논문 2021-2-1 http://dx.doi.org/10.29056/jsav.2021.12.01

얼굴 열화상 기반 감정인식을 위한 CNN 학습전략

이동환*, 유장희**†

Divide and Conquer Strategy for CNN Model in Facial Emotion Recognition based on Thermal Images

Donghwan Lee*, Jang-Hee Yoo**†

요 약

감정인식은 응용 분야의 다양성으로 많은 연구가 이루어지고 있는 기술이며, RGB 영상은 물론 열화상을 이용한 감정인식의 필요성도 높아지고 있다. 열화상의 경우는 RGB 영상과 비교해 조명 문제에 거의 영향을 받지 않는 장점이 있으나 낮은 해상도로 성능 높은 인식 기술을 필요로 한다. 본 논문에서는 얼굴 열화상 기반 감정인식의 성능을 높이기 위한 Divide and Conquer 기반의 CNN 학습전략을 제안하였다. 제안된 방법은 먼저 분류가 어려운 유사 감정 클래스를 confusion matrix 분석을 통해 동일 클래스 군으로 분류하도록 학습시키고, 다음으로 동일 클래스 군으로 분류된 감정 군을 실제 감정으로 다시 인식하도록 문제를 나누어서 해결하는 방법을 사용하였다. 실험을 통하여, 제안된 학습전략이 제시된 모든 감정을 하나의 CNN 모델에서 인식하는 경우보다 모든 실험에서 높은 인식성능을 보이는 것을 확인하였다.

Abstract

The ability to recognize human emotions by computer vision is a very important task, with many potential applications. Therefore the demand for emotion recognition using not only RGB images but also thermal images is increasing. Compared to RGB images, thermal images has the advantage of being less affected by lighting conditions but require a more sophisticated recognition method with low-resolution sources. In this paper, we propose a Divide and Conquer-based CNN training strategy to improve the performance of facial thermal image-based emotion recognition. The proposed method first trains to classify difficult-to-classify similar emotion classes into the same class group by confusion matrix analysis and then divides and solves the problem so that the emotion group classified into the same class group is recognized again as actual emotions. In experiments, the proposed method has improved accuracy in all the tests than when recognizing all the presented emotions with a single CNN model.

한글키워드 : 얼굴 감정인식, 열화상, CNN 학습전략, 분할 정복법

keywords : Facial emotion recognition, Thermal image, CNN training strategy, Divide and Conquer

1. 서 론

감정은 인간의 기억, 학습, 추론 및 문제 해결 등의 인지 과정에 많은 영향을 미친다[1]. 일상에 있어 중요한 요소인 감정인식에 관한 연구는 자

* 과학기술연합대학원대학교 ICT전공
** 한국전자통신연구원 인공지능연구소
† 교신저자: 유장희(email: jhy@etri.re.kr)
접수일자: 2021.11.12. 심사완료: 2021.12.07.
게재확정: 2021.12.20.

율주행 자동차[2], 온라인교육[3], 지능형 로봇[4] 등 다양한 분야에서 진행되고 있다. 감정 감지 및 인식 시장은 2020년 195억 달러로 연평균 11.3%의 성장률을 보이고 있으며, 2026년에는 371억 달러에 이를 것으로 전망되고 있다[5]. 인간의 감정을 인식하기 위해서는 표정, 음성, 언어, 행동 등 다양한 형태로 표출되는 감정 관련 정보를 이해하고 추출해야 한다. 특히, 표정은 감정을 나타내기 위한 핵심 요소 중 하나이며, 몸짓, 손짓보다 더욱 효과적인 시각적 신호로 알려져 있다[6].

표정에 기반한 감정인식은 RGB 얼굴 영상을 이용한 연구가 가장 많이 진행되고 있으나 조명 변화 등 극복해야 할 어려운 문제들을 가지고 있다[7]. 또한, 놀람 또는 두려움과 같은 유사 표정[8]은 감정인식을 더욱 어렵게 하고 있다. 최근 들어 COVID-19 확산 방지를 위한 체온 측정용 열 적외선 센서의 급속한 보급 및 관련 기술의 발전으로 얼굴의 열화상(thermal image)을 이용한 감정인식 연구의 필요성이 높아지고 있다[9]. 열화상은 조명에 영향을 거의 받지 않으며, 감정 변화에 따른 온도 정보를 포함하고 있다[10]. 즉, RGB 영상의 조명 문제를 극복할 수 있는 장점이 있으나 낮은 해상도 등의 문제로 더욱 높은 성능의 인식 방법을 요구한다.

얼굴 열화상 기반 감정인식은 열화상으로부터 추출된 특징을 활용한 기계학습 기반의 감정인식 방법에서 출발하였다. Kopaczka 등[11]은 HoG, LBP 등을 이용하여 추출한 특징을 기반으로 SVM, k-NN 등을 사용한 감정인식을 수행하였으며, 제안한 방법이 사람과 비교해 약 5.51% 높은 인식성능을 보였다. 최근에는 얼굴 열화상 기반의 감정인식에도 딥러닝을 적용하는 연구가 진행되고 있다. Elbarawy 등[12]은 2개의 콘볼루션 층(convolutional layer)과 2개의 Max-pooling 층으로 구성된 CNN(Convolutional Neural Networks)을 제안하였다. 제안된 기술은 기계학습 방법보다 약 6.7% 향상된 인식성능을 보였으며, 오토인코더(Autoencoder) 형태의 딥러닝 모델보다 약 6.7% 높은 정확도와 5배 빠른 처리 시간을 보였다.

한편, CNN 모델의 일반화 성능 향상을 위한 학습전략에 관한 많은 연구가 진행되고 있다. He 등[13]은 learning rate에 대한 batch size의 비율을 변화시키며, 상관관계를 분석하는 방법을 제시하였다. Wu 등[14]은 CNN의 Max-pooling과 완전 연결 층(fully-connected layer)에 드롭아웃(dropout)을 적용한 학습전략을 제안하였다. 또한, 성능 향상을 위해 데이터의 클래스 불균형 문제 해결이 필요하며, 이를 위해 다양한 데이터를 활용하는 방법, 가중치 부여 방법, 데이터 샘플링과 알고리즘을 결합한 방법 등에 관한 연구가 진행되었다[15]. 그러나, 많은 클래스의 데이터를 하나의 모델에서 한 번에 학습 및 추론하는 문제는 쉽게 해결하기 어려운 도전적 문제이다.

본 논문에서는 제시된 모든 감정의 분류를 하나의 모델에서 학습하는 일반적 CNN 모델의 성능 향상을 위해 Divide and Conquer 기반의 학습전략을 제안하였다. 제안된 감정인식 과정은 입력 영상에 대한 얼굴검출을 수행하고, 감정인식을 위해 최적화된 ResNet-18 기반 CNN 모델 기반의 학습을 수행하였다. 여기서, 분류가 어려운 유사 감정들을 동일 클래스 군으로 먼저 분류하도록 학습시키고, 이렇게 분류된 감정을 다시 실제 감정으로 인식하는 Divide and Conquer 전략을 적용하였다. 또한, 제안된 방법의 성능 검증을 위하여, 기존의 방법과 인식성능을 비교하였다. 그리고, 실험에서 제안된 Divide and Conquer 학습전략이 기존의 방법보다 성능 향상에 도움이 되었음을 확인할 수 있었다.

2. 연구 방법

본 연구에서 제안하는 감정인식 과정은 그림 1과 같다. 먼저 입력 영상에서 불필요한 영역을 제거하기 위해, 얼굴검출을 수행하였다. 그리고, 과적합 방지를 위하여 히스토그램 평활화, flip, rotation과 같은 데이터 증강기법을 적용하였다. 다음으로 ResNet-18 CNN 모델을 얼굴 열화상 기반 감정인식을 위해 최적화하고, 먼저 하나의 모델에서 모든 감정을 학습 및 추론하는 과정을 통해 confusion matrix를 생성하였다. 이를 통하여, 분류가 어려운 유사 감정을 동일 클래스로 그룹화하고, 학습을 통하여 1차 분류를 수행하였다. 그리고, 이렇게 분류된 유사 감정 군을 2차로 다시 학습하여, 실제 감정으로 인식하는 학습 및 추론을 수행하도록 하였다. 다음 각 절에서는 얼굴 열화상 데이터셋과 본 연구에서 제안된 감정인식의 각 단계를 자세히 설명하였다.

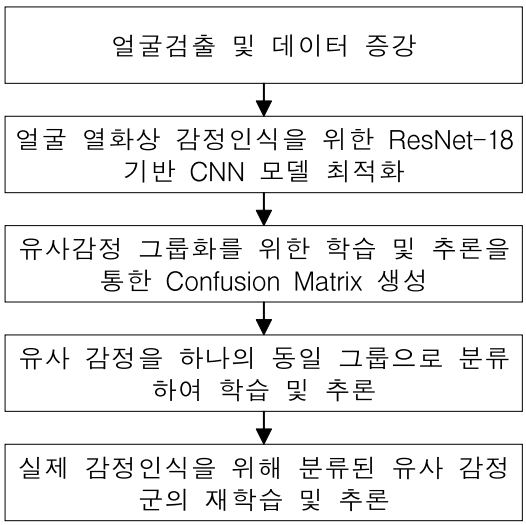

그림 1. 제안된 감정인식 과정
Fig. 1. Proposed Emotion Recognition Procedure

2.1 얼굴 열화상 데이터셋

열화상 카메라는 피사체의 표면으로부터 방출되는 열을 적외선으로 감지하여, 이를 영상으로 변환해주는 장치로 7㎛-14㎛의 장파장 적외선 센서를 사용한다. 따라서, 가시광선보다 파장이 훨씬 큰 에너지를 감지해야 하는 특성상 해상도를 크게 하기 어려운 문제가 있다. 즉, 열화상은 RGB 영상보다 해상도가 낮으며, 기본적으로 하나의 채널을 사용하는 회색(grey) 영상으로 변환된다. 또한 사람의 시각으로는 인지하기 쉽지 않은 부분이 있으므로 감정인식 데이터셋의 태깅을 위하여, RGB 영상과 함께 수집되는 경우가 많다[16-18]. 표 1은 얼굴 열화상 공개 데이터셋을 요약한 것으로 조명, 카메라와 참여자 간 거리, 얼굴 포즈, 안경 착용 등을 고려하여 수집되었다.

표 1. 얼굴 열화상 데이터셋
Table 1. Facial Thermal Image Datasets

데이터셋	IRIS[16]	NVIE[17]	RWTH[11]	Tufts[18]
공개연도	2003	2010	2018	2020
해상도	320 × 240	320 × 240	1024 × 768	336 × 256
촬영거리	198 cm	75 cm	90 cm	150 cm
참여자 수	30	105~112	90	113
열화상 수	2,640	18,312	2,935	1,582
감정 수	3종	7종	8종	4종
조명 방향	5	3	-	-
얼굴 포즈	11	1	1/9	1/9

- : Not enough information

IRIS[16] 데이터셋은 30명으로부터 획득한 얼굴의 RGB 영상 및 열화상으로 구성되었다. 놀람, 행복, 슬픔의 감정 영상과 왼쪽, 오른쪽 등 조명 변화에 따른 영상이 포함되어 있다. 그리고, 감정과 조명마다 좌에서 우로 변하는 얼굴 포즈 영상이 포함되어 있다. 안경은 참여자가 원하는 경우만 착용하도록 하였다. NVIE[17] 데이터셋은 자연스러운 감정변화 영상과 촬영자의 요청으로 연기한 감정 영상으로 구성되어 있다. 무표정, 행복, 화남, 역겨움, 공포, 슬픔, 놀람의 감정 영상은 정면, 왼쪽, 오른쪽의 조명 변화에 따라 수집되었으며, 안경을 작용한 영상과 착용하지 않은 영상

모두 존재한다. 영상 손상, 촬영 거부 등의 이유로 일부의 영상은 데이터셋에서 누락되어 사용할 수 없다.

RWTH[11] 데이터셋은 열화상으로만 구성되어 있다. 무표정, 행복, 슬픔, 놀람, 공포, 역겨움, 화남을 연기한 감정, 자연스럽게 표출된 감정 등의 영상과 S자 형태로 움직이는 얼굴 포즈 영상이 포함되어 있다. 얼굴 열화상에서 추출된 68개 특징점을 통해 입, 눈, 코 등 특정 영역을 분리하여 사용할 수 있다. Tufts[18] 데이터셋은 RGB, 열, NIR 등의 영상으로 구성되어 있다. 촬영자의 요청으로 연기한 무표정, 행복, 졸림, 놀람의 감정 영상과 안경을 착용한 영상이 포함되어 있다. 얼굴 포즈 영상은 참여자가 직접적으로 움직이지 않고, 카메라의 움직임에 의해 수집되었다. 두 개의 데이터셋 모두 정면에서 수집된 영상 폴더와 9방향의 포즈를 고려하여 촬영한 별도의 영상을 포함하고 있다.

2.2 전처리 및 데이터 증강

감정인식을 위한 특징 영역을 증가시키기 위해 얼굴 외의 불필요한 영역을 제거하였다. 이를 위해 RetinaFace 검출기를 사용하였으며, RetinaFace는 학습 앵커에 대한 얼굴 분류 손실, 얼굴 영역 회귀 손실, 얼굴 랜드마크 회귀 손실, 밀집 회귀 손실 최소화를 동시에 수행한다[19]. RGB 영상 학습기반의 얼굴 검출기이나 열화상 데이터셋에서 얼굴을 모두 검출할 수 있었다. 검출된 얼굴 영역의 크기는 사람마다 각기 다르게 나타나므로 검출된 영역의 평균 픽셀 수 기준으로 크기를 정규화하였다.

데이터 증강은 제한된 학습 데이터를 과도하게 학습하여, 학습하지 않은 데이터에 대한 일반화 성능의 감소 방지를 위한 기법이다[20]. 검출된 얼굴 영상의 대비(contrast) 향상을 위해 히스토그램 평활화(histogram equalization)를 적용하였다. 그림 2는 검출된 얼굴 영상과 히스토그램 평활화에 의해 생성된 얼굴 영상의 예를 나타내고 있다. 또한, 검출된 얼굴 영상에 대해 CNN 학습에 주로 사용되는 horizontal flip과 random rotation을 적용하였다.

그림 2. 히스토그램 평활화에 의해 생성된 영상 (a) 원본 영상, (b) 결과 영상
Fig. 2. Image using Histogram Equalization (a) Original Image (b) Histogram Equalized Image

2.3 ResNet-18 CNN 구조의 최적화

본 연구에서는 열화상 감정인식을 위한 학습 데이터의 수가 충분하지 않으므로 사전 학습된 ResNet[21] 중 가장 얕은 18개의 층으로 구성된 모델을 참고하였다. ResNet-18은 224×224 공간 크기(Spatial size)의 3채널 영상을 사전 학습한 모델이다. 학습 및 추론을 위한 열화상은 이보다 공간 크기가 작으며, 하나의 채널로 구성된다. 따라서, 첫 번째 콘볼루션 층의 필터 크기와 채널을 변경하였으며, 공간 크기를 감소시키지 않고 그대로 유지하였다. 두 번째 층은 학습을 통해 특징 맵을 추출하기 위해, Max-pooling 층 대신 콘볼루션 층을 사용하였다. 또한, 완전 연결 층의 노드 개수는 분류할 감정의 개수와 일치시켰다.

그림 3은 얼굴 열화상 감정인식을 위해 최적화한 CNN 모델의 구조를 나타낸 것이다. 그림에서 각 박스는 콘볼루션 층의 필터 크기와 출력 차원

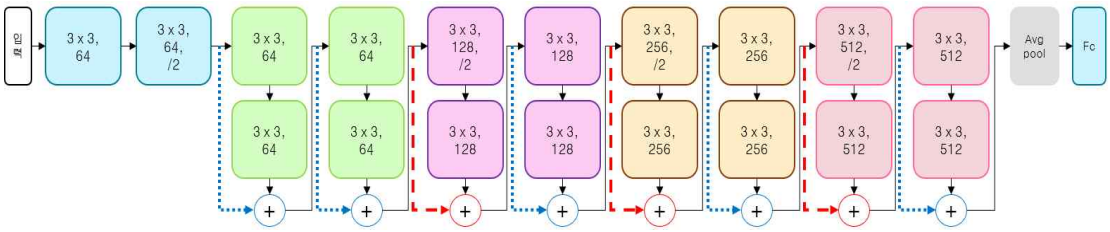

그림 3. 본 연구에서 사용된 CNN 구조
Fig. 3. CNN Architecture used in This Study

을 나타내며, 이전 층의 출력을 유지하여 다음 층에 더해주는 Residual 블록을 확인할 수 있다. 점선은 이전 층의 출력을 그대로 유지한다. 파선(Dashed line)은 이전 층의 특징 맵 개수를 2배로 증가시키며, 출력 공간 크기를 절반으로 감소시킨 출력을 유지한다.

2.4 Divide and Conquer 기반 CNN 학습전략

일반적인 CNN 학습 방법은 하나의 모델에서 제시된 모든 감정을 학습시켜 인식하는 것이다. 그러나, 분류할 감정의 개수가 증가할수록 유사 감정도 증가함으로 인식의 정확도가 감소하는 문제가 존재한다. 이러한 문제를 해결하기 위하여 적용한 Divide and Conquer는 하나의 문제를 보다 작은 여러 개의 문제로 나누어 해결하는 방법을 의미한다. 즉, 분류가 어려운 유사 감정들을 동일 클래스 군으로 먼저 분류하고, 이렇게 분류된 감정을 다시 실제 감정으로 인식할 수 있도록 문제를 나누어서 해결하는 것을 의미한다. 이 경우 하나의 CNN 모델이 아닌 나누어진 문제의 개수 만큼에 CNN 모델이 필요하므로 계산량은 증가할 수 있으나 보다 효과적으로 문제를 해결할 수 있을 것이다.

본 연구에서는 그림 3의 열화상 감정인식에 최적화된 CNN 모델을 이용하여, 인식의 성능을 향상시킬 수 있는 방법을 제안하였다. 제안하는 학습전략을 보다 자세히 설명 하기 위해, n개 감정으로 구성된 데이터셋 E를 다음과 같이 정의한다.

$$E = [C_1, C_2, ..., C_n] \tag{1}$$

여기서, C_n는 데이터셋을 구성하고 있는 감정 클래스를 나타낸다. 하나의 CNN 모델에서 제시된 n개의 감정을 학습시킨 후 추론 결과를 confusion matrix로 분석하여, 잘못 분류되는 감정을 확인할 수 있다.

즉, 유사도가 높은 감정 클래스는 거짓 양성(false positive)과 거짓 음성(false negative)의 개수가 많은 경우를 의미한다. 이를 통하여, 집합 E를 m개의 유사 클래스 군 G로 다음과 같이 정의할 수 있다.

$$G = [G_1, G_2, ..., G_m] \tag{2}$$

그리고, 분류된 유사 클래스 군 G를 다시 분류하여, 실제 감정인식을 수행해야 하는 집합 S_m을 다음과 같이 정의할 수 있다.

$$S_m = [G_{m,p}], p > 1 \tag{3}$$

여기서, p 는 각각의 유사 클래스 군 G에 실제 클래스 개수를 의미한다. 클래스 개수가 2개 이상의 경우에 감정인식을 수행한다. 식 (1)과 같이 제시된 모든 감정을 한 번에 학습 및 추론하는

CNN 학습 방법보다 식 (2), (3)과 같이 유사 감정 클래스 군으로 1차 분류한 후 각각의 클래스 군에서 2차적으로 감정을 인식하는 학습전략을 사용함으로써 감정인식 성능을 향상할 수 있을 것이다.

본 연구에서 제안하는 학습전략을 요약하면 그림 4와 같다. 앞에서 기술한 것과 같이 먼저 하나의 CNN 모델에서 제시된 모든 감정을 인식하는 ⓐ의 과정을 수행한다. 과정 ⓐ의 결과로 confusion matrix의 거짓 양성과 거짓 음성 개수 분석을 통해 유사 감정 클래스 군의 그룹화 과정 ⓑ를 수행한다. 그리고 분할된 유사 감정 클래스 군을 다시 분류하는 과정 ⓒ를 수행한다.

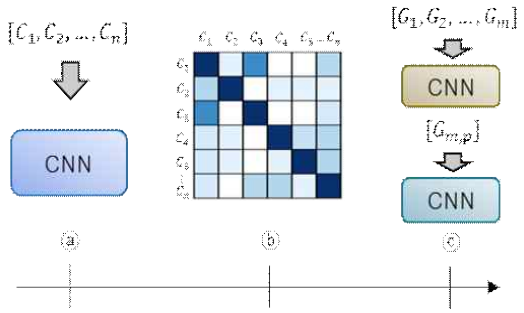

그림 4. Divide and Conquer 전략을 적용한 학습과정
Fig. 4. Procedure of the Divide and Conquer Training Strategy

3. 실험 및 검토

본 연구에서 제안된 얼굴 열화상 기반 감정인식을 위한 CNN 학습전략의 성능 평가를 위하여, 열화상 데이터셋들 중 가장 최근에 공개된 Tufrs DB[18]를 사용하여 실험을 진행하였다. Tufts 데이터셋에서 선글라스를 착용한 영상을 제외한 가용한 얼굴 열화상 원본 데이터의 수는 446장이다. 첫 번째 실험인 Divide and Conquer 기반 CNN 학습전략의 성능을 비교하기 위하여, 이들 중 약 20% 수준인 88장을 테스트 데이터로 사용하였다. 그리고, 나머지 358장의 원본 영상과 학습 프로그램에서 자동으로 이들 원본 영상을 flip과 rotation에 의해 증강한 716장을 포함하여, 총 1,074장의 영상을 학습에 사용하였다.

두 번째 실험은 Tufts 데이터셋을 이용하여 감정인식을 수행한 TERNet[22]과 가능한 유사한 조건에서 성능을 비교하기 위한 데이터셋을 구성하였다. 이를 위해, 먼저 원본 영상에 히스토그램 평활화를 적용하여 생성된 892개의 영상 중 20% 수준인 179장을 선택하여 테스트 데이터로 사용하도록 하였다. 그리고, 나머지 713장의 영상에 대해 첫 번째 실험에서와 같은 데이터 증강기법을 적용하여 생성된 1,426장의 영상을 포함한 총 2,139장의 영상을 학습에 사용하였다. 표 2는 앞서 기술한 것과 같이 실험에 사용한 데이터셋의 구성을 정리한 것이다.

표 2. 실험에 사용된 데이터셋
Table 2. Experimental Dataset

실험 구분	해상도 (정규화)	학습 영상 개수		테스트 영상 개수
		원본 + (전처리)	증강된 영상	
제안된 학습전략	130×160	358	716	88
선행연구와 성능 비교	130×160	713	1,426	179

제안된 CNN 모델의 구조 및 데이터 증강을 위해 PyTorch 및 OpenCV를 이용하였으며, NVIDIA RTX A6000 GPU를 사용하여 학습 및 추론을 하였다. CNN 모델을 위한 하이퍼파라미터로 learning rate는 0.001, epoch은 10,000으로 설정하고, optimizer는 Adam, 그리고 loss function은 L1을 사용하였다.

3.1 CNN 학습전략 비교 실험

제안된 Divide and Conquer 기반 CNN 학습전략이 일반적인 방법보다 성능을 높일 수 있는

지 확인하기 위한 실험을 수행하였다. 먼저 제안된 CNN 구조를 사용하여, 하나의 모델에서 제시된 모든 감정을 학습 및 추론하는 일반적 방법을 이용한 인식성능을 측정하고, 그림 5의 confusion matrix를 생성하였다. 그리고, confusion matrix의 분석을 통해 Neutral과 Smile이 오 분류가 높은 가장 유사 감정 군이며, Neutral과 Shock는 두 번째의 유사 감정 군임을 확인할 수 있다.

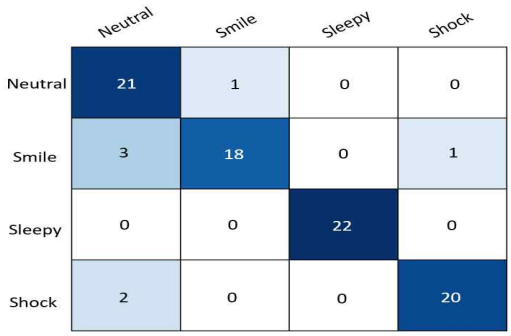

그림 5. 유사 감정 군 분류를 위한 Confusion Matrix
Fig. 5. Confusion Matrix for Similar Emotion Groups

그리고, Divide and Conquer 학습전략에 의해 유사 감정 군으로 분류된 클래스를 1차 분류한 후 실제 감정인식을 수행하였다. 실험에서 유사 감정으로 분류된 수가 비교적 많지 않음으로 유사 감정 군의 분류에 따른 성능을 확인하기 위하여, 다양한 유사 감정 군 분류에 따른 실험을 수행하였다. 표 3은 각각의 학습전략과 유사 감정 군 분류에 따른 인식성능을 측정한 실험 결과를 보여주고 있다.

표에서 제시된 실험 결과와 같이 제안된 CNN 구조를 사용하여, 제시된 모든 감정을 하나의 모델에서 학습한 인식성능은 92.05%를 나타내었다. 그리고, 첫 번째와 두 번째의 유사 감정 군 분류를 통한 인식성능은 각각 94.32%와 95.45%를 나타내고 있다. 결과적으로 제안된 Divide and Conquer 기반의 CNN 학습전략이 모든 경우 기존의 학습 방법보다 2.27~3.40% 높은 성능을 보였음을 확인할 수 있었다.

표 3. CNN 학습전략 실험 결과
Table 3. Experimental Results for CNN Training Strategy

사용된 학습 방법	유사 감정군 분류	인식성능 (%)
제안된 CNN 구조	(Neutral), (Smile), (Sleepy), (Shock)	92.05
제안된 CNN 구조에 Divide and Conquer 학습전략 사용	(Neutral+Smile), (Sleepy), (Shock)	94.32
	(Neutral+Shock), (Smile+Sleepy)	94.32
	(Neutral+Shock), (Smile), (Sleepy)	95.45

3.2 선행연구와의 비교 실험

본 연구에서 얼굴 열화상 감정인식을 위해 최적화한 CNN 구조와 이를 기반으로 Divide and Conquer 학습전략을 적용한 방법의 인식성능을 기존의 선행연구와 비교하는 실험을 수행하였다. 표 4는 선행연구와의 비교 실험 결과를 보여주고 있다.

표 4. 성능 비교 실험 결과
Table 4. Experimental Results for Performance Comparison

사용된 학습 방법	유사 감정군 분류	인식성능 (%)
TERNet[22]	(Neutral), (Smile), (Sleepy), (Shock)	96.20
제안된 CNN 구조	(Neutral), (Smile), (Sleepy), (Shock)	97.77
제안된 CNN 구조에 Divide and Conquer 학습전략 사용	(Neutral+Shock), (Smile), (Sleepy)	98.88

선행연구인 TERNet[22]의 경우는 다양한 전처리 기법 및 데이터 증강기법으로 생성된 영상을 포함한 총 5,876장의 영상 중 5,424장을 학습에 452장을 테스트에 사용하였다. 여기서, 테스트 데이터는 데이터 증강 후 무작위로 선택되어, 동일 참여자의 증강된 감정 영상이 학습과 테스트

양편에 존재하는 것을 배제하지 않았다. 여하튼, 제안된 방법은 선행연구보다 실험에 사용된 영상의 해상도가 낮고, 학습 데이터의 개수가 적음에도 1.57%와 2.68% 높은 성능을 보여주고 있다.

제안된 방법은 얼굴 열화상 감정인식 데이터셋에 대한 인식성능의 향상을 기대할 수 있으나 기존의 방법과 같이 하나의 CNN 모델이 아닌 분할된 문제 개수만큼의 CNN 모델에 대한 학습 및 추론 과정이 필요하며, 보다 많은 컴퓨팅자원을 필요로 한다. 또한, 표 3의 결과와 같이 유사 클래스의 분류에 따라 다소 인식성능의 차이를 보여주고 있어, 분류할 클래스가 많은 경우는 컴퓨팅자원뿐만 아니라 유사 클래스 분류에 어려움이 있을 수 있을 것으로 예상된다.

5. 결 론

본 논문에서는 RGB 영상과 비교해 상대적으로 작은 해상도를 갖는 얼굴 열화상을 이용한 감정인식 연구에 관하여 기술하였다. 이를 위해, ResNet-18 CNN 구조의 최적화와 Divide and Conquer 기반 CNN 학습전략에 관하여 기술하였다. 또한, 실험을 통하여, 제안된 학습전략이 기존의 방법보다 높은 성능을 보임을 확인하였다. 더불어, 선행연구와의 비교 실험에서도, 얼굴 열화상 감정인식을 위해 최적화한 CNN 구조의 인식성능 향상은 물론 제안된 Divide and Conquer 학습전략 또한 성능 향상에 기여할 수 있음을 확인하였다.

향후, 유사 감정 군에 대한 자동 분류와 분할된 문제의 개수만큼 증가하는 CNN을 하나의 모델로 통합하여, 더욱 쉽게 문제를 해결할 수 있는 CNN 구조에 관한 연구가 필요할 것이다. 또한, 복잡하고 많은 수의 데이터셋에 성공적인 적용을 통한 얼굴분석, 행동인식, 음성인식 등의 다양한 분야에 활용될 수 있도록 하는 연구가 필요할 것이다.

이 논문은 2021년도 정부(과학기술정보통신부)의 재원으로 수행된 연구임. (2019-0-00330, 영유아/아동의 발달장애 조기선별을 위한 행동·반응 심리인지 AI 기술 개발)

참 고 문 헌

[1] C. M. Tyng, H. U. Amin, M. N. M. Saad, and A. S. Malik, “The Influences of Emotion on Learning and Memory”, Frontiers in Psychology, Vol.8, pp.1-22, Aug. 2017. https://doi.org/10.3389/fpsyg.2017.01454

[2] S. Zepf, J. Hernandez, A. Schmitt, W. Minker, and R. W. Picard, “Driver Emotion Recognition for Intelligent Vehicles: A Survey”, ACM Computing Surveys, Vol.53, No.3, pp.1-30, June 2020. https://doi.org/10.1145/3388790

[3] E. Yadegaridehkordi, N. F. B. M. Noor, M. N. B. Ayub, H. B. Affal, and N. B. Hussin, “Affective Computing in Education: A Systematic Review and Future Research”, Computers & Education, Vol.142, pp.1-19, Dec. 2019. https://doi.org/10.1016/j.compedu.2019.103649

[4] Z. Liu et al., “A Facial Expression Emotion Recognition based Human-Robot Interaction System”, in IEEE/CAA Journal of Automatica Sinica, Vol.4, No.4, pp.668-676, Sep. 2017. https://doi.org/10.1109/JAS.2017.7510622

[5] MarketsandMarkets, “Emotion Detection and Recognition Market by Component (Solutions [Facial Expression Recognition, Speech & Voice Recognition] Services),

Technology, Application Area, End User, Vertical, Region-Global Forecast to 2026", https://www.marketsandmarkets.com/Market-Reports/emotion-detection-recognition-market-23376176.html, Mar. 2021.

[6] Z. Yu and C. Zhang, "Image based Static Facial Expression Recognition with Multiple Deep Network Learning", in Proceedings of the 2015 ACM on International Conference on Multimodal Interaction, pp.435-442, Seattle, USA, Nov. 2015. https://doi.org/10.1145/2818346.2830595

[7] I. M. Revina and W. R. S. Emmanuel, "A Survey on Human Face Expression Recognition Techniques", Journal of King Saud University - Computer and Information Sciences, Vol.33, No.6, pp.619-628, July 2021. https://doi.org/10.1016/j.jksuci.2018.09.002

[8] K. Zhao, J. Zhao, M. Zhang, Q. Cui, and X. L. Fu, "Neural Responses to Rapid Facial Expressions of Fear and Surprise", Frontiers in Psychology, Vol.8, pp.1-8, May 2017. https://doi.org/10.3389/fpsyg.2017.00761

[9] D. Poster et al., "A Large-scale, Time-synchronized Visible and Thermal Face Dataset", in Proceedings of the 2021 IEEE/CVF Winter Conference on Applications of Computer Vision, pp.1559-1568, Waikoloa, USA, Jan. 2021. https://doi.org/10.1109/WACV48630.2021.00160

[10] C. Ordun, E. Raff and S. Purushotham, "The Use of AI for Thermal Emotion Recognition: A Review of Problems and Limitations in Standard Design and Data", arXiv preprint arXiv:2009.10589, Sep. 2020. https://arxiv.org/abs/2009.10589

[11] M. Kopaczka, R. Kolk, and D. Merhof, "A Fully Annotated Thermal Face Database and Its Application for Thermal Facial Expression Recognition", in Proceedings of the 2018 IEEE International Instrumentation and Measurement Technology Conference, pp.1-6, Houston, USA, May 2018. https://doi.org/10.1109/I2MTC.2018.8409768

[12] Y. M. Elbarawy, N. I. Ghali, and R. S. El-Sayed, "Facial Expressions Recognition in Thermal Images based on Deep Learning Techniques", International Journal of Image, Graphics and Signal Processing, Vol.11, No.10, pp.1-7, Oct. 2019. https://doi.org/10.5815/ijigsp.2019.10.01

[13] F. He, T. Liu and D. Tao, "Control Batch Size and Learning Rate to Generalize Well: Theoretical and Empirical Evidence", in Advances in Neural Information Processing Systems, pp.1141-1150, Vancouver, Canada, Dec. 2019. https://papers.nips.cc/paper/2019/hash/dc6a70712a252123c40d2adba6a11d84-Abstract.html

[14] H. Wu and X. Gu, "Towards Dropout Training for Convolutional Neural Networks", Neural Networks, Vol.71, pp.1-10, Nov. 2015. https://doi.org/10.1016/j.neunet.2015.07.007

[15] J. M. Johnson and T. M. Khoshgoftaar, "Survey on Deep Learning with Class Imbalance", Journal of Big Data, Vol.6, No.1, pp.1-54, Mar. 2019. https://doi.org/10.1186/s40537-019-0192-5

[16] Besma Abid, "IRIS Thermal/Visible Face Database, IEEE OTCBVS WS Series Bench", vcipl-okstate.org/pbvs/bench/, June 2003.

[17] S. Wang et al., "A Natural Visible and Infrared Facial Expression Database for Expression Recognition and Emotion Inference", in IEEE Transactions on Multimedia, Vol.12, No.7, pp.682-691, Nov. 2010. https://doi.org/10.1109/TMM.2010.2060716

[18] K. Panetta et al., "A Comprehensive Database for Benchmarking Imaging Systems", in IEEE Transactions Pattern Analysis and Machine Intelligence, Vol.42, No.3, pp.509-520, Mar. 2020. https://doi.org/ 10.1109/TPAMI.2018.2884458

[19] J. Deng et al., “Retinaface: Single-stage Dense Face Localisation in the Wild”, arXiv preprint arXiv:1905.00641, May 2019. https://arxiv.org/abs/1905.00641
[20] C. Shorten and T. M. Khoshgoftaar, “A Survey on Image Data Augmentation for Deep Learning”, Journal of Big Data, Vol.6, No.1, pp.1-48, July 2019. https://doi.org/10.1186/s40537-019-0191-0
[21] K. He, X. Zhang, S. Ren and J. Sun, “Deep Residual Learning for Image Recognition”, in Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, pp.770-778, Las Vegas, USA, June 2016. https://doi.org/10.1109/CVPR.2016.90
[22] S. K. M, Kamath, R. Rajendran, Q. Wan, K. Panetta and S. S. Agaian, “TERNet: A Deep Learning Approach for Thermal Face Emotion Recognition”, Mobile Multimedia/Image Processing, Security, and Applications 2019, Vol.10993, pp.1-7, May 2019. https://doi.org/10.1117/12.2518708

저 자 소 개

이동환(Donghwan Lee)

2020.02 충남대학교 컴퓨터공학과 졸업
2020.03-현재 : 과학기술연합대학원대학교 ICT전공 석사과정
<주관심분야> 인공지능, 딥러닝, 영상인식

유장희(Jang-Hee Yoo)

1988.02 한국외국어대학교 물리학과 졸업
1990.02 한국외국어대학교 전산학과 석사
2004.07 영국 University of Southampton 전자 및 컴퓨터과학 박사
1989.11-현재: 한국전자통신연구원 인공지능연구소 책임연구원
2005.09~현재: 한국저작권위원회 감정인, 현)감정전문위원
2007.03-현재: 과학기술연합대학원대학교 전임교수, 캠퍼스 대표교수
2007.01~현재: 한국SW감정평가학회 이사
2014.09~현재: 경찰청 과학수사자문위원
2014.8~2015.8: University of Washington 방문학자
2018.3~2020.3: 국가지식재산위원회 전문위원
<주관심분야> 컴퓨터 비전, 인공지능, 생체인식, 휴먼모션분석, HCI 및 지능형 로봇