



# 基於工程圖說之 施工查驗表單 自動化檢核

陳弘翊／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

陳翊翔／國震中心與臺大土木合設 AI 中心 專案經理

林之謙／國立臺灣大學土木工程系 副教授

陳俊杉\*／國立臺灣大學土木工程系 特聘教授

以查驗表單進行現場查驗是營建工程中確保結構品質的關鍵步驟，然而表單內標準的檢核需人工對照繁複且篇幅龐大的施工圖說，不僅耗時費力且易出錯。本研究提出一結合多模態檢索增強生成（Retrieval-Augmented Generation, RAG）與代理式工作流程之規範自動化檢核（Automated Compliance Checking, ACC）系統。該系統導入為工程圖說特別設計之層級式詳圖檢索流程，針對表單內容檢索相關圖說與詳圖。以實際橋梁工程圖說與查驗表單進行實驗之結果顯示本系統於圖說頁面檢索、詳圖檢索與符合性檢核三項指標中分別達成 93.3%、92.4% 與 83.3% 的準確率，驗證其於工程圖說 ACC 應用之實用性，為工程領域實務導入基於圖說之 ACC 系統發展提供具體解決方案。

## 簡介

現場工程查驗是現代營建工程中確保結構品質的關鍵步驟。在施工過程中，每一步驟皆須依據相應的施工規範或工程圖說嚴格執行以確保工程品質<sup>[1]</sup>。為利現場查驗之執行，傳統上工程師會於查驗前預先準備包含各項查驗項目與查驗標準之查驗表單，而於現場查驗時，即可依據表單所列之查驗標準逐項進行檢核。然而查驗表單中所列之查驗標準是否正確與精準，攸關查驗作業的有效性。若表單內容出現錯誤或與實際規範不符，將導致現場查驗未能確實落實。一項在臺灣進行之研究指出，在公共工程施工查核品質管理制度常見缺失中，缺失排行第一名即為「品管自主檢查標準未訂量化標準及落實執行」，缺失比率高達 84.1%<sup>[2]</sup>。此類問題包括品管自主檢查表未落實執行，或檢查標準未訂量化、容許

誤差值，或未確實記載檢查值等。其根本原因在於查驗表單的製作過程需人工對照複雜且篇幅龐大的施工規範及圖說，以淡江大橋工程專案為例，該專案之施工圖說總計高達 1650 頁。此一流程不僅耗時費力，亦極易產生錯誤<sup>[3]</sup>。此情境突顯了針對工程查驗表單進行規範自動化檢核（Automated Compliance Checking, ACC），並結合文字規範與施工圖說進行檢核之必要性。

針對文字法規之 ACC，傳統方法多仰賴自然語言處理技術<sup>[3]</sup>與建築資訊模型<sup>[4]</sup>。而近期隨著大型語言模型（Large Language Models, LLMs）及其相關技術，如檢索增強生成（Retrieval-Augmented Generation, RAG）<sup>[5]</sup>、多模態 RAG<sup>[6]</sup>與代理型工作流程（Agentic Workflow）<sup>[7]</sup>之發展，進一步拓展了對營建法規語意理解的可能性。已有多項研究證實，結合生成式預訓練 Transformer 模型、提示工程與 RAG 技術之架構，可實現基於文字層面的工程規範 ACC<sup>[8-10]</sup>。除了文字規範

\* 通訊作者，dchen@ntu.edu.tw

外，圖說在多數應用領域中亦扮演關鍵角色，對於準確傳達設計意圖與技術要求具有不可或缺的重要性。在工程領域中，諸如鋼筋配筋規格、施工高程、鋼筋搭接長度等查驗項目，這類查驗項目不僅數量眾多且在一工程內各處標準不盡然相同，無法依靠純文字規範闡明所有細節，因此必須倚賴大量設計圖或施工圖等文件內之圖說或表格進行檢核。為解讀工程圖說資訊，傳統之電腦視覺技術包括可擷取圖面文字資訊之光學字元辨識<sup>[11]</sup>；可辨識圖形元素之卷積神經網路<sup>[12]</sup>；而注意力機制模型則可捕捉文字與圖像間之上下文關係<sup>[13]</sup>。近期研究亦開始探索運用多模態 LLMs 於工程文件理解上之潛力<sup>[14]</sup>。儘管已有相關技術進展，然而目前尚未有效連結至針對查驗表單進行施工圖說 ACC 之流程中。其主要挑戰包括：現有語言模型之上下文視窗（Context Window）限制，導致難以處理大型且複雜之施工圖；以及施工圖說間高度視覺相似性，致使檢索困難。故目前多模態 RAG 架構尚不足以支援針對查驗表單之 ACC。此現況凸顯出需開發具工程領域特化之檢索機制，方能實現基於工程圖說之施工查驗表單 ACC 系統。

為克服上述限制，本研究提出一套創新架構如圖 1 所示，整合層級式詳圖檢索與代理式反思流程，以實現基於工程圖說之施工查驗表單 ACC。層級式詳圖檢索嵌入工程圖說之領域特有結構規則，並由多個模組組成，能針對查驗表單內容，自設計圖說中層級式溯源並檢索相關詳圖。實驗驗證結果顯示，在使用真實施工圖說與查驗表單進行測試時，所提之架構於低難度之 RC 擋土牆工程分項工程資料集之圖說頁面檢索、詳圖檢索及整

體符合性檢核準確率均達 100.0%，更於高難度之全套管基樁分項工程資料集上分別達 93.3%、92.4% 與 83.3%，證實其有效性及可擴展性。本研究之主要貢獻於發展出一套具工程領域特化檢索能力之架構，可克服現有多模態 RAG 方法之限制，具體貢獻如下：

1. 導入一物件偵測模型以分割複雜施工圖面為具結構性的詳圖，使模型可於語言模型上下文視窗限制內，保留必要視覺與語意資訊進行處理。
2. 設計一與施工圖說結構對應之層級式檢索流程，即便面對圖說頁面與詳圖間高度視覺相似性，仍能精確檢索相關圖說頁面及詳圖，確保檢核品質。

本文其餘章節安排如下：第貳節說明本研究架構之方法與實驗配置；第參節討論消融實驗結果與檢核錯誤分佈及變化<sup>[15]</sup>；第肆節為結論與未來研究方向。

## 研究方法

### 系統運作流程

本基於工程圖說之查驗表單 ACC 系統旨在自施工圖檔中檢索正確圖說頁面及詳圖，並對查驗表單進行檢核，其整體系統運作流程如圖 2 所示。

### 輸入及資料預處理

#### 施工查驗表單

圖 3(a) 所示為一施工查驗表單範例，其內容包含表單名稱、工程名稱、檢查位置、管理項目以及抽查標準等資訊。為利自動化處理，查驗表單內容經人工轉換為 JSON 格式，作為系統輸入資料。

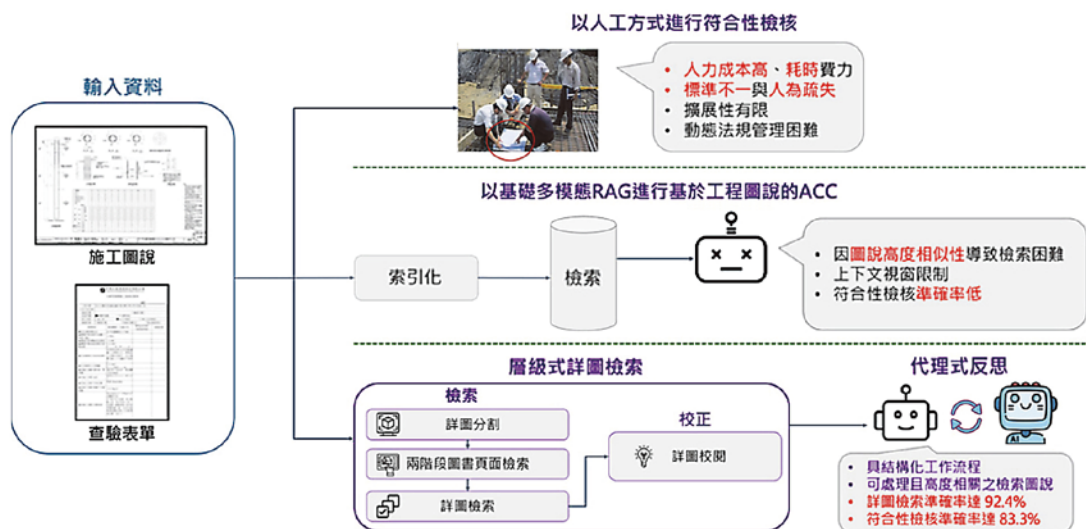


圖 1 本研究背景與所提架構概覽

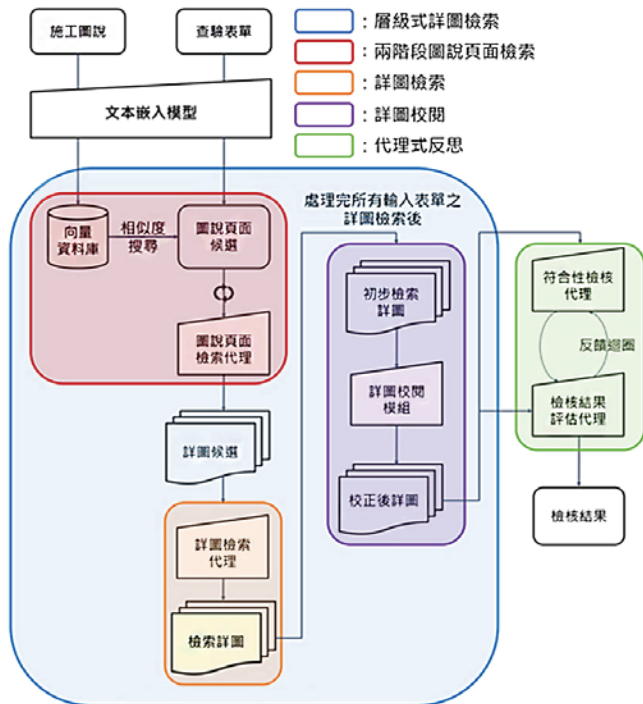


圖 2 系統運作流程概覽

# 施工圖說檔案

圖 3(b) 為一施工圖說頁面之範例，展示圖說頁面結構與詳圖配置。施工圖說檔案通常為 PDF 格式，作為本系統進行檢核的主要參考依據。其內部具備有助於系統進行有組織檢索之層級式結構，可分為三個層級：(1) 頂層：整份施工圖說檔案，作為整體設計資訊之載體；(2) 中層：單一圖說頁面，每頁具備唯一之圖

號與圖名；(3) 底層：頁面中內嵌之詳圖或表格，呈現施作細節與規格資訊。為支援系統流程之順利運作，需對輸入之施工圖說檔案進行預處理。具體步驟如下：首先將 PDF 文件中每一頁擷取並轉換為圖像檔案。隨後從各頁圖像中提取圖名與圖號，並使用 OpenAI 提供之「text-embedding-3-small」<sup>[16]</sup> 模型產生圖名的文字向量。最後將上述圖名、圖號與文字向量存入向量資料庫中（本研究採用 Neo4j 圖型資料庫進行實作）<sup>[17]</sup>。透過此前處理流程，系統可執行基於輸入查驗表單內容對於工程圖說頁面之相似度搜尋，並建立施工圖說檔案內各頁面與其詳圖間之關係，進而支援後續之層級式詳圖檢索與 ACC 作業。

## 層級式詳圖檢索

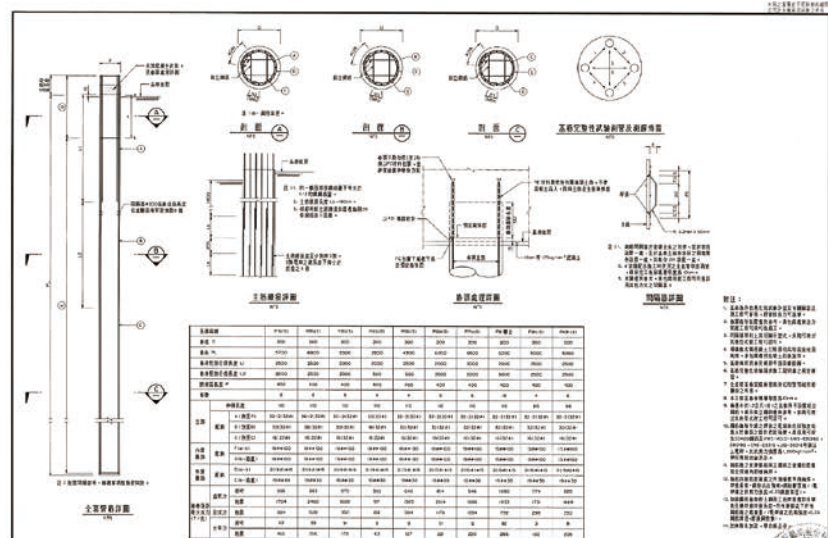
層級式詳圖檢索採用由上而下之層級式檢索策略以逐步縮小檢索範圍，檢索最符合查驗表單內容之圖說內容。該機制整合多個模組，包括相似度搜尋、詳圖分割、兩階段圖說頁面檢索、詳圖檢索以及詳圖校閱。以下各小節將分別介紹層級式詳圖檢索各組成模組之設計與運作細節。

## 相似度搜尋

系統將查驗表單資訊轉換為文字向量表示，並與在預處理階段儲存於向量資料庫中之圖名向量進行相似度比較，取得圖說頁面候選。

工程名稱		編號：	
流江大橋及其連絡路 5K+000~7K+035 新築工程			
分項工程名稱	檢查位置	檢查日期	
檢查時間	<input checked="" type="checkbox"/> 驗時前點點 <input type="checkbox"/> 不定期抽查		
施工工序	<input type="checkbox"/> 施工前 <input checked="" type="checkbox"/> 施工中 <input type="checkbox"/> 施工完成檢查		
檢查結果	<input type="checkbox"/> 檢查合格 <input checked="" type="checkbox"/> 有缺失需改正	對此檢查項目	
管理項目	抽查標準（定量定率）	實測抽查情形 （敘述抽查值）	檢查結果
*★定位-樁管定位	水平位置偏差不≥7.5cm		
*★鑽杆作業-套管全額入土時最大偏心	≤10cm		
*★鑽杆作業-鑽頭垂直高度	<1/200		
*★鑽杆作業-樁底沉澱及橋基標高差	$L = L_1 + D/2$ $m : D = \quad m$ 利用密實泵及空氣壓縮機之正循環方式，將精細之水泥漿泥排出，且於澆置混凝土 5 分鐘前不得停止。		
*★孔底處理-孔底淤泥排除	$L = L_1 + D/2$ $L = \quad m, D = \quad m$		
*★孔底處理-孔深檢驗	誤差不得加水，泵送機不得有堵塞		
*★混凝土澆置-預拌車、泵送機	保持在 13℃ 以上、32℃ 以下。		
*★混凝土澆置-溫度	厚度 22cm±4cm		
*★混凝土澆置-坍度試驗	≤0.15kg/m <sup>3</sup>		
*★混凝土澆置-氣壓值含量試驗	規定強度水中 350kgf/cm <sup>2</sup> 以試體製作組 每 20m <sup>3</sup> 至少取樣一組，除數未達 10m <sup>3</sup> 得併入同一組檢驗，餘數超過 10m <sup>3</sup> 時單獨取樣一組，且每支試體應檢驗混凝土抗壓強度試驗（組上）		
*★混凝土澆置-抗壓強度			

(a) 施工查驗表單範例



(b) 施工圖說頁面範例

圖 3 施工查驗表單及施工圖說頁面範例



詳圖分割

本研究將引入 LLMs 分析圖說頁面內容以進行精確檢索，然而將整頁施工圖說直接輸入至 LLMs 可能出現以下問題：圖說頁面所包含資訊量龐大，若完整輸入將導致總 token 過大，超過 LLMs 之上下文視窗上限。即便技術上能處理整頁圖說頁面內容，工程圖說的複雜性仍將造成極高的計算成本。由於 LLMs 仰賴自注意力機制進行語意理解，當輸入過長或結構過於複雜時，將導致模型效能下降，產出結果亦可能不具備足夠準確性與參考價值。為解決上述問題，本研究訓練並應用「Roboflow 3.0 Object Detection (Fast)」<sup>[18]</sup> 物件偵測模型，以自動化方式將整頁圖說頁面分割為獨立的詳圖與表格，如圖 4 所示。候選施工圖說在輸入至 LLMs 前將先行處理，透過物件偵測模型將其切割為可供後續檢索與分析之小區塊。這些被分割的詳圖與表格將被儲存並作為後續 LLM 處理的單位，藉此實現更高效率的圖說頁面語意分析與檢索。

兩階段圖說頁面檢索

兩階段圖說頁面檢索的處理流程設計為由兩個語言模型代理協同合作，以完成圖說頁面的檢索。第一個語言模型代理的任務是判斷給定圖說頁面之圖名或其各詳圖內容是否涵蓋了查驗表單中指定的檢查位置資訊。在這一階段，系統會將該頁圖說的圖名以及所有經物件偵測模型分割出的詳圖內容，連同查驗表單中標註的檢查位置資訊一起輸入至語言模型代理。語言模型代理根據這些資訊進行語意理解與比對，評估該圖說頁面與檢查位置的相關性，並判斷該頁是否包含所需資訊。若檢查位置驗證通過，第二個語言模型代理將進一步驗證這些詳圖是否包含足以驗證查驗表單中所列各查驗項目的必要元素。在此階段，系統將切割後的詳圖與各查驗項目一併輸入至語言模型代理中，語言模型代理根據這些資訊評估圖說頁面內容的相關性，確認是否可作為檢核的參考依據。透過此設計，即便初步相似度搜尋之結果無法完全判斷圖說內

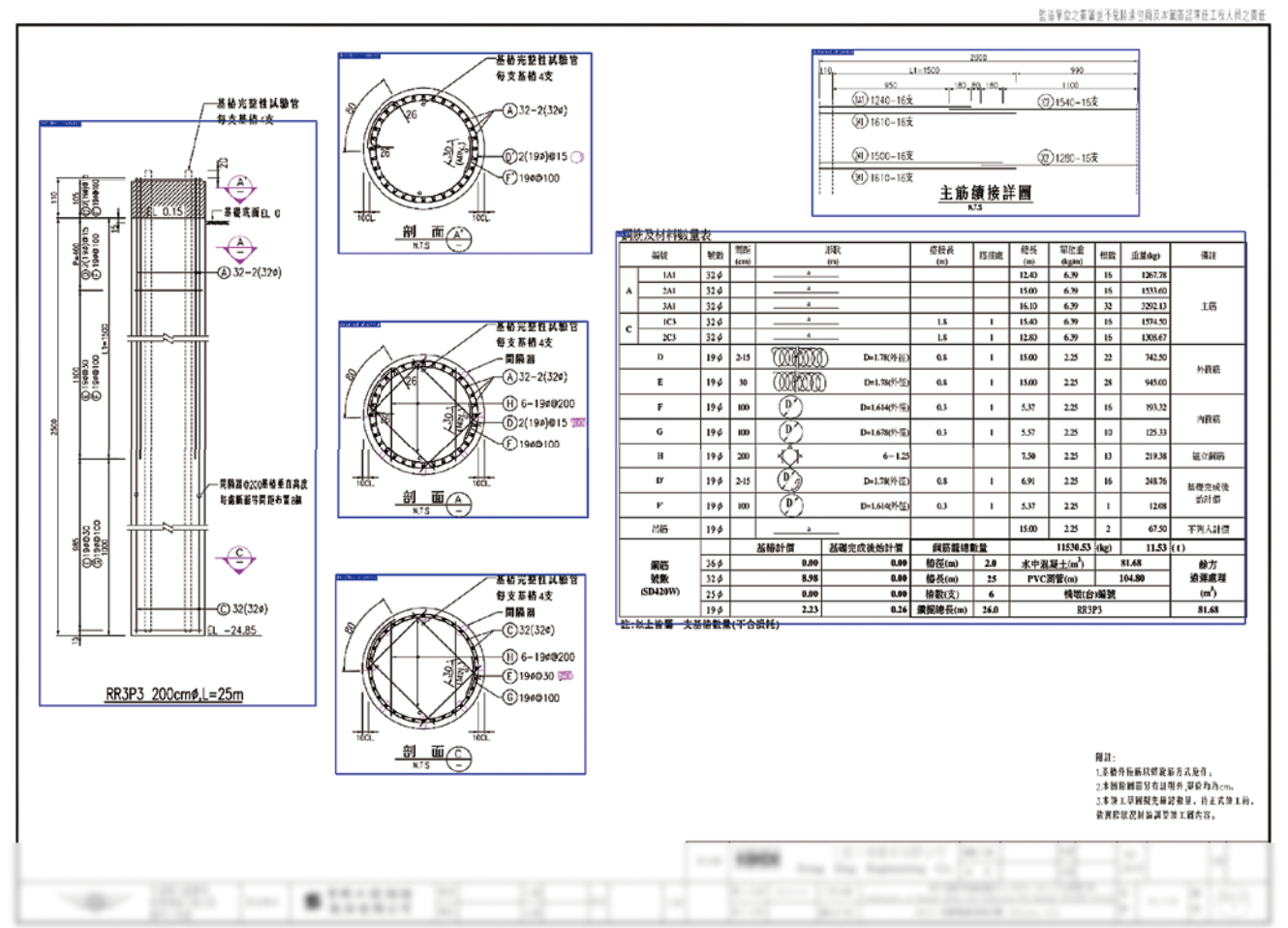


圖 4 透過一物件偵測模型實現之工程圖說詳圖分割範例

容細節，也能藉由語言模型代理的語意推理能力進一步篩選出更相關的圖面。

### 詳圖檢索

為了進一步檢索與輸入之查驗表單中指定的查驗項目最相關的詳圖或表格，本研究發展一詳圖檢索流程。在此流程中，自上一步驟檢索得之圖說頁面內每張詳圖將由一語言模型代理進行評估，判斷其和指定查驗項目間的相關性。語言模型代理會為每張詳圖提供一個介於 0 到 10 之間的相關性分數，用以表示該詳圖與指定查驗項目的契合程度，隨後選取分數最高的詳圖作為檢索結果。

### 詳圖校閱

詳圖檢索流程在多數情況下皆表現良好，但偶爾會出現評分不合導致檢索錯誤發生。這些錯誤源於 LLMs 固有的隨機性及不確定性。為了解決此問題，本研究提出了一詳圖校閱流程，用以校正錯誤檢索的詳圖。此方法不依賴額外的外部知識，而是利用預訓練模型提取的各詳圖內多數群體之平均影像特徵作為參考依據。圖 5 說明了本研究所提出的詳圖校閱流程，針對不同查驗表單的同一查驗項目，每張被檢索的詳圖會先透過預訓練的「ResNet-50」模型<sup>[19]</sup>提取影像特徵，並使用 K-means 分群為兩類。多數群體被視為正確檢索的詳圖，並將作為後續校正的參考依據。屬於少數群體的詳圖，將重新回到詳圖檢索流程。所有詳圖候選將再次進行特徵提取處理，並將其特徵與多數群特徵進行比較，計算平均歐氏距離（Euclidean Distance）。最後根據這些平均距離對詳圖候選進行重新排序，優先選擇與多數

群距離較近者做為校正後之檢索詳圖。此校閱流程能有效減少因 LLMs 的不確定性與隨機性所導致詳圖錯誤檢索錯誤，進而提升整體詳圖檢索結果的可靠性。

### 代理式反思

代理式反思的運作流程涉及兩個語言模型代理：符合性檢核代理與檢核結果評估代理。符合性檢核代理負責根據輸入查驗表單中指定的查驗項目與抽查標準，以透過層級式詳圖檢索得到的詳圖進行檢核。相對於符合性檢核代理，檢核結果評估代理負責評估符合性檢核代理所產出的檢核結果，以確保檢核過程的準確性與完整性。若檢核結果評估代理從結果中發現任何錯誤或說明不足之處，其將提供改善建議並傳回給符合性檢核代理，促使其進行修正後的檢核。

### 實驗驗證

為驗證所提出系統的可行性，本研究進行一系列消融實驗並驗證不同類型工程圖說，測試層級式詳圖檢索流程內各模組的可行性與可擴展性。以下各小節分別說明資料集、消融實驗配置及評估指標之細節。

### 資料集

本研究之實驗將透過來自一真實橋梁工程專案之資料集進行驗證，該橋梁工程專案的施工圖說共計 1650 頁，涵蓋多種分項工程。本研究選取其中兩項分項工程的現場查驗表單作為實驗對象進行消融實驗，測試層級式詳圖檢索流程中各模組對於檢核任務之可行性及可擴展性，其細節如下：

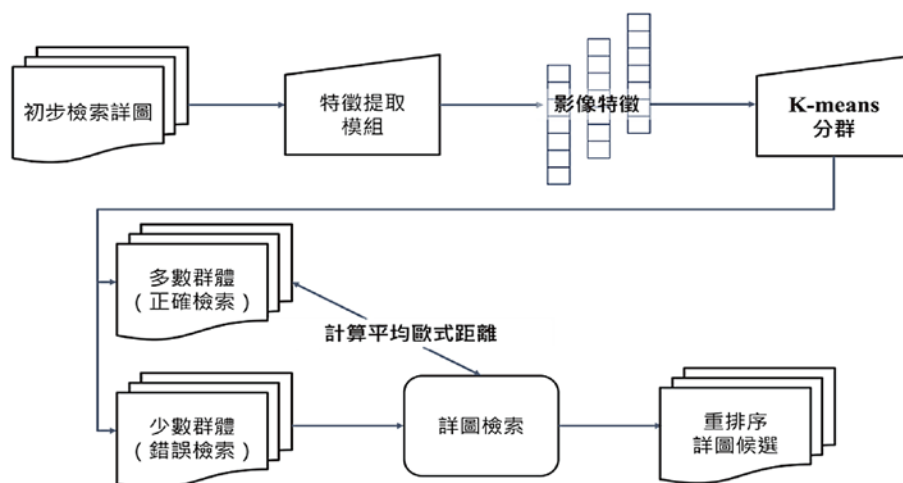


圖 5 詳圖校閱運作流程圖示意

1. RC 擋土牆工程分項工程資料集

此資料集包含 9 份針對不同配置之 RC 擋土牆工程施工的現場查驗表單，對應一頁工程圖說頁面，屬於難度較低之資料集。每份表單均包含表單名稱、工程名稱、檢查位置與五組需透過圖說進行檢核的查驗項目及對應抽查標準（共計 45 組）。

2. 全套管基樁分項工程資料集

此資料集包含 90 份針對不同全套管基樁施工位置的現場查驗表單，分佈於六個橋段，對應七頁工程圖說頁面，屬於難度較高之資料集。每份表單同樣包含表單名稱、工程名稱、檢查位置，以及五組需透過圖說進行檢核的查驗項目及對應抽查標準（共計 450 組）。

消融實驗配置

為驗證層級式詳圖檢索流程中各模組在圖說檢核與檢核任務中的貢獻，本研究設計五種不同的圖說檢索組合方式進行消融試驗並透過 OpenAI 的「o4-mini」<sup>[20]</sup> 進行實驗。所涉及的模組包括詳圖分割、兩階段圖說頁面檢索、詳圖檢索以及詳圖校閱，五種圖說檢索流程細節說明如下：

1. 基礎

透過一語言模型自相似度搜尋結果得之候選圖說中篩選出一圖說頁面作為檢索結果。

2. 詳圖分割

首先將候選圖說頁面進行詳圖分割，隨後透過一語言模型自候選圖說中篩選出一圖說頁面作為檢索結果。

3. 兩階段圖說頁面檢索

將圖說頁面檢索流程拆分為兩階段的驗證流程，先驗證指定檢查位置後再驗證圖說頁面內容。

4. 詳圖檢索

改為使用 LLMs 檢索最為相關的單一詳圖作為檢索結果。

5. 詳圖校閱

加入額外的校閱步驟以減少因 LLMs 在檢索正確詳圖時的不確定性所導致的錯誤。

評估指標

為評估所提出方法的整體效能與可行性，本研究採用三項指標作為評估依據，此三項指標說明如下：

1. 圖說頁面檢索準確率

本研究假設每份查驗表單在提供之施工圖說檔案中皆存在一最具關聯性的圖說頁面，可作為檢核之依據。圖說頁面檢索準確率即根據層級式詳圖檢索流程是否正確檢索出該正確圖說頁面進行計算。

2. 詳圖檢索準確率

針對每一查驗項目，本研究假設在提供之施工圖說檔案之最具關聯性的圖說頁面中存在一張最具關聯性的詳圖，可作為檢核之依據。詳圖檢索準確率即根據層級式詳圖檢索流程是否正確檢索出該正確詳圖進行計算。

3. 符合性檢核準確率

針對每一查驗項目，本研究假設在提供之施工圖說檔案皆定義有一正確的抽查標準。符合性檢核準確率即根據系統是否能正確判斷該查驗項目是否符合圖說要求並能否在發現抽查標準有誤時能正確修正該標準進行計算，亦為本研究提出架構之整體指標。

實驗結果與討論

消融實驗結果

表 1 彙整了以 RC 擋土牆資料集與全套管基樁資料集進行之消融實驗的準確率結果。綜合所有消融實驗結果，最佳的結果來自於整合所有四個模組的配置。此配置於難度較低之 RC 擋土牆分項工程資料集上達到 100.0% 的圖說頁面檢索準確率、100.0% 的詳圖檢索準確率以及 100.0% 的符合性檢核準確率，並於難度較高

表 1 消融實驗結果比較

實驗配置		RC 擋土牆			全套管基樁		
LLM	檢索方式	圖說頁面檢索	詳圖檢索	符合性檢核	圖說頁面檢索	詳圖檢索	符合性檢核
o4-mini	基礎	100.0	-	24.4	66.7	-	20.0
	+ 詳圖分割	100.0	-	100.0	74.4	-	62.9
	+ 兩階段圖說頁面檢索	100.0	-	100.0	93.3	-	80.4
	+ 詳圖檢索	100.0	100.0	100.0	93.3	80.2	74.2
	+ 詳圖校閱	100.0	100.0	100.0	93.3	92.4	83.3



之全套管基樁分項工程資料集達成了 93.3% 的圖說頁面檢索準確率、92.4% 的詳圖檢索準確率以及 83.3% 的符合性檢核準確率。

## 消融實驗結果分析

### 基礎

此檢索方式由相似度搜尋結果以大型語言模型重新篩選出一圖說頁面作為檢索結果，雖然在難度較低的 RC 擋土牆資料集中，此方法可達到 100.0% 的圖說頁面檢索準確率（由於最佳相似度結果即為正解），但在高難度的全套管基樁資料集中表現顯著不佳，圖說頁面檢索準確率僅為 66.7%，兩資料集之符合性檢核準確率更降至僅 24.4% 及 20.0%，為所有檢索策略中最低。此結果顯示單純且未經調整的檢索方式並不適用於圖說溯源檢核任務，需引入更進階之技術方能有效提升整體效能。

### 詳圖分割

當加入詳圖分割模組後，兩資料集之符合性檢核準確率有顯著提升。在 RC 擋土牆資料集中，由 24.4% 大幅提升至 100.0%。而在全套管基樁資料集中，亦由 20.0% 提升至 62.9%。此結果凸顯將施工圖說頁面切分為較小且可處理的小區塊對於模型推理能力的幫助，能有效降低處理圖說資訊的複雜度，使 LLMs 更能聚焦於與查驗項目及抽查標準高度相關的內容，提升整體檢核的準確率。

### 兩階段圖說頁面檢索

加入兩階段圖說頁面檢索機制後，全套管基樁資料集之圖說頁面檢索與符合性檢核準確率皆顯著提升。由於較簡易的 RC 擋土牆資料集在前述檢索策略下即已達成 100.0% 的圖說頁面檢索準確率，故此模組對該資料集無進一步改善空間。而在全套管基樁資料集中，圖說頁面檢索準確率由 74.4% 提升至 93.3%，符合性檢核準確率同步提升至 80.4%。此結果顯示將圖說頁面檢索任務拆解為兩個連續且相對簡化的步驟，能有效模擬實務中工程師尋找正確圖說頁面的推理流程，顯著提升語言模型在辨識正確施工圖說頁面時的準確率。

### 詳圖檢索

引入詳圖檢索模組後，在全套管基樁資料集中之符合性檢核準確率下降，由 80.4% 降至 74.2%。此結果顯示，雖然詳圖檢索旨在提供更聚焦的輸入內容，該步驟仍存在檢索錯誤詳圖的風險。若檢索錯誤詳圖，

反而可能導致缺乏關鍵資訊，進而影響檢核準確性。然而若進一步分析全套管基樁資料集中，在「成功檢索到正確圖說」的前提下的檢核表現，符合性檢核準確率由原先的 86.2%（即 93.3% 圖說頁面檢索準確率內達成 80.4% 符合性檢核準確率）提升至 92.5%（即 80.2% 詳圖檢索準確率內達成 74.2% 符合性檢核準確率），展現出提升趨勢。此一現象指出，儘管詳圖檢索步驟可能因錯誤檢索而拉低整體檢核準確率，其對於成功檢索到正確圖說的樣本而言，卻能顯著提升檢核表現。這是因為相較於提供整頁圖說資訊，將單一、聚焦的詳圖提供予語言模型代理，不僅可有效縮小分析範圍、降低自注意力機制的運算負擔，更有助於模型聚焦於與查驗項目及抽查標準直接相關的內容，從而減少推理錯誤，提升檢核判斷的精確性。

### 詳圖校閱

最終加入之詳圖校閱模組進一步提升了詳圖檢索準確率與整體符合性檢核準確率兩項關鍵指標的表現。在全套管基樁資料集中，詳圖檢索準確率提升至 92.4%，符合性檢核準確率則提升達 83.3%，皆為整體最佳表現。詳圖校閱步驟能有效降低單獨僅有詳圖檢索步驟所帶來的檢索錯誤，在減少檢索錯誤的同時，又保有將單一且最聚焦的詳圖提供予語言模型代理處理的優勢。此最後一步驟不僅強化了整體符合性檢核流程的穩健性與可靠性，也使本研究所提出之架構在所有測試的檢索方式中達成最佳的整體檢核效果。

## 結論

以施工查驗表單進行現場工程查驗是現代營建工程中確保結構品質的關鍵步驟。傳統上此流程仰賴工程師人工對照篇幅龐大且結構複雜之施工規範與圖說對查驗項目及標準逐條檢核，不僅耗時費力亦易出錯。儘管現有 LLMs 已具備初步理解施工規範與圖說之能力，然其尚未被有效導入至基於工程圖說之施工查驗表單 ACC 應用中。主要挑戰包括 LLMs 的上下文視窗限制，以及施工圖說間高度視覺相似性所造成之檢索困難。為解決上述問題，本研究提出一融合多模態 RAG 技術與代理型工作流程之創新架構，以層級式詳圖檢索逐層篩選出與查驗項目相關之圖說頁面與對應詳圖。系統以一物件偵測模型進行詳圖分割，以降

低視覺輸入的處理負擔，並於檢索流程之最終步驟導入詳圖校閱機制進一步確認詳圖內容之正確性。最終於符合性檢核階段透過代理式反思，結合兩語言模型代理之迭代驗證，提升檢核準確率。本研究以一實際橋樑工程案之施工圖說與查驗表單作為測試資料集進行驗證，結果顯示所提出系統於圖說頁面檢索、詳圖檢索與符合性檢核三項指標上，於較簡易之 RC 擋土牆資料集上全部達成 100.0% 的準確率，並於較高難度之全套管基樁資料集分別達成 93.3%、92.4% 與 83.3% 的準確率，展現了整體系統的可擴展性。消融實驗分析進一步驗證各模組對整體效能之貢獻：詳圖分割能有效降低語言模型之輸入複雜度；兩階段圖說頁面檢索提升圖說頁面辨識準確性；詳圖檢索有助於聚焦模型處理範圍，降低模型處理負擔；詳圖校閱則能顯著降低錯誤檢索導致之後續檢核失敗，提升整體系統穩健性與可靠性。本研究針對工程領域圖說在資訊結構與視覺呈現上之特殊性，設計出一具擴展性、穩健性與高準確性的層級式詳圖檢索流程，整合於基於工程圖說之施工查驗表單 ACC 架構中。此架構於處理圖說量龐大且內容複雜之大型建設專案時，展現出高度實用潛力。透過補足現有多模態 RAG 框架於工程圖說檢索應用上的缺口，為工程領域實務導入可行的 ACC 系統提供一項具體且具應用價值的技術基礎。

## 誌謝

本研究由國家科學及技術委員會與中興工程顧問股份有限公司透過「以自然語言處理擷取施工規範查驗項目並輔助查驗表單的設計與檢核」計畫 NSTC 112-2622-E-002-018 與 NSTC 113-2622-E-002-027 補助支持。作者亦感謝國震中心與臺大土木合設 AI 研究中心以及國家高速網路與計算中心提供運算與儲存資源。

## 參考文獻

1. Bureau of Land Management, MS 9103: Facility Construction, Manual section Rel. 9-402 MS 9103, U.S. Department of the Interior, Bureau of Land Management, Washington, D.C., Apr. 2014. Supersedes Rel. 9-325.
2. 張建信 (2021), 「以工程實務面探討公共工程施工三級品質管理制度之執行成效」, 碩士論文, 國立臺灣大學土木工程學研究所, 臺北。
3. Xu, X., Jeon, J., Zhang, Y., Yang, L., and Cai, H. (2021). "Automatic Generation of Customized Checklists for Digital Construction Inspection," *Transportation Research Record*, Vol. 2675, No. 5, pp. 418-435.
4. Goh, W., Jang, J., Park, S. I., Choi, B.-H., Lim, H., and Zi, G. (2024). "Automated Compliance Checking System for Structural Design Codes in a BIM Environment," *KSCE Journal of Civil Engineering*, Vol. 28, No. 10, pp. 4175-4189.
5. Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.-t., Rocktäschel, T., Riedel, S., and Kiela, D. (2020). "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," *Advances in Neural Information Processing Systems*, Vol. 33, pp. 9459-9474.
6. Langchain AI, Multi-Modal RAG, (2024). Available: [https://github.com/langchain-ai/langchain/blob/master/cookbook/Multi\\_modal\\_RAG.ipynb](https://github.com/langchain-ai/langchain/blob/master/cookbook/Multi_modal_RAG.ipynb) (accessed: 2025-02-11)
7. Weng, L., LLM-Powered Autonomous Agents, (2023). Available: <https://lilianweng.github.io>
8. Liu, X., Li, H., and Zhu, X. (2023). "A GPT-based Method of Automated Compliance Checking through Prompt Engineering".
9. Yang, F. and Zhang, J. (2024). "Prompt-based Automation of Building Code Information Transformation for Compliance Checking," *Automation in Construction*, Vol. 168, 105817.
10. Uhm, M., Kim, J., Ahn, S., Jeong, H., and Kim, H. (2025). "Effectiveness of Retrieval Augmented Generation-based Large Language Models for Generating Construction Safety Information," *Automation in Construction*, Vol. 170, 105926.
11. Villena Toro, J., Wiberg, A., and Tarkian, M. (2023). "Optical Character Recognition on Engineering Drawings to Achieve Automation in Production Quality Control," *Frontiers in Manufacturing Technology*, Vol. 3.
12. Jamieson, L., Moreno-García, C., and Elyan, E. (2024). "Towards Fully Automated Processing and Analysis of Construction Diagrams: AI-Powered Symbol Detection," *International Journal on Document Analysis and Recognition (IJDAR)*, Vol. 28, pp. 71-84.
13. Xuesong, X., Gang, X., Li, S., Xia, Z., Peixi, W., Yuanming, Z., and Zhenbo, C. (2025). "Associative Reasoning for Engineering Drawings using an Interactive Attention Mechanism," *Automation in Construction*, Vol. 170, 105942.
14. Doris, A.C., Grandi, D., Tomich, R., Alam, M.F., Ataei, M., Cheong, H., and Ahmed, F. (2024). "DesignQA: A Multimodal Benchmark for Evaluating Large Language Models' Understanding of Engineering Documentation".
15. Yue, X., Ni, Y., Zhang, K., Zheng, T., Liu, R., Zhang, G., Stevens, S., Jiang, D., Ren, W., Sun, Y., et al. (2024). "MMMU: A Massive Multi-Discipline Multimodal Understanding and Reasoning Benchmark for Expert AGI," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9556-9567.
16. OpenAI, text-embedding-3-small, (2025). Available: <https://platform.openai.com/docs/models/text-embedding-3-small>
17. Neo4j, Inc., Neo4j Graph Database, Version 2025.04 (2025).
18. Dwyer, B., Nelson, J., Hansen, T., et al. (2024). Roboflow (version 1.0) [software], Computer Vision.
19. He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep Residual Learning for Image Recognition," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770-778.
20. OpenAI, o4-mini (2025-04-16 version), (2025). Available: <https://platform.openai.com/docs/models/o4-mini> 