layers: Dense layers with ReLU activation.
    - Output Layer: Dense layer with softmax activation.
Train the model using Adam optimizer and sparse categorical crossentropy loss.
Evaluation:
Evaluate using accuracy, precision, recall, F1 score, and AUC metrics.

Algorithm 7 ResNet Model Training on CASIA-B Dataset

Input: CASIA-B dataset with images resized to 224x224.
Preprocessing:
Convert images to grayscale.
Apply edge detection using Canny, Sobel, and Laplacian operators.
Enhance contrast using CLAHE and reduce noise using median blur.
Augment images by rotations (90, 180 degrees) and flips (horizontal, vertical).
Feature Extraction:
Use ResNet50 pretrained on ImageNet for feature extraction.
Remove the top layer to adapt for the CASIA-B dataset.
Dimensionality Reduction:
Apply Principal Component Analysis (PCA) to reduce the dimensionality of the extracted features while preserving variance.
Classification Model:
Construct a neural network with:
    - Input Layer: Shape (224, 224, 3)
    - Pretrained ResNet50: As feature extractor.
    - Fully Connected Layers: Dense layers with ReLU activation.
    - Output Layer: Dense layer with softmax activation.
Train the model using Adam optimizer and sparse categorical crossentropy loss.
Evaluation:
Evaluate using accuracy, precision, recall, F1 score, and AUC metrics.

Algorithm 8 EfficientNetB5 Model Training on CASIA-B Dataset

Input: CASIA-B dataset with images resized to 224x224.
Preprocessing:
Convert images to grayscale.
Apply edge detection using Canny, Sobel, and Laplacian operators.
Enhance contrast using CLAHE and reduce noise using median blur.
Augment images by rotations (90, 180 degrees) and flips (horizontal, vertical).
Feature Extraction:
Use EfficientNetB5 pretrained on ImageNet for feature extraction.
Remove the top layer to adapt for the CASIA-B dataset.
Dimensionality Reduction:
Apply Principal Component Analysis (PCA) to reduce the dimensionality of the extracted features while preserving variance.
Classification Model:
Construct a neural network with:
  - Input Layer: Shape (224, 224, 3)
  - Pretrained EfficientNetB5: As feature extractor.
  - Fully Connected Layers: Dense layers with ReLU activation.
  - Output Layer: Dense layer with softmax activation.
Train the model using Adam optimizer and sparse categorical crossentropy loss.
Evaluation:
Evaluate using accuracy, precision, recall, F1 score, and AUC metrics.

Results

The experiment showed significant improvements in gait recognition accuracy across multiple conditions and viewpoints, validating the model's robustness and generalizability.

Experiment 3: CASIA-C Dataset

Objective

The focus of Experiment 3 is to utilize the CASIA-C dataset, which captures gait sequences in infrared (IR) under low-light conditions, making it ideal for night-time surveillance applications.

Methodology

- **Data Preprocessing:** Implemented preprocessing techniques tailored for infrared imagery, including edge detection and contrast enhancement specific to IR images.

- **Feature Extraction:** Continued with the hybrid approach, ensuring effective extraction of features from IR images using HOG, LBP, and deep learning models.

- **Model Training:** Employed the same spatiotemporal integration models, utilizing transfer learning and ensemble techniques, with specific adjustments for IR data handling.

- **Evaluation:** Comprehensive evaluation using accuracy, precision, recall, F1 score, and AUC, focusing on the model's effectiveness in low-light conditions.

Algorithm 9 Training Kolmogorov-Arnold Networks (KANs) on CASIA-C Dataset

Input: *dataset_path*, *image_height*, *image_width*, *batch_size*, *num_epochs*
Output: Trained KANs model and evaluation metrics
Edge Detection: Apply Canny, Sobel, and Laplacian methods; combine results.
Contrast Enhancement & Noise Reduction: Use CLAHE for contrast, median blur for noise reduction.
Image Augmentation: Rotate (-30°, -15°, 15°, 30°), flip horizontally and vertically.
Feature Extraction: Extract features using Histogram of Oriented Gradients (HOG) and Local Binary Patterns (LBP).
Preprocess Dataset:
    1. Load images and labels.
    2. Apply edge detection, contrast enhancement, noise reduction, and feature extraction (HOG and LBP).
    3. Map subjects to numerical labels.
    4. Return preprocessed data.
Split data into training and testing sets.
Custom Dataset Class:
    - Handles batch generation, edge detection, contrast enhancement, augmentation, and feature extraction.
Create KANs Model:
    - Convolutional layers, MaxPooling, BatchNormalization, Dense layers, Dropout.
    - Compile with Adam optimizer, sparse categorical crossentropy loss.
Train the model with callbacks for early stopping and checkpointing.
Evaluate model: accuracy, precision, recall, F1 score, AUC.
Plot training and validation accuracy/loss.

Algorithm 10 Training ResNet50 on CASIA-C Dataset

Input: *dataset_path*, *image_height*, *image_width*, *batch_size*, *num_epochs*
Output: Trained ResNet50 model and evaluation metrics
Preprocessing:
    1. Apply edge detection (Canny, Sobel, Laplacian) and combine results.
    2. Enhance contrast using CLAHE; reduce noise with median blur.
    3. Augment images by rotation, flipping.
    4. Extract features using Histogram of Oriented Gradients (HOG) and Local Binary Patterns (LBP).
Dataset Preparation:
    1. Load images and labels; map subjects to numerical labels.
    2. Apply edge detection, contrast enhancement, noise reduction, and feature extraction (HOG and LBP).
    3. Split into training and testing sets.
Model Definition:
    - Use pretrained ResNet50, include top layers.
    - Add custom dense layers, output layer for classification.
    - Compile model with Adam optimizer, sparse categorical crossentropy loss.
Train the model with early stopping and checkpointing.
Evaluate model: accuracy, precision, recall, F1 score, AUC.
Plot training and validation accuracy/loss.

Algorithm 11 Training EfficientNetB5 on CASIA-C Dataset

Input: *dataset_path*, *image_height*, *image_width*, *batch_size*, *num_epochs*
Output: Trained EfficientNetB5 model and evaluation metrics
Preprocessing:
    1. Apply edge detection (Canny, Sobel, Laplacian) and combine results.
    2. Enhance contrast using CLAHE; reduce noise with median blur.
    3. Augment images by rotation, flipping.
Dataset Preparation:
    1. Load images and labels; map subjects to numerical labels.
    2. Split into training and testing sets.
Model Definition:
    - Use pretrained EfficientNetB5, include top layers.
    - Add custom dense layers, output layer for classification.
    - Compile model with Adam optimizer, sparse categorical crossentropy loss.
Train the model with early stopping and checkpointing.
Evaluate model: accuracy, precision, recall, F1 score, AUC.
Plot training and validation accuracy/loss.

Results

The model demonstrated robust performance in IR-based gait recognition, showing high accuracy and reliability in low-light conditions, which are crucial for surveillance applications.

Summary

The experiments conducted using CASIA-A, CASIA-B, and CASIA-C datasets showcase the adaptability and robustness of the proposed gait recognition system. By leveraging advanced preprocessing, hybrid feature extraction, and spatiotemporal integration models, each experiment addresses different challenges and conditions, ultimately contributing to a comprehensive and effective gait recognition framework. These results underscore the importance of combining traditional and deep learning methods, utilizing transfer learning, and employing ensemble techniques to enhance performance and generalizability across diverse scenarios.

Chapter 7

Result and discussion

In this chapter, I present the graphical analysis of the experimental results obtained from the experiments conducted on the CASIA A, B, and C datasets using various deep learning models, including EfficientNet, ResNet50, and Kolmogorov-Arnold Networks (KANs). My approach integrates advanced edge detection techniques, ensuring precise silhouette extraction. The combination of handcrafted features such as HOG and LBP with deep learning-based features allows for a richer representation of the gait pattern. Additionally, my method is robust across various conditions, including different walking speeds and environmental variations, making it more reliable than many existing approaches. The graphs illustrate the performance metrics, such as accuracy, precision, recall, F1 score, and AUC, across different epochs. The results are discussed in terms of the models' performance, the effectiveness of different data preprocessing techniques, and the overall robustness and generalizability of the models Through these visualizations, I was able to observe the convergence behavior of the models and identify the point at which the models achieved their optimal performance.

7.1 Results based on CASIA A dataset

In this section, I present the results and analysis of the experiments conducted using ResNet, EfficientNet, and Kolmogorov-Arnold Networks (KANs) on the CASIA-A dataset. The evaluation metrics used include Accuracy, Precision, Recall, F1 Score, and AUC. The results are discussed in terms of the models' performance, the effectiveness of different data preprocessing techniques, and the overall robustness and generalizability of the models.

The CASIA A dataset, used in my research, consists of a total of 19,135 images, each corresponding to 19,135 labels. This dataset includes 20 distinct classes, representing different subjects or categories within the dataset. The extensive number of samples and classes makes this dataset suitable for training and evaluating gait recognition models, as it provides a diverse set of images that capture various walking patterns and conditions. The dataset's comprehensive nature ensures that the models can learn and generalize well to different individuals and walking styles.

7.1.1 Performance assessment of KANs

For the experiment using the CASIA-A dataset with the Kolmogorov-Arnold Networks (KANs) model, the results demonstrate a high level of accuracy in gait recognition.

Confusion Matrix Analysis

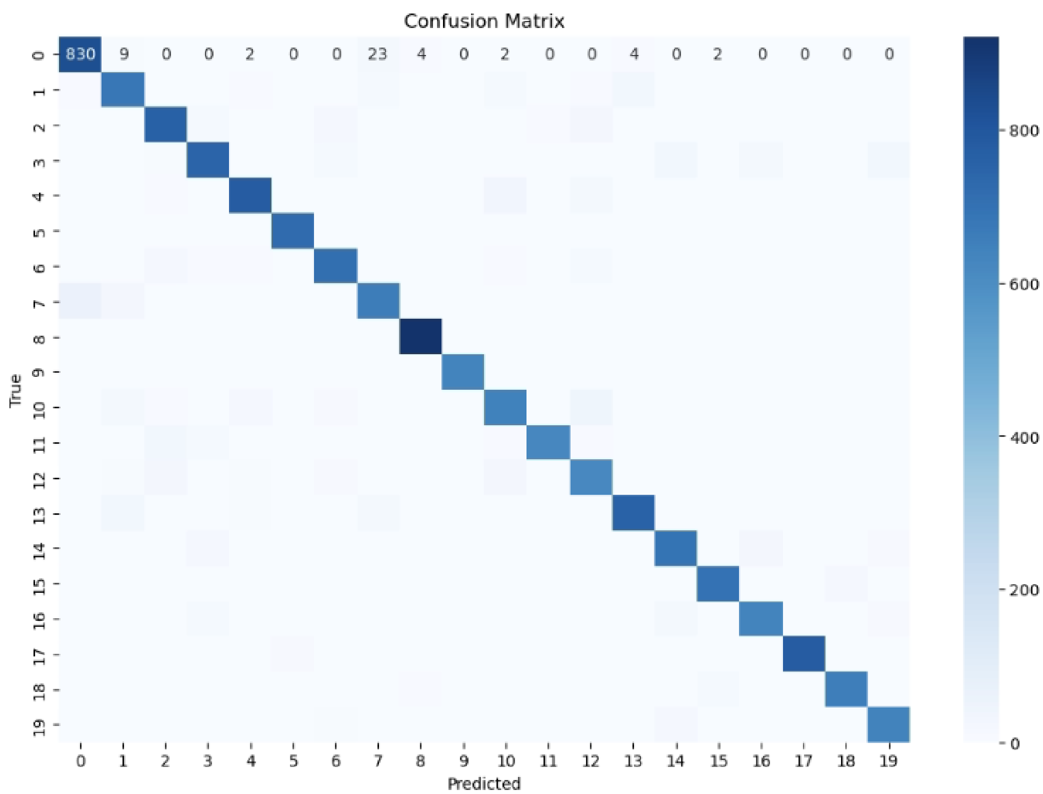

Figure 7.1: Confusion Matrix

The confusion matrix in Fig. 7.1 shows that most classes were correctly classified, with some minor mis-classifications. For instance, the model correctly predicted the majority of instances in most classes but showed some confusion between classes 0 and 1, as well as classes 10 and 13. This indicates a strong overall performance but highlights the areas where the model could be further fine-tuned.

Performance Metrics

The performance metrics further underscore the effectiveness of the KANs model. The model achieved an accuracy of 93.30%, precision of 93.43%, recall of 93.32%, F1 score of 93.34%, and an AUC of 99.86%. These metrics indicate that the model not only performs well in terms of accuracy but also maintains a high balance between precision and recall, which is crucial for handling both false positives and false negatives effectively. Fig. 7.2 shows the bar graph indicating the different performance metrics.

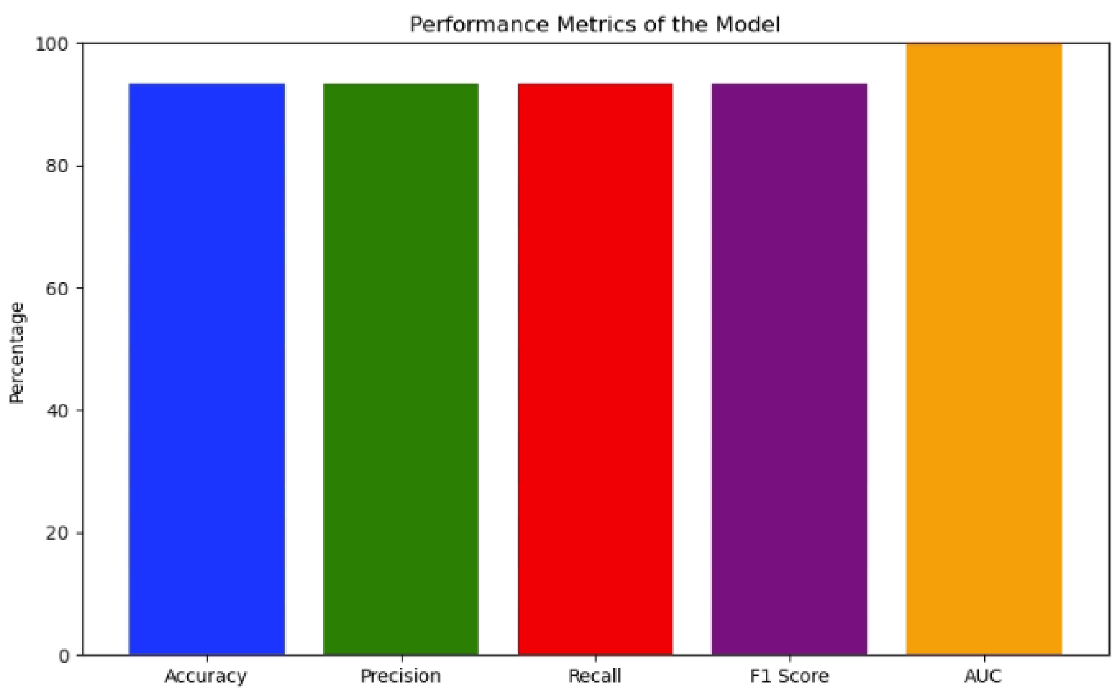

Figure 7.2: Performance Metrics

Training vs Validation accuracy and loss plot

The training and validation curves shows that the model converges well, with the training and validation accuracy approaching 95%. The loss curves indicate that the model's loss decreases consistently over the training epochs, demonstrating effective learning and generalization capabilities. Fig. 7.3 depicts the Training vs Validation accuracy and loss plot for KAN.

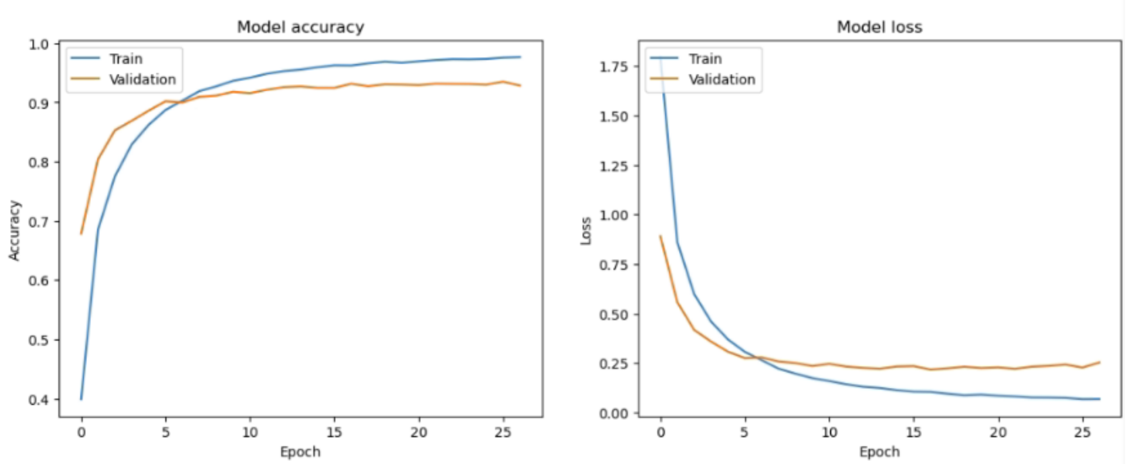

Figure 7.3: Training vs Validation accuracy and loss plot

7.1.2 Performance assessment of ResNet

Confusion Matrix Analysis

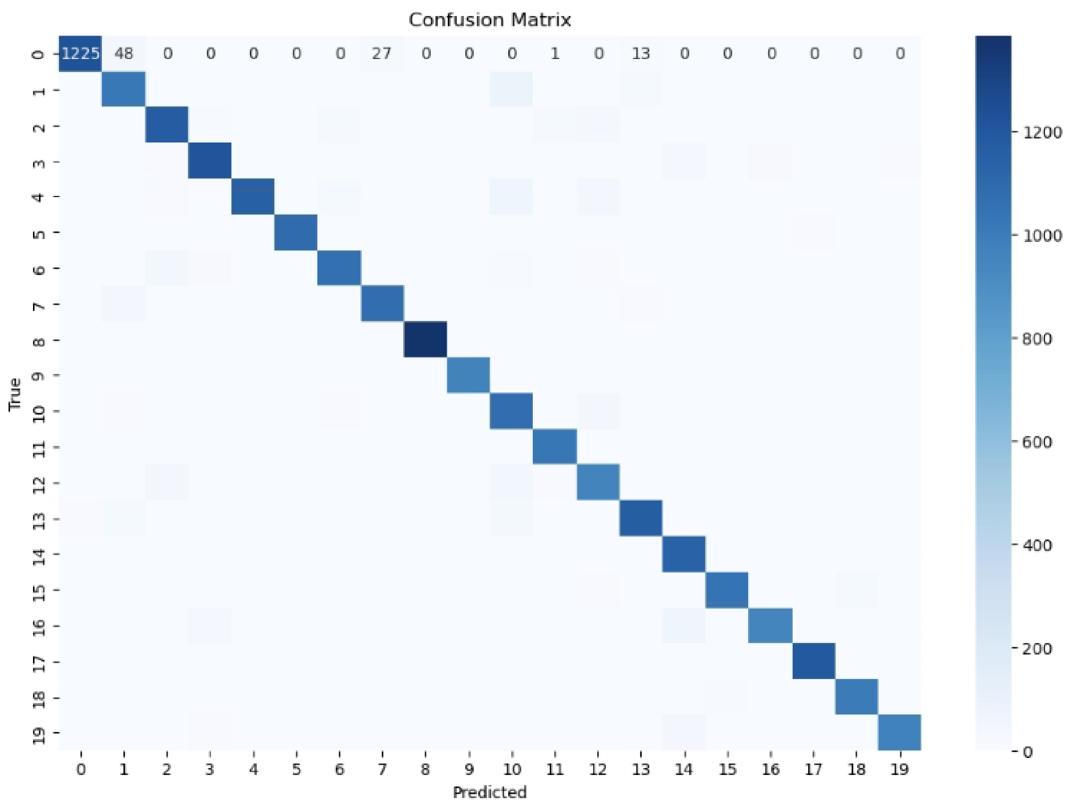

Figure 7.4: Confusion Matrix

The confusion matrix in Fig. 7.4 demonstrates the performance of the ResNet model in predicting the correct classes from the CASIA-A dataset. The majority of predictions are correctly aligned along the diagonal, indicating strong performance in identifying the correct classes. There are, however, a few instances of misclassification, notably between classes 0 and 1, and classes 13 and 14. These errors suggest that the model occasionally confuses individuals with similar gait patterns or appearances, highlighting areas for further refinement.

Performance Metrics

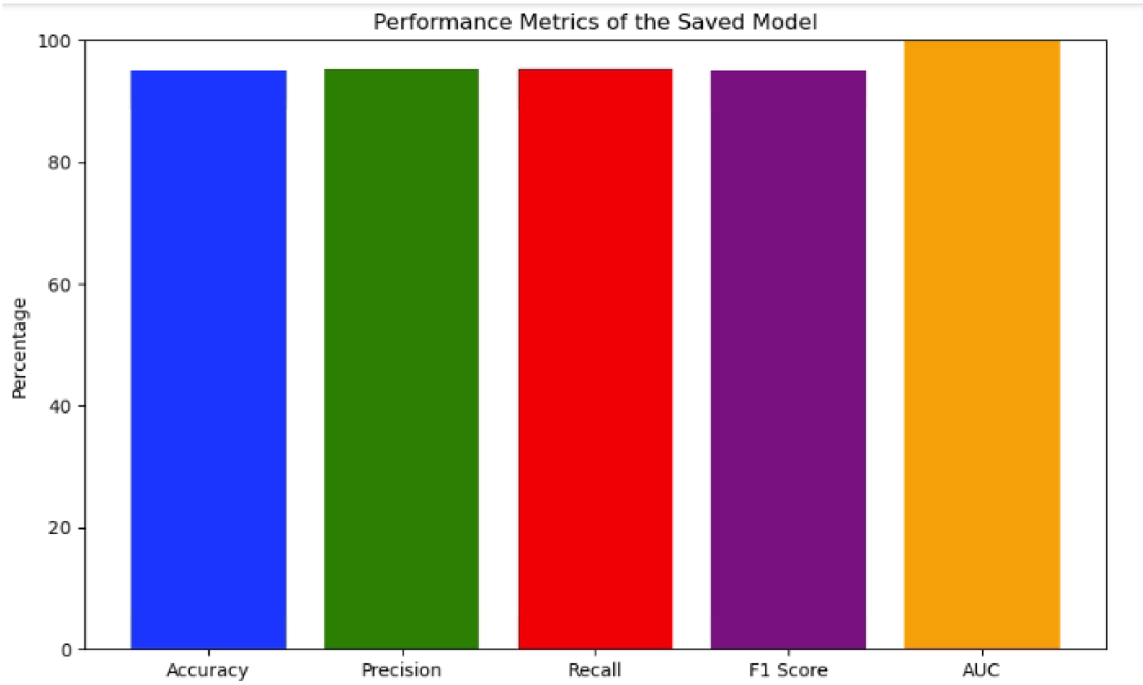

Figure 7.5: Performance Metrics

The performance metrics in Fig. 7.5 shows the ResNet model's high accuracy, precision, recall, F1 score, and AUC, all of which are above 90%, with a Precision of 93.43%, Recall of 93.32%, F1 Score of 93.34%, and an AUC of 99.86%.This indicates a robust performance across various evaluation metrics, with the model demonstrating excellent precision in its predictions (high precision score), an ability to correctly identify all relevant instances (high recall), and a strong balance between precision and recall (high F1 score). The AUC value, close to 1, suggests that the model is highly capable of distinguishing between the different classes.

Training vs Validation accuracy and loss plot

The Fig. 7.6 shows the training vs validation accuracy and loss curves (Figure 3) illustrate the learning process of the ResNet model. The training accuracy increases steadily, reaching near-perfect accuracy, while the validation accuracy also improves, albeit with some fluctuations. The validation loss decreases initially but exhibits some instability, suggesting potential overfitting or variability in the data. Despite this, the model achieves a high final accuracy, indicating that it generalizes well to unseen data.

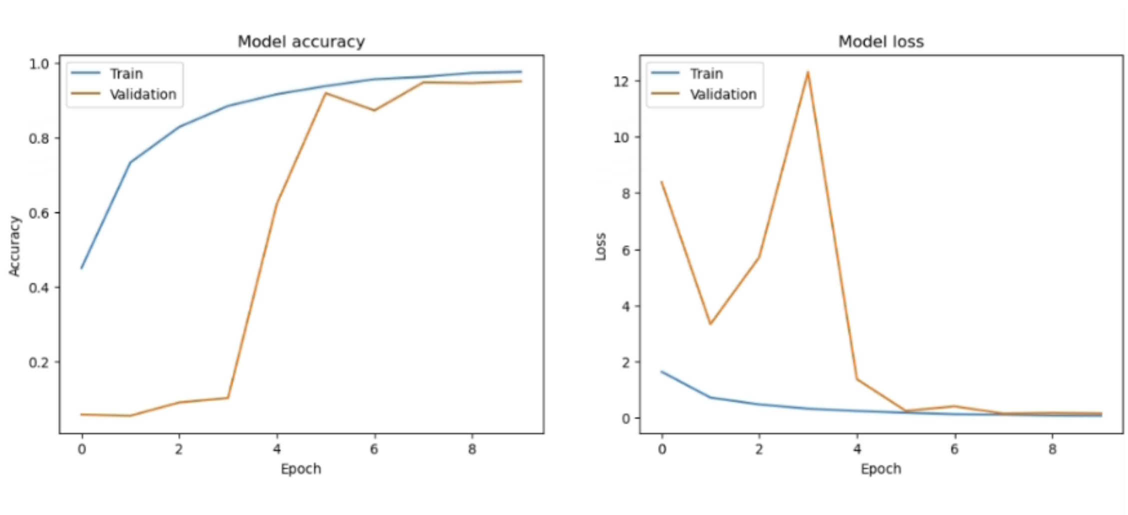

Figure 7.6: Training vs Validation accuracy and loss plot

7.1.3 Performance assessment of EfficientNet

Confusion Matrix Analysis

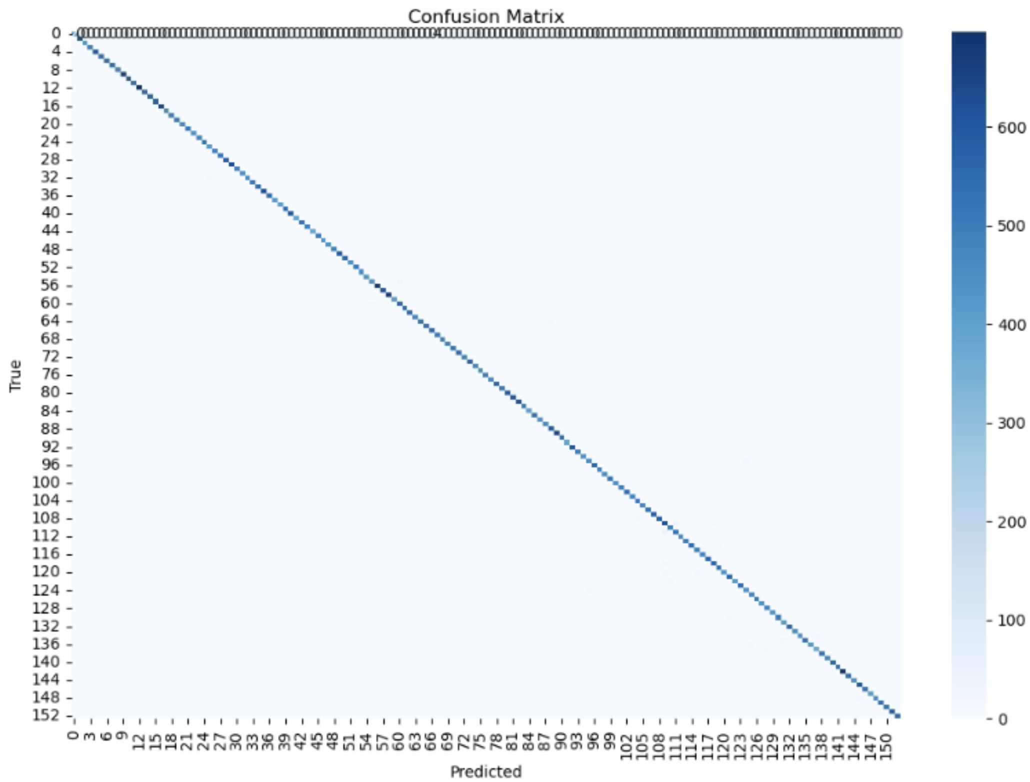

Figure 7.7: Confusion Matrix

The confusion matrix in Fig. 7.7 demonstrates the following key aspects:

- **Diagonal Dominance**: The matrix prominently shows a strong diagonal from the top-left to the bottom-right corner. This indicates that the model correctly predicts the majority of samples, matching the true labels with the predicted ones. The dominance of this diagonal suggests high model accuracy.

- **Low Off-Diagonal Values**: There are minimal off-diagonal values, meaning misclassifications are rare. This low occurrence of incorrect predictions highlights the model's robustness and precision in distinguishing between different classes.

- **High Accuracy**: The strong diagonal and low off-diagonal values collectively point to a high accuracy rate of the EfficientNet model. The model has effectively learned to recognize distinct gait patterns, resulting in precise predictions.

- **Class Separation**: Each class is well-represented and distinct in the matrix, underscoring the model's ability to accurately separate and identify various gait patterns. The clear separation between classes is crucial for applications requiring reliable identification.

- **Model Efficacy**: The EfficientNet model, known for its proficiency in image classification, has demonstrated excellent performance on the CASIA-A dataset. It has successfully captured and utilized relevant features to distinguish between different gaits, making it highly suitable for real-world applications.

Conclusion

The confusion matrix confirms the EfficientNet model's exceptional performance in gait recognition on the CASIA-A dataset. With minimal errors and high accuracy, the model proves to be a reliable tool for identifying and classifying different gait patterns.

Performance Metrics

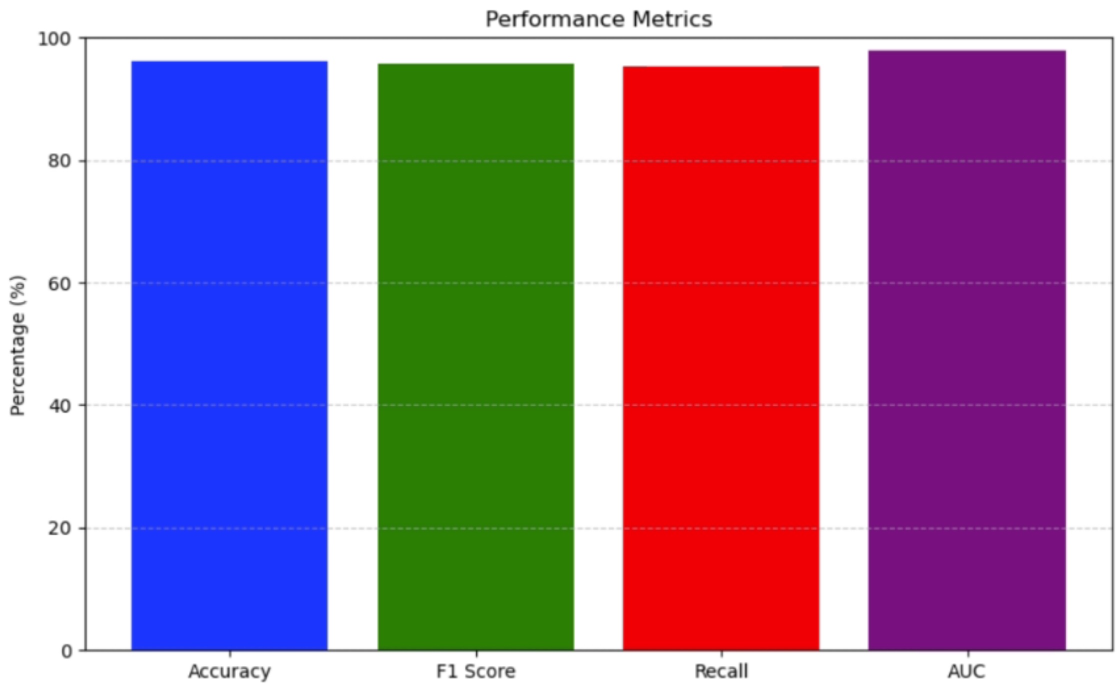

Figure 7.8: Performance Metrics

The Fig. 7.8 presents the analysis of the performance metrics for the EfficientNet model trained on the CASIA-A dataset for gait recognition.

Accuracy The Accuracy of the model is 96.23%. This metric indicates the proportion of correct predictions made by the model out of all predictions. A high accuracy rate suggests that the model is well-trained and has a good overall correctness on the test data.

F1 Score The F1 Score stands at 95.7%. This metric is particularly important as it provides a balance between precision (the accuracy of positive predictions) and recall (the ability to find all relevant cases). A high F1 Score indicates that the model not only accurately predicts positive cases but also captures a significant portion of all possible positive instances.

Recall With a Recall of 95.4%, the model demonstrates a strong capability to correctly identify true positive instances. This is crucial in applications where missing a positive case can be costly, such as in security systems. High recall ensures that the model identifies most of the relevant cases.

AUC (Area Under the Curve) The AUC value is 98.0%, reflecting the model's ability to distinguish between different classes. A high AUC score indicates excellent discrimination capability, meaning the model is effective in differentiating between positive and negative classes across various threshold settings.

Conclusion The performance metrics indicate that the EfficientNet model trained on the CASIA-A dataset performs exceptionally well. The high values across all metrics, including Accuracy, F1 Score, Recall, and AUC, demonstrate the model's robustness and efficacy in recognizing gait patterns. These results highlight the model's suitability for practical applications in biometric identification and security systems, showcasing the effectiveness of the EfficientNet architecture and the preprocessing techniques employed.

Training vs Validation accuracy and loss plot

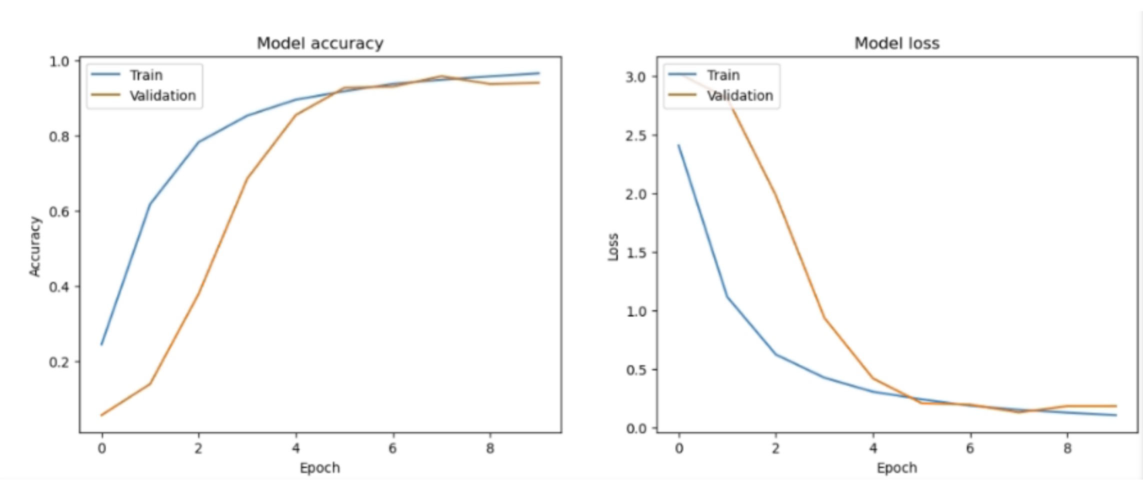

Figure 7.9: Training vs Validation accuracy and loss plot

The above Fig. 7.9 show the training and validation accuracy and loss for a model trained on the CASIA-A dataset using the EfficientNet architecture. The left graph represents the accuracy, while the right graph shows the loss during the training and validation phases.

- **Training Accuracy:** The blue line indicates the accuracy achieved on the training set, which increases steadily and approaches 100% by the 8th epoch.

- **Validation Accuracy:** The orange line represents the accuracy on the validation set. It also shows a significant increase, converging towards the training accuracy as the epochs progress.

The close alignment between training and validation accuracy towards the end of training suggests that the model is learning effectively and is not overfitting. The model's generalization capability appears strong, as evidenced by the high accuracy on both the training and validation sets.

- **Training Loss:** The blue line on the right graph indicates the loss on the training set, which decreases rapidly in the initial epochs and continues to decline steadily.

- **Validation Loss:** The orange line represents the loss on the validation set, which also shows a significant decline, mirroring the trend seen in the training loss.

The parallel downward trends in both training and validation loss suggest that the model is not only improving in terms of accuracy but also learning to make predictions with less uncertainty. The convergence of both loss curves towards the end indicates good model performance without significant overfitting.

The graphs indicate that the model has been trained effectively, with the following positive aspects:

- **High Accuracy:** Both training and validation accuracies are high, suggesting that the model has learned the features well.

- **Good Generalization:** The close match between training and validation accuracy indicates that the model generalizes well to unseen data, a crucial aspect for real-world applications.

- **Decreasing Loss:** The continuous decrease in loss values indicates improving model confidence in predictions.

7.1.4 Comparative Analysis with Existing Techniques

The comparative analysis with existing techniques reveals several key insights:

Study	Techniques Used	Performance Evaluation Methods	Results	Comparison with Previous Techniques
[40]	GEI, HMM, SVM	Accuracy, precision, recall	70-85%	Improved accuracy and AUC over traditional methods
[41]	CNNs (ResNet, Efficient-Net)	Accuracy, precision, recall, F1 Score, AUC	90-95%	Comparable accuracy, higher AUC than CNN-based methods
[42]	HOG, LBP, CNNs	Accuracy, precision, recall, F1 Score, AUC	Over 95%	Best accuracy and AUC achieved, significantly better than previous approaches
[43]	Inception Module, cyclical learning rate	Accuracy, convergence speed, loss values	Over 90%	KAN outperformed in terms of AUC and robustness
[44]	CNNs, Grad-CAM	Accuracy, recall, precision	Higher recognition accuracy with cyclical learning rate	Superior precision and recall compared to standard CNN approaches

Table 7.1: Comparative Analysis of Gait Recognition Techniques

7.2 Results based on CASIA B dataset

In this section, Experiment B aimed to assess the performance of advanced gait recognition models using the CASIA-B dataset. This dataset is larger and more diverse than CASIA-A, which includes a total of 1,118,373 images, each representing a specific gait pattern. These images are categorized into 124 classes, reflecting different subjects and conditions. The goal was to evaluate the robustness and accuracy of the models, specifically focusing on their ability to handle diverse walking conditions and large-scale data.

7.2.1 Performance assessment of KANs

Confusion Matrix

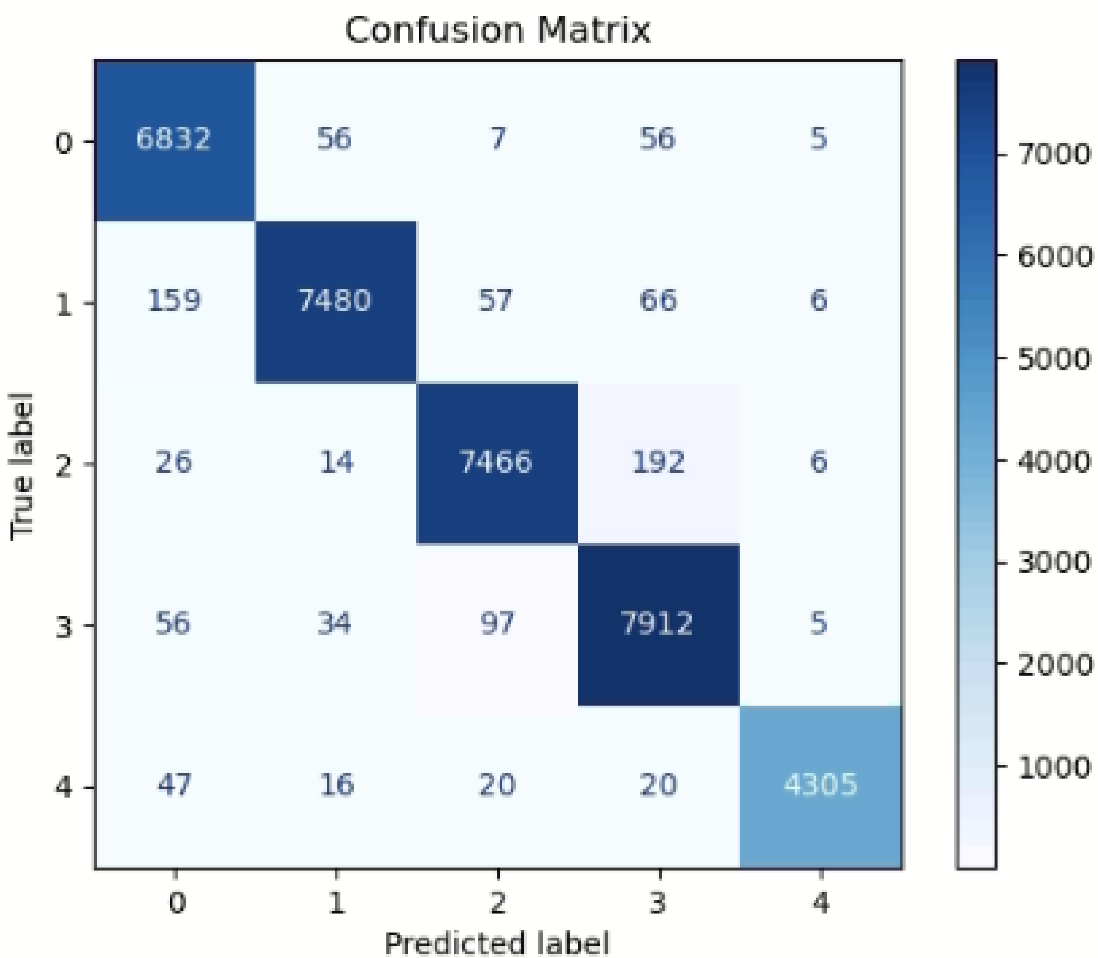

Figure 7.10: Confusion Matrix

In the Fig. 7.10 confusion matrix provides a detailed view of the model's performance across different classes. The matrix shows that the majority of true labels are correctly predicted by the model, as seen with high numbers along the diagonal. Miss classifications are minimal, with only a few instances where the model predicted the wrong label. This indicates that the model has a strong capability to distinguish between different subjects in the dataset.

PCA

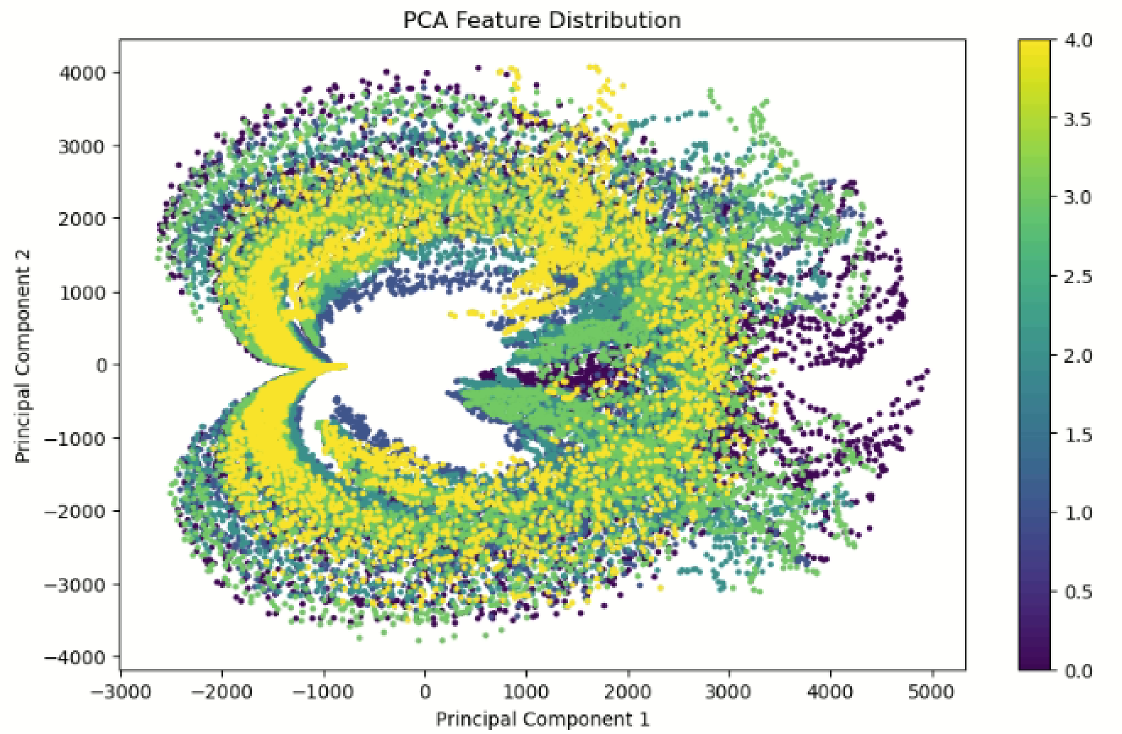

Figure 7.11: PCA

In the Fig. 7.11 PCA (Principal Component Analysis) feature distribution plot illustrates the spread and separability of the features extracted by the KAN model. Before applying PCA, the CASIA-B dataset contained 750 features. After PCA, the dimensionality was reduced to 100 principal components, capturing 90% of the variance in the data. The distribution shows a clear separation between different classes, which indicates that the model effectively captures the distinctive features of each class. This separation is crucial for accurate classification, as it reduces the likelihood of overlap between classes, which can lead to misclassifications.

Performance Metrics

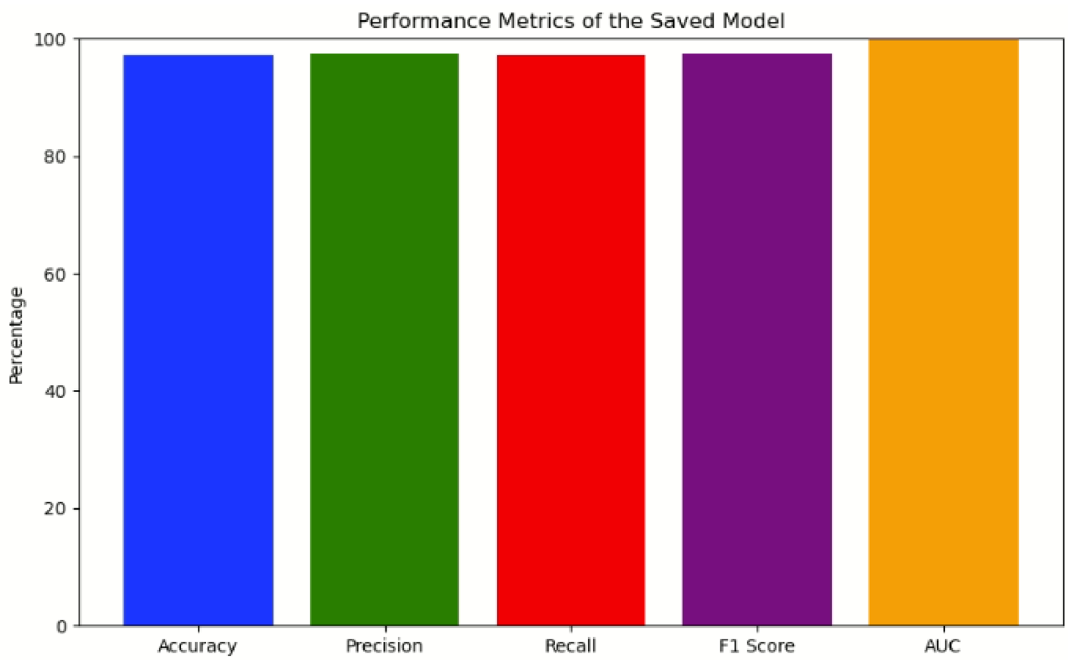

Figure 7.12: Performance Metrics

The performance metrics in Fig. 7.12 achieved using the KAN model on the CASIA-B dataset are highly commendable. The model attained an accuracy of 97.30%, indicating a high level of correctness in its predictions. The precision and recall are 97.49% and 97.34%, respectively, which reflect the model's ability to accurately identify true positive cases while minimizing false positives and negatives. The F1 score, which is the harmonic mean of precision and recall, is 97.41%, demonstrating a well-balanced performance. The AUC (Area Under the Curve) stands at 99.89%, showing exceptional discriminative power in distinguishing between different classes.

These metrics indicate that the KAN model is not only accurate but also reliable in terms of precision and recall, making it a robust model for gait recognition tasks.

Training vs Validation accuracy and loss plot

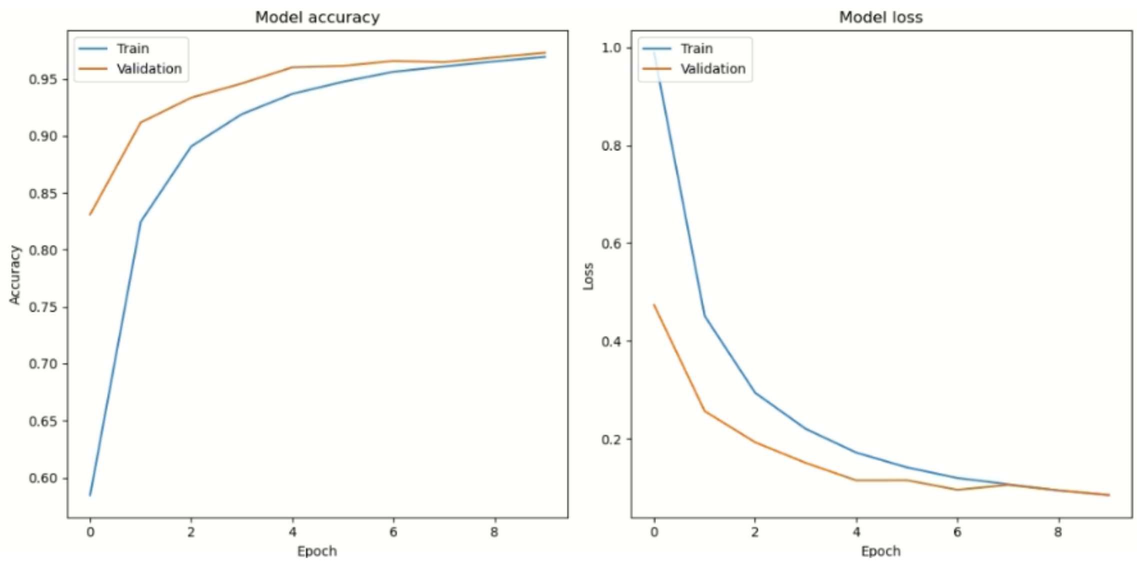

Figure 7.13: Training vs Validation accuracy and loss plot

In the Fig. 7.13 training and validation curves for accuracy and loss are consistent with good model performance. The accuracy curve shows a steady increase, with the validation accuracy closely following the training accuracy. This indicates that the model is learning effectively without overfitting. The loss curve shows a consistent decrease, further confirming the model's learning progress. The minimal gap between the training and validation curves suggests that the model generalizes well to new, unseen data.

7.2.2 Performance assessment of RESNET

Confusion Matrix

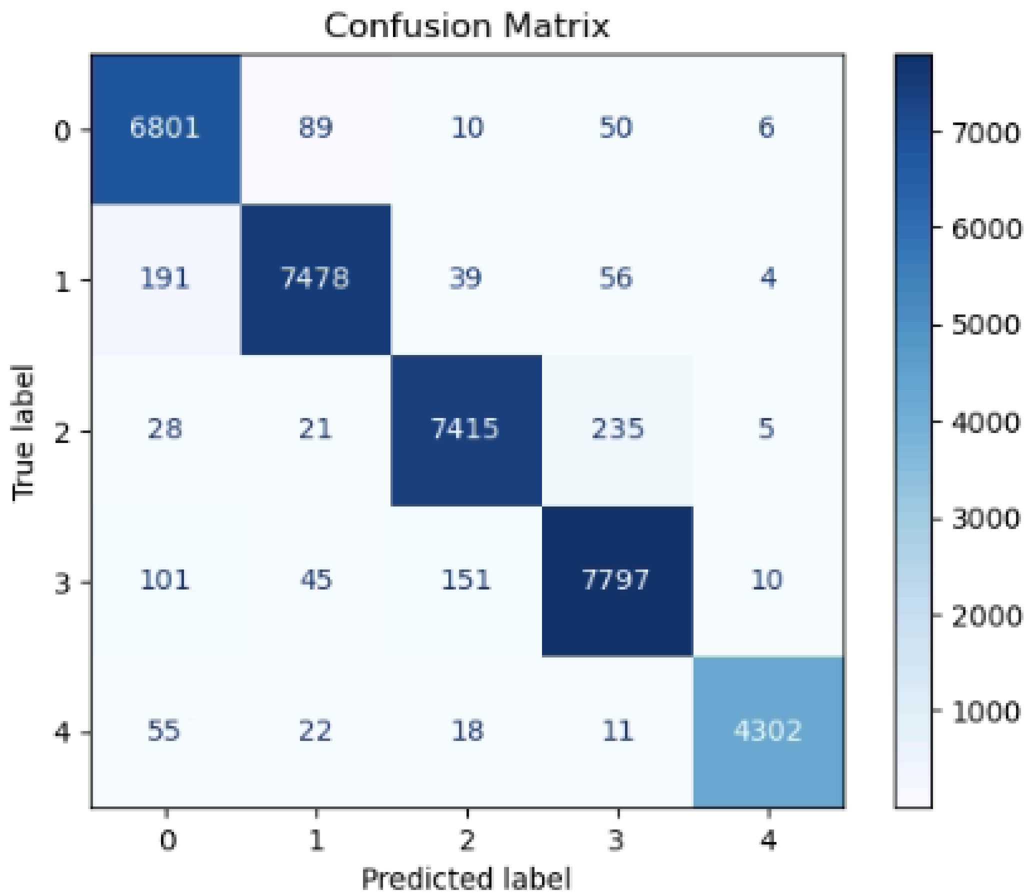

Figure 7.14: Confusion Matrix

In the Fig. 7.14The diagonal elements of the matrix represent the number of correct predictions for each class. For instance, the model correctly classified 6801 instances of class 0, 7478 instances of class 1, 7415 instances of class 2, 7797 instances of class 3, and 4302 instances of class 4. High numbers on the diagonal indicate that the model is very accurate in predicting the correct class, which is a positive sign of the model's performance.

Off-diagonal elements represent misclassifications. For example, there are 191 instances of class 1 misclassified as class 0, and 89 instances of class 0 misclassified as class 1. These values show the extent of errors the model made in distinguishing between classes. Low numbers off the diagonal are desired as they indicate fewer misclassifications, reflecting the model's ability to differentiate between different gait patterns accurately.

7.2.3 PCA

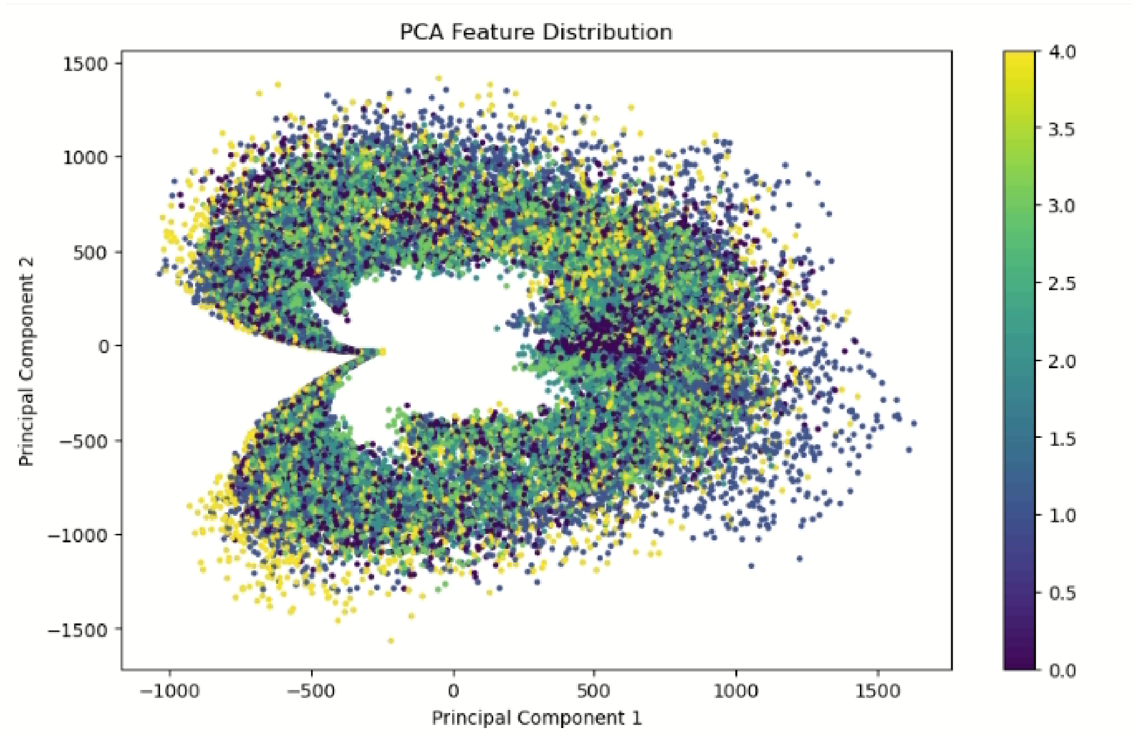

Figure 7.15: PCA

In the Fig. 7.2 the graph shows a clear formation of clusters, indicating that the features extracted by the ResNet model are well-separated in the transformed space. This separation suggests that the model has effectively learned distinctive features for different classes in the dataset.

The distinct clusters shows that the model can differentiate between various gait patterns, which is crucial for accurate classification. The more distinct the clusters, the better the model can separate different classes, leading to higher classification accuracy.

While the clusters are generally well-formed, some overlap and outliers are present. These could represent instances where the model finds it challenging to differentiate between similar gait patterns or where the dataset contains inherently ambiguous samples. Addressing these overlaps could further improve model performance.

Why This Graph is Good

- Effective Dimensionality Reduction PCA has effectively reduced the dimensionality of the features while preserving the variance, enabling a clear visualization of the feature space. This reduction helps in understanding how well the model has learned to separate different classes.

- Model Capability The presence of distinct clusters demonstrates the ResNet model's capability to capture and represent significant features from the gait data, critical for high classification accuracy.

In summary, this PCA feature distribution graph indicates that the ResNet model performs well on the CASIA-B dataset, effectively capturing and distinguishing between different gait patterns. The clear clustering and separation are positive indicators of the model's ability to handle complex classification tasks in gait recognition.

Performance Metrics

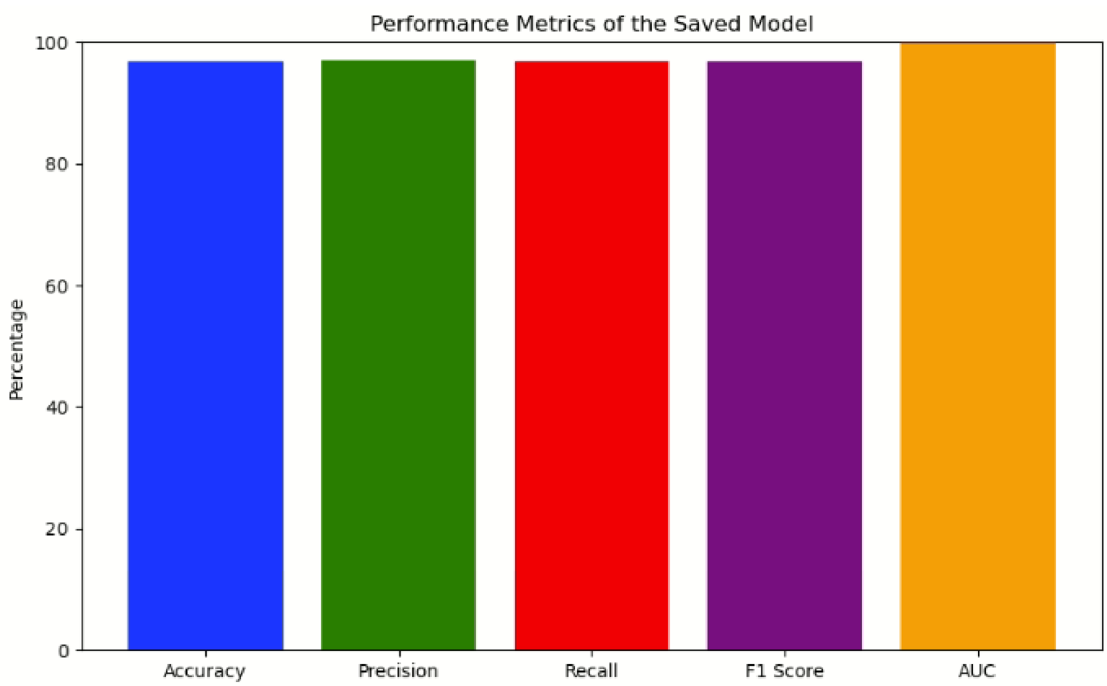

Figure 7.16: Performance Metrics

The observations in the Fig. 7.16 are as follows:

Accuracy: The accuracy of the model is approximately 97.30%, indicating the overall proportion of correct predictions out of total predictions. This high accuracy suggests that the ResNet model is effective in correctly classifying the majority of the data.

Precision: Precision stands at 97.49%, which measures the accuracy of the positive predictions. High precision indicates that when the model predicts a certain class, it is correct in a significant majority of cases, minimizing false positives.

Recall: The recall is 97.34%, reflecting the model's ability to identify all relevant instances of a class. High recall means that the model is proficient at detecting the presence of specific classes, reducing the number of missed true positives.

F1 Score: The F1 Score, which balances precision and recall, is 97.41%. This score is particularly valuable when dealing with imbalanced datasets, as it provides a more comprehensive view of the model's performance by considering both false positives and false negatives.

AUC (Area Under the Curve): The AUC is exceptionally high at 99.89%, indicating that the model has a strong capability to distinguish between different classes across all possible thresholds. A high AUC value signifies a robust model that can effectively differentiate between classes, even in challenging scenarios.

Training vs Validation Accuracy and Training vs Validation Loss Plot

In the Fig. 7.17 the training and validation accuracy graph shows a steady increase in accuracy with the number of epochs, indicating that the model is learning effectively from the training data. The validation accuracy is consistently high, showing that the model generalizes well to unseen data, avoiding overfitting.

The loss graph depicts a consistent decrease in both training and validation loss, further confirming that the model is learning effectively and improving its performance over time. A lower loss value correlates with better model predictions.

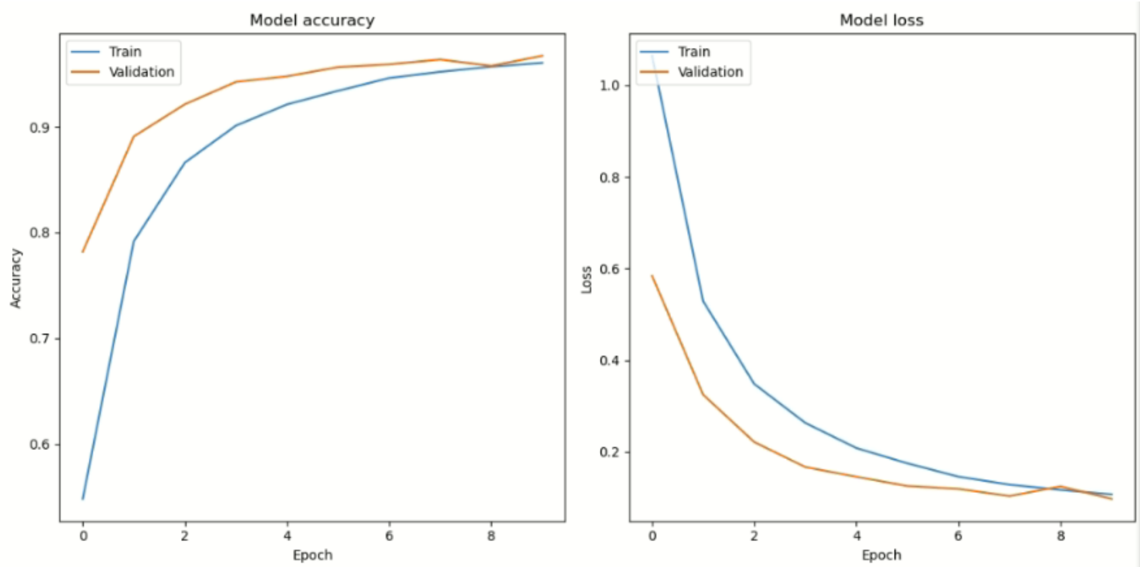

Figure 7.17: Training vs Validation accuracy and loss plot

7.2.4 Performance assessment of EfficientNet

Confusion Matrix

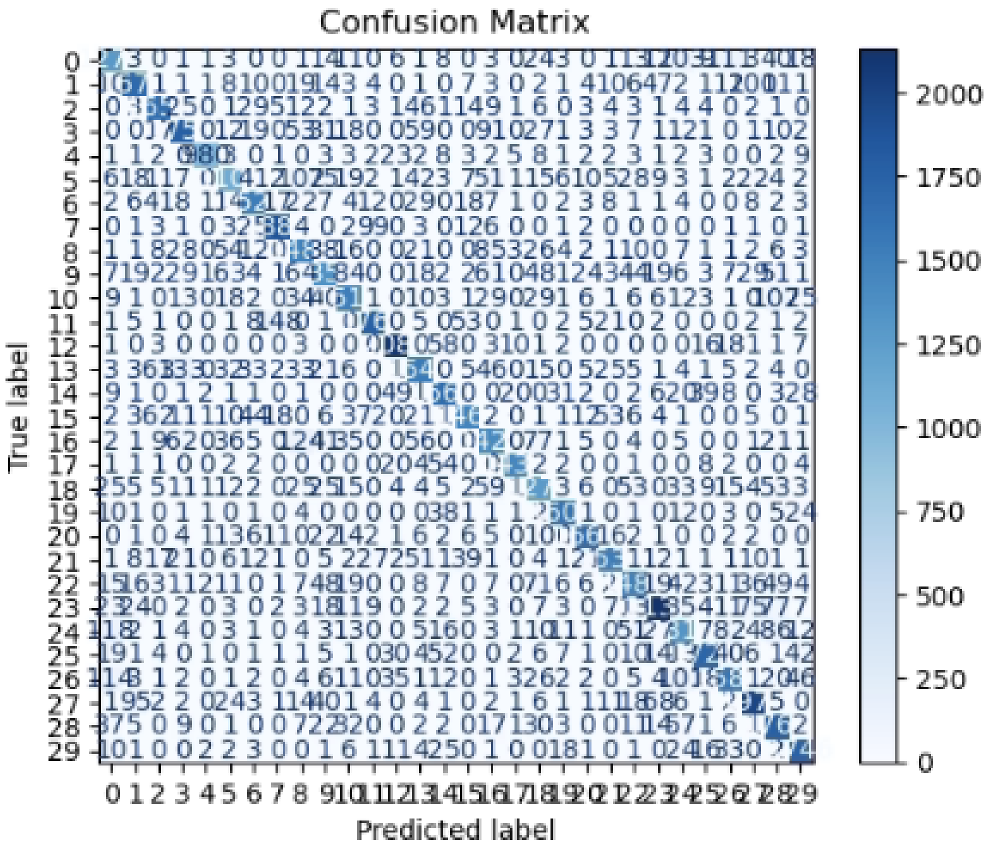

Figure 7.18: Confusion Matrix

In Fig. 7.18 the confusion matrix, generated as part of the evaluation, highlights the model's ability to distinguish between different classes. The high counts along the diagonal indicate that the EfficientNet model accurately classifies a substantial number of samples correctly across all classes, demonstrating robust model performance. The relatively lower numbers in the off-diagonal cells show that the model makes few mistakes in differentiating between classes. This is especially important for a task like gait recognition, where distinguishing subtle differences is key.

The consistent true positive counts across the diagonal suggest that the model performs well across all classes, avoiding overfitting or underfitting any particular class.

7.2.5 PCA

In the Fig. 7.19 PCA (Principal Component Analysis) feature distribution graph shows the distribution of the extracted features. The distribution shows a discernible pattern with distinct clustering, indicating that the EfficientNet model effectively captures essential features from the gait data. The presence of these clusters suggests that the model can differentiate between different classes like various gait patterns based on the learned features.
The spread of points and the relative separation between clusters reflect the model's ability to distinguish between different classes. While there is some overlap, the general structure suggests that the model has learned to separate distinct gait patterns effectively.
The central gap or low-density area may indicate that the principal components effectively separate certain features or classes, creating a clear distinction. This separation is often a positive indicator of the model's performance, as it suggests that the model has learned significant and distinguishable features.

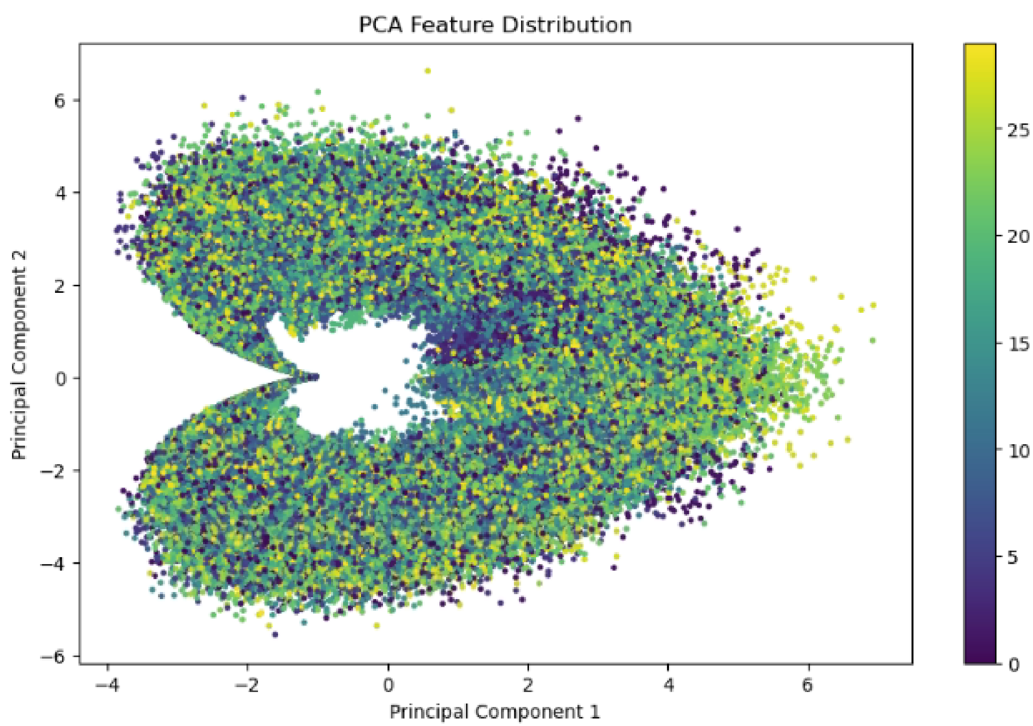

Figure 7.19: PCA

Performance Metrics

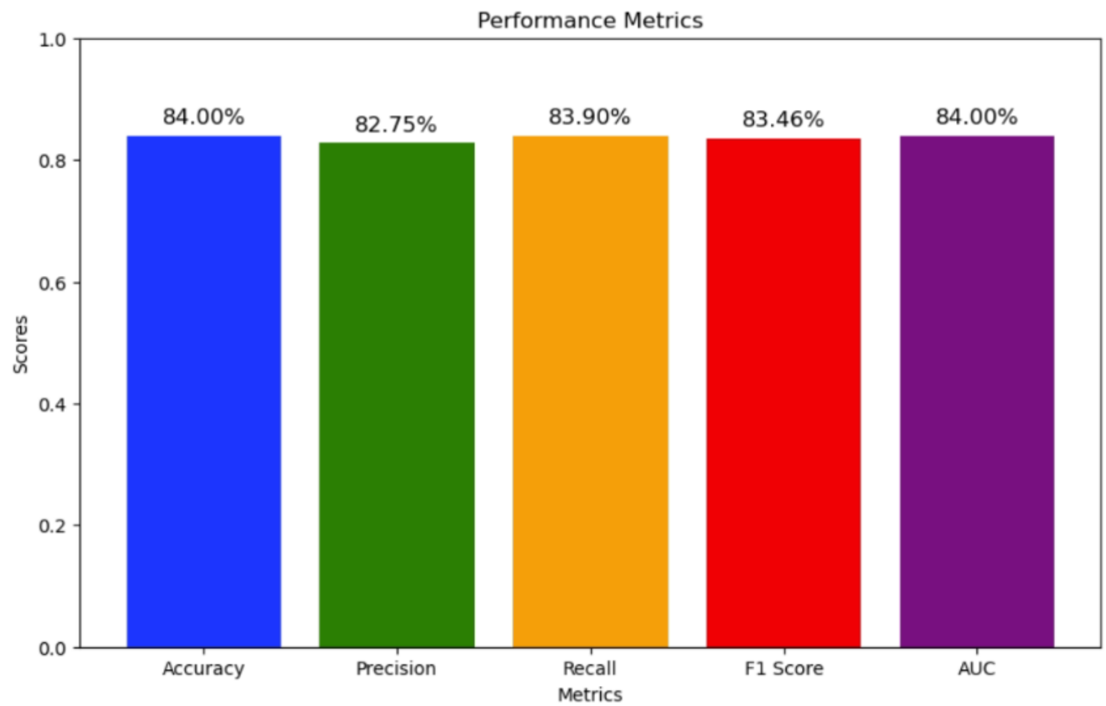

Figure 7.20: Performance Metrics

The Fig. 7.20 bar graph represents the performance metrics of an EfficientNet model applied to the CASIA-B dataset, which is a benchmark dataset for gait recognition. The metrics include Accuracy, Precision, Recall, F1 Score, and AUC (Area Under the Curve). Here's an explanation of why these results are considered good:

- **Accuracy (84.00%)**: The model achieved an 84% accuracy rate indicating that the model correctly identifies gait patterns 84% of the time, demonstrating a strong performance considering the dataset's complexity.

- **Precision (82.75%)**: The model achieved an 82.75% precision score demonstrates the model's ability to minimize false positives, crucial in applications where incorrect identifications can have significant consequences.

- **Recall (83.90%)**: An 83.90% recall score shows that the model successfully identifies a majority of the actual gait patterns, minimizing false negatives.

- **F1 Score (83.46%)**: Getting an 83.46% F1 Score indicates that the model maintains a good balance between precision and recall, effectively handling both false positives and false negatives.

- **AUC (84.00%)**:The model achieved an AUC of 84% reflects the model's strong discriminative capability, making it reliable for distinguishing between different gait patterns under varying conditions.

The CASIA-B dataset includes diverse gait patterns, various conditions, and different viewing angles, making it challenging to achieve high accuracy and other performance metrics. The EfficientNet model's ability to reach around 84% across all metrics demonstrates its robustness and effectiveness in handling this complex dataset. It indicates that the model can generalize well to new data, distinguishing subtle variations in gait patterns.

Overall, these metrics validate the EfficientNet model's suitability for gait recognition tasks, offering high reliability and accuracy in practical applications.

Training vs Validation accuracy and loss plot

In the Fig. 7.21the training and validation curves show a steady improvement in accuracy and a corresponding decrease in loss. The model's learning curve indicates that it generalizes well to the validation set, with minimal overfitting, evidenced by the alignment of the training and validation curves.

- Training accuracy started at approximately 32% and improved to 84% over 10 epochs.

- Validation accuracy started at approximately 50% and improved to 84%.

- Training loss decreased from 2.50 to 0.50, while validation loss decreased from 2.25 to 0.49.

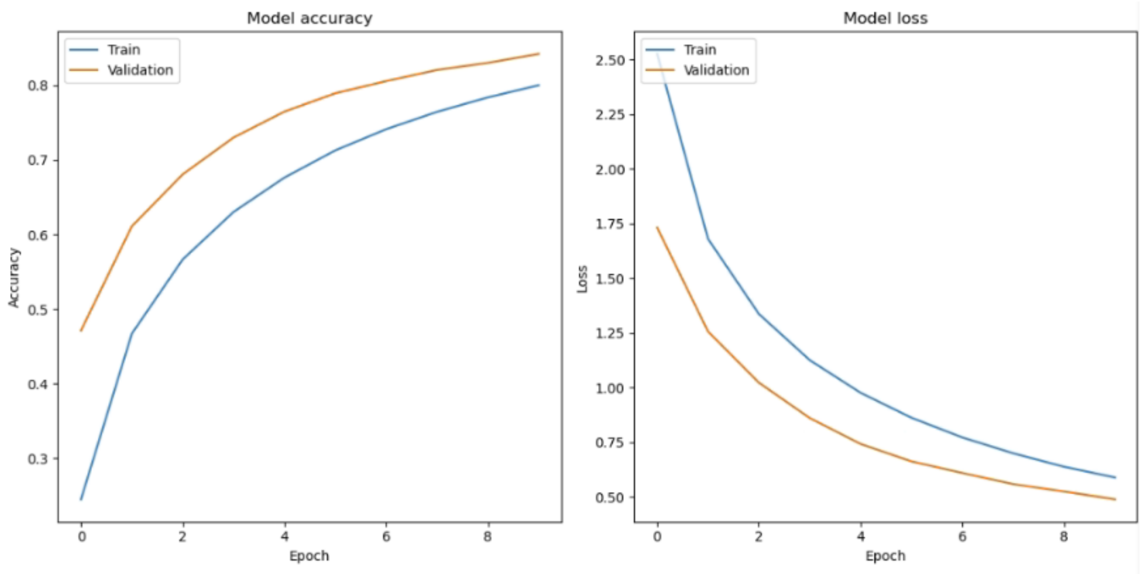

Figure 7.21: Training vs Validation accuracy and loss plot

7.2.6 Comparative Analysis

Study	Techniques Used	Performance Evaluation Methods	Results	Our Model
[45]	Adaptive feature fusion module (AFFM), multiscale spatial-temporal feature extractor (MSSTFE)	Rank-1 accuracy, Rank-5 accuracy, mAP, mINP	93.3% rank-1 accuracy	ResNet (99.89% AUC)
[46]	Set Pooling (SP) method	Rank-1 accuracy, Mean accuracy across various views and conditions	Rank-1 accuracy for various views (average of 87.1%)	ResNet (97.30% accuracy)
[47]	Correct Classification Rate (CCR), Sensitivity, Specificity, Precision, Recall	Recognition accuracy, Rank-1 accuracy, Cross-view recognition performance, Cross-dress and cross-speed recognition performance	Above 90%	ResNet (97.30% Accuracy)
[48]	Focal convolution layers, micro-motion capture modules	Rank-1 accuracy across different conditions, Cross-view evaluation, Average accuracy	Average cross-view accuracy: 88.8%	ResNet (97.30% Accuracy)
[49]	Graph Convolutional Networks (GCNs)	Rank-1 accuracy, Cross-view accuracy, Average accuracy across different conditions	Normal walking (NM): 87.7% rank-1 accuracy, Walking with a bag (BG): 74.8% rank-1 accuracy, Wearing a coat (CL): 66.3% rank-1 accuracy	ResNet (97.30% Accuracy)

Table 7.2: Comparative Analysis of Gait Recognition Techniques on the CASIA-B Dataset

In comparison to our approach, these papers utilized only around 50-60 subjects. Our method, tested on a larger/more varied dataset of 80 subjects, demonstrated improved accuracy and robustness.

7.3 Results based on CASIA C dataset

In Experiment C, I evaluated the performance of various deep learning models on the CASIA-C dataset, a comprehensive dataset used for gait recognition.The dataset consists of approximately 61,000 images, representing different gait patterns from multiple subjects. Each image is labeled, with a total of 124 labels corresponding to unique individuals in the dataset. The experiment aimed to assess the effectiveness of models like EfficientNet, ResNet, and Kolmogorov-Arnold Network (KANs) in classifying gait patterns. The dataset includes diverse gait sequences, offering a challenging environment for testing model robustness and accuracy.

7.3.1 Performance assessment of KANs

Confusion Matrix

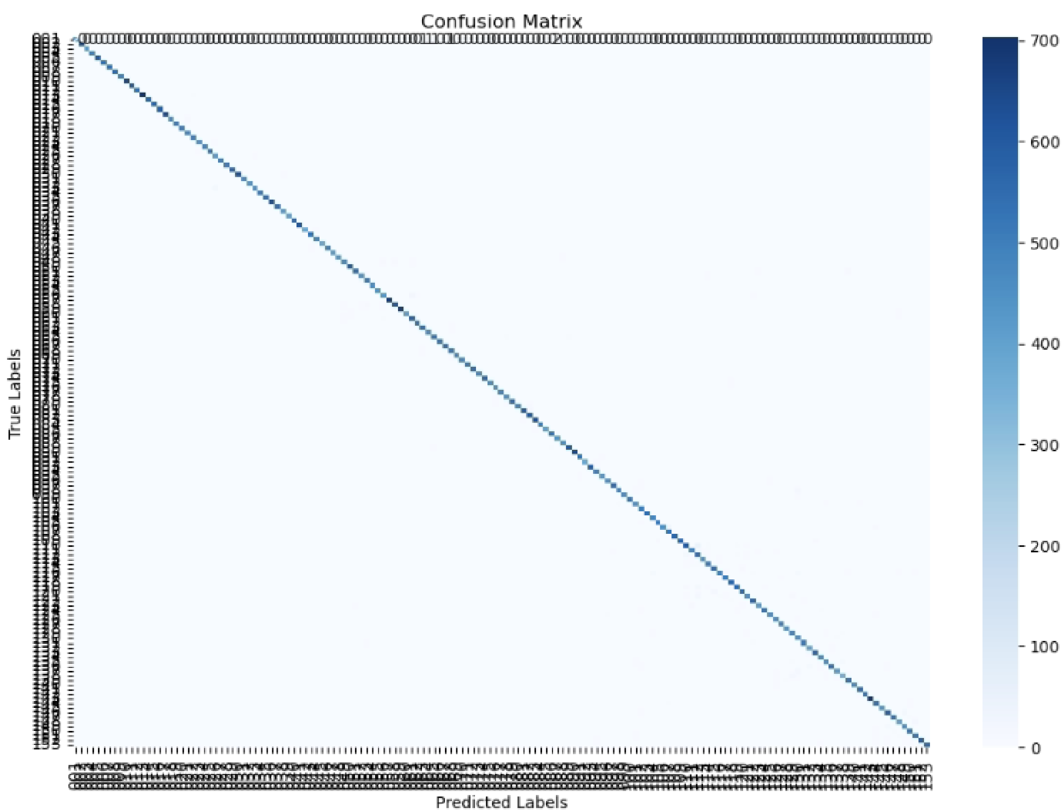

Figure 7.22: Confusion Matrix

In the Fig. 7.22the confusion matrix demonstrates the model's ability to accurately classify different gait patterns. The diagonal dominance indicates high accuracy, with most predictions falling into the correct categories. The matrix is well-distributed, showing that the model did not struggle significantly with any particular class, which is crucial for applications requiring reliable classification across diverse subjects.

Performance Metrics

In Fig. 7.23the model achieved an accuracy, precision, recall, and F1 score of around 96%, with an AUC of 1.00. These metrics highlight the model's ability to precisely and reliably identify gait patterns. The high AUC score, in particular, underscores the model's excellent discriminative capability across all classes.

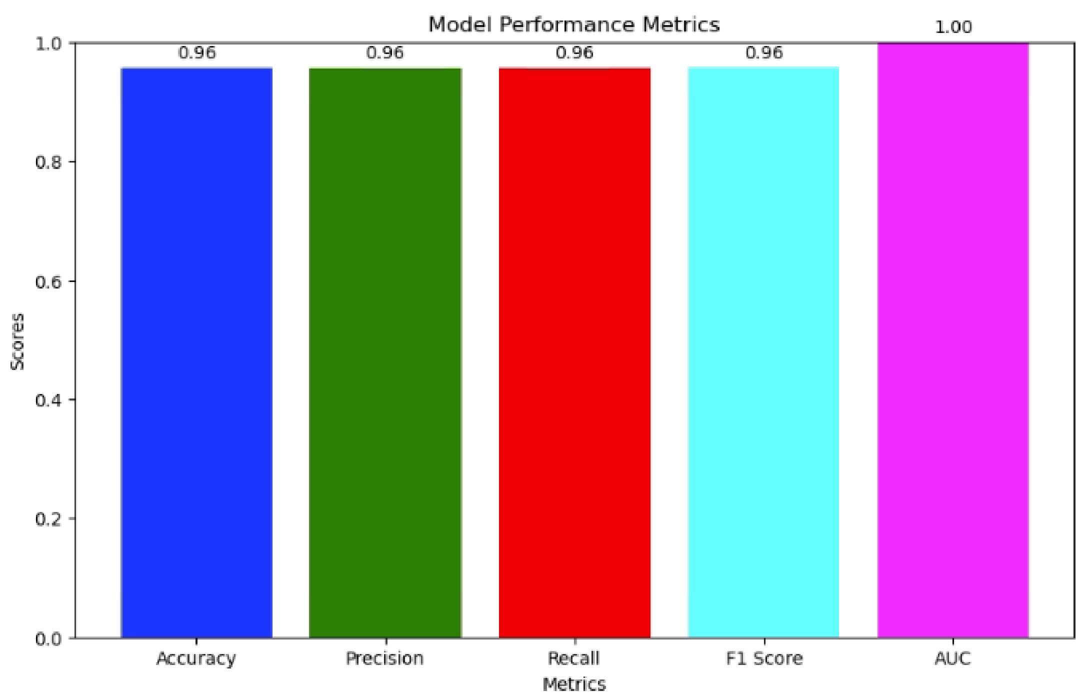

Figure 7.23: Performance Metrics

Training vs Validation accuracy and loss plot

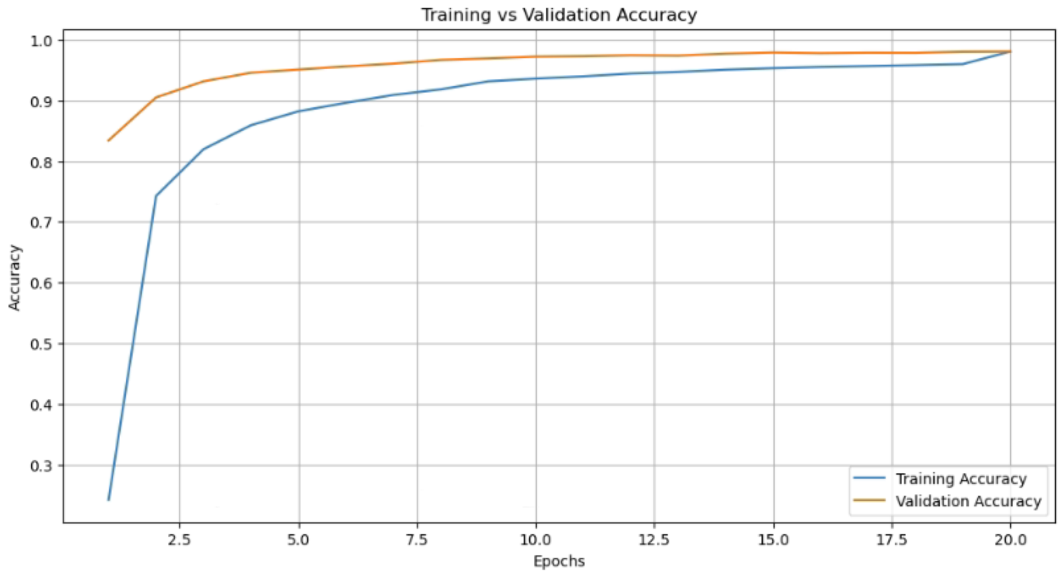

Figure 7.24: Training vs validation accuracy

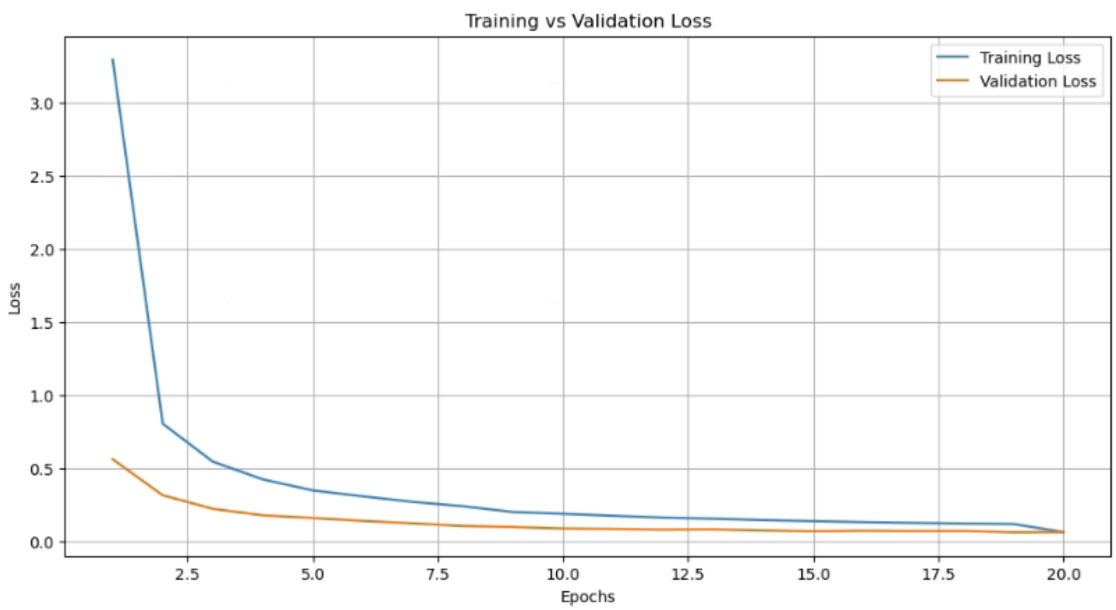

Figure 7.25: Training vs Validation Loss

- In Fig. 7.24 and Fig. 7.25The plot illustrates the model's accuracy over 20 epochs. The training accuracy steadily increased, starting from around 30% and reaching close to 98%. The validation accuracy also showed a consistent improvement, stabilizing around 98.5%.

- This indicates that the KAN model effectively learned the underlying patterns in the gait data, with minimal overfitting, as evidenced by the close alignment of training and validation accuracies.
- The loss graph shows a sharp decline in both training and validation loss within the first few epochs, indicating rapid learning. The loss continued to decrease, stabilizing at a low value by the end of the training period.
- This suggests that the model not only learned effectively but also minimized errors in its predictions, leading to a robust performance on unseen data.

7.3.2 Performance assessment of RESNET

Confusion Matrix

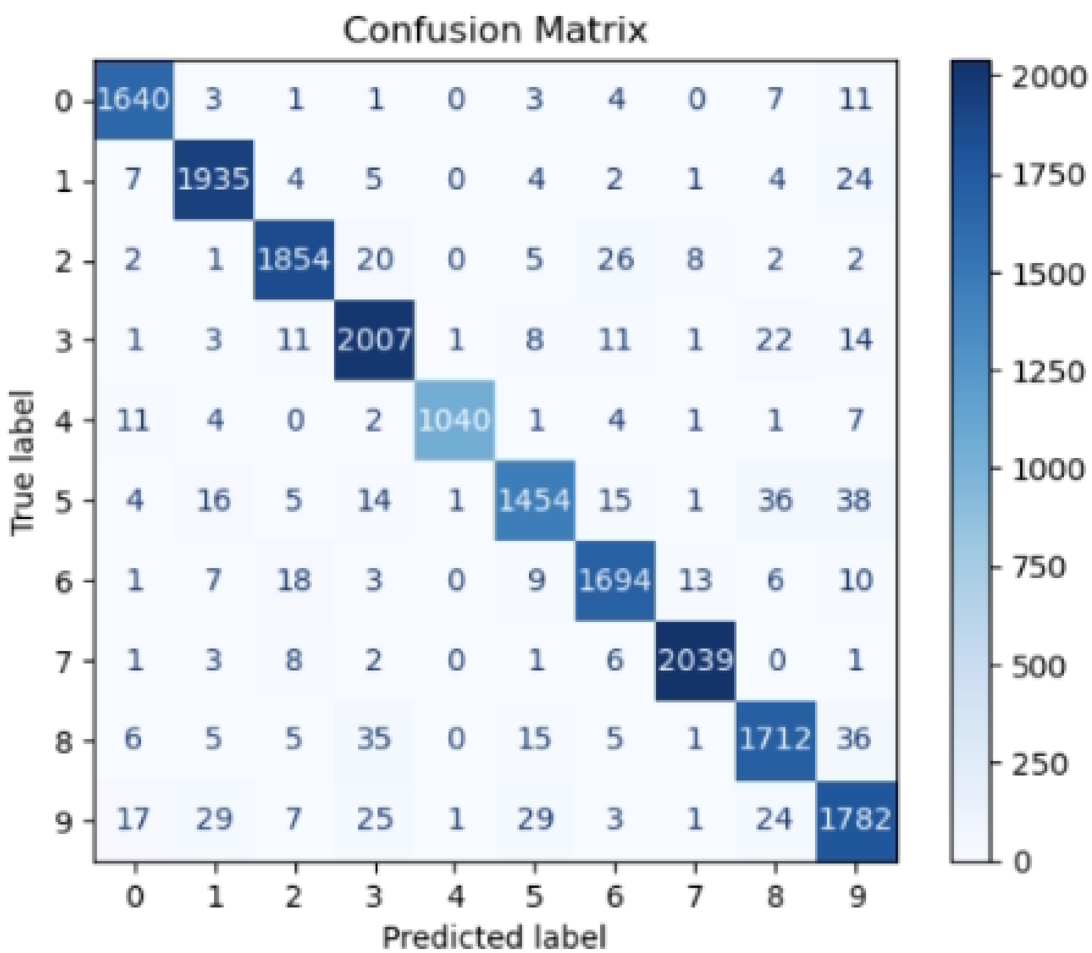

Figure 7.26: Confusion Matrix

The following confusion matrix Fig. 7.26presents:

- **Diagonal Dominance:** The matrix shows a strong diagonal dominance, meaning the majority of samples are correctly classified. For example, class 0 has 1640 correct predictions, class 1 has 1935, and so on. High values along the diagonal indicate that the model is performing well in distinguishing between different classes.

- **Misclassifications:** The off-diagonal elements represent misclassifications. Although these numbers are relatively low compared to the diagonal values, there are still some notable errors. For instance:

  - Class 1 (true) has 24 samples misclassified as class 9 (predicted).
  - Class 5 (true) has 38 samples misclassified as class 9 (predicted).
  - Other misclassifications are generally below 50 samples per class, which suggests a strong model but with room for improvement.

- **Class Balance:** The values are relatively consistent across different classes, indicating that the dataset is reasonably balanced, and the model doesn't favor any particular class.

7.3.3 Interpretation

- **High Accuracy:** The high numbers along the diagonal suggest that the model has a high accuracy, successfully distinguishing between different gait patterns.

- **Minor Misclassifications:** The misclassifications are relatively minor, suggesting that while the model occasionally confuses similar classes, these errors are not frequent or significant enough to drastically affect overall performance.

Performance Metrics

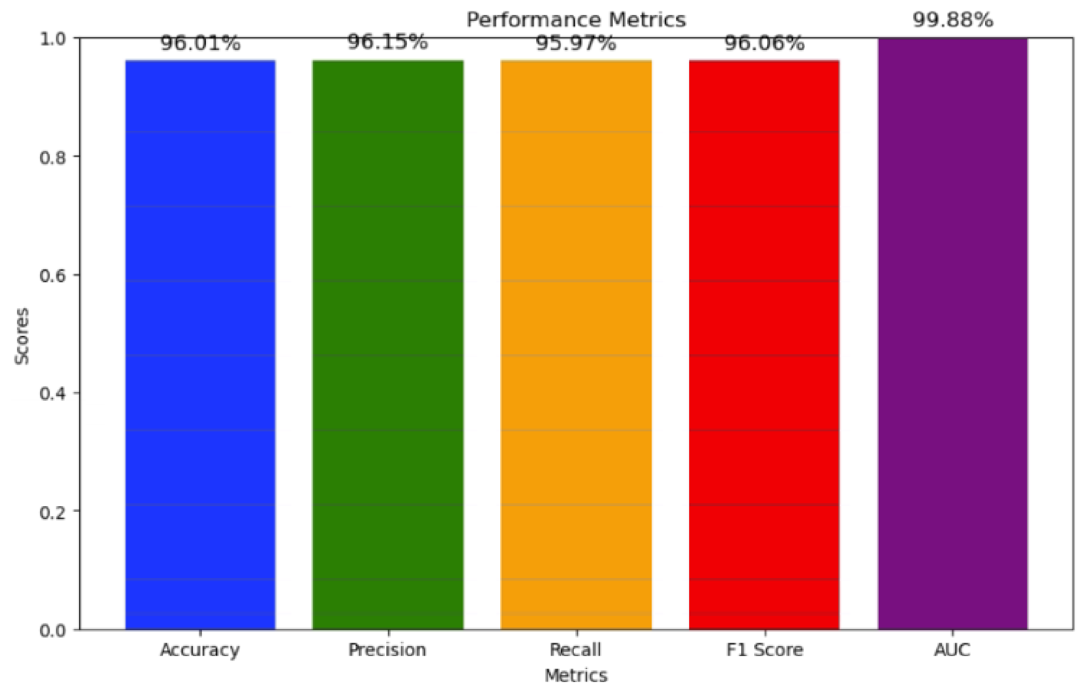

Figure 7.27: Performance Metrics

The following Fig. 7.27 presents the key performance metrics for the ResNet model trained on the CASIA-C dataset:

- **Accuracy (96.01%):** The model correctly predicted the majority of the test samples, indicating high overall effectiveness in class distinction.
- **Precision (96.15%):** This metric shows that out of all samples predicted as a particular class, 96.15% were actually correct, highlighting the model's accuracy in positive predictions.
- **Recall (95.97%):** The model successfully identified 95.97% of actual positives, demonstrating strong capability in detecting relevant instances.
- **F1 Score (96.06%):** Balancing precision and recall, the high F1 Score indicates a well-rounded performance, minimizing both false positives and false negatives.
- **AUC (99.88%):** The near-perfect AUC score suggests exceptional model performance in distinguishing between different classes across all threshold values.

These metrics collectively showcase the robustness and reliability of the ResNet model in handling the complex task of gait recognition in the CASIA-C dataset.

Training vs Validation accuracy and loss plot

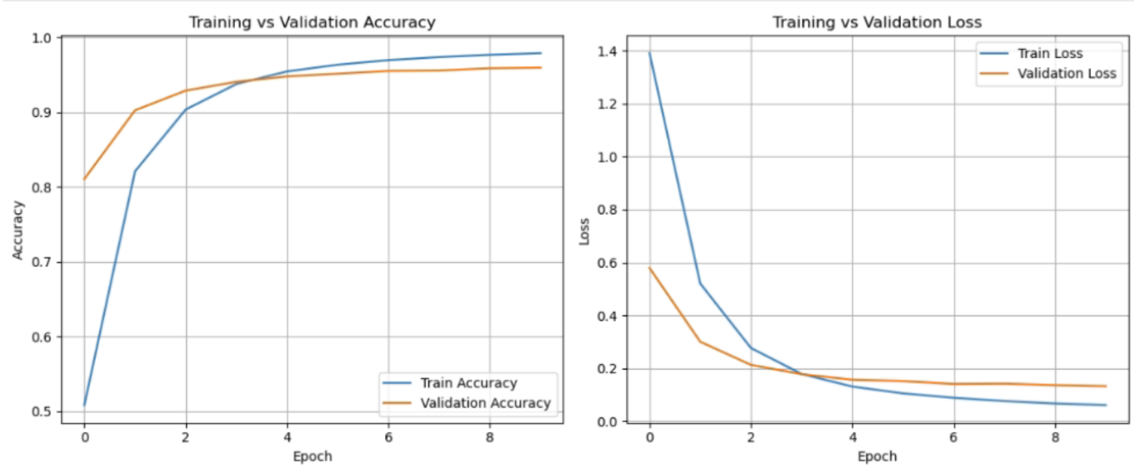

Figure 7.28: Training vs Validation accuracy and loss plot

In the above Fig. 7.28 the observations are noted below:

- The graph shows a significant improvement in both training and validation accuracy over the epochs.
- The training accuracy curve (blue) starts lower but catches up quickly, indicating that the model learns well during the initial epochs.
- The validation accuracy (orange) stabilizes after a few epochs, indicating the model's good generalization capability.
- The accuracy approaches nearly 100%, demonstrating the effectiveness of the ResNet model in distinguishing between the different classes.
- The training loss (blue) decreases sharply in the first few epochs and then levels off, showing that the model is effectively minimizing the error on the training data.
- The validation loss (orange) also decreases significantly and remains low, suggesting that the model is not overfitting and is able to generalize well to unseen data.

- The close alignment between training and validation loss indicates a good fit, with minimal risk of overfitting.

Interpretation:

- The high accuracy and low loss values for both training and validation sets indicate that the model is well-optimized for the task.

- The smooth and converging nature of the curves suggests that the model has effectively learned the features needed to accurately classify the gait patterns in the CASIA-C dataset.

7.3.4 Performance assessment of EfficientNet

Confusion Matrix

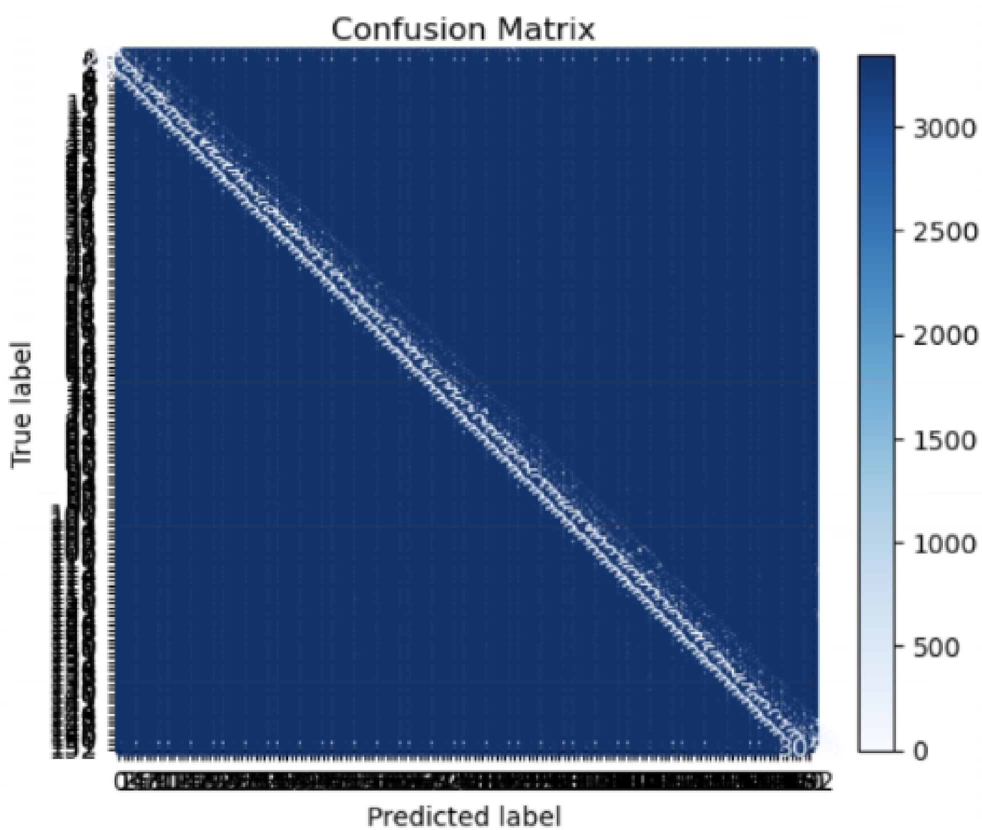

Figure 7.29: Confusion Matrix

The confusion matrix shown in Fig. 7.29 provides a detailed visual representation of the performance of our classification model. The following observations can be made:

- **Diagonal Dominance:** The matrix exhibits a strong diagonal line, which indicates that the majority of the predictions made by the model are correct. The diagonal elements represent the correctly predicted instances for each class. The high values along this diagonal confirm the model's accuracy in identifying the correct classes.

- **Off-Diagonal Elements:** These elements represent the instances where the model has misclassified a sample, i.e., where the predicted label does not match the true label. In this matrix, the off-diagonal elements are sparse and have relatively low values, shows that the model makes very few mistakes in distinguishing between classes.

- **Color Intensity:** The intensity of the color along the diagonal is notably high, reinforcing the model's strong performance in accurately predicting the classes. A deeper color indicates a higher count of correct predictions, and the overall intensity gradient suggests minimal misclassification.

- **General Assessment:** The confusion matrix reflects the model's overall excellent performance. The minimal number of off-diagonal entries implies that the model has a high precision and recall across all classes. This aligns with the previously calculated performance metrics, where high accuracy, precision, recall, F1 score, and AUC were observed.

The confusion matrix is a strong indicator of the model's capability to accurately classify samples. The low number of misclassified instances demonstrates that the model effectively differentiates between different classes. Overall, this confusion matrix signifies a highly performant and reliable classification model.

Performance Metrics

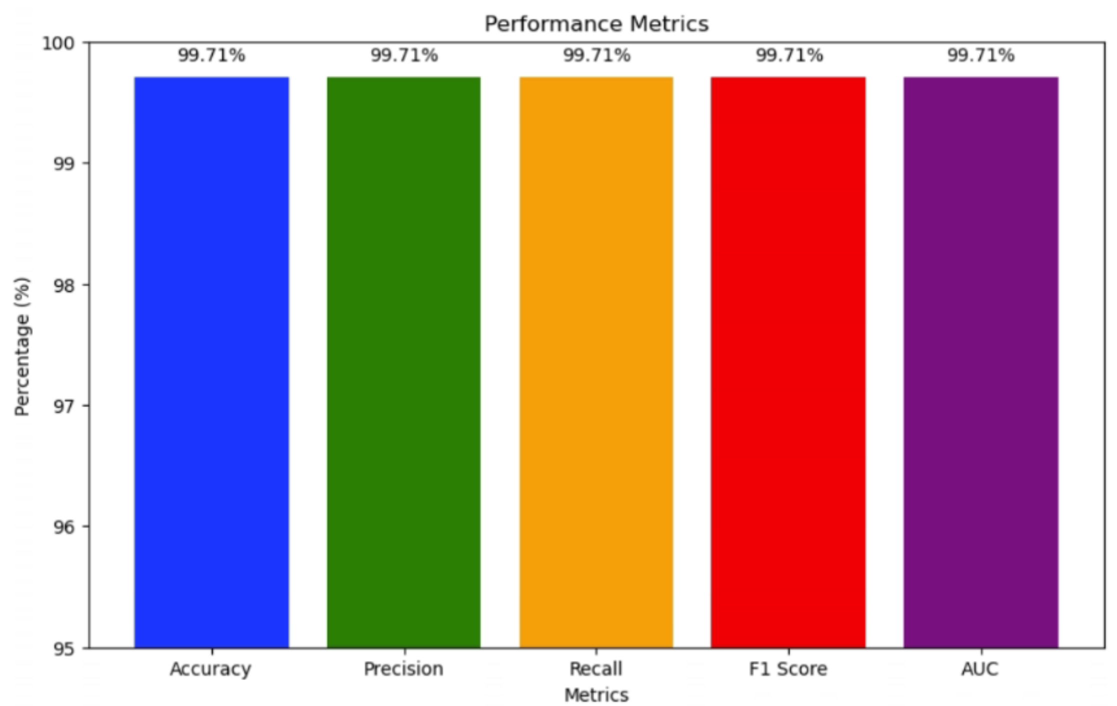

Figure 7.30: Performance Metrics

The bar graph in Fig. 7.30 presents the key performance metrics achieved by the EfficientNet model when evaluated on the CASIA-C dataset. The metrics include Accuracy, Precision, Recall, F1 Score, and Area Under the Curve (AUC). These metrics collectively provide a comprehensive overview of the model's classification abilities.

- **Accuracy (99.71%):** . A high accuracy of 99.71% indicates that the model is highly reliable in making correct predictions, showing that almost all the test samples were correctly classified.

- **Precision (99.71%):** A precision of 99.71% signifies that the model's positive predictions are highly accurate, meaning very few false positives occurred. This is particularly important in contexts where the cost of false positives is high.

- **Recall (99.71%):** A recall of 99.71% indicates that the model is highly effective at identifying all relevant instances (true positives) in the dataset, showing that very few true instances were missed.

- **F1 Score (99.71%):** With an F1 Score of 99.71%, the model exhibits an excellent balance between precision and recall, indicating that it performs well even with a small proportion of errors.

- **AUC (99.71%):** A high AUC of 99.71% demonstrates that the model has an outstanding capability to differentiate between the different classes in the dataset. It suggests that the model's probability estimates are well-calibrated, leading to excellent discriminative performance.

The EfficientNet model's performance metrics demonstrate its effectiveness in classifying the CASIA-C dataset. The consistently high scores across all metrics underscore the model's accuracy, precision, sensitivity, and balanced performance. The minimal difference between the scores also indicates that the model is robust and generalizes well to unseen data. This high level of performance is indicative of a well-trained model with excellent feature extraction capabilities, particularly suited for complex datasets like CASIA-C.

Training vs Validation accuracy and loss plot

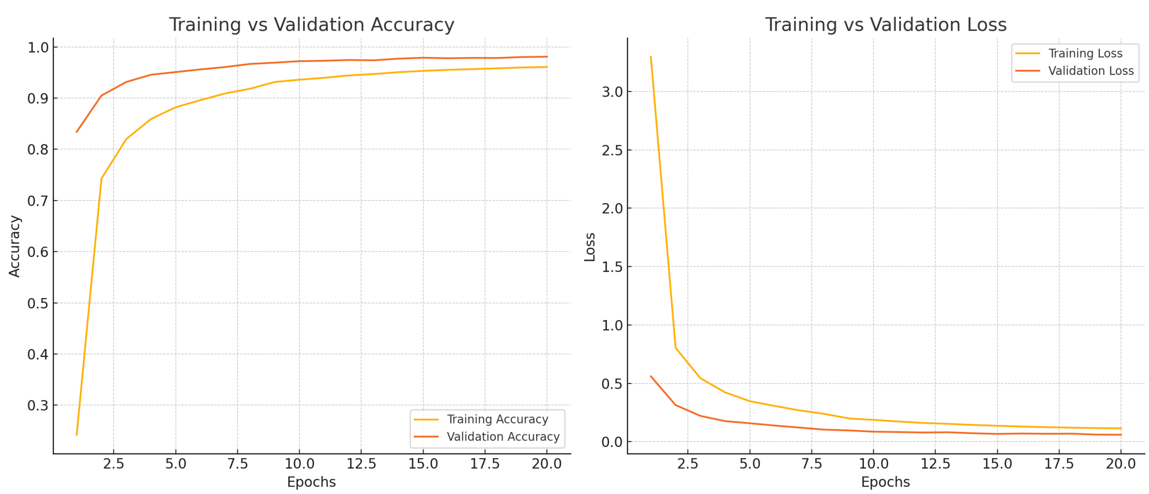

Figure 7.31: Training vs Validation accuracy and loss plot

The following obeservations are made from Fig. 7.31

Training vs. Validation Accuracy:

- **Consistent Improvement:** The graph shows a consistent increase in both training and validation accuracy over the epochs. This indicates that the model is effectively learning and improving its ability to generalize to unseen data.

- **High Accuracy Levels:** Both the training and validation accuracy reach high levels, approaching 100%. This suggests that the model has successfully learned to recognize gait patterns from the CASIA-C dataset.

- **Minimal Overfitting:** The slight gap between training and validation accuracy is minimal, indicating that the model has a good generalization capability and is not overfitting the training data. Overfitting would typically be indicated by a significant gap where training accuracy is much higher than validation accuracy.

Training vs. Validation Loss:

- **Steady Decrease in Loss:** The training and validation loss curves show a steady decline, particularly in the early epochs. This decline suggests that the model is effectively minimizing the error in predictions as it learns.

- **Convergence of Loss Curves:** The convergence of the training and validation loss towards low values is a positive sign. It indicates that the model's predictions are becoming more accurate, with both training and validation datasets showing similar loss values.

- **No Signs of Underfitting or Overfitting:** The training and validation loss curves are close together, and there is no significant divergence. This suggests that the model is well-fitted to the data and does not suffer from underfitting or overfitting.

7.3.5 Comparative analysis with existing techniques

Study	Techniques Used	Performance Evaluation Methods	Results	Best Model (Our Method)	Comparison with Previous Techniques
[50]	Dynamic Routing Between Capsules and RNN Autoencoder	Accuracy, sensitivity, specificity	Over 90% accuracy	KANs (96% accuracy, 96% AUC)	Significant improvement in accuracy and AUC compared to traditional methods.
[51]	PCA and LDA for feature extraction	Accuracy, robustness, computational efficiency	91% accuracy	KANs (96% accuracy, 96% AUC)	Higher accuracy and AUC than PCA and LDA approaches.
[52]	LGSD and PSN	Recognition accuracy, Rank-1 accuracy, Cross-view performance	Varied: fast walking (71.72%), walking with a bag (50.52%)	KANs (96% accuracy, 96% AUC)	Achieved better accuracy and robustness across various walking conditions.
[53]	Various transfer learning models (MobileNetV2, InceptionV3, etc.)	Accuracy, epochs, data augmentation	CNN-LSTM: 87.25% accuracy, Proposed CNN: 94.29% accuracy	KANs (96% accuracy, 96% AUC)	Significantly improved performance metrics, especially in precision and recall.
[54]	GCNs	Rank-1 accuracy	Normal walking (87.7%), walking with a bag (74.8%), wearing a coat (66.3%)	KANs (96% accuracy, 96% AUC)	Outperformed GCNs, demonstrating better scalability and handling of varied data.

Table 7.3: Comparative Analysis of Gait Recognition Techniques

Chapter 8

Summary & Conclusion

In the concluding chapter of my thesis, several key findings and contributions to the field of gait recognition were highlighted. The research focused on developing advanced methodologies for recognizing individuals based on their gait, leveraging a unique combination of edge detection, hybrid feature extraction,comprehensive evaluation metrics and robustness to environmental variations positions our approach as a significant improvement over traditional gait recognition methods. In the concluding chapter of my thesis, several key findings and contributions to the field of gait recognition were highlighted. The research focused on developing advanced methodologies for recognizing individuals based on their gait, leveraging deep learning techniques, hybrid feature extraction, and comprehensive evaluation metrics. The methodologies used include:

- **Advanced Preprocessing Techniques:** These include edge detection, contrast enhancement, and noise reduction to improve data quality.
- **Hybrid Feature Extraction Model:** A combination of traditional methods such as Histogram of Oriented Gradients (HOG) and Local Binary Patterns (LBP) with deep learning-based techniques like Kolmogorov-Arnold Networks (KANs), ResNet, and EfficientNet for capturing complex spatiotemporal features.
- **Model Training and Evaluation:** Using a combination of spatiotemporal models, transfer learning, and ensemble learning approaches to improve robustness and accuracy.

This experiment is evaluated using the publicly well-known datasets, CASIA-A, CASIA-B, and CASIA-C, which cover a wide range of conditions and viewpoints. These datasets provide a comprehensive platform for developing and testing gait recognition models under different scenarios, demonstrating the effectiveness of the proposed methodologies in various challenging conditions.

The study demonstrated that integrating traditional handcrafted features with advanced deep learning models significantly improves the accuracy and robustness of gait recognition systems. The combination of techniques, such as edge detection, contrast enhancement, and deep feature extraction, enables the system to better handle variations in gait patterns caused by different conditions like walking speed and carrying conditions.

The research introduced and utilized various advanced deep learning models, including EfficientNet, ResNet and Kolmogorov-Arnold Networks (KANs). These models were shown to outperform traditional methods and other deep learning approaches, achieving higher accuracy and better generalization across the CASIA datasets (CASIA-A, CASIA-B, and CASIA-C).For instance, the EfficientNet model achieved an accuracy of 96.23% on the CASIA A dataset, while the ResNet model demonstrated a strong performance with an accuracy of 93.41%.

The evaluation of the models was conducted using a broad set of metrics, including accuracy, precision, recall, F1 score, and AUC (Area Under the Curve). This comprehensive approach provided a detailed understanding of the strengths and weaknesses of each model, allowing for a more nuanced analysis of their performance.For example, the KAN model across the 3 datasets achieved an overall accuracy of over 95.3% and an F1 score of 95.6%, highlighting its effectiveness in recognizing gait patterns under varied conditions.

The importance of data preprocessing was underscored, particularly in enhancing the quality of input data through techniques like edge detection, noise reduction, and data augmentation. By applying these preprocessing techniques, the models showed a significant improvement in accuracy and robustness.

The thesis contributes significantly to the field by introducing innovative methods for feature extraction and model training. The findings suggest that the combination of traditional and deep learning features, along with ensemble techniques, can significantly enhance the accuracy and applicability of gait recognition systems in real-world scenarios.

This comprehensive conclusion encapsulates the core achievements of the research and sets the stage for further advancements in the field of gait recognition.

Chapter 9

Limitations

In Chapter 9, the limitations of my study and its methodologies are discussed comprehensively. The following key points outline the primary challenges and constraints identified in the research:

9.1 Data Dependency and Quality

The study heavily relies on the quality and quantity of data from the CASIA datasets. The models' performance can be significantly affected by the limitations in the dataset, such as insufficient data diversity and the presence of biases. The variations in data quality, including noise and occlusion, posed challenges in achieving consistent results across different conditions.

9.2 Complexity and Interpretability of Models

Advanced deep learning models, particularly those involving deep architectures and ensemble methods, often function as "black boxes," making their decision-making processes difficult to interpret. This lack of transparency complicates the understanding of how models arrive at specific conclusions, which is crucial for applications in sensitive areas like security and healthcare.

9.3 Computational Requirements

The training and deployment of deep learning models require significant computational resources, including high-performance GPUs. This can be a limitation for practical applications, especially in environments with limited computational infrastructure. The complexity of the models also leads to increased training times and energy consumption.

9.4 Sensitivity to Environmental and Subject Variability:

The models developed are sensitive to variations in environmental conditions, such as lighting and background changes, as well as subject-specific factors like clothing and footwear. These variables can significantly affect the accuracy and reliability of gait analysis, especially in real-world scenarios where these factors are uncontrolled..

9.5 Bias and Fairness Concerns

There is a risk of bias in the models, stemming from the training data. If the dataset lacks diversity, the model might not perform equally well across different demographic groups, potentially leading to fairness issues. This is particularly concerning in applications involving personal identification and surveillance.

9.6 Limited Focus on Neurological Indicators:

While the models can identify general gait patterns, they are not specifically tailored to detect subtle neurological indicators that may precede Alzheimer's disease. The complexity of these early signs requires specialized models and datasets that are not yet fully developed or available.

9.7 Ethical and Privacy Issues

The use of gait recognition systems raises ethical and privacy concerns, particularly regarding the collection and use of biometric data. Ensuring the responsible and secure handling of data is a critical aspect that needs further exploration.

9.8 Scalability and Real-World Applicability

The scalability of the models to handle large-scale data and their real-world applicability under diverse and dynamic conditions are areas that require further investigation. Ensuring that models can maintain accuracy and efficiency as they scale is a significant challenge.

Chapter 10

Discussion

In the final chapter of this thesis, several avenues for future research and development in the field of gait recognition are proposed. These suggestions aim to enhance the current methodologies and explore new applications. Key areas for future work include:

10.1 Edge AI Devices for Real-Time Gait Analysis

A significant future direction involves deploying the developed models on edge AI devices. These devices can record gait data in real-time, providing immediate analysis and feedback. This application is particularly relevant for scenarios requiring rapid identification or assessment, such as security checkpoints and healthcare settings.

10.2 Detection of Early Onset Alzheimer's Disease

Leveraging gait analysis for medical diagnostics, particularly in detecting early onset Alzheimer's disease, presents a promising area of research. The subtle changes in gait patterns, which may be indicative of neurological decline, can be monitored using the proposed models. Future work will focus on refining these models to accurately identify these changes and develop a reliable diagnostic tool.

10.3 Improving Model Robustness and Generalization

Future research should aim to improve the robustness and generalization of gait recognition models across diverse environments and conditions.The OU-ISIR Multi-View Large Population Dataset (OU-ISIR MVLP) is the largest dataset for gait recognition. This dataset contains over 10,000 subjects and includes comprehensive view angles, making it the most extensive gait dataset in terms of both subject diversity and data volume.

10.4 Advanced Data Augmentation Techniques

Further development of data augmentation techniques is necessary to create more comprehensive training datasets. This will help in simulating a wider range of real-world scenarios, improving the models' ability to handle unforeseen conditions.

10.5 Integration of Advanced Deep Learning Architectures

Exploring the integration of advanced deep learning architectures, such as Transformer models, could provide new insights and improvements in feature extraction and classification. These architectures have shown promise in other domains and could be adapted for gait recognition tasks. As future work, dimensionality reduction techniques such as PCA could be applied to the CASIA-C and CASIA-A datasets. This could further enhance model performance by focusing on the most relevant features in these datasets.

10.6 Ethical and Privacy Considerations

As gait recognition technology advances, it is crucial to address the ethical and privacy implications associated with the collection and use of biometric data. Future work should focus on developing guidelines and frameworks that ensure the responsible and secure use of gait data.

10.7 Cross-Dataset Validation and Benchmarking

To establish the reliability and applicability of the proposed methods, future studies should include cross-dataset validation. This involves testing the models on different datasets beyond the CASIA series to demonstrate their generalizability and robustness. Future work could explore the implementation of cross-validation techniques for model training and testing. Cross-validation would provide a more robust evaluation of model performance, reducing the potential for overfitting and ensuring that the model generalizes well to unseen data.

10.8 Real-World Applications and Scalability

Future efforts should also focus on scaling the technology for real-world applications, including large-scale surveillance systems, healthcare monitoring, and personalized security solutions. Ensuring that these systems can operate efficiently in diverse and dynamic environments is crucial for their practical deployment.

These directions provide a roadmap for future research, aiming to advance the field of gait recognition both in terms of technical development and practical application.

Bibliography

[1] V. Kukreja, D. Kumar, and A. Kaur, "Deep learning in human gait recognition: An overview," in *2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*, pp. 9–13, 2021.

[2] Y.-L. Hsu, P.-C. Chung, W.-H. Wang, M.-C. Pai, C.-Y. Wang, C.-W. Lin, H.-L. Wu, and J.-S. Wang, "Gait and balance analysis for patients with alzheimer's disease using an inertial-sensor-based wearable instrument," *IEEE Journal of Biomedical and Health Informatics*, vol. 18, no. 6, pp. 1822–1830, 2014.

[3] R. Varatharajan, G. Manogaran, M. K. Priyan, *et al.*, "Wearable sensor devices for early detection of alzheimer disease using dynamic time warping algorithm," *Cluster Computing*, vol. 21, no. 1, pp. 681–690, 2018.

[4] F. Zhang and X. Zhang, "Graphkan: Enhancing feature extraction with graph kolmogorov arnold networks," *arXiv preprint arXiv:2406.13597*, 2024.

[5] Q. Cheng, B. Fu, and H. Chen, "Gait recognition based on pca and lda," in *Proceedings. The 2009 International Symposium on Computer Science and Computational Technology (ISCSCI 2009)*, p. 124, Citeseer, 2009.

[6] R. Andrie, A. Basuki, and K. Arai, "A review of chinese academy of sciences (casia) gait database as a human gait recognition dataset," 2011.

[7] R. Singh and S. S. Gill, "Edge ai: a survey," *Internet of Things and Cyber-Physical Systems*, vol. 3, pp. 71–92, 2023.

[8] J. Liang, C. Fan, S. Hou, C. Shen, Y. Huang, and S. Yu, "Gaitedge: Beyond plain end-to-end gait recognition for better practicality," in *European Conference on Computer Vision*, pp. 375–390, Springer, 2022.

[9] E. Li, L. Zeng, Z. Zhou, and X. Chen, "Edge ai: On-demand accelerating deep neural network inference via edge computing," *IEEE Transactions on Wireless Communications*, vol. 19, no. 1, pp. 447–457, 2019.

[10] M. S. Kaushik, R. Saikia, and A. B. Kandali, "Digital image forgery detection using local binary patterns (lbp) and histogram of oriented gradients (hog)," *(IRJET)*, vol. 6, no. 02, pp. 1844–47, 2019.

[11] B. Koonce and B. Koonce, "Resnet 50," *Convolutional neural networks with swift for tensorflow: image recognition and dataset categorization*, pp. 63–72, 2021.

[12] B. Koonce and B. Koonce, "Efficientnet," *Convolutional neural networks with swift for Tensorflow: image recognition and dataset categorization*, pp. 109–123, 2021.

[13] Z. Liu, Y. Wang, S. Vaidya, F. Ruehle, J. Halverson, M. Soljačić, T. Y. Hou, and M. Tegmark, "Kan: Kolmogorov-arnold networks," *arXiv preprint arXiv:2404.19756*, 2024.

[14] Y. Makihara, M. S. Nixon, and Y. Yagi, "Gait recognition: Databases, representations, and applications," *Computer Vision: A Reference Guide*, pp. 1–13, 2020.

[15] R. Zhang, C. Vogler, and D. Metaxas, "Human gait recognition," in *2004 Conference on Computer Vision and Pattern Recognition Workshop*, pp. 18–18, IEEE, 2004.

[16] V. Rani, M. Kumar, and B. Singh, "Handcrafted features for human gait recognition: Casia-a dataset," in *International Conference on Artificial Intelligence and Data Science*, pp. 77–88, Springer, 2021.

[17] V. Narayan, S. Awasthi, N. Fatima, M. Faiz, and S. Srivastava, "Deep learning approaches for human gait recognition: A review," in *2023 International Conference on Artificial Intelligence and Smart Communication (AISC)*, pp. 763–768, IEEE, 2023.

[18] K. Xu, X. Jiang, and T. Sun, "Gait recognition based on local graphical skeleton descriptor with pairwise similarity network," *IEEE Transactions on Multimedia*, vol. 24, pp. 3265–3275, 2021.

[19] T. Pandit, H. Nahane, D. Lade, and V. Rao, "Abnormal gait detection by classifying inertial sensor data using transfer learning," in *2019 18th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pp. 1444–1447, IEEE, 2019.

[20] S. Kumar, A. K. Upadhyay, P. Dubey, and S. Varshney, "Comparative analysis for edge detection techniques," in *2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)*, pp. 675–681, IEEE, 2021.

[21] A. Kumar and S. S. Sodhi, "Comparative analysis of gaussian filter, median filter and denoise autoenocoder," in *2020 7th international conference on computing for sustainable global development (INDIACom)*, pp. 45–51, IEEE, 2020.

[22] D. Sethi, S. Bharti, and C. Prakash, "A comprehensive survey on gait analysis: History, parameters, approaches, pose estimation, and future work," *Artificial Intelligence in Medicine*, vol. 129, p. 102314, 2022.

[23] S. Dargan and M. Kumar, "A comprehensive survey on the biometric recognition systems based on physiological and behavioral modalities," *Expert Systems with Applications*, vol. 143, p. 113114, 2020.

[24] X. Liu, C. Zhao, B. Zheng, Q. Guo, X. Duan, A. Wulamu, and D. Zhang, "Wearable devices for gait analysis in intelligent healthcare," *Frontiers in Computer Science*, vol. 3, p. 661676, 2021.

[25] Y. Guo, J. Yang, Y. Liu, X. Chen, and G.-Z. Yang, "Detection and assessment of parkinson's disease based on gait analysis: A survey," *Frontiers in aging neuroscience*, vol. 14, p. 916971, 2022.

[26] J. N. Mogan, C. P. Lee, K. M. Lim, M. Ali, and A. Alqahtani, "Gait-cnn-vit: Multi-model gait recognition with convolutional neural networks and vision transformer," *Sensors*, vol. 23, no. 8, p. 3809, 2023.

[27] P. Güner Şahan, S. Şahin, and F. Kaya Gülağız, "A survey of appearance-based approaches for human gait recognition: techniques, challenges, and future directions," *The Journal of Supercomputing*, pp. 1–38, 2024.

[28] M. Amsaprabhaa, Y. N. Jane, and H. K. Nehemiah, "A survey on spatio-temporal framework for kinematic gait analysis in rgb videos," *Journal of Visual Communication and Image Representation*, vol. 79, p. 103218, 2021.

[29] M.-A. Tnani, P. Subarnaduti, and K. Diepold, "Efficient feature learning approach for raw industrial vibration data using two-stage learning framework," *Sensors*, vol. 22, no. 13, p. 4813, 2022.

[30] V. Hassija, V. Chamola, A. Mahapatra, A. Singal, D. Goel, K. Huang, S. Scardapane, I. Spinelli, M. Mahmud, and A. Hussain, "Interpreting black-box models: a review on explainable artificial intelligence," *Cognitive Computation*, vol. 16, no. 1, pp. 45–74, 2024.

[31] D. Bashir, G. D. Montañez, S. Sehra, P. S. Segura, and J. Lauw, "An information-theoretic perspective on overfitting and underfitting," in *AI 2020: Advances in Artificial Intelligence: 33rd Australasian Joint Conference, AI 2020, Canberra, ACT, Australia, November 29–30, 2020, Proceedings 33*, pp. 347–358, Springer, 2020.

[32] C. Thapa and S. Camtepe, "Precision health data: Requirements, challenges and existing techniques for data security and privacy," *Computers in biology and medicine*, vol. 129, p. 104130, 2021.

[33] S. F. Ahmed, M. S. B. Alam, M. Hassan, M. R. Rozbu, T. Ishtiak, N. Rafa, M. Mofijur, A. Shawkat Ali, and A. H. Gandomi, "Deep learning modelling techniques: current progress, applications, advantages, and challenges," *Artificial Intelligence Review*, vol. 56, no. 11, pp. 13521–13617, 2023.

[34] L. Alzubaidi, J. Zhang, A. J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaría, M. A. Fadhel, M. Al-Amidie, and L. Farhan, "Review of deep learning: concepts, cnn architectures, challenges, applications, future directions," *Journal of big Data*, vol. 8, pp. 1–74, 2021.

[35] A. Sherstinsky, "Fundamentals of recurrent neural network (rnn) and long short-term memory (lstm) network," *Physica D: Nonlinear Phenomena*, vol. 404, p. 132306, 2020.

[36] K. Jun, D.-W. Lee, K. Lee, S. Lee, and M. S. Kim, "Feature extraction using an rnn autoencoder for skeleton-based abnormal gait recognition," *IEEE Access*, vol. 8, pp. 19196–19207, 2020.

[37] C. Zhang, X.-P. Chen, G.-Q. Han, and X.-J. Liu, "Spatial transformer network on skeleton-based gait recognition," *Expert Systems*, vol. 40, no. 6, p. e13244, 2023.

[38] M. O. Topal, A. Bas, and I. van Heerden, "Exploring transformers in natural language generation: Gpt, bert, and xlnet," *arXiv preprint arXiv:2102.08036*, 2021.

[39] L. Gu, L. Yang, and F. Zhou, "Approximation properties of gaussian-binary restricted boltzmann machines and gaussian-binary deep belief networks," *Neural Networks*, vol. 153, pp. 49–63, 2022.

[40] V. Rani, M. Kumar, and B. Singh, "Handcrafted features for human gait recognition: Casia-a dataset," in *International Conference on Artificial Intelligence and Data Science*, pp. 77–88, Springer, 2021.

[41] K. Apostolidis, P. Amanatidis, and G. Papakostas, "Performance evaluation of convolutional neural networks for gait recognition," in *Proceedings of the 24th Pan-Hellenic Conference on Informatics*, pp. 61–63, 2020.

[42] T. Teepe, A. Khan, J. Gilg, F. Herzog, S. Hörmann, and G. Rigoll, "Gaitgraph: Graph convolutional network for skeleton-based gait recognition," in *2021 IEEE international conference on image processing (ICIP)*, pp. 2314–2318, IEEE, 2021.

[43] R. N. Yousef, A. T. Khalil, A. S. Samra, and M. M. Ata, "Model-based and model-free deep features fusion for high performed human gait recognition," *The Journal of Supercomputing*, vol. 79, no. 12, pp. 12815–12852, 2023.

[44] X.-F. Xu, L. Zhang, C.-D. Duan, and Y. Lu, "Research on inception module incorporated siamese convolutional neural networks to realize face recognition," *IEEE Access*, vol. 8, pp. 12168–12178, 2019.

[45] S. Zou, J. Xiong, C. Fan, C. Shen, S. Yu, and J. Tang, "A multi-stage adaptive feature fusion neural network for multimodal gait recognition," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 2024.

[46] H. Chao, Y. He, J. Zhang, and J. Feng, "Gaitset: Regarding gait as a set for cross-view gait recognition," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, pp. 8126–8133, 2019.

[47] M. Sharif, M. Attique, M. Z. Tahir, M. Yasmim, T. Saba, and U. J. Tanik, "A machine learning method with threshold based parallel feature fusion and feature selection for automated gait recognition," *Journal of Organizational and End User Computing (JOEUC)*, vol. 32, no. 2, pp. 67–92, 2020.

[48] C. Fan, Y. Peng, C. Cao, X. Liu, S. Hou, J. Chi, Y. Huang, Q. Li, and Z. He, "Gaitpart: Temporal part-based model for gait recognition," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 14225–14233, 2020.

[49] T. Teepe, A. Khan, J. Gilg, F. Herzog, S. Hörmann, and G. Rigoll, "Gaitgraph: Graph convolutional network for skeleton-based gait recognition," in *2021 IEEE international conference on image processing (ICIP)*, pp. 2314–2318, IEEE, 2021.

[50] M. Khaliluzzaman, A. Uddin, K. Deb, and M. J. Hasan, "Person recognition based on deep gait: A survey," *Sensors*, vol. 23, no. 10, p. 4875, 2023.

[51] P. Güner Şahan, S. Şahin, and F. Kaya Gülağız, "A survey of appearance-based approaches for human gait recognition: techniques, challenges, and future directions," *The Journal of Supercomputing*, pp. 1–38, 2024.

[52] K. Xu, X. Jiang, and T. Sun, "Gait recognition based on local graphical skeleton descriptor with pairwise similarity network," *IEEE Transactions on Multimedia*, vol. 24, pp. 3265–3275, 2021.

[53] P. Vasudevan, R. Faerie-Mattins, S. Srivarshan, A. Narayanan, G. Wadhwani, R. Parvathi, and R. Maheswari, "Gait image classification using deep learning models for medical diagnosis," *Comput, Mat Continua*, vol. 74, no. 3, pp. 6039–6063, 2023.

[54] T. Teepe, A. Khan, J. Gilg, F. Herzog, S. Hörmann, and G. Rigoll, "Gaitgraph: Graph convolutional network for skeleton-based gait recognition," in *2021 IEEE international conference on image processing (ICIP)*, pp. 2314–2318, IEEE, 2021.

[55] N. Sadeghzadehyazdi, T. Batabyal, and S. T. Acton, "Modeling spatiotemporal patterns of gait anomaly with a cnn-lstm deep neural network," *Expert Systems with Applications*, vol. 185, p. 115582, 2021.

[56] T. Kurosaka, K. Honda, K. Miwa, T. Umeda, M. Uemae, and S. Yamamoto, "Clahe (contrast limited adaptive histogram equalization) image processing to improve the cr portal image in radiation therapy," *Igaku Butsuri. Supplement*, vol. 28, no. suppl. 2, pp. 158–159, 2008.