# Image processing model based on SinGAN

**Abstract:** Generative adversarial network (GAN) is a new and effective method of training generative models in the field of image generation. Single generative adversarial network (SinGAN) is an unconditional GAN built on a picture. Based on the principle of generative confrontation network, this paper expounds the principle and basic structure of single generative confrontation network (SinGAN). Taking SinGAN as the research object, an image with semantic similarity to the original image was generated in the training process through examples, and the training and application of SinGAN were studied. Studies have shown that: The training or testing performed by SinGAN on the same picture, through the establishment of a model, can transfer the texture of the picture, improve the definition and other operations.

**Key words:** generative adversarial network, deep learning, image processing.

1. **Introduction**

Since McCarty proposed the term "artificial intelligence" in 1956, the development of artificial intelligence has experienced many twists and turns, and its development has been slow. In recent years, with the rapid growth of data volume and the substantial improvement of computer computing power, artificial intelligence has achieved rapid development and reached an unprecedented height, becoming the most popular topic today. Especially in the field of machine learning, its performance is more prominent, which makes many researchers pay more and more attention to machine learning. Machine learning is divided into supervised learning, unsupervised learning and semi-supervised learning according to whether the data set has a mark. Supervised learning relies on labeled data, but the acquisition of a large amount of labeled data is expensive, and in learning tasks such as data generation and strategy learning, the acquisition of these labeled data is not even feasible. Unsupervised learning is more in line with the concept of intelligence, and researchers generally believe that unsupervised learning will be one of the important development directions of artificial intelligence[1].

Generative models are the key technology in unsupervised learning tasks. Early generative models, such as deep belief networks(Deep belief network,DBN[2]), have good results, but have not received strong attention from people.

The generative adversarial network[3] is a generative model proposed by Dr. Godlow in 2014. It is based on statistics and probability theory, drawing on the adversarial ideas in game theory, and provides an efficient deep learning method for data generation. The sample method can obtain high-quality generated data through competition between the generated network and the discriminant network without or with a small amount of labeled data.

The guiding ideology of the GAN series of algorithms is the end-to-end[4] idea. After entering the data into the network, you can directly get the output of the result. Using the GAN generation algorithm. The network model will automatically perform feature extraction and learning through the hostile network method. Strong openness and high quality of generated images. In recent years, with the rapid development of deep learning, fields such as image generation, image style transfer, and image processing have become research hotspots[5-8].

GAN has made a huge leap in modeling high-dimensional distribution of visual data[9,10]. Especially when training with category-specific data sets (such as faces, boulevards, etc.), unconditional GAN has achieved remarkable success in generating realistic, high-quality samples. However, the distribution of building molds with multiple categories and highly diversified data sets (such as ImageNet) is still a major challenge, and usually needs to be adjusted to generate or train models for specific tasks based on another input signal.

A single generation adversarial network (Singan) [11]is a model with a simple architecture that can process common natural images that contain complex structures and textures, without having to rely on datasets with images of the same category. This is achieved by a pyramid-structured fully convolutional GAN, each GAN is responsible for capturing image distribution at different scales.

Singan brings GAN into a new field: learning unconditional generative models from a single natural image. Modeling the internal distribution of images in a single natural image has been recognized as a useful prior for many computer vision tasks. A single natural image usually has enough internal statistical information to enable the network to learn a strong generative model. This article will discuss the image generation model of SIGAN based on the introduction of GAN principles and basic structure.

## 2 The principle and basic structure of GAN

### 2.1 The principle of GAN

The core idea of GAN operation comes from the Nash equilibrium of game theory, and the two models gradually reach equilibrium through continuous competition and optimization [9].

### 2.2 The basic structure of GAN

The generative adversarial network (GAN) is composed of two parts: a generator and a discriminator. The purpose of the generator is to generate a false target, attempting to completely deceive the discriminator's recognition, as shown in Figure 1 [10]. The training strategy of GAN is to define a binary maximum and minimum game between two competing networks. The generator network G receives random noise sampled in the prior distribution Pz (z) and maps it to the new In the data space, the generated data G (z) is obtained; the real data x comes from the real data distribution Pd (x). The discriminator network D acts as a binary classifier, judges whether the input sample is real data x or generated data G (z), and uses the Sigmoid function to output a value between 0 and 1 to express the true and false of the data. G and D constantly improve their generating and discriminating ability in the training of confrontation

Until the discriminator cannot finally distinguish the real data from the generated data accurately, it is considered that the generator has reached the optimal, and at this time it has reached a dynamic Nash equilibrium. The input of generating model G is a random variable with uniform distribution following [-1, 1], and the output is a picture (or other). Therefore, the structure of the generated network is a deconvolution network. The discriminator improves its discriminating ability by learning the true and false goals, and does not allow the false goals to deceive itself. The two evolved with each other, and played with each other. One side evolved, the other lost, and finally stopped until the false target was very similar to the true target. For vivid images, researchers all agree on the following image metaphor [9, 10]: The generator (gener-ator, G) is an image forger, trying to learn the distribution of real images as much as possible, the purpose is to create fake images. The discriminator (discrimina-tor, D) is an image discriminator, receiving fakes and real images, the purpose of which is to distinguish them. In the literature [3], there is a more vivid description, that is, the relationship between the generator and the discriminator is described as "counterfeiters and policemen using counterfeit banknotes", the purpose of "counterfeiters" is to make counterfeit banknotes It is no different from real banknotes, and the purpose of the police is to distinguish real banknotes from counterfeit banknotes. Through the competition between them, counterfeiters have improved their counterfeiting ability, and at the same time, the police have also improved their discrimination ability. In the end, the probability that the police will distinguish between genuine and counterfeit banknotes will approach 0.5.

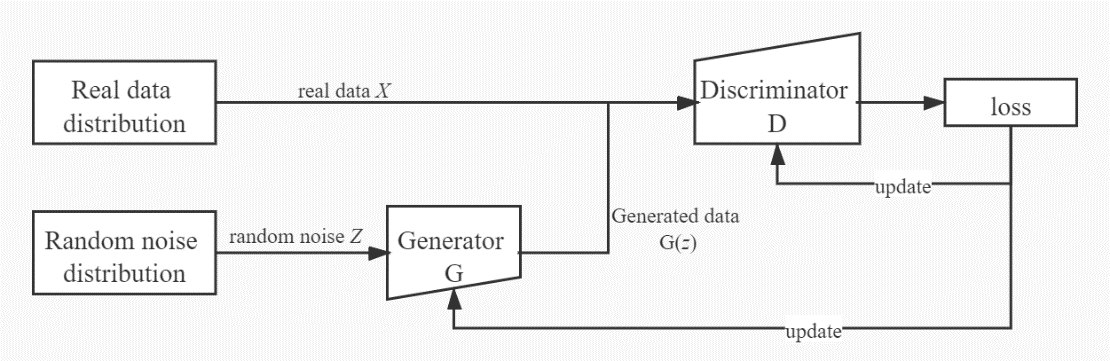


Fig 1: base structure of GAN.

## 3 The principle and basic structure of SinGAN

### 3.1 The principle of SinGAN

SINGAN is an adversarial network model based on GAN, which is an unconditional generative model learned from a single natural image [11]. The goal is to learn an unconditional generative confrontation model. This model can capture the internal data relationships of input training data (internal standards). Conceptually, this model is a bit like the setting of conditional GAN. The training data is an image block of a single image rather than the entire image. In order not only to generate textures but also to process more general natural images, a structural data that can capture complex images at different scales of the image is needed, similar to the need to obtain global attributes (shape and arrangement of large objects, images Details and texture information)

### 3.2 The basic structure of SinGAN

SinGAN is an unconditional generative model that can be learned from a single natural image. It is designed in an unconditional GAN ​​way and is in the form of cascaded Generator-Discriminator p air. Each GD is responsible for a scale. By learning the data distribution in patch, the network can finally output the structure and image of the target in the original image during the test, which is different from the real image of the original image [11]. Compared with the previous single-image GAN method, SinGAN is not only suitable for texture images, but also unconditional (new samples can be generated directly from the noise, without the need to introduce other conditional input as in Continiental GAN, or from image to image. translation). SinGAN is used to train the inherent distribution of patch in the captured image, and then can generate high-quality and diverse samples consistent with the visual content of the training image. SinGAN has a pyramid structure of SinGAN structure composed of N Gan networks. Its basic structure is shown in Fig. 2 [11], . From training to testing, it is based on the coree-to-fine thinking. From bottom to top, the scale gradually changes from rough to fine, which allows the generation of new sample images with arbitrary sizes and aspect ratios.All the generators have similar architecture, as depeticed in Figure 3.

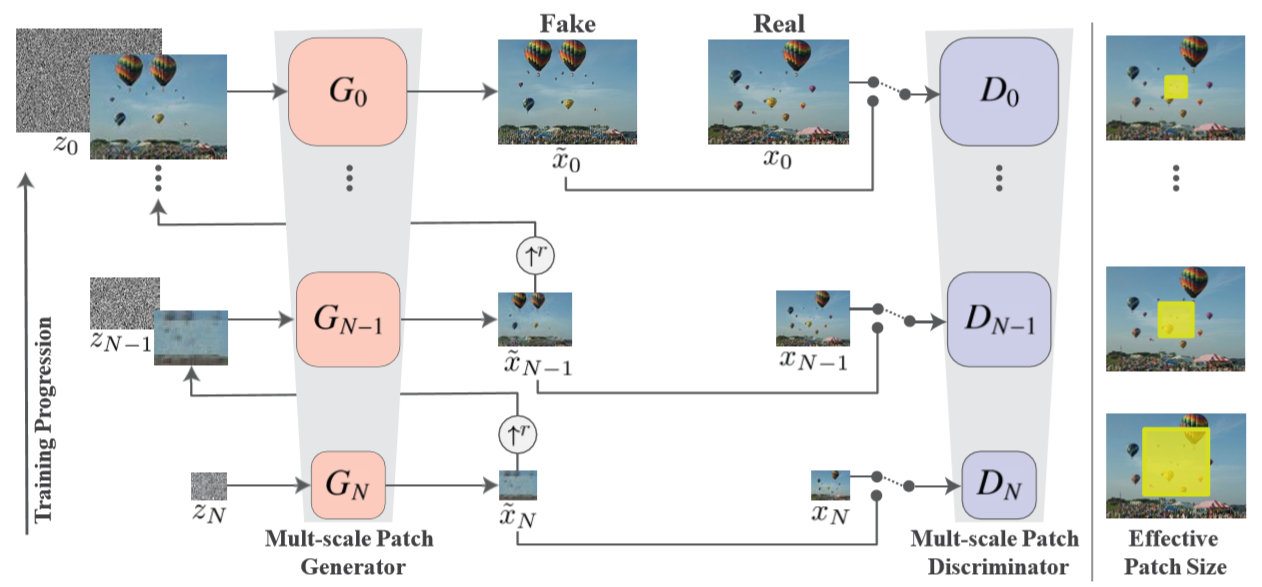


Fig 2: SinGAN’s multi-scale pipeline.

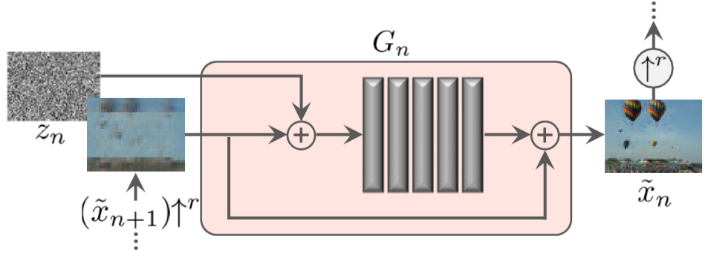


Fig 3: Single scale generation.

At n scales, the input of the Gn network is the generated image of the n + 1th G network. After upsampling, it is added to the corresponding random noise to learn the generated image samples, and the D network judges whether the generated image is true or false. In particular, each layer D network of SINGAN is judged based on patch. The classic patch D rises from the coarsest layer GN to the finest layer G 0. The receptive field of each D is fixed and is n × n. That is, in the coarsest GN, the size of the patch is 1/2 of the image. At this time, the GAN network can learn the global structure of the image, while in the finest G0, the GAN network learns the local details. In terms of discriminators, each level of discriminator uses a Markov discriminator, which is entirely composed of convolutional layers. The final output is a n × n matrix, and the average value of the output matrix is ​​taken as the judgment result. Compared with the ordinary discriminator, the main difference between the Markov discriminator and the ordinary discriminator is that the ordinary discriminator mainly judges whether a picture is "true" or "false", and the Markov discriminator first divides a picture into n × n Matrix, and then judge whether each small image is in the original picture. After judgment, you can get a matrix. Markov discriminator maintains high resolution and high detail for ultra-high resolution and clear picture. Simply put, the generators and discriminators in SinGAN form a pyramid structure. The generator at the bottom of the pyramid learns the overall structure of the image, while the generator at the upper layer learns the texture of the image, the bottom layer The input of the generator is a mixture of the original image and white noise, and the input of the next layer is a mixed sample of the output of the previous layer and random white noise. After the image is trained, SinGAN can generate various high-quality image samples at any size. These samples are semantically similar to the training image, but include new targets and structures, and SinGAN can be applied to various image processing tasks, such as images Drawing, editing, fusion, super-resolution reconstruction and animation.

## 4 Case Analysis

The following is an example of investigating the generating ability of SinGAN. During the training process, an image (the trail in the maple forest) that is semantically similar to the original image is generated. As mentioned above, after the image is trained, SinGAN can generate various high-quality image samples at any size, but it contains new targets and structures, as shown in Figure 4-Figure 6. In this example, the test conducted by the training experiment is the effect of the image fusion of SGAN at different scales. Figure 4 is the training target image, Figure 5 is the random sample image generated by training, and Figure 6 is the reference image. It can be seen that at low scale, the generator has learned the structure of the image, and at high scale, the generator has learned the texture of the image. When using only 2 layers of GAN training (Figure 7-Figure 8), due to the limited receptive field, the model can only learn local details and lack global information. As the number of training layers increases, Singan can learn more global information , But the capture of local details will become weak, as shown in Figure 9-Figure 11. It can be seen that it is almost indistinguishable from the reference image, but at the lower level, the model captures the overall structure of the image weakly, and can only see the texture of the original image.

Fig4:Training example Fig5:Generated sample

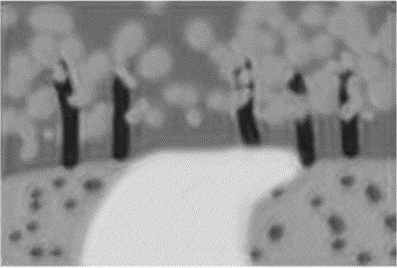
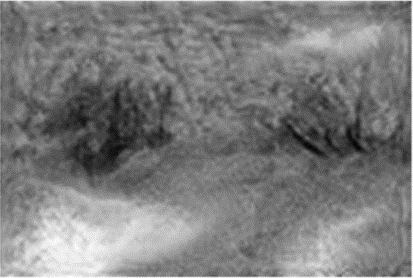
 

Fig6: Input image Fig7: 1 scales

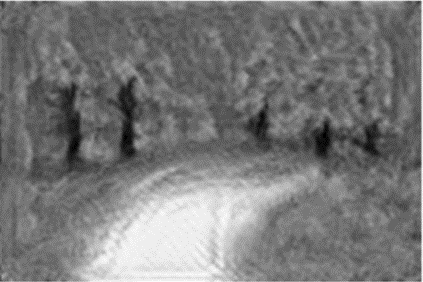
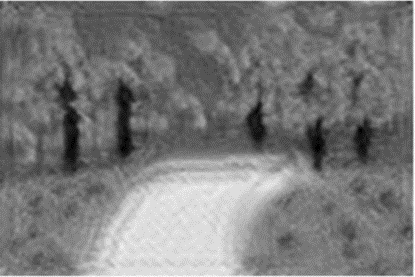
 

Fig8: 2 scales Fig9: 3 scales

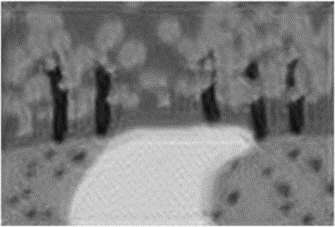
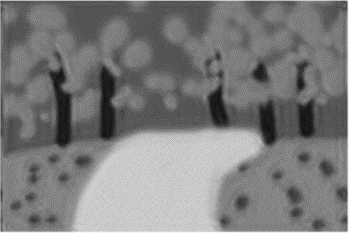
 

Fig10: 4 scales Fig11: 5 scales

## 5 Conclusion

In summary, Singan has the following innovations and advantages. First of all, two restrictions are broken: (1) Singan is trained on an image, regardless of the type of the image, whether the data set is a single type of data (such as a face), or multi-class classification data, only in Training on one image, that is, designing an unconditional GAN ​​network based on single natural image training (generated directly from noise as input); (2) using unconditional GAN, without modifying the model, directly applied to multiple image tasks. Secondly, the G network and D network in the GAN network have the same model structure and the same receptive field, and add a reconstruction loss to ensure that the GAN can be trained smoothly. Again, a pyramid-type GAN network of couse-to-fine is designed, and each layer learns the missing details of the previous layer.

On the other hand, the generated image is interpreted from a new perspective. In the past, GANs usually provided many images of a certain class as the training set, and then the generator learned the distribution of the same features in these samples. For example, human faces have eyes and mouths. Then use the human face as the training set. During the test, input noise, and the network can output a human face with facial features. However, SinGAN started from a new perspective, instead of learning the common features of "faces", instead learning the data distribution of a single face image, so that the network can not only generate this person's face, but also may have this from different perspectives. Human face. In other words, ordinary GAN can generate faces of different people, while SinGAN specializes the generation ability and can generate different states of a face.

However, SinGAN also has shortcomings: it takes a long time to train SinGAN, and a trained model can only process one picture, and its usability is poor. Therefore, it is very critical and urgent to study how to accelerate the SinGAN training, which is very important and significant for computer graphics and image processing and even the field of artificial intelligence.

## 6 References

1. Mantziaris, Charalampos, et al. "Intra-and intersegmental influences among central pattern generating networks in the walking system of the stick insect." Journal of neurophysiology 118.4 (2017): 2296-2310.
2. Hinton, Geoffrey E., Simon Osindero, and Yee-Whye Teh. "A fast learning algorithm for deep belief nets." Neural computation 18.7 (2006): 1527-1554.
3. Goodfellow, Ian, et al. "Generative adversarial nets." Advances in neural information processing systems. 2014.
4. Mo, Jeonghoon, and Jean Walrand. "Fair end-to-end window-based congestion control." IEEE/ACM Transactions on networking 8.5 (2000): 556-567.
5. Yoo, Donggeun, et al. "Pixel-level domain transfer." European Conference on Computer Vision. Springer, Cham, 2016.
6. Li, Chuan, and Michael Wand. "Precomputed real-time texture synthesis with markovian generative adversarial networks." European conference on computer vision. Springer, Cham, 2016.
7. Pathak, Deepak, et al. "Context encoders: Feature learning by inpainting." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.
8. 王坤峰,苟超,段艳杰,林懿伦,郑心湖,王飞跃.生成式对抗网络GAN的研究进展与展望[J].自动化学报,2017,43(03):321-332.
9. 王格格,郭涛,李贵洋.多层感知器深度卷积生成对抗网络[J].计算机科学,2019,46(09):243-249.
10. 朱秀昌,唐贵进.生成对抗网络图像处理综述[J].南京邮电大学学报(自然科学版),2019,39(03):1-12.
11. Shaham, Tamar Rott, Tali Dekel, and Tomer Michaeli. "Singan: Learning a generative model from a single natural image." Proceedings of the IEEE International Conference on Computer Vision. 2019.