GANs and Diffusion Models

Exploring Advanced Techniques in Generative AI.

Presentation Agenda

1. **Introduction**

   Generative Models

2. **GANs**

   Architecture, working, advantages, and challenges.

3. **Diffusion Models**

   Concept, mechanism, and applications.

4. **Comparative Analysis**

   GANs vs. Diffusion Models.

What Are Generative Models?

- Generate new data resembling a given dataset.

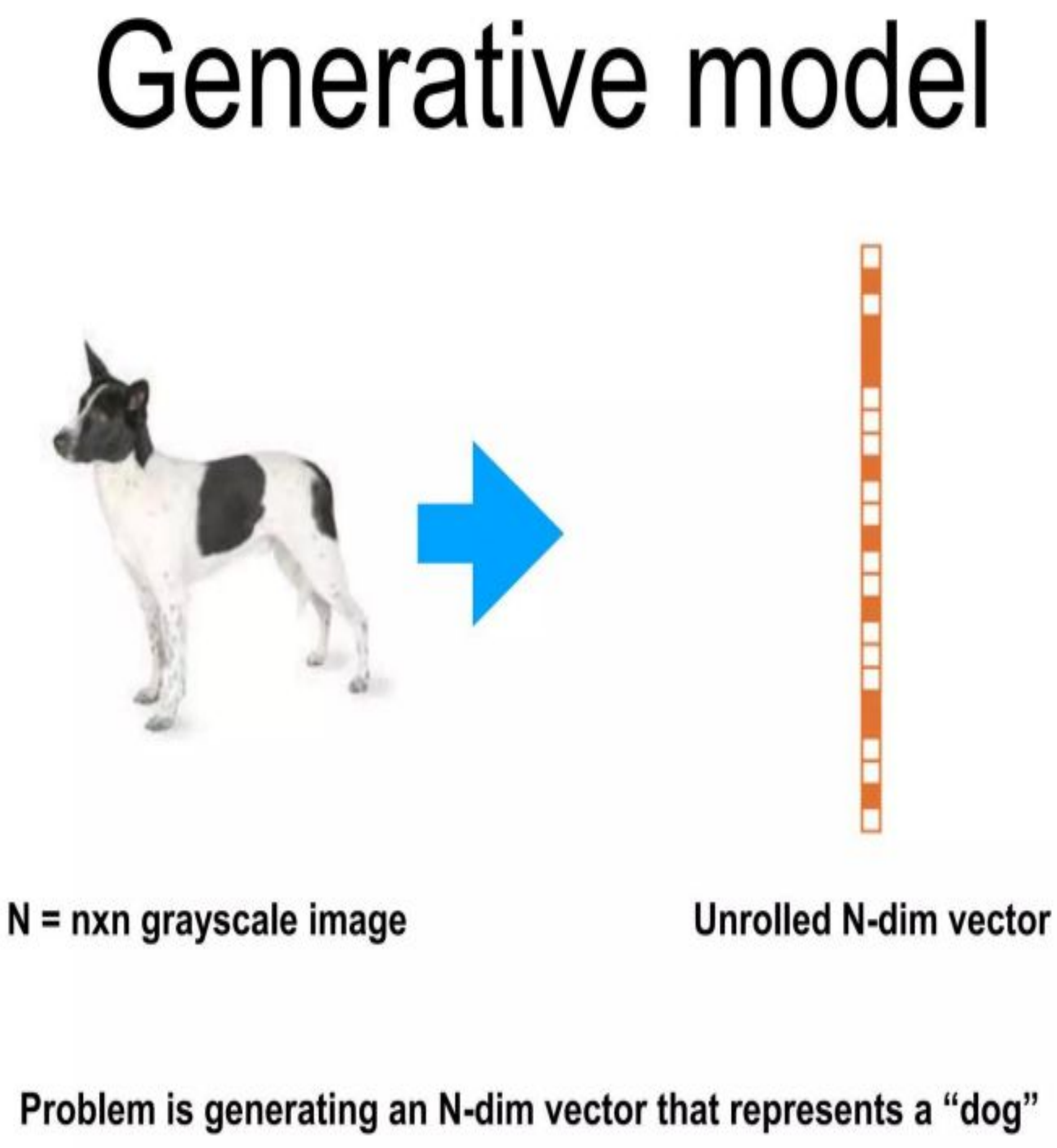
Types:

VAEs
Learn probabilistic mappings from data to latent space.

GANs
Adversarial training between generator and discriminator.

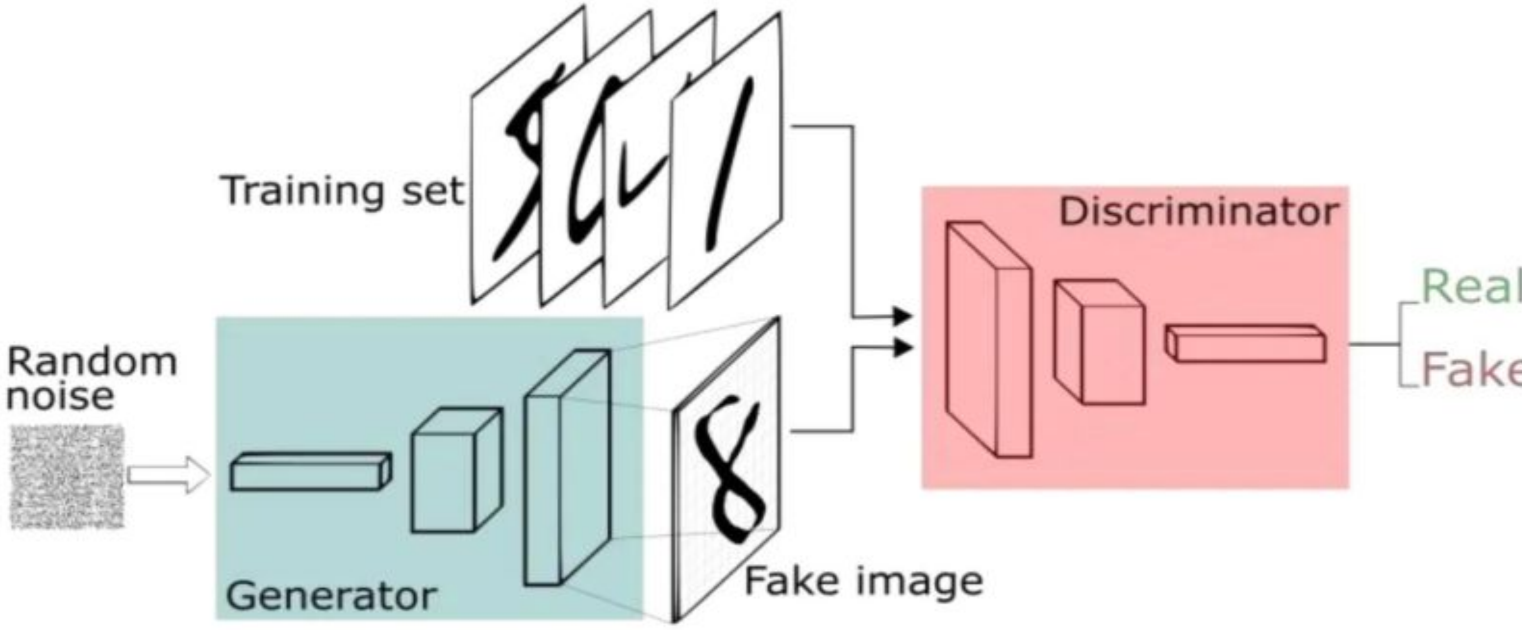
Diffusion Models:

Generate data by reversing a diffusion process

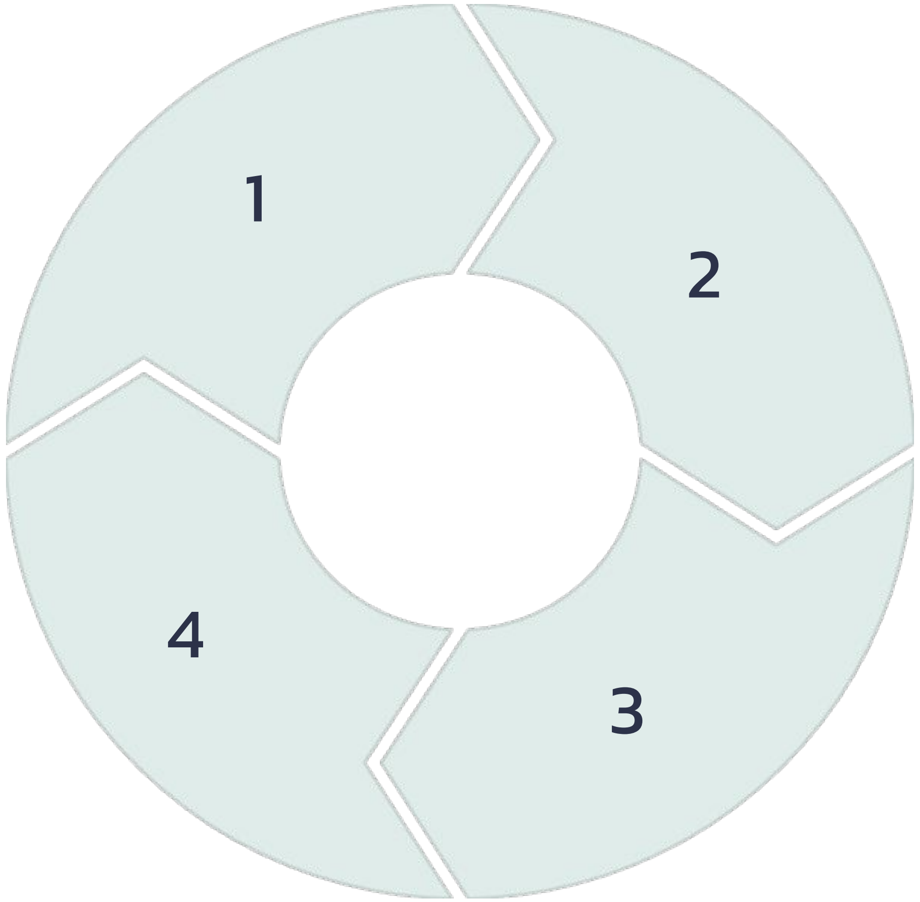


Understanding GANs

1. **Generator**
   Creates synthetic data mimicking real data.
2. **Discriminator**
   Distinguishes between real and generated data.
3. **Training**
   Generator produces data that fools the discriminator.



How GANs Work:



Generate

Synthetic data from noise.

Discriminator

Assesses whether the data is real (from the dataset) or fake (generated)

Adversarial Process:

Both networks improve iteratively;

Outcome

A generator capable of producing high-quality, realistic data.

How GANs Work:

Generator

generate new
instances of data

Discriminator

evaluates the authenticity
of the generated data

Discriminator Learning:

- Need to predict “1” for “true” images, and “0” for “fake” images

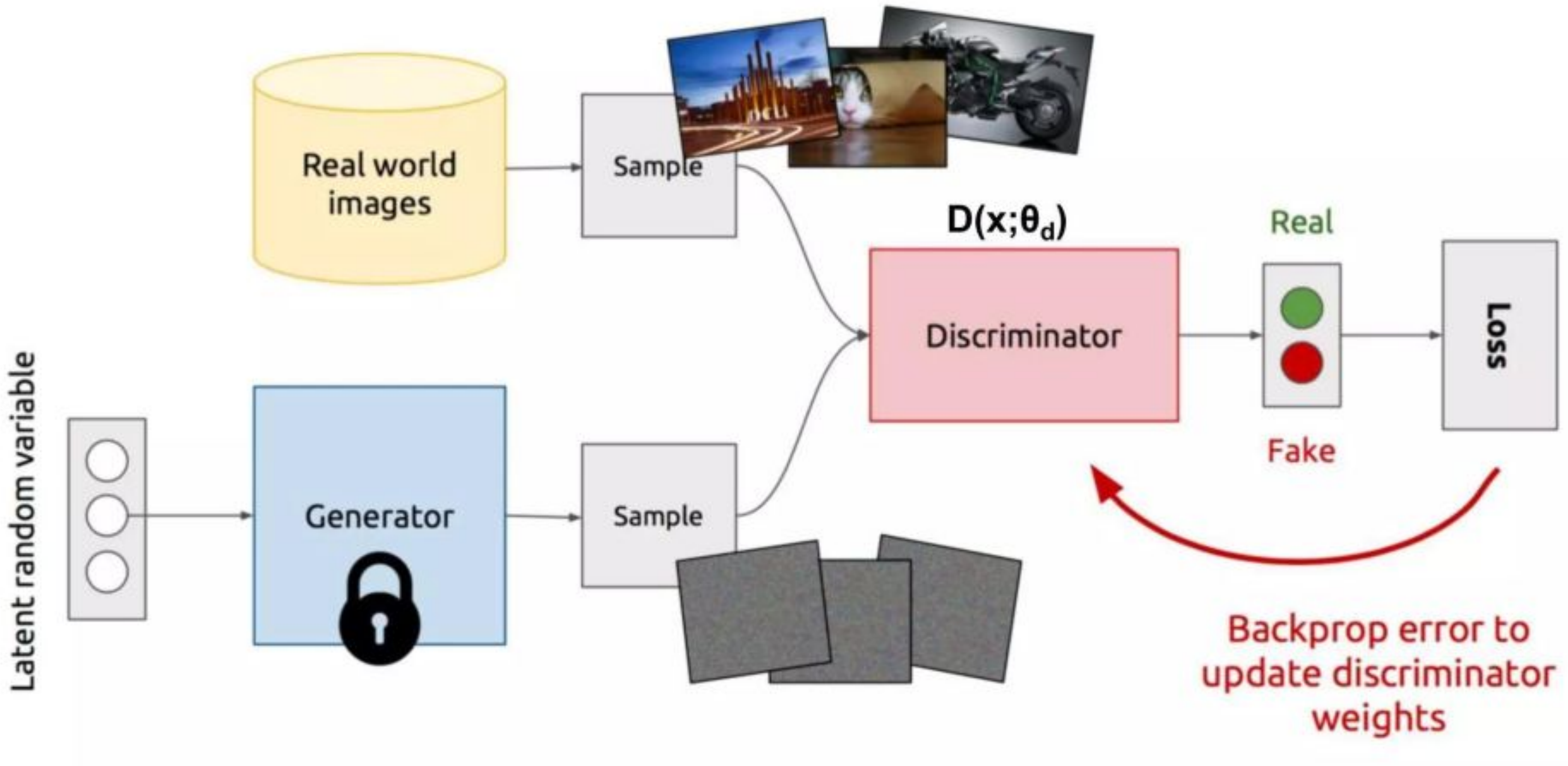
- So the loss function (negative of log loss) for the discriminator is

$$\textbf{V'} = \frac{1}{m}\sum_{i=1}^{m}\left[\log D\left(\boldsymbol{x}^{(i)}\right) + \log\left(1 - D\left(G\left(\boldsymbol{z}^{(i)}\right)\right)\right)\right]$$

- We want to maximize this loss function. In other words, perform **gradient ascent** as

$$\theta_d \leftarrow \theta_d + \eta \Delta V'(\theta_d)$$

Discriminator Learning:



Generator Learning:

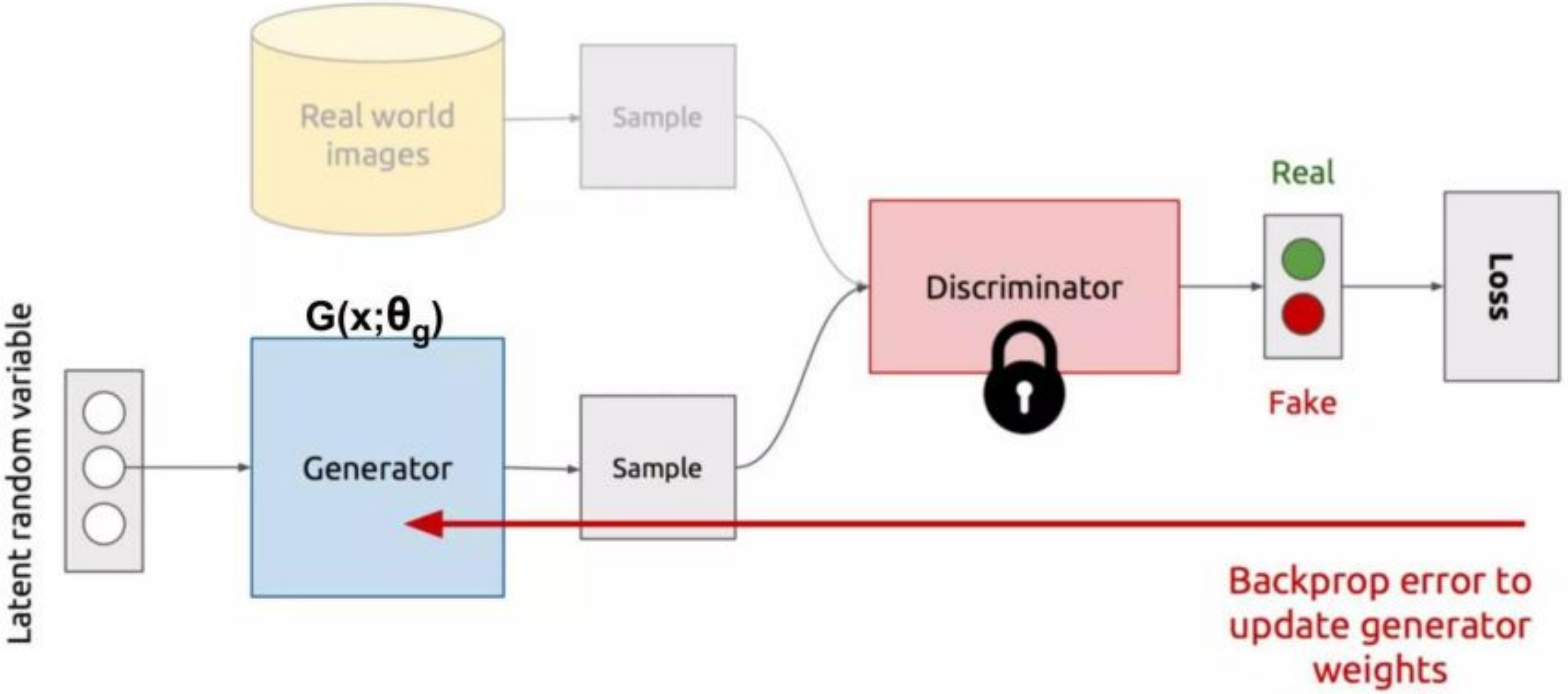
- Generator needs to fool the discriminator.
- It needs the discriminator to output “1” for “fake” images.
- So the loss function (log loss) for the generator is

$$V' = \frac{1}{m}\sum_{i=1}^{m}\log\left(1 - D\left(G\left(\boldsymbol{z}^{(i)}\right)\right)\right)$$

- We want to minimize this loss function. In other words, perform **gradient descent** as

$$\theta_g \leftarrow \theta_g - \eta \Delta V'(\theta_g)$$

Generator Learning:



Generator Learning:

GANs in a Nutshell:

GANs: Advantages & Challenges

Advantages

- High-quality data generation
- Versatility across various data types (images, text, etc.).

Challenges

- Training instability due to the adversarial setup.
- Mode collapse, where the generator produces limited variations.
- Difficulty in evaluating model performance objectively

Introduction to Diffusion Models

Concept

Generate data by learning to reverse a diffusion process that adds noise to data

Process:

1

Forward Diffusion

Gradually add noise to data, destroying structure.

2

Reverse Diffusion

Learn to reconstruct data by reversing the noise addition, generating new samples.

How Diffusion Models Work

1. **Data Preprocessing**

   Clean and prepare data for training.

2. **Forward Diffusion**

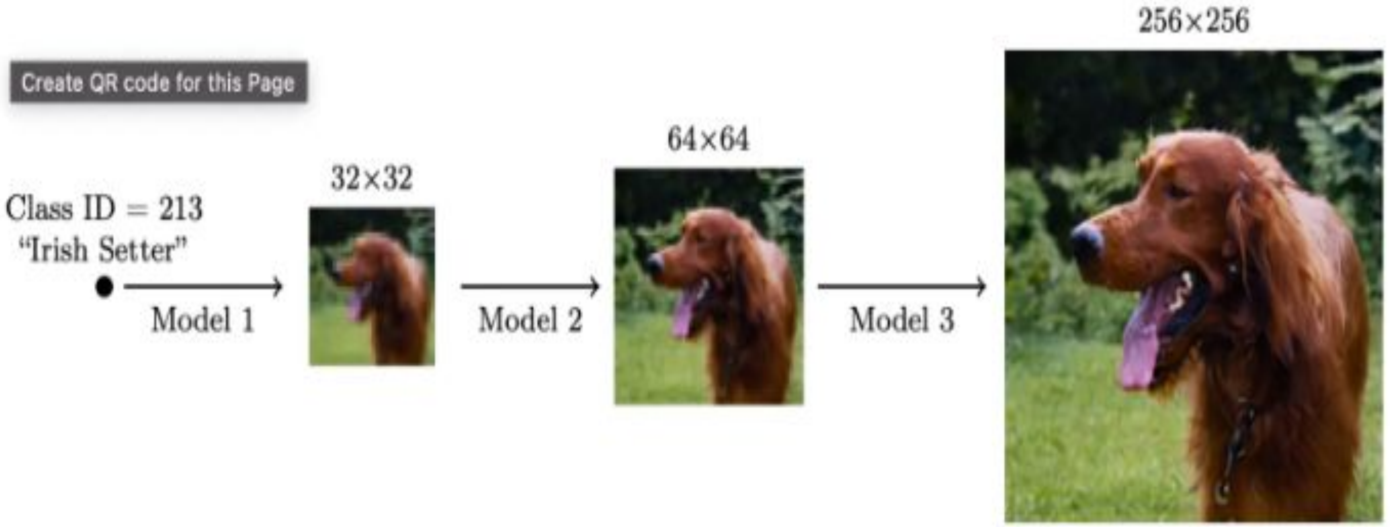
   Introduce noise incrementally to the data, creating a sequence of noisy data points.

3. **Reverse Diffusion**

   Train a model to reconstruct original data from noisy versions, effectively learning the data distribution.

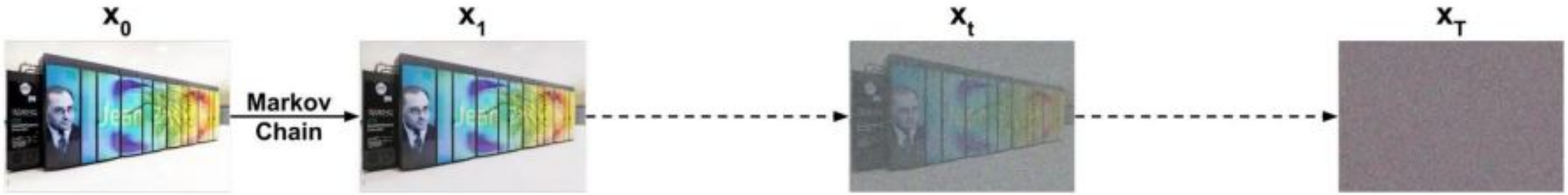
4. **Generation**

   Sampling from the learned distribution to produce new data instances.



A cascaded diffusion model comprising a base model and two super-resolution models.

Forward diffusion process:



This process will add noise to any image gradually

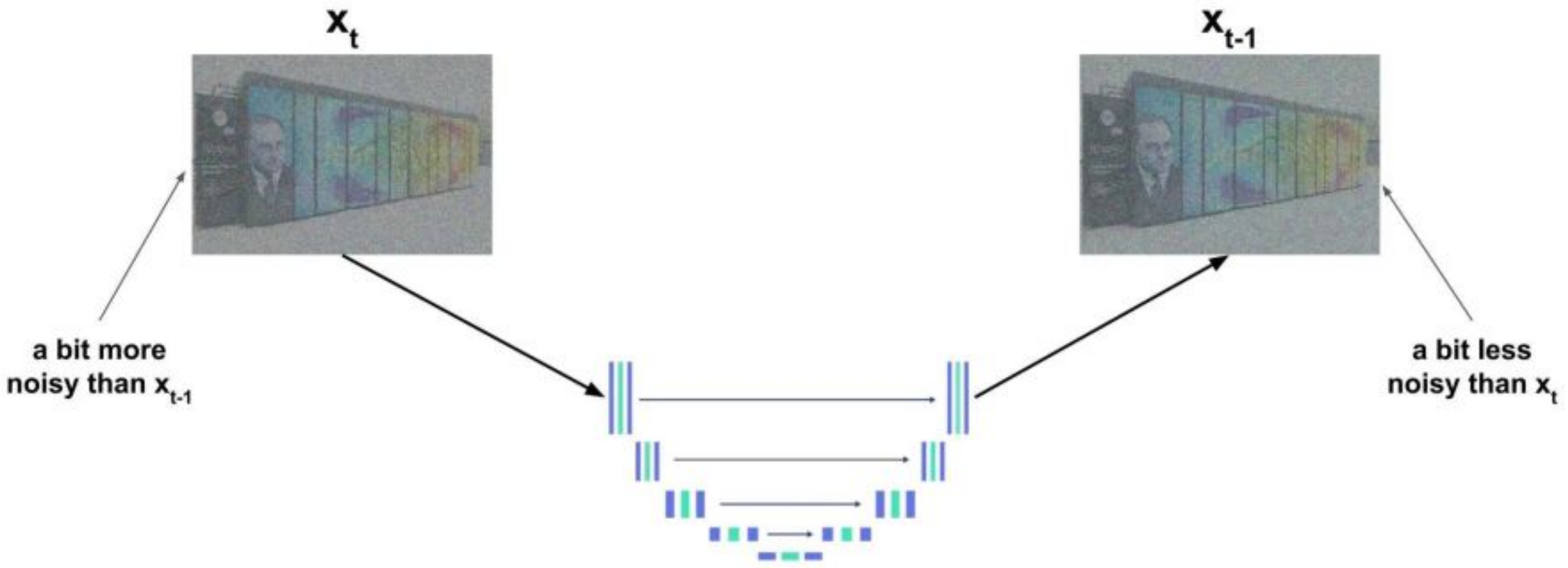
$0 \leq t \leq T$; T is a hyperparameter

Forward diffusion process:

Examples of images at different times t

Here we choose T=1000, but it can be different values (it's an hyperparameter)

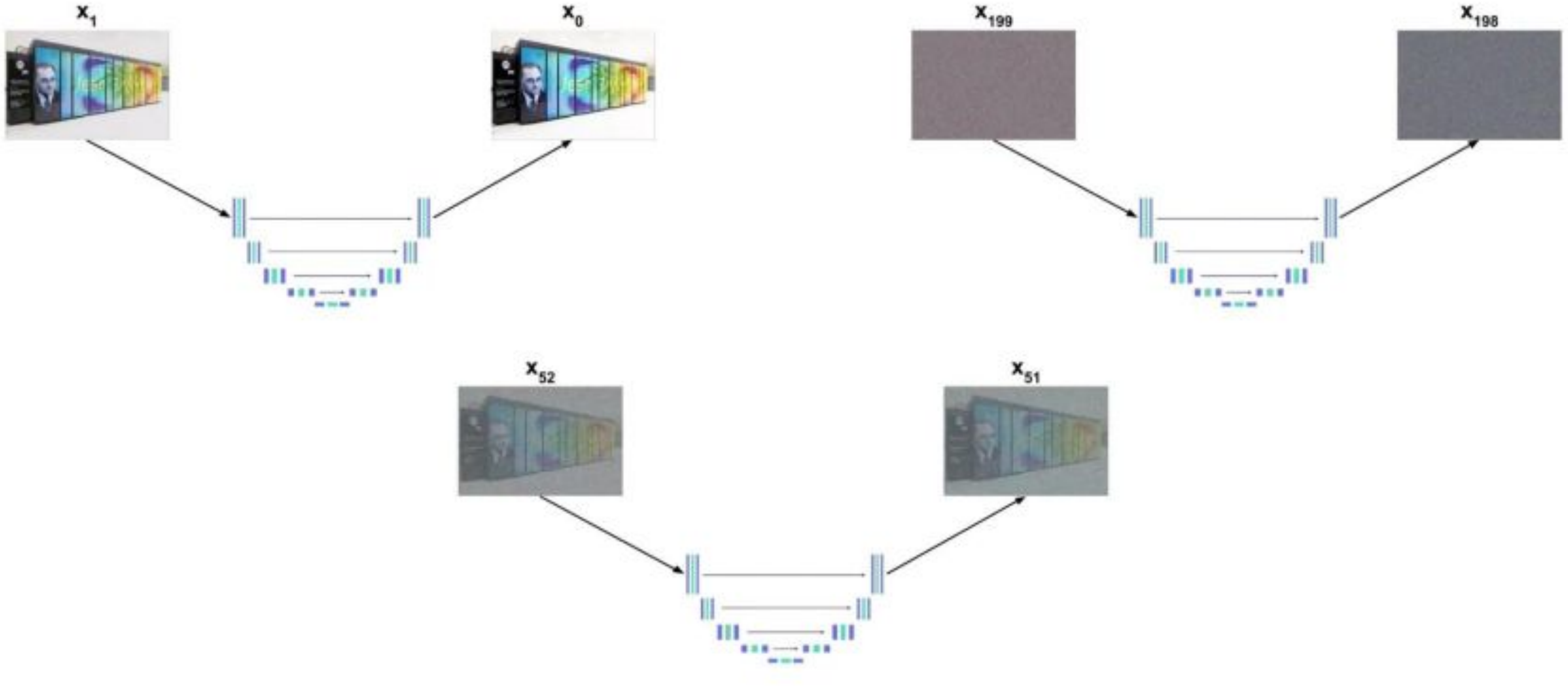
Reverse diffusion process:



We train a model to predict x_{t-1} from x_t

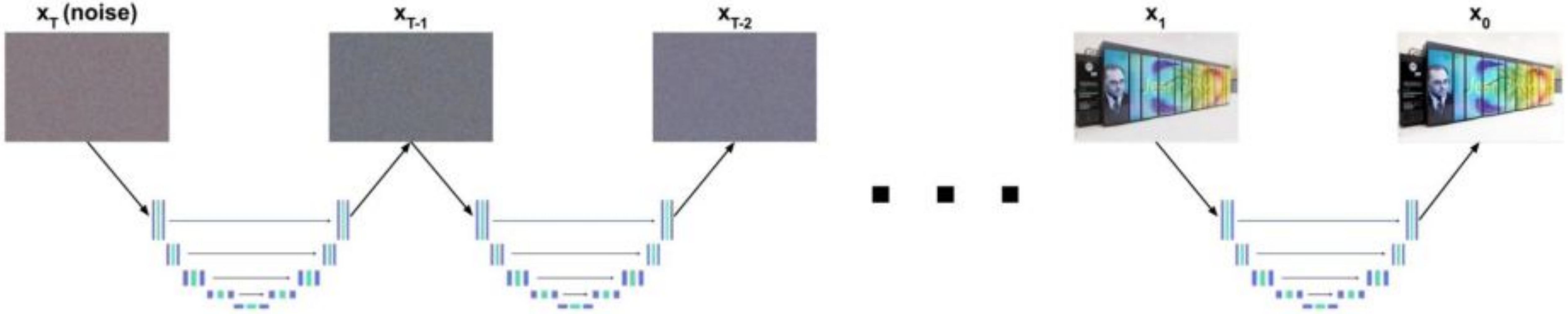
x_0 is any image from the dataset

Reverse diffusion process:



The same model must predict every x_{t-1} from x_t

Sampling process:

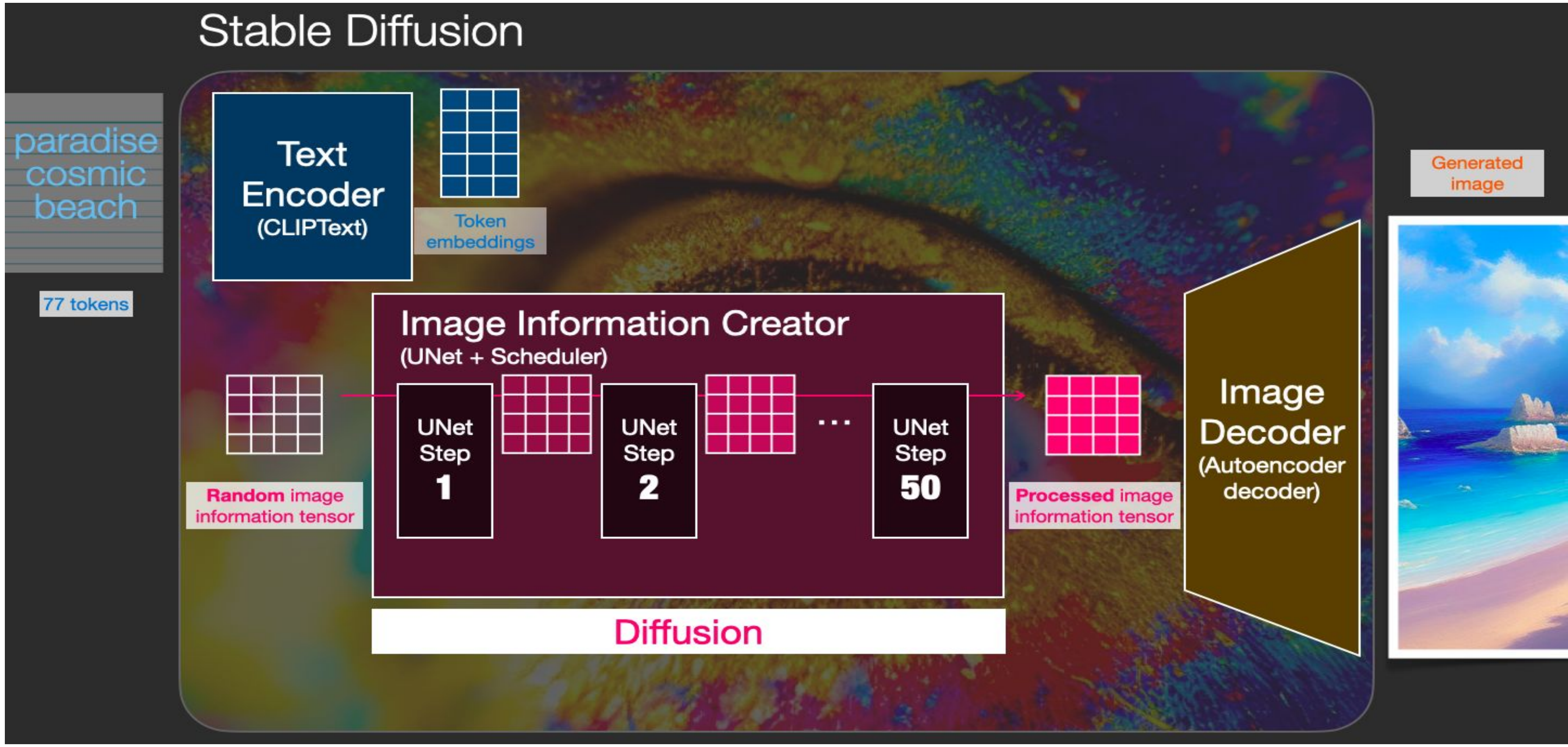


From a random noise we can generate an image

Stable Diffusion:

- Concept

A powerful generative model that creates images from text descriptions.



- Text-to-Image Generation:

Transforms natural language inputs into realistic images by reversing a diffusion process.

- Latent Space Representation

Operates in a compressed latent space to reduce computational complexity while maintaining high image quality.

Popular diffusion tools:

DALL-E 2:

Developed by OpenAI, DALL-E 2 is known for highly detailed and creative images from textual descriptions. It uses advanced diffusion techniques to produce images that are both imaginative and realistic, making it a popular tool in creative and artistic applications.

Popular diffusion tools:

DALL-E 3:

DALL-E 3 is the latest version of OpenAI image generation models and is a huge advancement over DALL-E 2. The most notable change is that this latest version isn't just an app but is integrated into ChatGPT. It also stands out with its image generation quality.

Popular diffusion tools:

SORA:

Sora's the latest model by OpenAI, and it's a game-changer. The AI community has been waiting for this drop since it's the first-ever text-to-video model by OpenAI. Sora can make 1080p videos in any resolution up to a minute long, and the videos it creates are scarily realistic.

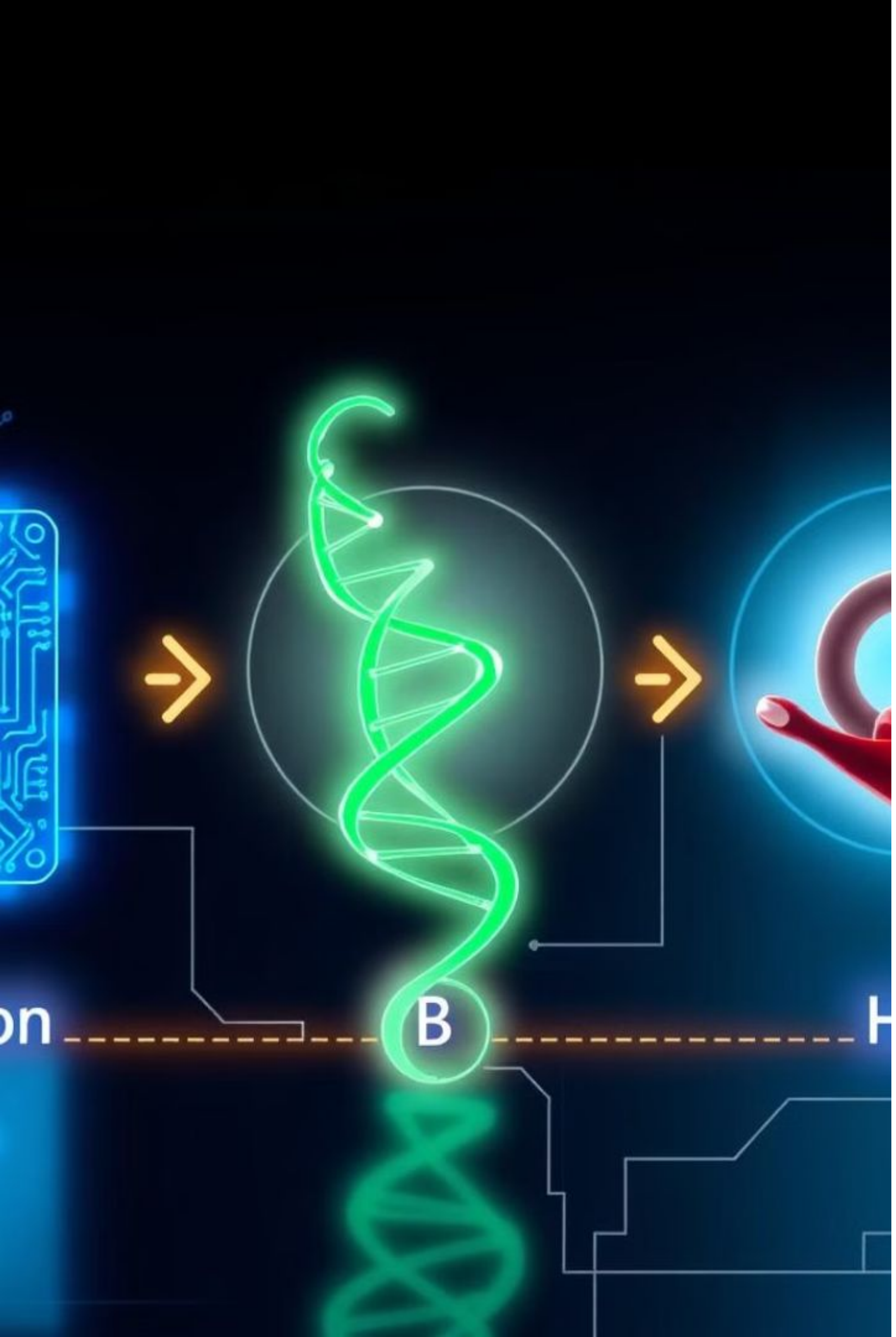
Applications of Diffusion Models

Image Generation

Audio Synthesis

Data Augmentation

Anomaly Detection



GANs vs. Diffusion Models

Training Stability	GANs: Prone to instability.	Diffusion: More stable.
Sample Quality	GANs: High quality, mode collapse.	Diffusion: Diverse, high-fidelity.
Computational Efficiency	GANs: Faster sampling.	Diffusion: Slower sampling.