### **Module 9 - Exercise**

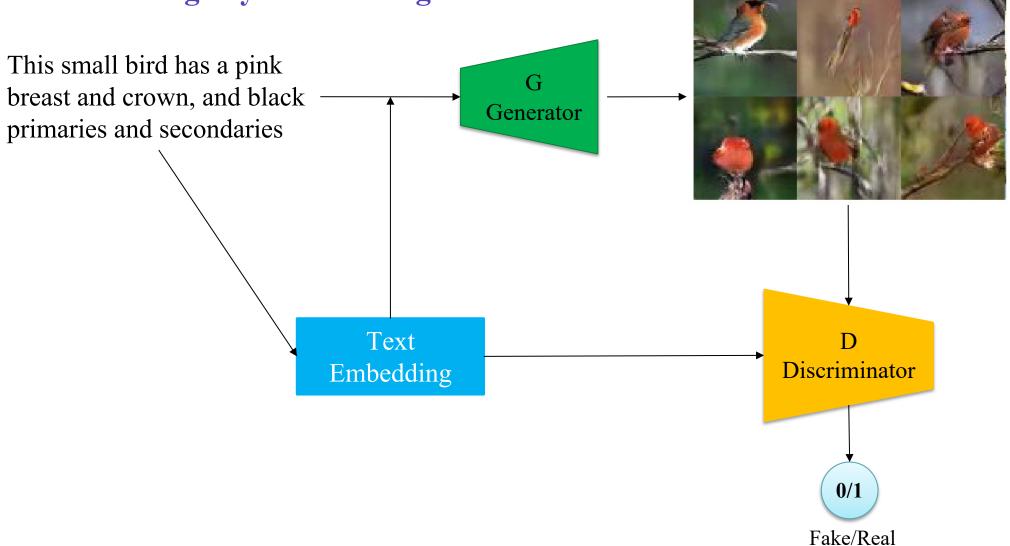
# Text to Image Synthesis using DCGAN

AI VIET NAM
Nguyen Quoc Thai



# **Objective**

### **Text to Image Synthesis using DCGAN**





# Outline

- > Introduction
- > DCGAN
- > Text to Image Synthesis using DCGAN



## Introduction



### **Text to Image**

This small bird has a pink breast and crown, and black primaries and secondaries





## Introduction



### **Text to Image**

The flower has petals that are bright pinkish purple with while stigma





### Introduction



### The milestones of text-to-image models and large models

2015.11 AlignDraw First TTI model using deep learning.

### 2016.5 GAN-CLS

First TTI model that achieves visually plausible result using GAN.

#### 2021.2 DALLE

First TTI model using autoregressive Transformer, with strong ability in generating zero-shot images.

### 2021.10 VO-Diffusion

First TTI model using Diffusion method, based on VQ-VAE.

### 2022.06 PARTI

A scaled-up, large autoregressive Transformer TTI model, with VQ-GAN structure for image processing, demonstrating toptier results in TTI generation.

### 2022.08 Stable Diffusion

An open-source latent diffusion model, widely used for content creation in research and commercial products, leveraging the capabilities of large models like BERT and CLIP.

2017.6 Transformer First model using pure-

attention architecture, opening up opportunities for large vision/language models.

### 2018.10 BERT

A pretrained Transformer renowned for remarkable text encoding capability, widely employed as a text encoder in TTI models.

#### 2021.2 CLIP

First languagevision model that learns the modality between language and image features.

### 2021.6 VQ-GAN

A model that merges the efficiency of GAN with the expressive power of Transformer based on VQ-VAE, extensively employed as an image encoder and decoder for TTI models. 2022.10 ChatGPT

A conversational large language model for dialogue system and question-answering, which could be potentially integrated with TTI models to leverage its strong prompt generating ability.



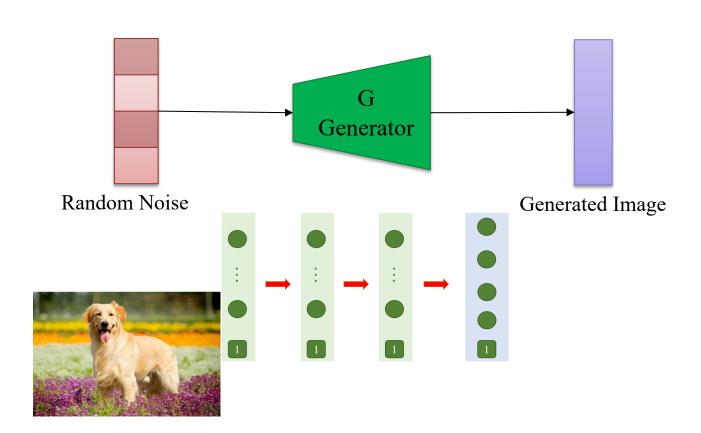
# Outline

- > Introduction
- > DCGAN
- > Text to Image Synthesis using DCGAN





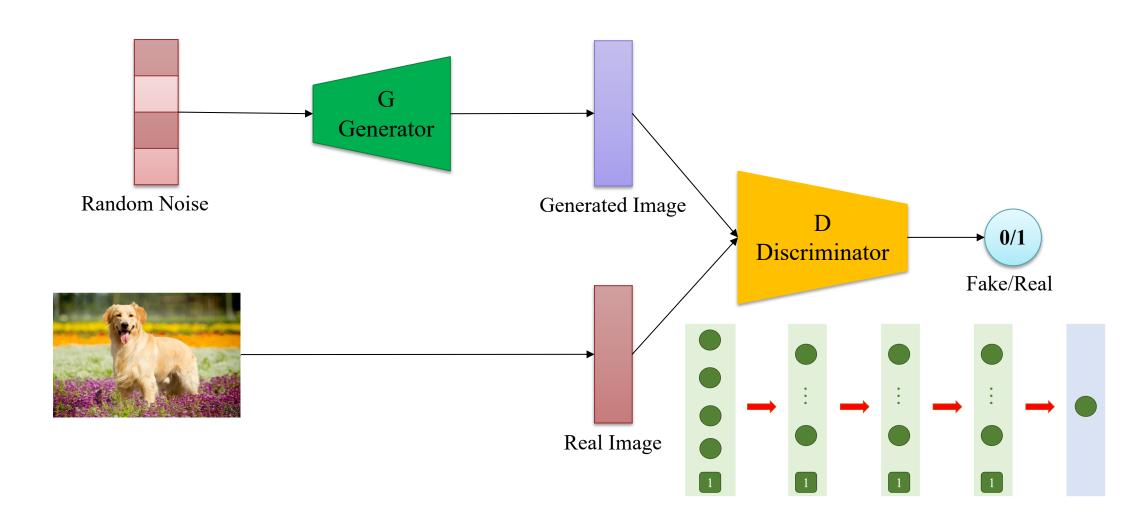
### **GAN** (Generative Adversarial Networks)



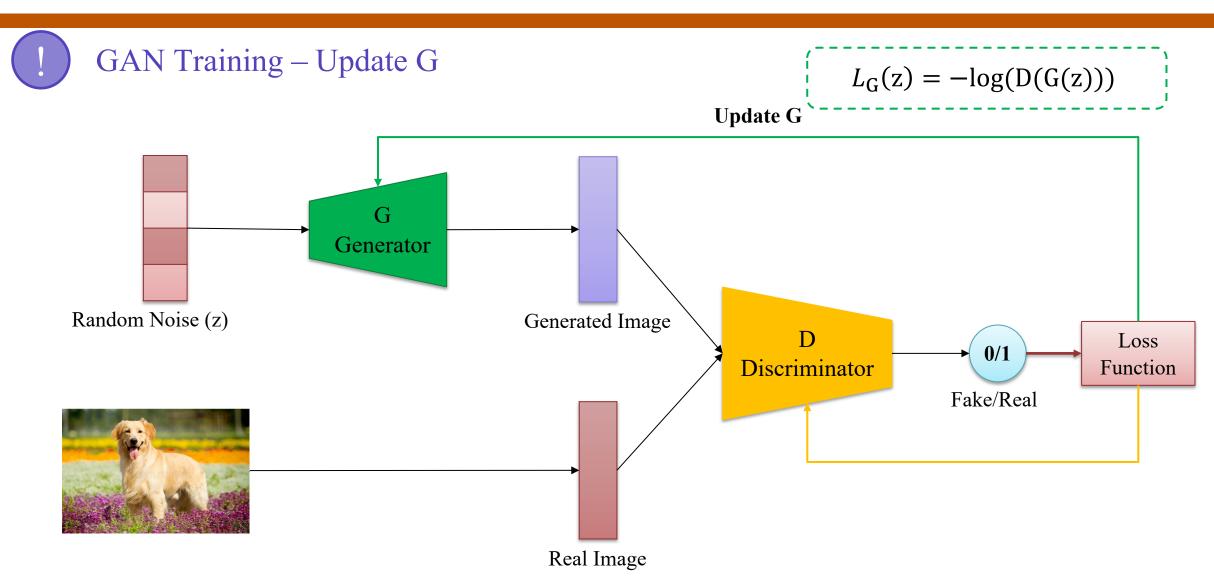




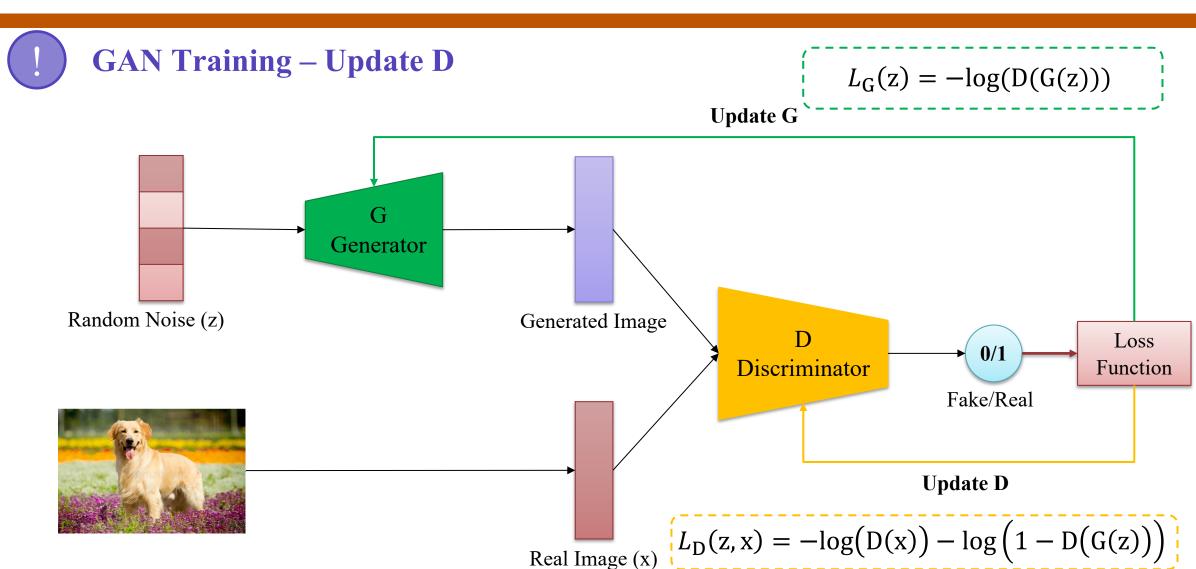
### **GAN** (Generative Adversarial Networks)







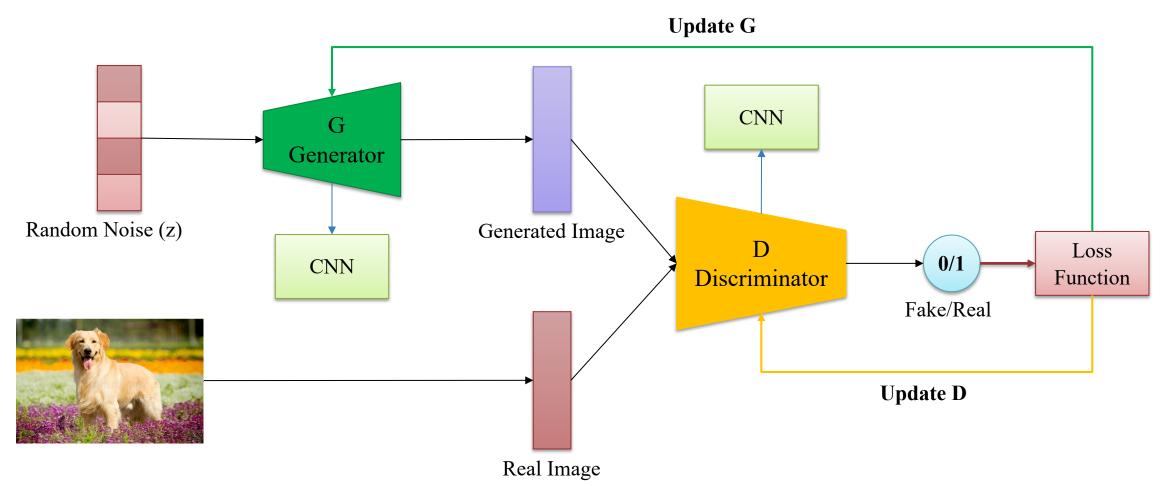




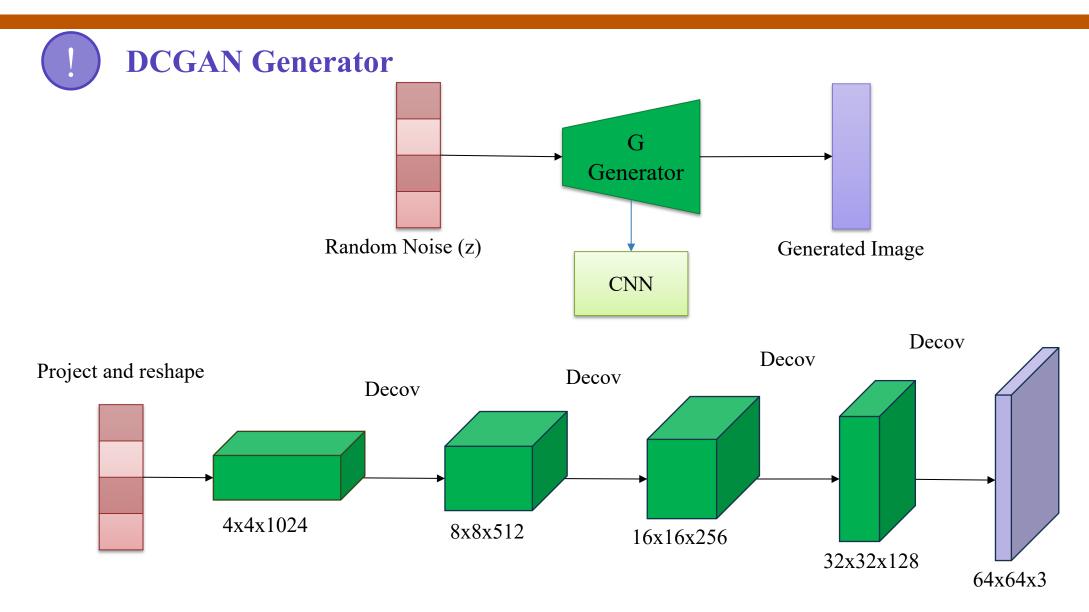




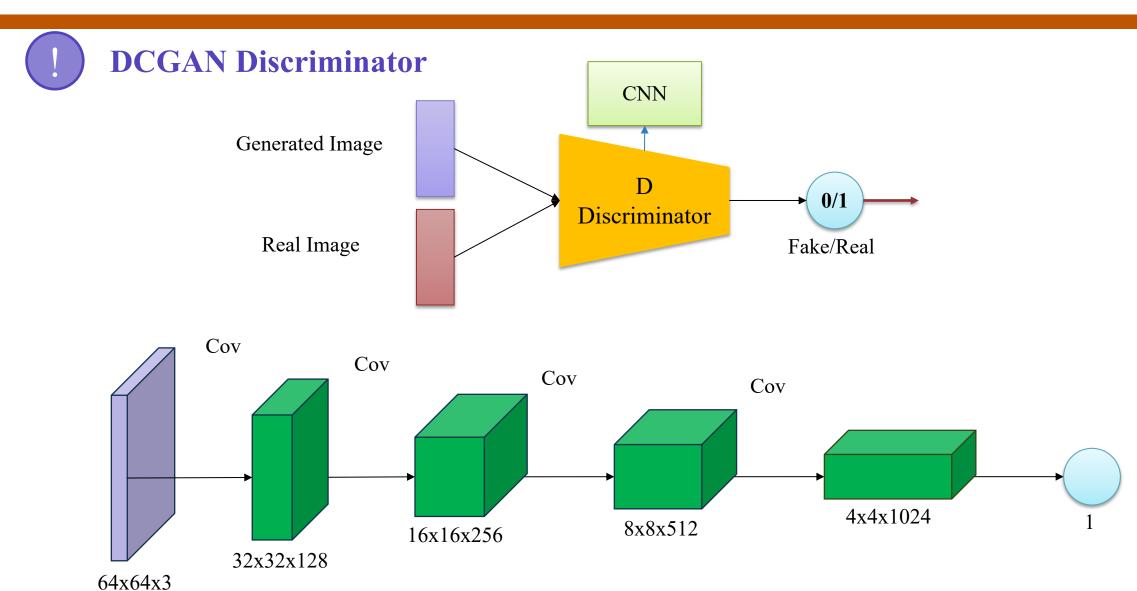
### **Deep Convolution GAN (DCGAN)**













# Outline

- > Introduction
- > DCGAN
- > Text to Image Synthesis using DCGAN





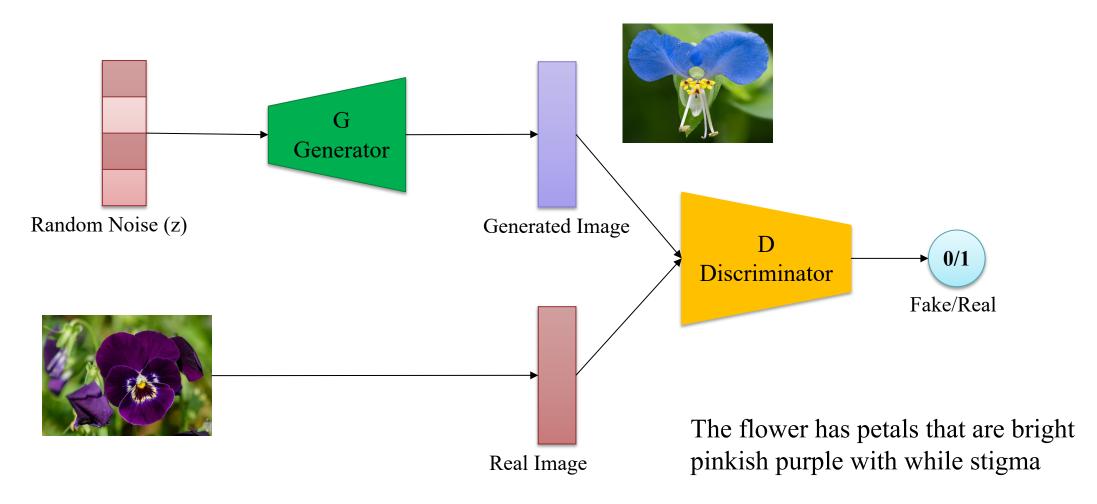
The flower has petals that are bright pinkish purple with while stigma







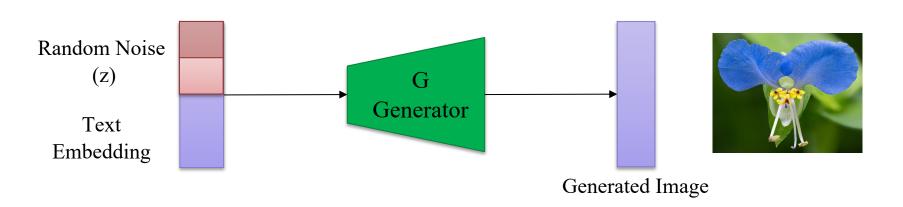
### **Approach**

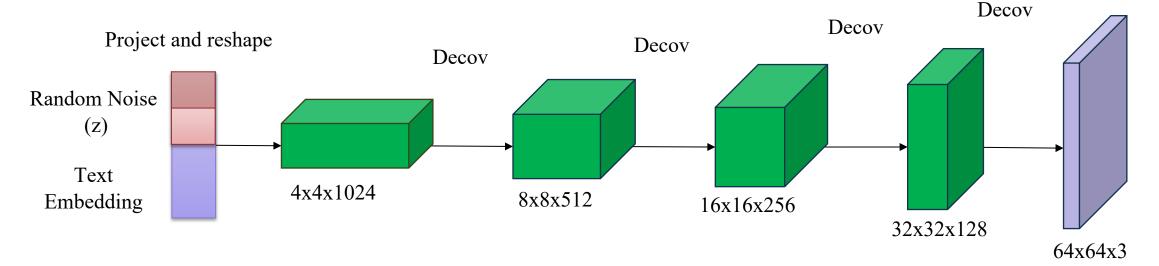






### **GANs with Join Distributions**

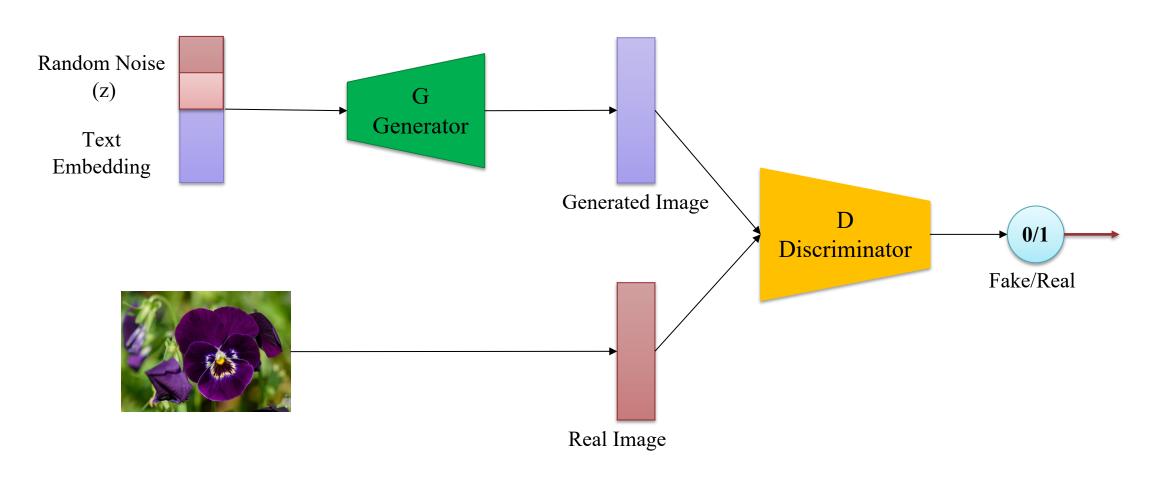




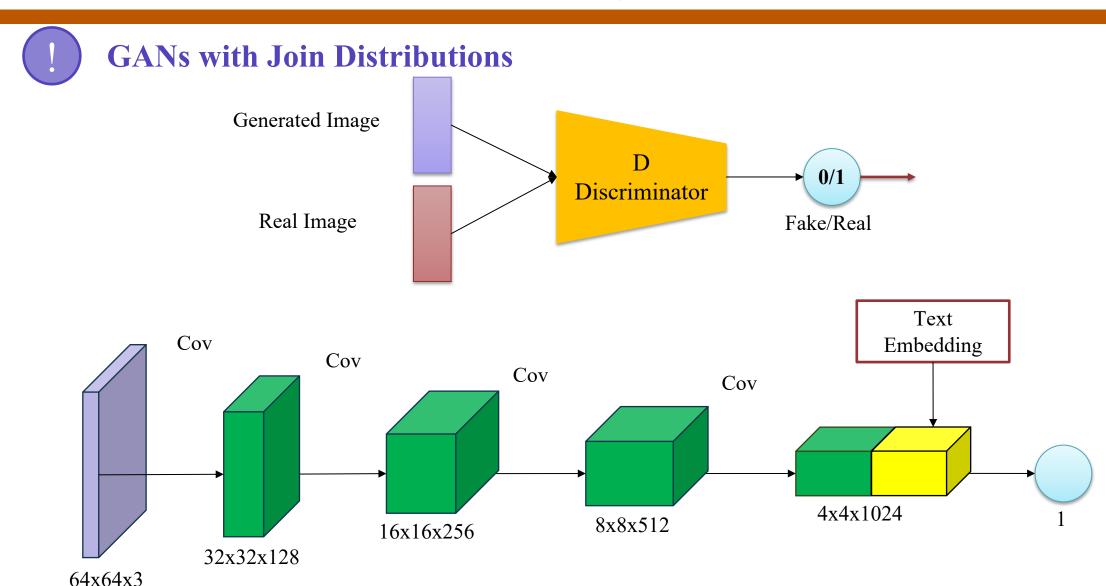




### **GANs with Join Distributions**





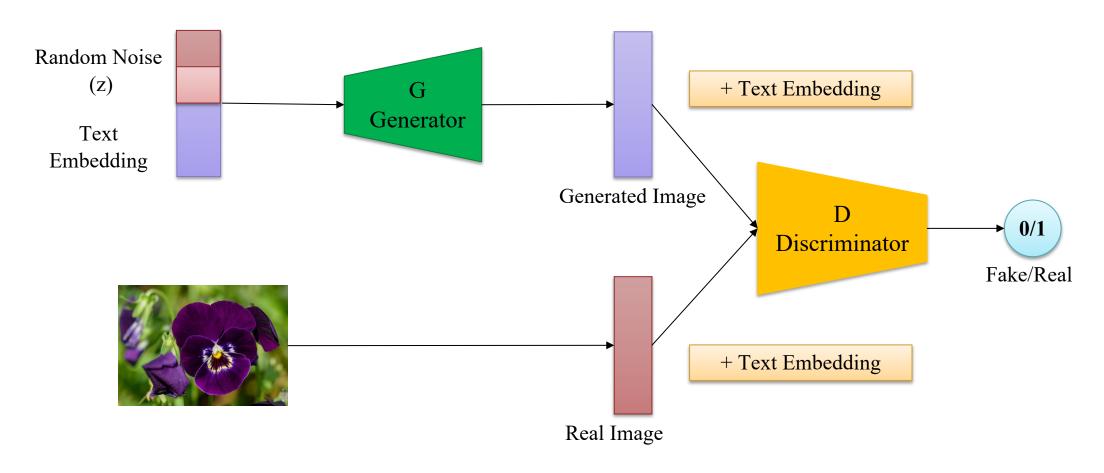






### **GANs with Join Distributions**

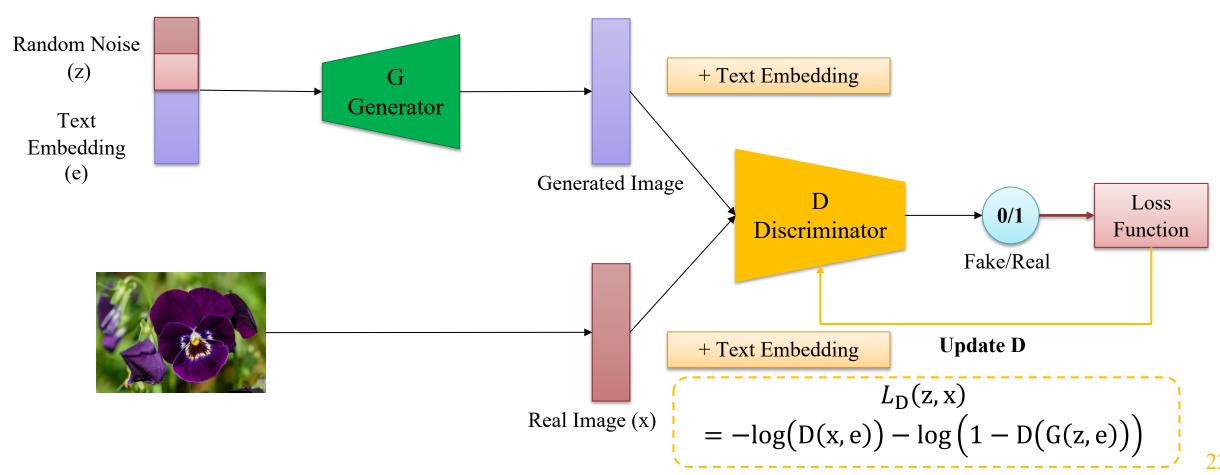
The flower has petals that are bright pinkish purple with while stigma







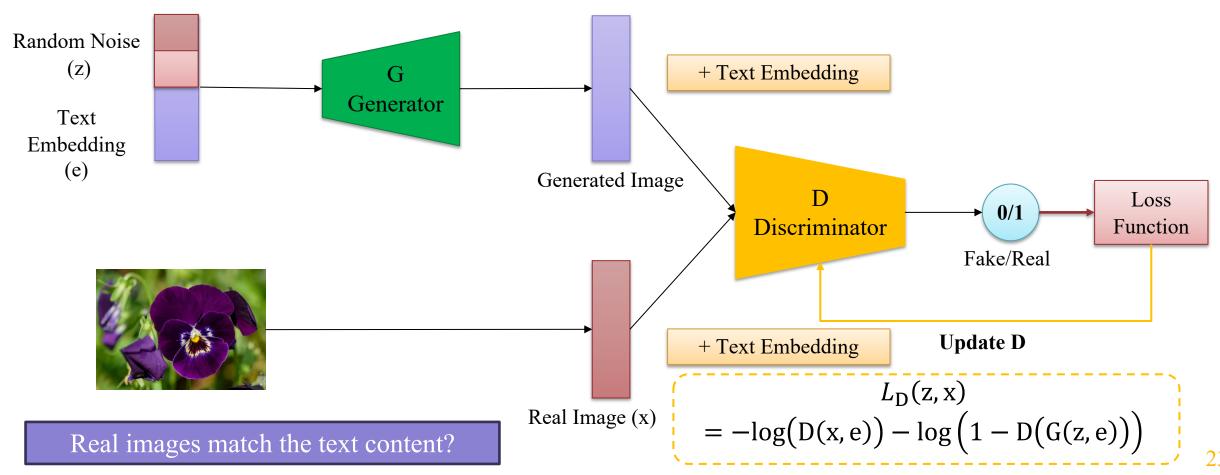
### **GANs with Join Distributions**



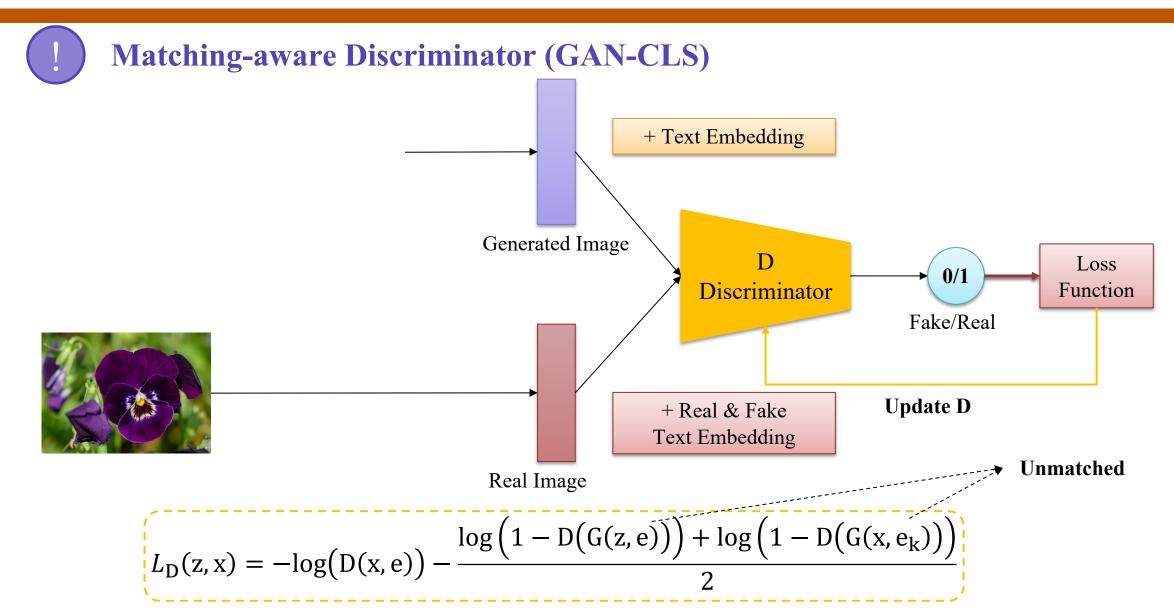




### **GANs with Join Distributions**



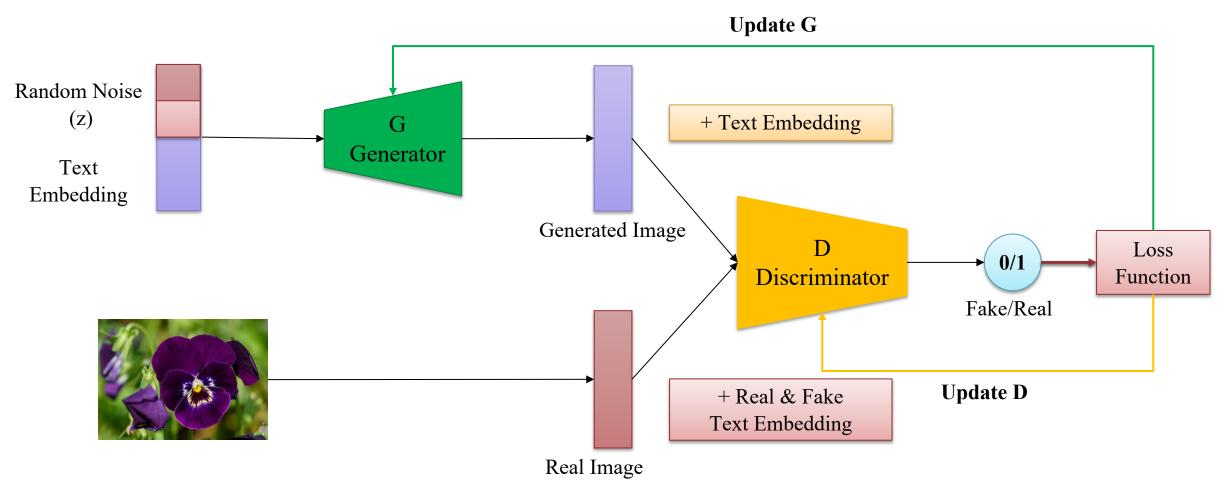




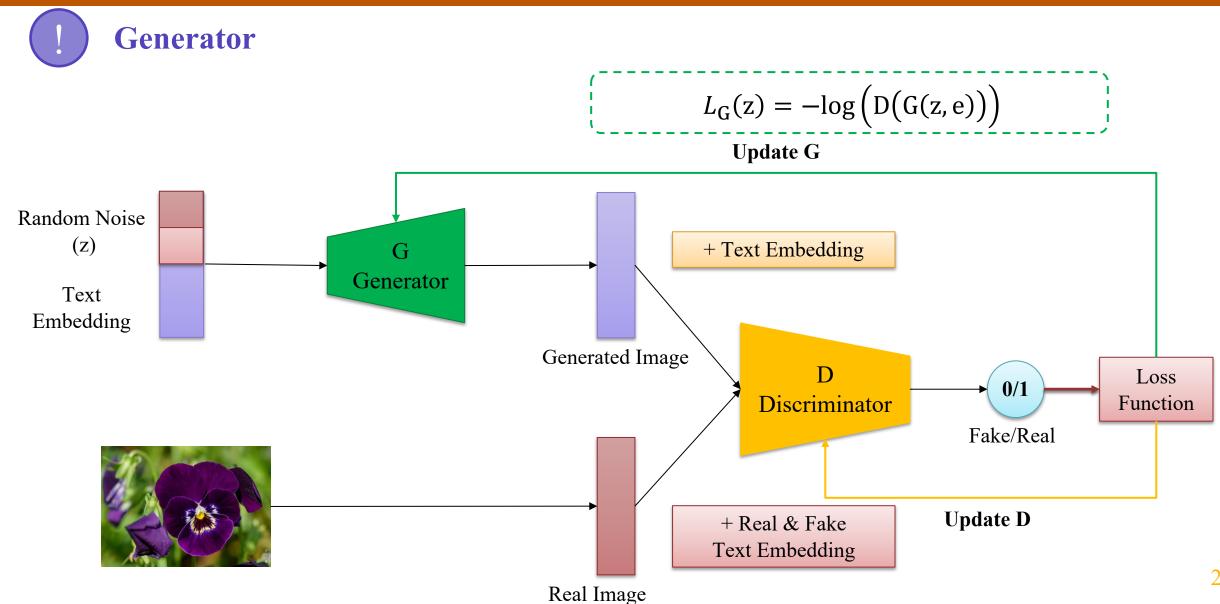




### **Matching-aware Discriminator (GAN-CLS)**



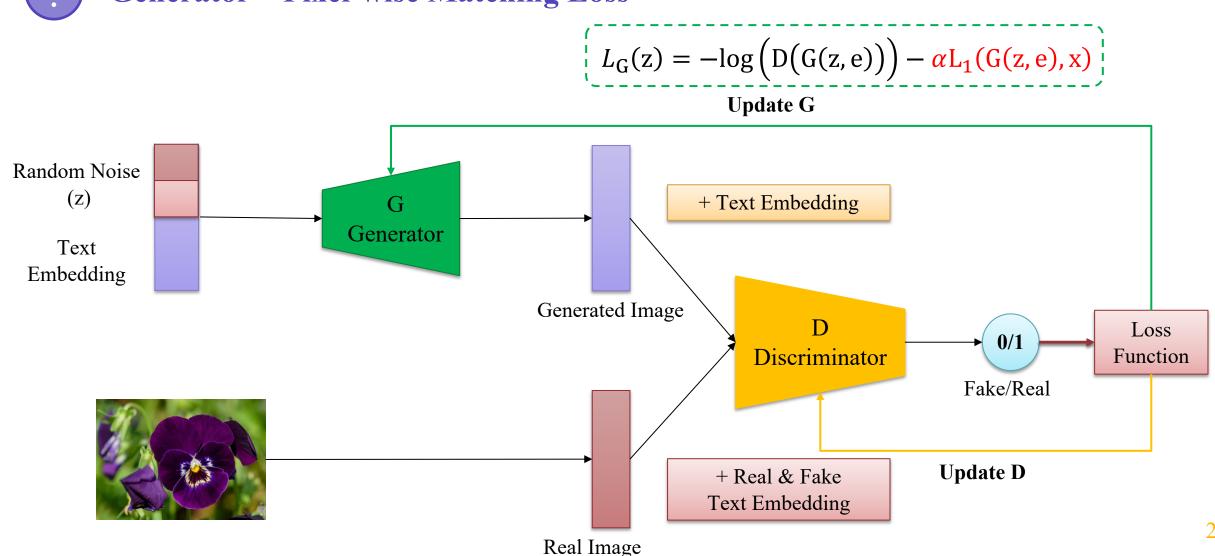








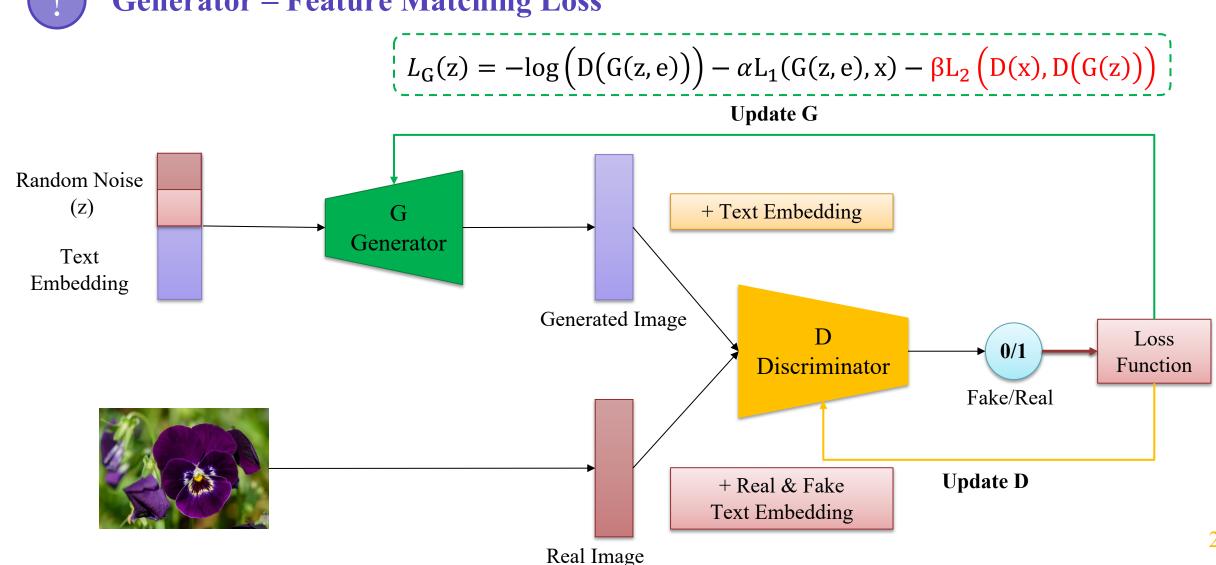
### **Generator – Pixel-wise Matching Loss**





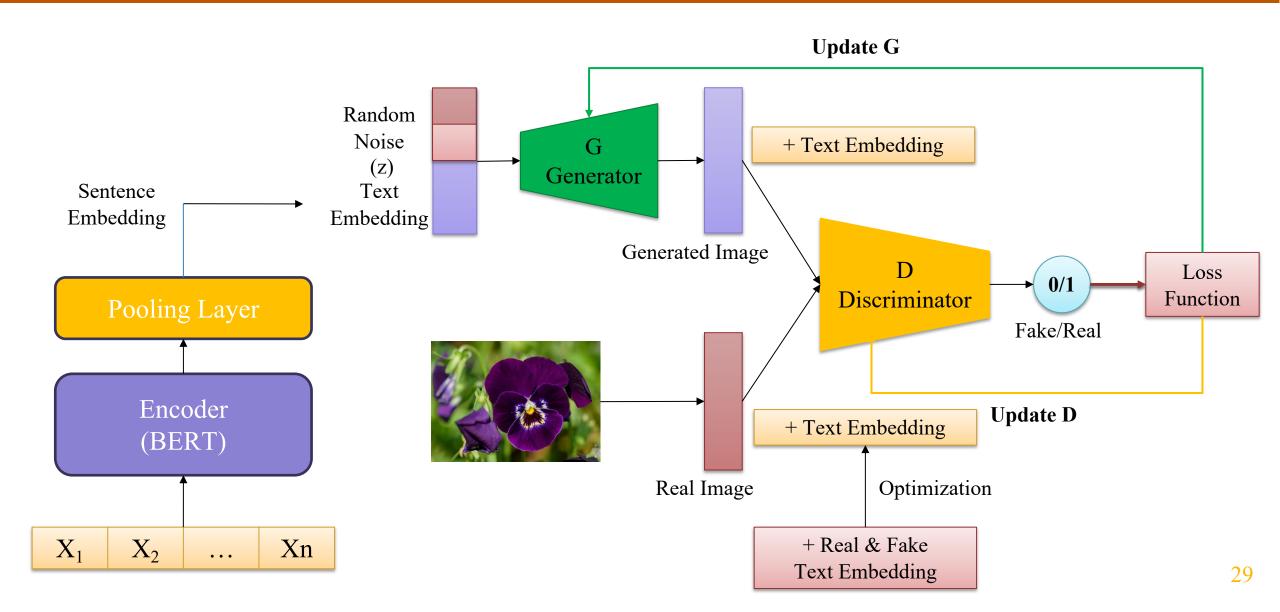


### **Generator – Feature Matching Loss**





# Summary





# Thanks!

Any questions?