Computer Methods and Programs in Biomedicine 215 (2022) 106621

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

journal homepage: www.elsevier.com/locate/cmpb

심층 학습을 기반으로 한 얼굴 표정 인식

후일궤Ge, Zhiyu Zhu*, Yuewei Dai*, Biao Wang, Xuedon g 우

Schoolof Electronic Information, Jiangsu University of Science and Technology, Zhenjiang 212003, China

t i c l e i n f o

Articlehistory: 2021년 9월 17일
수신 2021년 12월 22일 수정 202
2년 1월 3일 승인

*Keywords:*3D 얼굴 깊이 정보 딥
러닝 얼굴 표정 인식 타겟 탐지 합
성 신경망

b s t r a c t

Backgroundand objective: 얼굴 표정 인식 기술은 우리의 일상 생활에서 점점 더 중요한 역할을 할 것입니다. 자율 주행, 가상 현실 및 우리 삶에 통합된 모든 종류의 로봇은 얼굴 표정 인식 기술의 개발에 달려 있습니다. 컴퓨터 비전 분야의 많은 작업은 딥 러닝 기술과 합성곱 신경망을 기반으로 합니다. 이 논문은 생성된 대책 네트워크를 기반으로 하는 가려진 표정 인식 모델을 제안합니다. 이 모델은 가려진 얼굴 이미지 복원과 얼굴 인식의 두 모듈로 나뉩니다.

*Methods:*첫째, 본 논문에서는 최근 10년간의 심층적 얼굴 표정 인식 방법의 연구 현황과 관련 얼굴 표정 데이터베이스의 개발을 요약한다. 그런 다음 심층 학습을 기반으로 한 현재 얼굴 표정 인식 방법은 정적 얼굴 표정 인식과 동적 얼굴 표정 인식의 두 가지 범주로 나뉜다. 두 방법을 각각 소개하고 요약한다. 이 분야의 고급 심층적 표정 인식 알고리즘을 목표로 하여 일반적인 표정 데이터베이스에서 이러한 알고리즘의 성능을 비교하고 이러한 알고리즘의 강점과 약점을 자세히 분석한다.

Discussionand results: 얼굴 표정 인식 작업이 점차 통제된 실험실 환경에서 어려운 실제 환경으로 옮겨감에 따라, 딥 러닝 기술의 급속한 발전으로 딥 신경망은 차별적인 특징을 학습할 수 있으며, 점차 자동 얼굴 표정 인식 작업에 적용되고 있습니다. 현재의 딥 얼굴 표정 인식 시스템은 다음 두 가지 문제를 해결하기 위해 노력하고 있습니다. (1) 충분한 학습 데이터 부족으로 인한 과적합 (2) 실제 환경에서 표정과 관련이 없는 다른 변수는 간섭 문제를 일으킴

*Conclusion:*알고리즘 관점에서 볼 때 얼굴 동작 단위 모델, 쾌감 각성 차원 모델과 같은 다른 표정 모델과 오디오 모드, 3D 얼굴 깊이 정보, 인체 생리 정보 등의 다른 멀티모달 모델을 결합하면 표정 인식을 보다 실용적으로 만들 수 있습니다.

1. 서론

이미지 인식은 컴퓨터를 사용하여 이미지를 처리, 분석 및 이해하여 다양한 패턴의 대상 및 객체를 식별하는 기술입니다. 컴퓨터 비전 분야의 주요 연구 방향이며 이미지를 기반으로 한 지능형 데이터 수집 및 처리에서 중요한 역할을 합니다. 이미지 인식 기술은 특정 대상 객체(예: 손으로 쓴 글자, 제품 또는 얼굴)의 탐지 및 인식, 이미지 분류 및 표시, 주관적인 이미지 품질 평가를 효율적으로 완료할 수 있습니다. 현재 이미지 인식 기술은 광범위한 상업적 시장과 이미지 검색, 상품과 같은 인터넷 응용 제품에서 낙관적인 응용 전망을 가지고 있습니다.

영어: 성 추천, 사용자 행동 분석 및 얼굴 인식.게다가 UAV, 자율 주행 및 지능형 로봇과 같은 첨단 산업과 지질학, 의학 및 생물학과 같은 많은 분야에서 장기적인 개발 잠재력이 있습니다.초기 이미지 인식 시스템은 주로 방향성 기울기 히스토그램 [1]과 규모 불변 특징 변환 [2]을 사용한 다음 추출된 특징을 분류기에 입력하여 분류 및 인식했습니다.이러한 기능은 기본적으로 수동으로 설계됩니다.다른 인식 문제의 경우 추출된 특징은 시스템 성능에 직접적인 영향을 미칩니다.따라서 연구자는 시스템 성능을 개선하기 위해 더 나은 적응적 특징을 설계하기 위해 미해결 문제 영역을 추가로 연구해야 합니다.이 기간의 이미지 인식 시스템은 종종 특정 인식 작업을 위한 것이고 데이터 규모가 작고 일반화 능력이 부족하여 일상적인 응용 프로그램에서 이상적인 인식 효과를 달성하기 쉽지 않습니다.

* Corresponding authors.
E-mail addresses: zhuzy@just.edu.cn (Z. Zhu), dywjust@163.com (Y. Dai).

https://doi.org/10.1016/j.cmpb.2022.106621
0169-2607/

딥 러닝은 인공 신경망의 데이터를 기반으로 하는 표현 학습의 한 분야입니다.딥 러닝에는 지도 학습, 반지도 학습 및 비지도 학습이 포함됩니다. 딥 러닝에서 딥 신경망, 딥 신념 네트워크 및 순환 신경망은 음성 인식, 컴퓨터 비전, 오디오 인식, 무인 운전 및 자연어 처리에 널리 사용되었습니다 .리나 데크터는 1986년에 처음으로 딥 러닝을 도입했습니다.또한 이고르 아이젠버그는 2000년에 인공 신경망의 개념을 도입했습니다.사실, 알렉세이 이바크넨코와 라파는 1965년 초에 지도 피드포워드 학습 네트워크를 제안했습니다.1986년에 제프리 힌튼은 다층 퍼셉트론(MLS)의 역전파 알고리즘을 제안하고 비선형 변환을 위해 SIGMOD 활성화 함수를 사용하여 비선형 학습 및 분류 문제를 해결했습니다.2006년에 제프리 힌튼 등은 사전 학습 가중치는 모델을 초기화하고 지도 학습에 따라 모델을 미세 조정하는 데 사용되어야 한다고 지적했습니다. 딥 러닝의 추가 개발과 함께 1998년 lenet [3]의 제안은 합성곱 신경망(CNN)의 등장을 알렸습니다.

그러나 당시에는 하드웨어가 뒤떨어져 있었기 때문에 합성곱 신경망은 다른 기계 학습 방법(예: SVM[4])에 비해 유리하지 못했다. 컴퓨팅 장치의 발전으로 Hinton 등이 제안한 Alex net은 컴퓨터 비전 경연 대회 ILSVRC 2012에서 상당한 성과를 거두었다. 합성곱 신경망은 최근 수십 년 동안 급속한 발전을 이루었다.

2. 관련 연구

2.1.Review of face recognition technology

얼굴 인식은 접근이 쉽고 조작이 간편하며 특징이 다양하다는 특징이 있습니다. 지난 세기에 많은 학자들이 얼굴 인식을 연구했습니다. 그러나 네트워크가 미비하고 얼굴 이미지의 리소스가 제한적이며 사진 품질이 좋지 않아 많은 학자들이 주로 알고리즘 관점에서 연구했지만 인식 정확도가 낮아 인간의 눈 인식 효과와는 거리가 멉니다. 머신 러닝 기술이 점차 성숙해지면서 유전 알고리즘, 베이지안 분류기, 지원 벡터 머신과 같은 강력한 알고리즘이 많이 생겨났습니다. 이러한 알고리즘을 얼굴 인식 기술에 적용하면 얼굴 인식의 정확도가 어느 정도 향상되지만 특징 추출이 복잡하고 단일하여 인적 요인의 영향을 크게 받으므로 이러한 방법은 널리 사용되지 않습니다. 과학기술의 발전으로 많은 얼굴 이미지 리소스를 확보할 수 있게 되었고 딥 러닝의 급속한 발전은 얼굴 인식에 새로운 활력을 불어넣었고 인식 정확도가 크게 향상되었습니다.

Sun 등은 얼굴 정보를 나타내는 고급 특징 그룹을 학습하는 데 딥 러닝을 사용하는 DeepId [5] 얼굴 인식 방법을 제안했습니다. 얼굴 특징 정보를 학습한 후에는 분류기만 사용하여 분류하고 LFW 데이터에서 인식률이 더 좋습니다. Zhu 등은 개선된 DeepId 모델을 임베디드 시스템에 내장하여 좋은 인식 결과와 속도를 달성했습니다. DeepId2 [6]와 비교하여 DeepId2의 인식 정확도가 크게 향상되었습니다. 주된 이유는 특징 추출의 인식이 분류 정확도를 고려할 뿐만 아니라 인식 신호와 검증 신호로 각각 사용되는 클래스 간의 격차도 고려한 다음 두 신호에 가중치를 부여하여 공통 감독 학습을 위해 결합하기 때문입니다. DeepId2와 비교하여 DeepId2 + [7]은 폐색의 견고성을 개선합니다. Schroff 등 [8] 얼굴 인식 분야가 큰 진전을 이루었지만(DeepFace[9], DeepId2) 대규모 얼굴 검증 및 인식의 효과적인 구현은 여전히 심각한 과제에 직면해 있다고 믿습니다. 이 논문에서는 영어: 종단간 학습 이미지 코딩을 유클리드 공간으로 압축 유클리드 공간 메트릭 매핑한 다음 이 코딩을 기반으로 얼굴 인식을 수행합니다. 동시에 Facenet은 학습 프로세스에서 트리플 손실 함수를 사용하고 학습된 특징 벡터를 통해 동일한 클래스의 이미지와 다른 클래스의 이미지 간의 거리를 계산하므로 동일한 클래스 간의 거리가 감소하고 다른 클래스 간의 거리가 증가합니다. 우연히도 Wen et al.[10]은 중앙 손실 함수를 제안했으며 중앙 손실 함수와 기존 소프트맥스 손실 공동 감독 학습을 사용하면 동일한 범주 간의 거리를 효과적으로 줄이고 분류 및 인식 능력을 향상시킬 수 있습니다.

네트워크에서 학습한 특징을 더 구별하기 쉽도록 하기 위해 많은 학자들이 소프트맥스 손실을 기반으로 개선해 왔습니다.이는 훈련 효과를 단순화할 뿐만 아니라 더 나은 인식 효과를 달성합니다.예를 들어 L-소프트맥스 손실[11], A-소프트맥스 손실[12], Cosface[13]는 분류 각도 측면에서 제약 조건을 추가하고, 원래 분류 함수의 분류 간격을 변경하고, 특징에 대한 네트워크의 학습 능력을 향상시키고, 클래스 내에서 특징을 집계하고, 클래스 간에 분산시켜 매우 좋은 인식 효과를 달성합니다.Chen et al.[14]은 매우 효과적인 경량 얼굴 인식 네트워크 모델을 제안했는데, 모델 매개변수는 100만 미만이고 인식 정확도는 MobileNetV2[15]보다 높으며 속도도 크게 향상되었습니다.동시에 네트워크의 마지막 계층에서 가중 글로벌 보안 로컬 평균 풀링을 사용합니다. 이 방법은 모델의 학습 매개변수를 어느 정도 늘리지만, 공통 풀링으로 인해 정확도가 떨어지는 영향을 효과적으로 줄일 수 있습니다.

2.2.Research status of target detection at home and abroad

타겟 탐지는 컴퓨터 비전 분야에서 매우 중요한 문제입니다. 최근 몇 년 동안, 타겟 탐지의 속도와 정확도를 높이기 위해 합성 신경망을 기반으로 하는 많은 탐지기가 제안되었습니다. 그러나 타겟 탐지의 규모 변화는 여전히 모든 탐지기에 중요한 문제입니다. 기본적으로 두 가지 종류의 타겟 탐지 방법이 있습니다. 즉, 단일 단계 방법과 2단계 방법입니다.

2단계 방법은 대상 탐지 작업을 관심 영역을 추출한 다음 관심 영역에서 분류 및 회귀를 시작하는 두 단계로 나눕니다.R-CNN[16]은 에지 상자[17] 또는 선택적 검색[18]에서 생성된 영역을 사용합니다.사전 훈련된 합성곱 신경망에서 영역 기반 기능을 생성하고 분류에 SVM을 사용하는 것이 제안됩니다.Fast R-CNN[19]은 합성곱 신경망을 훈련하기 위해 종단 간 분류 및 위치 회귀 손실을 도입합니다.Fast R-CNN[20]은 지역 추천 선택적 검색이 있는 네트워크(RPN)로 대체됩니다.RPN은 후보 경계 상자(앵커 상자)를 생성하고 배경 영역을 필터링하는 데 사용된 다음 이러한 제안을 기반으로 또 다른 작은 네트워크를 분류 및 경계 상자 위치 회귀에 사용합니다. R-FCN [[30]]은 속도 빠른 RCNN의 관심 영역 풀링 대신 위치 촉각적 관심 영역 풀(PSROI)을 사용합니다. 이 방법은 검출기의 속도와 정확도를 높일 수 있습니다. 현재 가변 롤 네트워크 [[31]]에서 변형 가능 합성곱 및 변형 가능 PSROI가 제안되어 R-FCN의 정확도를 계속 높일 수 있습니다.

단일 단계 방식은 ROI 추출 프로세스를 제거하고 후보 앵커 프레임을 직접 분류하고 회귀시킵니다.YOLO[[32]]는 입력 이미지를 여러 그리드로 분리하고 이미지의 각 섹션을 찾아 분류합니다.이 개념을 사용하면 YOLO는 대상을 빠르게 감지할 수 있지만 정확도는 이상적이지 않습니다.YOLOv2[[33]]는 YOLO 네트워크를 개선하고 분류 및 감지의 공동 학습 전략을 제안하여 YOLOv2가 레이블이 지정된 객체 클래스 없이 객체 클래스 감지를 예측할 수 있도록 합니다.

탐지 데이터. 동시에 다중 스케일 훈련 방법을 사용하여 동일한 모델을 서로 다른 크기의 이미지에서 실행하여 속도와 정확도 간에 적절한 균형을 제공합니다. YOLOv3 [21]는 darknet-53을 백본 네트워크로 사용하고 잔여 구조는 새로운 기능을 추출할 때 가능한 한 원래 기능을 유지할 수 있으며 어느 정도 그래디언트 소실 문제를 제거할 수 있습니다. 동시에 풀링으로 인한 그래디언트의 부정적 효과를 줄이기 위해 풀링 대신 2의 합성곱 단계 크기를 사용하여 다운샘플링을 달성합니다. 작은 대상에 대한 알고리즘의 탐지 효과를 높이기 위해 YOLOv3는 FPN [[34]]과 유사한 업샘플링 퓨전 방법을 채택하여 기능 맵의 의미 정보를 늘리고 다중 스케일 탐지 방법을 채택합니다.

다중 레이블 객체를 지원하기 위해 Softmax 대신 로지스틱을 사용하여 범주를 예측합니다. 이러한 개선을 통해 속도와 정확도가 크게 향상되었습니다. SSD [[35]]는 또 다른 효과적인 단일 단계 대상 감지기입니다. 다양한 크기의 객체를 감지하기 위해 피처 피라미드를 구성합니다. 이런 식으로 큰 피처 맵은 작은 대상을 감지할 수 있고 작은 피처 맵은 큰 대상을 감지할 수 있습니다. 동시에 YOLO의 마지막 레이어와 달리 SSD는 감지에 합성곱 피처 맵을 사용합니다. SSD는 참조를 위해 Fast R-CNN의 앵커 프레임 설계 개념을 사용하고 각 픽셀에서 길이-너비 비율이 다른 사전 선택 프레임을 설정하여 어느 정도 학습의 어려움을 줄일 수 있습니다. 2단계 감지와 비교하여 SSD는 감지 정확도와 실시간 효율성이 더 높습니다.

그러나 SSD는 여전히 클래스 불균형 문제가 있어 작은 대상을 효과적으로 감지할 수 없습니다. 이러한 단점에 따라 일부 학자들은 개선을 이루었습니다. 업샘플링 방법을 통해 딥 피처 맵과 얕은 피처 맵이 융합되어 피처 맵의 의미 정보를 풍부하게 하고 작은 대상의 감지를 더욱 편리하게 만들었습니다. RFBnet [[36]]은 Inception [[37]]의 구조에 홀 합성곱을 추가하여 인간의 시각 수용 필드를 시뮬레이션하여 네트워크의 피처 추출 능력을 개선하고 속도를 보장하는 조건에서 감지 정확도를 향상시킵니다. 초점 손실 [[38]]은 그림에 사전 선택 상자가 많지만 그 중 일부에만 대상이 포함되어 있어 마이너스 샘플 수가 너무 많아 총 손실의 최고를 차지하고 대부분은 샘플 분류에 적합하여 범주 불균형의 문제가 발생한다고 생각합니다. 초점 손실은 쉽게 분류되는 샘플의 무게를 줄임으로써 일부 범주 불균형의 문제를 해결합니다. 동시에 간단하고 밀도가 높은 네트워크가 정확성과 속도를 보장하도록 설계되었습니다.

3. 방법

3.1.Depth target detection algorithm

컴퓨터 비전 분야에서 객체 인식 기술은 비디오와 사진에서 샘플 객체를 감지할 수 있는 알고리즘입니다. 최근의 타겟 인식 알고리즘은 대부분 다층 신경망과 딥 러닝 소프트웨어 프레임워크를 기반으로 하는 고성능 GPU 칩과 수천 개의 스트림 프로세서를 내장하고 있습니다. 따라서 딥 객체 감지(Deep OD)라고도 합니다. 객체 감지는 항상 컴퓨터 비전 분야에서 중요한 문제였습니다. 딥 러닝이 객체 감지에 적용되기 전에는 기존의 객체 감지 기술이 수동 기능 수집을 기반으로 했습니다. 기존의 타겟 감지 방법은 일반적으로 영역 선택, 기능 추출 및 회귀 분류의 세 단계로 나뉩니다. 그러나 몇 가지 문제가 있습니다. 영역 선택에서 전략 효과가 좋지 않고 시간 복잡도가 높으며 수동으로 추출한 기능의 견고성이 그다지 좋지 않습니다. 타겟 감지 분야에 딥 러닝을 적용하면 타겟 감지가 딥러닝은 2단계 타겟 감지와 단일 단계 타겟 감지의 두 가지 유형으로 구분할 수 있습니다.

2단계 타겟 탐지는 이름에서 알 수 있듯이 전체 타겟 탐지 프로세스가 두 단계로 나뉜다는 것입니다. 먼저 가능한 영역을 제안하고 동시에 특징을 추출한 다음 분류하고 회귀합니다. 2단계 타겟 탐지 방법은 영역 위치를 기반으로 합니다. 가장 일반적인 것은 R-CNN(Region-CNN) 시리즈입니다.

이름에서 알 수 있듯이 단일 단계 대상 감지는 단계별 학습 없이 감지 및 분류 작업을 한 번의 학습으로 완료하는 것입니다. 분류의 정확성을 보장하는 조건에서 성능이 크게 향상됩니다. 일반적인 단일 단계 대상 감지 알고리즘에는 YOLO, SSD, YOLOv2 및 YOLOv3가 포함됩니다.

응용 프로그램의 관점에서, 타겟 탐지는 중요한 타겟 탐지와 일반 타겟 탐지를 포함합니다. 이러한 것들 안에서, 일반 타겟 탐지는 객체 인식, 차량 탐지, 보행자 탐지, 얼굴 탐지 등을 포함합니다. 중요한 타겟 탐지는 인간의 시각에서 이미지로부터 중요한 타겟을 추출하는 효과를 모방해야 합니다. 대표적인 타겟 탐지 알고리즘에는 R-CNN, FastR-CNN, Fast R-CNN SSD(Single Shot MultiBox Defender), Yolo(You Only Look Once) 등이 있습니다. 이러한 알고리즘은 신경망의 토폴로지를 개선하여 표준 오픈 데이터 세트에 대한 인식률과 성능을 지속적으로 개선합니다.

3.2.Joint optimization method

본 논문은 Softmax 손실과 개선된 Island 손실의 공동 최적화 방법을 제안하는데, 이는 Softmax 손실에 기반한 표현 분류 작업과 개선된 Island 손실에 기반한 메트릭 학습 작업을 AlexNet-Emotion 네트워크에 통합합니다. Softmax 손실은 모델이 표현 특징을 학습하게 할 수 있고, 개선된 Island 손실은 모델이 학습한 표현 특징을 더욱 구별적이고 컴팩트하게 만들 수 있습니다. 다중 작업 학습을 통해 두 작업의 장점이 활용되어 패턴의 구별 능력과 견고성이 향상됩니다.

이 논문에서는 쉽고 사용 가능한 Island loss 함수를 선택하여 그 단점을 개선했습니다. 그 효과는 원래 함수보다 더 나은 것으로 증명되었습니다. Island loss의 주요 기능은 샘플을 해당 클래스 중심으로 밀어 넣고 클래스 중심을 서로 떨어뜨리는 것입니다. 샘플과 동족 클래스 중심 사이의 범위를 계산할 때 제곱 차이 손실 함수를 사용하여 모든 샘플의 손실과 역전파를 계산합니다. 모든 샘플의 손실로 인해 많은 계산 손실이 발생하고 분류하기 더 어려운 샘플에 주의를 기울이지 않는다는 번거로운 문제를 해결하기 위해. 이 논문은 island loss를 기반으로 개선된 Island loss라고 하는 온라인 어려운 샘플 마이닝 기술을 추가합니다. 클래스 중심의 반경을 초과하는 비교적 어려운 샘플의 손실만 계산함으로써 모델은 비교적 어려운 샘플을 추가로 조정하고 모델의 수렴 속도를 높일 수 있습니다. 샘플과 동족 클래스 사이의 중심 거리 L_c의 계산 공식은 다음과 같습니다.

$$L_c = \frac{1}{2}\sum_{i=1}^{m} max\left(||x_i - c_{y_i}||^2 - \tau,\ 0\right) \quad (1)$$

여기서 m은 배치 학습 샘플의 수이고, y_i는 샘플 $x_{(i)}$의 실수 레이블이며, c_{y_i}은 샘플 $x_{(i)}$에 해당하는 클래스 중심이고, x_i은 샘플 $x_{(i)}$의 FC7 계층의 특징 벡터이고, τ는 임계값입니다.

코사인 유사도는 클래스 중심 간 거리를 계산하는 데 사용됩니다. 값이 작을수록 클래스 중심 간 거리가 커집니다. Island Loss 함수의 개선된 공식은 다음과 같습니다.

다음과 같습니다:

$$L_{IL} = L_c + {}_1 \sum_{c_j \in K} \sum_{\substack{c_k \in K \\ c_{k \neq} c_j}} \left(\frac{c_k c_j}{||c_k||_2 \left|\left|c_j\right|\right|_2} + 1 \right) \tag{2}$$

여기서 k는 표현 유형입니다. • 벡터 점 곱셈, $\| c_j \|_2$은 c_j의 L2 노름, λ_1은 클래스 중심의 손실 가중치, c_k는 클래스 K의 클래스 중심입니다.

따라서 소프트맥스 손실과 개선된 아일랜드 손실의 공동 최적화 방법은 소프트맥스 손실과 개선된 아일랜드 손실의 장점을 활용하여 AlexNet-Emotion 네트워크를 드릴링하기 위해 제안되었으며, 이는 작은 클래스 간 분산과 큰 클래스 내 분산의 문제를 효과적으로 해결할 수 있습니다. 모델의 전체 손실 함수는 다음과 같습니다.

$$L = L_s + L_{IL} \tag{3}$$

$$L_s = -\frac{1}{m} \sum_{i=1}^{m} \log \frac{e^{z_{yi}}}{\sum_{j=1}^{k} e^{z_j}} = -\frac{1}{m} \sum_{i=1}^{m} \log \left(\frac{e^{w_{y_i}^T x_i + b_{y_i}}}{\sum_{j=1}^{k} e^{w_j^T x_i + b_j}} \right) \tag{4}$$

이 중 L_s는 Softmax 손실, λ은 개선된 Island 손실의 가중치, y_i은 샘플 $x_{(i}$)의 실수 범주 레이블, x_i는 샘플 F의 FC7 계층의 표현 특징 벡터, m과 K는 드릴링 샘플과 표현 범주의 계정, W_j과 b_j은 범주 j의 가중치와 편향, z_j은 범주 j의 점수이다.

표현 클래스 중심은 평균 샘플의 특징 벡터에 의해 계산되며, 이는 노이즈 간섭을 유발하므로, 랜덤 경사 하강법(SGD) 방식을 적용하여 클래스 중심을 업데이트합니다. j-타입 표현 센터의 업데이트 공식은 다음과 같습니다.

$$\Delta c_j = \frac{\sum_{i=1}^{m} \delta(y_i, j)\left(c_j - x_i\right)}{1 + \sum_{i=1}^{m} \delta(y_i, j)} + \frac{\lambda_1}{|K| - 1} \sum_{\substack{c_k \in N \\ c_k \neq c_j}} \frac{c_k}{\|c_k\|_2 \|c_k\|_2} - \left(\frac{c_k \cdot c_j}{c_{k2} \left|\left|c_j\right|\right|_2^3} \right) c_j \tag{5}$$

$$c_j^{t+1} = c_j^t - a c_j^t \tag{6}$$

여기서 a는 클래스 중심 업데이트 학습률이고 t는 t번째 반복입니다. $y_i = j$, $\delta(y_i, j) = 1$일 때; $y_i \neq j$, $\delta(y_i, j) = 0$일 때.

본 논문에서는 영상 프레임에서 표정 강도의 변화 과정을 효과적으로 검출하기 위하여 IRNN을 이용하여 영상 내 시간 정보를 추출한다.

구체적으로, 비디오를 N개의 이미지 프레임으로 변환하고, AlexNet-Emotion 네트워크를 사용하여 각 프레임의 공간적 특징을 추출하여 n개의 128차원 표현 특징 벡터를 얻은 다음, 이를 IRNN 네트워크로 전송합니다. IRNN은 은닉층을 통해 시퀀스의 각 프레임의 종속성을 학습하여 표현의 시간적 특징을 파악합니다. 각 순간의 은닉층 표현에는 이전 순간의 표현 변화 정보가 포함됩니다.

프레임 간의 표현 특징을 효과적으로 통합하기 어려운 번거로움을 해결하기 위해 각 프레임의 중요도를 구분하기 위한 어텐션 모듈을 도입했다. 현재 프레임의 은닉층 표현을 다른 프레임과 비교하여 현재 프레임의 중요도를 얻을 수 있다. 비디오에서 임의의 두 비디오 프레임의 시간 복잡도는 $O(n^2)$이다.

기계 번역에서 주의 메커니즘을 도입하여 비교 시간을 단축하는 데 영감을 받은 이 논문은 주의 메커니즘을 사용하여 IRNN 네트워크를 임베드합니다. 그림 4.1에서 볼 수 있듯이 IRNN 네트워크는 비디오 코딩 벡터 c를 생성합니다. 비디오 코딩 벡터 c와 은닉 계층 표현 h_i 사이의 유사도를 계산하여 각 프레임의 은닉 계층 표현에 가중치를 부여하고, 시간 복잡도가 O(n)인 비디오 수준 표현 특징을 얻기 위해 선형 가중치를 수행합니다. 계산 공식은 다음과 같습니다.

$$a_i = c^T V h_i \tag{7}$$

$$_i = \frac{\exp(a_i)}{\sum_{t=1}^{T} a_t} \tag{8}$$

$$f_v = \sum_{l=1}^{T} \xi_i h_i \tag{9}$$

여기서 t는 입력 IRNN 비디오 프레임의 수이고, f_v는 비디오 수준 표현 특징이고, ξ_i은 i번째 은닉층 표현의 가중치이고, V는 학습 매개변수입니다.

3.3.Occluded face recognition model based on dual discrimination countermeasure network

본 논문에서는 생성된 대책 네트워크를 기반으로 한 가려진 표정 인식 모델을 제안한다. 이 모델은 그림 5와 같이 가려진 얼굴 이미지 복원과 얼굴 인식의 두 모듈로 나뉜다.

4. 결과

4.1.Deep facial expression recognition based on static image

네트워크 정적 데이터 처리의 편의성과 가용성으로 인해 많은 연구가 표정 인식을 위한 시간 정보를 고려하지 않고 정적 이미지에 기반을 두고 있습니다. 비교적 작은 얼굴 표정 데이터베이스에 딥 네트워크를 직접 학습시키면 필연적으로 과적합 문제가 발생합니다. 이 문제를 해결하기 위해 많은 관련 연구에서는 추가 보조 데이터를 사용하여 사전 학습하고 자체 네트워크를 구축하거나 AlexNet(Krizhevsky et al., 2012), visualgeometry group(VGG, [27]), VGG-Face(Parkhi et al., 2015) 및 GoogLeNet [[37]]과 같은 효과적인 사전 학습 네트워크를 기반으로 직접 미세 조정합니다. CASIA WebFace(Yi 등, 2014), CFW(cellular face in the wild)(Zhang 등, 2012) 및 FaceScrub 데이터 세트(Ng 및 Winkler, 2014)와 같은 대규모 얼굴 인식 데이터베이스와 FER2013(Goodfellow 등, 2013) 및 TFD(Toronto to face 데이터베이스)와 같은 비교적 대규모 얼굴 표정 데이터베이스는 보조 학습에 적합합니다.Kaya 등(2017)은 얼굴 데이터에 사전 학습된 VGG 얼굴 모델이 객체 데이터에 사전 학습된 Imagenet 모델보다 얼굴 표정 인식 작업에 더 적합하다고 지적했습니다.[28] 또한 대규모 얼굴 데이터베이스에 대한 사전 학습과 추가 표정 데이터베이스에 대한 추가적인 미세 조정이 표정 인식률을 효과적으로 개선할 수 있다고 지적했습니다. 또한, [29]은 다단계 미세 조정 전략을 제안했습니다.첫 번째 단계에서는 추가 얼굴 표정 데이터베이스 FE R 2013을 사용하여 기존 사전 학습 모델을 미세 조정했습니다.두 번째 단계에서는 알려진 (EmotiW) 학습 세트를 사용하여 모델을 미세 조정하여 대상 데이터베이스에 더 적합하게 만들었습니다.Ding et al. (2017)은 사전 학습 모델에 보관된 얼굴 정보가 표정 인식 작업에 미치는 간섭을 제거하기 위해 Face et2ExpNet 프레임워크를 제안했습니다.

4.2.Deep facial expression recognition network based on dynamic image sequence

입력 시퀀스에서 연속된 프레임 간의 시간 상관 정보는 얼굴 표정 인식에 유익합니다. 이 섹션에서는 동적 시퀀스를 기반으로 하는 심층 시공간적 표정 인식 네트워크에 초점을 맞춥니다. 네트워크는 시간 창 기간 동안 특정 범위의 프레임을 사용합니다.

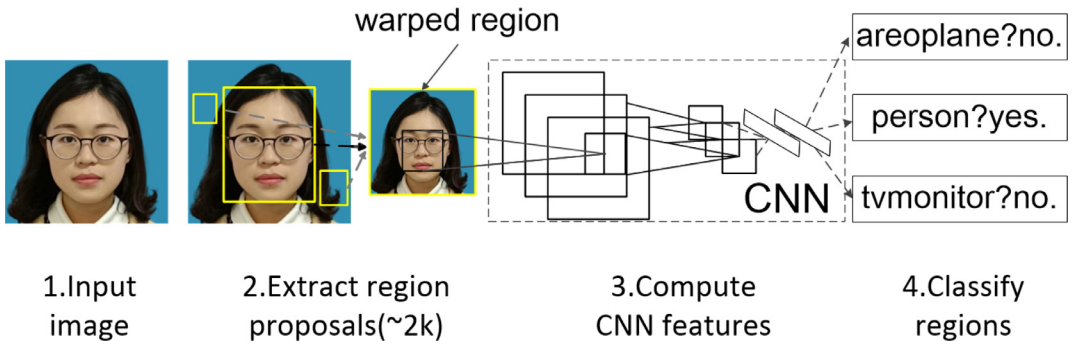

Fig. 1. Architeture of R-CNN.

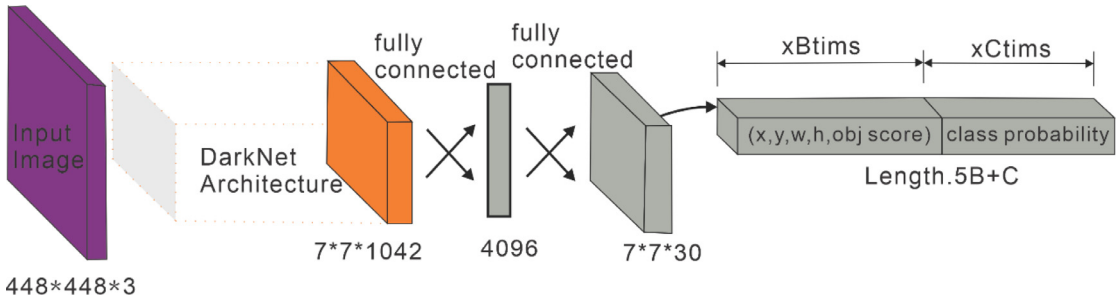

Fig. 2. Architeture of YOLO.

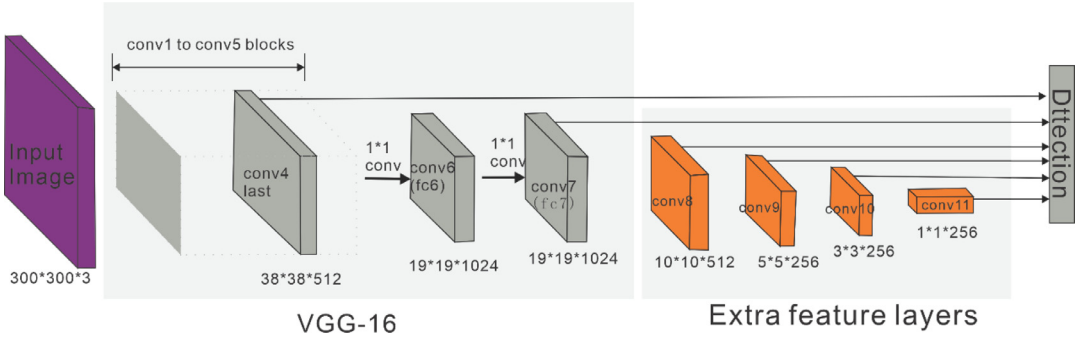

그림 3. SSD의 아키텍처.

별도의 입력으로 비디오 시퀀스의 시공간적 동작 모드를 고려하고 공간적 및 시간적 정보를 사용하여 보다 미묘한 표정을 캡처합니다.비디오 기반의 동적 얼굴 표정 인식은 비디오에서 사람의 표정을 인식하고 비디오를 기본 표정으로 분류하는 것입니다.각 비디오 프레임에는 표정 변화가 포함되어 있습니다.시간 정보를 사용하여 각 프레임의 특징을 비디오 수준 표현으로 효과적으로 통합하는 방법은 비디오 분류에서 어려운 문제입니다.초기 작업에서는 CNN을 사용하여 공간 정보를 추출하고 평균 집계를 통해 비디오 수준 표정 특징[22]을 생성했습니다.

최근 RNN은 연속적인 은닉층 표현을 통해 시퀀스에 대한 정보를 전파하고 시간 정보를 효과적으로 활용합니다. 일부 학자들은 마지막 순간의 은닉층 표현을 비디오 레벨 표현 특징으로 사용하는 CNN-RNN 시공간 모델을 추가로 제안합니다. 그러나 위에서 언급한 모델은 비디오 레벨 표현 특징을 생성하기 위해 표현 변화 정보를 효과적으로 사용할 수 없습니다. 이 문제를 해결하기 위해 AlexNet-Emotion 네트워크를 기반으로 주의 메커니즘을 기반으로 하는 동적 얼굴 표정 인식 모델이 제안되었습니다. 먼저 AlexNet-Emotion 네트워크를 사용하여 비디오 프레임의 공간 정보를 추출합니다. 그런 다음 순환 신경망을 사용하여 시간 비디오 프레임의 형성. 둘째, 주의 메커니즘을 사용하여 명확한 표현이 있는 일부 비디오 프레임에 더 많은 가중치를 부여하여 비디오 수준 표현 기능을 생성합니다. 마지막으로 분류는 전체 연결 계층[23]을 통해 실현됩니다.

4.3.Summary of target detection based on deep learning

2단계 타겟 감지이든 단일 단계 타겟 감지이든 타겟 감지 알고리즘의 성능은 일반적으로 감지 속도와 감지 정확도로 측정됩니다. 2단계 타겟 감지와 단일 단계 타겟 감지는 모두 속도와 정확도 간의 균형을 찾으려고 합니다. 그러나 두 방법의 강조점은 다릅니다. 단일 단계 타겟 감지 알고리즘은 감지 속도에 더 많은 주의를 기울이고 2단계 타겟 감지 알고리즘은 감지 정확도에 더 많은 주의를 기울입니다[24]. 또한 두 종류의 타겟 감지 알고리즘을 개발할 때 서로에게서 배우고 융합하여 더 나은 감지 결과를 얻습니다. 두 가지 유형의 타겟 감지 알고리즘은 적용 시나리오가 다릅니다. 예를 들어 실시간 타겟 감지, 단일 단계 타겟 감지가 더 좋습니다. 이미지 장면이 더 복잡한 경우 2단계 타겟 감지 알고리즘의 정확도가 필요합니다.

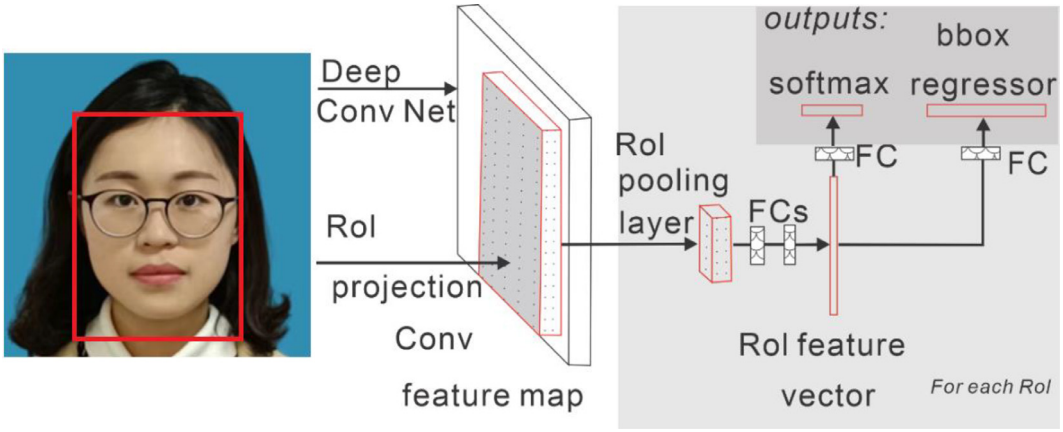

그림 4. Fast R-CNN의 아키텍처.

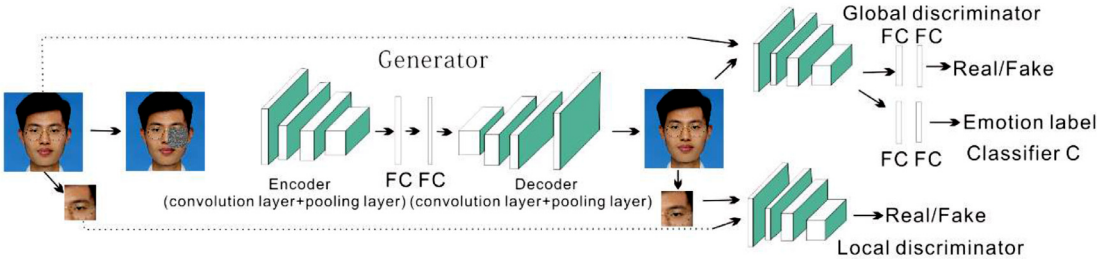

그림 5. 이중 판별 네트워크 기반 폐색 표현 인식 모델.

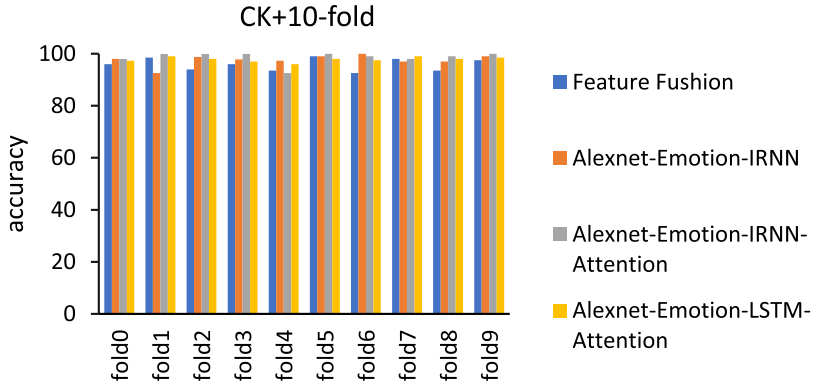

그림 6. CK + 데이터 세트를 기반으로 한 4개 실험 그룹의 10배 교차 검증 정확도 막대 그래프.

표 1 제안된 알고리즘과 다른 알고리즘의 비교 결과.

Algorithm	CK+	MMI
AUDN	93.7%	75.85%
DeRL	96.57%	72.67%
IL-CNN	94.39%	70.67%
Inception	93.2%	77.6%
IACNN	95.37%	71.55%
AlexNet-Emotion Iinit	94.41%	65.88%
AlexNet-Emotion Softmax loss	96.45%	76.89%
AlexNet-Emotion w/o Angle Variance	93.73%	71.84%
AlexNet-Emotion Pre	97.14%	78.68%

이 논문에서 제안한 알고리즘의 타당성을 더욱 높이기 위해 현재 널리 사용되는 알고리즘과 비교하여 비교 결과를 표 1에 나타냈습니다. Liu et al.은 액션 단위에서 영감을 받은 딥 네트워크(AUDN)를 분석했습니다. 첫째, 마이크로 액션 패턴(MAP)은 합성곱 계층과 풀링 계층을 통해 학습됩니다. 그런 다음 기능 그룹화를 사용하여 관련 맵과의 적응적 조합을 통해 더 큰 수용 필드를 시뮬레이션하여 더 추상적인 중간 계층 의미론을 생성합니다. 마지막으로 각 수용 필드에서 다중 계층 학습기를 사용하여 그룹화 하위 네트워크를 구성하여 표현 분석에 도움이 되는 고수준 표현을 추가로 생성합니다. [26]은 DeR L(deexpression residual learning) 알고리즘을 제안했습니다. 중립적 표현을 생성하는 과정에서 표현 정보가 필터링되고 표현 정보는 여전히 중간 계층에 유지됩니다. 중간 계층의 잔여 정보는 표현 기능을 생성하는 데 사용됩니다. Cai 등은 Island loss optimization [25] 모델을 사용하는 것을 제안했습니다. Mollahoseini 등은 두 개의 합성곱 계층, 최대 풀링 계층 및 네 개의 개시 계층으로 구성된 모델을 제안했습니다. 본 논문에서는 얼굴 주요점 각도 변화 특징 맵의 보조 정보를 제공하기 위해 AlexNet-Emotion 네트워크를 제안한다. 표 1의 비교 결과는 본 논문에서 제안하는 알고리즘의 정확성이 제안하는 알고리즘보다 더 높음을 보여준다.

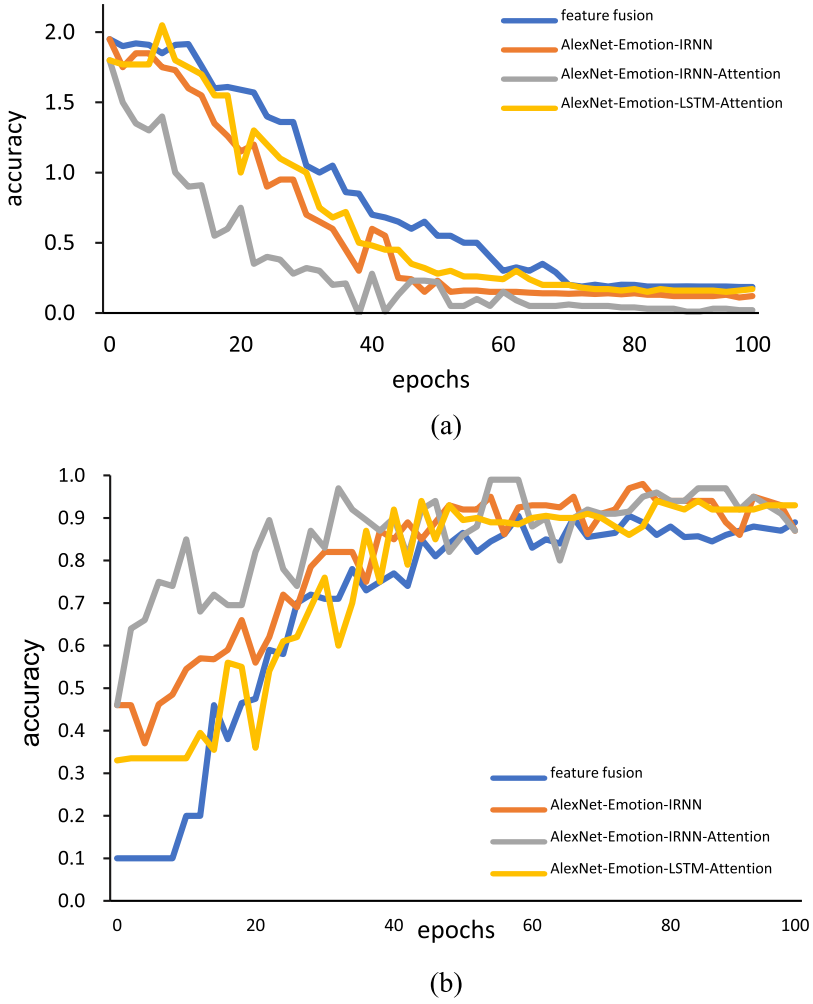

그림 7. (a) 4개 실험 그룹의 훈련 손실 (b) 테스트 세트의 정확도.

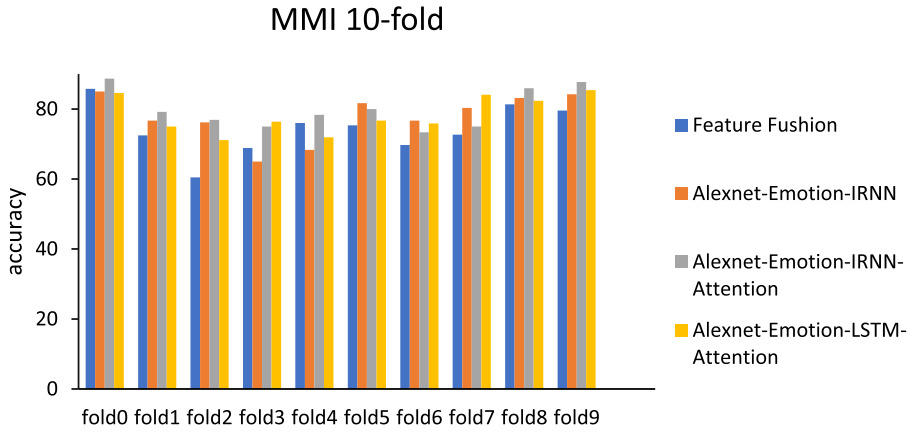

그림 8. MMI 데이터 세트에 대한 4개 실험 그룹의 10배 교차 검증 정확도 막대 그래프.

위 알고리즘보다 더 나은 성능을 보였으며, 이를 통해 AlexNet-Emotion 모델의 효과성이 검증되었습니다.

CK + 데이터 세트에 대한 4개 실험 그룹의 10배 교차 검증 정확도 막대 그래프(그림 6)가 제시됩니다. 나아가, 4개 실험 그룹의 훈련 손실과 테스트 세트의 정확도도 설명합니다(그림 7). 마지막으로, MMI 데이터 세트에 기반한 4개 실험 그룹의 10배 교차 검증 정확도 막대 그래프가 수행됩니다(그림 8).

5. 토론

본 논문은 먼저 얼굴 표정 인식의 배경 지식을 제안하고 얼굴 표정 인식 분야에서 데이터베이스와 알고리즘의 진화와 발전을 요약합니다. 이 분야에서 딥러닝이 주류 프레임워크가 되었다는 점을 지적합니다. 그런 다음 딥러닝 기반 표정 인식 알고리즘을 두 가지 범주로 나눕니다.

egories(정적 표현 인식 네트워크와 동적 표현 인식 네트워크). 현재 공통 표현 데이터베이스를 기반으로 하는 많은 알고리즘의 성능을 비교해보면, 동적 표현 네트워크가 정적 표현 네트워크보다 더 나은 인식 결과를 얻을 수 있는 경우가 많다는 것을 알 수 있다. 표현이 동적 과정에 있다는 사실과 동적 시퀀스에 기반한 심층 표현 인식 알고리즘이 효과적인 시간 정보를 포착하고 동일한 시퀀스에 있는 문자의 식별 정보를 활용할 수 있다는 사실을 감안할 때, 공간-시간 심층 네트워크가 표현 인식 분야의 주요 트렌드가 될 것이다.

데이터 집합의 관점에서, 정확한 표정 태그가 있는 다양한 훈련 데이터를 수집하면 표정 인식률을 근본적으로 개선할 수 있습니다. 알고리즘의 관점에서, 얼굴 동작 단위 모델, 쾌감 각성 차원 모델과 같은 다른 표정 모델과 오디오 모드, 3D 얼굴 깊이 정보, 인체 생리 정보와 같은 다른 멀티모달 모델을 결합하면 표정 인식을 보다 실용적으로 만들 수 있습니다.

6. 결론

컴퓨터 비전 분야에서 국내외 연구 현황을 바탕으로 얼굴 표정 인식 기술은 큰 진전과 발전을 이루었습니다. 그러나 연구자들이 해결해야 할 많은 과제와 어려움이 여전히 남아 있습니다. 예를 들어, 실제 장면에서 얼굴 표정 인식에 대한 연구와 다양한 표정 간에는 어느 정도 혼란이 있습니다. 얼굴 감정의 표현은 지역, 문화, 환경에 따라 다를 수 있습니다. 성격에 차이가 있습니다. 따라서 얼굴 표정 인식 기술은 더 나은 결과를 얻기 위해 다양한 측면에서 개선되어야 합니다. 딥 러닝은 매우 강력한 연구 방향을 제공합니다. 합성곱 신경망의 합성곱 계층, 풀링 계층, 전체 연결 계층을 사용하면 이 네트워크 구조가 스스로 관련 기능을 연구하고 선택하여 사용할 수 있습니다. 이 기능은 여러 연구에 많은 편의성을 제공하며 과거의 매우 어려운 모델링 프로세스를 생략할 수 있습니다. 또한 딥 러닝은 이미지 분류, 객체 감지, 포즈 추정 및 이미지 분할에서 큰 개선을 이루었습니다. 한편, 딥러닝은 광범위한 응용 분야와 강력한 다재다능성을 가지고 있으므로 다른 응용 분야로 확장하기 위해 계속해서 노력할 수 있습니다. 반면, 딥러닝은 여전히 탐색하고 발견해야 할 잠재적인 문제가 많이 있습니다(그림 1-3, 식 (1)-(9)).

경쟁 이익 선언

저자들은 이 논문에는 이해상충이 없다고 선언합니다.

승인

본 연구는 중국 국가자연과학기금(No. 62006102)의 지원을 받아 수행되었습니다.

참고문헌

[1] C. He, 얼굴 인식 기술 개요, Intell.Comput.Appl. 6(5)(2016) 112 – 114. [2] M.H. Zhu, S.T. Li, Y. Hua, 희소 표현에 기반한 얼굴 표정 인식 방법, Pattern Recognit.Artif.Intell. 27(8)(2014) 708 – 712. [3] Y.L. Xue, X. Mao, C. Catalin-Daniel, 폐색 조건에서의 강력한 얼굴 표정 인식 방법, J. Beijing Univ. Aeronaut. Astronaut. 36(4)(2010) 429 – 433. [4] Y. Chenglin, P. Hailong, 효과적인 컴퓨팅과 적대적 신경망을 기반으로 한 얼굴 인식 프레임워크와 소셜 로봇을 위한 머신 비전에서의 구현, Comput.Electr.Eng. 92(2021) 92 .

[5] Y. Sun, X. Wang, X. Tang, 10,000개 클래스 예측을 통한 얼굴 표현 심층 학습, IEEE 컴퓨터 비전 및 패턴 인식 컨퍼런스 회의록, 하와이 호놀룰루, 2014, pp. 1891 – 1898. [6] Y. Sun, Y. Chen, X. Wang 등, 공동 식별 검증을 통한 얼굴 표현 심층 학습, 스페인 바르셀로나에서 열린 신경 정보 처리 시스템 발전 회의록, 2014, pp. 1988 – 1996. [7] Y. Sun, X. Wang, X. Tang, 심층적으로 학습된 얼굴 표현은 희소하고 선택적이며 견고합니다.IEEE 컴퓨터 비전 및 패턴 인식 컨퍼런스 회의록, 호놀룰루, 하와이, 2015, pp. 2892 – 2900. [8] F. Schroff, D. Kalenichenko, J. Philbin, Facenet: 얼굴 인식 및 클러스터링을 위한 통합 임베딩, IEEE 컴퓨터 비전 및 패턴 인식 컨퍼런스 회의록, 호놀룰루, 하와이, 2015, pp. 815 – 823. [9] Y. Taigman, M. Yang, M.A. Ranzato, et al. , Deepface: 얼굴 검증에서 인간 수준 성능과의 격차 해소, IEEE 컴퓨터 비전 및 패턴 인식 컨퍼런스 회의록, 호놀룰루, 하와이, 2014, pp. 1701 – 1708. [10] Y. Wen, K. Zhang, Z. Li, et al., 딥 얼굴 인식을 위한 차별적 특징 학습 접근법, 유럽 컴퓨터 비전 컨퍼런스, Springer, Cham, 베를린, 2016, pp. 499 – 515. [11] W. Liu, Y. Wen, Z. Yu, et al., 합성곱 신경망을 위한 대마진 소프트맥스 손실, 국제 기계 학습 컨퍼런스 회의록, 2, ICML, 스톡홀름, 2016, p. 7. [12] F. Wang, J. Cheng, W. Liu 외, 얼굴 검증을 위한 가산 마진 소프트맥스, IEEE Signal Processing Letters, 25, PISCATAWAY, USA, 2018, pp. 926 – 930. [13] H. Wang, Y. Wang, Z. Zhou 외, CosFace: 심층 얼굴 인식을 위한 대마진 코사인 손실, IEEE 컴퓨터 비전 및 패턴 인식 컨퍼런스 논문집, 호놀룰루, 하와이, 2018, pp. 5265 – 5274. [14] S. Chen, Y. Liu, X. Gao 외. , MobileFaceNets: 모바일 기기에서 정확한 실시간 얼굴 검증을 위한 효율적인 CNN, 중국 생체 인식 컨퍼런스, Springer, Cham, Berlin, 2018, pp. 428 – 438. [15] M. Sandler, A. Howard, M. Zhu, et al., MobileNetV2: 역 잔차 및 선형 병목 현상, IEEE 컴퓨터 비전 및 패턴 인식 컨퍼런스 회의록, Honolulu, HI, 2018, pp. 4510 – 4520. [16] R. Girshick, J. Donahue, T. Darrell, et al., 객체 감지를 위한 R-CNN, IEEE 컴퓨터 비전 및 패턴 인식 컨퍼런스 회의록, 2014. [17] Zhu G., Porikli F., Li H. 에지 박스에서 무작위로 움직이는 객체 추적 제안.arXiv 사전 인쇄본 arXiv: 1507.08085, 2015. [18] J.R.R. Uijlings, K.E. A. Van De Sande, T. Gevers, et al., 객체 인식을 위한 선택적 검색, Int. J. Comput. Vis. 104(2) (2013) 154 – 171. [19] C. Lee, J.H. Kim, K.W. Oh, 객체 감지를 위한 더 빠른 R-CNN 모델 비교[C]//2016 제16회 제어 국제회의, 자동화 및 시스템(iccas), IEEE, 2016, pp. 107 – 110. [20] S. Ren, K. He, R. Girshick 외, Faster R-CNN: 영역 제안 네트워크를 통한 실시간 객체 감지를 향해, IEEE 패턴 분석 및 머신 인텔리전스 거래 39(6)(2016) 1137 – 1149. [21] X.X. Zhang, X. Zhu, 빠른 이미지 등록 및 개선된 YOLOv3 네트워크를 통한 항공 적외선 이미지 시퀀스에서 이동 차량 감지, Int. J. Remote Sens. 41(11)(2020) 4312 – 4335. [22] Nguyen H.T. 얼굴 인식을 위한 얼굴 특징 추출에 대한 기여. uni- versité de grenoble, 2014. [23] Z. Tang, G. Zhao, T. Ouyang, 단기 풍향 예측을 위한 2단계 딥 러닝 모델, Renew. Energy 173 (2021) 1005 – 1016, doi: 10.1016/j.renene.2021.04.041. [24] C. Lunerti, R.M. Guest, R. Blanco-Gonzalo, et al., 스마트폰의 얼굴 인식에 대한 환경적 영향, IEEE 51st International Carnahan Conference on Security Technology 회의록, IEEE, 2017. [25] Z.H. Tang, Y.Y. Li, X.Y. Chai, H.Y. Zhang, S.X. Cao, 발전소 보일러의 부하 제약 하에서 NOx 배출에 대한 적응 비선형 모델 예측 제어, J. Chem. Eng. Jpn. 53 (1) (2020) 36 – 44, doi: 10.1252/jcej.19we142. 영어: [26] L.C. Yan, B. Yoshua, H. Geoffrey, 심층 학습, Nature 521 (7553) (2015) 436 – 4 4 4 . [27] Simonyan K., Zisserman A., 대규모 이미지 인식을 위한 매우 깊은 합성곱 신경망. arXiv 사전 인쇄본 arXiv:1409.1556, 2014. [28] Knyazev B., Shvetsov R., Efremova N., et al. 비디오에서 감정 분류를 위한 대규모 얼굴 인식 데이터 세트에 대해 사전 학습된 합성곱 신경망. arXiv 사전 인쇄본 arXiv:1711.04598, 2017. [29] Yue-Hei Ng J, Hausknecht M, Vijayanarasimhan S, et al. 짧은 스니펫을 넘어서: 비디오 분류를 위한 딥 네트워크[C]//IEEE 컴퓨터 비전 및 패턴 인식 컨퍼런스 회의록. 2015: 4694-4702. [30] Gabrielsson S. 사기 저하 상황에서의 도덕적 노력: 정신과 입원 환자

[32] W. Chen, M. Fang, Y.H. Liu, et al., in: 다중 객체 추적을 기반으로 한 동적 거리 장면의 단안 의미 SLAM[C]//2017, IEEE, 2017, 599-604쪽. [33] Xu X., Dou P., Le H.A., et al. 3D 지원 2D 얼굴 인식이 딥 러닝을 만날 때: 포즈 불변 얼굴 인식을 위한 확장된 UR2D[J]. arXiv 사전 인쇄본 arXiv:1709.06532, 2017. [34] J. Wang, J. Ding, H. Guo, et al. , 마스크 OBB: 항공 이미지에서 다중 범주 객체 감지를 위한 의미 기반 마스크 지향 경계 상자 표현[J], 원격 감지 11(24)(2019) 2930. [35] Felsberg M., 컴퓨터 비전의 딥 러닝 혁명 5년 후: 온라인 이미지 및 비디오 분석을 위한 최신 방법[J]. 2017.

[36] Rezaei M, Azarmi M. Deepsocial: 사회적 거리두기 모니터링 및 코로나19 팬데믹 시 감염 위험 평가[J]. 응용 과학, 2020, 10(21): 7514. [37] S. Paul, L. Singh, 딥러닝의 발전에 대한 리뷰[C]//2015 IEEE 계산 지능 워크숍: 이론, 응용 및 미래 방향(WCI), IEEE(2015) 1 – 6. [38] H Ge, Y Dai, Z Zhu, et al., 개선된 생성적 대결 네트워크를 기반으로 한 강력한 얼굴 인식 알고리즘[J], 응용 과학 11(24)(2021) 11588.