

從異常偵測走向理解與行動：人工智慧於智慧製造檢測的技術演進與可信部署

From Anomaly Detection to Understanding and Action: The Evolution and Trustworthy Deployment of AI for Smart Manufacturing Inspection

林士智、賴尚宏

Shih-Chih Lin, Shang-Hong Lai

異常偵測是智慧製造品質檢測的關鍵技術，但實務仍面臨缺陷樣本稀少、未知異常難以事先定義，以及模型難以跨產品與跨感測條件泛化等挑戰。本文從人工規則、正常資料學習、合成異常、少樣本與零樣本方法出發，說明視覺基礎模型、語意－結構融合，以及統一與跨域模型的技術演進，並延伸討論視覺語言問答、代理式取證與可信部署。本文強調，任務準確度、異常定位、因果視覺依賴、誤報與漏檢，以及部署安全，應分別評估，才能建立可驗證、可追溯，並能協助工程師決策的異常理解系統。

Anomaly detection is an important technology for quality inspection in smart manufacturing, yet practical deployment remains difficult because defect samples are scarce, unknown anomalies cannot be exhaustively defined, and models often generalize poorly across products, sensors, and production conditions. This article reviews the evolution from rule-based inspection and normal-only learning to synthetic anomaly generation, few-shot and zero-shot detection, visual foundation models, semantic-structural fusion, and unified cross-domain systems. It further discusses the transition from anomaly detection to anomaly understanding through vision-language question answering, structured reasoning, and agentic evidence acquisition using zooming, normal-reference retrieval, and difference maps. Particular attention is given to trustworthy deployment. Task accuracy, anomaly localization, causal visual dependence, false alarms, fatal misses, and operational safety are related but distinct properties and should be evaluated separately. A reliable inspection system should therefore provide verifiable and traceable evidence, preserve data and model lineage, and support engineers in making risk-aware decisions rather than merely producing fluent explanations.

一、前言

在現代製造環境中，產品外觀、組裝品質與設備狀態，直接影響良率、可靠性與生產成本。常見的異常包括刮痕、裂縫、污染、凹陷、元件漏裝、位置偏移及晶圓圖案異常。這些缺陷所佔比例通常極低，卻可能在後續製程中被放大，進而造成重工、報廢、客訴，甚至安全風險。運用相機、顯微鏡、深度感測器與人工智慧建立自動光學檢測 (automatic optical inspection, AOI) 系統，已成為智慧製造的重要基礎。

異常偵測與一般影像分類有本質上的差異。分類模型可從大量標註影像中學習固定類別；工業現場卻往往只有充足的正常資料，真實缺陷不僅稀少，也無法涵蓋未來可能出現的所有未知異常。因此異常偵測的核心並非記住每一種缺陷，而是建立正常產品應有的外觀、結構與關係，再找出偏離正常狀態的證據。相關綜述將主要挑戰歸納為訓練資料稀少、感測模態多樣及異常層級複雜三項；近年的方法也逐步從針對每項產品個別訓練，轉向運用基礎模型先驗進行跨產品泛化^(1, 2)。

當異常偵測走入真實產線後，還必須進一步思考第四個問題：模型分數高，是否就代表能夠安全部署？現代異常分析不再只關注「能否找到缺陷」，還必須檢查模型是否使用正確的影像證據、能否適應新產品與資料漂移、是否會產生大量誤報，以及錯誤判斷會不會在後續診斷流程中持續放大。如圖 1 所示，技術目標正從規則式判定，逐步走向多模態理解、主動取證與可信部署。本文其餘章節安排如下：第二章界定異常分析的任務層級與實務挑戰；第三、四章依時間序回顧從人工規則到基礎模型的技術演進；第五章討論如何從偵測走向理解與主動取證；第六章聚焦可信部署的評估方法；第七章整理跨產業應用；第八、九章展望未來並作結。

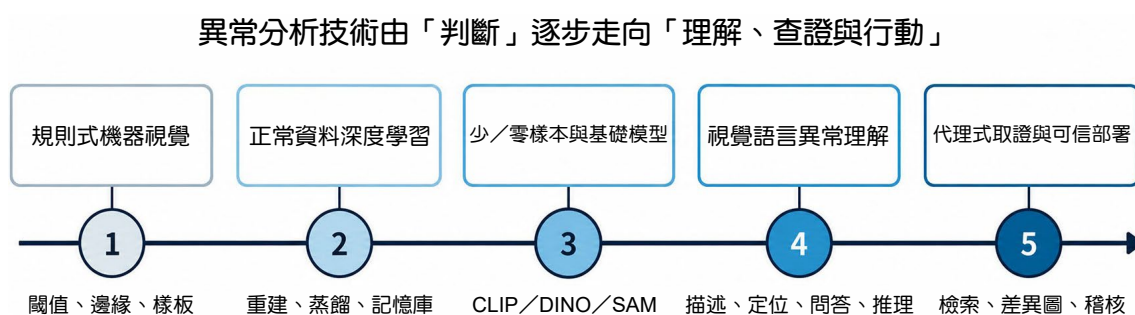


圖 1. 異常偵測與理解技術的演進。本文示意圖均由作者繪製。

二、異常偵測的任務層級與實務挑戰

2.1 從判斷、定位到異常理解

工業異常分析可分為四個層次。第一是影像層級判斷，輸出產品為正常或異常的判定與分數；第二是異常定位，以熱度圖或邊界框指出可疑區域；第三是像素層級分割，精確標示缺陷的形狀與範圍；第四是異常理解，進一步辨識異常類型、外觀特徵、與正常參考的差異及可能影響。這四項能力並不會自然同時成立：模型可能正確判定整張影像為異常，卻定位到錯誤區域；也可能產生看似合理的文字說明，卻未真正使用影像證據。

2.2 結構異常與邏輯異常

結構異常通常可從局部外觀發現，例如裂縫、孔洞、污染、微小刮痕或病灶；邏輯或語意異常則涉及元件的數量、位置與關係，例如漏件、錯裝、組裝順序不合理，或「每個元件單獨看來正常，但整體關係錯誤」。前者需要細緻的紋理與幾何特徵，後者則需理解產品整體結構及元件規則。CSAD 利用基礎模型建立元件區域，再分析元件分布、局部外觀與整體關係，說明異常偵測已從「像素是否不同」逐步走向「場景是否合理」⁽³⁾。圖 2 彙整了上述任務層級，以及結構異常與邏輯異常的差異。

2.3 真實產線使問題更為複雜

研究資料集多半在固定相機、光線與背景下建立，真實產線則可能面臨批次差異、設備漂移、材料反光、姿態變化、產品改版與新型缺陷。異常區域通常面積小、對比低，正常加工痕跡與缺陷外觀也可能十分相似。另一方面，高產量產線每天處理大量產品，即使誤報率看似不高，也可能造成可觀的複檢負擔。以每日檢測十萬件的電子組裝線為例，誤報率僅百分之一即產生一千件複檢案例；每件耗時三十秒，相當於每日額外投入八小時以上人力。反之，若每千件產品中有一件缺陷品未被攔截，每日即有上百件缺陷品流入客戶端。因此，模型不僅要有良好的平均表現，也必須在指定誤報率、固定精確率、實際生產節拍及長期資料漂移下，維持穩定的檢出能力⁽²⁾。

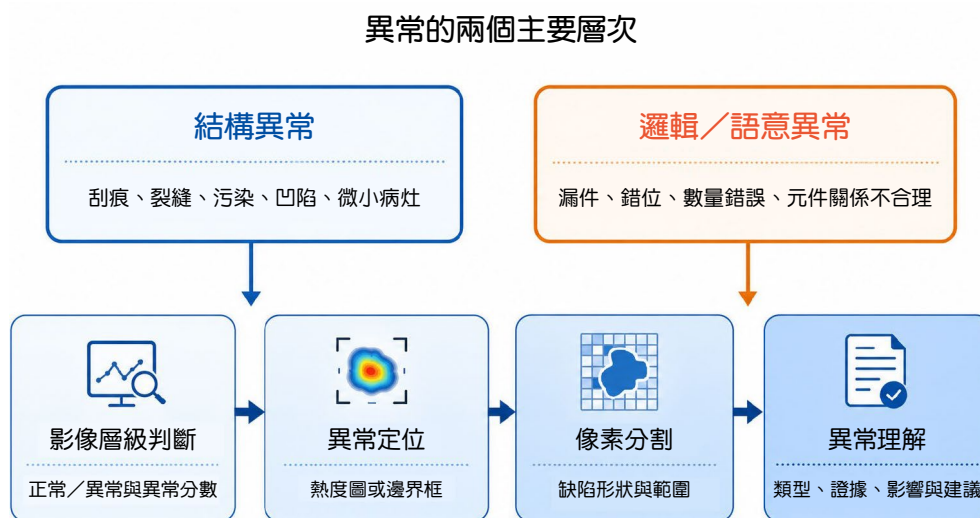


圖 2. 異常分析的任務層級，以及結構異常與邏輯異常的差異。

三、從人工規則走向正常資料深度學習

3.1 規則式機器視覺

早期檢測系統多採用灰階閾值、顏色範圍、邊緣、幾何尺寸、紋理統計與樣板比對。這類方法運算快速、容易解釋，在產品位置、光線與缺陷型態固定的任務中仍相當有效；但一旦產品、材料或環境改變，工程師就必須重新設定規則，而且也難以處理不規則、低對比或未知缺陷。

3.2 正常資料學習：重建、蒸餾與記憶庫

深度學習方法多以正常影像為主要學習資料。重建式方法先讓模型學習正常產品應有的外觀，再以輸入影像與重建結果之間的差異找出可疑區域；知識蒸餾方法則比較預訓練教師模型與只看過正常資料的學生模型之間的特徵差異；記憶庫方法會保存正常影像的局部特徵，測試時再尋找最相近的正常參考。MVTec AD 是廣泛使用的工業異常基準資料集，提供正常影像、缺陷影像與缺陷區域標註，常用於比較不同方法。PatchCore 的概念像建立一座「正常特徵資料庫」：先把正常影像切成許多局部區塊並保存其特徵，測試時若某個區塊找不到相近的正常樣本，就提高它的異常分數。DRAEM 則先在正常影像上人工製造缺陷，再訓練模型學習恢復正常外觀並同時標出異常位置⁽⁴⁻⁶⁾。這三類方法各有適用場域：重建式方法直觀，且可輸出模型所推定的正常外觀，便於檢視模型行為；蒸餾式方法毋須重建整張影像，推論速度通常較快；記憶庫方法在正常樣本充足時定位精準，惟記憶庫的容量與檢索時間會隨產品種類增加而成長。

3.3 模型容量並非越大越好

重建模型存在一項基本矛盾：解碼器過弱時，正常結構無法正確恢復，容易產生誤報；解碼器過強時，又可能連缺陷一起重建，使原本應該突出的異常差異反而消失。對晶圓或週期性紋理而言，模型既要記住正常結構的位置關係，又不能把異常細節原封不動地複製回去。作者團隊提出的 SQUAD，將連續影像特徵轉成有限的離散表示，可將其理解為限制模型僅能以一組貼近正常樣態的「基本構件」重建影像，藉此降低直接複製缺陷的機會⁽⁷⁾。圖 3 比較了兩種重建方式的實際差異。更一般而言，可透過限制模型容量、在中間特徵加入適度擾動，以及加強困難區域的學習，在「保留正常對應」與「避免複製缺陷」之間取得平衡。這也說明，異常偵測模型並不是越大越好。

3.4 合成異常補足資料不足

真實缺陷不足時，可在正常影像上合成刮痕、污染、孔洞、紋理替換或擴散生成缺陷，並同步取得像素層級遮罩。作者團隊的多合成動態加權方法整合不同生成方式，使模型能學習更多樣的異常型態⁽⁸⁾；LFQUIAD 進一步結合合成擴散修補與量化表示，處理少樣本情境⁽⁹⁾。然而，合成影像看來逼真，不表示一定有助於真實檢測；模型可能只學到生成工具留下的痕跡。因此，合成資料應同時評估外觀合理性、缺陷位置、真實測試效益，以及不同缺陷類別是否出現不對稱的誤報或漏檢代價。若模型對合成缺陷的分數遠高於真實缺陷，通常顯示其學到的是生成工具的痕跡，而非缺陷本身的特徵。

為使表 1 所示的演進幅度更加具體，可以最常用的 MVTec AD 基準之影像層級 AUROC 為例：針對單一產品類別訓練的記憶庫方法 PatchCore 約 99.1%⁽⁵⁾，重建判別式方法 DRAEM 約 98.0%⁽⁶⁾；改以單一模型同時處理十五類產品的統一模型 UniAD 約 96.5%，而完全不使用目標產品訓練資料的零樣本方法 WinCLIP 約 91.8%（兩者分別見 4.3 節與 4.1 節）。這組數字顯示，在單一類別且正常資料充足的設定下，公開基準表現已接近飽和；統一與零樣本方法則以數個百分點的差距，換取大幅降低的資料蒐集與模型維護成本。惟須強調，上述皆為公開基準上的排序指標，並不同於特定產線的誤報與漏檢表現，後者仍應依第六章的部署導向指標另行評估。

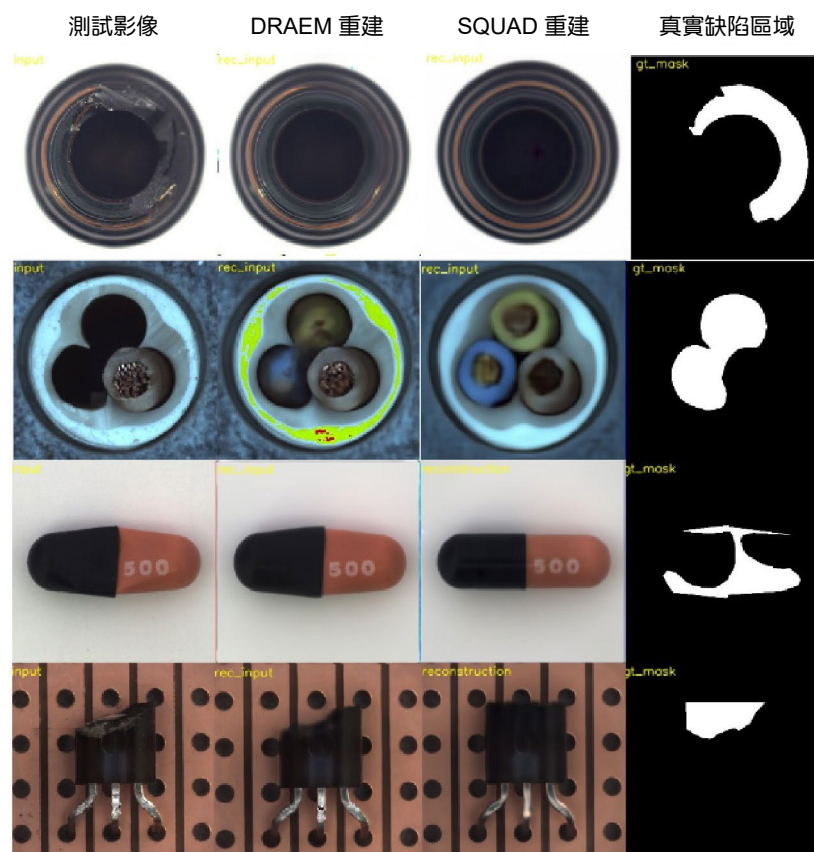


圖 3. 重建式異常偵測的實際案例。由左至右為測試影像、DRAEM 重建結果、SQUAD 重建結果與真實缺陷區域。當模型把異常一併重建時，缺陷差異可能被削弱；SQUAD 則嘗試恢復較接近正常的外觀，使異常更容易被發現。影像取自作者團隊 SQUAD 研究⁽⁷⁾，重新排版。

表 1. 視覺異常偵測技術演進與實務特性比較。

方法階段	主要資料	代表概念	優點	主要限制
規則式	少量樣本+人工規則	閾值、邊緣、樣板	快速、易解釋	產品或環境改變後需重設
正常資料學習	正常影像	重建、蒸餾、記憶庫	無須大量缺陷標註	容量與對應的權衡、逐類維護成本
合成異常	正常影像+人工缺陷	剪貼、紋理、擴散生成	可取得像素遮罩	合成痕跡偏差、類別風險不均
少／零樣本	少量或無目標資料	CLIP 語意、DINO 結構、正常參考	縮短新產品導入時間	微小缺陷感受力與跨域穩定性
統一與跨域	多類別／來源域資料	共享模型、凍結骨幹	降低模型維護成本	未知產線適應與來源域過度擬合
異常理解	影像+文字+知識	視覺語言模型 (VLM)、問答、結構化推理	可描述、定位與提出假設	幻覺、模板捷徑與證據不一致
主動取證代理	影像+參考+工具	放大、檢索、差異圖	模擬工程師查證流程	工具路徑與證據使用未必一致
可信部署	測試資料+風險成本	FA/FM、因果審計、穩定性	依操作風險選擇模型	校準、漂移與跨場域驗證

四、基礎模型驅動的下一波技術演進

大型預訓練模型改變了異常偵測長期依賴單一產品資料的方式。CLIP (Contrastive Language-Image Pre-training, 對比式語言－影像預訓練) 同時從大量影像與文字中學習，因此能把「正常」、「破損」或「缺少元件」等文字概念與影像內容對應起來；DINO 則是一類透過自監督學習 (self-supervised learning) 從大量影像中學習的視覺模型，較擅長保留形狀、邊界與局部紋理等結構資訊。簡言之，CLIP 著重於回答「這是什麼、語意是否相符」，DINO 則長於辨識「何處的結構與周圍相異」。分割基礎模型則可進一步協助切出物件與元件區域^(10, 11)。近期綜述將可供泛化的先驗整理為五類：視覺先驗、語言與推理先驗、幾何與多模態先驗、通用任務先驗，以及生成異常先驗⁽²⁾，如圖 4 所示。這些先驗可互補使用，但仍需在目標場域驗證是否真的能降低誤報，並保留對微小缺陷的感受能力。

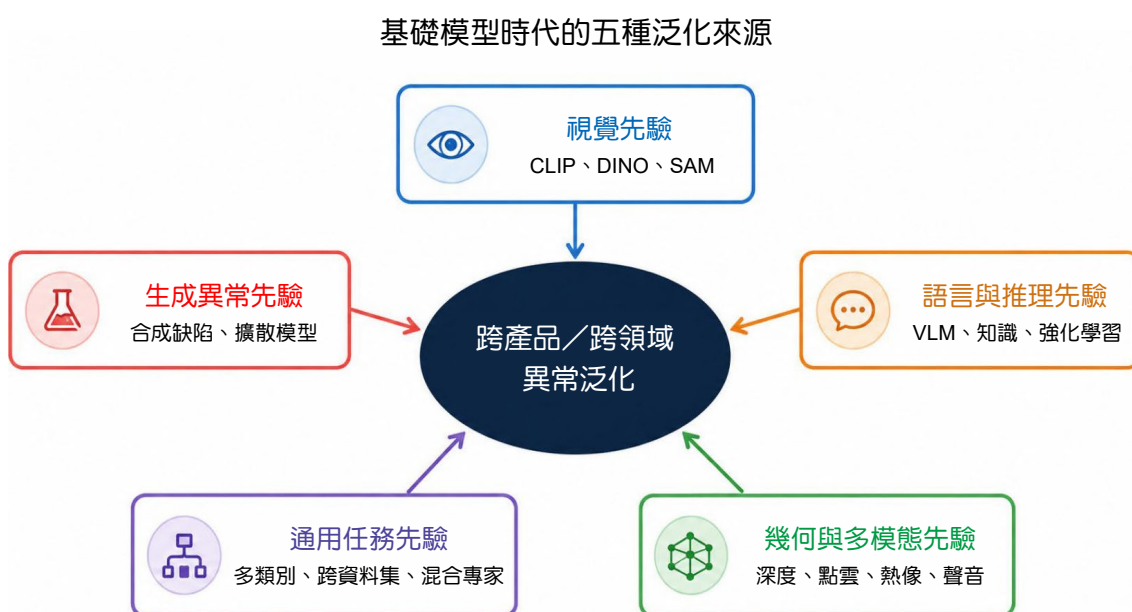


圖 4. 基礎模型時代的五種泛化來源。

4.1 少樣本與零樣本異常偵測

少樣本 (few-shot) 方法利用一張或少量正常參考影像，建立產品的正常表示；零樣本 (zero-shot) 方法則在完全沒有目標產品訓練資料時，直接借助預訓練知識進行判斷。WinCLIP 是把 CLIP 用於工業異常檢測的代表性方法：它利用「正常」與「異常」文字提示，再配合影像的局部區域特徵，判斷哪裡可能有缺陷⁽¹²⁾。AnomalyCLIP 進一步讓異常提示不必綁定特定產品類別，使同一套語意概念更容易跨產品使用⁽¹³⁾。作者團隊的 CLIP-FSQA 則將 CLIP 提供的語意資訊與離散量化表示結合，讓少量正常樣本也能形成較穩定的正常參考⁽¹⁴⁾。對產線而言，這類方法的實質意義在於縮短新產品的導入時間：過去每項新產品往往須蒐集數百張影像並重新訓練，如今僅需少量正常樣本或一段文字規格，即可先建立可用的初版模型，再隨資料累積逐步強化。

4.2 語意與結構必須互補

CLIP 擅長理解「正常」、「破損」或「缺少元件」等整體語意概念，但對很小的刮痕、低對比病灶或細微幾何變形，未必足夠敏感；DINO 類模型則較能保留局部形狀與紋理連續性。圖 5 示意兩者如何互補：由 CLIP 提供「正常／異常」的語意判斷，再由 DINO 補上局部結構證據，最後分別產生像素層級的異常圖與整張影像的異常分數。實作上，結構資訊只以小幅度加入，避免破壞 CLIP 原本的語意能力；同時把「找位置」與「判斷整張影像是否異常」分開處理，讓模型既能看清小缺陷，也不會因局部熱度圖的調整而意外改變整體判斷⁽¹⁵⁾。

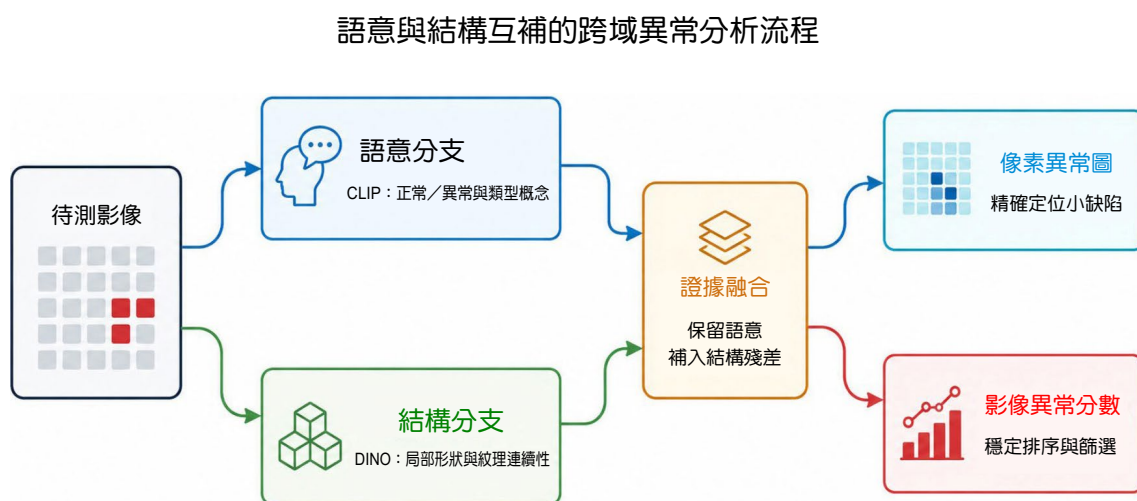


圖 5. 以語意與結構互補進行異常定位與影像排序的概念流程。

4.3 從單一產品走向統一與跨域模型

傳統作法通常為每項產品建立一個模型；當產品類別增加後，訓練、校正與版本維護成本也會快速上升。統一異常偵測的目標，是讓「一個模型」處理多種產品。UniAD 是此方向的代表架構之一，讓不同產品共用同一套模型⁽¹⁶⁾；作者團隊的 UniLEAD 則進一步採用較輕量的共享設計，讓不同產品共用大部分能力，只保留少量必要的專門處理，降低模型數量與維護成本⁽¹⁷⁾。當廠內產品達數十項時，統一模型可將數十套模型與校正紀錄收斂為一套共享能力加上少量產品專屬設定，版本管理與異常追蹤因而簡化。更進一步的跨域偵測，要求模型在來源產品上訓練後，能直接轉移至新產品、不同成像條件，甚至跨到醫療影像。透過凍結大型基礎模型、只訓練少量介面，以及分開處理語意與結構資訊，可降低模型過度記住來源資料的問題；作者團隊的 UFMADS 與 BSMAD 也顯示，這類輕量融合方式可延伸到跨器官或跨成像模態的異常偵測與分割^(15, 18)。

圖 6 以真實影像說明跨資料集的遷移結果：上半部為模型於 VisA 訓練後直接測試於 MVTEC AD，涵蓋瓶子、纜線、膠囊、螺絲與紋理表面等工業物件；下半部則為反向設定。雖然兩個資料集的產品外觀與成像條件差異很大，模型在未曾見過的目標資料集上，仍能將較高的異常反應集中在真正有問題的區域，而非背景或正常結構⁽¹⁹⁾。

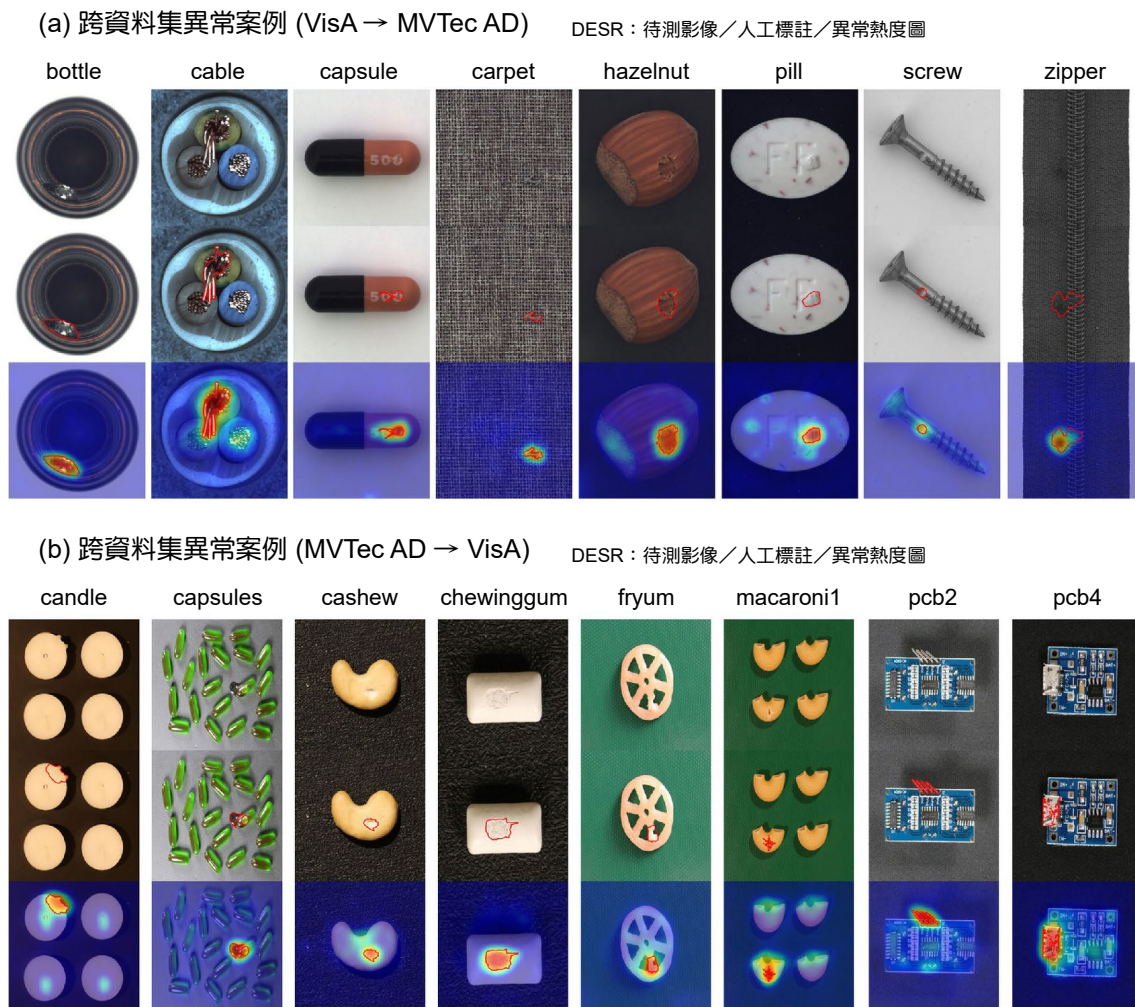


圖 6. 跨資料集異常偵測的真實案例。(a) 模型於 VisA 訓練、跨域測試於 MVTec AD 的結果；(b) 模型於 MVTec AD 訓練、跨域測試於 VisA 的結果。各欄由上而下為待測影像、人工標註的真實異常區域與模型異常熱度圖。影像取自作者團隊 DESR 研究⁽¹⁹⁾，重新排版。

4.4 幾何、多感測與生成先驗

單靠彩色影像，可能無法發現高度、內部結構或溫度異常。未來系統需整合深度、點雲、X 光、紅外線、聲音、振動與製程數值；由於不同模態在尺度、對位、缺失資料與取樣頻率上存在差異，單純串接特徵通常並不足夠。另一方面，語言控制的生成模型可依產品結構指定缺陷位置與型態，補足缺少的異常資料。真正有效的多模態與生成先驗，應在目標場域證明能降低誤報、提升小缺陷定位，並維持跨批次與跨設備的穩定性。

五、從異常偵測到異常理解

5.1 視覺語言模型讓檢測系統具備對話能力

傳統模型通常只輸出異常分數與熱度圖，缺陷類型及可能原因仍需由工程師自行判讀。視覺語言模型 (vision-language model, VLM) 則能同時理解影像與文字，因此使用者可以直接

用文字詢問「哪裡有異常？」或「這個缺陷可能是什麼？」。AnomalyGPT 將大型 VLM 導入工業檢測，使模型可回答異常是否存在、位於何處，並進行多輪對話⁽²⁰⁾。MMAD 則不是一個檢測模型，而是一套用來測試多模態大型模型是否真的理解工業異常的評估基準，涵蓋產品知識、異常辨識、定位與描述等七項能力⁽²¹⁾。作者團隊的知識－語言對齊研究，則進一步調整正常與異常的文字錨點，使語言概念與工業缺陷的視覺表徵對應得更穩定⁽¹⁹⁾。

5.2 異常理解必須連結可驗證證據

如同臨床診斷須指出病灶位置並說明診斷依據，理想系統應能回答五個問題：是否異常、異常是什麼、異常在哪裡、哪些影像證據支持判斷，以及應採取何種行動。以半導體為例，晶圓圖可設計成涵蓋缺陷存在、類型、位置、外觀與可能製程假設的多階段問答；在混合缺陷情境下，模型還必須輸出缺陷集合，並為每一類缺陷分別提供由影像計算而得的位置。模型提出的製程原因只能視為待驗證假設，而非已確認的物理根因。無論文字回答多流暢，或注意力是否落在正確位置，都不能單獨證明該位置真正決定了答案⁽²²⁾。

5.3 強化學習與結構化推理

近年方法開始採用群組相對策略最佳化 (group relative policy optimization, GRPO) 等強化學習技術。簡言之，GRPO 會使模型對同一問題產生多個候選回答，並以同一組候選的平均表現作為比較基準：表現優於群組平均的回答獲得強化，劣於平均者則被抑制。由於比較基準直接取自群組內部，不需再額外訓練一個價值評估模型，訓練流程相對簡單穩定。這種作法特別適合異常推理任務，因為此類任務的獎勵大多可由規則直接檢核——例如答案是否正確、框選位置是否與標註區域重疊、輸出格式是否完整——模型因此能在可驗證的回饋下，逐步保留較佳的推理與定位方式。AnomalyR1 代表從傳統監督式問答走向「同時推理與定位」的發展趨勢⁽²³⁾。不過，如果獎勵只看最後答案是否正確，模型仍可能靠類別比例、選項位置或固定模板猜答案，而沒有真正看懂影像。作者團隊近期提出的 EGVLR 因此把「看到了什麼證據」、「如何判斷」、「異常位於哪裡」與「最後答案」分開表示，再分別獎勵答案正確性、位置合理性與說明一致性；對正常影像則要求不要隨意框出缺陷。這種分階段訓練在 MMAD 上改善了需要視覺證據的推理與定位，也降低了無缺陷影像上的錯誤框選⁽²⁴⁾。圖 7 即為此類任務的實例：模型除須判定缺陷類型，亦須標示支持該判斷的影像位置，二者缺一皆不足以稱為真正的理解。

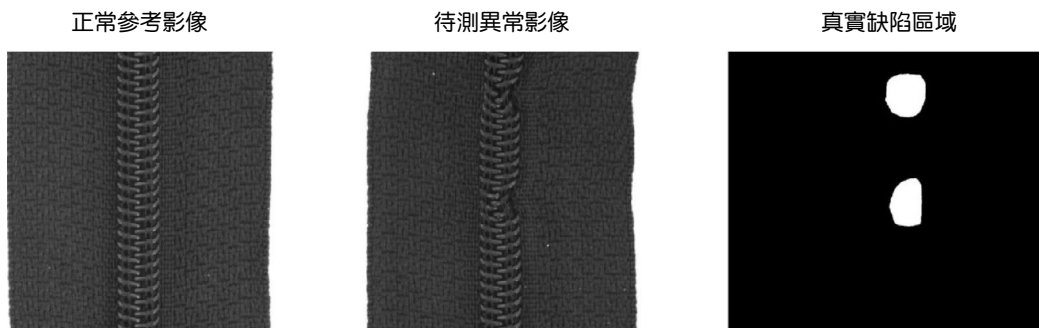


圖 7. 異常理解的實際案例。左為正常拉鍊，中為待測影像，可看到部分拉鍊齒被擠得更靠近；右為人工標註的真實缺陷區域。這類任務要求模型不只答對缺陷類型，也要指出支持判斷的位置。影像取自作者團隊 EGVLR 研究⁽²⁴⁾，重新排版。

5.4 代理式異常分析與主動取證

工程師面對難以判定的影像時，通常會放大可疑區域、尋找同型號的正常樣本、計算差異，或要求補充另一個視角。代理式人工智慧 (agentic AI) 希望讓模型也能依需要主動使用這些工具，而不是只看一次影像就立即下結論。ReAct是一種讓模型在「推理」與「採取行動」之間交替的通用框架⁽²⁵⁾；AnomalyClaw 則把這個概念用於異常檢測，讓模型透過多輪反駁、檢索與工具查證候選異常⁽²⁶⁾。然而，模型有呼叫工具，不代表它真的使用了工具提供的證據。多輪訓練仍需確認工具影像有進入模型的生成與評分流程，並透過遮蔽或打亂影像等測試，檢查答案是否真的會隨證據改變。

六、可信部署不能只看整體準確率

工業部署真正關心的是操作風險，而非單一排行榜分數。圖 8 整理四項應分開稽核的能力：任務準確度、視覺定位、因果視覺依賴與部署安全。同一模型可能在其中一項進步，另一項卻退步；高準確率、清楚的熱度圖、流暢的文字與完整的工具操作紀錄，都不能互相替代。可信評估還應納入跨階段一致性、端到端診斷正確性，以及不同訓練執行之間的穩定性。

可信賴異常理解：答案、證據、路徑與風險需分別稽核



圖 8. 可信賴的異常理解，需分別稽核任務準確度、視覺定位、因果依賴與部署安全。

6.1 誤報與漏檢具有不同成本

誤報 (false alarm, FA) 會讓正常產品進入複檢流程，降低產能或造成不必要的報廢；致命漏檢 (fatal miss, FM) 則可能使缺陷產品流出，影響客戶與安全。在高度不平衡的資料中，只要把所有產品都判為多數類別，仍可能獲得很高的整體準確率。例如在缺陷率為百分之一的產線上，一個一律判定為正常的模型即可達到百分之九十九的準確率，卻無法攔截任何缺陷品；若第一階段篩選錯誤，還可能觸發後續看似連貫、實則缺乏影像依據的類型、位置與製程說明。因此實際評估至少應報告 FA、FM、平衡準確率與固定誤報率下的召回率、信賴區間與 FA-FM 風險曲線。

這裡的「操作點 (operating point)」，是指模型實際上線時，因判定門檻設定而落在的那一組誤報率與漏檢率數值。同一個模型只要調整門檻，誤報率與漏檢率就會此消彼長，形成一條 FA-FM 曲線；門檻一旦選定，模型即固定在曲線上的某一點運作，這一點便是操作點。因此門檻不宜一律沿用 0.5，而應依產線的風險成本決定。

以一個簡化的例子說明。假設某產線導入初版模型後，量測得誤報率 12%、漏檢率 0.5%，複檢站因誤報過多而負荷過重；工程師加入抑制誤報的訓練介入後，誤報率降至 0.5%，漏檢率卻升至 3%。這並非單純的改善，而是操作點沿著曲線移動：誤報減少的同時，漏檢增加了六倍。若一件缺陷品流出的代價遠高於一次複檢，後者反而是更危險的設定。合理的作法，是先依誤報與漏檢各自的單位成本，在 FA-FM 圖上畫出可接受的範圍 (圖 9)，再檢查候選模型 (含以不同隨機種子重新訓練的結果) 是否都穩定落在此範圍內，而非只比較整體準確率。

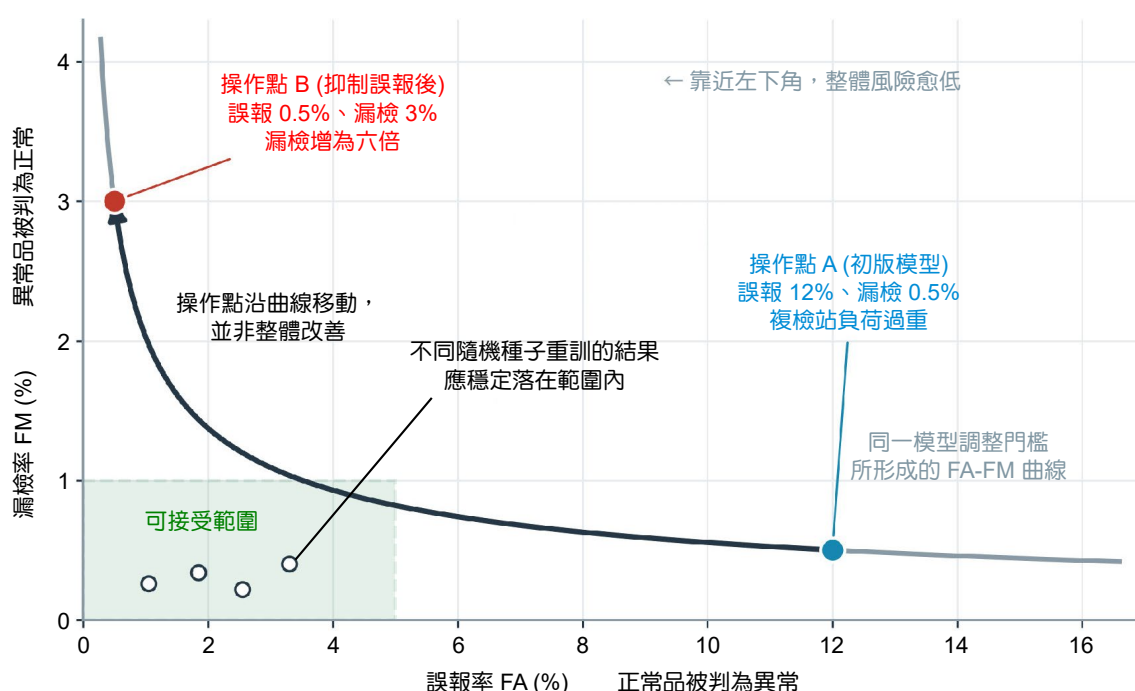


圖 9. FA-FM 風險平面與操作點示意 (數值為簡化範例，非實測結果)。同一模型調整判定門檻，誤報率與漏檢率即沿曲線此消彼長；由操作點 A 移動至 B 時，誤報大幅下降，漏檢卻增為六倍。可接受範圍依產線的誤報與漏檢單位成本劃定，候選模型 (含不同隨機種子的重訓結果) 應穩定落在範圍內。

6.2 影像層級與像素層級需分開評估

局部異常圖表現良好，不代表整張影像的判斷一定同樣可靠；反之，模型也可能正確判定產品為異常，但熱度圖卻集中在背景。常見的 AUROC (area under the ROC curve) 可理解為模型在不同判定門檻下區分正常與異常的整體能力；平均精確率則更能反映異常樣本較少時的排序品質，即依異常分數由高到低排列後，真正的異常樣本是否集中在前面。影像層級指標關心「整件產品判對了嗎」，像素層級與區域重疊指標則關心「缺陷位置找對了嗎」。因此，即使報告綜合分數，也應保留各項組成指標，避免平均值掩蓋定位或排序失敗。

6.3 因果視覺依賴與資料捷徑

本節要處理的問題，可類比為判斷學生是真正讀懂題目，還是只憑作答技巧猜中答案。單看分數無法分辨這兩種情形，必須改動題目本身，再看答案是否跟著改變。因果審計即依此邏輯進行：固定問題與正確答案，將正確答案直接提供給模型，讀取模型賦予該答案的機率作為基準，接著在其餘輸入完全不變的條件下更換影像，比較機率的變化。

常用的破壞方式有兩種。「遮蔽」是把影像換成空白，用以檢驗模型是否需要影像存在；「打亂」則是把像素在產品輪廓內重新排列，缺陷總量與外形維持不變、僅破壞空間排列，用以檢驗模型是否依賴缺陷的位置與形狀。若機率明顯下降，表示答案確實建立在影像證據之上；若幾乎不變，模型很可能不是看圖作答，而是依循文字端既有的統計規律——例如題目的問法通常對應哪一個答案，或固定的答題格式本身即足以推得答案，此即所謂的語言先驗 (language prior)。

要使審計具說服力，前後兩次測試應使用相同樣本與順序以進行配對比較，遮蔽與打亂的結果也須分開報告，並加入「不提供影像、僅給文字題目」等對照，界定流暢文字本身能證明的範圍。以晶圓圖問答為例，某模型在缺陷類型題上準確率極高，影像打亂後信心卻幾乎不變；加入反事實訓練資料——把打亂後的影像配上其實際對應的正確答案——重新訓練後，真實影像與打亂影像的信心差距才轉為顯著。惟審計至此仍未結束：視覺依賴改善不等於部署更安全，任何資料或損失函數的調整，都應回頭重新檢查各缺陷類別的 FA、FM 與多種子穩定性。

6.4 運算、資安與可追溯性

大型 VLM 的運算成本較高，高速產線可採分層式架構：先由小型模型進行篩選，再將困難案例交由大型模型比較與說明。例如產線每三秒產出一件產品時，小型模型可在數十毫秒內完成初判，僅將少數可疑案例送交大型模型，即可兼顧節拍與理解能力。製造影像涉及產品設計與製程機密，應優先考量廠內推論、權限控管、資料留存與模型版本管理。每次推論宜保存輸入影像、正常參考、提示與模型版本、門檻、熱度圖、工具紀錄、信心與人工覆核結果，形成可追溯的資料與決策鏈，以支援稽核、校準與持續改善。

七、智慧製造與跨領域應用

7.1 半導體與電子製造

半導體應用涵蓋晶圓圖、晶圓表面、積體電路顯微影像、印刷電路板與焊點檢測。異常可能是週期圖案中的局部破壞，也可能是多種缺陷同時存在。以晶圓圖為例，同一片晶圓可能同時出現中心圓斑、邊緣環狀與線性刮痕，各自對應不同製程站點；若模型只輸出單一類別，工程師仍須自行拆解混合訊號。未來系統可將缺陷存在、類型、位置、外觀與可能製程假設設計為多階段問答，並結合製程站點、設備參數、批次資訊與歷史缺陷地圖。模型產生的製程說明如 5.2 節所述，仍需搭配設備資料與工程師分析加以確認。

7.2 金屬、機械、組裝與包裝

金屬與機械零件需要檢測刮痕、裂縫、凹陷、加工不完整與尺寸變形；組裝檢測則關注元件的數量、方向與關係。食品、紡織與包裝產業可檢測異物、破損、印刷錯誤與封口不完整。二維影像適合辨識顏色與紋理差異，三維點雲與深度資訊可偵測幾何變形，多視角影像

則可降低遮蔽造成的漏檢。產品快速換線時，少樣本參考、統一模型與可調整的文字規格，有助於縮短導入時間。例如包裝產線每逢節慶改版，印刷圖樣與封膜材質常於短期內全面更換；若模型須重新蒐集數百張影像方能上線，改版初期便會出現檢測空窗。

7.3 醫療、遙測與設備監控

類似技術也可應用於核磁共振 (magnetic resonance imaging, MRI)、電腦斷層 (computed tomography, CT)、內視鏡、皮膚與病理影像，或延伸至衛星影像、熱像與設備聲音。跨領域測試可用來檢查模型是否真正學到通用的語意與結構；然而，不同領域在資料分布、法規與錯誤成本上差異極大。工業資料上的良好結果，不能直接等同於具備臨床或公共安全部署資格；仍需經過領域專家驗證、獨立資料測試、校準與風險管理。

八、未來發展方向

第一，模型將由「每項產品各自訓練」，走向可依正常參考、文字規格與感測條件動態適應的通用異常系統。第二，語意模型與稠密結構模型將更緊密整合，讓「理解異常概念」與「精準定位微小異常」同時成立。第三，影像、點雲、熱像、聲音與製程數值將形成多感測閉迴路，模型可在資訊不足時主動要求額外觀測。第四，異常理解將與智慧代理及具身巡檢結合，依序完成感知、比較、反駁、查詢、決策與行動。第五，可信賴評估將從平均準確率，擴展至因果證據、FA-FM 風險曲線、校準、長期漂移、跨批次與多種子穩定性。

真正的技術進步，不只是在既有資料集上增加少數分數，而在於降低新產品導入成本、減少誤報與漏檢、即使環境改變仍能維持可靠，並讓工程師能檢查模型究竟使用了哪些證據。異常偵測也將從獨立演算法，轉型為串接感測器、資料平台、製程知識與人員決策的品質工程系統。

九、結論

異常偵測技術已從人工規則，逐步發展至正常資料重建、知識蒸餾、特徵記憶庫、合成異常、少樣本與零樣本方法。視覺基礎模型提供可遷移的語意與結構知識，使統一模型、跨域檢測與多模態分析日益可行；視覺語言模型則讓系統能描述異常、回答問題、比較正常參考，並提出診斷假設。

然而，新一代系統的價值，不在於能否說得像專家，而在於其判斷是否建立在正確、可驗證且可追溯的視覺證據上。未來的智慧檢測必須同時處理任務準確度、空間定位、因果視覺依賴、工具證據是否真正被使用、誤報與漏檢、運算效率、資安及長期穩定性。換言之，異常分析的下一階段，不只是讓機器看見缺陷，而是讓機器理解差異、主動查證，並以可稽核且安全的方式協助人類採取行動。

參考文獻

1. Y. Cao, X. Xu, J. Zhang, Y. Cheng, X. Huang, G. Pang, and W. Shen, arXiv:2401.16402 (2024).
2. X. Jiang, B. Hu, and F. Zheng, “A Survey of Recent Advances in Industrial Anomaly Detection: From Normal-Only Training to Foundation-Model Priors”, Preprints, 202606.0719 (2026).
3. Y.-H. Hsieh and S.-H. Lai, “CSAD: Unsupervised Component Segmentation for Logical Anomaly Detection”, *British*

- Machine Vision Conference*, November 25-28, (2024).
4. P. Bergmann, M. Fauser, D. Sattlegger, and C. Steger, "MVTec AD-A Comprehensive Real-World Dataset for Unsupervised Anomaly Detection", *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, June 16-20, (2019).
 5. K. Roth, L. Pemula, J. Zepeda, B. Scholkopf, T. Brox, and P. Gehler, "Towards Total Recall in Industrial Anomaly Detection", *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, June 18-24 (2022).
 6. V. Zavrtanik, M. Kristan, and D. Skočaj, "DRAEM-A Discriminatively Trained Reconstruction Embedding for Surface Anomaly Detection", *IEEE/CVF International Conference on Computer Vision*, October 11-17, (2021).
 7. S.-C. Lin and S.-H. Lai, "SQUAD: Scalar Quantized Representation Learning for Unsupervised Anomaly Detection and Localization", *European Conference on Computer Vision Workshops*, September 29 - October 4, 166 (2024).
 8. S.-C. Lin, H.-W. Lee, Y.-H. Hsieh, C.-Y. Ho, and S.-H. Lai, "Masked Attention ConvNeXt UNet with Multi-Synthesis Dynamic Weighting for Anomaly Detection and Localization", *British Machine Vision Conference*, November 20-24 (2023).
 9. S.-C. Lin and S.-H. Lai, "LFQUIAD: Lookup-Free Quantized Autoencoder for Few-Shot Unsupervised Industrial Anomaly Detection via Synthetic Diffusion Inpainting", *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, June 11-12, (2025).
 10. A. Radford, J. W. Kim, C. Hallacy, et al., "Learning Transferable Visual Models from Natural Language Supervision", *International Conference on Machine Learning*, July 18-24, (2021).
 11. M. Caron, H. Touvron, I. Misra, et al., "Emerging Properties in Self-Supervised Vision Transformers", *IEEE/CVF International Conference on Computer Vision*, October 11-17, (2021).
 12. J. Jeong, Y. Zou, T. Kim, D. Zhang, A. Ravichandran, and O. Dabeer, "WinCLIP: Zero-/Few-Shot Anomaly Classification and Segmentation", *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, June 18-22, (2023).
 13. Q. Zhou, G. Pang, Y. Tian, S. He, and J. Chen, "AnomalyCLIP: Object-Agnostic Prompt Learning for Zero-Shot Anomaly Detection", *International Conference on Learning Representations*, May 7-11, (2024).
 14. S.-C. Lin and S.-H. Lai, "CLIP-FSQAE: CLIP-Guided Finite Scalar Quantized Autoencoder for Few-Shot Anomaly Detection", *IEEE International Conference on Image Processing*, September 14-17, 313 (2025).
 15. S.-C. Lin, J.-N. Hwang, and S.-H. Lai, "BSMAD: Bridging Semantic and Structural Manifolds for Robust Cross-Modality Medical Anomaly Detection", *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, June 3-4, (2026).
 16. Z. You, L. Cui, Y. Shen, K. Yang, X. Lu, Y. Zheng, and X. Le, "A Unified Model for Multi-Class Anomaly Detection", *Advances in Neural Information Processing Systems*, November 28 – December 9, (2022).
 17. S.-C. Lin and S.-H. Lai, "UniLEAD: A Unified and Lightweight Model for Anomaly Detection", *IEEE/CVF International Conference on Computer Vision Workshops*, October 19-20, 1483 (2025).
 18. S.-C. Lin, J.-N. Hwang, and S.-H. Lai, "UFMADS: A Unified Framework for Prompt-Guided Medical Anomaly Detection and Segmentation", *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, June 3-4, 1291 (2026).
 19. S.-C. Lin, "DESR: Dual-Encoder Semantic Retrieval for Open-Set Anomaly Discovery", *European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD)* (2026).
 20. Z. Gu, B. Zhu, G. Zhu, Y. Chen, M. Tang, and J. Wang, *Proceedings of the AAAI Conference on Artificial Intelligence*, **38** (3), 1932 (2024).
 21. X. Jiang, J. Li, H. Deng, et al., "MMAD: A Comprehensive Benchmark for Multimodal Large Language Models in Industrial Anomaly Detection", *International Conference on Learning Representations*, April 24-28 (2025).
 22. S. Jain and B. C. Wallace, "Attention is not Explanation", *Conference of the North American Chapter of the Association for Computational Linguistics*, June 2-7 (2019).
 23. Y. Chao, J. Liu, J. Tang, and G. Wu, "AnomalyR1: A GRPO-Based End-to-End MLLM for Industrial Anomaly Detection", arXiv:2504.11914 (2025).
 24. S.-C. Lin, Y.-H. Lu, D. Y. Ye, and S.-H. Lai, "EGVLR: Evidence-Grounded Vision-Language Reinforcement for Anomaly Reasoning", *European Conference on Computer Vision*, (2026).
 25. S. Yao, J. Zhao, D. Yu, et al., "ReAct: Synergizing Reasoning and Acting in Language Models", *International Conference on Learning Representations*, May 1-5, (2023).
 26. X. Jiang, Y. Zhao, Z. Yang, and F. Zheng, "AnomalyClaw: A Universal Visual Anomaly Detection Agent via Tool-Grounded Refutation", arXiv:2605.10397, (2026).

作者簡介

林士智先生為國立清華大學跨院國際博士班學位學程 博士候選人。

Shih-Chih Lin is currently a Ph.D. candidate in the International Intercollegiate Ph.D. Program at National Tsing Hua University.

賴尚宏先生為美國佛羅里達大學電機與電腦工程博士，現職為國立清華大學資訊工程學系教授兼電機資訊學院副院長。

Shang-Hong Lai received his Ph.D. in Electrical and Computer Engineering from the University of Florida. He is currently a professor in the Department of Computer Science and Associate Dean of the College of Electrical Engineering and Computer Science at National Tsing Hua University.