

열차 내 분실물과 분실자 관리를 위한 Deep OC-SORT 기반 다중객체 추적 및 매칭 알고리즘 구현

장현상*, 민지윤**, 김준서***, 박상민****, 신지혜*****, 김동규*****,
도경민*****, 유길상*****

Implementation of a Deep OC-SORT based Multi-Object Tracking and Matching Algorithm for Managing Lost Items and Owners in Trains

Hyun Sang Jang*, Ji Yun Min**, Jun Seo Kim***, Sang Min Park****, Ji Hye Shin*****,
Dong Gue Kim*****, Kyeong Min Do*****, and Gil Sang Yoo*****

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(RS-2023-00246191)
These authors contributed equally to this work. They share first authorship.

요약

최근 교통수단 및 공공시설에서 유실물 관리의 중요성이 강조되고 있다. 특히 KTX와 같은 대중교통에서는 유실물 발생 빈도가 높아 효율적이고 자동화된 관리 시스템의 필요성이 커지고 있다. 기존 유실물 관리 시스템은 직원의 개입이 필수적이며, 분실자가 스스로 유실 사실을 인지해야만 물품을 찾을 수 있다는 한계가 있다. 본 연구는 이러한 문제를 해결하기 위해 KTX 객차 내에서 실시간으로 유실물을 감지하고 추적할 수 있는 시스템을 제안한다. 이를 위해 Deep OC-SORT 기반의 실시간 다중 객체 추적 및 매칭 알고리즘을 개발하였다. 실험 결과, 검증 데이터에 대한 mAP(mean Average Precision) 값은 mAP50에서 0.995, mAP50-95에서 0.96으로 우수한 성능을 보여주었다. 실제 현장 적용 결과, 분실물과 소유자의 매칭이 정확하게 이루어지는 것으로 확인되었다. 본 시스템은 KTX뿐만 아니라 다양한 교통수단과 공공시설에서 유실물 관리의 효율성을 크게 향상시킬 것으로 기대된다.

Abstract

Recently, the importance of lost and found management in transportation systems and public facilities has been emphasized. In particular, in public transportation like KTX, where the frequency of lost items is high, the need for an efficient and automated management system has increased. Traditional lost and found systems require staff involvement and rely on the owner's awareness of the loss, limiting their effectiveness. This study proposes a real-time system for detecting and tracking lost items in KTX carriages to address these issues. A real-time multi-object tracking and matching algorithm based on Deep OC-SORT was developed to track objects and match them to their owners. Experimental results showed excellent performance with mean Average Precision(mAP) scores of 0.995 for mAP50 and 0.96 for mAP50-95 on the validation dataset. When applied in a real-world environment, the system successfully matched lost items with their owners. This system is expected to significantly improve lost and found management not only in KTX but also in other transportation systems and public facilities.

Keywords

deep OC-SORT, multi-object tracking, matching algorithm, YOLOv8, blind primed supervised

* 국민대학교 경제학과
- ORCID: <https://orcid.org/0009-0006-2152-2758>
** 강원대학교 수학과
- ORCID: <https://orcid.org/0009-0000-5828-8877>
*** 고려대학교 기계공학부
- ORCID: <https://orcid.org/0009-0003-7620-2880>
**** 고려대학교 산업경영공학부
- ORCID: <https://orcid.org/0009-0005-6187-7281>
***** 성신여자대학교 경영학과
- ORCID: <https://orcid.org/0009-0001-1230-957X>

***** 고려대학교 경영학과
- ORCID: <https://orcid.org/0009-0005-5924-7221>
***** 고려대학교 디자인조형학부
- ORCID: <https://orcid.org/0009-0008-5686-5894>
***** 고려대학교 정보대학 정보창의교육연구소(교신저자)
- ORCID: <https://orcid.org/0009-0002-1085-5355>
• Received: Nov. 07, 2024, Revised: Dec. 20, 2024, Accepted: Dec. 23, 2024
• Corresponding Author: Gilsang Yoo
Creative Informatics & Computing Institute, 145 Anam-ro, Seongbuk-gu, Seoul, Korea
Tel.: +82-2-3290-1674, Email: ksyoo@korea.ac.kr

I. 서 론

최근 유실물 관리 문제는 전 세계적으로 공공 교통수단과 시설에서 중요한 이슈로 떠오르고 있다. 특히 대중교통, 공항, 대형 시설 등과 같은 환경에서는 유실물 관리가 큰 과제로 부각되고 있다. 기존의 유실물 처리는 주로 관리자의 수작업에 의존해 왔으나, 이는 많은 인력과 시간이 소모되며 효율성이 떨어진다. 대한민국에서는 전국의 기차역과 열차에서 매년 약 24만 건의 유실물이 습득되고 있으며, 특히 경찰청 유실물 통합 포털 통계에 따르면 KTX(Korea Train eXpress) 서울역에서만 매주 1,200여 건의 유실물이 발견되고 있다. 이처럼 수백만 명의 승객이 이용하는 대중교통에서는 유실물 발생 빈도가 매우 높아 이 문제가 더욱 두드러지고 있다.

유실물 관리기관은 유실물 처리에 필요한 시간과 노동, 그리고 보관 공간 운영에 상당한 비용을 지출하고 있으며, 이를 해결하기 위한 다양한 방법들이 시도되고 있다[1]. 예를 들어, P. Choudhary et al.[2] 연구에서는 분실자와 유실물을 발견한 사람이 소통할 수 있는 스마트폰 애플리케이션을 개발하였다. 이 애플리케이션을 통해 분실자와 발견자가 서로 연락을 주고받으며, 분실물 청구에 필요한 세부 정보를 전달할 수 있다. 분실자는 물건을 잃어버린 위치를 간단히 태그하고, 키워드를 입력해 검색 과정을 보다 빠르고 효과적으로 진행할 수 있지만, 이 시스템은 분실자와 습득자가 앱을 사용해야만 작동하는 단점이 있다.

기존의 유실물 탐지 기술은 대부분 직원의 개입을 필요로 하고, 유실물 발생 후 처리 과정이 복잡하며 많은 시간이 소요된다. 또한, 분실자가 물건을 잃어버렸다는 사실을 인식해야만 처리 절차가 시작되므로 실시간 처리가 어렵다는 한계가 있다. 이러한 문제를 해결하기 위해 본 연구에서는 객체 탐지(Object detection), 추적(Tracking), 매칭(Matching), 이미지 캡셔닝(Image captioning) 기술을 결합한 실시간 유실물 탐지 및 관리 시스템을 제안하였다. 이 시스템은 CCTV 영상을 분석하여 승객과 소지품을 실시간으로 탐지하고, 유실물이 발생할 경우 자동으로 알림을 제공할 수 있다.

본 논문의 구성은 다음과 같다. 제2장에서는 객체 탐지 및 추적 기술을 중심으로 관련 연구를 설명하고, Fast R-CNN(Fast Region-based Convolutional Neural Network), DETR(Detection with Transformers) 모델, ByteTrack, TAM(Track Anything Model), Deep OC-SORT 등의 최신 알고리즘을 다루었다. 제3장에서는 본 연구에서 제안하는 다중 객체 추적 및 매칭 시스템에 대해 YOLOv8을 활용한 객체 탐지, Deep OC-SORT(Observation-Centric Simple Online and Realtime Tracking) 기반의 실시간 추적, 거리 기반 매칭 알고리즘, 그리고 BLIP(Bootstrapping Language-Image Pre-training)를 이용한 이미지 캡셔닝 과정을 설명하였다. 제4장에서는 제안된 시스템의 성능을 평가하고, 학습 및 검증 손실과 정확도 변화를 분석하였다. 마지막으로, 제5장에서는 본 연구의 결론과 향후 연구 과제를 제시하였다.

II. 관련 연구

2.1 Object detection 기술

Fast R-CNN은 이미지 내 객체 탐지와 위치 추정 작업의 효율성을 높이기 위해 설계된 모델이다[3][4]. 이전 모델들이 복잡한 다단계 파이프라인과 높은 계산 비용으로 인해 처리 속도가 느리고 비효율적이었던 문제를 해결하기 위해, Fast R-CNN은 단일 단계 훈련 알고리즘을 도입하여 이러한 문제를 개선하였다[5][6]. 특히, R-CNN과 SPPnet에서 사용되는 다단계 훈련 방식과 달리, Fast R-CNN은 네트워크의 모든 계층을 동시에 훈련할 수 있어 훈련 시간이 대폭 단축되었다. Fast R-CNN은 전체 이미지를 합성곱 신경망으로 처리한 후, 관심 영역(RoI, Region of Interest)을 선택하여 해당 영역의 특징을 추출하고, 이를 기반으로 객체의 클래스와 위치를 예측한다[7]. 이 과정에서 RoI 풀링 레이어를 도입하여, 고정된 크기의 특징 벡터를 추출함으로써 객체 제안 과정의 효율성을 크게 향상시켰다.

DETR(Detection with Transformers)은 set-based global loss와 bipartite matching 알고리즘을 활용하여 기존 객체 탐지 방식과 차별화된 접근을 제공하는 모델이다[8].

기존의 객체 탐지 모델들이 다수의 객체 경계 상자(Bounding box)를 예측하고, 그중에서 최종 예측을 선택하는 방식이었다면, DETR은 이러한 예측 과정을 간소화할 수 있는 장점이 있다[9]. DETR은 ResNet-50을 백본으로 사용하여 입력 이미지에서 특징 맵을 추출하고, Transformer 기반의 인코더-디코더 구조를 채택하여 이미지 내 객체 간의 상호작용을 파악한다. 이 모델은 self-attention 메커니즘을 통해 각 픽셀 간의 연관성을 학습함으로써 이미지 내 객체 간 관계를 보다 깊이 이해할 수 있다. DETR의 주요 장점 중 하나는 객체 탐지 과정에서 발생하는 중복된 예측을 제거할 수 있다는 점이다. 이는 set prediction 방식을 통해 이루어지며, 기존 모델들이 필요로 했던 복잡한 후처리 과정인 비최대 억제(Non-maximum suppression) 없이도 높은 성능을 제공한다. 이러한 방식으로 DETR은 간단하면서도 효율적인 객체 탐지 결과를 제공하여 후처리 과정을 최소화한다.

2.2 Tracking 관련 기술

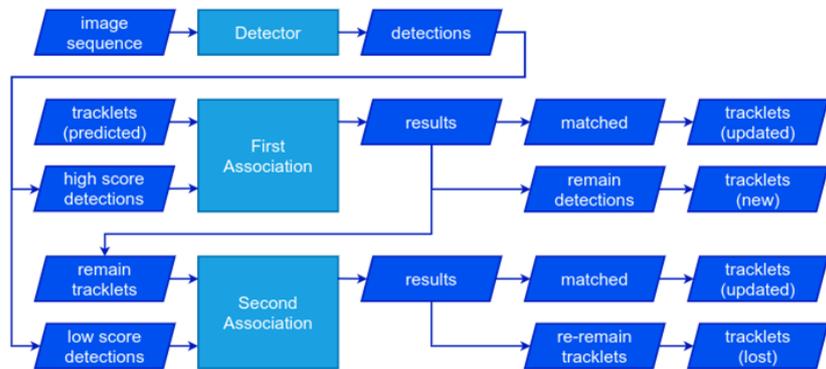
그림 1에서와 같이, ByteTrack Detection기반 모델은 단기 객체 추적에 최적화된 모델로, 주로 짧은 시간 동안의 객체 움직임을 효과적으로 추적하는데 강점을 가지고 있다[10]. 그러나 장기적인 추적이 필요한 상황에서는 몇 가지 한계를 보인다. 장기 추적에서는 객체의 위치와 상태가 시간에 따라 크게 변화하기 때문에, 단기 추적에 최적화된 ByteTrack 모델은 이러한 변화를 효율적으로 처리하

지 못할 가능성이 있다. 특히, 낮은 점수로 탐지된 객체의 경우, 그 신뢰도가 낮아지면 추적의 정확성에 부정적인 영향을 미칠 수 있다[11]. 또한, ByteTrack은 사전 학습된 객체 탐지 모델의 성능에 크게 의존하는 단점을 지니고 있다[12]. 다양한 환경과 조건에서 사전 학습된 모델이 안정적인 성능을 발휘하지 못할 경우, ByteTrack의 추적 성능 역시 저하될 가능성이 크다. 예를 들어, 조도가 낮은 환경이나 복잡한 배경에서는 객체 탐지가 어려워질 수 있으며, 이에 따라 추적이 중단되거나 부정확한 결과를 초래할 수 있다. 따라서 ByteTrack은 특정 조건에서 매우 효과적일 수 있지만, 모든 상황에서 일관된 성능을 보장하기는 어려운 한계가 존재한다.

이러한 한계는 장기적인 영상 분석이 요구되는 응용 분야에서 특히 중요한 고려 사항이 된다. 예를 들어, 공공장소의 감시 카메라 영상이나 교통 상황 모니터링과 같이 객체가 프레임 간에 장시간 나타나지 않을 수 있는 경우에는 ByteTrack의 추적 성능이 저하될 수 있다. 따라서 이러한 응용 분야에서는 ByteTrack 모델보다는 다른 모델이 더 적합할 수 있다.

TAM(Track Anything Model)은 이미지 내 객체를 분할하는 SAM(Segment Anything Model)과 마스크 기반 추적 모델인 XMem을 결합한 모델이다 [13][14]. 그림 2에서와 같이, 이 모델의 주요 장점은 객체가 프레임 밖으로 나갔다가 다시 들어와도 동일한 ID를 유지할 수 있다는 점이다. 그러나 이 모델은 객체 선택과 같이 사용자의 수동 초기화로 작동할 수 있다는 한계가 있으며, 자동 추적 기능을 구현하기 위해서는 추가적인 방법이 필요하다.

ByteTrack: Multi-Object Tracking by Associating Every Detection Box



© 2023 Luffca Inc.

그림 1. ByteTrack 다중 객체 추적 과정
Fig. 1. ByteTrack multi-object tracking process

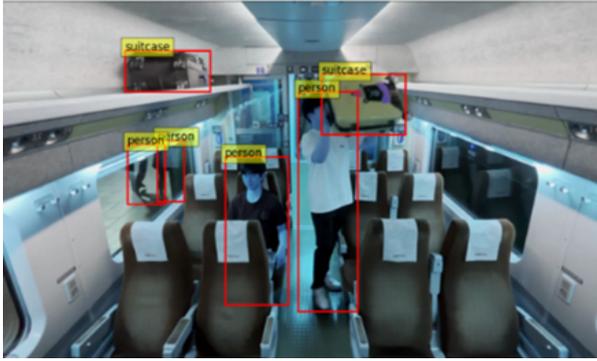


그림 2. autoTAM을 이용한 객체 탐지 결과
Fig. 2. Object detection results using autoTAM

본 연구에서는 TAM의 수동적 한계를 극복하기 위해 DETR(Detection Transformer)을 결합하여 자동 추적을 시도하였고 그 결과는 그림 3과 같다. 그림에도 불구하고, DETR은 객체 탐지와 분류에서 성공적인 성능을 보였으나, 추적 과정에서 ID 유지가 잘 이루어지지 않는 문제가 발생하였다. 또한, SAM과 XMem 기반의 TAM은 모델 크기가 커서 영상을 처리하는 데 상당한 시간이 소요된다는 단점이 있다.

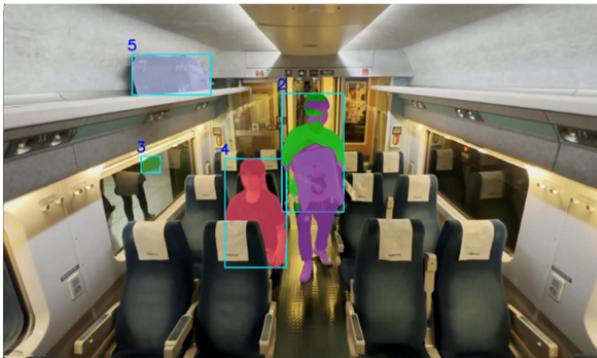


그림 3. DETR과 autoTAM을 이용한 객체 추적
Fig. 3. Object tracking using DETR and autoTAM

OC-SORT는 기존의 SORT(Simple Online and Realtime Tracking) 알고리즘을 기반으로 하며, 칼만 필터(Kalman Filter)를 사용하여 객체의 선형적인 움직임을 가정하고 추적을 수행한다. 하지만 SORT는 객체가 일시적으로 가려지거나 사라질 때 추적 성능이 저하되는 문제를 가지고 있다. 이를 개선한 Deep OC-SORT는 세 가지 주요 모듈인 카메라 움직임 보정(CMC, Camera Motion Compensation), 동적 외형(DA, Dynamic Appearance), 적응 가중치(AW, Adaptive Weighting)를 도입하였다. Deep OC-SORT

는 순수한 모션 기반 추적 알고리즘인 OC-SORT에 시각적 외형 정보를 추가하여 다중 객체 추적(MOT, Multi-Object Tracking) 성능을 크게 향상시켰다. 특히 2023년 MOT20 데이터셋에서 높은 성능을 기록한 모델로 평가받고 있다[15].

III. 제안한 다중객체 추적 및 매칭 모델

제안한 시스템은 KTX 객차 내 CCTV 데이터를 활용하여 실시간으로 유실물을 탐지하고, 승객이 가방을 두고 내리는 상황을 자동으로 인식하여 알림을 제공한다. 전체 시스템 구성은 그림 4와 같다.

먼저, 전처리된 영상은 YOLOv8을 사용하여 프레임 단위로 분석되며, Object Detection 단계를 통해 사람과 가방을 실시간으로 탐지한다. 이후, Deep OC-SORT 알고리즘을 활용하여 탐지된 사람과 가방을 지속적으로 추적하는 Tracking 단계를 수행한다. 이 알고리즘은 객체의 ID를 유지하면서 실시간으로 높은 추적 성능을 제공하며, 승객이 가방을 두고 떠날 때 이를 자동으로 인식한다.

다음으로, Matching 단계에서는 유클리드 거리 기반 매칭 알고리즘을 통해 실시간으로 승객과 가방의 소유 관계를 인식한다. 승객이 가방을 두고 화면 밖으로 벗어나면 해당 가방이 유실된 것으로 판단하여 즉시 알림을 제공한다. 마지막으로, BLIP 모델을 사용한 Image Captioning 단계를 통해 분실된 가방의 색상, 형태 등의 정보를 자동으로 생성하여 이를 승객에게 알림으로 제공한다.

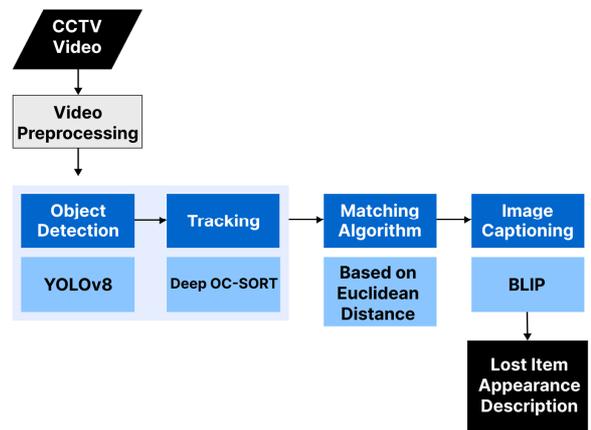


그림 4. 전체 시스템 구조도
Fig. 4. Overall system architecture

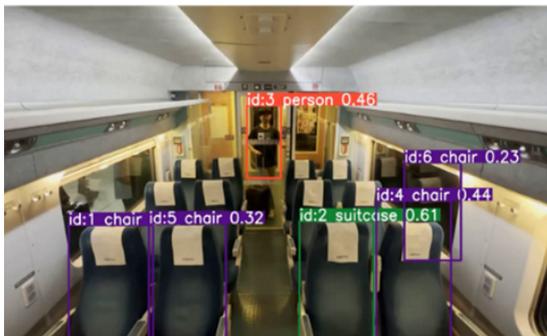
3.1 영상 전처리

본 연구에서는 COCO(Common Objects in Context) 데이터셋에서 가방이 포함된 2,737장의 이미지를 추출하고, KTX 내부에서 촬영한 30개의 영상을 활용하여 총 4,480장의 이미지를 생성하였다. 생성된 데이터셋은 수작업으로 라벨링하여 학습에 사용하였다.

3.2 다중객체 탐지

Object Detection 단계에서는 YOLOv8 모델을 사용하여 사람과 가방을 탐지한다[16][17]. YOLOv8은 약 3.2백만 개의 파라미터를 가진 경량화된 모델로 기본적인 탐지 성능은 우수하지만, 특정 작업에 대해 성능을 높이기 위해서는 추가적인 학습이 필요하다[18]. 이를 위해 전이 학습(Transfer learning)을 적용하였다. YOLOv8 모델의 백본은 약 1400만 장의 이미지로 학습된 ImageNet 가중치를 고정하고, 객체를 분류하고 위치를 예측하는 헤드 부분만 전이 학습의 조정을 통해 가방과 사람의 탐지 성능을 높였다. 과적합을 방지하기 위해 AdamW 옵티마이저를 사용하였으며, 배치 크기(batch size)는 8, 에포크(epoch)는 20, 학습률(lr, learning rate)은 0.01667의 파라미터를 적용하여 모델의 학습을 진행하였다. 이러한 과정을 통해, 본 시스템은 가방과 승객을 정확하게 탐지하고 실시간으로 유실물을 탐지할 수 있는 성능을 확보하였다.

그림 5와 같이, 모델 학습 후 테스트 영상을 통해 성능을 검증한 결과, 학습 이전에는 객체 ID가 최대 22까지 증가하며 동일 객체를 여러 번 탐지했으나, 학습 후에는 객체를 일관되게 추적할 수 있었다.



(a) 전이학습 이전
(a) Before transfer learning



(b) 전이학습 이후
(b) After transfer learning

그림 5. YOLOv8 전이학습 전후 비교
Fig. 5. Comparison of YOLOv8 before and after transfer learning

이는 모델이 학습을 통해 동일한 객체를 정확히 인식하고 추적할 수 있게 되었음을 보여준다.

3.3 객체 트래킹

사람과 가방을 정확하게 매칭하기 위해서는 객체의 ID가 지속적으로 유지되는 것이 중요하다. ID 유지가 불안정하면 사람과 가방 간의 관계를 정확하게 추적할 수 없어 매칭의 신뢰도가 떨어진다. Deep OC-SORT는 다중 객체 추적(MOT, Multi-Object Tracking) 알고리즘으로, 여러 객체를 동시에 추적하고 식별하는 데 뛰어난 성능을 보여준다. 이 알고리즘은 CCTV 모니터링, 스마트 시티 관리, 교통 시스템 분석 등 다양한 분야에서 객체를 인식하고 그 움직임을 분석하는 데 널리 활용되고 있다.

Deep OC-SORT에는 카메라 움직임 보정(CMC), 탐지 신뢰도에 따라 객체 외형 정보 반영 정도를 조절(DA), 그리고 외관 특징과 탐지 사이의 유사도에 따른 가중치 부여(AW) 등의 알고리즘이 추가되어 ID 유지 성능을 극대화하였다. CMC 모듈은 카메라의 움직임이 있는 장면에서 프레임 간 객체의 위치를 보다 정확하게 보정하기 위해 설계되었다. 식 (1)에서와 같이, 이 모듈은 칼만 필터의 상태 $x = [x_c, y_c, a, s, \dot{x}_c, \dot{y}_c, \dot{a}]$ 를 M_t 와 T_t 의 카메라 움직임을 반영하는 변환 행렬(Scaled rotation matrix)과 이동 벡터(Translation vector)를 적용한다.

$$\begin{cases} x_{new}[0:2] \leftarrow M_t x[0:2] + T_t, \\ x_{new}[4:6] \leftarrow M_t x[4:6], \\ P_{new}[0:2,0:2] \leftarrow M_t P[0:2,0:2] M_t^T \\ P_{new}[4:6,4:6] \leftarrow M_t P[4:6,4:6] M_t^T \end{cases} \quad (1)$$

DA 모듈은 탐지 신뢰도(detector confidence)에 따라 β 를 β_t 로 유동적으로 조정함으로써, 객체 외형 정보의 가중치를 반영한다. 식 (2)의 지수 이동 평균(EMA, Exponential Moving Average)을 사용하여 식(3)과 같이 이전 외형 정보와 새로운 외형 정보를 가중합 한다. β_t 는 탐지 신뢰도 $conf_{det}$ 에 기반해 동적으로 변화한다.

$$e_t = \beta e_{t-1} + (1 - \beta) e^{new} \quad (2)$$

$$\beta_t = \beta_f + (1 - \beta_f) \left(1 - \frac{conf_{det} - \sigma}{1 - \sigma}\right) \quad (3)$$

AW 모듈은 트랙과 탐지 간의 코사인 유사도로 계산된 유사도 행렬 A_c 을 기반으로, 트랙과 탐지 간의 변별력을 측정한다. 이 변별력에 따라 트랙-탐지 쌍의 결과가 잘 맞으면 더 높은 가중치를 부여한다. 이후, 가중치를 이용하여 최종 비용 행렬 $Cost$ 를 업데이트한다. 비용 행렬은 객체와 탐지된 박스 간의 매칭 비용을 나타내며 낮은 비용일수록 더 좋은 매칭을 의미한다. 따라서, AW 모듈은 외형 정보가 더 잘 반영되도록 하고, 트랙과 탐지의 매칭이 더욱 정확해진다. 식 (4)에서, $IoU[m,n]$ 은 트랙과 탐지 간의 겹침 정도에 대한 비용 행렬이고, $weight_b(m,n)$ 은 AW 모듈에 의해 조정된 가중치를 의미한다.

$$Cost[m,n] = IoU[m,n] + [a_w + weight_b(m,n)] A_c[m,n] \quad (4)$$

3.4 다중객체 매칭

매칭 알고리즘은 사람과 가방 간의 효율적인 매칭을 위해 유클리드 거리 기반의 가장 가까운 객체 선택 방식을 사용한다. 그림 6는 매칭 알고리즘의 동작 과정이다. 먼저, 탐지된 사람과 가방의 바운딩 박스 중심 좌표를 계산하고, 이를 바탕으로 객체 간

의 거리를 산출한다. 이 과정에서는 프레임 내 모든 사람과 가방의 중심 좌표가 계산되며, 각 사람에 대해 모든 가방과의 유클리드 거리가 계산된다.

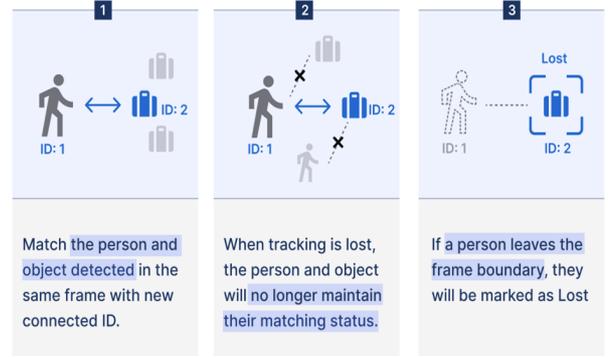


그림 6. Matching 알고리즘 동작 과정
Fig. 6. Matching algorithm process

매칭 과정에서는 각 사람에게 가장 가까운 가방을 매칭하는 것을 목표로 한다. 이를 위해, 각 사람에 대해 계산된 유클리드 거리를 기반으로 가장 가까운 가방을 선택하며, 이때 이미 다른 사람과 매칭된 가방은 중복으로 매칭되지 않도록 처리한다. 이는 한 번 형성된 사람-가방 쌍이 이후 프레임에서도 지속적으로 유지되도록 하는 중요한 단계이다. 매칭된 쌍은 연속된 프레임에서도 추적되며, 사람과 가방 간의 관계가 유지된다.

핵심 매칭 알고리즘은 식 (5)와 같이 사람과 가방 객체의 중심 좌표를 기반으로 유클리드 거리 $d(i,j)$ 를 계산하여 가장 가까운 객체를 선택한다.

$$d(i,j) = \sqrt{((x_{p_i} - x_{bag_j})^2 + (y_{p_i} - y_{bag_j})^2)} \quad (5)$$

식 (5)에서 (x_{p_i}, y_{p_i}) 는 사람 객체 i 의 중심 좌표, (x_{bag_j}, y_{bag_j}) 는 가방 객체 j 의 중심 좌표이다. 각 사람 객체 i 는 다음과 같이 가장 가까운 가방 객체 j 와 매칭 된다.

$$j = \arg \min_{j \in B, j \notin matched_bags} d(i,j) \quad (6)$$

식 (6)에서 B 는 모든 가방 객체의 집합, 이미 매칭된 가방의 집합은 $matched_bags$ 를 의미한다.

이 매칭 알고리즘은 단순히 현재 프레임에서만 동작하는 것이 아니라 이후 프레임에서도 지속적으로 유지되도록 설계되었다. 예를 들어, 사람이 프레임 밖으로 나가거나 시야에서 사라질 경우, 해당 가방은 소유자와의 매칭이 끊어져 유실된 것으로 처리된다. 이는 유실물 관리 시스템의 핵심 기능으로, 실시간으로 가방과 소유자가 분리되는 상황을 감지하고 이를 알림으로 제공함으로써 유실물을 방지할 수 있도록 한다.

3.5 Image captioning

Image Captioning 단계에서는 그림 7과 같이, 분실로 판단된 가방의 바운딩 박스 이미지를 입력받아 이미지 캡셔닝 모델인 BLIP에 적용한다. BLIP 모델은 이미지의 시각적 내용을 설명하는 텍스트를 생성하는 데 사용되며, 다양한 이미지 캡션 벤치마크에서 뛰어난 성능을 입증하였다. 이 모델은 특히 복잡한 장면보다는 개별 객체의 이미지를 설명하는 데 최적화되어 있어, 가방과 같은 개별 물체의 시각적 특성을 효과적으로 묘사한다[19].

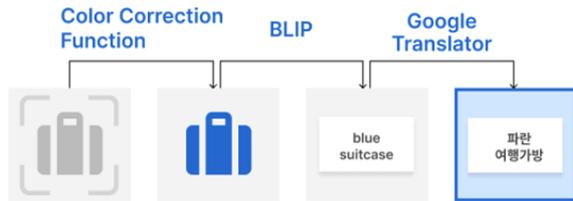


그림 7. BLIP 기반 이미지 캡셔닝 과정
Fig. 7. BLIP-based image captioning process

BLIP 모델이 생성한 캡션에는 가방의 외형적 특징, 모양, 및 기타 시각적 정보가 포함되며, 이러한 정보는 유실물의 정확한 식별과 회수에 중요한 역할을 한다. 예를 들어, 가방의 색상, 크기, 모양 등과 같은 세부 정보가 캡션에 포함되어, 유실물 식별 과정을 보다 용이하게 한다.

추가적으로, 가방의 색상을 더욱 명확하게 표현하기 위해 색 보정 함수를 사용하여 이미지의 색상을 분석하고, 이를 통해 가방의 주요 색상을 추출한다. 이렇게 추출된 색상 정보는 캡션에 포함되어 시각적 설명이 더욱 정확하고 직관적으로 제공된다. 마지막으로, BLIP 모델이 생성한 캡션과 색상 정보

는 한글로 번역되어 사용자가 쉽게 이해할 수 있도록 지원된다. 이 과정은 유실물 관리 시스템이 다양한 언어 환경에서 효과적으로 작동할 수 있도록 보장하며, 승객이 자신의 유실물을 신속하게 인식하고 회수할 수 있도록 돕는다.

결과적으로, BLIP 모델을 활용한 이미지 캡셔닝 단계는 유실물의 시각적 특성을 정확히 설명하여 유실물 관리 시스템의 효율성 및 사용자 편의성을 크게 향상시키는 중요한 역할을 한다.

IV. 실험 결과

4.1 모델 학습 및 평가

KTX 내부에서 사람과 가방의 탐지 정확도를 향상시키기 위해 실제 KTX 환경에서 촬영된 영상을 학습 데이터로 사용하였다. 이 데이터는 다양한 실제 상황을 반영하고 있어, 모델이 보다 정교하고 현실적인 조건에서 학습될 수 있도록 한다. 학습 과정에서 바운딩 박스 위치 예측 및 객체 분류에 대한 손실 값이 점진적으로 감소하였으며, 이는 모델이 점차적으로 사람과 가방을 더 정확하게 탐지하고 분류할 수 있게 되었음을 의미한다.

본 연구에서는 YOLOv8 모델을 사용하여 사람과 가방 탐지를 위한 학습을 진행하였으며, 그 결과는 다음과 같다.

그림 8과 같이 Epoch이 진행될수록 학습 손실은 꾸준히 감소하였다.

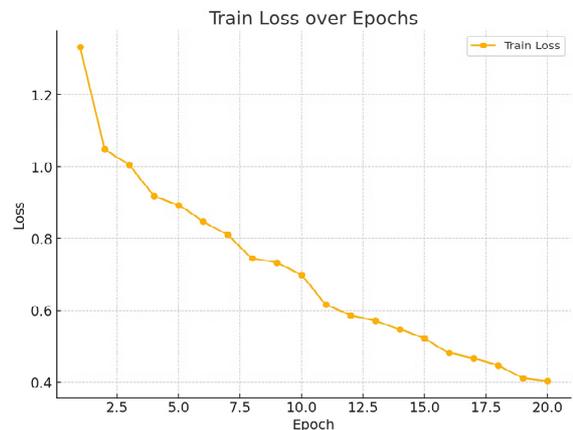


그림 8. 학습 손실 값의 에폭(Epoch)별 변화
Fig. 8. Train loss over epochs

초기 손실 값은 1.2에서 시작하여 Epoch 20에서 약 0.4까지 감소하였는데, 이는 모델이 훈련 데이터에 대한 성능을 점차적으로 개선하고 있음을 나타낸다.

그림 9와 같이, 검증 손실도(Validation loss)는 Epoch이 증가함에 따라 감소하는 경향을 보였다. Epoch 1에서 검증 손실 값은 1.349였으나 Epoch 10에서는 0.4183으로 감소하였다.

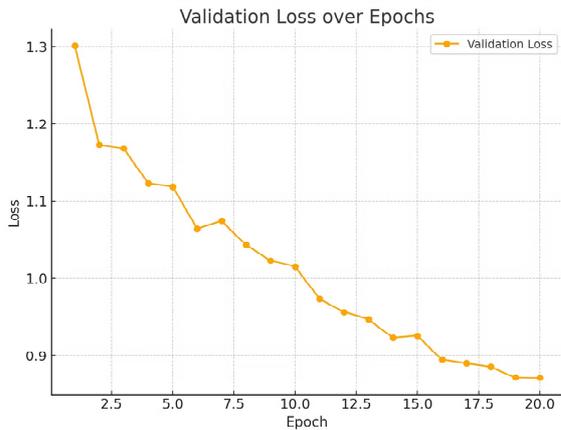


그림 9. 검증 손실 값의 에폭(Epoch)별 변화
Fig. 9. Validation loss over epochs

그림 10과 같이, 학습 정확도는 Epoch이 증가함에 따라 뚜렷하게 향상되었다. 초기 정확도는 0.656에서 시작하여 Epoch 10에 0.854에 도달하였고, 이후로도 지속적인 향상을 보였다. 이는 모델이 학습 데이터에 대해 점차 더 높은 예측 성능을 달성하고 있음을 시사한다.

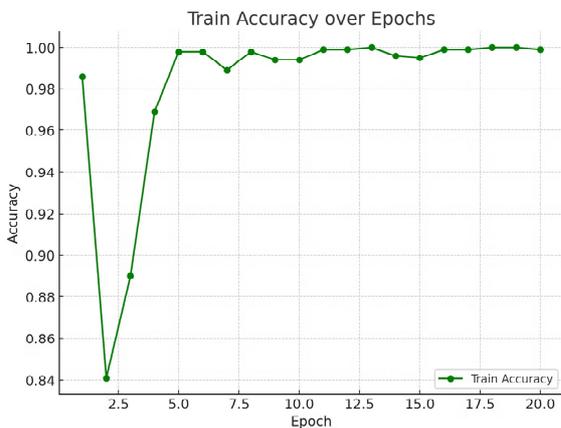


그림 10. 학습 정확도의 에폭(Epoch)별 변화
Fig. 10. Train accuracy over epochs

그림 11과 같이, 검증 데이터셋에 대한 정확도 또한 Epoch이 진행됨에 따라 안정적인 성능을 유지했다. Epoch 1에서 검증 정확도는 0.995로 매우 높은 값에서 시작했으며, Epoch 10에서 0.994로 비슷한 수준을 유지하고 있다. 이는 모델이 검증 데이터에 대해 매우 높은 성능을 보여주고 있으며 과적합이 크게 발생하지 않았음을 의미한다.

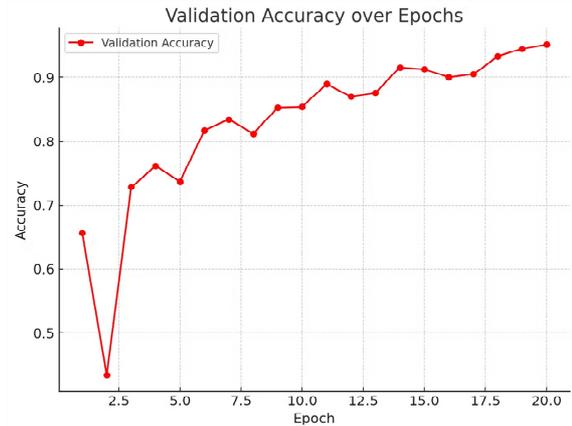


그림 11. 검증 정확도의 에폭(Epoch)별 변화
Fig. 11. Validation accuracy over epochs

4.2 모델 시뮬레이션

제안한 모델을 실제 현장에 적용한 결과는 그림 12와 같다.

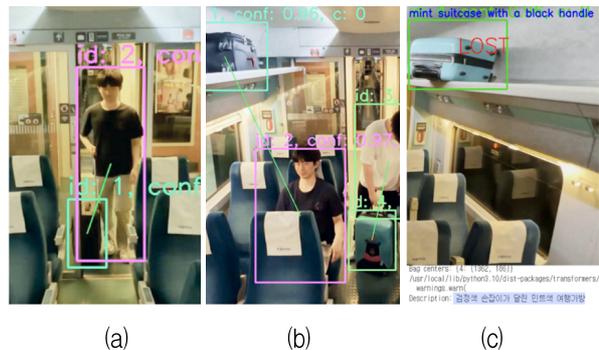


그림 12. 실험 결과
(a) 다중 객체 검출 및 추적 (b) 객체 매칭 (c) 분실 판정
Fig. 12. Experimental results (a) Multi-object detection and tracking (b) Object matching (c) Loss determination

그림 12의 (a)는 다중 객체 검출 및 추적 단계로, 승객이 출입문을 통해 들어오는 순간 사람과 물건을 동시에 탐지한다. 그림 12의 (b)는 객체 매칭 단

계로, 매칭 과정에서는 사람과 가방이 동시에 추적되며, 유클리드 거리를 기반으로 사람과 가장 가까운 가방을 선택하여 두 객체를 선으로 연결된다. 이를 통해 사람과 가방 간의 관계가 시각적으로 명확히 드러난다. 또한, BLIP 모델과 색 보정 함수를 사용하여 가방의 색상을 더 명확하게 표현하였고, 생성된 설명을 한국어로 번역하여 사용자에게 제공된다. 이 과정은 그림 12의 (c) 분실 판정 단계에서 영상에 직접 출력되어 유실물이 발생했을 때 빠르게 식별할 수 있도록 돕는다.

사람과 가방을 인식하도록 튜닝된 모델은 사람이 객실에 들어와 가방을 소지할 때, 이 둘을 정확히 탐지하고 이후에도 지속적으로 추적할 수 있다. 사람이 가방과 함께 있는 동안에는 두 객체가 가까운 거리에 있을 때 선으로 연결되어 매칭이 유지되며, 이후 프레임에서도 일관되게 추적 상태를 관리한다.

'LOST'라는 문구를 화면에 표시한다. 이와 동시에 가방의 바운딩 박스 이미지를 분석하여 '검정색 손잡이가 달린 민트색 여행가방과 같이 상세한 설명을 생성하고, 이를 영상에 출력한다. 이 기능은 유실물 관리 시스템에서 중요한 역할을 하며, 승객이 자신의 유실물을 쉽게 식별하고 회수할 수 있도록 지원한다.

V. 결론 및 향후 과제

본 논문에서는 KTX 객차 내에서 실시간으로 유실물을 감지하고 자동으로 알림을 제공하는 시스템을 제안하였다. 기존의 수작업 중심의 유실물 관리 방식과 달리, 본 시스템은 CCTV 영상을 활용하여 가방과 소유자 간의 관계를 인식하고, 분실 여부를 실시간으로 판단하여 승객이 하차하기 전에 유실물을 인지할 수 있도록 설계되었다.

본 연구에서 제안한 시스템은 다음과 같은 주요 성과를 달성하였다. 첫째, CCTV 영상 분석을 통해 사람과 가방을 정확하게 인식하고 추적함으로써, 가방을 두고 떠난 상황을 실시간으로 감지하였다. 둘째, Deep OC-SORT 기반의 다중 객체 추적 알고리즘을 통해 사람과 가방의 ID를 지속적으로 관리하며, 유클리드 거리 기반 매칭 알고리즘을 사용하여 사람과 가방의 관계를 정확하게 파악하였다. 셋째, 가방이 분실되었으면 자동으로 캡션을 생성하여 이를 승객에게 알림으로 제공함으로써 유실물 관리 과정을 자동화하였다. 이러한 시스템은 유실물 관리에 드는 행정적 부담과 비용을 줄이고, 인력의 효율적 배분을 가능하게 하였다.

향후 연구과제로는 현재 KTX에 국한된 시스템을 공항, 지하철, 대형 시설 등 다양한 교통수단과 공공장소로 확장하여 유실물 관리의 통합적 접근을 모색하고, 저조도 환경 및 혼잡한 상황에서도 높은 정확도를 유지할 수 있도록 모델의 성능 개선이 진행될 예정이다.

결론적으로, 본 논문에서 제안한 시스템은 기존의 수동적 유실물 관리 방식에서 벗어나 기술적 접근을 통해 유실물 관리의 효율성을 극대화할 수 있는 중요한 도구로 활용될 수 있다. CCTV 영상을

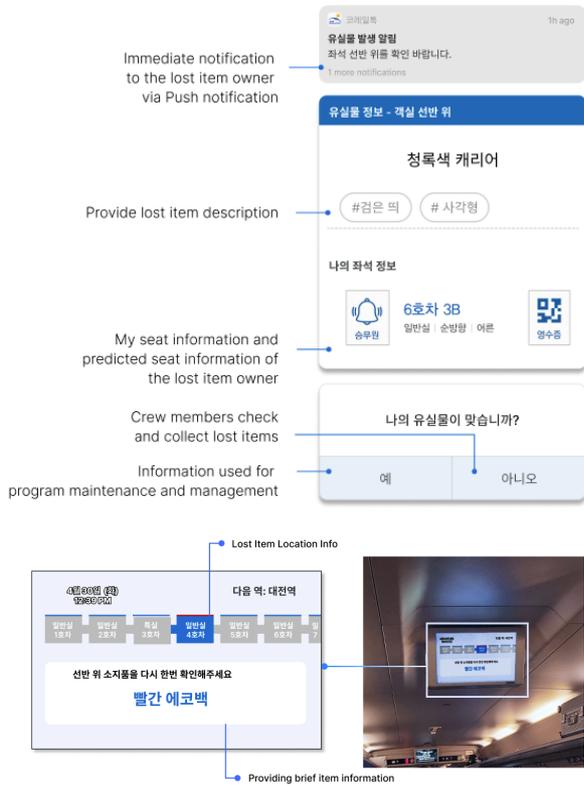


그림 13. 유실물 알림 및 관리 시스템 작동 과정
Fig. 13. Lost item notification and management system process

그림 13은 유실물 알림 및 관리 시스템 작동 과정이다. 만약 사람이 가방을 두고 화면을 벗어나면 시스템은 이를 가방이 분실된 것으로 인식하여

기반으로 승객과 그들의 소지품을 실시간으로 감지하고, 가방 등 유실물이 발생할 경우 이를 즉시 인식하여 자동 알림을 제공함으로써, 유실물 처리 과정의 신속성과 정확성을 크게 향상시킬 수 있다. 이로써 기존의 인력 및 행정력 소모를 최소화하고 관련 비용을 절감하는 효과를 기대할 수 있으며, 향후 다양한 교통수단과 공공시설로의 확대 적용을 통해 유실물 관리 시스템의 혁신을 이룰 수 있을 것으로 기대된다.

References

- [1] Y. Liu, K. Fang, E. Yang, S. Zhao, S. Liu, and R. He, "Lost-Found Item Net for Classification Based on Inception-Resnet", 2022 16th IEEE International Conference on Signal Processing (ICSP), Beijing, China, pp. 335-338, Oct. 2022. <https://doi.org/10.1109/ICSP56322.2022.9965294>.
- [2] P. Choudhary, A. K. Choudhary, A. P. Srivastava, and A. Singh, "Find Mine: Find the Lost Items via Mobile App", 2021 2nd International Conference on Intelligent Engineering and Management (ICIEM), London, United Kingdom, pp. 491-495, Apr. 2021. <https://doi.org/10.1109/ICIEM51511.2021.9445379>.
- [3] X. Liu, H. Wang, H. Jing, A. Shao, and L. Wang, "Research on Intelligent Identification of Rock Types Based on Faster R-CNN Method", IEEE Access, Vol. 8, pp. 21804-21812, Jan. 2020. <https://doi.org/10.1109/ACCESS.2020.2968515>.
- [4] Y. Li, S. Zhang, and W.-Q. Wang, "A Lightweight Faster R-CNN for Ship Detection in SAR Images", IEEE Geoscience and Remote Sensing Letters, Vol. 19, pp. 1-5, Dec. 2022. <https://doi.org/10.1109/LGRS.2020.3038901>.
- [5] X. Chen, et al., "Fast and Accurate Craniomaxillofacial Landmark Detection via 3D Faster R-CNN", IEEE Transactions on Medical Imaging, Vol. 40, No. 12, pp. 3867-3878, Dec. 2021. <https://doi.org/10.1109/TMI.2021.3099509>.
- [6] F. Fang, L. Li, H. Zhu, and J. -H. Lim, "Combining Faster R-CNN and Model-Driven Clustering for Elongated Object Detection", IEEE Transactions on Image Processing, Vol. 29, pp. 2052-2065, Oct. 2020. <https://doi.org/10.1109/TIP.2019.2947792>.
- [7] H. Cui, et al., "An Improved Combination of Faster R-CNN and U-Net Network for Accurate Multi-Modality Whole Heart Segmentation", IEEE Journal of Biomedical and Health Informatics, Vol. 27, No. 7, pp. 3408-3419, Jul. 2023. <https://doi.org/10.1109/JBHI.2023.3266228>.
- [8] Y.-M. Kim, Y.-H. Yoo, I.-U. Yoon, H. Myung, and J.-H. Kim, "Spatio-Temporal Deformable DETR for Weakly Supervised Defect Localization", IEEE Sensors Journal, Vol. 23, No. 17, pp. 19935-19945, Sep. 2023. <https://doi.org/10.1109/JSEN.2023.3298777>.
- [9] Y. Wang and J.-E. Ha, "Improved DETR With Class Tokens in an Encoder", IEEE Access, Vol. 12, pp. 129498-129510, Sep. 2024. <https://doi.org/10.1109/ACCESS.2024.3458461>.
- [10] N. N. Myo, A. Boonkong, K. Khampitak, and D. Hormdee, "Real-Time Surgical Instrument Segmentation Analysis Using YOLOv8 With ByteTrack for Laparoscopic Surgery", IEEE Access, Vol. 12, pp. 83091-83103, Jun. 2024. <https://doi.org/10.1109/ACCESS.2024.3412780>.
- [11] Y. Wang and V. Y. Mariano, "A Multi Object Tracking Framework Based on YOLOv8s and ByteTrack Algorithm", IEEE Access, Vol. 12, pp. 120711-120719, Aug. 2024. <https://doi.org/10.1109/ACCESS.2024.3450370>.
- [12] Y. G. Bihanda, C. Fatichah, and A. Yuniarti, "Multi-Vehicle Tracking and Counting Framework in Average Daily Traffic Survey Using RT-DETR and ByteTrack", IEEE Access, Vol. 12, pp. 121723-121737, Sep. 2024. <https://doi.org/10.1109/ACCESS.2024.3453249>.

- [13] J. Yang, M. Gao, Z. Li, S. Gao, F. Wang, and F. Zheng, "Track Anything: Segment Anything Meets Videos", arXiv preprint, arXiv:2304.11968, Apr. 2023. <https://doi.org/10.48550/arXiv.2304.11968>.
- [14] S. M. Noe, T. T. Zin, P. Tin, and I. Kobayashi, "Efficient Segment-Anything Model for Automatic Mask Region Extraction in Livestock Monitoring", 2023 IEEE 13th International Conference on Consumer Electronics - Berlin (ICCE-Berlin), Berlin, Germany, pp. 167-171, Sep. 2023. <https://doi.org/10.1109/ICCE-Berlin58801.2023.10375624>.
- [15] G. Maggolino, A. Ahmad, J. Cao, and K. Kitani, "Deep OC-Sort: Multi-Pedestrian Tracking by Adaptive Re-Identification", 2023 IEEE International Conference on Image Processing, Kuala Lumpur, Malaysia, pp. 3025-3029, Oct. 2023. <https://doi.org/10.1109/ICIP49359.2023.10222576>.
- [16] X. Liu, Y. Wang, D. Yu, and Z. Yuan, "YOLOv8-FDD: A Real-Time Vehicle Detection Method Based on Improved YOLOv8", IEEE Access, Vol. 12, pp. 136280-136296, Sep. 2024. <https://doi.org/10.1109/ACCESS.2024.3453298>.
- [17] J.-S. Kim and S.-J. Lee, "Improvement of Object Detection Performance in Satellite Images using Image Segmentation and Up-scaling", The Journal of Korean Institute of Information Technology, Vol. 22, No. 6, pp. 21-29, Jun. 2024. <https://doi.org/10.14801/jkiit.2024.22.6.21>
- [18] C. [You and H. Kong, "Improved Steel Surface Defect Detection Algorithm Based on YOLOv8", IEEE Access, Vol. 12, pp. 99570-99577, Jul. 2024. <https://doi.org/10.1109/ACCESS.2024.3429555>.
- [19] J. Li, D. Li, C. Xiong, & S. Hoi, "Blip: Bootstrapping language-image pre-training for unified vision-language understanding and generation," in International conference on machine learning, PMLR, pp. 12888-12900, Feb. 2022. <https://doi.org/10.48550/arXiv.2201.12086>

저자소개

장 현 상 (Hyun Sang Jang)



2020년 2월 : 국민대학교
경제학과(학사)
2024년 6월 : 고려대학교 지능정보
SW아카데미 4기 수료(640H)
관심분야 : 딥러닝, 경제학,
금융공학

민 지 윤 (Ji Yun Min)



2024년 2월 : 강원대학교
수학과(학사)
2024년 6월 : 고려대학교 지능정보
SW아카데미 4기 수료(640H)
관심분야 : 비전 LLM

김 준 서 (Jun Seo Kim)



2024년 2월 : 고려대학교
기계공학부(학사)
2024년 6월 : 고려대학교 지능정보
SW아카데미 4기 수료(640H)
관심분야 : 데이터분석, 딥러닝,
구조해석

박 상 민 (Sang Min Park)



2024년 2월 : 고려대학교
산업경영공학부 (학사)
2024년 6월 : 고려대학교 지능정보
SW아카데미 4기 수료(640H)
관심분야 : 데이터분석, 딥러닝,
컴퓨터비전

신 지 혜 (Ji Hye Shin)



2020년 3월 ~ 현재 :
성신여자대학교
경영학과(주전공),
통계학과(복수전공) 학사과정
2024년 6월 : 고려대학교 지능정보
SW아카데미 4기 수료(640H)
관심분야 : 딥러닝, 컴퓨터비전,
데이터사이언스, 데이터분석

김 동 규 (Dong Gue Kim)



2024년 2월 : 고려대학교
경영학과(학사)
2024년 6월 : 고려대학교 지능정보
SW아카데미 4기 수료(640H)
관심분야 : 딥러닝, 컴퓨터비전, 웹
개발

도 경 민 (Kyeong Min Do)



2024년 2월 : 고려대학교
디자인조형학부(제1전공),
컴퓨터학과(제2전공)(학사)
2024년 6월 : 고려대학교 지능정보
SW아카데미 4기 수료(640H)
관심분야 : 딥러닝, 사용자 경험,
UI/UX

유 길 상 (Gil Sang Yoo)



2010년 2월 : 중앙대학교
영상공학과(박사)
2010년 3월 ~ 현재 :
(사)한국컴퓨터게임학회 부회장
2011년 3월 ~ 현재 : 고려대학교
정보대학 정보창의교육연구소/
지능정보 SW아카데미 교수

2023년 3월 ~ 현재 : (사)한국미디어 아트산업협회
수석부회장

관심분야 : 데이터사이언스, 데이터 시각화, 빅데이터
분석, 3D영상 콘텐츠, 머신러닝, 딥러닝, 컴퓨터교육