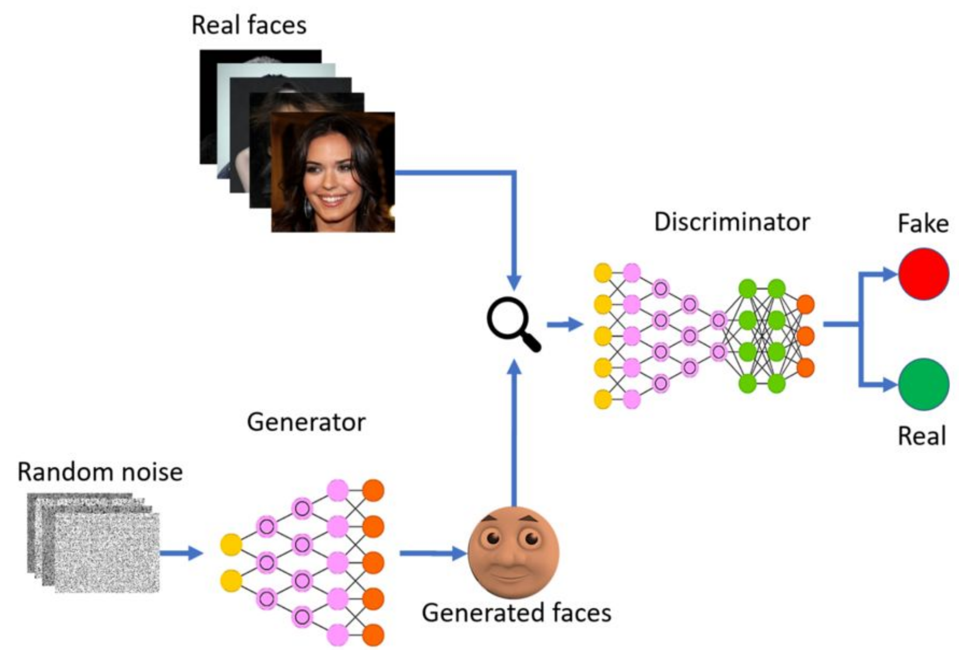
Style GAN (Generative Adversarial Network)



Traditional GAN

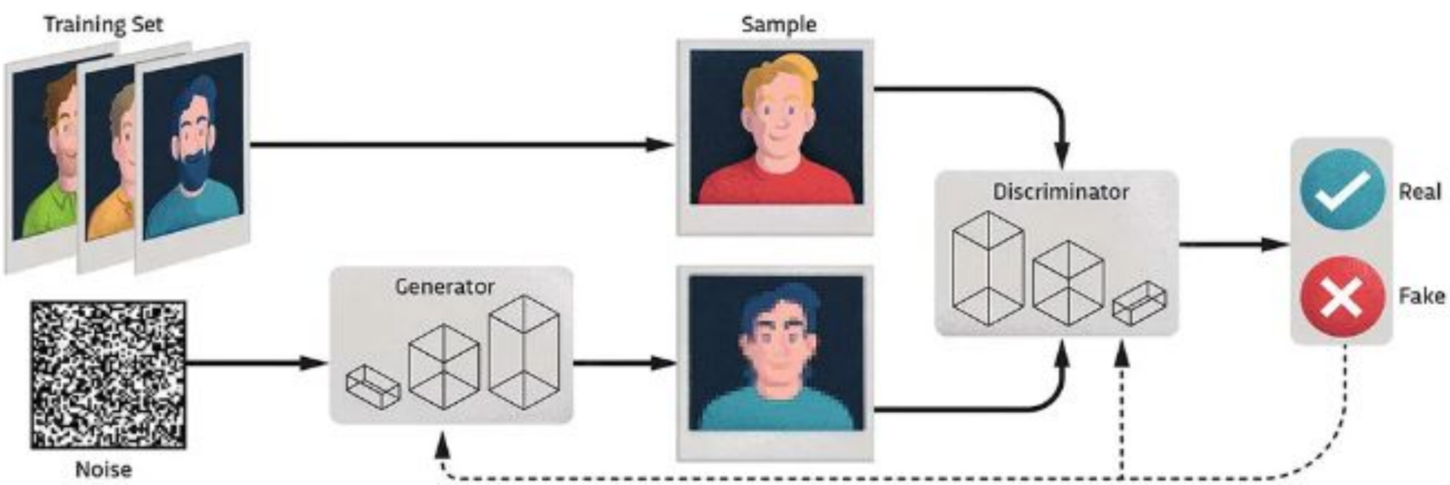
- An approach in generative modeling that generates a new set of data based on a training data given by the user.
- It involves two key components - 1) Generator and 2) Discriminator
- The basic idea behind the GAN's is to pit the two neural networks (Generator and Discriminator) against each other.
- **Generative** – It explains the process behind generating data in a way that's easy to grasp visually.

  Adversarial – The training of the model is done in an adversarial setting.

  Networks – use deep neural networks for training purposes.

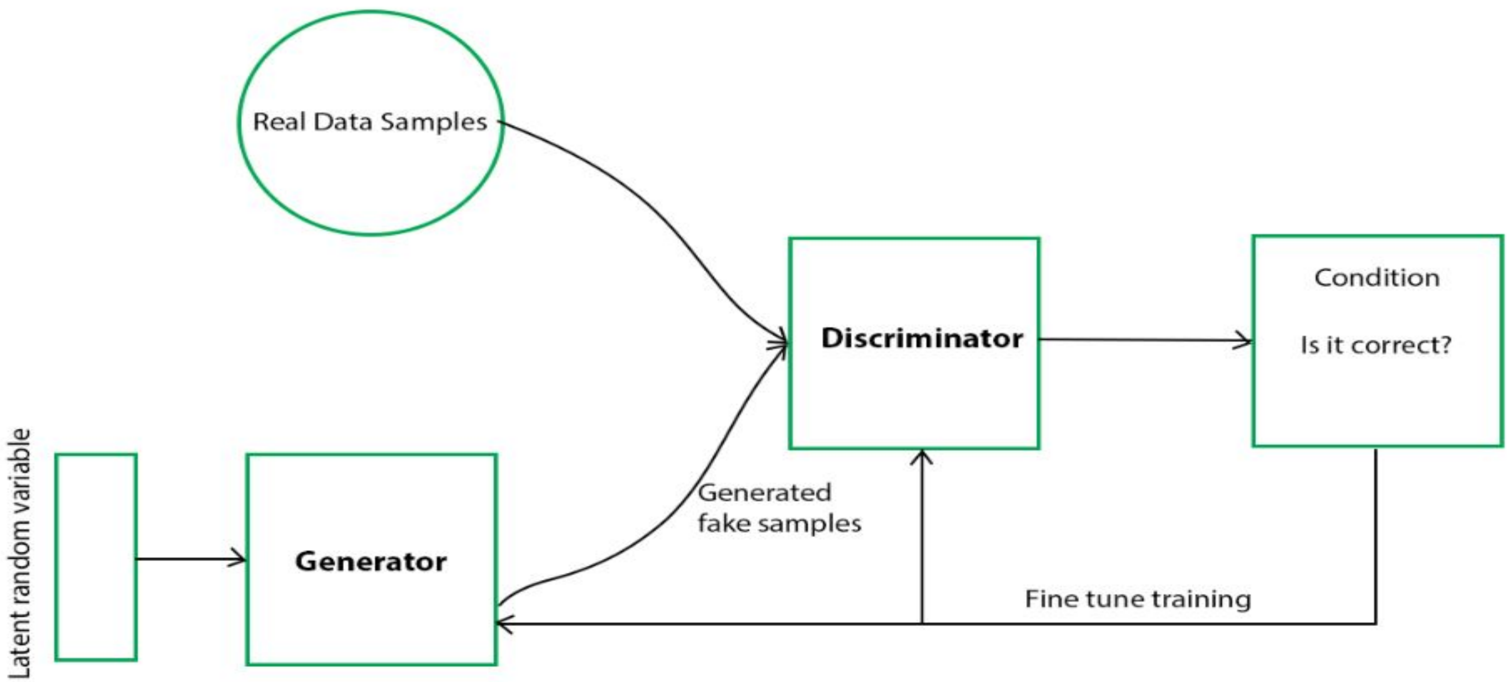
GENERATOR

- It will generate data that is fake data based on original(real) data. It is also a neural network that has hidden layers, activation, loss function.
- It’s aim is to generate the fake image based on feedback and make the discriminator fool that it cannot predict a fake image. And when the discriminator is made a fool by the generator, the training stops and we can say that a generalized GAN model is created.



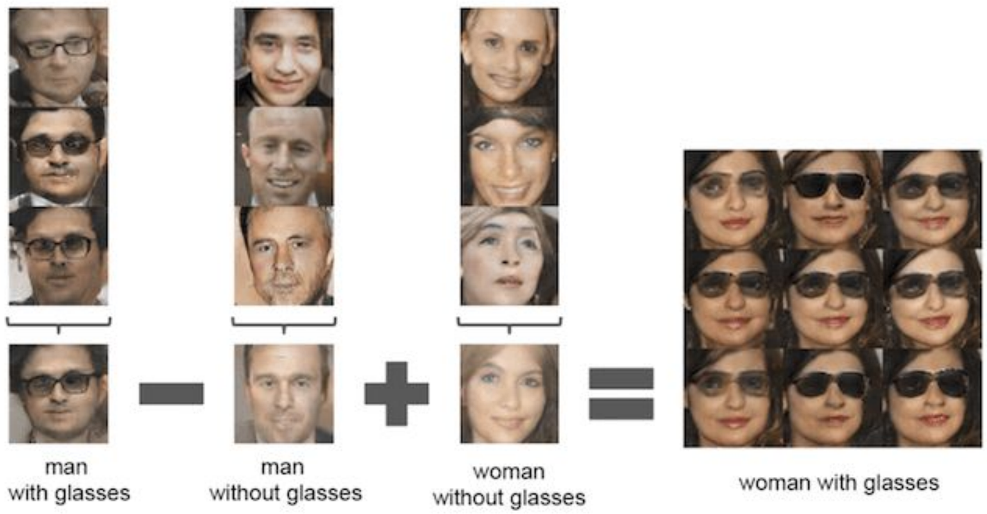
Discriminator

- The discriminator plays a crucial role in a GAN system by evaluating the authenticity of incoming data.
- It competes with the generator to improve the overall performance of the GAN.
- Its ability to differentiate between real and synthetic data enables GANs to produce outputs that closely resemble the input data.
- The output of the discriminator is a single scalar value representing the probability that the input data is real.(close to 0 or 1)

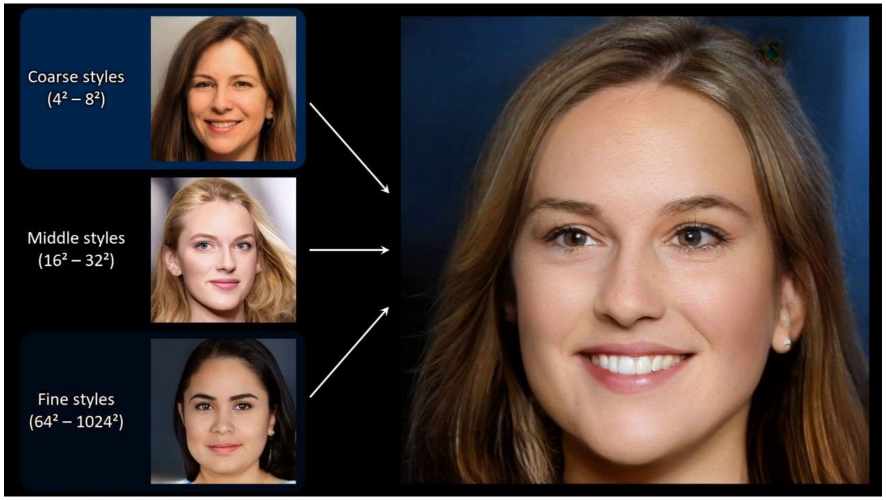


Applications of the traditional GAN

- **Generate Examples for Image Datasets**



- **Generate Realistic Photographs**



Applications of the traditional GAN

- **Text-to-Image Translation**

- **Face Aging**

Limitations of the traditional GAN

- **Limited Control** : Traditional GAN often lack control over the features of generated images.
- **Unrealistic Output Quality**:Produces images with distortions or unrealistic features.
- **Mode collapse** : Occurs when a GAN fails to capture the entire distribution of the training data, resulting in limited varieties of images.
- **Unrealistic global structure**: Traditional GANs may generate images with unrealistic global structures, such as bizarre backgrounds or irregular object placements.
- **Limited scalability to high resolutions**: Generating high-resolution images with traditional GANs is challenging due to memory constraints and training instability.

Style GAN

What is Style GAN?

- A type of generative adversarial network developed by NVIDIA.
- Generates realistic images using a real world data by taking into account about the minimalistic details like pose, texture, colour etc.
- Stochastic variation in the images – it is achieved by introducing noise inputs at various layers of in the network.

How is it different from the traditional GAN?

Difference in components:

- Mapping Network
- Progressive growing
- Adaptive Instance Normalization (AdaIN)

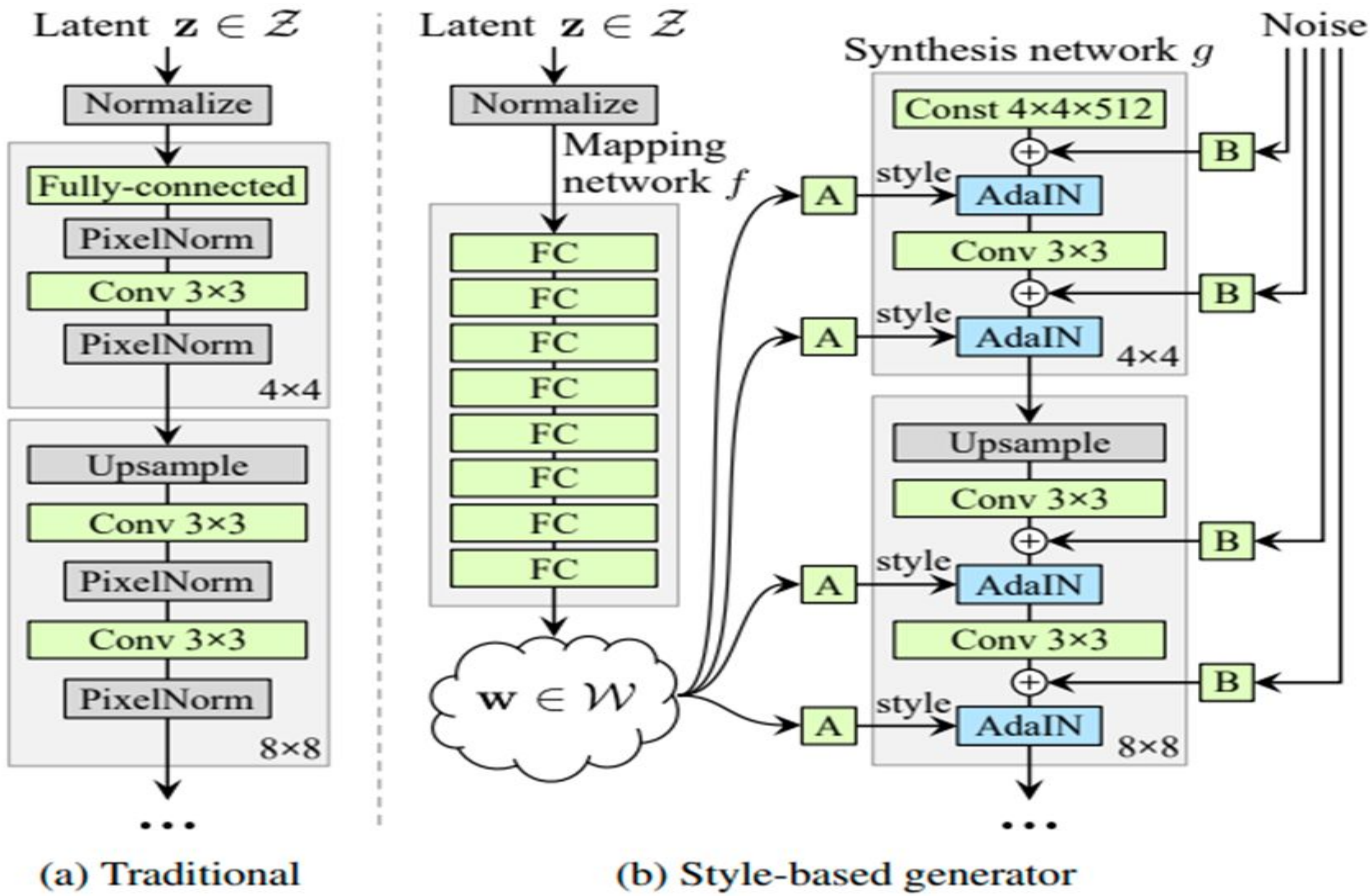
Difference in results:

- More control over the characteristics of the generated image.
- Produces higher quality images through normalization.(Mean and Standard Deviation)
- Addresses the modal collapse in the traditional GAN which is caused by limited variation of the output images.

Architecture: StyleGAN

- Baseline Progressive Growing GANs
- Bi-linear Sampling
- Mapping Network
- Eliminating traditional (Latent) input
- Enhancement of Noisy

Difference b/w the traditional and style GAN

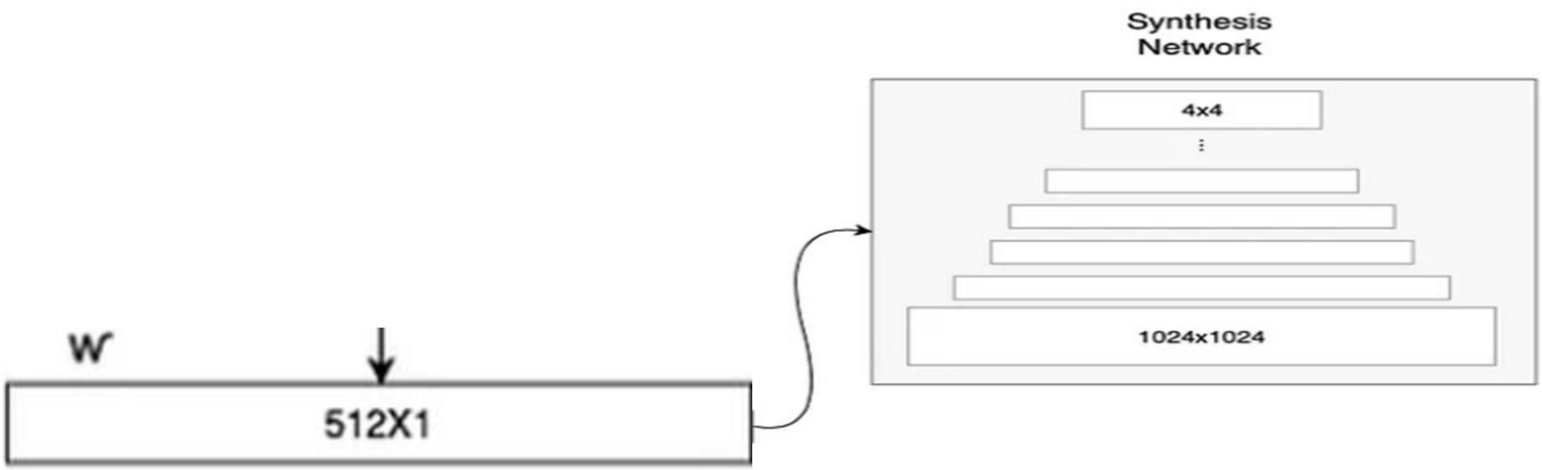


Working of Mapping Network

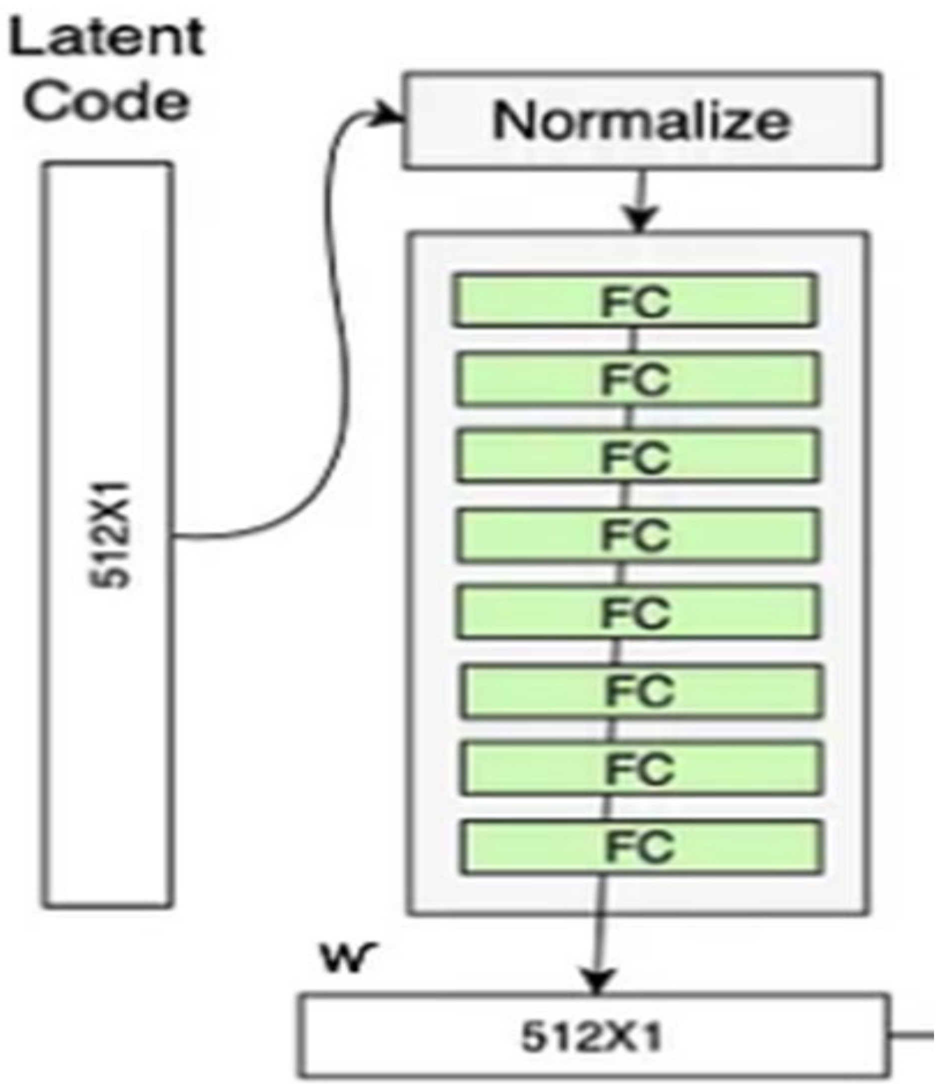
- The input vector is given to the mapping network .
- The goal of the mapping network is to encode the input vector into an intermediate vector.
- The output of the mapping network is of same size as that of input which is 512*1.
- The output vector is then given to the every convolution layer in the generator network through another fully connected layer A - ‘Affine Transformation’.
- The affine transformation converts the intermediate vector into a visual representation.

Synthesis/Generator Network:

- A constant vector of size 4*4*512 is used as an initial image(input of 4*4 level) instead of 100*1 used in the traditional GAN.
- Starts by training on a low resolution image and adds higher resolution layer in the next stages.
- This helps in generating high quality images with smooth transitions between different resolutions.

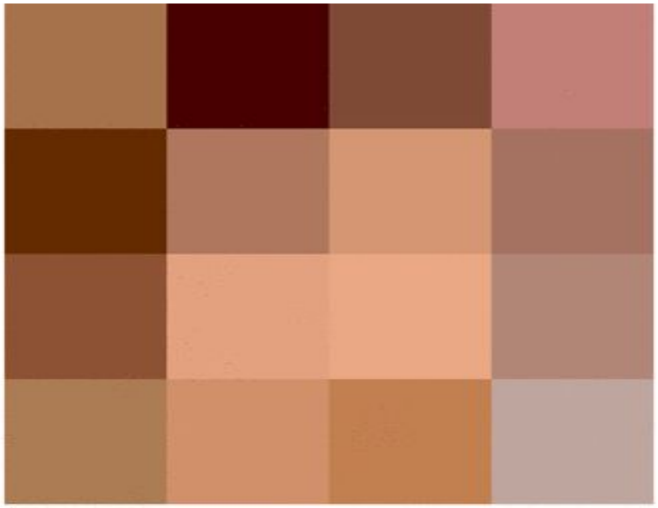


MAPPING NETWORK



How the Upsampling is done?

- Instead of using the transposed convolution layer like in traditional GAN, **bi-linear sampling** is used in the StyleGAN.
- It is an extension of the linear sampling.
- The weighted average of the surrounding pixels are calculated.The weights are based on the distance of each pixel from the desired position.

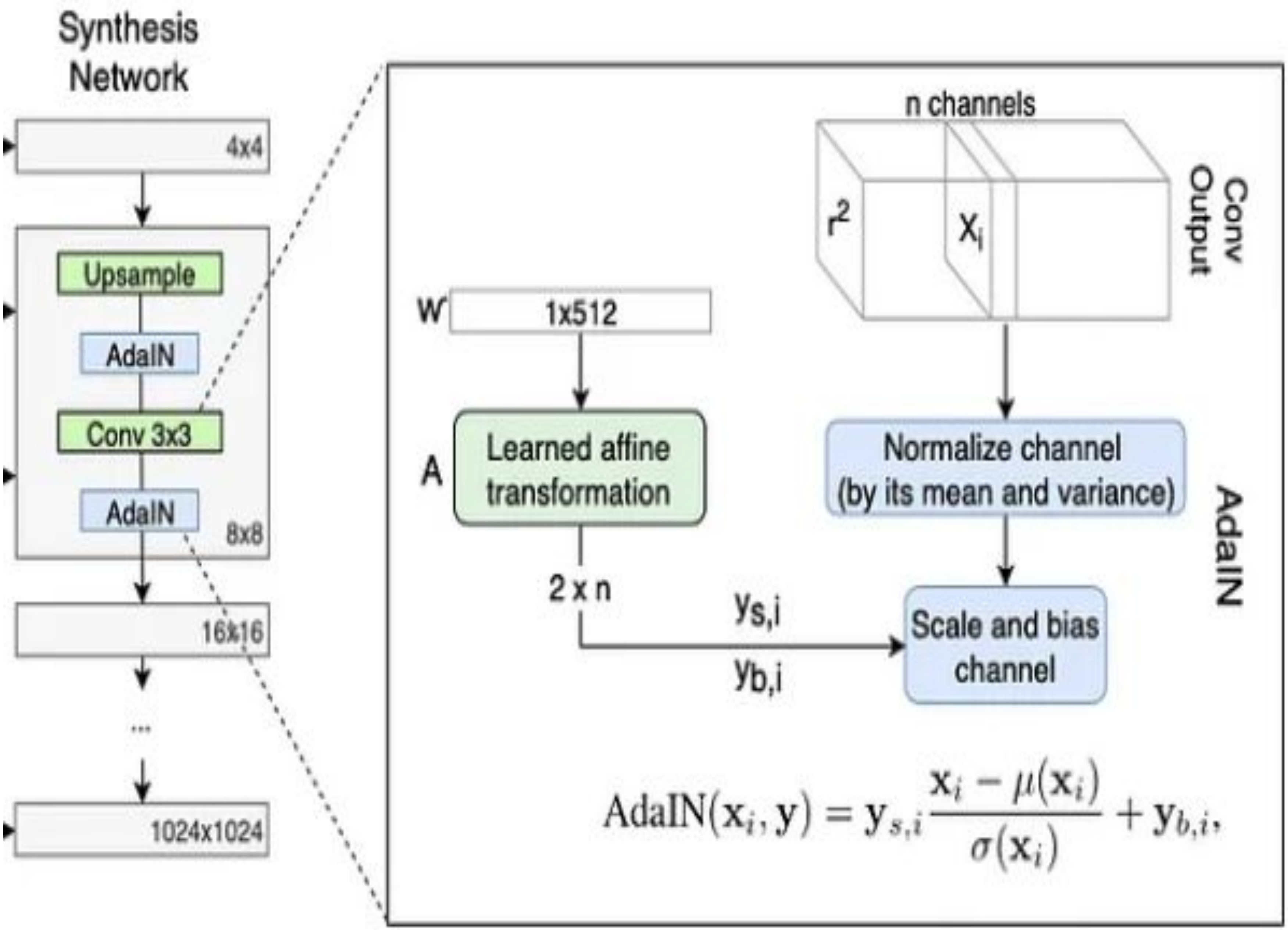


Training time: 0 days
4x4 resolution

AdaIN

- AdaIN is a normalization technique which helps in generating more diverse style images.
- Instead of fixing a constant value for the whole image, it adapts to the local statistics dynamically.
- By adjusting these values , AdaIN effectively transforms the style information from the reference image to the generated image.

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


Synthesis Network
4x4
Upsample
AdaIN
Conv 3x3
AdaIN
8x8
16x16
...
1024x1024
n channels
Conv Output
r2
Xi
W'
1x512
A
Learned affine transformation
Normalize channel (by its mean and variance)
AdaIN
2 x n
ys,i
yb,i
Scale and bias channel
$$\mathrm{AdaIN}(\mathbf{x}_i, \mathbf{y}) = \mathbf{y}_{s,i}\frac{\mathbf{x}_i - \mu(\mathbf{x}_i)}{\sigma(\mathbf{x}_i)} + \mathbf{y}_{b,i},$$

Gaussian Noise

- The Gaussian noise(B) accounts for the stochasticity factor.

- Features such as the exact placement of hairs, stubble, freckles, or skin pores can be considered as stochastic.

- These features are divided into 3 types:

1. **Coarse** - resolution of up to 82 - affects pose, general hair style, face shape, etc.
2. **Middle** - resolution of 162 to 322 - affects finer facial features, hair style, eyes open/closed, etc.
3. **Fine** - resolution of 642 to 10242 - affects color scheme (eye, hair and skin) and micro features.

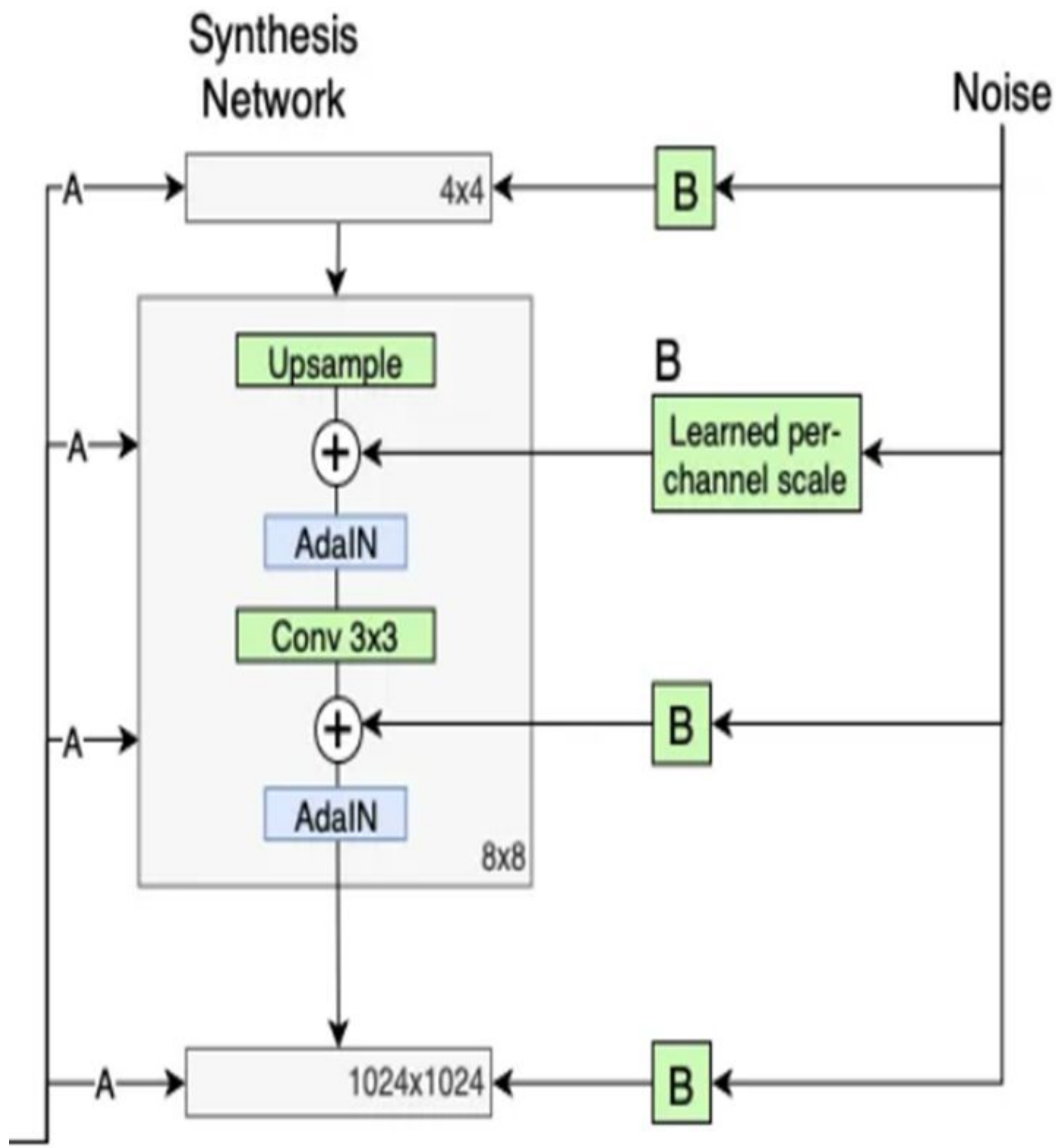
Gaussian Noise

Effect of noise inputs at different layers of our generator:

a. Noise is applied to all layers.
b. No noise.
c. Noise in fine layers only (64^2–1024^2)
d. Noise in coarse layers only (4^2 – 32^2).

Combination of both

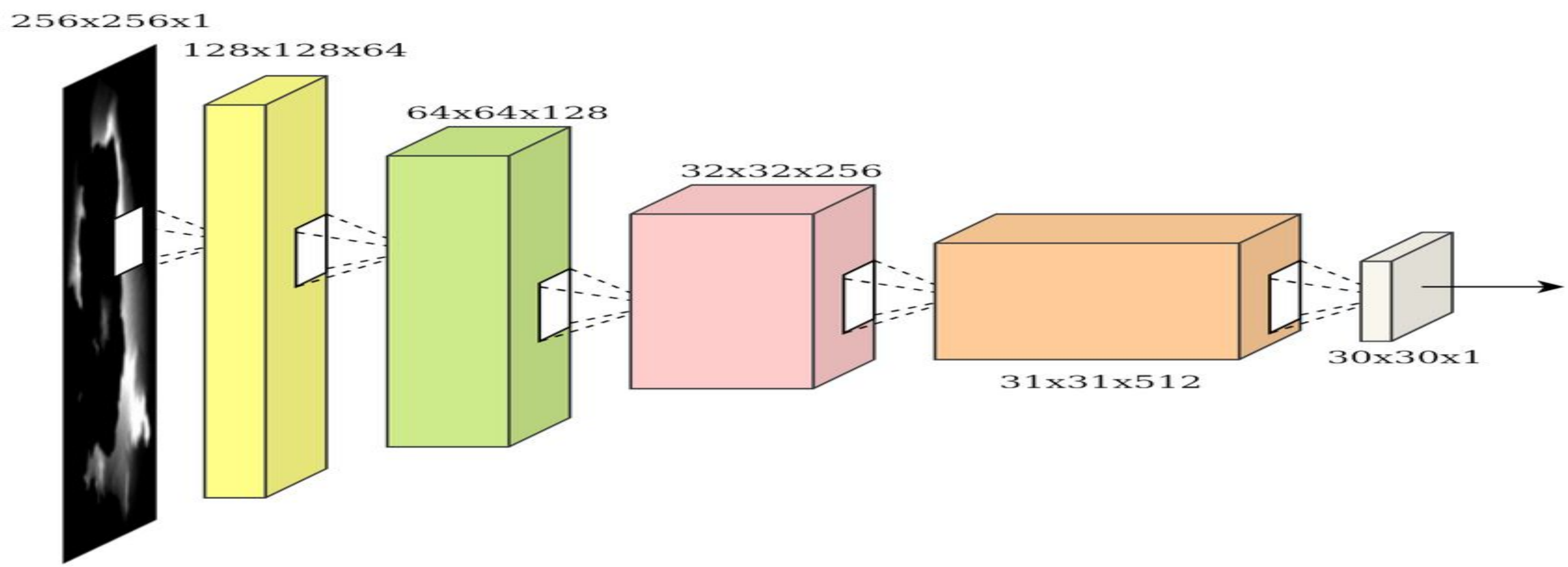
In combination, the AdaIN and Gaussian Noise contribute to the flexibility and realism of the generated images.



DISCRIMINATOR

- While there a lot of modifications in the generator , the paper proposed by NVIDIA didn’t discuss much about the discriminator.

- The traditional GAN generally uses a single discriminator that operates on the entire image, whereas the StyleGAN uses:
    a) a patch-based discriminator or
    b) Multi-Scale Discriminator.



DISCRIMINATOR

Multi-Scale Discriminator:

- It analyzes images at multiple scales or resolutions.
- Each scale of the discriminator is responsible for analyzing a particular level of detail in the image.

Patch-Based Discriminator:

- At each scale, the discriminator examines patches of the image rather than the entire image.
- These patches are typically small regions of the image, allowing the discriminator to capture local features and details.
- By analyzing patches, the discriminator can focus on finer details while still maintaining a fixed resolution.

Variants of the StyleGAN

- There are other extensions of StyleGAN like,

  1) StyleGAN2
  2) StyleGAN2-ADA –Adaptive discriminator augmentation
  3) StyleGAN3.
  4)StyleGAN-T

- **StyleGAN2**: Improved training stability and image quality

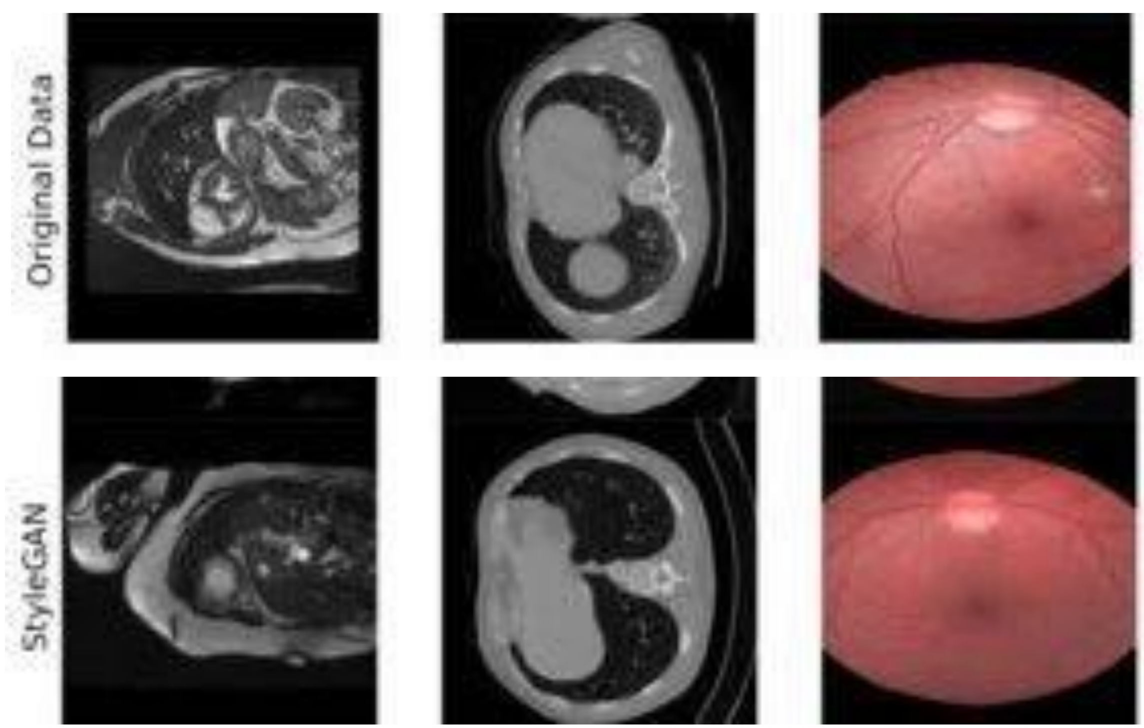
- **StyleGAN2-ADA**: Ability to adjust the discriminator dynamically during training.

- **StyleGAN3**: Ability to generate images with higher fidelity and consistency.
- **StyleGAN-T:** Introduces Text-to-image generation, merging NLP and computer vision, promising efficient and versatile image synthesis.

Applications

Even though the primary discussion in this presentation is based on image generating(face), there are many other applications of StyleGAN as follows:

- **Medical Imaging** - StyleGAN has been explored for generating synthetic medical images,including X-rays MRI scans, and histopathology slides.



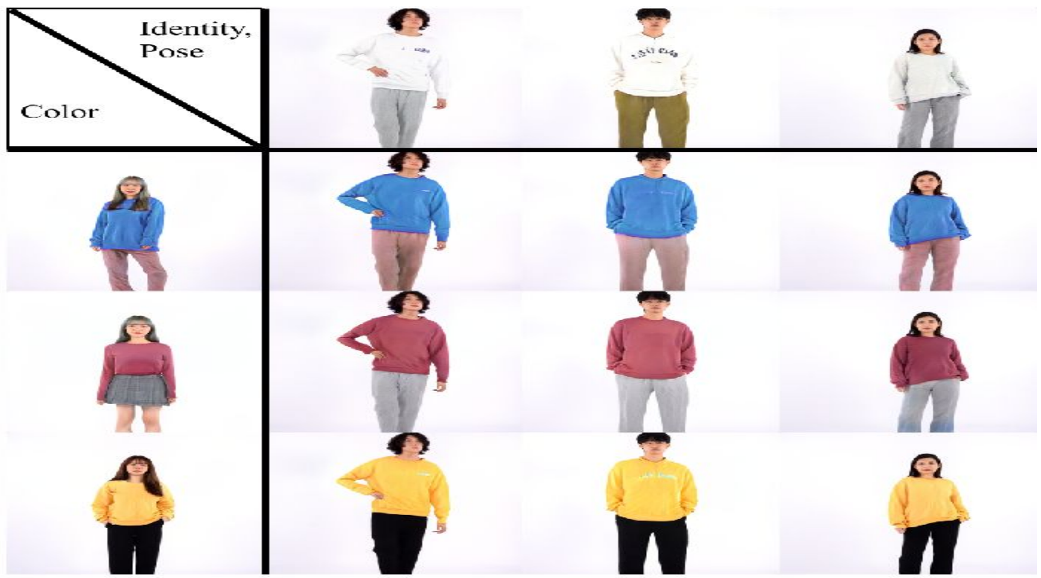
Applications

- **Interior Design** - StyleGAN can be used to generate interior design concepts, including furniture arrangements, room layouts, and decor ideas.

Figure 4. **Layer-wise analysis on object removal.** We remove

Applications

- **Fashion Design -** StyleGAN has been applied to generate fashion designs, including clothing, accessories, and textile patterns.



- **Video Game Development** - StyleGAN-generated assets, such as characters, environments, and textures, can be incorporated into video game development.

Implementation

Dataset :Flickr-Face-HQ(FFHQ)

Image Quality: The FFHQ dataset contains 52,000 high-quality PNG images at a resolution of 512×512, offering detailed facial representations.

Diversity: The dataset features a broad range of ages, ethnicities, and backgrounds, providing a diverse set of human facial images.

Accessories: It includes images with various accessories like eyeglasses, sunglasses, and hats, adding complexity to the dataset.

[**Code:**]()

Conclusion:

- **Controlled Image Style:** StyleGAN empowers precise manipulation of image styles, enhancing creative control.

- **Innovative Architecture:** StyleGAN's novel techniques like Mapping Network and AdaIN redefine synthetic image generation, offering unprecedented flexibility.

- **Remarkable Performance:** Demonstrating exceptional results, particularly in human face synthesis, StyleGAN showcases its capability for realistic image generation.

- **Versatile Applications:** From data augmentation to gaming and art, StyleGAN and StyleGAN-T offer diverse solutions for various industries and creative pursuits.

References:

- A Style-Based Generator Architecture for Generative Adversarial Networks
- analyticsvidhya.com/blog/2021/05/stylegan-explained-in-less-than-five-minutes/
- analyticsvidhya.com/blog/2021/10/an-end-to-end-introduction-to-generative-adversarial-networksgans/

Thank You