CLAIMS

1. An instance segmentation method, comprising:

performing feature extraction on an image via a neural network to output features at at least two different hierarchies;

extracting region features corresponding to at least one instance candidate region in the image from the features at the at least two different hierarchies, and fusing region features corresponding to a same instance candidate region, to obtain a first fusion feature of each instance candidate region; and

performing instance segmentation based on each first fusion feature, to obtain at least one of an instance segmentation result of a corresponding instance candidate region or an instance segmentation result of the image.

2. The method according to claim 1, wherein the performing feature extraction on the image via the neural network to output the features at the at least two different hierarchies comprises: performing feature extraction on the image via the neural network to output the features at the at least two different hierarchies via at least two network layers with different network depths in the neural network.

3. The method according to claim 1 or 2, further comprising: after outputting the features at the at least two different hierarchies, performing at least one retracing fusion on the features at the at least two different hierarchies to obtain a second fusion feature, wherein the one retracing fusion comprises: fusing the features at different hierarchies output by the network layers with the different network depths respectively based on a network depth direction of the neural network according to two different hierarchy directions in sequence; and

the extracting the region features corresponding to the at least one instance candidate region in the image from the features at the at least two different hierarchies comprises: extracting the region features corresponding to the at least one instance candidate region from the second fusion feature,

wherein the two different hierarchy directions comprise: a direction from high hierarchy features to low hierarchy features and a direction from the low hierarchy features to the high hierarchy features.

wherein the according to the two different hierarchy directions in sequence comprises: along the direction from the high hierarchy features to the low hierarchy features and the direction from the low hierarchy features to the high hierarchy features in sequence; or along the direction from the low hierarchy features to the high hierarchy features and the direction from the high hierarchy features to the low hierarchy features in sequence.

4. The method according to claim 3, wherein the fusing the features at the different hierarchies output by the network layers with the different network depths respectively along the direction from the high hierarchy features to the low hierarchy features and the direction from the low hierarchy features to the high hierarchy features in sequence comprises:

along a direction from deep to shallow of the network depth of the neural network, after upsampling features at a higher hierarchy output by a network layer with a deeper network depth in the neural network in sequence, fusing the upsampled features with features at a lower hierarchy output by a network layer with a shallower network depth, to obtain a third fusion feature; and

along the direction from the low hierarchy features to the high hierarchy features, after downsampling fusion features at a lower hierarchy in sequence, fusing the downsampled fusion features with fusion features at a higher hierarchy in the third fusion feature,

wherein the features at the higher hierarchy comprise: features output by the network layer with a deeper network depth in the neural network or features obtained by performing at least one feature extraction on the features output by the network layer with the deeper network depth.

5. The method according to claim 4, wherein the after upsampling features at the higher hierarchy output by the network layer with the deeper network depth in the neural network in sequence, fusing the upsampled features with the features at the lower hierarchy output by the network layer with the shallower network depth comprises: after upsampling the features at the higher hierarchy output by the network layer with the deeper network depth in the neural network in sequence, fusing the

upsampled features with adjacent features at the lower hierarchy output by the network layer with the shallower network depth.

6. The method according to claim 4 or 5, wherein the after downsampling fusion features at the lower hierarchy in sequence, fusing the downsampled fusion features with the fusion features at the higher hierarchy in the third fusion feature comprises: after downsampling the fusion features at the lower hierarchy in sequence, fusing the downsampled fusion features with adjacent fusion features at the higher hierarchy in the third fusion feature.

7. The method according to claim 3, wherein the fusing the features at the different hierarchies output by the network layers with the different network depths respectively along the direction from the low hierarchy features to the high hierarchy features and the direction from the high hierarchy features to the low hierarchy features in sequence comprises:

along a direction from shallow to deep of the network depth of the neural network, after downsampling features at a lower hierarchy output by a network layer with a shallower network depth in the neural network in sequence, fusing the downsampled features with features at the higher hierarchy output by the network layer with the deeper network depth, to obtain a fourth fusion feature; and

along the direction from the high hierarchy features to the low hierarchy features, after upsampling fusion features at the higher hierarchy in sequence, fusing the upsampled fusion features with fusion features at a lower hierarchy in the fourth fusion feature,

wherein the features at the lower hierarchy comprise: features output by the network layer with a shallower network depth in the neural network or features obtained by performing at least one feature extraction on the features output by the network layer with the shallower network depth.

8. The method according to claim 7, wherein the after downsampling features at the lower hierarchy output by the network layer with the shallower network depth in the neural network in sequence, fusing the downsampled features with the features at the higher hierarchy output by the network layer with the deeper network depth comprises:

after downsampling the features at the lower hierarchy output by the network layer with the shallower network depth in the neural network in sequence, fusing the downsampled features with adjacent features at the higher hierarchy output by the network layer with the deeper network depth.

9. The method according to claim 7 or 8, wherein the after upsampling fusion features at the higher hierarchy in sequence, fusing the upsampled fusion features with the fusion features at the lower hierarchy in the fourth fusion feature comprises: after upsampling the fusion features at the higher hierarchy in sequence, fusing the upsampled fusion features with adjacent fusion features at the lower hierarchy in the fourth fusion feature.

10. The method according to any one of claims 1 to 9, wherein the fusing the region features corresponding to the same instance candidate region comprises: respectively performing pixel-level fusion on a plurality of region features corresponding to the same instance candidate region,

wherein the performing the pixel-level fusion on the plurality of region features corresponding to the same instance candidate region comprises: respectively performing element-wise max on the plurality of region features corresponding to the same instance candidate region; or respectively performing element-wise averaging on the plurality of region features corresponding to the same instance candidate region; or respectively performing element-wise summing on the plurality of region features corresponding to the same instance candidate region.

11. The method according to any one of claims 1 to 10, wherein the performing the instance segmentation based on the each first fusion feature, to obtain at least one of the instance segmentation result of the corresponding instance candidate region or the instance segmentation result of the image comprises at least one of:

performing the instance segmentation on the instance candidate region corresponding to a first fusion feature based on the first fusion feature, to obtain the instance segmentation result of the corresponding instance candidate region; or,

performing the instance segmentation on the image based on the each first fusion feature, to obtain the instance segmentation result of the image.

12. The method according to any one of claims 1 to 11, wherein the performing the instance segmentation based on the each first fusion feature, to obtain the instance segmentation result of the image comprises:

respectively performing the instance segmentation on the instance candidate region corresponding to the each first fusion feature based on the each first fusion feature, to obtain the instance segmentation result of each instance candidate region; and

obtaining the instance segmentation result of the image based on the instance segmentation result of the each instance candidate region.

13. The method according to claim 11 or 12, wherein the performing the instance segmentation on the instance candidate region corresponding to a first fusion feature based on the first fusion feature, to obtain the instance segmentation result of the corresponding instance candidate region comprises:

performing pixel-level instance class prediction based on the first fusion feature, to obtain an instance class prediction result of the instance candidate region corresponding to the first fusion feature; performing pixel-level foreground and background prediction based on the first fusion feature, to obtain a foreground and background prediction result of the instance candidate region corresponding to the first fusion feature; and

obtaining the instance segmentation result of an instance candidate region corresponding to the first fusion feature based on the instance class prediction result and the foreground and background prediction result.

14. The method according to claim 13, wherein the performing pixel-level foreground and background prediction based on the first fusion feature comprises at least one of: predicting pixels belonging to a foreground in the instance candidate region corresponding to the first fusion feature based on the first fusion feature; or,

predicting pixels belonging to a background in the instance candidate region corresponding to the first fusion feature based on the first fusion feature,

wherein the foreground comprises portions corresponding to all instance classes, and the background comprises portions other than the portions corresponding

to all instance classes; or the background comprises portions corresponding to all instance classes, and the foreground comprises portions other than the portions corresponding to all instance classes.

15. The method according to claim 13 or 14, wherein the performing pixel-level instance class prediction based on the first fusion feature comprises:

performing feature extraction on the first fusion feature via a first convolutional network, the first convolutional network comprising at least one fully convolutional layer; and

performing pixel-level object class prediction via a first fully convolutional layer based on a feature output by the first convolutional network.

16. The method according to any one of claims 13 to 15, wherein the performing pixel-level foreground and background prediction based on the first fusion feature comprises:

performing feature extraction on the first fusion feature via a second convolutional network, the second convolutional network comprising at least one fully convolutional layer; and

performing pixel-level foreground and background prediction via a full connection layer based on a feature output by the second convolutional network.

17. The method according to any one of claims 13 to 16, wherein the obtaining an instance segmentation result of an instance candidate region corresponding to the first fusion feature based on the instance class prediction result and the foreground and background prediction result comprises: performing pixel-level summing on the object class prediction result of the instance candidate region corresponding to the first fusion feature and the foreground and background prediction result of the instance candidate region corresponding to the first fusion feature, to obtain the instance segmentation result of the instance candidate region corresponding to the first fusion feature,

wherein after obtaining the foreground and background prediction result of the instance candidate region corresponding to the first fusion feature, the method further comprises: reshaping the foreground and background prediction result into a

foreground and background prediction result with a dimension consistent with that of the instance class prediction result; and

the performing pixel-level summing on the object class prediction result of the instance candidate region corresponding to the first fusion feature and the foreground and background prediction result of the instance candidate region corresponding to the first fusion feature comprising: performing pixel-level summing on the instance class prediction result of the instance candidate region corresponding to the first fusion feature and the reshaped foreground and background prediction result of the instance candidate region corresponding to the first fusion feature.

18. An instance segmentation apparatus, comprising:

a neural network, configured to perform feature extraction on an image to output features at at least two different hierarchies;

an extraction module, configured to extract region features corresponding to at least one instance candidate region in the image from the features at the at least two different hierarchies;

a first fusion module, configured to fuse region features corresponding to a same instance candidate region, to obtain a first fusion feature of each instance candidate region; and

a segmentation module, configured to perform instance segmentation based on each first fusion feature, to obtain at least one of an instance segmentation result of a corresponding instance candidate region or an instance segmentation result of the image.

19. An electronic device, comprising:

a memory, configured to store a computer program; and

a processor, configured to execute the computer program stored in the memory, wherein when the computer program is executed, the method according to any one of claims 1 to 17 is implemented.

20. A computer-readable storage medium having a computer program stored thereon, wherein when the computer program is executed by a processor, the method according to any one of claims 1 to 17 is implemented.