한국인 감정 세분화 및 표정 인식을 위한 YOLOv8-Face와 ResNet-18 기반 감정 분석 모델의 성능 향상 연구

백서연, 김장현
성균관대학교 실감미디어공학과, 인터랙션사이언스학과/인간AI인터랙션융합학과
e-mail : qortjdus1999@g.skku.edu, alohakim@skku.edu

Enhancing the Performance of YOLOv8-Face and ResNet-18 Based Models for Emotion Segmentation and Facial Expression Recognition of Koreans

Seo Yeon Baek, Jang Hyun Kim

1. 서 론

감정 인식 기술은 다양한 분야에서 활용도가 높을 뿐만 아니라 인간과 기계 간의 상호작용 향상을 위하여 필수적이다. 특히, 한국 사회는 유교적 가치관에 의해 감정 표현을 억제하거나 복잡한 감정을 내포하는 경우가 많으므로, 한국인에 대한 감정 인식 연구가 지속해서 필요하다. 본 연구는 Susan David의 감정분류체계를 따른 한국인 표정 이미지 데이터 세트를 사용하여, 기존보다 더 정확하게 한국인의 감정 표현을 인식할 수 있는 모델 개발을 목표로 한다 [1].

2. 관련 연구

표정 이미지에서의 정확한 얼굴 위치 검출은 얼굴 인식 분야의 전처리 단계에서 중요한 역할을 하며, 크게 두 단계 구조(two-stage)와 단일 단계 구조(one-stage)로 분류할 수 있다. R-FCN(Region-based Fully Convolutional Networks)은 대표적인 두 단계 구조이다. 관심 영역인 RoI(Region of Interest) 제안 단계 이후 분류 단계가 수행되므로, 효율적이지만 복잡하다는 단점이 있다 [2]. 반면, 단일 단계 구조는 시스템 구조가 단순하고 속도가 매우 빠르다. 단일 단계 구조의 대표적인 예로는 YOLO(You Only Look Once)가 있으며, 특히 최신 모델인 YOLOv8-Face는 매우 높은 속도와 정확도 성능을 제공한다. 이는 지연 시간과 파라미터 수 측면에서 높은 성능을 발휘하므로, 본 연구의 목적에 부합하는 YOLOv8-Face 모델을 채택하였다.

3. 방법론

3.1 데이터 수집 및 전처리

AI 허브에서 제공된 '한국인 감정인식을 위한 복합 영상 데이터' 세트의 500,000장 이미지 중 49,000장의 이미지를 무작위로 선정하여 전처리하였다. YOLOv8-Face 모델을 통해 얼굴 영역을 자동으로 감지하고, 224x224픽셀로 크기 조정하여 추출되도록 하였다. 이는 학습 모델인 ResNet-18의 입력 크기가 224x224픽셀이므로, 학습 과정에서 불필요한 크기 조정 작업을 전처리 단계에서 처리하도록 했다. 또한 한 사람당 최대 15개의 이미지까지만 허용하여 각 클래스 내의 이미지 간 불균형을 최소화하였다. 마지막으로, 그림 1처럼 정답 레이블이 잘못 분류되거나, 배경에서 객체를 인식하는 오류가 존재했다. 이러한 오류는 데이터 세트의 정확성에 부정적인 영향을 주기 때문에, 전처리 후 수작업으로 모든 데이터를 다시 검토하였다.

(그림 1) 레이블 오분류 및 배경 객체 인식 예시

최종 데이터 세트는 7가지 감정(기쁨, 슬픔, 당황, 분노, 불안, 상처, 무표정)으로 구성되며, 각 감정 클래스별로 5,000장씩 분포하여 총 35,000장의 데이터를 확보하였다.

3.2 데이터 세트 분할

전처리로 확보한 35,000장의 데이터 세트는 21,000개의 훈련 데이터, 7,000개의 검증 데이터와 7,000개의 테스트 데이터로 분할 하였으며, 균형 잡힌 데이터 분할을 위해 이미지의 인물 ID를 기반으로 나누었다. 각 이미지 파일명에서 고유 ID를 추출하여 각 클래스 내에서 같은 ID를 가진 이미지들을 그룹화하였다. 이후 클래스별로 독립적인 분할을 통해 훈련(60%), 검증(20%), 테스트(20%) 세트에서 이미지 분포를 균일하게 유지하였다. 이를 통하여 모든 클래스와 클래스에 속한 이미지가 학습 과정에서 동등한 기회를 얻도록 하였다.

3.3 모델 학습

ResNet-18 모델을 사용하여 딥러닝 기반의 한국인 감정 표현 인식을 학습하도록 하였다. ResNet-18의 구조는 유지하되, 마지막 완전 연결 계층을 7개의 감정 클래스에 대응하여 ResNet-18의 최종 특징 공간인 512차원을 7차원으로 축소하

였다. 과적합 방지를 위해 드롭아웃과 데이터 증강 방법인 무작위 회전(±10도)을 수행하였으며, 안정적인 모델의 학습을 위해서 사용한 데이터 세트의 평균[0.5675, 0.4436, 0.3918]과 표준편차[0.2492, 0.2085, 0.1894]을 계산하여 정규화 과정을 추가하였다. 또한 그리드 서치를 수행하여, 초기 학습률은 0.0001, 가중치 감쇠는 0.001, 드롭아웃 비율은 0.1, 배치 크기는 32인 최적화된 하이퍼파라미터를 적용하였다. 손실 함수로는 Cross Entropy Loss를 설정하였고, 옵티마이저로 AdamW를 사용하였다. 마지막으로, StepLR 스케줄러를 사용하여 5 에포크마다 학습률이 이전보다 10%로 감소하도록 하였다. 전체 학습은 20 에포크 동안 진행되었으며, 3 에포크 연속으로 이전 검증 정확도보다 향상하지 않으면 조기 종료하였다.

모델 학습 결과, 1 에포크에서 훈련 손실과 정확도는 각각 0.9582와 63.64%였으며, 검증 손실과 정확도는 0.7130와 72.84%로 시작하였다. 20 에포크 동안 학습한 후에는 훈련 손실과 정확도가 0.0942와 97.28%이고, 검증 손실과 정확도는 0.2932와 92.02%로 개선되었다. 최종 검증 정확도가 92.02%이므로 모델이 꾸준히 향상되면서 학습되었으며, 효과적으로 일반화되었음을 나타낸다.

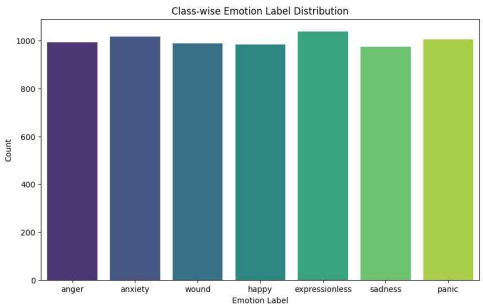
4. 실험 및 평가

4.1 모델 평가 실험 설정

모델 평가 또한 모델 훈련과 마찬가지로 ResNet-18 아키텍처를 사용하였다. 손실 함수로는 Cross Entropy Loss를 사용하였으며, ResNet-18의 기본 가중치로 모델을 초기화하고 훈련 모델을 통해 학습된 최종 모델 가중치를 통해 테스트 데이터 세트를 평가하였다.

4.2 모델 평가 결과 및 분석

각 감정 클래스별 레이블 분포를 분석한 결과, 그림 2와 같이 각 감정의 데이터 개수가 대부분 균등하게 분포하였다. 이는 각 감정에 대한 데이터 세트가 균형 잡혀 있음을 알 수 있다.

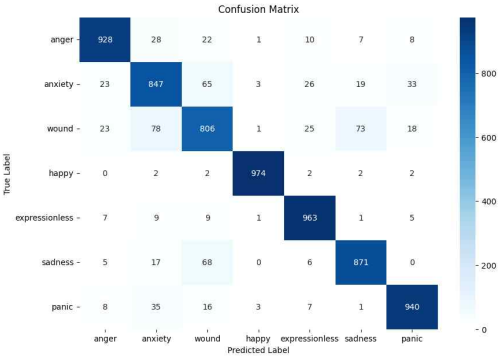


(그림 2) 감정 레이블 분포

또한 혼동 행렬 분석 결과, 그림 3과 같이 '행복'과 '무표정'의 분류 정확도는 매우 높았으며, 비교적 '불안'과 '상처'에 대한 정확도가 낮았다. '불안'이 '상처'로 65회 오분류 되었고, '상처'는 '불안'으로 78회 오분류 되었으며, 이는 '불안'과 '상처'의 표정 구분이 상대적으로 어려운 것을 시사한다.

마지막으로, 모델의 성능 지표 결과를 통하여 모델이 새로운 데이터에 대해 일반화가 잘 되었음을 확인하였다. 테스트 손실과 테스트 정확도는 각각 0.2905, 91.89%였으며, 정밀도는 0.9135, 재현율은 0.9187, F1-점수는 0.9151로, 모델이 대부분의 테스트 데이터를 올바르게 분류하였다.

결론적으로 낮은 테스트 손실과 높은 정확도를 보였으며, 정밀도와 재현율, F1-점수도 높아 모델이 높은 성능을 보인 것으로 나타났다.



(그림 3) 감정 클래스 간의 혼동 행렬

5. 결 론

본 논문에서는 한국인의 감정 표현 인식 모델 개발을 위하여 ResNet-18 모델 기반의 딥러닝 모델을 학습시키고 평가하였다. 7가지 감정(기쁨, 슬픔, 당황, 분노, 불안, 상처, 무표정)을 균등하게 배분하고 전처리하여 데이터 불균형을 최소화하였으며, 최적의 하이퍼파라미터를 통해 모델 학습을 진행하였다. 그 결과, 최종 모델의 테스트 정확도는 91.89%를 달성하였으며, 정밀도 0.9135, 재현율 0.9187, F1-점수 0.9151의 우수한 성능을 보였다. 본 연구의 결과를 기존 연구인 "한국인 표정 검출을 위한 딥러닝 모델 구조"와 비교해 보면, 해당 연구에서는 49,000장의 데이터로 84.37% (±0.17)의 정확도를 달성했지만, 본 연구에서는 35,000장의 데이터로 91.89%의 정확도를 달성하였다 [3]. 이는 본 연구에서 제안한 방법론이 더 적은 데이터로도 약 7.52% 높은 정확도를 달성했음을 보여준다. 또한 데이터 세트 출처인 AI 허브에서 제공된 정확도 80.85%보다도 훨씬 높은 정확도를 달성하였다. 본 연구는 한국 사회의 문화적 특성을 반영한 한국인 감정 인식 모델의 개발 가능성을 보여주는 동시에, 데이터 세트의 한계를 포함한다. 긍정적 감정인 '행복' 외에는 부정적 감정 데이터(슬픔, 당황, 분노, 불안, 상처, 무표정)이므로 다양한 감정을 다루지 못하였다. 향후 연구에서는 더 복잡하고 다양한 감정 데이터 세트들을 적용한 후속 연구가 필요하다.

ACKNOWLEDGMENT

Funding Statement: This study was supported by a National Research Foundation of Korea (NRF) (http://nrf.re.kr/eng/index) grant funded by the Korean government (RS-2023-00208278)

참 고 문 헌

[1] Susan David, "A List of Emotions" LinkedIn post, 2023.

[2] Jifeng Dai, Yi Li, Kaiming He, Jian Sun, "R-FCN: Object Detection via Region-based Fully Convolutional Networks," arXiv:1605.06409v3, 2023.

[3] Jiyoung Lee, Jiho Kim, Euna Lee, Hongchul Lee, "Deep Learning Model Structure for Korean Facial Expression Detection," Journal of KIIT, Vol. 21, No. 2, pp. 9-17, 2023.