

모델 크기별 데이터 증강 비율 탐구를 통한 YOLO 기반 의류 이미지 다중 카테고리 분류 연구[☆]

Exploring Data Augmentation Ratios for YOLO-Based Multi-Category Clothing Image Classification by Model Size

박 세 연¹ 황 성 아^{1*} 장 백 철^{1*}
Seyeon Park Sunga Hwang Beakcheol Jang

요 약

최근 여러 의류 쇼핑 플랫폼 및 의류 관련 산업에서 AI를 도입하여 소비자의 니즈를 충족시키고 구매력을 높이는 체계를 도입함에 따라, 의류의 카테고리 및 색상을 정확히 분류하는 필요성이 급증하고 있다. 본 연구는 구매자 리뷰 이미지를 사용하여 한 이미지 내 여러 카테고리의 다양한 의류와 해당 색상을 분류하는 딥러닝 모델을 개발함으로써 이와 같은 문제를 해결하고자 한다. 구매자 리뷰 이미지 데이터를 직접 크롤링하여 데이터 증강 등 다양한 전처리 과정을 거친 후, YOLOv10 모델을 이용하여 의류의 객체를 탐지하고 이를 카테고리별로 분류한다. 이후, 이미지의 색상을 더 잘 추출하기 위해 고려한 크롭 방법을 통해 의류 영역을 자르고, 색상 차트와의 유사도를 계산하여 가장 유사한 색상명을 추출하는 방법을 구현한다. 실험 결과, 본 연구의 접근 방식이 효과적임을 보여주며, 모델 크기 및 증강 비율이 높을수록 성능이 향상됨을 확인하였다. 사용한 모델은 의류 카테고리 및 색상 추출에서 모두 안정적인 성능을 기록하였으며, 그 신뢰성을 입증하였다. 제안된 시스템은 사용자 리뷰 이미지를 기반으로 한 정확한 의류 카테고리 및 색상 분류를 통해 고객 만족도 및 구매력을 향상할 뿐만 아니라, 자동화된 패션 분석에 대한 추가 연구의 기초를 마련한다. 또한, 패션 트렌드 분석, 재고 관리, 마케팅 전략 수립 등 관련 산업의 여러 분야에도 활용될 수 있는 확장성을 지닌다.

☞ 주제어 : 의류 분류, 객체 탐지, 색상 추출, 하이퍼파라미터 튜닝, 이미지 증강

ABSTRACT

With the recent adoption of AI by various clothing shopping platforms and related industries to meet consumer needs and enhance purchasing power, the necessity for accurate classification of clothing categories and colors has surged. This paper aims to address this issue by developing a deep learning model that classifies various clothing items and their colors within a single image using buyer review images. After directly crawling buyer review image data and performing various preprocessing steps such as data augmentation, we utilized the YOLOv10 model to detect clothing objects and classify them into categories. Subsequently, to improve color extraction, we implemented a cropping method to isolate clothing regions in the images and calculated the similarity with a color chart to extract the most similar color names. Our experimental results show that our approach is effective, with performance increasing with model size and augmentation scale. The employed model showed stable performance in both clothing category and color extraction, proving its reliability. The proposed system not only enhances customer satisfaction and purchasing power by accurately classifying clothing categories and colors based on user review images but also lays the foundation for further research in automated fashion analysis. Moreover, it possesses the scalability to be utilized in various fields of the related industry, such as fashion trend analysis, inventory management, and marketing strategy development.

☞ keyword : clothing classification, object detection, color extraction, hyper-parameter fine-tuning, image augmentation

1. 서 론

의류 기반의 이미지를 분류하는 기술은 의류 산업의 재고 및 공급망 관리를 더욱 효율적으로 운영하도록 하며 소비자에게 정확한 검색 결과를 제공하고 맞춤형 추천을 가능하게 한다[1]. 이는 소비자의 브랜드에 대한 신뢰도 및 구매율을 높이고 기업의 이익을 증가하는 등 유통과 소비에 밀접한 관련성을 가지고 있다.

¹ Yonsei University Graduate School of Information, Seoul, Korea.

* Corresponding author (bjang@yonsei.ac.kr)

[Received 13 July 2024, Reviewed 24 July 2024(R2 08 August 2024), Accepted 30 August 2024]

[☆] This work was supported by the National Research Foundation of Korea (NRF) funded by Korean Government under Grant RS-2023-00273751.

이에 따라 여러 의류 쇼핑 플랫폼, 의류 관련 산업 및 연구에서는 AI를 도입하여 이미지에 따라 의류의 세부 카테고리 및 색상을 분류하고, 분류한 카테고리를 바탕으로 소비자가 검색한 카테고리에 해당하는 의류만을 필터링하여 보여주거나 상·하의 상품을 매칭[2]하여 소비자의 니즈를 충족시키고 구매력을 높이는 체계를 도입하고 있다[3].

그러나 현재 도입하고 있는 대부분의 시스템은 한 이미지 내 메인 상품 하나에 대한 분류[4]를 진행하여 찾고 있는 비슷한 상품군을 추천해 준다[5, 6]. 또한 실제 구매자가 촬영한 이미지를 기반으로 적용하지 않고 상품의 메인 이미지를 통하여 의류의 카테고리 및 색상을 분류하고 있다[7]. 따라서 한 가지 상품에 대한 정보를 파악할 수 있으며 이에 대한 분류 성능은 높을 수 있으나, 해당 상품에 대한 정보만을 제공하고 이미지 내 2개 이상의 의류 카테고리가 존재할 경우, 이를 찾아내지 못한다는 한계점이 존재한다. 즉, 이미지 내 여러 가지의 의류를 분류하고 상·하의와의 매칭된 스타일을 보려는 목적에는 맞지 않는 기술이라고 할 수 있다. 또한, 보정된 상품 이미지로 인하여 실제 구매자가 경험한 색상과는 차이가 있을 수 있어 소비자의 만족도를 떨어뜨리기도 한다.

따라서 본 연구에서는 구매자가 직접 촬영한 리뷰 이미지를 기반으로 한 이미지 내 존재하는 여러 가지 의류의 카테고리 및 색상을 분류하는 시스템을 개발하여, 소비자에게 한 이미지 내 여러 의류 정보를 한 번에 제공하고, 동시에 보정되지 않은 구매자가 촬영한 이미지를 기반으로 보다 정확한 의류 정보를 제공하고자 한다.

이를 통하여 소비자의 니즈를 충족시키고 구매 경험을 향상한다. 나아가, 상·하의의 매칭된 스타일에 대한 정보를 제공하거나 카메라를 인식해 현재 입고 있는 옷을 실시간 분류하는 등, 의류 산업의 여러 방면에서 이 기술은 유용하며 폭넓게 활용될 수 있을 것으로 본다.

2. 선행 연구

2.1 의류 카테고리 및 색상 분류

온라인으로 의류를 구매하는 소비자가 늘어남에 따라, 의류 추천 시스템을 위한 의류 분류 모델의 개발이 활발히 연구되는 추세이다. [8]의 연구에서는 패션 이미지 분류를 위하여 비전 트랜스포머(ViT)[9] 기반의 모델을 개발하고 이의 결과로 8가지의 패션 카테고리로 분류하여 비슷한 의류를 추천해 주는 추천시스템을 구축한다. 그

결과 F1-스코어 95%로 안정적인 성능을 달성하였다. 그러나 학습 및 평가를 진행한 이미지 데이터셋이 상품 추천 시스템을 위해 구축되었기 때문에 실제 구매자의 이미지가 아닌 상품의 메인 컷으로만 구성된 Fashion-MNIST[10] 벤치마크 데이터셋이라는 점과, 분류 카테고리의 경우 *male, women, boys, girls, clothing, footwear, sporting goods, domestic items*로 구성되어 옷의 형태에 따른 세부 카테고리 및 분류와는 거리가 멀다는 점이 본 연구의 목적과 다르다.

[11]의 연구에서는 CNN을 활용하여 *T-shirt/top, Trouser, Pullover, Dress, Coat, Sandal, Shirt, Sneaker, Bag, Ankle boot*와 같이 10가지 카테고리로 의류를 세부적으로 분류한다. 그러나 해당 모델 또한 마찬가지로 Fashion-MNIST 벤치마크 데이터셋으로 학습 및 평가를 진행하였기 때문에 한 이미지 내에 여러 가지의 의류 카테고리 및 배경 등의 노이즈가 존재하는 경우 이를 식별하기에는 적절치 않다고 판단한다.

[12]의 연구에서는 이미지 내 상의 또는 하의 하나의 객체에 대해서만 속성을 분류한다는 기존 연구의 한계점을 해결하는 것을 목표로, ResNet[13]과 EfficientNet[14] 모델을 사용하여 의류 카테고리 및 소재 속성에 대한 125개의 라벨을 이용해 의류를 분류하고 있다. 이에 추가로 색상 속성까지 함께 분류할 수 있다면 더욱 완성도 높은 의류 분류 시스템이 될 것으로 보인다.

[15]의 연구에서는 Mask R-CNN[16]을 이용하여 의류의 색상을 분류한다. 의류 색상은 *black, blue, green, magenta, pink, purple, red, yellow, white, navy(deep blue)* 총 10가지의 범주로 나누어 분류하며, 단색 옷뿐만 아니라 무늬가 있는 옷 또한 전체적인 색상을 잘 감지한 결과를 보여준다. 그러나 의류가 아닌 얼굴 영역까지 의류라고 인식하여 얼굴에 대한 색상 값을 제시하거나, 색상의 차이 및 경계가 큰 경우 옷의 전체적인 모양을 보지 않고 오직 색상만으로 구분해서 한 가지 옷에 대하여 두 가지 이상의 색상의 결과값을 나타내거나, 한 이미지 내 상·하의가 같이 있더라도 하의의 영역이 더 적은 경우 상의의 색상만을 분류한 결과를 보여주는 등 그 성능에 있어 한계점을 보인다.

2.2 객체 탐지

객체 탐지(Object Detection)란 컴퓨터 비전(Computer Vision) 분야의 중요한 태스크 중 하나이며, 이미지 혹은 영상 내에서 찾고자 하는 유의미한 객체를 찾아내는 작

업을 말한다. 이미지 내에서 어떤 객체가 탐지되는지, 탐지된 특정 객체의 개수는 얼마나 많은지를 파악하기 위한 효과적인 방법으로써 이 기술을 사용하며, 나아가서 얼굴 인식이나 비디오 추적을 하는 데에도 활용한다.

객체 탐지 과정은 두 가지 스텝으로 나누어볼 수 있다. 첫 번째는 **Regional Proposal**로 객체의 위치를 찾는 **Localization** 과정으로, 바운딩 박스를 통해 객체의 위치를 파악하는 스텝이다. 이때 바운딩 박스란 객체가 존재하는 위치를 직사각형으로 표시한 것으로 x, y 좌표로 구성된 데이터이다. 두 번째는 **Classification** 과정으로, 찾아낸 객체가 어떤 라벨에 해당하는지 분류하는 과정이다. 이 두 가지 스텝을 통해 어떤 객체가 어느 위치에서 탐지되었는지를 알아낼 수 있다.

두 가지 스텝을 어떻게 수행하냐에 따라 객체 탐지 모델은 **1-stage detector**, **2-stage detector** 두 가지 종류로 나누어볼 수 있다. **2-stage detector**란 **Regional Proposal**과 **Classification** 과정을 순차적으로 진행하는 것을 말한다. **R-CNN**[17], **Fast R-CNN**[18], **Faster R-CNN**[19], **Mask R-CNN** 등 **R-CNN** 계열 모델이 이에 해당한다. 두 가지 과정을 순서대로 진행하기 때문에 정확도는 높지만, 시간이 오래 걸린다는 단점이 있다. 반면, **1-stage detector**란 두 과정을 동시에 진행하는 것을 말한다. **SSD**[20], **RetinaNet**[21], **RefineDet**[22], **YOLO**[23] 등의 모델이 이에 해당하며, 정확도는 **2-stage detector**에 비하여 비교적 낮을 수 있지만, 빠르다는 장점이 있어 실시간 객체 탐지에도 많이 사용된다.

이러한 선행 연구를 토대로, 본 연구에서는 객체 탐지 기술을 활용하여 의류 상품의 메인 컷만이 아닌 구매자가 직접 촬영한 진신 이미지 데이터셋을 활용하여 훈련 및 평가함으로써 다양한 배경과 명암 등의 환경 변화에도 강직한 모델을 구축하고자 한다. 이때 모델은 하나의 이미지 내 상의와 하의 등 여러 카테고리에 해당하는 의류 객체를 동시에 다중으로 탐지하고, 이에 대한 세부적인 의류 카테고리 및 색상을 분류할 수 있게 한다.

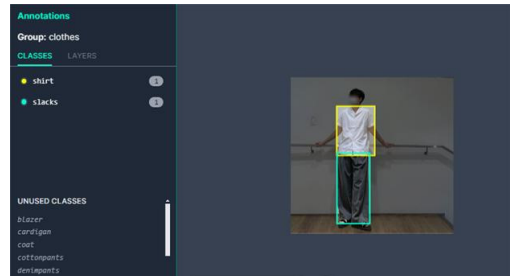
3. 연구 방법

3.1 데이터 및 전처리

3.1.1 데이터 수집

온라인 패션 플랫폼 M사는 2023년 기준 패션 업계에서 연 매출이 약 1조원을 달성하는 등 대한민국에서 영향

력 있는 이커머스 플랫폼 중 하나이며, 그만큼 다양한 상품과 구매자의 텍스트 및 이미지로 구성된 리뷰가 존재한다. 따라서 M사의 구매자 리뷰 이미지를 크롤링하여 다양한 형태의 의류에 대한 이미지를 수집한다. 이때 수집한 데이터셋은 2023년 10월 5일부터 9일 기준으로 크롤링해왔으며, 이미지 크롤링은 Python 언어 기반 **BeautifulSoup** 라이브러리를 사용하여 진행한다. 업로드된 구매자 리뷰 중 셔츠, 니트/스웨터, 맨투맨, 후드티셔츠, 긴소매 티셔츠, 반소매 티셔츠, 스커트, 데님팬츠, 코튼 팬츠, 트레이닝/조거팬츠, 슬랙스, 쇼트 팬츠, 롱코트, 재킷, 블레이저, 숏패딩/패딩베스트, 카디건을 포함한 17가지의 카테고리에 해당하는 진신 형태를 포함한 이미지 리뷰를 웹 스크래핑한다. 이때 각 카테고리에 해당되는 상품군 중 판매순으로 정렬하여 랜덤하게 약 30-40여 종의 상품을 선택한다. 이후 상품 각각에 대해 '스타일 후기'란의 1,2페이지에 해당하는 이미지 20여 장을 가져온다. 따라서 한 카테고리별 약 600-700장의 이미지가 수집되었다. 초상권 보호를 위하여 블러미(blur.me) 웹 프로그래밍을 이용하여 얼굴이 포함된 이미지에 대하여 얼굴 부분을 모자이크 처리한다.



(그림 1) 'Roboflow'를 이용한 라벨링 과정
(Figure 1) The labeling process with 'Roboflow'

3.1.2 데이터 라벨링

모자이크 처리 후, 그림 1과 같이 전체 이미지 데이터셋에 대해 '로보플로우(Roboflow)'를 이용해 바운딩박스를 생성하는 라벨링 작업을 한다. 로보플로우란 라벨링 툴로써 이미지 데이터에 대해 보다 쉽게 라벨링 및 데이터 증강, 모델의 성능을 측정할 수 있는 플랫폼이다. 라벨 값은 17가지의 의류 카테고리로, ['blazer', 'cardigan', 'coat', 'cottonpants', 'denimpants', 'hoodies', 'jacket', 'longsleeve', 'mtm', 'padding', 'shirt', 'shortpants', 'shortsleeve', 'skirt', 'slacks', 'sweater', 'trainingpants']이 이

에 해당한다. 이 라벨값을 통하여 실제 카테고리값을 정의함으로써 추후 모델을 훈련하고, 훈련된 모델을 기반으로 평가 데이터셋의 카테고리를 예측하고 해당 예측값의 정확한지를 판단하며 모델의 성능을 평가할 때 이용한다.

3.1.3 데이터 증강

라벨링된 이미지 데이터는 총 5,776개이다. 딥러닝 기반 모델로 학습시키기에는 데이터 개수가 부족하기 때문에 모델의 성능을 향상하고 과적합을 방지하기 위하여 이미지 증강(augmentation)을 진행한다. 이를 위해 학습:검증:평가 데이터셋의 비율을 8:1:1로 맞추고, 학습데이터셋에 대해 3배의 양으로 데이터를 증강했다. 데이터 증강의 결과 예시는 그림 2와 같다. 증강 기법으로는 그림 2의 좌측 위 이미지와 같이 최대 20% 이미지 확대, 그림 2의 우측 위 이미지와 같이 최대 15° 회전, 그림 2의 가운데 좌측 이미지와 같이 이미지 내 25% 정도 그레이 스케일 처리, 그림 2의 가운데 우측 이미지와 같이 -25%에서 25% 사이의 노출 조절, 그림 2의 아래 좌측 이미지와 같이 최

대 6%의 픽셀에 대해 5개의 이미지 컷아웃 처리, 그림 2의 아래 우측 이미지와 같이 최대 5% 노이즈값 부여 기법을 사용하였다.

이렇게 데이터 증강을 거친 데이터셋을 사용하였으며, 이때 증강한 비율에 따라 2배 증강한 학습 이미지 데이터의 수는 9,242장, 3배 증강한 학습 이미지 데이터 수는 13,863장이다.

3.2 연구 프로세스

본 연구의 전체 프로세스는 그림 3과 같다. 전처리를 완료한 이미지 데이터셋에 대하여 학습, 검증, 평가 데이터를 8:1:1 비율로 나누고, YOLOv10 사전학습 모델을 이용하여 학습 및 파인 튜닝한다. 이때 객체 탐지 및 앞서 소개한 17가지의 의류 카테고리를 분류하도록 하며 구축한 모델을 토대로 탐지된 객체의 바운딩박스에 따라 이미지를 각각 크롭한다. 이후 크롭한 이미지 각각에 대하여 유사도를 계산하여 가장 확률이 높은 색을 추정한다. 이를 통해 최종적으로 이미지를 입력값으로 넣으면 결과값으로 의류의 카테고리 및 색상을 추출하는 시스템을 구축할 수 있다. 구체적인 방법은 다음 절에서 단계별로 설명한다.

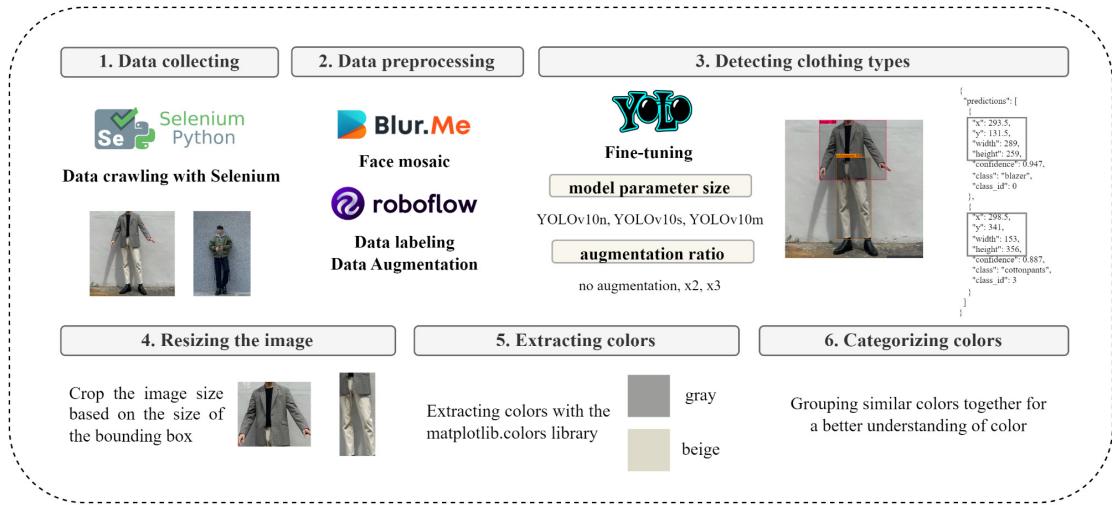
3.3 의류 객체 탐지 및 카테고리 분류

본 연구에서는 객체 탐지 모델로 Ultralytics의 YOLO 모델을 사용한다. YOLO는 CNN 기반의 객체 탐지 모형으로, 객체 탐지 관련 TASK에서 자주 사용하는 모델이다. YOLO는 1-stage detector 형태로 빠르다는 장점을 가지는 동시에, 다른 모델과 비교했을 때도 경쟁력 있는 성능을 보여준다. YOLO는 2016년 v1으로 시작하여, 2024년 5월에 출시된 YOLOv10[24]이 존재한다. 모델의 파라미터의 수에 따라 YOLOv10n, YOLOv10s, YOLOv10m 등의 종류가 존재하고, n이 가장 적은 크기의 파라미터를 가지며 속도가 빠르다. 반면, m일수록 파라미터의 크기가 커지므로 속도는 느려질 수 있으나 성능이 좋다는 특징을 가진다.

본 연구에서는 사전 학습된 YOLOv10을 사용하여 의류 객체의 위치와 카테고리를 탐지하는 모델을 구축한다. 구체적으로 YOLOv10n, YOLOv10s, YOLOv10m의 여러 크기의 모델을 다양한 파라미터 설정으로 비교하여 최적의 모델을 구축한다.



(그림 2) 이미지 증강 기법 적용
(Figure 2) Image augmentation results



(그림 3) 연구의 전체 프로세스
(Figure 3) The overall process of the research

3.4 탐지된 객체에 대한 이미지 크롭

객체 탐지가 완료된 이미지는 탐지된 객체가 속하는 의류 카테고리 및 이에 대한 바운딩박스 영역이 데이터로 저장된다. 이때 전신 사진의 경우 하나의 이미지 내에 2개 이상의 의류가 탐지될 수 있기 때문에, 이 각각의 의류 객체에 대한 색상을 추출하기 위해서는 각 객체의 바운딩박스 위치에 따라 이미지를 크롭(crop)하는 과정이 필요하다. 그러므로 YOLO의 내장함수인 `save_crop=True`를 이용하여 저장된 객체의 바운딩박스 영역에 따라 이미지를 크롭한다.

3.5 의류 색상 추출 및 색상 계열 분류

의류의 색상을 추출하는 과정에서는 크롭한 이미지 중 가장 많은 색상을 위주로 전체 색상을 결정하게 된다. 따라서 크롭 단계에서 사각형 형태의 바운딩박스 영역만큼 이미지를 자른다면, 의류와 배경 영역이 차지하는 비율이 비슷한 경우 다음 단계인 의류 색상 추출 과정에서 추출한 색상 결과값이 옷이 아닌 배경색으로 잘못 추출될 가능성이 존재한다. 이를 해결하고 이미지 내 의류의 색상을 더 잘 추출하기 위하여, 크롭한 이미지의 가로, 세로 길이의 80%로 가로 및 세로 길이를 조정하여 크롭한다. 이를 통하여 탐지된 의류 바운딩박스 안쪽 영역으로 크기를 줄여 자름으로써 전체 크롭한 영역 중 의류 색상 영

역의 비중을 크게 만들어 이에 대한 색상을 더 잘 추출하도록 만든다.

이후 크롭한 하나의 의류 이미지를 가지고 해당 의류의 색상을 추출한다. 색상 추출 방법은 파이썬의 색상을 추출하는 `extcolors` 라이브러리를 이용하여 픽셀의 RGB 값을 구하는 방식을 사용한다. `extcolors` 라이브러리의 `extract_from_image` 함수를 사용한다. 이 함수는 CIE76 알고리즘을 이용하여 색상 간 유사도를 파악하는 방식이다. 이를 사용하여 크롭한 이미지를 인풋값으로 넣어 이에 대한 결과값인 색상 RGB값을 얻는다.

이후 추출한 색상 RGB값을 `matplotlib.colors` 라이브러리의 `XKCD_COLORS` 색상 차트와 비교하여 가장 유사한 색상명을 추출하게 된다. 이때 이미지의 RGB 색상 형식과 맞추기 위하여 `XKCD_COLORS`의 HEX로 된 색상 형식을 RGB로 변환한다. `XKCD` 색상 차트는 총 949가지의 컬러명으로 구분된 컬러차트이다. 이를 활용하여 이미지 내 다양한 색상을 구분하게 된다.

본 연구에서는 더 적은 가짓수의 컬러명으로 구분된 `matplotlib.colors` 라이브러리의 `CSS4_COLORS` 색상 차트도 사용하여 결과를 비교해보았다. 그 결과 명도에 채도에 따라 비슷한 계열의 색상으로 분류하지 못하여 추출한 색상명이 육안으로 보았을 때 수용하기 어려운 경우가 대부분이었다. 따라서 본 연구에서는 의류 색상 추출 방법 과정에서 `XKCD` 색상 차트를 이용하여 보다 세부적

인 색상명으로 나타낸다.

$$\min(R^1 - R^2)^2 + (G^1 - G^2)^2 + (B^1 - B^2)^2 \quad (1)$$

XKCD의 RGB 각각의 값을 R^1, G^1, B^1 라 하고, 추출한 RGB값을 R^2, G^2, B^2 라 할 때 유사도를 구하여 색상명 추출 방식은 수식 1과 같다. XKCD 색상 차트의 각 색상별 RGB값과 이미지에서 추출한 RGB값에서 각 R값, G값, B값의 차이를 구한 후 제곱을 하여 합산하고, 그 값이 최소인 색상을 선택한다. 이렇게 나온 색상명은 결과값 데이터로 저장된다.

나아가 의류 색상 추출 결과로 얻게 된 의류의 색상 값에 대하여 계열 분류를 통해 비슷한 색상 군끼리 묶어 색상에 대한 이해력을 기존 대비 높일 수 있다.

앞서 설명한 의류 색상 추출 방법의 경우 XKCD 색상 차트를 이용하여 949가지의 색상명으로 분류한다. 이는 세부적인 색상명을 알기에는 적합하나, 육안으로 보기에 비슷한 색상마저도 다른 색상명, 즉 다른 카테고리로 분류하기 때문에 사용자가 비슷한 색상 군끼리 의류를 필터링하거나 확인하고 싶을 경우에는 어려움이 존재한다.

따라서 상위 분류 색상명인 11가지 및 이외 색상 카테고리 나누어 기존 세부 색상을 더 큰 범위로 묶어 분류한다. 상위 분류 카테고리는 의류 색상 중 가장 많은 black, white, grey, green, blue, orange, red, purple, pink, brown, yellow로 분류하고, 이에 속하지 않는 이외의 색은 데이터 중 차지하는 비중이 작으므로 모두 others에 속하도록 한다.

4. 실 험

4.1 실험 환경

실험에서 사용한 환경 및 하이퍼파라미터 설정은 다음과 같다. 배치 크기를 16로 설정하고, patience는 10, 최대 에포크 수는 100으로 설정한다. 이미지 사이즈는 640×640 으로, 옵티마이저는 'auto'로, 학습률은 0.01를 사용한다. 실험에서 YOLOv10 중 YOLOv10n, YOLOv10s, YOLOv10m를 사용하여 파라미터 수에 따른 변화를 살펴본다. 또한 이미지 증강의 효과 정도를 파악하기 위하여 증강 배율에 따라 증강하지 않은 이미지와 이미지 수를 2배 증강한 데이터셋, 이미지 수를 3배 증강한 데이터셋으로 변화를 주어 실험 결과를 비교 분석한다.

4.2 성능 평가지표

객체 탐지를 위한 평가 지표로는 분류 성능 평가지표인 Precision, Recall과 mAP를 주로 사용한다.

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$IoU = \frac{(PBA) \cap (ABA)}{(PBA) \cup (ABA)} \quad (4)$$

Precision이란 정밀도로 분류 모델이 positive로 판단한 것 중 실제 positive인 샘플의 비율을 말한다. 반면 Recall이란 재현율로 실제 positive 샘플 중 분류 모델이 positive로 판단한 비율을 의미한다. 이때 TP는 True Positive로 실제 양성을 양성이라 예측한 값, FP는 False Positive로 실제 음성을 양성으로 예측한 값, FN은 False Negative로 실제 양성인 값을 음성으로 예측한 값을 말한다. 이는 수식 2와 3을 통해 계산 가능하다.

mAP란 mean Average Precision이다. AP 곡선(Average Precision Curve)은 정밀도와 재현율을 고려한 종합 평가 지표로, Precision과 Recall Curve로 이루어진 그래프의 아래 영역이다. 실제 AP값은 0에서 1 사이값으로 이루어지며 1에 가까울수록 더 정확하다고 표현할 수 있다. 따라서 AP 값의 평균을 구한 값을 mAP라고 한다. mAP50은 IoU 임계값이 0.5에서 계산된 mAP를 말하는데, 이때 IoU 값이란 수식 4와 같이, 객체 탐지에서 예측한 바운딩박스 영역과 실제 객체의 영역의 교집합 부분을 두 합집합 부분으로 나눈 값이다. 이때 수식 4에서 PBA란 Prediction Bounding box Area로 예측 바운딩박스의 영역을 말하며, ABA란 Actual Bounding box Area로 실제 바운딩박스의 영역을 의미한다. 따라서 이 값이 1에 가까울수록 예측과 실제 영역이 일치한다는 것이므로 그만큼 예측 정확도가 높다고 판단한다. mAP50-95는 IoU 임계값을 0.5~0.95 범위로 구성하여 계산된 mAP를 의미한다. 이를 통하여 다양한 객체 탐지의 난이도를 고려한 모델의 성능을 종합적으로 파악할 수 있다.



(그림 4) 객체 탐지 및 분류 결과 예시
(Figure 4) Examples of object detection and classification results

4.3 실험 결과

4.3.1 의류 객체 탐지

3장의 연구 방법론에 따라 전처리된 데이터를 사전학습된 YOLOv10 모델로 학습하여 인풋값의 이미지 내에 어떤 카테고리의 의류가 어느 위치에서 탐지되었는지 찾아낸다. 이후 학습된 모델의 성능을 평가하여 성능이 좋은 모델을 최종 사용할 모델로 결정한다.

의류 객체 탐지 및 카테고리 분류 결과는 그림 4와 같다. 이미지 내 대부분의 의류에 대하여 알맞은 카테고리로 분류하였으며 카테고리명 우측의 mAP값 역시 0.9 이상을 기록한 사례가 많음을 알 수 있다.

표 1과 같이 YOLOv10n, YOLOv10s, YOLOv10m 등 파라미터 수에 변화를 준 다양한 모델을 기반으로 학습 및 평가하였으며, 이때 augmentation size에 변화를 준 조합으로 성능을 비교한다.

표 1에서 알 수 있듯이, 증강 전후의 성능 차이를 비교해 보면 증강 후의 성능이 같은 하이퍼파라미터 및 모델을 사용했을 때 증강 전의 성능보다 mAP50 값이 향상된 것을 볼 수 있다. 2배로 증강한 데이터셋의 경우 증강을 하지 않은 데이터셋으로 학습한 결과보다 mAP50의 경우 약 13%p 향상하였으며, mAP50-95의 경우 약 8%p 향상하였다. 또한 2배 증강한 데이터셋보다 3배 증강한 데이터셋으로 학습한 결과 mAP50에서 약 3%p 정도의 향상이

있었으며, mAP50-95의 경우 2%p의 향상도를 보여주었다. 이로써 증강을 한 데이터셋이 증강을 적용하지 않은 데이터셋보다 향상된 성능을 보여주며, 원본 데이터셋 개수의 2배 증강한 것보다 3배 증강했을 때 대체적으로 더 나은 성능임을 알 수 있다.

(표 1) 파라미터 수 및 하이퍼파라미터 튜닝 조합에 따른 mAP 성능 비교

(Table 1) Comparison of mAP performance for different combinations of parameter numbers and hyperparameter tuning

Model	#Params	Augmentation size	mAP50	mAP50-95
YOLOv10-N	2.3M	None	0.668	0.612
		x2	0.763	0.666
		x3	0.760	0.661
YOLOv10-S	7.2M	None	0.682	0.630
		x2	0.788	0.694
		x3	0.813	0.714
YOLOv10-M	15.4M	None	0.724	0.651
		x2	0.785	0.691
		x3	0.834	0.736

(표 2) YOLOv10m 튜닝 모델의 의류 카테고리별 성능 결과
(Table 2) Performance results by apparel category for the YOLOv10m tuned model

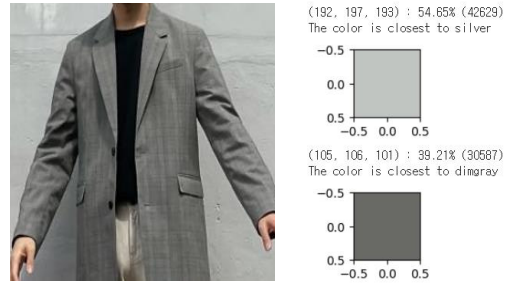
Class	Precision	Recall	mAP50	mAP50-95
blazer	0.931	0.792	0.919	0.804
cardigan	0.822	0.713	0.841	0.747
coat	0.747	0.833	0.862	0.785
cottonpants	0.887	0.688	0.859	0.767
denimpants	0.873	0.737	0.907	0.786
hoodies	0.711	0.457	0.682	0.587
jacket	0.789	0.504	0.738	0.655
longsleeve	0.854	0.544	0.802	0.703
mtm	0.878	0.6	0.799	0.712
padding	0.614	0.833	0.866	0.768
shirt	0.952	0.508	0.76	0.679
shortpants	0.919	0.75	0.874	0.712
shortsleeve	<u>0.958</u>	<u>0.907</u>	<u>0.978</u>	<u>0.856</u>
skirt	0.826	0.757	0.87	0.682
slacks	0.867	0.636	0.847	0.73
sweater	0.728	0.474	0.683	0.621
trainingpants	0.751	0.478	0.723	0.655
all	<u>0.837</u>	<u>0.623</u>	<u>0.834</u>	<u>0.736</u>

결과를 살펴보면, 가장 높은 mAP값을 달성한 모델 및 하이퍼파라미터 튜닝 조합은 YOLOv10m을 사용한 모델로 원본 개수보다 3배 증강한 데이터를 사용하였을 때였으며, 이때의 mAP50값은 0.834를 기록하였다.

해당 모델에 대하여 의류의 카테고리별 객체 탐지 결과 성능을 비교한 결과는 표 2와 같다. 전체 클래스 평균 precision 0.837, recall 0.623, mAP50 0.834, mAP50-95 0.736을 나타냈으며, 클래스별 성능의 경우 모든 클래스에 대하여 mAP50의 경우 0.68 이상, mAP50-95의 경우 0.62 이상의 결과값을 보여준다.

4.3.2 의류 색상 추출

의류 색상 추출 결과는 그림 5와 같다. 그림 5의 우측 위의 색상은 첫 번째로 가장 높은 유사도를 나타낸 색상이고, 우측 아래의 색상은 두 번째로 높은 유사도를 띤 색상이다. 결과를 보여주기 위하여 유사도가 높은 두 가지의 색상을 추출하였으며, 실제 결과로 저장되는 데이터는 유사도가 가장 높은 'silver' 색상이 저장된다.



(그림 5) 의류 색상 추출 예시

(Figure 5) Example of extracting apparel colors

(표 3) 의류 색상 계열 분류에 대한 분류 성능 결과
(Table 3) Performance for color range classification of apparel

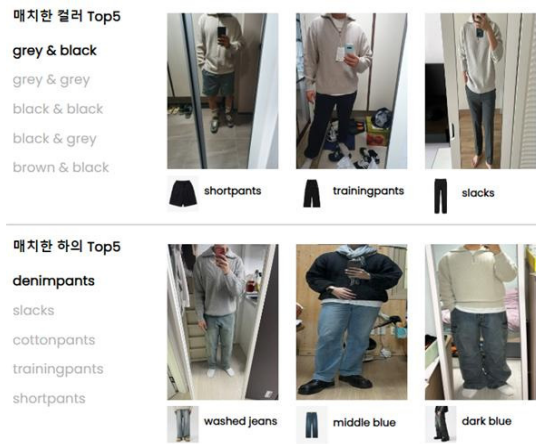
Color	True	False	Accuracy
black	42	8	0.84
white	<u>48</u>	<u>2</u>	<u>0.96</u>
grey	41	9	0.82
blue	47	5	0.94
orange	46	4	0.92
red	47	3	0.94
purple	41	9	0.82
pink	46	4	0.92
brown	42	8	0.84
yellow	45	5	0.90
others	44	6	0.88

추출한 색상의 계열 분류에 대한 정확도는 표 3과 같다. 분류한 각 색상 계열별로 50개의 이미지를 테스트셋으로 설정하여 others 색상을 포함한 총 12가지의 분류 카테고리에 대한 성능을 평가한다. 3명의 검토자가 분류된 이미지와 색상 계열에 대하여 육안으로 잘 분류되었는지의 여부를 판단한다. 3명이 모두 색상 계열이 옳게 분류되었다고 판단할 경우에만 잘 분류되었다는 결론을 내리며, 한 사람이라도 다른 색상으로 판단할 경우 해당 이미지에 대한 색상은 옳지 않게 분류된 것으로 본다. 최종적으로 각 색상에 대한 정확도는 옳게 분류된 이미지 수를 전체 이미지 수인 50으로 나눈 값으로 산출한다.

그 결과, 모든 색상에 대한 분류 정확도의 평균은 약 0.86 정도로 안정적인 성능을 보여주며, 흰색(white)에 대해서는 0.96의 가장 높은 정확도를 나타냈다.

4.4 시스템 및 활용 방안

전체 프로세스를 거친 이미지의 결과값은 다음과 같다. 이미지 입력값이 주어지면 이에 대하여 의류의 카테고리 분류하고 및 탐지된 의류의 영역에 따라 이미지가 크롭되고, 크롭된 이미지를 토대로 색상을 유추하게 된다. 따라서 결과값은 이미지 내 탐지된 의류 개수만큼의 의류 카테고리 및 세부 색상이 나타난다.



(그림 6) 리뷰 웹페이지 개발 예시

(Figure 6) Example of developing a review webpage

본 연구에서는 제시한 시스템의 활용 인사이트를 제공하기 위하여 그림 6과 같은 서비스를 고안해 보았다. 특정 의류에 대하여 많은 구매자가 매칭한 상하의 카테고리 및 색 조합을 나타내주는 웹 페이지 서비스이다. 웹 페이지는 HTML, CSS, Javascript 언어로 개발을 진행하였으며 매칭한 의류의 색상 및 종류를 한눈에 알기 쉽게 정리하여 보여줌으로써 소비자에게 특정 의류의 매칭 스타일을 보여주고 결과적으로 소비자의 사용자 경험을 증진하고 구매력의 향상 효과를 기대해 볼 수 있다.

5. 결론 및 향후 연구

본 연구에서는 객체 탐지 기술을 활용하여 구매자 리뷰 이미지를 기반으로 다중 의류 카테고리 및 색상을 분류하는 딥러닝 모델을 개발하고 이를 평가하였다. 이를 위하여 실제 온라인 패션 플랫폼 M사의 구매자 리뷰 이미지를 크롤링해 얼굴을 모자이크 처리하고 라벨링을 진행하였다. 이후 다양한 증강 기법을 이용하여 원본 이미

지 대비 2배 혹은 3배의 양으로 데이터를 증강하였다. 이 데이터를 이용하여 CNN 기반의 YOLOv10 모델을 통하여 의류 객체를 탐지 및 카테고리를 인식하고, 탐지된 객체의 영역에 따라 이미지를 크롭한 후 의류의 색상을 추출한다.

연구 결과, 본 모델은 다중 의류 카테고리 및 색상 분류에서 높은 정확도를 기록하였다. 특히, 다양한 카테고리 및 색상 조합에서도 일관된 성능을 보임으로써 그 실용성을 입증하였다.

본 연구의 의의는 다음과 같다. 첫째, 구매자 리뷰 이미지를 활용함으로써, 실제 사용자 환경에서의 모델 적용 가능성을 높였다. 둘째, 다중 카테고리 및 색상 분류 문제를 동시에 해결하는 모델을 제안함으로써, 의류 추천 시스템 등 다양한 응용 분야에서 활용될 수 있는 가능성을 제시하였다.

본 연구를 통해 개발된 모델은 온라인 쇼핑 플랫폼에서의 사용자 경험을 향상할 수 있으며, 보다 정확한 의류 추천 및 검색 기능을 제공할 수 있다. 이를 통해 기업은 판매 증대 및 고객 만족도를 높일 수 있을 것이다. 소비자 또한 한눈에 보기 쉬운 스타일 매치 등의 이미지로부터 추출한 인사이트 정보를 바탕으로 구매 의사결정에 도움을 받을 수 있다. 그뿐만 아니라 본 연구는 다른 분야로도 큰 확장 가능성을 가진다. 패션 트렌드 분석, 재고 관리, 마케팅 전략 수립 등에서 활용될 수 있을 것이며, 나아가 패션 플랫폼 시장 전반에 긍정적인 영향을 미칠 것으로 기대된다.

본 연구에서 개발한 시스템의 경우 객체 탐지 모델의 mAP50값은 0.834, mAP50-95값은 0.736이다. 탐지 성능이 낮은 클래스에 대하여 더욱 다양한 이미지를 추가 수집 및 증강하고, 더 효과적인 모델을 개발하여 모든 클래스에 대하여 시스템의 성능을 개선하는 것은 큰 의의를 가질 것으로 생각한다. 나아가 사용한 증강 기법에 대한 신뢰성 테스트를 진행하여 더욱 효과적인 증강 정도 또는 기법에 대한 연구가 필요할 것이다. 또한, 현재의 경우 색상 추출 과정에서 단색이 아닌 패턴이 있는 의류에 대하여 전체적인 색상을 정확히 추출하지 못하는 경우가 존재함을 확인하였다. 따라서 다양한 패턴과 질감을 고려한 색상 추출 방법에 대한 연구도 향후 필요할 것으로 보인다.

참고문헌(Reference)

[1] Shajini, Majuran, and Amirthalingam Ramanan, "An improved landmark-driven and spatial - channel attentive

- convolutional neural network for fashion clothes classification,” *The Visual Computer*, Vol. 37, No. 6, pp. 1517-1526, 2021.
<https://doi.org/10.1007/s00371-020-01885-7>
- [2] Hye-Suk Kim, “Classification and Combination of Fashion Items Using CNN-Based Deep Learning,” *Journal of Digital Contents Society*, Vol. 22, No. 3, pp. 475-482, 2021.
<http://dx.doi.org/10.9728/dcs.2021.22.3.475>
- [3] JW Yang, “Analysis of consumer purchasing behavior models according to contents of fashion shopping platform applications : Focused on the top three women’s clothing companies users’ rankings,” Master’s Thesis, Hongik University, 2022.
<https://dcoil.hongik.ac.kr/srch/srchDetail/000000027932>
- [4] Cychnerski, Jan, et al., “Clothes detection and classification using convolutional neural networks,” 2017 22nd IEEE international conference on emerging technologies and factory automation (ETF A), IEEE, 2017.
<https://doi.org/10.1109/ETF A.2017.8247638>
- [5] Abd Alaziz, Hadeer M., et al., “Enhancing Fashion Classification with Vision Transformer (ViT) and Developing Recommendation Fashion Systems Using DINOVA2,” *Electronics*, Vol. 12, No. 20, pp. 4263, 2023.
<https://doi.org/10.3390/electronics12204263>
- [6] Sivaranjani, Lingala, et al., “Fashion Recommendation System Using Machine Learning,” 2023 4th International Conference on Smart Electronics and Communication (ICOSEC), IEEE, 2023.
<https://doi.org/10.1109/ICOSEC58147.2023.10275967>
- [7] Nocentini, Olivia, et al., “Image classification using multiple convolutional neural networks on the fashion-MNIST dataset,” *Sensors*, Vol. 22, No. 23, pp. 9544, 2022.
<https://doi.org/10.3390/s22239544>
- [8] JH Cho, HS Kwon, Yoon KH, “Perspectives of ‘Ubiquitous Health Care System’ for Diabetes Management,” *Diabetes & Metabolism Journal*, Vol. 30, No. 2, pp. 87-95, 2006.
<http://dx.doi.org/10.4093/jkda.2006.30.2.87>
- [9] Dosovitskiy, Alexey, et al., “An image is worth 16x16 words: Transformers for image recognition at scale,” arXiv, 2020.
<https://doi.org/10.48550/arXiv.2010.11929>
- [10] Xiao, Han, Kashif Rasul, and Roland Vollgraf, “Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms,” arXiv, 2017.
<https://doi.org/10.48550/arXiv.1708.07747>
- [11] Mohanty, D. K. et al., “Modified Convolutional Neural Network for Fashion Classification,” 2022 International Conference on Advancements in Smart, Secure and Intelligent Computing (ASSIC), pp.1-9, 2022.
<https://doi.org/10.1109/ASSIC55218.2022.10088358>
- [12] HS Jeong, SY Lee, CK Lee, “Deep learning-based clothing attribute classification using fashion image data,” *Smart Media Journal*, Vol. 13, No. 4, pp. 57-64, 2024.
<http://doi.org/10.30693/SMJ.2024.13.4.57>
- [13] He, Kaiming, et al., “Deep residual learning for image recognition,” *Proceedings of the IEEE conference on computer vision and pattern recognition(CVPR)*, 2016.
<https://doi.org/10.48550/arXiv.1512.03385>
- [14] Tan, Mingxing, and Quoc Le, “Efficientnet: Rethinking model scaling for convolutional neural networks,” *International conference on machine learning(ICML)*, PMLR, pp. 6105-6114, 2019.
<https://doi.org/10.48550/arXiv.1905.11946>
- [15] Guo, Calvin H., “Search My Favorites by Color: Fashion Parsing through Color Classification,” *CS230: Deep Learning*, Winter 2020, Stanford University, CA, 2020.
- [16] He, Kaiming, et al., “Mask r-cnn,” *Proceedings of the IEEE international conference on computer vision (ICCV)*, pp. 2961-2969, 2017.
<https://doi.org/10.48550/arXiv.1703.06870>
- [17] Girshick, Ross, et al., “Rich feature hierarchies for accurate object detection and semantic segmentation,” *Proceedings of the IEEE conference on computer vision and pattern recognition(CVPR)*, pp. 580-587, 2014.
<https://doi.org/10.48550/arXiv.1311.2524>
- [18] Girshick, Ross, “Fast r-cnn,” *Proceedings of the IEEE international conference on computer vision(ICCV)*, pp. 1440-1448, 2015.
<https://doi.org/10.48550/arXiv.1504.08083>
- [19] Ren, Shaoqing, et al., “Faster r-cnn: Towards real-time object detection with region proposal networks,” *Advances in neural information processing systems*

- (NeurIPS), 2015.
<https://doi.org/10.48550/arXiv.1506.01497>
- [20] Liu, Wei, et al., "Ssd: Single shot multibox detector," Computer Vision - ECCV 2016: 14th European Conference, Part I 14. Springer International Publishing, 2016.
<https://doi.org/10.48550/arXiv.1512.02325>
- [21] Lin, Tsung-Yi, et al., "Focal loss for dense object detection," Proceedings of the IEEE international conference on computer vision(ICCV), pp. 2980-2988, 2017.
<https://doi.org/10.48550/arXiv.1708.02002>
- [22] Zhang, Shifeng, et al., "Single-shot refinement neural network for object detection," Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), pp. 4203-4212, 2018.
<https://doi.org/10.48550/arXiv.1711.06897>
- [23] Redmon, Joseph, et al., "You only look once: Unified, real-time object detection," Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), pp. 779-788, 2016.
<https://doi.org/10.48550/arXiv.1506.02640>
- [24] Wang, Ao, et al., "Yolov10: Real-time end-to-end object detection," arXiv, 2024.
<https://doi.org/10.48550/arXiv.2405.14458>

● 저 자 소 개 ●



박 세 연(Seyeon Park)

2018년~2023년 동국대학교 경영정보학과 학사
2023년~현재 연세대학교 정보대학원 비즈니스 빅데이터 분석 트랙 석사과정
관심분야 : Computer Vision, Deep Learning, Bigdata Analytics
E-mail : seyeon@yonsei.ac.kr



황 성 아(Sunga Hwang)

2019년~2023년 연세대학교 미래캠퍼스 정보통계학과 학사
2023년~현재 연세대학교 정보대학원 비즈니스 빅데이터 분석 트랙 석사과정
관심분야 : Natural Language Processing, Deep Learning
E-mail : sungahwang@yonsei.ac.kr



장 백 철(Beakcheol Jang)

2009년 North Carolina State University 컴퓨터공학과(공학박사)
2021년~현재 연세대학교 정보대학원 교수
관심분야 : Natural Language Processing, Artificial Intelligence, Bigdata Analytics
E-mail : bjang@yonsei.ac.kr