用於 TFT-LCD Array 製造中跨製程缺陷分類的新型多模態學習方法

在薄膜電晶體液晶顯示器 (TFT-LCD) 的製造過程中，多層陣列製程帶來的自動缺陷分類 (ADC) 面臨諸多挑戰。尤其是在分層陣列過程中存在的複雜識別模式，使得傳統的深度學習分類訓練策略難以達到理想的跨流程識別效果。面對這些問題，本文提出了一種新型的多模態學習方法，該方法不僅基於高效的知識工程技術，還引入了跨模態對比學習策略。透過此方法，除了傳統的視覺模式識別外，還能學習到細粒度的描述資訊，從而大幅提升識別性能。實驗結果顯示，在多種模型架構中，本研究所提出的訓練策略均顯著超越了傳統方法，達到了 0.92 % 至 7.89 % 的準確率增長。值得一提的是，此方法已獲得台灣一家 TFT-LCD 製造領導廠商的認可與驗證。本研究不僅在跨流程和多產品的缺陷分類領域取得了顯著進展，更為製造業的複雜識別任務指明了全新的研究方向。

劉奕、陳鴻文

關鍵詞：智慧製造、AOI、TFT-LCD、多模態機器學習、人工智慧

隨著薄膜電晶體液晶顯示器 (TFT-LCD) 製造技術的發展，自動缺陷分類 (Automated Defect Classification, ADC) 已成為確保產品高品質的核心領域半導體光學檢測技術。近年來，深度學習技術已被證明具有顛覆性的潛力，對此領域帶來了深遠的影響 (Chien et al., 2022; Lu & Su, 2021)。然而，TFT-LCD 的陣列 (Array) 製程，特別是逐層堆疊的特性和各製程階段如薄膜沉積、光阻塗覆、曝光顯影以及蝕刻等取像特徵差異，為深度學習技術帶來了巨大的挑戰。

因每層各製程階段的影像資料均有所不同，過去的作法為針對每個製程階段單獨開發專門的深度學習模型，而這種方法有兩種明顯的缺點：第一是由於需要為每個製程單獨開發模型，因此會有大量的模型需要維護，導致了 AI 應用的成本大幅增加。第二是這種多模型的策略會使每個模型的訓練樣本數量減少，從而可能導致模型的準確度不如預期。基於上述缺點，一個能夠跨製程及站點進行缺陷辨識的通用模型的需求顯得尤為迫切。但陣列製程的逐層堆疊特性，使得純粹依賴圖像資訊的 AI 模型在識別上仍然面臨困難。反觀，經過良好訓練的工程師能夠結合細粒度描述和缺陷圖像來達到更精確的辨識。

為了解決這些挑戰，本研究提出了一種基於知識工程和跨模態對比學習策略的多模態機器學習方法，旨在整合圖像資料以及來自場域專家的描述性訊息，以期望提高缺陷辨識的準確度。實驗結果表明，我們的策略成功地結合了影像缺陷與其技術描述，實現了高效的跨流程缺陷分類，並在不同的主流機器視覺模型架構下均取得辨識效果的提

升。本研究的主要貢獻如下：

一、為 TFT-LCD 陣列製程中的跨製程缺陷識別提出了一種新型的多模態學習策略，據我們所知此乃該領域首次之研究，此研究已投稿至 (Liu et al., 2023)。

二、我們的方法已得到了台灣一家 TFT-LCD 領導製造商的認證，並在實際應用資料中展示了其跨流程和多產品缺陷檢測的顯著優勢，為解決製造中的複雜識別問題提供了新方向。

文獻回顧

一、自動光學檢測與自動瑕疵分類

自動光學檢測 (Automated Optical Inspection; AOI) 在電子製造中扮演著極其重要的角色，作為一種關鍵的無損檢測技術 (Ebayyeh & Mousavi, 2020)。它的主要功能是及時攔截潛在缺陷，從而確保產品品質；其中，AOI 在製造業中的核心技術在於自動缺陷分類 (Automated Defect Classification; ADC) 演算法 (Chien et al., 2022)。這些精心設計的演算法具有缺陷檢測和分類、以及精確定位需要修復的產品的雙重功能。透過 ADC 演算法的自動化，對人工檢查的依賴大大減少，從而也降低了材料浪費的風險。

隨著深度學習技術的發展，監督式學習搭配卷積神經網路 (Convolutional Neural Network; CNN) 已成為 AOI 領域的重要研究方向 (He et al., 2015; Krizhevsky et al., 2012)。這些網路與基於規則的系統和傳統機器學習策略並列，均展現出優異的性能 (Chang et al., 2022; Chien et al., 2022)。 然而，盡管此方法獲得了業界的廣泛認可，但它仍然存在一些固有的限制，其中之一就是在模型的訓練階段必須預先確定缺陷的類型，這樣不可避免地固定了缺陷類別的數量，使得在後續應用中，加入新的缺陷類型成為了一大挑戰。此外，現有的 CNN 方法在分類缺陷時，往往將各缺陷類別視為獨立的實體，忽略了缺陷名稱中蘊含的豐富描述性資訊。這種做法不僅增加了分類的難度，更忽略了從特定領域知識中所能體現的不同缺陷之間的關聯。

基於上述的限制與挑戰，迫切需要開發新的策略，以實現更靈活、更具適應性和更全面的缺陷分類。我們的研究目的為在訓練過程中整合領域知識，並將缺陷視為相互關聯的實體，從而提供一個更加靈活、細緻且能適應不斷變化製造環境的缺陷識別方法。

二、多模態機器學習

近年來，多模態機器學習方法在眾多計算任務上顯示出潛力，包括影像分類、物件偵測和分割等 (Gu et al., 2022; Kim et al., 2021; Li et al., 2022; Radford et al., 2021; Xu et al., 2022)。這些方法透過融合多種資訊模式如結合視覺數據和文字描述來達到優化結果和更深層的資料理解。

其中，CLIP(Contrastive Language-Image Pre-training) 框架是本領域的一大突破 (Radford et al., 2021)。CLIP 能夠融合視覺和語言數據，這不僅打破了傳統有監督分類模型的局限，也展現在零樣本遷移學習中的卓越性能。事實上，它在視覺語言檢索任務中已被證實具有高效能，這成為了 AI 領域的一大進步。而除了理論上的成果，CLIP 也在多領域展現其實際應用潛力，例如在醫學領域的應用 (Eslami et al., 2021)。

這種文字、圖像的跨模態對比學習方法的後續研究中試圖透過更深入的文字描述優化視覺辨識結果。研究者利用其他的同義詞和類別定義資料庫，例如 WordNet 和 Wiktionary，來豐富文字輸入 (Shen et al., 2022)。而從 GPT-3(Brown et al., 2020) 中提取描述性屬性 (Pratt et al., 2023)，甚至利用語意投影來簡化屬性集也是該領域的研究趨勢 (Yan et al., 2023)。這些成果意味著，透過更詳盡且具體的描述性訊息可以進一步優化視覺 - 語言模型的效果，並提升在各種下游任務上的性能。

我們的研究受到 CLIP 成功的啟發，嘗試將其應用到困難的瑕疵辨識問題中。工程師在對瑕疵照片進行分類時，對於每種瑕疵類別都有其判斷的依據，透過將這些依據轉化為具體的描述性訊息，我們可以更有效地進行缺陷識別任務。

研究方法

本研究所提出的方法從根本上基於跨模態對比學習範式，以增強不同資訊模態之間的相互關係；利用包含描述性資訊生成和多模式學習的兩步驟和兩個編碼器方法，旨在開創一種透過直觀地整合視覺和描述性訊息表示來增強當前製造狀態的系統。在推理階段，首先使用描述編碼器 (descriptive encoder) 將預先定義的缺陷描述資訊提取到描述嵌入 (descriptive embedding) 中，然後使用視覺編碼器將需要分類的缺陷圖像提取到視覺嵌入中。 然後，透過相似度匹配可以找到與圖像語義最匹配的缺陷描述資訊。最後，將最匹配的描述資訊對應到缺陷代碼即可完成預測。**圖一**說明了本研究提出方法的總體摘要，詳細描述將在以下小節中介紹。

一、描述性訊息生成

為了將傳統的分類問題轉化為相似性配對問題，本研究設計了一系列步驟，將離散的缺陷類別轉換為可用於多模態學習的描述資訊。**圖二**演示了缺陷描述性訊息的產生過程。

(一) 透過領域知識定義屬性集

首先，利用該領域專家的知識來找出並定義與描述 TFT-LCD 陣列製造中常見的不同類型缺陷相關的屬性 (例如：瑕疵的型態、發生的製程階段)。這種屬性集 (attribute set) 是此缺陷辨識框架的基礎詞彙。

(二) 為定義的屬性指派屬性值

在定義屬性之後，每一個瑕疵類別的每個屬性都會被指派一個屬性值 (例如：此類別的瑕疵型態為「刮傷」、發生的製程階段為蝕刻) ，類別間的屬性值組合不能重複。透過預先定義好的屬性來描述各種瑕疵類別，建立起具有詳細語義描述的缺陷概況資料庫。

(三) 嵌入生成

為了處理已定義屬性的語義訊息，本研究使用查找表 (look-up table) 將所有缺陷的屬性對應到嵌入向量 (embedding vectors) 。這能夠產生對於各缺陷類別的整體表示 (holistic representation)，進而透過多模態學習系統進行下一步的處理。 缺陷類別 c 的描述嵌入可以計算如下：

$$E_c^T = \text{concatenate}\left[[CLS], e_{c,1}^T, e_{c,2}^T, \ldots e_{c,k}^T\right] \in \mathbb{R}^{k \times d} \quad (1)$$

$$e^{T}_{c,j} = \varphi_j(v_{c,j}) \in \mathbb{R}^d,\ j=1, 2, \ldots k \quad (2)$$

其中 $e^{T}_{c,j}$ 是可學習的嵌入，表示類別 c 的第 j 個屬性的描述資訊；$\varphi: \mathbb{X}_j \rightarrow \mathbb{R}^d$ 是表示預先定義屬性 j 的描述訊息的查找表；$v_{c,j}$ 是類別 c 的第 j 個屬性的屬性值；k 是定義屬性的數量，d 是嵌入維度。

二、多模態機器學習

這個階段利用雙編碼器架構：視覺編碼器以及描述編碼器，透過跨模態對比學習進行訓練，這種範式確保了對視覺和語意 (描述性訊息) 資料兩種模態之間關係的細微理解和表示。

(一) 編碼器

視覺編碼器可以是任何電腦視覺主幹網路 (backbone network)，例如卷積神經網路 (He et al., 2015; Krizhevsky et al., 2012) 或最近被廣泛使用的 Transformer-like 的架構 (Dosovitskiy et al., 2021; Liu et al., 2021; Vaswani et al., 2017)，透過替換最後一個全連接層以滿足所需的嵌入尺寸。 對於描述編碼器，我們使用多層 Transformer layers 來對瑕疵類別的描述資訊進行建模。 給定一個特定的圖像和標題嵌入對 (I, E^T) ，前饋計算可以表示如下：

$$V = f^{visual}(I) \in \mathbb{R}^d , \quad (3)$$

$$\mathrm{T} = f^{descriptive}(E^T) \in \mathbb{R}^{k \times d} t = T_0 \in \mathbb{R}^d, \quad (4)$$

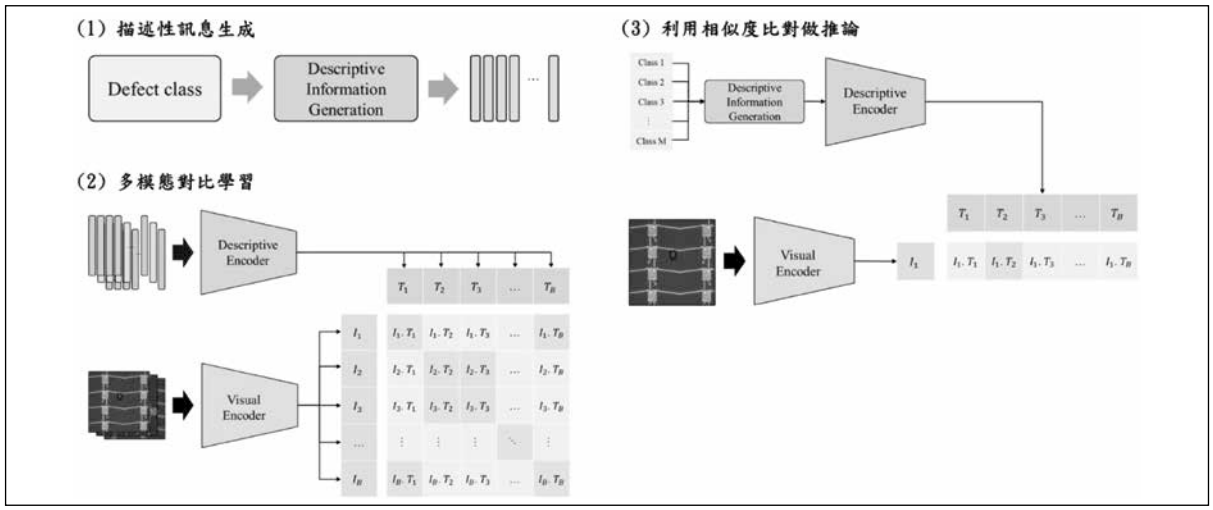
其中 e^{visual} 表示視覺編碼器 f^{visual} 提取的缺陷圖像的嵌入，T 表示從描述編碼器 $f^{descriptive}$ 提取的缺陷描述性嵌入，$t = T_0$ 表示對應 [CLS] token 的嵌入。

(二) 對比學習

與 (Radford et al., 2021) 中描述的原始 ITC 損失不同，本研究使用二元交叉熵損失來訓練模型，因為在一批樣本中對於某單模態實例可能對應多個正對 (相同的缺陷類別可能包含在一批採樣數據中)。 **圖一 (1)** 示範了一批採樣數據中的匹配標籤分配。 修正後的 ITC 損失可表示為：

$$L_{ITC} = \mathbb{E}_{(I,ET)\sim D}[BCE(y^{i2t}, p^{i2t}) + BCE(y^{t2i}, p^{i2t})] \quad (5)$$

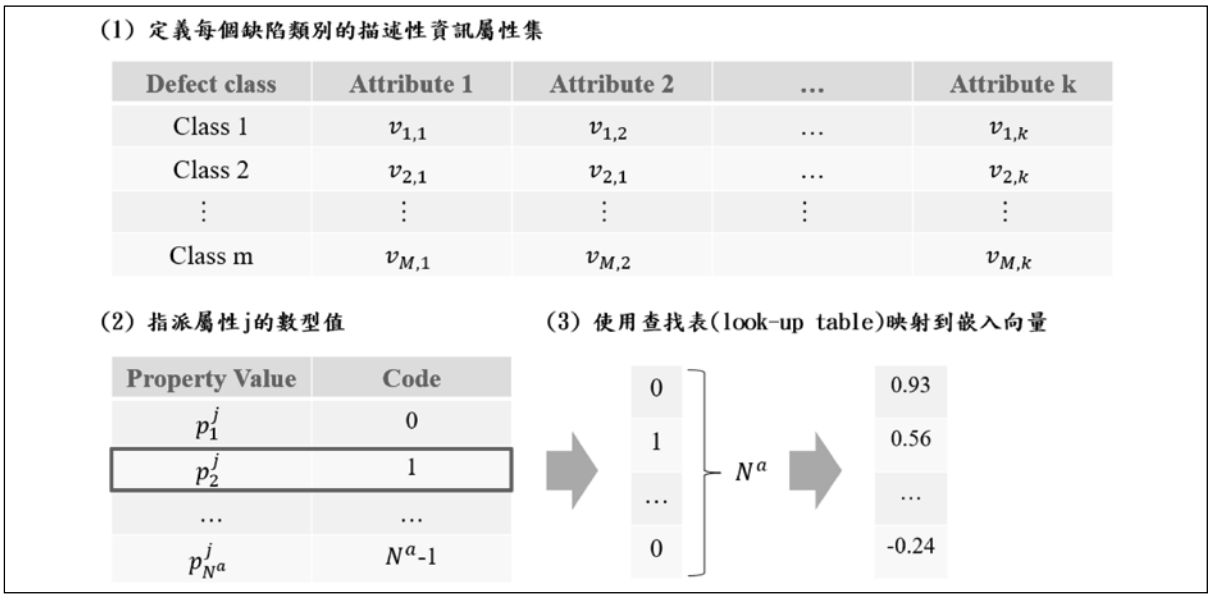
其中 p^{i2t} 和 p^{i2t} 是兩種模態之間的點積相似度；y^{i2t} 和 y^{t2i} 指的是匹配標籤，BCE 是二元交叉熵損失函數。



圖一　本研究提出方法總體摘要。利用跨模態對比學習，本研究提出的兩步驟方法透過雙編碼器系統利用視覺和語義表示來增強自動缺陷分類。 在推理階段，系統使用這些編碼器從缺陷描述和圖像中提取細節訊息，隨後透過相似性匹配識別最佳匹配，從而實現精確的缺陷類別預測

三、利用相似度比對做推論

當模型完成訓練後，首先將定義的缺陷類別屬性集轉化為描述性嵌入。在推論階段，這些描述性嵌入將與由視覺編碼器產生的視覺嵌入進行比對，以實現缺陷分類。擁有了這些描述和視覺嵌入，我們可以透過相似度比對找出與圖像語義最匹配的缺陷描述。接著，只需將這最匹配的描述對應至特定的缺陷類別，即可完成預測。**圖一** (3) 詳細描繪了完整的推論過程。



圖二 將離散缺陷類別轉換為可學習嵌入以供後續多模態學習的描述性資訊產生過程

實驗

為了驗證此方法的有效性，本研究利用了台灣一間 TFT-LCD 領導製造商的資料進行實驗。本實驗的目標是建立一個跨製程道別的自動瑕疵分類模型來準確的辨識缺陷類別，詳細的描述將在以下小節中介紹。

一、實驗數據說明

本研究使用的實驗資料集來自 TFT-LCD 陣列生產過程的 AOI 檢測站。這些影像涵蓋了多種流程和產品中的不同缺陷。圖 3 按照瑕疵類別展示了範例照片，可以觀察到某些缺陷類別間的視覺相似性。整體資料集包含了 161,252 個樣本和 27 個獨特的缺陷類別，更多細節請見**表一**。

表一 各缺陷類別的樣本數總表

缺陷類別	樣本數	缺陷類別	樣本數
C1	5594	C15	12488
C2	5625	C16	4614
C3	5226	C17	5296
C4	13348	C18	10270
C5	15867	C19	29031
C6	4580	C20	142
C7	11470	C21	335
C8	1070	C22	2425
C9	6820	C23	3614
C10	4191	C24	2001
C11	3699	C25	373
C12	4861	C26	277
C13	5130	C27	3268
C14	8201		
總樣本數		169816	

二、實驗設置

本研究使用兩個 NVIDIA TITAN RTX GPU 進行深度學習實驗，選擇 Pytorch 1.8 作為深度學習框架，並在 Python 環境下運行。

由於資料中每個缺陷的樣本數不均衡，本實驗以 0.8、0.1 和 0.1 的比例採用分層抽樣方法對資料進行訓練、驗證和測試的分割。影像首先調整到 384×384 的大小，然後裁切至 256×256 以供模型使用。

訓練策略上，本實驗選擇了對訓練影像進行了隨機裁切、旋轉和翻轉的增強。且設定了早停 (early stopping) 策略：當驗證損失在三個 epoch 不降時，學習率降為 0.1；而在十個 epoch 後損失仍不降時，則終止訓練。

(一) 基線方法 (baseline approach)

瑕疵辨識被視為傳統的多分類任務，採用交叉熵損失進行訓練。本實驗使用了多種經典的模型架構，如 AlexNet、ResNet、EfficientNet(Tan & Le, 2020) 和 ViT。使用 Adam 優化器 (Kingma & Ba, 2014) 來更新模型權重，初始學習率設為 5e-4，且設置了 1e-6 的權重衰減以避免過擬合。

(二) 提出方法 (proposed approach)

模型架構上以基線方法相同的視覺編碼器為基礎，但在最後全連接層進行了修改以符合嵌入維度需求。描述性編碼器使用兩層的 Transformer layers，且兩個編碼器嵌入的維度都設定為 128。本實驗利用四個屬性來描述所有缺陷類別，詳見**表二**。同基線方法，也使用了 Adam 優化器，但初始學習率和權重衰減設為 1e-4 和 5e-4。

表二　本實驗使用的屬性集 (attribute set)，各缺陷類別都有四個屬性的獨特組合。屬性值“None”表示該屬性對於類別沒有區別性的描述

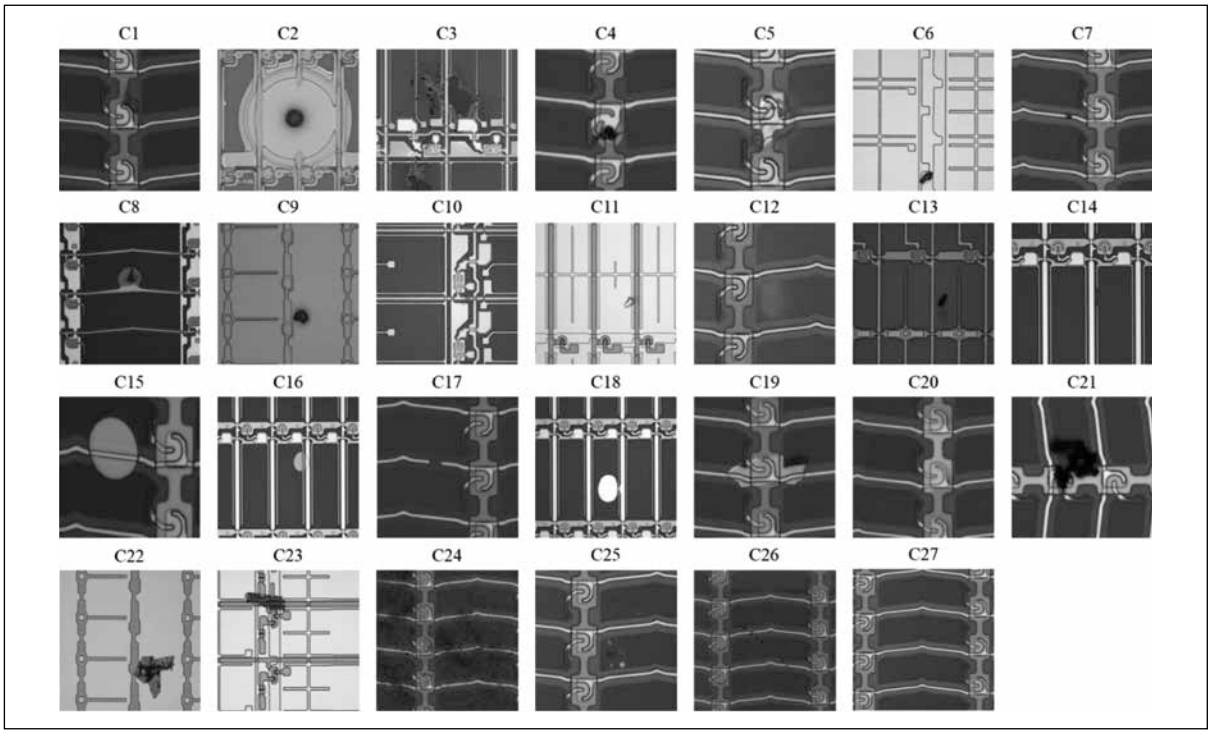
缺陷類別	屬性 1: 發生的層別	屬性 2: 發生的工序	屬性 3: 瑕疵型態	屬性 4: 變形型態
C1	1	Thim-Film	Particle	Circuit Open
C	1	Thim-Film	Splash	None
C3	2	Thim-Film	None	Critical
C4	2	Thim-Film	Residue	None
C5	2	Thim-Film	Hole	None
C6	2	Thim-Film	Particle	In Film
C7	3	Thim-Film	Particle	Circuit Open
C8	3	Thim-Film	Residue	None
C9	3	Thim-Film	Particle	In Film
C10	3	Thim-Film	Spray	None
C1	3	Thim-Film	Hole	None
C12	4	Thim-Film	Hole	None
C13	4	Thim-Film	Particle	In Film
C14	1	Photolithography	None	Circuit Open
C15	1	Photolithography	Residue	None
C16	2	Photolithography	Residue	None
C17	3	Photolithography	None	Circuit Open
C18	3	Photolithography	Residue	None
C19	2	Etching	Residue	None
C20	3	Etching	Residue	None
C21	None	None	Dust	None
C22	1	None	Flake	None
C23	3	None	Flake	None
C24	None	None	Sand	None
C25	None	None	Oil-like	None
C26	None	None	Glass Scratch	None
C27	None	None	None	None

三、實驗結果

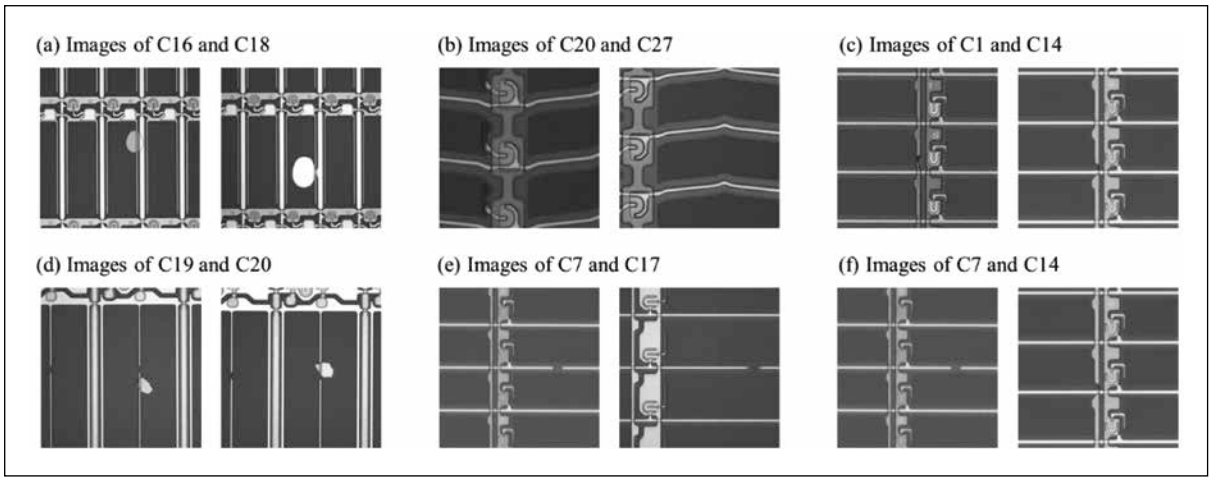
本研究使用了不同的隨機種子初始化模型權重以及分割資料，每個架構都進行了5次的實驗。**表三**展示了題出的新方法在各視覺編碼器上的表現優於基線模型，證明了方法的穩健性且與使用的模型架構無關。此外，**圖三**揭示了在視覺上相似但屬於不同類別的缺陷影像。本研究提出的方法利用**表二**的屬性集紀錄的訊息提取細節差異，從而在視覺特徵相似圖像中學習各瑕疵類別間的異同之處，達到了更佳的識別效果。

表三 不同視覺編碼器上的評估結果 (平均 ± 標準差)

視覺編碼器架構	基線方法 (單模態)	本研究提出方法 (多模態)
AlexNet	76.32% (±1.64)	84.21% (±1.76)
ResNet 18	89.43% (±0.86)	92.56% (±0.72)
ResNet 34	92.56% (±0.82)	93.48% (±0.52)
ResNet 50	92.93% (±0.34)	94.03% (±0.23)
EfficientNet b0	93.47% (±0.12)	95.66% (±0.13)
ViT-B16	92.72% (±0.67)	94.43% (±0.65)



圖三 從 AOI 機台取像的各類別缺陷照片



圖四 視覺上容易混淆的照片。有關兩個缺陷類別之間的細微差異資訊可以在表 2 中找到。(a) 兩張照片都有殘留 (residue type) 缺陷。 右側的電路因缺陷而變形，但左側的電路則沒有。(b) 左邊的缺陷 (residue type) 很小 (只佔整個影像的一小部分)，右邊沒有缺陷。(c)、(e) 兩張照片都是 circuit open 型缺陷，但左邊的縫隙不乾淨，有輕微的黑點，而右邊的斷口相對乾淨。(d) 兩張照片都有殘留缺陷。 所有電路和元件均未變形，但右圖缺陷周圍有金屬殘留痕跡。(f) 兩張照片都是 circuit open 型缺陷，但變形位置不同，對應的是「發生層別」屬性的差異

結論

本文提出了一種新型多模態方法，透過知識工程和跨模態對比學習策略，將傳統分類問題轉換為跨模態內容匹配任務。此策略旨在解決複雜的缺陷識別應用，並在不同模型架構下獲得更好的測試性能。為此，本研究設計了一種基於知識工程的描述性資訊生成技術，這技術以語義豐富的嵌入來取代離散的分類標籤，使模型能夠學習缺陷圖像之間的細粒度差異。實驗結果及其相關分析證明了該方法在 TFT-LCD 陣列製程的實際跨製程缺陷辨識任務中的有效性。

對於未來的研究，可以探索更有效地生成描述性資訊的方法，透過整合例如缺陷區域或面板上組件的位置標註等其他資訊，進一步改進所提出的多模態方法。此外，該方法的應用範疇有望擴展到更多場景，例如 TFT-LCD 製造中的 Color Filter/Cell 製程或是其他光刻工藝，這為製造中複雜的識別任務提供了新的研究方向。

參考文獻

1. Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., … Amodei, D. (2020). *Language Models are Few-Shot Learners* (arXiv:2005.14165). arXiv. http://arxiv.org/abs/2005.14165
2. Chang, Y.-C., Chang, K.-H., Meng, H.-M., & Chiu, H.-C. (2022). A Novel Multicategory Defect Detection Method Based on the Convolutional Neural Network Method for TFT-LCD Panels. *Mathematical Problems in Engineering*, *2022*, e6505372. https://doi.org/10.1155/2022/6505372
3. Chien, C.-F., Ling, Y.-M., Kao, S.-X., & Lin, C.-H. (2022). Image-Based Defect Classification for TFT-LCD Array via Convolutional Neural Network. *IEEE Transactions on Semiconductor Manufacturing*, *35*(4), 650–657. https://doi.org/10.1109/TSM.2022.3199856

4. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2021). *An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale* (arXiv:2010.11929). arXiv. http://arxiv.org/abs/2010.11929
5. Ebayyeh, A. A. R. M. A., & Mousavi, A. (2020). A Review and Analysis of Automatic Optical Inspection and Quality Monitoring Methods in Electronics Industry. *IEEE Access*, *8*, 183192–183271. https://doi.org/10.1109/ACCESS.2020.3029127
6. Eslami, S., de Melo, G., & Meinel, C. (2021). *Does CLIP Benefit Visual Question Answering in the Medical Domain as Much as it Does in the General Domain*? (arXiv:2112.13906). arXiv. https://doi.org/10.48550/arXiv.2112.13906
7. Gu, X., Lin, T.-Y., Kuo, W., & Cui, Y. (2022). *Open-vocabulary Object Detection via Vision and Language Knowledge Distillation* (arXiv:2104.13921). arXiv. https://doi.org/10.48550/arXiv.2104.13921
8. He, K., Zhang, X., Ren, S., & Sun, J. (2015). *Deep Residual Learning for Image Recognition* (arXiv:1512.03385). arXiv. http://arxiv.org/abs/1512.03385
9. Kim, W., Son, B., & Kim, I. (2021). *ViLT*: *Vision-and-Language Transformer Without Convolution or Region Supervision* (arXiv:2102.03334). arXiv. https://doi.org/10.48550/arXiv.2102.03334
10. Kingma, D. P., & Ba, J. (2014, December 22). *Adam*: *A Method for Stochastic Optimization*. arXiv.Org. https://arxiv.org/abs/1412.6980v9
11. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems*, *25*. https://proceedings.neurips.cc/paper_files/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html
12. Li, L. H., Zhang, P., Zhang, H., Yang, J., Li, C., Zhong, Y., Wang, L., Yuan, L., Zhang, L., Hwang, J.-N., Chang, K.-W., & Gao, J. (2022). *Grounded Language-Image Pre-training* (arXiv:2112.03857). arXiv. https://doi.org/10.48550/arXiv.2112.03857
13. Liu, Y., Lee, W.-T., Lu, H-P., Chen, H.-W. (2023). A Novel Multi-Modal Learning Approach for Cross-Process Defect Classification in TFT-LCD Array Manufacturing. (the article has been submitted to *IEEE Transactions on Semiconductor Manufacturing*)
14. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., & Guo, B. (2021). *Swin Transformer: Hierarchical Vision Transformer using Shifted Windows* (arXiv:2103.14030). arXiv. https://doi.org/10.48550/arXiv.2103.14030
15. Lu, H.-P., & Su, C.-T. (2021). CNNs Combined With a Conditional GAN for Mura Defect Classification in TFT-LCDs. *IEEE Transactions on Semiconductor Manufacturing*, *34*(1), 25–33. https://doi.org/10.1109/TSM.2020.3048631
16. Pratt, S., Covert, I., Liu, R., & Farhadi, A. (2023). *What does a platypus look like? Generating customized prompts for zero-shot image classification* (arXiv:2209.03320). arXiv. http://arxiv.org/abs/2209.03320
17. Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I. (2021). *Learning Transferable Visual Models From Natural Language Supervision* (arXiv:2103.00020). arXiv. https://doi.org/10.48550/arXiv.2103.00020
18. Shen, S., Li, C., Hu, X., Yang, J., Xie, Y., Zhang, P., Gan, Z., Wang, L., Yuan, L., Liu, C., Keutzer, K., Darrell, T., Rohrbach, A., & Gao, J. (2022). *K-LITE: Learning Transferable Visual Models with External Knowledge* (arXiv:2204.09222). arXiv. https://doi.org/10.48550/arXiv.2204.09222
19. Tan, M., & Le, Q. V. (2020). *EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks* (arXiv:1905.11946). arXiv. https://doi.org/10.48550/arXiv.1905.11946
20. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). *Attention Is All You Need* (arXiv:1706.03762). arXiv. https://doi.org/10.48550/arXiv.1706.03762
21. Xu, J., De Mello, S., Liu, S., Byeon, W., Breuel, T., Kautz, J., & Wang, X. (2022). GroupViT: Semantic Segmentation Emerges from Text Supervision. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 18113–18123. https://doi.org/10.1109/CVPR52688.2022.01760
22. Yan, A., Wang, Y., Zhong, Y., Dong, C., He, Z., Lu, Y., Wang, W., Shang, J., & McAuley, J. (2023). *Learning Concise and Descriptive Attributes for Visual Recognition* (arXiv:2308.03685). arXiv. http://arxiv.org/abs/2308.03685

作者簡介

劉奕 / 國立清華大學 跨院國際博士班學位學程

陳鴻文 / 國立清華大學 跨院國際博士班學位學程 / 指導老師