

## 폐플라스틱 재활용을 위한 최신 객체 탐지 모델의 성능 분석: Swin transformer와 비전 모델 평가

김도은 · 손준혁 · 김도경 · 안유찬<sup>†</sup>

계명대학교 화학공학과  
42601 대구광역시 달서구 달구벌대로 1095  
(2025년 9월 10일 접수, 2025년 11월 4일 수정본 접수, 2025년 11월 27일 채택)

### Performance Analysis of State-of-the-Art Object Detection Models for Plastic Waste Recycling: Evaluation of Swin Transformer and Vision Models

Doeun Kim, Junhyeok Son, Dogyeong Kim and Yuchan Ahn<sup>†</sup>

Department of Chemical Engineering, Keimyung University, 1095 Dalgubeol-daero, Dalseo-gu, Daegu, 42601, Korea  
(Received 10 September 2025; Received in revised from 4 November 2025; Accepted 27 November 2025)

#### 요 약

본 연구는 폐플라스틱 자동 분류를 위한 최신 인공지능 객체 탐지 모델의 성능을 평가하였다. Faster R-CNN, YOLOv8, YOLOv11, Swin Transformer 총 4개 모델을 대상으로 PET, PS, PP, PE 4종의 폐플라스틱 이미지 48,000개로 학습 및 테스트를 수행하였다. Optuna를 활용하여 하이퍼파라미터를 최적화하고, 정확도, 정밀도, 재현율, F1 score, 평균 정밀도 평균값(mAP), 추론 시간 등으로 성능을 평가하였다. Swin Transformer는 최고 정확도(0.988)와 mAP(0.988)를 기록하며 복잡하거나 오염된 플라스틱에서도 우수한 분류 성능을 나타냈다. YOLOv11은 가장 빠른 추론 속도(61.67 ms/이미지)를 보이며 실시간 처리 환경에 적합함을 확인하였다. 연구 결과, 모델 선택 시 정확도와 처리 속도 간의 트레이드오프가 존재함을 확인하였으며, 폐플라스틱 재활용을 위한 AI 기반 분류 시스템 구축에 실질적인 참고 자료를 제공한다.

**Abstract** – This study evaluates the performance of state-of-the-art AI-based object detection models for automated plastic waste classification. Four models—Faster R-CNN, YOLOv8, YOLOv11, and Swin Transformer—were trained and tested on a dataset of 48,000 images covering four plastic types (PET, PS, PP, PE). Hyperparameters were optimized using Optuna, and performance was assessed via accuracy, precision, recall, F1 score, mean Average Precision (mAP), and inference time. Swin Transformer achieved the highest accuracy (0.988) and mAP (0.988), demonstrating superior classification capability, particularly for complex or contaminated plastics. YOLOv11 exhibited the fastest inference speed (61.67 ms/image) with competitive accuracy, highlighting its suitability for real-time applications. These results reveal a trade-off between accuracy and processing speed, providing guidance for model selection based on application requirements. This study offers quantitative benchmarks and practical insights for deploying AI-driven plastic recycling systems.

Key words: Artificial intelligence, Object detection, Deep learning, Plastic waste, Swin transformer

#### 1. 서 론

##### 1-1. 연구 배경

플라스틱은 경제성, 다용성, 내구성 등의 특성으로 인해 급격한 산업 발전과 도시화 과정에서 다양한 분야에 폭넓게 사용되어 왔다. 전 세계 플라스틱 생산량은 연간 약 3억 5,900만 톤에 달하며, 이에

따른 폐기물 배출도 지속적으로 증가하고 있다[1]. 폐플라스틱은 자연환경에서 쉽게 분해되지 않고, 미세 플라스틱 형태로 공기, 수질, 토양 등 생태계 전반에 확산되어 환경오염과 생물 피해를 유발한다[1,2]. 따라서 환경 부담을 줄이고 자원 순환을 촉진하기 위한 효과적인 폐기물 관리 전략의 수립이 시급하며, 이중 재활용은 매립이나 소각보다 친환경적이고 지속 가능한 대안으로 주목받고 있다[3]. 폐플라스틱의 재활용 효율을 높이기 위해서는 다양한 재질과 오염 상태를 가진 폐기물을 정확히 분류하는 과정이 필수적이다[4]. 그러나 소비 후 배출되는 폐플라스틱은 폴리에틸렌 테레프탈레이트(Polyethylene Terephthalate, PET), 폴리프로필렌(Polypropylene, PP), 폴리스티렌(Polystyrene, PS), 폴리에틸렌(Polyethylene, PE) 등

<sup>†</sup>To whom correspondence should be addressed.

E-mail: yuchan.ahn@kmu.ac.kr

This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

서로 다른 물성의 재질이 혼합되어 배출되며, 이질적인 재질의 혼합은 재활용 원료의 품질 저하와 재활용 효율 감소를 초래한다[5]. 전통적으로는 작업자가 수작업으로 분류를 수행해왔으나, 이는 시간과 비용 소모가 크고, 미세플라스틱 노출에 따른 작업자 건강 위험도 존재한다[6]. 실제로 많은 재활용 시설에서는 여전히 인력 중심 분류에 의존하고 있는 현실이다[7].

### 1-2. 기존 분류 기술과 한계

폐플라스틱 분류를 자동화하기 위한 기술로는 부유선별, 광학 분류기를 사용하는 근적외선 Near-Infrared (NIR), 가시광선 등 센서 기반 분류 등이 활용된다[8]. NIR 기반 광학 분류기는 플라스틱 재질별로 서로 다른 스펙트럼 특성을 측정하여 재질을 식별하며 원격 및 빠른 측정과 같은 여러 장점으로 인해 NIR 기반 분류기는 다양한 형태의 폐플라스틱을 빠르고 효율적으로 식별, 선별하는 데 활용되고 있으며 대량 처리와 자동화 공정에 적합한 핵심 기술로 자리 잡고 있다[9]. 그러나 복합 재료, 오염물, 투명 필름 등 실제 폐기물의 다양성과 복잡성으로 인해 분류 정확도에는 한계가 있다. 예를 들어, 투명 PET와 투명 PP는 NIR 센서에서 스펙트럼이 유사해 혼동되기 쉽고[9], 색소나 라벨 오염은 분류 성능 저하를 유발한다[10]. 최근 늘어난 다층 복합 필름과 바이오 기반 소재는 기존 센서 기술로 구분이 어렵고[11], 단일 기준 분류법은 오분류율 증가 문제를 안고 있다. 이러한 분류 오류는 재활용 공정 전체의 품질 저하와 경제성 악화로 이어지며, 지속 가능한 자원 순환 체계의 불안정성을 초래할 수 있다. 따라서 분류의 정확도와 처리 효율을 모두 개선할 수 있는, 신속하고 반복성이 뛰어난 자동 분류 시스템 도입이 필요하다[12]. 최근 주목받는 인공지능 기술은 시각적 특징 추출과 재질 감별, 고속 처리에서 기존 기술의 한계를 극복할 수 있는 효과적인 대안이다[12]. 인공지능 기반 분류 시스템은 분류 정확도 향상뿐 아니라 작업 속도 증가, 인건비 절감, 작업자 안전 확보 등 재활용 공정의 지속 가능성 제고에 크게 기여할 수 있다[8,12].

### 1-3. 인공지능 기반 분류 기술의 발전

특히 딥러닝 기반 이미지 인식 기술은 폐플라스틱 분류 공정에서 핵심 기술로 자리잡았다[3,13,14]. 초기 전통적 기계학습은 폐플라스틱의 다양한 형태와 물성을 정밀하게 인식하는 데 한계가 있었으나, 합성곱 신경망(Convolutional Neural Network, CNN) 기반 모델의 도입으로 분류 정확도가 크게 향상되었다[15]. ResNet, VGGNet, DenseNet, MobileNet 등 다양한 CNN 모델들이 개발되었고[16,17], Faster Region-based CNN (Faster R-CNN), Mask Region-based CNN (Mask R-CNN), You Only Look Once (YOLO) 시리즈 등 객체 탐지 기술은 분류와 위치 추정을 통합하여 실시간 처리 가능성을 확대하였다[2,18,19]. 최근 Transformer 기반 모델인 Shifted Window Transformer (Swin Transformer)는 전역 문맥 정보를 활용해 복잡한 폐플라스틱 분류 문제에서 우수한 성능을 발휘하고 있다[20]. Transformer 기반의 객체 탐지 및 분류 모델들은 기존에 구분하기 어려웠던 오염되거나 불투명한 재질, 유사 색상 재질도 높은 정확도로 분류하고 있으며[21], 실제 폐기물 처리 현장에 적용 가능한 94~99%의 높은 정확도와 실시간 처리 성능을 보이며 산업 현장 적용 가능성을 입증하고 있다[22].

### 1-4. 연구 동향 및 한계

전이 학습 기반의 모델인 ResNet-50, DenseNet169, MobileNet\_v2

등은 WaDaBa, TrashNet, MixTrash 등 다양한 데이터셋에서 94~98% 이상의 높은 정확도를 기록하며, 복잡한 폐기물 환경에서도 기존 방법 대비 명확한 성능 향상을 보여주고 있다[22-24]. 이러한 AI 분류 시스템은 단순 자동화를 넘어 처리 속도 향상, 인건비 절감, 자원 회수율 증가 등 전방위적 효율성 개선을 가능하게 한다[25]. 그러나 대부분 연구가 개별 모델의 정성적 분석이나 정확도 중심 평가에 머무르고, 다양한 모델을 동일 조건에서 비교·분석한 연구는 부족하다. 실제 응용에서는 정확도뿐 아니라 추론 속도, 모델 경량성, 다양한 재질 대응력, 일관성 등 복합 지표에 기반한 정량 평가가 필수적이다[26,27].

### 1-5. 연구 목적

비교, 분석하는 것을 목표로 한다. 이를 위해 전통적인 CNN 기반 모델(Faster R-CNN), YOLO 계열(YOLOv8, YOLOv11), 그리고 Transformer 기반 모델(Swin Transformer)을 대상으로 동일한 폐플라스틱 이미지 데이터셋을 활용한 학습 및 테스트를 수행한다. 성능 평가는 정확도(Accuracy), 정밀도(Precision), 재현율(Recall), F1 Score, 모델 파라미터 수 등의 다양한 지표를 통해 이뤄지며, 각 모델에 대해 하이퍼파라미터 튜닝을 통해 최적 조건을 도출하고 일반화 가능성과 안정성을 검토한다. 여러 지표를 바탕으로 합리적인 모델 선택 기준을 제시함으로써 폐플라스틱 분류 분야의 연구 및 기술 개발에 실질적인 참고 자료를 제공한다. 또한, 데이터 기반의 분석을 통해 각 모델의 적용 가능성과 한계를 탐색하는 데 중점을 두며, 향후 실제 응용을 위한 기초 자료를 제공하고자 한다.

## 2. 실험

데이터셋 구축 및 전처리, 모델 선정 및 구조 정의, 하이퍼파라미터 최적화, 모델 학습 및 검증, 성능 평가 지표 정의의 다섯 단계로 구성된다. 먼저, AI Hub에서 제공하는 폐플라스틱 이미지 데이터셋을 활용하여 학습·검증·테스트 데이터를 구성하였다. 이후 Faster Region-based Convolutional Neural Network (Faster R-CNN), You Only Look Once version 8 (YOLOv8), You Only Look Once version 11 (YOLOv11), Shifted Window Transformer (Swin Transformer) 모델을 비교 대상으로 선정하였으며, 동일한 학습 환경에서 성능을 검증하였다. 또한, Optuna를 활용한 베이지안 최적화를 통해 각 모델별 하이퍼파라미터를 탐색하였고, 최종적으로 정확도, 정밀도, 재현율, F1-score, mean Average Precision (mAP) 지표를 활용하여 성능을 평가하였다.

### 2-1. 전체 실험 프로세스 개요

본 연구의 전체 실험 절차는 Fig. 1과 같다. 먼저 모델의 초기 성능을 확인하기 위하여 30 에포크(epoch)로 베이스라인 학습을 수행하였다. 이후 옵투나(Optuna)를 활용하여 하이퍼파라미터 탐색을 진행하였으며 각 트라이얼(trial)은 5 epoch씩 총 10회로 구성되었다. 이는 제한된 계산 자원 내에서 효율적인 탐색을 수행하기 위한 설정으로 전역 최적해보다는 상대적으로 우수한 파라미터 조합을 찾는 것을 목표로 하였다. 이 과정에서 성능이 낮은 trial은 3 epoch에서 조기 종료 커 탐색 효율을 높였다. 구체적으로, 본 연구에서는 Optuna의 조기 종료 코드를 사용하며 이전 trial들의 중간 성능값에 미치지 못하는 trial은 3 epoch 내에서 조기 중단시켰다. 그리고 “이는

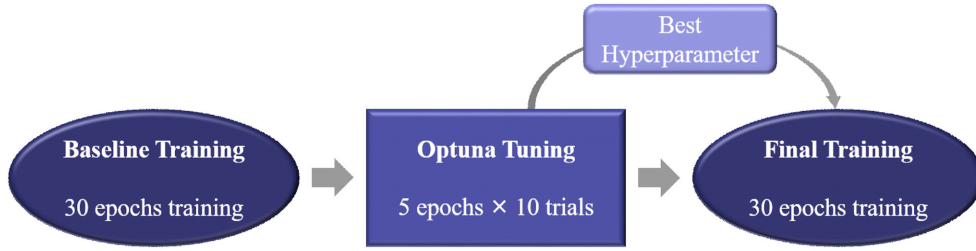


Fig. 1. Overview of the Experimental Process.

제한된 계산 자원 내에서 효율적인 탐색을 수행하기 위한 설정으로 전역 최적해보다는 상대적으로 우수한 파라미터 조합을 찾는 것을 목표로 하였다. 마지막으로 Optuna에서 도출된 최적의 하이퍼 파라미터를 적용하여 모델을 처음부터 다시 30 epoch 학습하여 최종 모델을 구축하였다.

2-2. 데이터셋

이 연구는 과학기술정보통신부의 재원으로 한국지능정보사회진흥원의 지원을 받아 구축된 ‘페플라스틱 이미지 데이터’를 활용하여 수행된 연구이며, 본 연구에 활용된 데이터는 AI 허브[28]에서 제공받았다. 데이터 셋은 총 48,000개의 이미지 형식 데이터로 구성되어 있으며 클래스는 PET, PS, PP, PE 총 4가지로 구성되어있다. 클래스 당 12,000개로 구성되어있다. 데이터 셋은 클래스 당 12,000개로 균등하게 구성되어 있어 클래스 불균형(class imbalance) 문제로 인한 성능 왜곡 가능성을 최소화하였다. 데이터는 학습, 검증, 테스트가 6 : 2 : 2 비율로 구성되어있으며 학습 데이터 28,800개, 검증 데이터 9,600개, 테스트 데이터 9,600개로 구성되어있다.

2-3. 실험 대상 모델

Faster R-CNN은 대표적인 Two-Stage Detector로, 먼저 Region Proposal Network (RPN)를 통해 CNN 특징 맵 위에서 객체가 있을 법한 후보 영역(region proposal)을 생성한다. 이후 각 후보 영역에 대해 ROI Pooling 또는 ROI Align을 적용하여 고정 크기의 특징 벡터를 얻고, 이를 기반으로 객체 클래스 분류와 바운딩 박스 회귀를 수행한다. 이러한 구조는 후보 영역 생성과 분류 과정을 하나의 네트워크에서 end-to-end로 학습할 수 있게 하여, 기존 R-CNN 대비 연산 속도를 크게 향상시키면서도 높은 정확도를 유지한다. 또한, RPN과 검출 네트워크가 공유하는 CNN 특징 맵 덕분에 메모리 효율과 학습 안정성이 개선되어, 다양한 크기와 형태의 객체를 정밀하게 탐지할 수 있다[29].

반면 YOLO (You Only Look Once) 계열은 One-Stage Detector로 분류되며, 입력 이미지를  $N \times N$  그리드로 나누고 각 셀에서 객체 존재 여부, 바운딩 박스 좌표, 클래스 확률을 동시에 예측한다. 이러한 방식은 후보 영역 생성 단계가 필요 없고, 전체 이미지를 한 번의 CNN 연산으로 처리할 수 있어 실시간 객체 탐지에 적합하다. YOLOv8은 기존 YOLO 구조를 개선하여 앵커 프리(anchor-free)

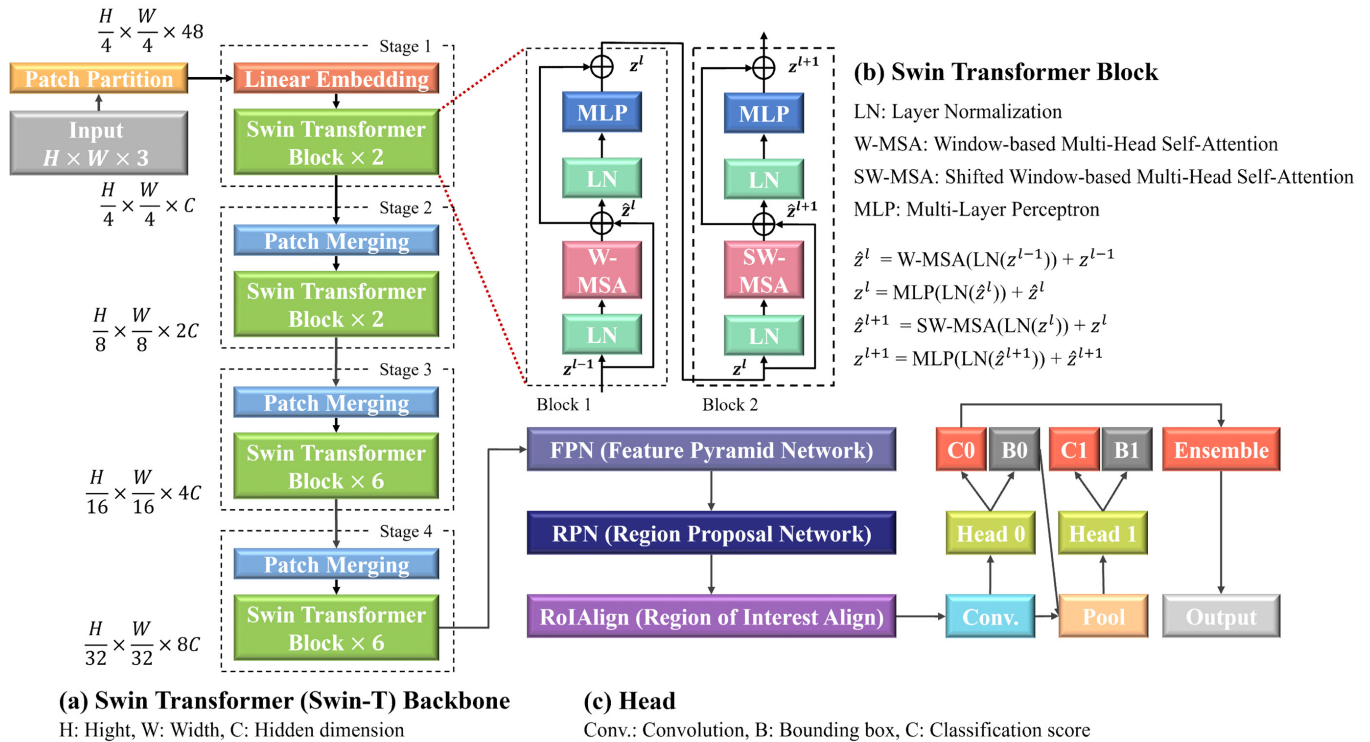


Fig. 2. Overall architecture of object detection framework based on swin transformer.

방식을 채택하고, 경량화된 CSPNet 변형과 C2f 모듈을 적용함으로써 연산 효율과 정확도를 동시에 높였다. 또한 Mosaic, MixUp 등의 데이터 증강 기법을 개선하고, 어텐션 메커니즘과 동적 합성곱을 활용하여 다양한 크기의 객체를 효과적으로 탐지할 수 있다. YOLOv11은 YOLOv8의 구조를 기반으로 C2PSA와 GELAN 등 향상된 레이어 구조와 최적화 기법을 추가하여 모델을 더욱 경량화하고 처리 속도를 개선하였다. 그 결과, 다양한 해상도의 입력에서도 안정적이고 일관된 탐지 성능을 유지할 수 있어, 실시간 탐지 환경뿐만 아니라 정확도가 중요한 응용에서도 강점을 가진다[30].

Fig. 2는 Swin Transformer 모델의 구조를 나타낸 그림이다. Swin Transformer는 Vision Transformer (ViT)의 구조를 개선하여 고해상도 이미지 처리에 효율적으로 적용할 수 있도록 계층적 비전 변환기(hierarchical vision transformer)이다. 기존 ViT는 이미지를 고정 크기의 패치로 분할하고 전역 self-attention을 수행하므로, 입력 크기가 커질수록 연산량이 제곱으로 증가하여 대규모 비전 작업에 비효율적이다. Swin Transformer는 이러한 한계를 극복하기 위해 로컬 윈도우 기반 self-attention과 윈도우 이동(Shifted Window) 기법을 적용하여 전역 문맥 정보를 점진적으로 통합한다. 또한 계층적으로 패치 병합(Patch Merging)을 수행해 해상도를 점차 축소하며 멀티스케일 특징을 추출할 수 있어, 복잡한 배경이나 작은 객체 탐지에서 강점을 보인다. 이러한 구조는 객체 탐지, 이미지 분류, 시맨틱 세그멘테이션 등 다양한 비전 과제에서 효율성과 정확도를 동시에 향상시키며, Faster R-CNN이나 YOLO 계열 모델의 백본으로 결합될 경우 장거리 의존성 캡처 능력과 표현력을 강화하여 탐지 성능을 극대화할 수 있다.

Swin Transformer는 총 네 단계로 구성된다. 먼저 입력 RGB 이미지를 중복되지 않는 패치로 분할한 후, 첫 번째 단계에서 선형 임베딩(Linear Embedding)을 수행하여 각 패치를 고정 크기의 벡터로 변환한다. 이는 분할된 이미지를 Transformer 학습에 적합한 형태로 매핑하는 과정이다. 변환된 데이터는 두 개의 블록으로 구성된 Swin Transformer 블록을 통과한다. 이후 2, 3, 4단계에서는 인접한 패치를 병합(Patch Merging)하여 해상도를 절반으로 축소하는 동시에 채널 수를 두 배로 증가시켜, 점진적으로 추상화된 특징 표현을 학습한다. 각 단계에서 Swin Transformer 블록의 개수는 항상 2의 배수로 구성되며, 예를 들어 Swin-T 모델에서는 Stage 1, 2, 4에 각각 2개의 블록이, Stage 3에는 6개의 블록이 배치된다[20].

Swin Transformer 블록은 기존 Transformer 블록의 멀티헤드 어텐션(Multi-Head Self-Attention, MSA) 모듈을 윈도우 기반 MSA(Window-based MSA)와 시프트 윈도우 기반 MSA(Shifted Window-based MSA)로 대체하고, 나머지 구성은 유지한다. 윈도우 기반 MSA는 전역 Self-Attention 대비 연산 복잡도를 선형적으로 줄여, 고해상도 입력에서도 효율적으로 동작한다. 블록은 레이어 정규화(Layer Normalization), MSA 모듈, GELU 활성화 함수를 포함한 MLP로 구성되며, 각 연산 후 Residual Connection을 적용하여 학습 안정성과 수렴 속도를 향상시킨다. 블록 1에서는 윈도우 기반 MSA를, 블록 2에서는 시프트 윈도우 기반 MSA를 사용하여 윈도우 경계에서의 정보 손실을 완화하고 패치 간 상호 연관성을 강화한다[20].

본 모델은 Swin Transformer를 백본으로 활용하여 입력 이미지로부터 다단계 특징 맵을 추출하고, 윈도우 기반 셀프 어텐션 메커니즘을 통해 전역 문맥 정보를 효과적으로 학습합니다. 추출된 특징 맵은 FPN을 통해 서로 다른 해상도의 특징이 융합되어 다양한

크기의 객체 탐지 성능을 향상시키며, 이는 작은 객체부터 큰 객체까지 포괄적인 탐지를 가능하게 합니다. 이후 Faster R-CNN 헤드에서는 RPN이 다양한 크기와 비율의 앵커 박스를 기반으로 후보 영역을 생성하고, RoIAlign을 통해 이 후보 영역의 특징을 정규화한 뒤 분류 및 회귀 브랜치에서 각각 객체 클래스 확률 예측과 바운딩 박스 좌표 보정을 수행하여 최종적으로 객체의 위치와 종류를 정밀하게 탐지한다.

본 연구는 폐플라스틱 분류 정확도 및 처리 효율 향상을 목표로, Swin Transformer 아키텍처를 기반으로 진행되었다. 특히, Swin Transformer 논문에서 제안된 Shifted-window 기법은 기존 Vision Transformer의 단점인 고해상도 이미지 처리의 비효율성을 개선하고 계층적 특징을 효과적으로 추출하는 데 유리하다고 판단하였다. 이러한 이유로, 복잡한 폐플라스틱 데이터를 처리하는 본 연구의 목적을 달성하는 데 가장 적합한 모델로 Swin Transformer를 채택하였다. 본 연구에서 실험 대상 모델인 ‘Swin Transformer’는 구체적으로 Swin Transformer를 백본(Backbone)으로, Faster R-CNN의 헤드를 결합한 형태를 사용하였으며, 본문 및 표, 그림에서는 편의상 ‘Swin Transformer’로 축약하여 표기하였음을 밝힌다.

#### 2-4. 학습 환경 및 설정

본 연구는 Windows 11 운영체제에서 수행되었으며 사용한 CPU 환경은 AMD Ryzen 7 7800X3D 8-Core Processor, NVIDIA GeForce RTX 4070 Ti GPU 그리고 32GB DDR4 RAM (5200MT/s, 2/4 슬롯 사용)에서 진행되었다.

모델은 Python 기반의 오픈 소스 딥러닝 라이브러리인 PyTorch를 주된 프레임워크로 사용하였다. PyTorch는 자동 미분 기능을 제공하여 모델 학습을 용이하게 하며, GPU 가속을 지원하여 높은 성능을 제공한다. Torchvision은 PyTorch에 통합된 기계 비전 패키지로, 이미지 처리 알고리즘 및 다양한 사전 학습된 모델을 제공한다.

#### 2-5. 하이퍼파라미터 튜닝

하이퍼파라미터는 기계학습 모델 훈련을 관리하는 데 사용하는 외부 구성 변수로 모델을 훈련하기 전에 수동으로 설정한다. 하이퍼파라미터는 모델 구조, 기능 및 성능을 직접 제어하기 때문에 튜닝을 통해 최적의 결과를 위한 모델의 성능 조정이 가능하다. 하이퍼파라미터 튜닝은 기계 학습의 필수적인 부분이며 적절한 파라미터 값을 설정하는 것이 중요하다. 하지만 적절한 하이퍼파라미터를 선택하기 위해 설정된 규칙이나 최적 또는 기본값은 없기 때문에 실험을 통해서 적절한 하이퍼파라미터를 설정이 필요하다.

Table 1은 Optuna 기반 하이퍼파라미터 튜닝을 위해 각 모델별 주요 하이퍼파라미터와 그 탐색 범위를 제시한 것이다. 이러한 하이퍼파라미터들은 각 모델의 성능에 큰 영향을 미치는 요소로, 최적의 조합을 찾기 위해 Optuna 프레임워크를 활용하여 탐색하였다. Faster R-CNN, YOLOv8, YOLOv11, Swin Transformer 모델에 대해 학습률(learning rate), 가중치 감쇠(weight decay), 배치 크기(batch size), 드롭아웃 비율(dropout rate), 모멘텀(momentum) 등 핵심적인 하이퍼파라미터를 설정하였으며, 모델 특성에 따라 Box Loss Gain, Classification Loss Gain, Mosaic 등의 추가 항목도 포함하였다. 각 모델별로 설정된 하이퍼파라미터의 탐색 범위는 문헌 및 사전 실험 결과를 참고하여 설정하였으며, 보다 효과적인 성능 향상을 유도할 수 있도록 설계되었다.

**Table 1. Optuna hyperparameter ranges for each detection model**

Models	Hyperparameter	Range
Faster R-CNN [31]	Learning rate	1e-4 - 5e-4
	Weight decay	1e-5 - 5e-5
	Batch	2, 4, 8
	Drop out rate	0.3 - 0.4
YOLO v8 [32, 33]	Initial learning rate	1e-5 - 1e-1
	Final learning rate	0.01 - 1.0
	Weight decay	1e-7 - 0.001
	Batch	4, 8, 16, 32, 64
	Momentum	0.6 - 0.98
	Box Loss Gain	0.02 - 0.2
	Classification Loss Gain	0.2 - 4.0
	Mosaic	0.0 - 1.0
YOLO v11 [32, 33, 34]	Initial learning rate	1e-5 - 1e-1
	Final learning rate	0.01 - 1.0
	Weight decay	1e-7 - 0.001
	Batch	4, 8, 16, 32, 64
	Momentum	0.6 - 0.98
	Box Loss Gain	0.02 - 0.2
	Classification Loss Gain	0.2 - 4.0
	Mosaic	0.0 - 1.0
Swin Transformer [35]	Learning rate	1e-4 - 5e-4
	Weight decay	1e-4 - 1e-2
	Batch	2, 4, 8
	Drop out rate	0.3 - 0.5

**2-6. 성능 평가 지표**

본 연구는 객체 탐지 기술을 활용하여 폐플라스틱을 분류하는 것을 목표로 하며, 탐지된 객체의 존재 유무와 더불어 정확한 클래스 식별을 중요하게 고려한다. 따라서 모델의 성능 평가는 탐지 정확도와 분류 정확도를 종합적으로 반영할 수 있는 정확도, 정밀도, 재현율, F1 score mAP, 추론시간을 사용한다.

혼동 행렬은 분류 모델의 성능을 시각적으로 평가하는 데 사용되는 도구이다. 이 행렬은 네 가지 주요 요소로 구성되며, 모델의 다양한 성능 지표 계산에 필수적이다. 특히 다중 클래스 분류 문제에서 유용하게 활용된다. 혼동행렬은 Table 2와 같이 구성되어있다.

**Table 2. Confusion matrix**

		Predicted Class	
		True	False
Actual Class	True	TP (True Positive)	FN (False Negative)
	False	FP (False Positive)	TN (True Negative)

- TP (True Positive): 실제 정답이 참인 데이터를 모델이 참이라고 예측한 경우
- FP (False Positive): 실제 정답이 거짓인 데이터를 모델이 참이라고 예측한 경우
- FN (False Negative): 실제 정답이 참인 데이터를 모델이 거짓이라고 예측한 경우
- TN (True Negative): 실제 정답이 거짓인 데이터를 모델이 거짓이라고 예측한 경우

**2-6-1. 정확도(Accuracy)**

정확도는 전체 데이터 중에서 모델이 올바르게 예측한 비율이다. 전체 예측 중 TP와 TN의 비율로 정확도는 1에 가까울수록 좋다.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

식 (1)을 보면 정확도는 모델이 올바르게 예측한 경우(TP+TN)를 전체 데이터(TP+TN+FP+FN)로 나눈 값이다. 정확도는 이해하기 쉽고 계산이 간단하지만 클래스 불균형이 심할 경우 정확도만으로는 모델 성능을 정확히 평가하기 어렵다. 따라서 좋은 모델인지 판단하기 위해서는 정밀도와 재현율을 함께 보는 것이 좋다.

**2-6-2. 정밀도(Precision)**

모델이 참이라고 예측한 데이터 중 실제 참인 것의 비율을 의미한다. 정확도와 유사하게 1에 가까울수록 좋다.

$$Precision = \frac{TP}{TP+FP} \tag{2}$$

정밀도 수식은 식 (2)와 같다. 모델이 참이라고 예측한 데이터(TP+FP) 중에서 실제 정답이 데이터(TP)의 비율을 나타낸다. 정밀도는 모델이 거짓을 참으로 잘못 예측한 경우를 줄이는 데 중점을 둔 지표이다.

**2-6-3. 재현율(Recall)**

재현율은 모델이 참을 찾을 수 있는 능력을 나타낸다. 실제 참인 데이터 중에서 모델이 참으로 정확히 예측한 비율을 의미한다. 재현율도 1에 가까울수록 좋지만 라벨에 따라 다른 성능을 보일 수 있다.

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

재현율은 식 (3)을 보면 실제 참인 데이터(TP+FN) 중에서 모델이 올바르게 예측한 데이터(TP)의 비율을 나타낸다. 정밀도와 재현율은 모두 TP를 높이는 데 초점을 맞추지만 재현율은 FN을 낮추고 정밀도는 FP를 낮추는데 초점을 맞춘다. 따라서 재현율과 정밀도는 서로 보완적인 지표로 분류 성능을 평가하는데 사용된다. 이때 어느 한쪽을 강제로 높이면 다른 하나의 수치는 떨어진다. 따라서 특정 응용에서 무엇이 더 중요한지에 따라 정밀도와 재현율 사이의 균형을 맞추는 것이 중요하다.

**2-6-4. F1 score**

F1 score는 정밀도와 재현율의 조화 평균으로 두 지표간의 균형을 나타낸다. 분류 대상 데이터 분포가 고르지 못한 경우 학습 데이터 분포에 따라 정확도 수치가 왜곡될 수 있어 F1 점수를 사용한다. F1 score는 데이터 값들이 서로 상호 의존적인 경우나 비율을 나타내는 경우에 유용하다.

$$F1\ score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{4}$$

F1 score는 정밀도와 재현율을 사용하여 계산한다. 식 (4)를 보면 계산에서 산술 평균이 아닌 조화 평균을 사용하여 두 지표의 균형을 객관적으로 평가한다. F1 score는 정밀도와 재현율이 모두 높을

때 높은 값을 가진다. 특히, 한쪽 지표가 낮은 경우 전체 점수도 낮아지므로, 두 지표를 모두 중요하게 고려해야 할 때 유용한 지표이다.

#### 2-6-5. 평균 정밀도 평균값(mean Average Precision, mAP)

mAP는 객체 탐지 또는 분류 모델의 전반적인 성능을 평가하기 위해, 각 클래스별 평균 정밀도(Average Precision, AP)를 계산한 뒤 그 평균을 구한 값이다. PR (Precision-Recall) 그래프는 알고리즘의 성능을 시각적으로 파악하는 데 유용하지만, 서로 다른 알고리즘 간 성능을 정량적으로 비교하기에는 한계가 있다. 따라서 클래스별 AP의 평균인 mAP를 보완 지표로 사용한다.

$$AP = \int_0^1 P(r)dr \quad (5)$$

식 (5)는 AP를 계산하는 식이다. AP는 인식 모델의 성능을 하나의 값으로 표현한 것으로 PR 그래프의 면적으로 계산된다. AP가 높을수록 모델의 성능이 우수하다는 것을 의미한다. 위 수식은 PR 곡선 아래 면적을 계산하는 적분식으로 P(r)은 재현율 r에 대한 정밀도의 함수이다.

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i \quad (6)$$

식 (6)은 mAP를 계산하는 식이다. mAP는 물체 클래스가 여러 개인 경우 각 클래스 당 AP를 구한 다음 AP의 평균을 구하여 성능을 평가한다. N은 전체 클래스의 개수이고 AP<sub>i</sub>는 i번째 클래스의 AP 값이다. mAP는 다중 클래스 환경에서 모델의 탐지 및 분류 성능을 종합적으로 평가하는 가장 대표적인 지표로 다양한 클래스를 정확히 탐지하고 분류해야 하는 객체 탐지 모델의 성능을 평가할 때 필수적으로 사용된다.

#### 2-6-6. 추론 시간(Inference time)

추론 시간은 모델이 한 개의 입력 이미지를 처리하는 데 소요되는 평균 시간이다.

$$\text{Inference time} = \frac{\text{Total Inference Time}}{\text{Number of Test Sample}} \quad (7)$$

식 (7)처럼 전체 테스트 데이터셋에 대한 추론 시간을 이미지 수로 나누어 계산하여 모델의 단일 이미지 처리 속도를 객관적으로 평가할 수 있다. 이 지표를 통해서 서로 다른 모델의 추론 속도를 비교하고 실시간 처리에 모델을 적용할 수 있는지 가능성을 평가할 수 있다. 또한 모델의 경량화, 최적화, 하드웨어 선택 등에서 효율성을 판단하는 지표로 활용된다.

### 3. 결과 및 고찰

본 연구에서는 폐플라스틱 자동 분류를 위해 적용한 네 가지 객체 탐지 모델 Faster R-CNN, YOLOv8, YOLOv11, Swin Transformer의 성능을 비교한 결과를 제시한다. 모든 모델은 동일한 환경과 조건에서 학습되었으며 Accuracy, Precision, Recall, F1 score, mAP, 학습 시간 등의 지표를 기준으로 평가되었다. 특히 Optuna 기반 하이퍼파라미터 최적화를 통해 각 모델의 성능을 최대화한 후 비교하여 모델별 성능 차이와 특성을 명확히 확인할 수 있었다.

#### 3-1. 모델 튜닝 결과

본 연구에서는 Optuna를 활용하여 Faster R-CNN, YOLOv8, YOLOv11, Swin Transformer 모델에 대한 하이퍼파라미터 최적화를 수행하였다. 최적의 하이퍼파라미터를 설정하기 위해서 Optuna의 병렬 좌표 그래프(Parallel Coordinates Plot) 기능을 사용하였다. Parallel Coordinates Plot에서 선이 많이 수렴되고 선의 색이 진할수록 더 좋은 성능을 나타낸다.

Fig. 3은 각 모델별 탐색 과정에서의 하이퍼파라미터 값과 목적 함수(Objective Value)의 관계를 나타낸 평행좌표 그래프이다. 탐색 횟수는 제한적이었으나, 이를 통해 학습률(learning rate), 배치 크기(batch size), 가중치 감쇠(weight decay), 드롭아웃 비율(dropout rate) 등의 조합이 성능에 미치는 영향을 확인할 수 있었다. 특히, 일부 파라미터에서는 뚜렷한 성능 향상 구간이 확인되었으며, 이를 기반으로 상대적으로 우수한 파라미터 조합을 최적값으로 설정하였다. 하지만 Swin Transformer의 경우 탐색 횟수가 충분하지 않아 전체적인 수렴은 관찰되지 않았지만, 제한된 탐색 내에서는 상대적으로 우수한 파라미터 조합을 확인할 수 있었다.

Table 3은 본 연구에서 수행한 하이퍼파라미터 튜닝 결과를 정리한 것이다. Faster R-CNN의 경우 Learning rate는  $3.50 \times 10^{-4}$ , Weight Decay는  $1.40 \times 10^{-5}$ , Batch는 2, Drop out rate는 0.354로 설정되었다. YOLO v8 모델은 Initial learning rate  $1.90 \times 10^{-5}$ , Final learning rate 0.372, Weight Decay  $9.00 \times 10^{-9}$ , Batch 4, Momentum 0.866 등의 값이 최적의 하이퍼파라미터이며, Box Loss Gain과 Classification Loss Gain은 각각 0.193과 1.079로 조정되었다. YOLO v11 모델은 Initial learning rate 0.070, Final learning rate 0.037, Weight Decay  $7.0 \times 10^{-9}$ , Batch 16, Momentum 0.672의 값이 선택되었고, 추가적으로 Box Loss Gain은 0.028, Classification Loss Gain은 1.045, Mosaic 값은 0.857로 설정되었다. 마지막으로 Swin Transformer 모델은 Learning rate  $1.36 \times 10^{-5}$ , Weight Decay  $5.63 \times 10^{-3}$ , Batch 4, Drop out rate 0.311에서 최적의 성능을 보였다. 이러한 결과는 각 모델별로 최적의 하이퍼파라미터 조합이 차이가 있음을 보여주며, 모델 특성에 따라 학습 전략이 달라져야 함을 보여준다.

#### 3-2. 모델 학습 결과

모델의 분류 성능을 보다 세부적으로 분석하기 위해 혼동행렬(confusion matrix)을 활용하였다. 혼동행렬은 실제 클래스와 예측 클래스 간의 분포를 시각적으로 나타내어, 각 모델이 특정 클래스에서 보이는 강점과 약점을 직관적으로 확인할 수 있는 도구이다. 본 연구에서는 Faster R-CNN, Swin Transformer, YOLOv8, YOLOv11 모델의 예측 결과를 바탕으로 혼동행렬을 도출하였으며, 이를 통해 클래스 간 오분류 양상을 비교·분석하였다.

Fig. 4는 총 9,600개의 테스트 이미지를 기반으로 작성되었습다. 테스트셋은 클래스당 2,400개의 이미지로 구성되어 있으며 각 이미지에는 3개의 객체가 존재하므로 클래스당 약 7,200개로 총 28,800개에 근접하는 객체를 평가하였다. Fig. 4는 각 모델별 튜닝 후 혼동행렬의 그림이다. 혼동 행렬 결과에 따르면, Faster R-CNN은 전반적으로 안정적인 성능을 보였으나 PP와 PS 클래스 간 오분류가 다수 발생하였다. YOLOv8과 YOLOv11은 PET와 PS 클래스 간 혼동이 일부 관찰되었으나, PP와 PE는 거의 완벽하게 분류하였다. Swin Transformer의 경우 전반적으로 모든 클래스에서 높은 분류 성능을 나타내며, 일부 PET과 PS 간의 혼동을 제외하면 가장

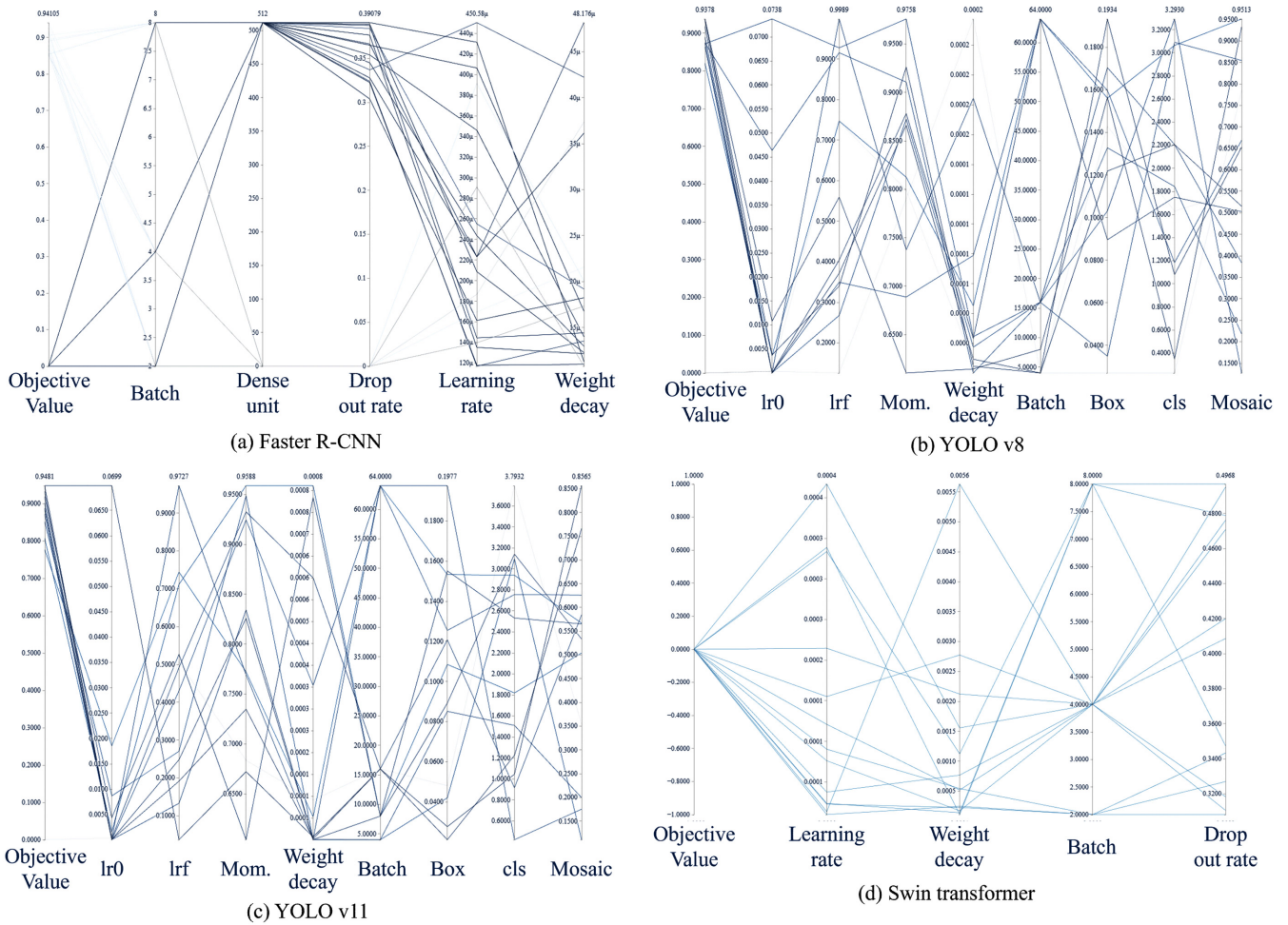


Fig. 3. Optuna-based parallel coordinates plot for hyperparameter tuning.

Table 3. Best-performing hyperparameters for each detection model

Models	Hyperparameter	Range
Faster R-CNN	Learning rate	$3.50 \times 10^{-4}$
	Weight decay	$1.40 \times 10^{-5}$
	Batch	2
	Drop out rate	0.354
YOLO v8	Initial learning rate	0.091
	Final learning rate	0.032
	Weight decay	$9.00 \times 10^{-6}$
	Batch	4
	Momentum	0.866
	Box Loss Gain	0.193
	Classification Loss Gain	1.079
	Mosaic	0.653
YOLO v11	Initial learning rate	0.070
	Final learning rate	0.037
	Weight decay	$7.00 \times 10^{-6}$
	Batch	16
	Momentum	0.672
	Box Loss Gain	0.028
	Classification Loss Gain	1.045
	Mosaic	0.857
Swin Transformer	Learning rate	$1.36 \times 10^{-5}$
	Weight decay	$5.63 \times 10^{-3}$
	Batch	4
	Drop out rate	0.311

우수한 결과를 보였다.

Table 4를 보면 정량적 성능 지표에서도 이러한 경향이 확인되었다. Swin Transformer는 0.988의 정확도, 0.987의 정밀도, 0.987의 재현율, 0.987의 F1 score, 0.988의 mAP로 네 가지 모델 중 가장 높은 성능을 기록하였다. YOLOv11과 YOLOv8은 각각 정확도 0.975, 0.970로 Swin Transformer보다는 낮지만, 비교적 빠른 추론 속도(61.67 ms/img, 64.90 ms/img)를 보여 실시간 응용 가능성이 높음을 확인하였다. 반면 Faster R-CNN은 상대적으로 낮은 정확도(0.971)와 가장 느린 추론 속도(78.54 ms/img)를 보이며 최신 모델 대비 성능상 한계를 드러냈다. Faster R-CNN의 이러한 한계는 2-Stage Detector로서 필연적으로 영역 제안(Region Proposal) 단계를 거치면서 발생하는 높은 계산 복잡도로 인해 최신 1-Stage Detector에 비해 느린 추론 속도를 보이는 구조적 특성에 기인한다. 또한, 혼동 행렬에서 관찰된 바와 같이 PP와 PS 클래스 간의 오분류가 다수 발생하였는데, 이는 실제 페플라스틱 이미지 데이터셋 내에서 두 재질의 시각적 특징 유사성이 높기 때문으로 분석된다. Faster R-CNN의 특징 추출 및 분류 능력이 최신 모델만큼 이러한 미묘한 시각적 차이를 효과적으로 구분하지 못했음을 시사하며, 이는 2-Stage 구조의 복잡성이 반드시 높은 정확도로 이어지지 않을 수 있음을 보여준다.

종합적으로 볼 때, Swin Transformer는 최고 수준의 분류 정확도를 제공하며, YOLOv11은 우수한 정확도와 더불어 가장 빠른 추론 속

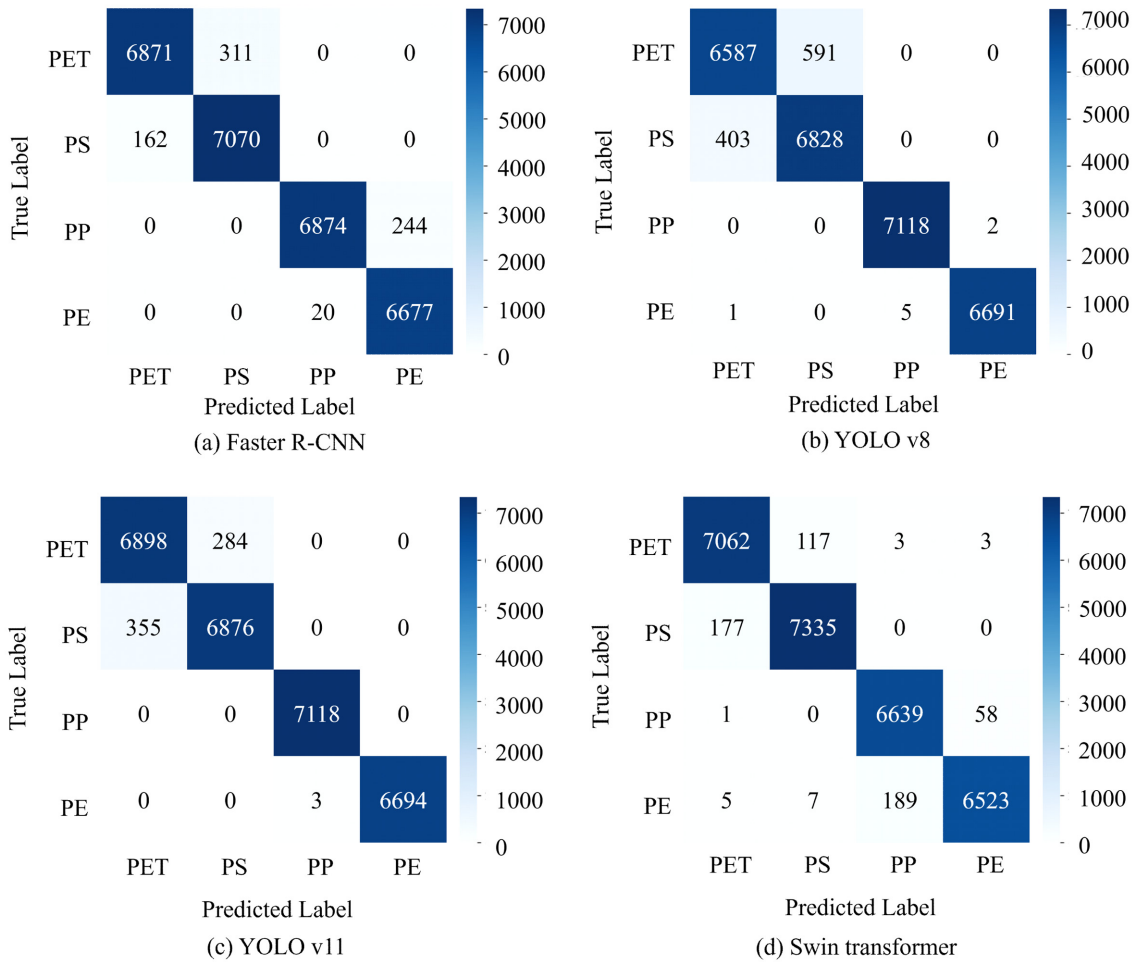


Fig. 4. Confusion matrices of each object detection model on the test set.

Table 4. Comprehensive evaluation of object detection models

	Accuracy	Precision	Recall	F1 Score	mAP 50-95	Inference time (ms/image)
Faster R-CNN	0.971	0.971	0.970	0.971	0.943	78.54
YOLO v8	0.970	0.971	0.965	0.968	0.974	64.90
YOLO v11	0.975	0.977	0.977	0.977	0.983	61.67
Swin transformer	0.988	0.987	0.987	0.987	0.988	63.85

도를 제공하여 실시간 페플라스틱 분류 작업에 적합하다. YOLOv8은 YOLOv11보다 다소 낮은 성능을 보였으나 여전히 균형 잡힌 성능을 나타냈다. 따라서 적용 목적에 따라 정확도를 최우선시하는 경우 Swin Transformer를, 실시간 처리와 효율성을 중시하는 경우 YOLOv11을 선택하는 것이 바람직하다.

3-3. 모델의 예측 바운딩 박스(Bounding Box) 및 컨피던스 스코어(Confidence Score)

다양한 객체 탐지 모델이 페플라스틱 이미지를 대상으로 수행한 탐지 결과를 바운딩 박스와 confidence score, 클래스 이름과 함께 시각화하였다. 각 그림은 Faster R-CNN, Swin Transformer, YOLOv8, YOLOv11의 성능 차이를 직관적으로 비교할 수 있도록 구성되었다.

Fig. 5에서 확인할 수 있듯이, Swin Transformer는 다양한 형태의 플라스틱을 안정적으로 탐지하며, 배경 잡음에도 비교적 강건한 성능을 보였다. Faster R-CNN은 다수의 바운딩 박스를 생성하여 탐지 정

확도는 높으나 불필요한 중복 박스가 나타나는 경향을 보였다. YOLOv8과 YOLOv11은 비교적 간결한 탐지 결과를 제공하며, 특히 YOLOv11은 높은 confidence score를 바탕으로 안정적인 객체 탐지를 수행하였다. 이러한 시각적 비교를 통해 각 모델의 탐지 특성과 한계를 명확히 확인할 수 있었다.

3-4. 결과에 대한 고찰 및 분석

실험 결과, Swin Transformer를 백본으로 활용한 모델이 복잡한 페플라스틱 분류 환경에서 가장 우수한 성능을 보였다. 이는 Swin Transformer의 계층적 특징 추출 능력과 윈도우 기반 셀프 어텐션 메커니즘(Window-based Self-Attention)이 다양한 크기와 형태를 가진 페플라스틱 객체를 효과적으로 탐지하고 분류하는 데 기여했음을 시사한다. 특히, 불규칙한 모양이나 배경 잡음이 많은 데이터셋에서도 안정적인 성능을 나타내어, 기존 합성곱 신경망(CNN) 기반 모델들이 가지는 한계를 보완할 수 있음을 확인하였다.



Fig. 5. Bounding box and confidence score visualization of waste plastic detection by models [28].

반면, YOLO 계열 모델들은 절대적인 정확도 면에서는 Swin Transformer보다 낮은 성능을 보였으나, 빠른 추론 속도와 경량성이라는 장점을 바탕으로 실시간 처리 환경에서의 활용 가능성을 입증하였다. 이는 대규모 폐플라스틱 선별 라인과 같이 실시간성이 중요한 응용 환경에서 특히 유용할 수 있으며, 고성능 GPU가 아닌 일반적인 산업용 장비에서도 적용 가능성이 높다는 점에서 의미가 있다.

또한 Faster R-CNN의 경우 높은 정확도와 안정적인 분류 성능을 보였으나, 연산량과 추론 속도 측면에서는 YOLO 계열 모델이나 Swin Transformer 기반 모델에 비해 상대적으로 불리하였다. 따라서 정확도가 최우선인 연구 환경에서는 Faster R-CNN이 적합할 수

있으나, 실제 산업적 적용을 고려할 경우에는 Swin Transformer나 YOLO 계열 모델이 더 합리적인 선택지가 될 수 있다.

또한 주요 오분류 사례를 분석한 결과 플라스틱 표면이 이물질로 심하게 오염되었거나 형태가 극심하게 변형(찌그러짐)된 경우 또는 여러 플라스틱이 겹쳐져 객체의 경계가 불분명한 경우에 주로 예측 오류가 발생하는 경향을 확인하였다. 이는 향후 데이터 전처리 과정에서 이러한 샘플들을 보강할 필요성을 시사한다.

각 모델의 정확도(Accuracy), 정밀도(Precision), 재현율(Recall), F1 점수(F1-Score), 평균 정밀도(mAP)와 같은 핵심 성능 지표를 종합적으로 비교한 결과, 본 연구는 폐플라스틱 자동 분류 시스템 구축 시 요구되는 목적인 정확도나 실시간 적용 가능성에 따라 합리

적인 모델 선택 기준을 제시할 수 있었다. 이는 향후 인공지능 기반 폐플라스틱 선별 기술의 실용화에 있어 중요한 참고 자료가 될 것으로 기대된다.

#### 4. 결 론

본 연구에서는 폐플라스틱 자동 분류를 위한 인공지능 기반 객체 탐지 모델의 성능을 비교하였다. CNN 기반 모델인 Faster R-CNN, YOLO 계열(YOLOv8, YOLOv11), 그리고 Transformer 기반 모델인 Swin Transformer를 동일 데이터셋에서 학습 및 평가하였으며, 정확도, 정밀도, 재현율, F1 score, mAP, 추론 시간 등 다양한 성능 지표를 활용하였다. 실험 결과, Swin Transformer는 모든 지표에서 우수한 성능을 나타내며 특히 높은 정확도(Accuracy 0.988)와 mAP (0.988)를 달성하여 복잡한 폐플라스틱 분류 환경에서도 안정적인 분류 능력을 입증하였다.

반면, YOLOv11은 정확도가 높으면서도 가장 빠른 추론 속도(61.67 ms/image)를 기록하여 실시간 자동 분류 시스템 구축에 적합함을 확인하였다. YOLOv8은 성능과 속도 면에서 균형을 보여 다양한 상황에서 활용 가능하며, Faster R-CNN은 상대적으로 낮은 정확도와 느린 처리 속도를 보여 최신 실시간 응용에는 한계가 있음을 나타냈다. 이러한 결과는 폐플라스틱 분류 모델 선택 시 정확도와 실시간 처리 속도 간 트레이드오프를 고려해야 함을 시사한다.

본 연구는 4개 클래스(PET, PP, PS, PE)를 대상으로 수행되었으나, 실제 재활용 환경에서는 오염, 투명 필름, 혼합 재질 등 보다 복잡한 데이터가 존재한다. 본 연구에서 사용된 데이터는 이러한 실제 현장의 변수들을 모두 반영하지 못한다는 한계가 있다. 향후 연구에서는 복합 재질 및 오염 환경을 포함한 데이터셋 확장과 함께 모델 경량화, 에너지 효율성, 실시간 하드웨어 적용 가능성 등 실제 산업 현장에서의 실용적인 지표들을 고려한 연구가 필요하다. 예를 들어, 본 연구에서 빠른 속도를 보인 YOLOv11 모델은 경량화 측면에서도 이점이 있어 상대적으로 저전력의 엣지 디바이스(edge device) 적용에 유리할 수 있으며, 이는 에너지 효율과 설비 비용 절감에 기여할 수 있다. 반면, 가장 높은 정확도를 보인 Swin Transformer 기반 모델(또는 Swin Transformer (Backbone) + Faster R-CNN)은 많은 파라미터 수로 인해 더 높은 연산 자원과 전력을 필요로 하므로, 유지보수 및 운영 비용 측면에서 다르게 평가될 수 있다. 본 연구 결과는 폐플라스틱 자동 분류 시스템 구축 및 최적 모델 선택을 위한 실질적 참고 자료로 활용될 수 있으며, 이러한 실용적 요소들을 포함한 후속 연구를 통해 지속 가능한 자원 순환 체계 확립에 기여할 수 있을 것으로 기대된다. 또한, 본 연구에서 확인된 바와 같이 모델별로 특정 클래스 분류에 강점을 보이는 경향이 다르므로 향후 여러 모델의 예측을 결합하는 앙상블(ensemble) 기법을 도입하여 전반적인 분류 정확도를 더욱 향상시키는 방안도 모색할 수 있을 것이다.

#### Reference

- Pilapitiya, P. G. C. N. T. and Ratnayake, A. S., "The World of Plastic Waste: A Review," *Clean. Mater.* **11**, 100220(2024).
- Dokl, M., Van Fan, Y., Vujanović, A., Pintarič, Z. N., Aviso, K. B., Tan, R. R., Pahor, B., Kravanja, Z. and Čuček, L., "A Waste Sep-

- aration System Based on Sensor Technology and Deep Learning: A Simple Approach Applied to a Case Study of Plastic Packaging Waste," *J. Clean. Prod.* **450**, 141762(2024).
- Nafiu, S. A., Ogunbusola, E. M., Adeyemi, A. G., Olatunji, T. L. and Oladimeji, A. I., "Waste Plastic Management: Recycling and the Environmental Health Nexus," *Clean. Mater.* **15**, 100291(2025).
- Lubongo, C. and Alexandridis, P., "Assessment of Performance and Challenges in Use of Commercial Automated Sorting Technology for Plastic Waste," *Recycling* **7**(2), 11(2022).
- Muzata, T. S., Matuana, L. M. and Rabnawaz, M., "Challenges in the Mechanical Recycling and Upcycling of Mixed Postconsumer Recovered Plastics (PCR): A Review," *Current Research in Green and Sustainable Chemistry* **8**, 100407(2024).
- Satav, A. G., Pradhan, S., Dange, S. P. and Bendre, P., "A State-of-the-Art Review on Robotics in Waste Sorting: Scope and Challenges," *International Journal on Interactive Design and Manufacturing (IJIDeM)* **17**(6), 2789-2806(2023).
- Kerdlap, P., Sakulwiriyanan, P. and Gheewala, S. H., "Life Cycle Cost Analysis of Distributed versus Centralized Plastic Sorting and Recycling," *Journal of Industrial Ecology*. **27**(1), 297-311(2023).
- Ruj, B., Pandey, V., Jash, P. and Srivastava, V. K., "Sorting of Plastic Waste for Effective Recycling," *Int. J. Appl. Sci. Eng. Res.*, **4**(4), 564-571(2015).
- Choi, J., Lim, B. and Yoo, Y., "Advancing Plastic Waste Classification and Recycling Efficiency: Integrating Image Sensors and Deep Learning Algorithms," *Appl. Sci.*, **13**(18), 10224(2023).
- Jobert, G. and Brenière, X., "Mid-infrared Spectrometer for Black Plastics Sorting Using a Broadband Uncooled Micro-bolometer Array," *Spectrosc. J.*, **3**(2), 13(2025).
- Chen, X., Yan, L., Yin, X., Zhao, Y. and Zeng, W., "Determining the Composition of Post-consumer Flexible Multilayer Plastic Packaging with Near-infrared Spectroscopy," *Waste Manag.*, **123**, 33-41(2021).
- Ramos, E., Lopes, A. G. and Mendonça, F., "Application of Machine Learning in Plastic Waste Detection and Classification: A Systematic Review," *Processes* **12**(8), 1632(2024).
- Fotovatikhah, F., Bazarghan, M., Pirani, M. and Malmir, M., "A Systematic Review of AI-based Techniques for Automated Waste Classification," *Sensors* **25**(10), 3181(2025).
- Fang, B., Yu, J., Chen, Z., Osman, A. I., Farghali, M., Ihara, I. and Yap, P. S., "Artificial Intelligence for Waste Management in Smart Cities: A Review," *Environ. Chem. Lett.*, **21**(4), 1959-1989(2023).
- Lubongo, C., Bin Daej, M. A. A. and Alexandridis, P., "Recent Developments in Technology for Sorting Plastic for Recycling: The Emergence of Artificial Intelligence and the Rise of the Robots," *Recycling* **9**(4), 59(2024).
- Zhao, X., Wang, L., Zhang, Y., Han, X., Deveci, M. and Parmar, M., "A Review of Convolutional Neural Networks in Computer Vision," *Artif. Intell. Rev.*, **57**(4), 99(2024).
- Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, L. and Tong, J., "Recent Advances in Convolutional Neural Networks," *Pattern Recognit.*, **77**, 354-377(2018).
- He, K., Gkioxari, G., Dollár, P. and Girshick, R., "Mask R-CNN," *Proc. IEEE Int. Conf. Comput. Vis.* 2017.
- Redmon, J., Divvala, S., Girshick, R. and Farhadi, A., "You Only Look Once: Unified, Real-time Object Detection," *Proc. IEEE*

- Conf. Comput. Vis. Pattern Recognit. 2016.
20. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S. and Guo, B., "Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows," Proc. IEEE/CVF Int. Conf. Comput. Vis. 2021.
  21. Wang, Z., Li, M., Zhou, Y. and Zhang, X., "Multi-category Sorting of Plastic Waste Using Swin Transformer: A Vision-based Approach," *J. Environ. Manag.*, **370**, 122742(2024).
  22. Kunwar, S., Owabumoye, B. R. and Alade, A. S., "Plastic Waste Classification Using Deep Learning: Insights from the WaDaBa Dataset," arXiv preprint arXiv:2412.20232 (2024).
  23. Ahmed, M. I. B., Rahman, A. and Osman, S., "Deep Learning Approach to Recyclable Products Classification: Towards Sustainable Waste Management," *Sustainability* **15**(14), 11138(2023).
  24. Alsabbagh, A. R. and Al-Kadi, O., "Comparative Analysis of Deep Convolutional Neural Networks for Detecting Medical Image Deepfakes," arXiv preprint arXiv:2406.08758 (2024).
  25. Ming, L. W., Foo, S. Y., Ng, A. and Yeap, G., "AI as a Driver of Efficiency in Waste Management and Resource Recovery," *Int. Trans. Artif. Intell.*, **2**(2), 128-134(2024).
  26. Chacón-Albero, O., Lopez, R. and Alvarez, M., "AI for Sustainable Recycling: Efficient Model Optimization for Waste Classification Systems," *Sensors* **25**(12), 3807(2025).
  27. Ahn, S., Junhyeok, and Yuchan, A., "AI-based Plastic Waste Sorting Method Utilizing Object Detection Models for Enhanced Classification," *Waste Manag.*, **193**, 273-282(2025).
  28. AIHub, "Plastic Waste Dataset," aihub.or.kr, <https://www.aihub.or.kr/> (Accessed 2025-01-10).
  29. Ren, S., He, K., Girshick, R. and Sun, J., "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *Advances in Neural Information Processing Systems* **28** (2015).
  30. Hidayatullah, P., Malik, H. and Gunawan, W., "YOLOv8 to YOLO11: A Comprehensive Architecture In-Depth Comparative Review," arXiv preprint arXiv: 2501.13400 (2025).
  31. Raiaan, M. A. K., Ghosh, S., Rahman, M. and Hossain, J., "A Systematic Review of Hyperparameter Optimization Techniques in Convolutional Neural Networks," *Decis. Anal. J.*, **11**, 100470 (2024).
  32. Ultralytics, "YOLO Documentation," docs.ultralytics.com, <https://docs.ultralytics.com/> (Accessed 2025-01-10).
  33. Ghate, D., Patil, A. and Pawar, V., "Advancing Arecanut Quality Grading: A Comparative Analysis of YOLO Models with Hyperparameter Optimization," (2025).
  34. Diphtho, R. A. and Basak, S., "Enhancing Dermatological Diagnosis Through Medical Image Analysis: How Effective Is YOLO11 Compared to Leading CNN Models?," *NDT* **3**(2), 11(2025).
  35. Praveena, S. and Veeramakali, T., "Multiple Hyperparameter Optimization in Image Processing Employing Swin Transformer for Enhanced Skin Lesion Segmentation," 2025 Int. Conf. Comput. Robot. Test. Eng. Eval. (ICRTEE), IEEE, 2025.

#### Authors

**Do-Eun Kim:** Undergraduate Student, Department of Chemical Engineering, Keimyung University, Daegu 42601, Korea; rlom0212@gmail.com

**Jun-Hyeok Son:** Master's Degree, Department of Chemical Engineering, Keimyung University, Daegu 42601, Korea; science2718@gmail.com

**Do-Gyeong Kim:** Undergraduate Student, Department of Chemical Engineering, Keimyung University, Daegu 42601, Korea; kdg081447@gmail.com

**Yu-Chan Ahn:** Professor, Department of Chemical Engineering, Keimyung University, Daegu 42601, Korea; yuchan.ahn@kmu.ac.kr