

Style GAN: 고화질 이미지 생성에 적합한 아키텍처 제안

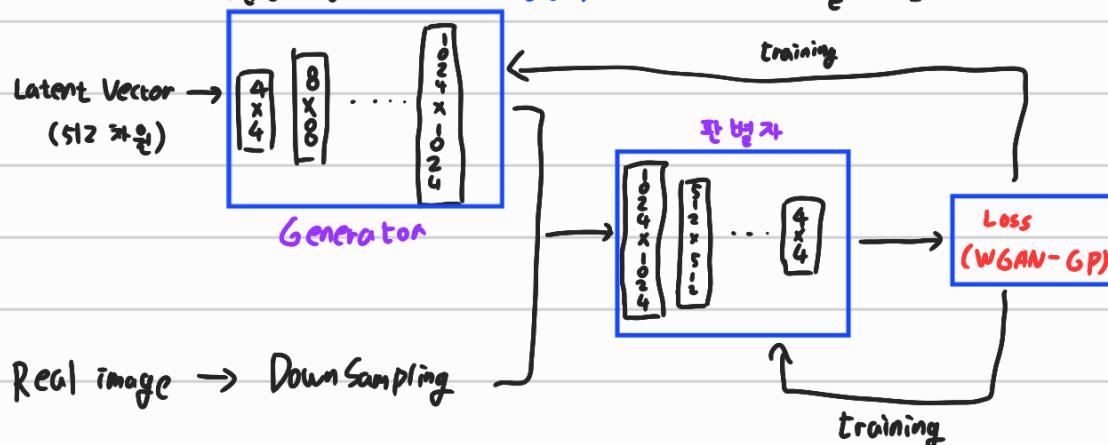
- ① PGGAN 베이스라인 ② Disentanglement 독성 향상 ③ FID라는 고해상도 얼굴데이터셋 발표
↳ 다양한 특징들이 잘 분류되어 있는 것
↳ 성별, 연령, 이론 등을

Related Work

GAN, DCGAN, WGAN-GP (loss function), PG GAN

PGGAN:

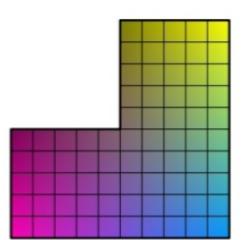
- Main Idea: 학습할 때 layer를 추가 시켜면서 고해상도 이미지 학습 성공
- Limitation: 이미지 특징 제어 어려움
- Architecture: 학습을 진행하는 과정에서 점진적으로 layer를 블리거나 끌어내감



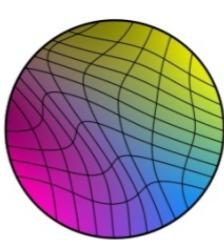
Style GAN 핵심 아이디어 (Mapping Network)

Latent z를 바로 Network에 넣는 것이 아니라 w latent vector로 매핑을 한 후에 사용

⇒ 도메인 z ⇒ 도메인 w로 mapping (\because 이미지의 특징을 추출하기 위해) \Rightarrow 각 이미지의 특징을 잘 분류할 수 있다.



(a) Distribution of features in training set



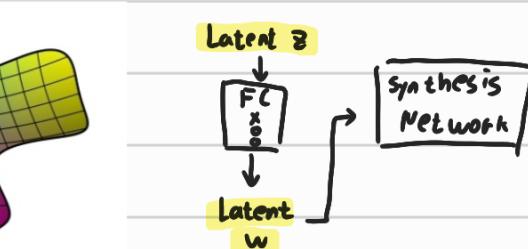
(b) Mapping from Z to features



(c) Mapping from W to features

원래 이미지 데이터셋
"세로축: 성별 (남성, 여성)
가로축: 머리길이 (한국인, 유럽인)"

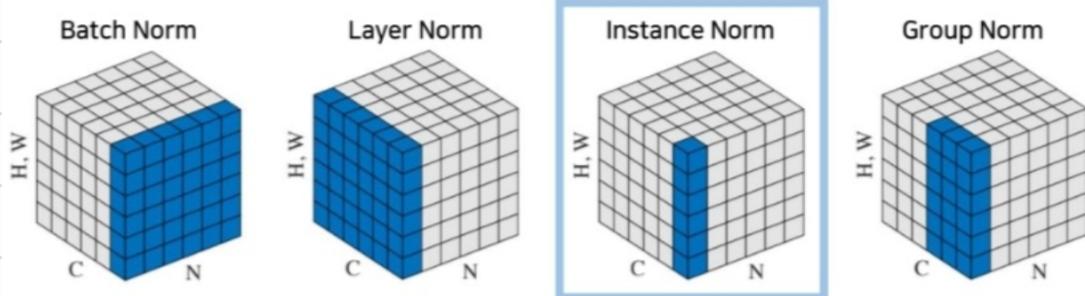
가우시안 Sampling z vector
특징의 변환 가능성
(entanglement) (nonlinear)



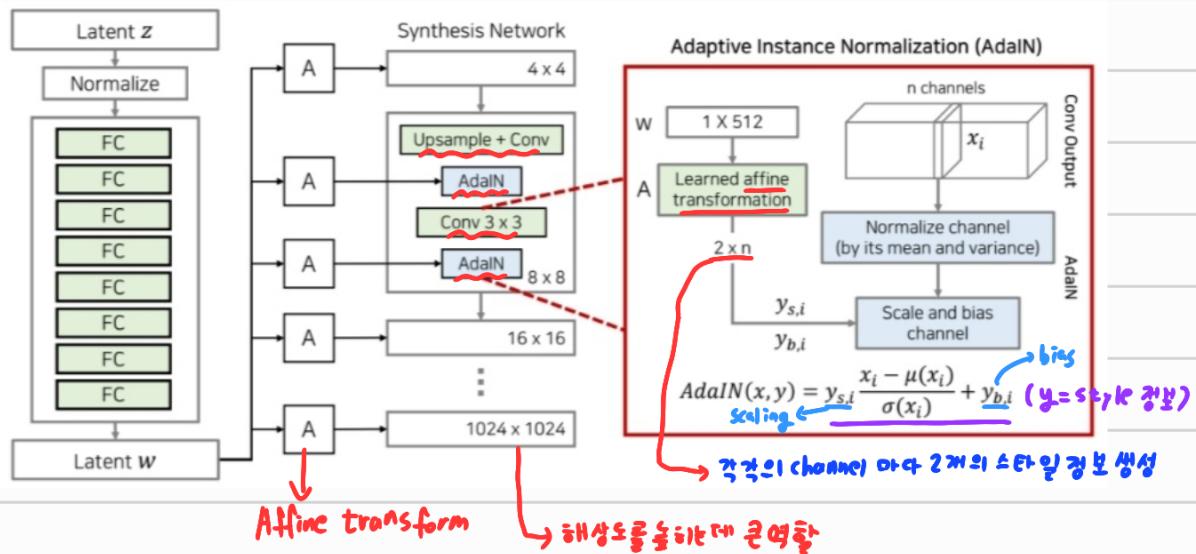
linear함. 특징분포를 따르지 않아도됨
→ 특징들이 분리될 수 있음
→ 벡터들이 linear함.
성질을 가진다. $f: z \rightarrow w$

Related Work : AdaIN

- 다른 원하는 데이터로 부터 style 정보를 가져와 적용 가능
- 학습 시킬 파라미터가 필요하지 않다 (가중치 사용 안함)
- feed-forward 방식의 style transfer에서 좋은 성능



- Style Module



Style GAN Removing Traditional Input

⇒ 다양한 style GAN의 경우 layer를 거치는 과정에서 적용될 수 있거나 함

∴ style 정보에 의해서 이미지의 다양성 보장

⇒ 초기 입력을 상수로 대체하여 진행

+ 주로 깨끗한 머리카락 배경 같은 환경에 의해 변화되는 것들을 처리하기 위해 Noise input을 입력으로 넣음

→ 각각의 layer마다 noise 정보 넣음

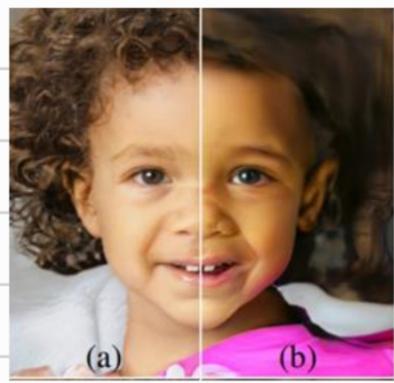
→ Noise 값도 별도의 Affine 변환을 거침

∴ style = high-level global attributes (ex. 얼굴형, 푸즈, 안경유무)

noise = stochastic variational (ex. 주근깨, 피부모공)

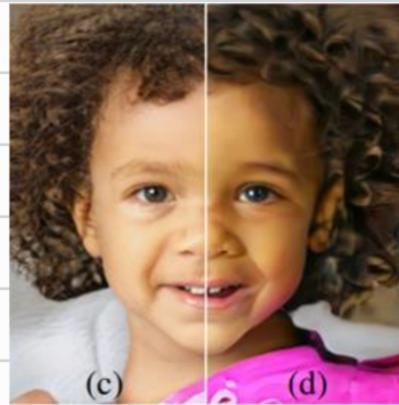
[coarse noise : 큰 크기의 머리 곱슬 거림, 배경]

[fine noise : 세밀한 크기의 머리 곱슬 거림]



a) 모든 레이어에 노이즈 적용

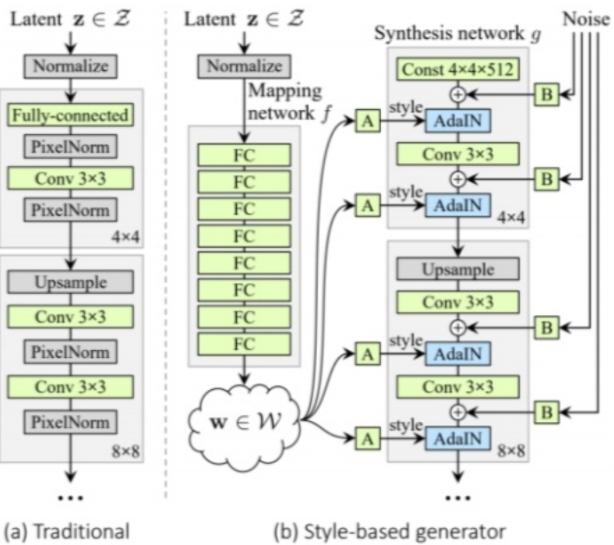
b) 노이즈 적용X



c) Fine layer 적용

d) Coarse hyper에 적용

StyleGAN 요약



a) Traditional : PGGAN

b) style GAN

∴ StyleGAN의 생성자는 더욱 linear 하며 잘 entangled 된다.

Style 정보가 입력되는 layer의 위치에 따라서 style이 미치는 영향력의 규모가 달라질 수 있다.



Coarse style: 전반적인 style 정보
ex) 얼굴 구조, 안경 유무

Middle style: 조금 더 세밀한 style
ex) 헤어스타일, 눈빛과 같은 것

Fine style: 동작이나 미세한 구조
ex) 색상과 미세한 구조

↑ Coarse style 적용할 때 맨위 4개는 Image B, 맨아래 4개는 이미지 A

FID : 전통적인 GAN 성능 평가 방식

Method	dataset	
	CelebA-HQ	FFHQ
A Baseline Progressive GAN [30]	7.79	8.04
B + Tuning (incl. bilinear up/down)	6.11	5.25
C + Add mapping and styles	5.34	4.85
D + Remove traditional input	5.07	4.88
E + Add noise inputs	5.06	4.42
F + Mixing regularization	5.17	4.40

⇒ 테크닉을 추가하면서
성능 향상

[Table] Frechet Inception Distance (FID) for various generator designs.

A) PG GAN B) Bilinear / down sampling (interpolation 테크닉 적용)

C) Mapping Network + AdaIn D) 초기 인코딩이 학습된 $4 \times 4 \times 512$ 상수텐서 이용

E) Noise & F) Mixing Regularization (style mixing technique)

- Mixing Regularization

: 인접한 layer 간의 상관관계 줄여기 위한 방법

↳ 다양한 스타일이 서로 잘 혼리 될 수 있도록 함

: Detail 1) 학습 과정에서 2개 입력 vector

2) crossover point 설정 (cross over한 특정 포인트를 기점으로하여 위쪽은 w_1 vector
아래에는 w_2 vector 사용)

: 스타일은 각 레이어에 대해서 지역화 함

- 성능 지표 제작 (Disentanglement 고려)

① 두 벡터를 interpolation 할 때 얼마나 급격하게 이미지 특징이 바뀌는지 → Path length

② latent space에서 attributes가 얼마나 선형적으로 분류될 수 있는지 평가 → Separability

⇒ W space 이 Z space 보다 이상적인 성질 갖고 있음

Path length

: 지점 t 와 $t+\epsilon$ 사이에서의 VGAE의 거리가 얼마나 먼지

$$\text{Loss}_W = E \left[\frac{1}{\epsilon^2} d \left(G \left(\text{lerp} \left(f(z_1), f(z_2); t \right) \right), G \left(\text{lerp} \left(f(z_1), f(z_2); t + \epsilon \right) \right) \right) \right]$$

↓ ↓ ↓ ↓ ↓
 w_1 w_2 w_1 w_2 ↗ 아주 작은 상수
 ↓ ↓
 이미지 생성 interpolation
 " " linear interpolation

Linear Separability

- CelebA-HQ: 얼굴마다 성별(gender) 등의 40개의 **binary attributes**가 명시되어 있는 데이터셋
 - 이를 이용해 40개의 분류(classification) 모델을 학습합니다.
- 하나의 속성(attribute)마다 200,000개의 이미지를 생성하여 분류 모델에 넣습니다.
 - 이후에 **confidence**가 낮은 절반은 제거하여 100,000개의 레이블이 명시된 **latent vector**를 준비합니다.
 - 이렇게 준비된 100,000개의 데이터를 학습 데이터로 사용합니다.
- 매 attribute마다 **linear SVM** 모델을 학습합니다.
 - 이때 전통적인(traditional) GAN에서는 z , Style GAN에서는 w 를 이용합니다.
- 각 linear SVM 모델을 활용하여 다음의 값을 계산합니다. ($i =$ 각 attribute의 인덱스)

$$\exp\left(\sum_i H(Y_i|X_i)\right) \xleftarrow{\text{entropy}}$$