

# 整合無人機與深度學習之瀝青鋪面裂縫自動檢測研究

劉鎧銘<sup>1\*</sup> 高書屏<sup>2</sup> 王豐良<sup>3</sup> 林志憲<sup>4</sup>

## 摘要

本研究結合無人機與深度學習物件偵測模型 YOLOv9，針對道路瀝青鋪面裂縫進行自動化檢測與分析。利用無人機高機動性與高解析度影像獲取能力，可快速覆蓋大面積道路區域並減少人工巡檢所需時間與成本。研究中首先以公開鋪面裂縫資料集 CrackForest 進行 YOLOv9 模型訓練，經過 100 個訓練週期後，模型在測試集上達成 mAP50 為 0.891 及 mAP50-90 為 0.550 的檢測精度，顯示其在小物體辨識與道路裂縫檢測上的優勢。最後，透過影像正射化與幾何校正，將檢測結果套疊於真實坐標，並進行裂縫寬度量分析，驗證本系統裂縫寬度量測精度優於約為 0.16mm，可作為道路維護決策與管理之有效工具。

**關鍵詞：**無人機、YOLOv9、裂縫檢測、鋪面裂縫、深度學習

## 1. 前言

道路為連結城市與鄉村、促進人流與物流的重要基礎設施，其使用狀況亦反映國家整體發展與治理效率。道路品質更常常被世界銀行視為衡量國家經濟發展程度的指標之一(Ragnoli *et al.*, 2018)。隨著全球都市化進程加快，車輛擁有率與使用頻率持續上升，鋪面長期承受高載重壓力，加以極端氣候與環境變化的影響，使得道路損壞與裂縫情形日益嚴重。一旦鋪面破損未及時修復，將可能導致交通事故風險升高與運輸效率下降，對社會經濟活動造成實質性損失。目前道路常見的損害類型包括裂縫與坑洞，其中多數路面劣化皆起始於微細裂縫，隨時間推移與車流重壓逐漸擴展，最終演變為大面積剝離及結構性破壞。若能及早發現裂縫並進行維修，不僅可有效抑制鋪面進一步劣化，亦能節省後續高額的整修成本與人力資源，延

長瀝青路面之使用壽命。因此，定期執行鋪面裂縫巡檢已被證實為維持路面品質最經濟有效的方法之一(Gavilán *et al.*, 2011)，裂縫之及時檢測與

養護亦成為道路管理維護中的核心課題。

隨著道路使用強度不斷提高，鋪面劣化問題日益嚴重。早期的裂縫檢測多仰賴人工目視巡檢，然而，該方式不僅效率低下，且檢測結果高度仰賴人員經驗與主觀判斷，容易導致誤判或遺漏。此外，作業人員需在車流中行走亦存在安全風險，限制了人工巡檢在高交通流量區域的可行性。為解決上述限制，車載影像系統應運而生。此類方法通常結合高解析度攝影機與雷射光達等感測設備，搭配電腦影像處理演算法，能在車輛行進間自動捕捉並辨識路面裂縫，提升效率與準確性。Chambon *et al.*(2009) 使用馬可夫隨機場模型，以及 Tsai *et al.*(2010)採用 Otsu 閾值法，皆在複雜背景下達成穩定辨識表現。Gavilán *et al.*(2011)更藉由車載系統成功分類多種不同的道路，顯示該技術具備處理不同背景的能力。然而，此類系統建置成本高昂，且於實地操作時可能干擾交通流量，使其在高使用率道路上的應用受到限制。近年無人機(UAV)技術的成熟為裂縫檢測帶來新契機。無人機具備低成本、快速部署與高機動性的特性，可於不影響交通情況下取得大面積、

<sup>1</sup> 國立中興大學土木工程學系 碩士

<sup>2</sup> 國立中興大學土木工程學系 教授

<sup>3</sup> 健行科技大學應用空間資訊系 助理教授

<sup>4</sup> 國立中興大學土木工程學系 博士

\* 通訊作者, E-mail: daniel1900513@gmail.com

收到日期：民國 114 年 09 月 03 日

修改日期：民國 114 年 10 月 15 日

接受日期：民國 115 年 02 月 25 日

高解析度之道路影像，顯著提升資料蒐集效率。為克服高空拍攝導致的解析度下降與背景複雜問題，研究者開始結合影像處理與深度學習演算法以提升辨識準確度。Zhou *et al.*(2014)融合道路顏色與邊界資訊進行形態學處理，有效提升裂縫辨識率；Zhu *et al.*(2022)則透過調整無人機飛行參數，優化影像解析度與辨識品質。綜上所述，UAV 搭配先進影像演算法已成為目前裂縫自動檢測的重要趨勢，兼具高效率、低干擾與良好辨識能力，極具實務應用潛力。然而要處理大量的高解析度影像，若是運用人工去處理這些資料的話可能也會導致檢測人員在長時間的工作下產生視覺疲勞，進而影響判斷能力，故利用自動判別是必須的，它可以有效的降低資料處理速度，並減少人工辨識的主觀性。目標偵測深度學習技術分為單階段演算法和雙階段演算法，單階段的檢測方法主要包含了 YOLO(You Only Look Once)系列和 SSD(Single Shot MultiBox Detector)兩者，輸入影像後可立即得到影像辨識的成果。雙階段演算法包含了卷積神經網路(Convolutional Neural Network, CNN)，是採用區域提議網路(Region Proposal Network, RPN)進行物件的框選後再進行辨識。然而 Cha *et al.*(2017)提出基於 CNN 之影像級與像素級裂縫檢測方法，顯著提升辨識準確率與自動化程度，雖然雙階段的演算法在一開始就有較高的精度，但因為需要先進行區域選擇在判別，故辨識速度上面會相對的較久，需要速度快的道路巡檢可能較不適用。

所以本研究以無人機結合影片拍攝的方式，能全面性地捕捉需要修補的裂縫位置；加上訓練一套單階段的目標偵測模型運用在瀝青道路裂縫檢測上的適用性，並利用該模型能快速且自動的辨識出地面上的裂縫，降低人為的誤差和主觀性，進而在提高道路檢測速度與準確性的同時，後續再透過裂縫量化分析，有效掌握裂縫的發展狀況與變化趨勢，為道路檢測和維護提供新的思路與解決方案。

## 2. 研究方法

### 2.1 研究流程

本研究預計要結合無人機與深度學習技術的瀝青道路裂縫檢測系統，整體流程分為模型建立與應用兩大部分，如圖 1 所示，詳細步驟說明如下：

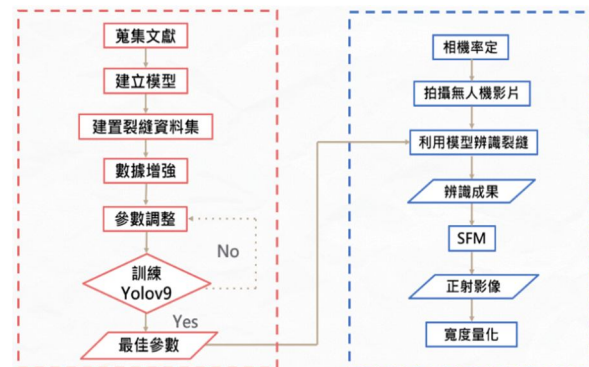


圖 1 研究流程圖

本研究首先進行相關文獻探討，根據目前路面裂縫偵測技術的發展現況，並分析常見之物件偵測模型與應用情境，基於這些以及文獻探討，選擇 YOLOv9 做為後續研究使用模型。第二部分會先用相機率定後消除相機所帶來的光學誤差，再利用影片拍攝的方式獲取研究標的數據，捕捉所需要修補的裂縫位置。利用訓練好的 YOLOv9 模型快速且自動的辨識裂縫，捕捉其位置，再將後續辨識成果較佳的影片提取有裂縫的幾幀影像，做後續量化使用。然而收集的無人機影像可能因無人機的姿態失真導致後續量化的精確性，故先利用 SfM 的技術將影像恢復成正射影像後，再用此影像做後續寬度量化，以確保其精確性，並提供道路養護一個可靠的評判標準。

### 2.2 研究資料收集與處理

#### 2.2.1 數據集介紹

本研究所使用的訓練資料集是使用 Crack Forest 的公開鋪面裂縫資料集，總共影像共有 10400 張，其影像大小為 600x600 像素，其鋪面裂縫為高解析度的車載拍攝路面裂縫，而它的裂縫標註形式如圖 2 所示，為利用許多 YOLO 訓練框所框出裂縫的輪廓。

10400 張裂縫影像依照 7:2:1 的比例隨機分配給訓練、驗證和測試，做後續模型訓練使用，訓練的

張數約為 7280 張，驗證的張數約為 2080 張，測試的張數約為 1040 張。

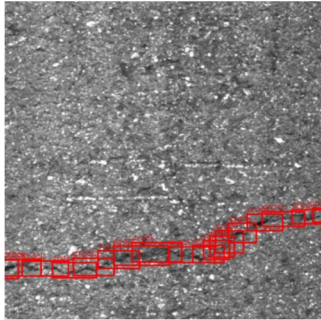


圖 2 裂縫資料集標註形式示意圖

### 2.2.2 測試數據

本研究測試影像由無人機 DJI Mavic 2 Pro 搭載攝影機 L1D-20c 進行拍攝。要進行實驗區域選擇前，首先要先知道目前道路裂縫檢測的規定為何，根據交通部公路總局所頒布的養護手冊規定(中華民國交通部，2020)，裂縫平均寬度在小於等於 6mm 時會被分為 L 級，其中在平均寬度大於 3mm 時會進行填封處理，故在選擇拍攝的實驗道路需要含有大於寬度 3mm 之裂縫，並預計選擇一塊背景相對乾淨的道路路面做研究使用，經挑選後選擇雲林縣的科工十六路，拍攝時間為 2025 年 4 月 4 號 14:00，此區域含有 3mm 寬度的裂縫，並且裂縫密集程度也較高。圖 3 為研究區域位置，框為所選擇的道路範圍。

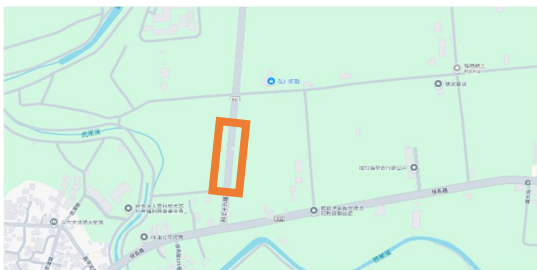


圖 3 研究區域圖

## 2.3 物件偵測網路

YOLO 模型為單階段偵測的物件偵測模型，其與二階段偵測的模型最大的不同為少了區域提議網絡(Region Proposal Network)技術。二階段偵測是採用區域提議網絡的技術，會先生成包含檢測目標的先驗框，然後以分類器判斷每個先驗框是否包含檢測目標及目標類別的信心值，調整信心的高低以防止低的誤判區域，然後得到檢測的結果。雖然生成候選區域再進行檢測的方式擁有較高的檢測精度，但檢測速度較慢。一階段架構直接於卷積神經網路生成特徵圖並進行預測框回歸並分類，雖然檢測精度會不如 RPN 的效果，但能大幅度的降低模型推論時間。而物件偵測的流程如圖 4 所示。

目標偵測模型的運作流程一般可分為四個主要階段，分別為 Input、Backbone、Neck 與 Head。首先，在 Input 階段，根據各模型架構需求，輸入統一尺寸影像作為分析資料。接著於 Backbone 層中，利用卷積神經網路對影像進行多層次特徵萃取，以保留影像中的關鍵語義與結構資訊。Neck 層則負責對來自 Backbone 的特徵進行進一步整合與壓縮，透過多尺度特徵融合機制提升模型對不同大小目標的辨識能力。最後，於 Head 層將融合後的特徵進行物件類別分類與邊界框回歸，以產出具定位資訊的預測結果。根據輸出策略不同，目標偵測模型可區分為 密集預測(Dense Prediction)與 稀疏預測(Sparse Prediction)。一階段模型如 YOLO 模型採用密集預測架構，於影像網格上設定大量錨點，並對每個網格進行同時的物件存在性與邊界框預測。相較之下，二階段模型如 Faster R-CNN 則透過 RPN 先選出潛在目標區域，再進行分類與精細定位，因此屬於稀疏預測方式，具有較高的準確性但推論速度較慢。

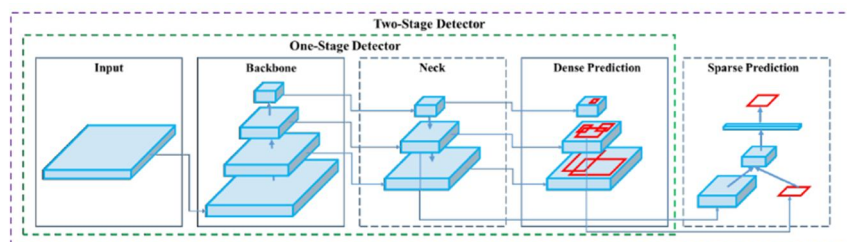


圖 4 物件偵測網路架構圖

### 2.4 YOLO 模型架構

首先 YOLO 最一開始的是由 Redmon *et al.* (2016)發表的論文，它的辨識流程如圖 5 所示是根據先將整張輸入影像會被劃分為一個  $S \times S$  的網格。每一個格子被指派負責偵測其中心點所落入的物件。每個格子會預測兩樣東西，一個是邊界框(bounding boxes)和信心值(confidence)，預測邊界框會先得知每個框包含位置座標(x, y, w, h)而信心值會給出每個一邊界框裡是屬於哪一個類別的可能性最高，其公式如式(1)。

$$\text{confidence} = p(\text{object}) * IOU_{pred}^{truth} \dots\dots (1)$$

$p(\text{object})$ 為物體存在的機率， $IOU_{pred}^{truth}$ 表示預測邊界框與地面真實邊界框，接下來加上邊界框和信心值的搭配顯示出機率分布圖。最後將這些系統將這些預測結果進行融合與非極大值抑制(Non-Maximum Suppression, NMS)，產生出最終的物件偵測框與其類別。

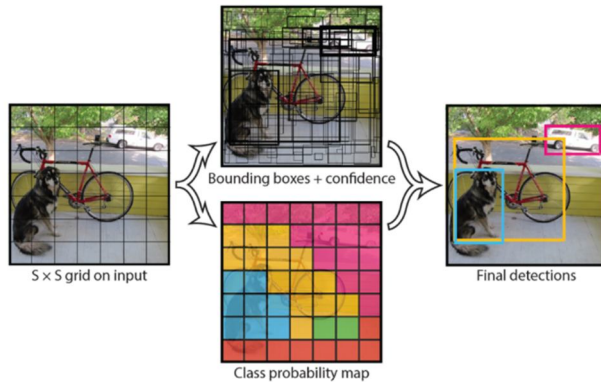


圖 5 YOLO 模型物件偵測圖

而由 Wang *et al.* (2024)發表 YOLOv9 是 YOLO 模型有國際期刊認證的最新模型，與 YOLOv7 的是同一位作者發表的，故在程式碼的邏輯大部分與 YOLOv7 是相同的，YOLOv7 已經可以達到相當高的精度了，所以 YOLOv9 提出方法解決深度神經網路，通过整合可程式梯度資訊(Programmable Gradient Information, PGI) 和多功能廣義高效層聚合網路(GELAN)架構，YOLOv9 不僅增強了模型的學習能力，確保了在整個檢測過程中保留關鍵信息，實現卓越的準確性和性能，解決訓練時遺失的許多

原始資訊。

圖 6 為 YOLOv9 的架構圖，採用由深至淺的模組化設計，其架構可分為 Backbone、Neck 與 Detection Head 三大部分。輸入影像首先透過一系列卷積層與可重參數化模組 RepNCSPELAN4 進行特徵提取，強化局部與全局資訊的捕捉，構成主幹網路(Backbone)。接著引入 GELAN 上下採樣結構以融合不同層級特徵，Neck 部分則整合語意與細節資訊，進一步強化特徵表達能力。偵測頭(Head)設有三個輸出尺度(P3、P4、P5)，對應於不同大小物件之偵測任務，每一路徑皆包含邊界框回歸(conv-reg)與分類(conv-cla)分支。另於 Backbone 中段引入 CBLIner 與 CBFuse 模組進行通道壓縮與特徵融合，兼顧模型表達力與運算效率。整體而言，YOLOv9 融合輕量化、多尺度與模組化設計，展現優異的即時目標偵測性能。

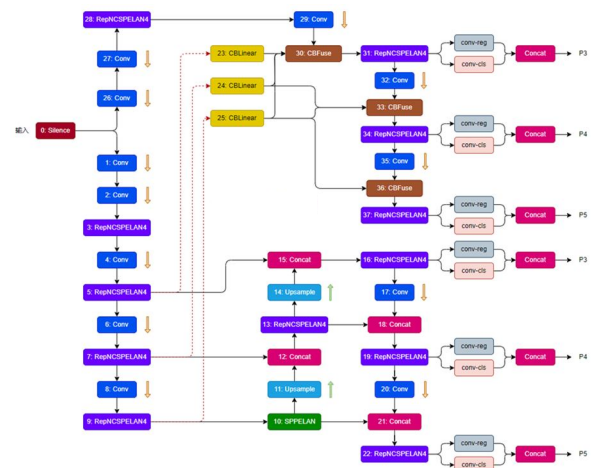


圖 6 YOLOv9 架構圖

目前 YOLOv9 運用在醫學和農業上，以 Chien *et al.* (2024)利用 YOLOv9 模型應用到骨折偵測任務中，作為電腦輔助診斷，幫助放射科醫師和外科醫師解讀 X 光影像，mAP50-95 值提升了 3.7%，可以有有效的幫助一些缺乏豐富經驗的醫生提早熟悉並解讀 X 光影像。而農業則是以 Vo *et al.* (2024) 用 YOLOv9 檢測成熟的番茄，相較於其他模型有較佳的表現，說明了對相對小的目標的可用性。故本研究也將運用此模型於鋪面裂縫檢測上。

### 2.5 模型性能評估指標

評估模型的指標為混淆矩陣，其示意圖如圖 7 所示，它是評判模型好壞的一個重要指標其中它將情況分為實際現地的情形和模型訓練完成所產生的情形，而這兩個經過排列組合可以分為 4 種不同的情況，分別為真陽性(True Positive, TP)、偽陽性(False Positive, FP)、真陰性(True Negative, TN)以及偽陰性(False Negative, FN)。利用這四種不同的情形可以去算出以下指標，分別為精確率(Precision)、召回率(Recall)、平均精確度(Mean Average Precision)。

混淆矩陣 (Confusion Matrix)	實際○	實際×
預測○	TP (True Positive)	FP (False Positive)
預測×	FN (False Negative)	TN (True Negative)

圖 7 混淆矩陣示意圖

精確率表示在所有被模型判定為正類的樣本中，實際為正類的比例，反映模型在正向預測中的準確程度。其計算方式為 TP 與預測為正類總數 TP + FP 之間的比值，如式(2)所示。

$$Precision = \frac{TP}{(TP+FP)} \dots\dots\dots (2)$$

召回率又稱為查全率，用以衡量模型對實際正類樣本的識別能力。其定義為真正例 TP 與實際正類總數 TP + FN 之間的比值，亦即在所有實際為正類的樣本中，被正確預測出來的比例，如式(3)所示。

$$Recall = \frac{TP}{(TP+FN)} \dots\dots\dots (3)$$

平均精確率為各類別平均精度 (Average Precision, AP) 之加權或非加權平均，用以綜合評估模型在多類別偵測任務中的整體性能。AP 則是指

由精確率與召回率所構成曲線下的面積，代表在不同召回率下模型的精準表現，其計算方式如式(4)。

$$AP = \int_0^1 p(r)dr \dots\dots\dots (4)$$

公式中變數 r 沿著精確率(Precision)曲線所進行之積分，亦即計算 Precision-Recall 曲線下之面積。

而 mAP 則是將所有類別所對應之 AP 值進行加總後取平均，反映模型在整體偵測任務中的平均表現，其計算方式如式(5)所示。

$$mAP = \frac{1}{k} \sum_i^k AP_i \dots\dots\dots (5)$$

其中，k 為類別的數量。

## 2.6 相機率定

相機率定(Camera Calibration)是計算相機在拍攝影像時，其內部幾何與光學參數以及相機與世界坐標系之間空間關係的過程。透過率定，可以將影像中的像素座標轉換為實際空間中的位置，並消除鏡頭造成的幾何畸變，使影像更符合真實世界的幾何特性。

因為本研究是利用影片的方式去拍攝道路，故利用影片方式率定，以確保後續研究拍攝時的像素等參數會是一致的。在把影片每 30 幀擷取一張影像，總共有 183 張影像。因為無人機的相機會自動追蹤物體，尤其在無人機做上下移動時特別明顯，故在拍攝時不適合上下移動無人機去拍攝率定板，故將影片分為三組去拍攝，利用上、中和下的拍攝角度以確保率定時有不同方位的拍攝角度去涵蓋，藉此提高率定的精度，圖 8、圖 9、圖 10 是率定拍攝不同角度的示意圖。

表 1 為率定完成的畸變差參數。



圖 8 由上往下拍攝示意圖



圖 9 置中拍攝示意圖

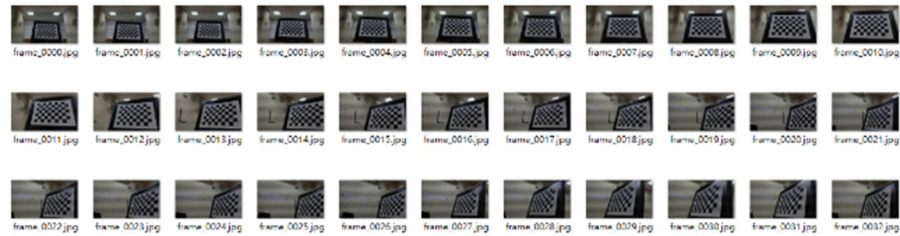


圖 10 由下往上拍攝示意圖

表 1 畸變差參數表

畸變差參數	
輻射畸變差 (k1, k2, k3)	(-0.0013, 0.0191, -0.0002)
偏心畸變差 (p1, p2)	(-0.0011, 0.0439)

## 2.7 SfM

運動回應結構(Structure from Motion, SfM)是一種基於多視角 2D 影像之計算機視覺技術，能在無需事先率定的情況下自動推估每張影像拍攝當下之相機內外方位參數，進而重建場景的三維結構。SfM 整合了特徵點偵測、特徵匹配、相機姿態估計與三維點雲重建等關鍵步驟，透過視差分析與幾何反算，實現由影像至模型的立體重建流程。其基本原理如圖 11 所示。

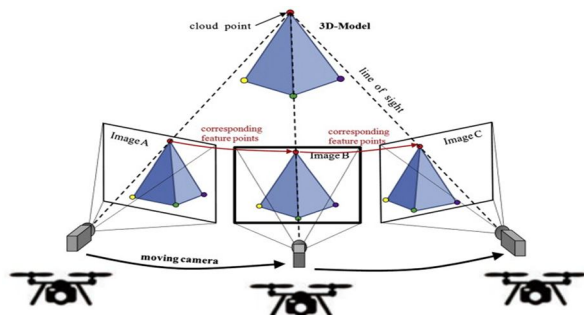


圖 11 SfM 演算法示意圖(楊明德&蔡慧萍, 2018)

在 SfM 中每張輸入影像需先透過特徵擷取技術偵測出具代表性的關鍵點。其中，Lowe(2004)提出的 SIFT(Scale-Invariant Feature Transform)演算法，因其對平移、旋轉與尺度變化的強大不變性及對光照與視角變形的容忍力，成為 SfM 中最具代表性的

特徵點擷取技術之一，它的步驟如下所示，分為四個步驟：

- (1) 尺度空間極值偵測：建立影像之尺度空間，並以高斯差分(DoG)逼近拉普拉斯運算，用以尋找不同尺度下之極值點，其公式如式(6)所示。

$$D(x,y,\sigma)=[G(x,y,k\sigma)-G(x,y,\sigma)]\times I(x,y) \\ =L(x,y,k\sigma)-L(x,y,\sigma) \dots\dots\dots (6)$$

- (2) 透過泰勒展開精細估算位置與尺度，並剔除邊緣或低對比特徵點，以提升匹配穩定性。
- (3) 為了讓特徵點具備旋轉不變性，SIFT 演算法會根據每個關鍵點鄰域內像素的梯度方向與強度來指定其主方向，其公式如式(7)、式(8)所示。

$$m(x,y) = \sqrt{L_x^2 + L_y^2} \dots\dots\dots (7)$$

$$\theta(x,y) = \tan^{-1}\left(\frac{L_y}{L_x}\right) \dots\dots\dots (8)$$

其中 $\theta(x,y)$ 為在位置 $(x,y)$ 的像素灰階變化的方向角度。 $\frac{L_y}{L_x}$ 為 y 與 x 方向上的灰階差比值。

- (4) 為了進一步強化特徵的不變性，SIFT 在每個關鍵點周圍建立一個 16x16 的區域，並將其劃分為 4x4 的子區塊。每個子區塊中會計算一組 8

個方向的梯度直方圖，最終將所有子區塊的直方圖組合為一個長度為 128 的特徵向量。此描述子具備平移、旋轉與尺度的不變性，並對光照變化與仿射變形具備良好魯棒性，有助於進行準確且穩定的特徵匹配。

## 2.8 U-net

U-Net 是由 Ronneberger *et al.* (2015)提出的一種卷積神經網路架構，針對 FCN 架構進行優化，使其在保持高精度定位的同時，更有效地提取影像的上下文資訊。特色是其能在訓練資料量有限的情況下，仍展現出優異的分割效能，提升模型對細微結構的辨識能力。圖 12 為 U-Net 的模型架構圖，它主要分為三個部份。

### (1) 編碼器(Encoder)：

位於 U-Net 架構的左側，編碼器功能類似於卷積神經網路中的特徵提取器，由多層卷積層、非線性激活函數、池化層與下採樣操作所組成。其目的為逐步壓縮輸入影像的空間尺寸，並提取高層次抽象特徵。隨著層數增加，特徵圖的解析度降低、通道數增加，能有效捕捉全局語意資訊與關鍵區域特徵。

### (2) 解碼器(Decoder)：

解碼器位於網路右側，負責將經過編碼器壓縮後的特徵圖逐層恢復至原始影像尺寸。解碼過程中通常採用上採樣或反卷積操作，並在每一層結合來自編碼器對應層的特徵，逐步重建空間細節。此機制有助於提升模型在邊界與小區域的分割準確性。

### (3) 跳躍連接(Skip Connections)：

U-Net 最大的特點在於其引入跳躍連接機制，將編碼器中各層的特徵圖直接傳遞至解碼器對應層。此設計能有效補償下採樣過程中損失的空間細節，使得解碼器在重建過程中能同時參考高階語意特徵與原始解析度特徵，進而提升分割結果的精確性與邊界還原能力。

## 3. 研究成果與分析

### 3.1 模型環境與超參數

本研究使用 Anaconda 建置環境去訓練 YOLOv9 模型，利用的環境與電腦配備參數如表 2、表 3 所示。

表 2 電腦配備表

電腦配備	
作業系統	Windows 10 專業版
處理器(CPU)	Intel® Core™ i7-8700K CPU@3.70GHz 3.70GHz
顯示卡(GPU)	NVIDIA GeForce GTX 1060 6GB
記憶體(RAM)	40.0 GB
tqdm	4.66.5

表 3 模型訓練環境參數表

環境參數	
套件項目	版本
CUDA	11.6
Matplotlib	3.7.5
Numpy	1.24.4
Pillow	9.5.0
Python	3.8.19
Opencv-python	4.10.0.84
Spyder	5.5.1
Torch-gpu	1.13.0+cu116
Torchvision	0.14.0+cu116
tqdm	4.66.5

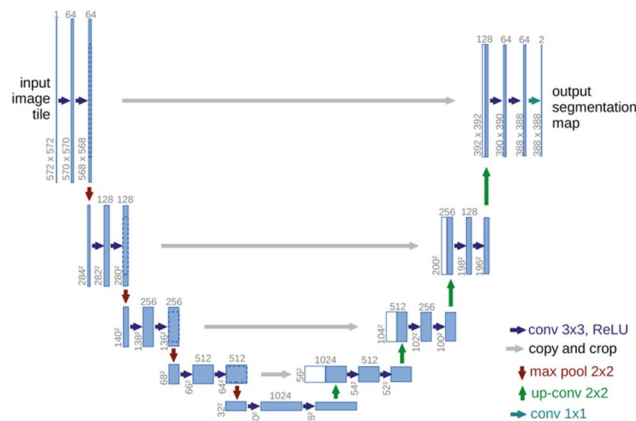


圖 12 U-Net 模型架構圖

在機器學習與深度學習模型中，超參數 (Hyperparameters)是在訓練過程開始前所設定的參數，其值不會透過模型學習自動調整，而是需透過經驗或實驗進行手動設置或調參。這些參數對模型的學習效率、收斂速度與最終準確度具有顯著影響。而表 4 是本研究訓練 YOLOv9 所使用的超參數表。

表 4 模型訓練超參數表

超參數	值
Initial learning rate 初始學習率	0.02
Final learning rate 最終學習率	0.02
Batch size 批次大小	2
Momentum 動量	0.937
Weight_decay 權重衰減	0.0005

### 3.2 飛行航高比較

本研究所需要偵測的裂縫寬度最小值為公路局養護手冊所頒佈 3mm 大小，然而為了需要有效的觀測到路面上的裂縫，必須要控制無人機的飛行高度讓相機所拍攝的影片有足夠的解析度，故利用公式(9)推導極限航高。

$$GSD = \frac{H \times p}{f} \dots\dots\dots (9)$$

其中 GSD(Ground sample distance)是地面採樣距離，為衡量解析度的指標， $H$  為航高、 $p$  為感光元件大小、 $f$  為鏡頭焦距。然而相機出場後感光元件大小可能會有失真等誤差存在，根據 Bandini *et al.* (2017)的算法，利用鏡頭視角來計算所需偵測的寬度，故本研究依照此參數來替代感光元件以計算極限航高，公式如(10)所示。

$$H = \frac{W \times GSD}{2 \times \tan \frac{HFOV}{2}} \dots\dots\dots (10)$$

因為要捕捉 3mm 的裂縫寬度，故設定其在影像上需要最少有 1 pixel，故 GSD 為 3(mm/pixel)，而 HFOV 為鏡頭視角，也就是本研究使用的相機規格表所得出的 77°， $W$  為影像寬度，其大小為 5472pixels，經計算後可得到極限航高大約為 10.32 m，後續選取道路上在裂縫上的白色石頭，上面最寬的裂縫，現實寬度約為 10.5mm 的裂縫，將影像展圖在小畫

家上並找出其位置，可以發現他約為 3.5 個 pixels，說明這組航高計算是可以當捕捉 3mm 裂縫寬度的極限航高使用。

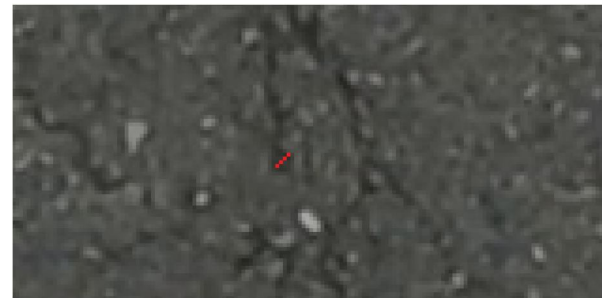


圖 13 裂縫影像示意圖

然而 10.32m 為可以辨識到研究所需的裂縫最小值，但是此解析度可能在 YOLO 模型裡的辨識率可能會相當不好，故本研究選取了三組航高去比較航高對於道路檢測的效率和辨識率。選取了三組航高包含了 10m，也就是極限航高，然後在選取了 7m 和 8.5m，其中因台灣有些道路有電線縱橫交錯的關係，7m 的航高剛好約為 2 層樓，可以高過路邊電線以防無人機被電線擋住路線造成道路巡檢時的安全疑慮。在 7m 和 10m 的中間取一個中間值確保兩者比較的值是有可信的，而非人為或是模型的表現浮動所導致的。然後將這三組航高所拍攝的影片進行全幅的影像拼接，對三組航高進行影像提取，以每十幀擷取一張影像以保持地面的資訊，7m 的航高總共擷取了 90 張影像，8.5m 的為 66 張影像，10m 的總共有 109 張影像，這些影像都是 3840x2160 的 jpg 檔，拼接的成果圖如圖 14 所示。

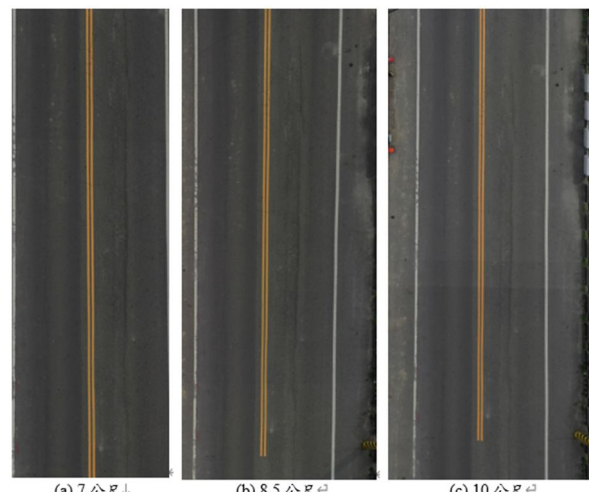


圖 14 全幅影像圖

為了要比較模型與實際辨識裂縫的準確性，首先要先找出真值以做比較，故利用全幅影像進行人工的裂縫標註，將人眼能辨識的裂縫框選出來，圖 15 則是人工框選後的結果。

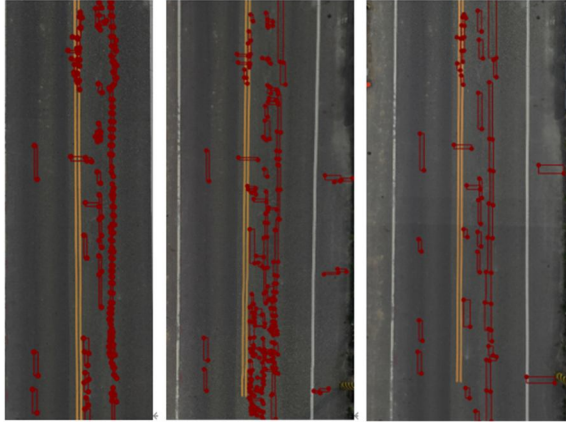


圖 15 人工標註的全幅影像圖

利用影片所有出現的候選框與上列三組航高進行 IOU 的計算，得到不同航高所得到的成功辨識率，將所有影片以及圖 15 成果進行候選框重疊率進行計算，其中以 7m 的航高所得到的影片的辨識率是最高的，為 85%，比 10m 航高的準確率高了 12%，而 8.5m 的為 75%，故後續選擇 7m 航高做後續的研究。

### 3.3 YOLOv9

以下為 YOLOv9 模型經過 94 小時訓練 100 個 epoch 的損失圖，其損失函數與效能指標如圖 16 所示。

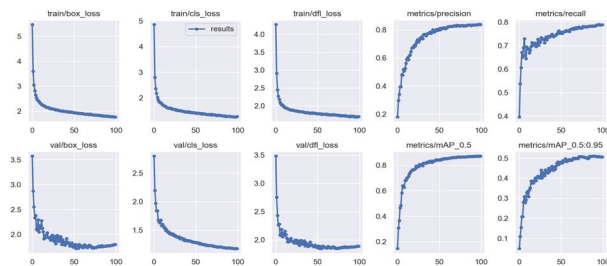


圖 16 YOLOv9 訓練損失圖

模型經過 100 個 epoch 的訓練，從損失曲線可觀察到，無論是訓練或驗證階段，box\_loss(邊界框損失)、cls\_loss(分類損失)與 dfl\_loss(分佈式 Focal Loss)皆呈現穩定下降趨勢，顯示模型在空間定位與

類別預測上均具良好收斂性。在評估指標方面表現如表 5 所示。

表 5 模型訓練指標表

Precision	Recall	mAP50	mAP50-95
0.845	0.815	0.891	0.55

Precision(準確率)與 Recall(召回率)皆隨訓練輪數增加而穩定上升，最終分別穩定於約 0.85 與 0.80，顯示模型在降低誤檢的同時，亦能有效涵蓋多數裂縫目標。mAP50 達到約 0.89，代表在 IoU 門檻為 0.5 時，模型已能準確定位裂縫之邊界框。然而，mAP50-95 進一步綜合評估模型在多個 IoU 門檻(0.5 至 0.95，每 0.05 一階)下的整體偵測表現，其數值穩定提升並收斂至約 0.55，顯示模型不僅能成功偵測裂縫位置，亦具備一定程度的邊界框定位穩定性。考量瀝青鋪面裂縫本身形狀細長且邊界不規則，高 IoU 門檻下對邊界框重疊要求更為嚴格，因此 mAP50-95 相較 mAP50 略低屬合理現象，亦反映模型在實際道路場景中仍能維持良好且穩定的整體偵測能力。

後續選取一塊鋪面進行模型的測試，圖 17 裡的一幀影片辨識，可以看到大多數的裂縫都有被辨識出來，說明模型的可行性。



圖 17 實驗區域辨識成果圖

### 3.4 SfM

為了評估對整體為確定後續量化精度準確性，本研究採用 Structure from Motion 技術進行三維重建與正射影像產製，其主要流程可分為以下四個。

#### (1) 特徵點選取

由多張 UAV 影像中提取局部不變特徵點，使

用 SIFT 演算法。每張影像的特徵資料儲存為 .pkl 格式，記錄了每一幀影像中關鍵點的描述子資訊，為後續匹配與重建提供基礎。

### (2) 特徵匹配與稀疏點雲重建

透過跨影像特徵匹配，並利用兩兩影像進行相對姿態估計與三角測量，逐步擴展場景點雲，產生稀疏三維模型。如圖 18 所示。

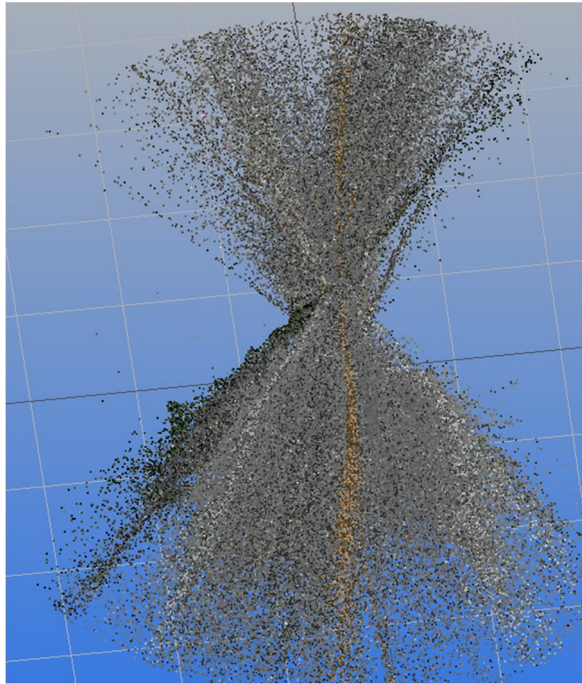


圖 18 粗點雲示意圖

```

frame_0000
R:
1.00000000 0.00000000 0.00000000
0.00000000 1.00000000 0.00000000
0.00000000 0.00000000 1.00000000
t:
0.00000000 0.00000000 0.00000000

frame_0001
R:
0.99999981 0.00010334 0.00060415
-0.00010318 0.99999996 -0.00025445
-0.00060418 0.00025439 0.99999979

frame_0002
R:
-0.99999997 -0.00016789 0.00015287
0.00016811 -0.99999998 0.00141554
0.00015263 0.00141537 0.99999899

frame_0003
R:
-0.99999853 -0.00031956 -0.00168644
0.00031463 -0.99999367 0.00292482
-0.00168737 0.00292429 0.99999430

frame_0004
R:
-0.99999983 -0.00020272 -0.00055204
0.00020136 -0.99999697 0.00245468
-0.00055254 0.00245457 0.99999683

frame_0005
R:
-0.99999966 -0.00023517 -0.00079270
0.00023343 -0.99999757 0.00219359
-0.00079322 0.00219340 0.99999728

```

圖 19 相片外方位參數圖



圖 20 正射影像圖

### (3) 相機外方位參數估算

在 SfM 過程中，同步求解每一張影像的外部參數(旋轉矩陣  $R$  與平移向量  $t$ )，建立全域相機位置與姿態資訊，如圖 19 所示，這些參數為後續多視角幾何建模與正射影像糾正提供依據。

### (4) 正射影像生成

結合已校正的相機姿態與三維點雲，將原始影像重新投影至平面座標系中，生成無透視變形的正射影像成果，如圖 20 所示。此成果可用於後續量測、比對與分析，具備高解析與幾何精度。

## 3.5 寬度量化

經過 YOLOv9 快速辨識後，可以得到他的類別 (crack) 和置信分數(小數點)，可以快速的抓出有裂縫的區域，將不重要的背景去除，留下裂縫的區域做後續量化之研究。圖 21 為擷取後的裂縫區域。

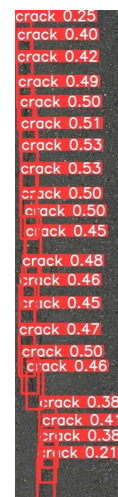


圖 21 裂縫辨識區域圖

擷取完後將辨識框進行去除，把完整的道路裂縫呈現出來後，然後進行 U-net 模型的遮罩提取，以利於做裂縫寬度量化，但是此圖所擷取出來的裂縫因為範圍太少讓模型無法找到影像上下文的資

訊，會使裂縫區域偵測不出來，導致輸出成全黑的遮罩圖，所以本研究將裂縫區域擷取的擴大了左右的圖片擷取，讓模型有足夠的影像大小去讀取上下文的資訊，能夠順利的產出遮罩圖，後續再利用遮罩圖去算平均裂縫寬度大小。圖 22 為擴大後的影像，圖 23 為 U-Net 所得出的遮罩圖。



圖 22 擴大後裂縫擷取圖

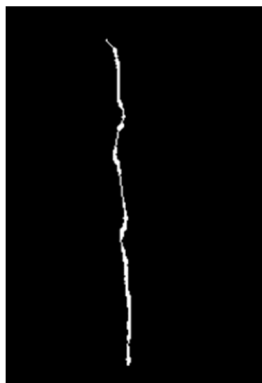


圖 23 U-net 辨識完遮罩圖

將裂縫的輪廓遮罩擷取出來之後，本研究參考 Miao & Srimahachota (2021) 提出的算法進行裂縫寬度的計算，該方法首先於裂縫骨架中選定一觀測點  $P$ ，並以此點為中心，向前後各取固定長度建立局部鄰域範圍。如圖 24，透過鄰域內的骨架點進行線性回歸分析，以獲得該區域裂縫的主要延伸方向，並建立一條擬合直線。接著，構建一條垂直於擬合線  $l$  的法線，並尋找此法線與裂縫邊緣的交點。最終，將兩點間的垂直距離作為該處裂縫寬度之估計值。

其中  $S$  為裂縫輪廓， $\delta$ 、 $\delta$  為鄰近值距離， $l$  為擬合線， $l'$  為法線， $P$  和  $P'$  為法線方向到裂縫輪廓距離。

然而根據此算法會算出各個節點的寬度 pixels 數，然而一條裂縫可能會有寬度大小不一的情形，

這會導致節點過少的話在計算平均寬度的時候漏掉裂縫的極端數據，造成寬度量化的不精確。過多的擷取節點也會造成資料過於冗餘，故本研究將選取三個節點數，分別為 25、50 和 75 個節點去分析同樣長度的裂縫他們計算平均裂縫寬度與真值的比值，評估出最適合的一個節點選取數量。圖 25 為五十個節點在裂縫上的示意圖。

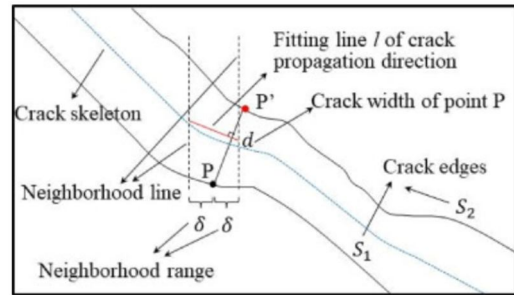


圖 24 Miao & Srimahachota(2021)提出寬度量量化方法示意圖

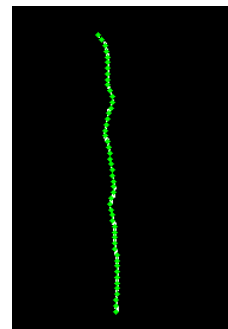


圖 25 五十個節點示意圖

表 6 為 50 個節點所算出來的寬度量量化表。

表 6 50 個節點寬度量量化表

編號	坐標 (x,y)	pixels	編號	坐標 (x,y)	pixels
1	(72,24)	1	26	(84,142)	2
2	(75,28)	1	27	(85,146)	1
3	(78,32)	3	28	(85,151)	2
4	(79,36)	3	29	(85,156)	2
5	(80,41)	3	30	(84,161)	5
6	(80,46)	2	31	(83,165)	3
7	(80,51)	2	32	(82,170)	1
8	(80,56)	2	33	(83,175)	2
9	(80,62)	2	34	(84,179)	2
10	(80,67)	2	35	(86,184)	4
11	(82,71)	3	36	(86,189)	3
12	(84,75)	1	37	(86,194)	3
13	(84,80)	3	38	(86,199)	2
14	(83,85)	3	39	(86,204)	2
15	(81,89)	3	40	(87,209)	4
16	(80,94)	1	41	(88,213)	3
17	(79,99)	1	42	(88,218)	2
18	(78,103)	2	43	(88,224)	2
19	(78,108)	3	44	(88,229)	3
20	(80,113)	4	45	(88,234)	2
21	(81,118)	3	46	(88,239)	3
22	(81,122)	1	47	(87,244)	1
23	(82,127)	1	48	(87,249)	2
24	(83,132)	1	49	(87,254)	3
25	(83,137)	1	50	(87,258)	2

經過 3 組不同節點所算出來的各節點裂縫寬度量化之結果，其中 25 節點的總裂縫寬度為 53 pixels，平均裂縫寬度大小為 2.12 pixels；50 節點的總裂縫寬度為 113 pixels，平均裂縫寬度大小為 2.26 pixels；75 節點的總裂縫寬度為 168 pixels，平均裂縫寬度大小為 2.24 pixels。計算飛行航高 10.32m 的時候 GSD 為 3mm/pixel，故飛行 7m 的時候 GSD 為 2.034mm/pixel，將平均裂縫寬度乘上 GSD 後就可以得到所算出來的裂縫寬度實際大小，如表 7 所示。

表 7 各節點精度表

已知值(mm)	節點數	pixels	寬度(mm)	絕對誤差(mm)
4.76	25	2.12	4.31	0.45
	50	2.26	4.60	0.16
	75	2.24	4.55	0.21

根據上面的數據，可以得到 25 個節點與真值的絕對誤差為 0.45 mm，50 個節點為 0.16mm，75 個節點的為 0.21mm，其中三組資料算出來的結果皆是相當的。

## 4. 結論與建議

### 4.1 結論

本研究首先利用 UAV 結合 YOLOv9 對路面先進行裂縫辨識，在選取最佳航高後，再利用 U-net 對裂縫進行寬度的量化以掌握整體裂縫生長。在研究裡可以發現 YOLOv9 在鋪面裂縫辨識上展現極佳效能，本研究導入 YOLOv9 模型進行瀝青鋪面裂縫辨識，透過 CrackForest 公開資料集訓練後，模型於 mAP50 達 0.891，mAP50-95 達 0.55，顯示模型在辨識精度與多尺度適應性上表現優異。YOLOv9 引入的 PGI 與 GELAN 主幹架構，使其在小裂縫與紋理背景較為複雜的場景下仍具良好辨識力，證實其作為高效輕量級檢測架構的可行性。

而飛行高度對辨識率有顯著影響，探討不同飛行高度對模型辨識結果的影響。考量無人機飛行穩定性與視野覆蓋範圍後，實驗結果顯示 7m 航高可達成最佳的準確率，約為 83%的準確率，1pixel 約為 2.034mm，同時保留兩線車道畫面，為後續大範圍鋪面巡檢提供實用參考依據。

最後利用不同節點數去計算平均裂縫寬度，可以得到 25 個節點與已知值的絕對誤差為 0.45 mm，50 個節點的絕對誤差為 0.16mm，75 個節點的絕對誤差為 0.21mm，其中以 50 個節點精度最好且不會造成過多的資料，說明此長度下進行裂縫寬度計算時三個節點數的擷取所算出來的誤差值都是相對小的，可以根據需求去選取。

### 4.2 建議

- (1) 增加多樣化的實景訓練資料以提升泛化能力建議未來可擴增資料來源，包括不同氣候、路況、照明條件下的鋪面影像，使模型具備更佳的適應性與魯棒性。尤其應涵蓋夜間、潮濕路面、起伏坡度與遮蔽陰影等複雜情境，以應對實際巡檢多變條件。
- (2) 推進系統化航線規劃與自動化調度能力，為實現完整無人巡檢任務，建議整合航線規劃模組與影像觸發控制器，使無人機能根據路網資訊自動規劃巡航軌跡，並依據飛行高度、鏡頭視角動態調整拍攝策略，以減少重疊冗餘並提升涵蓋效率。也可以利用更高解析度的機載相機去捕捉影片，探討更高的不同航高的影響與可行性。
- (3) 因為本研究僅針對公路局所頒布的養護手冊針對裂縫的寬度去做量化，後續也可以針對裂縫的長度進行量化，已對整體裂縫的走向以及發展有更進一步的了解和控制，為整體道路裂縫檢測提供一個更完善的流程。

### 參考文獻

- 中華民國交通部，2020。交通部公路總局公路養護手冊。[Ministry of Transportation and Communications, 2020. Highway Maintenance Manual by the Directorate General of Highways, Ministry of Transportation and Communications. (in Chinese)]
- 楊明德、蔡慧萍，2018。無人飛行載具發展現況與未來趨勢，土木水利，45(3)：23-29，DOI：10.6653/MoCICHE.201806\_45(3).0004。[Yang,

- M.D., and Tsai, H.P., 2018. An analysis of the development status and future trajectories of Unmanned Aerial Vehicle, *Civil and Hydraulic Engineering*, 45(3): 23-29, DOI: 10.6653/MoCICHE.201806\_45(3).0004. (in Chinese)]
- Bandini, F., Jakobsen, J., Olesen, D., Reyna-Gutierrez, J. A., and Bauer-Gottwein, P., 2017. Measuring water level in rivers and lakes from lightweight Unmanned Aerial Vehicles, *Journal of Hydrology*, 548: 237-250, DOI: 10.1016/j.jhydrol.2017.02.038.
- Cha, Y.J., Choi, W., and Büyüköztürk, O., 2017. Deep learning-based crack damage detection using convolutional neural networks, *Computer-Aided Civil and Infrastructure Engineering*, 32(5): 361-378, DOI: 10.1111/mice.12263.
- Chambon, S., Subirats, P., and Dumoulin, J., 2009. Introduction of a wavelet transform based on 2D matched filter in a Markov random field for fine structure extraction: Application on road crack detection, in *Proceedings of the SPIE Vol. 7251, Image Processing: Machine Vision Applications II*, San Jose, CA, USA, pp. 87-98, DOI: 10.1117/12.805437.
- Chien, C.T., Ju, R.Y., Chou, K.Y., and Chiang, J.S., 2024. YOLOv9 for fracture detection in pediatric wrist trauma X-ray images, *Electronics Letters*, 60(11): e13248, DOI: 10.1049/ell2.13248.
- Gavilán, M., Balcones, D., Marcos, O., Llorca, D.F., Sotelo, M.A., Parra, I., Ocaña, M., Aliseda, P., Yarza, P., and Amírola, A., 2011. Adaptive road crack detection system by pavement classification, *Sensors*, 11(10): 9628-9657, DOI: 10.3390/s111009628.
- Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints, *International Journal of Computer Vision*, 60: 91-110, DOI: 10.1023/B:VISI.0000029664.99615.94.
- Miao, P., and Srimahachota, T., 2021. Cost-effective system for detection and quantification of concrete surface cracks by combination of convolutional neural network and image processing techniques, *Construction and Building Materials*, 293: 123549, DOI: 10.1016/j.conbuildmat.2021.123549.
- Ragnoli, A., De Blasiis, M.R., and Di Benedetto, A., 2018. Pavement distress detection methods: A review, *Infrastructures*, 3(4): 58, DOI: 10.3390/infrastructures3040058.
- Redmon, J., Divvala, S., Girshick, R., and Farhadi, A., 2016. You only look once: Unified, real-time object detection, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, pp. 779-788, DOI: 10.1109/CVPR.2016.91.
- Ronneberger, O., Fischer, P., and Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation, in *Proceedings of the Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015*, Navab, N., Hornegger, J., Wells, W., and Frangi, A. (eds), *Lecture Notes in Computer Science*, Springer, Cham, vol. 9351, pp. 234-241, DOI: 10.1007/978-3-319-24574-4\_28.
- Tsai, Y.C., Kaul, V., and Mersereau, R.M., 2010. Critical assessment of pavement distress segmentation methods, *Journal of Transportation Engineering*, 136(1): 11-19, DOI: 10.1061/(ASCE)TE.1943-5436.0000051.
- Vo, H.T., Mui, K.C., Thien, N.N., and Tien, P.P., 2024. Automating tomato ripeness classification and counting with YOLOv9, *International Journal of Advanced Computer Science and Applications*, 15(4): 1120-1128.
- Wang, C.Y., Yeh, I.H., and Mark Liao, H.Y., 2024. Yolov9: Learning what you want to learn using programmable gradient information, in *Proceedings of the European Conference on*

- Computer Vision (ECCV), Leonardis, A., Ricci, E., Roth, S., Russakovsky, O., Sattler, T., Varol, G. (eds) , Lecture Notes in Computer Science, Springer, Cham, vol. 15089, DOI: 10.1007/978-3-031-72751-1\_1.
- Zhou, H., Kong, H., Wei, L., Creighton, D., and Nahavandi, S., 2014. Efficient road detection and tracking for unmanned aerial vehicle, *IEEE Transactions on Intelligent Transportation Systems*, 16(1): 297-309, DOI: 10.1109/TITS.2014.2331353.
- Zhu, J., Zhong, J., Ma, T., Huang, X., Zhang, W., and Zhou, Y., 2022. Pavement distress detection using convolutional neural networks with images captured via UAV, *Automation in Construction*, 133: 103991, DOI: 10.1016/j.autcon.2021.103991.

# Automatic Detection of Asphalt Pavement Cracks Using Unmanned Aerial Vehicles and Deep Learning

Kai-Ming Liu <sup>1\*</sup>    Szu-Pyng Kao <sup>2</sup>    Feng-Liang Wang <sup>3</sup>    Jhih-Sian Lin <sup>4</sup>

## Abstract

This study integrates Unmanned Aerial Vehicles (UAVs) with the deep learning object detection model YOLOv9 to conduct automated detection and analysis of cracks in asphalt pavement. Leveraging the high mobility and high-resolution imaging capabilities of UAVs, the system can rapidly cover large road areas and significantly reduce the time and labor costs associated with manual inspections. The YOLOv9 model was initially trained using the publicly available CrackForest dataset for pavement cracks. After 100 training epochs, the model achieved a detection accuracy of  $mAP50 = 0.891$  and  $mAP50-95 = 0.550$  on the test set, demonstrating strong performance in small object detection and pavement crack identification. Finally, through orthorectification and geometric correction, the detection results were projected onto real-world coordinates for crack width quantification. The results verify that the proposed system can serve as an effective tool for road maintenance decision-making and management.

**Keywords:** UAV, YOLOv9, Crack Detection, Pavement Crack, Deep Learning

---

<sup>1</sup> Master, Department of Civil Engineering, National Chung Hsing University

<sup>2</sup> Professor, Department of Civil Engineering, National Chung Hsing University

<sup>3</sup> Assistant Professor, Department of Applied Geoinformatics, Chien Hsin University of Science and Technology

<sup>4</sup> Ph.D., Department of Civil Engineering, National Chung Hsing University

\* Corresponding Author, E-mail: daniel1900513@gmail.com

Received Date: Sep. 03, 2025

Revised Date: Oct. 15, 2025

Accepted Date: Feb. 25, 2026