

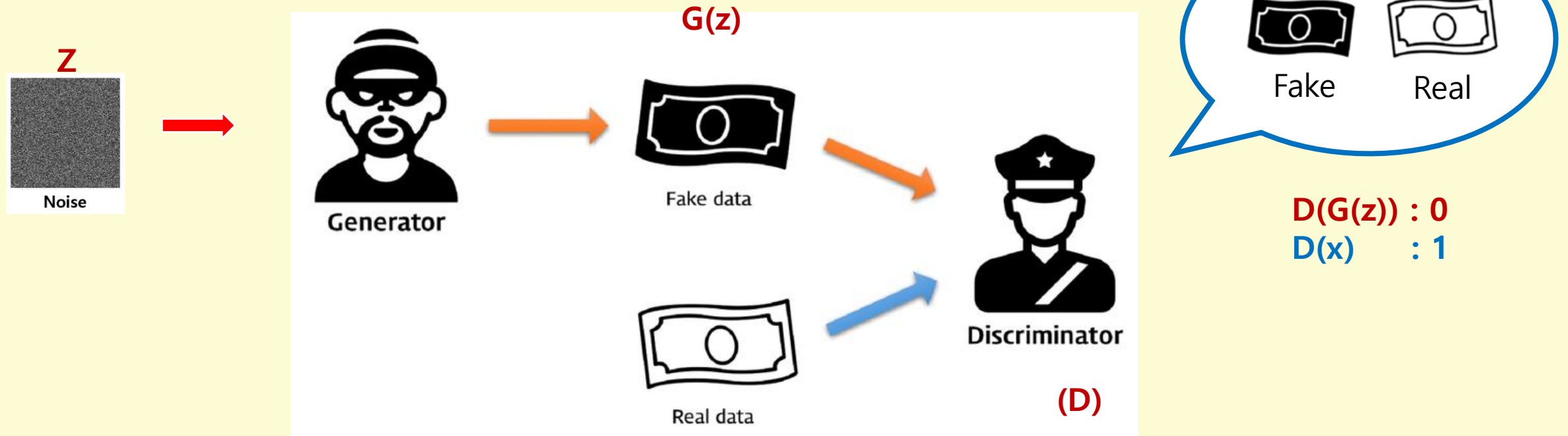
# *Analyzing and Improving the Image Quality of StyleGAN*

Team 연어유희

# 목차

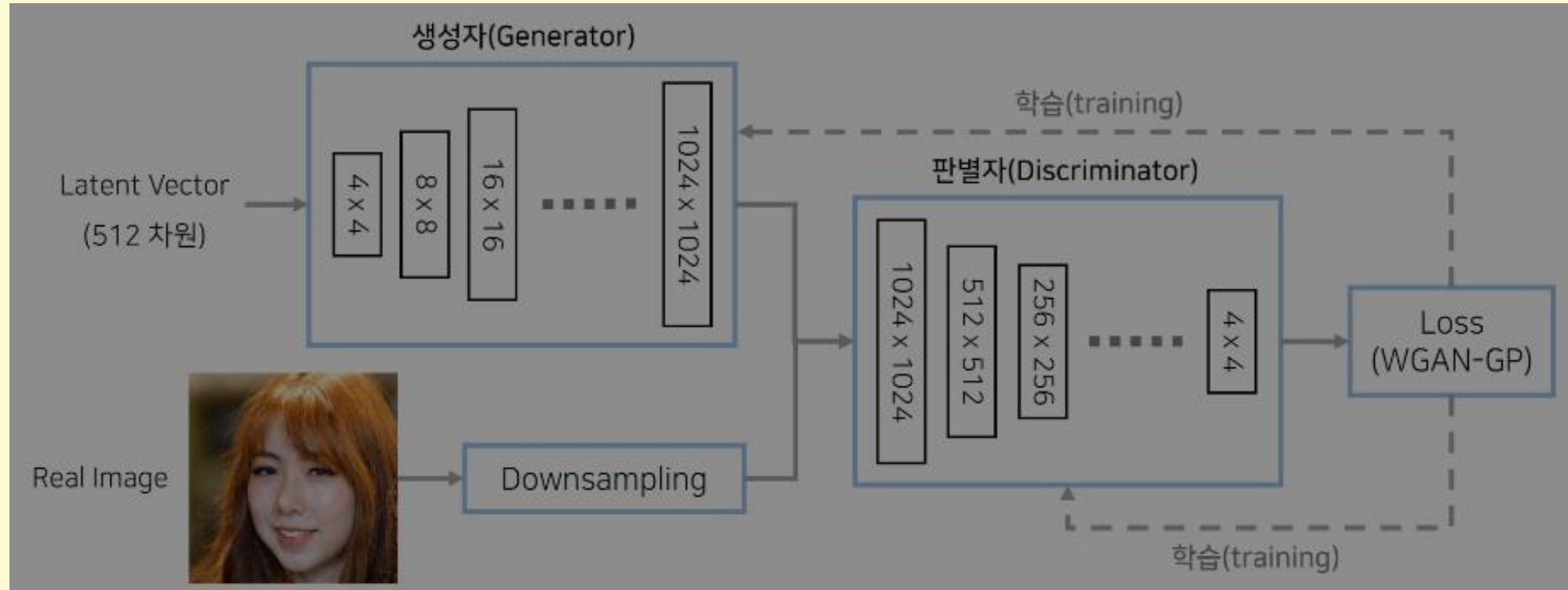
- ① GAN
- ② PGGAN
- ③ StyleGAN
- ④ StyleGAN2
- ⑤ Code & Output
- ⑥ Q & A

# GAN (Generative Adversarial Network)



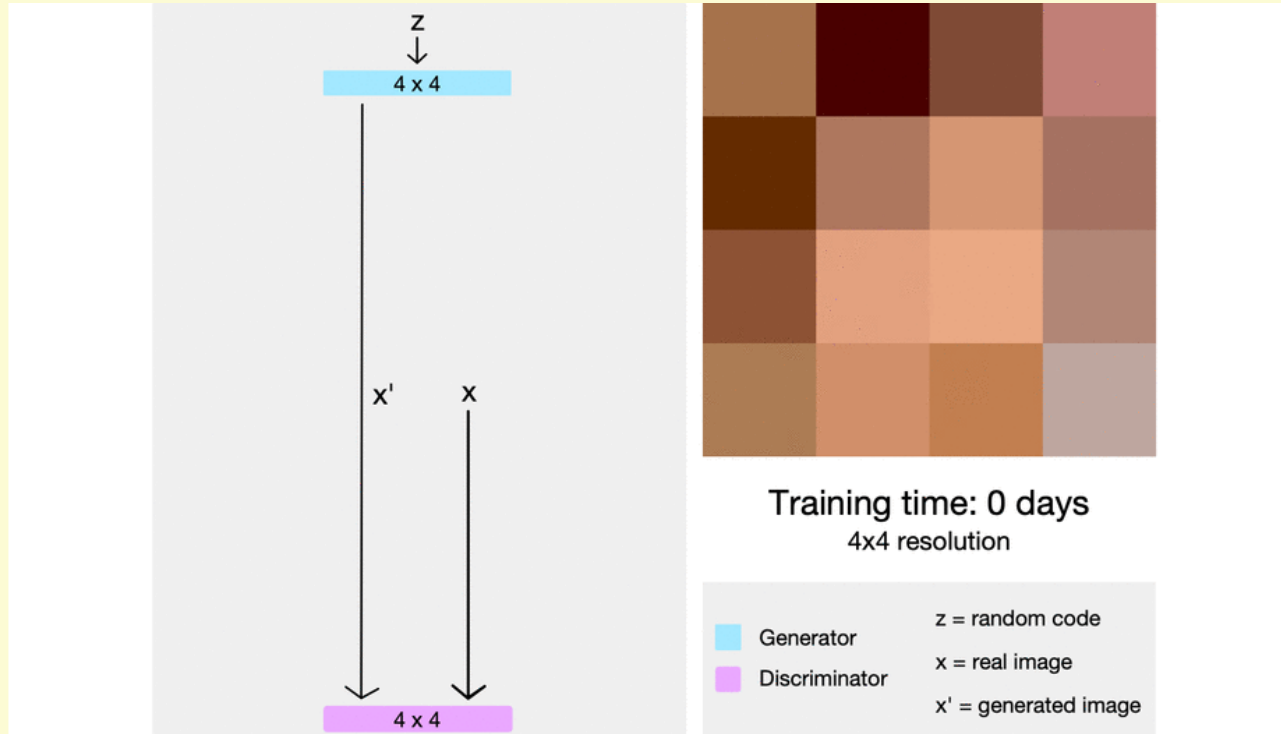
- X**
- ① 생성자(Generator), 판별자(Discriminator) 네트워크를 활용한 생성 모델
  - ② 생성자는 판별자를 활용하여 이미지 분포를 학습하고 이미지를 생성함

# PGGAN (Progressive Growing of GANs)



- ① 학습을 진행하는 과정에서 점진적(progressively)으로 네트워크의 레이어를 붙여 나감
- ② 판별자도 생성자와 마찬가지로 대칭적으로 여러 개의 layer를 붙여 나감
- ③ 한 번에 전체 크기의 이미지 feature들을 학습시키지 않음 ( $4 \times 4$ 의 저해상도로 start)

# PGGAN (Progressive Growing of GANs)



## Main Idea

- 학습 과정에서 레이어를 추가
- 고해상도 이미지 학습 성공 (1024 x 1024)

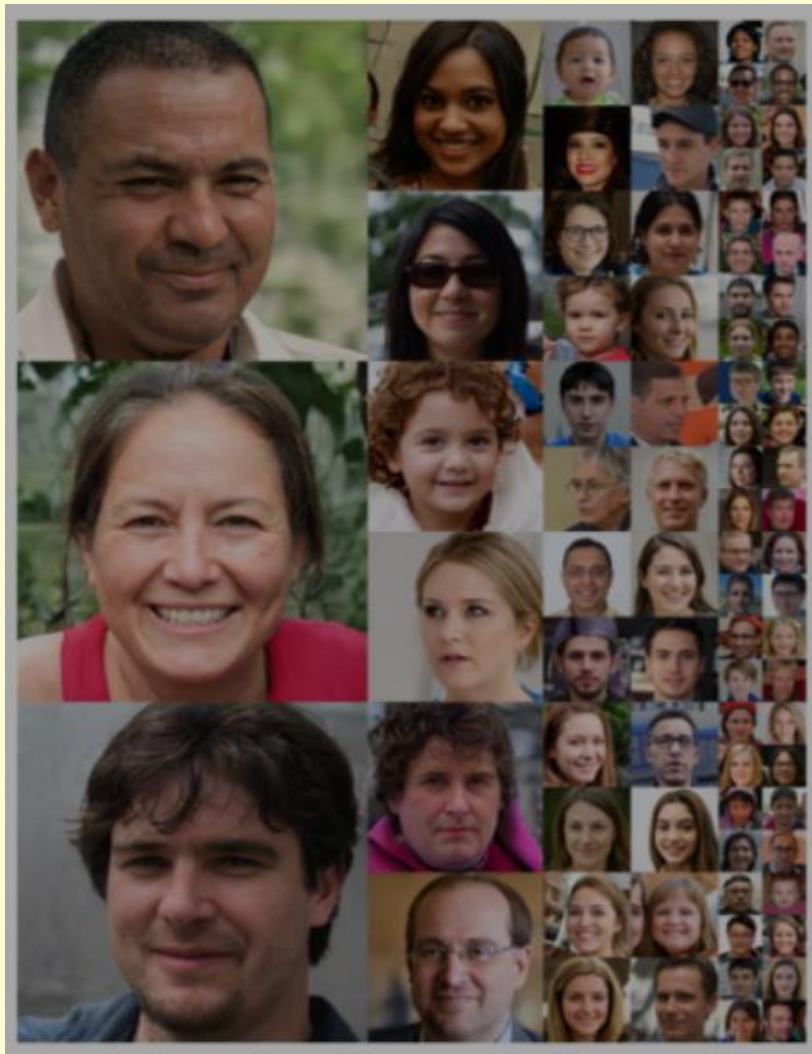
## Limitation

- 이미지의 특징 제어 어려움



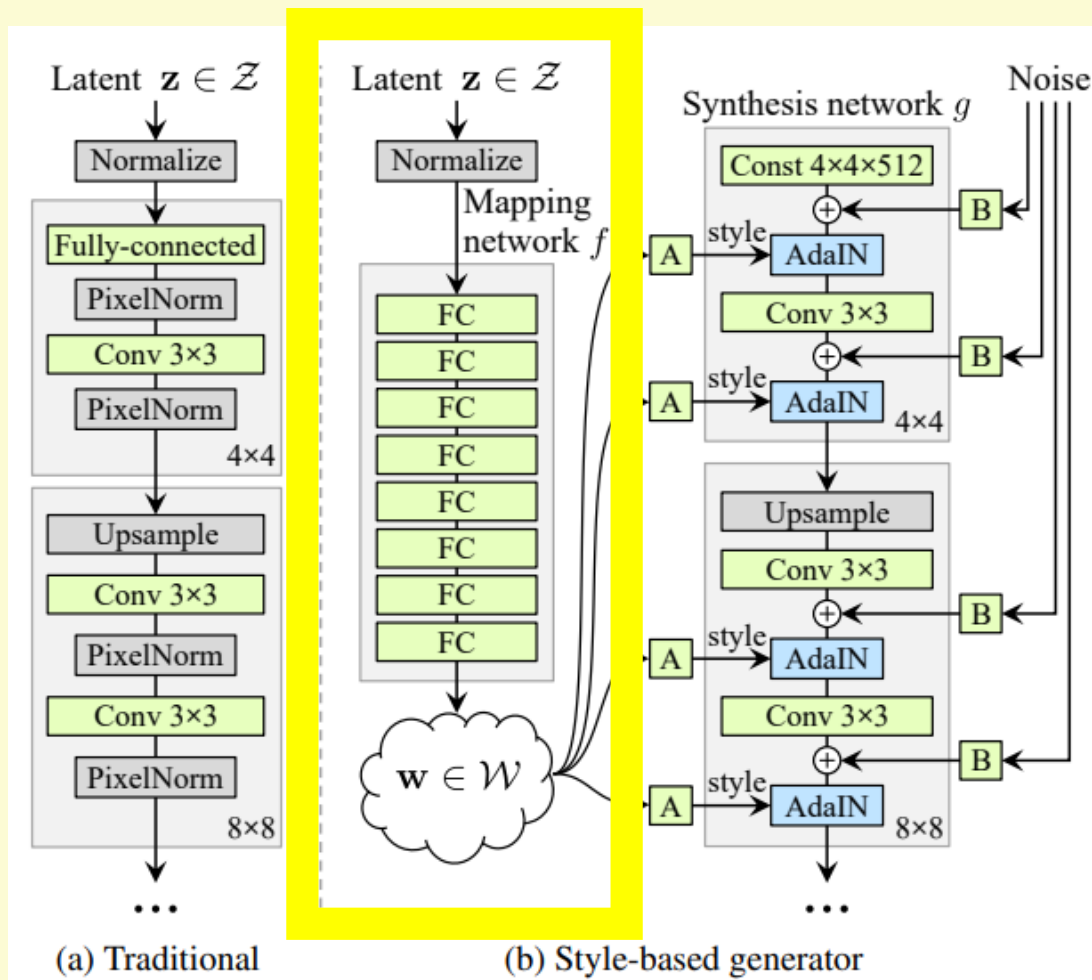
StyleGAN 에서 개선

# StyleGAN



- ① PGGAN 베이스라인 아키텍처의 성능 향상
- ② Disentanglement 특성 향상
- ③ 고해상도 얼굴 데이터셋(FFHQ) 발표

# StyleGAN Architecture



PGGAN

## 💡 Mapping Network

- 특정 space에서의 벡터를 다른 space에서의 벡터로 매핑을 시켜주는 네트워크

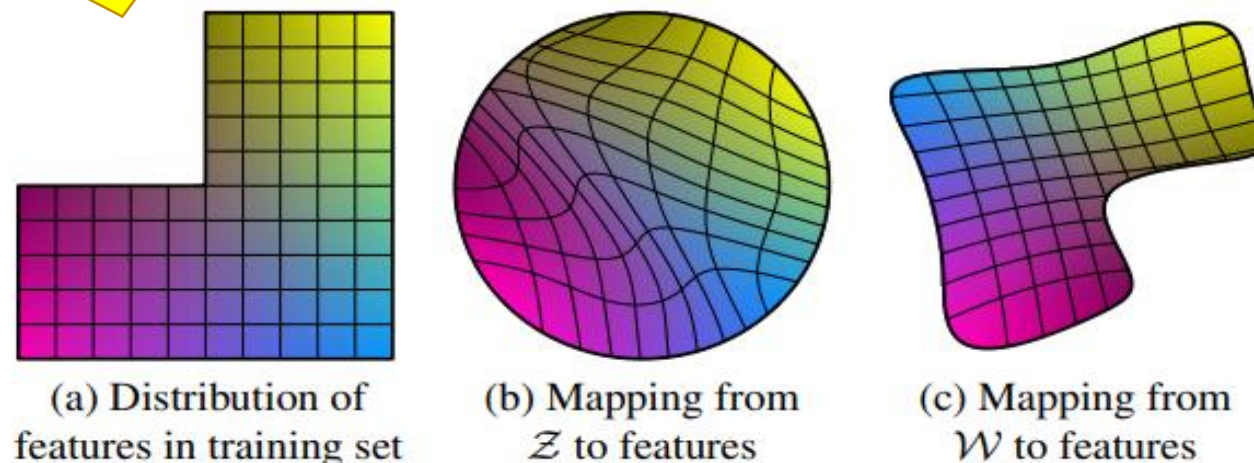
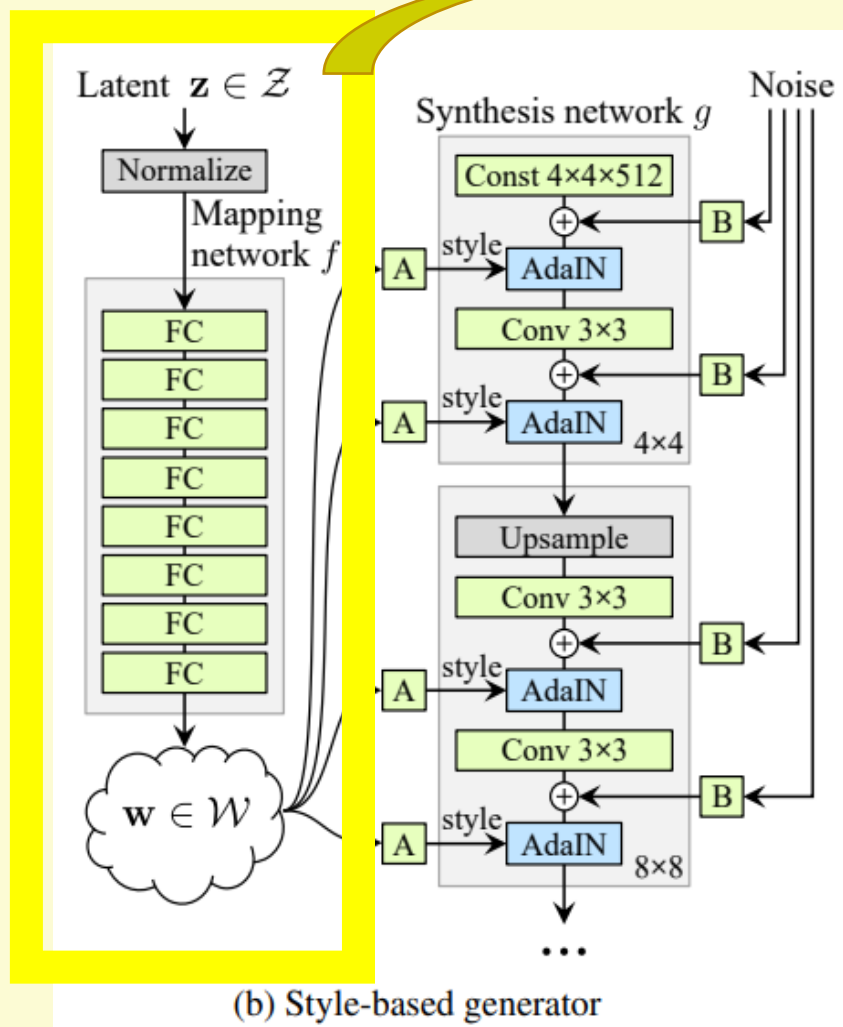
① Z도메인에서 W도메인으로 매핑 수행

② W vector를 각각의 Block에 넣어줌

③ Affine transformation을 해준 후 AdaIN에 넣음

④ Noise를 추가

## StyleGAN Architecture



(a) 실제 학습 데이터가 가지는 분포

ex) 세로축: 남성, 여성

가로축: 머리 길이

(b) 가우시안 분포를 따르는 latent vector를 샘플링

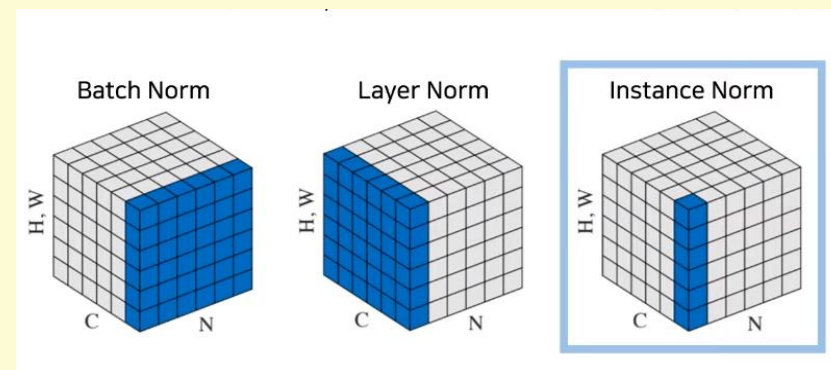
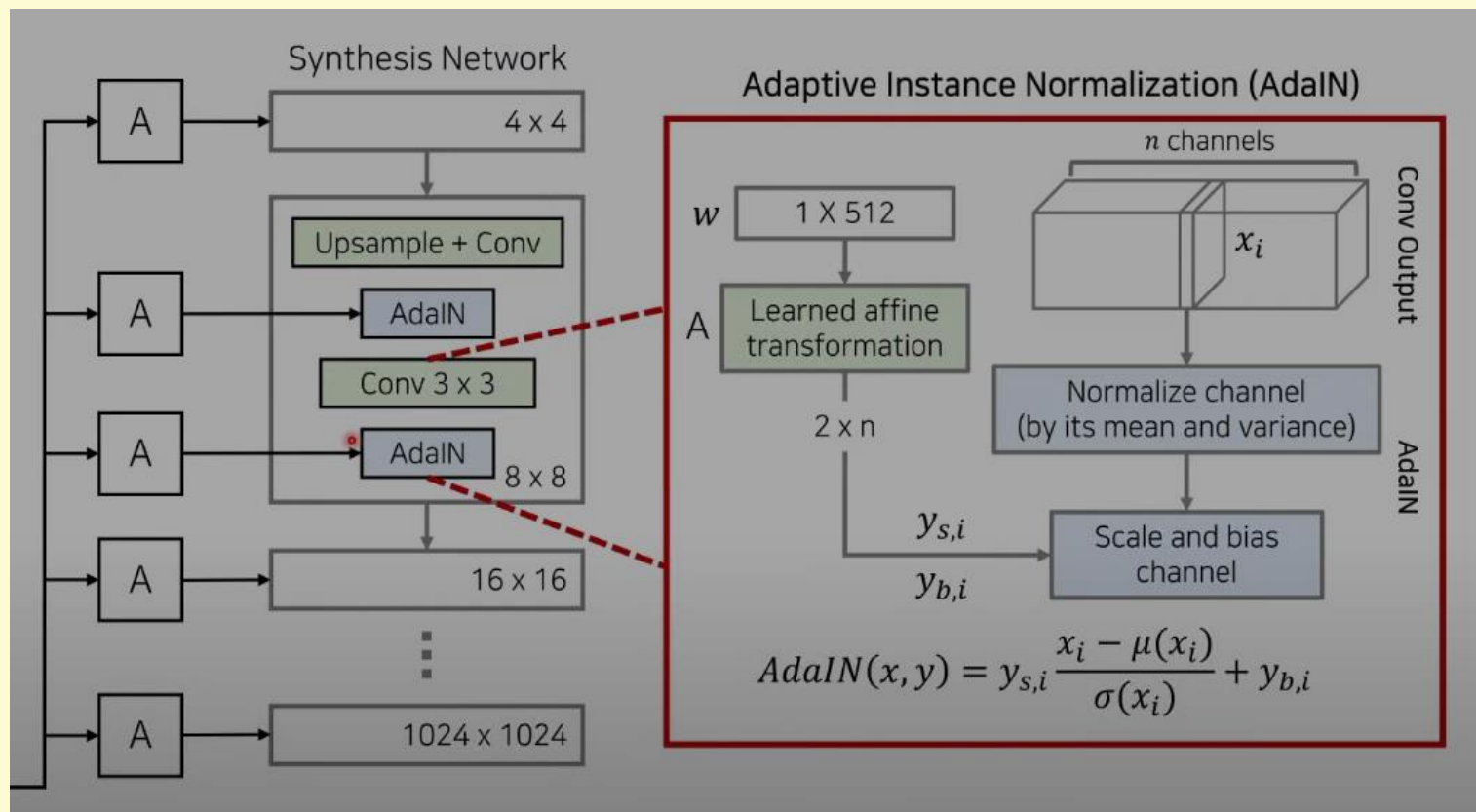
-> entangled됨

(c) Mapping network  $\mathcal{G}_M$  takes latent vector  $w$

-> 각각의 특징에 대해서 interpolate을 수행할 때 linear한 상태가 될 확률 up



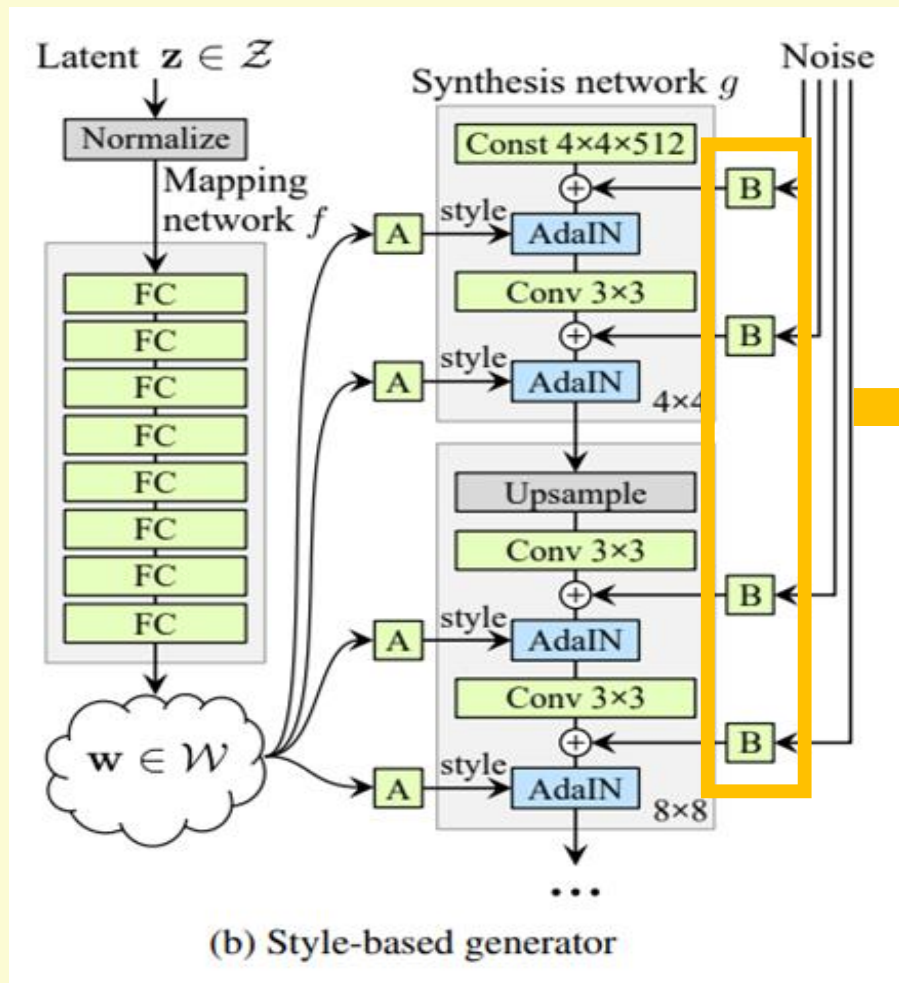
# StyleGAN Architecture – Adaptive Instance Normalization (AdaIN)



$$AdaIN(x_i, y) = y_{s,i} \frac{x_i - \mu(x_i)}{\sigma(x_i)} + y_{b,i}$$

각 채널 단위로 정규화를 수행한 뒤에 별도의 스타일 정보를 입력 받아서 feature상의 statistics을 바꾸는 방식

## StyleGAN Architecture – Stochastic Variation



### Stochastic Variation (확률론적 변이)

- 다양한 **확률적인 측면**을 컨트롤 할 수 있게 해줌  
ex) 주근깨, 머리카락의 배치, 여드름 등

# StyleGAN Architecture – Stochastic Variation

## Stochastic Variation (확률론적 변이)

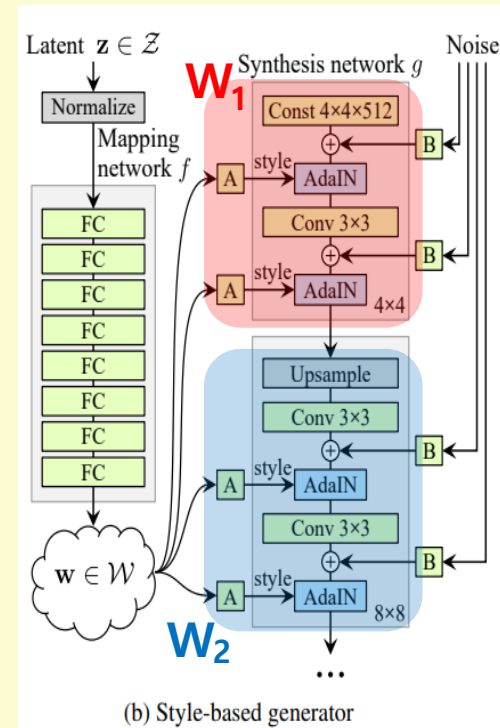
- 다양한 **확률적인 측면**을 컨트롤 할 수 있게 해 줌

ex) 주근깨, 머리카락의 배치, 여드름 등



(a) Generated image (b) Stochastic variation (c) Standard deviation

# StyleGAN Architecture – Mixing Regularization



인접한 layer 간의 style 상관관계를 줄여 나가는 것

- 각각의 스타일이 잘 localize되어서 다른 layer에 관여하지 않도록

crossover point 설정



# StyleGAN 한계점



Figure 1. Instance normalization causes water droplet -like artifacts in StyleGAN images. These are not always obvious in the generated images, but if we look at the activations inside the generator network, the problem is always there, in all feature maps starting from the 64x64 resolution. It is a systemic problem that plagues all StyleGAN images.

## 1) blob-like artifact 발생

training methods that eliminate them. First, we investigate the origin of common **blob-like artifacts**, and find that the generator creates them to circumvent a design flaw in its architecture. In Section 2, we redesign the normalization used



Figure 6. Progressive growing leads to “**phase**” artifacts. In this example the teeth do not follow the pose but stay aligned to the camera, as indicated by the blue line.

## 2) phase artifact (fixed position)

## StyleGAN2 등장



① 이전과 마찬가지로  $Z \rightarrow W$  매핑 네트워크를 사용

② 기존 아키텍처의 문제점 지적

1) **blob-like artifact** 발생

2) **phase artifact** 발생



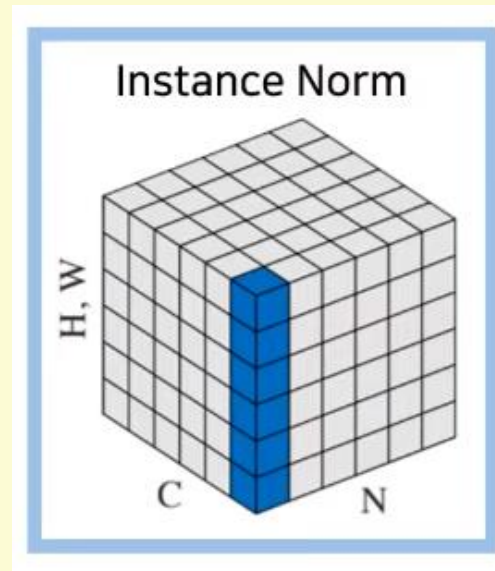
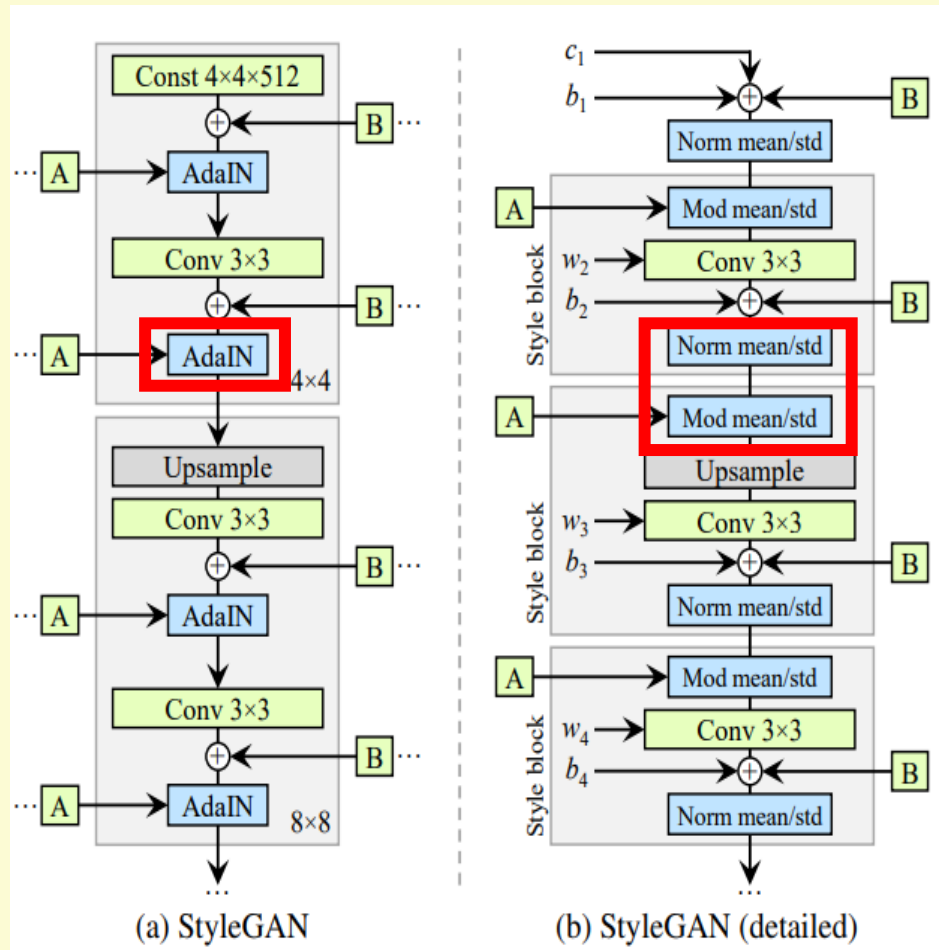
- AdaIN이 **원인**으로 추정
- 대략  **$64 \times 64$**  해상도에서부터 이러한 artifact가 보이기 시작

We begin by observing that most images generated by StyleGAN exhibit characteristic blob-shaped artifacts that resemble water droplets. As shown in Figure 1, even when the droplet may not be obvious in the final image, it is present in the intermediate feature maps of the generator.<sup>1</sup> The anomaly starts to appear around  $64 \times 64$  resolution, is present in all feature maps, and becomes progressively

We pinpoint the problem to the AdaIN operation that normalizes the mean and variance of each feature map separately, thereby potentially destroying any information found in the magnitudes of the features relative to each other. We

# StyleGAN2

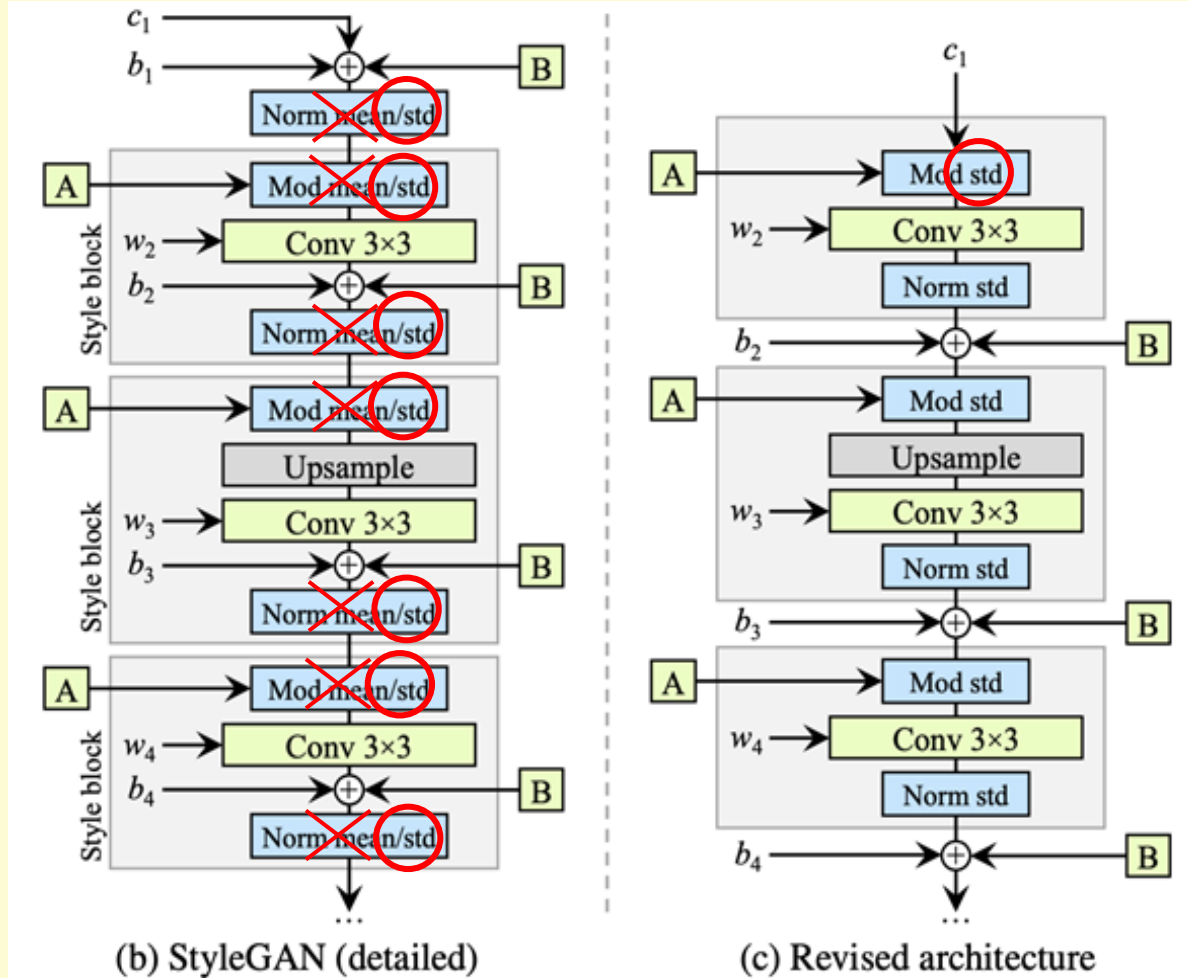
## 1) blob-like artifact 발생



원인) AdaIN

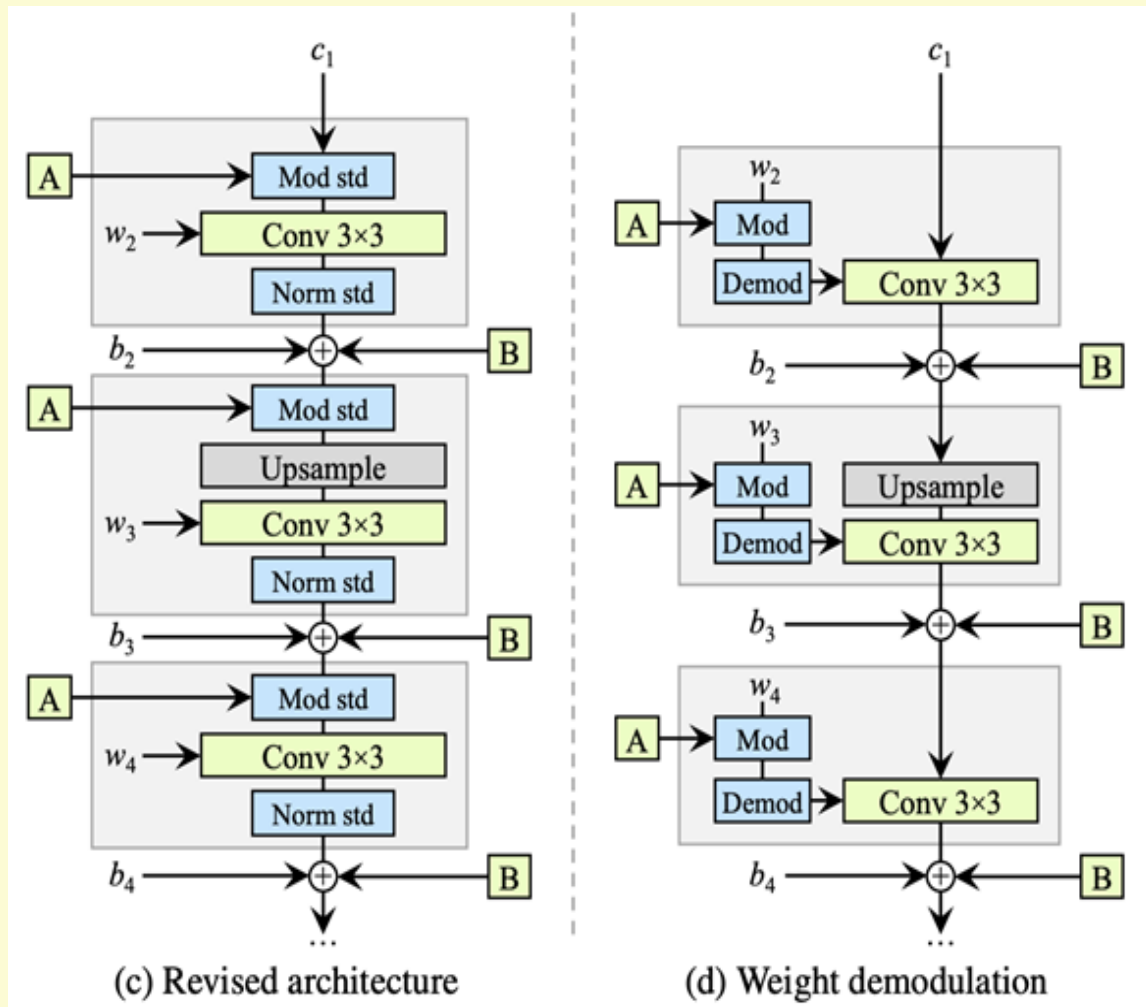
각 feature map의 평균과 분산을 개별적으로 normalize 하여 서로 연관된 feature들의 정보가 소실될 수 있음





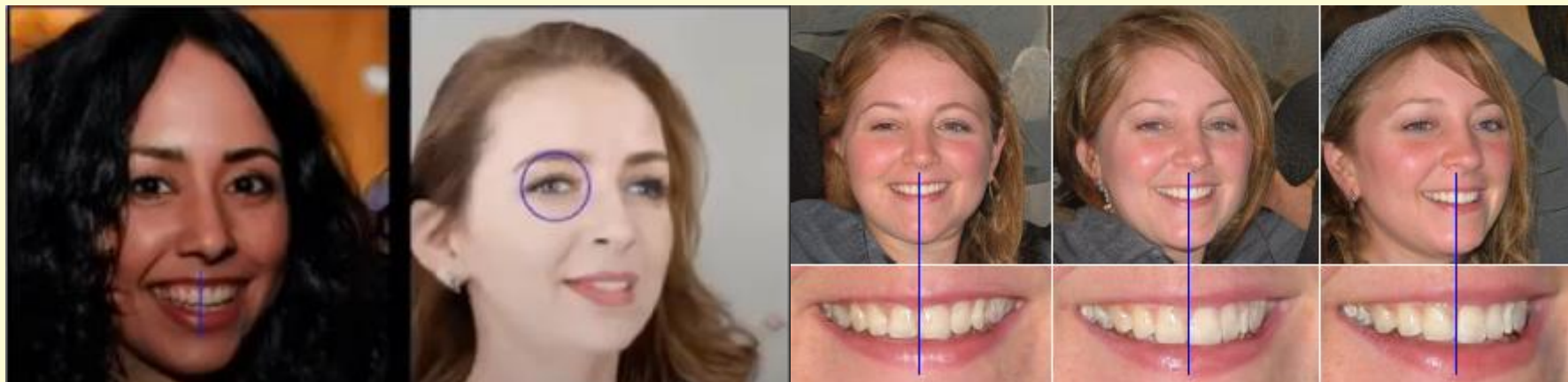
Normalization 단계를 제거

- ① Mean은 건드리지 말고 standard deviation(표준편차)만 변경해도 충분
- ② 최대한 style block 외부에서 feature map들의 값을 변경하고자 함



Normalization 단계를 제거  
Demodulation 대체

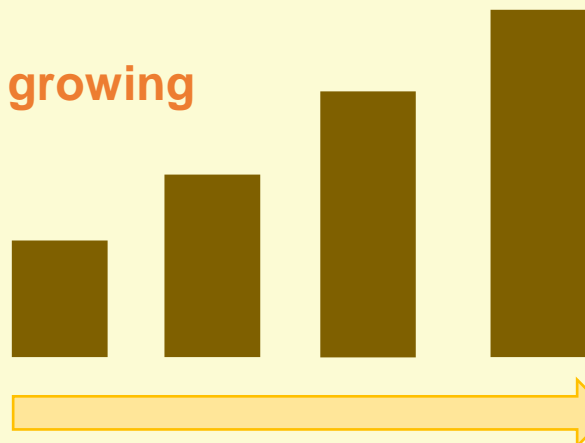
- ① Weight 값에 대하여 modulation 진행
- ② 직접적으로 정규화를 하지 않고 feature map의 statistics를 예측하여 정규화를 적용



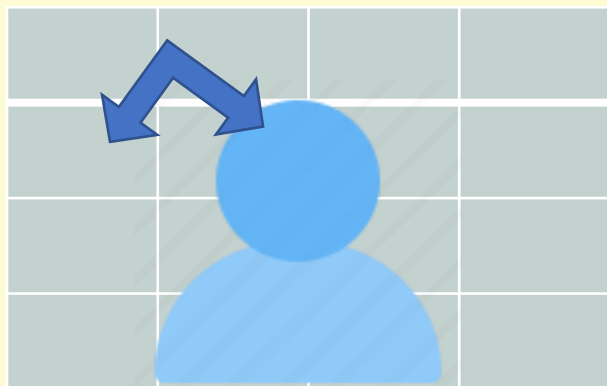
### Phase artifact란?

얼굴의 특정 부분이 **고정된 위치**를 갖는 문제 발생

**원인) Progressive growing**



- ① 각 resolution이 output resolution에 즉시 영향을 미침
- ② 최대한 높은 frequency detail을 유발함



높은 frequency detail 란?

인접한 픽셀 간의 값의 변화가 크다는 것

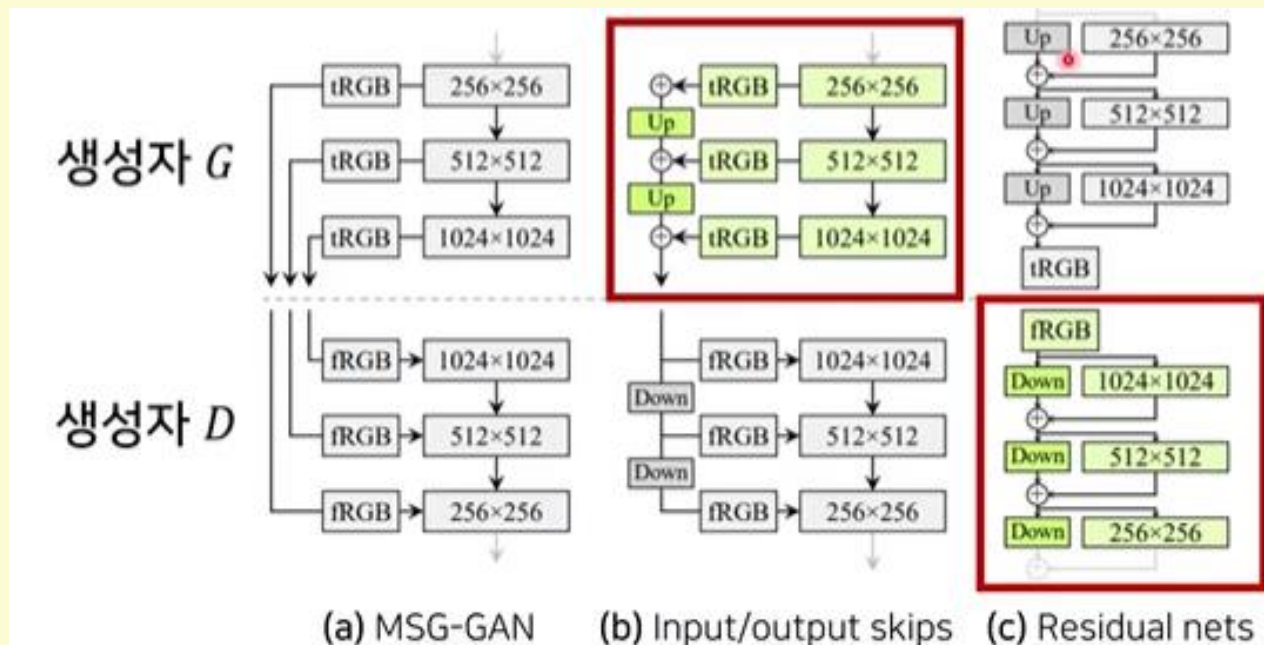


저해상도 부분에서 높은 frequency detail을 갖는 이미지를 만들어버렸다면 후반부에 있는 layer는 이미지 detail을 살리는 방식으로 결과 생성



Generated image

Feature map  $128^2$



단순한 feedforward 네트워크를 사용하는 방식을 다시 사용

FFHQ	D original		D input skips		D residual	
	FID	PPL	FID	PPL	FID	PPL
G original	4.32	265	4.18	235	3.58	269
G output skips	4.33	169	3.77	127	<b>3.31</b>	<b>125</b>
G residual	4.35	203	3.96	229	3.79	243

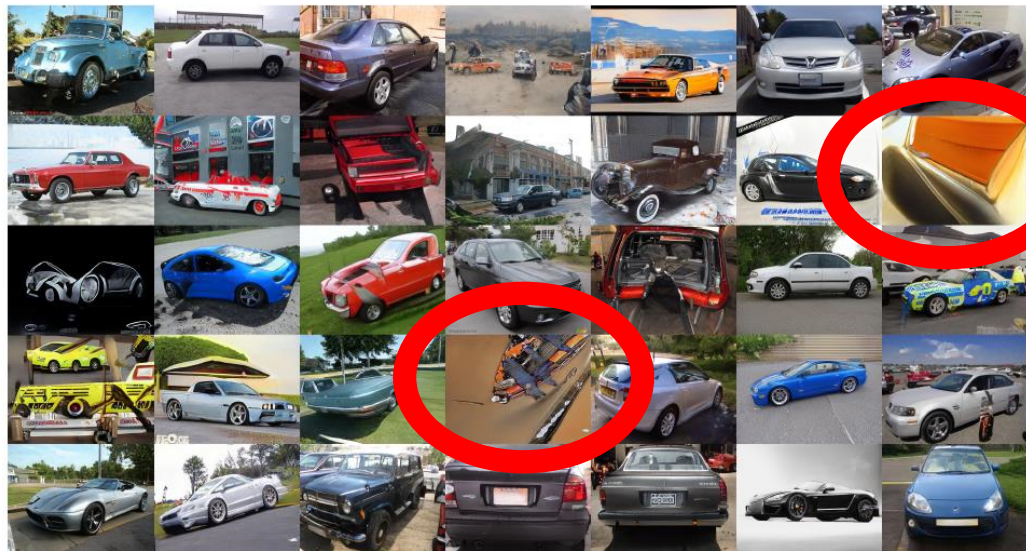
LSUN Car	D original		D input skips		D residual	
	FID	PPL	FID	PPL	FID	PPL
G original	3.75	905	3.23	758	3.25	802
G output skips	3.77	544	3.86	<b>316</b>	3.19	471
G residual	3.93	981	3.40	667	<b>2.66</b>	645

[결과]

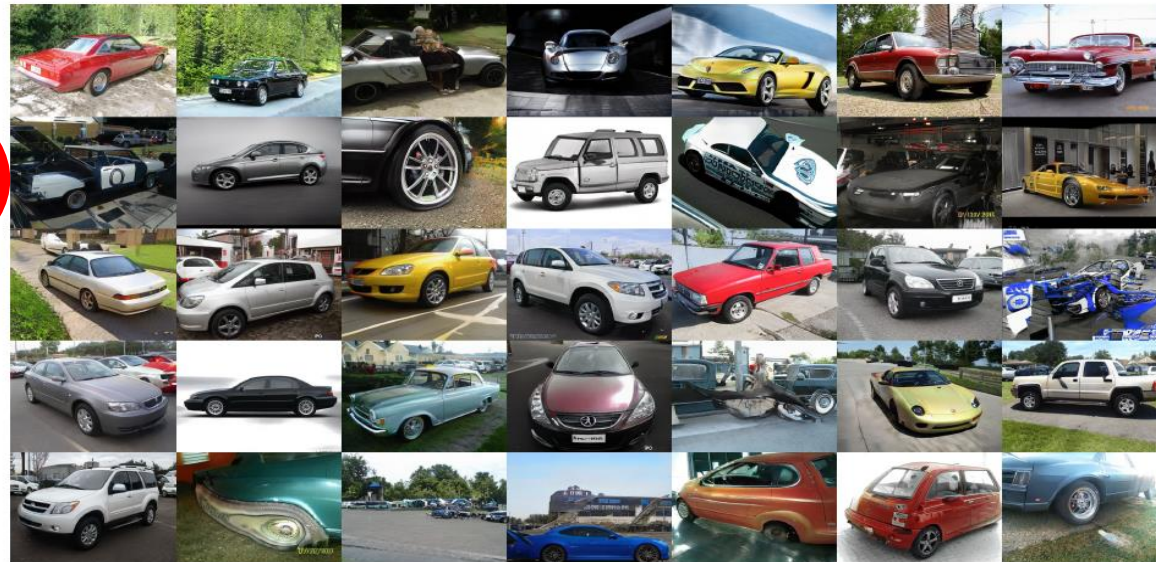
G는 output skips , D는 residual net을 쓸 때 우수



# StyleGAN2 Image Quality and Generator Smoothness



Model 1: FID=3.27, P=0.70, R=0.44, PPL=1485



Model 2: FID= 3.27, P=0.67, R=0.48, PPL=437

- ① FID, Precision & Recall score이 같더라도 **PPL**에 따라서 image quality가 달라질 수 있음
- ② **PPL**이 낮을수록 자동차 모양의 object를 잘 생성함



(a) Low PPL scores



(b) High PPL scores



**Code**

# code

```
@persistent_class
class MappingNetwork(torch.nn.Module):
    def __init__(self,
        z_dim,          # Input latent (Z) dimensionality, 0 = no latent.
        c_dim,          # Conditioning label (C) dimensionality, 0 = no label.
        w_dim,          # Intermediate latent (W) dimensionality.
        num_ws,         # Number of intermediate latents to output, None = do not broadcast.
        num_layers       = 8, # Number of mapping layers.
        embed_features    = None, # Label embedding dimensionality, None = same as w_dim.
        layer_features    = None, # Number of intermediate features in the mapping layers, None = same as w_dim.
        activation        = 'lrelu', # Activation function: 'relu', 'lrelu', etc.
        lr_multiplier     = 0.01, # Learning rate multiplier for the mapping layers.
        w_avg_beta        = 0.995, # Decay for tracking the moving average of W during training, None = do not track.
    ):

```



# code

```
def modulated_conv2d(  
    x,                # Input tensor of shape [batch_size, in_channels, in_height, in_width].  
    weight,           # Weight tensor of shape [out_channels, in_channels, kernel_height, kernel_width].  
    styles,           # Modulation coefficients of shape [batch_size, in_channels].  
    noise             = None, # Optional noise tensor to add to the output activations.  
    up                = 1,    # Integer upsampling factor.  
    down              = 1,    # Integer downsampling factor.  
    padding           = 0,    # Padding with respect to the upsampled image.  
    resample_filter   = None, # Low-pass filter to apply when resampling activations. Must be prepared beforehand by calling upfirdn2d.setup_filter().  
    demodulate        = True, # Apply weight demodulation?  
    flip_weight       = True, # False = convolution, True = correlation (matches torch.nn.functional.conv2d).  
    fused_modconv     = True, # Perform modulation, convolution, and demodulation as a single fused operation?  
):  
    batch_size = x.shape[0]  
    out_channels, in_channels, kh, kw = weight.shape  
    misc.assert_shape(weight, [out_channels, in_channels, kh, kw]) # [OIkk]  
    misc.assert_shape(x, [batch_size, in_channels, None, None]) # [NIHW]  
    misc.assert_shape(styles, [batch_size, in_channels]) # [NI]  
  
    # Pre-normalize inputs to avoid FP16 overflow.  
    if x.dtype == torch.float16 and demodulate:  
        weight = weight * (1 / np.sqrt(in_channels * kh * kw) / weight.norm(float('inf'), dim=[1,2,3], keepdim=True)) # max_Ikk  
        styles = styles / styles.norm(float('inf'), dim=1, keepdim=True) # max_I
```

# code

```
@persistence.persistent_class
class SynthesisBlock(torch.nn.Module):
    def __init__(self,
        in_channels,           # Number of input channels, 0 = first block.
        out_channels,          # Number of output channels.
        w_dim,                 # Intermediate latent (W) dimensionality.
        resolution,            # Resolution of this block.
        img_channels,          # Number of output color channels.
        is_last,               # Is this the last block?
        architecture = 'skip', # Architecture: 'orig', 'skip', 'resnet'.
        resample_filter = [1,3,3,1], # Low-pass filter to apply when resampling activations.
        conv_clamp = None,      # Clamp the output of convolution layers to +-X, None = disable clamping.
        use_fp16 = False,      # Use FP16 for this block?
        fp16_channels_last = False, # Use channels-last memory format with FP16?
        **layer_kwargs,        # Arguments for SynthesisLayer.
    ):
        assert architecture in ['orig', 'skip', 'resnet']
```

FFHQ	D original		D input skips		D residual	
	FID	PPL	FID	PPL	FID	PPL
G original	4.32	265	4.18	235	3.58	269
G output skips	4.33	169	3.77	127	<b>3.31</b>	<b>125</b>
G residual	4.35	203	3.96	229	3.79	243

LSUN Car	D original		D input skips		D residual	
	FID	PPL	FID	PPL	FID	PPL
G original	3.75	905	3.23	758	3.25	802
G output skips	3.77	544	3.86	<b>316</b>	3.19	471
G residual	3.93	981	3.40	667	<b>2.66</b>	645

# code

```
@persistence.persistent_class
class DiscriminatorBlock(torch.nn.Module):
    def __init__(self,
        in_channels,           # Number of input channels, 0 = first block.
        tmp_channels,         # Number of intermediate channels.
        out_channels,         # Number of output channels.
        resolution,           # Resolution of this block.
        img_channels,         # Number of input color channels.
        first_layer_idx,      # Index of the first layer.
        architecture = 'resnet', # Architecture: 'orig', 'skip', 'resnet'.
        activation = 'lrelu',  # Activation function: 'relu', 'lrelu', etc.
        resample_filter = [1,3,3,1], # Low-pass filter to apply when resampling activations.
        conv_clamp = None,     # Clamp the output of convolution layers to +-X, None = disable clamping.
        use_fp16 = False,     # Use FP16 for this block?
        fp16_channels_last = False, # Use channels-last memory format with FP16?
        freeze_layers = 0,    # Freeze-D: Number of layers to freeze.
    ):
    assert in_channels in [0, tmp_channels]
    assert architecture in ['orig', 'skip', 'resnet']
```

FFHQ	D original		D input skips		D residual	
	FID	PPL	FID	PPL	FID	PPL
G original	4.32	265	4.18	235	3.58	269
G output skips	4.33	169	3.77	127	<b>3.31</b>	<b>125</b>
G residual	4.35	203	3.96	229	3.79	243

LSUN Car	D original		D input skips		D residual	
	FID	PPL	FID	PPL	FID	PPL
G original	3.75	905	3.23	758	3.25	802
G output skips	3.77	544	3.86	<b>316</b>	3.19	471
G residual	3.93	981	3.40	667	<b>2.66</b>	645



**output**

## output

```
관리자: 명령 프롬프트
Microsoft Windows [Version 10.0.19044.1645]
(c) Microsoft Corporation. All rights reserved.

C:\WINDOWS\system32>activate stylegan2

(stylegan2) C:\WINDOWS\system32>cd:

(stylegan2) D:\>cd stylegan2-ada-pytorch-main

(stylegan2) D:\stylegan2-ada-pytorch-main>python dataset_tool.py
Usage: dataset_tool.py [OPTIONS]
Try 'dataset_tool.py --help' for help.

Error: Missing option '--source'.

(stylegan2) D:\stylegan2-ada-pytorch-main>python dataset_tool.py --source D:\data\cat --dest D:\data\cat-temp
100% | 114/114 [00:02<00:00, 54.86it/s]

(stylegan2) D:\stylegan2-ada-pytorch-main>
```

```
선택 관리자: 명령 프롬프트
(stylegan2) D:\stylegan2-ada-pytorch-main>python train.py --data D:\data\cat --outdir D:\data\results

Training options:
{
  "num_gpus": 1,
  "image_snapshot_ticks": 50,
  "network_snapshot_ticks": 50,
  "metrics": [
    "fid50k_full"
  ],
  "random_seed": 0,
  "training_set_kwargs": {
    "class_name": "training_dataset.ImageFolderDataset",
    "path": "D:\\data\\cat",
    "use_labels": false,
    "max_size": 114,
    "xflip": false,
    "resolution": 512
  },
  "data_loader_kwargs": {
    "pin_memory": true,
    "num_workers": 3,
    "prefetch_factor": 2
  },
  "G_kwargs": {
    "class_name": "training.networks.Generator",
    "z_dim": 512,
    "w_dim": 512,
    "mapping_kwargs": {
      "num_layers": 2
    },
    "synthesis_kwargs": {
      "channel_base": 32768,
      "channel_max": 512,
      "num_fp16_res": 4,
      "conv_clamp": 256
    }
  },
  "D_kwargs": {
    "class_name": "training.networks.Discriminator",
    "block_kwargs": {},
    "mapping_kwargs": {},
    "epilogue_kwargs": {
      "mbstd_group_size": 4
    },
  },
}
```



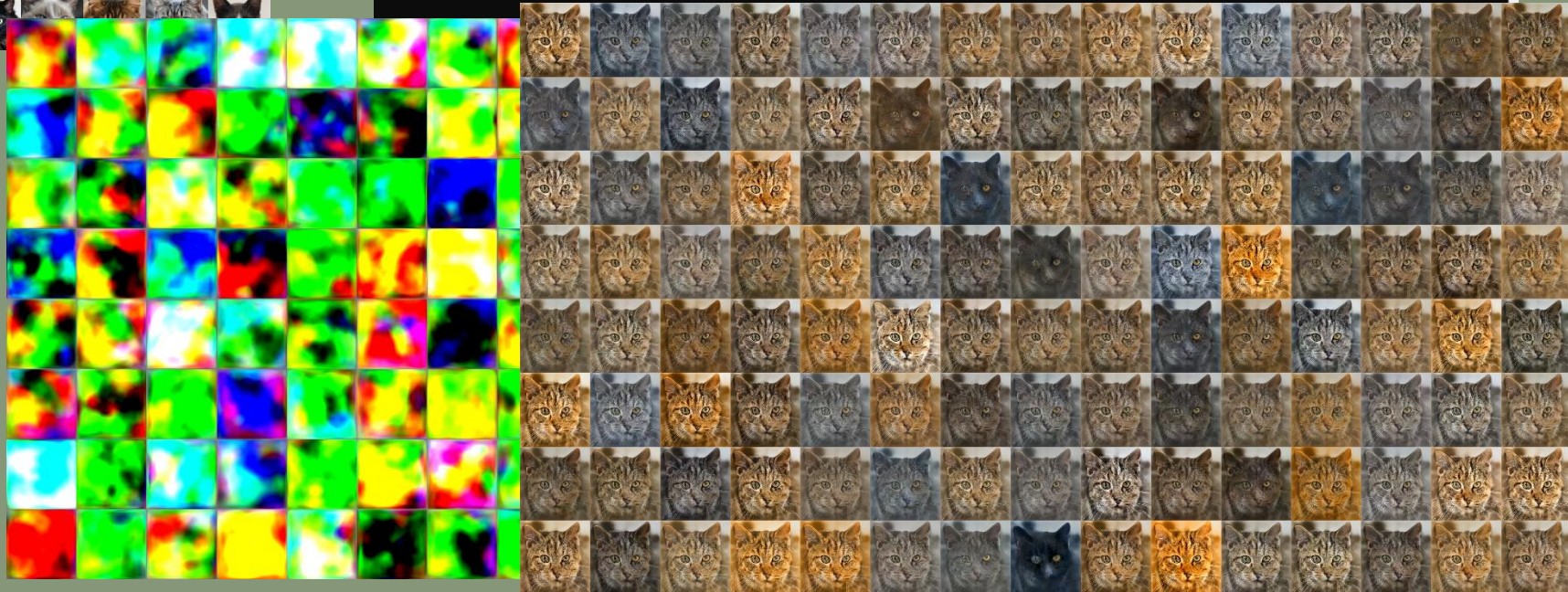
Real Images



output



Real Images



```
Setting up augmentation...
Distributing across 1 GPUs...
Setting up training phases...
Exporting sample images...
Initializing logs...
Training for 25000 kimg...

tick 0 kimg 0.0 time 43s sec/tick 8.4 sec/kimg 4200.97 maintenance 34.9 cnumen 2.66 gnumen 5.79 augment 0.000
Evaluating metrics...
[{"results": [{"fid50k_full": 390.44270353308017, "metric": "fid50k_full", "total_time": 1261.2876143360138, "total_time_str": "21m 01s", "num_gpus": 1, "snapshot_path": "network-snapshot-000000.pkl", "timestamp": 1650724922.136062}]]
```

