

컨볼루션 오토인코더를 이용한 마스크 착용 얼굴 이미지 생성

이승호*

Generation of Masked Face Image Using Deep Convolutional Autoencoder

Seung Ho Lee*

*Assistant Professor, Department of Future Technology, Korea University of Technology and Education, Cheonan, 31079 Korea

요약

코로나19 팬데믹으로 인해 마스크 착용이 일상화되면서 마스크 착용 얼굴을 식별하는 얼굴인식 연구에 대한 중요도가 높아지고 있다. 안정된 얼굴인식 성능을 위해서는 인식 대상에 대한 풍부한 학습용 이미지 확보가 필요하지만 인물 별로 마스크 착용 얼굴 이미지를 다량 확보하는 것은 쉽지 않다. 본 논문에서는 마스크 미착용 얼굴 이미지에 가상의 마스크 패턴을 합성하는 새로운 방법을 제안한다. 제안 방법은 동일 인물에 대해 마스크 미착용 얼굴 이미지와 마스크 착용 얼굴 이미지를 쌍으로 컨볼루션 오토인코더에 입력하여 얼굴과 마스크의 기하학적 관계를 학습한다. 학습이 완료된 컨볼루션 오토인코더는 학습에 사용되지 않은 새로운 마스크 미착용 얼굴 이미지에 가상의 마스크 패턴을 자연스러운 형태로 합성해준다. 제안 방법은 고속으로 대량의 마스크 착용 얼굴 이미지를 생성할 수 있으며, 얼굴 특징점 추출에 기반하는 마스크 합성 방법에 비해 실용적이다.

ABSTRACT

Researches of face recognition on masked faces have been increasingly important due to the COVID-19 pandemic. To realize a stable and practical recognition performance, large amount of facial image data should be acquired for the purpose of training. However, it is difficult for the researchers to obtain masked face images for each human subject. This paper proposes a novel method to synthesize a face image and a virtual mask pattern. In this method, a pair of masked face image and unmasked face image, that are from a single human subject, is fed into a convolutional autoencoder as training data. This allows learning the geometric relationship between face and mask. In the inference step, for a unseen face image, the learned convolutional autoencoder generates a synthetic face image with a mask pattern. The proposed method is able to rapidly generate realistic masked face images. Also, it could be practical when compared to methods which rely on facial feature point detection.

키워드 : 얼굴 마스크 합성, 얼굴 마스크 생성, 컨볼루션 오토인코더, 얼굴인식

Keywords : Facial Mask Synthesis, Facial Mask Generation, Convolutional Autoencoder, Face Recognition

Received 10 April 2022, Revised 15 April 2022, Accepted 29 April 2022

* Corresponding Author Seung Ho Lee (E-mail:leesh903@koreatech.ac.kr, Tel:+82-41-560-1116)

Assistant Professor, Department of Future Technology, Korea University of Technology and Education, Cheonan, 31079 Korea

Open Access <http://doi.org/10.6109/jkiice.2022.26.8.1136>

print ISSN: 2234-4772 online ISSN: 2288-4165

© This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
Copyright © The Korea Institute of Information and Communication Engineering.

I. 서론

코로나19 팬데믹으로 인해 마스크 착용이 일상화되면서 출입 통제나 스마트 기기의 사용자 인증 등 마스크 착용 얼굴을 식별하는 얼굴인식 연구에 대한 요구가 높아지고 있다[1]. 마스크 착용 얼굴에 대해 안정적인 인식 성능을 얻기 위해서는 학습용 데이터로 사용될 마스크 착용 얼굴 이미지들을 사전에 확보해야 하며, 이 데이터에 다양한 얼굴 특성(정면/비정면 얼굴, 무표정/표정 얼굴 등)이 포함되어야 한다[2]. 하지만 인물 별로 다수의 마스크 착용 얼굴 이미지를 확보하기는 현실적으로 어려워, 최근 얼굴 이미지에 가상의 마스크 이미지를 합성하는 연구가 수행되고 있다[1, 3]. [3]에서는 마스크 착용 시 얼굴 특징점 추출 정확도가 저하되는 문제를 해결하기 위해 마스크 착용 얼굴 합성을 통해 얼굴 특징점 추출용 학습 데이터를 확보하고자 하였다. 자연스러운 합성 결과를 위해, 입력 이미지에 대한 얼굴 특징점 추출 데이터로 얼굴 자세 추정(pose estimation)을 수행한 뒤 자세에 맞는 마스크 패턴을 적용하였다[3]. 하지만 이 방법에서는 얼굴 자세를 정면과 측면으로만 정의하여 매우 다양한 얼굴 자세가 존재하는 실제 응용에는 적합하지 못하다. [1]에서는 얼굴 특징 추출 데이터에 곧바로 마스크 합성이 적용된다. 영상처리 및 컴퓨터비전 라이브러리인 OpenCV[4]를 통해 지정된 특징점들을 다각형 선으로 잇고 그 도형을 색으로 채워 마스크를 완성하였다. 하지만 그림 5(가운데 열)에서 볼 수 있듯이 각진 모양으로 인해 합성된 마스크가 다소 부자연스럽다는 한계점이 있다.

본 논문에서는 위 방법들의 한계점을 극복하기 위해 새로운 마스크 착용 얼굴 이미지 생성 방법을 제안한다. 제안 방법은 컨볼루션 오토인코더[5]에 동일 인물의 마스크 미착용 얼굴 이미지와 마스크 착용 얼굴 이미지를 쌍으로 학습 데이터로 입력하여 얼굴과 마스크의 기하학적 관계를 학습한다. 따라서 얼굴의 각도, 눈, 코, 입의 위치 등에 적합한 형태로 마스크의 합성 크기, 위치, 모양이 정해진다. 학습이 완료된 컨볼루션 오토인코더는 학습에 사용되지 않은 새로운 마스크 미착용 얼굴 이미지에 가상의 마스크 패턴을 자연스러운 형태로 합성해 준다. 제안 방법은 마스크 합성을 위해 얼굴 특징점 추출이나 자세 추정을 필요로 하지 않으므로 실용적이다.

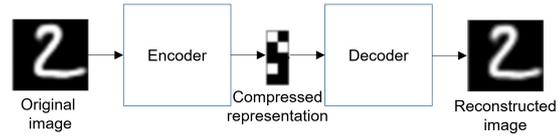


Fig. 1 An example of autoencoder used for image compression.

본 논문의 나머지 구성은 다음과 같다. 2장에서는 제안 방법의 이해를 돕기 위한 관련 기술들을 간략히 요약한다. 3장에서는 제안하는 마스크 착용 얼굴 이미지 생성 방법에 대해 설명한다. 4장에서는 실험결과를 바탕으로 장점 및 한계점에 대해 설명하고 5장에서 결론을 맺는다.

II. 관련 기술

본 장에서는 제안 방법 이해를 돕기 위해 오토인코더와 얼굴인식의 개념에 대해 간략히 설명한다.

2.1. 오토인코더(autoencoder)

오토인코더는 입력 데이터(=원본 데이터)를 압축 후 복원하여 출력했을 때 압축 전 입력 데이터와 출력(=복원 데이터)이 서로 최대한 유사해지도록 학습하는 신경망의 일종이다. 그림 1과 같이 오토인코더는 데이터를 압축하는 인코더와 복원하는 디코더로 구성되며, 신경망의 노드들은 완전 연결되어(fully connected) 있다. 인코더는 원본 데이터(그림 1의 Original image)를 더 낮은 차원의 공간으로 맵핑하며, 디코더는 낮은 차원의 공간에 투영된 압축 데이터로부터 원래의 데이터로 복원한다(그림 1의 Reconstructed image). 오토인코더의 응용 분야는 매우 다양하다. 주로 데이터 압축, 노이즈 제거, 원본 데이터 변형이나 합성을 통한 새로운 데이터 생성 등에 활용되고 있다.

2.2. 얼굴인식(face recognition)

얼굴인식은 카메라를 통해 입력 이미지가 취득되었을 때 입력 이미지 속 얼굴과 데이터베이스 내 저장된 학습용 얼굴 이미지와의 비교를 통해 어떤 인물인지 식별(identification)하는 기술이다[6, 7]. 얼굴인식은 스마트 감시시스템, 출입 통제 시스템, 스마트 기기의 사용자 인증 등 매우 다양한 용도로 활용되고 있다. 현재 활발

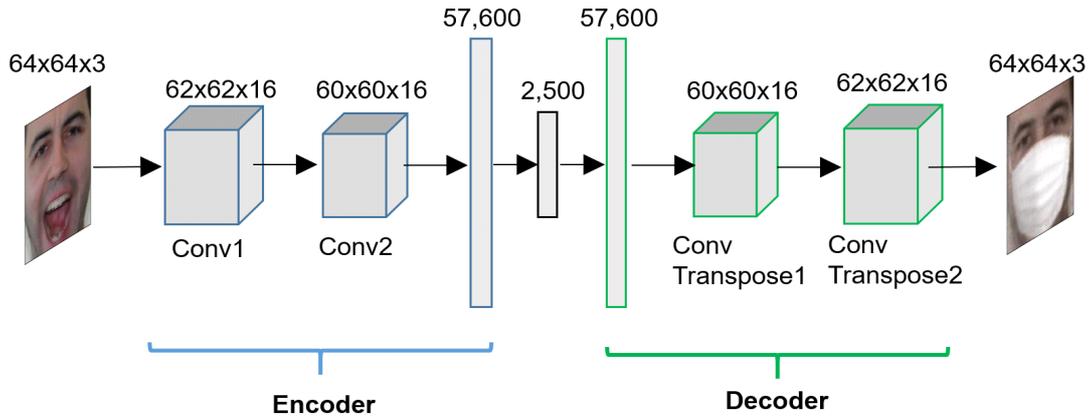


Fig. 2 Convolutional autoencoder network used in this paper to generate a synthetic masked face image.

히 연구되고 있는 딥러닝(예 : 컨볼루션 신경망[8, 9]) 기반 얼굴인식 모델은 다양한 환경에서 획득한 대용량의 얼굴 데이터를 통해 학습하는 것이 일반적이며, 특징추출 및 분류 네트워크를 동시에 최적화 함으로써 높은 정확도를 보이고 있다[2]. 하지만 아무리 많은 수의 학습용 얼굴 이미지를 확보하더라도 테스트용 얼굴 이미지와 특성이 크게 다른 경우에는 불일치(mismatch)로 인해 성능이 저하될 수 있다[2, 6]. 마스크 착용 얼굴을 인식하는 경우 인물 별로 충분한 수의 마스크 착용 얼굴 이미지를 확보하기 어려우므로 이 문제를 극복하기 위한 연구가 반드시 필요하다.

III. 마스크 착용 얼굴 이미지 생성 방법

제안 방법은 이미지 분석에 특화된 컨볼루션 신경망과 데이터 생성에 특화된 오토인코더의 장점이 결합된 컨볼루션 오토인코더를 사용한다. 컨볼루션 층은 얼굴과 마스크 사이의 위치, 크기, 형태 등 기하학적인 특징을 2차원적으로 분석하는데 효과적이다. 그림 2는 본 논문에서 사용된 컨볼루션 오토인코더의 신경망 구조를 보여준다. 본 신경망은 인코더와 디코더로 구성되며 입력 이미지의 크기는 64x64x3(64x64 픽셀 크기의 RGB 컬러 얼굴 이미지)이다. 인코더의 컨볼루션 층(Conv1, Conv2)의 필터 크기는 3x3x3이며 특징맵(feature map)의 개수는 각각 16이다. 컨볼루션 층의 활성화 함수로 Relu가 사용되었다. 입력 데이터가 두 개의 컨볼루션 층

을 통과하면 60x60 픽셀 크기의 특징맵 16개로 구성된 3차원 데이터는 57,600(=60x60x16)개의 값을 갖는 1차원 데이터로 변환된 후 완전 연결 계층(fully connected layer)로 입력된다. 57,600차원의 데이터는 다시 2,500차원의 데이터로 압축된다.

디코더는 인코더와 좌우 대칭 형태를 가진다. 압축된 2,500차원의 데이터는 완전 연결 계층에서 다시 57,600차원의 데이터로 확장된 후 2개의 Transposed Convolution 층(Conv Transpose1, Conv Transpose2)을 차례로 거치면서 업샘플링 되어 인코더에서의 데이터 사이즈로 복구된다. 디코더의 출력은 입력과 동일하게 64x64 픽셀 크기의 컬러 얼굴 이미지이며, 입력 얼굴 이미지에 가상의 마스크가 합성된 영상이다.

다음 3.1절과 3.2절에서는 제안 방법에서 사용한 컨볼루션 오토인코더의 학습 단계와 추론 단계에 대해 자세히 설명한다.

3.1. 학습 단계(Training step)

학습 단계의 목적은 마스크 미착용 얼굴 이미지를 입력했을 때 마스크가 합성된 얼굴 이미지를 출력할 수 있는 컨볼루션 오토인코더(그림 2)를 생성하는 것이다. 이를 위해 그림 3과 같이 마스크 미착용 얼굴 이미지와 마스크 착용 얼굴 이미지를 쌍(pair)으로 획득하고 각각 입력(원본 데이터)과 타겟으로 사용하였다. 그림 3에서 마스크 착용 얼굴 이미지를 획득하기 위해 모바일용 얼굴 인식 앱인 SNOW[10]에서 제공하는 마스크 생성 기능을 얼굴 이미지에 적용하여 가상의 마스크를 합성하였다.

본 논문에서는 GTA V Face Database[11]로부터 검출된 22명의 얼굴 이미지 301쌍(=총 602장)을 사용하였다. 얼굴이 포함된 이미지로부터 정렬된 얼굴 이미지를 획득하기 위해 양쪽 눈의 중심 좌표를 추출한 뒤 양쪽 눈이 수평이 되도록 얼굴영역을 회전한 뒤 정해진 크기로 잘라냈다. 각 인물마다 약 70%의 얼굴 이미지를 랜덤 선별하여 총 204쌍(=408장)을 컨볼루션 오토인코더 학습용으로 사용하였다. 컨볼루션 오토인코더 신경망의 가중치 갱신을 위해 Adam 옵티마이저를 사용하였으며 신경망 출력(즉, 입력으로부터 복원된 데이터)과 타겟 간 오차에 해당하는 손실함수(loss function)로 평균 제곱 오차(mean squared error)를 사용하였다. 에포크(epoch)는 160회로 설정하였다.

3.2. 추론 단계(inference step)

추론 단계에서는 학습이 완료된 컨볼루션 오토인코더에 마스크 미착용 얼굴 이미지를 입력하면 가상의 마스크 착용 얼굴 이미지를 출력한다. 이 기능을 통해 학습에 사용하지 않았던 얼굴 이미지들(예 : GTA V Face Database에서 추출된 301쌍 중 97쌍)에 대해 고속으로 마스크 착용 얼굴 이미지를 자동 생성할 수 있다. 여기서 마스크의 형상은 학습 단계에서 사용된 마스크 패턴과 유사하다.



Fig. 3 Face image pairs used in the training step (Unmasked face: input, Masked face: target)

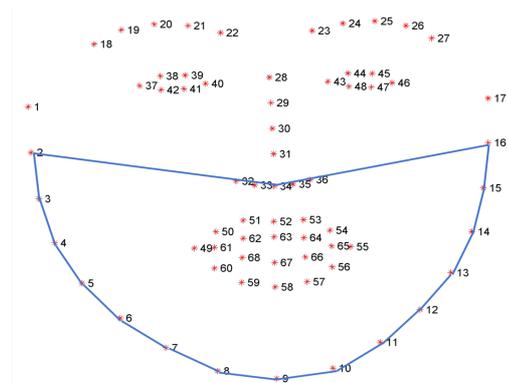


Fig. 4 Definition of facial mask pattern based on facial feature points (Method in [1]).

IV. 실험 결과

4.1. 마스크 얼굴 합성 결과

본 절에서는 마스크 미착용 얼굴 이미지를 입력으로 넣어 합성된 마스크 착용 얼굴 이미지를 출력하는 추론 단계의 결과를 확인한다. 비교를 위해 동일한 얼굴 이미지에 [1]의 방법을 적용하여 마스크 착용 얼굴 이미지를 생성하였다. 비교 방법[1]은 얼굴 특징점 추출에 기반한다. 구체적으로, 그림 4와 같이 얼굴 특징점 일부를 다각형의 형태로 연결하고 이 도형 안을 검정색으로 채운다.

그림 5는 제안 방법과 비교 방법의 마스크 착용 얼굴 합성 결과를 나타낸다. 비교 방법[1]의 경우, 다소 각진 마스크 형태로 인해 합성 결과가 다소 부자연스러웠다. 이에 비해 제안 방법은 학습된 마스크 패턴을 얼굴 각도 등 기하학적인 특성에 맞추어 표현함으로써 정면 및 측면 얼굴들에 대해 좀 더 자연스럽게 사실적인 결과를 얻을 수 있었다. 또한 입력 이미지(Original)의 인물을 구분하는데 중요한 고유 얼굴 특성들(예 : 눈썹, 눈매, 표정 등)이 합성된 결과 이미지에서도 잘 유지되는 것을 확인할 수 있었다.

제안 방법은 파이썬과 딥러닝 라이브러리인 케라스(Keras from Tensorflow)를 사용하여 구현되었다. i9-10900K CPU, 64GB RAM, 그래픽카드(RTX-3080) 1장을 사용하여 계산속도를 측정한 결과, 그림 2의 컨볼루션 오토인코더를 학습시키는데 소요된 시간은 약 18초

였다. 그림 6은 학습결과로서 에포크 160까지 손실함수 값의 변화 추이를 보여준다. 에포크 20까지는 급격히 손실함수 값이 낮아지고, 그 뒤로 서서히 수렴하는 안정적인 학습 결과를 보였다. 추론 단계에 소요되는 시간은 학습 단계에 비해 훨씬 짧다. 50장의 64x64 컬러 얼굴 이미지를 처리하는데 0.16초밖에 걸리지 않았다.



Fig. 5 Comparison results of synthesizing masked face images

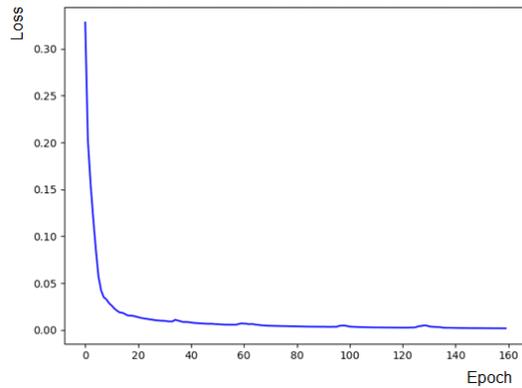


Fig. 6 Illustration of training graph (loss vs. epoch).

4.2. 장점과 한계점

[1, 3]을 비롯한 많은 방법들은 얼굴 특징점 추출 결과에 기반하여 가상의 마스크를 합성하였다. 하지만 비정면(non-frontal)의 촬영 얼굴, 저해상도 이미지 등의 문제로 인해 특징점 추출이 어려운 경우에는 마스크 합성도 불가능해질 수 있다. 반면에 제안 방법은 특징점 추출 과정 없이도 얼굴 이미지만 있으면 적용할 수 있으므로 활용도가 높고 제약이 적다. SNOW 앱[10]의 경우 이미지 단위로 마스크 합성을 처리하므로 다수 개의 이미지들에 대해 마스크 얼굴 이미지를 얻으려면 수동으로 선택한 이미지 한장씩 기능을 적용하여 결과를 얻어야 하는 번거로움이 있다. 하지만 제안 방법은 여러 장의 얼굴 이미지들을 매트릭스 형태로 한꺼번에 신경망에 입력하여 매우 빠른 속도로 마스크 착용 얼굴 이미지들을 생성할 수 있는 배치처리가 가능하다. 뿐만 아니라, 다소 제한된 종류의 마스크 패턴을 제공하는 SNOW 앱과 [1] 방법과는 다르게 제안 방법은 컨볼루션 오토인코더 추가 학습을 통해 새로운 마스크 패턴 학습이 가능하다. 비슷한 얼굴 촬영 조건(동일 인물, 동일한 조명 조건, 동일한 얼굴 각도와 표정 등)을 유지한 채 마스크 착용 및 미착용 얼굴을 각각 촬영하고 이 얼굴 이미지 쌍을 이용하여 컨볼루션 오토인코더를 학습한다면, 추론 단계에서 보다 사실적이고 자연스러운 합성 결과를 기대할 수 있을 것이다.

제안 방법의 한계점은 마스크와 얼굴 합성 결과 이미지의 선명도가 다소 낮다는 것이며 이는 향후 연구를 통해 개선이 필요하다.

V. 결 론

본 논문에서는 마스크 미착용 얼굴 이미지에 가상의 마스크 패턴을 합성하는 방법을 제안하였다. 제안 방법의 아이디어는 컨볼루션 오토인코더에 동일 인물에 대해 마스크 미착용 얼굴 이미지와 마스크 착용 얼굴 이미지를 쌍으로 입력하여 얼굴과 마스크의 기하학적인 관계를 학습하는 것이다. 제안 방법은 마스크 착용 얼굴을 대상으로 한 얼굴인식에서 풍부한 학습 이미지 데이터를 확보하는데 유용할 것이다. 또한 마스크를 착용하였던 범죄 용의자의 얼굴을 예측하는 몽타주 제작에도 활용 가능할 것으로 기대한다.

향후에는 보다 사실적이고 선명한 마스크 착용 이미지를 생성하기 위한 연구를 수행할 것이다. 이를 위해 새로운 손실함수를 설계하고, GAN(Generative Adversarial Networks)[12]과 같은 정교한 이미지 합성 딥러닝 기법을 제안 방법에 적용할 것이다. 또한 마스크 착용 이미지에 대한 얼굴인식 정확도 분석을 수행할 것이다.

References

[1] M. L. Ngan, P. J. Grother, and K. K. Hanaoka, "Ongoing Face Recognition Vendor Test (FRVT) Part 6A: Face Recognition Accuracy with Masks Using Pre-COVID-19 Algorithms," National Institute of Standards and Technology, Technical Report NISTIR 3811, Jul. 2020.

[2] H. I. Kim, J. Y. Moon, and J. Y. Park, "Research Trends for Deep Learning-Based High-Performance Face Recognition Technology," *Electronics and Telecommunications Trends*, vol. 33, no. 4, pp. 43-53, Aug. 2018.

[3] Y. Kim, X. Zhang, and J. Park, "Automatic Mask Face Data Synthesis System," in *Proceedings of Conference of the Korean Institute of Broadcast and Media Engineers*, Online, pp. 239-240, 2020.

[4] OpenCV-Python [cited 2022 April 10], [Internet]. Available: <https://docs.opencv.org/4.3.0/>.

[5] J. Masci, U. Meier, D. Ciresan, and J. Schmidhuber, "Stacked Convolutional Auto-Encoders for Hierarchical Feature Extraction," in *Proceedings of International Conference on Artificial Neural Networks*, Espoo, Finland, pp. 52-59, 2011.

[6] H. Kim, S. Lee, and Y. No, "Face Recognition in the Wild," *Information and Communications Magazine*, vol. 31, no. 4,

pp. 88-98, Mar. 2014.

[7] G. Hua, M. H. Yang, E. Learned-Miller, Y. Ma, M. Turk, D. J. Kriegman, and T. S. Huang, "Introduction to the Special Section on Real-world Face Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 10, pp. 1921-1924, Oct. 2011.

[8] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient Based Learning Applied to Document Recognition," in *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278-2324, Nov. 1998.

[9] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet Classification with Deep Convolutional Neural Networks," in *Proceedings of Advances in Neural Information Processing Systems*, Lake Tahoe: NV, USA, pp. 1097-1105, 2012.

[10] SNOW Face Recognition Application, [Internet]. Available: <https://play.google.com/store/apps/details?id=com.campmobile.snow&hl=ko&gl=US>.

[11] GTAV Face Database [cited 2022 April 10], [Internet]. Available: <https://gtav.upc.edu/en/research-areas/face-database>.

[12] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Nets," in *Proceedings of Advances Neural Information Processing Systems Conference*, Montreal, Canada, pp. 2672-2680, 2014.



이승호(Seung Ho Lee)

現 한국기술교육대학교 융합학과 교수
前 국방과학연구소 선임연구원
※관심분야: 머신러닝, 컴퓨터비전