CLAIMS

1. An object sequence recognition method, comprising:

acquiring a first image comprising an object sequence;

inputting the first image to an object sequence recognition network and performing feature extraction to obtain a feature sequence, supervision information in a training process of the object sequence recognition network at least comprising class supervision information of each sample object in a sample object sequence and sequence length supervision information of the sample object of each class in a sample object sequence; and

predicting a class of the object sequence based on the feature sequence using a classifier of the object sequence recognition network to obtain class information of the object sequence;

wherein the inputting the first image to an object sequence recognition network and performing feature extraction to obtain a feature sequence comprises:

performing feature extraction on the first image using a convolutional subnetwork in the object sequence recognition network to obtain a feature map; and

splitting the feature map to obtain the feature sequence;

wherein the performing feature extraction on the first image using a convolutional subnetwork in the object sequence recognition network to obtain a feature map comprises:

down-sampling the first image using the convolutional subnetwork in a length dimension of the first image in a first direction to obtain a first-dimensional feature, the first direction being different from an arrangement direction of objects in the object sequence;

extracting a feature in a length dimension of the first image in a second direction based on a length of the first image in the second direction to obtain a second-dimensional feature, the second direction being the same as the arrangement direction of the objects in the object sequence; and

obtaining the feature map based on the first-dimensional feature and the second-dimensional feature.

2. The method of claim 1, wherein the splitting the feature map to obtain the feature sequence comprises:

pooling the feature map in the first direction to obtain a pooled feature map; and

splitting the pooled feature map in the second direction to obtain the feature sequence.

3. The method of claim 1 or 2, wherein the predicting a class of the object sequence based on the feature sequence using a classifier of the object sequence recognition network to obtain class information of the object sequence comprises:

predicting a class corresponding to each feature in the feature sequence using the classifier of the object sequence recognition network;

determining a class of each object in the object sequence based on a prediction result of the class corresponding to each feature in the feature sequence;

determining a sequence length of target features of objects belonging to a same class in the feature sequence; and

obtaining the class information of the object sequence based on the class of each object in the object sequence and a sequence length of target features corresponding to the objects of each class.

4. A method for training an object sequence recognition network, comprising:

acquiring a sample image, the sample image comprising a sample object sequence and class labeling information of the sample object sequence;

inputting the sample image to an object sequence recognition network to be trained

and performing feature extraction to obtain a sample feature sequence;

performing class prediction on sample objects in the sample object sequence based on the sample feature sequence using a classifier of the object sequence recognition network to be trained to obtain a class prediction result of the sample object sequence, the class prediction result of the sample object sequence comprising class prediction information of each sample object in the sample object sequence;

determining a first loss and a second loss based on the class prediction result of the sample object sequence, the first loss being configured to supervise the class prediction result of the sample object sequence based on the class labeling information of the sample object sequence, and the second loss being configured to supervise the number of sample objects of each class in the sample object sequence based on the class labeling information of the sample object sequence; and

adjusting a network parameter of the object sequence recognition network to be trained according to the first loss and the second loss to make a loss of a classification result output by an adjusted object sequence recognition network satisfy a convergence condition;

wherein the inputting the sample image to an object sequence recognition network to be trained and performing feature extraction to obtain a sample feature sequence comprises:

performing feature extraction on the sample image using a convolutional subnetwork in the object sequence recognition network to be trained to obtain a sample feature map of the sample image; and

splitting the sample feature map to obtain the sample feature sequence;

wherein the performing feature extraction on the sample image using a convolutional subnetwork in the object sequence recognition network to be trained to obtain a sample feature map of the sample image comprises:

down-sampling the sample image using the convolutional subnetwork in a length dimension of the sample image in a first direction to obtain a first-dimensional sample feature, the first direction being different from an arrangement direction of the sample object sequence in the sample sequence;

extracting a feature in a length dimension of the sample image in a second direction based on a length of the sample image in the second direction to obtain a second-dimensional sample feature, the second direction being the same as the arrangement direction of the sample object sequence in the sample sequence; and

obtaining the sample feature map of the sample image based on the first-dimensional sample feature and the second-dimensional sample feature.

5. The method of claim 4, wherein the splitting the sample feature map to obtain the sample feature sequence comprises:

pooling the sample feature map in the first direction to obtain a pooled sample feature map; and

splitting the pooled sample feature map in the second direction to obtain the sample feature sequence.

6. The method of claim 4 or 5, wherein the adjusting a network parameter of the object sequence recognition network to be trained according to the first loss and the second loss to make a loss of a classification result output by an adjusted object sequence recognition network satisfy a convergence condition comprises:

performing weighted fusion on the first loss and the second loss to obtain a total loss; and

adjusting the network parameter of the object sequence recognition network to be trained according to the total loss to make the loss of the classification result output by the adjusted object sequence recognition network satisfy the convergence condition.

7. The method of claim 6, wherein the performing weighted fusion on the first loss and the second loss to obtain a total loss comprises:

assigning a first dynamic weight to the first loss to obtain a first dynamic loss, the first dynamic weight gradually decreasing with the increase of a training count and/or training time of the object sequence recognition network to be trained when the training count reaches a first threshold or the training time reaches first time;

assigning a second dynamic weight to the second loss to obtain a second dynamic loss, the second dynamic weight gradually increasing with the increase of the training count and/or training time of the object sequence recognition network to be trained when the training count reaches a second threshold or the training time reaches second time; and

fusing the first dynamic loss and the second dynamic loss to obtain the total loss.

8. A computer storage medium, in which a computer-executable instruction is stored, wherein when executed by a processor, the computer-executable instruction is configured to:

acquire a first image comprising an object sequence;

input the first image to an object sequence recognition network and perform feature extraction to obtain a feature sequence, supervision information in a training process of the object sequence recognition network at least comprising class supervision information of each sample object in a sample object sequence and sequence length supervision information of the sample object of each class in the sample object sequence; and

predict a class of the object sequence based on the feature sequence using a classifier of the object sequence recognition network to obtain class information of the object sequence;

wherein when inputting the first image to an object sequence recognition network and perform feature extraction to obtain a feature sequence, the computer-executable instruction is configured to:

perform feature extraction on the first image using a convolutional subnetwork in the object sequence recognition network to obtain a feature map; and

split the feature map to obtain the feature sequence

wherein when performing feature extraction on the first image using a convolutional

subnetwork in the object sequence recognition network to obtain a feature map, the computer-executable instruction is configured to:

down-sample the first image using the convolutional subnetwork in a length dimension of the first image in a first direction to obtain a first-dimensional feature, the first direction being different from an arrangement direction of objects in the object sequence;

extract a feature in a length dimension of the first image in a second direction based on a length of the first image in the second direction to obtain a second-dimensional feature, the second direction being the same as the arrangement direction of the objects in the object sequence; and

obtain the feature map based on the first-dimensional feature and the second-dimensional feature.

9. A computer storage medium, in which a computer-executable instruction is stored, wherein when executed by a processor, the computer-executable instruction is configured to:

acquire a sample image, the sample image comprising a sample object sequence and class labeling information of the sample object sequence;

input the sample image to an object sequence recognition network to be trained and perform feature extraction to obtain a sample feature sequence;

perform class prediction on sample objects in the sample object sequence based on the sample feature sequence using a classifier of the object sequence recognition network to be trained to obtain a class prediction result of the sample object sequence, the class prediction result of the sample object sequence comprising class prediction information of each sample object in the sample object sequence;

determine a first loss and a second loss based on the class prediction result of the sample object sequence, the first loss being configured to supervise the class prediction result of the sample object sequence based on the class labeling information of the sample object sequence, and the second loss being configured to supervise the number of sample objects of each class in the sample object sequence based on the class labeling information of the sample

object sequence; and

adjust a network parameter of the object sequence recognition network to be trained according to the first loss and the second loss to make a loss of a classification result output by an adjusted object sequence recognition network satisfy a convergence condition;

wherein when inputting the sample image to an object sequence recognition network to be trained and perform feature extraction to obtain a sample feature sequence, the computer-executable instruction is configured to:

perform feature extraction on the sample image using a convolutional subnetwork in the object sequence recognition network to be trained to obtain a sample feature map of the sample image; and

split the sample feature map to obtain the sample feature sequence;

wherein when performing feature extraction on the sample image using a convolutional subnetwork in the object sequence recognition network to be trained to obtain a sample feature map of the sample image, the computer-executable instruction is configured to:

down-sample the sample image using the convolutional subnetwork in a length dimension of the sample image in a first direction to obtain a first-dimensional sample feature, the first direction being different from an arrangement direction of the sample object sequence in the sample sequence;

extract a feature in a length dimension of the sample image in a second direction based on a length of the sample image in the second direction to obtain a second-dimensional sample feature, the second direction being the same as the arrangement direction of the sample object sequence in the sample sequence; and

obtain the sample feature map of the sample image based on the first-dimensional sample feature and the second-dimensional sample feature.

10. A computer device, comprising a memory and a processor, wherein a computer-executable instruction is stored in the memory; wherein when executing the computer-executable instruction in the memory, the processor is configured to:

acquire a first image comprising an object sequence;

input the first image to an object sequence recognition network and perform feature extraction to obtain a feature sequence, supervision information in a training process of the object sequence recognition network at least comprising class supervision information of each sample object in a sample object sequence and sequence length supervision information of the sample object of each class in the sample object sequence; and

predict a class of the object sequence based on the feature sequence using a classifier of the object sequence recognition network to obtain class information of the object sequence;

wherein when inputting the first image to the object sequence recognition network and performing the feature extraction to obtain the feature sequence, the processor is configured to:

perform feature extraction on the first image using a convolutional subnetwork in the object sequence recognition network to obtain a feature map; and

split the feature map to obtain the feature sequence;

wherein when performing the feature extraction on the first image using the convolutional subnetwork in the object sequence recognition network to obtain the feature map, the processor is configured to:

down-sample the first image using the convolutional subnetwork in a length dimension of the first image in a first direction to obtain a first-dimensional feature, the first direction being different from an arrangement direction of objects in the object sequence;

extract a feature in a length dimension of the first image in a second direction based on a length of the first image in the second direction to obtain a second-dimensional feature, the second direction being the same as the arrangement direction of the objects in the object sequence; and

obtain the feature map based on the first-dimensional feature and the second-dimensional feature.

11. A computer device, comprising a memory and a processor, wherein a computer-

executable instruction is stored in the memory; wherein when executing the computer-executable instruction in the memory, the processor is configured to:

acquire a sample image, the sample image comprising a sample object sequence and class labeling information of the sample object sequence;

input the sample image to an object sequence recognition network to be trained and perform feature extraction to obtain a sample feature sequence;

perform class prediction on sample objects in the sample object sequence based on the sample feature sequence using a classifier of the object sequence recognition network to be trained to obtain a class prediction result of the sample object sequence, the class prediction result of the sample object sequence comprising class prediction information of each sample object in the sample object sequence;

determine a first loss and a second loss based on the class prediction result of the sample object sequence, the first loss being configured to supervise the class prediction result of the sample object sequence based on the class labeling information of the sample object sequence, and the second loss being configured to supervise the number of sample objects of each class in the sample object sequence based on the class labeling information of the sample object sequence; and

adjust a network parameter of the object sequence recognition network to be trained according to the first loss and the second loss to make a loss of a classification result output by an adjusted object sequence recognition network satisfy a convergence condition;

wherein when inputting the sample image to an object sequence recognition network to be trained and perform feature extraction to obtain a sample feature sequence, the processor configured to:

perform feature extraction on the sample image using a convolutional subnetwork in the object sequence recognition network to be trained to obtain a sample feature map of the sample image; and

split the sample feature map to obtain the sample feature sequence;

wherein when performing feature extraction on the sample image using a convolutional

subnetwork in the object sequence recognition network to be trained to obtain a sample feature map of the sample image, the processor configured to:

down-sample the sample image using the convolutional subnetwork in a length dimension of the sample image in a first direction to obtain a first-dimensional sample feature, the first direction being different from an arrangement direction of the sample object sequence in the sample sequence;

extract a feature in a length dimension of the sample image in a second direction based on a length of the sample image in the second direction to obtain a second-dimensional sample feature, the second direction being the same as the arrangement direction of the sample object sequence in the sample sequence; and

obtain the sample feature map of the sample image based on the first-dimensional sample feature and the second-dimensional sample feature.

12. A computer program, comprising computer instructions executable by an electronic device, wherein when executed by a processor in the electronic device, the computer instructions are configured to:

acquire a first image comprising an object sequence;

input the first image to an object sequence recognition network and perform feature extraction to obtain a feature sequence, supervision information in a training process of the object sequence recognition network at least comprising class supervision information of each sample object in a sample object sequence and sequence length supervision information of the sample object of each class in the sample object sequence; and

predict a class of the object sequence based on the feature sequence using a classifier of the object sequence recognition network to obtain class information of the object sequence;

wherein when inputting the first image to the object sequence recognition network and performing the feature extraction to obtain the feature sequence, the computer instructions are configured to:

perform feature extraction on the first image using a convolutional subnetwork in the

object sequence recognition network to obtain a feature map; and

split the feature map to obtain the feature sequence;

wherein when performing the feature extraction on the first image using the convolutional subnetwork in the object sequence recognition network to obtain the feature map, the computer instructions are configured to:

down-sample the first image using the convolutional subnetwork in a length dimension of the first image in a first direction to obtain a first-dimensional feature, the first direction being different from an arrangement direction of objects in the object sequence;

extract a feature in a length dimension of the first image in a second direction based on a length of the first image in the second direction to obtain a second-dimensional feature, the second direction being the same as the arrangement direction of the objects in the object sequence; and

obtain the feature map based on the first-dimensional feature and the second-dimensional feature.

13. A computer program, comprising computer instructions executable by an electronic device, wherein when executed by a processor in the electronic device, the computer instructions are configured to:

acquire a sample image, the sample image comprising a sample object sequence and class labeling information of the sample object sequence;

input the sample image to an object sequence recognition network to be trained and perform feature extraction to obtain a sample feature sequence;

perform class prediction on sample objects in the sample object sequence based on the sample feature sequence using a classifier of the object sequence recognition network to be trained to obtain a class prediction result of the sample object sequence, the class prediction result of the sample object sequence comprising class prediction information of each sample object in the sample object sequence;

determine a first loss and a second loss based on the class prediction result of the sample object sequence, the first loss being configured to supervise the class prediction result of the sample object sequence based on the class labeling information of the sample object sequence, and the second loss being configured to supervise the number of sample objects of each class in the sample object sequence based on the class labeling information of the sample object sequence; and

adjust a network parameter of the object sequence recognition network to be trained according to the first loss and the second loss to make a loss of a classification result output by an adjusted object sequence recognition network satisfy a convergence condition;

wherein when inputting the sample image to an object sequence recognition network to be trained and perform feature extraction to obtain a sample feature sequence, the computer instructions are configured to:

perform feature extraction on the sample image using a convolutional subnetwork in the object sequence recognition network to be trained to obtain a sample feature map of the sample image; and

split the sample feature map to obtain the sample feature sequence;

wherein when performing feature extraction on the sample image using a convolutional subnetwork in the object sequence recognition network to be trained to obtain a sample feature map of the sample image, the computer instructions are configured to:

down-sample the sample image using the convolutional subnetwork in a length dimension of the sample image in a first direction to obtain a first-dimensional sample feature, the first direction being different from an arrangement direction of the sample object sequence in the sample sequence;

extract a feature in a length dimension of the sample image in a second direction based on a length of the sample image in the second direction to obtain a second-dimensional sample feature, the second direction being the same as the arrangement direction of the sample object sequence in the sample sequence; and

obtain the sample feature map of the sample image based on the first-dimensional sample feature and the second-dimensional sample feature.