A Style-Based Generator Architecture for Generative Adversarial Networks (StyleGAN)

Karras et al. 2019

Reviewed & Presented by Joon Hyeok Kim

Contents

1. What StyleGAN achieved
2. How did StyleGAN do that
3. Pros, cons, and updates of StyleGAN

1. What StyleGAN achieved

StyleGAN improved **disentanglement** by mapping **latent codes** to an intermediate space and controlling them as **styles** at different layers.

Step by step

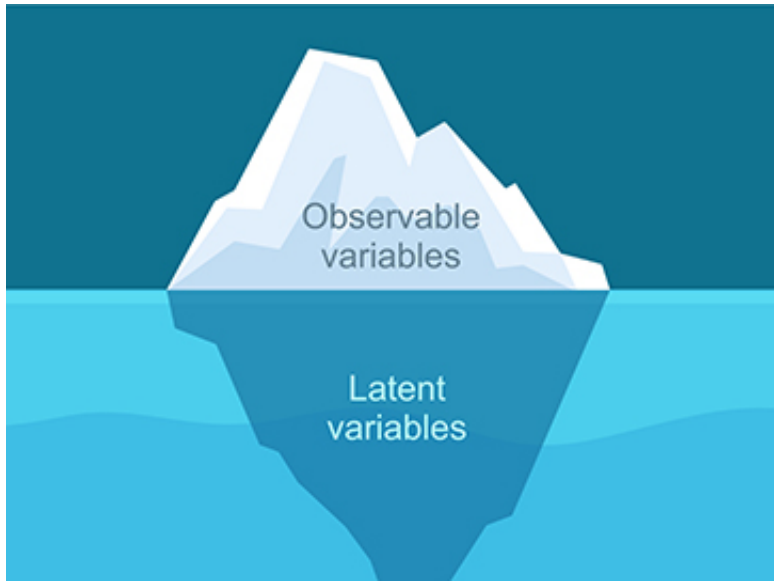
1-1 Latent Code

1-2 Disentanglement

1-1 . Latent Variable (Latent Code)

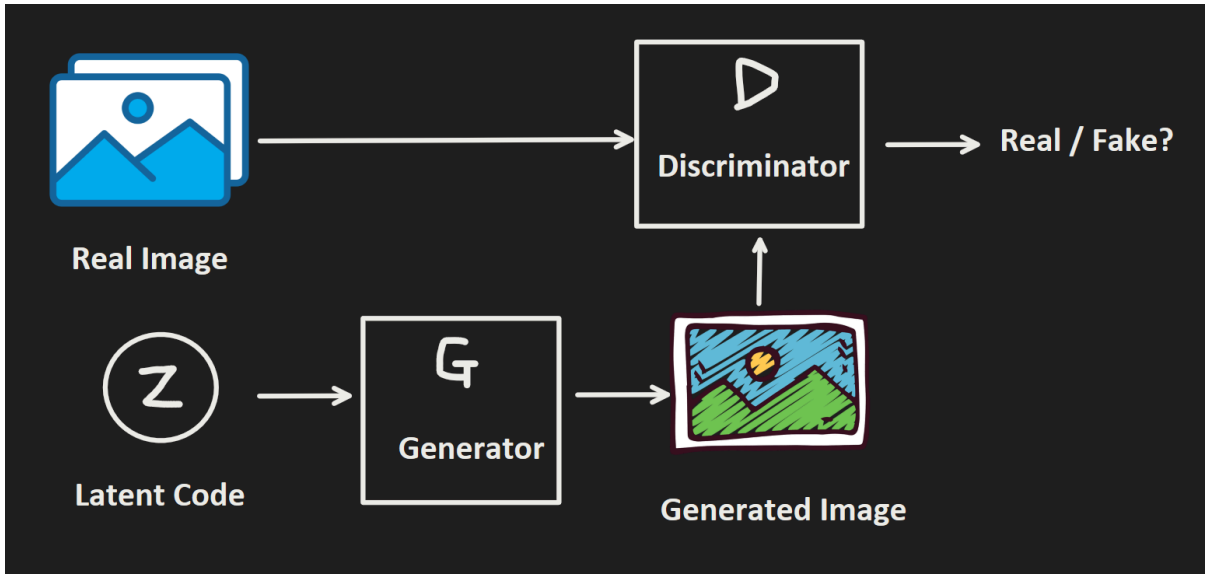
- What is the latent variable z?
  - hidden
  - believed to influence a set of observable outcomes
  - what we want to learn in VAE

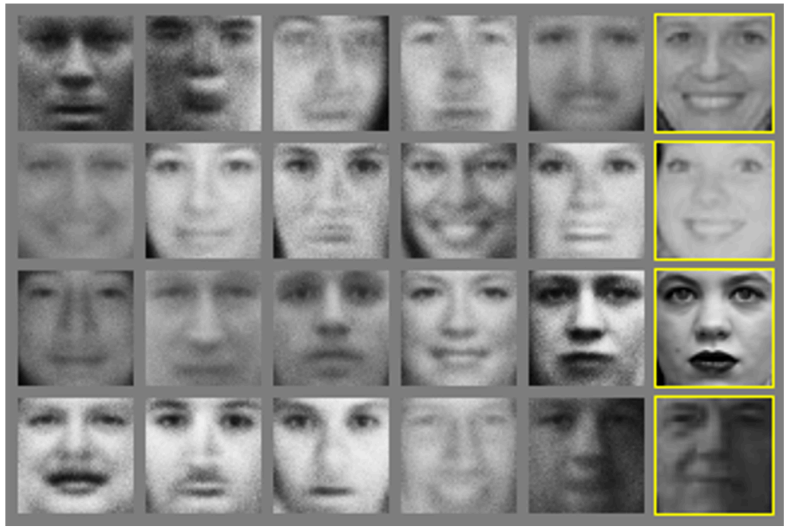
- Did GAN consider latent variables?



GAN (Goodfellow et al., 2014)

- GAN simply utilized **latent code** $z \sim \mathcal{N}(0, \mathbf{I})$
  - Corresponds to VAE's Gaussian prior of $p(z)$
- No meaningful semantics in z
- An arbitrary z' generates an image $G(z')$

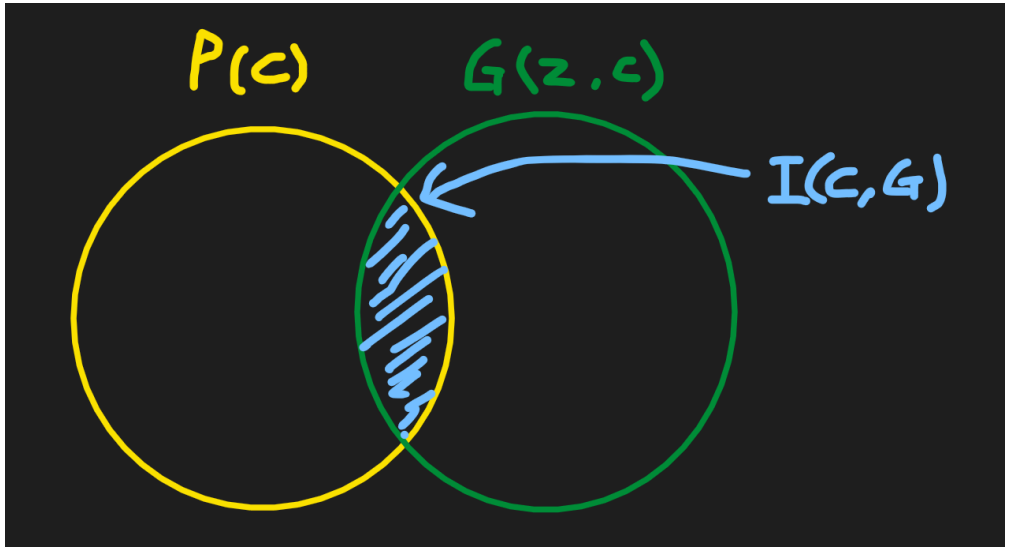


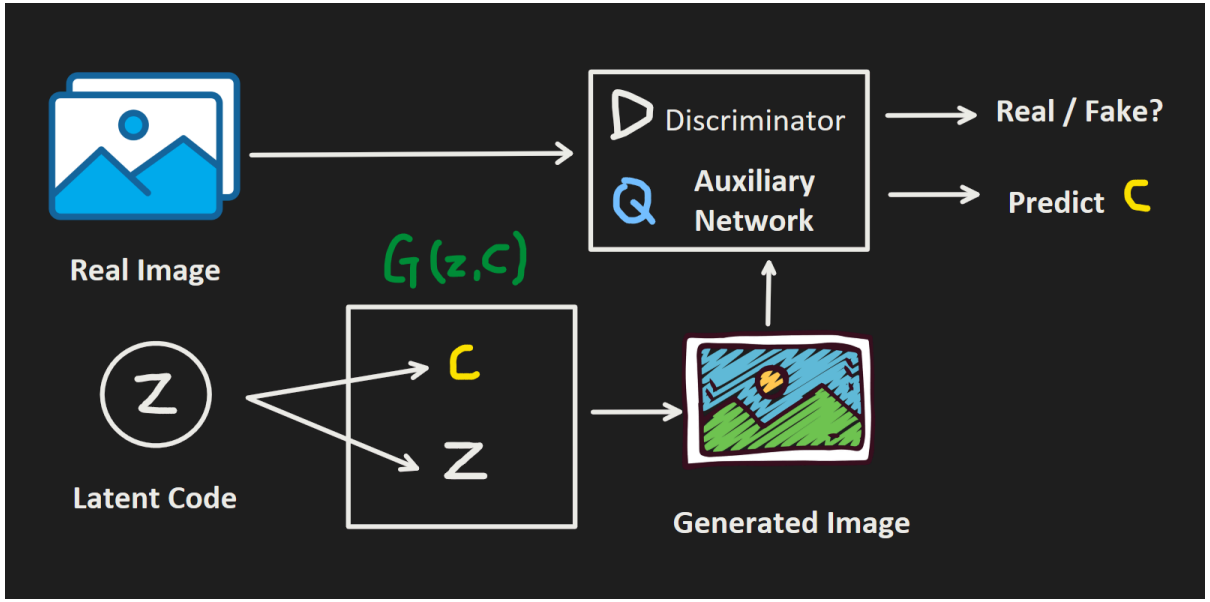


Info GAN (Chen et al., 2016)

- Info GAN decomposed latent code into **noise**(z) and **latent code** (c) using information theoretic approach.

$$\min_G \max_D V_I(D,G) = \underbrace{V(D,G)}_{\text{GAN original}} - \lambda \underbrace{I(c;G(z,c))}_{\text{Mutual Info.}} = \min_{G,Q} \max_D V(D,G) - \lambda L_1(G,Q)$$





Info GAN (Chen et al., 2016)

(a) Azimuth (pose)

(b) Elevation

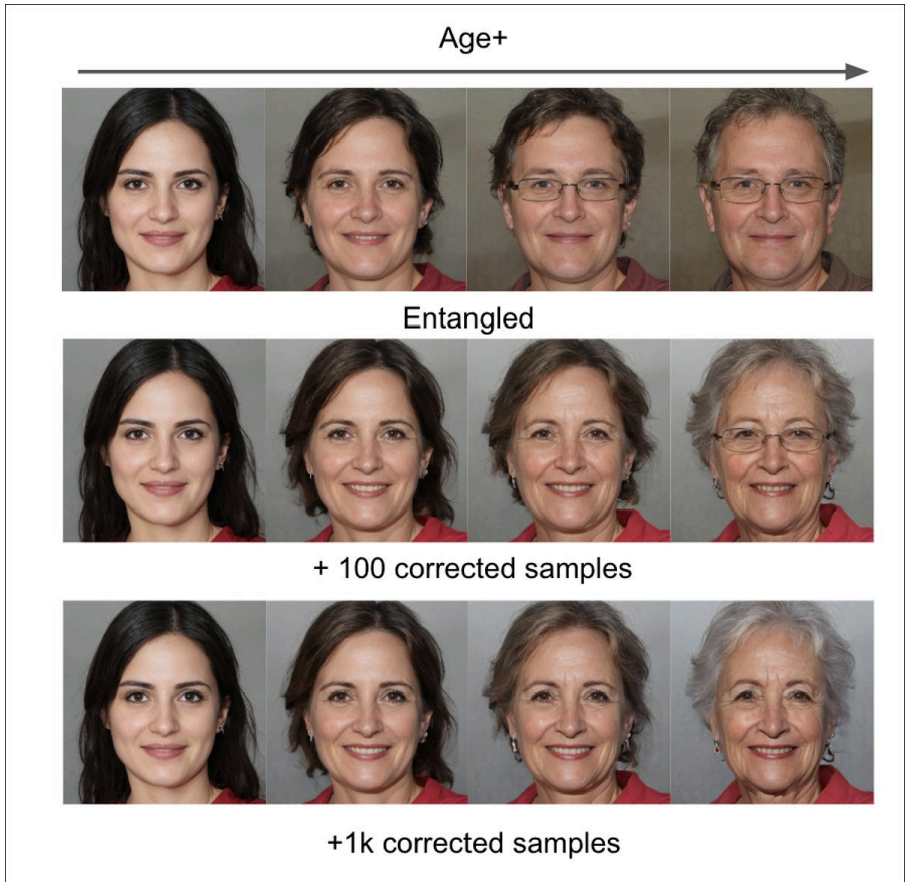
(c) Lighting

(d) Wide or Narrow

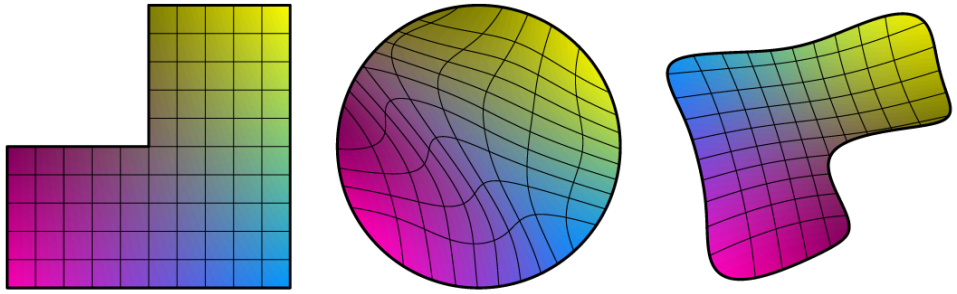
1-2 Disentanglement

Def.) different latent dimensions (or factors) control independent, semantically meaningful aspects of the generated data.

E.g. gender entangled with age



Graphical Description



(a) Distribution of features in training set

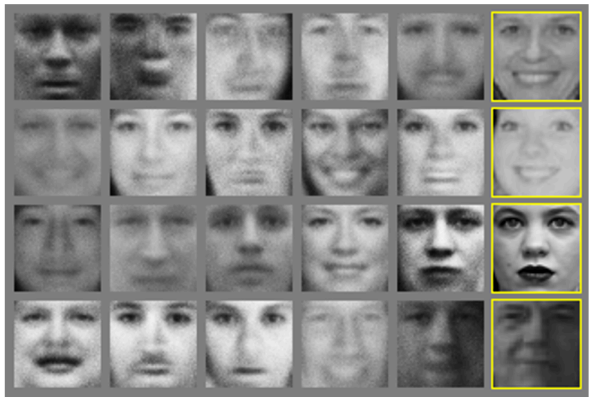
(b) Mapping from $\mathcal{Z}$ to features

(c) Mapping from $\mathcal{W}$ to features

Figure 6. Illustrative example with two factors of variation (image features, e.g., masculinity and hair length). (a) An example training set where some combination (e.g., long haired males) is missing. (b) This forces the mapping from $\mathcal{Z}$ to image features to become curved so that the forbidden combination disappears in $\mathcal{Z}$ to prevent the sampling of invalid combinations. (c) The learned mapping from $\mathcal{Z}$ to $\mathcal{W}$ is able to "undo" much of the warping.

StyleGAN (Karras et al., 2019)

GAN (2014)

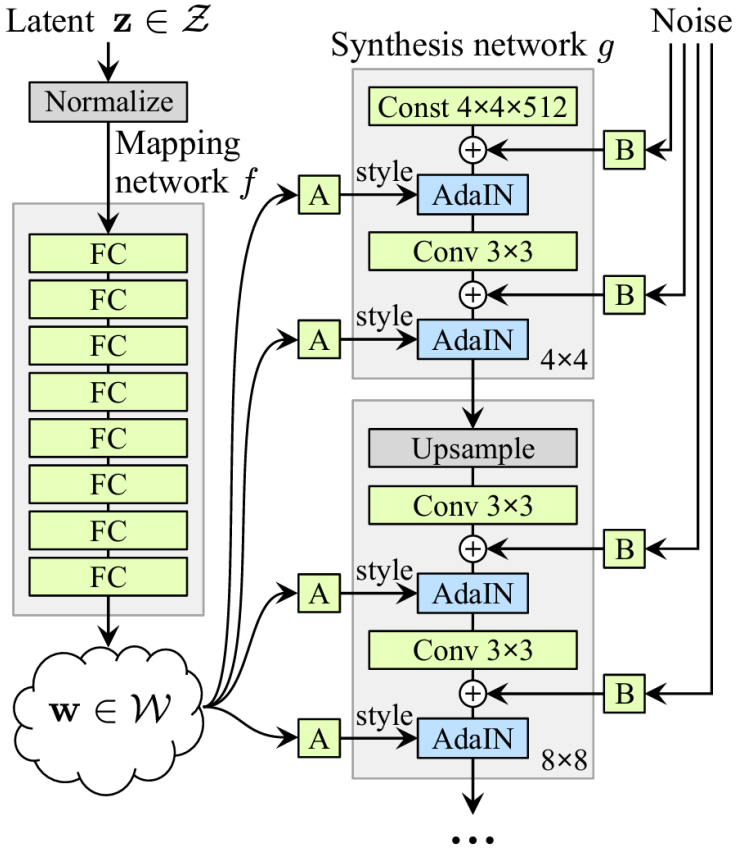


InfoGAN (2016)

(a) Azimuth (pose)

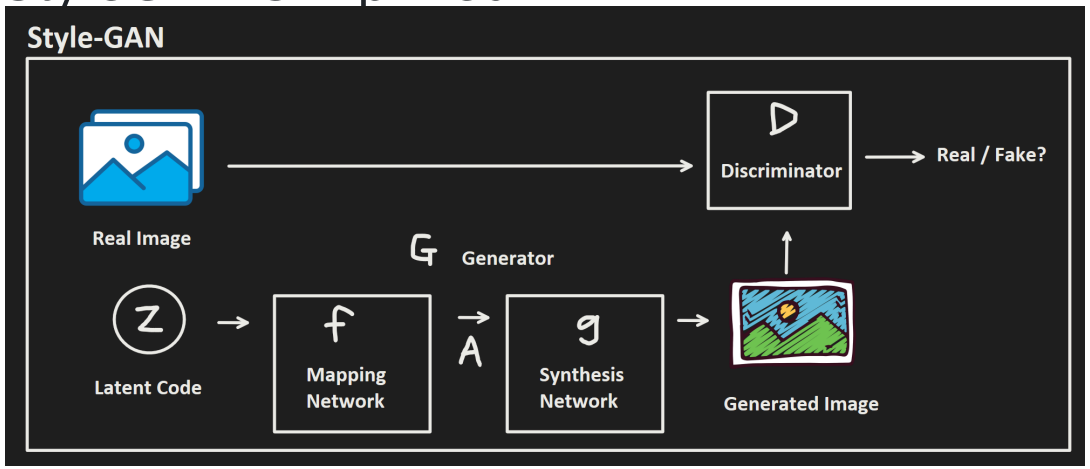
2. How did StyleGAN do that

- StyleGAN Original

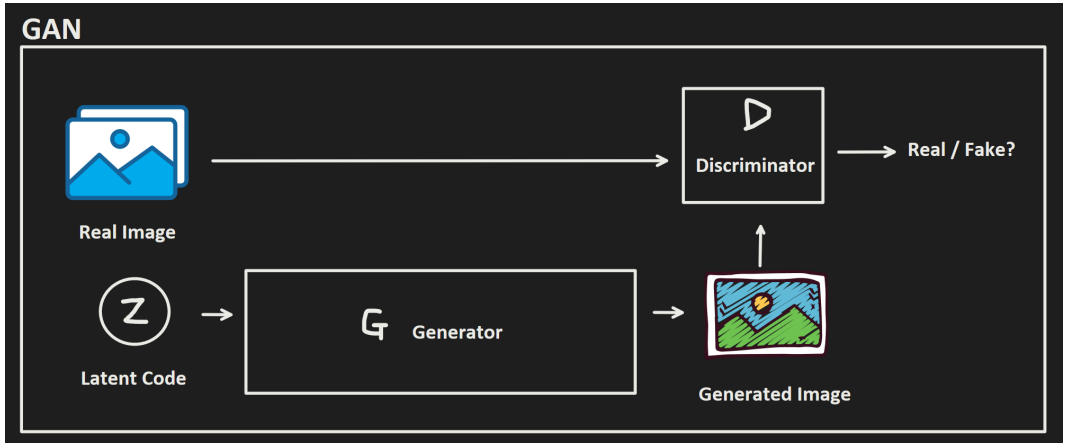


(b) Style-based generator

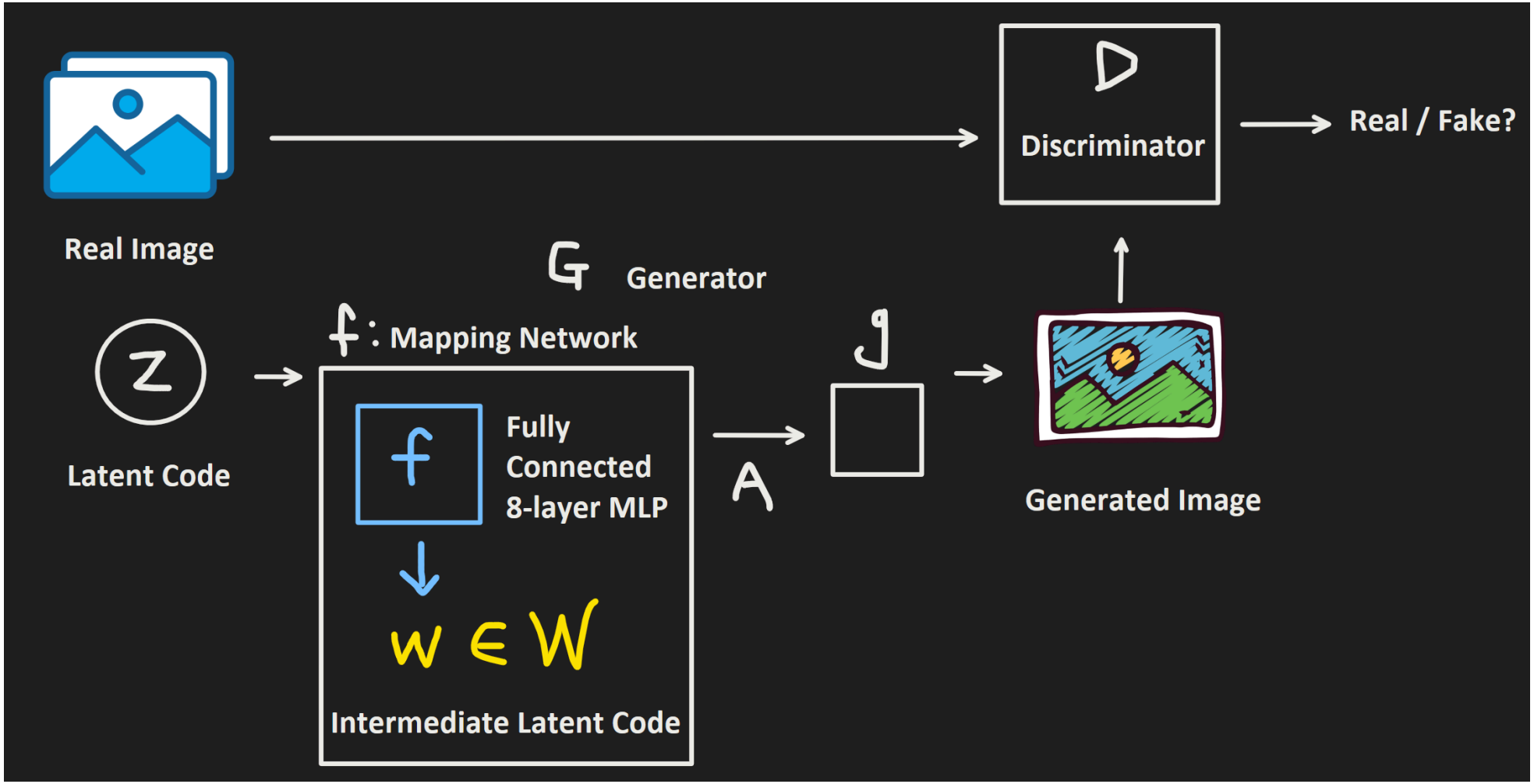
- StyleGAN Simplified



- GAN

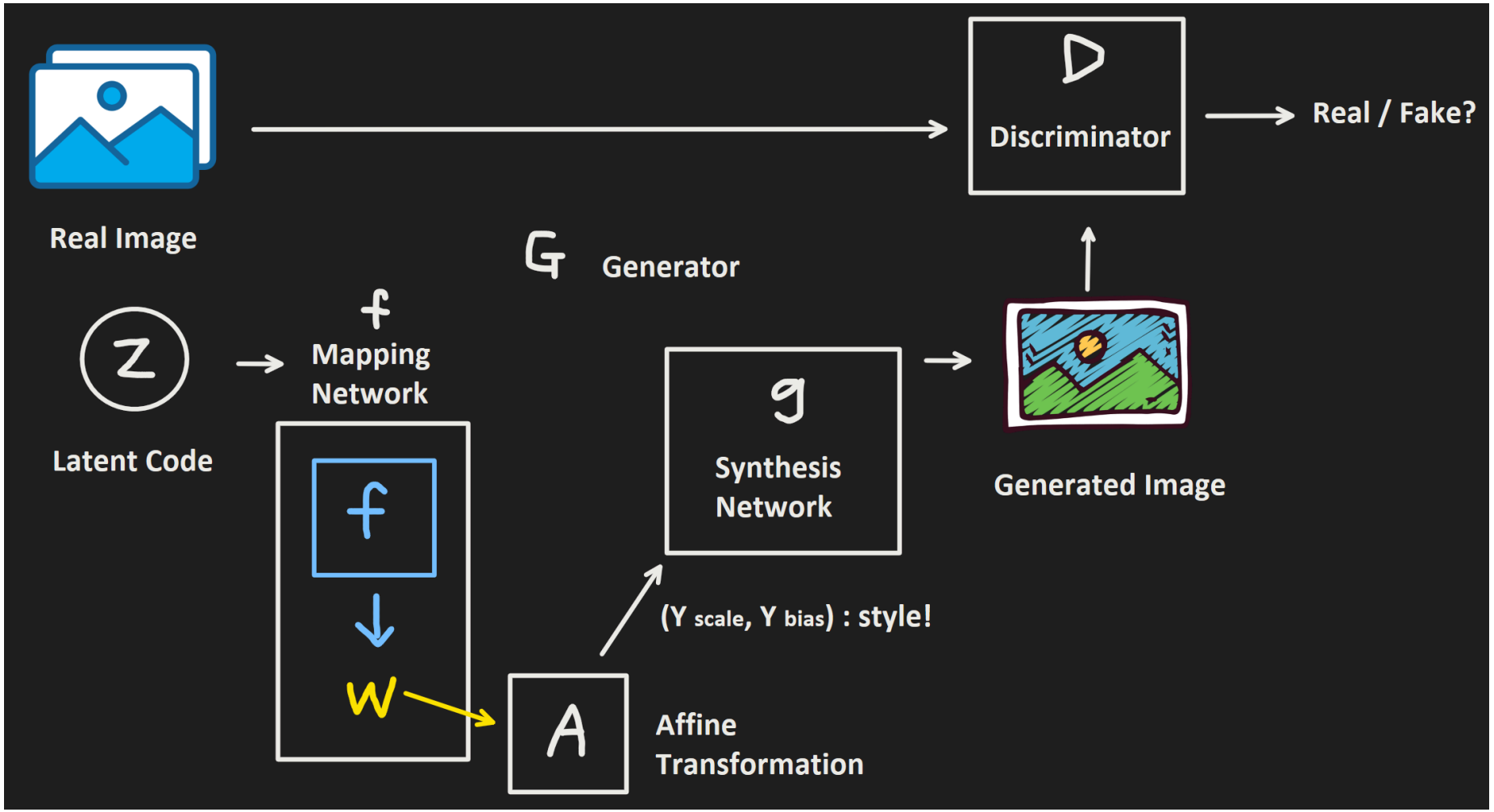


StyleGAN (Karras et al., 2019)



$$\mathbf{z} \in \mathcal{Z}, \quad \mathbf{w} \in \mathcal{W}, \quad \mathcal{Z}, \mathcal{W} \subseteq \mathbb{R}^{512}$$

StyleGAN (Karras et al., 2019)



$A_\ell : \mathcal{W} \to \mathcal{Y}_\ell, \quad A_\ell(\mathbf{w}) = \{(\mathbf{y}_{s,i,\ell}, \mathbf{y}_{s,i,\ell})\}_{i=1}^{C_\ell}, \quad C_\ell$ is the # of channel in layer ℓ

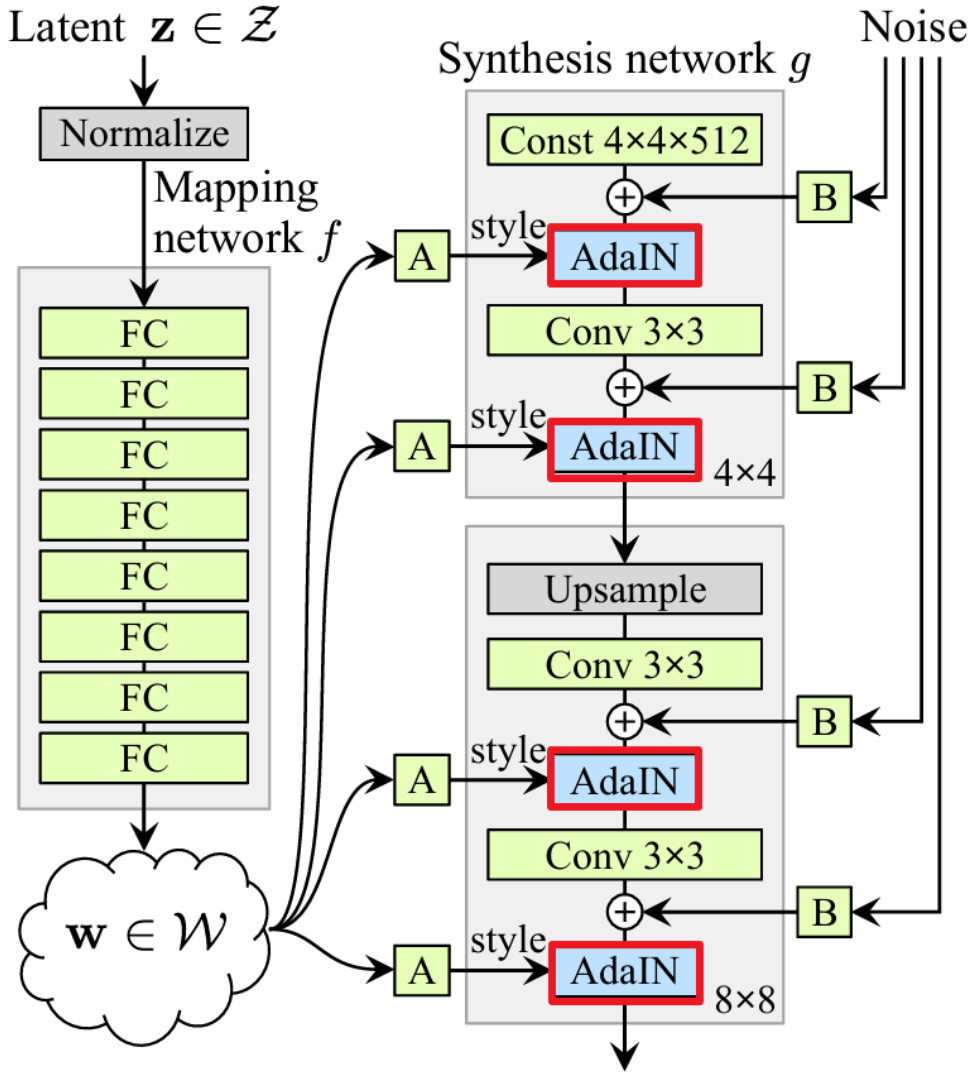
Synthetic Network (g) : AdaIN (Adaptive Instance Normalization)

$$\mathrm{AdaIN}(\mathbf{x}_i, \mathbf{y}) = \mathbf{y}_{s,i} \frac{\mathbf{x}_i - \mu(\mathbf{x}_i)}{\sigma(\mathbf{x}_i)} + \mathbf{y}_{b,i}$$

where $\mathbf{x}_i$ is the i-th channel (feature map) of the activation $\mathbf{x}$

Desc.

- AdaIN is a concept from style transfer.
- The normalization $\left(\frac{\mathbf{x}_i - \mu(\mathbf{x}_i)}{\sigma(\mathbf{x}_i)} \right)$ removes the original style, leaving only the content.
- The scale $\mathbf{y}_{s,i}$ and bias $\mathbf{y}_{b,i}$ reintroduce a new style.



Synthetic Network (g) : Progressive Growing

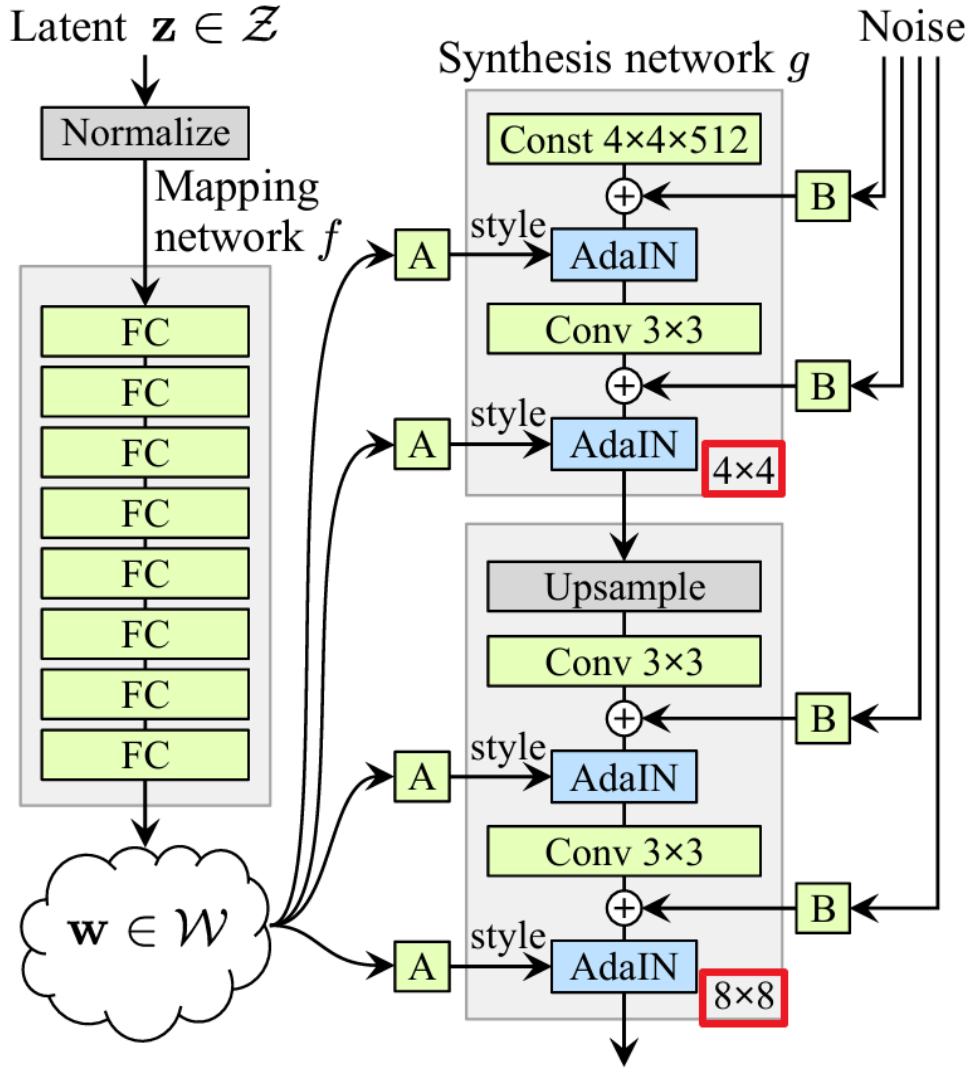
Def.)

Gradually growing the resolution of layers (upsampling)

- $(8 \times 8) \rightarrow (16 \times 16) \rightarrow \cdots \rightarrow (1024 \times 1024)$

Purpose

It stabilizes generating high-resolution 1024×1024 photo realistic image



Experiment : Style Mix

1. Start with two latent codes $\mathbf{z}_A, \mathbf{z}_B$
2. Input $\mathbf{z}_A, \mathbf{z}_B$ into the generator.
   - e.g.)
     $\mathbf{z}_A \rightarrow \mathbf{w}_A = f(\mathbf{z}_A)$
     $\rightarrow (\mathbf{y}_s, \mathbf{y}_b)_A = A(\mathbf{w}_A)$
     $\rightarrow \mathrm{AdaIN}(\mathbf{x}, \mathbf{y}_A)$
     $\rightarrow \cdots \rightarrow \mathrm{Image\ A}$
3. Mix $\mathbf{w}_A$ and $\mathbf{w}_B$
   - Coarse Mix : $\mathbf{w}_B$ from low resolution
   - Fine Mix : $\mathbf{w}_B$ only at high resolution

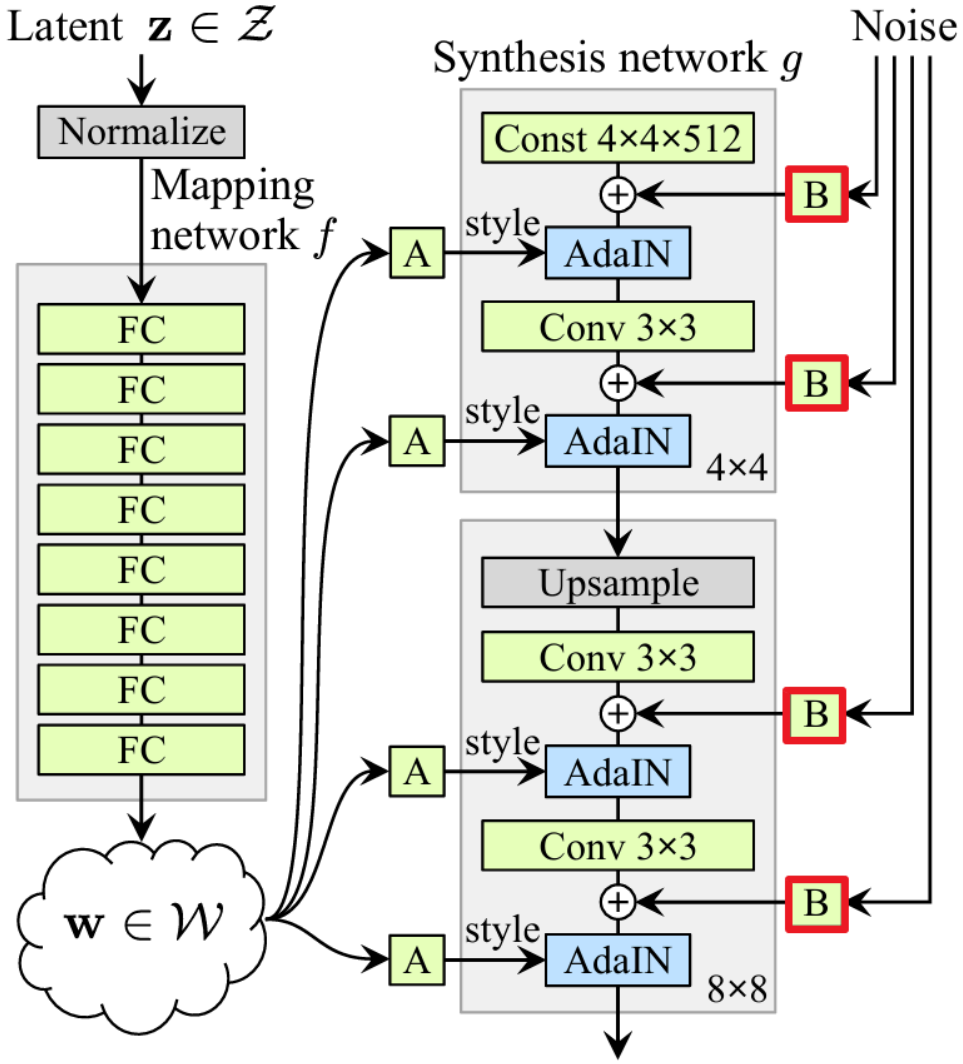
Synthetic Network (g) : Noise (Stochastic Variation)

Def.)

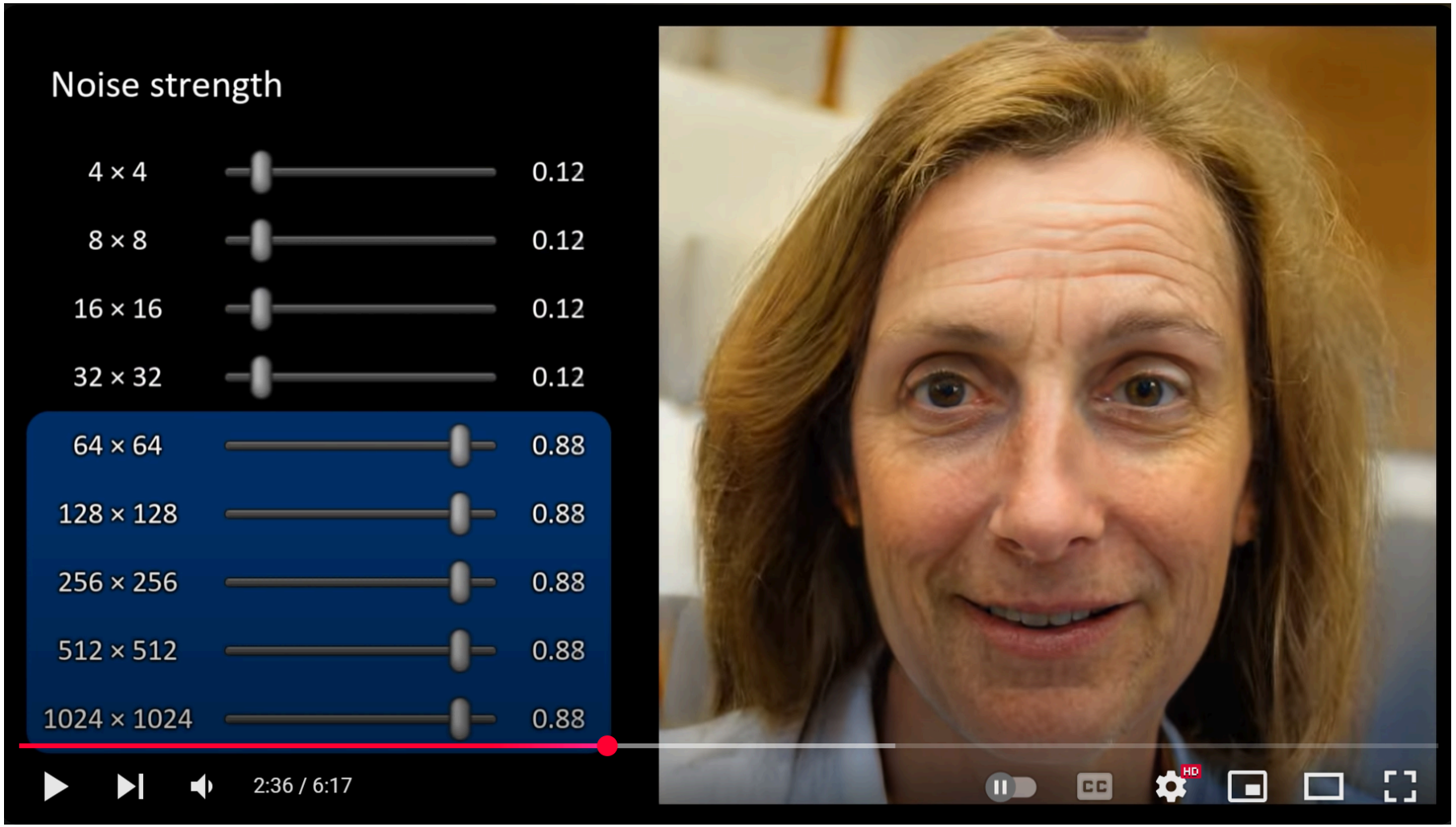
B_ℓ : the noise input to the ℓ-th AdaIN layer

Desc.)

- Single-channel images consisting of uncorrelated Gaussian noise
- Added independently to each pixel $\Rightarrow$ Local Effect!
  - cf.) Styles were complete feature maps scaled and biased with the same values $\Rightarrow$ Global Effect!
- e.g.) placement of hairs, stubble, freckles, etc



Experiment : Noise Strength Control



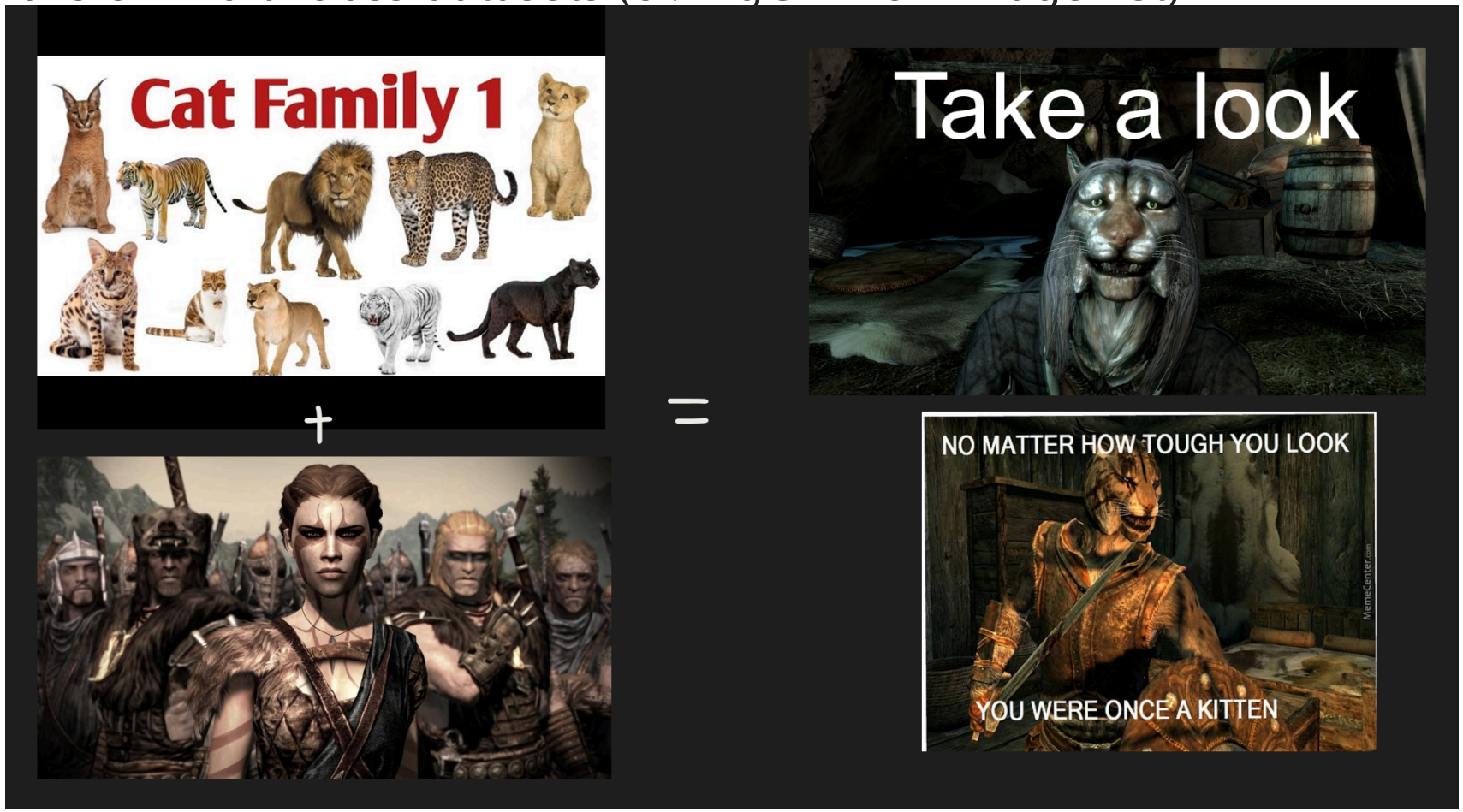
Pros, cons, and updates on StyleGAN

Advantage) StyleGAN...

- disentangled latent control (styles)
- enabled style mixing and stochastic variation
- improved interpretability of the latent dimension
  - e.g.) hair style, skin color, gender, age, etc
- achieved high quality image synthesis (ProGAN)

Disadvantage) StyleGAN...

- is unconditional, so does not support conditional image synthesis.
  - e.g.) "Generate a photo of a cat dancing" $\cdots (X)$
- fails on multi-class datasets (cf. BigGAN on ImageNet)



Disadvantage) StyleGAN...

- has progressive growing checkerboard, droplet artifacts

  - Solved in Style GAN 2 (Karras et al., 2020) by discarding AdaIN and Progressive Growing (...)
  - Instead it...
    - replaced AdaIN with Weight Demodulation
    - performed full-resolution training from the start

Disadvantage) StyleGAN...

- is very computationally expensive at high resolution and needs multi-GPU training

NVIDIA

Research | Research Labs | Publications | AI Playground | Research Areas | Careers | Licensing

People

Tero Karras

Problem?

Tero Karras works as a Senior Distinguished Research Scientist at NVIDIA Research, w
interests revolve around deep learning, generative models, and digital content creatior
generative models and has also had a pivotal role in the development of NVIDIA's RTX
design.

Research Area(s):

- Computer Graphics
- Generative AI
- Real-Time Rendering

Main Field of Interest:

Artificial Intelligence and Machine Learning

Questions

Thank you