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; and

updating the target model based on the first 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 comprises:

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

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

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

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, and the third loss value.

4. The method according to claim 3, 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.

5. The method according to claim 4, 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.

6. The method according to claim 5, 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.

~~7. The method according to any one of claims 1 to 6, wherein the obtaining a plurality of training samples comprises:~~

~~obtaining a first target image of a product, a text title of the product, and a second target image of the product, wherein the first target image comes from a provider of the product, and the second target image comes from a buyer of the product;~~

~~using the first target image as an input of a segmentation model to obtain a segmentation result, wherein the segmentation result indicates a contour of the product in the first target image;~~

~~extracting a target area of a third target image based on the segmentation result, wherein the~~

third target image is obtained by inputting the first target image into a generator, and the image provided by the provider of the product and the image provided by the buyer of the product are used for a training process of the generator;

performing random affine transformation on the target area to obtain a transformed target area;

combining the transformed target area and a preset background to obtain a fourth target image; and

obtaining a first-type training sample, a second-type training sample, and a third-type training sample, wherein each first-type training sample comprises the first target image and the text title, each second-type training sample comprises the second target image and the text title, and each third-type training sample comprises the fourth target image and the text title.

8. The method according to any one of claims 1 to 7, wherein the method further comprises:

performing a plurality of times of training on a hash model based on a plurality of image features extracted by the target model, to obtain a trained hash model, wherein any one of the plurality of times of training comprises:

obtaining a hash feature corresponding to each image feature;

performing clustering processing on a plurality of hash features, to obtain a plurality of clustered datasets;

obtaining a first hash feature from the plurality of hash features, wherein the first hash feature is any one of the plurality of hash features;

obtaining a first clustered dataset from the plurality of clustered datasets based on a similarity between a cluster center of each clustered dataset and the first hash feature, wherein a similarity between a cluster center of the first clustered dataset and the first hash feature is the highest in the plurality of clustered datasets;

obtaining an eighth loss value based on a difference between the cluster center of the first clustered dataset and the first hash feature; and

updating the hash model based on the eighth loss value.

9. The method according to claim 8, wherein 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, the seventh loss value, and the eighth loss value.

10. The method according to claim 8 or 9, wherein any one of the plurality of times of training further comprises:

applying a first constraint and a second constraint to a target hash feature, to obtain an updated target hash feature, wherein the target hash feature is a hash feature obtained based on a last

~~updated hash model, the first constraint is that quantities of third preset values and fourth preset values in the hash feature are average, the second constraint is that an element on a first diagonal of a target matrix is the fourth preset value and an element other than the first diagonal in the target matrix is the third preset value, an element in the i^{th} row and the j^{th} column in the target matrix represents an average value of a plurality of target elements, the plurality of target elements represent products of values of the i^{th} dimension of the hash features and values of the j^{th} dimension of the hash features, and i and j are positive integers;~~

~~comparing a value of each dimension of the target hash feature and a second mean, wherein the second mean is an average value of the third preset value and the fourth preset value; and~~

~~when the value of each dimension is less than the second mean, obtaining a ninth loss value based on a difference between the value of each dimension of the hash feature and the third preset value; or~~

~~when the value of each dimension is not less than the second mean, obtaining a ninth loss value based on a difference between the value of each dimension of the hash feature and the fourth preset value; and~~

~~the updating the hash model based on the eighth loss value comprises:~~

~~updating the hash model based on the eighth loss value and the ninth loss value.~~

~~11~~7. 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 updating an initial target model based on a first loss value, the first loss value is obtained based on a difference between a first vector and a second vector, a dimension of the first vector is the same as a dimension of the second vector, the first vector indicates an image feature of a first image, the second vector indicates a text feature of a first text, the first image is an image comprised in any one of a plurality of training samples, the first text describes a target object in the first image, each training sample comprises an image and a text, and the text describes a target object in the image.

~~12~~8. The method according to claim ~~11~~7, wherein the target model is specifically a model obtained by updating the initial target model based on the first loss value and a second loss value, the second loss value is obtained based on a difference between at least one attribute and a preset tag, and the at least one attribute is at least one attribute that is of the target object in the first image and that is predicted based on the image feature of the first image.

~~13~~9. The method according to claim ~~8~~12, wherein the target model is specifically obtained by updating the initial target model based on the first loss value, the second loss value, and a third loss value, the third loss value is obtained based on a difference between a first probability and a second probability, the first probability is predicted based on the image feature of the first image and is a probability that the target object in the first image belongs to each target object in an object set, and the second probability is predicted based on the text feature of the first text and is a probability that the first text describes each target object in the object set.

~~14~~10. The method according to claim ~~13~~9, wherein the target model is specifically a model obtained by updating the initial target model based on the first loss value, the second loss value, the third loss value, a fourth loss value, and a fifth loss value, the fourth loss value is obtained based on a difference between the first probability and a first preset value, 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, the fifth loss value is obtained based on a difference between the second probability and a second preset value, and the second preset value indicates a preset probability that the first text describes each target object in the object set.

~~15~~11. The method according to claim ~~14~~10, wherein the target model is specifically obtained by updating the initial 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 a sixth loss value, the sixth loss value is obtained based on a difference between the image feature of the first image and an image feature of a second image, and the second image and the first image present the same target object from different angles.

~~16~~12. The method according to claim ~~15~~11, wherein the target model is specifically obtained by updating the initial 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 a seventh loss value, the seventh loss value is obtained based on a difference between a preset threshold and a difference between a first distance and a second distance, the first distance is a distance between the first vector and a third vector, the second distance is a distance between the first vector and a fourth vector, 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.

~~17. The method according to any one of claims 11 to 16, wherein the plurality of training samples comprise a first-type training sample, a second-type training sample, and a third-type training sample, each first-type training sample comprises a first target image and a text title, each second-type training sample comprises a second target image and the text title, each third-type training sample comprises a fourth target image and the text title, the first target image comes from~~

~~a provider of a product, the second target image comes from a buyer of the product, the text title is a text title of the product, the fourth target image is obtained by combining a transformed target area and a preset background, the transformed target area is obtained through random affine transformation performed on a target area, the target area is obtained through extraction from a third target image based on a segmentation result, the third target image is obtained by inputting the first target image into a generator, the image provided by the provider of the product and the image provided by the buyer of the product are used for a training process of the generator, the segmentation result is a result obtained by using the first target image as an input of a segmentation model, and the segmentation result indicates a contour of the product in the first target image.~~

~~18. The method according to any one of claims 11 to 17, wherein the method further comprises:~~

~~inputting the image feature of the to-be-retrieved image into a hash model, to obtain a hash feature of the to-be-retrieved image; and~~

~~the obtaining a predicted image from an image database based on the image feature of the to-be-retrieved image comprises:~~

~~obtaining the predicted image from the image database based on the hash feature of the to-be-retrieved image, wherein the hash model is obtained by updating an initial hash model based on an eighth loss value, the eighth loss value is obtained based on a difference between a cluster center of a first clustered dataset and a first hash feature, the cluster center of the first clustered dataset is a cluster center with a highest similarity to the first hash feature in cluster centers of all clustered datasets in the plurality of pieces of clustered data, the plurality of clustered datasets are obtained through clustering processing performed on a plurality of hash features, and the first hash feature is any one of the plurality of hash features.~~

~~19. The method according to claim 18, wherein the target model is specifically obtained by updating the initial 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, the seventh loss value, and the eighth loss value.~~

~~20~~13. 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 ~~10~~12.

~~21~~14. 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 any one of claims ~~11~~ 7 to ~~19~~12.

~~22. A chip system, wherein the chip system comprises a processor and a communication interface, the processor obtains program instructions by using the communication interface, and when the program instructions are executed by the processor, the method according to any one of claims 1 to 10 is implemented.~~

~~23. A chip system, wherein the chip system comprises a processor and a communication interface, the processor obtains program instructions by using the communication interface, and when the program instructions are executed by the processor, the method according to any one of claims 11 to 19 is implemented.~~

~~24. A computer-readable storage medium, comprising a program, wherein when the program is executed by a processing unit, the method according to any one of claims 1 to 10 is performed.~~

~~25. A computer-readable storage medium, comprising a program, wherein when the program is executed by a processing unit, the method according to any one of claims 11 to 19 is performed.~~

~~26~~15. 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 ~~10~~12.

~~27. 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 11 to 19.~~