CLAIMS

1. A method for detecting an object, comprising:

predicting, by a deep convolutional neural network for target area frame detection, to obtain a plurality of fusion feature maps from a to-be-processed image, wherein the deep convolutional neural network comprises a first sub-network and a second sub-network; the first sub-network has at least one down-sampling layer, and the second sub-network has at least one up-sampling layer; the fusion feature maps are obtained through a first feature map and a second feature map; and the first feature map is obtained from the first sub-network, and the second feature map is obtained from the second sub-network; and

obtaining target area frame data according to the plurality of fusion feature maps.

2. The method according to claim 1, wherein the second sub-network is provided at the tail end of the first sub-network; the first sub-network has a plurality of first convolution layers and the at least one down-sampling layer; the down-sampling layer is provided between the plurality of first convolution layers; the second sub-network has a plurality of second convolution layers and the at least one up-sampling layer; the up-sampling layer is provided between the plurality of second convolution layers; the first convolution layers and the second convolution layers are provided symmetrically; and the at least one down-sampling layer and the at least one up-sampling layer are symmetrically provided, separately.

3. The method according to claim 2, wherein the at least one first convolution layer is provided with a first output branch for outputting the first feature map, and the second convolution layer is provided with a second output branch for outputting the second feature map.

4. The method according to claim 3, wherein the second sub-network further has a plurality of third convolution layers, and an input of the third convolution layer comprises the first output branch and the second output branch;

the predicting to obtain the plurality of fusion feature maps comprises:

separately obtaining the fusion feature maps from output ends of the plurality of third convolution layers.

5. The method according to any one of claims 1 to 4, wherein at least one point in the fusion feature map has frame fusion detection data and prediction accuracy information corresponding to a plurality of object detection frames;

the obtaining the target area frame data according to the plurality of fusion feature maps comprises:

separately obtaining target area frame data corresponding to the fusion feature maps according to the frame fusion detection data and prediction accuracy information in the at least one fusion feature map.

6. The method according to any one of claims 1 to 5, wherein the obtaining the target area frame data according to the plurality of fusion feature maps comprises:

separately obtaining primary target area frame data corresponding to the fusion feature maps;

iteratively executing the following object area frame regression operation until the iteration meets an iteration termination condition: adjusting the fusion feature maps; and obtaining new primary target area frame data from the adjusted fusion feature maps; and

using the primary target area frame data obtained through the iteration as target area frame data in the to-be-processed image.

7. The method according to claim 6, wherein the deep convolutional neural network further comprises a third sub-network; the third sub-network has a plurality of groups of fourth convolutional layers and a plurality of pooling layers; the plurality of groups of fourth convolutional layers corresponds to the third convolutional layers, respectively; the plurality of pooling layers corresponds to the plurality of groups of fourth convolutional layers, respectively; and an input of each of the pooling layers comprises the adjusted fusion feature maps and the primary target area frame data.

8. The method according to claim 7, wherein the object area frame regression operation comprises:

separately convoluting the current fusion feature maps through the fourth convolutional layers to obtain the adjusted fusion feature maps;

performing area pooling on the adjusted fusion feature maps through the pooling layers according to the current primary target area frame data to obtain a new fusion feature map; and

obtaining the new primary target area frame data from the new fusion feature map.

9. The method according to claim 8, wherein the third sub-network further has a fifth convolution layer provided at an output end of the pooling layer;

the obtaining the new primary target area frame data from the new fusion feature map comprises:

performing normalized convolution on the new fusion feature map through the fifth convolution layer; and

obtaining the new primary target area frame data from the fusion feature map subjected to normalized convolution.

10. The method according to any one of claims 7 to 9, wherein both the first sub-network and the second sub-network are Inception-Batch Normalization (Inception-BN) network structures, and the third sub-network is a Residual Network (ResNet) structure.

11. The method according to claim 1, comprising:

training the deep convolutional neural network for target area frame detection by:

inputting a sample image containing target area frame labeling information in the deep convolutional neural network for target area frame detection to detect to obtain a plurality of fusion feature maps corresponding to the sample image;

obtaining target area frame data of the sample image according to the plurality of fusion feature maps corresponding to the sample image;

determining first differential data for object frame detection according to the obtained target area frame data of the sample image and the target area frame labeling information; and

adjusting a network parameter of the deep convolutional neural network according to the first differential data.

12. The method according to claim 11, wherein the deep convolutional neural network further comprises a third sub-network; the third sub-network has a plurality of groups of fourth convolutional layers and a plurality of pooling layers; the plurality of groups of fourth convolutional layers corresponds to the third convolutional layers, respectively; the plurality of pooling layers corresponds to the plurality of groups of fourth convolutional layers, respectively; and an input of each of the pooling layers comprises the adjusted fusion feature maps and the primary target area frame data.

13. The method according to claim 12, further comprising:

iteratively executing the following target area frame regression training operation until the iteration meets an iteration termination condition:

separately convoluting the current fusion feature maps through the fourth convolutional layers to obtain the adjusted fusion feature maps;

performing area pooling on the adjusted fusion feature maps through the pooling layers according to the current primary target area frame data to obtain a new fusion feature map;

obtaining the new primary target area frame data from the new fusion feature map;

determining second differential data for object frame detection according to frame regression data between the unadjusted primary target area frame data and the new primary target area frame data, the new primary target area frame data, and corresponding target area frame labeling information; and

adjusting a network parameter of the third sub-network according to the second differential data.

14. The method according to any one of claims 11 to 13, wherein prior to inputting a sample image containing target area frame labeling information in a deep convolutional neural network for target area frame detection to detect to obtain a plurality of fusion feature maps, the method further comprises:

zooming the sample image so that a ground truth of at least one object area frame is covered by the object detection frame.

15. The method according to any one of claims 12 to 14, wherein the target area frame labeling information of the sample image comprises labeling information of a positive sample area frame and labeling information of a negative sample area frame;

the overlapping ratio of the positive sample area frame to the ground truth of the object area frame is not less than a first overlapping ratio value; the overlapping ratio of the negative sample area frame to the ground truth of the object area frame is not greater than a second overlapping ratio value; and the first overlapping ratio value is greater than the second overlapping ratio value.

16. The method according to claim 15, wherein the target area frame labeling information of the sample image further comprises labeling information of a neutral sample area frame; and the overlapping ratio of the neutral sample area frame to the ground truth of the object area frame is between the first overlapping ratio value and the second overlapping ratio value.

17. The method according to claim 16, wherein in all the sample images,

the proportion of a sum of the labeled positive sample area frames in the total number of frames of the positive sample area frames, the negative same area frames and the neutral sample area frames is not less than a predetermined first ratio, the first ratio is greater than 50%;

the proportion of a sum of the labeled negative sample area frames in the total number of frames is not greater than a predetermined second ratio; and

the proportion of a sum of the labeled neutral sample area frames in the total number of frames is not greater than a predetermined third ratio, the third ratio is not greater than half of the sum of the first ratio and the second ratio.

18. An apparatus for detecting an object, comprising:

a fusion feature map prediction module, configured to predict to obtain a plurality of fusion feature maps from a to-be-processed image through a deep convolutional neural network for target area frame detection, wherein the deep

convolutional neural network comprises a first sub-network and a second sub-network; the first sub-network has at least one down-sampling layer, and the second sub-network has at least one up-sampling layer; the fusion feature maps are obtained through a first feature map and a second feature map; and the first feature map is obtained from the first sub-network, and the second feature map is obtained from the second sub-network; and

a target area frame prediction module, configured to obtain target area frame data according to the plurality of fusion feature maps obtained by the fusion feature map prediction module.

19. An electronic device, comprising a processor and a memory;

the memory is configured to store at least one executable instruction, which enables the processor to execute the operations corresponding to the method for detecting an object according to any one of claims 1 to 17.

20. A computer-readable storage medium, configured to store computer-readable instructions, wherein when the instructions are executed, the operation in each step of the method for detecting an object according to any one of claims 1 to 17 is implemented.