

논문 2024-61-12-14

초분광 얼굴 데이터셋 증강을 통한 얼굴 인식 시스템 성능 향상

(Improving Face Recognition Performance through Hyperspectral Face Dataset Augmentation)

최영인*, 강주성*, 이지오*, 신영학**, 이흥노***

(Youngin Choi, Jusung Kang, Jioh Lee, Younghak Shin[©], and Heung-No Lee)

요약

RGB 이미지는 빨강, 초록, 파랑의 세 가지 기본 색상 채널만을 사용하여 정보를 표현하는 반면, 초분광 이미지는 더 넓은 스펙트럼을 활용하여 정보를 담아낸다. 이를 통해 초분광 얼굴 이미지는 얼굴의 세부적인 특징과 피부 상태를 더욱 정확하게 파악할 수 있는 세밀한 정보를 제공한다. 그러나 초분광 얼굴 이미지를 획득하기 위해서는 고가의 카메라, 정교한 장비 설치, 그리고 제한된 촬영 환경이 요구되어 충분한 양의 데이터셋을 확보하기가 어렵다. 이러한 이유로 현재 공개된 초분광 얼굴 데이터셋은 객체 및 샘플 수가 제한적이며, 이는 초분광 이미지를 사용한 얼굴 인식 시스템의 일반화에 어려움을 준다. 이를 해결하기 위해, 본 논문에서는 초분광 얼굴 데이터셋의 다양성을 증대시키는 증강 기법을 제안한다. 먼저, 이미지 생성 기법인 StyleGAN2-ADA를 사용하여 새로운 얼굴 객체를 증강한 후, Face Pose Augmentation을 적용하여 동일 객체에 대한 다양한 포즈를 생성함으로써 데이터의 다양성을 확대하였다. 초분광 이미지 기반 얼굴 인식 실험을 통해 제안된 데이터 증강 방법의 효용성을 네 가지 모델에 대해 검증한 결과, 모든 모델에서 증강된 초분광 얼굴 데이터셋을 추가로 사용한 경우 Accuracy가 향상되었으며, 최대 0.9907까지 성능이 증가하였다. 이를 통해 제안된 방법이 얼굴 인식 시스템의 성능 향상에 유의미한 기여를 할 수 있음을 입증하였다.

Abstract

RGB images use only three primary color channels to represent information, whereas hyperspectral images capture data using a broader spectrum. This allows hyperspectral face images to provide more detailed information, enabling a more accurate analysis of facial features and skin conditions. However, obtaining hyperspectral face images requires expensive cameras, precise equipment setup, and restricted environments, making it difficult to collect a sufficient amount of data. As a result, publicly available hyperspectral face datasets have limited objects and sample sizes, hindering the generalization of face recognition systems based on hyperspectral images. To address this issue, this paper proposes a data augmentation method to increase the diversity of hyperspectral face datasets. First, StyleGAN2-ADA, an image generation technique, was used to augment new facial objects. Next, Face Pose Augmentation was applied to generate diverse poses for the same object, further enhancing data variety. Through face recognition experiments based on hyperspectral images, the effectiveness of the proposed data augmentation method was validated across four models. The results showed that the models trained with the augmented hyperspectral face dataset exhibited improved accuracy, with performance increasing up to 0.9907. This demonstrates that the proposed approach can significantly enhance the performance of face recognition systems.

Keywords : Hyperspectral face recognition, StyleGAN2-ADA, Face pose augmentation, Data augmentation

*비회원, ***정회원, 광주과학기술원 전기전자컴퓨터공학부 및 AI 대학원(Department of Electrical Engineering and Computer Science and AI Graduate School, Gwangju Institute of Science and Technology)

**비회원, 목포대학교 컴퓨터공학과(Department of Computer Engineering, Mokpo National University)

※ 본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학ICT연구센터육성지원사업의 연구결과로 수행되었음. (IITP-2024-2021-0-01835).

© Corresponding Author(E-mail : younghak@mnu.ac.kr)

Received ; May 8, 2024

Revised ; September 13, 2024

Accepted ; September 13, 2024

I. 서론

얼굴은 각기 독특한 공간 구조, 기하학적 패턴, 그리고 특별한 질감을 지니고 있어 개인을 식별하는데 중요한 역할을 한다. 이러한 고유성은 생체 인식, 감시, 법의학 등 다양한 분야에서 얼굴 인식 시스템을 통한 개인 인증에 사용되고 있다^[1].

최근 초분광 얼굴 이미지를 활용한 얼굴 인식 시스템에 대한 연구가 활발히 진행되고 있다^[2]. 초분광 얼굴 이미지는 RGB 이미지보다 세밀한 정보를 담고 있어 얼굴의 세부적인 특징과 피부 상태를 더욱 정확하게 파악할 수 있다. RGB 이미지는 빨강, 초록, 파랑의 세 가지 기본 색상 채널만을 사용하여 색상 정보를 제공하는 반면, 초분광 이미지는 더 넓은 스펙트럼을 활용하여 정보를 담는다. 이는 얼굴의 세부적인 특성을 더 다양하고 정밀하게 파악할 수 있게 해주며, 결과적으로 얼굴 인식 시스템의 정확성과 신뢰성을 향상시킬 수 있다.

그러나 초분광 얼굴 데이터셋을 수집하는 과정에는 고가의 카메라, 정밀한 장비 설치, 그리고 촬영 환경^[1]의 제한이 필요하다. 이러한 한계로 인해 초분광 얼굴 데이터셋은 수집이 어렵고, 그 결과 현재 공개된 초분광 얼굴 데이터셋은 그 규모가 제한적이다. 예를 들어, CMU-HSFD^[3]는 54개의 객체를 포함하고 있으며, PolyU-HSFD^[4]는 48개, 그리고 UWA-HSFD^[5]는 78개의 객체로 구성되어 있다. 이는 RGB 기반의 얼굴 데이터셋에 비해 현저히 적은 수이며, 이로 인해 얼굴 인식 시스템의 일반화 성능을 저해할 수 있다.

본 연구에서는 이러한 초분광 얼굴 데이터셋의 제한을 극복하기 위해 딥러닝 기반의 초분광 얼굴 데이터셋 증강에 대한 연구를 수행하였다. StyleGAN2-ADA^[6]를 사용하여 새로운 얼굴 객체를 생성하고, 임베딩 모델을 학습하여 새로운 객체를 판별할 수 있는 방법을 제시하였다. 또한, Face Pose Augmentation^[7] 기술을 사용하여 동일한 객체를 다양한 포즈로 변형시킴으로써 데이터의 다양성을 증가시켰다. 이어서, 증강된 초분광 얼굴 데이터셋과 기존 초분광 얼굴 데이터셋을 함께 사용하여 모델을 학습시킨 후, 기존 데이터셋만을 사용하여 학습한 모델과 성능을 비교 분석하였다. 그 결과, 증강된 초분광 얼굴 데이터셋을 활용한 얼굴 인식 시스템 성능면에서 일관적으로 향상을 보임을 확인할 수 있었다.

본 논문의 구성은 다음과 같다. II장에서는 관련 연구와 사용된 기술을 다룬다. III장에서는 제안된 방법 상세히 설명하며, IV장에서는 실험 결과를 통해 제안된

방법의 유효성을 입증한다. 마지막으로 V장에서는 본 논문의 주요 내용을 요약한다.

II. 관련 연구

1. 초분광 얼굴 인식 시스템

X. Zhang 등은 VGG, ResNet, MobileFacedNet과 같은 다양한 모델을 사용하여 초분광 얼굴 인식 시스템을 학습하였다. 테스트 단계에서는 동일한 스펙트럼 대역 내에서 다른 얼굴 간의 인식 성능을 코사인 거리를 통해 평가하였다^[2]. 본 논문에서는 ArcFace^[8]를 추가 적용하여 얼굴 인식 성능을 더욱 향상시키고자 하였다.

초분광 얼굴 데이터셋은 여러 파장 대역의 정보를 포함하지만, 모든 대역을 학습에 활용할 경우 대역 간 상호 간섭으로 인해 성능 저하가 발생할 수 있다. 이를 해결하기 위해 최적의 파장 대역을 선택하는 연구들이 진행되었으나^[9, 10], 초분광 얼굴 이미지를 증강하여 인식 성능을 개선하려는 시도는 아직 충분하지 않다.

초분광 이미지 증강 기법은 주로 항공 영상 분야^[11]에서 널리 사용되며, 여러 스펙트럼 대역을 동시에 처리하고 대역 간 상관성을 반영하여 물체 탐지 및 분류 성능을 향상시키는 데 활용된다. 그러나 얼굴 인식에서는 특정 대역에서 얼굴의 중요한 특징이 더 두드러지므로, 적절한 대역을 선택하여 사용하는 것이 성능 향상에 매우 중요하다. 기존 데이터 증강 기법들^[12]은 주로 RGB 이미지를 기반으로 회전, 크기 변환 등과 같은 방법을 사용하여 데이터의 다양성을 확보하지만, 이는 세 개의 채널만을 사용하기 때문에 넓은 스펙트럼 대역을 포함한 초분광 이미지에는 적합하지 않다. 초분광 이미지는 각 대역이 고유한 정보를 가지고 있으며, 이를 반영한 증강 기법이 필요하다. 본 연구에서는 특정 파장 대역을 선택해 초분광 얼굴 이미지를 증강한 후, 이를 기존 데이터셋과 결합하여 학습한 모델이 기존 데이터셋만을 사용해 학습한 모델보다 더 우수한 인식 성능을 보였음을 입증하였다.

2. 초분광 얼굴 이미지 증강

가. StyleGAN2-ADA

StyleGAN2-ADA^[6]는 적은 양의 데이터로도 효과적인 학습이 가능하도록 설계된 적대적 생성 네트워크(Generative Adversarial Network; GAN)의 변형 모델이다. 이 모델은 데이터 증강 기법을 통해 판별자(discriminator)의 과적합 문제를 해결하며, 증강된 이

미지를 사용해 판별자와 생성자(generator)를 동시에 학습시킨다. 증강 범위는 증강 확률(augmentation probability)을 통해 조절되며, 판별자의 출력에 따라 적응적으로 조정된다. ADA(Adaptive Discriminator Augmentation)는 판별자가 훈련 데이터와 생성된 데이터 간의 출력 차이를 기반으로 증강 강도를 자동으로 조절함으로써, 판별자의 과적합을 방지하는 역할을 한다. ADA에서 사용되는 과적합 판단 기준은 식 (1)과 같이 정의되며, D_i , D_v , D_g 는 각각 훈련 데이터, 검증 데이터, 생성된 데이터에 대한 판별자의 출력을 의미한다. r_v 값이 0에 가까울수록 판별자가 훈련 데이터에 과적합되지 않고 있음을 나타내며, 1에 가까울수록 과적합의 위험이 높아지고 있음을 의미한다.

$$r_v = \frac{E[D_i] - E[D_v]}{E[D_i] - E[D_g]} \quad (1)$$

ADA는 이러한 지표를 바탕으로 판별자의 과적합을 모니터링하고, 증강 확률을 동적으로 조정하여 모델의 학습 안정성을 유지한다. 본 연구에서는 StyleGAN2-ADA를 통해 기존 초분광 얼굴 데이터셋에 없는 새로운 객체를 생성하였다.

나. Face Pose Augmentation

Face Pose Augmentation^[7]은 얼굴 이미지의 포즈를 변환하여 다양한 각도와 방향에서 얼굴을 생성하는 기술이다. 이 과정에서 주로 사용되는 주요 기술 중 하나는 3DDFA(3D Dense Face Alignment)이다. 3DDFA^[13]는 얼굴 이미지에서 3차원 포즈를 추정하는 기술로, 밀도가 높은 점을 이용하여 얼굴의 형태를 재구성한다. 3DDFA에서 추정된 3차원 얼굴 모델은 식 (2)과 같이 표현되며, 여기서 $\bar{S}$ 는 평균 얼굴 상태, A_{id} 와 A_{exp} 는 각각 신원(identity)과 표정(expression)의 주성분이며, α_{id} 와 α_{exp} 는 해당 계수를 나타낸다.

$$S = \bar{S} + A_{id}\alpha_{id} + A_{exp}\alpha_{exp} \quad (2)$$

Face Pose Augmentation에서는 3차원 포즈 추정을 기반으로, 주어진 자세 변환 값에 따라 얼굴 이미지를 조정하여 다양한 포즈를 생성한다. 3차원 모델은 식 (3)과 같이 단순화된 정사영(simplified orthographic projection)을 통해 2D 이미지 평면으로 투영되며, f 는 스케일 인자, P_r 은 정사영 행렬, $R(q)$ 은 q 로부터 유도된 회전 행렬, t_{2d} 는 평행 이동 벡터이다.

$$V(p) = f \cdot P_r \cdot R(q) \cdot S + t_{2d} \quad (3)$$

이 기술은 정면이 아닌 시각에서의 객체 특징을 살릴 수 있기 때문에, 데이터 다양성을 증강시키기 위해 사용되었다^[7]. 변환된 회전 행렬은 식 (4)과 같이 계산되며, ΔR 은 주어진 자세 변환 값이다.

$$R_{aug} = \Delta R \cdot R \quad (4)$$

본 연구에서는 Face Pose Augmentation을 통해 동일한 객체를 다양한 포즈로 변형하여 데이터 다양성을 증강하는 연구를 수행하였다.

III. 제안하는 방법

그림 1은 제안된 방법을 포함한 전체 플로우 차트를 시각화한 것이다. 초분광 얼굴 데이터셋은 전처리 과정을 거쳐 StyleGAN2-ADA의 학습 데이터와 Face Pose Augmentation의 입력 이미지로 사용하였다. 이를 통해 증강된 초분광 얼굴 데이터셋은 기존 데이터

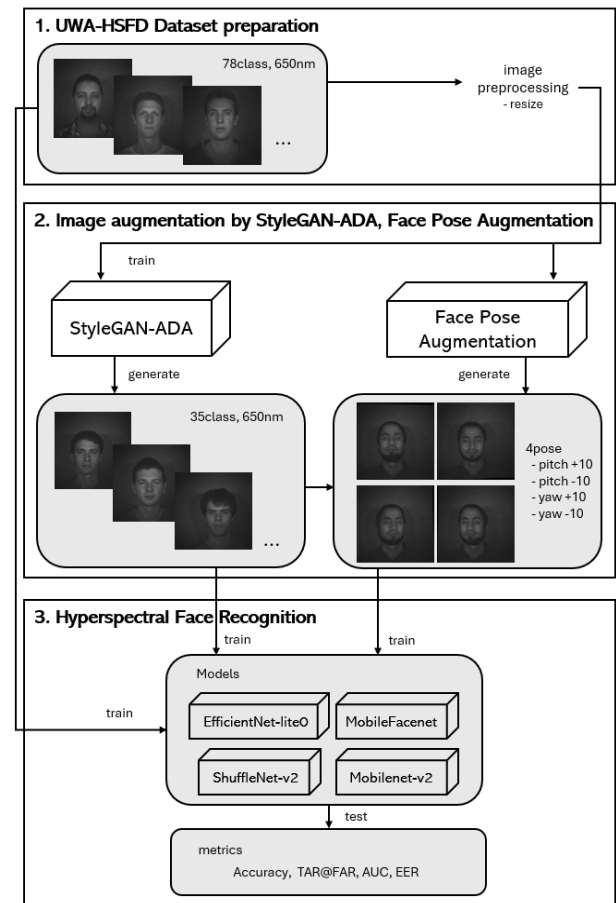


그림 1. 제안하는 방법이 포함된 전체 플로우 차트
Fig. 1. Full Flowchart Including Proposed Methods.

셋과 결합되어 얼굴 인식 시스템의 학습 데이터로 활용되었다.

1. 초분광 얼굴 이미지 증강 시스템

가. StyleGAN2-ADA를 통한 새로운 객체 증강

(1) 학습 데이터

새로운 객체를 증강하기 위해 사용된 데이터셋은 UWA-HSFD이다. 이 데이터셋은 400~720nm의 범위에서 10nm 간격으로 촬영된 33개의 파장 대역으로 구성되어 있으며, 78명의 주체가 여러 세션에서 촬영되었다. 학습에 사용된 데이터셋은 650nm 파장 대역을 기준으로 하였으며, 이 대역은 초분광 얼굴 인식에서 특히 우수한 성능을 보이는 것으로 알려져 있다^[14]. 또한, 초분광 얼굴 이미지를 촬영하는 과정에서 머리 움직임이나 눈 깜빡임으로 인해 파장 대역 간 정렬 오류가 발생할 수 있는데, 이를 고려하여 정렬이 비교적 정확하게 이루어진 650nm 대역을 선택하였다. 학습에 앞서 모든 이미지는 1024×1024 에서 256×256 으로 크기를 축소하여 전처리하였다.

(2) 모델 학습

전이 학습을 위해 Flickr-Faces-HQ Dataset (FFHQ) 70,000장의 이미지를 회색조(grayscale)로 변환하여 미리 학습된 모델(pretrained model)을 구축한 후, 이를 바탕으로 초분광 얼굴 데이터셋의 학습을 진행하였다. 회색조 영상은 색상 채널을 배제함으로써 복잡성을 줄이는 동시에, 주요한 얼굴 특징을 유지할 수 있어 효율적인 학습을 가능하게 한다. 이러한 접근법은 데이터 부족 문제를 보완하면서도, 650nm 대역을 포함한 특정 스펙트럼 밴드에서의 성능을 향상시킬 수 있는 가능성을 제시한다.

(3) 새로운 객체 판별법

StyleGAN2-ADA를 통해 생성된 객체가 기존 객체와 다르다는 것을 증명하기 위해, 기존 데이터셋만 사용하여 임베딩 모델 학습을 수행하였다. 임베딩 모델을 통해 얼굴 이미지 간의 유사성을 정량화하고, 이를 통해 객체를 분류하였다. 임베딩 모델은 MobileFacenet을 기반으로 학습되었으며, 동일한 객체 간의 거리는 최소화하고, 다른 객체 간의 거리는 최대화되도록 임베딩 공간이 학습되었다^[15, 16].

그림 2는 학습된 임베딩 모델을 사용하여 동일한 객

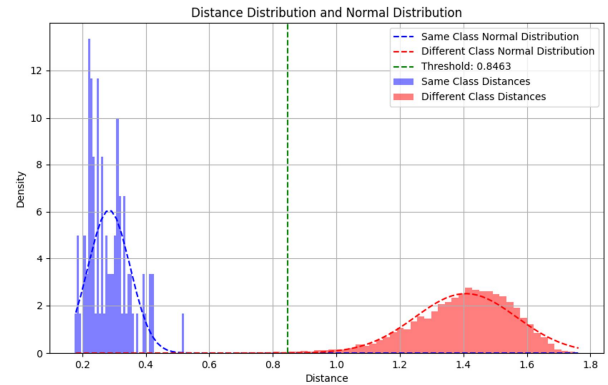


그림 2. 임베딩 모델을 통한 객체 분류 임계값 시각화

Fig. 2. Visualization of Classification Thresholds for Subjects using an Embedding Model.

체 간의 거리와 다른 객체 간의 거리를 시각화한 결과를 보여준다. 빨간색 분포는 다른 객체 간의 거리를, 파란색 분포는 동일한 객체 간의 거리를 나타낸다. 두 분포를 정규분포화한 후, 겹치는 지점에서의 값을 기준으로 이미지가 동일 객체인지 여부를 판별하는 임계값(threshold)을 도출하였다. 도출된 임계값은 0.8463이다. StyleGAN2-ADA로 생성된 이미지와 학습된 이미지 간의 거리를 계산한 결과, 0.8463 이상의 거리를 가지는 이미지는 서로 다른 객체로 분류되었다.

나. Face Pose Augmentation을 통한 동일한 객체의 다양한 포즈 증강

기존 초분광 얼굴 데이터셋과 StyleGAN2-ADA를 사용해 생성된 초분광 얼굴 데이터셋의 포즈를 증강하기 위해, 3DDFA를 사용하여 얼굴의 3차원 위치와 포즈를 추정하였다. 이러한 추정을 바탕으로, 주어진 자세 변환 값에 따라 얼굴 이미지를 조정하여 다양한 포즈의 이미지를 생성하였다. 증강된 이미지는 상하 회전(pitch)이 +10도와 -10도, 좌우 회전(yaw)이 +10도와 -10도로 구성된다. 즉, 한 객체에 대해 총 4가지 다른 시점에서의 이미지를 증강한 것이다. Face Pose Augmentation은 미리 학습된 모델을 사용하였으며, 입력 이미지의 크기는 256×256 이다.

2. 초분광 얼굴 인식 시스템

가. 데이터셋 구성

표 1은 학습 데이터셋과 테스트 데이터셋의 객체 수와 총 샘플 수를 나타낸다. 학습 데이터셋은 총 115개의 기존 초분광 얼굴 데이터셋과 750개의 기존 및 증강된 초분광 얼굴 데이터셋으로 구성되어 있다. 증강된

초분광 얼굴 데이터셋에는 35명의 새로운 객체가 포함되어 있으며, 각 객체마다 4가지 다양한 포즈의 얼굴 이미지가 포함되어 있다. 두 모델의 성능을 비교하고 평가하기 위해 동일한 테스트 데이터셋을 사용하였으며, 테스트 데이터셋은 15개의 객체와 총 30개의 샘플로 구성되어 있다.

표 1. 초분광 얼굴 인식 시스템의 학습 데이터셋 및 테스트 데이터셋

Table 1. Training and Testing Datasets for the Hyperspectral Face Recognition System.

	Training Dataset		Test Dataset
	Original Dataset	Original and Augmented Dataset	
Subjects	63	98	15
Total Samples	115	750	30

나. 모델 학습

모델을 학습하기 전에 MTCNN(Multi-task Cascaded Convolutional Networks)을 사용하여 얼굴 부분을 검출하고, 정렬된 얼굴 영역을 추출하였다. 추출된 얼굴 영역은 학습 데이터셋으로 사용되며, 특징 추출기를 통해 생성된 임베딩 벡터의 손실 함수를 최적화하는 방식으로 학습이 진행되었다. 이때 손실 함수는 동일한 객체 간의 거리를 최소화하고, 다른 객체 간의 거리를 최대화하는 방향으로 최적화한다. 본 연구에서 사용된 손실 함수는 ArcFace^[8]로, ArcFace는 특징 벡터를 단위 길이로 정규화한 후 객체 간의 각도에 마진을 추가하여, 소프트 맥스 함수를 통해 객체별 확률을 계산하고 크로스 엔트로피 손실을 최소화하도록 학습된다. 식 (5)은 ArcFace의 손실 함수식을 나타내며, N 은 샘플 수를 의미한다. θ_{y_i} 는 정답 객체 y_i 에 대한 각도이며, θ_j 는 y_i 가 아닌 다른 모든 객체에 대한 각도이다. n 은 y_i 를 제외한 객체 수이며, s 는 스케일링 파라미터, m 은 마진 파라미터이다.

$$L_1 = -\frac{1}{N} \sum_{i=1}^N \log \frac{e^{s(\cos(\theta_{y_i} + m))}}{e^{s(\cos(\theta_{y_i} + m))} + \sum_{j=1, j \neq y_i}^n e^{s \cos \theta_j}} \quad (5)$$

학습 과정은 그림 3에 나타나 있다. 사용된 DCNN은 MobileFaceNet^[17], Mobilenet-v2^[18], Shufflenet-v2^[19], EfficientNet-Lite0^[20]이다. 이 DCNN들은 이미지를 입력으로 받아 각 객체의 특징을 추출하고 임베딩

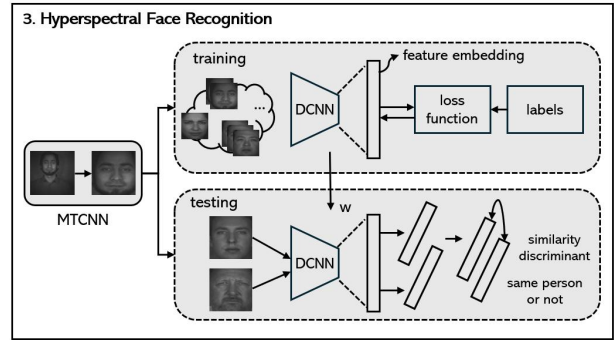


그림 3. 초분광 얼굴 인식 시스템을 위한 학습 및 테스트 과정

Fig. 3. Training and Testing Process for Hyperspectral Face Recognition System.

벡터로 변환한다. MobileFaceNet과 MobileNet-v2는 깊이별 분리 가능한 합성곱(Depthwise Separable Convolution)과 역잔차 블록(Inverted Residual Blocks)을 사용하여 파라미터 수와 연산량을 줄이면서도 모델의 표현력을 유지한다. ShuffleNet-v2는 채널 셔플(Channel Shuffle)과 그룹 합성곱(Group Convolution)을 통해 연산 복잡도를 낮추면서도 효율적인 특징 추출을 가능하게 한다. EfficientNet-Lite0는 복잡 스케일링(Compound Scaling)을 통해 네트워크의 깊이, 너비, 해상도를 최적화하고, 압축 및 활성화 블록(Squeeze-and-Excitation Blocks)을 통해 채널 간 상호작용을 강화하여 모델의 특징 추출 능력을 향상시킨다. 모든 모델은 SGD(Stochastic Gradient Descent) 최적화 알고리즘을 사용했으며, 학습률은 0.1, 모멘텀은 0.9로 설정하였다. 또한, 가중치 감쇠(weight decay)를 $5e-4$ 로 설정하여 과적합을 방지하였으며, MultiStepLR 알고리즘을 사용하여 20, 35, 45 에포크(epoch)마다 학습률을 0.1씩 감소시켜 모델의 수렴을 촉진하였다.

IV. 실험 및 실험 결과

1. 초분광 얼굴 이미지 증강 시스템

가. StyleGAN2-ADA를 통한 새로운 객체 증강

(1) 모델 평가

StyleGAN2-ADA로 학습된 모델을 평가하기 위해 생성된 이미지와 실제 이미지 간의 차이를 측정하는 지표로 FID-Score^[21]를 사용하였다. FID-Score는 값이 낮을수록 생성된 이미지가 실제 이미지와 더 유사함을 의미하며, 더 좋은 성능을 나타낸다. 학습 시 증강 확률을

0.6, 생성자와 판별자의 학습률은 0.0025로 설정되었다. 먼저, 전이 학습 없이 초분광 얼굴 데이터셋의 단일 스펙트럼 밴드를 사용해 학습한 결과, FID-Score는 54.61로 측정되었다. 이후 전이 학습을 적용하여 동일한 데이터셋을 재학습한 결과, FID-Score는 40.12로 감소하였다. 이는 전이 학습이 FID-Score를 낮춤으로써, 단일 밴드 초분광 이미지에서 StyleGAN2-ADA의 성능을 유의미하게 향상시켰음을 보여준다.

StyleGAN2-ADA의 입력으로는 기존 학습 데이터셋이 사용되었으며, 이를 바탕으로 새로운 객체들이 증강되었다. 생성된 객체의 차이를 분석하기 위해 학습된 임베딩 모델을 활용한 결과, 생성된 객체 50개 중 35개는 기존 초분광 얼굴 데이터셋과의 거리가 0.8463보다 컸고, 15개는 0.8463보다 작았다. 그림 4는 기존 초분광 얼굴 데이터셋과 증강된 초분광 얼굴 데이터셋의 예시를 보여준다. 그림 (a)에서는 왼쪽이 기존 초분광 얼굴 이미지를, 오른쪽이 기존 데이터셋과의 거리가 0.8463보다 큰 생성된 이미지를 나타낸다. 반면, 그림 (b)에서는 왼쪽이 기존 이미지를, 오른쪽이 거리가 0.8463보다 작은 이미지를 보여준다. 이를 통해 증강된 초분광 얼굴 이미지가 기존 데이터셋과 유사하거나 차이가 나는 예시를 시각적으로 확인할 수 있다.

나. Face Pose Augmentation을 통한 동일한 객체의 다양한 포즈 증강

그림 5는 기존 초분광 얼굴 데이터셋과 포즈가 증강된 초분광 얼굴 데이터셋의 예시를 보여준다. 그림 5의 (a)는 기존 초분광 얼굴 이미지를, (b), (c), (d)는 각각 얼굴을 위로 10도, 좌로 10도, 우로 10도 회전시킨 이미지를 나타낸다. 이 그림을 통해 기존 초분광 얼굴 데이터셋과 다른 포즈로 증강된 이미지를 확인할 수 있다.

2. 초분광 얼굴 인식 시스템

가. 모델 평가

ArcFace로 학습된 모델들을 평가하기 위해서 개방형 얼굴 검증(open-set face verification)을 수행하였다^[22]. 이는 학습 데이터셋에 없는 객체를 포함한 테스트 데이터셋을 사용하여, 두 얼굴 이미지가 동일한 사람인지 여부를 판별하는 작업을 진행하는 평가 방법이다. 학습된 특징 추출기를 사용하여 두 이미지의 임베딩 벡터를 추출하고, 이를 통해 두 얼굴이 동일한 사람인지 판단한다. 평가 지표로는 Accuracy, AUC 등을 사용하였으며, 자세한 내용은 표 2에 나와 있다.

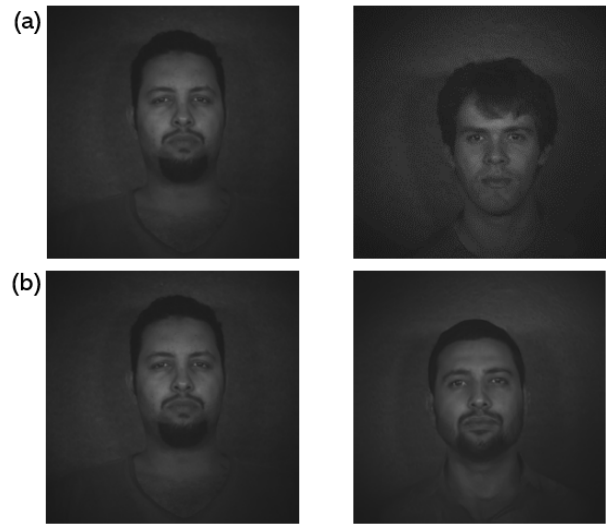


그림 4. (a) 기존 초분광 얼굴 이미지(왼쪽)와 거리가 0.8463보다 큰 생성된 이미지(오른쪽); (b) 기존 이미지(왼쪽)와 거리가 0.8463보다 작은 생성된 이미지(오른쪽)

Fig. 4. (a) Original Hyperspectral Face Image (left) and Generated Image with a distance greater than 0.8463 (right); (b) Original Image (left) and Generated Image with a distance less than 0.8463 (right).

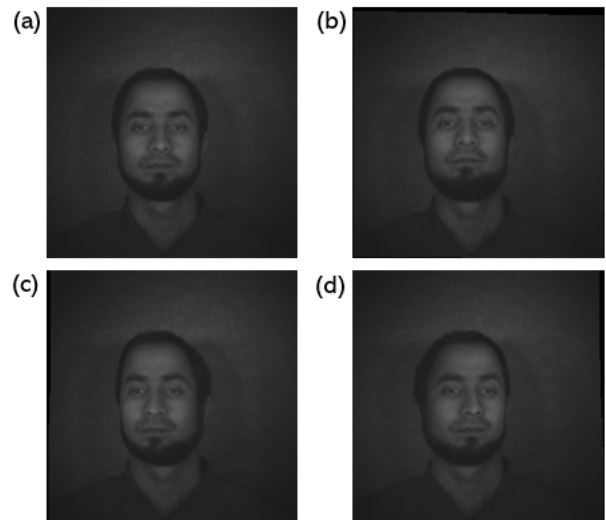


그림 5. 기존 초분광 얼굴 이미지 (a) 및 Face Pose Augmentation을 통해 증강된 초분광 얼굴 이미지 (b), (c), (d)

Fig. 5. Original Hyperspectral Face Image (a) and Augmented Hyperspectral Face Image via Face Pose Augmentation (b), (c), (d).

테스트 데이터셋은 동일한 객체 이미지 쌍 19개와 다른 객체 이미지 쌍 416개로 구성되었다. 표 3은 기존 초분광 얼굴 데이터셋으로 학습된 모델(base model)과 기존 및 증강된 초분광 얼굴 데이터셋을 사용하여 학습

표 2. 얼굴 검증을 평가하기 위해 사용된 메트릭
Table 2. Metrics Used to Evaluate Face Verification.

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 • Effective for imbalanced datasets by evaluating performance across thresholds
TAR@FAR	• True Accept Rate at a given 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

된 모델(proposed model)의 성능을 비교한 결과를 보여준다.

결과적으로, 네 가지 모델 모두 기존 초분광 얼굴 데이터셋만 사용하는 것보다 증강된 초분광 얼굴 데이터셋을 함께 사용한 모델에서 더 우수한 성능을 보였다. 특히, Shufflenet-v2의 AUC 성능은 0.7098에서 0.9619로 증가하여 0.2521 향상되었다. 가장 높은 AUC 성능은 Mobilenet-v2에서 나타났으며, 0.9906을 기록하였다. 이는, 모델이 높은 Accuracy와 민감도(sensitivity)를 지니고 있음을 의미한다.

Accuracy 성능은 평균적으로 0.0735 향상되었으며, 가장 높은 Accuracy는 MobileFacenet을 통해 0.9609로 나타났다. 또한, TAR@FAR 0.100에서는 Shufflenet-v2의 성능이 가장 크게 향상되어 0.6316 증가하였다. 마지막으로, EER은 평균적으로 0.1009 낮아졌으며, 가장 낮은 값은 Mobilenet-v2에서 0.0252로 나타났다. 이를 통해, 본 논문에서는 기존 및 증강된 초분광 얼굴 데이터셋을 사용한 모델이 더 우수한 성능을 발휘함을 입증하였다.

V. 결 론

본 논문에서는 제한적인 초분광 얼굴 데이터셋을 증강하기 위한 방법을 제안하였다. 딥러닝 기반의 이미지 생성 모델인 StyleGAN2-ADA를 활용하여 새로운 얼굴을 생성하였으며, 생성된 이미지 50장 중 기존 이미

표 3. 개방형 얼굴 검증을 통한 성능 평가
Table 3. Performance Evaluation via Open-set Face Verification.

Model	Metric	Base Model	Proposed Model
MobileFacenet	Accuracy(↑)	0.9448	0.9609
	AUC(↑)	0.9530	0.9854
	TAR@FAR(↑)	0.001	0.2632
		0.010	0.7368
		0.100	0.8421
Mobilenet-v2	EER(↓)	0.1606	0.1116
	Accuracy	0.9241	0.9517
	AUC	0.9467	0.9907
	TAR@FAR	0.001	0.0526
		0.010	0.4211
		0.100	0.8421
Shufflenet-v2	EER	0.1151	0.0252
	Accuracy	0.7103	0.8344
	AUC	0.7098	0.9619
	TAR@FAR	0.001	0.0526
		0.010	0.0526
		0.100	0.2105
EfficienNet-Lite0	EER	0.2770	0.1151
	Accuracy	0.7517	0.8781
	AUC	0.8467	0.9262
	TAR@FAR	0.001	0.0000
		0.010	0.1053
		0.100	0.5263
	EER	0.2170	0.1139

지와의 임베딩 벡터 거리를 기준으로 35개의 새로운 객체를 추가하였다. 또한, 동일 객체의 다양한 포즈를 증강하기 위해 Face Pose Augmentation 기술을 적용하여 각 객체당 4개의 포즈를 증강시켰다. 이를 통해 데이터의 다양성을 증대시켰고, 해당 데이터를 얼굴 인식 모델의 학습에 활용하였다. 실험 결과, 평가한 네 가지 모델 모두에서 증강된 초분광 얼굴 데이터셋을 함께 사용한 얼굴 인식 모델이 기존 데이터셋만을 사용한 모델보다 우수한 성능을 보였다.

그러나 본 연구는 선택된 일부 파장 대역만을 활용한 증강 방법에 제한이 있으며, 이는 연구의 한계로 남는다. 향후 연구에서는 여러 파장 대역을 통합한 증강 기법을 도입하고, 다양한 대역을 결합하여 보다 정교한 초분광 얼굴 인식 모델을 제안할 계획이다. 또한, 본 연구에서 사용된 테스트 데이터셋의 수가 제한적이었기

때문에, 증강된 초분광 얼굴 데이터셋이 얼굴 인식 성능에 미치는 영향을 15명의 객체로만 평가한 것은 연구의 한계로 남는다. 이러한 소규모 평가 결과만으로는 제안된 모델의 일반화 성능을 충분히 평가하기는 어렵다. 아울러, 본 연구에서 제안한 증강 기법이 CMU-HSFD, PolyU-HSFD 등과 같은 데이터셋에도 동일한 성능 향상을 보일 수 있을지는 추가적인 검증이 필요하다. 따라서, 향후 연구에는 다양한 데이터셋을 확보하여 증강 기법의 범용성과 일반화 성능을 검증할 계획이다.

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] X. Zhang and H. Zhao, "Hyperspectral-cube-based mobile face recognition: A comprehensive review," *Information Fusion*, vol. 74, pp. 132-150, 2021.
- [3] L. J. Denes, P. Metes, and Y. Liu, CMU-RI-TR-02-25, 2002.
- [4] W. Di, L. Zhang, D. Zhang, and Q. Pan, "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.
- [5] M. Uzair, A. Mahmood, and A. Mian, "Hyperspectral face recognition with spatospectral information fusion and PLS regression," *IEEE Transactions on Image Processing*, vol. 24, no. 3, pp. 1127-1137, 2015.
- [6] T. Karras, M. Aittala, J. Hellsten, S. Laine, J. Lehtinen, and T. Aila, "Training generative adversarial networks with limited data," *Advances in Neural Information Processing Systems*, vol. 33, pp. 12104-12114, 2020.
- [7] S. Cheng et al., "Towards pose-invariant lip-reading," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020: IEEE, pp. 4357-4361.
- [8] 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*, 2019, pp. 4690-4699.
- [9] Z. Xie, Y. Li, J. Niu, L. Shi, Z. Wang, and G. Lu, "Hyperspectral face recognition based on sparse spectral attention deep neural networks," *Optics Express*, vol. 28, no. 24, pp. 36286-36303, 2020.
- [10] V. Sharma, A. Diba, T. Tuytelaars, and L. Van Gool, "Hyperspectral CNN for image classification & band selection, with application to face recognition," *Technical report KUL/ESAT/PSI/1604, KU Leuven, ESAT, Leuven, Belgium*, 2016.
- [11] W. Li, C. Chen, M. Zhang, H. Li, and Q. Du, "Data augmentation for hyperspectral image classification with deep CNN," *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 4, pp. 593-597, 2018.
- [12] S. Yang, W. Xiao, M. Zhang, S. Guo, J. Zhao, and F. Shen, "Image data augmentation for deep learning: A survey," *arXiv preprint arXiv:2204.08610*, 2022.
- [13] J. Guo, X. Zhu, Y. Yang, F. Yang, Z. Lei, and S. Z. Li, "Towards fast, accurate and stable 3D dense face alignment," in *European Conference on Computer Vision*, vol. 20, pp. 152-168, 2020.
- [14] F. Taherkhani, J. Dawson, and N. M. Nasrabadi, "Deep sparse band selection for hyperspectral face recognition," *Hyperspectral Image Analysis: Advances in Machine Learning and Signal Processing*, pp. 319-350, 2020.
- [15] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 815-823, 2015.
- [16] Y. Taigman, M. Yang, M. A. Ranzato, and L. Wolf, "Deepface: Closing the gap to human-level performance in face verification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1701-1708, 2014.
- [17] S. Chen, Y. Liu, X. Gao, and Z. Han, "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 13*, pp. 428-438, 2018.
- [18] 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.
- [19] 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.
- [20] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *International Conference on Machine Learning*, pp. 6105-6114, 2019.
- [21] 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," *Advances in Neural Information Processing Systems*, vol. 30, 2017.
- [22] M. Wang and W. Deng, "Deep face recognition: A survey," *Neurocomputing*, vol. 429, pp. 215-244, 2021.

 저 자 소 개



최 영 인(비회원)
2022년 목포대학교 컴퓨터공학과
학사 졸업.
2024년 광주과학기술원 AI 대학원
석사 졸업.
2024년~광주과학기술원
AI 대학원 박사과정

<주관심분야: 인공지능, 기계학습 및 딥러닝, 컴
퓨터 비전, 초분광 이미징>



강 주 성(비회원)
2012년 아주대학교 전자공학부
학사 졸업.
2024년 광주과학기술원
정보통신공학부 박사 졸업.
2024년~한국전자통신연구원
선임연구원

<주관심분야: 인공지능, 추론분석, 강화학습, 온
디바이스AI>



이 지 오(비회원)
2022년 충북대학교 전자공학부
학사 졸업.
2022년~광주과학기술원
AI 대학원 석,박통합과정

<주관심분야: 인공지능, 기계학습 및 딥러닝, 신
호처리>



신 영 학(비회원)
2009년 광운대학교
전자통신공학과 학사 졸업.
2011년 광주과학기술원
정보통신공학부 석사 졸업.
2016년 광주과학기술원
정보통신공학부 박사 졸업.

2020년~목포대학교 컴퓨터공학과 교수
<주관심분야: 인공지능, 기계학습 및 딥러닝>



이 흥 노(정회원)
1993년 University of
California,
Los Angeles
전자공학과 학사 졸업.
1994년 University of
California,
Los Angeles
전자공학과 석사 졸업.

1999년 University of California, Los Angeles
전자공학과 박사 졸업.

2009년~현재 광주과학기술원
전기전자컴퓨터공학부 교수

<주관심분야: 정보이론, 신호처리, 통신/네트워
크 이론, 블록체인, 인공지능>