

FEATURE EXTRACTION MODEL PROCESSING METHOD AND APPARATUS, AND FEATURE EXTRACTION METHOD AND APPARATUS

RELATED APPLICATION

[0001] This application claims priority to Chinese Patent Application No. 2023109319666, filed on July 27, 2023 and entitled "FEATURE EXTRACTION MODEL PROCESSING METHOD AND APPARATUS, AND FEATURE EXTRACTION METHOD AND APPARATUS", which is incorporated by reference in its entirety.

FIELD OF THE TECHNOLOGY

[0002] The present disclosure relates to the field of computer technologies, and in particular, to a feature extraction model processing method and apparatus, a computer device, a storage medium, and a computer program product, and a feature extraction method and apparatus, a computer device, a storage medium, and a computer program product.

BACKGROUND OF THE DISCLOSURE

[0003] With the development of computer technologies, a feature extraction technology appears. The feature extraction technology is widely applied to fields such as text translation, image recognition, and identity authentication. For example, feature extraction is performed on a palm image, and an extracted palm feature is used for identity authentication.

[0004] In the traditional feature extraction technology, a palm print feature is extracted from a palm image through a neural network. If a small quantity of palm print features are extracted, a feature different from the obtained palm print features is extracted from another palm image. To be specific, for the same regions in a plurality of palm images, a feature of only one region is extracted, and the same region in another palm image is ignored.

[0005] However, there may be a detail difference between the same regions in the plurality of palm images, and directly ignoring the same region in the another palm image may cause inaccurate extraction of detailed palm print features, thereby affecting subsequent related applications of the extracted palm print features.

SUMMARY

[0006] According to various embodiments provided in the present disclosure, a feature extraction model processing method and apparatus, a computer device, a computer-readable storage medium, and a computer program, and a feature extraction processing method and apparatus, a computer device, a computer-readable storage medium, and a computer program are provided.

[0007] In an aspect, the present disclosure provides a feature extraction model processing method, performed by a computer device and including:

[0008] obtaining a plurality of sample palm images of a sample palm;

[0009] determining a first sample image region and a second sample image region other than the first sample image region in each sample palm image, each first sample image region including a preset part of the sample palm;

[0010] obtaining a first sample embedding feature of each first sample image region;

[0011] mapping each first sample embedding feature into a key space to obtain a key vector of the first sample embedding feature;

[0012] obtaining a second sample embedding feature of each second sample image region;

[0013] mapping each second sample embedding feature into a query space to obtain a query vector of the second sample embedding feature;

[0014] performing self-attention prediction according to the key vectors of the first sample embedding features and the query vectors of the second sample embedding features, to obtain a predicted self-attention weight corresponding to the first sample embedding feature and a predicted self-attention weight corresponding to the second sample embedding feature; and

[0015] fusing the first sample embedding features with the second sample embedding feature according to the predicted self-attention weight of the first sample embedding feature and the predicted self-attention weight of the second sample embedding features, to obtain a predicted palm feature jointly corresponding to the plurality of sample palm images;

[0016] obtaining a label palm feature jointly corresponding to the plurality of sample palm images;

[0017] and performing model training based on a predicted difference between the predicted palm feature and the label palm feature, to obtain a feature extraction model.

[0018] In another aspect, the present disclosure further provides a feature extraction model processing apparatus, including:

[0019] a sample obtaining module, configured to: obtain a plurality of sample palm images of a sample palm, and determine a first sample image region and a second sample image region other than the first sample image region in each sample palm image, each first sample image region including a preset part of the sample palm;

[0020] a key space mapping module, configured to: obtain a first sample embedding feature of each first sample image region, and map each first sample embedding feature into a key space to obtain a key vector of the first sample embedding feature;

[0021] a query space mapping module, configured to: obtain a second sample embedding feature of each second sample image region, and map each second sample embedding feature into a query space to obtain a query vector of the second sample embedding feature;

[0022] a weight prediction module, configured to perform self-attention prediction according to the key vectors of the first sample embedding features and the query vectors of the second sample embedding features, to obtain a predicted self-attention weight corresponding to the first sample embedding features and a predicted self-attention weight corresponding to the second sample embedding features;

[0023] a sample feature fusion module, configured to fuse the first sample embedding features with the second sample embedding feature according to the predicted self-attention weight of the first sample embedding features and the predicted self-attention weight of the second sample embedding features, to obtain a predicted palm feature jointly corresponding to the plurality of sample palm images; and

[0024] a training module, configured to: obtain a label palm feature jointly corresponding to the plurality of sample palm images, and perform model training based on a predicted difference between the predicted palm feature and the label palm feature, to obtain a feature extraction model.

[0025] In another aspect, the present disclosure further provides a computer device, including a memory and a processor, the memory having a computer program stored therein, and the processor, when executing the computer program, executing the operations of the method embodiments of the present disclosure.

[0026] In another aspect, the present disclosure further provides a computer-readable storage medium, having a computer program stored therein, the computer program, when executed by a processor, executing the operations of the method embodiments of the present disclosure.

[0027] In another aspect, the present disclosure further provides a computer program product, including a computer program, the computer program, when executed by a processor, executing the operations of the method embodiments of the present disclosure.

[0028] In another aspect, the present disclosure provides a feature extraction method, performed by a computer device and including:

[0029] obtaining a plurality of palm images of a palm;

[0030] determining, through the feature extraction model obtained by training by the feature extraction model processing method, a first image region and a second image region other than the first image region in each palm image, each first image region including a preset part of the palm;

[0031] obtaining a first embedding feature of each first image region and a second embedding feature of each second image region through the feature extraction model; and

[0032] fusing each first embedding feature with each second embedding feature according to a self-attention weight for each first embedding feature and a self-attention weight for each second embedding feature in the feature extraction model through the feature extraction model, to obtain a palm feature jointly corresponding to the plurality of palm images, the self-attention weights being obtained by training on the feature extraction model.

[0033] In another aspect, the present disclosure further provides a feature extraction apparatus, including:

[0034] an image obtaining module, configured to obtain a plurality of palm images of a palm;

[0035] a region determining module, configured to determine, through the feature extraction model obtained by training by the feature extraction model processing apparatus, a first image region and a second image region other than the first image region in each palm image, each first image region including a preset part of the palm;

[0036] a feature obtaining module, configured to obtain a first embedding feature of each first image region and a second embedding feature of each second image region through the feature extraction model; and

[0037] a feature fusion module, configured to fuse each first embedding feature with each second embedding feature according to a self-attention weight for each first embedding feature and a self-attention weight for each second embedding feature in the feature extraction model through

the feature extraction model, to obtain a palm feature jointly corresponding to the plurality of palm images, the self-attention weights being obtained by training on the feature extraction model.

[0038] In another aspect, the present disclosure further provides a computer device, including a memory and a processor, the memory having a computer program stored therein, and the processor, when executing the computer program, executing the operations of the method embodiments of the present disclosure.

[0039] In another aspect, the present disclosure further provides a computer-readable storage medium, having a computer program stored therein, the computer program, when executed by a processor, executing the operations of the method embodiments of the present disclosure.

[0040] In another aspect, the present disclosure further provides a computer program product, including a computer program, the computer program, when executed by a processor, executing the operations of the method embodiments of the present disclosure.

[0041] Details of one or more embodiments of the present disclosure are provided in the accompanying drawings and descriptions below. Other features and advantages of the present disclosure become clear with reference to the specification, the accompanying drawings, and the claims.

BRIEF DESCRIPTION OF THE DRAWINGS

[0042] To describe the technical solutions in the embodiments of the present disclosure or the traditional art more clearly, the following briefly describes the accompanying drawings required for describing the embodiments or the traditional art. Apparently, the accompanying drawings in the following descriptions show merely the embodiments of the present disclosure. A person of ordinary skill in the art may still derive other drawings from the disclosed accompanying drawings without creative efforts.

[0043] FIG. 1 is a diagram of an application environment of a feature extraction model processing method and a feature extraction method in an embodiment.

[0044] FIG. 2 is a flowchart of a feature extraction model processing method in an embodiment.

[0045] FIG. 3 is a schematic diagram of two sample palm images in different photographing angles in an embodiment.

[0046] FIG. 4 is a flowchart of mapping a first sample embedding feature into a key space and a second sample embedding feature into a query space in an embodiment.

[0047] FIG. 5 is a schematic diagram of mapping a sample palm image into a vector space in an embodiment.

[0048] FIG. 6 is a flowchart of mapping a first sample embedding feature and a second sample embedding feature into a value space in an embodiment.

[0049] FIG. 7 is a flowchart of determining a first sample image region and a second sample image region other than the first sample image region in each sample palm image in an embodiment.

[0050] FIG. 8 is a flowchart of a feature extraction model processing method in an embodiment.

[0051] FIG. 9 is a flowchart of a feature extraction model processing method in an embodiment.

[0052] FIG. 10 is a flowchart of parallel training on a feature extraction model in an embodiment.

[0053] FIG. 11 is a flowchart of a feature extraction method in an embodiment.

[0054] FIG. 12 is a schematic scenario diagram of an application scenario of a feature extraction method in an embodiment.

[0055] FIG. 13 is a structural block diagram of a feature extraction model processing apparatus in an embodiment.

[0056] FIG. 14 is a structural block diagram of a feature extraction apparatus in an embodiment.

[0057] FIG. 15 is a diagram of an internal structure of a computer device in an embodiment.

DESCRIPTION OF EMBODIMENTS

[0058] The technical solutions in embodiments of the present disclosure are clearly and completely described in the following with reference to the accompanying drawings in the embodiments of the present disclosure. Apparently, the described embodiments are merely some rather than all of the embodiments of the present disclosure. All other embodiments obtained by a person of ordinary skill in the art based on the embodiments of the present disclosure without creative efforts shall fall within the protection scope of the present disclosure.

[0059] Artificial intelligence (AI) is a theory, method, technology, and application system that uses a digital computer or a machine controlled by the digital computer to simulate, extend, and

expand human intelligence, perceive an environment, acquire knowledge, and use knowledge to obtain an optimal result. In other words, AI is a comprehensive technology in computer science and attempts to understand the essence of intelligence and produce a new intelligent machine that can react in a manner similar to human intelligence. AI is to study the design principles and implementation methods of various intelligent machines, to enable the machines to have the functions of perception, reasoning, and decision-making.

[0060] The AI technology is a comprehensive discipline, and relates to a wide range of fields including both hardware-level technologies and software-level technologies. The basic AI technologies generally include technologies such as a sensor, a dedicated AI chip, cloud computing, distributed storage, a big data processing technology, an operating/interaction system, and electromechanical integration. AI software technologies mainly include several major directions such as a computer vision (CV) technology, a speech processing technology, a natural language processing technology, and machine learning/deep learning.

[0061] Machine learning (ML) is a multi-field interdisciplinary subject involving the probability theory, statistics, the approximation theory, convex analysis, the algorithm complexity theory, and the like. ML specializes in studying how a computer simulates or implements a human learning behavior to obtain new knowledge or skills, and reorganize an existing knowledge structure, so as to keep improving its performance. ML is the core of AI, is a basic way to make the computer intelligent, and is applied to various fields of AI. ML and deep learning generally include technologies such as an artificial neural network, a belief network, reinforcement learning, transfer learning, inductive learning, and learning from demonstrations.

[0062] Solutions provided in the embodiments of the present disclosure relate to technologies such as natural language processing and ML of AI, and are specifically described by using the following embodiments.

[0063] A feature extraction model processing method provided in an embodiment of the present disclosure may be applied to an application environment shown in FIG. 1. A terminal 102 communicates with a server 104 through a network. A data storage system may store data that needs to be processed by the server 104. The data storage system may be integrated on the server 104, or placed on a cloud or other servers. Both the terminal 102 and the server 104 may be separately configured to perform the feature extraction model processing method provided in this embodiment of the present disclosure. The terminal 102 and the server 104 may further be cooperatively configured to perform the feature extraction model processing method provided in this embodiment of the present disclosure. When the terminal 102 and the server 104 are cooperatively configured to

perform the feature extraction model processing method provided in this embodiment of the present disclosure, the terminal 102 obtains a plurality of sample palm images of a sample palm from the server 104, and determines a first sample image region and a second sample image region other than the first sample image region in each sample palm image. Each first sample image region includes a preset part of the sample palm. The terminal 102 obtains a first sample embedding feature of each first sample image region, and maps each first sample embedding feature into a key space to obtain a key vector of each first sample embedding feature. The terminal 102 obtains a second sample embedding feature of each second sample image region, and maps each second sample embedding feature into a query space to obtain a query vector of each second sample embedding feature. The terminal 102 performs self-attention prediction according to each key vector and each query vector, to obtain a predicted self-attention weight corresponding to each first sample embedding feature and a predicted self-attention weight corresponding to each second sample embedding feature. The terminal 102 fuses each first sample embedding feature with each second sample embedding feature according to the predicted self-attention weight of each first sample embedding feature and the predicted self-attention weight of each second sample embedding feature, to obtain a predicted palm feature jointly corresponding to the plurality of sample palm images; and obtains a label palm feature jointly corresponding to the plurality of sample palm images, and performs model training based on a predicted difference between the predicted palm feature and the label palm feature, to obtain a feature extraction model. The feature extraction model may be configured on the terminal 102 or the server 104.

[0064] The terminal 102 may be, but is not limited to, various personal computers, notebook computers, smart phones, tablet computers, smart voice interaction devices, smart appliances, in-vehicle terminals, aircraft, portable wearable devices, and the like. The server 104 may be an independent physical server, or may be a server cluster or a distributed system formed by a plurality of physical servers, or may be a server that provides basic cloud computing services such as cloud services, cloud databases, cloud computing, cloud functions, cloud storage, network services, cloud communication, middleware services, domain name services, security services, content delivery networks (CDNs), big data, and artificial intelligence platforms.

[0065] In an embodiment, a feature extraction method may alternatively be applied to the application environment shown in FIG. 1. Both the terminal 102 and the server 104 may be separately configured to perform the feature extraction method provided in this embodiment of the present disclosure. The terminal 102 and the server 104 may further be cooperatively configured to perform the feature extraction method provided in this embodiment of the present disclosure.

[0066] Here, "a plurality of" and another quantity mentioned in the embodiments of the present disclosure all mean "at least two".

[0067] In an embodiment, as shown in FIG. 2, a feature extraction model processing method is provided. Application of the method to the computer device (the computer device may be the terminal or the server in FIG. 1) in FIG. 1 is taken as an example for description, including the following operations:

[0068] Operation S202: Obtain a plurality of sample palm images of a sample palm, and determine a first sample image region and a second sample image region other than the first sample image region in each sample palm image, each first sample image region including a preset part of the sample palm.

[0069] The sample palm is a palm of a sample object. The sample object may be a person or an animal. The sample palm may be a palm, a sole of a foot, or the like of the person, or may be a front paw, a back paw, or the like of the animal.

[0070] The sample palm images are images obtained by photographing the sample palm. The sample palm may include fingers, a center of palm, and a palm edge. The fingers include a thumb, an index finger, a middle finger, a ring finger, a little finger, and the like. The palm edge is a part located around the center of the palm. The sample palm includes a preset part, and the preset part may be a center-of-palm part of the sample palm.

[0071] In this embodiment, the computer device may obtain a plurality of sample palm images of the same sample palm of the sample object, for example, obtain a plurality of sample palm images of a left palm of a person or obtain a plurality of sample palm images of a right palm of a person.

[0072] The computer device determines the preset part in each sample palm image, determines a region, in which the preset part is located, in the sample palm image, and uses the region as the first sample image region. For each sample palm image, the computer device uses a region other than the first sample image region in the sample palm image as the second sample image region, to obtain the first sample image region and the second sample image region of each sample palm image.

[0073] In this embodiment, the computer device may photograph the same sample palm of the sample object in different angles through a camera, to obtain the plurality of sample palm images in different photographing angles. As shown in FIG. 3, two sample palm images in different photographing angles are shown.

[0074] In this embodiment, the computer device may take a video of the same sample palm of the sample object through the camera, and the sample object moves and turns the sample palm in the video taking process, so as to change a palm posture of the sample palm, to obtain a palm video. A plurality of video frames are obtained from the palm video. Each video frame is used as a sample palm image.

[0075] Operation S204: Obtain a first sample embedding feature of each first sample image region, and map each first sample embedding feature into a key space to obtain a key vector of each first sample embedding feature.

[0076] The key space means a projection space in a query-key-value (QKV) mechanism. A QKV projection operation is an operation in an attention mechanism, and is configured for projecting an input vector into three different spaces and converting the input vector into vectors in the three spaces. Three different spaces are a query space, a key space, and a value space. The value space is also referred to as a content space. Each first sample embedding feature includes a palm print feature of the first sample image region, and may further include a posture feature of the first sample image region.

[0077] In this embodiment, the computer device performs feature mapping on each first sample image region, to map the first sample image region into a vector space to obtain the first sample embedding feature of each first sample image region. The computer device maps each first sample embedding feature into the key space to obtain the key vector of each first sample embedding feature.

[0078] Operation S206: Obtain a second sample embedding feature of each second sample image region, and map each second sample embedding feature into a query space to obtain a query vector of each second sample embedding feature.

[0079] The query space is a projection space in the QKV mechanism. The query space is different from the key space. Each second sample embedding feature includes a palm print feature of the second sample image region, and may further include a posture feature of the second sample image region.

[0080] In this embodiment, the computer device respectively performs feature mapping on each second sample image region, to map the second sample image region into the vector space to obtain the second sample embedding feature of each second sample image region. The computer device maps each second sample embedding feature into the query space to obtain the query vector of each second sample embedding feature.

[0081] Operation S208: Perform self-attention prediction according to each key vector and each query vector, to obtain a predicted self-attention weight corresponding to each first sample embedding feature and a predicted self-attention weight corresponding to each second sample embedding feature.

[0082] The predicted self-attention weight is a self-attention weight determined by a model in a model training process. A final self-attention weight is obtained after the model training is completed. The self-attention weight is a weight that represents a relevance between features in a self-attention mechanism. The relevance between the features may be represented by a similarity between the features.

[0083] The self-attention mechanism is a common mechanism in computer vision and natural language processing, which establishes association and weight allocation between the features by calculating similarities between queries (Q), keys (K), and values (V). In this embodiment, the self-attention mechanism is applied to feature extraction of the plurality of palm images of the same sample palm, to obtain more feature information of the sample palm.

[0084] In this embodiment, the computer device determines, for each key vector, a content similarity between the key vector and each query vector, and performs the self-attention prediction according to each content similarity, to obtain the predicted self-attention weight corresponding to each first sample embedding feature and the predicted self-attention weight corresponding to each second sample embedding feature.

[0085] Further, each content similarity is normalized to obtain a predicted self-attention weight corresponding to each first sample embedding feature and the predicted self-attention weight corresponding to each second sample embedding feature.

[0086] In this embodiment, the computer device may determine a key similarity between the key vectors, and perform self-attention prediction according to each content similarity and each key similarity to obtain the predicted self-attention weight corresponding to each first sample embedding feature and the predicted self-attention weight corresponding to each second sample embedding feature.

[0087] In this embodiment, the computer device may determine a query similarity between the query vectors, and perform self-attention prediction according to each content similarity and each query similarity to obtain the predicted self-attention weight corresponding to each first sample embedding feature and the predicted self-attention weight corresponding to each second sample embedding feature.

[0088] Operation S210: Fuse each first sample embedding feature with each second sample embedding feature according to the predicted self-attention weight of each first sample embedding feature and the predicted self-attention weight of each second sample embedding feature, to obtain a predicted palm feature jointly corresponding to the plurality of sample palm images.

[0089] In this embodiment, the computer device fuses each first sample embedding feature with each second sample embedding feature according to the predicted self-attention weight of each first sample embedding feature and the predicted self-attention weight of each second sample embedding feature to obtain the predicted palm feature jointly corresponding to the plurality of sample palm images.

[0090] Further, the computer device performs weighted summation on each first sample embedding feature as well as each second sample embedding feature, and the corresponding predicted self-attention weights, to obtain the predicted palm feature jointly corresponding to the plurality of sample palm images.

[0091] Operation S212: Obtain a label palm feature jointly corresponding to the plurality of sample palm images, and perform model training based on a predicted difference between the predicted palm feature and the label palm feature, to obtain a feature extraction model.

[0092] The label palm feature is a pre-marked feature that is used as a reference result of model training.

[0093] In this embodiment, the computer device obtains the label palm feature jointly corresponding to the plurality of sample palm images, determines the predicted difference between the predicted palm feature and the label palm feature, performs model training based on the predicted difference between the predicted palm feature and the label palm feature, to adjust a model parameter and continue training, and obtains the feature extraction model when the training is completed. The feature extraction model obtained by training includes a self-attention weight.

[0094] In this embodiment, the computer device determines the predicted difference between the predicted palm feature and the label palm feature, and obtains a preset iteration condition; performs matching on the predicted difference and the preset iteration condition; and when the predicted difference satisfies the preset iteration condition, adjusts a model parameter and continues training, stops the training until the predicted difference determined during training does not satisfy the preset iteration condition, and obtains the feature extraction model.

[0095] The preset iteration condition is a condition in which a next iteration needs to be performed. The preset iteration condition may be a preset quantity of iterations, a preset difference

threshold, or the like. For example, when the predicted difference is greater than the difference threshold, the model parameter is adjusted and the training is continued. The training is stopped until the predicted difference determined in the training process is less than or equal to the difference threshold, and the feature extraction model on which the training is completed is obtained.

[0096] In this embodiment, the plurality of sample palm images of the sample palm are obtained. The first sample image region and the second sample image region other than the first sample image region in each sample palm image are determined, and each first sample image region includes the preset part of the sample palm, so that the same image regions in the plurality of sample palm images and different image regions in the plurality of sample palm images can be determined. The first sample embedding feature of each first sample image region is obtained, and each first sample embedding feature is mapped into the key space. A feature representation of the first sample image regions that are the same can be mapped from the vector space to the key space, so as to obtain the key vector of each first sample embedding feature. The second sample embedding feature of each second sample image region is obtained, and each second sample embedding feature is mapped into the query space. Feature representations of different first sample image regions can be mapped from the vector space to the query space, so as to obtain the query vector of each second sample embedding feature. The predicted self-attention weight corresponding to each first sample embedding feature of the plurality of sample palm images and the predicted self-attention weight corresponding to each second sample embedding feature of the plurality of sample palm images can be predicted according to the key vectors of the same sample image regions and the query vectors of different sample image regions. The predicted self-attention weights that are predicted represent an importance of each key vector and an importance of each query vector in fusion, namely, representing attention degrees of a model to the features in the same sample image regions, thus simultaneously considering the features in the same sample image regions and the features in different sample image regions. Each first sample embedding feature and each second sample embedding feature are fused according to the predicted self-attention weight of the first sample embedding feature and the predicted self-attention weight of the second sample embedding feature, so that the same features and different features in the plurality of sample palm images can be fused, so as to obtain the predicted palm feature jointly corresponding to the plurality of sample palm images. The attention degrees of the model to the features in the same sample image regions are different, so that the features extracted from the same sample image regions are not completely the same. Therefore, different detailed features existing in the same sample image regions can be extracted, thereby obtaining a larger amount of feature information. The label palm

feature jointly corresponding to the plurality of sample palm images is obtained. The label palm feature is used as a palm feature that is expected to be obtained through model training. The model training is performed based on a difference between the predicted palm feature obtained through model prediction and the palm feature that is expected to be obtained, so that the difference between the predicted palm feature and the palm feature that is expected to be obtained is gradually reduced during the training, thereby gradually improving extraction precision of the model and finally obtaining the feature extraction model.

[0097] The finally obtained feature extraction model includes the self-attention weights obtained during training. The self-attention weights represent the attention degrees to the features of the same image regions and the features of different image regions in the plurality of palm images, so that the features of the same image regions are paid attention to, and different detailed features existing in the same image regions can be extracted, thereby obtaining a larger amount of feature information.

[0098] In an embodiment, the performing self-attention prediction according to each key vector and each query vector, to obtain a predicted self-attention weight corresponding to each first sample embedding feature and a predicted self-attention weight corresponding to each second sample embedding feature includes:

[0099] determining a key similarity between the key vectors and a query similarity between the query vectors; determining, for each key vector, a content similarity between the key vector and each query vector; and performing self-attention prediction according to each key similarity, each query similarity, and each content similarity, to obtain the predicted self-attention weight corresponding to each first sample embedding feature and the predicted self-attention weight corresponding to each second sample embedding feature.

[00100] In this embodiment, the computer device calculates a similarity between every two key vectors. A similarity between two key vectors is referred to as a key similarity. The key similarity may be a cosine similarity or a Euclidean distance. The computer device may calculate a similarity between every two query vectors. A similarity between two query vectors is referred to as a query similarity. The query similarity may be a cosine similarity or a Euclidean distance.

[00101] For each key vector, a similarity between the key vector and each query vector is determined, and the similarity between the key vector and the query vector is a content similarity. The content similarity may be a cosine similarity or a Euclidean distance.

[00102] In this embodiment, for each first sample embedding feature, the predicted self-attention weight corresponding to the first sample embedding feature is determined according to the key similarity between the key vector of the first sample embedding feature and each of other key vectors and the content similarity between the key vector of the first sample embedding feature and each query vector. For each second sample embedding feature, the predicted self-attention weight corresponding to the second sample embedding feature is determined according to the query similarity between the query vector of the second sample embedding feature and another query vector and the content similarity between the query vector of the second sample embedding feature and each key vector.

[00103] The other key vectors are key vectors except the key vector of the second sample embedding feature among the key vectors of the second sample embedding features. The another query vector is a query vector except the query vector of the second sample embedding feature in the query vectors of the second sample embedding features.

[00104] Further, for each first sample embedding feature, the key similarity between the key vector of the first sample embedding feature and each of other key vectors and the content similarity between the key vector of the first sample embedding feature and each query vector are determined, and the determined key similarity and the determined content similarity are fused to obtain the predicted self-attention weight corresponding to the first sample embedding feature. For each second sample embedding feature, the query similarity between the query vector of the second sample embedding feature and each of other query vectors and the content similarity between the query vector of the second sample embedding feature and each key vector are determined, and the determined query similarity and the determined content similarity are fused to obtain the predicted self-attention weight corresponding to the second sample embedding feature.

[00105] In this embodiment, a cosine similarity between the key vectors is determined. The cosine similarity between two key vectors is used as the key similarity between the two key vectors. A cosine similarity between the query vectors is determined. The cosine similarity between two query vectors is used as the query similarity between the two query vectors.

[00106] In this embodiment, a distance between the key vectors is determined. The distance between two key vectors is used as the key similarity between the two key vectors. The distance between the two key vectors may be a Euclidean distance between the two key vectors.

[00107] A distance between the query vectors is determined. The distance between two query vectors is used as the query similarity between the two query vectors. The distance between the two query vectors may be a Euclidean distance between the two query vectors.

[00108] In this embodiment, for each iteration of model training, the computer device determines key vectors and query vectors in the iteration, so as to determine a key similarity between the key vectors in the iteration and a query similarity between the query vectors in the iteration. In the iteration, for each key vector, a content similarity between the key vector and each query vector is determined. Self-attention prediction is performed according to each key similarity, each query similarity, and each content similarity that are in the iteration, to obtain the predicted self-attention weight corresponding to each first sample embedding feature and the predicted self-attention weight corresponding to each second sample embedding feature.

[00109] In this embodiment, the key similarity between the key vectors and the query similarity between the query vectors are determined. The key similarity represents a degree of correlation between the key vectors. A higher key similarity indicates a large degree of correlation between the key vectors. The query similarity represents a degree of correlation between the query vectors. A high query similarity indicates a large degree of correlation between the query vectors. For each key vector, the content similarity between the key vector and each query vector is determined. The content similarity represents a degree of correlation between the key vector and the query vector. A high content similarity indicates a degree of correlation between the key vector and the query vector. The self-attention prediction is performed according to each key similarity, each query similarity, and each content similarity to obtain the predicted self-attention weight corresponding to each first sample embedding feature and the predicted self-attention weight corresponding to each second sample embedding feature, so that the corresponding weights can be calculated based on the degrees of correlation between the vectors, and the self-attention weights of all the features in the plurality of sample palm images can be obtained.

[00110] In an embodiment, the mapping each first sample embedding feature into a key space to obtain a key vector of each first sample embedding feature includes: obtaining an initial key weight, and mapping each first sample embedding feature into the key space based on the initial key weight, to obtain the key vector of each first sample embedding feature.

[00111] The mapping each second sample embedding feature into a query space to obtain a query vector of each second sample embedding feature includes: obtaining an initial query weight, and mapping each second sample embedding feature into the query space based on the initial query weight, to obtain the query vector of each second sample embedding feature.

[00112] The initial key weight represents a conversion coefficient for conversion from the vector space to the key space in the training process. The initial query weight represents a conversion coefficient for conversion from the vector space to the key space in the training process.

[00113] In this embodiment, the computer device obtains the initial key weight, and performs point multiplication on the initial key weight and each first sample embedding feature, to map each first sample embedding feature into the key space, to obtain the key vector of each first sample embedding feature.

[00114] The computer device obtains the initial query weight, and performs point multiplication on the initial query weight and each second sample embedding feature, to map each second sample embedding feature into the query space, to obtain the query vector of each second sample embedding feature.

[00115] In this embodiment, the computer device obtains the feature extraction model through a plurality of times of iterative training, and each iteration corresponds to its initial key weight and initial query weight. The initial key weight of each iteration is obtained based on adjustment on the initial key weight of a previous iteration. An initial query weight of each iteration is obtained based on adjustment on the initial query weight of the previous iteration.

[00116] For each iteration, the computer device obtains the initial key weight of the iteration, and maps each first sample embedding feature into the key space based on the initial key weight of the iteration, to obtain the key vector of each first sample embedding feature in the iteration.

[00117] For each iteration, the computer device obtains the initial query weight of the iteration, and maps each second sample embedding feature into the query space based on the initial query weight of the iteration, to obtain the query vector of each second sample embedding feature in the iteration.

[00118] The computer device determines the predicted self-attention weight corresponding to each first sample embedding feature in the iteration and the predicted self-attention weight corresponding to each second sample embedding feature in the iteration according to each key vector and each query vector that are in the iteration. Each first sample embedding feature and each second sample embedding feature are fused according to the predicted self-attention weights of the first sample embedding feature and the second sample embedding feature in the iteration, to obtain the predicted palm feature jointly corresponding to the plurality of sample palm images in the iteration. The initial key weight and the initial query weight that correspond to the iteration are adjusted based on the predicted difference, in the iteration, between the predicted palm feature and the label palm feature, and an updated key weight and an updated query weight that are in the iteration are obtained. The updated key weight is used as the initial key weight of a next iteration, and the updated query weight is used as the initial query weight of the next iteration, thereby entering the next iteration until the feature extraction model is obtained through training.

[00119] As shown in FIG. 4, an initial key weight is W_k , and a first sample embedding feature is X_1 . A key vector K is obtained by mapping the first sample embedding feature X_1 into a key space through the initial key weight W_k .

[00120] An initial query weight is W_Q , and a second sample embedding feature is X_2 . A query vector Q is obtained by mapping the second sample embedding feature X_2 into a key space through the initial query weight W_Q . A predicted self-attention weight α is determined according to each key vector K and each query vector Q.

[00121] In this embodiment, the initial key weight is obtained. Each first sample embedding feature is mapped into the key space based on the initial key weight, so that each first sample image region can be accurately projected from a vector space to the key space, to obtain the key vector of each first sample embedding feature. The initial query weight is obtained. Each second sample embedding feature is mapped into the query space based on the initial query weight, so that each second sample embedding feature can be accurately projected from the vector space to the query space, to obtain the query vector of each second sample embedding feature.

[00122] In an embodiment, the performing model training based on a predicted difference between the predicted palm feature and the label palm feature, to obtain a feature extraction model includes:

[00123] determining the predicted difference between the predicted palm feature and the label palm feature; when the predicted difference satisfies a preset iteration condition, determining a key similarity between the key vectors, and adjusting the initial key weight according to each key similarity, to obtain an updated key weight; determining a query similarity between the query vectors, and adjusting the initial query weight according to each query similarity, to obtain an updated query weight; and using the updated key weight as an initial key weight of a next iteration, using the updated query weight as an initial query weight of the next iteration, returning to and continuing to perform the operation of mapping each first sample embedding feature into the key space based on the initial key weight, stopping the operation until the determined predicted difference does not satisfy the preset iteration condition, and obtaining the feature extraction model.

[00124] In this embodiment, the computer device determines the predicted difference between the predicted palm feature and the label palm feature, and performs matching on the predicted difference and the preset iteration condition. When the predicted difference satisfies the preset iteration condition, the computer device determines the key similarity between the key

vectors, and adjusts the initial key weight according to each key similarity, to obtain the updated key weight. In addition, the computer device determines the query similarity between the query vectors, and adjusts the initial query weight according to each query similarity, to obtain the updated query weight.

[00125] After the updated key weight and the updated query weight are obtained, the next iteration is performed. The updated key weight is used as the initial key weight of the next iteration, and the updated query weight is used as the initial query weight of the next iteration.

[00126] In this embodiment, each iteration corresponds to its initial key weight and initial query weight. For each iteration, the computer device obtains the initial key weight and the initial query weight of the iteration, maps each first sample embedding feature into the key space based on the initial key weight of the iteration, to obtain the key vector of each first sample embedding feature in the iteration, and respectively maps each second sample embedding feature into the query space based on the initial query weight of the iteration, to obtain the query vector of each second image region in the iteration.

[00127] The computer device determines the predicted self-attention weight corresponding to each first sample embedding feature in the iteration and the predicted self-attention weight corresponding to each second sample embedding feature in the iteration according to each key vector and each query vector that are in the iteration, and fuses each first sample embedding feature with each second sample embedding feature based on the predicted self-attention weights in the iteration, to obtain the predicted palm feature jointly corresponding to the plurality of sample palm images in the iteration. The computer device determines the predicted difference between the predicted palm feature in the iteration and the label palm feature, determines the key similarity between the key vectors in the iteration when the predicted difference in the iteration satisfies the preset iteration condition, and adjusts the initial key weight in the iteration according to each key similarity, to obtain the updated key weight in the iteration; determines the query similarity between the query vectors in the iteration, and adjusts the initial query weight in the iteration according to each query similarity, to obtain the updated query weight in the iteration; and uses the updated key weight as the initial key weight of the next iteration of the iteration, and uses the updated query weight as the initial query weight of the next iteration of the iteration, thus entering the next iteration, so as to perform and continue to perform the operation of mapping each first sample embedding feature into the key space based on the initial key weight in the next iteration, stops the operation until the determined predicted difference does not satisfy the preset iteration condition, and obtains the feature extraction model.

[00128] That the preset iteration condition is not satisfied may be that a quantity of iterations does not satisfy a preset quantity of iterations, the predicted difference of the iteration does not satisfy a preset difference threshold, or the like. That the predicted difference of the iteration does not satisfy the preset difference threshold may be that the predicted difference of the iteration is less than or equal to the difference threshold.

[00129] In this embodiment, by determining the predicted difference between the predicted palm feature and the label palm feature, a difference between the predicted palm feature obtained by model prediction and the palm feature that is expected to be obtained may be determined. When the predicted difference satisfies the preset iteration condition, it indicates that the next iteration needs to be performed, and the key weight and the query weight are updated first. To be specific, the key similarity between the key vectors is determined. The key similarity represents a degree of correlation between the key vectors. A higher similarity indicates a larger degree of correlation. Therefore, the initial key weight can be adjusted according to each key similarity, to obtain the updated key weight. The query similarity represents a degree of correlation between the query vectors. Thus, the initial query weight can be adjusted according to each query similarity, to obtain the updated query weight. The updated key weight is used as the initial key weight of the next iteration, and the updated query weight is used as the initial query weight of the next iteration, thus entering the next iteration for continuous execution. Therefore, the key weight and the query weight are continuously updated in the iterations, and the predicted self-attention weights determined based on the key weight and the query weight are also continuously updated, so that the difference between the predicted palm feature and the palm feature that is expected to be obtained is gradually reduced in the iterations. The operation is stopped until the determined predicted difference does not satisfy the preset iteration condition. In this case, a final key weight, a final query weight, and final self-attention weights are obtained, thereby obtaining the feature extraction model.

[00130] In an embodiment, the method further includes:

[00131] obtaining an initial content weight, and mapping each first sample embedding feature and each second sample embedding feature into a value space based on the initial content weight, to obtain a content vector of each first sample embedding feature and a content vector of each second sample embedding feature.

[00132] The fusing each first sample embedding feature with each second sample embedding feature according to the predicted self-attention weight of each first sample embedding feature and the predicted self-attention weight of each second sample embedding feature, to obtain a predicted palm feature jointly corresponding to the plurality of sample palm images includes:

[00133] fusing the content vector of each first sample embedding feature with the content vector of each second sample embedding feature according to the predicted self-attention weight of each first sample embedding feature and the predicted self-attention weight of each second sample embedding feature, to obtain the predicted palm feature jointly corresponding to the plurality of sample palm images.

[00134] The initial content weight represents a conversion coefficient for conversion from the vector space to the value space in the training process.

[00135] In this embodiment, the computer device obtains the initial content weight, and performs point multiplication on the initial content weight and each first sample embedding feature to map each first sample embedding feature and each second sample embedding feature into the value space, to obtain a content vector of each first sample embedding feature and a content vector of each second sample embedding feature.

[00136] The computer device fuses the content vector of each first sample embedding feature and the content vector of each second sample embedding feature based on the predicted self-attention weight of each first sample embedding feature and the predicted self-attention weight of each second sample embedding feature, to obtain the predicted palm feature corresponding to the plurality of sample palm images.

[00137] In this embodiment, the computer device performs weighted summation on the content vector of each first sample embedding feature as well as the content vector of each second sample embedding feature, and the corresponding predicted self-attention weights, to obtain the predicted palm feature jointly corresponding to the plurality of sample palm images.

[00138] As shown in FIG. 5, a sample palm image is mapped into a vector space to obtain a corresponding sample embedding feature X . The sample embedding feature X includes a first sample embedding feature X_1 and a second sample embedding feature X_2 . To be specific, a first sample image region and a second sample image region are respectively mapped to the vector space to obtain the first sample embedding feature X_1 and the second sample embedding feature X_2 .

[00139] As shown in FIG. 6, a predicted self-attention weight α is determined according to each key vector K and each query vector Q .

[00140] The sample embedding feature X is mapped into the value space based on an initial content weight W_v . To be specific, each first sample embedding feature X_1 and each second sample embedding feature X_2 are respectively mapped into the value space, to obtain a content vector V .

[00141] The content vectors V are fused according to respective predicted self-attention weights, to obtain a predicted palm feature.

[00142] In this embodiment, the initial content weight is obtained. Each first sample embedding feature and each second sample embedding feature are respectively mapped into the value space based on the initial content weight, so that all embedding features of the plurality of sample palm images can be accurately mapped into the value space, to obtain the content vector of each first sample embedding feature and the content vector of each second sample embedding feature. The content vector of each first sample embedding feature and the content vector of each second sample embedding feature are fused according to the predicted self-attention weight of each first sample embedding feature and the predicted self-attention weight of each second sample embedding feature, so that features of the same parts in the plurality of sample palm images can be fused to different degrees, and attention degrees to the features of the same parts in the sample palm images are different. Therefore, differences between the features of the same parts in the plurality of sample palm images can be paid attention to, and these detailed features having differences can be obtained. Furthermore, with reference to features of other image regions that do not include the preset part, the predicted palm feature jointly corresponding to the plurality of sample palm images that are obtained contains a larger amount of information, helping to improve accuracy of recognizing and verifying the palm feature in different scenarios.

[00143] In an embodiment, the method further includes:

[00144] determining, for each key vector, a content similarity between the key vector and each query vector when the predicted difference satisfies the preset iteration condition; adjusting the initial content weight according to each content similarity, to obtain an updated content weight; and using the updated content weight as an initial content weight of a next iteration, and performing the returning to the operation of mapping each first sample embedding feature into the key space based on the initial key weight.

[00145] In this embodiment, when the predicted difference satisfies the preset iteration condition, for each key vector, the content similarity between the key vector and each query vector is determined. The content similarity may be a cosine similarity or a distance. To be specific, a cosine similarity between each key vector and each query vector is used as the content similarity, or a distance between each key vector and each query vector is used as the content similarity. The distance may be a Euclidean distance.

[00146] The computer device adjusts the initial content weight according to each content similarity, to obtain the updated content weight, and uses the updated content weight as the initial

content weight of the next iteration, and enters the next iteration to perform the operation of mapping each first sample embedding feature into the key space based on the initial key weight in the next iteration.

[00147] In this embodiment, the computer device obtains the feature extraction model through a plurality of times of iterative training, and each iteration corresponds to its initial key weight, initial query weight, and initial content weight. The initial content weight of each iteration is obtained based on adjustment on the initial content weight of a previous iteration.

[00148] For each iteration, the computer device obtains the initial content weight of the iteration, and maps each first sample embedding feature and each first sample embedding feature into the key space based on the initial content weight of the iteration, to obtain the content vector of each first sample embedding feature in the iteration and the content vector of each second sample embedding feature in the iteration.

[00149] After the predicted self-attention weight corresponding to each first sample embedding feature in the iteration and the predicted self-attention weight corresponding to each second sample embedding feature in the iteration are determined, the content vectors in the iteration are fused based on the predicted self-attention weights in the iteration, to obtain the predicted palm feature jointly corresponding to the plurality of sample palm images in the iteration. The computer device determines the predicted difference between the predicted palm feature in the iteration and the label palm feature, and adjusts the initial key weight in the iteration according to the key similarity between the key vectors in the iteration when the predicted difference in the iteration satisfies the preset iteration condition, to obtain the updated key weight in the iteration; adjusts the initial query weight in the iteration according to the query similarity between the query vectors in the iteration, to obtain the updated query weight in the iteration; adjusts the initial content weight in the iteration according to the content similarity between the query vectors in the iteration, to obtain the updated content weight in the iteration; and uses the updated key weight as the initial key weight of the next iteration of the iteration, uses the updated query weight as the initial query weight of the next iteration, and uses the updated content weight as the initial content weight of the next iteration, thus entering the next iteration, so as to perform and continue to perform the operation of mapping each first sample embedding feature into the key space based on the initial key weight in the next iteration, stops the operation until the determined predicted difference does not satisfy the preset iteration condition, and obtains the feature extraction model.

[00150] In this embodiment, when the predicted difference satisfies the preset iteration condition, for each key vector, the content similarity between the key vector and each query vector

is determined. The content similarity represents a degree of correlation between the key vector and the query vector, so as to adjust the initial content weight according to each content similarity, to obtain the updated content weight. The updated content weight is used as the initial content weight of the next iteration, and the returning to the operation of mapping each first sample embedding feature into the key space based on the initial key weight is performed, so as to continuously update the content weight in the iterations. Thus, the difference between the predicted palm feature and the palm feature that is expected to be obtained is gradually reduced in the iterations, thereby gradually improving extraction precision of a model and finally obtaining the feature extraction model.

[00151] In an embodiment, as shown in FIG. 7, the determining a first sample image region and a second sample image region other than the first sample image region in each sample palm image includes:

[00152] Operation S702: Extract a feature point of each sample palm image.

[00153] Operation S704: Recognize the preset part in each sample palm image based on the feature point of each sample palm image, and determine the first sample image region, in which the preset part is located, in each sample palm image.

[00154] In this embodiment, the computer device performs feature point extraction on each sample palm image, to obtain a plurality of feature points of each sample palm image.

[00155] For each sample palm image, the computer device recognizes the preset part based on the feature points of the sample palm image, determines a position, in the sample palm image, of each feature point forming the preset part, and obtains, according to the determined position, the first sample image region including the preset part.

[00156] Operation S706: Determine, for each sample palm image, the second sample image region other than the corresponding first sample image region in the sample palm image.

[00157] In this embodiment, the computer device uses a region other than the first sample image region in the sample palm image as the second sample image region, so as to obtain the first sample image region and the second sample image region in the sample palm image. According to the same processing, the computer device may obtain the first sample image region and the second sample image region of each sample palm image.

[00158] In this embodiment, the feature points of each sample palm image are extracted. The preset part in each sample palm image is recognized based on the feature points of each sample palm image, and the first sample image region, in which the preset part is located, in each sample palm image is determined, so that image regions of the same preset part included in the plurality of

sample palm images can be determined. The same or very similar features exist between these image regions. It facilitates further extraction of detailed features for the same or similar features, thereby improving accuracy of obtaining and recognizing the detailed features. For each sample palm image, the second sample image region other than the corresponding first sample image region in the sample palm image is determined, thus facilitating subsequently performing different processing on a sample image region that includes the preset part and a sample image region that does not include the preset part.

[00159] In an embodiment, the feature extraction model processing method includes a plurality of parts such as image obtaining, image mapping, self-attention mechanism processing, parallel training, and feature storage.

[00160] As shown in FIG. 8, the image obtaining part is configured for: obtaining a plurality of sample palm images of a palm of the same user. By obtaining the sample palm images in different angles and at different time points, diversified data including more information is obtained. To be specific, the diversified data of the plurality of palm images of the same user is obtained by photographing the palm of the user in a plurality of angles and a plurality of tenses through a plurality of cameras. The data includes palm information in different angles, at different time points, and in different dynamic conditions. This provides a richer information basis for subsequent feature extraction and processing.

[00161] The image mapping is intended to map a sample palm image into a particular vector space, so as to extract a feature representation of the sample palm image and reserve more information of the sample palm image. A redundant and sparse embedding mode may be used. To be specific, a vector may have a redundant feature, and the entire vector is sparse. In this way, the feature information of the sample palm image can be better reserved.

[00162] For example, a sample palm image is represented through a pixel matrix. Each element represents a pixel value of the sample palm image. The pixel matrix is a 5*5 matrix. A value of an element is in a range from 0 to 255. Therefore, the pixel matrix may be flattened into a 25-dimensional vector, and each pixel value is used as a feature. Dimension reduction is performed on the 25-dimensional vector, and the 25-dimensional vector is mapped to a vector representation with a low dimension. This may be specifically implemented by using a dimension reduction method such as principal component analysis (PCA) or an autocoder. In this way, information of a sample palm image can be compressed to a low dimension, and important features are reserved. Finally, a low-dimensional embedding feature vector of the sample palm image may be obtained. Each element represents a feature representation of a corresponding pixel of the sample palm image

in a low-dimensional space. The vector may be configured for a subsequent palm feature extraction, matching, or recognition task.

[00163] By using the redundant and sparse embedding mode, feature information of more sample palm images can be reserved. This mode allows existence of a redundant feature. To be specific, a plurality of embedding feature representations of the same feature are allowed to exist. Meanwhile, the entire vector is represented by using sparsity, so that a model can better extract and represent detailed features of the sample palm image, thereby improving accuracy and robustness of subsequent processing and recognition stages on the palm.

[00164] The self-attention mechanism processing is an attention processing unit for a plurality of sample palm images of a multi-head self-attention mechanism, and aims to calculate association and weight allocation between the plurality of sample palm images of the palm by using the self-attention mechanism, thereby obtaining more redundant information and increasing an amount of information of a palm feature. The unit projects the embedding feature into a plurality of subspaces by using a QKV mechanism, to learn the features of the sample palm images in the plurality of subspaces in parallel, and enhance a calculation effect by using the multi-head attention mechanism. To be specific, the self-attention mechanism is used to calculate degrees of correlation and weight allocation between the plurality of sample palm images. By converting embedding features of sample image regions of the sample palm images into query vectors, key vectors, and content vectors, and calculating similarities between the query vectors, the key vectors, and the content vectors, redundant information between the plurality of sample palm images can be obtained, and relevances between these sample palm images can be captured, thereby obtaining more redundant information and increasing an amount of information of a palm feature.

[00165] After predicted self-attention weights are calculated, feature fusion is performed on content vectors, to obtain a predicted palm feature. The predicted palm feature obtained by the feature fusion is used as an input of the parallel training part in FIG. 8.

[00166] As shown in FIG. 9, the parallel training uses a feedforward neural network as a basic model. The feedforward neural network is a common deep learning model, and performs information processing and feature extraction in a manner of transmitting a signal layer by layer. To avoid problems such as gradient explosion and enhance a training effect of a model, a residual network (ResNet) structure and a batch normalization technology are used during training. ResNet introduces cross-layer connection and residual learning, to solve problems of gradient disappearance and a limited expression capability during deep network training. In the batch normalization, each small batch of samples are normalized, so that an input distribution of a

network is more stable, and a training speed and a generalization capability of a model are improved.

[00167] In the parallel training, parallel computation is performed on each layer to reserve a feature of a sample palm image. By parallel computation, features of a plurality of sample palm images can be processed simultaneously, thereby avoiding information loss. This parallel computation mode helps to increase the training speed and reserve detailed features of the sample palm images.

[00168] As shown in FIG. 10, by using the feedforward neural network structure, the ResNet, the batch normalization technology, and the parallel computation mode, a parallel training module can increase a training speed of a palm swiping recognition system. To avoid problems such as gradient explosion, parallel computation is performed on each layer, and the palm features of the sample palm images are reserved, thereby improving computation efficiency. In addition, the training process can be accelerated. Meanwhile, the palm features can be effectively extracted and learned, thereby improving accuracy and performance of a palm swiping recognition system.

[00169] As shown in FIG. 10, a first sample image region and a second sample image region of a sample palm image may be mapped into a vector space, to obtain a sample embedding feature 1 and a sample embedding feature 2. The sample embedding feature 1 and the sample embedding feature 2 may be normalized in parallel, and the two embedding features that are normalized are respectively inputted to residual blocks to implement parallel processing in a residual network.

[00170] Or, the sample embedding feature 1 and the sample embedding feature 2 may be respectively inputted to hidden neurons of a hidden layer, so that a hidden neuron 1 processes the sample embedding feature 1, and a hidden neuron 2 processes the sample embedding feature 2, to implement parallel processing. A feature processed by the hidden neuron 1 is inputted to a residual block 1 of the residual network, and a feature processed by the hidden neuron 2 is inputted to a residual block 2 of the residual network, thereby obtaining the predicted palm feature 1 outputted by the residual block 1 and a predicted palm feature 2 outputted by the residual block 2, to implement parallel processing of the parts during training. A predicted palm feature corresponding to the sample palm image may be obtained by fusing the predicted palm feature 1 with the predicted palm feature 2.

[00171] In this embodiment, the sample embedding feature 1 and the sample embedding feature 2 may be embedding features respectively corresponding to two sample palm images, thereby implementing parallel training on a plurality of sample palm images.

[00172] In this embodiment, to improve calculation efficiency and explore different subspaces, parallel learning during the training allows features of a plurality of sample palm images to be processed simultaneously, to reserve more information and more features. A multi-head attention mechanism allows a model to learn and calculate degrees of correlation and importances between the sample palm images in the plurality of subspaces, thereby increasing an amount of information. By the combination of the parallel learning and the multi-head attention mechanism, redundant information between different sample palm images can be captured more comprehensively, so as to extract richer palm feature information to improve accuracy of a palm swiping recognition system.

[00173] In an embodiment, as shown in FIG. 11, a feature extraction method is provided. Application of the method to the computer device (the computer device may be the terminal or the server in FIG. 1) in FIG. 1 is taken as an example for description, including the following operations:

[00174] Operation S1102: Obtain a plurality of palm images of a palm.

[00175] The palm is a palm of a to-be-authenticated object. The to-be-authenticated object may be a person or an animal. The palm may be a palm, a sole of a foot, or the like of the person, or may be a front paw, a back paw, or the like of the animal.

[00176] The palm images are images obtained by photographing the palm of the to-be-authenticated object. The palm may include fingers, a center of palm, and a palm edge. The fingers include a thumb, an index finger, a middle finger, a ring finger, a little finger, and the like. The palm includes a preset part, and the preset part may be a center-of-palm part of the palm.

[00177] In this embodiment, the computer device may obtain a plurality of palm images of the palm of the to-be-authenticated object, for example, obtain a plurality of palm images of a left palm of a person or obtain a plurality of palm images of a right palm of a person.

[00178] In this embodiment, the computer device may photograph the same palm of the to-be-authenticated object in different angles through a camera, to obtain the plurality of palm images in different photographing angles.

[00179] In this embodiment, the computer device may take a video of the same palm of the to-be-authenticated object through the camera, and the to-be-authenticated object moves and turns the palm in the video taking process, so as to change a palm posture of the palm, to obtain a palm video. A plurality of video frames are obtained from the palm video. Each video frame is used as a palm image.

[00180] Operation S1104: Determine, through the feature extraction model obtained by training by the feature extraction model processing method, a first image region and a second image region other than the first image region in each palm image, each first image region including a preset part of the palm.

[00181] In this embodiment, the computer device inputs each palm image to the feature extraction model. The feature extraction model is obtained by training through the feature extraction model processing method.

[00182] The feature extraction model determines the preset part in each palm image, determines a region, in which the preset part is located, in the palm image, and uses the region as the first image region. For each palm image, a region other than the first image region in the palm image is used as the second image region, to obtain the first image region and the second image region of each palm image.

[00183] In this embodiment, the determining, through the feature extraction model obtained by training by the feature extraction model processing method, a first image region and a second image region other than the first image region in each palm image includes:

[00184] extracting a feature point of each palm image through the feature extraction model obtained by training by the feature extraction model processing method; through the feature extraction model, recognizing the preset part in each palm image based on the feature point of each palm image, and determine the first sample image region, in which the preset part is located, in each sample palm image; and determining, for each palm image, the second image region other than the corresponding first image region in the palm image through the feature extraction model.

[00185] Operation S1106: Obtain a first embedding feature of each first image region and a second embedding feature of each second image region through the feature extraction model.

[00186] In this embodiment, the computer device performs feature mapping on each first image region through the feature extraction model, to map the first image region into a vector space to obtain the first embedding feature of each first image region. The computer device performs feature mapping on each second image region through the feature extraction model, to map the second image region into the vector space to obtain the second embedding feature of each second image region.

[00187] Operation S1108: Fuse each first embedding feature with each second embedding feature according to a self-attention weight for each first embedding feature and a self-attention

weight for each second embedding feature in the feature extraction model through the feature extraction model, to obtain a palm feature jointly corresponding to the plurality of palm images.

[00188] In this embodiment, the feature extraction model obtains a plurality of self-attention weights through training. The plurality of self-attention weights include: the self-attention weight for each first embedding feature and the self-attention weight for each second embedding feature. The computer device fuses each first embedding feature with each second embedding feature based on the self-attention weights through the feature extraction model, to obtain the palm feature corresponding to the plurality of palm images.

[00189] Further, the computer device performs weighted summation on each first embedding feature as well as each second embedding feature, and the corresponding predicted self-attention weights, to obtain the palm feature jointly corresponding to the plurality of palm images.

[00190] In this embodiment, the feature extraction model obtained by training by the feature extraction model processing method includes a plurality of self-attention weights, and the plurality of self-attention weights represent attention degrees of the feature extraction model to features of the same image regions and different image regions in the plurality of palm images. The plurality of palm images of the palm are obtained. The first image region and the second image region other than the first image region in each palm image are determined through the feature extraction model. Each first image region includes the preset part of the palm, so that the same image region in the plurality of palm images and different image regions in the plurality of palm images can be determined. The first embedding feature of each first image region and the second embedding feature of each second image region are obtained through the feature extraction model, and the image regions are converted from an image space to a vector space to obtain the corresponding embedding features. Each first embedding feature and each second embedding feature are fused according to the self-attention weight for each first embedding feature and the self-attention weight for each second embedding feature in the feature extraction model through the feature extraction model, so that features of different image regions and features of the same image regions are both paid attention to. Thus, different detailed features existing in the same image regions can be extracted, and the obtained palm feature includes a larger amount of information. In addition, the extracted palm feature includes more detailed features. This can improve recognition accuracy of feature recognition using the palm feature in different scenarios.

[00191] In an embodiment, the feature extraction model further includes a content weight obtained by training. The method further includes: mapping each first embedding feature and each second embedding feature into a value space based on the content weight through the feature

extraction model, to obtain a content vector of each first embedding feature and a content vector of each second embedding feature.

[00192] The fusing each first embedding feature with each second embedding feature according to a self-attention weight for each first embedding feature and a self-attention weight for each second embedding feature in the feature extraction model through the feature extraction model, to obtain a palm feature jointly corresponding to the plurality of palm images includes: fusing the content vector of each first embedding feature with the content vector of each second embedding feature according to the self-attention weight for each first embedding feature and the self-attention weight for each second embedding feature in the feature extraction model through the feature extraction model, to obtain the palm feature jointly corresponding to the plurality of palm images.

[00193] The content weight represents a conversion coefficient for converting from the vector space to the value space.

[00194] In this embodiment, the feature extraction model obtains the plurality of self-attention weights and the plurality of content weights through training. The computer device performs point multiplication on the content weight and each first embedding feature through the feature extraction model, to map each first embedding feature and each second embedding feature into the value space, to obtain the content vector of each first embedding feature and the content vector of each second embedding feature.

[00195] The computer device fuses the content vector of each first embedding feature with the content vector of each second embedding feature according to the self-attention weight for each first embedding feature and the self-attention weight for each second embedding feature through the feature extraction model, to obtain the palm feature jointly corresponding to the plurality of palm images.

[00196] In this embodiment, the self-attention weights include weights corresponding to the content vectors. The computer device performs weighted summation on the content vector of each first embedding feature as well as the content vector of each second embedding feature, and the corresponding weights through the feature extraction model, to obtain the palm feature jointly corresponding to the plurality of palm images.

[00197] In this embodiment, through the feature extraction model, each first embedding feature and each second embedding feature are mapped into the value space based on the content weights, so that all embedding features of the plurality of palm images can be accurately mapped into the value space, to obtain the content vector of each first embedding feature and the content

vector of each second embedding feature. The content vector of each first embedding feature and the content vector of each second embedding feature are fused according to the self-attention weight for each first embedding feature and the self-attention weight for each second embedding feature, so that features of the same parts in the plurality of palm images can be fused to different degrees, and attention degrees to the features of the same parts in the palm images are different. Therefore, differences between the features of the same parts in the plurality of palm images can be paid attention to, and these detailed features having differences can be obtained. Furthermore, with reference to features of other image regions that do not include the preset part, the palm feature corresponding to the plurality of palm images that are obtained contains a larger amount of information, helping to improve accuracy of recognizing and verifying the palm feature in different scenarios.

[00198] In an embodiment, the feature extraction method is applied to a resource transfer scenario. The obtaining a plurality of palm images of a palm includes:

[00199] invoking, in response to a resource transfer trigger operation, a camera to photograph the palm of a to-be-authenticated object in different angles, to obtain the plurality of palm images.

[00200] The method further includes:

[00201] performing identity authentication on the to-be-authenticated object according to the palm feature, to obtain an authentication result; and obtaining resource transfer information when the authentication result indicates that the identity authentication on the to-be-authenticated object succeeds, and performing a resource transfer operation based on the resource transfer information.

[00202] In this embodiment, in the resource transfer scenario, a to-be-authenticated user may run a resource transfer application on the computer device, and taps to trigger the resource transfer operation after entering the resource transfer information based on the resource transfer application. The resource transfer application invokes, in response to the resource transfer trigger operation, the camera to photograph the palm of the to-be-authenticated object in different angles, to obtain the plurality of palm images.

[00203] The computer device inputs the plurality of palm images into the feature extraction model, and determines the first image region and the second image region other than the first image region in each palm image through the feature extraction model, each first image region including the preset part of the palm; obtains the first embedding feature of each first image region and the second embedding feature of each second image region through the feature extraction model; and fuses each first embedding feature with each second embedding feature based on the self-attention

weights through the feature extraction model, to obtain the palm feature corresponding to the plurality of palm images. The computer device performs identity authentication on the to-be-authenticated object according to the palm feature, to obtain the authentication result; and obtains the resource transfer information when the authentication result indicates that the identity authentication on the to-be-authenticated object succeeds, and performs the resource transfer operation based on the resource transfer information.

[00204] In this embodiment, each first embedding feature and each second embedding feature are mapped into the value space based on the content weight through the feature extraction model, to obtain the content vector of each first embedding feature and the content vector of each second embedding feature. The content vector of each first embedding feature and the content vector of each second embedding feature are fused according to the self-attention weight for each first embedding feature and the self-attention weight for each second embedding feature through the feature extraction model, to obtain the palm feature corresponding to the plurality of palm images. The computer device performs identity authentication on the to-be-authenticated object according to the palm feature. When the identity authentication on the to-be-authenticated object succeeds, the computer device obtains the resource transfer information and performs the resource transfer operation based on the resource transfer information.

[00205] In this embodiment, the feature extraction method is applied to the resource transfer scenario, so that in response to the resource transfer trigger operation, the camera is invoked to photograph the palm of the to-be-authenticated object in different angles, to obtain the plurality of palm images. Thus, the palm feature of the plurality of palm images is obtained through the feature extraction model. The obtained palm feature includes features, having detail differences, of the same parts in the plurality of palm images, so that the identity authentication on the to-be-authenticated object can be more accurate according to the palm feature. Therefore, the resource transfer operation is performed based on the resource transfer information, thereby effectively improving security of resource transfer.

[00206] In another embodiment, the feature extraction method may be applied to access control, gate control, or the like. As shown in FIG. 12, a user places a palm in a palm recognition region of a gate, the palm recognition region obtains a plurality of palm images of the palm of the user, and the plurality of palm images are processed through a feature extraction model, to obtain a palm feature. Identity authentication is performed on the user according to the palm feature. The gate will be opened when the authentication succeeds. When the authentication fails, the gate will not be opened, and a prompt is sent.

[00207] The feature extraction method may be applied to any application scenario in which feature extraction of a palm image needs to be performed, and is not limited to the examples of the foregoing embodiments.

[00208] In an embodiment, a feature extraction model processing method is provided, applied to a computer device and including:

[00209] obtaining a plurality of sample palm images of a sample palm, and extracting a feature point of each sample palm image;

[00210] next, recognizing a preset part in each sample palm image based on the feature point of each sample palm image, and determining a first sample image region, in which the preset part is located, in each sample palm image; and

[00211] then, determining, for each sample palm image, a second sample image region other than the corresponding first sample image region in the sample palm image.

[00212] Further, a first sample embedding feature of each first sample image region is obtained, and a second sample embedding feature of each second sample image region is obtained.

[00213] Further, an initial key weight, an initial query weight, and an initial content weight of a current iteration are obtained, and each first sample embedding feature is mapped into a key space based on the initial key weight of the current iteration, to obtain a key vector of each first sample embedding feature in the current iteration; each second sample embedding feature is mapped into a query space based on the initial query weight of the current iteration, to obtain a query vector of each second image region in the current iteration; and each first sample embedding feature and each second sample embedding feature are mapped into a value space based on the initial content weight of the current iteration, to obtain a content vector of each first sample embedding feature and a content vector of each second sample embedding feature in the current iteration. When the current iteration is a first iteration, the initial key weight, the initial query weight, and the initial content weight are all preset weights.

[00214] Then, in the current iteration, for each key vector, a content similarity between the key vector and each query vector is determined. Each content similarity is normalized to obtain a predicted self-attention weight corresponding to each first sample embedding feature and a predicted self-attention weight corresponding to each second sample embedding feature.

[00215] Next, the content vectors in the current iteration are fused based on the predicted self-attention weights in the current iteration, to obtain a predicted palm feature jointly corresponding to the plurality of sample palm images in the current iteration.

[00216] Further, a label palm feature jointly corresponding to the plurality of sample palm images are obtained, and a predicted difference between the predicted palm feature and the label palm feature in the current iteration.

[00217] Next, when the predicted difference of the current iteration is greater than a difference threshold, a key similarity between the key vectors in the current iteration is determined, and the initial key weight of the current iteration is adjusted according to each key similarity, to obtain an updated key weight in the current iteration. A query similarity between the query vectors in the current iteration is determined, and the initial query weight of the current iteration is adjusted according to each query similarity, to obtain an updated query weight in the current iteration. In the current iteration, for each key vector, a content similarity between the key vector and each query vector is determined. The initial content weight of the current iteration is adjusted according to each content similarity, to obtain an updated content weight in the current iteration.

[00218] Further, the updated key weight, the updated query weight, and the updated query weight in the current iteration are respectively used as an initial key weight, an initial query weight, and an initial content weight that correspond to a next iteration of the current iteration, thus entering the next iteration, so as to return to and continue to perform the operation of mapping each first sample embedding feature into a key space based on the initial key weight. The operation is stopped until the determined predicted difference is less than or equal to the difference threshold, to obtain a feature extraction model. A key weight, a query weight, a content weight, and a self-attention weight of the feature extraction model are obtained when a final iteration is stopped.

[00219] In this embodiment, a feature extraction method is further provided, including:

[00220] obtaining a plurality of palm images of a palm, and determining a first image region and a second image region other than the first image region in each palm image through a feature extraction model, each first image region including a preset part of the palm;

[00221] then, obtaining a first embedding feature of each first image region and a second embedding feature of each second image region through the feature extraction model; and mapping each first embedding feature and each second embedding feature into a value based on a content weight, to obtain a content vector of each first embedding feature and a content vector of each second embedding feature.

[00222] Further, the content vector of each first embedding feature and the content vector of each second embedding feature are fused according to a self-attention weight for each first embedding feature and a self-attention weight for each second embedding feature in the feature

extraction model through the feature extraction model, to obtain a palm feature jointly corresponding to the plurality of palm images.

[00223] In this embodiment, feature points of each sample palm image are extracted. The preset part in each sample palm image is recognized based on the feature points of each sample palm image, and a first sample image region, in which the preset part is located, in each sample palm image is determined, so that image regions of the same preset part included in the plurality of sample palm images can be determined. The same or very similar features exist between these image regions. It facilitates further extraction of detailed features for the same or similar features, thereby improving accuracy of obtaining and recognizing the detailed features. For each sample palm image, a second sample image region other than the corresponding first sample image region in the sample palm image is determined, so as to determine different image regions in the plurality of sample palm images.

[00224] Each first sample image region is accurately projected from a vector space into a key space based on an initial key weight, to obtain a key vector of each first sample embedding feature. Each second sample image region is accurately projected from the vector space into a query space based on an initial query weight, to obtain a query vector of each second sample embedding feature. Each first sample embedding feature and each second sample embedding feature are accurately projected from the vector space to the value space based on a content weight, to obtain a content vector corresponding to the first sample embedding feature and a content vector corresponding to the second sample embedding feature.

[00225] A predicted self-attention weight corresponding to each first sample embedding feature and a predicted self-attention weight corresponding to each second sample embedding feature can be predicted according to the key vectors of the same sample image regions and the query vectors of different sample image regions. The predicted self-attention weights that are predicted represent an importance of each key vector and an importance of each query vector in fusion, namely, representing attention degrees of a model to features in the same sample image regions, thus simultaneously considering the features in the same sample image regions and features in different sample image regions. Each first sample embedding feature and each second sample embedding feature are fused based on the predicted self-attention weights, so that the same and different features in the plurality of sample palm images can be fused, to obtain a predicted palm feature corresponding to the plurality of sample palm images. The attention degrees of the model to the features in the same sample image regions are different, so that the features extracted from the same sample image regions are not completely the same. Therefore, different detailed features

existing in the same sample image regions can be extracted, thereby obtaining a larger amount of feature information. A label palm feature corresponding to the plurality of sample palm images is obtained. The label palm feature is used as a palm feature that is expected to be obtained through model training, and a difference between the predicted palm feature obtained through model prediction and the palm feature that is expected to be obtained can be determined. When the predicted difference satisfies a preset iteration condition, it indicates that a next iteration needs to be performed, and the key weight and the query weight are updated first. To be specific, a key similarity between the key vectors is determined. The key similarity represents a degree of correlation between the key vectors. A higher similarity indicates a larger degree of correlation. Therefore, the initial key weight can be adjusted according to each key similarity, to obtain an updated key weight. A query similarity represents a degree of correlation between the query vectors. Thus, the initial query weight can be adjusted according to each query similarity, to obtain an updated query weight. A content similarity represents a degree of correlation between each query vector and each key vector, so that the initial content weight can be adjusted according to each content similarity, to obtain an updated content weight.

[00226] The updated key weight, the updated query weight, and the updated content weight are used as an initial key weight, an initial query weight, and an initial content weight of a next iteration, so as to enter the next iteration for continuous execution. Therefore, the key weight and the query weight are continuously updated in the iterations, and the predicted self-attention weights determined based on the key weight and the query weight are also continuously updated, so that the difference between the predicted palm feature and the palm feature that is expected to be obtained is gradually reduced in the iterations. The operation is stopped until the determined predicted difference does not satisfy the preset iteration condition. In this case, a final key weight, a final query weight, and final self-attention weights are obtained, thereby obtaining a feature extraction model.

[00227] The finally obtained feature extraction model includes a plurality of self-attention weights obtained during training. The plurality of self-attention weights represent the attention degrees to the features of the same image regions and the features of different image regions in the plurality of palm images, so that the features of the same image regions are paid attention to, and different detailed features existing in the same image regions can be extracted, thereby obtaining a larger amount of feature information.

[00228] Although the operations are displayed sequentially according to the instructions of the arrows in the flowcharts of the embodiments, these operations are not necessarily performed

sequentially according to the sequence instructed by the arrows. Unless otherwise explicitly specified in the present disclosure, execution of the operations is not strictly limited, and the operations may be performed in other sequences. Moreover, at least some operations in the flowchart involved in the above embodiments may include a plurality of operations or a plurality of stages. The operations or stages are not necessarily executed completely at the same moment and may be executed at different moments. The operations or stages are not necessarily sequentially executed, and may be executed alternately with other operations or at least some operations or stages of other operations.

[00229] Based on the same inventive concept, an embodiment of the present disclosure further provides a feature extraction model processing apparatus for implementing the above feature extraction model processing method. An implementation solution provided by the apparatus to solve the problems is similar to the implementation solution recorded in the above method. Therefore, for specific limitations on the one or more embodiments of the feature extraction model processing apparatus provided below, refer to the limitations on the feature extraction model processing method mentioned above, and details will not be elaborated here.

[00230] In an embodiment, as shown in FIG. 13, a feature extraction model processing apparatus is provided, including:

[00231] a sample obtaining module 1302, configured to: obtain a plurality of sample palm images of a sample palm, and determine a first sample image region and a second sample image region other than the first sample image region in each sample palm image, each first sample image region including a preset part of the sample palm;

[00232] a key space mapping module 1304, configured to: obtain a first sample embedding feature of each first sample image region, and map each first sample embedding feature into a key space to obtain a key vector of each first sample embedding feature;

[00233] a query space mapping module 1306, configured to: obtain a second sample embedding feature of each second sample image region, and map each second sample embedding feature into a query space to obtain a query vector of each second sample embedding feature;

[00234] a weight prediction module 1308, configured to perform self-attention prediction according to each key vector and each query vector, to obtain a predicted self-attention weight corresponding to each first sample embedding feature and a predicted self-attention weight corresponding to each second sample embedding feature;

[00235] a sample feature fusion module 1310, configured to fuse each first sample embedding feature with each second sample embedding feature according to the predicted self-attention weight of each first sample embedding feature and the predicted self-attention weight of each second sample embedding feature, to obtain a predicted palm feature jointly corresponding to the plurality of sample palm images; and

[00236] a training module 1312, configured to: obtain a label palm feature jointly corresponding to the plurality of sample palm images, and perform model training based on a predicted difference between the predicted palm feature and the label palm feature, to obtain a feature extraction model.

[00237] In this embodiment, the plurality of sample palm images of the sample palm are obtained. The first sample image region and the second sample image region other than the first sample image region in each sample palm image are determined, and each first sample image region includes the preset part of the sample palm, so that the same image regions in the plurality of sample palm images and different image regions in the plurality of sample palm images can be determined. The first sample embedding feature of each first sample image region is obtained, and each first sample embedding feature is mapped into the key space. A feature representation of the first sample image regions that are the same can be mapped from the vector space to the key space, so as to obtain the key vector of each first sample embedding feature. The second sample embedding feature of each second sample image region is obtained, and each second sample embedding feature is mapped into the query space. Feature representations of different first sample image regions can be mapped from the vector space to the query space, so as to obtain the query vector of each second sample embedding feature. The predicted self-attention weight corresponding to each first sample embedding feature and the predicted self-attention weight corresponding to each second sample embedding feature can be predicted according to the key vectors of the same sample image regions and the query vectors of different sample image regions. The predicted self-attention weights represent an importance of each key vector and an importance of each query vector in fusion, namely, representing attention degrees of a model to the features in the same sample image regions, thus simultaneously considering the features in the same sample image regions and the features in different sample image regions. Each first sample embedding feature and each second sample embedding feature are fused based on the predicted self-attention weights, so that the same and different features in the plurality of sample palm images can be fused, to obtain the predicted palm feature corresponding to the plurality of sample palm images. The attention degrees of the model to the features in the same sample image regions are different, so that the features extracted from the same sample image regions are not completely the same. Therefore,

different detailed features existing in the same sample image regions can be extracted, thereby obtaining a larger amount of feature information. The label palm feature corresponding to the plurality of sample palm images is obtained. The label palm feature is used as a palm feature that is expected to be obtained through model training. The model training is performed based on a difference between the predicted palm feature obtained through model prediction and the palm feature that is expected to be obtained, so that the difference between the predicted palm feature and the palm feature that is expected to be obtained is gradually reduced during the training, thereby gradually improving extraction precision of the model and finally obtaining the feature extraction model.

[00238] The finally obtained feature extraction model includes the self-attention weights obtained during training. The self-attention weights represent the attention degrees to the features of the same image regions and the features of different image regions in the plurality of palm images, so that the features of the same image regions are paid attention to, and different detailed features existing in the same image regions can be extracted, thereby obtaining a larger amount of feature information.

[00239] In an embodiment, the weight prediction module 1308 is further configured to: determine a key similarity between the key vectors and a query similarity between the query vectors; determine, for each key vector, a content similarity between the key vector and each query vector; and perform self-attention prediction according to each key similarity, each query similarity, and each content similarity, to obtain the predicted self-attention weight corresponding to each first sample embedding feature and the predicted self-attention weight corresponding to each second sample embedding feature.

[00240] In this embodiment, the key similarity between the key vectors and the query similarity between the query vectors are determined. The key similarity represents a degree of correlation between the key vectors. A higher key similarity indicates a large degree of correlation between the key vectors. The query similarity represents a degree of correlation between the query vectors. A high query similarity indicates a large degree of correlation between the query vectors. For each key vector, the content similarity between the key vector and each query vector is determined. The content similarity represents a degree of correlation between the key vector and the query vector. A high content similarity indicates a degree of correlation between the key vector and the query vector. The self-attention prediction is performed according to each key similarity, each query similarity, and each content similarity to determine the predicted self-attention weight corresponding to each first sample embedding feature and the predicted self-attention weight

corresponding to each second sample embedding feature, so that the corresponding weights can be calculated based on the degrees of correlation between the vectors, and the predicted self-attention weights of all the features in the plurality of sample palm images can be obtained.

[00241] In an embodiment, the key space mapping module 1304 is further configured to: obtain an initial key weight, and map each first sample embedding feature into the key space based on the initial key weight, to obtain the key vector of each first sample embedding feature.

[00242] The query space mapping module 1306 is further configured to: obtain an initial query weight, and map each second sample embedding feature into the query space based on the initial query weight, to obtain the query vector of each second sample embedding feature.

[00243] In this embodiment, the initial key weight is obtained. Each first sample embedding feature is mapped into the key space based on the initial key weight, so that each first sample image region can be accurately projected from a vector space to the key space, to obtain the key vector of each first sample embedding feature. The initial query weight is obtained. Each second sample embedding feature is mapped into the query space based on the initial query weight, so that each second sample image region can be accurately projected from the vector space to the query space, to obtain the query vector of each second sample embedding feature.

[00244] In an embodiment, the training module 1312 is further configured to: determine the predicted difference between the predicted palm feature and the label palm feature; when the predicted difference satisfies a preset iteration condition, determine a key similarity between the key vectors, and adjust the initial key weight according to each key similarity, to obtain an updated key weight; determine a query similarity between the query vectors, and adjust the initial query weight according to each query similarity, to obtain an updated query weight; and use the updated key weight as an initial key weight of a next iteration, using the updated query weight as an initial query weight of the next iteration, return to and continue to perform the operation of mapping each first sample embedding feature into the key space based on the initial key weight, stop the operation until the determined predicted difference does not satisfy the preset iteration condition, and obtain the feature extraction model.

[00245] In this embodiment, by determining the predicted difference between the predicted palm feature and the label palm feature, a difference between the predicted palm feature obtained by model prediction and the palm feature that is expected to be obtained may be determined. When the predicted difference satisfies the preset iteration condition, it indicates that the next iteration needs to be performed, and the key weight and the query weight are updated first. To be specific, the key similarity between the key vectors is determined. The key similarity represents a degree of

correlation between the key vectors. A higher similarity indicates a larger degree of correlation. Therefore, the initial key weight can be adjusted according to each key similarity, to obtain the updated key weight. The query similarity represents a degree of correlation between the query vectors. Thus, the initial query weight can be adjusted according to each query similarity, to obtain the updated query weight. The updated key weight is used as the initial key weight of the next iteration, and the updated query weight is used as the initial query weight of the next iteration, thus entering the next iteration for continuous execution. Therefore, the key weight and the query weight are continuously updated in the iterations, and the predicted self-attention weights determined based on the key weight and the query weight are also continuously updated, so that the difference between the predicted palm feature and the palm feature that is expected to be obtained is gradually reduced in the iterations. The operation is stopped until the determined predicted difference does not satisfy the preset iteration condition. In this case, a final key weight, a final query weight, and final self-attention weights are obtained, thereby obtaining the feature extraction model.

[00246] In an embodiment, the apparatus further includes:

[00247] a value space mapping module, configured to: obtain an initial content weight, and map each first sample embedding feature and each second sample embedding feature into a value space based on the initial content weight, to obtain a content vector of each first sample embedding feature and a content vector of each second sample embedding feature.

[00248] The sample feature fusion module 1310 is further configured to fuse the content vector of each first sample embedding feature with the content vector of each second sample embedding feature according to the predicted self-attention weight of each first sample embedding feature and the predicted self-attention weight of each second sample embedding feature, to obtain the predicted palm feature jointly corresponding to the plurality of sample palm images.

[00249] In this embodiment, the initial content weight is obtained. Each first sample embedding feature and each second sample embedding feature are mapped into the value space based on the initial content weight, so that all embedding features of the plurality of sample palm images can be accurately mapped into the value space, to obtain the content vector of each first sample embedding feature and the content vector of each second sample embedding feature. The content vector of each first sample embedding feature and the content vector of each second sample embedding feature are fused based on the predicted self-attention weights, so that features of the same parts in the plurality of sample palm images can be fused to different degrees, and attention degrees to the features of the same parts in the sample palm images are different. Therefore, differences between the features of the same parts in the plurality of sample palm images can be

paid attention to, and these detailed features having differences can be obtained. Furthermore, with reference to features of other image regions that do not include the preset part, the predicted palm feature corresponding to the plurality of sample palm images that are obtained contains a larger amount of information, helping to improve accuracy of recognizing and verifying the palm feature in different scenarios.

[00250] In an embodiment, the training module 1312 is further configured to determine, for each key vector, a content similarity between the key vector and each query vector when the predicted difference satisfies the preset iteration condition; adjust the initial content weight according to each content similarity, to obtain an updated content weight; and use the updated content weight as an initial content weight of a next iteration, and perform the returning to the operation of mapping each first sample embedding feature into the key space based on the initial key weight.

[00251] In this embodiment, when the predicted difference satisfies the preset iteration condition, for each key vector, the content similarity between the key vector and each query vector is determined. The content similarity represents a degree of correlation between the key vector and the query vector, so as to adjust the initial content weight according to each content similarity, to obtain the updated content weight. The updated content weight is used as the initial content weight of the next iteration, and the returning to the operation of mapping each first sample embedding feature into the key space based on the initial key weight is performed, so as to continuously update the content weight in the iterations. Thus, the difference between the predicted palm feature and the palm feature that is expected to be obtained is gradually reduced in the iterations, thereby gradually improving extraction precision of a model and finally obtaining the feature extraction model.

[00252] In an embodiment, the sample obtaining module 1302 is further configured to extract a feature point of each sample palm image; recognize the preset part in each sample palm image based on the feature point of each sample palm image, and determine the first sample image region, in which the preset part is located, in each sample palm image; and determine, for each sample palm image, the second sample image region other than the corresponding first sample image region in the sample palm image.

[00253] In this embodiment, the feature points of each sample palm image are extracted. The preset part in each sample palm image is recognized based on the feature points of each sample palm image, and the first sample image region, in which the preset part is located, in each sample palm image is determined, so that image regions of the same preset part included in the plurality of sample palm images can be determined. The same or very similar features exist between these

image regions. It facilitates further extraction of detailed features for the same or similar features, thereby improving accuracy of obtaining and recognizing the detailed features. For each sample palm image, the second sample image region other than the corresponding first sample image region in the sample palm image is determined, thus facilitating subsequently performing different processing on a sample image region that includes the preset part and a sample image region that does not include the preset part.

[00254] Based on the same inventive concept, an embodiment of the present disclosure further provides a feature extraction apparatus for implementing the above feature extraction method. An implementation solution provided by the apparatus to solve the problems is similar to the implementation solution recorded in the above method. Therefore, for specific limitations on the one or more embodiments of the feature extraction apparatus provided below, refer to the limitations on the feature extraction method mentioned above, and details will not be elaborated here.

[00255] In an embodiment, as shown in FIG. 14, a feature extraction apparatus 1400 is provided, including:

[00256] an image obtaining module 1402, configured to obtain a plurality of palm images of a palm;

[00257] a region determining module 1404, configured to determine, through the feature extraction model obtained by training by the feature extraction model processing apparatus, a first image region and a second image region other than the first image region in each palm image, each first image region including a preset part of the palm;

[00258] a feature obtaining module 1406, configured to obtain a first embedding feature of each first image region and a second embedding feature of each second image region through the feature extraction model;

[00259] a feature fusion module 1408, configured to fuse each first embedding feature with each second embedding feature according to a self-attention weight for each first embedding feature and a self-attention weight for each second embedding feature in the feature extraction model through the feature extraction model, to obtain a palm feature jointly corresponding to the plurality of palm images, the self-attention weights being obtained by training on the feature extraction model.

[00260] In this embodiment, the feature extraction model obtained by training by the feature extraction model processing method includes a plurality of self-attention weights, and the plurality

of self-attention weights represent attention degrees of the feature extraction model to features of the same image regions and different image regions in the plurality of palm images. The plurality of palm images of the palm are obtained. The first image region and the second image region other than the first image region in each palm image are determined through the feature extraction model. Each first image region includes the preset part of the palm, so that the same image region in the plurality of palm images and different image regions in the plurality of palm images can be determined. The first embedding feature of each first image region and the second embedding feature of each second image region are obtained through the feature extraction model, and the image regions are converted from an image space to a vector space to obtain the corresponding embedding features. Through the feature extraction model, fusion is performed according to the self-attention weights of each first embedding feature and each second embedding feature in the feature extraction model, so that features of different image regions and features of the same image regions are both paid attention to. Thus, different detailed features existing in the same image regions can be extracted, and the obtained palm feature includes a larger amount of information. In addition, the extracted palm feature includes more detailed features. This can improve recognition accuracy of feature recognition using the palm feature in different scenarios.

[00261] In an embodiment, the feature extraction model further includes a content weight obtained by training. The apparatus further includes:

[00262] a mapping module, configured to map each first embedding feature and each second embedding feature into a value space based on the content weight, to obtain a content vector of each first embedding feature and a content vector of each second embedding feature.

[00263] The feature fusion module is further configured to fuse the content vector of each first embedding feature with the content vector of each second embedding feature according to the self-attention weight for each first embedding feature and the self-attention weight for each second embedding feature in the feature extraction model through the feature extraction model, to obtain the palm feature jointly corresponding to the plurality of palm images.

[00264] In this embodiment, through the feature extraction model, each first embedding feature and each second embedding feature are mapped into the value space based on the content weights, so that all embedding features of the plurality of palm images can be accurately mapped into the value space, to obtain the content vector of each first embedding feature and the content vector of each second embedding feature. The content vector of each first embedding feature and the content vector of each second embedding feature are fused according to the self-attention weights, so that features of the same parts in the plurality of palm images can be fused to different

degrees, and attention degrees to the features of the same parts in the palm images are different. Therefore, differences between the features of the same parts in the plurality of palm images can be paid attention to, and these detailed features having differences can be obtained. Furthermore, with reference to features of other image regions that do not include the preset part, the palm feature corresponding to the plurality of palm images that are obtained contains a larger amount of information, helping to improve accuracy of recognizing and verifying the palm feature in different scenarios.

[00265] In an embodiment, the image obtaining module is further configured to invoke, in response to a resource transfer trigger operation, a camera to photograph the palm of a to-be-authenticated object in different angles, to obtain the plurality of palm images.

[00266] The apparatus further includes:

[00267] an authentication module, configured to perform identity authentication on the to-be-authenticated object according to the palm feature, to obtain an authentication result; and

[00268] a resource transfer module, configured to: obtain resource transfer information when the authentication result indicates that the identity authentication on the to-be-authenticated object succeeds, and perform a resource transfer operation based on the resource transfer information.

[00269] In this embodiment, the feature extraction method is applied to the resource transfer scenario, so that in response to the resource transfer trigger operation, the camera is invoked to photograph the palm of the to-be-authenticated object in different angles, to obtain the plurality of palm images. Thus, the palm feature of the plurality of palm images is obtained through the feature extraction model. The obtained palm feature includes features, having detail differences, of the same parts in the plurality of palm images, so that the identity authentication on the to-be-authenticated object can be more accurate according to the palm feature. Therefore, the resource transfer operation is performed based on the resource transfer information, thereby effectively improving security of resource transfer.

[00270] The modules in the feature extraction model processing apparatus and the modules in the feature extraction apparatus may be partially or completely implemented through software, hardware, or any combination thereof. The foregoing modules may be built in or independent of a processor of a computer device in a form of hardware, or may be stored in a memory of the computer device in a form of software, for the processor to invoke to execute operations corresponding to the foregoing modules.

[00271] In one embodiment, a computer device is provided. The computer device may be a terminal or a server. The computer device is, for example, a terminal. A diagram of an internal structure of the terminal may be shown in FIG. 15. The computer device includes a processor, a memory, an input/output interface, a communication interface, a display unit, and an input apparatus. The processor, the memory, and the input/output interface are connected through a system bus. The communication interface, the display unit, and the input apparatus are connected to the system bus through the input/output interface. The processor of the computer device is configured to provide computing and control capabilities. The memory of the computer device includes a non-volatile storage medium and an internal memory. The non-volatile storage medium stores an operating system and a computer program. The internal memory provides an environment for running of the operating system and the computer program in the non-volatile storage medium. The input/output interface of the computer device is configured to exchange information between the processor and an external device. The communication interface of the computer device is configured for conducting wired or wireless communication with an external terminal. The wireless communication may be implemented through wireless fidelity (WIFI), a mobile cellular network, near field communication (NFC), or another technology. The computer program implements the feature extraction model processing method when executed by the processor. The display unit of the computer device is configured to form visual pictures, which may be a display screen, a projection apparatus or a virtual reality imaging apparatus. The display screen may be a liquid crystal display screen or an e-ink display screen. The input apparatus of the computer device may be a touch layer covering the display screen, or may be a button, a trackball, or a touchpad arranged on a housing of the computer device, or may be an external keyboard, touchpad, a mouse or the like.

[00272] A person skilled in the art can understand that the structure shown in FIG. 15 is merely a block diagram of a partial structure related to a solution in the present disclosure, and does not constitute a limitation on the computer device to which the solution in the present disclosure is applied. Specifically, the computer device may include more or fewer components than those shown in the figure, or some components may be combined, or a different component layout may be used.

[00273] In an embodiment, a computer device is further provided, including a memory and a processor. The memory has a computer program stored therein. The processor, when executing the computer program, implements the operations in the foregoing method embodiments.

[00274] In an embodiment, a computer-readable storage medium is provided, having a computer program stored therein. The computer program, when executed by a processor, implements the operations in the foregoing method embodiments.

[00275] In an embodiment, a computer program product is provided, including a computer program. The computer program, when executed by a processor, implements the operations in the foregoing method embodiments.

[00276] Here, user information (including, but not limited to, user equipment information, user personal information, and the like) and data (including, but not limited to, data for analysis, stored data, displayed data, and the like) involved in the present disclosure are all information and data authorized by users or fully authorized by all parties, and collection, use, and processing of relevant data need to comply with relevant laws, regulations, and standards of relevant countries and regions.

[00277] A person of ordinary skill in the art may understand that all or some of procedures of the method in the foregoing embodiments may be implemented by a computer program instructing relevant hardware. The computer program may be stored in a non-volatile computer-readable storage medium. When the program is executed, the procedures of the foregoing method embodiments may be implemented. Any reference to a memory, a database, or other media used in the embodiments provided in the present disclosure can include at least one of a non-volatile memory and a volatile memory. The non-volatile memory may include a read-only memory (ROM), a magnetic tape, a floppy disk, a flash memory, an optical memory, a high-density embedded non-volatile memory, a resistive random access memory (ReRAM), a magnetoresistive random access memory (MRAM), a ferroelectric random access memory (FRAM), a phase change memory (PCM), a graphene memory, and the like. The volatile memory can include a random access memory (RAM), external cache memory, or the like. For the purpose of illustration rather than limitation, the RAM may be in various forms, such as a static random access memory (SRAM), or a dynamic random access memory (DRAM). The database involved in the embodiments of the present disclosure may include at least one of a relational database and a non-relational database. The non-relational database may include a distributed database based on a block chain, and is not limited thereto. The processor involved in the embodiments provided in the present disclosure may be a general-purpose processor, a central processing unit, a graphics processing unit, a digital signal processor, a programmable logic device, a quantum computing-based data processing logic device, or the like, but is not limited thereto.

[00278] Technical features of the foregoing embodiments may be randomly combined. To make description concise, not all possible combinations of the technical features in the foregoing embodiments are described. However, the combinations of these technical features shall be considered as falling within the scope recorded by this specification provided that no conflict exists.

[00279] The foregoing embodiments only describe several implementations of the present disclosure, which are described specifically and in detail, but cannot be construed as a limitation to the patent scope of the present disclosure. For a person of ordinary skill in the art, several transformations and improvements can be made without departing from the idea of the present disclosure. These transformations and improvements belong to the protection scope of the present disclosure. Therefore, the protection scope of the patent of the present disclosure shall be subject to the appended claims.

CLAIMS

What is claimed is:

1. A feature extraction model processing method, performed by a computer device and the method comprising:

obtaining a plurality of sample palm images of a sample palm;

determining a first sample image region and a second sample image region other than the first sample image region in each sample palm image, each first sample image region comprising a preset part of the sample palm;

obtaining a first sample embedding feature of each first sample image region;

mapping each first sample embedding feature into a key space to obtain a key vector of the first sample embedding feature;

obtaining a second sample embedding feature of each second sample image region;

mapping each second sample embedding feature into a query space to obtain a query vector of the second sample embedding feature;

performing self-attention prediction according to the key vectors of the first sample embedding features and the query vectors of the second sample embedding features, to obtain a predicted self-attention weight corresponding to the first sample embedding features and a predicted self-attention weight corresponding to the second sample embedding features; and

fusing the first sample embedding features with the second sample embedding features according to the predicted self-attention weight of the first sample embedding features and the predicted self-attention weight of the second sample embedding features, to obtain a predicted palm feature jointly corresponding to the plurality of sample palm images;

obtaining a label palm feature jointly corresponding to the plurality of sample palm images; and

performing model training based on a predicted difference between the predicted palm feature and the label palm feature, to obtain a feature extraction model.

2. The method according to claim 1, wherein the performing self-attention prediction according to the key vectors of the first sample embedding features and the query vectors of the second sample embedding features, to obtain a predicted self-attention weight corresponding to the first sample embedding features and a predicted self-attention weight corresponding to the second

sample embedding features comprises:

determining a key similarity between the key vectors and a query similarity between the query vectors;

determining, for each key vector, a content similarity between the key vector and each query vector; and

performing self-attention prediction according to the key similarity, the query similarity, and the content similarities, to obtain the predicted self-attention weight corresponding to the first sample embedding features and the predicted self-attention weight corresponding to the second sample embedding features.

3. The method according to claim 1 or 2, wherein the mapping each first sample embedding feature into the key space to obtain the key vector of the first sample embedding feature comprises:

mapping each first sample embedding feature into the key space based on an initial key weight of the first sample embedding feature to obtain the key vector of the first sample embedding feature; and

the mapping each second sample embedding feature into the query space to obtain the query vector of the second sample embedding feature comprises:

mapping each second sample embedding feature into the query space based on an initial query weight of the second sample embedding feature to obtain the query vector of the second sample embedding feature.

4. The method according to claim 3, wherein the performing model training based on the predicted difference between the predicted palm feature and the label palm feature, to obtain the feature extraction model comprises:

determining the predicted difference between the predicted palm feature and the label palm feature;

when the predicted difference satisfies a preset iteration condition, determining a key similarity between the key vectors, and adjusting the initial key weight according to each key similarity, to obtain an updated key weight;

determining a query similarity between the query vectors, and adjusting the initial query weight according to each query similarity, to obtain an updated query weight; and

using the updated key weight as an initial key weight of a next iteration, using the updated query weight as an initial query weight of the next iteration, returning to and continuing to perform

the operation of mapping each first sample embedding feature into the key space based on the initial key weight, stopping the operation until the determined predicted difference does not satisfy the preset iteration condition, and obtaining the feature extraction model.

5. The method according to claim 4, further comprising:

obtaining an initial content weight, and mapping each first sample embedding feature and each second sample embedding feature into a value space based on the initial content weight, to obtain a content vector of each first sample embedding feature and a content vector of each second sample embedding feature; and

the fusing the first sample embedding features with the second sample embedding features according to the predicted self-attention weight of the first sample embedding features and the predicted self-attention weight of the second sample embedding features, to obtain a predicted palm feature jointly corresponding to the plurality of sample palm images comprising:

fusing the content vector of each first sample embedding feature with the content vector of each second sample embedding feature according to the predicted self-attention weight of each first sample embedding feature and the predicted self-attention weight of each second sample embedding feature, to obtain the predicted palm feature jointly corresponding to the plurality of sample palm images.

6. The method according to claim 5, further comprising:

determining, for each key vector, a content similarity between the key vector and each query vector when the predicted difference satisfies the preset iteration condition;

adjusting the initial content weight according to each content similarity, to obtain an updated content weight; and

using the updated content weight as an initial content weight of a next iteration, and performing the returning to the operation of mapping each first sample embedding feature into the key space based on the initial key weight.

7. The method according to any one of claims 1 to 6, wherein the determining the first sample image region and the second sample image region in each sample palm image comprises:

extracting a feature point of the sample palm image;

recognizing the preset part in the sample palm image based on the feature point of the sample palm image;

determining the first sample image region, in which the preset part is located, in the sample

palm image; and

determining the second sample image region other than the first sample image region in the sample palm image.

8. A feature extraction method, performed by a computer device and comprising:

obtaining a plurality of palm images of a palm;

determining, through the feature extraction model obtained by training by the feature extraction model processing method according to any one of claims 1 to 7, a first image region and a second image region other than the first image region in each palm image, each first image region comprising a preset part of the palm;

obtaining a first embedding feature of each first image region and a second embedding feature of each second image region through the feature extraction model; and

fusing each first embedding feature with each second embedding feature according to a self-attention weight for each first embedding feature and a self-attention weight for each second embedding feature in the feature extraction model through the feature extraction model, to obtain a palm feature jointly corresponding to the plurality of palm images, the self-attention weights being obtained by training on the feature extraction model.

9. The method according to claim 8, wherein the feature extraction model further comprises a content weight obtained by training; and the method further comprises:

mapping each first embedding feature and each second embedding feature into a value space based on the content weight through the feature extraction model, to obtain a content vector of each first embedding feature and a content vector of each second embedding feature; and

the fusing each first embedding feature with each second embedding feature according to a self-attention weight for each first embedding feature and a self-attention weight for each second embedding feature in the feature extraction model through the feature extraction model, to obtain a palm feature jointly corresponding to the plurality of palm images comprises:

fusing the content vector of each first embedding feature with the content vector of each second embedding feature according to the self-attention weight for each first embedding feature and the self-attention weight for each second embedding feature in the feature extraction model through the feature extraction model, to obtain the palm feature jointly corresponding to the plurality of palm images.

10. The method according to any one of claims 8 to 9, wherein the feature extraction method is

applied to a resource transfer scenario; and the obtaining a plurality of palm images of a palm comprises:

invoking, in response to a resource transfer trigger operation, a camera to photograph the palm of a to-be-authenticated object in different angles, to obtain the plurality of palm images;

the method further comprises:

performing identity authentication on the to-be-authenticated object according to the palm feature, to obtain an authentication result; and

obtaining resource transfer information when the authentication result indicates that the identity authentication on the to-be-authenticated object succeeds, and performing a resource transfer operation based on the resource transfer information.

11. A feature extraction model processing apparatus, comprising:

a sample obtaining module, configured to: obtain a plurality of sample palm images of a sample palm, and determine a first sample image region and a second sample image region other than the first sample image region in each sample palm image, each first sample image region comprising a preset part of the sample palm;

a key space mapping module, configured to: obtain a first sample embedding feature of each first sample image region, and map each first sample embedding feature into a key space to obtain a key vector of the first sample embedding feature;

a query space mapping module, configured to: obtain a second sample embedding feature of each second sample image region, and map each second sample embedding feature into a query space to obtain a query vector of the second sample embedding feature;

a weight prediction module, configured to perform self-attention prediction according to the key vectors of the first sample embedding features and the query vectors of the second sample embedding features, to obtain a predicted self-attention weight corresponding to the first sample embedding features and a predicted self-attention weight corresponding to the second sample embedding features;

a sample feature fusion module, configured to fuse the first sample embedding features with the second sample embedding features according to the predicted self-attention weight of the first sample embedding features and the predicted self-attention weight of the second sample embedding features, to obtain a predicted palm feature jointly corresponding to the plurality of sample palm images; and

a training module, configured to: obtain a label palm feature jointly corresponding to the plurality of sample palm images, and perform model training based on a predicted difference between the predicted palm feature and the label palm feature, to obtain a feature extraction model.

12. A feature extraction apparatus, comprising:

an image obtaining module, configured to obtain a plurality of palm images of a palm;

a region determining module, configured to determine, through the feature extraction model obtained by training by the feature extraction model processing apparatus according to claim 11, a first image region and a second image region other than the first image region in each palm image, each first image region comprising a preset part of the palm;

a feature obtaining module, configured to obtain a first embedding feature of each first image region and a second embedding feature of each second image region through the feature extraction model; and

a feature fusion module, configured to fuse each first embedding feature with each second embedding feature according to a self-attention weight for each first embedding feature and a self-attention weight for each second embedding feature in the feature extraction model through the feature extraction model, to obtain a palm feature jointly corresponding to the plurality of palm images, the self-attention weights being obtained by training on the feature extraction model.

13. A computer device, comprising a memory and a processor, the memory having a computer program stored therein, and the processor, when executing the computer program, implementing the operations of the method according to any one of claims 1 to 10.

14. A computer-readable storage medium, having a computer program stored therein, the computer program, when executed by a processor, implementing the operations of the method according to any one of claims 1 to 10.

15. A computer program product, comprising a computer program, the computer program, when executed by a processor, implementing the operations of the method according to any one of claims 1 to 10.