

METHODS, APPARATUSES, DEVICES AND STORAGE MEDIA FOR DETECTING CORRELATED OBJECTS INVOLVED IN IMAGE

TECHNICAL FIELD

5 **[0001]** The present disclosure relates to computer technology, in particular to methods, apparatuses, devices, and storage media for detecting correlated objects involved in an image.

BACKGROUND

10 **[0002]** Intelligent video analysis technology can assist people to understand states of objects in physical space and their relationship with each other. In an application scenario of intelligent video analysis, it is expected to recognize an identity of a person corresponding to a body part involved in the video.

15 **[0003]** The correlation between a body part and a personnel identity can be identified by some intermediate information. The intermediate information can indicate an object that has a relatively definite correlation with respect to both the body part and the personnel identity. For example, face information can be used as biological information to identify a person's identity. When it is expected to confirm an identity of a person to whom the hand detected in the image belongs, it can be
20 determined by the face that is correlated with the hand. Here, the correlated object indicates that two objects have an attribution relationship with the same third object, or have the same identity information attribute. As two body parts are correlated objects, it can be considered that the two body parts belong to the same person.

[0004] By correlating the body parts in the image, it can further assist analyzing the

behaviors and states of the person in a multi-person scenario, and the relationship between a plurality of persons.

SUMMARY

5 **[0005]** In view of above, the present disclosure at least discloses a method for detecting correlated objects involved in an image. The method includes: detecting a face object, a hand object and a preset body part object involved in a target image, wherein the preset body part object represents a preset connection part between a face and a hand; respectively predicting correlation between the detected face object and
10 the detected preset body part object, and correlation between the detected preset body part object and the detected hand object, to obtain a first correlation prediction result between the face object and the preset body part object, and a second correlation prediction result between the preset body part object and the hand object; and correlated objects involved in the target image are determined based on the first
15 correlation prediction result and the second correlation prediction result.

[0006] The present disclosure also discloses an apparatus for detecting correlated objects involved in an image. The apparatus includes: a detector configured to detect a face object, a hand object and a preset body part object involved in a target image, where the preset body part object represents a preset part of a body connection part
20 between a face and a hand; a first correlation predicting unit configured to respectively predict correlation between the detected face object and the detected preset body part object, and correlation between the detected preset body part object and the detected hand object, to obtain a first correlation prediction result between the face object and the preset body part object, and a second correlation prediction result
25 between the preset body part object and the hand object; and a determining unit

configured to determine correlated objects involved in the target image, based on the first correlation prediction result and the second correlation prediction result.

[0007] The present disclosure also discloses an electronic device, including: a processor; a memory for storing executable instructions of the processor; wherein the processor is configured to invoke the executable instructions stored in the memory to implement the method for detecting correlated objects involved in an image according to any one of the above examples.

[0008] A non-transitory computer-readable storage medium, wherein the storage medium stores a computer program, and the computer program is configured to perform the method for detecting correlated objects involved in an image according to any one of the above examples.

[0009] In the above solutions, by using a preset body part object which refers to a preset connection part between a face and a hand as an intermediary, respectively predicting correlation between the face object and the preset body part object, and correlation between the preset body part object and the hand object. Then, based on a prediction result between the face object and the preset body part object, and a prediction result between the preset body part object and the hand object, correlation between the detected face object and the hand object is determined. Compared with directly predicting correlation between a face and a hand, by introducing the preset body part object which is closely correlated with both the face and the hand as an intermediary, correlation between the face object and the hand object can be determined with higher accuracy. In addition, less interference information can be introduced when predicting correlation between a face and a hand, and accuracy of predicting correlation can be improved.

[0010] It should be understood that the general description and the following

detailed description are only exemplary and explanatory, and cannot limit the present disclosure.

BRIEF DESCRIPTION OF THE DRAWINGS

5 **[0011]** In order to more clearly describe the technical solutions in one or more examples of the present disclosure or related technologies, accompanying drawings that to be used in the description of the examples or related technologies will be briefly introduced in the following. Apparently, the accompanying drawings in the following description are only some of the examples described in one or more
10 embodiments of the present disclosure. For those of ordinary skill in the art, other drawings can be obtained based on these drawings without creative labour.

[0012] FIG. 1 is a flowchart illustrating a method for detecting correlated objects involved in an image according to the present disclosure;

[0013] FIG. 2 is a flowchart illustrating a target detection method for detecting a
15 face object, a hand object, and a preset body part object involved in a target image according to the present disclosure;

[0014] FIG. 3 is a flowchart illustrating a method for predicting correlation between a face object and a preset body part object according to the present disclosure;

[0015] FIG. 4 is a flowchart illustrating a method for predicting correlation between
20 a face object and a preset body part object according to the present disclosure;

[0016] FIG. 5 is a flowchart illustrating a method for detecting correlated objects according to the present disclosure;

[0017] FIG. 6 is a flowchart illustrating a method for adjusting a prediction result according to the present disclosure;

25 **[0018]** FIG. 7 is a flowchart illustrating a method for training a neural network in

the method for detecting correlated objects according to the present disclosure;

[0019] FIG. 8 is a block diagram illustrating an apparatus for detecting correlated objects involved in an image according to the present disclosure;

[0020] FIG. 9 is a block diagram illustrating hardware structure of an electronic device according to the present disclosure.

DETAILED DESCRIPTION OF THE EXAMPLES

[0021] Examples will be described in detail herein, with the illustrations thereof represented in the drawings. When the following descriptions involve the drawings, the same number in different drawings refers to same element or similar elements unless otherwise indicated. The embodiments described in the following examples do not represent all embodiments consistent with the present disclosure. Rather, they are merely examples of devices and methods consistent with some aspects of the present disclosure and as detailed in the appended claims.

[0022] The terms used in the present disclosure are for the purpose of describing particular examples only, and are not intended to limit the present disclosure. Terms determined by “a”, “the” and “said” in their singular forms in the present disclosure and the appended claims are also intended to include plural forms, unless clearly indicated otherwise in the context. It should also be understood that the term “and/or” used herein includes any or all possible combinations of one or more of the correlated listed items. It should also be understood that, depending on the context, the word “if” as used herein may be interpreted as “when” or “upon” or “in response to determining”.

[0023] The present disclosure discloses a method for detecting correlated objects involved in an image. In the method, by using a preset body part object which refers

to a preset connection part between a face and a hand as an intermediary, correlation between a face object and the preset body part object, and correlation between the preset body part object and a hand object are respectively predicted. Then, based on a prediction result between the face object and the preset body part object, and a
5 prediction result between the preset body part object and the hand object, correlation between the detected face object and the hand object is determined. Compared with directly predicting correlation between a face and a hand, by introducing a preset body part object which is closely correlated with both the face and the hand as an intermediary, correlation between the face object and the hand object can be
10 determined with improved accuracy. In addition, less interference information can be introduced when predicting correlation between a face and a hand, and accuracy of predicting correlation can be improved.

[0024] Referring to FIG. 1, FIG. 1 is a flowchart illustrating a method for detecting correlated objects involved in an image according to the present disclosure. As shown
15 in FIG. 1, the method can include the following steps.

[0025] At S102, a face object, a hand object and a preset body part object are detected from a target image, where the preset body part object represents a preset connection part between a face and a hand.

[0026] At S104, correlation between the detected face object and the detected preset
20 body part object, and correlation between the detected preset body part object and the detected hand object are predicted respectively, to obtain a first correlation prediction result between the face object and the preset body part object, and a second correlation prediction result between the preset body part object and the hand object.

[0027] At S106, based on the first correlation prediction result and the second
25 correlation prediction result, correlated objects involved in the target image are

determined.

[0028] The detection method can be applied to an electronic device. The electronic device can execute the method by installing a software system corresponding to the method. In examples of the present disclosure, the type of the electronic device can be
5 a notebook computer, a computer, a server, a mobile phone, a PAD terminal, etc., which is not particularly limited in the present disclosure.

[0029] It is understandable that the method can be executed by either of a client device or a server device, or can be executed by both of the client device and the server device in cooperation.

10 **[0030]** For example, the method can be integrated in the client device. After receiving a correlated object detection request, the device can execute the method through the computing power provided by hardware of the device.

[0031] For another example, the method can be integrated into a server device. After receiving a correlated object detection request, the device can execute the method
15 through the computing power provided by hardware of the device.

[0032] For another example, the method can be divided into two steps: obtaining a target image and performing correlated object detection on the target image. Here, the step of obtaining a target image can be performed by the client device, and the step of performing correlated object detection on the target image can be performed by the
20 server device. The client device can initiate a correlated object detection request to the server device after obtaining the target image. After receiving the correlated object detection request, the server device can perform correlated object detection on the target image in response to the request.

[0033] The following description will be given with reference to an example in
25 which the execution entity is an electronic device (hereinafter referred to as the

device).

[0034] The target image refers to an image that needs image processing to extract useful information. The target image can involve several to-be-detected objects. For example, in a tabletop game scenario, the target image can involve some persons
5 around the tabletop, as well as the persons' face objects, hand objects, and preset body objects(such as elbows).

[0035] In some examples, the device can interact with a user to obtain the target image input by the user. For example, the device can provide a window for a user to input the target image to be processed through its equipped interface. Thus, the user
10 can complete the input of the target image based on this window.

[0036] In some examples, the device can also be connected to an image capture device deployed on-site in a to-be-captured scenario to obtain an image captured by the image capture device and take the image as a target image.

[0037] After the target image is obtained, the above S102 can be performed to detect
15 the face object, the hand object, and the preset body part object involved in the target image, where the preset body part object represents a preset connection part between a face and a hand.

[0038] The preset body part object can represent a preset connection part between a face and a hand. Compared with directly predicting correlation between a face and a
20 hand, by introducing a preset body part object that is more closely correlated with the face and the hand as an intermediary, correlation between the face object and the hand object can be determined with improved accuracy.

[0039] In some examples, the preset body part may refer to a preset body part object on the arm. In some examples, in order to improve the accuracy of predicting
25 correlation, the preset body part objects can include at least one of a shoulder object,

an elbow object, and a wrist object that are easier to be detected from the target image.

[0040] In this step, the target image can be input into a target object detecting network for target detection, to obtain the face object, the hand object and the preset
5 body part object involved in the target image.

[0041] It should be understood that the result of target detection on the target image can include position information of the face object, the hand object and the preset body part object. The position information can include a bounding box and position information of the bounding box. When the bounding box is a rectangular box, the
10 position information of the bounding box can include coordinate of at least one of the vertices, as well as length and width information of the bounding box.

[0042] The target object detecting network is used to perform target detection tasks. For example, the target object detecting network can be a neural network built based on a RCNN (Region Convolutional Neural Network), a FAST-RCNN (Fast a Region
15 Convolutional Neural Network) or a FASTER-RCNN (Faster Region Convolutional Neural Network).

[0043] In practice, before detecting target with the target object detecting network, the network can be trained based on some training samples with position label information of the face object, the hand object and the preset body part object, until
20 the network converges.

[0044] Referring to FIG. 2, FIG. 2 is a flowchart illustrating a target detection method for detecting a face object, a hand object, and a preset body part object involved in a target image according to the present disclosure. It should be noted that FIG. 2 only schematically illustrates the process of the target detection method, and
25 does not limit the present disclosure.

[0045] As shown in FIG. 2, the target object detecting network can be a FASTER-RCNN, and can include at least Backbone (a backbone network), RPN (Region Proposal Network, a candidate box generating network), and RCNN (Region-based Convolutional Neural Network).

5 **[0046]** Here, the backbone network can perform some convolution operations on the target image to obtain a target feature map of the target image. Then, the target feature map can be processed by the RPN network to obtain anchors (anchor boxes) respectively corresponding to the target objects involved in the target image. After that, the anchor boxes output by the RPN network and the target feature map output
10 by the backbone network can be processed by the RCNN network for bbox (bounding boxes) regression and classification to obtain bounding boxes respectively corresponding to the face object, the hand object, and the preset body part object involved in the target image.

[0047] It should be noted that in examples of the present disclosure, a same target
15 object detecting network can be used to detect three different types of body part objects, and as well as categories and positions of the face object, the hand object, and the preset body part object involved in sample images are respectively marked in the training. When performing the target detection task, the target object detecting network can output a detection result for different types of body part objects and body
20 objects.

[0048] After determining the bounding boxes respectively corresponding to the face object, the hand object, and the preset body part object, S104 can be performed to predict correlation between the detected face object and the detected preset body part object, and correlation between the detected preset body part object and the detected
25 hand object, to obtain a first correlation prediction result between the face object and

the preset body part object, and a second correlation prediction result between the preset body part object and the hand object.

[0049] The correlation predicting mentioned above specifically refers to detecting correlation between two objects. In practice, a probability or a confidence that two objects belong to the same body object can be calculated, to detect the correlation between the two objects. The two objects can include a face object and a preset body object, or a preset body object and a hand object.

[0050] In some examples, a probability that the detected face object and the detected preset body object belong to the same body object, and a probability that the detected preset body object and the detected hand object belong to the same body object can be calculated based on features such as distance, relative position relationship, colour correlation degree between two objects in the image, and prior knowledge such as the distance, relative position relationship, colour of two correlated objects in an actual scene, and be taken as the first correlation prediction result and the second correlation prediction result respectively.

[0051] In some examples, the correlation predicting can be performed through a correlation predicting model which is constructed based on a neural network, to obtain a confidence that two objects belong to the same body object.

[0052] In some examples, the degree of confidence can be quantified by a prediction score. The higher the prediction score is, the higher the probability that the two parts belong to the same body will be.

[0053] It should be understood that, in some cases, the target image can involve a plurality of face objects, a plurality of hand objects, and a plurality of preset body part objects. In the method of the present disclosure, the face objects can be randomly combined with the hand objects to form a plurality of first combinations, which are

respectively used to predict the correlation between the face object and the hand object. In this case, when determining the first correlation prediction result, S1042 can be performed first, in which each of the detected face objects are combined with each of the preset body part objects to obtain a plurality of second combinations.

5 **[0054]** Before performing S1042, a unique identifier can be created for each detected face object, each hand object, and each preset body part object.

[0055] In some examples, a unique identifier can be created for each object based on the category of each object and a list of integer numbers. For example, the created identifier can be a face object F1, a face object F2, a hand object H1, a preset body
10 part object E1, etc., where "F", "H" and "E" are used to identify the categories of the face object, the categories of the hand objects and the categories of the preset body part objects.

[0056] After the identifiers are created, each of the face objects can be used as the target face object in turn, and combined with each of the preset body part objects
15 according to the identifiers, to obtain a plurality of second combinations. It should be understood that all the object combination methods involved in the present application can refer to the combination method of the second combination described above, which will not be described in detail later.

[0057] After a plurality of second combinations are obtained, S1044 can be
20 performed, in which obtaining a first correlation prediction result between the face object and the preset body part object in each of the second combinations by, for each of the second combinations, predicting correlation on the face object and the preset body part object in the second combination based on the visual features of the face object and the preset body part object in the second combination.

25 **[0058]** In some examples, correlation predicting can be performed based on a

correlation predicting model. The correlation predicting model can be a regression model or a classification model constructed based on a visual feature extracting unit. The predicting model can include a fully connected layer, and finally output a correlation prediction score.

- 5 **[0059]** Here, the fully connected layer can be a calculating unit constructed based on algorithms such as linear regression and least square regression. The calculating unit can perform feature mapping on the visual features to obtain the corresponding correlation prediction score value.

- 10 **[0060]** In practice, the calculating unit can be trained based on several training samples with label information on correlations between face objects and preset body part objects.

- 15 **[0061]** When constructing training samples, several original images can be first obtained, face objects and preset body part objects involved in the original images are randomly combined with an annotation tool, to obtain a plurality of combinations.
- 20 Afterwards correlation labeling between the face object and the preset body part object in each combination is performed. If the face object in a combination is correlated to the preset body part object (belonging to the same person), it can be labelled with 1, otherwise it can be labelled with 0. Alternatively, when the original image is labelled, information of a person object (such as a person ID) to which a face object and a preset body part object belong can be labelled respectively. Thus, it can be determined whether the face object is correlated to the preset body part object in the combination based on whether the information of the person to which the face object belongs and the information of the person to which the preset body part object belongs is consistent.

- 25 **[0062]** Referring to FIG. 3, FIG. 3 is a flowchart illustrating a method for predicting

correlation between a face object and a preset body part object according to the present disclosure.

[0063] As an example, the correlation predicting model shown in FIG. 3 can include a visual feature extracting unit and a fully connected layer.

5 **[0064]** The visual feature extracting unit can obtain a feature region based on a bounding box and a target feature map corresponding to the target image.

[0065] For example, the visual feature extracting unit can be a RoI Align (Region of interest Align) unit or a RoI pooling (Region of interest pooling) unit.

[0066] The fully connected layer can be a unit constructed based on algorithms such
10 as linear regression and least square regression. This unit can perform feature mapping (matrix operation) on the feature region (pixel matrix) to obtain a corresponding correlation prediction score value.

[0067] When predicting with the correlation predicting model, the bounding boxes of the face object and the preset body part object in each second combination, and the
15 target feature map corresponding to the target image can be input to the visual feature extracting unit, to obtain the visual features corresponding to the face object and the preset body part object.

[0068] Then, the visual features are input into the fully connected layer to calculate the first correlation prediction result.

20 **[0069]** In some examples, in order to improve the accuracy of predicting correlation, when calculating the first correlation prediction score, for each of the second combinations, based on the visual features and the position features of the face object and the preset body part object in the second combination, correlation predicting model can be used to predict correlation of the face object with respect to the preset
25 body part object in the second combination, to obtain a first correlation prediction

result between the face object and the preset body part object in each of the second combinations. Here, the visual features include features such as colour and/or texture, and the position features include features such as coordinate positions, relative positional relationships with other objects, and the like.

5 **[0070]** Referring to FIG. 4, FIG. 4 is a flowchart illustrating a method for predicting correlation between a face object and a preset body part object according to the present disclosure.

10 **[0071]** As shown in FIG. 4, after the visual features corresponding to the face object and the preset body part object respectively are extracted by the visual feature extracting unit, feature splicing (for example, single-core convolution operation, superposition, etc.) is performed on the position features (for example, coordinates) of the face object and the position features of the preset body part object, to obtain a spliced feature.

15 **[0072]** After the spliced feature is obtained, the spliced feature can be input into the fully connected layer for performing feature mapping (matrix operation) to obtain the first correlation prediction result.

20 **[0073]** In correlation predicting, in addition to the visual features of the face object and the preset body part object, the position features respectively corresponding to the bounding boxes of the face object and the preset body part object are also used. Thus, information such as the potential positional relationship between body parts can be extracted, and the extracted information which is useful for predicting the correlation between body part objects can be introduced, thereby improving the accuracy of the correlation prediction result.

25 **[0074]** Then, S1046 can be performed to combine each of the detected preset body part objects with each of the hand objects to form a plurality of third combinations.

[0075] At S1048, a second correlation prediction result between the preset body part object and the hand object in each of the third combinations is obtained by, for each of the third combinations, predicting correlation of the hand object with respect to the preset body part object in the third combinations based on the visual features and position features of the preset body part object and the hand object in the third combination.

[0076] It should be understood that the description of step S1046-S1048 can refer to the description of step S1042-S1044, which is not described in detail here.

[0077] It should be noted that this application does not specifically limit the sequence of determining the first correlation prediction result and the second correlation prediction result. For example, S1042-S1044 can be performed first, or S1046-S1048 can be performed first, or predicting the first correlation prediction result and the second correlation prediction result can be performed simultaneously.

[0078] After obtaining a plurality of first correlation prediction results and a plurality of second correlation prediction results, the process can be proceeded to S106, in which correlated objects involved in the target image can be determined based on the first correlation prediction results and the second correlation prediction results.

[0079] In some examples, based on the first correlation prediction results and the second correlation prediction results, a face object and a hand object of which correlations with respect to the same preset body part object satisfy a preset condition can be determined as the correlated objects involved in the target image.

[0080] The preset condition can be set based on actual business requirements. In some examples, the preset condition may specify that the confidence of the correlation with respect to the same preset body part object reaches a preset threshold

(empirical threshold).

[0081] In some examples, the first correlation prediction results that reach the first preset threshold (empirical threshold) can be selected from the plurality of first correlation prediction results, and the face object and the preset body part object corresponding to the first correlation prediction result are determined as a pair of preliminarily correlated face object and preset body part object.

[0082] Afterwards, a number of face objects preliminarily correlated with the same preset body part object can be determined.

[0083] If the number of face objects preliminarily correlated with the same preset body part object is 1, it is determined that the face object is correlated with the hand object.

[0084] If the number of face objects preliminarily correlated with the same preset body part object is greater than 1, from the plurality of face objects preliminarily correlated with the preset body part object, a face object which has the largest correlation with the preset body part object is determined, and the face object is determined as a face object correlated with the preset body part object.

[0085] Then, based on a similar method, the hand object correlated with the preset body part object can be determined.

[0086] After the face object and the hand object correlated with the same preset body part object are determined, the face object and the hand object can be determined as a pair of correlated objects belonging to the same body object.

[0087] In the above solutions, since the preset body part object is the preset connecting part between a face and a hand, the preset body part object is closely related to the hand object and the face object. In predicting the correlation between the face object and the hand object, the actually correlated face object and the hand

object can be correlated through intermediate information, which can improve the accuracy of the detection result of the correlated object.

[0088] In some examples, in order to improve the accuracy of predicting correlation, when performing S106, correlation between the detected face object and the detected hand object can be predicted first to obtain a third correlation prediction result. Then, auxiliary information which is useful for predicting the correlation between the face object and the hand object can be extracted from the first correlation prediction result and the second correlation prediction result. Afterwards, based on the auxiliary information, the third correlation prediction result can be adjusted, and based on the adjusted third correlation prediction result, the correlated objects involved in the target image can be determined, thereby the accuracy of predicting correlation can be improved.

[0089] The above steps are described below with reference to the drawings.

[0090] Referring to FIG. 5, FIG. 5 is a flowchart illustrating a method for detecting correlated objects according to the present disclosure.

[0091] The process shown in FIG. 5 includes at least a target object detecting network, a first preset network and a second preset network. The target object detecting network is configured to detect face objects, hand objects, and preset body part objects involved in the target image. The first preset network is configured to predict correlation between a face object and a hand object to obtain a third correlation prediction result. The second preset network is configured to predict correlation between a face object and a preset body part object, and correlation between a preset body part object and a hand object, to obtain a first correlation prediction result and a second correlation prediction result respectively.

[0092] As shown in FIG. 5, in predicting the correlation, a target image can be input

into a target object detecting network, and target detecting is performed to obtain bounding boxes of face objects, hand objects, and preset body part objects involved in the target image.

[0093] Then, the first preset network can be used to predict correlation between the detected face objects and hand objects to obtain third correlation prediction results. It should be understood that the description of the step of predicting the third correlation prediction result can refer to the description of the steps of S1042-S1044, which will not be elaborated herein.

[0094] Then, the second preset network can be used to predict correlation between the detected face objects and the preset body part object, as well as the preset body part object and the hand objects respectively, to obtain first correlation prediction results regarding correlations between the face objects and the preset body part objects, and second correlation prediction results regarding correlations between the preset body part object and the hand objects.

[0095] After that, the third correlation prediction result can be adjusted based on the first correlation prediction result and the second correlation prediction result. In some optional implementations, the first correlation prediction result and the second correlation prediction result can be used to verify the third correlation prediction result. The credibility of the third correlation prediction result can be increased if the verification is passed, otherwise the credibility of the third correlation prediction result is reduced or the third correlation prediction result is adjusted to "not correlated".

[0096] As an example, if it is determined based on a first correlation prediction result that a face object F1 is correlated with a preset body part object E1, it is determined based on a second correlation prediction result that the preset body part

object E1 is not correlated with a hand object H1, and it is determined based on a third correlation prediction result that the face object F1 is correlated with the hand object H1, then the third correlation prediction result between the face object F1 and the hand object H1 can be determined to be: not correlated.

5 **[0097]** In some examples, the correlation prediction result can include a correlation prediction score.

[0098] Referring to FIG. 6, FIG. 6 is a flowchart illustrating a method for adjusting a prediction result according to the present disclosure.

[0099] As shown in FIG. 6, each of the detected preset body part objects can be used
10 in turn as a target body part object, and the following operations can be performed.

[0100] At S602, it is to determine a target face object of which a first correlation prediction score in the first correlation prediction result with respect to the target body part object is highest.

[0101] In some examples, the first correlation prediction scores corresponding to the
15 target body part objects can be sorted in descending order. The face object corresponding to the first correlation prediction score ranked first can be determined as the target face object.

[0102] Thus, the face object having the highest correlation with the target body part object can be obtained.

20 **[0103]** In some examples, candidate face objects can be determined in a way that the first correlation prediction score of each candidate face object with respect to the target body part object is greater than a preset threshold. Then, from the candidate face objects, the one having the highest first correlation prediction score is selected as the target face object.

25 **[0104]** The preset threshold is an empirical threshold. If the correlation prediction

score of two body objects reaches the preset threshold, it indicates that the two body objects are more likely to belong to the same person.

[0105] In the above examples, if the first correlation prediction score between the face object and the preset body part object is lower than the preset threshold, the candidate face object cannot be determined. In this case, it can indicate that the face object and the preset body part do not belong to the same person (it can be caused by the preset body part belonging to the same person as the face object being blocked), so there is no need to adjust the third correlation prediction score corresponding to the face object based on the first correlation prediction score. Thus, on the one hand, the amount of model calculation can be reduced and the efficiency of detecting correlated object can be improved; on the other hand, useless correction can be avoided and the accuracy of detecting correlated object can be improved.

[0106] Then, S604 can be performed, it is to determine a target hand object of which a second correlation prediction score in the second correlation prediction result with respect to the target body part object is highest.

[0107] In some examples, the second correlation prediction scores corresponding to the target body part objects can be sorted in descending order. The hand object corresponding to the second correlation prediction score ranked first can be determined as the target hand object.

[0108] Thus, the hand object having the highest correlation with the target body part object can be obtained.

[0109] In some examples, candidate hand objects can be determined in a way that the second correlation prediction score of the candidate hand object with respect to the target body part object is greater than a preset threshold. From the candidate hand objects, the one having the highest second correlation prediction score is determined

as the target hand object.

[0110] The preset threshold is an empirical threshold. If the correlation prediction score of two body objects reaches the preset threshold, it indicates that the two body objects are more likely to belong to the same person.

5 **[0111]** In the above examples, if the second correlation prediction score between the hand object and the preset body part object is lower than the preset threshold, then the candidate hand object cannot be determined. In this case, it can indicate that the hand object and the preset body part do not belong to the same person (it can be caused by the preset body part belonging to the same person as the hand object being blocked),
10 so there is no need to adjust the third correlation prediction score corresponding to the hand object based on the second correlation prediction score. Thus, on the one hand, the amount of model calculation can be reduced and the efficiency of detecting correlated object can be improved; on the other hand, useless correction can be avoided and the accuracy of detecting correlated object can be improved.

15 **[0112]** Finally, S606 can be performed, in which based on the first correlation prediction score of the target face object with respect to the target body part object, and the second correlation prediction score of the target body part object with respect to the target hand object, a third correlation prediction score in the third correlation prediction result between the target face object and the target hand object is adjusted.

20 **[0113]** In some examples, an average value of the first correlation prediction score of the target face object with respect to the target body part object, and the second correlation prediction score of the target hand object with respect to the target body part object can be determined first.

[0114] The adjusted third correlation prediction score is obtained by adding the
25 average value to the third correlation prediction score between the target face object

and the target hand object.

[0115] It should be noted here that there can be many ways to adjust the third correlation prediction score. For example, the sum of the first correlation prediction score, the second correlation prediction score, and the third correlation prediction score is directly determined as the adjusted third correlation prediction score. For another example, the sum of the third correlation prediction score and the first correlation prediction score or the second correlation prediction score is determined as the adjusted third correlation prediction score. In the present disclosure, the adjustment methods of the third correlation prediction score are not exhaustively listed.

[0116] After completing the adjustment of the third correlation prediction result, the correlation between the detected face object and the detected hand object can be determined based on the adjusted third correlation prediction result. Whether the face object in the image is correlated with the hand object can be determined by the third correlation prediction result. For example, whether the face object and the hand object are correlated can be determined by whether the correlation prediction score representing the third correlation prediction result exceeds a threshold.

[0117] At this step, it is also possible to select each from a plurality of the third correlation prediction scores in an order of the third correlation prediction scores from high to low, and for a current combination of the face object and the hand object corresponding to the selected third correlation prediction score perform the following first and second steps.

[0118] In the first step, based on the determined correlated objects involved in the target image, it is determined whether a number of hand objects that are correlated with the face object in the current combination reaches a first preset threshold, and it

is determined whether a number of face objects that are correlated with the hand object in the current combination reaches a second preset threshold.

[0119] The first preset threshold is an empirical threshold that can be set according to actual situations. Here, the first preset threshold can be 2.

5 **[0120]** The second preset threshold is an empirical threshold that can be set according to actual situations. Here, the second preset threshold can be 1.

[0121] In some examples, a combination with a correlation prediction score reaching a preset score threshold can be determined as current pair of objects, based on an order of the third correlation prediction scores from high to low.

10 **[0122]** In the examples of the present disclosure, a combination with a correlation prediction score reaching a preset score threshold can be determined as current pair of objects. Correlation determination is performed on the current pair of objects, thereby the accuracy of the correlation prediction result can be improved.

[0123] In some examples, a counter can be maintained for each face object and each
15 hand object. For any face object, if one hand object is determined as being correlated with the face object, the value of the counter corresponding to the face object is increased by 1. In this case, two counters can be used to determine whether the number of hand objects that are correlated with the face object reaches the first preset threshold, and to determine whether the number of face objects that are correlated
20 with the hand object in the current pair of objects reaches the second preset threshold.

[0124] In the second step, in response to that the number of hand objects that are correlated with the face object in the current combination is lower than the first preset threshold, and the number of face objects that are correlated with the hand object in the current combination is lower than the second preset threshold, the face object and
25 the hand object in the current combination are determined as correlated objects

involved in the target image.

[0125] According to the above solutions, in complex scenarios (for example, the target image involves a plurality of people with overlapping faces, limbs, and hands), it can avoid unreasonable prediction such as that one face object is predicted as being correlated with more than two hand objects or that one hand object is predicted as being correlated with more than one face object. For example, in a multiplayer tabletop game scenario, where hands or faces of different people may overlap or shield each other, the solutions can correlate hands with respective faces with a higher accuracy.

10 **[0126]** Since a face object and a hand object that are strongly correlated to the same preset body part object are very likely to belong to the same person in actual situations, predicting correlated objects based on the adjusted third correlation prediction score can effectively improve the accuracy of predicting correlation.

15 **[0127]** In some examples, the detection result of the correlated objects involved in the target image can be output.

[0128] For example, in a tabletop game scenario, a bounding box corresponding to the face object and the hand object indicated by the correlated objects can be output on an image output device (such as a display). By outputting the detection result of the correlated objects on the image output device, an observer can conveniently and intuitively determine the correlated objects involved in the target image displayed on the image output device, thereby facilitating further manual verification on the detection result of the correlated objects.

20 **[0129]** The following will describe an example in a tabletop game scenario. It should be understood that for implementation in other scenarios, reference can be made to the description of the tabletop game scenario example in the present

disclosure, which is not described in detail here.

[0130] In a tabletop game scenario, a game table is usually provided, and game participants surround the game table. An image capture device for capturing live images of a tabletop game can be deployed in the tabletop game scenario. The live
5 image can involve the faces, hands, and elbows of the game participants. In this scenario, it is expected to determine the hand and face that are correlated objects involved in the live image, so that the personal identity information to which the hand belongs can be determined based on the face correlated with the hand involved in the image.

10 **[0131]** Here, the hand and the face are correlated objects, or the hand and the face are correlated, which means that the two belong to the same body, that is, the two are the hand and the face of the same person.

[0132] In this scenario, a detection device for detecting correlation between a face and a hand can also be deployed. The device can obtain live images from the image
15 capture device and determine the correlated objects involved in the live images.

[0133] The detection device can be equipped with a trained face-elbow-hand object detecting network, a face-hand correlation predicting network, and a face-elbow-hand correlation predicting network. The input of the correlation predicting network can include the output of the face-elbow-hand object detecting network.

20 **[0134]** The face-elbow-hand object detecting network can include a neural network constructed based on the FASTER-RCNN network. The object detecting network can detect bounding boxes respectively corresponding to face objects, hand objects, and elbow objects from live images.

[0135] The face-hand correlation predicting network and the face-elbow-hand
25 correlation predicting network can be a neural network constructed based on a region

feature extracting unit and a fully connected layer.

[0136] The face-elbow-hand correlation predicting network can extract visual features corresponding to the face and the hand, and combine the position features of the bounding boxes corresponding to the face and the hand which are detected by the object detecting network, to predict a third correlation prediction score between the face and the hand.

[0137] The face-elbow-hand correlation predicting network can include a face-elbow correlation predicting network and an elbow-hand correlation predicting network. The face-elbow-hand correlation predicting network can respectively predict the first correlation prediction score between the detected face and the detected elbow, and the second correlation prediction score between the detected elbow and the detected hand.

[0138] In the examples of the present disclosure, the detection device can obtain live images from the image capture device in response to a user's operation or periodically.

[0139] Then, the object detecting network can be used to detect the face objects, the hand objects, and the elbow objects involved in the live image.

[0140] Then, any of detected face objects and any of detected hand objects can be combined to obtain a plurality of first combinations. After that, for each of the first combinations, the face-hand correlation predicting network is used to predict a correlation between the face and the hand in the first combination, to obtain a third correlation prediction score between the face and the hand in each of the first combinations.

[0141] Similarly, any of the detected faces and any of detected elbows can be combined to obtain a plurality of second combinations, and any of the detected

elbows and any of the detected hands can be combined to obtain a plurality of third combinations. First correlation prediction scores between the faces and the elbows in the second combinations, and second correlation prediction scores between the elbows and the hands in the third combinations, can be respectively predicted with the
5 face-elbow-hand correlation predicting network.

[0142] Since a face object and a hand object that are strongly correlated to the same elbow object are very likely to belong to the same person in actual situations, performing correlated object prediction based on the adjusted third correlation prediction scores can improve the accuracy of predicting correlation effectively.

10 **[0143]** After that, each of the detected elbows can be taken as the target elbow in turn, to perform the following step:

[0144] determining a target face object of which a first correlating prediction score in the first correlation prediction result with respect to the target elbow reaches a first preset threshold and has the largest value; determining a target hand object of which a
15 second correlating prediction score in the second correlation prediction result with respect to the target elbow reaches a second preset threshold and has the largest value. Then, an average of the determined first correlation prediction score and the determined second correlation prediction score is calculated, and a sum of the average and the third correlation prediction score between the target face and the target hand is
20 calculated, to obtain an adjusted third correlation prediction score.

[0145] In the detection device, a counter can also be maintained for each face object and each hand object. For any face object, if one hand object is determined as being correlated with the face object, the value of the counter corresponding to the face object is increased by 1. In this case, two counters can be used to determine whether
25 the number of hand objects that are correlated with the face object reaches the first

preset threshold, and to determine whether the number of face objects that are correlated with the hand object in the current pair of objects reaches the second preset threshold.

5 **[0146]** Further, it is possible to determine each of the combinations as the current combination in turn in the order of the third correlation prediction scores from high to low, and perform the following steps:

10 **[0147]** obtaining a first value from the counter corresponding to the face object in the current combination, and determining whether the first value reaches 2; and obtaining a second value from the counter corresponding to the hand object in the current combination, and determining whether the second value reaches 1.

15 **[0148]** If the first value is lower than 2 and the second value is lower than 1, it means that the number of hand object that correlated with the face object is less than 2, and the number of face object that correlated with the hand object is less than 1. Therefore, the face object and the hand object of the current objects pair can be determined as correlated objects involved in the live image.

20 **[0149]** According to the above solutions, in complex scenarios (for example, the target image involves a plurality of people with overlapping faces, limbs, and hands), it can avoid unreasonable prediction such as that one face object is predicted as being correlated with more than two hand objects or that one hand object is predicted as being correlated with more than one face object.

25 **[0150]** The detection device is also equipped with a display unit.

[0151] The display unit can output a bounding box involved the face object and the hand objected indicated by the correlated objects on the display mounted on the detection device. By outputting the detection result of the correlated objects on the display, an observer can determine the correlated objects involved in the live image

displayed on the image output device conveniently and intuitively, thereby facilitating further manual verification on the detection result of the correlated objects.

[0152] The detection device can also obtain live images in real time, and determine the correlated hand objects and face objects from the live images. The detection
5 device can recognize the action being performed or the area being touched by the hand object involved in the live image. If the detection device recognizes that the hand object performed actions such as fetching/releasing game props, or touching a preset game area, the personal identity of the relevant person can be determined based on the face object correlated with the hand object. After that, the identity of the
10 determined person can be output to facilitate the management of the tabletop game manager.

[0153] The solution of determining correlated objects involved in the target image according to the present disclosure has been described above. In the following, a method for training a target object detecting model and a correlation predicting
15 network used in the solution will be described.

[0154] In the present disclosure, in order to improve the accuracy of the determination results of correlated face object and hand object, each model can be trained in stages. The first stage is the training of the target object detecting network; and the second stage is the joint training of the target object detecting network and the
20 correlation predicting model.

[0155] Referring to FIG. 5, as shown in FIG. 5, a target object detecting network, a first preset network, and a second preset network can be included. The target object detecting network is configured to detect a face object, a hand object, and a preset body part object involved in the target image. The first preset network is configured to
25 predict a third correlation prediction result between the face object and the hand

object. The second preset network is used to predict the first correlation prediction result between the face object and the preset body part object, and the second correlation prediction result between the preset body part object and the hand object.

[0156] The first preset network includes a face-hand correlation detecting model; the second preset network includes a face-preset-body-part correlation detecting model and a preset-body-part-hand correlation detecting model. The target object detecting network, the face-hand correlation detecting model, the face-preset-body-part correlation detecting model, and the preset-body-part-hand correlation detecting model share the same backbone network.

10 **[0157]** Referring to FIG. 7, FIG. 7 is a flowchart illustrating a method for training a neural network in the method for detecting correlated objects according to the present disclosure.

[0158] As shown in FIG. 7, the method includes the following steps.

[0159] At S702, the target object detecting network is trained based on a first training sample set; where the first training sample set contains a plurality of training samples containing first label information; the first label information contains position label information of face objects, hand objects and preset body part objects. In some examples, the position label information can include position label information of bounding boxes.

20 **[0160]** At this step, the original image can be labelled with a true or false value by manual labelling or machine-aided labelling. For example, in a tabletop game scenario, after obtaining the original image, an image annotation tool can be used to label position label information respectively corresponding to the bounding boxes of the face object, the bounding box of the hand object, and the bounding box of the
25 preset body part object involved in the original image, to obtain several training

samples. It should be noted that when encoding the training samples, one-hot encoding and other methods can be used for encoding, and the specific encoding method is not limited in the present disclosure.

5 **[0161]** After that, the target object detecting network can be trained based on the preset loss function until the network converges.

10 **[0162]** Then at S704, joint training is performed on the target object detecting network, the face-hand correlation detecting model, the face-preset-body-part correlation detecting model, and the preset-body-part-hand correlation detecting model based on a second training sample set. The second training sample set includes
a plurality of training samples containing second label information. The second label information includes position label information of face objects, hand objects and preset body part objects, and label information on correlations between face objects, preset body part objects and hand objects.

15 **[0163]** When this step is performed, the original image can be labelled with a true or false value by manual labelling or machine-aided labelling. For example, after obtaining the original image, on one hand, an image annotation tool can be used to label position label information respectively corresponding to the bounding boxes of the face objects and the bounding boxes of the hand objects and the bounding boxes of the preset body part objects (such as elbows) involved in the original image; on the
20 other hand, the image annotation tool can be used to randomly combine the face objects and the preset body part objects involved in the original image, randomly combine the face objects and the hands objects involved in the original image, and randomly combine the preset body part objects and the hands objects involved in the original image, to obtain a plurality of combinations. Then, the two parts in each
25 combination are labelled with a correlation result. In some examples, if the two parts

in one combination has correlation (belonging to the same person), then 1 is labelled; otherwise, 0 is labelled.

[0164] After a second training sample set is determined, a joint learning loss function can be determined based on the loss functions respectively corresponding to
5 the models.

[0165] In some examples, the loss functions respectively corresponding to the models can be added to obtain the joint learning loss function.

[0166] It should be noted that, in the present disclosure, a hyper parameter such as regularization items can also be added to the joint learning loss function. The type of
10 hyper parameter to be added is not particularly limited here.

[0167] After that, the models can be jointly trained based on the joint learning loss function and the second training sample set until the models converge.

[0168] Since the supervised joint training method is used in the model training, the models can be trained simultaneously, so that the models can restrict and promote
15 each other in the training process. It can improve the convergence efficiency of the models on the one hand; on the other hand, promote the backbone network shared by all models to extract more beneficial features for predicting correlation, thereby improving the accuracy of predicting correlation.

[0169] Corresponding to any of the above implements, the present application also
20 proposes an apparatus 80 for detecting correlated objects involved in an image.

[0170] Referring to FIG. 8, FIG. 8 is a block diagram illustrating an apparatus for detecting correlated objects involved in an image according to the present disclosure.

[0171] As shown in FIG. 8, the apparatus 80 includes: a detector 81 configured to detect a face object, a hand object and a preset body part object involved in a target
25 image, where the preset body part object represents a preset connection part between a

face and a hand; a first correlation predicting unit 82 configured to respectively predict correlation between the detected face object and the detected preset body part object, and correlation between the preset body part object and the detected hand object, to obtain a first correlation prediction result between the face object and the preset body part object, and a second correlation prediction result between the preset body part object and the hand object; a determining unit 84 configured to, based on the first correlation prediction result and the second correlation prediction result, determine correlated objects involved in the target image.

[0172] In some illustrated examples, the apparatus 80 also includes: the second correlation predicting unit 83 configured to predict correlation between the detected face object and the detected hand object to obtain a third correlation prediction result. The determining unit 84 includes: an adjustment sub-unit configured to adjust the third correlation prediction result based on the first correlation prediction result and the second correlation prediction result; a determining sub-unit configured to determine correlated objects involved in the target image based on the adjusted third correlation prediction result.

[0173] In some illustrated examples, the target image involves a plurality of face objects and a plurality of hand objects. The second correlation predicting unit 83 is specifically configured to: combining each of the detected face objects with each of the detected hand objects to form a plurality of first combinations; obtaining a third correlation prediction result between the face object and the hand object in each of the first combinations by, for each of the first combinations, predicting correlation between the face object and the hand object in the first combination based on visual features and position features of the face object and the hand object in the first combination, predict correlation between the face object and the hand object in the

first combination.

[0174] In some illustrated examples, the target image also involves a plurality of preset body part objects. The first correlation predicting unit 82 is specifically configured to: combining each of the detected face objects with each of the detected
5 preset body part objects to form a plurality of second combinations; obtaining a first correlation prediction result between the face object and the preset body part object in each of the second combinations by, for each of the second combinations, predicting correlation on the face object and the preset body part object in the second combination based on visual features and position features of the face object and the
10 preset body part object in the second combination; combining each of the preset body part objects with each of the hand objects to form a plurality of third combinations; and obtaining a second correlation prediction result between the preset body part object and the hand object in each of the third combinations by, for each of the third combinations, predicting correlation between the preset body part object and the hand
15 object in the third combination based on visual features and position features of the preset body part object and the hand object in the third combination.

[0175] In some illustrated examples, the correlation prediction result includes a correlation prediction score. The determining sub-unit is specifically configured to: take each of the detected preset body part objects as a target body part object, and
20 perform the following operations: determining a target face object of which a first correlation prediction score in the first correlation prediction result with respect to the target body part object is highest; determining a target hand object of which a second correlation prediction score in the second correlation predication result with respect to the target body part object is highest; and based on the first correlation prediction
25 score of the target face object with respect to the target body part object, and the

second correlation prediction score of the target body part object with respect to the target hand object, adjusting a third correlation prediction score in the third correlation prediction result between the target face object and the target hand object.

[0176] In some illustrated examples, the determining sub-unit is specifically configured to: determine candidate face objects, where each of the candidate face objects has a first correlation prediction score with respect to the target body part object greater than a preset threshold; selecting one from the candidate face objects of which the first correlation prediction score with respect to the target body part object is highest as the target face object; and/or determining a target hand object of which a second correlation prediction score in the second correlation predication result with respect to the target body part object is highest comprises: determining candidate hand objects wherein each of the candidate hand objects has a second correlation prediction score with respect to the target body part object greater than a preset threshold; and selecting one from the candidate hand objects of which the second correlation prediction score with respect to the target body part object is highest as the target hand object.

[0177] In some illustrated examples, the determining sub-unit is specifically configured to:

determine an average value of the first correlation prediction score of the target face object with respect to the preset body part object, and the second correlation prediction score of the target hand object with respect to the target body part object; and obtain the adjusted third correlation prediction score by adding the average value to third correlation prediction score between the target face object and the target hand object.

[0178] In some illustrated examples, the determining sub-unit is specifically

configured to:

selecting each from a plurality of the third correlation prediction scores in an order of the third correlation prediction scores from high to low, and for a current combination of the face object and the hand object corresponding to the selected third correlation prediction score, based on determined correlated objects involved in the target image, determining a number of hand objects that are correlated with the face object in the current combination as a first number, and determining a number of face objects that are correlated with the hand object in the current combination as a second number; in response to that the first number is lower than a first preset threshold, and the second number is lower than a second preset threshold, determining the face object and the hand object in the current combination as correlated objects involved in the target image.

[0179] In some illustrated examples, the determining unit 84 is specifically configured to: based on the first correlation prediction result and the second correlation prediction result, determine a face object and a hand object of which correlations with respect to a same preset body part object satisfying a preset condition as correlated objects involved in the target image.

[0180] In some illustrated examples, the apparatus 80 also comprises: an output unit configured to output a detection result of the correlated objects involved in the target image.

[0181] In some illustrated examples, the preset body part comprises at least one of a shoulder object, an elbow object and a wrist object.

[0182] In some illustrated examples, the face object, the hand object, and the preset body part object involved in the target image are detected from the target image by a target object detecting network; the third correlation prediction result is detected by a

first preset network comprising a face-hand correlation detecting model; the first correlation prediction result and the second correlation prediction result are detected by a second preset network comprising a face-preset-body-part correlation detecting model and a preset-body-part-hand correlation detecting model; the target object
5 detecting network, the face-hand correlation detecting model, the face-preset-body-part correlation detecting model and the preset-body-part-hand correlating model are trained by: training the target object detecting network based on a first training sample set which comprises a plurality of training samples with respective first label information, wherein the first label information contains
10 respective position label information of face objects, hand objects and preset body part objects; and jointly training the target object detecting network, the face-hand correlation detecting model, the face-preset-body-part correlation detecting model, and the preset-body-part-hand correlation detecting model based on a second training sample set which comprises a plurality of training samples with respective second
15 label information, wherein the second label information includes respective position label information of face objects, hand objects and preset body part objects, and respective label information on correlations between face objects, preset body part objects and hand objects.

[0183] The examples of the apparatus for detecting correlated objects involved in an
20 image according to the present disclosure can be used in an electronic device. Correspondingly, the present disclosure discloses an electronic device, which can include a processor; a memory configured to store executable instructions of the processor. The processor is configured to invoke the executable instructions stored in the memory to implement the method for detecting correlated objects in an image as
25 shown in any of the examples.

[0184] Referring to FIG. 9, FIG. 9 is a block diagram of hardware of an electronic device according to the present disclosure.

[0185] As shown in FIG. 9, the electronic device can include a processor for executing instructions, a network interface for network connection, a memory for storing operating data for the processor, and a non-transitory memory for storing instructions corresponding to the apparatus for detecting correlated objects involved in an image.

[0186] The examples of the apparatus for detecting correlated objects involved in an image can be implemented by software, or can be implemented by hardware or a combination of software and hardware. Taking software implementation as an example, as a logical device, it is formed by reading the corresponding computer program instructions in the non-transitory memory into the memory through the processor of the electronic device where it is located. From a hardware perspective, in addition to the processor, memory, network interface, and non-transitory memory shown in FIG. 9, the electronic device in which the device is located in the example usually includes other hardware based on the actual functions of the electronic device, which will not be elaborated here.

[0187] It should be understood that, in order to improve the processing speed, the corresponding instructions of the apparatus for detecting correlated objects involved in an image can also be directly stored in the memory, which is not limited here.

[0188] The present disclosure provides a non-transitory computer-readable storage medium, the storage medium stores a computer program, and the computer program is configured to perform any method for detecting correlated objects involved in an image according to any of the above examples.

[0189] Those skilled in the art should understand that one or more examples of the

present disclosure can be provided as a method, a system, or a computer program product. Therefore, one or more examples of the present disclosure can adopt the form of a complete hardware example, a complete software example, or an example combining software and hardware. Moreover, one or more examples of the present disclosure can be in a form of a computer program product having one or more computer-usable storage media (which can include but not limited to disk storage, CD-ROM, optical storage, etc.) containing computer-usable program codes.

[0190] In the present disclosure, "and/or" means having at least one of the two. For example, "A and/or B" can include three schemes: A, B, and "A and B".

[0191] The various examples in the present disclosure are described in a progressive manner, and the same or similar parts between the various examples can be referred to each other, and each example focuses on the differences from other examples. In particular, as for the data processing device example, since it is basically similar to the method example, the description is relatively simple, and for related parts, reference can be made to the part of the description of the method example.

[0192] The above has described specific examples of the present disclosure. Other examples are within the scope of the appended claims. In some cases, the actions or steps described in the claims can be performed in a different order from that in the examples and still achieve desired results. In addition, the processes depicted in the drawings do not necessarily require the specific order or sequential order shown to achieve the desired result. In some examples, multitasking and parallel processing are also possible or can be advantageous.

[0193] The examples of the subject and functional operations described in the present disclosure can be implemented in the following: digital electronic circuits, tangible computer software or firmware, computer hardware that can include the

structures disclosed in the present disclosure and their structural equivalents, or one or more of them. The examples of the subject matter described in the present disclosure can be implemented as one or more computer programs, that is, one or more units in the computer program instructions that are encoded in the tangible non-transitory
5 program carrier to be executed by the data processing device or to control the operation of the data processing device. Alternatively or in addition, the program instructions can be encoded in 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 to be performed by
10 the data processing device. The computer storage medium can be a machine-readable storage device, a machine-readable storage substrate, a random or serial access memory device, or a combination of one or more of them.

[0194] The processing and logic flows described in the present disclosure can be executed by one or more programmable computers executing one or more computer
15 programs, to perform corresponding functions by operating according to input data and generating output. The processing and logic flow can also be executed by a dedicated logic circuit, such as FPGA (Field Programmable Gate Array) or ASIC (Application Specific Integrated Circuit), and the device can also be implemented as a dedicated logic circuit.

[0195] A computer suitable for executing a computer program can include, for example, a general-purpose and/or special-purpose microprocessor, 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 random access memory. The basic components of a computer can include a central processing unit for
20 implementing or executing instructions and one or more memory devices for storing
25

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, etc., or the computer will be operatively coupled with this mass storage device to receive data or send data to it, or both. However, the computer does not have to have such equipment. In addition, the computer can 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 universal serial bus (USB) flash drives are portable storage devices, to name a few.

[0196] Computer readable media suitable for storing computer program instructions and data can include all forms of non-transitory memory, media and memory devices, such as semiconductor memory devices (such as EPROMs, EEPROMs and flash memory devices), magnetic disks (such as internal hard disks or removable disks), magneto-optical disks, CD ROMs and DVD-ROM disks. The processor and the memory can be supplemented by or incorporated into a dedicated logic circuit.

[0197] Although the present disclosure contains many specific implementation details, these should not be construed as limiting any disclosed scope or claimed scope, but are mainly used to describe the features of specific disclosed examples. Certain features described in a plurality of examples in the present disclosure can also be implemented in combination in a single example. On the other hand, various features described in a single example can also be implemented in a plurality of examples separately or in any suitable sub-combination. In addition, although features can function in certain combinations as described above and even as originally claimed, one or more features from the claimed combination can in some cases be removed from the combination, and the claimed combination can refer to a sub-combination or a variant of the sub-combination.

[0198] Similarly, although operations are depicted in a specific order in the drawings, this should not be construed 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 processing can be advantageous. In addition, the separation of various system units and components in the examples should not be understood as requiring such separation in all examples, and it should be understood that the described program components and systems can usually be integrated in a single software product, or packaged into a plurality of software products.

[0199] Thus, specific examples of the subject matter have been described. Other examples are within the scope of the appended claims. In some cases, the actions recited in the claims can be performed in a different order and still achieve desired results. In addition, the processes depicted in the drawings are not necessarily in the specific order or sequential order shown in order to achieve the desired result. In some implementations, multitasking and parallel processing can be advantageous.

[0200] The above are only preferred examples of one or more examples of the present disclosure, and are not used to limit one or more examples of the present disclosure. Any modification, equivalent replacement, improvement within the spirit and principle of one or more examples of the present disclosure shall be included in the protection scope of one or more examples of the present disclosure.

WHAT IS CLAIMED IS:

1. A method for detecting correlated objects involved in an image, comprising:
 - detecting a face object, a hand object and a preset body part object involved in a
 - 5 target image, wherein the preset body part object represents a preset connection part
 - between a face and a hand;
 - respectively predicting correlation between the detected face object and the
 - detected preset body part object, and correlation between the detected preset body part
 - object and the detected hand object, to obtain a first correlation prediction result
 - 10 between the face object and the preset body part object, and a second correlation
 - prediction result between the preset body part object and the hand object;
 - predicting correlation between the detected face object and the detected hand
 - object to obtain a third correlation prediction result,
 - adjusting the third correlation prediction result based on the first correlation
 - 15 prediction result and the second correlation prediction result; and
 - determining correlated objects involved in the target image based on the adjusted
 - third correlation prediction result.
2. The method of claim 1, wherein
 - 20 the target image involves a plurality of face objects and a plurality of hand objects;
 - and
 - predicting correlation between the detected face object and the detected hand
 - object to obtain a third correlation prediction result comprises:
 - combining each of the detected face objects with each of the detected hand
 - 25 objects to form a plurality of first combinations;

obtaining a third correlation prediction result between the face object and the hand object in each of the first combinations by, for each of the first combinations, predicting correlation between the face object and the hand object in the first combination based on visual features and position features of the face object and the hand object in the first combination.

3. The method of claim 1 or 2, wherein the target image involves a plurality of preset body part objects; and respectively predicting correlation between the detected face object and the detected preset body part object, and correlation between the preset body part object and the hand object, to obtain a first correlation prediction result between the face object and the preset body part object, and a second correlation prediction result between the preset body part object and the hand object comprises:

combining each of the detected face objects with each of the preset body part objects to form a plurality of second combinations;

obtaining a first correlation prediction result between the face object and the preset body part object in each of the second combinations by, for each of the second combinations, predicting correlation on the face object and the preset body part object in the second combination based on visual features and position features of the face object and the preset body part object in the second combination;

combining each of the preset body part objects with each of the hand objects to form a plurality of third combinations; and

obtaining a second correlation prediction result between the preset body part object and the hand object in each of the third combinations by, for each of the third combinations, predicting correlation between the preset body part object and the hand

object in the third combination based on visual features and position features of the preset body part object and the hand object in the third combination.

4. The method of claim 1, wherein adjusting the third correlation prediction result
5 based on the first correlation prediction result and the second correlation prediction result comprises:

determining a target face object of which a first correlation prediction score in the first correlation prediction result with respect to the preset body part object is highest;

determining a target hand object of which a second correlation prediction score in
10 the second correlation predication result with respect to the preset body part object is highest; and

based on the first correlation prediction score of the target face object with respect to the preset body part object, and the second correlation prediction score of the target hand object with respect to the preset body part object, adjusting a third
15 correlation prediction score in the third correlation prediction result between the target face object and the target hand object.

5. The method of claim 4, wherein

determining a target face object of which a first correlation prediction score in the
20 first correlation prediction result with respect to the preset body part object is highest comprises:

determining candidate face objects, wherein each of the candidate face objects has a first correlation prediction score with respect to the target body part object greater than a preset threshold; and

25 selecting one from the candidate face objects of which the first correlation

prediction score with respect to the preset body part object is highest as the target face object; and/or

determining a target hand object of which a second correlation prediction score in the second correlation predication result with respect to the preset body part object is
5 highest comprises:

determining candidate hand objects, wherein each of the candidate hand objects has a second correlation prediction score with respect to the target body part object greater than a preset threshold; and

selecting one from the candidate hand objects of which the second
10 correlation prediction score with respect to the preset body part object is highest as the target hand object.

6. The method of claim 4 or 5, wherein based on the first correlation prediction score of the target face object with respect to the preset body part object, and the
15 second correlation prediction score of the target hand object with respect to the preset body part object, adjusting the third correlation prediction score in the third correlation prediction result between the target face object and the target hand object comprises:

determining an average value of the first correlation prediction score of the target
20 face object with respect to the preset body part object, and the second correlation prediction score of the target hand object with respect to the preset body part object; and

obtaining the adjusted third correlation prediction score by adding the average value to the third correlation prediction score between the target face object and the
25 target hand object.

7. The method of claim 6, wherein determining correlated objects involved in the target image based on the adjusted third correlation prediction result comprises:

selecting each from a plurality of the third correlation prediction scores in an
5 order of the third correlation prediction scores from high to low, and

for a current combination of the face object and the hand object corresponding to the selected third correlation prediction score:

based on determined correlated objects involved in the target image,
determining a number of hand objects that are correlated with the face object in the
10 current combination as a first number, and determining a number of face objects that
are correlated with the hand object in the current combination as a second number;
and

in response to that the first number is lower than a first preset threshold, and
the second number is lower than a second preset threshold, determining the face
15 object and the hand object in the current combination as correlated objects involved in
the target image.

8. The method of any one of claims 1-7, wherein determining correlated objects
involved in the target image based on the first correlation prediction result and the
20 second correlation prediction result comprises:

based on the first correlation prediction result and the second correlation
prediction result, determining a face object and a hand object of which correlations
with respect to a same preset body part object satisfying a preset condition as
correlated objects involved in the target image.

25

9. The method of any one of claims 1-8, further comprising:

outputting a detection result of the correlated objects involved in the target image.

10. The method of any one of claims 1-9, wherein the preset body part comprises
5 at least one of a shoulder object, an elbow object and a wrist object.

11. The method of any one of claims 1-10, wherein

the face object, the hand object, and the preset body part object involved in the
target image are detected from the target image by a target object detecting network;

10 the third correlation prediction result is detected by a first preset network
comprising a face-hand correlation detecting model;

the first correlation prediction result and the second correlation prediction result
are detected by a second preset network comprising a face-preset-body-part
correlation detecting model and a preset-body-part-hand correlation detecting model;

15 and

the target object detecting network, the face-hand correlation detecting model, the
face-preset-body-part correlation detecting model and the preset-body-part-hand
correlation detecting model are trained by:

training the target object detecting network based on a first training sample
20 set which comprises a plurality of training samples with respective first label
information, wherein the first label information contains respective position label
information of face objects, hand objects and preset body part objects; and

jointly training the target object detecting network, the face-hand correlation
detecting model, the face-preset-body-part correlation detecting model, and the
25 preset-body-part-hand correlation detecting model based on a second training sample

set which comprises a plurality of training samples with respective second label information, wherein the second label information contains respective position label information of face objects, hand objects and preset body part objects, and respective label information on correlations between face objects, preset body part objects and
5 hand objects.

12. An apparatus for detecting correlated objects involved in an image, comprising:

a detector configured to detect a face object, a hand object and a preset body part
10 object involved in a target image, wherein the preset body part object represents a preset connection part between a face and a hand;

a first correlation predicting unit configured to respectively predict correlation between the detected face object and the detected preset body part object, and correlation between the detected preset body part object and the detected hand object,
15 to obtain a first correlation prediction result between the face object and the preset body part object, and a second correlation prediction result between the preset body part object and the hand object; and

a second correlation predicting unit configured to predict correlation between the detected face object and the detected hand object to obtain a third correlation
20 prediction result;

a determining unit configured to, determine correlated objects involved in the target image based on the first correlation prediction result and the second correlation prediction result;

a determining unit includes: an adjustment sub-unit configured to adjust the third
25 correlation prediction result based on the first correlation prediction result and the

second correlation prediction result; a determining sub-unit configured to determine correlated objects involved in the target image based on the adjusted third correlation prediction result.

- 5 13. An electronic device comprising:
- a processor;
- a memory for storing executable instructions of the processor;
- wherein the processor is configured to invoke the executable instructions stored
- in the memory to implement the method for detecting correlated objects involved in
- 10 an image according to any one of claims 1 to 11.

14. A non-transitory computer-readable storage medium, wherein the storage
- medium stores a computer program, and the computer program is configured to
- perform the method for detecting correlated objects involved in an image according to
- 15 any one of claims 1 to 11.