



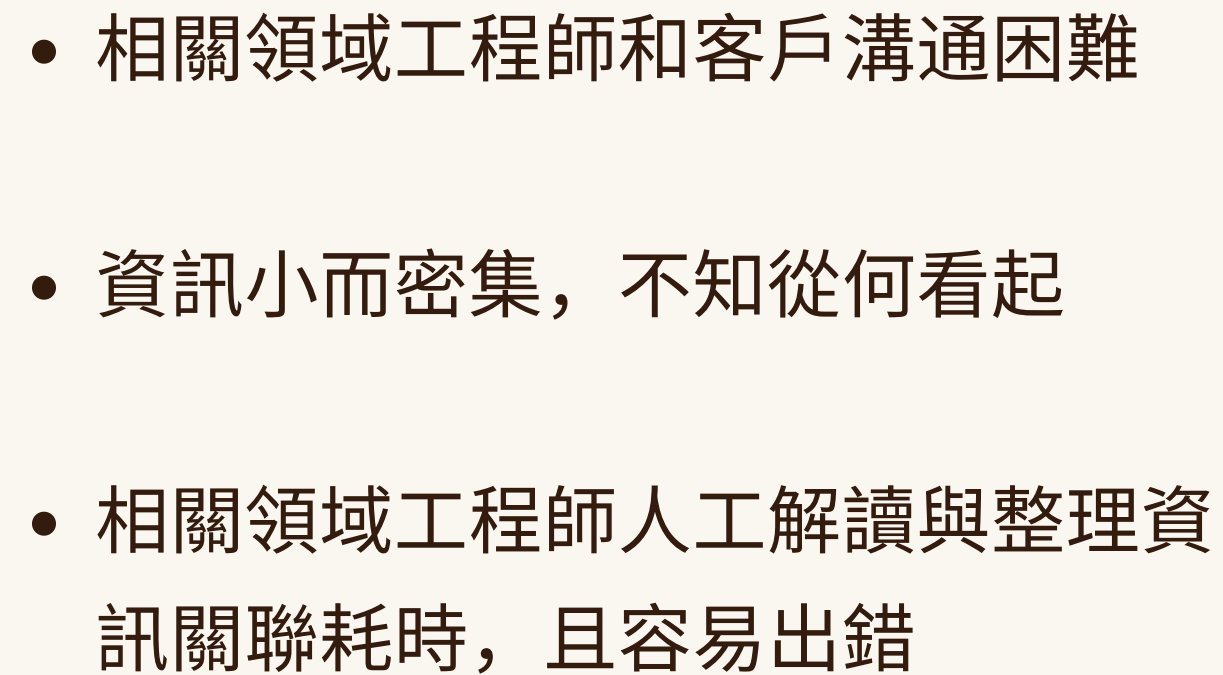
Towards Understanding Engineering Drawings with AI

指導教授：藍崑展

專題成員：柯冠先,陳奕帆,鄭永順
吳芸慧,賴鵬豐,馮家俊



情境介紹



研究動機與目的

Problem Statement

- 傳統工程與製造業中，人工解讀與整理圖紙中的符號與文字關聯極為耗時且容易出錯。
- 密集的尺寸線、極端長寬比的物件、以及局部貪婪演算法造成的搶位與錯配，導致自動化關聯辨識困難。

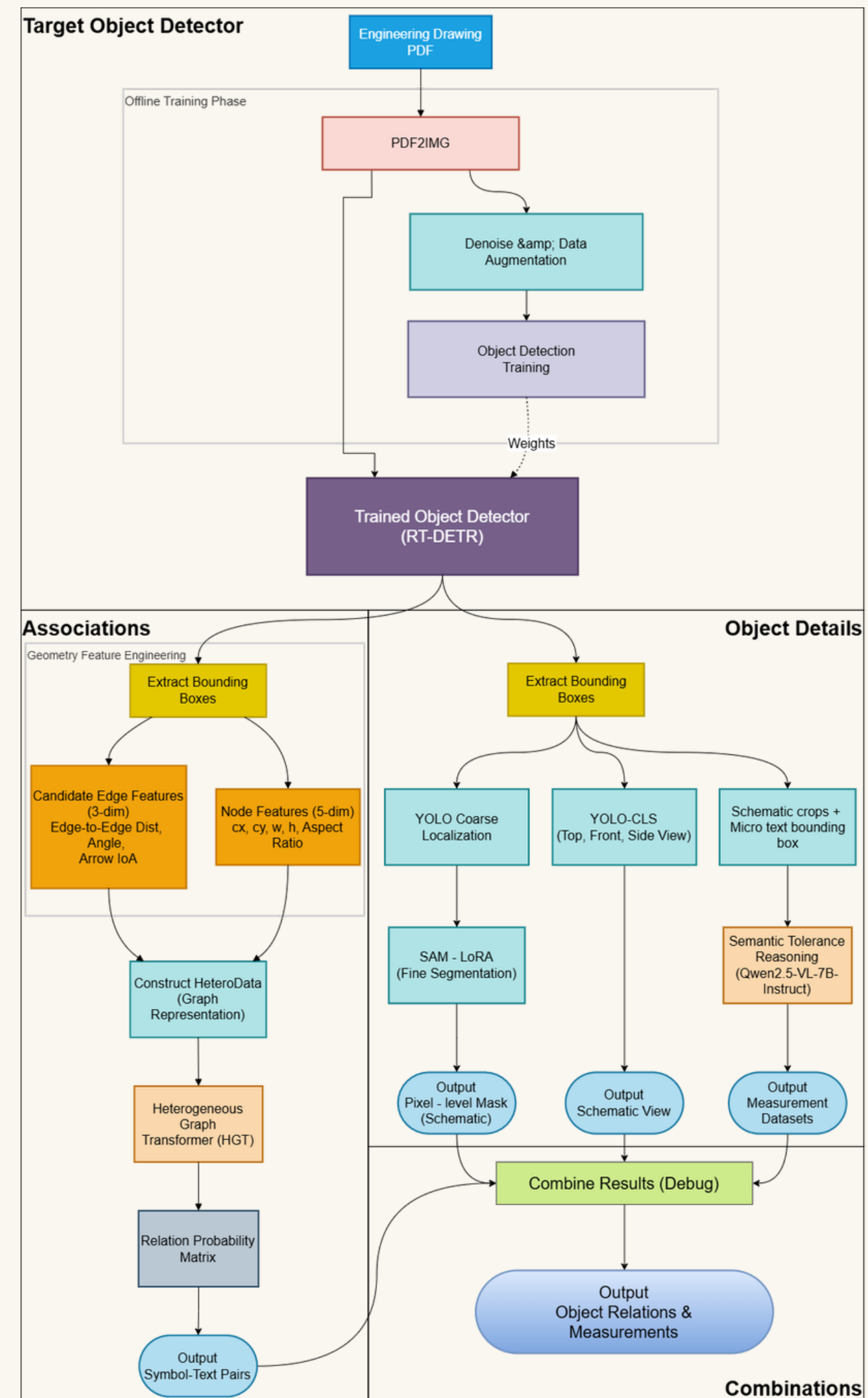
Goals

- 開發一套基於 AI 的自動化系統，直接輸入工程圖紙影像，自動偵測並輸出所有相互關聯的「符號與文字配對 (Symbol-Text Pairs)」。
- 突破空間距離限制，不論文字與符號在空間上是緊鄰或遠距離分布，皆能準確辨識。

ARCHITECTURE

系統架構圖

- Target Object Detector: 圖紙物件檢測模組 (RT-DETR)
 - 圖紙檢測內容物分為6大類別
- Object Details: 文本識別處理 (Qwen VLM)
 - 抓出尺寸公差值，為後續計算零件尺寸做應用
- Associations: 異質圖神經網路關聯 (HGT)
 - Schem ← Arrow ← Text 為主

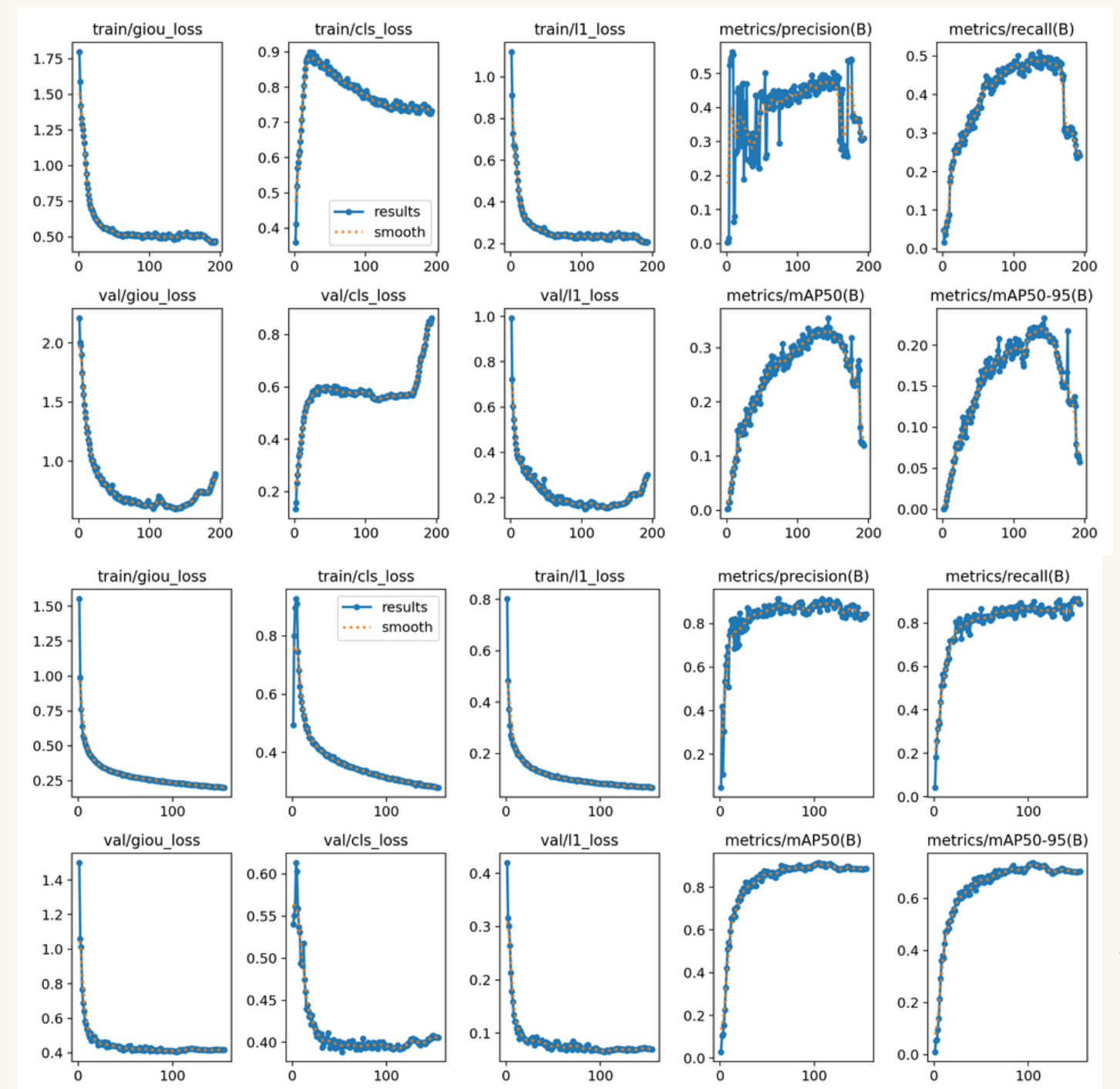


Object Detection



- YOLOV11使用CNN架構無法有效地偵測細長的物件，而使用transformer架構的RT-DETR能夠更好的看到細長的物件
- 透過將圖片中關鍵的部分擷取下來，當成新的data輸入模型進行訓練，能夠讓模型學習小物件細節，改善在downsample時細節消失的問題
- 使用多階段的訓練方式，逐漸提高解析度，解決在有限記憶體下，batch size不足模型難以收斂的問題

	YOLOV11	RT-DETR
mAP@0.5	0.707	0.876



RESEARCH DETAIL

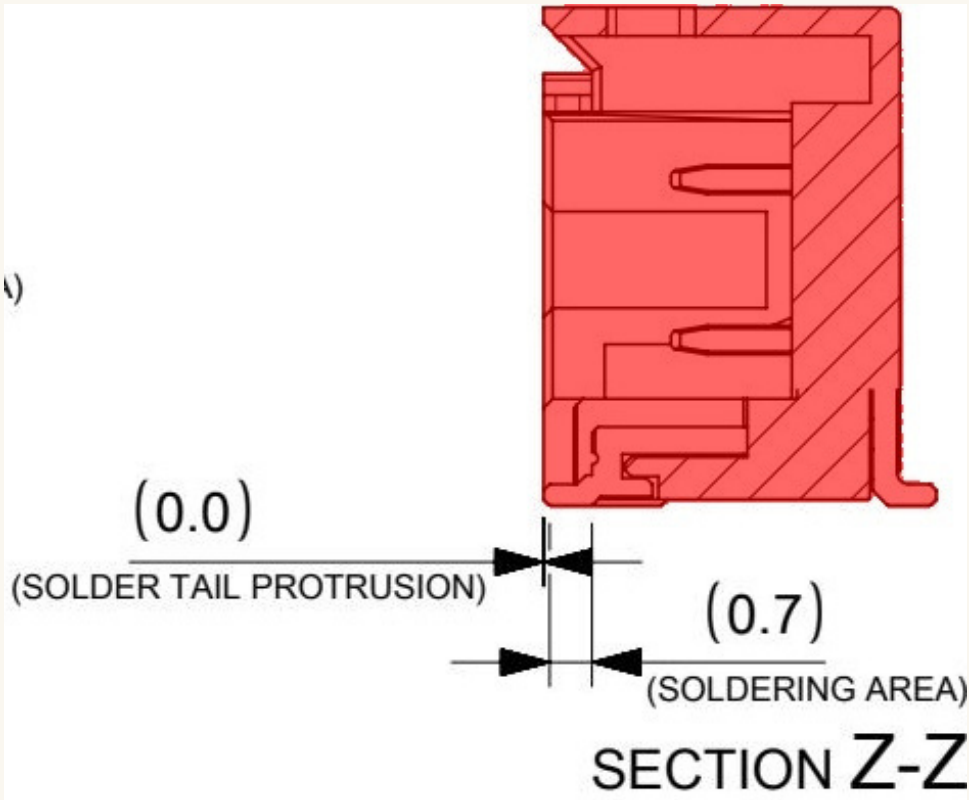
Segmentation



基於 LoRA 微調與兩階段級聯管線 (YOLO+SAM)，克服 RGB 模型在黑白工程圖面的領域差異，實現像素級分割。透過兩階段增強（雜訊、Copy-Paste）與損失函數優化，提升抗過擬合能力，有效壓縮推論延遲。

- **核心技術：**LoRA 微調跨越 Domain Gap、YOLO+SAM-LoRA 級聯加速。
- **數據與增強：**雜訊添加與跨階段 Copy-Paste 強化困難樣本。
- **最佳化：**BCE 與 Dice Loss 權重調整，提升微小特徵分割連續性。

image name	iou_without_Lora&YOLO	iou_without_Lora_with_YOLO	iou_LORA+YOLO
v1_GEN_Z_B62-10006-0041_1_crop_0_jpg.rf.i4dfiPvA6tq1YX0maSt8.jpg	0.561886051	0	0
v1_GEN_Z_B62-10006-0041_1_crop_1_jpg.rf.aUHNMKWEIM4gKYngj6Am.jpg	0.15784847	0.464735927	0.766477521
v1_GEN_Z_B62-10006-004F_1_crop_1_jpg.rf.AHEZi9qKdRiFh49pSruS.jpg	0.290689488	0.899133979	0.955564128
v1_batch1_B21_1_crop_0_jpg.rf.DW5AH8zjyyFprwYGjflh.jpg	0.663026092	0.933689277	0.973070712
v1_batch1_B21_1_crop_2_jpg.rf.RYIYdbNGBqzcYFgFdOg5.jpg	0.002531775	0.95861118	0.976432271
v1_batch1_B22_1_crop_1_jpg.rf.lIqPomDFivAyS1BAoYi8.jpg	0.10491464	0.875093557	0.981106615
v1_batch1_B22_6_crop_1_jpg.rf.6jq48NaLvXWlsLWXjgow.jpg	0.807157642	0.81138879	0.938259638
v1_batch1_B22_8_crop_0_jpg.rf.RdpKMRXOfGQ13qxJ9bM.jpg	0.789528155	0.505168944	0.51336836
v2_MICRO_FIT_B20-20002-0003_1_crop_6_jpg.rf.6oSajtMsl27qYalN6tms.jpg	0.694505495	0.772637619	0.884206799
v2_MICRO_FIT_B20-20002-0003_2_crop_4_jpg.rf.n8CVxmKpMmcu3X0N8BFp.jpg	0.440730159	0.703729169	0.920913691
v2_MICRO_FIT_B20-20002-0016_1_crop_0_jpg.rf.YsbusBnlMAWOvUuuGC7r.jpg	0.577085888	0.874375236	0.970555781
v2_MICRO_FIT_B20-20002-0016_1_crop_2_jpg.rf.tbdiS1ojSBieljGNrL0H.jpg	0.493637349	0.925026628	0.973546328
v2_MICRO_FIT_B20-20002-0016_2_crop_0_jpg.rf.5MFYrHED91cx927G8WZ1.jpg	0.254312042	0.915847333	0.925317919
v2_MICRO_FIT_B20-20002-0016_2_crop_1_jpg.rf.pGfhYfGWcliqs9Fhso6b.jpg	0.626268778	0.909750522	0.934509844
v2_MICRO_HI_B21-20003-0008_1_crop_1_jpg.rf.a0129Kh989G35FIUux8s.jpg	0.808555388	0.848723386	0.950536686
v2_MINITEK_B20-20007-000V_2_crop_1_jpg.rf.TV4mkv2cCrCPn0Kt9GV8.jpg	0.868014921	0.773878228	0.755294884
v2_MINITEK_B20-20007-001M_1_crop_0_jpg.rf.0ITk5PK3dADgiGcUM0sS.jpg	0.857498694	0.89514486	0.971768049
v2_MINITEK_B20-20007-001M_2_crop_1_jpg.rf.u0hZuYrQUSibL3nqJol0.jpg	0.829528495	0.94748976	0.973989238
v2_MULTI-TRAK_B62-10001-004X_1_crop_2_jpg.rf.cyQiLMfry8xyrCWedLxu.jpg	0.480619208	0.628618643	0.769659185
v2_MULTI-TRAK_B62-10001-004X_1_crop_5_jpg.rf.s66yyBzcx9qnRWP0jVzh.jpg	0.386284168	0.902147088	0.929548564
v2_MULTI-TRAK_B62-10001-004X_1_crop_6_jpg.rf.h7EUVjGrabMTV1jVGTbL.jpg	0.958783349	0.959587437	0.980297777
mIoU	0.554924107	0.785941789	0.859258285



RESEARCH DETAIL

Tolerance semantic parsing

-**目標**：突破傳統 OCR 限制，自動將「公差表」精準配對至零件尺寸

-**核心模型與工具**：RT-DETR (初始座標定位), Qwen2.5-VL 7B-instruct, llama.cpp (GGUF 本地端部署), Python

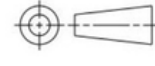
-**測試資料**：265 零件 (schematic) 以及相應的公差表

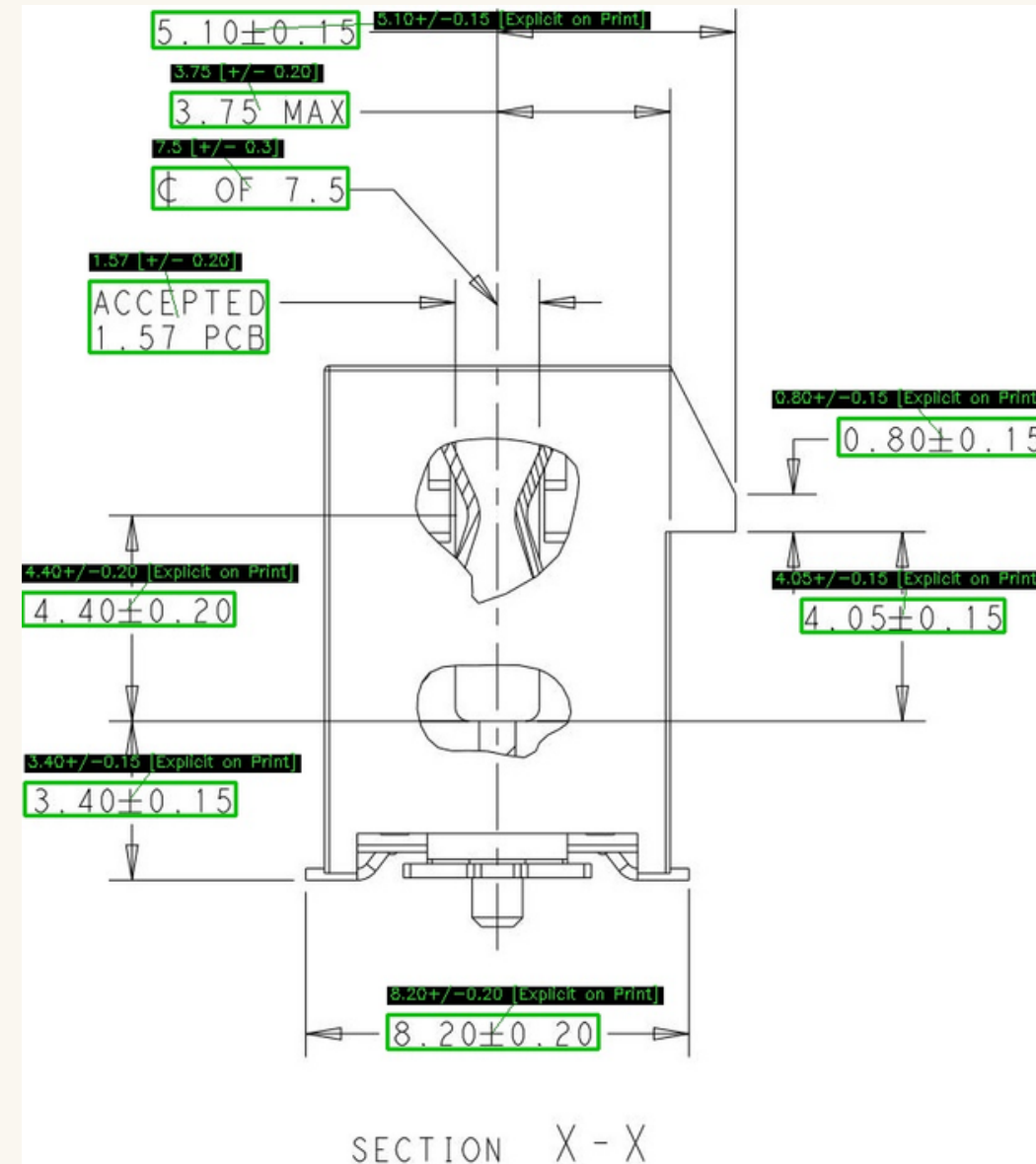
-**核心機制**：

Step 1 表單記憶 (Memorization): VLM 讀取公差表，建立全域 JSON 規則庫。

Step 2 動態匹配 (Hybrid Fusion): VLM 萃取尺寸，Regex 判斷小數位數配對公差。

Step 3 視覺化輸出 (QA Rendering): 讀取座標繪製圖框，透過防碰撞演算法確保標註不重疊。

THIS DRAWING IS A CONTROLLED DOCUMENT.		DWN 22OCT2019 CHK SANDERS YAN		TE Connectivity	
DIMENSIONS: mm		TOLERANCES UNLESS OTHERWISE SPECIFIED:		PRELIMINARY	
		0 PLC ± 1 PLC ±0.3 2 PLC ±0.20 3 PLC ±0.050 4 PLC ± ANGLES ± FINISH ±		APVD 22OCT2019 NAME PCI EXPRESS CARD EDGE CONNECTOR PRODUCT SPEC GEN5, STD	
MATERIAL -		WEIGHT -		SIZE A2	CAGE CODE 00779
				DRAWING NO C-2364427	RESTRICTED TO -
CUSTOMER DRAWING				SCALE 6:1	SHEET 1 OF 6 REV 9



公差資料放在零件數字
(implicit tolerance)

• **accuracy**

table memorization

82/97 tables = 84,5%

OCR parsing

220/265 schematic = 83,01%

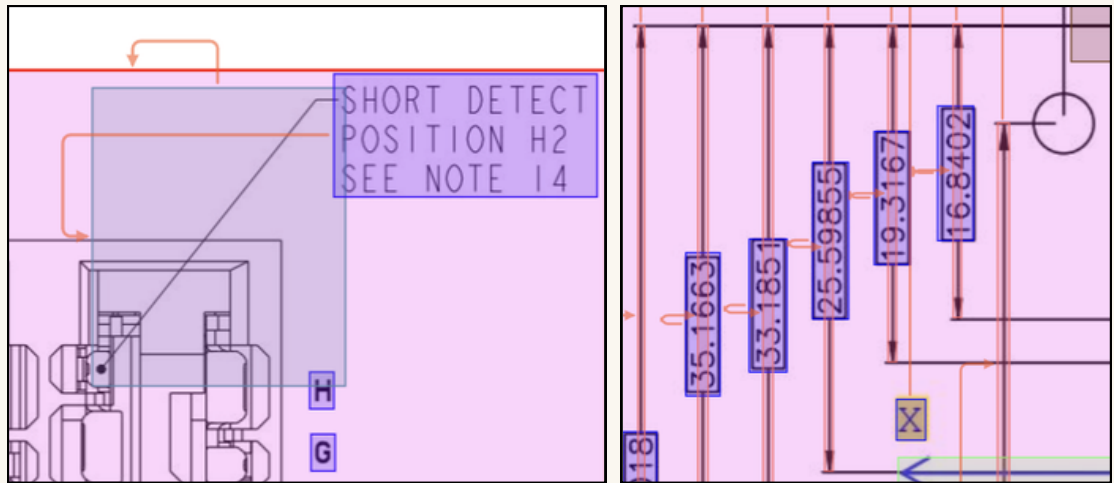
RESEARCH DETAIL

Associations

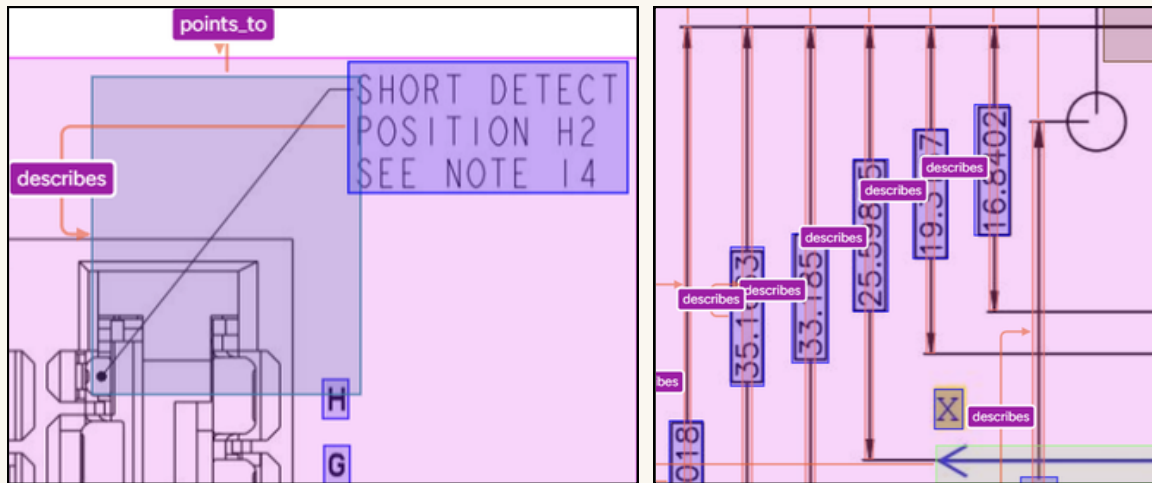
- 基礎目標：關聯零件 (*Schem*)、尺寸箭頭 (*Arrow*) 與文字 (*Text*)
- 模型：XGBoost, Heterogeneous Graph Transformer (HGT)
- 測試樣本數：正樣本 **2555** 筆標記
- HGT 優勢
 - 準確度 HGT 略高，整體表現更穩定。
 - 以 **Graph (多層級)** 做為輸出分類，而非單純的二維關聯。
 - 合理的多維輸出不會降低模型訓練成效，在後續關聯更多節點類別 (*BoundingBox*) 依然可**高彈性**調整。

	XGBoost (baseline)	HGT (Ours)
Accuracy	0.98	0.996 (<i>Similar results</i>)

Outputs of **XGBoost (2 types)**: *Relation (1)*, *None (0)*



Outputs of **HGT (3 types)**: *describes (1)*, *points_to (2)*, *None (0)*



研究貢獻

打造出一套高穩健度、兼具全局視野與商用落地價值的自動化工程圖紙理解系統。

優化方向

- 未來將持續優化在極端複雜或新型態工程圖紙下的識別表現
- 進一步提升跨領域自動理解的泛化能力。





Thank You For Listening



指導教授：藍崑展

專題成員：柯冠先,陳奕帆,鄭永順
吳芸慧,賴鵬豐,馮家俊