

CLAIMS

1. A method for image processing, comprising:

obtaining an image to be processed, a first convolution kernel and a second convolution kernel, a receptive field of the first convolution kernel being different from a receptive field of the second convolution kernel;

obtaining a first feature image by performing a convolution process on the image to be processed using the first convolution kernel, and obtaining a second feature image by performing a convolution process on the image to be processed using the second convolution kernel; and

obtaining a first crowd density image by performing a fusion process on the first feature image and second feature image.

~~2. The method of claim 1,~~ wherein the method further comprises: before obtaining the first crowd density image by performing the fusion process on the first feature image and second feature image,

obtaining a first self-attention image by performing a first feature extraction process on the image to be processed, and obtaining a second self-attention image by performing a second feature extraction process on the image to be processed, each of the first self-attention image and second self-attention image being used for representing scale information of the image to be processed, and the scale information represented by the first self-attention image being different from the scale information represented by the second self-attention image; and

determining a first weight of the first feature image based on the first self-attention image, and determining a second weight of the second feature image based on the second self-attention image, wherein

obtaining the first crowd density image by performing the fusion process on the first

feature image and second feature image comprises:

obtaining the first crowd density image by performing the fusion process on the first feature image and second feature image based on the first weight and second weight.

32. The method of claim 21, wherein obtaining the first crowd density image by performing the fusion process on the first feature image and second feature image based on the first weight and second weight comprises:

obtaining a third feature image by determining a dot product of the first weight and first feature image;

obtaining a fourth feature image by determining a dot product of the second weight and second feature image; and

obtaining the first crowd density image by performing a fusion process on the third feature image and fourth feature image.

43. The method of claim 2-1 or 32, wherein determining the first weight of the first feature image based on the first self-attention image, and determining the second weight of the second feature image based on the second self-attention image comprises:

obtaining a third self-attention image corresponding to the first self-attention image and a fourth self-attention image corresponding to the second self-attention image by performing a normalization process on the first self-attention image and second self-attention image; and

taking the third self-attention image as the first weight and the fourth self-attention image as the second weight.

54. The method of any one of claims 2-1 to 43, wherein the method further comprises: before obtaining the first feature image by performing the convolution process on the image to be processed using the first convolution kernel, and obtaining the second feature image by performing the convolution process on the image to be processed using the second convolution kernel,

obtaining a fifth feature image by performing a third feature extraction process on the image to be processed, wherein

obtaining the first feature image by performing the convolution process on the image to be processed using the first convolution kernel, and obtaining the second feature image by performing the convolution process on the image to be processed using the second convolution kernel comprises:

obtaining the first feature image by performing a convolution process on the fifth feature image using the first convolution kernel, and obtaining the second feature image by performing a convolution process on the fifth feature image using the second convolution kernel; and

obtaining the first self-attention image by performing the first feature extraction process on the image to be processed, and obtaining the second self-attention image by performing the second feature extraction process on the image to be processed comprises:

obtaining the first self-attention image by performing the first feature extraction process on the fifth feature image, and obtaining the second self-attention image by performing the second feature extraction process on the fifth feature image.

65. The method of any one of claims 1-54, wherein each of the first convolution kernel and second convolution kernel is an atrous convolution kernel, a size of the first convolution kernel is the same as a size of the second convolution kernel, a weight of the first convolution kernel is the same as a weight of the second convolution kernel, and a dilation rate of the first convolution kernel is different from a dilation rate of the second convolution kernel.

76. The method of claim 65, wherein the dilation rate of the first convolution kernel or the dilation rate of the second convolution kernel is a reference value.

87. The method of any one of claims 1-76, further comprising: determining a sum of pixel values in the first crowd density image to obtain a number of people in the image to be processed.

98. The method of any one of claims 1-87, wherein the method is applied to a crowd counting network; and

a training process of the crowd counting network comprises:

obtaining a sample image,

obtaining a second crowd density image by processing the sample image using the crowd counting network,

obtaining a network loss based on a difference between the sample image and second crowd density image, and

adjusting a parameter of the crowd counting network based on the network loss.

109. The method of claim 98, wherein the method further comprises: before obtaining the network loss based on the difference between the sample image and second crowd density image,

obtaining a true crowd density image of the sample image, wherein

obtaining the network loss based on the difference between the sample image and second crowd density image comprises:

obtaining the network loss based on a difference between the true crowd density image and second crowd density image.

[4110](#). The method of claim [98](#), wherein the method further comprises: before obtaining the second crowd density image by processing the sample image using the crowd counting network,

pre-processing the sample image to obtain at least one pre-processed image, wherein

obtaining the second crowd density image by processing the sample image using the crowd counting network comprises:

obtaining at least one third crowd density image by processing the at least one pre-processed image using the crowd counting network, there being a one-to-one correspondence between the pre-processed image and third crowd density image; and

obtaining the network loss based on the difference between the sample image and second crowd density image comprises:

obtaining the network loss based on a difference between a target image in the at least one pre-processed image and a third crowd density image corresponding to the target image.

[4211](#). The method of claim [4110](#), wherein the pre-processing comprises at least one of: intercepting an image of a predetermined size from the sample image, or performing a flipping process on the sample image or the image of the predetermined size.

[4312](#). A device for image processing, comprising:

an obtaining unit, configured to obtain an image to be processed, a first convolution kernel, and a second convolution kernel, a receptive field of the first convolution kernel being

different from a receptive field of the second convolution kernel;

a convolution processing unit, configured to obtain a first feature image by performing a convolution process on the image to be processed using the first convolution kernel, and obtain a second feature image by performing a convolution process on the image to be processed using the second convolution kernel; and

a fusion processing unit, configured to obtain a first crowd density image by performing a fusion process on the first feature image and second feature image.

~~14. The device of claim 13,~~wherein the method further ~~comprising~~comprises:

a feature extraction processing unit, configured to: before obtaining the first crowd density image by performing the fusion process on the first feature image and second feature image, obtain a first self-attention image by performing a first feature extraction process on the image to be processed, and obtain a second self-attention image by performing a second feature extraction process on the image to be processed, each of the first self-attention image and second self-attention image being used for representing a scale information of the image to be processed, and the scale information represented by the first self-attention image being different from the scale information represented by the second self-attention image; and

a first determining unit, configured to determine a first weight of the first feature image based on the first self-attention image, and determine a second weight of the second feature image based on the second self-attention image, wherein

the fusion processing unit is configured to:

obtain the first crowd density image by performing the fusion process on the first feature image and second feature image based on the first weight and second weight.

~~15~~13. The device of claim ~~14~~12, wherein the fusion processing unit is specifically configured to:

obtain a third feature image by determining a dot product of the first weight and first

feature image;

obtain a fourth feature image by determining a dot product of the second weight and second feature image; and

obtain the first crowd density image by performing a fusion process on the third feature image and fourth feature image.

[1614](#). The device of claim [1412](#) or [1513](#), wherein the first determining unit is configured to:

obtain a third self-attention image corresponding to the first self-attention image and a fourth self-attention image corresponding to the second self-attention image by performing a normalization process on the first self-attention image and second self-attention image; and

take the third self-attention image as the first weight and the fourth self-attention image as the second weight.

[1715](#). The device of any one of claims [1412-1614](#), wherein the feature extraction processing unit is further configured to obtain a fifth feature image by performing a third feature extraction process on the image to be processed, before obtaining the first feature image by performing the convolution process on the image to be processed using the first convolution kernel and obtaining the second feature image by performing the convolution process on the image to be processed using the second convolution kernel;

the convolution processing unit is configured to:

obtain the first feature image by performing a convolution process on the fifth feature image using the first convolution kernel, and obtain the second feature image by performing a convolution process on the fifth feature image using the second convolution kernel, and

the feature extraction processing unit is further configured to:

obtain the first self-attention image by performing the first feature extraction process

on the fifth feature image, and obtain the second self-attention image by performing the second feature extraction process on the fifth feature image.

~~1816.~~ The device of any one of claims ~~13-17~~11-15, wherein each of the first convolution kernel and second convolution kernel is an atrous convolution kernel, a size of the first convolution kernel is the same as a size of the second convolution kernel, a weight of the first convolution kernel is the same as a weight of the second convolution kernel, and a dilation rate of the first convolution kernel is different from a dilation rate of the second convolution kernel.

~~1917.~~ The device of claim ~~1816~~1816, wherein the dilation rate of the first convolution kernel or the dilation rate of the second convolution kernel is a reference value.

~~20. The device of any one of claims 13-19, further comprising a second determining unit, configured to obtain a number of people in the image to be processed by determining a sum of pixel values in the first crowd density image.~~

~~21. The device of any one of claims 12-20, wherein a method for image processing performed by the device is applied to a crowd counting network;~~

~~the device further comprises: a training unit, configured to perform a training process on the crowd counting network; and the training process of the crowd counting network comprises:~~

~~obtaining a sample image,~~

~~processing the sample image using the crowd counting network to obtain a second crowd density image,~~

~~obtaining a network loss based on a difference between the sample image and second~~

~~crowd density image, and~~

~~adjusting a parameter of the crowd counting network based on the network loss.~~

~~22. The device of claim 21, wherein the training unit is further configured to:~~

~~obtain a true crowd density image of the sample image based on an impulse function, a gaussian kernel and the sample image, before obtaining the network loss based on the difference between the sample image and second crowd density image; and~~

~~obtain the network loss based on a difference between the true crowd density image and second crowd density image.~~

~~23. The device of claim 21, wherein the training unit is further configured to:~~

~~obtain at least one pre-processed image by pre-processing the sample image, before obtaining the second crowd density image by processing the sample image using the crowd counting network;~~

~~obtain at least one third crowd density image by processing the at least one pre-processed image using the crowd counting network, there being a one-to-one correspondence between the pre-processed image and third crowd density image; and~~

~~obtain the network loss based on a difference between a target image in the at least one pre-processed image and a third crowd density image corresponding to the target image.~~

~~24. The device of claim 23, wherein the pre-processing comprises at least one of: intercepting an image of a predetermined size from the sample image, or performing a flipping process on the sample image or the image of the predetermined size.~~

~~25. A processor, configured to perform the method of any one of claims 1-12.~~

~~26~~18. Electronic equipment, comprising a processor and memory connected to each other, the memory being configured to store computer program codes comprising computer instructions that when executed by the processor, cause the electronic equipment to perform the method of any one of claims 1-~~12~~11.

~~27~~19. A computer-readable storage medium having stored thereon a computer program comprising program instructions that when executed by a processor of electronic equipment, cause the processor to perform the method of any one of claims 1-~~12~~11.

~~28~~20. A computer program product comprising instructions, wherein the computer program product, when run on a computer, causes the computer to perform the method of any one of claims 1-~~12~~11.