CLAIMS

1. A method for object detection, comprising:

obtaining a plurality of to-be-determined targets in a to-be-detected image;

determining, for at least one category, confidences of a plurality of to-be-determined targets respectively;

determining categories of the plurality of to-be-determined targets according to the confidences;

respectively determining position offset values corresponding to the categories of the plurality of to-be-determined targets;

respectively using the position offset values corresponding to the categories of the plurality of to-be-determined targets as position offset values of the plurality of to-be-determined targets; and

determining a category and position information of at least one to-be-determined target in the to-be-detected image according to the categories of the plurality of to-be-determined targets, the position offset values of the plurality of to-be-determined targets, and confidences of the categories of the plurality of to-be-determined targets.

~~2. The method according to claim 1,~~ wherein the operation of obtaining a plurality of to-be-determined targets in a to-be-detected image comprises:

obtaining the plurality of to-be-determined targets formed based on at least one predetermined region size in the to-be-detected image.

~~3. The method according to claim 2,~~ wherein the operation of obtaining the plurality of to-be-determined targets formed based on at least one predetermined region size in the to-be-detected image comprises:

obtaining a feature map of the to-be-detected image;

forming a reference box of a feature point in the feature map according to reference box configuration information, wherein the reference box configuration information is preset

and a number and sizes of reference boxes are defined by the reference box configuration information;

using the reference box of the feature point as one to-be-determined target; and

obtaining, respectively corresponding to a plurality of feature points in the feature map, the plurality of to-be-determined targets.

24. The method according to claim 1 ~~3~~, wherein the operation of obtaining a feature map of the to-be-detected image comprises:

inputting the to-be-detected image into a backbone network in a convolutional neural network;

inputting a feature map output by the backbone network into a filter layer in the convolutional neural network;

filtering the feature map by the filter layer according to a preset sliding window, and

using the filtered feature map as the feature map of the to-be-detected image.

35. The method according to claim 1 ~~2~~, wherein the operation of obtaining the plurality of to-be-determined targets formed based on at least one predetermined region size in the to-be-detected image comprises:

obtaining a feature map of the to-be-detected image;

pooling the feature map based on reference box configuration information to obtain a plurality of new feature maps; and

using the plurality of new feature maps as the plurality of to-be-determined targets.

46. The method according to any one of claims 1 ~~2~~ to 3 ~~5~~, wherein the predetermined region size comprises: nine predetermined region sizes formed based on three different lengths and three different aspect ratios; or nine predetermined region sizes formed based on three different widths and three different aspect ratios; or nine predetermined region sizes formed based on three different lengths and widths.

57. The method according to any one of claims 1 to 4 ~~6~~, wherein the category comprises: two object categories and one background category.

68. The method according to any one of claims 1 to 5 ~~7~~, wherein the operation of

determining, for at least one category, confidences of a plurality of to-be-determined targets respectively, and determining categories of the plurality of to-be-determined targets according to the confidences comprises:

for one to-be-determined target, calculating, for the at least one category, a confidence of the to-be-determined target respectively, and using a category corresponding to a highest confidence as a category of the to-be-determined target.

79. The method according to any one of claims 1 to 6~~8~~, wherein the operation of determining position offset values corresponding to the respective categories of the plurality of to-be-determined targets comprises:

for one to-be-determined target, calculating, for a category of the to-be-determined target, a top offset value, a bottom offset value, a left offset value, and a right offset value of the to-be-determined target .

8~~10~~. The method according to any one of claims 1 to 7~~9~~, wherein the position information of the to-be-determined target comprises: position information of a bounding box of the to-be-determined target.

9~~11~~. The method according to claim 8~~10~~, wherein the operation of determining a category and position information of at least one to-be-determined target in the to-be-detected image according to the categories of the plurality of to-be-determined targets, the position offset values of the plurality of to-be-determined targets, and confidences of the categories of the plurality of to-be-determined targets comprises:

selecting, from the plurality of to-be-determined targets, a plurality of to-be-determined targets with confidences meeting a predetermined confidence requirement;

forming position information of bounding boxes in the to-be-detected image according to position offset values of the plurality of selected to-be-determined targets; and

using categories of the plurality of selected to-be-determined targets as categories of the corresponding bounding boxes.

10~~12~~. The method according to any one of claims 1 to 9~~11~~, wherein the operation of

determining, for at least one category, confidences of a plurality of to-be-determined targets respectively, determining categories of the plurality of to-be-determined targets according to the confidences, respectively determining position offset values corresponding to the categories of the plurality of to-be-determined targets comprises:

using a convolutional neural network to determine, for at least one category, confidences of the plurality of to-be-determined targets respectively, determine categories of the plurality of to-be-determined targets according to the confidences, and respectively determine position offset values corresponding to the categories of the plurality of to-be-determined targets; and

the method further comprises: training the convolutional neural network, wherein the operation of training the convolutional neural network comprises:

obtaining, from an image sample set, an image sample annotated with information of at least one standard position and the category of the standard position;

obtaining a plurality of to-be-determined targets in the image sample;

determining, for at least one category, confidences of the plurality of to-be-determined targets respectively by one convolutional layer in the convolutional neural network;

determining categories of the plurality of to-be-determined targets according to the confidences;

respectively determining, by another convolutional layer in the convolutional neural network, position offset values corresponding to the categories of the plurality of to-be-determined targets;

respectively using the position offset values corresponding to the categories of the plurality of to-be-determined targets as position offset values of the plurality of to-be-determined targets;

calculating standard position offset values of the plurality of to-be-determined targets with respect to the corresponding standard position;

calculating a deviation between a position offset value of at least one of the plurality of to-be-determined targets with respect to a category of the corresponding standard position

and the corresponding standard position offset value by one loss layer in the convolutional neural network, and calculating a deviation between a confidence of at least one of the plurality of to-be-determined targets with respect to the category of the corresponding standard position and the category of the standard

position by one loss layer in the convolutional neural network;

back-propagating the deviations in the convolutional neural network; and

completing the training of the convolutional neural network until a number of image samples obtained from the image sample set reaches a predetermined number and the deviations are within a predetermined range.

11~~13~~. The method according to claim 10~~12~~, wherein the operation of obtaining a plurality of to-be-determined targets in the image sample comprises:

obtaining a plurality of to-be-determined targets formed based on at least one predetermined region size in the image sample.

12~~14~~. The method according to claim 10~~12~~ or 11~~13~~, wherein the standard position comprises: a standard bounding box.

13~~15~~. The method according to claim 12~~14~~, wherein the operation of calculating standard position offset values of the plurality of to-be-determined targets with respect to the corresponding standard position comprises:

determining standard bounding boxes having largest overlap areas respectively corresponding to the plurality of to-be-determined targets; and

respectively calculating standard position offset values of the plurality of to-be-determined targets with respect to the standard bounding boxes having the largest overlap areas.

14~~16~~. The method according to claim 12~~14~~ or 13~~15~~, wherein the at least one of the plurality of to-be-determined targets comprises: a positive to-be-determined target selected from a plurality of positive to-be-determined targets in the plurality of to-be-determined targets, and a negative to-be-determined target selected from a plurality of negative to-be-determined targets in the plurality of to-be-determined targets, wherein the selected positive to-be-determined target and the selected negative to-be-determined target meet a predetermined ratio; and

for one to-be-determined target, if a ratio of an overlap area between the to-be-determined target and a standard bounding box of the to-be-determined target having

a largest overlap area to an area covered by both the to-be-determined target and the standard bounding box is greater than a first ratio threshold, the to-be-determined target is a positive to-be-determined target; and if the ratio is smaller than a second ratio threshold, the to-be-determined target is a negative to-be-determined target.

1517. An apparatus for object detection, comprising:

a to-be-determined target obtaining module, configured to obtain a plurality of to-be-determined targets in a to-be-detected image;

a category and position offset value prediction module, configured to determine, for at least one category, confidences of a plurality of to-be-determined targets respectively, determine categories of the plurality of to-be-determined targets according to the confidences, and respectively determine position offset values corresponding to the categories of the plurality of to-be-determined targets;

a position offset determination module, configured to respectively use the position offset values corresponding to the categories of the plurality of to-be-determined targets as position offset values of the plurality of to-be-determined targets; and

a detection result determination module, configured to determine a category and position information of at least one to-be-determined target in the to-be-detected image according to the categories of the plurality of to-be-determined targets, the position offset values of the plurality of to-be-determined targets, and confidences of the categories of the plurality of to-be-determined targets.,

wherein the to-be-determined target obtaining module is further configured to:

obtain the plurality of to-be-determined targets formed based on at least one predetermined region size in the to-be-detected image,

wherein the to-be-determined target obtaining module is specifically configured to:

obtain a feature map of the to-be-detected image;

form a reference box of a feature point in the feature map according to reference box configuration information, wherein the reference box configuration information is preset and a number and sizes of reference boxes are defined by the reference box configuration

information;

use the reference box of the feature point as one to-be-determined target; and

obtain, respectively corresponding to a plurality of feature points in the feature map, the plurality of to-be-determined targets.

1618. An electronic device, comprising: a memory, configured to store a computer program; and

a processor, configured to execute the computer program stored in the memory, wherein when the computer program is executed, the method for object detection according to any one of claims 1 to 1416 is implemented.

1719. A computer-readable storage medium having a computer program stored thereon, wherein when the computer program is executed by a processor, the method for object detection according to any one of claims 1 to 1416 is implemented.

18~~20~~. A computer program product, comprising a computer-readable code, wherein when the computer-readable code runs in a device, a processor in the device executes instructions for implementing the method for object detection according to any one of claims 1 to 14 ~~16~~.