CLAIMS

1. A method for target object recognition, comprising:

obtaining an image area corresponding to an object of an image to be detected;

performing target object detection on the image area corresponding to the object of the image to be detected to obtain target object prediction information of the object, wherein the target object prediction information is confidence information that the detected object is a target object;

performing key point detection on the image area corresponding to the object of the image to be detected to obtain key point prediction information of the object, wherein the key point prediction information is confidence information that a key point of the detected object is a key point of the target object;

detecting deflection angle information of the object from the image area;

fusing the target object prediction information and the key point prediction information with the deflection angle information to obtain comprehensive prediction information of the object; and

recognizing the target object according to the comprehensive prediction information.

2. The method according to claim 1, wherein the performing key point detection on the object of the image to be detected to obtain the key point prediction information of the object comprises:

performing, using a neural network model for positioning a key point, key point detection on the object of the image to be detected to obtain the key point prediction information of the object.

3. The method according to claim 1, wherein the detecting deflection angle information of the object from the image area comprises:

detecting the deflection angle information of the object from the image area by using a neural network model for object classification.

4. The method according to any one of claims 1 to 3, wherein

the target object is a face;

the detecting the deflection angle information of the object from the image area comprises:

detecting a pitch angle of the face and/or a yaw angle of the face from the image area;

the fusing the target object prediction information and the key point prediction information with the deflection angle information to obtain the comprehensive prediction information of the object comprises:

normalizing the pitch angle of the face and/or the yaw angle of the face according to an applicable exponential function; and

multiplying the target object prediction information, the key point prediction information, and the normalized pitch angle of the face to obtain the comprehensive prediction information of the object;

or,

multiplying the target object prediction information, the key point prediction information, and the normalized yaw angle of the face to obtain the comprehensive prediction information of the object;

or,

multiplying the target object prediction information, the key point prediction information, the normalized pitch angle of the face, and the normalized yaw angle of the face to obtain the comprehensive prediction information of the object.

5. The method according to any one of claims 1 to 4, wherein the image to be detected is a video frame image;

after recognizing the target object according to the comprehensive prediction information, the method further comprising:

tracking the target object according to a result of recognizing the target object from multiple video frame images;

or,

select a video frame image having a highest comprehensive prediction quality from the multiple video frame images as a captured image according to comprehensive prediction information separately obtained for the multiple video frame images;

or,

selecting a predetermined number of video frame images from the multiple video frame images according to the comprehensive prediction information separately obtained for the multiple video frame images, and performing feature fusion on the selected video frame images.

6. An apparatus for target object recognition, comprising:

an image area obtaining module, configured to obtain an image area corresponding to the object of the image to be detected;

an object detection module, configured to perform target object detection on the image area corresponding to the object of the image to be detected to obtain target object prediction information of the object, wherein the target object prediction information is confidence information that the detected object is a target object;

a key point detection module, configured to perform key point detection on the image area corresponding to the object of the image to be detected to obtain key point prediction information of the object, wherein the key point prediction information is confidence information that a key point of the detected object is a key point of the target object;

a deflection angle detection module, configured to: detect deflection angle information of the object from the image area obtained by the image area obtaining module;

a prediction information fusion module, configured to fuse the target object prediction information obtained by the object detection module and the key point prediction information obtained by the key point detection module with the deflection angle information obtained by the deflection angle detection module to obtain comprehensive prediction information of the object; and

an object recognition module, configured to recognize the target object according to the comprehensive prediction information obtained by the prediction information fusion module.

7. The apparatus according to claim 6, wherein the key point detection module is configured to perform, using a neural network model for positioning the key point, key point detection on the object of the image to be detected to obtain the key point prediction information of the object.

8. The apparatus according to claim 6, wherein the deflection angle detection module is configured to detect the deflection angle information of the object from the image area by using the neural network model for object classification.

9. The apparatus according to any one of claims 6 to 8, wherein

the target object is a face;

the apparatus further comprises:

a face deflection angle detection module, configured to: before the target object prediction information and the key point prediction information is fused with the deflection angle information to obtain comprehensive prediction information of the object, detect a pitch angle of the face and/or a yaw angle of the face from the image area; and

the prediction information fusion module is configured to:

normalize the pitch angle of the face and/or the yaw angle of the face according to the applicable exponential function; and

multiply the target object prediction information, the key point prediction information, and the normalized pitch angle of the face to obtain the comprehensive prediction information of the object;

or,

multiply the target object prediction information, the key point prediction information, and the normalized yaw angle to obtain the comprehensive prediction information of the object;

or,

multiply the target object prediction information, the key point prediction information, the normalized pitch angle of the face, and the normalized yaw angle of the face to obtain the comprehensive prediction information of the object.

10. The apparatus according to any one of claims 6 to 9, wherein the image to be detected is a video frame image;

the apparatus further comprises:

an object tracking module, configured to after the target object is recognized according to the comprehensive prediction information, track the target object according to a result of recognizing the target object from multiple video frame images;

or,

a captured image selection module, configured to select a video frame image having a highest comprehensive prediction quality from the multiple video frame images as a captured image according to the comprehensive prediction information separately obtained for the multiple video frame images;

or,

a feature fusion module, configured to select a predetermined number of video frame images from the multiple video frame images according to the comprehensive prediction information separately obtained for the multiple video frame images, and perform feature fusion on the selected video frame images.

11. An electronic device, comprising: a processor, a memory, a communication element, and a communication bus, wherein the processor, the memory, and the communication element communicate with each other through the communication bus;

the memory is configured to store at least one executable instruction, wherein the executable instruction enables the processor to execute the method for target object recognition according to any one of claims 1 to 5.

12. A computer-readable storage medium, having computer program instructions stored thereon, wherein the program instructions, when being executed by a processor, implements the method for target object recognition according to any one of claims 1 to 5.

13. A computer program, comprising computer program instructions, wherein the program instructions, when being executed by a processor, implements the method for target object recognition according to any one of claims 1 to 5.