



(19) 中華民國智慧財產局

(12) 發明說明書公告本

(11) 證書號數：TW I716012 B

(45) 公告日：中華民國 110 (2021) 年 01 月 11 日

(21) 申請案號：108121923

(22) 申請日：中華民國 108 (2019) 年 06 月 24 日

(51) Int. Cl. : G06F3/048 (2013.01)

G06F16/50 (2019.01)

(30) 優先權：2018/08/24 中國大陸

201810975325.X

(71) 申請人：開曼群島商創新先進技術有限公司 (開曼群島) ADVANCED NEW TECHNOLOGIES CO., LTD. (KY)

開曼群島

(72) 發明人：徐娟 (CN)

(74) 代理人：林志剛

(56) 參考文獻：

CN 107358596A

US 9760789B2

US 9940509B2

US 2018/0082378A1

審查人員：易昶霈

申請專利範圍項數：12 項 圖式數：6 共 35 頁

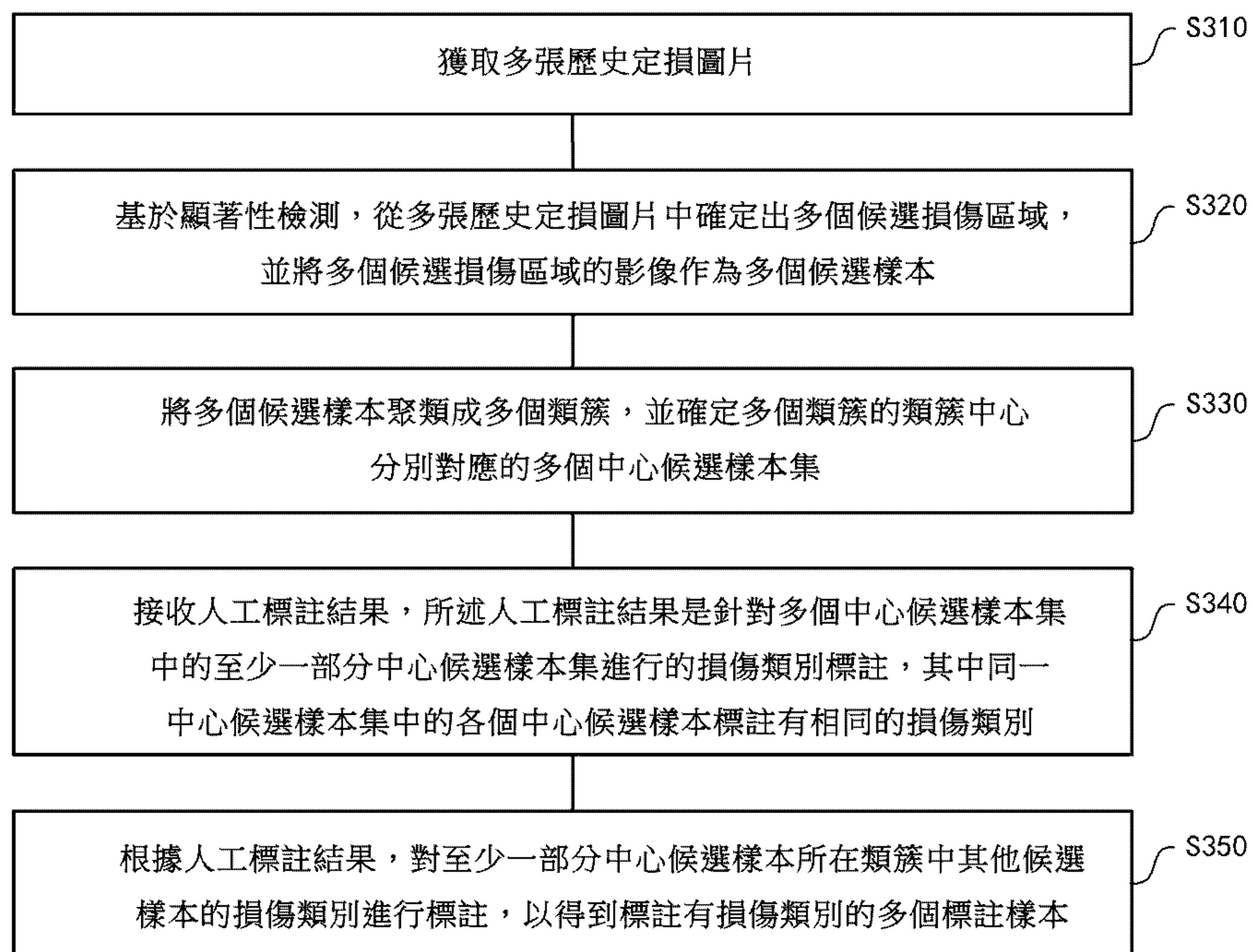
(54) 名稱

樣本標註方法、裝置、儲存媒體和計算設備、損傷類別的識別方法及裝置

(57) 摘要

本說明書實施例提供一種樣本標註方法，該方法包括：首先，獲取多張歷史定損圖片，並基於顯著性檢測，從多張歷史定損圖片中確定出多個候選損傷區域的影像作為多個候選樣本；接著，將多個候選樣本聚類成多個類簇，並確定多個類簇的類簇中心分別對應的多個中心候選樣本集；然後，接收人工標註結果，人工標註結果是針對多個中心候選樣本集中的至少一部分中心候選樣本集進行的損傷類別標註；接著，根據人工標註結果，對至少一部分中心候選樣本集所在類簇中其他候選樣本的損傷類別進行標註，以得到標註有損傷類別的多個標註樣本。

指定代表圖：



【圖 3】

I716012

**【發明摘要】****【中文發明名稱】**

樣本標註方法、裝置、儲存媒體和計算設備、損傷類別的識別方法及裝置

**【中文】**

本說明書實施例提供一種樣本標註方法，該方法包括：首先，獲取多張歷史定損圖片，並基於顯著性檢測，從多張歷史定損圖片中確定出多個候選損傷區域的影像作為多個候選樣本；接著，將多個候選樣本聚類成多個類簇，並確定多個類簇的類簇中心分別對應的多個中心候選樣本集；然後，接收人工標註結果，人工標註結果是針對多個中心候選樣本集中的至少一部分中心候選樣本集進行的損傷類別標註；接著，根據人工標註結果，對至少一部分中心候選樣本集所在類簇中其他候選樣本的損傷類別進行標註，以得到標註有損傷類別的多個標註樣本。

【指定代表圖】第(3)圖。  
【代表圖之符號簡單說明】無

【特徵化學式】無

## 【發明說明書】

### 【中文發明名稱】

樣本標註方法、裝置、儲存媒體和計算設備、損傷類別的識別方法及裝置

### 【技術領域】

本說明書實施例涉及模型訓練領域，具體地，涉及一種樣本標註方法及裝置、損傷類別的識別方法及裝置。

### 【先前技術】

在車險理賠場景，目前業界推出了用於理賠的用戶客戶端，根據用戶在客戶端中上傳的包括事故車輛的圖片，可以自動識別出事故車輛的損傷資訊，如損傷部件、損傷類別，並給出相關的理賠資訊，如維修方案、賠償金額等。這種方案因無需人工查勘定損核損，減少了保險公司的成本，提升了普通用戶的車險理賠體驗。

目前業界在識別圖片中事故車輛的損傷資訊時，通常是通過直接將該圖片與歷史資料庫中的海量圖片進行對比，找出相似圖片，以根據相似圖片對應的損傷資訊確定出事故車輛的損傷資訊。這種方法的計算量大，且由此確定出的損傷資訊的準確度不高。因此，需要一種合理的方法，可以更加快速、準確地確定出圖片中車輛的損傷資訊。

### 【發明內容】

在本說明書描述的一種樣本標註方法中，先對海量定損照片進行聚類，再根據接收的針對部分聚類中心的損傷類型進行人工標註的標註結果，對部分聚類中心所在類簇的其他圖片進行自動標識，以快速得到海量的弱標籤訓練樣本，進而可以基於這些弱標籤訓練樣本訓練損傷識別模型。

根據第一態樣，提供一種樣本標註方法，該方法包括：獲取多張歷史定損圖片；基於顯著性檢測，從所述多張歷史定損圖片中確定出多個候選損傷區域，並將所述多個候選損傷區域的影像作為多個候選樣本；將所述多個候選樣本聚類成多個類簇，並確定所述多個類簇的類簇中心分別對應的多個中心候選樣本集；接收人工標註結果，所述人工標註結果是針對所述多個中心候選樣本集中的至少一部分中心候選樣本集進行的損傷類別標註，其中同一中心候選樣本集中的各個中心候選樣本標註有相同的損傷類別；根據所述人工標註結果，對所述至少一部分中心候選樣本集所在類簇中其他候選樣本的損傷類別進行標註，以得到標註有損傷類別的多個標註樣本。

根據一個實施例，所述從所述多張歷史定損圖片中確定出多個候選損傷區域，包括：檢測所述多張歷史定損圖片中各張歷史定損圖片中的顯著性對象；確定所述顯著性對象所在的最小矩形區域，並將所述最小矩形區域作為對應的候選損傷區域。

根據一個實施例，所述將所述多個候選樣本聚類成多

個類簇，包括：基於影像特徵提取模型，確定與所述多個候選樣本中各個候選樣本對應的特徵向量，以得到多個特徵向量；對所述多個特徵向量進行聚類，得到所述多個類簇。

根據一個實施例，所述確定所述多個類簇的類簇中心分別對應的多個中心候選樣本集，包括：確定各個類簇中與對應的類簇中心距離最近的預定數量的中心候選樣本，並將所述預定數量的中心候選樣本構成對應的中心候選樣本集。

根據一個實施例，還包括：將所述多個標註樣本作為第一訓練樣本集，訓練損傷識別模型，所述損傷識別模型用於從圖片中識別出車輛損傷的損傷類別。

根據一個實施例，還包括：從所述多個候選樣本中選擇至少一部分候選樣本；基於所述損傷識別模型，預測所述至少一部分候選樣本中各個候選樣本對應的損傷類別；接收針對所述至少一部分候選樣本的損傷類別進行人工修正的人工修正結果；基於所述人工修正結果，標註出所述至少一部分候選樣本的損傷類別；將標註有損傷類別的至少一部分候選樣本作為第二訓練樣本集，進一步訓練所述損傷識別模型。

根據第二態樣，提供一種損傷類別的識別方法，該方法包括：獲取當前定損圖片；將所述當前定損圖片輸入上述第一態樣的任一實施例的損傷識別模型中，預測所述當前定損圖片對應的損傷類別。

根據第三態樣，提供一種樣本標註裝置，該裝置包括：獲取單元，用於獲取多張歷史定損圖片；第一確定單元，用於基於顯著性檢測，從所述多張歷史定損圖片中確定出多個候選損傷區域，並將所述多個候選損傷區域的影像作為多個候選樣本；聚類單元，用於將所述多個候選樣本聚類成多個類簇；第二確定單元，用於確定所述多個類簇的類簇中心分別對應的多個中心候選樣本集；接收單元，用於接收人工標註結果，所述人工標註結果是針對所述多個中心候選樣本集中的至少一部分中心候選樣本集進行的損傷類別標註，其中同一中心候選樣本集中的各個中心候選樣本標註有相同的損傷類別；標註單元，用於根據所述人工標註結果，對所述至少一部分中心候選樣本集所在類簇中其他候選樣本的損傷類別進行標註，以得到標註有損傷類別的多個標註樣本。

根據第四態樣，提供一種損傷類別的識別裝置，該裝置包括：獲取單元，用於獲取當前定損圖片；預測單元，用於將所述當前定損圖片輸入第一態樣或第三態樣的任一實施例中所述的損傷識別模型中，預測所述當前定損圖片對應的損傷類別。

根據第五態樣，提供了一種電腦可讀儲存媒體，其上儲存有電腦程式，當所述電腦程式在電腦中執行時，令電腦執行第一態樣或第二態樣所描述的方法。

根據第六態樣，提供了一種計算設備，包括記憶體和處理器，其特徵在於，所述記憶體中儲存有可執行程式

碼，所述處理器執行所述可執行程式碼時，實現第一態樣或第二態樣所描述的方法。

在本說明書實施例揭露的樣本標註方法中，先通過顯著性檢測識別出海量定損照片中的多個候選損傷區域，然後對這些候選損傷區域進行聚類，根據針對部分類簇中心所對應區域的損傷類型進行人工標註的標註結果，對部分類簇中心所在類簇的其他候選區域進行自動標識，如此可以快速得到大量的訓練樣本，進而可以根據這些訓練樣本訓練損傷識別模型。

### 【圖式簡單說明】

為了更清楚地說明本說明書揭露的多個實施例的技術方案，下面將對實施例描述中所需要使用的圖式作簡單地介紹，顯而易見地，下面描述中的圖式僅僅是本說明書揭露的多個實施例，對於本領域普通技術人員來講，在不付出創造性勞動的前提下，還可以根據這些圖式獲得其它的圖式。

圖1示出根據一個例子的車輛局部照片；

圖2示出根據一個例子的定損客戶端的界面變化示意圖；

圖3示出根據一個實施例的樣本標註方法流程圖；

圖4示出根據一個實施例的候選損傷區域的識別過程示意圖；

圖5示出根據一個實施例的損傷識別模型的優化訓練

方法流程圖；

圖6示出根據一個實施例的樣本標註裝置結構圖。

### 【實施方式】

下面結合圖式，對本說明書揭露的多個實施例進行描述。

本說明書實施例揭露一種樣本標註方法，基於此方法得到的訓練樣本可以用於訓練損傷識別模型。下面，首先對損傷識別模型的應用場景進行介紹。

損傷識別模型可以應用於提供給用戶使用的定損客戶端中。根據一個例子，在事故現場，用戶可以通過終端，如，手機、平板電腦等，拍攝現場照片，例如圖1中示出的車輛局部圖片，並將拍攝的照片上傳到終端中的定損客戶端，然後定損客戶端可以利用損傷識別模型，確定現場照片所對應的車輛損傷資訊，例如，如圖2所示，可以確定出車輛損傷類別為中度刮擦，進一步地，還可以給出與損傷資訊對應的維修方案和相關賠償金額，例如，維修方案為：補漆，保修賠付金額：120元。

對於損傷識別模型的訓練，在一種實施方案中，可以基於大量人工標註的樣本進行訓練。在這種方案中，可以先從保險公司獲取大量的歷史定損照片，如可以包括保險公司的定損人員在事故現場拍攝的照片，然後由標識人員標註出定損照片中的損傷對象，以及對應的損傷類別，再將這些標註樣本作為訓練樣本，訓練損傷識別模型。然

而，這種方案中人工標註成本過高，因此實施起來比較困難。

基於此，本說明書實施例揭露一種樣本標註方法，先通過顯著性檢測識別出海量定損照片中的多個候選損傷區域，然後對這些候選損傷區域進行聚類，根據針對部分類簇中心所對應區域的損傷類型進行人工標註的標註結果，對部分類簇中心所在類簇的其他候選區域進行自動標識，如此可以快速得到大量的訓練樣本，進而可以根據這些訓練樣本訓練損傷識別模型。下面，描述以上過程的具體實施步驟。

圖3示出根據一個實施例的樣本標註方法流程圖，所述方法的執行主體可以為具有處理能力的設備：服務器或者系統或者裝置。如圖3所示，該方法流程包括以下步驟：步驟S310，獲取多張歷史定損圖片；步驟S320，採用顯著性檢測，從多張歷史定損圖片中確定出多個候選損傷區域，並將多個候選損傷區域的影像作為多個候選樣本；步驟S330，將多個候選樣本聚類成多個類簇，並確定多個類簇的類簇中心分別對應的多個中心候選樣本集；步驟S340，接收人工標註結果，所述人工標註結果是針對多個中心候選樣本集中的至少一部分中心候選樣本集進行的損傷類別標註，其中同一中心候選樣本集中的各個中心候選樣本標註有相同的損傷類別；步驟S350，根據人工標註結果，對至少一部分中心候選樣本集所在類簇中其他候選樣本的損傷類別進行標註，以得到標註有損傷類別的多個標

註樣本。

首先，在步驟 S310，獲取多張歷史定損圖片。

在一個實施例中，可以從保險公司的資料庫中獲取歷史定損圖片。在另一個實施例中，還可以在與交通事故相關的論壇或新聞網站中獲取歷史定損圖片。

在一個實施例中，多張歷史定損圖片可能涵蓋反映事故車輛不同資訊的圖片，例如，反映損傷情況、車牌號、車輛型號等資訊的圖片。在一個例子中，歷史定損圖片可以包括反映車輛損傷情況的車身的局部或整體照片。在另一個例子中，歷史定損圖片可以包括拍攝的駕駛證圖片或行駛證圖片。換而言之，歷史定損圖片中有可能包含與損傷對象不直接相關的圖片內容。

在獲取多張歷史定損照片以後，接著，在步驟 S320，採用顯著性檢測，從多張歷史定損圖片中確定出多個候選損傷區域的影像作為多個候選樣本。

需要說明的是，顯著性檢測(或稱為視覺注意機制)是指通過智慧型演算法模擬人的視覺特點，提取影像中的顯著性對象(即人類感興趣的對象)。通常，顯著性對象具有一定的結構和紋理，色彩上具有較強的刺激，且顯著性對象和周圍區域具有較大的差異，基於這些簡單的顯著性檢測原則，可以實現顯著性檢測。

對於步驟 S310中獲取的歷史定損照片來說，因定損員在拍攝車輛損傷的過程中，其拍攝目的主要在於清晰地反映損傷對象，因此損傷對象通常位於圖片的中心部位，且

損傷對象與車身其他部位之間，在顏色、結構和紋理上通常具有較大的反差。因此，如果歷史定損圖片中包含損傷對象資訊，那麼可以通過對該圖片的中心區域採用顯著性檢測，快速定位出圖片中的損傷對象。

進一步地，根據一個具體的實施例，識別出多個候選損傷區域可以包括：首先採用顯著性檢測，識別多張歷史定損圖片中各張歷史定損圖片中的顯著性對象，然後確定出顯著性對象所在的最小矩形區域，並將此最小矩形區域作為對應的候選損傷區域。在一個例子中，確定出最小矩形區域可以包括：根據顯著性對象的座標資訊確定出最小矩形區域，例如，其中座標資訊可以包括顯著性對象中最上、最下、最左和最右四個點的座標資訊。根據一個具體的例子，根據圖4中(a)示出的車身局部影像，可以得到圖4中(b)示出的顯著性度量圖，其中白色部分為檢測到的顯著性對象，然後可以確定出圖4中(c)示出的包含顯著性對象的最小矩形區域，並將此作為候選損傷區域。

在一個實施例中，可以從某張歷史定損圖片中識別出一個或多個候選損傷區域。

以上，可以得到多個候選損傷區域，並可以將多個候選損傷區域的影像作為多個候選樣本。需要說明的是，由於歷史定損圖片中還可能包含其他車輛資訊，因此，採用顯著性檢測從圖片中識別的顯著性對象有可能包含非損傷對象，例如汽車其他無關部件，背景中的物體，用戶的證件等等。由此形成的多個候選樣本會有一部分樣本並不包

含損傷對象。

對於如上獲得的候選樣本，接著，在步驟 S330，將多個候選樣本聚類成多個類簇，並確定多個類簇的類簇中心分別對應的多個中心候選樣本集。

在一個實施例中，將多個候選樣本聚類成多個類簇可以包括：先基於影像特徵提取模型，確定與所述多個候選樣本中各個候選樣本對應的特徵向量，以得到多個特徵向量，再對多個特徵向量進行聚類，得到多個類簇。進一步地，需要說明的是，可以採用現有的影像特徵提取模型確定特徵向量，如，ImageNet訓練的ResNet模型，或方向梯度直方圖(Histogram of Oriented Gradient, HOG)特徵提取模型等，在此不作限定。

在一個實施例中，可以採用現有的多種聚類演算法，將多個候選樣本聚類成多個類簇。其中，多種聚類演算法可以包括 K-MEANS 演算法、K-MEDOIDS 演算法、CLARANS 演算法等劃分聚類演算法；BIRCH 演算法、CURE 演算法、CHAMELEON 演算法等層次聚類演算法；以及 DBSCAN 演算法、OPTICS 演算法、DENCLUE 演算法等密度聚類演算法。在一個具體的實施例中，在採用 K-MEANS 演算法時需要預先設定類簇的數量 K，且 K 的值越大，聚類得到的類簇數量越多，相應地分類精度越高，具體可以由工作人員結合業務經驗，以及聚類的反饋結果，逐次對 K 進行調整，以得到合適的多個類簇。

以上可以得到多個類簇，且各個類簇包括對應的各個

類簇中心，接著，可以確定各個類簇中心對應的各個中心候選樣本集。

根據一個實施例，可以確定各個類簇中與對應的類簇中心距離最近的預定數量的中心候選樣本，並將確定出的預定數量的中心候選樣本構成與該類簇對應的中心候選樣本集。需要說明的是，通常情況下，預定數量遠遠小於各個類簇中候選樣本的總數量。在一個例子中，一個類簇中包括1萬個候選樣本，而預定數量可以為50個或100個，等等。

根據一個實施例中採用的聚類演算法，類簇中心對應於數學上的一個向量，也就是各個圖片的特徵向量所映射到的向量空間中的一個點，稱為中心點。在這樣的情況下，可以將距離該中心點最近的預定數量的特徵向量所對應的多個候選樣本構成該類簇的中心候選樣本集。

由上可以得到與多個類簇中心分別對應的多個中心候選樣本集，然後，可以將這些中心候選樣本集提供給標識人員，以標註出對應的損傷類別。進一步地，在一個具體的實施例中，可以將與中心候選樣本集對應的歷史定損照片提供給標識人員，且歷史定損照片中標識出了與中心候選樣本集對應的候選損傷區域。在另一個具體的實施例中，可以只向標識人員展示與中心候選樣本集對應的候選損傷區域，例如，可以從歷史定損圖片中切割出與此候選損傷區域對應的子圖片，並將子圖片提供給標識人員。

接著，在步驟S340，接收人工標註結果，所述人工標

註結果是針對多個中心候選樣本集中的至少一部分中心候選樣本集進行的損傷類別標註，其中同一中心候選樣本集中的各個中心候選樣本標註有相同的損傷類別。

需要說明的是，通常情況下，之所以在各個類簇中選取多個中心候選樣本構成中心候選樣本集，並將其提供給工作人員以標註與該類簇對應的損傷類別，而不是只選擇類簇中心所對應的單個中心候選樣本提供給工作人員，是因為提供與各個類簇對應的多個中心候選樣本，可以使工作人員更加準確地對各個類簇的損傷類別進行判斷和標註。

如前所述，步驟 S320 中形成的多個候選樣本中會有一部分樣本並不包含損傷對象。經過步驟 S330 的聚類，包含相似顯著性對象的樣本會被聚類到同一類簇，類簇中心對應的中心候選樣本集可以反映該類簇中最為典型的顯著性對象。那麼相應地，在確定出的所有中心候選樣本集中，也會有部分中心候選樣本集中的顯著性對象不是損傷對象，而是其他對象，如車禍現場的垃圾桶、樹木、路人、證件，或未被損壞的車輛部件等。因此，標識人員可以先對全部的中心候選樣本集進行篩選，然後只對包括損傷對象的中心候選樣本集進行標識，也就是只針對多個中心候選樣本集中的至少一部分中心候選樣本集，進行損傷類別的標註。

在一個實施例中，對於多個中心候選樣本集中的任意兩個中心候選樣本集，在人工標註結果中，它們所對應的

損傷類別可以相同也可以不同。

根據一個具體的例子，假定從歷史定損圖片中識別出100萬個候選樣本，通過步驟S330的聚類，這些候選樣本被聚類為100個類簇，平均每個類簇1萬個樣本。接著，將這100個類簇對應的100個中心候選樣本集發送給標識人員進行標識，其中各個中心候選樣本集中包括100個中心候選樣本。標識人員可以從這100個樣本集中首先排除掉例如45個非損傷對象的樣本集，只對剩餘的55個中心候選樣本集，共計5500個中心候選樣本，進行標識。由於中心候選樣本集是該類簇中候選樣本的典型代表，因此在排除掉45個中心候選樣本集的同時，就意味著，這45個類簇所對應的所有候選樣本(45萬個)得到排除，不必作為標註對象。如此，通過聚類操作和簡單的篩選，排除掉大量的干擾樣本，極大減小人工成本。標識人員只需要對極少量的中心候選樣本集進行標註即可。

根據一個具體的例子，人工標註結果可以包括：中心候選樣本集A的損傷類別為輕度刮擦，中心候選集B和C的損傷類別為中度變形，中心候選樣本集D的損傷類別為重度變形等。

以上，可以得到多個中心候選樣本集中的至少一部分的損傷類別進行人工標註的人工標註結果。接著，然後，在步驟S350，根據人工標註結果，對至少一部分中心候選樣本集所在類簇中其他候選樣本的損傷類別進行標註，以得到標註有損傷類別的多個標註樣本。

在一個實施例中，對於標註有損傷類別的各個中心候選樣本集，可以將其所在類簇中其他候選樣本標註上與該中心候選樣本集相同的損傷類別。在一個具體的例子中，某個類簇中有1萬個候選樣本，其中心候選樣本集中的100個中心候選樣本被人工標註為重度變形，據此，其他的9900個候選樣本可以被自動標註為重度變形。如此，通過對少量中心候選樣本集進行標註，可以實現對其他候選樣本進行自動標識。

以上，可以基於輕量級的人工標註，得到標註有損傷類別的多個標註樣本。具體地，在以上100萬個候選樣本的例子中，在將候選樣本聚類為100個類簇之後，只需要對55個中心候選樣本集，共計5500個中心候選樣本，進行人工標註，然後對應地對類簇的其他樣本進行自動標識，就可以得到55萬個標註樣本。這樣標註的樣本可以稱為弱標籤樣本。

需要說明的是，在步驟S350之後還可以包括：將以上得到的多個標註樣本作為第一訓練樣本集，訓練損傷識別模型。

在一個實施例中，可以基於現有的分類演算法，訓練損傷識別模型。在一個例子中，分類演算法可以包括貝葉斯分類演算法、支持向量機、人工神經網路和整合學習等，在此不作限定。

通過執行本說明書實施例提供的樣本標註方法，實現了基於少量的人工標註，得到大量的標註樣本，進而基於

這些標註樣本可以得到具有一定準確度的損傷識別模型，解決了損傷檢測模型需要耗費大量人工標註成本的問題。

進一步地，為了提升上述損傷識別模型的準確度，還可以先使用這個基於海量弱標籤資料訓練的損傷識別模型，對一定數量的候選樣本進行演算法預標註，再由人工進行檢查更正，得到人工檢查修改後的訓練集，再基於此訓練集，進一步訓練損傷識別模型。下面，結合圖5對以上過程進行描述。

圖5示出根據一個實施例的損傷識別模型的優化訓練方法流程圖。如圖5所示，該方法流程包括以下步驟：

首先，在步驟S510，從多個候選樣本中選擇至少一部分候選樣本。

在一個實施例中，可以從前述步驟S320中確定出的多個候選樣本中，隨機選擇出預定數量的候選樣本，如此確定出的候選樣本中可能會包括顯著性對象不是損傷對象的負樣本，其中預定數量可以根據實際的精度需要靈活選擇。在另一個實施例中，可以從各個類簇中，按照固定比例，或者隨機地，選擇出預定數量的候選樣本。

以上，在選擇出至少一部分候選樣本後，接著，在步驟S520，基於損傷識別模型，預測至少一部分候選樣本中各個候選樣本對應的損傷類別。

具體地，將選擇出的至少一部分候選樣本輸入前述訓練得到的損傷識別模型中，以得到其中各個候選樣本對應的損傷類別。需要說明的是，因前述損傷識別模型是基於

弱標籤資料訓練得到，而這些弱標籤本身的準確度並不能達到很高的要求，因此訓練得到的初步的損傷識別模型的準確率也不會達到非常高的水平，預測出的損傷類別中會存在一些錯誤。

於是，進一步地，在一個實施例中，將步驟 S520 預測出損傷類別的候選樣本提供給標識人員，以使標識人員對預測出的損傷類別進行檢查修正。在步驟 S530，接收針對至少一部分候選樣本的損傷類別進行人工修正的人工修正結果。然後，在步驟 S540，基於人工修正結果，標註出至少一部分候選樣本的損傷類別。

接著，在步驟 S550，將標註有損傷類別的至少一部分候選樣本作為第二訓練樣本集，進一步訓練所述損傷識別模型。

如此，經過檢查修正這種低成本地人工標註，可以得到一定數量的更加精準的第二訓練樣本集，進而可以根據第二訓練樣本集，進一步訓練損傷識別模型，以提高損傷識別模型的準確率。

根據另一態樣的實施例，還提供一種損傷類別的識別方法，所述方法可以基於前述實施例中訓練得到的損傷識別模型進行。具體地，所述方法可以包括：首先，獲取當前定損圖片；然後，將當前定損圖片輸入前述任一實施例中提供的損傷識別模型中，預測當前定損圖片對應的損傷類別。

根據再一態樣的實施例，還提供一種樣本標註裝置。

圖 6 示出根據一個實施例的樣本標註裝置結構圖。如圖 6 所示，該裝置 600 包括：

獲取單元 610，用於獲取多張歷史定損圖片；

第一確定單元 620，用於基於顯著性檢測，從所述多張歷史定損圖片中確定出多個候選損傷區域的影像作為多個候選樣本；

聚類單元 630，用於將所述多個候選樣本聚類成多個類簇；

第二確定單元 640，用於確定所述多個類簇的類簇中心分別對應的多個中心候選樣本集；

接收單元 650，用於接收人工標註結果，所述人工標註結果是針對所述多個中心候選樣本集中的至少一部分中心候選樣本集進行的損傷類別標註，其中同一中心候選樣本集中的各個中心候選樣本標註有相同的損傷類別；

標註單元 660，用於根據所述人工標註結果，對所述至少一部分中心候選樣本集所在類簇中其他候選樣本的損傷類別進行標註，以得到標註有損傷類別的多個標註樣本。

根據一個實施例，所述第一確定單元 620 具體用於：

檢測所述多張歷史定損圖片中各張歷史定損圖片中的顯著性對象；

確定所述顯著性對象所在的最小矩形區域，並將所述最小矩形區域作為對應的候選損傷區域。

根據一個實施例，所述聚類單元 630 具體用於：

基於影像特徵提取模型，確定與所述多個候選樣本中各個候選樣本對應的特徵向量，以得到多個特徵向量；

對所述多個特徵向量進行聚類，得到所述多個類簇。

根據一個實施例，所述第二確定單元640具體用於：

確定各個類簇中與對應的類簇中心距離最近的預定數量的中心候選樣本，並將所述預定數量的中心候選樣本構成對應的中心候選樣本集。

根據一個實施例，還包括第一訓練單元670，用於將所述多個標註樣本作為第一訓練樣本集，訓練損傷識別模型，所述損傷識別模型用於從圖片中識別出車輛損傷的損傷類別。

根據一個實施例，還包括第二訓練單元680，用於：

從所述多個候選樣本中選擇至少一部分候選樣本；

基於所述損傷識別模型，預測所述至少一部分候選樣本中各個候選樣本對應的損傷類別；

接收針對所述至少一部分候選樣本的損傷類別進行人工修正的人工修正結果；

基於所述人工修正結果，標註出所述至少一部分候選樣本的損傷類別；

將標註有損傷類別的至少一部分候選樣本作為第二訓練樣本集，進一步訓練所述損傷識別模型。

總之，採用本說明書實施例提供的樣本標註裝置，實現了基於少量的人工標註，得到大量的標註樣本，進而基於這些標註樣本可以得到具有一定準確度的損傷識別模

型，解決了損傷檢測模型需要耗費大量人工標註成本的問題。

根據再一態樣的實施例，還提供一種損傷類別的識別裝置，所述裝置可以基於前述實施例中訓練得到的損傷識別模型進行。具體地，所述裝置可以包括：獲取單元，用於獲取當前定損圖片；預測單元，用於將當前定損圖片輸入前述任一實施例中提供的損傷識別模型中，預測當前定損圖片對應的損傷類別。

如上，根據又一態樣的實施例，還提供一種電腦可讀儲存媒體，其上儲存有電腦程式，當所述電腦程式在電腦中執行時，令電腦執行結合圖3或圖5所描述的方法。

根據又一態樣的實施例，還提供一種計算設備，包括記憶體和處理器，所述記憶體中儲存有可執行程式碼，所述處理器執行所述可執行程式碼時，實現結合圖3或圖5所描述的方法。

本領域技術人員應該可以意識到，在上述一個或多個示例中，本說明書揭露的多個實施例所描述的功能可以用硬體、軟體、韌體或它們的任意組合來實現。當使用軟體實現時，可以將這些功能儲存在電腦可讀媒體中或者作為電腦可讀媒體上的一個或多個指令或程式碼進行傳輸。

以上所述的具體實施方式，對本說明書揭露的多個實施例的目的、技術方案和有益效果進行了進一步詳細說明，所應理解的是，以上所述僅為本說明書揭露的多個實施例的具體實施方式而已，並不用於限定本說明書揭露的

多個實施例的保護範圍，凡在本說明書揭露的多個實施例的技術方案的基礎之上，所做的任何修改、等同替換、改進等，均應包括在本說明書揭露的多個實施例的保護範圍之內。

**【符號說明】**

S310、S320、S330、S340、S350：步驟

S510、S520、S530、S540、S550：步驟

600：裝置

610：獲取單元

620：第一確定單元

630：聚類單元

640：第二確定單元

650：接收單元

660：標註單元

670：第一訓練單元

680：第二訓練單元

## 【發明申請專利範圍】

### 【第 1 項】

一種樣本標註方法，包括：

獲取多張歷史定損圖片；

基於顯著性檢測，從所述多張歷史定損圖片中確定出多個候選損傷區域，並將所述多個候選損傷區域的影像作為多個候選樣本；

將所述多個候選樣本聚類成多個類簇，並確定所述多個類簇的類簇中心分別對應的多個中心候選樣本集；

接收人工標註結果，所述人工標註結果是針對所述多個中心候選樣本集中的至少一部分中心候選樣本集進行的損傷類別標註，其中同一中心候選樣本集中的各個中心候選樣本標註有相同的損傷類別；

根據所述人工標註結果，對所述至少一部分中心候選樣本集所在類簇中其他候選樣本的損傷類別進行標註，以得到標註有損傷類別的多個標註樣本；

將所述多個標註樣本作為第一訓練樣本集，訓練損傷識別模型，所述損傷識別模型用於從圖片中識別出車輛損傷的損傷類別；

從所述多個候選樣本中選擇至少一部分候選樣本；

基於所述損傷識別模型，預測所述至少一部分候選樣本中各個候選樣本對應的損傷類別；

接收針對所述至少一部分候選樣本的損傷類別進行人工修正的人工修正結果；

基於所述人工修正結果，標註出所述至少一部分候選樣本的損傷類別；

將標註有損傷類別的所述至少一部分候選樣本作為第二訓練樣本集，進一步訓練所述損傷識別模型。

**【第2項】**

根據申請專利範圍第1項所述的方法，其中，所述從所述多張歷史定損圖片中確定出多個候選損傷區域，包括：

檢測所述多張歷史定損圖片中各張歷史定損圖片中的顯著性對象；

確定所述顯著性對象所在的最小矩形區域，並將所述最小矩形區域作為對應的候選損傷區域。

**【第3項】**

根據申請專利範圍第1項所述的方法，其中，所述將所述多個候選樣本聚類成多個類簇，包括：

基於影像特徵提取模型，確定與所述多個候選樣本中各個候選樣本對應的特徵向量，以得到多個特徵向量；

對所述多個特徵向量進行聚類，得到所述多個類簇。

**【第4項】**

根據申請專利範圍第1項所述的方法，其中，所述確定所述多個類簇的類簇中心分別對應的多個中心候選樣本集，包括：

確定各個類簇中與對應的類簇中心距離最近的預定數量的中心候選樣本，並將所述預定數量的中心候選樣本構

成對應的中心候選樣本集。

**【第5項】**

一種損傷類別的識別方法，包括：

獲取當前定損圖片；

將所述當前定損圖片輸入損傷識別模型中，預測所述當前定損圖片對應的損傷類別，

其中，用以下步驟來訓練所述損傷識別模型：

將所述多個標註樣本作為第一訓練樣本集，訓練所述損傷識別模型，所述損傷識別模型用於從圖片中識別出車輛損傷的損傷類別；

從所述多個候選樣本中選擇至少一部分候選樣本；

基於所述損傷識別模型，預測所述至少一部分候選樣本中各個候選樣本對應的損傷類別；

接收針對所述至少一部分候選樣本的損傷類別進行人工修正的人工修正結果；

基於所述人工修正結果，標註出所述至少一部分候選樣本的損傷類別；

將標註有損傷類別的所述至少一部分候選樣本作為第二訓練樣本集，進一步訓練所述損傷識別模型。

**【第6項】**

一種樣本標註裝置，包括：

獲取單元，用於獲取多張歷史定損圖片；

第一確定單元，用於基於顯著性檢測，從在所述獲取單元獲取的所述多張歷史定損圖片中確定出多個候選損傷

區域，並將所述多個候選損傷區域的影像作為多個候選樣本；

聚類單元，用於將在所述第一確定單元確定的所述多個候選樣本聚類成多個類簇；

第二確定單元，用於確定在所述聚類單元中聚類的所述多個類簇的類簇中心分別對應的多個中心候選樣本集；

接收單元，用於接收人工標註結果，所述人工標註結果是針對在所述第二確定單元確定的所述多個中心候選樣本集中的至少一部分中心候選樣本集進行的損傷類別標註，其中同一中心候選樣本集中的各個中心候選樣本標註有相同的損傷類別；

標註單元，用於根據在所述接收單元接收的所述人工標註結果，對所述至少一部分中心候選樣本集所在類簇中其他候選樣本的損傷類別進行標註，以得到標註有損傷類別的多個標註樣本；

第一訓練單元，用於將在所述標註單元得到的所述多個標註樣本作為第一訓練樣本集，訓練損傷識別模型，所述損傷識別模型用於從圖片中識別出車輛損傷的損傷類別；

第二訓練單元，用於：

從在所述第一確定單元確定的所述多個候選樣本中選擇至少一部分候選樣本；

基於所述損傷識別模型，預測所述至少一部分候選樣本中各個候選樣本對應的損傷類別；

接收針對所述至少一部分候選樣本的損傷類別進行人工修正的人工修正結果；

基於所述人工修正結果，標註出所述至少一部分候選樣本的損傷類別；

將標註有損傷類別的所述至少一部分候選樣本作為第二訓練樣本集，進一步訓練所述損傷識別模型。

**【第7項】**

根據申請專利範圍第6項所述的裝置，其中，所述第一確定單元具體用於：

檢測在所述獲取單元獲取的所述多張歷史定損圖片中各張歷史定損圖片中的顯著性對象；

確定所述顯著性對象所在的最小矩形區域，並將所述最小矩形區域作為對應的候選損傷區域。

**【第8項】**

根據申請專利範圍第6項所述的裝置，其中，所述聚類單元具體用於：

基於影像特徵提取模型，確定與在所述第一確定單元確定的所述多個候選樣本中各個候選樣本對應的特徵向量，以得到多個特徵向量；

對所述多個特徵向量進行聚類，得到所述多個類簇。

**【第9項】**

根據申請專利範圍第6項所述的裝置，其中，所述第二確定單元具體用於：

確定各個類簇中與對應的類簇中心距離最近的預定數

量的中心候選樣本，並將所述預定數量的中心候選樣本構成對應的中心候選樣本集。

**【第10項】**

一種損傷類別的識別裝置，包括：

獲取單元，用於獲取當前定損圖片；

預測單元，用於將所述當前定損圖片輸入損傷識別模型中，預測所述當前定損圖片對應的損傷類別；

其中，所述識別裝置包括以下單元來訓練所述損傷識別模型：

第一訓練單元，用於將在所述標註單元得到的所述多個標註樣本作為第一訓練樣本集，訓練損傷識別模型，所述損傷識別模型用於從圖片中識別出車輛損傷的損傷類別；

第二訓練單元，用於：

從在所述第一確定單元確定的所述多個候選樣本中選擇至少一部分候選樣本；

基於所述損傷識別模型，預測所述至少一部分候選樣本中各個候選樣本對應的損傷類別；

接收針對所述至少一部分候選樣本的損傷類別進行人工修正的人工修正結果；

基於所述人工修正結果，標註出所述至少一部分候選樣本的損傷類別；

將標註有損傷類別的所述至少一部分候選樣本作為第二訓練樣本集，進一步訓練所述損傷識別模型。

**【第 11 項】**

一種電腦可讀儲存媒體，其上儲存有電腦程式，當所述電腦程式在電腦中執行時，令電腦執行申請專利範圍第 1 至 5 項中任一項的所述的方法。

**【第 12 項】**

一種計算設備，包括記憶體和處理器，其特徵在於，所述記憶體中儲存有可執行程式碼，所述處理器執行所述可執行程式碼時，實現申請專利範圍第 1 至 5 項中任一項所述的方法。

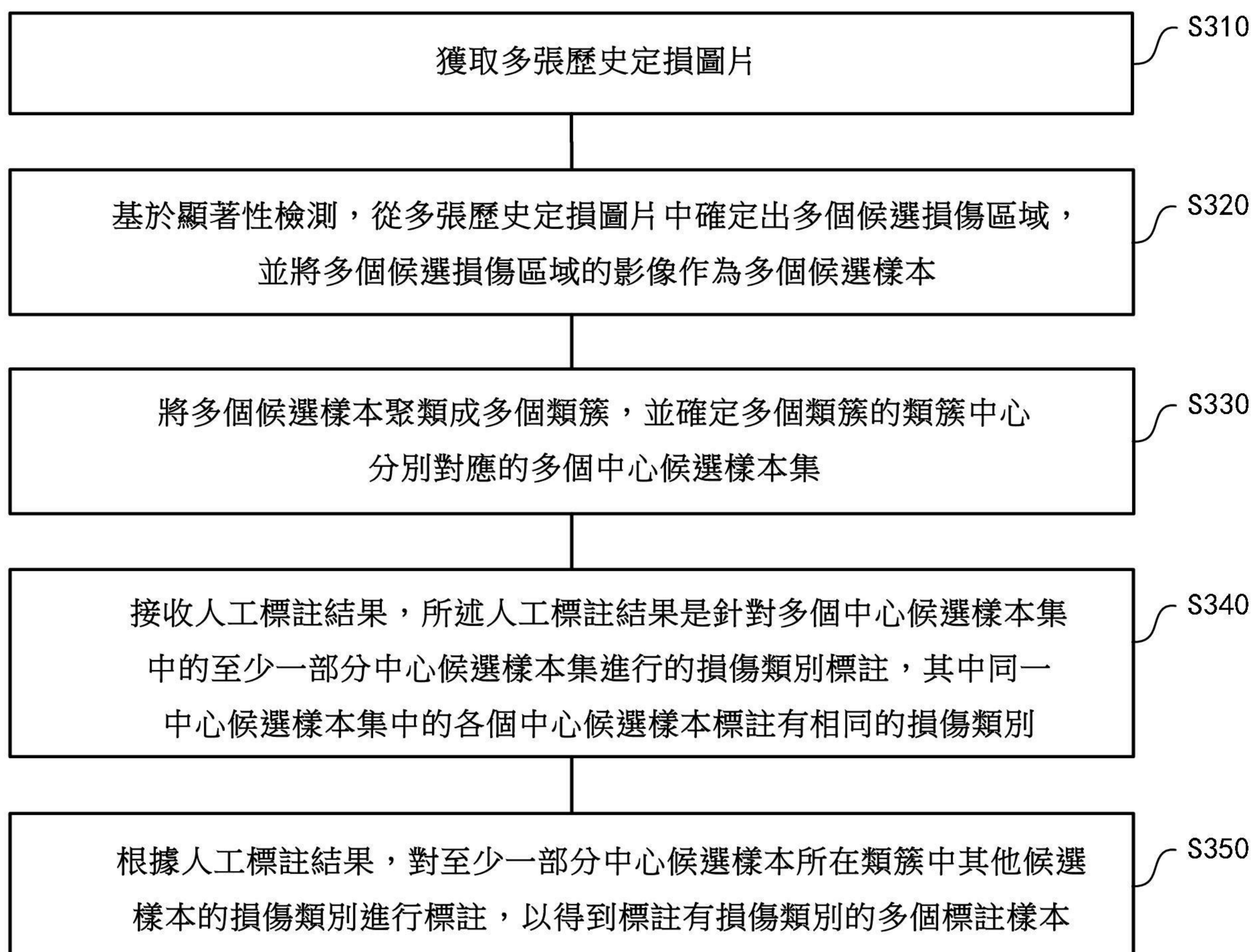
【發明圖式】



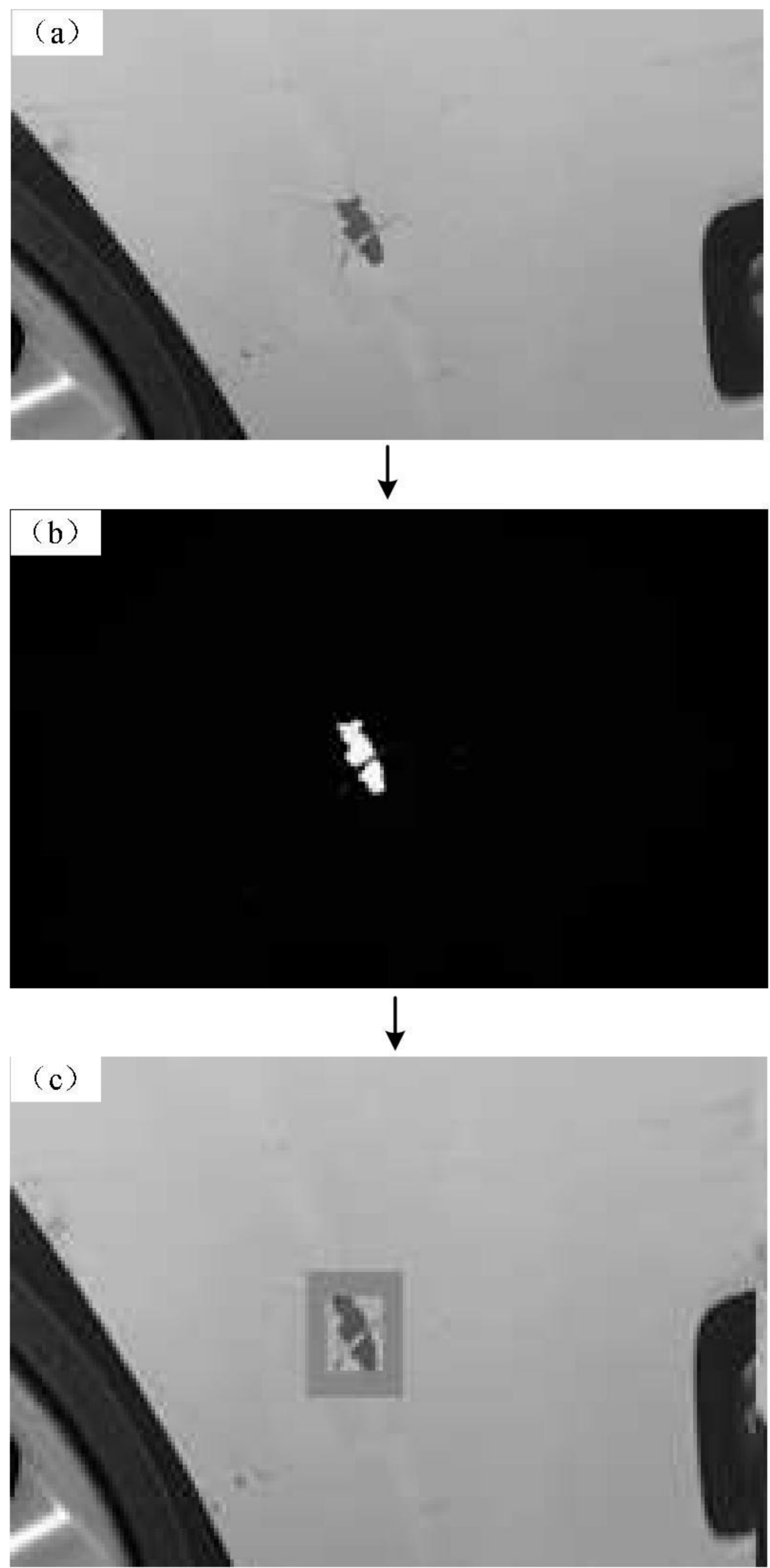
【圖 1】



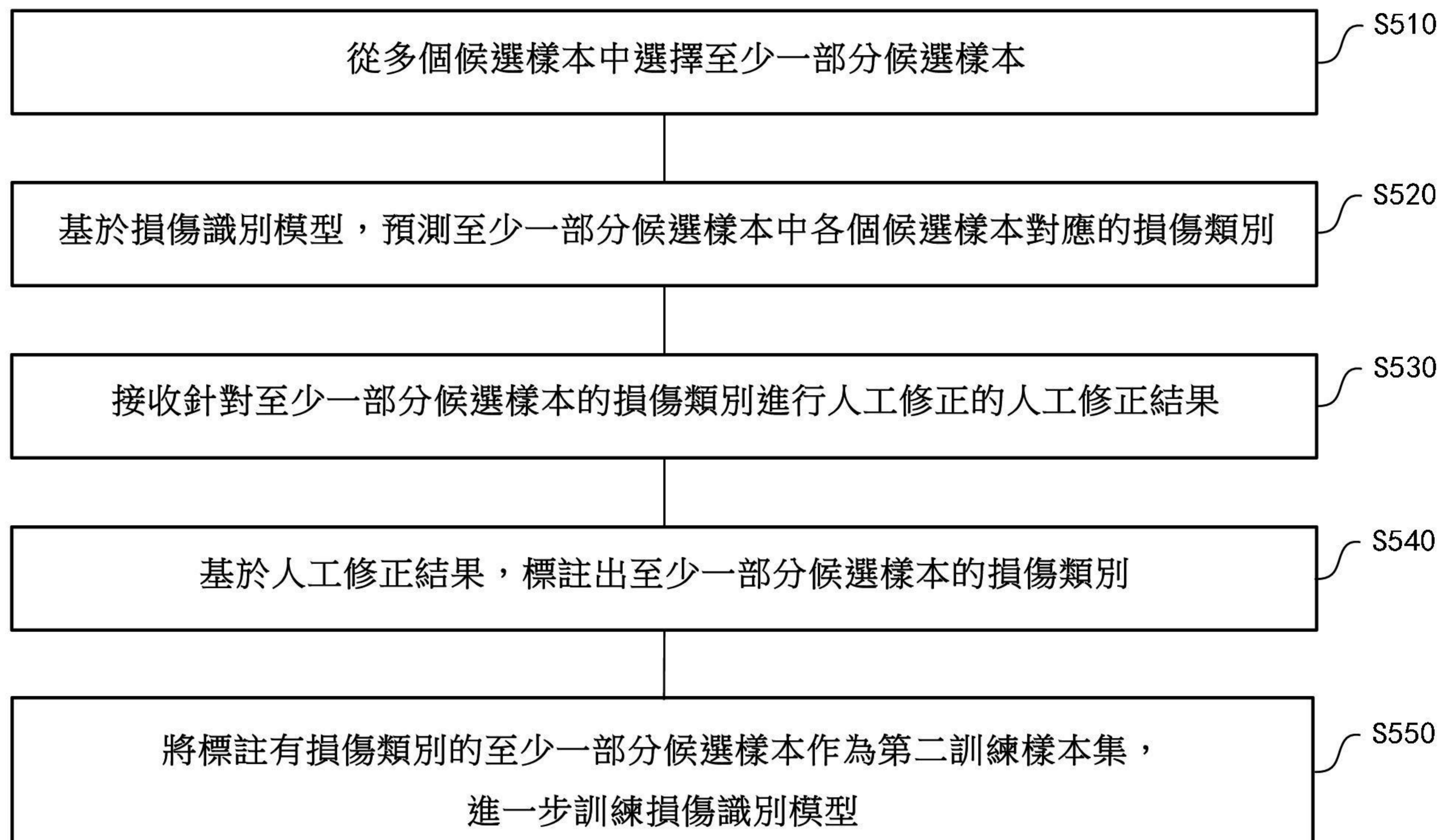
【圖 2】



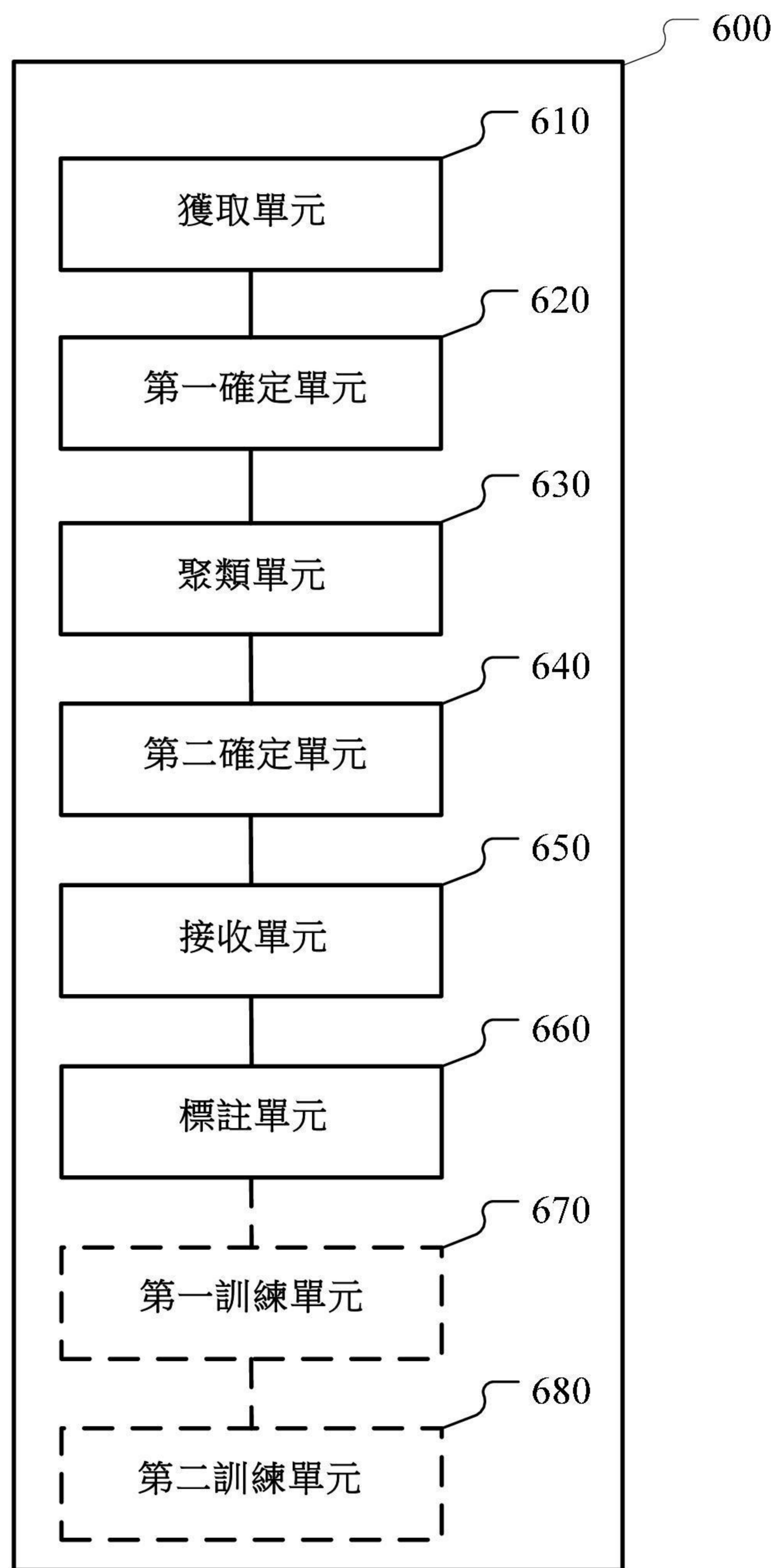
【圖 3】



【圖 4】



【圖 5】



【圖 6】