Image generation model comparison

Creating Art

THEN

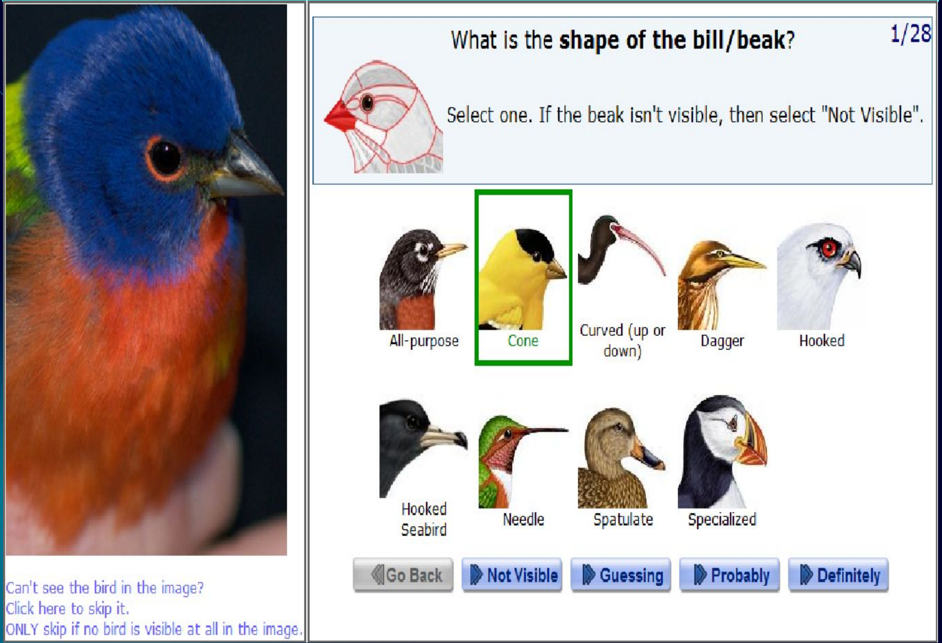
NOW

AI Artists
DALL·E 2
haha art generation go brrrrrrrrr

Problem Statement

Which deep learning model is the best for image generation?

CUB 200-2011

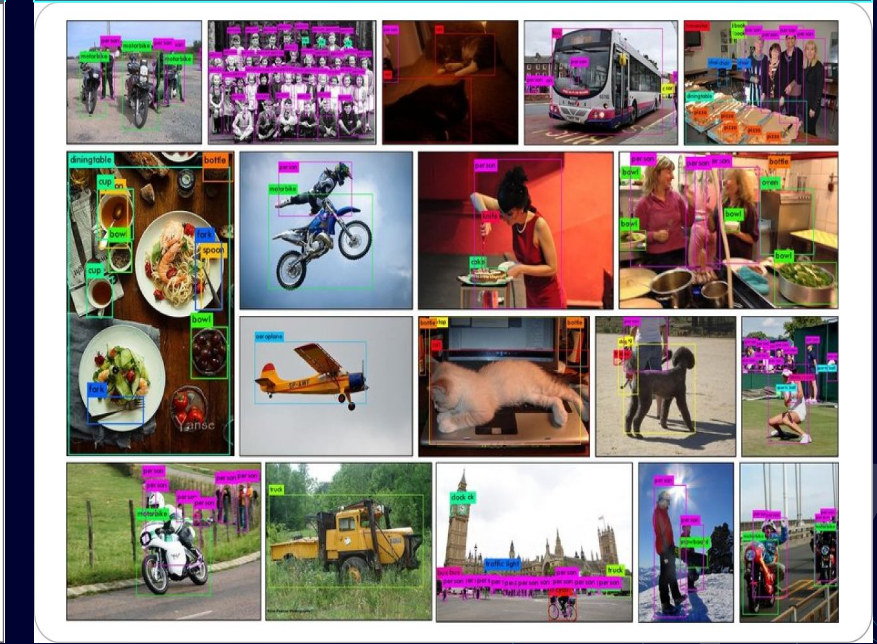


What is the **shape of the bill/beak?** 1/28

Select one. If the beak isn't visible, then select "Not Visible".

Can't see the bird in the image?
Click here to skip it.
ONLY skip if no bird is visible at all in the image.

COCO



taoxugit/AttnGAN

2 Contributors

78 Issues

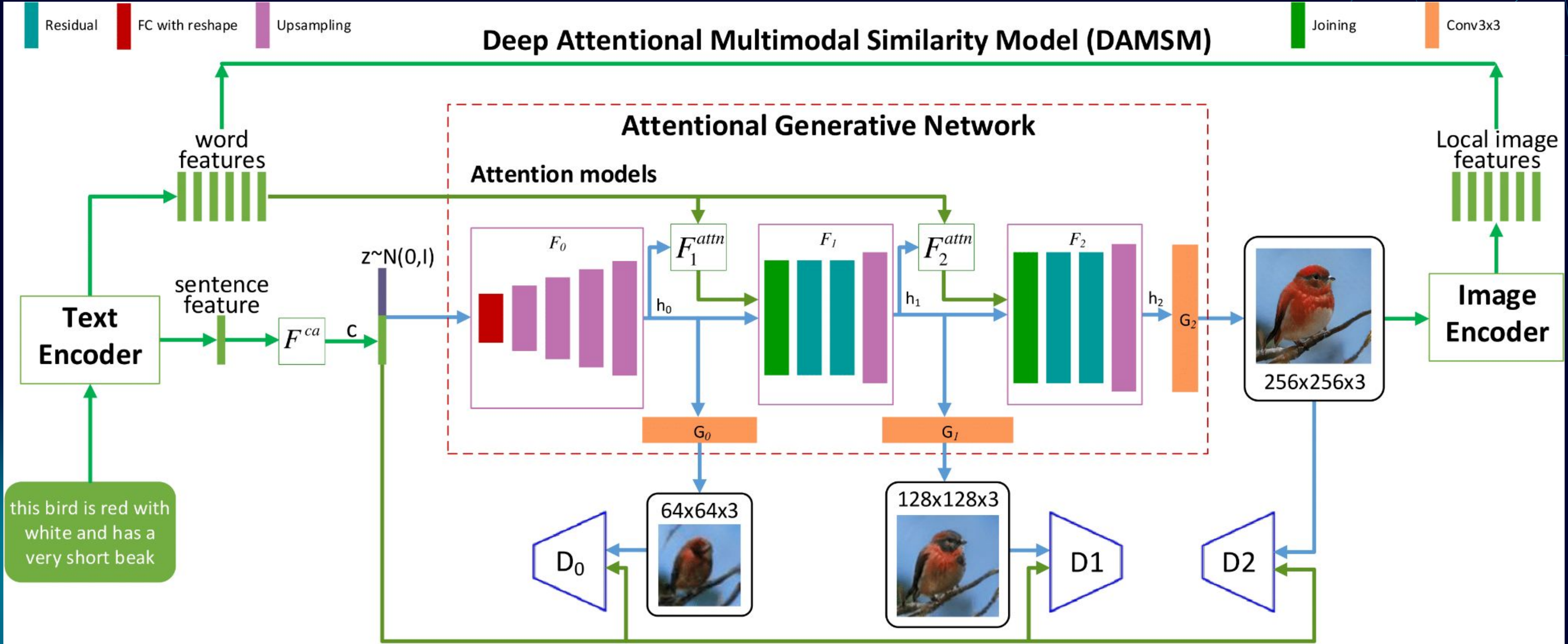
1k Stars

414 Forks

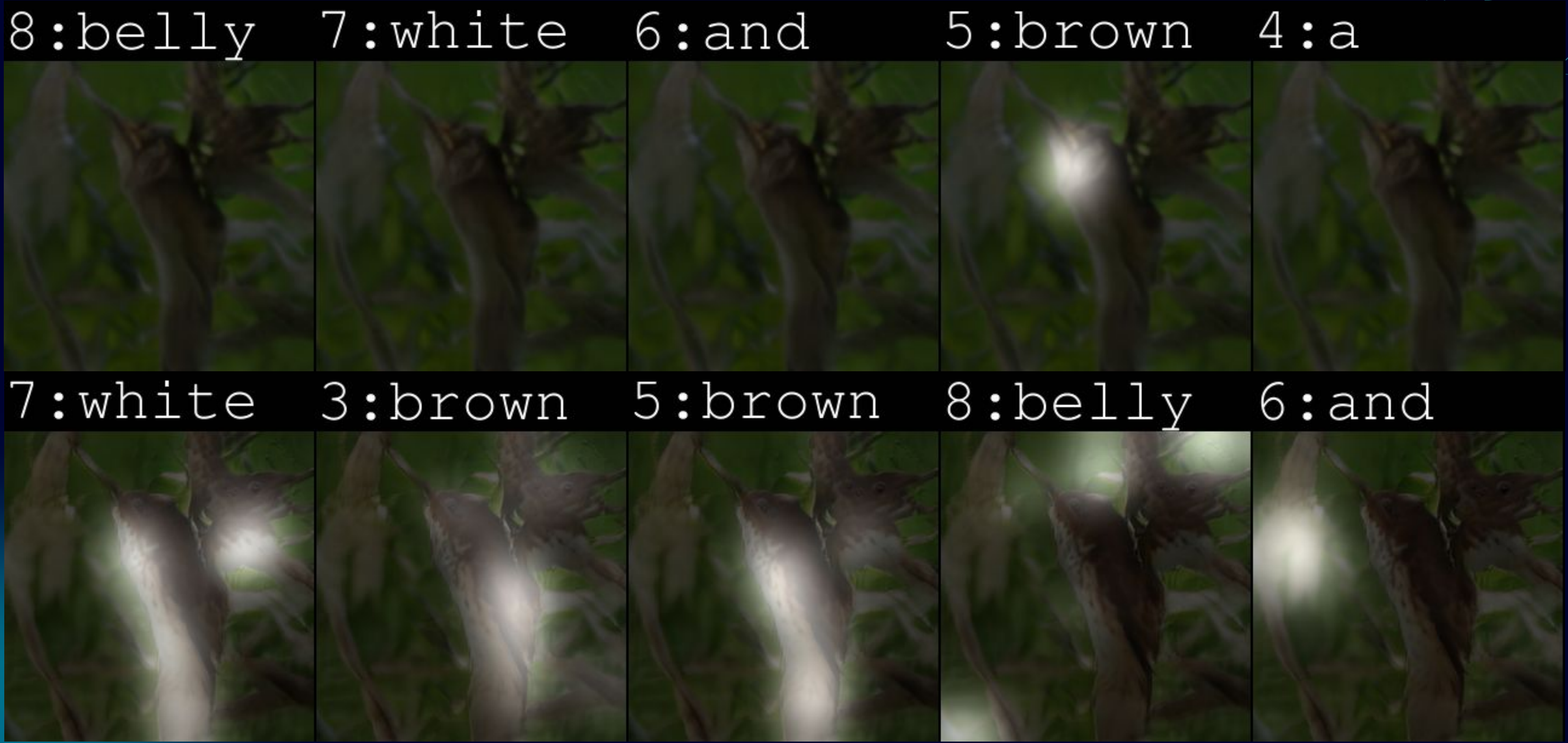
01

AttnGAN-Attentional Generative Adversarial Networks

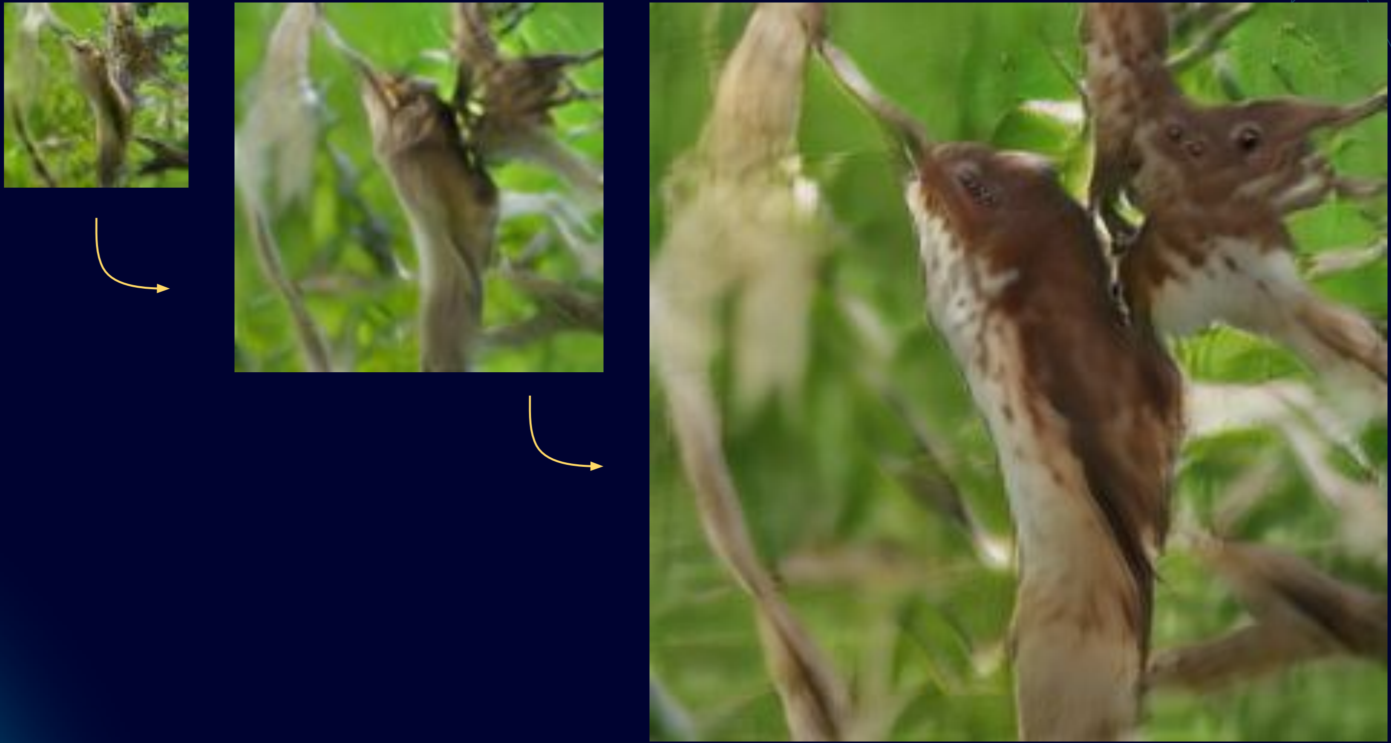
How the model work?



Match keywords with sub-region



Upscaling



ADVANTAGE

01 Produce detailed result

02 Fine-Grained Control

03 Improve text-image matching

tobran/**DF-GAN**

A Simple and Effective Baseline for Text-to-Image Synthesis (CVPR2022 oral)

2 Contributors

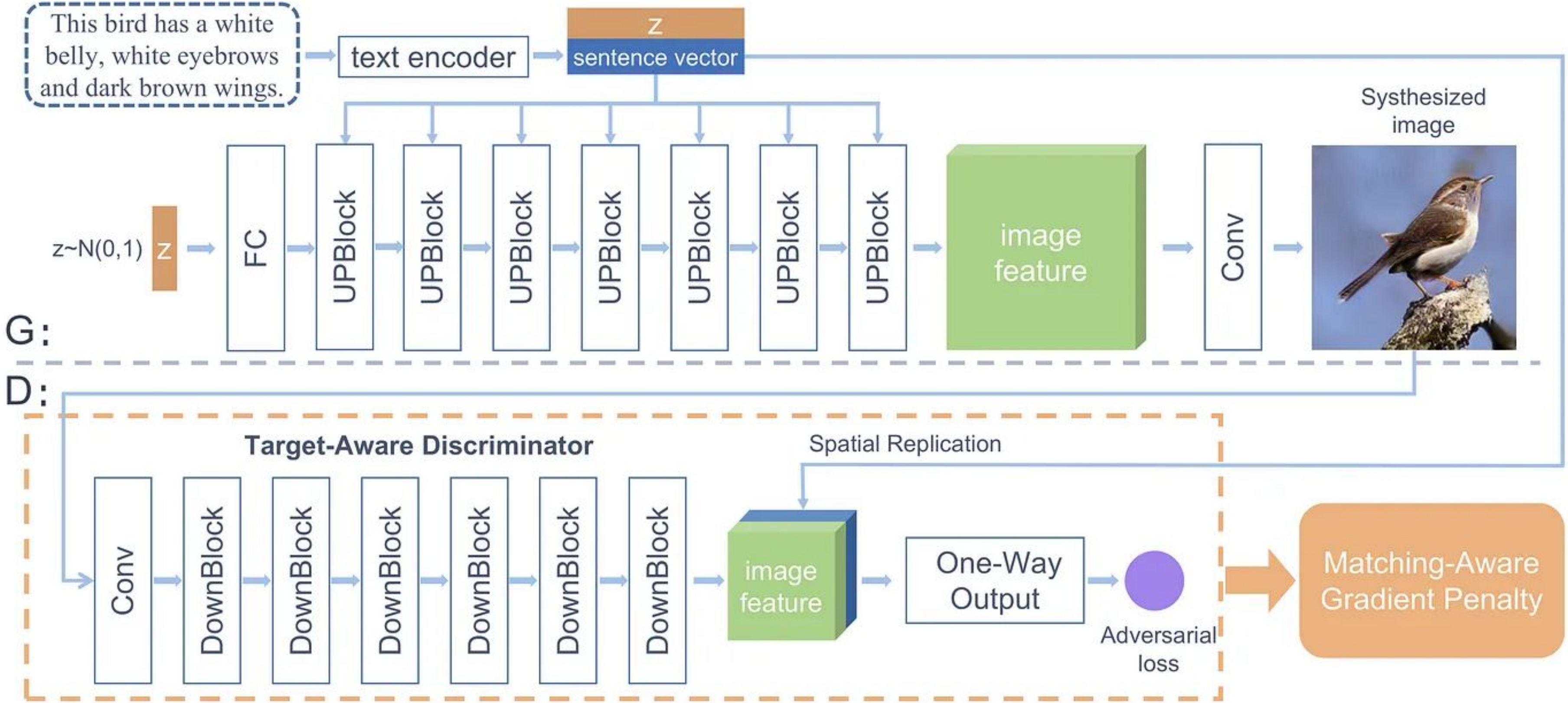
13 Issues

281 Stars

67 Forks

02

DF-GAN: Deep Fusion Generative Adversarial Networks



This bird has a white belly, white eyebrows and dark brown wings.

text encoder

z

sentence vector

Systhesized image

z~N(0,1)

z

FC

UPBlock

UPBlock

UPBlock

UPBlock

UPBlock

UPBlock

UPBlock

image feature

Conv

G:

D:

Target-Aware Discriminator

Spatial Replication

Conv

DownBlock

DownBlock

DownBlock

DownBlock

DownBlock

DownBlock

image feature

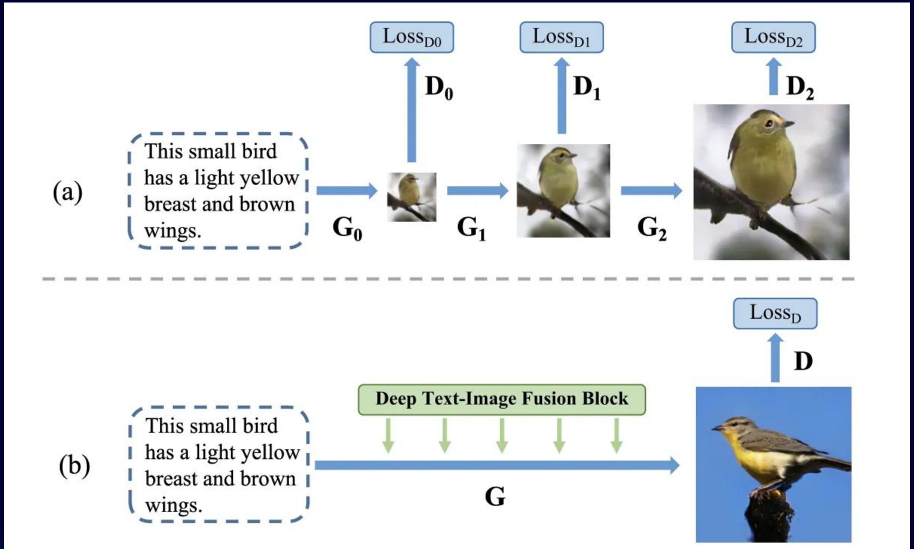
One-Way Output

Adversarial loss

Matching-Aware Gradient Penalty

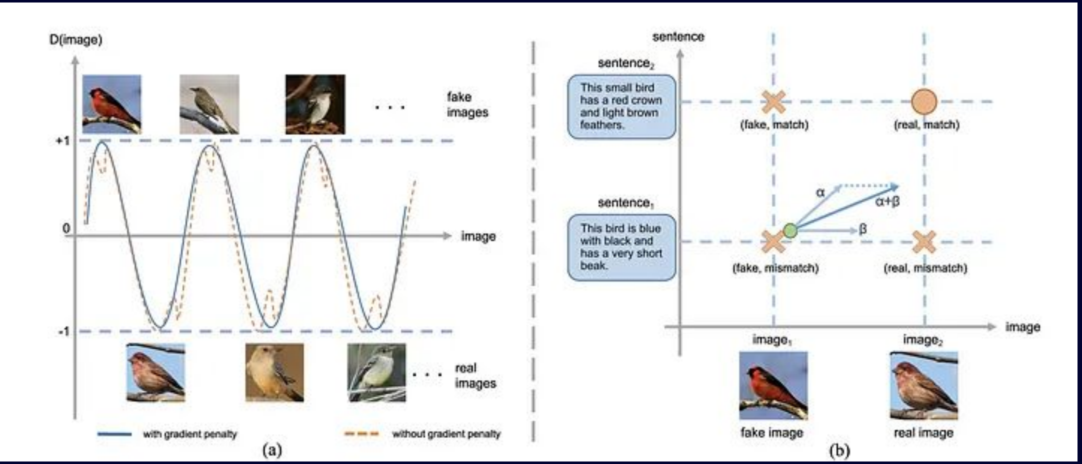
G: generator network D: discriminator network FC: fully connected layer UPBlock: upsample + residual block + DFBlock DownBlock: downsample + residual block

One-Stage Text-to-Image Backbone


Loss$_{D0}$
Loss$_{D1}$
Loss$_{D2}$
D_0
D_1
D_2
(a)
This small bird has a light yellow breast and brown wings.
G_0
G_1
G_2
Loss$_D$
D
Deep Text-Image Fusion Block
(b)
This small bird has a light yellow breast and brown wings.
G

- DF-GAN directly synthesizes high-resolution images from textual descriptions in a single step.
- DF-GAN employs hinge loss to stabilize the adversarial training process.

The Target-Aware Discriminator



(a)

(b)

- Introduce Gradient penalty to smoothens the surface for and around the real data points for smoother convergence
- Push the real and text-image consistent points to the minimum of the loss curve

Deep Text-Image Fusion Block (DFBlock)

- In typical text-to-image setups, this blending might be insufficient, resulting in images lacking detail or not matching the text accurately.
- The DFBlock addresses this by employing multiple layers to blend text and visual features more comprehensively.

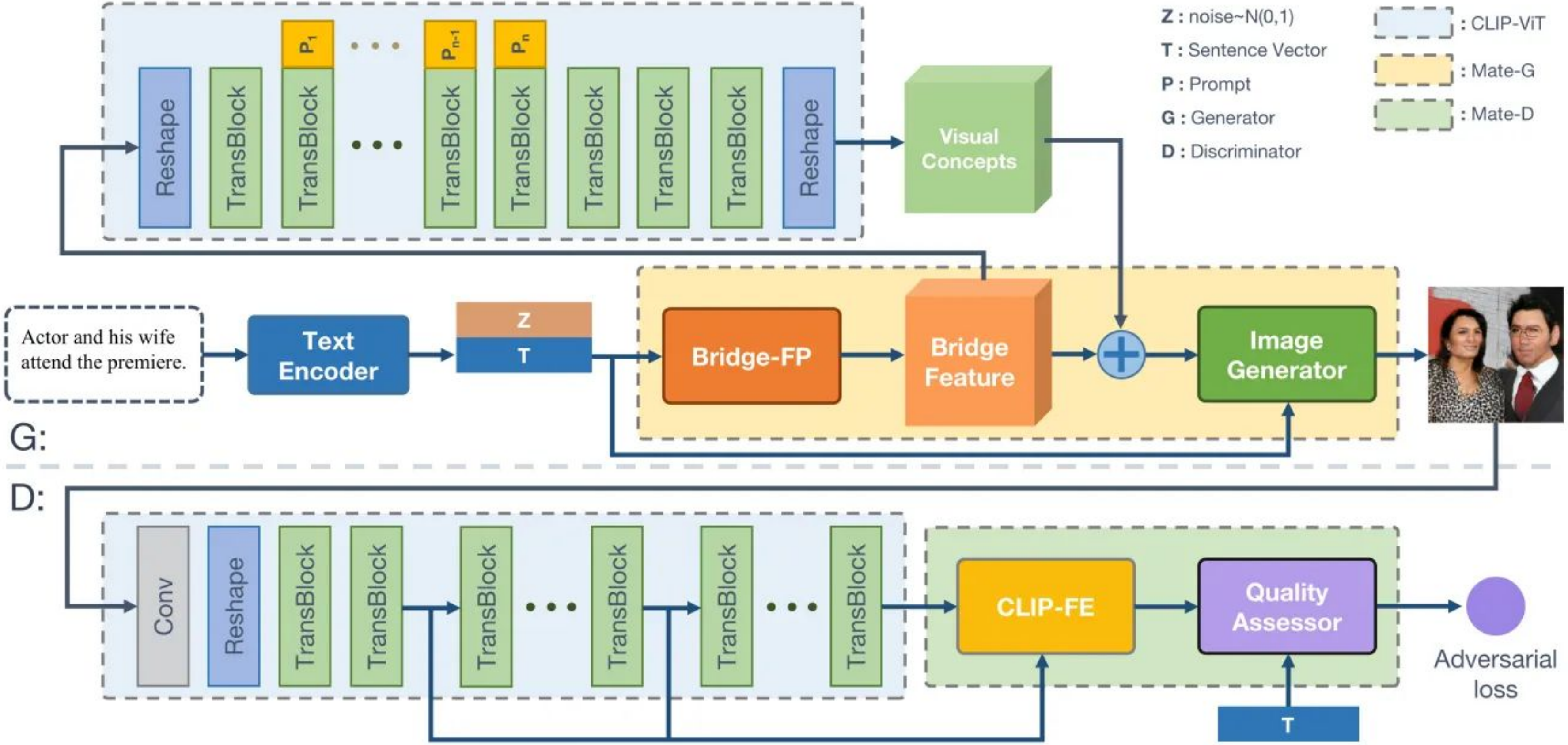
03

GALIP: Generative Adversarial CLIPs

CLIP (Contrastive Language-Image Pre-training)

Neural Network Model developed by OpenAI

- Generate Image from Textual Description
- Image classification



Z : noise~N(0,1)

T : Sentence Vector

P : Prompt

G : Generator

D : Discriminator

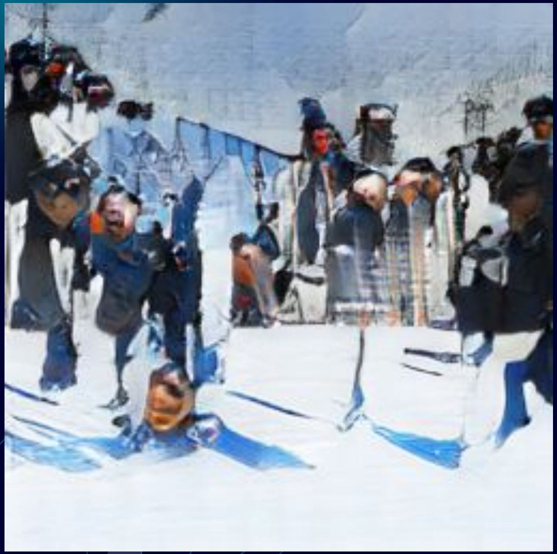
: CLIP-ViT

: Mate-G

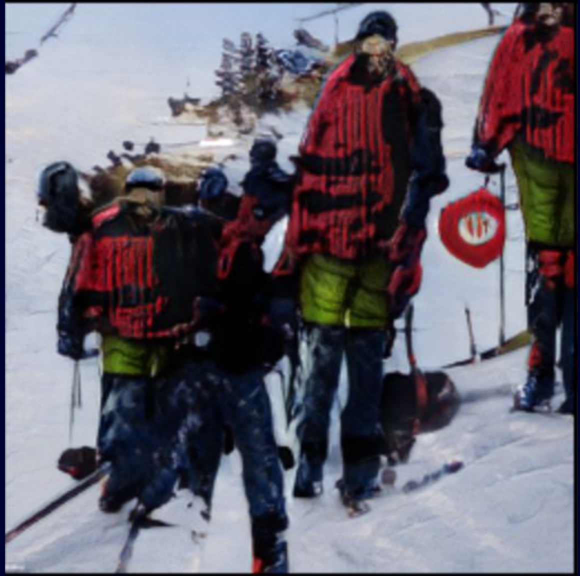
: Mate-D

Sample result from each model

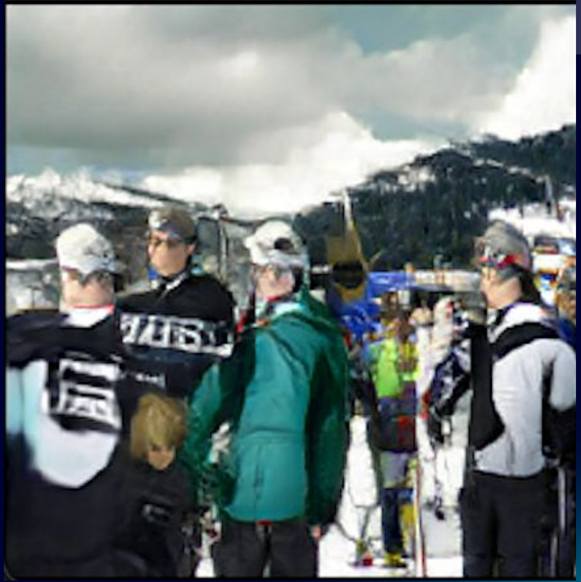
CAPTION “The skiers are standing next to a large crowd.”



AttnGAN



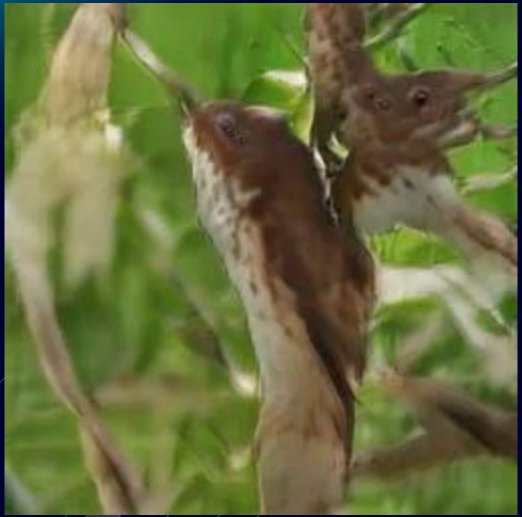
DF-GAN



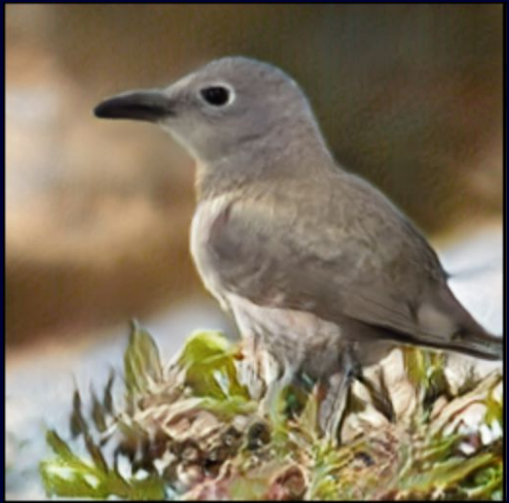
GALIP

Sample result from each model

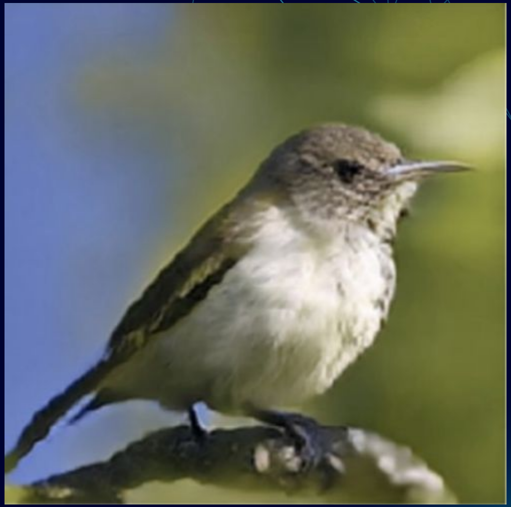
CAPTION “this bird had brown primaries, a brown crown, and white belly.”



AttnGAN



DF-GAN



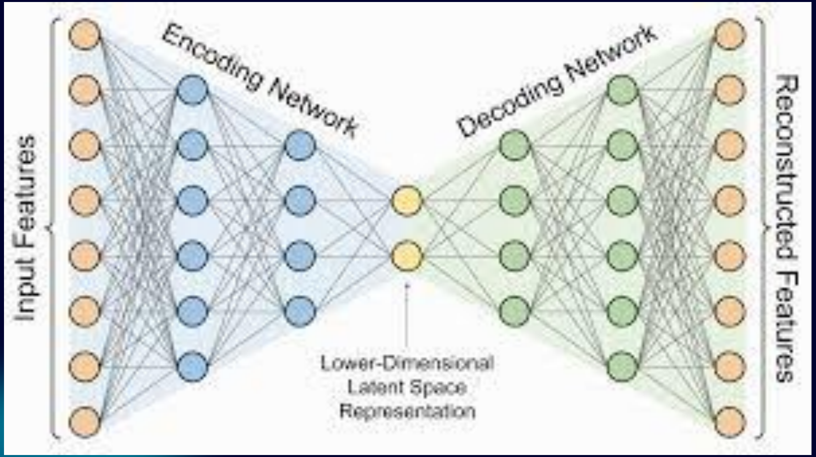
GALIP

04

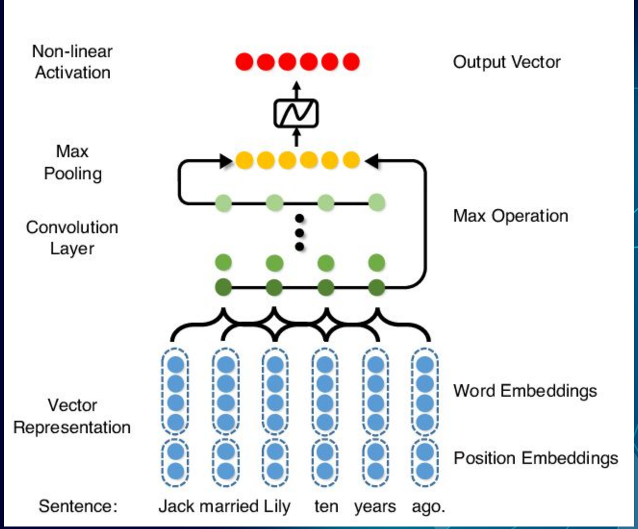
Stable Diffusion

Components

Auto-Encoder



Text-Encoder



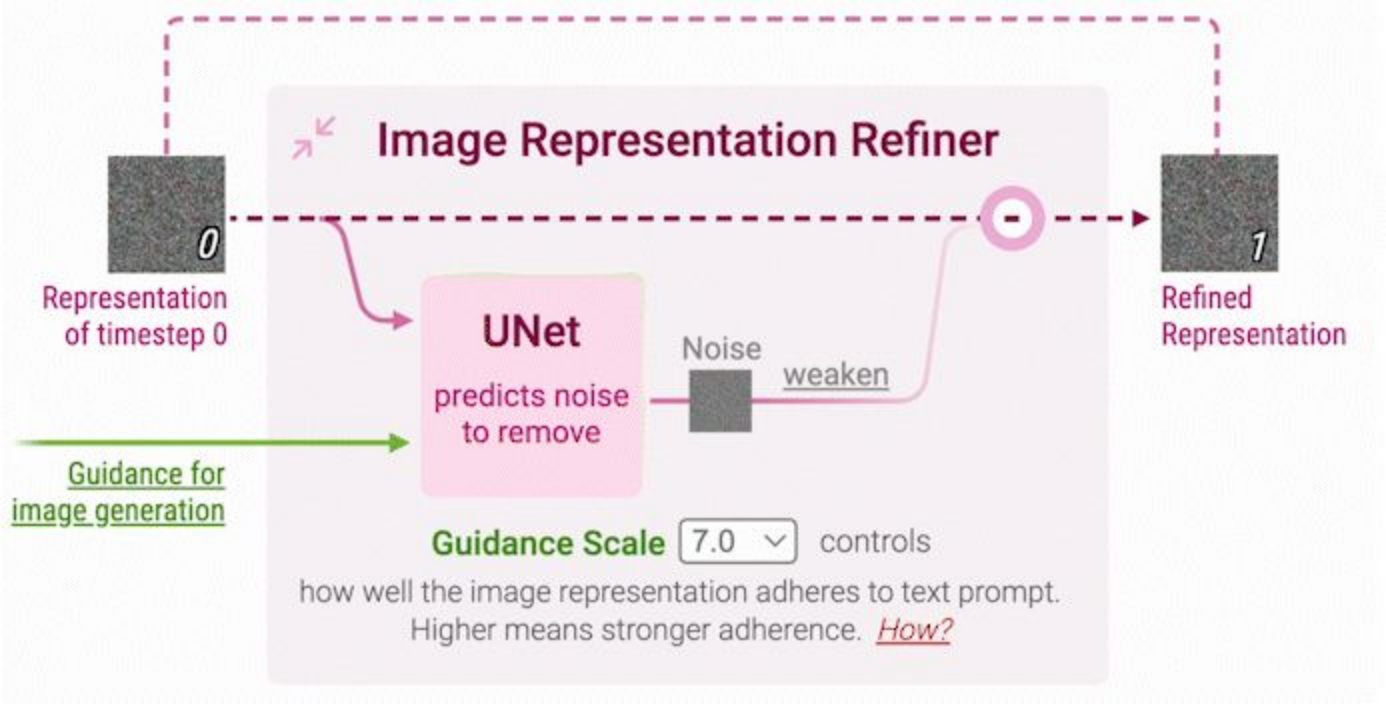
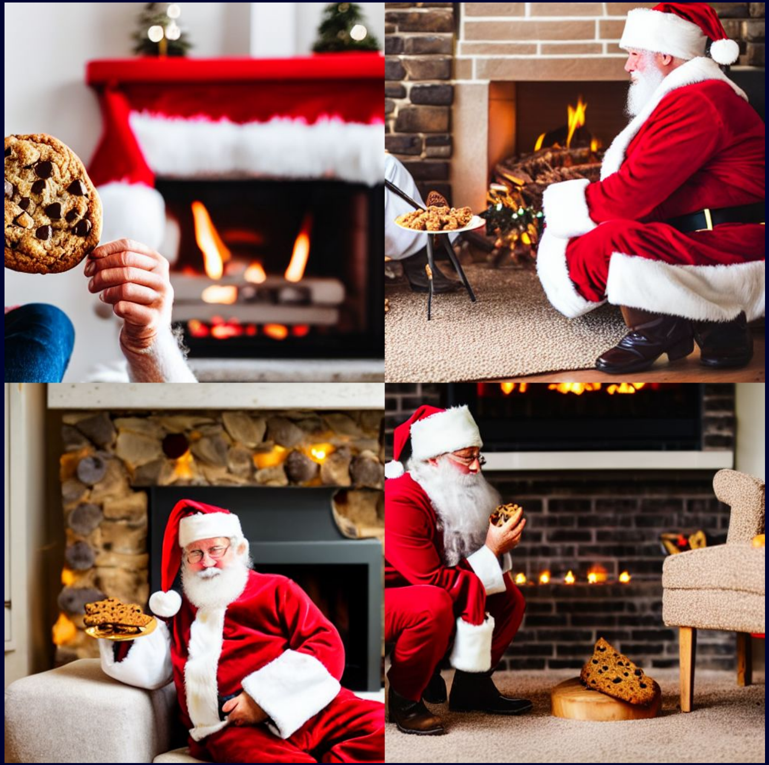
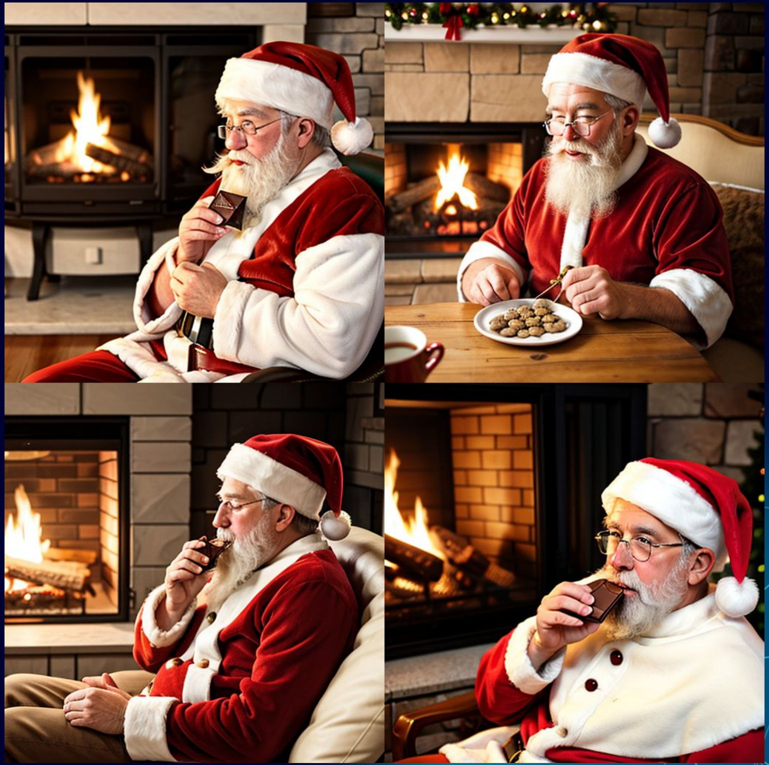


Image Representation Refiner
0
Representation of timestep 0
UNet
predicts noise to remove
Noise
weaken
1
Refined Representation
Guidance for image generation

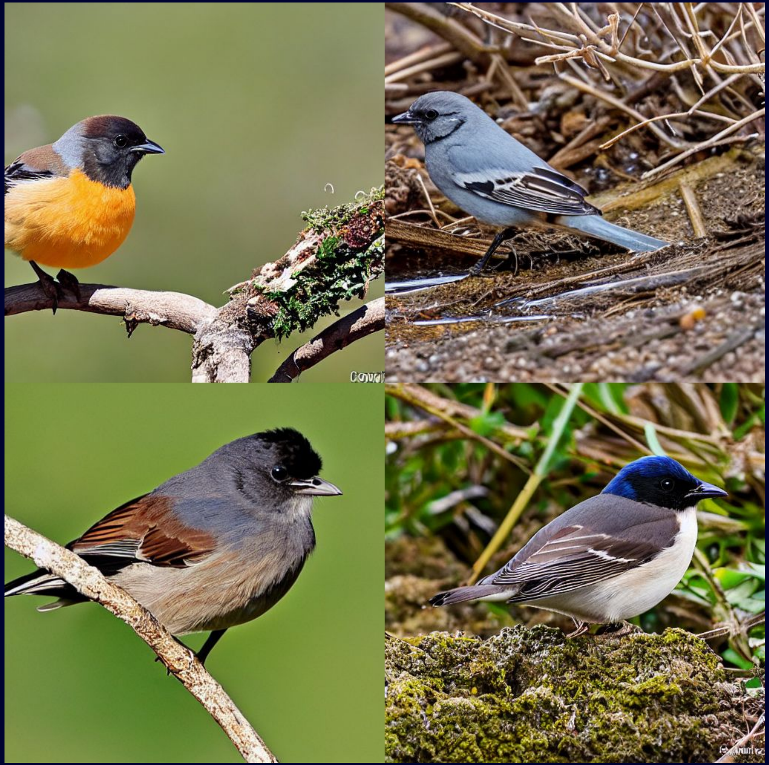
Guidance Scale 7.0 controls how well the image representation adheres to text prompt. Higher means stronger adherence. *How?*



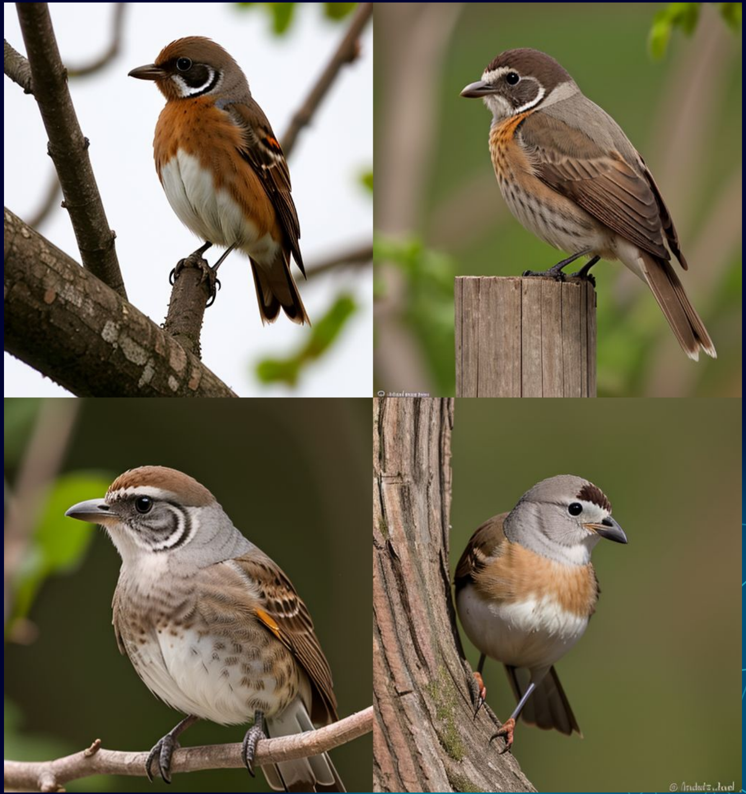
santa clause eating a chocolate chip cookie in front of a cozy fireplace
Negative prompt: ugly, bad anatomy, bad proportions, deformed, extra limbs, low quality, low res, mutated, missing limbs, disfigured, disgusting
Steps: 70, Sampler: DPM++ 2M Karras, CFG scale: 7, Seed: 422078003, Size: 512x512, Model hash: 6ce0161689, Model: v1-5-pruned-emaonly, Version: v1.8.0



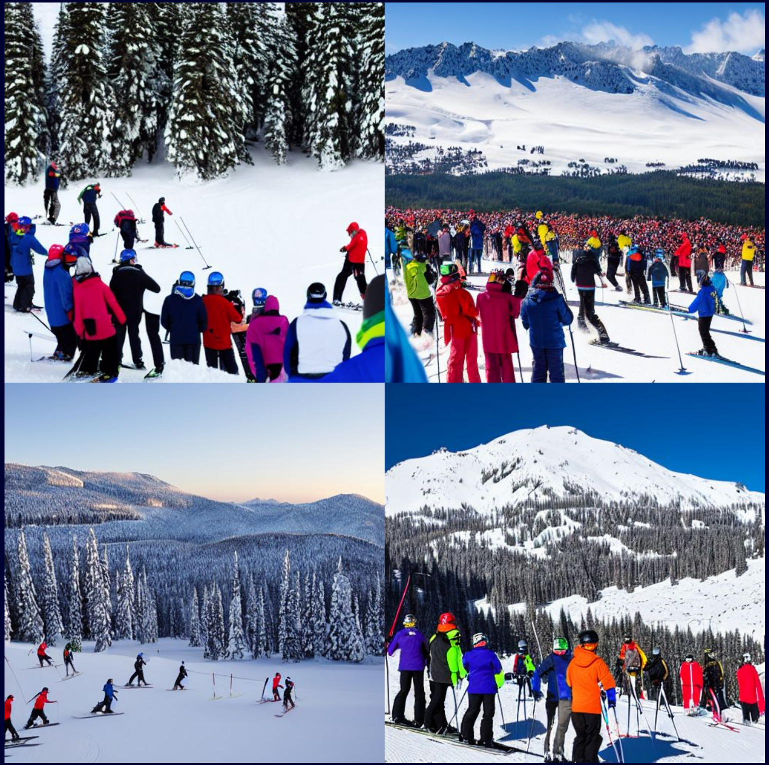
santa clause eating a chocolate chip cookie in front of a cozy fireplace
Negative prompt: ugly, bad anatomy, bad proportions, deformed, extra limbs, low quality, low res, mutated, missing limbs, disfigured, disgusting
Steps: 70, Sampler: DPM++ 2M Karras, CFG scale: 7, Seed: 3889559361, Size: 512x512, Model hash: 93ed864a22, Model: cyberrealistic_v42, VAE hash: c6a580b13a, VAE: vae-ft-mse-840000-ema-pruned.ckpt, Version: v1.8.0



this bird had brown primaries, a brown crown, and white belly
Negative prompt: ugly, deformed, extra limbs, low quality, low res, mutated, missing limbs, disfigured, disgusting
Steps: 70, Sampler: DPM++ 2M Karras, CFG scale: 7, Seed: 1085542589, Size: 512x512, Model hash: 6ce0161689, Model: v1-5-pruned-emaonly, Version: v1.8.0

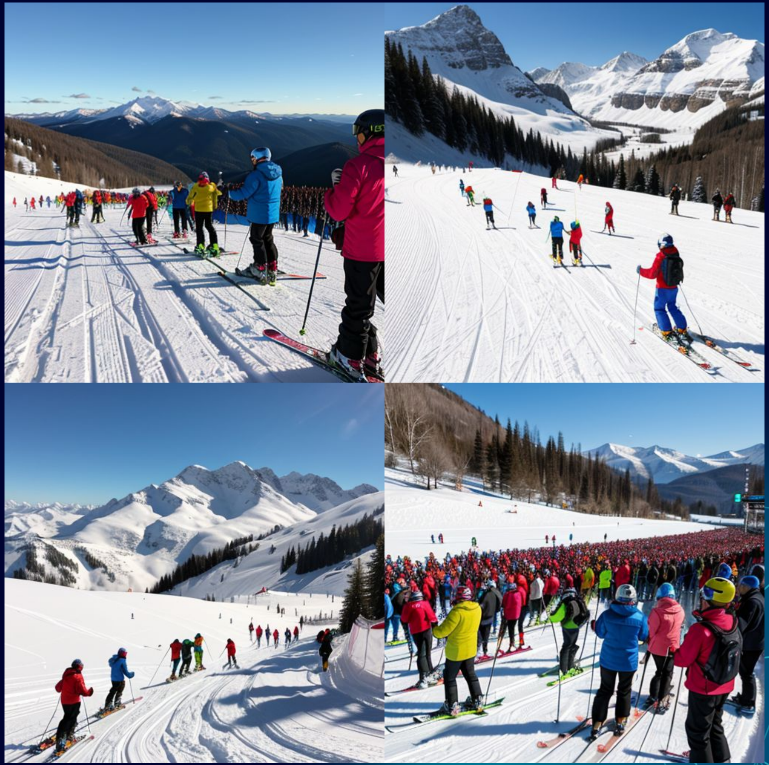


this bird had brown primaries, a brown crown, and white belly
Negative prompt: ugly, deformed, extra limbs, low quality, low res, mutated, missing limbs, disfigured, disgusting
Steps: 70, Sampler: DPM++ 2M Karras, CFG scale: 7, Seed: 250506351, Size: 512x512, Model hash: 93ed864a22, Model: cyberrealistic_v42, VAE hash: c6a580b13a, VAE: vae-ft-mse-840000-ema-pruned.ckpt, Version: v1.8.0



The skiers are standing next to a large crowd
Negative prompt: ugly, deformed, extra limbs, low quality, low res, mutated, missing limbs, disfigured, disgusting

Steps: 70, Sampler: DPM++ 2M Karras, CFG scale: 7, Seed: 1179824724, Size: 512x512, Model hash: 6ce0161689, Model: v1-5-pruned-emaonly, Version: v1.8.0



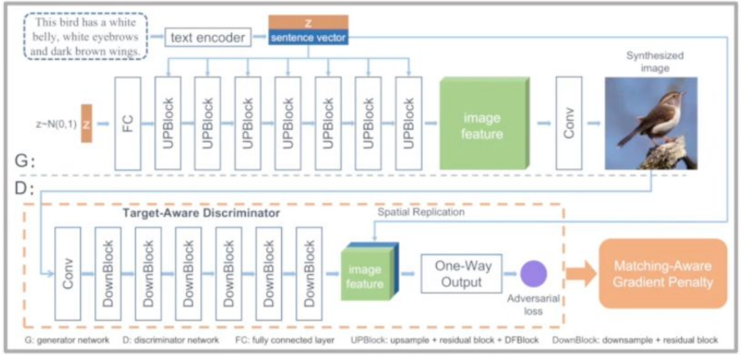
The skiers are standing next to a large crowd
Negative prompt: ugly, deformed, extra limbs, low quality, low res, mutated, missing limbs, disfigured, disgusting

Steps: 70, Sampler: DPM++ 2M Karras, CFG scale: 7, Seed: 435320631, Size: 512x512, Model hash: 93ed864a22, Model: cyberrealistic_v42, VAE hash: c6a580b13a, VAE: vae-ft-mse-840000-ema-pruned.ckpt, Version: v1.8.0

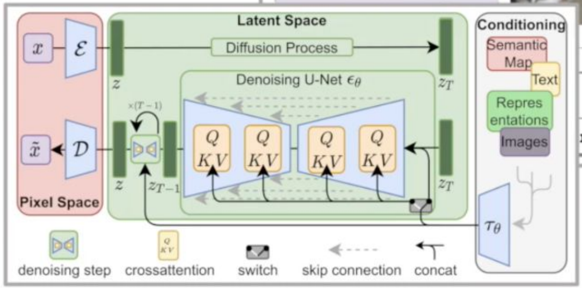
Diffusion model pros and cons

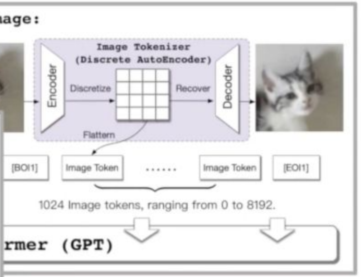
- Enhancement from low-quality source
- Feature- specific enhancement
- open-sourced

- Require lot of training data
- Time Consuming
- Synthesized visual features
- Computationally intensive



The framework of DF-GAN [1]





The framework of Cogview [3]

The framework of LDM [2]

□ GAN

- hard to synthesize complex images
- ✓ fast synthesis speed
- ✓ small model size
- ✓ meaningful latent space

□ AR and diffusion models

- ✓ more powerful generative capabilities
- slow synthesis speed
- large model size and hardware requirements
- lack a meaningful latent space

THANK YOU!

Resources

https://www.vision.caltech.edu/datasets/cub_200_2011/
https://paperswithcode.com/dataset/cub-200-2011
https://codeburst.io/understanding-attngan-text-to-image-convertor-a79f415a4e89
https://blog.segmind.com/stable-diffusion-deployment/#the-anatomy-of-stable-diffusion-
https://poloclub.github.io/diffusion-explainer/#:~:text=Stable%20Diffusion%20generates%20an%20image,quality%20of%20the%20image%20representation.

Citations

AttnGAN:

@article{Tao18attngan,
author = {Tao Xu, Pengchuan Zhang, Qiuyuan Huang, Han Zhang, Zhe Gan, Xiaolei Huang, Xiaodong He},
title = {AttnGAN: Fine-Grained Text to Image Generation with Attentional Generative Adversarial Networks},
Year = {2018},
booktitle = {{CVPR}}
}

Citations

DF-GAN:

```
@inproceedings{tao2022df,
          title={DF-GAN: A Simple and Effective Baseline for Text-to-Image Synthesis},
          author={Tao, Ming and Tang, Hao and Wu, Fei and Jing, Xiao-Yuan and Bao,
          Bing-Kun and Xu, Changsheng},
          booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision
                  and Pattern Recognition},pages={16515--16525}, year={2022}
          }
```

Citations

GALIP:

```
@inproceedings{tao2023galip,
        title={GALIP: Generative Adversarial CLIPs for Text-to-Image Synthesis},
        author={Tao, Ming and Bao, Bing-Kun and Tang, Hao and Xu, Changsheng},
        booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision
                   and Pattern Recognition},
        pages={14214--14223},
        year={2023}
       }
```

Citations

Stable Diffusion

```
authors:
    - given-names: AUTOMATIC1111
    title: ”Stable Diffusion Web UI”
    date-released: 2022-08-22
    url: ”https://github.com/AUTOMATIC1111/stable-diffusion-webui”
```