Thesis for Master's Degree

Hyperspectral Face Dataset Augmentation for Enhancing Face Recognition System Performance

Youngin Choi

Artificial Intelligence Graduate School

Gwangju Institute of Science and Technology

2024

석 사 학 위 논 문

얼굴 인식 시스템 성능 향상을 위한 초분광 얼굴 데이터셋 증강

최영인

A I 대 학 원

광 주 과 학 기 술 원

2024

Hyperspectral Face Dataset Augmentation for Enhancing Face Recognition System Performance

Advisor: Heung-No Lee
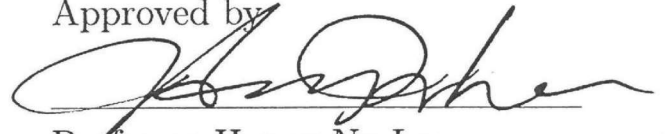
by

Youngin Choi

Artificial Intelligence Graduate School

Gwangju Institute of Science and Technology

A thesis submitted to the faculty of the Gwangju Institute of Science and Technology in partial fulfillment of the requirements for the degree of Master of Science in the Artificial Intelligence Graduate School

Gwangju, Republic of Korea

June 10, 2024

Approved by

Professor Heung-No Lee

Committee Chair

Hyperspectral Face Dataset Augmentation for Enhancing Face Recognition System Performance
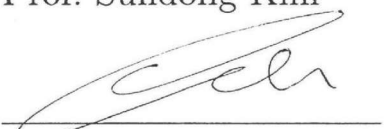
Youngin Choi

Accepted in partial fulfillment of the requirements for

the degree of Master of Science

June 10, 2024

Committee Chair _____
Prof. Heung-No Lee

Committee Member _____
Prof. Sundong Kim

Committee Member _____
Prof. Hoon Hahn Yoon

Abstract

RGB images represent information using only three primary color channels: red, green, and blue. In contrast, hyperspectral imaging utilizes a broader spectrum to convey information, allowing hyperspectral face images to capture a wider range of details, thereby enabling a more accurate assessment of facial features and skin conditions. However, acquiring hyperspectral face images and building a substantial dataset is challenging due to the need for expensive cameras, sophisticated setups, and controlled environments. Currently available hyperspectral face datasets are limited in the number of subjects and samples, making it difficult to generalize face recognition systems that utilize hyperspectral imaging.

To address these limitations, this paper proposes enhancing the diversity of hyperspectral face datasets through various augmentation techniques. Specifically, the image generation method StyleGAN2-ADA was employed to create new subjects and Face Pose Augmentation was used to generate various poses for the same subject, thereby increasing dataset diversity. Experiments were conducted based on the augmented hy-

perspectral face dataset to verify the impact of the proposed method on improving

the performance of face recognition systems. Experimental results demonstrate that

models applying the proposed augmentation method exhibit enhanced performance

compared to previous models, making a significant contribution to the advancement of

face recognition systems.

국 문 요 약

RGB 이미지는 빨강, 초록, 파랑의 세 가지 기본 색상 채널만을 사용하여 정보를 표현하는 반면, 초분광 이미징은 더 넓은 스펙트럼을 사용하여 정보를 표현한다. 이로써 초분광 얼굴 이미지는 보다 다양한 세부 정보를 포함하고 있어 얼굴의 특징과 피부 상태를 더욱 정확하게 파악할 수 있다. 그러나 고가의 카메라와 정교한 설치, 그리고 제한된 환경에서의 촬영 등으로 인해 초분광 얼굴 이미지를 획득하고 충분한 양의 데이터를 구축하는 것은 어려운 일이다. 현재 공개된 초분광 얼굴 데이터셋은 객체와 샘플 수가 제한적이며, 이는 초분광 이미징을 활용한 얼굴 인식 시스템의 일반화를 어렵게 만든다.

이에 본 논문에서는 이러한 한계를 극복하기 위해 초분광 얼굴 데이터셋을 다양하게 증강하여 데이터의 다양성을 확보하고자 한다. 먼저, 이미지 생성 기법인 StyleGAN2-ADA를 활용하여 새로운 객체를 생성하고, 더불어 Face Pose Augmentation을 이용하여 동일 객체에 대한 다양한 포즈를 변형시켜 데이터의 다양성을 향상시켰다. 이를 통해 증강된 초분광 얼굴 데이터셋을 기반으로 실험을 진행하고, 제안한 방법이 얼굴 인식 시스템의 성능 향상에 미치는 효과를 검증하였다. 실험 결과, 제안한 증강 방법을 적용한 모델의 성능이 이전에 비해 향상되었으며, 이는 얼굴 인식 시스템의 성능 향상에 유의미한 기여를 한 것으로 나타났다.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Motivation

Face recognition is a biometric technology used for identity verification, identifying individuals based on their unique physiological characteristics. Face recognition systems utilize computer vision and deep learning algorithms to extract facial features and compare them to a stored database to verify or authenticate identity. These systems play a crucial role in various fields. Firstly, in terms of security and user authentication, face recognition is used in smartphones, laptops, and access control systems as a replacement for passwords or PIN codes, providing a higher level of security. This addresses the vulnerabilities of traditional authentication methods, such as password theft or hacking. Additionally, face recognition systems are essential tools in public safety and surveillance. They can identify specific individuals in real-time within large crowds in public places such as airports, train stations, and stadiums, aiding in early detection of potential threats and significantly contributing to crime prevention and investigation [1]. Lastly, the importance of contactless authentication has been highlighted during situations like the recent COVID-19 pandemic. Face recognition can verify identity without physical contact, thereby reducing the risk of infection [2].

Hyperspectral imaging offers several advantages that overcome the limitations of

RGB images. Firstly, hyperspectral imaging provides higher spectral resolution. By collecting information from various wavelength bands, hyperspectral imaging allows for more accurate recognition of unique facial features through detailed spectral information obtained at each wavelength. This is beneficial for more accurately distinguishing skin texture, pigmentation changes, wrinkles, and other features. Secondly, hyperspectral imaging is less sensitive to changes in lighting and pose. This enables consistent face recognition performance under various lighting conditions and pose variations. Thirdly, hyperspectral imaging is resistant to spoofing techniques that deceive face recognition systems. Because hyperspectral imaging can capture information beneath the surface, it reduces recognition errors even when faces are covered or disguised. For example, by identifying patterns in the blood vessels beneath the skin or structural features within skin layers, hyperspectral imaging maintains high recognition performance despite external facial changes. Due to these advantages, hyperspectral imaging is considered an important technology that can replace or complement traditional RGB image-based face recognition systems, significantly enhancing their accuracy and reliability [3].

However, Acquiring hyperspectral face datasets presents several challenges. Hyperspectral imaging requires expensive cameras and sophisticated setups, and maintenance is also complex. Additionally, specific conditions and controlled environments are required for shooting. For instance, consistent lighting conditions and a stable environment are necessary, making it difficult to reflect the diverse environmental conditions of real-life scenarios. These limitations reduce the accessibility of dataset acquisition [4, 5].

Consequently, currently available hyperspectral face datasets have a limited number of subjects. In face recognition systems, the size and diversity of the dataset play a crucial role in training and evaluation. If the dataset is insufficient, it cannot adequately reflect diverse facial features and environmental conditions, resulting in decreased performance in practical applications and making generalization difficult. These limitations are significant factors that hinder the advancement of hyperspectral image-based face recognition technology.

1.2 Contribution

This paper explores the use of deep learning for dataset augmentation to overcome the limitations of hyperspectral face datasets and to ensure diversity. Specifically, StyleGAN2-ADA was employed to generate new hyperspectral face subjects, and an embedding model-based subject identification method was proposed to distinguish the new subjects. This approach allowed the overcoming of existing dataset limitations and the acquisition of diverse hyperspectral face data. Additionally, Face Pose Augmentation technology was used to transform the same subject into various poses, further increasing the diversity of the dataset.

Moreover, ensuring the diversity of the hyperspectral face dataset directly contributed to enhancing the performance of the face recognition system. The model trained with the augmented hyperspectral face dataset, along with the existing hyperspectral face dataset, was compared to the model trained only with the existing hyperspectral face dataset. The results showed that the model trained with the augmented

dataset exhibited higher recognition performance. This indicates that the augmented hyperspectral face dataset includes more diverse characteristics compared to the existing dataset, suggesting that data augmentation techniques can effectively address the limitations of constrained datasets.

The structure of this paper is as follows. Chapter 2 provides background knowledge on the differences between RGB images and hyperspectral imaging, applications of hyperspectral imaging, and data augmentation techniques used in the research. Chapter 3 details the proposed data augmentation methodology and designs a hyperspectral face recognition system to demonstrate the effectiveness of the data augmentation. Chapter 4 evaluates the results of the augmented dataset and the performance of the proposed hyperspectral face recognition system. Finally, Chapter 5 summarizes the key findings of the research, discusses the contributions and limitations of the study, and suggests directions for future research.

Chapter 2

Background

2.1 Overview of RGB and Hyperspectral imaging

RGB images represent images using three color channels: red, green, and blue. These images are widely used in the fields of computer vision and digital image processing and can be observed in digital cameras, smartphone cameras, monitors, televisions, and more. Each channel is represented by values ranging from 0 to 255, and combinations of these values can express various colors. RGB images provide color information in the most intuitive way, operating most similarly to the human visual system.

Hyperspectral imaging utilizes light wavelengths in the visible spectrum (500-700nm) and slightly into the near-infrared spectrum (700-1000nm) to analyze object characteristics. It consists of hyperspectral cubes with 2-dimensional spatial information and 1-dimensional spectral information, where the images within the cube represent spatial values captured at specific wavelengths. Hyperspectral imaging is utilized in various fields such as medical diagnosis, environmental monitoring, agriculture, food safety inspection, etc. Particularly in the field of face recognition, it enables more accurate recognition by analyzing structural features of the skin layer, vascular patterns, etc. It provides more information compared to RGB images and is used for more sophisticated image analysis and processing.
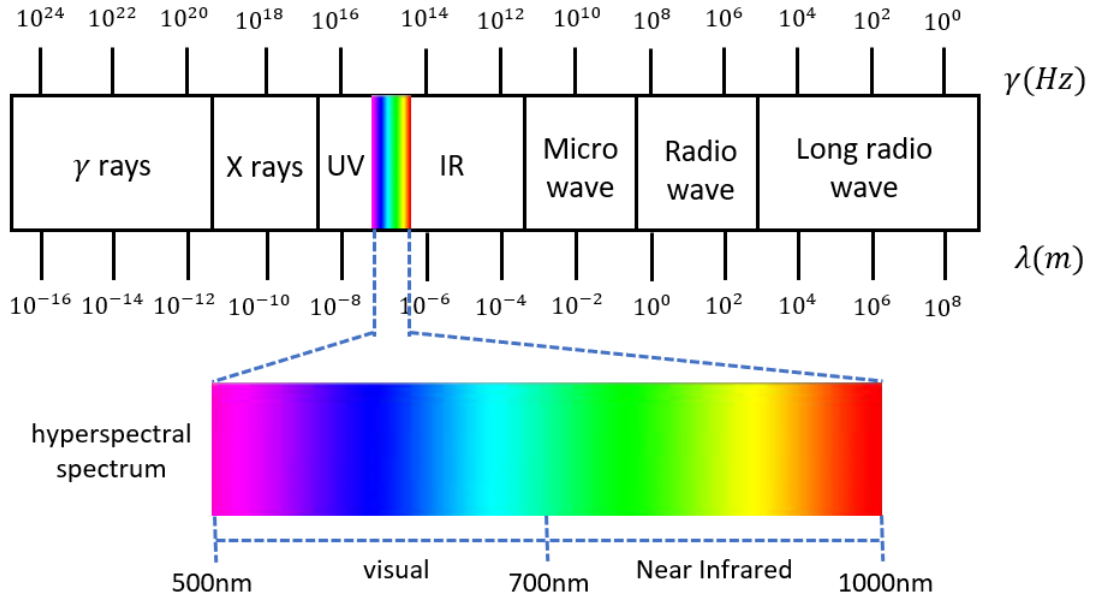
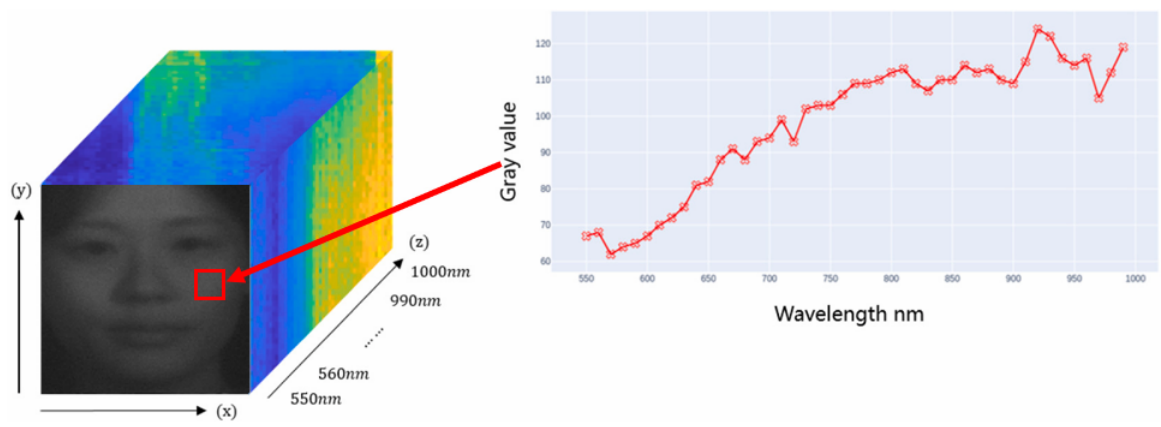Figure 2.1: Hyperspectral spectrum

Figure 2.2: Hyperspectral Cube

2.2 Applications of Hypersepctral Imaging

Hyperspectral imaging can be applied in various fields. The following are examples of applications of hyperspectral imaging. Firstly, it can be utilized in the medical diagnosis field. By simultaneously measuring tissue structure and internal chemical characteristics, it is used to monitor and assess the health status of tissues. This is beneficial for analyzing tissue characteristics of tumors or measuring the shape of blood vessels and oxygen saturation [6]. Secondly, hyperspectral imaging plays an important role in the food industry. It is used for ingredient analysis, detecting contaminants, and assessing freshness, thus improving food safety and quality to provide consumers with safe products [7]. Thirdly, in the agricultural sector, hyperspectral imaging is used to monitor the health status of crops and increase productivity. By analyzing the nutritional status of crops, presence of pests, and growth status, appropriate management and measures can be taken [8].

Especially in the field of face recognition, hyperspectral imaging enables the following applications. Firstly, it can measure skin pigmentation, oxygen saturation, and hemoglobin concentration, allowing for differentiation of subtle differences even among individuals with similar skin tones [9]. Secondly, hyperspectral imaging is effective in preventing spoofing using the unique spectral signature of the face. Characteristics such as vascular patterns beneath the skin or melanin distribution are difficult to tamper with, enhancing the security of face recognition systems by preventing forgery and tampering. This is particularly useful in high-security areas for identity verification or access control systems [10]. Lastly, hyperspectral imaging is less sensitive to

changes in lighting conditions, maintaining consistent performance under various lighting conditions. This helps face recognition systems maintain high accuracy in diverse environments such as outdoor, indoor, day, or night [11].

2.3   Hyperspectral Face Datasets

There are three main datasets commonly used for hyperspectral face recognition. Firstly, CMU-HSFD [12], collected by Carnegie Mellon University, covers a wavelength range from 450nm to 1100nm with 65 bands at 10nm intervals. The images have a resolution of 640x480 and include 54 subjects. This dataset was captured under various lighting conditions to account for the impact of light on subject characteristics. Secondly, PolyU-HSFD [13], collected by Hong Kong Polytechnic University, spans from 400nm to 720nm with 33 bands at 10nm intervals. Images were obtained by varying angles during capture, including frontal, left, and right views. However, the first 6 bands of this dataset have very low signal-to-noise ratios and are not suitable for face recognition. Additionally, 26 subjects have between 4 to 7 samples, while the remaining 23 subjects have 1 to 2 samples each. Lastly, UWA-HSFD [14] collected by the University of Western Australia covers a wavelength range from 400nm to 720nm with 33 bands at 10nm intervals. The images have a resolution of 1024x1024 and include 78 subjects. This dataset has adaptively different exposure times based on the signal strength at each wavelength, resulting in minimal noise interference, but it suffers from inter-band alignment issues due to subject movement.

While these datasets provide valuable resources for hyperspectral face recognition

research, they have certain limitations. The number of subjects in each dataset is limited. CMU-HSFD includes 54 subjects, PolyU-HSFD includes 48 subjects, and UWA-HSFD includes 78 subjects, indicating a lack of diversity in the data. This inadequacy in representing various races, ages, and genders may compromise the generalization performance of models. Additionally, each dataset has technical issues. For example, PolyU-HSFD has the low-signal band problem, and UWA-HSFD faces inter-band alignment issues due to subject movement and blinking. These technical problems can degrade the quality of the data. Furthermore, collecting hyperspectral datasets requires expensive equipment and sophisticated installation. The high cost and complexity of using and maintaining such equipment pose significant constraints on dataset construction.

2.4 Image Augmentation Techniques

As mentioned earlier, the size and diversity of hyperspectral facial datasets are limited due to the difficulty and cost of data collection. To address this, we considered using image augmentation techniques to increase the size and diversity of datasets at minimal cost. The augmentation techniques used in this paper for hyperspectral facial images are as follows.

StyleGAN2-ADA [15] is a variation of Generative Adversarial Network (GAN). GAN consists of two neural networks: a generator G and a discriminator D. The ob-

jective function of GAN is defined as:

$$V(G, D) = \mathbb{E}_{x \sim p_{\text{data}}(x)}[\log D(x)] + \mathbb{E}_{z \sim p_z(z)}[\log(1 - D(G(z)))]$$

Here, $p_{\text{data}}(x)$ represents the distribution of real data, and $p_z(z)$ represents the distribution of latent space. The objective function $V(G, D)$ controls the learning of the generator and discriminator while maintaining balance between them. The generator learns to maximize $\log(1 - D(G(z)))$ to make fake data look real, while the discriminator learns to maximize $\log D(x)$ to correctly distinguish between real and fake data. Ultimately, the generator generates images that are indistinguishable from real data, and the discriminator learns to classify fake images to the point where they cannot be distinguished anymore. Through this process, GAN learns to generate data similar to real data and learns the distribution of real data. However, training GAN with a small dataset can lead to the problem of discriminator overfitting and divergence. This means that the balance between the generator and the discriminator can be disrupted, and the distribution of real data may not be well learned.

StyleGAN2-ADA addresses these issues by mitigating discriminator overfitting through data augmentation. Only augmented images are used to train the discriminator and generator, and the augmentation range is adaptively adjusted based on the discriminator's output using augmentation probability p. This method enables effective learning even with a small dataset. Additionally, there are 18 variations across six categories. For instance, the dataset is diversified through methods such as image rotation, color changes, and size adjustments.

Face Pose Augmentation [16] was utilized to generate the same object in various poses. A comprehensive facial detection and augmentation system was implemented through facial detection, landmark detection, 3D facial alignment, and finally, the Face Pose Augmentator. Specifically, RetinaFace was used to detect faces, Face Alignment Network (FAN) identified key landmarks, and the 3D Dense Face Alignment (3DDFA) model predicted the 3D pose of the face. Finally, the Face Pose Augmentator augmented the facial pose based on the predicted 3D pose. As a result, this system provides a unified framework with high accuracy and performance in facial analysis and augmentation. By using pretrained models, facial alignment and pose augmentation can be performed without separate training processes.

2.5 Hyperspectral Face Recognition

Previous research on hyperspectral face recognition systems has focused on selecting optimal bands. Although hyperspectral face datasets consist of high-dimensional facial data, each band can be used as an individual image in the dataset. When training models, typically 30 images per person are used [3]. However, training with all wavelength bands can decrease classification performance due to high correlation between bands. Therefore, research has been conducted to select the optimal bands for face recognition. Based on prior studies, this paper conducts research on augmenting face recognition by selectively choosing the optimal wavelength bands.

Chapter 3

Proposed Method

3.1 Overview of the Proposed method

The proposed method in this paper consists of two main components. First, StyleGAN2-ADA is used to generate new subject faces, and an embedding model is employed to determine if the generated faces are new subjects. Additionally, Face Pose Augmentation is utilized to enhance the dataset with various face poses of the subjects. Second, the augmented dataset is used to train the hyperspectral face recognition system. Figure 3.1 illustrates the entire flowchart of the proposed method. Through preprocessing, the size of the hyperspectral face dataset is adjusted to be used as input for both the StyleGAN2-ADA training and the Face Pose Augmentation. To demonstrate the effectiveness of the augmented hyperspectral face dataset, the face recognition performance of models trained solely on the original hyperspectral face dataset was compared with models trained on both the original and augmented datasets, using the same test dataset for evaluation.
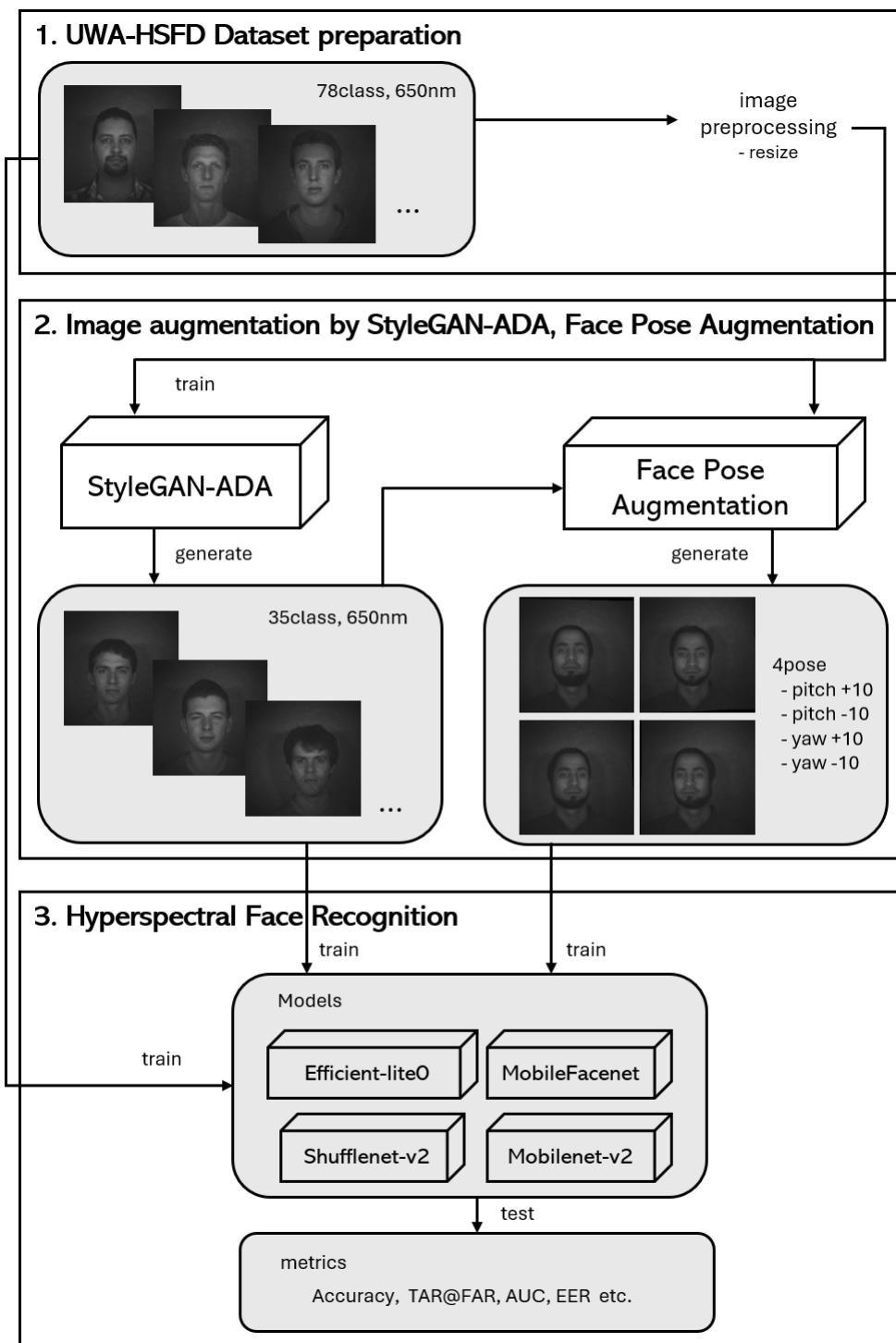
Figure 3.1: Flowchat of Proposed Method

3.2   Hyperspectral Face Image Augmentation System

3.2.1   Augmentation of New Subjects via StyleGAN2-ADA

The dataset used to augment new subjects is UWA-HSFD. We selected publicly available datasets from the internet, while two other datasets were restricted. The dataset used for training was based on the 650nm wavelength within the range of 400nm to 720nm. The process of capturing hyperspectral images took some time, resulting in slight head movements and eye blinking in some images. These movements caused minor alignment errors between wavelengths. Therefore, we selected the accurately aligned 650nm wavelength for use. Additionally, images were preprocessed before training by downsizing the resolution from 1024x1024 to 256x256.

The current hyperspectral face dataset comprises approximately 150 samples. Due to the inadequate quantity for training StyleGAN2-ADA, transfer learning was performed. A pretrained model was created by converting 70,000 images from the Flickr-Faces-HQ Dataset (FFHQ) to grayscale, as hyperspectral imaging is typically rendered in grayscale to visually represent reflectance data at specific wavelengths. Each grayscale image visually represents reflectance at the corresponding wavelength.

To demonstrate that the subjects generated by StyleGAN2-ADA differ from existing subjects, an embedding model was trained using only the existing hyperspectral face dataset. The similarity between generated images and existing face images was quantified, and subjects were classified using the trained embedding model. The embedding model was trained using MobileFacenet [17], which is based on the MobileNetV2 architecture designed to reduce model size and increase computational efficiency using

depthwise separable convolutions. Triplet loss [18] was employed in training with the objective of minimizing the distance between positive embeddings and anchor embeddings, while simultaneously maximizing the distance between negative embeddings and anchor embeddings. The triplet loss function is defined as:

$$L(A, P, N) = \max\left\{\left(||f(A) - f(P)||^2 - ||f(A) - f(N)||^2 + \alpha, 0\right\}\right.$$

Here, $f(\cdot)$ denotes the embedding function for each sample. A, P, and N represent the anchor, positive, and negative samples respectively. $||\cdot||$ represents the L2 norm, and $||\cdot||^2$ signifies the squared L2 norm of a vector. This loss formulation ensures that the distance between the positive embedding and the anchor embedding is at least α smaller than the distance between the negative embedding and the anchor embedding.

Figure 3.2 visualizes the distances between the existing hyperspectral face dataset using the trained embedding model, showing distances between the same subjects (blue distribution) and different subjects (red distribution). Distances were computed using cosine similarity, a measure that assesses similarity by calculating the cosine angle between two embedding vectors. It produces values ranging from -1 to 1, with a value closer to 1 indicating a higher degree of similarity between the vectors.

The equation for cosine similarity between two embedding vectors a and b is given by:

$$\text{cosine similarity} = \frac{a \cdot b}{||a|| \, ||b||}$$

Here, $a \cdot b$ represents the dot product of vectors a and b, while $||a||$ and $||b||$ represent
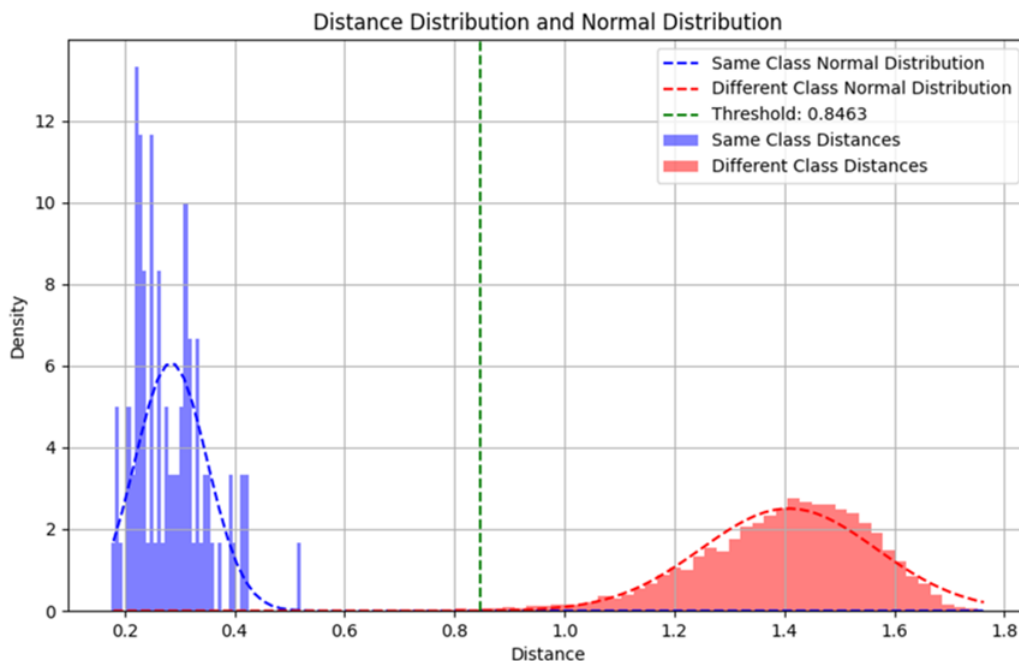
Figure 3.2: Visualization of Classification Thresholds for Subjects

the norms of vectors a and b respectively. Normalizing both distributions, we derived a threshold value of 0.8463 at the point of overlap between the two distributions, to determine if images belong to the same subject. Subsequently, distances between images generated by StyleGAN2-ADA and the existing hyperspectral face dataset were calculated, classifying images with distances of 0.8463 or higher as new subjects.

3.2.2 Augmentation of Various Poses of the Same Subject via Face Pose Augmentation

To augment the poses of existing hyperspectral face datasets and the datasets generated using StyleGAN2-ADA, 3DDFA was employed to estimate the 3D positions and poses of faces. Based on these estimates, face images were adjusted according to the given pose transformation values to generate images with various poses. The aug-

mented images include the following configurations: pitch rotations of +10 degrees and -10 degrees, and yaw rotations of +10 degrees and -10 degrees. In total, four different pose images were augmented for each subject. Face Pose Augmentation utilized a pre-trained model, with input image sizes of 256x256.

3.3 Hyperspectral Face Recognition System

The Table 3.1 presents the number of subjects and total samples in the training and testing datasets. Each training dataset consists of a total of 115 existing hyperspectral face images and 750 existing and augmented hyperspectral face images. The augmented hyperspectral face dataset includes 35 new subjects, each with additional face images depicting four different poses. To compare and evaluate the performance of the two models, the same testing dataset comprising 15 subjects and a total of 30 samples was utilized.

| | Training Dataset | | Test Dataset |
	Original Dataset	Original and Augmented Dataset	
Number of Objects	63	98	15
Total Samples	115	750	30

Table 3.1: Datasets for the Hyperspectral Face Recognition System

Before training the models, the Multi-task Cascaded Convolutional Networks (MTCNN) was employed to detect facial regions and extract aligned face patches. These extracted face patches were used as the training dataset, and a loss function was optimized for the embedding vectors generated by a specific extractor. The loss function utilized was ArcFace [19], which minimizes the distance between embeddings of the same iden-

tity while maximizing the distance between embeddings of different identities. ArcFace normalizes feature vectors to unit length, introduces a margin considering angles between subjects, and minimizes the cross-entropy loss by calculating probabilities for each subject using the softmax function. The Arcface loss function is defined as:

$$L = -\frac{1}{N} \sum_{i=1}^{N} \log \frac{e^{s(\cos(\theta_{y_i}+m))}}{e^{s(\cos(\theta_{y_i}+m))} + \sum_{j=1,j\neq y_i}^{C} e^{s\cos(\theta_j)}}$$

In this equation, L is the loss function, N is the number of samples in the dataset, s is a scaling factor used to amplify the angular margin, m is the angular margin, which is a hyperparameter, C is the total number of classes (subjects) in the dataset, θ_{y_i} represents the angle between the input feature vector and the weight vector corresponding to the true class y_i, and θ_j represents the angle between the input feature vector and the weight vector corresponding to the j-th class, where $j \neq y_i$.

For feature extraction, the models MobileFacenet, Mobilenet-v2 [20], Shufflenet-v2 [21], and Efficient-Lite0 [22] were utilized. MobileFaceNet employs depthwise separable convolutions to reduce model size and computational cost while generating high-dimensional embeddings to effectively measure facial similarity. Similarly, Mobilenet-v2 utilizes depthwise separable convolutions to significantly reduce parameters and computations, introduces residual connections for more stable and effective learning, expands input channels at the beginning of each block using extended layers to learn rich features, and then reduces them again for computational efficiency. Shufflenet-v2 utilizes channel shuffling to effectively mix features from different groups, designed with a parallel processing structure for high operational efficiency, avoiding complex

operations and replacing them with simpler ones to greatly improve execution speed. Lastly, Efficient-Lite0 optimizes efficiency and performance by balancing the network's width, depth, and resolution, utilizing Mobile Inverted Bottleneck Convolution blocks to enhance computational efficiency, and aiding in the automatic exploration of model architectures through Neural Architecture Search to achieve optimal performance. All four models are designed for efficient execution in mobile and embedded environments.

The trained models take images as input, extract features of each face, and transform them into embedding vectors. All models utilized the Stochastic Gradient Descent (SGD) optimization algorithm with a learning rate of 0.1 and a momentum of 0.9. Weight decay was set to 5e-4 to prevent overfitting, and the MultiStepLR algorithm was employed to decrease the learning rate by 0.1 at epochs 20, 35, and 45 to facilitate model convergence.

Chapter 4

Experiment and Results

4.1 Hyperspectral Face Image Augmentation System

4.1.1 Augmentation of New Subjects via StyleGAN2-ADA

To evaluate the models trained using StyleGAN2-ADA, the Fréchet Inception Distance (FID-Score) metric [23] was used to measure the difference between generated and real images. During training, the augmentation probability was fixed at 0.6. The FID-Score obtained without transfer learning, using only the hyperspectral face dataset, was 54.61. After transfer learning with the hyperspectral face dataset, the FID-Score improved to 40.12, indicating the benefit of transfer learning for training StyleGAN2-ADA. Furthermore, using a trained embedding model to discern new subjects, out of 50 subjects generated by StyleGAN2-ADA, 35 were determined to be new subjects based on a distance threshold of 0.8463 from the existing hyperspectral face dataset. Examples illustrating the original and augmented hyperspectral face datasets are depicted in the figures. Figure 4.1 represents images from the original hyperspectral face dataset, while Figure 4.2 depicts images generated by StyleGAN2-ADA that have a distance greater than 0.8463 from the original dataset. It is evident from the figures that the augmented hyperspectral face dataset differs from the original dataset. Figures 4.3 and 4.4 show images where the distance between the original dataset and the generated

dataset is less than 0.8463. In each figure, the image on the left belongs to the original dataset, while the image on the right belongs to the generated dataset. The distances between the images are 0.6943 and 0.6782, respectively.

4.1.2 Augmentation of Various Poses of the Same Subject via Face Pose Augmentation

Figure 4.5 presents examples of the original hyperspectral face dataset and the augmented hyperspectral face dataset with varied poses. The first figure represents the original hyperspectral face image, while the others depict images with pitch adjustments of -10 degrees, yaw adjustments of -10 degrees, and yaw adjustments of +10 degrees, respectively. It is evident from the figures that the augmented hyperspectral face dataset exhibits different poses compared to the original dataset.

4.2 Hyperspectral Face Recognition System

To evaluate the trained models, a dataset containing objects not present in the training dataset was used to assess the performance of open-set face recognition tasks, specifically face verification and face identification [24]. Both tasks utilized a test dataset comprising 19 pairs of images from the same subject and 416 pairs of images from different subjects. Figure 4.6 and 4.7 illustrate the process of training and evaluating face verification and face identification tasks, respectively.

For open-set face verification, evaluation metrics such as Accuracy and AUC were used, as detailed in the Table 4.1. Overall, the performance of models trained with
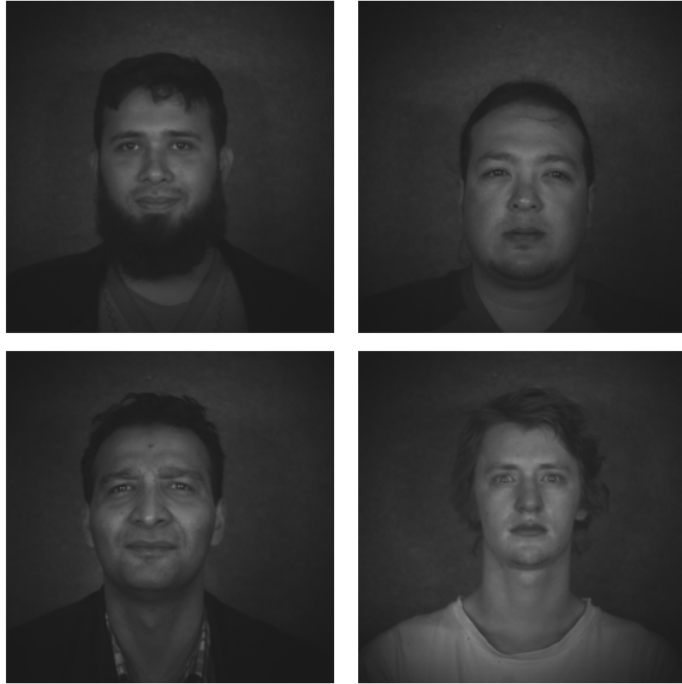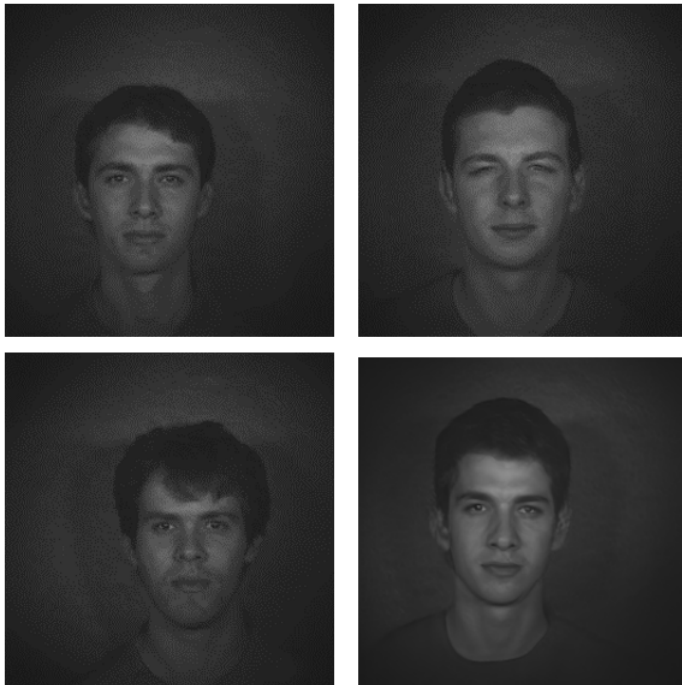
Figure 4.1: Original Dataset

Figure 4.2: Augmented Datset by StyleGAN2-ADA

Figure 4.3: Dataset with a distance less than 0.8463 (Distance : 0.6943)

Figure 4.4: Dataset with a distance less than 0.8463 (Distance : 0.6782)
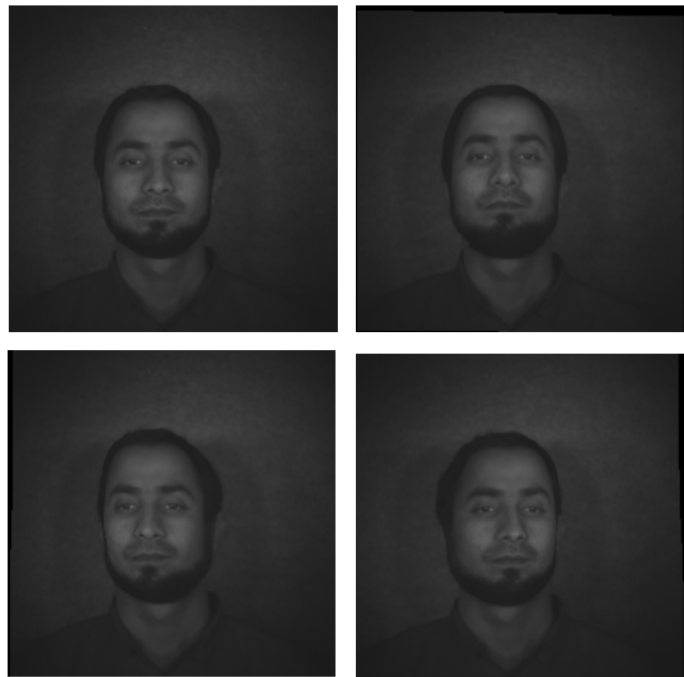

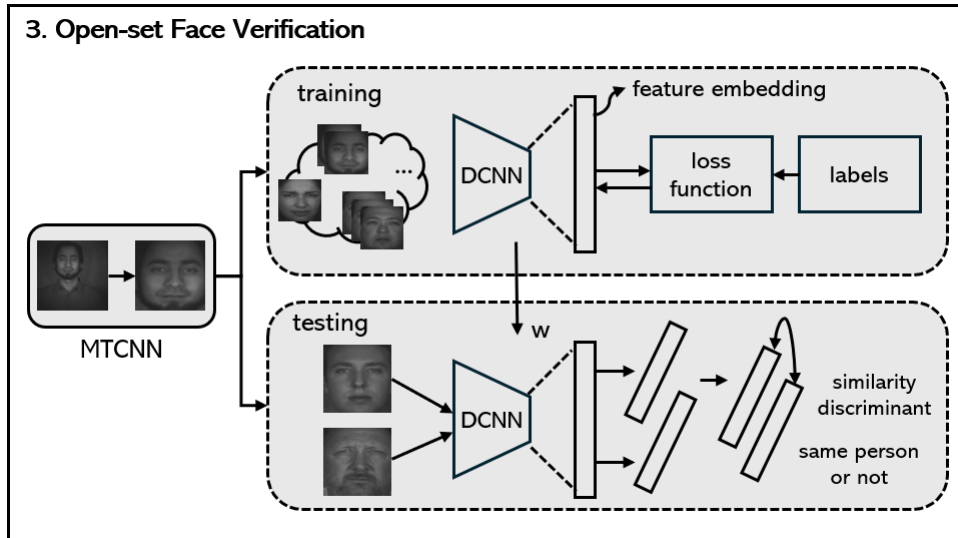Figure 4.5: Original and Augmented Dataset via Face Pose Augmentation
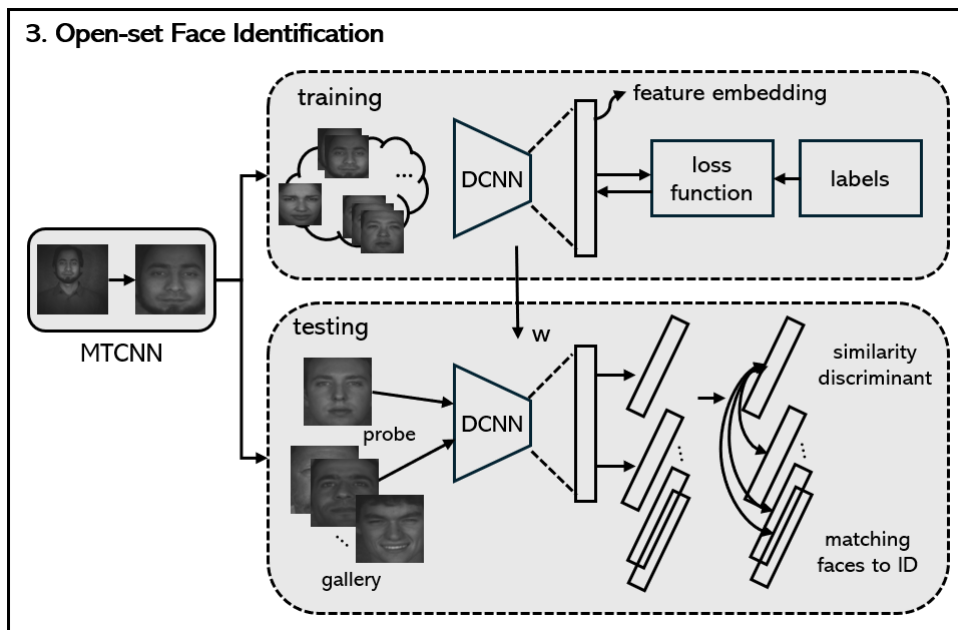
Figure 4.6: Open-set Face Verification

Figure 4.7: Open-set Face Identification

the augmented hyperspectral face dataset in conjunction with the original dataset surpassed those trained solely with the original dataset. Particularly noteworthy was the significant improvement in AUC performance for Shufflenet-v2, increasing from 0.7098 to 0.9619, indicating that the model exhibited high accuracy and sensitivity. This implies the model's ability not only to better identify the correct individuals in face verification tasks but also to accurately distinguish incorrect individuals.

Evaluation Metric	Description
Accuracy	• The ratio of correctly classified samples in the entire dataset
AUC	• Area Under the Curve • Represents the area under the ROC curve which indicates the performance of the model
TAR @ FAR	• True Accept Rate @ False Accept Rate • The ratio of correctly accepted genuine users without rejecting them in the system
EER	• Equal Error Rate • The point where the false accept rate and false reject rate are equal, indicating the error rate at this threshold

Table 4.1: Metrics Used to Evaluate Face Verification

Accuracy performance improved on average by 0.0735, with the highest Accuracy observed in MobileFacenet at 0.9609. Furthermore, the performance of Shufflenet-v2 significantly improved in TAR@FAR 0.100, reaching 0.6316. Lastly, the Equal Error Rate (EER) decreased by an average of 0.1009, with the lowest value observed in Mobilenet-v2 at 0.0252. Table 4.2 compares the verification performance of models trained with the original hyperspectral face dataset and those trained with both the original and augmented datasets.

For open-set face identification, the evaluation metric used was TPIR@FPIR. In

face recognition, it is crucial not only to correctly identify the right individuals but also to accurately reject incorrect individuals. This is important as misidentifying individuals in authentication systems can pose security risks. TPIR@FPIR is used to evaluate both aspects. Overall, models trained with the augmented hyperspectral face dataset alongside the original dataset outperformed those trained solely with the original dataset. On average, there was a performance improvement of approximately 0.1617 in TPIR@FPIR, with Shufflenet-v2 showing the most significant improvement with TPIR@FPIR 0.100 reaching approximately 0.353. Table 4.3 compares the identification performance of models trained with the original hyperspectral face dataset and those trained with both the original and augmented datasets.

model	metric		base model	proposed model
MobileFacenet	Accuracy (↑)		0.9448	**0.9609**
	AUC (↑)		0.9530	**0.9854**
		0.001	0.2632	**0.6316**
	TAR @ FAR (↑)	0.010	0.7368	**0.7895**
		0.100	0.8421	**0.9474**
	EER (↓)		0.1606	**0.1116**
Mobilenet-v2	Accuracy		0.9241	**0.9517**
	AUC		0.9467	**0.9907**
		0.001	0.0526	**0.6316**
	TAR @ FAR	0.010	0.4211	**0.6842**
		0.100	0.8421	**1.0000**
	EER		0.1151	**0.0252**
Shufflenet-v2	Accuracy		0.7103	**0.8344**
	AUC		0.7098	**0.9619**
		0.001	0.0526	**0.2105**
	TAR @ FAR	0.010	0.0526	**0.3684**
		0.100	0.2105	**0.8421**
	EER		0.2770	**0.1151**
Efficient-Lite0	Accuracy(↑)		0.7517	**0.8781**
	AUC(↑)		0.8467	**0.9262**
		0.001	0.0000	**0.0526**
	TAR @ FAR(↑)	0.010	0.1053	**0.1053**
		0.100	0.5263	**0.7895**
	EER(↓)		0.2170	**0.1139**

Table 4.2: Performance Evaluation via Open-set Face Verification

model	metric		base model	proposed model
MobileFacenet	TPIR @ FPIR (↑)	0.001	0.5882	**0.7941**
		0.010	0.8529	**0.8824**
		0.100	0.9118	**0.9706**
Mobilenet-v2	TPIR @ FPIR	0.001	0.4706	**0.7941**
		0.010	0.6765	**0.8235**
		0.100	0.9118	**1.000**
Shufflenet-v2	TPIR @ FPIR	0.001	0.4706	**0.5588**
		0.010	0.4706	**0.6471**
		0.100	0.5588	**0.9118**
Efficient-Lite0	TPIR @ FPIR	0.001	0.4706	**0.5294**
		0.010	0.5000	**0.5882**
		0.100	0.5882	**0.9118**

Table 4.3: Performance Evaluation via Open-set Face Identificaiton

Chapter 5

Conclusion

In this study, a hyperspectral face recognition system was proposed, augmented by generating new subjects using StyleGAN2-ADA and augmenting various poses using Face Pose Augmentation. Utilizing the UWA-HSFD dataset, new face images were generated and the limitations of the existing dataset were overcome through transfer learning. MobileFacenet was used to classify 70% of the generated subjects as new subjects. Additionally, by estimating the 3D position and pose of faces using 3DDFA, four different images per subject were created, enhancing the diversity of the dataset. Training various models on both the original and augmented datasets resulted in higher overall performance metrics such as Accuracy, AUC, TAR@FAR, and EER. Moreover, the identification metric TPIR@FPIR showed an average performance improvement of 0.1617. These results demonstrate the significant contribution of data augmentation through StyleGAN2-ADA and Face Pose Augmentation to enhancing the performance of hyperspectral face recognition systems.

However, this study has limitations in terms of the lack of diversity in the newly generated subjects. This reflects issues such as the predominantly male composition of the dataset, insufficient diversity in age groups and races, and the lack of diversity in lighting conditions. Consequently, most of the generated images consist of the same demographic of male subjects.

Additionally, as the training and augmentation were conducted using only a single band of the hyperspectral face dataset, the characteristics of faces across multiple wavelength bands were not reflected. Moreover, the dataset is limited to the visible light spectrum.

Furthermore, recent advancements in miniature hyperspectral cameras hold promise for commercialization due to their compact size and affordability. Therefore, the plan is to enhance the generalized performance of face recognition systems by acquiring datasets covering various environments, lighting conditions, poses, and accessories using these miniature hyperspectral cameras. Additionally, the aim is to evaluate facial characteristics using spectral barcodes from various wavelength bands.

References

1. W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM computing surveys (CSUR)*, vol. 35, no. 4, pp. 399–458, 2003.

2. W. Hariri, "Efficient masked face recognition method during the covid-19 pandemic," *Signal, image and video processing*, vol. 16, no. 3, pp. 605–612, 2022.

3. X. Zhang and H. Zhao, "Hyperspectral-cube-based mobile face recognition: A comprehensive review," *Information Fusion*, vol. 74, pp. 132–150, 2021.

4. S. Rao *et al.*, "Anti-spoofing face recognition using a metasurface-based snapshot hyperspectral image sensor: supplement," 2022.

5. Z. Pan, G. E. Healey, M. B. Prasad, and B. J. Tromberg, "Face recognition in hyperspectral images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 12, pp. 1552–1560, 2003.

6. V. V. Tuchin, "Tissue optics and photonics: biological tissue structures," *Journal of Biomedical Photonics & Engineering*, vol. 1, no. 1, pp. 3–21, 2015.

7. Y.-Z. Feng and D.-W. Sun, "Application of hyperspectral imaging in food safety inspection and control: a review," *Critical reviews in food science and nutrition*, vol. 52, no. 11, pp. 1039–1058, 2012.

8. A.-K. Mahlein, "Plant disease detection by imaging sensors–parallels and specific

demands for precision agriculture and plant phenotyping," *Plant disease*, vol. 100, no. 2, pp. 241–251, 2016.

9. G. Lu and B. Fei, "Medical hyperspectral imaging: a review," *Journal of biomedical optics*, vol. 19, no. 1, pp. 010901–010901, 2014.

10. A. Jenerowicz *et al.*, "Application of hyperspectral imaging in hand biometrics," in *Counterterrorism, Crime Fighting, Forensics, and Surveillance Technologies II*, vol. 10802, SPIE, 2018.

11. C. P. Huynh and A. Robles-Kelly, "Hyperspectral imaging for skin recognition and biometrics," in *2010 IEEE International Conference on Image Processing*, pp. 1285–1288, IEEE, 2010.

12. L. Denes, P. Metes, and Y. Liu, "Hyperspectral face database," tech. rep., Carnegie Mellon University, The Robotics Institute, Pittsburgh, Pennsylvania, 2002.

13. D. Wei *et al.*, "Studies on hyperspectral face recognition in visible spectrum with feature band selection," *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans*, vol. 40, no. 6, pp. 1354–1361, 2010.

14. M. Uzair, A. Mahmood, and A. Mian, "Hyperspectral face recognition with spatiospectral information fusion and pls regression," *IEEE Transactions on Image Processing*, vol. 24, no. 3, pp. 1127–1137, 2015.

15. T. Karras, S. Laine, and T. Aila, "Training generative adversarial networks with

limited data," in *Advances in neural information processing systems*, vol. 33, pp. 12104–12114, 2020.

16. J. Guo, X. Zhu, Y. Yang, F. Yan, Z. Lei, and S. Z. Li, "Towards fast, accurate and stable 3d dense face alignment," in *Proceedings of the European Conference on Computer Vision (ECCV 2020)*, pp. 152–168, 2020.

17. S. Chen *et al.*, "Mobilefacenets: Efficient cnns for accurate real-time face verification on mobile devices," in *Biometric Recognition: 13th Chinese Conference, CCBR 2018, Urumqi, China, August 11-12, 2018, Proceedings*, vol. 13, Springer International Publishing, 2018.

18. A. Hermans, L. Beyer, and B. Leibe, "In defense of the triplet loss for person re-identification," *arXiv preprint arXiv:1703.07737*, 2017.

19. J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "Arcface: Additive angular margin loss for deep face recognition," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 4690–4699, 2019.

20. M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4510–4520, 2018.

21. N. Ma, X. Zhang, H.-T. Zheng, and J. Sun, "Shufflenet v2: Practical guidelines for efficient cnn architecture design," in *Proceedings of the European conference on computer vision (ECCV)*, pp. 116–131, 2018.

22. M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International conference on machine learning*, pp. 6105–6114, PMLR, 2019.

23. M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "Gans trained by a two time-scale update rule converge to a local nash equilibrium," in *Advances in neural information processing systems*, vol. 30, 2017.

24. M. Wang and W. Deng, "Deep face recognition: A survey," *Neurocomputing*, vol. 429, pp. 215–244, 2021.