

CLAIMS

1. A method for recognizing a sequence in an image, comprising:

performing feature extraction on a to-be-processed image to obtain a first feature map of the to-be-processed image, wherein the to-be-processed image comprises a sequence formed by stacking a plurality of objects along a stacking direction;

determining a region feature of each segmented region of a plurality of segmented regions in the to-be-processed image based on the first feature map, wherein all segmented regions of the plurality of segmented regions are obtained by dividing the to-be-processed image into k regions along the stacking direction, k is a set number of objects stacked along the stacking direction, and k is an integer greater than 1; and

determining a category of each object in the sequence based on the region feature of each segmented region,

wherein the determining a category of each object in the sequence based on the region feature of each segmented region further comprises:

determining, based on the region feature of a first segmented region, a probability that an object in the first segmented region belongs to each of a plurality of object categories, wherein the first segmented region is any one of all the segmented regions, and

determining an object category with a maximum probability among the plurality of object categories as an object category to which the object in the first segmented region belongs.

2. The method according to claim 1, wherein the determining, based on the region feature of a first segmented region, a probability that an object in the first segmented region belongs to each of a plurality of object categories is implemented by a classification network, the classification network comprises a first convolution layer and a softmax layer, the number of convolution kernels in

the first convolution layer is equal to the number of set object categories, a size of the convolution kernels is $1 \times N$, N is a length of the region feature of each segmented region, and N is an integer greater than 1; and

the determining, based on the region feature of a first segmented region, a probability that an object in the first segmented region belongs to each of a plurality of object categories comprises:

performing convolution processing on the region feature of the first segmented region by using each convolution kernel in the first convolution layer respectively, to determine a category feature belonging to each of the plurality of object categories among features of the first segmented region; and

processing the category feature by using the softmax layer in the classification network, to determine a probability that the features of the first segmented region belong to each of the plurality of object categories.

3. The method according to claim 1, wherein the step of performing feature extraction on a to-be-processed image to obtain a first feature map of the to-be-processed image is implemented by a feature extraction network; the feature extraction network comprises a second convolution layer and a pooling layer; and

the performing feature extraction on a to-be-processed image to obtain a first feature map of the to-be-processed image comprises:

performing feature extraction on the to-be-processed image by using the second convolution layer, to obtain a second feature map; and

performing pooling processing on the second feature map by using the pooling layer, to obtain the first feature map, wherein

a feature dimension of the first feature map is less than a feature dimension of the second feature map.

4. The method according to claim 1, wherein the to-be-processed image comprises an image of a surface of an object constituting the sequence along the

stacking direction, and the stacking direction is a thickness direction of the object in the sequence.

5. The method according to claim 1, wherein at least one object among the plurality of objects in the sequence is a sheet-like object with a certain thickness.

6. The method according to claim 5, wherein a surface of the at least one object in the sequence along the stacking direction has a set identifier, and the identifier comprises at least one of a color, a texture, or a pattern.

7. The method according to claim 1, wherein the to-be-processed image is obtained by cropping the sequence from a collected image, and one end of the sequence is aligned with one edge of the to-be-processed image.

8. The method according to claim 7, wherein the plurality of objects in the sequence have a same thickness, and a length of the to-be-processed image along the stacking direction is k times the thickness of the objects in the sequence, or

the plurality of objects in the sequence have different thicknesses, and the length of the to-be-processed image along the stacking direction is k times a maximum thickness of an object of a possible category in the sequence.

9. The method according to claim 1, further comprising:

after the category of each object in the sequence is determined, determining a total value represented by the sequence based on a correspondence between the category and a value represented by the category.

10. An apparatus for recognizing a sequence in an image, comprising:

a processor; and

a memory configured to store processor executable instructions, wherein

the processor is configured to invoke the instructions stored in the memory,

to:

perform feature extraction on a to-be-processed image to obtain a first

feature map of the to-be-processed image, wherein the to-be-processed image comprises a sequence formed by stacking a plurality of objects along a stacking direction;

determine a region feature of each segmented region of a plurality of segmented regions in the to-be-processed image based on the first feature map, wherein all segmented regions of the plurality of segmented regions are obtained by dividing the to-be-processed image into k regions along the stacking direction, k is a set number of objects stacked along the stacking direction, and k is an integer greater than 1; and

determine a category of each object in the sequence based on the region feature of each segmented region, wherein the determining a category of each object in the sequence based on the region feature of each segmented region further comprises:

determining, based on the region feature of a first segmented region, a probability that an object in the first segmented region belongs to each of a plurality of object categories, wherein the first segmented region is any one of all the segmented regions, and

determining an object category with a maximum probability among the plurality of object categories as an object category to which the object in the first segmented region belongs.

11. The apparatus according to claim 10, wherein step of determining, based on the region feature of the first segmented region, the probability that an object in the first segmented region belongs to each of a plurality of object categories is implemented by a classification network, the classification network comprises a first convolution layer and a softmax layer, the number of convolution kernels in the first convolution layer is equal to the number of set object categories, a size of the convolution kernels is $1 \times N$, N is a length of the region feature of each segmented region, and N is an integer greater than 1; and

the determining, based on the region feature of the first segmented region, the probability that an object in the first segmented region belongs to each of a plurality

of object categories comprises:

performing convolution processing on the region feature of the first segmented region by using each convolution kernel in the first convolution layer respectively, to determine a category feature belonging to each set object category among features of the first segmented region; and

processing the category feature by using the softmax layer in the classification network, to determine a probability that the features of the first segmented region belong to each set object category.

12. The apparatus according to claim 10, wherein the step of performing feature extraction on the to-be-processed image to obtain the first feature map of the to-be-processed image is implemented by a feature extraction network; the feature extraction network comprises a second convolution layer and a pooling layer; and

performing feature extraction on the to-be-processed image to obtain the first feature map of the to-be-processed image comprises:

performing feature extraction on the to-be-processed image by using the second convolution layer, to obtain a second feature map; and

performing pooling processing on the second feature map by using the pooling layer, to obtain the first feature map, wherein

a feature dimension of the first feature map is less than a feature dimension of the second feature map.

13. The apparatus according to claim 10, wherein the to-be-processed image comprises an image of a surface of an object constituting the sequence along the stacking direction, and the stacking direction is a thickness direction of the object in the sequence.

14. The apparatus according to claim 10, wherein a surface of at least one object among the plurality of objects in the sequence along the stacking direction

has a set identifier, and the identifier comprises at least one of a color, a texture, or a pattern.

15. The apparatus according to claim 10, wherein the to-be-processed image is obtained by cropping the sequence from a collected image, and one end of the sequence is aligned with one edge of the to-be-processed image.

16. The apparatus according to claim 15, wherein the plurality of objects in the sequence have a same thickness, and a length of the to-be-processed image along the stacking direction is k times the thickness of the objects in the sequence, or

the plurality of objects in the sequence have different thicknesses, and the length of the to-be-processed image along the stacking direction is k times a maximum thickness of an object of a possible category in the sequence.

17. The apparatus according to claim 10, wherein the processor is further configured to:

after the category of each object in the sequence is determined, determine a total value represented by the sequence based on a correspondence between the category and a value represented by the category.

18. A non-transitory computer readable storage medium having computer program instructions stored thereon, wherein the computer program instructions upon execution by a processor cause the processor to perform operations, the operations comprising:

performing feature extraction on a to-be-processed image to obtain a first feature map of the to-be-processed image, wherein the to-be-processed image comprises a sequence formed by stacking a plurality of objects along a stacking direction;

determining a region feature of each segmented region of a plurality of segmented regions in the to-be-processed image based on the first feature map, wherein all segmented regions of the plurality of segmented regions are obtained by dividing the to-be-processed image into k regions along the stacking direction, k is a

set number of objects stacked along the stacking direction, and k is an integer greater than 1; and

determining a category of each object in the sequence based on the region feature of each segmented region, wherein the determining a category of each object in the sequence based on the region feature of each segmented region further comprises:

determining, based on the region feature of a first segmented region, a probability that an object in the first segmented region belongs to each of a plurality of object categories, wherein the first segmented region is any one of all the segmented regions, and

determining an object category with a maximum probability among the plurality of object categories as an object category to which the object in the first segmented region belongs.