

<https://doi.org/10.7236/JIIBC.2024.24.3.177>
JIIBC 2024-3-25

YOLOv4를 이용한 CCTV 영상 내 군중 밀집도 분석 서비스 개발

Development for Analysis Service of Crowd Density in CCTV Video using YOLOv4

황승연*, 김정준**

Seung-Yeon Hwang*, Jeong-Joon Kim**

요약 본 논문에서는 2022년 10월 29일 한국에서 발생한 이태원 압사 사고를 기반으로 미래에 발생할 수 있는 인파 사고에 대하여 군중 밀집으로 인한 위험을 미리 예측하고, 예방하기 위한 목적으로 작성되었다. 단일 CCTV 같은 경우 관리자가 실시간으로 현재 상황을 판별할 수 있지만, 하루 종일 해당 화면만 들여다볼 수 없기 때문에 CCTV 화각으로 촬영된 영상들을 학습한 YOLO v4를 이용하여 객체를 탐지하고, 정해진 군집의 수가 초과하는 순간에 알림을 통해 군중 밀집으로 인한 안전사고를 예방하게 된다. YOLO v4 모델을 사용하게 된 이유는 이전 YOLO 모델보다 더욱 높은 정확성과 빠른 속도로 개선되어, 객체 탐지 기법이 더 용이해졌기 때문이다. 본 서비스를 AI-Hub 사이트에 등재된 CCTV 영상 데이터로 테스트하는 과정을 거치게 된다. 현재 한국에 CCTV는 기하급수적으로 증가하였고, 이를 실제 CCTV에 적용한다면 앞으로 일어나게 될 군중 밀집으로 인한 사고를 비롯한 다양한 사고를 예방할 수 있을 것으로 기대한다.

Abstract In this paper, the purpose of this paper is to predict and prevent the risk of crowd concentration in advance for possible future crowd accidents based on the Itaewon crush accident in Korea on October 29, 2022. In the case of a single CCTV, the administrator can determine the current situation in real time, but since the screen cannot be seen throughout the day, objects are detected using YOLOv4, which learns images taken with CCTV angle, and safety accidents due to crowd concentration are prevented by notification when the number of clusters exceeds. The reason for using the YOLO v4 model is that it improves with higher accuracy and faster speed than the previous YOLO model, making object detection techniques easier. This service will go through the process of testing with CCTV image data registered on the AI-Hub site. Currently, CCTVs have increased exponentially in Korea, and if they are applied to actual CCTVs, it is expected that various accidents, including accidents caused by crowd concentration in the future, can be prevented.

Key Words : CCTV, Crowd Density, Deep Learning, Object Detection, YOLOv4

*준회원, 안양대학교 컴퓨터공학과
**정회원, 안양대학교 ICT융합학부 소프트웨어전공
접수일자 2024년 4월 1일, 수정완료 2024년 5월 9일
게재확정일자 2024년 6월 7일

Received: 1 April, 2024 / Revised: 9 May, 2024 /

Accepted: 7 June, 2024

**Corresponding Author: jjkim@anyang.ac.kr

Dept. ICT Convergence Engineering at Anyang University, Korea

I. 서 론

지난해 2022년 10월 29일, 이태원 대규모 압사 사고가 발생하여 총 159명의 사망자(사고 발생 시 사망자 : 156명, 사고 이후 사망자 : 3명)와 196명의 부상자가 발생하였다. 사고가 발생한 해밀톤 호텔 인근 골목은 길이 325m, 폭 5~7m에 불과하였고, 사건 발생 이태원 일대 10만여의 인파가 몰린 것으로 추정되었다.

이태원 참사를 수사했던 경찰청 특별수사본부가 당시 모습을 담은 폐쇄회로 CCTV를 공개하면서 어느 순간 밀집도와 정체가 심해져 통행이 불가능해지고 사람들이 동시다발로 넘어지는 등 사건의 전반적인 모습을 확인할 수 있었다.

이번 이태원 대규모 참사와 같은 압사 사고는 이번이 처음이 아니다. 1990년 이슬람 순례자들이 영국이며 1426명이 압사하는 역대 최악의 압사 사고가 발생했고 같은 달, 인도네시아 축구 경기장에서 125명이 압사당하여 사망자가 생겼다. 더 이상 이러한 끔찍한 사고가 발생되지 않기 위해서 다양한 방안이 모색되어야 한다고 생각했다.

최근 딥러닝 기술에 대한 연구는 나날이 발전해 나가고 있으며 그중에서도 객체를 인식하고 검출하는 연구가 활발히 진행되고 있다. 본 논문은 위와 같이 CCTV를 통해 현장 상황에서 발생할 수 있는 안전사고를 미리 사전에 예측하고 예방할 수 있는 방안을 마련하고자 한다. 제안하는 검출 모델은 대표적인 객체 탐지(Object Detection) 방법 중 하나인 YOLOv4^[1]를 활용하여 CCTV와 Deep Learning을 이용하여 객체를 판별하고 일정 수준의 인구가 넘어갈 경우 알림을 울려 사고를 방지할 수 있다.

본 논문의 구성은 다음과 같다. 2장에서는 객체 탐지 방안인 YOLOv4 방안을 비롯한 관련 연구에 대해 소개하고, 3장에서는 학습 과정과 시스템 설계 및 구현 과정에서 발생한 문제점 등을 설명한다. 4장에서는 구현한 시스템을 기반으로 결과를 분석, 평가하고 끝으로 5장에서는 논문을 요약하고 향후 연구 방향을 제시하는 결론을 맺는다.

II. 관련연구

1. YOLO v4 (You Only Look Once)

최근의 딥러닝 모델들은 실시간으로 동작하지 않고 학습할 때 많은 수의 GPU가 필요하고 mini-batch의 크기가 크다는 문제점이 야기되었다. YOLO v4는 기존의

YOLO v3 이후에 딥러닝의 정확도 및 속도 개선하고 기존의 사용하던 GPU만으로 빠른 학습이 가능하여 더 좋은 성능을 낼 수 있도록 극대화한 객체 검출 모델이다.

본 연구에서는 CCTV 기반의 영상 데이터로부터 YOLO v4를 이용하여 학습시키고, 학습된 YOLO v4를 통해 인구수를 판별하여 미리 정해놓은 위험 숫자를 넘겼을 때 자동으로 경고가 울릴 수 있게끔 설계하고자 한다. 그러기 위해 객체 검출 정확도가 높고 무엇보다 Real-time 동작이 되는 해당 모델을 이용하여 연구를 진행하였다.

다음 그림 1은 본 논문에서 제안하는 YOLO v4와 최신 딥러닝 기법을 비교하여 나타낸 것이다. 테스트 환경에서 65FPS의 추론 속도와 함께 MS COCO 데이터셋에서 43.5%의 평균 정밀도를 얻은 것을 볼 수 있다.

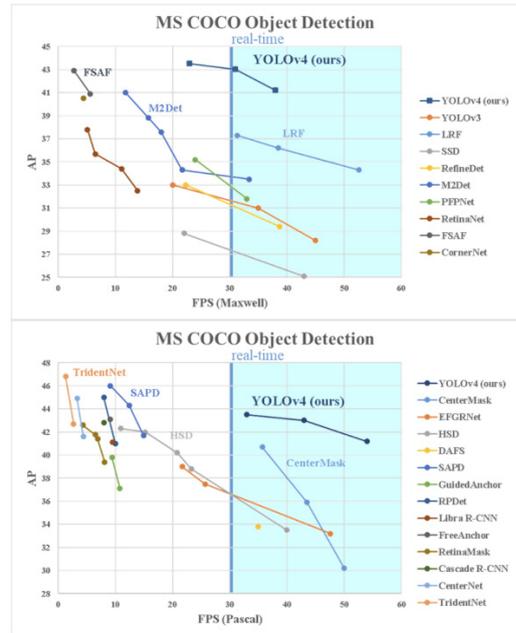


그림 1. FPS-AP 그래프

Fig. 1. FPS-AP graph

2. COCO Dataset (Common Objects in Context)^[2]

Object Detection 분야에서 대표적으로 PASCAL VOC, MS COCO 등이 있는데, COCO Dataset을 이용하여 본 YOLO 모델을 검증하였다. COCO Dataset은 Object Detection뿐만 아니라, Segmentation, Keypoint Detection 등을 위해 제공된 컴퓨터 비전 분야의 task를 목적으로 만들어진 데이터셋이다. COCO dataset은 다양한 크기의 물체가 존재하고, 이미지가 특

정 카테고리에 명확히 속해있지 않는다. 또한, 80개 class를 학습한 pre-trained weight를 제공하는데 이를 사용할 시 80개 클래스에 속한 object들에 대한 detection을 추가 학습 없이 사용할 수 있다. 다음 표 1은 본 연구에서의 인구밀집도를 탐지하기 위해 COCO 2017 dataset를 활용한 것을 나타낸다.

표 1. FPS-AP 그래프
 Table 1. FPS-AP graph

Category	Number of objects
TOTAL	164,000
Training set	118,000
Validation set	5,000
Test set	41,000

III. 시스템 설계 및 구현

1. 실험환경

본 연구에서 CCTV 영상 면적에 해당하는 인구밀집도를 탐지하는 방법으로 그림 2와 같은 흐름도를 제안한다. 본 연구에서 CCTV를 통해 보여지는 정해진 면적 공간에서 물체든 사람이든 모든 객체는 과밀하여 압사 사고로 이어질 수 있는 위험이 존재한다. 따라서 면적을 차지하는 모든 물체에 대하여 실시간으로 object를 판별해야 하기 때문에 COCO dataset을 YOLO v4를 이용하여 학습시킨다.

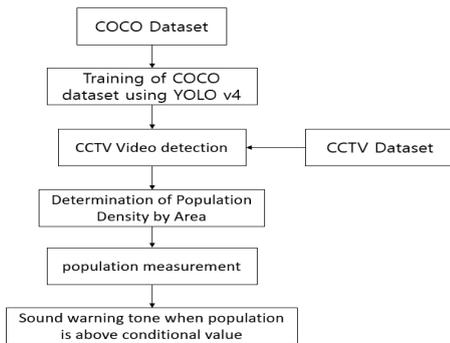


그림 2. 프로세스 흐름도
 Fig. 2. Process flow char

COCO dataset을 학습한 YOLO v4를 활용하여 CCTV 영상 속 객체를 탐지하게 된다면 모든 객체를 바운딩 박스(bounding box)로 표시하고, 지정된 클래스

로 세분화하여 물체 및 인구 탐지를 위해 레이블링 한다. 그다음으로 영상 속 면적에 따른 인구밀집도를 판별하기 위해 안전사고가 발생하기 전 예방할 수 있는 적정 인구수를 정한다. 영상 속 인구수가 선정된 적정 인구수보다 많을 경우 경고음이 울리게 된다.

다음 그림 3은 YOLO v4에 영상을 테스트시켜, 나타낸 output 예시이다. 모든 객체는 다음과 같이 바운딩 박스로 물체의 경계를 나타내고, 레이블로 다음 객체들의 정의와 0과 1사이 소수점으로 정확도를 나타낸다.



그림 3. 각 객체에 따른 바운딩 박스와 레이블
 Fig. 3. Bounding boxes and labels according to each object

2. 데이터 수집 및 전처리

CCTV 속 객체 인식을 위한 데이터를 테스트하기 위해서 AI 기술 및 제품·서비스 개발에 필요한 AI 인프라를 지원하는 AI 통합 플랫폼, AI-Hub에 데이터솔루션이 제작한 ‘유동인구 분석을 위한 CCTV 영상’ 데이터셋^[3]을 사용하였다. 총 7곳의 상권으로부터 CCTV 화각을 고려하여 수집된 데이터셋으로, 촬영 시간 및 촬영 장소의 다양성을 추구하며 비식별화를 통해 개인정보 문제를 해결한 학습용 데이터셋이라고 설명되어 있다. 다양한 장소와 상황에 따른 객체들이 구성되어 있어, 본 논문에서 인구 밀집을 테스트하고자 하는 목적으로 적합하다고 판단하여 해당 데이터셋을 활용되었다. 각 데이터는 크게 Training Data, Validation Data로 나누어져 있고, 레이블링 데이터와 원천데이터 항목으로 구분된다. 본 논문에서는 YOLO 모델로 직접 테스트시키기 위해 Training Data, Validation Data 각 항목의 원천데이터만 사용하여 객체를 탐지한다.

3. 인구밀집도 감지

AI-Hub에서 가져온 데이터에는 7곳의 상권으로부터 CCTV 영상을 수집한 것이기 때문에 다양한 상황에서의

인구수 및 밀집도 등을 확인할 수 있었다. 본 연구에서는 객체 인식이라는 명확한 전제로 유동인구에 따른 객체를 탐지하면 된다. 따라서 이를 탐지하기 위해 객체 탐지 모델인 YOLO v4를 사용한다.

이미지 및 영상에 포함된 객체를 안정적으로 검출했다면 안전사고를 방지하기 위해 사람을 대상으로 이를 카운팅해야 한다. 실시간으로 정확하게 객체를 추적하고 인식하기 위해서 YOLO v4 모델을 사용하였다. 피플 카운팅(People Counting) 분야는 아파트, 공항, 백화점 등 사람들이 자유롭게 공간을 지날 수 있는 다양한 분야에서 사용 가능하다. 제시한 모델은 사람들의 군집 영상을 분석하여 그에 해당하는 객체 수를 추출하게 된다.

IV. 실험 및 결과

1. 학습 환경

본 논문에서 제안하는 객체 검출 및 인식을 통한 군중 밀집도에 따른 경고음 출력 시스템에서 기존 이미지 분석을 영상으로 분석하고자 한다. 사진 테스트를 진행했던 것처럼 영상을 업로드하고 코드를 재작성하면 된다. 다음 모델을 구현하기 위해 다음과 같이 학습 환경을 구성한다.

본 논문은 윈도우 11 운영체제를 탑재한 PC를 이용하였다. 개발환경 구축을 위한 SW 버전들은 서로 호환성 있는 버전들로 구성해야 한다. Python 3.7, Tensorflow 2.3+, OpenCV, numpy 등을 설치하였고, 그 외 라이브러리 및 기존 코드의 경우 YOLO v4 논문과 AlexyAB 이 공식 Github에서 복제된 저장소를 참고하였다. 파일을 다운받아 실행하기 위해 반드시 모델을 설치해야 하는데 다운받은 make 파일의 옵션을 다크넷 하위폴더로 변경해 주었다.

2. 학습 데이터 및 모델

YOLO v4 모델은 80개의 객체를 탐지할 수 있는 COCO Dataset을 가지고 있다. 객체 인식 모델 학습을 위해 Darknet framework 환경을 구성하는 과정이 필요하다. pre-trained 클래스에서 YOLO v4를 실행시키고 객체 탐지를 하기 위해 미리 학습된 YOLO v4 weights 파일을 다운로드해서 이용한다면 추가 학습 없이 수행이 가능하다.

3. 모델 실험 및 결과

본 논문에서는 객체 탐지를 위한 테스트를 위해 차량이나 보행자, 7곳의 상권 등으로부터 CCTV 영상을 수집한 AI-Hub 사이트의 dataset을 활용하였다.

다음 그림 4, 5에서는 실제 CCTV 화각에서 촬영한 영상을 예시로 테스트한 것이다.



그림 4. YOLO v4를 이용한 테스트
Fig. 4. Testing with YOLO v4

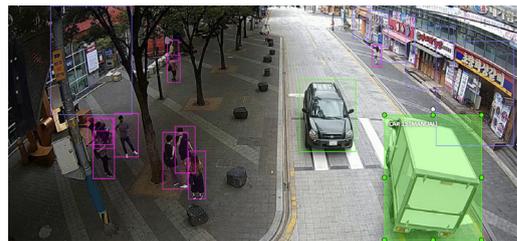


그림 5. 객체로 인식된 바운딩 박스
Fig. 5. Bounding box recognized as object

본 논문에서 제안하는 인구밀집도 위험 수준 경고 알고리즘에서 실시간으로 객체 인식을 하기 위하여 YOLO v4를 사용하여 임계값이 0.6 이상인 bounding box만 영상에 출력된 것을 알 수 있다. 인식하는 객체 Class는 그림 1에서 보듯 사람, 차, 건물(상가) 등의 사전 훈련된 가중치를 읽어들이는.

피플 카운팅을 위해 객체로 인식된 bounding box 해당 개수를 계산한다. 다음 그림 6과 그림 7은 피플 카운팅 이해를 돕기 위한 그림과 YOLO Object Counting System을 적용한 테스트 영상이다.

피플 카운트를 통해 일정 수준의 제한된 인원수를 넘겼을 경우 비프음이 출력되도록 하였다. 제한 인원수를 초과하여 넘기는 시간동안 계속해서 일정한 비프음을 울리도록 작동시키고, 제한 인원수보다 적은 경우 다시 경고음이 울리지 않도록 작동시켰다.

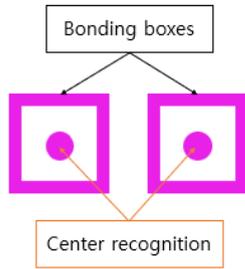


그림 6. 피플 카운팅을 위한 객체 인식 원리
Fig. 6. Object recognition principles for people counting

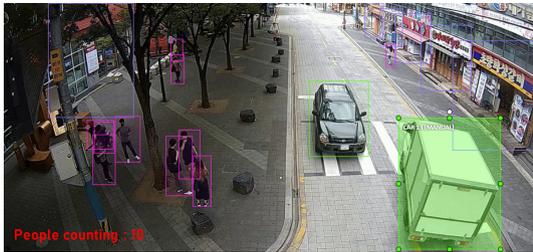


그림 7. 객체로 인식된 바운딩 박스들의 개수 카운팅
Fig. 7. Counting the number of bounding boxes recognized as objects

V. 결 론

본 논문에서는 COCO Dataset을 학습한 YOLO v4 모델을 AI-Hub 유동인구 분석을 위한 CCTV 영상 데이터로 테스트하여 객체를 검출하고 이를 기반으로 경고음이 울리도록 사고 방지 방안을 제안하였다. 딥러닝 기반의 Object Detection 기술인 YOLO v4을 이용하여 사람 및 물체 등을 탐지하고 일정 수치 이상의 객체 수가 초과할 경우 경고음을 울린다. 진행한 연구에 대하여 기존 데이터를 학습시키고 새로운 CCTV 데이터를 테스트 해 본 결과 물체가 크고 선명하면 정확히 판단하여 Accuracy가 상대적으로 80-90%에 달하고 이미지가 작고 해상도가 낮아 영상으로 잘 표현이 되지 않을 때는 20-30%에 달하는 것을 확인할 수 있었다.

본 논문의 연구자는 앞으로 사고 발생 시 무고한 희생자들이 발생하지 않기 위하여 앞으로 철저한 분석과 개선책이 필요하다고 생각하였다. 그 결과 서울 인근에 설치된 CCTV 수만 대략 8만 개이고, 이태원 참사 당시 상황을 알려 위기상황임을 인지할 수 있었던 방법 또한, CCTV 혹은 그 당시 시민들이 직접 찍었던 영상 제보였다. 이를 기반으로 본 논문을 작성하게 되었다. 향후 연구에서도 재난 및 여러 안전사고에 대하여 더 이상의 무

고한 피해자들이 속출되지 않도록 미리 탐지하고 예방하여 성능 향상에 도모할 계획이다. 앞으로 안전에 대한 인식과 그만큼의 투자를 통해 이러한 피해가 또다시 반복되지 않기를 바란다. 그러기 위해 향후 AI CCTV, CCTV 화각에서 보이는 면적에 따른 인구 밀집 분석 등 예방 체계와 안전망을 구축할 수 있도록 더 세부적인 방향을 잡고 연구하고자 한다.

References

- [1] Bochkovskiy, Alexey, Chien-Yao Wang, and Hong-Yuan Mark Liao. "Yolov4: Optimal speed and accuracy of object detection." arXiv preprint arXiv:2004.10934, 2020.
DOI: <https://doi.org/10.48550/arXiv.2004.10934>
- [2] Lin, Tsung-Yi, et al. "Microsoft coco: Common objects in context." Computer Vision-ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13. Springer International Publishing, 2014.
DOI: <https://doi.org/10.48550/arXiv.1405.0312>
- [3] D.I. Solution Co., Ltd., "CCTV image data for moving population analysis," AI-Hub, Dec 2022.
DOI: <https://aihub.or.kr/aihubdata/data/view.do?currMenu=115&topMenu=100&aihubDataSe=realm&dataSetSn=489>
- [4] Seung-Cheol Lim, Jae-Seung Go. "A Study on Design and Implementation of Driver's Blind Spot Assist System Using CNN Technique," The Journal of The Institute of Internet, Broadcasting and Communication (IIBC), Vol. 20, No. 2, pp. 149-155, 2020.
DOI: <https://doi.org/10.7236/IIBC.2020.20.2.149>.
- [5] Park, Y.-S., Lee, S.-D. and Choi, J.-H. (2021) "A Study on Establishment Method of Smart Factory Dataset for Artificial Intelligence," The Journal of The Institute of Internet, Broadcasting and Communication. Korea Internet Broadcasting and Communication Society, Vol. 21, No. 5, pp. 203-208.
DOI: <https://doi.org/10.7236/IIBC.2021.21.5.203>
- [6] Kim, June Seok, Hong, Il Young, "Analysis of Building Object Detection Based on the YOLO Neural Network Using UAV Images," Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography, Vol. 39, No. 6, pp. 381-392, 2021.
DOI: <https://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE10958031>
- [7] Taejun Lee, Minwoo Cho, Jiho Song, Chulhyun Hwang, Heokyoung Jung, "Construction of Training Data and Model Training for YOLOv4-based Factory Operation Safety Management," Journal of the Korea Information and Communication Association Comprehensive Conference, Vol. 25, No. 1, pp. 252-254, 2021.

DOI: <https://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE10565966>

- [8] Kumari, N., Ruf, V., Mukhametov, S., Schmidt, A., Kuhn, J., & Küchemann, S. "Mobile Eye-Tracking Data Analysis Using Object Detection via YOLO v4," Sensors (Basel, Switzerland), 21, 2021.
DOI: <https://doi.org/10.3390/s21227668>
- [9] Liu, Y., Gao, M., Zong, H., Wang, X., & Li, J. "Real-Time Object Detection for the Running Train Based on the Improved YOLO V4 Neural Network," Journal of Advanced Transportation, 2022.
DOI: <https://doi.org/10.1155/2022/4377953>
- [10] Yu, Jimin and Wei Zhang, "Face Mask Wearing Detection Algorithm Based on Improved YOLO-v4," Sensors (Basel, Switzerland) 21 (2021): n. pag.
DOI: <https://doi.org/10.3390/s21093263>

저 자 소 개

황 승 연(준회원)



- Seung-Yeon Hwang received his BS in Department of Computer Science at Korea Polytechnic University in 2019. He is currently studying MS in Department of Computer Science at Anyang University. His research interests include Database System, Big Data, Data Analysis, Machine Learning, etc.

김 정 준(정회원)



- Jeong Joon Kim received his BS and MS in Computer Science at Konkuk University in 2003 and 2005, respectively. In 2010, he received his PhD in at Konkuk University. He is currently a professor at the department of Computer Science at Anyang University. His research interests include Database Systems, Big Data, Semantic Web, Geographic Information Systems (GIS) and Ubiquitous Sensor Network (USN), etc.