CLAIMS

What is claimed is:

1. A method for training a face detection model, comprising:

acquiring a training face image;

performing three-dimensional reconstruction on the training face image based on a preset three-dimensional face model, and acquiring a training three-dimensional face model by: selecting M three-dimensional face models; performing principal component analysis on the M three-dimensional face models, and acquiring a principal component matrix; and performing three-dimensional reconstruction on each training face image by using the principal component matrix and an eigenvalue matrix corresponding to each training face image, and acquiring the training three-dimensional face model;

generating a training UV coordinate map comprising three-dimensional coordinates of the training three-dimensional face model based on the training three-dimensional face model; and

training a semantic segmentation network by using the training face image and the training UV coordinate map, and acquiring a face detection model by: extracting the training face image; inputting the training face image into the semantic segmentation network to extract a predicted UV coordinate map; calculating a loss rate of the predicted UV coordinate map by using a preset loss function and the training UV coordinate map; calculating a gradient by using the loss rate; judging whether the gradient satisfies a preset iteration condition; determining the semantic segmentation network as a face detection model based on a judgment result that the gradient satisfies the preset iteration condition; and performing gradient descent on a network parameter of the semantic segmentation network by using the gradient and a preset learning rate, and returning to the step of extracting the training face image based on a judgment result that the gradient does not satisfy the preset iteration condition, wherein the face detection model is configured to generate a UV coordinate map comprising three-dimensional coordinates.

~~2. The method according to claim 1, wherein performing the three-dimensional reconstruction on the training face image based on the preset three-dimensional face model, and acquiring the training three-dimensional face model comprises:~~

~~selecting M three-dimensional face models;~~

~~performing principal component analysis on the M three-dimensional face models,~~

~~and acquiring a principal component matrix; and~~

~~performing three-dimensional reconstruction on each training face image by using the principal component matrix and an eigenvalue matrix corresponding to each training face image, and acquiring the training three-dimensional face model.~~

~~3~~2. The method according to claim ~~2~~1, before performing principal component analysis on the M three-dimensional face models, and acquiring a principal component matrix, further comprising:

preprocessing selected three-dimensional face models; and

aligning preprocessed three-dimensional face models by using an optical flow method, and acquiring aligned three-dimensional face models,

wherein the preprocessing comprises at least one of: smoothing, hole-filling and coordinate correction.

~~4~~3. The method according to claim ~~2~~ 1 or ~~3~~2, wherein performing three-dimensional reconstruction on each training face image by using the principal component matrix and an eigenvalue matrix corresponding to each training face image, and acquiring the training three-dimensional face model comprises:

setting an initial eigenvalue matrix and an initial projection parameter matrix;

constructing an initial three-dimensional face model by using the initial eigenvalue matrix and the principal component matrix;

acquiring a projected face image of the initial three-dimensional face model in a two-dimensional space based on the initial projection parameter matrix;

calculating a difference value between the projected face image and the training face image;

iteratively optimizing the initial eigenvalue matrix and the initial projection parameter matrix by using a stochastic gradient descent method based on the difference value, acquiring the eigenvalue matrix and the projection parameter matrix, when the difference value converges and taking the eigenvalue matrix and the projection parameter matrix as a target eigenvalue matrix and a target projection parameter matrix, and taking the target eigenvalue matrix as the eigenvalue matrix corresponding to the training face image; and

generating the training three-dimensional face model by using the target eigenvalue matrix and the principal component matrix.

~~5~~4. The method according to claim ~~4~~3, wherein iteratively optimizing the initial eigenvalue matrix and the initial projection parameter matrix by using a stochastic gradient descent method based on the difference value, acquiring the eigenvalue matrix and the projection parameter matrix, when the difference value converges, and taking the eigenvalue matrix and the projection parameter matrix as a target eigenvalue matrix and a target projection parameter matrix comprises:

randomly selecting K feature points from the training face image and the projected face image, for first N rounds of iterative optimization;

calculating a difference value between the projected face image and the training face image based on the K feature points;

iteratively optimizing part of the eigenvalues in the initial eigenvalue matrix and the initial projection parameter matrix using the stochastic gradient descent method based on the difference value; and

after the N rounds of iterative optimization, optimizing all the eigenvalues in the eigenvalue matrix after the N rounds of iterations and the projection parameter matrix after the N rounds of iterations by using the stochastic gradient descent method, acquiring the eigenvalue matrix and the projection parameter matrix, when the difference value converges, and taking the eigenvalue matrix and the projection parameter matrix as the target eigenvalue matrix and the target projection parameter matrix.

~~6~~5. The method according to ~~any of~~ claim~~s~~ 1~~-3~~ or 2, wherein generating the training UV coordinate map comprising the three-dimensional coordinates of the training three-dimensional face model based on the training three-dimensional face model comprises:

acquiring a plurality of vertices of the training three-dimensional face model, each of the vertices having three-dimensional coordinates;

projecting the plurality of vertices onto a preset spherical surface and acquiring a plurality of projection points corresponding to the plurality of vertices on the preset spherical surface; and

unfolding the spherical surface comprising the plurality of projection points to generate a training UV coordinate map comprising the plurality of projection points, wherein each projection point in the training UV coordinate map is associated with three-dimensional coordinates of a corresponding vertex.

~~7. The method according to any of claims 1-3, wherein training the semantic segmentation network by using the training face image and the training UV coordinate map, and acquiring the face detection model comprises:~~

~~extracting the training face image;~~

~~inputting the training face image into the semantic segmentation network to extract a predicted UV coordinate map;~~

~~calculating a loss rate of the predicted UV coordinate map by using a preset loss function and the training UV coordinate map;~~

~~calculating a gradient by using the loss rate;~~

~~judging whether the gradient satisfies a preset iteration condition;~~

~~determining the semantic segmentation network as a face detection model based on a judgment result that the gradient satisfies the preset iteration condition; and~~

~~performing gradient descent on a network parameter of the semantic segmentation network by using the gradient and a preset learning rate, and returning to the step of extracting the training face image based on a judgment result that the gradient does not satisfy the preset iteration condition.~~

~~8~~6. The method according to claim ~~7~~1, wherein calculating the loss rate of the predicted UV coordinate map by using the preset loss function and the training UV coordinate map comprises:

acquiring a training UV coordinate map corresponding to the training face image; and

calculating the loss rate of the predicted UV coordinate map by using the training UV coordinate map, the predicted UV coordinate map, the preset loss function and a preset loss weight, wherein the preset loss weight is a preset loss weight of a feature point in the training UV coordinate map.

~~9~~7. A method for detecting a face key point, comprising:

acquiring a target face image;

inputting the target face image into a pre-trained face detection model to generate a UV coordinate map of the target face image, wherein the UV coordinate map comprises a plurality of pixel points, and each pixel point of the UV coordinate map comprises three-dimensional coordinates;

acquiring a UV template image comprising a pre-labeled face key point, wherein a resolution of the UV template image is the same as that of the UV coordinate map, and pixels in the UV template image and the UV coordinate map are in a one-to-one correspondence; and

determining a pixel point corresponding to the pre-labeled face key point in the UV coordinate map to detect three-dimensional coordinates of the face key point,

wherein the face detection model is trained by the method for training a face detection model according to any of claims 1-6.

~~10~~8. An apparatus for training a face detection model, comprising:

a training face image acquiring module configured to acquire a training face image;

a three-dimensional reconstruction module configured to perform three-dimensional reconstruction on the training face image based on a preset three-dimensional face model and acquire a training three-dimensional face model by: selecting M three-dimensional face models; performing principal component analysis on the M three-dimensional face models, and acquiring a principal component matrix; and performing three-dimensional reconstruction on each training face image by using the principal component matrix and an eigenvalue matrix corresponding to each training face image, and acquiring the training three-dimensional face model;

a training UV coordinate map generating module configured to generate a training UV coordinate map comprising three-dimensional coordinates of the training three-dimensional face model based on the training three-dimensional face model; and

a training module configured to train a semantic segmentation network by using the training face image and the training UV coordinate map and acquire a face detection model by: extracting the training face image; inputting the training face image into the semantic segmentation network to extract a predicted UV coordinate map; calculating a loss rate of the predicted UV coordinate map by using a preset loss function and the training UV coordinate map; calculating a gradient by using the loss rate; judging whether the gradient satisfies a preset iteration condition; determining the semantic segmentation network as a face detection model based on a judgment result that the gradient satisfies the preset iteration condition; and performing gradient descent on a network parameter of the semantic segmentation network by using the gradient and a preset learning rate, and returning to the step of extracting the training face image based on a judgment result that the gradient does not satisfy the preset iteration condition,

wherein the face detection model is configured to generate a UV coordinate map comprising three-dimensional coordinates.

~~11~~9. An apparatus for detecting a face key point, comprising:

a target face image acquiring module configured to acquire a target face image;

a UV coordinate map generating module configured to input the target face image into a pre-trained face detection model to generate a UV coordinate map of the target face image, wherein the UV coordinate map comprises a plurality of pixel points, and each pixel point of the UV coordinate map comprises three-dimensional coordinates;

a template image acquiring module configured to acquire a UV template image comprising a pre-labeled face key point, wherein a resolution of the UV template image is the same as that of the UV coordinate map, and pixels in the UV template image and the UV coordinate map are in a one-to-one correspondence; and

a face key point three-dimensional coordinates detecting module configured to determine a pixel point corresponding to the pre-labeled face key point in the UV coordinate map to detect three-dimensional coordinates of the face key point,

wherein the face detection model is trained by the method for training a face detection model according to any of claims 1-6.

~~12~~10. A device comprising:

at least one processor; and

a storage apparatus for storing at least one program,

wherein when the at least one program is executed by the at least one processor, the at least one processor is enabled to realize at least one of the method for training a face detection model according to any of claims 1-~~8~~ 6 and the method for detecting a face key point according to claim ~~9~~7.

~~13~~11. A computer-readable storage medium storing a computer program therein, wherein when the program is executed by a processor, at least one of the method for training a face detection model according to any of claims 1-~~8~~ 6 and the method for detecting a face key point according to claim ~~9~~ 7 is realized.