

TARGET RE-IDENTIFICATION METHOD AND APPARATUS, DEVICE, STORAGE MEDIUM, AND PROGRAM PRODUCT

CROSS-REFERENCE TO RELATED APPLICATION

[0001] The disclosure claims priority to Chinese Patent Application No. 202011022834.4 filed on September 25, 2020, the disclosure of which is hereby incorporated by reference in its entirety.

TECHNICAL FIELD

[0002] The disclosure relates to the field of computer technologies, and in particular, to a target re-identification method and apparatus, a device, a storage medium, and a program product.

BACKGROUND

[0003] Target re-identification is important in the field of computer vision and intelligent video surveillance, and is performed to determine positions of a same target captured by a same camera or different cameras. As the urbanization process advances and a quantity of urban cameras continues to increase, target re-identification plays an important role in practical applications in many fields, such as pedestrian walking behavior analysis, pedestrian and vehicle tracking across cameras, and detection of abnormal behavior of pedestrian vehicles. However, in practical applications, target re-identification may be interfered by other targets in a scenario. In addition, there may be factors such as extremely similar appearance features between targets, and these factors will adversely affect the target re-identification process.

SUMMARY

[0004] Embodiments of the disclosure provide a target re-identification technical solution.

[0005] According to an aspect of the embodiments of the disclosure, a target re-identification method is provided, including: acquiring a target object picture and a to-be-processed picture set, the target object picture including a target object; determining a candidate picture set from the to-be-processed picture set according to a first feature value of the target object picture and a second feature value set corresponding to the to-be-processed picture set, a value of similarity between any two pictures in the candidate picture set being greater than or equal to a preset similarity value; and identifying the first feature value and the second feature value set based on a trained graph association identification network, and determining a target picture set from the candidate picture set, where a first value of similarity between an object included in a picture in the target picture set and the target object is greater than or equal to a first value of similarity between an object included in a non-target picture and the target object, and the candidate picture set includes the target picture set and the non-target picture. In this way, a more accurate positive sample can be determined from the to-be-processed picture set, and interference of a negative sample can be reduced to obtain the target picture set, so that accuracy of a result of subsequently performing trajectory and behavior analysis on the object included in the picture in the target picture set based on attribute information of the picture in the target picture set is improved.

[0006] In some possible implementations, the graph association identification network may include a first graph structure establishment sub-network, a graph association update sub-network and a classifier that are serially connected. The identifying the first feature value and the second feature value set based on a trained graph association identification network, and determining a target picture set from the candidate picture set may include: inputting the first feature value and the second feature value set into the first graph structure establishment sub-network to obtain a first graph structure, where the first graph structure includes nodes and edges each of which is for connecting two nodes, a quantity of nodes is the same as a quantity of pictures in the candidate picture set, and the edge for connecting the two nodes is determined based on a similarity between the two nodes that are connected and a preset similarity; inputting the first graph structure into the graph

association update sub-network to obtain an updated and optimized second graph structure; determining the first value of similarity corresponding to each candidate picture in the candidate picture set through the classifier according to the second graph structure; and determining the target picture set based on the first value of similarity corresponding to each candidate picture and a similarity threshold. Compared with a conventional convolutional neural network, a graph convolutional neural network can better perform unique node classification and edge prediction on irregular graph data, and therefore can be more widely employed.

[0007] In some possible implementations, determining the first value of similarity corresponding to each candidate picture in the candidate picture set through the classifier according to the second graph structure may include: adding and fusing the first graph structure and the second graph structure to obtain a third graph structure; and determining the first value of similarity corresponding to each candidate picture in the candidate picture set through the classifier according to the third graph structure. The original first graph structure and the second graph structure are added and fused, so that impact on an entire graph structure due to unfavorable factors including parameter uncontrollability and instability in an optimization process can be reduced.

[0008] In some possible implementations, the graph association update sub-network may include an attention mechanism layer, a plurality of graph convolution layers, a plurality of activation layers, and a plurality of fully-connected layers. The attention mechanism layer, the plurality of graph convolution layers, the plurality of activation layers and the plurality of fully-connected layers are serially connected. The operation of inputting the first graph structure into the graph association update sub-network to obtain an updated and optimized second graph structure may include: inputting the first graph structure into the attention mechanism layer to obtain a weight vector of each node in the first graph structure; determining the weight vector of each node and the first graph structure as an input of a next layer of the attention mechanism layer; determining any currently processing layer in the plurality of graph convolution layers, the plurality of activation layers and the plurality of fully-connected layers as a current layer; using an output of a previous layer of the current layer as an input of the current layer, and obtaining an output of the current layer after calculation processing; and in a case that any one of the current layers has a corresponding output, obtaining the updated and

optimized second graph structure according to an output of the last layer in the graph association update sub-network. A quantity of layers of the graph association update sub-network and a position relationship thereof are adjusted, so that a network architecture can be flexibly established for various application scenarios, to obtain the second graph structure that better meets requirements.

[0009] In some possible implementations, the operation of determining a candidate picture set from the to-be-processed picture set according to a first feature value of the target object picture and a second feature value set corresponding to the to-be-processed picture set may include: determining, based on a feature code extraction network, a first feature value of the target object included in the target object picture; determining, based on the feature code extraction network, a second feature value of an object included in each picture in the to-be-processed picture set; determining, according to the second feature value and the first feature value, a second value of similarity corresponding to each picture; and determining the candidate picture set from the to-be-processed picture set according to the second value of similarity. The candidate picture set can be preliminarily and accurately determined from the candidate picture set based on a similarity between feature values, to pave the way for subsequent picture processing.

[0010] In some possible implementations, the operation of determining the candidate picture set from the to-be-processed picture set according to the second value of similarity may include: ranking second values of similarity corresponding to all to-be-processed pictures in a descending order of the values; and obtaining the candidate picture set based on to-be-processed pictures corresponding to N top-ranked second values of similarity. Ranking of the second values of similarity is performed to correspond to the step of selecting the preset N pictures, thereby improving diversity of this implementation solution.

[0011] In some possible implementations, the operation of determining the candidate picture set from the to-be-processed picture set according to the second value of similarity may include: ranking second values of similarity corresponding to all to-be-processed pictures in a descending order of the values; dividing the to-be-processed picture set into a first candidate picture set and a non-first candidate picture set based on to-be-processed pictures corresponding to N1 top-ranked second values of similarity, the first candidate picture set including the pictures corresponding to the N1 top-ranked

second values of similarity; determining N2 pictures from the non-first candidate picture set according to a second feature value of a picture in the first candidate picture set and a second feature value of a picture in the non-first candidate picture set, to form a second candidate picture set; and determining the candidate picture set based on the first candidate picture set and the second candidate picture set. Compared with selecting the candidate picture set at one time, in this implementation, the candidate picture set is gradually determined through two searches, so that more hard positive sample pictures can be inputted into the candidate picture set, thereby paving the way for improving accuracy of subsequent picture identification and improving diversity of this implementation solution.

[0012] In some possible implementations, the operation of determining N2 pictures from the non-first candidate picture set according to a second feature value of a picture in the first candidate picture set and a second feature value of a picture in the non-first candidate picture set, to form a second candidate picture set may include: determining any currently used picture in the first candidate picture set as a current picture; determining, according to a second feature value of the current picture and the second feature value of the picture in the non-first candidate picture set, a third value of similarity corresponding to each picture in the non-first candidate picture set; determining, according to the third value of similarity corresponding to each picture, a third candidate picture set corresponding to the current picture from the non-first candidate picture set; and in a case that each of the current pictures corresponds to a third candidate picture set, determining the N2 pictures according to the third candidate picture set corresponding to each of the current pictures, to form the second candidate picture set. Two searches are performed on the picture in the first candidate picture set, so that the picture in the first candidate picture set is used as a transition picture, and furthermore, more positive sample pictures can be obtained to determine the candidate picture set, thereby paving the way for improving accuracy of subsequent picture identification.

[0013] In some possible implementations, after the operation of determining a target picture set from the candidate picture set, the method further may include: determining attribute information of the picture in the target picture set; and performing, according to the attribute information, trajectory and behavior analysis on the object included in the

picture in the target picture set. In this way, the picture in the target picture set can be applied to actual scenarios according to the attribute information.

[0014] In some possible implementations, the attribute information may include a picture acquisition position and a picture acquisition time. The performing, according to the attribute information, trajectory and behavior analysis on the object included in the picture in the target picture set may include: ranking pictures in the target picture set according to the picture acquisition time; and performing, based on the picture acquisition position and ranked pictures, motion trajectory determination and behavior speculation on the object included in the picture in the target picture set. How to perform trajectory and behavior analysis on the object based on the included attribute information is specified, so that the obtained target picture set can be applied to specific scenarios, and an actual problem in life is resolved.

[0015] According to a second aspect of the embodiments of the disclosure, a target re-identification apparatus is provided, including: a picture acquiring module, configured to acquire a target object picture and a to-be-processed picture set, the target object picture including a target object; a candidate picture determining module, configured to determine a candidate picture set from the to-be-processed picture set according to a first feature value of the target object picture and a second feature value set corresponding to the to-be-processed picture set, a value of similarity between any two pictures in the candidate picture set being greater than or equal to a preset similarity value; and a target picture determining module, configured to: identify the first feature value and the second feature value set based on a trained graph association identification network, and determine a target picture set from the candidate picture set, where a first value of similarity between an object included in a picture in the target picture set and the target object is greater than or equal to a first value of similarity between an object included in a non-target picture and the target object, and the candidate picture set includes the target picture set and the non-target picture.

[0016] According to a third aspect of the embodiments of the disclosure, an electronic device is provided, including at least one processor and a memory communicably connected to the at least one processor, where the memory stores instructions executed by the at least one processor, and the at least one processor executes the instructions

stored in the memory, to implement the target re-identification method in any one implementation of the first aspect.

[0017] According to a fourth aspect of the embodiments of the disclosure, a computer readable storage medium is provided, having at least one instruction or at least one program stored thereon, the at least one instruction or the at least one program being loaded and executed by a processor to implement the target re-identification method in any one implementation of the first aspect.

[0018] According to a fifth aspect of the embodiments of the disclosure, a computer program product including instructions is provided, when running on a computer, the instructions causing the computer to perform any target re-identification method provided in the first aspect of the embodiments of the disclosure.

[0019] In the embodiments of the disclosure, two times of picture deletion and selection are performed, including: determining a candidate picture set from a to-be-processed picture set based on a first feature value of a target object picture and a second feature value set corresponding to the to-be-processed picture set, where a value of similarity between any two pictures in the candidate picture set is greater than or equal to a preset similarity value; and identifying the first feature value and the second feature value set based on a trained graph association identification network, and determining, from the candidate picture set, a target picture set that is relatively similar to a target object in the target object picture. In this way, a more accurate positive sample can be determined from the to-be-processed picture set, and interference of a negative sample can be reduced, to obtain the target picture set, so that accuracy of a result of subsequently performing trajectory and behavior analysis on an object included in a picture in the target picture set based on attribute information of the picture in the target picture set is improved.

[0020] It should be understood that the foregoing general descriptions and the following detailed descriptions are merely examples and explanations, but are not intended to limit the embodiments of the disclosure.

[0021] Exemplary embodiments are described in detail below according to the following reference accompanying drawings, and other features and aspects of the

embodiments of the disclosure become clear.

BRIEF DESCRIPTION OF THE DRAWINGS

[0022] To describe the technical solutions and the advantages in the embodiments of the specification or in the prior art more clearly, the following briefly introduces the accompanying drawings required for describing the embodiments or the prior art. Apparently, the accompanying drawings in the following description show merely some embodiments of the specification, and a person of ordinary skill in the art may still derive other drawings from these accompanying drawings without creative efforts.

[0023] FIG. 1 is a flowchart of a target re-identification method according to an embodiment of the disclosure.

[0024] FIG. 2 is a flowchart of a method for acquiring a candidate picture set according to an embodiment of the disclosure.

[0025] FIG. 3 is a flowchart of a method for acquiring a candidate picture set according to an embodiment of the disclosure.

[0026] FIG. 4 is a flowchart of a method for acquiring a second candidate picture set according to an embodiment of the disclosure.

[0027] FIG. 5 is a schematic diagram of acquiring a candidate picture set according to an embodiment of the disclosure.

[0028] FIG. 6 is a schematic structural diagram of a graph association identification network according to an embodiment of the disclosure.

[0029] FIG. 7 is a flowchart of a method for determining a target picture set based on a graph association identification network according to an embodiment of the disclosure.

[0030] FIG. 8 is a schematic diagram of a first graph structure according to an embodiment of the disclosure.

- [0031] FIG. 9 is a schematic diagram of a residual structure according to an embodiment of the disclosure.
- [0032] FIG. 10 is a flowchart of a method for training a graph association identification network according to an embodiment of the disclosure.
- [0033] FIG. 11A is an application flowchart of a target re-identification method according to an embodiment of the disclosure.
- [0034] FIG. 11B is a logic flowchart of a pedestrian re-identification method according to an embodiment of the disclosure.
- [0035] FIG. 11C is a learning framework of an association feature according to an embodiment of the disclosure.
- [0036] FIG. 11D is a schematic diagram of selecting a hard sample from a gallery according to an embodiment of the disclosure.
- [0037] FIG. 12 is a block diagram of a target re-identification apparatus according to an embodiment of the disclosure.
- [0038] FIG. 13 is a block diagram of an electronic device according to an embodiment of the disclosure.
- [0039] FIG. 14 is a block diagram of another electronic device according to an embodiment of the disclosure.

DETAILED DESCRIPTION

- [0040] The following clearly and completely describes the technical solutions in the embodiments of the specification with reference to the accompanying drawings in the embodiments of the specification. Apparently, the described embodiments are merely some but not all of the embodiments of the specification. All other embodiments obtained by a person of ordinary skill in the art based on the embodiments of the specification without creative efforts shall fall within the protection scope of the disclosure.

[0041] It should be noted that, the terms "first", "second", and so on in the specification, claims, and accompanying drawings of the disclosure are intended to distinguish between similar objects but do not necessarily indicate a specific order or sequence. It should be understood that, the data termed in such a way are interchangeable in proper circumstances so that the embodiments of the disclosure described herein can be implemented in other orders than the order illustrated or described herein. In addition, the terms "include", "contain", and any other variants thereof are intended to cover non-exclusive inclusion. For example, a process, method, system, product, or server that includes a series of steps or units is unnecessarily limited to those expressly listed steps or units, but may include other steps or units that are not expressly listed or that are inherent to such a process, method, system, product, or device.

[0042] The following will describe various exemplary embodiments, features, and aspects of the embodiments of the disclosure in detail with reference to the accompanying drawings. Like accompanying symbols in the accompanying drawings represent elements with like or similar functions. Although various aspects of the embodiments are illustrated in the accompanying drawing, the accompanying drawings are not necessarily drawn in proportion unless otherwise specified.

[0043] The special term "exemplary" here means "used as an example, an embodiment, or an illustration". Any embodiment described as "exemplary" here is not necessarily to be interpreted as superior to or better than other embodiments.

[0044] The term "and/or" herein describes only an association relationship for describing associated objects and represents that three relationships may exist. For example, A and/or B may represent the following three cases: only A exists, both A and B exist, and only B exists. In addition, the term "at least one" herein means any one of a plurality of cases or any combination of at least two of a plurality of cases. For example, including at least one of A, B, or C may mean including any one or more elements selected from a set consisting of A, B, and C.

[0045] In addition, for better illustration of the embodiments of the disclosure, various implementation details are given in the following specific implementations. A person skilled in the art should understand that the embodiments of the disclosure may also be implemented without some implementation details. In some examples, methods, means,

elements, and circuits well known to a person skilled in the art are not described in detail so as to highlight the subject matter of the embodiments of the disclosure.

[0046] According to the target re-identification solution provided in the embodiments of the disclosure, a target object picture and a to-be-processed picture set are acquired. The target object picture includes a target object. A candidate picture set is determined from the to-be-processed picture set according to a first feature value of the target object picture and a second feature value set corresponding to the to-be-processed picture set. A value of similarity between any two pictures in the candidate picture set is greater than or equal to a preset similarity value. The first feature value and the second feature value set are identified based on a trained graph association identification network, and a target picture set is determined from the candidate picture set. A first value of similarity between an object included in a picture in the target picture set and the target object is greater than or equal to a first value of similarity between an object included in a non-target picture and the target object. The candidate picture set includes the target picture set and the non-target picture. In this way, through the two times of picture deletion and selection, impact of various factors such as illumination and background complexity can be reduced, a more accurate positive sample can be determined from the to-be-processed picture set, and interference of a negative sample can be reduced, to obtain the target picture set, so that accuracy of performing trajectory and behavior analysis on the object included in the picture in the target picture set based on attribute information of the picture in the target picture set is improved.

[0047] In practical applications of the related art, because a picture in the to-be-processed picture set is affected by illumination intensity, a background disorder, or a viewing angle change of a picture acquiring device, a relatively large quantity of interfered negative samples are used or positive samples that are relatively difficult to be identified are neglected in an existing modeling process, so that network precision obtained through modeling is not high. Therefore, in an application process, picture selection accuracy is not high, and therefore accuracy of performing trajectory and behavior analysis on the object is affected. According to the target re-identification method provided in the embodiments of the disclosure, the candidate picture set is obtained by comparing feature values of the target object picture and an initial picture, and the target picture set with a higher similarity to the target object of the target object

picture is identified from the candidate picture set based on the graph association identification network, thereby improving accuracy of a to-be-analyzed picture. In this way, sufficient trajectory and behavior analysis can be performed on an object in the selected target picture set.

[0048] The technical solutions provided in the embodiments of the disclosure may be applied to expansion of application scenarios such as target re-identification and target identification of images or videos, which are not limited in the embodiments of the disclosure.

[0049] The target re-identification method provided in the embodiments of the disclosure may be performed by a terminal device, a server, or other types of electronic devices. The terminal device may be User Equipment (UE), a mobile device, a user terminal, a terminal, a cellular phone, a cordless phone, a Personal Digital Assistant (PDA), a handheld device, a computing device, an in-vehicle device, a wearable device, or the like. In some possible implementations, the target re-identification method may be implemented in a manner that a processor invokes computer readable instructions stored in a memory. The following uses an electronic device as an execution body to describe the target re-identification method in the embodiments of the disclosure. For example, the target re-identification method may be implemented in the manner that the processor invokes the computer readable instructions stored in the memory.

[0050] FIG. 1 is a flowchart of a target re-identification method according to an embodiment of the disclosure. As shown in FIG. 1, the method includes the following steps S10 and S20.

[0051] At S10, a target object picture and a to-be-processed picture set are acquired. The target object picture includes a target object.

[0052] In some implementations, the target object may include, but is not limited to, a vehicle, a pedestrian, or a combination of the vehicle and the pedestrian. The vehicle may be a car, a van, a motorcycle, a bicycle, or the like.

[0053] In some implementations, the target object picture may be obtained by using an electronic device, or the electronic device may obtain the target object picture from another device. For example, the electronic device may obtain the target object picture

from a device such as a photographing device or a monitoring device. In some implementations, the target object picture may be a frame in a video. Similarly, the to-be-processed picture set may be obtained by using the electronic device, or may be obtained by using another device and integrated into the electronic device.

[0054] In the embodiments of the disclosure, a target picture is determined from the to-be-processed picture set based on the target object picture, and trajectory and behavior analysis may be performed on an object included in the target picture based on the target picture. Therefore, the electronic device can purposely and selectively acquire some pictures to form the to-be-processed picture set. In some implementations, assuming that the target object picture is acquired by using a camera A, the electronic device may also acquire the picture by using the camera A, and/or acquire some pictures by using at least one camera disposed near the camera A, to form the to-be-processed picture set. In some implementations, assuming that the target object picture is photographed by using the camera A at a moment, the electronic device may also acquire a picture photographed by the camera A before or after the moment, and/or a picture photographed by another camera before or after the moment, to form the to-be-processed picture set. In some implementations, assuming that the target object picture is photographed by using the camera A at a moment, the electronic device may also acquire a picture photographed by the camera A before or after the moment, and/or a picture photographed before or after the moment by another camera disposed near the camera A, to form the to-be-processed picture set. In this way, because many interfered pictures are ruled out in advance, a large amount of calculation can be reduced and device overheads can be saved in a process in which the electronic device performs an operation on the to-be-processed picture set.

[0055] At S20, a candidate picture set is determined from the to-be-processed picture set according to a first feature value of the target object picture and a second feature value set corresponding to the to-be-processed picture set. A value of similarity between any two pictures in the candidate picture set is greater than or equal to a preset similarity value.

[0056] In some implementations, before the candidate picture set is determined from the to-be-processed picture set according to the first feature value of the target object picture and the second feature value set corresponding to the to-be-processed picture set, in the embodiments of the disclosure, pre-deletion and selection may be further

performed on the to-be-processed picture set. In the following descriptions, the target object is a pedestrian. In a process of acquiring the to-be-processed picture set, an object such as a person may not be included in an acquired picture because of an acquiring channel problem or another problem. For example, if a second feature value of a picture in the to-be-processed picture set is directly extracted, device overheads are greatly increased. Therefore, pre-deletion and selection may be performed on the picture in the to-be-processed picture set by using an object identification module disposed in the electronic device, and a picture that does not include a person is deleted, to obtain relatively clean picture data.

[0057] In the embodiments of the disclosure, the candidate picture set is determined from the to-be-processed picture set in a plurality of manners. In some implementations, a picture may be selected from the candidate picture set based on a gender of the target object in the target object picture, to form the candidate picture set. A gender of an object included in a picture in the candidate picture set is the same as the gender of the target object. In some implementations, a picture may also be selected from the candidate picture set based on a gender and a body type of the target object in the target object picture, to form the candidate picture set. A gender and a body type of an object included in a picture in the candidate picture set are the same as the gender and the body type of the target object.

[0058] In some implementations, in the embodiments of the disclosure, the candidate picture set may be acquired by extracting a feature value of a picture, a first feature value of the target object included in the target object picture is determined, a second feature value of an object included in the picture in the to-be-processed picture set is determined, a second value of similarity corresponding to each picture is determined according to the second feature value and the first feature value, and the candidate picture set is determined from the to-be-processed picture set according to the second value of similarity. However, in this manner, no requirement is imposed on the value of similarity between any two pictures in the candidate picture set. That is, in this manner, each picture in the candidate picture set may be only associated with the target object picture.

[0059] In both the foregoing two methods, a feature value of the picture in the to-be-processed picture set is directly compared with that of the target object picture to obtain a second value of similarity. However, in a picture or video photographing process, a

specific quantity of hard positive sample pictures and hard negative sample pictures may exist in the to-be-processed picture set due to various reasons such as illumination, a photographing background, and viewing angle conversion. For example, if these hard sample pictures are not considered in a process of determining the candidate picture set in a previous period, it is highly likely to adversely affect a subsequent picture identification process.

[0060] In the embodiments of the disclosure, a sample picture refers to each picture in the to-be-processed picture set, a positive sample picture refers to a sample picture, in the picture, including an object that is the same as the target object, and a negative sample picture refers to a sample picture, in the picture, including an object that is different from the target object. A hard positive sample picture means that although an object included in the picture is the same as the target object, the object cannot be easily identified by the electronic device due to photographing light, an object posture, or the like. A hard negative sample picture means that although an object included in the picture is different from the target object, the object is easy to be mistakenly considered as the same as the target object due to photographing light, an object posture, or the like. Considering the existence of the hard positive sample picture and the hard negative sample picture, to improve accuracy of subsequent picture identification, FIG. 2 is a flowchart of a method for acquitting a candidate picture set according to an embodiment of the disclosure. As shown in FIG. 2, the method includes the following steps S201 to S204.

[0062] At S201, a first feature value of a target object included in a target object picture is determined based on a feature code extraction network.

[0063] In some implementations, the target object picture is inputted into the feature code extraction network. The feature code extraction network is trained. First, the target object in the target object picture may be framed, and then feature extraction is performed on the framed target object to obtain the first feature value.

[0064] At S202, a second feature value of an object included in a picture in a to-be-processed picture set is determined based on the feature code extraction network.

[0065] In some implementations, an electronic device may uniformly input the picture in the to-be-processed picture set into the feature code extraction network, so that the

feature code extraction network can perform feature extraction on the object in the picture to obtain a second feature value of each picture.

[0066] In some other implementations, considering that a quantity of pictures in the to-be-processed picture set may be large, the feature code extraction network may be built in a plurality of electronic devices, and the to-be-processed picture set is divided into a plurality of subsets. Pictures in each subset are extracted by one electronic device, and then summarized to an initial electronic device.

[0067] The feature code extraction network may be trained by using an unsupervised, supervised, or semi-supervised learning method. In some implementations, in a training process of the feature code extraction network, each training picture that includes an object may be used as a category to perform multi-classification learning. After training, the last classification layer of the network is removed, and an output of the network is used as a feature code. In this implementation, for a feature extraction manner, reference is made to the foregoing feature extraction manner for the target object in the target object picture.

[0068] In some implementations, the first feature value and the second feature value may also be referred to as a first feature code and a second feature code. The first feature value and the second feature value may be outputted in a plurality of forms, for example, may be outputted in a form of vectors, or may be outputted in a form of a plurality of binary numbers. The form of the output may be determined based on an actual requirement, and details are not described herein again.

[0069] At S203, a second value of similarity corresponding to each to-be-processed picture is determined according to the second feature value and the first feature value.

[0070] In some implementations, the electronic device may calculate, based on each second feature value and the first feature value, the second value of similarity of each to-be-processed picture relative to the target object picture. For example, assuming that there are 10,000 pictures in the to-be-processed picture set, 10,000 second feature values corresponding to the 10,000 pictures and the first feature value corresponding to the target object picture may be obtained by using the feature code extraction network. Each second feature value and the first feature value are calculated according to a preset rule,

to obtain 10,000 second values of similarity.

[0071] At S204, a candidate picture set is determined from the to-be-processed picture set according to the second value of similarity.

[0072] In some implementations, the electronic device acquires a preset second similarity threshold, compares the obtained second value of similarity with the second similarity threshold, and determines pictures corresponding to second values of similarity greater than the second similarity threshold to form the candidate picture set, where a value of similarity between any two of the pictures in the candidate picture set is greater than or equal to a preset similarity value.

[0073] In some other implementations, the electronic device may rank second values of similarity corresponding to all to-be-processed pictures in a descending order of the values, and determine pictures corresponding to N top-ranked second values of similarity as pictures in the candidate picture set, where a value of similarity between any two of the pictures in the candidate picture set is greater than or equal to a preset similarity value. For example, N is 100, and 100 pictures are determined from the to-be-processed picture set to form the candidate picture set.

[0074] In the foregoing method, the value of similarity between the any two pictures in the candidate picture set is greater than or equal to the preset similarity value. In some implementations, a value of similarity between any two pictures may be obtained by calculating second feature values of the two pictures. In other words, in this implementation, not only an association between each candidate picture in the candidate picture set and the target object picture needs to be determined by calculating the first feature value and the second feature value, but also an association between pictures in the candidate picture set needs to be established by using the value of similarity of the any two pictures in the candidate picture set. In this way, a quantity of hard positive sample pictures in the candidate picture set can be increased as much as possible, and a quantity of hard negative sample pictures can be reduced. In some implementations, the preset similarity value may be set based on an actual situation.

[0075] In some other implementations, the electronic device may first determine a first candidate picture set from the to-be-processed picture set, determine a second

candidate picture set based on the basis of the first candidate picture set, and combine the foregoing two candidate picture sets into the candidate picture set. FIG. 3 is flowchart of a method for acquiring a candidate picture set according to an embodiment of the disclosure. As shown in FIG. 3, the method includes the following steps S301 to S303.

[0076] At S301, second values of similarity corresponding to all to-be-processed pictures are ranked in a descending order of the values.

[0077] At S302, a to-be-processed picture set is divided into a first candidate picture set and a non-first candidate picture set based on to-be-processed pictures corresponding to N1 top-ranked second values of similarity. The first candidate picture set includes the pictures corresponding to the N1 top-ranked second values of similarity.

[0078] Alternatively, a preset second similarity threshold may be acquired, the obtained second value of similarity is compared with the second similarity threshold, and a to-be-processed picture corresponding to a second value of similarity greater than the second similarity threshold is determined to form the first candidate picture set, and remaining pictures in the to-be-processed picture set form the non-first candidate picture set. In some implementations, no intersection set exists between the first candidate picture set and the non-first candidate picture set.

[0079] At S303, N2 pictures are determined from the non-first candidate picture set according to a second feature value of a picture in the first candidate picture set and a second feature value of a picture in the non-first candidate picture set, to form a second candidate picture set.

[0080] FIG. 4 is flowchart of a method for acquiring a second candidate picture set according to an embodiment of the disclosure. As shown in FIG. 4, the method includes the following steps S401 to S404.

[0081] At S401, any currently used picture in a first candidate picture set is determined as a current picture.

[0082] Assuming that N1 is 10, 10 pictures in the first candidate picture set are determined in a process of the first selection. Each of the 10 pictures undergoes processing at steps S401-S403.

[0083] At S402, a third value of similarity corresponding to each picture in a non-first candidate picture set is determined according to a second feature value of the current picture and a second feature value of a picture in the non-first candidate picture set.

[0084] Continuing with the assumption in which there are a total of 10,000 pictures in the to-be-processed picture set, because the first candidate picture set includes 10 pictures in the foregoing example, the non-first candidate picture set further includes 9990 pictures. In this step, an electronic device obtains, according to the second feature value of the current picture and second feature values of the 9990 pictures, third values of similarity of the 9990 pictures relative to the current picture.

[0085] At S403, a third candidate picture set corresponding to the current picture is determined from the non-first candidate picture set according to the third value of similarity corresponding to each picture.

[0086] In some implementations, the electronic device may preset a third similarity threshold, and determine a picture corresponding to a third value of similarity greater than the third similarity threshold into the third candidate picture set corresponding to the current picture.

[0087] In some implementations, the electronic device ranks the 9990 third values of similarity, and determines pictures corresponding to the first several third values of similarity into the third candidate picture set corresponding to the current picture.

[0088] At S404, in a case that each of the current pictures has a corresponding third candidate picture set, N2 pictures are determined according to the third candidate picture set corresponding to each of the current pictures, to form a second candidate picture set.

[0089] In this way, in a case that each of the current pictures has a corresponding third candidate picture set, that is, in a case that a picture in 10 pictures in the first candidate picture set meets a similarity requirement thereof, the N2 pictures are determined according to the third candidate picture set corresponding to each of the current pictures, to form the second candidate picture set.

[0090] In some implementations, there is a duplicate picture in the third candidate picture set corresponding to different pictures in the first candidate picture set. For the

duplicate picture, after a candidate picture set is formed, the candidate picture set is re-checked to delete the duplicate picture. In some implementations, after a candidate picture set is formed, the candidate picture set is re-checked to delete the duplicate picture, and the second candidate picture set may further be supplemented with pictures according to the third value of similarity until the N2 pictures that meet the requirement are determined. In some implementations, a sum of N1 and N2 may be N.

[0091] At S304, a candidate picture set is determined based on the first candidate picture set and the second candidate picture set.

[0092] In the embodiments of the disclosure, the first candidate picture set and the second candidate picture set may be combined to obtain a candidate picture set. A value of similarity between any two pictures in the candidate picture set is greater than or equal to a preset similarity value.

[0093] FIG. 5 is a schematic diagram of obtaining a candidate picture set according to an embodiment of the disclosure. As shown in FIG. 5, an initial picture set 51 is screened based on a target object picture, to obtain a picture in a first candidate set and a picture in a second candidate set, and the two parts of pictures form a candidate picture set 52. In the foregoing embodiment, the picture in the first candidate picture set may be considered as a transition picture. For example, the target object picture is a front picture of a pedestrian, the picture in the first candidate picture set may be a side picture of the pedestrian, and the picture in the second candidate picture set may be a back picture of the pedestrian. In comparison with the front picture, because there may be more similarities of the pedestrian in the back picture and the side picture, it is more likely that the back picture is determined based on the side picture. Therefore, in this implementation, an electronic device limits a maximum quantity of searches through two searches, uses a similarity relationship between pictures in a to-be-processed picture set instead of merely using a relationship between a picture and the target object picture, to dig a hard positive sample picture as much as possible, thereby improving a probability of a positive sample in the candidate picture set and paving the way for subsequent picture processing.

[0094] At S30, the first feature value and the second feature value set are identified based on a trained graph association identification network, and a target picture set is

determined from the candidate picture set. A first value of similarity between an object included in a picture in the target picture set and the target object is greater than or equal to a first value of similarity between an object included in a non-target picture and the target object. The candidate picture set includes the target picture set and the non-target picture.

[0095] In some implementations, the graph association identification network may include, but is not limited to, a deep learning network such as a convolutional neural network, a recurrent neural network, or a recursive neural network. Taking the convolutional neural network as an example, a large quantity of training data sets may be acquired. Each training data set includes a first feature value of a target object picture and a second feature value of a candidate picture, as well as an annotated target picture. Then, target picture identification training is performed on the convolutional neural network based on the large quantity of training data sets, and a parameter of the convolutional neural network is adjusted in training to enable a target picture outputted by the convolutional neural network to match with the annotated target picture, to obtain the graph association identification network.

[0096] In some implementations, the graph association identification network may include, but is not limited to a graph convolutional neural network. This is because in real life, there are many irregular data structures, typically a first graph structure, or referred to as a topology structure, such as a social network, a chemical molecular structure, or a knowledge graph. Even if in the case of a language, the language actually has a complex tree structure inside, which is also the first graph structure. In the case of a picture, when target identification is performed, only a part of key points in a two-dimensional picture are concerned, and these points also form a structure of a graph. The structure of the graph is generally very irregular and may be considered as data with an infinite dimension, so it has no translation invariance. A surrounding structure of each node may be unique. Data of this structure causes a poor application effect of a conventional convolutional neural network. However, the graph convolutional neural network subtly designs a method for extracting a feature from graph data, so that these features can be used, to perform node classification, graph classification, and link prediction on graph data, and can also be used, to incidentally obtain a graph embedding representation. This is widely and properly used.

[0097] FIG. 6 is a schematic structural diagram of a graph association identification network according to an embodiment of the disclosure. As shown in FIG. 6, the graph association identification network may include a first graph structure establishment sub-network 61, a graph association update sub-network 62, and a classifier 63. The first graph structure establishment sub-network 61, the graph association update sub-network 62 and the classifier 63 are serially connected. First, a first feature value and a second feature value set 601 are inputted into the first graph structure establishment sub-network 61 to obtain a first graph structure 602, and then the first graph structure 602 is inputted into the graph association update sub-network 62 to obtain a second graph structure 603, and finally the second graph structure 602 is inputted into the classifier 63 to obtain a target graph set.

[0098] FIG. 7 is a schematic diagram of a method for determining a target picture set based on a graph association identification network according to an embodiment of the disclosure. As shown in FIG. 7, the method includes the following steps S701 to S704.

[0099] At S701, the first feature value and the second feature value set are inputted into the first graph structure establishment sub-network to obtain a first graph structure. The first graph structure includes nodes and edges each of which is for connecting two nodes. A quantity of nodes is the same as a quantity of pictures in the candidate picture set. The edge for connecting the two nodes is determined based on a similarity between the two nodes that are connected and a preset similarity.

[00100] In some implementations, continuing with the example in which there are 100 pictures in the candidate picture set, the electronic device inputs the first feature value and the second feature value set into the first graph structure establishment sub-network. The first graph structure establishment sub-network calculates a difference between each second feature value and the first feature value to obtain an association feature value corresponding to each second feature value. Each association feature value refers to an association relationship between a corresponding picture and the target object picture. Each association feature value is defined as one node, and therefore 100 nodes may be determined. A similarity between any two nodes is determined based on association feature values corresponding to the any two nodes. According to a permutation and combination formula, a similarity between two nodes needs to be calculated $100 \times 99 / 2 = 4950$ times. If a similarity is greater than the preset similarity, an edge may be

formed between two nodes corresponding to the similarity. In this way, a first graph structure shown in FIG. 8 can be obtained. The first graph structure shown in FIG. 8 merely shows part of nodes 81 and sides 82 between nodes.

[00101] At S702, the first graph structure is inputted into the graph association update sub-network to obtain an updated and optimized second graph structure.

[00102] In some implementations, the graph association update sub-network may include a plurality of graph convolution layers, a plurality of activation layers, and a plurality of fully-connected layers. The plurality of graph convolution layers, the plurality of activation layers and the plurality of fully-connected layers are serially connected.

[00103] In some implementations, a quantity of graph convolution layers and a quantity of fully-connected layers may be the same, and one activation layer follows each graph convolution layer. For example, the following form may be presented: Graph convolution layer - Activation layer - Fully-connected layer - Graph convolution layer - Activation layer - Fully-connected layer - Graph convolution layer - Activation layer - Fully-connected layer ... Fully-connected layer.

[00104] In some implementations, there may be a different quantity of graph convolution layers and fully-connected layers, and one activation layer follows each graph convolution layer and each fully-connected layer. For example, the following form may be presented: Graph convolution layer - Activation layer - Graph convolution layer - Activation layer - Graph convolution layer - Activation layer ... Fully-connected layer - Activation layer - Fully-connected layer - Activation layer ... Fully-connected layer - Activation layer.

[00105] Quantities of graph convolution layers, fully-connected layers, and activation layers and position relationships thereof in the association update sub-network may be set based on an actual requirement. For example, nine graph convolution layers may be set.

[00106] In some implementations, an attention mechanism may be added to a graph convolution layer to enhance reasoning validity of the graph convolution layer, so that in the convolution process, an association between nodes corresponding to two positive samples is enhanced, and an association between nodes corresponding to a positive

sample and a negative sample is reduced. Thus, the graph association update sub-network includes an attention mechanism layer, a plurality of graph convolution layers, a plurality of activation layers, and a plurality of fully-connected layers. The attention mechanism layer, the plurality of graph convolution layers, the plurality of activation layers and the plurality of fully-connected layers are serially connected. The number of attention mechanism layers may be set based on an actual situation. In some implementations, there may be only one attention mechanism layer, and the attention mechanism layer may be disposed in front of the first picture convolution layer. In some implementations, one attention mechanism layer may be disposed in front of each graph convolution layer.

[00107] In some implementations, assuming that the attention mechanism layer is disposed only in front of the first graph convolution layer, that the first graph structure is inputted into the graph association update sub-network to obtain an updated and optimized second graph structure may be represented as follows. A weight vector of each node and the first graph structure are determined as an input of a next layer of the attention mechanism layer. Any currently processing layer in the plurality of graph convolution layers, the plurality of activation layers and the plurality of fully-connected layers is determined as a current layer. An output of a previous layer of the current layer is used as an input of the current layer, and an output of the current layer is obtained after calculation processing. In a case that any one of the current layers has a corresponding output, the updated and optimized second graph structure is obtained according to an output of the last layer in the graph association update sub-network. For the calculation process of disposing the attention mechanism layer in front of each graph convolution layer, reference is made to the foregoing calculation process, and details are not described herein again.

[00108] In some implementations, the gradient vanishing problem and the gradient exploding problem may exist due to a network depth in the deep learning network. Normalized initialization and batch normalization of data may be used, to resolve the problems related to the gradient. However, because the depth is deepened, another problem is caused, that is, network performance degrades. In other words, the network depth is deepened, and an error rate increases. Therefore, a residual structure may be used, to resolve the degradation problem, and network performance is also improved while the gradient problem is resolved. As shown in FIG. 9, the residual structure may

include a graph convolution layer 91, a batch normalization mechanism 92, and an activation layer 93. A result obtained after the inputted original data successively passes through the graph convolution layer 91 and the batch normalization mechanism 92 and the original data is added to be inputted into the activation layer to obtain final target data.

[00109] At S703, a first value of similarity corresponding to each candidate picture in the candidate picture set is determined by the classifier according to the second graph structure.

[00110] In some possible implementations, the first graph structure and the second graph structure may be added and fused to obtain a third graph structure, and the first value of similarity corresponding to each candidate picture in the candidate picture set is determined by the classifier according to the third graph structure. A value corresponding to an i^{th} node in the first graph structure and a value corresponding to an i^{th} node in the second graph structure may be added to obtain a value corresponding to an i^{th} node in the third graph structure, and the structures remain unchanged to obtain the third graph structure. Alternatively, a value corresponding to an i^{th} node in the first graph structure and a value corresponding to an i^{th} node in the second graph structure may be added and averaged to obtain a value corresponding to an i^{th} node in the third graph structure, and the structures remain unchanged to obtain the third graph structure. Alternatively, a value corresponding to an i^{th} node in the first graph structure and a value corresponding to an i^{th} node in the second graph structure may be weighted and added to obtain a value corresponding to an i^{th} node in the third graph structure, and the structures remain unchanged to obtain the third graph structure. The i^{th} node in the first graph structure, the i^{th} node in the second graph structure, and the i^{th} node in the second graph structure are all nodes corresponding to a same picture.

[00111] At S704, the target picture set is determined based on the first value of similarity corresponding to each candidate picture and a similarity threshold.

[00112] A first value of similarity between an object included in a picture in the target picture set and the target object is greater than or equal to a first value of similarity between an object included in a non-target picture and the target object.

- [00113] The embodiments of the disclosure further provide a method for training a graph association identification network. As shown in FIG. 10, the method includes the following steps S1001 to S1007.
- [00114] At S1001, an electronic device acquires a training sample data set. The training sample data set includes first feature values corresponding to a plurality of reference pictures, a second feature value set corresponding to each first feature value, and a first similarity value set corresponding to the second feature value set.
- [00115] At S1002, the electronic device constructs a preset machine learning network, and determines the preset machine learning network as a current machine learning network.
- [00116] At S1003, the electronic device performs, based on the current machine learning network, association identification on the first feature values and the second feature value set corresponding to each first feature value, to determine a predicted first similarity set.
- [00117] At S1004, the electronic device determines a loss value based on the first similarity value set corresponding to the second feature value set and the predicted first similarity set.
- [00118] At S1005, the electronic device determines whether the loss value is greater than a preset threshold.
- [00119] If it is determined that the loss value is greater than the preset threshold, step S1006 is proceeded. If it is determined that the loss value is less than or equal to the preset threshold, step S1007 is proceeded.
- [00120] At S1006, the electronic device performs back propagation based on the loss value, updates the current machine learning network to obtain an updated machine learning network, and re-determines the updated machine learning network as the current machine learning network, and S1003 is proceeded.
- [00121] At S1007, the electronic device determines the current machine learning network as a graph association identification network.

[00122] FIG. 11A is an application flowchart of a target re-identification method according to an embodiment of the disclosure. As shown in FIG. 11A, in addition to S10 to S30, the method further includes the following steps S40 and S50.

[00123] At S40, attribute information of the picture in the target picture set is determined.

[00124] The electronic device may determine the attribute information of the picture based on the picture. The attribute information may include a picture acquisition position and a picture acquisition time. The acquisition position may include, but is not limited to, position information of a device for photographing the picture, and may further include position information of a scenario presented in the picture. The picture acquisition time includes, but is not limited to, a picture photographing time.

[00125] At S50, trajectory and behavior analysis is performed on the object included in the picture in the target picture set according to the attribute information.

[00126] Because the determined object in the picture in the target picture set is basically considered as the same as the target object, pictures in the target picture set are temporally ranked according to the picture acquisition time. Motion trajectory determination and behavior speculation are performed on the object included in the picture based on the picture acquisition position and the ranked pictures. For example, places where the object passed through and things that the object did in a specific period of time are determined. Based on the things determined by the picture, things that the object may do later are speculated and analyzed to obtain an analysis result.

[00127] A person skilled in the art may understand that, in the specific implementation of the foregoing method, a writing sequence of the steps does not mean a strict execution sequence for constituting any limitation on an implementation process. An actual execution sequence of the steps should be determined based on functions and possible internal logic thereof.

[00128] There is a probe data set and a gallery data set in pedestrian re-identification, so that pictures belong to a same pedestrian are searched from all galleries for each target picture. However, in practical applications, modeling is complex during target re-identification due to factors such as strong illumination, a background disorder, and angle

conversion. Many existing methods are mainly limited to appearance feature information of a learning target. However, appearance of the target may be interfered with by another target and an environment background in a scenario.

[00129] Modeling in pedestrian re-identification is complex and may be affected by many factors. Pedestrian re-identification may be interfered with by other pedestrians in the scenario, and there may be extremely similar appearance features between pedestrians. In existing modeling, only a similarity relationship between two pedestrians is considered, and a relationship between a potentially hard positive sample and hard negative sample and a target pedestrian is ignored. This relationship is modeled by using a graph convolutional neural network in the embodiments of the disclosure, so that similarity information among all pedestrians in the gallery can be fully considered. In addition, an efficient single-trigger re-ranking algorithm is proposed, so that re-ranking can be implemented only by using a single target picture.

[00130] The embodiments of the disclosure are described by using re-identification prediction for pedestrians as an example. The following three steps may be performed. First, a feature network is trained to perform feature coding on all targets and gallery pictures. Then, for each target picture, a candidate target feature is searched from the gallery based on a similarity, an association feature is calculated, and a graph structure is established. Finally, the association feature is optimized by using the graph convolutional neural network, and a final candidate target sequence is predicted based on an optimized association feature. In the embodiments of the disclosure, similarity information between candidate targets may be fully used, an association feature between candidate targets is further optimized by using the graph convolutional neural network, and then re-ranking is performed based on an optimized feature to obtain a better prediction sequence. In a broader sense, the solutions provided in the embodiments of the disclosure are applicable to general search tasks. A better search sequence may be obtained by using this solution, which is more efficient in practical applications than a conventional re-ranking algorithm. Moreover, the algorithm is scalable and may be combined with the conventional re-ranking algorithm to obtain higher search precision.

[00131] In the embodiments of the disclosure, a framework of the graph convolutional neural network is used, to resolve a repositioning prediction problem of a target pedestrian. Many factors that affect pedestrian re-identification may be modeled by using

a powerful description capability of the graph convolutional neural network. Through a feature of graph convolution, an association feature of a candidate pedestrian similar to the target pedestrian is used as an input of a deep learning network, and the association feature is optimized and learned by using a graph convolution operation. A graph convolutional learning module for the association feature proposed in the embodiments of the disclosure may re-rank the found candidate pedestrian features to improve search precision. Moreover, a framework for graph convolutional learning may be decoupled with a preconfigured feature network or collaborative learning may be performed, and may be rapidly deployed in practice.

[00132] FIG. 11B is a logical flowchart of a pedestrian re-identification method according to an embodiment of the disclosure. As shown in FIG. 11B, the process includes the following steps S1101 to S1103.

[00133] At S1101, a target picture is inputted into a deep neural network to obtain a feature code.

[00134] In this process, a visual picture of all pedestrians in a scenario is used as an input of a target picture to obtain a visual feature code that is configured to describe all the pedestrians. The feature code is configured to calculate an association feature and is used as an input of the deep neural network.

[00135] In implementation, the following steps may be performed. First, in the entire scenario, a small target graph of each target pedestrian is extracted through target detection or the like. Then, for each small target graph, a feature extraction network is trained and the feature code is extracted. During training of the feature extraction network, each target pedestrian is used as one category to perform multi-classification learning. After training, the following classification layer is removed, and an output of the network is used as the feature code.

[00136] At S1102, an association feature between the target picture and a gallery picture is calculated based on the feature code, and a graph structure is established.

[00137] A feature code matrix obtained through the foregoing operation represents a visual feature of each target pedestrian picture. For a candidate feature found from a gallery, the graph structure is established based on the association feature to describe

information about a rich similarity relationship between candidate pedestrians.

[00138] First, hard gallery sampling is performed to mine hard positive samples as much as possible. Specifically, the sampling is performed in a manner where a maximum quantity of searches is limited through quadratic search and similarity relationships between galleries are used. Then a graph structure is established to represent overall information about a similarity relationship between the target picture and a candidate gallery. An association feature between the target picture and the candidate gallery is used as a node of the graph structure, and similarity information between candidate galleries determines a connection status of an edge. To simplify calculation, a representation form of the association feature is interpolation of a feature code of the target picture and the candidate gallery.

[00139] FIG. 11C is a learning framework of an association feature according to an embodiment of the disclosure. As shown in FIG. 11C, a target picture 111a and a gallery picture 111b are inputted into a feature extraction network 112 to obtain a target feature of the target picture and a feature of the gallery picture, and then association gallery pictures 113 similar to the target picture are obtained through a search process of the target picture to the gallery picture (also referred to as Probe to Gallery, P2G). Then, the association gallery pictures 113 are passed to an HGS sampler to obtain gallery candidate pictures 114, the gallery candidate pictures 114 and the target picture 111a are processed to obtain candidate pictures 115 that form a graph structure, and a target feature graph 111 is removed from the candidate pictures 115 of the graph structure to obtain a graph node 116a. Moreover, a feature matrix 116b between different gallery pictures in the association gallery pictures 113 is obtained through a search process of a gallery picture to a gallery picture (also referred to as Gallery to Gallery, G2G), then an image edge is determined by using the feature matrix 116b in a graph reasoning phase, and a graph structure 117 is generated in combination with a graph node 126a. Then an optimized graph structure 119 is obtained by using a Graph Convolution Network (GCN) 118. Finally, the graph structure 117 and the optimized graph structure 119 are inputted into a regression prediction network 120 for prediction, to obtain a prediction probability of each graph node.

[00140] FIG. 11D is a schematic diagram of selecting a hard sample from a gallery according to an embodiment of the disclosure. As shown in FIG. 11D, for the target

picture 111, the first four gallery pictures associated with the target picture in the association gallery picture 113 are selected by screening, then the first two gallery pictures that are most similar to the first four gallery pictures are selected by screening based on a similarity between gallery pictures, and the six gallery pictures form a gallery candidate picture 114. The gallery candidate picture 114 includes a hard negative sample A of the target picture.

[00141] At S1103, reasoning optimization is performed on the association feature by using a graph convolution network, and an optimized prediction sequence is output.

[00142] Graph convolution reasoning is performed by using a depth graph convolutional neural network. Unlike a conventional convolution network, a graph convolution structure can better reflect a topology relationship between neighboring nodes in semantics. In the embodiments of the disclosure, nine graph convolution layers are used. To further enhance reasoning validity, an attention mechanism is utilized to optimize each association feature, and a better search sequence may be obtained by using an optimized result. In particular, a residual structure is utilized to suppress training hardies due to network over-depth. For the optimized association feature, a final search sequence may be determined through further classification.

[00143] The related art is mainly based on some conventional convolution neural networks or shallow graph networks, and similarity information of a single sample in a gallery is mainly considered. Moreover, in a conventional re-ranking algorithm, because re-ranking usually needs to be simultaneously performed on a large quantity of target pictures, the running efficiency is slow and the practical value is low. However, in the embodiments of the disclosure, a depth graph convolutional neural network is utilized to perform target re-identification, and similarity information between samples is more fully utilized, thereby performing better comprehensive analysis on various influencing factors. In the embodiments of the disclosure, similarity information between all pedestrians in a gallery may be fully considered, and an association feature and a hard sample mining technology based on two searches are used, so that association information between samples can be more fully learned in high-dimensional feature space. The embodiments of the disclosure provide an efficient single-trigger re-ranking algorithm. Re-ranking can be implemented only by using a single target picture, so that it can be flexibly applied to an existing re-ranking, and performance is stably improved.

- [00144] The embodiments of the disclosure may be applied to a video surveillance scenario to predict positions of all pedestrians in a same camera or different cameras. Moreover, a motion trajectory of a target pedestrian across cameras within a period of time is analyzed based on a prediction result, to implement cross-camera target tracing, and some abnormal behavior in a scenario may be further detected.
- [00145] In the embodiments of the disclosure, a depth graph convolutional network is used, and similarity information between samples is more fully utilized, thereby performing better comprehensive analysis on various influencing factors. In the embodiments of the disclosure, an association feature and a hard sample mining technology based on two searches are used, so that association information between samples can be more fully learned in high-dimensional feature space. An algorithm module provided in the embodiments of the disclosure can be flexibly applied to an existing re-identification algorithm, thereby stably improving performance.
- [00146] It may be understood that the foregoing method embodiments mentioned in the embodiments of the disclosure may be combined with each other to form a combined embodiment without departing from the principle and the logic. Details are not described in the embodiments of the disclosure due to space limitation.
- [00147] In addition, the embodiments of the disclosure further provide a target re-identification apparatus, an electronic device, a computer readable storage medium, and a computer program product. The foregoing may be utilized to implement any target re-identification method provided in the embodiments of the disclosure. For corresponding technical solutions and descriptions, reference is made to the corresponding descriptions of the method part. Details are not described again.
- [00148] FIG. 12 is a block diagram of a target re-identification apparatus according to an embodiment of the disclosure. As shown in FIG. 11, the target re-identification apparatus includes a picture acquiring module 1201, a candidate picture determining module 1202, and a target picture determining module 1203.
- [00149] The picture acquiring module 1201 is configured to acquire a target object picture and a to-be-processed picture set. The target object picture includes a target object.

- [00150] The candidate picture determining module 1202 is configured to determine a candidate picture set from the to-be-processed picture set according to a first feature value of the target object picture and a second feature value set corresponding to the to-be-processed picture set. A value of similarity between any two pictures in the candidate picture set is greater than or equal to a preset similarity value.
- [00151] The target picture determining module 1203 is configured to: identify the first feature value and the second feature value set based on a trained graph association identification network, and determine a target picture set from the candidate picture set. A first value of similarity between an object included in a picture in the target picture set and the target object is greater than or equal to a first value of similarity between an object included in a non-target picture and the target object. The candidate picture set includes the target picture set and the non-target picture.
- [00152] In some possible implementations, the graph association identification network includes a first graph structure establishment sub-network, a graph association update sub-network, and a classifier. The first graph structure establishment sub-network, the graph association update sub-network and the classifier are serially connected. The target picture determining module is configured to: input the first feature value and the second feature value set into the first graph structure establishment sub-network to obtain a first graph structure, where the first graph structure includes nodes and edges each of which is for connecting two nodes, a quantity of nodes is the same as a quantity of pictures in the candidate picture set, and the edge for connecting the two nodes is determined based on a similarity between the two nodes that are connected and a preset similarity; input the first graph structure into the graph association update sub-network to obtain an updated and optimized second graph structure; determine the first value of similarity corresponding to each candidate picture in the candidate picture set through the classifier according to the second graph structure; and determine the target picture set based on the first value of similarity corresponding to each candidate picture and a similarity threshold.
- [00153] In some possible implementations, the target picture determining module is configured to: add and fuse the first graph structure and the second graph structure to obtain a third graph structure, and determine the first value of similarity corresponding to each candidate picture in the candidate picture set through the classifier according to the third graph structure.

[00154] In some possible implementations, the graph association update sub-network includes an attention mechanism layer, a plurality of graph convolution layers, a plurality of activation layers, and a plurality of fully-connected layers. The attention mechanism layer, the plurality of graph convolution layers, the plurality of activation layers and the plurality of fully-connected layers are serially connected. The target picture determining module is configured to: input the first graph structure into the attention mechanism layer to obtain a weight vector of each node in the first graph structure, determine the weight vector of each node and the first graph structure as an input of a next layer of the attention mechanism layer, determine any currently processing layer in the plurality of graph convolution layers, the plurality of activation layers and the plurality of fully-connected layers as a current layer, use an output of a previous layer of the current layer as an input of the current layer, and obtain an output of the current layer after calculation processing, and in a case that any one of the current layers has a corresponding output, obtain the updated and optimized second graph structure according to an output of the last layer in the graph association update sub-network.

[00155] In some possible implementations, the candidate picture determining module is configured to: determine, based on a feature code extraction network, a first feature value of the target object included in the target object picture; determine, based on the feature code extraction network, a second feature value of an object included in each picture in the to-be-processed picture set; determine, according to the second feature value and the first feature value, a second value of similarity corresponding to each picture; and determine the candidate picture set from the to-be-processed picture set according to the second value of similarity.

[00156] In some possible implementations, the candidate picture determining module is configured to: rank second values of similarity corresponding to all to-be-processed pictures in a descending order of the values, and obtain the candidate picture set based on to-be-processed pictures corresponding to N top-ranked second values of similarity.

[00157] In some possible implementations, the candidate picture determining module is configured to: rank second values of similarity corresponding to all to-be-processed pictures in a descending order of the values; divide the to-be-processed picture set into a first candidate picture set and a non-first candidate picture set based on to-be-processed pictures corresponding to N1 top-ranked second values of similarity, where the first

candidate picture set includes the pictures corresponding to the N1 top-ranked second values of similarity; determine N2 pictures from the non-first candidate picture set according to a second feature value of a picture in the first candidate picture set and a second feature value of a picture in the non-first candidate picture set, to form a second candidate picture set; and determine the candidate picture set based on the first candidate picture set and the second candidate picture set.

[00158] In some possible implementations, the candidate picture determining module is configured to: determine any currently used picture in the first candidate picture set as a current picture; determine, according to a second feature value of the current picture and the second feature value of the picture in the non-first candidate picture set, a third value of similarity corresponding to each picture in the non-first candidate picture set; determine, according to the third value of similarity corresponding to each picture, a third candidate picture set corresponding to the current picture from the non-first candidate picture set; and in a case that each of the current pictures corresponds to a third candidate picture set, determine the N2 pictures according to the third candidate picture set corresponding to each of the current pictures, to form the second candidate picture set.

[00159] In some possible implementations, the apparatus further includes an analysis module, configured to: determine attribute information of the picture in the target picture set; and perform, according to the attribute information, trajectory and behavior analysis on the object included in the picture in the target picture set.

[00160] In some possible implementations, the attribute information includes a picture acquisition position and a picture acquisition time. The analysis module is configured to: rank pictures in the target picture set according to the picture acquisition time; and perform, based on the picture acquisition position and ranked pictures, motion trajectory determination and behavior speculation on the object included in the picture.

[00161] In some embodiments, functions or modules included in the apparatus provided in the embodiments of the disclosure can be configured to perform the method described in the foregoing method embodiments. For practical implementation thereof, reference is made to the descriptions of the foregoing method embodiments. For brevity, details are not described here again.

[00162] The embodiments of the disclosure further provide a computer readable storage medium, having at least one instruction or at least one program stored thereon. The at least one instruction or the at least one program is loaded and executed by a processor to implement the foregoing method. The computer readable storage medium may be a non-volatile computer readable storage medium.

[00163] The embodiments of the disclosure further provide an electronic device, including a processor, and a memory configured to store processor executable instructions. The processor is configured to perform the foregoing method. The electronic device may be provided as a terminal, a server, or other forms of devices.

[00164] The embodiments of the disclosure provide a computer program product including instructions. When running on a computer, the instructions cause the computer to perform the target re-identification method provided in the embodiments of the disclosure.

[00165] FIG. 13 is a block diagram of an electronic device according to an embodiment of the disclosure. For example, the electronic device 1300 may be a terminal such as a mobile phone, a computer, a digital broadcast terminal, a message transceiver device, a game console, a tablet device, a medical device, a fitness device, a personal digital assistant, or a vehicle-mounted device.

[00166] Referring to FIG. 13, the electronic device 1300 may include one or more of the following components: a processing component 1302, a memory 1304, a power component 1306, a multimedia component 1308, an audio component 1310, an Input/Output (I/O) interface 1312, a sensor component 1314, and a communications component 1316.

[00167] The processing component 1302 generally controls an overall operation of the electronic device 1300, such as operations associated with display, a telephone call, data communication, a camera operation, and a recording operation. The processing component 1302 may include one or more processors 1320 to execute instructions, to complete all or some of the steps of the foregoing method. In addition, the processing component 1302 may include one or more modules, to facilitate interaction between the processing component 1302 and other components. For example, the processing

component 1302 includes a multimedia module, to facilitate interaction between the multimedia component 1308 and the processing component 1302.

[00168] The memory 1304 is configured to store data of various types to support operations on the electronic device 1300. Examples of these data include instructions, contact data, phone book data, a message, a picture, or a video of any application program or method operated on the electronic device 1300. The memory 1304 is implemented by any type of volatile or non-volatile storage device or a combination thereof, such as a Static Random Access Memory (SRAM), an Electrically Erasable Programmable Read-Only Memory (EEPROM), an Erasable Programmable Read-Only Memory (EPROM), a Programmable Read-Only Memory (PROM), a Read-Only Memory (ROM), a magnetic memory, a flash memory, a magnetic disk, or an optical disc.

[00169] The power component 1306 supplies power to the components of the electronic device 1300. The power component 1306 may include a power management system, one or more power supplies, and other components associated with power generation, management, and distribution for the electronic device 1300.

[00170] The multimedia component 1308 includes a screen between the electronic device 1300 and a user to provide an output interface. In some embodiments, the screen may include a Liquid Crystal Display (LCD) and a Touch Panel (TP). If the screen includes the touch panel, the screen is implemented as a touchscreen, to receive an input signal from the user. The touch panel includes one or more touch sensors to sense a touch, a slide, and a gesture on the touch panel. The touch sensor may not only sense a boundary of a touch action or a slide action, but also detect the duration and pressure related to the touch operation or the slide operation. In some embodiments, the multimedia component 1308 includes a front-facing camera and/or a rear-facing camera. In a case that the electronic device 1300 is in an operation mode, for example, a photographing mode or a video mode, the front-facing camera and/or the rear-facing camera may receive external multimedia data. Each front-facing camera or rear-facing camera is a fixed optical lens system or has a focal length and an optical zoom capability.

[00171] The audio component 1310 is configured to output and/or input an audio signal. For example, the audio component 1310 includes one microphone (MIC). In a

case that the electronic device 1300 is in an operation mode, such as a call mode, a recording mode, or a voice identification mode, the microphone is configured to receive an external audio signal. The received audio signal may be stored in the memory 1304 or sent via the communications component 1316. In some embodiments, the audio component 1310 further includes a loudspeaker, configured to output an audio signal.

[00172] The I/O interface 1312 provides an interface between the processing component 1302 and a peripheral interface module, and the peripheral interface module is a keyboard, a click wheel, a button, or the like. These buttons may include but are not limited to a home button, a volume button, a startup button, and a lock button.

[00173] The sensor component 1314 includes one or more sensors, and is configured to provide status evaluation in various aspects for the electronic device 1300. For example, the sensor component 1314 may detect an on/off state of the electronic device 1300 and relative positioning of the components, for example, the components are a display and a keypad of the electronic device 1300. The sensor component 1314 may also detect a position change of the electronic device 1300 or a component of the electronic device 1300, existence or nonexistence of contact between the user and the electronic device 1300, an orientation or acceleration/deceleration of the electronic device 1300, and a temperature change of the electronic device 1300. The sensor component 1314 may include a proximity sensor, configured to detect existence of a nearby object in a case that there is no physical contact. The sensor component 1314 may further include an optical sensor, such as a Complementary Metal-Oxide-Semiconductor (CMOS) image sensor or a Charge Coupled Device (CCD) image sensor, that is used in an imaging application. In some embodiments, the sensor component 1314 may further include an acceleration sensor, a gyro sensor, a magnetic sensor, a pressure sensor, or a temperature sensor.

[00174] The communications component 1316 is configured to facilitate wired or wireless communication between the electronic device 1300 and other devices. The electronic device 1300 may access a wireless network based on a communications standard, such as Wireless Fidelity (Wi-Fi), a 2nd Generation (2G) mobile communications technology, a 3rd Generation (3G) mobile communications system, or a combination thereof. In an exemplary embodiment, the communications component 1316 receives a broadcast signal or broadcast-related information from an external broadcast

management system through a broadcast channel. In an exemplary embodiment, the communications component 1316 further includes a Near Field Communication (NFC) module, to promote short-range communication. For example, the NFC module may be implemented based on a Radio Frequency Identification (RFID) technology, an Infrared Data Association (IrDA) technology, an Ultra Wideband (UWB) technology, a Bluetooth (BT) technology, and other technologies.

[00175] In an exemplary embodiment, the electronic device 1300 may be implemented by one or more of an Application Specific Integrated Circuit (ASIC), a Digital Signal Processor (DSP), a Digital Signal Processing Device (DSPD), a Programmable Logic Device (PLD), a Field Programmable Gate Array (FPGA), a controller, a microcontroller, a microprocessor, or other electronic components, and is configured to perform the foregoing method.

[00176] In an exemplary embodiment, a non-volatile computer readable storage medium, for example, the memory 1304 including computer program instructions, is further provided. The computer program instructions are executed by the processor 1320 of the electronic device 1300 to complete the foregoing method.

[00177] FIG. 14 is a block diagram of another electronic device according to an embodiment of the disclosure. For example, the electronic device 1400 may be provided as a server. Referring to FIG. 14, the electronic device 1400 includes a processing component 1422. In some implementations, the processing component 1422 includes one or more processors, and a storage resource represented by a memory 1432 that is configured to store instructions that may be executed by the processing component 1422, for example, an application program. The application program stored in the memory 1432 may include one or more modules that each correspond to a set of instructions. In addition, the processing component 1422 is configured to execute the instructions to perform the foregoing method.

[00178] The electronic device 1400 may further include: a power supply component 1426, configured to perform power management of the electronic device 1400; a wired or wireless network interface 1450, configured to connect the electronic device 1400 to a network; and an I/O interface 1458. The electronic device 1400 may operate an operating system stored in the memory 1432, such as Windows Server™, Mac OS X™, and

Unix™, Linux™, FreeBSD™, or similar systems.

[00179] In an exemplary embodiment, a non-volatile computer readable storage medium, for example, the memory 1432 including computer program instructions, is further provided. The computer program instructions are executed by the processing component 1422 of the electronic device 1400 to complete the foregoing method.

[00180] The embodiments of the disclosure may be a system, a method, and/or a computer program product. The computer program product may include a computer readable storage medium, on which computer readable program instructions used by the processor to implement various aspects of the embodiments of the disclosure are stored.

[00181] The computer readable storage medium is a tangible device that can maintain and store instructions used by an instruction execution device. For example, the computer readable storage medium may be, but not limited to, an electrical storage device, a magnetic storage device, an optical storage device, an electromagnetic storage device, a semiconductor storage device, or any appropriate combination thereof. The computer readable storage medium may include a portable computer disk, a hard disk, a Random Access Memory (RAM), a read-only memory, an Erasable Programmable Read-Only Memory (EPROM or flash memory), a static random access memory, a portable Compact Disc Read-Only Memory (CD-ROM), a Digital Versatile Disk (DVD), a memory stick, a floppy disk, a mechanical coding device such as a punched card storing an instruction or a protrusion structure in a groove, and any appropriate combination thereof. The computer readable storage medium used here is not interpreted as an instantaneous signal such as a radio wave or other freely propagated electromagnetic wave, an electromagnetic wave propagated by a waveguide or other transmission media (for example, an optical pulse transmitted by an optical fiber cable), or an electrical signal transmitted by a wire.

[00182] The computer readable program instruction described here is downloaded from a computer readable storage medium to each computing/processing device, or downloaded to an external computer or an external storage device via a network, such as the Internet, a local area network, a wide area network, and/or a wireless network. The network may include a copper transmission cable, optical fiber transmission, wireless transmission, a router, a firewall, a switch, a gateway computer, and/or an edge server. A

network adapter card or a network interface in each computing/processing device receives the computer readable program instruction from the network, and forwards the computer readable program instruction, so that the computer readable program instruction is stored in a computer readable storage medium in each computing/processing device. Computer program instructions for executing the operations in the embodiments of the disclosure may be compilation instructions, instruction set architecture (ISA) instructions, machine instructions, machine-related instructions, microcode, firmware instructions, status setting data, or source code or target code written in any combination of one or more programming languages. The programming languages include an object-oriented programming language such as Smalltalk or C++, and a conventional procedural programming language such as the "C" language or a similar programming language. The program readable program instructions can be completely executed on a user computer, partially executed on a user computer, executed as an independent software package, executed partially on a user computer and partially on a remote computer, or completely executed on a remote computer or a server. In the case of a remote computer, the remote computer may be connected to a user computer via any type of network, including a Local Area Network (LAN) or a Wide Area Network (WAN), or may be connected to an external computer (for example, connected via the Internet with the aid of an Internet service provider). In some embodiments, an electronic circuit such as a programmable logic circuit, a field programmable gate array, or a programmable logic array is personalized by using status information of the computer readable program instructions, and the electronic circuit can execute the computer readable program instructions to implement various aspects of the embodiments of the disclosure.

[00183] Various aspects of the embodiments of the disclosure are described here with reference to the flowcharts and/or block diagrams of the methods, apparatuses (systems), and computer program products according to the embodiments of the disclosure. It should be understood that each block in the flowcharts and/or block diagrams and a combination of the blocks in the flowcharts and/or block diagrams can be implemented with the computer readable program instructions.

[00184] These computer readable program instructions may be provided for a general-purpose computer, a dedicated computer, or a processor of another programmable data

processing apparatus to generate a machine, so that when the instructions are executed by the computer or the processors of other programmable data processing apparatuses, an apparatus for implementing a specified function/action in one or more blocks in the flowcharts and/or block diagrams is generated. These computer readable program instructions may also be stored in a computer readable storage medium, and these instructions instruct a computer, a programmable data processing apparatus, and/or other devices to work in a specific manner. Therefore, the computer readable medium having the instructions stored thereon includes a manufacture, and the manufacture includes instructions for implementing specified functions/actions in one or more blocks in the flowcharts and/or block diagrams. The computer readable program instructions may also be loaded onto a computer, other programmable data processing apparatuses, or other devices, so that a series of operations and steps are executed on the computer, the other programmable apparatuses, or the other devices, thereby generating computer-implemented processes. Therefore, the instructions executed on the computer, the other programmable apparatuses, or the other devices implement the specified functions/actions in the one or more blocks in the flowcharts and/or block diagrams.

[00185] The flowcharts and block diagrams in the accompanying drawings show architectures, functions, and operations that may be implemented by the systems, methods, and computer program products in the implementations in the embodiments of the disclosure. In this regard, each block in the flowcharts or block diagrams may represent a module, a program segment, or a part of instruction, and the module, the program segment, or the part of instruction includes one or more executable instructions for implementing a specified logical function. In some alternative implementations, functions marked in the block may also occur in an order different from that marked in the accompanying drawings. For example, two consecutive blocks are actually executed substantially in parallel, or are sometimes executed in a reverse order, depending on the involved functions. It should also be noted that each block in the block diagrams and/or flowcharts and a combination of blocks in the block diagrams and/or flowcharts may be implemented by using a dedicated hardware-based system configured to execute specified functions or actions, or may be implemented by using a combination of dedicated hardware and computer instructions.

[00186] The implementations in the embodiments of the disclosure are described

above. The foregoing descriptions are exemplary but not exhaustive, and are not limited to the disclosed embodiments. For a person of ordinary skill in the art, many modifications and variations are all apparent without departing from the scope and spirit of the described embodiments. The terms used herein are intended to best explain the principles of the implementations, practical applications, or improvements to the technologies in the market, or to enable other persons of ordinary skill in the art to understand the embodiments disclosed herein.

[00187] Industrial applicability

[00188] In the embodiments of the disclosure, a target object picture and a to-be-processed picture set are acquired. The target object picture includes a target object. A candidate picture set is determined from the to-be-processed picture set according to a first feature value of the target object picture and a second feature value set corresponding to the to-be-processed picture set. A value of similarity between any two pictures in the candidate picture set is greater than or equal to a preset similarity value. The first feature value and the second feature value set are identified based on a trained graph association identification network, and a target picture set is determined from the candidate picture set. A first value of similarity between an object included in a picture in the target picture set and the target object is greater than or equal to a first value of similarity between an object included in a non-target picture and the target object, and the candidate picture set includes the target picture set and the non-target picture. In this way, a more accurate positive sample can be determined from the to-be-processed picture set, and interference of a negative sample can be reduced to obtain the target picture set, so that accuracy of a result of subsequently performing trajectory and behavior analysis on the object included in the picture in the target picture set based on attribute information of the picture in the target picture set is improved.

CLAIMS

1. A target re-identification method, comprising:

acquiring a target object picture and a to-be-processed picture set, the target object picture comprising a target object;

determining a candidate picture set from the to-be-processed picture set according to a first feature value of the target object picture and a second feature value set corresponding to the to-be-processed picture set, a value of similarity between any two pictures in the candidate picture set being greater than or equal to a preset similarity value; and

identifying the first feature value and the second feature value set based on a trained graph association identification network, and determining a target picture set from the candidate picture set, wherein a first value of similarity between an object comprised in a picture in the target picture set and the target object is greater than or equal to a first value of similarity between an object comprised in a non-target picture and the target object, and the candidate picture set comprises the target picture set and the non-target picture.

2. The method of claim 1, wherein the graph association identification network comprises a first graph structure establishment sub-network, a graph association update sub-network and a classifier that are serially connected; and

identifying the first feature value and the second feature value set based on the trained graph association identification network and determining the target picture set from the candidate picture set comprises:

inputting the first feature value and the second feature value set into the first graph structure establishment sub-network to obtain a first graph structure, wherein the first graph structure comprises nodes and edges each of which is for connecting two nodes, a quantity of nodes is the same as a quantity of pictures in the candidate picture set, and the edge for connecting the two nodes is determined based on a similarity between the two nodes that are

connected and a preset similarity;

inputting the first graph structure into the graph association update sub-network to obtain an updated and optimized second graph structure;

determining the first value of similarity corresponding to each candidate picture in the candidate picture set through the classifier according to the second graph structure; and

determining the target picture set based on the first value of similarity corresponding to each candidate picture and a similarity threshold.

3. The method of claim 2, wherein determining the first value of similarity corresponding to each candidate picture in the candidate picture set through the classifier according to the second graph structure comprises:

adding and fusing the first graph structure and the second graph structure to obtain a third graph structure; and

determining the first value of similarity corresponding to each candidate picture in the candidate picture set through the classifier according to the third graph structure.

4. The method of claim 2, wherein the graph association update sub-network comprises an attention mechanism layer, a plurality of graph convolution layers, a plurality of activation layers, and a plurality of fully-connected layers;

the attention mechanism layer, the plurality of graph convolution layers, the plurality of activation layers and the plurality of fully-connected layers are serially connected; and

inputting the first graph structure into the graph association update sub-network to obtain the updated and optimized second graph structure comprises:

inputting the first graph structure into the attention mechanism layer to obtain a weight vector of each node in the first graph structure;

determining the weight vector of each node and the first graph structure as an input of

a next layer of the attention mechanism layer;

determining any currently processing layer in the plurality of graph convolution layers, the plurality of activation layers and the plurality of fully-connected layers as a current layer;

using an output of a previous layer of the current layer as an input of the current layer, and obtaining an output of the current layer after calculation processing; and

in a case that any one of the current layers has a corresponding output, obtaining the updated and optimized second graph structure according to an output of the last layer in the graph association update sub-network.

5. The method of any one of claims 1 to 4, wherein determining the candidate picture set from the to-be-processed picture set according to the first feature value of the target object picture and the second feature value set corresponding to the to-be-processed picture set comprises:

determining, based on a feature code extraction network, a first feature value of the target object comprised in the target object picture;

determining, based on the feature code extraction network, a second feature value of an object comprised in each to-be-processed picture in the to-be-processed picture set;

determining, according to the second feature value and the first feature value, a second value of similarity corresponding to each to-be-processed picture; and

determining the candidate picture set from the to-be-processed picture set according to the second value of similarity.

6. The method of claim 5, wherein determining the candidate picture set from the to-be-processed picture set according to the second value of similarity comprises:

ranking second values of similarity corresponding to all to-be-processed pictures in a descending order of the values; and

obtaining the candidate picture set based on to-be-processed pictures corresponding to N top-ranked second values of similarity.

7. The method of claim 5, wherein determining the candidate picture set from the to-be-processed picture set according to the second value of similarity comprises:

ranking second values of similarity corresponding to all to-be-processed pictures in a descending order of the values;

dividing the to-be-processed picture set into a first candidate picture set and a non-first candidate picture set based on to-be-processed pictures corresponding to N1 top-ranked second values of similarity, wherein the first candidate picture set comprises the to-be-processed pictures corresponding to the N1 top-ranked second values of similarity;

determining N2 pictures from the non-first candidate picture set according to a second feature value of a picture in the first candidate picture set and a second feature value of a picture in the non-first candidate picture set, to form a second candidate picture set; and

determining the candidate picture set based on the first candidate picture set and the second candidate picture set.

8. The method of claim 7, wherein determining the N2 pictures from the non-first candidate picture set according to the second feature value of the picture in the first candidate picture set and the second feature value of the picture in the non-first candidate picture set, to form the second candidate picture set comprises:

determining any currently used picture in the first candidate picture set as a current picture;

determining, according to a second feature value of the current picture and the second feature value of the picture in the non-first candidate picture set, a third value of similarity corresponding to each picture in the non-first candidate picture set;

determining, according to the third value of similarity corresponding to each picture, a third candidate picture set corresponding to the current picture from the non-first candidate

picture set; and

in a case that each of the current pictures has a corresponding third candidate picture set, determining the N2 pictures according to the third candidate picture set corresponding to each of the current pictures, to form the second candidate picture set.

9. The method of claim 1, further comprising: after determining the target picture set from the candidate picture set,

determining attribute information of the picture in the target picture set; and

performing, according to the attribute information, trajectory and behavior analysis on the object comprised in the picture in the target picture set.

10. The method of claim 9, wherein the attribute information comprises a picture acquisition position and a picture acquisition time; and

performing, according to the attribute information, trajectory and behavior analysis on the object comprised in the picture in the target picture set comprises:

ranking pictures in the target picture set according to the picture acquisition time; and

performing, based on the picture acquisition position and the ranked pictures, motion trajectory determination and behavior speculation on the object comprised in the picture in the target picture set.

11. A target re-identification apparatus, comprising:

a picture acquiring module, configured to acquire a target object picture and a to-be-processed picture set, the target object picture comprising a target object;

a candidate picture determining module, configured to determine a candidate picture set from the to-be-processed picture set according to a first feature value of the target object picture and a second feature value set corresponding to the to-be-processed picture set, a value

of similarity between any two pictures in the candidate picture set being greater than or equal to a preset similarity value; and

a target picture determining module, configured to: identify the first feature value and the second feature value set based on a trained graph association identification network, and determine a target picture set from the candidate picture set, wherein a first value of similarity between an object comprised in a picture in the target picture set and the target object is greater than or equal to a first value of similarity between an object comprised in a non-target picture and the target object, and the candidate picture set comprises the target picture set and the non-target picture.

12. A computer readable storage medium, having at least one instruction or at least one program stored thereon, the at least one instruction or the at least one program being loaded and executed by a processor to implement the target re-identification method of any one of claims 1 to 10.

13. An electronic device, comprising at least one processor and a memory communicatively connected to the at least one processor, wherein the memory stores instructions executed by the at least one processor, and the at least one processor executes the instructions stored in the memory, to implement the target re-identification method of any one of claims 1 to 10.

14. A computer program product, comprising a computer readable code, when running on a device, the computer readable code causing a processor in the device to execute the target re-identification method of any one of claims 1 to 10.