전염병 확산 방지를 위한 딥러닝 기반 얼굴 만지기 행동 인식 연구

조성만, 김민지, 최준명, 김태형, 박주영 *김남국
울산대학교, *서울아산병원
*namkugkim@gmail.com

Implementation of Face-Touching Action Recognition System based on Deep Learning for Preventing Contagious Diseases

Sungman Cho, Minjee Kim, Joonmyeong Choi, Taehyung Kim, Juyoung Park *Namkug Kim
*University of Ulsan College of Medicine, Asan Medical Center

요 약

무의식적인 손-얼굴의 접촉으로 인한 감염의 문제점을 해결하기 위해, 얼굴 만지기 행동을 인식할 필요가 있다. 본 연구는 최근 각광을 받는 딥러닝 기술을 이용하여 비디오 영상에서 얼굴 만지기 행동 인식에 대한 연구이다. 우선, 비디오 영상에서 얼굴 만지기와 관련된 11 가지 행동에 대한 시, 공간적 특징을 컨볼루션 신경망을 통해 추출한다. 추출된 정보는 각 행동 레이블로 인코딩되어 비디오 영상에서 얼굴 만지기 행동을 분류한다. 또한, 3D, 2D 컨볼루션 신경망의 대표 네트워크인 I3D, MobileNet v3 에 대해 비교 실험을 진행한다. 제안하는 시스템을 적용하여 인간의 행동을 분류하는 실험을 진행했을 때, 얼굴을 만지는 행동을 99% 의 확률로 구분했다. 이 시스템을 이용하여 일반인이 무의식적인 얼굴 만지기 행동에 대해서 정량적으로 또는 적시적으로 인식을 하여, 안전한 위생 습관을 확립하여 감염의 확산방지에 도움을 줄수 있기를 바란다.

1. 서론

호흡기 바이러스 감염의 확산을 막기 위해서는 손을 자주 씻고, 얼굴을 만지지 않는 등의 호흡기 위생 실천이 필수적이다. 호흡기 바이러스는 주변 물체의 표면에 일정 기간 살아있는 채로 존재할 수 있고, 이를 손으로 만진 후 얼굴을 만지게 되면 코나 입을 통해 호흡기 내부로 침투하여 감염이 이루어질 수 있다. 최근 세계적으로 유행하고 있는 COVID-19 또한 호흡기 바이러스이기 때문에, 보건 기관에서는 손을 자주 씻고, 눈과 코, 입을 만지지 말고 사회적 거리를 넓히는 등의 예방 철칙을 권고하고 있다[1, 2].

호흡기 위생을 실천해야 하는 중요성에도 불구하고, 이상의 예방 철칙에 대한 이행률은 여전히 낮은 수준이다. 먼저, 손 씻기의 중요도에 대한 인식과 실천이 미흡한 상태의 어려움이 있다. 올바른 손 씻기란 흐르는 물에 손 등, 손바닥, 엄지 손가락, 손톱 끝을 비누를 사용하여 일정 시간 이상 세척하는 것으로 정의된다[3]. 우리나라의 경우 2005 년도 올바른 손 씻기를 실천하는 경우가 17%였고, 2015 년도에는 26.2%로 점차 개선되었으나 선진국의 준수율이 42%~49%인 것에 비해 낮은 준수율을 보여주고 있다[4]. 또한, 대부분의 사람들이 얼굴을 만지는 행동을 무의식적으로 행한다는 점이 있다. 이전 연구들에 의하면 인간은 무의식적으로 한 시간에 평균 15.7 회~23 회로 얼굴을 만지는 것으로 발표되었다[5, 6].

무의식적인 얼굴 접촉을 예방하는 첫번째 단계는 얼굴을 만지는 순간을 인지하고 습관을 개선하는 것이다. 얼굴을 만지는 순간의 행동을 인식하여 알람을 울리는 등의 대응으로 얼굴을 만지는 행동의 감소 효과를 기대할 수 있다.

최근 딥러닝 기반 인공지능 기술의 급격한 발전에 힘입어, 비디오에서 인간의 행동을 인식하고 분류하는 비디오 분석 연구가 급속도로 증가하였다. 특히, I3D[7] 모델의 등장은 3D 컨볼루션 신경망 구조로 인간 행동의 시, 공간적 특징(Spatio-Temporal Feature)을 추출하여 다양한 공개 데이터 셋에서 높은 성능으로 인간의 행동을 분류할 수 있음을 보여주었다. 하지만, I3D 모델은 2D 구조에 비해 한 차원이 늘어난 만큼 파라미터가 많아 학습과 추론에 많은 시간이 소요된다. 이상의 이유로, I3D 와 같은 3D 컨볼루션 신경망 모델은 실시간성이 부족하기 때문에 현실 세계에 존재하는 비디오 문제에 적용하기에는 어려움이 있다.

본 논문에서는 무의식적인 손-얼굴의 접촉으로 인한 감염의 문제점을 해결하기 위해, 딥러닝 기반 얼굴 만지기 행동 인식 시스템을 제안한다. 구체적으로는 비디오 영상에서 얼굴 만지기를 포함한 11 가지 인간의 행동에 대해 2D 컨볼루션 신경망 MobileNet v3[8]을 통과하여 시,공간적 특징을 추출하고, 추출된 특징을 사용하여 얼굴을 만지는 행동과 그렇지 않은 행동으로 분류한다. 본 연구에서는 대표적인 3D 컨볼루션 신경망 I3D 를 사용한 얼굴 만지기 행동을 인식 결과를 기준치로 잡고, 2D 컨볼루션 신경망 MobileNet v3 를 사용했을

때의 추론 시간과 정확도를 비교한다.

본 논문의 구성은 다음과 같다. 2 절에서는 최근 딥러닝 기반 비디오 분석 분야 중 하나인 인간 행동 인식의 동향에 대해 살펴본 후, 3, 4 절에서는 본 논문에서 제안하는 기법을 설명하고, 5 절에서는 제안한 기법의 성능을 실험을 통해서 확인한다. 마지막으로 6 절에서는 본 논문에 대한 결론을 맺는다.

2. 관련 연구

행동 인식을 위한 딥러닝 신경망 설계는 2D 컨볼루션 신경망과 LSTM[9]을 같이 사용하는 방법[10]과, 3D 컨볼루션 신경망을 사용하는 방법[7]으로 나누어진다. 전자의 경우 2D 컨볼루션 신경망을 거치고 난 이후의 특징들을 시계열 데이터 학습에 사용되는 신경망인 LSTM 을 통과시켜 일정 시간 동안의 특징을 학습하고 이를 통해 행동을 인식한다. 구조가 간편하다는 장점이 있지만, 전체적인 시간 정보를 충분히 고려하기 어렵다는 단점이 있다.

3D 컨볼루션 신경망 기반의 방법론은 기존에 사용하던 2D 컨볼루션 신경망에 시간 정보를 더해 총 3 차원의 컨볼루션을 사용하는 방법을 말한다. 직접적으로 시간정보를 학습할 수 있다는 장점이 있지만, 2D 컨볼루션 신경망 방법 보다 학습해야할 파라미터가 훨씬 더 많다는 단점이 존재한다. 그림 1 은 2D 컨볼루션 신경망과 LSTM 을 혼합한 방법과, 3D 컨볼루션을 사용한 신경망의 구조도를 나타낸다.

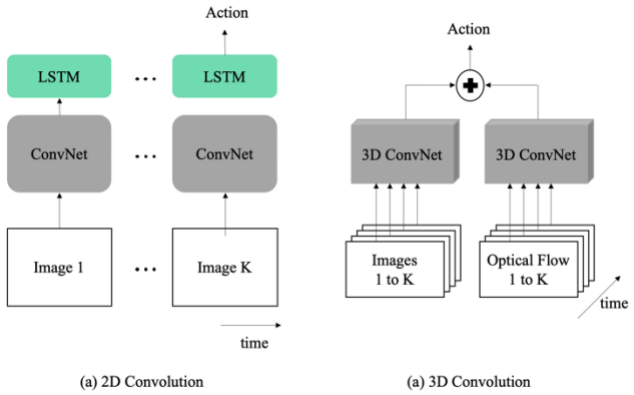

그림 1. 행동 인식을 위한 신경망 구조

3. 데이터 획득 및 전처리

학습에 필요한 데이터를 확보하기 위하여 약 50 명의 인원이 직접 촬영을 하였다. 서로 다른 10 개의 장소에서 비디오를 촬영하여 약 190,000 장의 학습 데이터를 생성하였다. 행동의 종류는 물 마시기, 마스크 벗기, 안경 만지기, 키보드 만지기 등 총 6 가지의 일반 행동과 전화 받기, 턱 괴기, 눈 비비기 등 총 5 가지의 얼굴을 만지는 행동으로 구분하였다. 그림 2 는 수집한 데이터를 보여준다.

그림 2. 학습에 사용된 데이터셋

3D 컨볼루션 신경망에서 활용 가능한 데이터를 구성하기 위하여, 동영상을 16 프레임 단위로 나누어 하나의 클립으로 만드는 작업을 수행하였다. 그림 3 은 데이터 전처리 구조를 나타낸다.

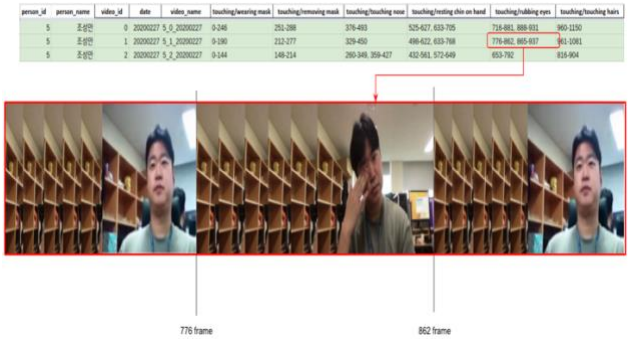

그림 3. 데이터 전처리 과정

4. 신경망 설계 및 구조

얼굴 만지는 행동을 인식하기 위하여 I3D 기반의 컨볼루션 신경망을 사용하는 방법과 MobileNet 기반의 컨볼루션 신경망을 사용하는 방법을 사용하였다.

I3D 기반의 컨볼루션 신경망에서는 학습 성능을 높이기 위하여 색상 왜곡, 임의 회전 클립 사이의 프레임 간격 조절(4~12) 등의 데이터 증강(data augmentation) 기법을 사용하였다. 옵티컬플로우를 추가해주는 방식은 오히려 정확도가 더 떨어지는 현상이 발생해서, 사용하지 않았다.

I3D 기반의 컨볼루션 신경망은 GPU 환경에서 동작할 때 빠른 속도로 높은 성능을 보이는 장점을 지니지만, 일반 CPU 환경에서 동작할 때는 속도가 많이 저하되는 문제점이 있다. 따라서, CPU 환경에서도 빠른 속도로 동작하는 신경망을 구성하기 위하여 계산량이 많은, MobileNet 기반의 컨볼루션 신경망 구조를 추가하였다.

그림 4 는 MobileNet 기반의 컨볼루션 신경망 구조에 입력으로 들어가는 3 채널 이미지를 만들어주는 과정을 나타낸다. 기존 RGB 영상을 Gray 로 만든 이후에 일정 간격을 기준으로 3 프레임을 겹쳐서 1 개의 3 채널 영상으로 만들어 준다. 이렇게 3 채널 영상이 만들어지게 되면, MobileNet 의 입력으로 들어간다.

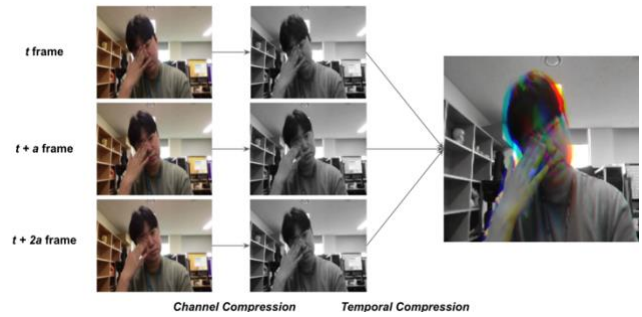

그림 4. MobileNet 기반 신경망 에서의 입력 이미지 전처리

그림 5 은 MobileNet 기반의 컨볼루션 신경망 구조에 입력으로 들어가는 3 채널 이미지를 만들어주는 과정을 나타낸다. 기존 RGB 영상을 Gray 로 만든 이후에 일정 간격을 기준으로 3 프레임을 겹쳐서 1 개의 3 채널 영상으로 만들어 준다. 이렇게 3 채널 영상이 만들어지게 되면, MobileNet v3 의 입력으로 들어간다.

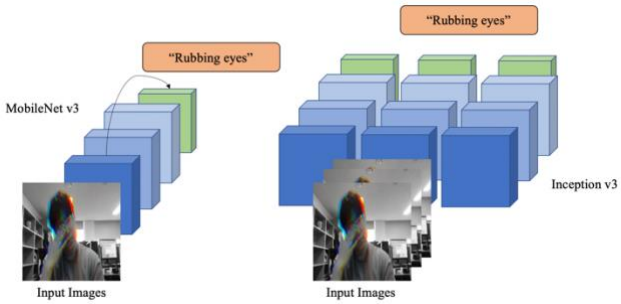

그림 5. MobileNet 기반(좌) 과 I3D 기반(우) 의 신경망 구조

5. 신경망 실험

I3D 기반 신경망의 정확도를 검사하기 위하여 선형 평가(Linear evaluation)를 실시하였다. 실험을 위하여 약 190,000 장의 전체 데이터의 10%인 19,000 장의 데이터를 테스트셋으로 설정하였다. 그림 6 은 11 개의 클래스에 대한 선형 평가 결과를 나타낸다.

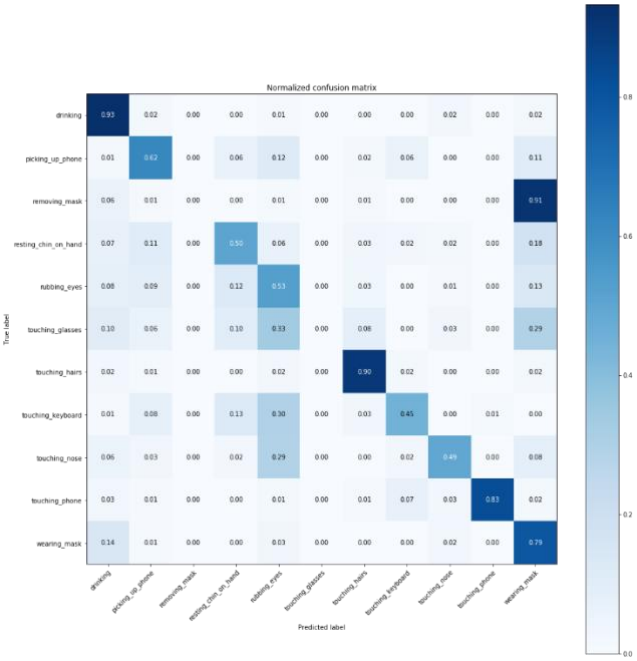

그림 6. 11 가지 클래스에 대한 선형 평가 결과

머리 만지기, 핸드폰 만지기, 마스크 쓰기, 물 마시기의 행동은 높은 정확도로 예측에 성공하였으나 마스크 벗기, 안경 만지기, 키보드 만지기 등에 대한 행동은 인식 성능이 낮은 결과를 확인할 수 있었다. 마스크를 쓰는 행동에 대해서는 마스크를 벗는 행동과 구분을 잘 하지 못하는 결과도 확인할 수 있었다.

높은 정확도 확보를 위해 11 개의 클래스에 대하여 얼굴을 만지는 행동, 얼굴을 만지지 않는 행동 2 가지로 구분하여 선형 평가도 진행하였다. 그림 7 은 얼굴을 만지는 행동과 그렇지 않은 행동 2 가지에 대한 선형 평가 결과를 나타낸다. 얼굴을 만지는 행동에는 전화 받기, 턱 괴기, 눈 비비기, 머리 만지기, 코 만지기 행동이 포함되어 있고, 얼굴을 만지지 않는 행동에는 마스크 쓰기, 마스크 벗기, 핸드폰 만지기, 물 마시기, 키보드 만지기, 안경 만지기 행동이 포함되어 있다.

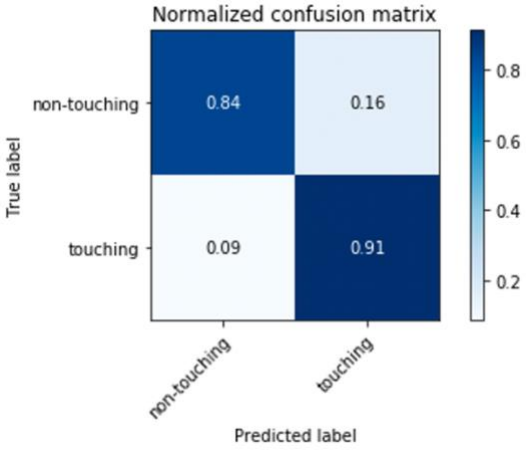

그림 7. I3D 기반 신경망의 선형 평가 결과

MobileNet 기반의 신경망은 테스트 세트에 대하여 90%의 정확도를 보이는 것을 확인하였다. MobileNet 기반의 방법은 Intel i7-6700 CPU 3.40GHz CPU 에서 0.07~0.09 초의 처리속도를 보이는 것을 확인하였고, I3D 기반의 방법은 같은 CPU 에서 1.4~1.5 초의 처리속도를 보이는 것을 확인하였다. 그림 8 은 MobileNet 을 사용하여 얼굴을 만지는 행동과 그렇지 않은 행동을 구분하는 실험의 선형 평가 결과를 나타낸다.

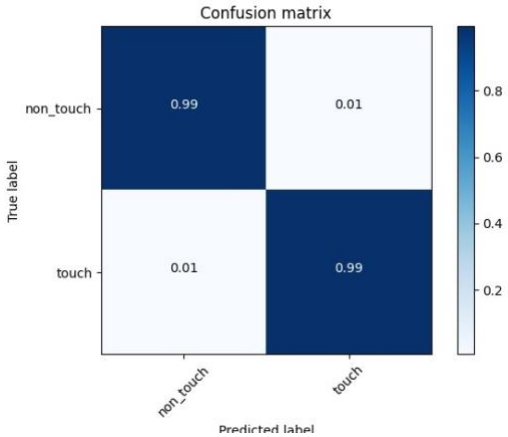

그림 8. MobileNet 기반 신경망의 선형 평가 결과

6. 결론

본 논문에서는 얼굴을 만지는 행동과 얼굴을 만지지 않는 행동의 구분을 위하여 11 개의 클래스를 가진 학습 데이터를 구축하였다. 또한, 행동 인식을 하기 위한 신경망의 구조로 I3D 기반의 방법과 MobileNet 기반의 방법 등 총 2 가지를 제안하였으며 두 방법 모두 90% 이상의 정확도로 얼굴 만지는 행동을 검출하는 것을 확인하였다. 테스트 세트에 대해서는 90% 이상의 정확도를 선보였으나, 데이터의 성격이 많이 다를 경우에는 정확도가 떨어지는 문제점을 파악하여, 향후 보다 강인한 신경망을 설계할 예정이다.

참고문헌

[1] C. f. D. C. a. P. (CDC). "Coronavirus Disease 2019 (COVID-19)." https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/prevention.html?CDC_AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Fcoronavirus%2F2019-ncov%2Fprepare%2Fprevention.html (accessed May 25th, 2020).

[2] W. H. O. (WHO). "Protecting yourself and others from the spread COVID-19." https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public (accessed May 25th, 2020).

[3] W. H. O. (WHO). "Clean hands protect against infection." https://www.who.int/gpsc/clean_hands_protection/en/ (accessed May 25th, 2020).

[4] M.-S. Lee, S. J. Hong, and Y.-T. Kim, "Handwashing with soap and national handwashing projects in Korea: focus on the National Handwashing Survey, 2006-2014," *Epidemiology and health,* vol. 37, 2015.

[5] Y. L. A. Kwok, J. Gralton, and M.-L. McLaws, "Face touching: A frequent habit that has implications for hand hygiene," *American journal of infection control,* vol. 43, no. 2, pp. 112-114, 2015.

[6] M. Nicas and D. Best, "A study quantifying the hand-to-face contact rate and its potential application to predicting respiratory tract infection," *Journal of occupational and environmental hygiene,* vol. 5, no. 6, pp. 347-352, 2008.

[7] J. Carreira and A. Zisserman, *Quo Vadis, Action Recognition? A New Model and the Kinetics Dataset*. 2017, pp. 4724-4733.

[8] A. Howard *et al.*, *Searching for MobileNetV3*. 2019, pp. 1314-1324.

[9] S. Hochreiter and J. Schmidhuber, "Long Short-term Memory," *Neural computation,* vol. 9, pp. 1735-80, 12/01 1997, doi: 10.1162/neco.1997.9.8.1735.

[10] A. Ullah, J. Ahmad, K. Muhammad, M. Sajjad, and S. Baik, "Action Recognition in Video Sequences using Deep Bi-directional LSTM with CNN Features," *IEEE Access,* vol. PP, pp. 1-1, 11/28 2017, doi: 10.1109/ACCESS.2017.2778011.