

사진 재현 및 편집 기술 연구

2017104019 임에단

지도교수 : 박경문 교수님

요약

최근 GAN을 기반으로 한 사진 합성 기술이 지속적으로 연구되고 있다. 실제 이미지 편집을 하는 기술인 GAN Inversion은 입력된 이미지를 복구할 수 있는 적절한 latent code를 찾아내는 것이 목표이며, 해당 latent code를 찾고 조작하면 이미지를 목표에 따라 편집할 수 있다. 이미지를 고화질로 재구성 및 편집하려면 이미지 세부 정보까지 보존해야 하는데, 재구성과 편집 사이에는 trade-off 문제가 있다. 최신의 GAN Inversion 프레임워크인 HFGI는 재현과 편집에서 균형을 맞춰 이 문제를 극복하였지만, 극단적인 misalignment의 이미지 사례를 처리하는 것에 있어서 어려움이 있다. 따라서 본 연구에서는 시점 변경이 큰 이미지를 편집할 때도 충분히 개선될 수 있도록 기술을 연구하고자 한다.

1. 서론

1.1. 연구배경

최근 GAN 기술이 발전함에 따라 latent space를 조작하여 이미지를 편집할 수 있는 방법들이 많이 연구되고 있다. 이미지를 편집할 때는 원하는 속성(예: 표정, 나이)을 수정하되, 다른 세부 사항들은 유지될 수 있어야 한다. 예를 들어 StyleGAN은 여러 이미지 생성모델 중에서 이미지를 재구성하고 편집하는 데 있어서 왜곡이 낮고 품질 또한 뛰어나다. 하지만 원본 이미지의 특징을 살리면서 특정 부분을 변형하는 부분에서는 완벽하지는 못하다. 인코더 기반의 GAN Inversion 기술에서는 실제 이미지를 저차원의 latent code로 압축하게 되면 필연적으로 정보 손실이 일어나게 되는데, 손실되는 정보는 주로 이미지의 세부 정보이다. 이미지를 고화질로 재구성 및 편집을 하려면 이러한 세부 정보까지 모두 보존해야 할 필요가 있다. 이 문제점을 해결하기 위해 다양한 프레임워크가 발표되고 있으며, 그중 재구성과 편집이 모두 개선이 잘 이루어진 High-Fidelity Gan Inversion 프레임워크는 극단적인 misalignment인 이미지 사례를 처리하는 것이 어려운 한계가 존재한다. 따라서 HFGI 프레임워크에서 시점 변경이 큰 이미지 또한 편집할 때에도 충분히 개선되어 보다 더 나은 기술이 될 수 있도록 연구를 진행하고자 한다.

1.2. 연구목표

인코더 기반의 large-rate GAN inversion system으로 이뤄진 HFGI 프레임워크에서의 한계점이었던 extreme misalignment의 이미지들이 편집할 때 충분히 개선될 수 있도록 기술을 연구하는 것이 본 연구목표이다.

2. 관련 연구

2.1. GAN Inversion Approach

2.1.1. GAN

GAN (Generative Adversarial Networks)은 생성자(Generator)와 구분자(Discriminator) 두 네트워크를 적대적(Adversarial)으로 학습시키는 비지도 학습 기반의 생성모델(Unsupervised generative model)이다. GAN은 생성자가 만든 가짜 데이터가 진짜 데이터와 비슷하여 판별자가 진위를 판별하지 못할 때까지 알고리즘을 개선하는 방식으로 학습을 진행한다. 이처럼 GAN으로 학습하는 생성자는 진짜 같은 가짜 데이터를 만들어내기 때문에, 유명 화가의 화풍을 입힌 이미지나 음성 변조 파일, 영상 등 다양한 콘텐츠 분야에서 활용되고 있다.

2.1.2. GAN Inversion

GAN Inversion은 GAN과는 반대로, 주어진 실제 이미지를 재구성할 수 있는 가장 적절한 latent code를 찾아내는 기술이다. 기존의 GAN Inversion 접근법으로는 세 가지로 분류가 가능하다. (1) Optimization-based (2) Encoder-based (3) Hybrid. 최적화 기반의 GAN Inversion 기술로는 I2S, PTI 등이 있으며 이미지 별로 최적화가 이루어지기 때문에 재구성의 품질이 뛰어나지만 추론에 있어서 시간이 많이 소모되고, 편집 능력이 떨어진다. 인코더 기반은 반대로 편집 능력이 좋고 빠른 추론이 가능하지만, 실제 이미지가 latent code로 압축되는 과정에서 정보 손실이 발생하기 때문에 재구성의 품질이 좋지 않으며 예시로 pSp와 e4e 등이 있다.

다른 접근법으로는 사용된 latent space에 따라 분류가 가능하다. Z는 단순하지만 특징의 entanglement에 있어서 문제가 있다. W와 W+는 보다 이미지의 실분포를 잘 설명해주고, 실제 이미지 요소에 따라 잘 나뉘어있기 때문에 편집이 용이하다. 그 외에 W+를 변환한 S와 비선형 매핑 네트워크의 마지막 활성화 계층으로 inversion하는 P 공간이 있다. 이러한 latent space는 실제 이미지가 필연적으로 세부 정보를 손실된다.

2.1.3. Rate-Distortion-Edit Trade-Offs

GAN Inversion 기술에서, Information bottleneck 이론에 따르면 깊은 압축으로 인해 손실되는 정보는 주로 이미지 세부 정보(high-frequency 패턴)이다. Low-rate latent code은 차원이 낮으므로 일부 정보가 불가피하게 손실되고, High-dimension으로 차원을 올리는 방법은 reconstruction 품질이 좋아지지만, 과적합 되기 쉬워지기 때문에 편집성이 떨어진다. 이처럼 Reconstruction과 Editability는 trade-off 관계에 있다. 따라서 이미지 생성 및 편집하는 기술은 editability의 성능을 손상시키지 않으면서 reconstruction의 성능(fidelity)을 올릴 수 있도록 균형이 잘 맞는 섬세한 시스템의 설계가 필요하다.

2.2. High-Fidelity GAN Inversion

Encoder-based 방식에 DCI 방법과 ADA 모듈이 추가된 새로운 GAN Inversion의 프레임워크로 이미지의 세부 정보가 잘 보존된 속성 편집을 가능하게 하는 High-Fidelity GAN Inversion이 연구되었다.

2.2.1 Distortion Consultation Inversion (DCI)

low-rate latent vector이 놓쳐버린 high-frequency한 이미지 세부 정보를 갖는 “distortion map”을 활용한다. Distortion map은 무시된 이미지 세부 정보를 다시 가져와 해당 논문에서 새로 고안된 consultation encoder를 통해 high-rate latent map에 투영되고 low-rate latent vector과 융합을 한다. 네트워크는 generation을 위한 참조로 이미지 세부 정보를 명시적으로 consult 하게 되며 기존의 기본 인코더가 보완이 된다.

2.2.2 Adaptive Distortion Alignment (ADA)

이미지를 속성 편집할 때 low-rate latent code W 는 다음과 같이 특정한 semantic 방향에 따라 이동한다 : $W^{edit} = W + \alpha N^{edit}$ 이 경우 general하게 작동하지만 하지만 distortion map에서 inversion과는 다르게 editing된 이미지는 misalignment가 발생하는 문제가 있다. ADA 모듈은 이러한 misalign을 잡아줄 수 있다. ADA는 encoder-decoder 구조이며, self-supervised learning을 한다. Ground Truth는 기존의 distortion map이고, 랜덤한 augment를 distortion map에 적용한 것과 GT의 차이를 L1 loss로 적용하여 최적화를 한다. alignment loss는 다음과 같다 : $L_{align} = \|\hat{\Delta} - \Delta\|_1$

3. 프로젝트 내용

3.1. 시나리오

본 프로젝트는 얼굴 영역을 위주로 실험할 것이기 때문에 FFHQ 데이터 셋을 훈련에 사용한다. 연구를 검증할 때는 CelebA-HQ 데이터 셋을 활용한다. 개선된 모델은 기존의 HFGI와 비교할 것이다. 생성된 이미지로 잘 개선되었는지 판단하기 위해 quantitative와 qualitative 평가를 할 예정이다. 기존의 HFGI 한계점을 극복하기 위해 종합적인 Losses 식이나 상수를 변형하거나, 편집하고자 하는 속성에 더 집중할 수 있도록 다양한 방법으로 시도해볼 예정이다.

3.2. 요구사항

인코더 기반의 large-rate GAN inversion system으로 이뤄진 HFGI 프레임워크에서 한계점이었던 extreme misalignment의 이미지들 또한 편집할 때 충분히 개선이 되어야 한다. 본 연구에서는 이미지를 편집하는 속성들 중에서도 예를 들어 “pose” 부분에 있어서 해당 프레임워크가 더욱 좋은 성능을 보여줄 수 있도록 요구된다.

4. 향후 일정 및 역할 분담

[향후 일정]

9/16 : 주제 및 향후 면담 일정 논의

9/19 : 주제 확정 및 GPU 서버 대여

9/26 : HFGI 논문 리뷰 발표

9/30 : 기초조사서 제출

10/3 : pSp, e4e, Restyle 논문 리뷰 발표, HFGI 코드 실행 및 결과 발표, 모델의 코드 리뷰

10/10 : InterStyle, HyperStyle, PTI 논문 리뷰 발표, HFGI 모델 개선 연구

10/14 : 논문 접수 및 제출

10/17~~ : HFGI 모델 개선 연구

10/28 : 중간보고서 제출

11/25 : 최종보고서 제출

11/28 : 논문 최종본 접수

12/5 : 오프라인 발표 평가

[팀원 구성 - 개인]

2017104019 컴퓨터공학과 임에딘

5. 결론 및 기대 효과

사진의 재현과 편집에 있어서 모두 성능이 뛰어난 GAN 반전은 계속해서 연구되고 있으며, 인코더 기반의 large-rate GAN inversion system으로 이뤄진 HFGI 프레임워크에서의 한계점이었던 extreme misalignment의 이미지들이 편집할 때 충분히 개선될 수 있도록 연구를 한다면 사진 재현 및 편집의 능력이 보다 더욱 향상될 것이다.

6. 참고문헌

[1] Alaluf, Yuval, et al. "Hyperstyle: Stylegan inversion with hypernetworks for real image editing." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022.

[2] Alaluf, Yuval, Or Patashnik, and Daniel Cohen-Or. "Restyle: A residual-based stylegan encoder via iterative refinement." *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021.

- [3] Richardson, Elad, et al. "Encoding in style: a stylegan encoder for image-to-image translation." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2021.
- [4] Tov, Omer, et al. "Designing an encoder for stylegan image manipulation." *ACM Transactions on Graphics (TOG)* 40.4 (2021): 1-14.
- [5] Wang, Tengfei, et al. "High-fidelity gan inversion for image attribute editing." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022.