

# 실내 문화시설 안전을 위한 딥러닝 기반 방문객 검출 및 동선 추적에 관한 연구

<sup>1</sup>신원섭, <sup>2\*</sup>노승민

## Deep Learning-based Approach for Visitor Detection and Path Tracking to Enhance Safety in Indoor Cultural Facilities

<sup>1</sup>Wonseop Shin, <sup>2\*</sup>Seungmin Rho

### 요약

포스트-코로나 시대에는 방역 조치의 중요성이 크게 강조되고 있으며, 이에 맞춰 딥러닝을 이용한 마스크 착용 상태 검출 및 다른 전염병 예방에 관련된 연구가 진행되고 있다. 그러나 질병 확산 방지를 위한 문화시설 관람객 탐지 및 추적 연구도 마찬가지로 중요하므로 이에 대한 연구가 진행되어야 한다. 본 논문에서는 사전 수집된 데이터 셋을 이용하여 컨볼루션 신경망 기반 객체 탐지 모델을 전이 학습시키고, 학습된 탐지 모델의 가중치를 다중 객체 추적 모델에 적용하여 방문객을 모니터링 한다. 방문객 탐지 모델은 Precision 96.3%, Recall 85.2% F1-Score 90.4%의 결과를 보여주었다. 추적 모델의 정량적 결과로 MOTA 65.6%, IDF1 68.3%, HOTA 57.2%의 결과를 보여주었으며, 본 논문의 모델과 다른 다중 객체 추적 모델 간의 정성적 비교에서 우수한 결과를 보여주었다. 본 논문의 연구는 포스트-코로나 시대의 문화시설 내 방역 시스템에 적용될 수 있을 것이다.

### Abstract

In the post-COVID era, the importance of quarantine measures is greatly emphasized, and accordingly, research related to the detection of mask wearing conditions and prevention of other infectious diseases using deep learning is being conducted. However, research on the detection and tracking of visitors to cultural facilities to prevent the spread of diseases is equally important, so research on this should be conducted. In this paper, a convolutional neural network-based object detection model is trained through transfer learning using a pre-collected dataset. The weights of the trained detection model are then applied to a multi-object tracking model to monitor visitors. The visitor detection model demonstrates results with a precision of 96.3%, recall of 85.2%, and an F1-score of 90.4%. Quantitative results of the tracking model include a MOTA (Multiple Object Tracking Accuracy) of 65.6%, IDF1 (ID F1 Score) of 68.3%, and HOTA (Higher Order Tracking Accuracy) of 57.2%. Furthermore, a qualitative comparison with other multi-object tracking models showcased superior results for the model proposed in this paper. The research of this paper can be applied to the hygiene systems within cultural facilities in the post-COVID era.

**Keywords:** COVID-19, Epidemic Prevention System, Deep Learning, Object Detection, Multi-object Tracking

<sup>1</sup> 중앙대학교 첨단영상대학원 엔터테인먼트-테크놀로지 전공 석사과정 (wonseop218@gmail.com)

<sup>2\*</sup> 교신저자 중앙대학교 산업보안학과 교수 (smrho@cau.ac.kr)

## I. 서론

포스트-코로나 시대로 들어오면서, 새로운 감염병 등장의 위험과 이를 위한 방역 시스템에 대한 관심이 증가하고 있다. 위드, 포스트-코로나란 With와 COVID-19의 합성어로 COVID-19와 함께 살아가는 시기를 말한다. 포스트-코로나는 Post와 COVID-19의 합성어로 감염병 확산이 끝난 후 시기를 말한다. 위드-코로나 시기에 감염자와 접촉자가 이동한 시설과 동선 등을 분석하여 해당 위치에 있었던 접촉자들에게 안내를 해야 했으며, 감염자가 방문한 시설과 위치에 방역을 하였다. 밀집 시설에서 감염자가 나타났다면, 시설 내에서 감염자가 이동한 경로에 대해서 추적이 필요하기 때문에 CCTV를 통해 녹화된 동영상을 통해 감염자를 찾고 동선을 분석해야 했다.

특히 방문객들이 많은 박물관, 극장과 같은 밀집 문화 시설 내 감염자를 찾고 추적하여 동선을 분석하는 것을 관리자들이 일일이 모니터링 하는 것은 현실적으로 한계가 있다. 또한 새로운 감염병이 등장하여 위드-코로나 시기와 같이 상황이 온다면 이에 맞는 준비된 방역 시스템을 도입해야 한다.

최근 딥러닝 기술의 발전으로 객체를 탐지하거나 추적하는 연구들이 진행되어왔다. CNN[1]을 이용하여 객체의 특징을 추출하여 이를 학습시켜 객체를 탐지 및 추적한다. 이러한 다양한 모델들의 학습하는 시간을 줄이기 위해 기존에 사전학습 된 모델에 데이터 셋을 이용하여 전이학습을 시키는 방법[2]을 이용하여 학습에 필요한 시간을 줄이고 과적합(Over Fitting) 되는 것을 방지한다. 또한 감염병 확산 방지를 막기 위해 CNN 기반의 딥러닝 모델을 이용한 마스크 착용 탐지 모델 연구[3][4]가 진행되어왔다. 하지만 딥러닝을 적용하여 문화시설 내에서 방문객을 탐지하고 추적하는 연구 또한 필요하다.

본 논문에서는 문화시설 내에 있는 방문객들을 탐지하고 추적하는 방법을 제안한다. 방문객을 탐지하기 위하여 학습데이터 셋을 이용하여 딥러닝 기반의 객체 탐지 모델을 전이 학습시키고 탐지하며 학습한 모델을 딥러닝 기반 추적 모델에 적용하여 방문객을 추적한다. 이를 통해 효율적으로 문화 시설 내에서 방문객을 탐지하고 추적하여 포스트-코로나 시대에 맞는 방역 시스템에 적용될 수 있는 방법을 제안한다.

## II. 관련 연구

### 2.1 Convolution Neural Network (CNN)

일반적인 심층 신경망은 1차원 형태의 데이터로 한정되어 있기 때문에 이미지 데이터를 1차원 형태로 평면화 시키는 과정에서 공간 정보 손실이 있고 또한 많은 가중치와 편향을 가지고 있기 때문에 모델을 학습 및 추론 비용이 높다. 때문에 이러한 문제를 해결하기 위하여 Convolution Neural Network(CNN)가 제안되었다.

CNN은 계층 구조를 가지고 있다. 합성곱 계층(Convolution layer)과 풀링 계층(Pooling layer)으로 이루어져 있다. 합성곱 계층과 풀링 계층을 이용하여 입력 데이터로부터 특징 맵(Feature Map)을 추출해내며 고차원 데이터의 형상을 그대로 유지할 수 있다. 필터를 공유하여 기존의 신경망 기법과 비교하면 더 적은 파라미터를 가지고 있기 학습에 필요한 비용이 적게 든다. 이러한 특징을 가지고 있는 CNN을 이용하여 이미지 처리와 컴퓨터 비전 분야에서 성공적으로 활용되고 있다.

### 2.2 객체 탐지 연구

R-CNN[5]은 CNN을 객체 탐지(Object Detection) 분야에 처음 적용한 알고리즘이다. Region-based Convolutional Neural Networks의 약자로 객체가 있을 것 같은 위치를 추출하여 객체를 분류하고 탐지한다. R-CNN은 Region Proposals 과정을 거치는데 Selective Search 기법을 이용하여 다양한 크기와 위치를 후보 영역으로 추출한다. 추출한 후보 영역들을 CNN의 input으로 입력하여 후보 영역의 특징 맵을 추출한다. 추출한 특징 맵을 Soft Vector Machine[6]을 이용해 객체를 분류하여 Bounding Box에 회귀하여 학습시켜 추출한 특징과 실제 Bounding Box(label)간의 차이를 학습하여 탐지할 수 있다. 하지만 Region Proposals 과정에서 추출한 후보 영역들마다

처리를 해야 하기 때문에 학습 비용이 많이 든다는 단점이 있어 이를 해결하기 위해 Fast R-CNN[7]과 Mask R-CNN[8] 등이 제안되었다.

Fast R-CNN은 R-CNN의 학습 비용이 많이 드는 문제를 해결하기 위하여 제안되었다. CNN을 이용하여 이미지의 특징 맵을 1번만 추출한 후 Region of Interest (ROI) Pooling을 적용하여 후보영역을 고정된 크기로 변환한다. 추출한 특징 맵을 이용하여 Multi-Task loss를 이용하여 객체의 탐지와 분류를 동시에 진행한다. 객체를 분류하기 위하여 Soft Max Loss 함수를 사용해 추출한 후보영역에 대한 객체 분류를 확률을 예측한 후 실제 객체와의 차이를 계산하여 Classification Loss를 구한다. 객체 탐지는 L1 Loss 함수를 이용하여 실제 객체의 Bounding Box와 예측한 Bounding Box의 차이를 계산하여 Box regression Loss를 구한다. 구한 2가지의 Loss를 조합하여 Multi-Task loss를 계산해 분류와 탐지를 동시에 학습을 할 수 있도록 한다. 이러한 특징을 통해 R-CNN의 문제점을 해결할 수 있었고, R-CNN보다 속도와 정확성에서 높은 결과를 보여주었다.

R-CNN 계열 모델들은 특징 추출과 객체 분류를 나누어 진행하는 2-Stage Detector 모델이다. 정확도는 높지만 속도가 느려 실시간으로 처리해야 하는 문제에는 적용하기 어렵다. 이에 반해 YOLO[9]는 특징 추출과 객체 분류를 동시에 하는 1-Stage Detector로 특징 추출과 분류를 한꺼번에 처리하여 속도가 빨라 실시간으로 객체를 탐지할 수 있다.

YOLO는 이미지 전체를 한 번만 본다. R-CNN 계열 모델들은 이미지를 후보영역으로 여러 장 나누어 CNN 모델에 이용하는데, YOLO는 이미지 전체를 이용하여 객체를 탐지한다. 입력 이미지는  $S \times S$ 의 그리드 셀로 이미지를 나눈다. 각각의 그리드 셀은 해당 객체의 클래스와 Bounding Box의 개수를 예측한다. 또한 그리드를 기반으로 수행되기 때문에 다양한 크기의 객체를 탐지할 수 있다. 하지만 작은 객체들이 여러 개 모여 있는 경우 정확한 탐지를 수행하지 못하는 단점이 있다. 그럼에도 불구하고 처리 속도가 빠르기 때문에 실시간으로 객체를 분류하고 탐지해야 하는 문제에서는 계속해서 활용되고 있다.

### 2.3 객체 추적 연구

객체 추적 (Object Tracking)은 비디오와 같은 연속적인 프레임으로 구성되어 있는 데이터에서 움직이는 객체나 다중 객체를 추적해야 한다. Multi-Object Tracking (MoT)는 여러 개의 객체를 추적하는 것을 말한다.

SORT[10] 모델은 대표적인 MoT 모델이다. Tracking by detection 방식 중 Online Tracking 방식을 적용하였다. Online Tracking 방식은 이전 프레임과 현재 프레임의 객체를 탐지한 결과를 가지고 연관성을 계산하여 추적을 수행한다. SORT에서는 Faster R-CNN[11] 프레임워크를 이용하여 객체를 탐지하였고 탐지된 객체를 추적 프레임워크에 전달하여 추적을 진행한다. 추적하기 위해 탐지된 객체의 위치를 예측해야 하는데, 추적이 진행되는 프레임의 객체 위치를 예측하기 위하여 칼만 필터(Kalman Filter)를 사용한다. 이전 프레임에서의 객체의 위치나 크기 와 같은 변수들을 칼만 필터를 이용하여 추적하며 객체의 상태를 예측한다. 또한 Intersection over Union(IoU)를 기반으로 헝가리안 알고리즘을 이용해 이전 프레임과 현재 프레임의 객체의 연관성을 계산한다. 계산한 연관성을 통해 프레임별 객체를 매칭 시켜 추적을 진행한다. 하지만 객체의 위치 정보와 크기를 이용하기 때문에 Occlusion에 취약하고 ID switching이 빈번히 발생하는 문제점이 있다.

SORT의 문제점을 해결하기 위하여 DeepSORT[12]가 제안되었다. DeepSORT는 SORT 모델에 CNN을 적용하였다. CNN을 이용하여 객체 시각적 특징 정보를 추출하고 이를 프레임 간 연관성 계산에 적용한다. 또한 시각적 특징 정보를 이용한 Re-identification(Re-ID)를 통해 객체가 Occlusion되었다가 다시 등장해도 동일한 ID를 할당해줄 수 있다. 객체가 Occlusion되기 직전의 프레임에서의 객체의 정보와 다시 비교하여 매칭 시키고 추적 상태를 업데이트를 하는 방식으로 Re-ID를 진행한다. CNN을 이용하여 객체를 추적하여 SORT보다 높은 정확도를 보여주었으며, ID-Switching도 비교적 적게 나타났다. 하지만 CNN을 이용하기 때문에 계산 비용이 SORT보다 오래 걸린다는 단점이 있다. 본 논문에서는 실시간으로 문화 시설 내 사람을 탐지하는 것을 위해 YOLO 모델을 사용하였고, 객체 탐지 모델의 가중치를 객체 추적 모델에 적용하여 문화시설 내 방문객을 추적하였다.

## III. 방문객 탐지 및 추적 모델

본 논문에서는 딥러닝 기반으로 문화시설 내에서 방문객을 탐지하고 추적할 수 있도록 하였다. 그림 1 은 본 논문에서 제안하는 방문객을 탐지하고 추적하기 위해 수행하는 파이프라인을 보여준다.

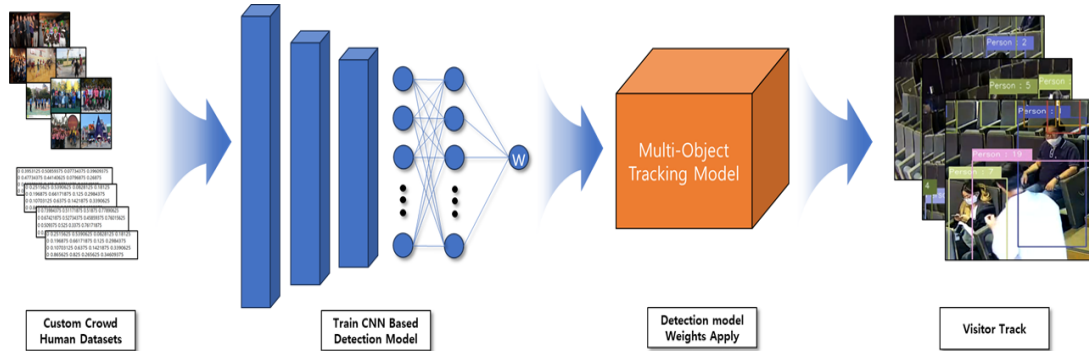


Figure 1. Visitor Detection • Tracking Full Process

그림 1. 방문객 탐지 • 추적 프로세스

### 3.1 방문객 탐지 모델

방문객을 탐지 모델을 학습하기 위하여 ‘Crowd Human’ 데이터 셋을 이용하여 진행하였다. Crowd Human 데이터 셋은 약 2 만 3 천개로 구성되어 있다. 데이터 셋의 Class 는 Face, Full-Body 로 구분되어 있으며 Face 는 0, Full-Body 는 1 의 값으로 labeling 되어 있다. Crowd Human 데이터 셋은 Bounding Box 가 image 에 보이는 부분만 Bounding Box 가 되어있지 않고 그림 2 와 같이 사람의 다리 부분이 Occlusion 되어도 해당 사람의 전체 비율에 맞춰 발끝 부분까지 Bounding Box 가 되어 있기 때문에 방문객의 특정 부분이 Occlusion 되어도 강건하게 탐지할 수 있을 것이라 판단하여 해당 데이터 셋을 이용하였다.

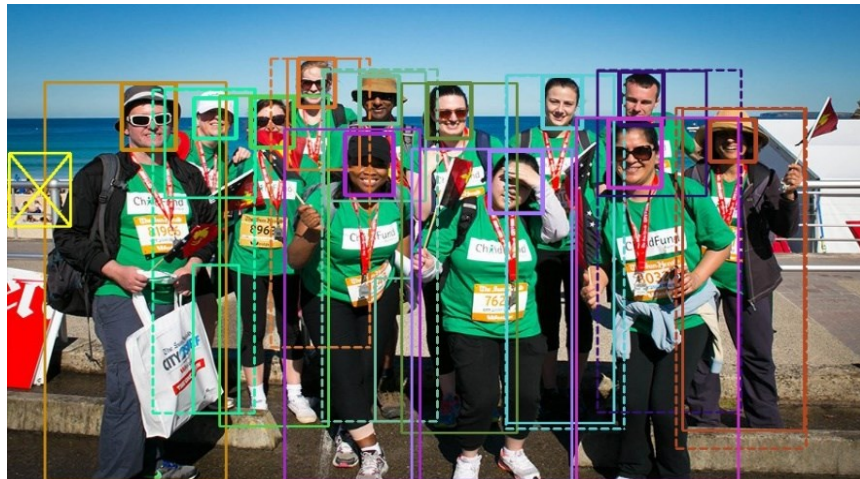


Figure 2. Example of Crowd human dataset

그림 2. Crowd human 데이터 셋 예시

본 논문의 연구에 맞게 label 이 0 인 Face Class 를 삭제하였고 Full-Body 의 label 값을 0 으로 수정하여 진행하였다. 또한 Crowd Human 데이터 셋의 labeling 파일이 odgt 형식으로 제공하여 YOLOv7 모델 학습을 위한 형식인 txt 형식으로 변형하였다.

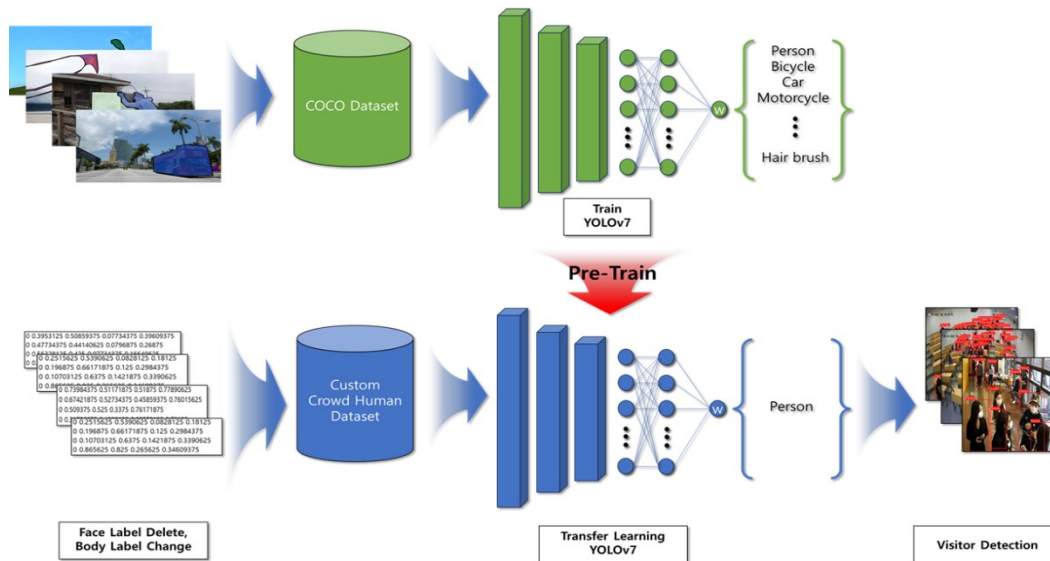


Figure 3. Visitor Detection Model Process

그림 3. 방문객 탐지 모델 프로세스

2022년에 YOLOv7은 기존 You Only Look Once(YOLO)의 새로운 버전의 모델로서, Extended efficient layer aggregation networks(E-ELAN)을 통해 확장하고, 복합 스케일링을 통해 계산량을 감소시킨다. 추가로 Trainable bag-of-freebies를 제안하여 정확도를 향상시킨 모델[13]이다. YOLOv7 모델은 6가지의 모델을 지원하는데, 그 중 모델 사이즈가 제일 작은 모델에 Transfer learning을 진행하였다.

그림 3은 방문객 탐지 모델 학습 파이프라인이다. COCO 데이터 셋으로 사전 학습된 모델을 기반으로 Hyper-Parameter를 튜닝하고 본 논문에서 이용한 Crowd Human 데이터 셋을 이용하여 Transfer learning을 통해 방문객 탐지 모델을 학습시켰다. 그림 4는 본 연구를 통해 학습된 방문객 탐지 모델의 추론 결과이다.



Figure 4. Detection model inference results

그림 4. 탐지 모델 추론 결과

### 3.2 방문객 추적 모델

본 논문에서 학습한 탐지 모델을 BoT-SORT 모델[14]에 적용시켰고 해당 모델의 추적 방식을 이용하여 방문객을 추적하였다. BoT-SORT의 칼만 필터는 기존의 SORT, Deep SORT의 칼만

필터와는 차별점이 있다. 두 모델의 칼만 필터의 상태 벡터는 타겟(객체)의 중심 수직 ( $x_c, x'_c$ )과 수평 픽셀위치( $y_c, y'_c$ ), Bounding Box 의 면적( $s$ ), 중횡비( $a$ )를 이용한다.

수식(1)은 SORT 모델이 사용하는 칼만 필터의 상태 벡터이다. 해당 칼만 필터 상태벡터를 이용하여 Bounding Box 를 예측하면 부정확하게 예측된다. 때문에 폭( $w$ )과 높이( $h$ )를 추가한 BoT-SORT 의 칼만 필터를 상태벡터로 이용하면 보다 정확하게 Bounding Box 를 예측할 수 있다. 수식(2)는 BoT-SORT 의 칼만 필터 상태 벡터이다.

$$X = [x_c, y_c, x'_c, y'_c, s, a] \quad (1)$$

$$X = [x_c, y_c, x'_c, y'_c, w, h, w', h'] \quad (2)$$

이를 통해 방문객 추적을 진행할 때 보다 정확하게 방문객을 Bounding 함으로써 객체 추적 및 매칭의 정확도를 높여 객체를 추적할 수 있다.

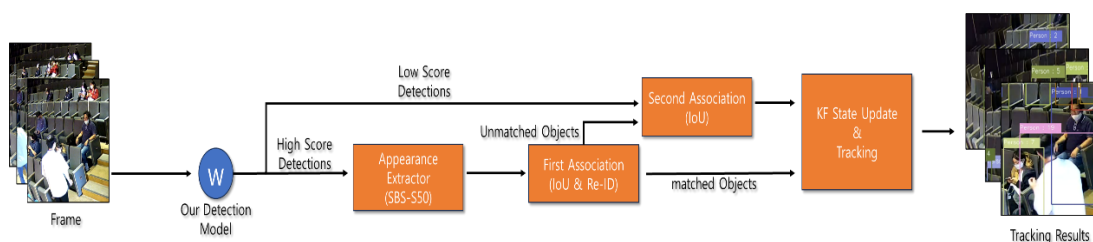


Figure 5. Visitor Tracking Model Process

그림 5. 방문객 추적 모델 프로세스

그림 5는 방문객 추적을 위한 프로세스 그림이다. 먼저 탐지 모델을 이용하여 프레임 내 객체를 탐지한다. 탐지된 객체의 결과가 높게 나온 프레임이라면 외관 정보를 추출하기 위해 SBS-S50[15]을 이용하여 외관정보를 추출한다. 추출한 정보로 IoU 와 Re-ID 를 기반으로 연관성을 계산하여 매칭 한다. IoU란 Intersection over Union 의 약자로 예측한 경계 상자 값과 labeling 한 정답 경계 상자 값의 겹쳐 있는 영역의 지표를 의미한다. 예측 값과 정답 값 간의 겹쳐 있는 영역이 클수록 IoU 값은 높게 측정된다.

낮은 탐지 결과를 가지고 있는 프레임의 경우, 특정 프레임에서 객체가 Occlusion 되었거나 매칭이 되지 않은 객체들과 함께 IoU를 기반으로 연관성을 계산하여 매칭 한다. 탐지 결과에 따른 연관성을 계산한 후 칼만 필터의 상태 벡터와 외관 정보를 업데이트하고 추적을 진행한다.

## IV. 실험 결과

### 4.1 방문객 탐지 모델 결과

본 논문에서는 Crowd Human 데이터 셋을 이용하여 CNN 기반 탐지 모델을 학습시켰고, 모델의 성능을 정량적으로 확인하였다. 방문객 탐지 모델의 학습 환경은 Intel(R) Core(TM) i7-10700K CPU, RAM 96GB, NVIDIA GeForce RTX 3090 GPU 1 개를 이용하여 학습하였다. 학습데이터 14,000 개, 검증 데이터 6,000 개, 테스트 데이터 3,000 개로 나누어 epoch 300 번, Adam Optimizer, learning rate 는 0.001 로 설정하였고 학습시간은 총 51 시간 소요되었다. 정량적인 평가는 Precision(P), Recall(R), F1-Score 지표를 이용하였다. 수식 (3), (4), (5)는 각각의 지표를 나타내는 수식이다.

$$\text{Precision(P)} = \frac{TP}{TP + FP} \quad (3)$$

$$\text{Recall(R)} = \frac{TP}{TP + FN} \quad (4)$$

$$\text{F1-Score} = 2 * \frac{P * R}{P + R} \quad (5)$$

P는 모델이 True 라고 분류한 결과 중 실제 True 인 비율을 의미한다. R은 재현율이라고 하며 실제 True 중에서 모델이 정확하게 예측한 비율을 의미한다. F1-Score는 P와 R의 조화평균이다. 표(1)은 방문객 탐지 모델을 평가한 표이다. 정확도 96.3%, 재현율 85.2%, F1-Score 90.4%로 평가하였다.

Table 1. Evaluation Results of Visitor Detection Model

표 1. 방문객 탐지 모델 평가 결과

Class	Precision (P)	Recall (R)	F1-Score
Person	96.3%	85.2%	90.4%

#### 4.2 방문객 추적 모델 결과

방문객 추적은 본 논문에서 이용한 모델을 성능을 정량적으로 평가하였고, 기존의 추적 모델들과 정성적으로 비교하였다. 표 (2)는 본 논문의 방문객 추적 모델을 평가한 지표이다. Multi-Object Tracking Accuracy(MOTA)[16], IDF1[17], High Order Tracking Accuracy(HOTA)[18]를 이용하여 평가하였다. MOTA는 객체 추적의 정확성을 측정하는데 사용되는 지표이며, IDF1은 객체 식별 정확성을 측정하기 위한 지표이다. HOTA는 객체 추적의 정확성을 측정하는 지표로, 객체의 위치, 크기, 속도 등 다양한 관점에서 객체 추적 성능을 평가한다. 본 논문의 추적 모델은 MOTA는 65.6%, IDF1은 68.3%, HOTA는 57.2%로 정량적으로 평가하였다.

Table 2. Evaluation Results of Visitor Tracking Model

표 2. 방문객 추적 모델 평가 결과

Class	MOTA	IDF1	HOTA
Person	65.6%	68.3%	57.2%

그림 6은 문화시설 내 관람객 입장 동영상을 이용하여 기존의 추적모델과 본 논문의 모델의 추적 성능을 비교한 결과이다. 그림 6(a)는 SORT 모델을 이용하여 추적한 결과이다. 짧은 프레임 동안 Occlusion 되었다가 다시 탐지되었을 때 새로운 ID로 바뀌는 것을 확인할 수 있었다. 그림 6(b), DeepSORT는 짧은 프레임 동안 Occlusion 되어도 계속해서 ID를 유지하였지만, 긴 프레임 동안 Occlusion 되어 있으면 ID가 바뀌는 것을 확인할 수 있었다. 그림 6(c)는 본 논문의 추적 모델 결과로, DeepSORT에서 ID가 바뀐 프레임과 비교해보았을 때 본 논문의 모델이 추적한 Person 7번의 ID가 바뀌지 않는 것을 확인할 수 있었다. SORT와 DeepSORT 모델에서는 추적된 객체들이 특정 프레임 동안 Occlusion 되면 ID가 바뀌었지만, 본 논문에서의 추적 모델은 비교한 두 모델보다 Occlusion에 강인하였다.

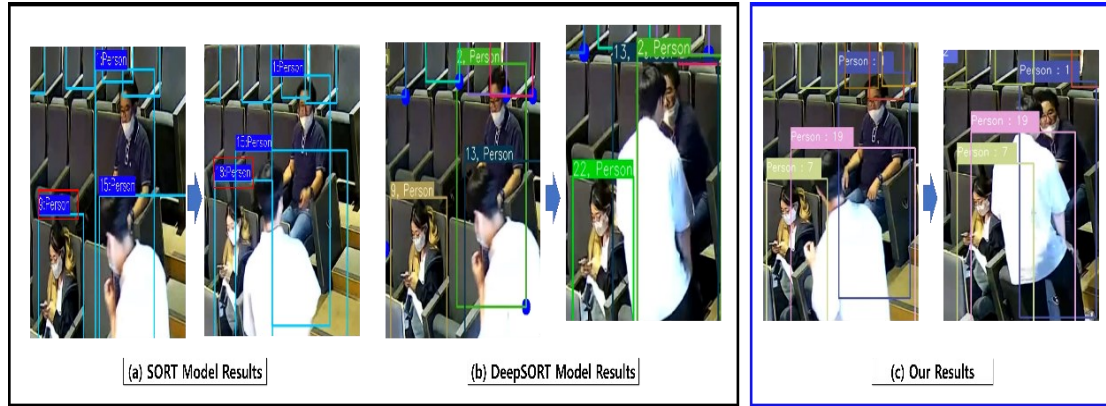


Figure 6. Results of Audience Tracking Model (a) SORT model Results, (b) DeepSORT model Results, (c) Our Results.

그림 6. 방문객 추적 모델 결과 (a) SORT model Results, (b) DeepSORT model Results, (c) Our Results

하지만, 본 논문에서의 모델도 Failure case 가 존재한다. 그림 7은 탐지 및 ID 할당 실패 사례를 나타낸다. 그림 7(a)는 방문객은 2 명이지만 4 명이 있는 것으로 탐지하였고 이에 맞춰 ID 도 할당되었다. 그림 7(b)는 ID 할당이 변경된 사례이다. 탐지된 방문객이 입장할 때는 23 번으로 할당되었지만, 9 번으로 재 할당되었다.

이러한 문제를 해결하기 위해 실제 문화시설 내 방문객 데이터를 데이터 셋에 추가하여 탐지 모델을 학습시켜 보다 향상된 탐지 모델 가중치를 추적 모델에 적용시키거나, 추가적인 추적 알고리즘 개발이 필요하다.



Figure 7. Failure Case of Detection & Re-ID (a) Failure Case of Detection, (b) Failure Case of Re-ID

그림 7. 탐지 및 ID 할당 실패 사례 (a) Failure Case of Detection, (b) Failure Case of Re-ID

## V. 결론

본 논문에서는 방문객을 추적하기 위하여, 먼저 CNN 기반 Detection 모델을 학습시켜 사람을 탐지하며 해당 탐지 모델을 이용하여 추적을 위한 MoT 모델에 적용하여 방문객을 추적하였다. Crowd Human 데이터 셋을 이용하여 Occlusion 상황이어도 보이지 않는 사람의 부분까지 인식하는 CNN 기반 Human Detection 모델을 학습시켰다. 학습시킨 모델을 Realtime 에 적합한 Human Tracking 모델에 적용하여 전이학습을 시켜 방문객의 모습이 가려져 있어도 방문객의 전체 형상을 추정하여 추적하였다. 본 논문에서의 방문객 탐지 모델은 높은 정확도와 속도를



보여주었으며, 기존의 추적 모델보다 정확하게 추적하는 것을 정성적으로 확인하였다. 이를 통해 실제 문화시설에 적용하여 문화시설 내 방문객을 추적할 수 있다.

하지만, 두 모델보다는 강인하게 추적하였으나 여전히 Occlusion 후 다시 추적하였을 때 해당 방문객의 ID가 바뀌는 문제가 발생하여 해당 문제를 해결하는 연구가 필요하다. 향후 문화시설 내에 기술을 적용하기 위해서, 앞서 서술한 폐색 문제를 해결하기 위한 추가 연구와 영화나 뮤지컬 등이 상영 중인 극장과 같은 Gray Scale 환경에서도 강인하게 추적하는 연구를 진행할 것이다. 또한 방문객을 추적한 후 문화시설 도면에 Mapping 하여 추적된 방문객의 동선을 표시하고 분석하는 연구를 진행할 것이다. 추가 연구들을 통해 보다 강건하게 방문객을 추적하고 동선을 분석하는 시스템을 개발하여 포스트 코로나 시대에 맞는 방역 시스템으로 활용될 수 있다.

## VI. 감사의 글

이 논문은 2023년도 중앙대학교 연구장학기금 지원에 의한 것임.

## VII. 참고문헌

- [1] LeCun, Y., Bengio, Y., and Hinton, G. "Gradient-based learning applied to document recognition." *Proceedings of the IEEE*, 86(11), pp. 2278-2324. doi: 10.1109/5.726791, 1998.
- [2] J. Talukdar, S. Gupta, P. S. Rajpura, and R. S. Hegde, "Transfer Learning for Object Detection using State-of-the-Art Deep Neural Networks," in *Proceeding of the 5th International Conference on Signal Processing and Integrated Networks*, Noida, pp. 78-83, doi: 10.1109/SPIN.2018.8474198, 2018.
- [3] S. H. Lee, H. G. Kwon, Y. J. Kim, J. S. Jeong, and H. J. Seo, "Development of CCTV for Identification of Maskless Wearers based on Deep Learning," in *Proceeding of the 28th Korea Society of Computer Information*, Korea, pp. 317-318, doi: 10.6109/jkiice.2021.25.1.44, 2020.
- [4] Sunwoo Shin, Woosung Jung, Taemin Lee and Sanghyun Seo. "A Study on Detecting Mask Wearing Status using Ensemble based on Deep Learning" *Journal of Digital Contents Society*. Vol 22, No.11 pp. 1931-1939, doi: 10.9728/dcs.2021.22.11.1931, November 2021.
- [5] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation," *2014 IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, USA, pp. 580-587, doi: 10.1109/CVPR.2014.81, 2014.
- [6] M. A. Hearst, S.T. Dumais, E. Osuna, J. Platt, and B. Scholkopf, "Support vector machines," in *IEEE Intelligent Systems and their Applications*, vol. 13, no. 4, pp. 18-28, doi: 10.1109/5254.708428, July-Aug. 1998.
- [7] R. Girshick, "Fast R-CNN," *2015 IEEE International Conference on Computer Vision (ICCV)*, Santiago, Chile, pp. 1440-1448, doi: 10.1109/ICCV.2015.169, 2015.
- [8] He, K., Gkioxari, G., Dollar, P., and Girshick, R. "Mask R-CNN" *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 2980-2988, doi: 10.1109/ICCV.2017.322., 2017.
- [9] Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. "You only look once: Unified, real-time object detection." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 779-788, doi: 10.1109/CVPR.2016.91, 2016.
- [10] Bewley, A., Ge, Z., Ott, L., Ramos, F., and Upcroft, B. "Simple online and realtime tracking." *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, pp. 3464-3468, doi: 10.1109/ICIP.2016.7533003, 2016.
- [11] Ren, S., He, K., Girshick, R., and Sun, J. "Faster R-CNN: Towards real-time object detection with region proposal networks." *Proceedings of the Neural Information Processing Systems (NIPS) Conference*, pp. 91-99, doi: 10.1109/ICCV.2015.169, 2015.
- [12] Woehlke, S., Büttner, L., Kreuzinger, N, and Leal-Taixé, L. "DeepSORT: A Simple Online and Realtime Tracking with Re-identification." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2573-2582, doi: 10.1109/CVPR42600.2020.00264, 2020.

- [13] Nir Aharon, Roy Orfaig, and Ben-Zion Bobrovsky, "BoT-SORT: Robust Associations Multi-Pedestrian Tracking" arXiv, doi: arXiv:2206.14651, 2022.
- [14] Wang, Chien-Yao & Bochkovski, Alexey, and Liao, Hong-yuan. "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors."arXiv, doi: 10.48550/arXiv.2207.02696, 2022.
- [15] He, L., Liao, X., Liu, W., Liu, X., Cheng, P, and Mei, T. (2020). "FastReID: A Pytorch Toolbox for General Instance Re-identification. arXiv, abs/2006.02631, 2020.
- [16] Bernardin, K., Stiefelhagen, R. "Evaluating Multiple Object Tracking Performance: The CLEAR MOT Metrics." J Image Video Proc 2008, 246309, doi: 10.1155/2008/246309, 2008.
- [17] Ristani, E., Solera, F., Zou, R., Cucchiara, R., Tomasi, C., "Performance Measures and a Data Set for Multi-target, Multi-camera Tracking." ECCV vol 9914, doi: 10.1007/978-3-319-48881-3\_2, 2016.
- [18] Luiten, J., Osep, A., Dendorfer, P., Torr, P., Geiger, A., Leal-Taixe, L., & Leibe, B, "HOTA: A Higher Order Metric for Evaluating Multi-Object Tracking." International Journal of Computer Vision, 129(2), pp. 548-578, doi: 10.1007/s11263-020-01390-7, 2021

## 저자소개



**신원섭 (Wonseop Shin)**

2023 년 ~ 현 재: 중앙대학교 첨단영상대학원 영상학과 엔터테인먼트 테크놀로지 전공 석사과정

관심분야 : Deep Learning, Object Detection, Computer Vision, Virtual Reality/Augmented Reality



**노승민 (Seungmin Rho)**

2013 년 ~ 2019 년 : 성결대학교 미디어소프트웨어학과 조교수

2019 년 ~ 2021 년 : 세종대학교 소프트웨어학과 조교수

2021 년 ~ 현재 : 중앙대학교 산업보안학과 부교수

관심분야 : Big Data Security, Artificial Intelligence Security, ContentSecurity