



以深度學習與數位學生 打造工地鋼筋查驗新法

陳翊翔／國震中心與台大土木合設AI研究中心 研究助理

莊仕杰、張鈞程、羅昱恆／國立臺灣大學土木工程學系電腦輔助工程組 碩士班研究生

黃琮煒／國立臺灣大學土木工程學系 大學部學生

邱永全／國震中心與台大土木合設AI研究中心 研究助理

林冠成／中興工程顧問股份有限公司三鶯捷運工程處 經理

黃志民／中興工程顧問股份有限公司研發及資訊部 協理

周頌安／中興工程顧問股份有限公司 總工程師

陳俊杉／國立臺灣大學土木工程學系 教授

現行工地鋼筋混凝土工程常因工程量體龐大、施作與查驗工程人員經驗不足、影像紀錄侷限性與高風險場域可及性等問題，而常有鋼筋綁紮組立品質難確保、工程查驗紀錄難運用於後續缺失改善與工程人員涉足高風險場域之工安問題等。本研究以深度學習與數位學生打造一工地鋼筋查驗新法，深度學習模型以中興工程顧問公司監造的三鶯捷運工程為實施案例，配合預鑄梁場與施工現場工程進度，取得大量鋼筋綁紮實際彩色影像與深度影像，並標註鋼筋組立查驗特徵作為模型的輸入資料，來訓練鋼筋特徵影像實例分割的深度學習模型，此模型於驗證集與測試集中皆有良好表現，再應用無人飛行載具執行鋼筋查驗作業。數位學生部分以運動回復結構（Structure from Motion, SfM）與多視角立體視覺（Multi-View Stereo, MVS）建立鋼筋組立的三維點雲，透過數位學生空間資訊投影、鋼筋影像間距特徵實例分割、影像像素與數位學生點雲的轉換關係，將鋼筋混凝土構件之鋼筋組立數位學生分割為具鋼筋間距特徵識別的「智慧學生」，此一智慧學生，將能透過自動化的真實尺度計算與相關統計數值應用於工地實務，以提升工程效率與品質。

關鍵字：深度學習、數位學生、工地查驗、鋼筋、影像實例分割、無人飛行載具

簡介

大型建物與公共工程多為鋼筋混凝土結構（Reinforced Concrete Structures），而影響鋼筋混凝土結構安全品質的關鍵則是鋼筋組立的施工品質。鋼筋組立是依照施工圖說所繪製的鋼筋號數、尺寸、形狀、間距及位置組合而成的框架。每個構造物依照其功能，或為結構安全、或為配合大型機電管線，所需組合之鋼筋數量及類型都相當繁雜。在工地現場進行澆注混凝土前，

查驗鋼筋框架是否符合設計圖要求與施工規範的規定，是件繁瑣但必要的步驟。以下分就工地鋼筋查驗現況、本研究所提出之鋼筋查驗新法、深度學習應用於工地與數位學生應用於工地，做一簡述。

工地鋼筋查驗現況

現行鋼筋查驗因工程量體龐大，僅能以人工進行目視檢查或是抽樣量測來確保鋼筋組立綁紮是否合乎規範，無法執行全面性地查驗；此外查驗品質的好壞

也相當仰賴現場查驗工程師的經驗，從以使查驗品質難以穩定；而現行查驗的紀錄影像雖能作為查驗結果的參考，但往往只能記錄鋼筋組立其中一角，無法掌握鋼筋組立全貌，不易於連結後續的鋼筋組立修正作業與查驗議題追蹤；另外大型建物與公共工程的施工現場往往相當複雜且具高風險，如深基礎開挖（如井基）可達數十公尺深，高架橋墩柱可達數十公尺高，在此高風險工作環境中，查驗人員常需吊掛繩索或進入吊箱至待檢核區段以目視或抽樣量測進行查驗作業（如圖 1），從業人員的安全實須獲得關注與改善。綜前所述，工地鋼筋查驗現況的痛點包含以下四項，無法全面性完成鋼筋查驗、鋼筋查驗品質難以穩定、查驗影像無法釐清鋼筋組立查驗全貌與從業人員於高風險環境的查驗作業安全。

工地鋼筋查驗新法

鑒於工地鋼筋查驗現況中所遭遇的困難，本研究提出以深度學習與數位孿生技術所打造的工地鋼筋查驗新法，能為現行工地鋼筋查驗現況所帶來的改變如圖 2 所示。深度學習模型的即時運算提供穩定的查驗品質並使全查驗成為可能，數位孿生則記錄工地鋼筋組立的細節並呈現全貌，使工地鋼筋查驗作業的議題追蹤得以實現。以下分別簡述深度學習與數位孿生於工地鋼筋查驗新法中扮演的腳色與作用。



(a) 查驗人員繫上安全繩執行鋼筋查驗的抽樣量測



(b) 查驗人員進入吊箱以目視執行鋼筋查驗作業

圖 1 工地鋼筋查驗的高風險作業環境

本研究使用 Mask Region-based Convolutional Neural Network（以下簡稱 Mask R-CNN）^[1] 實現深度學習中的影像實例分割技術，學習工地現場鋼筋組立的影像與深度影像中與查驗規範相關的鋼筋特徵，包含鋼筋組立的間距、搭接長度、彎鉤角度與彎鉤延伸長度，從而建立具鋼筋特徵識別能力的深度學習模型，並透過影像其相對應的空間資訊進行真實尺度計算，完成查驗作業。此深度學習模型即使在含有數量以百計的鋼筋間距的影像都能即時完成運算得到結果，使工地現場鋼筋組立的全查驗成為可能。深度學習可透過數據量增加而增強模型表現，經訓練完成後能提供穩定的鋼筋查驗品質。深度學習於本研究中作為鋼筋特徵識別的影像實例分割模型，將查驗所需的鋼筋重要特徵辨識後，與鋼筋組立數位孿生結合。深度學習相關細節將於第 2 章深度學習應用於鋼筋組立辨識進行詳述。

本研究透過電腦視覺中的運動回復結構（Structure from Motion, SfM）^[2-4] 與多視角立體視覺（Multi-View Stereo, MVS）^[5] 從工地鋼筋二維影像建立工地鋼筋組立的數位孿生，並利用數位孿生的空間資訊進行再投影取得鋼筋二維影像所對應的深度影像，並將影像與深度影像作為前述之深度學習模型之輸入進行鋼筋特徵識別，識別結果再透過影像像素與數位孿生的空間資訊關係進行轉換，完成鋼筋組立數位孿生的特徵識別，數位孿生在本研究中作為資訊載體，並提供深度學習模型所需的空間資訊。數位孿生相關細節將於第 3 章鋼筋組立數位孿生建立進行詳述。



圖 2 現行工地鋼筋查驗與本研究所提出的工地鋼筋查驗新法的比較

深度學習於工地應用現況

深度學習於工地的應用如圖 3 所示，所使用的深度學習架構以深度神經網路 (Deep Neural Network, DNN)、卷積神經網路 (Convolution Neural Network, CNN) [6]、循環神經網路 (Recurrent neural network, RNN) [7] 與生成對抗網路 (Generative Adversarial Network, GAN) [8] 等為主，透過對工程影像進行分類與描述、對機具或是工程人員進行偵測、追蹤與行為識別、對營建相關產業的歷時訊號或數據進行預測等，應用有工地安全 [9-12] 與建物破壞程度分類 [13]、路面檢測分類 [14] 與大型機具狀態識別 [15] 與營建成本預測 [16]。其中應用於影像相關的卷積神經網路是其中最廣泛應用的深度學習模型架構 [6]，多以物件的方框偵測 [17] 為主，像素等級的影像分割技術 [1,18] 則少有於工地的應用。

數位孿生應用於工地

數位孿生因具備自動、開放且可追蹤的特性，本研究將其納入成為實現工地檢核系統的工具。數位孿生在工業中一般被定義為三部分：實體資產、數位模型與兩者之間的同步連結，目的為進行實體資產的管控。根據 Sacks 等人於 2020 年發表之研究 [20]，數位孿生於工地之工作流程可分為四個面向討論，分別為實體 - 虛擬 (physical-virtual)、產品 - 流程 (product-

process)、意圖 - 狀態 (intent-status)、數據 - 資訊 - 知識 - 決策 (data-information-knowledge-decisions)，如圖 4 所示。實體 - 虛擬維度即數位孿生之基本定義，由實體資產與數位模型構成；產品 - 流程維度說明了數位孿生可用於紀錄施工階段的兩個狀態，第一為產物，即結構物，第二為施工的過程；意圖 - 狀態則代表了時間的維度，不同階段會有不同的產物，如 as-designed model (設計模型) 與 as-built model (現地模型)、as-planned process (計畫流程) 與 as-performed process (實際流程)；最後數據到決策的維度，傳達了數位孿生引入的核心意義，乃由大量的數據走向準確且迅速的決策。在 2020 年 Lin [21] 等人發表之最新研究引入數位孿生的概念於工地建造過程，透過自動比對 4D 建築資訊模型 (Building Information Modeling, BIM) 與 4D 現地點雲模型，達成自動、即時且主動的工程進度管控與圖資管理，並於兩個實際工地驗證此方法能有效改善既有營管架構。

深度學習應用於鋼筋組立辨識

用於影像像素等級分割的深度學習技術包含語意分割 (semantic segmentation) [18] 與實例分割 (instance segmentation) [1]，語意分割試圖在語意上理解圖中像素點的作用，描繪出物體的邊界與密集的程度，但實例分

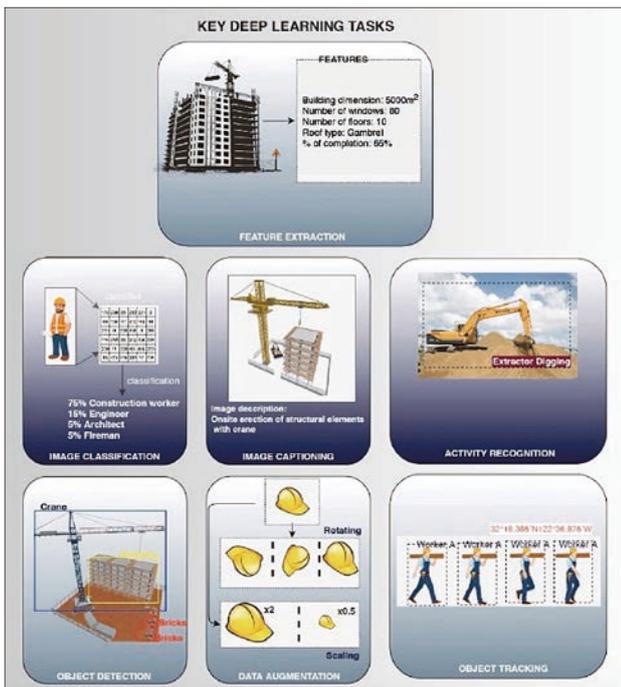


圖 3 深度學習於營建工程之應用 (摘自 [19])

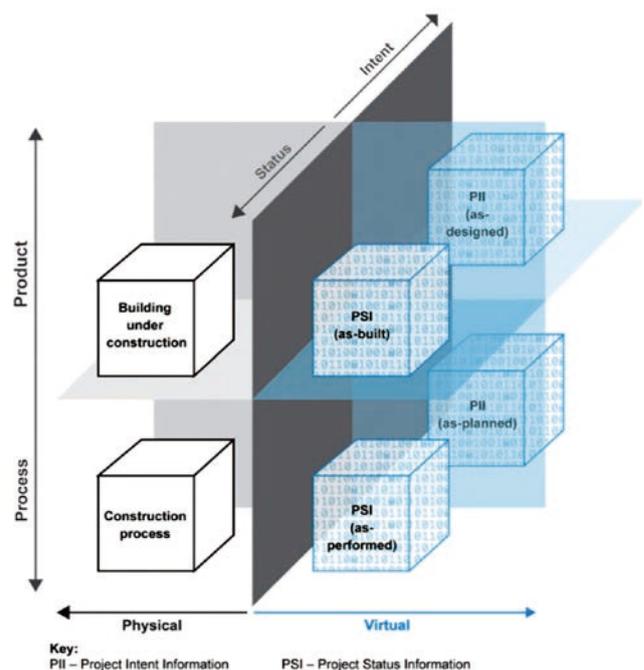


圖 4 數位孿生於工地應用之三個面向：實體 - 虛擬、產品 - 流程、意圖 - 狀態 (摘自 [20])

割技術可將類別相同之重疊實例 (instance) 進行獨立預測，為目前深度學習於電腦視覺技術中困難的任務之一。本研究使用 Mask R-CNN [1] 模型進行實例分割，其模型流程架構如圖 5。Mask R-CNN 使用了兩個階段來完成影像的實例分割。第一階段為使用 Region Proposal Network (以下簡稱 RPN) 來生成可能包含目標的區域。RPN 的運作原理為使用不同尺寸與長寬比的 anchor box 在特徵圖上掃描，藉此找到感興趣的區域 (Region of Interest, 以下簡稱 RoI)。第二階段首先將第一階段生成的 RoI 池化 (pooling)，模型經過 RoIAlign 使每個 RoI 的尺寸相同，才能夠快速地進行分類的預測，並調整 anchor box 的位置，使其更為準確。再使用卷積神經網路對調整完後的 RoI 生成遮罩 (mask)，在影像中以像素的等級對目標物進行分割預測。

本研究所辨識的對象為工地現場的鋼筋組立，由不同排列方式的鋼筋框架層層綁紮而成，僅靠 RGB 彩色圖像進行影像分割時，模型通常難以分辨複雜場景下的前景與背景，以及類似顏色像素的物品，再者 RGB 彩色圖像不具備能與真實尺度接軌的空間資訊，即使能正確識別出鋼筋特徵，但對於鋼筋組立的間距、搭接長度無法提供資訊進行計算。先前研究以深度影像與語意分割技術為主用於鋼筋間距查驗 [22]，本研究延續其脈絡，將模型輸入改為彩色與深度影像以增加其影像特徵，並以影像實例分割的模型 Mask R-CNN 取代原有的語意分割模型進行鋼筋組立的間距特徵預測，透過重疊的實例資訊可改善鋼筋間距計算上的表現，以下就彩色影像與深度影像的蒐集方式與資料處理作一簡要說明。

在深度感測器的迅速發展下，可攜帶式的彩色影像與深度影像 (RGB-Depth Image, 以下簡稱為 RGB-D) 相機逐漸普及，如 Microsoft Kinect 與 Intel RealSense，

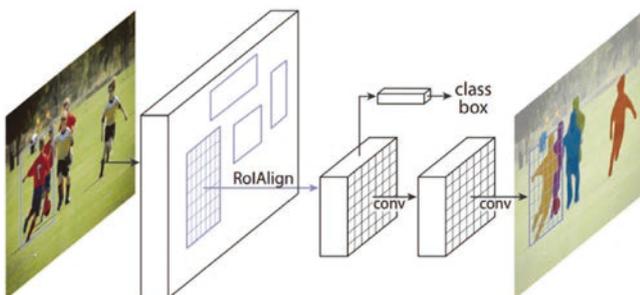


圖 5 Mask R-CNN 的流程架構 (摘自 [1])

搭載主動紅外線感測器，並可將深度影像與 RGB 影像同步顯示，藉此提供了大量的 RGB-D 資料集。目前 RGB-D 影像分割方法眾多，大致可分為三類。其一為將深度資訊以水平差異、對地面高度與表面法向量的角度 (horizontal disparity, height above ground, and the angle the pixel's local surface normal makes with the inferred gravity direction, 以下簡稱為 HHA) 編碼 [22]，並從 HHA 與 RGB 透過完整的 CNN 提取特徵 [23]；其二為將 RGB 影像與深度影像各自提取特徵後，將深度特徵與 RGB 特徵以交換網路 (transformation network) 學習 [24] 或融合 [25]，接著以反卷積層 (deconvolutional layers) 與反池化層 (unpooling layers) 重建標籤圖；其三為使用循環神經網路架構將 RGB-D 影像中的資訊循環學習以改善影像分割 [26]。本研究透過前述的第一種方法，將深度資訊以 HHA 編碼後與 RGB 影像作為訓練完成的 Mask R-CNN 模型輸入，來提取鋼筋組立特徵。

鋼筋組立影像實例分割模型訓練與成果

本研究以 Intel RealSense 深度相機於中興工程顧問公司監造的三鶯捷運工程的預鑄梁場與施工現場蒐集鋼筋組立的 RGB-D 影像，在鋼筋組立特徵識別中，本研究使用開源軟體 LabelMe [27] 一共標註了 248 張鋼筋組立影像裡的鋼筋框架結點與間距，結點意指鋼筋與鋼筋重疊交錯之處，以多邊形 (Polygon) 沿著交錯處邊界進行標註，間距則指結點與結點之間的連線，以線段進行標註，並將其中的 200 張影像作為模型的訓練集，28 張作為驗證集，20 張作為測試集。模型訓練曲線如圖 6 所示，本研究選取訓練過程中，訓練集的損失函數 (Loss function) 最低點的模型參數作為最終模型參數。

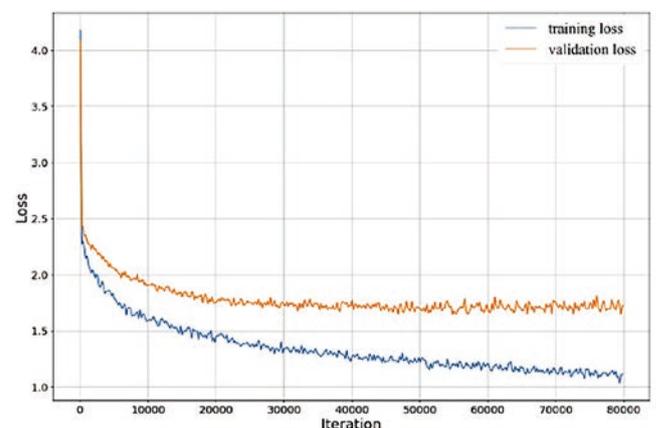


圖 6 鋼筋組立間距特徵影像實例分割模型訓練曲線

模型於測試集中的鋼筋組立間距特徵預測成果如圖 7(a) 所示，能成功將鋼筋框架的結點與間距特徵以像素等級進行實例分割。透過實例分割的結點與間距資訊結合深度影像中的空間資訊，就能完成鋼筋間距的真實尺度計算，計算結果如圖 7(b) 所示。

無人飛行載具執行鋼筋組立查驗

本研究以 DJI 的無人飛行載具 (Phantom 4 Pro V2) 搭配 Intel RealSense 深度相機 D415 與 Nvidia Jetson Nano 邊緣計算核心打造用於工地鋼筋查驗的無人飛行載具原型，以高為 160 公分的鋼筋混凝土構件之鋼筋組立為對象蒐集 RGB-D 影像，並透過前述的鋼筋組立間距特徵模型提取特徵，完成鋼筋間距真實尺度計算，實驗情形與計算結果如圖 8。



(a) 鋼筋組立結點與間距的實例分割結果



(b) 鋼筋間距真實尺度計算結果

圖 7 鋼筋結點與間距影像實例分割與間距計算結果

鋼筋組立數位孿生建立與深度投影

數位孿生建立與深度投影流程

本研究使用開源軟體 Colmap [4,28,29] 以 SfM [2-4] 與 MVS [5] 技術從鋼筋組立二維影像重建鋼筋組立的三維點雲，步驟如圖 9，先透過相機收集鋼筋組立的二維影像，經由 SfM 提取影像與影像間的特徵，建立鋼筋組立的稀疏點雲與二維影像在空間中對應的姿態，再藉由 MVS 配對影像像素點與影像姿態建構密集點雲模型，最後將密集點雲對應的空間資訊進行投影，取得每個鋼筋組立二維影像姿態所對應的深度資訊。



(a) 無人飛行載具執行鋼筋查驗實驗的鋼筋間距真實尺度計算結果



(b) 無人飛行載具執行鋼筋查驗實驗的側拍影像

圖 8 鋼筋組立間距特徵識別深度學習模型應用於無人飛行載具執行鋼筋查驗

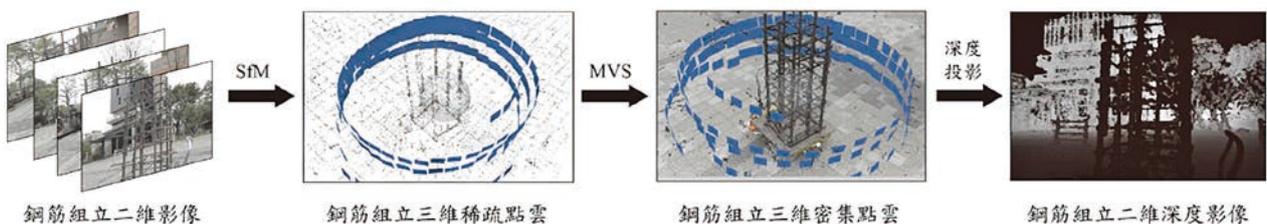
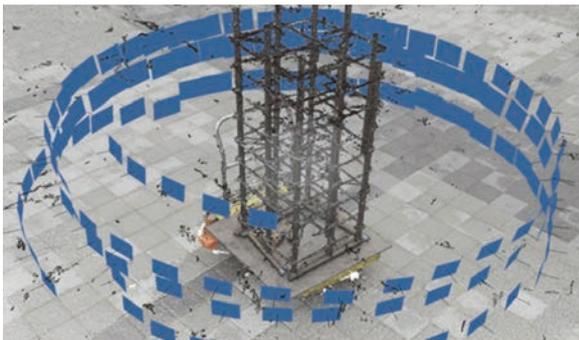


圖 9 鋼筋組立數位孿生與深度投影流程圖

本研究以高為 160 公分高的鋼筋混凝土構件鋼筋組立為目標物進行數位孿生的建立，如圖 10(a) 所示，拍攝人員離目標鋼筋組立距離約為 2 公尺處進行環繞拍攝，一共取得 176 張鋼筋組立二維影像，再依前述所提的 SfM 與 MVS 方法取得影像姿態與密集點雲，完成此鋼筋混凝土構件鋼筋組立的數位孿生建置，如圖 10(b)。



(a) 鋼筋組立數位孿生拍攝實驗影像



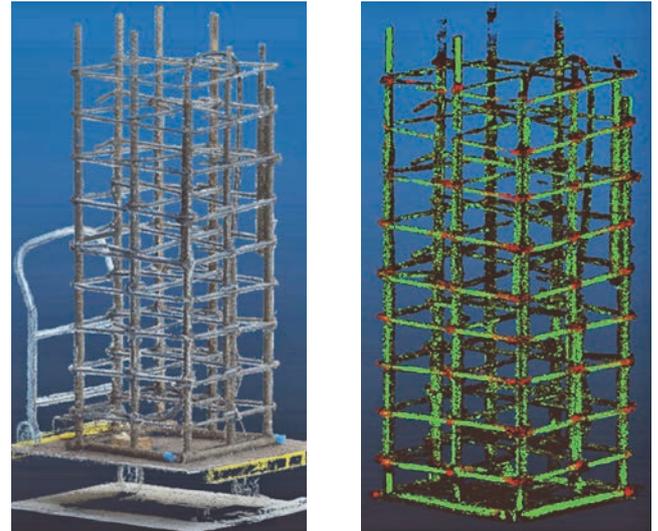
(b) 鋼筋組立數位孿生結果，藍色方框為所使用的影像姿態

圖 10 鋼筋組立數位孿生建立實驗照與結果

鋼筋組立數位孿生間距特徵識別成果

利用此鋼筋混凝土構件的鋼筋組立數位孿生的空間資訊，對 176 張的二維鋼筋組立影像的像平面進行投影，取得對應的 176 張二維鋼筋組立的深度影像，組成此鋼筋混凝土構件鋼筋組立的 RGB-D 資料集。並將此資料集作為第 2 章所述的鋼筋組立間距特徵深度學習模型的輸入進行特徵識別，識別結果透過 MVS 所建立的二維影像像素與數位孿生的點雲座標的關係進行轉換，完成此鋼筋混凝土構件的鋼筋組立數位孿生的特徵識別。

數位孿生與數位孿生特徵識別成果比較如圖 11，透過鋼筋組立二維影像所建立的數位孿生，雖可完整涵蓋此鋼筋組立的完整資訊，但資訊量過於龐大如以人工進行識別則過於耗時而無實益，欲以傳統演算法設計自動化處理流程則易淪為個案處理，難以廣泛應用。因此本



(a) 透過鋼筋組立二維影像以 SfM 與 MVS 所建立的數位孿生 (b) 將原始數位孿生透過間距特徵深度學習模型識別後的特徵數位孿生，紅色代表鋼筋與鋼筋交會的結點，綠色則為鋼筋間距。

圖 11 鋼筋組立原始數位孿生與特徵識別之數位孿生

研究透過鋼筋組立的 RGB-D 資料的影像實例分割技術，以像素等級識別影像特徵，並透過數位孿生建立流程中影像像素與點雲座標的關係，自動化完成數位孿生的特徵識別。與僅以影像進行識別相比，數位孿生特徵識別可改善影像識別由於空間資訊不足無法將識別特徵於整體鋼筋組立中進行定位的缺點。本研究所發展之特徵識別之數位孿生，可根據查驗項目自動進行特徵識別與真實尺度計算與相關統計數值應用於工地鋼筋查驗，將於鋼筋查驗自動化發展具重要貢獻。

結論與展望

本研究透過鋼筋組立的 RGB-D 資料收集與間距特徵標註，成功完成鋼筋組立間距特徵識別影像實例分割的深度學習模型訓練，於驗證集與測試集中皆有良好表現。本研究並以鋼筋混凝土構件之鋼筋組立為目標進行拍攝，透過 SfM 與 MVS 技術建立此鋼筋組立的數位孿生。後續透過數位孿生空間資訊投影、鋼筋間距特徵實例分割、影像像素與數位孿生點雲的轉換關係，成功將鋼筋混凝土構件之鋼筋組立數位孿生分割為具鋼筋間距特徵識別的「智慧孿生」，此一智慧孿生將能透過自動化的真實尺度計算與相關統計數值應用於工地鋼筋查驗。此一由深度學習與數位孿生所打造的查驗新法，僅需工地鋼筋組立影像就能以穩定查驗品質提供工地鋼筋組立的全面查驗，搭配數位孿生

確立查驗位置來實現鋼筋組立修正作業與查驗議題追蹤，搭配無人飛行載具則可輔助高風險場域的鋼筋查驗作業，增進工地安全。

本研究於後續規劃中，將完成其他重要鋼筋查驗特徵之深度學習識別模型，並搭配數位孿生技術打造出各式鋼筋查驗特徵識別之智慧孿生。並發展此一智慧孿生與設計（建築資訊模型、設計圖或是施工圖說）直接進行比對的查驗方法來邁向工地查驗自動化，此一技術帶來的革新後續將能搭配精實營建作法，創造產業的真實利基以驅動營建產業的數位轉型。

誌謝

本研究感謝科技部與中興工程顧問公司在「以人工智慧輔助工地鋼筋查驗」產學合作研究案（MOST 109-2622-E-002-016-CC2）的研究經費贊助，並感謝國立臺灣大學土木工程系的韓仁毓教授、張書瑋副教授、張家銘副教授、陳柏華副教授、林之謙助理教授與國立臺灣科技大學營建工程系紀乃文專案助理教授對本研究的指導。

參考文獻

- He, K., G. Gkioxari, P. Dollár, and R. Girshick, Mask R-CNN. *IEEE Trans Pattern Anal Mach Intell*, 2020. 42(2): p. 386-397.
- Lowe, D.G. Object recognition from local scale-invariant features. in *Proceedings of the Seventh IEEE International Conference on Computer Vision*. 1999.
- Bay, H., A. Ess, T. Tuytelaars, and L. Van Gool, Speeded-Up Robust Features (SURF). *Computer Vision and Image Understanding*, 2008. 110(3): p. 346-359.
- Schönberger, J. and J.-M. Frahm. Structure-from-Motion Revisited. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
- Goesele, M., B. Curless, and S.M. Seitz. Multi-View Stereo Revisited. in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*. 2006.
- Krizhevsky, A., I. Sutskever, and G.E. Hinton, ImageNet classification with deep convolutional neural networks, in *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*. 2012, Curran Associates Inc.: Lake Tahoe, Nevada. p. 1097-1105.
- Mikolov, T., M. Karafiát, L. Burget, J. ernocký, and S. Khudanpur. Recurrent neural network based language model. in *Eleventh annual conference of the international speech communication association*. 2010.
- Goodfellow, I.J., J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, Generative adversarial networks. *arXiv preprint arXiv:1406.2661*, 2014.
- Nath, N.D., A.H. Behzadan, and S.G. Paal, Deep learning for site safety: Real-time detection of personal protective equipment. *Automation in Construction*, 2020. 112: p. 103085.
- Fang, Q., H. Li, X. Luo, L. Ding, H. Luo, T.M. Rose, and W. An, Detecting non-hardhat-use by a deep learning method from far-field surveillance videos. *Automation in Construction*, 2018. 85: p. 1-9.
- Yang, K., C.R. Ahn, and H. Kim, Deep learning-based classification of work-related physical load levels in construction. *Advanced Engineering Informatics*, 2020. 45: p. 101104.
- Yu, Y., H. Li, X. Yang, L. Kong, X. Luo, and A.Y.L. Wong, An automatic and non-invasive physical fatigue assessment method for construction workers. *Automation in Construction*, 2019. 103: p. 1-12.
- Chi, N.-W., J.-P. Wang, J.-H. Liao, W.-C. Cheng, and C.-S. Chen, Machine learning-based seismic capability evaluation for school buildings. *Automation in Construction*, 2020. 118: p. 103274.
- Gopalakrishnan, K., S.K. Khaitan, A. Choudhary, and A. Agrawal, Deep Convolutional Neural Networks with transfer learning for computer vision-based data-driven pavement distress detection. *Construction and Building Materials*, 2017. 157: p. 322-330.
- Rashid, K.M. and J. Louis, Times-series data augmentation and deep learning for construction equipment activity recognition. *Advanced Engineering Informatics*, 2019. 42: p. 100944.
- Raffei, M.H. and H. Adeli, Novel machine-learning model for estimating construction costs considering economic variables and indexes. *Journal of construction engineering management*, 2018. 144(12): p. 04018106.
- Redmon, J. and A. Farhadi, Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*, 2018.
- Chen, L.-C., Y. Zhu, G. Papandreou, F. Schroff, and H. Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. in *Proceedings of the European conference on computer vision (ECCV)*. 2018.
- Akinosho, T.D., L.O. Oyedele, M. Bilal, A.O. Ajayi, M.D. Delgado, O.O. Akinade, and A.A. Ahmed, Deep learning in the construction industry: A review of present status and future innovations. *Journal of Building Engineering*, 2020. 32: p. 101827.
- Sacks, R., I. Brilakis, E. Pikas, H.S. Xie, and M. Girolami, Construction with digital twin information systems. *Data-Centric Engineering*, 2020. 1.
- Lin, J.J. and M. Golparvar-Fard, Visual and Virtual Production Management System for Proactive Project Controls. *University of Illinois at Urbana-Champaign*, 2020.
- Gupta, S., R. Girshick, P. Arbeláez, and J. Malik. Learning Rich Features from RGB-D Images for Object Detection and Segmentation. in *European conference on computer vision*. 2014. Cham: Springer.
- Long, J., E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2015.
- Wang, J., Z. Wang, D. Tao, S. See, and G. Wang. Learning Common and Specific Features for RGB-D Semantic Segmentation with Deconvolutional Networks. in *European Conference on Computer Vision*. 2016. Cham: Springer.
- Hazirbas, C., L. Ma, C. Domokos, and D. Cremers. FuseNet: Incorporating Depth into Semantic Segmentation via Fusion-Based CNN Architecture. in *Asian conference on computer vision*. 2016. Cham: Springer.
- Abdulnabi, A.H., B. Shuai, Z. Zuo, L. Chau, and G. Wang, Multi-modal Recurrent Neural Networks With Information Transfer Layers for Indoor Scene Labeling. *IEEE Transactions on Multimedia*, 2018. 20(7): p. 1656-1671.
- Russell, B.C., A. Torralba, K.P. Murphy, and W. Freeman, LabelMe: a database and web-based tool for image annotation. *International journal of computer vision*, 2008. 77(1-3): p. 157-173.
- Schönberger, J.L., E. Zheng, J.-M. Frahm, and M. Pollefeys. Pixelwise view selection for unstructured multi-view stereo. in *European Conference on Computer Vision*. 2016. Springer.
- Schönberger, J.L., T. Price, T. Sattler, J.-M. Frahm, and M. Pollefeys. A vote-and-verify strategy for fast spatial verification in image retrieval. in *Asian Conference on Computer Vision*. 2016. Springer.