

METHOD, APPARATUS AND DEVICE FOR RECOGNIZING STACKED OBJECTS, AND COMPUTER STORAGE MEDIUM

TECHNICAL FIELD

[0001] Embodiments of this disclosure relate to, but not limited to, the technical field of computer vision, and in particular, to a method, apparatus and device for recognizing stacked objects, and a computer storage medium.

BACKGROUND

[0002] In some scenes, many products need to be produced or used in batches, and some of these products are stackable objects. Stackable objects can be stacked along a certain stacking direction to form an object sequence.

[0003] In a video analysis scene, since objects forming a stack are similar in appearance, for example, they may all be flaky objects of the same size, and the number of the objects forming the stack is uncertain, it is difficult to recognize the composition of sequence objects, i.e., each object in the stack, on the basis of images or videos.

SUMMARY

[0004] Embodiments of this disclosure provide a method, apparatus and device for recognizing stacked objects, and a computer storage medium.

[0005] The first aspect provides a method for recognizing stacked objects, including: a to-be-recognized image is obtained, the to-be-recognized image including an object sequence formed by stacking at least one object; edge detection is performed on the object sequence on the basis of the to-be-recognized image to determine an edge segmentation image of the object sequence, the edge segmentation image including edge

information of each object forming the object sequence; and the category of each object in the object sequence is recognized on the basis of the to-be-recognized image and the edge segmentation image.

[0006] In some embodiments, the operation in which the category of each object in the object sequence is recognized on the basis of the to-be-recognized image and the edge segmentation image includes: each object in the object sequence of the to-be-recognized image is segmented on the basis of the edge segmentation image to obtain a sub-image corresponding to each object; and category recognition is performed on each sub-image to obtain the category of each object.

[0007] In this case, each object in the object sequence of the to-be-recognized image is segmented on the basis of the edge segmentation image to obtain the sub-image corresponding to each object, and category recognition is performed on each sub-image to obtain the category corresponding to each sub-image. Therefore, the category of each object in the object sequence can be determined accurately on the basis of the category corresponding to each sub-image.

[0008] In some embodiments, the operation in which each object in the object sequence of the to-be-recognized image is segmented on the basis of the edge segmentation image to obtain a sub-image corresponding to each object includes: first position information of each object in the to-be-recognized image is determined on the basis of the edge information of each object forming the object sequence; and each object in the to-be-recognized image is segmented on the basis of the first position information to obtain each sub-image.

[0009] In this case, the first position information of each object in the to-be-recognized image is determined on the basis of the edge information of each object forming the object sequence, and then each object in the object sequence of the to-be-recognized image is segmented on the basis of the first position information to obtain each sub-image. Therefore, position information of each object in the object sequence in the edge segmentation image can be located accurately on the basis of the first position information, and then the object sequence of the to-be-recognized image is segmented on the basis of the first position information to obtain each sub-image, so as to accurately

match each object in the object sequence to accurately determine the category of each object in the object sequence.

[0010] In some embodiments, the operation in which category recognition is performed on each sub-image to obtain the category of each object includes: category recognition is performed on each sub-image to obtain at least two categories and at least two confidences having one-to-one correspondence to the at least two categories; and in a case where the difference between the highest confidence and the second highest confidence in the at least two confidences is greater than a threshold, the category corresponding to the highest confidence is determined as the category of the object corresponding to the sub-image.

[0011] In this case, in a case where the difference between the highest confidence and the second highest confidence is greater than the threshold, the category corresponding to the highest confidence is determined as the category of the object corresponding to the sub-image. Therefore, the category of each object in the object sequence can be determined accurately.

[0012] In some embodiments, the method further includes at least one of the following operations.

[0013] In a case where the difference is less than or equal to the threshold and the category corresponding to the highest confidence is the same as the categories of sub-images corresponding to two second objects adjacent to a first object corresponding to the sub-image in the object sequence, the category corresponding to the second highest confidence is determined as the category of the first object corresponding to the sub-image.

[0014] In a case where the difference is less than or equal to the threshold and the category corresponding to the second highest confidence is the same as the categories of sub-images corresponding to two second objects adjacent to a first object corresponding to the sub-image in the object sequence, the category corresponding to the highest confidence is determined as the category of the first object corresponding to the sub-image.

[0015] In a case where the difference is less than or equal to the threshold and the category corresponding to the second highest confidence is different from the categories of sub-images corresponding to two second objects adjacent to a first object corresponding to the sub-image in the object sequence, the category corresponding to the highest confidence is determined as the category of the first object corresponding to the sub-image.

[0016] In this case, in a case where the difference is less than or equal to the threshold, if the category corresponding to the highest confidence determined on the basis of each sub-image is the same as the categories of two sub-images adjacent to each sub-image, the category corresponding to the second highest confidence is determined as the category of the object corresponding to each sub-image; if the category corresponding to the second highest confidence determined on the basis of each sub-image is the same as the categories of two sub-images adjacent to each sub-image, the category corresponding to the highest confidence is determined as the category of the object corresponding to each sub-image, so as to eliminate the effect of the sub-images adjacent to each sub-image on determination of the category of each sub-image; and if the category corresponding to the second highest confidence is different from the categories of two sub-images adjacent to each sub-image, the category corresponding to the highest confidence is determined as the category of the object corresponding to each sub-image. Therefore, the category of the object corresponding to each sub-image can be determined accurately.

[0017] In some embodiments, the edge segmentation image includes a mask image representing the edge information of each object, and/or, the edge segmentation image has the same size as the to-be-recognized image.

[0018] In this case, the edge segmentation image includes the mask image representing the edge information of each object, so that the edge information of each object can be determined easily on the basis of the mask image; and the edge segmentation image has the same size as the to-be-recognized image, so that the edge position of each object in the to-be-recognized image can be determined accurately on the basis of the edge position of each object in the edge segmentation image.

- [0019] In some embodiments, the edge segmentation image is a binarized mask image, pixels of a first pixel value in the edge segmentation image correspond to pixels of the edge of each object in the to-be-recognized image, and pixels of a second pixel value in the edge segmentation image correspond to pixels of the non-edge part of each object in the to-be-recognized image.
- [0020] In this case, the edge segmentation image is the binarized mask image, so that whether each pixel point in the binarized mask image is located on the edge of each object in the object sequence can be determined depending on whether each pixel point is the first pixel value or the second pixel value. Therefore, the edge of each object in the object sequence can be determined easily.
- [0021] In some embodiments, the operation in which edge detection is performed on the object sequence on the basis of the to-be-recognized image to determine an edge segmentation image of the object sequence includes: the to-be-recognized image is input into a trained edge detection model to obtain an edge detection result for each object in the object sequence, the edge detection model being obtained by training based on a sequence object image including object edge annotation information; and the edge segmentation image of the object sequence is generated according to the edge detection result.
- [0022] In this case, the edge detection result for each object in the object sequence in the to-be-recognized image is determined on the basis of the trained edge detection model, and the trained edge detection model is obtained by training based on the sequence object image including object edge annotation information. Therefore, the edge detection result for each object in the object sequence can be determined easily and accurately through the trained edge detection model.
- [0023] In some embodiments, the operation in which category recognition is performed on each sub-image to obtain the category of each object includes: each sub-image is input into a trained object classification model to obtain the category of each corresponding object; where the object classification model is obtained by training based on single-object images, and the single-object images are obtained after segmenting the sequence object image according to the edge detection result for each object.

- [0024] In this case, the category of each object in the object sequence is determined on the basis of the trained object classification model, and the trained object classification model is obtained by training based on the single-object images. Therefore, the category of each object in the object sequence can be determined easily and accurately through the trained object classification model.
- [0025] In some embodiments, the object has a value attribute corresponding to the category; and the method may further include: the total value of objects in the object sequence is determined on the basis of the category of each object and the corresponding value attribute.
- [0026] In this case, the total value of objects in the object sequence is determined on the basis of the category of each object and the corresponding value attribute. Therefore, it is convenient to count the total value of stacked objects, for example, it is convenient to detect and determine the total value of stacked tokens.
- [0027] The second aspect provides an apparatus for recognizing stacked objects, including: an obtaining unit, configured to obtain a to-be-recognized image, the to-be-recognized image including an object sequence formed by stacking at least one object; a determination unit, configured to perform edge detection on the object sequence on the basis of the to-be-recognized image to determine an edge segmentation image of the object sequence, the edge segmentation image including edge information of each object forming the object sequence; and a recognition unit, configured to recognize the category of each object in the object sequence on the basis of the to-be-recognized image and the edge segmentation image.
- [0028] In some embodiments, the recognition unit is further configured to: segment each object in the object sequence of the to-be-recognized image on the basis of the edge segmentation image to obtain a sub-image corresponding to each object; and perform category recognition on each sub-image to obtain the category of each object.
- [0029] In some embodiments, the recognition unit is further configured to: determine first position information of each object in the to-be-recognized image on the basis of the edge information of each object forming the object sequence; and segment each object in

the to-be-recognized image on the basis of the first position information to obtain each sub-image.

[0030] In some embodiments, the recognition unit is further configured to: perform category recognition on each sub-image to obtain at least two categories and at least two confidences having one-to-one correspondence to the at least two categories; and in a case where the difference between the highest confidence and the second highest confidence in the at least two confidences is greater than a threshold, determine the category corresponding to the highest confidence as the category of the object corresponding to the sub-image.

[0031] In some embodiments, the recognition unit is further configured to implement at least one of following operations.

[0032] In a case where the difference is less than or equal to the threshold and the category corresponding to the highest confidence is the same as the categories of sub-images corresponding to two second objects adjacent to a first object corresponding to the sub-image in the object sequence, the category corresponding to the second highest confidence is determined as the category of the first object corresponding to the sub-image.

[0033] In a case where the difference is less than or equal to the threshold and the category corresponding to the second highest confidence is the same as the categories of sub-images corresponding to two second objects adjacent to a first object corresponding to the sub-image in the object sequence, the category corresponding to the highest confidence is determined as the category of the first object corresponding to the sub-image.

[0034] In a case where the difference is less than or equal to the threshold and the category corresponding to the second highest confidence is different from the categories of sub-images corresponding to two second objects adjacent to a first object corresponding to the sub-image in the object sequence, the category corresponding to the highest confidence is determined as the category of the first object corresponding to the sub-image.

[0035] In some embodiments, the edge segmentation image includes a mask image representing the edge information of each object, and/or, the edge segmentation image has the same size as the to-be-recognized image.

[0036] In some embodiments, the edge segmentation image is a binarized mask image, pixels of a first pixel value in the edge segmentation image correspond to pixels of the edge of each object in the to-be-recognized image, and pixels of a second pixel value in the edge segmentation image correspond to pixels of the non-edge part of each object in the to-be-recognized image.

[0037] In some embodiments, the determination unit is further configured to: input the to-be-recognized image into a trained edge detection model to obtain an edge detection result for each object in the object sequence, the edge detection model being obtained by training based on a sequence object image including object edge annotation information; and generate the edge segmentation image of the object sequence according to the edge detection result.

[0038] In some embodiments, the recognition unit is further configured to: input each sub-image into a trained object classification model to obtain the category of each corresponding object; and the object classification model is obtained by training based on single-object images, and the single-object images are obtained after segmenting the sequence object image according to the edge detection result for each object.

[0039] In some embodiments, the object has a value attribute corresponding to the category; and the determination unit is further configured to: determine the total value of objects in the object sequence on the basis of the category of each object and the corresponding value attribute.

[0040] A third aspect provides a device for recognizing stacked objects, including a memory and a processor. The memory stores a computer program capable of running on the processor, and when the processor executes the computer program, the steps of the foregoing method are implemented.

[0041] A fourth aspect provides a computer storage medium. The computer storage medium stores one or more programs, and the one or more programs can be executed by one or more processors so as to implement the steps of the foregoing method.

[0042] In the embodiments of this disclosure, the category of each object in the object sequence is recognized on the basis of the to-be-recognized image and the edge segmentation image, where not only the edge information of each object determined on the basis of the edge segmentation image, but also feature information of each object in the object sequence in the to-be-recognized image is considered. Therefore, the determined category of each object in the object sequence in the to-be-recognized image has high accuracy.

BRIEF DESCRIPTION OF THE DRAWINGS

[0043] The accompanying drawings described herein are used to provide a further understanding of this disclosure, and form a part of the present application. The exemplary embodiments and descriptions thereof in this disclosure are used to explain this disclosure and do not limit this disclosure in any improper way.

[0044] FIG. 1 illustrates a schematic structural diagram of a system for recognizing stacked objects provided in embodiments of this disclosure.

[0045] FIG. 2 illustrates a schematic diagram of an implementation process of a method for recognizing stacked objects provided in embodiments of this disclosure.

[0046] FIG. 3 illustrates a schematic diagram of an implementation process of another method for recognizing stacked objects provided in embodiments of this disclosure.

[0047] FIG. 4 illustrates a schematic diagram of an implementation process of yet another method for recognizing stacked objects provided in embodiments of this disclosure.

[0048] FIG. 5 illustrates a schematic diagram of an implementation process of still another method for recognizing stacked objects provided in embodiments of this disclosure.

[0049] FIG. 6 illustrates a schematic diagram of a process framework of a method for recognizing stacked objects provided in embodiments of this disclosure.

[0050] FIG. 7 illustrates a schematic structural diagram of the composition of an apparatus for recognizing stacked objects provided in embodiments of this disclosure.

[0051] FIG. 8 illustrates a schematic diagram of a hardware entity of a device for recognizing stacked objects provided in embodiments of this disclosure.

DETAILED DESCRIPTION

[0052] The technical solution of this disclosure will be specifically described hereinafter in detail through embodiments with reference to the accompanying drawings. The following specific embodiments may be combined with each other, and the same or similar concepts or processes may not be described in some embodiments.

[0053] It is to be noted that the terms "first", "second" and the like in the embodiments of this disclosure are used to distinguish between similar targets and are not necessarily used to describe a particular order or sequence.

[0054] In addition, the technical solutions recited in the embodiments of this disclosure can be arbitrarily combined without causing conflicts.

[0055] "At least one" and "at least one frame" in the embodiments of this disclosure may respectively refer to "one or at least two" and "one frame or at least two frames". "Multiple" and "multiple frames" in the embodiments of this disclosure may respectively refer to "at least two" and "at least two frames". "At least one image frame" in the embodiments of this disclosure may refer to continuously captured images, or, may refer to non-continuously captured images. The number of images may be determined depending on an actual situation, and is not limited in the embodiments of this disclosure.

[0056] To solve the problem of human resource waste caused by manually determining the category of each object in an object sequence formed by stacking, a computer vision method is proposed to recognize each object in the object sequence. For example, the following two solutions are proposed.

[0057] First solution: after an image captured for an object sequence is obtained, image features may be extracted using Convolutional Neural Networks (CNNs), then sequence modeling is performed on the features using a Recurrent Neural Network (RNN), classification prediction and duplication deletion is performed on each feature slice using a Connectionist Temporal Classification (CTC) loss function to perform an output result, and the category of each object in the object sequence can be determined on the basis of the output result. However, the main problems of the method are that part of training of RNN sequence modeling is time-consuming, the model can only be supervised by CTC loss alone, and the prediction effect is limited.

[0058] Second solution: after an image captured for an object sequence is obtained, image features can be extracted using CNNs, then attention centers are generated in combination with a vision attention mechanism, and a corresponding result is predicted for each attention center and others superfluous information is ignored. However, the main problem of the method is that the attention mechanism requires a lot of computation and memory usage.

[0059] Therefore, there is no related algorithm for specifically solving the problem of recognizing each object in the object sequence formed by stacking. Although the foregoing two methods can be applied to object sequence recognition, because the sequence of the object sequence is usually long, the objects forming the stack are similar in appearance, and the number of stacked objects is uncertain, the use of the foregoing two methods cannot achieve high-accuracy prediction of the category of each object in the object sequence.

[0060] FIG. 1 illustrates a schematic structural diagram of a system for recognizing stacked objects provided in embodiments of this disclosure. As illustrated in FIG. 1, a system 100 for recognizing stacked objects may include a camera assembly 101, a device 102 for recognizing stacked objects, and a management system 103.

[0061] In some implementations, the camera assembly 101 may include multiple cameras, the multiple cameras can photograph a surface for object placement from different angles, the surface for object placement may be the surface of a game table or a storage table, etc., and one camera assembly 101 may correspond to one surface for object placement. For example, the camera assembly 101 may include three cameras. The first camera may be a bird's eye view camera, and the first camera may be mounted on the top of the surface for object placement. The second camera and the third camera are respectively mounted on the sides of the surface for object placement, and an included angle between the second camera and the third camera is a set included angle, for example, the set included angle may range from 30 degrees to 120 degrees, and the set included angle may be 30 degrees, 60 degrees, 90 degrees, 120 degrees, etc. The second camera and the third camera may be arranged on the surface for object placement, so as to capture object and player statuses on the surface for object placement from the side view.

[0062] In some implementations, the device 102 for recognizing stacked objects may correspond to only one camera assembly 101, in which case one device 102 for recognizing stacked objects may correspond to one surface for object placement. In other implementations, the device 102 for recognizing stacked objects may correspond to multiple camera assemblies 101, in which case one device 102 for recognizing stacked objects may correspond to multiple surfaces for object placement. The device 102 for recognizing stacked objects and the surface for object placement may be arranged in a designated space (such as a game place). For example, the device 102 for recognizing stacked objects may be an edge-side device, and the device 102 for recognizing stacked objects may be connected to a server in the designated space, so that the server can control the edge-side device, and the original structure and function of the server are not affected. In other implementations, the device 102 for recognizing stacked objects may be the server in the designated space or may be arranged on the cloud.

[0063] The camera assembly 101 may be communicably connected to the device 102 for recognizing stacked objects. In some implementations, the camera assembly 101 may capture real-time images periodically or non-periodically, and send the captured real-time images to the device 102 for recognizing stacked objects. For example, in a

case where the camera assembly 101 includes multiple cameras, the multiple cameras can capture real-time images once every target duration, and send the captured real-time images to the device 102 for recognizing stacked objects. The multiple cameras can capture the real-time images at the same time or at different times. In other implementations, the camera assembly 101 may capture a real-time video, and send the real-time video to the device 102 for recognizing stacked objects. For example, in a case where the camera assembly 101 includes multiple cameras, the multiple cameras can separately send captured real-time videos to the device 102 for recognizing stacked objects, so that the device 102 for recognizing stacked objects can crop real-time images from the real-time videos. The real-time images respectively cropped from multiple real-time videos each time may be real-time images obtained at the same time. The real-time images in the embodiments of this disclosure may be any one or multiple of the following images. In other embodiments, the device 102 for recognizing stacked objects may acquire images or videos from other video sources, and the obtained images or videos may be real-time or pre-stored.

[0064] The device 102 for recognizing stacked objects may analyze, on the basis of real-time images, behaviors of the object on the surface for object placement in the designated space and a target nearby the surface for object placement (such as a game participant including a game master and/or player), so as to determine whether the behavior of the object complies with regulations or is proper.

[0065] The device 102 for recognizing stacked objects may be communicatively connected to the management system 103. The management system may include a display device. In a case where the device 102 for recognizing stacked objects determines that the behavior of the object is not proper, the device 102 for recognizing stacked objects may send warning information to the management system 103 arranged on the surface for object placement and corresponding to the object of which the behavior is not proper, so that the management system 103 can send a warning corresponding to the warning information.

[0066] In the embodiments corresponding to FIG. 1, it is illustrated that the camera assembly 101, the device 102 for recognizing stacked objects, and the management

system 103 are independent of each other. However, in other embodiments, the camera assembly 101 and the device 102 for recognizing stacked objects may be integrated together, or, the device 102 for recognizing stacked objects and the management system 103 may be integrated together, or, the camera assembly 101, the device 102 for recognizing stacked objects, and the management system 103 are integrated together.

[0067] The method for recognizing stacked objects in the embodiments of this disclosure may be applied in a game, entertainment, or competitive scene, and the objects may include tokens, game cards, game chips and the like in this scene. No specific limitation is made thereto in this disclosure.

[0068] FIG. 2 illustrates a schematic diagram of an implementation process of a method for recognizing stacked objects provided in embodiments of this disclosure. As illustrated in FIG. 2, the method is applied to an apparatus for recognizing stacked objects. The method includes the following operations.

[0069] At S201, a to-be-recognized image is obtained. The to-be-recognized image includes an object sequence formed by stacking at least one object.

[0070] In some implementations, the apparatus for recognizing stacked objects may include a device for recognizing stacked objects. In other implementations, the apparatus for recognizing stacked objects may include a processor or a chip, and the processor or the chip may be applied to the device for recognizing stacked objects. The device for recognizing stacked objects may include one or a combination of at least two of: a server, a mobile phone, a pad, a computer having a wireless transceiver function, a palm computer, a desktop computer, a personal digital assistant, a portable media player, a smart speaker, a navigation apparatus, a wearable device such as a smart watch, smart glasses, and a smart necklace, a pedometer, a digital TV, a Virtual Reality (VR) terminal device, an Augmented Reality (AR) terminal device, a wireless terminal in industrial control, a wireless terminal in self-driving, a wireless terminal in remote medical surgery, a wireless terminal in smart grid, a wireless terminal in transportation safety, a wireless terminal in smart city, a wireless terminal in smart home, or a vehicle, a vehicle-mounted device, a vehicle-mounted module and the like in an Internet of things system.

[0071] A camera mounted on a side of a surface for object placement may implement photographing for the object sequence to obtain a captured image, the camera may implement photographing once every set duration, the captured image may be an image currently captured by the camera, or the camera may capture a video, and the captured image may be an image cropped from the video. The to-be-recognized image may be determined on the basis of the captured image. In a case where one camera is used to photograph the object sequence, an image captured by the one camera is determined as the captured image. In a case where at least two cameras are used to photograph the object sequence, images captured by the at least two cameras are respectively determined as at least two captured image frames. The to-be-recognized image may include one image frame or at least two image frames, and the at least two image frames may be determined respectively on the basis of the at least two captured image frames. In other embodiments, the to-be-recognized image may be determined on the basis of an image acquired from another video source. For example, the acquired image may be directly stored in a video source, or, the acquired image may be cropped from a video stored in a video source.

[0072] In some implementations, the captured image or the acquired image may be directly determined as the to-be-recognized image.

[0073] In other implementations, the captured image or the acquired image may be processed by at least one of scaling, cropping, denoising, noise addition, grayscale, rotation, and normalization, so as to obtain the to-be-recognized image.

[0074] In yet some implementations, the captured image or the acquired image can be subjected to object detection to obtain a bounding box (such as rectangular box) of an object, and then the captured image is cropped on the basis of the bounding box of the object, so as to obtain the to-be-recognized image. For example, in a case where one captured image includes one object sequence, one to-be-recognized image is determined on the basis of the one captured image. For another example, in a case where one captured image includes at least two object sequences, one to-be-recognized image can be determined on the basis of the one captured image, the one to-be-recognized image includes at least two object sequences, or, at least two to-be-recognized images having

one-to-one correspondence to the at least object sequences can be determined on the basis of the one captured image. In other implementations, the captured image can be cropped after at least one of the following processing, or, the captured image can be cropped and then processed by at least one of the following processing: scaling, cropping, denoising, noise addition, grayscale, rotation, and normalization, so as to obtain the to-be-recognized image.

[0075] In still some implementations, the to-be-recognized image is obtained by cropping from the captured image or the acquired image, and at least one edge of the object sequence in the to-be-recognized image may be respectively aligned with at least one edge of the to-be-recognized image. For example, one or more edges of the object sequence in the to-be-recognized image are respectively aligned with one or more edges of the to-be-recognized image.

[0076] In the embodiments of this disclosure, there may be one or at least two object sequences, the one or at least two object sequences may be formed by stacking at least one object, and each object sequence may refer to a stack of objects formed by stacking along one stacking direction. One object sequence may include objects stacked regularly or include objects stacked irregularly.

[0077] The object in the embodiments of this disclosure may be at least one of a flaky object, a blocky object, or a bagged object. The objects in the object sequence may include objects of the same shape or objects of different shapes. Any two adjacent objects in the object sequence may be in direct contact, for example, one object is placed on another object; or, any two adjacent objects in the object sequence may be adhered by another material, and another material is a glue, an adhesive or any other materials having adhesion functions.

[0078] In a case where the object is the flaky object, the flaky object is the object having a certain thickness, and the thickness direction of the object may be the stacking direction of the object.

[0079] One surface (or called as one side surface) of at least one object in the object sequence along the stacking direction has a set appearance identifier for recognizing the

category of the object. In the embodiments of this disclosure, there may be different appearance identifiers on side surfaces of different objects in the object sequence of the to-be-recognized image, for distinguishing the different objects. The appearance identifier may include at least one of size, color, pattern, texture, text on the surface or the like. The side surface of the object may be parallel to the stacking direction (or the thickness direction of the object).

[0080] The object in the object sequence may be a cylinder, a prism, a circular truncated cone, a truncated pyramid, or another regular or regular flaky object. In some implementation scenes, the object in the object sequence may be the token. The object sequence may be formed by stacking multiple tokens in the longitudinal or horizontal direction. Because tokens of different categories have different coin values or face values, and at least one of the sizes, colors, patterns, or coin value symbols of the tokens having different coin values may be different, in the embodiments of this disclosure, according to an obtained to-be-recognized image including at least one token, the category of the coin value corresponding to each token in the to-be-recognized image may be detected to obtain a coin value classification result of the token. In some embodiments, the token may include a game chip, and the coin value of the token may include the chip value of the chip.

[0081] At S202, edge detection is performed on the object sequence on the basis of the to-be-recognized image to determine an edge segmentation image of the object sequence, where the edge segmentation image includes edge information of each object forming the object sequence.

[0082] In some embodiments, the operation in which edge detection is performed on the object sequence on the basis of the to-be-recognized image to determine an edge segmentation image of the object sequence may include: the to-be-recognized image is input into an edge segmentation model (or called as an edge segmentation network), performing edge detection on the object sequence in the to-be-recognized image through the edge segmentation model, and the edge segmentation image of the object sequence is output through the edge segmentation model. The edge segmentation network may be a segmentation model for the edge of each object in the object sequence.

[0083] The edge segmentation model in the embodiments of this disclosure may be a trained edge segmentation model. For example, an initial edge segmentation model may be trained through a training sample to determine the trained edge segmentation model. The training sample may include multiple annotated images, each annotated image includes an object and annotation information for the boundary contour of each object, or each annotated image includes an object sequence and annotation information for the boundary contour of each object in the object sequence.

[0084] The edge segmentation network may include one of a Richer Convolutional Features (RCF) for edge detection network, a Holistically-nested Edge Detection (HED) network, a Canny edge detection network, evolved networks of these networks, or the like.

[0085] The pixel size of the edge segmentation image may be the same as the pixel size of the to-be-recognized image. For example, in a case where the pixel size of the to-be-recognized image is 800×600 or $800 \times 600 \times 3$, 800 is the pixel size of the to-be-recognized image in the width direction, 600 is the pixel size of the to-be-recognized image in the height direction, 3 is the number of channels of the to-be-recognized image, the channels include three channels, i.e., red, green, and blue (RGB) channels, and the pixel size of the edge segmentation image is 800×600 .

[0086] The purpose of performing edge segmentation on the to-be-recognized image is to perform binary classification on each pixel in the to-be-recognized image, and to determine whether each pixel in the to-be-recognized image is an edge pixel of the object. In a case where a certain pixel in the to-be-recognized image is an edge pixel of the object, an identifier value of the corresponding pixel in the edge segmentation image may be determined as a first value; and in a case where a certain pixel in the to-be-recognized image is not an edge pixel of the object, an identifier value of the corresponding pixel in the edge segmentation image may be determined as a second value. The first value is different from the second value. The first value may be 1 and the second value may be 0; or, the first value may be 0 and the second value may be 1. In this case, the identifier value of each pixel in the edge segmentation image is the first value or the second value, and thus the edge of each object in the object sequence of the

to-be-recognized image may be determined on the basis of the positions of the first value and the second value in the edge segmentation image. In some implementations, the edge segmentation image may be called as an edge mask.

[0087] At S203, the category of each object in the object sequence is recognized on the basis of the to-be-recognized image and the edge segmentation image.

[0088] The edge of each object in the object sequence can be obtained through the edge segmentation image, and then an area correspondingly defined by the edge of each object in the to-be-recognized image can be obtained by combining the edge segmentation image and the to-be-recognized image. Therefore, the category of each object in the object sequence can be recognized by determining the category of the object in each defined area in the to-be-recognized image.

[0089] In some implementations, image classification can be performed on each defined area in the to-be-recognized image to obtain the category of the object in each area. For example, the category of each area can be determined through a classification neural network.

[0090] In other implementations, a feature object in each area can be detected by feature object detection, and the category of each area is determined on the basis of the detected feature object. The feature object may be an object having at least one of a set shape, a set color, a set texture, a set size, or a set number.

[0091] In yet some implementations, image semantic segmentation can be performed for each defined area in the to-be-recognized image to obtain an image subjected to semantic segmentation, and the category of the object in each area is determined on the basis of the image subjected to semantic segmentation. For example, each pixel of each defined area in the to-be-recognized image can be classified through a semantic segmentation model to obtain an image subjected to semantic segmentation, and the object category of each pixel is determined on the basis of the image subjected to semantic segmentation to determine the category of each defined area. For example, the category corresponding to the maximum number of pixel points in each defined area may be determined as the category of each defined area.

[0092] In the embodiments of this disclosure, the number of objects in the object sequence of the to-be-recognized image may further be obtained through the edge segmentation image.

[0093] In a case where the objects are the tokens, the objects of different categories may indicate that the values (or face values) of the tokens are different.

[0094] In some implementations, in the case of obtaining the category of each object in the object sequence, the apparatus for recognizing stacked objects can output the category of each object in the object sequence, or, can output the identifier value corresponding to the category of each object in the object sequence. In some implementations, the identifier value corresponding to the category of each object may be the value of the object. In a case where the objects are the tokens, the category of each object may be represented by the value of the token.

[0095] For example, the category of each object or the identifier value corresponding to the category of each object can be output to a management system for display by the management system. For another example, the category of each object or the identifier value corresponding to the category of each object can be output to a behavior analysis means in the device for recognizing stacked objects, so that the behavior analysis means can determine, on the basis of the category of each object or the identifier value corresponding to the category of each object, whether objects around the surface for object placement comply with regulations.

[0096] In some implementations, the behavior analysis means can determine increase or decrease in the number and/or the total value of tokens of each placement area. The placement area may be an area for placing tokens on the surface for object placement. For example, in a game clearing stage, in a case where it is determined that the tokens in one placement area decrease and the hand of a player appears, it is determined that the player has moved the tokens, a warning is output to the management system, and then the management system sends the warning.

[0097] In the embodiments of this disclosure, the category of each object in the object sequence is recognized on the basis of the to-be-recognized image and the edge

segmentation image, where not only the edge information of each object determined on the basis of the edge segmentation image, but also feature information of each object in the object sequence in the to-be-recognized image is considered. Therefore, the determined category of each object in the object sequence in the to-be-recognized image has high accuracy.

[0098] FIG. 3 illustrates a schematic diagram of an implementation process of another method for recognizing stacked objects provided in embodiments of this disclosure. As illustrated in FIG. 3, the method is applied to an apparatus for recognizing stacked objects. The method includes the following operations.

[0099] At S301, a to-be-recognized image is obtained, the to-be-recognized image including an object sequence formed by stacking at least one object.

[00100] At S302, edge detection is performed on the object sequence on the basis of the to-be-recognized image to determine an edge segmentation image of the object sequence.

[00101] At S303, each object in the object sequence of the to-be-recognized image is segmented on the basis of the edge segmentation image to obtain a sub-image corresponding to each object.

[00102] In some embodiments, S303 may be implemented in the following manner: first position information of each object in the to-be-recognized image is determined on the basis of the edge information of each object forming the object sequence; and each object in the to-be-recognized image is segmented on the basis of the first position information to obtain each sub-image.

[00103] In some implementations, the first position information may be determined on the basis of the contour of the edge segmentation image. For example, the number of objects in the object sequence may further be determined on the basis of the edge segmentation image or on the basis of the contour or boundary position of the edge segmentation image, and then the first position information of each object in the object sequence in the edge segmentation image or the to-be-recognized image is determined on the basis of the number of objects in the object sequence.

- [00104] After the number of objects in the object sequence is obtained, the number of objects in the object sequence can be output. For example, the number of objects in the object sequence can be output to the management system or the analysis means for display by the management system, or so that the analysis means determines, on the basis of the number of objects in the object sequence, whether the behavior of the object complies with regulations.
- [00105] In some implementations, no matter the sizes of objects of different categories are the same or different, the contour or boundary position of each object in the object sequence can be determined on the basis of the edge segmentation image, and number information of the objects in the object sequence can be determined on the basis of the contour or boundary position of each object.
- [00106] In some other implementations, in a case where the sizes of objects of different categories are the same, the total height of the object sequence and the width of any object can be determined on the basis of the edge segmentation image. Because the ratio of the height to the width of one object is fixed, the number information of objects in the object sequence can be determined on the basis of the total height of the object sequence and the width of any object.
- [00107] In yet some other implementations, in a case where the sizes of objects of different categories are the same, the total height and a photographing parameter of the object sequence can be determined on the basis of the edge segmentation image. Because the height of each object in the object sequence under the same photographing parameter is fixed, the number information of objects in the object sequence can be determined on the basis of the total height and the photographing parameter of the object sequence.
- [00108] In a case where the to-be-recognized image is one image frame, one edge segmentation image frame can be obtained on the basis of one to-be-recognized image frame, and the number information of objects in the object sequence is determined on the basis of the one edge segmentation image frame.
- [00109] In a case where the to-be-recognized image includes at least two image frames, at least two to-be-recognized image frames can be obtained on the basis of at least two

captured image frames. The at least two captured image frames can be obtained by photographing the object sequence from different angles at the same time point; corresponding at least two edge segmentation image frames can be obtained on the basis of the at least two to-be-recognized image frames; and the number information of objects in the object sequence can be determined on the basis of the at least two edge segmentation image frames. In other embodiments, the at least two to-be-recognized image frames can be spliced to obtain a spliced image, and the number information of objects in the object sequence is determined on the basis of the edge segmentation image corresponding to the spliced image.

[00110] The first position information may be one-dimensional coordinate information or two-dimensional coordinate information. In some implementations, the first position information of each object in the edge segmentation image or the to-be-recognized image may include: start position information and/or end position information of the edge of each object in the stacking direction in the edge segmentation image or the to-be-recognized image. In other implementations, the first position information of each object in the edge segmentation image or the to-be-recognized image may include: start position information and end position information of the edge of each object in the stacking direction, and start position information and end position information of the edge of each object in a direction perpendicular to the stacking direction in the edge segmentation image or the to-be-recognized image.

[00111] For example, the width direction of the edge segmentation image may serve as x-axis, the height direction of the edge segmentation image may serve as y-axis, the stacking direction may be the direction of y-axis, and the start position information and the end position information of the edge of each object in the stacking direction may be coordinate information on y-axis or may be coordinate information on x-axis or y-axis. In other implementations, the first position information of each object in the edge segmentation image or the to-be-recognized image may include: position information of the edge of each object or key points on the edge of each object in the edge segmentation image or the to-be-recognized image.

[00112] Each sub-image may be rectangular. In some implementations, each image obtained by segmenting each object is rectangular, and each rectangular image is then determined as each sub-image. In other implementations, each image obtained by segmenting each object is non-rectangular (for example, at least one edge is arc-shaped), and then pixel filling and/or stretching processing is performed on each non-rectangular image to obtain each rectangular sub-image.

[00113] In some implementations, the image sizes of any two sub-images may be the same. In other implementations, the image sizes of multiple different sub-images may be different. The input size of the sub-image may be the same as the size available for input to a network. In a case where the size of a certain sub-image is different from the size available for input to the network, size transform processing can be performed on the sub-image, so that the size of the sub-image subjected to size transform processing may be the same as the size available for input to the network.

[00114] In this manner, the first position information of each object in the to-be-recognized image is determined on the basis of the edge information of each object forming the object sequence, and then each object in the object sequence of the to-be-recognized image is segmented on the basis of the first position information to obtain each sub-image. Therefore, position information of each object in the object sequence in the edge segmentation image can be located accurately on the basis of the first position information, and then the object sequence of the to-be-recognized image is segmented on the basis of the first position information to obtain each sub-image, so as to accurately match each object in the object sequence to accurately determine the category of each object in the object sequence.

[00115] At S304, category recognition is performed on each sub-image to obtain the category of each object.

[00116] In some implementations, S304 may be implemented in the following manner: performing category recognition on each sub-image to obtain at least two categories and at least two confidences having one-to-one correspondence to the at least two categories; and in a case where the difference between the highest confidence and the second highest confidence in the at least two confidences is greater than a threshold, determining the

category corresponding to the highest confidence as the category of the object corresponding to each sub-image.

[00117] For example, category recognition may be performed on each sub-image through a classification network (or called as a classification model). The classification network may include a CNN. The CCN may include at least one of: AlexNet, GoogLeNet, ResNet, LeNet, a Visual Geometry Group (VGG), a Generative Adversarial Networks (GAN), a Region-CNN (R-CNN) or the like.

[00118] For another example, feature detection may be performed on each sub-image through a detection network, and the category of each sub-image is determined on the basis of the detected feature. Taking an example, for one sub-image, in a case where the detection network detects a feature having a first category, a first ratio of pixel points corresponding to the feature of the first category to all pixel points of the sub-image is determined, and a first confidence is determined on the basis of the first ratio; and in a case where the detection network detects a feature having a second category, a second ratio of pixel points corresponding to the feature of the second category to all pixel points of the sub-image is determined, and a second confidence is determined on the basis of the second ratio until all categories in the sub-image and the confidence corresponding to each category are determined. In some implementations, the first confidence may be the highest confidence, and the second confidence may be the second highest confidence. In some implementation scenes, the feature having the first category may be at least one pink rectangle, and the feature of the second category may be at least one brown rectangle.

[00119] In some implementations, the apparatus for recognizing stacked objects may include one target network (the target network in the embodiments of this disclosure may be the classification network or the detection network), and the one target network may process each sub-image in sequence to obtain at least two categories of each sub-image and corresponding at least two confidences. In other embodiments, the apparatus for recognizing stacked objects may include at least two target networks, and the at least two target networks may process, in parallel, sub-images obtained by segmentation to obtain at least two categories of each sub-image and corresponding at least two confidences.

The at least two target networks may be the same (including the same network structure and the same network parameter). In this case, the at least two target networks implement parallel processing on the sub-image obtained by segmentation, so that the speed of determining the category of each object in the object sequence can be greatly increased. For example, in a case where the apparatus for recognizing stacked objects may include 5 target networks, each object in the to-be-recognized image is segmented to obtain 10 sub-images, so that the 5 target networks can process the first 5 sub-images in parallel and then process the last 5 sub-images.

[00120] In this case, in a case where the difference between the highest confidence and the second highest confidence is greater than the threshold, the category corresponding to the highest confidence is determined as the category of the object corresponding to the sub-image. Therefore, the category of each object in the object sequence can be determined accurately.

[00121] In the embodiments of this disclosure, each object in the object sequence of the to-be-recognized image is segmented on the basis of the edge segmentation image to obtain the sub-image corresponding to each object, and category recognition is performed on each sub-image to obtain the category corresponding to each sub-image. Therefore, the category of each object in the object sequence can be determined accurately on the basis of the category corresponding to each sub-image.

[00122] In some implementations, the method may further include: in a case where the difference between the highest confidence and the second highest confidence in the at least two confidences is less than or equal to the threshold, the category corresponding to the highest confidence and the category corresponding to the second highest confidence are output. For example, the apparatus for recognizing stacked objects may output the category corresponding to the highest confidence and the category corresponding to the second highest confidence to the management system, so that a game administrator can select a correct category through the management system. For example, the management system can output the position of at least one object in the object sequence as well as the category corresponding to the highest confidence of each object, the category corresponding to the second highest confidence and other categories, so that the game

administrator selects the correct category of each object and then the management system can output the correct category to the apparatus for recognizing stacked objects. In some implementations, the apparatus for recognizing stacked objects may train a target model on the basis of the correct category.

[00123] In this manner, the game controller can assist the apparatus for recognizing stacked objects in implementing category determination on each object in the object sequence, so as to improve the accuracy of the determined category of each object in the object sequence.

[00124] In some embodiments, after the at least two categories and the at least two confidences are obtained respectively corresponding to the at least two categories by performing category recognition on each sub-image, and the difference between the highest confidence and the second highest confidence is determined, the method for recognizing stacked objects may further include the following operation: in a case where the difference is less than or equal to the threshold and the category corresponding to the highest confidence is the same as the categories of sub-images corresponding to two second objects adjacent to a first object corresponding to the sub-image in the object sequence, the category corresponding to the second highest confidence is determined as the category of the first object corresponding to the sub-image.

[00125] One or two sub-images adjacent to each sub-image are determined on the basis of the position of each sub-image in the object sequence. For example, in a case where the object corresponding to a certain sub-image is located at an end of the object sequence (for example, the object corresponding to the certain sub-image is the one at the bottom or on the top of the object sequence), it is determined that the sub-image has one adjacent sub-image. In other cases, it is determined that the sub-image has two adjacent sub-images, and thus the categories of one or two sub-images adjacent to the sub-image can be determined.

[00126] In some other embodiments, after obtaining the at least two categories and the at least two confidences respectively corresponding to the at least two categories by performing category recognition on each sub-image, and determining the difference between the highest confidence and the second highest confidence, the method for

recognizing stacked objects may further include the following step: in a case where the difference is less than or equal to the threshold and the category corresponding to the second highest confidence is the same as the categories of sub-images corresponding to two second objects adjacent to a first object corresponding to the sub-image in the object sequence, determining the category corresponding to the highest confidence as the category of the first object corresponding to the sub-image.

[00127] In yet some other embodiments, after obtaining the at least two categories and the at least two confidences respectively corresponding to the at least two categories by performing category recognition on each sub-image, and determining the difference between the highest confidence and the second highest confidence, the method for recognizing stacked objects may further include the following step: in a case where the difference is less than or equal to the threshold and the category corresponding to the second highest confidence is different from the categories of sub-images corresponding to two second objects adjacent to a first object corresponding to the sub-image in the object sequence, determining the category corresponding to the highest confidence as the category of the first object corresponding to the sub-image.

[00128] In the embodiments of this disclosure, in a case where the difference is less than or equal to the threshold, if the category corresponding to the highest confidence determined on the basis of each sub-image is the same as the categories of two sub-images adjacent to each sub-image, the category corresponding to the second highest confidence is determined as the category of the object corresponding to each sub-image; if the category corresponding to the second highest confidence determined on the basis of each sub-image is the same as the categories of two sub-images adjacent to each sub-image, the category corresponding to the highest confidence is determined as the category of the object corresponding to each sub-image, so as to eliminate the effect of the sub-images adjacent to each sub-image on determination of the category of each sub-image; and if the category corresponding to the second highest confidence is different from the categories of two sub-images adjacent to each sub-image, the category corresponding to the highest confidence is determined as the category of the object corresponding to each sub-image. Therefore, the category of the object corresponding to each sub-image can be determined accurately.

[00129] In some embodiments, the edge segmentation image includes a mask image representing the edge information of each object, and/or, the edge segmentation image has the same size as the to-be-recognized image.

[00130] In the embodiments of this disclosure, the edge segmentation image and the to-be-recognized image have the same size, for example, the pixel sizes of the edge segmentation image and the to-be-recognized image may be the same. That is, the edge segmentation image has the same number of pixel points in the width direction and the height direction as the to-be-recognized image.

[00131] In this case, the edge segmentation image includes the mask image representing the edge information of each object, so that the edge information of each object can be determined easily on the basis of the mask image; and the edge segmentation image has the same size as the to-be-recognized image, so that the edge position of each object in the to-be-recognized image can be determined accurately on the basis of the edge position of each object in the edge segmentation image.

[00132] In some embodiments, the edge segmentation image is a binarized mask image, pixels of a first pixel value in the edge segmentation image correspond to pixels of the edge of each object in the to-be-recognized image, and pixels of a second pixel value in the edge segmentation image correspond to pixels of the non-edge part of each object in the to-be-recognized image.

[00133] The pixel size of the edge segmentation image may be $N \times M$. That is, the edge segmentation image may include $N \times M$ pixel points, and the pixel value of each pixel point in the $N \times M$ pixel points is the first pixel value or the second pixel value. For example, in a case where the first pixel value is 0 and the second pixel value is 1, pixels having the pixel value of 0 are pixels on the edge of each object, and pixels having the pixel value of 1 are pixels on the non-edge part of each object. The pixels on the non-edge part of each object may include pixels not located on the edge of each object in the object sequence, and may further include background pixels of the object sequence.

[00134] In this case, the edge segmentation image is the binarized mask image, so that whether each pixel point in the binarized mask image is located on the edge of each

object in the object sequence can be determined depending on whether each pixel point is the first pixel value or the second pixel value. Therefore, the edge of each object in the object sequence can be determined easily.

[00135] FIG. 4 illustrates a schematic diagram of an implementation process of another method for recognizing stacked objects provided in embodiments of this disclosure. As illustrated in FIG. 4, the method is applied to an apparatus for recognizing stacked objects. The method includes the following steps.

[00136] At S401, a to-be-recognized image is obtained, the to-be-recognized image including an object sequence formed by stacking at least one object.

[00137] At S402, the to-be-recognized image is input into a trained edge detection model to obtain an edge detection result for each object in the object sequence.

[00138] The edge detection model is obtained by training based on a sequence object image including object edge annotation information.

[00139] The edge detection result includes the result of whether each pixel in the to-be-recognized image is the edge pixel of the object.

[00140] At S403, the edge segmentation image of the object sequence is generated according to the edge detection result.

[00141] The pixel value of each pixel point in the edge segmentation image may be the first pixel value or the second pixel value. In a case where the pixel value of a certain pixel point is the first pixel value, it indicates that the pixel point is an edge pixel point of the object; and in a case where the pixel value of a certain pixel point is the second pixel value, it indicates that the pixel point is a non-object edge point. The non-object edge point may be a point inside the object or a point on the background of the object sequence.

[00142] At S404, the category of each object in the object sequence is recognized on the basis of the to-be-recognized image and the edge segmentation image.

[00143] In some embodiments, S404 may be implemented in the following manner: segmenting each object in the object sequence of the to-be-recognized image on the basis of the edge segmentation image to obtain a sub-image corresponding to each object; and performing category recognition on each sub-image to obtain the category of each object.

[00144] The performing category recognition on each sub-image to obtain the category of each object includes: inputting each sub-image into a trained object classification model to obtain the category of each corresponding object; and the object classification model is obtained by training based on single-object images, and the single-object images are obtained after segmenting the sequence object image according to the edge detection result for each object.

[00145] The single-object image may include one object. In some embodiments, the single-object image may only have one object. In other embodiments, the single-object image not only may have one object, but also may have some backgrounds and/or shields.

[00146] The trained object classification model may be obtained by training an initial object classification model through multiple single-object images.

[00147] The single-object image may be determined in the following manner: acquiring a captured image of the object sequence, and segmenting each object in the object sequence in the captured image to determine the single-object image corresponding to each object. In a case where the images obtained after segmentation have different sizes as that of the image available for input to the initial object classification model, size transform can be performed on the images obtained after segmentation to obtain the single-object images having the same size as the image available for input to the initial classification model. That is, the single-object image has the same size as the image available for input to the initial object classification model.

[00148] In some embodiments, the method for segmenting each object in the object sequence in the captured image may include: displaying the captured image using any display device capable of displaying, and performing an operation by a user for the displayed captured image to segment each object in the object sequence. In the

embodiments of this disclosure, the edge detection result for each object may be a detection result determined manually.

[00149] In some embodiments, in a case where the captured image is acquired, the captured image can be processed by at least one of: scaling, cropping, denoising, noise addition, grayscale, rotation, and normalization, so as to obtain a processed image; and then each object in the processed image is segmented.

[00150] In this manner, the category of each object in the object sequence is determined on the basis of the trained object classification model, and the trained object classification model is obtained by training based on the single-object images. Therefore, the category of each object in the object sequence can be determined easily and accurately through the trained object classification model.

[00151] In the embodiments of this disclosure, the edge detection result for each object in the object sequence in the to-be-recognized image is determined on the basis of the trained edge detection model, and the trained edge detection model is obtained by training based on the sequence object image including object edge annotation information. Therefore, the edge detection result for each object in the object sequence can be determined easily and accurately through the trained edge detection model.

[00152] The embodiments of this disclosure may further provide a method for recognizing stacked objects. The method includes: obtaining a to-be-recognized image, the to-be-recognized image including an object sequence formed by stacking at least one object, and the to-be-recognized image including at least two image frames; determining an edge segmentation image of the object sequence in each image on the basis of the at least two image frames; determining the number of objects in each edge segmentation image in the at least two edge segmentation images respectively corresponding to the at least two image frames; determining the number of objects in the edge segmentation image corresponding to the maximum number as the number of objects in the object sequence; and recognizing the category of each object in the object sequence on the basis of the image corresponding to the maximum number and the edge segmentation image corresponding to the maximum number.

[00153] In some implementations, each image frame in the at least two image frames may be obtained by photographing the object sequence. In some implementations, each edge segmentation image corresponding to each image frame in the at least two image frames can be determined. For example, in a case where the at least two image frames include an image A and an image B, an edge segmentation image A can be determined on the basis of the image A, and an edge segmentation image B can be determined on the basis of the image B.

[00154] In a case where at least two edge segmentation image frames include an edge segmentation image A and an edge segmentation image B, if the number of objects in the object sequence determined on the basis of the edge segmentation image A is 10, and the number of objects in the object sequence determined on the basis of the edge segmentation image B is 20, 20 is determined as the number of objects in the object sequence.

[00155] In these embodiments, the number of objects in the edge segmentation image corresponding to the maximum number is determined as the number of objects in the object sequence, so that as many objects in the object sequence can be determined as possible, thereby improving the accuracy of the determined number of objects in the object sequence. Moreover, the category of each object in the object sequence is recognized on the basis of the image corresponding to the maximum number and the edge segmentation image corresponding to the maximum number, so that the categories of as many objects in the object sequence can be recognized as possible, thereby improving the accuracy of the determined category of each object in the object sequence.

[00156] This disclosure may further provide a method for recognizing stacked objects. The method includes: obtaining a to-be-recognized image, the to-be-recognized image including an object sequence formed by stacking at least one object, and the to-be-recognized image including at least two image frames; determining an edge segmentation image of the object sequence in each image on the basis of the at least two image frames; and synthesizing at least two edge segmentation image frames respectively corresponding to the at least two image frames to obtain a synthesized image, determining the number of objects in the object sequence on the basis of the synthesized

image, and outputting the number of objects in the object sequence. In some embodiments, at least two image frames may further be spliced to obtain a spliced image; and the category of each object in the object sequence is recognized on the basis of the spliced image and the synthesized image.

[00157] This disclosure may further provide a method for recognizing stacked objects. The method includes: obtaining a to-be-recognized image, the to-be-recognized image including an object sequence formed by stacking at least one object, and the to-be-recognized image including at least two image frames; splicing the at least two image frames to obtain a spliced image, and determining an edge segmentation image of the object sequence in the spliced image; and determining the number of objects in the object sequence on the basis of the edge segmentation image of the object sequence in the spliced image, and outputting the number of objects in the object sequence. In some embodiments, the category of each object in the object sequence may further be recognized on the basis of the spliced image and the edge segmentation image of the object sequence in the spliced image.

[00158] For example, in a case where the at least two image frames include an image A and an image B, if objects at the upper part of the image A are shielded and objects at the lower part of the image B are shielded, the lower part of the image A and the upper part of the image B can be spliced to obtain the spliced image where no object is shielded.

[00159] The spliced image obtained by splicing at least two image frames is acquired, and the category of each object in the object sequence is recognized on the basis of the spliced image and the synthesized image.

[00160] The approach of recognizing the category of each object in the object sequence on the basis of the spliced image and the synthesized image or the approach of recognizing the category of each object in the object sequence on the basis of the spliced image and the edge segmentation image of the object sequence in the spliced image may be similar to the approach of recognizing the category of each object in the object sequence on the basis of the to-be-recognized image and the edge segmentation image in S203.

- [00161] FIG. 5 illustrates a schematic diagram of an implementation process of still another method for recognizing stacked objects provided in embodiments of this disclosure. As illustrated in FIG. 5, the method is applied to an apparatus for recognizing stacked objects. The method includes the following steps.
- [00162] At S501, a to-be-recognized image is obtained, the to-be-recognized image including an object sequence formed by stacking at least one object.
- [00163] At S502, edge detection is performed on the object sequence on the basis of the to-be-recognized image to determine an edge segmentation image of the object sequence.
- [00164] At S503, the category of each object in the object sequence is recognized on the basis of the to-be-recognized image and the edge segmentation image.
- [00165] In some embodiments, the object has a value attribute corresponding to the category. Different categories may have the same value attribute or different value attributes.
- [00166] At S504, the total value of objects in the object sequence is determined on the basis of the category of each object and the corresponding value attribute.
- [00167] The apparatus for recognizing stacked objects may be configured with a mapping relationship between the category of an object and the value of the object, so that the value attribute of each object can be determined on the basis of the mapping relationship and the category of each object.
- [00168] In a case where the objects include tokens, the determined value of each object may be the face value of the token.
- [00169] The obtained values of the objects may be added to obtain the total value of the objects in the object sequence.
- [00170] In some implementations, a surface for object placement may include multiple placement areas, and objects may be placed on at least one placement area in the multiple placement areas, so that the category of each object in an object sequence placed in each

placement area can be determined on the basis of the to-be-recognized image. One or more object sequences may be placed on one placement area. For example, the category of each object in the object sequence in each placement area may be determined on the basis of the edge segmentation image and a semantic segmentation image.

[00171] When the category of each object in the object sequence in each placement area is obtained, the value attribute of each object in the object sequence in each placement area can be determined, and then the total value of objects in each placement area is determined on the basis of the value attribute of each object in the object sequence in each placement area.

[00172] In some implementations, whether an action of a game participant complies with regulations can be determined by combining change in the total value of objects in each placement area with the action of the game participant.

[00173] In a case where the total value of objects in each placement area is obtained, the total value of objects in each placement area can be output to the management system for display by the management system. For another example, the total value of objects in each placement area may be output to a behavior analysis means in the device for recognizing stacked objects, so that the behavior analysis means can determine, on the basis of the change in the total value of objects in each placement area, whether objects around the surface for object placement comply with regulations.

[00174] In the embodiments of this disclosure, the total value of objects in the object sequence is determined on the basis of the category of each object and the corresponding value attribute. Therefore, it is convenient to count the total value of stacked objects, for example, it is convenient to detect and determine the total value of stacked tokens.

[00175] FIG. 6 illustrates a schematic diagram of a process framework of a method for recognizing stacked objects provided in embodiments of this disclosure. As illustrated in FIG. 6, a to-be-recognized image may be an image 61 or may include an image 61, the to-be-recognized image is input into an edge detection network to obtain an edge segmentation image, and the edge segmentation image may be an image 62 or may include an image 62.

[00176] The contour of each object in the object sequence can be determined on the basis of the image 62, so that the number of objects in the object sequence and the start position and/or end position of each object in the object sequence on y-axis of the image 62 can be determined. In some implementations, the start position and the end position of each object in the object sequence on x-axis of the image 62 can further be obtained.

[00177] The image 61 can be segmented on the basis of the start position and the end position of each object in the image 62 on y-axis of the image 62 or on the basis of the start position and the end position of each object in the image 62 on y-axis and x-axis of the image 62, so as to obtain multiple sub-images. Each sub-image may be an image 63 or may include an image 63, the sub-image may only include an object area, or, the sub-image may not only include an object area but also include a background area. Then, each image 63 is input into a classification network, and the classification network classifies each image 63 to obtain the category of the object corresponding to each image 63.

[00178] For example, FIG. 6 illustrates that the category of the object corresponding to each image 63 includes (6, 6, 6, 5, 5, 5). That is, 3 categories having the identifier value of 6 can be determined and 3 categories having the identifier value of 5 can be recognized.

[00179] The method for recognizing stacked objects provided in the embodiments of this disclosure includes two parts, i.e., edge detection and object classification.

[00180] In edge detection, a deep learning edge detection model (corresponding to the foregoing edge detection network) is used, the input is an object sequence image (corresponding to the foregoing to-be-recognized image), and the output is an object edge image (an edge mask) (corresponding to the foregoing edge segmentation image). The size of the edge mask is consistent with the size of the input object sequence image, the edge mask is a picture having the pixel value of 0 or 1, the pixel value 1 represents that a pixel is located on the edge of an object, and if a pixel is not located on the edge, the pixel value is 0.

[00181] In object classification, the output edge mask is used to segment the object sequence image according to an edge line in an object edge image to obtain a small image (corresponding to the foregoing sub-image) of each object. An object classifier (corresponding to the foregoing classification network) is used to perform category recognition on each object. The object classifier may be a CNN classifier commonly used in deep learning; ResNet18 serves as a network infrastructure; and n categories are divided, n is greater than or equal to 1, and n is the total number of object categories. Finally, classification results are arranged according to the position of each object, and a recognition result for the object sequence in the picture can be obtained.

[00182] On the basis of the foregoing embodiments, embodiments of this disclosure provide an apparatus for recognizing stacked objects. Units included in the apparatus and modules included in the units can be implemented by a processor in the device for recognizing stacked objects, and of course, can also be implemented by a specific logic circuit.

[00183] FIG. 7 illustrates a schematic structural diagram of the composition of an apparatus for recognizing stacked objects provided in embodiments of this disclosure. As illustrated in FIG. 7, an apparatus 700 for recognizing stacked objects includes: an obtaining unit 701, configured to obtain a to-be-recognized image, the to-be-recognized image including an object sequence formed by stacking at least one object; a determination unit 702, configured to perform edge detection on the object sequence on the basis of the to-be-recognized image to determine an edge segmentation image of the object sequence, the edge segmentation image including edge information of each object forming the object sequence; and a recognition unit 703, configured to recognize the category of each object in the object sequence on the basis of the to-be-recognized image and the edge segmentation image.

[00184] In some embodiments, the recognition unit 703 is further configured to: segment each object in the object sequence of the to-be-recognized image on the basis of the edge segmentation image to obtain a sub-image corresponding to each object; and perform category recognition on each sub-image to obtain the category of each object.

- [00185] In some embodiments, the recognition unit 703 is further configured to: determine first position information of each object in the to-be-recognized image on the basis of the edge information of each object forming the object sequence; and segment each object in the to-be-recognized image on the basis of the first position information to obtain each sub-image.
- [00186] In some embodiments, the recognition unit 703 is further configured to: perform category recognition on each sub-image to obtain at least two categories and at least two confidences having one-to-one correspondence to the at least two categories; and in a case where the difference between the highest confidence and the second highest confidence in the at least two confidences is greater than a threshold, determine the category corresponding to the highest confidence as the category of the object corresponding to the sub-image.
- [00187] In some embodiments, the recognition unit 703 is further configured to implement at least one of:
- [00188] in a case where the difference is less than or equal to the threshold and the category corresponding to the highest confidence is the same as the categories of sub-images corresponding to two second objects adjacent to a first object corresponding to the sub-image in the object sequence, determining the category corresponding to the second highest confidence as the category of the first object corresponding to the sub-image;
- [00189] in a case where the difference is less than or equal to the threshold and the category corresponding to the second highest confidence is the same as the categories of sub-images corresponding to two second objects adjacent to a first object corresponding to the sub-image in the object sequence, determining the category corresponding to the highest confidence as the category of the first object corresponding to the sub-image; and
- [00190] in a case where the difference is less than or equal to the threshold and the category corresponding to the second highest confidence is different from the categories of sub-images corresponding to two second objects adjacent to a first object corresponding to the sub-image in the object sequence, determining the category

corresponding to the highest confidence as the category of the first object corresponding to the sub-image.

[00191] In some embodiments, the edge segmentation image includes a mask image representing the edge information of each object, and/or, the edge segmentation image has the same size as the to-be-recognized image.

[00192] In some embodiments, the edge segmentation image is a binarized mask image, pixels of a first pixel value in the edge segmentation image correspond to pixels of the edge of each object in the to-be-recognized image, and pixels of a second pixel value in the edge segmentation image correspond to pixels of the non-edge part of each object in the to-be-recognized image.

[00193] In some embodiments, the determination unit 702 is further configured to: input the to-be-recognized image into a trained edge detection model to obtain an edge detection result for each object in the object sequence, the edge detection model being obtained by training based on a sequence object image including object edge annotation information; and generate the edge segmentation image of the object sequence according to the edge detection result.

[00194] In some embodiments, the recognition unit 703 is further configured to: input each sub-image into a trained object classification model to obtain the category of each corresponding object; and the object classification model is obtained by training based on single-object images, and the single-object images are obtained after segmenting the sequence object image according to the edge detection result for each object.

[00195] In some embodiments, the object has a value attribute corresponding to the category; and the determination unit 702 is further configured to: determine the total value of objects in the object sequence on the basis of the category of each object and the corresponding value attribute.

[00196] The description of the foregoing apparatus embodiments is similar to the description of the foregoing method embodiments, and has beneficial effects similar to those of the method embodiments. For technical details not recited in the apparatus

embodiments of this disclosure, please refer to the description of the method embodiments of this disclosure for understanding.

[00197] It is to be noted that, in the embodiments of this disclosure, when being implemented in form of software functional modules and sold or used as an independent product, the foregoing method for recognizing stacked objects may also be stored in a computer storage medium. Based on such an understanding, the technical solutions of the embodiments of this disclosure substantially or parts making contributions to the conventional art may be embodied in form of a software product, and the computer software product is stored in a storage medium, including a plurality of instructions configured to enable one device for recognizing stacked objects to execute all or part of the method in each embodiment of this disclosure.

[00198] FIG. 8 illustrates a schematic diagram of a hardware entity of a device for recognizing stacked objects provided in embodiments of this disclosure. As illustrated in FIG. 8, a hardware entity of a device 800 for recognizing stacked objects includes a processor 801 and a memory 802. The memory 802 stores a computer program capable of running on the processor 801, and the processor 801 executes the program to implement the steps of the method in any one of the foregoing embodiments.

[00199] The memory 802 stores the computer program capable of running on the processor, is configured to store instructions and applications executable to the processor 801, can also cache data (such as image data, video data, voice communication data, and video communication data) to be processed or processed by the processor 801 and modules in the device 800 for recognizing stacked objects, and can be implemented as a FLASH or a Random Access Memory (RAM).

[00200] When the processor 801 executes the program, the steps of any one of the foregoing methods for recognizing stacked objects are implemented. The processor 801 generally controls the overall operation of the device 800 for recognizing stacked objects.

[00201] Embodiments of this disclosure provide a computer storage medium. The computer storage medium stores one or more programs, and the one or more programs

may be executed by one or more processors so as to implement the steps of the method for recognizing stacked objects in any one of the foregoing embodiments.

[00202] It is to be noted here that: the description of the foregoing storage medium and device embodiments is similar to the description of the foregoing method embodiments, and has beneficial effects similar to those of the method embodiments. For technical details not recited in the storage medium and device embodiments of this disclosure, please refer to the description of the method embodiments of this disclosure for understanding.

[00203] The foregoing apparatus for recognizing stacked objects, a chip, or a processor may include any one or a combination of more than one of: an Application Specific Integrated Circuit (ASIC), a Digital Signal Processor (DSP), a Digital Signal Processing Device (DSPD), a Programmable Logic Device (PLD), a Field Programmable Gate Array (FPGA), a Central Processing Unit (CPU), a Graphics Processing Unit (GPU), an embedded Neural-network Processing Unit (NPU), a controller, a micro-controller, or a microprocessor. It can be understood that an electronic device for implementing the function of the foregoing processor may also be another device, and no specific limitation is made thereto in the embodiments of this disclosure. The foregoing apparatus for recognizing stacked objects, the chip, or the processor may implement or execute the methods, the steps, and the logic block diagrams recited in the embodiments of this disclosure. The steps of the methods recited with reference to the embodiments of this disclosure may be directly implemented by a hardware decoding processor, or implemented by a combination of hardware and software modules in a decoding processor. The software module may be located in a mature storage medium in the art, such as the RAM, a flash memory, a Read Only Memory (ROM), a Programmable Read-Only Memory (PROM), an electrically erasable programmable memory, a register, or the like. The storage medium is located in the memory, and the processor reads information in the memory and completes the steps in the foregoing methods in combination with hardware of the processor.

[00204] The foregoing computer storage medium/memory may be the ROM, the PROM, an Erasable Programmable Read-Only Memory (EPROM), an Electrically

Erasable Programmable Read-Only Memory (EEPROM), a Ferromagnetic Random Access Memory (FRAM), a flash memory, a magnetic surface memory, an optical disk, or a Compact Disc Read-Only Memory (CD-ROM), and may also be various terminals including any one or any combination of the foregoing memories, such as a mobile phone, a computer, a pad device, and a personal digital assistant.

[00205] It should be understood that the expression "one embodiment" or "an embodiment" or "embodiments of this disclosure" or "the foregoing embodiments" or "some implementations" or "some embodiments" mentioned throughout the description means that a particular feature, structure, or characteristic described in connection with a certain embodiment is included in at least one embodiment of the disclosure. Thus, the phrase "in one embodiment" or "in an embodiment" or "embodiments of this disclosure" or "the foregoing embodiments" or "some implementations" or "some embodiments" in various places throughout the description is not necessarily to refer to the same embodiment. Furthermore, the particular features, structures, or characteristics may be combined in one or more embodiments in any proper manner. It should be understood that, in the embodiments of this disclosure, the size of a serial number of the steps of each process does not mean a sequential order to execute. The execution order of each process should be determined in terms of its function and internal logic, and should not be intended to limit an implementation process of the embodiments of this disclosure. The serial numbers of the foregoing embodiments of this disclosure are only for description, and do not represent any advantages and disadvantages of the embodiments.

[00206] Unless otherwise specified, any step in the embodiments of this disclosure is executed by the apparatus for recognizing stacked objects or may be executed by the processor of the apparatus for recognizing stacked objects. Unless otherwise specified, the embodiments of this disclosure do not limit the sequence in which the apparatus for recognizing stacked objects performs the steps. In addition, the methods used for processing data in different embodiments may be the same or different. It is to be further noted that any step in the embodiments of this disclosure can be executed independently by the apparatus for recognizing stacked objects. That is, when executing any step of the foregoing embodiments, the apparatus for recognizing stacked objects does not rely on the execution of other steps.

[00207] In the description of this disclosure, it should be understood that the terms "central", "transversal", "longitudinal", "length", "width", "thickness", "upper", "lower", "front", "rear", "left", "right", "vertical", "horizontal", "top", "bottom", "inner", "outer", "clockwise", "anticlockwise" and the like indicate orientations and position relationships which are based on the illustrations in the accompanying drawings, and these terms are merely for ease and brevity of the description, instead of indicating or implying that the apparatus or elements shall have a particular orientation and shall be structured and operated based on the particular orientation. Therefore, the terms cannot be understood as limitations to this disclosure. In addition, the expressions "first" and "second" are merely intended for description, and cannot be understood as indicating or implying relative importance or implicitly specifying the number of indicated technical features. Therefore, The features defined with "first" and "second" may explicitly or implicitly include one or more of the features. In the description of this disclosure, the "multiple" means two or more than two, unless specifically defined otherwise.

[00208] In this disclosure, unless specified or limited otherwise, a structure in which a first feature is "on" or "below" a second feature may include a structure in which the first feature is in direct contact with the second feature, and may also include a structure in which the first feature and the second feature are not in direct contact with each other, but are contacted via an additional feature formed therebetween. Furthermore, a first feature "on", "above", or "on top of" a second feature may include a structure in which the first feature is right or obliquely "on", "above", or "on top of" the second feature, or just means that the first feature is at a height higher than that of the second feature; while a first feature "below", "under", or "on bottom of" a second feature may include a structure in which the first feature is right or obliquely "below", "under", or "on bottom of" the second feature, or just means that the first feature is at a height lower than that of the second feature.

[00209] It should be understood that the recited device and method in the embodiments provided in this disclosure may be implemented in other manners. The device embodiments described above are merely exemplary. For example, the unit division is merely logical function division and may be actually implemented in other division manners. For example, a plurality of units or components may be combined or integrated

into another system, or some features may be ignored or not performed. In addition, the displayed or discussed mutual couplings or direct couplings or communicational connections among the components may be implemented by means of some interfaces. The indirect couplings or communicational connections between the devices or units may be electrical, mechanical, or in other forms.

[00210] The units described as separate components may or may not be physically separate, and the components displayed as units may or may not be physical units both, may be located at one position, or may be distributed on a plurality of network units. Some or all of the units may be selected according to actual needs to achieve the objectives of the solutions of the embodiments.

[00211] In addition, the functional units in the embodiments of this disclosure may be all integrated into one processing unit, or each of the units may separately serve as an independent unit, or two or more units are integrated into one unit, and the integrated unit may be implemented in the form of hardware, or may also be implemented in the form of a hardware and software functional unit.

[00212] The methods recited in the method embodiments provided in this disclosure can be arbitrarily combined without causing conflicts so as to obtain new method embodiments.

[00213] The features recited in product embodiments provided in this disclosure can be arbitrarily combined without causing conflicts so as to obtain new product embodiments.

[00214] The features recited in the method or device embodiments provided in this disclosure can be arbitrarily combined without causing conflicts so as to obtain new method or device embodiments.

[00215] A person of ordinary skill in the art may understand that all or some steps for implementing the foregoing method embodiments may be completed by a program by instructing related hardware; the foregoing program may be stored in a computer storage medium; when the program is executed, steps including the foregoing method embodiments are performed. Moreover, the foregoing storage medium includes various

media capable of storing program code, such as a mobile storage device, the ROM, the magnetic disk, or the optical disk.

[00216] Or when the foregoing integrated unit of this disclosure is implemented in the form of a software functional module and sold or used as an independent product, the integrated unit may be stored in one computer storage medium. Based on such an understanding, the technical solutions of the embodiments of this disclosure substantially or parts making contributions to the conventional art may be embodied in form of a software product, and the computer software product is stored in a storage medium, including a plurality of instructions configured to enable one computer device (which may be a personal computer, a server, or a network device) to execute all or part of the method in each embodiment of this disclosure. Moreover, the foregoing storage medium includes various media capable of storing program code, such as the mobile storage device, the ROM, the magnetic disk, or the optical disk.

[00217] In the embodiments of this disclosure, descriptions of the same steps and the same content in different embodiments may be referred to each other. In the embodiments of this disclosure, the term "and" does not cause any effect on the sequence of the steps. For example, the description that the apparatus for recognizing stacked objects executes A and executes B may indicate that the apparatus for recognizing stacked objects executes A before executing B, or the apparatus for recognizing stacked objects executes B before executing B, or the apparatus for recognizing stacked objects executes A and B at the same time.

[00218] It should be noted that the accompany drawings in the embodiments of this disclosure are only to illustrate the schematic position of each device on the apparatus for recognizing stacked objects, and do not represent the real position in the apparatus for recognizing stacked objects. The real position of each device or each area may be correspondingly changed or deviated according to actual conditions (such as the structure of the apparatus for recognizing stacked objects). Moreover, the proportions of different parts in the apparatus for recognizing stacked objects in the drawings do not represent the real proportions.

[00219] The singular forms "one", "the", and "this" used in the embodiments of this disclosure and the appended claims are also intended to include multiple forms, unless other meanings are clearly represented in the context.

[00220] It should be understood that the term "and/or" used herein only describes an association relation between associated objects, indicating that three relations may exist, for example, A and/or B may indicate three conditions, i.e., A exists separately, A and B exist at the same time, and B exists separately. In addition, the character "/" in the text generally represents that the preceding and latter associated objects are in an "or" relation.

[00221] It should be noted that in the embodiments involved in this disclosure, all steps or part of the steps can be executed, as long as a complete technical solution can be formed.

[00222] The descriptions above are only implementations of this disclosure. However, the scope of protection of this disclosure is not limited thereto. Within the technical scope recited by this disclosure, any variation or substitution that can be easily conceived of by those skilled in the art should all fall within the scope of protection of this disclosure. Therefore, the scope of protection of this disclosure should be determined by the scope of protection of the claims.

CLAIMS

1. A method for recognizing stacked objects, comprising:

obtaining a to-be-recognized image, wherein the to-be-recognized image comprises an object sequence formed by stacking at least one object;

performing edge detection on the object sequence on the basis of the to-be-recognized image to determine an edge segmentation image of the object sequence, wherein the edge segmentation image comprises edge information of each object forming the object sequence; and

recognizing the category of each object in the object sequence on the basis of the to-be-recognized image and the edge segmentation image.

2. The method of claim 1, wherein the recognizing the category of each object in the object sequence on the basis of the to-be-recognized image and the edge segmentation image comprises:

segmenting each object in the object sequence of the to-be-recognized image on the basis of the edge segmentation image to obtain a sub-image corresponding to each object; and

performing category recognition on each sub-image to obtain the category of each object.

3. The method of claim 2, wherein the segmenting each object in the object sequence of the to-be-recognized image on the basis of the edge segmentation image to obtain a sub-image corresponding to each object comprises:

determining first position information of each object in the to-be-recognized image on the basis of the edge information of each object forming the object sequence; and

segmenting each object in the to-be-recognized image on the basis of the first position information to obtain each sub-image.

4. The method of claim 2 or 3, wherein the performing category recognition on each sub-image to obtain the category of each object comprises:

performing category recognition on each sub-image to obtain at least two categories and at least two confidences having one-to-one correspondence to the at least two categories; and

in a case where a difference between a highest confidence and a second highest confidence in the at least two confidences is greater than a threshold, determining the category corresponding to the highest confidence as the category of the object corresponding to the sub-image.

5. The method of claim 4, wherein the method further comprises at least one of:

in a case where the difference is less than or equal to the threshold and the category corresponding to the highest confidence is the same as the categories of sub-images corresponding to two second objects adjacent to a first object corresponding to the sub-image in the object sequence, determining the category corresponding to the second highest confidence as the category of the first object corresponding to the sub-image;

in a case where the difference is less than or equal to the threshold and the category corresponding to the second highest confidence is the same as the categories of sub-images corresponding to two second objects adjacent to a first object corresponding to the sub-image in the object sequence, determining the category corresponding to the highest confidence as the category of the first object corresponding to the sub-image; or

in a case where the difference is less than or equal to the threshold and the category corresponding to the second highest confidence is different from the categories of sub-images corresponding to two second objects adjacent to a first object corresponding to the sub-image

in the object sequence, determining the category corresponding to the highest confidence as the category of the first object corresponding to the sub-image.

6. The method of any one of claims 1 to 5, wherein the edge segmentation image comprises a mask image representing the edge information of each object, and/or, the edge segmentation image has the same size as the to-be-recognized image.

7. The method of claim 6, wherein the edge segmentation image is a binarized mask image, pixels of a first pixel value in the edge segmentation image correspond to pixels of the edge of each object in the to-be-recognized image, and pixels of a second pixel value in the edge segmentation image correspond to pixels of the non-edge part of each object in the to-be-recognized image.

8. The method of any one of claims 2 to 5, wherein the performing edge detection on the object sequence on the basis of the to-be-recognized image to determine an edge segmentation image of the object sequence comprises:

inputting the to-be-recognized image into a trained edge detection model to obtain an edge detection result for each object in the object sequence, wherein the edge detection model is obtained by training based on a sequence object image comprising object edge annotation information; and

generating the edge segmentation image of the object sequence according to the edge detection result.

9. The method of claim 8, wherein the performing category recognition on each sub-image to obtain the category of each object comprises:

inputting each sub-image into a trained object classification model to obtain the category of each corresponding object;

wherein the object classification model is obtained by training based on single-object images, and the single-object images are obtained after segmenting the sequence object image according to the edge detection result for each object.

10. The method of any one of claims 1 to 9, wherein the object has a value attribute corresponding to the category; and the method further comprises:

determining a total value of objects in the object sequence on the basis of the category of each object and the corresponding value attribute.

11. An apparatus for recognizing stacked objects, comprising:

an obtaining unit, configured to obtain a to-be-recognized image, wherein the to-be-recognized image comprises an object sequence formed by stacking at least one object;

a determination unit, configured to perform edge detection on the object sequence on the basis of the to-be-recognized image to determine an edge segmentation image of the object sequence, wherein the edge segmentation image comprises edge information of each object forming the object sequence; and

a recognition unit, configured to recognize the category of each object in the object sequence on the basis of the to-be-recognized image and the edge segmentation image.

12. A device for recognizing stacked objects, comprising a memory and a processor, wherein the memory stores a computer program capable of running on the processor, and when executing the computer program, the processor is configured to:

obtain a to-be-recognized image, wherein the to-be-recognized image comprises an object sequence formed by stacking at least one object;

perform edge detection on the object sequence on the basis of the to-be-recognized image to determine an edge segmentation image of the object sequence, wherein the edge

segmentation image comprises edge information of each object forming the object sequence;
and

recognize the category of each object in the object sequence on the basis of the to-be-recognized image and the edge segmentation image.

13. The device of claim 12, wherein when recognizing the category of each object in the object sequence on the basis of the to-be-recognized image and the edge segmentation image, the processor is configured to:

segment each object in the object sequence of the to-be-recognized image on the basis of the edge segmentation image to obtain a sub-image corresponding to each object; and

perform category recognition on each sub-image to obtain the category of each object.

14. The device of claim 13, wherein when segmenting each object in the object sequence of the to-be-recognized image on the basis of the edge segmentation image to obtain a sub-image corresponding to each object, the processor is configured to:

determine first position information of each object in the to-be-recognized image on the basis of the edge information of each object forming the object sequence; and

segment each object in the to-be-recognized image on the basis of the first position information to obtain each sub-image.

15. The device of claim 13 or 14, wherein when performing category recognition on each sub-image to obtain the category of each object, the processor is configured to:

perform category recognition on each sub-image to obtain at least two categories and at least two confidences having one-to-one correspondence to the at least two categories; and

in a case where a difference between a highest confidence and a second highest confidence in the at least two confidences is greater than a threshold, determine the category corresponding to the highest confidence as the category of the object corresponding to the sub-image.

16. The device of claim 15, wherein the processor is further configured to perform at least one of following operations:

in a case where the difference is less than or equal to the threshold and the category corresponding to the highest confidence is the same as the categories of sub-images corresponding to two second objects adjacent to a first object corresponding to the sub-image in the object sequence, determining the category corresponding to the second highest confidence as the category of the first object corresponding to the sub-image;

in a case where the difference is less than or equal to the threshold and the category corresponding to the second highest confidence is the same as the categories of sub-images corresponding to two second objects adjacent to a first object corresponding to the sub-image in the object sequence, determining the category corresponding to the highest confidence as the category of the first object corresponding to the sub-image; or

in a case where the difference is less than or equal to the threshold and the category corresponding to the second highest confidence is different from the categories of sub-images corresponding to two second objects adjacent to a first object corresponding to the sub-image in the object sequence, determining the category corresponding to the highest confidence as the category of the first object corresponding to the sub-image.

17. The device of any one of claims 12 to 16, wherein the edge segmentation image comprises a mask image representing the edge information of each object, and/or, the edge segmentation image has the same size as the to-be-recognized image.

18. The device of claim 17, wherein the edge segmentation image is a binarized mask image, pixels of a first pixel value in the edge segmentation image correspond to pixels of the edge of each object in the to-be-recognized image, and pixels of a second pixel value in the edge segmentation image correspond to pixels of the non-edge part of each object in the to-be-recognized image.

19. A computer storage medium, wherein the computer storage medium stores one or more programs, and when executed by one or more processors, the one or more programs are configured to:

 obtain a to-be-recognized image, wherein the to-be-recognized image comprises an object sequence formed by stacking at least one object;

 perform edge detection on the object sequence on the basis of the to-be-recognized image to determine an edge segmentation image of the object sequence, wherein the edge segmentation image comprises edge information of each object forming the object sequence; and

 recognize the category of each object in the object sequence on the basis of the to-be-recognized image and the edge segmentation image.

20. A computer program, comprising computer instructions executable by an electronic device, wherein when executed by a processor in the electronic device, the computer instructions are configured to:

 obtain a to-be-recognized image, wherein the to-be-recognized image comprises an object sequence formed by stacking at least one object;

 perform edge detection on the object sequence on the basis of the to-be-recognized image to determine an edge segmentation image of the object sequence, wherein the edge segmentation image comprises edge information of each object forming the object sequence; and

recognize the category of each object in the object sequence on the basis of the to-be-recognized image and the edge segmentation image.