

CLAIMS

1. A face recognition method, comprising:

acquiring an image to be recognized; and

recognizing the image to be recognized based on a cross-modal face recognition network to obtain a recognition result of the image to be recognized, wherein the cross-modal face recognition network is obtained by training based on face image data in different modalities,

wherein a process of obtaining the cross-modal face recognition network by training based on the face image data in the different modalities comprises:

training a first modal network based on a first image set and a second image set to obtain a second modal network, wherein an object in the first image set belongs to a first category, and an object in the second image set belongs to a second category;

selecting a first number of images in the first image set and a second number of images in the second image set according to a preset condition, and obtaining a third image set according to the first number of images and the second number of images; and

training the second modal network based on the third image set to obtain the cross-modal face recognition network,

wherein the first modal network comprises a first feature extraction branch, a second feature extraction branch and a third feature extraction branch, and

wherein training the first modal network based on the first image set and the second image set to obtain the second modal network comprises:

training the first modal network by inputting the first image set to the first feature extraction branch, inputting the second image set to the second feature extraction branch and inputting a fourth image set to the third feature extraction branch, wherein images in the fourth image set are images collected in a same scenario or images collected in a same

collection manner; and

determining a trained first feature extraction branch or a trained second feature extraction branch or a trained third feature extraction branch as the second modal network.

~~2. The method of claim 1, wherein a process of obtaining the cross-modal face recognition network by training based on the face image data in the different modalities comprises:~~

~~obtaining the cross-modal face recognition network by training based on a first modal network and a second modal network.~~

~~3. The method of claim 2, further comprising: before obtaining the cross-modal face recognition network by training based on the first modal network and the second modal network,~~

~~training the first modal network based on a first image set and a second image set, wherein an object in the first image set belongs to a first category, and an object in the second image set belongs to a second category.~~

~~4. The method of claim 3, wherein training the first modal network based on the first image set and the second image set comprises:~~

~~training the first modal network based on the first image set and the second image set to obtain the second modal network;~~

~~selecting a first number of images in the first image set and a second number of images in the second image set according to a preset condition, and obtaining a third image set according to the first number of images and the second number of images; and~~

~~training the second modal network based on the third image set to obtain the cross-modal face recognition network.~~

52. The method of claim 41, wherein the preset condition comprises any one of: the first number is the same as the second number, a ratio of the first number to the second number is equal to a ratio of a number of images in the first image set to a number of images in the second image set, or the ratio of the first number to the second number is equal to a ratio of a number of persons in the first image set to a number of persons in the second image set.

~~6. The method of claim 2 or 4, wherein the first modal network comprises a first feature extraction branch, a second feature extraction branch and a third feature extraction branch, and~~

~~wherein training the first modal network based on the first image set and the second image set to obtain the second modal network comprises:~~

~~training the first modal network by inputting the first image set to the first feature extraction branch, inputting the second image set to the second feature extraction branch and inputting a fourth image set to the third feature extraction branch, wherein images in the fourth image set are images collected in a same scenario or images collected in a same collection manner; and~~

~~determining a trained first feature extraction branch or a trained second feature extraction branch or a trained third feature extraction branch as the second modal network.~~

73. The method of claim 61, wherein training the first modal network by inputting the first image set to the first feature extraction branch, inputting the second image set to the second feature extraction branch and inputting the fourth image set to the third feature extraction branch comprises:

inputting the first image set, the second image set and the fourth image set to the first feature extraction branch, the second feature extraction branch and the third feature extraction branch respectively to obtain a first recognition result, a second recognition result and a third

recognition result respectively;

acquiring a first loss function of the first feature extraction branch, a second loss function of the second feature extraction branch and a third loss function of the third feature extraction branch; and

adjusting a parameter of the first modal network according to the first image set as well as the first recognition result and the first loss function, the second image set as well as the second recognition result and the second loss function, and the fourth image set as well as the third recognition result and the third loss function to obtain an adjusted first modal network, wherein the parameter of the first modal network comprises a parameter of the first feature extraction branch, a parameter of the second feature extraction branch and a parameter of the third feature extraction branch, and the parameters of the three branches of the adjusted first modal network are identical to each other.

84. The method of claim 73, wherein the image in the first image set comprises first labeling information, the image in the second image set comprises second labeling information, the image in the fourth image set comprises third labeling information, and

wherein adjusting the parameter of the first modal network according to the first image set as well as the first recognition result and the first loss function, the second image set as well as the second recognition result and the second loss function, and the fourth image set as well as the third recognition result and the third loss function to obtain the adjusted first modal network comprises:

obtaining a first gradient according to the first labeling information, the first recognition result, the first loss function and an initial parameter of the first feature extraction branch, obtaining a second gradient according to the second labeling information, the second recognition result, the second loss function and an initial parameter of the second feature extraction branch, and obtaining a third gradient according to the third labeling information, the third recognition result, the third loss function and an initial parameter of the third feature extraction branch; and

determining an average value of the first gradient, the second gradient and the third gradient as a back propagation gradient of the first modal network, and adjusting the parameter of the first modal network through the back propagation gradient, so that the parameter of the first feature extraction branch, the parameter of the second feature extraction branch and the parameter of the third feature extraction branch are identical to each other.

95. The method of claim 41 or 52, wherein selecting the first number of images in the first image set and the second number of images in the second image set according to the preset condition to obtain the third image set comprises:

selecting f images from the first image set and the second image set respectively so that a number of persons in the f images is a threshold, to obtain the third image set; or,

selecting m images and n images from the first image set and the second image set respectively so that a ratio of m to n is equal to the ratio of the number of the images in the first image set to the number of the images in the second image set and the numbers of persons in the m images and the n images are both the threshold, to obtain the third image set; or,

selecting s images and t images from the first image set and the second image set respectively so that a ratio of s to t is equal to the ratio of the number of the persons in the first image set to the number of the persons in the second image set and the numbers of persons in the s images and the t images are both the threshold, to obtain the third image set.

106. The method of claim 31, wherein training the second modal network based on the third image set to obtain the cross-modal face recognition network comprises:

sequentially performing feature extraction processing, linear transformation and nonlinear transformation on an image in the third image set to obtain a fourth recognition result; and

adjusting a parameter of the second modal network according to the image in the third

image set, the fourth recognition result and a fourth loss function of the second modal network to obtain the cross-modal face recognition network.

~~417~~. The method of any one of claims 1 to ~~5, 7, 8-4~~ and ~~106~~, wherein the first category and the second category correspond to different races.

~~428~~. A face recognition device, comprising:

an acquisition unit, configured to acquire an image to be recognized; and

a recognition unit, configured to recognize the image to be recognized based on a cross-modal face recognition network to obtain a recognition result of the image to be recognized, wherein the cross-modal face recognition network is obtained by training based on face image data in different modalities,

wherein the recognition unit comprises:

a training subunit, configured to:

train a first modal network based on a first image set and a second image set to obtain a second modal network, wherein an object in the first image set belongs to a first category, and an object in the second image set belongs to a second category,

select a first number of images in the first image set and a second number of images in the second image set according to a preset condition and obtain a third image set according to the first number of images and the second number of images; and

train the second modal network based on the third image set to obtain the cross-modal face recognition network,

wherein the first modal network comprises a first feature extraction branch, a second feature extraction branch and a third feature extraction branch, and the training subunit is further configured to:

train the first modal network by inputting the first image set to the first feature extraction branch, input the second image set to the second feature extraction branch and input a fourth image set to the third feature extraction branch, wherein images in the fourth image set are images collected in a same scenario or images collected in a same collection manner; and

determine a trained first feature extraction branch or a trained second feature extraction branch or a trained third feature extraction branch as the second modal network.

~~13. The device of claim 12, wherein the recognition unit comprises:~~

~~a training subunit, configured to obtain the cross-modal face recognition network by training based on a first modal network and a second modal network.~~

~~14. The device of claim 13, wherein the training subunit is further configured to:~~

~~train the first modal network based on a first image set and a second image set, wherein an object in the first image set belongs to a first category, and an object in the second image set belongs to a second category.~~

~~15. The device of claim 14, wherein the training subunit is further configured to:~~

~~train the first modal network based on the first image set and the second image set to obtain the second modal network;~~

~~select a first number of images in the first image set and a second number of images in the second image set according to a preset condition and obtain a third image set according to the first number of images and the second number of images; and~~

~~train the second modal network based on the third image set to obtain the cross-modal face recognition network.~~

~~169.~~ The device of claim ~~158~~, wherein the preset condition comprises any one of: the first number is the same as the second number, a ratio of the first number to the second number is equal to a ratio of a number of images in the first image set to a number of images in the second image set, or the ratio of the first number to the second number is equal to a ratio of a number of persons in the first image set to a number of persons in the second image set.

~~17. The device of claim 13 or 15, wherein the first modal network comprises a first feature extraction branch, a second feature extraction branch and a third feature extraction branch, and the training subunit is further configured to:~~

~~train the first modal network by inputting the first image set to the first feature extraction branch, input the second image set to the second feature extraction branch and input a fourth image set to the third feature extraction branch, wherein images in the fourth image set are images collected in a same scenario or images collected in a same collection manner; and~~

~~determine a trained first feature extraction branch or a trained second feature extraction branch or a trained third feature extraction branch as the second modal network.~~

~~1810.~~ The device of claim ~~171~~, wherein the training subunit is further configured to:

~~input the first image set, the second image set and the fourth image set to the first feature extraction branch, the second feature extraction branch and the third feature extraction branch respectively to obtain a first recognition result, a second recognition result and a third recognition result respectively;~~

~~acquire a first loss function of the first feature extraction branch, a second loss function of the second feature extraction branch and a third loss function of the third feature extraction branch; and~~

adjust a parameter of the first modal network according to the first image set as well as the first recognition result and the first loss function, the second image set as well as the second recognition result and the second loss function, and the fourth image set as well as the third recognition result and the third loss function to obtain an adjusted first modal network, wherein the parameter of the first modal network comprises a parameter of the first feature extraction branch, a parameter of the second feature extraction branch and a parameter of the third feature extraction branch, and the parameters of the three branches of the adjusted first modal network are identical to each other.

[1911](#). An electronic device, comprising a memory and a processor, wherein the memory has stored thereon computer-executable instructions that, when executed by the processor, cause the method of any one of claims 1-[117](#) to be implemented.

[2012](#). A computer-readable storage medium having stored thereon a computer program that, when executed by a processor, causes the method of any one of claims 1-[117](#) to be implemented.