WHAT IS CLAIMED IS:

1. A method for object detection implemented on a computing device including at least one processor and a storage device, comprising:

acquiring a group of images of an object captured by at least three imaging apparatuses, the at least three imaging apparatuses are different types, the group of images including at least three images of different modalities with a same visual angle and overlapping area, wherein the at least three imaging apparatuses are fixed on at least three fixing positions of a supporting assembly and shooting axes of the at least three imaging apparatuses are in a same vertical plane; and

determining a recognition result of the object based on the group of images according to an object recognition model, the recognition result including at least one of a position of the object or a category of the object,

wherein the object recognition model includes a fusion feature extraction submodel and an object recognition submodel, and the determining the recognition result of the object based on the group of images according to the object recognition model includes:

obtaining a fusion feature image based on the group of images according to the fusion feature extraction submodel, the fusion feature image including at least one fusion feature of the group of images; and

determining the recognition result based on the fusion feature image according to the object recognition submodel,

wherein the obtaining the fusion feature image includes:

obtaining a stacked matrix by concatenating matrixes corresponding to the group of images using a matrix concatenation function, the matrixes corresponding to the group of images including matrixes of a plurality of feature images of the group of images or matrixes of images in the group of images;

determining a fused matrix by reducing a dimension of the stacked matrix; and

determining the fusion feature image based on the fused matrix.

2. The method of claim 1, wherein the obtaining the fusion feature image includes:

obtaining a fusion image by fusing the group of images; and

determining the fusion feature image by extracting features from the fusion image.

3. The method of claim 1, wherein the obtaining the fusion feature image includes:

obtaining the plurality of feature images of the group of images by extracting image features from each image of the group of images; and

determining the fusion feature image by fusing the plurality of feature images.

4. The method of claim 1, wherein the obtaining the fusion feature image includes:

obtaining the plurality of feature images of the group of images by extracting image features from each image of the group of images;

obtaining a preliminary fusion feature image by fusing the plurality of feature images; and

determining the fusion feature image by extracting features from the preliminary fusion feature image.

5. The method of any one of claims 1-4, wherein the fusion feature extraction submodel includes a plurality of convolution blocks connected in series, each of a portion of the plurality of convolution blocks including at least one convolution layer and a pooling layer, and the last convolution layer including at least one convolution layer.

6. The method of any one of claims 1-5, wherein the fusion feature extraction submodel includes:

a first convolution block including at least one convolution layer and a pooling layer;

a second convolution block including at least one convolution layer and a pooling layer;

a third convolution block including at least one convolution layer and a pooling layer;

a fourth convolution block including at least one convolution layer and a pooling layer; and

a fifth convolution block including at least one convolution layer.

7. The method of claim 1, wherein the object recognition model includes a plurality of object recognition submodels, each of the plurality of object recognition submodels corresponding to an image of a modality; and

wherein the obtaining the recognition result of the object based on the group of images according to the object recognition model includes:

obtaining a plurality of candidate recognition results and confidence scores of the plurality of candidate recognition results according to the plurality of object recognition submodels based on the group of images, each of the plurality of candidate recognition results corresponding to one of the plurality of recognition submodels; and

determining the recognition result based on the plurality of candidate recognition results and the confidence scores.

8. The method claim 7, wherein the determining the recognition result based on the plurality of candidate recognition results and the confidence scores includes:

determining a weight of each of the plurality of candidate recognition results based on environment information; and

determining the recognition result based on the weight of each of the plurality of candidate recognition results, the plurality of candidate recognition results, and the confidence scores.

9. The method of any one of claims 1-8, wherein

the weight is determined for each of the confidence scores of the plurality of candidate recognition results,

the group of images includes a color image, an infrared image, and a polarization image,

the environment information includes weather information,

if weather is sunny or cloudy, a weight for the color image exceeds a weight for the polarization image, and a weight for the polarization image exceeds a weight for the infrared image, and

if the weather is rainy, snowy, or foggy, the weight for the infrared image exceeds the weight for the color image, and the weight for the color image exceeds the weight for the polarization image.

10. The method of claim 7, wherein the determining the recognition result based on the plurality of candidate recognition results and the confidence scores includes:

determining an average confidence score of the confidence scores; and

when the average confidence score is larger than a threshold, determining that the object is a specific category and outputting a bounding box of the object.

11. The method of claim 7, wherein the determining the recognition result based on the plurality of candidate recognition results and the confidence scores includes:

determining a sum of the confidence scores; and

when the sum of the confidence scores is larger than a threshold, determining that the object is a specific category and outputting a bounding box of the object.

12. The method of claim 1, wherein the at least three imaging apparatuses include a visible light camera, an infrared camera, and a polarization camera.

13. The method of claim 1, wherein pixels in each of the at least three images correspond to each other one by one.

14. A system for object detection, comprising:

at least one storage device including a set of instructions; and

at least one processor configured to communicate with the at least one storage device, wherein, when the set of instructions are executed, the at least one processor is configured to instruct the system to perform operations, including:

acquiring a group of images of an object captured by at least three imaging apparatuses, the at least three imaging apparatuses are different types, the group of images including at least three images of different modalities with a same visual angle and overlapping area, wherein the at least three imaging apparatuses are

fixed on at least three fixing positions of a supporting assembly and shooting axes of the at least three imaging apparatuses are in a same vertical plane; and

determining a recognition result of the object based on the group of images according to an object recognition model, the recognition result including at least one of a position of the object or a category of the object,

wherein the object recognition model includes a fusion feature extraction submodel and an object recognition submodel, and the determining the recognition result of the object based on the group of images according to the object recognition model includes:

obtaining a fusion feature image based on the group of images according to the fusion feature extraction submodel, the fusion feature image including at least one fusion feature of the group of images; and

determining the recognition result based on the fusion feature image according to the object recognition submodel,

wherein the obtaining the fusion feature image includes:

obtaining a stacked matrix by concatenating matrixes corresponding to the group of images using a matrix concatenation function, the matrixes corresponding to the group of images including matrixes of a plurality of feature images of the group of images or matrixes of images in the group of images;

determining a fused matrix by reducing a dimension of the stacked matrix; and

determining the fusion feature image based on the fused matrix.

15. The system of claim 14, wherein the obtaining the fusion feature image includes:

obtaining a fusion image by fusing the group of images; and

determining the fusion feature image by extracting features from the fusion image.

16. The system of claim 14, wherein the obtaining the fusion feature image includes:

obtaining the plurality of feature images of the group of images by extracting image features from each image of the group of images; and

determining the fusion feature image by fusing the plurality of feature images.

17. The system of claim 14, wherein the obtaining the fusion feature image includes:

obtaining the plurality of feature images of the group of images by extracting image features from each image of the group of images;

obtaining a preliminary fusion feature image by fusing the plurality of feature images; and

determining the fusion feature image by extracting features from the preliminary fusion feature image.

18. The system of any one of claims 14-17, wherein the fusion feature extraction submodel includes a plurality of convolution blocks connected in series, each of a portion of the plurality of convolution blocks including at least one convolution layer and a pooling layer, and the last convolution layer including at least one convolution layer.

19. The system of any one of claims 14-18, wherein the fusion feature extraction submodel includes:

a first convolution block including at least one convolution layer and a pooling layer;

a second convolution block including at least one convolution layer and a pooling layer;

a third convolution block including at least one convolution layer and a pooling layer;

a fourth convolution block including at least one convolution layer and a pooling layer; and

a fifth convolution block including at least one convolution layer.

20. A non-transitory computer-readable storage medium storing computer instructions, and when a computer reads the computer instructions in the storage medium, the computer executes a method for object detection, comprising:

acquiring a group of images of an object captured by at least three imaging apparatuses, the at least three imaging apparatuses are different types, the group of images including at least three images of different modalities with a same visual angle and overlapping area, wherein the at least three imaging apparatuses are fixed on at least three fixing positions of a supporting assembly and shooting axes of the at least three imaging apparatuses are in a same vertical plane; and

determining a recognition result of the object based on the group of images according to an object recognition model, the recognition result including at least one of a position of the object or a category of the object,

wherein the object recognition model includes a fusion feature extraction submodel and an object recognition submodel, and the determining the recognition result of the object based on the group of images according to the object recognition model includes:

obtaining a fusion feature image based on the group of images according to the fusion feature extraction submodel, the fusion feature image including at least one fusion feature of the group of images; and

determining the recognition result based on the fusion feature image according to the object recognition submodel,

wherein the obtaining the fusion feature image includes:

obtaining a stacked matrix by concatenating matrixes corresponding to the group of images using a matrix concatenation function, the matrixes corresponding to the group of images including matrixes of a plurality of feature images of the group of images or matrixes of images in the group of images;

determining a fused matrix by reducing a dimension of the stacked matrix; and

determining the fusion feature image based on the fused matrix.