

METHOD AND APPARATUS FOR SAMPLE LABELING, AND METHOD AND APPARATUS FOR IDENTIFYING DAMAGE CLASSIFICATION

5

TECHNICAL FIELD

[0001] Embodiments of the present specification relate to the field of model training, and in particular to a method and apparatus for sample labeling, and a method and apparatus for identifying a damage classification.

BACKGROUND

[0002] In a vehicle insurance claim settlement scenario, user clients for claim settlement are commonly used in the current industry, which automatically identifies damage information such as damage components, and damage classification, and provides relevant claim information, such as repair schemes, amounts of compensation and the like according to an image of an accident vehicle uploaded by a user. Due to the absence of manual damage inspection, loss assessment, and loss verification, this scheme reduces the cost of insurance companies and improves the experience of ordinary users in vehicle insurance claim settlement.

[0003] In the current industry, when identifying damage information of the accident vehicle in an image, the image is usually compared with massive images in a historical database to identify a similar image in order to determine the damage information for the accident vehicle based on damage information corresponding to the similar image. This method requires a large quantity of calculations, and damage information determined based on this method has low accuracy. As such, there is a need for an appropriate method that determines the damage information for a vehicle in an image more quickly and accurately.

SUMMARY

[0004] In one method for sample labeling according to the present disclosure, massive loss assessment images are clustered first. Next, a received labeling result of manually labeling a portion of cluster centers with damage classifications is used to automatically label other images in class clusters containing the portion of cluster centers, thereby quickly obtaining massive weak label training samples, and accordingly a damage identification model can be trained on the basis of the weak label training samples.

[0005] According to a first aspect, a method for sample labeling is provided. The method comprises: acquiring a plurality of historical loss assessment images; determining, on the basis of a saliency detection, a plurality of candidate damage areas from the plurality of historical loss assessment images, and regarding images of the plurality of candidate damage areas as a plurality of candidate samples; clustering the plurality of candidate samples into a plurality of class clusters, and determining a plurality of center candidate sample sets respectively corresponding to class cluster centers of the plurality of class clusters; receiving a manual labeling result, the manual labeling result being obtained by performing damage classification labeling on at least a portion of the center candidate sample sets from the plurality of center candidate sample sets, wherein respective center candidate samples in the same center candidate sample set are labeled with the same damage classification; and performing, according to the manual labeling result, damage classification labeling on other candidate samples in the class clusters to which the at least a portion of center candidate sample sets belong, and obtaining a plurality of labeled samples labeled with damage classifications.

[0006] According to one embodiment, determining a plurality of candidate damage areas from the plurality of historical loss assessment images comprises: detecting a salient object in each of the plurality of historical loss assessment images; and determining a minimum rectangular area where the salient object is

located, and regarding the minimum rectangular area as a corresponding candidate damage area.

5 **[0007]** According to one embodiment, clustering the plurality of candidate samples into a plurality of class clusters comprises: determining, on the basis of an image feature extraction model, a feature vector corresponding to each candidate sample of the plurality of candidate samples so as to obtain a plurality of feature vectors; and clustering the plurality of feature vectors to obtain the plurality of class clusters.

10 **[0008]** According to one embodiment, determining a plurality of center candidate sample sets respectively corresponding to class cluster centers of the plurality of class clusters comprises: determining a predetermined number of center candidate samples in each class cluster closest to the corresponding class cluster center, and constructing a corresponding center candidate sample set using the predetermined number of center candidate samples.

15 **[0009]** According to one embodiment, the method further comprises: use the plurality of labeled samples as a first training sample set to train a damage identification model, the damage identification model being used to identify damage classifications of vehicle damages from images.

20 **[0010]** According to one embodiment, the method further comprises: selecting at least a portion of candidate samples from the plurality of candidate samples; predicting, on the basis of the damage identification model, a damage classification corresponding to each candidate sample from the at least a portion of candidate samples; receiving a manual correction result of manual corrections applied to the damage classifications of the at least a portion of candidate samples; labeling the at least a portion of candidate samples with damage classifications on the basis of the manual correction result; and use the at least a portion of candidate samples labeled with the damage classifications as a second training sample set to perform further training on the damage identification model.

30 **[0011]** According to a second aspect, a method for identifying a damage

classification is provided. The method comprises: acquiring a current loss assessment image; and inputting the current loss assessment image to the damage identification model of any of the embodiments of the first aspect, and predicting a damage classification corresponding to the current loss assessment image.

5 **[0012]** According to a third aspect, an apparatus for sample labeling is provided. The apparatus comprises: an acquisition unit configured to acquire a plurality of historical loss assessment images; a first determining unit configured to determine, on the basis of saliency detection, a plurality of candidate damage areas from the plurality of historical loss assessment images, and regard images of
10 the plurality of candidate damage areas as a plurality of candidate samples; a clustering unit configured to cluster the plurality of candidate samples into a plurality of class clusters; a second determining unit configured to determine a plurality of center candidate sample sets respectively corresponding to class cluster centers of the plurality of class clusters; a receiving unit configured to
15 receive a manual labeling result, the manual labeling result being obtained by performing damage classification labeling on at least a portion of the center candidate sample sets from the plurality of center candidate sample sets, wherein respective center candidate samples in the same center candidate sample set are labeled with the same damage classification; and a labeling unit configured to
20 perform, according to the manual labeling result, damage classification labeling on other candidate samples in the class clusters to which the at least a portion of the center candidate sample sets belong so as to obtain a plurality of labeled samples labeled with damage classifications.

[0013] According to a fourth aspect, an apparatus for identifying a damage
25 classification is provided. The apparatus comprises: an acquisition unit configured to acquire a current loss assessment image; and a predicting unit configured to input the current loss assessment image to the damage identification model of any of the embodiments of the first aspect or the third aspect and predict a damage classification corresponding to the current loss assessment image.

30 **[0014]** According to a fifth aspect, a computer-readable storage medium

with a computer program stored thereon is provided. The computer program, when executed on a computer, causes the computer to execute the method described in the first aspect or the second aspect.

5 **[0015]** According to a sixth aspect, a computing device is provided. The computing device comprises a memory and a processor, and is characterized in that the memory stores executable code, and the processor, when executing the executable code, implements the method described in the first aspect or the second aspect is implemented.

10 **[0016]** In the method for sample labeling disclosed in the embodiments of the present specification, a plurality of candidate damage areas of massive loss assessment images are identified by saliency detection firstly, then the candidate damage areas are clustered, and according to the labeling result where the damage classifications of the areas corresponding to a portion of class cluster centers are manually labeled, other candidate areas of the class clusters where the portion of
15 class cluster centers are located are automatically labeled. As such, massive training samples can be obtained, and thus the damage identification model may be trained using these training samples.

BRIEF DESCRIPTION OF THE DRAWINGS

20 **[0017]** To describe technical schemes of the embodiments disclosed herein more clearly, the accompanying drawings for describing the embodiments are briefly introduced below. It is apparent that the accompanying drawings described below are only the embodiments disclosed herein, and those of ordinary skill in the art may still acquire other drawings from these accompanying
25 drawings without creative efforts.

[0018] FIG. 1 illustrates an image of a part of a vehicle according to an embodiment;

[0019] FIG. 2 illustrates a schematic diagram of the changes in an interface of a loss assessment client according to an embodiment;

30 **[0020]** FIG. 3 illustrates a flowchart of a method for sample labeling

according to an embodiment;

[0021] FIG. 4 illustrates a schematic workflow of the identification of a candidate damage area according to an embodiment;

5 [0022] FIG. 5 illustrates a flowchart of an optimized training method of a damage identification model according to an embodiment; and

[0023] FIG. 6 illustrates a structural diagram of an apparatus for sample labeling according to an embodiment.

DETAILED DESCRIPTION

10 [0024] The plurality of embodiments disclosed by the specification are described below with reference to the accompanying drawings.

[0025] Embodiments of the present specification discloses a method for sample labeling, and training samples obtained based on the method can be used to train a damage identification model. An application scenario for the damage
15 identification model is first introduced below.

[0026] The damage identification model can be applied to a loss assessment client provided to a user. According to one example, at the scene of the accident, a user can capture an image, such as a partial image of a vehicle as shown in FIG 1, of the accident scene through a terminal, such as a mobile phone,
20 a tablet computer, etc., and upload the captured image to the loss assessment client in the terminal; then the loss assessment client can use the damage identification model to determine vehicle damage information corresponding to the image of accident scene. For example, as shown in FIG. 2, the loss assessment client can determine that vehicle damage classification is medium
25 grade scratch; further, it can provide a repair scheme corresponding to the damage information and the associated compensation amount, for example, the repair scheme is paint repair and repair compensation amount is RMB 120.

[0027] In one embodiment, training of the damage identification model can be performed based on a large number of manual labeling samples. In such
30 an scheme, a large number of historical loss assessment images can be acquired

from the insurance company, such as images captured by loss adjusters of insurance company at the accident scene, then damage objects of the loss assessment images and corresponding damage classifications can be labeled by labeling staff, and these labeled samples are used as training samples to train the damage identification model. However, such a scheme has a very high cost of manual labeling, and therefore, it is difficult to implement.

[0028] Based on the above description, embodiments of the present specification disclose a method for sample labeling, wherein a plurality of candidate damage areas within massive loss assessment images are identified by saliency detection firstly, then the candidate damage areas are clustered, and according to the labeling result where the damage classifications of the areas corresponding to a portion of class clustering centers are manually labeled, other candidate areas of the class clusters where the portion of class cluster centers are located are automatically labeled. As such, massive training samples can be obtained, and thus the damage identification model may be trained using the training samples. The detailed implementation steps of the above process are described below.

[0029] FIG. 3 illustrates a flowchart of a method for sample labeling according to an embodiment, and the executive subject of the method may be a device with processing capabilities: a server, a system, or an apparatus. As shown in FIG. 3, the process of the method comprises the following steps: step S310, acquiring a plurality of historical loss assessment images; step S320, employing saliency detection to determine a plurality of candidate damage areas from the plurality of historical loss assessment images, and use images of the plurality of candidate damage areas as a plurality of candidate samples; step S330, clustering the plurality of candidate samples into a plurality of class clusters, and determining a plurality of center candidate sample sets respectively corresponding to class cluster centers of the plurality of class clusters; step S340, receiving a manual labeling result, the manual labeling result being obtained by performing damage classification labeling on at least a portion of center candidate sample sets

of the plurality of center candidate sample sets, wherein respective center candidate samples of the same center candidate sample set are labeled with the same damage classification; and step S350, performing, according to the manual labeling result, performing damage classification labeling on other candidate samples in the class clusters to which the at least a portion of center candidate sample sets belong to obtain a plurality of labeled samples labeled with damage classifications.

[0030] Firstly, at step S310, a plurality of historical loss assessment images are acquired.

10 [0031] In one embodiment, historical loss assessment images can be acquired from the database of an insurance company. In another embodiment, historical loss assessment images can also be acquired from forums or news sites related to traffic accidents.

[0032] In one embodiment, a plurality of historical loss assessment images may contain images reflecting different information of an accident vehicle, for example, images reflecting information such as damage, vehicle plate number, vehicle model, etc. In one example, historical loss assessment images may comprise partial or global images reflecting vehicle damage of the vehicle bodies. In another example, historical loss assessment images may comprise a captured image of a driver's license or a vehicle license. In other words, the historical loss assessment images may contain image content not directly related to damage object.

[0033] After acquiring a plurality of historical loss assessment images, then at step S320, by employing saliency detection, images of a plurality of candidate damage areas are determined from the plurality of historical loss assessment images as a plurality of candidate samples.

[0034] It should be noted that the saliency detection (or referred to as a visual attention mechanism) refers to extracting a salient object (i.e., an object of human interest) in an image by simulating a human's visual characteristics with an intelligent algorithm. In general, salient objects have a certain structure and

texture, a strong stimulation in color, and the salient objects are quite different from surrounding areas. Based on these simple saliency detection principles, saliency detection can be realized.

[0035] For the historical loss assessment images acquired at step S310, because while the loss adjuster is capturing an image of vehicle damage, his purpose is to clearly reflect the damage object, the damage object is usually located in the central part of image, and there is usually a large difference between the damage object and other parts of vehicle body in color, structure, and texture. Therefore, if a historical loss assessment image contains damage object information, the damage object in the image can be quickly located by employing saliency detection for the central region of the image.

[0036] Further, according to one specific embodiment, identifying a plurality of candidate damage areas may comprise: firstly identifying a salient object in each of the plurality of historical loss assessment images by employing saliency detection, then determining a minimum rectangular area where the salient object is located, and using the minimum rectangular area as a corresponding candidate damage area. In one example, determining a minimum rectangular area may comprise: determining a minimum rectangular area based on coordinate information of the salient object, for example, the coordinate information may comprise the coordinates of the uppermost, the lowermost, the leftmost, and the rightmost points in the salient object. According to one specific example, the saliency map shown in (b) of FIG. 4 may be obtained according to the partial image of the vehicle body shown in FIG. 4, wherein the white part is the detected salient object, and then the minimum rectangular area containing the salient object shown in (c) of FIG. 4 can be determined and used as a candidate damage area.

[0037] In one embodiment, one or a plurality of candidate damage areas may be identified from a particular historical loss assessment image.

[0038] According to above description, a plurality of candidate damage areas may be obtained, and images of the plurality of candidate damage areas may

be used as a plurality of candidate samples. It should be noted that, since other vehicle information may also be included in the historical loss assessment images, salient objects identified from the images by employing saliency detection may comprise non-damage objects, such as other unrelated components of vehicles, objects in the background, certificates of users, etc. Therefore, a portion of the plurality of candidate samples formed may comprise no damage object.

[0039] For the candidate samples obtained as above, then at step S330, the plurality of candidate samples are clustered into a plurality of class clusters, and a plurality of center candidate sample sets respectively corresponding to class cluster centers of the plurality of class clusters are determined.

[0040] In one embodiment, clustering a plurality of candidate samples into a plurality of class clusters may comprise: firstly, determining, on the basis of an image feature extraction model, a feature vector corresponding to each of the plurality of candidate samples so as to obtain a plurality of feature vectors; and then clustering the plurality of feature vectors to obtain the plurality of class clusters. Further, it should be noted that an existing image feature extraction model can be employed to determine the feature vectors, e.g., an ImageNet trained ResNet model, or a Histogram of Oriented Gradient (HOG) feature extraction model, which is not limited here.

[0041] In one embodiment, a plurality of candidate samples may be clustered into a plurality of class clusters by employing a variety of existing clustering algorithms. Among them, a variety of clustering algorithms may comprise partitioning clustering algorithms such as K-MEANS algorithm, K-MEDOIDS algorithm, CLARANS algorithm, etc.; hierarchical clustering algorithms such as BIRCH algorithm, CURE algorithm, CHAMELEON algorithm, etc.; and density-based clustering algorithms such as DBSCAN algorithm, OPTICS algorithm, and DENCLUE algorithm, etc. In one specific embodiment, when employing K-MEANS algorithm, the number of clusters K needs to be preset, and the larger the value of K, the more class clusters can be obtained by clustering, and the higher the resolution of classification will be.

Specifically, a staff member can gradually adjust K based on operating experience and the feedback result of the clustering, in order to obtain a plurality of appropriate class clusters.

5 **[0042]** According to above description, a plurality of class clusters may be obtained, and each class cluster comprises a corresponding class cluster center, and then each center candidate sample set corresponding to each class cluster center can be determined.

10 **[0043]** According to one embodiment, a predetermined number of center candidate samples in each class cluster closest to the corresponding class cluster center are determined, and the determined predetermined number of center candidate samples are used to construct a corresponding center candidate sample set. It should be noted that, in general, the predetermined number is significantly less than the total number of candidate samples in each class cluster. In one example, one class cluster comprises 10,000 candidate samples, and the
15 predetermined number may be 50 or 100, and so on.

20 **[0044]** According to one clustering algorithm employed in one embodiment, the class cluster center corresponds to a vector in mathematics, that is, a point in the vector space mapped by the feature vector of each image, referred to as a center point. In such a case, a plurality of candidate samples
20 corresponding to a predetermined number of feature vectors closest to the center point are used to construct the center candidate sample sets of the class cluster.

25 **[0045]** According to method above, a plurality of center candidate sample sets corresponding to the plurality of class cluster centers may be obtained, and then these center candidate sample sets may be provided to the labeling staff to
25 label corresponding damage classifications. Further, in one specific embodiment, historical loss assessment images corresponding to center candidate sample sets may be provided to labeling staff, and candidate damage areas corresponding to the center candidate sample sets may be identified in the historical loss assessment images. In another specific embodiment, only the candidate damage
30 areas corresponding to center candidate sample sets may be displayed to the

labeling staff, for example, a sub-image corresponding to the candidate damage area may be cut from a historical loss assessment image and provided to the labeling staff.

5 **[0046]** Then, at step S340, a manual labeling result is received, the manual labeling result being obtained by performing damage classification labeling on at least a portion of center candidate sample sets from the plurality of center candidate sample sets, wherein respective center candidate samples within the same center candidate sample set are labeled with the same damage classification.

10 **[0047]** It should be noted that, in general, instead of only selecting a single center candidate sample corresponding to the class cluster center and providing it to the operator, a plurality of center candidate samples in each class cluster are selected to construct a center candidate sample set and provided to the operator to label the damage classification corresponding to the class cluster. The reason for this is that providing a plurality of center candidate samples corresponding to each
15 class cluster can allow the operator to determine and label the damage classification of each class cluster more accurately.

20 **[0048]** As previously mentioned, a portion of the plurality of candidate samples formed at step S320 may comprise no damage object. After the clustering at step S330, samples containing similar salient objects will be clustered into the same class cluster, and the center candidate sample set corresponding to the class cluster center can reflect the most typical salient object in the class cluster. Among all the determined center candidate sample sets, there will also be some salient objects in the center candidate sample sets that are not
25 damage objects, but other objects in accident scene, such as trash cans, trees, passersby, certificates, or undamaged vehicle components, etc. Accordingly, labeling staff can filter all the center candidate sample sets firstly, and then only the center candidate sample sets comprising damage objects are labeled, that is, the damage classifications of only at least a portion of center candidate sample sets from a plurality of center candidate sample sets are labeled.

30 **[0049]** In one embodiment, in the manual labeling result, any two center

candidate sample sets of the plurality of center candidate sample sets can correspond to the same or different damage classifications.

[0050] According to one specific example, assuming that 1,000,000 candidate samples are identified from historical loss assessment images, which are clustered into 100 class clusters with an average of 10,000 samples per class cluster by clustering at step S330. Next, 100 center candidate sample sets corresponding to the 100 class clusters are provided to labeling staff for labeling, where each center candidate sample set comprises 100 center candidate samples. The labeling staff can first exclude 45 sample sets with non-damage objects from the 100 sample sets, and only label the remaining 55 center candidate sample sets with a total of 5500 center candidate samples. Since the center candidate sample set is a typical representative of candidate samples in the class cluster, excluding 45 center candidate sample sets means that all candidate samples (450,000) corresponding to the 45 class clusters are excluded and do not need to be labeled. Accordingly, large amounts of interference samples are excluded by clustering operations and simple filtering, greatly reducing labor costs. The labeling staff only needs to label a very small number of center candidate sample sets.

[0051] According to one specific example, the manual labeling result may comprise: the damage classification of center candidate sample set A is light scratching, the damage classification of center candidate sample sets B and C is medium grade deformation, and the damage classification of center candidate sample set D is severe deformation, and the like.

[0052] According to above description, a manual labeling result of manually labeling damage classifications of at least a portion of the plurality of center candidate sample sets can be obtained. Then, at step S350, according to the manual labeling result, damage classifications of other candidate samples in the class clusters to which the at least a portion of center candidate sample sets belong are labeled to obtain a plurality of labeled samples labeled with damage classifications.

[0053] In one embodiment, other candidate samples in class clusters

where center candidate sample sets labeled with damage classifications are located may be labeled with the same damage classification as the center candidate sample set. In one specific example, there are 10,000 candidate samples in a class cluster in which 100 center candidate samples in the center candidate sample set are manually labeled as severe deformation, and accordingly, the other 9900 candidate samples may be automatically labeled as severe deformation. Therefore, by labeling a small number of center candidate sample sets, other candidate samples can be automatically labeled.

[0054] According to above description, a plurality of labeled samples labeled with damage classifications are obtained based on lightweight manual labeling. Specifically, in the above example of 1,000,000 candidate samples, only 55 center candidate sample sets, a total of 5500 center candidate samples, are manually labeled after clustering the candidate samples into 100 class clusters, and then accordingly other samples in class clusters are automatically labeled, thus obtaining 550,000 labeled samples. Samples labeled this way can be referred to as weak labeled samples.

[0055] It should be noted that, after step S350, the method can include using the plurality of labeled samples obtained above as a first training sample set to train a damage identification model.

[0056] In one embodiment, the damage identification model may be trained based on the existing classification algorithms. In one example, the classification algorithms may comprise Bayesian classification algorithm, support vector machine, artificial neural network, integrated learning, and so on, which are not limited here.

[0057] By performing the method for sample labeling provided by the embodiments of this specification, a large number of labeled samples are obtained based on a small amount of manual labeling, which in turn can result in a damage identification model with certain accuracy, thus solving the problem of the damage detection model requiring a large amount of manual labeling cost.

[0058] Further, in order to improve the accuracy of the above-mentioned

damage identification model, a certain number of candidate samples can be pre-labeled with damage identification model trained by massive weak label data, then manual check and correction are performed to obtain a training set having undergone manual check and modification, and then based on this training set, the
5 damage identification model is further trained. The above process is described below with reference to FIG. 5.

[0059] FIG. 5 illustrates a flowchart of an optimized method for training a damage identification model according to an embodiment. As shown in FIG. 5, the method may comprise the following steps.

10 [0060] Firstly, at step S510, at least a portion of candidate samples are selected from a plurality of candidate samples.

[0061] In one embodiment, a predetermined number of candidate samples may be randomly selected from the plurality of candidate samples identified at aforementioned step S320, and the candidate samples determined this way may
15 comprise negative samples in which the salient object is not damage object, and the predetermined number may be flexibly selected according to actual accuracy requirements. In another embodiment, a predetermined number of candidate samples may be selected from each class cluster in accordance with a fixed proportion or randomly.

20 [0062] After selecting at least a portion of candidate samples, at step S520, a corresponding damage classification for each of the at least a portion of candidate samples is predicted based on the damage identification model.

[0063] Specifically, the selected at least a portion of the candidate samples are input to the aforementioned damage identification model resulting from
25 training to obtain a corresponding damage classification for each candidate sample. It should be noted that because the aforementioned damage identification model is obtained based on training using weak label data, and the accuracy of these weak labels does not meet a very high requirement, the accuracy of the initial damage identification model obtained by training does not reach a very
30 high level, and there may be some errors in the predicted damage classifications.

[0064] Accordingly, further, in one embodiment, the candidate samples with the damage classification predicted at step S520 are provided to labeling staff, so that the labeling staff can check and correct the predicted damage classification. At step S530, a manual correction result of manual corrections
5 applied to the damage classifications of the at least a portion of candidate samples is received. Then, at step S540, based on the manual correction result, the at least a portion of candidate samples are labeled with damage classifications.

[0065] Next, at step S550, the at least a portion of candidate samples labeled with damage classifications are used as a second training sample set to
10 perform further training on the damage identification model.

[0066] As such, a certain number of more accurate second set of training samples can be obtained by such low-cost manual labeling of check and correction, and then the damage identification model can be further trained based on the second set of training samples to improve the accuracy of the damage
15 identification model.

[0067] According to an embodiment of another aspect, provided is a method for identifying a damage classification that can be performed based on the damage identification model trained in the previous embodiments. Specifically, the method may comprise: firstly, acquiring a current loss assessment image; and
20 then, inputting the current loss assessment image into the damage identification model provided in any of the previous embodiments to predict with respect a damage classification corresponding to the current loss assessment image.

[0068] According to an embodiment of another aspect, provided is an apparatus for sample labeling. FIG. 6 illustrates a structural diagram of an
25 apparatus for sample labeling according to an embodiment. As shown in FIG. 6, apparatus 600 comprises:

[0069] an acquisition unit 610 configured to acquire a plurality of historical loss assessment images;

[0070] a first determining unit 620 configured to determine, on the basis
30 of saliency detection, a plurality of candidate damage areas from the plurality of

historical loss assessment images, and use images of the plurality of candidate damage areas as a plurality of candidate samples;

[0071] a clustering unit 630 configured to cluster the plurality of candidate samples into a plurality of class clusters;

5 [0072] a second determining unit 640 configured to determine a plurality of center candidate sample sets respectively corresponding to class cluster centers of the plurality of class clusters;

[0073] a receiving unit 650 configured to receive a manual labeling result, the manual labeling result being obtained by performing damage classification labeling on at least a portion of the plurality of center candidate sample sets, wherein respective center candidate samples in the same center candidate sample set are labeled with the same damage classification; and

[0074] a labeling unit 660 configured to perform, according to the manual labeling result, damage classification labeling on other candidate samples in the class clusters to which the at least a portion of center candidate sample sets belong to obtain a plurality of labeled samples labeled with damage classifications.

[0075] According to one embodiment, the first determining unit 620 is specifically configured to:

[0076] detect a salient object in each of the plurality of historical loss assessment images; and

[0077] determine a minimum rectangular area where the salient object is located and use the minimum rectangular area as a corresponding candidate damage area.

[0078] According to one embodiment, the clustering unit 630 is specifically configured to:

[0079] determine, on the basis of an image feature extraction model, a feature vector corresponding to each of the plurality of candidate samples so as to obtain a plurality of feature vectors; and

[0080] cluster the plurality of feature vectors to obtain the plurality of class clusters.

[0081] According to one embodiment, the second determining unit 640 is specifically configured to:

5 [0082] determine a predetermined number of center candidate samples in each class cluster closest to the corresponding class cluster center, and using the predetermined number of center candidate samples to construct a corresponding center candidate sample set.

[0083] According to one embodiment, the apparatus further comprises a first training unit 670 configured to use the plurality of labeled samples as a first training sample set to train a damage identification model, the damage
10 identification model being used to identify damage classifications of damages on vehicles from images.

[0084] According to one embodiment, the apparatus further comprises a second training unit 680 configured to:

15 [0085] select at least a portion of candidate samples from the plurality of candidate samples;

[0086] based on the damage identification model, predict a damage classification corresponding to each of the at least a portion of candidate samples;

[0087] receive a manual correction result of manual corrections applied to the damage classifications of the at least a portion of candidate samples;

20 [0088] label the at least a portion of candidate samples with damage classifications based on the manual correction result; and

[0089] use the at least a portion of candidate samples labeled with the damage classifications as a second training sample set to perform further training on the damage identification model.

25 [0090] In summary, by employing the apparatus for sample labeling provided by the embodiments of this specification, a large number of labeled samples can be obtained by manually labeling a small number of samples, such that a damage identification model with certain accuracy can be obtained on the basis of these labeled samples, thereby solving the problem of a damage detection
30 model requiring large amounts of manual labeling costs.

[0091] According to an embodiment of yet another aspect, provided is an apparatus for identifying a damage classification that can be operated based on the damage identification model trained in the previous embodiments. Specifically, the apparatus may comprise: an acquisition unit configured to acquire a current
5 loss assessment image; and a predicting unit configured to input the current loss assessment image into the damage identification model provided in any of the previous embodiments to predict a damage classification corresponding to the current loss assessment image.

[0092] As mentioned above, according to an embodiment of yet another
10 aspect, a computer-readable storage medium with a computer program stored thereon is further provided. The computer program, when executed on a computer, causes the computer to execute the method described with reference to FIG. 3 or FIG. 5.

[0093] According to an embodiment of yet another aspect, a computing
15 device comprising a memory and a processor is further provided. The memory stores executable code, and the processor, when executing the executable code, implements the method described with reference to FIG. 3 or FIG. 5.

[0094] Those skilled in the art may be aware that in the aforementioned
one or a plurality of examples, the functions described in a plurality of
20 embodiments disclosed by the present specification can be implemented by hardware, software, firmware, or any combination thereof. When implemented by software, these functions may be stored in a computer-readable medium, or transmitted as one or a plurality of instructions or as one or a plurality of pieces of code in the computer-readable medium.

[0095] The objectives, the technical schemes, and the beneficial effects of
25 the embodiments disclosed by the present specification are further described in detail in the foregoing specific implementation. It should be understood that the foregoing descriptions are merely specific implementation of embodiments disclosed by the present specification, and are not intended to limit the protection
30 scope of the embodiments disclosed by the present specification. Any

modification, equivalent replacement, and improvement made on the basis of the technical schemes of the embodiments of the present specification shall fall within the protection scope of the embodiments of the present specification.

What Is Claimed Is:

1 1. A method for sample labeling, comprising:
2 acquiring a plurality of historical loss assessment images;
3 determining, on the basis of saliency detection, a plurality of candidate
4 damage areas from the plurality of historical loss assessment images, and using
5 images of the plurality of candidate damage areas as a plurality of candidate
6 samples;
7 clustering the plurality of candidate samples into a plurality of class
8 clusters, and determining a plurality of center candidate sample sets respectively
9 corresponding to class cluster centers of the plurality of class clusters;
10 receiving a manual labeling result, the manual labeling result being
11 obtained by performing damage classification labeling on at least a portion of the
12 plurality of center candidate sample sets, wherein respective center candidate
13 samples in the same center candidate sample set are labeled with the same
14 damage classification; and
15 performing, according to the manual labeling result, damage classification
16 labeling on other candidate samples in the class clusters to which the at least a
17 portion of center candidate sample sets belong to obtain a plurality of labeled
18 samples labeled with damage classifications.

1 2. The method according to claim 1, wherein determining a plurality
2 of candidate damage areas from the plurality of historical loss assessment images
3 comprises:
4 detecting a salient object in each of the plurality of historical loss
5 assessment images; and
6 determining a minimum rectangular area where the salient object is
7 located, and using the minimum rectangular area as a corresponding candidate
8 damage area.

1 3. The method according to claim 1, wherein clustering the plurality
2 of candidate samples into a plurality of class clusters comprises:
3 determining, on the basis of an image feature extraction model, a feature
4 vector corresponding to each of the plurality of candidate samples to obtain a
5 plurality of feature vectors; and
6 clustering the plurality of feature vectors to obtain the plurality of class
7 clusters.

1 4. The method according to claim 1, wherein determining a plurality
2 of center candidate sample sets respectively corresponding to class cluster centers
3 of the plurality of class clusters comprises:
4 determining a predetermined number of center candidate samples in each
5 class cluster closest to the corresponding class cluster center, and using the
6 predetermined number of center candidate samples to construct a corresponding
7 center candidate sample set.

1 5. The method according to any one of claims 1-4, further
2 comprising:
3 using the plurality of labeled samples as a first training sample set to train
4 a damage identification model, the damage identification model being used to
5 identify damage classifications of damages on vehicles from images.

1 6. The method according to claim 5, further comprising:
2 selecting at least a portion of candidate samples from the plurality of
3 candidate samples;
4 predicting a damage classification corresponding to each of the at least a
5 portion of candidate samples on the basis of the damage identification model;
6 receiving a manual correction result of manual corrections applied to the
7 damage classifications of the at least a portion of candidate samples;

8 labeling the at least a portion of candidate samples with damage
9 classifications on the basis of the manual correction result; and
10 using the at least a portion of candidate samples labeled with the damage
11 classifications as a second training sample set to perform further training on the
12 damage identification model.

1 7. A method for identifying a damage classification, comprising:
2 acquiring a current loss assessment image; and
3 inputting the current loss assessment image into the damage identification
4 model of claim 5 or 6 to predict a damage classification corresponding to the
5 current loss assessment image.

1 8. An apparatus for sample labeling, comprising:
2 an acquisition unit configured to acquire a plurality of historical loss
3 assessment images;
4 a first determining unit configured to determine, on the basis of saliency
5 detection, a plurality of candidate damage areas from the plurality of historical
6 loss assessment images, and use images of the plurality of candidate damage areas
7 as a plurality of candidate samples;
8 a clustering unit configured to cluster the plurality of candidate samples
9 into a plurality of class clusters;
10 a second determining unit configured to determine a plurality of center
11 candidate sample sets respectively corresponding to class cluster centers of the
12 plurality of class clusters;
13 a receiving unit configured to receive a manual labeling result, the manual
14 labeling result being obtained by performing damage classification labeling on at
15 least a portion of the plurality of center candidate sample sets, wherein respective
16 center candidate samples in the same center candidate sample set are labeled with
17 the same damage classification; and

18 a labeling unit configured to perform, according to the manual labeling
19 result, damage classification labeling on other candidate samples in the class
20 clusters to which the at least a portion of center candidate sample sets belong so
21 as to obtain a plurality of labeled samples labeled with damage classifications.

1 9. The apparatus according to claim 8, wherein the first determining
2 unit is specifically configured to:

3 detect a salient object in each of the plurality of historical loss assessment
4 images; and

5 determine a minimum rectangular area where the salient object is located,
6 and use the minimum rectangular area as a corresponding candidate damage area.

1 10. The apparatus according to claim 8, wherein the clustering unit is
2 specifically configured to:

3 determine, on the basis of an image feature extraction model, a feature
4 vector corresponding to each of the plurality of candidate samples so as to obtain
5 a plurality of feature vectors; and

6 cluster the plurality of feature vectors to obtain the plurality of class
7 clusters.

1 11. The apparatus according to claim 8, wherein the second
2 determining unit is specifically configured to:

3 determine a predetermined number of center candidate samples in each
4 class cluster closest to the corresponding class cluster center, and use the
5 predetermined number of center candidate samples to construct a corresponding
6 center candidate sample set.

1 12. The apparatus according to any one of claims 8-11, further
2 comprising a first training unit configured to use the plurality of labeled samples
3 as a first training sample set to train a damage identification model, the damage

4 identification model being used to identify damage classifications of damages on
5 vehicles from images.

1 13. The apparatus according to claim 12, further comprising a second
2 training unit configured to:
3 select at least a portion of candidate samples from the plurality of
4 candidate samples;
5 predict a damage classification corresponding to each of the at least a
6 portion of candidate samples on the basis of the damage identification model;
7 receive a manual correction result of manual corrections applied to the
8 damage classifications of the at least a portion of candidate samples;
9 label the at least a portion of candidate samples with damage
10 classifications on the basis of the manual correction result; and
11 use the at least a portion of candidate samples labeled with the damage
12 classifications as a second training sample set to perform further training on the
13 damage identification model.

1 14. An apparatus for identifying a damage classification, comprising:
2 an acquisition unit configured to acquire a current loss assessment image;
3 and
4 a predicting unit configured to input the current loss assessment image
5 into the damage identification model of any one of claims 5, 6, 12 and 13 to
6 predict a damage classification corresponding to the current loss assessment
7 image.

1 15. A computer-readable storage medium, on which a computer
2 program is stored, wherein the computer program, when executed on a computer,
3 causes the computer to execute the method according to any one of claims 1-7.

1 16. A computing device, comprising a memory and a processor,
2 characterized in that the memory stores executable code, and the processor, when
3 executing the executable code, implements the method according to any one of
4 claims 1-7.