IMAGE PROCESSING METHOD AND APPARATUS, TERMINAL, AND STORAGE MEDIUM

TECHNICAL FIELD

[0001] The disclosure relates to the technical field of image processing, and relates, but is not limited, to an image processing method and apparatus, a terminal, and a storage medium.

BACKGROUND

[0002] Image stylization based on samples aims to achieve that the style of a rendered image is consistent with the style of a specified art image sample. It is widely applied in daily production and life, such as avatar design and animation poster advertising production. In the related art, when an image generation model is used to render style images, the image generation model can only learn the overall style of a data set, and can hardly accurately simulate the artistic portrait style of a specified sample in the data set.

SUMMARY

[0003] The embodiments of the disclosure provide an image processing method and apparatus, a terminal, and a storage medium.

[0004] In a first aspect, an embodiment of the disclosure provides a training method of an image generation model. The image generation model includes a trained first generator and an extrinsic style control module to be trained. The first generator is pre-trained with a real sample set and is capable of generating a real image based on an intrinsic style code.

[0005] The method includes that: a style sample set and the real sample set are obtained; the extrinsic style control module is trained based on the real sample set so that a first fused image generated by the image generation module may reconstruct a second fused image generated by the first generator; and after training of the extrinsic style control module is completed, the image generation model is trained based on the style sample set and a randomly sampled intrinsic style code till a first convergence condition is met.

[0006] In a second aspect, an embodiment of the disclosure provides an image processing method. The method includes that: a style image and a content image to be fused are obtained; extraction is performed on the content image to obtain an intrinsic intermediate style code of the content image; extraction is performed on the style image to obtain an extrinsic intermediate style code of the style image; and a fused image is generated based on the intrinsic intermediate style code of the content image and the extrinsic intermediate style code of the style image.

[0007] In a third aspect, an embodiment of the disclosure provides a training apparatus of an image generation module. The image generation model includes a trained first generator and an extrinsic style control module to be trained. The first generator is

pre-trained with a real sample set and may generate a real image based on an intrinsic style code. The apparatus includes: a first obtaining module configured to obtain a style sample set and the real sample set; a first training module configured to train the extrinsic style control module based on the real sample set so that a first fused image generated by the image generation module may reconstruct a second fused image generated by the first generator; and a second training module configured to, after training of the extrinsic style control module is completed, train the image generation model based on the style sample set and a randomly sampled intrinsic style code till a first convergence condition is met.

[0008] In a fourth aspect, an embodiment of the disclosure provides an image processing apparatus. The apparatus includes: a first obtaining module configured to obtain a style image and a content image to be fused; a first extraction module configured to perform extraction on the content image to obtain an intrinsic intermediate style code of the content image; a second extraction module configured to perform extraction on the style image to obtain an extrinsic intermediate style code of the style image; and a generation module configured to generate a fused image based on the intrinsic intermediate style code of the content image and the extrinsic intermediate style code of the style image.

[0009] In a fifth aspect, an embodiment of the disclosure provides a computer device, which includes: a memory and a processor. The memory stores a computer program capable of running in the processor. The processor executes the computer program to implement the steps in the above method.

[0010] In a sixth aspect, an embodiment of the disclosure provides a computer storage medium storing one or more programs, which are executable by one or more processors to implement the steps in the above method.

[0011] In the training method of the image generation model provided by the embodiment of the disclosure, on one hand, the image generation model is trained by using a progressive transfer learning method. The progressive transfer learning method includes two stages. Firstly, in the first stage, the extrinsic style control module is trained with a real sample data set, and the first generator is pre-trained with the real sample set, so that the style transfer in the real sample set can be realized. Taking a StyleGAN as the first generator as an example, a real face sample set is used to train the extrinsic style control module to realize the style transfer in real face data. Secondly, in the second stage, the image generation model, especially the extrinsic style control module, is trained by using the style sample set to realize the style transfer of the extrinsic style control module between face data and style data. On the other hand, the first generator in the image generation model is pre-trained, so that the image generation model may transfer and learn with a small data set to generate a high-resolution image, which can effectively reduce the problem of requiring a large amount of training data due to learning of the mapping relationship between real images and style images.

[0012] In the related art, the first generator, such as a StyleGAN, can only render the overall style of a single field, and can hardly render the specified sample style. The image generation model obtained by the above training method introduces the control on the extrinsic style by constructing the extrinsic style control module, so that the image generation model can control the intrinsic style and the extrinsic style at the same time, which is more suitable for processing a stylization task based on samples.

BRIEF DESCRIPTION OF THE DRAWINGS

[0013] In order to describe the technical solutions of the embodiments of the disclosure more clearly, the drawings required for the descriptions of the embodiments will be simply introduced below. It is apparent that the drawings described below are only some embodiments of the disclosure. Other drawings may further be obtained by those of ordinary skill in the art according to these drawings without creative work.

[0014] FIG. 1A illustrates a composition structure diagram of a StyleGAN model.

[0015] FIG. 1B illustrates a flowchart of an image processing method provided in an embodiment of the disclosure.

[0016] FIG. 2A illustrates an implementation flowchart of S200 in an embodiment of the disclosure.

[0017] FIG. 2B illustrates an implementation flowchart of S210 in an embodiment of the disclosure.

[0018] FIG. 2C illustrates an implementation flowchart of S213 in an embodiment of the disclosure.

[0019] FIG. 2D illustrates an implementation flowchart of S230 in an embodiment of the disclosure.

[0020] FIG. 3 illustrates an implementation flowchart of S310 in an embodiment of the disclosure.

[0021] FIG. 4 illustrates a composition structure diagram of an image generation model provided in an embodiment of the disclosure.

[0022] FIG. 5 illustrates an implementation flowchart of an image processing method provided in an embodiment of the disclosure.

[0023] FIG. 6A illustrates an implementation flowchart of an image processing method provided in an embodiment of the disclosure.

[0024] FIG. 6B illustrates an implementation flowchart of an image processing method provided in an embodiment of the disclosure.

[0025] FIG. 7 illustrates a schematic diagram of a hardware entity of a computer device provided in an embodiment of the disclosure.

DETAILED DESCRIPTION

[0026] In order to make the purpose, technical solutions, and advantages of the embodiments of the disclosure clearer, the technical solutions in the embodiments of the disclosure will be clearly and completely described below in combination with the drawings in the embodiments of the disclosure. It is apparent that the described embodiments are not all embodiments but part of embodiments of the disclosure. The following embodiments are adopted to describe the disclosure rather than limit the scope of the disclosure. All other embodiments obtained by those of ordinary skill in the art on the basis of the embodiments in the disclosure without creative work shall fall within the

scope of protection of the disclosure.

[0027] "Some embodiments" involved in the following descriptions describes a subset of all possible embodiments. However, it can be understood that "some embodiments" may be the same subset or different subsets of all the possible embodiments, and may be combined without conflicts.

[0028] It is to be pointed out that term "first/second/third" involved in the embodiments of the disclosure is only for distinguishing similar objects and does not represent a specific sequence of the objects. It can be understood that "first/second/third" may be interchanged to specific sequences or orders if allowed to implement the embodiments of the disclosure described herein in sequences except the illustrated or described ones.

[0029] Those skilled in the art can understand that, unless otherwise defined, all the terms (including technical terms and scientific terms) used herein have the same meanings usually understood by those of ordinary skill in the art of the embodiments of the disclosure. It should also be understood that terms defined in, for example, a general dictionary, should be understood to have the same meanings as those in the context of the conventional art, and may not be explained as idealized or too formal meanings, unless otherwise specifically defined like those here.

[0030] Before the embodiments of the disclosure are described, the nouns and terms involved in the embodiments of the disclosure are described. The nouns and terms involved in the embodiments of the disclosure are applicable to the following explanation.

[0031] An Adaptive Instance Normalization (AdaIN) layer is a network module. The module is configured to normalize image features according to the channel dimension, so that the mean and variance of the image features in each channel are normalized to be between 0 and 1. The normalized mean and variance are adjusted to the mean and variance of a target style, so as to realize transfer of the style.

[0032] A latent code is a latent variable sample space.

[0033] Perceptual loss is used for evaluating the difference between a feature image of an original image and a feature image of a reconstructed image.

[0034] A Style-Based Generator Architecture for Generative Adversarial Networks (StyleGAN) is a style-based generative network capable of mapping a random style code into a face image. The StyleGAN may convert a generation space of the generative network from the face field to the artistic portrait field through hundreds of artistic portrait images and hours of training time.

[0035] An intrinsic style may be understood as a content style of the image, such as a real face style of the pre-trained StyleGAN.

[0036] An extrinsic style may be understood as a pattern and a color style of the image, such as an artistic portrait style composed of a small-scale artistic portrait data set.

[0037] Portrait stylization is the process of converting a sample artistic portrait style into a target face. It is widely applied in industries related to comics, animation, posters and advertising.

[0038] A content image, namely an image that provides content in the stylization image process, is generally an image input by a user, such as a real face image. In the

process described later, the capital letter I is used to represent the content image. In different contexts, I may refer to the real face image. For the training process, I may refer to a real sample.

[0039] A style image may be an image performing sample stylization by any object, such as a face, an animal, landscape, etc. Because a face style image is the most extensive, in order to facilitate understanding, the following embodiments of the disclosure will take face stylization as an example. At this time, the style image may be understood as an artistic portrait. In the process described hereinafter, the capital letter S is used to represent the style image. In different contexts, S may refer to the artistic portrait. For the training process, S may refer to a style sample.

[0040] Image stylization based on samples aims to render a face picture and achieve that the style of the rendered picture is consistent with that of a specified artistic portrait sample. It is widely applied in daily production and life, such as avatar design and animation poster advertising production. At present, the mainstream technology in the industry adopts an image conversion method to train a network to learn the mapping relationship between the real face image I and the artistic portrait S. This kind of method relies on a large amount of training data and is only applicable to a small-scale image, which is difficult to meet the actual needs.

[0041] At present, there is also a technology that adopts the StyleGAN to perform transfer learning on the small-scale artistic portrait data set to achieve high-resolution face stylization. However, the technology can only learn one overall style of the data set and is difficult to accurately simulate the style of an artistic portrait sample specified in the data set. How to implement a high-resolution face stylization technology based on a sample on small-scale data is a difficult problem.

[0042] Before the embodiments of the disclosure are introduced, the StyleGAN model is introduced at first. As shown in FIG. 1A, the model 10 includes a mapping network 11 and a generative network 12.

[0043] The mapping network 11 may map an input style code of a content image or a randomly sampled style code z into an intrinsic intermediate style code (which may be understood as an intermediate vector), and the intrinsic intermediate style code will be subsequently transmitted to the generative network 12 as 18 control vectors of the generative network 12, so that different elements of the control vectors may control different visual features. The style code of the content image may be obtained after the content image I is coded by a coder 111.

[0044] The mapping network 11 includes eight Full Connection (FC) layers, and the size of the intrinsic intermediate style code $\hat{W}$ output by the mapping network 11 is the same as that of input of the StyleGAN model. The general input size may be (512×1) or (512×18), namely a 512-dimension vector or eighteen 512-dimension vectors. If the input is a 512-dimension vector, an input layer will be copied into 18 same vectors.

[0045] The generative network 12 includes nine generation modules for transformation from 4*4 to 1024*1024, that is, the modules correspond to transformation of 4*4, 8*8, 16*16, 32*32, 64*64, 128*128, 256*256, 512*512 and 1024*1024, respectively. Each module, except the first generation module and the last generation module, is affected by two control vectors (input of affine transformation A in the figure), the module is affected once by one control vector after upsampling, the module is

affected once by the other control vector after convolution, and the affecting mode adopts a style adjustment module (the AdaIN layer). The first module does not have upsampling operation and only has convolution operation, so that the first module only uses one control vector. Besides the control vectors corresponding to the upsampling operation and the convolution operation, the last module further includes a control vector corresponding to conversion of features to a 3-channel image. Therefore, the intrinsic intermediate style code $\hat{W}$ is totally transformed into 18 control vectors transmitted to the generative network 12.

[0046] In other words, the first generation module includes an affine transformation module, a convolution layer and a style adjustment module. The first generation module does not have an upsampling module. Other generation modules include the upsampling module, the affine transformation module, the convolution layer and the style adjustment module.

[0047] The implementation process of the first generation module for sequential style adjustment is shown in FIG. 1A: after the intrinsic intermediate style code $\hat{W}$ passes through a learnable affine transformation module A (which may be an FC layer), it is expanded into a zoom factor y_(s,i) and a deviation factor y_(b,i) through the style adjustment module, and the zoom factor is multiplied by features output by the standardized convolution layer, and then summed with the deviation factor to complete a process of affecting the style image by the intrinsic intermediate style code $\hat{W}$ once.

[0048] The lower the resolution of a StyleGAN middle layer is, the coarser the affected features is. These features may be divided into two types: 1) coarse features: the resolution of which is not higher than 32*32, and the number of corresponding resolution layers is 1 to 7, postures, hairstyles, facial shapes and the like are affected, facial features, hairstyles, opening or closing of eyes, and the like are affected; 2) fine features: the resolution of which is 64*64 to 1024*1024, and the number of corresponding resolution layers is 8 to18, color (eyes, hair and skin) and microscopic features are affected.

[0049] The StyleGAN can show good performance in the aspect of generating artistic portraits through transfer learning of limited data. However, the method may only learn one overall style of the data set and is difficult to accurately simulate the style of an artistic portrait sample specified in the data set. In order to facilitate understanding, compared with the StyleGAN, it may be understood as a one-way style generative network.

[0050] To solve the above problems, an embodiment of the disclosure provides an image generation model at first. Compared with the StyleGAN, the image generation model may be understood as a dual style generative network (DualStyleGAN), that is, an extrinsic style control module is added on the basis of the StyleGAN, the DualStyleGAN is constructed, and the extrinsic style is learned with the small-scale data through the progressive transfer learning method to realize the face stylization based on the specified artistic portrait sample. The method can flexibly control the dual style of an original face domain and an extended artistic portrait domain.

[0051] The disclosure provides a training method of an image generation model. The image generation model includes a trained first generator and an extrinsic style control module to be trained. The first generator is pre-trained with a real sample set and may generate a real image based on an intrinsic style code. Referring to FIG. 1B, the method

at least includes S100, S200 and S300 below.

[0052] At S100, a style sample set and the real sample set are obtained.

[0053] The real sample set may be a real image data set, which is obtained by shooting of a real face data set, a real animal data set, a real landscape data set, etc. The first generator is pre-trained by using the real sample set. For example, the StyleGAN may be pre-trained by using the real face data set.

[0054] The style sample set may be an artistic image data set, and artistic images in the artistic image data set may be artistic portraits, artistic landscapes, cartoon images, comic images, etc. For example, images in the artistic image data set may be different forms of images obtained through modification modes such as rendering and/or deformation.

[0055] At S200, the extrinsic style control module is trained based on the real sample set to enable a first fused image generated by the image generation model to reconstruct a second fused image generated by the first generator.

[0056] Here, the first generator is pre-trained with the real sample set. In S200, the extrinsic style control module is trained with the real sample data set, so that style transfer in the real sample set is realized. Taking the StyleGAN as the first generator as an example, the real face sample set is used to train the extrinsic style control module to realize the style transfer in face data.

[0057] The extrinsic style control module is trained based on the real sample set. In the training process, two associated inputs (two image codes z_1 and z_2) are used for the first generator and the image generation model, so that the first fused image generated by the image generation model may be particularly similar to the second fused image generated by the first generator. It is sufficient to reconstruct the second fused image by using the first fused image, that is, it can be considered that the training of the extrinsic style control module with the real sample set is completed. The two associated inputs may be that one is the same and the other is associated. The two image codes z_1 and z_2 are obtained by random sampling in the real image code set. When the number of image codes in the real image code set is large enough, the two image codes z_1 and z_2 may also be understood as random noise.

[0058] At S300, after training of the extrinsic style control module is completed, the image generation model is trained based on the style sample set and a randomly sampled intrinsic style code till a first convergence condition is met.

[0059] In some embodiments, S300 includes S310 to S330 below.

[0060] At S310, iterative training is performed on the image generation model based on the style sample set and the randomly sampled intrinsic style code to obtain each style image generated by the image generation model.

[0061] At S320, the adversarial loss, the second reconstruction loss, the extrinsic style loss and the intrinsic style loss are respectively determined and accumulated based on each generated style image, the style sample corresponding to each style image and the intrinsic style code to obtain the accumulative sum.

[0062] The adversarial loss is adversarial loss L_{adv} of the first generator, the second reconstruction loss L_{dist} is loss of the generated style image corresponding to the

corresponding style sample, the extrinsic style loss L_{sty} is loss of the generated style image corresponding to the corresponding style sample, and the intrinsic style loss L_{con} is loss of the generated style image corresponding to the corresponding intrinsic style code.

[0063] Here, in S310, the loss function (the first loss function) of the image generation model refers to a formula (1), which includes the adversarial loss L_{adv}, the second reconstruction loss L_{dist}, the extrinsic style loss L_{sty} and the intrinsic style loss L_{con} of the StyleGAN:

$$L_{adv} + L_{dist}(G(z_i, z_e, 1), S) + L_{sty}(G(z, z_e, 1), S) + L_{con}(G(z, z_e, 1), g(z)) \quad (1);$$

where z is a randomly sampled intrinsic style code, style loss L_{sty} measures the consistency of a style conversion result and a style image S in terms of style, a common loss function such as contextual loss or feature matching loss in the style transfer technology may be adopted. Content loss L_{con} measures the consistency of the style conversion result and a content image g(z) in terms of content, and a common content loss function such as identity loss or perceptual loss may be used. In addition, the content loss in the method may be realized by L2 regularization of parameters of the residual module.

[0064] At S330, training on the image generation model is completed in the case where the accumulated sum meets the first convergence condition.

[0065] After training in S200 is completed, the extrinsic style control module realizes style transfer in face data. In S300, the image generation module, especially the extrinsic style control module, is trained by using the style sample set to realize the style transfer of the extrinsic style control module between the face data and the style data.

[0066] In the related art, the first generator, such as a StyleGAN, can only render the overall style of a single field, which is difficult to render the specified sample style. In the embodiments of the disclosure, the extrinsic style control module needs to be trained with the style sample set, and the first generator is pre-trained with the real sample set. Because the extrinsic style control module added to the image generation model will change the generation space of the first generator, it is difficult to apply a fine-tuning strategy of transfer learning, and the embodiments of the disclosure adopt a progressive transfer learning method to train the image generation model. The trained image generation model introduces the control on the extrinsic style by constructing the extrinsic style control module, so that the image generation model can control the intrinsic style and the extrinsic style at the same time, which is more suitable for processing a stylization task based on samples.

[0067] In S200, the extrinsic style control module is trained based on the real sample set so that the first fused image generated by the image generation model may reconstruct the second fused image of the first generator. As shown in FIG. 2A, S200 includes S210, S230, S250 and S270.

[0068] At S210, each second fused image is generated based on the real sample set by using the first generator.

[0069] In the implementation process, as shown in FIG. 2B, S210 includes S211 to S213. At S211, an image code of each real sample in the real sample set is respectively determined to obtain a real image code set. In the implementation process, the image code of the real sample may be an image feature vector and an image feature matrix, and

the image code may be at least one M-dimension vector, such as a 512-dimension or 1024-dimension vector. At S212, each pair of real image codes is obtained from the real image code set. During implementation, various real image codes in the real image code set may be combined to obtain at least one pair of real image codes, in which each pair of real image codes may be recorded as a first real image code z_1 and a second real image code z_2. At S213, each corresponding second fused image is generated based on each pair of real image codes by using the first generator. During implementation, $\hat{W}$ is generated based on the first real image code z_1 and the second real image code z_2. The second fused image, recorded as $g(\hat{W}')$, is generated by using the first generator.

[0070] At S230, each first fused image is generated based on the real sample set by using the image generation model.

[0071] The first fused image and the second fused image are coded based on the same pair of real image codes in the real sample set, namely the first real image code z_1 and the second real image code z_2. In the case where S210 includes S211 to S213, S230 includes that: each corresponding first fused image recorded as $G(z_1, E(g(z_2)), 1)$ is generated based on each pair of real image codes by using the image generation model.

[0072] At S250, first reconstruction loss between each first fused image and the corresponding second fused image is determined.

[0073] The first reconstruction loss may be recorded as $L_{dist}(G(z_1, E(g(z_2)), 1), g(\hat{W}'))$.

[0074] In terms of step execution, there is no strict sequence between S210 and S230. During implementation, S210 and S230 may be executed before S250.

[0075] At S270, iterative training is performed on the extrinsic style control module based on the first reconstruction loss till a second convergence condition is met.

[0076] Description is performed by taking the StyleGAN as an example. A loss function (a second loss function) for performing iterative training on the extrinsic style control module refers to a formula (2). The second loss function includes adversarial loss L_{adv} and the first reconstruction loss L_{dist} of the StyleGAN:

$$L_{adv} + L_{dist}(G(z_1, E(g(z_2)), 1), g(\hat{W}')) \qquad (2);$$

where an extrinsic style code $E(g(z_2))$ is obtained by inputting z_2 into the StyleGAN to obtain $g(z_2)$ and then coding $g(z_2)$ by using a coder E.

[0077] In some embodiments, the first generator includes a first mapping network and a generative network. In S213, each corresponding second fused image is generated based on each pair of real image codes by using the first generator. As shown in FIG. 2C, S213 includes S2131a to S2133a.

[0078] At S2131a, an intermediate style code of each real image code in each pair of real image codes is respectively obtained based on the first mapping network.

[0079] Each intermediate style code includes N element feature sequences. The N element feature sequences may be N vectors of M dimensions. N is an integer greater than or equal to 2. The value of M is the i-th power of 2, and i is a natural number greater than or equal to 1, such as 18 512-dimension vectors (when the first generator is a StyleGAN) or 18 1024-dimension vectors.

[0080] At S2132a, for each pair of real image codes, first k element features and last (N-k) element features are respectively extracted from two intermediate style codes for fusion to obtain a fused intermediate style code.

[0081] k is an integer greater than or equal to 1 and less than N. It is assumed that each pair of real image codes includes the first real image code z_1 and the second real image code z_2, the first k element features are extracted for z_1, and the last (N-k) element features are extracted for z_2.

[0082] At S2133a, a corresponding second fused image is generated based on each fused intermediate style code by using the generative network.

[0083] Description is performed by taking the StyleGAN as an example. Firstly, a data set is generated by using a style exchange technology of the StyleGAN, a pair of image codes is randomly selected, for example, two image codes z_1 and z_2 are sampled, 18 intermediate style codes $\hat{W}_1$ and $\hat{W}_2$ are respectively obtained after passing through the first mapping network, first k intermediate style codes of $\hat{W}_1$ and last (18-k) intermediate style codes of $\hat{W}_2$ are extracted to form a style code $\hat{W}'$, a style fused image $g(\hat{W}')$ is obtained through the generative network, and $g(\hat{W}')$ is the second fused image.

[0084] In some other embodiments, S213 in which each corresponding second fused image is generated based on each pair of real image codes by using the first generator includes S1231b to S2133b.

[0085] At S2131b, for each pair of real image codes, it repeats k times and (N-k) times respectively. For example, for z_1 and z_2 in each pair of real image codes, it repeats k times and (18-k) times respectively.

[0086] At S2132b, each pair of real image codes repeated k times and (N-k) times are combined to obtain a fused real image code containing N element features. For example, z_1 repeated k times and z_2 repeated (18-k) times are connected in series to form a fused style code ž containing 18 style codes.

[0087] At S2133b, a corresponding second fused image is generated based on each fused real image code by using the first generator.

[0088] Description is performed by taking a StyleGAN as an example. A data set is generated by using a style exchange technology of the StyleGAN at first, and a DualStyleGAN is trained in face data to realize style transfer. Two style codes z_1 and z_2 are randomly sampled, and are repeated k times and (18-k) times respectively to be connected in series to form a fused style code ž containing 18 style codes, a style fused image g(ž) is obtained through the StyleGAN, and g(ž) is the second fused image.

[0089] In the case where S213 includes S2131a to S2133a, or where S213 includes S2131b to S2133b, as shown in FIG. 2D, S230 in which each corresponding first fused image is generated based on each pair of real image codes by using the image generation model includes S231 and S233.

[0090] At S231, a corresponding real image is generated based on a second real image code in each pair of real image codes by using the first generator.

[0091] Each pair of real image codes includes a first real image code and a second real image code. The first real image code is a real image code corresponding to that the first k element features are extracted. The second real image code is a real image code

corresponding to that the last (N-k) element features are extracted.

[0092] At S232, for each pair of real image codes, each generated real image is coded by using a coder to obtain a corresponding extrinsic style code.

[0093] At S233, a corresponding first fused image is generated based on a first real image code in each pair of real image codes and a corresponding extrinsic style code by using an image generation model to be trained.

[0094] In some embodiments, the method further includes that: in the case where the first generator is a StyleGAN, the value of k is set to be 7, 6 and 5 in turn; and for each value of k, the intermediate style code is obtained based on the first mapping network to obtain a fused intermediate style code, and a second fused image is obtained based on the generative network. With the training of a DualStyleGAN, k is gradually reduced from 7 to 5, so that the extrinsic style control module may gradually learn structure style features. During implementation, k in each batch is same. The value of k is modified uniformly after training for a period of time.

[0095] In some embodiments, as shown in FIG. 3, S310 in which iterative training on the image generation model is performed based on the style sample set and the randomly sampled intrinsic style code to obtain each style image generated by the image generation model includes S311 to S313.

[0096] At S311, an extrinsic style code and an intrinsic style code of each style sample are determined.

[0097] In some embodiments, S311 includes S3111 and S3112. At S3111, each style sample is coded by using a coder to obtain an extrinsic style code of a corresponding style sample. At S3112, de-stylization is performed on each style sample respectively to obtain an intrinsic style code of each style sample.

[0098] During implementation, S3112 includes steps a to c: at the step a, an optimized style code of each style sample is determined, and the optimized style code may reconstruct a corresponding style sample through a trained second generator; at the step b, each optimized style code is reconstructed respectively by adopting a trained first generator to obtain a reconstructed style sample, and the first generator and the second generator are obtained by training on a real image data set and a style image data set respectively; and at the step c, each reconstructed style sample is coded by using the coder to obtain an intrinsic style code of a corresponding style sample.

[0099] In some embodiments, the method further includes that: a first generator g pre-trained with a real sample set is obtained; the first generator is trained with a style sample set to obtain a second generator g'; and in the implementation process, the first generator is finely tuned to obtain a second generator g'. In the embodiment of the disclosure, g may be the whole StyleGAN (including a mapping network and a generative network), instead of just referring to the generative network.

[00100] For example, assuming that a style sample is represented by S, an extrinsic style code of the style sample S may be represented by z_e, and an intrinsic style code of the style sample S may be represented by z_i. For each style sample S, the style sample is coded by using a coder E, so that an extrinsic style code $z_e=E(S)$ may be obtained. Then, the style code z_e is updated by using a method based on optimization and referring to a formula (3), so that S may be reconstructed through g':

$$z' = \arg\min_{z} L_{dist}(g'(z), S) \qquad (3);$$

where $L_{dist}(x,y)$ measures the similarity between two images x and y, and a common loss function, such as L_1 distance and perceptual loss, may be adopted. Finally, a reconstructed image g(z') is coded again by using the coder E to obtain a style code zi=E(g(z')) after de-stylization of the S. In this way, z_i and z_e constitute the intrinsic style code and the extrinsic style code of an artistic image S.

[00101] At S312, a first target style image generated by the image generation model is obtained by taking the intrinsic style code and the extrinsic style code of each style sample as input of the first generator and input of the extrinsic style control module respectively.

[00102] At S313, a second target style image generated by the image generation model is obtained by taking the randomly sampled intrinsic style code and the extrinsic style code of each style sample as input of the first generator and input of the extrinsic style control module respectively.

[00103] In some embodiments, S320 in which second reconstruction loss, extrinsic style loss and intrinsic style loss are determined respectively based on each generated style image, the style sample corresponding to each style image and the intrinsic style code includes S321 to S323.

[00104] At S321, the second reconstruction loss is determined based on each first target style image and a corresponding style sample. At S322, the extrinsic style loss is determined based on each second target style image and a corresponding style sample. At S323, the intrinsic style loss is determined based on each second target style image and a corresponding random real image, where the random real image is generated by the first generator based on the randomly sampled intrinsic style code.

[00105] An embodiment of the disclosure provides an image generation model. Compared with the StyleGAN, the image generation model may be understood as a DualStyleGAN, that is, an extrinsic style control module is added on the basis of the StyleGAN to construct the DualStyleGAN, and the extrinsic style is learned on small-scale data through a progressive transfer learning method to realize the face stylization based on specified artistic portrait samples. The method can flexibly control the dual style of an original face domain and an extended artistic portrait domain.

[00106] A composition structure of the image generation model refers to FIG. 4. The image generation module includes an intrinsic style control module 110, a generative network 120 and an extrinsic style control module 130.

[00107] The intrinsic style control module 110 includes a first input coder 111 and a first mapping network 112. The first input coder 111 is configured to code a real face image I to obtain an intrinsic style code. The first mapping network 112 is configured to map the intrinsic style code to obtain an intrinsic intermediate style code.

[00108] The generative network 120 is configured to generate a high-resolution image based on the intrinsic intermediate style code.

[00109] The extrinsic style control module 130 includes a second input coder 131 and a second mapping network 132. The second input coder 131 is configured to code a style image S to obtain an extrinsic style code. The second mapping network 132 is configured

to map the extrinsic style code to obtain an extrinsic intermediate style code.

[00110] In an achievable mode, the intrinsic style control module 110 and the generative network 120 form a StyleGAN, which remains fixed during transfer learning.

[00111] Also referring to FIG. 4, the extrinsic style control module includes the second mapping network 132 having the same structure as the first mapping network 112, an extrinsic structure style control module (T_s) 133, a residual module (ModRes) 134 and an extrinsic color style control module (T_c) 135. In the final training process of the image generation model, these modules have the following functions.

[00112] The first mapping network 112 is configured to correspondingly obtain the intrinsic intermediate style code of each style sample respectively based on the intrinsic style code z of each style sample, and correspondingly obtain the randomly sampled intrinsic intermediate style code based on the randomly sampled intrinsic style code z.

[00113] The second mapping network 132 is configured to respectively obtain the extrinsic intermediate style code of each extrinsic style code, and each extrinsic intermediate style code includes N element feature sequences.

[00114] The extrinsic structure style control module 133 is configured to determine a structure style code of each style sample based on a part of the extrinsic intermediate style code of each style sample.

[00115] The extrinsic color style control module 135 is configured to extract a color style code of each style sample based on the other part of the extrinsic intermediate style code of each style sample.

[00116] The residual module 134 is configured to perform structure style adjustment on a target image generated by the generative network based on the structure style code of each style sample.

[00117] The generative network 120 is configured to generate a first target style image based on the structure style code and the color style code of each style sample and the intrinsic intermediate style code of each style sample, and generate a second target style image based on the structure style code and the color style code of each style sample and the randomly sampled intrinsic intermediate style code.

[00118] In FIG. 4, the extrinsic structure style control module (T_s) 133, the residual module 134 and the extrinsic color style control module (T_c) 135 may be training parts of the image generation model. The convolution layer and the affine transformation module in each generation module, the FC layer in the first mapping network 112 and the second mapping network 132, the first input coder 111 and the second input coder 131 may be fixed part of the image generation model, and the style adjustment module and the upsampling module in each generation module may be arithmetic operation modules of the image generation model.

[00119] In the embodiments of the disclosure, description is performed by taking a StyleGAN as a first generator as an example. Aiming at the problem that an existing face stylization method is difficult to realize high-resolution face stylization based on a sample with small-scale data, the embodiments of the disclosure provide a portrait stylization method based on dual styles, an extrinsic style control module is added on the basis of the StyleGAN to construct a brand new DualStyleGAN, and the extrinsic style is

learned with the small-scale data through a progressive transfer learning method to realize the high-resolution face stylization based on examples. Because the addition of the extrinsic style control module will change the pre-training generation space of the StyleGAN, it is difficult to apply a fine-tuning strategy of transfer learning, and therefore, it is necessary to adopt a progressive transfer learning method to train the network.

[00120] According to the image processing method provided by the embodiments of the disclosure, on one hand, a DualStyleGAN network model, which includes an intrinsic style channel and an extrinsic style channel, is provided. The content of a portrait is represented through the intrinsic style channel. The artistic style of the portrait is represented through the extrinsic style channel. The extrinsic style channel of DualStyleGAN enables the network model to adjust the content and style of an artistic portrait hierarchically, so as to accurately imitate the style of the sample. On the other hand, a progressive transfer learning method is provided. The generation space of the network model can be smoothly converted to the target domain through the transfer learning method.

[00121] The image processing method provided in an embodiment of the disclosure includes two stages.

[00122] In the first stage, de-stylization is performed on a style data set to obtain face-style group training data.

[00123] In the second stage, the DualStyleGAN is constructed and trained by using the progressive transfer learning method through three steps of initialization, style transfer in face data, and style transfer between a face and style data sequentially, so as to learn style features of a small-scale style data set and realize natural and accurate face stylization.

[00124] The first stage will be introduced below. De-stylization is performed on the style data set. The first stage includes S1.1 and S1.2.

[00125] At S1.1, transfer learning of the StyleGAN and training of a StyleGAN coder are performed.

[00126] Firstly, the StyleGAN (which may be understood as a first generator) g pre-trained with a real face data set is obtained. The StyleGAN g is finely tuned with an artistic portrait data set to obtain a fine tuned model g’ (which may be understood as a second generator). That is, the StyleGAN g pre-trained with the real face data set is finely tuned with the artistic portrait data set to obtain a fine tuned model g’. Then, one StyleGAN coder E is trained to map a face image into a style code of the StyleGAN. The style code may reconstruct the input face image through the StyleGAN g.

[00127] At S1.2, de-stylization of an artistic portrait is performed. Face de-stylization aims to restore a realistic face from the artistic portrait to form face-style group data, and perform supervision by using the data.

[00128] Firstly, for each artistic portrait S, an extrinsic style code of the S is obtained as $z_e=E(S)$.

[00129] Then, the style code z is updated by using a method based on optimization and referring to a formula (3), so that the S may be reconstructed through g’:

$$z' = \arg\min_{z} L_{dist}(g'(z), S) \qquad (3);$$

where an initial value of z is z_e, an optimized loss function updates the variable z from z_e to z', $L_{dist}(x,y)$ measures the similarity between two images x and y, and common loss functions, such as L_1 distance and perceptual loss, may be adopted.

[00130] z'=arg min_z(f(z)) is a mathematical description way, which represents the optimized variable z (the randomly sampled style code z may also be understood as a variable). The final optimization result is z', and f(z') is the z that minimizes f(z). The method based on optimization is to update the input z of the network through the loss function, instead of updating the network itself. The purpose of this step is to obtain the style code z' which is capable of reconstructing S. The reconstructed S may be defined as S', and S'= g'(z').

[00131] Finally, the reconstructed g(z') is coded by using the coder E again to obtain a style code z_i=E(g(z')) after de-stylization of the S. Here, z_i and z_e constitute the intrinsic style code and the extrinsic style code of the artistic image S.

[00132] In the second part, a DualStyleGAN is constructed and trained. The second part includes S2.1 and S2.2.

[00133] At S2.1, the DualStyleGAN is constructed, and for the 18-layer StyleGAN, as shown in FIG. 4, the extrinsic style control module 130 on the right side is constructed to form a DualStyleGAN G.

[00134] Firstly, an artistic portrait S passes through a coder E to obtain an extrinsic style code z_e, and then passes through a mapping network to obtain 18 extrinsic intermediate style codes. The mapping network is a multi-layer perceptron, which inputs 18 512-dimension vectors, namely 18 style codes with the length of 512.

[00135] At S2.1.1, an extrinsic structure style is controlled.

[00136] First seven extrinsic intermediate style codes in the 18 extrinsic intermediate style codes are coded into extrinsic structure style codes through a linear layer T_s 133. The extrinsic structure style codes adjust the structure of a face to match the artistic portrait, such as big eyes and big mouth. Residual features are obtained from features of first seven convolution layers of the StyleGAN by adjusting the residual module 134, and the residual features are multiplied by the weight w_i ($i < 8$) and then added into the features to complete adjustment of a face structure.

[00137] The residual module includes two convolution layers and two AdaIN layers (the style adjustment module in the figure). Adjustment parameters of the structure style are obtained from the extrinsic structure style codes through the linear layer in the AdaIN and are used for structure style control of the features. The input of the residual module consists of two parts: the features output by the convolution layers, and the extrinsic structure style codes. The first seven extrinsic intermediate style codes are in one-to-one correspondence to the first seven convolution layers. The first 4*4 generation module includes one convolution layer, the last 1024*1024 generation module includes three convolution layers, and the seven generation modules in the middle each include two convolution layers.

[00138] At S2.1.2, an extrinsic color style is controlled.

[00139] Firstly, the last eleven extrinsic intermediate style codes in the 18 extrinsic intermediate style codes are coded into the extrinsic color style codes through the linear

layer T_c 135, then subject to weighted summation with the intrinsic intermediate style codes obtained by the intrinsic style control module 110 of the StyleGAN by means of w_i ($i > 7$), and input into the style adjustment module of the StyleGAN, so as to realize the color style control of the features.

[00140] Therefore, the DualStyleGAN generates an artistic portrait according to the intrinsic style code z_i describing facial features and the extrinsic style code z_e describing the artistic style, and controls the styles of different scales by adjusting the weight w_i of different layers, which is formally expressed as $G(z_i, z_e, w)$. $G(z_i, z_e, w)$ represents the image output by the DualStyleGAN. For example, setting the weight $w_i = 0$ ($i > 7$) may retain the intrinsic color style, and setting $w_i < 1$ may reduce the degree of the extrinsic style and retain the appearance features of the face. The lowercase w represents the weight, such as w_i, while the uppercase W with an caret above $\hat{W}$ represents the intermediate style code (which may be understood as an intermediate vector).

[00141] At S2.2, the DualStyleGAN is trained: addition of the extrinsic style control module will change the pre-training generation space of the StyleGAN, it is difficult to apply a fine-tuning strategy of transfer learning. Therefore, a progressive transfer learning method is adopted. The progressive transfer learning method includes the following three steps.

[00142] At S2.2.1, an extrinsic style control module is initialized.

[00143] The convolution kernel of the convolution layer in the initialization residual module is a random value close to 0, and a parameter matrix of the initialization linear layer T_c is a unit matrix. The above initialization makes the residual feature value close to 0, and the extrinsic intermediate style code remains unchanged after T_c, so as to keep the structure almost unchanged, so that the StyleGAN added with the extrinsic style control module still maintains the original generation space.

[00144] At S2.2.2, style transfer in face data is that: a data set is generated by using a style exchange technology of the StyleGAN, and the DualStyleGAN is trained with the face data to realize style transfer.

[00145] Firstly, the data set is generated by using the style exchange technology of the StyleGAN, two style codes z_1 and z_2 are randomly sampled, 18 intermediate style codes $\hat{W}_1$ and $\hat{W}_2$ are respectively obtained after passing through the mapping network, first k intermediate style codes of $\hat{W}_1$ and last (18-k) intermediate style codes of $\hat{W}_2$ are extracted to form a style code $\hat{W}'$, and a style fused image $g(\hat{W}')$ is obtained through the generative network.

[00146] Then, the DualStyleGAN is trained to reconstruct $g(\hat{W}')$ according to the intrinsic style code z_1 and the extrinsic style code $E(g(z_2))$. As the training goes on, k is gradually reduced from 7 to 5, so that the extrinsic style control module gradually learns structure style features.

[00147] A loss function here refers to a formula, including adversarial loss L_{adv} and reconstruction loss L_{dist} of the StyleGAN:

$$L_{adv} + L_{dist}(G(z_1, E(g(z_2)), 1), g(\hat{W}')) \quad (2)$$

where only the extrinsic style control module of the DualStyleGAN is trained, other modules remain unchanged, and the weight w is all set to 1. An extrinsic style code

$E(g(z_2))$ is obtained by inputting z_2 into the StyleGAN to obtain $g(z_2)$ and then coding $g(z_2)$ by using the coder E.

[00148] In some other embodiments, S2.2.2 may also be implemented in the following way: the data set is generated by using the style exchange technology of the StyleGAN first, and the DualStyleGAN is trained in the face data to realize style transfer. Two style codes z_1 and z_2 are randomly sampled, and respectively repeated k times and (18-k) times to be connected in series to form a fused style code ž containing 18 style codes, and a style fused image g(ž) is obtained through the StyleGAN. Then, the DualStyleGAN is trained to reconstruct g(ž) according to the intrinsic style code z_1 and the extrinsic style code $E(g(z_2))$. As the training goes on, k is gradually reduced from 7 to 5, so that the extrinsic style control module gradually learns structure style features.

[00149] A loss function here refers to a formula (4), including adversarial loss L_{adv} and third reconstruction loss L_{dist} of the StyleGAN:

$$L_{adv} + L_{dist}(G(z_1, E(g(z_2)), 1), g(\check{z})) \quad (4)$$

where only the extrinsic style control module of the DualStyleGAN is trained, other modules remain unchanged, and the weight w is set to all 1.

[00150] At S2.2.3, style transfer between a face and style data is as follows.

[00151] The DualStyleGAN is trained by taking paired training data (z_i, z_e, S) obtained in S1.2 as supervision to realize face stylization. Referring to a formula (1), the loss function includes adversarial loss L_{adv}, reconstruction loss L_{dist}, style loss L_{sty} and content loss L_{con} of the StyleGAN:

$$L_{adv} + L_{dist}(G(z_i, z_e, 1), S) + L_{sty}(G(z, z_e, 1), S) + L_{con}(G(z, z_e, 1), g(z)) \quad (1);$$

where z is a randomly sampled intrinsic style code, the style loss L_{sty} measures the consistency of a style conversion result and a style image S in style, a common loss function, such as contextual loss or feature matching loss, in the style transfer technology may be adopted. The content loss L_{con} measures the consistency of the style conversion result and a content image g(z) in content, and a common content loss functions, such as identity loss or perceptual loss, may be adopted. In addition, the content loss in the method may be realized by L_2 regularization of parameters of the residual module.

[00152] In the image generation model as shown in FIG. 4, z represents random noise, namely a randomly sampled image code. As mentioned earlier, during implementation, the image code may be 18 512-dimension vectors. Because the sampling is ramdom, the image code may be regarded as random noise z mathematically. In the image generation model, the intrinsic style code on the left side may be obtained by coding the content image through the first input coder, and may also be random noise. Similarly, the extrinsic style code on the right side may be obtained by coding the style image S through the second input coder, and may also be obtained after the random noise z passes through a sampling network N_c.

[00153] In some other embodiments, as shown in FIG. 4, the image generation module may also include the sampling network N_c, and in that way, the training part of the image generation module also includes the sampling network N_c. During implementation, the training method of the image generation model also includes that: the sampling network N_c is trained, distribution of extrinsic style codes is learned and modeled, and the random

noise z is mapped into the extrinsic style code. Therefore, the trained sampling network N_c is used for mapping the random noise z into the extrinsic style code. At this time, the extrinsic style code output by the sampling network N_c may be recorded as $N_c(z)$. The input of the DualStyleGAN is z_1 and z_2. $N_c(z_2)$ is obtained after z_2 passes through the sampling network N_c, and in that way, a random artistic portrait generated by the DualStyleGAN may be recorded as $G(z_1, N_c(z_2), w)$.

[00154] In the embodiment of the disclosure, in an aspect, a network framework may perform transfer learning with a small data set to generate a high-resolution image. Therefore, the problems that a large amount of training data is required while training the network due to the learning of the mapping relationship between the real face image I and the artistic portrait S and it is only suitable for a small-size image may be effectively reduced. In another aspect, the control on the extrinsic style is introduced by constructing the extrinsic style control module, so that the DualStyleGAN can control the intrinsic style and the extrinsic style at the same time. It is more suitable for processing a face stylization task based on a sample. Therefore, the problem that the StyleGAN can only render the overall style of a single field and is difficult to render a specified sample style is avoided. In a further aspect, through targeted network initialization, intra-domain training and inter-domain training, the progressive transfer learning method robustly guides the DualStyleGAN to learn image distribution of an artistic portrait, which may avoid the problem that the transfer learning technology finely tunes the StyleGAN with the target data set and it is not suitable for the DualStyleGAN that changes the network architecture.

[00155] An embodiment of the disclosure further provides an image processing method. As shown in FIG. 5, the method includes the following operations.

[00156] At S501, a style image and a content image to be fused are obtained.

[00157] When the method is implemented, it may be a computer executable program, which is embodied as a client on a user side. When a user selects two images as the style image and content image to be fused, for example, the user selects two images from an album, and the client receives the operation that the user selects the images, that is, the client determines the style image and content image to be fused from the selected two images. Or, the user selects one content image, but the user does not need to select a style image. For example, the user may select a style, such as an impressionist style, so that the device may obtain a van Gogh style image from the local or a server.

[00158] At S502, extraction is performed on the content image to obtain an intrinsic intermediate style code of the content image.

[00159] During implementation, the operation of performing extraction on the content image to obtain the extrinsic intermediate style code of the content image includes that: an extrinsic style code of the content image may be extracted by using a coder; and the extrinsic style code of the content image is extracted by using a mapping network to obtain an extrinsic intermediate style code of the content image.

[00160] At S503, extraction is performed on the style image to obtain an extrinsic intermediate style code of the style image.

[00161] During implementation, the operation of performing extraction on the style image to obtain the extrinsic intermediate style code of the style image includes that: an

extrinsic style code of the style image may be extracted by using the coder; and the extrinsic style code of the style image may be extracted by using the mapping network to obtain the extrinsic intermediate style code of the style image.

[00162] In FIG. 4, the input of the mapping network is the style code of the image. In the process of implementation, after the user inputs the content image I and the style image S, the coder codes the content image I and the style image S respectively to obtain the style code of the content image I and the style code of the style image S, and then the mapping network converts them into the intrinsic intermediate style code of the content image I and the extrinsic intermediate style code of the style image S respectively.

[00163] At S504, a fused image is generated based on the intrinsic intermediate style code of the content image and the extrinsic intermediate style code of the style image.

[00164] In some embodiments, the operation of generating the fused image based on the intrinsic intermediate style code of the content image and the extrinsic intermediate style code of the style image includes that: the structure style code of the style image is determined based on a part of the extrinsic intermediate style code of the style image; the color style code of the style image is extracted based on the other part of the extrinsic intermediate style code of the style image; and the fused image is generated based on the structure style code and the color style code of the style image and the intrinsic intermediate style code of the content image.

[00165] In some embodiments, each extrinsic intermediate style code includes N element feature sequences.

[00166] The operation of determining the structure style code of the style image based on a part of the extrinsic intermediate style code of the style image includes that: the structure style code of the style image is determined based on first k element features in the extrinsic intermediate style code of the style image.

[00167] The operation o f extracting the color style code of the style image based on the other part of the extrinsic intermediate style code of the style image includes that: the color style code of the style image is extracted based on last (N-k) element features of the extrinsic intermediate style code of the style image.

[00168] The method provided by the embodiment of the disclosure includes a plurality of user scenes: 1) an artistic portrait is customized for a user: through the technology, the mapping relationship between a portrait and an artistic portrait such as cartoon, caricature, satire is established, which supports the user to generate creative avatars of various styles based on their own photos, and supports artistic face rendering of online meetings. 2) Artistic portrait data is generated: the technology may be used to generate a random artistic portrait. The generated image data may be used for avatar design, animation poster advertising and other graphic products, as well as subsequent depth learning and analysis.

[00169] Based on the foregoing embodiments, an embodiment of the disclosure provides a training apparatus of an image generation model, which includes each module, each module includes each unit, and each unit includes each sub-unit, which may be realized by a processor in a computer device. Of course, it may also be realized through a specific logic circuit. In the process of implementation, the processor may be a Central Processing Unit (CPU), a Microprocessor (MPU), a Digital Signal Processor (DSP) or a

Field Programmable Gate Array (FPGA) and the like.

[00170] The image generation model includes a trained first generator and an extrinsic style control module to be trained. The first generator is pre-trained with a real sample set and is capable of generating a real image based on an intrinsic style code. As shown in FIG. 6A, the apparatus 610 includes a first obtaining module 611, a first training module 612 and a second training module 613.

[00171] The first obtaining module 611 is configured to obtain a style sample set and the real sample set.

[00172] The first training module 612 is configured to train the extrinsic style control module based on the real sample set so that a first fused image generated by the image generation model may reconstruct a second fused image generated by the first generator.

[00173] The second training module 613 is configured to, after training of the extrinsic style control module is completed, train the image generation model based on the style sample set and a randomly sampled intrinsic style code till a first convergence condition is met.

[00174] In some embodiments, the first training module includes: a first generation unit, which is configured to generate each second fused image based on the real sample set by using the first generator; a second generation unit, which is configured to generate each first fused image based on the real sample set by using the image generation model, where the first fused image and the second fused image are based on the same pair of real image codes in the real sample set; a first determination unit, which is configured to determine first reconstruction loss between each first fused image and a corresponding second fused image; and a first training unit, which is configured to perform iterative training on the extrinsic style control module based on the first reconstruction loss till a second convergence condition is met.

[00175] In some embodiments, the first generation unit includes: a first determination sub-unit, which is configured to respectively determine image codes of each real sample in the real sample set to obtain a real image code set; a first obtaining sub-unit, whici s configured to obtain each pair of real image codes from the real image code set; and a first generation sub-unit, which is configured to generate each corresponding second fused image based on each pair of real image codes by using the first generator. Correspondingly, the second generation unit is configured to generate each corresponding first fused image based on each pair of real image codes by using the image generation model.

[00176] In some embodiments, the first generator includes a first mapping network and a generative network. The first generation sub-unit is configured to: respectively obtain intermediate style codes of each real image code in each pair of real image codes based on the first mapping network, where each intermediate style code includes N element feature sequences, N is an integer greater than or equal to 2; for each pair of real image codes, respectively extract first k element features and last (N-k) element features from two intermediate style codes for fusion to obtain a fused intermediate style code, where k is an integer greater than or equal to 1 and less than N; and generate corresponding second fused image based on each fused intermediate style code by using the generative network.

[00177] In some embodiments, the apparatus further includes a setting module, which is configured to set the value of k to be 7, 6 and 5 in turn in the case where the first generator is a StyleGAN. Each value of k is realized by using the first generation sub-unit: the intermediate style code is obtained based on the first mapping network to obtain the fused intermediate style code, and the second fused image is obtained based on the generative network.

[00178] In some embodiments, the second generation unit includes: a second generation sub-unit, which is configured to generate a corresponding real image based on a second real image code in each pair of real image codes by using the first generator, where the second real image code is a real image code corresponding to that the last (N-k) element features are extracted; a coding sub-unit, which is configured to code each generated real image by using a coder for each pair of real image codes to obtain a corresponding extrinsic style code; and a third generation sub-unit, which is configured to generate a corresponding first fused image based on the first real image code in each pair of real image codes and a corresponding extrinsic style code by using the image generation model to be trained, wherein the first real image code is a real image code corresponding to that the first k element features are extracted.

[00179] In some embodiments, the second training module includes: a first training unit, which is configured to perform iterative training on the image generation module based on the style sample set and the randomly sampled intrinsic style code to obtain each style image generated by the image generation model; a second determination unit, which is configured to respectively determine and accumulate adversarial loss, second reconstruction loss, extrinsic style loss and intrinsic style loss based on each generated style image and the style sample and the intrinsic style code corresponding to each style image to obtain the accumulated sum, wherein the adversarial loss is adversarial loss of the first generator, the second reconstruction loss is loss of the generated style image corresponding to the corresponding style sample, the extrinsic style loss is loss of the generated style image corresponding to the corresponding style sample, and the intrinsic style loss is loss of the generated style image corresponding to the corresponding intrinsic style code; and a third determination unit, which is configured to determine that training of the image generation model is completed in the case where the accumulated sum meets the first convergence condition.

[00180] In some embodiments, the first training unit includes: a second determination sub-unit, which is configured to determine an extrinsic style code and an intrinsic style code of each style sample; and a second obtaining sub-unit, which is configured to: obtain a first target style image generated by the image generation model by respectively taking the intrinsic style code and the extrinsic style code of each style sample as the input of the first generator and the input of the extrinsic style control module, and obtain a second target style image generated by the image generation model by respectively taking the randomly sampled intrinsic style code and the extrinsic style code of each style sample as the input of the first generator and the input of the extrinsic style control module.

[00181] The second determination unit includes: a third determination sub-unit, which is configured to determine the second reconstruction loss based on each first target style image and a corresponding style sample; a fourth determination sub-unit, which is configured to determine the extrinsic style loss based on each second target style image

and a corresponding style sample; and a fifth determination sub-unit, which is configured to determine the intrinsic style loss based on each second target style image and a corresponding random real image. The random real image is generated by the first generator based on the randomly sampled intrinsic style code.

[00182] In some embodiments, the second determination sub-unit is configured to code each style sample by using the coder to obtain the extrinsic style code of a corresponding style sample, and respectively perform de-stylization on each style sample to obtain the intrinsic style code of each style sample.

[00183] In some embodiments, the extrinsic style control module includes a second mapping network having the same structure as the first mapping network, an extrinsic structure style control module, an extrinsic color style control module and a residual module. The obtaining sub-unit is configured to correspondingly obtain the intrinsic intermediate style code of each style sample respectively based on the intrinsic style code of each style sample, and correspondingly obtain the randomly sampled intrinsic intermediate style code based on the randomly sampled intrinsic style code by using the first mapping network, and respectively obtain the extrinsic intermediate style code of each extrinsic style code by using the second mapping network; and each extrinsic intermediate style code includes N element feature sequences. The extrinsic structure style control module is configured to determine the structure style code of each style sample based on a part of the extrinsic intermediate style code of each style sample. The extrinsic color style control module is configured to determine the color style code of each style sample based on the other part of the extrinsic intermediate style code of each style sample. The residual module is configured to perform structure style adjustment on the target image generated by the generative network based on the structure style code of each style sample. The generative network is configured to generate the first target style image based on the structure style code and the color style code of each style sample and the intrinsic intermediate style code of each style sample, and generate the second target style image based on the structure style code and the color style code of each style sample and the randomly sampled intermediate style code.

[00184] Based on the foregoing embodiments, an embodiment of the disclosure provides an image processing apparatus, which includes each module, each module includes each unit, which may be realized by a processor in a computer device. Of course, it may also be realized through a specific logic circuit. In the process of implementation, the processor may be a CPU, an MPU, a DSP or an FPGA and the like.

[00185] FIG. 6B illustrates a composition structure diagram of an image processing apparatus provided by an embodiment of the application. As shown in FIG. 6B, the apparatus 620 includes a first obtaining module 621, a first extraction module 622, a second extraction module 623 and a generation module 624.

[00186] The first obtaining module 621 is configured to obtain a style image and a content image to be fused.

[00187] The first extraction module 622 is configured to perform extraction on the content image to obtain an intrinsic intermediate style code of the content image.

[00188] The second extraction module 623 is configured to perform extraction on the style image to obtain an extrinsic intermediate style code of the style image.

[00189] The generation module 624 is configured to generate a fused image based on the intrinsic intermediate style code of the content image and the extrinsic intermediate style code of the style image.

[00190] In some embodiments, the generation module includes: a fourth determination unit, which is configured to determine the structure style code of the style image based on a part of the extrinsic intermediate style code of the style image; a fifth determination unit, which is configured to extract the color style code of the style image based on the other part of the extrinsic intermediate style code of the style image; and a third generation unit, which is configured to generate the fused image based on the structure style code and the color style code of the style image and the intrinsic intermediate style code of the content image.

[00191] In some embodiments, each extrinsic intermediate style code includes N element feature sequences. The fourth determination unit is configured to determine the structure style code of the style image based on the first k element features in the extrinsic intermediate style code of the style image. The fifth determination unit is configured to extract the color style code of the style image based on the last (N-k) element features of the extrinsic intermediate style code of the style image.

[00192] The above descriptions about the apparatus embodiments are similar to those about the method embodiments and beneficial effects similar to those of the method embodiments are achieved. Technical details undisclosed in the apparatus embodiments of the disclosure may be understood with reference to those about the method embodiments of the disclosure.

[00193] It is to be noted that, in the embodiments of the disclosure, the method may also be stored in a computer-readable storage medium when implemented in form of a software function module and sold or used as an independent product. Based on such an understanding, the technical solutions of the embodiments of the disclosure substantially or parts making contributions to the related art may be embodied in form of a software product. The computer software product is stored in a storage medium, including a plurality of instructions configured to enable a device to execute all or part of the method in each embodiment of the disclosure. The storage medium includes: various media capable of storing program codes such as a U disk, a mobile hard disk, a Read Only Memory (ROM), a magnetic disk or an optical disk. As a consequence, the embodiments of the disclosure are not limited to any target hardware and software combination.

[00194] FIG. 7 illustrates a schematic diagram of a hardware entity of a computer device provided in an embodiment of the disclosure. As shown in FIG. 7, the hardware entity of the device 700 for detecting game currencies on a game table includes a processor 701 and a memory 702, wherein the memory 702 stores a computer program capable of running on the processor 701, and the processor 701 implements the steps in the method in any embodiment when executing the program. In some implementation modes, the device 700 for receiving and losing the game currencies on the game table may be a detection device described in any embodiment.

[00195] The memory 702 stores the computer program capable of running in the processor. The memory 702 is configured to store an instruction and application executable for the processor 701, may also cache data (for example, image data) to be processed or having been processed by the processor 701 and each module in the device

700 for detecting the game currencies on the game table, and may be implemented by a flash or a Random Access Memory (RAM).

[00196] The processor 701 executes the program to implement the steps of any method above. The processor 701 usually controls the overall operation of the device 700 for detecting the game currencies on the game table.

[00197] An embodiment of the disclosure provides a computer storage medium storing one or more programs which may be executed by one or more processors to implement the steps of the method in any abovementioned embodiment.

[00198] It is to be pointed out here that the above descriptions about the storage medium and device embodiments are similar to those about the method embodiment, and beneficial effects similar to those of the method embodiment are achieved. Technical details undisclosed in the storage medium and device embodiments of the disclosure are understood with reference to those about the method embodiment of the disclosure.

[00199] The processor may include at least one of a target Application Specific Integrated Circuit (ASIC), a DSP, a Digital Signal Processing Device (DSPD), a Programmable Logic Device (PLD), an FPGA, a CPU, a controller, a microcontroller, and a microprocessor. It can be understood that other electronic devices may also be configured to realize functions of the processor, and no specific limits are made in the embodiments of the disclosure.

[00200] The computer storage medium/the memory may be a memory such as a ROM, a Programmable Read-Only Memory (PROM), an Erasable Programmable Read-Only Memory (EPROM), an Electrically Erasable Programmable Read-Only Memory (EEPROM), a Ferromagnetic Random Access Memory (FRAM), a flash memory, a magnetic surface memory, an optical disk, or a Compact Disc Read-Only Memory (CD-ROM), or may be any terminal including one or any combination of the abovementioned memories, such as a mobile phone, a computer, a tablet device, and a personal digital assistant.

[00201] It is to be understood that "one embodiment" or "an embodiment" or "the embodiment of the disclosure" or "the abovementioned embodiment" or "some implementation modes" or "some embodiments" mentioned in the whole specification means that target features, structures or characteristics related to the embodiment are included in at least one embodiment of the disclosure. Therefore, "in one embodiment" or "in an embodiment" or "the embodiment of the disclosure" or "the abovementioned embodiment" or "some embodiments" appearing everywhere in the whole specification does not always refer to the same embodiment. In addition, these target features, structures or characteristics may be combined in one or more embodiments freely as appropriate. It is to be understood that, in each embodiment of the disclosure, a magnitude of a sequence number of each process does not mean an execution sequence and the execution sequence of each process should be determined by its function and an intrinsic logic and should not form any limit to an implementation process of the embodiments of the disclosure. The sequence numbers of the embodiments of the disclosure are adopted not to represent superiority-inferiority of the embodiments but only for description.

[00202] If not specified, when the detection device executes any step in the embodiments of the disclosure, the processor of the device executes the step. Unless

otherwise specified, the sequence of execution of the following steps by the detection device is not limited in the embodiments of the disclosure. In addition, the same method or different methods may be used to process data in different embodiments. It is also to be noted that any step in the embodiments of the disclosure may be executed independently by the detection device, namely the detection device may execute any step in the abovementioned embodiments independent of execution of the other steps.

[00203] In some embodiments provided by the disclosure, it is to be understood that the disclosed device and method may be implemented in another manner. The device embodiment described above is only schematic, and for example, division of the units is only logic function division, and other division manners may be adopted during practical implementation. For example, multiple units or components may be combined or integrated into another system, or some characteristics may be neglected or not executed. In addition, coupling or direct coupling or communication connection between each displayed or discussed component may be indirect coupling or communication connection, implemented through some interfaces, of the device or the units, and may be electrical and mechanical or adopt other forms.

[00204] The units described as separate parts may or may not be physically separated, and parts displayed as units may or may not be physical units, and namely may be located in the same place, or may also be distributed to multiple network units. Part of all of the units may be selected according to a practical requirement to achieve the purposes of the solutions of the embodiments. In addition, each function unit in each embodiment of the disclosure may be integrated into a processing unit, each unit may also serve as an independent unit and two or more than two units may also be integrated into a unit. The integrated unit may be implemented in a hardware form and may also be implemented in form of hardware and software function unit.

[00205] The methods disclosed in some method embodiments provided in the disclosure may be freely combined without conflicts to obtain new method embodiments. The characteristics disclosed in some product embodiments provided in the disclosure may be freely combined without conflicts to obtain new product embodiments. The features disclosed in several methods or device embodiments provided by the disclosure may be combined arbitrarily without conflict to obtain new method embodiments or device embodiments.

[00206] Those of ordinary skill in the art should know that all or part of the steps of the method embodiment may be implemented by related hardware instructed through a program, the program may be stored in a computer-readable storage medium, and the program is executed to execute the steps of the method embodiment. The storage medium includes: various media capable of storing program codes such as a mobile storage device, a ROM, a magnetic disk or a compact disc.

[00207] Or, the integrated unit of the disclosure may also be stored in a computer-readable storage medium when implemented in form of a software function module and sold or used as an independent product. Based on such an understanding, the technical solutions of the embodiments of the disclosure substantially or parts making contributions to the related art may be embodied in form of a software product. The computer software product is stored in a storage medium, including a plurality of instructions configured to enable a computer device (which may be a personal computer, a server cluster, a server or a tablet computer) to execute all or part of the method in each

embodiment of the disclosure. The storage medium includes various media capable of storing program codes such as a mobile hard disk, a ROM, a magnetic disk, or an optical disc. The storage medium includes: various media capable of storing program codes such as a mobile hard disk, a ROM, a magnetic disk or a compact disc.

[00208] In the embodiments of the disclosure, the descriptions about the same steps and the same contents in different embodiments may refer to those in the other embodiments. In the embodiments of the disclosure, term "and" does not influence the sequence of the steps.

[00209] The above are only the implementation modes of the disclosure and not intended to limit the scope of protection of the disclosure. Modifications or replacements are apparent to those skilled in the art within the technical scope disclosed by the disclosure, and these modifications or replacements shall fall within the scope of protection of the disclosure. Therefore, the scope of protection of the disclosure should be subject to the appended claims.

CLAIMS

1. A training method of an image generation model, wherein the image generation model comprises a trained first generator and an extrinsic style control module to be trained; the first generator is pre-trained with a real sample set and is capable of generating a real image based on an intrinsic style code;

wherein the method comprises:

obtaining a style sample set and the real sample set;

training the extrinsic style control module based on the real sample set to enable a first fused image generated by the image generation model to reconstruct a second fused image generated by the first generator; and

after training of the extrinsic style control module is completed, training the image generation model based on the style sample set and a randomly sampled intrinsic style code till a first convergence condition is met.

2. The method according to claim 1, wherein training the extrinsic style control module based on the real sample set to enable the first fused image generated by the image generation model to reconstruct the second fused image of the first generator comprises:

generating each second fused image based on the real sample set by using the first generator;

generating each first fused image based on the real sample set by using the image generation model, wherein the first fused image and the second fused image are coded based on the same pair of real image codes in the real sample set;

determining first reconstruction loss between each first fused image and the corresponding second fused image; and

performing iterative training on the extrinsic style control module based on the first reconstruction loss till a second convergence condition is met.

3. The method according to claim 2, wherein generating each second fused image based on the real sample set by using the first generator comprises:

respectively determining an image code of each real sample in the real sample set to obtain a real image code set;

obtaining each pair of real image codes from the real image code set; and

generating each corresponding second fused image based on each pair of real image codes by using the first generator;

correspondingly, generating each first fused image based on the real sample set by using the image generation model comprises:

generating each corresponding first fused image based on each pair of real image codes by using the image generation model.

4. The method according to claim 3, wherein the first generator comprises a first mapping network and a generative network; wherein generating each corresponding second fused image based on each pair of real image codes by using the first generator comprises:

respectively obtaining an intermediate style code of each real image code in each pair of real image codes based on the first mapping network, wherein each intermediate style code comprises N element feature sequences, and N is an integer greater than or equal to 2;

for each pair of real image codes, respectively extracting first k element features and last (N-k) element features from two intermediate style codes for fusion to obtain a fused intermediate style code, wherein k is an integer greater than or equal to 1 and less than N; and

generating the corresponding second fused image based on each fused intermediate style code by using the generative network.

5. The method according to claim 4, further comprising:

in a case where the first generator is a Style-Based Generator Architecture for Generative Adversarial Networks (StyleGAN), setting the value of k as 7, 6 and 5 in turn; and

for each value of k, obtaining the intermediate style code based on the first mapping network to obtain the fused intermediate style code, and obtaining the second fused image based on the generative network.

6. The method according to any one of claims 3 to 5, wherein generating each corresponding first fused image based on each pair of real image codes by using the image generation model comprises:

generating a corresponding real image based on a second real image code in each pair of real image codes by using the first generator, wherein the second real image code is a real image code corresponding to that the last (N-k) element features are extracted;

for each pair of real image codes, coding each generated real image by using a coder to obtain a corresponding extrinsic style code; and

generating a corresponding first fused image based on a first real image code in each pair of real image codes and a corresponding extrinsic style code by using an image generation model to be trained, wherein the first real image code is a real image code corresponding to that the first k element features are extracted;

7. The method according to any one of claims 1 to 6, wherein training the image generation model based on the style sample set and the randomly sampled intrinsic style code till the first convergence condition is met comprises:

performing iterative training on the image generation model based on the style sample set and the randomly sampled intrinsic style code to obtain each style image generated by the image generation model;

respectively determining and accumulating adversarial loss, second reconstruction loss, extrinsic style loss and intrinsic style loss based on each generated style image and the style sample and the intrinsic style code corresponding to each style image to obtain an accumulated sum, wherein the adversarial loss is adversarial loss of the first generator, the second reconstruction loss is loss of the generated style image corresponding to the corresponding style sample, the extrinsic style loss is loss of the generated style image corresponding to the corresponding style sample, and the intrinsic style loss is loss of the generated style image corresponding to the corresponding intrinsic style code; and

in a case wherer the accumulated sum meets the first convergence condition, completing training on the image generation model.

8. The method according to claim 7, wherein performing iterative training on the image generation model based on the style sample set and the randomly sampled intrinsic style code to obtain each style image generated by the image generation model comprises:

determining an extrinsic style code and an intrinsic style code of each style sample;

obtaining a first target style image generated by the image generation model by respectively taking the intrinsic style code and the extrinsic style code of each style sample as input of the first generator and input of the extrinsic style control module; and

obtaining a second target style image generated by the image generation model by respectively taking the randomly sampled intrinsic style code and the extrinsic style code of each style sample as input of the first generator and input of the extrinsic style control module;

wherein respectively determining the second reconstruction loss, the extrinsic style loss and the intrinsic style loss based on each generated style image, the style sample corresponding to each style image and the intrinsic style code comprises:

determining the second reconstruction loss based on each first target style image and a corresponding style sample;

determining the extrinsic style loss based on each second target style image and a corresponding style sample; and

determining the intrinsic style loss based on each second target style image and a corresponding random real image,

wherein the random real image is generated by the first generator based on the randomly sampled intrinsic style code.

9. The method according to claim 8, wherein determining the extrinsic style code and the intrinsic style code of each style sample comprises:

respectively coding each style sample by using the coder to obtain the extrinsic style code of a corresponding style sample; and

respectively performing de-stylization on each style sample to obtain the intrinsic style code of each style sample.

10. The method according to claim 9, wherein respectively performing de-stylization on each style sample to obtain the intrinsic style code of each style sample comprises:

determining an optimized style code of each style sample, wherein the optimized style code is capable of reconstructing a corresponding style sample through a trained second generator;

respectively reconstructing each optimized style code by using the trained first generator to obtain a reconstructed style sample, wherein the first generator and the second generator are obtained by training on a real image data set and a style image data set, respectively; and

coding each reconstructed style sample by using the coder to obtain the intrinsic style code of a corresponding style sample.

11. The method according to claim 10, comprising:

obtaining the first generator pre-trained with the real sample set; and

training the first generator on the style sample set to obtain the second generator.

12. The method according to any one of claims 8 to 11, wherein the extrinsic style control module comprises a second mapping network having the same structure as the first mapping network, an extrinsic structure style control module, an extrinsic color style control module and a residual module;

wherein obtaining the first target style image generated by the image generation model by respectively taking the intrinsic style code and the extrinsic style code of each style sample as input of the first generator and input of the extrinsic style control module, and obtaining the second target style image generated by the image generation model by respectively taking the randomly sampled intrinsic style code and the extrinsic style code of each style sample as input of the first generator and input of the extrinsic style control module comprise:

correspondingly obtaining the intrinsic intermediate style code of each style sample respectively based on the intrinsic style code of each style sample, and correspondingly obtaining the randomly sampled intrinsic intermediate style code based on the randomly sampled intrinsic style code by using the first mapping network;

respectively obtaining the extrinsic intermediate style code of each extrinsic style code by using the second mapping network, wherein each extrinsic intermediate style code comprises N element feature sequences;

determining a structure style code of each style sample based on a part of the extrinsic intermediate style code of each style sample by utilizing the extrinsic structure style control module;

extracting a color style code of each style sample based on the other part of the extrinsic intermediate style code of each style sample by utilizing the extrinsic color style control module;

performing structure style adjustment on a target image generated by the generative network based on the structure style code of each style sample by utilizing the residual module;

generating the first target style image based on the structure style code and the color style code of each style sample and the intrinsic intermediate style code of each style sample by utilizing the generative network; and

generating the second target style image based on the structure style code and the color style code of each style sample and the randomly sampled intrinsic intermediate style code.

13. An image processing method, comprising:

obtaining a style image and a content image to be fused;

performing extraction on the content image to obtain an intrinsic intermediate style code of the content image;

performing extraction on the style image to obtain an extrinsic intermediate style code of the style image; and

generating a fused image based on the intrinsic intermediate style code of the content image and the extrinsic intermediate style code of the style image.

14. The method according to claim 13, wherein generating the fused image based on the intrinsic intermediate style code of the content image and the extrinsic intermediate style code of the style image comprises:

determining a structure style code of the style image based on a part of the extrinsic intermediate style code of the style image;

extracting a color style code of the style image based on the other part of the extrinsic intermediate style code of the style image; and

generating the fused image based on the structure style code and the color style code of the style image and the intrinsic intermediate style code of the content image.

15. The method according to claim 14, wherein each extrinsic intermediate style code comprises N element feature sequences;

wherein determining the structure style code of the style image based on a part of the extrinsic intermediate style code of the style image comprises: determining the structure style code of the style image based on first k element features in the extrinsic intermediate style code of the style image; and

wherein extracting the color style code of the style image based on the other part of the extrinsic intermediate style code of the style image comprises: extracting the color style code of the style image based on last (N-k) element features of the extrinsic intermediate style code of the style image.

16. A training apparatus of an image generation model, wherein the image generation model comprises a trained first generator and an extrinsic style control module to be trained; the first generator is pre-trained with a real sample set and is capable of generating a real image based on an intrinsic style code;

wherein the apparatus comprises:

a first obtaining module, configured to obtain a style sample set and the real sample set;

a first training module, configured to train the extrinsic style control module based on the real sample set to enable a first fused image generated by the image generation model to reconstruct a second fused image generated by the first generator; and

a second training module, configured to, after training of the extrinsic style control module is completed, train the image generation model based on the style sample set and a randomly sampled intrinsic style code till a first convergence condition is met.

17. An image processing apparatus, comprising:

a first obtaining module, configured to obtain a style image and a content image to be fused;

a first extraction module, configured to perform extraction on the content image to obtain an intrinsic intermediate style code of the content image;

a second extraction module, configured to perform extraction on the style image to obtain an extrinsic intermediate style code of the style image; and

a generation module, configured to generate a fused image based on the intrinsic intermediate style code of the content image and the extrinsic intermediate style code of the style image.

18. A computer device, comprising a memory and a processor, wherein

the memory stores a computer program capable of running in the processor;

the processor executes the computer program to implement the steps in the method of any one of claims 1 to 12 or implement the steps in the method of any one of claims 13 to 15.

19. A nonvolatile computer-readable storage medium, storing one or more programs which are executable by one or more processors to implement the steps in the method of any one of claims 1 to 12 or implement the steps in the method of any one of claims 13 to 15.

20. A computer program product, comprising a non-transitory computer-readable storage medium having a computer program stored thereon, wherein when read and executed by a computer, the computer program implements the steps in the method of any one of claims 1 to 12 or the steps in the method of any one of claims 13 to 15.