

---

저자 (Authors)	이기탁, 김창식, 김영진 Keetark Lee, Changsik Kim, Young-Jin Kim
출처 (Source)	<a href="#">한국정보과학회 학술발표논문집</a> , 2018.6, 299-301(3 pages)
발행처 (Publisher)	<a href="#">한국정보과학회</a> The Korean Institute of Information Scientists and Engineers
URL	<a href="http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE07502987">http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE07502987</a>
APA Style	이기탁, 김창식, 김영진 (2018). Convolutional Neural Network 기반의 자동출석 확인 시스템. 한국정보과학회 학술발표논문집, 299-301
이용정보 (Accessed)	송실대학교 219.255.***.171 2020/06/15 16:24 (KST)

---

### 저작권 안내

DBpia에서 제공되는 모든 저작물의 저작권은 원저작자에게 있으며, 누리미디어는 각 저작물의 내용을 보증하거나 책임을 지지 않습니다. 그리고 DBpia에서 제공되는 저작물은 DBpia와 구독계약을 체결한 기관소속 이용자 혹은 해당 저작물의 개별 구매자가 비영리적으로만 이용할 수 있습니다. 그러므로 이에 위반하여 DBpia에서 제공되는 저작물을 복제, 전송 등의 방법으로 무단 이용하는 경우 관련 법령에 따라 민, 형사상의 책임을 질 수 있습니다.

### Copyright Information

Copyright of all literary works provided by DBpia belongs to the copyright holder(s) and Nurimedia does not guarantee contents of the literary work or assume responsibility for the same. In addition, the literary works provided by DBpia may only be used by the users affiliated to the institutions which executed a subscription agreement with DBpia or the individual purchasers of the literary work(s) for non-commercial purposes. Therefore, any person who illegally uses the literary works provided by DBpia by means of reproduction or transmission shall assume civil and criminal responsibility according to applicable laws and regulations.

## Convolutional Neural Network 기반의 자동출석 확인 시스템

이기탁<sup>O</sup> 김창식 김영진  
아주대학교 전자공학과

tak5868@ajou.ac.kr, kyscs30@ajou.ac.kr, youngkim@ajou.ac.kr

## Convolutional Neural Network based Automatic Attendance Check System

Keetark Lee<sup>O1)</sup> Changsik Kim Young-Jin Kim

Department of Electrical and Computer Engineering, Ajou University

## 요 약

그래픽 처리 장치의 성능향상으로 인해 역전파 알고리즘이 탄력을 받은 이후, 딥러닝은 금융, 헬스케어, 이미지 분류, 자연어처리 등 다양한 분야에 사용되고 있다. 본 논문에서는 Convolutional Neural Network을 이용한 얼굴 탐지 및 얼굴 인식 기법을 사용하여 기존의 자동출석 확인 시스템에서 가지고 있는 문제인 병목현상, 부정출석, 대리출석 문제를 해결한다. 이를 위해 본 논문에서는 강의실 모든 지역을 촬영할 수 있는 카메라 서버시스템 및 수강생 정보에 등록되어있지 않는 자를 판별할 수 있게 하는 open-set 얼굴인식의 사용을 제안하고, 이를 통합한 결과물을 직접 구현하여 실제 강의실 환경에서 테스트함으로써 그 실용성을 보인다.

## 1. 서 론

최근 얼굴 인식 기술에서 딥러닝 기술이 효과적으로 사용되고 있다. 이미지 내 분석 대상의 복잡한 특징(feature)을 추출하는 가중치(weight)를 학습하기 위해서는 데이터로부터 고 수준 특징을 추출하는 능력이 필요한데 딥러닝은 이를 실현시켜주고 있다. 얼굴 인식을 위한 딥러닝 기법으로는 CNN(Convolutional Neural Network)이 사용되며 이는 컨볼루션(convolution) 층과 max-pooling 층이 교차하는 구조이다 [1]. CNN이 이미지 처리 분야에 자리 잡기 시작한 것은 2012년 AlexNet[2]이 등장한 이후이다. 그 뒤로, 층을 더 쌓아 성능을 높이려는 시도가 계속되었다. 3x3 필터를 깊게 쌓아 높은 성능을 낸 VGGNet [3], 1x1 컨볼루션 필터를 사용하여 차원 축소 효과를 내는 Inception 모듈을 사용한 GoogleNet [4] 등이 있다. 딥러닝을 이용해 CNN을 학습한 결과물은 data-driven 가중치로서 나타나고, 사람의 얼굴로만 구성된 학습세트(training set)를 이용해 VGGNet을 학습하여 사람 얼굴 인식에 좀 더 특화된 data-driven 가중치를 학습한 연구[5]가 존재한다. 이러한 방법론을 통해 CNN은 높은 얼굴 인식률을 확보하여 얼굴 인식에 성공적으로 적용되었다[6]. 특히 가변적인 얼굴 패턴을 견고하게 탐지할 수 있기 때문에 얼굴 인식에서 효과적인 기법으로 통한다[7].

현대의 많은 대학교들은 수업과 출석의 효율성을 위해 RFID(Radio-Frequency-Identification), 비콘(beacon) 기반의 자동출석 확인 시스템을 구축하고 있다. 하지만 현존하는 자동출석 확인 시스템 중 다음에 같은 세 가지 중요한 문제점을 모두 해결하는 시스템은 존재하지 않는다. 첫째, 타인의 출석을 대신 해주는 대리출석 문제. 둘째, 많은 수의 인원을 일괄적으로 단번에 출석 처리할 수 없는 병목현상 문제. 셋째, 출석 확인 후 수업은 듣지 않는 부정출석 문제이다. 이러한 문제점을 지닌 기존의 자동출석 확인 시스템을 대체하기 위해 얼굴 인식 기반

의 출석 확인 시스템은 다양한 형태로 연구되어 왔다. [8]은 RFID와 얼굴 인식을 결합하여 출석 확인을 수행한다. 이는 대리 출석을 방지할 수 있으나 병목 현상의 가능성이 있다. [9]은 패턴매칭(pattern matching)을 이용해 얼굴 인식을 수행하며 지속적으로 출석을 확인하여 출석 신뢰도를 높인다. [10]은 Center-Face를 사용하여 얼굴 인식을 수행하고 지속적인 출석 확인이 가능하다. 그러나 [9]의 방법은 정확도가 낮을 수 있다는 단점이 존재하고 [9]과 [10]의 방법은 청강생, 도강생 등의 수강생 정보에 등록되어있지 않은 사람에 대한 분별력이 전혀 없으며 조도 등의 노이즈 환경을 고려하지 않는다.

따라서 본 논문에서는 기존의 출석 확인 시스템이 가지고 있는 구조적인 문제점을 해결하고, 청강생 등 알려지지 않은 샘플에 대한 분별력을 가지는 얼굴 인식 기반 자동출석 확인 시스템을 제안한다. 또한 이를 임베디드 보드에 직접 구현하여 실제 강의실 환경에서 테스트함으로써 그 실용성을 보인다.

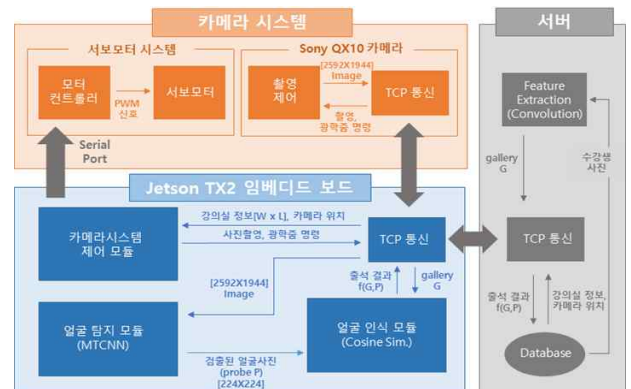


그림 1 제안하는 시스템의 구조도

본 논문의 구성은 다음과 같다. 2장에서는 제안하는 시스템의 전체적인 구조와 이를 구성하는 하위 기법과 하위 시스템들을 설명한다. 3장에서는 2장에서 설명한 알고리즘을 실제 구현한 결과를 보이며, 마지막으로 4장에서 결론을 짓는다.

\* 이 성과는 2018년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2018R1A2B6005466).

## 2. CNN 기반의 자동출석 확인 시스템

본 논문에서 제안하는 자동출석 확인 시스템의 동작은 그림 1과 같은 구조로 진행된다. 전체 시스템은 크게 3개의 하위 시스템인 카메라 시스템, 임베디드 보드, 서버로 구성되어 있다. 임베디드 보드로는 NVIDIA Jetson TX2를 사용하였는데 4개의 모듈로 구성되며 통신 모듈을 제외한 3개의 모듈은 각각 2.1, 2.2, 2.3절에서 소개한다.

### 2.1 카메라 시스템 제어 모듈

카메라 시스템은 2개의 서보모터와 1개의 카메라로 구성되어 있다. 카메라 시스템 제어모듈은 강의실 전면부에 위치한 카메라가 강의실 전체를 빠짐없이 촬영하는 것을 목적으로 하며 그림 1과 같이 서버로부터 강의실의 가로, 세로 길이 정보 및 카메라의 위치  $(x_c, y_c, z_c)$  정보를 받는다. 전체 강의실은 여러 개의 분할구역(segment)로 나뉘어져서 촬영되며 분할구역의 수는 강의실 면적에 비례한다.  $i$ 번째 분할구역의 중심 좌표를  $(x_i, y_i)$ 라고 하였을 때 종 방향 각도, 횡 방향 각도는 식(1), (2)와 같이 결정되며 광학 줌(optical zoom)의 강도는 식(3)과 같이 결정된다.

$$\text{종 방향 각도} = \arctan\left(\frac{x_i - x_c}{z_c}\right) \quad (1)$$

$$\text{횡 방향 각도} = \arccos\left(\frac{\sqrt{(x_i - x_c)^2 + z_c^2}}{\sqrt{(x_i - x_c)^2 + (y_i - y_c)^2 + z_c^2}}\right) \quad (2)$$

$$\text{광학 줌} \propto \sqrt{(x_i - x_c)^2 + (y_i - y_c)^2 + z_c^2} \quad (3)$$

종 방향 회전 각도와 횡 방향 회전 각도는 그림 1과 같이 serial port를 통해 모터 컨트롤러에게 전달되며 광학 줌의 강도는 TCP통신을 통해 카메라로 전달된다.

### 2.2 얼굴 탐지 모듈

얼굴 인식의 전처리 단계로 얼굴 정렬을 추가함으로써 괄목할만한 성능향상을 얻은 연구[11]에 입각하여 얼굴 탐지와 얼굴 정렬을 동시에 수행할 수 있는 MTCNN(Multi-Task cascaded Convolutional Neural Network) 얼굴 탐지 기법[12]을 이용한다.

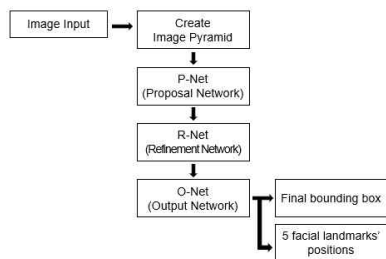


그림 2 MTCNN 얼굴 탐지 기법의 파이프라인

MTCNN의 파이프라인은 그림 2와 같다. 먼저 3단계의 직렬 프레임워크의 입력으로써 이미지 피라미드(image pyramid)를 생성한다. 이후 과정은 다음과 같다.

P-Net(Proposal Network) : 후보(candidate) window와 이것의 bounding box 회귀 벡터를 얻기 위해 P-Net을 수행한다.

R-Net(Refinement Network) : 여러 잘못된 후보들을 제외하며 bounding box 회귀를 통해 calibration을 수행하고, NMS(Non Maximum Suppression)로 후보들을 병합한다.

O-Net(Output Network) : R-Net과 과정이 유사하지만 O-Net을 통해 더욱 정밀한 얼굴 탐지를 수행한다. 특히 이 네트워크는 5개의 얼굴 대표점(facial landmark)의 위치를 출력한다.

### 2.3 얼굴 인식 모듈

실제 강의실 환경에서는 청강생, 도강생, 조교 등 수강생 정보에는 등록되어 있지 않은 인물들이 존재한다. 이들의 샘플이 검출되어 얼굴 인식 모듈의 입력으로 들어갈 경우, 출력이 closed-set이라고 가정한 대표적인 기존의 분류기들인 Support Vector Machine[13], Multi Layer Perceptron[13], Random Forest[14]를 사용할 시 다른 누군가 부득이하게 출석이 인정된다. 실제로 얼굴 탐지 모듈의 FP(False Positive)까지 고려한다면 이는 매우 치명적이다. 때문에 실제 강의실 환경에서 사용되는 얼굴 인식 모듈의 출력은 open-set임을 가정[15]할 필요가 있다. open-set 얼굴 인식[15]에서는 두 가지 용어가 정의되는데 첫째, 시스템이 이미 알고 있는 사람의 biometric 샘플 집합인 갤러리(gallery)와 둘째, 시스템의 입력으로 들어온 inference 대상인 프로브(probe)이다.

$$\text{gallery } G = \{G^1, G^2, \dots, G^n\} \quad (4)$$

식(4)에서  $n$ 은 갤러리에 등록된 사람의 수를 나타낸다. 이 때, 사람  $a$ 의 갤러리 템플릿(template)  $G^a$ 는 식(5)와 같이 구성된다.

$$G^a = \{G_0^a, G_1^a, G_2^a\}, \text{ where } a \in \{1, 2, \dots, n\} \quad (5)$$

식 (5)에서  $G_i^a$ 는 컨볼루션 층에서 추출된 사람  $a$ 의 특징 벡터를 의미한다. 얼굴 인식 과정은 시스템의 입력으로 들어온 프로브와 갤러리의 다음과 같은 유사도(similarity) 계산을 통해 이루어진다. 먼저 첫 번째로, 세 개의 특징 벡터를 식 (6)과 같이 요소 간(elementwise) 평균을 취한다. 특징벡터는 VGGNet 기반 얼굴 인식용 data-driven 컨볼루션 가중치를 사용하여 추출하였다 [5].

$$\bar{G}^a = \frac{1}{3} \sum_{k=0}^2 G_k^a \quad (6)$$

이 때, 갤러리 템플릿  $G^a$ 와 프로브  $P$  간의 유사도  $S(G^a, P)$ 는 식(7)과 같으며 유사도는 코사인 유사도[16]를 통해 계산된다.

$$S(G^a, P) = \cos(\bar{G}^a, P) = \frac{\sum_{i=1}^d (\bar{G}_i^a \times P_i)}{\sqrt{\sum_{i=1}^d (\bar{G}_i^a)^2} \times \sqrt{\sum_{i=1}^d (P_i)^2}} \quad (7)$$

식 (7)에서  $d$ 는 특징 벡터의 차원을 의미한다. 이 때, 얼굴 인식 모듈은 함수  $f$ 로 표현되며 얼굴 인식 모듈의 출력은 식(8)과 같다.

$$f(G, P) = \begin{cases} \operatorname{argmax}_a S(G^a, P) & \max_a S(G^a, P) \geq \tau \\ \text{unknown} & \max_a S(G^a, P) < \tau \end{cases} \quad (8)$$

unknown 출력은 프로브  $P$ 가 수강생 목록에 포함된 사람이 아니라는 것을 의미한다. 식(8)에서 임계값(threshold)  $\tau$ 는 얼굴 인식 모듈의 FAR(False Alarm Rate) 성능[15]과 DIR(Detection and Identification Rate) 성능[15]의 트레이드-오프(trade-off) 매개변수(parameter)를 의미한다.  $\tau$ 값이 낮을수록 수강생 목록에 포함되어 있지 않은 사람을 거르지 못할 확률인 FAR이 올라가면서, 수강생 목록에 포함되어 있는 사람이 올바르게 출석 확인될 확률인 DIR도 올라간다. 즉, FAR성능은 안 좋아지고, DIR성능은 좋아진다.  $\tau$ 값이 높은 경우는 그 반대의 경우에 해당한다. 본 논문에서는 자동출석시스템을 위한 적절한 FAR을 0.1로 보았다. 이를 위해, 우리는 [17]의 방법처럼 LFW(Labeled Faces in the Wild) 데이터베이스[18]를 이용하여  $\tau$ 를 무차별(brute force)적으로 대입하였을 때, FAR이 0.1인 지점을 찾았으며, 해당 지점을 만드는  $\tau$ 는 0.915이고 DIR성능은 0.85를 보여주었다. 또한,  $\tau=0$ (closed-set 얼굴 인식)일 경우 DIR성능은 0.97을 보여주었다.

## 3. 실험 결과

### 3.1 closed-set 얼굴 인식( $\tau=0$ )

청강생, 도강생, 조교의 존재 및 얼굴 탐지 모듈의 FP 샘플

이 발생하는 경우를 아예 배제함으로써 더 높은 얼굴 인식 TP(True Positive) 확률을 원하는 경우, 사용자 매개변수인  $\tau$ 를  $\tau=0$ 으로 설정함으로써 제안하는 시스템을 closed-set 얼굴 인식을 수행하는 기기로서 사용할 수 있다. 본 절에서는 강의실 창가부근에서 발생하는 illumination 효과를 고려하여 도전적인 테스트세트 구성을 위해 강한 햇빛 아래에서 촬영된 얼굴사진 총 5세트, 37장을 준비하였다.  $\tau=0$ 으로 설정 후 얼굴 인식 테스트 결과 92%성능을 보여주어 이는 [17] 및 2.3절에서 수행했던 LFW 데이터베이스의 결과보다 도전적인 테스트세트 구성으로 인해 다소 하락한 모습이었다.

### 3.2 open-set 얼굴 인식( $\tau=0.915$ )



그림 3 테스트를 위한 강의실 환경

앞서 제안한 시스템의 open-set 얼굴 인식의 실용성을 직접 확인하기 위해 제안하는 시스템을 직접 임베디드 보드에 구현하여 실제 강의실 환경에서 테스트 하였다. 강의실 전체 사진을 촬영하는 카메라는 칠판 위쪽에 부착 되어있고 피촬영자들은 카메라를 바라보는 것이 아닌 칠판을 바라보고 있도록 하면서, 일부러 턱을 괴거나 강의실 구석에 앉는 등 도전적인 실험 환경을 구성하였다. 그림 3은 구성된 실험 환경을 2.1절에서 언급된 카메라 제어 모듈을 통해 촬영한 사진이다.

표 1 제안하는 시스템의 테스트 결과

인식결과 \ 사람	사람A	사람B	사람C	사람D
test 1	O	O	O	O
test 2	X (unknown)	O	O	O
test 3	O	O	O	X (unknown)

표 1은 테스트 결과를 나타낸다. 총 3번의 전체 결과 중 2번의 오인식을 보였다. 2번의 오인식은 모두 사람의 label을 잘못 판단하는 경우가 아닌, 사람의 label은 맞았지만 유사도 값이  $\tau$ 를 넘지 못하여 unknown으로 인식한 결과였다. 이러한 결과는  $\tau=0$ 인 경우 높은 얼굴 인식 성능을 보여주는 결과와 맥락을 같이하는데, open-set 얼굴 인식에서 오인식의 주된 원인은 프로브의 label을 잘못 판단하는 것이 아닌, 프로브가 unknown이라고 판단하는 것에 있다는 결과였다. 총 2번의 오인식 결과 중, 사람A의 오인식 샘플은 강의실 앞문 쪽에 앉아 지나치게 옆모습이 촬영된 경우였고, 사람D의 오인식 샘플은 카메라 초점이 맞지 않은 채로 촬영된 경우로 나타났다.

### 3.3 수행시간 분석

표 2는 얼굴 탐지와 얼굴 인식 프로그램을 NVIDIA Jetson TX2 임베디드 보드에서 수행하였을 때의 수행시간을 측정한 결과이다. 얼굴 탐지와 얼굴 인식 각각 20장 이미지 및 20장의 프로브를 모듈의 입력으로 넣고 수행되는 시간을 평균 하였고, 그 결과 얼굴 탐지의 수행시간이 얼굴 인식에 비해 약 19배 긴

것으로 확인되었다.

표 2 임베디드 보드에서의 프로그램 수행시간

	얼굴 탐지 [2592*1944 크기 이미지]	얼굴 인식 [224*224 크기 프로브]
수행시간	3.1초	0.16초

## 4. 결론 및 향후 연구

본 논문에서는 기존 출석 확인 시스템의 문제점을 해결하는 CNN기반 자동출석 확인 시스템을 제안하고, 이를 직접 구현하여 실제 강의실 환경에서 그 실용성을 보였다. 실험 결과 illumination 환경에서 일반적인 환경에 비해 DIR 성능은 다소 떨어지는 것으로 나타났으며, 지나치게 옆모습이 촬영되는 경우, 카메라 초점이 맞지 않는 경우 오인식이 발생하는 결과를 보였다. 또한 얼굴 인식에 비해 얼굴 탐지의 수행시간이 매우 긴 것으로 나타났다. 이를 위해 강의실 창가의 illumination에 좀 더 강한 하능 갤러리 템플릿 구성 연구, 옆모습이 촬영되는 것을 고려한 3차원 얼굴 정렬에 관한 연구, 카메라 초점을 고려한 광학 줌 설정기법 및 보다 빠른 출석 확인을 위해 얼굴 탐지 기법 가속화에 대한 연구를 진행할 예정이다.

## 5. 참고문헌

- [1] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-Based Learning Applied to Document Recognition," *Proceedings of the IEEE*, 1998.
- [2] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," In *Advances in Neural Information Processing Systems*, pp. 1097-1105, 2012.
- [3] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. International Conference on Learning Representations (ICLR)*, Apr. 2015.
- [4] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, USA, pp. 1-9, June. 2015.
- [5] O. M. Parkhi, A. Vedaldi, A. Zisserman, "Deep Face Recognition," *British Machine Vision Conference*, 2015
- [6] A. R. Syafeeza, M. Khalil-Hani, S. S. Liew, and R. Bakhteri, "Convolutional neural network for face recognition with pose and illumination variation," *International Journal of Engineering & Technology*(0975-4024), vol. 6, no. 1, 2014.
- [7] R. D. Albu, "Human Face Recognition Using Convolutional Neural Networks," *Journal of Electrical & Electronics Engineering*, 2009.
- [8] S. Pss, and M. Bhaskar, "RFID and pose invariant face verification based automated classroom attendance system," in *Proc. International Conference on Microelectronics, Computing and Communications (MicroCom)*, Durgapur, India, pp. 1-6, Jan. 2016.
- [9] V. Soinya, R. S. Sri, K. S. Titty, R. Ramakrishnan, and S. Sivakumar, "Attendance automation using face recognition biometric authentication," in *Proc. International Conference on Power and Embedded Drive Control (ICPEDC)*, 2017.
- [10] R. Fu, D. Wang, D. Li, and Z. Luo, "University Classroom Attendance Based on Deep Learning," in *Proc. 10th International Conference on Intelligent Computation Technology and Automation (ICICTA)*, Changsha, China, pp. 128-131, Oct. 2017.
- [11] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "DeepFace: Closing the gap to Human-Level Performance in Face Verification," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Columbus, OH, USA, pp. 1701-1708, June. 2014.
- [12] Zhang, K., Zhang, Z., Li, Z., and Qiao, Y., "Joint face detection and alignment using multi-task cascaded convolutional networks," *arXiv preprint arXiv:1604.02878*, Apr. 2016.
- [13] M. Mohammed, M. B. Khan, "Machine Learning Algorithms and Applications," CRC Press, 2017
- [14] L. Breiman, "Random Forests," *Machine Learning*, 2001
- [15] S.Z. Li and A.K. Jain, *Handbook of Face Recognition*, eds. Springer, 2005
- [16] P.Dangeti, "Statistics for Machine Learning," Packt, 2017
- [17] M. Gunther, S. Cruz, E. M. Rudd, and T. E. Boulton, "Toward Open-Set Face Recognition," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2017.
- [18] E. Learned-Miller, G. B. Huang, A. RoyChowdhury, H. Li, and G. Hua, "Labeled faces in the wild: A survey," in *Advances in Face Detection and Facial Image Analysis*, pp. 189-248, 2016.