




KUBIG CONTEST

[CV분반]
StyleGAN과 StyleCLIP을
이용한 사용자 인터페이스
기반 이미지 조작 Project



0. Contents

1. Concept of StyleGAN and StyleCLIP

1.1 What is StyleGAN

1.2 What is StyleCLIP

2. Project

2.1 Image2StyleGAN

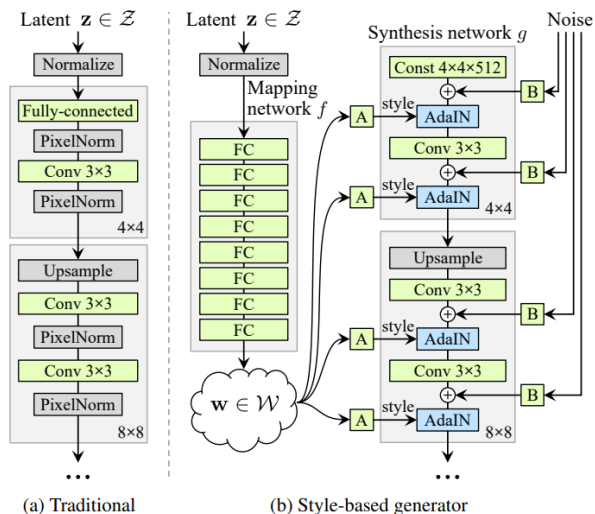
2.2 Latent Interpolation

2.3 StyleCLIP

3. Complement

1.1 What is StyleGAN ?

- Image Generation Model
- Input으로 random한 **latent vector**를 받아서 **학습 이미지와 비슷한 이미지**를 생성
- 기존 GAN 모델과 다르게 **Mapper network, AdaIN** 등 새로운 방법론을 활용해, 생성하는 이미지 특징들이 Disentangle하게 금 학습함과 동시에, 고해상도의 이미지를 생성할 수 있음.



1.2 What is StyleCLIP ?

- Text-Driven Manipulation of StyleGAN Imagery
- **StyleCLIP = StyleGAN + CLIP**

Input:



Text prompt: "Mohawk hairstyle" "Without makeup"

"Cute cat"

"Lion"

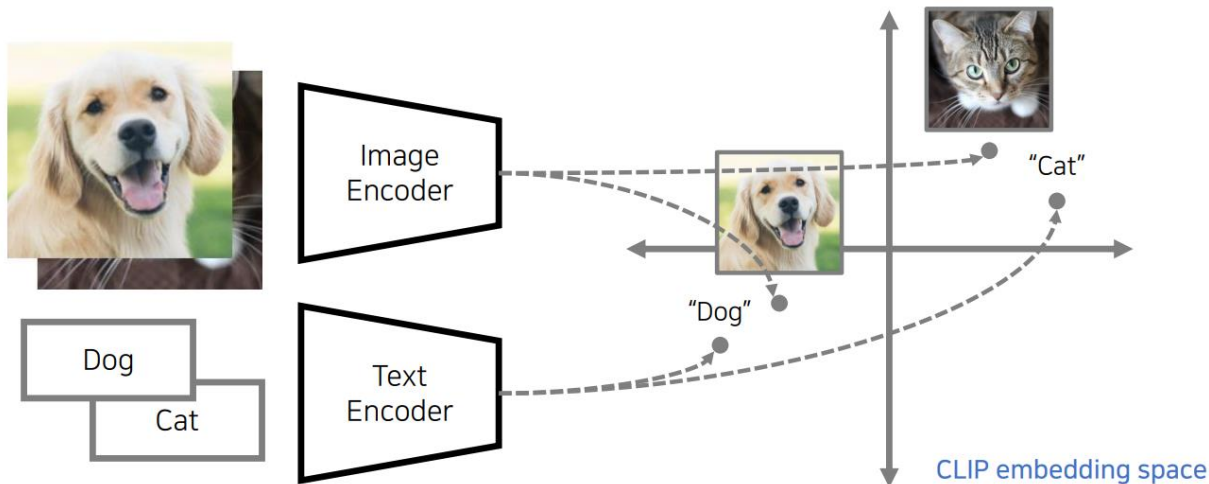


Result:



1.2 What is StyleCLIP ?

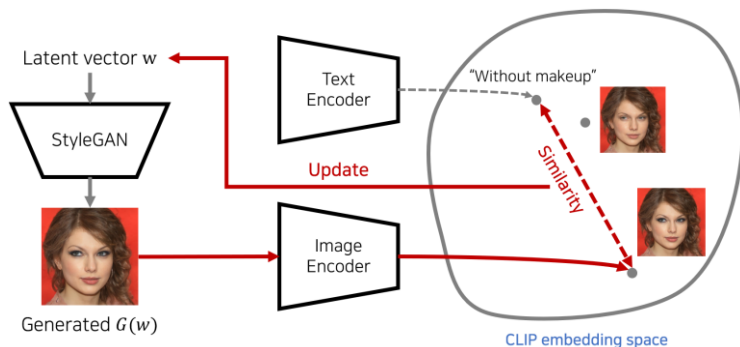
- CLIP은 대용량 데이터셋으로 이미지 인코더와 텍스트 인코더를 함께 훈련시킴
- 구한 latent vector 이용해 만든 이미지와 사용자가 넣어주는 text를 CLIP에 넣어서 loss 계산
- **이미지**와 **텍스트**를 동시에 받아 둘을 각각의 인코더를 거친 후 같은 **임베딩 스페이스 상 거리를 측정**하는 도구 -> 이미지와 텍스트 사이의 **유사도 계산** 가능 (**CLIP loss**)



1.2 What is StyleCLIP ?

1. Latent Optimizations

- 기본 이미지의 형태를 유지하면서 사용자가 넣어준 텍스트를 잘 표현하는 이미지 만들 수 있도록 latent vector 업데이트



$$\arg \min_{w \in \mathcal{W}} D_{\text{CLIP}}(G(w), t) + \lambda_{\text{L2}} \|w - w_s\|_2 + \lambda_{\text{ID}} \mathcal{L}_{\text{ID}}(w)$$

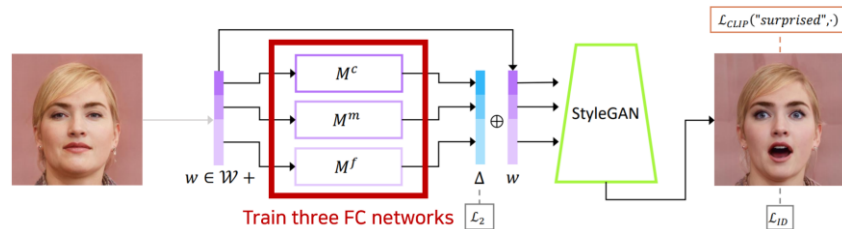
For manipulation
For similarity to the input image

$$\mathcal{L}_{\text{ID}}(w) = 1 - \langle R(G(w_s)), R(G(w)) \rangle$$

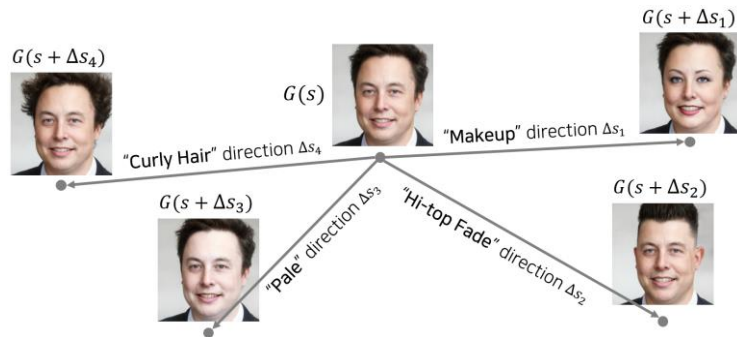
R : Pretrained ArcFace network

D_{clip} : Cosine distance between the CLIP embeddings

2. Latent Mapper



3. Global Directions



2.1 Image2StyleGAN - Concept

- 특정 이미지가 주어졌을때, 이를 representation할 수 있는 **latent vector**를 찾는 Task
- Pretrained StyleGAN, VGG16을 이용해 생성한 이미지와 원본 이미지와 loss 계산
 >> **latent vector** 업데이트
- 이렇게 찾은 특정 이미지의 **latent vector**를 통해 **이미지의 Style 변경**, 또는 **다른 이미지로 전환 가능**

2.1 Image2StyleGAN - Results

- 구글에서 많은 연예인들의 사진들을 찾을 수 있었고, 결과적으로 다양한 사진에 대한 latent vector를 얻을 수 있었음.
- Latent vector들 중 상당수는 입력 이미지를 잘 represent했지만, 일부는 그렇지 못했음.

Latent vector 잘 찾은 경우



Latent vector 잘 찾지 못한 경우



2.1 Image2StyleGAN – Result Analysis

➤ Representation 좋은 latent vector 가지는 이미지의 특징

1. 보통 **서양인**
2. 사진에 **얼굴**이 차지하는 비율이 **80~90%**
3. 옆모습보다 **정면**으로 찍은 얼굴

Model pretrain에 쓰인 FFHQ 데이터에 의한 특징



2.2 Latent Interpolation - Concept

- 이미지에 대응하는 Latent vector를 변화시키며 **이미지의 변화 양상** 관찰
- 앞서 Image2StyleGAN을 통해 **2개**의 이미지에 대한 **latent vector** 찾음(latent_A, latent_B)
- 두 latent vector를 계수합이 1인 **linear combination**으로 표현 ($a \cdot \text{latent_A} + b \cdot \text{latent_B}$, $a + b = 1$)
- 계수 값 변화에 따른 이미지 변화 관찰



Example of latent Interpolation

2.2 Latent Interpolation – Results

로제와 제니의 latent Interpolation

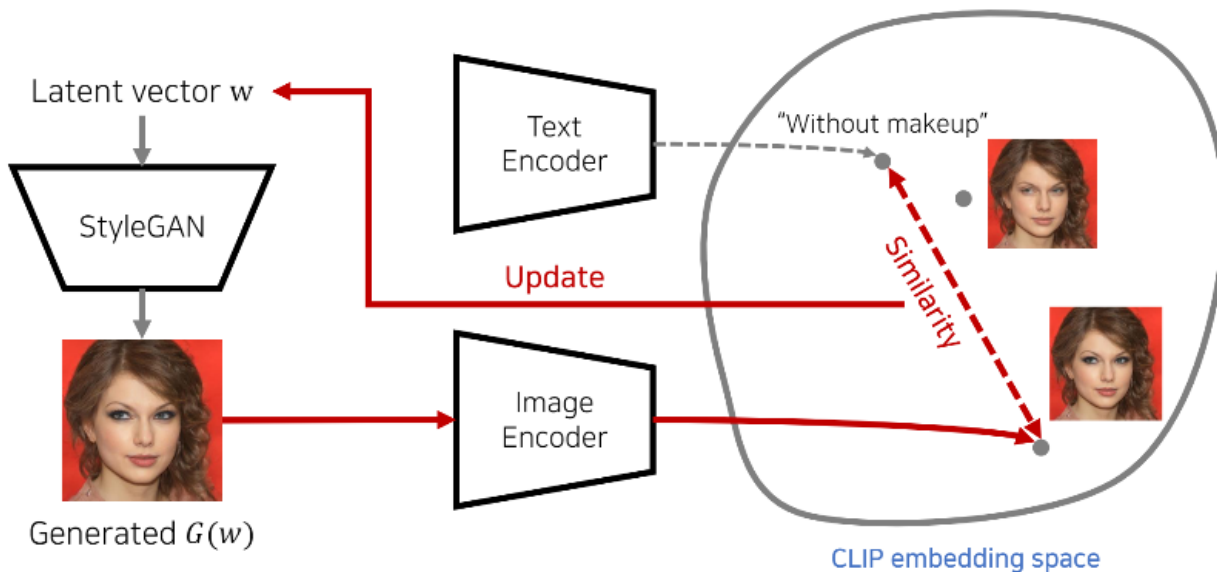


오바마와 조커의 latent Interpolation



2.3 StyleCLIP - Concept

- StyleCLIP을 구현하는 3가지 방법 중 첫번째 방법 **Latent optimization** 사용
- 앞서 Image2StyleGAN에서 찾은, 이미지를 잘 representation하는 latent vector 사용



2.3 StyleCLIP – Results (Good case)

- 실험결과 대체로 어느정도 text를 잘 반영하는 이미지를 생성가능했음.

Latent vector: Jablonski
Text: really angry face



Latent vector: 엠마 스톤
Text: smile face



2.3 StyleCLIP – Results (Bad case)

- 일부 latent vector에서는 이미지 왜곡이 생기면서 잘 작동하지 않는 것을 확인할 수 있었음.

Latent vector: 로제
Text: really angry face



Latent vector: 제니
Text: smile face



2.3 StyleCLIP – Result Analysis

➤ 실험하면서 알게 된 사실 및 분석

1. 원본 이미지를 아무리 좋은 quality 로 representation 하는 **latent vector**를 사용하더라도, StyleCLIP에서 잘 작동하지는 않는 사실을 알았음.
2. StyleCLIP에서 역시 **한국인 이미지**보다 **서양인 이미지**에서 더 잘 작동하는 것을 확인할 수 있었는데, 아마도 StyleGAN을 pretrain할 때 사용한 데이터의 bias문제라고 판단됨.
3. 향후 한국인 이미지 관련 프로젝트를 한다면 다른 pretrain dataset을 사용할 필요가 있어보임.
(ex: 한국인 face dataset)

3. Complement

- 추가 보완점으로 ‘인물 정면 사진 + 한국인 이미지’를 소재로 StyleCLIP 모델 학습에 도전해보는 방안 언급
- AI Hub “한국인 안면 이미지 소개” <https://aihub.or.kr/aidata/73>

AI Hub

개방 데이터 ▾ 외부 데이터 ▾ 활용 사례 ▾ 개발 지원 ▾ 경진대회 ▾ 게시판 ▾ 로그인 회원가입

개방 데이터

비전

음성/자면어
교육
국토환경
농축수산
안전
자율주행
헬스케어

한국인 안면 이미지 소개

소개 다운로드

데이터셋명	한국인 안면 이미지	데이터 유형	이미지
데이터 분야	비전	담당자명	조창현(한국과학기술연구원)
구축기관	한국과학기술연구원	전화번호	02-958-6650
가공기관		이메일	klface@imrc.kist.re.kr
접수기관	데이터 관련 문서	구축년도	2017년, 2018년, 2019년
구축 데이터량	2,860만	최종수정일자	2018.01.02
버전	1.0		
소개	한국인 얼굴 특성에 맞는 각도, 조도, 가림, 표정, 해상도 등으로 한국형 얼굴 데이터 베이스를 구축한 이미지 데이터 제공		
주요 키워드	얼굴 특징, 얼굴 인식, 얼굴 생성, 얼굴 변환, 얼굴 판별, 한국인 얼굴, 얼굴 이미지, 얼굴 포즈 이미지, 얼굴 표정 이미지		
저작권 및 이용정책	본 데이터는 과학기술정보통신부가 주관하고 한국지능정보사회진흥원등이 지원하는 "인공지능 학습용 데이터 구축사업"으로 구축한 데이터입니다. [데이터 이용정책 상세보기]		
데이터설명서	자료보기	구축활용가이드	자료보기
샘플데이터	다운로드	교육활용동영상	
지력도구	다운로드	AI모델	

연구가능 학습용 데이터 다운로드 프로그램 설치

Windows & Mac용

> 간단 사용설명서
> 백열 설치설명서 가이드

Ubuntu Ver.18.04용

> 간단 사용설명서



구현코드 및 참고논문

- https://github.com/ChoiDae1/StyleGAN_Project README.md 파일 참고.



Thank You 😊

