

다중 객체 감지 및 추적을 위한 YOLOv9 과 BoT-SORT 알고리즘의 Re-ID 성능 분석

Performance Analysis of Re-ID in YOLOv9 and BoT-SORT Algorithm for Multi-object Detection and Tracking

김선영¹, 강창호^{2*}

(Sun Young Kim¹ and Chang Ho Kang^{2,*})

¹School of Mechanical Engineering, Kunsan National University

²Department of Artificial Intelligence and Robotics, Sejong University

Abstract: This paper proposes a method for integrating YOLOv9 and BoT-SORT algorithms for multi-object tracking in a multi-camera environment. Particularly, it focuses on analyzing the impact of re-identification (Re-ID) techniques on the overall tracking system performance. Three Re-ID methods—MobileNetV2, OSNet, and OSNet-AIN—are evaluated using the MSMT17 dataset and then applied to the BoT-SORT algorithm to compare multi-object tracking performances. Experimental results show that BoT-SORT with OSNet-AIN achieves the best performance in HOTA, MOTA, and IDF1 metrics. This demonstrates that effective Re-ID techniques can significantly contribute to improving object tracking performance in complex real-world environments.

Keywords: YOLOv9, BoT-SORT, Re-ID, multi-object detection and tracking

I. 서론

자율 감시 시스템, 자율 주행 차량, 스마트 환경의 발전은 여러 관점에서 동시에 시각 데이터를 분석하고 해석하는 강건하고 효율적인 방법에 대한 수요를 촉발하였다. 이 분야는 여러 카메라의 시점을 통해 객체를 감지하고 추적하는 연구를 다루며, 컴퓨터 비전 응용의 정확성과 신뢰성을 향상시키는 것을 목표로 한다[1,2]. 다시점(multi-view) 다중 카메라를 이용한 객체 감지 및 추적의 복잡성은 가림 현상, 조명 변화, 카메라 교정 차이, 실시간 처리의 필요성과 같은 다양한 요인에서 비롯된다. 이러한 문제를 해결하기 위해서는 다양한 시점의 정보를 융합하고, 중첩되는 객체 궤적을 처리하며, 동적 환경 조건에 적용할 수 있는 정교한 알고리즘이 필요하다.

기존의 연구자들은 이러한 문제를 해결하기 위한 다양한 접근 방법을 제안하였다[3]. 합성곱 신경망(CNN, Convolutional Neural Network)과 같은 딥러닝 기반 방법은 여러 시점에서의 객체 감지 및 추적에 뛰어난 성능을 보여주었다. 또한, 센서 교정, 특성 매칭, 데이터 연관성을 포함한 융합 기술이 다양한 카메라에서 정보를 원활하게 통합하는 데 탐구되었다. 정교한 감시 및 지능형 시스템에 대한 수요가 계속 증가함에 따라, 강건한 다시점 다중 카메라 객체 감지 및 추적 알고리즘 개발은 컴퓨터 비전 및 관련 분야의 연구자들에게 중요한 쟁점이다. Faster R-CNN [4]과 Mask R-CNN [5]과 같은 영역 기반(region-based) CNN 아키텍처는 이미지 내 객체를 정확하게 식별하고 분류하는 데 능숙하며, 이어지는 추적 작업을 위한

견고한 기반을 마련하였다. 순환 신경망(RNN, Recurrent Neural Network) [6]과 장단기 메모리(LSTM, Long Short-Term Memory) [7]는 시간적 의존성을 포착하는 반면, 삼(Siamese) 네트워크[8]는 온라인 객체 추적을 용이하게 한다. 일부 제안된 방식은 객체 감지와 추적에 딥러닝 기법을 사용하여 하나의 구조로 통합한다. 이러한 예로는 객체 감지와 온라인 추적의 강점을 결합하여 강건한 성능을 달성하는 TWD (Tracking-While Detecting) 프레임워크가 있다.

본 논문은 다중 카메라의 다시점 환경에서 객체를 감지하고 추적하는 방법을 제안한다. YOLOv9 (You Only Look Once version 9) [9]은 각 시점에서 객체를 식별하는 데 사용되고 시간에 따라서 대상의 이동을 정밀하게 추적하기 위해서 BoT-SORT (Simple Online and Realtime Tracking) [10]기법을 사용하여 계속적으로 대상을 감시 및 추적할 수 있도록 하였다. 그리고 BoT-SORT 기법에 추적 성능을 향상시킬 수 있는 재식별(Re-ID, Re-Identification) [11] 기법을 적용하였다.

개인 Re-ID는 서로 다른 카메라 뷰에서 동일한 개인을 매칭하는 것을 목표로 하는 컴퓨터 비전 작업이다. 이는 다중 카메라 감시 시스템, 스마트 시티, 그리고 보안 응용 프로그램에서 중요한 역할을 한다. Re-ID 기술의 발전은 크게 특징 공학에서 딥러닝 기반의 종단간(end-to-end) 특징 표현 학습으로 패러다임이 변화해왔다. 최근의 Re-ID 연구는 더욱 복잡하고 효과적인 네트워크 구조에 초점을 맞추고 있다. 특히, 다중 스케일 특징 학습, 주의 메커니즘, 그리고 부분 기반 모

* Corresponding Author

Manuscript received September 22, 2024; revised November 7, 2024; accepted December 9, 2024

김선영: 국립군산대학교 기계공학부 부교수(sykim77@kunsan.ac.kr, ORCID[®] 0000-0002-6189-0268)

강창호: 세종대학교 AI로봇학과 조교수(kcguri@sejong.ac.kr, ORCID[®] 0000-0002-9899-3076)

※ 본 연구는 과학기술정보통신부의 재원으로 한국연구재단의 지원을 받아 수행되었음(No. 2021R1C1C1009219).

※ 이 논문은 2024 제39회 제어·로봇·시스템학회 학술대회에 초안이 발표되었음.

델 등이 주목받고 있다. 이러한 맥락에서 OSNet (Omni-Scale Network) [12]은 주목할 만한 혁신을 보여준다. 본 논문에서는 YOLOv9과 BoT-SORT 알고리즘에 OSNet 기반 Re-ID 기법을 적용하여 영상 객체 감지 및 추적 성능을 기존 기법들과 비교 분석하는 것을 중점으로 하였다.

본 연구에서 제안하는 YOLOv9과 BoT-SORT를 결합한 다중 객체 추적 시스템은 YOLOv9의 향상된 객체 검출 성능과 BoT-SORT의 강력한 추적 알고리즘이 결합되어, 특히 작은 객체나 부분적으로 가려진 객체에 대해서도 우수한 검출 및 추적 성능을 보여준다. BoT-SORT의 IoU (Intersection Over Union) 거리와 Re-ID 특징의 코사인 거리를 결합한 새로운 융합 방식은 객체의 공간적 위치 정보와 외관 특징을 동시에 고려하여 혼잡한 환경에서도 안정적인 추적이 가능하다. 또한, FastReID 라이브러리의 SBS (stronger baseline) 모델과 ResNeSt50 백본을 통한 강력한 특징 표현, 그리고 지수 이동 평균(EMA, Exponential Moving Average) 메커니즘의 도입으로 장기 추적 성능이 크게 향상되었다.

나머지 논문 구성은 다음과 같다. 2장에서는 실시간 객체 감지 알고리즘 중 하나인 YOLOv9에 대해 설명하고 3장에서 다중 객체 추적 알고리즘 중 하나인 BoT-SORT 알고리즘에 대해서 설명하고, 전체 알고리즘 성능 향상을 위해 변경된 구조 및 특징에 대해 자세히 기술한다. 특히 네트워크 기반의 Re-ID 기법을 분석한다. 그리고 4장에서 이를 검증하기 위한 시뮬레이션 결과를 정리하고 5장에서 결론을 맺는다.

II. YOLOv9

YOLOv9은 실시간 객체 감지 분야에서 주목할 만한 발전을 이루었다. 이 모델은 PGI (Programmable Gradient Information)와 GELAN (Generalized Efficient Layer Aggregation Network) 같은 혁신적인 기술을 도입하여 효율성, 정확성, 적응성에서 뛰어난 개선을 보여주며 MS COCO 데이터셋에서 새로운 기준을 설정하였다.

실시간 객체 감지를 위한 최적화 과정에서, YOLOv9은 깊은 신경망(deep neural network)에서 흔히 발생하는 정보 손실 문제를 극복하기 위해 혁신적인 접근 방식을 채택한다. 이 모델은 PGI와 다채 다능한 GELAN 아키텍처를 통합함으로써, 감지 과정 전반에 걸쳐 중요한 정보를 보존하고, 이를 통해 뛰어난 정확도와 성능을 달성한다. 이러한 기술적 통합은 YOLOv9의 학습 능력을 크게 향상시키고, 중요한 데이터의 연속성을 유지한다.

YOLOv9의 설계 핵심은 정보 병목 원리(information bottleneck principle)와 가역 함수(reversible functions)의 적용에 있다. 정보 병목 원리는 데이터가 네트워크의 연속적인 레이어를 통과 하면서 정보 손실이 증가하는 문제를 지적한다. 이에 대응하여, YOLOv9은 PGI를 구현하여 네트워크의 깊은 부분에 걸쳐 필수 데이터의 손실을 최소화하고, 이를 통해 더 안정적인 그라디언트(gradient) 생성과 모델 수렴을 달성한다. 또한, 가역 함수를 사용하여 특히 더 깊은 레이어에서 정보의 열화를 방지하고, 객체 감지 작업에 필수적인 데이터를 보존한다.

이러한 설계는 특히 경량 모델에서 중요한데, 경량 모델은 과소 매개변수화로 인해 피드포워드(feedforward) 과정에서

중요 정보를 잃기 쉽다. YOLOv9은 PGI와 가역 함수를 활용하여, 간소화된 모델에서도 정확한 객체 감지를 위해 필수적인 정보가 효과적으로 보존되고 활용되도록 한다. 이를 통해, YOLOv9은 속도나 정확도를 희생하지 않고 다양한 응용 프로그램에 적용할 수 있는 GELAN의 설계와 함께 뛰어난 매개변수 활용과 계산 효율성을 달성한다. 이러한 기능은 YOLOv9을 다양한 환경에서 신뢰할 수 있는 객체 감지 솔루션으로 만들어 준다.

III. BoT-SORT 알고리즘

1. BoT-SORT

BoT-SORT는 다중 객체 추적 분야에서 최신 성능을 달성한 알고리즘으로, 기존의 추적 기법들을 개선하고 새로운 아이디어를 통합하여 높은 정확도와 안정성을 제공하며 주요 처리과정은 그림 1과 같이 나타낼 수 있다. BoT-SORT의 핵심 구조는 여러 주요 처리 단계로 구성되어 있다. 먼저, 입력 프레임에 대해 YOLO 검출기를 사용하여 객체를 감지한다. 동시에 연속된 프레임 간의 카메라 움직임을 추정하고, 이를 바탕으로 칼만(Kalman) 필터를 사용하여 기존 트랙렛(tracklet)의 위치를 예측한다. 검출된 객체에 대해서는 딥 외관 특징 추출기(SBS-S50)를 사용하여 Re-ID 특징을 추출한다. 이후 첫 번째 연관 단계에서는 고신뢰도 검출에 대해 IoU와 Re-ID 특징을 결합하여 기존 트랙렛과 연관시킨다. 두 번째 연관 단계에서는 저신뢰도 검출에 대해 IoU 기반 연관을 수행한다. 이러한 연관 결과를 바탕으로 트랙렛을 관리하는데, 이 과정에서 트랙렛을 업데이트하고, 새로운 트랙렛을 생성하거나 비활성 트랙렛을 제거한다. 최종적으로 이 모든 과정을 거쳐 처리된 트랙을 결과로 출력한다. 이 전체 과정은 각 프레임마다 반복되어 연속적인 다중 객체 추적을 수행하며, BoT-SORT 알고리즘의 효과적인 추적 성능을 가능하게 한다. 이 알고리즘은 ByteTrack [13]을 기반으로 하지만, 여러 가지 중요한 개선 사항을 도입하여 성능을 크게 향상시켰다.

BoT-SORT의 주요 혁신 중 하나는 개선된 칼만 필터의 사용이다. 기존의 많은 추적 알고리즘들이 단순한 상태 벡터를 사용한 반면, BoT-SORT는 더 정교한 상태 벡터를 도입하였다. 이 벡터는 객체의 중심 좌표($x_c(k), y_c(k)$) 뿐만 아니라 너비 $w(k)$ 와 높이 $h(k)$, 그리고 각각의 변화율($\dot{x}_c(k), \dot{y}_c(k), \dot{w}(k), \dot{h}(k)$)을 포함한다. 구체적으로, 상태변수 벡터는 다음과 같이 정의된다.

$$x_k = [x_c(k), y_c(k), w(k), h(k), \dot{x}_c(k), \dot{y}_c(k), \dot{w}(k), \dot{h}(k)]^T \quad (1)$$

이러한 접근 방식은 경계 상자(bounding box)의 크기 변화를 더 정확하게 예측할 수 있게 하여, 객체 추적의 정확도를 향상시킨다.

또 다른 중요한 혁신은 카메라 모션 보상(CMC, Camera Motion Compensation) 기법의 도입이다. 많은 추적 알고리즘들이 정적 카메라를 가정하거나 단순한 모션 모델을 사용하는 반면, BoT-SORT는 프레임 간 카메라 움직임을 명시적으로 고려한다. 이를 위해 인접한 프레임 사이의 어파인(affine) 변환 행렬을 계산하고, 이를 사용하여 객체의 예측 위치를 보정한다.

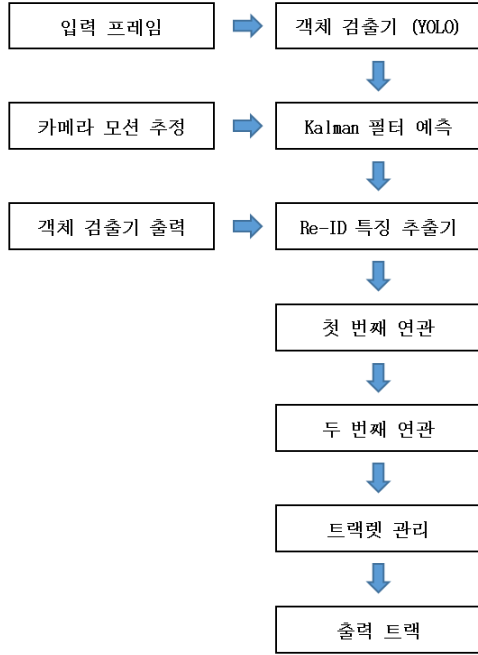


그림 1. BoT-SORT 주요 처리과정.

Fig. 1. The main process of BoT-SORT.

BoT-SORT의 또 다른 주요 특징은 IoU 거리와 Re-ID 특징의 코사인(cosine) 거리를 결합한 새로운 융합 방식이다. 이 방법은 객체의 공간적 위치와 외관 정보를 동시에 고려하여 더 강력한 데이터 연관성을 제공한다. 방식은 기존의 단순한 가중 합 방식보다 더 유연하고 효과적인 특징 융합을 가능하게 한다. 또한, BoT-SORT는 최신의 Re-ID 모델을 통합하여 객체의 외관 특징을 더 정확하게 추출하고 활용한다. FastReID 라이브러리의 SBS 모델을 사용하며, 이는 ResNeSt50을 백본(backbone)으로 하여 강력한 특징 표현을 제공한다. ResNeSt(Split-Attention Networks) [14]는 2022년에 제안된 새로운 아키텍처로, 기존 ResNet의 장점을 유지하면서 성능을 개선한 모델이다. ResNeSt는 특징 맵(feature map)을 여러 그룹으로 나누고, 각 그룹에 대해 별도의 어텐션(attention)을 적용하여 이를 통해 더 풍부한 특징 표현이 가능해진다. ResNeSt50은 일반적인 ResNet50과 비교했을 때, 비슷한 계산 복잡도를 유지하면서도 더 좋은 성능을 보이며, 특히 이미지 분류나 객체 탐지 같은 다양한 컴퓨터 비전 작업에서 우수한 성능을 보여준다.

마지막으로, BoT-SORT는 특징 업데이트에 EMA 메커니즘을 도입하여 시간에 따른 객체 외관의 변화를 효과적으로 반영한다. 이는 장기 추적 성능을 향상시키는 데 기여한다. 개선된 요소들의 조합으로 인해 BoT-SORT는 기존의 추적 알고리즘들보다 더 높은 정확도와 강건성을 제공하며, 특히 복잡한 실제 환경에서의 다중 객체 추적 문제에 효과적으로 대응할 수 있다.

2. Re-ID 알고리즘

BoT-SORT는 다중 객체 추적을 위한 강력한 알고리즘으로, 이미 Re-ID 기법을 효과적으로 활용하고 있다. 이 알고리즘은 ByteTrack을 기반으로 하여 여러 가지 개선 사항을 통합하였으며, 그 중 Re-ID 기능은 추적 성능 향상에 중요한 역할을 한다.

BoT-SORT에서 Re-ID 기법의 활용은 다음과 같은 방식으로 이루어진다. 먼저, FastReID 라이브러리의 SBS 모델을 Re-ID 특징 추출기로 사용한다. 이 모델은 ResNeSt50을 백본으로 사용하여 객체의 외관 특징을 효과적으로 추출한다. 추적된 객체의 외관 특징은 지수 이동 평균 메커니즘을 통해 업데이트된다. 이는 다음 수식으로 표현된다.

$$e_i^k = \alpha e_i^{k-1} + (1-\alpha) f_i^k \quad (2)$$

여기서 e_i^k 는 i 번째 트랙의 k 프레임에서의 평균 외관 상태, f_i^k 는 현재 매칭된 검출의 외관 임베딩, 그리고 α 는 모델 업데이트 방향으로 기본적으로 0.9로 설정한다.

BoT-SORT의 주요 혁신 중 하나는 IoU 거리와 Re-ID 특징의 코사인 거리를 결합한 새로운 융합 방식이다. 이 방식은 다음 수식으로 정리할 수 있다.

$$d \cos(i, j) = \begin{cases} 0.5 \times d \cos(i, j), & \text{if } d \cos(i, j) < \theta_{emb} \\ & \& diou(i, j) < \theta_{iou} \\ 1, & \text{otherwise} \end{cases} \quad (3)$$

여기서 $d \cos(i, j)$ 는 코사인 거리이고 $diou(i, j)$ 는 IoU 거리, 그리고 θ_{emb} 는 외관 임계값, 그리고 θ_{iou} 는 IoU 임계값이다. 최종 비용 행렬의 요소($C(i, j)$)는 다음과 같이 정의된다.

$$C(i, j) = \min(diou(i, j), d \cos(i, j)) \quad (4)$$

이를 통해 모션 정보와 외관 정보를 효과적으로 결합하여 더 정확한 객체 연관성을 제공한다.

계산 효율성을 위해 BoT-SORT는 높은 신뢰도를 가진 검출에 대해서만 선택적으로 Re-ID 특징을 추출한다. 이러한 Re-ID 특징은 주로 높은 신뢰도를 가진 검출과 트랙렛을 연관시키는 첫 번째 단계에서 사용되며, 헝가리안(Hungarian) 알고리즘을 통해 전역 최적 매칭을 찾는 데 활용된다. BoT-SORT에서 Re-ID 기법의 활용은 여러 가지 이점을 제공한다. 외관 정보를 활용함으로써 혼잡한 환경이나 객체 간 겹침이 발생하는 상황에서도 안정적인 추적이 가능하다. 또한, Re-ID 특징은 객체가 일시적으로 사라졌다가 다시 나타나는 상황에서 재식별을 가능하게 하여 장기 추적 성능을 개선한다. 더불어, 외관 정보를 활용하여 유사한 모션을 가진 다른 객체들 사이의 구분을 개선함으로써 ID 스위치를 감소시킨다. 결론적으로, BoT-SORT에서의 Re-ID 기법 활용은 추적의 정확성과 강건성을 크게 향상시키는 핵심 요소이다. 이는 복잡한 실제 환경에서의 다중 객체 추적 성능을 개선하는 데 중요한 역할을 한다.

최근의 Re-ID 연구는 더욱 복잡하고 효과적인 네트워크 구조에 초점을 맞추고 있다. 특히, 다중 스케일 특징 학습, 주의 메커니즘, 그리고 부분 기반 모델 등이 주목받고 있다. 이러한 맥락에서 OSNet은 주목할 만한 혁신을 보여준다. 본 논문에서는 다중 객체 추적을 위한 Re-ID 모듈로 세 가지 네트워크 MobileNetV2, OSNet, 그리고 OSNet-AIN의 구조를 비교 분석한다. 이들 네트워크는 각각 고유한 특성을 가지고 있으며, 다중 객체 추적 작업에서 서로 다른 장단점을 제공한다.

MobileNetV2는 모바일 및 임베디드 비전 애플리케이션을 위해 설계된 경량 CNN 구조이다. 이 네트워크의 핵심은 깊이별 분리 합성곱(depthwise separable convolution) 연산과 반전된 잔차(inverted residual) 구조의 사용이다. 깊이별 분리 합성곱은 표준 합성곱을 깊이별 합성곱과 점별 합성곱(pointwise convolution)으로 분해하여 계산 효율성을 크게 향상시킨다. 반전된 잔차 구조는 저차원 표현을 사용하여 메모리 효율성을 개선한다. MobileNetV2는 2.2M의 파라미터를 가지며, 계산 효율성과 정확도 사이의 균형을 제공한다.

OSNet은 다중 스케일 특징 학습을 위해 특별히 설계된 네트워크 구조이다. 이 네트워크의 주요 특징은 다양한 수용 영역 크기를 가진 여러 컨볼루션 스트림으로 구성된 옴니 스케일(omni-scale) 잔차 블록이다. 이 블록은 다음과 같이 표현될 수 있다.

$$\tilde{x} = \sum_{t=1}^T AG(F_t(x)) \quad (5)$$

여기서 $F_t(x)$ 는 t번째 스트림의 변환 함수이고, AG는 어그리게이션 게이트(aggregation gate)이다. OSNet은 2.2M의 파라미터를 가지며, 경량성을 유지하면서도 다양한 스케일의 특징을 효과적으로 학습할 수 있다.

OSNet-AIN은 OSNet을 기반으로 하되, 인스턴스 정규화(IN, Instance Normalization) 레이어를 최적의 위치에 추가한 구조이다. IN 레이어의 위치는 신경망 아키텍처 조사(NAS, Neural Architecture Search)를 통해 자동으로 결정된다. 이 과정은 다음과 같은 최적화 문제로 표현될 수 있다.

$$\min_{\alpha, \theta} E[L(x, \theta, f_{\pi}(z))] \quad (6)$$

여기서 α 는 아키텍처 파라미터, θ 는 네트워크 가중치, L 은 손실 함수이다. OSNet-AIN은 2.2M의 파라미터를 가지며, 도메인 일반화 능력이 크게 향상되어 다양한 환경에서의 Re-ID 성능을 개선한다.

OSNet 기반의 Re-ID 기법의 성능 분석을 위해서 BoT-SORT에 이 방법을 적용하여 다중 객체 특히 여러 사람을 다양한 영상 시각에서 추적할 수 있는지 여러 시뮬레이션을 수행하였다.

IV. 시뮬레이션 및 결과

1. 시뮬레이션 설정

COCO 데이터셋으로 사전에 학습된 YOLOv9을 이용해서 그림 2와 같이 동영상 속 사람을 추적하는 실험을 진행하면서 제안한 알고리즘의 성능을 확인하였다. 또한 Re-ID에 사용된 알고리즘들은 MSMT17 데이터셋을 이용하여 학습을 수행하였다. 이 데이터셋은 4,101명의 보행자의 보행 영상이 담겨있고 총 126,411개의 이미지이고 15개의 카메라로 촬영된 것이다. 모든 모델은 동일한 학습 설정을 사용하여 공정한 비교를 가능하게 하였다. 옵티마이저로는 Adam의 변형인 AMSGrad를 채택하였으며, 초기 학습률은 0.0015로 설정하였다. 학습률은 처음 10 에폭 동안 선형적으로 증가한 후, 코사인 어닐링 스케줄을 따라 감소하도록 설계되었다. 배치 크기는

64로 고정되었으며, 총 150 에폭 동안 학습을 진행하였다. 가중치 감소는 $5e-4$ 로 설정되어 과적합을 방지하였다. 데이터 증강 기법으로는 무작위 뒤집기, 무작위 자르기, 그리고 무작위 지우기를 적용하여 모델의 일반화 능력을 향상시켰다. 손실 함수로는 레이블 스무딩이 적용된 교차 엔트로피 손실을 사용하여 학습의 안정성을 높였다. 입력 이미지 크기는 256×128 로 통일하여 처리하였다. 이러한 학습 설정은 MSMT17 데이터셋에서 최적의 성능을 얻기 위해 경험적으로 선택되었다. 성능평가는 Rank-1 정확도 (R1)와 mAP (Mean Average Precision)를 사용하였다. R1은 쿼리 이미지와 가장 유사한 갤러리 이미지가 올바른 매치(match)일 확률을 의미하는 것이고 mAP는 전체 랭킹에 대한 평균 정밀도를 의미한다.

2. 시뮬레이션 결과 분석

그림 2에서 인식된 객체의 번호와 경계 상자 그리고 시간에 따른 추적 결과를 상자 색과 동일하게 표시하였고 동영상에서의 대상 궤적과 비교해 보면서 성능을 분석하였다. 표 1은 BoT-SORT에 적용하기 전 Re-ID 기능에 대한 평가이다. 표 1의 결과를 통해서 OSNet과 OSNet-AIN이 MobileNetV2보다 모든 성능 지표에서 우수한 성능을 보이며, 특히 OSNet-AIN이 가장 높은 성능을 달성함을 알 수 있다. 이는 옴니 스케일 특징 학습과 최적화된 IN 레이어 배치가 Re-ID 작업에 효과적임을 보여준다. 이러한 성능 차이가 BoT-SORT와 같은 다중 객체 추적 시스템에서 어떻게 나타나는지 평가하는 것이 본 논문의 다음 단계 시뮬레이션이다. 이 시뮬레이션의 결과는 표 2의 결과에 정리하였다. 그리고 각 경우의 추적 결과 영상을 궤적과 함께 그림 3에 표시하였다.

표 2에서는 각 Re-ID 기법을 BoT-SORT 알고리즘에 적용한 다중 객체 추적 성능을 HOTA (Higher Order Tracking Accuracy), MOTA (Multiple Object Tracking Accuracy), 그리고 IDF1 (ID F1-score)의 세 가지 주요 지표로 평가하였다. HOTA는 검출, 연관, 위치 정확도를 종합적으로 평가하는 지표로, 전반적인 추적 성능을 나타낸다. MOTA는 거짓 양성, 거짓 음성, ID 스위치를 고려한 추적 정확도를 측정하며, IDF1은 추적의 일관성을 평가하는 지표이다. 이러한 다양한 측면의 평가를 통해 각 Re-ID 기법의 종합적인 추적 성능을 분석할 수 있다. 분석 결과, OSNet-AIN을 적용한 BoT-SORT가 HOTA 34.8, MOTA 29.1, IDF1 39.8로 모든 지표에서 가장 우수한 성능을 보였다. OSNet과 MobileNetV2를 적용한 경우도 유사한 순위로 성능

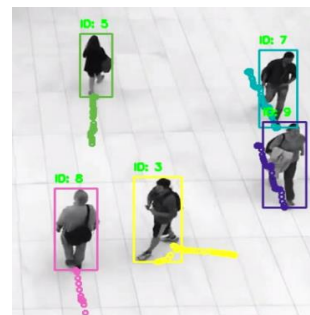
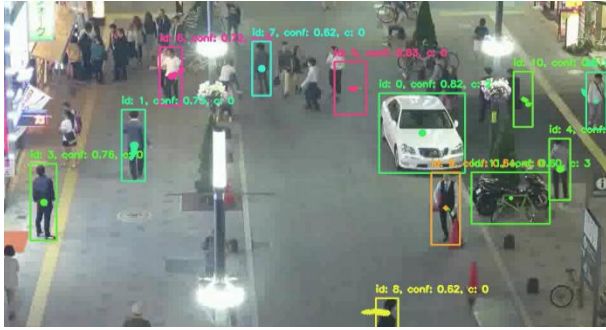
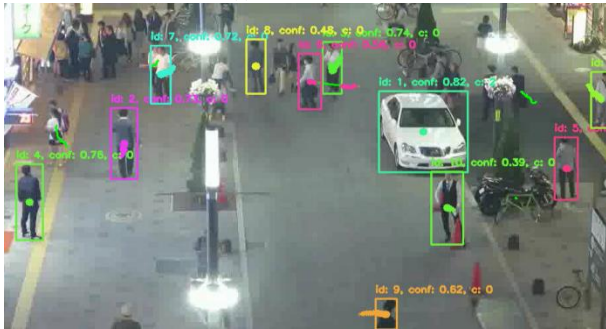


그림 2. 다시점 영상 정보 기반 다중 객체 추적 결과.
Fig. 2. Multi-object tracking results based on multi-view video information.

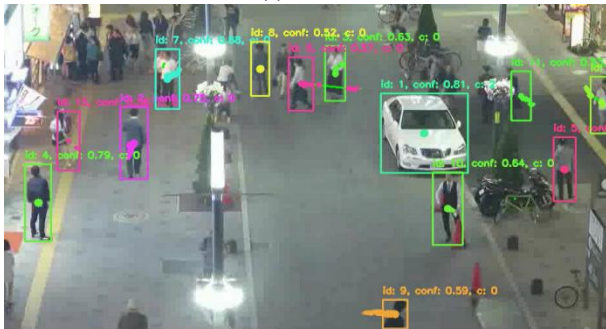
차이를 나타냈으며, 특히 OSNet-AIN은 MobileNetV2 대비 HOTA에서 0.9, MOTA에서 0.3, IDF1에서 1.7의 향상을 보여주었다. 이는 OSNet-AIN의 효과적인 특징 추출 능력이 복잡한 다중 객체 추적 환경에서도 일관되게 우수한 성능을 제공할 수 있음을 시사한다.



(a) MobileNetV2.



(b) OSNet.



(c) OSNet-AIN.

그림 3. Re-ID 기법들을 적용한 다중 객체 추적 결과.
Fig. 3. Multi-object tracking results applying Re-ID techniques.

표 1. Re-ID 기법 성능 비교.

Table 1. Performance comparison of Re-ID techniques.

기법	파라미터수	R1	mAP
MobileNetV2	2.2 M	50.9	27.0
OSNet	2.2 M	71.0	43.3
OSNet-AIN	2.2 M	79.1	55.1

표 2. Re-ID 기법 적용 후 BoT-SORT 추적 결과.

Table 2. BoT-SORT tracking results after applying Re-ID techniques.

기법	HOTA	MOTA	IDF1
MobileNetV2	33.9	28.8	38.1
OSNet	34.0	28.9	38.5
OSNet-AIN	34.8	29.1	39.8

V. 결론

본 논문에서는 다중 카메라의 다시점 환경에서 객체를 감지하고 추적하는 방법을 YOLOv9과 BoT-SORT를 사용하여 구현하였다. 특히, Re-ID 기법의 성능 향상이 전체 추적 시스템에 미치는 영향을 중점적으로 분석하였다.

먼저, MSMT17 데이터셋을 사용하여 MobileNetV2, OSNet, OSNet-AIN의 세 가지 Re-ID 기법의 성능을 비교하였다. 모든 모델이 2.2M의 동일한 파라미터 수를 가짐에도 불구하고, OSNet-AIN이 Rank-1 정확도와 mAP에서 가장 우수한 성능을 보였다. 이어서 이 Re-ID 기법들을 BoT-SORT 알고리즘에 적용하여 다중 객체 추적 성능을 평가하였다. HOTA, MOTA, IDF1의 세 가지 주요 지표를 사용하여 평가한 결과, OSNet-AIN을 적용한 BoT-SORT가 모든 지표에서 최고 성능을 달성하였다.

이러한 결과는 효과적인 Re-ID 기법의 중요성을 강조한다. OSNet-AIN의 우수한 성능은 복잡한 실제 환경에서의 객체 추적에 크게 기여할 수 있음을 시사한다. 특히, 동일한 파라미터 수를 가진 모델들 간의 성능 차이는 네트워크 구조 최적화의 중요성을 부각시킨다. 향후 연구에서는 OSNet-AIN의 구조를 더욱 최적화하거나, 다양한 환경에서의 성능 평가를 통해 이 접근 방식의 일반화 능력을 검증할 필요가 있다. 또한, 실시간 처리 능력 향상과 더 복잡한 시나리오에서의 성능 개선을 위한 연구가 필요할 것이다.

REFERENCES

- [1] S. Y. Kim and C. H. Kang, "YOLOv9 and deepsort algorithm for multiple object detection and tracking," *Proc. of ICROS 2024 Conference (in Korean)*, Daejeon, Korea, pp. 874-875, July 2024.
- [2] S. Y. Kim and C. H. Kang, "Adversarial Auto-encoder Based Preprocessing Algorithm for Improving Image Identification and Navigation Accuracy," *Journal of Institute of Control, Robotics and Systems (in Korean)*, vol. 28, no. 11, pp. 999-1005, 2022.
- [3] C. H. Kang and S. Y. Kim, "Comparative evaluation of physics-informed, hamiltonian, and lagrangian neural networks for 2D single target tracking," *Journal of Institute of Control, Robotics and Systems (in Korean)*, vol. 30, no. 3, pp. 288-293, 2024.
- [4] 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, 2017.
- [5] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," *Proc. of the IEEE International Conference on Computer Vision (ICCV)*, pp. 2961-2969, 2017.
- [6] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," *Advances in Neural Information Processing Systems*, pp. 3104-3112, 2014.
- [7] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735-1780, 1997.
- [8] G. Koch, R. Zemel, and R. Salakhutdinov, "Siamese neural networks for one-shot image recognition," *ICML Deep Learning Workshop*, vol. 2, 2015.
- [9] C. Y. Wang, I. H. Yeh, and H. Y. M. Liao, "YOLOv9: Learning what you want to learn using programmable gradient information," *arXiv preprint arXiv:2402.13616*, 2024.

- [10] Y. Zhang, J. Sun, Y. Xie, L. Qiao, Z. Chi, and X. Du, "BoT-SORT: Robust associations multi-pedestrian tracking," arXiv preprint arXiv:2206.14651, 2022.
- [11] K. Zhou, Y. Yang, A. Cavallaro, and T. Xiang, "Learning generalisable omni-scale representations for person re-identification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 8, pp. 2766-2780, 2021.
- [12] K. Zhou, Y. Yang, A. Cavallaro, and T. Xiang, "Omni-Scale feature learning for person re-identification," *Proc. of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 3702-3712, 2019.
- [13] Y. Zhang, P. Sun, Y. Jiang, D. Yu, F. Weng, Z. Yuan, and X. Wang, "Bytetrack: Multi-object tracking by associating every detection box," *European Conference on Computer Vision*, pp. 1-21, 2022.
- [14] H. Zhang, C. Wu, Z. Zhang, Y. Zhu, H. Lin, Z. Zhang, Y. Sun, T. He, J. Mueller, R. Manmatha, M. Li, and A. Smola, "Resnest: Split-attention networks," *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2736-2746, 2022.



김선영

1999년 국민대학교 전자공학과 졸업. 2015년 서울대학교 공과대학 기계항공공학부 대학원 졸업(공학석사). 2019년 동 대학원 졸업(공학박사). 2019년~2020년 세종대학교 지능기전공학부 박사후 연구원. 2020년~2023년 국립군산대학교 기계공학부 조교수. 2023년~현재 국립군산대학교 기계공학부 부교수. 관심분야는 항법 시스템, 필터링, 자율주행, 딥러닝, 다중 센서 융합, 다중 목표물 추적, 목표물 검출 및 식별.



강창호

2009년 세종대학교 기계항공우주공학과 졸업. 2016년 서울대학교 공과대학 기계항공공학부 대학원 졸업(공학박사). 2016년~2018년 BK21 플러스 서울대학교 융합지식기반 창조형 기계항공인재양성사업단 박사후 연구원. 2018년~2019년 고려대학교 공학연구소 연구교수. 2019년~2024년 국립금오공과대학교 기계시스템공학과(항공기계전자융합전공) 조교수. 2024년~현재 세종대학교 AI로봇학과 조교수. 관심분야는 항법 시스템, 딥러닝, 디지털 신호 처리, 비선형 필터링, 다중 목표물 추적 기법.