

# 마스크 착용 얼굴인식

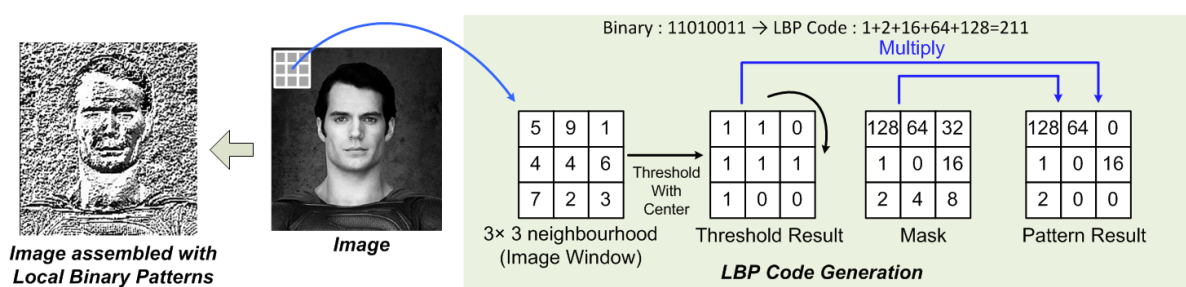
## Abstract(+ LBP feature에 대해)

- 마스크에 가려진 부분을 탐지하고, 가려지지 않은 부분의 LBP feature를 찾기 위해
- MTCNN을 통해 LBP feature를 찾고, 그 feature를 SVM의 입력으로 사용하여 얼굴인식을 수행할 예정

## LBP(Local Binary Pattern) feature?

- 얼굴 이미지의 질감을 효과적으로 나타내고 얼굴 인식의 강건성을 향상시키는 feature를 의미

절차 - 이름처럼 지역적인(Local) 이진(Binary) 패턴(Pattern)을 계산



- 좌측 흑백 그림의 한 3 \* 3 사각형 영역을 살펴보았을 때, 9개의 픽셀값이 존재
- 9개의 값 중 중앙에 위치한 값(4)을 기준으로, 주변 8개의 값들과 비교하여 4보다 크거나 같으면 1, 작으면 0으로 이진화 한다.
- 이진화된 9개의 값들을 중점을 제외하고, 좌측 상단부터 시계방향 순서대로 숫자를 하나씩 추출하여 2진수를 얻는다.
- 2진수를 10진수로 변환한 뒤, 해당 값을 3 \* 3 사각형 영역의 픽셀값으로 대체한다.

참조:

- <https://www.youtube.com/watch?v=h-z9-bMtd7w>
- <https://ckyrkou.medium.com/object-detection-using-local-binary-patterns-50b165658368>

## Face Detection And Occlusion Area Detection Caused by Masks

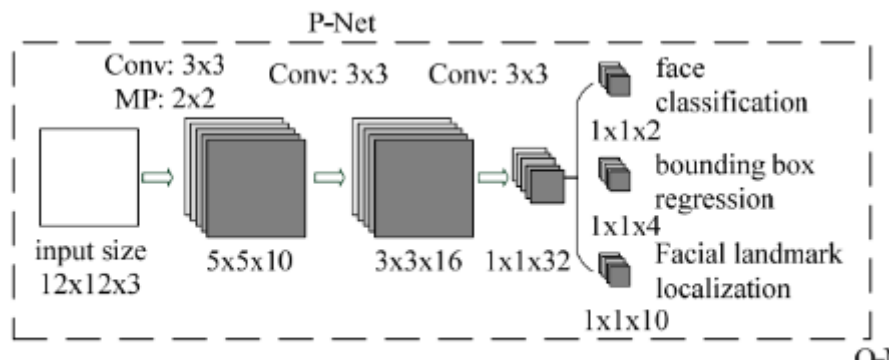
- 마스크 착용 여부를 탐지하기 위해서는 얼굴탐지 결과 상에서 마스크 탐지가 수행되어야 한다.
- 오리지널 MTCNN 알고리즘은 얼굴탐지를 위해 고안된 프레임워크이므로 얼굴과 마스크탐지에 그대로 적용하기에는 부적절하다. 따라서 이 논문에서 개선된 버전의 MTCNN 알고리즘을 제안한다.

### A. (original) MTCNN 얼굴탐지 방법

# (Multi-task Cascaded Convolutional Neural Networks)

- 빛, 각도, 얼굴표정변화에 강건하다.
- 3개의 Coarse to fine subnets으로 구성

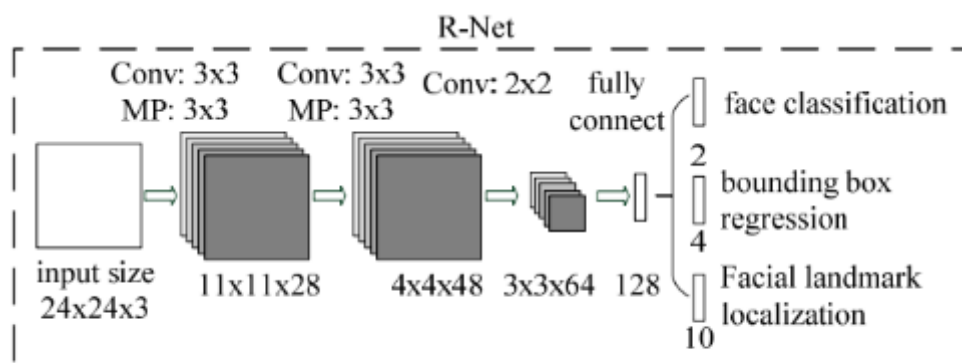
## 1. P-net



P-Net Structure

- 임의의 사이즈로 주어지는 얼굴을 탐지하기 위해서 인풋 데이터를 단계별로 resize하여 이미지 피라미드를 만든다.
- 다양한 스케일의 이미지로부터 얼굴을 탐지한다.
  - 예) 300\*200 크기의 이미지가 입력되면 이를 200x166, 100 x 66, 30 x 20 크기로 리사이즈한 이미지의 list를 만든다. (작은 얼굴도 검출하기 위해)
- 12\*12\*3 크기의 작은 이미지를 입력으로 받아 convolution layer만을 거쳐, 해당 영역이 얼굴인지를 나타내는 값들을 리턴
  - face classification
  - 얼굴 영역의 좌측 상단 꼭지점 x, y
  - 박스와 너비, 크기를 나타내는 4개의 bounding box regression 값
  - 양쪽 눈, 코, 입꼬리 x, y 좌표를 나타내는 10개의 landmark localization 값
- 12\*12의 윈도우를 설정하여 작은 얼굴도 잘 탐지하며 찾은 얼굴 영역을 원래 이미지 크기로 되돌린다.
- 탐지한 박스들을 대상으로 Non-Maximum-Suppression(NMS)과 Bounding box regression(BBR)을 적용

## 2. R-net

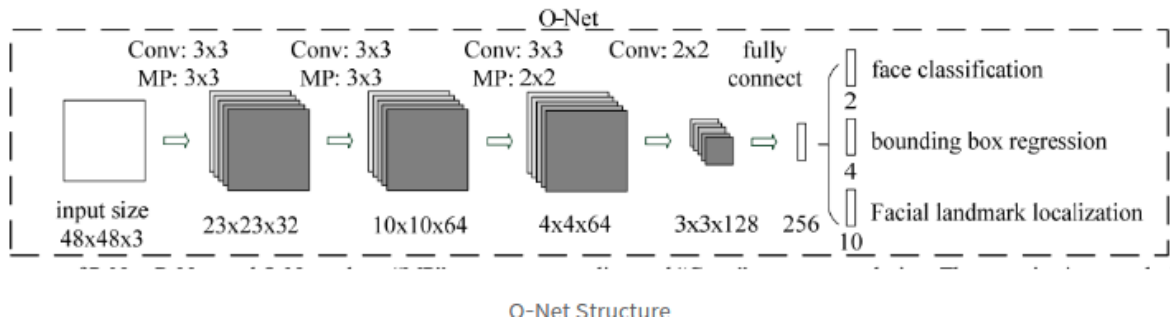


R-Net Structure

- P-net으로부터 도출된 후보 바운딩박스 중 전체 얼굴을 커버할 수 있는 바운딩박스를 추린다.

- 앞서 구한 박스들을 24\*24크기로 조정한 뒤, R-net을 통해 bounding box regression을 더 정교하게 수행하는 것.
- P-net과는 다르게 Fully connected layer를 사용
- p-net과 마찬가지로 찾아낸 박스를 원래 크기로 되돌린 다음, NMS와 BBR을 수행

### 3. O-net



- R-net을 통해 찾아낸 박스들을 모두 48\*48크기로 조정한 뒤, R-net으로부터 도출된 바운딩박스에 추가적인 adjustment를 수행한다.

이와 같은 과정을 통해 얼굴이 정확히 탐지되며, 마지막 두 subnets이 전체 이미지를 스캔할 필요가 없어 다른 알고리즘에 비해 속도가 빠르고 효과적이다.

참조:

- <https://yeomko.tistory.com/16>

## B. 개선된 MTCNN에 대한 전반적 구조

### B. The overall architecture of the improved MTCNN

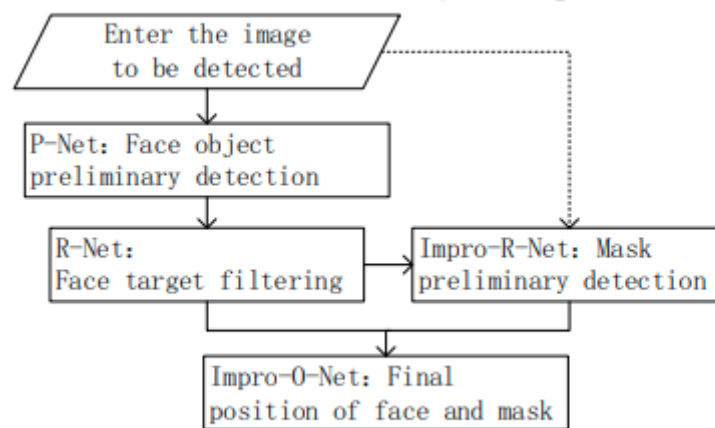


Fig. 1. Overall network architecture of improved MTCNN algorithm

- 마스크 탐지를 위한 프로세스가 추가되었다.
- P-net은 얼굴에 대한 preliminary extraction을 수행
- R-net은 P-Net이 잘못 탐지된 얼굴타겟을 필터링할 때 얼굴타겟이 적절한지 확인해주는 작업을 수행
- Impro-R-Net은 남아있는 얼굴 타겟들 중, 마스크 타겟을 초기 추출작업을 수행
- Impro-O-Net은 얼굴과 마스크 타겟을 최종 결정

## C. 마스크 탐지에 기반한 (개선된) R-Net

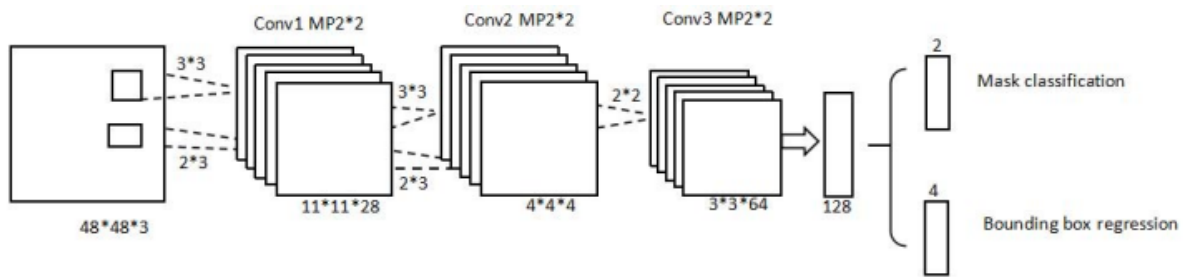


Fig. 2. Impro-R-Net network structure

- Fully connected layers를 가진 일반적인 CNN 네트워크
- 3개의 합성곱 네트워크로 구성되며(Conv1, Conv2, Conv3), 각각 maximum pooling layers에 연결됨
- 여러 종류의 마스크에 적용되기 위해 개선된 R-net은 세 합성곱 네트워크에 대해 1:1, 2:1 영상비(aspect ratio)의 커널을 사용
- 다음과 같은 두 종류의 아웃풋을 출력
  - the classification of the target area
  - the regression of the bounding box of the target area
- 마스크와 얼굴의 shape와 scale이 다르기 때문에, 해당 단계에서 정확한 얼굴을 가려내고 마스크를 처음 추출하기 위해서 두 가지의 conv layer(R-Net, Impro-R-Net)가 사용되며, 두 가지 아웃풋이 출력된다.
  - R-net은 얼굴을 필터링 후 -> 남은 얼굴들 출력
  - Impro-R-net은 얼굴 내의 마스크를 탐지 -> 탐지된 마스크 출력

## D. 피라미드 풀링에 기반한 (개선된) O-Net

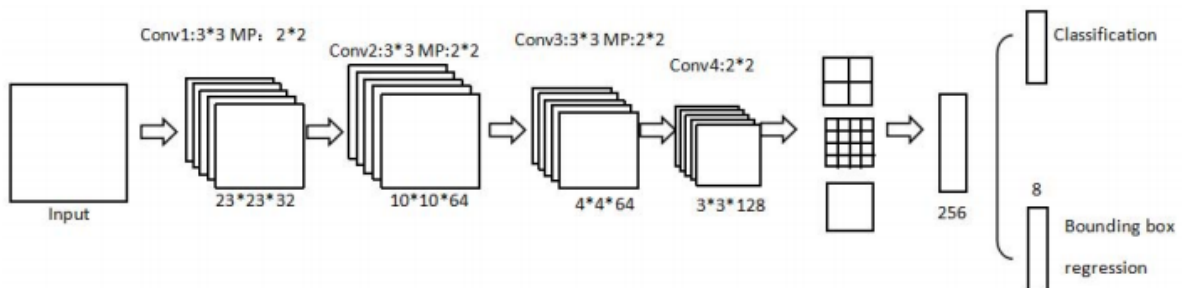


Fig. 3. Impro-O-Net network structure

- 마스크와 얼굴의 shape와 scale이 다르기 때문에, fully connected layers가 있는 일반적인 conv layer가 사용된다면 얼굴과 마스크는 스케일링 되어야하고, 이것은 이미지를 왜곡시켜 분류 정확도에 영향을 끼칠 것이다.
- 또한 두개의 CNN이 얼굴과 마스크를 각각 분류한다면 전체 알고리즘의 효율성이 떨어질 것이므로 개선된 O-net을 제안하였다.
- 네트워크가 다른 스케일의 인풋을 받기 위해서 피라미드 풀링에 기반한 합성곱 신경망을 디자인하였다.
- 다른 스케일의 피쳐맵을 고정된 차원으로 정규화하는 RollPooling layer라는 새로운 층을 통해, 이미지를 스케일링할 필요 없이 마스크와 얼굴 이미지를 같은 네트워크상에서 공유할 수 있다. (따라서 스케일링으로 인한 이미지 왜곡이 발생하지 않는다.)
- 다음의 두 가지 아웃풋을 출력하며 각각 다른 타입을 포함한다.
  - classification output
    - 얼굴, 마스크, 배경에 대한 likelihood score를 포함
  - regression output
    - 얼굴, 마스크에 대한 경계(border) fine-tuning을 포함

- border는 x, y, w, h로 나타나며 x, y는 인풋의 중점에 대한 오프셋이며 w, h는 인풋 경계에 대한 스케일링 정도를 나타냄

### 피라미드 풀링

- 3-way pooling method가 적용되었으며, pooling ratio는 각각 1\*1, 2\*2, 4\*4이다.
- 인풋 피쳐맵의 사이즈에 관계없이 피라미드 풀링 적용 후에는  $(1 + 4 + 16) * 128$ 차원의 피쳐가 생성된다.

## 가려지지 않은 부분에 대한 Texture feature 추출과 얼굴 인식

---

- 가려지지 않은 부분에 대해서 Local Binary Patterns를 적용하여 texture feature를 얻는다.
- 이렇게 얻어진 feature를 서포트 벡터머신의 입력으로 사용하여 얼굴인식을 수행한다.