DL Competition3 Team16 Report

周聖諺 Shen-Yen Chou

余孟旂 Meng-Chi Yu

馮翔荏 FENG-XIANG REN

1.

a) caption: white petals that become yellow as they go to the center where there is an orange stamen.

b) caption: the flower has a large bright orange petal with pink anther

c) caption: the pedicel on this flower is purple with a green sepal and rose colored pet als

d) caption: this flower is white and yellow in color with petals that are rounded at the edges

e) caption: lavender and white pedal and yellow small flower in the middle of the ped als

2. Models we’ ve tried

(1) MirrorGAN

The model consists of three modules: a semantic text embedding module (STEM), a global-local collaborative attentive module for cascaded image generation (GLAM), and a semantic text regeneration and alignment module (STREAM). STEM generat es word- and sentence-level embeddings. GLAM has a cascaded architecture for ge nerating target images from coarse to fine scales, leveraging both local word attenti on and global sentence attention to progressively enhance the diversity and semanti c consistency of the generated images. STREAM seeks to regenerate the text descri ption from the generated image, which semantically aligns with the given text descr iption.

After we tried self-attention-gan, we found a better version which improved the loss function. But we observed that it took much time to train, so we gave up this mode l.

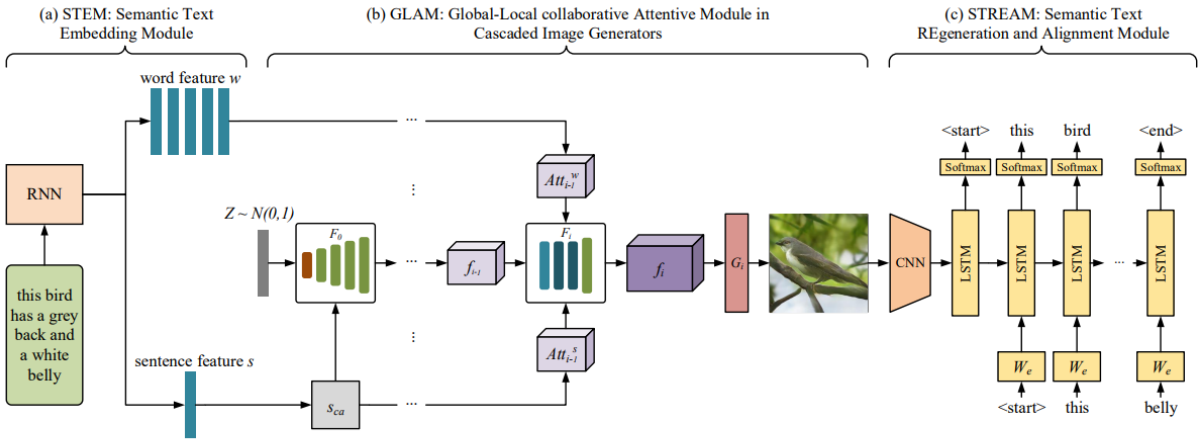

Figure 2: Schematic of the proposed MirrorGAN for text-to-image generation.

(2) StackGAN

The Stage-I GAN sketches the primitive shape and basic colors of the object based on the given text description, yielding Stage-I low resolution images. The Stage-II GAN takes Stage-I results and text descriptions as inputs, and generates high resolu tion images with photo realistic details. The Stage-II GAN is able to rectify defects and add compelling details with the refinement process.

Since it is the earliest improvements of *Generative Adversarial Text to Image Synth esis* and the model isn’ t too complex, we decide to give it a try.

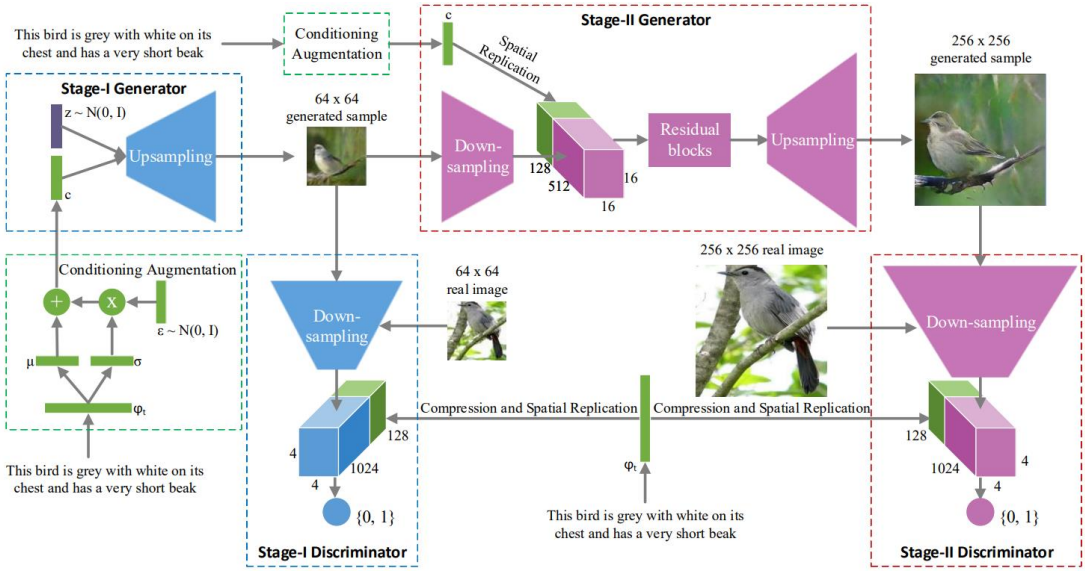

Figure 2. The architecture of the proposed StackGAN. The Stage-I generator draws a low resolution image by sketching rough shape and basic colors of the object from the given text and painting the background from a random noise vector. The Stage-II generator generates a high resolution image with photo-realistic details by conditioning on both the Stage-I result and the text again.

(3) DFGAN

They propose a novel fusion module called Deep Text-Image Fusion Block which c an exploit the semantics of text descriptions effectively and fuse text and image feat ures deeply during the generation process.

Since it is quite simple and has great results, we give it a try. However, the generate d image of it is diverse largely. Sometimes the images are awesome and sometimes are really bad.

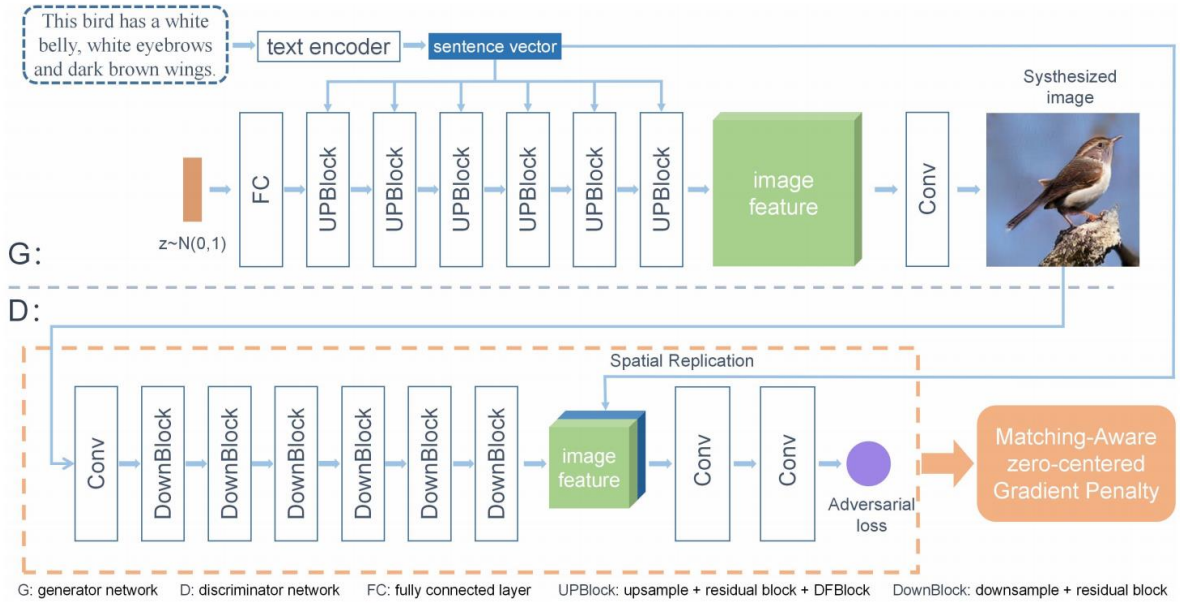

Fig. 2. The architecture of DF-GAN proposed for text-to-image synthesis. Our DF-GAN generates high-resolution images directly by one pair of generator and discriminator.

3. Experiments

(1) Data-Efficient GANs with DiffAugment

We refer to the paper and repo.

We propose Differentiable Augmentation (DiffAugment), a simple method that improves the data efficiency of GANs by imposing various types of differentiable augmentations on both real and fake samples. It can effectively stabilize training, and leads to better convergence.

However, after we applied this augmentation to our network, we got a worse result.

(2) BERT Sentence Embedding

We’ ve tried to use pre-trained Siamese-BERT(BERT-BASE-NLI) to embed the given sentences. However, it doesn't give a better result.

We’ ve thought that we should fine-tune the BERT to get a better result. However, due to lack of time, we didn’ t do that.

(3) Batch Size Tuning:

Since our result was not good, we tried a different batch-size (1024). Although it was faster than the original batch-size (64), the generated image looked worse than the original.

As a result, we still used batch-size 64.