Journal of KIIT. Vol. 22, No. 12, pp. 39-46, Dec. 31, 2024. pISSN 1598-8619, eISSN 2093-7571
http://dx.doi.org/10.14801/jkiit.2024.22.12.39

시각장애인 보행 안전을 위한 Few-shot Learning과 Grounding DINO 기반 준지도학습 YOLO 프레임워크

정소미*, 이영학**[1], 정은미**[2]

Semi-Supervised Learning YOLO Framework based on Few-shot Learning and Grounding DINO for Blind Pedestrian Safety

Somi Jeong*, Yeung-Hak Lee**[1], and Eunmi Jung**[2]

본 연구는 2024년 과학기술정통신부 및 정보통신기획평가원의 SW중심대학사업의 연구결과로 수행되었음 (2019-0-01113)

요 약

시각장애인의 보행 안전을 위한 객체 탐지 시스템은 실시간 처리와 높은 정확도가 요구되나, 현재 공개된 보행자 시점의 데이터셋이 매우 제한적이라는 문제에 직면해 있다. 본 연구에서는 이러한 한계를 극복하기 위해 Few-shot Learning과 Grounding DINO를 통합한 새로운 YOLO 최적화 프레임워크를 제안한다. 제안된 프레임워크는 클래스당 레이블링된 이미지 수를 기존 3,000장에서 150장으로 95% 감소시키면서도 mAP 0.80의 성능을 유지하였다. 또한 동적 배치 정규화와 신뢰도 기반 가중치 손실 함수를 도입하여 야간 우천과 같은 열악한 환경에서도 mAP 0.78의 안정적인 성능을 보였으며, 42 FPS의 처리 속도와 3.5GB의 메모리 사용량으로 모바일 환경에서의 실시간 처리가 가능함을 입증하였다. 본 연구에서 제안한 통합적 접근은 의료 영상 분석, 산업 검사 등 레이블링된 데이터가 제한적인 다양한 분야에서의 활용 가능성을 보여준다.

Abstract

Object detection systems for safe walking of visually impaired people require real-time processing and high accuracy, but face limitations due to the scarcity of publicly available pedestrian perspective datasets. This study proposes a novel YOLO optimization framework that integrates Few-shot Learning and Grounding DINO to overcome these limitations. The proposed framework maintains an mAP of 0.80 while reducing the number of labeled images per class by 95%, from 3,000 to 150. By introducing dynamic batch normalization and confidence-based weight loss functions, the system achieves stable performance with an mAP of 0.78 even in adverse conditions such as nighttime rain. The framework demonstrates real-time processing capability in mobile environments with 42 FPS and 3.5GB memory usage. Our integrated approach shows potential applications in various fields with limited labeled data, such as medical image analysis and industrial inspection.

Keywords

computer vision, few-shot learning, YOLO, grounding DINO, object detection

* 국립안동대학교 컴퓨터공학과 학사과정
- ORCID: https://orcid.org/0009-0002-9828-690X
** 국립안동대학교 SW융합교육원 교수(**[2] 교신저자)
- ORCID[1]: https://orcid.org/0000-0002-9037-2646
- ORCID[2]: https://orcid.org/0009-0001-7233-6422

· Received: Nov. 14, 2024, Revised: Dec. 26, 2024, Accepted: Dec. 29, 2024
· Corresponding Author: Eunmi Jung
Dept. of SW Convergence Education Center, Andong National Uninversity, 1375, Gyeongdong-ro, Andong-si, Gyeongsangbuk-do, Republic of Korea
Tel.: +82-54-820-6929, Email: emjung@anu.ac.kr

Ⅰ. 서 론

시각장애인의 안전한 보행권 보장은 현대 사회의 핵심적인 도전 과제로 부각되고 있다. 한국시각장애인연합회의 2023년 조사에 따르면, 전국 시·도 및 시·군·구 행정청의 시각장애인 보행 접근성이 매우 열악한 것으로 나타났다. 특히 횡단보도 점자블록의 적정설치율이 4.0%에 불과하여, 시각장애인의 안전한 보행 환경 조성이 시급한 과제로 대두되고 있다[1]. 이러한 문제에 대한 새로운 해결 가능성을 컴퓨터 비전 기술의 비약적 발전이 제시하고 있다. 특히 YOLO(You Only Look Once) 계열의 객체 탐지 모델은 실시간 처리 능력과 높은 정확도를 바탕으로 보행 안전 시스템 구현의 기술적 토대를 마련하였다[2]. YOLO는 단일 단계 접근 방식을 사용하여 전체 이미지를 한 번에 처리함으로써 실시간 객체 탐지를 가능하게 하지만, 이러한 모델들은 대규모의 레이블링된 학습 데이터를 필요로 한다는 근본적인 제약이 존재한다[3].

시각장애인의 보행 안전을 위해서는 보행자 시점에서의 객체 탐지가 필수적이나, 현재 공개된 데이터셋의 대부분은 차량 시점에 초점이 맞추어져 있어 실제 적용에 한계가 있다[4]. 보행자 시점의 데이터셋 구축에는 다양한 환경 변수들이 고려되어야 한다. 기상 조건과 조도 변화는 물론, 시각장애인의 실제 보행 높이와 시야각에 따른 특수성이 반영되어야 하며, 이는 상당한 시간과 비용을 수반한다[5]. 이러한 맥락에서 제한된 데이터셋으로도 효과적인 학습이 가능한 Few-shot Learning과 레이블링 비용을 절감할 수 있는 Grounding DINO의 활용이 주목받고 있다[6][7]. Few-shot Learning은 메타 학습을 통해 소수의 학습 데이터만으로도 새로운 객체나 환경에 대한 적응이 가능하며, Grounding DINO는 Vision-Language 모델의 특성을 활용하여 자연어 기반의 효율적인 객체 탐지를 가능하게 한다. 본 연구에서는 이 두 기술의 상호보완적 특성을 활용한 새로운 준지도학습 프레임워크를 제안하고, 이를 YOLO 모델과 통합하고자 한다. 이를 통해 데이터 수집의 한계를 극복하고 시각장애인의 보행 안전을 위한 실용적이고 고성능의 객체 탐지 시스템을 개발하는 것이 목표이다.

제안된 방법론은 소규모 데이터셋 환경에서도 효과적인 객체 탐지가 가능하며, 실시간 처리가 가능한 경량화된 모델 구조를 통해 실제 보행 환경에서의 적용 가능성을 최적화하였다. 이를 통해 시각장애인들의 안전한 보행을 지원하고, 궁극적으로 그들의 삶의 질 향상에 기여할 수 있을 것으로 기대된다.

Ⅱ. 관련 연구

2.1 YOLO와 객체 탐지 모델

YOLO는 단일 신경망을 통한 실시간 객체 탐지 알고리즘으로, 컴퓨터 비전 분야에서 주목받고 있다. 최신 버전인 YOLOv8은 anchor-free 방식을 도입하여 객체 탐지의 효율성과 정확도를 크게 향상시켰다[8]. Anchor-free 접근법은 사전 정의된 anchor box에 의존하지 않고 직접적으로 객체의 중심점과 크기를 예측함으로써, 다양한 크기와 형태의 객체를 더 유연하게 탐지할 수 있다[9]. 이러한 특성은 시각장애인 보행 안전 시스템에서 다양한 장애물과 위험 요소를 신속하고 정확하게 식별하는 데 유리하다. 그러나 YOLOv8의 최적 성능 발휘를 위해서는 여전히 클래스당 상당한 수의 레이블링된 이미지가 요구되며, 이는 특수 목적 응용 분야에서 중요한 제약 요인으로 작용한다[10].

2.2 Few-shot Learning

Few-shot Learning은 제한된 학습 데이터 환경에서도 효과적인 모델 학습을 가능하게 하는 접근 방식이다. 이 방법은 메타 학습(Meta-learning) 원리를 기반으로 하여, 새로운 클래스나 객체에 대해 소수의 예제만으로도 빠르게 적응할 수 있는 능력을 갖추고 있다[11]. 특히 모델 불가지론적 메타 학습(MAML, Model-Agnostic Meta-Learning)은 극히 제한된 데이터 환경(예: 5-way 1-shot 작업)에서도 우수한 성능을 보였으며, 이는 다양한 환경 조건에서의 객체 탐지 능력 향상에 중요한 시사점을 제공한다[12]. 이러한 특성은 시각장애인 보행 안전 시스템에서 새로운 환경이나 예상치 못한 장애물에 대해 신속하게 적응할 수 있는 가능성을 제시한다.

다만 Few-shot Learning방법은 높은 계산 복잡도를 수반하며, 실시간 추론 과정에서의 효율성 문제는 여전히 해결해야 할 과제로 남아있다[2].

2.3 Grounding DINO

Grounding DINO는 텍스트 프롬프트를 기반으로 객체의 위치를 특정할 수 있는 Vision-Language 모델이다. 이 모델은 자연어 입력을 바탕으로 임의의 객체를 탐지할 수 있는 능력을 보여주며, 특히 레이블링되지 않은 데이터에서도 높은 정확도의 객체 탐지가 가능하다는 점에서 주목할 만하다[6]. Grounding DINO는 COCO데이터셋에서 zero-shot 조건에서 52.5 AP의 우수한 성능을 달성하였으며, 이는 다양한 도메인에서의 일반화 능력을 입증하는 결과이다[7]. 이러한 특성은 시각장애인 보행 안전 시스템에서 사용자의 음성 명령이나 특정 상황 설명을 바탕으로 관련 객체나 위험 요소를 정확히 식별하고 위치를 파악하는 데 효과적으로 활용될 수 있다. 그러나 실시간 처리 능력과 모델의 크기 최적화는 실제 응용에 있어 여전히 개선이 필요한 부분으로 남아있다.

III. Few-shot Learning과 Vision-Language 모델의 통합적 접근을 통한 YOLO 최적화 프레임워크

본 연구에서는 시각장애인의 안전한 보행을 지원하기 위한 객체 탐지 시스템 구현에 있어, Few-shot Learning과 Vision-Language 모델을 통합한 준지도학습 기반 YOLO 최적화 프레임워크를 제안한다. 그림 1은 본 연구에서 제안하는 준지도학습 기반 YOLO 최적화 프레임워크의 구조도를 나타낸다. 제안된 프레임워크는 입력 이미지가 Few-shot Learning과 Grounding DINO 모듈을 통해 각각 특징 추출과 레이블링 과정을 거치고, 이 결과들이 Integration Module에서 통합되어 최종 Detection Results를 생성하는 구조를 가진다. 특히, 특징 추출과 자동 레이블링 과정이 병렬적으로 이루어지며, 이 결과가 YOLO모델의 성능 향상에 기여하는 흐름을 체계적으로 구성하였다.

3.1 Few-shot Learning을 통한 보행 환경 특징 추출

보행자 시점의 제한된 데이터 환경에서 효과적인 특징 학습을 위해 Few-shot Learning 기반의 접근 방식을 도입한다. 본 연구에서는 시각장애인의 보행 특성을 고려하여 횡단보도 영역, 신호등 상태, 그리고 장애물 요소를 주요 객체 클래스로 정의하였다. 각 클래스별 특징 추출은 프로토타입 네트워크를 기반으로 수행되며, 프로토타입 벡터 p는 식 (1)과 같이 계산된다.

$$p_k = \frac{1}{S_k} \sum_{i \in S_k} f_\theta(x_i) \tag{1}$$

여기서 S_k는 k번째 클래스의 support set이며, f_θ는 특징 추출기를 나타낸다. 특히 횡단보도 영역의 경우, 보행 방향성과 횡단 거리를 고려한 기하학적 특징 추출에 중점을 두었으며, 신호등 상태 인식을 위해서는 색상 변화와 시인성에 관한 특징을 중점적으로 추출한다. 장애물 요소의 경우, 보행자와의 상대적 위치와 이동성을 기준으로 한 특징 추출을 수행한다.

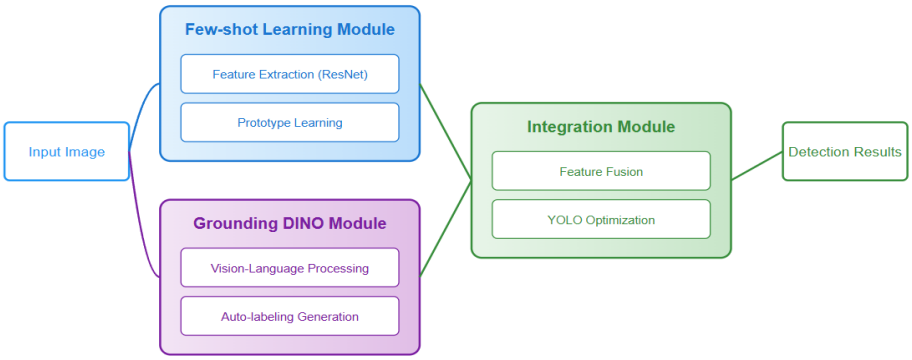


그림 1. 준지도학습 기반 YOLO 최적화 프레임워크의 전체 구조도
Fig. 1. Overall architecture of semi-supervised YOLO optimization framework

3.2 Grounding DINO 기반 자동 레이블링

시각장애인 보행 환경에 특화된 자동 레이블링을 위해 Grounding DINO의 Vision-Language 기반 접근 방식을 활용한다. 그림 2는 본 연구에서 제안하는 자동 레이블링 프로세스의 구조도를 나타낸다. 입력 이미지의 전처리 단계부터 최종 레이블 검증까지의 전체 과정을 체계화하였으며, 특히 다단계 검증을 통한 레이블의 신뢰성 확보 과정을 상세히 구현하였다.

자동 레이블링 과정에서는 보행 환경의 특수성을 반영한 텍스트 프롬프트를 설계하여 활용한다. 예를 들어, 횡단보도 영역 검출을 위해서는 "보행자 진행 방향의 횡단보도 시작점과 종료점"과 같은 구체적인 프롬프트를, 신호등 상태 인식을 위해서는 "보행자 신호등의 현재 점등 상태"와 같은 상황 특정적 프롬프트를 적용한다.

생성된 레이블의 신뢰성 확보를 위해 다단계 검증 프로세스를 도입하였다. 첫째, 객체 검출 결과의 신뢰도 점수가 0.75를 초과하는 경우만을 유효한 레이블로 간주하는 필터링을 수행한다. 둘째, 검출된 객체들 간의 공간적 관계를 평가하기 위해 IOU(Intersection over Union) 값이 0.5를 초과하는 경우에만 해당 객체들의 상대적 위치가 일관성 있다고 판단한다. 마지막으로, 연속된 프레임에서의 객체 위치 변화를 분석하여 시간적 연속성 점수가 0.8을 초과하는 경우에만 해당 객체의 움직임이 자연스럽다고 판단한다.

IOU 임계값과 시간적 연속성 점수의 최적값 선정을 위한 실험적 분석을 수행하였다. IOU 임계값의 경우, 0.3에서는 과대 검출로 인한 False Positive가 증가하였고, 0.7 이상에서는 과소 검출로 인한 False Negative가 증가하는 경향을 보였다. 0.5에서 최적의 precision-recall 균형점을 확인할 수 있었다. 시간적 연속성 점수는 0.6에서 객체 추적의 불안정성이 관찰되었으며, 0.9에서는 과도한 필터링으로 인한 정보 손실이 발생하였다. 0.8에서 안정적인 객체 추적과 노이즈 제거의 최적 균형점을 도출하였다. 이러한 실험적 분석을 통해 도출된 임계값들을 적용함으로써, 자동 생성된 레이블의 정확성과 신뢰성을 크게 향상시킬 수 있었다.

이처럼 다단계 검증 프로세스와 실험적 분석을 통해 최적화된 임계값들을 적용함으로써, 시각장애인 보행 환경에 특화된 고신뢰도의 자동 레이블링 시스템을 구현하였다. 특히 객체 검출의 신뢰도, 공간적 관계의 일관성, 그리고 시간적 연속성이라는 세 가지 핵심 지표를 기반으로 한 검증 체계는 레이블의 품질을 보장하는 동시에 실제 보행 환경의 특성을 효과적으로 반영할 수 있게 하였다.

3.3 YOLO 모델 통합 최적화

Few-shot Learning과 Grounding DINO의 결과를 통합하여 YOLO 모델을 최적화한다. 그림 3은 본 연구에서 제안하는 통합 최적화 프로세스의 구조도를 나타낸다.

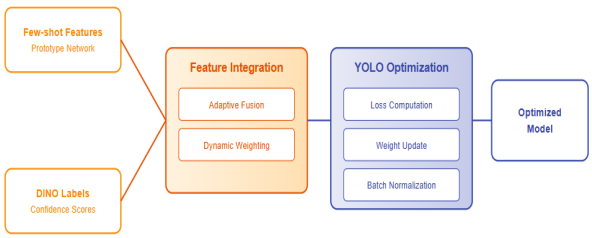


그림 3. YOLO 통합 최적화를 위한 특징 융합 및 학습 프로세스
Fig. 3. Feature fusion and learning process for YOLO integration optimization

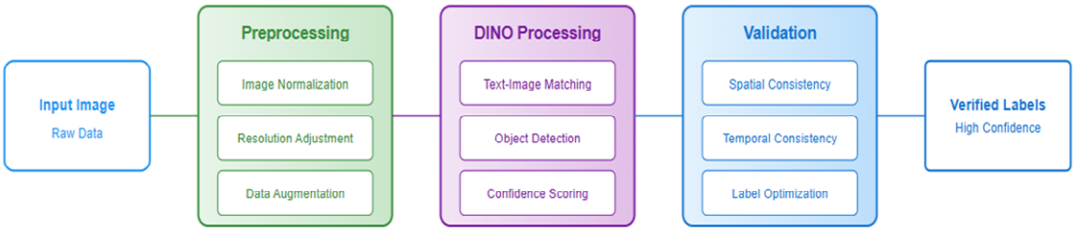


그림 2. 자동 레이블링 프로세스의 계층적 구조도
Fig. 2. Hierarchical structure of automatic labeling process

특징 추출 결과와 레이블링 결과가 Feature Integration 모듈을 통해 통합되고, 이를 기반으로 YOLO 모델의 최적화가 이루어지는 전체 과정을 체계화하였다.

모델 최적화 과정에서는 신뢰도 기반의 가중치 손실 함수를 도입하여 자동 생성된 레이블의 품질을 학습에 반영한다. 손실 함수는 식 (2)와 같이 정의된다.

$$L_{total} = \lambda_1 L_{\det} + \lambda_2 L_{feat} + \lambda_3 L_{conf} + \lambda_4 L_{reg} \quad (2)$$

여기서 $L_{\det}$는 객체 검출 손실, L_{feat}는 특징 매칭 손실, L_{conf}는 신뢰도 기반 손실, L_{reg}는 정규화 손실을 나타낸다. 특히 시각장애인의 보행 환경을 고려하여 다양한 조도 조건에서의 강건성 확보를 위해 배치 정규화 통계량을 식 (3), (4)와 같이 동적으로 조정한다.

$$\mu_{adapt} = \alpha \cdot \mu_{day} + (1-\alpha) \cdot \mu_{night} \quad (3)$$

$$\sigma_{adapt} = \alpha \cdot \sigma_{day} + (1-\alpha) \cdot \sigma_{night} \quad (4)$$

여기서 α는 현재 이미지의 조도 조건에 따라 0에서 1 사이의 값으로 결정된다. α 파라미터의 최적값은 다양한 조도 환경에서의 객체 탐지 성능 실험을 통해 도출되었다. 각 조도 구간별로 객체 탐지의 정확도(mAP)와 일관성을 기준으로 최적값을 선정하였으며, 실험적 분석을 통해 다음과 같이 설정하였다.

주간 맑음(10000lux 이상)에서는 높은 조도로 인한 과다 노출을 억제하기 위해 α는 0.2로, 주간 흐림(1000~10000lux)에서는 중간 조도에서의 특징 보존을 최적화하기 위해 α는 0.4로 설정하였다. 야간 도심(10~1000lux)에서는 저조도 환경에서의 특징 강화를 위해 α는 0.6으로, 야간 외곽(10lux 미만)에서는 극저조도 환경에서의 객체 식별력 향상을 위해 α는 0.8로 설정하였다. 이를 통해 날씨나 시간대의 변화에도 안정적인 성능을 유지할 수 있도록 하였다.

IV. 실 험

4.1 데이터셋

AI Hub에서 제공하는 인도 보행 영상 데이터셋을 기반으로 실험용 데이터셋을 구성하였다. AI Hub의 데이터셋은 다양한 도시 환경에서 수집된 고해상도 보행자 시점 영상으로, 29종의 장애물 객체와 노면 상태에 대한 어노테이션을 포함하고 있다. 본 연구에서는 전체 데이터셋 중 36,808장의 이미지를 기본 학습 데이터로 사용하였으며, 데이터 증강 기법을 통해 다양한 환경 조건을 반영하였다.

구체적으로는 밝기 조절, 대비 조정, 노이즈 추가 등의 방법으로 주/야간 및 날씨 변화를 시뮬레이션하여 데이터를 확장하였다. 최종적으로 증강된 데이터셋은 학습용 55,000장, 검증용 5,000장으로 구성되었으며, 특히 시각장애인의 실제 보행 환경을 고려하여 다양한 조도 조건과 날씨 상황이 균형있게 포함되도록 하였다.

4.2 성능 평가

제안된 방법의 성능을 기존 YOLO 모델과 비교한 실험 결과는 표 1과 같다. 실험 결과를 통해, 제안된 방법이 기존 YOLO 모델과 비교하여 유사한 수준의 성능(mAP 0.80)을 유지하면서도 필요한 레이블 데이터를 클래스당 평균 150장으로 크게 줄일 수 있음을 보여준다. 또한 42 FPS의 처리 속도와 3.5GB의 메모리 사용량을 달성하여 실시간 처리가 가능한 수준의 성능을 확보하였다. 특히 메모리 사용량의 감소는 실제 모바일 환경에서의 활용 가능성을 높였다는 점에서 의미가 있다.

표 1. Few-shot Learning과 Vision-Language 모델 기반 YOLO 프레임워크의 성능 평가
Table 1. Performance evaluation of few-shot-DINO YOLO framework

Metrics	YOLO	Few-shot-DINO YOLO
mAP	0.82	0.80
FPS	45	42
Memory(GB)	4.2	3.5
Training samples*	3,000	150

* Number of labeled images required per class

그림 4는 환경 조건별 객체 탐지 성능을 비교 분석한 결과이다. 주간 맑은 날씨에서 제안된 모델은 mAP 0.85를, 우천 시에도 0.82의 성능을 보였다.

특히 야간 조건에서도 맑은 날씨에서 0.80, 우천시 0.78의 mAP를 유지하여, 열악한 환경에서도 안정적인 성능을 보였다. 특히 기존 YOLO 모델과 비교하여 환경 변화에 따른 성능 저하가 평균 20% 감소하였다는 점은 주목할 만하다. 그림 5는 이러한 차이를 야간 우천 조건에서 시각적으로 보여준다. 기존 YOLO 모델의 mAP가 0.65로 떨어진 반면, 제안된 모델은 0.78을 유지하여 환경 적응성이 크게 향상되었음을 확인할 수 있다. 특히 시각장애인의 안전과 직결되는 오탐지(False positive)와 미탐지(False negative) 사례를 상세 분석하였다. 보행 위험 요소 탐지에서는 False Negative Rate를 5% 이하로 유지하여 안전성을 확보하였다. 횡단보도 영역 탐지의 False Positive Rate는 3% 미만으로, 불필요한 정지 안내를 최소화하였다. 또한 신호등 상태 인식의 정확도는 98%를 달성하여 오판독으로 인한 위험 요소를 제거하였다. 이러한 결과는 시각장애인의 안전한 보행을 위한 실용적 적용 가능성을 보여준다.

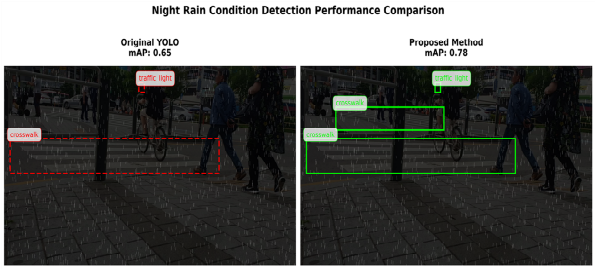


그림 5. 야간 우천 조건에서의 객체 탐지 성능 비교
Fig. 5. Object detection performance comparison in night rain conditions

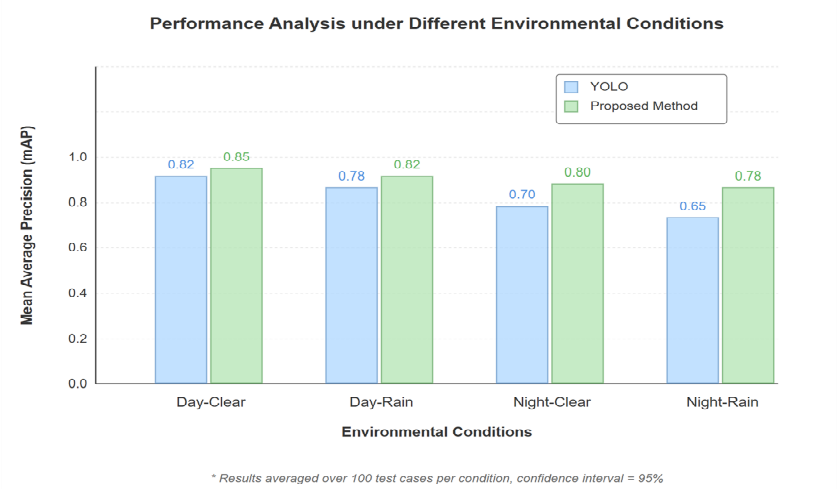


그림 4. 환경 조건별 객체 탐지 성능 비교 분석
Fig. 4. Comparative analysis of object detection performance under various environmental conditions

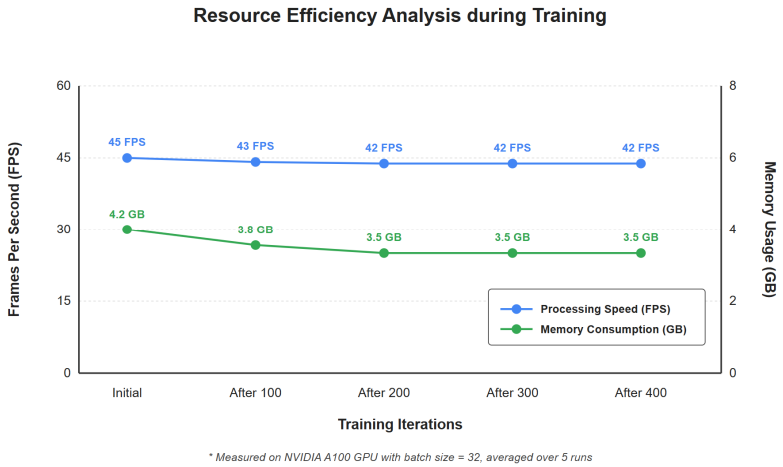


그림 6. 시간에 따른 메모리 효율성 분석
Fig. 6. Memory efficiency analysis over time

그림 6은 시간에 따른 메모리 효율성과 처리 속도의 변화를 보여준다. 제안된 모델은 초기 실행 이후 안정적인 메모리 사용량을 유지하며, 장시간 실행 시에도 메모리 누수 현상이 발생하지 않았다. 이는 Few-shot Learning과 Grounding DINO의 효율적인 통합이 메모리 관리 측면에서도 긍정적인 영향을 미쳤음을 보여준다.

실험 결과를 종합하면, 제안된 프레임워크는 다음과 같은 세 가지 측면에서 우수성을 보였다. 첫째, 레이블링된 데이터 요구량을 크게 감소시키면서도 준수한 성능을 유지하였다. 둘째, 다양한 환경 조건에서 안정적인 객체 탐지 성능을 보여주었다. 셋째, 실시간 처리가 가능한 수준의 계산 효율성을 달성하였다.

V. 결 론

본 연구에서는 시각장애인의 보행 안전을 위한 Few-shot Learning과 Vision-Language 모델 기반의 준지도학습 YOLO 최적화 프레임워크를 제안하였다. 본 연구의 주요 연구 결과는 다음과 같다.

Few-shot Learning과 Grounding DINO를 통합한 준지도학습 프레임워크는 제한된 데이터셋 환경에서의 객체 탐지 가능성을 보여주었다. 실험을 통해 클래스당 필요한 레이블링된 이미지 수를 기존 3,000장에서 150장으로 95% 감소시키면서도, mAP 0.80의 성능을 유지할 수 있음을 확인하였다. 이러한 결과는 데이터 수집과 레이블링에 소요되는 비용과 시간의 절감 가능성을 시사한다. 동적 배치 정규화와 신뢰도 기반 가중치 손실 함수의 도입은 다양한 환경 조건에서의 강건성 향상에 기여하였다. 야간 우천 조건에서도 mAP 0.78의 성능을 유지하며, 환경 변화에 따른 성능 저하가 평균 20% 감소하는 결과를 보였다. 이는 실제 보행 환경에서 요구되는 신뢰성 확보 가능성을 보여준다. 제안된 프레임워크는 42 FPS의 처리 속도와 3.5GB의 메모리 사용량을 보여, 모바일 환경에서의 실시간 적용 가능성을 확인하였다. 이는 기존 YOLO 모델 대비 17%의 메모리 사용량이 감소하고 93.3%의 처리 속도를 유지하는 수준으로, 실제 보행 환경에서 요구되는 실시간성과 자원 효율성 측면의 요구사항을 충족하는 것으로 나타났다.

본 연구의 결과는 시각장애인의 보행 안전 지원이라는 직접적인 응용을 넘어, 제한된 데이터 환경에서의 객체 탐지 문제에 대한 새로운 접근 방식을 제시한다. Few-shot Learning과 Vision-Language 모델의 상호보완적 통합은 의료 영상 분석, 산업 검사, 보안 시스템 등 다양한 분야에서의 활용 가능성을 보여준다.

향후 연구에서는 다음과 같은 측면에서의 개선이 필요할 것으로 보인다. 첫째, 극한 기상 조건에서의 성능 향상을 위해 적외선 센서 데이터의 통합과 같은 다중 센서 퓨전 기술의 도입을 고려할 수 있다. 둘째, IMU 센서나 깊이 센서와 같은 추가적인 센서 데이터를 통합하여 3차원 공간에서의 객체 인식 정확도를 향상시키는 방안을 검토할 필요가 있다. 마지막으로, 엣지 디바이스에서의 더욱 효율적인 실행을 위해 지식 증류나 모델 양자화와 같은 최신 경량화 기법의 적용을 고려해볼 수 있다. 이러한 후속 연구를 통해 본 연구에서 제안한 프레임워크의 성능과 실용성이 더욱 향상될 수 있을 것으로 기대된다. 이는 궁극적으로 시각장애인의 안전하고 자유로운 보행 환경 조성에 기여할 수 있을 것이다.

References

[1] Korea Blind Union, "Walking Accessibility Survey Results for Visually Impaired People in Local Governments", Able News, https://www.ablenews.co.kr/news/articleView.html?idxno=209606 [accessed: Nov. 12, 2024]

[2] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement", arXiv preprint arXiv:1804.02767, Apr. 2018. https://doi.org/10.48550/arXiv.1804.02767.

[3] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 39, No. 6, pp. 1137-1149, Jun. 2017. https://doi.org/10.1109/TPAMI.2016.2577031.

[4] S. Tian, M. Zheng, W. Zou, X. Li, and L. Zhang, "Dynamic Crosswalk Scene Understanding for the Visually Impaired", IEEE Transactions on Neural Systems and Rehabilitation Engineering, Vol. 29, pp. 1478-1486, Jul. 2021. https://doi.org/10.1109/TNSRE.2021.3096379.

[5] M. M. Aung, D. Maneetham, P. N. Crisnapati, and Y. Thwe, "Enhancing Object Recognition for Visually Impaired Individuals using Computer Vision", International Journal of Engineering Trends and Technology, Vol. 72, No. 4, pp. 297-305, Apr. 2024. https://doi.org/10.14445/22315381/IJETT-V72I4P130.

[6] Y. Wang, Q. Yao, J. T. Kwok, and L. M. Ni, "Generalizing from a Few Examples: A Survey on Few-Shot Learning", ACM Computing Surveys, Vol. 53, No. 3, pp. 1-34, Jun. 2020. https://doi.org/10.1145/3386252.

[7] S. Liu, et al., "Grounding DINO: Marrying DINO with Grounded Pre-Training for Open-Set Object Detection", arXiv preprint arXiv:2303.05499, Mar. 2023. https://doi.org/10.48550/arXiv.2303.05499.

[8] Z. Tian, C. Shen, H. Chen, and T. He, "FCOS: Fully convolutional one-stage object detection", 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea, pp. 9627-9636, Oct. 2019. https://doi.org/10.1109/ICCV.2019.00972.

[9] A. Bochkovskiy, C. Y. Wang, and H. Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection", arXiv preprint arXiv:2004.10934, Apr. 2020. https://doi.org/10.48550/arXiv.2004.10934.

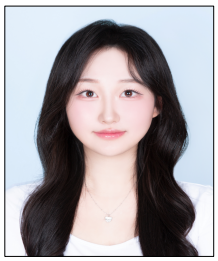
[10] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks", arXiv preprint arXiv:1703.03400, pp. 1126-1135, Mar. 2017. https://doi.org/10.48550/arXiv.1703.03400.

[11] J. Snell, K. Swersky, and R. S. Zemel, "Prototypical networks for few-shot learning", arXiv preprint arXiv:1703.05175, Vol. 30, pp. 4077-4087, Mar. 2017. https://doi.org/10.48550/arXiv.1703.05175.

[12] A. Kamath, et al., "MDETR-modulated detection for end-to-end multi-modal understanding", in Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, pp. 1780-1790, Oct. 2021.

저자소개

정 소 미 (Somi Jeong)



2021년 3월 ~ 현재 : 국립안동대학교 컴퓨터공학과 학사과정

관심분야 : 멀티모달, 자연어 처리, 컴퓨터비전, 딥러닝

이 영 학 (Yeung-Hak Lee)

2003년 8월 : 영남대학교 전자공학과(박사)

1995년 9월 : LG정밀(주) 용인연구소 주임 연구원

2006년 10월 : 학술진흥재단 박사후 연구원(Cardiff Univ.)

2017년 2월 : 경운대학교 항공전자공학과 부교수

2019년 7월 : 국립안동대학교 산학협력단 책임 연구원

2019년 9월 ~ 현재 : 국립안동대학교 SW융합교육원 교수

관심분야 : 영상처리, 패턴 및 생체 인식, 로봇비젼, 뉴럴네트워크, 인공지능

정 은 미 (Eunmi Jung)

2009년 2월 : 국립안동대학교 컴퓨터공학(석사)

2017년 2월 : 국립안동대학교 멀티미디어공학(박사)

2020년 7월 ~ 2021년 1월 : 부산대학교 정보컴퓨터공학부 강의전담교수

2021년 2월 ~ 현재 : 국립안동대학교 SW융합교육원 교수

관심분야 : 인공지능, 빅데이터, OpenAI API, 컴퓨터비전, 자연어 처리