

mask generater for occlusion-aware face recognition

Anonymous CVPR submission

Paper ID ****

Abstract

딥러닝 기반의 얼굴 인식 분야는 지난 몇년동안 네트워크 구조 개선 및 다양한 손실함수 제안을 통해 다양한 방식으로 발전되고 연구되어왔다. 그러나 기존의 CNN 기반의 얼굴 인식 네트워크에는 여러 한계점이 존재한다. 특히 폐색영역에 대한 처리과정 없이 잘 정제된 데이터셋으로 훈련된 네트워크는 실제환경에서의 정확도에 문제가 발생한다. 본 논문에서는 폐색영역을 인지하여 처리하는 *ene-to-end* 뉴럴 네트워크를 제안한다. 임의의 폐색영역을 추출하는 네트워크를 구성함으로써 기존의 문제점을 해결하고 제안하는 프레임워크가 기존의 네트워크에 비해 일반화된 성능을 가짐을 보인다.

1. Introduction

컴퓨터 비전 분야 중 얼굴 인식은 수십 년 동안 다양한 방식으로 발전되고 연구되어왔다. 특히 여러 생체인증 방식인 지문, 정맥, 홍채 등의 방식과 비교하였을 때, 간단하면서도 추가적인 행동없이 인식하는 장점을 갖고 있다. 또한 얼굴 인식은 신원 확인뿐 아니라 다양한 마케팅과 인구 통계 등의 어플리케이션에 활용할 수 있는 잠재력을 갖고 있어 많은 연구자들에 의해 관심을 받아 왔다.

최근 딥러닝 분야의 성공적인 발전은 컴퓨터 비전 분야의 폭발적인 성능향상과 깊은 관계가 있다. 얼굴 인식 분야도 마찬가지로 Deep Convolutional Neural Networks(DCNNs)을 기반으로 수년동안 많은 연구가 진행되어왔다. 네트워크 구조 개선 및 특징들을 구별하는 능력을 강화하기 위해 여러 손실함수들을 제안하는 등의 연구들은 특정 공공 데이터셋에서 사람의 인지 능력을 뛰어넘는 결과를 생성하였다.

기존의 딥러닝 모델을 활용한 얼굴 인식방식은 많은 성능향상을 이뤄 높은 정확도를 기록하였지만, 이를 활용한 실제 환경에서의 시스템에서는 많은 정확도 하락이 발생한다. 이는 정제되지 않은 실제 영상에는 포즈, 얼굴 표정, 조명 변화, 다른 오브젝트에 의한 폐색영역과 같은 요소들이 존재하기 때문이다. 이러한 요소들 중에서 폐색영역은 오랜 기간 동안 해결하기 어려운 문제로 남아있다. 실제 영상에 존재하는 폐색영역에는 선글라스, 마스크, 모자 등과 같이 간단한 액세서리부터 책, 컵과 같은 일반적인 물체들에 의해 발생한다는 문제 때문이다. 기존의 DCNNs을 활

용한 얼굴 인식은 입력 영상으로부터 여러 Conv 레이어를 통해 비교적 낮은 클래스 내 거리와 높은 클래스 간 거리의 특성을 가지도록 학습하여 클래스 간의 구별력을 높이는 데 초점이 맞춰져 있었다. 하지만 폐색영역을 구별하지 못하고 이 또한 학습하므로 일반적으로 기존의 네트워크는 정확도 하락이 발생 할 수밖에 없다는 한계점이 있다.

이와 같은 한계점을 해결하고자, 본 논문에서는 폐색된 영역을 인지하여 처리하는 *end-to-end* 네트워크를 제안한다. 제안하는 프레임워크는 두개의 단계를 통해 진행된다. 첫 번째로는 폐색된 영역을 추출하는 마스크 생성 모듈과 두 번째는 기존의 얼굴 인식 분야에서 널리 쓰이는 CNN 기반의 뉴럴 네트워크이다. 마스크 생성기 모듈은 원본 영상과 이에 대응하는 폐색영역을 추가한 영상을 입력으로 하여 입력으로부터 마스킹된 영역을 추출하는 것을 목표로 한다. 추출된 마스킹 영역은 폐색영역이 존재하는 영상과 함께 뉴럴 네트워크의 입력으로 사용되며 결과적으로 폐색되지 않은 영역으로부터 얼굴 인식을 수행하도록 한다.

2. Related Work

이 section에서는 가려진 얼굴 인식을 위한 다양한 기법들에 대해 소개한다.

2.1. General face detection

face detection은 일반적으로 제한되지 않은 환경에서 캡처 된 얼굴을 감지한다. 그러나 큰 포즈 변화, 다양한 조명, 저해상도 및 occlusion 등으로 인해 어려움을 겪고 있다. 일반적으로 face detection에 대한 접근 방식은 크게 세가지로 rigid templates, deformable part models (DPM), deep convolutional neural networks (DCNNs)로 나뉜다.

첫번째 방식으로 Viola-Jones face detector[1]는 계단식 분류기를 학습시키고 실시간 효율성으로 우수한 성능을 달성할 수 있는 rigid templates에 일반적이다. 이와 대조적으로, 두번째 방식인 DPM 기반 얼굴 감지는 높은 계산 복잡성의 비용을 기반으로 훨씬 더 나은 성능을 달성할 수 있다[2]. 마지막으로 가장 유망한 방식은 DCNN을 이용한 방법이다. neural network를 이용한 face detection은 2.2와 2.3에서 자세히 다룬다.

2.2. Neural network based face detection

컴퓨터 비전에 있어 neural network를 이용한 face detection 연구는 과거에서부터 계속되어 왔다. 1994년 초 Vaillant et al.[3]은 face detection을 위해 neural network를 적용하였다. 이 연구에서는 image window에서 얼굴의 존재 여부를 감지하고 가능한 모든 위치에서 네트워크로 전체 image를 스캔하기 위해 convolution neural network를 훈련시킬 것을 제안했다. 1996년 Rowley et al.[4]는 upright frontal face을 위해 retinally하게 연결된 neural network를 제시했다. 2002년 Garcia et al.[5]은 복잡한 이미지에서 사람의 정면 얼굴을 감지하는 neural network를 개발했다. 2005년 Osadchy et al.[6]는 얼굴감지 및 포즈를 동시에 추정하기 위한 convolution neural network를 개발하였다.

2.3. Convolutional neural network

Convolutional neural network(CNN)는 image recognition과 classification 분야에서 효과적으로 입증된 신경망이다. CNN은 데이터의 feature를 깊이 학습함으로써 다양한 비전 작업에 적합하다. face detection 연구에 있어서도 강력한 CNN의 기능을 인정받아 많은 연구가 이어지고 있다. CNN은 많은 layers로 구성되어진 feed forward neural networks의 한 종류로 입력 이미지를 가져와 필터 또는 커널과 convolution하여 특징을 추출하는 딥러닝 알고리즘이다. 일반적인 CNN의 architecture는 convolution, pooling, rectified linear unit(ReLU), fully connected layers로 구성되어진다.

CNN을 기반으로 몇 가지 물체 감지 방법이 얼굴 감지 작업에 적용되었다. 예를 들어, Lin et al.[7]은 LeNet CNN 구조의 출력 노드 수를 얼굴 감지를 위해 수정하는 NLet-Net 방법을 제안하여 테러리스트의 가려진 얼굴을 감지한다. FaceCraft[8]은 빠른 얼굴 감지를 위해 R-CNN 방법의 감지율을 향상시키는 region proposal network(RPN)과 빠른 R-CNN 모델을 공동으로 훈련한다. 그러나 R-CNN은 이미지 피라미드에서 특징을 추출하기 때문에 시간이 많이 걸리는 문제점이 있다. MTCNN[9]은 얼굴 감지와 정렬의 성능을 높이기 위해 탐지와 정렬 작업 간의 고유한 상관 관계를 활용하는 계단식 다중 작업 프레임 워크를 사용한다. MTCNN은 얼굴 감지하기 위해 여러 CNN networks를 사용하므로 계산 효율성이 떨어진다. SAFD[10]는 스케일 제안 네트워크를 사용하여 테스트 이미지에서 얼굴의 가능한 스케일을 예측하여 CNN 기반 얼굴 검출기의 성능을 효과적으로 향상시킬 수 있다. 가장 최근에는 CNN을 이용해 작은 얼굴을 찾는 것이 face detection에서 인기를 얻었으며 우수한 성능을 달성했다.[11]

3. Proposed method

최근 딥 러닝 기반 얼굴 인식 기술은 혁신적인 성능을 보이며 여러 분야에 활용되고 있다. 그러나 다양한 환경에서 성능을 내기가 어렵다. 특히 기존 얼굴 인식 연구는 얼굴 내 폐색영역을 세심히 다루지 않고 있다. 얼굴 인식 분야에서 폐색영역을 다루는 방법은 크게 두가지가 있다.

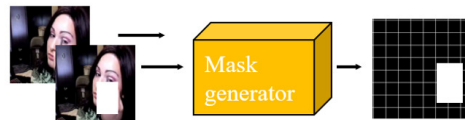


Figure 1. MASK generator.

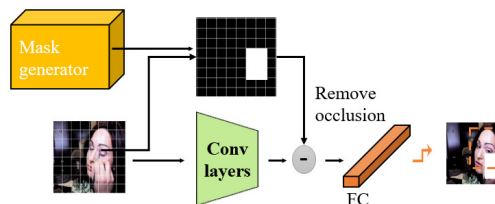


Figure 2. Pipeline of occlusion aware net.

첫째 얼굴에 위치한 폐색영역 또한 하나의 정보로 사용하여 얼굴의 특징을 인식하는 방법이다. 둘째 폐색영역을 제외한 얼굴의 일부 특징만을 이용하여 얼굴을 인식한다. 폐색영역이 제거된 얼굴의 특징으로부터 얼굴 인식을 수행한다. 그러나 폐색영역으로 인해 현실에서 얼굴인식의 성능은 매우 저하된다. 첫째, 선글라스, 모자, 마스크 등으로 인해 얼굴의 특징을 파악하기 어려우며 외부 객체로 인해 얼굴인식 성능이 매우 저하된다. 둘째, 특정 물체에 의해 얼굴이 반 이상 가려지는 경우 얼굴인식을 위한 데이터 확보가 어렵다. 이러한 다양한 폐색영역으로 인한 얼굴인식 성능 저하를 효과적으로 다루기 위해 본 논문에서는 다양한 폐색영역 특징을 적극적으로 반영하는 얼굴인식을 위한 딥 러닝 기반 폐색영역 인식 네트워크 (deep learning based occlusion-aware net for face recognition)을 제안한다. Figure. 2는 우리의 폐색영역 인식 네트워크 파이프라인을 보인다. 폐색영역에 독립적인 얼굴인식시스템을 위해 학습과정에서 데이터셋에 폐색영역을 추가하여 네트워크가 폐색영역이 있는 얼굴을 인식하는 것을 학습한다. 이 때 폐색영역을 만들어내는 마스크 생성기는 FC layer(Fully connected layer) 앞에 위치하여 배경과 폐색영역에 대한 마스크 정보를 이용하여 convolution layers에서 출력된 Activations에서 폐색영역을 버린 후 FC layer를 거쳐 얼굴 인식을 수행한다. 이러한 과정을 통해 폐색영역에 독립적인 얼굴 인식이 가능하다. 그러나 기존 [1] 연구에서는 폐색영역을 파악하기 위해 세그멘테이션 기법을 사용하고 있다. 그러나 이러한 방법은 메모리, 계산 비용을 상승시켜 latency와 성능 간 트레이드 오프를 저하시킨다. 우리는 빠른 추론 속도로 얼굴 인식 리얼타임이 수행 가능한 마스크 생성기를 제안한다.

3.1. Mask Generator

그림 2는 제안하는 마스크 생성기의 예시이다. 우리는 이미지를 NxN 블록으로 나누어 이미지에 폐색영역을 적용한 마스크를 만들어낸다. 이때 각 블록마다 임베딩을 적용하여 마스크정보를 따로 룩업테이블에 저장하도록 한다.

4. Expected results/improvements

4.1. Expected results

이 논문에서 우리는 폐색영역을 미리 고려한 학습을 통해 폐색영역으로 인한 얼굴인식 성능저하를 최소화하고자 한다. 또한 기존 연구와 다르게 마스킹 계산에 대한 복잡도가 매우 낮아 리얼타임을 충족할 것이다. 또한 임베딩을 통한 록업테이블을 만들어 마스크정보를 저장할 것이다. 우리가 제안하는 마스크 생성기는 폐색영역이 주는 오류정보를 효율적으로 제거하면서 보다 더 일반화된 성능을 제공한다.

4.2. improvements

기존의 폐색영역이 존재하는 얼굴영상을 처리하는 방법에는 크게 Occlusion aware face recognition, Occlusion recovery based face recognition 방식이 존재한다. 본 논문에서는 폐색영역 인지 기반의 얼굴인식을 수행하였으나 현재까지의 실험상으로는 폐색영역을 구별하여 네트워크를 구성하는 것만으로는 기존의 네트워크와 비교하여 큰 성능향상을 확인하기 힘들었다. 이에 따라 폐색영역 인지를 바탕으로 부분영역만으로 추출한 특성을 활용하여 네트워크 개선이 필요하다는 결론을 갖게 되었다. 마스크 생성 모듈 및 전체 네트워크를 다양한 데이터셋에서 훈련시키면서 결과를 확인하여 개선방향에 대해서 확인 예정

References