

A Study on Traffic Vulnerable Detection Using Object Detection-Based Ensemble and YOLOv5

Hyun-Do Lee*, Sun-Gu Kim*, Seung-Chae Na*, Ji-Yul Ham*, Chanhee Kwak**

*Student, Dept. of Industrial Data Science, Kangnam University, Yongin, Korea

*Student, Dept. of Industrial Data Science, Kangnam University, Yongin, Korea

*Student, Dept. of Industrial Data Science, Kangnam University, Yongin, Korea

*Student, Dept. of Industrial Data Science, Kangnam University, Yongin, Korea

**Assistant Professor, Dept. of AI Convergence, Kangnam University, Yongin, Korea

[Abstract]

Despite the continuous efforts to mitigate pedestrian accidents at crosswalks, the problem persists. Vulnerable groups, including the elderly and disabled individuals are at a risk of being involved in traffic incidents. This paper proposes the implementation of object detection algorithm using the YOLO v5 model specifically for pedestrians using assistive devices like wheelchairs and crutches. For this research, data was collected and utilized through image crawling, Roboflow, and Mobility Aids datasets, which comprise of wheelchair users, crutch users, and pedestrians. Data augmentation techniques were applied to improve the model's generalization performance. Additionally, ensemble techniques were utilized to mitigate type 2 errors, resulting in 96% recall rate. This demonstrates that employing ensemble methods with a single YOLO model to target transportation-disadvantaged individuals can yield accurate detection performance without overlooking crucial objects.

▶ **Key words:** Object Detection, YOLO, Ensemble, Traffic Vulnerable Detection, NMS, WBF

[요 약]

횡단보도에서 발생하는 교통사고를 해결하기 위한 시도가 계속되고 있지만, 교통사고는 끊임 없이 일어나는 상황이다. 특히 노인, 장애인 등의 교통약자들은 교통사고에 노출될 위험이 더 크다. 이에 대한 문제점을 주의 깊게 볼 필요가 있다. 본 논문은 교통 약자 중 휠체어, 목발과 같은 보조 기구를 이용하는 보행자를 위해 YOLO v5 모델을 활용한 객체 탐지 기술을 제안한다. 휠체어, 목발 사용자 그리고 보행자의 이미지 크롤링, Roboflow와 Mobility Aids 데이터를 수집하였다. 일반화 성능을 높이기 위해 데이터 증강 기법을 활용하였다. 더하여 Type 2 error를 줄이기 위해 앙상블 기법을 이용하여 Recall이 96%인 높은 성능 수치를 얻었다. 이를 통해 교통약자를 목표로 YOLO 내 단일 모델을 앙상블 할 시, 객체를 놓치지 않고 정확한 탐지 성능을 보여준다는 것을 입증하였다.

▶ **주제어:** 객체 탐지, YOLO, 앙상블, 교통약자 탐지, NMS, WBF

- First Author: Hyun-Do Lee, Corresponding Author: Chanhee Kwak
- *Hyun-Do Lee (hanse19911@naver.com), Dept. of Industrial Data Science, Kangnam University
- *Sun-Gu Kim (2ksun99@gmail.com), Dept. of Industrial Data Science, Kangnam University
- *Seung-Chae Na (nsc0073@gmail.com), Dept. of Industrial Data Science, Kangnam University
- *Ji-Yul Ham (hramsm123@gmail.com), Dept. of Industrial Data Science, Kangnam University
- **Chanhee Kwak (chk@kangnam.ac.kr), Dept. of AI Convergence, Kangnam University
- Received: 2023. 10. 23, Revised: 2023. 12. 21, Accepted: 2023. 12. 22.

I. Introduction

장애인, 고령자, 임산부, 어린이 등 거동에 불편을 느끼는 교통약자들은 보행시 여러 어려움을 겪는다. 특히 횡단 보도에서 도로를 건널 때, 일반인 기준으로 설정된 시간 안에 횡단을 완료하기는 매우 어려운 일이다. 만약 횡단보도에 대기 중이거나 횡단 중인 교통약자들을 발견할 수 있다면, 이들이 안전하게 길을 건널 수 있도록 시스템적인 도움을 제공할 수 있을 것이다. 또한 교통약자에 대한 데이터를 모아 교통약자 보호구역 등을 지정하는 것도 가능할 것이다. 이를 위해서는 교통약자를 정확하게 분류하는 것이 필수적이다. 이에 본 연구는 교통약자를 정확하게 탐지할 수 있는 모델을 제안한다.

본 연구가 제안하는 모델의 탐지 대상은 휠체어, 목발 사용자인 교통약자와 일반 보행자로 총 3종이다. 탐지 대상을 정확하게 분류하기 위해 다양한 각도의 CCTV 데이터를 학습에 활용하였으며, 모델 과적합을 방지하기 위해 데이터를 증강하여 일반화 성능이 강화된 모델을 구현하였다. 낮은 탐지율의 모델을 구현하기 위해 앙상블 기법을 활용하여 교통약자 탐지 시의 오류 최소화를 목표로 한다.

본 연구는 다음과 같이 구성되었다. 먼저 2장에서는 객체 탐지, 앙상블 기법 그리고 교통약자를 위한 횡단보도 복지에 대해 설명한다. 그 후, 3장에서는 본 연구가 제안하는 교통약자 탐지 모델을 기술한다. 4장에서는 탐지에 대한 실험 결과를 기술한다.

II. Preliminaries

1. You Only Look Once(YOLO)

객체 탐지는 컴퓨터 비전의 핵심 기술로, 이미지나 비디오에서의 객체의 위치와 종류를 탐지하고 식별하는 기술이다. 객체 탐지 기법은 다양한 산업 분야에서 활용되고 있으며, 최근 코로나 팬데믹으로 인해 마스크 착용 여부 식별에 활용되는 등 그 적용 범위를 넓혀가고 있다. 객체 탐지 모델의 구조는 다양하지만, 일반적으로 입력 데이터를 작은 단위로 분할하고, 합성곱 신경망(Convolutional Neural Network, CNN)을 거쳐 분할된 데이터로부터 특징을 추출한다. 이후, 객체 탐지 알고리즘을 통해 객체가 존재할 수 있는 영역에 bounding box(bbox)를 생성한다. 이 과정에는 객체의 위치를 식별하는 Localization 문제와 객체를 각 클래스에 맞게 분류하는 Classification 문제로 나뉘어 객체를 탐지한다.

객체 탐지 모델의 구조는 크게 1-Stage Detector와 2-Stage Detector로 구분된다. 1-Stage Detector는 특성 추출 후 Classification과 Localization 과정을 한 번에 진행하는 Detector다. 대표적인 모델로는 YOLO와 SSD, Retina Net 등이 존재한다. 2-Stage Detector는 Region Proposal을 우선 선별한 상태에서 특성을 추출하여 Classification과 Localization 과정을 구분해서 차례대로 진행하는 Detector다. 이에 해당하는 모델로는 R-CNN, Fast R-CNN, Faster R-CNN이 있다. 1-Stage 방식은 탐지 속도가 빠르지만, 상대적으로 낮은 정확도를 보여주고, 2-Stage 방식은 구조상 탐지 속도가 느리지만 비교적 정확한 성능을 보여준다.

이 중, YOLO[1]는 2016년에 등장한 모델로 기존 객체 탐지의 문제인 느린 속도를 개선하고자 제안된 객체 탐지 방법론이다. YOLO는 전체 이미지에서 얻은 feature map을 활용해서 bbox를 regression 개념으로 예측하고 모든 클래스에 대한 확률을 계산하면서 학습하는 모델이다. Input image data를 grid로 나눈 뒤 bbox regression stage와 class probability map을 계산하는 단계가 병렬적으로 진행되면서 최종 detection을 수행한다. 이 과정에서 신뢰도 측정을 위해 confidence score를 사용한다. 아래 식 (1)은 classification score와 Localization score를 모두 반영할 수 있는 YOLO 모델의 confidence score다. 이 score는 추후에 최종 bbox를 채택하는 NMS(Non-Maximum Suppression) 과정과 앙상블 과정에서 가장 높은 값의 box를 선정할 때 기준 점수로 사용된다. 여기서 IOU 값은 실제 bbox와 예측한 bbox의 합집합에서 교집합을 나눈 값을 나타내는 평가지표이다.

$$(1) \Pr(\text{Class} | \text{Object}) * \Pr(\text{Object}) * IOU_{pred}^{truth} = \Pr(\text{Class}) * IOU_{pred}^{truth}$$

객체 탐지 기술을 응용하여 교통약자를 검출하는 연구는 다양하게 존재하는데, Vasquez[2]는 Fast R-CNN 모델을 이용하여 보행자, 휠체어 사용자, 휠체어를 미는 사람, 목발 사용자 그리고 보행 보조 장비 사용자 탐지 연구를 수행하였다. 해당 연구에서는 깊이 데이터를 활용하여 각 사람의 위치와 속도를 추적하고, 히든 마르코프 모델(HMM)을 사용하여 각 클래스를 탐지하였다. 다만, 특수한 환경의 데이터 구축과 3D 개념의 detection이 요구되기 때문에 경량화된 모델 구현을 하기에는 어려움이 존재한다. 경량화되지 않은 객체 탐지 모델은 신호등 시스템에 탑재하기에 번거롭고 돌발 상황이 빈번하게 발생하는 보행 환경 속에서 빠르게 교통약자를 찾아내야 하는 목표를 달성하는 데에 제약이 있다.

따라서 본 논문에서는 YOLO 모델을 활용하고, [2] 에서 활용한 데이터 중 깊이 데이터를 제외한 일부의 데이터를 활용함으로써 경량화된 교통약자 탐지 서비스를 제안한다. 나아가 단일 YOLO 모델 간 Ensemble을 시도함으로써 다양한 환경에서도 객체 탐지가 가능한 일반화된 시스템을 제안한다.

2. Ensemble Model

머신러닝에서 Ensemble이란 하나의 문제를 다양한 시각에서 볼 수 있는 기법이다. 다시 말해, 여러 개의 모델을 생성하고 그 예측을 결합하여 보다 정확한 예측을 얻는 것을 말한다. Ensemble은 연산이 많은 하나의 모델을 사용하지 않고 적은 연산의 모델 여러 개를 조합하여 더 정확한 예측을 하는 데 도움을 준다. 객체 탐지에 이를 접목한다면, 연산량이 적은 다수의 모델로 학습을 한 뒤, 마지막에 각 모델이 검출한 결과를 모두 결합하여 최종 bbox를 탐지할 수 있다. 이를 통해 일반화된 모델을 구축하고 다양한 환경에서 성능을 보장할 수 있다. Lanfa[8]은 YOLO v5 모델 중 YOLO v5 n, YOLO v5 s, YOLO v5 m, YOLO v5 l을 앙상블하여 기존 단일 모델보다 철도 환경에서 플라스틱 폐기물을 탐지하는데 높은 성능을 보여주었다. Lanfa[8]에서 크기가 큰 모델보다 앙상블한 모델이 더 높은 정확도를 보여주었다.

빠르면서 정확한 탐지를 해야 하는 본 연구에서는 Lanfa[8]에서 아이디어를 얻어 YOLO v5 앙상블을 수행하였다. 단일 모델로 야기되는 과적합을 최소화하고 오탐지율의 최소화를 목적으로 앙상블 수행을 채택하였다. 앙상블 기법은 아래 두 방식을 사용한다.

2.1. Non-Maximum Suppression(NMS)

NMS는 YOLO에서 최종 bbox를 고를 때 사용하는 기법으로, 여러 bbox 중 하나만 남기고 나머지를 제거하는 방법을 말한다. 이는 다양한 객체 탐지 모델에서 bbox를 채택할 때 사용된다.

NMS 과정은 모든 bbox 중 지정한 threshold보다 작은 confidence score인 bbox를 제거한다[3]. 이후, 남은 bbox들을 confidence score 기준으로 내림차순하고, 가장 높은 score의 bbox 하나를 선정하여 다른 bbox들과의 IOU 값을 구한다. 해당 단계는 bbox들 간의 IOU 값을 계산하는 것이므로 기준 bbox와 가장 겹치는 bbox를 제거한다. 즉, 기준 bbox와 IOU가 threshold 이상인 bbox들을 제거할 수 있게 된다. 위 과정을 순차적으로 시행하여 모든 bbox를 비교하면서 제거한다.

2.2. Weight Boxes Fusion(WBF)

WBF[4]는 예측된 bbox들을 융합하여 새로운 bbox를 만들어 내는 기법을 말한다. WBF의 세부 절차는 다음과 같다. 우선 각 모델의 예측된 결과들을 confidence score 기준으로 내림차순 정렬한다. 그 후, 정렬된 순으로 bbox들 간의 IOU를 계산하여 동일한 객체 클러스터에 들어갈 수 있는지 판단한다. 즉, bbox의 IOU가 threshold 값 이상일 경우 해당 객체 클러스터에 추가한다. 이렇게 찾은 객체 클러스터에 하나의 bbox가 새로 추가될 때마다, 평균 bbox 좌표와 confidence score를 연산한다. 이때 각 bbox의 confidence score를 가중치로 하여 높은 score를 가진 bbox에 큰 비중을 둔다. 이 과정을 통해, 융합된 bbox 좌표를 얻는다. 모든 객체에 해당하는 연산이 끝나면, 융합된 bbox의 confidence score에 객체 클러스터의 수를 곱하고, 모델 개수로 나누어 가중치를 부여한다. 결과적으로 모델들이 예상한 객체 위치 중 예측된 빈도가 높은 부분에 가중치를 부여하며 앙상블을 마친다.

3. Research on Transportation Disadvantaged

보행 중 교통사고 예방을 위해 많은 기술이 개발 및 상용화되고 있다. 교통약자를 위한 시설물인 음성 안내 버튼과 신호등 보행 시간 안내 서비스가 대표적이다. 음성 안내 버튼은 보행신호와 시간을 음성으로 알려줘 도움이 필요한 보행자에게 신호를 전달하기 위한 목적으로 설계되었다. 하지만 휠체어와 같은 보행 보조기기를 사용하거나 신체활동에 제한이 있는 교통약자들의 경우, 이용이 어렵다. 신호등 보행 시간 안내 서비스 역시 거동이 불편한 사람들에게 적절한 시간을 제공하지 못하는 단점이 존재한다. 이처럼 기존의 기술과 시설물들은 능동적인 형태의 복지 서비스로 설계되어 사용자의 물리적인 접촉을 필요로 한다. 현재 기술과 시설물들은 교통약자의 안전과 편의를 제공하는 환경을 조성하는데 부족함이 있다.

관련 연구로 [5]는 교통약자의 안전을 고려한 보행신호 시간 산정 모형과 민감도 분석을 통해 적합한 시간을 얻는 연구를 하였다. [6]은 차량 지체를 최소화하는 동시에 교통약자의 안전을 확보하는 방법을 제시하였다. 하지만, 교통약자 보행에 적합한 시간만을 연구하였으며 수행 방식으로는 능동적인 형태인 보행자 감응 제어만을 활용하였다. 보행자 감응 제어 방식은 보행자의 요청 이후 돌아오는 보행자 서비스 시간대에 선택적으로 작동되는 방식이다. 이 방식으로는 실시간 제어가 힘들고 교통약자는 직접 요청해야 하는 과정을 거쳐야 한다. [7]에서 취약한 도로 사용자를 감지하는 것은 사고를 방지하는 데 큰 관련이 있다고

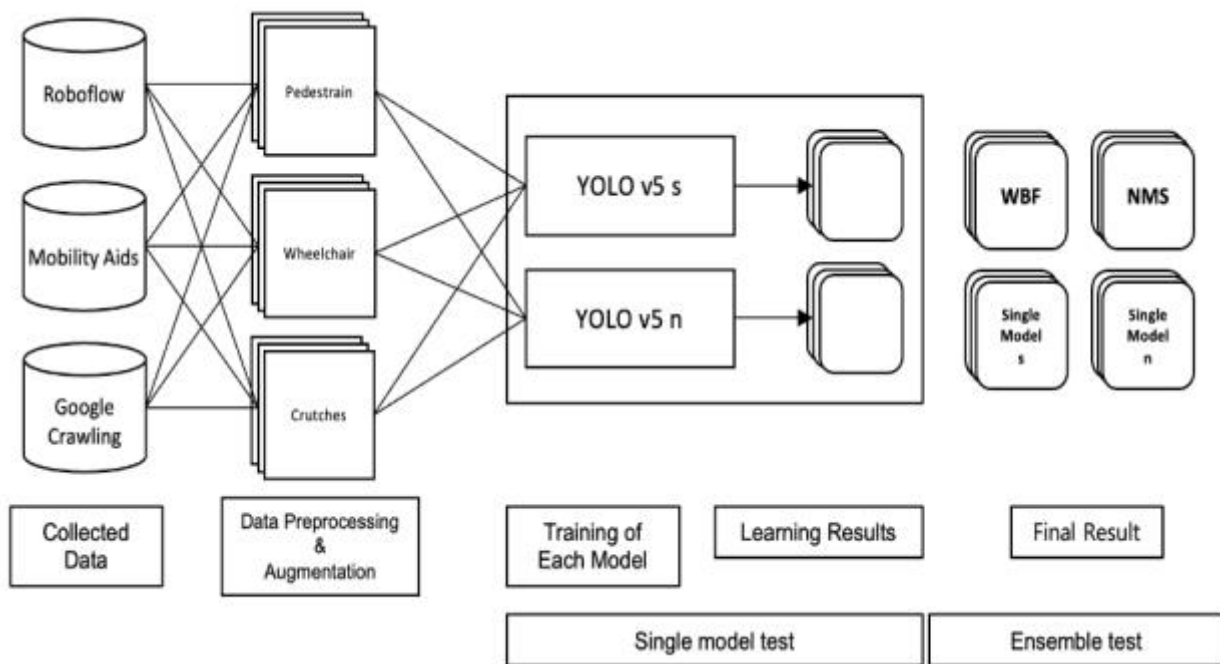


Fig. 1. Schematic Diagram of the Detection System

말한다. 해당 논문에서는 취약한 도로 사용자를 자전거 운전자로 정의하여 탐지한다. 본 연구는 [7] 연구에서 긍정적인 연구 결과를 토대로 취약 도로 이용자를 교통약자인 휠체어 사용자, 목발 사용자로 정의한다.

본 연구에서는 교통약자를 탐지하여 실시간 지능형 보행 신호 시스템에 기여할 수 있는 방법을 설계하였다. 해당 연구는 적은 연산으로 교통약자를 보다 확실하게 탐지할 수 있는 모델을 제안한다.

III. The Proposed Scheme

3.1. System Overview

[Fig 1]은 본 연구에서 제안하고자 하는 교통약자 탐지 시스템의 구조다. 관련 이미지 데이터를 수집하여 레이블링과 정제를 거치고, 단일 YOLO v5를 학습하였다. 나아가 일반화를 위해 다양한 조합의 모델을 앙상블 기법과 증강 기법을 활용하여 성능을 향상시켰다.

3.2. Training Dataset

본 연구에서는 도움이 필요한 교통약자를 휠체어 사용자(Wheelchair)와 목발 사용자(Crutches)로 정의하였으며 일반 보행자(Pedestrian) 객체를 추가시켜 총 3개의 class를 학습 데이터로 구성하였다.

본 데이터는 구글 이미지 카테고리에서 크롤링하여

YOLO 좌표에 맞게 직접 680개의 데이터를 레이블링했다. 레이블링 도구로는 labelImg 툴을 활용하였다. 이에 더해 Roboflow와 Mobility Aids 데이터[2]를 추가로 수집해 학습용 데이터를 구축하였다. Roboflow는 컴퓨터 비전에 필요한 데이터셋을 제공하는 사이트로, 사용자가 데이터를 업로드하거나 타인이 업로드한 데이터를 가지고 쓸 수 있는 기능을 제공한다. Mobility Aids 데이터셋은 프랑크푸르트 대학교 공과대학 시설과 프랑크푸르트의 한 병원에서 수집된 데이터로, 사람과 이동 휠체어와 목발 탐지를 위한 데이터 셋이다. 해당 데이터 셋은 VOC 좌표로 구성되어 있어 YOLO 좌표로 바꿔주는 과정을 거쳤다. 모든 데이터는 휠체어, 목발, 보조 기구에만 레이블링된 것이 아닌 해당 기구를 사용하는 사람과 함께 bounding box로 레이블링되어 있다.



(a) pedestrian (b) crutches (c) wheelchair

Fig. 2. Labeled images of people with disabilities and pedestrians

수집한 데이터를 총 3,543장의 이미지 데이터셋으로 구축하였다. 객체별 비율은 목발 사용자 33.3%, 휠체어 사용자 33.2% 그리고 보행자 33.5%로 구성하였다.

추가로 안정적인 일반화 성능을 위해 증강 기법(Augmentation)을 활용했다. 기존 수집된 3,543장의 이미지 데이터 셋에서 Albumentations 패키지를 활용하여 최종 6,017장의 이미지 데이터를 확보했다. 증강으로는 밝기, 좌우 변환, 각도 30도 이내의 회전을 진행했다. 데이터 누수를 피하고자 기존 3,543장 이미지 데이터에서 학습 데이터와 테스트 데이터를 7:3의 비율로 나누고 학습 데이터에만 증강을 진행했다.

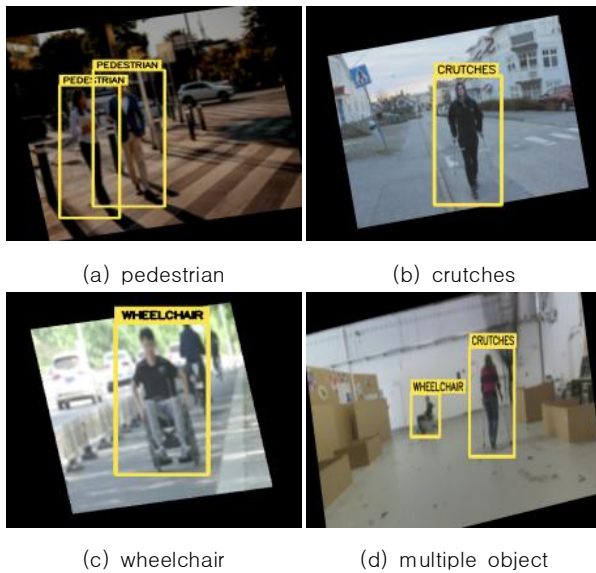


Fig. 3. Augmented data (a) pedestrian, (b) crutches, (c) wheelchair, (d) multiple objects

3.3. The Proposed Model

증강까지 진행된 학습 데이터를 YOLO 모델을 통해 학습했다. 학습된 모델은 휠체어 사용자, 목발 사용자 그리고 보행자의 특징을 탐색하여 해당 객체를 검출한다. YOLO 모델 중에서는 YOLO v5 n과 YOLO v5 s 모델을 선택하여 비교했다. 단일 모델을 통해 결과를 도출하였고, 앙상블 기법과 단일 모델의 성능을 비교했다.

YOLO v5는 YOLO v4 모델이 발표된 지 2개월 만에 공개된 모델로 CSPNet 기반의 Backbone으로 연결되어 있다. CSPNet[9]는 컴퓨팅 자원이 부족한 환경에서 감당하기 어려운 연산을 완화할 수 있는 모델이다. 이 모델의 목적은 Gradient Combination이 많이 생성되는 동안 연산량을 줄이는 것으로 초기 계층에서 두 개의 파트로 나눈 뒤, 한 파트는 DenseNet 연산이 진행되고 남은 파트는 마지막 계층에서 합침으로써 낮은 연산량을 달성했다. YOLO v5는

이전 버전과 달리 Darknet 대신 PyTorch를 통해 설계됐다. Bottleneck CSP를 이용하여 각 계층의 연산을 동일하게 분배해서 연산 병목현상을 없앴다. 다른 YOLO 모델과 다르게 모델의 크기별로 x(xlarge), l(large), m(medium), s(small), n(nano)로 나누어 구성했다.

현재 수집된 데이터를 파라미터 양이 많은 모델로 학습하게 되면 과적합 현상이 일어날 우려가 있다. 이를 대비하여, YOLO v5 중 적은 연산을 가진 s와 n을 활용하여 학습을 진행했다. 테스트 환경으로는 Google Colab Pro 내에 있는 A100 GPU를 사용했다. 각 모델을 평가하기 위한 지표로는 mAP(Mean Average Precision), 재현율(Recall)과 정밀도(Precision) 등이 있다. mAP는 재현율 값이 변화됨에 따라 바뀌는 정밀도 값들의 평균을 말하는 AP를 객체별로 구하고, 모든 객체의 AP에 대한 평균을 구한 것을 뜻한다. 여기서 재현율(Recall)이란 실제 양성 집단 중 양성으로 예측된 집단의 비율이다. 보통 자율주행처럼 즉각적인 판단과 조치를 요구하는 서비스에서는 재현율 값에 더 민감하게 결과를 확인하고 이를 높여야 한다. 정밀도(Precision)란, 양성으로 예측된 집단 중 실제 양성 집단의 비율을 뜻한다. AP는 Average Precision의 약자로, 재현율을 0부터 1까지 0.1 단위로 증가시킬 때 Precision이 감소하게 되는데 각 재현율 값마다 Precision 값을 계산하여 평균을 낸 값이다. mAP 0.5란 IOU 값의 임계값이 0.5 이상일 때의 결괏값들을 활용하여 mAP를 구한 것이다.

본 연구에서는 NMS 기법을 YOLO 모델 내부에 있는 기능으로 구현하였으며, WBF 기법은 직접 구현을 통해 실험을 진행한 후 평가지표를 계산했다. 평가지표로는 mAP 0.5와 Recall 값을 정의하였다. 교통약자를 올바르게 판단하지 못하는 상황인 Type 2 error를 피하기 위해, 실제 양성 집단 중 예측한 양성 집단의 비율을 계산하는 Recall 값을 중점적으로 확인했다.

3.4. Experiment Results

증강된 데이터 6,017장을 사용하였고 이미지 크기는 448×448 로 하이퍼 파라미터를 지정하였다. Epoch의 수는 50으로 두고 실험을 진행했다. WBF에서 IOU의 임계값을 0.5로 두었다. Benchmark 모델로는 RetinaNet 모델을 활용하여 성능 비교를 진행했다. 총 5가지의 실험 평가 결과[10]는 [Table 1]과 같다.

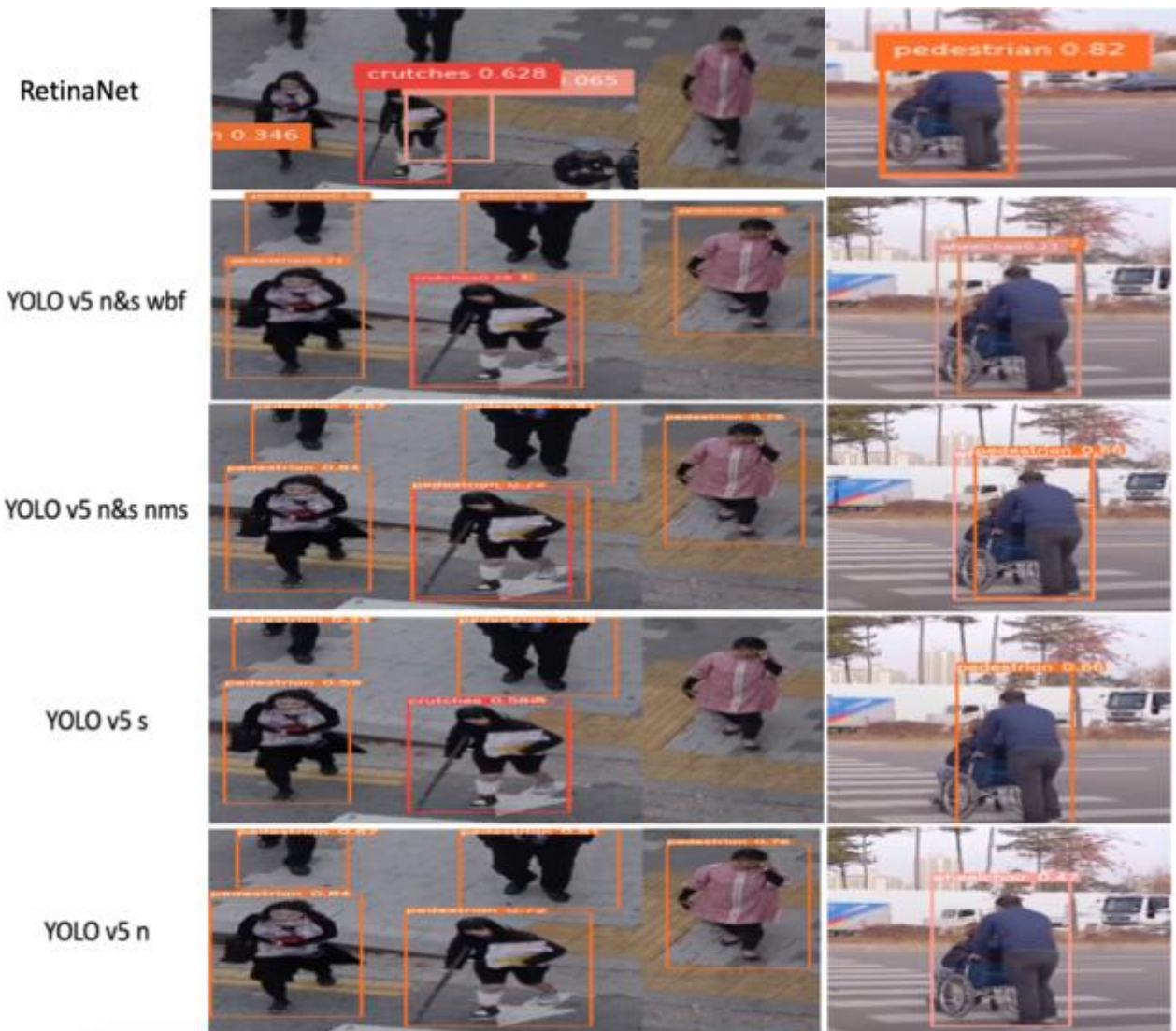


Fig. 4. Compare the results of 4 YOLO model inference

Table 1. Compare the results of 4 YOLO models and RetinaNet

| Model | mAP @0.5 | Recall |
|-----------------|----------|--------|
| RetinaNet | 0.90 | 0.88 |
| YOLO v5 s | 0.94 | 0.91 |
| YOLO v5 n | 0.91 | 0.87 |
| YOLO v5 n&s WBF | 0.93 | 0.96 |
| YOLO v5 n&s NMS | 0.95 | 0.92 |

평가지표의 결과[Table 1]를 보면, mAP 0.50 기준으로 n&s NMS, s, n&s WBF, n, RetinaNet 순으로 높다. Recall 기준으로는 n&s WBF, n&s NMS, s, RetinaNet, n순으로 높다. n&s WBF는 Recall이 우수해야 하는 본 연구 목적에 적합한 성능을 보여줬다. mAP @0.5는 단일 모델 s보다 낮으나, 교통약자를 탐지하지 못하는 경우를 최소화해야 하므로 비교적 mAP @0.5가 낮더라도 Recall이

높아서 목적에 맞는 우수한 모델로 채택하였다. n&s NMS는 Recall과 mAP @0.5에서 단일 모델보다 더 높은 지표를 보여줌으로써 단일 모델보다 양상불한 모델이 더 성능이 높다는 것을 증명하였다.

각 모델이 실제 횡단보도 환경의 영상에서는 어떻게 탐지하는지 성능을 비교했다. Youtube에서 횡단보도 내 휠체어 또는 목발을 사용하는 사람이 있는 대략 3분의 영상 5건을 수집하여 각 모델에 추론시켰다. 직접 영상 속 적합한 부분을 선정하여 비교 실험을 진행했다. [Fig 4]는 각 모델의 추론 결과를 보여준다. [Fig 4]의 단편적인 사진 두 장을 통해서 모델의 평가를 설명하기에는 부족하다. 하지만 [Table 1]과 [Fig 4]를 종합적으로 보았을 때, Type 2 error인 실제 양성을 음성으로 판정하는 것을 줄여야 하는 본 연구에서는 두 단일 모델은 부족한 면을 보여주는 것이 사실이다. WBF와 NMS의 모델은 단일 모델의 양상

블을 통해 Type 2 error가 적게 나타난다. 이후에도 다양한 횡단보도 배경의 영상에서 실험해 본 결과, 단일 모델보다 앙상블 모델이 Type 2 error가 적음을 확인할 수 있었다. 본 연구 결과를 통해 앙상블 모델이 단일 모델보다 본 연구의 목적에 맞게, 교통약자를 교통약자가 아니라고 탐지하는 경우가 적음을 입증하였다.

각 모델의 수행 시간은 i7 cpu와 RTX 3070의 환경에서 동일한 Youtube 영상으로 측정했다. WBF[4]는 해당 논문에서도 말하듯이 real time에서는 적합하지 않다. 추후에 real time이외에 해당 모델이 활용된다면, 안전한 탐지 성능을 보여줄 것이다. 본 연구에서는 real time 진행 시 NMS 앙상블 모델을 활용한다. 실행 속도로는 0.007초대가 나왔으며, 단일 모델인 n과 s 모두 0.003초대가 나왔다. 단일 모델보다는 당연히 느린 속도를 보이지만, Type 2 error의 최소화를 감안하면, 충분히 실행하기에 적합하다고 볼 수 있다.

IV. Conclusions

본 연구는 보행자, 휠체어 사용자 그리고 목발 사용자를 횡단보도 속 역동적인 환경에서 검출하고자 했다. 이 과정에서 학습 정확도 향상을 위해 이미지 증강 기법을 활용하였다. 모델 구성에 있어서는 YOLO v5 모델들을 NMS와 WBF를 활용하여 앙상블하여, 단일 모델의 과적합 문제를 해결하고 Type 2 error를 최소화할 수 있는 안정성을 갖춘 모델을 설계하였다. 실제 횡단보도 환경 속 영상에서 단일 모델과 앙상블한 모델을 비교하였을 때, 작은 모델들을 합쳐 연산을 최소화하고 본 연구 목적에 충족시키는 결과를 보였다. 연산량이 적은 모델들인 s와 n의 선택과 증강 기법을 통해 과적합을 최소화하여 하였고, 더 높은 Recall 값을 추구하기 위해 앙상블 기법을 활용하였다. 그 결과, 단일 모델에서 감지하지 못하거나 오탐지된 객체들을 보다 올바르게 탐지할 수 있었다.

본 연구의 한계점은 다음과 같다. 먼저 서비스를 제공하는 횡단보도 환경의 CCTV 각도에서 바라보는 데이터를 활용하지 못하였다. 다양한 각도의 데이터를 활용할 수 있다면 훨씬 정교한 모델 구축이 가능할 것이다. 그러나 대부분의 CCTV가 고정된 장소에서 촬영 데이터를 수집하기 때문에, 본 연구의 제안 모델은 의미 있는 결과라 할 수 있다. 아울러 향후 교통약자 검출 연구들에서는 교통 신호 체계와 연동하여 자동으로 적합한 보행 시간을 제공하는 교통 체계 개발 등이 가능할 것으로 기대한다.

REFERENCES

- [1] Redmon, Joseph, et al. "You only look once: Unified, real-time object detection." Proceedings of the IEEE conference on computer vision and pattern recognition, pp.779-788, June 2016, DOI:10.48550/arXiv.1506.02640
- [2] Vasquez, Andres, et al. "Deep detection of people and their mobility aids for a hospital robot." 2017 European Conference on Mobile Robots (ECMR). IEEE, pp.4-5, September 2017, DOI: 10.1109/ECMR.2017.8098665
- [3] 6 people including Hantam Combi, Sweat and Sweat Deep Learning Computer Vision Encyclopedia <https://wikidocs.net/142645>
- [4] Solovyev, Roman, Weimin Wang, and Tatiana Gabruseva. "Weighted boxes fusion: Ensembling boxes from different object detection models." Image and Vision Computing 107, pp. 2-3, Feb 2021, DOI:10.1016/j.imavis.2021.104117
- [5] Deoksu Hwang, et al. "Development of Pedestrian Signal Timing Models Considering the Characteristics of Weak Pedestrians." Journal of the Korean Society of Transportation, Vol. 21, No. 1, pp. 182-184, February 2008.
- [6] Junil Ryu, Wonchul Kim, and Hyungchul Kim. "A Priority Signal Control Strategy for Vulnerable Considering Traffic Flow - Focusing on crosswalks in coordinated arterial sections" Journal of the Korean ITS Society, Vol. 13, No. 4, pp. 15-16, October 2014. DOI: 10.12815/kits.2014.13.4.012.
- [7] Garcia-Venegas, Marichelo, et al. "On the safety of vulnerable road users by cyclist detection and tracking." Machine Vision and Applications 32.5 (2021): 109, August 2021.
- [8] Liu, Lanfa, et al. "Yolo-Based Multi-Model Ensemble for Plastic Waste Detection Along Railway Lines." IGARSS 2022-2022 IEEE International Geoscience and Remote Sensing Symposium. IEEE, pp. 7658-7661, Sep 2022. DOI:10.1109/IGARSS46834.2022.9883308
- [9] Wang, Chien-Yao, et al. "CSPNet: A new backbone that can enhance learning capability of CNN." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops. pp. 3-7, Nov 2019, DOI:10.48550/arXiv.1911.11929
- [10] Ye-Chan Yun, Young-Tae Kwak, "Development of A Uniform And Casual Clothing Recognition System For Patient Care In Nursing Hospitals", Journal of the Korea Society of Computer and Information ,Vol. 25, No. 12, pp. 45-53, Dec 2020 DOI:10.9708/jksoci.2020.25.12.045
- [11] Jaeyong Kang, and Jeonghwan Kwak. "Classification of brain tumors in MRI images using deep learning-based feature ensembles." Journal of the Korean Computer Information Society 26.7, Vol. 26, No. 7 pp. 37-44, March 2021, .DOI:10.3390/s21062222

Authors



Hyun-Do Lee is currently a student for the B.S. degrees in Industrial Data Science of Kangnam University, Korea. He is interested in automatic drive, computer vision, and SLAM.



Sun-Gu Kim is currently a student for the B.S. degrees in Data Science of Kangnam University, Korea. He is interested in computer vision, and multi modal data analysis.



Seung-Chae Na is currently a student for the B.S. degrees in Data Science of Kangnam University, Korea. He is interested in data-driven operations management.



Ji-Yul Ham is currently a student for the B.S. degrees in Data Science of Kangnam University, Korea. He is interested in anomaly detection, time series analysis, and image quality inspection.



Chanhee Kwak received the B.S. in computer science, M.S, and Ph.D. degrees in management engineering in KAIST, Korea. He joined the faculty of the department of AI convergence at Kangnam University,

Yong-in in 2020. He is currently an assistant professor in the department of AI convergence at Kangnam university, Yongin, Korea. His research interests include data analytic systems and machine learning techniques.