

얼굴의 감정 분류를 위한 DFLM(Deep learning facial landmark model)

DFLM(Deep learning facial landmark model) for facial emotion classification

정 이 나¹
Yi-na Jeong

심 용 보¹
Yong-bo Sim

손 수 략^{1*}
Su-rak Son

요 약

전자기기는 일상에 깊게 자리 잡아, 대부분 시간을 차지할 정도로 많이 쓰인다. 이런 전자기기를 이용해 인간의 심리상태를 판단하고 상황에 맞는 피드백을 주는 것은 삶에 많은 도움을 줄 수 있다. 기존 얼굴 감정 분류 기법으로 CNN을 이용한 방식이 연구되어 있는데 해당 방식은 수만 개의 픽셀을 입력 데이터로 받기 때문에 많은 학습 데이터가 필요하다는 단점이 있다. 본 논문에서는 이를 해결하기 위해 얼굴의 특징점을 입력 데이터로 받아 얼굴의 감정을 분류하는 딥러닝 모델 DFLM(Deep learning facial landmark model)을 제안한다. 제안한 DFLM은 얼굴에 68개의 특징점을 입력으로 하여 완전 연결 신경망을 구성해 7가지 표정을 분류할 수 있다. 시뮬레이션 결과 CNN 모델에 비해 빠른 학습이 가능하고, 동작 시간이 짧아 실시간으로 예측할 필요가 있는 분야에 도움이 될 것이다.

☞ 주제어 : 얼굴 특징점, 딥러닝, 얼굴 표정 분류

ABSTRACT

Electronic devices are deeply entrenched in our daily lives and are used so much that they take up most of our time. Using these electronic devices to judge a person's psychological state and give feedback appropriate to the situation can help a lot in life. A method using CNN has been studied as an existing facial emotion classification technique, but this method has the disadvantage that it requires a lot of training data because it receives tens of thousands of pixels as input data. In this paper, to solve this problem, we propose a deep learning model DFLM (Deep learning facial landmark model) that receives facial landmarks as input data and classifies facial emotions. The proposed DFLM can classify 7 facial expressions by constructing a fully connected neural network with 68 landmarks as inputs. As a result of the simulation, it is possible to learn faster than a CNN model, and its operation time is short, so it will be helpful in areas that need to make predictions in real time.

☞ keyword : facial landmark, deeplearning, facial emotion classification

1. 서 론

현재 전자기기는 일상 깊이 자리 잡아, 삶의 모든 영역에 인간과 전자기기가 공존하고 있다. 인간의 건강을 파악하거나 [1], 아이의 이상징후를 감지하거나 [2], 현재 기분을 파악해 맞춤형 서비스를 제공하는 등의 서비스는 삶의 질 향상에 많은 도움이 된다. 인간의 건강과 관련한 연구 중 다양한 환경에서 사람의 안전을 위해 통증을 감지하는 연구가 있다. 이 중에서도 표정은 사람의 상태 및

감정에 대한 가장 직접적이고 시각적인 표현이다 [1]. 기존의 체온측정 시스템은 딥러닝 기반 얼굴검출 알고리즘이 사용되어 얼굴의 체온측정에는 활용될 수 있지만, 의사소통이 어려운 영유아의 경우 이상징후를 인지하기 힘들다. 기존의 체온측정 시스템을 개선하여 영상에서 영유아를 검출한 후 얼굴을 추출해 표정 인식의 핵심인 눈, 코, 입을 선택적 샤프닝 필터를 적용하여 표정 인식을 진행해 인식률을 개선했다 [2]. 영화, 드라마, 게임 등 다양한 콘텐츠에 노출되는 사용자의 솔직한 평가나 반응을 알기 위한 설문 조사는 보통 콘텐츠를 감상한 후에 평가 내리는 방식이라 콘텐츠 도중의 감정, 평가에 편향, 주관 등이 개입될 수 있다. 여기서 실시간으로 감상자의 표정으로 평가해 객관적인 결과를 도출할 수 있다 [3]. 다양한 분야에서 연구되고, 활용되는 표정 분류 모델에는 CNN

¹ Software Department, Catholic Kwandong University, Gangneung-si, Gangwon-do, 25602, Korea.

* Corresponding author (sonsur@naver.com)

[Received 18 Oct 2022, Reviewed 09 Jan 2023, Accepted 17 Jan 2023]

(Convolution Neural Network) 딥러닝 모델을 이용한 분류 모델이 있다 [4]. 하지만 CNN 모델의 경우 일반적으로 수만 개의 픽셀을 입력으로 받아 이를 처리하기 위해 많은 은닉층과 노드를 요구하기 때문에 파라미터가 많아져 많은 학습 데이터가 필요하다. 본 논문에서는 얼굴 감정 분류를 위한 모델에서 이러한 문제를 해결하기 위해 입력 데이터를 픽셀 기반의 영상이 아닌 얼굴의 특징점으로 대체해 68개의 좌표를 입력으로 하는 모델을 구성하여 적은 데이터와 빠른 학습이 가능한 얼굴 감정 분류 모델인 DFLM(Deeplearning facial landmark model)을 제안한다.

2. 관련 연구

얼굴의 특징점을 이용해 얼굴을 인식하는 기법이 있다 [5]. Dlib에서 68개의 특징점을 이용해 높은 정확도로 얼굴을 인식하는 기능을 제공한다. 본 논문에서는 Dlib의 특징점 추출 기능을 이용해 표정 분류를 위한 데이터셋을 생성한다.

얼굴의 특징점은 사람의 얼굴에 대한 다양한 정보를 알아낼 수 있다. 특징점을 가지고 인간의 표정을 분류하는 기존 연구를 확인할 수 있는데, 기존 연구를 보면 얼굴 랜드마크 간의 거리를 계산하여 추출된 특징을 사용해 관계를 분석하고 5가지 표정으로 분류한다 [6]. 다른 논문에서는 랜드마크 기반 앙상블 네트워크를 활용해 표정을 분류하는 알고리즘을 제안한다. 얼굴 이미지에서 랜드마크 정보를 이용하여 앙상블 학습을 통해서 4가지의 감정 상태로 표정을 분류한다 [7]. 얼굴에 일정한 위치에 좌표 값들을 뽑는 특징점을 추출하는 방법들이 연구되어 왔다. 특징점을 추출한다는 것은 영상 속 사람의 얼굴을 인식할 수 있으며 이는 영상 촬영 시 얼굴 자동 초점 기능, 열화상 카메라에서 얼굴 온도를 감지, 사람의 표정으로 감정을 추론하는 등의 여러 기술의 토대가 된다 [8]. 최근 딥러닝 기술의 발전과 함께 얼굴의 특징점 검출 방법의 성능이 크게 향상되었다. 히트맵 회귀 방식은 효율적이고 강력한 방법으로 널리 사용되고 있으나 여러 단점이 존재해 적분 회귀 방법을 결합하여 특징점을 추출하는 방식을 제안해 검출 성능을 크게 향상했다 [9]. 모바일에서의 실시간 동작을 위해 딥러닝 기반의 얼굴 검출 및 얼굴 랜드마크 알고리즘을 기존 데스크탑 상에서 설계된 딥 네트워크 구조를 크게 변경하지 않고 모바일 상에서 실시간으로 동작시키기 위한 방법을 제안했다 [10].

CNN 기반으로 표정을 통해 감정분류를 하는 다양한 선행 연구들이 있다. 감성과 힐링이라는 주제를 바탕으로

딥러닝을 이용한 사용자의 표정을 인식하고 음악을 재생해줘 해당 감정에 맞게 감정을 더 증폭시켜 줄 수 있도록 해준다. 해당 연구에서는 mmist 모델을 기반으로 2개 층으로 구현해 행복, 슬픔, 놀라움 3가지 표정으로 분류한다 [11]. CNN을 이용한 또 다른 논문에서는 기존 표정 데이터베이스의 단점을 보완하여 질 좋은 다양한 데이터베이스를 이용해 ‘무표정’, ‘행복’, ‘슬픔’, ‘화남’, ‘놀람’, ‘역겨움’에 대해서 분류하는 모델을 제안한다 [4]. 다양한 CNN 아키텍처의 예측 성능 비교를 위해 AlexNet, GoogLeNet, Vggnet, Resnet의 성능을 비교 분석한 논문을 보면 각 모델을 100 epoch로 학습하여 epoch마다 학습 정확도를 확인한다. 이때 각 정확도는 CNN 아키텍처 망의 깊이가 깊어질수록 높아지는 경향을 보인다 [12]. 저해상도 이미지의 작은 데이터 세트를 분류하기 위해서는 새로운 소규모 심층 합성곱 신경망 모델을 개발하는 연구가 있다. 기존 심층 합성곱 신경망에 비해 일부의 메모리만 사용하지만, FER2013 및 FERPlus 데이터 세트에서 매우 유사한 결과를 보여준다 [13]. 이를 통해 인공지능 모델에서 입력 데이터의 해상도, 자료의 개수, 분류하고자 하는 라벨에 따라 최적화된 모델의 구조가 다를 수 있다는 것을 알 수 있다.

3. 본 론

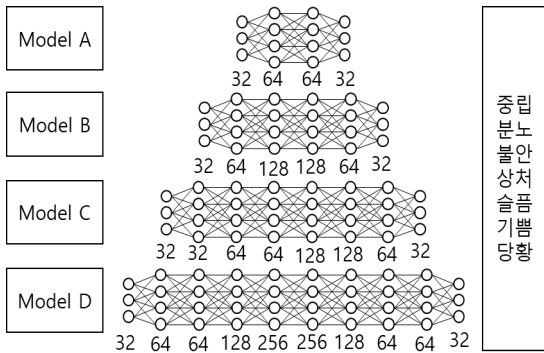
3.1 모델 구성

DFLM은 얼굴의 특징점을 68개로 짝은 좌표 데이터로 입력받아 인간의 표정을 분류하는 딥러닝 모델이다. 입력 데이터로 68개의 특징점을 받는다. 완전 연결 신경망으로 구성된 은닉층을 통과해 7가지 표정으로 분류한다.

DFLM에서 최적의 인식률을 낼 수 있는 은닉층의 개수를 알기 위해 Model A, B, C, D의 레이어를 그림 1과 같이 DFLM의 은닉층을 서로 다른 깊이로 구성하여 4개의 모델을 설계한다. Model A는 4개의 은닉층을 가지고 각각 32, 64, 64, 32개의 노드를 가진다. 다음으로는 6개의 은닉층을 가지고 각각 32, 64, 128, 128, 64, 32개의 노드를 가진 Model B가 있다. Model C는 8개의 은닉층을 가지고 각각 32, 32, 64, 64, 128, 128, 64, 32개의 노드를 가진다. 마지막으로 Model D는 10개의 은닉층을 가지고 각각 32, 64, 64, 128, 256, 256, 128, 64, 32, 32개의 노드로 구성된다. 각 레이어의 활성화 함수는 ReLU를 이용한다. 마지막 레이어는 출력 라벨 개수와 같은 7개의 노드를 가진 softmax를 활성화 함수로 하는 레이어를 가진다. softmax

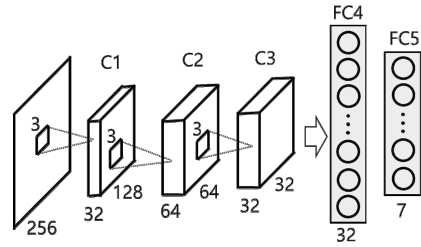
를 사용하는 이유는 분류 문제에서 사용되는 CEE(Cross Entropy Error, 교차 엔트로피 에러)를 이용했을 때 softmax와 CEE를 묶어 미분하면 매우 빠르게 미분값을 알아낼 수 있기 때문이다. 마지막 softmax 레이어는 트레이닝 과정에서 CEE의 미분 처리 속도를 높이기 위해 존재하기 때문에 실제 학습이 완료된 모델을 사용할 때는 사용하지 않는다.

DFLM은 CNN 기반 모델과 달리 얼굴 특징점에 68개의 좌푯값을 입력으로 하는 완전 연결 신경망으로 구성된 딥러닝 모델이다. 실험 환경에서 최적의 모델을 찾기 위해 DFLM의 Model A, B, C, D를 시뮬레이션하여 가장 높은 성능을 보이는 모델로 최종 결정한다.



(그림 1) DFLM(Deeplearning facial landmark model)의 모델 A, B, C, D의 구조
(Figure 1) DFLM(Deeplearning facial landmark model) model A, B, C, D structure

그림 2와 같이 비교를 위한 CNN 기반 모델은 Convolution을 수행하는 레이어 3개와 하나의 Fully connected layer로 구성된다. 각 레이어의 활성화 함수는 ReLU를 이용하고, 필터는 3x3 사이즈로 stride를 1씩 옮겨 가며 Convolution 연산을 수행한다. FC4는 완전 연결 신경망으로 C3에서의 출력인 3차원 데이터를 처리할 수 없어 1차원으로 재구성(reshape) 하고, ReLU 활성화 함수를 이용해 출력한다. 마지막 FC5 레이어에서는 7가지의 표정으로 분류된 정보를 가지기 위해 7개의 노드를 가지고 softmax를 활성화 함수로 한다. FC5 레이어는 학습할 때에 CEE의 처리 속도를 위해 존재하기 때문에 트레이닝 시에만 사용되고 실제 예측할 때에는 사용하지 않는다.



(그림 2) 비교를 위한 CNN 모델의 구조
(Figure 2) CNN model structure

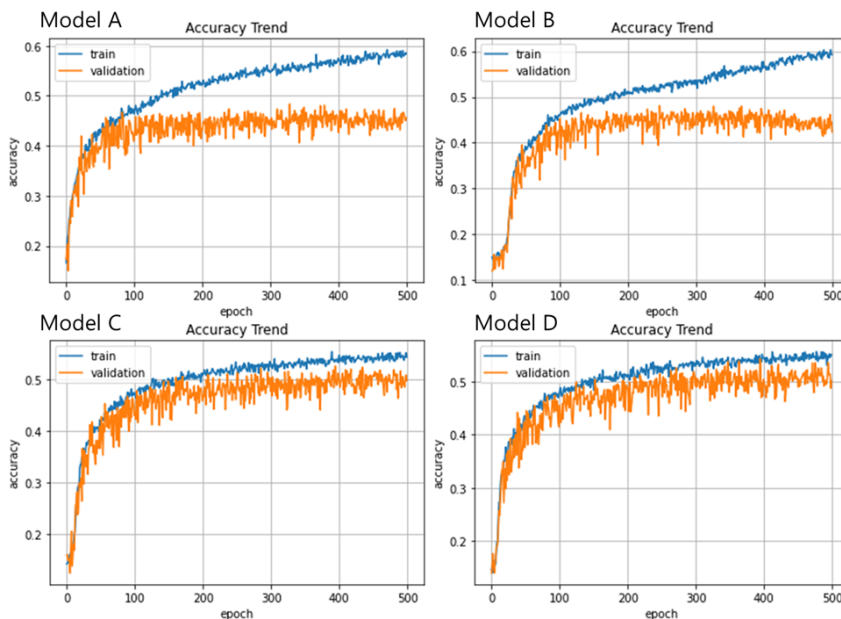
4. 시뮬레이션

4.1 실험설계

표정 분류를 위한 모델을 구성하기 위해 트레이닝을 위한 데이터 세트를 수집해야 한다. 데이터는 표정이 드러나는 사람의 얼굴 영상 데이터를 이용한다. 수집한 얼굴 영상 데이터는 AI-Hub의 한국인 감정인식을 위한 복합 영상에서 제공한 중립, 분노, 불안, 상처, 슬픔, 기쁨, 당황으로 7가지의 표정으로 분류되어 있다 [14].

본 논문에서 제안하는 DFLM과 기존 CNN 기반 모델의 비교를 위해서는 수집한 데이터와 라벨을 그대로 이용할 수 없다. 기존 CNN 모델은 영상의 가로, 세로 크기만큼의 2차원 색상 데이터를 입력하는 방식을 이용하지만, DFLM은 얼굴의 랜드마크를 이용하기 때문에 68개의 좌푯값이 들어가게 된다.

학습을 위해 수집한 7종류의 감정에 대한 표정을 담은 얼굴 영상에서 얼굴 영역을 탐색한다. 얼굴 영역을 기준으로 특징점을 추출해 68개의 좌푯값을 가지는 데이터로 만들어 준다. 이렇게 만들어진 68개의 좌푯값은 DFLM의 훈련을 위한 데이터에 활용된다. CNN 모델은 원초 데이터에서 얼굴을 탐색하여 얼굴 영역만큼 잘라낸 새로운 얼굴 이미지를 가로, 세로 256의 크기로 변환(resize) 해준다. 기존 RGB 컬러로 구성된 색상 이미지를 Gray로 바꿔 주어 256 크기로 회색조(GrayScale)된 이미지를 생성한다. 생성한 이미지 중 표정별로 1,000장의 이미지를 사용해 총 7,000장의 얼굴 이미지를 사용한다. 학습 시 데이터의 70%는 학습 데이터(training data)로 사용하고 나머지 30%는 검증 데이터(validation data)로 사용해 학습이 제대로 이루어지고 있는지 검증한다. 테스트 데이터(test data)는 학습이 종료된 이후, 학습에 사용되지 않은 1000장의 이미지를 학습된 모델에 입력 데이터로 넣어 결과를 예측



(그림 3) DFLM의 Model A, B, C, D 인식률
(Figure 3) Recognition rate of the DFLM model A, B, C and D

하게 하여 모델의 성능을 평가할 때 이용한다.

DFLM의 성능은 속도와 정확도, F1 Score로 평가한다. 또한, 비교군을 위해 표정 분류를 위해 사용되는 딥러닝 모델인 CNN 기반의 학습한 모델과 DFLM을 비교하여 성능을 검증한다. 실험 환경은 윈도우 11 (64bit) 운영체제의 PC에서 동작한다. CPU는 i5-12400F, 메모리는 8GB의 메모리 2개를 연결해 듀얼 채널로 구성된 16GB 용량으로 구성되어 있다. GPU는 RTX3070Ti를 사용한다. python3 언어와 keras 라이브러리를 이용해 학습 모델을 구축한다.

4.2 정확도

정확도는 간단하게 성능을 측정하는 방법으로, 올바르게 예측된 데이터의 수를 전체 데이터의 수로 나눈 값이다. 예측한 답이 맞는 경우의 수에 예측과 답이 맞는 경우, 예측과 답이 틀린 경우를 합한 경우의 수를 나눈 것을 의미한다.

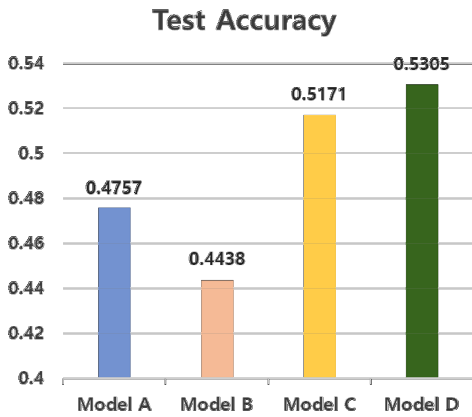
그림 3을 DFLM의 Model A, B, C, D가 학습될 때 epoch마다 학습 결과와 검증 결과의 정확도(accuracy)를 확인할 수 있다. 그림 3을 보면 학습이 진행될수록 모델

의 학습 정확도와 검증 정확도의 차이가 벌어지는 것을 볼 수 있다. 학습 정확도는 지속해서 증가하지만, 검증 정확도가 증가하는 폭이 작아 이 차이가 벌어진다는 것은 과적합 (overfitting)이 발생했다고 할 수 있다. 과적합은 학습 데이터를 과하게 잘 학습하는 것을 뜻한다. 학습 데이터는 실제 데이터의 부분집합이므로 학습 데이터에 대해서는 오차가 감소하지만 실제 데이터에 대해서는 오히려 오차가 증가한다. 그림 3에서 각 모델 A, B, C, D를 보면 학습 정확도와 검증 정확도의 차이가 Model D로 갈수록 줄어드는 경향을 보인다. 실험을 통해 레이어를 많이 가져갈수록 과적합을 줄이는데 유리하는 것을 알 수 있다.

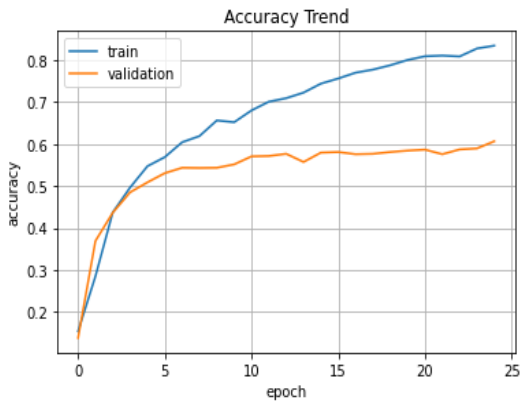
그림 4는 500회 반복 학습 후 각 모델에 테스트 데이터 (test data)를 이용해 정확도를 비교한 그래프이다. 테스트 데이터를 이용해 정확도를 확인해본 결과 model A는 0.4757의 정확도를 보이고, model B는 0.4438의 정확도, model C는 0.5171, model D는 0.5305의 정확도를 보여 model D의 인식률이 가장 높은 것을 알 수 있다. 해당 지표에 따라 model D로 갈수록 정확도가 높아지는 경향을 보인다. 그림 5와 같이 CNN으로 학습한 모델에서 정확도

는 0.5307로 DFLM과 유사한 수치를 보였으나 학습 데이터와 검증 데이터의 차이가 계속 벌어지며 더 이상 학습이 되지 않는 것을 알 수 있다. CNN 모델의 특성상 학습 데이터의 표본 수가 적을 때 과적합 문제가 발생하는데, 학습 정확도와 검증 정확도의 차이가 벌어져 CNN 모델에서 학습 중에 과적합 문제가 발생했음을 알 수 있다.

본 논문에서 진행한 실험 환경에서 epoch 3회부터 학습 정확도와 검증 정확도가 벌어지기 시작하여 약 0.45의 정확도를 달성한 시점부터 학습 정확도와 검증 정확도의 차이가 벌어지기 시작하고, 25회의 epoch가 진행되며 학습 정확도는 0.9까지 도달했으나 검증 정확도는 0.5 정도에서 멈춰 더 이상 상승하지 않는 것을 알 수 있다.



(그림 4) 테스트 데이터를 통해 나온 각 모델의 정확도
(Figure 4) Accuracy of each model from the test data



(그림 5) 테스트 데이터로 확인한 정확도
(Figure 5) Accuracy confirmed by test data

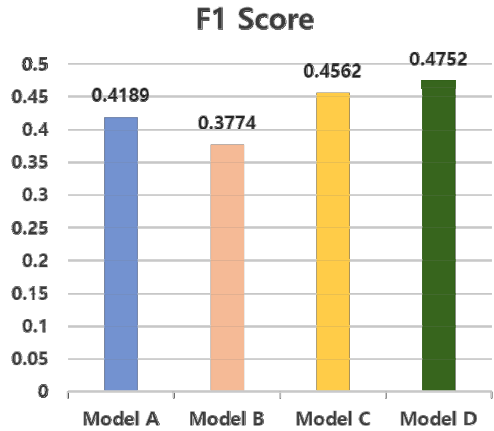
4.3 F1 Score

$$F_1 = 2 * \frac{1}{\frac{1}{r} + \frac{1}{p}} = \frac{p * r}{p + r}$$

(수식 1) F1 Score

p = precision, r = recall

F1 Score는 정밀도와 재현율의 조화 평균이다. 본 논문에서는 모델의 성능평가에 정확도와 F1 Score 두 가지를 이용해 모델의 인식 성능을 확인한다. 정확도로 모델의 성능을 평가했을 때 발생하는 여러 문제점이 있기 때문이다 [15].



(그림 6) Model A, B, C, D의 F1 Score
(Figure 6) F1 Score of Model A, B, C and D

정확도는 데이터에 따라 매우 잘못된 통계를 나타낼 수 있다. 예를 들어 암에 걸렸는지 판단하는 의료 인공지능의 경우를 생각해 보면, 이런 경우에 항상 음성이라고 하면 높은 정확도를 낼 수 있다. 암에 걸린 사람 보다 걸리지 않은 사람이 압도적으로 많기 때문에 무조건 음성 판정하면 높은 정확도를 보일 수 있는 것이다. 하지만, 높은 accuracy를 가졌지만 이런 모델은 아무런 의미가 없다. 이런 상황에 사용할 수 있는 통계치가 재현율(recall)이다. 재현율은 실제로 true인 데이터를 true라고 인식한 데이터의 수를 의미한다. 양성인 사람이 10명일 때 모두 음성이라고 대답하는 모델은 양성인 사람을 0명으로 예측할 것이기 때문에, 0/10이 되어 0의 재현율을 가지게 되고, 이 모델에 문제가 있다고 판단한다.

정밀도(precision)는 재현율과 상반되는 관계에 놓여있다. 위의 예에서 항상 양성판정을 내리는 모델이 있다고 했을 때, 재현율은 1이 된다. 재현율의 수치가 1로 매우 높게 나타났지만, 이 모델은 항상 음성으로 판정하는 모델과 같이 별 의미 없는 모델이다. 이런 상황에 도움을 주는 통계치가 정밀도이다. 정밀도는 모델이 true로 예측한 데이터 중 실제로 true인 데이터를 말한다. 즉 양성인 사람이 10명일 때 100명을 양성으로 답하면 10/100으로 낮은 정밀도를 가지게 되어 해당 모델에 문제가 있다고 판단 할 수 있다. 본 논문에서는 이런 정밀도와 재현율을 조합한 통계치인 F1 score를 이용한다. F1 Score는 정밀도와 재현율을 조합한 조화 평균이다. 조화 평균을 쓰는 이유는 정밀도와 재현율이 0에 가까울수록 F1 score도 동일하게 낮은 값을 갖도록 하기 위함이다.

그림 6에서 각 모델의 f1 score를 볼 수 있다. Model A는 0.4189, Model B는 0.3774, Model C는 0.4562, Model D는 0.4752에 가까운 수치를 보인다. 모델 A에서 D로 갈수록 F1 Score가 높은 경향을 보인다. 표 1에서 모델별 인식률과 F1 Score를 보면 Model D의 성능이 가장 좋다는 것을 알 수 있다.

(표 1) 각 학습 모델의 성능 지표

(Table 1) Performance metrics for each model

	Model A	Model B	Model C	Model D
정확도	0.4757	0.4438	0.5171	0.5305
F1 Score	0.4189	0.3774	0.4562	0.4752

4.4 학습 시간과 동작 시간

인공지능에서 모델의 학습에 필요한 시간과 학습된 모델이 예측값을 내는 데까지 걸리는 시간은 매우 중요한 지표라고 할 수 있다. 학습에 걸리는 시간을 비교해 봤을 때 DFLM의 경우 한 epoch에 걸리는 학습 시간이 1초 정도 소요되지만, CNN 모델의 경우 약 100초의 시간이 소요된다. 학습 뿐만이 아닌 모델을 평가하기 위한 테스트에서 CNN 모델의 경우 8s 128ms/step이라는 시간이 필요했고, DFLM은 0s 4ms/step 만에 테스트 데이터의 평가를 완료했다. 테스트 시간을 보면 제안하는 모델이 실제로 사용될 때 매우 빠른 속도로 동작할 수 있음을 알 수 있다.

5. 결 론

표정 분석을 위한 기존의 연구를 보면 CNN 모델을 기반으로 얼굴의 영상을 이용해 표정을 분류했다. 입력 데이터를 줄일 수 있는 얼굴 특징점을 이용한 표정 분류 모델이 기존 연구에서 사용하던 CNN 기법보다 적은 학습 데이터로도 비슷한 인식률을 유지하면서 빠른 학습을 할 수 있어 좋은 성능을 보인다. DFLM에서 각 모델을 학습할 때 적은 레이어를 가진 모델에서는 빠르게 과적합이 발생하는 것을 알 수 있었다. 시뮬레이션 결과 DFLM의 정확도는 0.51이 나오고, F1 score는 0.4752가 나왔고, CNN 모델의 정확도는 0.5307이다. DFLM 모델은 0s 4ms/step 만에 테스트 데이터의 평가를 완료했지만, CNN 모델에서는 8s 128ms/step이라는 시간이 필요하다. 입력 데이터를 줄여 모델 구성을 경량화하면 빠른 처리가 가능하다는 것을 알 수 있다. 학습에 사용된 표본의 수가 부족해 전체적인 정확도가 낮았고, 학습 시 정확도의 편차가 크게 나타나 이를 개선할 필요가 있다. 추후 연구를 통해 학습 데이터를 늘리고 완전 연결 신경망이 아닌 다양한 기법의 딥러닝 모델을 활용하는 등의 방법을 이용해 정확도를 높일 수 있을 것이라 기대한다.

참고문헌(Reference)

- [1] MinsongKi, Yeonggw Choi, "Extreme Pain Detection based on Human Facial Expressions", Journal of Digital Contents Society Vol.21, No.2, pp.415-422, 2020. <https://doi.org/10.9728/dcs.2020.21.2.415>
- [2] Yun-Su Kim, Su-In Lee, Jong-Won Seok, "The improved facial expression recognition algorithm for detecting abnormal symptoms in infants and young children", Journal of IKEEE, Vol.25, No.3, pp.430-436, 2021. <https://doi.org/10.7471/ikeee.2021.25.3.430>
- [3] Kyeong Tae Kim, Jae Young Choi, "Weighted Soft Voting Classification for Emotion Recognition from Facial Expressions on Image Sequences", Journal of Korea Multimedia Society, 20 (8), pp.1175-1186, 2017. <https://www.dbpia.co.kr/journal/articleDetail?nodeId=N ODE07236596>
- [4] In-kyu Choi, Ha-eun Ahn, Jisang Yoo, "Facial Expression Classification Using Deep Convolutional Neural Network", Journal of Electrical Engineering and

- Technology, 13 (1), 485-492, 2017.
<http://doi.org/10.5370/JEET.2018.13.1.485>
- [5] A. Juhong and C. Pintavirooj, “Face recognition based on facial landmark detection”, in Proc. of 2017 10th Biomedical Engineering International Conference (BMEiCON), pp.1-4, 2017.
<https://doi.org/10.1109/BMEiCON.2017.8229173>
- [6] Jin Hee Bae, Bo Hyeon Wang, Joon S. Lim, “Study for Classification of Facial Expression using Distance Features of Facial landmarks”, Journal of IKEEE, 25 (4), pp.613-618, 2021.
<https://doi.org/10.7471/ikeee.2021.25.4.613>
- [7] Young-Eun An, Ji-Min Lee, Min-Gu Kim, Sung-Bum Pan, “Classification of Facial Expressions Using landmark-based Ensemble Network”, Journal of Digital Contents Society, 23 (1), pp.117-122, 2022.
<https://doi.org/10.9728/dcs.2022.23.1.117>
- [8] Jungsu Gu, Ho Chul Kang, “Facial landmark Detection by Stacked Hourglass Network with Transposed Convolutional Layer”, Journal of Korea Multimedia Society, 24 (8), pp.1020-1025, 2021.
<https://doi.org/10.9717/kmms.2021.24.8.1020>
- [9] Do Yeop Kim, Ju Yong Chang, “Integral Regression Network for Facial landmark Detection”, Journal of Broadcast Engineering, 24(4), pp.564-572, 2019.
<https://doi.org/10.5909/JBE.2019.24.4.564>
- [10] Myoung-Kyu Sohn, Sang-Heon Lee, Hyunduk Kim, “Analysis and implementation of a deep learning system for face and its landmark detection on mobile applications”, Proceedings of Symposium of the Korean Institute of communications and Information Sciences, pp.920-921, 2021.
- [11] Kyung-Seob Yoon, SangWon Lee, “Music player using emotion classification of facial expressions”, Proceedings of the Korean Society of Computer Information Conference, 27 (1), pp.243-246, 2019.
<https://koreascience.kr/article/CFKO201909258121757.page>
- [12] Wonjae Choi, Taehwan Kim, Hyunseok Bae, Jungsun Kim, “Comparison of Emotion Prediction Performance of CNN Architectures”, 한국정보과학회 학술발표논문집, pp.1029-1031, 2019.
- [13] Sirojiddin Salimov, YOOJaehung, “A Design of Small Scale Deep CNN Model for Facial Expression Recognition using the Low Resolution Image Datasets”, The Journal of the Korea institute of Electronic Communication Sciences, 16 (1), pp.75-80, 2021. <https://doi.org/10.13067/JKIECS.2021.16.1.75>
- [14] AI-Hub, <https://www.aihub.or.kr/aihubdata/data/view.do?currMenu=115&topMenu=100&aihubDataSe=realm&dataSetSn=82>
- [15] F1 Score,
<https://bhcbboy100.medium.com/머신러닝-분류-평가-지표-이해하기-정확도-정밀도-재현율-f1-스코어-6bf91535a01a>

◎ 저 자 소개 ◎



정 이 나(Yi-na Jeong)

2011년 가톨릭관동대학교 컴퓨터공학과(공학사)
2012년 가톨릭관동대학교 대학원 컴퓨터공학과(공학석사)
2018년 가톨릭관동대학교 대학원 컴퓨터공학과(공학박사)
2018년~현재 가톨릭관동대학교 소프트웨어학과 교수
관심분야 : Sensor Network, IT security, Network security
E-mail : lupinus07@nate.com



심 용 보(Yong-bo Sim)

2021년 가톨릭관동대학교 소프트웨어학과(공학사)
2021년~현재 가톨릭관동대학교 대학원 소프트웨어학과 수료
관심분야 : CNN, C++, 물리엔진, etc.
E-mail : simyongbo@cku.ac.kr



손 수 락(Su-rak Son)

2018년 가톨릭관동대학교 컴퓨터공학과(공학사)
2020년 가톨릭관동대학교 대학원 컴퓨터공학과(공학석사)
2022년 가톨릭관동대학교 대학원 컴퓨터공학과(공학박사)
2022년~현재 가톨릭관동대학교 강사
관심분야 : 빅데이터, 인공지능, 프로그래밍 언어, etc.
E-mail : sonsur@naver.com