

CLAIMS

1. A method for generating a neural network model, comprising:

5 sending, by a first device, an indication about a structure of a subnetwork model to a second device, wherein the subnetwork model is determined by adjusting a structure of a hypernetwork model;

receiving, by the first device, a parameter of the subnetwork model from the second device, wherein the parameter of the subnetwork model is determined by the second device based on the indication and the hypernetwork model;

10 training, by the first device, the subnetwork model based on the received parameter of the subnetwork model; and

sending, by the first device, a parameter of the trained subnetwork model to the second device for the second device to update the hypernetwork model.

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

15 obtaining, by the first device, a preconfigured parameter of the hypernetwork model; and determining, by the first device, the structure of the subnetwork model based on the preconfigured parameter and by adjusting the structure of the hypernetwork model.

3. The method according to claim 2, wherein the obtaining a preconfigured parameter of the hypernetwork model comprises:

20 locally training, by the first device, the hypernetwork model to determine a local parameter of the hypernetwork model;

sending, by the first device, the local parameter to the second device; and

receiving, by the first device, the preconfigured parameter from the second device, wherein the preconfigured parameter is determined by the second device based on at least the local parameter received from the first device.

25 4. The method according to claim 2, wherein the determining the structure of the subnetwork model comprises:

initializing the subnetwork model to a hypernetwork model with the preconfigured parameter; and

30 iteratively updating the subnetwork model by performing the following operations at least once:

adjusting a structure of a plurality of layers of the subnetwork model to obtain a plurality of candidate network models;

selecting one candidate network model based on accuracy of the plurality of candidate network models to update the subnetwork model; and

if the subnetwork model meets a constraint of the first device, determining the structure of the subnetwork model.

5. The method according to claim 4, wherein the adjusting a structure of a plurality of layers comprises:

5 deleting parameters related to some nodes of one layer in the plurality of layers to obtain one of the plurality of candidate network models.

6. The method according to claim 5, further comprising:

determining the some nodes based on a predetermined quantity percentage.

7. The method according to claim 1, wherein the constraint comprises: a calculation amount
10 of the subnetwork model is less than a first threshold, or a quantity of parameters of the subnetwork model is less than a second threshold.

8. The method according to claim 7, wherein the first threshold and the second threshold are both associated with performance of the first device.

9. The method according to claim 1, wherein the indication is in a form of a mask, and the
15 mask indicates whether the subnetwork model has a corresponding parameter of the hypernetwork model.

10. A method for generating a neural network model, comprising:

receiving, by a second device, an indication about a structure of a subnetwork model from a
plurality of first devices, wherein the subnetwork model is determined by adjusting a structure of
20 a hypernetwork model;

determining, by the second device, a parameter of the subnetwork model based on the
indication and the hypernetwork model;

sending, by the second device, the parameter of the subnetwork model to the plurality of first
devices for the plurality of first devices to separately train the subnetwork model;

25 receiving, by the second device, a parameter of the trained subnetwork model from the
plurality of first devices; and

updating, by the second device, the hypernetwork model by using the received parameter.

11. The method according to claim 10, wherein the indication is in a form of a mask, and the
mask indicates whether the subnetwork model has a corresponding parameter of the hypernetwork
30 model.

12. The method according to claim 10, wherein the receiving a parameter of the trained
subnetwork model further comprises:

receiving, from the plurality of first devices, a change in the parameter of the trained
subnetwork model.

35 13. An apparatus for generating a neural network model, comprising:

a sending unit, configured to send an indication about a structure of a subnetwork model to a second device, wherein the subnetwork model is determined by adjusting a structure of a hypernetwork model;

5 a receiving unit, configured to receive a parameter of the subnetwork model from the second device, wherein the parameter of the subnetwork model is determined by the second device based on the indication and the hypernetwork model; and

a training unit, configured to train the subnetwork model based on the received parameter of the subnetwork model, wherein

10 the sending unit is further configured to send a parameter of the trained subnetwork model to the second device for the second device to update the hypernetwork model.

14. The apparatus according to claim 13, wherein the receiving unit is further configured to obtain a preconfigured parameter of the hypernetwork model, and the apparatus further comprises:

a model determining unit, configured to determine the structure of the subnetwork model based on the preconfigured parameter and by adjusting the structure of the hypernetwork model.

15 15. An apparatus for generating a neural network model, comprising:

a receiving unit, configured to receive an indication about a structure of a subnetwork model from a plurality of first devices, wherein the subnetwork model is determined by adjusting a structure of a hypernetwork model;

20 a unit for determining a parameter of a subnetwork model, configured to determine a parameter of the subnetwork model based on the indication and the hypernetwork model;

a sending unit, configured to send the parameter of the subnetwork model to the plurality of first devices for the plurality of first devices to separately train the subnetwork model; and

a hypernetwork update unit, wherein

25 the receiving unit is further configured to receive a parameter of the trained subnetwork model from the plurality of first devices; and

the hypernetwork update unit is configured to update the hypernetwork model by using the received parameter.