

## Automatic Fashion Item Labeling System Using YOLO and a High-Level Object Detection Model

Jun-oh Lim\*, Woo-jin Choi\*\*, Bong-jun Choi\*\*\*

\*Student, Dept. of Software, Dongseo University, Busan, Korea

\*\*Ph. D., School of Fashion and Textiles, Hong Kong Polytechnic University, Hung Hom, Hong Kong

\*\*\*Professor, Dept. of Software, Dongseo University, Busan, Korea

### [Abstract]

This paper propose an automatic labeling system for fashion items in images by combining one of the object detection models, YOLO(You Only Look Once), with a high-level classification object detection model. After detecting the primary fashion items, TOP and BOTTOM, in an image, the system analysis the bounding boxes of the detected objects and removes redundant or unnecessary bounding boxes through preprocessing to extract bounding boxes with accurate location information. The extracted bounding boxes are compared to the classes defined by the high-level object detection model with coordinate normalization to perform automatic labeling by matcing the input fashion item types. The system's performance was evaluated on 10,000 fashion images and corresponding test data, and 8,192 images were found to be correctly labeled. This demonstrates a significant improvement in efficiency over manual labeling methods, showing the system's practical contribution to large-scale fashion image data processing.

▶ **Key words:** Labeling, Fashion, YOLO(You Only Look Once), Bounding Box, Object Detection

### [요 약]

본 논문에서는 객체 탐지 모델 중 하나인 YOLO(You Only Look Once)와 대분류 객체 탐지 모델을 결합하여, 이미지 속 패션 아이템의 자동 레이블링 시스템을 제안한다. 본 시스템은 이미지 내 주요 패션 아이템인 상의(TOP)와 하의(BOTTOM)을 탐지한 후, 탐지된 객체의 경계 상자를 분석하고 전처리 과정을 통해 중복되거나 불필요한 경계 상자를 제거함으로써, 정확한 위치 정보를 가진 경계 상자를 추출한다. 추출된 경계 상자는 좌표 정규화와 함께 대분류 객체 탐지 모델에서 정의한 클래스와 비교하여, 입력된 패션 아이템 종류와 매칭함으로써 자동 레이블링을 수행한다. 10,000장의 패션 이미지와 텍스트 데이터로 본 시스템의 성능을 평가한 결과, 8,192장의 이미지에서 정확한 레이블링을 확인하였다. 이러한 결과는 기존 수작업 레이블링 방식보다 효율성을 크게 향상시켰으며, 대규모 패션 이미지 데이터 처리에 있어 실질적인 기여를 할 수 있음을 보여준다.

▶ **주제어:** 레이블링, 패션, YOLO, 경계 상자, 객체 탐지

- First Author: Jun-oh Lim, Corresponding Author: Bong-jun Choi
- \*Jun-oh Lim (wnsdh03219@gmail.com), Dept. of Software, Dongseo University
- \*\*Woo-jin Choi (woojin.choi@polyu.edu.hk), School of Fashion and Textiles, Hong Kong Polytechnic University
- \*\*\*Bong-jun Choi (bongjun.choi@dongseo.ac.kr), Dept. of Software, Dongseo University
- Received: 2024. 09. 27, Revised: 2024. 10. 25, Accepted: 2024. 10. 28.

## I. Introduction

최근 패션에 대한 관심이 급증함에 따라 전 세계적으로 패션 및 의류 시장이 빠르게 성장하고 있다. 이에 다양한 패션 커머스 플랫폼이 등장하고 있으며, 소비자들의 온라인 패션 소비를 촉진하기 위해 패션 아이템의 디지털 전환이 가속화되고 있다[1]. 이러한 환경에서 방대한 패션 이미지를 효율적으로 관리하고 활용하는 것이 중요한 이슈로 떠오르고 있으며, 특히 온라인 서비스에서는 인공지능(AI) 기술을 활용한 자동화가 필수 요소로 자리 잡고 있다[2].

패션 아이템은 비정형 데이터에 속하며, 이를 정형 데이터인 텍스트와 결합하는 과정에서 레이블링 작업이 필요하다. 그러나 현재 대부분의 레이블링 작업은 수작업에 의존하고 있어, 데이터셋의 규모가 커질수록 막대한 시간과 인력, 비용이 소요된다. 대표적인 사례로, ILSVRC 대회 ImageNet 데이터 구축 과정에서는 약 4만 9천여 명의 인력이 동원되어 1,400만 장의 이미지에 대한 레이블링 작업이 4년에 걸쳐 이루어진 바 있다. 패션 분야 역시 대규모 데이터와 세부적인 카테고리(스타일, 색상, 패턴, 소재 등)를 처리해야 하므로, 레이블링 작업의 복잡성이 매우 높다.

또한, 이러한 수작업 방식은 주관적인 판단에 의존해 일관성과 정확성을 유지하기 어려우며, 결과적으로 인공지능 학습의 질을 저하시킬 수 있다. 이에 따라 YOLO(You Only Look Once)와 같은 객체 탐지 모델을 활용한 자동화된 레이블링 시스템에 대한 연구가 활발히 진행되고 있다[3]. YOLO는 실시간 객체 탐지에 적합한 모델로, 이미지 내 객체의 위치와 클래스를 동시에 예측할 수 있어 높은 처리 속도를 제공한다. 그러나 YOLO를 중심으로 한 기존 연구들은 주로 일반적인 객체 탐지에 초점을 맞추었기 때문에, 패션 아이템의 특수성을 반영한 레이블링에는 한계가 있다. 특히, 세부적인 패션 아이템 종류 구분이나 경계 상자의 정밀한 분석에 대한 연구는 부족하였다.

이외에도, 패션 아이템 레이블링 자동화 분야에서 기존 연구들은 DeepFashion과 같은 대규모 데이터셋을 활용하여 의류 및 액세서리의 착용 위치나 스타일을 분석하는 기술을 제안하였다. 하지만, 이들 연구는 주로 고정된 카테고리 내에서의 분류에 중점을 두었으며, 세부적인 패션 아이템 특징을 고려한 정확한 경계 상자 추출이나 중복된 레이블 처리에 대한 해결책은 제시하지 못했다.

본 논문에서는 이러한 기존 연구의 한계점을 극복하기 위해 YOLO 객체 탐지 모델과 대분류 객체 탐지 모델을 결합한 패션 아이템 자동 레이블링 시스템을 제안한다. 본 시스템은 패션 아이템 중 상의(TOP)와 하의(BOTTOM)를

빠른 속도로 탐지하고, 경계 상자 분석을 통해 정확한 레이블링을 수행한다. 특히, 기존 연구에서 다루지 않았던 경계 상자 분석 및 전처리 과정을 통해 중복되거나 불필요한 데이터를 필터링함으로써, 레이블링 작업의 효율성과 정확성을 크게 향상시켰다. 또한, 본 연구는 10,000장의 패션 이미지에 대해 실험을 진행하였으며, 8,192장의 이미지에서 정확한 레이블링 성능을 확인하였다.

## II. Preliminaries

### 1. Related works

#### 1.1 YOLO(You Only Look Once)

YOLO는 Single Shot 계열의 객체 탐지 모델로, 객체의 위치와 클래스를 동시에 예측할 수 있는 빠른 처리 속도를 자랑한다. 기존의 R-CNN, Fast R-CNN, Faster R-CNN과 같은 CNN 기반 객체 탐지 모델들은 이미지 내에서 객체의 후보 영역을 먼저 생성한 후, 해당 영역을 분류하는 방식으로 동작하기에 높은 정확도를 제공하지만, 처리 속도가 느리다는 단점이 있다[4].

반면, YOLO는 객체의 위치 예측과 분류 과정을 하나의 End-to-End 방식으로 통합하여, 단 한 번의 평가 과정에서 경계 상자와 클래스 확률을 동시에 계산한다. 또한, YOLO는 이미지를 SxS 크기의 그리드로 분할하고, 각 그리드 셀이 경계 상자와 객체 존재 확률을 예측하는 회귀 문제로 처리한다. 이러한 구조 덕분에 YOLO는 높은 처리 속도를 자랑하며, 실시간 객체 탐지가 필요한 응용 분야에서 특히 유용하다[5].

이를 토대로, 본 연구에서는 최신 YOLO 계열의 모델인 YOLOv8을 채택한다. YOLOv8은 이전 버전보다 경량화되었으며 다양한 크기의 객체를 효율적으로 탐지할 수 있도록 다중 스케일 예측 및 더욱 복잡한 네트워크 구조, 앵커 프리 탐지를 도입하였다[6]. 특히, YOLOv8은 성능 향상을 위한 최적화 기술이 적용되어, 작은 객체를 탐지하는데 있어서 기존 한계를 극복하였으며, 패션 이미지처럼 다양한 크기와 복잡한 구성을 가지는 데이터셋에 적합하다. 이를 통해, 본 연구는 더욱 빠르고 정확한 패션 아이템 탐지를 가능하게 한다.

#### 1.2 Object Detection In Fashion

객체 탐지 기술은 다양한 산업 분야에서 활용되고 있으며, 패션 산업에서도 매우 중요한 역할을 하고 있다. 패션 아이템은 의류, 액세서리, 신발 등 다양한 패션 요소를 포

합하고 있으며, 이를 자동으로 인식하고 분류하는 기술이 필수적이다[7]. 패션 산업에서 객체 탐지 기술이 중요한 이유는, 패션 아이템의 디지털화와 자동화가 온라인 상거래에서 소비자의 구매 경험이 직접적인 영향을 미치기 때문이다. 소비자들은 다양한 스타일, 색상, 패턴, 소재 등을 고려하여 의류를 선택하는데, 이러한 정보를 자동으로 탐지하고 분류할 수 있는 기술이 필요하다. 패션 아이템에서 정확한 객체 탐지가 가능해야, 트렌드 분석, 개인화 추천, 그리고 재고 관리와 같은 다양한 응용 분야에서 활용될 수 있다[8].

패션 아이템은 스타일, 색상, 패턴, 소재와 같은 세부 정보를 포함하고 있기 때문에, 일반적인 객체 탐지 기술과는 다른 정밀한 레이블링 작업이 필요하다. 예를 들어, 동일한 드레스라도 색상과 패턴, 실루엣에 따라 서로 다른 레이블링이 필요할 수 있다. 그러나, 이러한 레이블링 작업을 수작업으로 처리하는 것은 비효율적이므로, 자동화된 레이블링 기술이 필수적으로 요구된다[9].

## 2. Previous Fashion Item Labeling Works

기존 패션 이미지 레이블링 연구는 대규모 패션 이미지 데이터셋을 활용하여 의류 및 액세서리의 착용 위치나 스타일 분석에 중점을 두었다. 대표적으로 DeepFashion 데이터셋은 의류와 액세서리의 착용 위치와 스타일 분석에 중요한 자료로 사용되었으며, 신체 구조 정보를 포함한 객체 탐지 방법이 적용되었다[10]. 또한, 일부 연구에서는 2단계 이상의 분류 모델을 사용하거나[11], 패션 아이템의 신체 착용 위치를 고려한 관절 키포인트 설정을 통해 보다 세밀한 분류와 레이블링을 수행하기도 했다[12].

최근에는, 패션 아이템 레이블링에 계층적 접근 방식을 채택한 연구가 제안된 바 있다[13]. 이 연구에서는 이미지 분할, 기본 카테고리 분류, 세부 분류 등의 단계를 설정하여 패션 이미지를 세분화한 후, 상의, 하의, 신발 등의 주요 카테고리 및 색상 및 질감 같은 속성을 탐지하였다. 특히, 계층적 분류 시스템을 통해 상의 길이, 치마 길이, 소매 길이 등의 세부 속성을 분류하여 패션 아이템을 보다 정확하게 레이블링할 수 있었다.

그러나 이러한 연구들은 주로 고정된 카테고리 내에서의 분류에 중점을 두었으며, 다양한 패션 아이템의 세부적 특징을 충분히 다루지 못했다. 또한, 경계 상자 처리 과정에서 발생할 수 있는 중복된 레이블 문제에 대한 해결책도 부족하였다.

본 연구는 이러한 한계를 극복하기 위해 YOLOv8 모델과 대분류 객체 탐지 모델을 도입하여, 패션 아이템의 세

부적인 카테고리 분류와 경계 상자 분석 및 전처리를 수행하는 데 초점을 맞추었다. 이를 통해 기존 연구의 한계를 보완하고, 보다 정밀한 레이블링을 가능하게 하는 자동화된 시스템을 제안한다.

## III. The Proposed Scheme

본 논문에서 제안하는 패션 아이템 자동 레이블링 시스템은 패션 이미지와 해당 이미지에 대응하는 패션 아이템 종류를 포함한 텍스트 데이터를 입력값으로 사용한다. 입력 이미지는 다양한 해상도와 크기를 가질 수 있으며, 텍스트 데이터에는 각 이미지의 이름과 해당 패션 아이템의 세부 정보가 포함된다. 또한, 각 패션 아이템의 상위 카테고리를 기반으로 한 대분류 객체 탐지 모델을 활용하여 전반적인 객체 탐지 작업을 수행한다. 해당 객체 탐지 모델은 기존의 다른 모델로도 대체 가능하며, 본 연구에서는 상의(TOP)와 하의(BOTTOM)를 클래스로 분류하는 단순한 모델로 구성된다. 그림 1은 본 시스템의 개념도를 보여준다.

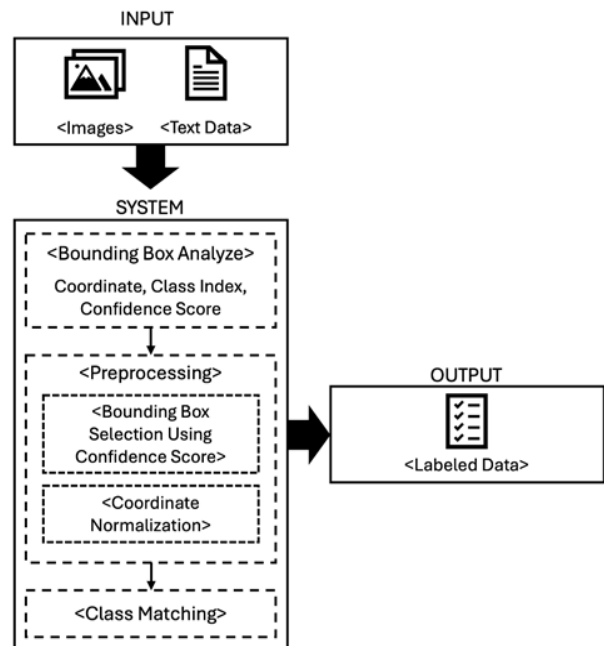


Fig. 1. Fashion Item Auto Labeling System

첫 번째 단계로, 입력된 패션 이미지는 대분류 객체 탐지 모델을 통해 상의와 하의의 예측을 수행하고, 탐지된 상의와 하의 각각의 경계 상자를 분석하는 과정을 거친다. 두 번째 단계로, 경계 상자 분석을 통해 추출한 정보와 신뢰도 점수(Confidence Score)를 바탕으로 가장 적합한 경계 상자를 선별한다. 신뢰도 점수는 경계 상자가 정확하

게 탐지되었는지를 나타내며, 이 과정에서 신뢰도가 낮은 경계 상자나 잘못 탐지된 경계 상자는 제거되고, 상의와 하의가 누락된 이미지는 자동으로 레이블링에서 제외된다. 세 번째 단계로, 남은 경계 상자는 좌표 정규화를 거쳐 각 경계 상자의 클래스 인덱스와 입력된 텍스트 데이터를 매칭한다. 상의(TOP)와 하의(BOTTOM)를 나타내는 클래스 인덱스를 바탕으로 탐지된 경계 상자와 해당 패션 아이템의 종류가 연결되며, 이를 통해 최종적으로 패션 아이템의 자동 레이블링이 수행된다. 네 번째 단계로, 최종적으로 레이블링이 완료된 패션 아이템의 데이터는 YOLO-TXT 형식으로 저장된다. 이 데이터는 각 이미지에 대한 경계 상자의 정규화된 좌표와 해당 패션 아이템의 클래스 정보를 포함하며, 시스템에서 패션 아이템의 위치와 종류를 명확하게 구분할 수 있도록 구성된다.

Table 1. System Input Value

Fashion Image	Fashion Item Type Text (TOP)	Fashion Item Type Text (BOTTOM)
image1	'T-Shirt'	'Jeans'
image2	'Sweater'	'Slacks'
image3	'Shirt'	'Sports'
image4	'Kneet'	'Cotton'

## 1. Data Collection

본 연구에서는 10,000장의 패션 이미지와 각 이미지에 대응하는 패션 아이템 종류가 포함된 텍스트 데이터를 사용해 실험을 진행하며, 데이터는 국내 최대 패션 편집숍인 M\*\*\*\*\*에서 수집한다. 데이터 수집 시, 실험의 정확도를 높이기 위해 상의와 하의 구분이 불분명한 이미지와 한 벌로 이루어진 옷(예: 원피스), 그리고 2인 이상이 포함된 이미지는 제외한다. 이를 통해 불필요한 잡음을 제거하고, 패션 아이템의 정확한 구분 및 레이블링이 가능하도록 한다.

## 2. Bounding Box Analysis

### 2.1 Bounding Box Coordinate

객체 탐지 모델에서 생성된 경계 상자는 AABB(Axis-Aligned Bounding Box), 즉 축에 맞춰 정렬된 직사각형 상자로 나타난다. AABB는 경계 상자의 각 변이 X, Y, Z축과 평행하게 정렬된 형태로, 객체를 감싸는 직사각형 형태로 정의된다. YOLO에서는 이러한 AABB를 [x\_center, y\_center, width, height]의 네 가지 값으로 표현하며, 이 좌표를 통해 객체의 위치와 크기를 나타낸다.

경계 상자의 중심 좌표인 x\_center와 y\_center는 각각 경계 상자의 x축과 y축 중심을 의미하며, 이는  $(x1 + x2)$

/ 2와  $(y1 + y2) / 2$ 로 계산한 값을 이미지의 너비와 높이로 나누어 정규화된 좌표로 변환한 값이다. 경계 상자의 너비(width)와 높이(height) 또한 경계 상자의 실제 크기를 이미지의 너비와 높이로 나누어 정규화한다. 이 정규화 과정을 통해 다양한 해상도의 이미지에서도 일관된 좌표 값을 얻을 수 있다.

YOLO 모델은 이미지에 대한 예측 결과를 'Results' 객체에 저장하며, 해당 객체에서 'boxes' 속성을 통해 탐지된 경계 상자의 정보를 얻을 수 있다. 'Boxes' 객체에는 여러 속성이 포함되어 있으며, 이 중 'xyxy' 속성을 통해 탐지된 경계 상자의 좌표를 추출한다. 이 좌표는  $nx4(x1, y1, x2, y2)$  크기의 2차원 배열 형태로 반환되며, 추출된 좌표는 본 시스템에서 좌표 정규화 과정을 거쳐 최종적으로 사용된다.

### 2.2 Bounding Box Confidence Score

'Boxes' 객체는 탐지된 경계 상자의 좌표 외에도, 클래스 인덱스('cls')와 신뢰도 점수('conf')를 반환한다. 이 속성들은 각각  $1 \times n$  크기의 1차원 배열 형태로 저장되며, n은 탐지된 경계 상자의 수를 의미한다. 신뢰도 점수는 0에서 1사이의 값으로, 해당 경계 상자가 정확하게 객체를 포함하고 있는지와 그 객체가 특정 클래스에 속하는지를 예측하는 정도를 나타낸다.

YOLO는 신뢰도 점수를 계산하기 위해, 먼저  $S \times S$  크기의 그리드(grid)로 분할한다. 이미지 내 객체의 중심이 특정 그리드 셀에 속하게 되면, 해당 그리드 셀이 그 객체를 탐지하는 책임을 지게 된다. 각 그리드 셀은 다양한 크기와 비율을 가진 여러 개의 경계 상자를 예측하며, 이 경계 상자들에 대해 개별적인 신뢰도 점수를 부여한다. 이러한 신뢰도 점수는 수식1과 같이 정의된다.

$$P_c = P_c(\text{Object}) * IOU_{pred}^{truth}$$

Eqs. 1. Confidence Score

이때,  $P_r(\text{Object})$ 는 경계 상자 안에 객체가 존재할 확률을 의미하며,  $IOU_{pred}^{truth}$ 는 예측된 경계 상자(predicted bounding box)와 실제 경계(ground truth bounding box) 사이의 일치도를 나타낸다. IOU(Intersection Over Union)는 두 경계 상자 간의 교집합 영역을 합집합 영역으로 나눈 값으로, 두 상자의 정확한 일치 정도를 정량적으로 평가하는 지표이다.

이러한 신뢰도 점수는 경계 상자가 올바르게 탐지되었는지, 그리고 해당 객체가 특정 클래스에 속하는지를 판단

하는 중요한 요소로 사용된다. 신뢰도 점수가 높을수록, 해당 경계 상자가 더 정확하게 탐지된 것이며, 이를 바탕으로 경계 상자 선별을 위한 전처리 과정에서 잘못된 탐지 결과를 걸러내는 데 활용한다.

### 3. Preprocessing for Bounding Box Selection

본 시스템에 입력된 대량의 패션 이미지는 대분류 객체 탐지 모델을 통해 상의와 하의의 경계 상자를 탐지한다. 그러나 탐지 과정에서 발생할 수 있는 오류를 방지하고, 최종 레이블링의 정확도를 높이기 위해 두 단계의 전처리 과정이 수행된다.

#### 3.1 Handling Images with Missing Detections

첫 번째 전처리 단계에서는 탐지 결과에서 상의(TOP)와 하의(BOTTOM) 중 하나라도 누락된 이미지를 처리한다. 이는 객체 탐지 모델이 정상적으로 작동하지 않거나, 이미지 품질 문제로 인해 탐지에 실패한 경우이다. 이러한 경우, 상의 또는 하의가 탐지되지 않은 이미지는 레이블링 실패로 간주되어 이후 처리 과정에서 제외된다. 이를 통해 불완전한 탐지 결과로 인한 오류를 사전에 방지하며, 본 시스템을 통한 레이블링 성공률에도 반영한다.

#### 3.2 Bounding Box Selection Using Confidence Score

본 시스템의 두 번째 전처리 단계에서는 신뢰도 점수를 활용하여 최종적으로 선택할 경계 상자를 선별한다. 신뢰도 점수는 YOLO 모델이 각 객체에 대해 얼마나 정확하게 탐지했는지를 나타내는 중요한 지표로, 각 클래스에 대한 탐지 결과의 신뢰성을 평가하는 데 사용된다. 본 연구에서는 신뢰도 점수를 기준으로 각 클래스에서 가장 신뢰도가 높은 경계 상자를 선택하여, 잘못된 탐지를 최소화하고 정확한 객체만을 선택하는 것을 목표로 한다.

먼저, 시스템은 각 이미지에서 탐지된 경계 상자에 대해 'Boxes' 객체에서 클래스 인덱스(cls)와 신뢰도 점수(conf) 정보를 추출한다. 각 이미지에서 탐지된 고유한 클래스 인덱스를 추출한다. 이때, 클래스 인덱스는 YOLO 모델이 상의(TOP)를 나타내는 클래스 인덱스 0과 하의(BOTTOM)을 나타내는 클래스 인덱스 1로 구분된다. 이후, 시스템은 각각의 클래스에 대해 동일한 클래스 인덱스를 가진 경계 상자들 중 신뢰도 점수가 가장 높은 경계 상자만을 선택한다. 이를 통해 각 클래스에 대해 가장 적합한 경계 상자 하나만 남기고 나머지는 제거하게 된다.

이를 통해 시스템은 잘못된 탐지를 최소화하고, 정확한 객체 탐지를 기반으로 한 신뢰도 높은 레이블링을 보장한다.

특히, 신뢰도 점수가 낮은 경계 상자는 제거되어 탐지 오류나 잘못된 레이블링을 방지한다. 이 과정은 이미지 내에서 중복된 경계 상자 문제를 해결하고, 각각의 클래스에 대해 하나의 대표 경계 상자만을 선택함으로써, 최종 레이블링 성능을 크게 향상시킨다.

또한, 이 단계에서의 경계 상자 선택은 경계 상자 좌표 정규화와 텍스트 데이터와의 매칭 작업을 준비하는 중요한 과정이다. 신뢰도 점수 기반 경계 상자 선별은 본 연구의 자동 레이블링 시스템의 정확도를 높이는 데 핵심적인 역할을 하며, 이를 통해 패션 아이템의 위치 정보와 종류가 정확하게 레이블링될 수 있도록 한다.

결과적으로, 본 시스템은 각 클래스별로 신뢰도 점수가 가장 높은 경계 상자만을 선택하여 최종 레이블링 성능을 최적화한다. 이러한 신뢰도 점수 기반 경계 상자 선별 기법은 특히 대규모 데이터셋에서 정확성과 효율성을 동시에 달성할 수 있는 방법론으로, 향후 다양한 응용 분야에서 유용하게 활용될 수 있을 것이다. 그림 2는 해당 과정의 결과를 시각화하여 보여준다.

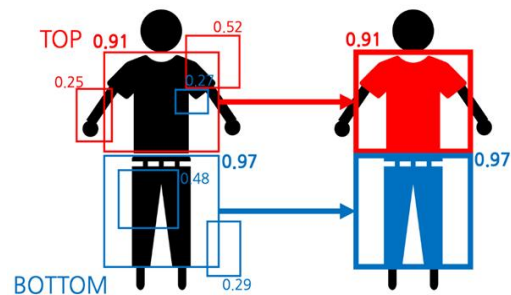


Fig. 2. Bounding Box Selection Using Confidence Score

### 4. Bounding Box Coordinate Normalization

YOLO 모델은 탐지된 객체의 경계 상자를  $[x\_center, y\_center, width, height]$  형식으로 정규화된 좌표로 표현한다. 이 정규화 과정은 객체의 절대적인 위치와 크기를 이미지 크기와 비교하여 상대적인 비율로 변환하는 작업이다. 이렇게 변환된 좌표를 사용함으로써, 이미지의 해상도나 크기에 상관없이 일관된 방식으로 객체의 위치를 표현할 수 있게 된다.

구체적으로, 경계 상자의 절대 좌표인  $[x1, y1, x2, y2]$ 는 '2.1 Bounding Box Coordinate'에서 설명된 바와 같이, 경계 상자의 왼쪽 상단과 오른쪽 하단의 좌표이다. 이 절대 좌표는 이미지의 너비(img\_width)와 높이(img\_height)를 사용해 정규화되며, 이를 통해  $x\_center$ 와  $y\_center$ 는 경계 상자의 중심 좌표로, width와 height

는 경계 상자의 너비와 높이를 나타낸다. 이러한 정규화 과정은 수식2를 통해 이루어진다.

$$x_{center} = \frac{(x1 + x2)}{2}, y_{center} = \frac{(y1 + y2)}{2}$$

$$width = \frac{x2 - x1}{img\_width}, height = \frac{y2 - y1}{img\_height}$$

Eqs. 2. Coordinate Normalization

정규화된 좌표는 객체의 위치와 크기를 이미지 크기와 비교하여 계산되므로, 서로 다른 해상도와 크기를 가진 이미지에서도 일관성을 유지할 수 있다. 이는 다양한 해상도의 이미지에서 모델이 객체를 탐지할 때 정확하고 일관된 결과를 얻을 수 있도록 돕는 중요한 과정이다. 그림3은 해당 경계 상자 좌표들의 위치와 필요성을 보여준다.

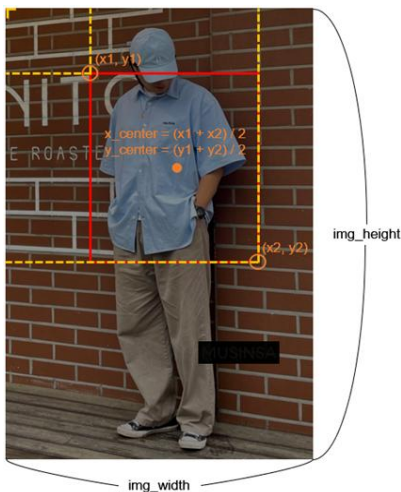


Fig. 3. Coordinate Normalization

## 5. Matching Fashion Item Types and Bounding Box Classes

본 논문에서 제안하는 패션 아이템 레이블링 시스템은 대분류 객체 탐지 모델을 활용하여 탐지된 경계 상자와 각 패션 아이템 종류 간의 매칭을 수행한다. 각 경계 상자의 클래스 인덱스를 이용해 대분류 객체 탐지 모델에서 정의된 클래스와 입력된 텍스트 데이터를 연결하며, 이를 통해 탐지된 객체가 어느 패션 아이템 종류에 속하는지를 정확하게 파악할 수 있다.

우선, 시스템은 '3.2 Bounding Box Using Confidence Score' 단계에서 추출한 각 경계 상자의 클래스 인덱스를 사용하여 경계 상자가 어떤 패션 아이템을 나타내는지 결정한다. 각 경계 상자의 클래스 인덱스는 대분류 객체 탐지 모델에서 사전 정의된 클래스와 동일하

며, 이를 통해 상의(TOP)와 하의(BOTTOM) 같은 주요 패션 아이템을 분류하는 기반이 된다. 이후, 시스템은 클래스 인덱스와 클래스 이름 간의 매핑 정보를 불러온다. 이 매핑 정보는 각 클래스가 어떤 패션 아이템을 나타내는지 명확히 하기 위한 자료로, 예를 들어, 클래스 인덱스 0은 상의(TOP)를, 클래스 인덱스 1은 하의(BOTTOM)을 나타낸다. 이를 바탕으로, 입력된 텍스트 데이터에 포함된 패션 아이템 종류와 탐지된 경계 상자를 연결한다.

이러한 매칭 과정은 다음과 같은 방식으로 진행된다. 탐지된 경계 상자의 클래스 인덱스가 0일 경우, 시스템은 이를 상의로 인식하고, 텍스트 데이터에서 상의에 해당하는 정보를 가져와 매칭한다. 즉, 상의를 나타내는 경계 상자의 클래스 인덱스인 TOP(0)은 텍스트 데이터에서 상의를 나타내는 패션 아이템 종류와 연결된다. 마찬가지로, 클래스 인덱스가 1일 경우 하의로 인식하며, 하의에 해당하는 정보를 매칭한다. 이를 통해 각 경계 상자가 어떤 패션 아이템에 해당하는지를 명확하게 식별할 수 있다.

또한, 이 과정에서 시스템은 중복된 클래스 인덱스가 발생하지 않도록 관리한다. 예를 들어, 서로 다른 이미지에서 동일한 패션 아이템 종류가 포함되어 있을 시, 해당 패션 아이템 종류는 같은 클래스 인덱스를 가져야 한다. 이를 위해 시스템은 각각의 텍스트 데이터에서 추출된 모든 패션 아이템 종류를 클래스 집합 리스트에 담아 패션 아이템 종류가 중복 입력되는 것을 방지한다.

최종적으로, 클래스 매칭이 완료된 각각의 상의와 하의 경계 상자는 입력된 텍스트 데이터의 패션 아이템 종류에 해당하는 클래스를 가지게 된다. 이후, 해당 경계 상자들의 정규화된 좌표와 함께 새로운 레이블 데이터로 생성된다. 예를 들어, 표 1에서 정의된 입력값 중 'image1'에 해당하는 패션 아이템인 'T-Shirt'와 'Jeans'는 각각 상의와 하의에 해당하는 경계 상자와 매칭되며, 2와 3이라는 새로운 클래스 인덱스를 부여받는다. 또한, 각각의 경계 상자의 정규화된 좌표를 포함하게 된다. 그림4는 이러한 과정을 거친 결과를 시각화하여 보여준다.

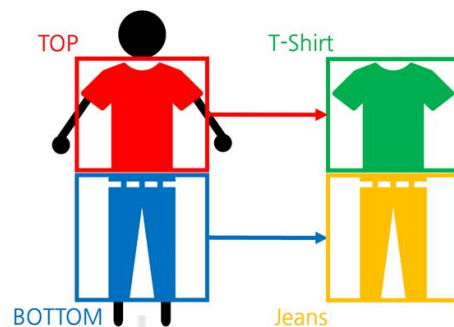


Fig. 4. Result of Class Matching

이러한 레이블 데이터는 YOLO-TXT 형식으로 출력되며, 각 이미지 내에서 탐지된 패션 아이템의 정확한 위치와 종류를 포함한다. 이러한 레이블링 과정은 이미지마다 반복 수행되며, 시스템이 탐지한 패션 아이템의 위치와 종류를 명확히 정의하여 자동 레이블링을 완성한다.

Table 2. System Output Value

Class Index	x_center	y_center	width	height
2	0.53..	0.25...	0.21...	0.15...
3	0.20...	0.13...	0.94...	0.59...
4	0.58...	0.48...	0.31...	0.83...
5	0.42...	0.85...	0.13..	0.53...
6	0.59...	0.13...	0.24...	0.94...
7	0.24...	0.14...	0.45...	0.32...
8	0.53...	0.24...	0.14...	0.85...
9	0.75...	0.68...	0.23...	0.19...

## IV. Conclusions

본 논문에서는 패션 아이템에 대한 자동 레이블링 시스템을 제안하고, '1. 데이터 수집' 단계에서 수집한 10,000장의 패션 이미지와 텍스트 데이터로 실험을 진행하여 시스템 성능을 평가하였다. 본 연구의 목적은 대량의 패션 이미지에서 객체 탐지 모델을 사용하여 상의와 하의를 자동으로 분류하고, 이를 바탕으로 정확하고 빠른 레이블링 시스템을 구축하는 데 있다. 이를 통해 패션 산업에서의 데이터 처리 효율성을 향상시키고, 수작업 레이블링의 한계를 극복하고자 하였다.

시스템 개발은 Python 기반으로 이루어졌으며, 접근성과 편의성을 고려하여 FastAPI를 사용한 백엔드 API 서버구축과 Svelte를 활용한 웹 기반 플랫폼을 통해 실험을 진행하였다. 또한, HP Z1 G9 4E883AV 워크스테이션(Intel Core i9-12900 프로세서)을 사용하여 로컬 환경에서 실행되었다.

실험 결과, 제안된 시스템은 8,192장의 이미지에서 성공적으로 레이블링을 수행하였으며, 이는 81.92%의 성공률을 의미한다. 또한, 이미지 1장당 평균 처리 시간은 약 0.15초로, 10,000장의 이미지를 처리하는 데 약 25분이 소요되었다. 이러한 결과는 기존의 수작업 레이블링 방식에 비해 빠른 처리 속도를 나타내며, 특히 대규모 데이터셋을 처리하는 데 있어 자동화된 시스템의 효율성과 실용성을 입증하였다.

본 연구는 기존의 수작업 기반 패션 아이템 레이블링 작업에서 자동화된 레이블링 시스템으로의 전환을 목표로 하였으며, 이를 성공적으로 구현하였다는 점에서 선행 연구와 차별화된다. 기존 연구들이 제한된 객체나 단순한 카

테고리만을 다루었던 것과 달리, 본 연구는 상의와 하의를 구분하는 대분류 객체 탐지 모델을 결합하여 패션 아이템의 세부적인 분류를 자동화하였다. 또한, YOLO 모델을 활용한 신뢰도 점수 기반 경계 상자 선별 기법은 레이블링 정확도를 높이는 데 큰 기여를 하였으며, 클래스 매칭과 더불어 빠른 처리 속도와 높은 성공률을 제공하는 중요한 차별점이다.

그럼에도 불구하고, 본 연구에는 몇 가지 한계점이 존재한다. 첫째, 실험에서 사용된 대분류 객체 탐지 모델이 상의와 하의의 두 가지 패션 아이템에만 초점을 맞추고 있으며, 액세서리나 기타 다양한 카테고리의 패션 아이템을 포함하지 못했다. 둘째, 원피스와 같은 한벌 옷이나 복수의 인물이 포함된 이미지에서는 탐지 정확도가 떨어지는 한계를 보여 실험을 진행하지 못하였다. 이는 대분류 객체 탐지 모델의 성능에 따라 결과가 달라질 수 있으며, 더 세부적인 클래스를 가진 모델을 사용할 경우 레이블링 가능한 패션 아이템의 종류가 증가한다는 것을 의미한다.

결론적으로, 본 연구는 객체 탐지 기술을 패션 아이템 레이블링에 적용하여 자동화와 효율성 향상에 기여하였으며, 다른 이미지 처리 분야에도 응용할 수 있는 중요한 기술적 발전을 제시하였다. 본 시스템은 패션 커머스, 이미지 검색, 개인 맞춤형 스타일링 서비스 등 다양한 패션 관련 인공지능 기술에서 활용될 수 있으며, 대규모 패션 이미지를 빠르고 정확하게 처리할 수 있는 효율적인 시스템을 제공한다는 점에서 실무적 의의가 크다.

향후 연구에서는 더 다양한 패션 아이템에 대한 클래스 확장과 함께, 복잡한 이미지 환경에서의 탐지 성능을 향상시키는 것이 필요하다. 이를 위해, 경계 상자 외의 이미지에 대한 고도화된 전처리 기술과 정밀한 객체 탐지 모델을 도입하고, 학습 데이터의 확장을 통해 탐지 성능을 개선해야 할 것이다. 또한, 패션 아이템의 인체 착용 위치를 고려한 키포인트 설정 및 세그멘테이션(Segmentation) 기술을 도입하여, 한벌로 이루어진 옷과 외투를 명확히 구분하고, 다수의 객체가 포함된 복잡한 이미지에서의 정확한 탐지를 위한 알고리즘 개선도 중요한 연구 과제가 될 것이다.

## ACKNOWLEDGEMENT

This research was supported by the MIST (Ministry of Science, ICT, Korea, under the National Program for Excellence in SW), supervised by the IITP(Institute of Information & communications Technology Planning & Evaluation) in 2024 (2019-0-01817)

## REFERENCES

- [1] A. A. Ogunyemi, I. J. Diyalou, I. O. Awoyelu, K. O. Bakare, and A. O. Oluwatope, "Digital Transformation of the Textile and Fashion Design Industry in the Global South: A Scoping Review", Towards new e-Infrastructure and e-Services for Developing Countries (AFRICOMM 2022), pp. 391-413, Tallinn, Estonia, June 2023. DOI: 10.1007/978-3-031-34896-9\_24.
- [2] A. P. Periyasamy, and S. Periyasami, "Rise of Digital Fashion and Metaverse: Influence on Sustainability", Digital Economy and Sustainable Development, Vol. 1, Article 16, pp. 1-15, September 2023. DOI: 10.1007/s44265-023-00016-z
- [3] I. Garcia-Aguilar, J. Garcia-Gonzalez, R. M. Luque-Baena, and E. Lopez-Rubio, "Automated Labeling of Training Data for Improved Object Detection in Traffic Videos by Fine-Tuned Deep Convolutional Neural Networks", Pattern Recognition Letters, Vol. 169, pp.24-32, January 2023. DOI: 10.1016/j.patrec.2023.01.015.
- [4] J. Huang, V. Rathod, C. Sun, M. Zhu, A. Korattikara, A. Fathi, I. Fischer, Z. Wojna, Y. Song, S. Guadarrama, and K. Murphy, "Speed/Accuracy Trade-Offs for Modern Convolutional Object Detectors", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 39, No. 6, pp. 1137-1149, June 2017. DOI: 10.1109/TPAMI.2016.2577031.
- [5] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 39, No. 6, pp. 1137-1149, June 2017. DOI: 10.1109/TPAMI.2016.2577031.
- [6] M. Hussain, "YOLOv5, YOLOv8 and YOLOv10: The Go-To Detectors for Real-time Vision", arXiv preprint arXiv:2407.02988, pp. 1-15, July 2024. DOI: 10.48550/arXiv.2407.02988.
- [7] L. Wang, Y. Zhang, and H. Li, "Fashion Object Detection for Tops & Bottoms", IEEE Transactions on Multimedia, Vol. 26, No. 4, pp. 1234-1245, April 2024. DOI: 10.1109/TMM.2024.3056789.
- [8] J. Doe, A. Smith, and B. Johnson, "Efficient Fine Tuning for Fashion Object Detection", Journal of Fashion Technology and Textile Engineering, Vol. 12, No. 3, pp. 145-158, March 2024. DOI: 10.1007/s44265-024-00045-y.
- [9] P. Gutierrez, P.-A. Sondag, P. Butkovic, M. Lacy, J. Berges, F. Bertrand, and A. Knudson, "Deep Learning for Automated Tagging of Fashion Images", Lecture Notes in Computer Science, Vol. 11131, pp. 3-11, January 2019. DOI: 10.1007/978-3-030-11015-4\_1.
- [10] Y. Ge, R. Zhang, X. Tang, and P. Luo, "DeepFashion2: A Versatile Benchmark for Detection, Pose Estimation, Segmentation, and Re-Identification of Clothing Images", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 43, No. 10, pp.3587-3601, October 2021. DOI: 10.1109/TPAMI.2020.2992036.
- [11] C. H. Lee, and C. W. Lin, "A Two-Phase Fashion Apparel Detection Method Based on YOLOv4", Applied Sciences, Vol. 11, No. 9, pp. 3782, April 2021. DOI: 10.3390/app11093782.
- [12] Jolim, Wjchoi, Bjchoi, "Automated Fashion Clothing Image Labeling System", Lecture Notes in Computer Science, Vol. 14532, pp. 3-8, February 2024. DOI: 10.1007/978-3-031-53830-8\_1.
- [13] B. E. Köktürk Güzel, "A Hierarchical Approach to Automated Fashion Product Tagging", Available at SSRN, [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=4705612](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4705612)

## Authors



Jun-oh Lim pursuing a B.S. degree in Software at Dongseo University, Korean, with an expected graduation date in February 2025. Junoh Lim joined Dongseo University, Korea, in 2018.

He is currently a student in the Department of Software at Dongseo University. He is interested in artificial intelligence, web application development, and software engineering.



Woo-jin Choi received the B.S. degrees in English-Korean Interpreting and Translation Studies from Kyunghee University, Korea, in 2016. She earned her M.S. and Ph. D. degrees in Department of Fashion and Textiles (Fashion Retailing) from Seoul National University, Korea, in 2018 and 2023, respectively.

Woojin Choi is postdoctoral fellow at the School of Fashion and Textiles at Hong Kong Polytechnic University in Hong Kong. Dr. Choi's research interests lie in fashion informatics, fashion retail technology, and fashion consumer behavior and psychology. Her recent research focuses on the diffusion of fashion trends using a data-mining approach.



Bong-jun Choi received the B.S. degrees in Computer Engineering from Dongseo University, Korea in 2013. He earned his M.S. degree in Computer Science from Yonsei University in 2015.

Bongjun Choi is Assistant Professor at Dongseo University in Busan, South Korea. He Studied Artificial Intelligence at Aalborg University and LG Electronics. His research interests lie in Bigdata, AI and AI convergence.