

OBJECT DETECTION METHOD AND APPARATUS, AND ELECTRONIC DEVICE

TECHNICAL FIELD

5 **[0001]** The present disclosure relates to the field of machine learning technology, and
in particular, to an object detection method and apparatus, and an electronic device.

BACKGROUND

[0002] Target detection is an important part of intelligent video analysis. For example,
10 humans, animals and the like in video frames or scene images may be used as detection
targets. In the related art, a target detector such as a Faster RCNN (Region
Convolutional Neural Network) may be used to acquire target detection boxes from the
video frames or scene images.

[0003] However, in dense scenes, different targets may occlude each other. Take a
15 scene with relatively dense crowds of people as an example, human body parts such as
arms, hands and legs may be occluded between different people. In this case, use of the
conventional detector may cause false detection of the human body. For example, there
are only two people in a scene image originally, but three human body boxes are
detected from the scene image, this situation is usually called “false positive”.
20 Inaccurate target detection may lead to errors in subsequent processing based on the
detected targets.

SUMMARY

[0004] In view of this, the present disclosure provides at least an object detection

method and apparatus, and an electronic device, so as to improve the accuracy of target detection in dense scenes.

[0005] In a first aspect, there is provided an object detection method, including: detecting a face object and a body object from an image to be processed; determining a matching relationship between the detected face object and body object; and in response to determining that the body object matches the face object based on the matching relationship, determining the body object as a target object detected.

[0006] In a second aspect, there is provided an object detection apparatus, including: a detection processing module, configured to detect a face object and a body object from an image to be processed; a matching processing module, configured to determine a matching relationship between the detected face object and body object; and a target object determination module, configured to, in response to determining that the body object matches the face object based on the matching relationship, determining the body object as a target object detected.

[0007] In a third aspect, there is provided an electronic device including a memory and a processor, the memory is configured to store computer instructions executable on the processor, and the processor is configured to perform the method of any of the embodiments of the present disclosure when executing the computer instructions.

[0008] In a fourth aspect, there is provided a computer-readable storage medium on which a computer program is stored, the computer program, when executed by the processor, causes the processor to perform the method of any of the embodiments of the present disclosure.

[0009] The object detection method and apparatus, and electronic device according to the embodiments of the present disclosure assist in the detection of the body object by

using the detection of the matching relationship between the body object and the face object, and use the body object that has a matching face object as the detected target object. On one hand, since the detection accuracy of the face object is relatively high, the detection accuracy of the body object can also be improved by using the face object to assist in the detection of the body object; on the other hand, the face object belongs to the body object, thus the detection of the face object can assist in positioning the body object. This solution can reduce the occurrence of “false positive” or false detection, improving the detection accuracy of the body object.

10 **BRIEF DESCRIPTION OF DRAWINGS**

[0010] In order to illustrate the technical solutions in one or more embodiments of the present disclosure more clearly, the accompanying drawings used in the embodiments will be briefly introduced below. Obviously, the drawings in the following description merely illustrate some embodiments of one or more embodiments of the present disclosure. For those ordinary skilled in the art, other drawings may also be obtained from these drawings without any creative efforts.

[0011] FIG. 1 illustrates a flowchart of an object detection method according to at least one embodiment of the present disclosure;

[0012] FIG. 2 illustrates a schematic diagram of detection boxes for a body object and a face object according to at least one embodiment of the present disclosure;

[0013] FIG. 3 illustrates a schematic diagram of an architecture of a network used in an object detection method according to at least one embodiment of the present disclosure;

[0014] FIG. 4 illustrates a schematic structural diagram of an object detection apparatus

according to at least one embodiment of the present disclosure;

[0015] FIG. 5 illustrates a schematic structural diagram of an object detection apparatus according to at least one embodiment of the present disclosure.

5

DETAILED DESCRIPTION

[0016] In order for those skilled in the art to better understand the technical solutions in one or more embodiments of the present disclosure, the technical solutions in one or more embodiments of the present disclosure will be clearly and completely described below in conjunction with the accompanying drawings in one or more embodiments of the present disclosure. Apparently, the described embodiments are merely a part of the
10 embodiments of the present disclosure, rather than all of the embodiments. All other embodiments obtained by those ordinary skilled in the art based on one or more embodiments of the present disclosure without any creative efforts shall fall within the protection scope of the present disclosure.

15 [0017] When detecting targets in dense scenes, “false positive” may sometimes occur. For example, in a game place with relatively dense people, many people gather in the place to play games. Occlusions between people such as leg occlusion and arm occlusion may occur in images captured from the game place. Such occlusions between human bodies may lead to the occurrence of “false positive”. In order to improve the
20 accuracy of target detection in the dense scenes, embodiments of the present disclosure provide an object detection method, which can be applied to detect individual human bodies in a crowded scene as target objects for detection.

[0018] FIG. 1 illustrates a flowchart of an object detection method according to at least one embodiment of the present disclosure. As shown in FIG. 1, the method

includes steps 100, 102 and 104.

[0019] At step 100, a face object and a body object are detected from an image to be processed.

[0020] The image to be processed may be an image of a dense scene, and a
5 predetermined target object is expected to be detected from the image. In an example, the image to be processed may be an image of a multiplayer game scene, and the purpose of detection is to detect the number of people in the image to be processed, then each people in the image may be regarded as a target object to be detected.

[0021] In this step, each face object and body object included in the image to be
10 processed may be detected. In an example, when detecting the face object and the body object from the image to be processed, object detection may be performed on the image to be processed to obtain detection boxes for the face object and the body object from the image. For example, feature extraction may be performed on the image to be processed to obtain image features, and then the object detection may be performed
15 based on the image features to obtain the detection box for the face object and the detection box for the body object.

[0022] FIG. 2 schematically illustrates a plurality of detected detection boxes. As shown in FIG. 2, a detection box 21 includes a body object, and a detection box 22 includes another body object. A detection box 23 includes a face object, and a detection
20 box 24 includes another face object.

[0023] At step 102, a matching relationship between the detected face object and body object is determined.

[0024] In this step, the detected face object may include at least one face object and the detected body object may include at least one body object. Based on each detected

detection box obtained at step 100, each detected face object may be combined with each detected body object to obtain at least one face-and-body combination, and the matching relationship may be determined for each combination. For example, in the example of FIG. 2, the matching relationship between the detection box 21 and the detection box 23 may be detected, the matching relationship between the detection box 22 and the detection box 24 may be detected, the matching relationship between the detection box 21 and the detection box 24 may be detected, and the matching relationship between the detection box 22 and the detection box 23 may be detected.

[0025] The matching relationship represents whether the face object matches the body object. For example, a face object and a body object belonging to the same person may be determined to be a match. In an example, the body object included in the detection box 21 and the face object included in the detection box 23 belong to the same person in the image, and match each other. In contrast, the body object included in the detection box 21 and the face object included in the detection box 24 do not belong to the same person, and do not match each other.

[0026] In practical implementations, the above-mentioned matching relationship may be detected in various ways. In an exemplary embodiment, position information and/or visual information of the face object and the body object may be determined according to detection results for the face object and the body object; and the matching relationship between the face object and the body object may be determined according to the position information and/or the visual information.

[0027] The position information may indicate a spatial position of the face object and the body object in the image, or a spatial distribution relationship between the face object and the body object. The visual information may indicate visual feature

information of each object in the image, which is generally an image feature, for example, image features of the face object and the body object in the image obtained by extracting visual features from the image.

[0028] In an example, for each face object, the detection box for the body object that satisfies a preset position overlapping relationship with the detection box for the face object may be determined as a target detection box, according to position information of the detection boxes for the detected body object and face object, and the body object in the target detection box may be determined as the body object that matches the face object. In an example, the position overlapping relationship may be preset as follows:

the detection box for the face object overlaps with the detection box for the body object, and a ratio of an overlapping area to an area of the detection box for the face object reaches 90% or more. The detection box for each face object detected at step 100 may be combined in pairs with the detection box for each body object detected at step 100, and it is detected whether two detection boxes in a pair satisfy the above-mentioned preset overlapping relationship. If the two detection boxes satisfy the above-mentioned preset overlapping relationship, then it is determined that the face object and the body object respectively included in the two detection boxes match each other.

[0029] In another example, the matching relationship between the face object and the body object may also be determined according to the visual information of the face object and the body object. For example, the image features, that is, the visual information, of the detected face object and body object, may be obtained based on the face object and the body object, and the visual information of the face object and the body object may be combined to determine whether the face object matches the body object. In an example, a neural network may be trained to detect the matching

relationship according to the visual information, and the trained neural network may be used to draw a conclusion as to whether the face object matches the body object according to the input visual information of the two.

[0030] In yet another example, the matching relationship between the face object and the body object may also be detected according to a combination of the position information and the visual information of the face object and the body object. In an example, the visual information of the face object and the body object may be used in combination with the position information of the two to determine whether the face object matches the body object. For example, the spatial distribution relationship between the face object and the body object, or the position overlapping relationship between the detection box for the face object and the detection box for the body object may be combined with the visual information to comprehensively determine whether the face object matches the body object by using a trained neural network. The trained neural network may include a visual information matching branch and a position information matching branch. The visual information matching branch is configured to match the visual information of the face object and the body object, the position information matching branch is configured to match the position information of the face object and the body object, and the matching results of the two branches may be combined to draw a conclusion whether the face object and the body object match each other. Alternatively, the trained neural network may adopt an “end-to-end” model to process the visual information and the position information of the face object, and the visual information and the position information of the body object to obtain the matching relationship between the face object and the body object.

[0031] At step 104, in response to determining that the body object matches the face

object based on the matching relationship, the body object is determined as a target object detected.

[0032] In this step, based on the detection of the matching relationship at step 102, if a body object has a matching face object in the image, the body object may be determined as the detected target object. Otherwise, if a body object does not have a matching face object in the image, it may be determined that the body object is not the final detected target object.

[0033] In addition, based on the detection of the matching relationship between the face object and the body object, if it is determined that a body object does not have a matching face object based on the detected matching relationship, the detection box for the body object may be removed. For example, it is assumed that a detection box for a body object is detected from the image, the detection box is located in a preset edge area of the image which may be a predefined area within a certain range from an edge of the image, and there is no face object in the image matching the body object in the detection box, the body object in the detection box is not regarded as the detected target object. Optionally, this detection box located in the preset edge area of the image may be removed.

[0034] In other examples, if the body object has no matching face object due to the detection box for the body object being at the edge of the image, the body object in the detection box may also be determined as the target object. For example, in the case that it is determined based on the detection of the matching relationship that the body object in the detection box does not have a matching face object, it may be further determined whether the detection box is located in the preset edge area of the image. When it is determined that the detection box is located in the preset edge area, the body object may

be determined as the detected target object though there is no face object in the image matching the body object. In practical implementations, whether to regard the body object in this case as the final detected target object may be flexibly determined according to actual business requirements. For example, in a people-counting sense, the
5 body object in this case may be retained as the final detected target object.

[0035] In addition, before detecting the above-mentioned matching relationship, it may also be detected whether the face object is occluded by other face objects or any body object. In the case that the face object is not occluded by other face objects and any body object, an operation of determining the matching relationship between the face
10 object and the detected body object may be performed. Otherwise, if a detected face object is occluded by other face objects, or the detected face object is occluded by any body object in the image, the face object may be deleted from the detection results. For example, in a scene of a multiplayer table game, due to a large number of people participating in the game, there may be situations where different people occlude each
15 other, including body occlusion or even partial occlusion of the face. In this case, if a face is occluded by bodies or faces of other people, the detection accuracy of the face object may be reduced, and thus the detection accuracy of the body object may also be affected when the face object is used to assist in detection of the body object. However, as described above, in the case that it is determined that the face object is not occluded
20 by other bodies or faces, the detection accuracy of the face object itself is relatively high, and thus use of the face object to assist in the detection of the body object may assist in improving the detection accuracy of the body object.

[0036] Furthermore, if it is detected that the detection box for the face object satisfies the preset position overlapping relationship with the detection box for the body object,

and the face object is not occluded by other face objects and body objects, then it may be determined that the face object matches the body object. For example, with reference to FIG. 2, the body object in the detection box 21 satisfies the preset position overlapping relationship with the face object in the detection box 23, and the face object
5 in the detection box 23 is not occluded by other face objects and body objects, then it is determined that the body object in the detection box 21 and the face object in the detection box 23 match each other, and the body object in the detection box 21 is the detected target object.

[0037] The object detection method according to the embodiments of the present
10 disclosure assists in the detection of the body object by using the detection of the matching relationship between the body object and the face object, and uses the body object that has a matching face object as the detected target object. On one hand, since the detection accuracy of the face object is relatively high, the detection accuracy of the body object can also be improved by using the face object to assist in the detection of
15 the body object; on the other hand, the face object belongs to the body object, thus the detection of the face object can assist in positioning the body object. This solution can reduce the occurrence of “false positive” or false detection, improving the detection accuracy of the target object.

[0038] In addition, in a crowded scene, a plurality of human bodies may be crossed or
20 occluded each other. In a traditional human detection method, the crossed bodies of different people might be detected as the body object. The object detection method according to the present disclosure may match the detected body object with the face object, which can effectively filter out such a false-positive body object and provide a more accurate body object detection result.

[0039] FIG. 3 illustrates a schematic diagram of an architecture of a network used in an object detection method according to at least one embodiment of the present disclosure. As shown in FIG. 3, the network used for target detection may include a feature extraction network 31, an object detection network 32, and a matching detection
5 network 33.

[0040] The feature extraction network 31 is configured to perform feature extraction on the image to be processed (an input image in FIG. 3) to obtain a feature map of the image. In an example, the feature extraction network 31 may include a backbone network and a FPN (Feature Pyramid Network). The image to be processed may be
10 processed through the backbone network and the FPN in turn, to extract the feature map.

[0041] For example, the backbone network may use VGGNet, ResNet, etc. The FPN may convert the feature map obtained from the backbone network into a feature map with a multi-layer pyramid structure. The backbone network, as a backbone part of the target detection network, is configured to extract the image features. The FPN, as a neck
15 part of the target detection network, is configured to perform a feature enhancement processing, which may enhance shallow features extracted by the backbone network.

[0042] The object detection network 32 is configured to perform object detection based on the feature map of the image, to acquire at least one face box and at least one body box from the image to be processed. The face box is the detection box containing the
20 face object, and the body box is the detection box containing the body object.

[0043] As shown in FIG. 3, the object detection network 32 may include an RPN (Region Proposal Network) and an RCNN (Region Convolutional Neural Network). The RPN may predict an anchor box (anchor) for each object based on the feature map output from the FPN, and the RCNN may predict a plurality of bounding boxes (bbox)

based on the feature map output from the FPN and the anchor box, where the bounding box includes a body object or a face object. As mentioned above, the bounding box containing the body object is the body box, and the bounding box containing the face object is the face box.

5 **[0044]** The matching detection network 33 is configured to detect the matching relationship between the face object and the body object based on the feature map of the image, and the body object and the face object in the bounding boxes output from the RCNN.

10 **[0045]** The aforementioned object detection network 32 and matching detection network 33 may be equivalent to detectors in an object detection task, and configured to output the detection results. The detection results in the embodiments of the present disclosure may include a body object, a face object, and a matching pair. The matching pair is a pair of body object and face object that match each other.

15 **[0046]** It should be noted that the network structure of the aforementioned feature extraction network 31, object detection network 32, and matching detection network 33 is not limited in the embodiments of the present disclosure, and the structure shown in FIG. 3 is merely an example. For example, the FPN in FIG. 3 may not be used, but the feature map extracted by the backbone network may be directly used by the RPN/RCNN or the like to make a prediction for the position of the object. For another
20 example, FIG. 3 illustrates a framework of a two-stage target detection network, which is configured to perform object detection by using the feature extraction network and the object detection network. In practical implementations, a one-stage target detection network may also be used, and in this case, there is no need to provide an independent feature extraction network, and the one-stage target detection network may be used as

the object detection network in this embodiment to achieve feature extraction and object detection. When the one-stage target detection network is used, a body object and a face object, after obtained, may then be used to predict a matching pair.

[0047] For the network structure shown in FIG. 3, the network may be trained firstly,
5 and then the trained network may be used to detect a target object in the image to be processed. The training and application process of the network will be described below.

[0048] Sample images may be used for network training. For example, a sample image set may be acquired, and each sample image in the sample image set may be input to the feature extraction network 31 shown in FIG. 3 to obtain the extracted feature map
10 of the image. Then, the object detection network 32 detects and acquires at least one face box and at least one body box from the sample image according to the feature map of the image. Then, the matching detection network 33 acquires the pairwise matching relationship between the detected face box and body box. For example, any face box may be combined with any body box to form a face-and-body combination, and it is
15 detected whether the face object and the body object in the combination match each other. A detection result for the matching relationship may be referred to as a predicted value of the matching relationship, and a true value of the matching relationship may be referred to as a label value of the matching relationship. Finally, a network parameter of at least one of the feature extraction network, the object detection network, and the
20 matching detection network may be adjusted according to a difference between the label value and the predicted value of the matching relationship. The network training may be ended until a predetermined network training end condition is satisfied, and the trained network structure shown in FIG. 3 for target detection may be obtained.

[0049] After the network training is completed, for example, if the number of human

bodies needs to be detected from a certain image to be processed, where different people occlude each other, then the image to be processed may be processed according to the network architecture shown in FIG. 3. The trained feature extraction network 31 may firstly extract a feature map of the image, and then the trained object detection network 32 may acquire a face box and a body box from the image, and the trained matching detection network 33 may detect the matching face object and body object to obtain a matching pair. Then, the body object that has not successfully matched the face object may be removed, and is not regarded as the detected target object. If the body object does not have a matching face object, it may be considered that the body object is a “false positive” body object.

[0050] The object detection method according to the embodiments of the present disclosure assists in the detection of the body object by using the detection of the face object with a high accuracy, and an correlation relationship between the face object and the body object, such that the detection accuracy of the body object may be improved, and the false detection caused by occlusions between objects may be solved.

[0051] In some embodiments, the detection result for the target object in the image to be processed may be saved. For example, in a multiplayer game, the detection result may be saved in a cache for the multiplayer game, so as to analyse a game status, changes in players, etc. according to the cached information. Alternatively, the detection result for the target object in the image to be processed may be visually displayed, for example, the detection box of the detected target object may be drawn and shown in the image to be processed.

[0052] In order to implement the object detection method of any of the embodiments of the present disclosure, FIG. 4 illustrates a schematic structural diagram of an object

detection apparatus according to at least one embodiment of the present disclosure. As shown in FIG. 4, the apparatus includes a detection processing module 41, a matching processing module 42 and a target object determination module 43.

[0053] The detection processing module 41 is configured to detect a face object and a
5 body object from an image to be processed.

[0054] The matching processing module 42 is configured to determine a matching relationship between the detected face object and body object.

[0055] The target object determination module 43 is configured to, in response to determining that the body object matches the face object based on the matching
10 relationship, determine the body object as a target object detected.

[0056] In an example, the detection processing module 41 may be further configured to perform object detection on the image to be processed to obtain detection boxes for the face object and the body object from the image.

[0057] In an example, the target object determination module 43 may be further
15 configured to remove the detection box for the body object, in response to determining that there is no face object in the image matching the body object based on the matching relationship.

[0058] In an example, the target object determination module 43 may be further configured to determine the body object as the detected target object, in response to
20 determining that there is no face object in the image matching the body object based on the matching relationship, and the body object being located in a preset edge area of the image.

[0059] In an example, the matching processing module 42 may be further configured to determine position information and/or visual information of the face object and the

body object according to detection results for the face object and the body object; and determine the matching relationship between the face object and the body object according to the position information and/or the visual information.

[0060] In an example, the position information may include position information of the detection boxes. The matching processing module 42 may be further configured to: for each face object, determine the detection box for the body object that satisfies a preset position overlapping relationship with the detection box for the face object as a target detection box, according to the position information of the detection boxes, and determine the body object in the target detection box as the body object that matches the face object.

[0061] In an example, the matching processing module 42 may be further configured to determine the matching relationship between the detected face object and body object, in response to the detected face object being not occluded by the detected body object and other face objects.

[0062] In an example, the detected face object may include at least one face object and the detected body object may include at least one body object. The matching processing module 42 may be further configured to combine each of the detected face object with each of the detected body object to obtain at least one face-and-body combination, and determine the matching relationship for each of the combination.

[0063] In an example, as shown in FIG. 5, the apparatus may further include a network training module 44.

[0064] The detection processing module 41 may be further configured to perform the object detection on the image to be processed using an object detection network to obtain the detection boxes for the face object and the body object from the image.

[0065] The matching processing module 42 may be further configured to determine the matching relationship between the detected face object and body object using a matching detection network.

5 [0066] The network training module 44 may be configured to detect at least one face box and at least one body box from a sample image through the object detection network to be trained; acquire a predicted value of a pairwise matching relationship between the detected face box and body box through the matching detection network to be trained; and adjust a network parameter of at least one of the object detection network and the matching detection network, based on a difference between the
10 predicted value and a label value of the matching relationship.

[0067] The object detection apparatus according to the embodiments of the present disclosure assists in the detection of the body object by using the detection of the matching relationship between the body object and the face object, and uses the body object that has a matching face object as the detected target object, making the detection
15 accuracy of the body object higher.

[0068] The present disclosure also provides an electronic device including a memory and a processor, the memory is configured to store computer instructions executable on the processor, and the processor is configured to perform the method of any of the embodiments of the present disclosure when executing the computer instructions.

20 [0069] The present disclosure also provides a computer-readable storage medium on which a computer program is stored, the computer program, when executed by the processor, causes the processor to perform the method of any of the embodiments of the present disclosure.

[0070] Those skilled in the art should understand that one or more embodiments of the

present disclosure may be provided as a method, a system, or a computer program product. Therefore, one or more embodiments of the present disclosure may adopt the form of a complete hardware embodiment, a complete software embodiment, or an embodiment combining software and hardware. Moreover, one or more embodiments
5 of the present disclosure may adopt the form of a computer program product implemented on one or more computer-usable storage media (including but not limited to disk storage, CD-ROM, optical storage, etc.) containing computer-usable program codes.

[0071] As used herein, “and/or” means having at least one of the two, for example, “A
10 and/or B” includes three schemes: A, B, and “A and B”.

[0072] The various embodiments in the present disclosure are described in a progressive manner, and the same or similar parts between the various embodiments may be referred to each other. Each embodiment focuses on the differences from other embodiments. In particular, as for the data processing device embodiment, since it is
15 basically similar to the method embodiment, the description thereof is relatively simple, and reference may be made to the partial description of the method embodiment for the related parts.

[0073] The foregoing describes specific embodiments of the present disclosure. Other embodiments are within the scope of the appended claims. In some cases, the actions
20 or steps recited in the claims may be performed in a different order than in the embodiments and may still achieve desired results. In addition, the processes depicted in the drawings do not necessarily require the specific order or sequential order shown in order to achieve the desired results. In some embodiments, multitasking and parallel processing are also possible or may be advantageous.

[0074] The embodiments of the subject matter and functional operations described in this disclosure may be implemented in: digital electronic circuits, tangibly embodied computer software or firmware, computer hardware including the structures disclosed in this disclosure and structural equivalents thereof, or a combination of one or more of
5 them. Embodiments of the subject matter described in the present disclosure may be implemented as one or more computer programs, that is, one or more modules of the computer program instructions encoded on a tangible non-transitory program carrier to be executed by a data processing device or to control the operation of the data processing device. Alternatively or additionally, the program instructions may be
10 encoded on artificially generated propagated signals, such as machine-generated electrical, optical or electromagnetic signals, which are generated to encode information and transmit it to a suitable receiver device for execution by the data processing device. The computer storage medium may be a machine-readable storage device, a machine-readable storage substrate, a random or serial access memory device,
15 or a combination of one or more of them.

[0075] The processing and logic flows described in the present disclosure may be executed by one or more programmable computers executing one or more computer programs to perform corresponding functions by operating according to input data and generating output. The processing and logic flows may also be executed by a dedicated
20 logic circuit, such as FPGA (Field Programmable Gate Array) or ASIC (Application Specific Integrated Circuit), and the device may also be implemented as the dedicated logic circuit.

[0076] Computers suitable for executing computer programs include, for example, general-purpose and/or special-purpose microprocessors, or any other type of central

processing unit. Generally, the central processing unit will receive instructions and data from a read-only memory and/or a random access memory. The basic components of a computer include a central processing unit for implementing or executing instructions and one or more memory devices for storing instructions and data. Generally, the computer will also include one or more mass storage devices for storing data, such as magnetic disks, magneto-optical disks, or optical disks, or the computer will be operatively coupled to the mass storage device to receive data from or transmit data to it, or both. However, the computer does not have to have such a device. In addition, the computer may be embedded in another device, such as a mobile phone, a personal digital assistant (PDA), a mobile audio or video player, a game console, a global positioning system (GPS) receiver, or a portable storage device such as a universal serial bus (USB) and a flash drive, for example.

[0077] Computer-readable media suitable for storing computer program instructions and data include all forms of non-volatile memory, media and memory device, including, for example, semiconductor memory devices (such as EPROMs, EEPROMs, and flash memory devices), magnetic disks (such as internal Hard disks or removable disks), magneto-optical disk and CD ROM and DVD-ROM disk. The processor and the memory may be supplemented by or incorporated into a dedicated logic circuit.

[0078] Although the present disclosure contains many specific implementation details, these should not be construed as limiting the scope of any disclosure or the scope of protection, but are mainly used to describe the features of detailed embodiments of the specific disclosure. Certain features described in multiple embodiments within the present disclosure may also be implemented in combination in a single embodiment. On the other hand, various features described in a single embodiment may also be

implemented in multiple embodiments separately or in any suitable sub-combination.

In addition, although features may function in certain combinations as described above and even initially claimed as such, one or more features from the claimed combination may in some cases be removed from the combination, and the claimed combination
5 may be directed to a sub-combination or a variant of the sub-combination.

[0079] Similarly, although operations are depicted in a specific order in the drawings, this should not be understood as requiring these operations to be performed in the specific order shown or sequentially, or requiring all illustrated operations to be performed, to achieve the desired result. In some cases, multitasking and parallel
10 processing may be advantageous. In addition, the separation of various system modules and components in the above embodiments should not be understood as requiring such separation in all embodiments, and it should be understood that the described program components and systems may usually be integrated together in a single software product, or packaged into multiple software products.

[0080] The above descriptions are only some embodiments of one or more
15 embodiments of the present disclosure, and are not intended to limit one or more embodiments of the present disclosure. Any modification, equivalent replacement, improvement, etc. made within the spirit and principle of one or more embodiments of the present disclosure shall be included within the protection scope of one or more
20 embodiments of the present disclosure.

WHAT IS CLAIMED IS:

1. An object detection method, comprising:
 - detecting a face object and a body object from an image to be processed;
 - 5 determining a matching relationship between the detected face object and body object; and
 - in response to determining that the body object matches the face object based on the matching relationship, determining the body object as a target object detected.
- 10 2. The method of claim 1, wherein the detecting a face object and a body object from an image to be processed comprises:
 - performing object detection on the image to obtain detection boxes for the face object and the body object from the image.
- 15 3. The method of claim 2, further comprising:
 - removing the detection box for the body object, in response to determining that there is no face object in the image matching the body object based on the matching relationship.
- 20 4. The method of claim 1, further comprising:
 - determining the body object as the detected target object, in response to determining that there is no face object in the image matching the body object based on the matching relationship, and the body object being located in a preset edge area of the image.

5. The method of claim 1, wherein the determining a matching relationship between the detected face object and body object comprises:
- determining position information and/or visual information of the face object and the body object according to detection results for the face object and the body object; and
 - determining the matching relationship between the face object and the body object according to the position information and/or the visual information.
6. The method of claim 5, wherein the position information comprises position information of the detection boxes;
- the determining the matching relationship between the face object and the body object according to the position information and/or the visual information comprises:
 - for each face object, determining the detection box for the body object that satisfies a preset position overlapping relationship with the detection box for the face object as a target detection box, according to the position information of the detection boxes; and
 - determining the body object in the target detection box as the body object that matches the face object.
7. The method of claim 1, wherein the determining a matching relationship between the detected face object and body object comprises:
- determining the matching relationship between the detected face object and body object, in response to the detected face object being not occluded by the detected body object and other face objects.

8. The method of claim 1, wherein the detected face object comprises at least one face object and the detected body object comprises at least one body object,
the determining a matching relationship between the detected face object and body
5 object comprises:
combining each of the detected face object with each of the detected body object
to obtain at least one face-and-body combination, and determine the matching
relationship for each of the combination.
- 10 9. The method of any of claims 1-8, wherein
the detecting a face object and a body object from an image to be processed
comprises:
performing object detection on the image using an object detection network to
obtain detection boxes for the face object and the body object from the image;
15 the determining a matching relationship between the detected face object and body
object comprises:
determining the matching relationship between the detected face object and body
object using a matching detection network;
and wherein, the object detection network and the matching detection network are
20 trained by:
detecting at least one face box and at least one body box from a sample image
through the object detection network to be trained;
acquiring a predicted value of a pairwise matching relationship between the
detected face box and body box through the matching detection network to be trained;

and

adjusting a network parameter of at least one of the object detection network and the matching detection network, based on a difference between the predicted value and a label value of the matching relationship.

5

10. An object detection apparatus, comprising:

a detection processing module, configured to detect a face object and a body object from an image to be processed;

a matching processing module, configured to determine a matching relationship
10 between the detected face object and body object; and

a target object determination module, configured to, in response to determining that the body object matches the face object based on the matching relationship, determine the body object as a target object detected.

15 11. The apparatus of claim 10, wherein the detection processing module is further configured to perform object detection on the image to obtain detection boxes for the face object and the body object from the image.

12. The apparatus of claim 11, wherein the target object determination module is
20 further configured to remove the detection box for the body object, in response to determining that there is no face object in the image matching the body object based on the matching relationship.

13. The apparatus of claim 10, wherein the target object determination module is

further configured to determine the body object as the detected target object, in response to determining that there is no face object in the image matching the body object based on the matching relationship, and the body object being located in a preset edge area of the image.

5

14. The apparatus of claim 10, wherein the matching processing module is further configured to determine position information and/or visual information of the face object and the body object according to detection results for the face object and the body object; and determine the matching relationship between the face object and the
10 body object according to the position information and/or the visual information.

15. The apparatus of claim 14, wherein the position information comprises position information of the detection boxes;

the matching processing module is further configured to:

15 for each face object, determine the detection box for the body object that satisfies a preset position overlapping relationship with the detection box for the face object as a target detection box, according to the position information of the detection boxes; and

determine the body object in the target detection box as the body object that matches the face object.

20

16. The apparatus of claim 10, wherein the matching processing module is further configured to determine the matching relationship between the detected face object and body object, in response to the detected face object being not occluded by the detected body object and other face objects.

17. The apparatus of claim 10, wherein the detected face object comprises at least one face object and the detected body object comprises at least one body object;

the matching processing module is further configured to combine each of the
5 detected face object with each of the detected body object to obtain at least one face-
and-body combination, and determine the matching relationship for each of the
combination.

18. The apparatus of any of claims 10-17, wherein

10 the detection processing module is further configured to perform object detection
on the image using an object detection network to obtain detection boxes for the face
object and the body object from the image;

the matching processing module is further configured to determine the matching
relationship between the detected face object and body object using a matching
15 detection network;

and wherein, the apparatus further comprises a network training module configured
to: detect at least one face box and at least one body box from a sample image through
the object detection network to be trained; acquire a predicted value of a pairwise
matching relationship between the detected face box and body box through the
20 matching detection network to be trained; and adjust a network parameter of at least
one of the object detection network and the matching detection network, based on a
difference between the predicted value and a label value of the matching relationship.

19. An electronic device comprising a memory and a processor, the memory is

configured to store computer instructions executable on the processor, and the processor is configured to perform the method of any of claims 1-9 when executing the computer instructions.

- 5 20. A computer-readable storage medium on which a computer program is stored, wherein the computer program, when executed by the processor, causes the processor to perform the method of any of claims 1-9.