

감정노동자를 위한 딥러닝 기반의 스트레스 감지시스템의 설계

옥유선 · 조우현*

부경대학교

Stress Detection System for Emotional Labor Based On Deep Learning Facial Expression Recognition

Yu-Seon Og · Woo-hyun Cho*

Pukyong National University

E-mail : yuseon@pukyong.ac.kr / whcho@pknu.ac.kr

요 약

서비스 산업의 성장과 함께 감정노동자의 스트레스가 사회적 문제로 인식되어 2018년 감정노동자 보호법이 시행되었다. 그러나 실질적인 감정노동자 보호 시스템의 부족으로 스트레스 관리를 위한 디지털 시스템이 필요한 시점이다. 본 논문에서는 대표적인 감정노동자인 고객 상담사를 위한 딥러닝 기반 스트레스 감지 시스템을 제안한다. 시스템은 실시간 얼굴검출 모듈, 한국인 감정 이미지 중심의 이미지 빅데이터를 딥러닝한 감정분류 FER 모듈, 마지막으로 스트레스 수치만을 시각화하는 모니터링 모듈로 구성된다. 이 시스템을 통하여 감정노동자의 스트레스 모니터링과 정신질환 예방을 목표로 설계하였다.

ABSTRACT

According to the growth of the service industry, stresses from emotional labor workers have been emerging as a social problem, thereby so-called the Emotional Labor Protection Act was implemented in 2018. However, insufficient substantial protection systems for emotional workers emphasizes the necessity of a digital stress management system. Thus, in this paper, we suggest a stress detection system for customer service representatives based on deep learning facial expression recognition. This system consists of a real-time face detection module, an emotion classification FER module that deep-learned big data including Korean emotion images, and a monitoring module that only visualizes stress levels. We designed the system to aim to monitor stress and prevent mental illness in emotional workers.

키워드

Deep Learning, Emotional Labor, Stress, Facial Expression Recognition, CNN

1. 서 론

감정노동은 고객과의 상호작용이 필요한 직무로 서비스 산업에서 그 비중이 증가하고 있다. 안전보건공단은 근로환경실태 조사에 의하면 전국 감정노동자는 2011년 740만 명, 2017년 1,330만 명[1]으로 대폭 증가하였다. 이러한 추세와 함께 감정노동자의 스트레스에 기인한 심리 건강과 정신질환에 대한 사회문제[1]가 대두되어 2018년 ‘감정노동자 보호법(산업안전보건법 제41조)’이 실시되었다. 그러나, ‘콜센터 노동자의 노동건강 실태’[2]에 의

하면 대표적인 감정노동자인 고객 상담사의 경우 1300여 명 중 80% 이상이 우울증 고위험군으로 판정되었다. 또한 사업주에 대한 조사와 처분이 신고에 의해서만 실시되기에 문제 발생 후에야 조치가 이루어지고 있다. 따라서, 고객 상담사의 스트레스 변화를 실시간으로 관리하는 시스템이 필요한 실정이다. 이에, 이 논문에서는 실시간 스트레스 감지 시스템을 제안한다. 이미지 처리 딥러닝 기술인 CNN 모델에 한국인 이미지 데이터 중심의 빅데이터를 학습하여 감정분류 모듈을 설계하였다. 이를 통해 분류된 실시간 감정을 스트레스 수치로 환산하였고 시각화로 모니터링이 가능토록 하였다.

* corresponding author

II. 관련 연구

2.1 Convolutional Neural Network (CNN)

이미지 분석을 위한 특징적인 패턴을 찾는 데 유용한 알고리즘으로서 CNN은 LECUN, Yann, et al. (1988)[3]에 의하여 제안되었다. 기존의 Fully Connected Layer로 구성된 신경망으로 이미지를 처리할 경우 공간 정보 유실 문제로 특징추출과 정확도 향상이 어렵다. 이미지 공간정보를 유지한 채로 특징추출을 할 수 있는 컨볼루션 연산과 연산량 증가를 방지하기 위해 특징과 채널 수를 유지한 채 출력 크기를 줄이는 풀링(Pooling) 레이어를 적용한다. CNN은 컨볼루션/풀링 레이어를 반복한 특징추출(Feature Extraction)파트와 Fully Connected Layer를 이용한 분류(Classification)파트로 구성된다.

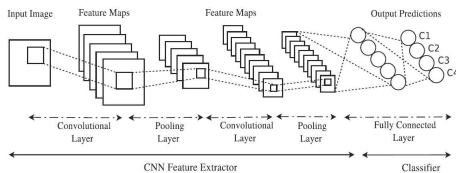


Fig. 4. Architecture of a standard CNN.

그림 1. 스탠다드 CNN의 구조

2.2 얼굴 감정인식(Facial Expression Recognition)

CNN기반 얼굴 감정인식의 선행연구는 대표적으로 얼굴 외관적 특징과 얼굴 랜드마크 포인트에서 기하학적 특징을 수집해 두 정보를 조인트 파인튜닝 (Joint Fine-Tuning)하는 기법[4]과 ILSVRC[5]에서 평가된 CNN 모델을 확보된 데이터셋에 맞게 변형 후 업데이트하는 기법이[6] 있다. 본 논문에서는 후자의 기법을 사용하여 데이터셋을 학습한다.

2.3 스트레스 인식(Stress Recognition)

딥러닝 얼굴 감정인식(FER)의 선행연구를 기반으로 스트레스를 분류하는 연구가 진행되어 왔다. 먼저 2019년 ZHANG, Jin, et al.[7]은 CK+, Oulu-CASIA, KMU-FED 데이터셋 기반 스트레스 검출 모듈을 제안하였다. Almeida, J., & Rodrigues, F.[8]은 CK+, KDEF, Net Images 데이터를 VGG16, VGG19, InceptionResNetV2 모델에 학습하여 angry, disgust, fear를 Stress 감정으로 정의하고 neutral, happy, sad, surprise 감정을 Not-Stress로 이진 분류하는 모듈을 제시하는 등 다양한 데이터셋 기반 스트레스 검출 모듈이 제안되었다. 그러나, 대부분의 연구는 오직 서양인 또는 아시아인 데이터가 소규모 포함된 데이터셋 중심으로 진행되었다. 다른 인종 보다 같은 인종의 얼굴 구성정보가 더 전체적으로 처리된다는 연구 결과[9]에 기반하여 본 논문에서는 감정인식을 위해 한국인 데이터 중심으로 학습한다.

III. 시스템 설계

이 장에서는 제안하는 시스템 설계의 프로세스를 소개한다. 먼저 학습 데이터셋을 확보하고 전처리 후, CNN 모델에 학습하여 FER 모듈을 생성한다. 스트레스 감지를 위해 얼굴검출 모듈에서 실시간 이미지를 추출하고, 학습된 FER 모듈을 통해 4가지의 감정으로 분류한다. 분류된 감정별로 스트레스 수치를 누적 연산하여 그 변화를 시각화한다.

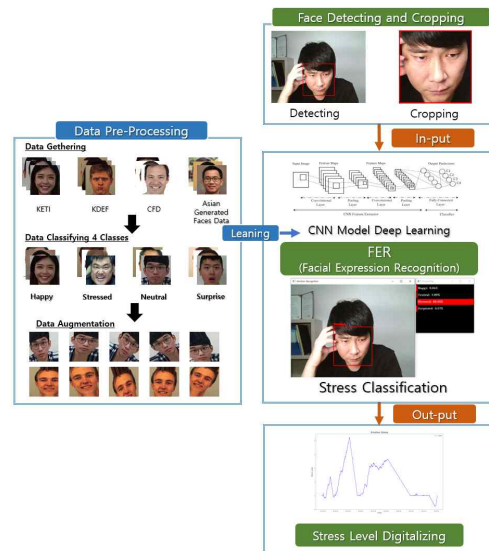


그림 2. 스트레스 감지 시스템의 프로세스 전반

3.1 학습 데이터셋 구성

CNN 모델 학습을 위해서 한국인 데이터인 한국전자기술연구원 지능정보 플래그십 R&D데이터의 감정분류용 데이터셋(KETI)[10]을 중심으로 KDEF [11], CFD[12-14], Asian Generated Faces Dataset(AGFD)[15] 데이터를 확보하였다. 최대 7개의 감정분류 데이터셋에서 angry, disgust, sad는 Stress로 통합, fear와 surprise는 Surprise로 통합하였다. 결과적으로 4개 감정(Happy, Neutral, Stress, Surprise)으로 분류하여 총 9,397개의 데이터셋을 구성하였다.

표 1. CNN모델 학습을 위한 감정별 이미지 개수

Class	Happy	Neutral	Stress	Surprise	Total
DataSet	1,937	2,812	2,787	1,861	9,397

3.2 데이터 전처리(Data Pre-Processing)

CNN 모델의 학습 효율성을 위해 얼굴 이미지에서 감정이 표현된 영역만 필요하다. 추후 서술할 얼굴검출(Face Detection and Cropping) 모듈로 얼굴 영역만 추출한다. 또한, FER 모듈 성능에 문제를 일으키는 과적합(Over-fitting)의 방지를 위해 데이터셋에 이미지증강(Data Augmentation)기법으로

학습 데이터를 증가시킨다. 사용된 기법은 30° Rotation, Width/Height Shift, Shear, Zoom, Horizontal_flip이며, Shape은 100X100, 3채널로 구성하였다.

3.3 얼굴검출(Face Detection and Cropping) 모듈

감정분류를 위해 FER 모듈에 입력되는 얼굴 이미지 추출을 위한 단계로써 (P Viola, M Jones, 2001)에 의해 제안된 Haar Cascade[16]를 이용한다. 이 알고리즘은 얼굴 영역 간의 밝기 차를 통해 특징을 생성하여 cascade 함수로 검출한다.

3.4 FER(Facial Expression Recognition) 모듈

감정분류를 위한 FER 모듈은 실시간 측정에 적합해야 하며 충분한 정확도를 확보해야 한다. 이에 ILSVRC[5] 참가하여 검증된 CNN 모델 중에서 2015년 참가한 InceptionV3(Top-1 Acc:78.8%)[17], 2016년 참가한 Xception(Top-1 Acc:79%)[18], 그리고 2018년 참가한 MobileNetV2(Top-1 Acc:74.7%)[19], 총 3개 모델에 데이터셋을 학습 후 비교·분석한다.

InceptionV3[17]는 2014년 ILSVRC[5] 대회에서 우수한 GoogLeNet(InceptionV1)이 복잡한 구조 때문에 다양한 변형이 힘들다는 점에서 Inception 모델의 연산량을 줄이는 최적화에 대한 방향성 제시를 통하여 소개되었다. 기존 모델의 Filter를 3x3 Convolution으로 Factorizing을 적용하여 유사한 구조에도 깊어진 Network와 함께 연산량을 감소시켰다. 여기에 Optimizer 변경, Label Smoothing, Factorized7-7, BN-auxiliary를 적용해 성능을 높였다.

Xception[18]모델은 총 14개의 모듈로 36개의 컨볼루션 레이어가 존재하며 residual connection을 사용한다. 특히 Inception 모델 대신 작동방식은 비슷하지만, 연산량이 감소하는 Depthwise Separable Convolution을 수정하여 사용하는 Extreme Inception 모델이다. 이를 통해 더 효과적으로 cross-channel correlations와 spatial correlations를 구분하여 계산할 수 있어 Inception 모델 대비 적은 연산 처리가 가능하다. 이 Xception에서 수정된 Depthwise Separable Convolution을 모든 Convolution에 대체한 모델이 MobileNetV2이다.[19] 또한 전체 Convolution 채널은 줄이고 block 내부에서만 증가 시켜 연산량과 파라미터를 줄였다. 마지막으로 ReLU6를 사용하여 연산속도를 획기적으로 줄이면서 보다 좋은 성능을 확보한 모델이다. 위 3가지 모델의 전체 또는 일부 레이어의 학습 여부, Classify 레이어, Learning Rate, Optimizer를 다양하게 변경하여 실험을 진행하였다.

표2. FER 모듈을 위한 CNN 학습실험 기준

Test	Model	Learning Rate	Optimizer	Classify layer	Pre-trained Model Trainable
1	InceptionV3	0.0001	Adam	Classify layer #1	All
2					None
3		0.001		All	
4		None			
5	MobileNetV2	0.0001	Adam	Classify layer #2	All
6					None
7		0.001		All	
8		None			
9	Xception	0.01	SDG	Classify layer #1	All
10	InceptionV3			Classify layer #2	All
11	Xception	0.0001	Adam	Classify layer #1	63 layer False
12	Xception				36 layer False

Classify layer #1 : Model output - Classify Layer - Output
 Classify layer #2 : Model output - Dense - Dropout - Classify Layer - Output

3.5 스트레스 데이터 수치화(Digitalize) 모듈

FER 모듈에서 분류된 감정 결과를 바탕으로 표 3의 조건에 따라 스트레스 수치를 누적 연산한다. 연산된 정보는 시간 정보와 함께 CSV파일에 저장하고 스트레스 모니터링을 위하여 시각화한다.

표 3. 감정 상태별 스트레스 수치 및 연산기준

감정상태	현재 스트레스 지수	증감수치	증감
Happy	> 0	- 0.02	감소
	≤ 0	- 0.005	감소
Neutral	> 0	- 0.005	감소
	≤ 0	-	변화없음
Stress	All	+ 0.02	증가
Surprise	All	+ 0.005	증가

IV. 실험 결과 및 테스트

본 연구의 스트레스 감지 시스템의 핵심인 FER 모듈은 실험 결과 정확도 98.67%, 손실률 0.04의 성능으로 Test #1의 Xception모델이 채택되었다.

표 4. FER 모듈 실험 정확도 및 손실률 결과

모델	Accuracy		Loss		Top Accuracy Test Model
	최소	최대	최소	최대	
Xception	62.78%	98.67%	0.9076	0.0435	Test #1
InceptionV3	53.84%	97.38%	1.0959	0.0675	Test #5
MobileNetV2	61.84%	97.59%	0.9177	0.0643	Test #9

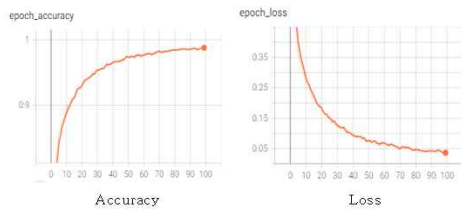


그림 3. Test #1 Xception 학습 결과 그래프

FER 모듈의 실제 인식률을 확인하기 위하여 Windows10, AMD Ryzen7 4800H CPU 환경에서 실시간 테스트를 진행하였다.

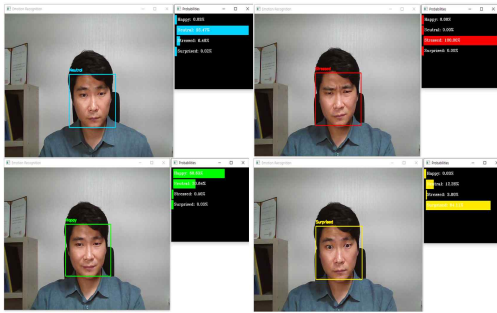


그림 4. 실시간 인식을 테스트

얼굴의 감정표현이 명확한 경우 90% 이상, 비교적 불명확한 경우 65% 이상의 정확성을 보여 해당 표정을 1순위 감정으로 검출 할 수 있었다. 또한 스트레스 데이터 수치화 모듈을 통하여 그림5와 같이 직관적으로 볼 수 있도록 스트레스 수치 변화를 그래프로 시각화 하였다. FER 모듈에서 감정 상태분류를 위해 웹캠으로 이미지를 생성하지만, 개인정보와 업무감시 이슈의 회피를 위해 이미지의 출력과 저장 없이 감정분류 결과만 저장한다. 또한, 고객 상담 시에만 동작하여 상담마다 스트레스 수치를 초기화하므로 그래프를 통하여 상담 건별 스트레스 증감현황을 모니터링할 수 있다.

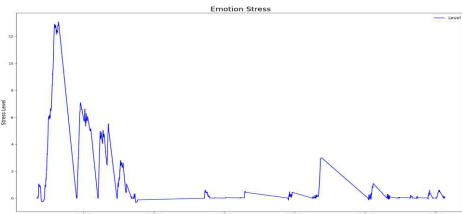


그림 5. 스트레스 수치 모니터링 그래프

V. 결 론

앞서 서술한 바와 같이 한국인 데이터셋을 중심으로 딥러닝하여 스트레스 감지 FER 모듈 생성하고, 이에 기반하여 감정노동자를 위한 스트레스 감지 시스템을 설계하였다. 시스템은 통제 가능한 조건에서는 Top-1 Acc 98%의 높은 학습 정확도를 보여주고 있다. 그러나, 실제 테스트에서는 조명, 성별, 얼굴형에 따라 학습된 FER 모듈별로 인식 정확도가 상이하다는 한계점이 있다. 특히 FER 모델의 정확도에 가장 크게 작용하는 요소는 적절한 조명이다. 따라서, 사용자별로 높은 정확도의 FER 모듈을 채택할 수 있도록 하고, 조명을 확보하여 실사용 데이터를 수집한 후 추후 연구를 진행하게 된다면 개선된 정확도와 활용성을 가진 시스템으로 발전이 가능할 것이다. 이를 바탕으로 고객 상담사뿐만 아니라 다양한 감정노동자의 보호 시스템으로써의 발전을 기대해 볼 수 있다.

References

- [1] 서울특별시, “서울시 감정노동보호 가이드라인 (개정판)”, 2020
- [2] “‘감정노동’ 콜센터 노동자 10명 중 8명 우울증 위험” Hankyoreh news. https://www.hani.co.kr/arti/society/society_general/996594.html.
- [3] LECUN, Yann, et al. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 1998, 86.11: 2278-2324.
- [4] JUNG, Heechul, et al. “Joint fine-tuning in deep neural networks for facial expression recognition.” In: Proceedings of the IEEE international conference on computer vision. p. 2983-2991. 2015.
- [5] “Imagenet” IMAGENET. 2021-03-11 update, <https://image-net.org/index.php>.
- [6] MOLLAHOSSEINI, Ali; CHAN, David; MAHOOR, Mohammad H. “Going deeper in facial expression recognition using deep neural networks.” In: 2016 IEEE Winter conference on applications of computer vision (WACV). IEEE, 2016.
- [7] ZHANG, Jin, et al. “Detecting negative emotional stress based on facial expression in real time.”, 2019 IEEE 4th International Conference on Signal and Image Processing (ICSIP). IEEE, p. 430-434. 2019.
- [8] Almeida, José, and Fátima Rodrigues. “Facial Expression Recognition System for Stress Detection with Deep Learning.” 2021.
- [9] MICHEL, Caroline; CALDARA, Roberto; ROSSION, Bruno. “Same-race faces are perceived more holistically than other-race faces.” Visual Cognition, 14.1: 55-73. 2006.
- [10] 한국전자기술연구원 지능정보플래그십 R&D 데이터 <https://aihub.or.kr/opendata/keti-data/recognition-visual/KETI-01-001>
- [11] E. Lundqvist, D., Flykt, A., & Ohman, A. “The Karolinska Directed Emotional Faces - KDEF, CD ROM from Department of Clinical Neuroscience” Psychology section, Karolinska Institutet, ISBN 91-630-7164-9. 1998.
- [12] Ma, Correll, & Wittenbrink “The Chicago Face Database: A Free Stimulus Set of Faces and Norming Data.” Behavior Research Methods, 47, 1122-1135. 2015.
- [13] Ma, Kantner, & Wittenbrink “Chicago Face Database: Multiracial Expansion.” Behavior Research Methods. 53.3. 1289-1300. 2020.
- [14] Lakshmi, Anjana, et al. “The India Face Set: International and Cultural Boundaries Impact Face Impressions and Perceptions of Category

- Membership.” *Frontiers in psychology* 12 :161. 2021.
- [15] Copyright(c) 2018, BUPT_GWY
<http://www.seeprettyface.com>
- [16] Viola, Paul, and Michael Jones. “Rapid object detection using a boosted cascade of simple features.” *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001. Ieee*, p. I-I. 2001.
- [17] Szegedy, Christian, et al. “Rethinking the inception architecture for computer vision.” *Proceedings of the IEEE conference on computer vision and pattern recognition*. p. 2818-2826. 2016.
- [18] Chollet, Francois. “Xception: Deep learning with depthwise separable convolutions.” *Proceedings of the IEEE conference on computer vision and pattern recognition*. p. 1251-1258. 2017.
- [19] Sandler, Mark, et al. “Mobilenetv2: Inverted residuals and linear bottlenecks.” *Proceedings of the IEEE conference on computer vision and pattern recognition*. p. 4510-4520. 2018.