특집논문 (Special Paper)
방송공학회논문지 제29권 제6호, 2024년 11월 (JBE Vol.29, No.6, November 2024)
https://doi.org/10.5909/JBE.2024.29.6.783
ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

깊이 추정 기반 데이터 정제를 통한 인물 검출 성능 향상

송 준 호[a)], 홍 민 수[b)], 김 영 길[a)‡], 김 정 래[a)‡]

Improving Person Detection Performance via Depth Estimation-Based Data Refinement

Junho Song[a)], Minsoo Hong[b)], Younggil Kim[a)‡], and Jeong-Rae Kim[a)‡]

요 약

본 논문에서는 객체의 경계 상자들 간의 가림 현상으로 인해 객체 검출 성능이 저하되는 문제를 해결하기 위하여 깊이 추정 모델을 활용한 데이터 정제 방법을 이용하여 객체를 탐지하는 방법을 제안한다. 구체적으로, Stable diffusion을 활용한 깊이 정보 추출 방법인 Marigold 모델을 사용하여 2D 이미지로부터 깊이 정보를 추출하고, 이를 통하여 겹쳐진 객체 중에서 가장 앞에 위치한 객체의 데이터를 보존하여 객체 검출 성능을 향상시킨다. 기존 뮤직뱅크 데이터세트를 대상으로 수행한 연구를 확장하여 공개된 데이터세트인 DanceTrack과 CrowdHuman 데이터세트를 대상으로 실험을 진행한 결과, 두 데이터세트 모두에서 mAP가 향상되었으며, 특히 중간 크기 객체의 검출 성능이 크게 향상되었다. 이러한 결과는 제안된 데이터 정제 방법의 일반성을 입증하였으며, 다양한 데이터세트에 대한 적용 가능성을 시사한다.

Abstract

In this paper, we introduce a method to refine data using a depth estimation model, aiming to improve object detection performance that is often hindered by occlusion. We utilize the Marigold model to extract depth information from 2D images. By keeping only the data of the foremost object in cases where objects overlap, the method enhances detection accuracy. We expanded our previous research, which was based on the Music Bank dataset, by conducting experiments on the Dancetrack and CrowdHuman datasets. The results showed an increase in mAP scores for both datasets, with a notable improvement in detecting medium-sized objects. These outcomes demonstrate the general applicability of proposed data refinement method and suggest its potential for broader dataset applications.

Keyword : Artificial Intelligence, Deep Learning, Object Detection, Depth Estimation

a) 서울시립대학교(University of Seoul)
b) 한국방송공사(Korean Broadcasting System)
‡ Corresponding Author : 김영길(Younggil Kim), 김정래(Jeong-Rae Kim)
E-mail: ygkim72@uos.ac.kr, jrkim@uos.ac.kr
Tel: +82-2-6490-2340, +82-2-6490-2616
ORCID: https://orcid.org/0000-0001-7066-0555, https://orcid.org/0000-0002-3261-7238
※이 논문의 연구 결과 중 일부는 한국방송·미디어공학회 2024년 하계학술대회에서 발표한 바 있음.
※본 연구는 문화체육관광부 및 한국콘텐츠진흥원의 2024년도 문화체육관광 연구개발사업으로 수행되었음(연구개발과제명: 공연 콘텐츠의 고해상도(8K/16K) 서비스를 위한 AI 기반 영상확장 및 서비스 기술개발, 연구개발과제번호: RS-2024-00395886, 기여율: 100%).
· Manuscript August 30, 2024; Revised September 24, 2024; Accepted September 24, 2024.

Ⅰ. 서 론

객체 탐지(Object Detection)는 컴퓨터 비전의 핵심 분야 중 하나로서 자율주행, 감시 시스템, 증강 현실 등 다양한 응용 분야에서 사용되며, 그 목적은 이미지 속 객체를 정확하게 식별하고, 그 객체의 위치를 추정하는 것이다. 그러나 이러한 목표를 달성하는 과정에서 어려움이 존재하며, 그 중에서도 가림 현상(Occlusion)은 객체 탐지에서 가장 큰 문제로 꼽힌다.

가림 현상은 하나의 객체가 다른 객체에 의해 부분적으로 또는 완전히 가려지는 현상을 의미하며, 이는 객체 탐지 모델이 가려진 객체를 올바르게 인식하지 못하게 하거나, 잘못된 위치와 크기의 경계 상자(Bounding Box)를 생성하게 만드는 주요 원인이 된다. 특히 가림 현상이 발생할 경우, 가려진 객체의 특징이 충분히 표현되지 않아 탐지 성능이 크게 저하될 수 있다. 이로 인해 뒤에 위치한 객체의 경계 상자가 오히려 앞에 있는 객체의 정보를 많이 포함하게 되고, 이는 탐지 결과의 신뢰도를 크게 저하시킬 수 있다. 예를 들어, 사람이 다른 사람이나 물체에 의해 가려진 경우, 모델은 뒤에 있는 사람을 일부만 인식하거나, 앞에 있는 사람의 일부를 뒤에 있는 사람의 경계 상자에 포함시키는 오류를 범할 수 있다. 이러한 오류는 객체 탐지의 정확도와 정밀도를 크게 저하시킨다.

가림 현상 문제를 해결하기 위해 다양한 접근법들이 제안되어 왔다. 첫 번째로, 부분 기반 모델(Part-based Model)은 객체를 여러 부분으로 나누어 각 부분을 독립적으로 탐지한 후 이를 결합하여 전체 객체를 재구성한다. 이를 통해 일부가 가려져 있어도 나머지 부분을 통해 객체를 탐지할 수 있다. 두 번째로 다중 스케일 피라미드(Multiscale Pyramid) 방식은 다양한 크기와 스케일로 특징을 추출하여 가림 현상에 덜 민감한 탐지를 가능하게 한다. 이 방법은 작은 크기의 객체나 부분적으로 가려진 객체를 탐지하는 데 유리하다. 데이터 증강 기법도 자주 활용되며, 가림 현상을 인위적으로 추가하여 모델이 다양한 가림 현상 상황에 대응할 수 있도록 한다.

최근에는 합성곱 신경망(CNN) 기반 객체 탐지 모델이 발전하면서 여러 계층 특징 맵을 결합하여 강인한 객체 탐지를 가능하게 하였으며, 어텐션 메커니즘(Attention Mechanism)을 활용해 가림 현상이 덜한 영역에 더 높은 가중치를 부여하는 방식도 제안되었다. 그러나 이러한 접근법들조차도 여전히 뒤에 있는 객체의 경계 상자가 앞의 객체 정보를 과도하게 포함하는 문제를 완전히 해결하지는 못했다.

이에 본 연구에서는 가림 현상이 발생하는 상황에서 깊이 추정 모델을 활용한 깊이 맵(Depth Map) 계산을 통해 앞의 객체와 뒤의 객체를 효과적으로 구분하고 앞의 객체 데이터를 보존하는 데이터세트 정제 방식을 제안하고자 한다. 제안하는 방법은 가림 현상이 있는 상황에서 앞에 있는 객체를 탐지함으로써 객체 검출에서의 mAP(mean Average Precision) 성능을 향상시켰으며 이를 통해 본 연구는 객체 탐지에서 가림 현상 문제를 해결하는 데 기여하고자 한다.

Ⅱ. 관련 연구

1. 객체 탐지

객체 탐지는 이미지에서 객체의 위치와 클래스를 동시에 예측하는 기술로, 딥러닝 기반의 다양한 모델들이 개발되어 왔다. 그 중에서도 YOLO[1]는 객체 탐지 분야에서 널리 사용되는 실시간 모델로, 이미지 전체를 한 번에 처리하여 객체의 위치와 클래스를 예측한다. 즉, YOLO는 이미지에서 격자(grid)를 생성하고, 각 격자 셀에 대해 경계 상자와 그 안의 객체 클래스를 예측한다. 이 모델의 주요 장점은 높은 처리 속도와 비교적 간단한 아키텍처에 있으며, 한 번의 네트워크 전달로 객체 탐지를 수행하므로 매우 빠르게 동작할 수 있다. 이러한 특성 덕분에 YOLO는 실시간 응용 프로그램에서 자주 사용되며, 다양한 버전으로 발전하면서 성능이 지속적으로 향상되고 있으며, 간편한 접근성으로 인하여 객체 탐지 방법으로 인기가 많다.

또한, 최근에는 트랜스포머(Transformer) 아키텍처를 도입한 DETR(Detection Transformers)[2]이 개발되었다. DETR은 CNN 기반 객체 탐지 모델들과는 다른 접근 방식을 채택하며, 이미지의 특징 맵을 추출한 후, 트랜스포머를 통해 이미지 내 모든 픽셀 간의 관계를 학습하고 이를 기반으로 객체의 위치와 클래스를 예측한다. 이 모델은 객체 탐지를 일종의 집합 예측 문제로 정의하여, 앵커 박스(Anchor

Box)나 NMS(Non-Maximum Suppression)와 같은 전통적인 기법들을 사용하지 않는다. DETR의 주요 장점은 단순한 아키텍처와 높은 예측 성능에 있으며, 복잡한 장면에서도 객체 간의 상호작용을 잘 포착할 수 있다. 그러나 DETR은 수렴 속도가 느리고, 초기 위치 예측이 부정확할 경우 경계 상자와 클래스 레이블 예측이 불안정해질 수 있는 한계가 있다.

이러한 DETR의 한계를 극복하기 위해 개발된 모델이 Co-DETR(Conditional Detection Trans-formers)[3]이다. Co-DETR은 DETR의 트랜스포머 구조를 계승하면서, 조건부 쿼리(Conditional Queries)라는 새로운 메커니즘을 도입하여 객체 탐지의 정밀도를 크게 향상시켰다. 조건부 쿼리는 객체의 위치와 크기에 대한 초기 정보를 제공하여, 학습 효율성을 높이고 더 정밀한 경계 상자 예측을 가능하게 한다. 이러한 구조적 개선 덕분에 Co-DETR은 다양한 벤치마크 데이터세트에서 SOTA(State-of-the-Art) 성능을 달성했으며, 복잡한 장면에서도 매우 높은 탐지 정확도를 보인다. 본 연구에서는 Co-DETR 모델을 베이스라인으로 사용하여 연구를 진행하였다.

2. 단안 깊이 추정

단안 깊이 추정(Monocular Depth Estimation)은 단일 이미지에서 깊이 정보를 추정하는 기술로, 초기에는 주로 기하학적 모델이나 규칙 기반 접근법에 의존하였다. 그러나 딥러닝의 발전과 함께 이 기술도 큰 변화를 겪었다. 2014년 딥러닝을 활용한 최초의 단안 깊이 추정 모델이 제안되었으며[4], 단일 이미지에서 다중 스케일의 깊이 정보를 예측하는 방식을 도입하였다. 이 연구는 심층 신경망을 통해 넓은 범위의 깊이 정보를 효과적으로 추정할 수 있음을 보여주었으며, 단안 깊이 추정의 기초를 다진 중요한 연구로 평가받고 있다. 이후 DCNF-FCSP[5] 모델은 조건부 무작위 필드(Conditional Random Fields, CRFs)를 딥러닝과 결합하여 이미지의 구조적 정보를 유지하면서도 깊이를 더 세밀하게 예측할 수 있는 모델이다. 이러한 접근은 딥러닝 기반 예측의 정확도를 높이는 데 크게 기여하였다.

또한, 2020년 발표된 MiDaS(Mixed Depth Scales)[6] 모델이 있다. MiDaS는 다양한 깊이 데이터세트에서 학습된 모델로, 여러 해상도와 다양한 장면에서도 안정적인 성능을 발휘할 수 있다. 이 모델은 특히 일반화 성능이 뛰어나, 다양한 응용 환경에서 깊이 정보를 일관되게 예측할 수 있다. MiDaS의 핵심은 다양한 데이터세트를 통해 학습함으로써 특정 환경이나 데이터세트에 치우치지 않는 깊이 추정 능력을 갖추었다는 것이다. 이러한 강점 덕분에 MiDaS는 다양한 응용 분야에서 널리 사용되고 있다.

최근에는 Marigold[7]라는 모델이 등장하여 깊이 추정의

그림 1. Marigold를 사용한 깊이 추정
Fig. 1. Depth estimation using Marigold

새로운 접근법을 제시하였다. Marigold는 Stable Diffusion 기반의 모델을 활용하여 깊이를 추정하는 방법으로, 이미지에서 깊이를 추정하는 작업을 조건부 디퓨전 과제로 재구성한다. 이 모델은 트랜스포머가 아닌, 대규모 이미지 생성 모델(Stable Diffusion)의 잠재 공간에서 학습을 수행하며, 이를 통해 깊이 정보를 효과적으로 추출한다. Marigold는 다양한 조건에서 깊이 추정 성능을 크게 향상시켰으며, 특히 훈련 데이터로 사용되지 않은 실제 이미지에서도 우수한 성능을 발휘한다. 또한, 제로샷(Zero-Shot) 전이 학습을 통해 다양한 데이터세트에서 SOTA(State-of-the-Art) 성능을 기록하였다. 본 연구에서는 제로샷 성능이 뛰어난 Marigold 모델을 사용하여 객체의 깊이를 추정하였다(그림 1).

Ⅲ. 깊이 추정 기반 데이터 정제 방법

깊이 추정 기반 데이터 정제 방법에 대한 순서도는 그림 2에서 확인할 수 있다. 먼저, 깊이 추정 모델을 사용하여

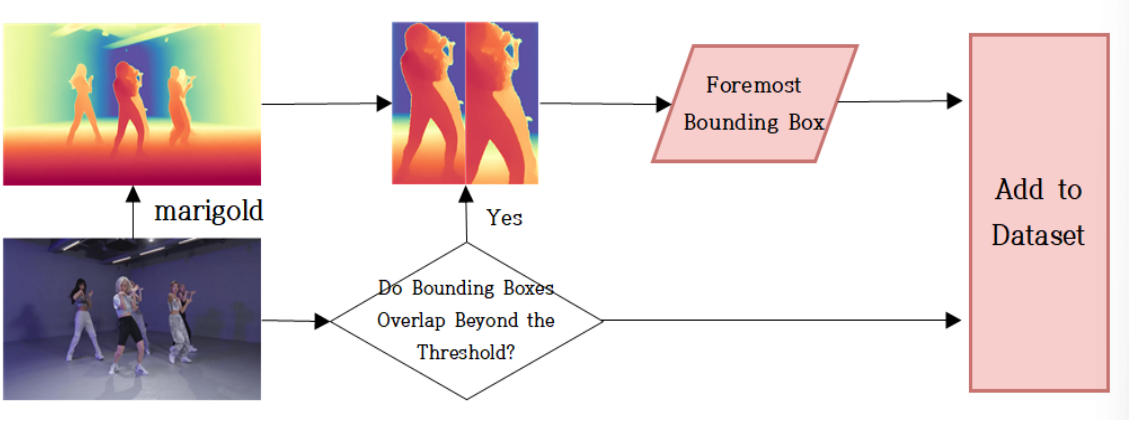

그림 2. 데이터 정제 순서도
Fig. 2. Data refinement process

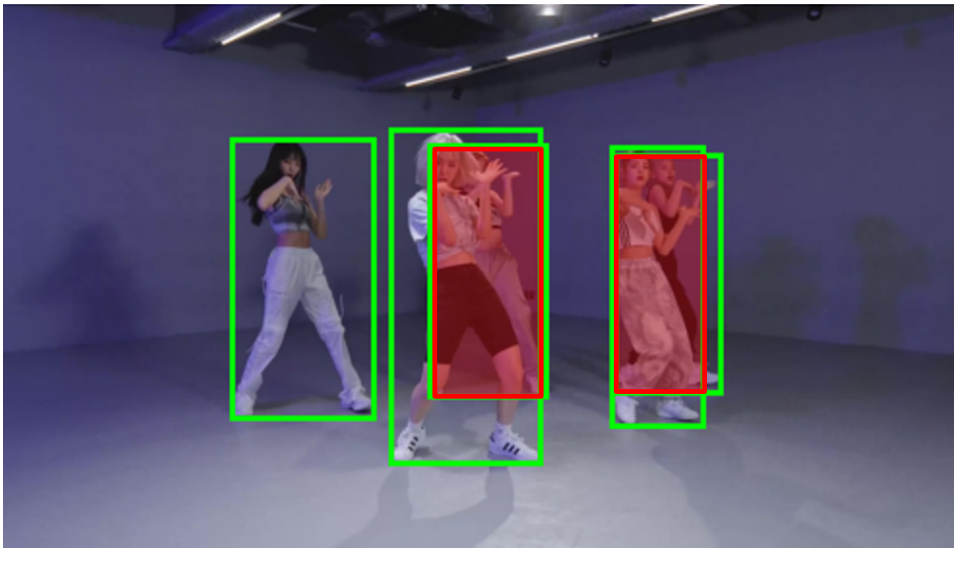

그림 3. 객체의 경계 상자마다 IoU 값을 계산하여 임곗값보다 많이 겹치는 경계 상자 추출
Fig. 3. Extracting bounding boxes with IoU exceeding the threshold for each object

이미지의 깊이 맵(Depth map)을 추출한다. 이 과정에서 깊이 추정 모델은 각 픽셀에 깊이 정보를 포함한 이미지를 생성한다. 이후, 각 객체의 경계 상자에 대해 IoU(Intersection over Union)를 계산하고 임곗값 이상의 경계 상자들을 찾는다. 이렇게 찾아낸 경계 상자들에 대해서는 추출된 깊이 지도에서 해당 상자의 관심 영역(ROI, Region Of Interest)을 추출한다(그림 3).

추출한 관심 영역에서 더 앞쪽에 위치한 경계 상자를 보존하는 것이 주된 목적이다. 가림 현상이 발생하는 경우, 뒤의 객체의 경계 상자는 앞의 객체 정보를 더 많이 포함하여 검출 시 오차가 커지게 된다. 따라서 검출 오차를 최소화하기 위해 IoU 임곗값 이상 겹치는 경계 상자에 대해 깊이

그림 4. 앞쪽 객체의 경계 상자(좌)와 뒤쪽 객체의 경계 상자(우)
Fig. 4. Foreground bounding boxes (left) and background bounding boxes (right)

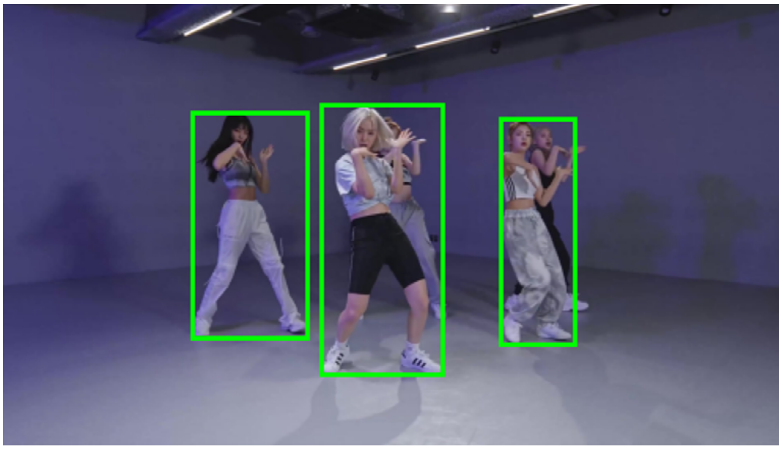

그림 5. 깊이 추정을 통하여 데이터 정제를 한 이미지의 경계 상자
Fig. 5. Bounding boxes of an image after data refinement using depth estimation

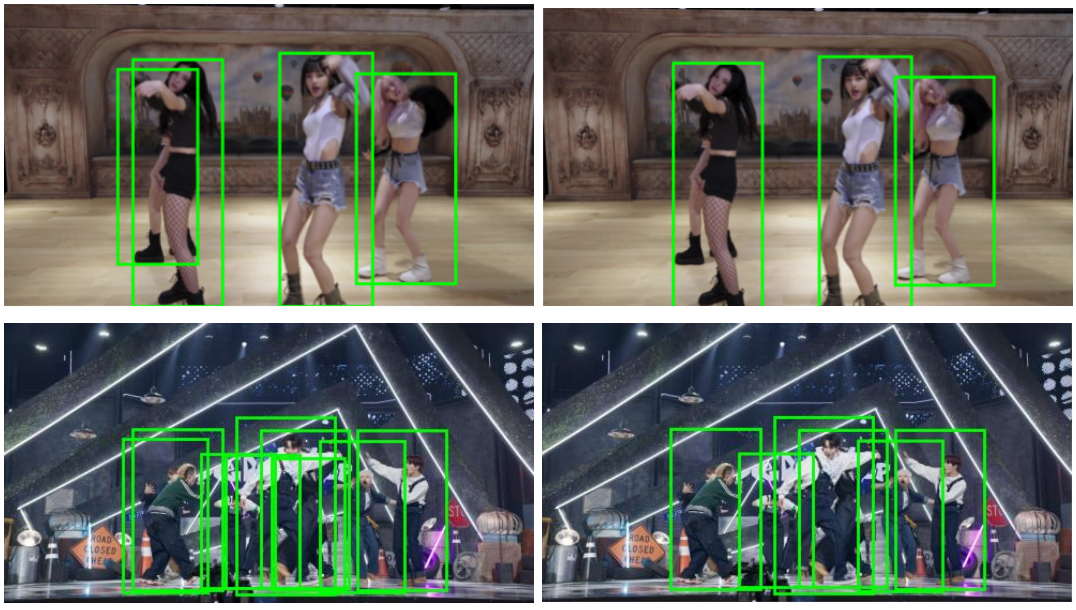

그림 6. 정제 전(왼쪽)과 후(오른쪽)의 경계 상자
Fig. 6. Bounding boxes before refinement (left) and bounding boxes after refinement (right)

맵을 기반으로 가장 앞의 경계 상자만 남기고 뒤 경계 상자를 제거하는 방식으로 데이터 정제를 수행한다. 이를 통해 검출 오차가 큰 뒤의 경계 상자를 데이터세트에서 배제함으로써 객체 검출율을 향상시킨다. Marigold 모델을 통과한 깊이 맵을 보면, 화면에서 앞쪽에 있는 객체일수록 픽셀 값이 0에 가깝게 나타나고, 반대로 배경에 속하는 부분일수록 픽셀 값이 1에 가깝게 나타난다. 이러한 특성을 활용하여, 객체가 겹쳐 있는 상황에서도 어느 객체가 더 앞에 위치해 있는지를 판별할 수 있고 객체 간의 깊이 관계를 명확하게 구분할 수 있다.

또한, 배경의 영향을 최소화하고 객체의 픽셀 값만을 비교하기 위해 배경 부분의 픽셀 값을 제외하여 객체 간 깊이 비교에서 고려하지 않았다. 이를 위해, 각 ROI 내에서 픽셀 값의 분포를 분석한 후, 상위 3사분위(Q3) 이상의 픽셀 값을 None으로 설정하여 배제하였다. 그런 다음, 각 ROI의 평균 픽셀 값을 계산하고, 이 값이 더 작은 ROI를 앞에 있는 객체의 경계 상자로 판단하여 추출하였다. 이 방법을 통해 특정 임곗값 이상 겹치는 객체들 중에서 가장 앞에 있는 객체를 효과적으로 식별하고 추출할 수 있었다(그림 5).

그림 6은 가림 현상이 발생한 다른 이미지에 대해서 데이터 정제를 수행한 예제이다. 제안한 방식으로 구한 ROI 내의 픽셀 값을 비교했을 때 더 작은 픽셀 값을 가진 ROI가 앞에 있는 객체의 ROI로 식별할 수 있다(그림 7, 8).

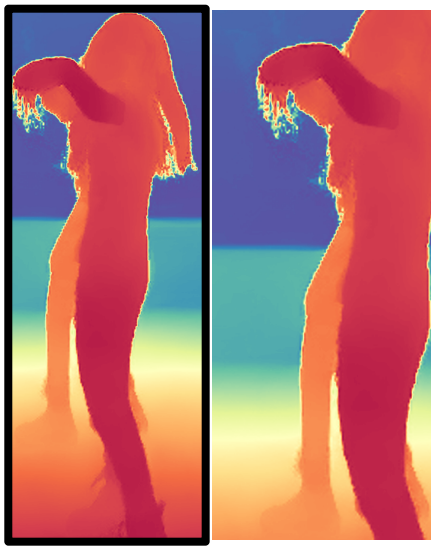

그림 7. 추출한 ROI의 depth map 평균 값 비교(왼쪽: 0.395, 오른쪽: 0.435)
Fig. 7. Comparison of the average values of the extracted ROI's depth maps (left: 0.395, right: 0.435)

그림 8. 추출한 ROI의 depth map 평균 값 비교(왼쪽: 0.279, 오른쪽: 0.312)
Fig. 8. Comparison of the average values of the extracted ROI's depth maps (left: 0. 279, right: 0. 312)

제안한 방식은 이전 연구[10]에서 소규모 데이터세트인 KBS 뮤직뱅크 이미지를 대상으로 실험을 진행한 결과, 뮤직뱅크 무대 이미지에서 인물 검출 성능이 향상됨을 확인하였다. 본 논문에서는 더 큰 규모의 데이터세트인 CrowdHuman과 DanceTrack에 적용하여 실험을 진행하였으며, 이를 통해 대규모 공공 데이터에서의 검출 성능 향상을 입증하고, 제안한 방식의 일반성을 확고히 하고자 한다.

Ⅳ. 실험

1. 데이터세트

본 연구에서는 깊이 추정을 통한 데이터 정제의 효과를 평가하기 위해 DanceTrack[8] 데이터세트와 CrowdHuman[9] 데이터세트를 사용했다. DanceTrack은 다양한 움직임을 특징으로 하는 객체 추적 데이터세트로, 참가자들이 유사한 복장을 입고 있어 시각적 식별이 어려우며, 복잡한 동작과 빈번한 위치 변경이 이루어지는 환경을 제공한다. CrowdHuman은 공공장소에서 촬영된 이미지로 구성된 대규모 인간 탐지 데이터세트로, 높은 밀도의 인간 이미지와 다양한 포즈가 특징이다. 다양한 환경에서 깊이 정보를 활용한 데이터 정제 방식이 객체 추적 알고리즘의 성능 향상에 얼마나 기여하는지 실험적으로 검증한다.

2. 실험 환경

인물 검출을 위한 모델로 Co-DETR을 선정하였고, 데이터세트의 겹친 경계 상자 정제를 위해 3장에서 제안한대로 Marigold를 사용하여 깊이 추정 기반 데이터 정제를 수행한다. 두 데이터세트의 학습 및 추론을 위해 리눅스 기반의 NVIDIA RTX 3090을 사용하였고, IoU의 임곗값은 0.7로 하여 데이터 정제를 수행했다. 빠른 실험을 위해 Co-DETR은 Tiny 모델을 사용하고, MMdetection[11]에서 제공하는 기본 파라미터 값을 그대로 사용하였다.

3. 검출 성능(mAP)

객체 탐지 분야에서 일반적으로 사용하는 mAP를 평가 지표로 사용한다. 세분화된 성능 확인을 위해 각각 IoU 임곗값이 50%(mAP@50)와 75%(mAP@75)일 때의 검출 성능을 평가하고 객체 크기에 따라 mAP@m, mAP@L로 나눈다. 비정제 데이터세트와 비교했을 때 본 논문에서 제안한 깊이 추정 기반 정제 데이터로 학습한 경우에서 mAP에서 더 높은 성능을 보이는 것을 확인할 수 있다(표 1). 이는 겹치는 객체의 적합한 경계 상자를 찾는 것이 객체 검출 성능 향상에 영향을 미친다는 것을 의미한다.

가장 앞쪽의 객체만 검출하면 나머지 객체에 대한 인물

표 1. 비정제 데이터세트와 깊이 추정을 통한 데이터세트의 객체 검출 성능 비교
Table 1. Comparison of performance between unrefined and depth estimation refined datasets

Data	model	mAP	mAP@50	mAP@75	mAP@m	mAP@L
Dance Track	baseline	0.711	0.9320	0.7510	0.4780	0.7250
	proposal	0.736	0.9440	0.7760	0.5310	0.7480
Crowd Human	baseline	0.526	0.822	0.585	0.494	0.686
	proposal	0.557	0.846	0.624	0.530	0.716

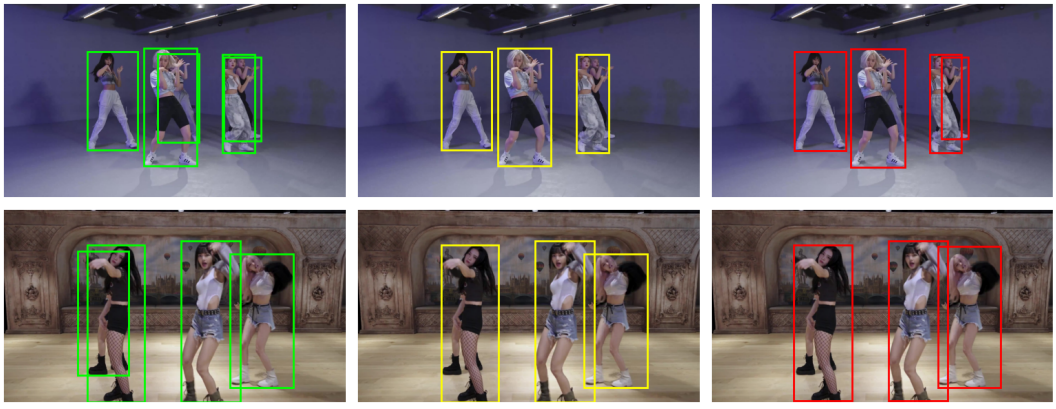

그림 9. 왼쪽부터 원본 데이터의 경계 상자, 정제한 데이터의 경계 상자, YOLO로 검출한 데이터의 경계 상자
Fig. 9. Bounding boxes of the original data (left), bounding boxes of the refined data (center), and bounding boxes of the data detected by YOLO (right)

검출 손실 여부에 대한 우려가 있었으나, 실험 결과 제안한 방법은 Dancetrack 데이터세트에서는 기존의 비정제 데이터를 사용한 방식에 비해 mAP에서 0.711에서 0.736까지 인물 검출 성능이 향상되었다. 특히 중간 크기 객체에 대한 객체 검출 성능인 mAP_m이 크게 향상되었다. CrowdHuman 데이터세트에서는 mAP에서 0.526에서 0.557까지 성능 향상을 이뤘으며 모든 mAP 지표에서 성능이 향상되었다. 추가로 YOLO 모델을 사용하여 인물 검출을 해보았을 때, 가림 현상이 일어나는 경우 앞에 있는 객체에 경계 상자가 그려지는데 이는 뒷 객체의 경계 상자가 잘못된 정보 혹은 중복된 정보를 가지고 있는 것으로 판단된다(그림 9).

4. 결과 사진

CrowdHuman 데이터세트에 대해 다양한 크기의 객체에 대해 과하게 겹친 상황에서 가장 앞의 경계 상자를 잘 검출

그림 10. CrowdHuman : 비정제 데이터세트 탐지 결과(좌), 제안한 정제 데이터세트 탐지 결과(우)
Fig. 10. Detection results on the unrefined dataset (left) and refined dataset (right) in CrowdHuman Dataset

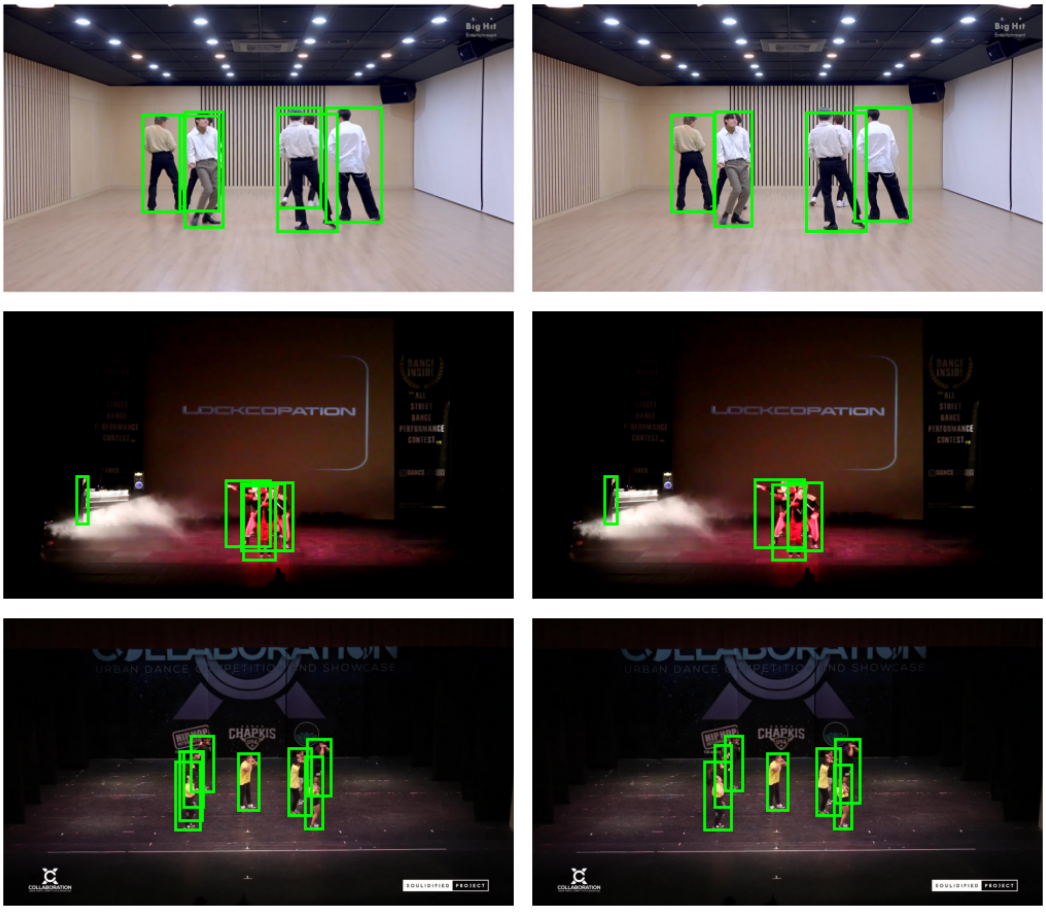

그림 11. DanceTrack : 비정제 데이터세트 탐지 결과(좌), 정제 데이터세트 탐지 결과(우)
Fig. 11. Detection results on the unrefined dataset (left) and refined dataset (right) in DanceTrack Dataset

해내는 것을 확인할 수 있고(그림 10), DanceTrack 데이터세트에 대해서도 정면에서 봤을 때와 위에서 봤을 때 모두 과하게 겹친 상황에서 가장 앞의 경계 상자만 잘 검출해내는 것을 알 수 있다(그림 11).

V. 결 론

본 논문에서는 이전 연구에서 수행했던 깊이 추정을 통해 뮤직뱅크 이미지 데이터세트를 정제하고, 이를 이용해 학습한 객체 검출 모델의 검출 성능을 향상시켰던 것을 공공 데이터에도 적용하여 인물 검출 성능을 향상시킴을 보임으로써 우리의 데이터 정제 방법의 일반성을 입증하였다. 깊이 추정 모델을 활용하여 겹치는 경계 상자 중 가장 앞의 상자를 보존하여 가림 현상으로 인한 객체 검출 성능 하락 문제를 해결할 수 있었다.

실험 결과, 제안한 방법은 비정제 데이터세트로 학습한 것에 비해 DanceTrack에서 mAP를 0.711에서 0.736으로 성능 향상을 보였으며, CrowdHuman에서는 mAP를 0.526에서 0.557까지 성능 향상을 보였다. 특히 중간 크기 객체의 검출 성능이 가장 크게 향상되었으며, 이러한 데이터 정제 방식이 다양한 데이터세트에 대해 검출 성능을 향상시킴을 기대한다.

참 고 문 헌 (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. 2016.
doi: https://doi.org/10.1109/cvpr.2016.91

[2] Carion, Nicolas, et al. “End-to-end object detection with transformers.” European conference on computer vision. Cham: Springer International Publishing, 2020.

[3] Zong, Zhuofan, Guanglu Song, and Yu Liu. “Detrs with collaborative hybrid assignments training.” Proceedings of the IEEE/CVF international conference on computer vision. 2023.
doi: https://doi.org/10.1109/iccv51070.2023.00621

[4] Eigen, David, Christian Puhrsch, and Rob Fergus. “Depth map prediction from a single image using a multi-scale deep network.” Advances in neural information processing systems 27 (2014).

[5] Liu, Fayao, et al. “Learning depth from single monocular images using deep convolutional neural fields.” IEEE transactions on pattern analysis and machine intelligence 38.10 (2015): 2024-2039.
doi: https://doi.org/10.1109/tpami.2015.2505283

[6] Ranftl, René, et al. “Towards robust monocular depth estimation: Mixing datasets for zero-shot cross-dataset transfer.” IEEE transactions on pattern analysis and machine intelligence 44.3 (2020): 1623-1637.
doi: https://doi.org/10.1109/tpami.2020.3019967

[7] Ke, Bingxin, et al. “Repurposing diffusion-based image generators for monocular depth estimation.” Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024.

[8] Sun, Peize, et al. “Dancetrack: Multi-object tracking in uniform appearance and diverse motion.” Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022.
doi: https://doi.org/10.1109/cvpr52688.2022.02032

[9] Shao, Shuai, et al. “CrowdHuman: A benchmark for detecting human in a crowd.” arXiv preprint arXiv:1805.00123 (2018).

[10] J. Song, Y. Ra, B. Park, H. Choi, M. Hong, “Improving K-POP Fancam Person Detection Performance via Diffusion Model-Based Depth Estimation” 2024 The Korean Institute of Broadcast and Media Engineers Summer Conference, Jeju, Korea, pp. 313-316, 2024.

[11] Chen, Kai, et al. “MMDetection: Open mmlab detection toolbox and benchmark.” arXiv preprint arXiv:1906.07155 (2019).

저 자 소 개

송 준 호

- 2018년 ~ 현재 : 서울시립대학교 수학과 재학
- ORCID : https://orcid.org/0009-0001-2251-4112
- 주관심분야 : 컴퓨터 비전, 머신러닝, 데이터사이언스, 인공지능

홍 민 수

- 2019년 2월 : 연세대학교 전자전기공학부 석사
- 2019년 ~ 2021년 : 한국전자기술연구원 연구원
- 2021년 ~ 현재 : KBS 미디어기술연구소 연구원
- ORCID : https://orcid.org/0009-0006-6215-1682
- 주관심분야 : 컴퓨터 비전, 영상처리, 인공지능

저 자 소 개

김 영 길

- 2001년 8월 : 한국과학기술원 전자공학 박사
- 2001년 ~ 2003년 : SK 하이닉스
- 2003년 ~ 현재 : 서울시립대학교 전자전기컴퓨터공학부 교수
- ORCID : https://orcid.org/0000-0001-7066-0555
- 주관심분야 : 이동통신, 신호처리

김 정 래

- 1997년 2월 : 서울대학교 수리과학부 석사
- 2004년 8월 : 서울대학교 수리과학부 박사
- 2010년 ~ 현재 : 서울시립대학교 수학과 교수
- ORCID : https://orcid.org/0000-0002-3261-7238
- 주관심분야 : 수치해석, 시스템생물학, 머신러닝