



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2025-0093118
(43) 공개일자 2025년06월24일

- | | |
|---|---|
| <p>(51) 국제특허분류(Int. Cl.)
 $G06V\ 40/16$ (2022.01) $G06V\ 10/24$ (2022.01)
 $G06V\ 10/44$ (2022.01) $G06V\ 10/77$ (2022.01)
 $G06V\ 10/774$ (2022.01)</p> <p>(52) CPC특허분류
 $G06V\ 40/174$ (2022.01)
 $G06V\ 10/242$ (2023.08)</p> <p>(21) 출원번호 10-2024-0037223
 (22) 출원일자 2024년03월18일
 심사청구일자 2024년03월18일</p> <p>(30) 우선권주장
 1020230183464 2023년12월15일 대한민국(KR)</p> | <p>(71) 출원인
 고려대학교 산학협력단
 서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)</p> <p>(72) 발명자
 고한석
 서울특별시 용산구 이촌로88길 30, 106동 301호(이촌동, 삼성리버스위트아파트)</p> <p>구본화
 서울특별시 성북구 북악산로 831, 201동 706호(정릉동, 정릉 힐스테이트1차)</p> <p>이보경
 서울특별시 강서구 방화대로5다길 34-5, 601호(공향동, 태경럭셔리하우스)</p> <p>(74) 대리인
 송인호, 윤형근, 최관락</p> |
|---|---|

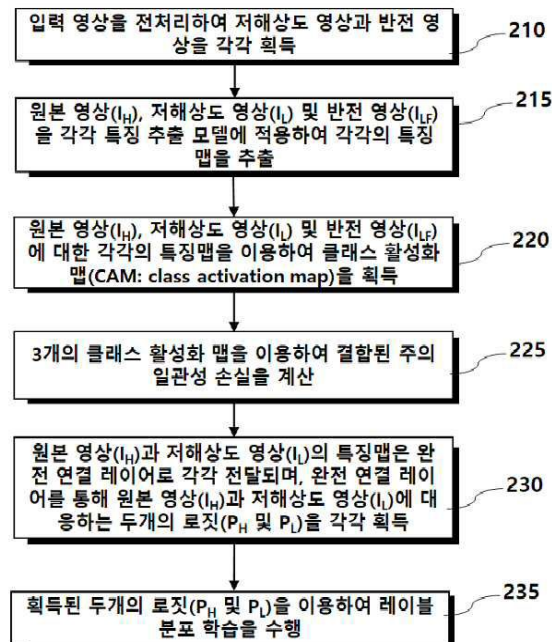
전체 청구항 수 : 총 7 항

(54) 발명의 명칭 다중 스케일 저해상도 얼굴 표정 인식 모델을 위한 학습 방법 및 그 장치

(57) 요약

다중 스케일 저해상도 얼굴 표정 인식 모델을 위한 학습 방법 및 그 장치가 개시된다. 다중 스케일 저해상도 얼굴 표정 인식 모델을 위한 학습 방법은 (a) 입력 영상을 전처리하여 저해상도 영상과 반전 영상을 각각 획득하는 단계; (b) 상기 입력 영상, 상기 저해상도 영상 및 상기 반전 영상을 딥러닝 모델에 각각 적용하여 제1 특 (뒷면에 계속)

대표도 - 도2



징맵, 제2 특징맵 및 제3 특징맵을 각각 추출하는 단계; (c) 상기 제1 특징맵, 상기 제2 특징맵 및 상기 제3 특징맵을 이용하여 제1 클래스 활성화 맵, 제2 클래스 활성화 맵 및 제3 클래스 활성화 맵을 각각 획득하는 단계; (d) 상기 제1 특징맵 및 상기 제2 특징맵을 기초로 제1 로짓과 제2 로짓을 각각 획득하는 단계; 및 (e) 상기 제1 클래스 활성화 맵, 상기 제2 클래스 활성화 맵 및 상기 제3 클래스 활성화 맵을 이용하여 결합된 주의 일관성 손실값을 계산하고, 상기 제1 로짓과 상기 제2 로짓을 이용하여 레이블 분포 학습(label distribution learning)을 수행하는 단계를 포함한다.

(52) CPC특허분류

G06V 10/454 (2023.08)

G06V 10/7715 (2023.08)

G06V 10/774 (2023.08)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711186077
과제번호	2023R1A2C2005916
부처명	과학기술정보통신부
과제관리(전문)기관명	한국연구재단
연구사업명	개인기초연구(과기정통부)
연구과제명	인간-컴퓨터 상호작용을 위한 딥러닝 기반 감성 지능형 에이전트 기술
과제수행기관명	고려대학교
연구기간	2023.03.01 ~ 2024.02.29

명세서

청구범위

청구항 1

- (a) 입력 영상을 전처리하여 저해상도 영상과 반전된 영상을 각각 생성하는 단계;
- (b) 상기 입력 영상, 상기 저해상도 영상 및 상기 반전된 영상을 딥러닝 모델에 각각 적용하여 제1 특징맵, 제2 특징맵 및 제3 특징맵을 각각 추출하는 단계;
- (c) 상기 제1 특징맵, 상기 제2 특징맵 및 상기 제3 특징맵을 이용하여 제1 클래스 활성화 맵, 제2 클래스 활성화 맵 및 제3 클래스 활성화 맵을 각각 획득하는 단계;
- (d) 상기 제1 특징맵 및 상기 제2 특징맵을 기초로 제1 로짓과 제2 로짓을 각각 획득하는 단계; 및
- (e) 상기 제1 클래스 활성화 맵, 상기 제2 클래스 활성화 맵 및 상기 제3 클래스 활성화 맵을 이용하여 결합된 주의 일관성 손실값을 계산하고, 상기 제1 로짓과 상기 제2 로짓을 이용하여 레이블 분포 학습(label distribution learning)을 수행하는 단계를 포함하는 다중 스케일 저해상도 얼굴 표정 인식 모델의 학습 방법.

청구항 2

제1 항에 있어서,

상기 저해상도 영상에 상응하는 제2 로짓의 추정 결과가 상기 입력 영상에 상응하는 추정된 클래스 및 논리적 레이블의 추정 클래스와 상이하면, 상기 저해상도 영상은 어려운 샘플로 결정되되, 상기 어려운 샘플에 대한 분류 난이도 지표는 제1 값(제로)로 설정되며,

상기 저해상도 영상이 어려운 샘플이 아닌 경우 분류 난이도 지표는 제2 값(1)로 설정되는 것을 특징으로 하는 다중 스케일 저해상도 얼굴 표정 인식 모델의 학습 방법.

청구항 3

제2 항에 있어서,

상기 분류 난이도 지표에 따라 최종 손실값이 상이하게 도출되되,

상기 최종 손실값은 상기 분류 난이도 지표를 기초로 상기 결합된 주의 일관성 손실값 및 레이블 분포 학습 손실값 중 적어도 하나를 이용하여 도출되는 것을 특징으로 하는 다중 스케일 저해상도 얼굴 표정 인식 모델의 학습 방법.

청구항 4

제1 항에 있어서,

상기 결합된 주의 일관성 손실값은 하기 수학식을 이용하여 계산되는 것을 특징으로 하는 다중 스케일 저해상도 얼굴 표정 인식 모델의 학습 방법.

$$L_{ca} = \frac{1}{K} \sum_{k=1}^K D_a(A_{Ck}, A_{Lk})$$

$$s_A = \frac{\delta_{ALF}}{\delta_{ALF} + \delta_{ALH}}$$

여기서, $A_C = ((1 - s_A) \cdot A_H + s_A \cdot \text{Flip}(A_{LF}))$ 이며, 이고,

$\delta_{A_{LH}} = \|A_L - A_H\|_1$ 이며, $\delta_{A_{LF}} = \|A_L - Flip(A_{LF})\|_1$ 이고, $Flip(\cdot)$ 는 수평 반전 함수를 나타내고, A_H 는 상기 입력 영상에 상응하는 상기 제1 클래스 활성화 맵을 나타내고, A_L 은 상기 저해상도 영상에 상응하는 상기 제2 클래스 활성화 맵을 나타내며, A_{LF} 은 상기 반전된 영상에 상응하는 상기 제3 클래스 활성화 맵을 나타내고, $\|\cdot\|_1$ 는 L1-놈 연산을 나타내며, D_a 는 두 주의 맵 사이의 거리를 계산하는 함수를 나타내고, K는 클래스 인덱스를 나타냄.

청구항 5

제3 항에 있어서,

상기 레이블 분포 학습 손실값은 하기 수학식을 이용하여 계산되는 것을 특징으로 하는 다중 스케일 저해상도 얼굴 표정 인식 모델의 학습 방법.

$$L_{lcl} = D_p(P_L, P_C)$$

여기서, $D_p(\cdot)$ 는 교차 엔트로피 손실 함수를 나타내고, $P_C = (s_P \cdot P_H + (1 - s_P) \cdot T)$ 는 유사도 s_P 를 가지고 P_H 와 T를 결합하여 생성된 저해상도에 대한 타겟 레이블 분포를 나타내며, s_P 는 $\frac{\delta_{LT}}{\delta_{LT} + \delta_{LH}}$ 로 계산되고, $\delta_{P_{LH}} = \|P_L - P_H\|_1$ 이며, $\delta_{P_{LT}} = \|P_L - T\|_1$ 이고, P_H 는 제1 로짓을 나타내고, P_L 은 제2 로짓을 나타내며, T는 원-핫 논리 레이블을 나타냄.

청구항 6

제1 항에 있어서,

상기 (a) 단계는,

상기 입력 영상을 랜덤하게 다운샘플링한 후 다시 업샘플링하여 상기 저해상도 영상을 생성하는 단계; 및

상기 저해상도 영상을 좌우 반전시켜 저해상도의 상기 반전된 영상을 생성하는 단계를 포함하는 것을 특징으로 하는 다중 스케일 저해상도 얼굴 표정 인식 모델의 학습 방법.

청구항 7

적어도 하나의 명령어를 저장하는 메모리; 및

상기 메모리에 저장된 명령어를 실행하는 프로세서를 포함하되,

상기 프로세서에 의해 실행된 명령어는 각각,

- (a) 입력 영상을 전처리하여 저해상도 영상과 반전된 영상을 각각 생성하는 단계-상기 반전된 영상은 상기 저해상도 영상을 반전한 영상임;
- (b) 상기 입력 영상, 상기 저해상도 영상 및 상기 반전된 영상을 특징 추출 모델에 각각 적용하여 제1 특징맵, 제2 특징맵 및 제3 특징맵을 각각 추출하는 단계;
- (c) 상기 제1 특징맵, 상기 제2 특징맵 및 상기 제3 특징맵을 이용하여 제1 클래스 활성화 맵, 제2 클래스 활성화 맵 및 제3 클래스 활성화 맵을 각각 획득하는 단계;
- (d) 상기 제1 특징맵 및 상기 제2 특징맵을 기초로 제1 로짓과 제2 로짓을 각각 획득하는 단계; 및
- (e) 상기 제1 클래스 활성화 맵, 상기 제2 클래스 활성화 맵 및 상기 제3 클래스 활성화 맵을 이용하여 결합된

주의 일관성 손실값을 계산하고, 상기 제1 로짓과 상기 제2 로짓을 이용하여 레이블 분포 학습(label distribution learning)을 수행하는 단계를 수행하는 것을 특징으로 하는 컴퓨팅 장치.

발명의 설명

기술 분야

[0001] 본 발명은 다중 스케일 저해상도 얼굴 표정 인식 모델을 위한 학습 방법 및 그 장치에 관한 것이다.

배경 기술

[0003] 얼굴 표정 인식은 인간-컴퓨터 상호 작용 및 다양한 컴퓨터 비전 작업에서 중요하다. 딥러닝 기반 얼굴 표정 인식 연구는 주로 데이터 셋의 노이즈 라벨 문제를 해결하는데 초점을 맞추고 있으며, 결과적으로 실제 환경에서 흔히 발생하는 저해상도(LR) 얼굴 영상을 처리할 때 성능이 제한되는 단점이 있다.

[0004] 다중 스케일 저해상도 영상은 카메라 장비 품질의 한계와 타겟과 카메라 렌즈 사이의 거리와 같은 요인에 의해 발생된다. 고해상도 영상과 비교하여 저해상도 얼굴 영상은 근육 운동으로 인한 주름과 같은 중요한 디테일이 부족하여 정확한 얼굴 표정 인식을 위한 중요한 단서가 부족하다.

[0005] 종래의 일관성 제약은 다운샘플링 증강을 사용할 때 저해상도 얼굴 표정 인식의 성능 향상에 한계가 있다. 다운샘플링 중에 얼굴 디테일이 왜곡되기 때문에 증강은 원본 영상을 손상시킬 수 있으며, 결과적으로 네트워크가 원래 목표의 감정적 특징이 효과적으로 추출되지 않는 얼굴 영상을 훈련하도록 만들 수 있다.

[0006] 도 1은 다운 샘플링으로 인해 눈물과 미세한 눈 주름이 사라지면서 왼쪽에서 볼 수 있는 것처럼 분노 얼굴과 유사한 외관을 가지게 된다. 따라서, 얼굴 표정 인식 모델의 예측은 왼쪽에서 볼 수 있는 것처럼 원래 레이블인 "슬픔"이 아닌 "분노"로 레이블 된다. 종래의 일관성 손실을 적용할 때, 이러한 어려운 샘플의 표현이 고해상도 표현과 유사하게 되어 과적합 문제가 발생한다. 이는 네트워크가 얼굴 식별과 같은 다양한 표현을 기반으로 일관성을 달성하도록 학습하기 때문에 발생한다.

[0007] 종래의 일관성 손실을 사용할 때, 상술한 어려운 샘플이 원본 샘플과 유사하게 표현되어 원래 샘플과 유사하게 취급되므로 일반화된 모델의 학습이 방해될 수 있다.

발명의 내용

해결하려는 과제

[0009] 본 발명은 다중 스케일 저해상도 얼굴 표정 인식 모델을 위한 학습 방법 및 그 장치를 제공하기 위한 것이다.

[0010] 또한, 본 발명은 다중 스케일 저해상도의 얼굴 영상의 표정 인식에 대해 높은 인식률을 보이며, 훈련 데이터에서 인식이 불가능한 샘플을 정의 및 판단하고, 인식이 불가능한 샘플을 학습에 반영하여 강인한 인식이 가능한 다중 스케일 저해상도 얼굴 표정 인식 모델을 위한 학습 방법 및 그 장치를 제공하기 위한 것이다.

[0011] 또한, 본 발명은 다중 스케일 저해상도 영상에 대한 특징맵의 선형 행동(linear behavior)을 학습할 수 있는 다중 스케일 저해상도 얼굴 표정 인식 모델을 위한 학습 방법 및 그 장치를 제공하기 위한 것이다.

[0012] 또한, 본 발명은 레이블 분포 학습(label distribution learning)을 통해 저해상도 영상이 적절한 타겟 분포(target distribution)으로 학습할 수 있는 다중 스케일 저해상도 얼굴 표정 인식 모델을 위한 학습 방법 및 그 장치를 제공하기 위한 것이다.

과제의 해결 수단

[0014] 본 발명의 일 측면에 따르면, 다중 스케일 저해상도 얼굴 표정 인식 모델을 위한 학습 방법이 제공된다.

[0015] 본 발명의 일 실시예에 따르면, (a) 입력 영상을 전처리하여 저해상도 영상과 반전된 영상을 각각 생성하는 단계; (b) 상기 입력 영상, 상기 저해상도 영상 및 상기 반전된 영상을 딥러닝 모델에 각각 적용하여 제1 특징맵, 제2 특징맵 및 제3 특징맵을 각각 추출하는 단계; (c) 상기 제1 특징맵, 상기 제2 특징맵 및 상기 제3 특징맵을 이용하여 제1 클래스 활성화 맵, 제2 클래스 활성화 맵 및 제3 클래스 활성화 맵을 각각 획득하는 단계; (d) 상기 제1 특징맵 및 상기 제2 특징맵을 기초로 제1 로짓과 제2 로짓을 각각 획득하는 단계; 및 (e) 상기 제1 클래스 활성화 맵, 상기 제2 클래스 활성화 맵 및 상기 제3 클래스 활성화 맵을 이용하여 결합된 주의 일관성 손실값을 계산하고, 상기 제1 로짓과 상기 제2 로짓을 이용하여 레이블 분포 학습(label distribution learning)을 수행하는 단계를 포함하는 다중 스케일 저해상도 얼굴 표정 인식 모델을 위한 학습 방법이 제공될 수 있다.

[0016] 상기 (a) 단계는, 상기 입력 영상을 랜덤하게 다운샘플링한 후 다시 업샘플링하여 상기 저해상도 영상을 생성하는 단계; 및 상기 저해상도 영상을 좌우 반전시켜 저해상도의 상기 반전된 영상을 생성하는 단계를 포함할 수 있다.

[0017] 상기 저해상도 영상에 상응하는 제2 로짓의 추정 결과가 상기 입력 영상에 상응하는 추정된 클래스 및 논리적 레이블의 추정 클래스와 상이하면, 상기 저해상도 영상은 어려운 샘플로 결정되되, 상기 어려운 샘플에 대한 분류 난이도 지표는 제1 값(제로)로 설정되며, 상기 저해상도 영상이 어려운 샘플이 아닌 경우 분류 난이도 지표는 제2 값(1)로 설정될 수 있다.

[0018] 상기 분류 난이도 지표에 따라 최종 손실값이 상이하게 도출되되, 상기 최종 손실값은 상기 분류 난이도 지표를 기초로 상기 결합된 주의 일관성 손실값 및 레이블 분포 학습 손실값 중 적어도 하나를 이용하여 도출될 수 있다.

[0019] 상기 결합된 주의 일관성 손실값은 하기 수학적식을 이용하여 계산되되,

$$L_{ca} = \frac{1}{K} \sum_{k=1}^K D_a(A_{Ck}, A_{Lk})$$

[0020]

[0021] 여기서, $A_C = ((1 - s_A) \cdot A_H + s_A \cdot \text{Flip}(A_{LF}))$ 이며, $s_A = \frac{\delta_{A_{LF}}}{\delta_{A_{LF}} + \delta_{A_{LH}}}$ 이고,

$\delta_{A_{LH}} = \|A_L - A_H\|_1$ 이며, $\delta_{A_{LF}} = \|A_L - \text{Flip}(A_{LF})\|_1$ 이고, $\text{Flip}(\cdot)$ 는 수평 반전 함수를 나타내고, A_H 는 상기 입력 영상에 상응하는 상기 제1 클래스 활성화 맵을 나타내고, A_L 은 상기 저해상도 영상에 상응하는 상기 제2 클래스 활성화 맵을 나타내며, A_{LF} 은 상기 반전된 영상에 상응하는 상기 제3 클래스 활성화 맵을 나타내고, $\|\cdot\|_1$ 는 L1-놈 연산을 나타내며, D_a 는 두 주의 맵 사이의 거리를 계산하는 함수를 나타내고, K는 클래스 인덱스를 나타낸다.

[0022] 상기 레이블 분포 학습 손실값은 하기 수학적식을 이용하여 계산되되,

$$L_{ldl} = D_p(P_L, P_C)$$

[0023]

[0024] 여기서, $D_p(\cdot)$ 는 교차 엔트로피 손실 함수를 나타내고, $P_C = (s_P \cdot P_H + (1 - s_P) \cdot T)$ 는 유사도 s_P 를 가지고 P_H 와 T를 결합하여 생성된 저해상도에 대한 타겟 레이블 분포를 나타내며, $s_P = \frac{\delta_{LT}}{\delta_{LT} + \delta_{LH}}$

계산되고, $\delta_{P_{LH}} = \|P_L - P_H\|_1$ 이며, $\delta_{P_{LT}} = \|P_L - T\|_1$ 이고, P_H 는 제1 로짓을 나타내고, P_L 은 제2 로짓을 나타내며, T는 원-핫 논리 레이블을 나타낸다.

[0026] 본 발명의 다른 측면에 따르면, 다중 스케일 저해상도 얼굴 표정 인식 모델의 학습 방법을 수행하는 장치가 제공된다.

[0027] 본 발명의 일 실시예에 따르면, 적어도 하나의 명령어를 저장하는 메모리; 및 상기 메모리에 저장된 명령어를

실행하는 프로세서를 포함하되, 상기 프로세서에 의해 실행된 명령어는 각각, (a) 입력 영상을 전처리하여 저해상도 영상과 반전된 영상을 각각 생성하는 단계; (b) 상기 입력 영상, 상기 저해상도 영상 및 상기 반전된 영상을 특징 추출 모델에 각각 적용하여 제1 특징맵, 제2 특징맵 및 제3 특징맵을 각각 추출하는 단계; (c) 상기 제1 특징맵, 상기 제2 특징맵 및 상기 제3 특징맵을 이용하여 제1 클래스 활성화 맵, 제2 클래스 활성화 맵 및 제3 클래스 활성화 맵을 각각 획득하는 단계; (d) 상기 제1 특징맵 및 상기 제2 특징맵을 기초로 제1 로짓과 제2 로짓을 각각 획득하는 단계; 및 (e) 상기 제1 클래스 활성화 맵, 상기 제2 클래스 활성화 맵 및 상기 제3 클래스 활성화 맵을 이용하여 결합된 주의 일관성 손실값을 계산하고, 상기 제1 로짓과 상기 제2 로짓을 이용하여 레이블 분포 학습(label distribution learning)을 수행하는 단계를 수행하는 것을 특징으로 하는 컴퓨팅 장치가 제공될 수 있다.

발명의 효과

- [0029] 본 발명의 일 실시예에 따른 다중 스케일 저해상도 얼굴 표정 인식 모델을 위한 학습 방법 및 그 장치를 제공함으로써, 다중 스케일 저해상도의 얼굴 영상의 표정 인식에 대해 높은 인식률을 보이며, 훈련 데이터에서 인식이 불가능한 샘플을 정의 및 판단하고, 인식이 불가능한 샘플을 학습에 반영하여 강인한 얼굴 표정 인식이 가능하다.
- [0030] 또한, 본 발명은 다중 스케일 저해상도 영상에 대한 특징맵의 선형 행동(linear behavior)를 학습할 수 있다.
- [0031] 또한, 본 발명은 레이블 분포 학습(label distribution learning)을 통해 저해상도 영상이 적절한 타겟 분포(target distribution)으로 학습할 수 있다.

도면의 간단한 설명

- [0033] 도 1은 종래 기술을 설명하기 위해 도시한 도면.
- 도 2는 본 발명의 일 실시예에 따른 다중 스케일 저해상도 얼굴 표정 인식 모델의 학습 방법을 나타낸 순서도.
- 도 3은 본 발명의 일 실시예에 따른 다중 스케일 저해상도 얼굴 표정 인식 모델의 전체 프레임워크를 도시한 도면.
- 도 4는 종래와 본 발명의 일 실시예에 따른 다운 샘플링한 테스트 데이터 셋에 대한 얼굴 표정 인식률을 비교한 결과를 도시한 도면.
- 도 5는 원본 얼굴 영상에 대한 종래 기술과 본 발명의 일 실시예에 따른 얼굴 표정 인식 결과를 비교한 도면.
- 도 6은 종래와 본 발명의 일 실시예에 따른 클래스 활성화 맵의 시각화를 비교한 도면.
- 도 7은 종래와 본 발명의 일 실시예에 따른 특징 맵에 대한 t-SNE의 시각화 결과를 비교한 도면.
- 도 8은 본 발명의 일 실시예에 따른 컴퓨팅 장치의 내부 구성을 개략적으로 도시한 블록도.

발명을 실시하기 위한 구체적인 내용

- [0034] 본 명세서에서 사용되는 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 명세서에서, "구성된다" 또는 "포함한다" 등의 용어는 명세서상에 기재된 여러 구성 요소들, 또는 여러 단계들을 반드시 모두 포함하는 것으로 해석되지 않아야 하며, 그 중 일부 구성 요소들 또는 일부 단계들은 포함되지 않을 수도 있고, 또는 추가적인 구성 요소 또는 단계들을 더 포함할 수 있는 것으로 해석되어야 한다. 또한, 명세서에 기재된 "...부", "모듈" 등의 용어는 적어도 하나의 기능이나 동작을 처리하는 단위를 의미하며, 이는 하드웨어 또는 소프트웨어로 구현되거나 하드웨어와 소프트웨어의 결합으로 구현될 수 있다.
- [0035] 이하, 첨부된 도면들을 참조하여 본 발명의 실시예를 상세히 설명한다.

- [0037] 도 2는 본 발명의 일 실시예에 따른 다중 스케일 저해상도 얼굴 표정 인식 모델의 학습 방법을 나타낸 순서도이고, 도 3은 본 발명의 일 실시예에 따른 얼굴 표정 인식 모델의 전체 프레임워크를 도시한 도면이다.

- [0038] 단계 210에서 컴퓨팅 장치(200)는 입력 영상을 전처리하여 저해상도 영상과 반전 영상을 각각 획득한다.
- [0039] 이에 대해 보다 상세히 설명하기로 한다. 도 3에는 본 발명의 일 실시예에 따른 얼굴 표정 인식 모델의 전체 프레임워크가 도시되어 있다.
- [0040] 우선, 컴퓨팅 장치(200)는 224 x 224 해상도의 원본 얼굴 영상을 입력 영상(편의상 I_H 라 칭하기로 함) 입력받은 후 랜덤하게 다운샘플링한 후 이를 다시 원본 해상도로 업샘플링하여 저해상도 영상(I_L)을 생성한다.
- [0041] 이어, 컴퓨팅 장치(200)는 저해상도 영상(I_L)을 좌우 반전시켜 저해상도 반전 영상(I_{LF})을 생성할 수 있다.
- [0042] 저해상도 영상(I_L)과 반전 영상(I_{LF})는 특징 추출 모델에 입력되기 전에 원본 영상(I_H)의 차원과 일치하도록 양선형 보간을 통해 업샘플링될 수 있다.
- [0043] 단계 215에서 컴퓨팅 장치(200)는 원본 영상(I_H), 저해상도 영상(I_L) 및 반전 영상(I_{LF})을 각각 특징 추출 모델에 적용하여 각각의 특징맵을 추출한다.
- [0044] 예를 들어, 특징 추출 모델은 컨볼루션 기반 네트워크일 수 있다. 본 발명의 일 실시예에서는 특징 추출 모델이 CNN인 것을 가정하여 이를 중심으로 설명하나, 이외에도 컨볼루션 연산을 기반으로 영상에 대한 특징맵을 추출하는 네트워크인 경우 제한 없이 적용될 수 있다.
- [0045] 단계 220에서 컴퓨팅 장치(200)는 원본 영상(I_H), 저해상도 영상(I_L) 및 반전 영상(I_{LF})에 대한 각각의 특징맵을 이용하여 클래스 활성화 맵(CAM: class activation map)을 획득한다.
- [0046] 클래스 활성화 맵은 CNN이 예측한 클래스 점수를 시각화하고 CNN이 식별한 영역을 지역화하는데 효과적이다. 이러한 클래스 활성화 맵은 가짜 라벨을 생성하고 훈련하는데 사용될 수 있다.
- [0047] 클래스 활성화 맵은 수학적 식 1과 같이 계산될 수 있다.

수학적 식 1

$$A_k = \sum_{c=1}^C W_{k,c} F_c.$$

- [0048]
- [0049] 여기서, C는 마지막 컨볼루션 레이어의 특징 맵($F \in \mathbb{R}^{C \times H \times W}$)의 채널 수를 나타내고, $W \in \mathbb{R}^{K \times C}$ 는 완전 연결 레이어의 가중치를 나타낸다.
- [0050] 예를 들어, 컴퓨팅 장치(200)는 특징 추출 모델(예를 들어, CNN)의 최종 컨볼루션 레이어와 완전 연결 레이어 사이의 연결 가중치를 클래스 활성화 가중치로서 추출하고, 원본 영상(I_H)이 특징 추출 모델을 통과하여 생성된 특징맵과 추출된 클래스 활성화 가중치를 결합하여 원본 영상(I_H)에 대한 클래스 활성화 맵(A_H)을 생성할 수 있다. 이와 같은 방식으로 컴퓨팅 장치(200)는 저해상도 영상(I_L) 및 반전 영상(I_{LF}) 각각의 특징맵을 이용하여 클래스 활성화 맵(A_L, A_{LF})을 각각 생성할 수 있다.
- [0051] 단계 225에서 컴퓨팅 장치(200)는 3개의 클래스 활성화 맵을 이용하여 결합된 주의 일관성 손실(combined attention consistency loss)을 계산한다.
- [0052] 이에 대해 보다 상세히 설명하기로 한다.
- [0053] 종래 기술들은 불일치 문제를 해결하기 위해, A_H 와 A_L 사이의 주의 일관성 손실 함수를 사용하였으며, 이는 수학적 식 2를 이용하여 계산될 수 있다.

수학식 2

$$L_a = \frac{1}{K} \sum_{k=1}^K D_a(A_{H_k}, A_{L_k})$$

[0054]

[0055] 여기서, D_a 는 L1 또는 L2 놈과 같이 두 어텐션 맵(즉, 클래스 활성화 맵) 사이의 거리를 계산하기 위한 함수이다.

[0056] 마찬가지로, 증강된 입력에 대한 예측을 위해 일관성 손실 함수가 사용될 수 있다. 이를 수학식으로 나타내면 수학식 3과 같다.

수학식 3

$$L_p = D_p(P_L, P_T)$$

[0057]

[0058] 여기서, D_p 는 두 카테고리 분포 사이의 거리를 포착할 수 있는 손실 함수를 나타내고, P_T 는 예를 들어, 원-핫 논리적 레이블 및 원본 입력(P_H)에 대한 예측과 같은 타겟 확률 분포를 나타낸다.

[0059] 수학식 2와 같은 주의 일관성 손실 함수는 과적합 문제를 야기할 수 있다. 이는 모델이 주의 맵 관점에서 다중 해상도 샘플을 동등하게 처리하도록 조건화되기 때문에 발생한다.

[0060] 균일한 처리가 어려운 샘플에도 확장되어 모델이 얼굴 영상 식별 특징 맵을 학습하는 방향으로 포커스가 이동되어, 구별적인 얼굴 표정 특징을 학습하는 능력이 저하되는 문제점이 있다.

[0061] 따라서, 본 발명의 일 실시예에서는 결합된 주의 일관성 손실 함수를 도출한다. 결합된 주의 일관성 손실 함수의 목적은 다중 스케일 저해상도 영상에서 주의 맵을 적절한 타겟 주의 맵과 일치시키는 것이다. A_{LF} 와 A_H 이 A_L 에 대한 유용한 레퍼런스로 이용될 수 있다.

[0062] 이를 위해, 컴퓨팅 장치(200)는 A_L 와 A_H 사이의 L1 거리($\delta_{A_{LH}} = \|A_L - A_H\|_1$)를 계산하고, A_L 와 A_{LF} 사이의 L1 거리($\delta_{A_{LF}} = \|A_L - \text{Flip}(A_{LF})\|_1$)를 각각 계산한다. 여기서, $\text{Flip}(\cdot)$ 는 수평 반전 연산을 나타낸다.

$$s_A = \frac{\delta_{A_{LF}}}{\delta_{A_{LF}} + \delta_{A_{LH}}}$$

[0063] A_L 와 A_H 사이의 유사도는 s_A 와 같이 계산될 수 있으며, A_L 와 A_{LF} 사이의 유사도는 $1 - s_A$ 와 같이 계산될 수 있다.

[0064] 결합된 주의 일관성 손실 함수의 값은 수학식 4와 같이 계산될 수 있다.

수학식 4

$$L_{ca} = \frac{1}{K} \sum_{k=1}^K D_a(A_{C_k}, A_{L_k})$$

[0065]

[0066] 여기서, $A_C = ((1 - s_A) \cdot A_H + s_A \cdot \text{Flip}(A_{LF}))$ 는 결합된 주의 맵을 나타내고, 결합된 주의 일관성은 A_L 와 결합된 주의 맵 간의 차이를 최소화할 수 있다. $D_a(\cdot)$ 는 L1 놈을 나타내고, K 는 클래스 인덱스를 나타낸다. 큰

s_A 는 클래스 활성화 맵이 초기 훈련 단계에서 불균형임을 의미하며, 이는 주의 맵(A_L 와 A_F)이 동일한 해상도 영상임에도 불구하고 서로 다르기 때문이다.

[0067] 따라서, A_F 이 결합된 주의 맵의 큰 부분을 차지하면 A_L 와 A_F 이 서로 가까워진다. 반면, 작은 s_A 는 주의 맵이 균형을 이루고 있음을 나타낸다. 결합된 주의 일관성 손실 함수의 값(L_{ca})은 $LR(A_L)$ 로부터의 판별적 표현의 지식을 주의 맵으로 전달하는 데 중점을 둔다.

[0068] 종래의 주의 일관성 손실과 달리, 본 발명의 일 실시예에 따른 방법은 레이블 분포 학습의 관점에서 적절한 타겟 주의 맵을 제공할 수 있다. 이는 s_A 의 균형 잡힌 성격에 의해 달성될 수 있으며, 단일 측면으로 편향되지 않는다.

[0069] 결과적으로, 본 발명의 일 실시예에 따르면 다중 스케일 저해상도 영상에서 유래된 주의 맵은 왜곡 정도에 따라 고해상도 주의 맵으로부터 거리를 유지할 수 있다. 결합된 주의 일관성 손실 함수는 모델이 다중 해상도 저해상도 영상에서 특징맵에 대해 샘플 증강 관점에서 선형적인 동작을 학습하도록 할 수 있다.

[0070] 단계 230에서 컴퓨팅 장치(200)는 원본 영상(I_H)과 저해상도 영상(I_L)의 특징맵은 완전 연결 레이어로 각각 전달되며, 완전 연결 레이어를 통해 원본 영상(I_H)과 저해상도 영상(I_L)에 대응하는 두개의 로짓(P_H 및 P_L)을 각각 획득한다.

[0071] 단계 235에서 컴퓨팅 장치(200)는 획득된 두개의 로짓(P_H 및 P_L)을 이용하여 레이블 분포 학습을 수행한다.

[0072] 이에 대해 보다 상세히 설명하기로 한다.

[0073] 레이블 분포 학습은 모호한 샘플, 저품질 영상, 일관되지 않은 주석 및 잘못된 주석으로 인한 잡음 레이블 문제를 위한 효과적인 솔루션이다. 많은 얼굴 표현 인식 모델에 통합되어 있기에, 저해상도 샘플은 모호하고 품질이 낮은 것으로 간주될 수 있다.

[0074] 본 발명의 일 실시예에 따르면, 레이블 분포 학습을 사용하여 어려운 샘플 범주에 속하지 않은 쉬운 샘플을 위한 타겟 분포를 결정할 수 있다.

[0075] 컴퓨팅 장치(200)는 두개의 로짓을 이용하여 L1 거리를 도출할 수 있다. 즉, $\delta_{P_{LH}} = \|P_L - P_H\|_1$ 와 $\delta_{P_{LT}} = \|P_L - T\|_1$ 와 같이 두 개의 로짓을 이용하여 L1 거리를 각각 도출할 수 있다. 여기서, T는 원-핫 논리 레이블을 나타낸다.

[0076] 또한, 컴퓨팅 장치(200)는 두개의 로짓에 대한 유사도 s_P 를 $\frac{\delta_{LT}}{\delta_{LT} + \delta_{LH}}$ 와 같이 계산할 수 있다.

[0077] 컴퓨팅 장치(200)는 레이블 분포 학습 손실 함수를 수학적식 5와 같이 계산할 수 있다.

수학적식 5

[0078]
$$L_{lcl} = D_p(P_L, P_C)$$

[0079] 여기서, $D_p(\cdot)$ 는 교차 엔트로피 손실 함수를 나타내고, $P_C = (s_P \cdot P_H + (1 - s_P) \cdot T)$ 는 유사도 s_P 를 가지고 P_H 와 T를 결합하여 생성된 저해상도에 대한 타겟 레이블 분포를 나타낸다.

[0080] 약간 저하가 있는 영상을 처리할 때, 레이블 분포 학습 손실 함수 L_{lcl} 의 최소화는 P_L 을 P_H 와 T 사이에서 공유되는 분포에 더 가까운 분포로 수렴하도록 유도할 수 있다.

[0081] 차별적인 특징맵을 학습하기 위해, 본 발명의 일 실시예에 따르면 고해상도 영상에 대한 논리적 레이블을 사용

하는 교차-엔트로피 손실 함수를 사용하며, $L_{ce} = D_p(P_H, T)$ 와 같이 정의될 수 있다.

L_{ca} 만 사용할 때, A_L 을 표현하는데 부족한 점을 보상하기 위해, 악성 학습을 활용하여 A_H 와 구별할 수 없는 A_L 을 생성할 수 있다.

판별자 $D(\cdot)$ 에 대한 적대적 손실 함수는 $L_{adv_d} = -\log D(A_H) - \log(1 - D(A_L))$ 와 같이 정의될 수 있고, 생성자에 대한 적대적 손실 함수는 수학식 6과 같이 나타낼 수 있다.

수학식 6

$$L_{adv_g} = \log(1 - D(A_L))$$

어려운 샘플을 구별하기 위해, 컴퓨팅 장치(200)는 두개의 로짓(P_H 및 P_L) 및 논리적 레이블 T 간의 표현 클래스 결과를 비교한다.

예를 들어, P_L 의 최대 인덱스(추정 결과)가 P_H 및 T 의 표현 클래스 결과와 다르면, 저해상도 영상은 어려운 샘플로 간주될 수 있다. 이후, 저해상도 영상이 어려운 샘플로 식별될 때, 분류 난이도 지표는 $\tau = 0$ 로 정의될 수 있다. 분류 난이도 지표는 수학식 7과 같이 도출될 수 있다.

수학식 7

$$\tau = \begin{cases} 0 & \text{if } C(P_L) \neq C(P_H) \text{ and } C(P_L) \neq C(T), \\ 1 & \text{otherwise,} \end{cases}$$

여기서, $C(\cdot) = \operatorname{argmax}_k(\cdot)$ 는 추정된 클래스가 추출되는 것을 나타낸다.

최종 손실 함수는 수학식 8과 같이 도출될 수 있다.

수학식 8

$$L_T = \begin{cases} L_{ce} + \lambda_{ca} L_{ca} + \lambda_{ldl} L_{ldl} + \lambda_{adv} L_{adv_g}, & \text{if } \tau = 1 \\ L_{ce} + \lambda_{ca} L_{ca}, & \text{otherwise} \end{cases}$$

여기서, λ_{ca} , λ_{ldl} 및 λ_{adv} 는 하이퍼 파라미터를 나타낸다.

본 발명의 일 실시예에 따르면, 모델이 어려운 샘플로 과적합되는 것을 방지하기 위해 분류 난이도 지표를 통합할 수 있다. L_{ca} 는 다중 스케일 저해상도 영상에서 각 표현 사이의 연결을 내재적으로 설정할 수 있다.

도 4는 종래와 본 발명의 일 실시예에 따른 다운 샘플링한 테스트 데이터 셋에 대한 얼굴 표정 인식률을 비교한 결과를 도시한 도면이다.

도 4에 도시된 바와 같이, 본 발명의 일 실시예에 따른 방법(HSAC)이 종래 기술과 비교하여 다운 샘플링 팩터에서 가장 좋은 성능을 보이는 것을 알 수 있다.

도 5는 원본 얼굴 영상에 대한 종래 기술과 본 발명의 일 실시예에 따른 얼굴 표정 인식 결과를 비교한 도면이

다. 도 5에서 보여지는 바와 같이, 본 발명이 종래 기술과 비교하여 얼굴 표정 인식 정확도를 뛰어넘은 것을 알 수 있다.

[0096] 도 6은 종래와 본 발명의 일 실시예에 따른 클래스 활성화 맵의 시각화를 비교한 도면이다. 도 6에서 보여지는 바와 같이, 종래의 일관성 학습을 적용한 기술들과 비교하여 본 발명의 일 실시예에 따른 다중 스케일 저해상도 얼굴 표정 인식 모델의 학습 방법이 감정을 보여주는 영역에만 집중하고 있는 것을 알 수 있다.

[0097] 도 7은 종래와 본 발명의 일 실시예에 따른 특징 맵에 대한 t-SNE의 시각화 결과를 비교한 도면이다. 종래의 일관성 학습(a)는 어려운 샘플(검정색 점)이 클래스 분포의 중심에 포함되는 것을 알 수 있다. 반면, 도 7의 (b)에서 보여지는 바와 같이 본 발명의 일 실시예에 따른 다중 스케일 저해상도 얼굴 표정 인식 모델의 학습 방법은 어려운 샘플의 분포가 클래스 분포의 외곽에 위치하는 것을 알 수 있다.

[0099] 도 8은 본 발명의 일 실시예에 따른 다중 스케일 저해상도 얼굴 표정 인식 모델의 학습 방법을 수행하는 컴퓨팅 장치의 내부 구성을 개략적으로 도시한 블록도이다.

[0100] 도 8을 참조하면, 본 발명의 일 실시예에 따른 컴퓨팅 장치(200)는 메모리(810) 및 프로세서(820)를 포함하여 구성된다.

[0101] 메모리(810)는 다중 스케일 저해상도 얼굴 표정 인식 모델의 학습 방법을 수행하기 위한 적어도 하나의 명령어를 저장한다.

[0102] 프로세서(820)를 메모리(810)에 저장된 명령어를 실행한다. 프로세서(820)에 의해 실행된 명령어는 입력 영상을 전처리하여 저해상도 영상과 반전된 영상을 각각 생성하고, 입력 영상, 저해상도 영상 및 반전된 영상을 특징 추출 모델에 각각 적용하여 제1 특징맵, 제2 특징맵 및 제3 특징맵을 각각 추출하며, 제1 특징맵, 상기 제2 특징맵 및 상기 제3 특징맵을 이용하여 제1 클래스 활성화 맵, 제2 클래스 활성화 맵 및 제3 클래스 활성화 맵을 각각 획득하고, 제1 특징맵 및 제2 특징맵을 기초로 제1 로짓과 제2 로짓을 각각 획득하며, 기 제1 클래스 활성화 맵, 상기 제2 클래스 활성화 맵 및 상기 제3 클래스 활성화 맵을 이용하여 결합된 주의 일관성 손실값을 계산하고, 제1 로짓과 상기 제2 로짓을 이용하여 레이블 분포 학습(label distribution learning)을 수행하는 일련의 과정을 수행할 수 있다.

[0103] 이에 대한 상세한 설명은 도 2 내지 도 3을 참조하여 설명한 바와 동일하므로 중복되는 설명은 생략하기로 한다.

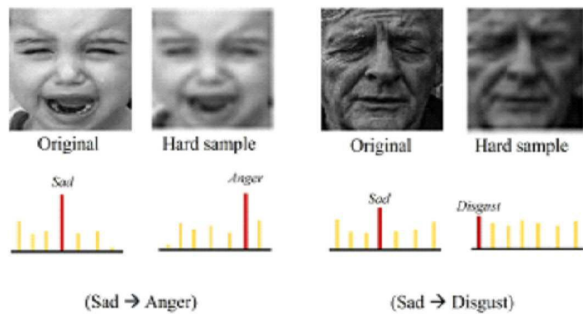
[0105] 본 발명의 실시 예에 따른 장치 및 방법은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 컴퓨터 판독 가능 매체에 기록되는 프로그램 명령은 본 발명을 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 분야 통상의 기술자에게 공지되어 사용 가능한 것일 수도 있다. 컴퓨터 판독 가능 기록 매체의 예에는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기매체(magnetic media), CD-ROM, DVD와 같은 광기록 매체(optical media), 플롭티컬 디스크(floptical disk)와 같은 자기-광 매체(magneto-optical media) 및 롬(ROM), 램(RAM), 플래시 메모리 등과 같은 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함된다. 프로그램 명령의 예에는 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드를 포함한다.

[0106] 상술한 하드웨어 장치는 본 발명의 동작을 수행하기 위해 하나 이상의 소프트웨어 모듈로서 작동하도록 구성될 수 있으며, 그 역도 마찬가지이다.

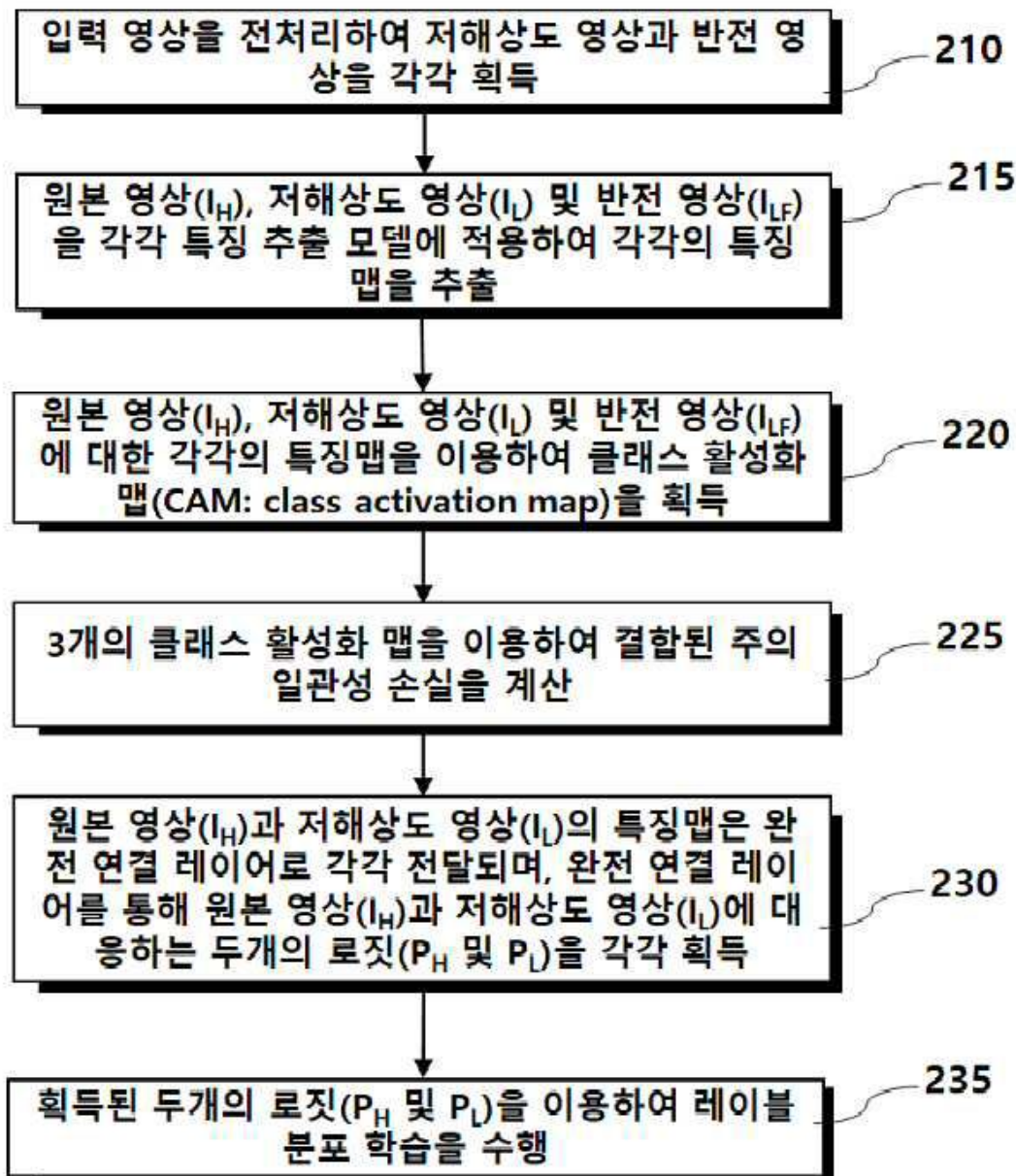
[0107] 이제까지 본 발명에 대하여 그 실시 예들을 중심으로 살펴보았다. 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자는 본 발명이 본 발명의 본질적인 특성에서 벗어나지 않는 범위에서 변형된 형태로 구현될 수 있음을 이해할 수 있을 것이다. 그러므로 개시된 실시 예들은 한정적인 관점이 아니라 설명적인 관점에서 고려되어야 한다. 본 발명의 범위는 전술한 설명이 아니라 특허청구범위에 나타나 있으며, 그와 동등한 범위 내에 있는 모든 차이점은 본 발명에 포함된 것으로 해석되어야 할 것이다.

도면

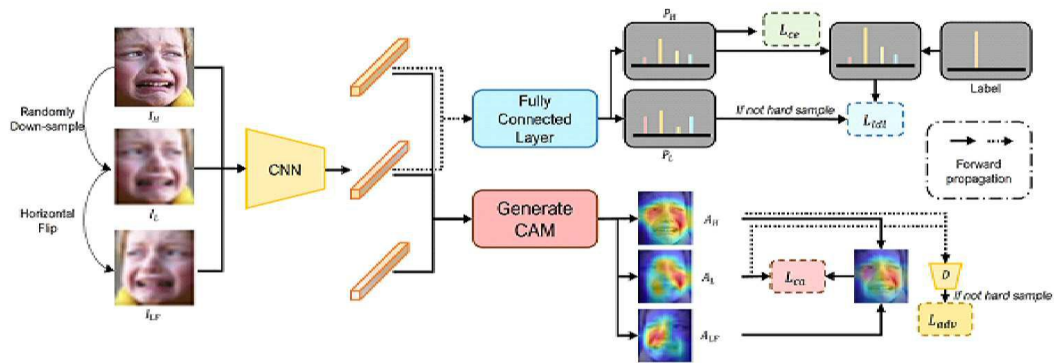
도면1



도면2



도면3



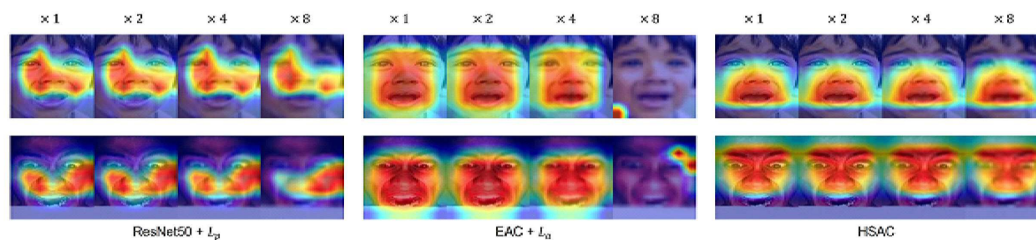
도면4

Dataset	Down-sampling Factors	RN50	RN50 + L_p	RN50 + RCAN	FSR-FER	EAC	EAC + L_a	HSAC
RAF-DB	$\times 2$ (%)	88.75	88.75	86.44	84.02	90.06	89.90	90.58
	$\times 4$ (%)	87.03	88.10	86.54	80.02	89.47	88.55	90.03
	$\times 8$ (%)	76.96	83.83	76.11	65.97	81.94	84.62	86.25
	Avg (%)	84.24	86.89	83.03	76.67	87.15	87.69	88.95
AffectNet-7	$\times 2$ (%)	62.80	63.09	59.03	-	65.20	65.26	65.43
	$\times 4$ (%)	61.26	62.77	58.63	-	64.37	64.31	65.17
	$\times 8$ (%)	49.89	57.83	42.26	-	61.31	63.03	63.40
	Avg (%)	57.98	61.23	53.30	-	63.62	64.20	64.66
SFEW2.0	$\times 2$ (%)	49.01	46.78	42.57	55.14	56.19	53.47	59.16
	$\times 4$ (%)	49.01	47.28	43.81	49.64	55.94	54.21	58.91
	$\times 8$ (%)	45.79	46.53	38.12	40.00	53.47	53.96	55.20
	Avg (%)	47.93	46.86	41.50	48.26	55.20	53.88	57.75

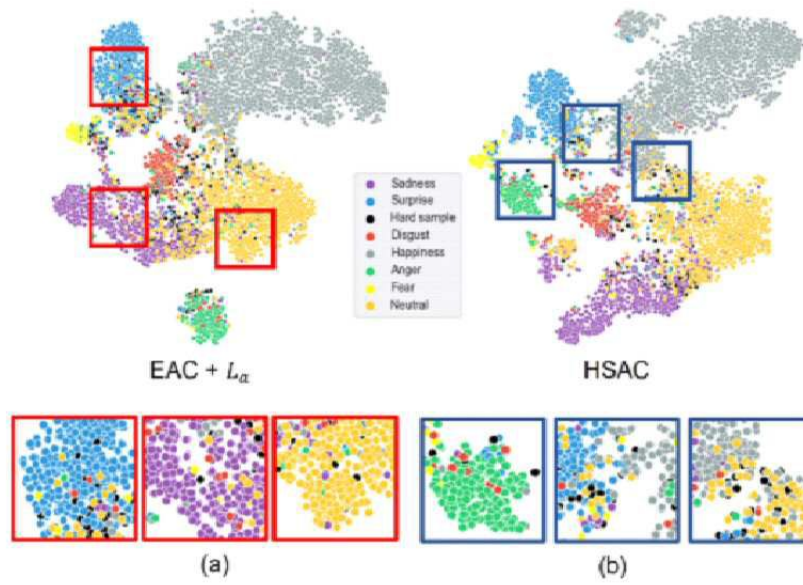
도면5

Dataset	DACL	MViT	SCN	EfficientFace	DMUE	RUL	EAC	HSAC
RAF-DB	87.78%	88.62%	88.14%	88.36%	89.42%	88.98%	90.35%	90.97%
AffectNet-7	65.20%	64.57%	-	63.70%	63.11%	-	65.23%	65.29%
SFEW2.0	-	-	-	-	58.34%	-	56.68%	58.42%

도면6



도면7



도면8

200

