I759156

(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 I759156 B**

(45)公告日： 中華民國 111 (2022) 年 03 月 21 日

(21)申請案號：110110527 (22)申請日： 中華民國 110 (2021) 年 03 月 24 日

(51)Int. Cl.： *G06T7/33 (2017.01)* *G06T1/40 (2006.01)*
G06K9/46 (2006.01) *G06N3/02 (2006.01)*

(30)優先權：2021/01/19 美國 63/138968

(71)申請人：福邦科技國際股份有限公司 (中華民國) MICROMAX INTERNATIONAL CORP. (TW)
臺北市內湖區民權東路 6 段 160 號 5 樓
國立臺灣科技大學 (中華民國) NATIONAL TAIWAN UNIVERSITY OF SCIENCE AND TECHNOLOGY (TW)
臺北市大安區基隆路 4 段 43 號

(72)發明人：花凱龍 HUA, KAI-LUNG (TW) ；陳永耀 CHEN, YUNG-YAO (TW) ；鍾昕燁 JHONG, SIN-YE (TW) ；陳佑丞 CHEN, YO-CHENG (TW) ；林八林 LIN, BA-LIN (TW) ；林子永 LIN, TZYY-YZONG (TW) ；溫承書 WEN, CHENG-SHU (TW) ；王彥博 WANG, YEN-PO (TW) ；陳俊榮 CHEN, CHUN-JUNG (TW) ；楊東行 YANG, TUNG-SHIN (TW) ；呂文翔 LU, WEN-HSIANG (TW) ；黃祺佳 HUANG, CHYI-JIA (TW)

(74)代理人：高玉駿；楊祺雄

(56)參考文獻：

TW 201917566A TW 202032425A
CN 108734274A

審查人員：潘世光

申請專利範圍項數：7 項 圖式數：4 共 30 頁

(54)名稱

影像物件辨識模型的訓練方法及影像物件辨識模型

(57)摘要

一種影像物件辨識模型的訓練方法，該影像物件辨識模型包含第一、第二及第三深度神經網路以及一連接第一、第二及第三深度神經網路的特徵融合層；將複數組訓練用影像其中的每一組訓練用影像中的可見光影像和熱影像各別對應輸入第一和第二深度神經網路，以對第一和第二深度神經網路進行訓練，且該特徵融合層接受由該第一和第二深度神經網路各別輸出之經過特徵處理的各該可見光影像和經過特徵處理的各該熱影像，並將兩者融合成一融合影像後輸入第三深度神經網路，以對第三深度神經網路進行訓練，而獲得完成訓練的該影像物件辨識模型。

指定代表圖：

符號簡單說明：

S1~S3:步驟

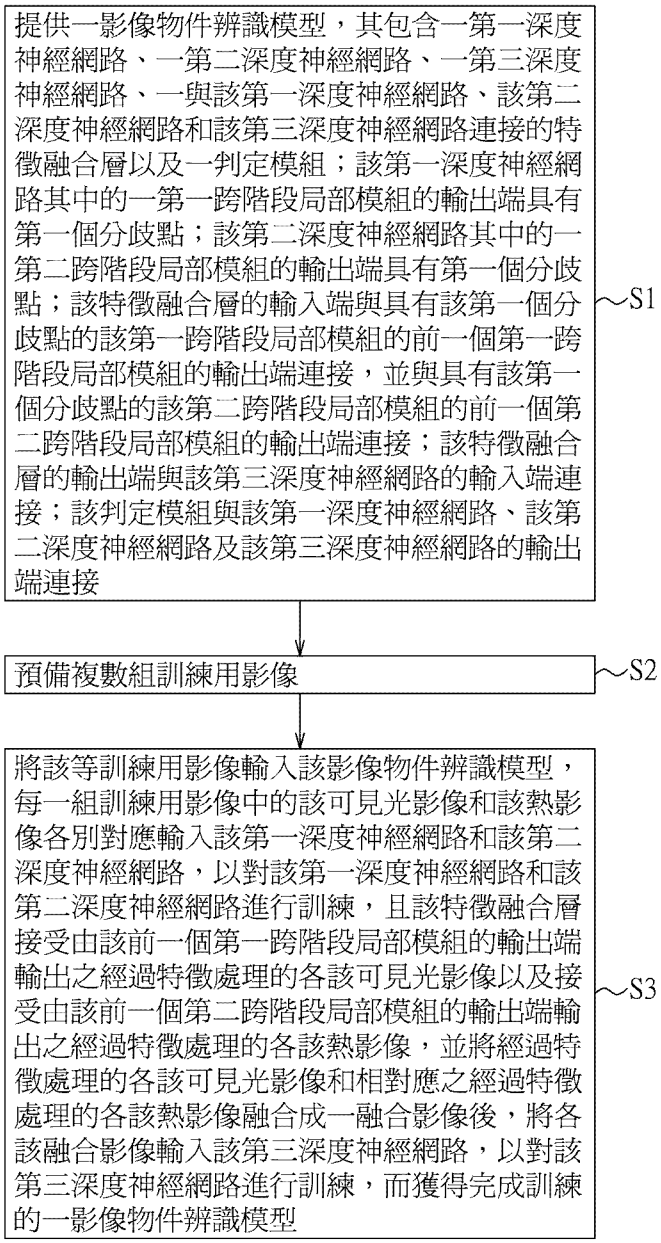

圖 1

公告本

I759156

【發明摘要】

【中文發明名稱】 影像物件辨識模型的訓練方法及影像物件辨識模型

【中文】

一種影像物件辨識模型的訓練方法，該影像物件辨識模型包含第一、第二及第三深度神經網路以及一連接第一、第二及第三深度神經網路的特徵融合層；將複數組訓練用影像其中的每一組訓練用影像中的可見光影像和熱影像各別對應輸入第一和第二深度神經網路，以對第一和第二深度神經網路進行訓練，且該特徵融合層接受由該第一和第二深度神經網路各別輸出之經過特徵處理的各該可見光影像和經過特徵處理的各該熱影像，並將兩者融合成一融合影像後輸入第三深度神經網路，以對第三深度神經網路進行訓練，而獲得完成訓練的該影像物件辨識模型。

【指定代表圖】：圖（1）。

【代表圖之符號簡單說明】

S1~S3 ……步驟

【發明說明書】

【中文發明名稱】 影像物件辨識模型的訓練方法及影像物件辨識模型

【技術領域】

【0001】 本發明是有關於一種影像物件辨識模型及其訓練方法，特別是指一種能根據同一成像時間獲得之同一場景的熱影像與可見光影像進行影像物件辨識的影像物件辨識模型及其訓練方法。

【先前技術】

【0002】 可見光相機(RGB Camera)在天候良好、光線明亮時，其拍攝範圍內之物件成像效果良好，但在光線昏暗，如夜晚無光源處，其成像效果則與光線強弱成反比。而在雨、雪、霧等天候不良或有煙、塵的環境時，則易遭遮蔽且無法穿透，成像效果不佳，以致影響辨識影像中之物件的識別率。熱感攝影機(或稱紅外線相機，Thermal Camera)在天候不佳或光線昏暗環境下，其成像效果較可見光相機佳，但熱感攝影機僅能描繪物件的外型，不能顯示物件的細節輪廓，例如無法顯示人臉的細部特徵，且當所拍攝的相鄰物件溫度相近時，熱感攝影機易混淆相鄰物件而影響辨識影像中之物件的識別率。

【0003】 因此，為解決上述問題，傳統採用上述兩種影像進行影像中之物件辨識的方法會設定一個切換機制，例如白天使用可見光

相機拍攝的可見光影像進行物件辨識，晚上則切換至使用熱感攝影機拍攝的熱影像進行物件辨識；但此種做法需要特別考慮時段而且過度依賴單一種影像，例如即使在晚上但燈火通明的地方，可見光影像的成像效果未必較熱影像差，反之，即使在晚上但溫度差異不大的環境，例如冬天或冰天雪地的地方，熱影像的成像效果亦不見得較可見光影像佳。

【0004】 因此，若能同時採用上述兩種影像進行影像物件辨識，可利用影像互補的效果，而不需考量時段或環境的變化對應切換不同的影像辨識機制，並可進行全天候的影像辨識。

【發明內容】

【0005】 因此，本發明之目的，即在提供一種影像物件辨識模型的訓練方法及一種影像物件辨識模型，其同時採用內容重疊的熱影像與可見光影像進行影像物件辨識，利用影像互補的效果，達到全天候影像辨識。

【0006】 於是，本發明一種影像物件辨識模型的訓練方法，由一電腦執行，並包括：該電腦執行一影像物件辨識模型，該影像物件辨識模型包含一第一深度神經網路、一第二深度神經網路、一第三深度神經網路、一與該第一深度神經網路、該第二深度神經網路和該第三深度神經網路連接的特徵融合層以及一判定模組。

【0007】 該第一深度神經網路包含一第一特徵提取層，該第一

特徵提取層包含複數串接的第一跨階段局部模組，該等第一跨階段局部模組其中的一第一跨階段局部模組的輸出端具有第一個分歧點；該第二深度神經網路包含一第二特徵提取層，該第二特徵提取層包含複數串接的第二跨階段局部模組，該等第二跨階段局部模組其中的一第二跨階段局部模組的輸出端具有第一個分歧點；該特徵融合層的輸入端與具有該第一個分歧點的該第一跨階段局部模組的前一個第一跨階段局部模組的輸出端連接，並與具有該第一個分歧點的該第二跨階段局部模組的前一個第二跨階段局部模組的輸出端連接；該特徵融合層的輸出端與該第三深度神經網路的輸入端連接；該判定模組與該第一深度神經網路、該第二深度神經網路及該第三深度神經網路的輸出端連接。

【0008】 於該電腦中預備複數組訓練用影像，每一組訓練用影像包含在同一時間拍攝且內容重疊的一可見光影像與一熱影像。

【0009】 該電腦將該等訓練用影像的每一組訓練用影像中的該可見光影像和該熱影像各別對應輸入該第一深度神經網路和該第二深度神經網路，以對該第一深度神經網路和該第二深度神經網路進行訓練，且該特徵融合層接受由該前一個第一跨階段局部模組的輸出端輸出之經過特徵處理的各該可見光影像以及接受由該前一個第二跨階段局部模組的輸出端輸出之經過特徵處理的各該熱影像，並將經過特徵處理的各該可見光影像和相對應之經過特徵處理

的各該熱影像融合成一融合影像後，將各該融合影像輸入該第三深度神經網路，以對該第三深度神經網路進行訓練，而獲得完成訓練的一影像物件辨識模型，使得一組待辨識影像中的一待辨識可見光影像和一待辨識熱影像被該電腦各別對應輸入完成訓練的該影像物件辨識模型的該第一深度神經網路和該第二深度神經網路後，完成訓練的該影像物件辨識模型的該第一深度神經網路、該第二深度神經網路及該第三深度神經網路的輸出端將分別輸出與該組待辨識影像相關的複數個候選物件資訊至該判定模組，使該判定模組能根據該等候選物件資訊，辨識出該待辨識可見光影像中的物件。

【0010】 在本發明的一些實施態樣中，該第一深度神經網路包含依序串接的一第一輸入層、該第一特徵提取層、一第一多尺度層及一第一預測層，且該第一跨階段局部模組的該第一個分歧點與該第一多尺度層連接；該第二深度神經網路包含依序串接的一第二輸入層、該第二特徵提取層、一第二多尺度層及一第二預測層，且該第二跨階段局部模組的該第一個分歧點與該第二多尺度層連接；該第三深度神經網路包含依序串接的一第三特徵提取層、一第三多尺度層及一第三預測層；該判定模組與該第一預測層、該第二預測層及該第三預測層的輸出端連接；每一組訓練用影像中的該可見光影像和該熱影像各別由對應的該第一輸入層和該第二輸入層輸入，以對該第一深度神經網路和該第二深度神經網路進行訓練，且該融合

影像被輸入至該第三特徵提取層，以對該第三深度神經網路進行訓練；且該待辨識可見光影像和該待辨識熱影像被各別對應輸入完成訓練的該影像物件辨識模型的該第一深度神經網路的該第一輸入層和該第二深度神經網路的該第二輸入層後，完成訓練的該影像物件辨識模型的該第一深度神經網路的該第一預測層、該第二深度神經網路的該第二預測層及該第三深度神經網路的該第三預測層的輸出端分別輸出與該組待辨識影像相關的複數個候選物件資訊至該判定模組。

【0011】 在本發明的一些實施態樣中，每一組訓練用影像包含的該熱影像是預先根據相對應的該可見光影像進行影像校正，而能與該可見光影像良好地融合的校正後熱影像；且該組待辨識影像中的該待辨識熱影像是預先根據該待辨識可見光影像進行影像校正，而能與該待辨識可見光影像良好地融合的校正後待辨識熱影像。

【0012】 此外，本發明一種影像物件辨識模型，其係根據上述之影像物件辨識模型的訓練方法訓練而成，而能接受包含在同一時間拍攝且內容重疊的一待辨識可見光影像與一待辨識熱影像的一組待辨識影像，以根據該待辨識可見光影像與該待辨識熱影像辨識出該待辨識可見光影像中的物件。

【0013】 再者，本發明一種影像物件辨識模型，其接受包含在同一時間拍攝且內容重疊的一可見光影像與一熱影像的一組影像，以

根據該可見光影像與該熱影像辨識該可見光影像中的物件，並包括：一第一深度神經網路，其接受該可見光影像輸入，並包含一第一特徵提取層，該第一特徵提取層包含複數串接的第一跨階段局部模組，該等第一跨階段局部模組其中的一第一跨階段局部模組的輸出端具有第一個分歧點；一第二深度神經網路，其接受該熱影像輸入，並包含一第二特徵提取層，該第二特徵提取層包含複數串接的第二跨階段局部模組，該等第二跨階段局部網路其中的一第二跨階段局部模組的輸出端具有第一個分歧點；一第三深度神經網路；一特徵融合層，其輸入端與具有該第一個分歧點的該第一跨階段局部模組的前一個第一跨階段局部模組的輸出端連接，並與具有該第一個分歧點的該第二跨階段局部模組的前一個第二跨階段局部模組的輸出端連接，且該特徵融合層的輸出端與該第三深度神經網路的輸入端連接，該特徵融合層接受由該前一個第一跨階段局部模組的輸出端輸出之經過特徵處理的該可見光影像以及接受由該前一個第二跨階段局部模組的輸出端輸出之經過特徵處理的該熱影像，並將經過特徵處理的該可見光影像和相對應之經過特徵處理的該熱影像融合成一融合影像，再將該融合影像輸入該第三深度神經網路；及一判定模組，其與該第一深度神經網路、該第二深度神經網路及該第三深度神經網路的輸出端連接，且該第一深度神經網路、該第二深度神經網路及該第三深度神經網路的輸出端分別輸出與

該組影像相關的複數個候選物件資訊至該判定模組，該判定模組根據該等候選物件資訊，辨識該可見光影像中的物件。

【0014】 在本發明的一些實施態樣中，該第一深度神經網路包含依序串接的一第一輸入層、該第一特徵提取層、一第一多尺度層及一第一預測層，且該第一跨階段局部模組的該第一個分歧點與該第一多尺度層連接；該第二深度神經網路包含依序串接的一第二輸入層、該第二特徵提取層、一第二多尺度層及一第二預測層，且該第二跨階段局部模組的該第一個分歧點與該第二多尺度層連接；該第三深度神經網路包含依序串接的一第三特徵提取層、一第三多尺度層及一第三預測層；該判定模組與該第一預測層、該第二預測層及該第三預測層的輸出端連接；且該可見光影像和該熱影像被各別對應輸入該第一輸入層和該第二輸入層，該融合影像被輸入至該第三特徵提取層，該第一預測層、該第二預測層及該第三預測層的輸出端分別輸出與該組影像相關的複數個候選物件資訊至該判定模組。

【0015】 在本發明的一些實施態樣中，該熱影像是預先根據該可見光影像進行影像校正，而能與該可見光影像良好地融合的校正後熱影像。

【0016】 本發明之功效在於：除了運用該第一深度神經網路和該第二深度神經網路分別對一組輸入的可見光影像和熱影像進行

物件偵測及辨識外，還藉由該特徵融合層獲取該第一深度神經網路的該第一特徵提取層輸出之經過特徵處理的該可見光影像以及獲取該第二深度神經網路的該第二特徵提取層輸出之經過特徵處理的該熱影像，並將兩者融合成該融合影像後，將該融合影像輸入該第三深度神經網路，使對該融合影像進行物件偵測及辨識，使得第一、第二及第三深度神經網路分別輸出複數個候選物件資訊至該判定模組，使該判定模組能根據該等候選物件資訊，辨識該可見光影像中的物件而提升物件辨識能力，而且藉由同時採用在同一時間拍攝的可見光影像及熱影像進行影像物件辨識，可同時取得這兩種影像的特徵，而利用影像特徵互補的效果，進行全天候的影像辨識並提升物件辨識率，使影像物件辨識不致受限於時段、天候或環境的變化。

【圖式簡單說明】

【0017】 本發明之其他的特徵及功效，將於參照圖式的實施方式中清楚地顯示，其中：

圖 1 是本發明影像物件辨識模型的訓練方法的一實施例的主要流程；

圖 2 是本實施例的影像物件辨識模型的架構方塊示意圖；

圖 3 顯示本實施例的第一深度神經網路和第二深度神經網路的各層的組成方塊示意圖；及

圖 4 是本實施例的影像物件辨識模型提取影像特徵的過程示意圖。

【實施方式】

【0018】 在本發明被詳細描述之前，應當注意在以下的說明內容中，類似的元件是以相同的編號來表示。

【0019】 參閱圖 1 所示，是本發明影像物件辨識模型的訓練方法的一實施例的主要流程步驟，由一電腦執行，首先，如圖 1 的步驟 S1，本實施例要預先提供(預備)待訓練的一影像物件辨識模型 100 給該電腦執行，且如圖 2 所示，該影像物件辨識模型 100 包含一第一深度神經網路 1、一第二深度神經網路 2、一第三深度神經網路 3、一連接該第一深度神經網路 1、該第二深度神經網路 2、該第三深度神經網路 3 的特徵融合層 4 以及一判定模組 5。而本實施例的該影像物件辨識模型 100 是基於 YOLOv4 物件偵測模型進行開發，因此以下以 YOLOv4 架構進行說明。值得一提的是，本實施例的該影像物件辨識模型 100 並不限於 YOLOv4 物件偵測模型，也可以使用其它的物件偵測方法，例如但不限於 YOLOv1 、YOLOv2、YOLOv3、R-CNN, Fast R-CNN, Faster R-CNN, Mask R-CNN、Scaled-YOLOv4、DetectoRS 及 EfficientDet 等具有深度學習的人工智慧模型。

【0020】 該第一深度神經網路 1 包含一第一特徵提取層 11，該第一特徵提取層 11 包含一第一跨階段局部網路 (Cross Stage Partial Network，以下簡稱 CSPNet)，該第一 CSPNet 的主要目的是使網路架構能夠獲取更豐富的梯度融合信息並降低計算量，且如圖 3 所示 YOLOv4 之架構方塊圖可知，該第一 CSPNet 具有第一個分歧點 P1，具體而言，該第一 CSPNet 是由複數個串連的第一跨階段局部(CSP)模組 111(以下簡稱第一 CSP 模組 111)組成，該等第一 CSP 模組 111 其中的一個第一 CSP 模組 111 的輸出端具有第一個分歧點 P1。且在本實施例中，如圖 2 所示，該第一深度神經網路 1 是採用但不限於 YOLOv4 架構，所以該第一深度神經網路 1 主要由第一輸入層 10、第一特徵提取層 11、第一多尺度層 12 及第一預測層 13 組成，且第一特徵提取層 11 中的該等第一 CSP 模組 111 實際上為依序串連的 CSP1、CSP2、CSP8、CSP8 和 CSP4 等特徵提取網路，且該分歧點 P1 是第一個 CSP8 的輸出端，該分歧點 P1 除了與第二個 CSP8 連接外，也與第一多尺度層 12 連接，而且第一輸入層 10 和第一個第一 CSP 模組 111(即 CSP1)之間還串連一第一激活層 210(CBM，YOLOv4 網路結構中的最小元件，由 Conv+Bn+Mish 激活函數三者組成)。此外，由於第一輸入層 10、第一特徵提取層 11、第一多尺度層 12 及第一預測層 13 的具體細部架構和功能已是習知技術且非本案技術重點所在，且可

參見公開之 YOLOv4 的相關文獻或介紹，故在此不予贅述。

【0021】 如同第一深度神經網路 1，該第二深度神經網路 2 同樣包含一第二特徵提取層 21，該第二特徵提取層 21 包含一第二跨階段局部網路(CSPNet)，且如圖 3 所示 YOLOv4 之架構方塊圖可知，該第二 CSPNet 具有第一個分歧點 P2，具體而言，如圖 2 所示，該第二 CSPNet 是由複數個串連的第二跨階段局部(CSP)模組 211(以下簡稱第二 CSP 模組 211)組成，該等第二 CSP 模組 211 其中的一個第二 CSP 模組 211 的輸出端具有第一個分歧點 P2。且在本實施例中，如圖 2 所示，該第二深度神經網路 2 是採用但不限於 YOLOv4 架構，所以該第二深度神經網路 2 主要由第二輸入層 20、第二特徵提取層 21、第二多尺度層 22 及第二預測層 23 組成，且第二特徵提取層 21 中的該等第二 CSP 模組 211 實際上為依序串連的 CSP1、CSP2、CSP8、CSP8 和 CSP4 等特徵提取網路，且該分歧點 P2 是第一個 CSP8 的輸出端，該分歧點 P2 除了與第二個 CSP8 連接外，也與第二多尺度層 22 連接。而且第二輸入層 20 和第一個第二 CSP 模組 211(即 CSP1)之間還串連一第二激活層 210(CBM)。而由於第二輸入層 20、第二特徵提取層 21、第二多尺度層 22 及第二預測層 23 的具體細部架構和功能並非本案技術重點所在，且可參見公開之 YOLOv4 的相關文獻或介紹，故在此不予贅述。

【0022】 該特徵融合層 4 的輸入端與具有該第一個分歧點 P1 的該第一 CSP 模組 111(即第一個 CSP8)的前一個第一 CSP 模組 111(即 CSP2)的輸出端 OP1 連接，並與具有該第一個分歧點 P2 的該第二 CSP 模組 211(即第一個 CSP8)的前一個第二 CSP 模組 211(即 CSP2)的輸出端 OP2 連接；該特徵融合層 4 的輸出端與該第三深度神經網路 3 的輸入端連接；且在本實施例中，該第三深度神經網路 3 採用但不限於 YOLOv4 的大部分架構，因此，如圖 2 所示，該第三深度神經網路 3 由第三特徵提取層 31、第三多尺度層 32 及第三預測層 33 組成，第三特徵提取層 31 包含複數第三跨階段局部(CSP)模組 311，且該等第三 CSP 模組 311 實際上為依序串連的 CSP8、CSP8 和 CSP4 等特徵提取網路。由於第三特徵提取層 31、第三多尺度層 32 及第三預測層 33 的具體細部架構和功能並非本案技術重點所在，且可參見 YOLOv4 的相關文獻或介紹，故在此不予贅述。

【0023】 該判定模組 5 與該第一深度神經網路 1、該第二深度神經網路 2 及該第三深度神經網路 3 的輸出端連接，具體而言，該判定模組 5 是與該第一深度神經網路 1 的該第一預測層 13、該第二深度神經網路 2 的該第二預測層 23 及該第三深度神經網路 3 的該第三預測層 33 的輸出端連接。

【0024】 且如圖 1 的步驟 S2，本實施例要於該電腦中預備複數

組訓練用影像，每一組訓練用影像包含在同一時間拍攝且內容重疊的一可見光影像及一熱影像，且該熱影像是已預先經過校正而與該可見光影像尺寸一致且能良好地重疊(套疊)在一起的校正後影像，其校正方法可以參考但不限於台灣第 110104936 號專利申請案；另外說明的是，該熱影像原始的解析度通常是 640x512，而該可見光影像原始的解析度通常是 1920x1080、1280x720 或 640x512 等，但為了配合該影像物件辨識模型 100 要求的輸入影像尺寸，例如 224x224、416x416 或 608x608 等，在使用該影像物件辨識模型 100 時，本實施例會將要輸入該影像物件辨識模型 100 的每一組訓練用影像(該可見光影像和該熱影像)的大小調整(Resize)為模型能接受的尺寸，例如 416x416。且本實施例最終產生的物件辨識結果並不會呈現在調整大小後的該可見光影像上，而是呈現在原始的該可見光影像上或者呈現在融合前述雙影像的一融合影像上。

【0025】 然後，如圖 1 的步驟 S3，該電腦將該等訓練用影像的每一組訓練用影像中的該可見光影像和該熱影像各別對應輸入該第一深度神經網路 1 的該第一輸入層 10 和該第二深度神經網路 2 的該第二輸入層 20，以對該第一深度神經網路 1 和該第二深度神經網路 2 進行訓練和深度學習；具體而言，該第一深度神經網路 1 的該第一特徵提取層 11 和該第二深度神經網路 2 的該第二特徵提取層 21 皆使用 CSPDarknet53 神經網絡架構，該第一特徵提取層

11(又稱骨幹層(Backbone)主要對經由第一輸入層 10 輸入的可見光影像 61 進行特徵提取，將特徵去蕪存菁，例如圖 3 所示，可見光影像 61 經過該第一特徵提取層 11 的 5 個第一 CSP 模組(CSP1、CSP2、CSP8、CSP8、CSP4)依序進行特徵提取後，該第一特徵提取層 11 輸出大小為 13x13 的第一特徵圖 61’；同理，該第二特徵提取層 21(又稱骨幹層(Backbone)主要對經由第二輸入層 20 輸入的熱影像 62 進行特徵提取，將特徵去蕪存菁，例如圖 3 所示，熱影像 62 經過該第二特徵提取層 21 的 5 個第二 CSP 模組(CSP1、CSP2、CSP8、CSP8、CSP4) 依序進行特徵提取後，該第二特徵提取層 21 輸出大小為 13x13 的第二特徵圖 62’。且在圖 4 中，416x416x3 是指將影像分成三路輸入 CBM 模組 110、210，416x416x32 是指 CBM 模組 110、210 輸出由 32 個大小為 416x416 的特徵圖所組成的圖層，208x208x128 是指 CSP1 模組 111、211 輸出由 128 個大小為 208x208 的特徵圖所組成的圖層，依此類推。

【0026】 且如圖 4 所示，該特徵融合層 4 接受該第一深度神經網路 1 之由該前一個第一 CSP 模組 111(CSP2)輸出之經過特徵處理(即特徵提取)的各該可見光影像 610(即圖 4 上方 CSP2 輸出之大小為 104x104 的可見光影像特徵圖)以及接受由該前一個第二 CSP 模組 211(CSP2)的輸出之經過特徵處理(即特徵提取)的各該

熱影像 620(即圖 4 下方 CSP2 輸出之大小為 104x104 的熱影像特徵圖)，並將經過特徵處理的各該可見光影像 610 和相對應之經過特徵處理的各該熱影像 620 融合成一融合影像 63，再將各該融合影像 63 輸入該第三深度神經網路 3，以對該第三深度神經網路 3 進行訓練和深度學習，具體而言，第三深度神經網路 3 的該第三特徵提取層 31(又稱骨幹層(Backbone)將對該融合影像 63 進行特徵提取，將特徵去蕪存菁，例如圖 4 所示，該融合影像 63 經過該第三特徵提取層 31 的 3 個 CSP 模組(CSP8、CSP8、CSP4) 依序進行特徵提取後，該第三特徵提取層 31 輸出大小為 13x13 的第三特徵圖 63’。

【0027】 接著，第一、第二及第三特徵圖 61’、62’、63’被分別對應輸入至該第一多尺度層 12、該第二多尺度層 22 及該第三多尺度層 32，其中各該多尺度層 12、22、32 又稱頸部層(Neck)，其主要由多尺度模組所組成，用來增強模型多尺度(小物件)的偵測能力，以擴大感受野以及融合不同尺度特徵圖的信息，以更好地進行特徵融合。而本實施例的各該多尺度層 12、22、32 是採用但不限於 YOLOv4 中的 SPP (Spatial Pyramid Pooling)和 PANet (Path Aggregation Network)架構。因此，該第一多尺度層 12、該第二多尺度層 22 及該第三多尺度層 32 分別針對第一、第二及第三特徵圖 61’、62’、63’進行進一步的特徵提取，並分別輸出第一、第

二及第三最終特徵圖至相對應的第一預測層 13、第二預測層 23 和第三預測層 33，且本實施例的第一預測層 13、第二預測層 23 和第三預測層 33 是採用 YOLOv4 中的 Dense Prediction，且基於 YOLO head 進行開發，因此，該第一預測層 13、第二預測層 23 和第三預測層 33 能分別根據輸入的第一、第二和第三最終特徵圖中的影像特徵進行候選框偵測以及物件辨識並分別輸出複數個候選物件資訊，每一個候選物件資訊至少包含一物件候選框及其對應的一信心指數(分數或機率)。且該等候選物件資訊被分別輸入該判定模組 5。

【0028】 該判定模組 5 在本實施例中是採用 DIOU-NMS 演算法，其中 DIOU 的全文為 Distance Intersection over Union，NMS 的全文為 Non-Max Suppression，而 DIOU-NMS 演算法的主要原理為利用信心指數來判斷哪一個物件候選框是最佳的候選框。且由於 DIOU-NMS 演算法已是一習知演算法，且非本案主要重點所在，故在此不予詳述。藉此，該判定模組 5 將根據 DIOU-NMS 演算法之原理，從該等候選物件資訊中選出最佳的候選物件資訊，並將選出的一或一個以上的最佳候選物件資訊(包含物件的候選框及其信心指數)標註於各該可見光影像中。

【0029】 因此，該電腦藉由上述複數組訓練用影像反覆訓練該影像物件辨識模型 100，將使該影像物件辨識模型 100 的辨識率逐

漸提升並收斂至一目標值，而獲得完成訓練的該影像物件辨識模型100，藉此，當一組待辨識影像中的一待辨識熱影像和一待辨識可見光影像被該電腦各別對應輸入完成訓練的該影像物件辨識模型100的該第一深度神經網路1和該第二深度神經網路2後，完成訓練的該影像物件辨識模型100的該第一深度神經網路1、該第二深度神經網路2及該第三深度神經網路3的輸出端(即第一預測層13、第二預測層23和第三預測層33的輸出端)將分別輸出與該組待辨識影像相關的複數個候選物件資訊至該判定模組5，其中每一個候選物件資訊包含框選物件的候選框及其信心指數，且該判定模組5將根據該等候選物件資訊，辨識出該待辨識可見光影像中的物件，並於輸出的該待辨識可見光影像中，將辨識的物件框選並標註其類別(例如人、車(汽車、卡車、機車、公車等)、動物(狗、貓、馬等)、植物等)。值得一提的是，本實施例也可應用但不限於台灣第110104936號專利申請案提供的雙影像融合方法，將該待辨識熱影像和該待辨識可見光影像融合成一融合影像後輸出，並根據影像辨識結果，將該融合影像中被辨識的物件框選並標註其類別。

【0030】 綜上所述，上述實施例除了運用第一深度神經網路1和第二深度神經網路2分別對一組輸入的可見光影像和熱影像進行物件偵測及辨識外，還藉由該特徵融合層4連接第一深度神經網路1的該第一特徵提取層11與第二深度神經網路2的該第二特徵提取層

21，以獲取該第一特徵提取層11中之一第一CSP模組111輸出之經過特徵處理的該可見光影像(特徵圖)610以及獲取該第二特徵提取層21中之一第二CSP模組211輸出之經過特徵處理的該熱影像(特徵圖)620，並將經過特徵處理的該可見光影像610和相對應之經過特徵處理的該熱影像620融合成一融合影像63後，將該融合影像63輸入該第三深度神經網路3，使對該融合影像63進行物件偵測及辨識，使得第一、第二及第三深度神經網路1、2、3分別輸出複數個候選物件資訊至該判定模組5，使該判定模組5能根據該等候選物件資訊，辨識出該可見光影像中的物件，而且，本實施例的影像物件辨識模型100藉由同時採用在同一時間拍攝的可見光影像及熱影像進行影像物件辨識，可同時取得這兩種影像的特徵，而利用影像特徵互補的效果，進行全天候的影像辨識並提升物件辨識率，使影像物件辨識不致受限於時段、天候或環境的變化，也不需根據時段、天候或環境變化不斷地切換不同的影像辨識機制，確實達到本發明的功效與目的。

【0031】 惟以上所述者，僅為本發明之實施例而已，當不能以此限定本發明實施之範圍，凡是依本發明申請專利範圍及專利說明書內容所作之簡單的等效變化與修飾，皆仍屬本發明專利涵蓋之範圍內。

【符號說明】

【0032】

100……影像物件辨識模型

1……第一深度神經網路

10 ……第一輸入層

11 ……第一特徵提取層

111 ……第一跨階段局部(CSP)模組

12 ……第一多尺度層

13 ……第一預測層

2……第二深度神經網路

20 ……第二輸入層

21 ……第二特徵提取層

211 ……第二跨階段局部(CSP)模組

22 ……第二多尺度層

23 ……第二預測層

3……第三深度神經網路

31 ……第三特徵提取層

311 ……第三跨階段局部(CSP)模組

32…… 第三多尺度層

33…… 第三預測層

4 …… 特徵融合層

5 …… 判定模組

61…… 可見光影像

61’…… 第一特徵圖

610 …… 經過特徵處理的可見光影像

62…… 熱影像

62’…… 第二特徵圖

620 …… 經過特徵處理的熱影像

63…… 融合影像

63’…… 第三特徵圖

P1、P2 …. 第一個分歧點

OP1、OP2 輸出端

S1~S3…… 步驟

【發明申請專利範圍】

【請求項1】一種影像物件辨識模型的訓練方法，由一電腦執行，並包括：

該電腦執行一影像物件辨識模型，該影像物件辨識模型包含一第一深度神經網路、一第二深度神經網路、一第三深度神經網路、一與該第一深度神經網路、該第二深度神經網路和該第三深度神經網路連接的特徵融合層以及一判定模組；該第一深度神經網路包含一第一特徵提取層，該第一特徵提取層包含複數串接的第一跨階段局部模組，該等第一跨階段局部模組其中的一第一跨階段局部模組的輸出端具有第一個分歧點；該第二深度神經網路包含一第二特徵提取層，該第二特徵提取層包含複數串接的第二跨階段局部模組，該等第二跨階段局部模組其中的一第二跨階段局部模組的輸出端具有第一個分歧點；該特徵融合層的輸入端與具有該第一個分歧點的該第一跨階段局部模組的前一個第一跨階段局部模組的輸出端連接，並與具有該第一個分歧點的該第二跨階段局部模組的前一個第二跨階段局部模組的輸出端連接；該特徵融合層的輸出端與該第三深度神經網路的輸入端連接；該判定模組與該第一深度神經網路、該第二深度神經網路及該第三深度神經網路的輸出端連接；

於該電腦中預備複數組訓練用影像，每一組訓練用影像包含在同一時間拍攝且內容重疊的一可見光影像與一熱影像；及

該電腦將該等訓練用影像的每一組訓練用影像中的該可見光影像和該熱影像各別對應輸入該第一深度神經網路和該第二深度神經網路，以對該第一深度神經網路和該第二深度神經網路進行訓練，且該特徵融合層接受由該前一個第一跨階段局部模組的輸出端輸出之經過特徵處理的各該可見光影像以及接受由該前一個第二跨階段局部模組的輸出端輸出之經過特徵處理的各該熱影像，並將經過特徵處理的各該可見光影像和相對應之經過特徵處理的各該熱影像融合成一融合影像後，將各該融合影像輸入該第三深度神經網路，以對該第三深度神經網路進行訓練，而獲得完成訓練的一影像物件辨識模型，使得一組待辨識影像中的一待辨識可見光影像和一待辨識熱影像被該電腦各別對應輸入完成訓練的該影像物件辨識模型的該第一深度神經網路和該第二深度神經網路後，完成訓練的該影像物件辨識模型的該第一深度神經網路、該第二深度神經網路及該第三深度神經網路的輸出端將分別輸出與該組待辨識影像相關的複數個候選物件資訊至該判定模組，使該判定模組能根據該等候選物件資訊，辨識出該待辨識可見光影像中的物件。

【請求項2】如請求項1所述影像物件辨識模型的訓練方法，其中，該第一深度神經網路包含依序串接的一第一輸入層、該第一特徵提取層、一第一多尺度層及一第一預測層，且該第一跨階段局部模組的該第一個分歧點與該第一多尺度層連接；該第二深度神經網路包含依序串接的一第二輸入

層、該第二特徵提取層、一第二多尺度層及一第二預測層，且該第二跨階段局部模組的該第一個分歧點與該第二多尺度層連接；該第三深度神經網路包含依序串接的一第三特徵提取層、一第三多尺度層及一第三預測層；該判定模組與該第一預測層、該第二預測層及該第三預測層的輸出端連接；每一組訓練用影像中的該可見光影像和該熱影像各別由對應的該第一輸入層和該第二輸入層輸入，以對該第一深度神經網路和該第二深度神經網路進行訓練，且該融合影像被輸入至該第三特徵提取層，以對該第三深度神經網路進行訓練；且該待辨識可見光影像和該待辨識熱影像被各別對應輸入完成訓練的該影像物件辨識模型的該第一深度神經網路的該第一輸入層和該第二深度神經網路的該第二輸入層後，完成訓練的該影像物件辨識模型的該第一深度神經網路的該第一預測層、該第二深度神經網路的該第二預測層及該第三深度神經網路的該第三預測層的輸出端分別輸出與該組待辨識影像相關的複數個候選物件資訊至該判定模組。

【請求項3】如請求項1所述影像物件辨識模型的訓練方法，其中每一組訓練用影像包含的該熱影像是預先根據相對應的該可見光影像進行影像校正，而能與該可見光影像良好地融合的校正後熱影像；且該組待辨識影像中的該待辨識熱影像是預先根據該待辨識可見光影像進行影像校正，而能與該待辨識可見光影像良好地融合的校正後待辨識熱影像。

【請求項4】一種影像物件辨識模型，其係根據請求項1至3其中任一項所述影像物件辨識模型的訓練方法訓練而成，而能接受包含在同一時間拍攝且內容重疊的一待辨識可見光影像與一待辨識熱影像的一組待辨識影像，以根據該待辨識可見光影像與該待辨識熱影像辨識出該待辨識可見光影像中的物件。

【請求項5】一種影像物件辨識模型，其接受包含在同一時間拍攝且內容重疊的一可見光影像與一熱影像的一組影像，以根據該可見光影像與該熱影像辨識該可見光影像中的物件，並包括：

一第一深度神經網路，其接受該可見光影像輸入，並包含一第一特徵提取層，該第一特徵提取層包含複數串接的第一跨階段局部模組，該等第一跨階段局部模組其中的一第一跨階段局部模組的輸出端具有第一個分歧點；

一第二深度神經網路，其接受該熱影像輸入，並包含一第二特徵提取層，該第二特徵提取層包含複數串接的第二跨階段局部模組，該等第二跨階段局部網路其中的一第二跨階段局部模組的輸出端具有第一個分歧點；

一第三深度神經網路；

一特徵融合層，其輸入端與具有該第一個分歧點的該第一跨階段局部模組的前一個第一跨階段局部模組的輸出端連接，並與具有該第一個分歧點的該第二跨階段局部模組的前一個第二跨階段局部模組的輸出端連接，

且該特徵融合層的輸出端與該第三深度神經網路的輸入端連接，該特徵融合層接受由該前一個第一跨階段局部模組的輸出端輸出之經過特徵處理的該可見光影像以及接受由該前一個第二跨階段局部模組的輸出端輸出之經過特徵處理的該熱影像，並將經過特徵處理的該可見光影像和相對應之經過特徵處理的該熱影像融合成一融合影像，再將該融合影像輸入該第三深度神經網路；及

一判定模組，其與該第一深度神經網路、該第二深度神經網路及該第三深度神經網路的輸出端連接，且該第一深度神經網路、該第二深度神經網路及該第三深度神經網路的輸出端分別輸出與該組影像相關的複數個候選物件資訊至該判定模組，該判定模組根據該等候選物件資訊，辨識該可見光影像中的物件。

【請求項6】如請求項5所述的影像物件辨識模型，其中，該第一深度神經網路包含依序串接的一第一輸入層、該第一特徵提取層、一第一多尺度層及一第一預測層，且該第一跨階段局部模組的該第一個分歧點與該第一多尺度層連接；該第二深度神經網路包含依序串接的一第二輸入層、該第二特徵提取層、一第二多尺度層及一第二預測層，且該第二跨階段局部模組的該第一個分歧點與該第二多尺度層連接；該第三深度神經網路包含依序串接的一第三特徵提取層、一第三多尺度層及一第三預測層；該判定模組與該第一預測層、該第二預測層及該第三預測層的輸出端連接；且該可見光影像和該熱影像被各別對應輸入該第

一輸入層和該第二輸入層，該融合影像被輸入至該第三特徵提取層，該第一預測層、該第二預測層及該第三預測層的輸出端分別輸出與該組影像相關的複數個候選物件資訊至該判定模組。

【請求項7】如請求項5所述的影像物件辨識模型，其中該熱影像是預先根據該可見光影像進行影像校正，而能與該可見光影像良好地融合的校正後熱影像。

【發明圖式】

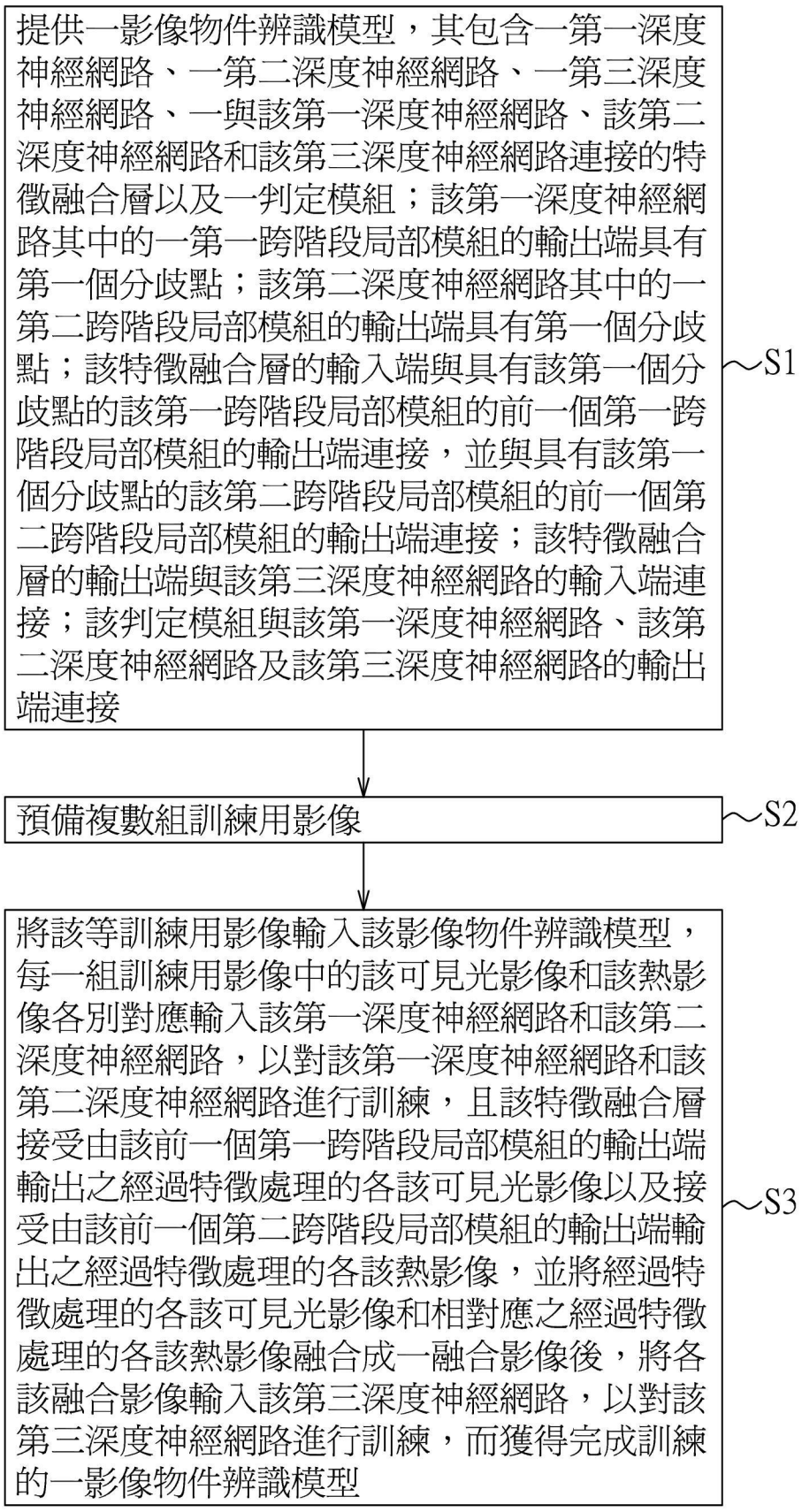
提供一影像物件辨識模型，其包含一第一深度神經網路、一第二深度神經網路、一第三深度神經網路、一與該第一深度神經網路、該第二深度神經網路和該第三深度神經網路連接的特徵融合層以及一判定模組；該第一深度神經網路其中的一第一跨階段局部模組的輸出端具有第一個分歧點；該第二深度神經網路其中的一第二跨階段局部模組的輸出端具有第一個分歧點；該特徵融合層的輸入端與具有該第一個分歧點的該第一跨階段局部模組的前一個第一跨階段局部模組的輸出端連接，並與具有該第一個分歧點的該第二跨階段局部模組的前一個第二跨階段局部模組的輸出端連接；該特徵融合層的輸出端與該第三深度神經網路的輸入端連接；該判定模組與該第一深度神經網路、該第二深度神經網路及該第三深度神經網路的輸出端連接
S1
預備複數組訓練用影像
S2
將該等訓練用影像輸入該影像物件辨識模型，每一組訓練用影像中的該可見光影像和該熱影像各別對應輸入該第一深度神經網路和該第二深度神經網路，以對該第一深度神經網路和該第二深度神經網路進行訓練，且該特徵融合層接受由該前一個第一跨階段局部模組的輸出端輸出之經過特徵處理的各該可見光影像以及接受由該前一個第二跨階段局部模組的輸出端輸出之經過特徵處理的各該熱影像，並將經過特徵處理的各該可見光影像和相對應之經過特徵處理的各該熱影像融合成一融合影像後，將各該融合影像輸入該第三深度神經網路，以對該第三深度神經網路進行訓練，而獲得完成訓練的一影像物件辨識模型
S3

圖 1

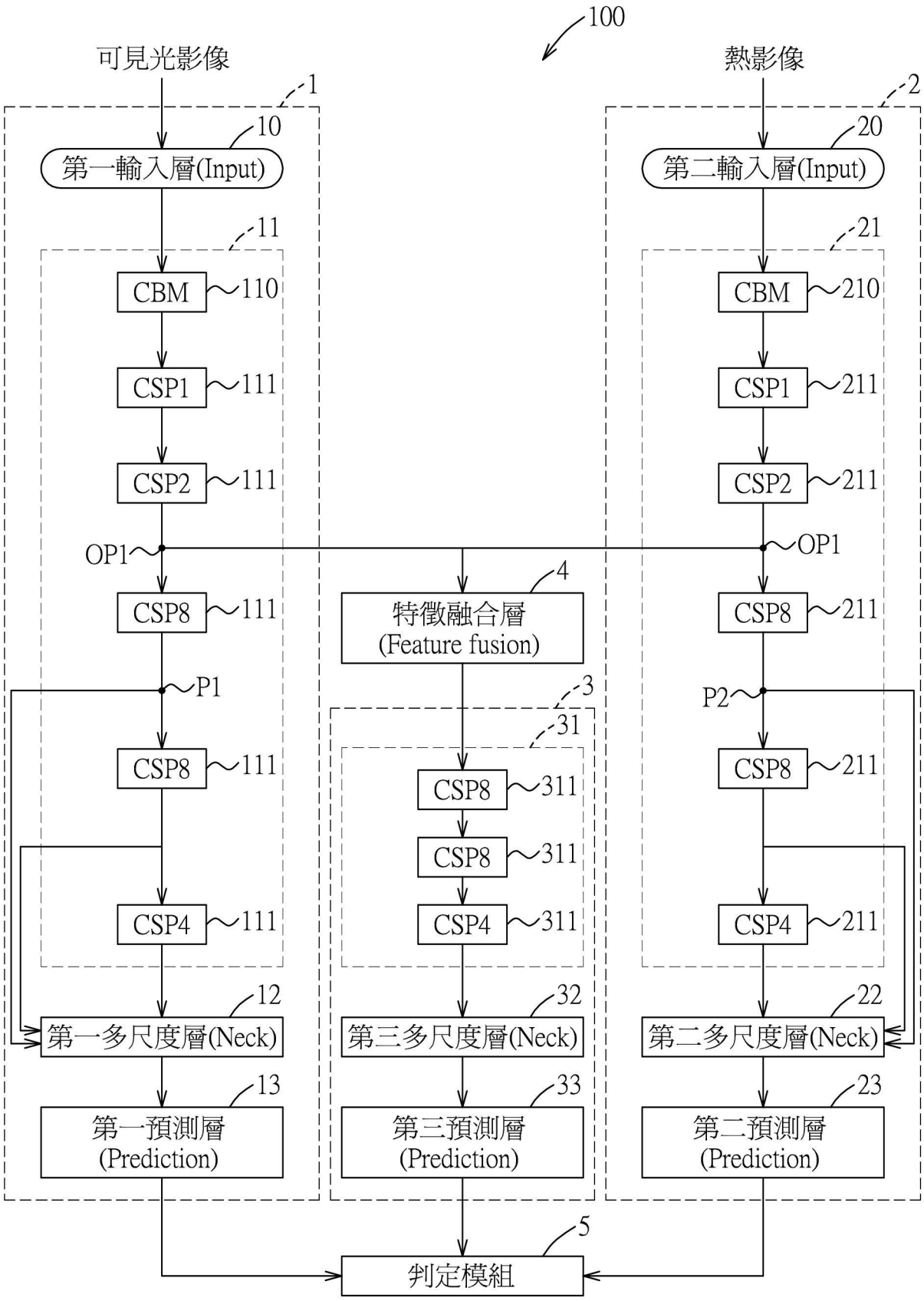
100
可見光影像
熱影像
1
2
10
20
第一輸入層(Input)
第二輸入層(Input)
11
21
CBM
110
210
CSP1
111
211
CSP2
111
211
OP1
OP1
4
特徵融合層
(Feature fusion)
CSP8
111
211
P1
P2
3
31
CSP8
311
CSP8
311
CSP4
111
311
211
12
32
22
第一多尺度層(Neck)
第三多尺度層(Neck)
第二多尺度層(Neck)
13
33
23
第一預測層
(Prediction)
第三預測層
(Prediction)
第二預測層
(Prediction)
5
判定模組

圖 2

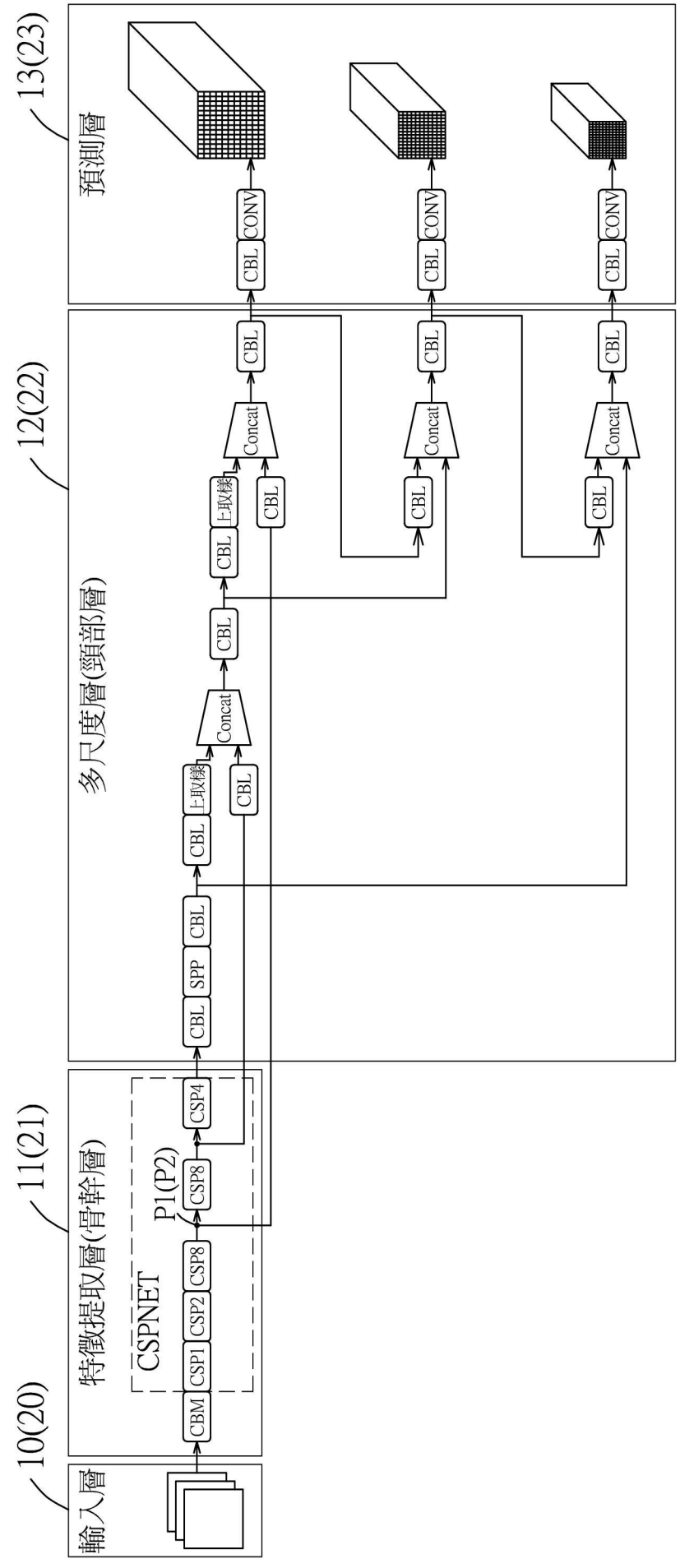
10(20)
輸入層
11(21)
特徵提取層(骨幹層)
CSPNET
P1(P2)
CBM
CSP1
CSP2
CSP8
CSP8
CSP4
12(22)
多尺度層(頸部層)
CBL
SPP
CBL
CBL
上取樣
CBL
Concat
CBL
CBL
上取樣
CBL
Concat
CBL
CBL
Concat
CBL
13(23)
預測層
CBL
CONV
CBL
CONV
CBL
CONV

圖 3

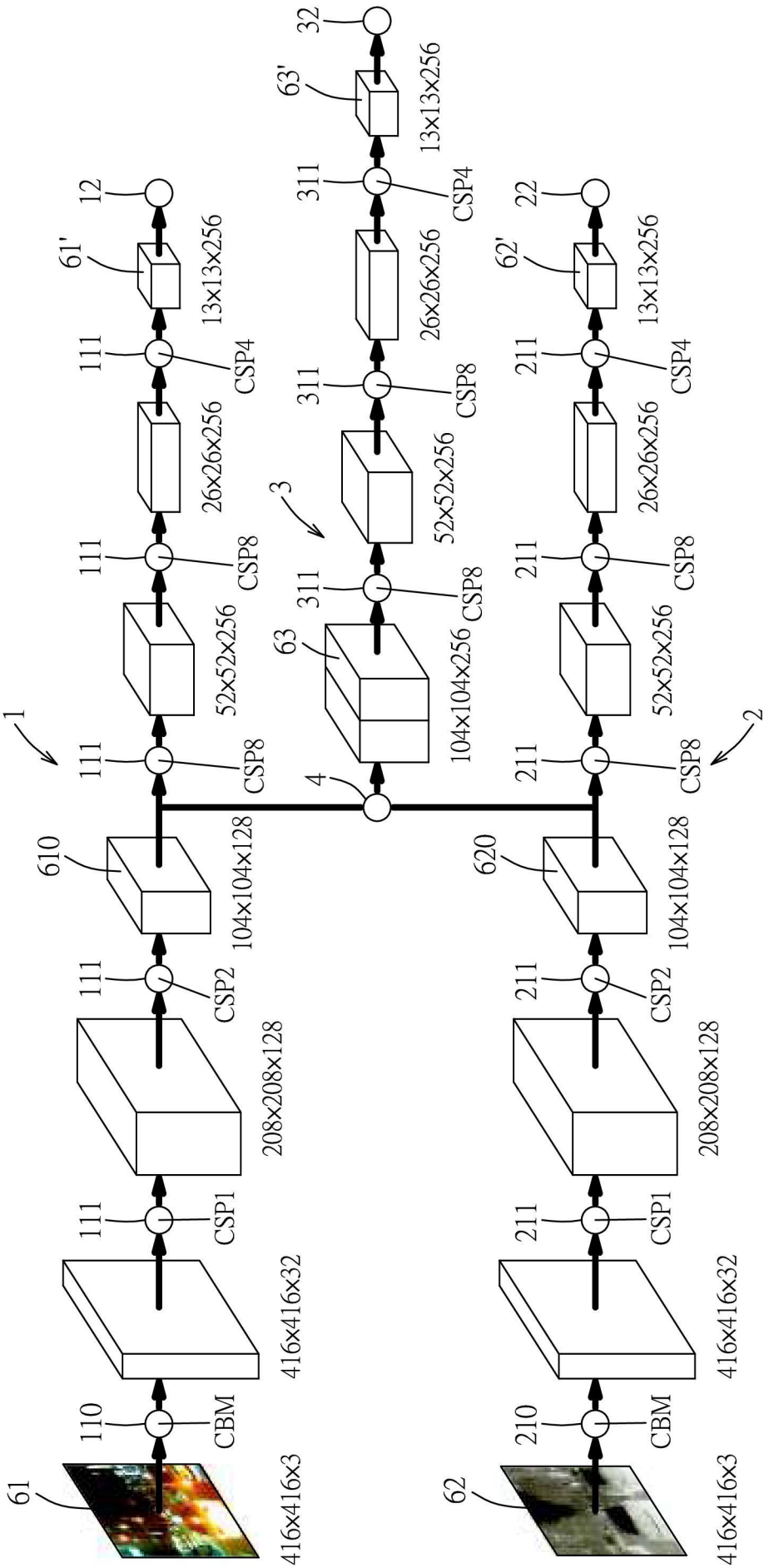
61
416x416x3
110
CBM
416x416x32
111
CSP1
208x208x128
111
CSP2
104x104x128
610
111
CSP8
1
52x52x256
111
CSP8
26x26x256
111
CSP4
61'
13x13x256
12
4
63
104x104x256
311
CSP8
3
52x52x256
311
CSP8
26x26x256
311
CSP4
63'
13x13x256
32
62
416x416x3
210
CBM
416x416x32
211
CSP1
208x208x128
211
CSP2
104x104x128
620
211
CSP8
2
52x52x256
211
CSP8
26x26x256
211
CSP4
62'
13x13x256
22

圖 4