CLAIMS

1. A semantic segmentation model training method, comprising:

performing, by a semantic segmentation model, image semantic segmentation on at least one unlabeled image to obtain a preliminary semantic segmentation result as a category of the at least one unlabeled image;

obtaining, by a convolutional neural network based on the category of the at least one unlabeled image and a category of at least one labeled image, sub-images respectively corresponding to at least two images and features corresponding to the sub-images, wherein the at least two images comprise the at least one unlabeled image and the at least one labeled image, and the at least two sub-images carry the categories of the corresponding images; and

training the semantic segmentation model on the basis of the categories of the at least two sub-images and feature distances between the at least two sub-images.

2. The method according to claim 1, wherein the training the semantic segmentation model on the basis of the categories of the at least two sub-images and feature distances between the at least two sub-images comprises:

establishing a patch graph according to category relations between the sub-images, wherein the patch graph comprises nodes and an edge, wherein the nodes comprise the sub-images, and the edge comprises a feature distance between any two sub-images; and

training the semantic segmentation model to enable the feature distance between two sub-images of a same category in the patch graph to be lower than a first preset value, and the feature distance between two sub-images of different categories to be greater than a second preset value.

3. The method according to claim 2, wherein the establishing a patch graph according to category relations between the sub-images comprises:

selecting at least one sub-image as a reference node, and for at least one

reference node:

using a sub-image of a same category as the reference node as a positively correlated node, using a sub-image of a different category from the reference node as a negatively correlated node, and separately establishing a positive correlation connection between the reference node and at least one positively correlated node, and a negative correlation connection between the reference node and at least one negatively correlated node; and

forming a sparse connectivity patch graph according to the at least one reference node, the positively correlated node of the reference node, the negatively correlated node of the reference node, the positive correlation connection, and the negative correlation connection.

4. The method according to claim 2 or 3, wherein the training the semantic segmentation model comprises:

training the semantic segmentation model by a gradient back propagation algorithm, so as to minimize an error of the convolutional neural network, wherein the error is a triplet loss of the features of the corresponding sub-images obtained based on the convolutional neural network.

5. The method according to claim 4, wherein the training the semantic segmentation model by a gradient back propagation algorithm comprises:

calculating a maximum error by a loss function according to the feature distance between the sub-images in the established patch graph; performing gradient back propagation on the maximum error to calculate an error of at least one layer in the convolutional neural network;

calculating a gradient of parameters of at least one layer according to the error of the at least one layer, and modifying the parameters of a corresponding layer in the convolutional neural network according to the gradient;

calculating an error according to the distance between sub-images outputted by the convolutional neural network with the optimized parameters, and taking the

error as the maximum error;

iteratively implementing following operations until the maximum error is lower than or equal to a preset value:

performing gradient back propagation on the maximum error to calculate an error of at least one layer in the convolutional neural network; and

calculating the gradient of parameters of at least one layer according to the error of the at least one layer, and modifying the parameters of the corresponding layer in the convolutional neural network according to the gradient.

6. The method according to claim 4 or 5, wherein the training the semantic segmentation model comprises:

obtaining parameters of the convolutional neural network based on a training result of the convolutional neural network; and

initializing parameters in the semantic segmentation model based on the obtained parameters of the convolutional neural network.

7. The method according to any one of claims 1-6, wherein the obtaining, by a convolutional neural network based on the category of the at least one unlabeled image and a category of at least one labeled image, sub-images respectively corresponding to at least two images and features corresponding to the sub-images comprises:

in response to movement of a select box with a preset size on the at least two images, performing determination on pixels in the select box, and when a proportion of pixels of a same category among the pixels in the select box is greater than or equal to a preset value, outputting the image in the select box as a sub-image, and labeling the sub-image as said category; and

obtaining features corresponding to the sub-images by the convolutional neural network.

8. The method according to claim 7, further comprising: when the proportion of the

pixels of the same category among the pixels in the select box is less than the preset value, discarding the select box.

9. The method according to claim 7 or 8, wherein the obtaining features corresponding to the sub-images by the convolutional neural network comprises:

performing feature extraction respectively on the unlabeled image and the labeled image by the convolutional neural network, so as to obtain features maps respectively corresponding to the unlabeled image and the labeled image; and

obtaining, based on a position and size of the select box corresponding to the sub-image, the corresponding feature in the select box from the corresponding feature map, so as to determine the features corresponding to the sub-image.

10. The method according to any one of claims 1-9, further comprising: before performing, by the semantic segmentation model, image semantic segmentation on the at least one unlabeled image,

using stochastic gradient descent to train the semantic segmentation model until a preset convergence condition is satisfied.

11. The method according to any one of claims 1-10, further comprising: before obtaining, by the convolutional neural network based on the category of the at least one unlabeled image and a category of the at least one labeled image, the sub-images respectively corresponding to the at least two images and the features corresponding to the sub-images,

using stochastic gradient descent to train the convolutional neural network until a preset convergence condition is satisfied.

12. A semantic segmentation model training apparatus, comprising:

a segmentation unit, configured to perform, by a semantic segmentation model, image semantic segmentation on at least one unlabeled image to obtain a preliminary semantic segmentation result as a category of the at least one unlabeled image;

a sub-image extraction unit, configured to obtain, by a convolutional neural network based on the category of the at least one unlabeled image and a category of at least one labeled image, sub-images respectively corresponding to at least two images and features corresponding to the sub-images, wherein the at least two images comprise the at least one unlabeled image and the at least one labeled image, and the at least two sub-images carry the categories of the corresponding images; and

a training unit, configured to train the semantic segmentation model on the basis of the categories of the at least two sub-images and feature distances between the at least two sub-images.

13. The apparatus according to claim 12, wherein the training unit comprises:

a patch graph establishing module, configured to establish a patch graph according to category relations between the sub-images, wherein the patch graph comprises nodes and an edge, wherein the nodes comprise the sub-images, and the edge comprises a feature distance between any two sub-images; and

a model training module, configured to train the semantic segmentation model to enable the feature distance between two sub-images of a same category in the patch graph to be lower than a first preset value, and the feature distance between two sub-images of different categories to be greater than a second preset value.

14. The apparatus according to claim 13, wherein the patch graph establishing module comprises:

a reference selection module, configured to select at least one sub-image as a reference node;

a connection relation establishing module, configured to, for at least one reference node, use a sub-image of a same category as the reference node as a positively correlated node, use a sub-image of a different category from the reference node as a negatively correlated node, and separately establish a positive correlation connection between the reference node and at least one positively correlated node, and a negative correlation connection between the reference node and at least one

negatively correlated node; and

a connection map establishing module, configured to form a sparse connectivity patch graph according to the at least one reference node, the positively correlated node of the reference node, the negatively correlated node of the reference node, the positive correlation connection, and the negative correlation connection.

15. The apparatus according to claim 13 or 14, wherein the model training module comprises:

a network training module, configured to train the semantic segmentation model by a gradient back propagation algorithm, so as to minimize an error of the convolutional neural network, wherein the error is a triplet loss of the features of the corresponding sub-images obtained based on the convolutional neural network.

16. The apparatus according to claim 15, wherein the network training module is specifically configured to:

calculate a maximum error by a loss function according to the feature distance between the sub-images in the established patch graph;

perform gradient back propagation on the maximum error to calculate an error of at least one layer in the convolutional neural network;

calculate a gradient of parameters of at least one layer according to the error of the at least one layer, and modify the parameters of a corresponding layer in the convolutional neural network according to the gradient;

calculate an error according to the distance between sub-images outputted by the convolutional neural network with the optimized parameters, and take the error as the maximum error;

iteratively implement following operations until the maximum error is lower than or equal to a preset value:

performing gradient back propagation on the maximum error to calculate an error of at least one layer in the convolutional neural network; and calculating the gradient of parameters of at least one layer according to the error of the at least one

layer, and modify the parameters of the corresponding layer in the convolutional neural network according to the gradient.

17. The apparatus according to claim 15 or 16, wherein the model training module further comprises:

a segmentation model training module, configured to obtain parameters of the convolutional neural network based on a training result of the convolutional neural network; and initialize parameters in the semantic segmentation model based on the obtained parameters of the convolutional neural network.

18. The apparatus according to any one of claims 12-17, wherein the sub-image extraction unit is configured to, in response to movement of a select box with a preset size on the at least two images, perform determination on pixels in the select box, and when the proportion of pixels of the same category among the pixels in the select box is greater than or equal to a preset value, output the image in the select box as a sub-image, and label the sub-image as said category; and obtain features corresponding to the sub-images by the convolutional neural network.

19. A computer storage medium, configured to store computer-readable instructions, wherein the instructions are executed to implement operations of the semantic segmentation model training method according to any one of claims 1-11.

20. A computer program, comprising a computer-readable code, wherein when the computer-readable code is run on a device, a processor in the device executes instructions for implementing the operations of the semantic segmentation model training method according to any one of claims 1-11.