The Summary of AEGAN

Overview

When generating images directly from the piror distribution(input noise) the quality of images is not good(64*64 and 128*128 preserve good but for 256 * 256 and 512 * 512 is limited)
So they use the **embedding** extracted from an aoutoencoder to **bridge** the distribution gap between the input noise and real data. The embeddings often contain rich structural information and help to generalize well with high resolution images. **After** using the autoencoder to extract the low-dimensional embedding, they learn a GAN in the **embedding space** to exploit the structural information. **Moreover**, they devise a denoiser network to remove artifacts/noises and refine the photo-realistic details.
The difficulties of generating high resolution images:

- It is hard to directly learn a mapping between prior distrbution(**input noise**) and the **distrubution of high-dimensional real data**. Since high-dimensional data often lie on some low-dimensional manifold. (That is the reason why they use low-dimensional embedding to uncover the image's structural information, which acts as a bridge to connect the prior distribution with distribution of real data)
- Second, regarding this issue is that there is no additional knowledge, such as label or semantic information obtained from real data, to help the model training.

Based on those problems, they propose the Auto-Embedding Generative Adversarial Networks for high resolution image synthesis.

Architecture

The proposed method consists of threee components :

- autoencoder
- embedding generative model
- denoiser network

Embedding generator G_E and the decoder F with a stack of up-sampling blocks which determined by upscaling factor (deconv followed by residual block).
For the encoder H, which consists of stack of down-sampling blocks (conains a strided deconvolutional layer followed by a residual blocks)to encode RGB images to the low-dimensional embeddings.
For the discriminators D_R and D_E,those models with a stack of strided CNN to down-sample the input images or embeddings.
The denoiser network follows the encoder-decoder design. The input generated images are fed into several down-sampling modules (CNN followed by BN and LearkyReLU) until it has size of 512. A

series of upsampling modules(DeConv followed by BN and ReLU) and then used to generate 512 * 512 RGB HR

Learning Embedding By AutoEncoder

The aim is that seeking to bridge the distribution gap between the input noise and the real images using a latent embedding extracted from an autoencoder

- encoder H which maps the high resolution images into a low dimensional embedding [H is fully CNN that extracts the high-level features of the data]
- decoder F which translated the latent embedding back to high resolution images. [F is a fully deconvolutional model that recovers the high resolution images]

Adversarial Embedding Generator

With the extracted embedding which contains image structural information, they seek to exploit it to improve the training of GANs. They define a generative model to match the meaningful embedding extracted from real data.

Adversarial Denoiser Network

This aims to figure out the artifacts in generated images. They observe that the decoded images often encounter visual noisy artifacts after going through the pipeline of the GAN and the decoder. This denoiser network follows the encoder-decoder design. The important thing is that the stride operation of convolution can extract the primary features of the data and discard pixel-pixel noises.

Training step

First Step

Training the autoencoder by minimizing the reconstruction loss to extract the low-dimensional embedding The learned embedding is able to capture the image structural information and recover the high resolution images

Second step

Fixing the autoencoder model, the autoencoder acts as a bridge that connects all the three components. They train an embedding generatove model and a denoiser network in a alternating manner

Last step

The fine-tuning step which can ensures the coherence of the whole model can achieve better performance

The advantages of using autoencoder

two advantages of extracting the embedding from AE

- AE can extracts the high-level features which preserve the primary data characteristics or structual information to reconstruct the original images. It is helpful to train a **generator** on the extracted features to produce meaningful samples.
- Secondly, the generator only need to learn a **mapping** from the **input noise** to extracted **low-dimensional embedding** (which come from autoencoder extract the high-level features) rather than the high-dimensional images, which greatly facilitates the training of deep generative models.