
應用深度學習研製智慧行動辨識不同災害
類型影像之研究

**A Study of Applying Deep Learning in
Developing Intelligent Mobile Recognition
for Different Types of Disasters Images
成果報告書**

執行單位：國立虎尾科技大學

執行期間：109年02月06日至109年12月31日

計畫主持人：蔡鴻旭 教授

行政院農業委員會水土保持局 編印

中華民國 109 年 12 月

(本報告書內容及建議純屬執行單位意見，僅供本局施政參考)

應用深度學習研製智慧行動辨識不同災害類型影像 之研究

摘要

首先，本計畫已發展基於卷積類神經網路以圖找圖技術，初步建置影像內容檢索(Content-based image retrieval; CBIR)系統。以水土保持局建置歷史影像平台(image repository platform - IRP)中影像集為例。使用者輸入一張查詢影像(query image)，CBIR 系統檢索後輸出一些影像其內容相似於查詢影像。系統可進一步提供輸出影像在 IRP 中的文字註解。目前已完成發展跨平台執行影像內容檢索功能，並優化 CBIR 系統之檢索效能及效率。另外，IRP 儲存歷年土砂災害照片，本計畫也發展出智慧型影像辨識技術可自動化判別土石流、崩塌等不同災害類型，未來可應用於水土保持相關業務之分類、辨識、預測等相關業務。然後，本計畫已發展出智慧型影像實例分割技術，應用於影像視覺表徵因子中綠視率與水視率之量化估計。

關鍵詞：影像內容檢索、影像分類、水土保持、深度學習、機器學習、災害影像分類

A Study of Applying Deep Learning in Developing Intelligent Mobile Recognition for Different Types of Disasters Images

Abstract

First, this project has developed an image content-based retrieval (CBIR) system which can retrieve images for a query image based on image contents. The access of the CBIR system supports a cross-platform operation including mobile access. The design of the CBIR system based on convolution neural network (CNN). The image set of the project comes from image repository platform (IRP) built by Soil and Water Conservation Bureau (SWCB). Users send a query image to the CBIR system. The images with similar content with that of the query image are displayed on the screen of devices. Then, users select one of similar images to get its corresponding descriptions with text annotations from the IRP. Moreover, the project has developed mobile computations for cross-platform image retrieval. Additionally, this project has developed an intelligent image classification system based on deep learning techniques, which can predict types of disaster images such as debris flow or landslide. The system can be applied in predicating disaster images. The results of predication can be employed to automatically determine what affairs for soil and water conservation will be responsible for disaster images. Furthermore, the project has developed an intelligent image instance segmentation technique for quantitatively estimating the

green looking ratio and the water looking ratio of visual preference of an image.

Keywords : Image Content Retrieval, Image Classification, Soil and Water Conservation, Deep Learning, Machine Learning, Disaster Image Classification

目次

摘要.....	I
Abstract.....	II
目次.....	IV
表次.....	VI
圖次.....	VIII
第一章 計畫內容	1-1
第一節 計畫動機	1-1
第二節 計畫目的	1-3
第三節 背景	1-6
一、 影像檢索及分類	1-6
二、 影像切割	1-6
三、 深度卷積編碼器-解碼器	1-8
四、 Scikit-image.....	1-10
五、 影像檢索及影像分類量化指標	1-10
第二章 工作項目與執行方法	2-1
第一節 工作項目	2-1
第二節 工作方法	2-2
第三節 工作執行步驟	2-3
一、 訓練影像資料庫收集與建置之研究	2-3
二、 影像前處理分析之研究	2-5
三、 以深度學習技術建置影像檢索模型之研究	2-9
四、 以深度學習技術建置不同災害類型影像分類模型之研究	2-22

五、 模擬實驗驗證模型之研究	2-27
六、 行動化模組之研究	2-57
七、 水土保持設施工程美學照片之內涵分析之量化估計 ...	2-63
第三章 執行成果及效益	3-1
第一節 執行成果	3-1
第二節 執行效益	3-2
一、 基於 IRP 影像檢索系統效益	3-2
二、 智慧型影像分類技術效益	3-2
三、 影像實例分割技術效益	3-2
第三節 結果與討論	3-4
第四章 結論及建議	4-1
第一節 結論	4-1
第二節 建議	4-2
參考文獻.....	參-1
附錄.....	附錄-1
附錄一、 期初工作執行計畫書審查意見	附錄-1
附錄二、 期中計畫書審查意見	附錄-5
附錄三、 Case 3 查詢結果案例說明	附錄-9
附錄四、 拍攝類別及 CCTV 災害類別影像.....	附錄-12
附錄五、 Labelme 操作過程及說明	附錄-23
附錄六、 實例分割影像與估計 WLR 與 GLR	附錄-31
附錄七、 期末報告計畫書審查意見	附錄-36

表次

表 1-1 二元分類混淆矩陣	1-11
表 2-1 各層의影像大小及深度層數	2-12
表 2-2 各層의影像大小及深度層數	2-13
表 2-3 RESNET50V2 各層架構	2-26
表 2-4 說明 CASE 1 檢索效能，用 ARR 呈現，訓練樣本 D_1 ，測試樣本 $\Gamma_{D_1^1}$	2-34
表 2-5 說明 CASE 1 檢索效能，用 ARR 呈現，訓練樣本 D_1 ，測試樣本 $\Gamma_{D_1^2}$	2-35
表 2-6 說明 CASE 1 檢索效率，用 ART(秒/張)呈現，所有樣本 D_1 ，查詢樣本 $\Gamma_{D_1^1}$	2-36
表 2-7 說明 CASE 1 檢索效率，用 ART(秒/張)呈現，所有樣本 D_1 ，查詢樣本 $\Gamma_{D_1^2}$	2-37
表 2-8 雲端伺服器主要硬體規格，訓練智慧型影像檢索系統	2-37
表 2-9 網頁伺服器主要硬體規格，用於影像檢索計算	2-37
表 2-10 說明 CASE 2 一般化檢索效能，用 ARR 呈現，訓練樣本 D_1^1 ，一般化測試樣本 D_1^2 ，測試樣本 $\Gamma_{D_1^1}$	2-40

表 2-11 說明 CASE 2 一般化檢索效能，用 ARR 呈現，訓練樣本 D_1^1 ， 一般化測試樣本 D_1^2 ，測試樣本 $\Gamma_{D_1^2}$	2-41
表 2-12 列出拍攝類別影像樣本及 CCTV 災害類別影像中訓練集、 驗證集及測試集的張數	2-45
表 2-13 拍攝類別訓練集和測試集的混淆矩陣和 AR	2-46
表 2-14 CCTV 災害類別訓練集和測試集的混淆矩陣和 AR	2-50
表 2-15 列出分類拍攝類別，分類正確的部份案例	2-51
表 2-16 列出分類拍攝類別，預測分類錯誤的部份案例	2-52
表 2-17 列出分類 CCTV 災害類別，分類正確的部份案例。	2-53
表 2-18 列出分類 CCTV 災害類別，分類錯誤的部份案例	2-54
表 2-19 收集水土保持設施工程美學照片及其來源	2-63
表附錄 4-1 拍攝類別中的 CCTV 的影像樣本	附錄-12
表附錄 4-2 拍攝類別中的高空拍攝的影像樣本	附錄-14
表附錄 4-3 CCTV 災害類別中的土石流影像樣本	附錄-15
表附錄 4-4 CCTV 災害類別中的設施崩塌影像樣本	附錄-17
表附錄 4-5 CCTV 災害類別中的山坡崩塌影像樣本	附錄-19
表附錄 4-6 CCTV 災害類別中的其他影像樣本	附錄-21
表附錄 6-1 舉例說明原圖、實例分割結果及估計 WLR 與 GLR，訓 練影像集(含驗證影像集)。	附錄-31
表附錄 6-2 舉例說明原圖、實例分割結果及估計 WLR 與 GLR，測 試影像集。	附錄-34

圖次

圖 1-1 跨平台行動載具執行影像檢索的系統架構圖	1-3
圖 1-2 智慧型影像分類判別技術架構圖	1-4
圖 1-3 運用 MASK RCNN 來實施實例分割流程	1-5
圖 1-4 實施實例分割，將每個氣球輪廓分割出來，每個汽球用不同 顏色來區分。	1-7
圖 1-5 MASK RCNN 架構圖 (圖片來源: [27])	1-8
圖 1-6 DCED 架構	1-9
圖 1-7 ENCODER-DECODER 模型，HIDDEN LAYER 用 CNN 來設計	1-10
圖 2-1 本計畫實施方法構想圖	2-1
圖 2-2 本計畫系統架構含七項主題與各議題間的關聯流程圖	2-2
圖 2-3 IRP 影像 API	2-4
圖 2-4 IRP 災害影像命名方式	2-4
圖 2-5 物件偵測技術偵測出人物移除這些物件	2-5
圖 2-6 物件偵測技術偵測出人形及人臉的流程	2-6
圖 2-7 各災害類別資料夾內相似災害影像檔案名稱	2-8
圖 2-8 各災害類別資料夾內相似災害影像檔案名稱	2-8
圖 2-9 圖找圖影像檢索範例	2-9
圖 2-10 細部 CNN 架構設計 ENCODER 擷取影像	2-11
圖 2-11 細部 CNN 架構設計 DECODER，還原影像	2-12
圖 2-12 基於 CNN 設計 AUTOENCODER 概念圖，擷取影像特徵	2-14
圖 2-13 建置影像特徵資料庫 F_{D_1}	2-15
圖 2-14 影像特徵比對的演算法	2-16

圖 2- 15 將 F_{D_1} 特徵庫，以 APNN 演算法建立 n_{APNN} 顆二元樹的流程圖	2-17
圖 2- 16 將 F_{D_1} 特徵庫，以 APNN 演算法建立 n_{APNN} 顆二元樹的流程圖	2-19
圖 2- 17 表示提升人工校正相似影像集分組效率	2-21
圖 2- 18 RESNET 架構.....	2-24
圖 2- 19 RESNET-V1 與 RESNET-V2 殘差模塊中處理過程差異比較	2-25
圖 2- 20 降低人工校正相似影像分組所需時間及提升 ARR 正確率之 流程圖	2-29
圖 2- 21 CASE1 實驗流程	2-30
圖 2- 22 表 Γ_{D_1} 與 Γ_{D_2} 影像數量對組數的統計直方圖	2-32
圖 2- 23 CASE2 實驗流程	2-38
圖 2- 24 CASE 3 實驗流程.....	2-42
圖 2- 25 本實驗人工所收集影像類別.....	2-44
圖 2- 26 拍攝類別中訓練集(TRAIN)及驗證集(VVALIDATION)，訓練過程 中訓練效能(LOSS)的變化。.....	2-46
圖 2- 27 CCTV 災害類別中訓練集(TRAIN)及驗證集(VVALIDATION)，訓 練過程中訓練效能(LOSS)的變化。	2-50
圖 2- 28 智慧型影像檢索系統建置成本分析	2-56
圖 2- 29 以圖找圖影像檢索功能已發展於各式載具，以手機說明其概 念.....	2-57
圖 2- 30 本計畫已發展出跨平台以圖找圖檢索行動化模組之系統流程	2-59
圖 2- 31 電腦版本的使用者操作頁面流程	2-60

圖 2- 32 IPAD 版本的使用者操作頁面流程	2-61
圖 2- 33 手機版本的使用者操作頁面流程	2-61
圖 2- 34 舉一例說明交互查詢檢索成果	2-62
圖 2- 36 (A)收集影像;(B)人工標註各實例;(C) MASK RCNN 輸出實例 結果;(D) 輸出水視率及綠視率估計值。	2-64
圖 2- 37 估計 GLR 與 WLR 的流程	2-65
圖 2- 38 計算 LCLASS, LBOX, LMASK 架構圖	2-67
圖 2- 39 MASK RCNN 架構圖 (圖來源[30]).....	2-69
圖 2- 40 (A)訓練過程中針對訓練影像集之總 LOSS, L(LOSS 表示)的情 形, X 軸為 EPOCHS 個數; (B)-(D)分別為三項 LCLASS (MRCNN_CLASS_LOSS 表示), LBOX (MRCNN_BBOX_LOSS 表示), LMASK(MRCNN_MASK_LOSS 表示)的情況。	2-71
圖 2- 41 (A)訓練過程中針對驗證影像集總 LOSS, L(VAL_LOSS 表示)的 情形, X 軸為 EPOCHS 個數; (B)-(D)分別為三項 LCLASS(VAL_MRCNN_CLASS_LOSS 表 示), LBOX(VAL_MRCNN_BBOX_LOSS), LMASK(VAL_MRCNN_MASK_LOSS 表示)的情況。	2-72
圖附錄 3-1 案例 1.....	附錄-9
圖附錄 3-2 案例 2.....	附錄-9
圖附錄 3-3 案例 3.....	附錄-10
圖附錄 3-4 案例 4.....	附錄-10
圖附錄 3-5 案例 5.....	附錄-11
圖附錄 5-1 操作畫面 1.....	附錄-23
圖附錄 5-2 操作畫面 2.....	附錄-24
圖附錄 5-3 操作畫面 3	附錄-24

圖附錄 5-4 操作畫面 4.....	附錄-25
圖附錄 5-5 操作畫面 5.....	附錄-25
圖附錄 5-6 操作畫面 6.....	附錄-26
圖附錄 5-7 操作畫面 7.....	附錄-26
圖附錄 5-8 操作畫面 8.....	附錄-27
圖附錄 5-9 操作畫面 9.....	附錄-27
圖附錄 5-10 操作畫面 10.....	附錄-28
圖附錄 5-11 操作畫面 11.....	附錄-28
圖附錄 5-12 操作畫面 12.....	附錄-29
圖附錄 5-13 操作畫面 13.....	附錄-29
圖附錄 5-14 操作畫面 14.....	附錄-30

第一章 計畫內容

第一節 計畫動機

由於衛星遙測、Google MAP、UAV 照片、監控攝影機串流、智慧型手機等等技術快速地進步，因此，愈來愈普遍地可快速地取得影像照片，農業水土保持相關影像照片逐年地爆炸性地增加及儲存，使用者面臨一個大問題就是如何能快速地尋找到所需的影像照片？早期做法是用人力輸入文字註解方式來註記或描述照片，然後利用文字比對來搜尋相關的所需的影像照片，可是面臨爆炸性地增加影像照片情況下，用人力輸入文字註解方式是無法適用現行大量影像的狀況。也就是說，針對影像內容相似歷史影像查詢，單用文字檢索方式有下列缺點，耗費人力、輸入註解需大量時間、人力輸入有錯誤，不能精確地查詢出內容相似的影像[2, 22]。因此，本計畫首先發展以圖找圖技術可以來解決上述的問題，本計畫已建置出基於 IRP 影像的影像內容檢索系統，可輔助使用者快速地且自動化地查詢出影像內容類似之照片(影像)。即，當使用者輸入一張查詢影像(query image)進入系統搜尋檢索 IRP 現有影像資料庫，即可輸出一些影像其內容非常類似於查詢影像內容。此系統可輔助水土保持局已經建置 IRP，提供另一項輔助功能-此系統自動化檢索內容相關的圖片，然後，再以此檢索出影像來查詢 IRP，得到其對應的文字註解。此研究結果，除可讓使用者更方便了解影像的文字註解，亦可輔助使用者進一步核對其文字註解作業的正確性。對 IRP 有增值應用績效。本計畫也已發展在跨平台行動載具上執行上述影像檢索與顯示其對應文字註解等功能。

另外，本計畫也建置出智慧型影像分類技術可自動化判別影像內容屬於土石流、崩塌等不同災害類型樣態，接著，本計畫已發展影像物件識別系統，針對水土保持影像內容分析，可萃取影像中的物件與識別其物件類型，以上三項研究成果將可應用於水土保持相關業務之分類、辨識、預測、工程照片美學內涵分析之量化估計等相關事項[1, 3, 4, 6, 26, 27]。最後，本計畫已發展行動載具上執行影像檢索技術[18,25]。

第二節 計畫目的

本年度計畫目標為完成下列工作項目。

- 建置跨平台行動載具執行影像檢索:基於 IRP 影像搜尋系統(以水土保持局建置歷史影像平台(IRP)蒐集之歷年土砂災害照片為範圍)。圖 1-1 說明跨平台行動載具執行影像檢索的系統架構圖。

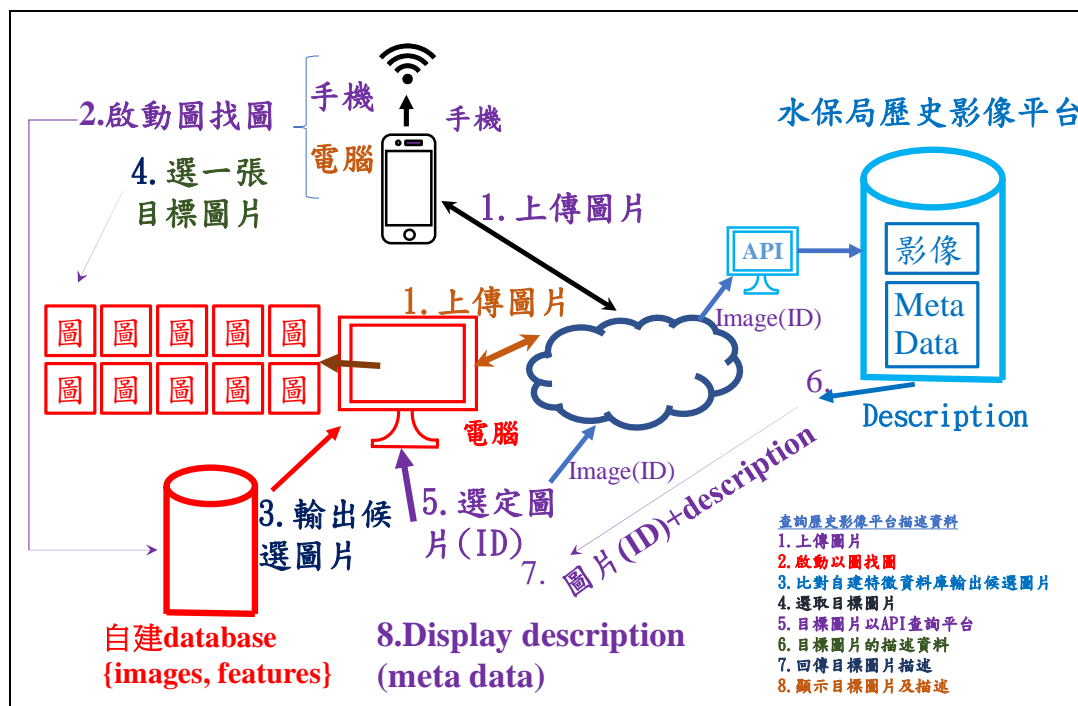


圖 1-1 跨平台行動載具執行影像檢索的系統架構圖

圖 1-1 使用情境流程步驟如下所述。

- 步驟1: 使用各種平台(電腦或行動載具)開啟網頁上傳查詢影像。
- 步驟2: 將上傳影像比對至影像資料庫中相似內容影像。
- 步驟3: 比對影像後將相似影像輸出，以網頁方式呈現(可跨平台)。
- 步驟4: 在網頁中點選目標影像。
- 步驟5: 透過 IRP 提供 API，將查詢目標影像的文字註解描述資訊。
- 步驟6: 從影像平台取得影像文字註解描述資訊。

步驟7: 將取得的影像資訊回傳至網頁中。

步驟8: 網頁顯示目標影像描述，如：行政區、拍攝時間、影像敘述等。

- 發展智慧型影像辨識技術: 可自動化判別 CCTV、高空攝影影像類別，接著，針對 CCTV 影像來自動化判別土石流、崩塌(不同構造物崩塌)等不同災害類型(以水土保持局建置歷史影像平台蒐集之歷年土砂災害照片為範圍)。訓練卷積類神經網路來實作智慧型分類器，完成上述影像內容之判別分類。

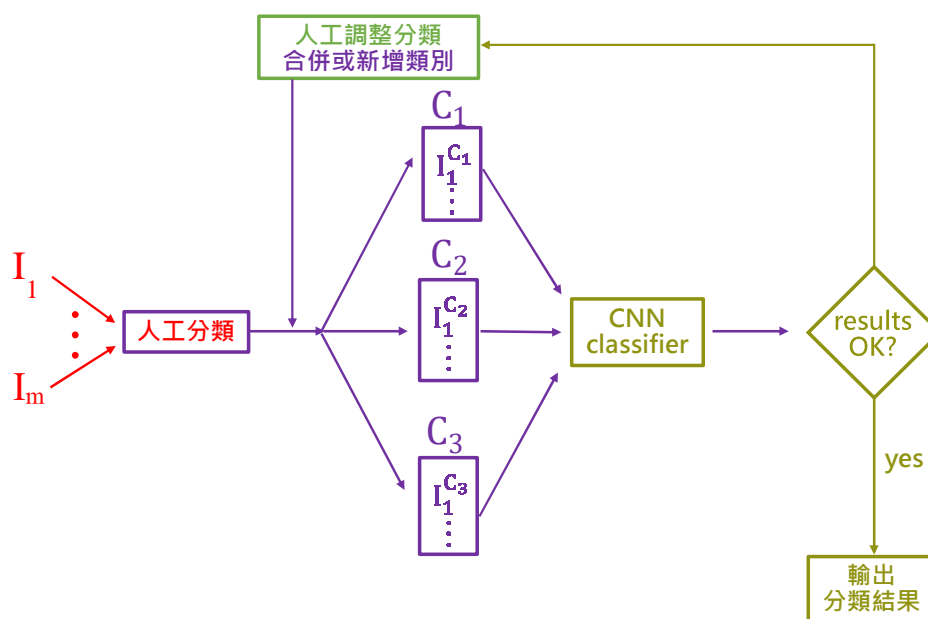


圖 1-2 智慧型影像分類判別技術架構圖

- 發展智慧型影像內容分析技術: 利用 Mask RCNN 實作實例分割 (instance segmentation) [9]，圖 1-3 運用 Mask RCNN 來實施實例分割流程。首先，針對照片中的各實例(物件)進行標記(label)，

然後，利用 RPN 來估計每個實例的錨框(anchor)，接著，進行實例分割及實例識別，最後計算，綠元素實例及水元素實例各佔整張影像的比例。

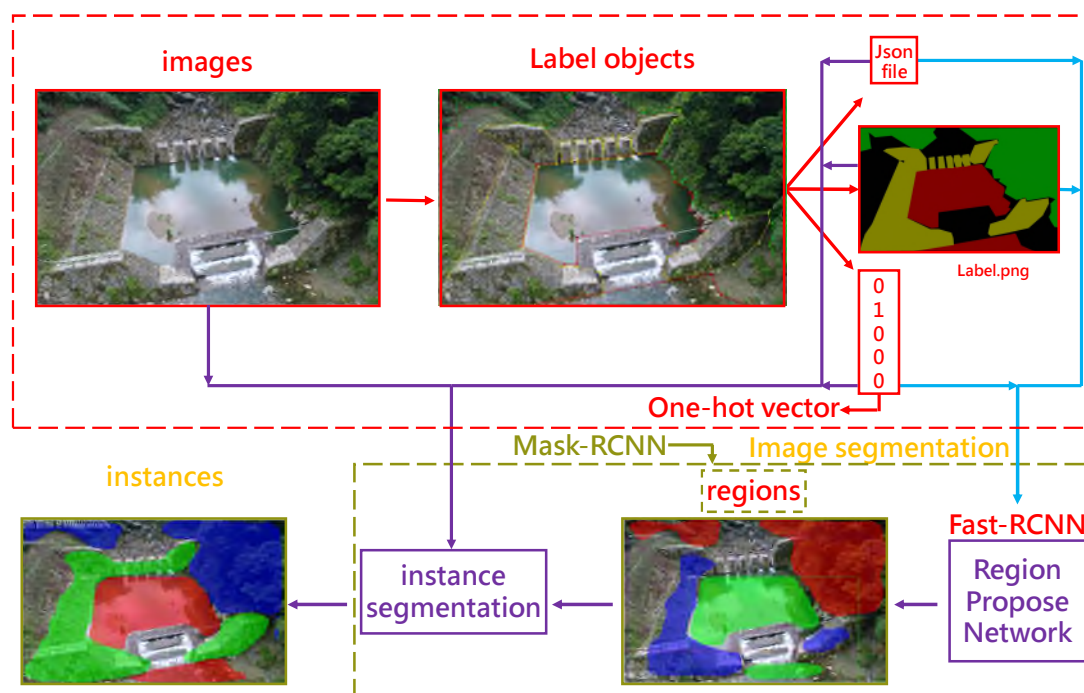


圖 1-3 運用 Mask RCNN 來實施實例分割流程

第三節 背景

一、影像檢索及分類

目前影像檢索已經用在許多領域[7, 16, 17, 24]。在[24]論文中，提出一種技術基於 CBIR 的特徵來研究影像檢索的技術，此種技術說明如下，步驟一：假設檢索域類似於用於訓練 CNN 模型的原始數據集，將新數據輸入預先訓練的模型取得最後三層激活值，步驟二：使用距離度量學習或相似性學習算法，可伸縮圖像相似性學習的在線算法 Online Algorithm for Scalable Image Similarity Learning (“OASIS”)，步驟三：重新調整訓練 CNN 模型，分別為使用額外註解資訊以及使用類標籤進行優化。

在[7]論文中，提出 NSVMBCNN(Novel SVM-Based Convolutional Neural Network)分類器，可以將紋理特徵(GLCM, NSF)得準確率提高到最佳結果，使用顏色特徵 (IIBCF) 和紋理特徵(GLCM)提取有效特徵，利用 K-均值算法進行分群完成分類標籤。

在[16]論文中，基於 CBIR 技術整理常使用的特徵用於影像檢索，如：視覺、顏色、形狀、紋理、空間等特徵，在提取特徵前需做影像前處理，如 RGB 轉 YCbCr 顏色空間、局部基本圖案(LBP)、顏色特徵(CIF)及計算特徵向量的相似度方式等。

二、影像切割

在[9]論文中提到，提出一種技術為像素導向分割(pixel-wise segmentation)此模組稱為 Mask RCNN，Decoder 可以將低維度影像特

徵向量對應到輸入的影像解析度的影像特徵向量，來做像素導向分類。主要用來理解道路和室內場景的有效架構，能夠對道路場景中的較小類別進行切割，同時還能對整個影像進行平滑分割，取得室內場景中具有各種形狀和大小進行細分，此種影像分割稱為實例分割(instance segmentation)[30]-[32]。圖 1-4 表示 Mask RCNN 可以實施實例分割，將每個氣球分割出來，每個汽球用不同顏色來區分，即，找到氣球在影像中的位置，且標示出其輪廓。每個氣球的虛線外框稱之為錨框(anchor)。圖 1-5 說明 Mask RCNN 的架構，包含兩個主要的單元，Fast RCNN 與實例分割。首先，用 Fast RCNN 來產出每個實例之錨框。接著，每個錨框再做物件識別及實例分割。

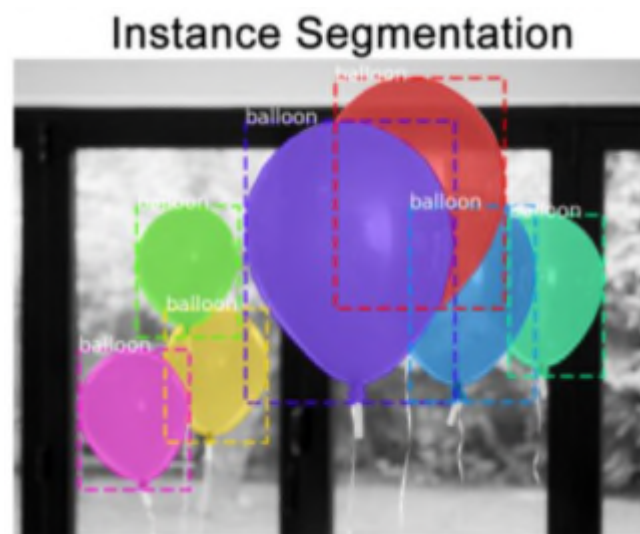


圖 1-4 實施實例分割，將每個氣球輪廓分割出來，每個汽球用不同顏色來區分。

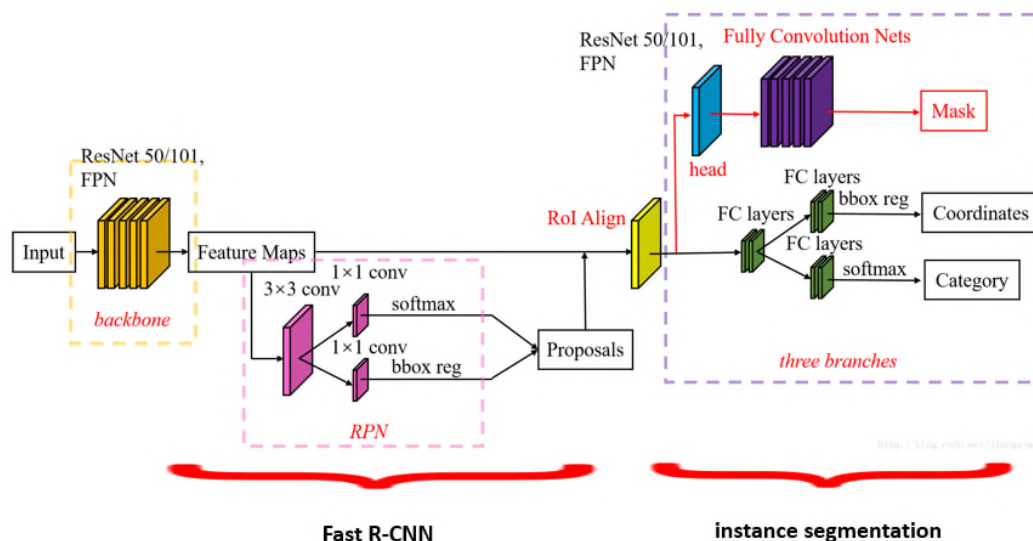


圖 1-5 Mask RCNN 架構圖 (圖片來源: [27])

三、深度卷積編碼器-解碼器

在[28]中，提到放射學中的最新應用，分為以下幾類：分類、分割、檢測和其他等。分類工作為透過相應標籤訓練數據訓練模型，利用醫學影像中描述的目標病變，並且將這些病變分為兩類或更多類。分割工作是使用 CNN 分類器用於計算器官或解剖結構的概率的公式的結構。偵測工作為放射科醫生的常見任務是檢測異常在醫學圖像中，其他為使用 CNN 實現 Encoder-Decoder 的架構對影像進行去除雜訊。在本計畫中訓練 Deep Convolutional Encoder-Decoder (DCED) 擷取影像特徵，再設計影像檢索技術。

DCED 是多層神經網路的一種非監督式學習算法，主要用於資料分類、視覺化。首先，DCED 架構可分成 Encoder(編碼器)和 Decoder(解碼器)兩部分，如圖 1-6 所示。

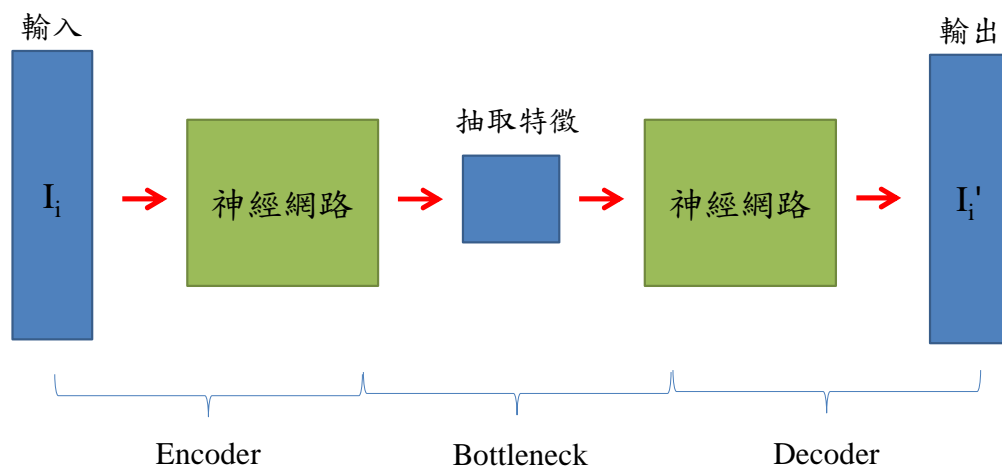


圖 1-6 DCED 架構

圖 1-7 表示 Encoder 的方法可以將資料壓縮，再透過 Decoder 的方式還原資料，壓縮的過程就是用更精簡的方式保存了資料。將影像 I_i 輸入到 Encoder-Decoder 模型中，透過 Encoder 壓縮成維度 γ 的特徵向量，再透過 Decoder 將維度 γ 的特徵向量，還原成原來高維度的影像 I_i' ，而 Decoder 所輸出的影像 I_i' 會與原影像 I_i 有差異，所以 Encoder-Decoder 模型的設計要盡可能降低輸入影像 I_i 及輸出影像 I_i' 的差異，假設有 n 張影像， I_i 與 I_i' 誤差公式為：

$$\text{LOSS} = \sum_{i=1}^n (I_i - I_i')^2$$

Encoder-Decoder 模型設計，最小化 LOSS，即：

$$\text{Minimize LOSS} = \sum_{i=1}^n (I_i - I_i')^2$$

最小化 LOSS 的工作即計算出權重 W_1 、 W_2 。

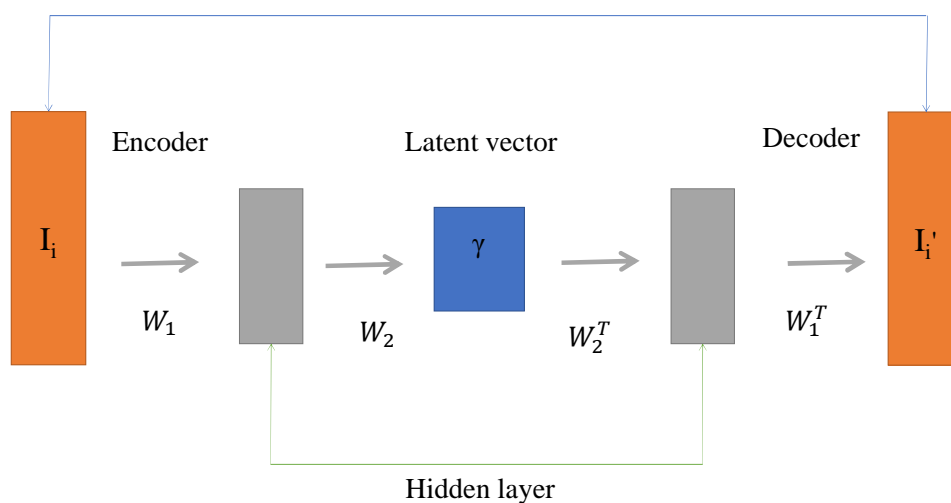


圖 1- 7 Encoder-Decoder 模型，hidden layer 用 CNN 來設計

四、Scikit-image

在[23]中提到，scikit-image 將圖像表示為一個標準 NumPy 數組，每個陣列都有一個維度，例如 2 個用於 2D 灰度圖像或 (M, N, 3) 具有 M 個垂直像素和 N 個水平像素的 RGB 彩色圖像，其中包括許多影像處理演算法，從視覺到醫學圖像分析的整個圖像處理研究中都有廣泛的應用。

五、影像檢索及影像分類量化指標

混淆矩陣(Confusion matrix)可用於二元分類及多元分類之效能評估指標，針對二元分類簡述如下。

表 1-1 二元分類混淆矩陣

		真實情況	
		Positive	Negative
預測結果	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

表 xx 為二元分類混淆矩陣，其中 TP、TN、FP、FN 介紹如下。

- True Positive (TP) 為真實情況為「Positive」，分類模型預測結果為「Positive」的個數
- True Negative (TN) 為真實情況為「Negative」，分類模型預測結果為「Negative」的個數
- False Positive (FP) 為真實情況為「Negative」，分類模型預測結果為「Positive」的個數
- False Negative (FN) 為真實情況為「Positive」，分類模型預測結果為「Negative」的個數

多分類情況的混淆矩陣類似二分類的混淆矩陣，如表1-1，對角線上的個數為分類模型預測結果等同於真實情況的結果。有關分類及檢索常用的量化指標有 Accuracy Rate、Precision Rate、Recall Rate、Average Recall Rate，其公式如下所示：

- Accuracy Rate (AR) $= \frac{TP+TN}{TP+TN+FP+FN}$

• Precision Rate (PR) = $\frac{TP}{TP+FP}$

• Recall Rate (RR) = $\frac{TP}{TP+FN}$

• Average Recall Rate (ARR) = $\frac{\sum_{i=1}^m RR_i}{m}$

第二章 工作項目與執行方法

第一節 工作項目

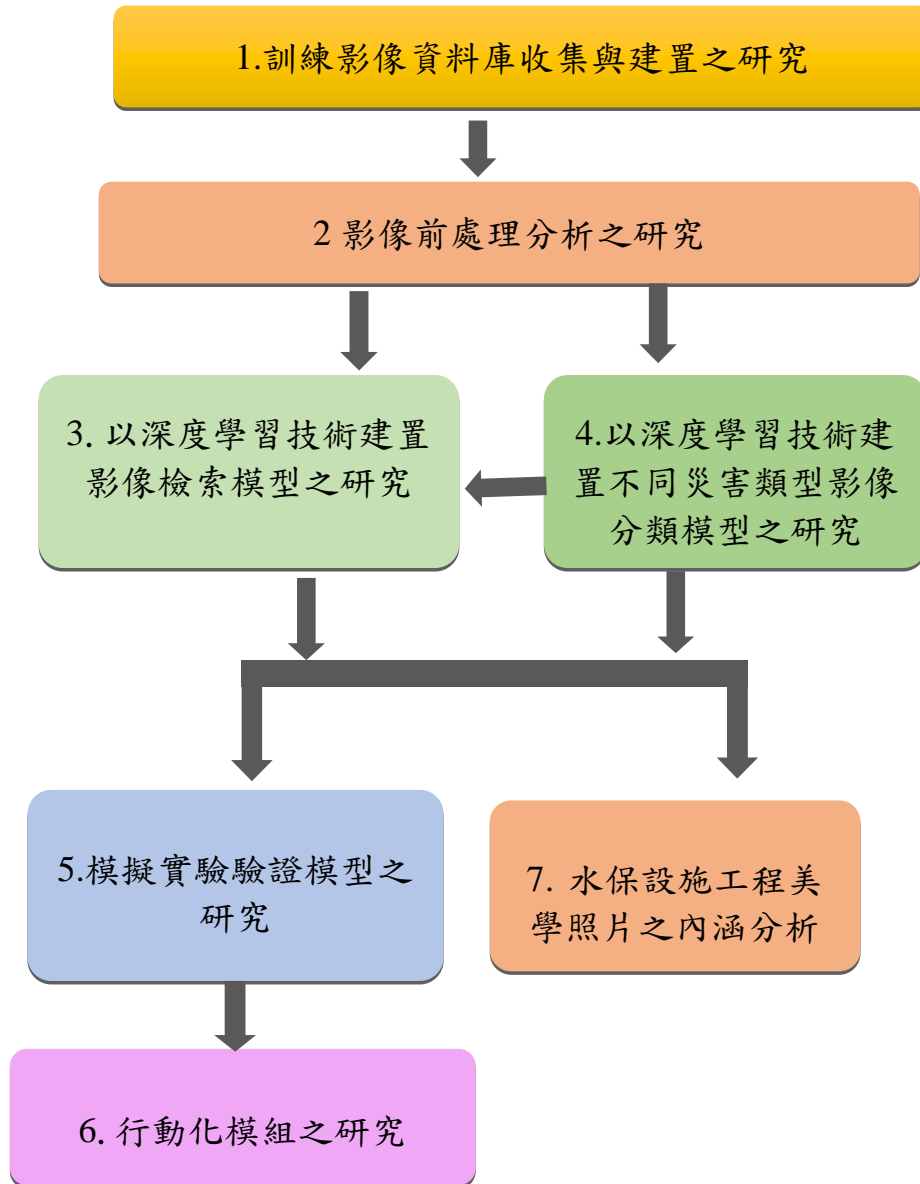


圖 2-1 本計畫實施方法構想圖

上圖為本計畫實施方法的構想圖，共有七項主題，其關聯如上圖。

本計畫欲研究七項主題將細述於第二節。

第二節 工作方法

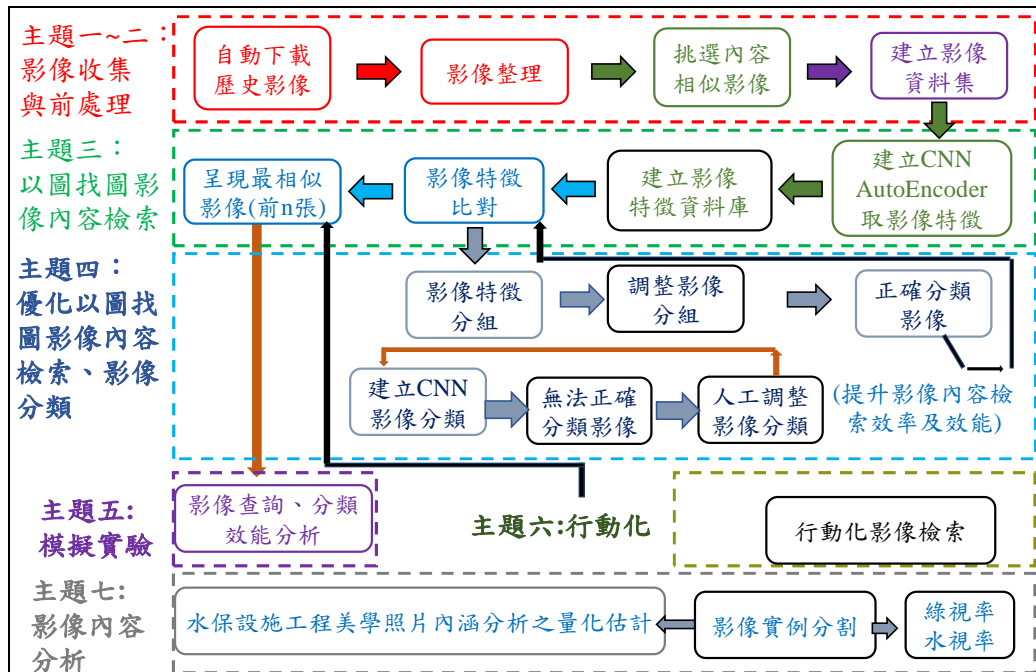


圖 2-2 本計畫系統架構含七項主題與各議題間的關聯流程圖

今年各主題的說明如下所述。上圖說明各主題間與各議題間的關聯流程圖。

主題一：訓練影像資料庫收集與建置之研究

主題二：影像前處理分析之研究

主題三：以深度學習技術建置影像檢索模型之研究

主題四：以深度學習技術建置不同災害類型影像分類模型之研究

主題五：模擬實驗驗證模型之研究

主題六：行動化影像檢索模組之研究

主題七：水保設施工程美學照片內涵分析之量化估計

第三節 工作執行步驟

一、訓練影像資料庫收集與建置之研究

本主題由 WWW 網路利用爬蟲機器人程式取得水土保持局開放水土保持相關影像資料庫蒐集水土保持相關影像建置資料庫，接著透過 YOLO 物件偵測技術偵測人形及人臉並設定一閾值(threshold)，若是人形及人臉在整張圖片內容所佔比例過高則會移除下載影像，將再建置篩選後資料庫。因水土保持局對於影像的描述可能會對相同的災害情況以不同的文字表達，無一定規範，使用關鍵字來進行影像分類，難以精確的分類，因此，本計畫將以人工比對，來挑選極相似災害影像內容進行災害分類及分群為本議題對影像的第一階段分類及分群，作為主題三及四的發展與效能測試使用。

(一)議題 1-1:自動下載歷史影像

本議題已發展自動下載模組，可下載 IRP 中收集影像。首先，建置水土保持相關影像 CSV 檔案，欄位包括事件 ID、經緯度、影像角度、縣市名稱、城鎮名稱、影像日期、地標、災害描述、影像網址，圖 2-3 所示說明 IRP 的 API。本議題已發展自動下載模組可透過 WWW 網路，發展爬蟲機器人程式，從 <https://photo.swcb.gov.tw/Repository/Database>，IRP 下載影像儲存至本地資料庫，影像收集於資料庫 D₀。相關技術參考[15]。

(二)議題 1-2: 建置本地端的影像資料庫

目前，本計畫針對目前 IRP 所提供之災害相關影像檔。在本地端建置影像資料庫，影像檔案命名方式，採用 IRP 所提供災害影像照片詮釋資料(meta data)內 PhotoType:照片分類及 EventID:照片識別碼。例如: type_0_EventID_635787957042643110。如圖 2-4 所示。



圖 2- 3 IRP 影像 API

type_0_EventID_636196609310965001

PhotoType:照片分類 EventID:照片識別碼

圖 2- 4 IRP 災害影像命名方式

二、影像前處理分析之研究

(一)議題 2-1:影像整理

本議題針對議題 1-1 中 IRP 歷史影像資料庫做初步分析。發現下列狀況，某些影像中包含人形、人臉等物件，這些物件會影響影像辨識及影像檢索的正確性。因此，本議題使用 YOLO 物件偵測技術(以卷積類神經網路發展出來的物件偵測)，來進行偵測人形及人臉，將由議題 1-1 IRP 歷史影像資料庫內，各影像中偵測人形及人臉，並以方框框選區塊[5, 13, 14, 19]。圖 2-5 說明偵測人形及人臉的結果。



圖 2-5 物件偵測技術偵測出人物移除這些物件

本計畫訓練之卷積類神經網路主要將災害視為重要的特徵，但對於人形、人臉標記等物件將視為雜訊，這些人形、人臉物件對於深度類神經網路來說可能會導致降低其效能，並可能使類神經網路誤判認為人物在該影像為重要的特徵。圖 2-6 說明本議題發展物件偵測技術偵測出人形及人臉，進而篩選適合影像的系統流程圖。

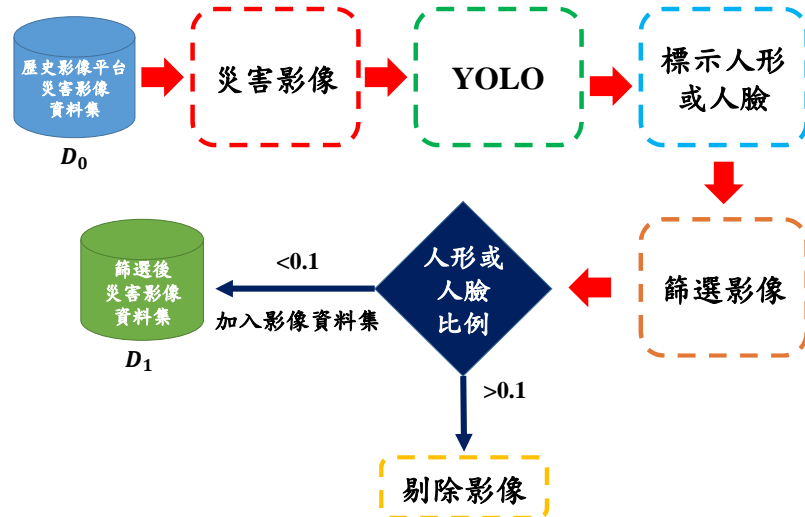


圖 2-6 物件偵測技術偵測出人形及人臉的流程

本議題中，物件偵測技術在偵測人形及人臉並設定一閾值 (threshold)，若是人形及人臉區域在整張影像所佔的比例 (region rate) 過高則會剔除影像。如下列公式所示，本議題用的 threshold 設為 0.1。

$$\text{region rate} = \frac{\text{人形及人臉區域}}{\text{整張影像}} > 0.1$$

將未被剔除影像收集於篩選資料庫 D_1 。

(二) 議題 2-2: 人工整理影像

本議題將**議題 1-2** 影像整理後各災害影像經由人工逐一比對與挑選，將相似災害影像進行影像分類及高相似影像內容之影像分群，接著，以程式將各個災害影像自動化分類資料夾名稱及災害影像檔名寫入 csv 檔，如圖 2-7 所示。分類資料夾名稱做為訓練類神經網路災害類別，災害影像檔名則是用於檢驗檢索出的災害影像相是否有檢索出由人工所挑選相似災害影像如圖 2-8 所示。本計畫預計實驗的測試影像數量說明如下。

- 透過**議題 1-1** 總共下載 23,440 張數收集於資料庫 D₀。
- 經由**議題 1-2** 物件偵測技術偵測圖片內人形或人臉篩選掉 289 張。
- 剩餘 23,151 張，收集於篩選資料庫 D₁。

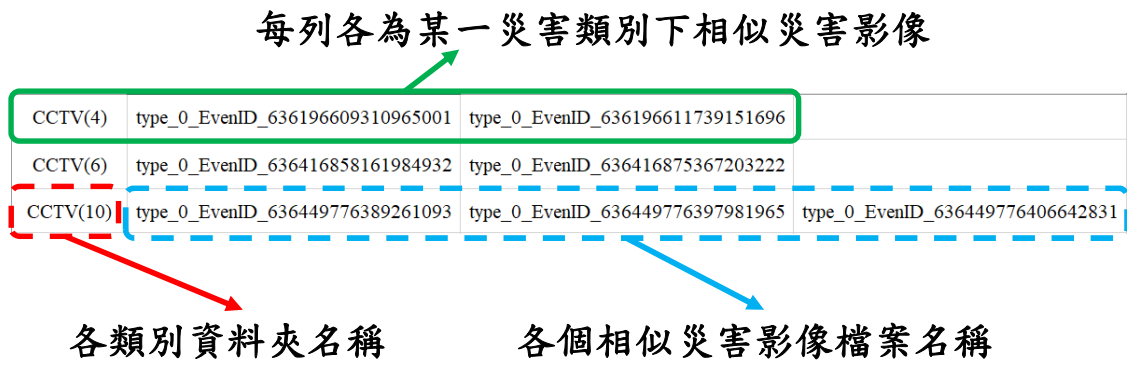


圖 2-7 各災害類別資料夾內相似災害影像檔案名稱

CCTV(4)	type_0_EventID_636196609310965001	type_0_EventID_636196611739151696	
			
CCTV(6)	type_0_EventID_636416858161984932	type_0_EventID_636416875367203222	
			
CCTV(10)	type_0_EventID_636416875367203222	type_0_EventID_636449776397981965	type_0_EventID_636449776406642831
			

圖 2-8 各災害類別資料夾內相似災害影像檔案名稱

三、以深度學習技術建置影像檢索模型之研究

本主題已發展以圖找圖的演算法模組來設計影像檢索技術，利用比對影像特徵相似度實現以圖找圖的功能。其流程簡述如下。首先，設計 AutoEncoder 來提取影像特徵(向量)，將欲查詢影像(query image)輸入至 AutoEncoder 並提取查詢特徵向量，並計算其特徵相似度，輸出影像比對後前 n_s 張影像。本主題採用最鄰近搜索法[29]，搜尋最接近查詢特徵向量的所有影像特徵向量。圖 2-9 說明以圖找圖影像檢索範例。

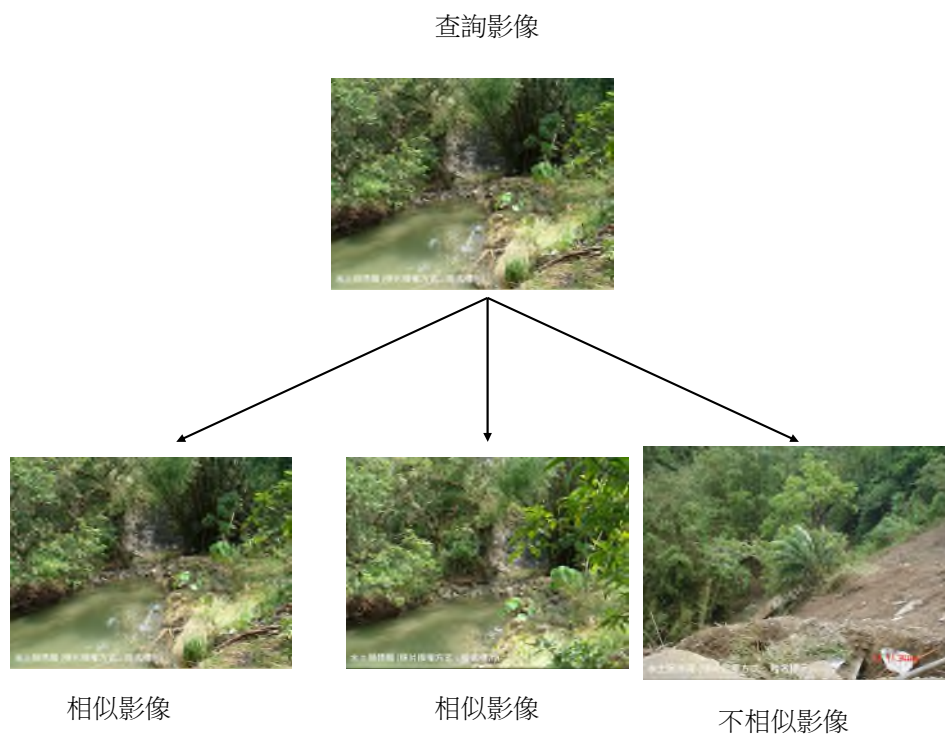


圖 2-9 圖找圖影像檢索範例

計算影像特徵向量是本主題的重點項目之一，因此，本主題建置深度類神經網路 AutoEncoder [8, 12, 18, 21]，此深度類神經網路會將輸入的影像轉換為特徵向量，此特徵向量會與相對應的影像存放至影像特徵資料庫，接者，已發展有效率的相似內容比對演算法，採用最鄰近搜索法，搜尋影像內容最接近查詢影像特徵前 n_s 張影像。

(一) 議題 3-1: 設計 DCED 之研究

圖 2-10~圖 2-11 舉例說明以 CNN 架構設計 DCED。主要是以 Encoder 與 Decoder 模組設計為主，詳細說明如下。

- Encoder 架構：用於提取影像特徵。其 CNN 模型由 11 層組成，其中 6 層是卷積層，五層是由池化層所組成，如圖 2-10 所示。輸入一張 3 維影像到模型中進行特徵提取，透過卷積層提取特徵，由池化層對影像進行降維，將影像特徵較重要的部分保留下來，減少模型參數的數量，減少過度擬合的情形。圖 2-10 中 32@256*256 表示影像大小為 256*256，共有 32 層。表 2-1 列出各層的影像大小(size of output images)及深度層數(即 number of filters)。
- Decoder: 利用影像特徵還原影像，其 CNN 模型由 11 層組成，其中 6 層是反卷積層，五層是由反池化層所組成，如圖 2-11。將 Encoder 所輸出影像特徵，透過反卷積層與反池

化層將特徵估計原影像，即求 Minimize Loss。盡量將影像特徵缺失的部分進行調整，最終可能得到的影像會因為一些細節無法恢復，導致與原影像中有些差異，此演算法屬於失真的資料壓縮類型。表 2-2 列出各層的影像大小及深度層數。

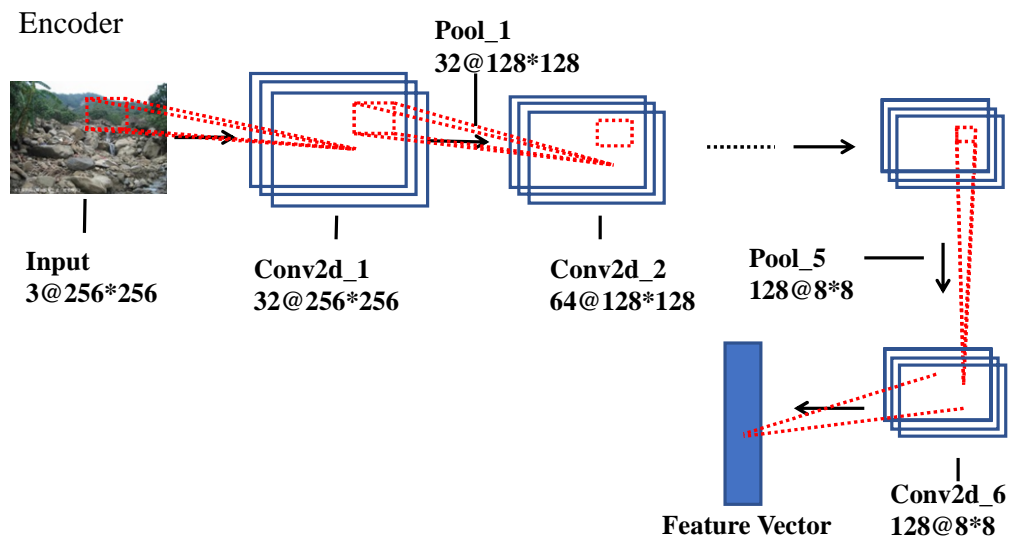


圖 2- 10 細部 CNN 架構設計 Encoder 擷取影像

表 2-1 各層的影響大小及深度層數

<i>Layer</i>	<i>Type</i>	<i>Number of Filters</i>	<i>size of Output Image</i>
0	Input layer	3	256*256
1	Convolution layer	32	256*256
2	MaxPooling layer	32	128*128
3	Convolution layer	64	128*128
4	MaxPooling layer	64	64*64
5	Convolution layer	64	64*64
6	MaxPooling layer	64	32*32
7	Convolution layer	128	32*32
8	MaxPooling layer	128	16*16
9	Convolution layer	128	16*16
10	MaxPooling layer	128	8*8
11	Convolution layer	128	8*8

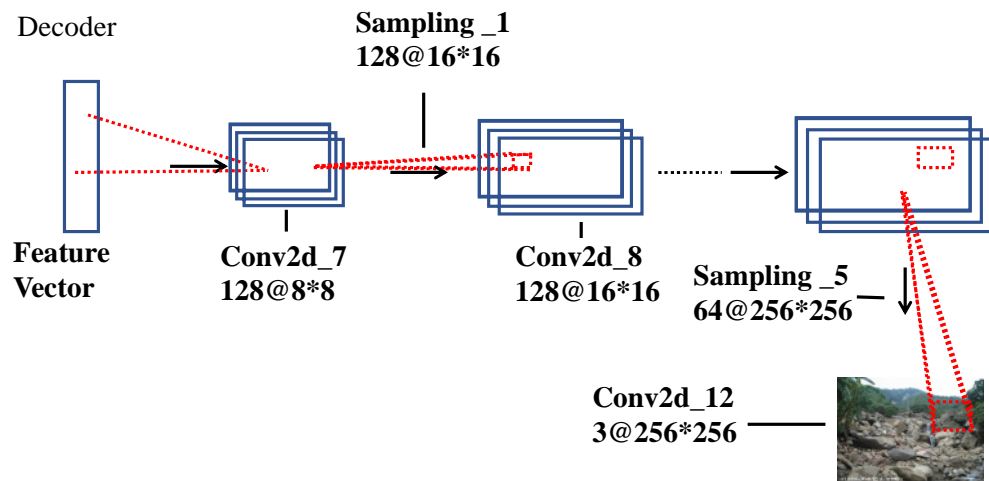


圖 2-11 細部 CNN 架構設計 Decoder，還原影像

表 2-2 各層의影像大小及深度層數

<i>Layer</i>	<i>Type</i>	<i>Number of Filters</i>	<i>Size of Output Image</i>
12	<i>Convolution layer</i>	128	8*8
13	<i>UpSampling layer</i>	128	16*16
14	<i>Convolution layer</i>	128	16*16
15	<i>UpSampling layer</i>	128	32*32
16	<i>Convolution layer</i>	128	32*32
17	<i>UpSampling layer</i>	128	64*64
18	<i>Convolution layer</i>	64	64*64
19	<i>UpSampling layer</i>	64	128*128
20	<i>Convolution layer</i>	64	128*128
21	<i>UpSampling layer</i>	64	256*256
22	<i>Convolution layer</i>	3	256*256

(二) 議題 3-2:AutoEncoder 提取影像特徵之研究

本議題將應用議題 3-1 結果，設計深度類神經網路來建置 AutoEncoder 稱為自動編碼器，AutoEncoder 須經由訓練，才能提取輸入影像較佳的特徵並轉換為特徵向量，較佳的特徵向量能獨特地表示該影像內容資訊，圖 2-12 說明設計概念圖，本議題將應用 CNN 設計 AutoEncoder 來擷取影像特徵。

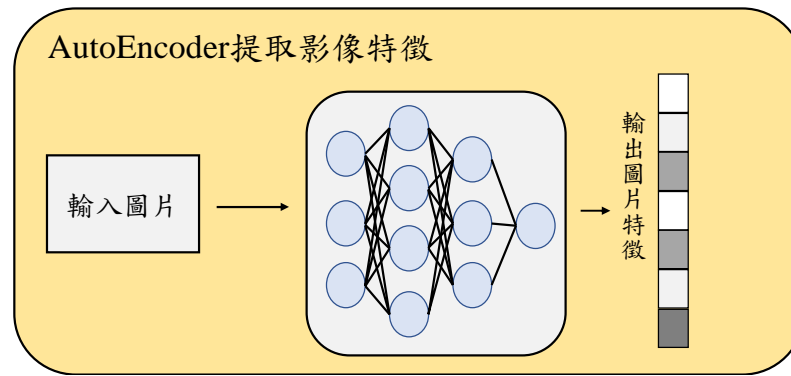


圖 2-12 基於 CNN 設計 AutoEncoder 概念圖，擷取影像特徵

本議題所設計的 AutoEncoder 的細部 CNN 架構，如圖 2-10 中的 Encoder，可用來擷取影像。其功能是将影像 I_k 映射至特徵空間，假設 m 維特徵空間，其映射行為如下列所示。

$$\begin{array}{ccc} \textit{image} & \xrightarrow{\quad\quad\quad} & \textit{feature} \\ I_k & & F_{I_k} \in [0,1]^m \end{array}$$

影像 I_k 經 AutoEncoder 映射至 $F_{I_k} \in [0,1]^m$ 。一般而言，1 維特徵空間的值可以被映射至 $[0, 1]$ ，這是一種標準化程序。即，影像 I_k 的特徵可標示為 F_{I_i} ，如下所示。

$$F_{I_i} = \{f_{1,I_i}, f_{2,I_i}, f_{3,I_i}, \dots, f_{m,I_i}\}, f_{j,I_i} \in [0,1]$$

(三) 議題 3-3:建置影像特徵資料集

本議題將應用議題 3-2 結果，將 D_1 中的每張影像 I_k 以 AutoEncoder 擷取其特徵向量 F_{I_k} ，將此特徵向量與影像以字典 (key, value) 形式儲存於影像特徵資料集 F_{D_1} 。圖 2-13 說明本議題將建置影像特徵資料庫。

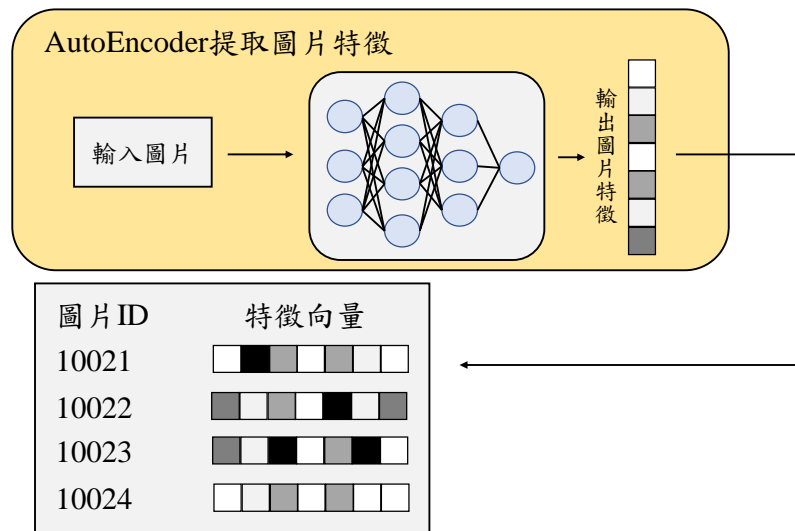


圖 2-13 建置影像特徵資料庫 F_{D_1}

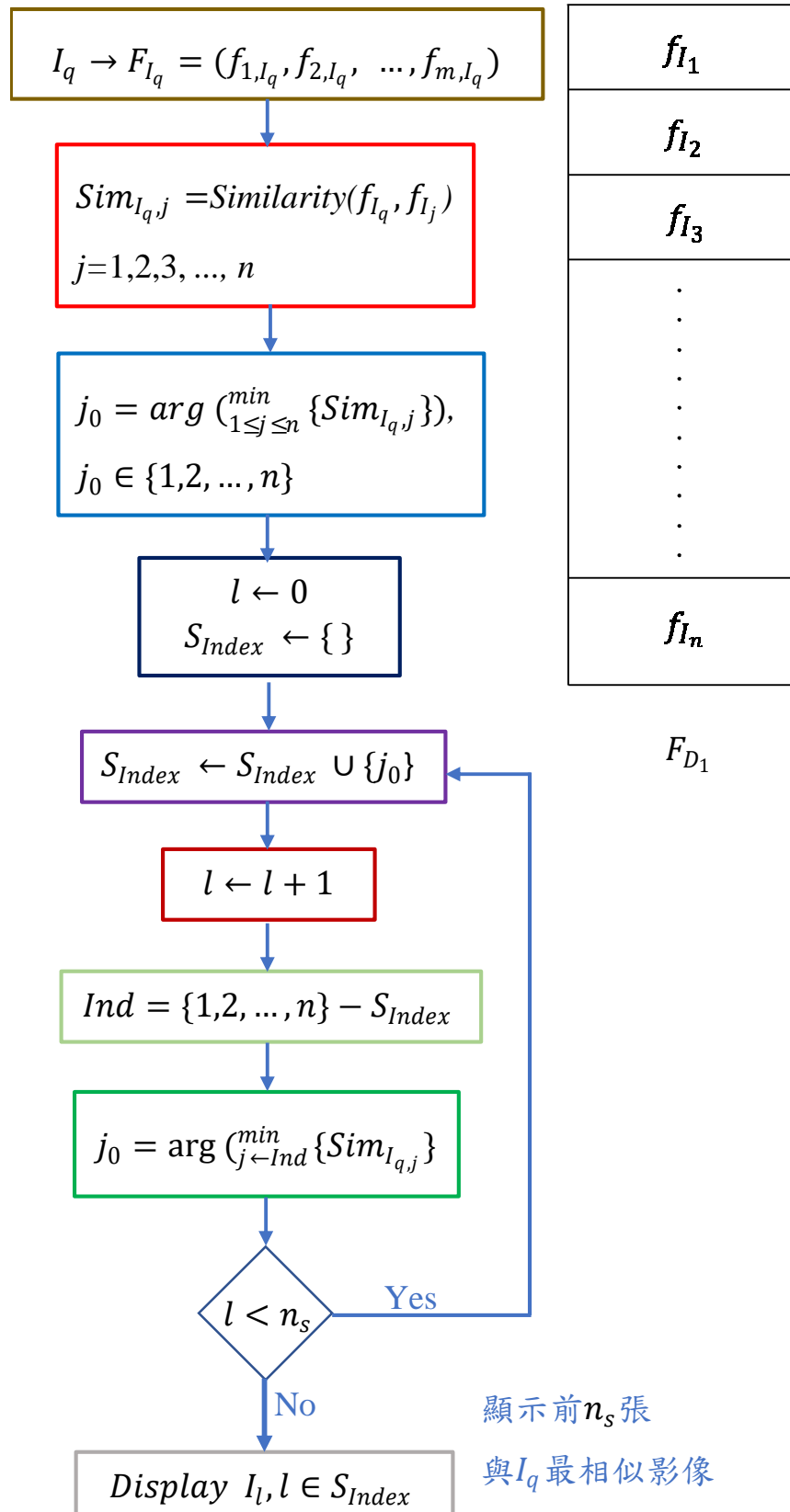


圖 2-14 影像特徵比對的演算法

(四) 議題 3-4: 影像特徵比對之研究

圖 2-14 說明影像特徵比對的演算法，令查詢影像 I_q 的特徵為 F_{I_q} ，重複比對 F_{I_q} 與 F_{D_1} 每個特徵。最後輸出前 n_s 張與 I_q 最相似影像。假設輸入是 n 張影像，這演算法的計算複雜度，在最差比對形況下是需要下列計算複雜度 $O(n)$ 。

$$g(n) = O(n_s n) \approx O(n) \text{ 因 } n \gg n_s \text{。}$$

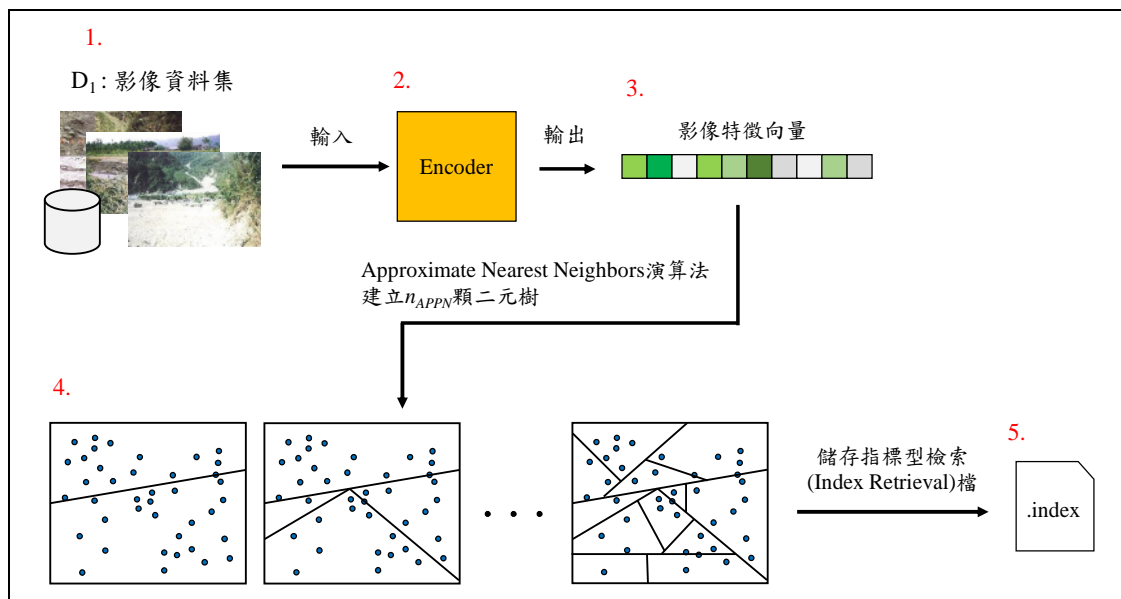


圖 2-15 將 F_{D_1} 特徵庫，以 APNN 演算法建立 n_{APNN} 顆二元樹的流程圖

(五) 議題 3-5: 提升影像特徵比對效率之研究

本議題嘗試提升議題 3-4 的影像特徵比對演算法的查詢效率，即加速查詢效率。因此，本計畫已使用 Approximate

Nearest Neighbors (APNN)演算法進行影像特徵向量比對 [29]。圖 2-15 說明建置二元樹的步驟流程。在進行影像特徵向量比對前，先將 D_1 影像集中的所有影像輸入至 Encoder，並輸出每一張影像的影像特徵向量，將所有影像特徵向量 F_{D_1} 映射至多維空間，以 APNN 演算法建立 n_{APNN} 顆二元樹，作法簡述於後。首先，隨機選取多維空間的兩個點，從這兩個點之間的連線畫一條法向量(建立一顆二元樹)，接著從分割後的空間再隨機尋找兩個點重複上述動作，執行 n_{APNN} 次(建立 n_{APNN} 顆二元樹)，最後，儲存成指標型檢索(Index Retrieval)檔。其演算法如下所示。

步驟 1：輸入影像資料庫 D_1

步驟 2：將影像資料庫 D_1 內所有影像輸入至 Encoder。

步驟 3：Encoder 輸出所有影像特徵向量資料庫 F_{D_1} 。

步驟 4：將 F_{D_1} 中所有影像特徵向量，使用 APNN 演算法建

立 n_{APNN} 顆二元樹，

4-1、首先找兩點，作出兩點連線之法向量，分成兩群向量。

4-2、再這兩群向量中，各自再分成兩群。建立出一顆 2 元樹。

4-3、重複遞迴式地用直線分割，繼續分割，直到建立

n_{APNN} 顆二元樹。

步驟 5：儲存成指標型檢索(Index Retrieval)檔，輸出指標

型檢索檔。

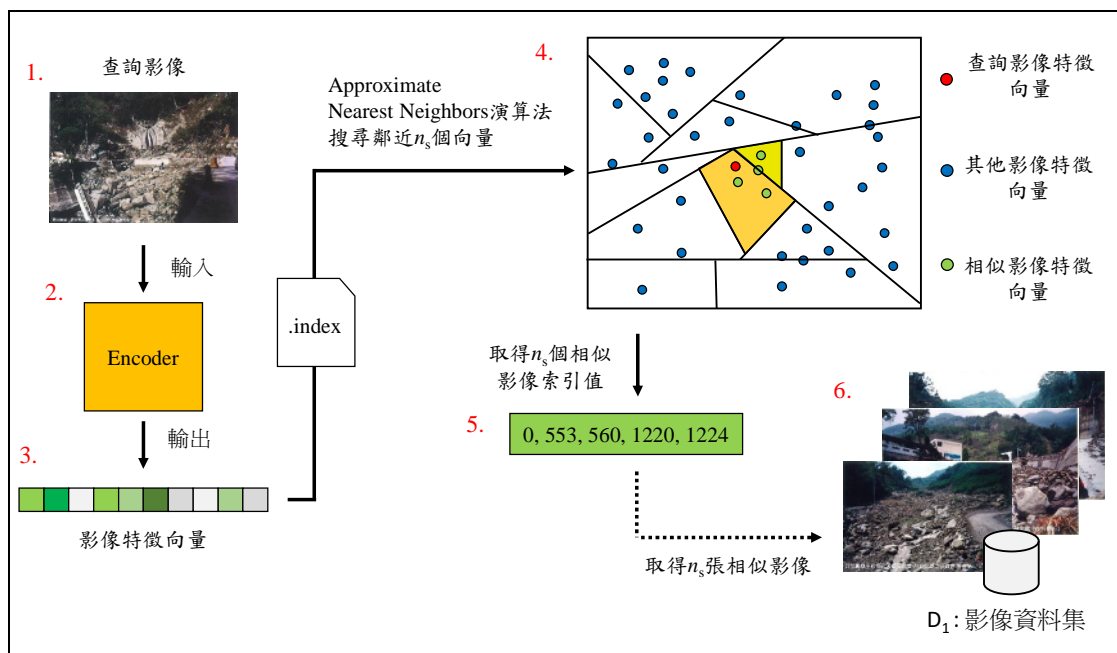


圖 2-16 將 F_{D_1} 特徵庫，以 APNN 演算法建立 n_{APNN} 顆二元樹的流程圖

當進行影像檢索時，輸入查詢影像 I_q 至經過影像訓練的 Encoder 並輸出該影像特徵向量 F_{I_q} ，以 F_{I_q} 在指標型檢索(Index Retrieval)檔中以 APNN 演算法檢索鄰近 n_s 個影像特徵向量，APNN 演算法檢索完成後會輸出該影像特徵向量的索引值，依照索引值，可從 D_1 取得對應的 n_s 張相似影像，即完成影像檢

索功能。其演算法如下所示

步驟 1：輸入查詢影像 I_q 至 Encoder。

步驟 2：Encoder 輸出該影像特徵向量 F_{I_q} 。

步驟 3： F_{I_q} 在指標型檢索(Index Retrieval)檔內進行 APNN

演算法，輸出 n_s 個相似影像索引值。

步驟 4：依 n_s 個相似影像索引值，至 D_1 中輸出對應 n_s 張相

似影像。

假設輸入是 n 張影像，APNN 演算法的計算複雜度，在最差比

對的形況下是需要下列計算複雜度 $O(\log n)$ 。

$$g(n) = O\left(\log \frac{n}{n_{APNN}}\right) \approx O(\log n) \text{ 因 } n \gg n_{APNN}$$

因此，本計畫用 APNN 演算法來檢索前 n 張相似影像內容，

必能提升**議題 3-4**的特徵比對演算法之查詢效率，因為

$O(\log n) \ll O(n)$ 。

(六) 議題 3-6: 提升人工校正相似影像集效率

本議題主要提升人工校正相似影像集分組效率，由於本計畫

沒有分組好的相似影像，所以必須由人工來挑出相似影像，

分組的相似影像將用於本計畫中測試影像檢索效能，以 ARR

為主。首次由人工挑出的分組相似影像可能有許多錯誤，如：

同組內仍有影像內容相似差異較大的問題存在，由於分組的組數頗多，可能上千組以上，若由人工重複檢視這些分組中的影像所需時間較多，所以本議題提出如圖 2-17 的流程，減少由人工需透過重複逐一檢視分組所需的時間。

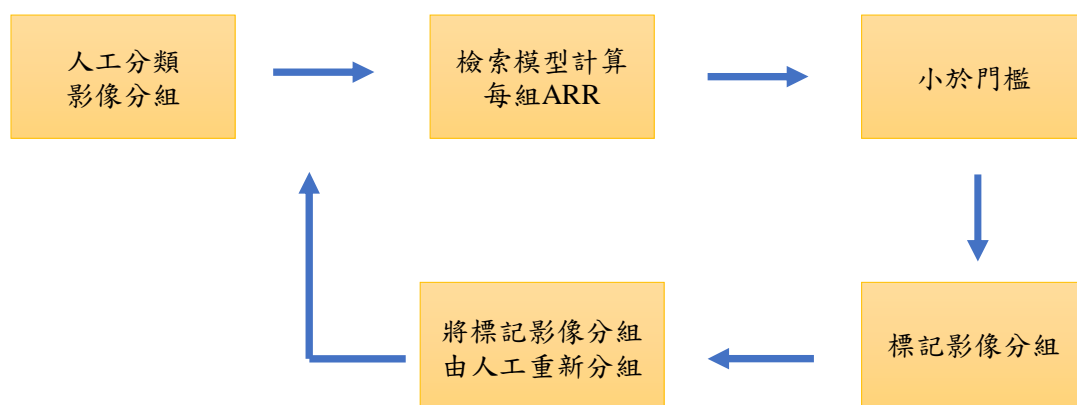


圖 2-17 表示提升人工校正相似影像集分組效率

四、以深度學習技術建置不同災害類型影像分類模型之研究

本主題發展智慧型影像分類技術，可自動化判別 CCTV、高空攝影影像類別，接著，針對 CCTV 影像來自動化判別土石流、崩塌(不同構造物崩塌)等不同災害類型(以水土保持局建置歷史影像平台蒐集之歷年土砂災害照片為範圍)。訓練卷積類神經網路來實作智慧型分類器，完成上述影像內容之判別分類。本主題設計 CNN 之深度類神經網路來發展智慧型影像分類，進行影像災害類型分類。將訓練影像資料集中影像輸入並進行訓練，以反向傳播方式調整類神經網路內的權重，凝合訓練影像。訓練完成後，即可提供給使用者進行影像災害類型分類。目前的問題是--IRP 並未提供每張影像的類別標籤，因此，本計畫嘗試人工判讀方式來分類影像(即，人工的方式對影像作觀察判斷，給予標籤類別)。藉由分類的效果做調整，新增類別或合併。

本主題採用深度殘差網路 (Deep Residual Networks, ResNet) [10, 11, 33, 34]來發展智慧型影像分類。其特色是在 plain 結構上引入殘差模塊 (residual block)的概念。在早期的 plain 結構中，期望深度層數越多，模型獲取影像特徵能力越強。然而隨層數到達某定值後，正確率呈現飽和狀態後，準確率開始下降，此問題稱為退化問題 (Degradation problem)，也就是說，訓練集和驗證集的 loss 隨者 epoch 的數量增多

而同時有上升的趨勢。因此，ResNet 透過學習殘差方式，而非去學習直接映射方式。就權重更新而言，不用主體部份的變化量調整權重，從而微小變化(變化量的變化)，得到較多的變化量來調整權重，可用來解決層數過多引發的退化問題。圖 2-18 說明 ResNet 架構。本計畫採用 ResNet-v2 [33]。其差別在於殘差模塊中的處理過程。ResNet-v2 將激活函數提前，使權重變化更加劇烈。圖 2-19 說明 ResNet-v1 與 ResNet-v2 殘差模塊中處理過程差異比較。表 2-3 列出 Resnet50v2 各層架構。

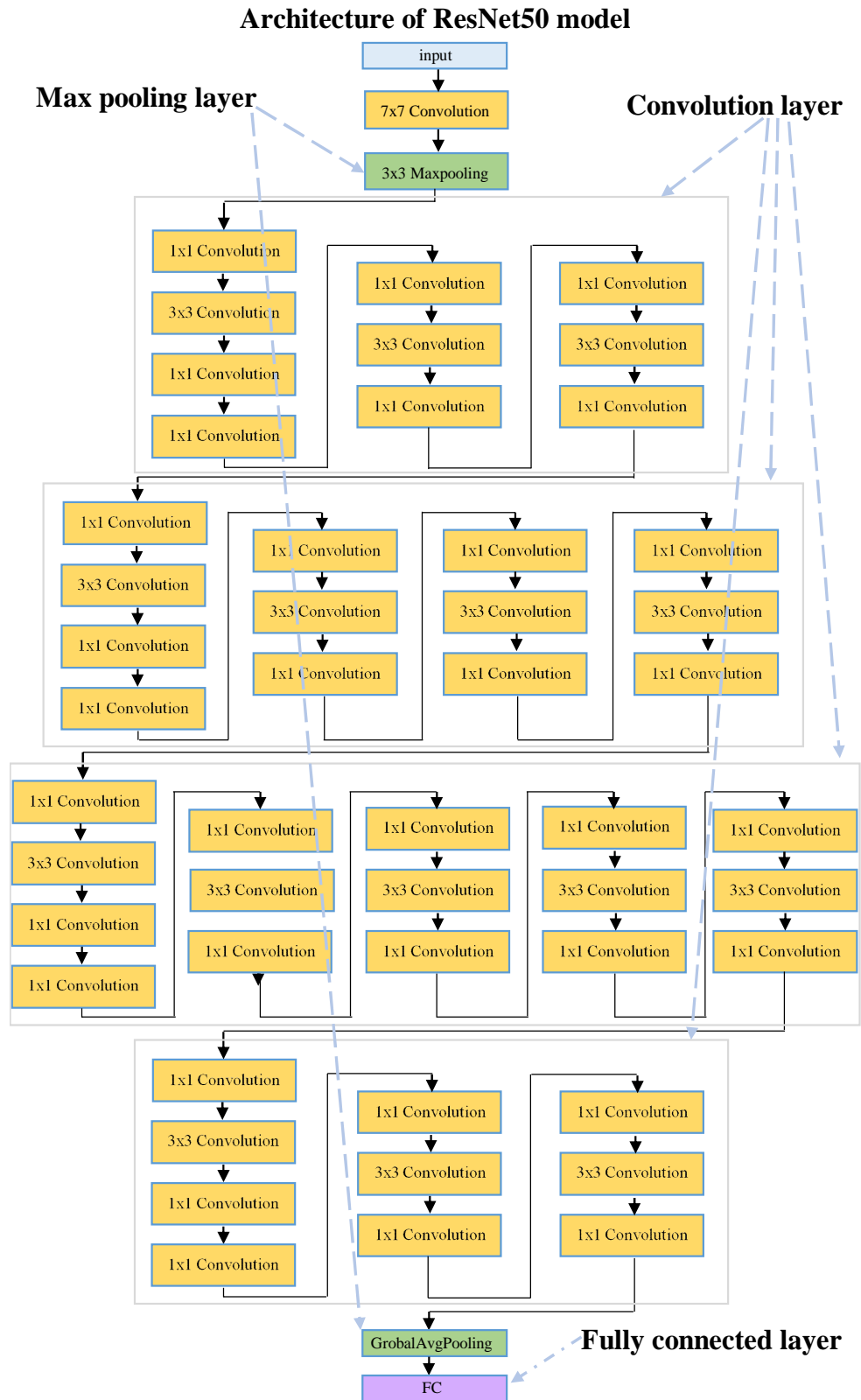


圖 2- 18 ResNet 架構

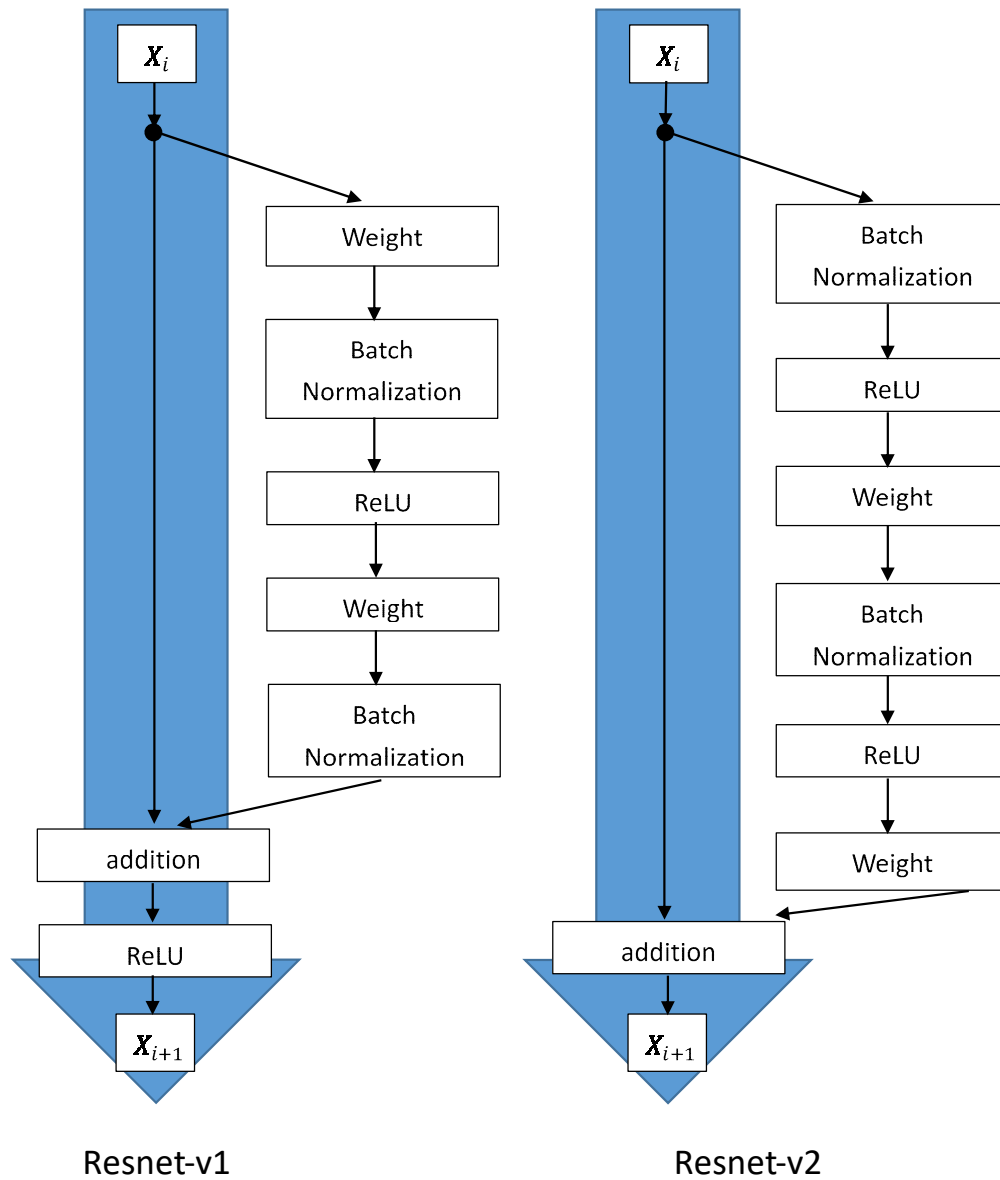


圖 2-19 ResNet-v1 與 ResNet-v2 殘差模塊中處理過程差異比較。

表 2- 3Resnet50v2 各層架構

Layer Name	Feature Map Output Size	Resnet50v2
Conv1	112×112	7×7 , 64, stride 2
Conv2_x	56×56	3×3 max pool, stride 2 $\begin{bmatrix} 1 \times 1,64 \\ 3 \times 3,64 \\ 1 \times 1,256 \end{bmatrix} \times 3$
Conv3_x	28×28	$\begin{bmatrix} 1 \times 1,128 \\ 3 \times 3,128 \\ 1 \times 1,512 \end{bmatrix} \times 4$
Conv4_x	14×14	$\begin{bmatrix} 1 \times 1,256 \\ 3 \times 3,256 \\ 1 \times 1,1024 \end{bmatrix} \times 6$
Conv5_x	7×7	$\begin{bmatrix} 1 \times 1,512 \\ 3 \times 3,512 \\ 1 \times 1,2048 \end{bmatrix} \times 3$
	1×1	Global Average Pool, FC

五、模擬實驗驗證模型之研究

本主題將以準確率(Accuracy)、召回率(Recall)、精確率(Precision)，做為評估分類模型分類能力的指標，並利用混淆矩陣來分析影像分類模型之分類能力效能。利用召回率(Recall)、精確率(Precision)估影像檢索模型之檢索能力。

(一) 議題 5-1: 模擬實驗分析智慧型以圖找圖演算法之檢索能力

本議題已完成電腦模擬實驗分析智慧型以圖找圖演算法之影像檢索效能，電腦模擬實驗步驟流程減述如下: 首先，利用議題 2-1 篩選後影像資料集 D_1 ，並將 D_1 以每 5000 張影像等切割成五個子影像資料集， $D_1 = \cup_{z=1}^5 D_1^z$ ，然後，議題 2-2 經由人工判斷極相似影像內容分組存檔以 $\Gamma_{D_1^z}$ 代表第 z 個子影像資料集 D_1^z 之分組集，此分組集 $\Gamma_{D_1^z}^A$ 定義如下，

$$\Gamma_{D_1^z} = \{I_{D_1^z}^{h,k} | h = 1, 2, \dots, n_{D_1^z}^{\text{row}}, k = 1, 2, \dots, n_{D_1^z}^{\text{row,col}}\}$$

分組集 $\Gamma_{D_1^z} = \cup_{h=1}^{n_{D_1^z}^{\text{row}}} \Gamma_{D_1^z}^h$ ，每組 $\Gamma_{D_1^z}^h$ 中各影像是人工判讀後極相似的影像內容，分組集 $\Gamma_{D_1^z}$ 共有 $n_{D_1^z}^{\text{row}}$ 組，每組 $\Gamma_{D_1^z}^h$ 內共有 $n_{D_1^z}^{\text{row,col}}$ 張影像， $I_{D_1^z}^{h,k}$ 表示在第 z 個子影像資料集 D_1^z 之分組集中，用人工判讀後極為相似影像記錄檔 $\Gamma_{D_1^z}$ 中第 h 列(此處的紀錄檔中列是

代表分組集中的組)上的第 k 張影像，以矩陣表示 $\Gamma_{D_1^z}$ 極相似影像分組記錄，矩陣表示式如下所示。

$$\begin{bmatrix} I_{D_1^z}^{1,1} & I_{D_1^z}^{1,2} & \dots & I_{D_1^z}^{1,n_{D_1^z}^{1,col}} \\ I_{D_1^z}^{2,1} & I_{D_1^z}^{2,2} & \dots & I_{D_1^z}^{2,n_{D_1^z}^{2,col}} \\ \vdots & \vdots & \ddots & \vdots \\ I_{D_1^z}^{n_{D_1^z}^{row},1} & I_{D_1^z}^{n_{D_1^z}^{row},2} & \dots & I_{D_1^z}^{n_{D_1^z}^{row},n_{D_1^z}^{row,col}} \end{bmatrix}$$

$\Gamma_{D_1^z}$ 極相似影像分組記錄表示

本議題已利用 $\Gamma_{D_1^z}$ 來計算平均召回率(Average Recall Rate, ARR)做為評估影像檢索模型能力指標，也提供平均檢索反應時間(Average Response Time, ART)。

本議題驗證運用**議題 3-6**的設計，的確可以減少由人工需透過重複逐一檢視全部分組所需的時間，也提升影像檢索的召回率。圖 2-20 表示降低人工校正相似影像分組所需時間及提升 ARR 正確率之流程圖，其步驟簡述如下：

步驟 1：人工分組影像資料庫 D_1^z 。

步驟 2：透過 AutoEncoder 進行影像檢索。

步驟 3：計算每組 ARR，並設置門檻值 h ，將高於門檻值 h

分成 $D_1^{z,1}$ ，低於門檻值 h 分成 $D_1^{z,2}$ 兩個影像資料集。

步驟 4：如果 $D_1^{z,2}$ 為空集合，則結束演算法。

步驟 5：如果 $D_1^{z,2}$ 不為空集合，則將 $D_1^{z,2}$ 影像資料庫重新進行人工分組。

步驟 6：將 $D_1^{z,1}$ 與新分組的 $D_1^{z,2}$ 聯集成新的 D_1^z 影像分組，跳至步驟 2。

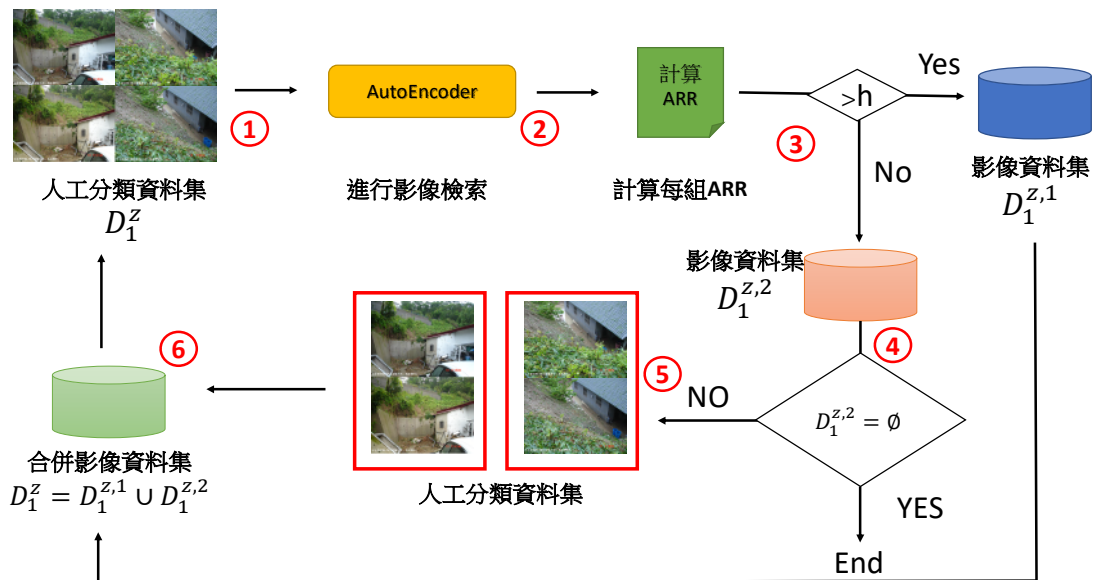


圖 2- 20 降低人工校正相似影像分組所需時間及提升 ARR 正確率之流程圖

本議題已設計模擬實驗，驗證影像檢索效能共有三 Cases，其做法依序簡述如下：

•Case 1: 使用 D_1 進行訓練 AutoEncoder 模型並透過相似影像內

容分組存檔 $\Gamma_{D_1^{1-2}}$ ，主要評估極相似影像內容被同時檢索出的

召回率效能。

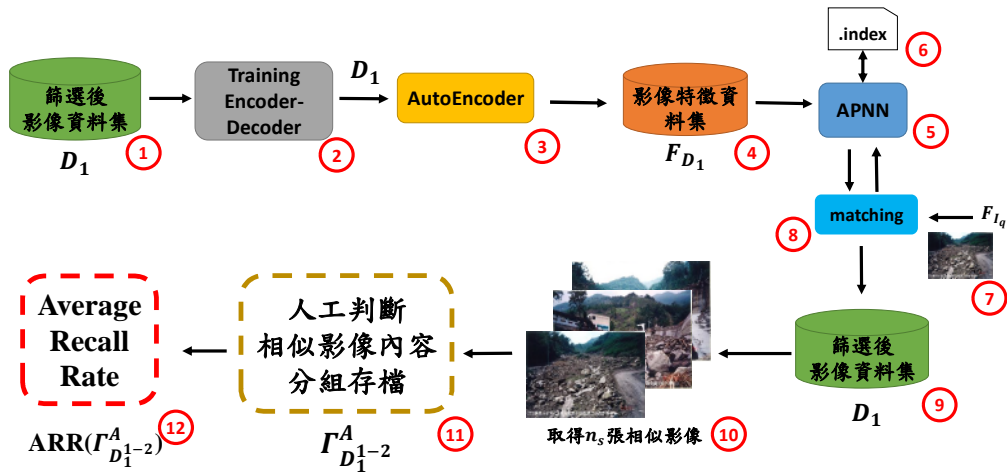


圖 2- 21 Case1 實驗流程

圖 2-21 表示 Case 1 實驗流程步驟，其說明簡述如下：

步驟 1：輸入影像資料庫 D_1 。

步驟 2：將影像資料庫 D_1 內所有影像訓練 Encoder-Decoder。

步驟 3：AutoEncoder 輸出所有影像特徵向量。

步驟 4：建立所有影像特徵向量資料庫 F_{D_1} 。

步驟 5：將 F_{D_1} 中所有影像特徵向量，使用 APNN 演算法建

立 n_{APNN} 顆二元樹。

步驟 6：儲存成指標型檢索(Index Retrieval)檔，輸出指標型

檢索檔。

步驟 7：輸入查詢影像 I_q 至 AutoEncoder 取得 F_{I_q} 。

步驟 8： F_{I_q} 在指標型檢索(Index Retrieval)檔內進行 APNN

演算法，輸出 n_s 個相似影像索引值。

步驟 9：依 n_s 個相似影像索引值，至 D_1 中輸出對應 n_s 張相似影像

步驟 10：取得 n_s 張相似影像。

步驟 11： $\Gamma_{D_1^{1-2}}$ 前 10000 張影像透過人工判斷相似影像內容分組存檔， D_1^{1-2} 為 $D_1^{1-2} = D_1^1 \cup D_1^2$ 。

步驟 12：透過 $\Gamma_{D_1^{1-2}}$ 與 n_s 個相似影像比對，計算出 $ARR(\Gamma_{D_1^{1-2}})$ 。

圖 2-21 說明 Case 1 實驗流程的量化數據資料表 2-3~2-6，表 2-3~2-4 呈現影像檢索效能，表 2-5~2-6 呈現影像檢索效率。影像檢索效能是以 ARR (平均召回率)來呈現，表 2-5~2-6 是以平均查詢反應時間(秒/張)。表 2-7 說明本計畫所使用的主要硬體規格來訓練單機版智慧型影像檢索系統。表 2-8 說明本計畫所使用的主要硬體規格來呈現跨平台智慧型影像檢索計算，所有的操作以網頁瀏覽器為主。圖 2-22 表示 $\Gamma_{D_1^1}$ 與 $\Gamma_{D_1^2}$ 中影像數量對組數的分佈直方圖，紅色表示 $\Gamma_{D_1^1}$ 的分佈，

靛藍色表示 $\Gamma_{D_1^2}$ 的分佈。

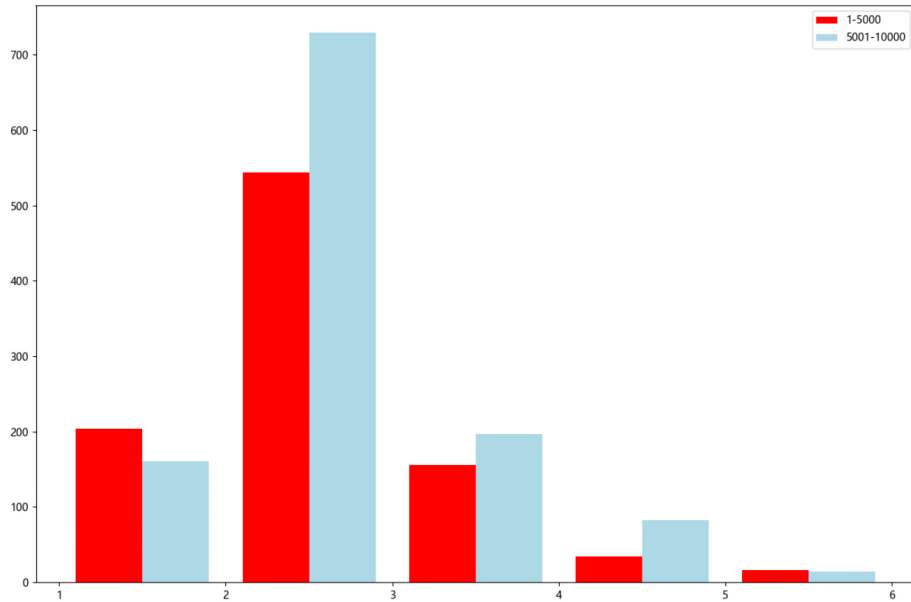


圖 2-22 表 $\Gamma_{D_1^1}$ 與 $\Gamma_{D_1^2}$ 影像數量對組數的統計直方圖

有關表 2-4~2-7 說明如下所述。

- 表 2-4 說明 Case 1 檢索效能，用 ARR 呈現，訓練樣本 D_1 ，測試樣本 $\Gamma_{D_1^1}$ ， $|\Gamma_{D_1^1}|=1988$ 。
- 表 2-5 說明 Case 1 檢索效能，用 ARR 呈現，訓練樣本 D_1 ，測試樣本 $\Gamma_{D_1^2}$ ， $|\Gamma_{D_1^2}|=2610$ 。
- 表 2-6 說明 Case 1 檢索效率(用單機版測量)，用平均查詢反應時間(秒/張)呈現，所有樣本 D_1 ，查詢樣本 $\Gamma_{D_1^1}$ ， $|\Gamma_{D_1^1}|=1988$ 。
- 表 2-7 說明 Case 1 檢索效率(用單機版測量)，用平均查詢反

應時間(秒/張)呈現，所有樣本 D_1 ，查詢樣本 $\Gamma_{D_1^2}$ ， $|\Gamma_{D_1^2}|=2610$ 。

由表 2-4~2-7 獲得以下的觀察。

- 表 2-4~2-5 呈現出當影像檢索輸出的張數增加，ARR 也會增加。
- 表 2-4~2-5 呈現出當二元樹的數量增加，ARR 卻沒有增加。
- 表 2-4~2-5 呈現出**議題 3-6**設計成效，表 2-4~2-5 第一列為沒有使用**議題 3-6**人工首次分組影像，第二列為運用**議題 3-6**人工分組影像，結果比較第二列的 ARR 高於第一列 ARR。
- 表 2-6~2-7 呈現出當影像檢索輸出的張數增加，ART(秒/張)也會微量增加。
- 表 2-6~2-7 呈現出當二元樹的數量增加，ART(秒/張)也會微量增加。

表 2-4 說明 Case 1 檢索效能，用 ARR 呈現，訓練樣本 D_1 ，測試樣本 Γ_{D_1}

n_{APNN} 二元樹數量	ARR	n_s : 影像檢索輸出的張數			
		10	20	40	80
20		0.464	0.496	0.529	0.574
		0.599	0.631	0.668	0.706
40		0.464	0.496	0.529	0.574
		0.599	0.631	0.668	0.706
60		0.464	0.496	0.529	0.574
		0.599	0.631	0.668	0.706
80		0.464	0.496	0.529	0.574
		0.599	0.631	0.668	0.706
100		0.464	0.496	0.529	0.574
		0.599	0.631	0.668	0.706
120		0.464	0.496	0.529	0.574
		0.599	0.631	0.668	0.706
1000		0.464	0.496	0.529	0.574
		0.599	0.631	0.668	0.706

表 2-5 說明 Case 1 檢索效能，用 ARR 呈現，訓練樣本 D_1 ，測試樣本 $\Gamma_{D_1^2}$

n_{APNN} 二元樹數量	n_s : 影像檢索輸出的張數			
	10	20	40	80
20	0.465	0.49	0.522	0.554
	0.592	0.618	0.645	0.678
40	0.465	0.49	0.522	0.554
	0.592	0.618	0.645	0.678
60	0.465	0.49	0.522	0.554
	0.592	0.618	0.645	0.678
80	0.465	0.49	0.522	0.554
	0.592	0.618	0.645	0.678
100	0.465	0.49	0.522	0.554
	0.592	0.618	0.645	0.678
120	0.465	0.49	0.522	0.554
	0.592	0.618	0.645	0.678
1000	0.465	0.49	0.522	0.554
	0.592	0.618	0.645	0.678

表 2-6 說明 Case 1 檢索效率，用 ART(秒/張)呈現，所有樣本 D_1 ，查詢樣本 Γ_{D_1}

ART(秒/張) n_{APNN} 二元樹數量	n_s :影像檢索輸出的張數			
	10	20	40	80
20	0.016	0.016	0.016	0.016
	0.019	0.018	0.018	0.018
40	0.016	0.016	0.016	0.017
	0.018	0.018	0.018	0.018
60	0.016	0.016	0.016	0.017
	0.019	0.019	0.018	0.019
80	0.016	0.016	0.017	0.017
	0.019	0.018	0.018	0.018
100	0.016	0.016	0.017	0.017
	0.018	0.017	0.017	0.017
120	0.016	0.016	0.017	0.017
	0.017	0.018	0.018	0.019
1000	0.017	0.017	0.018	0.019
	0.019	0.019	0.02	0.021

表 2-7 說明 Case 1 檢索效率，用 ART(秒/張)呈現，所有樣本 D_1 ，查詢樣本 $\Gamma_{D_1^2}$

ART(秒/張) n_{APNN} 二元樹數量	n_s :影像檢索輸出的張數			
	10	20	40	80
20	0.021	0.021	0.021	0.021
	0.024	0.023.	0.023	0.023
40	0.021	0.021	0.021	0.021
	0.024	0.023	0.023	0.023
60	0.021	0.021	0.021	0.021
	0.024	0.023	0.023	0.024
80	0.021	0.021	0.021	0.021
	0.023	0.023	0.023	0.024
100	0.021	0.021	0.021	0.022
	0.024	0.023	0.023	0.023
120	0.021	0.021	0.021	0.022
	0.023	0.022	0.022	0.023
1000	0.022	0.022	0.023	0.024
	0.023	0.023	0.023	0.025

表 2-8 雲端伺服器主要硬體規格，訓練智慧型影像檢索系統

CPU	Intel® Core™ i7-9700
GPU	AMD Radeon(TM) Pro WX7100 Graphics 8GB
RAM	48GB

表 2-9 網頁伺服器主要硬體規格，用於影像檢索計算

CPU	Intel® Core™ i7-5820K
RAM	16GB

·Case 2: 使用 D_1^1 訓練 AutoEncoder 模型，利用 AutoEncoder 擷

取 D_1^{1-2} ($D_1^{1-2} = D_1^1 \cup D_1^2$) 的影像特徵 $F_{D_1^{1-2}}$ ，再用

$\Gamma_{D_1^{1-2}}$ ($\Gamma_{D_1^{1-2}} = \Gamma_{D_1^1} \cup \Gamma_{D_1^2}$) 檢索極相似影像，主要評估使用 D_1^1 訓

練 AutoEncoder 之一般化能力(Generalization capability)，並與 Case 1 結果比較，觀察是否用小樣本就可訓練 AutoEncoder 擷取影像特徵。

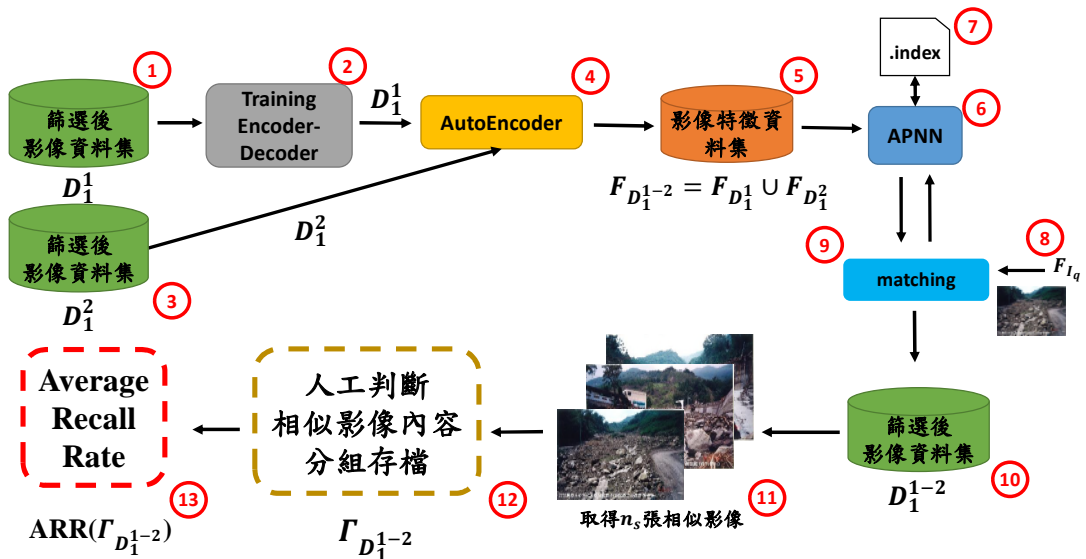


圖 2- 23Case2 實驗流程

圖 2-23 表示 Case 2 實驗流程步驟，其說明簡述如下：

步驟 1：輸入影像資料庫 D_1^1 。

步驟 2：將影像資料庫 D_1^1 影像訓練 Encoder-Decoder。

步驟 3：將 D_1^{1-2} 影像輸入至 AutoEncoder。

步驟 4：AutoEncoder 輸出 D_1^{1-2} 影像特徵向量資料庫 $F_{D_1^{1-2}}$ 。

步驟 5：建立所有影像特徵向量資料庫 $F_{D_1^{1-2}}$ 。

步驟 6：將 $F_{D_1^{1-2}}$ 中所有影像特徵向量，使用 APNN 演算法

建立 n_{APNN} 顆二元樹。

步驟 7：儲存成指標型檢索(Index Retrieval)檔，輸出指標型檢索檔。

步驟 8：輸入查詢影像 I_q 至 AutoEncoder 取得 F_{I_q} 。

步驟 9： F_{I_q} 在指標型檢索(Index Retrieval)檔內進行 APNN 演算法，輸出 n_s 個相似影像索引值。

步驟 10：依 n_s 個相似影像索引值，至 D_1^{1-2} 中輸出對應 n_s 張相似影像。

步驟 11：取得 n_s 張相似影像。

步驟 12： $\Gamma_{D_1^{1-2}}$ 前 10000 張影像透過人工判斷相似影像內容分組存檔。

步驟 13：透過 $\Gamma_{D_1^{1-2}}$ 與 n_s 個相似影像比對，計算出 $ARR(\Gamma_{D_1^{1-2}})$ 。

圖 2-23 說明 Case 2 實驗流程的量化數據資料表 2-10~2-11，主要測試 AutoEncoder 一般化能力。表 2-10 說明 Case 2 一般化檢索效能，用 ARR 呈現，訓練樣本 D_1^1 ，一般化測試樣本 D_1^2 ，測試樣本 $\Gamma_{D_1^1}$ ，另外，表 2-11 是用測試樣本 $\Gamma_{D_1^2}$ 。由表 2-10~2-11 與 Case 1 類似，輸出張數增加，ARR 增加，但是，二元樹數量增加，ARR 不會增加。比較表 2-4 ~ 2-5 與

表 2-10~ 2-11 可得到一個觀察結果，小樣本訓練比較大樣本訓練檢索效果略低，用卷積神經網路(CNN)所實現出來的 AutoEncoder 的確具有一般化能力。另外觀察，表 2-10~2-11 呈現出議題 3-6 設計成效，表 2-10~2-11 第一列為沒有使用議題 3-6 人工首次分組影像，第二列為運用議題 3-6 人工分組影像，結果比較第二列的 ARR 高於第一列 ARR。

表 2-10 說明 Case 2 一般化檢索效能，用 ARR 呈現，訓練樣本 D_1^1 ，一般化測試樣本 D_1^2 ，測試樣本 $\Gamma_{D_1^1}$

ARR n_{APNN} 二元樹數量	n_s :影像檢索輸出的張數			
	10	20	40	80
20	0.449	0.477	0.51	0.552
	0.563	0.58	0.606	0.643
40	0.449	0.477	0.51	0.552
	0.563	0.58	0.606	0.643
60	0.449	0.477	0.51	0.552
	0.563	0.58	0.606	0.643
80	0.449	0.477	0.51	0.552
	0.563	0.58	0.606	0.643
100	0.449	0.477	0.51	0.552
	0.563	0.58	0.606	0.643
120	0.449	0.477	0.51	0.552
	0.563	0.58	0.606	0.643
1000	0.449	0.477	0.51	0.552
	0.563	0.58	0.606	0.643

表 2-11 說明 Case 2 一般化檢索效能，用 ARR 呈現，訓練樣本 D_1^1 ，一般化測試樣本 D_1^2 ，測試樣本 $T_{D_1^2}$

ARR n_{APNN} 二元樹數量	n_s : 影像檢索輸出的張數			
	10	20	40	80
20	0.455	0.478	0.506	0.537
	0.572	0.596	0.621	0.649
40	0.455	0.478	0.506	0.537
	0.572	0.596	0.621	0.649
60	0.455	0.478	0.506	0.537
	0.572	0.596	0.621	0.649
80	0.455	0.478	0.506	0.537
	0.572	0.596	0.621	0.649
100	0.455	0.478	0.506	0.537
	0.572	0.596	0.621	0.649
120	0.455	0.478	0.506	0.537
	0.572	0.596	0.621	0.649
1000	0.455	0.478	0.506	0.537
	0.572	0.596	0.621	0.649

- Case 3: 先使用 D_1^1 訓練 AutoEncoder 模型且用 APNN 建立 D_1^1 的影像索引檔(Index)，接著，利用非 D_1^1 內影像作為查詢影像進行影像檢索，並且經 5 位觀察者以人工判斷影像內容相似度，用以觀察此模擬實驗結果。主要評估同時使用 AutoEncoder 及 APNN 之一般化能力。

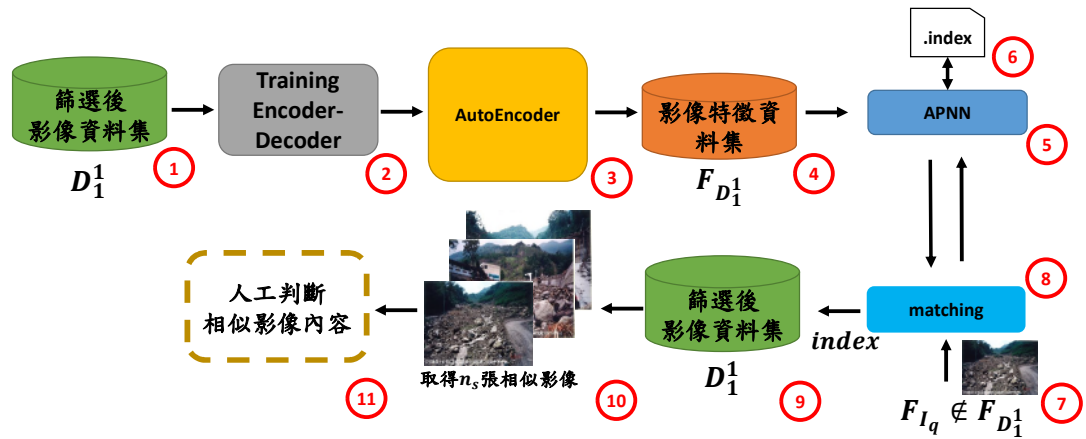


圖 2- 24 Case 3 實驗流程

圖 2-24 表示 Case 3 實驗流程步驟，其說明簡述如下：

步驟 1：輸入影像資料庫 D_1^1 。

步驟 2：將影像資料庫 D_1^1 影像訓練 Encoder-Decoder。

步驟 3：將 D_1^1 影像輸入至 AutoEncoder。

步驟 4：建立所有影像特徵向量資料庫 $F_{D_1^1}$ 。

步驟 5：將 $F_{D_1^1}$ 中所有影像特徵向量，使用 APNN 演算法建

立 n_{APNN} 顆二元樹。

步驟 6：儲存成指標型檢索(Index Retrieval)檔，輸出指標型

檢索檔。

步驟 7：輸入查詢影像 I_q ($I_q \notin D_1^1$) 至 AutoEncoder 取得 F_{I_q} ，

$F_{I_q} \notin F_{D_1^1}$ 。

步驟 8： F_{I_q} 在指標型檢索(Index Retrieval)檔內進行 APNN

演算法，輸出 n_s 個相似影像索引值。

步驟 9：依 n_s 個相似影像索引值，至 D_1^1 中輸出對應 n_s 張相似影像。

步驟 10：取得 n_s 張相似影像。

步驟 11： n_s 張相似影像結果透過 5 位人工審查投票，判斷影像內容是否相似。

附錄三提供 5 個 Case 3 的查詢結果案例，說明同時使用 AutoEncoder 及 APNN 之一般化能力。經 5 位觀察者以人工判斷檢索結果的影像內容其相似度大多近似於查詢影像，因此，本計畫發展的智慧型影像檢索系統可具 AutoEncoder 及 APNN 之一般化能力之效能令人滿意。

(二) 議題 5-2: 模擬實驗分析智慧型深度類神經網路分類災害影像模組之分類能力

本議題主要是要評估主題四所發展出的影像分類模型之分類效能。本議題利用隨機取樣方式，由 m' 類影像集 NDS_1' , NDS_2' , ..., $NDS_{m'}$ 中測試資料集，分別有各類別取得 k 張影像，然後分別當作欲分類影像，辨識影像屬於不同災害或構造物類型，算出其(平均)準確率(Accuracy)、(平均)混淆矩陣。觀察其結果，進而分析本計畫所提出智慧型影像分類模組之改善方式提高其分類效能。圖 2-25 說明本議題模擬實驗中人工所收集影像類別，分別有兩大類，高空攝影與 CCTV，然後，CCTV 又分四類:土石流、設施崩塌、山坡崩塌與其他。

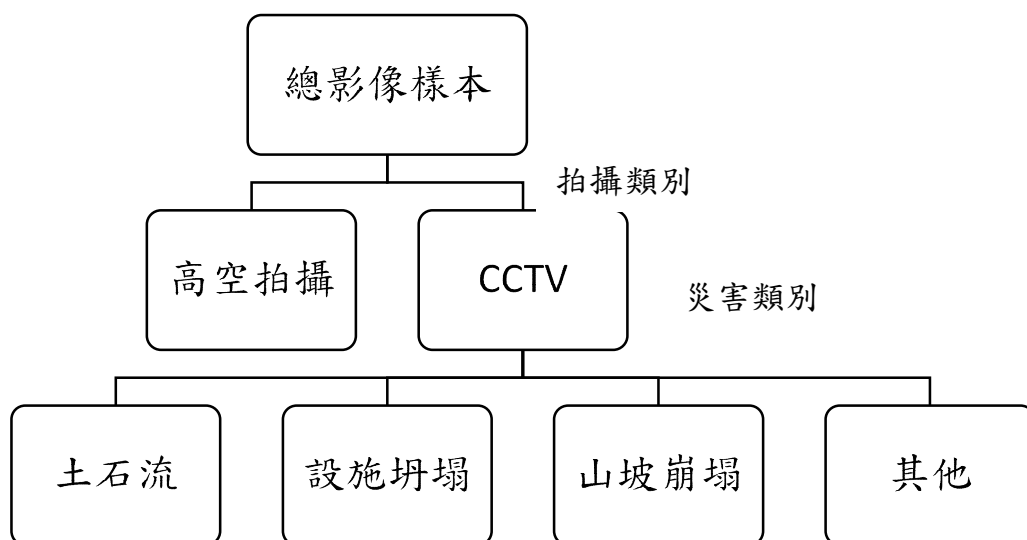


圖 2-25 本實驗人工所收集影像類別

表 2-12 列出拍攝類別影像樣本及 CCTV 災害類別影像中訓練集、驗證集及測試集的張數

拍攝類別影像樣本總張數 3688 張，訓練集和驗證集共 2400 張，測試集 1288 為張，共分為二類。			
訓練集	驗證集	測試集	類別
2160	240	1288	2
1080	120	644	高空攝影
1080	120	644	CCTV
CCTV 災害類別影像樣本總張數 1900 張，訓練集和驗證集共 880 張，測試集為 920 張，共分為四類。			
訓練集	驗證集	測試集	類別
792	88	920	4
198	22	380	土石流
198	22	55	設施崩塌
198	22	94	山坡崩塌
198	22	391	其他

表 2- 13 拍攝類別訓練集和測試集的混淆矩陣和 AR

拍攝類別 影像訓練集	預測標籤 -CCTV	預測標籤- 高空拍攝	AR
正確標籤 -CCTV	1080	0	1.000
正確標籤- 高空拍攝	7	1073	0.994
			0.997
epochs	197 (共 200，沒有提前終止)		
拍攝類別 影像測試集	預測標籤 -CCTV	預測標籤- 高空拍攝	AR
正確標籤 -CCTV	590	54	0.916
正確標籤- 高空拍攝	39	605	0.939
			0.927

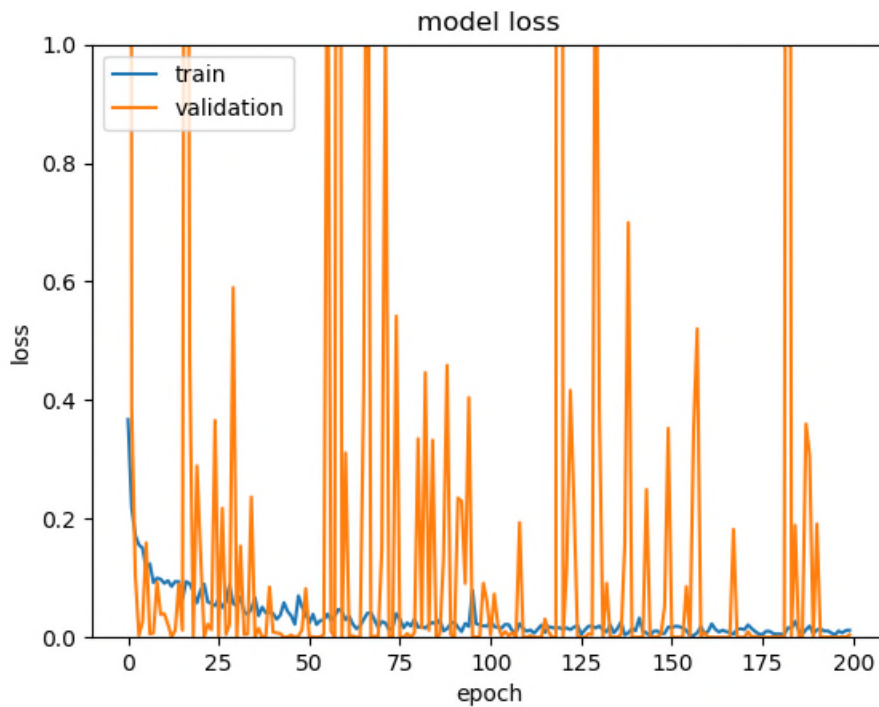


圖 2- 26 拍攝類別中訓練集(train)及驗證集(validation)，訓練過程中訓練效能(loss)的變化。

表 2-12 列出拍攝類別影像樣本及 CCTV 災害類別影像中訓練集、驗證集及測試集的張數。附錄四有列出每類的部分影像，每類提供 10 張影像。表 2-13 列出拍攝類別訓練集和測試集的混淆矩陣和 AR，設定訓練世代為 200 epochs。觀察表 2-13，發現本計畫所提出智慧型影像分類技術，訓練階段，對拍攝類別影像有較佳的效能，99.7% 的正確率(AR)。而對測試集而言也有 92.7% 的 AR。圖 2-26 說明拍攝類別中訓練集(train)及驗證集(validation)，訓練過程中訓練效能(loss)的變化。觀察圖 2-26，驗證影像集的分類結果差異相當大。但訓練影像集的分類結果比較一致。本實驗提前終止條件為目前 epoch 的 loss 值比前 50 次 epoch 之 loss 平均值沒有低於 10^{-6} ，明顯下降。此案例實驗時，沒有達到上述終止條件，最佳的分類效能發生在第 197 epoch。訓練時採用 10-fold，每次 10-fold 的訓練需約為 120 分鐘，共計約需 1200 分鐘 (20 小時)。

表 2-14 列出 CCTV 災害類別訓練集和測試集的混淆矩陣和 AR，訓練第 58 epoch 就符合終止條件。觀察表 2-14，發現本計畫所提出智慧型影像分類技術，在訓練階段中，對 CCTV 災害類別影像分類效能頗佳，有 94.2% 的 AR。但是對測試集

確只有 61.6% 的 AR。後續有討論說明。圖 2-27 說明拍攝類別中訓練集(train)及驗證集(validation)，訓練過程中訓練效能(loss)的變化。觀察圖 2-27，驗證影像集和訓練集的 loss 值在訓練初期差異極大，到中期以後(約第 20 epoch)，逐漸趨於一致穩定。本案例，訓練期間在第 58 epoch 結束後滿足終止條件。訓練時採用 10-fold，每次 10-fold 的訓練需約為 45 分鐘，共計約需 450 分鐘 (7 小時 30 分鐘)。表 2-15 及表 2-16 分別列出分類拍攝類別時，分類正確與錯誤的部份案例。表 2-17 及表 2-18 分別列出分類 CCTV 災害類別時，分類正確與錯誤的部份案例。

觀察本議題上述結果，歸納下列說明。

- 在拍攝影像類別的錯誤預測中，高空拍攝影像被誤判為 CCTV 影像，誤判的可能原因如下：
 - ◆ 高空拍攝時，拍攝位置離山體、建築物較近，易被視為 CCTV 影像。
 - ◆ 高空拍攝水域影像時，因為大範圍水域是平坦影像內容，易被視為 CCTV 影像。
- CCTV 影像被誤判為高空拍攝影像，誤判的可能原因如下：
 - ◆ 拍攝位置離山體、建築物較遠的照片，易被視為高空拍

攝影像。

- ◆拍攝位置高於拍攝目標，易被視為高空拍攝影像。
- ◆因此推斷本計畫所提出智慧型影像分類技術，對物體遠近，且對平坦水域影像內容相對不敏感。
- 在災害類別的錯誤預測中，土石流、設施崩塌和山坡崩塌三類影像，發現大都有混雜其他兩類的影像內容，而預測為其他的樣本較少，因此推斷其三類間的特徵差異性不高造成分類測試影像時效能不高。
- 依分類結果再重複經人工檢視影像來修正人工分類錯誤，但仍有不易分類判別的影像內容。

表 2-14 CCTV 災害類別訓練集和測試集的混淆矩陣和 AR

災害類別 影像訓練集	預測- 土石流	預測- 設施崩塌	預測- 山坡崩塌	預測- 其他	AR
標籤-土石流	190	2	4	2	0.960
標籤-設施崩塌	7	179	5	7	0.904
標籤-山坡崩塌	1	6	189	2	0.955
標籤-其他	3	5	2	188	0.949
					0.942
epoch	58 (第 58 epoch 結束後滿足終止條件)				
災害類別 影像測試集	預測- 土石流	預測- 設施崩塌	預測- 山坡崩塌	預測- 其他	正確 率
標籤-土石流	178	75	57	30	0.524
標籤-設施崩塌	6	28	18	3	0.509
標籤-山坡崩塌	12	14	60	8	0.638
標籤-其他	44	53	18	276	0.706
					0.616

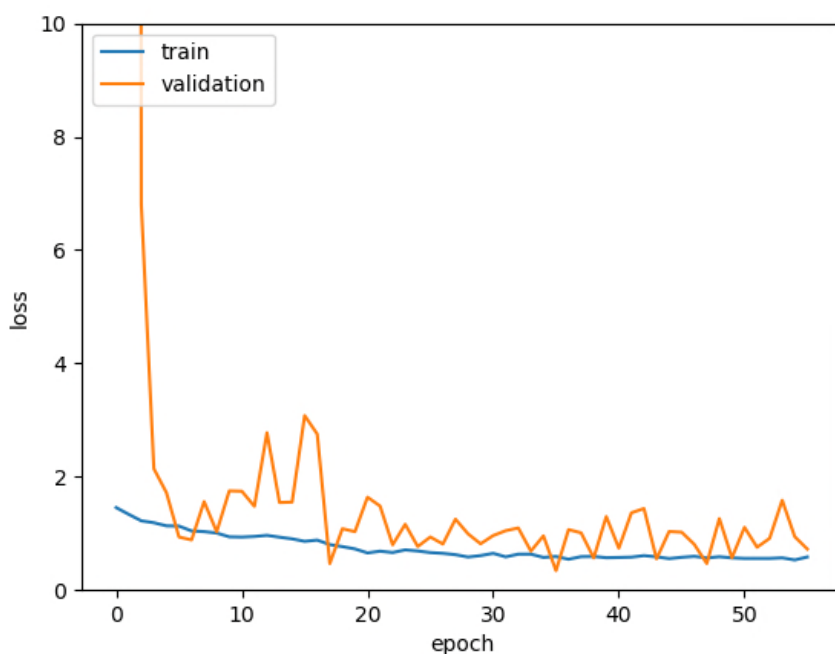


圖 2-27 CCTV 災害類別中訓練集(train)及驗證集(validation)，訓練過程中訓練效能(loss)的變化。



表 2-15 列出分類拍攝類別，分類正確的部份案例

高空拍攝- 正確預測		
		
	正確標籤-高空拍攝 預測標籤-高空拍攝	正確標籤-高空拍攝 預測標籤-高空拍攝
CCTV- 正確預測		
		
	正確標籤- CCTV 預測標籤- CCTV	正確標籤- CCTV 預測標籤- CCTV

表 2-16 列出分類拍攝類別，預測分類錯誤的部份案例



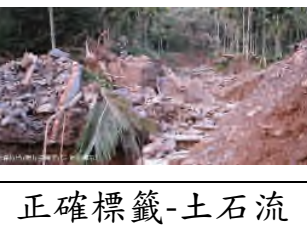





高空拍攝- 錯誤預測	 <p>水土保持局 (照片拍攝方式，姓名標示)</p>	 <p>水土保持局 (照片拍攝方式，姓名標示)</p>
	 <p>水土保持局 (照片拍攝方式，姓名標示)</p>	 <p>水土保持局 (照片拍攝方式，姓名標示)</p>
	正確標籤-高空拍攝 預測標籤- CCTV	正確標籤-高空拍攝 預測標籤- CCTV
CCTV- 錯誤預測	 <p>水土保持局 (照片拍攝方式，姓名標示)</p>	
	 <p>水土保持局 (照片拍攝方式，姓名標示)</p>	
	正確標籤- CCTV 預測標籤- 高空拍攝	正確標籤- CCTV 預測標籤- 高空拍攝

表 2-17 列出分類 CCTV 災害類別，分類正確的部份案例。

正確預測- 土石流		
		
	正確標籤-土石流 預測標籤-土石流	正確標籤-土石流 預測標籤-土石流
正確預測- 設施崩塌		
		
	正確標籤-設施崩塌 預測標籤-設施崩塌	正確標籤-設施崩塌 預測標籤-設施崩塌
正確預測- 山坡崩塌		
		
	正確標籤-山坡崩塌 預測標籤-山坡崩塌	正確標籤-山坡崩塌 預測標籤-山坡崩塌

正確預測- 其他		
		
	正確標籤-其他 預測標籤-其他	正確標籤-其他 預測標籤-其他

表 2- 18 列出分類 CCTV 災害類別，分類錯誤的部份案例

錯誤預測- 土石流		
		
	正確標籤-土石流 預測標籤-設施崩塌	正確標籤-土石流 預測標籤-山坡崩塌
錯誤預測- 設施崩塌		
		
	正確標籤-設施崩塌	正確標籤-設施崩塌

	預測標籤-土石流	預測標籤-山坡崩塌
錯誤預測- 山坡崩塌		
		
	正確標籤-山坡崩塌 預測標籤-土石流	正確標籤-山坡崩塌 預測標籤-設施崩塌
錯誤預測- 其他		
		
	正確標籤-其他 預測標籤-土石流	正確標籤-其他 預測標籤-設施崩塌

(三) 議題 5-3: 影像檢索模型成本分析

本議題針對本計畫所發展的影像檢索模型及影像分類模型探討其成本分析，要如何建立可靠、而且數量龐大的訓練資料庫？需要多少人力？時間？何種索引最理想？也是本議題須研究的題材。

圖 2-28 說明目前本計畫已發展的智慧型影像檢索系統的建置提供粗略的成本分析，包含建置時間及人力。有關程式設計人力成本約 8 個月/人。

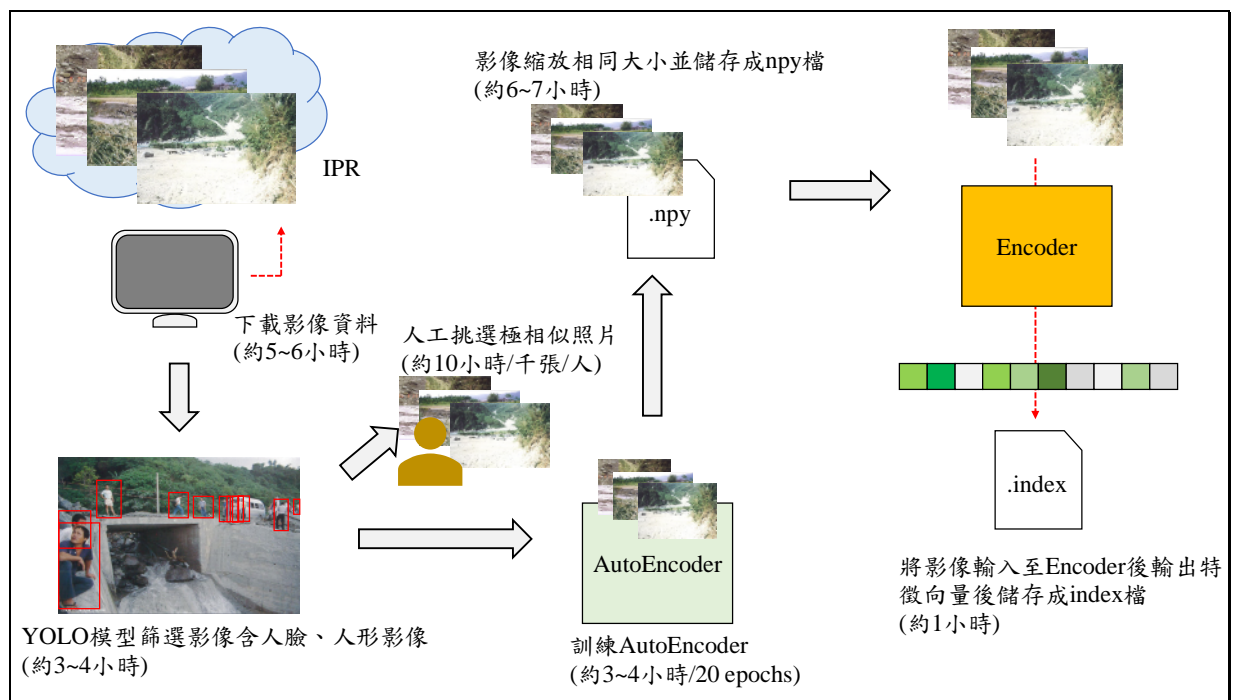


圖 2- 28 智慧型影像檢索系統建置成本分析

六、行動化模組之研究

本主題企圖將**主題三中的議題 3-4** 所訓練完成的深度類神經網路影像檢索模型將以跨平台方式來實現行動化影像檢索運算。主要以網頁瀏覽器來上傳查詢影像進行影像檢索，在雲端伺服器完成檢索的運算。可在各式載具執行影像檢索功能[18]。已完成影像檢索功能，直接使用行動載具上傳影像執行影像檢索功能。

(一)議題 6-1:智慧型以圖找圖行動化檢索模組之研究

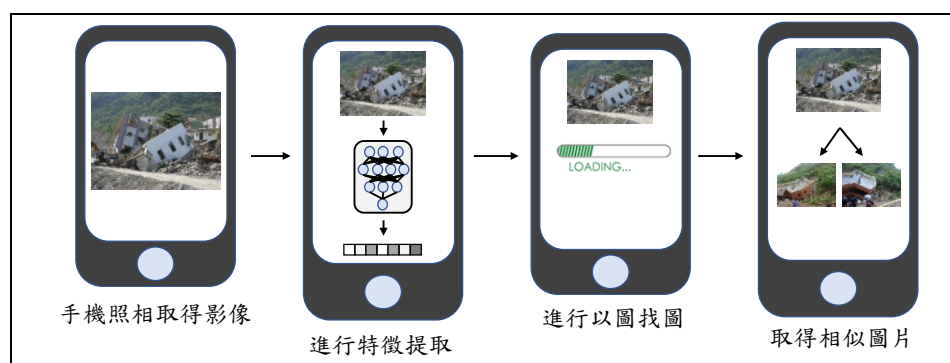


圖 2-29 以圖找圖影像檢索功能已發展於各式載具，以手機說明其概念

在圖 1-1 已說明跨平台行動載具執行影像檢索的系統架構圖，本議題已發展出圖 1-1 雛型系統，在圖 2-29 說明以圖找圖影像檢索功能已發展於各式載具，以手機說明其概念。圖 2-30 說明本計畫已發展出跨平台以圖找圖檢索行動化模組之系統流程。圖 2-30 以網頁端使用者介面操作流程為例。當使用者從網頁端上傳查詢影

像 I_q ， I_q 會經由伺服端輸入至 Encoder 並輸出查詢影像的特徵向量，此特徵向量會接著進行 APNN 演算法檢索鄰近 k 個影像特徵向量，本計畫設定 $k = n_s$ ，APNN 演算法檢索完成後會輸出該影像特徵向量的索引值，依照索引值從 D_1 取得對應的 k 張相似影像回傳至伺服端，伺服端接著以此 k 張影像的 id 查詢影像文字註解描述(至 IRP 查詢)，最後將 k 張相似影像與影像描述顯示於使用者網頁端。

步驟 1：在網頁端操作上傳查詢影像 I_q 。

步驟 2： I_q 傳回至雲端伺服端(本計畫建置)。

步驟 3：將查詢影像輸入至 Encoder。

步驟 4：Encoder 輸出影像特徵向量。

步驟 5：影像特徵向量在指標型檢索(Index Retrieval)檔內進行
APNN 演算法。

步驟 6：輸出 k 個相似影像索引值。

步驟 7：依照索引值取得 k 張相似影像。

步驟 8：於網頁端顯示 k 張相似影像。

步驟 9：挑一張相似影像至 IRP 查詢其文字註解描述。

步驟 10：於網頁端顯示相似影像與其影像文字註解描述。

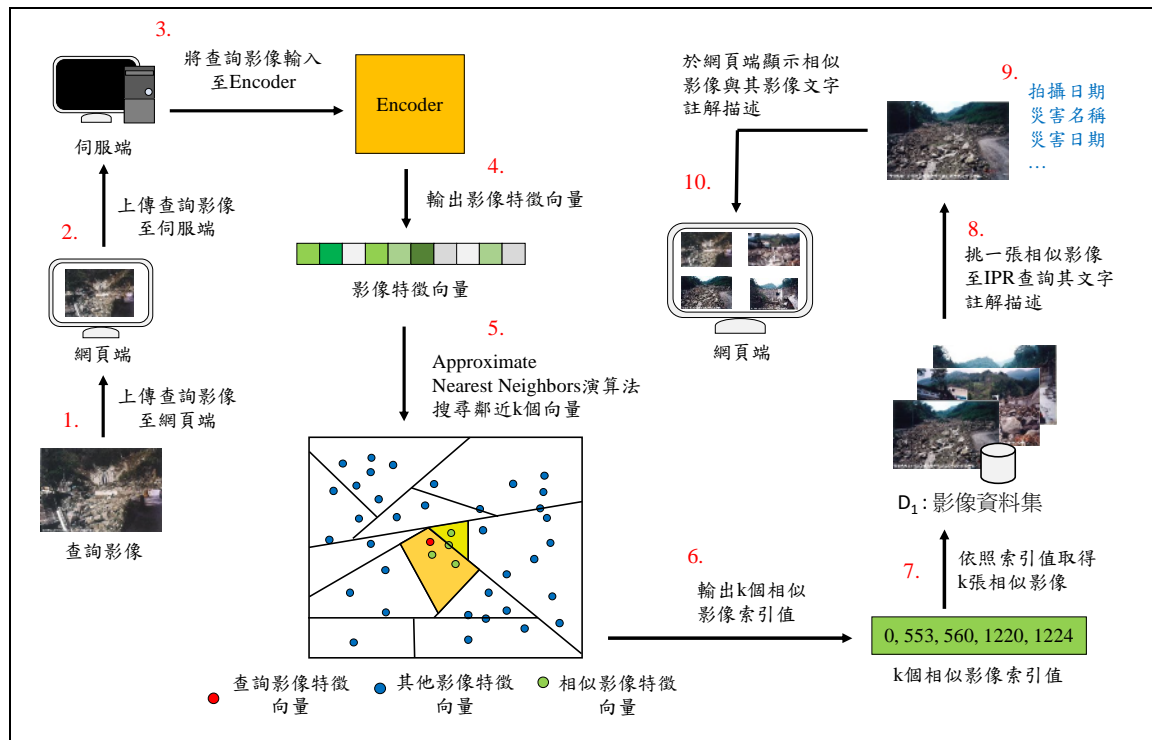


圖 2-30 本計畫已發展出跨平台以圖找圖檢索行動化模組之系統流程

本計畫已發展出跨平台以圖找圖檢索行動化系統，圖 2-31 ~ 2-33 為使用者操作頁面流程，分別為電腦版本、ipad 版、手機版。其網址為 <http://140.130.33.227:8887/>。其 QRcode 如下所示：

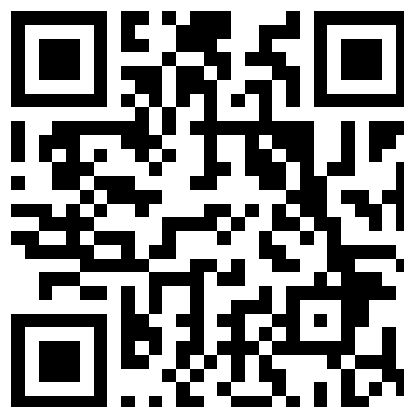


圖 2- 31 電腦版本的使用者操作頁面流程



圖 2- 32 ipad 版本的使用者操作頁面流程



圖 2- 33 手機版本的使用者操作頁面流程

(二)議題 6-2:使用端影像檢索評估之研究

- 本計畫已發展出的跨平台影像檢索系統可供多人同時連線查詢。
- 交互查詢檢索，圖 2-34 舉一例說交互查詢檢索成果，舉例，上傳 A_1 影像進行檢索，查詢產出影像 A_1, \dots, A_{10} ，接著，將 A_2, \dots, A_{10} 其中一張影像進行交互查詢測試，例如上傳 A_{10} 影像進行檢索，分別以 A_1 與 A_7 各自進行影像檢索，查詢結果有 7 張相同影像，以上可以說明本計畫所提出的影像檢索系統具影像檢索穩定度。雖查詢結果相似影像效能可能不高，但交互查詢測試可以說明系統是否具有檢索的穩定度。

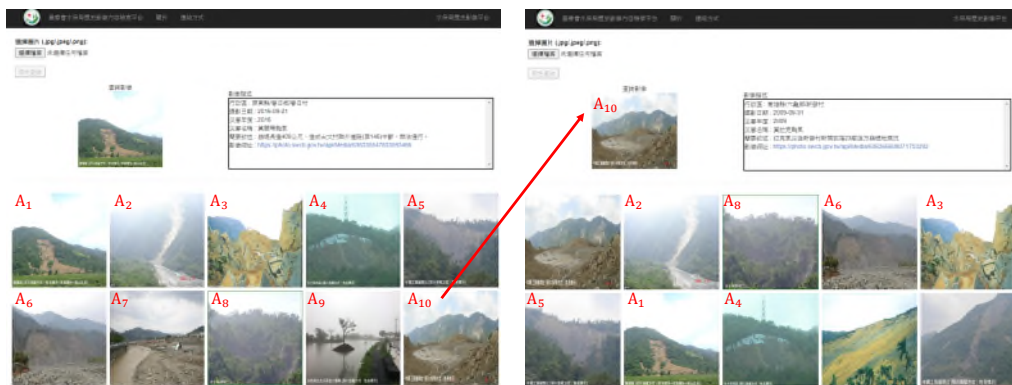


圖 2-34 舉一例說明交互查詢檢索成果

七、水土保持設施工程美學照片之內涵分析之量化估計

本主題主要參考[4]的研究成果，在其研究成果說明中利用問卷來分析視覺偏好度對表徵因子及感知因子的關聯，視覺偏好度主要有兩個元素，視覺性水量及視覺性綠量。本主題主要研究重點是發展深度學習模型來估算影像中的水視率(water looking rate, WLR)及綠視率(green looking rate, GLR)，降低人工問卷上主觀不一致問題。所以議題 7-1 將蒐集[18]所提及照片及網路上相關照片。議題 7-2 發展深度學習模型來估算影像中的水視率(WLR)及綠視率(RLR)。

(一)議題 7-1:收集水土保持設施工程美學照片

表 2-19 說明本計畫收集水土保持設施工程美學照片及其來源，總共 277 張照片。優良農業工程獎有 176 張、金質獎有 34 張、IPR 有 30 張。由[4]、自製及其他有 37 張。

表 2-19 收集水土保持設施工程美學照片及其來源

照片來源	優良農業工程獎	金質獎	IPR	[4]、自製、其他
數量	176	34	30	37

(二)議題 7-2:建置深度學習模型估算水視率及綠視率之研究

透過議題 7-1 所蒐集照片，可以透過影像實例分割模型來分析一張照片，主要作法是透過所訓練深度學習模型-Mask

RCNN 來實施影像實例分割，接著，估算出影像中的水視率及綠視率，可參考[9, 26, 30]來研究。本議題主要運用 Mask RCNN 來實施實例分割(instance segmentation)[9]。然後，估算水視率及綠視率。圖 3-36 說明本議題的說明。首先，圖 3-36 (a)表收集影像當輸入，接著，圖 3-36 (b)表人工標註各實例，然後，圖 3-36 (c)表利用 Mask RCNN 輸出實例分割結果，圖 3-36 (d)表輸出水視率及綠視率估計值，顯示於圖 3-36 (c)左下角的文字。

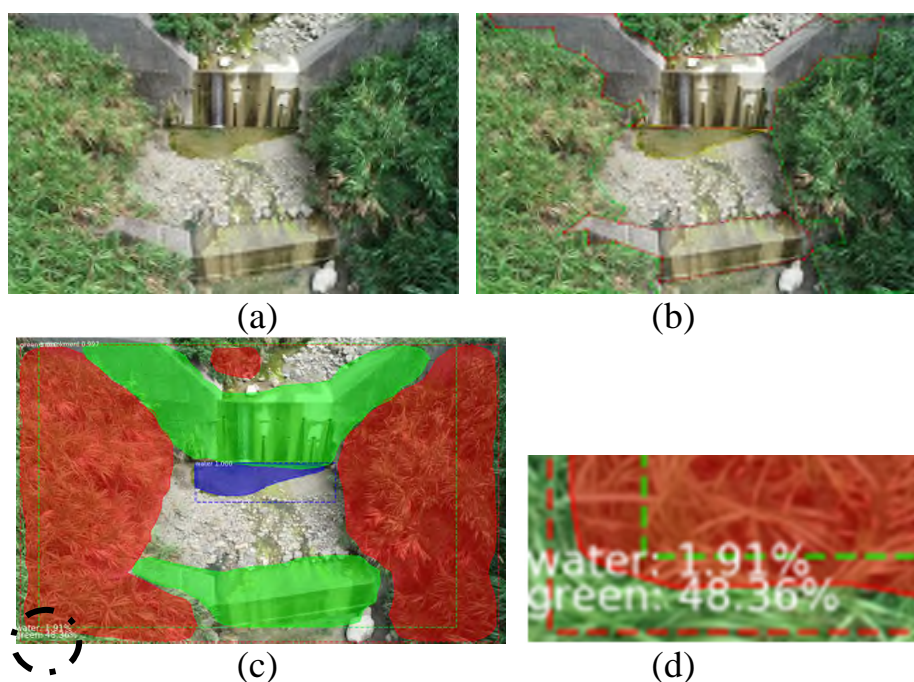


圖 2- 35 (a)收集影像; (b)人工標註各實例; (c) Mask RCNN 輸出實例結果; (d) 輸出水視率及綠視率估計值。

有關人工標註各實例的軟體為 labelme [20]，其操作步驟可詳閱附錄五。提供多邊形的標註方式來逼近影像中實例。本實驗中用人工標註五類實例。

- 天空(sky)用深藍色表示。
- 護岸(embankment)用紫色表示。
- 樹、草木、草皮等用綠色表示。
- 水(water)用紅色表示。
- 土石崩塌(collapse) 用淺藍色表示。

以上顏色用人工標註五類實例後建立 mask 影像用並非 Mask RCNN 所輸出實例分割影像中個實例分割的顏色。

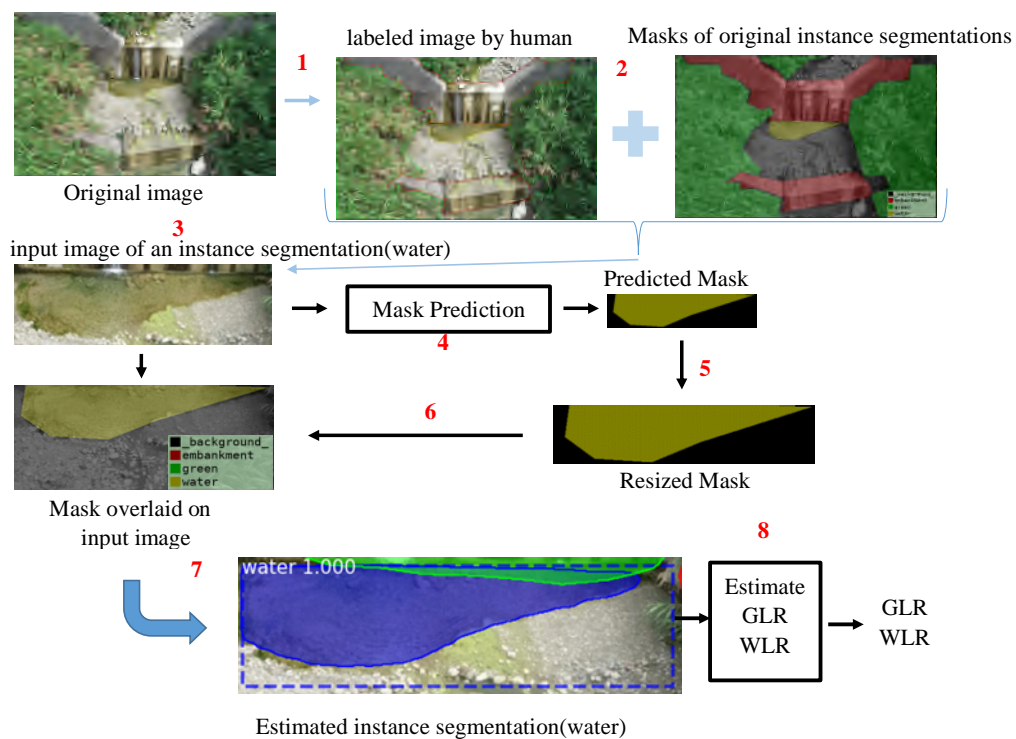


圖 2- 36 估計 GLR 與 WLR 的流程

圖 2-37 說明本議題利用 Mask RCNN 來估計 GLR 與 WLR 的流程。首先利用 Mask RCNN，輸出各實例分割，再利用 python 計算任意多邊形面積[31]演算法或區域的像素個數除整

張影像像素個數。計算水視分割區域及綠視分割區域之 WLR 與 GLR。以上兩種演算法會有相同計算結果。在圖 2-37 中其估計 GLR 與 WLR 的流程說明如下。

步驟 1: 輸入原影像 I

步驟 2: 用 labelme，人工標註 I 中的實例，Ilabel。

步驟 3: 輸出 2 個相關檔案, Jason 檔紀錄標註多邊形的頂點座標、每個實例 mask 檔(png 檔)。

步驟 4: Mask RCNN 取 ROI(錨框)來計算每個實例。

步驟 5: 估計出每個實例的輪廓位置後，預估其 mask。

步驟 6: 放大 mask 並套疊實例輪廓區域。

步驟 7: 重複步驟 1-6 來調整 Mask RCNN 權重，結束訓練 Mask RCNN。即可得到每個實例分割結果。

步驟 8: 利用[31]計算水視分割區域及綠視分割區域之 WLR 與 GLR。

在訓練 Mask RCNN 過程中，需要用到 Loss 來調整 Mask RCNN 的權重，圖 2-38 說明計算 Lclass, Lbox, Lmask 架構圖。此 Loss 由三項 loss 值加總。如下公式。

$$\text{LOSS} = L_{\text{class}} + L_{\text{box}} + L_{\text{mask}}。$$

- L_{class} 表實例分類錯誤。
- L_{box} 表實例錨框框在錯誤區域之錯誤。
- L_{mask} 表 Mask 套疊實例之錯誤。

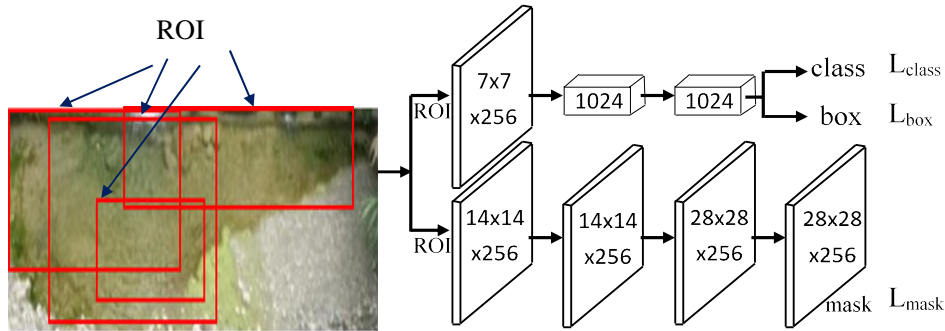


圖 2-37 計算 L_{class} , L_{box} , L_{mask} 架構圖

就各 L_{class} 、 L_{box} 、 L_{mask} 三項計算，下面說明之。

- L_{class} : 用 L_{cls} 代表。

$$\frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*)$$

$$L_{cls}(p_i, p_i^*) = -\log [p_i^* p_i + (1 - p_i^*)(1 - p_i)]$$

$L_{cls}(p_i, p_i^*)$ 中的 p_i 為 Anchor 預測為目標的機率， p_i^* 為 Ground truth 的標籤，0 是 negative label、1 是 positive label，此為一個標準的二分類交叉熵損失，對每個 anchor 計算 log 的損失，求和除以總 Anchor 數(N_{cls})。

- L_{box} : 表 Bounding box 的 loss，在此用 L_{reg} 表示。

$$\lambda \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}(t_i, t_i^*)$$

$$L_{reg}(t_i, t_i^*) = \sum_{i \in \{x, y, w, h\}} \text{smooth}_{L1}(t_i - t_i^*)$$

$$\text{smooth}_{L1}(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ |x| - 0.5 & \text{otherwise} \end{cases}$$

t_i 代表預測框的四個座標點， t_i^* 為 ground truth 的座標點，其中乘上 p_i^* 表示只關心 positive anchor 的回歸情況。因為在實際計算過程中，通常 N_{cls} 和 N_{reg} 差距過大，因此使用參數 λ 平衡兩者。

L_{mask} : 在 Mask Branch 中，對於新增加的 mask branch，其對於每個 region proposals 的輸出維度是 $K * m * m$ ，其中 $m * m$ 表示 mask 的大小， K 表示 K 個類別，因此這每個 region proposal 一共生成 K 個 binary mask，在得到預測 mask 後，對 mask 的每個像素點值求 sigmoid 函數值（即所謂的 per-pixel sigmoid），得到的結果作為 L_{mask} （交叉熵損失函數）的輸入之一。

(三)議題 7-3:模擬實驗驗證估算水視率及綠視率之研究

在圖 2-39 中說明 Mask RCNN 架構圖，其中用 CNN 來抽取輸入影像的 feature map，再利用 RPN 來計算實例的錨框 (ROI)，接著利用 ROI Align 來計算錨框內實例的輪廓(mask)，同時利用全連接類神經網路來判別實例的類別，另外，以線性回歸的方式將候選區域的 bounding box(矩形錨框)和原目標物的 bounding box(矩形錨框)盡量對齊。

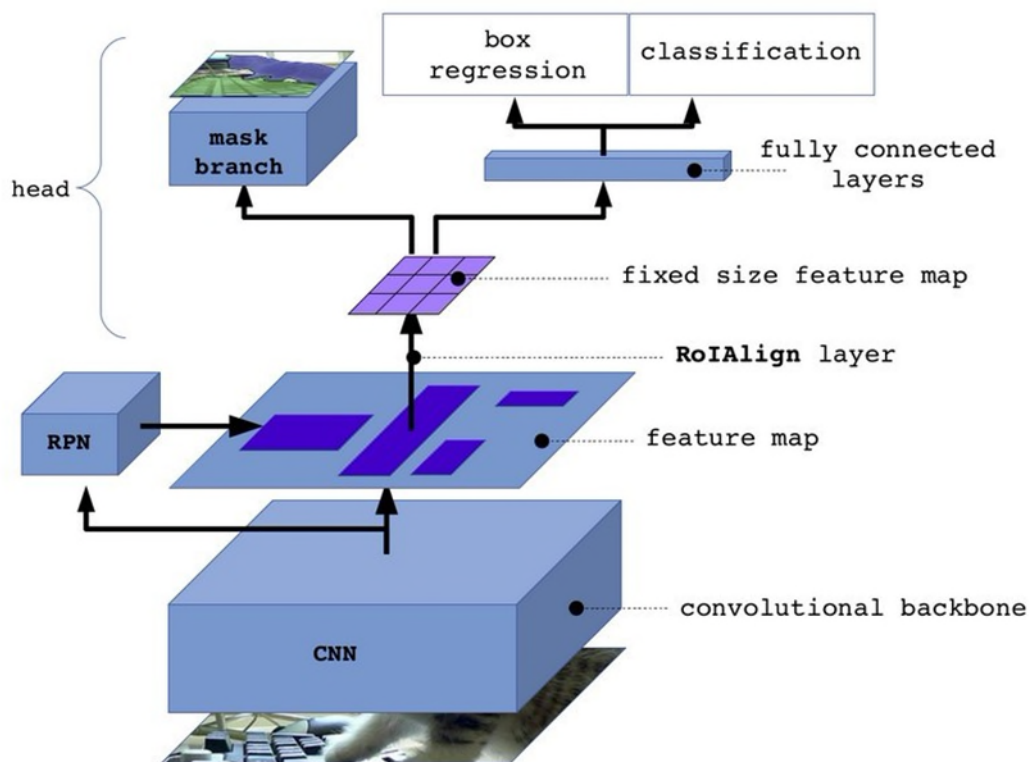
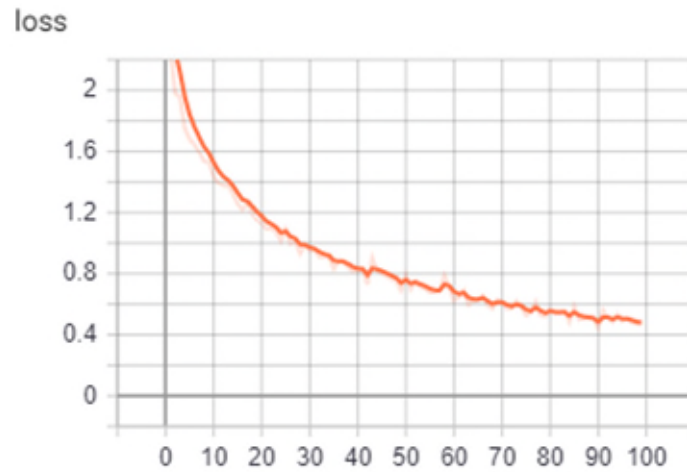


圖 2- 38 Mask RCNN 架構圖 (圖來源[30])

本實驗中所使用的實驗影像如議題 7-1 所述。訓練影像集與驗證影像集中影像個數為 243 張，其比例為 8:2。測試影像集中影像個數為 43 張。圖 2-40(a)表示訓練過程中針對訓練影像集之總 loss, $L(\text{loss})$ 表示)的情形，x 軸為 epochs 個數，圖 2-40 (b)-(d)分別為三項 L_{class} (mrcnn_class_loss 表示), L_{box} (mrcnn_bbox_loss 表示), L_{mask} (mrcnn_mask_loss 表示)的情況。圖 2-41 (a) 表示訓練過程中針對驗證影像集總 loss, $L(\text{val_loss})$ 表示)的情形，x 軸為 epochs 個數，圖 2-41 (b)-(d)分別為三項 L_{class} ($\text{val_mrcnn_class_loss}$ 表示), L_{box} ($\text{val_mrcnn_bbox_loss}$), L_{mask} ($\text{val_mrcnn_mask_loss}$ 表示)的情況。

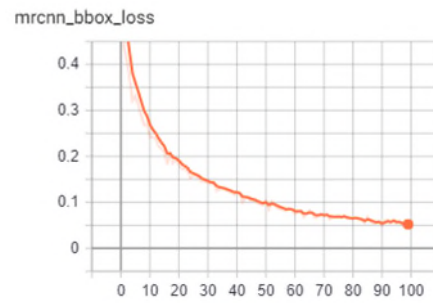
以訓練影像集(含驗證影像集)為例，表附錄 6-1 舉例說明原圖、實例分割結果及其預測水視分割區域及綠視分割區域之 WLR 與 GLR。以測試影像集為例，表附錄 6-2 舉例說明原圖、實例分割結果及其預測水視分割區域及綠視分割區域之 WLR 與 GLR。結果顯示訓練影像集(含驗證影像集)預估的 WLR 與 GLR 精準度高於測試影像集的預估。本實驗對兩類影像集所估計 WLR 與 GLR，也與人工視覺判定接近。



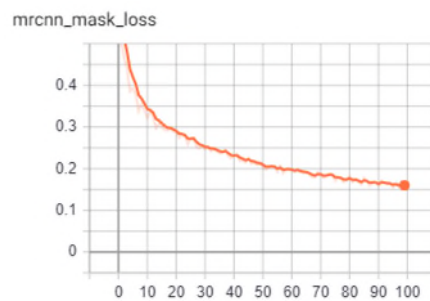
(a)



(b)



(c)



(d)

圖 2- 39 (a)訓練過程中針對訓練影像集之總 loss, L (loss 表示)的情形, x 軸為 epochs 個數; (b)-(d)分別為三項 L_{class} (mrcnn_class_loss 表示), L_{box} (mrcnn_bbox_loss 表示), L_{mask} (mrcnn_mask_loss 表示)的情況。

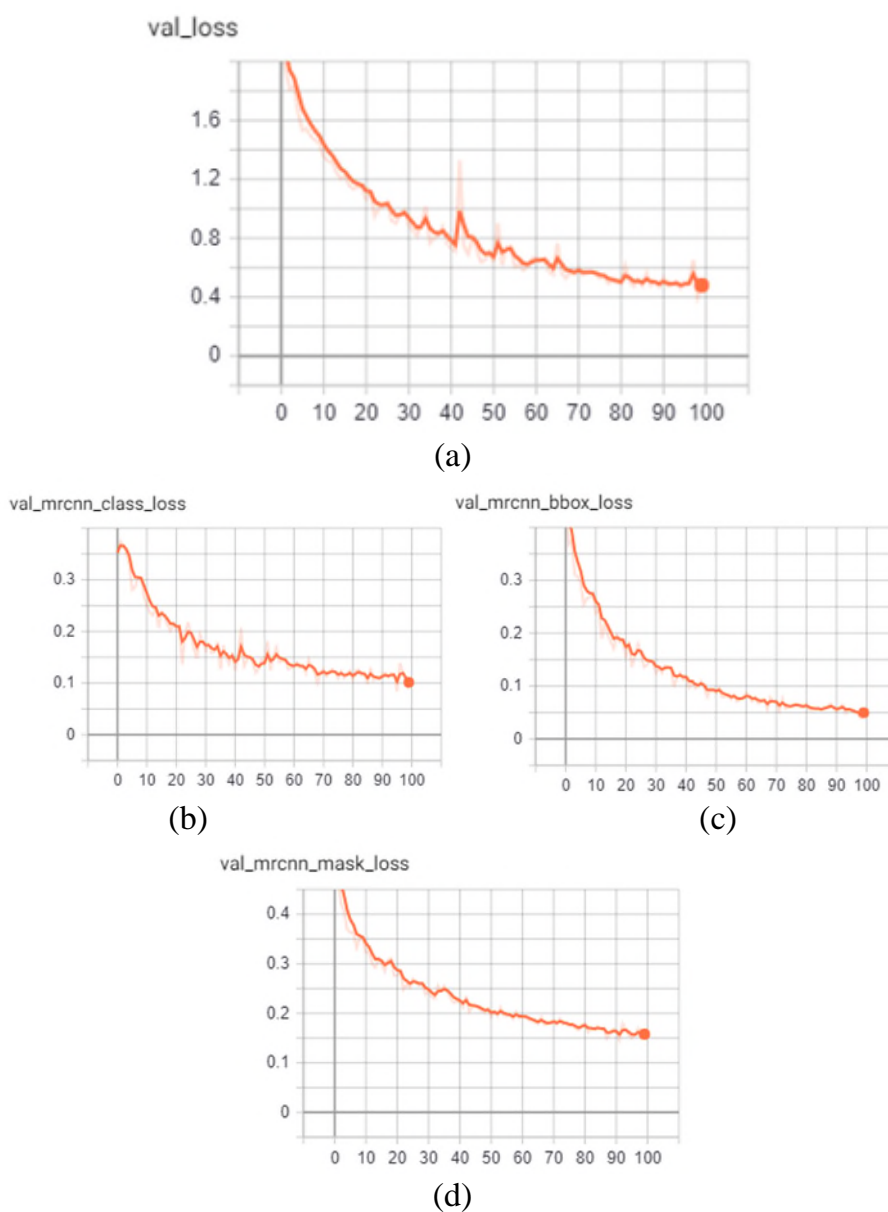


圖 2-40 (a)訓練過程中針對驗證影像集總 loss, $L(\text{val_loss})$ 表示)的情形, x 軸為 epochs 個數; (b)-(d)分別為三項 $L_{\text{class}}(\text{val_mrcnn_class_loss})$ 表示), $L_{\text{box}}(\text{val_mrcnn_bbox_loss})$, $L_{\text{mask}}(\text{val_mrcnn_mask_loss})$ 表示)的情況。

第三章 執行成果及效益

第一節 執行成果

本計畫可交付下列成果項目。

- 基於 IRP 影像搜尋(檢索)系統(以水土保持局建置歷史影像平台(IRP)蒐集之歷年土砂災害照片為範圍)之操作說明及展示影片各一份。行動載具上執行影像檢索之操作說明及展示影片各一份。
- 智慧型影像分類技術可自動化判別 CCTV、高空攝影兩類別照片，針對 CCTV 照片再自動化判別土石流、崩塌(不同構造物崩塌)等不同災害類型(以水土保持局建置歷史影像平台(IRP)蒐集之歷年土砂災害照片為範圍)之操作說明及展示影片各一份。
- 發展影像實例分割技術估算影像中的水視率及綠視率，可以應用於水土保持設施工程美學照片主要內涵分析以視覺偏好為主，其操作說明及展示影片各一份。
- 上述系統效能及效率分析結果報告一份。

第二節 執行效益

一、基於IRP影像檢索系統效益

水土保持局歷史影像平臺的檢索介面需透過輸入文字檢索，本計畫提供另一種檢索操作介面，利用以圖找圖檢索歷史影像平臺中的影像，輸出一些影像其內容相似於查詢影像。系統可進一步提供輸出影像在歷史影像平臺中的文字註解，可提升歷史影像平臺的加值利用。另外，已完成發展跨平台執行影像內容檢索功能，可以提供使用者更多元的查詢方式，有助於推廣歷史影像平臺的加值利用。此外，未來可以將歷史影像平臺中影像文字註解為空白的情況整理，提供給水土保持局相關單位參考，協助減少人工涉入回填充白影像文字註解，不需要透過人工逐一檢查所有影像。

二、智慧型影像分類技術效益

智慧型影像分類技術可自動化判別土石流、崩塌等不同災害類型，未來可應用於水土保持相關業務之分類、辨識、預測等相關業務。

三、影像實例分割技術效益

本計畫發展影像實例分割技術估算出影像中的水視率及綠視率。可以協助[18]視覺偏好度分析工作中人工問卷上主觀不一致的問題。也可以針對更多水土保持設施工程美學照片來實施內涵分析之量

化估計，也可以減少透過問卷獲取視覺偏好度之量化數據，所耗費的人力及時間。

第三節 結果與討論

由本計畫的模擬實驗結果可得出下列觀察，導致影像檢索及影像分類效能不易提升。

- 影像內容物件雜多(例如：土石流、各類崩塌、建築物、各類構造物、高空攝影圖等等)，物件與物件之間的空間複雜，較難萃取出影像獨特的特徵。
- 影像內容的樣態與影像顏色雜多，以崩塌為例，就有道路崩塌、邊坡崩塌、山坡崩塌、橋樑塌陷、建築物塌陷等等，以上舉例的崩塌樣態頗多，無法詳細分類。崩塌又混雜其他災害類型，例如，土石流、機具、車輛等，造成影像內容樣態雜多。
- 由 IRP 下載有 20000 多張實驗影像，目前沒有已發展工具來幫忙自動分類影像，需要透過人工的方式分類影像相當耗時，所以實驗影像中只能提取局部影像來模擬實驗，例如取 5000 張來發展影像檢索及影像分類模型之訓練與測試。
- 訓練深度學習的模型需要規格較高的硬體環境，例如 GPU 顯示卡的記憶體容量有限，無法設計較深多層模型，導致難以實驗來證明是否有更佳效能的深度學習模型。另外深度學習模型參數組合龐大，每項組合需要花費長時間來訓練模型，挑選更佳效能的

模型。

- 同個場景由不同角度所拍攝的影像數量較少。因同種場景的訓練影像樣本數較少，導致深度學習模型無法提升辨識的檢索及分類效能。

本計畫的研究限制如下所述。

- 由於 IRP 影像數量較為龐大，人工辨識與建置模型時間較長，所以本計畫限制影像為 IRP 中取 10000 張影像進行影像檢索模型訓練與人工辨識。
- 由於 IRP 相似影像內容較少，為了評估影像檢索的召回率，需要利用人工建置相似影像的分組，每組內影像需極相似影像內容，所以造成每組內的影像數量不多。
- 由於 10000 張影像內容雜多而且一張影像中有好幾種類別，因此不易分成多類，如：CCTV、坑洞、車子、建築物、橋樑、崩塌、土石流、隧道、道路坑洞、高空拍攝等等，所以本計畫研究影像分類的問題有下列限制，如先分成 2 類 CCTV 與高空拍攝。再將 CCTV 災害類別分成 4 類土石流、設施崩塌、山坡崩塌、其他。

第四章 結論及建議

第一節 結論

針對水土保持局歷史影像平台(IRP)內的影像為主，本計畫已產出 3 項成果，除了兩項成果，建置跨平台行動載具版影像搜尋系統、影像分類技術。另外，本計畫另一項成果，除收集 IRP 內的影像也收集網路影像，針對水土保持設施工程美學照片之建置深度學習模型分析影像內涵，此深度學習模型實施影像實例分割技術，可推估美學照片中的表徵因子的水視率及綠視率。最後，提供建置上述系統之效率、效能、成本分析、觀察及研究限制。

第二節 建議

本計畫實驗影像範圍是以水土保持局建置歷史影像平台(IRP)蒐集之歷年土砂災害照片為範圍，由於該平台，影像內容上傳時的文字註解比較難以詮釋影像內容審。初步分析此資料集，仍有些照片內容無關土砂災害及影像內容雜多(包含多項災害類別混合)，將與水土保持局討論是否可以提供照片內容的判斷原則?提供計畫執行建置系統時篩選實驗影像的依據。可能主辦及協辦機關是技術研究發展組。

參考文獻

- [1] B.-M. Chang, H.-H. Tsai, and C.-Y. Yen, "SVM-PSO based rotation-invariant image texture classification in SVD and DWT Domains," *Engineering Applications of Artificial Intelligence*, Vol. 52, pp. 96-107, June 2016.
- [2] B.-M. Chang, H.-H. Tsai, and W.-L. Chou, "Using visual features to design a content-based image retrieval method optimized by particle swarm optimization algorithm," *Engineering Applications of Artificial Intelligence*, Vol. 26, No. 10, pp. 2372-2382, Nov. 2013.
- [3] N. Chaudhuri and I. Bose, "Exploring the role of deep neural networks for post-disaster decision support," *Decision Support Systems*, Vol. 130, 113234, 2020.
- [4] S.-C. Chen and J.-C. Chen, "Scenic Beauty Estimation of Sediment Control Facilities of Torrent," 2019.
https://tech.swcb.gov.tw/Content/Upload/Innovation/%E6%BA%A%E6%B5%81%E9%98%B2%E7%A0%82%E8%A8%AD%E6%96%BD%E7%BE%8E%E8%B3%AA%E8%A9%95%E4%BC%B0%E9%A9%97%E6%94%B6_%E5%A0%B1%E5%91%8A%E6%9B%B8.pdf.
- [5] F. Chollet, "keras-team/keras: Deep Learning for humans - GitHub," Retrieved from <https://github.com/keras-team/keras>, 2015.
- [6] S.-H. Chuang, M.-S. Chen, T.-Y. Chou, and P.-W. Huang, "The study of a sequence image for debris flow events identification," *Journal of Chinese Soil and Water Conservation*, Vol. 38, No. 3, pp. 259-274, 2007.
- [7] M. J. J. Ghrabat, G. Ma, I. Y. Maolood, S. S. Alresheedi, and Abduljabbar, Z. A. "An effective image retrieval based on optimized genetic algorithm utilized a novel SVM-based convolutional neural network classifier," *Human-centric Computing and Information Sciences*, 9(1), 31, 2019.
- [8] P. Görgel, and A. Simsek, "Face recognition via deep stacked

- denoising sparse autoencoders (DSDSA),” *Applied Mathematics and Computation*, Vol. 355, pp. 325–342, 2019.
- [9] K. He, G. Gkioxari, P. Dollar, and R. Girshick, “Mask R-CNN,” 2018. <https://arxiv.org/pdf/1703.06870.pdf>.
- [10] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” *CoRR* abs/1512.03385, 2015.
- [11] K. He, X. Zhang, S. Ren, and J. Sun, “Identity mappings in deep residual networks,” In *European conference on computer vision* (pp. 630-645). Springer, Cham, 2016.
- [12] G. E. Hinton, “Reducing the dimensionality of data with neural networks,” *Science*, Vol. 313(5786), pp. 504 – 507, 2006.
- [13] Y. Jamtsho, P. Riyamongkol, and R. Waranusast, “Real-time Bhutanese license plate localization using YOLO,” *ICT Express* 2019.
- [14] E. N. Kajabad and S. V. Ivanov, “People detection and finding attractive areas by the use of movement detection analysis and deep learning approach,” *Procedia Computer Science*, Vol. 156, pp. 327–337, 2019.
- [15] R. Kenneth, *Requests: HTTP for Humans™*. Retrieved from <https://requests.readthedocs.io/en/master/>, 2011.
- [16] A. Latif, A. Rasheed, U. Sajid, J. Ahmed, N. Ali, N. I. Ratyal, and T. Khalil, “Content-based image retrieval and feature extraction: a comprehensive review,” *Mathematical Problems in Engineering*, 2019.
- [17] A. Qayyum, S. M. Anwar, M. Awais, and M. Majid, “Medical image retrieval using deep convolutional neural network,” *Neurocomputing*, 266, 8-20, 2017.
- [18] N. Rahim, J. Ahmad, K. Muhammad, A. K. Sangaiah, and S. W. Baik, “Privacy-preserving image retrieval for mobile devices with deep features on the cloud,” *Computer Communications*, Vol. 127, pp. 75–85, 2018.

- [19] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You Only Look Once: Unified, Real-Time Object Detection,” 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
- [20] B. C. Russell, A. Torralba, K. P. Murphy, W. T. Freeman. “LabelMe: a Database and Web-based Tool for Image Annotation,” *International Journal of Computer Vision*, 77(1-3):157-173, 2008.
- [21] M. Singh, S. Nagpal, M. Vatsa, and R. Singh, “Are you eligible? Predicting adulthood from face images via Class Specific Mean Autoencoder,” *Pattern Recognition Letters*, Vol. 119, pp. 121 – 130, 2019.
- [22] H.-H. Tsai, B.-M. Chang, and S.-H. Liu, “Rotation invariant image retrieval using PSO and SVR,” *Applied Soft Computing*, Vol. 17, pp. 127-139, Apr. 2014.
- [23] S. Van der Walt, J. L. Schönberger, J. Nunez-Iglesias, F. Boulogne, J. D. Warner, N. Yager, and T. Yu, “Scikit-image: image processing in Python,” *PeerJ*, 2, e453, 2014.
- [24] J. Wan, D.-Y. Wang, S. C. H. Hoi, P.-C. Wu, J.-K. Zhu, Y.-D. Zhang, and J.-T Li, “Deep learning for content-based image retrieval: A comprehensive study,” In *Proceedings of the 22nd ACM international conference on Multimedia* (pp. 157-166), New York, 2014, November.
- [25] N. Wang, X. Zhao, P. Zhao, Y. Zhang, Z. Zou and J. Ou, “Automatic damage detection of historic masonry buildings based on mobile deep learning,” *Automation in Construction*, Vol. 103, pp. 53-66, 2019.
- [26] J.-H. Wu, C.-H. Tseng, L.-C. Chen, S.-. Lo, and F.-P. Lin, “Automated image identification method for flood disaster monitoring in riverine environments: a case study in Taiwan,” 2015 AASRI International Conference on Industrial Electronics and Applications, London, UK, 2015.

- [27] C. Yeum, S. Dyke, and J. Ramirez, “Visual data classification in post-event building reconnaissance,” *Engineering Structures*, Vol. 155, pp. 16-24, 2018.
- [28] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, “Convolutional neural networks: an overview and application in radiology,” *Insights into Imaging*, 9(4), 611-629, 2018.
- [29] Annoy Tool, <https://github.com/spotify/annoy>.
- [30] Mask R-CNN structure,
https://blog.csdn.net/qq_36758461/article/details/99723099.
- [31] https://blog.csdn.net/weixin_43399503/article/details/90521073.
- [32] <https://blog.csdn.net/remanented/article/details/79564045>.
- [33] <https://arxiv.org/pdf/1512.03385.pdf>.
- [34] <https://arxiv.org/pdf/1603.05027.pdf>

附錄

附錄一、期初工作執行計畫書審查意見

項次	審查意見	回覆說明
期初需求訪談內容與意見：		
一	本計畫擬建置智慧型影像辨識技術,以自動化辨識土石流、崩塌等不同災害類型、構造類型,目標有實用價值,惟工作執行計畫書缺乏明確工作細部及步驟描述,建議縮小研究範圍,釐清工作內容,提出可行之研究步驟及流程,量化預期成果。	<ol style="list-style-type: none"> 1.在期中報告書第三節中對各項工作有提供細部說明及步驟流程。 2.第三節第五小節有提供量化成果。 3.目前成果已經發展以圖找圖影像內容檢索系統之計算核心模組。 4.已完成跨平台行動化影像內容檢索系統。 5.已完成影像特徵比對的效能分析,以計算複雜度為分析手段。
二	本研究欲同時辨識不同遙測影像,包括衛星影像、UAV 照片、CCTV 影像、手機相片。惟各種影像之呈現皆不同,無法用同一模型適用不同影像,應先選擇某一種影像來源較為可行。	<ol style="list-style-type: none"> 1.在期中報告書第三節中有說明對影像進行篩選及前處理(包括影像正規化),使得影像來源格式相異度可稍微降低。 2.目前使用 CCTV、衛星、手機照片進行實驗驗證,後續會先利用影像分類的方式再進行影像檢索,可避免不同類別的影像無法使用同一個模型進行檢索的情形。 3.本研究利用卷積類神經網路的非監督式學習方法來擷取影像特徵,依影像特徵來進行影像內容檢索,此方式可降低影像格式不一致的情況之效能。

項次	審查意見	回覆說明
三	本案富創新議題,具跨域整合研究,有關災害影像、水保設施工程美學照片之分類及特徵資料建置,宜有水保或景觀專才人員參與協助。	1.謝謝委員 2.會持續請教水保局的技術研究發展小組相關人員
四	所欲辨識之類別應先定義,且須與水保工程專業人員先溝通,以確定產出之成果。	會持續請教水保局的技術研究發展小組相關人員
五	欲以 CNN 之深度類神經網路來做影像災害類型分類,但計畫書中對網路架構仍屬一般型描述,並未對其層數、神經元數、參數數量及初步測試結果作描述,難以確定其可行性。	1.本研究以 CNN 擷取影像特徵細部的描述在第三節的第三小節。 2.後續對影像災害類型分類也將提供如第三節的第三小節 CNN 架構及參數的說明。
六	欲開發使用者於行動載具輸入所拍攝影像,CNN 即可立刻進行分類,但未說明其分類器安裝於手機本體或需透過網路於雲端辨識,其可行性尚有許多問題待釐清。	1.影像檢索的功能,因影像數量龐大手機記憶體有限,所以本研究設計透過於網路雲端辨識,使用跨平台的行動影像檢索系統。 2.影像分類的功能,因僅輸出分類的類別,所以可以由手機本體預估分類的類別輸出,不需透過網路於雲端辨識。
七	欲從影像中估算水量及綠量,用於分析工程美學,此項工作描述過於抽象,應定義其量化步驟,並說明如何評估水量及綠量辨識,及計算其比例,因為現場可能缺乏控制點及絕對尺寸,可行性待釐清。	1.本計畫預計利用影像分割的技術來分割一張影像中有水量及綠量的區域。 2.影像的成像會因攝影技術及環境的影響,所以本研究僅針對影像內容來估計水量及綠量的比例,暫不考慮控制點及絕對尺寸等因素。

項次	審查意見	回覆說明
八	本案為應用研究,宜針對應用缺口提出改善式精進說明。	<p>1.本計畫已發展以圖找圖的技術可以讓使用者輸入一張水土保持相關查詢影像(query image),本系統搜尋檢索 IRP 中影像資料庫,輸出一些影像內容與查詢影像內容非常類似的影像,然後,再選檢索出影像來查詢 IRP,得到其對應的文字註解描述。</p> <p>2.此應用缺口,除可讓使用者更了解影像的文字註解描述,亦可輔助使用者進一步核對其文字註解作業的正確性,研究中發現有許多影像在 IRP 有許多影像文字註解欄位是空白,本系統可以自動化檢索有空白欄位的影像,提供給相關人員補齊文字註解,減少大量人工檢核影像中欄位空白的作業工作量。</p>
九	災害影像如何精確分類,例如崩塌與土石流之關係,有些是複合型災害。	影像分類系統會輸出影像分類後屬於各類別的可能性程度,此種方式有別於 winner-take-all 影像分類決策模式。
十	建議補充照片災害分類各類 type 相關定義及關鍵字搜尋分類方式。	本計畫規劃嘗試用 semi-classification 的方式來發展。其說明如下。研究過程中,對於未分類影像輸入或者分類錯誤的影像,需要藉由人工來做調整建置新的類別。可是不能全部影像皆人工分類提供標籤。因此,主題四首先用主題三的影像檢索成果搭配人工觀察判斷,針對小樣

項次	審查意見	回覆說明
		<p>本影像集來進行類別分類，接著，建置訓練 CNN 深度類神經網路，做影像災害類型分類器，後續，針對剩餘的大樣本影像集，利用已訓練 CNN 來分類大樣本影像集。因此，將設計兩階段 CNN，第一階段分類器分類如果無法分類，收集於一類，主題四會再將無法分類的影像以人工的方式對影像作觀察判斷，給予標籤類別，可能是現有標籤或新增標籤。圖 2-19 說明 semi-classification 的流程架構。</p>
十一	<p>「摘要」有關農業水土保持影像之說明不妥,宜修正。內文亦有多處宜再檢核。</p>	<ol style="list-style-type: none"> 1.謝謝委員。 2.不妥處已修正。

附錄二、期中計畫書審查意見

項次	審查意見	回覆說明
期中需求訪談內容與意見：		
一	1.本計畫目前僅做到以深度學習技術建立物件辨識基本功能，尚未實測不同災害類型分類功能，且災害類別須先定義，列出資料庫中各種災害類別的影像內容與數目。	<p>不同災害類型分類功能已經有研究成果，請參閱： 第二章第三節的</p> <ul style="list-style-type: none"> •四、以深度學習技術建置不同災害類型影像分類模型之研究。 •5-2:模擬實驗分析智慧型深度類神經網路分類災害影像 模組之分類能力。
二	2.議題 5-2、5-3、6-2 未說明現有的工作進展，及未來執行步驟	<p>有關類類神經網路分類災害影像也有研究成果，請參閱下列議題</p> <ul style="list-style-type: none"> •5-2:模擬實驗分析智慧型深度類神經網路分類災害影像 模組之分類能力。 •5-3:影像檢索及分類模型成本分析。 •6-2:使用端影像檢索評估之研究。

項次	審查意見	回覆說明
三	3.議題 7-1，須利用問卷分析視覺偏好度對表徵因子及感知因子關聯，目前尚未看到問卷形式，未說明如何做前測、後測及一致性分析。	本研究不是利用問卷前、後測及一致性分析。主要參考[18]的研究成果，在[18]研究結果已經用問卷分析視覺偏好度對表徵因子及感知因子關聯，本研究主要希望用深度學習技術來推估表徵因子中的視覺偏好度的水視率及綠視率的估計，此研究成果請參閱第二章第三節的 七、水土保持設施工程美學照片之內涵分析之量化估計。
四	4.報告之第四章結論及建議太過簡單，僅用各四行文字帶過，P.4-2 最後一包話甚至未完，實須重新撰寫。	期末報告書中已經修正。
五	5.目前進度離計畫目標尚遠，請加強在水保工程內容之使用性，以免在期末無法達成計畫目標及交付預計成果。	已達成計畫目標，請參閱期末報告書。
六	6.請與水保工程人員多溝通，以符合本案為”應用研究”之屬性。	於收到期中報告審查意見後，請教水土保持局技術研究發展組的專家同仁。
七	7.中英文摘要及文章內文中 ICBR 及 CBIR 系統請再檢核確認。	期末報告書中已經更正。

項次	審查意見	回覆說明
八	<p>8.本計畫就期中成果與報告書所述，架構為影像辨識相關技術應用於水土保持局歷史影像平台的加值應用，藉由平台圖像文字等資料授權及 Api 方式，實作以下三點：一、圖像搜尋。二、物件識別與鏈結。三、RWD 介面。</p> <p>應用研究立意甚佳，惟規模過大，就階段成果而言，建議在搜尋流程功能，或是影像辨識分類這兩塊功能擇一深究，其餘部分列出研究與實作限制之相關討論，並實作出可操作之流程即可。並且因 AI、影像辨識、標籤分類等技術迭代速度極快，應用研究建議縮小執行規模後，整理上述相關階段之研究限制，影像辨識訓練結果之解析極限、異常結果等整理成結果與討論章節，以利未來相關領域參考應用。</p>	<p>遵照辦理。</p> <ul style="list-style-type: none"> •請參閱期末報告書第三章第三節的結果與討論章節(含研究限制)。 •請參閱期末報告書第二章第三節的六、行動化模組之研究或連到下列連結體驗操作。 <p>http://140.130.33.227:8887/</p>
九	<p>9.本計畫實作、主機架設於雲林之系統介面建議如下：水土保持局請以機關全銜稱之，農業委員會若有需要也請用全銜，並將實驗室 Logo 與水土保持局 Logo 同時呈現，另系統雖屬創新研究計畫及開放授權之資料再應用，但現行政府機關資訊系統需考量資料法源與正確性，平台名稱宜以創新研究計畫之成果為名，避免混淆與維運更新之權責。實作系統名稱建議更改：影像檢索平台</p> <p>關於說明頁籤內，再行補充本平台為行政院農業委員會水土保持局平台相關計畫之成果或實作說明、研究團隊參與同仁等細部說明，避免</p>	<p>系統已更正，請參閱期末報告書或連到下列連結體驗操作。</p> <p>http://140.130.33.227:8887/</p>

項次	審查意見	回覆說明
	誤解為水土保持局官方正式系統。連結至歷史影像平台部分全銜亦同。	
十	10.RWD 查詢流程，結構完成度佳，未來細節 UI 可精進。圖片超連結至歷史影像平台功能，建議可 HOVER 時上層加按鈕，避免整個圖片縮圖變成感應區，容易誤觸直接跳轉，手機版則可將按鈕置於內框文末。另一作法可將說明做 MsgBox，點選跳出後鎖定背景，或隔離感應區，並設計跳出鈕，以上操作建議。	系統已更正，請參閱期末報告書或連到下列連結體驗操作。 http://140.130.33.227:8887/

附錄三、Case 3 查詢結果案例說明



圖附錄 3-1 案例 1



圖附錄 3-2 案例 2



圖附錄 3-3 案例 3



圖附錄 3-4 案例 4



圖附錄 3-5 案例 5

附錄四、拍攝類別及 CCTV 災害類別影像

表附錄 4-1 拍攝類別中的 CCTV 的影像樣本

拍攝類別-CCTV 樣本	
 <p>水土保持局 (照片授權方式：姓名標示)</p>	 <p>水土保持局 (照片授權方式：姓名標示)</p>
 <p>水土保持局 (照片授權方式：姓名標示)</p>	 <p>水土保持局 (照片授權方式：姓名標示)</p>
 <p>水土保持局 (照片授權方式：姓名標示)</p>	 <p>水土保持局 (照片授權方式：姓名標示)</p>
 <p>水土保持局 (照片授權方式：姓名標示)</p>	 <p>水土保持局 (照片授權方式：姓名標示)</p>



表附錄 4-2 拍攝類別中的高空拍攝的影像樣本

拍攝類別-高空拍攝樣本













表附錄 4-3 CCTV 災害類別中的土石流影像樣本





表附錄 4-4 CCTV 災害類別中的設施崩塌影像樣本

災害類別-設施崩塌樣本	
	
	
	
	



表附錄 4-5 CCTV 災害類別中的山坡崩塌影像樣本

災害類別-山坡崩塌樣本	
	
	
	



表附錄 4- 6 CCTV 災害類別中的其他的影像樣本

災害類別-其他樣本	
	
	
	
	



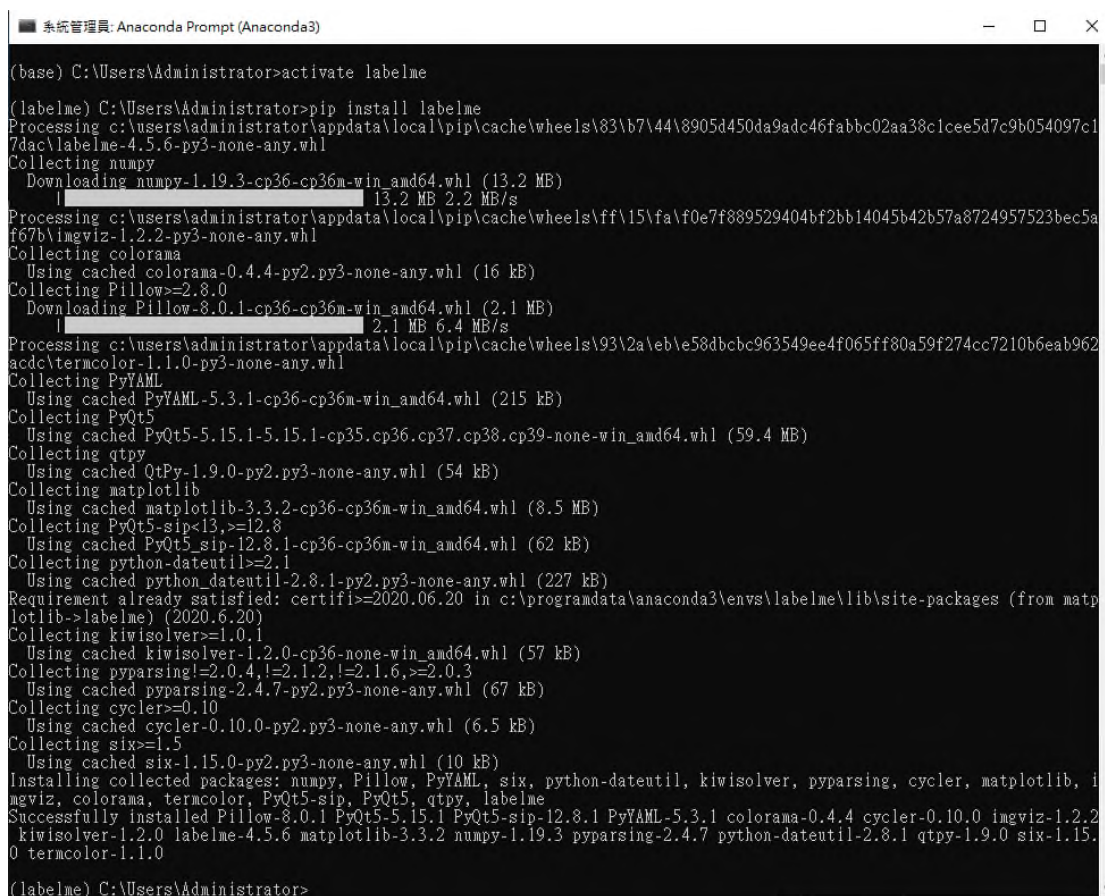
附錄五、Labelme 操作過程及說明

步驟一：打開 Anaconda Prompt->並輸入：

conda create --name=labelme python=3.6(建立新環境)

activate labelme(進入環境)

pip install labelme(下載 labelme)



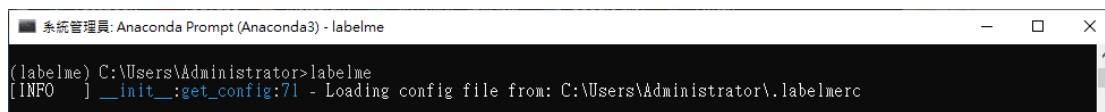
```
(base) C:\Users\Administrator>activate labelme

(labelme) C:\Users\Administrator>pip install labelme
Processing c:\users\administrator\appdata\local\pip\cache\wheels\83\b7\44\8905d450da9adc46fabbc02aa38c1cee5d7c9b054097c17dac\labelme-4.5.6-py3-none-any.whl
Collecting numpy
  Downloading numpy-1.19.3-cp36-cp36m-win_amd64.whl (13.2 MB)
    |#####| 13.2 MB 2.2 MB/s
Processing c:\users\administrator\appdata\local\pip\cache\wheels\ff\15\fa\f0e7f889529404bf2bb14045b42b57a8724957523bec5af67b\imgviz-1.2.2-py3-none-any.whl
Collecting colorama
  Using cached colorama-0.4.4-py2.py3-none-any.whl (16 kB)
Collecting Pillow>=2.8.0
  Downloading Pillow-8.0.1-cp36-cp36m-win_amd64.whl (2.1 MB)
    |#####| 2.1 MB 6.4 MB/s
Processing c:\users\administrator\appdata\local\pip\cache\wheels\93\2a\eb\58dbc963549ee4f065ff80a59f274cc7210b6eab962acd\termcolor-1.1.0-py3-none-any.whl
Collecting PyYAML
  Using cached PyYAML-5.3.1-cp36-cp36m-win_amd64.whl (215 kB)
Collecting PyQt5
  Using cached PyQt5-5.15.1-5.15.1-cp35.cp36.cp37.cp38.cp39-none-win_amd64.whl (59.4 MB)
Collecting qtpy
  Using cached QtPy-1.9.0-py2.py3-none-any.whl (54 kB)
Collecting matplotlib
  Using cached matplotlib-3.3.2-cp36-cp36m-win_amd64.whl (8.5 MB)
Collecting PyQt5-sip<13,>=12.8
  Using cached PyQt5-sip-12.8.1-cp36-cp36m-win_amd64.whl (62 kB)
Collecting python-dateutil>=2.1
  Using cached python_dateutil-2.8.1-py2.py3-none-any.whl (227 kB)
Requirement already satisfied: certifi>=2020.06.20 in c:\programdata\anaconda3\envs\labelme\lib\site-packages (from matplotlib->labelme) (2020.6.20)
Collecting kiwisolver>=1.0.1
  Using cached kiwisolver-1.2.0-cp36-none-win_amd64.whl (57 kB)
Collecting pyparsing!=2.0.4,!2.1.2,!2.1.6,>=2.0.3
  Using cached pyparsing-2.4.7-py2.py3-none-any.whl (67 kB)
Collecting cyclер>=0.10
  Using cached cyclер-0.10.0-py2.py3-none-any.whl (6.5 kB)
Collecting six>=1.5
  Using cached six-1.15.0-py2.py3-none-any.whl (10 kB)
Installing collected packages: numpy, Pillow, PyYAML, six, python-dateutil, kiwisolver, pyparsing, cyclер, matplotlib, imgviz, colorama, termcolor, PyQt5-sip, PyQt5, qtpy, labelme
Successfully installed Pillow-8.0.1 PyQt5-5.15.1 PyQt5-sip-12.8.1 PyYAML-5.3.1 colorama-0.4.4 cyclер-0.10.0 imgviz-1.2.2 kiwisolver-1.2.0 labelme-4.5.6 matplotlib-3.3.2 numpy-1.19.3 pyparsing-2.4.7 python-dateutil-2.8.1 qtpy-1.9.0 six-1.15.0 termcolor-1.1.0

(labelme) C:\Users\Administrator>
```

圖附錄 5-1 操作畫面 1

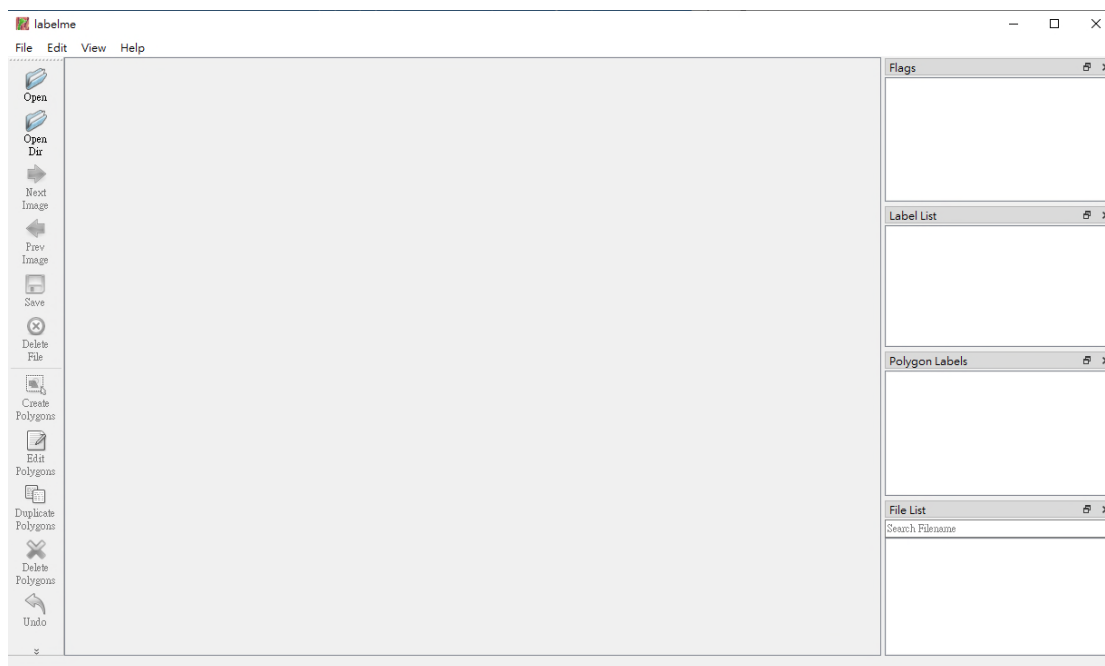
步驟二：進入 labelme



```
系統管理員: Anaconda Prompt (Anaconda3) - labelme
(labelme) C:\Users\Administrator>labelme
[INFO ] __init__:get_config:71 - Loading config file from: C:\Users\Administrator\labelmerc
```

圖附錄 5-2 操作畫面 2

此為 labelme 介面



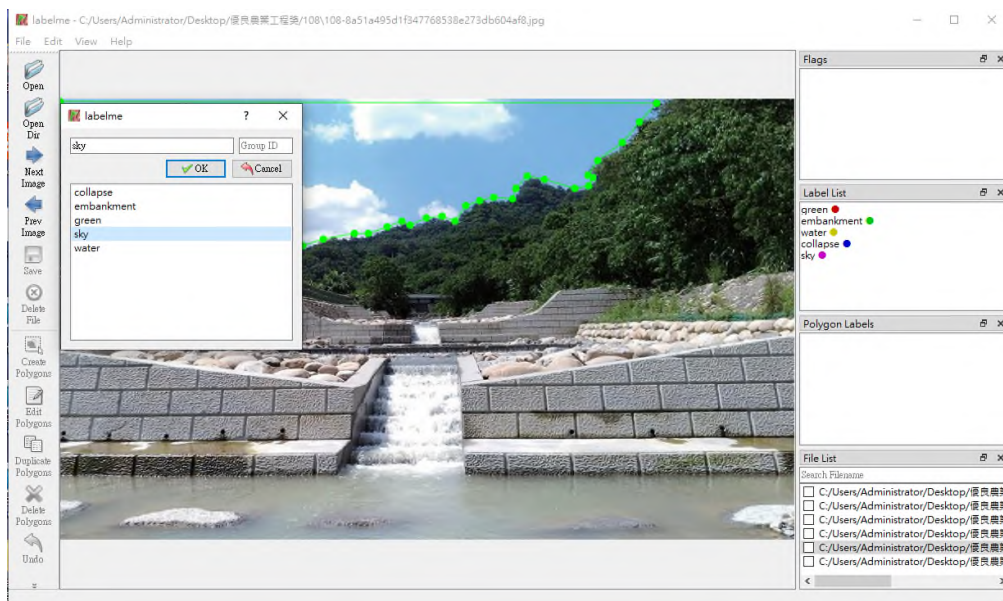
圖附錄 5-3 操作畫面 3

步驟三：找到一張圖片後放入資料夾，並使用介面左方的 OpenDir 將資料夾匯入



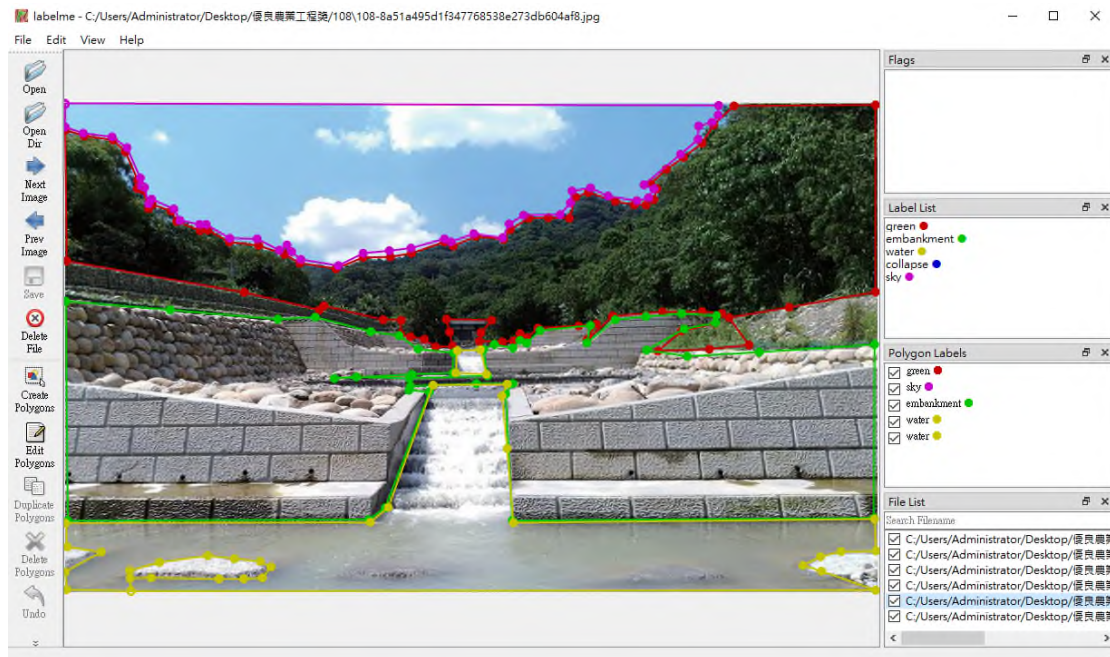
圖附錄 5- 4 操作畫面 4

步驟四：點擊介面左方的 CreatePolygons->劃出你要的圖片範圍->標註名字



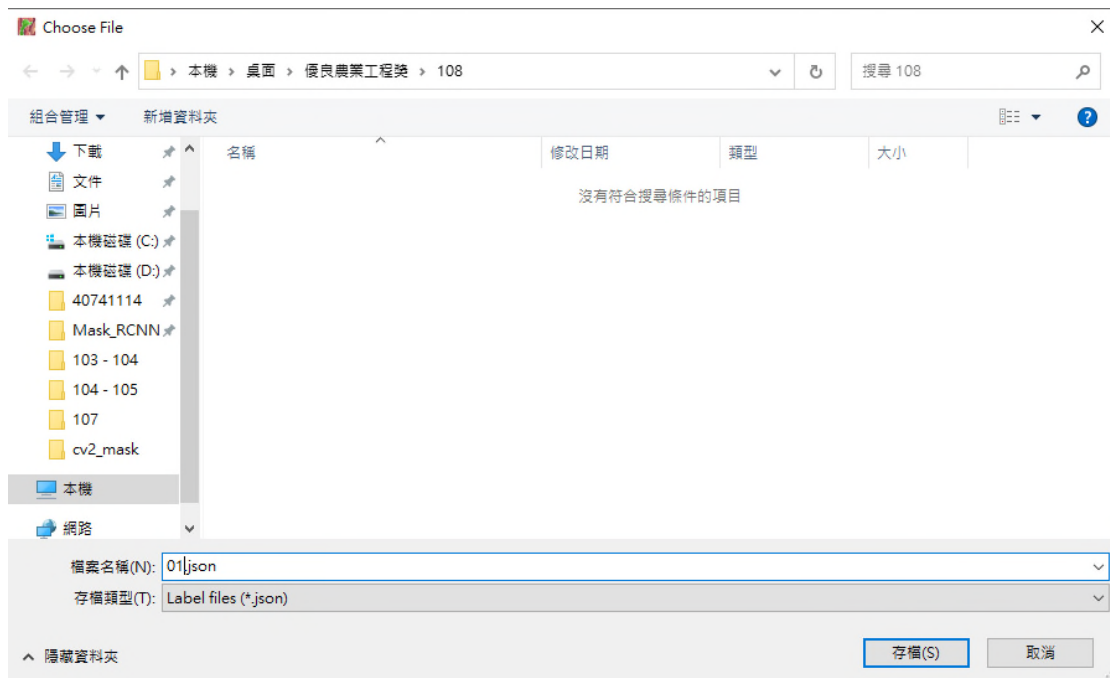
圖附錄 5- 5 操作畫面 5

步驟五：將水保圖片分為五個類別，分別為：green 樹木；草皮；草地、embankment 堤防、sky 天空、water 水、collapse 崩塌



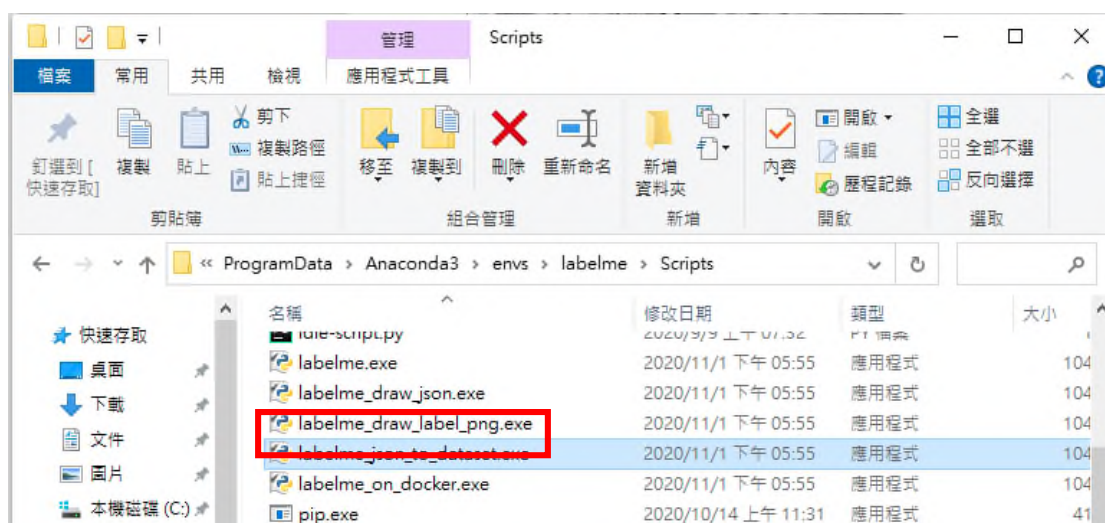
圖附錄 5- 6 操作畫面 6

步驟六：File->Save(副檔名為 json)



圖附錄 5- 7 操作畫面 7

尋找處理 label 產生的 json 檔，將 label 自帶處理檔 exe 加入環境變數，此檔案通常會在 anaconda 裡面



圖附錄 5- 8 操作畫面 8

步驟七：為了方便處理，將此處理檔加入至環境變數，先至本機，並右鍵內容



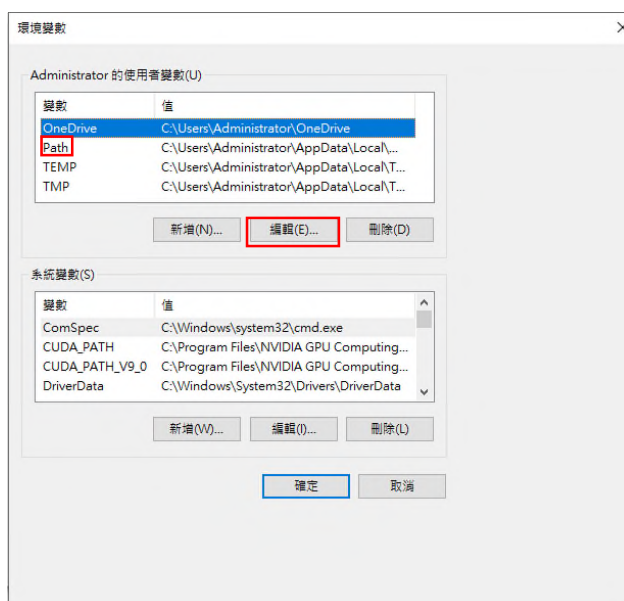
圖附錄 5- 9 操作畫面 9

打開進階系統設定



圖附錄 5- 10 操作畫面 10

打開環境變數->點 PATH->編輯->貼上處理檔路徑



圖附錄 5- 11 操作畫面 11

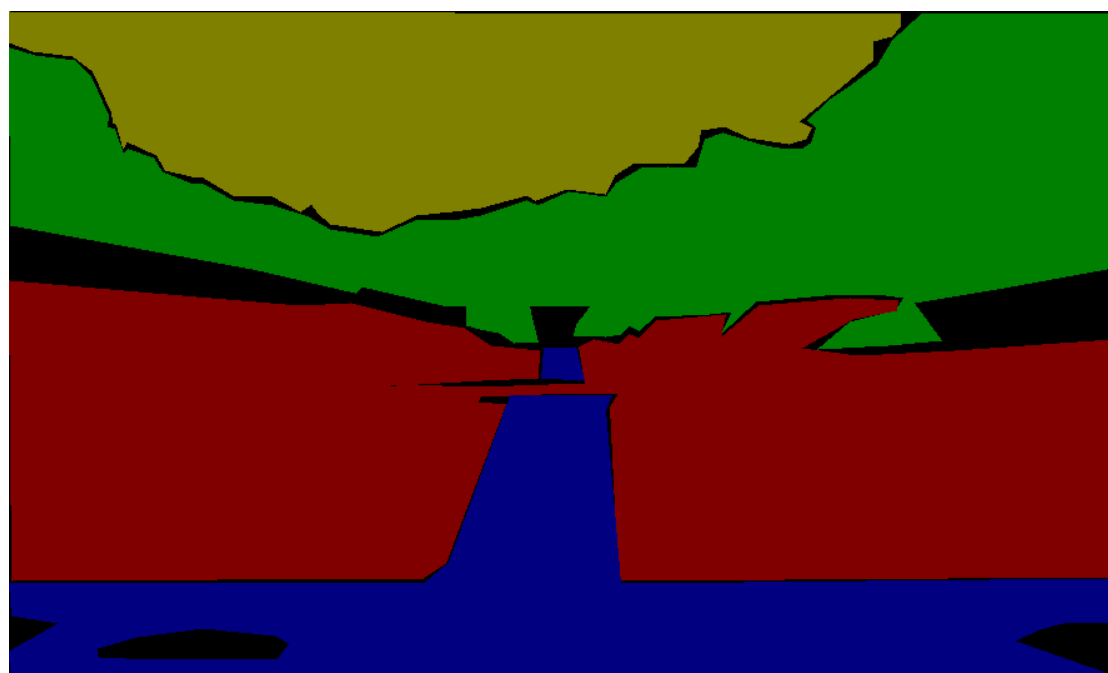
步驟八：打開 Anaconda Prompt→輸入 labelme_json_to_dataset 01.json



```
■ 獲取系統管理員: Anaconda Prompt (Anaconda3) - □ ×
(labelme) C:\Users\Administrator>cd C:\Users\Administrator\Desktop\優良農業工程獎\108
(labelme) C:\Users\Administrator\Desktop\優良農業工程獎\108>labelme_json_to_dataset 01.json
```

圖附錄 5- 12 操作畫面 12

出現 01.json 資料夾，裡面有被標註的圖片 png 檔




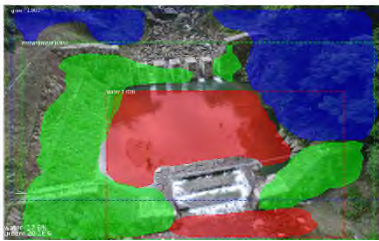

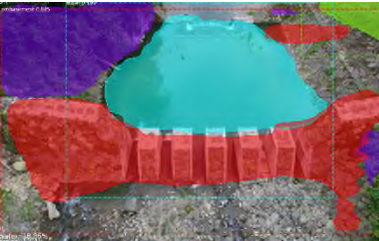

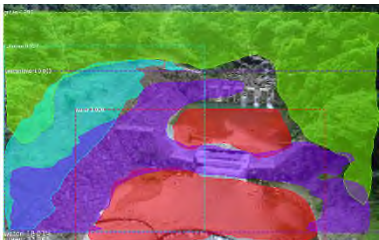

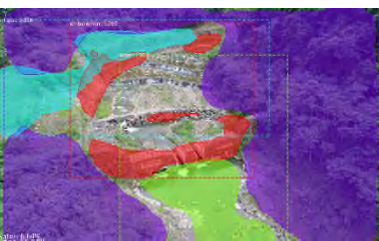

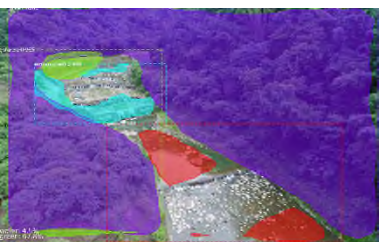
圖附錄 5- 13 操作畫面 13


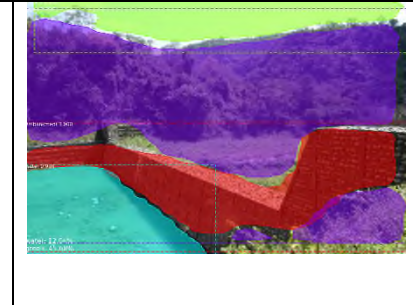

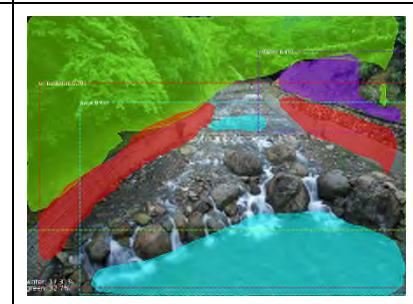

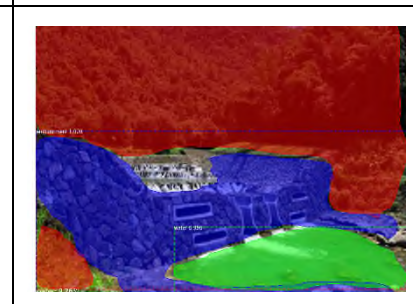

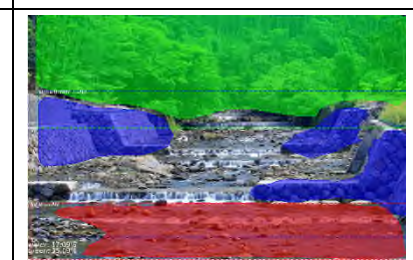

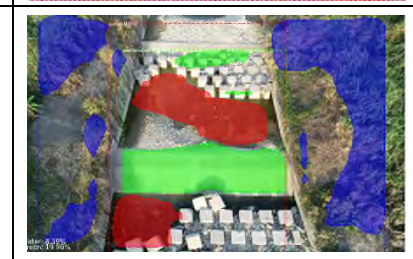

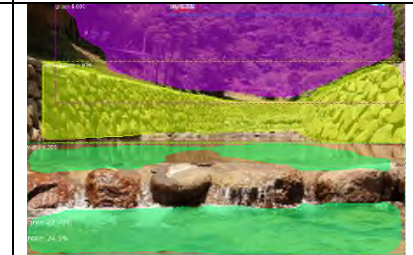



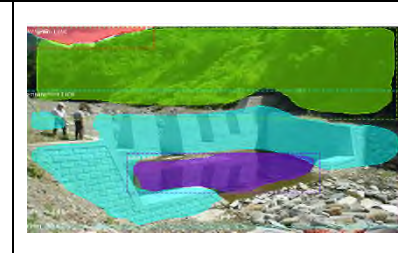

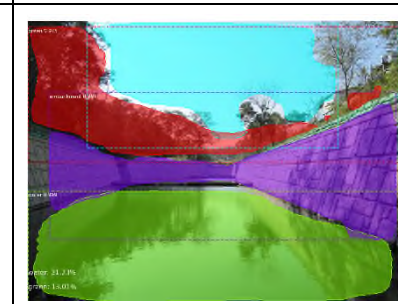

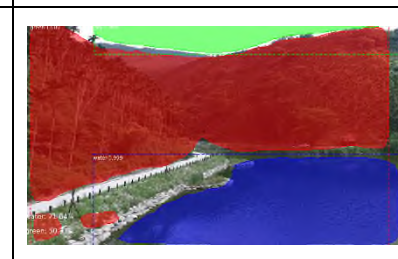

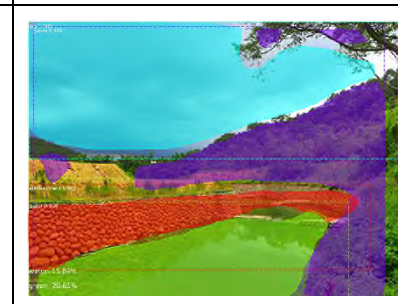

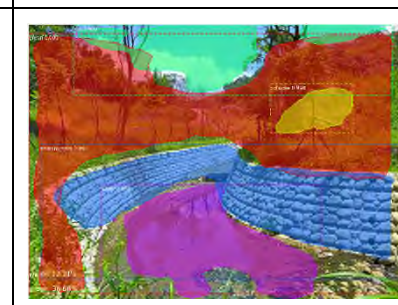

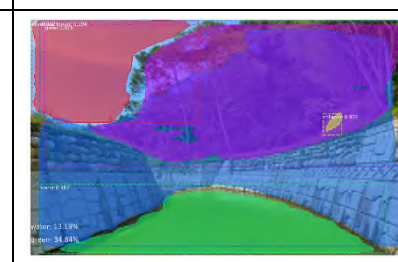
圖附錄 5-14 操作畫面 14

附錄六、實例分割影像與估計 WLR 與 GLR

表附錄 6-1 舉例說明原圖、實例分割結果及估計 WLR 與 GLR，
訓練影像集(含驗證影像集)。


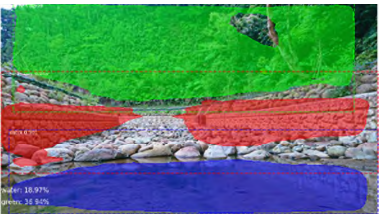

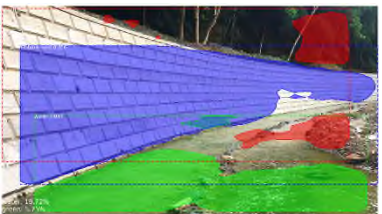



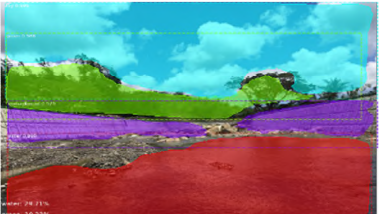


輸入影像	實例分割影像	WLR	GLE
		17.6%	20.16%
		18.86%	12.4%
		18.03%	33.36%
		6.64%	55.84%
		4.5%	67.8%




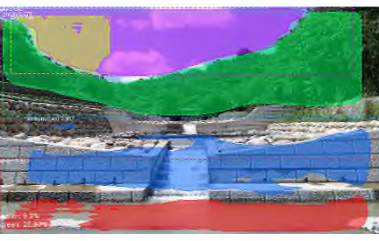

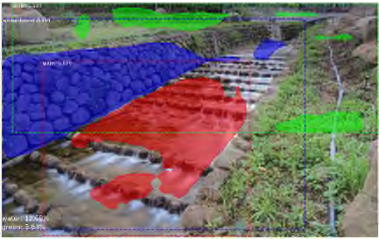

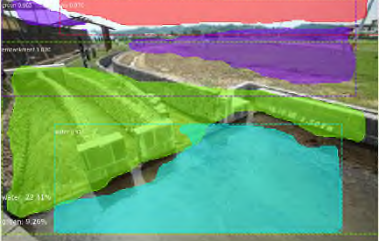

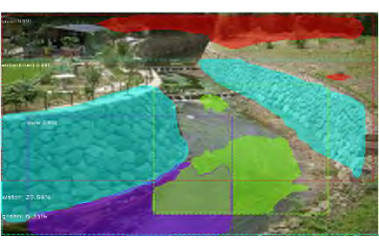

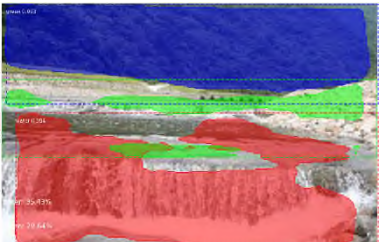
		<p>12.04%</p>	<p>45.68%</p>
		<p>17.31%</p>	<p>32.7%</p>
		<p>9.26%</p>	<p>48.27%</p>
		<p>17.09%</p>	<p>35.09%</p>
		<p>8.59%</p>	<p>19.98%</p>
		<p>27.71%</p>	<p>24.5%</p>

		<p>6.41%</p>	<p>30.67%</p>
		<p>31.23%</p>	<p>13.01%</p>
		<p>21.84%</p>	<p>50.87%</p>
		<p>15.84%</p>	<p>20.61%</p>
		<p>12.21%</p>	<p>38.68%</p>
		<p>13.19%</p>	<p>34.84%</p>

		14.83%	33.57%
---	--	--------	--------

表附錄 6-2 舉例說明原圖、實例分割結果及估計 WLR 與 GLE，測試影像集。

輸入影像	實例分割影像	WLR	GLE
		18.97%	36.94%
		18.72%	5.75%
		11.62%	35.92%
		29.71%	10.23%
		12.66%	28.48%

		<p>31.19%</p>	<p>10.42%</p>
		<p>9.9%</p>	<p>20.97%</p>
		<p>12.68%</p>	<p>3.84%</p>
		<p>22.41%</p>	<p>9.26%</p>
		<p>20.84%</p>	<p>6.35%</p>
		<p>35.43%</p>	<p>28.64%</p>

附錄七、期末報告計畫書審查意見

項次	審查意見	回覆說明
期中需求訪談內容與意見：		
一	1. 目前自動辨識功能，似乎是傳送至電腦辨識，在手機上，無訊號情況，是否已可用手機上的處理能力去辨識物件？	本計畫成果中，對於在手機上，無訊號情況是沒有支援此功能。因為影像(特徵)資料庫需龐大記憶容量，無法放置於手機記憶容量中，所以手機在無信號的情況，無法辨識物件。
二	2. 影像資料之判釋受影像品質、拍攝角度等多項因子影響，又災害類型並非單一類型，對於自動判釋工作相當不易，建議對於影像判釋提出未來執行的方向。	1.對於災害類型影像中，常有非單一災害類型，未來研究方向會研究多標籤(multi-label)類別影像內容分類器來克服非單一類型的情形。利用影像實例分割來判釋影像中的物件，來實現多標籤類別影像內容分類器是一個未來研究方向。 2.影像拍攝角度因子影響問題，可以利用影像處理模擬產生各類拍攝時的影響因子，例如拍攝角度、拍攝震動模糊、拍攝焦距等等，產生模擬影像集，來訓練影像內容分類器或影像內容檢索。另外，利用生成對抗網路來產生模擬影像集，也是另一項為來研究方向。