

<https://doi.org/10.7236/JIIBC.2024.24.1.141>

JIIBC 2024-1-21

얼굴 영역 추출 시 여유값의 설정에 따른 개성 인식 모델 정확도 성능 분석

Performance Analysis for Accuracy of Personality Recognition Models based on Setting of Margin Values at Face Region Extraction

구욱*, 한규원**, 김봉재***

Qiu Xu*, Gyuwon Han**, Bongjae Kim***

요약 최근 개인의 성향을 반영한 맞춤형 서비스가 각광 받고 있다. 이와 관련하여 개인의 개성을 인식하고 활용하고자 하는 연구가 지속적으로 이루어지고 있다. 각 개인의 개성을 인식하고 평가하는 방법은 다수가 있지만, OCEAN 모델이 대표적으로 사용된다. OCEAN 모델로 각 개인의 개성을 인식할 때 언어적, 준언어적, 비언어적 정보를 이용하는 멀티 모달리티 기반 인공지능 모델이 사용될 수 있다. 본 논문에서는 비언어적 정보인 사용자의 표정을 기반으로 OCEAN을 인식하는 인공지능 모델에서 영상 데이터에서 얼굴 영역을 추출할 때 지정하는 얼굴 영역 여유값(Margin)에 따른 개성 인식 모델 정확도 성능을 분석한다. 실험에서는 2D Patch Partition, R2plus1D, 3D Patch Partition, 그리고 Video Swin Transformer에 기반한 개성 인식 모델을 사용하였다. 얼굴 영역 추출 시 여유값을 60으로 사용했을 때 1-MAE 성능이 0.9118로 가장 우수하였다. 따라서 개성 인식 모델의 성능을 최적화하기 위해서는 적절한 여유값을 설정해야 함을 확인하였다.

Abstract Recently, there has been growing interest in personalized services tailored to an individual's preferences. This has led to ongoing research aimed at recognizing and leveraging an individual's personality traits. Among various methods for personality assessment, the OCEAN model stands out as a prominent approach. In utilizing OCEAN for personality recognition, a multi modal artificial intelligence model that incorporates linguistic, paralinguistic, and non-linguistic information is often employed. This paper examines the impact of the margin value set for extracting facial areas from video data on the accuracy of a personality recognition model that uses facial expressions to determine OCEAN traits. The study employed personality recognition models based on 2D Patch Partition, R2plus1D, 3D Patch Partition, and Video Swin Transformer technologies. It was observed that setting the facial area extraction margin to 60 resulted in the highest 1-MAE performance, scoring at 0.9118. These findings indicate the importance of selecting an optimal margin value to maximize the efficiency of personality recognition models.

Key Words : Artificial Intelligence, Face Extraction, OCEAN, MTCNN, Personality Recognition

*준회원, 충북대학교 전기·전자·정보·컴퓨터학부

**준회원, 충북대학교 컴퓨터공학과

***정회원, 충북대학교 컴퓨터공학과

접수일자 2023년 11월 27일, 수정완료 2024년 1월 18일
게재확정일자 2024년 2월 9일

Received: 27 November, 2023 / Revised: 18 January, 2024 /

Accepted: 9 February, 2024

*Corresponding Author: bjkim@chungbuk.ac.kr

Department of Computer Engineering, Chungbuk National University, Korea

I. 서 론

현재 인공지능 분야의 발전으로 점점 다양한 서비스가 가능해지고 개인화되고 있다. 이와 같은 지능화 및 개인화 서비스에서 사용자의 개성을 인식하고 이를 기반으로 최적의 콘텐츠나 사용자 경험을 제공할 수 있다. 이와 같은 개성 인식 기술에는 비언어적 정보인 사람의 표정과 같은 시각적 정보를 활용할 수 있다. 또한 인공지능 모델을 기반으로 시각적 정보를 활용하는 다양한 연구가 이루어지고 있다.

개성 인식 연구에서 OCEAN 모델은 개인의 개성을 측정하고 계량화할 수 있는 이론이다. 즉, OCEAN 모델은 개인의 개성을 이해하고 설명하는데 사용되는 이론이다. 이 모델은 다음과 같은 다섯 가지 주요 특성으로 구성되어 있다. 개방성(Openness)은 새로운 경험에 대한 개방성과 창의성을 나타낸다. 성실성(Conscientiousness)은 조직력, 신중함, 그리고 신뢰성을 의미한다. 외향성(Extraversion)은 대인 관계에서의 활동성과 긍정적인 감정을 의미한다. 친화성(Agreeableness)은 협조적이고 친절한 성향을 의미한다. 마지막으로 신경성(Neuroticism)은 감정의 안정성과 스트레스에 대한 반응을 나타낸다. 여러 연구에서 OCEAN 이론이 개인의 행복, 정신적 건강, 종교성, 정체성 그리고 가족, 친구, 연인과의 관계, 직업 선택, 직무 만족도, 성과, 사회 참여, 범죄 행동, 정치적 입장 등 다양한 요소를 잘 예측한다는 것이 입증되었다^[1]. 따라서 각 개인의 개성을 OCEAN 모델을 이용하여 정량화할 수 있다. 이런 정량화된 개성 정보를 이용하여 각 개인의 특성에 맞는 개인화된 서비스가 가능하다.

이와 같은 이유로 다양한 인공지능 모델과 기술을 이용하여 OCEAN을 예측하는 연구가 활발하게 이루어지고 있습니다. 최근에는 단일 모달리티 정보보다 멀티 모달리티 정보를 이용한 개성 인식 모델도 활발하게 연구되고 있다^[2]. 멀티 모달리티 정보를 활용할 때는 언어적, 준언어적, 비언어적 정보를 활용한다. 주로 얼굴과 표정 정보, 음성 정보, 발화에 따른 텍스트 정보가 사용된다.

멀티 모달리티에 기반하는 개성 인식 모델에서는 화자의 얼굴과 표정에 각 개인의 개성이 많이 포함되기 때문에 얼굴 및 표정 정보와 같은 비언어적 시각 정보에 초점을 둔 개성 인식 모델이 기본적으로 포함된다. 이때 순수 얼굴에 대한 정보만을 활용하는 것 보다 머리 스타일과 같은 얼굴 이외의 특징을 함께 사용하는 것이 OCEAN 기반 개성 인식 모델의 정확도 성능을 높이는 데 도움이 될 수 있다^[3]. 따라서 개성 분석을 원하는 화자가 있는 영

상에서 화자의 얼굴 정보를 추출할 때 얼굴 영역 추출 시 설정하는 여유값을 적절하게 선택하는 것이 필요하다.

본 연구에서는 얼굴 영역 추출 시 여유값의 설정에 따른 개성 인식 모델의 정확도 성능을 분석했다. 데이터 전처리 과정에서 얼굴 영역 추출을 위해 MTCNN을 사용했다. MTCNN은 이미지나 비디오에서 얼굴 특성을 빠르게 인식하는 데 효과적이다. MTCNN에서 설정할 수 있는 여유값은 얼굴 이미지 주변의 배경 크기를 조절하는 데 사용될 수 있다. 예를 들어 여유값을 0으로 설정하면 MTCNN은 얼굴 영역만을 정확히 추출한다. 여유값을 높이면 얼굴 주변의 배경이 더 많이 포함된 얼굴 영역 이미지를 얻을 수 있다. 본 논문에서는 여유값을 0에서 80까지 변화시키면서 개성 인식 모델의 1-MAE 측면에서의 정확도를 분석하였다. 실험 결과, 여유값이 60일 때 1-MAE 측면의 성능이 최적이었으며, 그 성능은 0.9118이다. 반대로 최악의 경우는 여유값이 0일 때였으며, 1-MAE 측면에서의 성능은 0.9096이다. 따라서 개성 인식 모델의 성능을 최적화하기 위해서는 얼굴 영역 추출 시 적절하게 여유값을 설정하는 것이 필요하다는 것을 실험적으로 검증하고 확인하였다. 이와 같은 최적화를 통해 개성 인식에 기반한 개인화 서비스의 성능과 사용자 경험을 향상시키는데 도움이 될 수 있을 것이다.

이후 본 논문의 구성은 다음과 같다. 2장에서는 관련 연구에 대하여 설명한다. 3장에서는 본 연구에서 사용한 개성 인식 모델에 대해서 자세하게 설명한다. 4장에서는 실험에서 사용한 데이터셋(Dataset)과 데이터셋의 전처리 방법을 자세하게 설명한다. 또한, 얼굴 영역 추출 시 여유값 설정에 따른 개성 인식 모델의 1-MAE 측면의 성능을 비교 분석한다. 마지막으로 5장에서는 본 논문의 결론을 맺는다.

II. 관련 연구

1. MTCNN 모델

MTCNN^[4]은 얼굴 감지를 위한 신경망 모델이다. MTCNN을 이용하여 개성 인식 모델에 사용하는 얼굴 영역을 추출하는데 사용할 수 있다. 이 모델은 처음에는 간단한 네트워크로 대략적인 얼굴 위치를 빠르게 파악하고, 이후 보다 정교한 네트워크를 통해 얼굴을 상세히 감지한다. P-Net, R-Net, 그리고 O-Net을 통해 빠르고 정확하게 얼굴을 감지한다. 먼저, 이미지 피라미드를 사용하여 원본 이미지의 크기를 다양하게 조절하고 P-Net

네트워크로 다수의 후보 영역을 생성한다. 이후 R-Net 네트워크로 후보 영역을 선별하며, 마지막으로 O-Net을 활용해 남은 후보 영역을 세밀하게 선별한다^[5].

2. Face_recognition 라이브러리

Face_recognition^[6]은 세계에서 가장 단순하면서도 효과적인 얼굴 인식 라이브러리 중 하나로 알려져 있으며, Labeled Faces in the Wild 얼굴 데이터셋에 대한 테스트에서 99.38%의 정확도를 보였다. 그러나 Face_recognition의 경우, 어린이와 아시아인에 대한 얼굴 인식 정확도는 상대적으로 낮다. Face_recognition을 이용하여 이미지 내의 얼굴 위치를 파악할 수 있지만, 라이브러리 자체에서 얼굴 영역을 추출하고 저장하는 별도의 기능을 제공하지 않는다. 또한, 이 라이브러리는 얼굴의 주요 특징점을 감지할 수 있는 기능을 제공하여 해당 얼굴이 누구인지를 판별할 수 있다. 따라서 다양한 얼굴 인식을 통한 사용자 인증 및 확인에 기반한 다양한 서비스에 활용될 수 있다.

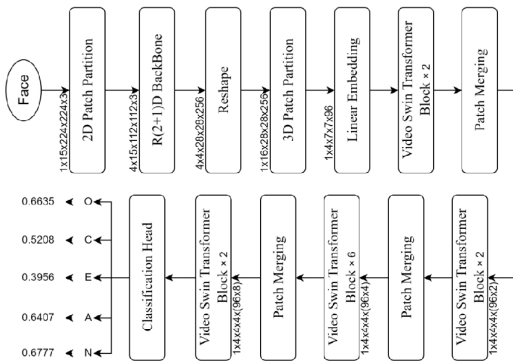


그림 1. 그림 1. 사용한 모델 구조
 Fig. 1. The model architecture

3. Dlib

Dlib 라이브러리는 Davis King에 의해 개발되어 널리 사용되는 컴퓨터 비전 및 머신 러닝 소프트웨어 라이브러리로, 효과적인 얼굴 영역 탐지 기능을 제공한다. 또한, Dlib는 사전 훈련된 얼굴 특징 감지 모델을 포함하고 있으며, 얼굴의 68개 주요 지점을 효율적으로 식별할 수 있게 해준다^[7]. Dlib의 주요 장점은 사용의 용이성과 유연성으로, NumPy와 OpenCV 등 다른 파이썬 라이브러리와 호환성이 뛰어나다는 것이다. 그러나 사용자가 최종적으로 추출된 얼굴을 영역과 관련 정보를 직접 후

처리하는 단계가 필요하다.

III. 사용한 개성 인식 모델

본 논문에서 사용한 개성 인식 모델은 그림 1과 같다. 그림 1에서 확인할 수 있듯이, 본 논문에서 사용한 개성 인식 모델의 주요 구성 요소는 2D Patch Partition, R2plus1D, 3D Patch Partition, Video Swin Transformer이다. 또한 얼굴 영역 추출 시 여유값의 설정에 따른 개성 인식 모델의 정확도에 초점을 두기 위하여 멀티 모달리티에 기반하는 개성 인식 모델이 아니라 얼굴 및 표정 정보에 기반하는 단일 모달리티에 기반하는 개성 인식 모델이다.

그림 1과 같이, 전처리를 통해 추출된 얼굴 영역 이미지는 2D Patch Partition을 거쳐 동일한 크기의 네 부분으로 분할한다. 분할된 각 4개의 영역은 순서대로 R2plus1D Backbone에 입력된다. 이후 각 4개 영역의 R2plus1D Backbone의 출력 결과가 다시 하나로 합쳐진다. 이후 합쳐진 데이터는 Reshape과 3D Patch Partition을 거쳐 Video Swin Transformer로 입력된다. 마지막으로 OCEAN 기반 인식 결과를 출력한다. 출력된 결과는 OCEAN 각각 0부터 1 사이로 나타난다.

1. 2D Patch Partition

추출된 얼굴 영역 224×224 크기의 데이터를 4개의 영역으로 균등하게 나눈다. 따라서 각 서브 영역의 크기는 각각 크기는 112×112 이다. Tensor 형식의 데이터를 다루기 위해 이미지의 높이와 너비를 나누기 위해 언폴드(Unfold) 방법을 사용하였다.

2. R2plus1D

R2plus1D는 2D Convolution과 1D Convolution을 사용하여 전통적인 3D Convolution을 대체할 수 있는 기법이다. 시간과 공간 두 부분을 두 단계로 분리하여 연산을 처리하는 개념이다. 3D Convolution에서 사용되는 필터는 $t \times d \times d$ 형태로, t 는 시간 차원을, d 는 공간 차원의 폭과 높이를 나타낸다. $(2 + 1)D$ Convolution은 공간을 나타내는 2D Convolution과 시간을 나타내는 1D Convolution으로 분할된다. 2D Convolution의 필터 크기는 $1 \times d \times d$ 이며, 1D Convolution의 필터 크기는 $t \times 1 \times 1$ 이다^[8]. R2plus1D 모델에는 네 개의

잔여 블록 그룹(Residual Block Groups)이 있으며, 사용한 모델에서는 앞의 세 개의 잔여 블록 그룹만 사용하여 출력 채널 수를 256으로 설정하여 사용하였다.

3. 3D Patch Partition

R2Plus1D을 통해 얻은 데이터는 256 크기를 가지는 벡터가 $16 \times 28 \times 28$ 로 구성된다. 각 $4 \times 4 \times 4$ 를 3D Patch를 통해 하나의 토큰으로 처리하고, 각 토큰은 96 크기의 벡터로 변환된다. 따라서 그림 1과 같이 3D Patch Partition 과정을 거친 후의 각 토큰은 96 크기의 벡터이며, 토큰은 $\frac{16}{4} \times \frac{28}{4} \times \frac{28}{4}$ 개가 된다.

4. Video Swin Transformer

Transformer는 자연어 처리 분야에서 뛰어난 성능을 보여주었다. Transformer는 비전(Vision) 분야에서도 뛰어난 성과를 보이고 있다. 또한 학습에 많은 데이터가 필요하다는 단점이 있지만, 전통적인 Convolutional Neural Network 대비 더 뛰어난 성능을 보여준다. 이와 같은 이유로 이미지 정보와 같은 시각 정보에 기반하여 다양한 인식과 분류를 수행하는 인공지능 모델의 구현은 Convolutional Neural Network에서 Transformer 기반 모델로 전환되고 있다. 최근에는 데이터를 효율적으로 활용하는 Vision Transformer 훈련을 위해 증류(Distillation) 방법을 사용하는 Vision Transformer의 변형이 제안되었다. CNN의 피라미드 구조나 전체 데이터에 셀프 어텐션(All-to-All Self-attention)을 적용하는 대신에 추상적 표현(Abstract Representation)을 학습하는 기법도 포함되어 있다^[9]. Video Swin Transformer는 Vision Transformer를 기반으로 하여 개발되었고, 사전 훈련된(Pre-trained) 모델을 계속 활용한다^[10]. Vision Transformer는 저해상도의 특징 맵을 생성하며, 전역 셀프 어텐션(Self-attention Globally) 연산 때문에 계산량이 크게 증가한다는 단점도 존재한다^[11]. 하지만, 비디오 정보 기반 인식 및 분류 분야에서 Video Swin Transformer는 매우 뛰어난 성능을 보이고 있다.

IV. 성능분석

1. 데이터셋 및 데이터 전처리 방법

가. 데이터셋

얼굴 영역 추출 시 여유값의 설정에 따른 개성 인식 모델의 정확도를 분석하기 위해 사용한 데이터셋은 First Impressions V2 (CVPR'17)이다^[12]. 데이터셋은 YouTube에서 수집된 평균 15초 지속 시간을 가지는 10,000개의 비디오로 구성되어 있다. 각 비디오는 한 사람이 카메라를 직접 응시하며 영어로 말하는 형태이다. 각 데이터는 OCEAN 값이 0~1 사이로 라벨링 되어 있다. 10,000개의 데이터는 학습(Training), 검증(Validation), 테스트(Test)용으로 3:1:1의 비율로 나누어졌다. 실험에는 학습과 검증 데이터만 사용하였다. 전체 데이터셋에는 15초 미만의 비디오도 포함되어 있다. 본 연구에서는 총 프레임 수가 128 프레임 이상인 비디오만을 선별하여 사용했다. 최종적으로 학습 데이터로는 5,979개, 검증 데이터로는 1,993개의 비디오가 사용되었다.

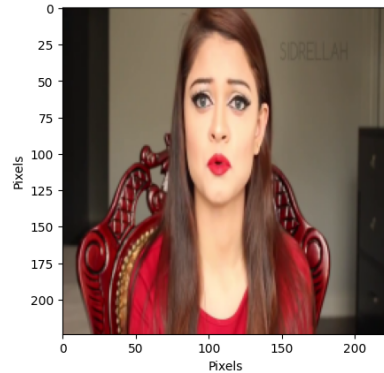


그림 2. 원본 프레임
Fig. 2. Original frame

나. 데이터 전처리

각 비디오에서 그림 2와 같이 프레임을 추출했다. 이후, MTCNN을 사용하여 얼굴 이미지를 추출하고 모델에 입력에 적합하도록 224×224 크기로 조절했다. MTCNN의 여유값 매개변수를 조절하여 학습에 사용할 수 있도록 전처리했다. 그림 3은 얼굴 추출 시 사용한 여유값의 변화에 따른 이미지 차이를 보여준다. 여유값이 증가함에 따라 얼굴 주변의 배경 부분이 더 포함되는 형태로 추출되는 것을 확인할 수 있다.

2. 실험 환경

표 1은 성능분석에서 사용된 실험 환경의 상세정보를 보여준다. 실험에 사용된 컴퓨팅 노드는 AMD Ryzen 9 7950X3D가 장착되어 있다. GPU는 NVIDIA GeForce

RTX 4090를 사용했다. 또한 RAM 용량은 128GB이고, 운영체제는 Ubuntu 22.04.2를 사용했다. 사용된 Python의 버전은 3.10.11이며, PyTorch와 Torchvision의 버전은 각각 2.0.1과 0.15.2이다. 마지막으로 CUDA 버전은 11.8.0을 사용했다.

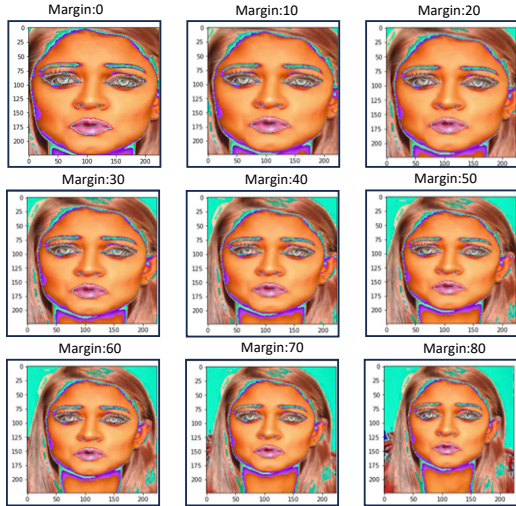


그림 3. 각 여유값에 따른 추출된 이미지 예
 Fig. 3. Images corresponding to each Margin value

표 1. 실험 환경
 Table 1. Experimental environment

항목	세부정보
CPU	AMD Ryzen 9 7950X3D 16-Core Processor
GPU	NVIDIA GeForce RTX 4090
RAM	128 GB
OS	Ubuntu 22.04.2
Python	3.10.11
Pytorch	2.0.1
Torchversion	0.15.2
CUDA	11.8.0

3. 평가지표 및 학습 환경 설정

개성 인식 모델의 정확도 평가 지표로 사용한 것은 1-MAE(Mean Absolute Error)이다. 평가지표인 1-MAE는 수식 1과 같은 형태로 계산한다. 여기서 y_i 는 i 번째 비디오의 라벨링 되어 있는 OCEAN 값을 의미하고, $f(x^i)$ 는 해당 데이터의 OCEAN 예측값이다.

$$1 - MAE = 1 - \frac{1}{n} \sum_{i=1}^n |y_i - f(x^i)| \quad (1)$$

Optimizer로는 AdamW를 사용하였으며, 총 120 Epochs 동안 학습을 진행하였다. 학습률(Learning rate)은 0.00003로 설정하였다. 배치 크기(Batch Size)를 8로 설정하였다. R2plus1D Backbone은 Kinetics-400 데이터셋으로 사전 학습(Pre-trained)된 모델을 사용하였다.

표 2. 하이퍼파라미터와 값 설정 정보
 Table 2. Hyper-Parameters and Their Values

Hyper-Parameters	Values
Epoch	120
Num_Workers	5
Batch Size	8
Learn Rate	0.00003
Loss Function	MAE Loss
Optimizer	AdamW optimizer

4. 실험 결과

그림 4는 여유값에 따른 개성 인식 모델의 1-MAE 성능 결과를 보여준다. 그림 4에서 확인할 수 있듯이 얼굴 영역 추출 시 사용한 여유값이 0일 때 1-MAE 성능이 가장 좋지 않았다. 여유값이 0일 때의 1-MAE는 0.9096이다. 또한 여유값이 0보다 큰 경우에는 모두 0보다는 좋은 성능을 보여주었다. 여유값이 60일 때 1-MAE 성능이 가장 높았으며, 이때 1-MAE 값이 0.9118이다.

따라서 얼굴 영역의 정보만을 이용하여 개성 인식 모델과 OCEAN 모델을 기반으로 개성은 인식할 때, 순수한 얼굴 영역을 이용하는 것은 1-MAE 측면의 성능 향상에 도움이 되지 않는다는 것을 알 수 있다. 기존 연구에 따르면 얼굴 인식 기술에서 밝기 정보도 중요 정보로 사용한다^{[13][14]}. 따라서 얼굴 주변부의 헤어 스타일, 밝기 정보 등이 OCEAN 모델 기반 개성 인식 모델의 성능 향상에 도움이 된다는 것을 알 수 있다.

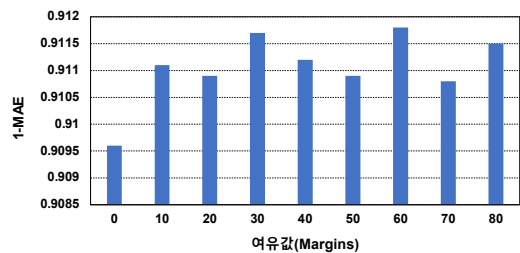


그림 4. 여유값에 따른 개성 인식 모델의 1-MAE 성능
 Fig. 4. 1-MAE Performance of the Personality Recognition Model according to Margins

요약하면, 1-MAE 측면의 성능을 최대화하기 위해서는 얼굴 영역 추출 시 설정하는 여유값이 개성 인식 모델의 정확도 성능에 큰 영향을 줄 수 있음을 확인할 수 있었다. 따라서 1-MAE 측면의 성능을 최적화하기 위해서는 얼굴 영역 추출 시 적절한 여유값으로 설정할 필요가 있다.

V. 결 론

OCEAN 모델 기반 개성 인식은 매우 중요한 분야이다. OCEAN 모델 기반의 개성 인식이 중요한 이유는 사용자의 개성을 파악하여 이를 기반으로 사용자에게 좀 더 개인화된 서비스 제공이 가능하기 때문이다.

본 논문에서는 얼굴 영역을 추출할 때 여유값이 개성 인식 모델의 정확도 성능에 미치는 영향을 분석했다. 이를 위해 2D Patch Partition + R2plus1D + 3D Patch Partition + Video Swin Transformer에 기반한 개성 인식 모델을 사용했다. 실험 결과, 여유값이 0일 때 가장 성능이 좋지 않았으며, 1-MAE는 0.9096이다. 반면, 여유값이 60일 때 가장 좋은 성능을 얻을 수 있었으며, 1-MAE 값은 0.9118이다. 개성 인식 성능을 높이기 위해서는 순수 얼굴 정보만 사용하는 것 보다는 얼굴 주변부의 정보를 같이 사용하는 것이 효과적임을 확인할 수 있었다. 또한 개성 인식 모델의 성능을 최적화하기 위해서는 얼굴 영역 추출 시 적절한 여유값으로 설정하여야 한다.

References

- [1] Yongjun Kim, "Feature Selection for Creative People Based on Big 5 Personality traits and Machine Learning Algorithms", *The Journal of the Institute of Internet, Broadcasting and Communication*, Vol. 19, No 1, pp. 97-102, 2019.
DOI: <https://doi.org/10.7236/JIIBC.2019.19.1.97>
- [2] Xiaodong Duan, Qingguo Zhan, Shaohui Zhan, Yingying Yu, Luning Chang, Yuangang Wang, "Multimodal Apparent Personality Traits Analysis of Short Video using Swin Transformer and Bi-directional Long Short-Term Memory Network", *2022 4th International Conference on Frontiers Technology of Information and Computer*, pp. 1003-1008, 2022.
DOI: <https://doi.org/10.1109/ICFTIC57696.2022.10075178>
- [3] Chanyoung Jo, Hyunjun Jung, "Multimodal Emotion Recognition System using Face Images and Multidimensional Emotion-based Text", *Journal of KIIT*, Vol. 21, No. 5, pp. 39-47, 2023.
DOI: <http://dx.doi.org/10.14801/jkiit.2023.21.5.39>
- [4] Kaipeng Zhang, Zhanpeng Zhang, Zhifeng Li, Yu Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks", *IEEE signal processing letters*, Vol 23, No. 10, pp. 1499-1503, 2016.
DOI: <https://doi.org/10.48550/arXiv.1604.02878>
- [5] Jia Xiang, Gengming Zhu, "Joint face detection and facial expression recognition with MTCNN", *2017 4th international conference on information science and control engineering*, pp. 424-427, 2017.
DOI: <https://doi.org/10.1109/ICISCE.2017.95>
- [6] Face Recognition Library.
https://github.com/ageitgey/face_recognition
- [7] Nataliya Boyko, Oleg Basystiuk, Nataliya Shakhovska, "Performance evaluation and comparison of software for face recognition, based on dlib and opencv library", *2018 IEEE Second International Conference on Data Stream Mining & Processing*, pp. 478-482, 2018.
DOI: <https://doi.org/10.1109/DSMP.2018.8478556>
- [8] Du Tran, Heng Wang, Lorenzo Torresani, Jamie Ray, Yann LeCun, Manohar Paluri, "A Closer Look at Spatiotemporal Convolutions for Action Recognition", *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6450-6459, 2018.
DOI: <https://doi.org/doi:10.1109/cvpr.2018.00675>
- [9] Chun-Fu (Richard) Chen, Quanfu Fan, Rameswar Panda, "Crossvit: Cross-attention multi-scale vision transformer for image classification", *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 357-366, 2021.
DOI: <https://doi.org/10.48550/arXiv.2103.14899>
- [10] Ze Liu, Jia Ning, Yue Cao, Yixuan Wei, Zheng Zhang, Stephen Lin, Han Hu, "Video swin transformer", *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 3202-3211, 2022.
DOI: <https://doi.org/10.48550/arXiv.2106.13230>
- [11] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, Baining Guo, "Swin transformer: Hierarchical vision transformer using shifted windows", *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 10012-10022, 2021.
DOI: <https://doi.org/10.48550/arXiv.2103.14030>
- [12] First Impressions V2 (CVPR'17) Dataset.
<https://chalearnlap.cvc.uab.cat/dataset/24/description/>
- [13] L. Torres, J.Y. Reutter, L. Lorente, "The importance of the color information in face recognition", *Proceedings 1999 International Conference on Image Processing*, Vol. 3, pp. 627-631, 1999.
DOI: <https://doi.org/10.1109/ICIP.1999.817191>
- [14] Youngjoon Cho1, Jongwon Kim, "A Study on The Classification of Target-objects with The Deep-learning Model in The Vision-images", *Journal of the Korea*

Academia-Industrial cooperation Society, Vol. 22, No.
2, pp. 20-25, 2021.
DOI: <https://doi.org/10.5762/KAIS.2021.22.2.20>

저 자 소 개

구 욱(준회원)



- 2017년 9월 ~ 2021년 6월 : Linyi University College of Information Science and Engineering
- 2022년 9월 ~ : 충북대학교 전기·전자·정보·컴퓨터학부
- 관심분야 : 인공지능, 기계학습

한 규 원(준회원)



- 2019년 3월 ~ : 충북대학교 컴퓨터공학과
- 관심분야 : 컴퓨터 비전

김 봉 재(정회원)



- 2021년 3월 ~ : 충북대학교 컴퓨터공학과 부교수
- 2016년 3월 ~ 2021년 2월 : 선문대학교 컴퓨터공학부 조교수
- 2015년 1월 ~ 2016년 2월 : 한국전자기술연구원 임베디드·SW 연구센터 선임연구원
- 2014년 9월 ~ 2014년 12월 : 삼성전자 SW 개발팀 (메모리) 책임연구원
- 2014년 2월 : 서울대학교 전기·컴퓨터공학부 박사
- 2008년 2월 : 광운대학교 컴퓨터공학부 학사
- 관심분야 : 시스템소프트웨어, 인공지능, 지능형시스템

※ 이 논문은 2023년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (2022-0-00043, 개성 형성이 가능한 에이전트 플랫폼 기술 개발)