

MTCNN을 활용한 군중 속 특징 추출

박진우* · 김민주 · 김시현 · 장동환 · 이성진 · 문상호
부산외국어대학교

Extracting Feature in the Crowd using MTCNN

Park-jin Woo* · Minju Kim · Sihyun Kim · Donghwan Jang · Sung-jin Lee · Sang-ho Moon
Busan University of Foreign Studies
E-mail: pjwpjw715@gmail.com

요 약

2020년에 경찰청에서 조사한 자료에 따르면 38,496건 중에 미해결 사건이 161건으로, 미해결 원인 중 가장 높은 성인실종자의 경우는 대부분 단순 가출로 평가된다. CCTV를 통해 영상을 통해 찾는 경우에도 수많은 사람이 지나가는 상황에 인물들의 얼굴을 일일이 확인하고 진술의 특징들로만 인물을 찾아내야 하기 때문에 긴 시간이 걸리고 정확도가 낮아 수사에 많은 시간이 소요되는 것이 확인되었다. 본 논문은 MTCNN을 활용하여 CCTV 속 인물 추출에 대한 연구를 진행하였다. MTCNN으로 학습된 얼굴들과 입고 있는 옷의 특징을 동시에 분석을 시작하여 겹쳐지는 경우의 인물들만 추출하여 관계자에게 확인이 가능하도록 한다. 향후 실종자의 특징들을 좁혀 정확도를 높이기 위해 더 다양한 특징 검출 학습을 목표로 한다.

ABSTRACT

According to the National Police Agency, 161 out of 38,496 unsolved cases as of 2020. Most of the adult missing persons, the highest of the unsolved causes, are evaluated as simple runaway, which takes a long time to investigate. Even if search through CCTV, it can take a long time and the accuracy can be somewhat low because you have to check the faces of the characters one by one and find the characters only with the characteristics of the statements.

This paper utilizes MTCNN to conduct research on character extraction in CCTV. We initiate simultaneous analysis of the features of faces learned with MTCNN and the clothes we are wearing, so that only the overlapping characters are extracted so that they can be identified to the related parties. For aim to learn more diverse feature detection to narrow down the features of missing persons in the future and increase their accuracy.

키워드

MTCNN, Character extraction, CCTV, Feature

1. 서 론

2020년 경찰청에서 조사한 자료에 따르면 해마다 정상아동, 지적장애, 치매환자의 실종 신고접수건은 근 5년 이래 38,000건을 넘어간다. 미발견 사례로 2020년도에만 161건으로 전년 대비 약 4.8배를 보이고 있다[1].

경찰청에서는 실종자 탐색을 위해 CCTV를 통한 동선 파악과 실종자 상태를 확인한다. 이 때 사용

되는 일반 CNN모델을 활용한 실종자 탐색의 경우 시각적 변화에 따라 실제 환경에 적용되기 어려움이 있어 인식이 되지 않은 경우 신고 당시 특징을 가지고 경찰의 육안으로 분석을 해야 한다.

본 연구에서는 실종자 탐색에 시간 절약과 높은 인식률을 위해 MTCNN(Multi-Task Cascaded Convolution Neural Network)을 활용한 소프트웨어를 개발하고자 한다.

적은 특징으로도 인식률을 높이기 위해 MTCNN 모델의 여러 규제에 변화를 주어 CCTV의 위치에

* speaker

서 거리에 관계없이 얼굴을 추출하여 실시간으로 보여주는 것을 목적으로 한다.

II. MTCNN

MTCNN은 기존 CNN보다 더 얼굴 인식률을 높이기 위해 제작되었다. 인식된 얼굴을 CNN에서는 Convolution Layer를 통해 연산이 된다. 이 방식은 연산도 복잡하고 영상의 크기에 따라 시간의 소요가 달라진다[2].

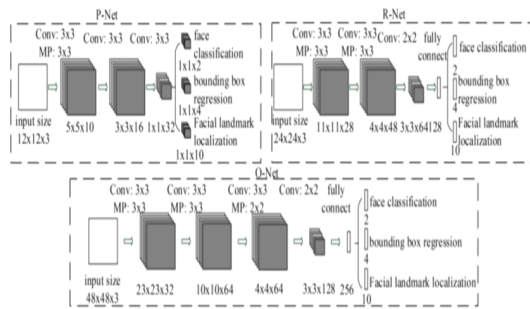


그림 1. MTCNN 구조[2]

그림 1은 MTCNN의 전체적인 구조를 보여주는 그림이다. MTCNN에서는 이미지 피라미드를 통해 인식한 얼굴을 P-Net, R-Net, O-Net이라는 계단식 Network를 걸쳐 인식한다.

이미지 피라미드는 서로 다른 크기의 얼굴을 인식하여 Network에 전달하기 위해서이다. 인식된 얼굴은 12x12 크기로 조정하여 P-Net에 전달한다.

P-Net에서는 12x12크기에 맞춰져 있는 Network이다. Fully Connected Layer를 거치지 않고 Convolution Layer를 통해 인식되어 얼굴이라고 판단된 영역을 원래의 크기에서도 인식이 가능하도록 좌표를 변환한다. 여기서 얼굴이라 인식된 영역에 관한 bounding box가 많이 남아 있으므로 Non-Maximum Suppression(NMS)을 거쳐 중복된 box를 줄인다. 이 때 줄여진 bounding box와 얼굴 분류와 눈, 코와 양 입 고리를 반환한다.

추려진 bounding box들은 24x24 크기로 조정되어 R-Net에 보내진다. R-Net에서는 Convolution Layer를 거쳐 bounding box들을 NMS를 거쳐 중복된 상자를 줄인 후 Fully Connected Layer를 거쳐 O-Net에 전달된다.

48x48크기로 조정되어 전달받은 O-Net에서는 다시 Convolution Layer와 Fully Connected Layer를 통해 신뢰성 낮은 bounding box를 제거함으로써 한 개의 bounding box 남게 된다.

최종적으로 P-Net, R-Net, O-Net 3가지 단계를 모두 거치고 나면 최종 결과 값은 bounding box와 신뢰성과 눈, 코, 입꼬리 값이 전달된다.

III. MTCNN 설계 및 구현

본 논문에서는 위의 표1의 상황에서 진행되고, MTCNN을 사용하여 얼굴 인식률을 높이는 연구를 진행했다. 진행 방식은 이미지 피라미드를 통해 약 12만장의 사진들을 P-Net의 크기에 맞추어 조정된 뒤 얼굴 인식이 되는지의 Label을 정한 뒤 학습을 시킨다. 이 때 활성화 함수에 따른 정확도와 손실률을 대조하여 가장 효율적인 활성화 함수를 고르도록 한다.

표 1. 실험에 사용된 환경

OS	Windows 10 64bit
CPU	Intel(R) Core(TM) I5-8250U
RAM	8GB
Language	Python 3.7.10
Package	Tensorflow-gpu 1.15.2
GPU	NVIDIA 10.0.0

준비된 사진들을 12x12크기에 맞게 조정된 뒤 얼굴을 인식시켜 해당 부분에서 5가지 특징점을 잡아준다. 잡아준 데이터들은 딥러닝을 위해 얼굴 인식 데이터와 특징 데이터를 병합한 뒤, 수치화한다.

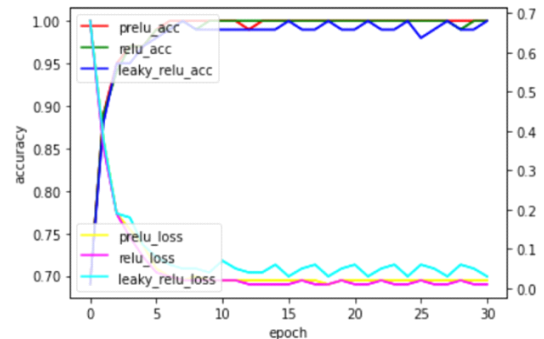


그림 2. 활성화 함수에 따른 P-Net결과

표 2는 활성화 함수에 따른 정확도와 손실률을 비교한 결과물이다. P-Net에서 3가지 활성화 함수 중에서도 Relu, Leaky Relu, PRelu를 비교한 결과 정확도와 분류 손실률이 가장 낮은 Relu를 통해 나온 모델이 가장 적합하다. 경계 상자의 손실률은 R-Net과 O-Net에서 선정되어 걸러질 예정이라 P-Net에서 나온 손실률은 최종모델에 큰 영향을 미치지 않을 것으로 판단된다.

IV. 결 론

본 논문에서는 MTCNN 중에서도 P-Net의 활성화 함수의 조정을 통해 최적의 P-Net의 구현에 대해 연구했다. 이미지 피라미드를 거쳐 나온 P-Net을 통하여 나온 결과 값은 매우 높은 정확도와 낮은 손실률을 보이고 있다. 해당 결과물은 R-Net과 O-Net를 거칠 때 더 높은 정확도를 전달하는 것에 큰 영향을 줄 것으로 본다.

본 연구는 최종적으로 R-Net과 O-Net의 조정을 통해 높은 정확도를 보이기 위해 연구될 예정이고 해당 모델은 CCTV와 같이 많은 연산을 소화하지 못하는 마이크로컨트롤러 유닛에 적용하여 더 우수한 인식률과 빠른 연산을 위해 연구할 예정이다.

References

- [1] Korea Police Agency. 실종아동등 신고접수 및 처리 현황 [Internet]. Available : https://www.index.go.kr/potal/main/EachDtlPageDetail.do?idx_cd=1610
- [2] Kwang-Min Nam, Yong-Jin Jung, “Cascade CNN with CPU with CPU-FPGA Architecture for Real-time Face Detection”, Vol. 21, No. 4, pp. 388-396, April. 2017.
- [3] Kaipeng Zhang, Zhanpeng Zhang,, Zhipeng Li. “Joint Face Detection and Alignment using Multi-task Cascaded Convolution Network” in *IEEE Signal Processing Letters*, Vol. 23, No. 16290010, pp. 1499-1503, August, 2016.