

CLAIMS

1. A method for text recognition, comprising:

performing feature extraction processing on an image to be detected to obtain a plurality of semantic vectors, wherein each of the plurality of semantic vectors corresponds to one of multiple characters of a text sequence in the image to be detected; and

sequentially performing recognition processing on the plurality of semantic vectors through a convolutional neural network to obtain a recognition result of the text sequence,

wherein sequentially performing the recognition processing on the plurality of semantic vectors through the convolutional neural network to obtain the recognition result of the text sequence comprises:

processing priori information of a target semantic vector through the convolutional neural network to obtain a weight parameter of the target semantic vector, wherein the target semantic vector is one of the plurality of semantic vectors, and the priori information comprises at least one of: a text recognition result corresponding to a previous semantic vector of the target semantic vector, or a start character; and

determining a text recognition result corresponding to the target semantic vector according to the weight parameter and the target semantic vector.

2. The method of claim 1, wherein processing the priori information of the target semantic vector through the convolutional neural network to obtain the weight parameter of the target semantic vector comprises:

performing encoding processing on the target semantic vector through at least one first convolutional layer of the convolutional neural network to obtain a first vector of the target semantic vector;

performing encoding processing on the priori information of the target semantic vector through at least one second convolutional layer of the convolutional neural network to obtain a second vector corresponding to the priori information; and

determining the weight parameter based on the first vector and the second vector.

3. The method of claim 2, wherein performing the encoding processing on the priori information of the target semantic vector through the at least one second convolutional layer of

the convolutional neutral network to obtain the second vector corresponding to the priori information comprises:

responsive to that the priori information comprises the text recognition result corresponding to the previous semantic vector of the target semantic vector, performing word embedding processing on the text recognition result corresponding to the previous semantic vector to obtain a feature vector corresponding to the priori information; and

encoding the feature vector through the at least one second convolutional layer of the convolutional neutral network to obtain the second vector.

4. The method of claim 2 or 3, wherein performing the encoding processing on the priori information of the target semantic vector through the at least one second convolutional layer of the convolutional neutral network to obtain the second vector corresponding to the priori information comprises:

encoding an initial vector corresponding to the start character in the priori information through the at least one second convolutional layer of the convolutional neutral network to obtain the second vector.

5. The method of any one of claims 1 to 4, wherein determining the text recognition result corresponding to the target semantic vector according to the weight parameter and the target semantic vector comprises:

obtaining an attention distribution vector corresponding to the target semantic vector based on the weight parameter and the target semantic vector; and

decoding the attention distribution vector through at least one de-convolutional layer of the convolutional neutral network to determine the text recognition result corresponding to the target semantic vector.

6. The method of any one of claims 1 to 5, wherein performing the feature extraction processing on the image to be detected to obtain the plurality of semantic vectors comprises:

performing feature extraction on the image to be detected to obtain feature information; and

performing down-sampling processing on the feature information to obtain the plurality of semantic vectors.

7. A device for text recognition, comprising:

an extraction module, configured to perform feature extraction processing on an image to be detected to obtain a plurality of semantic vectors, wherein each of the plurality of semantic vectors corresponds to one of multiple characters of a text sequence in the image to be detected; and

a recognition module, configured to sequentially perform recognition processing on the plurality of semantic vectors through a convolutional neural network to obtain a recognition result of the text sequence,

wherein the recognition module is configured to:

process priori information of a target semantic vector through the convolutional neural network to obtain a weight parameter of the target semantic vector, wherein the target semantic vector is one of the plurality of semantic vectors, and the priori information comprises at least one of: a text recognition result corresponding to a previous semantic vector of the target semantic vector, or a start character; and

determine a text recognition result corresponding to the target semantic vector according to the weight parameter and the target semantic vector.

8. The device of claim 7, wherein the recognition module is configured to:

perform encoding processing on the target semantic vector through at least one first convolutional layer of the convolutional neural network to obtain a first vector of the target semantic vector;

perform encoding processing on the priori information of the target semantic vector through at least one second convolutional layer of the convolutional neural network to obtain a second vector corresponding to the priori information; and

determine the weight parameter based on the first vector and the second vector.

9. The device of claim 8, wherein the recognition module is configured to:

perform, responsive to that the priori information comprises the text recognition result corresponding to the previous semantic vector of the target semantic vector, word embedding processing on the text recognition result corresponding to the previous semantic vector to obtain a feature vector corresponding to the priori information; and

encode the feature vector through the at least one second convolutional layer of the convolutional neural network to obtain the second vector.

10. The device of claim 7 or 8, wherein the recognition module is configured to:
 encode an initial vector corresponding to the start character in the priori information through the at least one second convolutional layer of the convolutional neutral network to obtain the second vector.
11. The device of any one of claims 7 to 10, wherein the recognition module is configured to:
 obtain an attention distribution vector corresponding to the target semantic vector based on the weight parameter and the target semantic vector; and
 decode the attention distribution vector through at least one de-convolutional layer of the convolutional neutral network to determine the text recognition result corresponding to the target semantic vector.
12. The device of any one of claims 7 to 11, wherein the extraction module is configured to:
 perform feature extraction on the image to be detected to obtain feature information;
and
 perform down-sampling processing on the feature information to obtain the plurality of semantic vectors.
13. An electronic device, comprising:
 a processor; and
 a memory configured to store instructions executable by the processor,
 wherein the processor is configured to execute the instructions stored in the memory to perform the method of any one of claims 1 to 6.
14. A computer-readable storage medium, having stored thereon computer program instructions that, when executed by a processor, cause the processor to perform the method of any one of claims 1 to 6.