

Image Generation Method and Device Based on Neural Network

Field of the Invention

This application relates to the field of image processing, in particular to an image
5 generation method based on a neural network and an image generation apparatus based on the
neural network.

Background of the Invention

Since the introduction of generative adversarial networks (GANs), they have been widely
10 applied to the field of image processing, especially image generation, and have made
breakthroughs in the past few years. Among them, StyleGAN is currently one of the most
advanced depth model that can generate high-resolution and super-realistic images. However,
the quality of the images is greatly reduced due to the prevalence of artifacts in the generated
images. Therefore, how to prevent the artifacts of the StyleGAN model in the generated
15 images has become an urgent problem to be solved.

Summary of the Invention

The embodiments of this application provide an image generation method based on a
neural network, an image generation apparatus based on the neural network, image generation
20 device based on the neural network, and a computer readable medium.

In a first aspect, an image generation method based on a neural network is provided and
includes:

receiving an initial input used to generate the image; and based on the initial input,
implementing M stages of image generation processes to generate the image with a preset
25 target resolution; wherein, in the M stages of the image generation processes, a resolution of
an image generated in each stage based on the initial input is gradually increased compared
with the previous stage from the first stage to the Mth stage until the image reaching the preset
target resolution is output in the Mth stage; and the each stage of the image generation process
includes at least one style control processing for affecting global information of the generated
30 image, and, during the style control processing, pixel-instance normalization processing is
performed on the image, and the pixel-instance normalization processing includes a pixel
normalization processing and an instance normalization processing.

Compared with an image processing method in the prior art, the technical solutions of
this application enable the generated images to have better visual quality, meanwhile, no

artifact is generated, and the overall quality of the generated images is improved. In the process of implementing image generation, only simple solutions are required for eliminating artifacts without the undue loss of image quality, and hardware configurations are optimized better, so that internal structures of a computer system are more simplified.

5 In one possible implementation of the above first aspect, the above method further includes: the initial input is a constant.

In one possible implementation of the above first aspect, the pixel-instance normalization processing includes: performing weighted combination on the pixel normalization processing and the instance normalization processing, as shown in a following formula:

$$10 \quad y = \rho \cdot y^p + (1 - \rho) \cdot y^I,$$

wherein y represents the pixel-instance normalization processing; y^p represents the pixel normalization processing; y^I represents the instance normalization processing; and $\rho \in [0,1]^C$, wherein ρ represents an affect degree of the pixel normalization processing on the images, and $1 - \rho$ represents an affect degree of the instance normalization processing on the images.

15 In one possible implementation of the above first aspect, the above method further includes: based on an input initial vector, implementing the style control processing for the M stages of image generation processes.

In one possible implementation of the above first aspect, based on an input initial vector, implementing the style control processing for the M stages of image generation processes further includes: encoding the initial vector into an intermediate vector through a mapping network in the neural network model, and in the M stages, processing the intermediate vector as N control vectors; and the N control vectors are configured to control visual features corresponding to the images, and by taking advantage of at least one style control processing, N control vectors affect the images, to control M stages of the image generation processes, wherein $N \geq M$.

In one possible implementation of the above first aspect, the above method further includes: the initial vector is a random vector.

In one possible implementation of the above first aspect, the above method further includes: the neural network model including an input layer, the mapping network, a generating network and an output layer, and the generating network further including M network layers; wherein at least one of the M network layers includes an up-sampling layer, a first end modulation layer, a convolutional layer and a second end modulation layer, each of the first end modulation layer and the second end modulation layer includes at least one style

control processing, and the initial input passes through the M network layers to generate the images.

In a second aspect, an image generation method based on a neural network is provided and includes: generating a first image; obtaining a second image serving as a training sample; and inputting the first image and the second image into a discrimination unit to distinguish the first image from the second image, and adjusting the generation process of the first image based on the result of the distinction; wherein the process of generating the first image includes: receiving an initial input used to generate the image; and based on the initial input, implementing M stages of image generation processes to generate the first image with a preset target resolution; wherein, in the M stages of the image generation processes, a resolution of an image generated in each stage based on the initial input is gradually increased compared with the previous stage from the first stage to the Mth stage until the first image reaching the preset target resolution is output in the Mth stage; and the each stage of the image generation process includes at least one style control processing for affecting global information of the generated image, and, during the style control processing, pixel-instance normalization processing is performed on the image, and the pixel-instance normalization processing includes a pixel normalization processing and an instance normalization processing.

In one possible implementation of the above second aspect, the pixel-instance normalization processing includes: performing weighted combination on the pixel normalization processing and the instance normalization processing, as shown in a following formula:

$$y = \rho \cdot y^p + (1 - \rho) \cdot y^I,$$

wherein y represents the pixel-instance normalization processing; y^p represents the pixel normalization processing; y^I represents the instance normalization processing; and $\rho \in [0,1]^C$, wherein ρ represents an affect degree of the pixel normalization processing on the images, and $1 - \rho$ represents an affect degree of the instance normalization processing on the images.

In one possible implementation of the above second aspect, the above method further includes: in the pixel-instance normalization processing, ρ is trainable and the value of ρ is adjusted according to the train automatically.

In a third aspect, an image generation apparatus based on a neural network is provided and includes: a receiving unit, receiving an initial input used to generate the image; and a generation unit, including M generation subunits, wherein based on the initial input, the M generation subunits sequentially implement image generation processes to generate the image

with a preset target resolution; wherein, in the M generation subunits, a resolution of an image generated based on the initial input is gradually increased compared with the previous stage from the first generation subunit to the Mth generation subunit until the image reaching the preset target resolution is output in the Mth generation subunit; and the each of the generation subunits includes at least one style control module for affecting global information of the generated image, the style control module performs pixel-instance normalization processing on the image, and the pixel-instance normalization processing includes a pixel normalization processing and an instance normalization processing.

In one possible implementation of the above third aspect, the pixel-instance normalization processing includes: performing weighted combination on the pixel normalization processing and the instance normalization processing, as shown in a following formula:

$$y = \rho \cdot y^p + (1 - \rho) \cdot y^l,$$

wherein y represents the pixel-instance normalization processing; y^p represents the pixel normalization processing; y^l represents the instance normalization processing; and $\rho \in [0,1]^C$, wherein ρ represents an affect degree of the pixel normalization processing on the images, and $1 - \rho$ represents an affect degree of the instance normalization processing on the images.

In one possible implementation of the above third aspect, the above apparatus further includes: based on an input initial vector, implementing the style control processing for the M stages of image generation processes.

In one possible implementation of the above third aspect, based on an input initial vector, implementing the style control processing for the M stages of image generation processes further includes: encoding the initial vector into an intermediate vector through a mapping network in the neural network model, and in M stages, processing the intermediate vector into N control vectors; and the N control vectors are configured to control visual features corresponding to the images, and by taking advantage of at least one style control processing, N control vectors affect the images, to control M stages of the image generation processes, wherein $N \geq M$.

In one possible implementation of the above third aspect, the above apparatus further includes: the neural network model including an input layer, the mapping network, a generating network and an output layer, and the generating network further including M network layers; wherein at least one of the M network layers includes an up-sampling layer, a first end modulation layer, a convolutional layer and a second end modulation layer, each of

the first end modulation layer and the second end modulation layer includes at least one style control processing, and the initial input passes through the M network layers to generate the images.

In a fourth aspect, an image generation apparatus based on a neural network is provided and includes: a first generation unit, generating a first image; an obtaining unit, obtaining a second image serving as a training sample; and a discrimination unit, inputting the first image and the second image to distinguish the first image from the second image, and adjusting the generation process of the first image based on the result of the distinction; wherein the first generation unit includes: a receiving unit, receiving an initial input used to generate the image; a generation unit, including M generation subunits, wherein based on the initial input, the M generation subunits sequentially implement image generation processes to generate the image with a preset target resolution; wherein, in the M generation subunits, a resolution of the image generated based on the initial input is gradually increased compared with the previous stage from the first generation subunit to the Mth generation subunit until the image reaching the preset target resolution is output in the Mth generation subunit; and each of the generation subunits includes at least one style control module for affecting global information of the generated images, the style control module performs pixel-instance normalization processing on the images, and the pixel-instance normalization processing includes a pixel normalization processing and an instance normalization processing.

In one possible implementation of the above fourth aspect, the pixel-instance normalization processing includes: performing weighted combination on the pixel normalization processing and the instance normalization processing, as shown in a following formula:

$$y = \rho \bullet y^p + (1 - \rho) \bullet y^l,$$

wherein y represents the pixel-instance normalization processing; y^p represents the pixel normalization processing; y^l represents the instance normalization processing; and $\rho \in [0,1]^C$, wherein ρ represents an affect degree of the pixel normalization processing on the images, and $1 - \rho$ represents an affect degree of the instance normalization processing on the images.

In one possible implementation of the above fourth aspect, the above apparatus further includes: in the pixel-instance normalization processing, ρ is trainable and the value of ρ is adjusted according to the train automatically.

In a fifth aspect, image generation device based on a neural network is provided and includes:

a memory, configured to store an instruction executed by one or more processors of a system, and

a processor, serving as one of the processors of the system and configured to execute the instruction to implement the image generation method based on the neural network of any one of the first aspect or the second aspect.

In a sixth aspect, a computer readable storage medium encoded through a computer program is provided and characterized in that an instruction is stored on the computer readable medium, and when the instruction is executed on a computer, the computer executes the image generation method based on the neural network of any one of the first aspect or the second aspect.

Brief Description of Drawings

Fig. 1 shows a schematic structural diagram of a neural network model for generating images according to some embodiments of this application;

Fig. 2 shows a structural block diagram of electronic device according to some embodiments of this application;

Fig. 3 shows a flow chart of an image generation method based on a neural network according to some embodiments of this application;

Fig. 4 shows a flow chart of implementing steps of M stages of generation processes of a first image according to some embodiments of this application;

Fig. 5 shows a structural block diagram of a neural network model according to some embodiments of this application;

Fig. 6 shows a schematic diagram of an architecture of a network layer and style control processing thereof according to some embodiments of this application;

Fig. 7 shows a schematic diagram of an artifact generation process according to some embodiments of this application;

Fig. 8 shows a schematic structural diagram of an image generation apparatus based on a neural network according to some embodiments of this application;

Fig. 9 shows a schematic structural diagram of another image generation apparatus based on a neural network according to some embodiments of this application; and

Fig. 10 shows a schematic structural diagram of electronic device according to some embodiments of this application.

Detailed Description of the Embodiments

The illustrative embodiments of this application include, but are not limited to, an image generation method based on a neural network, an image generation method based on the neural network, image generation device based on the neural network, and a computer readable medium.

5 It can be understood that for example, the terms "module" and "unit" used in this text can refer to or include an application-specific integrated circuit (ASIC), an electronic circuit, a processor (shared, dedicated or grouped) executing one or more software or firmware programs and/or a memory, a combinational logic circuit, and/or other appropriate hardware components providing described functions, or can serve as parts of these hardware
10 components.

 It can be understood that in all the embodiments of this application, the processor may be a microprocessor, a digital signal processor, a microcontroller, etc., and/or any combination thereof. According to another aspect, the processor may be a single-core processor, a multi-core processor, etc., and/or any combination thereof.

15 It can be understood that the image generation method based on the neural network and provided by this application can be implemented on various electronic device, including but not limited to, a server, a distributed server cluster consisting of a plurality of servers, a mobile phone, a tablet computer, a laptop computer, a desk computer, a wearable apparatus, a head-mounted display, mobile email device, a play station portable, a portable music player,
20 reader device, a personal digital assistant, virtual reality or augmented reality device, a television with one or more processors being embedded or coupled thereinto, etc..

 In particular, the image generation method based on the neural network and provided by this application is applicable to edge apparatus, edge computing is a distributed open platform (architecture) that integrates the core capabilities of networks, computing, storage and
25 applications on the network edge side close to an object or a data source, provides edge intelligent services nearby, and can meet key demands in the aspects of real-time service, data optimization, application intelligence, security and privacy protection, etc.. For example, the edge apparatuses may be apparatuses capable of performing edge computing on image data in videos in the ends close to video data sources (intelligent network cameras) in video
30 monitoring systems.

 As shown in Fig .1, according to some embodiments of this application, a neural network model for generating images is shown, which belongs to a generative adversarial network (GAN). The GAN is a machine learning architecture. Compared with a traditional neural network model, the GAN is a brand new unsupervised network architecture capable of

learning from mistakes by themselves and reducing the probability of future errors. As shown in Fig. 1, the GAN includes two independent networks, both of which are the targets of mutual antagonism. The first network is a generating network, which is configured to generate random images similar to real images and regard them as false images. The second network is a discrimination unit needing to be trained, which is configured to distinguish whether they are real images or false images.

In the training process, the goal of the generating network is to draw the false images that are very close to the real images to deceive the discrimination unit. Elements in potential spaces of training data are selected to be combined as an initial input, and random noise is added. For example, as the input of the generating network, an image of a cat can be selected, and then a third eye is added on the cat, thereby serving as a false image. As an image discriminator, the discrimination unit receives the real images and the false images generated by the generating network, and its task is to judge whether the images generated by the generating network belong to the real images or the false images. If the discrimination unit determines that the images generated by the generating network are the false images, parameters of the generating network need to be adjusted to enable the generated false images to be more realistic. If the discrimination unit determines that the images generated by the generating network are the real images, after the discrimination unit generates errors, network parameters of the discrimination unit do not need to be changed during training, only the final errors computed according to a loss function are transmitted back to the generating network for retraining, and thus the generating network updates its parameters, thereby completing training of the generating network. The whole training process continues all the time until the two enter into a balanced and harmonious state.

The result of training is the generating network with the high image generation quality and the discriminant unit with the strong image discrimination ability. The former can be used for machine creation (such as automatic drawing of "cat" and "dog"), and the latter can be used for machine sorting (such as automatic judging of "cat" and "dog").

According to the image generation method of the implementation of this application, as the initial input used to generate the image of the generating network, only one initial input (such as a constant with the dimension of $4 \times 4 \times 512$ shown in the figure) needs to be selected.

In addition, in the multi-stage image generation process of the generating network of the images, style control processing for affecting global information of the generated images is

added. In style control processing, pixel-instance normalization processing is performed on the images and include pixel normalization processing and instance normalization processing. In the process that the errors computed according to the loss function are transmitted back to the generating network for retraining, the ratio of affects of the pixel normalization processing and the instance normalization processing in the style control processing on image results can be adjusted.

It can be understood that the above description of the technical solution for image generation of the neural network model shown in Fig. 1 is only exemplary and not restrictive.

Fig. 2 shows a structural block diagram of electronic device 100 according to some embodiments of this application. Specifically, as shown in Fig. 2, the electronic device 100 includes one or more processors 104, a system control logic 108 connected with at least one of the processors 104, a system memory 112 connected with the system control logic 108, a nonvolatile memory (NVM) 116 connected with the system control logic 108, and a network interface 120 connected with the system control logic 108.

In some embodiments, each processor 104 may include one or more single-core or multi-core processors. In some embodiments, each processor 104 may include any combination of a general-purpose processor and a special-purpose processor (such as a graphics processor, an application processor and a baseband processor). In the embodiments that the electronic device 100 adopts an evolved node B (eNB) or radio access network (RAN) controller, the processors 104 can be configured to execute various corresponding embodiments.

In some embodiments, the processors 104 can be configured to generate images based on a neural network model. Specifically, for example, in the process of processing the images at each layer of the neural network model, the processors 104 continuously adjust the neural network model according to obtained image information until an image feature map with high accuracy is obtained. The processors 104 can further periodically update the above neural network model so as to better adapt to changes in various actual demands for the neural network model.

In some embodiments, the system control logic 108 may include any appropriate interface controller so as to provide any appropriate interface for at least one of the processors 104 and/or any appropriate device or component communicating with the system control logic 108.

In some embodiments, the system control logic 108 may include one or more memory controllers so as to provide an interface connected to the system memory 112. The system

memory 112 can be configured to load and store data and/or an instruction. In some embodiments, the system memory 112 of the electronic device 100 may include any appropriate NVM, such as an appropriate dynamic random access memory (DRAM). In some embodiments, the system memory 112 can be configured to load or store an instruction for
5 implementing image generation based on a neural network, or the system memory 112 can be configured to load or store an instruction for implementing an application program for image generation through the above neural network model. The processors 104 can implement the image generation method according to the embodiments of this application by operating the instruction.

10 The NVM/memory 116 may include one or more tangible and non-temporary computer readable media for storing data and/or instructions. In some embodiments, the NVM/memory 116 may include any appropriate NVM such as a flash memory and/or any appropriate nonvolatile storage device, such as at least one of a hard disk drive (HDD), a compact disc (CD) drive and a digital versatile disc (DVD) drive. The NVM/memory 116 may further be
15 configured to store the neural network model used in the above image generation process.

The NVM/memory 116 may include parts of storage resources on a apparatus provided with the electronic device 100, or it may be accessed by device, but not necessarily part of the device. For example, the NVM/memory 116 can be accessed through a network via the network interface 120.

20 In particular, the system memory 112 and the NVM/memory 116 may respectively include: a temporary copy and a permanent copy of an instruction 124 respectively. The instruction 124 may include: an instruction causing the electronic device 100 to implement the image generation method in this application when executed by at least one of the processors 104. In some embodiments, the instruction 124, hardware, firmware and/or software
25 components thereof can be additionally/substitutively arranged in the system control logic 108, the network interface 120 and/or the processors 104.

The network interface 120 may include a transceiver, which is configured to provide a wireless interface for the electronic device 100 and then to communicate with any other appropriate device (such as a front-end module and an antenna) through one or more networks.
30 In some embodiments, the network interface 120 can be integrated into other components of the electronic device 100. For example, the network interface 120 can be integrated into the processors 104, the system memory 112, the NVM/memory 116, and at least one of firmware apparatuses (not shown) with instructions. When at least one of the processors 104 executes

the instruction, the electronic device 100 implements the image generation method based on the neural network in this application.

The network interface 120 may further include any appropriate hardware and/or firmware so as to provide a multi-input and multi-output wireless interface. For example, the
5 network interface 120 may be a network adapter, a wireless network adapter, a telephone modem and/or a wireless modem.

In some embodiments, at least one of the processors 104 can be packaged with logics of one or more controllers for the system control logic 108 to form a system in package (SiP). In some embodiments, at least one of the processors 104 can be integrated on the same die
10 together with the logics of one or more controllers for the system control logic 108 to form a system on chip (SoC).

The electronic device 100 may further include: input/output (I/O) device 132. The I/O device 132 may include a user interface so that a user can interact with the electronic device 100. The design of a peripheral component interface also enables a peripheral component to
15 interact with the electronic device 100. In some embodiments, the electronic device 100 further includes a sensor, which is configured to determine at least one of an environment condition and position information relevant to the electronic device 100.

According to some embodiments of this application, an image generation method 300 based on a neural network is provided. As shown in Fig. 3, the method 300 includes the
20 following steps:

Step 301: receiving an initial input used to generate the image, further more, the initial input may be a constant.

Specifically, in step 301, because features of StyleGAN generated images are controlled by an intermediate vector and style control processing, the initial input can be ignored and is
25 substituted with the constant. Taking the constant as the initial input can bring the following technical effects: firstly, the probability of generating some abnormal images due to an improper value of the initial input can be reduced; and secondly, feature entanglement is reduced advantageously, and a network is easier to learn in the case where the intermediate vector is used only without depending on the entanglement with input vectors.

Step 302: based on the initial input, implementing M stages of image generation
30 processes to generate a first image with a preset target resolution.

Specifically, in step 302, in the M stages of image generation processes, the M stages from the first stage to the Mth stage are sequentially connected, and an output image of the previous stage serves as the input image of the current stage. An image resolution of each

stage is gradually increased compared with the previous stage until the image reaching the preset target resolution is output at the M^{th} stage. Specifically, the preset target resolution can be set to $1024*1024$.

Step 303: outputting the first image with a preset target resolution.

5 Specifically, in step 303, the generated first image reaching the preset target resolution is output. The first image serves as a result of the image generated by a generating network. By comparing with a second image input into a discrimination unit, the quality of the generated image can be improved continuously. Optionally, the first image with the preset target resolution of $1024*1024$ is output as the first image generated by the generating network.

10 Step 304: distinguishing the first image from the second image, and adjusting the generation process of the first image.

Specifically, in step 304, after preliminary training, the discrimination unit processes a real image, such as reducing the size, so that it conforms to an input of the discrimination unit and is regarded as the second image. When the generating network inputs the generated first
15 image into the discrimination unit as a sample image, the discrimination unit judges whether the image generated by the generating network belongs to a real image or a false image based on the difference between the first image and the second image. If the discrimination unit determines that the first image generated by the generating network is the false image, parameters of the generating network are adjusted to enable the generated false image to be
20 more realistic; if the the discrimination unit determines that the first image generated by the generating network is the real image, after the discrimination unit generates errors, the final errors computed according to a loss function are transmitted back to the generating network for retraining, and thus the generating network updates its parameters.

The specific process of image generation implemented in step 302 above is further
25 explained in combination with Fig. 4 below. As shown in Fig. 4, the process of image generation in step 302 can exemplary include the following steps:

Step 3021: encoding the input initial vector into an intermediate vector.

Specifically, in step 3021, a mapping network encodes the input vector input into the intermediate vector. If there is no mapping network, there will be a feature entanglement
30 phenomenon among N control vectors obtained later. For example, when the control vectors conduct style control processing on a resolution of $8*8$ (assuming that they can control the generation angle of a human face), control contents on a resolution of $32*32$, namely corresponding visual features (such as the skin color), will be found to be changed as well, and this phenomenon is called feature entanglement. The mapping network provides a

learning channel for feature entanglement of input vectors. The ability to control the visual features of generated images using only the input vectors is very limited, so it must follow the probability density of training data. For example, if images of black-haired people are more common in the dataset, more input values will be mapped to the feature. Therefore, a model cannot map parts of inputs (elements in the vectors) to other features, which will result in feature entanglement. However, by using the mapping network, the model can generate a vector that does not have to follow the distribution of the training data, and the correlation between the vectors can be reduced. The mapping network consists of 8 fully connected layers, and the intermediate vector output by the mapping network is the same as the initial vector input by an input layer in size, wherein the initial vector can be a 512×1 random vector.

Step 3022: Processing the intermediate vector into N control vectors.

Specifically, in step 3022, processing the intermediate vector as N control vectors. For example, when the resolution of the generated images is changed from 4×4 into 8×8 , changed gradually and finally changed to be 1024×1024 serving as a preset target resolution, it can be known that 9 generation stages are needed at the moment, namely $M=9$, in each stage, the image is affected by two corresponding control vectors, one of the control vectors affects the generated image once on a first end modulation layer behind an up-sampling layer, and the other control vector affects the generated image once on a second end modulation layer behind a convolutional layer, each of the first end modulation layer and the second end modulation layer comprises at least one style control processing, affecting modes are both style control processing, at the moment, $N=18$, that is, the intermediate vector is processed as 18 control vectors. The 18 control vectors respectively control 18 different visual features of the generated images, which can be divided into three types:

1. Rough visual features: when the resolution does not exceed 8×8 , the visual features such as gestures, general hair styles and face shapes are affected;
2. Medium visual features: when the resolution is 16×16 to 32×32 , the visual features such as finer face features, hair styles and opening or closing of eyes are affected; and
3. High-quality visual features: when the resolution is 64×64 to 1024×1024 , the visual features such as colors (the eyes, hair and skin) and microscopic features are affected.

In M stages, by taking advantage of at least one style control processing, N control vectors affect the images, and the N control vectors are configured to control visual features

corresponding to the images, wherein $N \geq M$. Specifically, N may be twice M , and as mentioned above, when $M=9$, $N=18$.

The above value of M is only an exemplary illustration according to the embodiment of the application. The value of M is determined by practical application, for example, other cases which can optionally generate $256*256$ (corresponding to $M=7$) or $512*512$ (corresponding to $M=8$) images as false faces.

Step 3023: N control vectors affecting the images by taking advantage of at least one style control processing

Specifically, in step 3023, the 18 control vectors are subjected to style control processing in the 9 image generation stages to affect the images, thus in the image generation process from the resolution of $4*4$ to the preset target resolution of $1024*1024$, the visual features of the images are refined from rough to high quality, artifacts are eliminated in the process, and the quality of the generated images is further improved.

Fig. 5 shows a structural block diagram of a neural network model according to some embodiments of this application. The neural network model consists of the input layer, the mapping network and the generating network, wherein the mapping network consists of 8 fully connected layers, the generating network consists of 9 network layers (including first to ninth network layers), resolutions of images generated by the 9 network layers are gradually increased from low resolution to high resolution. For example, the first network layer generates a $4*4$ image, the generated $4*4$ image is output to the second network layer, and after processing, the second network layer outputs a $8*8$ image, which is conducted in turn until the ninth network layer generates a $1024*1024$ image. Each network layer includes at least one time of style control processing for affecting global information of the generated images. As shown in Fig. 5, an initial input of 512×1 is input into the input layer firstly, then the initial input is input into the mapping network and decoded into an intermediate vector through the 8 fully connected layers in the mapping network, and the intermediate vector is processed as 18 control vectors for controlling visual features of the images, thereby participating in the image generation process that affects the generating network. The 18 control vectors affect the resolutions of the images through style control processing in the 9 network layers respectively to generate different visual features, and finally the image with the resolution being a preset target resolution of $1024*1024$ is generated to be output.

All the network layers in Fig. 5 correspond to all the image generation stages in the image generation process as mentioned above respectively, that is, when there are M image generation stages, correspondingly, there are M network layers are arranged.

Fig. 6 shows a schematic diagram of an architecture of network layers and style control processing thereof according to some embodiments of this application. According to the specific embodiments of this application, each network layer includes an up-sampling layer, a first end modulation layer, a convolutional layer and a second end modulation layer. The first end modulation layer and the second end modulation layer each include at least one time of style control processing (adaptive instance normalization, also called AdaIN). As shown in Fig. 6, taking style control processing behind the convolutional layers as an example, the specific processing flow is as follows: an intermediate vector performs weighted combination on pixel normalization processing and instance normalization processing through learned affine transformation to be subjected to weighted summation together with standardized convolution, and thus the process of affecting an original output x_i by control vectors is completed once. The reason why this affect way can realize style control is mainly that global information of images is affected by N control vectors (a mean value $\mu_c^{(l)}$ and a standard deviation $\sigma_c^{2(l)}$ after transformation) corresponding to the intermediate vector (noting that standardization erases visibility to local information of the images), and reservation of key information for generating a face is decided by the up-sampling layers and the convolutional layers. Therefore, the N control vectors corresponding to the intermediate vector realize style control processing of the images by affecting style information of the generated images.

According to the embodiments of this application, each of the M network layers includes two end modulation layers, and correspondingly, each network layer includes two times of style control processing. However, such setting only serves as the illustrative description, and this application is not limited to this.

In some embodiments, pixel-instance normalization processing includes: weighted combination is performed on the pixel normalization processing and the instance normalization processing, as shown in the following formula:

$$y = \rho \cdot y^p + (1 - \rho) \cdot y^l,$$

Wherein y represents the pixel-instance normalization processing; y^p represents the pixel normalization processing; y^l represents the instance normalization processing; and $\rho \in [0,1]^C$,

wherein ρ represents an affect degree of the pixel normalization processing on the images, and $1 - \rho$ represents an affect degree of the instance normalization processing on the images.

Specifically, input images are represented as $x \in R^{C \times H \times W}$, wherein C represents the number of channels, and H and W represent the height and width of the images respectively.

5 A scalar x_{chw} represents the chw^{th} element in x , wherein c represents an index of the channels, and hw represents space coordinates.

A pixel normalization operation (also called PN operation) has the effect of avoiding gradient explosion of a generator. Normalization is conducted along dimensions of the channels so that each position of a feature map can have a unit length. Specifically, the PN

10 operation $\eta^P(\bullet)$ can be represented as $y^P = \eta^P(x)$, wherein

$$y_{chw}^P = x_{chw} / \sqrt{\frac{1}{C} \sum_{i=1}^C x_{i hw}^2 + \varepsilon} \quad \forall c, h, w \quad (1)$$

An instance normalization operation (also called IN operation) is a normalization mode very widely applied to the generating network. It only normalize the two dimensions of height and width of the feature map and has a significant affect on the styles of the images.

15 Specifically, the IN operation $\eta^I(\bullet)$ can be represented as $y^I = \eta^I(x)$, wherein

$$y_{chw}^I = \frac{x_{chw} - \mu_c^{(I)}}{\sqrt{\sigma_c^{2(I)} + \varepsilon}} \quad (2)$$

Wherein a mean value $\mu_c^{(I)}$ and a standard deviation $\sigma_c^{2(I)}$ can be represented as

$$\mu_c^{(I)} = \frac{1}{HW} \sum_{h=1}^H \sum_{w=1}^W x_{chw} \quad (3)$$

$$\sigma_c^{2(I)} = \frac{1}{HW} \sum_{h=1}^H \sum_{w=1}^W (x_{chw} - \mu_c^{(I)})^2 \quad (4)$$

20 So-called pixel-instance normalization processing (also called PIN operation) of the present invention can be represented as $y = \eta^{PI}(x)$, wherein weighted combination is performed on pixel normalization processing and instance normalization processing and is as follows:

$$y = \rho \bullet y^P + (1 - \rho) \bullet y^I \quad (5)$$

25 $\rho \in [0,1]^C$, wherein ρ represents an affect degree of the pixel normalization processing on the images, $1 - \rho$ represents an affect degree of the instance normalization processing on the images, the magnitude of ρ is trainable, $\bullet$ represents a product of information

channels, and subsequent style modification layers are reserved as affine transformation. After the pixel-instance normalization processing, an output is $y' = \gamma \bullet y + \beta$, wherein $\gamma, \beta \in R^C$, as those skilled in the art know, are all learnable parameters.

Therefore, the pixel-instance normalization processing in the style control processing is the weighted combination of the pixel normalization processing and the instance normalization processing, when $\rho=0$, only the instance normalization processing is conducted; and when $\rho=1$, only the pixel normalization processing is conducted. The affect degrees of different normalization processing on the images are different, and thus the quality of the generated images is different. For example, when artifacts are generated, the images are seriously degraded, the quality of the images is significantly reduced, and by using pixel-instance normalization processing in style control, the artifacts generated in the image generation process can be successfully eliminated, thereby improving the visual quality of the generated images.

Researches show that almost all facial images generated by the StyleGAN contain such artifacts. Fig. 7 shows a schematic diagram of an artifact generation process according to some embodiments of this application. Taking facial image generation as an example, as shown in the first line in Fig. 7, in the early stage of image generation, image generation includes channels with the artifacts, while other channels synthesize facial image features. Then, as shown in the second line in Fig. 7, the artifacts with high pixels are propagated into the following generated feature graphs along with step-by-step increase of image resolutions. These feature graphs affect generation of a large number of visual features in subsequent layers, and the artifacts are generally distributed randomly without regularity. Finally, as shown in the third line in Fig. 7, with the increase of the image generation resolutions, the sizes of the artifacts are magnified and are more prominent than a face area, and thus the artifacts are generated in the finally generated images.

For this purpose, a neural network is trained based on a Flickr-Faces-HQ dataset, and the neural network used for image generation is called a StyleGAN-FFHQ network model. Specifically, the StyleGAN-FFHQ network model is modified to test which component referenced in the StyleGAN causes the generation of the artifacts compared with a PRGAN. As shown in Table 1, test results show that the instance normalization of style control processing in StyleGAN leads to artifacts in the image generation process. Therefore, the present invention applies the pixel normalization processing to the style control processing of the StyleGAN for weighted combination with the instance normalization processing, and the

configuration of the StyleGAN is initialized so as to train the StyleGAN-FFHQ through the pixel-instance normalization processing. Test results show that images generated by an improved the StyleGAN-FFHQ model show better visual quality without generating any artifact.

5 Table 1

Test	Artifacts ?
Use CelebA dataset	Yes
Remove mapping network	Yes
Use traditional input rather than a constant	Yes
Remove random noise	Yes
Remove blurring	Yes
Remove style transfer	Yes
Use pixel normalization processing	No

In some embodiments, in the pixel-instance normalization processing, ρ is trainable and the value of ρ is adjusted according to the train automatically.

10 Compared with an image processing method in the prior art, the technical solutions in this application enable the generated images to have better visual quality, meanwhile, no artifact is generated, and the overall quality of the generated images is improved. In the process of implementing image generation, only simple solutions are required for eliminating the artifacts without the undue loss of image quality, thus computational accuracy is significantly improved, and hardware configurations are optimized better, so that internal
15 structures of a computer system are more simplified.

According to some embodiments of this application, an image generation apparatus 400 based on a neural network is provided. Fig. 8 shows a schematic structural diagram of the image generation apparatus based on the neural network according to some embodiments of this application. As shown in Fig. 8, the image generation apparatus 400 based on the neural
20 network includes:

a receiving unit 401, receiving an initial input used to generate the image; and

a generation unit 402, including M generation subunits, wherein based on the initial input, the M generation subunits sequentially implement image generation processes to generate the image with a preset target resolution.

In the M generation subunits, a resolution of an image generated based on the initial input is gradually increased compared with the previous stage from the first generation subunit to the Mth generation subunit until the image reaching the preset target resolution is output in the Mth generation subunit.

5 Each generation subunit includes at least one style control module for affecting global information of the generated images, the style control module performs pixel-instance normalization processing on the images, and pixel-instance normalization processing includes a pixel normalization processing and an instance normalization processing.

10 In some embodiments, the pixel-instance normalization processing includes: weighted combination performed on the pixel normalization processing and the instance normalization processing, as shown in the following formula:

$$y = \rho \cdot y^p + (1 - \rho) \cdot y^l,$$

wherein y represents the pixel-instance normalization processing; y^p represents the pixel normalization processing; y^l represents the instance normalization processing; and $\rho \in [0,1]^C$, wherein ρ represents an affect degree of the pixel normalization processing on the images, and $1 - \rho$ represents an affect degree of the instance normalization processing on the images.

In some embodiments, based on an input initial vector, implementing the style control processing for the M stages of image generation processes.

20 In some embodiments, encoding the initial vector into an intermediate vector through a mapping network in the neural network model, and in M stages, processing the intermediate vector as N control vectors; and the N control vectors are configured to control visual features corresponding to the images, and by taking advantage of at least one style control processing, N control vectors affect the images, to control M stages of the image generation processes, 25 wherein $N \geq M$.

According to some embodiments of this application, another image generation apparatus 500 based on a neural network is provided. Fig. 9 shows a schematic structural diagram of the another image generation apparatus based on the neural network according to some embodiments of this application. As show in Fig. 9, the image generation apparatus 500 based 30 on the neural network includes:

- a first generation unit 501, generating a first image;
- an obtaining unit 502, obtaining a second image serving as a training sample; and

a discrimination unit 503, inputting the first image and the second image to distinguish the first image from the second image, and adjusting the generation process of the first image based on the result of the distinction;

wherein the first generation unit 501 includes:

5 a receiving unit, receiving an initial input used to generate the image;

a generation unit, including M generation subunits, wherein based on the initial input, the M generation subunits sequentially implement image generation processes to generate the image with a preset target resolution.

10 In the M generation subunits, a resolution of an image generated based on the initial input is gradually increased compared with the previous stage from the first generation subunit to the Mth generation subunit until the image reaching the preset target resolution is output in the Mth generation subunit.

Each generation subunit includes at least one style control module for affecting global information of the generated images, and, during the style control module performs pixel-instance normalization processing on the images, and pixel-instance normalization processing includes a pixel normalization processing and an instance normalization processing.

15 In some embodiments, the pixel-instance normalization processing includes: weighted combination performed on the pixel normalization processing and the instance normalization processing, as shown in the following formula:

$$y = \rho \bullet y^p + (1 - \rho) \bullet y^l,$$

20 wherein y represents the pixel-instance normalization processing; y^p represents the pixel normalization processing; y^l represents the instance normalization processing; and $\rho \in [0,1]^C$, wherein ρ represents an affect degree of the pixel normalization processing on the images, and $1 - \rho$ represents an affect degree of the instance normalization processing on the images.

In some embodiments, in the pixel-instance normalization processing, ρ is trainable and the value of ρ is adjusted according to the train automatically.

30 According to some embodiments of this application, image generation device based on a neural network. It can be understood that electronic device for image generation based on the neural network corresponds to an image generation method based on the neural network and provided by this application. The above technical details in the specific description of the image generation method based on the neural network and provided by this application are

still applicable to the electronic device for image generation based on the neural network. Please refer to the above for the specific description, which will not be described in detail.

According to some embodiments of this application, a computer readable storage medium encoded through a computer program is provided. It can be understood that the computer readable storage medium encoded through the computer program corresponds to an image generation method based on a neural network and provided by this application. The above technical details in the specific description of the image generation method based on the neural network and provided by this application are still applicable to the computer readable storage medium encoded through the computer program. Please refer to the above for the specific description, which will not be described in detail.

It can be understood that exemplary application for image generation based on the neural network and provided by the embodiments of this application includes, but is not limited to, image generation in the field of artificial intelligence.

Fig. 10 shows a schematic structural diagram of electronic device 900 according to some embodiments of this application. The electronic device 900 can also execute the image generation method disclosed by the above embodiments of this application. In Fig. 10, similar parts have the same drawing labels. As shown in Fig. 10, the electronic device 900 may include a processor 910, a power module 940, a memory 980, a mobile communication module 930, a wireless communication module 920, a sensor module 990, an audio module 950, a camera 970, an interface module 960, a key 901, a display screen 902 and the like.

It can be understood that the structure indicated in the embodiments of this application does not constitute a specific qualification for the electronic device 900. In other embodiments of this application, the electronic device 900 may include more or fewer parts than shown in the figure, or combinations of some parts, or disassembly of some parts, or different arrangements of the parts. The parts shown in the figure can be implemented through hardware, software, or combinations of the software and the hardware.

The processor 910 may include one or more processing units, for example, may include processing units or processing circuits such as a central processing unit (CPU), a graphics processing unit (GPU), a digital signal processor (DSP), a micro-programmed control unit (MCU), an artificial intelligence (AI) processor or a field programmable gate array (FPGA). Different processing units may be independent apparatuses and may also be integrated into one or more processors. A memory unit can be arranged in the processor 910 to store an instruction and data. In some embodiments, the memory unit in the processor 910 is a cache memory 980. The memory 980 mainly includes a program storage area 9801 and a data

storage area 9802, wherein the program storage area 9801 can store an operation system and an application program required for at least one function (such as audio playing and image recognition). The image generation method based on the neural network and provided by the embodiments of this application can be regarded as the application program capable of realizing image generation and other functions in the program storage area 9801. A neural network model is stored in the above data storage area 9802.

The power module 940 may include a power supply, a power management part and the like. The power supply may be a battery. The power management part is configured to manage charging of the power supply and power supplying of the power supply to other modules. In some embodiments, the power management part includes a charging management module and a power management module. The charging management module is configured to receive a charging input from a charger. The power management module is configured to be connected with the power supply. The charging management module is connected with the processor 910. The power management module receives inputs of the power supply and/or the charging management module to supply power to the processor 910, the display screen 902, the camera 970, the wireless communication module 920 and the like.

The mobile communication module 930 may include, but is not limited to, an antenna, a power amplifier, a filter, a low noise amplify (LNA) and the like. The mobile communication module 930 can receive electromagnetic waves through the antenna and conducts processing such as filtering and amplifying on the received electromagnetic waves, and the electromagnetic waves are transmitted to a modulation-demodulation processor for demodulation. The mobile communication module 930 can further amplify signals modulated through the modulation-demodulation processor, and the signals are converted into electromagnetic waves through the antenna to be radiated out. In some embodiments, at least parts of functional modules of the mobile communication module 930 can be arranged in the processor 910. In some embodiments, at least parts of functional modules of the mobile communication module 930 can be arranged in the same apparatus together with at least parts of modules of the processor 910. A wireless communication technology may include a global system for mobile communications (GSM), general packet radio service (GPRS), code division-multiple access (CDMA), wideband code division multiple access (WCDMA), time-division code division multiple access (TD-SCDMA), long term evolution (LTE), bluetooth (BT), a global navigation satellite system (GNSS), wireless local area networks (WLAN), near field communication (NFC), frequency modulation (FM), and/or an infrared (IR) technology and the like. The GNSS may include a global positioning system (GPS), a

global navigation satellite system (GLONASS), a beidou navigation satellite system (BDS), a quasi-zenith satellite system (QZSS) and/or satellite based augmentation systems (SBAS).

The wireless communication module 920 may include an antenna, and receives and transmits electromagnetic waves through the antenna. The wireless communication module
5 920 can provide solutions of wireless communications applied to the electronic device 900 and including wireless local area networks (WLAN) (such as wireless fidelity (Wi-Fi) networks), bluetooth (BT), a global navigation satellite system (GNSS), frequency modulation (FM), a near field communication (NFC) technology, an infrared (IR) technology and the like. The electronic device 900 can communicate with networks and other device through the
10 wireless communication technology.

In some embodiments, the mobile communication module 930 and the wireless communication module 920 of the electronic device 900 can also be located in the same module.

The display screen 902 is configured to display a man-machine interaction interface, an
15 image, a video and the like. The display screen 902 includes a display panel. The display panel can adopt a liquid crystal display (LCD), an organic light-emitting diode (OLED), an active-matrix organic light emitting diode (AMOLED), a flex light-emitting diode (FLED), a mini-led, a micro-Led, a micro-oled, quantum dot light emitting diodes (QLED) and the like.

The sensor module 990 may include a proximity optical sensor, a pressure sensor, a gyro
20 sensor, an air pressure sensor, a magnetic sensor, an acceleration sensor, a range sensor, a fingerprint sensor, a temperature sensor, a touch sensor, an ambient light sensor, a bone conduction sensor and the like.

The audio module 950 is configured to convert digital audio information into an analog audio signal for output, or to convert an analog audio input into a digital audio signal. The
25 audio module 950 can further be configured to encode and decode the audio signal. In some embodiments, the audio module 950 can be arranged in the processor 910, or parts of functional modules of the audio module 950 are arranged in the processor 910. In some embodiments, the audio module 950 may include a loudspeaker, a receiver, a microphone and a headphone interface.

30 The camera 970 is configured to capture static images or videos. An object generates an optical image through a lens to be projected onto a photosensitive element. The photosensitive element converts an optical signal into an electric signal and then transmits the electric signal to image signal processing (ISP) to be converted into a digital image signal. The electronic

device 900 can realize a shooting function through ISP, the camera 970, a video codec, a graphic processing unit (GPU), the display screen 902, an application processor and the like.

The interface module 960 includes an external memory interface, a universal serial bus (USB) interface, a subscriber identification module (SIM) card interface and the like. The external memory interface can be configured to be connected with an external memory card, such as a Micro SD card, to expand the storage capacity of the electronic device 900. The external memory card communicates with the processor 910 through the external memory interface to realize a data storage function. The USB interface is used for communication of the electronic device 900 with other electronic device. The SIM card interface is configured to communicate with an SIM card installed in the electronic device 900, such as reading phone numbers stored in the SIM card or writing the phone numbers into the SIM card.

In some embodiments, the electronic device 900 further include the key 901, a motor, an indicator and the like. The key 901 may include a volume button, an on/off button and the like. The motor is configured to enable the electronic device 900 to generate a vibration effect, such as generating vibration when the electronic device 900 of a user is called to remind the user to answer a call from the electronic device 900. The indicator may include a laser indicator, a radio-frequency indicator, an LED indicator and the like.

All the embodiments of the mechanism disclosed by this application can be implemented in hardware, software, firmware or a combination of these implementation methods. The embodiments of this application can be implemented as a computer program executed on a programmable system, or program codes. The programmable system includes at least one processor, a memory system (including a volatile and nonvolatile memory and/or a memory element), at least one input apparatus and at least one output apparatus.

The program codes can be applied to an input instruction to execute all functions described by this application and generate output information. The output information can be applied to the one or more output apparatuses in a known manner. For the purpose of this application, a processing system includes any system provided with processors such as the DSP, a micro-controller, an application-specific integrated circuit (ASIC) or a microprocessor.

The program codes can be implemented in a high-level programmed language or a programming language facing an object so as to communicate with the processing system. The program codes can also be implemented in an assembly language or a machine language as needed. In fact, the mechanism described in this application is not limited to the scope of

any particular programming language. In any case, the language can be a compiled language or an interpreted language.

In some cases, the disclosed embodiments can be implemented through the hardware, the firmware, the software, or any combination thereof. The disclosed embodiments can further
5 be implemented as instructions borne by or stored in one or more temporary or non-temporary machine readable (such as computer readable) storage media, which can be read and executed by the one or more processors. For example, the instructions can be distributed through networks or other computer readable media. Therefore, the machine readable media may include any mechanism configured to store or transmit information in a machine (such as a
10 computer) readable form, which include, but is not limited to, a floppy disk, a light disk, an optical disk, read-only memories (CD-ROMs), a magneto-optical disk, a read-only memory (ROM), a random access memory (RAM), an erasable programmable read-only memory (EPROM), an electrically erasable programmable read-only memory (EEPROM), a magnetic card or an optical card, a flash memory, or a tangible machine readable memory for
15 transmitting information through propagation signals (such as a carrier signal, an infrared signal and a digital signal) in electricity, light, sound or other forms through the Internet. Therefore, the machine readable media include any type of machine readable media suitable for storing or transmitting electronic instructions or information in a machine (such as the computer) readable form.

In the drawings, some structural or method features can be shown in a particular arrangement and/or order. However, it should be understood that such a specific arrangement and/or order may not be required. Rather, in some embodiments, these features can be arranged in a manner and/or order different from that shown in the illustrative drawings. In addition, the inclusion of the structural or method features in a particular figure does not
20 imply that such features are required in all embodiments, and in some embodiments, these features may be excluded or may be combined with other features.

It should be noted that all units/modules mentioned in all the device embodiments of this application are logic units/modules. Physically, a logic unit/module may be a physical unit/module, may be part of the physical unit/module, and may also be implemented as a
30 combination of a plurality of physical units/modules. The physical implementation of these logic units/modules themselves is not the most important, and the combination of functions realized by these logic units/modules is the key to solving the technical problems presented in this application. In addition, in order to highlight the innovative part of this application, the above device embodiments of this application do not introduce units/modules that are not

closely related to solving the technical problems presented in this application, which does not indicate that there are no other units/modules in the above device embodiments.

It should be noted that in the examples and specifications of this application, relationship terms such as first and second classes are used only to distinguish one entity or operation from another entity or operation and do not necessarily require or imply any such actual relationship or sequence between these entities or operations. Furthermore, The term "including", "containing" or any of its other variants is intended to cover non-exclusive containing, so that a process, method, article or device including a set of elements includes not only those elements but also other elements not explicitly listed, or also includes elements inherent to the process, method, article or device. In the absence of further restrictions, the elements defined by the statement "including a" does not preclude the existence of additional identical elements in the process, method, article or device that includes the elements.

Although this application has been illustrated and described by reference to some preferred embodiments of this application, the ordinary skilled in the art should be aware that various changes may be made to it in form and detail without deviating from the spirit and scope of this application.

Claims

1. An image generation method based on a neural network, characterized by comprising:
receiving an initial input used to generate the image; and
based on the initial input, implementing M stages of image generation processes to
5 generate the image with a preset target resolution;
wherein, in the M stages of image generation processes, a resolution of the image
generated in each stage based on the initial input is gradually increased compared with the
previous stage from the first stage to the Mth stage until the image reaching the preset target
resolution is output in the Mth stage; and
10 the each stage of image generation process comprises at least one style control
processing for affecting global information of the generated image, and, during the style
control processing, pixel-instance normalization processing is performed on the image, and
the pixel-instance normalization processing comprises a pixel normalization processing and
an instance normalization processing.
15
2. The method according to claim 1, characterized in that, the pixel-instance normalization
processing comprises:
performing weighted combination on the pixel normalization processing and the instance
normalization processing, as shown in a following formula:
20
$$y = \rho \cdot y^p + (1 - \rho) \cdot y^I,$$

wherein y represents the pixel-instance normalization processing; y^p represents the pixel
normalization processing; y^I represents the instance normalization processing; and $\rho \in [0,1]^C$,
wherein ρ represents an affect degree of the pixel normalization processing on the images,
and $1 - \rho$ represents an affect degree of the instance normalization processing on the images.
25
3. The method according to claim 1, characterized in that, the initial input is a constant.
4. The method according to claim 1, characterized in that, based on an input initial vector,
implementing the style control processing for the M stages of image generation processes.
30
5. The method according to claim 4, characterized in that, based on an input initial vector,
implementing the style control processing for the M stages of image generation processes
comprises:

encoding the initial vector into an intermediate vector through a mapping network in the neural network model, and in M stages, processing the intermediate vector as N control vectors; and the N control vectors are configured to control visual features corresponding to the images, and by taking advantage of at least one style control processing, N control vectors
5 affect the images, to control M stages of the image generation processes, wherein $N \geq M$.

6. The method according to claim 5, characterized in that, the initial vector is a random vector.

7. The method according to claims 1-6, characterized in that, the neural network model
10 comprises an input layer, the mapping network, a generating network and an output layer, and the generating network further comprises M network layers;

wherein at least one of the M network layers comprises an up-sampling layer, a first end modulation layer, a convolutional layer and a second end modulation layer, each of the first end modulation layer and the second end modulation layer comprises at least one style control
15 processing, and the initial input passes through the M network layers to generate the images.

8. An image generation method based on a neural network, characterized by comprising:

generating a first image;

obtaining a second image serving as a training sample; and

20 inputting the first image and the second image into a discrimination unit to distinguish the first image from the second image, and adjusting the generation process of the first image based on the result of the distinction;

wherein the process of generating the first image comprises:

receiving an initial input used to generate the image; and

25 based on the initial input, implementing M stages of image generation processes to generate the first image with a preset target resolution;

wherein, in the M stages of image generation processes, a resolution of an image generated in each stage based on the initial input is gradually increased compared with the previous stage from the first stage to the M^{th} stage until the first image reaching the preset
30 target resolution is output in the M^{th} stage; and

the each stage of image generation process comprises at least one style control processing for affecting global information of the generated image, and, during the style control processing, pixel-instance normalization processing is performed on the image, and

the pixel-instance normalization processing comprises a pixel normalization processing and an instance normalization processing.

9. The method according to claim 8, characterized in that the pixel-instance normalization processing comprises:

performing weighted combination on the pixel normalization processing and the instance normalization processing, as shown in a following formula:

$$y = \rho \cdot y^p + (1 - \rho) \cdot y^l,$$

wherein y represents the pixel-instance normalization processing; y^p represents the pixel normalization processing; y^l represents the instance normalization processing; and $\rho \in [0,1]^c$, wherein ρ represents an affect degree of the pixel normalization processing on the images, and $1 - \rho$ represents an affect degree of the instance normalization processing on the images.

10. The method according to claim 9, characterized in that in the pixel-instance normalization processing, the value of ρ is adjusted according to the train automatically.

11. An image generation apparatus based on a neural network, characterized by comprising:

a receiving unit, receiving an initial input used to generate the image; and

a generation unit, comprising M generation subunits, wherein based on the initial input, the M generation subunits sequentially implement image generation processes to generate the image with a preset target resolution;

wherein, in the M generation subunits, a resolution of the image generated based on the initial input is gradually increased compared with the previous stage from the first generation subunit to the M^{th} generation subunit until the image reaching the preset target resolution is output in the M^{th} generation subunit; and

the each of the generation subunits comprises at least one style control module for affecting global information of the generated image, and, the style control module performs pixel-instance normalization processing on the image, and the pixel-instance normalization processing comprises a pixel normalization processing and an instance normalization processing.

12. The apparatus according to claim 11, characterized in that, the pixel-instance normalization processing comprises:

performing weighted combination on the pixel normalization processing and the instance normalization processing, as shown in a following formula:

$$y = \rho \cdot y^p + (1 - \rho) \cdot y^l,$$

wherein y represents the pixel-instance normalization processing; y^p represents the pixel normalization processing; y^l represents the instance normalization processing; and $\rho \in [0,1]^C$, wherein ρ represents an affect degree of the pixel normalization processing on the images, and $1 - \rho$ represents an affect degree of the instance normalization processing on the images.

13. The apparatus according to claim 11, characterized in that, based on an input initial vector, implementing the style control processing for the M stages of image generation processes.

14. The apparatus according to claim 13, characterized in that, based on an input initial vector, implementing the style control processing for the M stages of image generation processes comprises:

15 encoding the initial vector into an intermediate vector through a mapping network in the neural network model, and in M stages, processing the intermediate vector as N control vectors;and

the N control vectors are configured to control visual features corresponding to the images, and by taking advantage of at least one style control processing, N control vectors affect the images,to control M stages of the image generation processes, wherein $N \geq M$.

15. The apparatus according to claims 11-14, characterized in that, the neural network model comprises an input layer, the mapping network, a generating network and an output layer, and the generating network further comprises M network layers;

25 wherein at least one of the M network layers comprises an up-sampling layer, a first end modulation layer, a convolutional layer and a second end modulation layer, the first end modulation layer and the second end modulation layer each comprise at least one style control processing, and the initial input passes through the M network layers to generate the images.

30 16. An image generation apparatus based on a neural network, characterized by comprising:
a first generation unit, generating a first image;
an obtaining unit, obtaining a second image serving as a training sample; and
a discrimination unit, inputting the first image and the second image to distinguish the

first image from the second image, and adjusting the generation process of the first image based on the result of the distinction;

wherein the first generation unit comprises:

a receiving unit, receiving an initial input used to generate the image;

5 a generation unit, comprising M generation subunits, wherein based on the initial input, the M generation subunits sequentially implement image generation processes to generate the image with a preset target resolution;

Wherein, in the M generation subunits, a resolution of an image generated based on the initial input is gradually increased compared with the previous stage from the first
10 generation subunit to the Mth generation subunit until the image reaching the preset target resolution is output in the Mth generation subunit; and

the each of the generation subunits comprises at least one style control module for affecting global information of the generated images, and, the style control module performs pixel-instance normalization processing on the images, and the pixel-instance normalization
15 processing comprises a pixel normalization processing and an instance normalization processing.

17. The apparatus according to claim 16, characterized in that the pixel-instance normalization processing comprises:

20 performing weighted combination on the pixel normalization processing and the instance normalization processing, as shown in a following formula:

$$y = \rho \bullet y^p + (1 - \rho) \bullet y^l,$$

wherein y represents the pixel-instance normalization processing; y^p represents the pixel normalization processing; y^l represents the instance normalization processing; and $\rho \in [0,1]^c$,

25 wherein ρ represents an affect degree of the pixel normalization processing on the images, and $1 - \rho$ represents an affect degree of the instance normalization processing on the images.

18. The apparatus according to claim 16, characterized in that in the pixel-instance normalization processing, the value of ρ is adjusted according to the train automatically.

30

19. Image generation device based on a neural network, characterized by comprising:

a memory, configured to store an instruction executed by one or more processors of a system, and

a processor, serving as one of the processors of the system and configured to execute the instruction to implement the image generation method of any one of claims 1-10.

20. A computer readable storage medium encoded through a computer program, characterized
5 in that an instruction is stored on the computer readable medium, and when the instruction is executed on a computer, the computer executes the image generation method of any one of claims 1-10.