

MobileNetV2를 이용한 고 밀집 실내환경에서의 사람 검출 시스템 기법

최수정, 임유진
숙명여자대학교 IT공학과
suzzang77@sookmyung.ac.kr, yujin91@sookmyung.ac.kr

Human Detection System in High Density Indoor Environment Using MobileNetV2

SooJeong Choi, Yujin Lim
Dept. of IT Engineering, Sookmyung Women's University

요 약

최근 인공지능 기술의 발달에 따라 여러 분야에 인공지능 기술이 활발히 응용되고 있다. 그중 안전 관리 분야에서 사람 인식을 통한 안전 관리 시스템의 지속적인 개발이 요구되고 있다. 그러나 실내 한정된 공간에서 사람들의 밀집도가 높은 경우 오브젝트의 중복도가 높아져 인식 성능이 낮아질 수 있다. 이를 해결하기 위해 본 논문은 사람의 밀집도가 높은 실내 환경에서 기존 객체 인식 기법의 성능을 분석하였다. 그리고 이러한 제한적인 환경에서 최적의 좋은 성능을 보일 수 있는 SSDLite와 MobileNetV2 모델을 기반으로 soft-NMS 기법을 적용하여 성능을 분석하였다.

1. 서론

최근 들어 인공지능 기술이 발달하면서 다양한 분야에 적용되고 있다. 그중에서도 안전 관리 분야에 대한 인공지능 기술의 적용이 많은 관심을 받고 있다. 예를 들어, 공사 현장에 인공지능 기술을 도입한 스마트 안전 관리 시스템[1]과 건물의 전반적인 모니터링을 담당하는 스마트 빌딩 시스템[2] 등을 들 수 있다.

최근 산업재해 유형별 분포도[3]를 보면 ‘넘어짐’이 약 20,659명으로 가장 많았고 다음으로 ‘업무상 질병’, ‘떨어짐’ 순으로 나타났다. 전년 대비 감소율이 약 0.79%로 크지 않은 것으로 보아 작업자의 안전을 위한 시스템이 지속해서 개발될 필요가 있다. 그뿐만 아니라 전 세계적으로 스마트 빌딩 시장[4]이 점차 커지는 경향이므로 사람 인식 기술을 통한 안전 관리의 필요성 또한 증가할 것으로 예상된다.

하지만 그림1에서처럼 실내의 한정된 공간에서 오브젝트(사람)들의 밀집도가 높은 경우 이미지 내 오브젝트 중복성이 높아져 인공지능 기반 객체 인식 알고리즘의 성능이 낮아질 수 있다. 이를 해결하기 위한 기법 중 하나인 NMS(non-max

suppression)[5]는 이미지에서 다양한 크기의 객체들을 인식한 bounding box들 중에서 가장 정확한 box 하나를 선택하고 나머지 box들은 억제하는 기법이다. 반면 soft-NMS[5]는 box를 억제하는 것이 아니라 confidence를 감소시켜 중복된 오브젝트 검출에 더 용이하다.



그림 1. 제한된 공간에 밀집된 사람들

기존의 연구는 환경의 제한 없이 soft-NMS와 NMS의 객체인식 성능을 비교했는데 그 결과 약 2% 정도의 AP(Average Precision) 차이가 난다는 걸 알 수 있다[5]. 이에 본 논문에서는 실내 안전 관리를 위한 시스템을 고려하여 오브젝트들의 밀집도

가 높은 제한된 환경에서의 기존 객체인식 기법들의 성능을 분석하였다. 실내 환경으로 가정했기 때문에 검출되는 오브젝트는 사람으로 한정했다. 또한 객체인식 모듈은 원거리에 위치한 서버가 아닌 실내 공간 내 카메라와 함께 설치되어 있는 프로세싱 유닛에서 작동한다고 가정했다.

따라서 본 논문에서는 객체 인식률이 높으면서도 상대적으로 요구하는 처리량이 적은 SSDLite[6]와 MobileNetV2[6] 모델을 기반으로 이미지 내 중복성이 높은 객체들의 인식을 향상을 위해 soft-NMS[5]를 적용하여 다양한 실험 시나리오에서 객체인식 성능을 분석하였다.

2. 성능 분석

본 논문에서는 오브젝트 중복성이 높은 데이터셋인 Crowdhuman Dataset[7]을 이용하여 실험을 진행했다. 1000개의 이미지 데이터에서 사람들의 중복성 정도를 기반으로 상대적으로 낮은 1단계부터 4단계로 나눠서 실험했다. 4단계로 갈수록 사람의 수가 많아지면서 이미지가 겹치는 사람의 수도 증가한다. 그리고 객체인식 모델로는 tensorflow에서 제공하는 COCO pre-trained model인 SSDLite MobileNet V2[8]을 사용하였다. 이미지 내의 객체 중복성 정도에 따른 성능을 soft-NMS Gaussian 기법과 Linear 기법으로 나누어 실험하였다.

Precision은 오브젝트를 인식한 결과가 얼마나 실제 오브젝트와 일치하는지를 나타내는 지표이고, Recall은 실제 오브젝트를 기준으로 얼마나 인식했는지를 나타내는 지표이다. 아래 식에서 TP는 True Positive를 나타내며, FP는 False Positive를, FN은 False Negative를 나타낸다.

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

그림2는 Precision을 측정된 결과로 3단계 데이터를 사용했을 때 Precision이 상대적으로 높게 나왔다. 그리고 Gaussian 기법으로 측정했을 때의 TP가 Linear 기법보다 약 4%-14% 정도 증가했지만, FP 또한 약 13%-19% 정도로 보다 더 증가했기 때문에 전체적으로 Linear 기법이 Gaussian 기법보다 높게

나왔다.

그림3은 Recall을 측정된 결과로 1단계 데이터를 사용했을 때 각 기법의 Recall이 높게 나온 것을 볼 수 있다. 식(2)에서 TP와 FN을 더한 것은 실제 이미지 내의 오브젝트의 개수를 의미한다. 이미지마다 실제 오브젝트의 개수는 항상 일정하므로 Gaussian 기법으로 측정했을 때의 TP가 크므로 식(2)에 의해 전체적으로 Gaussian 기법이 Linear 기법 보다 높게 나왔다.

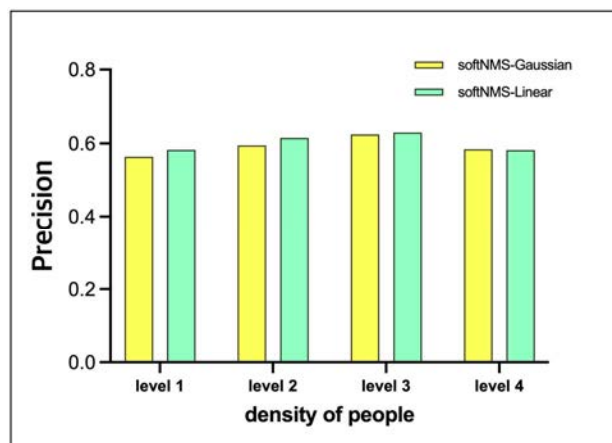


그림 2. 이미지내 사람 중복도에 따른 Precision 측정 결과

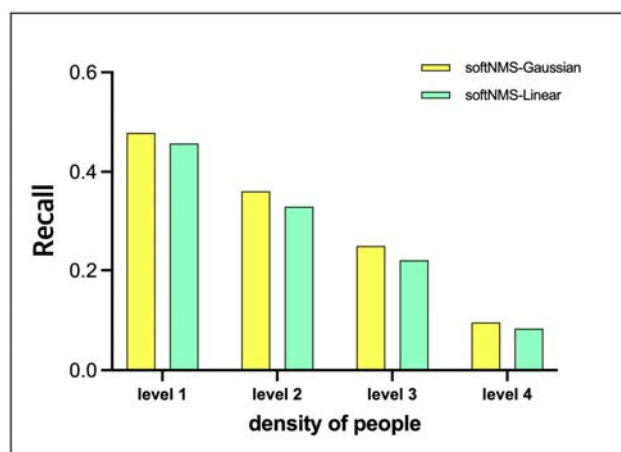


그림 3. 이미지내 사람 중복도에 따른 Recall 측정 결과

Precision과 Recall은 서로 상반관계에 있으므로 전체적인 객체인식 시스템의 성능을 평가하기는 어렵다. 그림4는 이를 해결하기 위한 Precision과 Recall 값을 종합하여 알고리즘을 평가하는 AP를 나타내고 있다. 객체 이미지 중복성 정도에 따른 person AP를 soft-NMS Gaussian 기법과 Linear 기법을 이용했을 때로 나누어 나타내었다. 중복 정도가 가장 낮

왔던 1단계에서 가장 인식률이 좋았고 Gaussian 기법이 약 35.49%, Linear 기법이 약 33.97%였다. 전체적으로 Gaussian 기법이 Linear 기법 보다 인식 성능이 높은 걸 확인할 수 있다.

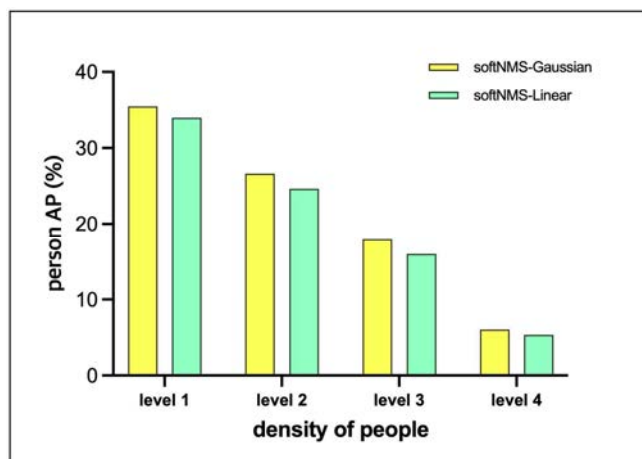


그림 4. 이미지 내 사람 중복도에 따른 AP 측정 결과

3. 결론 및 향후 연구

본 논문은 사람들의 밀집도가 높은 제한된 공간의 환경에서 객체인식 기법으로 soft-NMS를 사용하여 사람 인식 성능을 분석하였다. 객체가 중복된 환경에서 인식 성능이 NMS보다 성능이 좋은 것으로 알려져 있는 soft-NMS를 적용하여 실험을 하였다. 그러나 사람의 밀집 단계가 높아질수록 인식률이 현저히 떨어지는 것을 알 수 있었다.

soft-NMS 외에도 adaptive-NMS 및 NOH-NMS와 같은 다양한 NMS 기법이 존재한다[9]. AP 측면에서 NOH-NMS가 가장 높았고, 그다음으로 Adaptive-NMS, Soft-NMS 순으로 높았다. 함께 사용된 모델이 MobileNet이 아니라 ResNet인 걸 감안하더라도 NOH-NMS의 성능이 좋다는 걸 알 수 있다.

향후 연구에서는 다양한 NMS 기법들을 참고하여 제한된 공간에서 사람들이 밀집되어 있을 때 사람 인식률을 높이는 알고리즘을 연구하고자 한다.

사사문구

본 연구는 2022년도 중소벤처기업부의 기술개발사업 및 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (S3247940, No. 2021R1F1A1047113).

참고문헌

- [1]내 손안에 서울, "AI가 위험 감지...민간 공사장·건축물에 스마트 안전관리 도입," 2021.09.13.
- [2]Asem Elshimi, "변화하는 스마트 빌딩... IoT를 활용한 에너지 소비 혁신," 테크월드뉴스, 2022.09.09.
- [3]고용노동부, "2020년 산업재해현황분석"
- [4]연구개발특구진흥재단, "글로벌 시장동향보고서-스마트 빌딩 시장," 2021.10.
- [5]N. Bodla, B. Singh, R. Chellappa and L. S. Davis, "Soft-NMS - Improving Object Detection with One Line of Code," IEEE International Conference on Computer Vision (ICCV), pp. 5562-5570, Venice, 22-29 October 2017.
- [6]M. Sandler, A. Howard, M. Zhu, A. Zhmoginov and L. -C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4510-4520, Salt Lake City, UT, USA, 18-23 June 2018.
- [7]CrowdHuman - A Benchmark for Detecting Human in a Crowd, <https://www.crowdhuman.org>
- [8]TensorFlow 1 Detection Model Zoo, https://github.com/tensorflow/models/blob/master/research/object_detection/g3doc/tf1_detection_zoo.md
- [9]P. Zhou, C. Zhou, P. Peng, J. Du, X. Sun, X. Guo, and F. Huang, "Noh-NMS: Improving Pedestrian Detection by Nearby Objects Hallucination," ACM International Conference on Multimedia, pp. 1967 - 1975, Seattle WA USA, 12 October, 2020.