CLAIMS

What is claimed is:

1. A model training method, comprising:

obtaining a plurality of training samples, wherein each training sample comprises an image and a text, and the text describes a target object in the image; and

inputting the plurality of training samples into a target model, so that the target model performs the following procedure until a preset stop condition is met:

extracting an image feature of a first image and a text feature of a first text, wherein the first image is an image comprised in any one of the plurality of training samples, and the first text describes a target object in the first image;

obtaining a first loss value based on a difference between a first vector and a second vector, wherein a dimension of the first vector is the same as a dimension of the second vector, the first vector indicates the image feature of the first image, and the second vector indicates the text feature of the first text;

predicting, based on the image feature of the first image, a first probability that the target object in the first image belongs to each target object in an object set, wherein the object set comprises target objects comprised in all images in the plurality of training samples;

predicting, based on the text feature of the first text, a second probability that the first text describes each target object in the object set; and

obtaining a third loss value based on a difference between the first probability and the second probability; and

updating the target model based on the first loss value and the third loss value.

2. The method according to claim 1, wherein the method further comprises:

predicting at least one attribute of the target object in the first image based on the image feature of the first image; and

obtaining a second loss value based on a difference between the at least one attribute and a preset tag, wherein the preset tag is obtained through word segmentation preprocessing performed on the first text; and

the updating the target model based on the first loss value and the third loss value comprises:

updating the target model based on the first loss value, the second loss value and the third loss value.

3. The method according to claim 2, wherein the method further comprises:

obtaining a fourth loss value based on a difference between the first probability and a first preset value, wherein the first preset value indicates a preset probability that the target object in

the first image belongs to each target object in the object set; and

obtaining a fifth loss value based on a difference between the second probability and a second preset value, wherein the second preset value indicates a preset probability that the first text describes each target object in the object set; and

the updating the target model based on the first loss value comprises:

updating the target model based on the first loss value, the second loss value, the third loss value, the fourth loss value, and the fifth loss value.

4. The method according to claim 3, wherein the method further comprises:

obtaining a sixth loss value based on a difference between the image feature of the first image and an image feature of a second image, wherein the second image and the first image present a same target object from different angles; and

the updating the target model based on the first loss value comprises:

updating the target model based on the first loss value, the second loss value, the third loss value, the fourth loss value, the fifth loss value, and the sixth loss value.

5. The method according to claim 4, wherein the method further comprises:

obtaining a first distance between the first vector and a third vector, and obtaining a second distance between the first vector and a fourth vector, wherein the third vector indicates the image feature of the second image, the fourth vector indicates an image feature of a third image, and a target object comprised in the third image is different from that comprised in the first image; and

obtaining a seventh loss value based on a difference between a preset threshold and a difference between the first distance and the second distance; and

the updating the target model based on the first loss value comprises:

updating the target model based on the first loss value, the second loss value, the third loss value, the fourth loss value, the fifth loss value, the sixth loss value, and the seventh loss value.

6. An image retrieval method, comprising:

obtaining a to-be-retrieved image;

extracting an image feature of the to-be-retrieved image by using a target model; and

obtaining a predicted image from an image database based on the image feature of the to-be-retrieved image, wherein a similarity between an image feature of the predicted image and the image feature of the to-be-retrieved image reaches a threshold, wherein

the target model is a model obtained by training an initial target model based on the model training method according to any one of claims 1 to 5.

7. A model training apparatus, comprising:

a memory, configured to store computer-readable instructions; and

a processor coupled to the memory, configured to execute the computer-readable instructions

in the memory, so as to perform the method according to any one of claims 1 to 5.

8. An image retrieval device, comprising:

a memory, configured to store computer-readable instructions; and

a processor coupled to the memory, configured to execute the computer-readable instructions in the memory, so as to perform the method according to claim 6.

9. A computer program product, wherein when the computer program product is run on a computer, the computer is enabled to perform the method according to any one of claims 1 to 6.