

NETWORK TRAINING METHOD AND APPARATUS, TARGET DETECTION METHOD AND APPARATUS, AND ELECTRONIC DEVICE

CROSS-REFERENCE TO RELATED APPLICATION

- [0001] The application is based upon and claims priority to Chinese patent application No. 202010681178.2 filed on July 15, 2020, the disclosure of which is hereby incorporated by reference in its entirety.

TECHNICAL FIELD

- [0002] The disclosure relates to the technical field of computers, and in particular to a network training method and apparatus, a target detection method and apparatus, and an electronic device.

BACKGROUND

- [0003] Computer vision is an important direction of artificial intelligence technologies. In computer vision processing, it is usually necessary to detect targets (such as pedestrians and objects) in images or videos. Target detection on large-scale long-tailed data has important applications in many fields, such as abnormal object detection, abnormal behavior detection and emergency alarm in urban monitoring. However, due to the huge amount of long-tailed data and the phenomenon of serious imbalance between positive and negative samples, that is, most data pictures are background pictures, and only a small part of pictures contain detectable targets, which results in poor effect of the target detection on the long-tailed data through the target detection mode in the related art.

SUMMARY

- [0004] The embodiments of the disclosure provide solutions for network training and target detection.
- [0005] According to an aspect of the embodiments of the disclosure, a network training method is provided, which includes the following operations.
- [0006] An unannotated first sample image is input into a target detection network for processing to obtain a target detection result of the first sample image, where the target detection result includes an image area, feature information and class probability of a target in the first sample image. The category confidence of the target is determined according to the class probability of the target. For a first target with the category confidence in the target greater than or equal to a first threshold value, the first sample image where the first target is located is taken as an annotated second sample image and is added to a training set, where the annotation information of the second sample image includes the image area of the first target and the category corresponding to the category confidence of the first target, and the training set includes an annotated third sample image. For a second target whose category confidence in the target is less than the first threshold value, feature-relevance mining is performed on the second target according to the feature information of a third target in the third sample image, a fourth target and the first sample image where the fourth target is located are determined from the second target through feature-relevance mining, and the first sample image where the fourth

target is located is taken as the fourth sample image and is added to the training set. The target detection network is trained according to the annotation information of the fourth sample image, as well as the second sample image, the third sample image and the fourth sample image in the training set.

[0007] In a possible implementation mode, the operation that the target detection network is trained according to the annotation information of the fourth sample image, as well as the second sample image, the third sample image and the fourth sample image in the training set includes the following operations.

[0008] The first number of samples are respectively determined from a positive sample image of each category according to the category of the target in positive sample images of the training set, where the positive sample image is a sample image including the target. The positive sample image of each category is sampled according to the first number of the samples in the positive sample image of each category to obtain a plurality of fifth sample images. A negative sample image of the training set is sampled to obtain a plurality of sixth sample images, where the negative sample image is a sample image that does not include the target in the image. The target detection network is trained according to the fifth sample images and the sixth sample images.

[0009] In a possible implementation mode, the operations that feature-relevance mining is performed on the second target according to the feature information of the third target in the third sample image, and the fourth target and the first sample image where the fourth target is located are determined from the second target through feature-relevance mining include the following operations. The information entropy of the second target is determined according to the class probability of the second target; a fifth target is selected from the second target according to the category confidence and the information entropy of the second target; the second number of sample images to be mined in each category is respectively determined according to the category of the third target in the third sample image and the total number of sample images to be mined; and the fourth target and the first sample image where the fourth target is located are determined from the fifth target according to the feature information of the third target in the third sample image, the feature information of the fifth target and the second number of the sample images to be mined in each category.

[0010] In a possible implementation mode, the operation that the fifth target is selected from the second target according to the category confidence and the information entropy of the second target includes the following operations. The second target is sorted according to the category confidence and the information entropy of the second targets respectively, the third number of sixth targets and the fourth number of seventh targets are selected; and the sixth targets and the seventh targets are combined to obtain the fifth target.

[0011] In a possible implementation mode, the operation that the second number of the sample images to be mined in each category is respectively determined according to the category of the third target in the third sample image and the total number of the sample images to be mined includes the following operations. The proportion of the third target of each category is determined according to the category of the third target in the third sample image; the sampling proportion of each category is determined according to the proportion of the third target of each category; and the second number of the sample images to be mined in each category is respectively determined according to the sampling proportion of each category.

[0012] In a possible implementation mode, the operation that the fourth target and the

first sample image where the fourth target is located are determined from the fifth target according to the feature information of the third target in the third sample image, the feature information of the fifth target and the second number of the sample images to be mined in each category includes the following operations. The third target at the smallest distance from each fifth target among the third target of the first category is respectively determined according to the distance between the feature information of the third target of the first category and the feature information of each fifth target, and is taken as an eighth target, where the first category is any one of categories of the third target; and the target at the largest distance among the eighth targets is determined as the fourth target.

- [0013] In a possible implementation mode, the operation that the fourth target and the first sample image where the fourth target is located are determined from the fifth target according to the feature information of the third target in the third sample image, the feature information of the fifth target and the second number of the sample images to be mined in each category further includes the following operations. The determined fourth target is added to the third target of the first category, and the determined fourth target is removed from the unannotated fifth target.
- [0014] In a possible implementation mode, the method further includes the following operation. The third sample image is input into the target detection network for processing to obtain the feature information of the third target in the third sample image.
- [0015] In a possible implementation mode, before the operation that the unannotated first sample image is input into the target detection network for processing to obtain the target detection result of the first sample image, the method further includes the following operation.
- [0016] The target detection network is pre-trained through the annotated third sample image.
- [0017] In a possible implementation mode, the first sample image includes a long-tailed image.
- [0018] According to an aspect of the embodiments of the disclosure, a target detection method is provided, including the following operation. An image to be processed is input into a target detection network for processing to obtain a target detection result of the image to be processed, the target detection result includes the position and category of a target in the image to be processed, and the target detection network is obtained by training according to the above network training method.
- [0019] According to an aspect of the embodiments of the disclosure, a network training apparatus is provided, which includes a target detection part, a confidence determination part, an annotation part, a feature mining part and a training part.
- [0020] The target detection part is configured to input an unannotated first sample image into a target detection network for processing to obtain a target detection result of the first sample image, and the target detection result includes an image area, feature information and class probability of a target in the first sample image.
- [0021] The confidence determination part is configured to determine the category confidence of the target according to the class probability of the target.
- [0022] The annotation part is configured to, for a first target with the category confidence in the target greater than or equal to a first threshold value, take the first sample image where the first target is located as an annotated second sample image and add the annotated second sample image to a training set, where the annotation information of the second sample image includes the image area of the first target and the category corresponding to the category confidence of the first target, and the training set

includes an annotated third sample image.

- [0023] The feature mining part is configured to, for a second target with the category confidence in the target less than the first threshold value, perform feature-relevance mining on the second target according to the feature information of a third target in the third sample image, determine a fourth target and the first sample image where the fourth target is located from the second target through feature-relevance mining, take the first sample image where the fourth target is located as the fourth sample image, and add the fourth sample image to the training set.
- [0024] The training part is configured to train the target detection network according to the annotation information of the fourth sample image, as well as the second sample image, the third sample image and the fourth sample image in the training set.
- [0025] In a possible implementation mode, the training part includes a sampling number determination subpart, a first sampling subpart, a second sampling subpart and a training subpart. The sampling number determination subpart is configured to respectively determine the first number of samples from a positive sample image of each category according to the category of the target in positive sample images of the training set, and the positive sample image is a sample image including the target. The first sampling subpart is configured to sample the positive sample image of each category according to the first number of the samples in the positive sample image of each category to obtain a plurality of fifth sample images. The second sampling subpart is configured to sample a negative sample image of the training set to obtain a plurality of sixth sample images, where the negative sample image is a sample image that does not include the target. The training subpart is configured to train the target detection network according to the fifth sample images and the sixth sample images.
- [0026] In a possible implementation mode, the feature mining part includes an information entropy determination subpart, a target selection subpart, a mining number determination subpart, and a target and image determination subpart. The information entropy determination subpart is configured to determine information entropy of the second target according to the class probability of the second target. The target selection subpart is configured to select a fifth target from the second target according to the category confidence and information entropy of the second target. The mining number determination subpart is configured to respectively determine the second number of sample images to be mined in each category according to the category of the third target in the third sample image and the total number of sample images to be mined. The target and image determination subpart is configured to determine the fourth target and the first sample image where the fourth target is located from the fifth target according to the feature information of the third target in the third sample image, the feature information of the fifth target and the second number of the sample images to be mined in each category.
- [0027] In a possible implementation mode, the target selection subpart is configured to respectively sort the second target according to the category confidence and information entropy of the second target, select the third number of sixth targets and the fourth number of seventh targets; and combine the sixth targets and the seventh targets to obtain the fifth target.
- [0028] In a possible implementation mode, the mining number determination subpart is configured to determine the proportion of the third target of each category according to the category of the third target in the third sample image, determine the sampling proportion of each category according to the proportion of the third target of each

category, and respectively determine the second number of the sample images to be mined in each category according to the sampling proportion of each category.

- [0029] In a possible implementation mode, the target and image determination subpart is configured to respectively determine the third target at the smallest distance from each fifth target among the third target of the first category according to the distance between the feature information of the third target of the first category and the feature information of each fifth target, take the third target as an eighth target, and determine the target with the largest distance among the eighth targets as the fourth target. The first category is any one of categories of the third target.
- [0030] In a possible implementation mode, the target and image determination subpart is further configured to add the determined fourth target to the third target of the first category, and remove the determined fourth target from the unannotated fifth target.
- [0031] In a possible implementation mode, the apparatus further includes a feature extraction part, which is configured to input the third sample image into the target detection network for processing to obtain the feature information of the third target in the third sample image.
- [0032] In a possible implementation mode, before the target detection part, the apparatus further includes a pre-training part, which is configured to pre-train the target detection network through the annotated third sample image.
- [0033] In a possible implementation mode, the first sample image includes a long tail image.
- [0034] According to an aspect of the embodiments of the disclosure, a target detection apparatus is provided, which includes a detection processing part configured to input an image to be processed into a target detection network for processing to obtain a target detection result of the image to be processed, where the target detection result includes the position and category of a target in the image to be processed, and the target detection network is obtained by training according to the above network training method.
- [0035] In a possible implementation mode, before the operation that the first number of the samples from the positive sample images of each category is respectively determined according to the category of the target in the positive sample image of the training set, the method further includes the following operation. The positive sample images and the negative sample image in the training set are sampled to obtain the same or similar number of positive sample images and negative sample images.
- [0036] In a possible implementation mode, the total number of the sample images to be mined is 5% to 25% of the total number of the first sample images.
- [0037] In a possible implementation mode, the operation that the sixth targets and the seventh targets are combined to obtain the fifth target includes the following operations. The same target as the seventh targets is removed from the sixth targets to obtain the remaining target in the sixth targets different from the seventh targets; and the remaining target and the seventh targets are taken as the fifth target.
- [0038] In a possible implementation mode, the method further includes the following operation. The feature-relevance mining of the first category is ended when the number of the fourth sample images of the first category reaches the second number of the sample images to be mined of the first category,.
- [0039] In a possible implementation mode, after the operations that the third target at the smallest distance from each fifth target among the third target of the first category is respectively determined according to the distance between the feature information of the third target of the first category and the feature information of each fifth target, and the

third target is taken as an eighth target, the method further includes the following operation. When the number of the first sample images where the fourth target is located reaches the second number of the sample images to be mined of the first category, the determination of the eighth target is ended.

[0040] In a possible implementation mode, after the operations that the third target at the smallest distance from each fifth target among the third targets of the first category is respectively determined according to the distance between the feature information of the third target of the first category and the feature information of each fifth target, and the third target is taken as an eighth target, the method further includes the following operation. When the number of the first sample images where the fourth target is located does not reach the second number of the sample images to be mined of the first category and a set storing the feature information of the fifth target is empty, the determination of the eighth target is ended.

[0041] In a possible implementation mode, the operation that the third sample image is input into the target detection network for processing to obtain the feature information of the third target in the third sample image includes the following operations. The third sample image is input into the target detection network to obtain a feature vector output by a hidden layer of the target detection network; and the feature vector is determined as the feature information of the third target.

[0042] According to an aspect of the embodiments of the disclosure, an electronic device is provided, including a processor, and a memory configured to store a processor executable instruction. The processor is configured to call the instruction stored in the memory to implement the above network training method or the above target detection method.

[0043] According to an aspect of the embodiments of the disclosure, a computer-readable storage medium is provided. A computer program instruction may be stored on the computer-readable storage medium. The computer program instruction may be executed by a processor to implement the above network training method or the above target detection method.

[0044] According to the embodiments of the disclosure, the target detection result of the unannotated sample image may be obtained through the target detection network. According to the target detection result, pseudo-labeling and feature-relevance mining are performed respectively, and high-value sample images are annotated and collected and added to the training set. The target detection network is trained according to the expanded training set, so as to expand the number of positive sample data in the training set, alleviate the imbalance between positive and negative samples, and improve the training effect of the target detection network.

[0045] It is to be understood that the above general descriptions and detail descriptions below are merely exemplary and explanatory, which may not limit the embodiments of the disclosure. According to the following detailed descriptions made to exemplary embodiments with reference to the drawings, other features and aspects of the embodiments of the disclosure may become clear.

BRIEF DESCRIPTION OF THE DRAWINGS

[0046] The drawings here are incorporated in the specification as a part of the specification. These drawings show embodiments that are in accordance with the disclosure, and used together with the specification to describe technical solutions of the embodiments of the disclosure.

- [0047] FIG. 1 is a flowchart of a network training method according to the embodiments of the disclosure.
- [0048] FIG. 2 is a schematic diagram of a processing process of a network training method according to the embodiments of the disclosure.
- [0049] FIG. 3 is a block diagram of a network training apparatus according to the embodiments of the disclosure.
- [0050] FIG. 4 is a block diagram of an electronic device according to the embodiments of the disclosure.
- [0051] FIG. 5 is a block diagram of an electronic device according to the embodiments of the disclosure.

DETAILED DESCRIPTION

- [0052] Each exemplary embodiment, feature and aspect of the disclosure will be described below with reference to the drawings in detail. The same reference signs in the drawings represent components with the same or similar functions. Although each aspect of the embodiments is shown in the drawings, the drawings are not required to be drawn to scale, unless otherwise specified.
- [0053] Herein, special term "exemplary" refers to "use as an example, embodiment or description". Herein, any "exemplarily" described embodiment may not be explained to be superior to or better than other embodiments.
- [0054] The term "and/or" in this specification describes only an association relationship for describing associated objects and represents that three relationships may exist. For example, A and/or B may represent the following three cases: Only A exists, both A and B exist, and only B exists. In addition, term "at least one" in the disclosure represents any one of multiple or any combination of at least two of multiple. For example, including at least one of A, B, or C may represent including any one or more elements selected from a set formed by A, B, or C.
- [0055] In addition, for describing the embodiments of the disclosure better, many specific details are presented in the following specific implementation modes. It is understood by those skilled in the art that the embodiments of the disclosure may still be implemented even without some specific details. In some examples, methods, means, components and circuits known very well to those skilled in the art are not described in detail, so as to highlight the subject of the embodiments of the disclosure.
- [0056] FIG. 1 is a flowchart of a network training method according to the embodiments of the disclosure. As illustrated in FIG. 1, the network training method includes the following operations.
- [0057] In S11, an unannotated first sample image is input into a target detection network for processing to obtain a target detection result of the first sample image, and the target detection result includes an image area, feature information and class probability of a target in the first sample image.
- [0058] In S12, the category confidence of the target is determined according to the class probability of the target.
- [0059] In S13, for a first target with the category confidence greater than or equal to a first threshold value in the target, the first sample image where the first target is located is taken as an annotated second sample image and added to a training set, the annotation information of the second sample image includes the image area of the first target and the category corresponding to the category confidence of the first target, and the training set includes an annotated third sample image.

- [0060] In S14, for a second target with the category confidence less than the first threshold value in the target, feature-relevance mining is performed on the second target according to the feature information of a third target in the third sample image, a fourth target and the first sample image where the fourth target is located are determined from the second target through feature-relevance mining, and the first sample image where the fourth target is located is taken as the fourth sample image and is added to the training set.
- [0061] In S15, the target detection network is trained according to the annotation information of the fourth sample image, as well as the second sample image, the third sample image and the fourth sample image in the training set.
- [0062] In a possible implementation mode, the method may be executed by electronic devices, such as a terminal device or a server. The terminal device may be User Equipment (UE), a mobile device, a user terminal, a terminal, a cell phone, a cordless phone, a Personal Digital Assistant (PDA), a hand-held device, a computing device, a vehicle device, a wearable device, etc. The method may be implemented by the processor by calling a computer readable instruction stored in the memory. Or, the method may be executed through a server.
- [0063] For example, the first sample image may be an image acquired by an image collection device (such as a camera). The first sample image may include a large-scale long-tailed image, that is, most of the images are background images, and a small part of the images include detectable targets. The detectable targets may include, for example, a human body, a human face, a vehicle, an object, and the like. For example, in the field of security, an image of a certain geographical area may be collected through a camera, and people may pass through the geographical area only in a small part of the time, so that most of the collected images are background images, and only a small part of the images include human faces and/or human bodies. In this case, a plurality of collected images may form a long-tailed data set. The embodiments of the disclosure do not limit the acquisition mode of the first sample image and the category of targets in the first sample image.
- [0064] In a possible implementation mode, a target detection network may be preset to detect the position (i.e. a detection frame) and category of the target in the image. The target detection network may be, for example, a convolutional neural network, and the embodiments of the disclosure do not limit the network structure of the target detection network.
- [0065] In a possible implementation mode, before S11, the method further includes that: the target detection network is pre-trained through the annotated third sample image. That is, a training set may be preset, which includes the annotated third sample image, and the annotation information of the third sample image may include the detection frame and category of the target in the image. According to the training set, the target detection network may be pre-trained in the way of the related art, so that the target detection network has a certain detection accuracy.
- [0066] However, the target detection network after being pre-trained has poor detection effect on large-scale long-tailed images. Therefore, the target detection network may be further trained by using the unannotated first sample image through an active learning manner.
- [0067] In a possible implementation mode, in S11, an unannotated first sample image is input into a target detection network for processing to obtain a target detection result of the first sample image. The target detection result may include an image area, feature information and class probability of a target in a first sample image. An image area where

the target is located may be a detection frame in the image. The feature information of the target may be, for example, a feature vector output from a hidden layer (such as a convolution layer) of the target detection network. The class probability of the target may represent the class posterior probability that the target belongs to each category.

- [0068] In a possible implementation mode, the target in the first sample image may also be called an instance, and one or more targets may be detected in each first sample image. In actual processing, the magnitude of the number of detected targets may be several to dozens of times the magnitude of the number of images.
- [0069] In a possible implementation mode, in S12, according to the class probability of the target, the maximum value of the class probability may be obtained and determined as the category confidence of the target.
- [0070] In a possible implementation mode, in S13, for a target (which may be called a first target) with the category confidence greater than or equal to a first threshold value, a first sample image where the first target is located may be taken as an annotated sample image (which may be called a second sample image) and added to a training set. The image area of the first target is taken as the annotated image area, and the category corresponding to the category confidence of the first target is taken as the annotated category of the first target. The same second sample image may be annotated multiple times for a plurality of first targets in the second sample image. The first threshold value is, for example, 0.99, and the embodiments of the disclosure do not limit the value of the first threshold value.
- [0071] In a possible implementation mode, the processing process of S13 may be called pseudo-labeling. That is, the image where the target with relatively high confidence is located is taken as a high-value sample, and a target detection reasoning result is directly taken as a target annotation result. In this way, the amount of positive sample data in the training set may be expanded to solve the problem of difficulty in collection of positive samples.
- [0072] In a possible implementation mode, in S14, for the target (which may be called the second target) with the category confidence less than the first threshold value, the feature-relevance mining may be performed on the second target according to the feature information of the target (which may be called the third target) in the annotated third sample image in the training set, and the target (which may be called the fourth target) meeting the requirements may be mined from the second target. For example, the distance or correlation between the feature information of the third target and the feature information of the second target may be calculated, a preset number of targets may be selected according to the distance or correlation, and the selected preset number of targets may be taken as the fourth target.
- [0073] In a possible implementation mode, the first sample image where the mined fourth target is located may be taken as the fourth sample image and added to the training set, so as to complete the processing process of feature-relevance mining. In this way, the amount of sample data in the training set may be further expanded.
- [0074] In a possible implementation mode, the annotation information of the fourth sample image may be obtained by human annotation, such as human determination of the detection frame and category of the target in the fourth sample image. There is no limitation to this in the embodiments of the disclosure.
- [0075] In a possible implementation mode, in S15, after the annotation information of the fourth sample image is obtained, the target detection network may be trained according to the second sample image, the third sample image and the fourth sample

image in the training set.

- [0076] According to the embodiments of the disclosure, the target detection result of the unannotated sample image may be obtained through the target detection network. According to the target detection result, pseudo-labeling and feature-relevance mining are performed respectively, and high-value sample images are annotated and collected and added to the training set. The target detection network is trained according to the expanded training set, so as to expand the amount of positive sample data in the training set, alleviate the imbalance between positive and negative samples, and improve the training effect of the target detection network.
- [0077] In a possible implementation mode, after the processing in S11, the target detection result of each first sample image is obtained; and after the processing in S12, the category confidence of the target in each first sample image is obtained. In S13, the sample image where the first target with the category confidence greater than or equal to the first threshold value is located may be added to the training set, and the annotated second sample image may be obtained by pseudo-labeling. In S14, the second target with the category confidence less than the first threshold value may be mined.
- [0078] In a possible implementation mode, the S14 may include the following operations.
- [0079] The information entropy of the second target is determined according to the class probability of the second target.
- [0080] A fifth target is selected from the second target according to the category confidence and information entropy of the second target.
- [0081] The second number of sample images to be mined in each category is respectively determined according to the category of the third target in the third sample image and the total number of sample images to be mined.
- [0082] A fourth target and a first sample image where the fourth target is located are determined from the fifth target according to the feature information of the third target in the third sample image, the feature information of the fifth target and the second number of the sample images to be mined in each category.
- [0083] For example, according to the class probability of the second target, the information entropy of the second target may be obtained by calculation to represent the uncertainty of the second target. That is, the greater the information entropy of the second target is, the greater the uncertainty of the second target is; on the contrary, the smaller the information entropy of the second target is, the smaller the uncertainty of the second target is. The embodiments of the disclosure do not limit the calculation mode of information entropy.
- [0084] In a possible implementation mode, according to the category confidence and information entropy of the second target, the target (which may be called the fifth target) meeting a certain condition may be selected from a plurality of second targets, such as the target with relatively high category confidence, the target with relatively high information entropy.
- [0085] In a possible implementation mode, the step that the fifth target is selected from the second target according to the category confidence and information entropy of the second target may include the following operations.
- [0086] The second target is sorted according to the category confidence and information entropy of the second targets respectively, and the third number of sixth targets and the fourth number of seventh targets are selected.
- [0087] The sixth targets and the seventh targets are combined to obtain the fifth target.

- [0088] That is, a plurality of second targets are sorted according to the category confidence of the second targets; and according to the sorting result, a preset third number of targets (which may be called the sixth target) are selected from the plurality of second targets. Similarly, the plurality of second targets are sorted according to the information entropy of the second targets; and according to the sorting result, a preset fourth number of targets (which may be called the seventh target) are selected from the plurality of second targets. The third number and the fourth number may be $3 \times K$ respectively. K represents the number of sample images to be mined. For example, the value of K is 10000. In actual processing, the value of K may be 5% to 25% of the total number of unannotated first sample images. The embodiments of the disclosure do not limit the value of the K and the number relationship among the third number, the fourth number and the K .
- [0089] It should be understood that, those skilled in the art may set the values of the number K , the third number and the fourth number of sample images to be mined according to the actual situation, and the third number and the fourth number may be different, which is not limited by the embodiments of the disclosure.
- [0090] In a possible implementation mode, the selected sixth targets and seventh targets may be combined, and the plurality of targets after combination may be taken as the fifth target, so as to remove the possible repetitive targets. In actual processing, about $6 \times K$ fifth targets may be obtained.
- [0091] The above processing mode may be called bootstrapping. In this way, a certain number of positive and negative samples with relatively high possibility may be selected from the second target at the same time for subsequent feature-relevance mining, so as to reduce the amount of calculation of feature-relevance mining and improve the processing efficiency.
- [0092] In a possible implementation mode, the step that the second number of sample images to be mined in each category is respectively determined according to the category of the third target in the third sample image and the total number of sample images to be mined may include the following operations.
- [0093] The proportion of the third target of each category is determined according to the category of the third target in the third sample image.
- [0094] The sampling proportion of each category is determined according to the proportion of the third target of each category.
- [0095] The second number of sample images to be mined in each category is respectively determined according to the sampling proportion of each category.
- [0096] For example, the proportion f_c of the third target in each category may be determined according to the category of the third target in the annotated third sample image in the training set, and the sampling proportion $\hat{r}_c$ of each category may be calculated according to the proportion f_c by the following formulas:

$$r_c = \max(f_c, t \cdot \exp(f_c / t - 1)) \quad (1)$$

$$\hat{r}_c = \frac{r_c}{\sum_{i \in C} r_i} \quad (2)$$

In formulas (1) and (2), r_c represents the sampling value of the category c ; t is a hyper-parameter, for example, has a value of 0.1; C represents the number of categories; and r_i represents the sampling value of the i -th category in C categories.

- [0097] Through the processing of the formulas (1) and (2), the sampling proportion

corresponding to the category with a smaller proportion may be increased, and the sampling proportion corresponding to the category with a larger proportion may be reduced, so as to alleviate the problem of the imbalance between the quantities of samples of different categories, so as to improve the training effect of the network.

- [0098] In a possible implementation mode, the second number of sample images to be mined in each category may be determined according to the sampling proportion $\hat{f}_c$ of each category and the total number (K) of sample images to be mined. Therefore, feature-relevance mining may be performed according to the second number.
- [0099] In a possible implementation mode, the method further includes that: the third sample image is input into the target detection network for processing to obtain the feature information of the third target in the third sample image.
- [00100] That is, the annotated third sample image in the training set may be input into the target detection network, and the hidden layer (such as a convolution layer) of the target detection network outputs the feature information of the third sample image, such as a feature vector. In this way, the features of the third sample image may be obtained for subsequent feature-relevance mining.
- [00101] In a possible implementation mode, the operation that a fourth target and a first sample image where the fourth target is located are determined from the fifth target according to the feature information of the third target in the third sample image, the feature information of the fifth target and the second number of the sample images to be mined in each category includes the following operations.
- [00102] The third target at the smallest distance from each fifth target among the third target of the first category is respectively determined according to the distance between the feature information of the third target of the first category and the feature information of each fifth target, the third target is taken as an eighth target, where the first category is any one of the categories of the third target.
- [00103] The target at the largest distance among the eighth targets is determined as the fourth target.
- [00104] For example, after the second number of sample images to be mined in each category is determined, the corresponding number of sample images may be mined from the sample images where the fifth target is located by using a k-center mode. For any category (which may be called the first category) of the plurality of categories of the third target, the distance between the feature information of the third target of the first category and the feature information of each fifth target may be calculated, which may be, for example, Euclidean distance. For any fifth target, the third target at the smallest distance from the fifth target among the third targets in the first category may be determined, so that the third target at the smallest distance from each fifth target may be determined respectively, which may be called the eighth target.
- [00105] In a possible implementation mode, the target at the largest distance may be selected from eighth targets and determined as the fourth target obtained by feature-relevance mining at this time. The formula is:

$$u = \arg \max_{j \in \hat{D}_{UL}} \min_{l \in \hat{D}_L^c} \text{dist}(f_j, g_l) \quad (3)$$

In the formula (3), u represents the fourth target obtained by feature-relevance mining; $\text{dist}(f_j, g_l)$ represents the distance between the feature information f_j of the j-th fifth target and the feature information g_l of the first third target of the first category c; $\hat{D}_{UL}$

represents a set of feature information of the fifth target; and $\hat{D}_L^c$ represents a set of feature information of the third target of the first category c.

- [00106] In a possible implementation mode, the first sample image where the fourth target is located may be determined, and the sample image is added to the training set as the fourth sample image, so as to complete the feature-relevance mining process.
- [00107] In a possible implementation mode, the step that a fourth target and a first sample image where the fourth target is located are determined from the fifth target according to the feature information of the third target in the third sample image, the feature information of the fifth target and the second number of the sample images to be mined in each category further includes the following operations.
- [00108] The determined fourth target is added to the third target of the first category, and the determined fourth target is removed from the unannotated fifth target.
- [00109] That is, the fourth target obtained by the feature-relevance mining this time is taken as the annotated target, and the fourth target is removed from the unannotated target. In this case, the feature information of the fourth target may be added to the set $\hat{D}_L^c$ of feature information of the third target of the first category c and removed from the set $\hat{D}_{UL}$ of feature information of the fifth target. In this way, in the next feature-relevance mining, the two updated sets may be mined through the formula (3), and the above process may be repeated.
- [00110] In a possible implementation mode, when the number of the fourth sample images of the first category reaches the second number of the first category, or when the number of the fourth sample images of the first category does not reach the second number and the fifth target is exhausted (the set $\hat{D}_{UL}$ is empty), the feature-relevance mining of the first category may be completed.
- [00111] In this way, feature-relevance mining may be performed for each category respectively, and finally a sufficient number of fourth sample images (usually K sample images) may be obtained, so as to further expand the number of sample images in the training set and alleviate the imbalance between positive and negative samples.
- [00112] In a possible implementation mode, the mined fourth sample image may be subjected to human annotation to obtain the annotation information of the fourth sample image. Since there may be both a positive sample image (that is, the image includes the fourth sample image of the target) and a negative sample image (that is, the image does not include the fourth sample image in the target) in the fourth sample image, the annotation information of the fourth sample image may include sample category information that the image is a positive sample image or a negative sample image, the image frame where the target in the positive sample image is located and the category of the target.
- [00113] In a possible implementation mode, after human annotation is completed, the target detection network may be trained according to the annotation information of the fourth sample image, as well as the second sample image, the third sample image and the fourth sample image in the training set in S15.
- [00114] S15 may include that: the first number of samples from the positive sample images of each category is respectively determined according to the category of the target in the positive sample image of the training set, where the positive sample image is a sample image including the target.

- [00115] The positive sample image of each category is sampled according to the first number of samples in the positive sample images of each category to obtain a plurality of fifth sample images.
- [00116] The negative sample image of the training set is sampled to obtain a plurality of sixth sample images, where the negative sample image is a sample image that does not include the target.
- [00117] The target detection network is trained according to the fifth sample images and the sixth sample images.
- [00118] For example, the target detection network may be trained by resampling, and the sampling frequency of data with relatively low frequency in the data may be increased by resampling, so as to improve the performance of the network for these data and further improve the imbalance between positive and negative samples.
- [00119] In a possible implementation mode, the positive sample images and negative sample images in the training set (including the second sample image, the third sample image and the fourth sample image) may be sampled respectively, so that the number of the positive sample images and the number of the negative sample images after sampling are the same or similar.
- [00120] In a possible implementation mode, for the positive sample images, the total number of samples of the positive sample images may be preset. According to the category of the target in the positive sample images in the training set, the first number of samples from the positive sample image of each category is determined respectively.
- [00121] Similar to the previous processing, the proportion of the target in each category may be determined according to the categories of the target in the positive sample images. According to this proportion, the sampling proportion of each category may be calculated by the following formula:

$$R_h = \begin{cases} t_1 * e^{\frac{q_h - 1}{t_1}} & 0 < q_h \leq t_1 \\ q_h & t_1 < q_h < 1 \end{cases} \quad (4)$$

In the formula (4), R_h represents the sampling proportion of the positive sample image of the h -th category; q_h represents the proportion of the target in the h -th category; and t_1 is a hyper-parameter, for example, has a value of 0.1.

- [00122] Through the processing with the formula (4), the sampling proportion corresponding to the category with a smaller proportion may be increased and the sampling proportion corresponding to the category with a larger proportion may be reduced, so as to alleviate the problem of imbalance between the quantities of positive sample images of different categories, so as to improve the training effect of the network.
- [00123] In a possible implementation mode, the first number of the positive sample images of each category may be determined according to the sampling proportion of the positive sample images of each category and the total sampling number of the positive sample images.
- [00124] In a possible implementation mode, for any category, the first number of positive sample images may be randomly sampled from the positive sample images of the category according to the first number of the category as the fifth sample image. The fifth sample image of the total number of samples may be obtained by sampling the positive sample images of each category respectively.
- [00125] In a possible implementation mode, for the negative sample images, the

negative sample images in the training set may be directly sampled randomly according to the preset total sampling number to obtain the sixth sample image with the total sampling number. The total sampling number of the negative sample image may be the same as or different from the total sampling number of the positive sample image, which is not limited by the embodiments of the disclosure.

- [00126] In a possible implementation mode, the target detection network is trained according to the fifth sample image and the sixth sample image. That is, the fifth and sixth sample images are input into the target detection network respectively to obtain the target detection results of the fifth and sixth sample images. According to the target detection results and annotation information, the loss of the target detection network is determined. The parameter of the target detection network is inversely adjusted according to the loss. After several rounds of iteration, when the preset condition (such as network convergence) is met, the trained target detection network is obtained.
- [00127] In this way, the detection effect of the trained target detection network for long-tailed images may be significantly improved.
- [00128] In a possible implementation mode, before S11, the step of pre-training the target detection network through the annotated third sample image may also be executed through the above resampling training mode, so as to improve the pre-training effect of the target detection network.
- [00129] In actual application, the whole processing process of S11 to S15 may be repeated to realize continuous incremental training. That is, when the unannotated sample image is collected again, the target detection network after this training at this time may be used as the initial target detection network, and the expanded training set at this time may be used as the initial training set to repeat the processing process of pseudo-labeling, feature-relevance mining and resampling training, so as to continuously improve the performance of the target detection network.
- [00130] FIG. 2 is a schematic diagram of a processing process of a network training method according to the embodiments of the disclosure. As illustrated in FIG. 2, a data source includes a large number of unannotated first sample images 20. The first sample images 20 are input into the target detection network for prediction to obtain a target detection result 21 of each first sample image 20, including the image area (not shown), feature vector and class probability of the target in the first sample image.
- [00131] As illustrated in FIG. 2, in the example, the target detection network may include a CNN backbone network 211, a Feature-map Pyramid Network (FPN) 212, and a fully connected network 213. The fully connected network 213 is for example a bbox head. After the first sample images 20 are input into the target detection network, they are processed through the CNN backbone network 211 and the FPN 212 to obtain feature maps 214 of the first sample images, and the feature maps 214 are processed through the fully connected network 213 to obtain the target detection result 21.
- [00132] In the example, the category confidence of the target may be determined according to the class probability of the target. For the first target whose category confidence is greater than or equal to the first threshold value (for example, 0.99), the first sample images where these first targets are located are determined as second sample images 22, and pseudo-labeling is performed on the second sample images 22, that is, the image area of the first target and the category corresponding to the category confidence of the first target are taken as the annotation information of the second sample images 22. The annotated second sample images 22 are added to the training set 25 to expand the positive samples in the training set.

- [00133] In the example, for the second target whose category confidence is less than the first threshold value, a certain number of fifth targets are selected by a bootstrapping method to obtain the sample images 23 where the fifth target is located. According to the feature vector (not shown) of the third target in the annotated third sample images in the training set, feature-relevance mining is performed on the fifth target to determine the fourth target and the first sample images where the fourth target is located as the fourth sample images 24. The fourth sample images 24 are subjected to human annotation and added to the training set 25, so as to further expand the annotated images in the training set.
- [00134] In the example, after two extensions, the training set 25 includes the annotated second sample images, the annotated third sample images and the annotated fourth sample images. The training set 25 is resampled, the quantities of positive and negative samples are balanced, and the quantities of the positive samples in different categories are balanced to obtain the resampled training set 26. Then, according to the resampled training set 26, the target detection network is trained to complete the whole processing process.
- [00135] According to the embodiments of the disclosure, a target detection method is further provided, including the following operations.
- [00136] An image to be processed is input into a target detection network for processing to obtain a target detection result of the image to be processed, the target detection result includes the position and category of a target in the image to be processed. The target detection network is obtained by training according to the above network training method.
- [00137] That is, the target detection network trained by the above method may be deployed to realize the target detection of the image to be processed. The image to be processed may be, for example, an image collected by an image collection device (such as a camera), and the image may include a target to be detected, such as a human body, a human face, a vehicle and an object. There is no limitation to this in the embodiments of the disclosure.
- [00138] In a possible implementation mode, the image to be processed may be input into the target detection network for processing to obtain the target detection result of the image to be processed. The target detection result includes the position and category of the target in the image to be processed, such as the detection frame where the human face in the image to be processed is located and the identity corresponding to the human face.
- [00139] In this way, the detection accuracy of target detection may be improved and the target detection of large-scale long-tailed image data may be realized.
- [00140] According to the network training method of the embodiments of the disclosure, an active learning mining method is used to mine potential unannotated data, a weakly-supervised learning method is used to assist in annotating the unannotated data, and the number of positive sample data is expanded, so as to solve the problem of large data scale and difficult positive sample collection in large-scale long-tailed detection, and it alleviates the imbalance between positive and negative samples to a certain degree. It effectively improves the performance of a model in the environment of limited annotation and computing resources.
- [00141] According to the network training method of the embodiments of the disclosure, the target detection network is trained by resampling, which may solve the negative impact of the imbalance between positive and negative samples on network training, and alleviate the negative impact of the imbalance between different categories

of positive samples on network training, so that the target detection network may effectively converge and improve the network performance during training.

[00142] According to the network training method of the embodiments of the disclosure, the active learning method may mine potential high-value samples that are helpful for model improvement in a large amount of unannotated data, effectively improve the model performance in the environment of limited annotation and computing resource, and save a lot of manpower and computing costs required for the application of a deep learning model in new businesses. The resampling method may effectively train the target detection network in the case of unbalanced samples without too much manual parameter adjustment intervention, and save the labor cost required for the application of the deep learning model in new business.

[00143] The network training method according to the embodiments of the disclosure may be applied to the fields of intelligent video analysis, security, etc. Under limited manual and computing resources, the method may be used to detect potential targets in intelligent video analysis or intelligent monitoring online, and quickly iterate and improve the corresponding detection network. With less manpower and computing cost, it can quickly meet the performance requirements required by the business, and continuously improve the network performance in the future.

[00144] The network training method of the embodiments of the disclosure may be applied to online intelligent video analysis or intelligent monitoring, so as to quickly iterate and improve the potential target detection application in intelligent video analysis or intelligent monitoring online with limited manpower and computing resources, so as to quickly meet the performance requirements required by the business with less manpower and computing cost, and may continue to improve the performance of the model in the future.

[00145] It can be understood that various method embodiments mentioned in the embodiments of the disclosure may be combined to form combined embodiments without departing from principles and logics. For saving the space, elaborations are omitted in the embodiments of the disclosure. It can be understood by those skilled in the art that, in the above-mentioned method of the specific implementation modes, a specific execution sequence of each step should be determined by functions and probable internal logic thereof.

[00146] In addition, the embodiments of the disclosure further provide a network training apparatus, a target detection apparatus, an electronic device, a computer-readable storage medium, and a program. All of the above can be used to implement any network training method or target detection method provided by the embodiments of the disclosure. Corresponding technical solutions, description, and the corresponding records referring to a method part will not be elaborated herein.

[00147] FIG. 3 is a block diagram of a network training apparatus according to the embodiments of the disclosure. The apparatus includes a processor (not shown in FIG. 3) for executing a program part stored in a memory (not shown in FIG. 3). As illustrated in FIG. 3, the program part stored in the memory includes a target detection part 31, a confidence determination part 32, an annotation part 33, a feature mining part 34 and a training part 35.

[00148] The target detection part 31 is configured to input an unannotated first sample image into a target detection network for processing to obtain a target detection result of the first sample image, and the target detection result includes an image area, feature information and class probability of a target in the first sample image.

- [00149] The confidence determination part 32 is configured to determine the category confidence of the target according to the class probability of the target.
- [00150] The annotation part 33 is configured to, for a first target with the category confidence in the target greater than or equal to a first threshold value, take the first sample image where the first target is located as an annotated second sample image and add the annotated second sample image to a training set, where the annotation information of the second sample image includes the image area of the first target and the category corresponding to the category confidence of the first target, and the training set includes an annotated third sample image.
- [00151] The feature mining part 34 is configured to, for a second target with the category confidence in the target less than the first threshold value, perform feature-relevance mining on the second target according to the feature information of a third target in the third sample image, determine a fourth target and the first sample image where the fourth target is located from the second target through feature-relevance mining, take the first sample image where the fourth target is located as the fourth sample image, and add the fourth sample image to the training set.
- [00152] The training part 35 is configured to train the target detection network according to the annotation information of the fourth sample image, as well as the second sample image, the third sample image and the fourth sample image in the training set.
- [00153] In a possible implementation mode, the training part includes a sampling number determination subpart, a first sampling subpart, a second sampling subpart and a training subpart. The sampling number determination subpart is configured to respectively determine the first number of samples from a positive sample image of each category according to the category of the target in positive sample images of the training set, where the positive sample image is a sample image including the target. The first sampling subpart is configured to sample the positive sample image of each category according to the first number of the samples in the positive sample image of each category to obtain a plurality of fifth sample images. The second sampling subpart is configured to sample a negative sample image of the training set to obtain a plurality of sixth sample images, where the negative sample image is a sample image that does not include the target. The training subpart is configured to train the target detection network according to the fifth sample images and the sixth sample images.
- [00154] In a possible implementation mode, the feature mining part includes an information entropy determination subpart, a target selection subpart, a mining number determination subpart and a target and image determination subpart. The information entropy determination subpart is configured to determine information entropy of the second target according to the class probability of the second target. The target selection subpart is configured to select a fifth target from the second target according to the category confidence and information entropy of the second target. The mining number determination subpart is configured to respectively determine the second number of sample images to be mined in each category according to the category of the third target in the third sample image and the total number of sample images to be mined. The target and image determination subpart is configured to determine the fourth target and the first sample image where the fourth target is located from the fifth target according to the feature information of the third target in the third sample image, the feature information of the fifth target and the second number of the sample images to be mined in each category.
- [00155] In a possible implementation mode, the target selection subpart is configured

to sort the second target according to the category confidence and the information entropy of the second target respectively, and select a third number of sixth targets and a fourth number of seventh targets; and combine the sixth targets and the seventh targets to obtain the fifth target.

- [00156] In a possible implementation mode, the mining number determination subpart is configured to determine the proportion of the third target of each category according to the category of the third target in the third sample image, determine the sampling proportion of each category according to the proportion of the third target of each category, and respectively determine the second number of the sample images to be mined in each category according to the sampling proportion of each category.
- [00157] In a possible implementation mode, the target and image determination subpart is configured to respectively determine the third target at the smallest distance from each fifth target among the third targets of the first category according to the distance between the feature information of the third target of the first category and the feature information of each fifth target, take the third target as an eighth target, and determine the target with the largest distance among the eighth targets as the fourth target. The first category is any one of the categories of the third target.
- [00158] In a possible implementation mode, the target and image determination subpart is further configured to add the determined fourth target to the third target of the first category, and remove the determined fourth target from the unannotated fifth target.
- [00159] In a possible implementation mode, the apparatus further includes a feature extraction part, which is configured to input the third sample image into the target detection network for processing to obtain the feature information of the third target in the third sample image.
- [00160] In a possible implementation mode, the apparatus further includes a pre-training part, which is configured to pre-train the target detection network through the labeled third sample image.
- [00161] In a possible implementation mode, the first sample image includes a long-tailed image.
- [00162] In a possible implementation mode, the sampling number determination subpart is further configured to sample the positive sample images and the negative sample image in the training set to obtain the same or similar number of positive sample images and negative sample image before the first number of samples from the positive sample image of each category is respectively determined according to the category of the target in the positive sample images of the training set.
- [00163] In a possible implementation mode, the total number of the sample images to be mined is 5% to 25% of the total number of the first sample images.
- [00164] In a possible implementation mode, the target selection subpart is further configured to remove the same target as the seventh targets from the sixth target to obtain the remaining target in the sixth targets different from the seventh targets, and take the remaining target and the seventh targets as the fifth target.
- [00165] In a possible implementation mode, the method further includes that: after the third target at the smallest distance from each fifth target among the third target of the first category is respectively determined according to the distance between the feature information of the third target of the first category and the feature information of each fifth target, and the third target is taken as an eighth target, when the number of the first sample image where the fourth target is located reaches the second number of the sample images to be mined of the first category, the determination of the eighth target is ended.

- [00166] In a possible implementation mode, the target and image determination subpart is further configured to, after the third target at the smallest distance from each fifth target among the third target of the first category is respectively determined according to the distance between the feature information of the third target of the first category and the feature information of each fifth target, and the third target is taken as an eighth target, end the determination of the eighth target when the number of the first sample images where the fourth target is located does not reach the second number of the sample images to be mined of the first category and a set storing the feature information of the fifth target is empty.
- [00167] In a possible implementation mode, the feature extraction part is further configured to input the third sample image into the target detection network to obtain a feature vector output by a hidden layer of the target detection network, and determine the feature vector as the feature information of the third target.
- [00168] According to an aspect of the disclosure, a target detection apparatus is provided, which includes a detection processing part, which is configured to input an image to be processed into a target detection network for processing to obtain a target detection result of the image to be processed, the target detection result includes the position and category of a target in the image to be processed, where the target detection network is obtained by training according to the above network training method.
- [00169] In some embodiments, the functions or parts of the apparatus provided by the embodiments of the disclosure may be used to execute the method described in the above method embodiments, and its specific implementation may refer to the description of the above method embodiment. For simplicity, it will not be elaborated herein.
- [00170] In some embodiments, the "part" may also be a part of the circuit, a part of the processor, a part of the program or software, etc., of course, it may also be a unit, a module or a non modular.
- [00171] The embodiments of the disclosure further provide a computer-readable storage medium, which stores a computer program instruction. The above method is implemented when the computer program instruction is executed by a processor. The computer-readable storage medium may be a nonvolatile computer readable storage medium.
- [00172] The embodiments of the disclosure further provide an electronic device, including a processor, and a memory configured to store a processor executable instruction. The processor is configured to call the instruction stored in the memory to implement the above method.
- [00173] The embodiments of the disclosure further provide a computer program product, including a computer-readable code. When the computer-readable code runs on a device, a processor in the device executes an instruction to implement the network training method or a target detection method provided by any one of the above embodiments.
- [00174] The embodiments of the disclosure further provide another computer program product configured to store a computer-readable instruction. When the instruction is executed, a computer executes the operations of the network training method or the target detection method provided by any one of the previous embodiments.
- [00175] The electronic device may be provided as a terminal, a server, or other forms of devices.
- [00176] FIG. 4 is a block diagram of an electronic device 800 according to the embodiments of the disclosure. For example, the electronic device 800 may be a mobile

phone, a computer, a digital broadcast terminal, a message receiving and transmitting device, a gaming console, a tablet device, a medical device, fitness equipment, a personal digital assistant, or the like.

- [00177] Referring to FIG. 4, the electronic device 800 may include one or more of the following components: a processing component 802, a memory 804, a power component 806, a multimedia component 808, an audio component 810, an Input/Output (I/O) interface 812, a sensor component 814, and a communication component 816.
- [00178] The processing component 802 typically controls an overall operation of the electronic device 800, such as the operations associated with displays, telephone calls, data communications, camera operations, and recording operations. The processing component 802 may include one or more processors 820 to execute instructions to perform all or part of the steps in the above-mentioned method. Moreover, the processing component 802 may include one or more modules which facilitate interaction between the processing component 802 and the other components. For example, the processing component 802 may include a multimedia module to facilitate interaction between the multimedia component 808 and the processing component 802.
- [00179] The memory 804 is configured to store various types of data to support the operation of the electronic device 800. Examples of such data include instructions for any applications or methods operated on the electronic device 800, contact data, phonebook data, messages, pictures, video, etc. The memory 804 may be implemented by any type of volatile or non-volatile memory devices, or a combination thereof, such as a Static Random Access Memory (SRAM), an Electrically Erasable Programmable Read-Only Memory (EEPROM), an Erasable Programmable Read-Only Memory (EPROM), a Programmable Read-Only Memory (PROM), a Read-Only Memory (ROM), a magnetic memory, a flash memory, and a magnetic disk or a compact disk.
- [00180] The power component 806 provides power for various components of the electronic device 800. The power component 806 may include a power management system, one or more power supplies, and other components associated with generation, management and distribution of power for the electronic device 800.
- [00181] The multimedia component 808 includes a screen providing an output interface between the electronic device 800 and a user. In some embodiments, the screen may include a Liquid Crystal Display (LCD) and a Touch Panel (TP). If the screen includes the TP, the screen may be implemented as a touch screen to receive input signals from the user. The TP includes one or more touch sensors to sense touches, swipes, and gestures on the TP. The touch sensors may not only sense a boundary of a touch or swipe action but also detect a duration and pressure associated with the touch or swipe action. In some embodiments, the multimedia component 808 includes a front camera and/or a rear camera. A front camera and/or a rear camera may receive external multimedia data when the electronic device 800 is in an operation mode, for example, a photographing mode or a video mode. Each of the front camera and the rear camera may be a fixed optical lens system or have focusing and optical zooming capabilities.
- [00182] The audio component 810 is configured to output and/or input an audio signal. For example, the audio component 810 includes a Microphone (MIC). The MIC is configured to receive an external audio signal when the electronic device 800 is in the operation mode, such as a call mode, a recording mode, and a voice recognition mode. The received audio signal may further be stored in the memory 804 or sent through the communication component 816. In some embodiments, the audio component 810 further includes a speaker configured to output the audio signal.

- [00183] The I/O interface 812 provides an interface between the processing component 802 and a peripheral interface module, and the peripheral interface module may be a keyboard, a click wheel, a button, etc. The button may include, but is not limited to: a home button, a volume button, a starting button, and a locking button.
- [00184] The sensor component 814 includes one or more sensors configured to provide state assessment in various aspects for the electronic device 800. For example, the sensor component 814 may detect an on/off state of the electronic device 800 and relative positioning of components. For example, the components are a display and a small keyboard of the electronic device 800. The sensor component 814 may also detect a change in a position of the electronic device 800 or a component of the electronic device 800, presence or absence of contact between the user and the electronic device 800, orientation or acceleration/deceleration of the electronic device 800 and a change in temperature of the electronic device 800. The sensor component 814 may include a proximity sensor configured to detect presence of an object nearby without any physical contact. The sensor component 814 may also include a light sensor, such as a Complementary Metal Oxide Semiconductor (CMOS) or a Charge Coupled Device (CCD) image sensor, configured for use in an imaging application. In some embodiments, the sensor component 814 may also include an acceleration sensor, a gyroscope sensor, a magnetic sensor, a pressure sensor or a temperature sensor.
- [00185] The communication component 816 is configured to facilitate wired or wireless communication between the electronic device 800 and another device. The electronic device 800 may access a communication-standard-based wireless network, such as a Wireless Fidelity (Wi-Fi) network, the 2nd-Generation (2G) wireless communication technology, or the 3rd-Generation (3G) wireless communication technology, or a combination thereof. In one exemplary embodiment, the communication component 816 receives a broadcast signal or broadcast associated information from an external broadcast management system through a broadcast channel. In an exemplary embodiment, the communication component 816 further includes a Near Field Communication (NFC) module to facilitate short-range communication. For example, the NFC module may be implemented based on a Radio Frequency Identification (RFID) technology, an Infrared Data Association (IrDA) technology, an Ultra Wide Band (UWB) technology, a Bluetooth (BT) technology and another technology.
- [00186] In an exemplary embodiment, the Device 800 may be implemented by one or more ASICs (Application Specific Integrated Circuits), Digital Signal Processors (DSPs), Digital Signal Processing Devices (DSPDs), Programmable Logic Devices (PLDs), Field Programmable Gate Arrays (FPGAs), controllers, micro-controllers, microprocessors or other electronic elements, and is configured to execute the above-mentioned method.
- [00187] In the exemplary embodiment, a nonvolatile computer-readable storage medium is also provided, for example, a memory 804 including a computer program instruction. The computer program instruction may be executed by a processor 820 of an electronic device 800 to implement the above-mentioned method.
- [00188] FIG. 5 is a block diagram of an electronic device 1900 according to the embodiments of the disclosure. For example, the electronic device 1900 may be provided as a server. Referring to FIG. 5, the electronic device 1900 includes a processing component 1922, which further includes one or more processors, and a memory resource represented by a memory 1932, configured to store an instruction, e.g., an APP, executable for the processing component 1922. The APP stored in the memory 1932 may include one or more modules, each of which corresponds to a set of instructions. In

addition, the processing component 1922 is configured to execute instructions to execute the above-mentioned method.

[00189] The electronic device 1900 may further include a power component 1926 configured to execute power management of the electronic device 1900, a wired or wireless network interface 1950 configured to connect the electronic device 1900 to a network, and an I/O interface 1958. The electronic device 1900 may be operated based on an operating system stored in the memory 1932, for example, Windows Server™, Mac OS X™ launched by Apple Inc, Unix™, Linux™, FreeBSD™ or the like.

[00190] In the illustrative embodiment, a nonvolatile computer-readable storage medium is also provided, for example, a memory 1932 including a computer program instruction. The computer program instruction may be executed by a processor 1922 of an electronic device 1900 to implement the above-mentioned method.

[00191] The embodiments of the disclosure may be systems, methods, and/or computer program products. The computer program products may include a computer-readable storage medium, in which a computer-readable program instruction configured to enable a processor to implement each aspect of the embodiments of the disclosure is carried.

[00192] The computer-readable storage medium may be a physical device capable of retaining and storing an instruction used by an instruction execution device. The computer-readable storage medium may be, but not limited to, an electric storage device, a magnetic storage device, an optical storage device, an electromagnetic storage device, a semiconductor storage device, or any appropriate combination thereof. More specific examples (non-exhaustive list) of the computer-readable storage medium include a portable computer disk, a hard disk, a RAM, a ROM, an EPROM (or a flash memory), an SRAM, a Compact Disc Read-Only Memory (CD-ROM), a Digital Video Disk (DVD), a memory stick, a floppy disk, a mechanical encoding device, for example, a punched card or in-slot raised structure with an instruction stored therein, and any appropriate combination thereof. Herein, the computer-readable storage medium is not explained as a transient signal, for example, a radio wave or another freely propagated electromagnetic wave, an electromagnetic wave propagated through a wave guide or another transmission medium (for example, a light pulse propagated through an optical fiber cable) or an electric signal transmitted through an electric wire.

[00193] The computer-readable program instruction described here may be downloaded from the computer-readable storage medium to each computing/processing device or downloaded to an external computer or an external storage device through a network such as the Internet, a Local Area Network (LAN), a Wide Area Network (WAN) and/or a wireless network. The network may include a copper transmission cable, optical fiber transmission, wireless transmission, a router, a firewall, a switch, a gateway computer and/or an edge server. A network adapter card or network interface in each computing/processing device receives the computer-readable program instruction from the network and forwards the computer-readable program instruction for storage in the computer-readable storage medium in each computing/processing device.

[00194] The computer program instruction configured to execute the operations of the disclosure may be an assembly instruction, an Instruction Set Architecture (ISA) instruction, a machine instruction, a machine related instruction, a microcode, a firmware instruction, state setting data or a source code or target code edited by one or any combination of more programming languages, the programming language including an object-oriented programming language such as Smalltalk and C++ and a conventional procedural programming language such as "C" language or a similar programming

language. The computer-readable program instruction may be completely executed in a computer of a user or partially executed in the computer of the user, executed as an independent software package, executed partially in the computer of the user and partially in a remote computer, or executed completely in the remote server or a server. Under the condition that the remote computer is involved, the remote computer may be concatenated to the computer of the user through any type of network including an LAN or a WAN, or may be concatenated to an external computer (for example, concatenated by an Internet service provider through the Internet). In some embodiments, an electronic circuit such as a programmable logic circuit, an FPGA, or a Programmable Logic Array (PLA) may be customized by use of state personal information of a computer-readable program instruction, and the electronic circuit may execute the computer-readable program instruction, thereby implementing each aspect of the disclosure.

[00195] Herein, each aspect of the disclosure is described with reference to flowcharts and/or block diagrams of the method, device (system) and computer program product according to the embodiments of the disclosure. It is to be understood that each block in the flowcharts and/or the block diagrams and a combination of each block in the flowcharts and/or the block diagrams may be implemented by computer-readable program instructions.

[00196] These computer-readable program instructions may be provided for a universal computer, a dedicated computer or a processor of another programmable data processing device, thereby generating a machine to further generate a device that realizes a function/action specified in one or more blocks in the flowcharts and/or the block diagrams when the instructions are executed through the computer or the processor of the other programmable data processing device. These computer-readable program instructions may also be stored in a computer-readable storage medium, and through these instructions, the computer, the programmable data processing device and/or another device may work in a specific mode, so that the computer-readable medium including the instructions includes a product including instructions for implementing each aspect of the function/action specified in one or more blocks in the flowcharts and/or the block diagrams.

[00197] These computer program instructions can also be loaded to a computer or another programmable data processing device, so that a series of operating steps are performed on the computer or the another programmable data processing device to produce a computer-implemented process, and therefore instructions executed on the computer or the another programmable data processing device provide steps for implementing functions specified in one or more flows in the flowcharts and/or one or more blocks in the block diagrams.

[00198] The flowcharts and block diagrams in the drawings illustrate probably implemented system architectures, functions and operations of the system, method, and computer program product according to multiple embodiments of the disclosure. On this aspect, each block in the flowcharts or the block diagrams may represent part of a module, a program segment or an instruction, and part of the module, the program segment or the instruction includes one or more executable instructions configured to realize a specified logical function. In some alternative implementations, the functions marked in the blocks may also be realized in a sequence different from those marked in the drawings. For example, two continuous blocks may actually be executed substantially concurrently and may also be executed in a reverse sequence sometimes, which is determined by the involved functions. It is further to be noted that each block in the block

diagrams and/or the flowcharts and a combination of the blocks in the block diagrams and/or the flowcharts may be implemented by a dedicated hardware-based system configured to execute a specified function or operation or may be implemented by a combination of a special hardware and a computer instruction.

[00199] The computer program product may be specifically realized by means of hardware, software or a combination thereof. In an optional embodiment, the computer program product is specifically embodied as a computer storage medium, and in another optional embodiment, the computer program product is specifically embodied as software products, such as a Software Development Kit (SDK).

[00200] Each embodiment of the disclosure has been described above. The above descriptions are exemplary, non-exhaustive and also not limited to each disclosed embodiment. Many modifications and variations are apparent to those of ordinary skill in the art without departing from the scope and spirit of each described embodiment of the disclosure. The terms used herein are selected to explain the principle and practical application of each embodiment or improvements in the technologies in the market best or enable others of ordinary skill in the art to understand each embodiment disclosed herein.

INDUSTRIAL APPLICABILITY

[00201] The embodiments of the disclosure relates to a network training method and apparatus, a target detection method and apparatus, and an electronic device. The network training method includes that: an unannotated sample image is input into a target detection network for processing to obtain a target detection result, and the result includes an image area, feature information and class probability of a target; the category confidence of the target is determined according to the class probability of the target; for a first target whose category confidence is greater than or equal to a first threshold value, the sample image where the first target is located is taken as an annotated image and is added to a training set; for a second target whose category confidence is less than the first threshold value, feature-relevance mining on the second target is executed, a fourth target is determined from the second target, and a sample image where the fourth target is located is added to the training set; and the target detection network is trained according to the sample image in the training set. The embodiments of the disclosure may improve the training effect of the target detection network.

CLAIMS

1. A network training method, comprising:

inputting an unannotated first sample image into a target detection network for processing to obtain a target detection result of the first sample image, the target detection result comprising an image area, feature information and class probability of a target in the first sample image;

determining category confidence of the target according to the class probability of the target;

for a first target with the category confidence greater than or equal to a first threshold value in the target, taking the first sample image where the first target is located as an annotated second sample image, and adding the second sample image to a training set, wherein annotation information of the second sample image comprises the image area of the first target and a category corresponding to the category confidence of the first target, and the training set comprises an annotated third sample image;

for a second target with the category confidence less than the first threshold value in the target, performing feature-relevance mining on the second target according to the feature information of a third target in the third sample image, determining a fourth target and the first sample image where the fourth target is located from the second target through feature-relevance mining, taking the first sample image where the fourth target is located as the fourth sample image, and adding the fourth sample image to the training set; and

training the target detection network according to the annotation information of the fourth sample image, as well as the second sample image, the third sample image and the fourth sample image in the training set.

2. The method according to claim 1, wherein training the target detection network according to the annotation information of the fourth sample image, as well as the second sample image, the third sample image and the fourth sample image in the training set comprises:

respectively determining a first number of samples from a positive sample image of each category according to the category of the target in positive sample images of the training set, wherein the positive sample image is a sample image comprising the target;

sampling the positive sample image of the each category according to the first number of the samples in the positive sample image of the each category to obtain a plurality of fifth sample images;

sampling a negative sample image of the training set to obtain a plurality of sixth sample images, wherein the negative sample image is a sample image that does not comprise the target; and

training the target detection network according to the fifth sample images and the sixth sample images.

3. The method according to claim 1 or 2, wherein performing feature-relevance mining on the second target according to the feature information of the third target in the third sample image, and determining the fourth target and the first sample image where the fourth target is located from the second target through feature-relevance mining comprise:

determining information entropy of the second target according to the class probability of the second target;

selecting a fifth target from the second target according to the category confidence and the information entropy of the second target;

respectively determining a second number of sample images to be mined in each category according to the category of the third target in the third sample image and a total number of sample images to be mined; and

determining the fourth target and the first sample image where the fourth target is located from the fifth target according to the feature information of the third target in the third sample image, feature information of the fifth target and the second number of the sample images to be mined in the each category.

4. The method according to claim 3, wherein selecting the fifth target from the second target according to the category confidence and the information entropy of the second target comprises:

sorting the second target according to the category confidence and the information entropy of the second target respectively , and selecting a third number of sixth targets and a fourth number of seventh targets; and

combining the sixth targets and the seventh targets to obtain the fifth target.

5. The method according to claim 3 or 4, wherein respectively determining the second number of the sample images to be mined in the each category according to the category of the third target in the third sample image and the total number of the sample images to be mined comprises:

determining a proportion of the third target of each category according to the category of the third target in the third sample image;

determining the sampling proportion of each category according to the proportion of the third target of the each category; and

respectively determining the second number of the sample images to be mined in the each category according to the sampling proportion of the each category.

6. The method according to any one of claims 3 to 5, wherein determining the fourth target and the first sample image where the fourth target is located from the fifth target according to the feature information of the third target in the third sample image, the feature information of the fifth target and the second number of the sample images to be mined in the each category comprises:

respectively determining the third target at a smallest distance from each fifth target among the third target of a first category according to the distance between the feature information of the third target of the first category and the feature information of the each fifth target, and taking the third target as an eighth target, wherein the first category is any one of categories of the third target; and

determining a target at a largest distance among the eighth target as the fourth target.

7. The method according to claim 6, wherein determining the fourth target and the first sample image where the fourth target is located from the fifth target according to the feature information of the third target in the third sample image, the feature information of the fifth target and the second number of the sample images to be mined in the each category further comprises:

adding the determined fourth target to the third target of the first category, and removing the determined fourth target from the unannotated fifth target.

8. The method according to any one of claims 1 to 7, further comprising:

inputting the third sample image into the target detection network for processing to obtain the feature information of the third target in the third sample image.

9. The method according to any one of claims 1 to 8, wherein before the step of inputting the unannotated first sample image into the target detection network for processing to obtain the target detection result of the first sample image, the method further comprises:

pre-training the target detection network through the annotated third sample image.

10. The method according to any one of claims 1 to 9, wherein the first sample image comprises a long-tailed image.

11. The method according to claim 2, wherein before the step of respectively determining the first number of samples from the positive sample image of the each category according to the category of the target in the positive sample images of the training set, the method comprises:

sampling the positive sample images and the negative sample image in the training set to obtain a same or similar number of positive sample images and negative sample images.

12. The method according to claim 3, wherein the total number of the sample images to be mined is 5% to 25% of a total number of the first sample images.

13. The method according to claim 4, wherein combining the sixth targets and the seventh targets to obtain the fifth target comprises:

removing a same target as the seventh targets from the sixth targets to obtain a remaining target in the sixth targets different from the seventh targets; and

taking the remaining target and the seventh targets as the fifth target.

14. The method according to claim 6, wherein after the steps of respectively determining the third target at the smallest distance from the each fifth target among the third target of the first category according to the distance between the feature information of the third target of the first category and the feature information of the each fifth target, and taking the third target as an eighth target, the method further comprises:

ending the determination of the eighth target in a case where a number of the first sample image where the fourth target is located reaches the second number of the sample images to be mined of the first category.

15. The method according to claim 7, wherein after the steps of respectively determining the third target at the smallest distance from each fifth target among the third target of the first category according to the distance between the feature information of the third target of the first category and the feature information of the each fifth target, and taking the third target as an eighth target, the method further comprises:

ending the determination of the eighth target in a case where the number of the first sample images where the fourth target is located does not reach the second number of the sample images to be mined of the first category and a set storing the feature information of the fifth target is empty.

16. The method according to claim 8, wherein inputting the third sample image into the target detection network for processing to obtain the feature information of the third target in the third sample image comprises:

inputting the third sample image into the target detection network to obtain a feature vector output by a hidden layer of the target detection network; and
determining the feature vector as the feature information of the third target.

17. A target detection method, comprising:

inputting an image to be processed into a target detection network for processing to obtain a target detection result of the image to be processed, wherein the target detection result comprises a position and a category of a target in the image to be processed;
wherein the target detection network is obtained by training according to the network training method according to any one of claims 1-10.

18. A network training apparatus, comprising:

a target detection part, configured to input an unannotated first sample image into a target detection network for processing to obtain a target detection result of the first sample image, wherein the target detection result comprises an image area, feature information and class probability of a target in the first sample image;

a confidence determination part, configured to determine category confidence of the target according to the class probability of the target;

an annotation part, configured to, for a first target with the category confidence greater than or equal to a first threshold value in the target, take the first sample image where the first target is located as an annotated second sample image and add the annotated second sample image to a training set, wherein annotation information of the second sample image comprises the image area of the first target and a category corresponding to the category confidence of the first target, and the training set includes an annotated third sample image;

a feature mining part, configured to, for a second target with the category confidence less than the first threshold value in the target, perform feature-relevance mining on the second target according to the feature information of a third target in the third sample image, determine a fourth target and the first sample image where the fourth target is located from the second target through feature-relevance mining, take the first sample image where the fourth target is located as the fourth sample image, and add the fourth sample image to the training set; and

a training part, configured to train the target detection network according to the annotation information of the fourth sample image, as well as the second sample image, the third sample image and the fourth sample image in the training set.

19. An electronic device, comprising:

a processor; and

a memory, configured to store a processor executable instruction;

wherein the processor is configured to call the instruction stored in the memory to implement the network training method according to any one of claims 1 to 16 or implement the target detection method according to claim 17.

20. A computer-readable storage medium, having a computer program instruction stored thereon, wherein the computer program instruction is executed by a processor to implement the network training method according to any one of claims 1 to 16 or implement the target detection method according to claim 17.