I800767

(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 I800767 B**

**(45)公告日： 中華民國 112 (2023) 年 05 月 01 日**

(21)申請案號：109139875 (22)申請日： 中華民國 109 (2020) 年 11 月 16 日

(51)Int. Cl.： *G06F11/07 (2006.01)* *G06F11/30 (2006.01)*
*G01N21/88 (2006.01)* *G06T1/40 (2006.01)*

(71)申請人：財團法人工業技術研究院 (中華民國) INDUSTRIAL TECHNOLOGY RESEARCH INSTITUTE (TW)
新竹縣竹東鎮中興路 4 段 195 號

(72)發明人：趙怡翔 CHAO, YI-HSIANG (TW) ；謝志宏 HSIEH, CHIH-HUNG (TW) ；石明于 SHIH, MING-YU (TW)

(74)代理人：洪澄文

(56)參考文獻：

| | | | |
|---|---|---|---|
| TW | 202001681A | TW | 202024612A |
| TW | 202028849A | CN | 110097103A |
| US | 2018/0336471A1 | | |

審查人員：林育弘

申請專利範圍項數：13 項 圖式數：3 共 27 頁

(54)名稱

具有生成對抗網路架構之異常偵測裝置和異常偵測方法

(57)摘要

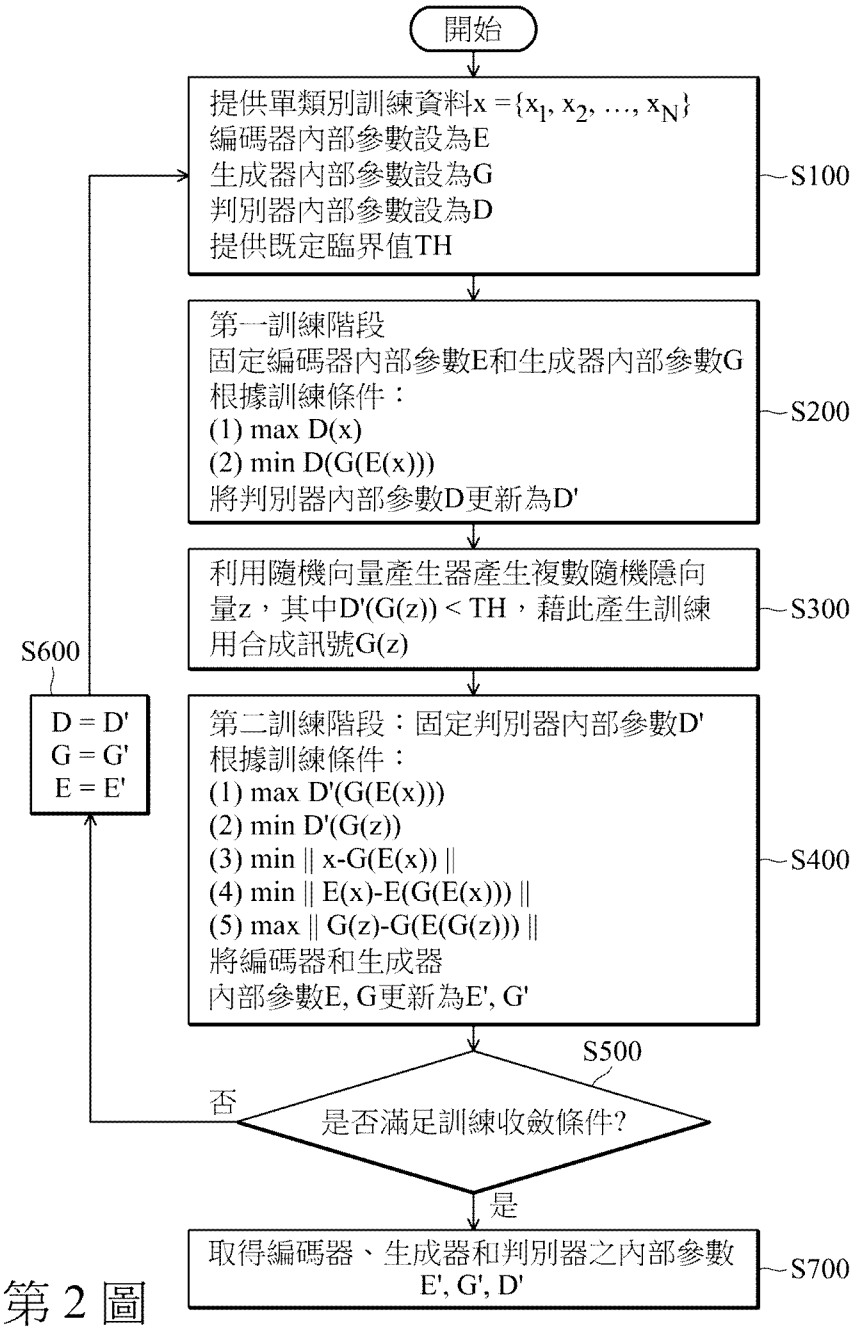
一種具有生成對抗網路架構之異常偵測裝置，其利用一組複數正常訊號所構成之單類別訓練資料進行異常偵測模型訓練。異常偵測裝置包括編碼器、生成器、判別器以及隨機向量產生器，而異常偵測模型訓練包括依序將隨機向量產生器所產生的隨機隱向量輸入至生成器以生成與正常訊號相同維度之複數合成訊號，將合成訊號依序輸入至判別器以產生對應之複數判別值，並且判別值小於一既定臨界值時，並將對應之上述合成訊號選擇為訓練用合成異常訊號，其中係利用上述正常訊號做為無異常訊號。

An anomaly detection device with a generative adversarial network architecture, which uses one-class training data composed of multiple normal signals to train an anomaly detection model. The anomaly detection device includes an encoder, a generator, a discriminator, and a random vector generator. In the training phase of anomaly detection model, the random latent vectors generated by the random vector generator are sequentially inputted to a generator to generate the synthesized signals with the same dimension as the normal signals. The synthesized signals are sequentially inputted to a discriminator to generate the corresponding discriminant values. When the corresponding discriminant values are less than a predetermined threshold value, the corresponding synthesized signals are selected as the training synthesized anomalous signals, and the normal signals are opposite to anomalous signals.

指定代表圖：

符號簡單說明：

S100~S700:步驟




第 2 圖

專利案號：109139875

公告本

I800767

# 【發明摘要】

【中文發明名稱】具有生成對抗網路架構之異常偵測裝置和異常偵測方法

【英文發明名稱】ANOMALY DETECTION APPARATUS AND ANOMALY DETECTION METHOD BASED ON GENERATIVE ADVERSARIAL NETWORKS

【中文】

一種具有生成對抗網路架構之異常偵測裝置，其利用一組複數正常訊號所構成之單類別訓練資料進行異常偵測模型訓練。異常偵測裝置包括編碼器、生成器、判別器以及隨機向量產生器，而異常偵測模型訓練包括依序將隨機向量產生器所產生的隨機隱向量輸入至生成器以生成與正常訊號相同維度之複數合成訊號，將合成訊號依序輸入至判別器以產生對應之複數判別值，並且判別值小於一既定臨界值時，並將對應之上述合成訊號選擇為訓練用合成異常訊號，其中係利用上述正常訊號做為無異常訊號。

【英文】

An anomaly detection device with a generative adversarial network architecture, which uses one-class training data composed of multiple normal signals to train an anomaly detection model. The anomaly detection device includes an encoder, a generator, a discriminator, and a random vector

generator. In the training phase of anomaly detection model, the random latent vectors generated by the random vector generator are sequentially inputted to a generator to generate the synthesized signals with the same dimension as the normal signals. The synthesized signals are sequentially inputted to a discriminator to generate the corresponding discriminant values. When the corresponding discriminant values are less than a predetermined threshold value, the corresponding synthesized signals are selected as the training synthesized anomalous signals, and the normal signals are opposite to anomalous signals.

【指定代表圖】　　第2圖

【代表圖之符號簡單說明】

S100~S700：步驟。

【特徵化學式】

無。

# 【發明說明書】

【中文發明名稱】　具有生成對抗網路架構之異常偵測裝置和異常偵測方法

【英文發明名稱】ANOMALY DETECTION APPARATUS AND ANOMALY DETECTION METHOD BASED ON GENERATIVE ADVERSARIAL NETWORKS

【技術領域】

【0001】本發明是關於一種具有排他學習能力之生成對抗網路架構為基礎的異常偵測裝置和異常偵測方法。

【先前技術】

【0002】在產業應用人工智慧進行瑕疵檢測的導入初期，可能會遇到瑕疵與正常影像數量極不平均，甚至僅有正常樣本，瑕疵樣本的數量不足。一般而言，瑕疵樣態多變不固定，要窮舉所有瑕疵樣本進行訓練變得困難，因此，難以使用監督式學習方式建立準確率高的瑕疵偵測模型。

【0003】因此，如何由正常樣本影像，無需瑕疵樣本影像之情況下，也能建立瑕疵偵測模型，為本領域可精進的方向之一。

【發明內容】

【0004】為了達到上述的想法，本揭露內容之一態樣提供了

一種具有生成對抗網路架構之異常偵測裝置，其利用一組複數正常訊號所構成之單類別訓練資料，進行異常偵測模型訓練。其包括：一編碼器，用以將其輸入之訊號編碼為向量輸出；一生成器，連接至上述編碼器，其用以將其輸入之向量生成一與上述正常訊號相同維度之重建訊號；一判別器，連接至上述生成器，其用以判斷其輸入之訊號為真實或虛假而輸出一判別值；以及一隨機向量產生器，用以產生複數隨機隱向量。其中，上述異常偵測模型訓練包括一第一訓練階段和一第二訓練階段，係依序將上述隨機隱向量輸入至上述生成器生成與上述正常訊號相同維度之複數合成訊號，上述合成訊號係依序輸入至上述判別器，產生對應之複數判別值，並且上述判別值小於一既定臨界值時，將對應之上述合成訊號選擇為訓練用合成異常訊號，其中係利用上述正常訊號做為無異常訊號。

【0005】 為了達到上述的想法，本揭露內容之一態樣提供了一種異常偵測方法，其適用於一包含處理器和記憶體之系統，上述記憶體包含可以執行於上述處理器之複數指令，並且使得上述處理器組態為一可以實現異常偵測方法之生成對抗網路，其上述生成對抗網路之訓練方法包括下列步驟：提供一組複數正常訊號所構成之單類別訓練資料；在第一訓練階段中，固定上述生成對抗網路中之編碼器和生成器之內部參數後，進行其中判別器之訓練，並且依序將上述單類別訓練資料之上述正常訊號輸入至上述編碼器，產生對應之複數第一隱向量，再將對應之上述第一隱向量輸入至上述生成器，用以產生對應之複數第一重建訊號；在第一訓練階段中，依序

將對應之上述第一重建訊號輸入至上述判別器，以其對應之判別值愈小為目標，對於上述判別器內部參數進行調整，並且依序將上述單類別訓練資料之上述正常訊號輸入至上述判別器，以其對應之判別值愈大為目標，對於上述判別器內部參數進行調整；以及產生複數隨機隱向量輸入至上述生成器生成與上述正常訊號相同維度之複數合成訊號，上述合成訊號係依序輸入至上述判別器，產生對應之複數判別值，並且上述判別值小於一既定臨界值時，將對應之上述合成訊號選擇為訓練用合成異常訊號。

【0006】 於本揭露內之態樣中，上述既定臨界值係使得上述訓練用合成異常訊號相異於上述正常訊號之方式所設定。上述正常訊號可以是自動光學檢測系統中利用光學儀器所取得之待測物的表面狀態影像。

【圖式簡單說明】

【0007】

第1圖表示本發明實施例中用以實現異常偵測方法之生成對抗網路的方塊圖；

第2圖表示本發明實施例中用以實現異常偵測方法之生成對抗網路的訓練流程圖；以及

第3圖表示用以實現本發明實施例之異常偵測裝置和異常偵測方法的電腦系統方塊圖。

【實施方式】

【0008】 以下說明係為完成發明的較佳實現方式，其目的在於描述本發明的基本精神，但並不用以限定本發明。實際的發明內容必須參考之後的申請專利範圍。

【0009】 必須了解的是，使用於本說明書中的”包含”、”包括”等詞，係用以表示存在特定的技術特徵、數值、方法步驟、作業處理、元件以及/或組件，但並不排除可加上更多的技術特徵、數值、方法步驟、作業處理、元件、組件，或以上的任意組合。

【0010】 於申請專利中使用如”第一”、"第二"、"第三"等詞係用來修飾申請專利中的元件，並非用來表示之間具有優先權順序，先行關係，或者是一個元件先於另一個元件，或者是執行方法步驟時的時間先後順序，僅用來區別具有相同名字的元件。

【0011】 第1圖表示本發明實施例中用以實現異常偵測方法之生成對抗網路100的方塊圖。如第1圖所示，本實施例之生成對抗網路100中包括一編碼器10、一生成器20、一判別器30、一隨機向量產生器40及一訓練邏輯50。生成對抗網路100可以利用例如筆電、平板、手機或其他具有運算能力的電子裝置加以實現。以一般電腦系統為例進行說明。一般電腦系統至少包含一處理器及一記憶體，其中記憶體可以用來儲存複數指令所構成的程式碼，其中程式碼可以載入至處理器中加以執行，用以實現生成對抗網路100中的編碼器10、生成器20、判別器30、隨機向量產生器40和訓練邏輯50。其中，記憶體可由唯讀記憶體、快閃記憶體、軟碟、硬碟、光

碟、隨身碟、磁帶、可由網路存取之資料庫或熟悉此技藝者可輕易思及具有相同功能之儲存媒體加以實現。

【0012】 在實施例中，編碼器10、生成器20、判別器30、隨機向量產生器40及訓練邏輯50也可以各自或分別由體積電路加以實現，例如微控制器(micro controller)、微處理器(microprocessor)、數位訊號處理器(Digital Signal Processor，DSP )、現場可程式化邏輯閘陣列(Field Programmable Gate Array，FPGA)、特殊應用積體電路(Application Specific Integrated Circuit，ASIC)或一邏輯電路。於一實施例中，編碼器10、生成器20、判別器30、隨機向量產生器40及訓練邏輯50也可以各自以軟體或韌體加以實現。

【0013】 在實施例中，做為單類別訓練資料的正常訊號$x = \{x_1, x_2, \ldots, x_N\}$，其中N為一正整數，表示做為訓練用之正常訊號的樣本數。本實施例之正常訊號x可以是二維影像、音訊、三維影像或其他可被編碼的資料。以二維影像為例，正常訊號x可以是布料紋理表面之影像資料、或者是一面板表面之影像資料、或者是自動光學檢測系統中利用光學儀器所取得之待測物的表面狀態影像。其中二維影像的正常訊號x為了處理方便，可以統一調校為具有相同維度的影像資料，例如32*32像素大小的二維影像。

【0014】 在實施例中，編碼器10、生成器20和判別器30可利用類神經網路模型加以實現，類神經網路模型例如是卷積神經網路(Convolutional Neural Network, CNN)、循環神經網路

(Recurrent Neural Network, RNN)或其他類神經網路模型等。

【0015】 於實施例中，編碼器(Encoder)10用來將輸入之訊號編碼成對應的隱向量(latent vector)，為方便表示，以符號E代表編碼器10之類神經網路內部參數，以函數E()表示編碼器的輸出。

【0016】 於實施例中，生成器(Generator)20用以將輸入的隱向量生成為對應的重建訊號，為方便表示，以符號G代表生成器20之類神經網路內部參數，以函數G()表示生成器的輸出。在實施例中，生成器20所產生的重建訊號會與正常訊號具有相同維度。

【0017】 於實施例中，判別器(Discriminator)30用以將輸入的訊號(例如正常訊號x、重建訊號或由後述隨機隱向量所生成的合成訊號)經過計算後得出一介於0到1之間的實數值(判別值)。其中此判別值愈大(接近1)代表其判斷輸入的訊號為真實訊號的可能性愈高，判別值愈小(接近0)代表其判斷輸入的訊號判別為虛假訊號的可能性愈高。為方便表示，以符號D代表判別器30之類神經網路內部參數，以函數D()表示判別器的輸出。於一般生成對抗網路中，生成器20和判別器30之間為對抗關係，亦即訓練的目的係使得判別器30對於真實訊號和虛假訊號可以分別輸出接近1和接近0的判別值，而同時生成器20的輸出在判別器30的判斷下則會接近真實訊號。

【0018】 於實施例中，隨機向量產生器40透過隨機亂數產生對應之複數個隨機隱向量z。這些隨機隱向量z在本實施例中則是用來產生不同於正常訊號x的異常訊號。處理上是依序將隨機隱向量z

輸入至生成器20生成與正常訊號x相同維度的複數個合成訊號G(z)，再將合成訊號G(z)依序輸入至判別器30，產生對應之複數判別值D(G(z))，並且當判別值D(G(z))小於一既定臨界值TH時，將對應的合成訊號選擇為訓練用合成異常訊號。亦即利用判別器30來評估對應之合成訊號G(z)是否為正常類別訊號，再回推隨機隱向量z是否屬於E(x)分佈的可能性，目標是選擇出不在E(x)分佈的z，此即為本實施例中的排他性(exclusive)觀念。藉此，雖然原始訓練資料中為單類別訓練資料(僅含正常訊號)，但是可以利用正常訊號x做為無異常訊號，並且利用訓練用合成異常訊號G(z)做為異常訊號，進行異常偵測模型訓練。

【0019】 如第1圖所示，訓練邏輯50可以根據正常訊號x、編碼器10、生成器20和判別器30的輸出，依據既定訓練條件進行編碼器10、生成器20、判別器30內部參數E、G、D的調校。當訓練邏輯50判斷滿足一訓練收斂條件時，則輸出此時的編碼器10、生成器20、判別器30內部參數E、G、D。其中訓練邏輯50則如第2圖所示，以下配合圖式詳細說明。於一實施例中，收斂條件可以事先定義，例如為執行迴圈100次。

【0020】 第2圖表示本發明實施例中用以實現異常偵測方法之生成對抗網路100的訓練流程圖。

【0021】 如第2圖所示，在步驟S100中，為生成對抗網路100的初始狀態。單類別訓練資料的正常訊號$x = \{x_1, x_2, \ldots, x_N\}$，例如可以是N個正常的待測物表面影像資料，影像資料維度為32*32

個像素。此時，編碼器10內部參數設為E，生成器20內部參數設為G，判別器30內部參數設為D。提供既定臨界值TH係使得訓練用合成異常訊號G(z)相異於上述正常訊號x之方式所設定，其介於0與1之間。例如可以設為0.5或0.7等。

【0022】 在步驟S200中，進行第一訓練階段，固定編碼器10的內部參數E和生成器20的內部參數G，對於判別器30的內部參數D進行調校。調校方式是根據訓練條件(1)和(2)：

$$\max\ D(x) \qquad (1)$$

$$\min\ D(G(E(x))) \qquad (2)$$

【0023】 根據訓練條件(1)，是依序將單類別訓練資料之正常訊號x輸入至判別器30，並且以其對應之判別值D(x)愈大為目標；另一方面，根據訓練條件(2)，依序將單類別訓練資料之正常訊號x輸入至編碼器10，產生對應之複數第一隱向量E(x)，再將對應之第一隱向量E(x)輸入至生成器20，用來產生對應的複數第一重建訊號G(E(x))，再將第一重建訊號G(E(x))輸入至判別器30，以其對應之判別值D(G(E(x)))愈小為目標。根據訓練條件(1)和(2)，則可以對於判別器30的內部參數D進行調校，更新為內部參數D'。

【0024】 在步驟S300中，利用隨機向量產生器40產生複數個隨機隱向量z，其中$D'(G(z)) < TH$，如前所述，藉此產生訓練用合成異常訊號G(z)。在本實施例之實施上，訓練用合成異常訊號G(z)的數量可以小於N(亦即正常訊號樣本數)。

【0025】 在步驟S400中，進行第二訓練階段，固定判別器30的內部參數D'，對於編碼器10的內部參數E和生成器20的內部參數

G進行調校。調校方式是根據訓練條件(3)~(7)：

$$\max\ D'(G(E(x))) \qquad (3)$$

$$\min\ D'(G(z)) \qquad (4)$$

$$\min \| x-G(E(x)) \| \qquad (5)$$

$$\min \| E(x)-E(G(E(x))) \| \qquad (6)$$

$$\max \| G(z)-G(E(G(z))) \| \qquad (7)$$

【0026】 根據訓練條件(3)，第一重建訊號G(E(x))輸入至判別器30，調校內部參數E和內部參數G的方式，使得其對應之判別值D'(G(E(x)))愈大為目標。如前所述，其與訓練條件(2)為對抗關係。如前所述，訓練用合成異常訊號G(z)係做為不同於正常訊號x之異常訊號，因此根據訓練條件(4)，訓練用合成異常訊號G(z)輸入至判別器30，並且調校內部參數E和內部參數G的方式，以其對應之判別值D'(G(z))愈小為目標。

【0027】 另外，根據訓練條件(5)、(6)、(7)，則是分別以第一誤差值愈小、第二誤差值愈小、第三誤差值愈大為目標，調校內部參數E和內部參數G。其中第一誤差值係計算上述正常訊號x與上述第一重建訊號G(E(x))間之距離值；第二誤差值係計算第一隱向量E(x)與第二隱向量E(G(E(x)))間之距離值；第三誤差值係計算訓練用合成異常訊號G(z)與第二重建訊號G(E(G(z)))間之距離值。必須說明的是，上述訓練目標為最大化(max)或最小化(min)係為相對之概念，於實作時可以依據實施上的一致性，加入負號或數學運算以便調整實作方式，並不違背本發明實施例之精神。

【0028】 必須特別說明的是，根據訓練條件(7)，本實施例中，訓練用合成異常訊號G(z)(代表非正常訊號)以及其經過編碼器10

和生成器20所產生的第二重建訊號G(E(G(z)))之間的重建誤差最大化，藉此達到本實施例中排他性生成對抗網路(exclusive generative adversarial network)架構的特性。在完成第二訓練階段後，則將編碼器10的內部參數E和生成器20的內部參數G，更新為E’和G’。

【0029】 在步驟S500中，則是判斷是否滿足訓練收斂條件。例如可以預先設定訓練執行迴圈為100次。

【0030】 在步驟S600中，當尚未滿足訓練收斂條件時，經由步驟S600回到步驟S100。在步驟S600中，將編碼器10、生成器20和判別器30的目前內部參數E’、G’、D’設定給編碼器10、生成器20和判別器30的內部參數E、G、D。

【0031】 在步驟S700中，當滿足訓練收斂條件時，則可以取得此時之編碼器10、生成器20和判別器30的內部參數E、G、D。藉此，即可以取得本實施例之排他性生成對抗網路，用以實現異常偵測裝置和異常偵測方法。

【0032】 第3圖表示用以實現本發明實施例之異常偵測裝置和異常偵測方法的電腦系統130的方塊圖。此電腦系統130架構可實施於桌上型電腦、筆記型電腦或其他具備運算能力的電子裝置。處理單元310可使用多種方式實施，例如以專用硬體電路或通用硬體(例如，單一處理器、具平行處理能力的多處理器、圖形處理器或其他具運算能力的處理器)，並且在執行程式碼或軟體時，提供前述的異常偵測功能。此系統架構另包含記憶體350用以儲存執行過程中需要的資料，例如，變數、資料表等，以及儲存裝置340，用以儲存各式各樣的電子檔案，例如，網頁、文件、音訊檔、視訊檔等。

此系統架構另包含通訊介面360，讓處理單元310可藉以跟其他電子裝置進行溝通。通訊介面360可以是區域網路通訊模組或無線區域網路通訊模組。輸入裝置330可包含鍵盤、滑鼠、觸控面板等。使用者可按壓鍵盤上的硬鍵來輸入字元，藉由操作滑鼠來控制鼠標，或者是在觸控面板製造手勢來控制執行中的應用程式。手勢可包含單擊、雙擊、單指拖曳、多指拖曳等，但不限定於此。顯示單元320可包含顯示面板(例如，薄膜液晶顯示面板、有機發光二極體面板或其他具顯示能力的面板)，用以顯示輸入的字元、數字、符號、拖曳鼠標的移動軌跡、繪製的圖案或應用程式所提供的畫面，提供給使用者觀看。

【0033】 以下以一實際範例說明本發明實施例之內容。下述表一為未採用本實施例之排他性生成對抗網路的布料紋理異常偵測情形(分成自動預測與人工標註的情形)，表二為採用採用本實施例之排他性生成對抗網路的布料紋理異常偵測情形(分成自動預測與人工標註的情形)，兩者為對照組，假設實驗所採用的訓練正常影像(真實資料x)為1000張布料影像，而測試正常影像有9730張，測試異常影像有3336張。

| 總正確率(%) | 總錯誤率(%) | 未採用排他性生成對抗網路(自動預測) | | | |
|---|---|---|---|---|---|
| 93.98 | 6.02 | 異常 | 正常 | 測試影像張數 | 召回率(recall)/精確率(precision)(%) |
| 人工標註 | 異常 | 3336 | 0 | 3336 | 100/80.91 |
| | 正常 | 787 | 8943 | 9730 | FNR/TNR(%) |
| 加總 | | 4123 | 8943 | 13066 | 0/91.91 |

表一

| 總正確率(%) | 總錯誤率(%) | 採用排他性生成對抗網路(自動預測) | | | |
|---|---|---|---|---|---|
| 96.14 | 3.86 | 異常 | 正常 | 測試影像張數 | 召回率(recall)/精確率(precision) (%) |
| 人工標註 | 異常 | 3336 | 0 | 3336 | 100/86.88 |
| | 正常 | 504 | 9226 | 9730 | FNR/TNR(%) |
| 加總 | | 3840 | 9226 | 13066 | 0/94.82 |

表二

【0034】 其中，以表一為例，召回率(recall)是將自動預測異常正確的張數(3336)除以人工標註異常的總張數(3336)而得。精確率是將自動預測異常正確的張數(3336)除以自動預測異常的總張數(4123)而得。

【0035】 於一實施例中，偽陰率FNR(false negative rate，FNR)是將自動預測異常錯誤的張數(0)除以人工標註異常的總張數(3336) 而得，亦即1 - 召回率。真陰率TNR(true negative rate，TNR)是將自動預測正常正確的張數(8943)除以人工標註正常的總張數(9730)而得。

【0036】 於一實施例中，總正確率是將自動預測正確的總張數(3336+8943)除以總測試影像數(13066)而得。總錯誤率為1 - 總正確率。

【0037】 表二的計算方式與表一相同，故此處不贅述之。由表一及表二可看出當召回率(recall)為100%的相同情況下，未採用本實施例之排他性生成對抗網路的精確率為80.91%，採用本實施

例之排他性生成對抗網路的精確率為86.88%，因此採用本實施例之排他性生成對抗網路提升了在判斷資料正常或異常時的精準度。

【0038】 在一些領域中，例如半導體領域的相關產品取像做異常偵測時，由於半導體領域的良率較高，因此要找到正常的產品取像作為訓練資料較容易，但不容易找到異常的產品取像作為訓練資料。本發明實施例所示之異常偵測裝置及異常偵測方法，可藉由隨機亂數產生之隨機隱向量經生成器生成與正常訊號相同維度之合成訊號。此等合成訊號係依序輸入至上述判別器，產生對應之複數判別值，當判別值小於一既定臨界值時，將對應之合成訊號選擇為訓練用合成異常訊號。利用正常訊號做為無異常訊號，並且利用訓練用合成異常訊號做為異常訊號，則可進行異常偵測模型訓練。換言之，本發明實施例所示之異常偵測裝置及異常偵測方法，可以收集正常真實資料以訓練模型，也可以生成異常資料以訓練模型，達到產生用以訓練資料處理裝置的異常的訓練資料之目的，藉此建立更能精準判斷輸入資料為正常或異常資料的異常偵測模型。

【0039】 本揭露內容之一態樣提供了一種具有生成對抗網路架構之異常偵測裝置，其利用一組複數正常訊號所構成之單類別訓練資料，進行異常偵測模型訓練。其包括：一編碼器，用以將其輸入之訊號編碼為向量輸出；一生成器，連接至上述編碼器，其用以將其輸入之向量生成一與上述正常訊號相同維度之重建訊號；一判別器，連接至上述生成器，其用以判斷其輸入之訊號為真實或虛假而輸出一判別值；以及一隨機向量產生器，用以產生複數隨機隱向量。其中，上述異常偵測模型訓練包括一第一訓練階段和一第二訓練階段，係依序將上述隨機隱向量輸入至上述生成器生成與上述正

常訊號相同維度之複數合成訊號，上述合成訊號係依序輸入至上述判別器，產生對應之複數判別值，並且上述判別值小於一既定臨界值時，將對應之上述合成訊號選擇為訓練用合成異常訊號。

【0040】 根據本揭露內之一樣態，編碼器、生成器、判別器，係分別由類神經網路所構成。根據本揭露內之一樣態，判別值愈大代表其輸入訊號判別為真實，愈小代表其輸入訊號判別為虛假。根據本揭露內之一樣態，異常偵測模型訓練利用訓練用合成異常訊號做為異常訊號，並且利用正常訊號做為無異常訊號，以進行異常偵測模型訓練。根據本揭露內之一樣態，在第一訓練階段中，係將編碼器和生成器之內部參數固定後，進行判別器之訓練，並且依序將單類別訓練資料之正常訊號輸入至編碼器，產生對應之複數第一隱向量，再將對應之第一隱向量輸入至生成器，用以產生對應之複數第一重建訊號。根據本揭露內之一樣態，在第一訓練階段中，依序將對應之第一重建訊號輸入至判別器，以其對應之判別值愈小為目標，對於判別器內部參數進行調整，並且依序將單類別訓練資料之正常訊號輸入至判別器，以其對應之判別值愈大為目標，對於判別器內部參數進行調整。

【0041】 根據本揭露內之一樣態，在第二訓練階段中，係將判別器之內部參數固定後，進行編碼器和生成器之訓練，並且依序將第一重建訊號輸入至編碼器，產生複數第二隱向量，再將訓練用合成異常訊號輸入至編碼器，產生複數第三隱向量，再將第三隱向量輸入至上述生成器，產生複數第二重建訊號。根據本揭露內之一樣態，依序計算正常訊號與第一重建訊號間之第一誤差值，以及計算第一隱向量與第二隱向量間之第二誤差值，以及計算訓練用合成

異常訊號與第二重建訊號間之第三誤差值。根據本揭露內之一樣態，在第二訓練階段中，係依序將對應之第一重建訊號輸入至判別器，以其對應之判別值愈大為目標進行編碼器和生成器內部之參數調整，並且依序將訓練用合成異常訊號輸入至判別器，以其對應之判別值愈小為目標進行編碼器和生成器內部之參數調整，並且以第一誤差值和第二誤差值愈小為目標進行編碼器和生成器內部之參數調整，並且以第三誤差值愈大為目標進行編碼器和生成器內部之參數調整。

【0042】 根據本揭露內之一樣態，正常訊號可以是一布料紋理表面之影像資料、或者是一面板表面之影像資料、或者是自動光學檢測系統中利用光學儀器所取得之待測物的表面狀態影像。

【0043】 本揭露內容之一態樣提供了一種異常偵測方法，其適用於一包含處理器和記憶體之系統，上述記憶體包含可以執行於上述處理器之複數指令，並且使得上述處理器組態為一可以實現異常偵測方法之生成對抗網路，其上述生成對抗網路之訓練方法包括下列步驟：提供一組複數正常訊號所構成之單類別訓練資料；在第一訓練階段中，固定上述生成對抗網路中之編碼器和生成器之內部參數後，進行上述生成對抗網路中之判別器之訓練，並且依序將上述單類別訓練資料之上述正常訊號輸入至上述編碼器，產生對應之複數第一隱向量，再將對應之上述第一隱向量輸入至上述生成器，用以產生對應之複數第一重建訊號；在上述第一訓練階段中，依序將對應之上述第一重建訊號輸入至上述判別器，以其對應之判別值愈小為目標，對於上述判別器內部參數進行調整，並且依序將上述單類別訓練資料之上述正常訊號輸入至上述判別器，以其對應之判

別值愈大為目標，對於上述判別器內部參數進行調整；以及產生複數隨機隱向量輸入至上述生成器生成與上述正常訊號相同維度之複數合成訊號，上述合成訊號係依序輸入至上述判別器，產生對應之複數判別值，並且上述判別值小於一既定臨界值時，將對應之上述合成訊號選擇為訓練用合成異常訊號。

【0044】 根據本揭露內之一樣態，異常偵測方法更包括於第二訓練階段中，固定判別器之內部參數後，進行編碼器和生成器之訓練，並且依序將第一重建訊號輸入至編碼器，產生複數第二隱向量，再將訓練用合成異常訊號輸入至編碼器，產生複數第三隱向量，再將第三隱向量輸入至生成器，產生複數第二重建訊號，其中，依序計算正常訊號與第一重建訊號間之第一誤差值，以及計算第一隱向量與第二隱向量間之第二誤差值，以及計算訓練用合成異常訊號與第二重建訊號間之第三誤差值。

【0045】 根據本揭露內之一樣態，異常偵測方法更包括在第二訓練階段中，係依序將對應之第一重建訊號輸入至判別器，以其對應之判別值愈大為目標進行編碼器和生成器內部之參數調整，並且依序將訓練用合成異常訊號輸入至判別器，以其對應之判別值愈小為目標進行編碼器和生成器內部之參數調整，並且以第一誤差值和第二誤差值愈小為目標進行編碼器和生成器內部之參數調整，並且以第三誤差值愈大為目標進行編碼器和生成器內部之參數調整。

【0046】 根據本揭露內之一樣態，異常偵測方法更包括當未滿足既定訓練收斂條件時，以編碼器、生成器和判別器之目前內部參數，跳回至第一訓練階段；以及當滿足既定訓練收斂條件時，以編碼器、生成器和判別器之目前內部參數設定生成對抗網路。

【符號說明】

【0047】

100:生成對抗網路

10:編碼器

20:生成器

30:判別器

40:隨機向量產生器

50:訓練邏輯

S100, S200, S300, S400, S500, S600, S700:步驟

130~系統

310:處理單元

320:顯示單元

330:輸入裝置

340:儲存裝置

350:記憶體

360:通訊介面

第 109139875 號 修正日期:111.03.30 修正本

# 【發明申請專利範圍】

【請求項1】 一種具有生成對抗網路架構之異常偵測裝置，其利用一組複數正常訊號所構成之單類別訓練資料進行異常偵測模型訓練，其包括：

一編碼器，用以將其輸入之訊號編碼為向量輸出；

一生成器，連接至上述編碼器，其用以將其輸入之向量生成一與上述正常訊號相同維度之重建訊號；

一判別器，連接至上述生成器，其用以判斷其輸入之訊號為真實或虛假而輸出一判別值；以及

一隨機向量產生器，用以產生複數隨機隱向量；

其中，上述異常偵測模型訓練包括一第一訓練階段和一第二訓練階段，係依序將上述隨機隱向量輸入至上述生成器生成與上述正常訊號相同維度之複數合成訊號，上述合成訊號係依序輸入至上述判別器，產生對應之複數判別值，並且上述判別值小於一既定臨界值時，將對應之上述合成訊號選擇為上述第二訓練階段所需使用的訓練用合成異常訊號；

其中在上述第一訓練階段中，係將上述編碼器和上述生成器之內部參數固定後，進行上述判別器之訓練，並且依序將上述單類別訓練資料之上述正常訊號輸入至上述編碼器，產生對應之複數第一隱向量，再將對應之上述第一隱向量輸入至上述生成器，用以產生對應之複數第一重建訊號；

其中在上述第二訓練階段中，係將上述判別器之內部參數固定

後，進行上述編碼器和上述生成器之訓練，並且依序將上述第一重建訊號輸入至上述編碼器，產生複數第二隱向量，再將上述訓練用合成異常訊號輸入至上述編碼器，產生複數第三隱向量，再將上述第三隱向量輸入至上述生成器，產生複數第二重建訊號，其中，依序計算上述正常訊號與上述第一重建訊號間之第一誤差值，以及計算上述第一隱向量與上述第二隱向量間之第二誤差值，以及計算上述訓練用合成異常訊號與上述第二重建訊號間之第三誤差值；以及

其中在上述第二訓練階段中，係依序將對應之上述第一重建訊號輸入至上述判別器，以其對應之判別值愈大為目標進行上述編碼器和上述生成器內部之參數調整，並且依序將訓練用合成異常訊號輸入至上述判別器，以其對應之判別值愈小為目標進行上述編碼器和上述生成器內部之參數調整，並且以上述第一誤差值和上述第二誤差值愈小為目標進行上述編碼器和上述生成器內部之參數調整，並且以上述第三誤差值愈大為目標進行上述編碼器和上述生成器內部之參數調整。

【請求項2】 如請求項1之異常偵測裝置，其中，上述編碼器、上述生成器、上述判別器，係分別由類神經網路所構成。

【請求項3】 如請求項1之異常偵測裝置，其中上述判別值愈大代表其輸入訊號判別為真實，愈小代表其輸入訊號判別為虛假。

【請求項4】 如請求項3之異常偵測裝置，其中上述異常偵測模型訓練利用上述訓練用合成異常訊號做為異常訊號，並且利用上述

正常訊號做為無異常訊號，以進行上述異常偵測模型訓練。

【請求項5】 如請求項4所述之異常偵測裝置，其中，在上述第一訓練階段中，依序將對應之上述第一重建訊號輸入至上述判別器，以其對應之判別值愈小為目標，對於上述判別器內部參數進行調整，並且依序將上述單類別訓練資料之上述正常訊號輸入至上述判別器，以其對應之判別值愈大為目標，對於上述判別器內部參數進行調整。

【請求項6】 如請求項1所述之異常偵測裝置，其中上述正常訊號係為一布料紋理表面之影像資料。

【請求項7】 如請求項1所述之異常偵測裝置，其中上述正常訊號係為一面板表面之影像資料。

【請求項8】 如請求項1所述之異常偵測裝置，其中上述正常訊號係為自動光學檢測系統中利用光學儀器所取得之待測物的表面狀態影像。

【請求項9】 如請求項1所述之異常偵測裝置，其中上述既定臨界值係使得上述訓練用合成異常訊號相異於上述正常訊號之方式所設定。

【請求項10】 一種異常偵測方法，其適用於一包含處理器和記憶體之系統，上述記憶體包含可以執行於上述處理器之複數指令，並且使得上述處理器組態為一可以實現異常偵測方法之生成對抗網路，其上述生成對抗網路之訓練方法包括下列步驟：

提供一組複數正常訊號所構成之單類別訓練資料；

在第一訓練階段中，固定上述生成對抗網路中之編碼器和生成器之內部參數後，進行上述生成對抗網路中之判別器之訓練，並且依序將上述單類別訓練資料之上述正常訊號輸入至上述編碼器，產生對應之複數第一隱向量，再將對應之上述第一隱向量輸入至上述生成器，用以產生對應之複數第一重建訊號；

在上述第一訓練階段中，依序將對應之上述第一重建訊號輸入至上述判別器，以其對應之判別值愈小為目標，對於上述判別器內部參數進行調整，並且依序將上述單類別訓練資料之上述正常訊號輸入至上述判別器，以其對應之判別值愈大為目標，對於上述判別器內部參數進行調整；以及

產生複數隨機隱向量輸入至上述生成器生成與上述正常訊號相同維度之複數合成訊號，上述合成訊號係依序輸入至上述判別器，產生對應之複數判別值，並且上述判別值小於一既定臨界值時，將對應之上述合成訊號選擇為第二訓練階段所需使用的訓練用合成異常訊號；

於上述第二訓練階段中，固定上述判別器之內部參數後，進行上述編碼器和上述生成器之訓練，並且依序將上述第一重建訊號輸入至上述編碼器，產生複數第二隱向量，再將上述訓練用合成異常訊號輸入至上述編碼器，產生複數第三隱向量，再將上述第三隱向量輸入至上述生成器，產生複數第二重建訊號， 其中，依序計算上述正常訊號與上述第一重建訊號間之第一誤差值，以及計算上述第一隱向量與上述第二隱向量間之第二誤差值，以及計

算上述訓練用合成異常訊號與上述第二重建訊號間之第三誤差值；以及

在上述第二訓練階段中，係依序將對應之上述第一重建訊號輸入至上述判別器，以其對應之判別值愈大為目標進行上述編碼器和上述生成器內部之參數調整，並且依序將上述訓練用合成異常訊號輸入至上述判別器，以其對應之判別值愈小為目標進行上述編碼器和上述生成器內部之參數調整，並且以上述第一誤差值和上述第二誤差值愈小為目標進行上述編碼器和上述生成器內部之參數調整，並且以上述第三誤差值愈大為目標進行上述編碼器和上述生成器內部之參數調整。

【請求項11】 如請求項10之異常偵測方法，更包括：

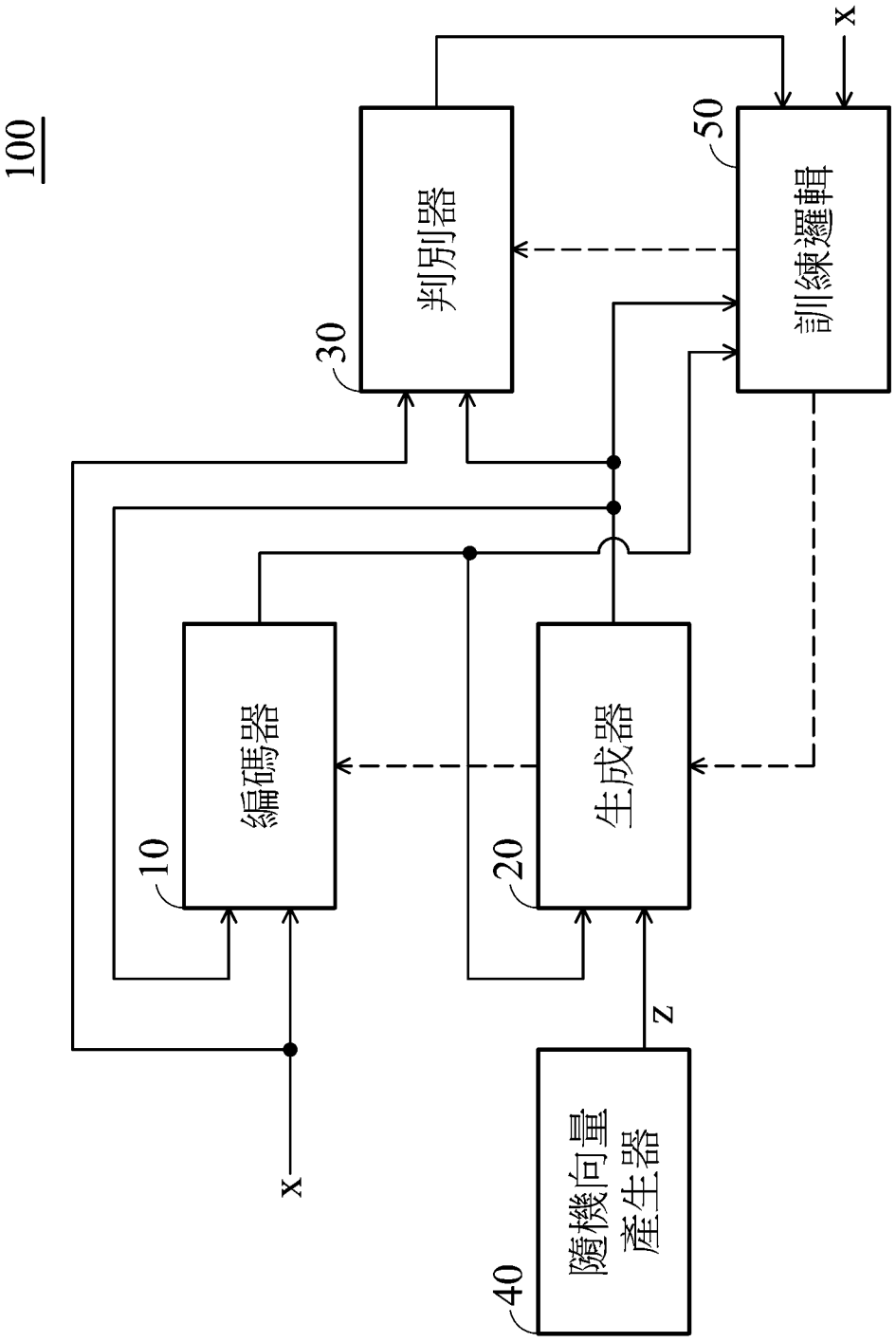
當未滿足既定訓練收斂條件時，以上述編碼器、上述生成器和上述判別器之目前內部參數，跳回至上述第一訓練階段；以及

當滿足既定訓練收斂條件時，以上述編碼器、上述生成器和上述判別器之目前內部參數設定上述生成對抗網路。

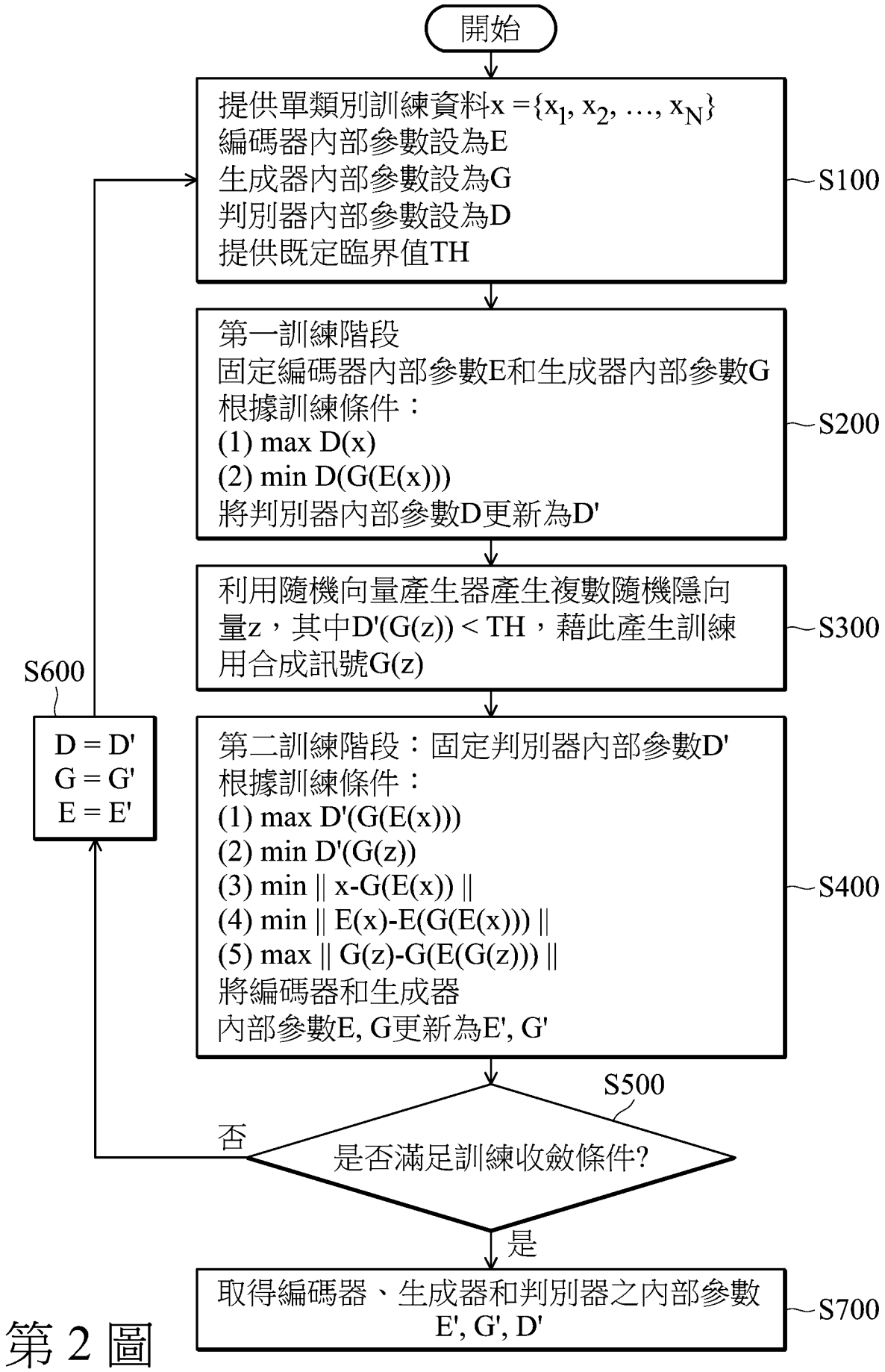
【請求項12】 如請求項10之異常偵測方法，其中，上述編碼器、上述生成器、上述判別器，係分別由類神經網路所構成。

【請求項13】 如請求項10所述之異常偵測方法，其中上述既定臨界值係使得上述訓練用合成異常訊號相異於上述正常訊號之方式所設定。

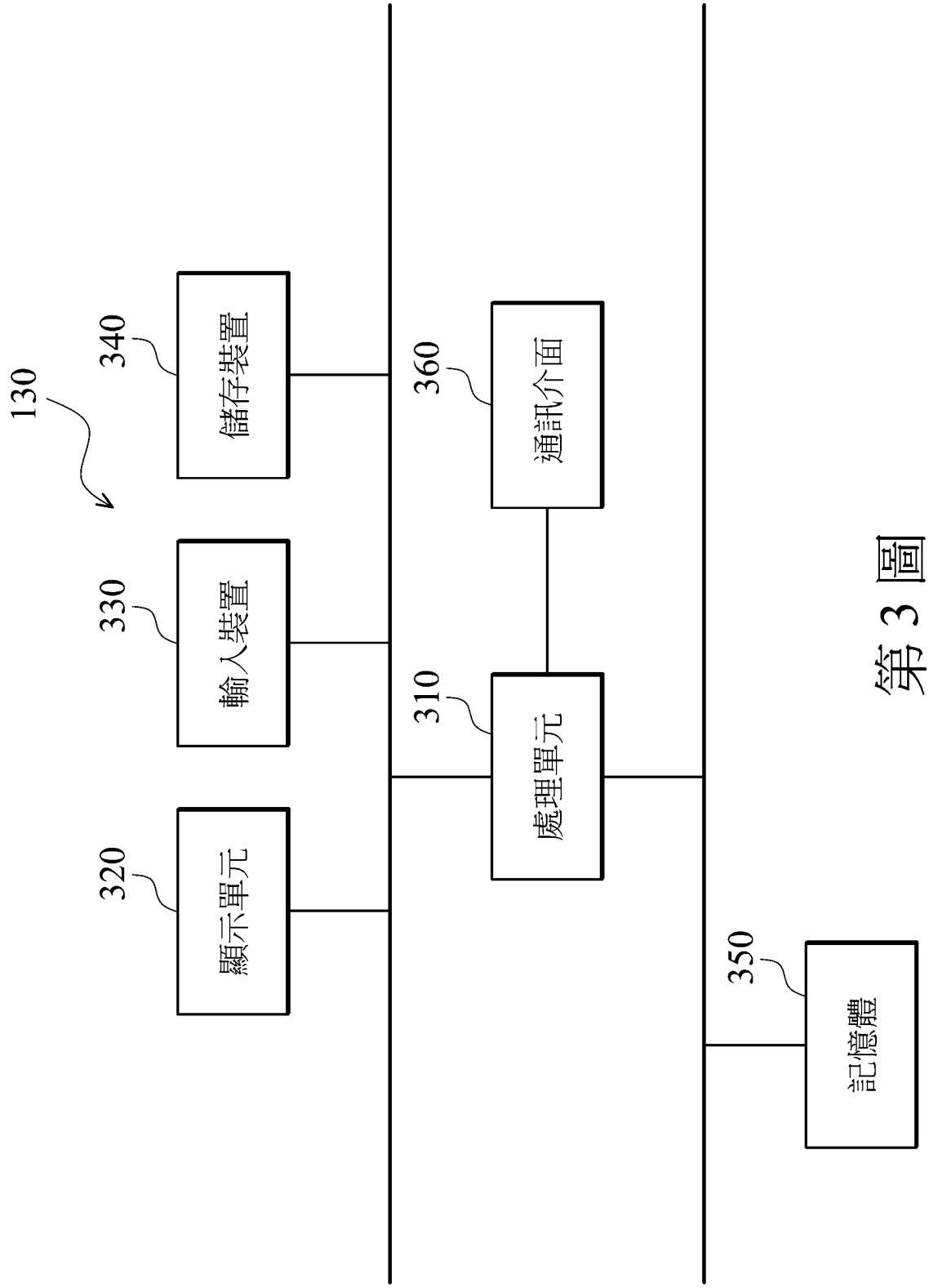
# 【發明圖式】




第 1 圖


開始
提供單類別訓練資料x ={x1, x2, …, xN}
編碼器內部參數設為E
生成器內部參數設為G
判別器內部參數設為D
提供既定臨界值TH
S100
第一訓練階段
固定編碼器內部參數E和生成器內部參數G
根據訓練條件：
(1) max D(x)
(2) min D(G(E(x)))
將判別器內部參數D更新為D'
S200
利用隨機向量產生器產生複數隨機隱向量z，其中D'(G(z)) < TH，藉此產生訓練用合成訊號G(z)
S300
第二訓練階段：固定判別器內部參數D'
根據訓練條件：
(1) max D'(G(E(x)))
(2) min D'(G(z))
(3) min || x-G(E(x)) ||
(4) min || E(x)-E(G(E(x))) ||
(5) max || G(z)-G(E(G(z))) ||
將編碼器和生成器
內部參數E, G更新為E', G'
S400
S500
是否滿足訓練收斂條件?
否
是
S600
D = D'
G = G'
E = E'
取得編碼器、生成器和判別器之內部參數
E', G', D'
S700

第 2 圖




第 3 圖