한국정보전자통신기술학회논문지(jkiiect)'24-6, Vol.17 No.3
논문 24-17-03-121 http://dx.doi.org/10.17661/jkiiect.2024.17.3.121

Online to Offline 상점의 자동화 : 초소형 깊이의 Yolov8과 특징점 기반의 상품 인식

시종욱*, 김대민**, 김성영***

Automation of Online to Offline Stores: Extremely Small Depth-Yolov8 and Feature-Based Product Recognition

Jongwook Si*, Daemin Kim**, Sungyoung Kim***

요 약 디지털 기술의 급속한 발전과 코로나19 팬데믹으로 인해 온라인 상거래가 크게 성장하면서, 소상공인들이 이러한 시장 변화에 적극적으로 대응할 수 있는 지원 방안의 필요성이 대두되었다. 이에 본 논문은 O2O(Online to Offline) 전략을 활용해 실제 매장 진열대에 전시된 상품들을 자동으로 촬영하고 이를 이용해 가상 상점을 만들 수 있는 기초적인 기술을 제시한다. 본 연구의 핵심은 진열된 상품의 위치와 이름을 정확히 파악하여 인식하는 것이며, 이를 위해 단일 클래스를 대상으로 하며 YOLOv8에 기반한 경량화 모델인 ESD-YOLOv8을 제안한다. 검출된 상품은 특징점 기반의 기술을 통해 상품명이 식별되며, 이는 새 상품을 사진 형태로 추가함으로써 신속하게 갱신할 수 있는 능력을 갖추고 있다. 실험을 통해 상품명 인식은 74.0%의 정확도, 위치 검출은 0.3M개의 파라미터만으로 F2-Score 기준 92.8%의 성능을 보였다. 이를 통해 제안된 방법이 높은 성능과 최적화된 효율성을 갖추고 있음을 확인하였다.

Abstract The rapid advancement of digital technology and the COVID-19 pandemic have significantly accelerated the growth of online commerce, highlighting the need for support mechanisms that enable small business owners to effectively respond to these market changes. In response, this paper presents a foundational technology leveraging the Online to Offline (O2O) strategy to automatically capture products displayed on retail shelves and utilize these images to create virtual stores. The essence of this research lies in precisely identifying and recognizing the location and names of displayed products, for which a single-class-targeted, lightweight model based on YOLOv8, named ESD-YOLOv8, is proposed. The detected products are identified by their names through feature-point-based technology, equipped with the capability to swiftly update the system by simply adding photos of new products. Through experiments, product name recognition demonstrated an accuracy of 74.0%, and position detection achieved a performance with an F2-Score of 92.8% using only 0.3M parameters. These results confirm that the proposed method possesses high performance and optimized efficiency.

Key Words : Online to Offline(O2O), Automation, Object Detection, Feature Matching, YOLO

1. 서론

디지털 기술의 발전은 온라인 상거래를 크게 촉진하고 있다. 특히 코로나19로 인한 오프라인 활동의 제한은 온라인 거래의 급격한 성장을 불러왔다. 2020년 코로나19 유행이 시작된 이후, 온라인 쇼핑 거래액은

This work was supported by the Technology development Program(S3344882) funded by the Ministry of SMEs and Startups(MSS, Korea)
* Dept. Computer·AI Convergence Engineering, Kumoh National Institute of Technology
** Dept. Computer Engineering, Kumoh National Institute of Technology
** Dept. Computer Engineering, Kumoh National Institute of Technology (Corresponding Author)
Received May 10, 2024 Revised May 15, 2024 Accepted May 19, 2024

2020년과 2021년에 각각 19.5%, 23.5% 증가했다는 통계는 온라인 판매의 중요성을 더욱 강조한다[1]. 그러나 디지털 접근성이 제한적인 소상공인들에게 있어서는 이러한 변화는 어려움을 동반하기도 한다.

대형 유통업체들의 확장으로 인해 2020년을 기준으로 우리나라 전통시장의 수가 1,401개로 감소하였다. 이는 2006년의 1,610개에서 209개(약 13%)가 감소한 수치이다. 점포 수 역시 2016년의 225,725개에서 2020년에는 207,145개로, 약 18,580개(8.2%)의 점포가 줄어들었다는 사실은 상당히 주목할 만하다. 이와 대조적으로, 동네 슈퍼마켓이나 전문소매점과 비교했을 때 백화점과 대형마트의 판매액은 같은 기간 동안 두 배에 가깝게 증가하였다[2].

이러한 환경에서 소상공인들이 온라인 시장 변화에 효과적으로 대응할 수 있도록, 인공지능 기반의 자동화된 온라인 상점 구축의 필요성이 대두되고 있다. 이러한 전략을 O2O(Online to Offline)이라 하며[3-4], 오프라인 상품을 온라인 상에서 거래할 수 있는 플랫폼을 의미한다. 본 논문에서는 O2O 전략을 활용하여 실제 진열대에 있는 상품들을 촬영하여 가상 상점을 구성하는 데 기반이 되는 상품 인식 기술을 제안한다. 이러한 O2O 상점은 자동 촬영으로 상품을 인식하고 이를 바탕으로 개인 온라인 상점을 구축하도록 한다.

본 연구는 진열대에 있는 상품들을 인식하기 위해 객체 검출 기술을 활용한다. 진열대 상품은 다양한 행사나 신제품 출시 등으로 인해 그 위치와 구성이 자주 변경되는 특성을 가지고 있다. 기존의 객체 인식 모델을 사용할 경우, 모델을 지속적으로 재학습하고 파라미터를 조정해야 하므로, 신제품을 인식하는 데 시간이 많이 소요되는 단점이 있다. 이에 본 논문에서는 진열대 사진에서 상품의 위치를 검출하기 위한 목적으로 단일 클래스를 사용하는 객체 검출 모델을 채택하고, 검출된 결과를 기반으로 영상 처리 기술의 특징점 추출 및 매칭 과정을 통해 구체적인 상품을 식별할 수 있는 시스템을 제안한다. 이 방식은 새로운 상품이 추가될 때 간단히 사진을 후보 데이터에 추가하는 것만으로도 신속하게 대응할 수 있어 효율적이다. 본 논문에서는 대표적인 1단계 객체 감지기인 YOLOv8[5]을 기반으로 파라미터를 줄여 O2O 상점을 위한 최적화를 목적으로 개선된 모델을 제안한다. 기존의 모델은 여러 클래스에 대하여 고려하였기 때문에, 파라미터 수가 많고 모델의 구조가 깊기에 추론 속도가 느린 단점이 있다. 하지만, 제안 방법에서는 클래스 하나에 대해서만 고려하기 때문에 O2O 상점에 적합하고 빠른 속도로 추론이 가능한 경량화된 모델이 필요하다.

2절에서는 관련 연구에 대해 소개하고 3절에서 본 논문에서 제안하는 O2O 상점의 자동화 방법에 대해 소개한다. 4절에서는 객체의 위치 검출과 특징점 추출 및 매칭의 정확도를 측정하고 분석한다. 마지막으로 5절에서는 결론과 추후 연구에 대해 소개한다.

2. 관련 연구

2.1 객체 검출 모델 및 활용

객체 탐지 기술에는 두 가지 유형이 있는데, 1단계 객체 탐지기(1-stage object detector)와 2단계 객체 탐지기(2-stage object detector)로 분류된다. 1단계의 경우 속도를 우선시하므로 간단한 구조로 구성된 것이 특징이며, 2단계 객체 탐지기는 입력 이미지에서 주요 객체 후보를 선별하는 단계와 이후 객체의 정확한 위치와 종류를 예측하는 단계로 구성된다. 각 유형별로 대표적인 모델들이 존재한다.

대표적인 1단계 모델로는 YOLO[6], SSD[7], RetinaNet[8]이 있다. YOLO[6]는 영상을 그리드로 나눈 뒤, 그리드 셀 별로 경계 상자와 클래스 확률을 예측한다. 하나의 통합 모델에서 객체의 종류와 위치를 동시에 계산함으로써 처리 속도가 빠른 것이 특징이다. 하지만, 작은 객체에 대한 정확도가 상대적으로 낮으며 이를 개선한 최신 기술로 제안 방법에서 사용할 YOLOv8[5]이 있다. SSD[7]는 여러 스케일의 특징맵을 통해 다양한 크기의 객체를 탐지하는 기술이다. 각 특징 맵에서 클래스 점수와 위치 좌표를 추정하고, NMS를 통해 최종 경계 상자를 결정한다. YOLO[6]와 비교하였을 때, 작은 객체에 대한 성능이 개선된 모델이다. RetinaNet[8]은 손실 함수에 초점을 두어 일반적인 1단계 객체 탐지기의 낮은 성능 문제를 해결한 기술이다. 이는 소수의 실제 객체에 대한 학습을 강화

하여 2단계 탐지기에 준하는 성능을 달성하였다.

대표적인 2단계 모델로는 R-CNN[9], Fast R-CNN[10], Faster R-CNN[11]이 있다. R-CNN[9]은 Selective Search를 사용해 영상에서 후보 영역을 추출하여 CNN을 통한 특징맵을 생성하며, 이를 SVM을 사용해 객체를 분류하는 기술이다. 하지만, CNN 입력을 위한 영상의 크기 변형으로 인해 손실이 있고 후보 영역만큼 CNN을 반복해야 한다는 단점이 있다. Fast R-CNN[10]은 R-CNN[9]을 개선한 방법으로 이미지 전체에 CNN을 적용해 특징 맵을 생성하고, 이를 바탕으로 후보 영역을 계산한다. 그리고 ROI Pooling을 통해 특징 벡터를 추출하여 Fully Connected 층과 Softmax를 통해 객체를 분류하는 방식이다. Faster R-CNN[11]은 Fast R-CNN[10]을 개선한 연구로, RPN(Region Proposal Network)을 사용하여 더욱 빠르게 후보 영역을 제안하도록 수정하였다. RPN은 슬라이딩 윈도우와 앵커 박스를 사용해 후보 영역을 예측하며, NMS로 최종 후보를 선정한다. 이와 같이, 각 객체 탐지 모델은 특유의 방식으로 영상 내 객체를 식별하며, 다양한 상황에 적용할 수 있다[12-13].

2.2 특징점 검출 및 매칭

대표적인 특징점 검출 방법에는 SIFT[14], SURF[15], ORB[16]가 있다.

SIFT[14]는 영상의 크기가 변해도 동일한 특징을 일관되게 식별할 수 있는 특징을 지닌 기술이다. 이 기술은 영상에서 변화에 강한 키포인트를 탐색하고, 이들 키포인트에 대한 디스크립터를 생성하여 영상의 고유한 특성을 파악한다.

SURF[14]는 SIFT[13]의 내용을 기반으로 속도를 향상한 기술이다. LoG 대신 박스 필터를 사용하여 근사화 함으로써 처리 속도를 개선한 방식이다. 그리고 키포인트 탐색과 디스크립터 계산 과정 모두에서 속도 향상을 위한 여러 최적화를 도입하였다.

ORB[15]는 FAST 기술로 초기 키포인트를 신속하게 탐지한 후, Harris 코너 검출기를 활용해 중요도가 높은 상위 N개의 키포인트를 선별한다. 이어서 영상의 스케일 변화에 강인한 특성을 확보하기 위해 다양한 크기의 이미지 피라미드를 적용하며, 회전에 대한 불변성을 제공하기 위해 수정된 BRIEF[17] 디스크립터를 사용한다.

이렇게 추출된 특징들은 매칭기를 통해 두 영역을 매칭 함수를 통해 비교하여 유사도를 평가하게 된다. 대표적인 매칭기로는 BFMatcher(Brute-Force Matcher)와 FLANNMatcher(Fast Library for Approximate Nearest Neighbors Matcher)가 있다. BFMathcer는 모든 디스크립터를 비교하여 평가하는 방식이며, FLANN Matcher는 모든 디스크립터를 비교하는 것이 아니라 이웃하는 디스크립터끼리 비교하는 방식이다.

3. O2O 상점의 자동화를 위한 방법

3.1 상품의 위치 검출

YOLOv8[5]은 객체 탐지를 위해 널리 사용되는 모델 중 하나로 크게 두 부분, 즉 Backbone과 Neck, Head로 나뉜다. Backbone 부분에서는 입력 데이터에서 중요한 특징들을 추출하는 역할을 한다. 그리고 Neck 부분에서는 추출된 특징에 대하여 upsample과 concatenation 과정을 거친다. 이렇게 병합된 특징들은 Head 부분에서 분류나 탐지와 같은 작업에 사용한다. 기존 YOLO모델은 anchor를 이용하여 객체를 검출하는 방식이었지만, 이와 달리 YOLOv8은 anchor를 사용하지 않는다. 이로 인해 anchor와 관련된 크기, 종횡비를 고려하지 않아 많은 계산을 필요로 하지 않기 때문에, 훈련 과정이 빠르다는 장점이 있다. 대신, 객체의 중심을 예측하여 크기와 형태를 동시에 감지하는 방식으로 진행한다.

모델의 핵심 구조에 대해 살펴보면, 합성곱 층에서는 일반적으로 배치 정규화와 SiLU라는 활성화 함수가 사용된다. YOLOv8은 크게 두 부분인 backbone과 head로 나누어진다. backbone에서는 입력 영상의 특징을 추출하는 역할을 하며, C2f(Cross Stage Partial Bottleneck with 2 convolutions faster)와 SPPF(Spatial Pyramid Pooling-Fast) 과정이 포함된다. 그리고 head에서는 backbone에서 추출한 특징을 통해 실제 객체에 대한 검출 과정을 진행한다.

C2f는 1x1 합성곱을 거쳐 데이터를 분리한 뒤,

그림 1. O2O 상점에 특화된 제안하는 ESD-YOLOv8 모델의 구조

Fig. 1. Proposed ESD-YOLOv8 Model Architecture Specialized for O2O Stores

bottleneck 구조를 통과시키고, 이를 다시 병합하는 과정을 포함한다. 이때, bottleneck을 통과하기 전의 초기 입력값과 통과한 후의 출력값을 병합하는 이유는 합성곱 층을 통과하며 발생할 수 있는 데이터의 손실을 최소화하기 위해서이다. SPPF는 1x1 합성곱을 거친 후, 맥스 풀링을 세 번, 두 번, 한 번 및 전혀 거치지 않은 네 가지 상태를 각각 만든 다음, 이를 하나로 합치는 과정을 거친다. 이러한 과정은 특징들을 고정된 맵에 풀링함으로써, 연산하는 속도를 가속화하는 역할을 한다.

객체 검출 과정에서는 3x3 합성곱을 두 번 거친 후 1x1 합성곱을 통해 바운딩 박스(bbox)와 클래스(cls)가 결정된다. 이 과정을 통해 YOLOv8은 입력 영상 내의 객체들을 정확하게 탐지하고 분류할 수 있다. 제안하는 방법은 YOLOv8을 O2O 상점에 특화된 모델을 만드는 것이며 이에 따라 여러 가지 부분을 수정한다. 서론에서 소개한 바와 같이 객체의 위치 검출만을 위해 사용하기 때문에 다양한 상품들은 "상품"이라는 하나의 클래스로 통합되며, 그림 1과 같이 O2O 상점에서 효율적으로 사용할 수 있는 경량화된 모델을 제안한다. 기존 모델의 깊이를 초소형으로 구성하여 실제 환경에서 빠르게 추론할 수 있도록 하는 것이 주된 목표이다. 전체적인 구조는 기존의 YOLOv8과 동일하며 모든 계층에서의 출력에서 깊이를 조정한 ESD-YOLOv8을 제안한다. 각 계층에서의 출력층의 크기는 표 1과 같다.

기존 YOLOv8과 같은 15, 18, 21번째 계층의 출력인 특징맵을 통해 객체를 검출하는 것은 동일하지만, 이 특징맵들의 크기는 (80,80,64), (40,40,32), (20,20,32)으로 매우 얕은 깊이임을 알 수 있다. 모든 층의 채널 개수가 Large 모델의 6.25%, Nano 모델의 25%로만 구성되도록 조정한다. 또한, C2F 내 Bottleneck의 반복을 제거한다. 모델 이외의 설정 값은 YOLOv8의 기본값과 동일하게 유지된다.

3.2 상품명 인식

제안하는 ESD-Yolov8을 통해 진열대 영상에서 각 상품의 좌표를 나타내는 위치 정보를 추출하고 식별된 영상과 정답 상품 영상 간 유사성을 평가하기 위해 특징 매칭 기법을 적용하여 상품명을 판단한다. 이 과정에서, 영상의 특징점과 해당 디스크립터를 추출 후, 추출된 정보를 정합하여 유사한 객체를 식별한다. 이를 통해 가장 유사한 영상의 매칭결과에 따라 상품명을 분류할 수 있다. 이 과정에서 KnnMatch 함수를 사용하여 매칭 쌍의 거리 비율을 활용한 필터링을 적용한다. 이 방법을 통해 부정확한 매칭을 효과적으로 제거하고, 결과적으로 가장 유사성이 높은 영상과의 매칭에서 높은 수의 매칭점을 확보한다. 그림 2는 이에 대한 예시를 나타내며 오른쪽 사진과 가장 유사한 특징점을 검출하여 매칭하는 과정을 살펴보면 실제 정답 사진에 가장 많은 매칭점이 있음을 확인할 수 있다. 매칭 쌍의 거리를 고려하여 잘못된 매칭을 최소화하였기 때문에 정확한 특징점에 대한 매칭 결과를 보다 정확하게 활용할 수 있다.

표 1. ESD-Yolov8의 세부사항

Table 1. Detailed information of ESD-YOLOv8

Num	Layer	Output	Detect
0	Conv	320x320x4	
1	Conv	160x160x8	
2	C2f	160x160x8	
3	Conv	80x80x16	
4	C2f	80x80x16	
5	Conv	40x40x32	
6	C2f	40x40x32	
7	Conv	20x20x64	
8	C2f	20x20x64	
9	SPPF	20x20x64	
10	Upsample	40x40x64	
11	Concat	40x40x96	
12	C2f	40x40x64	
13	Upsample	80x80x32	
14	Concat	80x80x64	
15	**C2f**	**80x80x64**	O
16	Conv	40x40x16	
17	Concat	40x40x32	
18	**C2f**	**40x40x32**	O
19	Conv	20x20x32	
20	Concat	20x20x96	
21	**C2f**	**20x20x32**	O

본 논문에서는 특히 SURF[15]을 통한 특징점과 디스크립터를 계산하고 매칭을 진행한다. 매칭 간에 사용되는 매칭 함수는 KnnMatch를 채택한다. 이는 k-nn 알고리즘을 통해 구성된 매칭 함수로 k개의 최근접 이웃 개수만큼 디스크립터에서 찾아서 반환하는 기능을 한다. 이러한 방법의 경우 때때로 잘못된 매칭 결과를 포함할 수 있는데, 이를 해결하기 위해 매칭 함수가 반환하는 각 매칭 쌍의 거리 비율을 검증하여 오류를 필터링하는 방식을 채택하였다. 일반적으로, 두 번째로 가까운 이웃과의 거리를 첫 번째 이웃과의 거리로 나눈 비율을 계산하여, 설정된 임계값 τ에 따라 좋은 매칭을 선별한다.

그림 2. 특징점을 이용한 상품명 인식

Fig. 2. Product name using feature points

4. 실험 및 결과

4.1 데이터 세트

본 논문에서는 진열대에서 "상품"으로 라벨링한 객체를 감지하기 위해 편의점, 슈퍼마켓 등에서 직접 수집한 448장의 영상 데이터를 활용하여 학습과 평가를 수행한다. 수집한 데이터는 진열대의 하나의 열에 5~7개의 상품이 나타나도록 촬영하여 데이터 세트의 분포를 일치 시킨다. 모든 상품은 과자와 라면 두 가지 카테고리로 분류되며, 모든 데이터는 640x640 픽셀의 크기를 갖는다. 그리고 각 샘플에는 여러 개의 인스턴스가 포함되어 있다. 진열된 상품 데이터를 수집하는 것은 매우 어려운 상황이므로, 학습 데이터의 비율은 전체 데이터의 90% 이상으로 설정한다. 표 2는 본

논문에서 사용된 데이터 세트의 영상과 인스턴스의 개수를 보여준다.

학습 데이터에 대해 다양한 영역에 위치한 상품들을 직접 어노테이션하며, 그림 3은 좌푯값과 객체의 가로 및 세로 길이에 관한 그래프를 나타낸다. 객체의 가로와 세로 길이는 앞에서 언급한 바와 같이, 주로 영상 크기의 10% 정도 부근에 위치하는 것으로 나타났으며, 세로 길이가 가로보다 더 긴 객체가 더 많이 포함되어 있는 것을 볼 수 있다. 객체의 크기가 표준화된 점은 데이터의 일관성을 유지하고 모델이 더욱 효과적으로 학습할 수 있다.

표 2. 학습 및 평가를 위한 데이터 개수
Table 2. Number of data for train and test

	Images	Instances
Train	418	10,017
Test	30	645
Sum	448	10,662

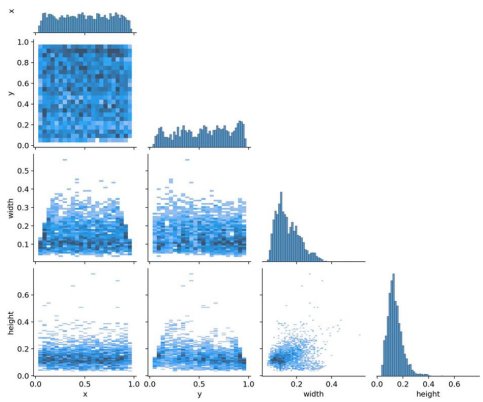


그림 3. 데이터세트의 상품 좌표와 크기에 대한 분포
Fig. 3. Distribution of product coordinates and sizes in the dataset

4.2 실험 환경 및 학습 세부 사항

본 논문에서는 Windows 10 운영체제를 기반으로 하며, GTX 1080 Ti 2대를 사용하여 ESC-YOLOv8 모델의 학습 및 평가를 진행한다. ESD-YOLOv8 모델은 사전 학습된 모델을 사용하지 않으므로, 초기 상태에서 무작위로 초기화된 채로 1000 epoch 동안 학습한다. Python 버전은 3.9.17을 사용하며, 딥러닝 모델을 위한 PyTorch는 1.9.1을 이용한다.

평가를 위해 제공된 다양한 크기의 YOLOv8 모델은 사전 훈련된 모델을 파인 튜닝하여 사용되며, 추가 학습 epoch는 50으로 설정한다. 입력 크기, 레이어 수, 파라미터 수, 그리고 FLOPs 값은 표 3과 같다. 제안하는 방법의 경량화 작업은 Nano 모델과 비교하였을 때 파라미터 수가 10% 수준임을 보인다.

4.3 성능 평가 및 결과

제안된 방법의 우수성을 확인하기 위해 상품을 올바르게 감지했는지를 평가하는 지표로 혼동 행렬을 사용한다. 또한, 객체 검출 모델에서 하나의 클래스에 대한 평균적인 정밀도를 계산하기 위해 F2-score와 AP(Average precision) 사용하여 평가한다. AP는 0.5와 0.5:0.95 두 가지 기준으로 계산하여 제시한다.

표 3에서는 추가로 다양한 크기의 YOLOv8 모델과 ESD-YOLOv8의 성능을 비교한 결과를 보여준다. 이는 Recall을 보다 강조하는 F2 Score를 통해 평가한다. 이 결과를 통해 제안된 방법이 Large 모델과 비교하여 5.5% 상승했음을 확인할 수 있다. 그러나 AP를 통해 측정한 결과에서는 제안된 방법이 가장 우수하다고 말하기는 어렵다. 또한, 일반적으로는 YOLO 모델에서 파라미터의 개수가 많아질수록(Nano에서 Xlarge로 갈수록) AP는 증가하는 형태를 보이나, 표

표 3. 상품 검출을 위한 기존 YOLOv8과 제안 모델의 구조와 성능 비교
Table 3. Comparison of the YOLOv8 structures for product detection and the proposed model

	Size	layers	params	FLOPs	F2-Score	AP	
						@0.5	@.5:.95
YOLOv8n[5]	640	168	3.0M	8.1G	0.906	**0.977**	0.668
YOLOv8s[5]	640	168	11.1M	28.4G	0.910	0.969	0.695
YOLOv8m[5]	640	218	25.8M	78.7G	0.892	0.970	**0.700**
YOLOv8l[5]	640	268	43.6M	164.8G	0.873	0.958	0.682
YOLOv8x[5]	640	268	68.1M	257.4G	0.910	0.970	**0.700**
ESD-YOLOv8 (Ours)	640	**158**	**0.3M**	**1.3G**	**0.928**	0.964	0.665

3의 결과에서는 큰 차이가 없음을 알 수 있다. mAP는 객체 검출 모델에서 많이 사용되는 지표이지만 주로 Precision에 중점을 두며, 지금과 같은 상황에서는 클래스가 하나이기 때문에 AP에 해당한다. 따라서, 클래스가 적기 때문에 해당 지표의 신뢰도가 낮다고 평가된다. 제안된 방법의 경량화된 특성과 하나의 클래스임을 고려할 때, 상대적으로 다소 낮은 AP에도 불구하고 큰 차이가 없기 때문에 제안하는 모델이 이점이 있다고 볼 수 있다.

제안된 방법은 Precision과 Recall 중에서 Recall이 상대적으로 중요하다. 이는 객체 검출 모델이 객체의 위치를 판단하고, 검출된 영역을 특징점을 통해 분류하기 때문에 정확하게 위치를 검출하는 것이 좋지만, 일부라도 정답 영역을 검출하는 것이 성능에 영향을 미치기 때문이다. 일부라도 영역을 찾지 못하면, 특징점을 통한 분류 과정을 진행할 수 없으므로 Recall이 매우 중요하다. 즉, False Negative(FN)의 비용이 더 큰 상황이기 때문에 Recall에 가중을 두어야 한다. 먼저, 표 4는 ESD-YOLOv8에 대한 상품이 아닌 영역을 모두 배경으로 하는 혼동 행렬을 제시한다. False Positive(FP)의 비율이 다소 높지만 FN의 가중을 고려해야 하는 상황에서 보면 높은 성능을 보인다고 할 수 있다. 이 혼동 행렬을 통해 평가한 결과로 Precision이 91.2%, Recall이 93.2%임을 보인다.

표 4. ESC-YOLOv8의 상품 검출에 대한 혼동 행렬
Table 4. Confusion Matrix for Product Detection using ESC-YOLOv8

		Act.	
		Product	Background
Pred.	Product	TP: 601	FP: 53
	Background	FN: 44	

특징 검출기와 매칭기를 결합하여 상품명을 분류하는 실험에서, 각각의 진열대 사진에 대한 상품 후보 사진들을 지정해두고 진행한다. 특징 검출기에는 SIFT[14], SURF[15], ORB[16]를, 특징 매칭기에는 BFMatcher와 FLANNMatcher를 사용하여 모든 조합을 통해 성능을 평가한다. 모든 매칭 함수는 τ를 0.7로 고정한 KnnMatch 함수를 사용한다.

표 5에 따르면, 특징 검출기와 매칭기의 여섯 가지 조합에 대한 정확도를 나타내고 있으며, 매칭 함수는 고정한 채로 각 조합별 상품명의 분류 성능을 보인다. 이는 위치 검출 결과에 위치한 상품에 대하여 실제 인식된 상품명이 올바른지에 대한 정확도를 의미한다. 분석 결과, SURF와 BFMatcher 조합이 74%의 정확도로 가장 우수한 성능을 보였으며, 이는 다른 모든 조합 중 최고의 성능을 의미한다. 동일한 조건에서, FLANNMatcher와 SURF 조합 또한 같은 정확도를 보여주었다. 이러한 결과는 SURF가 특징 검출기로서 뛰어난 성능을 나타내며, 매칭기의 종류가 SURF의 성능에 큰 영향을 미치지 않음을 시사한다. SIFT 또한 BFMatcher와 FLANNMatcher에 대해 동일하게 57.6%의 정확도를 보여주었는데, 이는 SURF보다 낮지만 ORB에 비해 높은 정확도이다. 이는 SIFT가 SURF와 비교했을 때는 다소 낮은 성능을 보인다는 것을 나타낸다. ORB의 경우, BFMatcher와의 조합에서 19.3%로 낮은 정확도를 보였으며, FLANNMatcher와의 조합에서는 가장 낮은 성능인 17.3%로 나타났다. 이는 ORB가 제안하는 실험의 조건에서는 다른 두 특징 검출기에 비해 상대적으로 낮은 성능을 가짐을 의미한다. 또한, KnnMatch 함수에서 매칭 쌍의 거리 비율을 이용한 필터링은 효과적으로 잘못된 매칭을 줄이고 정확한 매칭 결과를 도출하는 데 큰 역할을 하는 것을 확인할 수 있다.

표 5. 특징 추출기와 매칭기 조합의 성능 비교
Table 5. Performance Comparison of Combination with Feature Descriptor and Matcher

	SIFT[14]	SURF[15]	ORB[16]
BFMatcher	0.576	**0.740**	0.193
FLANNMatcher	0.576	**0.740**	0.173

그림 4는 제안 방법의 최종 결과로, 제안하는 ESD-YOLOv8 모델을 사용하여 객체의 위치를 인식하고, 인식된 각 객체는 빨간색 상자로 시각적으로 구분하여 나타낸 것이다. 이어서, 각각의 진열대에 대한 상품 데이터에 있는 상품 영상들과 비교를 진행하고 가장 높은 유사도를 보이는 상품의 이름을 감지한 객

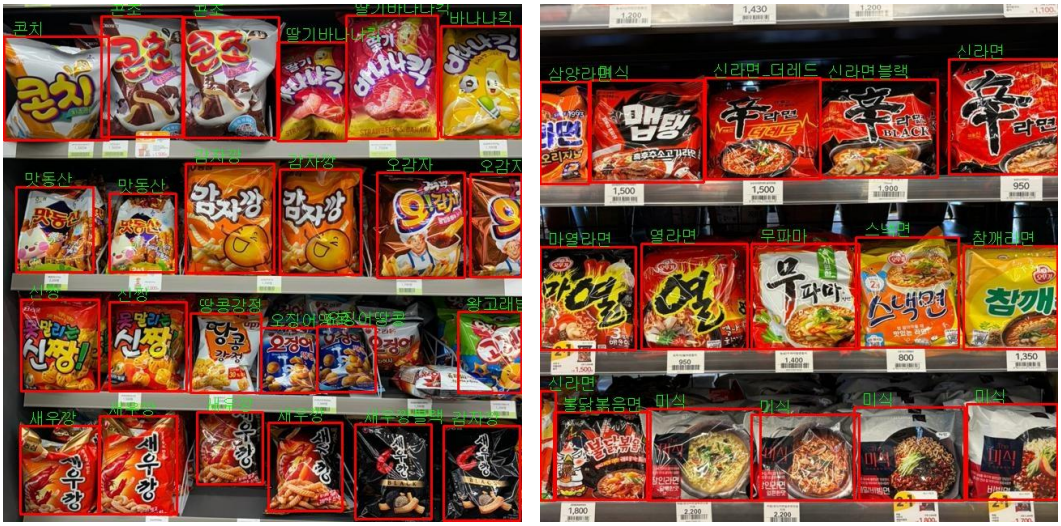


그림 4. ESD-YOLOv8과 특징점을 이용한 진열대 상품의 검출 결과 예시

Fig. 4. Example of Product Detection on Shelves Using ESD-YOLOv8 and Feature Points

체 위에 초록색으로 표기한 것이다. 이를 통해, 객체의 위치가 한 번 인식되면, 추가적인 학습 과정 없이도 상품 후보군 데이터에만 정보를 저장함으로써, 새로운 데이터가 추가될 때마다 효율적으로 상품명을 분류할 수 있음을 의미한다.

5. 결론 및 향후 과제

본 논문에서는 O2O 상점의 자동화를 하여 딥러닝과 영상처리 기술을 융합한 시스템을 제안하였다. 이 방법은 진열대에 상품이 추가되더라도 추가적인 학습이 필요 없다는 것이 큰 장점이다. 실험 결과, O2O 상점에 최적화된 제안 모델이 Yolov8s 대비 10%의 파라미터만으로도 높은 성능을 보임을 확인하였다. 또한, 상품명 인식을 위해 다양한 조합을 평가한 결과, 검출기로 SURF를 사용할 때 가장 높은 정확도를 나타냈으며 매칭기와의 조합은 성능 차이가 없음을 보였다.

향후 연구에서는 상품의 다양한 각도와 조명 조건에서도 견고하게 인식할 수 있는 알고리즘의 개선이 요구된다. 또한, 실시간 데이터 처리 능력과 상품 후보들을 데이터베이스로 관리하여 시스템의 범용성과 적용 범위를 넓히는 것이 중요하다. 이러한 문제점들을 해결한다면, 제안 방법이 소상공인들이 온라인과 오프라인의 경계를 넘나들며 비즈니스를 성장시키는 데 더욱 큰 도움을 줄 수 있을 것으로 기대한다.

REFERENCES

[1] https://www.sisaweek.com/news/articleView.html?idxno=155389

[2] https://www.segye.com/newsView/20221212516736

[3] D. Kim, J. Si, S. Lee, and S. Kim, "Calculation of Product Location Based on Object Detection and Product name recognition through Image Similarity Measurement", Proceedings of KIIT Conference, pp.494-495, 2023.

[4] D. Kim, J. Si, and S. Kim, "Feature Point Matching for Product Name Recognition in O2O Stores", Proceedings of KSCI Conference, pp.79-80, 2024.

[5] YOLOv8, https://github.com/ultralytics/ultralytics

[6] J. Redmon, S. Divvala, and R. Girshick, "You Only Look Once: Unified, Real-Time Object Detection", In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 779-788, 2016.

[7] W. Liu, A. C. Berg, *et al.*, "SSD: Single Shot MultiBox Detector", European Conference on Computer Vision, pp. 21-37, 2016.
[8] T. Y. Lin, P. Dollár *et al.*, "Focal loss for dense object detection", In Proceedings of the IEEE international conference on computer vision, pp. 2980-2988, 2017.
[9] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Region-Based Convolutional Networks for Accurate Object Detection and Segmentation", IEEE transactions on pattern analysis and machine intelligence, Vol 38, No. 1, pp. 142-158, 2015.
[10] R. GIRSHICK, "Fast r-cnn", Proceedings of the IEEE international conference on computer vision, pp.1440-1448, 2015.
[11] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks", Vol. 39, No. 6, pp.1137-1149, 2015.
[12] J. Si, G. Kim, J. Kim, and S. Kim, "Enhanced Location-based Facility Management in Mobile Environments using Object Recognition and Augmented Reality", The Journal of Korean Institute of Information Technology, Vol. 21, No. 11, pp. 183-192, 2023.
[13] J. Si, M. Kim, and S. Kim, "Converting Close-Looped Electronic Circuit Image with Single I/O Symbol into Netlist", The Journal of Korean Institute of Information Technology, Vol. 19, No. 8, pp. 1-10, 2021.
[14] G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints", International Journal of Computer Vision, Vol. 60, No. 2, pp. 91-110, 2004.
[15] H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: Speeded Up Robust Features", Computer Vision and Image Understanding, Vol. 110, No. 3, pp. 346-359, 2008.
[16] E. Rublee, V. Rabaud, K. Konolige and G. Bradski, "ORB: An efficient alternative to SIFT or SURF", International Conference on Computer Vision, pp. 2564-2571, 2011.
[17] M. Calonder, V. Lepetit, C. Strecha, and P. Fua. "Brief: Binary robust independent elementary features", European Conference on Computer Vision, pp. 778-792, 2010.

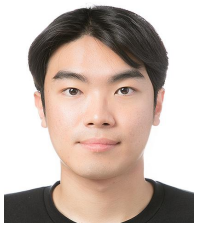
저자약력

시 종 욱 (Jongwook Si)

- 2020년 8월 : 국립금오공과대학교 컴퓨터공학과(공학사)
- 2022년 2월 : 국립금오공과대학교 컴퓨터공학과(공학석사)
- 2022년 3월~현재 : 국립금오공과대학교 컴퓨터·AI융합 공학과 대학원(박사과정수료)
- 2023년 9월~현재 : 국립금오공과대학교 인공지능공학과 강사

〈관심분야〉 : 컴퓨터비전, 영상처리, 이상감지, 생성형 AI, 딥러닝

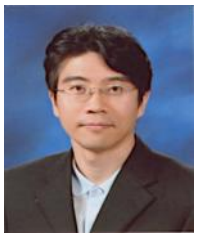
김 대 민 (Daemin Kim)



- 2020년 3월~현재 : 국립금오공과대학교 컴퓨터공학과 재학

〈관심분야〉 : 멀티미디어, 영상처리, 객체검출, 영상분할

김 성 영 (Sungyoung Kim)



- 1994년 2월 : 부산대학교 컴퓨터공학과(공학사)
- 1996년 2월 : 부산대학교 컴퓨터공학과(공학석사)
- 2003년 8월 : 부산대학교 컴퓨터공학과(공학박사)
- 2004년~현재 : 국립금오공과대학교 컴퓨터공학과 교수

〈관심분야〉 : 영상처리, 컴퓨터비전, 기계학습, 딥러닝, 메타버스