

패션 요소 검출을 위한 Mask R-CNN 딥러닝

김 희 숙¹

¹광주과학기술원 대우교수

Mask R-CNN deep learning for fashion element detection

Hye-Suk Kim¹

¹Lecture Professor, Gwangju Institute of Science and Technology, Gwangju Cheomdangwagi-ro 123, Korea

[요 약]

본 논문에서는 개인별 맞춤형 인공지능 패션 코디네이션 서비스를 제공할 수 있는 프레임워크를 마련하기 위해서 iMaterialist Fashion Attribute Dataset에서 제공되는 패션 이미지 데이터 세트를 대상으로 Mask R-CNN 딥러닝 알고리즘을 이용하여 패션 요소를 검출하였다. 총 8 에폭으로 패션 요소 검출을 위한 딥러닝을 수행한 결과, 학습 데이터의 손실은 Lcls 0.53, Lbox 0.38, Lmask 0.35로 나타났다. 그리고 검증 데이터의 손실은 Lcls 0.54, Lbox 0.33, Lmask 0.36으로 나타났다. 개인이 소유하고 있는 패션 이미지 데이터 세트를 기반으로 색상, 계절, 재질, 트렌드, 브랜드명 등의 다양한 조건들을 추가한 후 본 논문에서 구현한 딥러닝 방식을 이용하여 파인 튜닝할 경우 효과적인 개인별 인공지능 패션 코디네이션 서비스가 될 것이라 기대한다.

[Abstract]

In this paper, in order to prepare a framework that can provide personalized artificial intelligence fashion coordination services, fashion elements were detected using the Mask R-CNN deep learning algorithm targeting the fashion image data set provided in the iMaterialist Fashion Attribute Dataset. As a result of performing deep learning to detect fashion elements with a total of 8 epochs, the loss of training data was found to be Lcls 0.53, Lbox 0.38, and Lmask 0.35. And the loss of Validation data was Lcls 0.54, Lbox 0.33, Lmask 0.36. Effective personal artificial intelligence fashion when fine-tuning using the deep learning method implemented in this paper after adding various conditions such as color, season, material, trend, and brand name based on the fashion image data set owned by an individual It is expected that it will be a coordination service.

색인어 : R-CNN, Fast R-CNN, Faster R-CNN, Mask R-CNN, 인공지능, 패션

Key word : R-CNN, Fast R-CNN, Faster R-CNN, Mask R-CNN, Artificial Intelligence, Fashion

<http://dx.doi.org/10.9728/dcs.2021.22.4.689>



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.

Received 21 February 2021; **Revised** 12 March 2021

Accepted 29 March 2021

***Corresponding Author; Hye-Suk Kim**

Tel: +82-62-530-0147

E-mail: idishskim@naver.com

I. 서 론

패션 디자이너는 자신이 디자인한 패션 요소가 어떤 용도로 사용되었을 때 가장 효과적인 연출이 될 것인가를 생각하고 제품을 디자인한다. 반면에 패션 디자인과 관계없는 일반인들은 대부분 디자이너의 의도와 상관없이 자신의 주관대로 또는 즉흥적으로 패션 요소를 매칭하여 코디네이션하는 경우가 일반적이다. 하지만 최근 패션 요소의 시각적 분석[1] 분야가 관심을 받기 시작하였고 스마트 옷장[2]이 등장하였으며 더불어 패션 큐레이션 서비스(fashion curation service)가 패션 상품의 소비 과정에서 나타나는 의복 행동과 관련하여 패션 비즈니스에 직접적인 영향력을 제공하고 있다. 패션 큐레이션은 큐레이션 영역에서 패션 산업 분야에 특화된 큐레이션 행위로 패션 상품에 관하여 큐레이션 주체의 취향이 반영된 가치 있는 정보를 재구성하거나 생산하는 것을 의미한다[3]. 현재 여러 쇼핑몰에서는 이와 같은 패션 큐레이션 서비스를 통해서 소비자에게 상품 구매 동기와 의사 결정 과정에 영향력을 제공하고 있다.

대규모의 이미지 데이터 세트를 기반으로 시각적 분석을 통한 인공지능 기술이 빠른 속도로 발전하고 있다. 최근 패션 분야에서는 iMaterialist Fashion Attribute Dataset[4]이라는 새로운 이미지 데이터 세트가 발표되었다. 데이터 세트는 총 228개의 세분화 된 속성으로 8개 그룹을 포함하는 레이블로 완성되었으며 100만 개 이상의 패션 이미지가 제공되고 있다. 각 이미지에는 전문가가 패션 속성별로 레이블링을 하였고 이를 기반으로 편리하게 패션 관련 데이터를 분석하고 예측할 수 있는 다양한 딥러닝을 수행할 수 있게 되었다. 본 연구에서는 개인별 맞춤형 인공지능 패션 코디네이션 서비스를 제공할 수 있는 프레임워크(framework)를 마련하고자 iMaterialist Fashion Attribute Dataset에서 제공되는 패션 이미지 데이터 세트를 이용하여 기계가 스스로 패션 요소를 검출하고 분할 할 수 있는 딥러닝을 실행하였다.

본 논문의 구성은 다음과 같다. 2장에서는 다양한 이미지 인식 및 객체 분할에 활용되고 있는 딥러닝 알고리즘을 살펴보고 3장에서는 본 연구에서 사용한 패션 요소 검출을 위한 Mask R-CNN[5] 딥러닝 알고리즘을 이용한 패션 요소 검출 방법에 대해 설명한다. 이어서 4장에서는 구현된 딥러닝 방식의 손실(loss)을 평가하고 5장에서는 결론 및 향후 연구 방향에 대해 서술한다.

II. 관련 연구

컴퓨터 비전 분야에서 머신러닝 및 딥러닝 알고리즘을 이용한 이미지 인식은 일반적으로 이미지 내에 포함된 핵심 객체가 무엇인지를 알아내는 것이다. 이를 위하여 1998년 LeNet-5를 발표하면서 소개된 CNN(convolutional neural network) 알고리즘을 인공지능망에 적용함으로써 이미지를 더욱 효과적으로 인식할 수 있게 되었다[6]. CNN 등장 이후, 이를 기반으로 확장된 CNN 관련 딥러닝 알고리즘 R-CNN(Regions with CNN)[7], Fast R-CNN[8], Faster R-CNN[9], Mask R-CNN[5] 등이 발표되

어 이미지 인식 분야는 놀라운 성과를 보여주고 있다. 본 장에서는 R-CNN 초기 모델 및 Fast R-CNN, Faster R-CNN 등과 같이 향상된 R-CNN 모델에 대해 설명한다.

2-1 R-CNN 초기 모델

R-CNN은 객체 검출 과정 초기에 사용되었던 모델로 CNN에 region proposal을 추가하여 객체가 존재할만한 영역을 제안하고 제안된 영역에서 객체를 검출할 수 있도록 2014년도에 소개되었다[7]. 배경과 객체들로 이루어진 복잡한 상황의 이미지에서 R-CNN 알고리즘을 이용하면 아래의 그림 1과 같이 객체를 검출하고 분류할 수 있게 된다. 그림 1은 하나의 이미지에서 주요 객체들을 bounding box로 표현하고 더불어 bounding box의 객체가 무엇을 의미하는 것인지를 출력하는 모습을 R-CNN으로 구현한 것을 나타낸다.

R-CNN의 신경망 모델 구조는 이미지 입력, 객체가 존재할만한 후보 영역 추출, CNN 특징 계산, bounding box 기반의 영역 분류를 통해 객체를 검출할 수 있도록 구현된다. 즉, 영역별로 탐색 부분을 나누어 배경은 제외하고 객체가 존재할 영역을 대상으로 CNN을 실행시키는 방식을 의미하며 그림 2는 R-CNN 신경망 모델의 개략적인 흐름을 보여준다.

본 연구에서 활용하고자 하는 Mask R-CNN의 초기 모델인 R-CNN 알고리즘은 다음과 같이 3단계로 처리한 후 마지막으로 bounding box 선형 회귀(linear regression)를 통해 앞에서 보여준 그림 1과 같이 출력되었다.

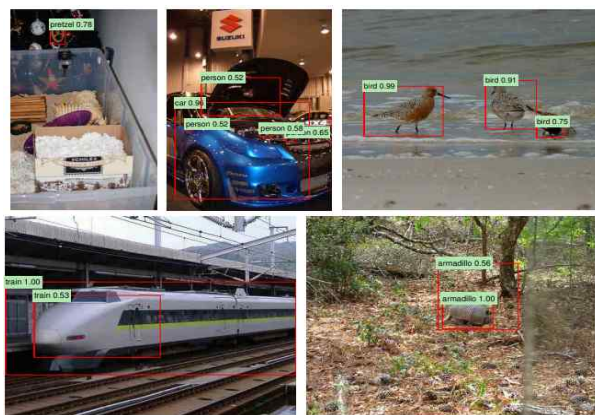


그림 1. R-CNN을 이용하여 검출된 객체 예시
Fig. 1. Example of objects detected using R-CNN

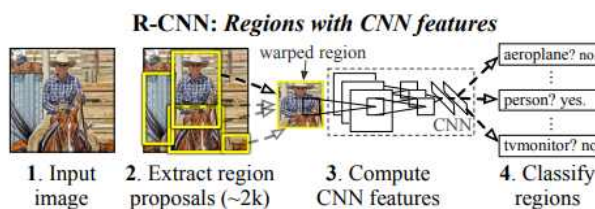


그림 2. R-CNN의 모델 구조
Fig. 2. Model structure of R-CNN

1단계는 이미지 중에서 객체로 인식될 수 있는 region proposal 즉, 영역 제안을 선택적 탐색(selective search)[10]이라는 규칙 기반 알고리즘(rule-based algorithm)을 적용하여 약 2,000개의 region proposal을 생성하는 작업을 수행한다. 선택적 탐색 알고리즘은 질감, 색상, 강도 등을 기준으로 주변 픽셀과의 유사도를 측정하여 영역을 나누고 이를 기준으로 객체가 존재할 가능성이 높은 위치를 추론하는 방식이다. 그림 3은 픽셀 유사도 정보를 기반으로 선택적 탐색을 수행하여 여러 가지 다양한 윈도우(window)를 생성하고 있음을 나타낸다[10]. 생성된 다양한 윈도우는 R-CNN의 region proposal에 해당하며 서로 독립적이지 않고 중복된 영역이 존재하고 있다.

2단계는 region proposal 영역을 대상으로 와핑(wrapping) 처리를 수행한다. 이는 객체가 될 수 있는 후보 영역 즉, 윈도우로 생성된 region proposal은 각각 크기가 다르기 때문에 고정된 픽셀 사이즈 227 x 227로 와핑을 수행하여 CNN 모델의 입력으로 사용하기 위함이다. R-CNN에서는 AlexNet을 개량한 CNN 모델을 파인 튜닝(fine tuning)하여 4,096차원의 특징 벡터를 추출하였다. 파인 튜닝에서는 그림 4에서 제시한 것처럼 ground truth에 해당하는 이미지들을 가져와 학습하도록 하였다[11].

3단계는 CNN을 통과하여 나온 4,096차원의 특징 벡터를 활용하여 각각의 레이블을 기준으로 학습시켜 놓은 서포트 벡터 머신(SVM; support vector machine)[12]을 통과시켜 분류하는 작업을 수행한다.



그림 3. 선택적 탐색 예시
Fig. 3. Examples of selective search

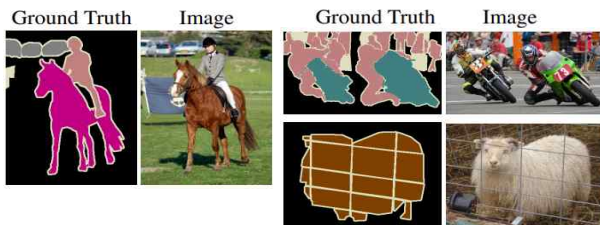


그림 4. ground truth의 예시
Fig. 4. Examples of ground truth

서포트 벡터 머신은 분류의 결정 경계(decision boundary)를 위해 기준이 되는 선을 정의하는 머신러닝 모델이다. R-CNN[7]에서는 서포트 벡터 머신을 통과하여 어떤 객체에 해당하는지에 대한 확률값을 의미하는 score를 가지게 된다. 이 경우, 객체 영역에 여러 개의 윈도우가 겹쳐있다면 가장 score가 높은 것으로 분류하게 된다.

이처럼 3단계의 처리로 객체를 검출하고 마지막으로 검출된 객체를 그림 1에서 보여준 것과 같이 bounding box로 출력한다. 이때 검출된 객체의 bounding box를 원본 이미지에서 정확한 위치에 정확한 크기로 표현하기 위해서 bounding box 선형 회귀를 이용한다. bounding box 선형 회귀를 위해서 검출된 bounding box와 ground truth에 해당하는 중심점 위치 x, y 그리고 너비(width), 높이(height)를 기반으로 검출된 객체의 bounding box가 최대한 ground truth와 일치하도록 이동시키는 학습을 수행한다. 학습을 수행하면서 손실이 최소가 되는 가중치(weight)를 곱해서 검출된 bounding box를 조정하는 방식이다.

2-2 향상된 R-CNN 모델

R-CNN[7]은 객체 검출을 위해 잘 작동하지만 2,000개의 region proposal로 제안된 영역마다 CNN을 수행하므로 속도가 저하되고, multi-stage pipelines의 모델로 한 번에 학습시키지 못하는 문제점이 있다. 이러한 문제점을 해결하기 위해서 R-CNN[7]을 발전시킨 Fast R-CNN[8], Faster R-CNN[9] 등이 소개되었다.

1) Fast R-CNN

Fast R-CNN은 초기의 R-CNN의 문제점을 해결하기 위해서 제안된 방식으로 그림 5와 같이 프로세스가 구성된다.

1단계는 R-CNN과 달리 맨 처음 전체 이미지를 딱 한 번 CNN에 통과시켜 feature map을 추출하는 작업을 수행한다. 그리고 R-CNN에서와 마찬가지로 선택적 탐색을 통해 region proposal에 해당하는 RoI(Region of Interest)를 찾는다.

2단계는 선택적 탐색으로 찾았던 RoI를 와핑 없이 feature map 크기에 맞춰서 투사시킨다.

3단계는 투사시킨 RoI에 대해 RoI Pooling을 진행하여 고정된 크기의 feature vector를 얻는다.

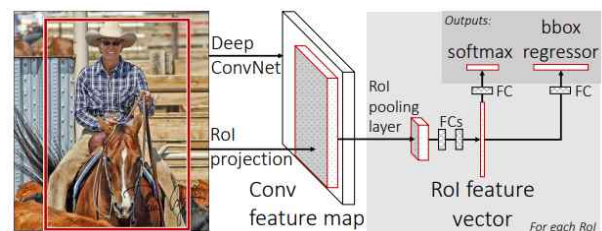


그림 5. Fast R-CNN의 모델 구조
Fig. 5. Model structure of Fast R-CNN

즉, 앞 단계에서 이미지를 CNN에 통과시켜 feature map을 추출한 후 공간 피라미드 풀링(SPP; Spatial Pyramid Pooling)[13]으로 고정된 크기 vector를 만들고 이를 FC layer의 input으로 입력한다. 따라서 CNN을 통과한 feature map에서 2,000개의 region proposal을 만들고 region proposal마다 공간 피라미드 풀링 신경망에 넣어 고정된 크기의 feature vector를 얻어낸다. 이 작업을 통해 R-CNN에서 2,000개의 region proposal마다 수행했던 2,000번의 CNN 연산이 1번으로 줄어들어 속도를 개선할 수 있게 된다.

4단계는 feature vector가 FC layer를 통과한 뒤 2개의 프로세스로 나뉘도록 설계하여 첫 번째 프로세스는 softmax를 통과하여 RoI에 대해 object classification을 수행하도록 하고, 두 번째 프로세스는 bounding box 회귀를 통해 selective search로 찾은 bounding box의 위치를 조정하도록 한다. 이처럼 bounding box regression layer를 softmax layer에 평행하게 두어 bounding box 좌표들이 출력될 수 있도록 처리된다.

R-CNN 초기 모델은 image feature를 생성하는 CNN, class를 예측하는 SVM 그리고 regression model이 bounding box를 찾아내는 것을 각각 나누었지만, Fast R-CNN[8]은 3개의 모델을 하나의 네트워크로 구현되도록 하였다.

2) Faster R-CNN

Fast R-CNN의 성과에도 불구하고 region proposal 부분에 병목 지점이 존재하고 있다. 객체의 위치를 찾는 과정 중 첫 번째 과정은 바로 객체가 있을 만한 후보 영역의 다양한 bounding box를 생성하는 것이다. Fast R-CNN에서는 region proposal을 찾기 위해서 selective search라는 과정을 통과하는데 이 과정에서 병목 현상이 발생하였다. 이러한 병목 현상을 해결하기 위해서 Faster R-CNN[9]이 제안되었다.

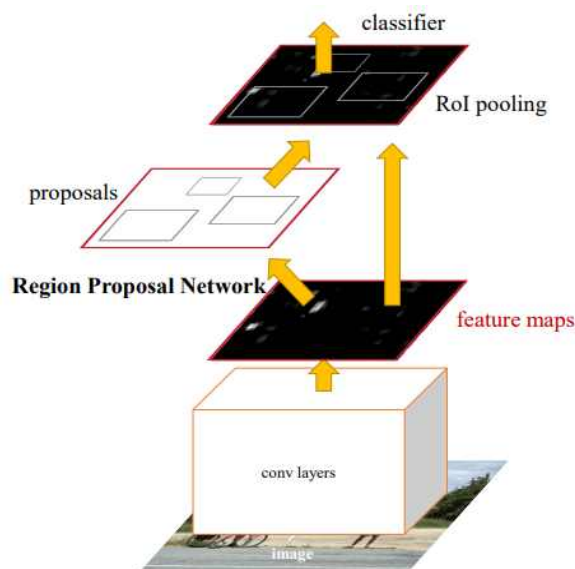


그림 6. Faster R-CNN의 모델 구조
Fig. 6. Model structure of Faster R-CNN

표 1. R-CNN, Fast R-CNN, Faster R-CNN의 입력 및 출력 비교
Table 1. Input and output comparison of R-CNN, Fast R-CNN, and Faster R-CNN

	input	outputs
R-CNN	windows of the image corresponding to objects	bounding boxes + labels for each object in the image + New bounding box coordinates for the object in the window
Fast R-CNN	images with region proposals	Object classifications of each region along with tighter bounding boxes
Faster R-CNN	images	Classifications and bounding box coordinates of objects in the images

Faster R-CNN은 각 이미지에서 분류의 첫 단계인 CNN을 통과하여 얻어진 feature map을 기반으로 영역을 제안하는 방식이다. 즉, CNN 결과를 selective search 알고리즘 대신 region proposal에 바로 이용하는 기법이다. 그림 6은 어떻게 CNN이 region proposal과 classification을 동시에 해냈는지를 보여준다. 단지 하나의 CNN만 학습시키면 region proposal이 생성되고 convolutional feature map이 Fast R-CNN처럼 region proposal에 이용될 수 있도록 하는 특징을 나타내고 있다.

표 1은 R-CNN, Fast R-CNN, Faster R-CNN을 수행하기 위한 input과 output 상태를 표현하고 있다. R-CNN의 input은 이미지에서 객체에 해당하는 윈도우가 사용되고, output으로 윈도우 안의 객체에 대한 새로운 bounding box의 좌표들이 출력된다. 빨라진 속도와 단순해진 Fast R-CNN은 input으로 region proposal이 생성된 이미지들이 사용되고, output으로 각 region에 있는 객체의 class와 꼭 맞는 bounding box들이 출력된다. 이어서 더욱 속도를 높인 영역 제안 방법인 Faster R-CNN은 input으로 region proposal이 필요 없는 원본 이미지를 사용하였고, output으로 분류 및 이미지 내부 객체들의 bounding box 좌표를 출력하였음을 나타낸다.

III. 패션 요소 검출을 위한 Mask R-CNN

3-1 패션 이미지 데이터 세트

패션과 컴퓨터 비전 사이에서 패션 산업 분야의 인공지능 서비스를 추진하기 위해서 웹 크롤링을 활용한 iMaterialist Fashion Attribute Dataset[4]가 구성되었다. 본 연구에서는 iMaterialist Fashion Attribute Dataset[4]의 복잡한 구조와 모호함을 표준화하기 위해서 27개의 주요 의류 항목 및 19개의 의류 부품을 포함하여 총 46개로 정의된 의류 개체와 294개의 세분화된 의류 속성이 포함될 수 있도록 구성한 데이터 세트[14]를 활용하여 이미지에서 패션 요소를 분류하여 검출하는 딥러닝을 수행한다.

표 2는 준비된 패션 이미지 데이터 세트를 기반으로 46개로 정의한 의류 개체(주요 의류 항목 27개 + 의류 부품 19개)를 나타낸다. 46개로 정의된 의류 개체의 super category는 상체(upperbody), 하체(lowerbody), 전신(wholebody), 머리(head), 목(neck), 팔과 손(arms and hands), 허리(waist), 다리와 발(legs and feet), 기타(others), 의류 부품(garment parts), 여밈(closures), 장식(decorations)으로 구성된다.

표 2. 46개의 타입으로 정의된 의류 개체

Table 2. Clothing objects defined as 46 types

id	name	super category
0	shirt, blouse	upperbody
1	top, t-shirt, sweatshirt	upperbody
2	sweater	upperbody
3	cardigan	upperbody
4	jacket	upperbody
5	vest	upperbody
6	pants	lowerbody
7	shorts	lowerbody
8	skirt	lowerbody
9	coat	wholebody
10	dress	wholebody
11	jumpsuit	wholebody
12	cape	wholebody
13	glasses	head
14	hat	head
15	headband, head covering, hair accessory	head
16	tie	neck
17	glove	arms and hands
18	watch	arms and hands
19	belt	waist
20	leg warmer	legs and feet
21	tights, stockings	legs and feet
22	sock	legs and feet
23	shoe	legs and feet
24	bag, wallet	others
25	scarf	others
26	umbrella	others
27	hood	garment parts
28	collar	garment parts
29	lapel	garment parts
30	epaulette	garment parts
31	sleeve	garment parts
32	pocket	garment parts
33	neckline	garment parts
34	buckle	closures
35	zipper	closures
36	applique	decorations
37	bead	decorations
38	bow	decorations
39	flower	decorations
40	fringe	decorations
41	ribbon	decorations
42	rivet	decorations
43	ruffle	decorations
44	sequin	decorations
45	tassel	decorations

표 3. 294개로 세분화된 의류 속성

Table 3. Clothing properties subdivided into 294 types

no	name	super category
0	classic (t-shirt)	nickname
1	polo (shirt)	nickname
2	undershirt	nickname
3	henley (shirt)	nickname
4	ringer (t-shirt)	nickname
.	.	.
.	.	.
289	peacock	animal
290	zebra	animal
291	giraffe	animal
292	toile de jouy	textile pattern
293	plant	textile pattern

표 3은 294개로 세분화된 의류 속성을 나타낸다. 294개의 세분화된 의류 속성의 super category는 닉네임(nickname), 실루엣(silhouette), 허리 라인(waistline), 길이(length), 목 라인 타입(neckline type), 오픈 타입(opening type), 비섬유 소재형 타입(non-textile material type), 가죽(leather), 섬유 마감 및 제조 기술(textile finishing and manufacturing techniques), 섬유 패턴(textile pattern), 동물(animal)로 구성된다.

3-2 Mask R-CNN을 활용한 패션 요소 검출

본 연구에서는 Faster R-CNN을 기반으로 생성된 bounding box를 픽셀 수준의 검출까지 가능할 수 있도록 아래와 같이 9단계의 Mask R-CNN[5] 알고리즘을 사용하여 패션 요소를 검출할 수 있도록 구현하였다.

1단계 : 선형 보간법(bilinear interpolation)을 사용해서 이미지를 resize 한다.

2단계 : 패딩(padding)을 사용해서 Backbone network의 input을 위해서 1024 x 1024의 input size로 조정한다.

3단계 : ResNet-101을 통해 각 레이어에서 feature map(C1, C2, C3, C4, C5)을 생성한다.

4단계 : FPN을 통해 이전에 생성된 feature map에서 P2, P3, P4, P5, P6 feature map을 생성한다.

5단계 : 최종 생성된 feature map에 각각 RPN을 적용하여 분류하고 bounding box regression 출력 값을 도출한다.

6단계 : 출력으로 얻은 bounding box 회귀 값을 원래 이미지로 projection 시켜서 anchor box를 생성한다.

7단계 : Non-max-suppression을 통해 생성된 anchor box 중 score가 가장 높은 anchor box를 제외하고 모두 삭제한다.

8단계 : 각각 크기가 서로 다른 anchor box들을 RoI 정렬을 통해 size를 맞춘다.

9단계 : Fast R-CNN에서의 분류, bounding box regression branch와 더불어 mask branch에 anchor box 값을 통과시킨다.

그림 7은 표 2에서 46개로 정의된 의류 개체에 대해 Mask

R-CNN을 적용하여 추출된 의류 개체의 일부를 보여주고 있다. 그림 7에서 (a)는 표 2의 id 10에 해당하는 dress, (b)는 표 2의 id 0에 해당하는 shirt와 id 6에 해당하는 pants, (c)는 표 2의 id 11에 해당하는 jumpsuit를 보여준다.

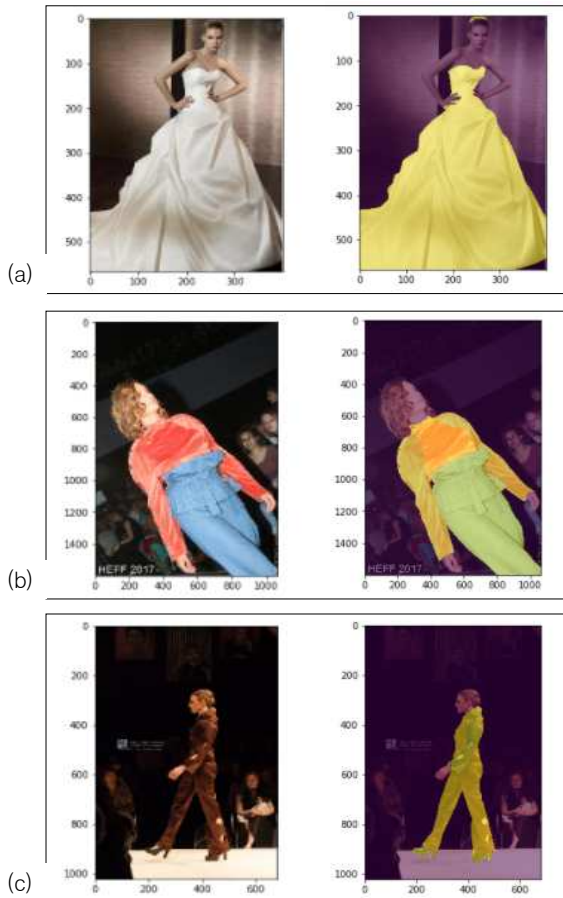


그림 7. Mask R-CNN을 이용하여 분리된 패션 요소
Fig. 7. Fashion elements separated using Mask R-CNN



그림 8. Mask R-CNN으로 검출된 패션 요소
Fig. 8. Fashion items detected by Mask R-CNN

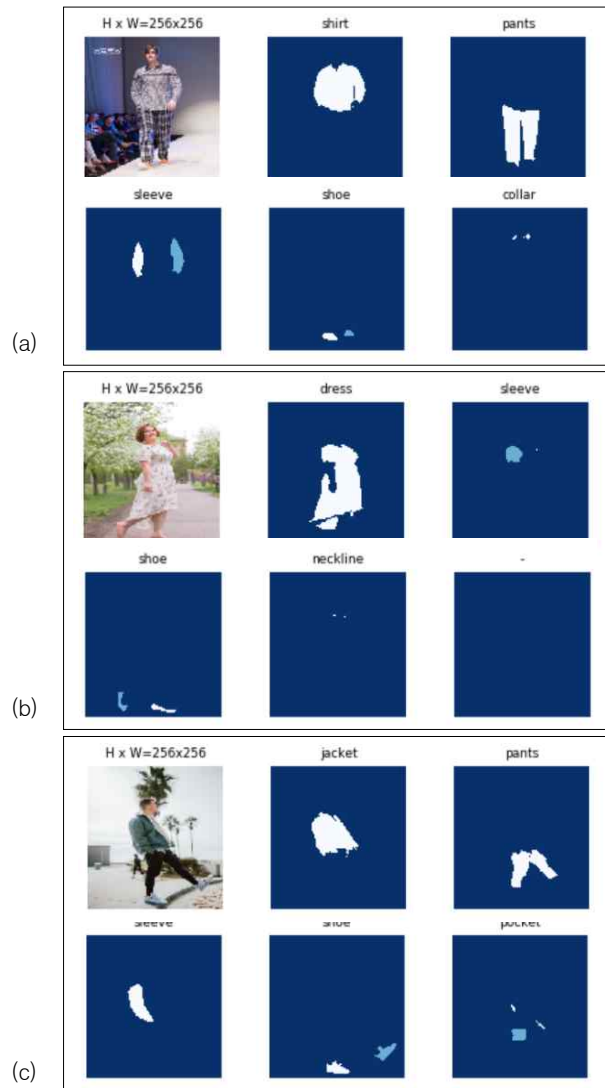


그림 9. 46개로 정의된 의류 개체의 super category 검출
Fig. 9. Super category detection of 46 defined clothing objects

그림 8은 Mask R-CNN의 마지막 단계에서 bounding box regression branch와 더불어 mask branch에 anchor box 값을 통과시킨 결과를 보여준다. 27개의 주요 의류 항목 및 19개의 의류 부품을 포함하여 46개로 정의된 의류 개체 클래스로 지도 학습된 결과에 따라 테스트 이미지를 대상으로 픽셀 단위까지 검출된 패션 요소를 출력하고 있으며 top, neckline, sleeve, applique, pocket, watch로 인식된 패션 요소를 확인할 수 있다.

그림 9에서 (a)는 46개로 정의된 의류 개체 레이블을 기준으로 shirt, pants, sleeve, shoe, collar로 분리되었고, (b)는 dress, sleeve, shoe, neckline으로, (c)는 jacket, pants, sleeve, shoe, pocket으로 분리되었음을 보여준다.

IV. Mask R-CNN 수행 후 손실 평가

Mask R-CNN의 손실 함수는 L_{cls} (classification loss) + L_{box} (bounding-box loss) + L_{mask} (average binary cross-entropy loss)로 계산된다[5]. 여기에서 L_{cls} 는 Softmax Cross Entropy, L_{box} 는 bounding box regression, L_{mask} 는 Binary Cross Entropy를 의미한다. 준비된 데이터 세트를 기반으로 총 8번의 에폭으로 딥러닝을 수행한 결과 산출된 L_{cls} , L_{box} , L_{mask} 를 표 4에 제시한다. 8 에폭 수행 후 학습 데이터의 L_{cls} 는 0.53, L_{box} 는 0.38, L_{mask} 는 0.35로 나타났다. 그리고 검증 데이터의 L_{cls} 는 0.54, L_{box} 는 0.33, L_{mask} 는 0.36으로 나타났으며 표 4의 손실 함수 결과를 시각화하면 그림 10과 같다.

표 4. 훈련 데이터셋과 검증 데이터셋을 기반으로 Mask R-CNN을 이용한 딥러닝 실행 후 손실

Table 4. Loss after deep learning execution using Mask R-CNN based on training dataset and validation dataset

epoch	train_mrcnn_class_loss	train_mrcnn_bbox_loss	train_mrcnn_mask_loss
1	1.238566	0.733075	0.68254
2	0.782887	0.569196	0.51264
3	0.814755	0.527152	0.439102
4	0.669339	0.449124	0.410754
5	0.644507	0.423773	0.389027
6	0.610574	0.394137	0.380548
7	0.618088	0.382417	0.358057
8	0.535927	0.382605	0.356228

epoch	val_mrcnn_class_loss	val_mrcnn_bbox_loss	val_mrcnn_mask_loss
1	0.832738	0.626864	0.587931
2	0.856314	0.594667	0.517932
3	0.684806	0.506972	0.49388
4	0.639155	0.535202	0.373547
5	0.696562	0.430287	0.464404
6	0.513316	0.370181	0.368127
7	0.735358	0.383918	0.368852
8	0.542074	0.336219	0.361719

패션 요소 검출에 관한 기존 방식[15]은 패션 이미지 데이터 세트를 티셔츠/탑, 바지, 폴오버, 드레스, 코트, 샌들, 셔츠, 스니커즈, 가방, 앵클 부츠와 같이 총 10개의 요소로 레이블링한 후 CNN LeNet-5 알고리즘을 이용하여 패션 요소를 검출하는 딥러닝 작업을 수행하였다. 반면, 본 연구에서는 표 2에 제시한 46개 의류 개체로 세분화하여 레이블링된 패션 이미지 데이터 세트를 기반으로 Mask R-CNN 딥러닝을 수행하여 패션 요소를 검출하였다. 그 결과 그림 8에서 제시한 것과 같이 의류 개체에 해당하는 패션 요소를 픽셀 단위까지 검출할 수 있었고, 의류 개체 레이블이 원본 이미지 위에 출력되어 직관적으로 편리하게 패션 요소가 구별되고 있음을 확인할 수 있었다.

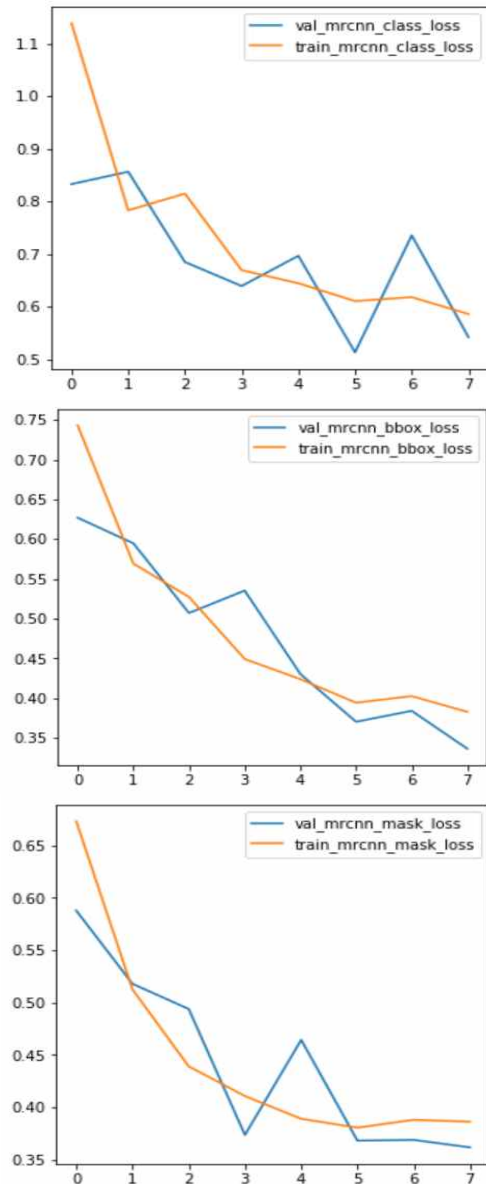


그림 10. Mask R-CNN 수행 후 손실 그래프

Fig. 10. Loss graph after performing Mask R-CNN

V. 결 론

삶의 질이 높아지고 자신의 개성을 표현하고자 하는 욕구가 강해지면서 의류 및 패션에 대한 관심도가 상승하고 있다. 이에 따라 스마트 옷장이 등장하였을 뿐만 아니라 의류 청정기와 같은 의류 관리 제품 또한 관심을 받기 시작하였다. 패션 온라인 쇼핑몰에서는 의류 및 패션 소품들을 판매할 때 인공지능 맞춤형 서비스를 제공하며 소비자에게 다가가고 있다. 하지만 구입한 의류나 패션 소품들을 적절하게 코디네이션하는 부분은 소비자 개인의 몫으로 남아있는 상황이다.

본 연구에서는 개인별 패션 코디네이션에 관한 인공지능 서비스를 제공하기 위해 패션 이미지에 Mask R-CNN 딥러닝 알고리즘을 적용하여 46개의 요소로 패션 요소가 검출되는 시스템을 구현하였다. 향후 개인이 소유하고 있는 패션 소품이나 의류를 대상으로 효과적인 패션 코디네이션을 제안하기 위해서 색상, 계절, 재질, 트렌드, 브랜드명 등의 다양한 요소들을 데이터 세트에 포함할 계획이다. 이처럼 패션에 관한 전문적인 지식 요소를 포함하여 파인 튜닝한 딥러닝으로 확장할 경우 개인의 개성을 존중하며 장소와 목적에 적합한 인공지능 패션 코디네이션 서비스가 완성될 수 있을 것으로 기대한다.

참고문헌

- [1] N. Park, and Y. Choi, "Color & Texture Attribute Classification System of Fashion Item Image for Standardizing Learning Data in Fashion AI," *The Journal of the Korean Society of Clothing and Textiles*, Vol. 44, No. 2, pp. 354-368, 2020
- [2] S. Y. Heo, S. W. Park, S. H. Jung, and J. H. Woo, "Implementation of Smart Wardrobe based on IoT," in *Proceeding of the Korea Information Processing Society Conference*, pp. 949-952, November 2018
- [3] J. U. Kim, H. J. Jung, Y. S. Kim, and K. W. Oh, "Impacts of Fashion Curation Users' Shopping Orientation, Usage Motives and Preferred Image Types on Fashion Product Purchase Intentions," *The Journal of the Korean Society of Clothing and Textiles*, Vol. 41, No. 5, pp. 796-808, 2017
- [4] S. Guo, W. Huang, X. Zhang, P. Srikhanta, Y. Cui, Y. Li, and S. Belongie, "The imaterialist fashion attribute dataset." in *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*. pp. 1-10, Jun, 2019.
- [5] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask R-CNN," in *Proceedings of the IEEE international conference on computer vision*. pp. 2961-2969, 2017
- [6] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," in *Proceeding of the IEEE*, pp. 2278-2324, 1998
- [7] 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*. pp. 1-21, 2014.
- [8] R. Girshick, "Fast r-cnn," in *Proceedings of the IEEE international conference on computer vision*. pp. 1440-1448, 2015.
- [9] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE transactions on pattern analysis and machine intelligence*, Vol. 39, No. 6, pp. 1137-1149, 2016.
- [10] J. R. R. Uijlings, K. E. A. van de Sande, T. Gevers, and A. W. M. Smeulders, "Selective search for object recognition," *The Journal of computer vision*, Vol. 104, No. 2, pp. 154-171, April 2013.
- [11] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 3431-3440, 2015.
- [12] Drucker, Harris, Christopher JC Burges, Linda Kaufman, Alex J. Smola, and Vladimir Vapnik. "Support vector regression machines," In *Advances in neural information processing systems*, pp. 155-161, 1997.
- [13] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE transactions on pattern analysis and machine intelligence*, Vol. 37, No. 9, pp. 1904-191, 2015.
- [14] iMat Fashion 2020 of the Seventh Workshop on Fine-Grained Visual Categorization [Internet]. Available: <https://sites.google.com/view/fgvc7/competitions/imatfashion2020>
- [15] K. Mohammed, A. Ahmed, and M. Hadeer, "Classification of Garments from Fashion MNIST Dataset Using CNN LeNet-5 Architecture," in *Proceedings of the International Conference on Innovative Trends in Communication and Computer Engineering*, pp. 238-243, 2020.



김희숙(Hye-Suk Kim)

1999년 : 전남대학교 대학원
전산통계학과 (이학석사)
2009년 : 전남대학교 대학원
전산학과 (이학박사)

2018년~ 현재 : GIST(Gwangju Institute of Science and Technology) Lecture Professor

※ 관심분야 : 인공지능, 딥러닝, 머신러닝, 데이터과학, 영상처리, 웹콘텐츠, UX, UI 등