

## U-net D 개요.

### \* Introduction

GAN은 많은 성장을 이루었지만, 전처리적인 의미론적 통일성이, 장거리 구조와 세부 구조에 대해 부족하다.

문제 중 하나는 D가 G가 실제 이미지를 더 잘 생성할 수 있도록 loss function처럼 행동하도록,

data distribution을 modeling 한다.

현재 D는 단순히 G를 구별하기 위한 representation만 확보한다.

따라서 전처리 구조의 detail을 모두 확보하지는 못하고, 이로 인해aliasing이 생기거나, 전처리 구조의 문제가 생긴다.

↳ D의 입장에서 짧은 구조의 의미론은 다른 task로 간주될 수 있다.

↳ 단순히 실제와 가짜를 구분하기 위해 학습하기 때문이다.

따라서 real과 fake의 global과 local을 모두 보는 구조를 제시한다.

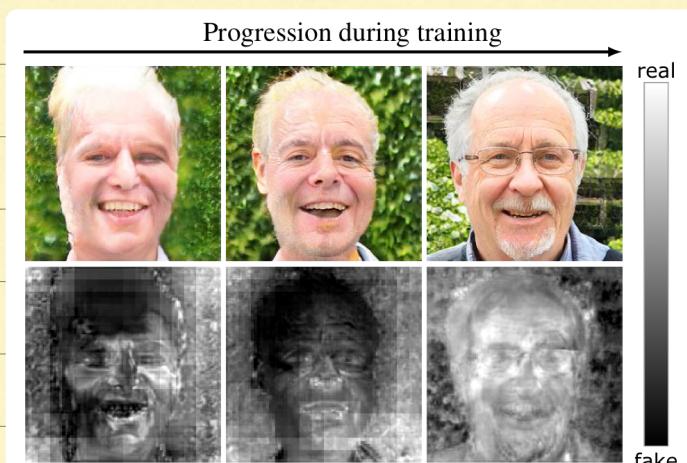


Figure 1: Images produced throughout the training by our U-Net GAN model (top row) and their corresponding per-pixel feedback of the *U-Net discriminator* (bottom row). The synthetic image samples are obtained from a fixed noise vector at different training iterations. Brighter colors correspond to the discriminator confidence of pixel being real (and darker of being fake). Note that the U-Net discriminator provides very detailed and spatially coherent response to the generator, enabling it to further improve the image quality, e.g. the unnaturally large man's forehead is recognized as fake by the discriminator and is corrected by the generator throughout the training.

Encoder에서는 기존 역할을 하고 decoder는 fake/real에 대한 image를 나눠낸다.

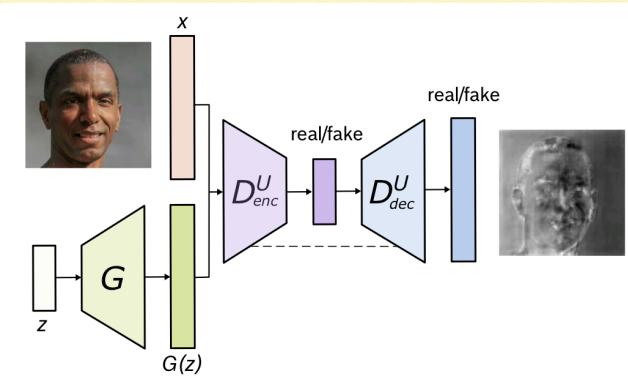


Figure 2: U-Net GAN. The proposed U-Net discriminator classifies the input images on a global and local *per-pixel* level. Due to the skip-connections between the encoder and the decoder (dashed line), the channels in the output layer contain both high- and low-level information. Brighter colors in the decoder output correspond to the discriminator confidence of pixel being real (and darker of being fake).

CutMix Augmentation을 사용했다.

	Real	Fake
Original images		
Real/fake ratio r	0.28	0.68
Mask M		
CutMix images		
$D_{dec}^U$ segm. map		
$D_{enc}^U$ class. score	0.31	0.60
	0.36	0.43

Figure 3: Visualization of the CutMix augmentation and the predictions of the U-Net discriminator on CutMix images. 1st row: real and fake samples. 2nd&3rd rows: sampled real/fake CutMix ratio  $r$  and corresponding binary masks M (color code: white for real, black for fake). 4th row: generated CutMix images from real and fake samples. 5th&6th row: the corresponding real/fake segmentation maps of  $D^U$  with its predicted classification scores.

→ CutMix이 encoder 결과는 fake?

decoder는 fake나 real

→ consistency regularization을 위해

CutMix로 pixel들이 일관되게 예측되지 않도록 한다.

→ 이를 통해 D가 유타를 더 구조에 짐작할 수 있도록 하고

설정이 더 유타적이다.

그리고 decoder의 localization이 도움을 준다.

BigGAN의 다른 평가는 FFHQ, CelebA, COCO-animal 등 BigGAN의 FID, IS 평가.

## \* Related work

- GAN

- Mix & Cut Regularization.

Mixup, Manifold mixup, Cutout 등이 있다.

논문에서는 CutMix transformation과 consistency regularization을 제시한다.

↳ localization 품질 증가, real과 fake 사이의 구별 안되는 차이에 집중하도록 유도합니다.

## \* UNet GAN Model

> 기본 vanilla GAN은 G와 D가 별개이며 학습이 됩니다.

논문에서는 UNet에서 G에서 D를 제안하는데 기본 D는 전드레인 알고 encoder로 써둔다.

↳ global, local을 모두 볼 수 있어 G가 더 많은 정보를 쓴다.

Decoder에서 pixel 별로 feedback을 줄 수 있다.

↳ 또한 CutMix로 D의 일관되지 않은 예측의 penalty를 주는 새로운 consistency regularization을 제안합니다.

↳ localization 품질 증가, real과 fake 사이의 구별 안되는 차이에 집중하도록 유도합니다.

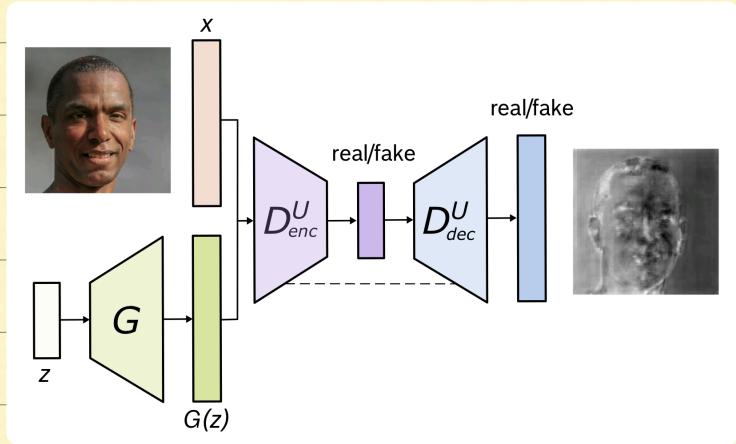
## \* UNet-based discriminator.

> 기본 UNet처럼 encoder는 global을 봐고, decoder는 input과 matching 시켜 localization을 결정하도록 한다.

Skip connection은 fine grained 한 detail을 가능하게 한다.

이런 D를  $D^4$ 라고 표기한다.

↳  $D^4$ 는  $D^0$ 의 pixel 별로 real, fake를 판별하는 loss가 추가된다.



$$\mathcal{L}_{D^u} = \mathcal{L}_{D_{enc}^u} + \mathcal{L}_{D_{dec}^u}$$

$$\hookrightarrow \mathcal{L}_{D_{enc}^u} = -\mathbb{E}_x [\log D_{enc}^u(x)] - \mathbb{E}_z [\log (1 - D_{enc}^u(G(x)))]$$

$$\hookrightarrow \mathcal{L}_{D_{dec}^u} = -\mathbb{E}_x \left[ \sum_{i,j} \log [D_{dec}^u(x)]_{i,j} \right] - \mathbb{E}_z \left[ \sum_{i,j} \log (1 - [D_{dec}^u(G(x))]_{i,j}) \right] \rightarrow \text{전체 pixel} \text{에 걸쳐서}$$

Generator의 대한 loss

$$\mathcal{L}_G = -\mathbb{E}_z [\log D_{dec}^u(G(z)) + \sum_{i,j} \log [D_{dec}^u(G(x))]_{i,j}]$$

→ Consistency regularization.

같은 훈련된  $D^u$ 의 핵심 당 fake/real 결정은 어느 class이며 동일해야 한다.

하지만 보장되지는 않는다.

이를 위해  $D^u$ 는 class별로 각각의 성능이 아니라 전반적 성능을 살피면 의미를 적어도 광범위한

차이를 보여야 한다.

$D_{dec}^u \rightarrow$  CutMix를 예측하기 힘으로써, 일정한 확률로 consistency regularization을 제안한다.

↳ 다른 class 뒤에 차운다.

↳  $D_{dec}^u$ 가 잘 판별하는 경우, 전자는 진짜를 예측하게 한다.

	Real	Fake	
Original images			
Real/fake ratio $r$	0.28	0.68	0.31
Mask M			
CutMix images			
$D_{dec}^U$ segm. map			
$D_{enc}^U$ class. score	0.31	0.60	0.36
			0.43

사용한 training sample 52

$$\tilde{x} = \text{mix}(x, G(x), M)$$

$$= M \odot x + (1 - M) \odot G(z)$$

↳ decoder는 pixel 베로 ICML fake/reals를 학습했습니다.

encoder는 fake로 학습.

↳ Generator가 부족하거나, artifact가 생길 수 있음.

Mask M은 사용한 decoder의 대상 GCL입니다.

이제 다음은  $\mathcal{L}$ 은

$$\mathcal{L}_{D_{dec}}^{\text{cons}} = \| D_{dec}^u \left( \text{mix}(x, G(z), M) \right) - \text{mix}(D_{dec}^u(x), D_{dec}^u(G(x)), M) \|^2$$

전체적인  $\mathcal{L}$ 은

$$\mathcal{L}_D = \mathcal{L}_{D_{enc}} + \mathcal{L}_{D_{dec}} + \lambda \mathcal{L}_{D_{dec}}^{\text{cons}}$$

또한 디蹲 WGAN을 같이 사용하는 경우가 있다.

## \* Implementation.

실装의 경우 BigGAN이 했다.

각 BigGAN의 Decoder의 경우 ch=64이  $16 \times 16 \times 4$  까지 down sampling을 시킨 후 pooling을 진행했다.

UNet의 경우 G의 구조를 볼 수 있다.

각 dec. enc layer 사이에 concat은, dec의 출력이 ch=16이 되었을 때 1x1 conv로 1channel이 되게 한다.

또한 BigGAN의 다른 차원 모든 layer BN이 동일한 latent z를 갖도록 했다.

↳ BigGAN은 사용적인 구조를 쓰는 듯?

self-attention 사용.

CutMix sample의 minibatch의 것은 확률  $P_{mix}$ 는 초기값이 0.45였지만 0.5로 바뀌었다.

는 A 네트가 GAN 이미지를 img를 생성하는데 시간을 짧아짐에 대해서.

또한 CutMix image는 minibatch 내 sample은 만들어진다.

Mask M의 경우 ratio  $r$ 의 대신 image를 random crop하여 만들어진다.  
↑  
넓이비율

그리고  $\lambda=1$ 로 선정함.

\* Experiments

Method	FFHQ				COCO-Animals			
	Best		Median		Best		Median	
	FID↓	IS↑	FID↓	IS↑	FID↓	IS↑	FID↓	IS↑
BigGAN [5]	11.48	3.97	12.42	4.02	16.37	11.77	16.55	11.78
U-Net GAN	<b>7.48</b>	<b>4.46</b>	<b>7.63</b>	<b>4.47</b>	<b>13.73</b>	<b>12.29</b>	<b>13.87</b>	<b>12.31</b>

Table 1: Evaluation results on FFHQ and COCO-Animals. We report the best and median FID score across 5 runs and its corresponding IS, see Section 4.2 for discussion.

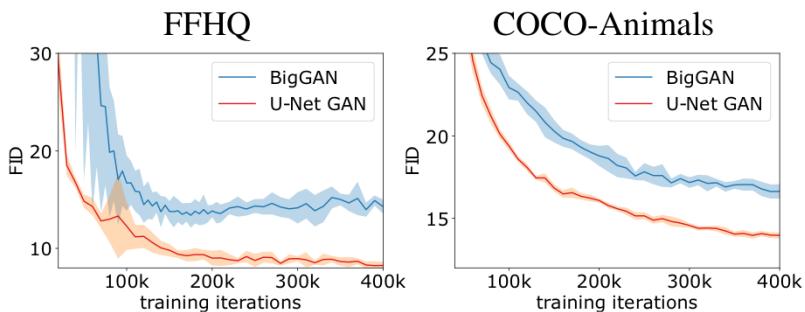


Figure 6: FID curves over iterations of the BigGAN model (blue) and the proposed U-Net GAN (red). Depicted are the FID mean and standard deviation across 5 runs per setting.

⇒ Barelike GAN ဆုံး၏။



Figure 5: Images generated with U-Net GAN trained on COCO-Animals with resolution  $128 \times 128$ .

⇒ Unconditional condition မရှိစေ ခဲ့သူ။

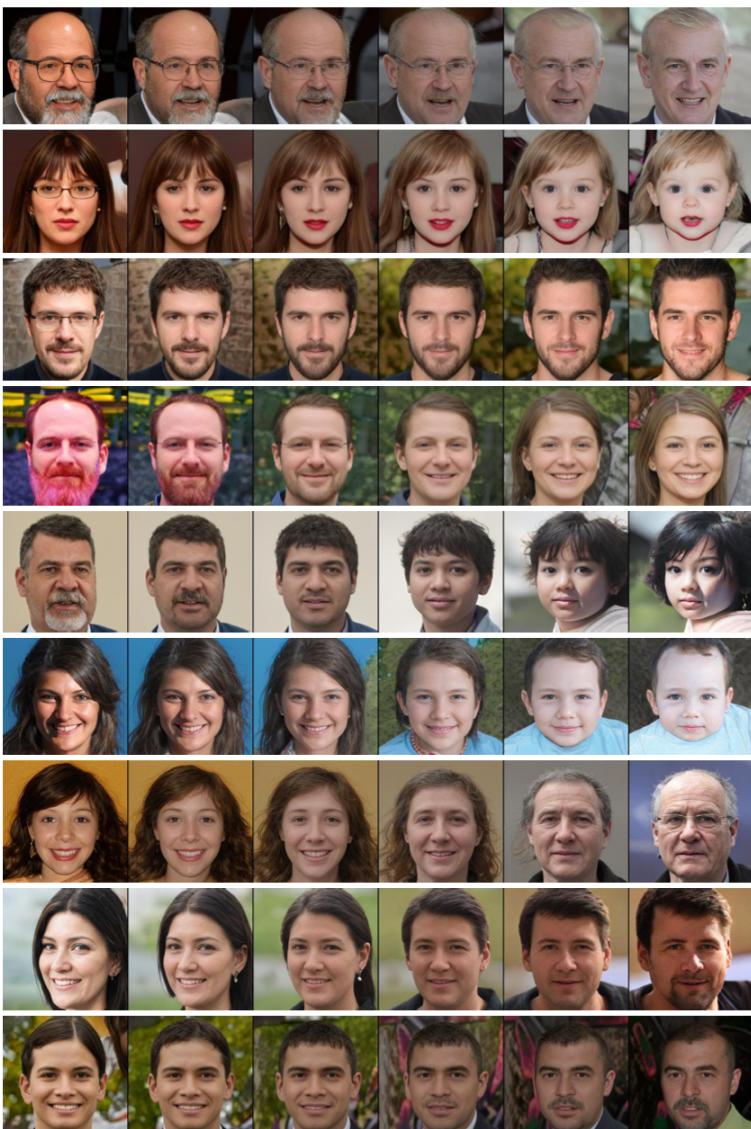


Figure 4: Images generated with U-Net GAN trained on FFHQ with resolution  $256 \times 256$  when interpolating in the latent space between two synthetic samples (left to right). Note the high quality of synthetic samples and very smooth interpolations, maintaining *global* and *local* realism.

⇒ 실제 공간 거리의 interpolation에도 훌륭한 품질.

= 또한 train sample의 뿐 아니라 많았으나, D의 globality에 의해 훈련에서도 성립

↗ Ablation study

Method	COCO-Animals	FFHQ
BigGAN [5]	16.55	12.42
U-Net based discriminator	15.86	10.86
+ CutMix augmentation	14.95	10.30
+ Consistency regularization	<b>13.87</b>	<b>7.63</b>

Table 2: Ablation study of the U-Net GAN model on FFHQ and COCO-Animals. Shown are the median FID scores. The proposed components lead to better performance, on average improving the median FID by 3.7 points over BigGAN [5]. See Section 4.2 for discussion.

↗ Compare with SOTA

Method	FID ↓	IS ↑
PG-GAN [19]	7.30	–
COCO-GAN [27]	5.74	–
BigGAN [5]	4.54	3.23
U-Net GAN	<b>2.95</b>	<b>3.43</b>

Table 3: Comparison with the state-of-the-art models on CelebA ( $128 \times 128$ ). See Section 4.2 for discussion.

\* Discriminator response visualize

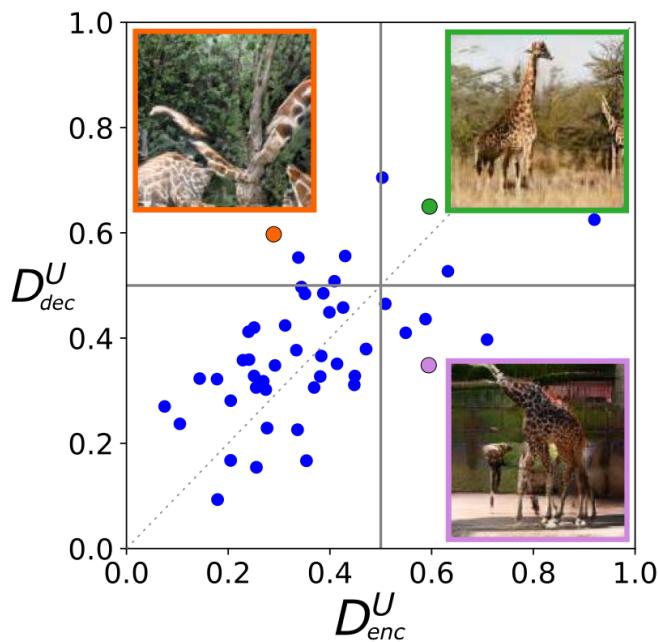


Figure 7: Visualization of the predictions of the encoder  $D_{enc}^U$  and decoder  $D_{dec}^U$  modules during training, within a batch of 50 generated samples. For visualization purposes, the  $D_{dec}^U$  score is averaged over all pixels in the output. Note that quite often decisions of  $D_{enc}^U$  and  $D_{dec}^U$  are not coherent with each other. As judged by the U-Net discriminator, samples in the upper left consist of locally plausible patterns, while not being globally coherent (example in orange), whereas samples in the lower right look globally coherent but have local inconsistencies (example in purple: giraffe with too many legs and vague background).

⇒ Dec Enc는 종종 다른 real/fake 결과를 내놓는다

↳ 이로 성별적이고 디자인한 feed back

\* Characterizing the training Dynamics

실험의 collapse와 관련된 일련의 예를 살펴보겠습니다.

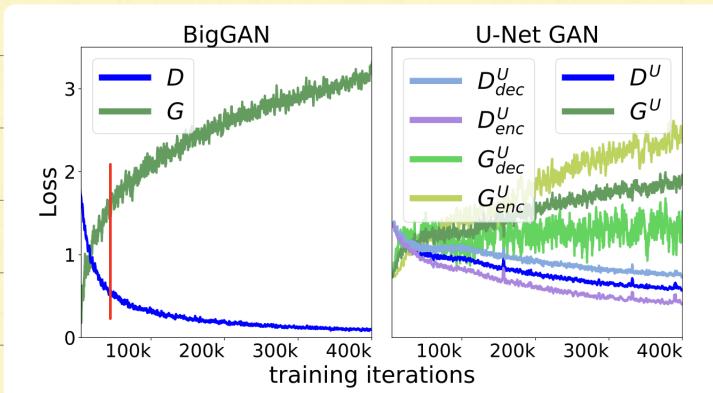


Figure 8: Comparison of the generator and discriminator loss behavior over training for U-Net GAN and BigGAN. The generator and discriminator loss of U-Net GAN is additionally split up into its encoder- and decoder components.

=> UNet의 경우 D의 손실이 증가하기 때문에 D의 손실은 더 긴 흐름을 할 수 있고

BigGAN의 경우 더욱 급방으로 D의 손실을 줄여나간다.