

비정형 패션 이미지 검색을 위한 MASK R-CNN 선형처리 기반 CNN 분류 학습모델 구현

(Implementation of CNN-based Classification Training Model for
Unstructured Fashion Image Retrieval using Preprocessing with
MASK R-CNN)

조 승 아¹⁾, 이 하 영²⁾, 장 혜 림²⁾, 김 규 리²⁾, 이 현 지²⁾, 손 봉 기³⁾, 이 재 호^{4)*}
(Seunga Cho, Hayoung Lee, Hyeelim Jang, Kyuri Kim, Hyeon-Ji Lee, Bong-Ki Son, Jaeho Lee)

요 약 본 논문에서는 패션 분야의 비정형 데이터 검색을 위한 패션 아이템별 세부 컨포넌트 이미지 분류 알고리즘을 제안한다. 코로나-19 환경으로 인하여 최근 AI 기반 쇼핑몰이 증가하는 추세이다. 하지만 기존의 키워드 검색과 사용자 서핑 행위 기반 개인 맞춤형 스타일 추천으로는 정확한 비정형 데이터 검색에는 한계가 있다. 본 연구는 다양한 온라인 쇼핑 사이트에서 크롤링한 이미지를 사용하여 Mask R-CNN을 활용한 전처리를 진행한 후, CNN을 통해 패션 아이템별 컴포넌트에 대한 분류를 진행하였다. 셔츠의 카라 및 패턴과 청바지의 핏, 워싱 및 컬러에 대한 분류를 진행하였으며, 다양한 전이학습 모델을 비교 분석한 후 가장 높은 정확도가 나온 Densenet121 모델을 사용하여 셔츠의 카라는 93.28%, 셔츠의 패턴은 98.10%의 정확도를 도달하였으며, 청바지의 핏은 Notched, Spread, Straight 3가지의 클래스의 경우 91.73%, Regular 핏을 추가한 4가지의 클래스의 경우 81.59%, 청바지의 색상은 93.91%, 청바지의 Washing은 91.20%, 청바지의 Demgae는 92.96%의 정확도를 도출하였다.

핵심주제어: 비정형 데이터, 이미지 분류, 신경망, Mask R-CNN

Abstract In this paper, we propose a detailed component image classification algorithm by fashion item for unstructured data retrieval in the fashion field. Due to the COVID-19 environment, AI-based online shopping malls are increasing recently. However, there is a limit to accurate unstructured data search with existing keyword search and personalized style recommendations based on user surfing behavior. In this study, pre-processing using Mask R-CNN was conducted using images crawled from online shopping sites and then classified components for each fashion item through CNN. We obtain the accuracy for collar of the shirt's as 93.28%, the pattern of the shirt as 98.10%, the 3 classes fit of the jeans as 91.73%, And, we further obtained one for the 4 classes fit of jeans as 81.59% and the color of the jeans as 93.91%. At the results for the decorated items, we also obtained the accuracy of the washing of the jeans as 91.20% and the damage of jeans accuracy as 92.96%.

Keywords: Unstructured Data, Image Classification, Neural Network, Mask R-CNN

* Corresponding Author: izeho@duksung.ac.kr
Manuscript received November 09, 2022 / revised
December 16, 2022 / accepted December 25, 2022

1) 덕성여자대학교 소프트웨어전공

2) 덕성여자대학교 IT미디어공학전공

3) 서원대학교 소프트웨어학부 컴퓨터공학전공

4) 덕성여자대학교 소프트웨어전공, 교신저자

1. 서 론

코로나-19의 환경 변화로 인해 오프라인 대비 AI를 이용한 온라인 쇼핑몰이 증가하는 추세이다. 현재 많이 사용되고 있는 AI 기반 온라인 쇼핑몰에서는, 패션 스타일 검색과 개인 맞춤 스타일 제공이 있다. 패션 스타일 검색은 키워드 기반의 검색이 대부분이다. 하지만 이는 일반 소비자가 사용하는 넓은 범위에 관한 용어와 디자이너들이 사용하는 용어에 대한 차이가 있기 때문에 실질적인 소비자 요구를 만족시키지 못하고 있다.

개인 맞춤형 스타일 제공 또한, 소비자들의 무의미한 서핑 행위를 기반으로 추천 기능을 제공하기 때문에 소비자들의 요구를 충족시키는 것에 한계점이 있다. 이러한 문제점으로 인하여, 본 연구는 비정형 데이터인 이미지 분류 모델에 관한 연구를 진행하였다.

비정형 데이터란 정량 데이터와 달리 그림이나 영상, 소리처럼 형태와 구조가 복잡해 정형화되지 않은 데이터를 말하며 유형이 불규칙하고 의미 파악이 모호 하기 때문에 기존의 컴퓨터 처리 방식을 적용하기 어렵다. 하지만 텍스트가 주된 정보의 유형이었던 과거와 달리, 다양한 멀티미디어 자료들의 증가에 따라 텍스트를 활용한 이미지 검색한계로 인하여 구글 이미지 검색시스템과 같은 이미지 검색 시스템이 주목을 받고 있다.

특히 패션 분야에서는 이미지 검색은 사용자들이 정확한 옷의 소재나 디자인을 설명할 수 없는 경우와 정확한 패션 전문 용어를 모를 경우에 대한 문제점을 해결할 수 있기 때문에 현재 많은 연구가 이루어지고 있다. 텍스트 검색에 비해 이미지 검색을 통한 쇼핑이 실제 구매로 이어질 확률이 높으며, 텍스트 검색으로 인한 한계점을 극복할 수 있다. 일반적인 이미지 검색은 사진을 입력을 하면 객체 위치 식별 기술을 통해 객체를 인식하고 속성 학습을 통하여 각 객체를 구분하고 분석을 하는 알고리즘을 가지고 있는 기술이다.

본 논문에서는 정확한 패션 트렌드 분석을 위해 셔츠의 카라 및 패턴, 청바지의 핏, 워싱 및 컬러와 같은 패션 아이템에 대한 세부 컴포넌트 단위의 분류를 진행하였으며, 패션 의류별 정확

하고 빠른 트렌드 분석을 위한 알고리즘을 제안한다. 분류의 정확도를 높이기 위해 데이터셋에 대한 여러가지 전처리 방식을 비교 분석 한 결과, 가장 높은 정확도를 보인 Mask R-CNN을 활용하여 데이터셋에 대한 전처리를 진행하였다.

2. 연구배경

Mask R-CNN은 이미지 분류를 수행하는 CNN과 위치를 추정하는 Regional Proposal 알고리즘을 연결한 모델인 R-CNN 모델 중 하나이다. 바운딩 박스(B-box)를 사용하여 객체를 찾는 객체 감지를 위한 다른 모델들과 달리, Mask R-CNN은 이미지내의 객체를 실제 테두리로 찾는 이미지 감지를 위한 모델이다.

Mask R-CNN은 기존 객체 감지에서 사용된 Faster R-CNN에 마스크 브랜치를 추가해 분류, 바운딩 박스 회귀, 객체 예측 마스크를 동시에 처리할 수 있으며, Mask R-CNN의 주된 기능은 영상 분할이다. 이는 영상이나 이미지에 있는 의미 있는 부분들을 구별해 주며, 이미지를 영역별로 나눠준 다음에 정확하게 객체의 경계선까지 추출해 주는 기능이다. 또한 클래스가 같더라도 다른 개체이면 이를 구분하여 인식해준다.

앞서 언급한 마스크 브랜치는 분류, 바운딩 박스 회귀, 분할 마스크의 3가지의 브랜치로 이루어져 있다. Mask R-CNN은 작은 크기의 FCN (Feature Pyramid Network)이며 Pixel-To-Pixel 방식으로써 픽셀별로 k개의 클래스 각각에 대해 물체의 유무를 판단한다. 기존의 다른 모델들과 달리 클래스 예측과 분할 마스크가 독립적으로 일어나고 위치정보를 보존할 수 있으며, 이전 모델에 있었던 느리고 부정확하다는 단점과 개체가 겹치는 문제를 해결한 모델이다.

전처리 과정을 거친 데이터셋을 본 연구에서 구현한 CNN모델의 입력데이터로 설정하여 분류를 진행하였다. 분류기 모델인 CNN은 높은 정확도를 위해 대량의 데이터가 필요하지만, 대량의 데이터를 얻는 것은 경제적인 부담이 크기 때문에 이를 해결하기 위해 전이학습을 사용하였다. 전이학습은 이미지넷과 같은 대형 데이터셋에서

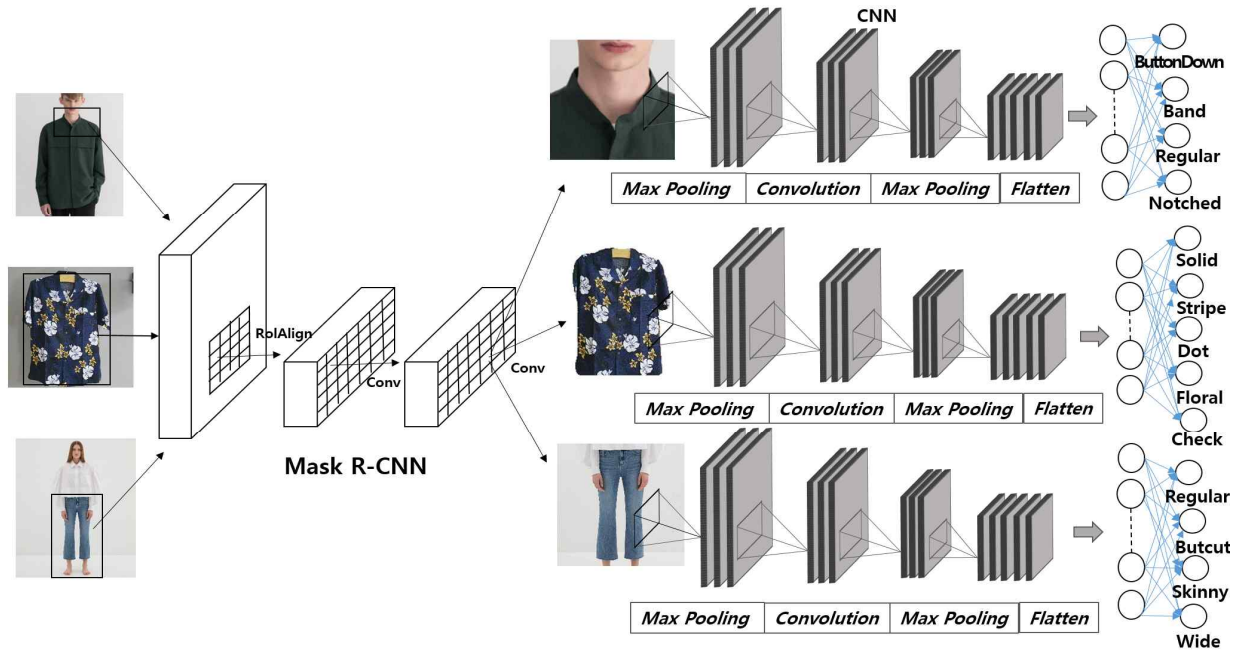


Fig. 1 Component Classification Model Structure by Fashion Item

훈련된 이미 공개되어있는 모델의 가중치를 도입하여 해결하고자 하는 모델에 적합하게 재설정하여 사용하는 기법이다. 전이학습을 사용하면 작은 데이터셋으로 원하는 결과를 도출해 낼 수 있으며 전이학습을 사용하지 않는 모델에 비해 정확도가 높으며 속도도 빠르다는 장점을 가지고 있어 많이 사용되는 기법 중 하나이다.

3. 모델 설계

3.1 제안 시스템 개요

본 논문에서 제안하는 패션 아이템 검출 모델을 위해 인터넷 쇼핑몰사이트에서 크롤링을 통해 데이터를 구축한 후, 딥러닝 기법인 Mask R-CNN과 CNN을 사용하여 해당 데이터에 대한 분류를 진행하였다. 기존의 패션 검출 시스템과는 달리, 텍스트 기반이 아닌 이미지 기반의 분류를 위한 모델을 설계하였으며, 단순히 패션 아이템이 아닌 패션 아이템 별 세부 컨포넌트에 대한 분류를 진행하였다. 모델의 정확도를 높이기 위해 여러 전처리 방법에 대한 연구를 진행하였

으며, 데이터셋의 규모에 대한 한계점을 해결하기 위해 전이학습에 대한 연구 또한 진행하였다. 그림 1은 패션 아이템별 컴포넌트 분류 모델에 관한 구성도이다.

3.2 모델 구성

인터넷 쇼핑몰에서 크롤링한 이미지에 Mask R-CNN을 이용하여 패션 아이템의 세부 컴포넌트인 셔츠의 카라 및 상의 부분과 청바지의 하의 부분을 크롭하여 전처리를 진행한 후, 이에 대한 결과값을 분류기 모델에 대한 입력값으로 사용하였다. 분류기 모델은 DenseNet121을 활용하여 CNN모델의 전이학습을 진행하였고, 다른 CNN 층을 동결한 후 분류기 부분만 학습하는 방법과 동결을 해제하여 전체 학습하는 미세조정 기법을 사용한 모델을 구현하였다. 해당 모델을 통과한 각각의 이미지에 대한 예측의 정확도를 측정하기 위해 예측 모델을 사용하여 검토를 진행하였다.

3.3 Mask R-CNN 기반 데이터 전처리

Table 1 Comparison of preprocessing models for datasets

Model	COCO format	Mask R-CNN	Mask R-CNN + Alpha	Mask R-CNN (Bbox)
Accuracy	72%	82.14%	82.63%	84.97%

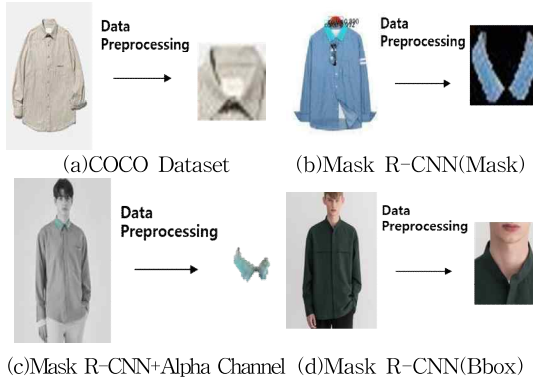


Fig. 2 Comparison of pretreatment models

데이터셋은 대형 온라인 쇼핑 사이트에서 HTML 태그 분석 후 파이썬 API를 이용한 크롤링을 통해 구축하였으며, Labelme Tool을 사용하여 직접 마우스로 해당 물체에 대해 라벨링 작업을 하였다. 그렇게 라벨링을 한 데이터셋을 사용하여 여러 가지 방식의 전처리를 진행하였다. 적절한 입력 이미지의 크기를 찾기 위해 여러가지 입력 이미지의 크기로 모델을 학습하였고, 그 결과 이미지의 형태가 왜곡되지 않은 정사각형 비율인 224×224의 크기의 이미지를 사용하였다. 표 1은 전처리 모델에 대한 비교분석 표이다.

그림 2(a)는 객체 감지와 분할을 위한 데이터셋인 COCO 데이터 형식을 이용한 전처리 방식이다. 모델은 코드를 수정하여 사용하였으며, 데이터셋을 COCO 형식으로 변환하는 동시에 바운딩 박스를 크롭하는 알고리즘을 구현하였다. 그 후, 바운딩 박스로 크롭한 이미지를 카라 유형별로 분류한 후, 2개의 컨볼루션 층과 풀링 층으로 이루어진 CNN 학습을 통해 분류를 진행하여 정확도 72%를 달성하였다. 이는 가장 간단한 전처리 방식이지만, 배경 노이즈 문제점이 존재한다.

그림 2(b)는 Mask R-CNN을 이용하여 예측된 바운딩 박스 중 크기가 가장 큰 바운딩 박스 영역

에 마스크를 적용하는 방법이다. 이러한 형태로 획득한 데이터셋을 224×224로 이미지 크기를 변경한 후 카라 유형별로 이미지를 분류하여 82.14%의 정확도를 도출하였다. 앞선 그림 2(a)의 문제점인 배경 노이즈는 감소하였지만 손과 얼굴 등 카라 외의 다른 객체를 검출하는 문제점과 검은 배경의 어두운 카라에 대한 분류의 어려움이 존재한다.

그림 2(c)는 그림 2(b)에서 사용한 RGB 3개의 채널을 사용한 Mask R-CNN의 모델에 알파 채널을 추가한 RGBA의 4개 채널을 사용하여 카라 이외의 영역을 모두 투명 이미지로 적용하여 5개의 컨볼루션 및 풀링 층으로 이루어진 CNN 모델을 통해 학습하여 정확도 82.63%를 달성하였다. 그림 2(c)의 문제점인 어두운색 카라 분류의 어려움은 해결되었지만, 손과 얼굴 등 카라 외의 다른 영역이 탐지되면서 카라 형태가 온전히 나오지 못하는 어려움은 해결되지 못하였다.

그림 2(d)는 Mask R-CNN을 이용하여 가장 큰 값을 가지는 ROIS만큼 원본 이미지를 바운딩 박스로 크롭하는 방법이다. 배경은 특별한 전처리 방법을 적용하지 않고 원본 그대로 사용하였으며 224×224의 크기로 변경하여 84.97%의 정확도를 도출하였다.

위의 연구 결과, 데이터셋에 대한 전처리는 Mask R-CNN을 활용하여 바운딩 박스로 물체를 인식하는 전처리 방식을 사용하여, 실험 데이터셋에 대해 0.96의 mAP와 0.94의 IoU의 결과를 도출하였다.

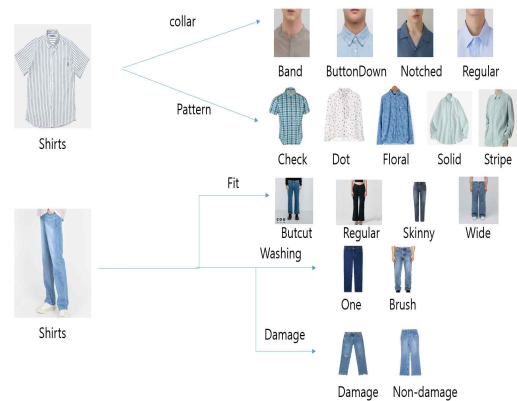


Fig. 3 Detailed Component Classification System by Fashion Item

그림 3는 패션 아이템별 세부 컴포넌트 클래스이다. 초기 셔츠 카라에 대하여 Notched, Spread, Straight 3개의 클래스로 연구를 진행하였으나, 예측 결과 Spread와 Straight를 구분하기 어렵다고 판단하여, 이 두 클래스를 통합하여 Regular로 분류한 후, Band와 Button Down을 추가하여 4개의 클래스로 진행하였다. 셔츠의 패턴 부분은 Check, Dot, Floral, Solid, Stripe 5개의 클래스, 청바지의 핏 부분은 Butcut, Skinny, Wide, Regular 4개의 클래스로 진행을 하였다.

3.4 이미지 분류 모델

본 연구에서는 Python 3.6, Keras 2.2.0, Tensorflow-gpu 1.10, H5py 2.10.0을 사용하여 연구를 진행하였으며, 분류기 모델인 CNN의 정확도를 향상 시키기 위해 전이학습에 관한 실험을 진행하였다. VGGNet, ResNet, DenseNet 3가지 모델에 대하여 비교 분석을 행한 결과, 가장 높은 정확도를 보인 DenseNet121을 사용하여 전이학습을 진행하였다. 표 2은 3개의 모델에 대한 비교 분석 표이다.

Table 2 Comparison of Transfer Learning Models for CNN Models

CNN Model	VGGNet	ResNet	DenseNet
Test Accuracy	56.89%	43.71%	78.57%
Validation Accuracy	56.89%	55.69%	76.19%

VGGNet은 Oxford의 Visual Geometry Group 연구팀이 개발한 CNN구조이며, 3×3필터를 사용한 단순 네트워크 구조를 가지고 있으며 층이 깊어질수록 필터가 2배로 규칙적으로 상승하여 비율이 체계적이라는 특징으로 인하여 기울기 소실 문제를 해결하였다. 하지만 파라미터의 수가 너무 많다는 문제점이 있다.

Resnet은 마이크로소프트에서 개발한 CNN모델이며 입력에서 바로 출력으로 연결되는 Skip Connection의 개념을 도입하였다. 입력에서 바로 출력으로 연결되기 때문에 파라미터 수의 증가로

인한 과적합을 해결하였고 기울기 소실에 대한 문제도 해결하였다.

Densenet은 Skip connection을 사용한다는 점에서 Resnet과 동일 하지만, 입력과 출력 피쳐맵을 합산했던 Resnet과 달리 연속으로 쌓아 점점 누적시킨다는 특징을 가지고 있다. 이로 인하여, 층이 깊어지면서 소실되었던 정보를 보존할 수 있으며 Resnet보다 더욱 향상된 성능을 가지고 있다.

본 연구에서는 Resnet50을 수정하여 8개의 컨볼루션 층과 2개의 컨볼루션 층마다 Skip connection을 갖는 모델을 사용하였다. 이 모델은 전체 206, 574, 466개의 파라미터가 도출되었으며, 학습 결과 43.71%의 정확도가 나왔다. 또한 Densenet121을 사용하여 학습하였을 경우 전체 7,039,554개의 파라미터가 도출되었으며, Resnet과 동일한 데이터셋을 사용하여 학습한 결과 78.57%의 정확도를 도출하였다. 이러한 연구 결과로 인하여 본 논문에서는 분류 모델로 Densenet121을 사용하였다. 정확도 향상을 위하여 전이학습 가중치를 동결시키고 분류기만 학습하는 방법과 동결을 해제한 후 전체 모델을 새로 학습하는 두 가지 방법을 사용하였다.

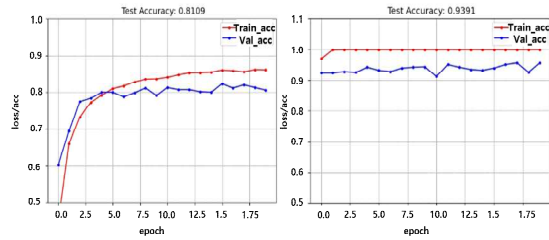
적은 양의 데이터셋을 사용할 경우 학습 데이터의 양을 늘리는 기법중 하나인 Image Data Generator를 사용하여 학습을 진행하였으며, 세부 파라미터인 rescale= 1/255를 사용하여 0과 1사이로 사이즈를 변경을 해주었고, fill_mode='nearest'를 사용하여 이미지의 여백이 생겼을 경우 여백을 주변 픽셀로 채우도록 설정을 하였다. 레이어의 앞단은 이미지넷으로 학습된 Densenet121 모델을 차용하였으며 Global Average Pooling을 사용하여 필터 수를 줄였고, 활성화 함수인 소프트맥스를 사용한 출력층을 추가하여 모델을 구성하였다. 또한 과적합을 방지하기 위해 Early Stopping을 사용하여 측정 시점에서 멈추도록 설계하였다. 손실함수는 교차 엔트로피를 사용하였으며 최적화 함수는 아담을 사용하였다.

4. 성능 분석

4.1 셔츠의 카라 및 패턴에 대한 성능분석

Table 3 Datasets for shirt collars

Dataset	Band	Button Down	Noched	Regular	Total
Train	534	618	736	737	2625
Validation	70	100	152	154	476
Test	70	100	152	154	476



(a) Freeze Convolutional Base (b) Train Entire Model

Fig. 4 Accuracy of shirt collar

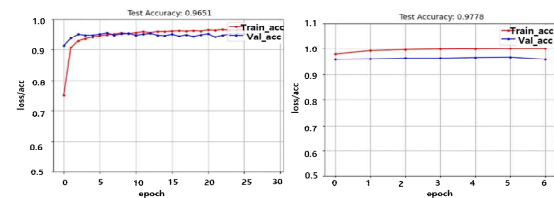
표 3과 그림 4는 셔츠의 카라에 대한 데이터셋 분류 기준 및 정확도이다. 패션 아이템별 컴포넌트 단위인 셔츠의 카라는 Band, Button Down, Noched, Regular 4개의 클래스로 분류를 하였다. 실험 데이터셋 2625장, Validation 데이터셋 476장 훈련 데이터셋 476장을 이용하여 학습을 진행하였다. Mask R-CNN을 통해 전처리를 진행한 데이터 셋을 사용하였으며, 이미지의 형태가 왜곡 되지 않도록 224×224 형태의 데이터셋을 사용하였다.

Densenet 121을 이용하여 전이학습을 진행하였으며, 미세조정(Fine-Tuning)기법 중 분류기만 학습하여 81.09%의 정확도를 도출하였으며, 동결을 해제한 후 전체 모델을 새로 학습하여 93.91%의 정확도를 도출하였다. 이러한 미세조정 기법은 기존 모델을 얼마나 재조정할 것인지 결정을 해야 하므로, 작은 학습률을 사용해야 하므로 분류기만 학습할 때는 0.0001의 학습률, 그리고 전체 모델을 새로 학습할 때는 10^{-5} 의 학습률로 학습을 진행하였다. 동결 해제 후 모델 전체를 학습할 경우에는 얼마 되지 않아 정확도가 100% 가까이 나와 과적합의 우려가 있지만 Validation 정확도도 함께 올라가는 것을 통해 과적합이 일어나지 않음을 알 수 있다.

연구 초반에는 Notched, Spread, Straight 3가지의 클래스 분류를 진행하여 정확도 84.97%를 도달하였고, 예측 결과 Notched는 96.87%, Straight는 88.54%의 정확도를 도달하였지만 Spread가 20.83%로 현저히 낮았다. 이 문제를 해결하기 위한 데이터셋 분석 과정에서 Spread와 Straight는 카라의 벌어진 각도를 제외한 모든 특징이 같아 데이터셋에서 셔츠의 각도에 따라 달라질 수도 있고, 육안으로도 구분하기 어렵다고 판단되어 둘을 분류하는 것이 의미가 없다고 판단을 하였다. 따라서 둘을 합쳐 Regular로 분류하였고 Band 카라를 추가하여 연구를 진행하였다. 또한 3가지 클래스로 분류를 진행을 하였을 때, 정확도 향상을 위해 검은색 배경을 활용하는 전처리 방식을 시도하였으나, 분류기만을 학습시켰을 때 정확도 71.28%, 전체를 학습시켰을 때 83.93%의 정확도로 인하여 적절한 전처리 방법이 아니라고 판단을 하였다.

Table 4 Datasets for shirt patterns

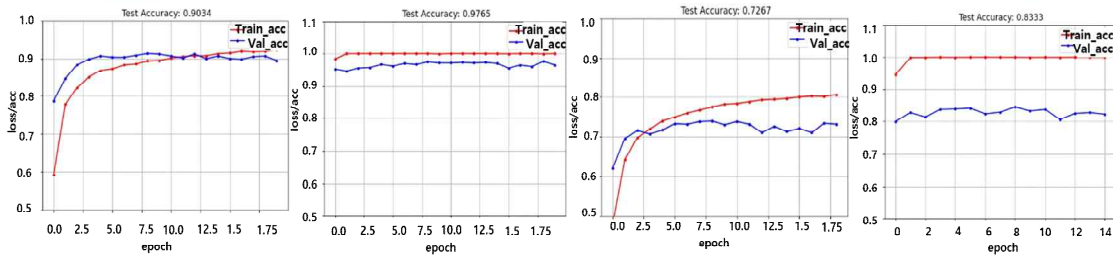
Dataset	Check	Dot	Floral	Solid	Stripe	Total
Train	700	606	700	700	600	3306
Validation	150	100	150	150	80	630
Test	150	100	150	150	80	630



(a) Freeze Convolutional Base (b) Train Entire Model

Fig. 5 Accuracy for shirt patterns

표 4와 그림 5는 셔츠 패턴에 관한 데이터셋 분류 및 정확도이다. 셔츠의 패턴 부분은 Check, Dot, Floral, Solid, Stripe 5개의 클래스로 분류를 진행하였으며 훈련 데이터셋 3306개, Validation 데이터 셋 630, 실험 데이터셋 630개를 이용하여 학습을 진행하였다. 분류기 부분만 학습을 할 때는 에포크 30으로 진행을 하였고, 전체 모델을



(a) Freeze Convolutional Base (b) Train Entire Model (c) Freeze Convolutional Base (d) Train 1 to 52 Layer
 Fig. 6 Accuracy of Jeans Fit((a),(b): Accuracy for 3 classes, (c),(d): Accuracy for 4 classes)

미세조정하여 학습할 때에는 에포크 6를 사용하여 학습을 진행하였다. 또한, 과적합(Overfitting)을 해결하기 위해 Early Stopping, 배치 정규화를 적용하고 여러 파라미터 조정을 통해 높은 정확도를 도출하였다.

4.2 청바지 핏 및 컬러에 대한 성능 분석

Table 5 3 Classes Dataset for Jeans Fit

Dataset	Butcut	Skinny	Wide	Total
Train	748	755	801	2304
Validation	126	126	131	383
Test	126	126	131	383

Table 6 4 Classes Dataset for Jeans Fit

Dataset	Butcut	Regular	Skinny	Wide	Total
Train	748	805	755	801	3109
Validation	126	133	126	131	516
Test	126	133	126	131	516

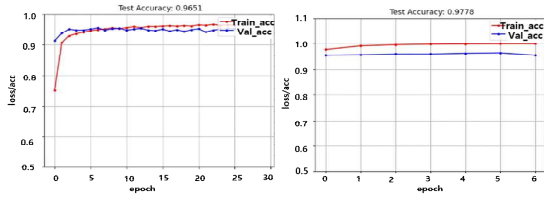
그림 6과 표 5, 6은 청바지 핏에 관한 데이터셋 분류 기준 및 정확도이다. 패션 아이템별 컴포넌트 단위인 청바지의 핏의 경우 연구 초기에는 Skinny, Wide, Regular, Butcut, Baggy 5개의 클래스로 진행을 하였지만 Baggy의 경우 Wide와 육안으로도 구별이 안되는 부분이 많아 Baggy에 대한 부분은 제외를 하였다. 이러한 환경을 바탕으로 Butcut, Wide, Skinny 3개의 클래스와 Regular를 추가한 4개의 클래스 두 가지로 분류를 진행을 하였다. Butcut, Skinny, Wide 3개의 클래스 분류 모델의 경우 훈련 데이터셋 2304개, Validation 데이터셋 383개, 실험 데이터

셋 383개를 이용하여 학습을 진행하였다. 분류기만 학습할 경우 0.0002의 학습률을 사용하였고, 전체 모델을 미세조정하여 학습하였을 때는 10^{-5} 의 학습률을 사용하였다. 그림 6(a)와 6(c)는 분류기 부분만 학습할 경우의 정확도이고, 그림 6(b)와 6(d)는 동결을 해제한 후 전체를 학습한 경우의 정확도이다.

Regular를 추가한 4개의 클래스 분류 모델의 경우, 훈련 데이터셋 3109개, Validation 데이터셋 516개 실험 데이터셋 516개를 사용하여 학습을 진행하였다. Regular 핏이 모호한 부분이 존재하여 Regular 부분의 예측 결과에 대한 향상 및 과적합 해결을 중점으로 연구를 진행하였다.

분류기만 학습을 했을 경우 0.0001 학습률을 사용하였으며 에포크 20 Step_per_epoch 775, Validation step 155를 사용하여 학습을 진행하였다. 또한 전체 모델을 미세조정하여 학습하였을 경우는 10^{-5} 의 학습률과 Early Stopping를 사용하여 5 Epoch, 775 Step_per_epoch 와 155 Validation을 이용하여 학습한 결과 86.4%를 도달하였지만, 과적합에 대한 문제가 있었다. 이를 해결하기 위해 Dropout 층을 추가하였으며 전체를 미세조정한 것이 아닌, 일부 레이어는 동결시켜놓은 채로 학습을 진행을 하였으며, 손실함수의 변경 및 여러가지 파라미터 조정 등을 통한 연구를 진행하였다. 그 결과, 분류기 부분만 학습을 한 경우 0.0001 학습률, 18 에포크로 학습하여 정확도 72.67%를 도출하였고, 1층부터 52층만 동결한 후 학습한 경우 10^{-5} 의 학습률과 Early Stopping을 사용한 14 에포크로 정확도 83.33%

를 도출하였으며, 과적합을 해결하였다.



(a) Freeze Convolutional Base (b) Train Entire Model

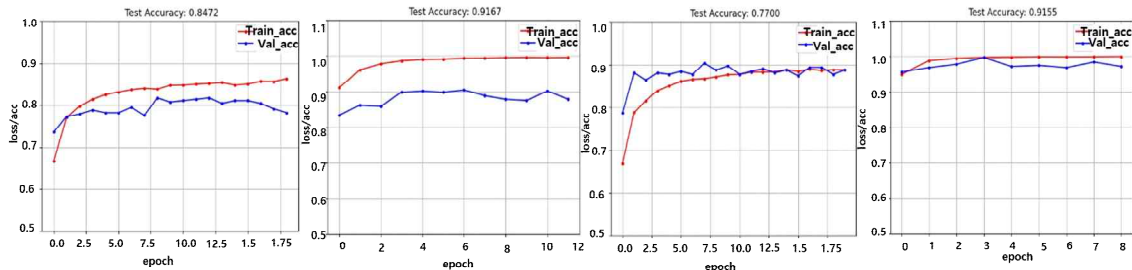
Fig. 7 Accuracy of Jeans Color

Table 7 Colors Datasets for Jeans

Dataset	Light	Middle	Deep	Total
Train	800	777	800	2377
Validation	133	126	133	392
Test	133	126	133	392

표 7과 그림 7은 청바지 컬러에 대한 데이터셋 분류 기준 및 정확도이다. 청바지 컬러는 Light, Middle, Deep 3개의 클래스로 나누어 학습을 진행하였고, 2377개의 훈련 데이터셋, 392개의 Validation 데이터셋 그리고 392개의 실험 데이터셋을 사용하였다. 분류기만을 학습한 경우 0.0001의 학습률을 사용하였고, 30 에포크, 395 Step_per_epoch 그리고 79 Validation step을 사용하여 학습을 하였고, 전체 모델을 미세조정하여 학습을 시킨 경우 10^{-5} 의 학습률, Early Stopping을 사용한 13 에포크, 395 Step_per_epoch, 그리고 79의 Validation step으로 학습하여 조금의 과적합이 일어났지만 높은 정확도를 도출하였다.

4.3 청바지 워싱에 대한 성능 분석



(a) Freeze Convolutional Base (b) Train Entire Model (c) Freeze Convolutional Base (d) Train Entire Model

Fig. 8 Accuracy of Jeans Washing((a), (b) Accuracy for Washing (c), (d) Accuracy for Damage)

Table 8 Damage Datasets for Jeans

Dataset	Damage	Non Damage	Total
Train	840	867	1707
Validation	105	108	213
Test	105	108	213

Table 9 Washing Datasets for Jeans

Dataset	One	Brush	Total
Train	879	858	1737
Validation	109	107	216
Test	109	107	216

그림 8과 표 8, 9는 청바지 Damage 및 Washing에 관한 데이터셋 분류 기준 및 정확도이다. 패션 아이템별 컴포넌트 단위인 청바지의 Washing은 Damage 부분과 Washing 부분으로 나누어 진행을 하였다. Damage 부분은 Damage, Non-damage 두개의 클래스로 나누었고, 훈련 데이터셋 1737개와 Validation 데이터셋 216개, 그리고 실험 데이터셋 216개를 사용하였다. Washing은 One과 Brush 2개의 클래스로 나누었고, 훈련 데이터셋 1707개, Validation 데이터셋 213개 그리고 실험 데이터셋 213개를 사용하였다. 연구 초기엔 Damage, One, Brush 3개의 클래스로 나누어 86.71%의 90%인 반면, Damage는 65%로 낮은 정확도가 나왔다. 이는, One과 Brush에 Damage가 있다고 판단이 되어 Damage 부분과 Washing 부분을 나누어 진행하였다. 또한 정확도 향상을 위해 다양한 전처리 방법을 시도하였다.

육안으로도 확인이 정확하게 되는 Damage와는 다르게 살짝의 스크래치가 있는 Damage는 예측이 어렵다고 판단이 되어 이미지 Edge Detection 중 하나인 Canny Edge Detection을 사용하였다. 그 결과, 반바지를 Damage라고 잘못 분류를 하는 문제점은 해결을 하였지만, 모델의 성능이 향상되지는 않았다. 명도 대비에 대한 전처리에 대한 연구도 진행을 하였다. 색상 이미지 명도 대비와 흑백 이미지 명도 대비 두 가지에 대한 연구를 진행하였는데, 색상 명도 대비보다는 흑백 명도 대비의 결과가 더 좋았기 때문에 Canny Edge Detection과 흑백 명도 대비를 같이 적용을 한 것과 Canny Edge Detection을 적용하지 않은 것 두 가지에 대하여 연구를 진행하였다. 그 결과, Canny Edge Detection을 입힌 것에 대한 예측결과는 Damage 부분에서는 86.5%에서 89.6%로 정확도 향상이 있었지만, Non-damage 부분에서는 94.8%에서 71.9%로 정확도가 낮아졌기 때문에 적절한 전처리 방식이 아니라고 판단을 하였다.

또한 Canny edge detection을 적용하지 않은 전처리 방식을 적용한 결과, Damage 부분에서는 91.7%, Non - damage 부분에서는 89.6%의 정확도로 앞서 Canny Edge Detection을 적용한 전처리 방식에 대한 문제점을 개선하였지만, 모델 성능 향상에 크게 기여하지 못했다고 판단을 하여 다른 전처리 방식을 사용하지 않고, Mask R-CNN을 이용한 하의 부분을 크롭한 이미지를 사용하였다. 과적합을 해결하기 위한 연구도 진행하였는데, 최적화 함수 및 손실함수와 활성화 함수의 변경과 여러 가지 파라미터의 변경 등의 시도를 하였지만, 과적합에 대한 문제를 완전하게 해결하지는 못하여 추후에는 모델의 안정화에 대하여 계속해서 연구를 진행할 것이다. 여러가지 시도 후 최적화 함수를 Rmsprop를 사용하였을 때 과적합이 개선 되었지만, 높은 정확도를 가진 모델은 아니었다. 따라서 Washing 부분의 경우 정확도가 가장 높은 모델을 살펴볼 때, 손실함수는 평균 제곱 오차, 최적화 함수는 아담과 활성화함수는 소프트맥스를 사용한 모델이었으며, 분류기만 학습하였을 경우는 학습률을 0.001로 학습을 하였을 때, 84.72%의 정확도를 도출하

였으며, 전체를 미세 조정된 모델에서는 10^{-5} 의 학습률을 사용하여 학습을 진행하여 정확도 91.55%를 도출하였다.

Damage의 경우 손실함수는 평균 제곱 오차, 최적화함수는 아담, 활성화함수는 소프트맥스를 사용하여 분류기만 학습한 경우는 0.001의 학습률을 사용하여 77%의 정확도를 도출하였으며, 전체 층에 대한 미세조정의 경우는 10^{-5} 의 학습률을 사용하여 학습을 시켜 92.96%의 정확도를 도출하였다.

5. 결론

본 논문에서는 패션 디자이너들의 빠른 트렌드 분석을 위하여, 패션 아이템별 세부 컴포넌트 단위로 분류하는 모델을 개발하였다. 최적화 모델을 설계하기 위해 전처리에 관한 연구를 진행하였으며, 그 과정에서 여러 전처리 방식을 비교 분석하여 가장 높은 정확도를 보인 Mask R-CNN을 선정하였다. 분류기 모델의 성능을 높이기 위해서 여러 CNN 모델의 비교 분석을 진행하였으며, 그 결과 가장 높은 결과를 보인 DenseNet121 전이학습을 진행하였다. 또한, 미세 조정 기법을 사용하여 다른 층은 동결하고 분류기만 학습하는 방법과 동결을 해제하고 전체를 학습시키는 방법을 이용하여 높은 성능을 가진 CNN 모델을 구현하였으며, 예측 모델을 이용하여 CNN 모델에 대하여 검증하였다. 추후에는 모델 성능 향상을 위해 데이터 전처리 방식에 대하여 연구를 진행할 것이며, Generative Adversarial Networks (GAN)을 활용하여 데이터셋을 생성한 후 그 결과값을 입력데이터로 사용하는 알고리즘에 대한 연구를 진행할 예정이다.

References

- K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in Proceedings of the IEEE International Conference on Computer Vision.

- Venice, Italy, pp. 2961–2969, 2017.
- R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Rich feature hierarchies for accurate object detection and semantic segmentation,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus: OH, USA, pp. 580–587, 2014
- S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks,” Advances in Neural Information Processing Systems 28, Montréal, Canada, 2015.
- J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, and L. Fei-Fei, “Imagenet: A large-scale hierarchical image database,” in Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, Miami: FL, USA, pp. 248–255. 2009.
- K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas: Nevada, USA, pp. 770–778, 2016.
- G. Huang, Z. Liu, L. V. D. Maaten, and K. Q. Weinberger, “Densely connected convolutional networks.” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu: Hawaii, USA, pp. 4700–4708, 2017.
- B. C. Russell, A. Torralba, K. P. Murphy, and W. T. Freeman, “LabelMe: a database and web-based tool for image annotation.” International Journal of Computer Vision , vol. 77, no. 1, pp. 157–173, 2008.
- T. Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, “Microsoft COCO: Common Objects in Context,” in Proceedings of the European Conference on Computer Vision, Zürich, Switzerland, pp. 740–755, 2014
- F. C. Akyon, labelme2coco project [Internet], 2022. Available: <https://github.com/fcakyon/labelme2coco/>.
- K. Simonyan, and A. Zisserman, “Very Deep Convolutional Networks For Large-Scale Image Recognition,” arXiv preprint arXiv:1409.1556, 2014.
- L. Min, Q. Chen, S. Yan, “Shuicheng. Network in network,” arXiv preprint arXiv:1312.4400, 2013.
- W. Rong, Z. Li, W. Zhang, and L. Sun, “An improved CANNY edge detection algorithm.” in Proceedings of the IEEE International Conference on Mechatronics and Automation. Tianjin, China, pp. 577–582, 2014.
- I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative Adversarial Nets,” Communications of the ACM, vol. 63, no. 11, pp. 139–144, 2020.
- M. H. Lee, K. W. Nam, and C. W. Lee, “Crack Detection on the Road in Aerial Image using Mask R-CNN.” Journal of the Korea Industrial Information Systems Research, vol. 24, no. 3, pp. 23–29, 2019.
- J. H. Lee, B. M. Kim, and Y. S. Shin, “Effects of Preprocessing and Feature Extraction on CNN-based Fire Detection Performance.” Journal of the Korea Industrial Information Systems Research, vol. 23, no. 4, pp. 41–53. 2018.
- J. W. Jo, M. H. Lee, H. R. Lee, Y. S. Chung, J. H. Baek, K. H. Kim, and C. W. Lee, “LeafNet: Plants Segmentation using CNN.” Journal of the Korea Industrial Information Systems Research, vol. 24, no. 4, pp. 1–8, 2019.



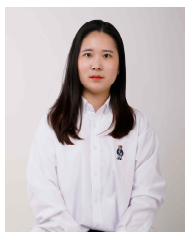
조 승 아 (Seunga Cho)

- (현재) 덕성여자대학교 소프트웨어전공 학사과정 재학
- 관심분야: Image Processing, CNN, Neural Network



손 봉 기 (Bong-Ki Son)

- 충북대학교 전자계산학과 석사
- 충북대학교 전자계산학과 박사
- (현재) 서원대학교 소프트웨어학부 컴퓨터공학전공 교수
- 관심분야: Deep Learning, Image Processing, Big Data



이 하 영 (Hayoung Lee)

- (현재) 덕성여자대학교 IT미디어공학과 학사과정 재학
- 관심분야: Deep Learning, Computer Vision, Signal Processing, Image Processing.



이 재 호 (Jaeho Lee)

- 고려대학교 전자컴퓨터공학과 공학석사
- 고려대학교 전기전자컴퓨터공학과 공학박사
- 전 LG전자 CTO 차세대표준



장 혜 립 (Hyelim Jang)

- (현재) 덕성여자대학교 IT미디어공학과 학사과정 재학
- 관심분야: Machine Learning, GAN, Neural Network

- 연구소 선임연구원
- 전 서원대학교 정보통신공학과 조교수
- (현재) 덕성여자대학교 소프트웨어전공 조교수
- 관심분야: WPAN, MAC, Bluetooth, Wi-Fi, Localization, NLP, Machine Learning.



김 규 리 (Kyuri Kim)

- (현재) 덕성여자대학교 IT미디어공학과 학사과정 재학
- 관심분야: Deep Learning, Computer Vision, Signal Processing, Image Processing.



이 현 지 (Hyeon-Ji Lee)

- (현재) 덕성여자대학교 IT미디어공학과 학사과정 재학
- 관심분야: Localization, Machine Learning, Deep Learning