US 20220318946A1

(19) United States

(12) Patent Application Publication (10) Pub. No.: US 2022/0318946 A1

QIN et al. (43) Pub. Date: Oct. 6, 2022

(54) METHOD FOR IMAGE SHAPE TRANSFORMATION BASED ON GENERATIVE ADVERSARIAL NETWORK

(71) Applicant: **UNIVERSITY OF ELECTRONIC SCIENCE AND TECHNOLOGY OF CHINA**, Chengdu (CN)

(72) Inventors: **Zhen QIN**, Chengdu (CN); **Yi DING**, Chengdu (CN); **Tianming ZHUANG**, Chengdu (CN); **Fuhu DENG**, Chengdu (CN); **Zhiguang QIN**, Chengdu (CN)

(73) Assignee: **UNIVERSITY OF ELECTRONIC SCIENCE AND TECHNOLOGY OF CHINA**, Chengdu (CN)

(21) Appl. No.: **17/450,692**

(22) Filed: **Oct. 13, 2021**

(30) **Foreign Application Priority Data**

Mar. 31, 2021 (CN) 202110347463.5

Publication Classification

(51) **Int. Cl.**

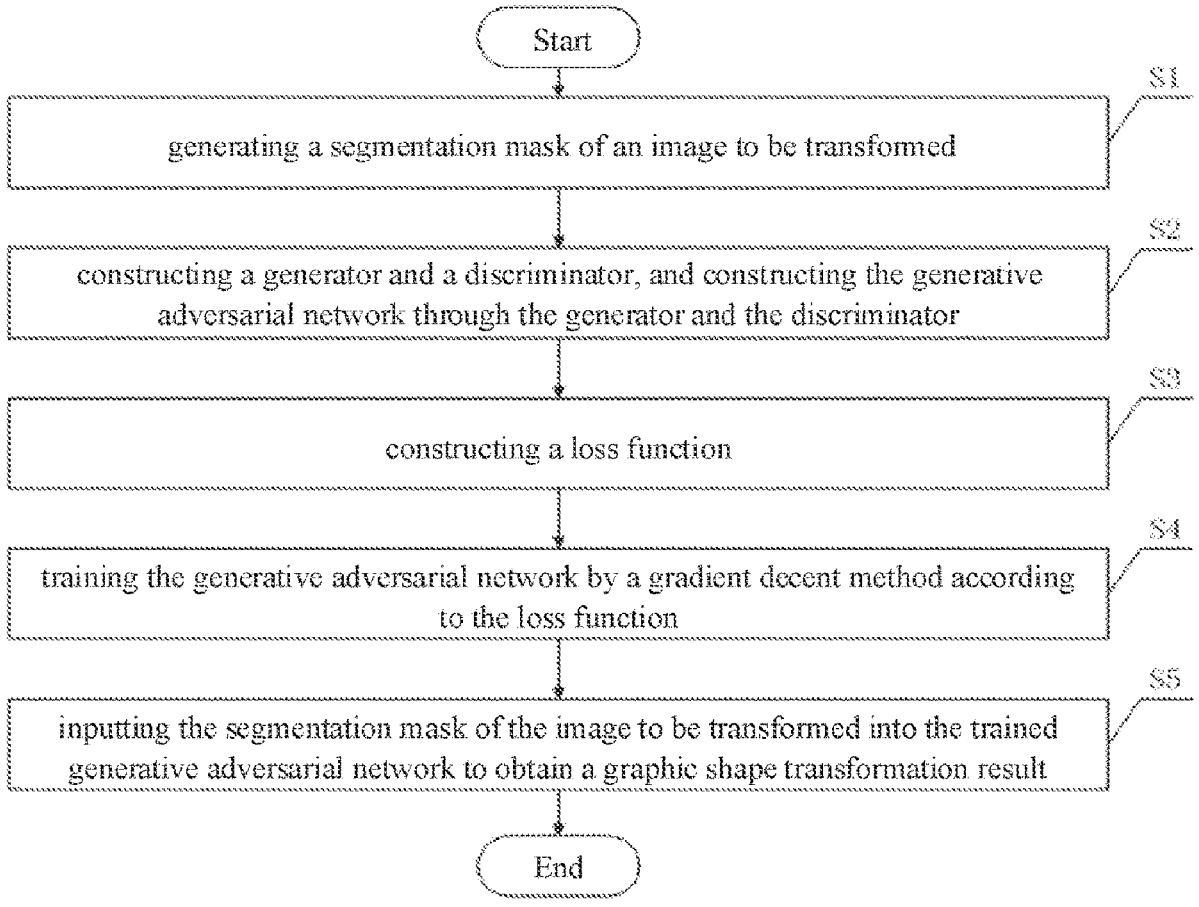
G06T 3/00	(2006.01)
G06N 3/04	(2006.01)
G06T 7/11	(2006.01)

(52) **U.S. Cl.**

CPC ***G06T 3/00*** (2013.01); ***G06N 3/0454*** (2013.01); ***G06T 7/11*** (2017.01)

(57) **ABSTRACT**

A method for image shape transformation based on a generative adversarial network includes the following steps: generating a segmentation mask of an image to be transformed; constructing a generator and a discriminator, and constructing the generative adversarial network through the generator and the discriminator; constructing a loss function, and training the generative adversarial network by a gradient descent method according to the loss function; and inputting the segmentation mask of the image to be transformed into the trained generative adversarial network to obtain an image shape transformation result.



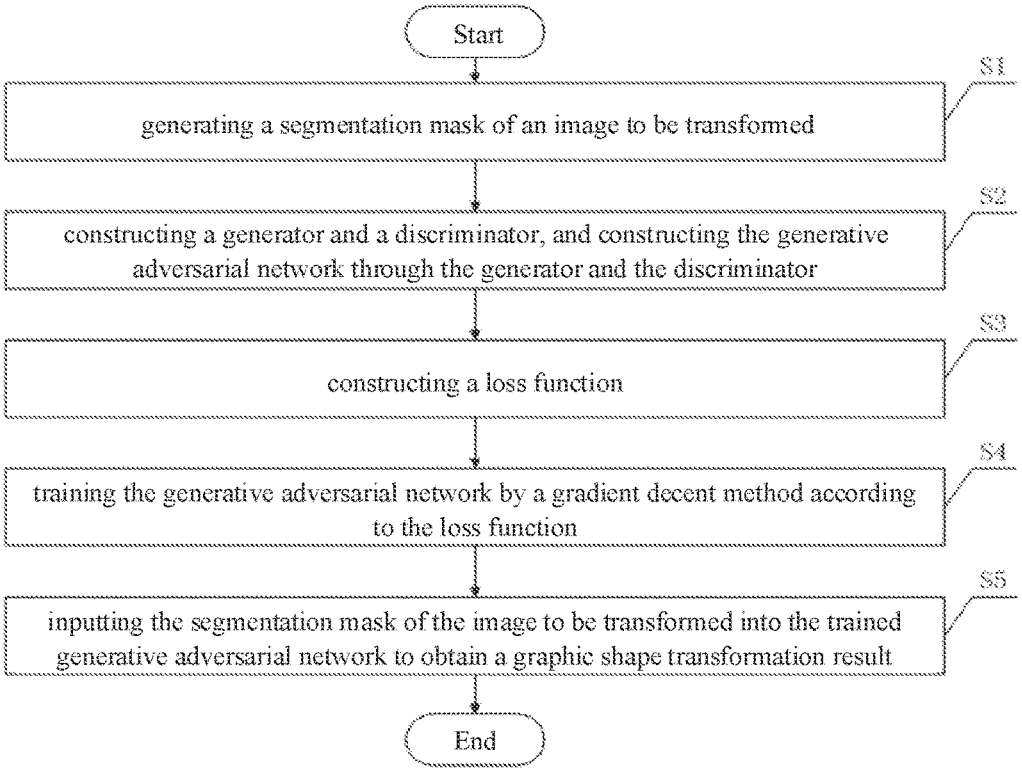

Start
S1
generating a segmentation mask of an image to be transformed
S2
constructing a generator and a discriminator, and constructing the generative adversarial network through the generator and the discriminator
S3
constructing a loss function
S4
training the generative adversarial network by a gradient decent method according to the loss function
S5
inputting the segmentation mask of the image to be transformed into the trained generative adversarial network to obtain a graphic shape transformation result
End

FIG. 1

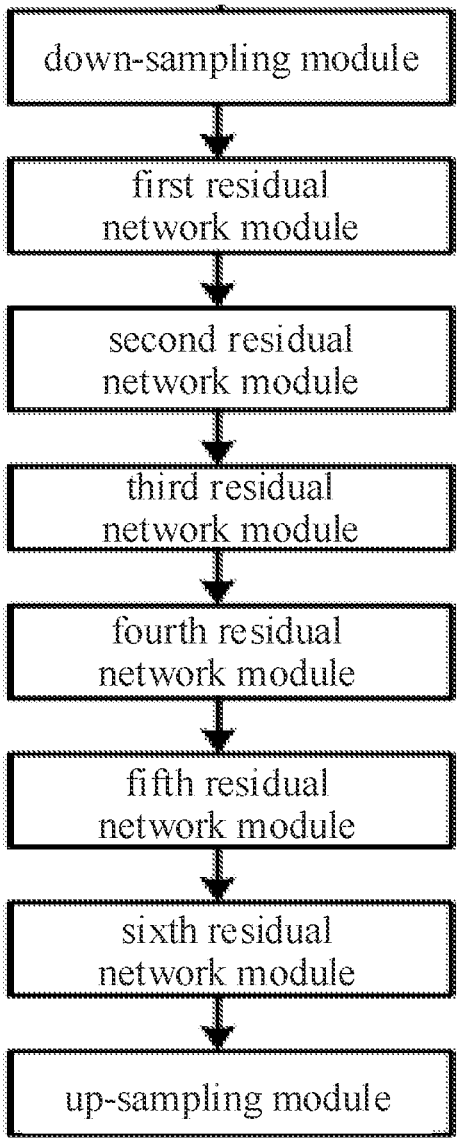

down-sampling module
first residual network module
second residual network module
third residual network module
fourth residual network module
fifth residual network module
sixth residual network module
up-sampling module

FIG. 2

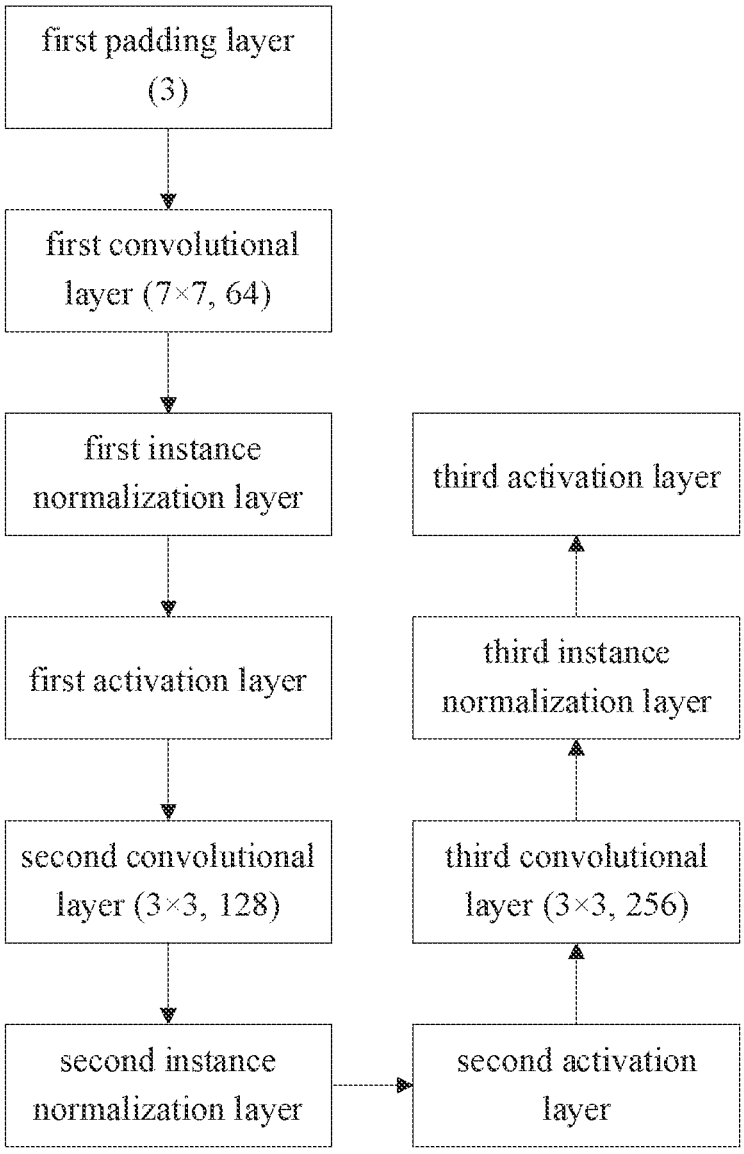

first padding layer (3)
first convolutional layer (7×7, 64)
first instance normalization layer
first activation layer
second convolutional layer (3×3, 128)
second instance normalization layer
second activation layer
third convolutional layer (3×3, 256)
third instance normalization layer
third activation layer

FIG. 3

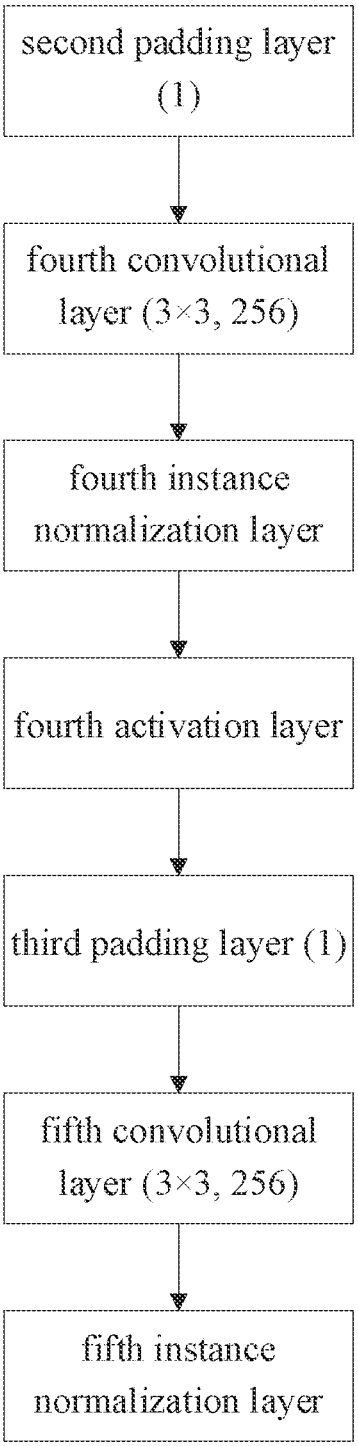

second padding layer (1)
fourth convolutional layer (3×3, 256)
fourth instance normalization layer
fourth activation layer
third padding layer (1)
fifth convolutional layer (3×3, 256)
fifth instance normalization layer

FIG. 4

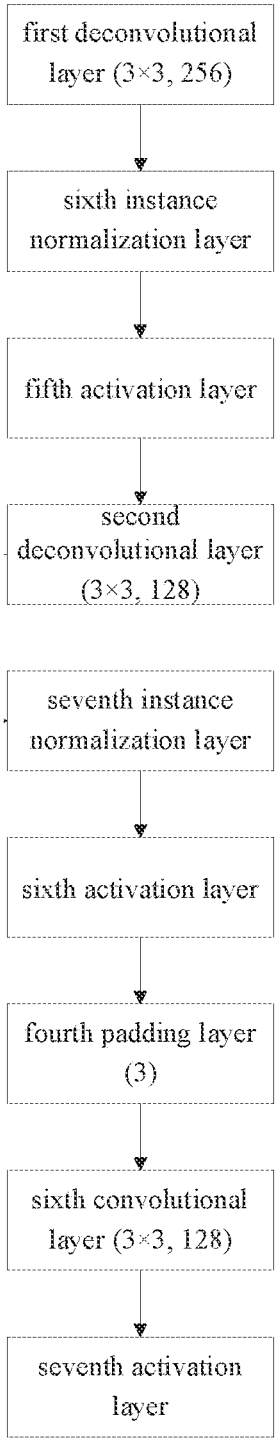

first deconvolutional layer (3×3, 256)
sixth instance normalization layer
fifth activation layer
second deconvolutional layer (3×3, 128)
seventh instance normalization layer
sixth activation layer
fourth padding layer (3)
sixth convolutional layer (3×3, 128)
seventh activation layer

FIG. 5

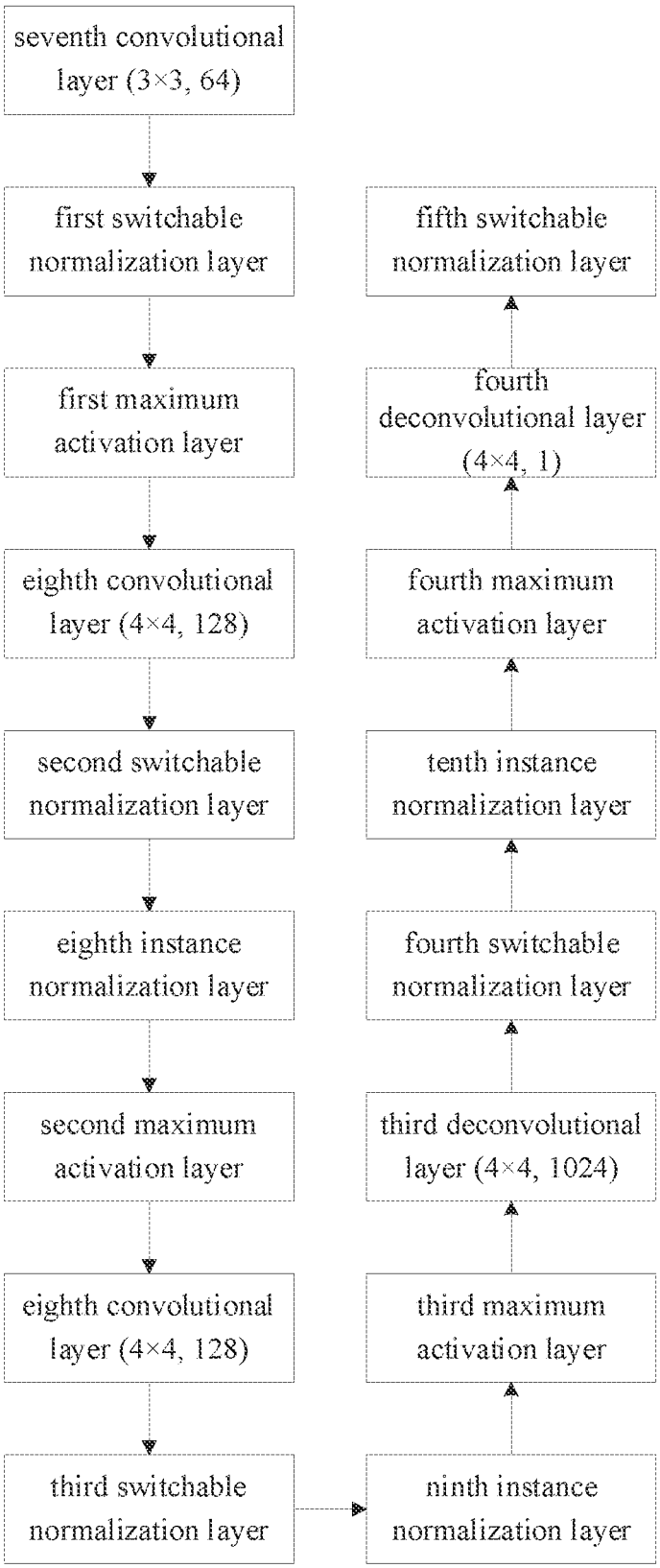

seventh convolutional layer (3×3, 64)
first switchable normalization layer
first maximum activation layer
eighth convolutional layer (4×4, 128)
second switchable normalization layer
eighth instance normalization layer
second maximum activation layer
eighth convolutional layer (4×4, 128)
third switchable normalization layer
ninth instance normalization layer
third maximum activation layer
third deconvolutional layer (4×4, 1024)
fourth switchable normalization layer
tenth instance normalization layer
fourth maximum activation layer
fourth deconvolutional layer (4×4, 1)
fifth switchable normalization layer

FIG. 6

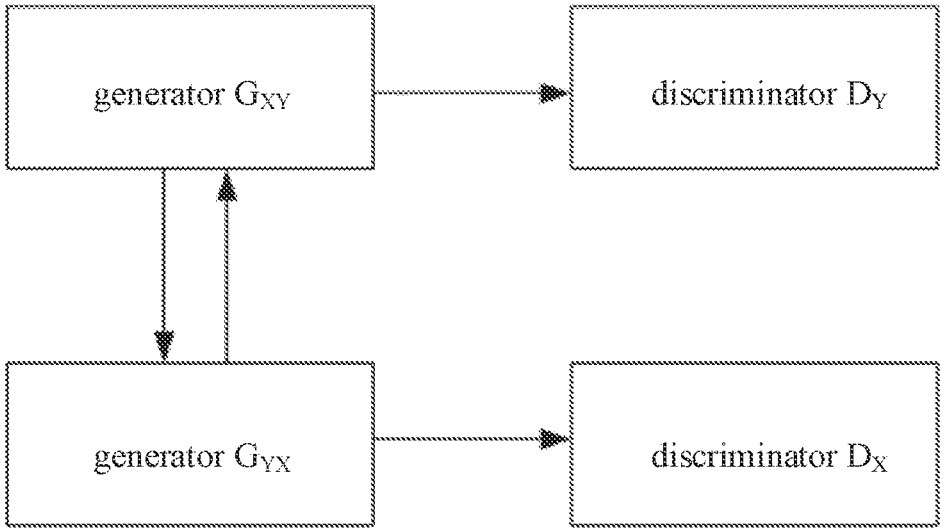

generator G_{XY}
discriminator D_Y
generator G_{YX}
discriminator D_X

FIG. 7

METHOD FOR IMAGE SHAPE TRANSFORMATION BASED ON GENERATIVE ADVERSARIAL NETWORK

CROSS REFERENCE TO THE RELATED APPLICATIONS

[0001] This application is based upon and claims priority to Chinese Patent Application No. 202110347463.5, filed on Mar. 31, 2021, the entire contents of which are incorporated herein by reference.

TECHNICAL FIELD

[0002] The present invention relates to the field of image translation, and specifically relates to a method for image shape transformation based on a generative adversarial network.

BACKGROUND

[0003] Adversarial neural network models have achieved remarkable success in many applications, such as image inpainting, semantic segmentation, image caption, video generation, and style conversion. Image translation is one of the main research topics in computer vision. In recent years, the development of generative adversarial networks has led to significant advancements in the field of image translation. When paired training data from two different domains are given, the model can be trained in a supervised manner by using a conditional generative adversarial network. When there is a lack of paired data sets, image-to-image translation can be realized by using an unsupervised cycle generative adversarial network and self-consistency loss.

[0004] However, technologies in this field mostly focus on the conversion task between the styles of two similar shapes, such as season transfer, selfie-to-amine, and style conversion, but lack satisfactory performance in the conversion task between objects with different shapes.

SUMMARY

[0005] In view of the above-mentioned shortcomings in the prior art, the present invention provides a method for image shape transformation based on a generative adversarial network to solve the problems identified in the prior art.

[0006] In order to achieve the above-mentioned objective, the present invention adopts the following technical solution, A method for image shape transformation based on a generative adversarial network includes the following steps:

[0007] S1: generating a segmentation mask of an image to be transformed;

[0008] S2: constructing a generator and a discriminator, and constructing the generative adversarial network through the generator and the discriminator;

[0009] S3: constructing a loss function;

[0010] S4: training the generative adversarial network by a gradient descent method. according to the loss function; and

[0011] S5: inputting the segmentation mask of the image to be transformed into the trained generative adversarial network to obtain an image shape transformation result.

[0012] Further, step S1 specifically includes: generating, for the image to be transformed, a segmentation mask m, in a source domain X and a segmentation mask m_y in a target domain

[0013] Further, in step S2, the generator includes a down-sampling module, a first residual network module, a second residual network module, a third residual network module, a fourth residual network module, a fifth residual network module, a sixth residual network module and an up-sampling module that are connected in sequence.

[0014] The down-sampling module includes a first padding layer, a first convolutional layer, a first instance normalization layer, a first activation layer, a second convolutional layer, a second instance normalization layer, a second activation layer, a third convolutional layer, a third instance normalization layer and a third activation layer that are connected in sequence.

[0015] Each of the residual network modules includes a second padding layer, a fourth convolutional layer, a fourth instance normalization layer, a fourth activation layer, a third. padding layer, a fifth convolutional layer and a fifth instance normalization layer that are connected in sequence.

[0016] The up-sampling module includes a first deconvolutional layer, a sixth instance normalization layer, a fifth activation layer, a second deconvolutional layer, a seventh instance normalization layer, a sixth activation laver, a fourth padding layer, a sixth convolutional layer and a seventh activation layer that are connected in sequence.

[0017] Further, in step S2, the discriminator includes a seventh convolutional layer, a first switchable normalization layer, a first maximum activation layer, an eighth convolutional layer, a second switchable normalization layer, an eighth instance normalization layer, a second maximum activation layer, an eighth convolutional layer, a third switchable normalization layer, a ninth instance normalization layer, a third maximum activation layer, a third deconvolutional layer, a fourth switchable normalization layer, a tenth instance normalization layer, a fourth maximum activation layer, a fourth deconvolutional layer and a fifth switchable normalization layer that are connected in sequence.

[0018] Further, in step S2, a method for constructing the generative adversarial network through the generator and the discriminator specifically includes:

[0019] S2.1: constructing a generator G_{XY} for converting a given image in the source domain X to an image in the target domain Y, and constructing a generator G_{YX} for converting a given image in the target domain Y to an image in the source domain X;

[0020] S2.2: constructing a discriminator D_Y for predicting whether the image is an image in the target domain, and constructing a discriminator D_X for predicting whether the image is an image in the source domain; and

[0021] S2.3: connecting the generator G_{XY} to the generator G_{YX}, connecting the generator G_{XY} to the discriminator D_Y, and connecting the generator G_{YX} to the discriminator D_X to construct the generative adversarial network.

[0022] Further, in step S2.3, a one-cycle generation process of the generative adversarial network includes a source domain cycle generation process and a target domain cycle generation process.

[0023] The source domain cycle generation process specifically includes:

[0024] S2.3.1.1: inputting a source domain image x and its segmentation mask m_x into the generator G_{XY} to obtain a first target domain generated image y' and its segmentation mask m'_y , which are denoted as $G_{XY}(x,m_x)$;

[0025] S2.3.1.2: inputting the target domain generated image y' and its segmentation mask into the generator G_{YX} to obtain a first source domain generated image and its segmentation mask, which are denoted as $G_{YX}(G_{XY}(x,m_x))$

[0026] S2.3.1.3: inputting the first target domain generated image y' and its mask m'_y as well as a target domain image y and its segmentation mask m_y into the discriminator D_Y for discrimination, and inputting $G_{YX}(G_{XY}(x,m_x))$ denoting the first source domain generated image and its segmentation mask into the discriminator D_X for discrimination, so as to complete the source domain cycle generation process.

[0027] The target domain cycle generation process specifically includes:

[0028] S2.3.2.1: inputting the target domain image y and its segmentation mask m_y into the generator G_{YX} to obtain a second source domain generated image X' and its segmentation mask m'_x, whiCh are denoted as $G_{YX}(y,m_y)$;

[0029] S2.3.2.2: inputting the second source domain generated image x' and its segmentation mask m'_x into the generator G_{XY} to obtain a second target domain generated image and its segmentation mask, which are denoted as $G_{XY}(G_{YX}(y,m_y))$;

[0030] S2.3.2.3: inputting the second source domain generated image x' and its segmentation mask m'_x, as well as the source domain nage x and its segmentation mask m_x into the discriminator D_X for discrimination, and inputting the second target domain generated image and its segmentation mask $G_{XY}(G_{YX}(y,m_y))$ into the discriminator D_Y for discrimination, so as to complete the target domain cycle generation process.

[0031] Further, in step S3, the loss function $\mathcal{L}_{total}$ is specifically expressed as:

$$\mathcal{L}_{total}=\lambda_{adv}\mathcal{L}_{adv}+\lambda_{cyc}\mathcal{L}_{cyc}+\lambda_{idt}\mathcal{L}_{idt}+\lambda_{ctx}\mathcal{L}_{ctx}+\lambda_{fs}\mathcal{L}_{fs};$$

[0032] wherein $\mathcal{L}_{adv}$ represents an adversarial loss function, $\mathcal{L}_{cyc}$ represents a cycle consistency loss function, $\mathcal{L}_{idt}$ represents an identity loss function, $\mathcal{L}_{ctx}$ represents a context-preserving loss function, $\mathcal{L}_{is}$ represents a feature similarity loss function, and λ_{adv}, λ_{cyc}, λ_{idt}, λ_{ctx} and λ_{fs} represent weights of $\mathcal{L}_{adv}$, $\mathcal{L}_{cyc}$, $\mathcal{L}_{idt}$, $\mathcal{L}_{ctx}$ and $\mathcal{L}_{fs}$ in the loss function $\mathcal{L}_{total}$, respectively.

[0033] Further, the adversarial loss function $\mathcal{L}_{adv}$ is specifically expressed as:

$$\mathcal{L}_{adv}=(D_X(x,m_x)-1)^2+D_X(G_{YX}(y,m_y))^2+(D_Y{}^*y,m_y)-1)^2+D_Y(G_{XY}(x,m_x))^2;$$

wherein, $D_X(x, m_x)$ represents a discriminant output of the discriminator D_x for the source domain image x and its segmentation mask m_x, $D_X(G_{YX}(y,m_y))$ represents a discriminant output of the discriminator D_X for $G_{YX}(y,m_y)$ denoting the source domain generated image and its segmentation mask, $G_{YX}(y,m_y)$ represents the source domain generated image and its mask generated by the target domain image y and its segmentation mask m_y through the generator G_{YX}, $D_Y(y,m_y)$ represents a discriminant output of the discriminator D_Y for the target domain image y and its segmentation mask m_y, $D_Y(G_{XY}(x,m_x))$ represents a discriminant output of the discriminator D_Y for $G_{XY}(x,m_x)$ denoting the target domain generated image and its segmentation mask, and $G_{XY}(x,m_x)$ represents the target domain generated image and its segmentation mask generated by the source domain image x and its segmentation mask m_x through the generator G_{XY}.

[0034] The cycle consistency loss function $\mathcal{L}_{cyc}$ is specifically expressed as:

$$\mathcal{L}_{cyc}=\|G_{YX}(G_{XY}(x,m_x))-(x,m_x)\|_1+\|G_{XY}(G_{XY}(y,m_y))-(y,m_y\|_1;$$

[0035] wherein $G_{YX}(G_{XY}(x,m_x))$ represents the source domain generated image and its segmentation mask generated by $G_{XY}(x, m_x)$ through the generator G_{YX}, $G_{XY}(G_{YX}(y, m_y))$ represents the target domain generated image and its segmentation mask generated by $G_{YX}(y,m_y)$ through the generator G_{XY} and represents $\|*\|_1$ 1-norm.

[0036] The identity loss function $\mathcal{L}_{dt}$ is specifically expressed as:

$$\mathcal{L}_{idt}=\|G_{XY}(y,m_y)-(y,m_y)\|_1+\|G_{YX}(x,m_x)-(x,m_x)\|_1;$$

[0037] wherein $G_{XY}(y,m_y)$ represents the segmentation mask of the first target domain generated image y obtained after the source domain image x and its segmentation mask m_x are input into the generator G_{XY}, and $G_{YX}(x,m_x)$ represents the segmentation mask of the second source domain generated image x' obtained after the target domain image y and its segmentation mask m_y are input into the generator G_{YX}.

[0038] The context-preserving loss function $\mathcal{L}_{ctx}$ is specifically expressed as:

$$\mathcal{L}_{ctx}=\|\omega(m_x,m'_y)e(x-y')\|_1+\|\omega(m_y,m'_x)e(y-x')\|_1;$$

[0039] wherein $\omega(m_x, m'_y)$ represents a minimum value obtained by subtracting one from elements in binary object masks through segmentation masks m_x and m'_y, and $\omega(m_y, m'_x)$ represents a minimum value obtained by subtracting one from elements in binary object masks through segmentation masks m_y and m'_x; y' represents the target domain generated image generated by the source domain image x through the generator G_{XY}, and x' represents the source domain generated image generated by the target domain image y and its segmentation mask m_y through the generator G_{YX}.

[0040] The feature similarity loss function $\mathcal{L}_{fs}$ is specifically expressed as:

$$\begin{aligned}\mathcal{L}_{fs} &= FS(y,y')\\ &= \frac{1}{N}\sum_j \max_i FS_{ij}\\ &= \frac{1}{N}\sum_j \max_i\left(W_{ij}\Big/\sum_k \omega_{ik}\right)\\ &= \frac{1}{N}\sum_j \max_i\left[\exp\left(\frac{1-\overline{d_{ij}}}{h}\right)\Big/\sum_k \omega_{ik}\right];\end{aligned}$$

[0041] wherein FS(y,y') represents the similarity between the image y and the image y', and

$$\max_i FS_{ij}$$

represents a feature i of the image y' that is most similar to a feature j of the image y'; N represents a total number of features of the image y' which is the same as a total number of features of the image y'; h represents a bandwidth parameter,

$$\exp\left(\frac{1-\overline{d_{ij}}}{h}\right)$$

represents an exponential operation from distance d_{ij} to W_{ij}, d_{ij} represents a normalized similarity distance, W_{ij} represents a similarity, $W_{ij}/\Sigma_k w_{ik}$ represents a normalized similarity, and w_{ik} represents a similarity value of a k^{th} W_{ij}.

[0042] The advantages of the present invention are as follows.

[0043] (1) The present invention provides a method for image shape transformation based on a generative adversarial network, which realizes the transformation between object images with different shapes.

[0044] (2) The present invention uses a cycle generator and a discriminator to learn cross-domain mappings, generates an image with a closer basic distribution based on a target instance, and can effectively learn complex segmentation guidance attributes related to shapes and positions.

[0045] (3) The present invention proposes a feature similarity loss function to clearly establish a similarity comparison between a source image and a target image.

[0046] (4) With low complexity but high image conversion efficiency, the present invention can efficiently process specific images in pictures to perform image transformations with large shape differences, and thus can be used in animation production, poster design and other fields to enhance the reality of image transformation, while reducing labor costs and workload.

BRIEF DESCRIPTION OF THE DRAWINGS

[0047] FIG. **1** is a flowchart of a method for image shape transformation based on a generative adversarial network proposed by the present invention.

[0048] FIG. **2** is a schematic diagram of a generator in the present invention.

[0049] FIG. **3** is a schematic diagram of a down-sampling module in the present invention.

[0050] FIG. **4** is a schematic diagram of a residual network module in the present invention.

[0051] FIG. **5** is a schematic diagram of an up-sampling module in the present invention.

[0052] FIG. **6** is a schematic diagram of a discriminator in the present invention.

[0053] FIG. **7** is a schematic diagram of a generative adversarial network in the present invention.

DETAILED DESCRIPTION OF THE EMBODIMENTS

[0054] The following describes the specific embodiments of the present invention to help those skilled in the art understand the present invention, but it should be clear that the present invention is not limited to the scope of the specific embodiments, for those ordinarily skilled in the art, as long as various changes are within the spirit and scope of the present invention defined and determined by the appended claims, these changes are obvious, and all inventions and creations that utilize the concept of the present invention shall fall within the scope of the present invention.

[0055] The embodiments of the present invention will be described in detail below with reference to the drawings.

[0056] As shown in FIG. **1**, a method for image shape transformation based on a generative adversarial network includes the following steps:

[0057] S**1**: a segmentation mask of an image to be transformed is generated;

[0058] S**2**: a generator and a discriminator are constructed, and the generative adversarial network is constructed through the generator and the discriminator;

[0059] S**3**: a loss function is constructed;

[0060] S**4**: the generative adversarial network is trained by a gradient descent method according to the loss function; and

[0061] S**5**: the segmentation mask of the image to be transformed is input into the trained generative adversarial network to obtain an image shape transformation result.

[0062] In the present embodiment, the present invention can be applied to game design, animation design, graphic design, medical imaging, and style transfer. In step S**1**, the image to be transformed can be a medical image to be transformed, an animated image with a shape to be transformed during the animation design, a game character or architectural image with a shape to be transformed during the game design, or an image to be transformed during the graphic design.

[0063] Step S**1** specifically includes: generating, for the image to be transformed, a segmentation mask m_x in a source domain X and a segmentation mask m_y in a target domain Y.

[0064] As shown in FIG. **2**, in step S**2**, the generator includes the down-sampling module, the first residual network module, the second residual network module, the third residual network module, the fourth residual network module, the fifth residual network module, the sixth residual network module and the up-sampling module connected in sequence.

[0065] As shown in FIG. **3**, the down-sampling module includes the first padding layer ReflectionPad2d(3), the first convolutional layer, the first instance normalization layer, the first activation layer ReLU, the second convolutional layer, the second instance normalization layer, the second activation layer ReLU, the third convolutional layer, the third instance normalization layer and the third activation layer ReLU connected in sequence.

[0066] As shown in FIG. **4**, each of the residual network modules includes the second padding layer ReflectionPad2d (f), the fourth convolutional layer, the fourth instance normalization layer, the fourth activation layer ReLU, the third padding layer ReflectionPad2d(1), the fifth convolutional layer and the fifth instance normalization layer connected in sequence.

[0067] As shown in FIG. **5**, the up-sampling module includes the first deconvolutional layer, the sixth instance normalization layer, the fifth activation layer ReLU, the second deconvolutional layer, the seventh instance normalization layer, the sixth activation layer Rail, the fourth padding layer ReflectionPad2d(3), the sixth convolutional layer and the seventh activation layer ReLU connected in sequence.

[0068] As shown in FIG. **6**, in step S**2**, the discriminator includes the seventh convolutional layer, the first switchable normalization layer, the first maximum activation layer LeakyReLU, the eighth convolutional layer, the second switchable normalization layer, the eighth instance normalization layer, the second maximum activation layer

LeakyReLU, the eighth convolutional layer, the third switchable normalization layer, the ninth instance normalization layer, the third maximum activation layer LeakyReLU, the third deconvolutional layer, the fourth switchable normalization layer, the tenth instance normalization layer, the fourth maximum activation layer LeakyReLU, the fourth deconvolutional layer and the fifth switchable normalization layer connected in sequence.

[0069] In step S**2**, a method for constructing the generative adversarial network through the generator and the discriminator specifically includes:

[0070] S**2**.**1**: a generator G_{XY} for converting a given image in the source domain X to an image in the target domain Y is constructed, and a generator G_{YX} for converting a given image in the target domain Y to an image in the source domain X is constructed;

[0071] S**2**.**2**: a discriminator D_Y for predicting whether the image is an image in the target domain is constructed, and a discriminator D_X for predicting whether the image is an image in the source domain is constructed; and

[0072] S**2**.**3**: as shown in FIG. **7**, the generator G_{XY} is connected to the generator G_{YX}, the generator G_{XY} is connected to the discriminator D_Y, and the generator G_{YX} is connected. to the discriminator D_X, so as to construct the generative adversarial network.

[0073] In the present embodiment, the generator includes three modules: a down-sampling module, a residual network module and an up-sampling module.

[0074] The down-sampling module converts the input feature vector (1, 4, 256, 256) into a feature vector (1, 256, 64, 64) through a four-layer convolution operation. The residual network module includes six blocks, where the input and output feature vectors have the same dimension. The up-sampling module converts the input feature vector (1, 512, 64. 64) into a feature vector (1, 3, 256, 256) through a five-layer convolution operation.

[0075] The discriminator includes two modules: a down-sampling module and a classifier module. The down-sampling module converts the input feature vector (1, 3, 256, 256) into a feature vector (1, 256, 32, 32) through a three-layer convolution operation.

[0076] In step S**2**.**3**, a one-cycle generation process of the generative adversarial network includes a source domain cycle generation process and a target domain cycle generation process.

[0077] The source domain cycle generation process specifically includes:

[0078] S**2**.**3**.**1**.**1**: a source domain image x and its segmentation mask m_x are input into the generator G_{XY} to obtain a first target domain generated image y' and its segmentation mask m'_y , which are denoted as $G_{XY}(x,m_x)$;

[0079] S**2**.**3**.**1**.**2**: the target domain generated image y' and its segmentation mask m'_y are input into the generator G_{XY} to obtain a first source domain generated image and its segmentation mask, which are denoted as $G_{YX}(G_{XY}, (x,m_x))$

[0080] S**2**.**3**.**1**,**3**: the first target domain generated image y and its mask m_y as well as a target domain image y and its segmentation mask m_y are input into the discriminator D_Y for discrimination, and $G_{YX}(G_{XY}(x,m_x))$ denoting the first source domain generated image and its segmentation mask is input into the discriminator D_X for discrimination, so as to complete the source domain cycle generation process.

[0081] The target domain cycle generation process specifically includes:

[0082] S**2**.**3**.**2**.**1**: the target domain image y and its segmentation mask m_y are input into the generator G_{YX} to obtain a second source domain generated image x' and its segmentation mask m'_x, which are denoted as $G_{YX}(y,m_y)$;

[0083] S**2**.**3**.**2**.**2**: the second source domain generated image x' and its segmentation mask m'_x are input into the generator G_{XY} to obtain a second target domain generated image and its segmentation mask, which are denoted as $G_{XY}(G_{YX}*y,m_y))$;

[0084] S**2**.**3**.**2**.**3**: the second source domain generated image x' and its segmentation mask m'_x as well as the source domain image x and its segmentation mask m_x are input into the discriminator D_X for discrimination, and $G_{XY}(G_{YX}(y,m_y))$ denoting the second target domain generated image and its segmentation mask is input into the discriminator D_Y for discrimination, so as to complete the target domain cycle generation process.

[0085] In step S**3**, the loss function $\mathcal{L}_{total}$ is specifically expressed as:

$$\mathcal{L}_{total}=\lambda_{adv}\mathcal{L}_{adv}+\lambda_{cyc}\mathcal{L}_{cyc}+\lambda_{idt}\mathcal{L}_{idt}+\lambda_{ctx}\mathcal{L}_{ctx}+\lambda_{fs}\mathcal{L}_{fs};$$

[0086] wherein $\mathcal{L}_{adv}$, represents an adversarial loss function, represents a cycle consistency loss function, $\mathcal{L}_{idt}$ represents an identity loss function, $\mathcal{L}_{ctx}$ represents a context-preserving loss function, $\mathcal{L}_{fs}$ represents a feature similarity loss function, and λ_{adv}, λ_{cyc}, λ_{idt}, λ_{ctx} and λ_{fs} represent weights of $\mathcal{L}_{adv}$, $\mathcal{L}_{cyc}$, $\mathcal{L}_{idt}$, $\mathcal{L}_{ctx}$ and $\mathcal{L}_{fs}$ in the loss function $\mathcal{L}_{total}$, respectively.

[0087] The adversarial loss function $\mathcal{L}_{adv}$ is specifically expressed as:

$$\mathcal{L}_{adv}=(D_X(x,m_x)-1)^2+D_X(G_{YX}(y,m_y))^2+(D_y(y,m_y)-1)^2+D_Y(G_{XY}(x,m_x))^2;$$

[0088] wherein, $D_X(x,m_x)$ represents a discriminant output of the discriminator D_X for the source domain image x and its segmentation mask m_x, $D_X(G_{YX}(y,m_y))$ represents a discriminant output of the discriminator D_X for $G_{YX}(y,m_y)$ denoting the source domain generated image and its segmentation mask, $G_{YX}(y,m_y)$ represents the source domain generated image and its mask generated by the target domain image y and its segmentation mask m through the generator G_{YX}, $D_Y(y,m_y)$ represents a discriminant output of the discriminator D_Y for the target domain image y and its segmentation mask m_y, $D_Y(G_{XY}(x,m_x))$ represents a discriminant output of the discriminator D_Y for the target domain generated image and its segmentation mask $G_{XY}(x,m_x)$ $G_{XY}(x,m_x)$ represents the target domain generated image and its segmentation mask generated by the source domain image x and its segmentation mask m_x through the generator G_{XY}.

[0089] The cycle consistency loss function $\mathcal{L}_{cyc}$ is specifically expressed as:

$$\mathcal{L}_{cyc}=\|G_{XY}(G_{XY}(x,m_x))-(x,m_x)\|_1+\|G_{XY}(G_{YX}(G_{YX}(y,m_y))-(y,m_y)\|_1;$$

[0090] wherein G_{YX} $(G_{XY}(x,m_x))$ represents the source domain generated image and its segmentation mask generated by $G_{XY}(x,m_x)$ through the generator G_{YX}, $G_{XY}(G_{YX}(y,m_y))$ represents the target domain generated image and its segmentation mask generated by $G_{YX}(y,m_y)$ through the generator G_{XY}, and $\|*\|_1$ represents 1-norm.

[0091] The identity loss function $\mathcal{L}_{idt}$ is specifically expressed as:

$$\mathcal{L}_{idt}=\|G_{XY}(y,m_y)-(y,m_y\|_1+\|G_{YX}(x,m_x)-(x,m_x)\|_1;$$

[0092] wherein $G_{XY}(y,m_y)$ represents the segmentation mask of the first target domain generated image y obtained after the source domain image x and its segmentation mask m_x are input to the generator G_{XY}, and $G_{YX}(x,m_x)$ represents the segmentation mask of the second source domain generated image x' obtained after the target domain image y and its segmentation mask m_y are input into the generator G_{YX}.

[0093] The context-preserving loss function $\mathcal{L}_{ctx}$ is specifically expressed as:

$$\mathcal{L}_{stx}=\|\omega(m_x,m'_y)e(x-y')\|_1+\|\omega(m_y,m'_x)e(y-x')\|_1;$$

[0094] wherein $\omega(m_x,m'_y)$ represents a minimum value obtained by subtracting one from elements in binary object masks through segmentation masks m_x and m'_y, and $\omega(m_y,m'_x)$ represents a minimum value obtained by subtracting one from elements in binary object masks through segmentation masks m_y and m'_x; y' represents the target domain generated image generated by the source domain image x through the generator G_{XY}, and x' represents the source domain generated image generated by the target domain image y and its segmentation mask m_y through the generator G_{YX}.

[0095] The feature similarity loss function $\mathcal{L}_s$ is specifically expressed as:

$$\begin{aligned}\mathcal{L}_{fs} &= FS(y,y')\\ &= \frac{1}{N}\sum_j \max_i FS_{ij}\\ &= \frac{1}{N}\sum_j \max_i\left(W_{ij}\Big/\sum_k \omega_{ik}\right)\\ &= \frac{1}{N}\sum_j \max_i\left[\exp\left(\frac{1-\overline{d_{ij}}}{h}\right)\Big/\sum_k \omega_{ik}\right];\end{aligned}$$

[0096] wherein FS(y,y') represents the similarity between the image y and the image y', and

$$\max_i FS_{ij}$$

represents a feature i of the image y that is most similar to a feature j of the image y'; N represents the total number of features of the image y , which is the same as the total number of features of the image y'; h represents a bandwidth parameter,

$$\exp\left(\frac{1-\overline{d_{ij}}}{h}\right)$$

represents an exponential operation from distance d_{ij} to W_{ij}, d_{ij} represents a normalized similarity distance, W_{ij} represents a similarity, $W_{ij}/\Sigma_k w_{ik}$ represents normalized similarity, and w_{ik} represents a similarity value of the k^{th} W_{ij}.

[0097] The similarity FS(y,y') between the two images is calculated by using these high-level features. Specifically, in a forward process, each layer generates a feature map, a real image y' obtained from real training data includes the feature y_i, while a composite image y' includes the feature y'_j. The content attribute and style attribute of the feature y'_j are consistent with those of a real domain data set. Assuming that the two images have the same number N of features, where N=|R|=|F|. The most similar feature y_i of each y'_j is found, that is, $\max FS_{ij}$. Then, all similarity values of y'_j are added to calculate a context similarity value between the two images. Finally, the context similarity value is divided by N to obtain an average similarity FS(y,y').

[0098] The advantages of the present invention are as follows.

[0099] (1) The present invention provides a method for image shape transformation based on a generative adversarial network, which realizes the transformation between objet images with different shapes.

[0100] (2) The present invention uses a cycle generator and a discriminator to learn cross-domain mappings, generates an image with a closer basic distribution based on a target instance, and can effectively learn complex segmentation guidance attributes related to shapes and positions.

[0101] (3) The present invention proposes a feature similarity loss function to clearly establish a similarity comparison between a source image and a target image.

[0102] (4) With low complexity but high image conversion efficiency, the present invention can efficiently process specific images in pictures to perform image transformations with large shape differences, and thus can be used in animation production, poster design and other fields to enhance the reality of image transformation, while reducing labor costs and workload.

What is claimed is:

1. A method for an image shape transformation based on a generative adversarial network, comprising the following steps:

S**1**: generating a segmentation mask of an image to be transformed;

S**2**: constructing a generator and a discriminator, and constructing the generative adversarial network through the generator and the discriminator,

S**3**: constructing a loss function;

S**4**: training the generative adversarial network by a gradient descent method according to the loss function to obtain a trained generative adversarial network; and

S**5**: inputting the segmentation mask of the image to be transformed into the trained generative adversarial network to obtain an image shape transformation result.

2. The method of claim **1**, wherein

step S**1** specifically comprises:

generating, for the image to be transformed, a segmentation mask m_x in a source domain X and a segmentation mask m_y in a target domain Y.

3. The method of claim **1**, wherein

in step S**2**, the generator comprises a down-sampling module, a first residual network module, a second residual network module, a third residual network module, a fourth residual network module, a fifth residual network module, a sixth residual network module and an up-sampling module connected in sequence;

the down-sampling module comprises a first padding layer, a first convolutional layer, a first instance normalization layer, a first activation layer, a second convolutional layer, a second instance normalization layer, a second activation layer, a third convolutional layer, a third instance normalization layer and a third activation layer connected in sequence;

each of the first residual network module, the second residual network module, the third residual network module, the fourth residual network module, the fifth residual network module and the sixth residual network module comprises a second padding layer, a fourth convolutional layer, a fourth instance normalization layer, a tburth activation layer, a third padding layer, a fifth convolutional layer and a fifth instance normalization layer connected in sequence; and

the up-sampling module comprises a first deconvolutional layer, a sixth instance normalization layer, a fifth activation layer, a second deconvolutional layer, a seventh instance normalization layer, a sixth activation layer, a fourth padding layer, a sixth convolutional layer and a seventh activation layer connected in sequence.

4. The method of claim I, wherein

in step S**2**, the discriminator comprises a seventh convolutional layer, a first switchable normalization layer, a first maximum activation layer, a top eighth convolutional layer, a second switchable normalization layer, an eighth instance normalization layer, a second maximum activation layer, a bottom eighth convolutional layer, a third switchable normalization layer, a ninth instance normalization layer, a third maximum activation layer, a third deconvolutional layer, a fourth switchable normalization layer, a tenth instance normalization layer, a fourth maximum activation layer, a fourth deconvolutional layer and a fifth switchable normalization layer connected in sequence.

5. The method of claim **1**, wherein

in step S**2**, a method for constructing the generative adversarial network through the generator and the discriminator specifically comprises:

S**2**.**1**: constricting a generator G_{XY} for converting a given image in a source domain X to an image in a target domain Y, and constructing a generator G_{YX} for converting a given image in the target domain Y to an image in the source domain X;

S**2**.**2**: constructing a discriminator D_Y for predicting whether the image is an image in the target domain, and constructing a discriminator D_X for predicting whether the image is an image in the source domain; and

S**2**.**3**: connecting the generator G_{XY} to the generator G_{YX}, connecting the generator G_{XY} to the discriminator D_Y, and connecting the generator G_{YX} to the discriminator D_X, to construct the generative adversarial network.

6. The method of claim **5**, wherein

in step S**2**.**3**, a one-cycle generation process of the generative adversarial network comprises a source domain cycle generation process and a target domain cycle generation process;

the source domain cycle generation process specifically comprises:

S**2**.**3**.**1**.**1**: inputting a source domain image x and a segmentation mask m_x of the source domain image x into the generator G_{XY} to obtain a first target domain generated image y and a segmentation mask m'_y of the first target domain generated image y', wherein the first target domain generated image y' and the segmentation mask m'_y are denoted as $G_{XY}(x,m_x)$;

S**2**,**3**.**1**.**2**: inputting the first target domain generated image y and the segmentation mask m'_y of the first target domain generated image y into the generator G_{YX} to obtain a first source domain generated image and a segmentation mask of the first source domain generated image, wherein the first source domain generated image and the segmentation mask of the first source domain generated image are denoted as $G_{YX}(G_{XY},(x,m_x))$;

S**2**,**3**.**1**.**3**: inputtingthe first target domain generated image y', the segmentation mask m'_y of the first target domain generated image y', a target domain image y, and a segmentation mask m_y of the target domain image y into the discriminator D_y for a first discrimination, and inputting $G_{YX}(x,m_x)$ denoting the first source domain generated image and the segmentation mask of the first source domain generated image into the discriminator D_X for a second discrimination, to complete the source domain cycle generation process;

the target domain cycle generation process specifically comprises:

S**2**.**3**.**2**,**1**: inputting the target domain image y and the segmentation mask m_y of the target domain image y into the generator G_{YX} to obtain a second source domain generated image x' and a segmentation mask m'_x of the second source domain generated image x', wherein the second source domain generated image x' and the segmentation mask m'_x of the second source domain generated image x' are denoted as $G_{YX}(y,m_y)$;

S**2**.**3**.**2**.**2**: inputting the second source domain generated image x' and the segmentation mask m'_x of the second source domain generated image x' into the generator G_{XY} to obtain a second target domain generated image and a segmentation mask of the second target domain generated image, wherein the second target domain generated image and the segmentation mask of the second target domain generated image are denoted as $G_{XY}(G_{YX}(y,m_y))$;

S**2**.**3**.**2**.**3**: inputting the second source domain generated image x', the segmentation mask m'_x of the second source domain generated image x', the source domain image x, and the segmentation mask m_x of the source domain image x into the discriminator D_X for a third discrimination, and inputting $G_{XY}(G_{YX}(y,m_y))$ denoting the second target domain generated image and the segmentation mask of the second target domain generated image into the discriminator D_Y for a fourth discrimination, to complete the target domain cycle generation process;

7. The method of claim **1**, wherein the loss function $\mathcal{L}_{total}$ step S**3** is specifically expressed as:

$$\mathcal{L}_{total}=\lambda_{adv}\mathcal{L}_{adv}+\lambda_{cyc}\mathcal{L}_{cyc}+\lambda_{idt}\mathcal{L}_{idt}+\lambda_{ctx}\mathcal{L}_{ctx}+\lambda_{fs}\mathcal{L}_{fs};$$

wherein $\mathcal{L}_{adv}$ represents an adversarial loss function, $\mathcal{L}_{cyc}$ represents a cycle consistency loss function, $\mathcal{L}_{idt}$ represents an identity loss function, $\mathcal{L}_{ctx}$ represents a context-preserving loss function, $\mathcal{L}_{fs}$ represents a feature similarity loss function, and λ_{adv}, λ_{cyc}, λ_{idt}, λ_{ctx} and λ_{fs} represent weights of $\mathcal{L}_{adv}$, $\mathcal{L}_{cyc}$, $\mathcal{L}_{idt}$, $\mathcal{L}_{ctx}$ and $\mathcal{L}_{fs}$ in the loss function $\mathcal{L}_{total}$, respectively.

8. The method of claim **7**, wherein

the adversarial loss function is specifically expressed as:

$$\mathcal{L}_{adv}=(D_X(x,m_x)-1)+D_X(G_{YX}(y,m_y))^2+(D_Y(y,m_y)-1)^2+D_Y(G_{XY}(x,m_x))^2;$$

wherein, $D_X(x,m_x)$ represents a discriminant output of a discriminator D_X for a source domain image x and a segmentation mask m, of the source domain image x,

$D_X(G_{YX}(y,m_y))$ represents a discriminant output of the discriminator D_X for $G_{YX}(y,m_y)$denoting a second source domain generated image and a segmentation mask of the second source domain generated image, $G_{YX}(y,m_y)$ represents the second source domain generated image and the segmentation mask of the second source domain generated. image generated by a target domain image y and a segmentation mask m_y of the target domain image y through a generator G_{YX}, $D_Y(y, m_y)$ represents a discriminant output of a discriminator D_Y for the target domain image y and the segmentation mask m_y of the target domain image y, $D_Y(G_{XY}(x,m_x)$ represents a discriminant output of the discriminator D_Y for $G_{XY}(x,m_x)$ denoting a first target domain generated image and a segmentation mask of the first target domain generated image, $G_{XY}(x,m_x)$represents the first targe domain generated image and the segmentation mask of the first target domain generated image generated by the source domain image x and the segmentation mask m_x of the source domain image x through a generator G_{XY};

the cycle consistency loss function $\mathcal{L}_{cyc}$ is specifically expressed as:

$$\mathcal{L}_{cyc}=\|G_{YX}(G_{XY}(x,m_x))-(x,m_x)\|_1+\|G_{XY}(G_{YX}(G_{YX}(y,m_y))-(y,m_y)\|_1;$$

wherein $G_{YX}(G_{XY}(x,m_x)$ represents the first source domain generated image and the segmentation mask of the first source domain generated image generated by $G_{XY}(x,m_x)$ through the generator G_{YX}, $G_{XY}(G_{YX}(y,m_y))$ represents the second target domain generated image and the segmentation mask of the second target domain generated image generated by G_{YX} (y,m_y) through the generator G_{XY}, and $\|*\|_1$ represents a 1-norm;

the identity loss function $\mathcal{L}_{idt}$ is specifically expressed as:

$$\mathcal{L}_{idt}=\|G_{XY}(y,m_y)-(y,m_y)\|_1+\|G_{YX}(x,m_x)-(x,m_x)\|_1;$$

wherein $G_{XY}(y,m_y)$ represents the segmentation mask of the first target domain generated image y obtained after the source domain image x and the segmentation mask m, of the source domain image x are input to the generator G_{XY}, and $G_{YX}(x,m_x)$ represents the segmentation mask of the second source domain generated image x' obtained after the target domain image y and the segmentation mask m_y of the target domain image y are input into the generator G_{YX};

the context-preserving loss function $\mathcal{L}_{ctx}$ is specifically expressed as:

$$\mathcal{L}_{ctx}=\|\omega(m_x,m'_y)e(x-y')\|_1+\|\omega(m_y,m'_x)e(y-x')\|_1;$$

wherein $\omega(m_x,m'_y)$ represents a minimum value obtained by subtracting one from elements in binary obi cc masks through the segmentation mask m_x of the source domain image x and the segmentation mask of m'_y the first target domain generated image y', and $\omega(m_y,m'_x)$ represents a minimum value obtained by subtracting one from the elements in the binary object masks through the segmentation mask m_y of the target domain image y and the segmentation mask m_x of the second source domain generated image x'; y' represents the first target domain generated image generated by the source domain image x through the generator G_{XY}, x' represents the second source domain generated image generated by the target domain image y and the segmentation mask m_y, of the target domain image y through the generator G_{YX};

the feature similarity loss function $\mathcal{L}_{fx}$ is specifically expressed as:

$$\begin{aligned}\mathcal{L}_{fs} &= FS(y, y')\\ &= \frac{1}{N}\sum_j \max_i FS_{ij}\\ &= \frac{1}{N}\sum_j \max_i\left(W_{ij}\Big/\sum_k \omega_{ik}\right)\\ &= \frac{1}{N}\sum_j \max_i\left[\exp\left(\frac{1-\overline{d_{ij}}}{h}\right)\Big/\sum_k \omega_{ik}\right];\end{aligned}$$

wherein FS(y,y') represents a similarity between an image y and an image y', and $\max FS_{ij}$ represents a feature i of the image y, wherein the feature i of the image y is most similar to a feature j of the image y'; N represents a total number of features of the image y and N is the same as a total number of features of the image y'; h represents a bandwidth parameter,

$$\exp\left(\frac{1-\overline{d_{ij}}}{h}\right)$$

represents an exponential operation from a distance d_{ij} to W_{ij}, d_{ij} represents a normalized similarity distance, W_{ij} represents a similarity, $w_{ij}/\Sigma_k w_{ik}$ represents a normalized similarity, and w_{ik} represents a similarity value of a k^{th} W_{ij}.

* * * * *