

IMAGE RECOGNITION MODEL GENERATION METHOD AND APPARATUS, COMPUTER DEVICE AND STORAGE MEDIUM

CROSS-REFERENCE TO RELATED APPLICATIONS

5 **[0001]** This application claims priority to Chinese Patent Application No. 202010862911.0, titled "METHOD AND APPARATUS FOR GENERATING IMAGE RECOGNITION MODEL, COMPUTER DEVICE AND STORAGE MEDIUM", filed on August 25, 2020 with the China National Intellectual Property Administration (CNIPA), which is incorporated herein by reference in its entirety.

10

FIELD

[0002] The present disclosure relates to a method and an apparatus for generating an image recognition model, a computer device and a storage medium.

15 **BACKGROUND**

[0003] In deep learning, image recognition technology has made great progresses. These progresses cannot be achieved without large-scale data sets, such as ImageNet and COCO. Generally, these large-scale data sets are balanced in terms of categories. But in an actual scenario, acquired data usually follows a long-tailed distribution, that is, a small amount of data belongs to a large quantity of categories, while a large amount of data belongs to a small quantity of categories.

20

SUMMARY

[0004] A method for generating an image recognition model is provided according to a first aspect of the present disclosure. The method includes:

25

acquiring a sample image set, the sample image set includes multiple sample image subsets, quantities of images included in the multiple sample image subsets decrease in turn, and each of the multiple sample image subsets has a same quantity of image categories;

training a to-be-trained image recognition model based on the sample image set, to

acquire a loss value of the to-be-trained image recognition model, the to-be-trained image recognition model includes multiple branchy neural networks, and each of the branchy neural networks is used to recognize a corresponding image, the loss value includes a target classification loss value and classification loss values respectively corresponding to the multiple branchy neural networks, the target classification loss value is a loss value of the to-be-trained image recognition model for the sample image set, and for each of the multiple branchy neural networks, the classification loss value corresponding to the branchy neural network is a loss value of the branchy neural network for sample image subsets corresponding to the branchy neural network; and

adjusting parameters of the to-be-trained image recognition model according to the loss value until the loss value is lower than a preset threshold, and determining the to-be-trained image recognition model when the loss value is lower than the preset threshold as a trained image recognition model.

[0005] An apparatus for generating an image recognition model is provided according to a second aspect of the present disclosure. The apparatus includes:

an acquisition module, configured to acquire a sample image set, the sample image set includes multiple sample image subsets, quantities of images included in the multiple sample image subsets decrease in turn, and each of the multiple sample image subsets includes a same quantity of image categories;

a training module, configured to train a to-be-trained image recognition model based on the sample image set, to acquire a loss value of the to-be-trained image recognition model, the to-be-trained image recognition model includes multiple branchy neural networks, and each of the branchy neural networks is used to recognize a corresponding image, the loss value includes a target classification loss value and classification loss values respectively corresponding to the multiple branchy neural networks, the target classification loss value is a loss value of the to-be-trained image recognition model for the sample image set, and for each of the multiple branchy neural networks, the classification loss value corresponding to the branchy neural network is a loss value of the branchy neural network for sample image subsets corresponding to the branchy neural network; and

an adjustment module, configured to adjust parameters of the to-be-trained image recognition model according to the loss value until the loss value is lower than a preset

threshold, and determine the to-be-trained image recognition model when the loss value is lower than the preset threshold as a trained image recognition model.

[0006] A computer device is provided according to a third aspect of the present disclosure. The computer device includes a memory and a processor. The memory stores a computer program. The processor, when executing the computer program, implements the following steps:

acquiring a sample image set, the sample image set includes multiple sample image subsets, quantities of images included in the multiple sample image subsets decrease in turn, and each of the multiple sample image subsets has a same quantity of image categories;

training a to-be-trained image recognition model based on the sample image set, to acquire a loss value of the to-be-trained image recognition model, the to-be-trained image recognition model includes multiple branchy neural networks, and each of the branchy neural networks is used to recognize a corresponding image, the loss value includes a target classification loss value and classification loss values respectively corresponding to the multiple branchy neural networks, the target classification loss value is a loss value of the to-be-trained image recognition model for the sample image set, and for each of the multiple branchy neural networks, the classification loss value corresponding to the branchy neural network is a loss value of the branchy neural network for sample image subsets corresponding to the branchy neural network; and

adjusting parameters of the to-be-trained image recognition model according to the loss value until the loss value is lower than a preset threshold, and determining the to-be-trained image recognition model when the loss value is lower than the preset threshold as a trained image recognition model.

[0007] A computer-readable storage medium is provided according to a fourth aspect of the present disclosure. The computer-readable storage medium stores a computer program. The computer program, when being executed by a processor, causes the following steps to be implemented:

acquiring a sample image set, the sample image set includes multiple sample image subsets, quantities of images included in the multiple sample image subsets decrease in turn, and each of the multiple sample image subsets has a same quantity of image categories;

training a to-be-trained image recognition model based on the sample image set, to acquire a loss value of the to-be-trained image recognition model, the to-be-trained image recognition model includes multiple branchy neural networks, and each of the branchy neural networks is used to recognize a corresponding image, the loss value includes a target
5 classification loss value and classification loss values respectively corresponding to the multiple branchy neural networks, the target classification loss value is a loss value of the to-be-trained image recognition model for the sample image set, and for each of the multiple branchy neural networks, the classification loss value corresponding to the branchy neural network is a loss value of the branchy neural network for sample image subsets corresponding
10 to the branchy neural network; and

adjusting parameters of the to-be-trained image recognition model according to the loss value until the loss value is lower than a preset threshold, and determining the to-be-trained image recognition model when the loss value is lower than the preset threshold as a trained image recognition model.

15 **[0008]** Details of one or more embodiments of the present disclosure are provided in the drawings and descriptions below. Other features and advantages of the present disclosure become apparent from the specification, drawings and claims.

BRIEF DESCRIPTION OF THE DRAWINGS

20 **[0009]** In order to illustrate the embodiments of the present disclosure or technical solutions in the conventional technology more clearly, drawings to be used in the description of the embodiments or the conventional technology are introduced simply hereinafter. It is apparent that the drawings described below only illustrate some embodiments of the present disclosure. For those skilled in the art, other drawings may be obtained according to these drawings
25 without any creative work.

[0010] Figure 1 is a diagram showing an application environment of a method for generating an image recognition model according to an embodiment of the present disclosure;

[0011] Figure 2 is a schematic flow chart of a method for generating an image recognition model according to an embodiment of the present disclosure;

[0012] Figure 3 is a schematic structural diagram of a branchy neural network according to an embodiment of the present disclosure;

[0013] Figure 4 is a schematic flow chart of steps of training a to-be-trained image recognition model to obtain a loss value according to an embodiment of the present disclosure;

[0014] Figure 5 is a schematic flow chart of steps of determining a loss value of a to-be-trained image recognition model according to an embodiment of the present disclosure;

[0015] Figure 6 is a schematic flow chart of a method for acquiring a sample image subset and a sample image set according to an embodiment of the present disclosure;

[0016] Figure 7 is a schematic structural diagram of an apparatus for generating an image recognition model according to an embodiment of the present disclosure; and

[0017] Figure 8 is a schematic diagram showing an internal structure of a computer device according to an embodiment of the present disclosure.

DETAILED DESCRIPTION

[0018] Data with a long-tailed distribution is used to train a neural network, such that the neural network can better recognize a small quantity of categories to which a large amount of image data belongs, and cannot accurately recognize a large quantity of categories to which a small amount of image data belongs. It can be seen that if the long-tailed distribution is ignored in the generation of image recognition model, the image recognition model has a greatly reduced performance in an actual practice. Therefore, an image recognition model acquired by a method for generating an image recognition model according to the conventional technology has a poor recognition effect.

[0019] In order to make the objectives, technical solutions and advantages of the present disclosure clearer, the present disclosure is described in detail in conjunction with the drawings and the embodiments hereinafter. It should be understood that the specific embodiments described herein are merely used for explaining the present disclosure, and are not intended to limit the present disclosure.

[0020] A method for generating an image recognition model according to the present

disclosure may be applied to an application environment shown in Figure 1. A terminal 11 communicates with a server 12 through a network. The server 12 acquires a sample image set sent by the terminal 11 through the network. The sample image set includes multiple sample image subsets. The quantities of images included in the multiple sample image subsets decrease in turn, and each of the multiple sample image subsets includes a same quantity of image categories. The server 12 trains a to-be-trained image recognition model based on the sample image set, to acquire a loss value of the to-be-trained image recognition model. The to-be-trained image recognition model includes multiple branchy neural networks. Each of the branchy neural networks is used to recognize a corresponding image. The loss value includes a target classification loss value and classification loss values respectively corresponding to the multiple branchy neural networks. The target classification loss value is a loss value of the to-be-trained image recognition model for the sample image set, and for each of the multiple branchy neutral networks, the classification loss value corresponding to the branchy neutral network is a loss value of the branchy neural network for sample image subsets corresponding to the branchy neural network. The server 12 adjusts parameters of the to-be-trained image recognition model according to the loss value until the loss value is lower than a preset threshold, and the to-be-trained image recognition model when the loss value is lower than the preset threshold is determined as a trained image recognition model. The terminal 11 may send a to-be-recognized image to the server 12, and acquire a recognition result returned by the server 12.

[0021] The terminal 11 may be, but is not limited to, various personal computers, laptops, smartphones, tablets and portable wearable devices. The server 12 may be implemented by an independent server or a server cluster formed by multiple servers.

[0022] As shown in Figure 2, a method for generating an image recognition model is provided according to an embodiment. The method is illustrated with an example that the method is applied to the server 12 in Figure 1, and the method includes the following steps 21 to 23.

[0023] In step 21, a sample image set is acquired. The sample image set includes multiple sample image subsets. The quantities of images included in the multiple sample image subsets decrease in turn, and each of the multiple sample image subsets has a same quantity of image categories.

[0024] The sample image set is a data set including all sample images, and is formed by multiple sample image subsets. Each of the sample image subsets includes one or more image categories of sample images, and the multiple sample image subsets have different image categories. Moreover, the total quantity of images in each of the multiple sample image subsets is different, and the quantities of images in the multiple sample image subsets decrease in turn.

[0025] For example, in the sample images, there are 100 sample images belonging to an image category A, 80 sample images belonging to an image category B, 60 sample images belonging to an image category C, 40 sample images belonging to an image category D, 20 sample images belonging to an image category E, and 10 sample images belonging to an image category F. The sample images belonging to the image category A and the image category B may form a sample image subset including 180 sample images. The sample images belonging to the image category C and the image category D may form a sample image subset including 100 sample images. The sample images belonging to the image category E and the image category F may form a sample image subset including 30 sample images. It can be seen that the quantities of images in the above three sample image subsets decrease in turn, and each of the three sample image subsets has a same quantity of image categories.

[0026] The server may directly acquire, from the terminal, the sample image set including multiple sample image subsets whose quantities of images decreases in turn. Alternatively, the server may acquire, from the terminal, a large amount of sample images, classify the sample images based on an image category corresponding to each of the sample images, to acquire a sample image set including multiple sample image subsets whose quantities of images decreases in turn. The sample image set may be formed by sample images that follow a long-tailed distribution, that is, a large amount of images belong to a small quantity of image categories, while a small amount of images belong to a large quantity of categories. Or, the sample image set may be formed by sample images that follow a normal distribution. The distribution feature of categories of sample images in the sample image set is not limited herein.

[0027] In this step 21, by acquiring the sample image set including multiple sample image subsets whose quantities of images decreases in turn, the sample images are pre-processed,

such that the sample images are sorted according to the image categories and belong to corresponding sample image subsets, so as to facilitate the subsequent feature learning of branchy neural networks. Moreover, an image category to which a small quantity of images belongs can be fully trained during the training process, thereby avoiding an ignorance of long-tailed data in the conventional training process of the neural network, and improving a generation effect of the image recognition model.

[0028] In step 22, the to-be-trained image recognition model is trained based on the sample image set, to acquire a loss value of the to-be-trained image recognition model. The to-be-trained image recognition model includes multiple branchy neural networks. Each of the branchy neural networks is used to recognize a corresponding image. The loss value includes a target classification loss value and classification loss values respectively corresponding to the multiple branchy neural networks. The target classification loss value is a loss value of the to-be-trained image recognition model for the sample image set, and for each of the branchy neural networks, the classification loss value corresponding to the branchy neural network is a loss value of the branchy neural network for sample image subsets corresponding to the branchy neural network.

[0029] The branchy neural networks are established through 1×1 convolutions, so that the establishment of the branchy neural networks can be completed with a few additional parameters. Multiple branchy neural networks are established in the to-be-trained image recognition model, and the branchy neural networks may divide parameters of the to-be-trained image recognition model into two parts. A part of the parameters are shared parameters for extracting a common feature of the sample images, and the other parameters are exclusive parameters for further extracting the sample images in sample image subsets corresponding to the branchy neural network based on the shared parameters. The exclusive parameters are corresponding parameters of a branchy neural network.

[0030] After the branchy neural networks are established, a correspondence between the branchy neural networks and the sample image subsets is determined. The correspondence may be determined according to the quantities of the branchy neural networks and the sample image subsets. Generally, there are three branchy neural networks and three sample image subsets, and it is specified that one of the branchy neural networks corresponds to three sample image subsets, the second branchy neural network corresponds to the last two sample

image subsets among the three sample image subsets, and the third branchy neural network corresponds to the last sample image subset among the three sample image subsets. The last sample image subset is a sample image subset including the minimum quantity of images.

[0031] For example, a sample image set includes three sample image subsets, which are respectively head classes (head data, abbreviated as H), medium classes (medium data, abbreviated as m) and tail classes (tail data, abbreviated as m). The head classes include the top third image categories with the maximum quantity of images, the medium classes include the middle third image categories with the middle quantity of images, and the tail classes include the last third image categories with the minimum quantity of images. Three branchy neural networks N_{h+m+t} , N_{m+t} and N_t are established through 1×1 convolutions, as shown in Figure 3. N_{h+m+t} corresponds to all the sample image subsets, and is used to classify images in all the sample image subsets according to image categories. N_{m+t} corresponds to two of the sample image subsets, and is used to classify, according to image categories, images in the sample image subsets of the medium classes and the tail classes that include the relatively small quantity of images. N_t corresponds to one of the sample image subsets, and is used to classify, according to image categories, images in the sample image subset of the tail classes that includes the minimum quantities of images. It can be seen that the three branchy neural networks N_{h+m+t} , N_{m+t} and N_t may lead the learning of image categories in sample image subsets corresponding to respective branchy neural networks through respective exclusive parameters. Moreover, the tail classes including a small quantity of images have a correspondence with each of the three branchy neural networks, while the head classes including a large quantity of images have a correspondence with only one of the branchy neural networks, which realizes the usage of the long-tailed data, so that the image categories with different quantities of images tend to be trained evenly.

[0032] The loss value of the to-be-trained image recognition model includes the classification loss value and the target classification loss value. For each of the branchy neural networks, a classification loss value corresponding to the branchy neural network is a loss value of the branchy neural network for sample image subsets corresponding to the branchy neural network. The target classification loss value is a loss value of the to-be-trained image recognition model for the sample image set. The loss value of the to-be-trained image recognition model may be adjusted. During training, the loss value of the to-be-trained image

recognition model may be acquired according to the multiple classification loss values and the target classification loss value, to determine the training degree of the entire image recognition model.

[0033] For each of the branchy neural networks, the classification loss value is a loss value of the branchy neural network for sample image subsets corresponding to the branchy neural network, that is, a loss value of the branchy neural network N_{h+m+t} for the sample image subsets of head classes, medium classes and tail classes, or a loss value of the branchy neural network N_t for the sample image subset of tail classes. The target classification loss value is a loss value of an image category outputted by the to-be-trained image recognition model for the entire sample image set, that is, a loss value of an image category acquired by fusing image categories outputted by multiple branchy neural networks recognizing the sample image set for the sample image set. A difference between the classification loss value and the target classification loss value lies in different objects considered in calculation of the loss value. The classification loss value is a loss value acquired by comparing an image category outputted by the branchy neural network with an actual image category of the sample image in a sample image subset corresponding to the branchy neural network. The target classification loss value is a loss value acquired by comparing an image category outputted by the entire to-be-trained image recognition model with an actual image category of the sample image in the sample image set. The image category outputted by the entire to-be-trained image recognition model is a result obtained by fusing image categories outputted by the multiple branchy neural networks.

[0034] In this step 22, images are recognized by the branchy neural networks, to acquire the target classification loss value and classification loss values respectively corresponding to the branchy neural networks, and the image recognition model is trained, to acquire the loss value of the to-be-trained image recognition model, so that an image category to which a small quantity of images belongs can be fully trained in the training process, thereby avoiding an ignorance of long-tailed data in the conventional training process of a neural network, and improving a generation effect of the image recognition model.

[0035] In step 23, parameters of the to-be-trained image recognition model are adjusted according to the loss value until the loss value is lower than a preset threshold, and the to-be-trained image recognition model when the loss value is lower than the preset threshold

is determined as a trained image recognition model.

[0036] The server adjusts, according to the calculated loss value, parameters of the to-be-trained image recognition model, including but not limited to parameters of a convolution layer, a pooling layer, a normalization layer and the like, such as a weight and a bias. Under normal conditions, after many training iterations, each loss value gradually decreases and approaches a constant value. The preset threshold may be set to be close to the constant value. In a case that the loss value is lower than the preset threshold, it is determined that the training of the image recognition model is completed.

[0037] In this step 23, the parameters of the image recognition model are continuously adjusted according to the loss value, and a training degree of the image recognition model is determined according to a difference between the loss value and the preset threshold. The parameters of the image recognition model are adjusted until the calculated loss value is lower than the preset threshold, and at this time it is determined that the training of the image recognition model is completed. In this way, the generation effect of the image recognition model is improved.

[0038] In the above method for generating an image recognition model, the sample image set is acquired. The sample image set includes multiple sample image subsets whose quantities of images decrease in turn, and each of the multiple sample image subsets has a same quantity of image categories. The to-be-trained image recognition model is trained based on the sample image set, to acquire a loss value of the to-be-trained image recognition model. The to-be-trained image recognition model includes multiple branchy neural networks, and each of the branchy neural networks is used to recognize a corresponding image. The loss value includes a target classification loss value and classification loss values respectively corresponding to the multiple branchy neural networks. The target classification loss value is a loss value of the to-be-trained image recognition model for the sample image set. For each of the branchy neural networks, the classification loss value is a loss value of the branchy neural network for sample image subsets corresponding to the branchy neural network. The parameters of the to-be-trained image recognition model are adjusted according to the loss value until the loss value is lower than a preset threshold. The to-be-trained image recognition model when the loss value is lower than the preset threshold is determined as a trained image recognition model. In the present disclosure, through setting multiple sample image subsets

whose quantities of images decrease in turn and the branchy neural networks for recognizing images of the corresponding sample image subsets, an image category to which a small quantity of images belongs can be fully trained in the training process, thereby avoiding an ignorance of long-tailed data in the conventional training process of a neural network, and
5 improving the generation effect of the image recognition model.

[0039] In an embodiment, as shown in Figure 4, step 22 of training the to-be-trained image recognition model based on the sample image set, to acquire the loss value of the to-be-trained image recognition model includes the following steps 41 to 43.

[0040] In step 41, the multiple sample image subsets in the sample image set are uniformly
10 sampled, to acquire a sample image input sequence.

[0041] In step 42, a sample image of the sample image input sequence is inputted into the to-be-trained image recognition model, to acquire an image category of the sample image.

[0042] In step 43, the loss value of the to-be-trained image recognition model is determined according to the image category of the sample image and an actual image category of the
15 sample image.

[0043] The multiple sample image subsets in the sample image set are uniformly sampled by the server, to acquire mini-batch data. The mini-batch data is taken as the sample image input sequence and inputted into the to-be-trained image recognition model for training, to acquire the image category of the sample image outputted by the image recognition model.
20 The actual image category of the sample image is acquired. The image category of the sample image and the actual image category are inputted into a preset loss function, to calculate the loss value of the image recognition model.

[0044] In the embodiment, through uniform sampling, the quantities of sample images of the respective image categories are relatively even in the sample image input sequence, so that
25 the determined loss value of the to-be-trained image recognition model is more accurate, and the generation effect of the image recognition model is improved.

[0045] In an embodiment, the to-be-trained image recognition model further includes a basic neural network. The basic neural network is connected to the branchy neural networks.

[0046] The above step 42 of inputting the sample image of the sample image input
30 sequence into the to-be-trained image recognition model, to acquire the image category of the

sample image is performed as follows. The sample image is inputted into the to-be-trained image recognition model, such that the basic neural network acquires a first image feature of the sample image, and the branchy neural networks acquire a second image feature of the sample image based on the first image feature and determine the image category of the sample image in the sample image set based on the second image feature.

[0047] The basic neural network is used to extract feature information of the sample image in the sample image set, that is, to extract a common feature of all image categories in the sample image set as the first image feature. The branchy neural networks acquire the first image feature extracted by the basic neural network, and then extract and output the second image feature. The second image feature outputted by the branchy neural networks is fused through a classifier to acquire the image category of the sample image. Parameters of the basic neural network are shared parameters, which may be used by each of the branchy neural networks. A type and a structure of the basic neural network are not limited herein.

[0048] In an embodiment, as shown in Figure 5, the above step 43 of determining the loss value of the to-be-trained image recognition model according to the image category of the sample image and the actual image category of the sample image includes the following steps 51 to 54.

[0049] In step 51, a loss value of the sample image in the sample image set is determined according to the image category of the sample image in the sample image set and the actual image category of the sample image.

[0050] In step 52, a loss value corresponding to the sample image set is acquired according to loss values of the sample image in the sample image set determined by the multiple branchy neural networks, and the loss value corresponding to the sample image set is determined as the target classification loss value.

[0051] In step 53, for each of the multiple branchy neural networks, loss values of all sample images in sample image subsets corresponding to the branchy neural network are acquired, and a sum of the loss values of all sample images in the sample image subset is determined as the classification loss value corresponding to the branchy neural network.

[0052] In step 54, the loss value of the to-be-trained image recognition model is calculated according to the target classification loss value and the classification loss values

corresponding to the multiple branchy neural networks.

[0053] For example, the sample image set includes three sample image subsets, that is, head classes, medium classes, and tail classes. The target classification loss value is acquired by comparing an image category outputted by the entire to-be-trained image recognition model with an actual image category of the sample image in the sample image set. The image category outputted by the entire to-be-trained image recognition model is a result acquired by fusing image categories outputted by the multiple branchy neural networks. Therefore, the target classification loss value is calculated as follows. Image categories corresponding to the sample image in the sample image set are outputted by the three branchy neural networks, and all the image categories and the actual image category are inputted into a loss function, to acquire a loss value. The acquired loss value is the target classification loss value and is calculated by the following equation:

$$L_f = J(F_{net}(X), Y), \text{ where } F_{net}(X) = N_{\bar{h}+\bar{m}+\bar{t}}(X) + N_{\bar{m}+\bar{t}}(X) + N_{\bar{t}}(X);$$

where, L_f represents the target classification loss value, J represents a cross entropy loss function, F_{net} represents the to-be-trained image recognition model, X represents an sample image in the sample image input sequence, Y represents an actual image category of the sample image, h , m and t respectively represent a first sample image subset, a second sample image subset and a third sample image subset whose quantities of images decreases in turn, $N_{\bar{h}+\bar{m}+\bar{t}}$, $N_{\bar{m}+\bar{t}}$, $N_{\bar{t}}$ respectively represent three branchy neural networks corresponding to the three sample image subsets, and a corner mark of each of the three branchy neural networks $N_{\bar{h}+\bar{m}+\bar{t}}$, $N_{\bar{m}+\bar{t}}$, $N_{\bar{t}}$ represents one or more sample image subsets corresponding to the branchy neural network.

[0054] For each of the branchy neural networks, the classification loss value is a loss value of the branchy neural network for sample image subsets corresponding to the branchy neural network, rather than for the whole sample image set. For example, since the branchy neural network $N_{\bar{h}+\bar{m}+\bar{t}}$ corresponds to the first sample image subset, the second sample image

subset and the third sample image subset, a loss value corresponding to the branchy neural network $N_{\bar{h+m+t}}$ is equivalent to a loss value of the branchy neural network $N_{\bar{h+m+t}}$ for the whole sample image set. Since the branchy neural network $N_{\bar{t}}$ simply corresponds to the third sample image subset, a loss value corresponding to the branchy neural network $N_{\bar{t}}$ is only required to be calculated according to an actual image category of an sample image in the third sample image subset. A sum of classification loss values calculated by all the branchy neural networks is acquired and determined as a final prediction result of the image category. The calculation is specifically as the following equation:

$$L_i = \sum_{x_i \in X} J(N_{\bar{h+m+t}}(x_i), y_i) + \sum_{x_i \in S_{m+t}} J(N_{\bar{m+t}}(x_i), y_i) + \sum_{x_i \in S_t} J(N_{\bar{t}}(x_i), y_i),$$

where, L_i represents the sum of the classification loss values corresponding to the multiple branchy neural networks, S_{m+t} represents a subset of X and includes sample images in the sample image input sequence which belong to the second sample image subset and the third sample image subset, and S_t represents another subset of X and includes sample images in the sample image input sequence which belong to the third sample image subset.

[0055] The loss value of the to-be-trained image recognition model is calculated by the classification loss values and the target classification loss value as the following equation:

$$L_{all} = (1-\alpha)L_f / n_1 + \alpha L_i / n_2,$$

where, L_{all} represents the loss value of the to-be-trained image recognition model, α represents a hyper-parameter, n_1 represents the quantity of the sample images in X , n_2 represents a total quantity of the sample images in X , S_{m+t} and S_t .

[0056] It should be noted that the hyper-parameter α in the equation of L_{all} may be adjusted for data sets with different long-tailed degrees. Moreover, in a case that a data set is in a normal distribution state, that is, each of image categories has almost the same quantity of images, the hyper-parameter α is set to zero for normal operation.

[0057] According to the above embodiment, the target classification loss value and the classification loss values are calculated according to a difference between the image category

of the sample image and the actual image category of the sample image, and the loss value of the to-be-trained image recognition model is further acquired, so as to adjust the parameters of the to-be-trained image recognition model, such that an image category including a small quantity of images can be fully trained in the training process, avoiding an ignorance of long-tailed data in the conventional training process of a neural network, and improving a generation effect of the image recognition model.

[0058] In an embodiment, as shown in Figure 6, before the sample image set is acquired in step 21, the method further includes the following steps 61 to 63.

[0059] In step 61, sample images are acquired, and a quantity of images in each of the image categories is determined according to image categories of the sample images.

[0060] In step 62, a sort order of the image categories is acquired according to the quantity of images in each of the image categories, and the image categories are classified into multiple category combinations according to the sort order. Each of the multiple category combinations includes a same quantity of image categories.

[0061] In step 63, sample image subsets corresponding to the multiple category combinations are acquired according to the multiple category combinations and sample images corresponding to the image categories in the multiple category combinations, and multiple sample image subsets are combined into the sample image set.

[0062] The server first acquires sample images from the terminal, identifies image categories of the sample images, classifies the sample images into corresponding image categories, and calculates the quantity of sample images corresponding to each of the image categories. The image categories are sorted in a descending order according to the quantity of sample images corresponding to each of the image categories, to acquire a sort order. The image categories are equally classified into multiple category combinations according to the quantity of the branchy neural networks and the quantity of the image categories. For example, the quantity of the branchy neural networks is three, and the quantity of the image categories is six. Two image categories may be classified into one group, so as to acquire three category combinations. The sample image subsets corresponding to the category combinations are acquired according to the category combinations and the sample images corresponding to each of the category combinations, and the multiple sample image subsets are combined into the sample image set.

[0063] According to the embodiment, image categories are sorted in a descending order or an ascending order according to the quantity of sample images corresponding to each of the image categories, and are equally classified according to the sort order, to acquire sample image subsets each including the same quantity of image categories, thus to realize pre-processing of the sample images. Moreover, each of the branchy neural networks corresponds to one or more sample image subsets according to the long-tailed data distribution feature, so that an image category including a small quantity of images can be fully trained in the training process, thereby avoiding an ignorance of long-tailed data in the conventional training process of a neural network, and improving a generation effect of the image recognition model.

[0064] It is to be understood that, although steps in the flow charts of Figure 2, Figures 4 to 6 are sequentially showed according to an indication of arrows, the steps are not necessarily sequentially performed in the sequence indicated by the arrows. Unless otherwise explicitly specified in this specification, execution of the steps is not strictly limited in a certain order, and the steps may be performed in other sequences. In addition, at least some steps in Figure 2, Figures 4 to 6 may include multiple steps or multiple stages. The steps or the stages are not necessarily performed at the same moment, and instead may be performed at different moments. The steps or the stages are not necessarily performed sequentially, and instead may be performed in turn or alternately with other step or at least some of steps or stages of other step.

[0065] In an embodiment, as shown in Figure 7, an apparatus for generating an image recognition model is provided. The apparatus includes an acquisition module 71, a training module 72 and an adjustment module 73.

[0066] The acquisition module 71 is configured to acquire a sample image set. The sample image set includes multiple sample image subsets whose quantities of images decrease in turn, and each of the multiple sample image subsets has a same quantity of image categories.

[0067] The training module 72 is configured to train a to-be-trained image recognition model based on the sample image set, to acquire a loss value of the to-be-trained image recognition model. The to-be-trained image recognition model includes multiple branchy neural networks, and each of the branchy neural networks is used to recognize a corresponding image. The loss value includes a target classification loss value and

classification loss values respectively corresponding to the multiple branchy neural networks. The target classification loss value is a loss value of the to-be-trained image recognition model for the sample image set. For each of the multiple branchy neural networks, the classification loss value corresponding to the branchy neural network is a loss value of the branchy neural network for sample image subsets corresponding to the branchy neural network.

[0068] The adjustment module 73 is configured to adjust parameters of the to-be-trained image recognition model according to the loss value until the loss value is lower than a preset threshold, and determine the to-be-trained image recognition model when the loss value is lower than the preset threshold as a trained image recognition model.

[0069] In an embodiment, the training module 72 is further configured to uniformly sample multiple sample image subsets in the sample image set, to acquire a sample image input sequence; input a sample image of the sample image input sequence into the to-be-trained image recognition model, to acquire an image category of the sample image; and determine the loss value of the to-be-trained image recognition model according to the image category of the sample image and an actual image category of the sample image.

[0070] In an embodiment, the training module 72 is further configured to input the sample image into the to-be-trained image recognition model, such that a basic neural network acquires a first image feature of the sample image, and the branchy neural networks acquire a second image feature of the sample image based on the first image feature and determine the image category of the sample image in the sample image set based on the second image feature.

[0071] In an embodiment, the training module 72 is further configured to determine a loss value of the sample image in the sample image set according to the image category of the sample image in the sample image set and the actual image category of the sample image; acquire a loss value corresponding to the sample image set according to loss values of the sample image in the sample image set determined by the multiple branchy neural networks, and determine the loss value corresponding to the sample image set as the target classification loss value; for each of the multiple branchy neural networks, acquire loss values of all sample images in the sample image subsets corresponding to the branchy neural network, and determine a sum of the loss values of all sample images in the sample image subset as the

classification loss value corresponding to the branchy neural network; and calculate the loss value of the to-be-trained image recognition model according to the target classification loss value and classification loss values corresponding to the multiple branchy neural networks.

[0072] In an embodiment, the acquisition module 71 is further configured to acquire sample images, and determine a quantity of images in each of the image categories according to the image categories of the sample images; acquire a sort order of the image categories according to the quantity of images in each of the image categories, and classify the image categories into multiple category combinations according to the sort order. Each of the multiple category combinations includes a same quantity of image categories. The acquisition module 71 is further configured to acquire sample image subsets corresponding to the multiple category combinations according to the multiple category combinations and the sample images corresponding to the image categories in the multiple category combinations; and combine multiple sample image subsets into the sample image set.

[0073] For the specific description of the apparatus for generating an image recognition model, reference may be made to the above description of the method for generating an image recognition model, which is not repeated herein. The modules in the above apparatus for generating an image recognition model may be implemented in whole or in part by software, hardware, or a combination thereof. Each of the above modules may be embedded in or independent of the processor in the computer device in a hardware form, or may be stored in a memory in the computer device in a software form, so as to be called by the processor to perform the operations corresponding to the above modules.

[0074] A computer device is provided according to an embodiment. The computer device may be a server, and an internal structure of the computer device may be shown in Figure 8. The computer device includes a processor, a memory, and a network interface that are connected with a system bus. The processor of the computer device is used to provide computing and control capabilities. The memory of the computer device includes a non-volatile storage medium and an internal memory. The non-volatile storage medium stores an operating system, a computer program, and a database. The internal memory provides an environment for the operation of the operating system and the computer program in the non-volatile storage medium. The database of the computer device is used to store data generated by the image recognition model. The network interface of the computer device is

used to communicate with an external terminal through network connection. The computer program, when being executed by the processor, implements the method for generating an image recognition model.

[0075] Those skilled in the art may understand that the structure shown in Figure 8 is only a block diagram of a partial structure related to the solution of this disclosure, and is not considered as a limitation to the computer device to which the solution of this disclosure is applied. Specifically, the computer device may include more or fewer components than those shown in Figure 8, or may combine with some components, or may arrange a different component.

[0076] A computer device is provided according to an embodiment. The computer device includes a memory and a processor. The memory stores a computer program, and the processor, when executing the computer program, implements steps of the method according to the above embodiments.

[0077] A computer-readable storage medium is provided according to an embodiment. The computer-readable storage medium stores a computer program. The computer program, when being executed by a processor, implements steps of the method according to the above embodiments.

[0078] Those skilled in the art may understand that all or some of the procedures of the method in the foregoing embodiments may be implemented by a computer program instructing relevant hardware. The computer program may be stored in a non-volatile computer-readable storage medium. When the computer program is executed, the procedures of the foregoing method embodiments may be implemented. Any reference to a memory, a storage, a database, or other medium used in the embodiments provided in this disclosure may include at least one of a non-volatile and volatile memory. The non-volatile memory may include a read-only memory (ROM), a magnetic tape, a floppy disk, a flash memory or an optical memory. The volatile memory may include a random access memory (RAM), or an external cache memory. As an illustration rather than a limitation, RAM may be in multiple forms, such as a static random access memory (SRAM) or a dynamic random access memory (DRAM).

[0079] The technical features in the foregoing embodiments may be randomly combined. For concise description, not all possible combinations of the technical features in the above

embodiments are described. However, the combinations of the technical features are to be considered as falling within the scope recorded in this specification provided that they do not conflict with each other.

[0080] The foregoing embodiments only show several implementations of the present disclosure and are described in detail, but they are not to be construed as a limit to the patent scope of the present disclosure. It should be noted that those skilled in the art may further make variations and improvements without departing from the ideas of the present disclosure, and all the variations and improvements fall within the protection scope of the present disclosure. Therefore, the protection scope of this present disclosure shall be subject to the protection scope of the appended claims.

CLAIMS

1. A method for generating an image recognition model, comprising:

acquiring a sample image set, wherein the sample image set comprises a plurality of
5 sample image subsets, wherein quantities of images comprised in the plurality of sample
image subsets decrease in turn, and each of the plurality of sample image subsets has a same
quantity of image categories;

training a to-be-trained image recognition model based on the sample image set, to
acquire a loss value of the to-be-trained image recognition model, wherein the to-be-trained
10 image recognition model comprises a plurality of branchy neural networks, and each of the
branchy neural networks is used to recognize a corresponding image, the loss value comprises
a target classification loss value and classification loss values respectively corresponding to
the plurality of branchy neural networks, the target classification loss value is a loss value of
the to-be-trained image recognition model for the sample image set, and for each of the
15 plurality of branchy neural networks, the classification loss value corresponding to the
branchy neural network is a loss value of the branchy neural network for sample image
subsets corresponding to the branchy neural network; and

adjusting parameters of the to-be-trained image recognition model according to the loss
value until the loss value is lower than a preset threshold, and determining the to-be-trained
20 image recognition model when the loss value is lower than the preset threshold as a trained
image recognition model.

2. The method according to claim 1, wherein the training the to-be-trained image
recognition model based on the sample image set, to acquire the loss value of the
to-be-trained image recognition model comprises:

25 uniformly sampling the plurality of sample image subsets in the sample image set, to
acquire a sample image input sequence;

inputting a sample image of the sample image input sequence into the to-be-trained
image recognition model, to acquire an image category of the sample image; and

determining the loss value of the to-be-trained image recognition model according to the
30 image category of the sample image and an actual image category of the sample image.

3. The method according to claim 2, wherein the to-be-trained image recognition model further comprises a basic neural network, the basic neural network is connected to the branchy neural networks,

wherein the inputting the sample image of the sample image input sequence into the to-be-trained image recognition model, to acquire the image category of the sample image comprises:

inputting the sample image into the to-be-trained image recognition model, such that the basic neural network acquires a first image feature of the sample image, and the branchy neural networks acquire a second image feature of the sample image based on the first image feature and determine the image category of the sample image in the sample image set based on the second image feature.

4. The method according to claim 3, wherein the determining the loss value of the to-be-trained image recognition model according to the image category of the sample image and the actual image category of the sample image comprises:

determining a loss value of the sample image in the sample image set according to the image category of the sample image in the sample image set and the actual image category of the sample image;

acquiring, according to loss values of the sample image in the sample image set determined by the plurality of branchy neural networks, a loss value corresponding to the sample image set, and determining the loss value corresponding to the sample image set as the target classification loss value;

for each of the plurality of branchy neural networks, acquiring loss values of all sample images in sample image subsets corresponding to the branchy neural network, and determining a sum of the loss values of all sample images in the sample image subset as the classification loss value corresponding to the branchy neural network; and

calculating the loss value of the to-be-trained image recognition model according to the target classification loss value and the classification loss values corresponding to the plurality of branchy neural networks.

5. The method according to claim 4, wherein the sample image set comprises three sample image subsets whose quantities of images decrease in turn, and the to-be-trained

image recognition model comprises three branchy neural networks, wherein

the target classification loss value is calculated by the following equation:

$$L_f = J(F_{net}(X), Y), \text{ where } F_{net}(X) = N_{h+m+t}^{\equiv}(X) + N_{m+t}^{\equiv}(X) + N_t^{\equiv}(X);$$

wherein, L_f represents the target classification loss value, J represents a cross entropy loss function, F_{net} represents the to-be-trained image recognition model, X represents the sample image in the sample image input sequence, Y represents the actual image category of the sample image, h , m and t respectively represent a first sample image subset, a second sample image subset and a third sample image subset whose quantities of images decrease in turn, $N_{h+m+t}^{\equiv}$, $N_{m+t}^{\equiv}$, $N_t^{\equiv}$ respectively represent three branchy neural networks corresponding to the three sample image subsets, and a corner mark of each of the $N_{h+m+t}^{\equiv}$, $N_{m+t}^{\equiv}$, $N_t^{\equiv}$ represents one or more sample image subsets corresponding to the branchy neural network; wherein

the classification loss values corresponding to the plurality of branchy neural networks are calculated by the following equation:

$$L_i = \sum_{x_i \in X} J(N_{h+m+t}^{\equiv}(x_i), y_i) + \sum_{x_i \in S_{m+t}} J(N_{m+t}^{\equiv}(x_i), y_i) + \sum_{x_i \in S_t} J(N_t^{\equiv}(x_i), y_i),$$

wherein, L_i represents a sum of the classification loss values corresponding to the plurality of branchy neural networks, S_{m+t} represents a subset of X and comprises sample images in the sample image input sequence which belong to the second sample image subset and the third sample image subset, and S_t represents another subset of X and comprises sample images in the sample image input sequence which belong to the third sample image subset.

6. The method according to claim 5, wherein the loss value of the to-be-trained image recognition model is calculated by the following equation:

$$L_{all} = (1-\alpha)L_f / n_1 + \alpha L_i / n_2,$$

wherein, L_{all} represents the loss value of the to-be-trained image recognition model, α represents a hyper-parameter, n_1 represents a quantity of sample images in X , n_2 represents a total quantity of the sample images in X , S_{m+t} and S_t .

7. The method according to claim 1, wherein before acquiring the sample image set, the

method further comprises:

acquiring sample images, and determining, according to image categories of the sample images, a quantity of images in each of the image categories;

5 acquiring a sort order of the image categories according to the quantity of images in each of the image categories, and classifying the image categories into a plurality of category combinations according to the sort order, wherein each of the plurality of category combinations comprises the same quantity of image categories; and

10 acquiring sample image subsets corresponding to the plurality of category combinations according to the plurality of category combinations and sample images corresponding to the image categories in the plurality of category combinations, and combining the plurality of sample image subsets into the sample image set.

8. An apparatus for generating an image recognition model, comprising:

15 an acquisition module, configured to acquire a sample image set, wherein the sample image set comprises a plurality of sample image subsets, wherein quantities of images comprised in the plurality of sample image subsets decrease in turn, and each of the plurality of sample image subsets comprises a same quantity of image categories;

20 a training module, configured to train a to-be-trained image recognition model based on the sample image set, to acquire a loss value of the to-be-trained image recognition model, wherein the to-be-trained image recognition model comprises a plurality of branchy neural networks, and each of the branchy neural networks is used to recognize a corresponding image, the loss value comprises a target classification loss value and classification loss values respectively corresponding to the plurality of branchy neural networks, the target classification loss value is a loss value of the to-be-trained image recognition model for the sample image set, and for each of the plurality of branchy neural networks, the classification
25 loss value corresponding to the branchy neural network is a loss value of the branchy neural network for sample image subsets corresponding to the branchy neural network; and

30 an adjustment module, configured to adjust parameters of the to-be-trained image recognition model according to the loss value until the loss value is lower than a preset threshold, and determine the to-be-trained image recognition model when the loss value is lower than the preset threshold as a trained image recognition model.

9. A computer device, comprising;

a memory for storing a computer program; and

a processor, configured to, when executing the computer program, implement steps of the method according to any one of claims 1 to 7.

5 10. A computer-readable storage medium storing a computer program, wherein the computer program, when being executed by a processor, causes steps of the method according to any one of claims 1 to 7 to be implemented.