

CLAIMS

1. A system reinforcement learning method, comprising:

processing an input image based on a first network of a system to obtain a first result;

inputting the first result into a second network of the system to obtain a second result;

5 and

obtaining a reinforcement operation based on the second result by using a reinforcement network, adjusting the first result based on the reinforcement operation, and obtaining a target result based on the adjusted first result.

10 2. The method according to claim 1, wherein the obtaining the reinforcement operation based on the second result by using the reinforcement network, adjusting the first result based on the reinforcement operation, and obtaining the target result based on the adjusted first result comprises:

15 obtaining the reinforcement operation by using the reinforcement network based on the second result outputted by the second network, and adjusting the first result based on the reinforcement operation to obtain a first intermediate result;

inputting the first intermediate result into the second network, obtaining the second result based on the first intermediate result, and inputting the second result into the reinforcement network; and

20 outputting the second result as the target result in response to a preset condition being met.

3. The method according to claim 2, wherein the reinforcement operation comprises at least one adjustment action; and

25 the obtaining the reinforcement operation by using the reinforcement network based on the second result outputted by the second network, and adjusting the first result based on

the reinforcement operation to obtain the first intermediate result comprises:

obtaining at least one adjustment action probability based on the second result by using the reinforcement network, and determining the at least one adjustment action based on the at least one adjustment action probability; and

5 adjusting the first result based on the at least one adjustment action to obtain the first intermediate result.

4. The method according to claim 2, wherein the adjusting the first result by using the reinforcement network based on the second result outputted by the second network to obtain
10 the first intermediate result comprises:

obtaining at least one adjustment action probability based on the second result and the first result by using the reinforcement network, and determining at least one adjustment action based on the at least one adjustment action probability; and

 adjusting the first result based on the at least one adjustment action to obtain the first
15 intermediate result.

5. The method according to claim 4, wherein the obtaining the at least one adjustment action probability based on the second result and the first result by using the reinforcement network, and determining the at least one adjustment action based on the at least one adjustment action
20 probability comprises:

obtaining a connection result based on the second result and the first result; and

obtaining the at least one adjustment action probability based on the connection result by using the reinforcement network, and determining the at least one adjustment action based on the adjustment action probability.

25

6. The method according to any one of claims 1 to 5, further comprising: training the system

based on a sample image, wherein the sample image comprises an annotated sample target result,

wherein the training the system based on the sample image comprises:

5 processing the sample image by using the first network of the system to obtain a first sample result;

 inputting the first sample result into the second network of the system to obtain a second sample result;

 obtaining a reinforcement operation based on the second sample result by using the reinforcement network, adjusting the first sample result based on the reinforcement operation,
10 and obtaining a predicted target result based on the adjusted first sample result; and

 respectively adjusting parameters of the second network and the reinforcement network based on the predicted target result and the sample target result.

7. The method according to claim 6, wherein the obtaining the reinforcement operation based
15 on the second sample result by using the reinforcement network, adjusting the first sample result based on the reinforcement operation, and obtaining the predicted target result based on the adjusted first sample result comprises:

 obtaining the reinforcement operation based on the second sample result by means of the reinforcement network, and adjusting the first sample result based on the reinforcement
20 operation to obtain a second intermediate sample result;

 obtaining an intermediate predicted result by means of the second network based on the second intermediate sample result, and inputting the intermediate predicted result into the reinforcement network; and

 outputting the intermediate predicted result as the predicted target result in response to
25 meeting the preset condition,

wherein the adjusting the parameters of the reinforcement network based on the predicted

target result and the sample target result comprises:

determining a first loss based on the intermediate predicted result and the sample target result; and

5 adjusting the parameters of the reinforcement network based on at least one the first loss.

8. The method according to claim 7, wherein the adjusting the parameters of the reinforcement network based on the at least one first loss comprises:

10 adjusting the parameters of the reinforcement network based on the at least one first loss according to an order in which the at least one first loss is obtained.

9. The method according to claim 7 or 8, wherein the determining the first loss based on the intermediate predicted result and the sample target result comprises:

15 determining a current intermediate reward based on the intermediate predicted result and the sample target result;

determining a current predicted reward based on the intermediate predicted result by using a scoring network, wherein the scoring network and the reinforcement network are configured to share part of a network structure;

20 determining a next intermediate predicted result based on the intermediate predicted result by using the reinforcement network and the second network, and determining a next predicted reward based on the next intermediate predicted result by using the scoring network; and

determining the first loss based on the current intermediate reward, the current predicted reward, and the next predicted reward.

25

10. A system reinforcement learning apparatus, comprising:

a first network module, configured to process an input image to obtain a first result;

a second network module, configured to process the first result to obtain a second result;

5 a reinforcement network module, configured to obtain a reinforcement operation based on the second result; and

an executor, configured to adjust the first result based on the reinforcement operation and obtain a target result based on the adjusted first result.

10 11. The apparatus according to claim 10, wherein the reinforcement network module comprises:

an adjustment module, configured to adjust the first result based on the second result outputted by the second network module to obtain a first intermediate result; and

15 a result output module, configured to input the first intermediate result into the second network module, obtain the second result based on the first intermediate result, input the second result into the adjustment module, and output the second result as the target result in response to a preset condition being met.

20 12. The apparatus according to claim 11, wherein the adjustment module is configured to: obtain at least one adjustment action probability based on the second result by using the reinforcement network module and determine at least one adjustment action based on the at least one adjustment action probability; and adjust the first result based on the at least one adjustment action to obtain the first intermediate result.

25 13. The apparatus according to claim 11, wherein the reinforcement network module is configured to obtain at least one adjustment action probability based on the second result and

the first result, and determine the at least one adjustment action based on the adjustment action probability; and

the executor is configured to adjust the first result based on the at least one adjustment action to obtain the first intermediate result.

5

14. The apparatus according to claim 13, wherein the reinforcement network module is configured to: obtain a connection result based on the second result and the first result; and obtain the at least one adjustment action probability based on the connection result by using the reinforcement network, and determine the at least one adjustment action based on the at least one adjustment action probability.

10

15. The apparatus according to any one of claims 10 to 14, further comprising: a training module, configured to train a system comprising the first network, the second network, and the reinforcement network based on a sample image, wherein the sample image comprises an annotated sample target result,

15

wherein the training module is configured to: input the sample image into the first network module, the second network module, the reinforcement network module, and the executor to obtain a predicted target result; and respectively adjust parameters of the second network and the reinforcement network based on the predicted target result and the sample target result.

20

16. The apparatus according to claim 15, wherein the reinforcement module is configured to obtain a reinforcement operation based on a second sample result obtained by the second network module;

the executor is configured to adjust a first sample result obtained by the first network module based on the reinforcement operation to obtain a second intermediate sample result; and

25

the second network module is configured to obtain an intermediate predicted result based on the second intermediate sample result, input the intermediate predicted result into the reinforcement module, and output the intermediate predicted result as the predicted target result in response to meeting the preset condition,

5 wherein the training module comprises:

a first loss module, configured to determine a first loss based on the intermediate predicted result and the sample target result; and

a first parameter adjustment module, configured to adjust the parameters of the reinforcement network based on at least one the first loss.

10

17. The apparatus according to claim 16, wherein the first parameter adjustment module is configured to adjust the parameters of the reinforcement network based on the at least one first loss according to an order in which the at least one first loss is obtained,

wherein the first loss module comprises:

15 a reward determination module, configured to determine a current intermediate reward based on the intermediate predicted result and the sample target result; determine a current predicted reward based on the intermediate predicted result by using a scoring network, wherein the scoring network and the reinforcement network are configured to share part of a network structure; determine a next intermediate predicted result based on the intermediate
20 predicted result by using the reinforcement network and the second network; and determine a next predicted reward based on the next intermediate predicted result by means of the scoring network; and

a loss determination module, configured to determine the first loss based on the current intermediate reward, the current predicted reward, and the next predicted reward.

25

18. An electronic device, comprising: a memory configured to store executable instructions; and

a processor, configured to communicate with the memory to execute the executable instructions, to implement operations of the system reinforcement learning method according to any one of claims 1 to 9.

5 19. A computer storage medium, configured to store computer readable instructions, wherein the instructions, when being executed, implement operations of the system reinforcement learning method according to any one of claims 1 to 9.

10 20. A computer program, comprising computer readable codes, wherein the computer readable codes, when running on a device, cause a processor in the device to execute instructions for implementing the system reinforcement learning method according to any one of claims 1 to 9.