

IMAGE GENERATION MODEL TRAINING METHOD, GENERATION METHOD, APPARATUS, AND DEVICE

[0001] The present disclosure is based upon and claims priority to Chinese Patent Application No. 202110143927.0, filed on February 2, 2021, and entitled "METHOD AND APPARATUS FOR TRAINING IMAGE GENERATION MODEL, METHOD AND APPARATUS FOR GENERATING IMAGES, AND DEVICES THEREOF," the content of which is herein incorporated by reference in its entirety.

TECHNICAL FIELD

[0002] Embodiments of the present disclosure relate to the technical field of machine learning and image processing, and in particular, relates to a method and apparatus for training an image generation model, a method and apparatus for generating images, and devices thereof.

BACKGROUND OF THE INVENTION

[0003] At present, machine learning is being more and more widely used in the technical field of image processing, for example, changing styles of images based on machine learning.

[0004] In the related art, a large number of training sample pairs are first acquired. Each of the training sample pairs includes a real face image and a corresponding face comic image. A large number of iterative trainings are performed on a learning model using a large number of training sample pairs, and the trained learning model is an image generation model.

SUMMARY OF THE INVENTION

[0005] Some embodiments of the present disclosure provide a method and apparatus for training an image generation model, a method and apparatus for generating images, and devices thereof.

The technical solutions are as follows.

[0006] According to one aspect of the embodiments of the present disclosure, a method for training an image generation model is provided. The method includes:

[0007] acquiring a first transformation model by training, wherein the first transformation model is configured to generate a first training image according to a first noise sample, the first training image being an image of a first style;

[0008] acquiring a reconstruction model by training based on the first transformation model, wherein the reconstruction model is configured to map an original image sample to a latent variable corresponding to the original image sample;

[0009] acquiring a second transformation model by training, wherein the second transformation model is configured to generate a second training image according to a second noise sample, the second training image being an image of a second style;

[0010] generating a grafted transformation model by grafting the first transformation model with the second transformation model; and

[0011] generating the image generation model based on the reconstruction model and the grafted transformation model, wherein the image generation model is configured to transform an image to be transformed of the first style into a target image of the second style.

[0012] According to another aspect of the embodiments of the present disclosure, a method for generating images is provided. The method includes:

[0013] generating a latent variable corresponding to an image to be transformed by inputting the image to be transformed of a first style into a reconstruction model; and

[0014] generating, based on the latent variable corresponding to the image to be transformed, a target image corresponding to the image to be transformed using a grafted transformation model, wherein the target image is an image of a second style;

[0015] wherein the grafted transformation model is a model generated by grafting a first transformation model with a second transformation model, the first transformation model is configured to generate an image of the first style according to a first noise sample, and the second transformation model is configured to generate an image of the second style according to a second noise sample.

[0016] According to still another aspect of the embodiments of the present disclosure, an apparatus for training an image generation model is provided. The apparatus includes:

[0017] a model training module, configured to acquire a first transformation model by training, wherein the first transformation model is configured to generate a first training image according to a first noise sample, the first training image being an image of a first style, wherein

[0018] the model training module is further configured to acquire a reconstruction model by training based on the first transformation model, wherein the reconstruction model is configured to map an original image sample to a latent variable corresponding to the original image sample, and

[0019] the model training module is further configured to acquire a second transformation model by training, wherein the second transformation model is configured to generate a second training

image according to a second noise sample, the second training image being an image of a second style; and

[0020] a model generation module, configured to generate a grafted transformation model by grafting the first transformation model with the second transformation model, wherein

[0021] the model generation module is further configured to generate the image generation model based on the reconstruction model and the grafted transformation model, wherein the image generation model is configured to transform an image to be transformed of the first style into a target image of the second style.

[0022] According to still another aspect of the embodiments of the present disclosure, an apparatus for generating images is provided. The apparatus includes:

[0023] a variable generation module, configured to generate a latent variable corresponding to an image to be transformed by inputting the image to be transformed of a first style into a reconstruction model; and

[0024] an image generation module, configured to generate, based on the latent variable corresponding to the image to be transformed, a target image corresponding to the image to be transformed using a grafted transformation model, wherein the target image is an image of a second style;

[0025] wherein the grafted transformation model is a model generated by grafting a first transformation model with a second transformation model, the first transformation model is configured to generate an image of the first style according to a first noise sample, and the second transformation model is configured to generate an image of the second style according to a second noise sample.

[0026] According to still another aspect of the embodiments of the present disclosure, a computer device is provided. The computer device includes a processor and a memory. The memory stores one or more computer programs therein. The one or more computer programs, when loaded and run by the processor of the computer device, cause the computer device to perform the method for training the image generation model as described above, or perform the method for generating the images as described above.

[0027] According to still another aspect of the embodiments of the present disclosure, a computer-readable storage medium storing one or more computer programs therein is provided. The one or more computer programs, when loaded and run by a processor, cause the processor to perform the method for generating the image generation model as described above, or perform the method for generating the images as described above.

[0028] According to still another aspect of the embodiments of the present disclosure, a computer program product is provided. The computer program product, when loaded and executed by a processor, causes to perform the method for training the image generation model as described above, or perform the method for generating the images as described above.

BRIEF DESCRIPTION OF DRAWINGS

[0029] FIG. 1 is a flowchart of a method for training an image generation model according to some embodiments of the present disclosure;

[0030] FIG. 2 is a schematic diagram of grafting models according to some embodiments of the present disclosure;

[0031] FIG. 3 is a flowchart of a method for training an image generation model according to some embodiments of the present disclosure;

[0032] FIG. 4 is a schematic structural diagram of a first transformation model according to some embodiments of the present disclosure;

[0033] FIG. 5 is a schematic structural diagram of a reconstruction model according to some embodiments of the present disclosure;

[0034] FIG. 6 is a flowchart of a method for generating images according to some embodiments of the present disclosure;

[0035] FIG. 7 is a schematic diagram of a method for generating images according to some embodiments of the present disclosure;

[0036] FIG. 8 is a block diagram of an apparatus for training an image generation model according to some embodiments of the present disclosure;

[0037] FIG. 9 is a block diagram of an apparatus for training an image generation model according to some embodiments of the present disclosure; and

[0038] FIG. 10 is a block diagram of an apparatus for generating images according to some embodiments of the present disclosure.

DETAILED DESCRIPTION

[0039] Face transformation is one of the common image processing methods in computer vision. The image generation model trained by using the method according to the embodiments of the present disclosure is applicable to a target application (such as a video application, a social application, a live streaming application, a shopping application, or an image processing application). A transformed target image is generated by inputting an image to be transformed (a picture or a video) into the target application. The image generation model is written to the target

application, or to a server that provides background services to the target application. Optionally, the face transformation is applicable to scenarios such as caricaturing faces, stylizing faces, and editing face attributes (such as changing the age or the gender from the vision), which are not specifically limited herein.

[0040] In the method according to some embodiments of the present disclosure, each step is performed by a computer device, which is an electronic device having capabilities of calculating, processing, and storing data. The computer device is a terminal, such as a personal computer (PC), a tablet computer, a smartphone, a wearable device, an intelligent robot, or the like, or a server. The server is an independent physical server, a server cluster, or a distributed system consisting of a plurality of physical servers, or a cloud server providing cloud computing services.

[0041] The technical solutions of the present disclosure are described and illustrated hereinafter by several embodiments.

[0042] FIG. 1 is a flowchart of a method for training an image generation model according to some embodiments of the present disclosure. Referring to FIG. 1, in the embodiments, the description is given using a scenario where the method is applicable to the computer device described above as an example. The method includes the following steps (step 101 to step 105).

[0043] In step 101, a first transformation model is acquired by training. The first transformation model is configured to generate a first training image according to a first noise sample. The first training image is an image of a first style.

[0044] In some embodiments, the corresponding first training image is acquired by inputting the first noise sample into the first transformation model, and then parameters of the first transformation model are adjusted based on the corresponding first training image. Thereafter, the next iterative training is performed on the first transformed model by inputting the remaining first noise samples into the first transformation model until the training of the first transformation model is completed. The first noise sample is vectorized and input into the first transformation model in vector form or matrix form.

[0045] In the embodiments, the first transformation model is configured to generate the first training image that conforms to a first feature. The first feature is a feature of a person, an object, or a scene that objectively exists. In some examples, the first feature is a real face feature. That is, the first training image is an image that simulates a real face. In the training process, a weight parameter of the first transformation model is adjusted by discriminating a possibility that the face in the first training image conforms to the real face feature, such that the first transformation model is capable of generating an image that conforms to the real face feature.

[0046] Optionally, the first noise sample is a randomly generated noise. In some embodiments, a probability density function of the first noise sample subjects to the Gaussian distribution.

[0047] In step 102, a reconstruction model is acquired by training based on the first transformation model. The reconstruction model is configured to map an original image sample to a latent variable corresponding to the original image sample.

[0048] Optionally, the latent variable is a variable that fails to be observed. In some embodiments, the reconstruction model is capable of mapping, by encoding, the original image sample to the corresponding latent variable, such that the high-dimensional original image sample is transformed into the lower-dimensional latent variable, which reduces computational complexity and computation time of the model, and thus the time cost of training the model is further improved.

[0049] In step 103, a second transformation model is acquired by training. The second transformation model is configured to generate a second training image according to a second noise sample. The second training image is an image of a second style.

[0050] In some embodiments, the corresponding second training image is acquired by inputting the second noise sample into the second transformation model, and then parameters of the second transformation model are adjusted based on the corresponding second training image. Thereafter, the next iterative training is performed on the second transformed model by inputting the remaining second noise samples into the second transformation model until the training of the second transformation model is completed.

[0051] In the embodiments, the second style is a comic style (such as a pop comic style or a sketch comic style), an ink painting style, a brush painting style, an oil painting style, a watercolor painting style, a drawing style, an abstract painting style, or an image painting style, which is not limited herein.

[0052] Optionally, during the training process, an initial weight parameter of the second transformation model is the weight parameter of the first transformation model. That is, the second transformation model is further acquired by training based on the first transformation model.

[0053] In step 104, a grafted transformation model is generated by grafting the first transformation model with the second transformation model.

[0054] Optionally, the grafted transformation model is generated by grafting different portions of the weight parameters respectively selected from the first transformation model and the second transformation model. The grafted transformation model has both the characteristic of the first transformation model to generate or maintain the first feature of the image and the characteristic of the second transformation model to generate the image of the second style.

[0055] In some embodiments, the grafted transformation model is generated by grafting n layers of weight network in a plurality of layers of weight network of the first transformation model with m layers of weight network in a plurality of layers of weight network of the second transformation model. The number of layers of the n layers of weight network and the m layers of weight network are different. n is a positive integer, and m is a positive integer. Optionally, the first transformation model and the second transformation model have the same or similar structure and have the same number of layers of weight network. The grafted transformation model is generated by splicing the last n layers of weight network of the first transformation model after the first m layers of weight network of the second transformation model. Optionally, the grafted transformation model has the same number of layers as the weight network of the first transformation model or the second transformation model. The grafted transformation model has the same or similar structure as the first transformation model or the second transformation model.

[0056] It should be noted that n is 2, 3, 4, 5, 6, 7, 8, 9, and the like. Optionally, the specific value of n is defined by those skilled in the art according to the actual needs, which is not limited herein. m is 2, 3, 4, 5, 6, 7, 8, 9, and the like. Optionally, the specific value of m is defined by those skilled in the art according to the actual needs, which is not limited herein.

[0057] Optionally, the sum of n and m is the number of layers of the weight network of the first transformation model or the second transformation model.

[0058] In some examples, as illustrated in FIG. 2, the first transformation model 21 and the second transformation model 22 have the same structure, and the weight networks of the two models both have 14 layers. The grafted transformation model 25 is acquired by splicing the last six layers of weight network 23 of the first transformation model 21 after the first eight layers of weight network 24 of the second transformation model 22.

[0059] In other embodiments, the grafted transformation model is generated by performing a summing or averaging or difference operation on weight parameters of the plurality of layers of weight network of the first transformation model and corresponding weight parameters of the plurality of layers of weight network of the second transformation model. Optionally, the first transformation model, the second transformation model, and the grafted transformation model have the same or similar structure, and the weight networks thereof have exactly the same structure and number of layers. In some examples, an i^{th} weight parameter of the weight network of the grafted transformation model is acquired by performing the summing or averaging, or difference operation on an i^{th} weight parameter of the weight network of the first transformation model and an i^{th} weight parameter of the weight network of the second transformation model, wherein i is a positive integer; and the like, all weight parameters of the weight network of the grafted

transformation model are acquired. Optionally, the summing operation is a weighted summing operation, and weight coefficients used to calculate the weight parameters of the weight network of the grafted transformation model are the same.

[0060] In step 105, an image generation model is generated based on the reconstruction model and the grafted transformation model.

[0061] Optionally, the image generation model is configured to transform an image to be transformed of the first style into a target image of the second style. In some embodiments, the image generation model is acquired by combining the reconstruction model with the grafted transformation model. The image to be transformed is an image of the first style, and the target image is an image of the second style. The target image is generated by inputting the image to be transformed into the image generation model, such that the image to be transformed is transformed from the first style to the second style, and thus the image to be transformed achieves the style transformation.

[0062] In some embodiments, step 105 includes the following sub-steps.

[0063] 1. The combined transformation model is acquired by combining the reconstruction model with the grafted transformation model.

[0064] 2. A fourth training sample set is acquired. The fourth training sample set includes at least one original image sample and an image of the second style corresponding to the at least one original image sample.

[0065] 3. The image generation model is generated by fine-tuning the combined transformation model using the fourth training sample set.

[0066] In some embodiments, the combined transformation model is acquired by splicing the grafted transformation model and the reconstruction model; thereafter, the image generation model is acquired by fine-tuning a weight parameter of the combined transformation model by model training.

[0067] In summary, in the technical solutions according to the embodiments of the present disclosure, the first transformation model and the second transformation model are separately trained. The first transformation model has the advantage of well maintaining the target feature of the image, and the second transformation model has the advantage of generating the image of the second style in an impressive style. In the training process, the first transformation model and the second transformation model are trained separately for their own target advantage, and the grafted transformation model is generated by grafting the trained first transformation model and the trained second transformation model. The image generation model is generated based on the reconstruction model and the grafted transformation model. In this way, the complexity of the

model training is greatly reduced, such that the required training samples and the times of iterations of the training are fewer, and thus the time cost of the model training is saved.

[0068] FIG. 3 is a flowchart of a method for training an image generation model according to some embodiments of the present disclosure. Referring to FIG. 3, in the embodiments, the description is given using a scenario where the method is applicable to the computer device described above as an example. The method includes the following steps (step 301 to step 312).

[0069] In step 301, a first training sample set is acquired. The first training sample set includes a plurality of first noise samples.

[0070] Optionally, the first noise sample is a noise sample employed in training a first transformation model. In some embodiments, the first training sample set is acquired by generating a plurality of random first noise samples by a random noise generation function.

[0071] Optionally, the first transformation model is a style generative adversarial network-V2 (StyleGAN-V2) model. As illustrated in FIG. 4, the first transformation model includes a first mapping network 41 and a first synthesis network 42. The mapping network is capable of untwisting the input. The first mapping network 41 includes eight fully connected layers (FC). The first synthesis network 42 includes at least one learnable affine transformation layer, at least one modulation model (mod-demod), at least one up-sample layer, at least one constant layer, and at least one noise parameter layer. Optionally, the modulation model uses a first equation, a second equation, and a third equation as follows for calculation.

[0072] First equation

$$w'_{ijk} = s_i \cdot w_{ijk}$$

[0073] Second equation

$$\sigma_j = \sqrt{\sum_{i,k} w'_{ijk}{}^2}$$

[0074] Third equation

$$w''_{ijk} = w'_{ijk} / \sqrt{\sum_{i,k} w'_{ijk}{}^2 + \epsilon}$$

[0075] i, j, and k indicate numerals of latent variables of the modulation model in different dimensions. w_{ijk} indicates a latent variable numbered (i, j, k). w'_{ijk} indicates a calculation result of the first equation. σ_j indicates a calculation result of the second equation. ϵ indicates a very small number. w''_{ijk} indicates a calculation result of the third equation.

[0076] In some embodiments, the first transformation model includes the first mapping network and the first synthesis network.

[0077] In step 302, latent variables respectively corresponding to the plurality of first noise samples are acquired by inputting the plurality of first noise samples into the first mapping network.

[0078] In some embodiments, the first mapping network includes a plurality of fully connected layers. The plurality of first noise samples are processed by the plurality of fully connected layers upon being input into the first mapping network. Then the latent variables respectively corresponding to the plurality of first noise samples are acquired.

[0079] In step 303, first training images respectively corresponding to the plurality of first noise samples are acquired by inputting the latent variables respectively corresponding to the plurality of first noise samples into the first synthesis network.

[0080] In some embodiments, after the latent variables respectively corresponding to the plurality of first noise samples have been affine transformed, modulated, and up-sampled in the first synthesis network upon being input into the first synthesis network, the first training images respectively corresponding to the plurality of first noise samples are acquired.

[0081] In step 304, based on the first training images respectively corresponding to the plurality of first noise samples, a weight parameter of the first transformation model is adjusted.

[0082] Optionally, upon acquiring the first training images respectively corresponding to the plurality of first noise samples, the weight parameter of the first transformation model is adjusted by discrimination results of the first training images respectively corresponding to the plurality of first noise samples.

[0083] In some embodiments, the first transformation model includes a first discrimination network. Step 304 further includes the following sub-steps.

[0084] 1. First discrimination losses respectively corresponding to the plurality of first noise samples are acquired by inputting the first training images respectively corresponding to the plurality of first noise samples into the first discrimination network.

[0085] 2. The weight parameter of the first transformation model is adjusted based on the first discrimination losses respectively corresponding to the plurality of first noise samples.

[0086] In this implementation, the first training images respectively corresponding to the plurality of first noise samples are discriminated using the first discrimination network in the first transformation model, and the generation qualities of the first training images respectively corresponding to the plurality of first noise samples are determined, such that the first discrimination losses respectively corresponding to the plurality of first noise samples are

determined. Based on this, the weight parameter of the transformation model is adjusted. Optionally, in adjusting the weight parameter of the transformation model, a weight parameter of the first discrimination network is also adjusted. Optionally, the first discrimination network is a generative adversarial network (GAN), and the first discrimination loss is expressed as GAN_loss.

[0087] In some examples, the first training image is a face image simulated based on the first noise sample, and the first discrimination network is configured to discriminate a probability that the first training image conforms to a real face image. For example, the more complete the face in the first training image, the smaller the first discrimination loss; and for example, the closer the ratio of the facial features of the face in the first training image is to the ratio of the facial features in the real face, the smaller the first discrimination loss is.

[0088] In some embodiments, upon the completion of the training of the first transformation model, the first mapping network of the first transformation model is removed, and the rest of the first transformation model other than the first mapping network is determined as the first transformation model.

[0089] In step 305, a second training sample set is acquired. The second training sample set includes a plurality of original image samples.

[0090] Optionally, the original image sample is an image of a first style. In some embodiments, the original image sample is an unprocessed real face image.

[0091] In step 306, latent variables respectively corresponding to the plurality of original image samples are generated by inputting the plurality of original image samples into a reconstruction model.

[0092] In some embodiments, upon being input into the reconstruction model, the plurality of original image samples are separately encoded by the reconstruction model, and the reconstruction model generates the latent variables respectively corresponding to the plurality of original image samples.

[0093] In some examples, as illustrated in FIG. 5, the reconstruction model includes an input layer 51, a plurality of residual network (resblock) layers 52, and a fully connected layer 53.

[0094] In step 307, reconstructed images respectively corresponding to the plurality of original image samples are generated by inputting the latent variables respectively corresponding to the plurality of original image samples into the first transformation model.

[0095] In some embodiments, the reconstructed images respectively corresponding to the plurality of original image samples are generated by directly inputting the latent variables respectively corresponding to the plurality of original image samples into the first synthesis network of the first transformation model. Optionally, the plurality of original image samples and the reconstructed

images respectively corresponding to the plurality of original image samples are images of the first style.

[0096] It should be noted that in training the reconstruction model, the first transformation model is a model that has been trained and parameters of the first transformation model do not change.

[0097] In step 308, losses of the reconstruction model respectively corresponding to the plurality of original image samples are determined based on the plurality of original image samples and the reconstructed images respectively corresponding to the plurality of original image samples.

[0098] In some embodiments, a plurality of sub-losses in the loss of the reconstruction model are determined based on the plurality of original image samples and outputs of the reconstructed images respectively corresponding to the plurality of original image samples in networks or functions.

[0099] In some embodiments, step 308 further includes the following sub-steps.

[00100] 1. A first sub-loss is determined based on an output result acquired by inputting each of the reconstructed images respectively corresponding to the plurality of original image samples into the first discrimination network. The first sub-loss indicates a first characterization of the reconstructed image.

[00101] Optionally, the first characterization is a degree of conformity of the reconstructed image to a feature that the reconstructed image needs to contain. In some examples, in the case that the reconstructed image is a face image, the first characterization is a degree of conformity of the reconstructed image to the facial feature. The closer the face in the reconstructed image is to a real face, the smaller the first sub-loss is.

[00102] 2. A second sub-loss is determined based on an output result acquired by inputting each of the plurality of original image samples and each of the reconstructed images respectively corresponding to the plurality of original image samples into a perceptual network. The second sub-loss indicates a first degree of conformity of the original image sample to the reconstructed image corresponding to the original image sample with respect to a target feature.

[00103] Optionally, whether the reconstructed image is an image generated according to an original image sample is determined based on the target feature. In some examples, the target feature is a face identification feature, and the face identification feature is configured to distinguish different faces. A probability that the original image sample and the reconstructed image corresponding to the original image sample are face images of the same person is determined by comparing a face identification feature in the original image sample to a face identification feature in the reconstructed image corresponding to the original image sample by the perceptual network, such that the second sub-loss is determined. Optionally, the second sub-

loss is a perceptual loss. Optionally, the second sub-loss is calculated with reference to a fourth equation as follows.

[00104] Fourth equation

$$\text{Perceptual loss} = E((VGG(x) - VGG(G(x)))^2)$$

[00105] Perceptual loss indicates the second sub-loss. x indicates the original image sample. $VGG(x)$ indicates an output result acquired by inputting the original image sample into a visual geometry group network-16 (VGG-16). $G(x)$ indicates the reconstructed image corresponding to the original image sample. $VGG(G(x))^2$ indicates an output result acquired by inputting the reconstructed image corresponding to the original image sample into the VGG-16.

[00106] 3. A third sub-loss is determined based on an output result acquired by inputting each of the plurality of original image samples and each of the reconstructed images respectively corresponding to the plurality of original image samples into a regression function. The third sub-loss indicates a second degree of conformity of the original image sample to the reconstructed image corresponding to the original image sample with respect to the target feature.

[00107] Optionally, the regression function is an L1 function or an L2 function. The third sub-loss is determined by the output result acquired by inputting the original image sample and the reconstructed image corresponding to the original image sample into the L function. Optionally, the third sub-loss is expressed as L1_loss. Optionally, the third sub-loss is calculated with reference to a fifth equation as follows.

[00108] Fifth equation:

$$L1_loss = E(x - G(x))$$

[00109] L1_loss indicates the third sub-loss. x indicates the original image sample. $VGG(x)$ indicates the output result acquired by inputting the original image sample into the VGG-16.

[00110] 4. The loss of the reconstruction model is determined based on the first sub-loss, the second sub-loss, and the third sub-loss.

[00111] In some examples, the loss of the reconstruction model is calculated by directly summing or weighted summing the first sub-loss, the second sub-loss, and the third sub-loss. In some embodiments, in weighted summing the first sub-loss, the second sub-loss, and the third sub-loss, values of weights respectively corresponding to the first sub-loss, the second sub-loss, and the third sub-loss are defined by those skilled in the art according to the actual needs, which are not limited herein.

[00112] In some examples, the loss of the reconstruction model is calculated with reference to a sixth equation as follows.

[00113] Sixth equation:

$$\text{Loss} = \text{GAN}_{\text{loss}} + \text{Perceptual loss} + \text{L1}_{\text{loss}}$$

[00114] Loss indicates the loss of the reconstruction model. GAN_{loss} indicates the first sub-loss. Perceptual loss indicates the second sub-loss. L1_{loss} indicates the third sub-loss.

[00115] In step 309, a weight parameter of the reconstruction model is adjusted based on the losses of the reconstruction model respectively corresponding to the plurality of original image samples.

[00116] Optionally, based on the losses of the reconstruction model respectively corresponding to the plurality of original image samples, the weight parameter of the reconstruction model is adjusted, such that the loss of the reconstruction model is as small as possible. In some embodiments, when the times of iterative training reach a predetermined number of times, the training of the reconstruction model is completed. In other embodiments, in the case that the losses of the reconstruction model are continuously less than a reconstruction loss threshold, the training of the reconstruction model is complete. Optionally, the specific value of the reconstruction loss threshold is defined by those skilled in the art according to the actual needs, which is not limited herein.

[00117] In step 310, a second transformation model is acquired by training. The second transformation model is configured to generate a second training image according to a second noise sample. The second training image is an image of a second style.

[00118] The content of step 310 is the same or similar to the contents of steps 301 to 304 in the embodiments of the present disclosure, which is not described herein any further.

[00119] In step 311, a grafted transformation model is generated by grafting the first transformation model with the second transformation model.

[00120] The content of step 311 is the same or similar to the content of step 104 in the embodiments described above, which is not described herein any further.

[00121] In step 312, an image generation model is generated based on the reconstruction model and the grafted transformation model. The image generation model is configured to transform an image to be transformed of the first style into a target image of the second style.

[00122] The content of step 312 is the same or similar to the content of step 105 in the embodiments described above, which is not described herein any further.

[00123] In summary, in the technical solutions according to the embodiments of the present disclosure, the first transformation model is acquired first by training, and then the reconstruction model is acquired by training based on the first transformation model. In this way, the time required to train the reconstruction model is reduced, such that the overall time cost of the model training is further reduced.

[00124] In the embodiments of the present disclosure, the weight parameter of the first transformation model that has been completely trained is determined as the initial parameter of the second transformation model when training, such that the second transformation model is capable of retaining some of the characteristics of the first transformation model, which is convenient for improving the compatibility between the first transformation model and the second transformation model upon grafting.

[00125] A method for generating images according to some embodiments of the present disclosure is described hereinafter in conjunction with FIG. 6 and FIG. 7. In the embodiments, the description is given using a scenario where the method is applicable to the computer device described above as an example. As illustrated in FIG. 6, the method includes the following steps (step 601 to step 602).

[00126] In step 601, as illustrated in FIG. 7, a latent variable 73 corresponding to an image to be transformed 71 is generated by inputting the image to be transformed 71 of the first style into a reconstruction model 72.

[00127] Optionally, the image to be transformed 71 is an unprocessed real face image.

[00128] In step 602, a target image 75 corresponding to the image to be transformed 71 is generated based on the latent variable 73 corresponding to the image to be transformed 71 using a grafted transformation model 74. The target image 75 is an image of the second style.

[00129] Optionally, the grafted transformation model 74 is a model generated by grafting a first transformation model with a second transformation model. The first transformation model is configured to generate a first training image according to a first noise sample, wherein the first training image is an image of the first style. The second transformation model is configured to generate a second training image according to a second noise sample, wherein the second training image is an image of the second style.

[00130] In some embodiments, a weight network, which is from the second transformation model, in the grafted transformation model 74 first stylizes the image to be transformed 71 based on the latent variable 73 corresponding to the image to be transformed 71, such that the image to be transformed 71 is transformed into an image of the second style. Thereafter, a weight network, which is from the first transformation model, of the grafted transformation model 74 adjusts a target feature in the stylized image to be transformed 71, such that a final target image 75 is acquired. For the description of the target features, reference may be made to step 308 in the embodiments of FIG. 3 above, which is not described herein any further.

[00131] In summary, in the technical solutions according to the embodiments of the present disclosure, the target image 75 is generated by the image generation model acquired by grafting

the first transformation model with the second transformation model. The image generation model (the grafted transformation model 74) has both the advantage of the first transformation model in well maintaining the target feature of the image and the advantage of the second transformation model in generating the image of the second style which is an impressive style. Therefore, the image generation model is capable of generating the target image 75 in an impressive style under the premise of well maintaining the target feature of the image to be transformed, which improves the transformation effect of the image generation model.

[00132] Apparatus embodiments of the present disclosure used to perform the method embodiments of the present disclosure are described hereinafter. For details not disclosed in the apparatus embodiments of the present disclosure, reference may be made to the method embodiments of the present disclosure, which are not described herein any further.

[00133] FIG. 8 is a block diagram of an apparatus for training an image generation model according to some embodiments of the present disclosure. Referring to FIG. 8, the apparatus has a function for practicing the above embodiments illustrating the method for training the image generation model. The function is implemented by hardware or is implemented by hardware executing corresponding software. The apparatus 800 includes a model training module 810 and a model generation module 820.

[00134] The model training module 810 is configured to acquire a first transformation model by training. The first transformation model is configured to generate a first training image according to a first noise sample. The first training image is an image of a first style.

[00135] The model training module 810 is further configured to acquire a reconstruction model by training based on the first transformation model. The reconstruction model is configured to map an original image sample to a latent variable corresponding to the original image sample.

[00136] The model training module 810 is further configured to acquire a second transformation model by training. The second transformation model is configured to generate a second training image according to a second noise sample. The second training image is an image of a second style.

[00137] The model generation module 820 is configured to generate a grafted transformation model by grafting the first transformation model with the second transformation model.

[00138] The model generation module 820 is further configured to generate the image generation model based on the reconstruction model and the grafted transformation model. The image generation model is configured to transform an image to be transformed of the first style into a target image of the second style.

[00139] In summary, in the technical solutions according to the embodiments of the present disclosure, the first transformation model and the second transformation model are separately trained. The first transformation model has the advantage of well maintaining the target feature of the image, and the second transformation model has the advantage of generating the image of the second style which is an impressive style. In the training process, the first transformation model and the second transformation model are trained separately for their own target advantage, and the grafted transformation model is generated by grafting the trained first transformation model and the trained second transformation model. The image generation model is generated based on the reconstruction model and the grafted transformation model. In this way, the complexity of the model training is greatly reduced, such that the required training samples and the times of iterations of the training are fewer, and thus the time cost of the model training is saved.

[00140] In some exemplary embodiments, the first transformation model includes a first mapping network and a first synthesis network. As illustrated in FIG. 9, the model training module 810 includes a sample acquisition sub-module 811, a variable generation sub-module 812, an image generation sub-module 813, and a parameter adjustment sub-module 814.

[00141] The sample acquisition sub-module 811 is configured to acquire a first training sample set. The first training sample set includes a plurality of first noise samples.

[00142] The variable generation sub-module 812 is configured to acquire latent variables respectively corresponding to the plurality of first noise samples by inputting the plurality of first noise samples into the first mapping network.

[00143] The image generation sub-module 813 is configured to acquire first training images respectively corresponding to the plurality of first noise samples by inputting the latent variables respectively corresponding to the plurality of first noise samples into the first synthesis network.

[00144] The parameter adjustment sub-module 814 is configured to adjust a weight parameter of the first transformation model based on the first training images respectively corresponding to the plurality of first noise samples.

[00145] In some exemplary embodiments, the first transformation model includes a first discrimination network. As illustrated in FIG. 9, the parameter adjustment sub-module 814 is configured to:

[00146] acquire first discrimination losses respectively corresponding to the plurality of first noise samples by inputting the first training images respectively corresponding to the plurality of first noise samples into the first discrimination network; and

[00147] adjust the weight parameter of the first transformation model based on the first discrimination losses respectively corresponding to the plurality of first noise samples.

[00148] In some exemplary embodiments, as illustrated in FIG. 9, the model training module 810, includes a loss determination sub-module 815.

[00149] The sample acquisition sub-module 811 is further configured to acquire a second training sample set. The second training sample set includes a plurality of original image samples.

[00150] The variable generation sub-module 812 is further configured to generate latent variables respectively corresponding to the plurality of original image samples by inputting the plurality of original image samples into the reconstruction model.

[00151] The image generation sub-module 813 is further configured to generate reconstructed images respectively corresponding to the plurality of original image samples by inputting the latent variables respectively corresponding to the plurality of original image samples into the first transformation model. The plurality of original image samples and the reconstructed images respectively corresponding to the plurality of original image samples are images of the first style.

[00152] The loss determination sub-module 815 is configured to determine, based on the plurality of original image samples and the reconstructed images respectively corresponding to the plurality of original image samples, losses of the reconstruction model respectively corresponding to the plurality of original image samples.

[00153] The parameter adjustment sub-module 814 is further configured to adjust the weight parameter of the reconstruction model based on the losses of the reconstruction model respectively corresponding to the plurality of original image samples.

[00154] In some exemplary embodiments, the first transformation model includes the first discrimination network. The loss determination sub-module 815, as illustrated in FIG. 9, is configured to:

[00155] determine a first sub-loss based on an output result acquired by inputting each of the reconstructed images respectively corresponding to the plurality of original image samples into the first discrimination network, wherein the first sub-loss indicates a first characterization of the reconstructed image;

[00156] determine a second sub-loss based on an output result acquired by inputting each of the plurality of original image samples and each of the reconstructed images respectively corresponding to the plurality of original image samples into a perceptual network, wherein the second sub-loss indicates a first degree of conformity of the original image sample to the reconstructed image corresponding to the original image sample with respect to a target feature;

[00157] determine a third sub-loss based on an output result acquired by inputting each of the plurality of original image samples and each of the reconstructed images respectively corresponding to the plurality of original image samples into a regression function, wherein the

third sub-loss indicates a second degree of conformity of the original image sample to the reconstructed image corresponding to the original image sample with respect to the target feature; and

[00158] determine the loss of the reconstruction model based on the first sub-loss, the second sub-loss, and the third sub-loss.

[00159] In some exemplary embodiments, during the training process, an initial weight parameter of the second transformation model is the weight parameter of the first transformation model.

[00160] In some exemplary embodiments, the model generation module 820 is configured to generate the grafted transformation model by grafting n layers of weight network in a plurality of layers of weight network of the first transformation model with m layers of weight network in a plurality of layers of weight network of the second transformation model, wherein layers of the n layers of weight network and the m layers of weight network are different, n is a positive integer, and m is a positive integer; or to generate the grafted transformation model by performing a summing or averaging or difference operation on weight parameters of the plurality of layers of weight network of the first transformation model and corresponding weight parameters of the plurality of layers of weight network of the second transformation model.

[00161] In some exemplary embodiments, the model generation module 820 is configured to:

[00162] acquire a combined transformation model by combining the reconstruction model and the grafted transformation model;

[00163] acquire a fourth training sample set, wherein the fourth training sample set includes at least one original image sample and an image of the second style corresponding to the at least one original image sample; and

[00164] generate the image generation model by fine-tuning the combined transformation model using the fourth training sample set.

[00165] FIG. 10 is a block diagram of an apparatus for generating images according to some embodiments of the present disclosure. Referring to FIG. 10, the apparatus has a function for practicing the above embodiments illustrating the method for generating the images. The function is implemented by hardware or by hardware executing corresponding software. The apparatus 1000 includes a variable generation module 1010 and an image generation module 1020.

[00166] The variable generation module 1010 is configured to generate a latent variable corresponding to an image to be transformed by inputting the image to be transformed of the first style into a reconstruction model.

[00167] The image generation module 1020 is configured to generate, based on the latent variable corresponding to the image to be transformed, a target image corresponding to the image to be

transformed using a grafted transformation model. The target image is an image of the second style. The grafted transformation model is a model generated by grafting a first transformation model with a second transformation model. The first transformation model is configured to generate an image of the first style according to a first noise sample, and the second transformation model is configured to generate an image of the second style according to a second noise sample.

[00168] In summary, in the technical solutions according to the embodiments of the present disclosure, the target image is generated by the image generation model acquired by grafting the first transformation model with the second transformation model. The image generation model (the grafted transformation model) has both the advantage of the first transformation model in well maintaining the target feature of the image and the advantage of the second transformation model in generating the image of the second style which is an impressive style. Therefore, the image generation model is capable of generating the target image in an impressive style under the premise of well maintaining the target feature of the image to be transformed, which improves the transformation effect of the image generation model.

[00169] It should be noted that description is only given to the division of the functional modules for the apparatus according to the above embodiments. In practice, the functions of the apparatus may be assigned to and implemented by different functional modules according to actual needs. That is, in terms of internal structure, the apparatus is divided into different functional modules to implement a part or all of the functions as described above. In addition, the apparatus according to the above embodiments is based on the same concept as the method embodiments as described above, and the specific implementation process of the apparatus is detailed in the method embodiments, which is not described herein any further.

[00170] In some exemplary embodiments, a computer device is provided. The computer device includes a processor and a memory, wherein the memory stores one or more computer programs therein. The one or more computer programs, when loaded and run by the processor of the computer device, cause the computer device to perform the method for training the image generation model as described above.

[00171] In some exemplary embodiments, a computer device is provided. The computer device includes a processor and a memory, wherein the memory stores one or more computer programs therein. The one or more computer programs, when loaded and run by the processor of the computer device, cause the computer device to perform the method for generating the images as described above.

[00172] In some exemplary embodiments, a computer-readable storage medium is provided. The computer-readable storage medium stores one or more computer programs therein. The one or

more computer programs, when loaded and run by a processor, cause the processor to perform the method for training the image generation model as described above.

[00173] In some exemplary embodiments, a computer-readable storage medium is provided. The computer-readable storage medium stores one or more computer programs therein. The one or more computer programs, when loaded and run by a processor, cause the processor to perform the method for generating the images as described above.

[00174] In some exemplary embodiments, a computer program product is provided. The computer program product, when loaded and executed by a processor, causes the processor to perform the method for training the image generation model as described above.

[00175] In some exemplary embodiments, a computer program product is provided. The computer program product, when loaded and executed by a processor, causes the processor to perform the method for generating the images as described above.

[00176] It should be noted that the term "a plurality of" herein refers to two or more. The symbol "/" generally denotes an "OR" relationship between contextual objects.

[00177] Described above are merely exemplary embodiments of the present disclosure, and are not intended to limit the present disclosure. Therefore, any modifications, equivalent substitutions, improvements, and the like made within the spirit and principles of the present disclosure shall be included in the protection scope of the present disclosure.

Claims

What is claimed is:

1. A method for training an image generation model, comprising:

acquiring a first transformation model by training, wherein the first transformation model is configured to generate a first training image according to a first noise sample, the first training image being an image of a first style;

acquiring a reconstruction model by training based on the first transformation model, wherein the reconstruction model is configured to map an original image sample to a latent variable corresponding to the original image sample;

acquiring a second transformation model by training, wherein the second transformation model is configured to generate a second training image according to a second noise sample, the second training image being an image of a second style;

generating a grafted transformation model by grafting the first transformation model with the second transformation model; and

generating the image generation model based on the reconstruction model and the grafted transformation model, wherein the image generation model is configured to transform an image to be transformed of the first style into a target image of the second style.

2. The method according to claim 1, wherein

the first transformation model comprises a first mapping network and a first synthesis network; and

training the first transformation model comprises:

acquiring a first training sample set, wherein the first training sample set comprises a plurality of first noise samples;

acquiring latent variables respectively corresponding to the plurality of first noise samples by inputting the plurality of first noise samples into the first mapping network;

acquiring first training images respectively corresponding to the plurality of first noise samples by inputting the latent variables respectively corresponding to the plurality of first noise samples into the first synthesis network; and

adjusting a weight parameter of the first transformation model based on the first training images respectively corresponding to the plurality of first noise samples.

3. The method according to claim 2, wherein

the first transformation model comprises a first discrimination network; and

adjusting the weight parameter of the first transformation model based on the first training images respectively corresponding to the plurality of first noise samples comprises:

acquiring first discrimination losses respectively corresponding to the plurality of first noise samples by inputting the plurality of first training images respectively corresponding to the plurality of first noise samples into the first discrimination network; and

adjusting the weight parameter of the first transformation model based on the first discrimination losses respectively corresponding to the plurality of first noise samples.

4. The method according to claim 1, wherein training the reconstruction model based on the first transformation model comprises:

acquiring a second training sample set, wherein the second training sample set comprises a plurality of original image samples;

generating latent variables respectively corresponding to the plurality of original image samples by inputting the plurality of original image samples into the reconstruction model;

generating reconstructed images respectively corresponding to the plurality of original image samples by inputting the latent variables respectively corresponding to the plurality of original image samples into the first transformation model, wherein the plurality of original image samples and the reconstructed images respectively corresponding to the plurality of original image samples are images of the first style;

determining losses of the reconstruction model respectively corresponding to the plurality of original image samples based on the plurality of original image samples and the reconstructed images respectively corresponding to the plurality of original image samples; and

adjusting a weight parameter of the reconstruction model based on the losses of the reconstruction model respectively corresponding to the plurality of original image samples.

5. The method according to claim 4, wherein

the first transformation model comprises a first discrimination network; and

determining the losses of the reconstruction model respectively corresponding to the plurality of original image samples based on the plurality of original image samples and the reconstructed images respectively corresponding to the plurality of original image samples comprises:

determining a first sub-loss based on an output result acquired by inputting each of the reconstructed images respectively corresponding to the plurality of original image samples into the first discrimination network, wherein the first sub-loss indicates a first characterization of the reconstructed image;

determining a second sub-loss based on an output result acquired by inputting each of the plurality of original image samples and each of the reconstructed images respectively corresponding to the plurality of original image samples into a perceptual network, wherein the second sub-loss indicates a first degree of conformity of the original image sample to the reconstructed image corresponding to the original image sample with respect to a target feature;

determining a third sub-loss based on an output result acquired by inputting each of the plurality of original image samples and each of the reconstructed images respectively corresponding to the plurality of original image samples into a regression function, wherein the third sub-loss indicates a second degree of conformity of the original image sample to the reconstructed image corresponding to the original image sample with respect to the target feature; and

determining the loss of the reconstruction model based on the first sub-loss, the second sub-loss, and the third sub-loss.

6. The method according to any one of claims 1 to 5, wherein during a training process, an initial weight parameter of the second transformation model is a weight parameter of the first transformation model.

7. The method according to any one of claims 1 to 5, wherein generating the grafted transformation model by grafting the first transformation model with the second transformation model comprises:

generating the grafted transformation model by grafting n layers of weight network in a plurality of layers of weight network of the first transformation model with m layers of weight network in a plurality of layers of weight network of the second transformation model; wherein layers of the n layers of weight network and the m layers of weight network are different, n is a positive integer, and m is a positive integer; or

generating the grafted transformation model by performing a summing or averaging or difference operation on weight parameters of the plurality of layers of weight network of the first transformation model and corresponding weight parameters of the plurality of layers of weight network of the second transformation model.

8. The method according to any one of claims 1 to 5, wherein generating the image generation model based on the reconstruction model and the grafted transformation model comprises:

acquiring a combined transformation model by combining the reconstruction model and the grafted transformation model;

acquiring a fourth training sample set, wherein the fourth training sample set comprises at least one original image sample and an image of the second style corresponding to the at least one original image sample; and

generating the image generation model by fine-tuning the combined transformation model using the fourth training sample set.

9. A method for generating images, comprising:

generating a latent variable corresponding to an image to be transformed by inputting the image to be transformed of a first style into a reconstruction model; and

generating, based on the latent variable corresponding to the image to be transformed, a target image corresponding to the image to be transformed using a grafted transformation model, wherein the target image is an image of a second style;

wherein the grafted transformation model is a model generated by grafting a first transformation model with a second transformation model, the first transformation model is configured to generate an image of the first style according to a first noise sample, and the second transformation model is configured to generate an image of the second style according to a second noise sample.

10. An apparatus for training an image generation model, comprising:

a model training module, configured to acquire a first transformation model by training, wherein the first transformation model is configured to generate a first training image according to a first noise sample, the first training image being an image of a first style, wherein

the model training module is further configured to acquire a reconstruction model by training based on the first transformation model, wherein the reconstruction model is configured to map an original image sample to a latent variable corresponding to the original image sample, and

the model training module is further configured to acquire a second transformation model by training, wherein the second transformation model is configured to generate a second training image according to a second noise sample, the second training image being an image of a second style; and

a model generation module, configured to generate a grafted transformation model by grafting the first transformation model with the second transformation model, wherein

the model generation module is further configured to generate the image generation model based on the reconstruction model and the grafted transformation model, wherein the image generation model is configured to transform an image to be transformed of the first style into a target image of the second style.

11. An apparatus for generating images, comprising:

a variable generation module, configured to generate a latent variable corresponding to an image to be transformed by inputting the image to be transformed of a first style into a reconstruction model; and

an image generation module, configured to generate, based on the latent variable corresponding to the image to be transformed, a target image corresponding to the image to be transformed using a grafted transformation model, wherein the target image is an image of a second style;

wherein the grafted transformation model is a model generated by grafting a first transformation model with a second transformation model, the first transformation model is configured to generate an image of the first style according to a first noise sample, and the second transformation model is configured to generate an image of the second style according to a second noise sample.

12. A computer device, comprising: a processor and a memory, the memory storing one or more computer programs therein, wherein the one or more computer programs, when loaded and run by the processor of the computer device, cause the computer device to perform the method for training the image generation model as defined in any one of claims 1 to 8, or perform the method for generating the images as defined in claim 9.

13. A computer-readable storage medium, storing one or more computer programs therein, wherein the one or more computer programs, when loaded and run by a processor, cause the processor to perform the method for training the image generation model as defined in any one of claims 1 to 8, or perform the method for generating the images as defined in claim 9.

14. A computer program product, when loaded and executed by a processor, causes the processor to perform the method for training the image generation model as defined in any one of claims 1 to 8, or perform the method for generating the images as defined in claim 9.