CLAIMS

1. A method for estimating a monocular depth, comprising:

obtaining an image to be processed;

inputting the image to be processed into a trained monocular depth estimation network model, to obtain an analysis result of the image to be processed, wherein the monocular depth estimation network model is trained by supervision of a disparity map output by a first binocular matching neural network model, and a training process of the first binocular matching neural network model comprises:

training a second binocular matching neural network model according to obtained synthetic sample data to obtain a trained second binocular matching network model; and adjusting a parameter of the trained second binocular matching neural network model according to obtained real sample data, to obtain the first binocular matching neural network model; and

outputting the analysis result of the image to be processed;

wherein the method further comprises:

obtaining a synthesized binocular picture with a depth tag as the obtained synthetic sample data, wherein the synthesized binocular picture comprises a synthesized left picture and a synthesized right picture, and

wherein obtaining the synthesized binocular picture with the depth tag comprises:

constructing a virtual 3D scene by means of the rendering engine; mapping the 3D scene into a binocular picture by means of two virtual cameras; obtaining the depth data of the synthesized binocular picture according to the position where the virtual 3D scene is constructed, the direction in which the virtual 3D scene is constructed, and the lens focus of the virtual camera; and annotating the binocular picture according to the depth data to obtain the synthesized binocular picture.

2. The method according to claim 1, wherein the training a second binocular matching neural network model according to the obtained synthetic sample data comprises:

training the second binocular matching neural network model according to the synthesized binocular picture to obtain the trained second binocular matching neural network model, wherein outputs of the trained second binocular matching neural network model are a disparity map and an occlusion map, the disparity map describing a disparity distance between each pixel point in the synthesized left picture and a corresponding pixel point in the synthesized right picture, the disparity distance being in units of pixels, and the occlusion map describing whether a respective pixel point in the synthesized right picture corresponding to each pixel point in the synthesized left picture is occluded by an object.

3. The method according to claim 1, wherein the adjusting a parameter of the trained second binocular matching neural network model according to the obtained real sample data, to obtain the first binocular matching neural network model comprises:

performing supervised training on the trained second binocular matching neural network model according to the obtained real binocular data with a depth tag, to adjust a weight of the trained second binocular matching neural network model to obtain the first binocular matching neural network model.

4. The method according to claim 1, wherein the adjusting a parameter of the trained second binocular matching neural network model according to the obtained real sample data, to obtain the first binocular matching neural network model, further comprises:

performing unsupervised training on the trained second binocular matching neural network model according to the obtained real binocular data without a depth tag, to adjust a weight of the trained second binocular matching neural network model to obtain the first binocular matching neural network model.

5. The method according to claim 4, wherein the performing unsupervised training on the trained second binocular matching neural network model according to the obtained real binocular data without a depth tag, to adjust the weight of the trained

second binocular matching neural network model to obtain the first binocular matching neural network model comprises:

performing unsupervised training on the trained second binocular matching neural network model according to the real binocular data without a depth tag by using a loss function, so as to adjust the weight of the trained second binocular matching neural network model to obtain the first binocular matching neural network model.

6. The method according to claim 3, wherein the real binocular data with a depth tag comprises a left picture and a right picture, and correspondingly, the training process of the monocular depth estimation network model comprises:

obtaining the left picture or right picture in the real binocular data with a depth tag as a training sample; and

training the monocular depth estimation network model according to the left picture or right picture in the real binocular data with a depth tag.

7. The method according to claim4 or 5, wherein the real binocular data without a depth tag comprises a left picture and a right picture, and correspondingly, the training process of the monocular depth estimation network model comprises:

inputting the real binocular data without a depth tag into the first binocular matching neural network model to obtain a corresponding disparity map;

determining a depth map corresponding to the disparity map according to the corresponding disparity map, a lens baseline distance of a camera that captures the real binocular data without a depth tag, and a lens focus of the camera that captures the real binocular data without a depth tag; and

performing supervision on the monocular depth estimation network model according to the depth map corresponding to the disparity map by using the left picture or right picture in the real binocular data without a depth tag as sample data, to train the monocular depth estimation network model.

8. The method according to claim6 or 7, wherein the analysis result of the image to be processed comprises a disparity map output by the monocular depth estimation network model, and correspondingly, the method further comprises:

determining a depth map corresponding to the disparity map according to the disparity map output by the monocular depth estimation network model, the lens baseline distance of the camera that captures a picture input into the monocular depth estimation network model, and the lens focus of the camera that captures the picture input into the monocular depth estimation network model; and

outputting the depth map corresponding to the disparity map.

9. A monocular depth estimation apparatus, comprising: an obtaining module, an execution module, an output module, a first training module, a second training module and a first obtaining module, wherein

the obtaining module is configured to obtain an image to be processed;

the execution module is configured to input the image to be processed into a trained monocular depth estimation network model, to obtain an analysis result of the image to be processed, wherein the monocular depth estimation network model is trained by supervision of a disparity map output by a first binocular matching neural network model; and

the output module is configured to output the analysis result of the image to be processed,

the first training module is configured to train a second binocular matching neural network model according to the obtained synthetic sample data to obtain a trained second binocular matching network model;

the second training module is configured to adjust a parameter of the trained second binocular matching neural network model according to the obtained real sample data, to obtain the first binocular matching neural network model, and

the first obtaining module is configured to obtain a synthesized binocular picture with a depth tag as the obtained synthetic sample data, wherein the synthesized binocular picture comprises a synthesized left picture and a synthesized

right picture,

wherein the apparatus further comprises: a constructing module, a mapping module, a second obtaining module and a third obtaining module, wherein

the constructing module is configured to construct a virtual 3D scene by means of the rendering engine; the mapping module is configured to map the 3D scene into a binocular picture by means of two virtual cameras;

the second obtaining module is configured to obtain the depth data of the synthesized binocular picture according to the position where the virtual 3D scene is constructed, the direction in which the virtual 3D scene is constructed, and the lens focus of the virtual camera; and

the third obtaining module is configured to annotate the binocular picture according to the depth data to obtain the synthesized binocular picture.

10. The apparatus according to claim 9, wherein the first training module comprises:

a first training unit, configured to train the second binocular matching neural network model according to the synthesized binocular picture to obtain the trained second binocular matching neural network model, wherein outputs of the trained second binocular matching neural network model are a disparity map and an occlusion map, the disparity map describing a disparity distance between each pixel point in the synthesized left picture and a corresponding pixel point in the synthesized right picture, the disparity distance being in units of pixels, and the occlusion map describing whether a respective pixel point in the right picture corresponding to each pixel point in the synthesized left picture is occluded by an object.

11. The apparatus according to claim 9, wherein the second training module comprises a second training unit, configured to perform supervised training on the trained second binocular matching neural network model according to the obtained real binocular data with a depth tag, to adjust a weight of the trained second binocular matching neural network model to obtain the first binocular matching neural network model.

12. The apparatus according to claim 9, wherein the second training unit is further configured to perform unsupervised training on the trained second binocular matching neural network model according to the obtained real binocular data without a depth tag, to adjust the weight of the trained second binocular matching neural network model to obtain the first binocular matching neural network model.

13. The apparatus according to claim 12, wherein the second training unit comprises a second training component, configured to perform unsupervised training on the trained second binocular matching neural network model according to the real binocular data without a depth tag by using a loss function, to adjust the weight of the trained second binocular matching neural network model to obtain the first binocular matching neural network model.

14. The apparatus according to claim 11, wherein the real binocular data with a depth tag comprises a left picture or a right picture, and correspondingly, the apparatus further comprises a third training module, configured to obtain the left or right picture in the real binocular data with a depth tag as a training sample; and train the monocular depth estimation network model according to the left picture or right picture in the real binocular data with a depth tag.

15. A device for estimating a monocular depth, comprising a memory and a processor, wherein the memory stores computer programs capable of running in the processor, and when the programs being executed, the processor implements steps of the method for estimating a monocular depth according to any one of claims 1 to 8.

16. A computer readable storage medium, having stored thereon computer programs that, when being executed by a processor, implement steps of the method for estimating a monocular depth according to any one of claims 1 to 8.