

TRAINING METHOD AND APPARATUS FOR FACE RECOGNITION MODEL, ELECTRONIC DEVICE AND STORAGE MEDIUM

The present application claims the priority to a Chinese patent application No. 202111637930.4 filed with the China National Intellectual Property Administration on December 29, 2021 and entitled “Training method and apparatus for face recognition model, electronic device and storage medium”, which is incorporated herein by reference in its entirety.

TECHNICAL FIELD

The present application relates to the field of deep learning technology, and in particular to a training method and apparatus for a face recognition model, an electronic device and a storage medium.

BACKGROUND

With the rapid development of a computer and deep learning technology, a deep learning model is increasingly widely used in the field of face recognition. Currently, a face recognition model based on deep learning needs to be trained using the full face data, and the trained face recognition model can perform identity recognition based on a person's face feature information. However, the performance of the face recognition model trained using source domain data is poor in a target domain.

Due to the characteristics of catastrophic forgetting of model training, the face recognition model will forget the recognition performance in the source domain after trained using target domain data separately.

In the case of that there is only a small amount of storage space can store a small amount of source domain data, how to train a face recognition model in a target domain, so that the trained face recognition model cannot only maintain the performance of the source domain but also improve its performance in the target, so as to improve a recognition capability and an accuracy of the face recognition model, is an urgent problem to be solved.

SUMMARY

Embodiments of the present application provide a training method and apparatus for a face recognition model, an electronic device and a storage medium, which can at least maintain the performance of the face recognition model in the source domain and improve its performance in the

target domain. The specific technical solutions are as follows:

In a first aspect, an embodiment of the present application provides a training method for a face recognition model, the method includes:

acquiring source domain face features and an initialization recognition model, wherein the initialization recognition model is obtained based on the training of full face image samples in a source domain, and the source domain face features are face features of the full face image samples obtained through the initialization recognition model;

acquiring target face image samples in a target domain, wherein identity labels corresponding to the target face image samples are unknown; and

adjusting a portion of model parameters of the initialization recognition model based on the target face image sample and the source domain face features, until the initialization recognition model converges, so as to obtain the face recognition model for the source domain and the target domain.

Optionally, acquiring the source domain face features includes:

screening the full face image samples based on a preset screening strategy, to obtain screened full face image samples, wherein the preset screening strategy ensures that the number of pieces of identity information corresponding to the screened full face image samples is not less than a preset number when the number of the screened full face image samples remains unchanged;

inputting the screened full face image samples into the initialization recognition model, to obtain face features output by an intermediate layer of the initialization recognition model; and

determining the source domain face features based on the face features output by the intermediate layer.

Optionally, the initialization recognition model includes a fixed parameter portion and a to-be-trained portion;

wherein adjusting a portion of model parameters of the initialization recognition model based on the target face image samples and the source domain face features, includes:

inputting the target face image samples into the fixed parameter portion and the to-be-trained portion, to obtain first prediction labels, and determining a first classification loss based on the first

prediction labels and pseudo labels corresponding to the target face image samples;

inputting the source domain face features into the to-be-trained portion, to obtain second prediction labels, and determining a second classification loss based on the second prediction labels and identity labels corresponding to the source domain face features;

inputting the source domain face features into the to-be-trained portion and an initial portion corresponding to the to-be-trained portion respectively, to obtain estimated features and initial features, and determining a constraint loss based on the estimated features and the initial features, wherein the initial portion is a corresponding model portion when a model parameter of the to-be-trained portion is fixed to a model parameter trained based on the full face image samples; and

adjusting the model parameter of the to-be-trained portion based on the first classification loss, the second classification loss and the constraint loss.

Optionally, before inputting the target face image samples into the fixed parameter portion and the to-be-trained portion to obtain first prediction labels, the method further includes:

clustering the target face image samples, and determining a pseudo label corresponding to each target face image sample, wherein the pseudo label is used to identify an identity of a person to whom the corresponding target face image sample belongs.

Optionally, adjusting the model parameter of the to-be-trained portion based on the first classification loss, the second classification loss and the constraint loss, includes:

calculating a loss function value L based on the first classification loss, the second classification loss and the constraint loss, according to the following formula:

$$L = L_{c1} + L_{c2} + \lambda L_{kd}$$

wherein, L_{c1} is the first classification loss, L_{c2} is the second classification loss, L_{kd} is the constraint loss, and λ is a preset parameter; and

adjusting the model parameter of the to-be-trained portion based on the loss function value.

Optionally, determining a constraint loss based on the estimated features and the initial features, includes:

calculating the constraint loss L_{kd} based on the estimated features and the initial features, according to the following formula:

$$L_{kd} = \frac{1}{n} \sum_{i=1}^n (F_i - \bar{F}_i)^2$$

wherein, n is the number of the source domain face features, F_i is an initial feature corresponding to i -th source domain face feature, and $\bar{F}_i$ is an estimated feature corresponding to i -th source domain face feature.

Optionally, determining the source domain face features based on the face features output by the intermediate layer, includes:

performing a dimension reduction processing on the face features output by the intermediate layer to obtain dimension reduced face features as the source domain face features;

before adjusting a portion of model parameters of the initialization recognition model based on the target face image samples and the source domain face features, the method further includes:

performing a dimension restoration processing on the source domain face features to obtain restored source domain face features.

Optionally, the method further includes:

acquiring a face image to be recognized in the target domain; and

identifying the face image to be recognized based on the face recognition model to determine an identity corresponding to the face image to be recognized.

In a second aspect, an embodiment of the present application provides a training apparatus for a face recognition model, the apparatus includes:

an initialization training module, configured for acquiring source domain face features and an initialization recognition model, wherein the initialization recognition model is obtained based on the training of full face image samples in a source domain, and the source domain face features are face features of the full face image samples obtained through the initialization recognition model;

an acquisition module for target domain samples, configured for acquiring target face image samples in a target domain, wherein identity labels corresponding to the target face image samples are unknown; and

an incremental training module, configured for adjusting a portion of model parameters of the initialization recognition model based on the target face image samples and the source domain face features, until the initialization recognition model converges, so as to obtain the face recognition model for the source domain and the target domain.

Optionally, the initialization training module includes:

a sample screening unit, configured for screening the full face image samples based on a preset screening strategy to obtain screened full face image samples, wherein the preset screening strategy ensures that the number of pieces of identity information corresponding to the screened full face image samples is not less than a preset number when the number of the screened full face image samples remains unchanged;

a feature acquisition unit, configured for inputting the screened full face image samples into the initialization recognition model to obtain face features output by an intermediate layer of the initialization recognition model; and

a feature determination unit, configured for determining the source domain face features based on the face features output by the intermediate layer.

Optionally, the initialization recognition model includes a fixed parameter portion and a to-be-trained portion;

the incremental training module includes:

a first input unit, configured for inputting the target face image samples into the fixed parameter portion and the to-be-trained portion to obtain first prediction labels, and determining a first classification loss based on the first prediction labels and pseudo labels corresponding to the target face image samples;

a second input unit, configured for inputting the source domain face features into the to-be-trained portion to obtain second prediction labels, and determining a second classification loss based on the second prediction labels and based on identity labels corresponding to the source domain face features;

a third input unit, configured for inputting the source domain face features into the to-be-trained portion and an initial portion corresponding to the to-be-trained portion respectively, to obtain estimated features and initial features, and determining a constraint loss based on the

estimated features and the initial features, wherein the initial portion is a corresponding model portion when a model parameter of the to-be-trained portion is fixed to a model parameter trained based on the full face image samples; and

a parameter adjustment unit, configured for adjusting the model parameter of the to-be-trained portion based on the first classification loss, the second classification loss and the constraint loss.

Optionally, the apparatus further includes:

a clustering module for target face image samples, configured for clustering the target face image samples and determining a pseudo label corresponding to each target face image sample before inputting the target face image samples into the fixed parameter portion and the to-be-trained portion to obtain first prediction labels, wherein the pseudo label is used to identify an identity of a person to whom the corresponding target face image sample belongs.

Optionally, the parameter adjustment unit includes:

a loss function value calculation sub-unit, configured for calculating a loss function value L based on the first classification loss, the second classification loss and the constraint loss, according to the following formula:

$$L = L_{c1} + L_{c2} + \lambda L_{kd}$$

wherein, L_{c1} is the first classification loss, L_{c2} is the second classification loss, L_{kd} is the constraint loss, and λ is a preset parameter; and

a parameter adjustment sub-unit, configured for adjusting the model parameter of the to-be-trained portion based on the loss function value.

Optionally, the third input unit includes:

a constraint loss calculation sub-unit, configured for calculating the constraint loss L_{kd} based on the estimated features and the initial features, according to the following formula:

$$L_{kd} = \frac{1}{n} \sum_{i=1}^n (F_i - \bar{F}_i)^2$$

wherein, n is the number of the source domain face features, F_i is an initial feature corresponding to i -th source domain face feature, and $\bar{F}_i$ is an estimated feature corresponding to

i-th source domain face feature.

Optionally, the feature determination unit includes:

a feature dimension reduction sub-unit, configured for performing a dimension reduction processing on the face features output by the intermediate layer to obtain dimension reduced face feature as the source domain face features.

The apparatus further includes:

a feature restoration module, configured for performing a dimension restoration processing on the source domain face features to obtain restored source domain face features before adjusting a portion of model parameters of the initialization recognition model based on the target face image samples and the source domain face features.

Optionally, the apparatus further includes:

an acquisition module of a face image to be recognized, configured for acquiring a face image to be recognized in the target domain; and

an identity determination module, configured for identifying the face image to be recognized based on the face recognition model to determine an identity corresponding to the face image to be recognized.

In a third aspect, an embodiment of the present application provides an electronic device including a processor, a communication interface, a memory, and a communication bus, wherein the processor, the communication interface, and the memory communication with each other over the communication bus,

the memory, configured for storing a computer program,

the processor, when executing the program stored in the memory, configured for implementing any of the method steps described in the first aspect.

In a fourth aspect, an embodiment of the present application provides a computer-readable storage medium storing a computer program thereon, when executed by a processor, the computer program causes the processor to implement any of the method steps described in the first aspect.

In a fifth aspect, an embodiment of the present application provides a computer program product containing instructions, that when running on a computer, the computer program product

causes the computer to carry out any of the method steps described in the first aspect.

Beneficial effects of the embodiments of the present application:

in the solution provided by the embodiments of the present application, the electronic device can acquire source domain face features and an initialization recognition model, wherein the initialization recognition model is obtained based on the training of full face image samples in a source domain, and source domain face features are face feature of the full face image samples obtained through the initialization recognition model; acquire target face image samples in a target domain, wherein identity labels corresponding to the target face image samples are unknown; adjust a portion of model parameters of the initialization recognition model based on the target face image samples and the source domain face features, until the initialization recognition model converges, so as to obtain the face recognition model for the source domain and the target domain. After the initialization model is trained by using the full face image samples in the source domain, a portion of the source domain face features are stored, and a portion of the parameters of the initialization model are fixed. Furthermore, after the initialization model is further trained by using the target face image samples in the target domain and the source domain face features, the face recognition model for the source domain and the target domain is obtained. This face recognition model not only maintains the recognition capability of the full face images in the source domain, but also can accurately recognize the target face images in the target domain. When the face recognition model cannot be trained by using face data of the source domain and the target domain simultaneously, the face images can be recognized accurately, thereby improving the recognition capability and accuracy of the face recognition model. Of course, it is not necessary for any product or method implementing the present application to achieve all the above-mentioned advantages simultaneously.

BRIEF DESCRIPTION OF DRAWINGS

In order to more clearly describe the technical solutions of the embodiments of the present application and the prior art, the accompanying drawings that need to be used in the embodiments and the prior art are briefly described below. Obviously, the accompanying drawings described below are for only some embodiments of the present application. Those skilled in the art can also obtain other accompanying drawings based on these accompanying drawings without any creative efforts.

Figure 1 is a flowchart of a training method for a face recognition model provided by an embodiment of the present application;

Figure 2 is a specific flowchart of acquiring source domain face features in step S101 in the embodiment shown in Figure 1;

Figure 3 is a specific flowchart of adjusting a portion of model parameters of an initialization recognition model in step S103 in the embodiment shown in Figure 1;

Figure 4 is a schematic diagram of training the initialization recognition model using target face image samples and source domain face features based on the embodiment shown in Figure 1;

Figure 5 is a schematic diagram of feature extraction and feature dimension reduction processing based on the embodiment shown in Figure 1;

Figure 6 is a specific flowchart of determining an identity of a face image based on the embodiment shown in Figure 1;

Figure 7 is a schematic structural diagram of a training apparatus for a face recognition model provided by an embodiment of the present application;

Figure 8 is a specific structural schematic diagram of an incremental training module in the embodiment shown in Figure 7;

Figure 9 is a schematic structural diagram of an electronic device provided by an embodiment of the present application.

DETAILED DESCRIPTION

In order to make the objective, technical solutions, and advantages of the present application clearer and more understandable, the present application will be further described in detail below with reference to the accompanying drawings and embodiments. Obviously, the described embodiments are only some, and not all, of the embodiments of the present application. All other embodiments obtained based on the embodiments of the present disclosure by those skilled in the art fall into the scope of protection of the present disclosure.

In order to enable the trained face recognition model to not only maintain the performance of a source domain but also improve its performance in a target domain in a scenario where face data of the source domain and the target domain cannot be used at the same time to train the face

recognition model, embodiments of the present application provide a training method and apparatus for a face recognition model, an electronic device and a storage medium. The following first describes a training method for a face recognition model provided by the embodiments of the present application.

The training method for a face recognition model provided by an embodiment of the present application can be applied to any electronic device that can perform face recognition model training, for example, it can be various computing devices used for model training, it can be servers corresponding to gates at the entrances and exits of various parks, it can be servers for face recognition devices in security systems, etc., which are not specifically limited here. For clarity of description, it is called an electronic device subsequently.

As shown in Figure 1, a training method for a face recognition model, includes:

S101, acquiring source domain face features and an initialization recognition model,

wherein the initialization recognition model is obtained based on the training of full face image samples in a source domain, and source domain face features are face features of the full face image samples obtained through the initialization recognition model;

S102, acquiring target face image samples in a target domain,

wherein identity labels corresponding to the target face image samples are unknown;

S103, adjusting a portion of model parameters of the initialization recognition model based on the target face image samples and the source domain face features, until the initialization recognition model converges, so as to obtain the face recognition model for the source domain and the target domain.

It can be seen that in the solution provided by the embodiment of the present application, the electronic device can acquire source domain face features and an initialization recognition model, wherein the initialization recognition model is obtained based on the training of full face image samples in a source domain, and source domain face features are face features of the full face image samples obtained through the initialization recognition model; target face image samples in a target domain are acquired, wherein identity labels corresponding to the target face image samples are unknown; a portion of model parameters of the initialization recognition model is adjusted based on the target face image samples and the source domain face features, until the initialization

recognition model converges, so as to obtain the face recognition model for the source domain and the target domain. After the initialization model is trained by using the full face image samples in the source domain, a portion of the source domain face features are stored, and a portion of the parameters of the initialization model are fixed. Furthermore, after the initialization model is further trained by using the target face image samples in the target domain and the source domain face features, the face recognition model for the source domain and the target domain is obtained. This face recognition model cannot only maintain the recognition capability of the full face images in the source domain, but also can recognize the target face images in the target domain accurately. When the face recognition model cannot be trained by using face data of the source domain and the target domain simultaneously, the face images can be recognized accurately, thereby improving the recognition capability and accuracy of the face recognition model.

The face recognition model based on deep learning has excellent face recognition performance in conventional scenes. However, in some special scenes, such as child face recognition, low-quality face recognition and face recognition of people wearing masks, etc., there is still room for improvement in face recognition performance. For example, a convolutional neural network is a commonly used deep learning network for face recognition, that when training, it cannot be trained with face images in different scenes simultaneously due to issues such as data privacy or training resources. The training of convolutional neural networks also has the characteristics of catastrophic forgetting, which means that a face recognition model that has acquired a portion of face recognition capabilities through training forgets or loses the portion of face recognition capabilities previously acquired when learning to recognize a new face image.

In order for the face recognition model to accurately recognize a face image, the model needs to be trained with full face data. However, due to data privacy and other issues, it may not be possible to obtain the full face data, so that the model can only be trained with non-full face data, but training with such non-full face data, the face recognition performance of the resulting face recognition model will be very poor. For example, when the face recognition model is a model based on a convolutional neural network and the face data is images of faces wearing masks, due to the characteristics of catastrophic forgetting, if the face recognition model is trained only using these images of faces wearing masks as conventional training, the resulting face recognition model can hardly recognize face images without masks, and the recognition capability and accuracy are very poor.

When the face recognition model cannot be trained using face data of a source domain and a target domain simultaneously, in order to accurately recognize a face image in the target domain, the electronic device can train the face recognition model by domain adaptation incremental learning. In the embodiment of the present application, the face data used to train the face recognition model can be divided into full face image samples in the source domain and target face image samples in the target domain, wherein the full face image samples are face images with known identity information; the target face image samples in the target domain are face image samples of persons to be recognized, and the identity labels corresponding to the target face image samples are unknown.

Domain adaptation is a transfer learning method in which a data distribution corresponding to the source domain and the target domain are different, but the tasks corresponding to both are the same. For example, in this embodiment, although the distributions of the full face image samples and the target face image samples are different, the tasks of both are training the face recognition model. Incremental learning refers to a learning method that can continuously learn new knowledge from new samples and save most of the knowledge that has been learned. For example, in this embodiment, after being trained by using full face image samples in the source domain, the initialization model is further trained using target face image samples in the target domain to obtain a face recognition model which cannot only maintain the recognition capability of the face images in the source domain, but also can improve the recognition ability of the target face images in the target domain.

In the above step S101, the electronic device can acquire source domain face features and an initialization recognition model, wherein the initialization recognition model is obtained based on the training of full face image samples in the source domain. The full face image samples can be CASIA, VGGFace2 or MS1MV2, etc., which is not limited here. Identity information of the full face image samples in the source domain is known, that is, identity labels of the full face image samples are known. When training the initialization model, after inputting the full face image samples in the source domain into the initialization model, the initialization model can obtain prediction labels of the full face image samples. Based on the prediction labels and the identity labels of the full face image samples, the electronic device can calculate a classification loss according to the following formula:

$$L_c = \frac{1}{N} \sum_i -\log \frac{e^{s(\cos(\theta_{y_i,i})-m)}}{e^{s(\cos(\theta_{y_i,i})-m)} + \sum_{j \neq y_i} e^{s(\cos(\theta_{j,i}))}}$$

wherein L_c is a cross-entropy loss used, m is the size of a margin, s is the size of a scale value, θ is an angle between the weight and the feature, i is an index of an input sample, y_i is a label corresponding to the input sample with index i , and N is the number of the samples.

Furthermore, based on the classification loss calculated by the above formula, the electronic device can reduce the classification loss continuously by adjusting model parameters of the face recognition model, until the number of iterations of the full face image samples in the source domain reaches a preset number, and determine that the initialization model converges, so as to obtain the initialization recognition model. Of course, it is also reasonable to determine the convergence of the initialization model based on the convergence of the loss function of the initialization model so as to obtain the initialization recognition model. In this way, the trained initialization recognition model has the ability to recognize face images in the source domain, that is, it has the ability to recognize the faces of ordinary people.

In order for the face recognition model to maintain the ability to recognize face images in the source domain after incremental training, the face features of the full face image samples in the source domain can be extracted, and the initialization recognition model can be trained by using the source domain face features during the incremental training, so that the initialization recognition model can better maintain the recognition capability of face images in the source domain. In one implementation, the electronic device can store only a small amount of source domain face features, for example, when there is only a small amount of storage space, the electronic device can input a randomly selected portion of the full face image samples into the initialization recognition model to acquire face features output by the initialization recognition model as source domain face features.

The initialization recognition model obtained through initialization training already has the ability to recognize face images in the source domain. In order to improve the recognition capability of face images in the target domain, the face images in the target domain can be used to perform incremental training on the initialization recognition model. After the electronic device acquires the source domain face features and the initialization recognition model, it can acquire the target face image samples in the target domain, that is, execute the above step S102.

The target face image samples in the target domain can be collected by the electronic device,

or can be input into the electronic device by an external device, which is not limited here. Identity labels corresponding to the target face image samples are unknown, that is, specific identity information corresponding to the target face image samples in the target domain cannot be determined. For example, the target face image samples in the target domain obtained by the electronic device are face images of multiple employees in an industrial park. Due to privacy protection, real identities of these face images cannot be determined. The electronic device can cluster these face images, and record the label of each class of face images as “Employee A”, “Employee B”, etc., so as to determine the identities of the employee during the face recognition.

Furthermore, in the above step S103, the electronic device can adjust a portion of model parameters of the initialization recognition model based on the target face image samples and the source domain face features until the initialization recognition model converges so as to obtain the face recognition model for the source domain and the target domain.

In order to enable the initialization recognition model to maintain the recognition capability of face images in the source domain and improve the recognition capability of target face images in the target domain, the initialization recognition model can include a fixed portion of parameters and an adjustable portion of the parameters. The model parameters of the fixed portion of parameters remain unchanged, so that the recognition capability of the initialization recognition model for face images in the source domain can be maintained.

Furthermore, the electronic device can input the target face image samples in the target domain and the source domain face features into the initialization recognition model, and train the initialization recognition model. The electronic device can adjust a portion of model parameters of the initialization recognition model, that is, the model parameters of the adjustable portion of the parameters can be adjusted, until the initial recognition model converges so as to obtain the face recognition model for the source domain and the target domain. The face recognition model for the source domain and the target domain then has the capability to recognize the target face images in the target domain.

By employing the solution provided by the embodiment of the present application, the electronic device can train the face recognition model by using full face image samples in the source domain, and the resulting initialization recognition model has the capability to recognize face images in the source domain. During training the initialization recognition model by using the

target face image samples in the target domain, by fixing a portion of parameters of the initialization recognition model and training the initialization recognition model again by using the source domain face features, the resulting face recognition model for the source domain and the target domain cannot only maintain the capability to recognize face images in the source domain, but also recognize the target face images in the target domain accurately. In the case that the face recognition model cannot be trained with face data of the source domain and the target domain simultaneously, the face recognition model for the source domain and the target domain can be obtained by using only a small amount of source domain face features. The trained face recognition model cannot only maintain the performance of the source domain, but also improve its performance in the target domain, which can recognize the face images accurately, thereby improving the recognition capability and accuracy of the face recognition model.

As an implementation of the embodiments of the present application, as shown in Figure 2, acquiring the source domain face features can include:

S201, screening the full face image samples based on a preset screening strategy to obtain screened full face image samples,

wherein the preset screening strategy can ensure that the number of pieces of identity information corresponding to the screened full face image samples is not less than a preset number when the number of the screened full face image samples remains unchanged.

During the initial training, the data size of the full face image samples in the source domain used to train the face recognition model is usually very large, which is not conducive to storage. Furthermore, in the step of adjusting a portion of model parameters of the initialization recognition model based on the target face image samples and the source domain face features, it is not necessary to use source domain face features corresponding to all of the full face image samples in the source domain, and the recognition capability of the initialization recognition model for face images in the source domain can also be maintained. Therefore, in order to reduce the storage space required for data storage and the amount of calculation required for face recognition model training, the electronic device can screen the full face image samples in the source domain.

For example, for the full face image samples in the source domain corresponding to each piece of identity information, it is only necessary to extract source domain face features contained in three full face image samples to maintain the recognition capability of the initialization recognition model

for face images in the source domain. Then three full face image samples corresponding to each piece of identity information can be screened for subsequent extraction of source domain face features.

The electronic device can screen the full face image samples based on the preset screening strategy to obtain screened full face image samples. In the case that the size of the storage space remains unchanged, that is, the number of the screened full face image samples remains unchanged, the higher the intra-class richness of the full face image samples, the better the initialization recognition model can maintain its capability to recognize face images in the source domain. That is to say, it is possible to obtain full face image samples corresponding to multiple identity information as many as possible, and the full face image samples of different states corresponding to each piece of identity information does not need to be too many, that is, the intra-class richness can be appropriately reduced, so as to better maintain the recognition capability of the target domain face recognition model for face images in the source domain.

Therefore, the above-mentioned preset screening strategy can be a strategy that ensures that the number of pieces of identity information corresponding to the screened full face image samples is not less than a preset number when the number of the screened full face image samples remains unchanged, wherein the preset number can be set according to requirements such as storage space and calculation amount, which is not specifically limited here.

In one implementation, the electronic device can randomly select full face image samples corresponding to a certain amount of identity information from all of the full face image samples, and then randomly select a certain number of full face image samples from the full face image samples corresponding to each piece of identity information. For example, the electronic device can first randomly select full face image samples corresponding to 1,000 piece of identity information, and then randomly select 10 full face image samples from full face image samples corresponding to each piece of identity information.

In another implementation, for the full face image samples corresponding to each piece of identity information, the electronic device can select full face image samples based on the distance between the face features of the respective full face image samples and a feature center, such as select a certain number of the full face image samples that are closest to the feature center. Wherein, the feature center is a classifier vector corresponding to the identity information to which these full

face image samples belong. The closer the distance to the feature center is, the higher the accuracy of the real face characteristics of a person whose identity information the face feature identifier of full face image samples belongs to. Therefore, by employing such full face image samples in subsequent feature extraction, it is more beneficial to the recognition capability of the trained face recognition model.

S202, inputting the screened full face image samples into the initialization recognition model to obtain face features output by an intermediate layer of the initialization recognition model.

Once the screened full face image samples are determined, the electronic device can input the screened full face image samples into the initialization recognition model to obtain face features output by the intermediate layer of the initialization recognition model. Wherein, the intermediate layer of the initialization recognition model is neither the first layer nor the last layer in the initialization recognition model. The feature extraction operation can be represented by a formula: $F_i=f(x_i)$, wherein F_i is an extracted face feature, x_i is a full face image sample input into the initialization recognition model, and $f(x)$ is the function on which feature extraction of the initialization recognition model based.

As an implementation, the initialization recognition model can include multiple layers, for example, the initialization recognition model is a residual network including four residual blocks, which during the incremental training of the initialization training, model parameters of the first three residual blocks can be fixed, and only model parameter of the fourth residual block is adjusted. Then, the electronic device can obtain face features output by the third residual block of the initialization recognition model.

S203, determining the source domain face features based on the face features output by the intermediate layer.

After obtaining the face features output by the intermediate layer of the initialization recognition model, the electronic device can determine the source domain face features. In one implementation, the electronic device can take the face features output by the intermediate layer as the source domain face features. In another implementation, the electronic device can perform dimension reduction processing on the face features output by the intermediate layer to obtain dimension reduced face features as source domain face features for easy storage.

In this embodiment, the electronic device can screen the full face image samples and determine

the source domain face features based on the screened full face image samples. The source domain face features is conducive to maintaining the recognition capability of the target domain face recognition model for face images in the source domain. By screening the full face image samples in the source domain, the data storage space and calculation amount required for face recognition model training are reduced without reducing the recognition capability of the face recognition model for face images in the source domain.

As an implementation of the embodiments of the present application, the above-mentioned initialization recognition model can include a fixed parameter portion and a to-be-trained portion.

In order to maintain the recognition capability of the face recognition model for face images in the source domain, after the initialization recognition model is obtained based on the training of full face image samples in the source domain, the initialization recognition model can be divided into a fixed parameter portion and a to-be-trained portion. During training the initialization recognition model by using the target face image samples and the source domain face features, the model parameter of the fixed parameter portion is no longer adjusted, so as to maintain the recognition capability of the face recognition model for face images in the source domain. The model parameter of the to-be-trained portion is adjusted, so that the trained face recognition model has a strong recognition for face images in the target domain.

In one implementation, the source domain face features are face features output by the intermediate layer of the initialization recognition model. Then, the fixed parameter portion of the initialization recognition model can be consistent with the intermediate layer outputting the source domain face features, that is, a model portion that processes the source domain face features. For example, the initialization recognition model is a residual network including four residual blocks, and the source domain face features is output by a third residual block. Then, the fixed parameter portion of the initialization recognition model can include the first three residual blocks, and the to-be-trained portion can include the fourth residual block and a classifier.

Correspondingly, as shown in Figure 3, adjusting a portion of model parameters of the initialization recognition model based on the target face image sample and the source domain face features can include:

S301, inputting the target face image samples into the fixed parameter portion and the to-be-trained portion to obtain first prediction labels, and determining a first classification loss

based on the first prediction labels and pseudo labels corresponding to the target face image samples.

Based on the target face image samples and the source domain face features, the initialization recognition model is incrementally trained, specifically, trained by using a teacher-student network, wherein the student network (that is, the face recognition model) can use a small amount of stored features (that is, the above source domain face features) to train the teacher network (that is, the initialization recognition model) incrementally, so that the student network can obtain performance similar to that of the teacher network, that is, to maintain the capability to recognize face images in the source domain. During the incremental training, the student network is further trained using target face image samples in the target domain, so that it can have the capability to recognize target face images in the target domain.

Since the identity labels corresponding to the target face image samples in the target domain are unknown, the electronic device can determine a pseudo label corresponding to each target face image sample after acquiring the target face image samples in the target domain. The pseudo label is used to identify an identity, but not the true identity, of a person to whom the corresponding target face image sample belongs. For example, the pseudo label may be A, B, C, or 11, 12, 13, etc., which is not specifically limited here.

In order to improve the recognition capability of the face recognition model for face images in the target domain, the electronic device can input each target face image sample into the fixed parameter portion and the to-be-trained portion, to obtain the first prediction labels, and then based on the difference between the pseudo labels corresponding to the target face image samples and the first prediction labels, the electronic device can calculate a classification loss corresponding to the target face image samples, that is, a first classification loss. This first classification loss can characterize the difference between the recognition result of the current face recognition model for face images in the target domain and the real result.

S302, inputting the source domain face features into the to-be-trained portion to obtain second prediction labels, and determining a second classification loss based on the second prediction labels and identity labels corresponding to the source domain face features.

In order to maintain the recognition capability of the face recognition model for face images in the source domain, the electronic device can input the source domain face features into the

to-be-trained portion, to obtain the second prediction labels output by the to-be-trained portion. Since the identity labels of full face image samples corresponding to the source domain face features are known, the electronic device can calculate a classification loss corresponding to the source domain face features, that is, the second classification loss, based on the difference between the second prediction labels and the identity labels corresponding to the source domain face features. This second classification loss can characterize the difference between the recognition result of the current face recognition model for the face images in the source domain and the real result.

S303, inputting the source domain face features into the to-be-trained portion and an initial portion corresponding to the to-be-trained portion respectively to obtain estimated features and initial features, and determining a constraint loss based on the estimated features and the initial features;

wherein the initial portion is a corresponding model portion when a model parameter of the to-be-trained portion is fixed to a model parameter trained based on the full face image samples.

The electronic device can fix the model parameter of the to-be-trained portion, and the fixed model parameter is the model parameter of the to-be-trained portion of the initialization recognition model trained based on the full face image samples, after the model parameter of the to-be-trained portion of the initialization recognition model is fixed, the portion is the initial portion. For example, the initialization recognition model is a residual network including four residual blocks. The model parameters of the first three residual blocks are fixed, and only the model parameter of the fourth residual block is adjusted, the to-be-trained portion includes the fourth residual block, when the model parameter of the fourth residual block is fixed to the parameter of the fourth residual block corresponding to the initialization model, which is the initial portion.

The electronic device can input the source domain face features into the to-be-trained portion, and the to-be-trained portion can further extract the source domain face features based on the current model parameter, to obtain the estimated features. The source domain face features are input into the initial portion corresponding to the to-be-trained portion, and the initial portion can further extract the source domain face features based on the fixed model parameter, to obtain the initial features.

Furthermore, the electronic device can calculate the constraint loss based on the difference between the estimated features and the initial features. This constraint loss can characterize the

difference between the face features extracted by the to-be-trained portion and the face features extracted by the corresponding portion of the initialization model whose model parameter is fixed, and which can be used as a constraint to supervise the training of the face recognition model. The execution sequence of the above steps S301 to S303 is not specifically limited.

S304, adjusting the model parameter of the to-be-trained portion based on the first classification loss, the second classification loss and the constraint loss.

After obtaining the first classification loss, the second classification loss and the constraint loss above, the electronic device can adjust the model parameter of the to-be-trained portion based on the first classification loss, the second classification loss and the constraint loss, so as to train the initialization recognition model until the number of iterations of the target face image samples and source domain face features reaches the preset number, and it is determined that the initialization identification model converges.

In another implementation, the loss function value of the face recognition model can be calculated based on the first classification loss, the second classification loss and the constraint loss, and the model parameter of the to-be-trained portion is adjusted based on the loss function value until the loss function of the face recognition model converges, and it is determined that the initialization identification model converges.

Wherein, the specific way of adjusting the model parameter can be a gradient descent algorithm, a stochastic gradient descent algorithm, etc., which is not specifically limited and explained here.

The first classification loss can characterize the difference between the recognition result of the current face recognition model for face images in the target domain and the real result, the second classification loss can characterize the difference between the recognition result of the current face recognition model for face images in the source domain and the real result, and the constraint loss can characterize the difference between face features extracted by the to-be-trained portion and face features extracted by the corresponding portion of the initialization model whose model parameter is fixed. Therefore, adjusting the model parameter of the to-be-trained portion based on the first classification loss, the second classification loss and the constraint loss can make the difference between the recognition result of the face recognition model for face images in the target domain and the real result smaller and smaller, and maintain the accuracy of the recognition result for face

images in the source domain.

In this embodiment, the electronic device can train the initialization recognition model by using the source domain face features, so as to maintain the recognition capability of the face recognition model for face images in the source domain. The electronic device can train the initialization recognition model by using the target face image samples in the target domain, so as to improve the recognition capability of the face recognition model for face images in the target domain. By fixing a portion of the model parameters of the initialization recognition model, the recognition capability of the face recognition model for face images in the source domain is better maintained, and the recognition capability of the face recognition model for target face images in the target domain is improved.

With reference to Figure 4, the following is an example of the process of adjusting a portion of model parameters of the initialization recognition model based on the target face image samples and the source domain face features provided by an embodiment of the present application, wherein *conv* is a convolution layer of the initialization recognition model, *bn* represents batch normalization of data, *relu* and *tanh* are activation functions used in an activation function layer, *residual block* is a residual block of the initialization recognition model, the fixed parameter portion includes first three layers of residual blocks, and the to-be-trained portion includes the fourth layer of residual block and a classifier, the initial portion is the fourth layer residual block whose model parameter is fixed.

The electronic device can input the target face image samples into the fixed parameter portion and the to-be-trained portion, to obtain the first prediction labels, and determine the first classification loss based on the first prediction labels and the pseudo labels corresponding to the target face image samples.

The electronic device can perform a dimension restoration processing on the dimension reduced face features, to obtain the restored source domain face features, and input the restored source domain face features into the to-be-trained portion, to obtain the second prediction labels, and determine the second classification loss based on the second prediction labels and based on identity labels corresponding to the source domain face features.

The electronic device can input the restored source domain face features into the to-be-trained portion and the initial portion corresponding to the to-be-trained portion respectively, to obtain

estimated features and initial features, and determine the constraint loss based on the estimated features and the initial features.

Furthermore, the electronic device can obtain the loss function of the face recognition model based on the first classification loss, the second classification loss and the constraint loss, adjust the model parameter of the to-be-trained portion until the initialization recognition model converges, so as to obtain the face recognition model for the source domain and the target domain.

Since the specific implementations of each of the above processes have been described in each of the above embodiments, they will not be described again here. In this embodiment, by fixing a portion of the model parameters of the initialization recognition model and training the initialization recognition model again by using the source domain face features, the resulting face recognition model for the target domain maintains the capability to recognize face images in the source domain, and improves the capability to recognize face images in the target domain. In the case that the face recognition model cannot be trained by using face data of the source domain and the target domain simultaneously, the face images can be accurately recognized, thereby improving the recognition capability and accuracy of the face recognition model.

As an implementation of the embodiments of the present application, before inputting the target face image samples into the fixed parameter portion and the to-be-trained portion to obtain the first prediction labels, the above method can further include:

clustering the target face image samples and determining a pseudo label corresponding to each target face image sample.

Since identity labels corresponding to the target face image samples in the target domain are unknown, the initialization recognition model cannot be trained by using such target face image samples. Therefore, the electronic device can cluster the target face image samples in the target domain and determine the pseudo label corresponding to each target face image sample.

During the clustering process, the target face image samples can be classified according to their identities based on the similarity of face features, thereby obtaining multiple groups of target face image samples, wherein each group of target face image samples belongs to the same person, the electronic device can label each group of target face image samples with a pseudo label to identify the identity of the person to whom the corresponding target face image sample belongs.

In one implementation, a k-means++ (K-means) clustering algorithm can be used to cluster the target face image samples to obtain multiple categories, and determine a pseudo label of each target face image sample included in each of the categories. The electronic device can further cluster target face image samples with algorithms such as maximum expectation clustering of Gaussian mixture model, agglomerative hierarchical clustering, mean shift clustering, etc., which are not specifically limited here.

For example, the target face image samples in the target domain are face images of workers in a factory, the target face image samples are clustered and divided into multiple groups, and each group of target face image samples corresponds to face images of one worker. The electronic device can determine that the pseudo label of each group of target face image samples is A, B, C, or 11, 12, 13, etc., which are not limited here.

In this embodiment, the electronic device can cluster the target face image samples and determine a pseudo label corresponding to each target face image sample, so that when the true identity of the person to whom the target face image sample belongs cannot be known, an accurate pseudo label can also be determined to identify the identity of the person to whom the corresponding target face image sample belongs.

As an implementation of the embodiments of the present application, adjusting the model parameter of the to-be-trained portion based on the first classification loss, the second classification loss and the constraint loss can include:

calculating a loss function value L based on the first classification loss, the second classification loss and the constraint loss, according to the following formula,

$$L = L_{c1} + L_{c2} + \lambda L_{kd}$$

and adjusting the model parameter of the to-be-trained portion based on the loss function value.

Wherein, L_{c1} is the above first classification loss, L_{c2} is the above second classification loss, L_{kd} is the above constraint loss, and λ is a preset parameter.

By summing the first classification loss, the second classification loss and the constraint loss of the preset parameter, the obtained loss function value can accurately characterize the difference between the recognition result of the face images in the target domain and the real result, the

difference between the recognition result of the face images in the source domain and the real result and the difference between the face features extracted by the to-be-trained portion and the face features extracted by the corresponding portion of the initialization model whose model parameter is fixed. Therefore, the electronic device can use the above formula to calculate the loss function value L , and adjust the model parameter of the to-be-trained portion based on the loss function value L , to obtain a face recognition model with strong recognition capability.

Wherein, the value of the preset parameter λ can be set according to the change of the loss function value during the training process combined with actual experience, which is not specifically limited here.

In this embodiment, the electronic device can calculate the loss function value according to the formula based on the first classification loss, the second classification loss and the constraint loss, and adjust the model parameter of the to-be-trained portion based on the loss function value, so as to cause the initialization recognition model to converge. Through the above formula, the electronic device can accurately calculate the loss function value, thereby enhancing the training effect of the initialization recognition model, improving the recognition effect of the face recognition model for face images in the source domain, and making the face recognition model have better performance in recognizing face images in the target domain.

As an implementation of the embodiments of the present application, determining the constraint loss based on the estimated features and the initial features can include:

calculating the constraint loss L_{kd} based on the estimated features and the initial features, according to the following formula:

$$L_{kd} = \frac{1}{n} \sum_{i=1}^n (F_i - \bar{F}_i)^2$$

wherein, n is the number of the source domain face features, F_i is an initial feature corresponding to i -th source domain face feature, and $\bar{F}_i$ is an estimated feature corresponding to i -th source domain face feature.

By calculating the difference between an initial feature corresponding to each source domain face feature and an estimated feature corresponding to the source domain face feature, the difference between each face feature extracted by the to-be-trained portion and the face feature

extracted by the corresponding portion of the initialization model whose model parameter is fixed can be obtained. Furthermore, by calculating the variance of the difference values corresponding to the respective source domain face features, the obtained constraint loss can accurately characterize the degree of difference between the estimated features and the initial features. Therefore, the electronic device can use the above formula to calculate the constraint loss so as to ensure that an accurate loss function value can be calculated.

In this embodiment, the electronic device can calculate the constraint loss of each source domain face feature after model training, and thus calculate the constraint loss of the initialization recognition module. Through the above formula, the electronic device can accurately calculate the constraint loss and use the constraint loss as supervision to adjust the model parameter of the to-be-trained portion, thereby obtaining a more accurate face recognition model.

As an implementation of the embodiments of the present application, determining the source domain face features based on the face features output by the intermediate layer can include:

performing a dimension reduction processing on the face features output by the intermediate layer to obtain dimension reduced face features as the source domain face features.

Since the source domain face features are output by the intermediate layer of the initialization recognition model, the feature dimension is high, resulting in a large storage space required for storing the source domain face features. In order to save the storage space, the electronic device can perform a dimension reduction processing on the face features output by the intermediate layer, thereby significantly reducing the storage space required for storing the source domain face features while basically maintaining the amount of feature information.

For example, a PCA (Principal Component Analysis) dimension reduction method and the like can be used to perform a dimension reduction processing on the face features output by the intermediate layer, and the dimension reduced face features can be obtained as source domain face features. The core operation of the PCA dimension reduction is SVD (Singular Value Decomposition). SVD can be formulated as: $AA^T = U\Sigma^2U^T$, wherein A is a matrix to be decomposed, A^T is a transposed matrix of A, U is a left singular matrix of A, U^T is a transposed matrix of U, and Σ is a diagonal matrix containing the corresponding eigenvalues. The PCA dimension reduction method can delete certain relevant dimensions in the original data, and maximize the amount of information carried by the data while reducing the dimension of the data.

In one implementation, the process of obtaining source domain face features from full face image samples in the source domain and performing a dimension reduction processing can be shown in Figure 5, wherein the initialization recognition model can be a convolutional neural network, *conv* is a convolution layer of the convolutional neural network, *bn* represents batch normalization to the data, *relu* and *tanh* are activation functions used in an activation function layer, and *residual block* is a residual block of the convolutional neural network. The fixed parameter portion includes first three layers of residual blocks. The module parameter of the fourth layer of residual block can be changed to training the initialization model. After inputting the full face image samples in the source domain into the convolutional neural network, the electronic device can extract face features output by the intermediate layer of the fixed parameter portion, and then perform a dimension reduction processing to obtain the dimension reduced face features.

Correspondingly, before adjusting a portion of model parameters of the initialization recognition model based on the target face image samples and the source domain face features, the method can further include:

performing a dimension restoration processing on the source domain face features to obtain restored source domain face features.

In order to maintain the recognition capability of the face recognition model for face images in the target domain as much as possible, since the source domain face features are obtained through dimension reduction, the electronic device can perform a dimension restoration processing on the source domain face features to obtain the restored source domain face features before adjusting a portion of model parameters of the initialization recognition model based on the target face image samples and the source domain face features. The dimension of the restored source domain face features is the same as that of the face features output by the intermediate layer of the fixed parameter portion.

In this embodiment, the electronic device can perform a dimension reduction processing on the face features output by the intermediate layer, and perform a dimension restoration processing on the dimension reduced source domain face features before adjusting a portion of model parameters of the initialization recognition model based on the target face image samples and the source domain face features. As a result, the storage space required for storing the source domain face features can be significantly reduced without affecting the recognition capability of the face

recognition model for face images in the target domain.

As an implementation of the embodiments of the present application, as shown in Figure 6, the above method can further include:

S601, acquiring a face image to be recognized in the target domain.

After training the initialization recognition model and obtaining the face recognition model for the target domain, the face recognition model can recognize face images in the target domain while maintaining the capability to recognize face images in the source domain. The electronic device can deploy the face recognition model in an actual application scenario, that is, a target domain scenario, and then acquire a face image to be recognized in the target domain.

For example, when the face recognition model is used for face recognition at the entrance and exit gates of a park, an original model in the gate can be replaced by the face recognition model, this face recognition model has better performance in the face recognition of park personnel and basically maintains the capability to recognize a face of non-park personnel. Then, when a person wants to enter or exit the gate, the electronic device can collect a face image of the person as the face image to be recognized in the target domain.

S602, identifying the face image to be recognized based on the face recognition model to determine an identity corresponding to the face image to be recognized.

Furthermore, the electronic device can identify the face image to be recognized based on the face recognition model, and determine the identity corresponding to the face image to be recognized. After identifying the face image to be recognized, the electronic device can determine the identity identifier corresponding to the face image to be recognized, and perform different operations based on different identifiers. For example, the electronic device can control the gate to open or keep closed based on the identity identifier corresponding to the face image to be recognized.

For example, during training, a pseudo label corresponding to the target face image sample is “person A”, “person B”, etc. After the electronic device recognizes that the identity corresponding to the face image to be recognized is “person B”, it can be determined that the person is a person in the industrial park who has an access authority, then the electronic device can control the gate to open. If it is recognized that the identity corresponding to the face image to be recognized is XX Liu, it can be determined that the person is not a person in the industrial park, but someone outside

the park who has no access authority, then the electronic device can control the gate to remain closed.

In this embodiment, the electronic device can recognize the target face image in the target domain while maintaining the capability to recognize the face image in the source domain. Furthermore, whether it is a person corresponding to the source domain or the target domain, the face image to be recognized can be recognized based on the face recognition model, and the identity corresponding to the face image to be recognized can be accurately determined.

Corresponding to the above training method for a face recognition model, an embodiment of the present application further provides a training apparatus for the face recognition model. The training apparatus for the face recognition model provided by the embodiment of the present application is described below.

As shown in Figure 7, the training apparatus for the face recognition model includes:

an initialization training module 701, configured for acquiring source domain face features and an initialization recognition model,

wherein the initialization recognition model is obtained based on the training of full face image samples in the source domain, and the source domain face features are face feature of the full face image samples obtained through the initialization recognition model;

an acquisition module for target domain samples 702, configured for acquiring target face image samples in a target domain,

wherein identity labels corresponding to the target face image samples are unknown.

An incremental training module 703, configured for adjusting a portion of model parameters of the initialization recognition model based on the target face image samples and the source domain face features, until the initialization recognition model converges, so as to obtain the face recognition model for the source domain and the target domain.

It can be seen that in the solution provided by the embodiments of the present application, the electronic device can acquire source domain face features and an initialization recognition model, wherein the initialization recognition model is obtained based on the training of full face image samples in a source domain, and source domain face features are face feature of the full face image samples obtained through the initialization recognition model; acquire target face image samples in

a target domain, wherein identity labels corresponding to the target face image samples are unknown; adjust a portion of model parameters of the initialization recognition model based on the target face image samples and the source domain face features, until the initialization recognition model converges, so as to obtain the face recognition model for the source domain and the target domain. After the initialization model is trained by using the full face image samples of the source domain, a portion of the source domain face features are stored, and a portion of the parameters of the initialization model are fixed. Furthermore, after the initialization model is further trained by using the target face image samples in the target domain and the source domain face features, the face recognition model for the source domain and the target domain is obtained. This face recognition model not only maintains the recognition capability of the full face images in the source domain, but also can accurately recognize target face images in the target domain. The face images can be recognized accurately when the face recognition model cannot be trained by using face data of the source domain and the target domain simultaneously, thereby improving the recognition capability and accuracy of the face recognition model.

As an implementation of the embodiments of the present application, the above initialization training module 701 can include:

a sample screening unit, configured for screening the full face image samples based on a preset screening strategy to obtain screened full face image samples,

wherein the preset screening strategy ensures that the number of pieces of identity information corresponding to the screened full face image samples is not less than a preset number when the number of the screened full face image samples remains unchanged;

a feature acquisition unit, configured for inputting the screened full face image samples into the initialization recognition model to obtain face features output by an intermediate layer of the initialization recognition model; and

a feature determination unit, configured for determining the source domain face features based on the face features output by the intermediate layer.

As an implementation of the embodiments of the present application, the above initialization recognition model includes a fixed parameter portion and a to-be-trained portion.

As shown in Figure 8, the above incremental training module 703 can include:

a first input unit 801, configured for inputting the target face image samples into the fixed parameter portion and the to-be-trained portion to obtain first prediction labels, and determining a first classification loss based on the first prediction labels and based on pseudo labels corresponding to the target face image samples;

a second input unit 802, configured for inputting the source domain face features into the to-be-trained portion to obtain second prediction labels, and determining a second classification loss based on the second prediction labels and based on identity labels corresponding to the source domain face features;

a third input unit 803, configured for inputting the source domain face features into the to-be-trained portion and an initial portion corresponding to the to-be-trained portion respectively, to obtain estimated features and initial features, and determining a constraint loss based on the estimated features and the initial features;

wherein the initial portion is a corresponding model portion when a model parameter of the to-be-trained portion is fixed to a model parameter trained based on the full face image samples; and

a parameter adjustment unit 804, configured for adjusting the model parameter of the to-be-trained portion based on the first classification loss, the second classification loss and the constraint loss.

As an implementation of the embodiments of the present application, the above apparatus can further include:

a clustering module for target face image samples, configured for clustering the target face image samples and determining a pseudo label corresponding to each target face image sample before inputting the target face image samples into the fixed parameter portion and the to-be-trained portion to obtain the first prediction labels,

wherein the pseudo label is used to identify an identity of the person to whom the corresponding target face image sample belongs.

As an implementation of the embodiments of the present application, the above parameter adjustment unit 804 can include:

a loss function value calculation sub-unit, configured for calculating a loss function value L

based on the first classification loss, the second classification loss and the constraint loss, according to the following formula:

$$L = L_{c1} + L_{c2} + \lambda L_{kd}$$

wherein L_{c1} is the first classification loss, L_{c2} is the second classification loss, L_{kd} is the constraint loss, and λ is a preset parameter; and

a parameter adjustment sub-unit, configured for adjusting the model parameter of the to-be-trained portion based on the loss function value.

As an implementation of the embodiments of the present application, the above third input unit 803 can include:

a constraint loss calculation sub-unit, configured for calculating the constraint loss L_{kd} based on the estimated features and the initial features, according to the following formula:

$$L_{kd} = \frac{1}{n} \sum_{i=1}^n (F_i - \bar{F}_i)^2$$

wherein, n is the number of the source domain face features, F_i is an initial feature corresponding to i -th source domain face feature, and $\bar{F}_i$ is an estimated feature corresponding to i -th source domain face feature.

As an implementation of the embodiment of the present application, the above feature determination unit can include:

a feature dimension reduction sub-unit, configured for performing a dimension reduction processing on the face features output by the intermediate layer to obtain dimension reduced face features as the source domain face features.

The above apparatus can further include:

a feature restoration module, configured for performing a dimension restoration processing on the source domain face features to obtain restored source domain face features before adjusting a portion of model parameters of the initialization recognition model based on the target face image samples and the source domain face features.

As an implementation of the embodiments of the present application, the above apparatus can

further include:

an acquisition module of a face image to be recognized, configured for acquiring a face image to be recognized in the target domain; and

an identity determination module, configured for identifying the face image to be recognized based on the face recognition model to determine an identity corresponding to the face image to be recognized.

An embodiment of the present application further provides an electronic device, as shown in Figure 9, which includes a processor 901, a communication interface 902, a memory 903 and a communication bus 904, wherein the processor 901, the communication interface 902, and the memory 903 communicate with each other over the communication bus 904,

the memory 903, is configured for storing a computer program;

the processor 901, is configured for implementing the method steps described in any one of the above embodiments when executing the program stored in the memory 903.

The communication bus mentioned by the above electronic device can be a Peripheral Component Interconnect (PCI) bus or an Extended Industry Standard Architecture (EISA) bus, etc.. The communication bus can be divided into an address bus, a data bus, a control bus, etc.. For ease of presentation, only one thick line is used in the figure, but it does not mean that there is only one bus or one type of bus.

The communication interface is used for communication between the above electronic device and other devices.

The memory can include a Random Access Memory (RAM) or a Non-Volatile Memory (NVM), such as at least one disk memory. Optionally, the memory can also be at least one storage device located far away from the aforementioned processor.

The above processor can be a general-purpose processor, including a Central Processing Unit (CPU), a Network Processor (NP), etc.. It can also be a Digital Signal Processor (DSP), an Application Specific Integrated Circuit (ASIC), a Field-Programmable Gate Array (FPGA) or other programmable logic devices, discrete gate or transistor logic devices, discrete hardware

components.

In yet another embodiment provided by the present application, a computer-readable storage medium is further provided, which stores a computer program, that when executed by a processor, implements the method steps described in any one of the above embodiments.

In yet another embodiment provided by the present application, a computer program product containing instructions is further provided, which, when run on a computer, cause the computer to carry out the method steps described in any one of the above embodiments.

In the above embodiments, it can be implemented in whole or in part by software, hardware, firmware, or any combination thereof. When implemented by software, it can be implemented in whole or in part in the form of a computer program product. The computer program product includes one or more computer instructions. When the computer program instructions are loaded and executed on a computer, the processes or functions described in the embodiments of the present application are generated in whole or in part. The computer can be a general-purpose computer, a special-purpose computer, a computer network, or other programmable device. The computer instructions can be stored in a computer-readable storage medium or transmitted from one computer-readable storage medium to another computer-readable storage medium, for example, the computer instructions can be transmitted from a website site, computer, server, or data center to another website site, computer, server or data center over wired (such as coaxial cable, optical fiber, digital subscriber line (DSL)) or wirelessly (such as infrared, wireless, microwave, etc.). The computer-readable storage medium can be any available media that can be accessed by a computer or a data storage device such as a server, data center, etc. that includes an integration of one or more available media. The available media can be a magnetic media (e.g., floppy disk, hard disk, magnetic tape), optical media (e.g., DVD), or semiconductor media (e.g., Solid State Disk (SSD)), etc.

It should be noted that relational terms such as first and second and the like herein are used merely to distinguish one entity or operation from another, and do not necessarily require or imply any such actual relationship or order between these entities or operations. Furthermore, the terms “include”, “comprise” or any other variation thereof are intended to encompass non-exclusive inclusion, so that a process, method, article or device containing a set of elements includes not only those elements, but also other elements not expressly listed, or other elements inherent to such

process, method, article or device. Without further limitation, elements defined by the phrase “including one ...” does not preclude the presence of additional identical elements in a process, method, article, or device including said element.

The various embodiments in this specification are described in a related manner. The same and similar parts between the various embodiments can be referred to each other, and each embodiment focuses on the differences from other embodiments. In particular, for the embodiments of the apparatus, electronic device, computer-readable storage medium, and computer program product, the description is relatively simply since it is basically similar to the embodiment of the method, and the relevant points can be referred to the partial description of the method embodiments.

The above embodiments are only preferred embodiments of the present application and are not intended to limit the present application. Any modification, equivalent replacement, improvement, etc. made within the spirit and principles of the present application shall be included in the protection scope of the present application.

CLAIMS

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

acquiring source domain face features and an initialization recognition model, wherein the initialization recognition model is obtained based on the training of full face image samples in a source domain, and the source domain face features are face features of the full face image samples obtained through the initialization recognition model;

acquiring target face image samples in a target domain, wherein identity labels corresponding to the target face image samples are unknown; and

adjusting a portion of model parameters of the initialization recognition model based on the target face image samples and the source domain face features, until the initialization recognition model converges, so as to obtain the face recognition model for the source domain and the target domain.

2. The method according to claim 1, wherein acquiring the source domain face features comprises:

screening the full face image samples based on a preset screening strategy, to obtain screened full face image samples, wherein the preset screening strategy ensures that the number of pieces of identity information corresponding to the screened full face image samples is not less than a preset number when the number of the screened full face image samples remains unchanged;

inputting the screened full face image samples into the initialization recognition model, to obtain face features output by an intermediate layer of the initialization recognition model; and

determining the source domain face features based on the face features output by the intermediate layer.

3. The method according to claim 1, wherein the initialization recognition model comprises a fixed parameter portion and a to-be-trained portion;

wherein adjusting a portion of model parameters of the initialization recognition model based on the target face image samples and the source domain face features, comprises:

inputting the target face image samples into the fixed parameter portion and the to-be-trained portion, to obtain first prediction labels, and determining a first classification loss based on the first

prediction labels and pseudo labels corresponding to the target face image samples;

inputting the source domain face features into the to-be-trained portion, to obtain second prediction labels, and determining a second classification loss based on the second prediction labels and identity labels corresponding to the source domain face features;

inputting the source domain face features into the to-be-trained portion and an initial portion corresponding to the to-be-trained portion respectively, to obtain estimated features and initial features, and determining a constraint loss based on the estimated features and the initial features, wherein the initial portion is a corresponding model portion when a model parameter of the to-be-trained portion is fixed to a model parameter trained based on the full face image samples; and

adjusting the model parameter of the to-be-trained portion based on the first classification loss, the second classification loss and the constraint loss.

4. The method according to claim 3, wherein before inputting the target face image samples into the fixed parameter portion and the to-be-trained portion to obtain first prediction labels, the method further comprises:

clustering the target face image samples, and determining a pseudo label corresponding to each target face image sample, wherein the pseudo label is used to identify an identity of a person to whom the corresponding target face image sample belongs.

5. The method according to claim 3, wherein adjusting the model parameter of the to-be-trained portion based on the first classification loss, the second classification loss and the constraint loss, comprises:

calculating a loss function value L based on the first classification loss, the second classification loss and the constraint loss, according to the following formula:

$$L = L_{c1} + L_{c2} + \lambda L_{kd}$$

wherein, L_{c1} is the first classification loss, L_{c2} is the second classification loss, L_{kd} is the constraint loss, and λ is a preset parameter; and

adjusting the model parameter of the to-be-trained portion based on the loss function value.

6. The method according to claim 3, wherein determining a constraint loss based on the

estimated features and the initial features, comprises:

calculating the constraint loss L_{kd} based on the estimated features and the initial features, according to the following formula:

$$L_{kd} = \frac{1}{n} \sum_{i=1}^n (F_i - \bar{F}_i)^2$$

wherein, n is the number of the source domain face features, F_i is an initial feature corresponding to i -th source domain face feature, and $\bar{F}_i$ is an estimated feature corresponding to i -th source domain face feature.

7. The method according to any one of claims 2 to 6, wherein determining the source domain face features based on the face features output by the intermediate layer, comprises:

performing a dimension reduction processing on the face features output by the intermediate layer to obtain dimension reduced face features as the source domain face features;

before adjusting a portion of model parameters of the initialization recognition model based on the target face image samples and the source domain face features, the method further comprises:

performing a dimension restoration processing on the source domain face features to obtain restored source domain face features.

8. The method according to any one of claims 1 to 6, wherein the method further comprises:

acquiring a face image to be recognized in the target domain; and

identifying the face image to be recognized based on the face recognition model to determine an identity corresponding to the face image to be recognized.

9. A training apparatus for a face recognition model, comprising:

an initialization training module, configured for acquiring source domain face features and an initialization recognition model, wherein the initialization recognition model is obtained based on the training of full face image samples in a source domain, and the source domain face features are face features of the full face image samples obtained through the initialization recognition model;

an acquisition module for target domain samples, configured for acquiring target face image samples in a target domain, wherein identity labels corresponding to the target face image samples

are unknown; and

an incremental training module, configured for adjusting a portion of model parameters of the initialization recognition model based on the target face image samples and the source domain face features, until the initialization recognition model converges, so as to obtain the face recognition model for the source domain and the target domain.

10. A computer-readable storage medium storing a computer program thereon, when executed by a processor, the computer program causes the processor to implement the method steps according to any one of claims 1-8.

11. A computer program product containing instructions, when running on a computer, the computer program product causes the computer to carry out the method steps according to any one of claims 1-8.