

GAN 을 활용한 가상 피팅 서비스 개발 연구

박상희*, 노주영**, 송상연*, 신승화*, 김기천*

건국대학교 컴퓨터공학과*

건국대학교 응용통계학과**

e-mail : {plsbyh*, jooyoungro**, arsene96*, seu678*}@konkuk.ac.kr

A Study on Virtual Fitting Service Using GAN

Sang-Hee Park*, Joo-Young Ro**, Sang-Yeon Song*, Seung-Hwa Shin*, Keecheon Kim*

Department of Computer Science and Engineering, Konkuk University*

Department of Applied Statistics, Konkuk University**

요 약

사회가 발전함에 따라 사람들의 의류 소비 패턴 매장 매출보다 모바일 쇼핑물 매출이 늘고 있다. 스마트 뱅킹, 쇼핑물 애플리케이션 등 모바일 서비스가 일상생활로 스며들면서 모바일로 의류를 구매하는 것은 쉬워졌다. 하지만, 온라인이라는 특성상 옷을 택배로 받고, 입어야 옷이 어울리는 지 아닌 지를 판단할 수 있다는 고질적인 문제점이 있다. 이러한 문제점은 반품 또는 교환으로 이루어지고 이는 쇼핑물과 소비자 모두에게 굉장히 낭비되는 비용이다. 본 논문에서는 사람의 사진에 옷을 입힌 사진을 제공함으로써, 사람이 옷을 실제로 입지 않더라도 그 때의 fit 을 제공하고자 한다. 이때, 단순한 합성이 아니라, 딥러닝 중 GAN(Generative Adversarial Network)를 사용해 기존 기술의 문제점을 해결하고자 한다.

1. 서론

스마트 뱅킹, 쇼핑물 애플리케이션 등 모바일 서비스가 발달함에 따라 의류 소비 패턴 매장 매출보다 모바일 쇼핑물 매출이 늘고 있다. 매출이 늘어남에 따라 필연적으로 제품의 교환 및 환불의 수도 역시 늘고 있다. 또한 SNS 마켓 같은 경우의 피해 품목 중 의류·섬유·신변용품이 전체의 87.5%로 가장 높은 비율을 차지하고 있다. 아래 [표 1]에서 의류와 섬유·신변용품의 환불 상담 건수의 비중이 5.7%로 다른 품목들 중 가장 높다는 것을 확인할 수 있다.

(단위: 건)

순위	품 목 명	2018. 4		2019. 3		2020. 1		2020. 4	
		건 수	비 율	건 수	비 율	건 수	비 율	전 년 동 월 대비	전 월 대비
1	의류·섬유	3,099	4.8%	3,197	5.3%	3,489	5.7%	12.6%	9.1%
2	이동전화서비스	1,809	2.8%	1,788	3.0%	1,869	3.0%	3.3%	4.5%
3	헬스장·휘트니스센터	1,477	2.3%	1,602	2.7%	1,713	2.8%	16.0%	6.9%

[표 1] 전년 월 대비 상담 건수 증감률[1]

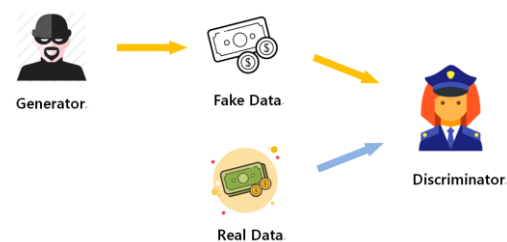
크기나 모양이 정해진 가전제품과 같은 품목들과 달리 의류라는 특성상, 자신에게 어울리는 지 아닌 지는 입어봐야 알 수 있다는 문제점이 있다. 이와 같은 특성으로 인해 선 구매 후, 교환 및 환불 방식은 쇼핑물과 소비자 모두에게 시간과 배송비가 낭비되는 문제점이 있다.

이러한 문제점을 해결하기 위해서 GAN 을 사용한 가상피팅서비스가 나왔지만 낮은 화질, 얼굴 소실 등

성능이 좋지 못했다. 실제로 이러한 낮은 성능으로 인해 실사용까지 이어지지 않았다. 따라서 실제 이용률 증대를 위해, 고화질 및 성능 개선을 한 새로운 딥러닝 네트워크 구조에 대한 연구가 필요하다.

2. 관련 연구

2.1. GAN

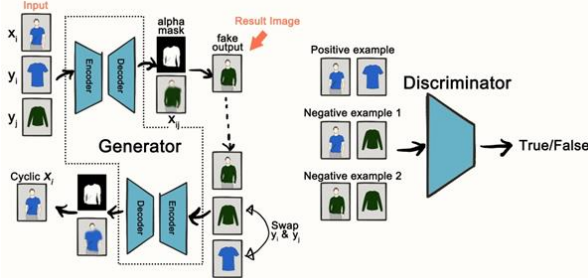


(그림 1) GAN 의 학습 과정 원리

GAN 이란 Generative Adversarial Network 의 약자다. Generative 와 Adversarial 두 단어의 뜻을 생각해보면 쉽게 GAN 의 특징을 파악할 수 있다. Generative 는 생성모형을 이용하여 그럴듯한 가짜를 생성하는 특징을 나타낸다. Adversarial 은 두 가지의 모델을 적대적으로 경쟁시키며 발전시킨다는 것을 뜻한다. 이러한 구조와 비슷한 대표적인 예로 위조지폐범(Generator)와 경찰(Discriminator)이 있다. 위조지폐범은 경찰을 속이기 위해 점점 진짜와 같이 만들고, 경찰은 위조지폐를 구별하기 위해 점점 더 세밀하게 검사한다. 그리고

시간이 많이 흐르면 위조지폐는 진짜 지폐와 구별할 수 없을 정도로 유사해진다. 즉, GAN 이란 두 가지 모델이 존재하여 한 쪽은 속이고, 다른 쪽은 구별하려고 반복하면서 구별할 수 없을 정도의 완벽한 생성 모델을 만들고자 하는 두 모델이 서로 경쟁하는 형태의 학습 모델을 말한다.

2.2. Conditional Analogy GAN



(그림 2) 본 논문의 구조도

Zalando 연구팀이 CAGAN 을 활용하여 모델이 옷을 입은 사진과 옷 단독 사진, 그리고 합성하고 싶은 옷 단독 사진 총 3 가지 이미지를 입력으로 넣어 합성하고 싶은 옷을 모델이 입은 사진을 생성하는 연구를 진행하였다. [3] 이 모델에서 네트워크는 CycleGAN 을 사용하였다. (그림 2)의 구조의 Generator 에서는 모델에 옷을 합성하고, 다시 합성한 사진을 원본 사진으로 복원하여 원본 사진과 비교해서 학습한다. Discriminator 는 기존에 입은 옷을 A, 합성하고 싶은 옷을 B 라고 할 때, 모델이 A 를 입은 사진과 A 사진, 모델이 A 를 입은 사진과 B 사진, 모델이 B 를 입은 사진과 B 사진, 총 3 가지 종류의 쌍을 입력으로 넣어 학습한다. Generator 는 Encoder-Decoder 구조보다 형태와 해상도를 유지하는 면에서 좋은 결과를 보이는 U-Net 구조를 사용하고, Discriminator 는 여러 층을 쌓은 Conv2d(2-Dimensional Convolution Layer)로 구성된다. Loss Function 에서는 모델이 옷을 입은 사진 원본과 합성한 옷을 입은 사진을 복원한 사진의 차이를 loss 로 사용한 Cycle-consistency loss 를 추가하여 학습이 진행될수록 (그림 2)에서의 x, y 와 같이 서로 다른 도메인 사이의 이미지 변환이 잘 되도록 구성하였다.

CAGAN 은 다른 종류의 옷도 적용이 가능하다. 예를 들어, 긴팔 티셔츠에서 반팔 티셔츠를 합성할 수도 있다. 하지만 사람의 얼굴이 비뚤어지고 고해상도의 이미지를 생성하기 힘들다는 문제점이 있다. 우리는 사용자에게 가상으로 옷을 입은 모습을 보여줘야 하기 때문에 사람의 얼굴이나 관절이 왜곡되면 안된다. 그래서 고해상도의 이미지에서도 좋은 성능을 보인다는 StyleGAN 을 적용해볼 예정이다.

3. 모델

3.1. Data Set



(그림 3) 이미지 데이터

데이터는 온라인 쇼핑몰을 대상으로 크롤링해 수집한다. 4 가지 종류가 하나의 데이터 셋을 이루며 (a) 모델 사진 (b)는 옷 사진을 입력 받고 이는 원본에 해당한다. 학습된 모델을 통해 (c) (d)를 생성해낸다. (a)와 (c)는 동일한 모델이어야 Discriminator 가 Generator 를 통해 생성된 사진을 판별 가능하다.

3.2. Style GAN

최근 NVIDIA 에서 발표한 논문으로 높은 성능을 가진 GAN 기반 모델이다. [6] 학습 과정 중 discriminator 와 generator 모델을 저해상도에서 고해상도로 점진적으로 확장함으로써 상당한 고화질 이미지 합성이 가능하다. Progressive growing GAN 의 generator network 에 초점을 맞추어, 2 가지 요소인 style mapping, style mixing 와 PGGAN 의 특징인 progressive learning 을 통해 성능을 향상하였다.

3.2.1. Style Mapping

기존의 방법에서는 이미지 생성시 input vector(z)로부터 직접 이미지를 생성하였다. 이 경우 고정된 input distribution 에 학습 이미지의 distribution 을 맞춰야 한다는 한계가 있다. 그러나 mapping network 를 사용하여 intermediate vector(w)로 먼저 변환한 후 이미지를 생성한다면 학습 데이터를 훨씬 유동적인 공간에 mapping 가능하고, w 를 이용하여 visual attribute 를 조절하기가 용이하다.

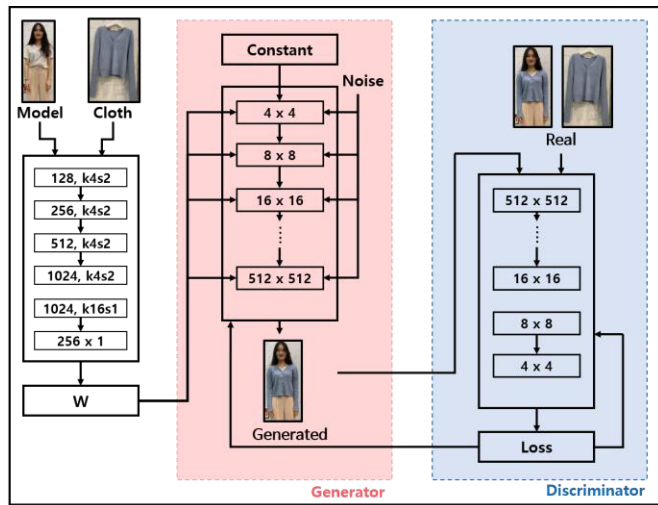
3.2.2. Style Mixing

Mapping network 로부터 나온 intermediate vector w 는 synthesis network 의 모든 layer 에 대한 style 을 표현하도록 학습된다. 이는 여러 layer 에 대한 style 이 correlate 되는 위험을 발생시킬 수 있다. 이에 대한 해결책으로 학습 시 우선 두개의 input vector 로부터 w(1), w(2)를 만들어낸다. 그리고 style 이 교체되는 layer 를 매번 random 하게 결정해 줌으로써 연속된 두 layer 간의 style 의 correlate 현상을 막을 수 있다.

3.2.3. Progressive Learning

기존의 방법에서는 이미지의 해상도 부분에서 미흡한 면이 있었다. 이것을, 처음에는 low resolution 에서 시작하여 high resolution 로의 점진적 증가를 통하여 고해상도 이미지 생성이 가능하다.

3.3. 제안 방법



(그림 3) 본 논문의 구조도

CAGAN 을 통해 데이터의 입력과 실현가능성을 봤지만, CAGAN 은 입력 이미지의 크기가 128x96 이라는 한계가 있어 고화질 이미지를 생성하지 못한다. 하지만 StyleGAN 은 입력 이미지의 크기가 1024x1024 까지 가능하다. 또한, Style Mapping, Stochastic Variation, Style mixing 등의 기법을 통해 고화질 이미지 생성을 할 수 있어 StyleGAN 을 사용하기로 했다.

구조는 Style GAN 의 구조와 거의 유사하다. 크게 condition 이미지를 축소 학습하는 Mapping Network 와 합성 이미지를 생성하는 Generator, condition 이미지와 실제/합성 이미지를 통해 구별하는 Discriminator 로 구성했다. 변경된 부분은 Mapping Network 과 Discriminator 의 입력이다.

StyleGAN 의 경우 Z 를 Mapping Network 를 통해 나온 W 를 학습에 사용한다. 하지만, 본 논문의 경우 Model, Cloth 이미지를 합성하는 것이 목적이기에 Mapping Network 에 condition 으로 Model, Cloth 이미지를 사용한다. 또한, latent variable 인 Z 를 사용해 이미지를 생성하는 것이 아니기에 Z 를 제거한다. 이미지는 256x256 이므로 CNN 을 통해 크기를 256x1 까지 축소한다. 그리고, Discriminator 에 Condition 인 Cloth 이미지를 Generated/Real 이미지와 같이 입력한다.

CNN 입력 데이터는 512x512x3 이미지 2 개이고 CNN 을 통해 256x1 로 크기를 줄인다. Generator 와 Discriminator 에서는 4x4, 8x8, 16x16, ..., 512x512 순으로 점진적으로 학습을 진행한다. 이 때, 입력 이미지 또한 화질을 4x4 와 같이 줄여 입력으로 넣는다. 학습은 Generator 가 생성한 이미지와 모델이 실제로 입은 이미지를 비교하여 학습한다. 학습이 끝나면 Mapping Network 에 Model 과 Cloth 이미지만을 입력하여 합성된 512x512x3 이미지를 Generator 가 생성한

다.

현재 데이터의 개수는 10,000(pair)개를 사용하고 있고, Google Colab 에서 학습하고 있다. Epoch 은 1000 이고, batch_size 와 learning rate 는 입력 이미지의 크기에 따라 달라져, size 는 4x4~16x16: 128, 32x32 ~64x64: 32, 128x128: 16, 256x256~512x512: 4 이고, rate 는 ~128x128: 0.0015, 256x256~: 0.002 로 설계했다.

4. 결론

본 논문은 웹과 앱으로 사용자가 지정된 input 을 upload 하면 학습된 모델을 통해 이용자가 가상으로 옷을 입는 모습을 보여주는 서비스를 제공하기 위해 StyleGAN 의 고화질 이미지 생성 능력을 활용하여 시스템을 설계 구현을 목표로 하였다. 많은 사람이 모바일 쇼핑물을 이용함에 따라, 교환 및 환불의 수도 늘어간다. 이전부터 “의류를 구매하기 전에 입어보면 좋을 텐데..., 입었을 때의 느낌을 알고 싶다.” 라는 바람이 있어왔다. 하지만, 기존의 기술들은 소비자와 쇼핑물의 만족을 이끌어내지 못했다. 결과적으로 소비자와 쇼핑물의 만족을 이끌어내 교환 및 환불로 인한 비용을 줄일 수 있을 거라 기대한다.

Acknowledgement

"본 연구는 과학기술정보통신부 및 정보통신기획평가원의 SW 중심대학지원사업의 연구 결과로 수행되었음"(No.2018-0-00213, SW 중심대학(건국대학교))

참고문헌

- [1] 소비자상담센터, 2019.05.13, "2019 년 4 월 소비자 빅데이터 트렌드", <http://www.1372.go.kr/board2/board.ccn?nMenuCode=15&gSiteCode=2&boardCode=3034&mode=view&boardSeq=2069404&mpart=>
- [2] 한국소비자원, 2019.08.08, "190807_SNS 마켓 실태조사_보도자료", http://www.kca.go.kr/brd/m_32/view.do?seq=2614&multi_itm_seq=0
- [3] Cloth Swapping with Deep Learning: Implement Conditional Analogy GAN in Keras. Shao-An Lu. 2017.10.26. <https://shaoanlu.wordpress.com/2017/10/26/reimplement-conditional-analogy-gan-in-keras/>
- [4] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. C. Courville, and Y. Bengio. Generative adversarial nets. In Advances in Neural Information Processing Systems 27, 2014. 2, 3
- [5] Jetchev, Nikolay, and Urs Bergmann. "The conditional analogy gan: Swapping fashion articles on people

images." *Proceedings of the IEEE International Conference on Computer Vision*. 2017.

[6] Karras, Tero, Samuli Laine, and Timo Aila. "A style-based generator architecture for generative adversarial networks." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2019.