

CLAIMS

1. A key point detection method, comprising:

obtaining a plurality of first feature maps at a plurality of scales for an input image, scales of the plurality of first feature maps having a multiple relationship;

performing forward processing on each of the plurality of first feature maps by using a first pyramid neural network to obtain a plurality of second feature maps in one-to-one correspondence to the plurality of first feature maps, wherein each of the plurality of second feature maps has the same scale as that of a first feature map corresponding to the second feature map;

performing reverse processing on each of the plurality of second feature maps by using a second pyramid neural network to obtain a plurality of third feature maps in one-to-one correspondence to the plurality of second feature maps, wherein each of the plurality of third feature maps has the same scale as that of a second feature map corresponding to the third feature map; and

performing feature fusion processing on each of the plurality of third feature maps, and obtaining a position of each key point in the input image by using a feature map subjected to the feature fusion processing.

2. The method according to claim 1, wherein the obtaining the plurality of first feature maps at the plurality of scales for the input image comprises:

adjusting the input image to a first image of a preset specification; and

inputting the first image to a residual neural network, and performing downsampling processing at different sampling frequencies on the first image to obtain a plurality of first feature maps at different scales.

3. The method according to claim 1, wherein the forward processing comprises first convolution processing and first linear interpolation processing, and the reverse processing comprises second convolution processing and second linear interpolation processing.

4. The method according to any one of claims 1 to 3, wherein the performing forward processing on each of the plurality of first feature maps by using a first pyramid neural network to obtain a plurality of second feature maps in one-to-one correspondence to the plurality of first feature maps comprises:

performing convolution processing on a first feature map C_n in first feature maps $C_1...C_n$ by using a first convolution kernel to obtain a second feature map F_n corresponding to the first feature map C_n , wherein n represents a number of the first feature maps, and n is an integer greater than 1;

performing linear interpolation processing on the second feature map F_n to obtain a first intermediate feature map F'_n corresponding to the second feature map F_n , wherein the first intermediate feature map F'_n has the same scale as that of a first feature map C_{n-1} ;

performing convolution processing on first feature maps $C_1...C_{n-1}$ other than the first feature map C_n by using a second convolution kernel to obtain second intermediate feature maps $C'_1...C'_{n-1}$ respectively in one-to-one correspondence to the first feature maps $C_1...C_{n-1}$, wherein each of the second intermediate feature maps has the same scale as that of a first feature map corresponding to the second intermediate feature map; and

obtaining second feature maps $F_1...F_{n-1}$ and first intermediate feature maps $F'_1...F'_{n-1}$ based on the second feature map F_n and each of the second intermediate feature maps $C'_1...C'_{n-1}$, wherein the second feature map F_i is obtained by performing superposition processing on the second intermediate feature map C'_i and the first intermediate feature map F'_{i+1} , the first intermediate feature map F'_i is obtained by performing linear interpolation

processing on its corresponding second feature map F_i , and the second intermediate feature map C'_i has the same scale as that of the first intermediate feature map F'_{i+1} , wherein i is an integer greater than or equal to 1 and less than n .

5. The method according to any one of claims 1 to 4, wherein the performing reverse processing on each of the plurality of second feature maps by using the second pyramid neural network to obtain the plurality of third feature maps in one-to-one correspondence to the plurality of second feature maps comprises:

performing convolution processing on a second feature map F_1 in second feature maps $F_1...F_m$ by using a third convolution kernel to obtain a third feature map R_1 corresponding to the second feature map F_1 , wherein m represents a number of the second feature maps, and m is an integer greater than 1;

performing convolution processing on second feature maps $F_2...F_m$ by using a fourth convolution kernel to obtain respective third intermediate feature maps $F''_2...F''_m$, wherein each of the third intermediate feature maps has the same scale as that of a second feature map corresponding to the third intermediate feature map;

performing convolution processing on the third feature map R_1 by using a fifth convolution kernel to obtain a fourth intermediate feature map R'_1 corresponding to the third feature map R_1 ; and

obtaining third feature maps $R_2...R_m$ and fourth intermediate feature maps $R'_2...R'_m$ by using the third intermediate feature maps $F''_2...F''_m$ and the fourth intermediate feature map R'_1 , wherein a third feature map R_j is obtained by superposition processing of a third intermediate feature map F''_j and a fourth intermediate feature map R'_{j-1} , and the fourth intermediate feature map R'_{j-1} is

obtained by performing convolution processing on its corresponding third feature map R_{j-1} by using a fifth convolution kernel, wherein j is greater than 1 and less than or equal to m .

6. The method according to any one of claims 1 to 5, wherein the performing feature fusion processing on each of the plurality of third feature maps, and obtaining the position of each key point in the input image by using the feature map subjected to the feature fusion processing comprises:

performing feature fusion processing on each of the plurality of third feature maps to obtain a fourth feature map; and

obtaining the position of each key point in the input image based on the fourth feature map.

7. The method according to claim 6, wherein the performing feature fusion processing on each of the plurality of third feature maps to obtain the fourth feature map comprises:

adjusting each of the plurality of third feature maps to a plurality of feature maps of the same scale by using linear interpolation; and

connecting the plurality of feature maps of the same scale to obtain the fourth feature map.

8. The method according to claim 6 or 7, wherein before the performing feature fusion processing on each of the plurality of third feature maps to obtain the fourth feature map, the method further comprises:

respectively inputting a first group of third feature maps to different bottleneck block structures, and perform convolution processing on the first group of third feature maps to respectively obtain updated third feature maps, each of the bottleneck block structures comprising a different number of convolution modules, wherein the plurality of third feature maps comprise the first group of third feature maps and a second group of third feature maps, and

the first group of third feature maps and the second group of third feature maps each comprises at least one third feature map.

9. The method according to claim 8, wherein the performing feature fusion processing on each of the plurality of third feature maps to obtain the fourth feature map comprises:

adjusting each of the updated third feature maps and the second group of third feature maps to feature maps of the same scale by using linear interpolation; and

connecting the feature maps of the same scale to obtain the fourth feature map.

10. The method according to any one of claims 6 to 9, wherein the obtaining the position of each key point in the input image based on the fourth feature map comprises:

performing dimension reduction processing on the fourth feature map by using a fifth convolution kernel; and

determining positions of key points of the input image by using a fourth feature map subjected to the dimension reduction processing.

11. The method according to any one of claims 6 to 9, wherein the obtaining the position of each key point in the input image based on the fourth feature map comprises:

performing dimension reduction processing on the fourth feature map by using a fifth convolution kernel;

performing purification processing on features in the fourth feature map subjected to the dimension reduction processing by using a convolution block attention module to obtain a purified feature map; and

determining the positions of the key points of the input image by using a purified feature map.

12. The method according to any one of claims 1 to 11, further comprising: training the first pyramid neural network by using a training image data set,

comprising:

performing the forward processing on a plurality of first feature maps corresponding to each image in the training image data set by using the first pyramid neural network, to obtain a plurality of second feature maps corresponding to each image in the training image data set;

determining the obtained key points by using each second feature map;

obtaining a first loss of each key point according to a first loss function; and

reversely regulating each convolution kernel in the first pyramid neural network by using the first loss until a number of trainings reaches a set first threshold number of times.

13. The method according to any one of claims 1 to 12, further comprising: training the second pyramid neural network by using a training image data set, comprising:

performing, by using the second pyramid neural network, the reverse processing on the plurality of second feature map corresponding to each image in the training image data set output by the first pyramid neural network, to obtain a plurality of third feature map corresponding to each image in the training image data set;

determining the obtained key points by using each plurality of third feature maps;

obtaining a second loss of each key point according to a second loss function; and

reversely regulating convolution kernels in the second pyramid neural network by using the second loss until the number of trainings reaches a set second threshold number of times; or reversely regulating the convolution kernels in the first pyramid neural network and the convolution kernels in the second pyramid neural network by using the second loss until the number of

trainings reaches the set second threshold number of times.

14. The method according to any one of claims 1 to 13, wherein the feature fusion processing is performed on each of the plurality of third feature maps by using a feature extraction network, and

before the performing feature fusion processing on each of the plurality of third feature maps by using a feature extraction network, the method further comprises: training the feature extraction network by using the training image data set, comprising:

performing, by using the feature extraction network, the feature fusion processing on the plurality of third feature maps corresponding to each image in the training image data set output by the second pyramid neural network, and identifying key points of each image in the training image data set by using a feature map subjected to the feature fusion processing;

obtaining a third loss of each key point according to a third loss function; and

reversely regulating parameters of the feature extraction network by using a third loss value until the number of trainings reaches a set third threshold number of times; or reversely regulating, by using the third loss function, parameters of the convolution kernel in the first pyramid neural network, parameters of the convolution kernel in the second pyramid neural network, and parameters of the feature extraction network, until the number of training times reaches a set third threshold number of times.

15. A key point detection apparatus, comprising:

a multi-scale feature obtaining module, configured to obtain a plurality of first feature maps at a plurality of scales for an input image, scales of the plurality of first feature maps having a multiple relationship;

a forward processing module, configured to perform forward processing on each of the plurality of first feature maps by using a first pyramid neural network to obtain a plurality of second feature maps in one-to-one correspondence to the plurality of

first feature maps, wherein each of the plurality of second feature maps has the same scale as that of a first feature map corresponding to the second feature map;

a reverse processing module, configured to perform reverse processing on each of the plurality of second feature maps by using a second pyramid neural network to obtain a plurality of third feature maps in one-to-one correspondence to the plurality of second feature maps, wherein each of the plurality of third feature maps has the same scale as that of a second feature map corresponding to the third feature map; and

a key point detecting module, configured to perform feature fusion processing on each of the plurality of third feature maps, and obtain a position of each key point in the input image by using a feature map subjected to the feature fusion processing.

16. The apparatus according to claim 15, wherein the multi-scale feature obtaining module is configured to adjust the input image to a first image of a preset specification; and input the first image to a residual neural network, and perform downsampling processing at different sampling frequencies on the first image to obtain a plurality of first feature maps at different scales.

17. The apparatus according to claim 15 or 16, wherein the forward processing module is configured to perform convolution processing on a first feature map C_n in first feature maps $C_1...C_n$ by using a first convolution kernel to obtain a second feature map F_n corresponding to the first feature map C_n , wherein n represents a number of the first feature maps, and n is an integer greater than 1;

perform linear interpolation processing on the second feature map F_n to obtain a first intermediate feature map F_n^1 corresponding to the second feature map F_n , wherein the first intermediate feature map F_n^1 has the same scale as

that of a first feature map C_{n-1} ;

perform convolution processing on first feature maps $C_1...C_{n-1}$ other than the first feature map C_n by using a second convolution kernel to obtain second intermediate feature maps $C'_1...C'_{n-1}$ respectively in one-to-one correspondence to the first feature maps $C_1...C_{n-1}$, wherein each of the second intermediate feature maps has the same scale as that of a first feature maps corresponding to the second intermediate feature map; and

obtain second feature maps $F_1...F_{n-1}$ and first intermediate feature maps $F'_1...F'_{n-1}$ based on the second feature map F_n and each of the second intermediate feature maps $C'_1...C'_{n-1}$, wherein the second feature map F_i is obtained by performing superposition processing on the second intermediate feature map C'_i and the first intermediate feature map F'_{i+1} , the first intermediate feature map F'_i is obtained by performing linear interpolation on its corresponding second feature map F_i , and the second intermediate feature map C'_i has the same scale as that of the first intermediate feature map F'_{i+1} , wherein i is an integer greater than or equal to 1 and less than n .

18. The apparatus according to any one of claims 15 to 17, wherein the reverse processing module is configured to:

perform convolution processing on a second feature map F_1 in second feature maps $F_1...F_m$ by using a third convolution kernel to obtain a third feature map R_1 corresponding to the second feature map F_1 , wherein m represents a number of the second feature maps, and m is an integer greater than 1;

perform convolution processing on second feature maps $F_2...F_m$ by using a

fourth convolution kernel to obtain respective third intermediate feature maps $F'_2...F'_m$, wherein each of the third intermediate feature maps has the same scale as that of a second feature map corresponding to the third intermediate feature map;

perform convolution processing on the third feature map R_1 by using a fifth convolution kernel to obtain a fourth intermediate feature map R'_1 corresponding to the third feature map R_1 ; and

obtain third feature maps $R_2...R_m$ and fourth intermediate feature maps $R'_2...R'_m$ by using the third intermediate feature maps $F'_2...F'_m$ and the fourth intermediate feature map R'_1 , wherein a third feature map R_j is obtained by superposition processing of a third intermediate feature map F'_j and a fourth intermediate feature map R'_{j-1} , and the fourth intermediate feature map R'_{j-1} is obtained by performing convolution processing on its corresponding third feature map R_{j-1} by using a fifth convolution kernel, wherein j is greater than 1 and less than or equal to m .

19. An electronic device, comprising:

a processor; and

a memory configured to store instructions executable by the processor,

wherein the processor is configured to:

obtain a plurality of first feature maps at a plurality of scales for an input image, scales of the plurality of first feature maps having a multiple relationship;

perform forward processing on each of the plurality of first feature maps by using a first pyramid neural network to obtain a plurality of second feature maps in one-to-one correspondence to the plurality of first feature maps, wherein each

of the plurality of second feature maps has the same scale as that of a first feature map corresponding to the second feature map;

perform reverse processing on each of the plurality of second feature maps by using a second pyramid neural network to obtain a plurality of third feature maps in one-to-one correspondence to the plurality of second feature maps, wherein each of the plurality of third feature maps has the same scale as that of a second feature map corresponding to the third feature map; and

perform feature fusion processing on each of the plurality of third feature maps, and obtaining a position of each key point in the input image by using a feature map subjected to the feature fusion processing.

20. A computer-readable storage medium, having stored thereon computer program instructions that, when being executed by a processor, implements the method according to any one of claims 1 to 14.