

样本标注方法及装置、损伤类别的识别方法及装置

技术领域

[01]本说明书实施例涉及模型训练领域，具体地，涉及一种样本标注方法及装置、损伤类别的识别方法及装置。

5 背景技术

[02]在车险理赔场景，目前业界推出了用于理赔的用户客户端，根据用户在客户端中上传的包括事故车辆的图片，可以自动识别出事故车辆的损伤信息，如损伤部件、损伤类别，并给出相关的理赔信息，如维修方案、赔偿金额等。这种方案因无需人工查勘定损核损，减少了保险公司的成本，提升了普通用户的车险理赔体验。

- 10 [03]目前业界在识别图片中事故车辆的损伤信息时，通常是通过直接将该图片与历史数据库中的海量图片进行对比，找出相似图片，以根据相似图片对应的损伤信息确定出事故车辆的损伤信息。这种方法的计算量大，且由此确定出的损伤信息的准确度不高。因此，需要一种合理的方法，可以更加快速、准确地确定出图片中车辆的损伤信息。

发明内容

- 15 [04]在本说明书描述的一种样本标注方法中，先对海量定损照片进行聚类，再根据接收的针对部分聚类中心的损伤类型进行人工标注的标注结果，对部分聚类中心所在类簇的其他图片进行自动打标，以快速得到海量的弱标签训练样本，进而可以基于这些弱标签训练样本训练损伤识别模型。

- [05]根据第一方面，提供一种样本标注方法，该方法包括：获取多张历史定损图片；基
20 于显著性检测，从所述多张历史定损图片中确定出多个候选损伤区域，并将所述多个候选损伤区域的图像作为多个候选样本；将所述多个候选样本聚类成多个类簇，并确定所述多个类簇的类簇中心分别对应的多个中心候选样本集；接收人工标注结果，所述人工标注结果是针对所述多个中心候选样本集中的至少一部分中心候选样本集进行的损伤类别标注，其中同一中心候选样本集中的各个中心候选样本标注有相同的损伤类别；根
25 据所述人工标注结果，对所述至少一部分中心候选样本集所在类簇中其他候选样本的损伤类别进行标注，以得到标注有损伤类别的多个标注样本。

- [06]根据一个实施例,所述从所述多张历史定损图片中确定出多个候选损伤区域,包括:检测所述多张历史定损图片中各张历史定损图片中的显著性对象;确定所述显著性对象所在的最小矩形区域,并将所述最小矩形区域作为对应的候选损伤区域。
- [07]根据一个实施例,所述将所述多个候选样本聚类成多个类簇,包括:基于图像特征提取模型,确定与所述多个候选样本中各个候选样本对应的特征向量,以得到多个特征向量;对所述多个特征向量进行聚类,得到所述多个类簇。
- 5 [08]根据一个实施例,所述确定所述多个类簇的类簇中心分别对应的多个中心候选样本集,包括:确定各个类簇中与对应的类簇中心距离最近的预定数量的中心候选样本,并将所述预定数量的中心候选样本构成对应的中心候选样本集。
- 10 [09]根据一个实施例,还包括:将所述多个标注样本作为第一训练样本集,训练损伤识别模型,所述损伤识别模型用于从图片中识别出车辆损伤的损伤类别。
- [10]根据一个实施例,还包括:从所述多个候选样本中选择至少一部分候选样本;基于所述损伤识别模型,预测所述至少一部分候选样本中各个候选样本对应的损伤类别;接收针对所述至少一部分候选样本的损伤类别进行人工修正的人工修正结果;基于所述人工修正结果,标注出所述至少一部分候选样本的损伤类别;将标注有损伤类别的至少一部分候选样本作为第二训练样本集,进一步训练所述损伤识别模型。
- 15 [11]根据第二方面,提供一种损伤类别的识别方法,该方法包括:获取当前定损图片;将所述当前定损图片输入上述第一方面的任一实施例的损伤识别模型中,预测所述当前定损图片对应的损伤类别。
- 20 [12]根据第三方面,提供一种样本标注装置,该装置包括:获取单元,用于获取多张历史定损图片;第一确定单元,用于基于显著性检测,从所述多张历史定损图片中确定出多个候选损伤区域,并将所述多个候选损伤区域的图像作为多个候选样本;聚类单元,用于将所述多个候选样本聚类成多个类簇;第二确定单元,用于确定所述多个类簇的类簇中心分别对应的多个中心候选样本集;接收单元,用于接收人工标注结果,所述人工标注结果是针对所述多个中心候选样本集中的至少一部分中心候选样本集进行的损伤类别标注,其中同一中心候选样本集中的各个中心候选样本标注有相同的损伤类别;标注单元,用于根据所述人工标注结果,对所述至少一部分中心候选样本集所在类簇中其他候选样本的损伤类别进行标注,以得到标注有损伤类别的多个标注样本。
- 25 [13]根据第四方面,提供一种损伤类别的识别装置,该装置包括:获取单元,用于获取

当前定损图片；预测单元，用于将所述当前定损图片输入第一方面或第三方面的任一实施例中所述的损伤识别模型中，预测所述当前定损图片对应的损伤类别。

[14]根据第五方面，提供了一种计算机可读存储介质，其上存储有计算机程序，当所述计算机程序在计算机中执行时，令计算机执行第一方面或第二方面所描述的方法。

- 5 [15]根据第六方面，提供了一种计算设备，包括存储器和处理器，其特征在于，所述存储器中存储有可执行代码，所述处理器执行所述可执行代码时，实现第一方面或第二方面所描述的方法。

[16]在本说明书实施例披露的样本标注方法中，先通过显著性检测识别出海量定损照片中的多个候选损伤区域，然后对这些候选损伤区域进行聚类，根据针对部分类簇中心所

- 10 对应区域的损伤类型进行人工标注的标注结果，对部分类簇中心所在类簇的其他候选区域进行自动打标，如此可以快速得到大量的训练样本，进而可以根据这些训练样本训练损伤识别模型。

附图说明

[17]为了更清楚地说明本说明书披露的多个实施例的技术方案，下面将对实施例描述中
15 所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本说明书披露的多个实施例，对于本领域普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据这些附图获得其它的附图。

[18]图 1 示出根据一个例子的车辆局部照片；

[19]图 2 示出根据一个例子的定损客户端的界面变化示意图；

- 20 [20]图 3 示出根据一个实施例的样本标注方法流程图；

[21]图 4 示出根据一个实施例的候选损伤区域的识别过程示意图；

[22]图 5 示出根据一个实施例的损伤识别模型的优化训练方法流程图；

[23]图 6 示出根据一个实施例的样本标注装置结构图。

具体实施方式

- 25 [24]下面结合附图，对本说明书披露的多个实施例进行描述。

[25]本说明书实施例披露一种样本标注方法，基于此方法得到的训练样本可以用于训练

损伤识别模型。下面，首先对损伤识别模型的应用场景进行介绍。

[26]损伤识别模型可以应用于提供给用户使用的定损客户端中。根据一个例子，在事故现场，用户可以通过终端，如，手机、平板电脑等，拍摄现场照片，例如图 1 中示出的车辆局部图片，并将拍摄的照片上传到终端中的定损客户端，然后定损客户端可以利用

5 损伤识别模型，确定现场照片所对应的车辆损伤信息，例如，如图 2 所示，可以确定出车辆损伤类别为中度刮擦，进一步地，还可以给出与损伤信息对应的维修方案和相关赔偿金额，例如，维修方案为：补漆，保修赔付金额：120 元。

[27]对于损伤识别模型的训练，在一种实施方案中，可以基于大量人工标注的样本进行训练。在这种方案中，可以先从保险公司获取大量的历史定损照片，如可以包括保险公司的定损人员在事故现场拍摄的照片，然后由打标人员标注出定损照片中的损伤对象，

10 以及对应的损伤类别，再将这些标注样本作为训练样本，训练损伤识别模型。然而，这种方案中人工标注成本过高，因此实施起来比较困难。

[28]基于此，本说明书实施例披露一种样本标注方法，先通过显著性检测识别出海量定损照片中的多个候选损伤区域，然后对这些候选损伤区域进行聚类，根据针对部分类簇中心所对应区域的损伤类型进行人工标注的标注结果，对部分类簇中心所在类簇的其他

15 候选区域进行自动打标，如此可以快速得到大量的训练样本，进而可以根据这些训练样本训练损伤识别模型。下面，描述以上过程的具体实施步骤。

[29]图 3 示出根据一个实施例的样本标注方法流程图，所述方法的执行主体可以为具有处理能力的设备：服务器或者系统或者装置。如图 3 所示，该方法流程包括以下步骤：

20 步骤 S310，获取多张历史定损图片；步骤 S320，采用显著性检测，从多张历史定损图片中确定出多个候选损伤区域，并将多个候选损伤区域的图像作为多个候选样本；步骤 S330，将多个候选样本聚类成多个类簇，并确定多个类簇的类簇中心分别对应的多个中心候选样本集；步骤 S340，接收人工标注结果，所述人工标注结果是针对多个中心候选样本集中的至少一部分中心候选样本集进行的损伤类别标注，其中同一中心候选样本集中的各个中心候选样本标注有相同的损伤类别；步骤 S350，根据人工标注结果，对至少一部分中心候选样本集所在类簇中其他候选样本的损伤类别进行标注，以得到标注有损伤类别的多个标注样本。

25

[30]首先，在步骤 S310，获取多张历史定损图片。

[31]在一个实施例中，可以从保险公司的数据库中获取历史定损图片。在另一个实施例

中，还可以在与交通事故相关的论坛或新闻网站中获取历史定损图片。

[32]在一个实施例中，多张历史定损图片可能涵盖反映事故车辆不同信息的图片，例如，反映损伤情况、车牌号、车辆型号等信息的图片。在一个例子中，历史定损图片可以包括反映车辆损伤情况的车身的局部或整体照片。在另一个例子中，历史定损图片可以包
5 括拍摄的驾驶证图片或行驶证图片。换言之，历史定损图片中有可能包含与损伤对象不直接相关的图片内容。

[33]在获取多张历史定损照片以后，接着，在步骤 S320，采用显著性检测，从多张历史定损图片中确定出多个候选损伤区域的图像作为多个候选样本。

[34]需要说明的是，显著性检测（或称为视觉注意机制）是指通过智能算法模拟人的视
10 觉特点，提取图像中的显著性对象（即人类感兴趣的对象）。通常，显著性对象具有一定的结构和纹理，色彩上具有较强的刺激，且显著性对象和周围区域具有较大的差异，基于这些简单的显著性检测原则，可以实现显著性检测。

[35]对于步骤 S310 中获取的历史定损照片来说，因定损员在拍摄车辆损伤的过程中，其拍摄目的主要在于清晰地反映损伤对象，因此损伤对象通常位于图片的中心部位，且
15 损伤对象与车身其他部位之间，在颜色、结构和纹理上通常具有较大的反差。因此，如果历史定损图片中包含损伤对象信息，那么可以通过对该图片的中心区域采用显著性检测，快速定位出图片中的损伤对象。

[36]进一步地，根据一个具体的实施例，识别出多个候选损伤区域可以包括：首先采用显著性检测，识别多张历史定损图片中各张历史定损图片中的显著性对象，然后确定出
20 显著性对象所在的最小矩形区域，并将此最小矩形区域作为对应的候选损伤区域。在一个例子中，确定出最小矩形区域可以包括：根据显著性对象的坐标信息确定出最小矩形区域，例如，其中坐标信息可以包括显著性对象中最上、最下、最左和最右四个点的坐标信息。根据一个具体的例子，根据图 4 中（a）示出的车身局部图像，可以得到图 4 中（b）示出的显著性度量图，其中白色部分为检测到的显著性对象，然后可以确定出
25 图 4 中（c）示出的包含显著性对象的最小矩形区域，并将此作为候选损伤区域。

[37]在一个实施例中，可以从某张历史定损图片中识别出一个或多个候选损伤区域。

[38]以上，可以得到多个候选损伤区域，并可以将多个候选损伤区域的图像作为多个候选样本。需要说明的是，由于历史定损图片中还可能包含其他车辆信息，因此，采用显著性检测从图片中识别的显著性对象有可能包含非损伤对象，例如汽车其他无关部件，

背景中的物体，用户的证件等等。由此形成的多个候选样本会有一部分样本并不包含损伤对象。

[39]对于如上获得的候选样本，接着，在步骤 S330，将多个候选样本聚类成多个类簇，并确定多个类簇的类簇中心分别对应的多个中心候选样本集。

- 5 [40]在一个实施例中，将多个候选样本聚类成多个类簇可以包括：先基于图像特征提取模型，确定与所述多个候选样本中各个候选样本对应的特征向量，以得到多个特征向量，再对多个特征向量进行聚类，得到多个类簇。进一步地，需要说明的是，可以采用现有的图像特征提取模型确定特征向量，如，ImageNet 训练的 ResNet 模型，或方向梯度直方图（Histogram of Oriented Gradient, HOG）特征提取模型等，在此不作限定。
- 10 [41]在一个实施例中，可以采用现有的多种聚类算法，将多个候选样本聚类成多个类簇。其中，多种聚类算法可以包括 K-MEANS 算法、K-MEDOIDS 算法、CLARANS 算法等划分聚类算法；BIRCH 算法、CURE 算法、CHAMELEON 算法等层次聚类算法；以及 DBSCAN 算法、OPTICS 算法、DENCLUE 算法等密度聚类算法。在一个具体的实施例中，在采用 K-MEANS 算法时需要预先设定类簇的数量 K，且 K 的值越大，聚类得到的
- 15 类簇数量越多，相应地分类精度越高，具体可以由工作人员结合业务经验，以及聚类的反馈结果，逐次对 K 进行调整，以得到合适的多个类簇。

[42]以上可以得到多个类簇，且各个类簇包括对应的各个类簇中心，接着，可以确定各个类簇中心对应的各个中心候选样本集。

- [43]根据一个实施例，可以确定各个类簇中与对应的类簇中心距离最近的预定数量的中心候选样本，并将确定出的预定数量的中心候选样本构成与该类簇对应的中心候选样本集。需要说明的是，通常情况下，预定数量远远小于各个类簇中候选样本的总数量。在一个例子中，一个类簇中包括 1 万个候选样本，而预定数量可以为 50 个或 100 个，等等。
- 20

- [44]根据一个实施例中采用的聚类算法，类簇中心对应于数学上的一个向量，也就是各个图片的特征向量所映射到的向量空间中的一个点，称为中心点。在这样的情况下，可以将距离该中心点最近的预定数量的特征向量所对应的多个候选样本构成该类簇的中心候选样本集。
- 25

[45]由上可以得到与多个类簇中心分别对应的多个中心候选样本集，然后，可以将这些中心候选样本集提供给打标人员，以标注出对应的损伤类别。进一步地，在一个具体的

实施例中，可以将与中心候选样本集对应的历史定损照片提供给打标人员，且历史定损照片中标识出了与中心候选样本集对应的候选损伤区域。在另一个具体的实施例中，可以只向打标人员展示与中心候选样本集对应的候选损伤区域，例如，可以从历史定损图片中切割出与此候选损伤区域对应的子图片，并将子图片提供给打标人员。

- 5 [46]接着，在步骤 S340，接收人工标注结果，所述人工标注结果是针对多个中心候选样本集中的至少一部分中心候选样本集进行的损伤类别标注，其中同一中心候选样本集中的各个中心候选样本标注有相同的损伤类别。

- [47]需要说明的是，通常情况下，之所以在各个类簇中选取多个中心候选样本构成中心候选样本集，并将其提供给工作人员以标注与该类簇对应的损伤类别，而不是只选择类簇中心所对应的单个中心候选样本提供给工作人员，是因为提供与各个类簇对应的多个中心候选样本，可以使工作人员更加准确地对各个类簇的损伤类别进行判断和标注。

- [48]如前所述，步骤 S320 中形成的多个候选样本中会有一部分样本并不包含损伤对象。经过步骤 S330 的聚类，包含相似显著性对象的样本会被聚类到同一类簇，类簇中心对应的中心候选样本集可以反映该类簇中最为典型的显著性对象。那么相应地，在确定出的所有中心候选样本集中，也会有部分中心候选样本集中的显著性对象不是损伤对象，而是其他对象，如车祸现场的垃圾桶、树木、路人、证件，或未被损坏的车辆部件等。因此，打标人员可以先对全部的中心候选样本集进行筛选，然后只对包括损伤对象
- 15 的中心候选样本集进行打标，也就是只针对多个中心候选样本集中的至少一部分中心候选样本集，进行损伤类别的标注。

- 20 [49]在一个实施例中，对于多个中心候选样本集中的任意两个中心候选样本集，在人工标注结果中，它们所对应的损伤类别可以相同也可以不同。

- [50]根据一个具体的例子，假定从历史定损图片中识别出 100 万个候选样本，通过步骤 S330 的聚类，这些候选样本被聚类为 100 个类簇，平均每个类簇 1 万个样本。接着，将这 100 个类簇对应的 100 个中心候选样本集发送给打标人员进行打标，其中各个中心候选样本集中包括 100 个中心候选样本。打标人员可以从这 100 个样本集中首先排除掉例如 45 个非损伤对象的样本集，只对剩余的 55 个中心候选样本集，共计 5500 个中心候选样本，进行打标。由于中心候选样本集是该类簇中候选样本的典型代表，因此在排除掉 45 个中心候选样本集的同时，就意味着，这 45 个类簇所对应的所有候选样本（45 万个）得到排除，不必作为标注对象。如此，通过聚类操作和简单的筛选，排除掉大量
- 25 30 的干扰样本，极大减小人工成本。打标人员只需要对极少量的中心候选样本集进行标注

即可。

[51]根据一个具体的例子，人工标注结果可以包括：中心候选样本集 A 的损伤类别为轻度刮擦，中心候选集 B 和 C 的损伤类别为中度变形，中心候选样本集 D 的损伤类别为重度变形等。

- 5 [52]以上，可以得到多个中心候选样本集中的至少一部分的损伤类别进行人工标注的人工标注结果。接着，然后，在步骤 S350，根据人工标注结果，对至少一部分中心候选样本集所在类簇中其他候选样本的损伤类别进行标注，以得到标注有损伤类别的多个标注样本。

- [53]在一个实施例中，对于标注有损伤类别的各个中心候选样本集，可以将其所在类簇
10 中其他候选样本标注上与该中心候选样本集相同的损伤类别。在一个具体的例子中，某个类簇中有 1 万个候选样本，其中心候选样本集中的 100 个中心候选样本被人工标注为重度变形，据此，其他的 9900 个候选样本可以被自动标注为重度变形。如此，通过对少量中心候选样本集进行标注，可以实现对其他候选样本进行自动打标。

- [54]以上，可以基于轻量级的人工标注，得到标注有损伤类别的多个标注样本。具体地，
15 在以上 100 万个候选样本的例子中，在将候选样本聚类为 100 个类簇之后，只需要对 55 个中心候选样本集，共计 5500 个中心候选样本，进行人工标注，然后对应地对类簇的其他样本进行自动打标，就可以得到 55 万个标注样本。这样标注的样本可以称为弱标签样本。

- [55]需要说明的是，在步骤 S350 之后还可以包括：将以上得到的多个标注样本作为第
20 一训练样本集，训练损伤识别模型。

[56]在一个实施例中，可以基于现有的分类算法，训练损伤识别模型。在一个例子中，分类算法可以包括贝叶斯分类算法、支持向量机、人工神经网络和集成学习等，在此不作限定。

- [57]通过执行本说明书实施例提供的样本标注方法，实现了基于少量的人工标注，得到
25 大量的标注样本，进而基于这些标注样本可以得到具有一定准确度的损伤识别模型，解决了损伤检测模型需要耗费大量人工标注成本的问题。

[58]进一步地，为了提升上述损伤识别模型的准确度，还可以先使用这个基于海量弱标签数据训练的损伤识别模型，对一定数量的候选样本进行算法预标注，再由人工进行检查更正，得到人工检查修改后的训练集，再基于此训练集，进一步训练损伤识别模型。

下面，结合图 5 对以上过程进行描述。

[59] 图 5 示出根据一个实施例的损伤识别模型的优化训练方法流程图。如图 5 所示，该方法流程包括以下步骤：

[60] 首先，在步骤 S510，从多个候选样本中选择至少一部分候选样本。

- 5 [61] 在一个实施例中，可以从前述步骤 S320 中确定出的多个候选样本中，随机选择出预定数量的候选样本，如此确定出的候选样本中可能会包括显著性对象不是损伤对象的负样本，其中预定数量可以根据实际的精度需要灵活选择。在另一个实施例中，可以从各个类簇中，按照固定比例，或者随机地，选择出预定数量的候选样本。

[62] 以上，在选择出至少一部分候选样本后，接着，在步骤 S520，基于损伤识别模型，

- 10 预测至少一部分候选样本中各个候选样本对应的损伤类别。

[63] 具体地，将选择出的至少一部分候选样本输入前述训练得到的损伤识别模型中，以得到其中各个候选样本对应的损伤类别。需要说明的是，因前述损伤识别模型是基于弱标签数据训练得到，而这些弱标签本身的准确度并不能达到很高的要求，因此训练得到的初步的损伤识别模型的准确率也不会达到非常高的水平，预测出的损伤类别中会存在

- 15 一些错误。

[64] 于是，进一步地，在一个实施例中，将步骤 S520 预测出损伤类别的候选样本提供给打标人员，以使打标人员对预测出的损伤类别进行检查修正。在步骤 S530，接收针对至少一部分候选样本的损伤类别进行人工修正的人工修正结果。然后，在步骤 S540，基于人工修正结果，标注出至少一部分候选样本的损伤类别。

- 20 [65] 接着，在步骤 S550，将标注有损伤类别的至少一部分候选样本作为第二训练样本集，进一步训练所述损伤识别模型。

[66] 如此，经过检查修正这种低成本地人工标注，可以得到一定数量的更加精准的第二训练样本集，进而可以根据第二训练样本集，进一步训练损伤识别模型，以提高损伤识别模型的准确率。

- 25 [67] 根据另一方面的实施例，还提供一种损伤类别的识别方法，所述方法可以基于前述实施例中训练得到的损伤识别模型进行。具体地，所述方法可以包括：首先，获取当前定损图片；然后，将当前定损图片输入前述任一实施例中提供的损伤识别模型中，预测当前定损图片对应的损伤类别。

[68]据再一方面的实施例，还提供一种样本标注装置。图 6 示出根据一个实施例的样本标注装置结构图。如图 6 所示，该装置 600 包括：

[69]获取单元 610，用于获取多张历史定损图片；

[70]第一确定单元 620，用于基于显著性检测，从所述多张历史定损图片中确定出多个
5 候选损伤区域的图像作为多个候选样本；

[71]聚类单元 630，用于将所述多个候选样本聚类成多个类簇；

[72]第二确定单元 640，用于确定所述多个类簇的类簇中心分别对应的多个中心候选样本集；

[73]接收单元 650，用于接收人工标注结果，所述人工标注结果是针对所述多个中心候
10 选样本集中的至少一部分中心候选样本集进行的损伤类别标注，其中同一中心候选样本集中的各个中心候选样本标注有相同的损伤类别；

[74]标注单元 660，用于根据所述人工标注结果，对所述至少一部分中心候选样本集所在类簇中其他候选样本的损伤类别进行标注，以得到标注有损伤类别的多个标注样本。

[75]根据一个实施例，所述第一确定单元 620 具体用于：

15 [76]检测所述多张历史定损图片中各张历史定损图片中的显著性对象；

[77]确定所述显著性对象所在的最小矩形区域，并将所述最小矩形区域作为对应的候选损伤区域。

[78]根据一个实施例，所述聚类单元 630 具体用于：

[79]基于图像特征提取模型，确定与所述多个候选样本中各个候选样本对应的特征向量，
20 以得到多个特征向量；

[80]对所述多个特征向量进行聚类，得到所述多个类簇。

[81]根据一个实施例，所述第二确定单元 640 具体用于：

[82]确定各个类簇中与对应的类簇中心距离最近的预定数量的中心候选样本，并将所述预定数量的中心候选样本构成对应的中心候选样本集。

25 [83]根据一个实施例，还包括第一训练单元 670，用于将所述多个标注样本作为第一训练样本集，训练损伤识别模型，所述损伤识别模型用于从图片中识别出车辆损伤的损伤类别。

- [84]根据一个实施例，还包括第二训练单元 680，用于：
- [85]从所述多个候选样本中选择至少一部分候选样本；
- [86]基于所述损伤识别模型，预测所述至少一部分候选样本中各个候选样本对应的损伤类别；
- 5 [87]接收针对所述至少一部分候选样本的损伤类别进行人工修正的人工修正结果；
- [88]基于所述人工修正结果，标注出所述至少一部分候选样本的损伤类别；
- [89]将标注有损伤类别的至少一部分候选样本作为第二训练样本集，进一步训练所述损伤识别模型。
- [90]总之，采用本说明书实施例提供的样本标注装置，实现了基于少量的人工标注，得
- 10 到大量的标注样本，进而基于这些标注样本可以得到具有一定准确度的损伤识别模型，解决了损伤检测模型需要耗费大量人工标注成本的问题。
- [91]根据再一方面的实施例，还提供一种损伤类别的识别装置，所述装置可以基于前述实施例中训练得到的损伤识别模型进行。具体地，所述装置可以包括：获取单元，用于获取当前定损图片；预测单元，用于将当前定损图片输入前述任一实施例中提供的损伤
- 15 识别模型中，预测当前定损图片对应的损伤类别。
- [92]如上，根据又一方面的实施例，还提供一种计算机可读存储介质，其上存储有计算机程序，当所述计算机程序在计算机中执行时，令计算机执行结合图 3 或图 5 所描述的方法。
- [93]根据又一方面的实施例，还提供一种计算设备，包括存储器和处理器，所述存储器
- 20 中存储有可执行代码，所述处理器执行所述可执行代码时，实现结合图 3 或图 5 所描述的方法。
- [94]本领域技术人员应该可以意识到，在上述一个或多个示例中，本说明书披露的多个实施例所描述的功能可以用硬件、软件、固件或它们的任意组合来实现。当使用软件实现时，可以将这些功能存储在计算机可读介质中或者作为计算机可读介质上的一个或多个
- 25 指令或代码进行传输。
- [95]以上所述的具体实施方式，对本说明书披露的多个实施例的目的、技术方案和有益效果进行了进一步详细说明，所应理解的是，以上所述仅为本说明书披露的多个实施例的具体实施方式而已，并不用于限定本说明书披露的多个实施例的保护范围，凡在本说

说明书披露的多个实施例的技术方案的基础之上，所做的任何修改、等同替换、改进等，均应包括在本说明书披露的多个实施例的保护范围之内。

权利要求书

- 1、一种样本标注方法，包括：
获取多张历史定损图片；
基于显著性检测，从所述多张历史定损图片中确定出多个候选损伤区域，并将所述
- 5 多个候选损伤区域的图像作为多个候选样本；
将所述多个候选样本聚类成多个类簇，并确定所述多个类簇的类簇中心分别对应的多个中心候选样本集；
接收人工标注结果，所述人工标注结果是针对所述多个中心候选样本集中的至少一部分中心候选样本集进行的损伤类别标注，其中同一中心候选样本集中的各个中心候选
- 10 样本标注有相同的损伤类别；
根据所述人工标注结果，对所述至少一部分中心候选样本集所在类簇中其他候选样本的损伤类别进行标注，以得到标注有损伤类别的多个标注样本。
- 2、根据权利要求 1 所述的方法，其中，所述从所述多张历史定损图片中确定出多个候选损伤区域，包括：
15 检测所述多张历史定损图片中各张历史定损图片中的显著性对象；
确定所述显著性对象所在的最小矩形区域，并将所述最小矩形区域作为对应的候选损伤区域。
- 3、根据权利要求 1 所述的方法，其中，所述将所述多个候选样本聚类成多个类簇，包括：
20 基于图像特征提取模型，确定与所述多个候选样本中各个候选样本对应的特征向量，以得到多个特征向量；
对所述多个特征向量进行聚类，得到所述多个类簇。
- 4、根据权利要求 1 所述的方法，其中，所述确定所述多个类簇的类簇中心分别对应的多个中心候选样本集，包括：
25 确定各个类簇中与对应的类簇中心距离最近的预定数量的中心候选样本，并将所述预定数量的中心候选样本构成对应的中心候选样本集。
- 5、根据权利要求 1-4 任一项所述的方法，还包括：
将所述多个标注样本作为第一训练样本集，训练损伤识别模型，所述损伤识别模型用于从图片中识别出车辆损伤的损伤类别。
- 30 6、根据权利要求 5 所述的方法，还包括：
从所述多个候选样本中选择至少一部分候选样本；

基于所述损伤识别模型，预测所述至少一部分候选样本中各个候选样本对应的损伤类别；

接收针对所述至少一部分候选样本的损伤类别进行人工修正的人工修正结果；

基于所述人工修正结果，标注出所述至少一部分候选样本的损伤类别；

- 5 将标注有损伤类别的至少一部分候选样本作为第二训练样本集，进一步训练所述损伤识别模型。

7、一种损伤类别的识别方法，包括：

获取当前定损图片；

将所述当前定损图片输入权利要求 5 或 6 所述的损伤识别模型中，预测所述当前定

- 10 损图片对应的损伤类别。

8、一种样本标注装置，包括：

获取单元，用于获取多张历史定损图片；

第一确定单元，用于基于显著性检测，从所述多张历史定损图片中确定出多个候选损伤区域，并将所述多个候选损伤区域的图像作为多个候选样本；

- 15 聚类单元，用于将所述多个候选样本聚类成多个类簇；

第二确定单元，用于确定所述多个类簇的类簇中心分别对应的多个中心候选样本集；

接收单元，用于接收人工标注结果，所述人工标注结果是针对所述多个中心候选样本集中的至少一部分中心候选样本集进行的损伤类别标注，其中同一中心候选样本集中的各个中心候选样本标注有相同的损伤类别；

- 20 标注单元，用于根据所述人工标注结果，对所述至少一部分中心候选样本集所在类簇中其他候选样本的损伤类别进行标注，以得到标注有损伤类别的多个标注样本。

9、根据权利要求 8 所述的装置，其中，所述第一确定单元具体用于：

检测所述多张历史定损图片中各张历史定损图片中的显著性对象；

确定所述显著性对象所在的最小矩形区域，并将所述最小矩形区域作为对应的候选

- 25 损伤区域。

10、根据权利要求 8 所述的装置，其中，所述聚类单元具体用于：

基于图像特征提取模型，确定与所述多个候选样本中各个候选样本对应的特征向量，以得到多个特征向量；

对所述多个特征向量进行聚类，得到所述多个类簇。

- 30 11、根据权利要求 8 所述的装置，其中，所述第二确定单元具体用于：

确定各个类簇中与对应的类簇中心距离最近的预定数量的中心候选样本，并将所述

预定数量的中心候选样本构成对应的中心候选样本集。

12、根据权利要求 8-11 任一项所述的装置，还包括第一训练单元，用于将所述多个标注样本作为第一训练样本集，训练损伤识别模型，所述损伤识别模型用于从图片中识别出车辆损伤的损伤类别。

5 13、根据权利要求 12 所述的装置，还包括第二训练单元，用于：

从所述多个候选样本中选择至少一部分候选样本；

基于所述损伤识别模型，预测所述至少一部分候选样本中各个候选样本对应的损伤类别；

接收针对所述至少一部分候选样本的损伤类别进行人工修正的人工修正结果；

10 基于所述人工修正结果，标注出所述至少一部分候选样本的损伤类别；

将标注有损伤类别的至少一部分候选样本作为第二训练样本集，进一步训练所述损伤识别模型。

14、一种损伤类别的识别装置，包括：

获取单元，用于获取当前定损图片；

15 预测单元，用于将所述当前定损图片输入权利要求 5、6、12、13 中任一项所述的损伤识别模型中，预测所述当前定损图片对应的损伤类别。

15、一种计算机可读存储介质，其上存储有计算机程序，当所述计算机程序在计算机中执行时，令计算机执行权利要求 1-7 中任一项的所述的方法。

16、一种计算设备，包括存储器和处理器，其特征在于，所述存储器中存储有可执行代码，所述处理器执行所述可执行代码时，实现权利要求 1-7 中任一项所述的方法。