

출입 통제에 활용 가능한 딥러닝 기반 마스크 착용 판별

이승호
한국기술교육대학교 융합학과

Deep learning based face mask recognition for access control

Seung Ho Lee

Department of Future Technologies, Korea University of Technology and Education

요약 전 세계적으로 유행하며 수많은 확진자와 사망자를 발생시킨 코로나바이러스-19(COVID-19)는 일상에서 사람 간 전염이 가능하여 국민들을 불안과 공포에 떨게 하고 있다. 감염을 최소화하기 위해서는 건물 출입시 마스크 착용이 필수적이지만 일부 사람들은 여전히 마스크 없이 얼굴을 노출시킨 채 건물에 출입하고 있다. 본 논문에서는 효율적인 출입 통제를 위해 얼굴에 마스크를 착용했는지 여부를 자동으로 판별하는 방법을 제안한다. 제안 방법은 양쪽 눈 영역을 검출하고 눈 위치를 참조하여 마스크 착용 영역(양쪽 눈 아래 얼굴 영역)을 예측한다. 이 때 마스크 착용 영역을 보다 정확히 예측하기 위해 양쪽 눈 위치가 수평이 되도록 얼굴 영역을 회전하여 정렬한다. 정렬된 얼굴 영역에서 추출된 마스크 착용 영역은 이미지 분석에 특화된 딥러닝 기법인 CNN(Convolutional neural network)을 통해 마스크 착용 여부(착용 또는 미착용)를 최종 판별한다. 총 186장의 테스트 이미지에 대해 실험한 결과, 98.4%의 판별 정확도를 보였다.

Abstract Coronavirus disease 2019 (COVID-19) was identified in December 2019 in China and has spread globally, resulting in an ongoing pandemic. Because COVID-19 is spread mainly from person to person, every person is required to wear a facemask in public. On the other hand, many people are still not wearing facemasks despite official advice. This paper proposes a method to predict whether a human subject is wearing a facemask or not. In the proposed method, two eye regions are detected, and the mask region (i.e., face regions below two eyes) is predicted and extracted based on the two eye locations. For more accurate extraction of the mask region, the facial region was aligned by rotating it such that the line connecting the two eye centers was horizontal. The mask region extracted from the aligned face was fed into a convolutional neural network (CNN), producing the classification result (with or without a mask). The experimental result on 186 test images showed that the proposed method achieves a very high accuracy of 98.4%.

Keywords : Face Mask Detection, Deep Learning, Convolutional Neural Network, Eye Detection, OpenCV

1. 서론

2019년에 중국에서 처음 발발한 코로나바이러스-19(COVID-19)는 전 세계에서 확진자가 폭증하고 있으며 수많은 사망자를 발생시켰다. 코로나바이러스의 전파경로는 비말과 콧물, 가래 등의 호흡기 분비물로 알려

져 있어 마스크 착용이 감염 예방에 매우 중요하다. 하지만 국가 차원의 마스크 착용 권고에도 불구하고 여전히 일부 사람들이 마스크를 착용하지 않은 채 건물에 출입하는 일이 발생하고 있어 많은 감염 위험이 우려되고 있다. 건물 출입 시 마스크 착용을 검사하는 곳이 많지만 수많은 사람들을 일일이 눈으로 확인하는 것은 많은 인

이 논문은 2019년도 한국기술교육대학교 교수 교육연구진흥과제 지원에 의하여 연구되었음.

*Corresponding Author : Seung Ho Lee(Korea University of Technology and Education)

email: leesh903@koreatech.ac.kr

Received May 22, 2020

Revised July 7, 2020

Accepted August 7, 2020

Published August 31, 2020

력과 비용을 발생시킬 수 있으므로 비효율적이다. 따라서 효율적인 출입 통제를 위해 자동으로 마스크 착용 여부를 판별하는 방법 개발이 시급하다. 마스크 착용 여부를 판별하려면 먼저 얼굴 영역을 검출해야 한다. 얼굴 영역을 검출하는 방법은 비올라 존스 알고리즘[1]을 포함하여 수많은 알고리즘들이 개발되어 왔다. 하지만 마스크를 착용한 얼굴의 경우 얼굴 대부분의 영역이 가려지므로 기존의 방법들을 적용해서는 검출이 어렵다. 이러한 이유로 정부서울청사는 출입 통제를 위한 안면인식 시스템을 일시 정지하고 청사 직원들이 마스크를 착용한 채 입장할 수 있도록 조치를 취하였다[2].

본 논문에서 제안하는 마스크 착용 판별 방법은 마스크 착용/미착용 얼굴 모두를 검출할 수 있도록 양쪽 눈 영역 검출에 기반한다. 양쪽 눈 영역을 검출하고 눈 위치를 참조하여 얼굴 영역을 추출한다. 그리고 마스크 착용 영역(양쪽 눈 아래 얼굴 영역)을 보다 정확히 예측하기 위해 양쪽 눈높이가 동일해지도록 얼굴 영역을 회전하여 정렬한다. 정렬된 얼굴 영역에서 추출된 마스크 착용 영역은 이미지 분석에 특화된 딥러닝 기법인 CNN(convolutional neural network)[3, 10]을 통해 마스크 착용 여부(착용 또는 미착용)를 최종적으로 판별한다.

다양한 각도로 촬영된 100장의 마스크 착용자 이미지들 중 93장에서 정렬된 얼굴 영역 및 마스크 착용 영역이 성공적으로 검출된 것을 확인하였다. 상기 과정에서 검출된 (마스크 착용자에 대한) 마스크 착용 영역 이미지 93장에 마스크 미착용자에 대한 마스크 착용 영역 이미지 93장을 추가하여 총 186장의 이미지에 대해 k겹 교차 검증(k-cross validation)을 적용하여 마스크 착용 여부 판별(착용 또는 미착용) 테스트를 수행하였고, 실험 결과 98.4%의 판별 정확도를 보였다.

본 논문의 나머지 구성은 다음과 같다. 2장에서는 마스크 착용 영역 추출 과정과 마스크 착용 여부 판별과정으로 구성된 제안 방법을 자세하게 설명하고, 3장에서는 두 가지 과정에 대한 실험결과를 보인다. 4장에서 제안 방법의 유용성을 분석하고 4장에서 결론을 맺는다.

2. 제안 방법

제안 방법은 사람 키 높이 정도에 설치된 카메라를 응시하면서 한명씩 차례로 입장하는 출입 통제 환경에 적합하다. Fig. 1은 제안하는 마스크 착용 판별 방법의 전체 흐름도를 나타낸다. 카메라에 촬영된 출입자의 사진이

제안 방법의 입력이다. 본 논문에서는 이를 입력 이미지(input image)라고 부른다. 입력 이미지에 눈 영역 검출(Eye Region Detection)을 적용하여 양쪽 눈 좌표를 출력한다. 양쪽 눈이 수평이 되도록 얼굴 정렬(Face Alignment)을 수행하여 정렬된 얼굴 영역을 출력한다. 다음으로 정렬된 얼굴 영역에서 상대적인 위치로 마스크 착용 영역을 이미지로 출력한다. CNN 분류기(CNN Classification)가 마스크 착용 영역 이미지를 분석하여 마스크 착용 여부를 판별한다. 2.1절과 2.2절에서는 각각 마스크 착용 영역 추출과 마스크 착용 여부 판별에 대해 자세히 설명한다.

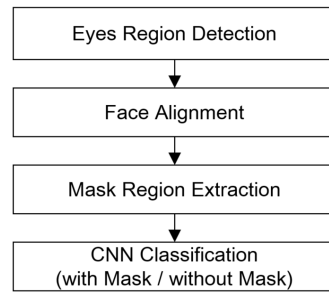


Fig. 1. Block diagram of proposed method.

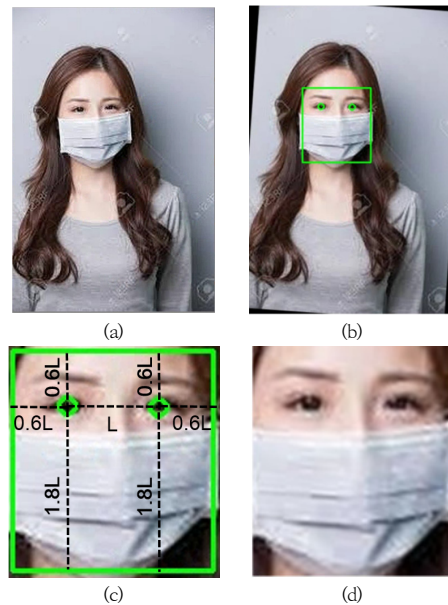


Fig. 2. An example of face alignment. (a) Input image. (b) Input image rotation based on two eyes centers. (c) Determination of aligned face region. (d) Cropped image of the aligned face region.

2.1 마스크 착용 영역 추출

Fig. 2(a)는 입력 이미지의 예를 보여준다. 제안 방법은 입력 이미지에서 양쪽 눈을 검출하기 위해 Haar cascade 분류기[4]를 사용한다. Haar cascade 분류기는 얼굴 등 객체 검출을 위해 [1]에서 제안된 효과적인 기법이다. 본 논문에서는 대표적인 영상처리 라이브러리인 OpenCV[5]에서 제공하는 사전 학습된 눈 검출용 Haar cascade 분류기 모델을 사용한다. 눈 검출용 분류기 모델의 파일명은 haarcascade_eye.xml과 haarcascade_eye_tree_eyeglasses.xml이다. 먼저 눈 검출 기본 모델인 haarcascade_eye.xml을 사용하여 눈 검출을 수행하고, 만약 검출 개수가 2개가 아닌 경우(양쪽 눈 검출 실패를 의미)에는, 안경 쓴 얼굴에 특화된 눈 검출 모델인 haarcascade_eye_tree_eyeglasses.xml을 사용하여 다시 한 번 눈 검출을 반복 수행한다. haarcascade_eye_tree_eyeglasses.xml 모델을 사용해도 검출 개수 2개를 충족하지 못한 경우에는 입력 이미지의 사이즈가 작은 것이 미검출의 원인일 수도 있다. 따라서 입력 이미지 사이즈를 가로 세로 2배로 키워서 위의 과정을 다시 한 번 수행한다.

검출 개수 2개를 충족한 경우에는 양쪽 눈이 검출된 것으로 간주하고 얼굴 영역 정렬을 수행한다. Fig. 2(b)와 같이 양쪽 눈높이가 같아지도록 입력 이미지를 회전 시킨다. 다음으로 양쪽 눈 사이의 거리(L)를 기준으로 삼아 상대적인 크기로 얼굴 영역을 결정한다(Fig. 2(c)). Fig. 2(d)는 정렬된 얼굴 영역을 잘라내어 획득한 얼굴 이미지를 보여준다.

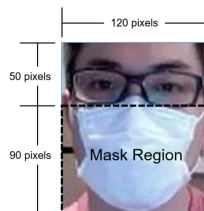


Fig. 3. Illustration of mask region extraction from an aligned face region image.

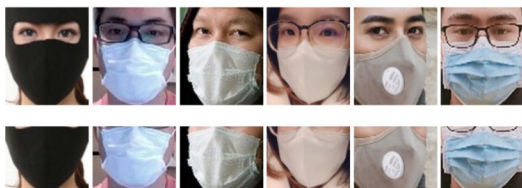


Fig. 4. Images of aligned faces with masks (upper row) and images of corresponding mask regions (lower row).

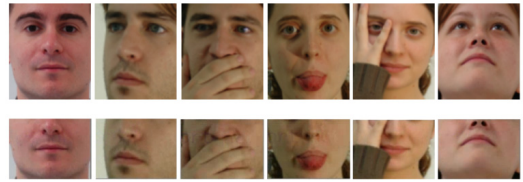


Fig. 5. Images of aligned faces without masks (upper row) and images of corresponding mask regions (lower row).

정렬된 얼굴 영역 이미지는 마스크 착용 영역 추출을 위해 고정된 픽셀 크기(세로 140 픽셀, 가로 120 픽셀)로 리사이징 된다. Fig. 3에서 점선으로 표시한 것처럼, 눈 아래 세로 90 픽셀, 가로 120 픽셀의 영역이 마스크 착용 영역이 된다. Fig. 4와 Fig. 5는 정렬 얼굴 영역(윗줄)으로부터 추출된 마스크 착용 영역(아랫줄)에 대한 예시 이미지들을 나타낸다.

2.2 CNN 기반 마스크 착용 여부 판별

CNN은 이미지 인식에 특화된 딥러닝 기법이다. 컨볼루션 계층(convolution layer)에서는 이미지를 분별해 낼 수 있는 시각적 특징을 스스로 학습한다. 구체적으로, 학습 단계(training stage)에서는 학습용 입력 (예 : 마스크 착용 영역) 및 정답 클래스(예 : 마스크 착용 여부)를 입력 받아 공간적 필터들에 대한 가중치 값(커널 값)들을 결정한다. 예측 단계(prediction stage)에서는, 학습 단계에서 학습된 공간적 필터들을 테스트용 입력 이미지에 적용하여 필터링 결과에 해당하는 이차원의 특징 맵(feature map)들을 획득한다. 이 특징맵 데이터는 완전 연결 계층(fully connected layer)로 전달되고 분석되어 최종적으로 클래스 예측값이 출력된다.

Fig. 6은 본 논문에서 사용된 CNN 모델 구조를 나타낸다. 이 CNN 모델은 마스크 착용 영역에 대한 컬러 이미지를 입력 받는다. 첫 번째 컨볼루션 계층은 5x5 픽셀 사이즈의 필터 32개로 구성된다. 두 번째 컨볼루션 계층은 5x5 픽셀 사이즈의 필터 128개로 구성된다.

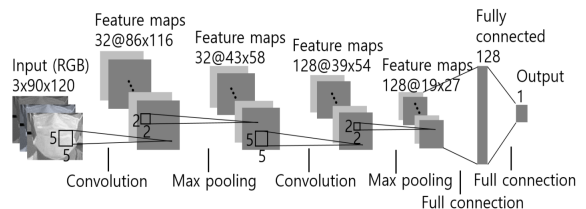


Fig. 6. Convolutional neural network model used in this paper.

이 두 가지 컨볼루션 계층은 필터링을 통해 각각 32개, 128개의 특징맵들을 출력한다. 또한 각 컨볼루션 계층 다음에는 특징맵의 공간적 크기를 줄이고, 신경망의 가중치 학습량을 감소시키기 위해 2x2 픽셀 크기의 맥스 풀링 계층(max pooling layer)이 적용된다. 전형적인 CNN 모델과 유사하게, 완전 연결 계층이 신경망의 마지막 부분에 배치된다. 두 번째 맥스 풀링 계층을 통해 획득한 128개 특징맵의 이차원 데이터는 특징맵의 픽셀값들이 연결된 일차원 데이터로 변환되어 128개의 노드로 구성된 완전 연결 계층에 입력된다. 출력층(output layer)은 한 개의 노드로 구성되고 0-1 사이의 값을 출력하여 최종적으로 마스크 착용, 미착용을 판별한다.

3. 실험결과

제안 방법은 크게 마스크 착용 영역 추출 과정과 CNN 기반 마스크 착용 여부 판별 과정으로 구성된다. 실험결과도 이 두 가지 과정으로 구분하여 자세히 설명한다.

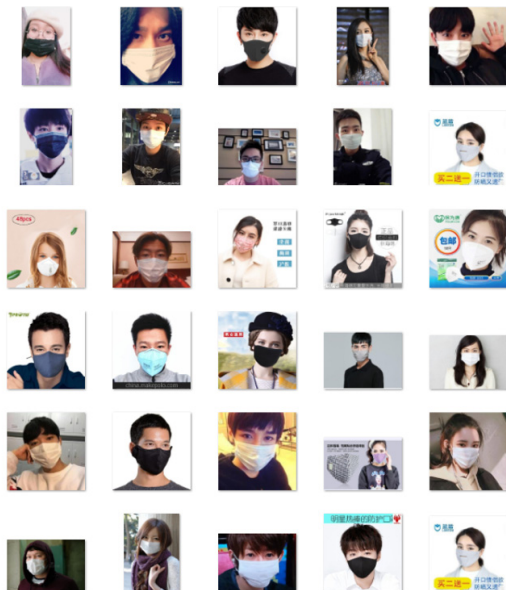


Fig. 7. Examples of input images in which the mask regions are successfully extracted (see Fig. 8).



Fig. 8. Examples of mask region images extracted from the input images in Fig. 7.



Fig. 9. Examples of input images in which the mask regions are not successfully extracted.

3.1 마스크 착용 영역 추출 실험결과

입력 이미지로부터 얼굴의 마스크 착용 영역을 정확하게 추출하는지를 검증하기 위해 Real-World Masked Face Dataset(RMFD)[6]에서 100장의 컬러 이미지를 수집하였다([7]에서 다운로드 가능). 수집된 이미지들은 사람 한 명을 포함하며 다양한 얼굴 각도를 포함한다. 이미지 사이즈는 세로 256 픽셀 x 가로 142 픽셀부터 1000 픽셀 x 667 픽셀 크기까지 다양하다.

입력 이미지 100장 중 93장의 이미지에서 마스크 착용 영역이 성공적으로 추출되었다. Fig. 7과 Fig. 8은 입력 이미지들 중 일부와 해당 입력 이미지들에서 추출된 마스크 영역 이미지들을 보여준다. Fig. 9는 마스크 착용 영역 추출에 실패한 7장의 입력 이미지들을 보여준다. 이들은 눈을 너무 작게 뜨거나 마스크를 너무 눈 가까이에 착용하여 눈 영역 검출에 실패한 이미지들이다.

Table 1. Results for k-fold validation (k=5).

Dataset	Accuracy (%)
1st dataset	97.37
2nd dataset	97.37
3rddataset	100.00
4th dataset	97.22
5th dataset	100.00
Average	98.39

i7-8700K CPU, 48GB 램 PC에서 마스크 착용 영역 검출에 대한 계산시간을 측정 분석하였다. 총 100장의 이미지를 처리하는데 7.2초가 소요되어 장당 평균 0.072 초 수준의 처리속도를 보였다. 가장 사이즈가 작은 이미지(256 픽셀 x 147 픽셀)의 경우 장당 0.04초, 가장 사이즈가 큰 이미지(1000 픽셀 x 667 픽셀)의 경우 장당 0.36초가 소요되었다. 100장의 이미지 중 해상도가 너무 낮은 이미지의 경우, 눈 검출 알고리즘이 요구하는 최소 픽셀 사이즈를 충족하지 못하여 2.1절에서 설명된 '이미지 확대 후 재검출'이 적용된 경우도 있었다. 하지만 특정 사이즈 이상(실험에 사용된 RMFD 이미지의 경우 300 픽셀 x 300 픽셀)에서는 이미지 확대 없이도 양쪽 눈 및 마스크 착용 영역이 모두 성공적으로 검출된 것을 확인하였다. 장당 처리시간을 일정 시간 이내로 보장하기 위해 '이미지 확대 후 재검출'을 입력 이미지의 사이즈에 따라 선택적으로 미적용 설정 할 수 있다. 예를 들어, 본 논문에서 사용된 RMFD 이미지들의 특성을 고려한다면, 가로, 세로가 둘 다 300 픽셀 이상이면 '이미지 확대 후 재검출' 기능을 미적용 할 수 있다.

제안 방법의 효용성을 입증하기 위해, 얼굴검출에서 가장 널리 사용되고 있는 방법 중 하나인 비올라 존스 얼굴검출 알고리즘[1]으로 동일한 실험 데이터셋(마스크 착용 얼굴 이미지 100장)에 대해 얼굴 검출을 수행하였다. 눈 검출과 마찬가지로 파이썬용 OpenCV에서 제공하는 사전 학습된 얼굴 검출용 분류기 모델을 사용하였다[9]. 얼굴 검출용 분류기 모델의 파일명은 haarcascade_frontalface_default.xml이다. 비올라 존스 얼굴검출 알고리즘에서 본 실험 데이터셋에 가장 높은 성능을 보이는 scaleFactor=1.1, minNeighbors=4로 파라미터 값을 설정하였다. 여기서 scaleFactor는 다양한 크기의 얼굴을 검출하기 위해 입력 이미지 사이즈를 줄여나가는 비율이다[9]. 그리고 minNeighbors는 다양한 이미지 사이즈에서 검출된 후보 얼굴 영역들이 동일 위치에 그룹핑 되었을 때 얼굴 영역으로 판정하기 위한 최소 후보

얼굴 영역의 개수이다[9]. 얼굴검출 결과는 다음과 같다. 총 100장 중 55장의 이미지에서 얼굴이 정확하게 검출되었다. 나머지 45장의 이미지는 미검출(아무 것도 검출 안 됨) 또는 오검출(얼굴이 아닌 대상 또는 영역을 검출)에 해당한다. 제안 방법에 비해 현저하게 얼굴검출 성능이 낮은 건 비올라 존스 얼굴검출 알고리즘에서 중요하게 고려되는 코나 입 영역에 대한 특성이 마스크에 의해 가려져 얼굴검출 분류기의 분별력을 저해시키기 때문인 것으로 보인다.

3.2 CNN 기반 마스크 착용 여부 판별 실험결과

두 번째 실험의 목적은 마스크 착용 얼굴과 미착용 얼굴을 정확하게 구별할 수 있는지를 검증하는 것이다. 3.1 절에서 검출된 93장의 마스크 착용 영역 이미지들에 마스크 미착용자 6명에 대한 마스크 착용 영역 이미지들 93장을 추가하여 총 186장의 이미지를 확보하였다. 마스크 미착용자들에 대한 입력 이미지들은 무료로 공개된 GTAV Face Database[8]에서 2.1절의 마스크 착용 영역 추출 방법을 적용하여 획득하였다. Fig. 5는 마스크 미착용자에 대한 정렬 얼굴 영역 및 마스크 착용 영역들의 예시를 보여준다. 다양한 얼굴 각도를 포함하며 입이나 눈 등 얼굴의 일부 영역이 손으로 가려져 있는 경우도 포함하는 것을 확인할 수 있다.

실험에 사용된 CNN 모델(Fig. 6)은 파이썬과 대표적인 딥러닝 라이브러리인 케라스(Keras)로 구현하였다. 컨볼루션 계층들에 대한 활성화 함수로 렐루(ReLU)[10, 11]를 사용하였고, 출력층에 대한 활성화 함수로 소프트맥스(Softmax)[12]를 사용하였다. 손실함수로는 평균 제곱 오차(Mean Squared Error)를 사용하였고, 에포크(Epoch)[13]와 배치사이즈(Batch size)[13]는 각각 30과 10으로 설정하였다.

정확도를 측정하기 위해 k겹 교차 검증을 적용하였다. 본 논문에서는 k=5로 설정하였다. 5개로 구성된 데이터셋에 대한 개별 정확도와 이들을 평균한 전체 정확도를 표 1에 나타냈다. 98.4%의 높은 전체 정확도를 것을 확인할 수 있다. CNN 기반 마스크 착용 여부 판별에 소요되는 시간은 장당 평균 약 0.015초이다.

4. 결론

본 논문에서는 출입 통제를 위해 얼굴에 마스크를 착용했는지 여부를 자동으로 판별하는 방법을 제안하였다.

눈 검출을 통해 검출 및 정렬된 얼굴 영역에서 마스크 착용 영역을 추출한 뒤, 이미지에 특화된 딥러닝 기반인 CNN를 이용하여 마스크 착용 여부(착용 또는 미착용)를 최종 판별한다. 마스크 착용 영역 검출과 마스크 착용 여부 판별 효용성에 대해 각각 실험을 통해 검증하였다. 본 논문에서 제안한 방법은 코로나바이러스와 같은 전염병이 유행인 시기에 적은 인력으로 마스크 미착용자를 색출하는데 유용하게 활용될 수 있을 것으로 기대한다.

본 논문에서 제안하는 방법은 사람 키 높이 정도에 설치된 카메라를 응시하면서 한명씩 차례로 입장하는 출입 통제 환경에 적합하다. 향후 연구계획으로 다수 명의 사람에 대해 한꺼번에 마스크 착용 여부를 판별할 수 있는 방법에 대한 연구를 수행할 것이다.

References

- [1] P. Viola, M. J. Jones, "Robust-real time face detection," International Journal of Computer Vision, Vol.57, No.2, pp.137-154, May 2004.
DOI: <https://doi.org/10.1023/B:VISI.0000013087.49260.fb>
- [2] Newspim.com [cited 2020 July 6], Available From: <http://www.newspim.com/news/view/20200311000350> (accessed Jul. 6, 2020)
- [3] Y. Lecun, L. Bottou, Y. Bengio, P. Haffner, "Gradient based learning applied to document recognition," Proceedings of the IEEE, Vol.86, No.11, pp.2278-2324, Nov. 1998.
DOI: <https://doi.org/10.1109/5.726791>
- [4] Haar-cascade classifier [cited 2020 July 6], Available From: https://docs.opencv.org/3.4/db/d28/tutorial_cascade_classifier.html (accessed Jul. 6, 2020)
- [5] OpenCV-Python [cited 2020 July 6], Available From: <https://docs.opencv.org/4.3.0/> (accessed Jul. 6, 2020)
- [6] Z. Wang, G. Wang, B. Huang, Z. Xiong, Q. Hong, H. Wu, P. Yi, K. Jiang, N. Wang, Y. Pei, et al., "Masked face recognition dataset and application," arXiv preprint, arXiv:2003.09093, Mar. 2020.
- [7] RMFD Download [cited 2020 July 6], Available From: <https://github.com/X-zhangyang/Real-World-Masked-Face-Dataset> (accessed Jul. 6, 2020)
- [8] GTAV Face Database [cited 2020 July 6], Available From: <https://francesctarres.wordpress.com/gtav-face-database/> (accessed Jul. 6, 2020)
- [9] Viola-Jones face detector in Python [cited 2020 July 6], Available From: https://docs.opencv.org/2.4/modules/objdetect/doc/cascade_classification.html (accessed Jul. 6, 2020)

- [10] A. Krizhevsky, I. Sutskever, G. E. Hinton, "Imagenet classification with deep convolutional neural networks," In Advances in Neural Information Processing Systems, pp.1097-1105, 2012.
DOI: <https://doi.org/10.1145/3065386>
- [11] B. Xu, N. Wang, T. Chen, M. Li, "Empirical evaluation of rectified activations in convolutional network," arXiv preprint arXiv:1505.00853, Nov. 2015.
- [12] Wikipedia- Softmax function [cited 2020 July 6], Available From: https://en.wikipedia.org/wiki/Softmax_function (accessed Jul. 6, 2020)
- [13] I. Kandel, M. Castelli, "The effect of batch size on the generalizability of the convolutional neural networks on a histopathology dataset," ICT Express (available online) May 2020.
DOI: <https://doi.org/10.1016/j.ict.2020.04.010>

이 승 호(Seung Ho Lee)

[정회원]



- 2011년 2월 : KAIST 전기및전자공학과 (공학석사)
- 2015년 8월 : KAIST 전기및전자공학과 (공학박사)
- 2015년 9월 ~ 2018년 2월 : 국방과학연구소(ADD) 선임연구원
- 2018년 3월 ~ 현재 : 한국기술교육대학교 조교수

<관심분야>

머신러닝/딥러닝, 컴퓨터비전, 패턴인식