
컴퓨터공학 종합설계

수정에 용이한 패션 생성기 (Editable Fashion Generator)

제출일	2021. 06. 13	전 공	컴퓨터공학
과 목	컴퓨터공학 종합설계	담당교수	이상철
학 번	12181566	이 름	권예경
학 번	12181606	이 름	박서형
학 번	12183894	이 름	정예림

목 차

I. 서론	3
II. 관련연구	3
III. 본론	4
IV. 실험결과	7
V. 결론	10
VI. 참고문헌	10

1. 서론

전 세계 패션 시장규모는 41조를 넘어서서 꾸준히 성장하고 있다. 그리고 AI는 이제 더 이상 인간에게 뒤지지 않을 만큼 크게 발전되었다. 그렇다면 창의적이고 독창적인 디자인이 중요한 패션을 AI 관점에서 바라보면 어떨까?

이러한 생각에서 고안한 본 팀의 설계 주제는 '수정에 용이한 패션 생성기'이다. 옷을 입고 있는 인물의 사진을 넣으면, AI 모델 관점에서 새롭게 변형된 스타일의 옷으로 재탄생시키는 기술을 개발하고자 하였다.

이것으로 기대할 수 있는 첫 번째 효과는 환경적인 측면에서 리사이클링이다. 해마다 기업들은 상품 개발을 위해 수많은 샘플들을 만들고 이를 버린다. 그러나 본 팀의 모델을 이용한다면, 직접 원단을 이용해 샘플 제작 없이도 새로운 디자인의 옷을 입고 있는 모델의 사진을 보면서 디자인의 상품성을 판단할 수 있게 된다.

결국 이러한 리사이클링과 함께 인간의 영역으로 여겨졌던 디자인을 AI가 할 수 있게 된다면 결과적으로 전체적인 생산 비용을 크게 낮출 수도 있다. 그리고 사용자의 측면에서 디자인 감각은 있지만, 패션 제작의 진입장벽이 높아 시도하지 못하던 많은 사람들이 쉽게 패션 디자인을 접하게 되는 매개체로 사용될 수도 있다.

2. 관련연구

다음은 관련 연구 및 이번 설계의 타겟이 된 논문이다.

2.1. 'Be your own prada: Fashion synthesis with structural coherence.'

본 팀이 타겟하는 논문은 2017 IEEE에 역선택되었던 'Be your own prada'[6]라는 논문이다. 해당 논문은 기존 이미지에 변형하고 싶은 스타일을 자연어로 넣어주면 그에 맞게 디자인을 바꿔주는 모델인데, 저희 팀은 기존 논문에 비해 심미적으로도 우수하며 패션 수정을 segmentation map을 이용해 더 직관적으로 할 수 있다는 차별성을 가지고자 하였다.

2.2. 'Fashion-AttGAN: Attribute-aware fashion editing with multi-objective GAN.'

또 다른 기존 연구 중 하나인 Fashion-AttGan 연구에서는 옷의 색깔을 바꾸는 정도에 그쳤다. 그러나 본 팀의 모델은 색깔뿐 아니라 패턴 그리고 형태까지도 변화를 주고자 하였다.

2.3. CycleGAN

CycleGAN[1]을 사용하면 이미지 대 이미지 변환에서 짝지어진 예시 없이 X 도메인으로부터 얻은 이미지를 타겟 도메인 Y 로 바꿀 수 있다. 이 때 adversarial loss를 이용해, $G(X)$ 로부터의 이미지 데이터의 분포를 Y 로부터의 이미지 데이터의 분포가 구분 불가능하도록 $G : X \rightarrow Y$ 를

학습시키고, $F(G(X))$ 가 X 와 유사해지도록 cycle consistency loss를 이용한다.

2.4. Style Transfer

의미를 명시적으로 나타내는 이미지 표현의 부족으로 이미지의 의미를 다른 스타일로 변형하는 것은 어려웠다. 이러한 한계를 CNN에서 파생된 이미지 표현을 사용하여 해결한다. 영상의 콘텐츠와 스타일을 분리하고 이를 재조합하여 임의의 사진의 내용과 예술작품의 스타일이 결합된 새로운 이미지 제작이 가능해졌다. [7]

2.5. Segmentation-map과 fashion synthesis

Segmentation-map과 fashion synthesis를 활용한 패션 합성 관련 연구[6]도 존재한다. 사람의 이미지와 새롭게 바꿀 의상을 묘사한 문장이 주어지면 의상이 원하는 대로 수정됨과 동시에 사람의 자세가 바뀌지 않고 유지된다. 이 과제를 해결하기 위해 첫 번째 GAN 생성기를 사용하여 문장 조건에 맞는 segmentation map을 생성하고 두 번째 GAN으로 생성된 맵에서 최종 이미지로 변형시킨다.

2.6. Attention

FE-GAN[8]과 Design-GAN[7] 모두 attention 기법을 활용한 패션 편집 네트워크이다. FE-GAN은 스케치와 색상 스트로크를 조작하여 새로운 패션을 생성하는데 이때 attention 정규화 레이어가 합성된 이미지의 품질을 향상시켜 네트워크의 decoder에서 여러 척도로 추가 적용된다. Design-GAN은 랜드마크 attention 및 유사성 제약 메커니즘을 활용하여 패션 카테고리 변환을 달성하는 네트워크이다. Attention 메커니즘 덕분에 CycleGAN과 InstaGAN과 비교했을 때 최종 이미지에서 실제 피부에 가깝게 표현되고 바뀐 의상의 정확도가 더 높게 측정되었음을 확인할 수 있다.

3. 본론

우선 학습에 사용한 데이터셋이다. 본 프로젝트에서는 'Deepfashion'[3]을 데이터셋으로 활용했다. 80만 장 중 비슷한 특징의 주석이 달린 데이터를 개별 카테고리로 재구성하여 6가지 특징(Denim, Floral, Striped, Tie-Dye, Baroque, Mesh)을 가지는 카테고리로 레이블링 및 1500장을 선별하여 이용하였다.

그 중에서 본 설계에서는 새롭게 변형시킬 룩의 카테고리로 데님, 스트라이프, 메쉬, 타이다이, 플로럴, 바로크 6가지에 한정하여 설계를 진행하였고, 각각의 특징은 다음과 같다.

- 플로럴 : 화사한 색감과 채도가 비교적 높은 색상들로 구성된 꽃들이 그려진 디자인, 룩을 의미한다.

- 데님 : 데님 소재의 옷감으로 만들어진 다양한 룩을 지칭한다.
- 스트라이프 : 줄무늬가 지배적인 디자인의 요소로 작용하는 룩들을 총칭하는 표현이다.
- 바로크 : 17세기의 종교양식에서 비롯한 특유의 어둡고 칙칙한 분위기와 교회의 디자인 양식에서 비롯한 무늬들로 이루어진 패션룩이다.
- 타이다이 : 서양에서 비교적 최근에 나타난 룩으로, 끈으로 의상을 묶어서 염료를 묻혀, 규칙적이지 않은 염료가 염색된 형태의 디자인으로 구성된 패션룩이다.
- 매쉬 : 매쉬소재의 시스루가 가미된 모든 패션 룩의 총칭이다.

본 팀 네트워크의 큰 줄기는 다음과 같다. fashion image로부터 segmentation map을 생성하도록 학습하고 segmentation map으로부터 특정 룩의 attribute을 갖는 fashion image를 생성하도록 학습하는 autoencoder를 만들었다

기존 CycleGAN에서는 X 도메인으로부터 얻은 영상을 Y 도메인으로 바꿀 때, G: X -> Y 생성기와 F: Y -> X 생성기를 동시에 학습했다. 하지만 본 설계에서는 X에 해당하는 패션 영상과 Y에 해당하는 segmentation map이 상이하므로 각각의 생성자를 따로 학습하는 방향으로 진행하였다. 본 네트워크에서는 (1), (2), (3) 수식의 Loss 함수를 기반으로 연산이 이루어진다. 이때 판별기에 대한 함수는 D이다.

$$\mathcal{L}_{cyc}(G) = \mathbb{E}_{x \sim p_{data}(x)} [\|F(G(x)) - x\|_2] \quad (1)$$

$$\mathcal{L}_{iden}(G) = \mathbb{E}_{y \sim p_{data}(y)} [\|G(y) - y\|_2] \quad (2)$$

$$\mathcal{L}_{GAN}(G, D_Y, X, Y) = \mathbb{E}_{y \sim p_{data}(y)} [\log D_Y(y)] + \mathbb{E}_{x \sim p_{data}(x)} [\log(1 - D_Y(G(x)))] \quad (3)$$

$$\mathcal{L}(G, D_Y) = \mathcal{L}_{GAN}(G, D_Y, X, Y) + \lambda_1 \mathcal{L}_{cyc}(G) + \lambda_2 \mathcal{L}_{iden}(G) \quad (4)$$

$$\mathcal{L}(F, D_X) = \mathcal{L}_{GAN}(F, D_X, Y, X) + \lambda_1 \mathcal{L}_{cyc}(F) + \lambda_2 \mathcal{L}_{iden}(F) \quad (5)$$

$$G^*, F^* = \arg \min_{G, F} \max_{D_X, D_Y} \{\mathcal{L}(G, D_X) + \mathcal{L}(F, D_Y)\} \quad (6)$$

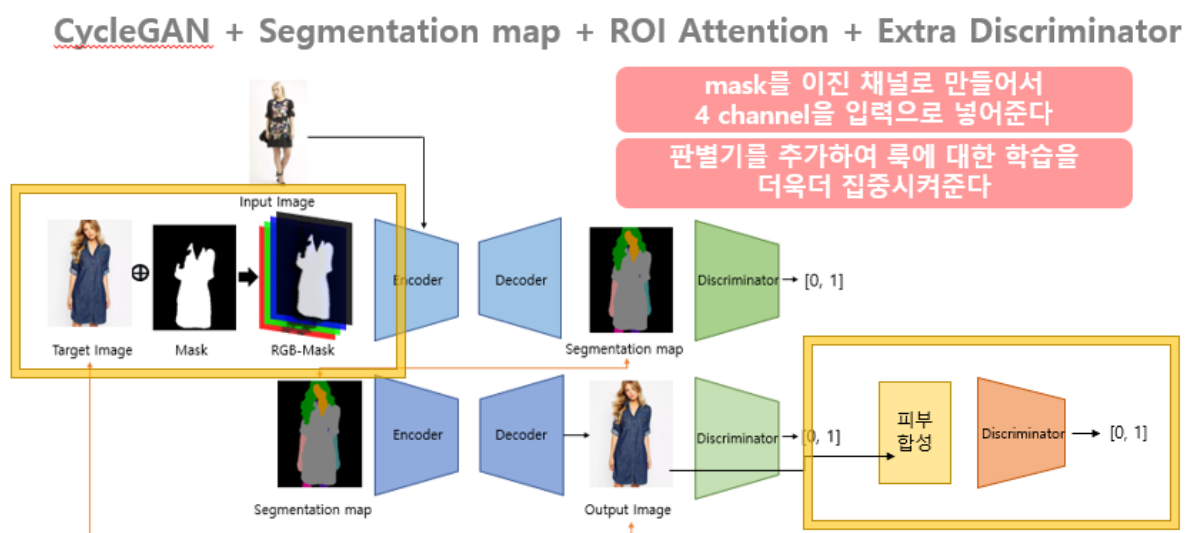
(4), (5) 수식처럼 실험적으로 각 Loss의 가중치를 구한 후, 각 생성자의 학습에는 각 수식이 개별적으로 사용된다. 마지막으로 (6) 수식과 같이 판별기와 생성기가 적대적으로 함께 학습된다.

패션을 변환, 생성하기 위해서 목적 패션의 특징 정보가 있어야 하고 이때 문자[6]나 영상[7]으로 전하는 방법이 가장 보편적이다. 하지만 본 프로젝트에서는 특정 패션 룩의 보편적인 특징을 학습을 통해 전달해 주는 방법을 채택하였다. 결과적으로 어떠한 특징을 갖는 패션이 입력으로 들

어와도 segmentation map이 생성되고 그 map의 수정이 수동적으로 이루어지더라도 학습된 특징의 패션을 생성하는 네트워크를 구현하고자 하였다.

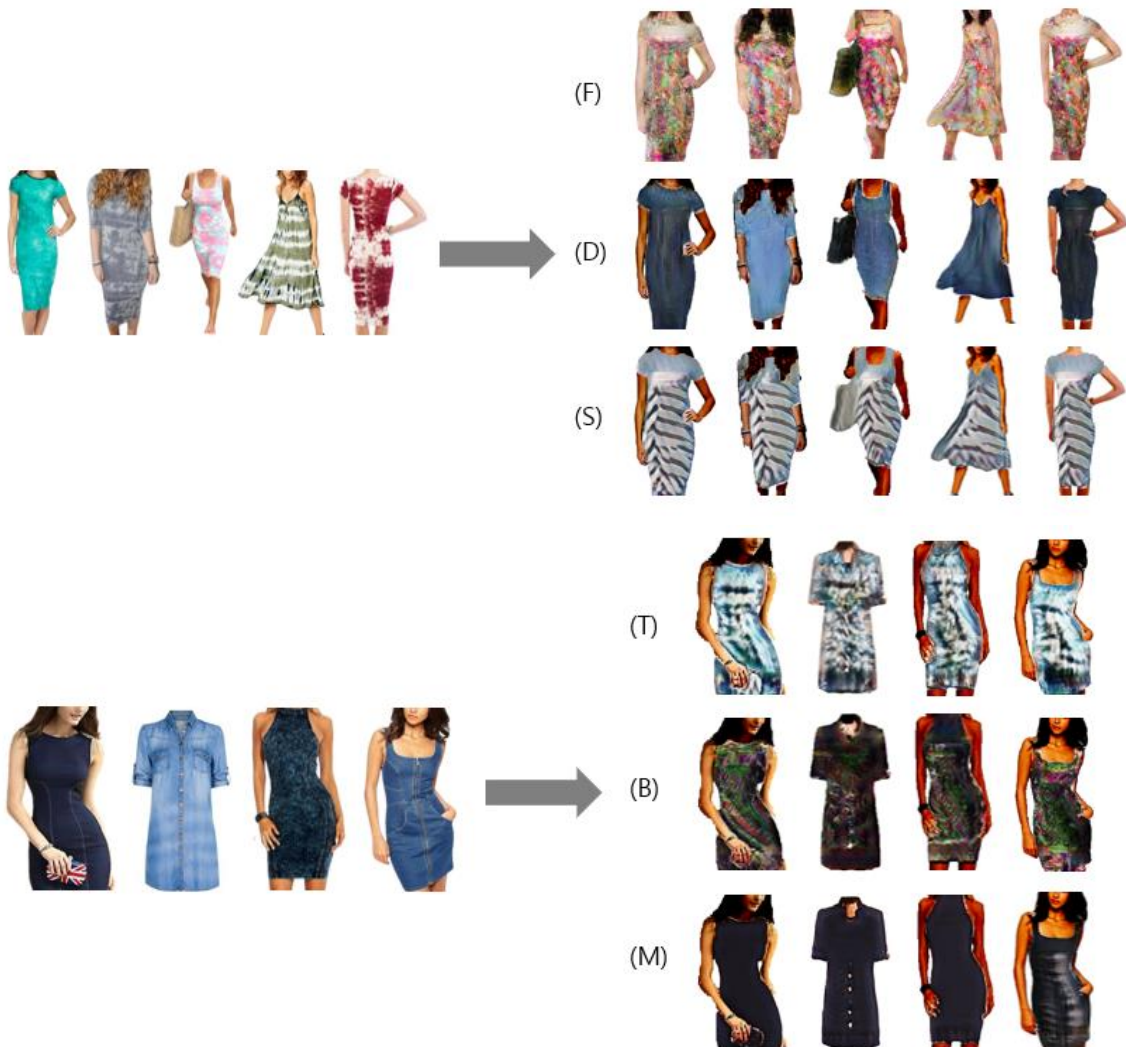
Mask 기반으로 ROI를 진행한 선행연구[5]에서는 mask를 이진 채널로 만들어서 도합 4채널을 입력으로 넣어준다. 이를 차용하여 패션 영상 도메인에서 4채널로 입력을 넣음으로써 특정 룩의 특징을 효과적으로 학습하게 유도한다. 이때 mask를 통한 패션의 특징에 더욱 집중하도록 구현하기 위해 mask에 대한 정보를 latent space 단위에서 concatenation 연산을 진행하지 않고 영상 단위에서 4채널로 학습을 진행하였다.

최종적으로 본 팀은 CycleGAN의 입력단에 기존 RGB 3채널에 mask를 이진 채널로 만들어 더해줘서 4 channel로 넣어주어 특정 패션 룩의 특징을 효과적으로 학습하게 유도하고자 하였다. 그리고 output image에 원래 이미지에서의 피부를 합성시켜 다시 한 번 더 판별기에 넣어주어서 옷의 특징을 보다 잘 잡아내도록 하였다.



4. 실험결과

다음은 본 팀의 네트워크를 통해 생성한 디자인 결과물이다. 왼쪽 영상이 입력 영상이고 오른쪽이 각 스타일에 맞게 변형된 영상이다. 위에서부터 플로럴, 데님, 스트라이프, 타이다이, 바洛克, 메쉬 순이다.



다음은 기존 CycleGAN과 본 팀의 네트워크 실험결과를 비교한 것이다. 플로럴의 화사함, 스트라이프의 패턴을 더 잘 추출하여 반영된 것을 볼 수 있고 masking과 segmentation map 덕분에 옷 테두리에 대한 분명함이 증가하였음을 확인할 수 있었다.



다음은 정량적인 평가 지표로 본 팀의 네트워크 성능을 측정해본 결과이다. 첫번째로 사용한 지표는 PSNR점수이다. PSNR점수는 두 그룹 사이의 화질 손실 정도를 나타내는 지표로 gan에 의해 생성된 이미지의 품질을 평가하는데 사용된다. 점수가 높을수록 성능이 좋다는 것인데 CycleGAN으로 만들어낸 이미지보다 본 팀의 네트워크로 만들어낸 이미지의 점수가 높았다

PSNR Score	Denim	Floral	Striped
Zhu, et al. (CycleGAN)	2.1089	1.8295	1.9970
Ours	2.2947	2.0662	3.7672

두번째 지표는 이미지 품질 평가(image quality assessment) 분야에서 가장 영향력이 있는 SSIM 지수이다. 영상 품질을 측정하는 구조적 유사도 지수로 인간의 시각적 화질 차이 및 유사도를 평가하는데 사용된다. 역시나 점수가 높을수록 성능이 좋다는 뜻인데, 다음 결과를 통해 본 팀 모델의 지수가 CycleGAN에 비해 상대적으로 높을 것을 확인할 수 있다.

SSIM	Denim	Floral	Striped
Zhu, et al. (CycleGAN)	0.0065	0.0056	0.0069
Ours	0.0183	0.0170	0.0161

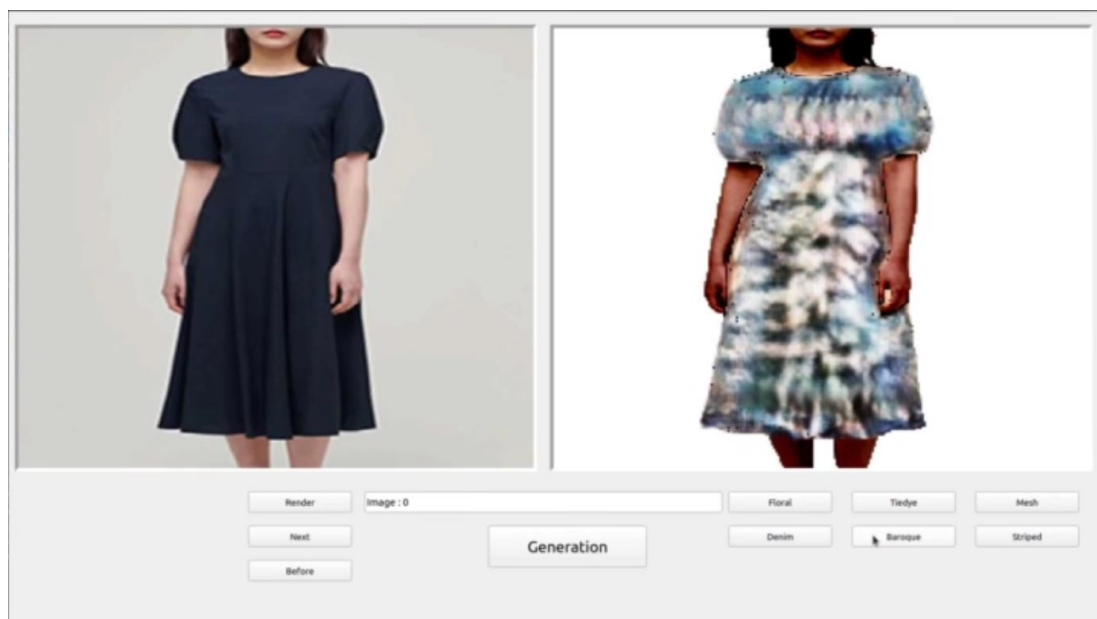
세번째 지표는 작년 IEEE에 게재된 논문 'Image Quality Assessment: Unifying Structure and Texture Similarity'[10]에서 제안된 DISTS metric이다. 구조의 왜곡에 민감하면서 텍스처 리샘플링을 용인해주는 FR-IQA방법이다. 좋은 품질일수록 0에 가까운 작은 점수를 반환하는데 본 팀 모델의 데님과 플로럴에서 더 낮은 점수를 반환해 높은 성능을 보여주었다.

SSIM	Denim	Floral	Striped
Zhu, et al.(CycleGAN)	0.0336	0.0337	0.0423
Ours	0.0300	0.0288	0.0538

마지막으로 다양한 pre-trained network를 통해 classification을 했을 때 각 네트워크 구조에서 생성된 사진의 비중을 추가함에 따라 정확도가 높아짐 또한 확인할 수 있었다. 이는 분별이 확실해질 만큼 룩에 대한 특징이 잘 반영되었다는 뜻으로 볼 수 있다.

Accuracy	Res50	Vgg16	Dense121
Real <u>Img</u>	0.82	0.55	0.68
<u>FakeDenim</u>	0.84	0.63	0.77
<u>FakeFloral</u>	0.93	0.76	0.88
<u>FakeStripe</u>	0.88	0.89	0.80
All Fake	0.99	0.88	0.96

본 팀은 이러한 과정을 거쳐 만들어진 네트워크의 실질적인 시연을 위해 PyQT5을 이용하여 다음과 GUI를 제작하여 배포하였다.



5. 결론

이번 종합설계에서는 segmentation map과 RA-CycleGAN을 기반으로 수정이 가능하면서도 다양한 룩의 특성을 반영하여 기존 디자인을 새롭게 생성하는 패션 변환 방법을 제안한다. 본 프로젝트에서 제안한 네트워크는 기존 CycleGAN과 비교했을 때 특징에 대한 학습의 강화로 재현성이 우수하다는 장점이 있다.

한계점으로는 mode collapse 현상으로 인해 generation 시 패션 디자인의 다양성을 표현하기 어렵다는 것이다. 이 문제를 해결하기 위해 WGAN의 크리티크 개념을 도입하여 립시츠 함수를 통해 안정적인 학습을 유도할 수 있고, 최종적으로 이를 통해 현재 네트워크의 문제점인 mode collapse를 해결할 수 있을 것이다. 이를 통해 더 다양한 디자인, 특히 스트라이프 혹은 플로럴처럼 mode collapse가 심한 디자인에서 빛을 발할 것이다.

6. 참고문헌

- [1] Zhu, Jun-Yan, et al. "Unpaired image-to-image translation using cycle-consistent adversarial networks." *Proceedings of the IEEE international conference on computer vision*. 2017.
- [2] Chen, Xinyuan, et al. "Attention-gan for object transfiguration in wild images." *Proceedings of the European Conference on Computer Vision (ECCV)*. 2018.
- [3] Liu, Ziwei, et al. "Deepfashion: Powering robust clothes recognition and retrieval with rich annotations." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
- [4] Li, Peike, et al. "Self-correction for human parsing." *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2020).
- [5] Song, Chunfeng, et al. "Mask-guided contrastive attention model for person re-identification." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018.
- [6] Zhu, Shizhan, et al. "Be your own prada: Fashion synthesis with structural coherence." *Proceedings of the IEEE international conference on computer vision*. 2017.
- [7] Gatys, Leon A., Alexander S. Ecker, and Matthias Bethge. "Image style transfer using convolutional neural networks." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
- [8] 강혜원, 의상사회심리학, 3장, 2012.
- [9] Victoria Gomelsky, Technology Could Turn You Into a Tiffany, The New York Times, 2021.
- [10] Ding, Ma, et al. "Image Quality Assessment: Unifying Structure and Texture Similarity" *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2021).

[11] Dong, Haoye, et al. "Fashion editing with adversarial parsing learning." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020.

[12] Lang, Yining, et al. "Design-gan: Cross- category fashion translation driven by landmark attention." ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2020.