

CLAIMS

1. An image processing method, comprising:

determining a plurality of image feature maps of a target image that correspond to different preset scales;

determining, based on the plurality of image feature maps, a first probability that each pixel in the target image belongs to a foreground and a second probability that each pixel in the target image belongs to a background; and

performing panoramic segmentation on the target image based on the plurality of image feature maps, the first probability that each pixel in the target image belongs to the foreground, and the second probability that each pixel in the target image belongs to the background.

2. The method according to claim 1, wherein determining the plurality of image feature maps of the target image that correspond to different preset scales comprises:

performing feature extraction on the target image to obtain a first feature map for each of the different preset scales;

splicing the first feature map for each of the different preset scales to obtain a first spliced feature map;

extracting an image feature from the first spliced feature map to obtain a second feature map corresponding to a maximum preset scale in the different preset scales; and

determining, based on the first feature map for each of the different preset scales and the second feature map corresponding to the maximum preset scale, the plurality of image feature maps of the target image that correspond to the different preset scales.

3. The method according to claim 2, wherein determining, based on the first feature map for each of the different preset scales and the second feature map corresponding to the maximum preset scale, the plurality of image feature maps of the target image that correspond to the different preset scales comprises:

for each of the different preset scales except the maximum preset scale,

determining, based on the first feature map for a preset scale in the different preset scales that is adjacent to the preset scale and that is greater than the preset scale and the second feature map corresponding to the maximum preset scale, a second feature map corresponding to the preset scale; and

determining, based on the first feature map corresponding to the preset scale and the second feature map corresponding to the preset scale, an image feature map of the target image that corresponds to the preset scale.

4. The method according to claim 2, wherein splicing the first feature map for each of the different preset scales to obtain a first spliced feature map comprises:

for each of the different preset scales except the maximum preset scale, performing upsampling processing on the first feature map for the preset scale to obtain a first upsampled feature map; wherein a scale of each first upsampled feature map is the maximum preset scale; and

splicing the first feature map corresponding to the maximum preset scale and each first upsampled feature map to obtain the first spliced feature map.

5. The method according to any one of claims 1 to 4, wherein determining, based on the plurality of image feature maps, the first probability that each pixel in the target image belongs to the foreground and the second probability that each pixel in the target image belongs to the background comprises:

for each of the different preset scales except the maximum preset scale, performing upsampling processing on the image feature map for the preset scale to obtain an upsampled image feature map; wherein a scale of each upsampled image feature map is the maximum preset scale;

splicing the image feature map corresponding to the maximum preset scale and each upsampled image feature maps to obtain a second spliced feature map; and

determining, based on the second spliced feature map, the first probability that each pixel in the target image belongs to the foreground and the second probability that each pixel in the target image belongs to the background.

6. The method according to claim 5, wherein performing panoramic segmentation on the target image based on the plurality of image feature maps, the first probability that each pixel in the target image belongs to the foreground, and the second probability that each pixel in the target image belongs to the background comprises:

determining semantic segmentation logits according to the second spliced feature map and the second probability that each pixel in the target image belongs to the background; wherein a higher second probability that a pixel in the target image belongs to the background, a larger first scaling ratio corresponding to the pixel, and a first scaling ratio corresponding to a pixel in the target image is a ratio of a value corresponding to the pixel

in the semantic segmentation logits to a value corresponding to the pixel in the second spliced feature map;

determining an initial bounding box, an instance class, and instance segmentation logits of each object in the target image according to the second spliced feature map and the first probability that each pixel in the target image belongs to the foreground; wherein a higher first probability that a pixel in the target image belongs to the foreground, a larger second scaling ratio corresponding to the pixel, and a second scaling ratio corresponding to a pixel in the target image is a ratio of a value corresponding to the pixel in the instance segmentation logits to a value corresponding to the pixel in the second spliced feature map;

for each object, determining semantic segmentation logits corresponding to the object from the semantic segmentation logits according to the initial bounding box and the instance class of the object;

determining panoramic segmentation logits of the target image according to the semantic segmentation logits and the instance segmentation logits that correspond to each object; and

determining a bounding box and an instance class of an object in the background and the foreground of the target image according to the panoramic segmentation logits of the target image.

7. The method according to claim 6, wherein determining the semantic segmentation logits according to the second spliced feature map and the second probability that each pixel in the target image belongs to the background comprises:

determining a foreground-background classification feature map by using the first probability that each pixel in the target image belongs to the foreground and the second probability that each pixel in the target image belongs to the background;

extracting an image feature from the foreground-background classification feature map to obtain a feature map;

enhancing feature pixels in the feature map and corresponding to the background in the target image, and weakening feature pixels in the feature map and corresponding to the foreground in the target image, to obtain a first processed feature map;

fusing the first processed feature map with the second spliced feature map to obtain a fused feature map; and

determining the semantic segmentation logits based on the fused feature map.

8. The method according to claim 6, wherein determining an initial bounding box, an instance class, and instance segmentation logits of each object in the target image according to the second spliced feature map and the first probability that each pixel in the target image belongs to the foreground comprises:

determining a foreground-background classification feature map by using the first probability that each pixel in the target image belongs to the foreground and the second probability that each pixel in the target image belongs to the background;

extracting an image feature from the foreground-background classification feature map to obtain a feature map;

enhancing feature pixels in the feature map and corresponding to the foreground in the target image, and weakening feature pixels in the feature map and corresponding to the background in the target image, to obtain a second processed feature map;

fusing the second processed feature map with a region of interest corresponding to each object in the second spliced feature map to obtain a fused feature map; and

determining the initial bounding box of each object, the instance class of each object, and the instance segmentation logits of each object based on the fused feature map.

9. The method according to any one of claims 1 to 8, wherein the image processing method is performed by a neural network, the neural network is trained by using a sample image, and the sample image comprises an instance class annotated for an object and mask information annotated for the object.

10. The method according to claim 9, wherein the neural network is trained by using the following steps comprising:

determining a plurality of sample image feature maps of the sample image that correspond to the different preset scales, a first sample probability that each pixel in the sample image belongs to the foreground, and a second sample probability that each pixel in the sample image belongs to the background;

performing panoramic segmentation on the sample image according to the plurality of sample image feature maps, the first sample probability that each pixel in the sample image belongs to the foreground, and the second sample probability that each pixel in the sample image belongs to the background, to output an instance class and mask information of each object in the sample image;

determining a network loss function based on the mask information of each object in the sample image that is outputted by the neural network and mask information annotated for each object; and

adjusting a network parameter in the neural network by using the network loss function.

11. The method according to claim 10, wherein determining the network loss function based on the mask information of each object in the sample image that is outputted by the neural network and mask information annotated for each object comprises:

determining identical information between the mask information of each object in the sample image that is outputted by the neural network and the mask information annotated for each object, to obtain mask intersection information;

determining combined information between the mask information of each object in the sample image that is outputted by the neural network and the mask information annotated for each object, to obtain mask union information; and

determining the network loss function based on the mask intersection information and the mask union information.

12. An image processing apparatus, comprising:

a feature map determining module, configured to determine a plurality of image feature maps of a target image that correspond to different preset scales;

a foreground-background processing module, configured to determine, based on the plurality of image feature maps, a first probability that each pixel in the target image belongs to a foreground and a second probability that each pixel in the target image belongs to a background; and

a panoramic analysis module, configured to perform panoramic segmentation on the target image based on the plurality of image feature maps, the first probability that each pixel in the target image belongs to the foreground, and the second probability that each pixel in the target image belongs to the background.

13. The apparatus according to claim 12, wherein the feature map determining module is configured to:

perform feature extraction on the target image to obtain a first feature map for each of the different preset scales;

splice the first feature map for each of the different preset scales to obtain a first spliced feature map;

extract an image feature from the first spliced feature map to obtain a second feature map corresponding to a maximum preset scale in the different preset scales; and

determine, based on the first feature map for each of the different preset scales and the second feature map corresponding to the maximum preset scale, the plurality of image feature maps of the target image that correspond to the different preset scales.

14. The apparatus according to claim 13, wherein when determining, based on the first feature map for each of the different preset scales and the second feature map corresponding to the maximum preset scale, the plurality of image feature maps of the target image that correspond to the different preset scales, the feature map determining module is configured to:

for each of different preset scales except the maximum preset scale,

determine, based on the first feature map for a preset scale that is adjacent to the preset scale in the different preset scales and that is greater than the preset scale and the second feature map corresponding to the maximum preset scale, a second feature map corresponding to the preset scale; and

determine, based on the first feature map corresponding to the preset scale and the second feature map corresponding to the preset scale, an image feature map of the target image that corresponds to the preset scale.

15. The apparatus according to claim 13, wherein when splicing the first feature map for each of the different preset scales to obtain the first spliced feature map, the feature map determining module is configured to:

for each of the different preset scales except the maximum preset, perform upsampling processing on the first feature map for the preset scale, to obtain a first upsampled feature map; wherein a scale of each first upsampled feature map is the maximum preset scale; and

splice the first feature map corresponding to the maximum preset scale and each first upsampled feature map, to obtain the first spliced feature map.

16. The apparatus according to any one of claims 12 to 15, wherein the foreground-background processing module is configured to:

for each of the different preset scales except the maximum preset scale, perform upsampling processing on the image feature map for the preset scale, to obtain an upsampled image feature map; wherein a scale of each upsampled image feature map is the maximum preset scale;

splice the image feature map corresponding to the maximum preset scale and each unsampled image feature maps to obtain a second spliced feature map; and

determine, based on the second spliced feature map, the first probability that each pixel in the target image belongs to the foreground and the second probability that each pixel in the target image belongs to the background.

17. The apparatus according to claim 16, wherein the panoramic analysis module is configured to:

determine semantic segmentation logits according to the second spliced feature map and the second probability that each pixel in the target image belongs to the background; wherein a higher second probability that a pixel in the target image belongs to the background, a larger first scaling ratio corresponding to the pixel, and a first scaling ratio corresponding to a pixel in the target image is a ratio of a value corresponding to the pixel in the semantic segmentation logit to a value corresponding to the pixel in the second spliced feature map;

determine an initial bounding box, an instance class, and an instance segmentation logits of each object in the target image according to the second spliced feature map and the first probability that each pixel in the target image belongs to the foreground; wherein, a higher first probability that a pixel in the target image belongs to the foreground, a larger second scaling ratio corresponding to the pixel, and a second scaling ratio corresponding to a pixel in the target image is a ratio of a value corresponding to the pixel in the instance segmentation logits to a value corresponding to the pixel in the second spliced feature map;

for each object, determine semantic segmentation logits corresponding to the object from the semantic segmentation logits according to the initial bounding box and the instance class of the object;

determine panoramic segmentation logits of the target image according to the semantic segmentation logits and the instance segmentation logits that correspond to each object; and

determine a bounding box and an instance class of an object in the background and the foreground of the target image according to the panoramic segmentation logit of the target image.

18. An electronic device, comprising a processor, a storage medium, and a bus, the storage medium storing machine readable instructions executable by the processor, when the electronic device runs, the processor communicating with the storage medium by using the

bus, and the processor executing the machine readable instructions to perform the image processing method according to any one of claims 1 to 11.

19. A computer readable storage medium storing a computer program, wherein when run by a processor, the image processing method according to any one of claims 1 to 11 is performed.

20. A computer program stored in a storage medium, wherein when run by a processor, the image processing method according to any one of claims 1 to 11 is performed.