인공지능 학습용 패션 데이터셋 최근 동향 조사

Hailin Jin, Zhegao Piao, 구영현, *유성준
세종대학교
*sjyoo@sejong.ac.kr

A Survey of Fashion Datasets for AI Training

Hailin Jin, Zhegao Piao, Yeong Hyeon Gu, *Seong Joon Yoo
Sejong University

요 약

패션산업은 매년 1 조원씩 성장(연평균 2.1%)하며 많은 연구자들의 관심을 받고 있다. 전통적인 패션산업은 점차 디지털화되어 선진적인 컴퓨터 비전 기술을 적용해 소비자들에게 더 좋은 쇼핑 서비스를 제공하고 있다. 본 논문에서는 2014 년부터 2019 년 사이에 구축된 대표적인 패션 데이터셋을 연도별로 정리하고 각 데이터셋에 포함된 주석(annotation)의 특징을 정리했다. 또한 데이터셋이 패션 상품 검출(Fashion detection), 패션 이미지 생성(Fashion image generation), 가상 피팅(Virtual try-on) 그리고 패션 의류 분할(Fashion Clothing segmentation) 등 연구에서의 활용될 수 있는 여부에 대해 분석했다.

1. 서론

패션은 자신을 표현하는 수단이며 개인의 특정한 감각, 실루엣, 스타일, 디자인을 가진 의복의 조합으로 현대 사회에서 필수적인 요소이다. 또한 2020 년인 올해 세계 패션시장은 약 356 조 1 천억 원(약 3,000 억 달러)으로 예상되며 2025 년에는 올해 대비 20% 증가한 약 427 조원 시장의 규모로 확대될 것으로 보인다.[1]

인공지능은 최근 5 년간 매우 빠른 속도로 발전했으며 차세대 IT 핵심분야 중 하나로 컴퓨터비전, 자연어처리, 음성인식, 로봇 등 다양한 분야에 적용되고 있다. 패션 산업에서도 전자상거래, 개성화 디자인, 패션 추천 등의 인공지능기술을 활용한 연구를 통해 소비자에게 더욱 편리한 서비스를 제공하고 있다. 패션 분야의 연구는 주로 컴퓨터 비전 또는 자연어 처리 기술을 통해 진행한다. 그 중 본 논문에서는 컴퓨터비전 기술에 사용되는 패션 데이터셋에 대해 조사하고 정리했다.

컴퓨터비전 연구의 어려운 문제점 중의 하나는 연구목적에 적합한 고품질의 데이터셋을 구축하는 것이다. 데이터셋의 품질은 모델의 최종 결과에 직접적인 영향을 줄 수 있기 때문에 컴퓨터 비전 분야에서 매우 중요한 역할을 한다.

컴퓨터비전 분야에서 많이 사용되고 있는 이미지 데이터셋으로는 ImageNet[2]이 있다. ImageNet 은 WordNet 계층에 따라 구성된 이미지 데이터셋으로 총 15,000,000 장 이미지와 20,000 개 이상의 카테고리로 구성되었고 그 중 필터링을 거친 1,200,000 장의 이미지와 1,000 개의 카테고리가 가장 많이 사용된다. ImageNet 은 2 년 동안 167 개국의 48,940 명의 근로자가 인터넷에서 수집한 10 억 장의 이미지를 정리, 분류 그리고 분석한 결과로 컴퓨터 비전 분야의 발전에 크게 기여했다. ImageNet 은 다양한 카테고리를 포함하지만 패션과 관련된 카테고리가 적다는 문제점이 있다. 따라서 전문적인 인공지능 기반 패션 모델 연구를 위한 데이터셋이 필요하다.

본 논문에서는 2014 년부터 2019 년 사이에 패션 분석 연구에 사용된 이미지 데이터셋을 조사해 아래와 같은 방법으로 요약하였다.

- 패션 데이터셋의 구축 목적 및 적용 분야

* 교신저자

- 패션 데이터셋의 이미지 개수, 해상도 등의 주석 정보를 집계
- 패션 데이터셋의 구축 연도 및 트렌드
- 패션 데이터셋의 특징

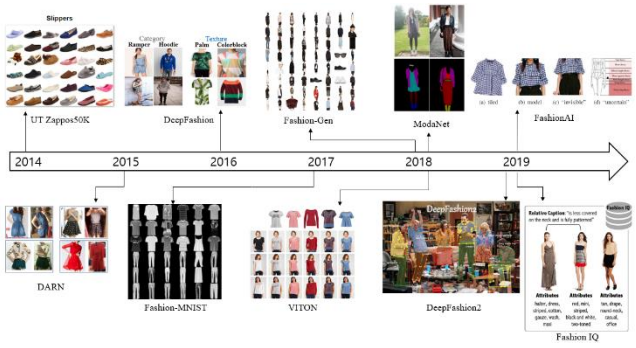

그림 1. 패션 데이터셋 타임라인&섬네일

2. 패션 데이터셋

본 장에서는 2014년부터 2019년 사이에 구축된 대표적인 패션 데이터셋의 역사를 정리하고 기술한다.

그림 1 은 본 논문에서 조사한 데이터셋들을 타임라인(Timeline)과 섬네일 (Thumbnail)로 표현한 것이다.

2.1 UT Zapos50K (2014)

UT Zapos50K[3]는 Zappos.com(온라인 신발 및 의류 매장)에서 수집한 신발 데이터셋으로 약 50,025 장의 이미지를 신발, 샌들, 슬리퍼, 부츠 등 4 가지 주요 카테고리로 정리했다. 그 밖에도 기능성 종류와 개별 브랜드와 같은 데이터도 포함되어 있으며 지금까지도 패션 영역에서 많이 사용되는 신발 데이터셋이다. 이 데이터셋은 모두 흰색 배경이며 중심 위치에는 신발이 배치되어 있다. 또한 신발은 모두 동일한 방향(왼쪽)으로 정렬되어 있어 데이터 분석과 활용에 매우 편리하다.

UT Zapos50K 데이터셋의 구축 목적은 신발 상품들의 상대 비교(pairwise comparison)를 진행하기 위해 구축되었다. 상대 비교란 여러 후보에 대한 주관적 평가를 통해 신제품 개발이나 서비스 개선을 위한 방법으로 널리 사용되고 있다. 즉 2 개의 입력 이미지에서 시각적 특성을 더 강하게 나타내는 이미지를 찾는 것이 목적이다. 이 데이터셋은 이미지의 배경이 흰색이고 단일 이미지에 한 개의 아이템을 갖고 있으며 카테고리 및 브랜드 분류, 미세 속성 학습 (fine-grained attribute learning), 성별 스타일 매칭 그리고 제로 샷(zero-shot) 학습 등의 영역에서도 많이 적용된다.

2.2 DARN (2015)

DARN[4]은 온라인 이미지와 오프라인 이미지의 (online-offline image pair) 집합으로 구성된 이미지 데이터셋이다. 온라인 이미지란 100 개 이상의 의류 속성값으로 구성된 프로토타입(prototype) 이미지이고 오프라인 이미지란 프로토타입 이미지에 대응한 사용자 이미지이다.

DARN 데이터셋은 450,000 장의 온라인 쇼핑 이미지와 90,000 장 오프라인 이미지, 총 540,000 장의 800x500 해상도를 가진 이미지로 구성되었다. 이 데이터셋은 20 가지의 의류 카테고리와 179 개의 의류 속성을 포함하고 있으며 모든 이미지는 실제 온라인 쇼핑 웹사이트에서 수집했다.

DARN 데이터셋은 모달리티(modality)의 다양성을 반영하고 있기 때문에 패션 이미지 검색 영역에서 많이 활용하고 있을 뿐만 아니라 패션 이미지 분류와 시각적 유사도 계산 분야에서도 활발히 사용되고 있다.

2.3 DeepFashion (2016)

DeepFashion[5]은 총 800,000 장의 이미지로 구성된 대규모 패션 의류 데이터셋이다. 이 데이터셋은 다양한 각도와 배경을 포함하며 그 중 단일 상품 이미지와 상품을 입고 있는 사용자 이미지도 있다. 또한 DeepFashion 은 50 개의 카테고리와 1,000 개 속성 그리고 경계 박스 (bounding box) 및 의류 랜드마크 등 패션 아이템과 관련한 풍부한 주석을 포함하고 있다.

DeepFashion 데이터셋을 기반으로 속성 예측(Attribute Prediction), 소비자-쇼핑 의류 검색 (Consumer-to-shop Clothes Retrieval), 매장 의류 검색 (In-shop Clothes Retrieval) 그리고 랜드마크 탐지 (Landmark Detection) 등 4 개의 벤치마크 (benchmark)를 제시한다.

2.4 Fashion-MNIST (2017)

Fashion-MNIST[6]는 60,000 장의 훈련 이미지와 10,000 장의 테스트 이미지로 구성된 패션 데이터셋이다. 각 이미지는 28x28 의 그레이스케일(Grayscale)로 구성되었고 0 에서 255 사이의 픽셀 값을 가진다. 그리고 10 개의 패션 카테고리가 포함되며 0 에서 9 까지의 숫자로 표시한다.

Fashion-MNIST 데이터셋은 MNIST[7] 손글씨 데이터셋을 대체하는 이미지 데이터셋으로 이미지 분류 영역에서 “Hello, World”와 같은 의미를 가진다. 패션 영역 뿐만 아니라 다른 기계학습/딥러닝 알고리즘의 입문용도로 많이 사용되기 때문에 우수한 호환성을 제공한다.

2.5 VITON (2018)

VITON[8] 데이터셋은 가상피팅을 위해 구축된 데이터셋이다.

여성 모델이 상품을 착용한 이미지와 해당 상품의 정면 이미지를 쌍으로 수집했고 그 양은 약 19,000 장이다(그림 2). 해당 이미지를 256x192 의 해상도로 resize 한 후 노이즈가 있거나 파싱 결과가 없는 것을 제거하여 최종적으로 16,253 쌍의 데이터를 구축했다. 그 중에 14,221 개를 훈련셋으로 사용하고 남은 2,032 개를 테스트셋으로 사용한다. 현재 가장 우수한 가상피팅 데이터셋으로 패션 연구에서 많이 사용한다.

그림 2. VITON

2.6 Fashion-Gen (2018)

Fashion-Gen[9]은 293,008 장의 이미지 (훈련 이미지: 260,480, 검증 이미지: 32,528, 테스트 이미지: 32,525)로 구성되었으며 48 개의 메인 카테고리 및 121 개의 fine-grained 카테고리를 포함하고 있다.

그림 3 Fashion-Gen

Fashion-Gen 의 특징은 동일한 스튜디오 조건에서 촬영된 1360 x 1360 의 높은 해상도를 가진 FHD(Full High Definition) 이미지라는 것이다. 모든 패션 아이템은 카테고리에 따라 1~6 개의 서로 다른 각도로 촬영되었다. 이는 동일한 패션 상품을 여러 각도로 촬영한 최초의 고화질 데이터셋이다. 또한 각 아이템은 그림 3 과 같이 전문가가 제공한 텍스트 설명이 포함되어 있다.

Fashion-Gen 데이터셋은 텍스트로부터 이미지를 생성(Text to image generation) 하는 문제를 해결하기 위해 사용한 데이터셋이다. 또한 높은 해상도와 다양한 각도를 가진 특성을 이용해 패션 이미지 분류, 패션 의류 검색 등과 같은 다양한 분야에서도 활용되고 있다.

2.7 ModaNet (2018)

ModaNet[10]은 Paperdoll[11] 데이터셋을 기반으로 구축된 대규모 데이터셋이다. 이 데이터셋에는 픽셀 레벨의 분할, 경계 박스, 폴리곤(Polygon) 그리고 13 개의 카테고리를 포함한 55,176 개의 완전한 주석(fully-annotated)을 포함한 이미지로 구성되어 있다.

ModaNet 데이터셋은 패션 이미지의 객체 검출, 분할, 폴리곤 예측 그리고 색상 속성을 이용한 프로토타입 예측 등 분야에서 광범위하게 사용이 가능하다.

ModaNet 데이터셋은 새로운 컴퓨터 비전 연구를 위한 벤치마크 주석 세트를 제공한다. 또한 기타 다른 데이터셋과 비교할 때 픽셀 레벨로 된 주석과 폴리곤이 있어 특히 패션 이미지 분할 영역에서 많이 사용한다.

2.8 DeepFashion2 (2019)

DeepFashion[5] 데이터셋은 풍부한 주석을 가지고 있어 패션 이미지 분석에 편리하지만 몇 가지 단점이 있다. 첫 번째로 각 이미지마다 단일 의류 아이템만 있어 정보량이 부족하다. 두 번째는 랜드마크를 가지는 이미지 데이터셋이 적어서 마지막으로 픽셀 레벨의 주석이 없어서 분할 등 연구에 적용할 수 없다.

이러한 문제를 해결하기 위해 보완된 데이터셋이 바로 DeepFashion2[12]이다. 이 데이터셋은 의류 탐지, 포즈 추정, 의류 분할 그리고 아이템 검색 등 4 가지 종류의 다용도 벤치마크로 구성된다. DeepFashion2 는 491,000 개의 다양한 이미지를 갖고 있으며 이미지의 각 항목에는 카테고리, 스타일, 경계 박스, 픽셀당 mask 등 13 개 종류의 주석이 있다. 그 외에 873,000 개는 온라인 이미지와 오프라인 이미지 쌍(DARN[4]과 같은 구조)으로 되어있기 때문에 다양하게 응용이 가능하다.

DeepFashion2 데이터셋은 풍부한 데이터와 태그로 패션 연구의 발전에 기여하며 다른 도전적인 작업에서도 우수한 잠재력을 보여준다. 예를 들어, GAN(Generative Adversarial Network) 기법을 이용한 의류 이미지 생성과 의류 멀티 도메인 학습(explore multi-domain learning) 영역에서도 사용이 가능하다.

2.9 FashionAI (2019)

세분화된 속성 인식은 패션에 대한 이해가 매우 중요하지만 이와 관련된 전문적이고 종합적인 패션 데이터셋이 부족하다. FashionAI[13] 데이터셋은 이러한 문제점을 해결하기 위해 고안되었다. 이 데이터셋은 총 357,000 개의 고품질 이미지로 구성되어 있고 24 개의 핵심 포인트, 6 가지 카테고리에서 적용되는 245 개 여성 의류 라벨, 그리고 68 개 속성이 포함되어 있다.

FashionAI 데이터셋의 특징은 트리(tree)처럼 구성된 계층적(Hierarchical) 구조이다. 계층적 구조란 그림 4 와 같은 속성

트리에 있는 모든 root 및 leaf 노드가 속성 차원과 속성 값으로 지정되는 것이다.

DeepFashion[5]과 비교했을 때 FashionAI 데이터셋은 더 정확한 콘셉트와 완전한 속성을 가진다. 또한 패션 전문가들의 확인 절차를 적용했기 때문에 데이터 품질을 보증할 수 있다. 이로 인해 FashionAI 데이터셋은 좋은 일반화(generalization) 능력을 가지고 실용성이 뛰어나며 특히 패션 의미 (semantics) 연구에 많은 도움이 된다.

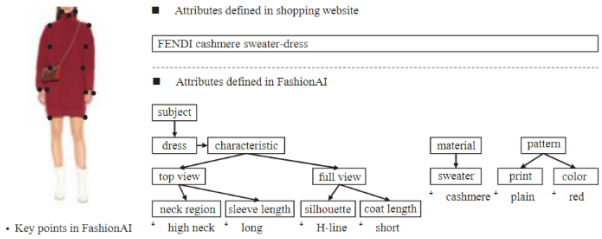

그림 4. FashionAI

2.10 Fashion IQ (2019)

Fashion IQ[14]는 자연어 기반 패션 이미지 검색을 위해 생성된 데이터셋이다. 이 데이터셋은 총 3 개의 카테고리 (dresses, tops&tees, shirts)를 포함한 77,684 장의 이미지로 구성되어 있고 그 중 49,464 장 이미지에 사이드 정보(side information)가 포함되며 60,272 개의 상대 캡션들이 있다. 사이드 정보는 질감, 직물, 형태, 구성 및 스타일 등 5 가지 주석으로 그룹화된 1,000 개 라벨로 구성된다. 그리고 상대 캡션은 참조 이미지와 사용자가 지정한 타깃 이미지의 차별성을 자연어로 표현한 설명을 의미한다.

Fashion IQ 는 상대적 캡션과 제품 설명에서 파생된 구석을 모두 사용할 수 있는 최초의 제품 지향적 데이터셋이며 자연어 피드백을 통합한 검색 시스템에 대한 연구에서 적극적으로 활용되고 있다.

3. 패션 데이터셋 비교

본 장에서는 패션 데이터셋의 주석과 기능(적용 분야) 별로 비교를 진행한다. 패션 데이터셋 간의 차이를 직관적으로 보기 위해 각 데이터셋의 이미지 개수, 해상도 그리고 주석 정보를 정리해 표기했다(표 1). 표 1 에서 볼 수 있는 바와 같이 이미지 개수가 가장 많은 데이터셋은 DeepFashion 이고 해상도가 가장 높은 데이터셋은 Fashion-Gen 이다. 쌍을 이루는 데이터셋은 DARN, DeepFashion, VITON, DeepFashion2 가 있지만 이 가운데서 DeepFashion2 의 데이터 개수가 가장 많이 포함되며 주석의 종류가 가장 풍부하다.

표 2 는 데이터셋을 classification, detection, segmentation, retrieval, generation, virtual try-on 6 개 분야의 활용 여부를 정리한 표이다. 비교한 결과를 볼 때 2019 년에 구축된 DeepFashion2 데이터셋은 classification, detection, segmentation, retrieval, generation 분야에 모두 적용이 가능하지만 가상 피팅 연구에서는 DeepFashion2 데이터셋을 사용할 수 없다. 기타 데이터셋들은 특정 목적 또는 특정 분야의 연구를 위해 구축되었기 때문에 많은 제한성을 갖는다. 또한 최신 데이터셋일 수록 특정 task 에 제한되지 않고 대부분 분야에 적용할 수 있게끔 구축되고 있다.

표 1. 패션 데이터셋의 주석 비교

	images	resolution	categories	attributes	landmarks	pairs	bounding box
UT Zappos50K	50,025	150x100	4	4	×	×	×
DARN	540,000	800x500	20	179	×	91,390	✓
DeepFashion	800,000	varying sizes	50	1000	✓	251,000	×
Fashion-MNIST	70,000	28x28	10	×	×	×	×
Fashion-Gen	325,536	1360x1360	48	×	×	×	×
VITON	16,253	256x192	×	×	×	16,253	×
ModaNet	55,176	varying sizes	13	✓	×	×	✓
DeepFashion2	491,000	varying sizes	13	✓	✓	873,000	✓
FashionAI	357,000	varying sizes	6	245	×	×	×
Fashion IQ	77,684	varying sizes	3	1000	×	×	×

표 2. 패션 데이터셋의 적용 분야 비교

	classification	detection	segmentation	retrieval	generation	virtual try-on
UT Zappos50K (2014)	✓	×	×	✓	×	×
DARN (2015)	✓	✓	×	✓	×	×
DeepFashion (2016)	✓	✓	×	✓	×	×
Fashion-MNIST (2017)	✓	×	×	×	×	×
Fashion-Gen (2018)	✓	✓	×	✓	✓	×
VITON (2018)	×	×	×	×	✓	✓
ModaNet (2018)	✓	×	✓	×	✓	×
DeepFashion2 (2019)	✓	✓	✓	✓	✓	×
FashionAI (2019)	✓	×	×	×	×	×
Fashion IQ (2019)	✓	✓	×	×	×	×

4. 결론

본 논문에서는 최근 2014 년부터 2019 년까지 구축된 패션 데이터셋의 역사와 각 데이터셋에 포함된 주석 종류와 특징을 조사했다. 조사한 내용을 기반으로 봤을 때 패션 데이터셋은 크게 두 가지 유형으로 나눌 수 있다. 첫 번째 유형인 단일 영역의 데이터셋은 해당 영역에서 많이 활용되지만 다른 영역에서는 사용이 불가능하다. 두 번째 유형인 범용적 기준 데이터셋은 여러 task 에서 적용이 가능해 최근에 많은 각광을 받고 있다.

패션 데이터셋의 발전 추세를 볼 때 향후 패션산업의 수요를 해결하기 위해 데이터셋의 종류는 더욱 다양해질 것이다. 또한 데이터셋의 주석 종류가 많아지게 되고 주석의 비율이 높아지게 되며 품질도 더욱 향상될 것으로 전망한다.

Acknowledgement

이 논문은 2019 년도 과학기술정보통신부의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임.

(2019-0-00136, 스마트시티 산업 생산성 혁신을 위한 AI 융합 기술 개발)

References

[1] 칸타 "올해 글로벌 패션 시장 규모 357 조". 패션포스트 https://fpost.co.kr/board/bbs/board.php?bo_table=newsinnews&wr_id=1010, 2020.02.12

[2] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. Imagenet: A large-scale hierarchical image database. In CVPR, 2009

[3] A. Yu and K. Grauman. "Fine-Grained Visual Comparisons with Local Learning". In CVPR, 2014.

[4] J. Huang, R. S. Feris, Q. Chen, and S. Yan. Cross-domain image retrieval with a dual attribute-aware ranking network. In ICCV, 2015

[5] Liu, Ziwei and Luo, Ping and Qiu, Shi and Wang, Xiaogang and Tang, Xiaoou. DeepFashion: Powering Robust Clothes Recognition and Retrieval with Rich Annotations. In CVPR, 2016

[6] Han Xiao and Kashif Rasul and Roland Vollgraf. Fashion-MNIST: a Novel Image Dataset for Benchmarking Machine Learning Algorithms. arXiv preprint arXiv:1708.07747, 2017

[7] Y. LeCun, L. Bottou, Y. Bengio and P. Haffner: Gradient-Based Learning Applied to Document Recognition. In IEEE, 1998

[8] Han, Xintong and Wu, Zuxuan and Wu, Zhe and Yu, Ruichi and Davis, Larry S. VITON: An Image-based Virtual Try-on Network. In CVPR, 2018

[9] N. Rostamzadeh, S. Hosseini, T. Boquet, W. Stokowiec,Y. Zhang, C. Jauvin, and C. Pal. Fashion-gen: The generative fashion dataset and challenge. arXiv preprint arXiv:1806.08317, 2018.

[10] Shuai Zheng and Fan Yang and M. Hadi Kiapour and Robinson Piramuthu. ModaNet: A Large-Scale Street Fashion Dataset with Polygon Annotations. ACM Multimedia, 2018

[11] Kota Yamaguchi, M. Hadi Kiapour, and Tamara L. Berg. Paper Doll Parsing:Retrieving Similar Styles to Parse Clothing Items. In ICCV, 2013

[12] Yuying Ge and Ruimao Zhang and Lingyun Wu and Xiaogang Wang and Xiaoou Tang and Ping Luo. A Versatile Benchmark for

Detection, Pose Estimation, Segmentation and Re-Identification of Clothing Images. In CVPR, 2019

[13] Xingxing Zo, Xiangheng Kong, Waikeung Wong, Congde Wang, Yuguang Liu and Yang CaoA. Hierarchical Dataset for Fashion Understanding. In CVPR, 2019

[14] Guo, Xiaoxiao and Wu, Hui and Gao, Yupeng and Rennie, Steven and Feris, Rogerio. The Fashion IQ Dataset: Retrieving Images by Combining Side Information and Relative Natural Language Feedback. arXiv preprint arXiv:1905.12794, 2019