

透過自我調整合成與自注意力從單一缺陷數據檢測混合型晶圓缺陷

Inspection of mixed-type wafer defects from single-defect data via adaptive synthesis and self-attention CNN

王子晨

國立臺灣科技大學自動化及控制研究所

學號：M11412013

Abstract—本研讀報告針對論文 *Inspection of mixed-type wafer defects from single-defect data via adaptive synthesis and self-attention CNN* (Measurement, 256, 2025) 進行系統性分析，涵蓋研究目的、方法、實驗結果、程式復現與個人閱讀感想。此期刊旨在解決半導體晶圓檢測中混合型缺陷 (Mixed-type defects) 標註數據稀缺的問題。作者提出了一種利用單一缺陷數據合成混合缺陷圖像的自適應方法，並結合自注意力機制 (Self-Attention) 的 CNN 模型進行分類。實驗結果證實，即使僅使用合成數據進行訓練，該模型在 MixedWM38 數據集上仍能達到與使用真實數據訓練相當的 95% 準確率。

I. 題目與文獻資訊

- 論文名稱：Inspection of mixed-type wafer defects from single defect data via adaptive synthesis and self-attention CNN [1]
- 論文出處：Measurement, 256 (2025) 118306
- 作者：Nguyen Duc Thuan
- 年份：2025/06

II. Introduction

A. 研究目的與背景

半導體製造是一個包含設計、晶圓製造、測試與封裝等多個階段的複雜過程。其中，晶圓檢測 (Wafer Inspection) 位於製造與封裝之間，是確保最終電子元件性能與可靠性的關鍵防線，其核心任務是在產品部署前檢測並剔除次級品。

隨著半導體製程技術的快速進步，對於缺陷檢測的精度要求也日益嚴格。晶圓缺陷主要可分為兩大類：

- 單一缺陷 (Single Defects)：包含點缺陷（如原子缺失）、裂紋、刮痕或表面汙染等特定的局部瑕疵。
- 混合型缺陷 (Mixed-type Defects)：當多種單一缺陷相互作用或重疊時產生，例如群聚缺陷 (Clustered defects) 或光刻過程中的重疊誤差 (Overlay defects)，其模式遠比單一缺陷複雜。

儘管針對單一缺陷的檢測技術已相對成熟，但混合型缺陷的檢測仍面臨巨大挑戰。主要原因在於數據的稀缺性與標註的高昂成本：

- 數據稀缺 (Data Scarcity)：在真實製造環境中，混合型缺陷往往沒有被系統性地記錄或標註，導致訓練樣本嚴重不足。

- 標註困難：收集並精確標註這些複雜的混合模式需要大量的人力與時間資源，造成了模型訓練數據的巨大缺口。

基於上述痛點，本研究提出了一種創新的解決方案，旨在：

- 解決無標註數據問題：基於混合缺陷可由單一缺陷合成的假設，提出一種自適應數據合成方法。利用單一缺陷的 Heat maps 提取感興趣區域，在無需真實混合數據的情況下，合成出保留特徵且具物理合理性的混合缺陷樣本。

- 強化特徵交互學習：開發結合自注意力機制的卷積神經網路 (SACNN)。由於混合缺陷涉及複雜的空間重疊，傳統卷積層難以捕捉全局依賴性，自注意力機制能使模型聚焦於關鍵區域，有效模擬缺陷間的交互作用。

B. 相關技術與現況

晶圓圖缺陷識別技術經歷了從統計方法到深度學習的演進，但各階段技術在面對混合型缺陷時均存在局限性。

1) 傳統機器學習方法：早期技術依賴統計方法與空間分析來識別缺陷聚類。隨後，支持向量機、隨機森林與 k-近鄰算法被廣泛應用。這些方法通常依賴手工提取的特徵，如灰度共生矩陣或局部二值模式。這些方法在單一缺陷分類上表現尚可，但難以捕捉混合型缺陷中複雜的空間特徵與交互作用，且缺乏擴展性。

2) 深度學習與卷積神經網路：CNN 的出現使特徵提取自動化，顯著提升了檢測精度。先進架構如 ResNet、EfficientNet 以及專為混合缺陷設計的 MER-Net 和 Dual-Head CNN，均展現了優異的性能。這些模型通常假設擁有充足的標註數據。當缺乏混合型缺陷的標註樣本時，這些監督式學習模型的泛化能力會大幅下降。

3) 生成式模型與數據增強：為了解決數據不足，研究者嘗試使用生成對抗網絡 (GANs) 如 DCGAN、CycleGAN，以及近期的擴散模型 (Diffusion Models) 如 WMDiff 來生成合成數據。雖然這些方法能緩解數據不平

衡，但它們通常仍需要「部分」真實的混合缺陷數據進行訓練，且 GAN 訓練過程常有不穩定的問題。

4) 現有的合成策略：現有的簡單疊加或合成方法涉及複雜的預處理步驟，且生成的圖像往往不自然，無法準確反映真實世界的缺陷分佈。這會導致模型學習到錯誤的特徵，進而降低在真實數據上的泛化能力。

III. Methods

現有技術在缺乏標註數據的混合缺陷檢測上仍有顯著缺口，這也正是本研究試圖透過自適應 ROI 合成與自注意力機制來填補的關鍵領域。此研究提出了一種創新的兩階段框架，旨在解決半導體製造中混合型缺陷（Mixed-type defects）數據稀缺的問題。第一階段為晶圓圖合成，利用單一缺陷數據透過自適應方法生成混合缺陷樣本。第二階段為自注意力卷積神經網路（SACNN），利用合成數據訓練模型以進行高精度的缺陷分類。

A. 晶圓圖合成

為了生成逼真的混合缺陷晶圓圖，此研究設計了一套自適應合成流程，包含感興趣區域（ROI）提取、權重分配、增強合併及自適應量化四個步驟，如 fig.1 所示。

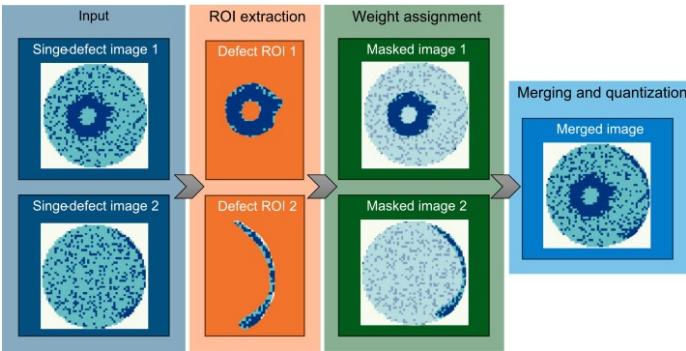


Fig. 1. Framework for synthesizing mixed-type defect wafer maps from single-defect images

1) 感興趣區域提取：對單一缺陷的二值化晶圓圖 $I_i(x, y)$ 進行連通分量標記（Connected Component Labeling），以識別潛在的缺陷模式：

$$L_i(x, y) = \text{label}(I_i(x, y)) \quad (1)$$

在 Eq.(1) 中， $L_i(x, y)$ 將每個連通區域賦予一個唯一的整數標籤。 $I_i(x, y)$ 為輸入的單一缺陷二值化晶圓圖。 $L_i(x, y)$ 為標記後的圖像，其中每個連通區域被賦予一個唯一的整數標籤。為了去除噪聲，應用尺寸濾波器保留面積大於特定閾值 A_{min} 的區域：

$$L_i^{\text{refined}}(x, y) = \begin{cases} L_i(x, y), & \text{if } \text{Area}(L_i = k) > A_{min} \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

Eq.(2) 中， $\text{Area}(L_i = k)$ 計算第 k 個連通分量的像素數量。此研究根據最小可觀測缺陷（約佔晶圓面積 1%）將 A_{min} 設定為 25， $L_i^{\text{refined}}(x, y)$ 為去除噪聲後的標記圖。接著定義 ROI 遷罩：

$$ROI_i(x, y) = \begin{cases} 1, & \text{if } L_i^{\text{refined}}(x, y) > 0 \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

於 Eq.(3) 中 $ROI_i(x, y)$ 若該像素屬於缺陷區域則為 1，否則為 0。為了確保缺陷區域的連續性並平滑邊界，對 ROI_i 進行形態學運算：

$$ROI_i^{\text{final}}(x, y) = (ROI_i(x, y) \oplus B_{\text{dil}}) \ominus B_{\text{ero}} \quad (4)$$

Eq.(4) 中， \oplus 表示膨脹（Dilation）， \ominus 表示侵蝕（Erosion）。 B_{dil} 與 B_{ero} 為半徑 1 至 3 像素的圓盤形之形態學運算的結構元素，用於填補空隙並平滑邊緣。此外，計算缺陷質心 (x_c, y_c) 以支援後續的空間分析：

$$(x_c, y_c) = \left(\frac{\sum_{(x,y) \in ROI_i} x}{|ROI_i|}, \frac{\sum_{(x,y) \in ROI_i} y}{|ROI_i|} \right) \quad (5)$$

Eq.(5) 中 (x_c, y_c) 為缺陷的質心座標。其中 $|ROI_i|$ 代表缺陷區域的總像素數。

2) 權重分配：為每個缺陷像素分配權重，結合了強度資訊與空間距離。越靠近中心或強度越高的像素權重越大，這有助於模擬真實缺陷「中心強、邊緣弱」的特性，系統根據像素強度與距離質心的遠近分配權重 $W_i(x, y)$ ：

$$W_i(x, y) = \alpha \times \frac{I'_i(x, y)}{\max I'_i} + \beta \times \exp \left(-\frac{(x - x_c)^2 + (y - y_c)^2}{2\sigma^2} \right) \quad (6)$$

Eq.(6) 中， $I'_i(x, y)$ 為二階導數強度圖，用於強調邊緣特徵。 α 與 β 為調節強度項與距離項影響力的係數，此研究設定 $(\alpha, \beta) = (1, 1)$ 以平衡強度與空間影響。 σ 為控制高斯分佈範圍的空間尺度參數，設定為 26（約為缺陷邊界框的一半），以控制權重隨距離衰減的程度。

3) 增強與合併：為了增加數據多樣性，對單一缺陷熱圖 $H_i(x, y)$ 進行隨機幾何變換（旋轉、翻轉、平移）得到 $H_i^{\text{aug}}(x, y)$ ：

$$H_i^{\text{aug}}(x, y) = \mathcal{T}(H_i(x, y)) \quad (7)$$

Eq.(7) 中 \mathcal{T} 為幾何變換函數，包含旋轉、翻轉與平移。Eq.(8) 利用 Eq.(6) 計算的權重圖將 N 個增強後的熱圖進行加權疊加，生成混合熱圖 $H_{\text{merge}}(x, y)$ ：

$$H_{\text{merge}}(x, y) = \sum_{i=1}^N W_i(x, y) \times H_i^{\text{aug}}(x, y) \quad (8)$$

Eq.(8) 確保了高權重區域（缺陷核心）在合併後仍被強調。

4) 自適應量化：為了將混合熱圖轉換為二值化晶圓圖，研究引入自適應閾值 $\tau_{adaptive}$ ，根據混合熱圖的統計特性動態調整：

$$\tau_{adaptive} = \mu_{H_{merge}} + \lambda \sigma_{H_{merge}} \quad (9)$$

Eq.(9) 中， $\mu_{H_{merge}}$ 與 $\sigma_{H_{merge}}$ 分別為混合熱圖的均值與標準差， λ 為調節閾值靈敏度的係數，設定為 0.5 以平衡假陽性與假陰性率。最終的二值化圖 $B(x, y)$ 定義為：

$$B(x, y) = \begin{cases} 1, & \text{if } H_{merge}(x, y) > \tau_{adaptive} \\ 0, & \text{otherwise} \end{cases} \quad (10)$$

Eq.(10) 中， $B(x, y)$ 為最終的二值化晶圓圖，1 表示缺陷區域，0 表示非缺陷區域。確保僅保留顯著高於背景噪聲的缺陷模式。

B. 自注意力 CNN

為了有效分類具有複雜空間交互作用的混合型缺陷，本研究在卷積神經網路中引入了自注意力機制（Self-Attention Mechanism）。此架構稱為 SACNN，如 fig.2 所示，包含多層卷積層與自注意力模組，旨在捕捉缺陷間的全局依賴性。模型包含三個卷積層，分別使用 3×3 卷積核與 ReLU 激活函數。在原本的前兩個卷積層後嵌入了自注意力層，使模型能動態捕捉遠距離特徵之間的關聯，而不僅僅是局部特徵。

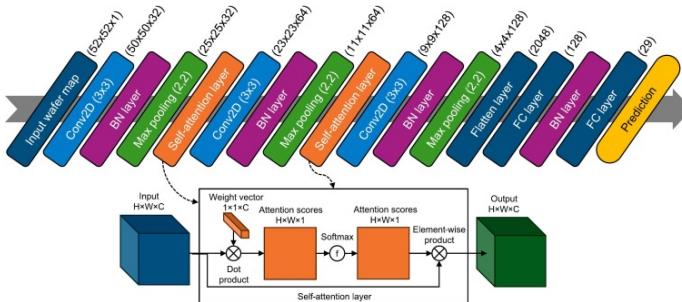


Fig. 2. The architecture of SACNN model

1) 自注意力機制：自注意力層透過計算特徵圖中不同位置的注意力分數，強化關鍵區域的特徵響應。標準化後的注意力權重 α_i 計算如 Eq.(11) 所示：

$$\alpha_i = \frac{\exp(score(X_i))}{\sum_j \exp(score(X_j))} \quad (11)$$

Eq.(11) 中， X 為輸入特徵圖， $score(X)$ 為經線性變換後的注意力分數。計算出的權重 α_i 隨後用於對輸入特徵進行加權 ($Y_i = \alpha_i \times X_i$)，從而放大與缺陷模式最相關的特徵區域。

2) 損失函數：模型訓練採用交叉熵損失函數（Cross-Entropy Loss）進行優化：

$$Z = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^C y_{ij} \log \hat{y}_{ij} \quad (12)$$

Eq.(12) 中， N 為樣本數， C 為類別數， y_{ij} 為真實標籤， \hat{y}_{ij} 為模型預測的機率。此外，模型應用 L_2 正則化以防止過擬合。

C. 定量指標

為了評估合成晶圓圖與真實晶圓圖的相似度，本研究採用了三種指標。分別為結構相似性指數（SSIM）、峰值信噪比（PSNR）與弗雷歇距離（Fréchet Inception Distance, FID）。

1) 結構相似性 (SSIM): 結構相似性 (SSIM) 衡量亮度、對比度與結構的相似性，計算公式如 Eq.(13) 所示：

$$SSIM = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (13)$$

Eq.(13) 中， μ 代表平均值， σ 代表變異數， σ_{xy} 代表共變異數。 C_1 與 C_2 為穩定常數，防止分母為零。

2) 峰值信噪比 (PSNR): 峰值信噪比 (PSNR) 評估圖像重建質量，計算公式如 Eq.(14) 所示：

$$PSNR = 20 \times \log_{10} \left(\frac{MAX_I}{\sqrt{MSE}} \right) \quad (14)$$

Eq.(14) 中， MAX_I 為圖像像素的最大可能值， MSE 為均方誤差。

3) 弗雷歇距離 (FID): 弗雷歇距離 (FID) 衡量兩個分佈之間的差異，特別適用於評估生成圖像的質量。計算公式如 Eq.(15) 所示：

$$FID = \|\mu_r - \mu_s\|^2 + Tr(\Sigma_r + \Sigma_s - 2(\Sigma_r \Sigma_s)^{0.5}) \quad (15)$$

Eq.(15) 中， (μ_r, Σ_r) 與 (μ_s, Σ_s) 分別為真實與生成圖像的特徵均值與協方差矩陣。

IV. Results

實驗首先驗證了合成數據與真實數據的相似度，接著評估了所提出的自注意力 CNN (SACNN) 模型在混合型缺陷分類上的表現，並與現有的其他方法進行了全面的比較。

A. 實驗設置與數據集

此研究的實驗基於公開的 MixedWM38 數據集進行，這是一個廣泛應用於半導體晶圓檢測研究的標準數據集。該數據集包含 38,015 張解析度為 52×52 像素的晶圓圖。這些晶圓圖涵蓋了 38 種不同的缺陷模式。其中包括 1 種正常模式、8 種單一缺陷模式（如 Center, Donut, Edge-Loc, Edge-Ring, Loc, Scratch 等），以及 29 種混合型缺陷模式。混合型缺陷由 2 至 4 種單一缺陷組合而成，具體包含 13 種雙重混合 (2-mixed)、12 種三重混合 (3-mixed) 以及 4 種四重混合 (4-mixed) 缺陷。

fig.3 展示了數據集中各類缺陷的範例圖像，從簡單的單一缺陷到複雜的多重疊加缺陷，顯示了數據集的多樣性與挑戰性。

為了驗證本研究提出的「零樣本 (Zero-shot)」混合缺陷檢測能力，實驗設計採取了嚴格的數據劃分策略：所有的單一缺陷圖像僅用於合成數據的生成（即訓練階段），而所有真實的混合型缺陷圖像則完全保留用於模型的驗證與測試。這確保了模型在訓練過程中從未見過真實的混合缺陷數據，從而能公正地評估合成方法的有效性與模型的泛化能力。

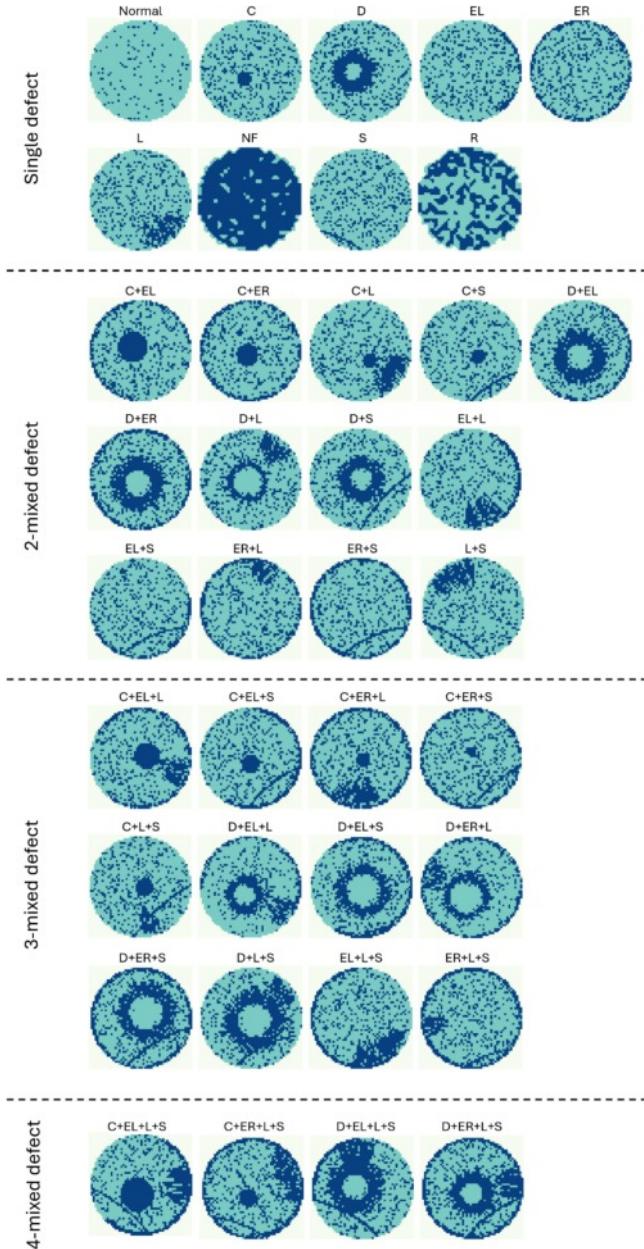


Fig. 3. Example images from the MixedWM38 dataset illustrating the 38 defect types

B. 合成數據品質評估

為了評估本研究提出的自適應合成方法的品質，實驗從定性視覺比較與定量指標分析兩個方面進行了驗證。

1) 定性視覺比較: fig.4展示了不同合成方法(Simple merging, ROI merging, Kim et al.[2], Kang et al.[3], 以及本研究提出的 Proposed method)合成之混合缺陷晶圓圖與對應的真實混合缺陷圖像的視覺比較。結果顯示，簡單合併法(Simple merging)僅是將單一缺陷直接疊加，忽略了空間與強度的連貫性，導致缺陷重疊處出現不自然的邊界。

ROI 合併法雖改善了空間連貫性，但缺乏自適應權重，仍產生突兀的過渡。相比之下，此研究的方法利用自適應 ROI 提取與權重分配，生成的混合缺陷圖在缺陷交互區域

展現出平滑且自然的過渡，在視覺上最接近真實的晶圓缺陷模式。

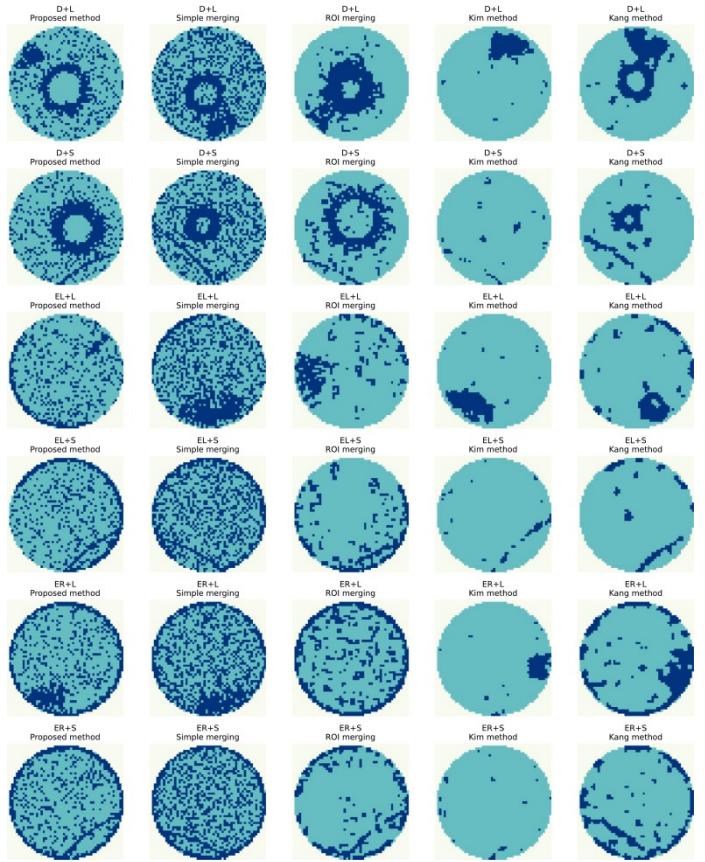


Fig. 4. Visual comparison of wafer map synthesis results across different methods for various defect combinations

2) 定性視覺比較：為了量化合成數據的品質，實驗計算了 SSIM、PSNR 與 FID 三個指標，結果如 TABLE I 所示。

TABLE I
Quantitative comparison of wafer map synthesis methods

Method	SSIM	PSNR	FID
Proposed method	0.91	32.5dB	18.2
Simple merging	0.75	26.7dB	45.8
ROI merging	0.81	28.2dB	38.1
Kim et al.[2]	0.69	24.5dB	52.7
Kang et al.[3]	0.72	21.5dB	49.3

TABLE I中，本研究提出的方法在所有指標上均優於現有的合成方法。SSIM 與 PSNR：本研究提出的方法達到了最高的 SSIM (0.91) 與 PSNR (32.5 dB)。這表明合成圖像在亮度、對比度與結構細節上最接近真實圖像，且失真程度最低。相比之下，Kim et al. 與 Kang et al. 的方法 SSIM 僅約 0.69 至 0.72，顯示其生成的圖像結構與真實數據存在顯著差異。

FID：本研究方法的 FID 得分為 18.2，顯著低於 Simple merging (45.8) 與 Kim et al. (52.7)。FID 分數越低代表合成數據在特徵空間的分佈越接近真實數據分佈，

這證實了本研究生成的數據具有更高的逼真度與特徵一致性。

C. 分類效能評估

本節評估 SACNN 模型在混合缺陷分類上的表現，並分析自注意力機制的影響。

1) 基準模型比較（使用真實數據訓練與測試）：為了建立效能基準，研究首先使用真實混合缺陷數據訓練了兩個模型：標準 CNN 與引入自注意力機制的 SACNN。

標準 CNN 之準確率如 TABLE II 與 fig.5 所示，標準 CNN 的整體準確率為 93%。然而，混淆矩陣顯示該模型在處理複雜的 3 混合與 4 混合缺陷時，容易發生誤判。

TABLE II

Accuracy performance of benchmark method 1 (No merging with standard CNN)

Class	Precision	Recall	F1-Score	Support
C + EL	0.93	0.94	0.94	1000
C + ER	0.95	0.94	0.94	1000
C + L	0.89	0.97	0.93	1000
C + S	0.92	0.96	0.94	1000
D + EL	0.96	0.87	0.91	1000
D + ER	0.93	0.93	0.93	1000
D + L	0.92	0.97	0.94	1000
D + S	0.84	1.00	0.91	1000
EL + L	0.96	0.93	0.95	1000
EL + S	0.96	0.94	0.95	1000
ER + L	0.96	0.95	0.95	1000
ER + S	0.95	0.97	0.96	1000
L + S	0.94	0.98	0.96	1000
C + EL + L	0.94	0.84	0.89	1000
C + EL + S	0.97	0.91	0.94	2000
C + ER + L	0.90	0.96	0.93	1000
C + ER + S	0.88	0.98	0.93	1000
C + L + S	0.90	0.94	0.92	1000
D + EL + L	0.99	0.92	0.95	1000
D + EL + S	0.89	0.80	0.84	1000
D + ER + L	0.93	0.99	0.96	1000
D + ER + S	0.87	0.94	0.91	1000
D + L + S	0.92	0.87	0.90	1000
EL + L + S	0.89	0.88	0.89	1000
ER + L + S	0.93	0.92	0.93	1000
C + EL + L + S	0.95	0.78	0.86	1000
C + ER + L + S	0.88	0.93	0.90	1000
D + EL + L + S	0.95	0.86	0.90	1000
D + ER + L + S	0.93	0.96	0.94	1000
Accuracy			0.93	30,000
Macro Avg	0.93	0.93	0.92	30,000
Weighted Avg	0.93	0.93	0.92	30,000

此研究提出的 SACNN 在引入自注意力機制後，如 TABLE III 與 fig.6 所示，模型的整體準確率提升至 95%，F1-Score 亦達到 0.95。特別是在 D+EL+L、D+ER+S 等複雜類別上，SACNN 的精確率與召回率均有顯著提升，證明了自注意力機制能有效捕捉缺陷間的全局關聯。

2) 可視化分析 (Grad-CAM): 為了進一步解釋 SACNN 的優越性，如 fig.7 展示了標準 CNN 與 SACNN 的 Grad-CAM 熱力圖對比。在包含多個重疊缺陷的複雜樣本（如 C+EL+S）中，標準 CNN 的激活區域較為發散，容易受到背景噪聲干擾或遺漏部分缺陷。相反，SACNN 的激活圖呈現出更聚焦且清晰的特徵響應，能精確定位所有相關的缺陷區域。這視覺化地證實了自注意力機制能幫助模型在複雜背景下「聚焦」於關鍵特徵。

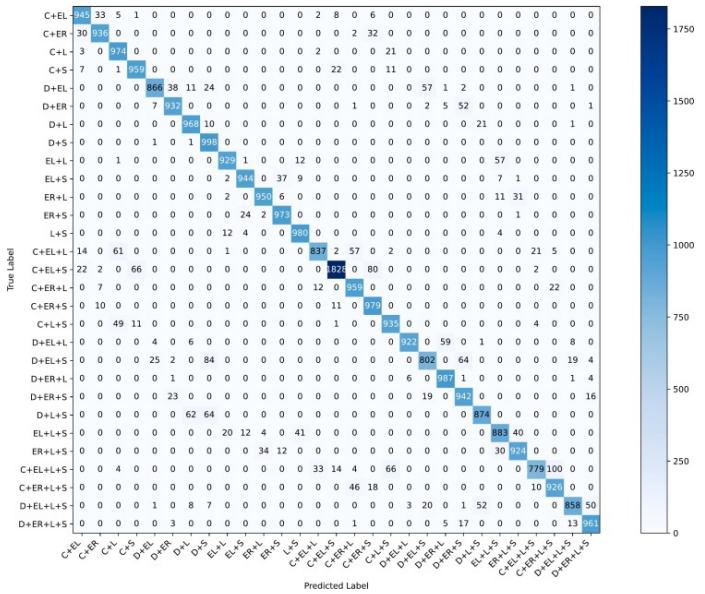


TABLE III
Accuracy performance of benchmark method 2 (No merging with SACNN)

Class	Precision	Recall	F1-Score	Support
C + EL	0.95	0.96	0.96	1000
C + ER	0.92	0.99	0.96	1000
C + L	0.94	0.95	0.95	1000
C + S	0.94	0.99	0.96	1000
D + EL	0.96	0.93	0.95	1000
D + ER	0.92	0.99	0.95	1000
D + L	0.95	0.98	0.97	1000
D + S	0.92	0.98	0.95	1000
EL + L	0.95	0.99	0.97	1000
EL + S	0.98	0.96	0.97	1000
ER + L	0.97	0.97	0.97	1000
ER + S	0.95	0.99	0.97	1000
L + S	0.95	0.98	0.97	1000
C + EL + L	0.91	0.92	0.91	1000
C + EL + S	0.98	0.94	0.96	2000
C + ER + L	0.96	0.92	0.94	1000
C + ER + S	0.94	0.96	0.95	1000
C + L + S	0.92	0.97	0.95	1000
D + EL + L	0.98	0.96	0.97	1000
D + EL + S	0.91	0.88	0.90	1000
D + ER + L	0.97	0.96	0.96	1000
D + ER + S	0.91	0.95	0.93	1000
D + L + S	0.95	0.93	0.94	1000
EL + L + S	0.99	0.88	0.93	1000
ER + L + S	0.95	0.96	0.96	1000
C + EL + L + S	0.95	0.88	0.91	1000
C + ER + L + S	0.95	0.92	0.93	1000
D + EL + L + S	0.99	0.86	0.92	1000
D + ER + L + S	0.94	0.96	0.95	1000
Accuracy			0.95	30,000
Macro Avg	0.95	0.95	0.95	30,000
Weighted Avg	0.95	0.95	0.95	30,000

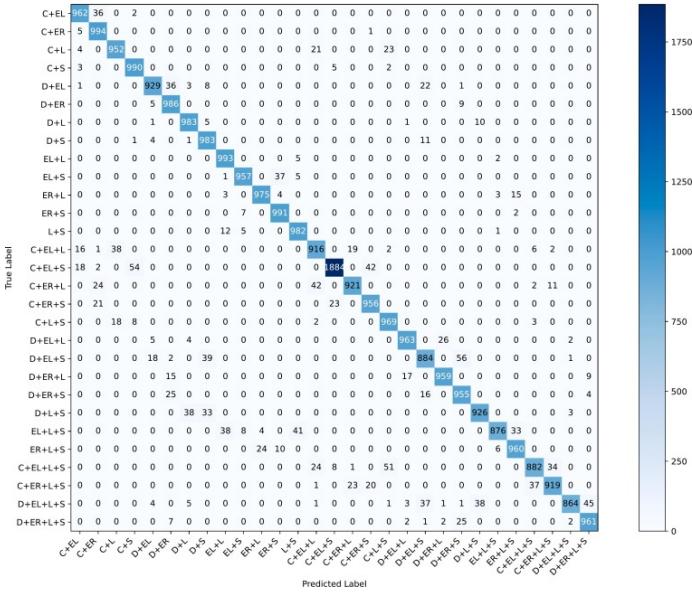


Fig. 6. Confusion matrix of benchmark method 2 (No merging with SACNN).

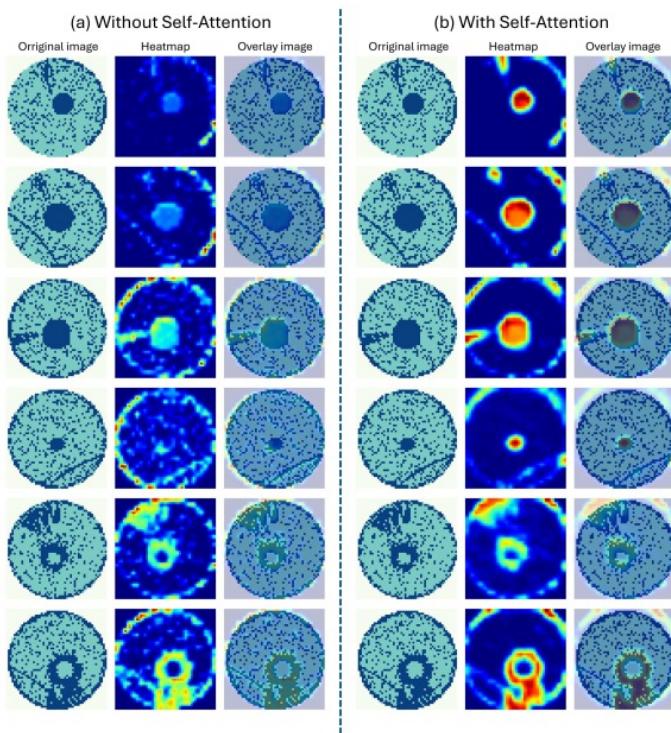


Fig. 7. Grad-CAM visualization results for the standard CNN (a) and SACNN (b) on mixed-type wafer defect images

3) 合成數據訓練之有效性：這部分為此研究最核心的成果。利用本研究提出的合成數據訓練 SACNN 模型，並在真實數據上進行測試。

訓練過程如 fig.8展示了訓練過程中的損失與準確率曲線。模型收斂迅速且穩定，驗證集準確率在 20 個 epoch 內即超過 90%，顯示合成數據具有良好的特徵代表性。

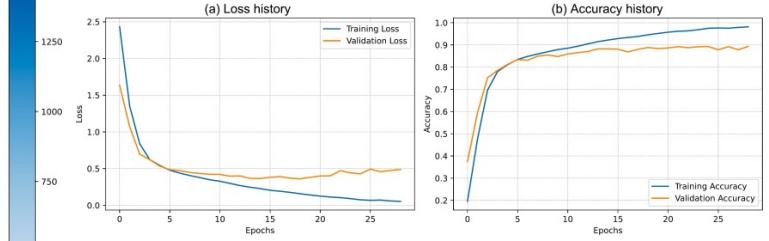


Fig. 8. The training process of the SACNN model on synthetic data. (a) Loss history, and (b) Accuracy history

分類結果如 TABLE IV 所示，使用合成數據訓練的模型在真實測試集上達到了 95% 的整體準確率，此結果與使用真實數據訓練的基準模型 TABLE IV 完全一致。這是一個極具意義的發現，意味著合成數據足以替代真實數據進行模型訓練。

TABLE IV

Accuracy performance of the proposed method across all classes

Class	Precision	Recall	F1-score	Support
C + EL	0.98	0.95	0.96	1000
C + ER	0.96	0.98	0.97	1000
C + L	0.92	0.99	0.95	1000
C + S	0.97	0.95	0.96	1000
D + EL	0.98	0.89	0.93	1000
D + ER	0.91	0.99	0.95	1000
D + L	0.95	0.98	0.97	1000
D + S	0.93	0.97	0.95	1000
EL + L	0.97	0.99	0.98	1000
EL + S	0.98	0.96	0.97	1000
ER + L	0.95	0.99	0.97	1000
ER + S	0.95	0.99	0.97	1000
L + S	0.96	0.98	0.97	1000
C + EL + L	0.97	0.89	0.93	1000
C + EL + S	0.96	0.96	0.96	2000
C + ER + L	0.93	0.98	0.95	1000
C + ER + S	0.92	0.99	0.96	1000
C + L + S	0.97	0.92	0.94	1000
D + EL + L	0.95	0.98	0.96	1000
D + EL + S	0.90	0.89	0.90	1000
D + ER + L	0.98	0.94	0.96	1000
D + ER + S	0.90	0.96	0.93	1000
D + L + S	0.95	0.90	0.93	1000
EL + L + S	0.99	0.86	0.92	1000
ER + L + S	0.93	0.95	0.94	1000
C + EL + L + S	0.92	0.91	0.92	1000
C + ER + L + S	0.96	0.94	0.95	1000
D + EL + L + S	0.88	0.91	0.90	1000
D + ER + L + S	0.97	0.88	0.92	1000
Accuracy			0.95	30,000
Macro avg	0.95	0.95	0.95	30,000
Weighted avg	0.95	0.95	0.95	30,000

在具體類別上，模型展現了極高的穩健性。對於 3-mixed defects，F1-score 達到 0.94。對於最具挑戰性的 4-mixed defects，F1-score 仍維持在 0.91。fig.9的混淆矩陣

進一步顯示，各類別的誤判率極低，證明模型具有極佳的泛化能力。

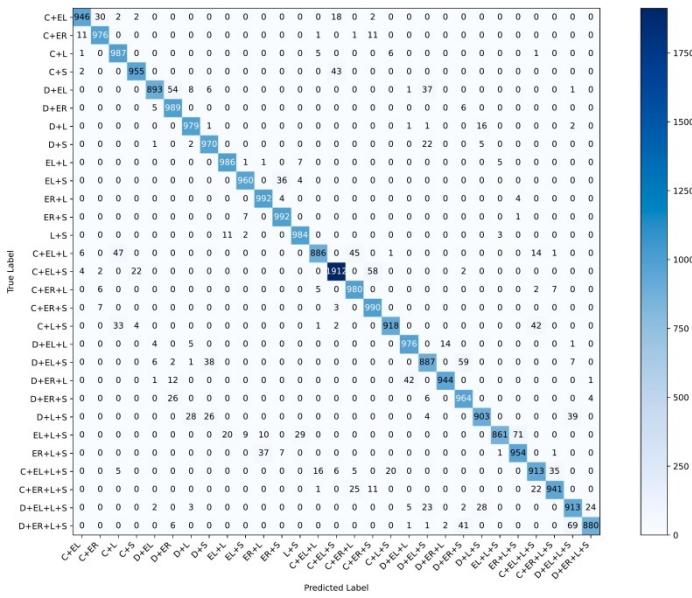


Fig. 9. Confusion matrix of the proposed method (SACNN trained on synthetic data)

4) 與其他合成方法的比較：為了驗證本研究合成策略的優越性，實驗將其與 Simple merging、ROI merging 以及文獻中的其他方法進行了比較。

使用 Simple merging(如 TABLE V 所示、fig.10) 與 ROI merging (如 TABLE VI 所示、fig.11) 數據訓練的模型，其整體準確率分別僅為 40% 與 37%。這些方法在 3 混合與 4 混合缺陷上的表現極差，F1-score 經常接近於 0，顯示簡單合成無法讓模型學習到區分複雜缺陷所需的特徵。

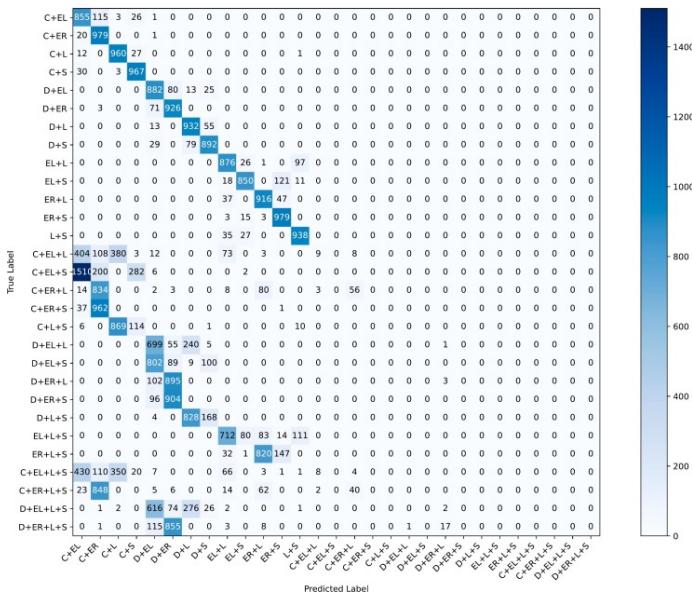


Fig. 10. Confusion matrix of the simple merging method

TABLE VII 總結了不同方法的綜合比較。當使用 Kang et al. 與 Kim et al. 的合成邏輯訓練 SACNN 時，

TABLE V
Accuracy performance of the simple merging method across all classes

Class	Precision	Recall	F1-Score	Support
C + EL	0.26	0.85	0.39	1000
C + ER	0.24	0.98	0.38	1000
C + L	0.37	0.96	0.54	1000
C + S	0.67	0.97	0.79	1000
D + EL	0.25	0.88	0.40	1000
D + ER	0.24	0.93	0.38	1000
D + L	0.39	0.93	0.55	1000
D + S	0.70	0.89	0.79	1000
EL + L	0.47	0.88	0.61	1000
EL + S	0.85	0.85	0.85	1000
ER + L	0.46	0.92	0.61	1000
ER + S	0.75	0.98	0.85	1000
L + S	0.80	0.94	0.86	1000
C + EL + L	0.41	0.01	0.02	1000
C + EL + S	0.00	0.00	0.00	2000
C + ER + L	0.52	0.06	0.10	1000
C + ER + S	0.00	0.00	0.00	1000
C + L + S	0.00	0.00	0.00	1000
D + EL + L	0.00	0.00	0.00	1000
D + EL + S	0.00	0.00	0.00	1000
D + ER + L	0.13	0.00	0.01	1000
D + ER + S	0.00	0.00	0.00	1000
D + L + S	0.00	0.00	0.00	1000
EL + L + S	0.00	0.00	0.00	1000
ER + L + S	0.00	0.00	0.00	1000
C + EL + L + S	0.00	0.00	0.00	1000
C + ER + L + S	0.00	0.00	0.00	1000
D + EL + L + S	0.00	0.00	0.00	1000
D + ER + L + S	0.00	0.00	0.00	1000
Accuracy			0.40	30,000
Macro Avg	0.26	0.41	0.28	30,000
Weighted Avg	0.25	0.40	0.24	30,000

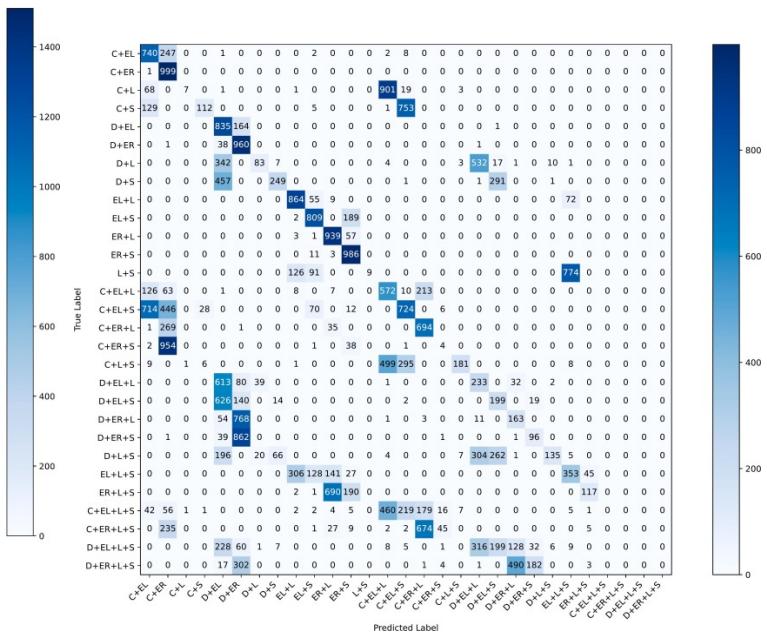


Fig. 11. Confusion matrix of the ROI merging method

TABLE VI
Accuracy performance of the ROI merging method across all classes

Class	Precision	Recall	F1-Score	Support
C + EL	0.40	0.74	0.52	1000
C + ER	0.31	1.00	0.47	1000
C + L	0.78	0.01	0.01	1000
C + S	0.76	0.11	0.20	1000
D + EL	0.24	0.83	0.38	1000
D + ER	0.29	0.96	0.44	1000
D + L	0.58	0.08	0.15	1000
D + S	0.73	0.25	0.37	1000
EL + L	0.66	0.86	0.75	1000
EL + S	0.69	0.81	0.74	1000
ER + L	0.51	0.94	0.66	1000
ER + S	0.65	0.99	0.78	1000
L + S	1.00	0.01	0.02	1000
C + EL + L	0.23	0.57	0.33	1000
C + EL + S	0.36	0.36	0.36	2000
C + ER + L	0.39	0.69	0.50	1000
C + ER + S	0.05	0.00	0.01	1000
C + L + S	0.90	0.18	0.30	1000
D + EL + L	0.17	0.23	0.19	1000
D + EL + S	0.21	0.20	0.20	1000
D + ER + L	0.20	0.16	0.18	1000
D + ER + S	0.29	0.10	0.14	1000
D + L + S	0.88	0.14	0.23	1000
EL + L + S	0.29	0.35	0.32	1000
ER + L + S	0.68	0.12	0.20	1000
C + EL + L + S	0.00	0.00	0.00	1000
C + ER + L + S	0.00	0.00	0.00	1000
D + EL + L + S	0.00	0.00	0.00	1000
D + ER + L + S	0.00	0.00	0.00	1000
Accuracy			0.37	30,000
Macro Avg	0.42	0.37	0.29	30,000
Weighted Avg	0.42	0.37	0.29	30,000

準確率僅約 38% 至 40%，遠低於此研究的 95%。這突顯了此研究提出的自適應合成策略在保留缺陷特徵與物理合理性方面的關鍵優勢，使其成為目前解決混合型晶圓缺陷數據稀缺問題的最佳方案。

TABLE VII
Classification accuracy across different mixed-defect categories for various synthesis methods

Merging Method	Model	All mixed defects
Simple merging	Standard CNN	0.37
Simple merging	SACNN	0.4
ROI merging	Standard CNN	0.37
ROI merging	SACNN	0.41
Kang et al.	Standard CNN	0.38
Kang et al.	SACNN	0.4
Kang et al.	Kang et al.	0.72
Kim et al.	Standard CNN	0.37
Kim et al.	SACNN	0.4
Kim et al.	Kim et al.	0.93
Proposed method	Proposed model	0.95

V. Implementation and Validation

這部分我自行撰寫程式碼重建了論文所述的自適應數據合成管道 (Adaptive Synthesis Pipeline) 與自注意力卷積神經網路 (SACNN)。我採用消融研究策略，將實驗分為

四個層次。首先驗證模型架構的有效性，其次評估合成數據的品質指標，最後探討分類效能落差及其成因。

A. 模型架構驗證

在評估合成數據之前，首先確保自行實作的 SACNN 模型架構具備學習混合缺陷特徵的能力，排除模型實作錯誤的可能性。此階段直接使用 MixedWM38 真實數據集進行監督式學習。將有多重缺陷之真實晶圓圖共 30000 張，依照 14:3:3 的比例隨機劃分為訓練集、驗證集與測試集。訓練歷程如 fig.12 所示。Loss 與 Accuracy 曲線均證明模型收斂良好且無過擬合跡象。

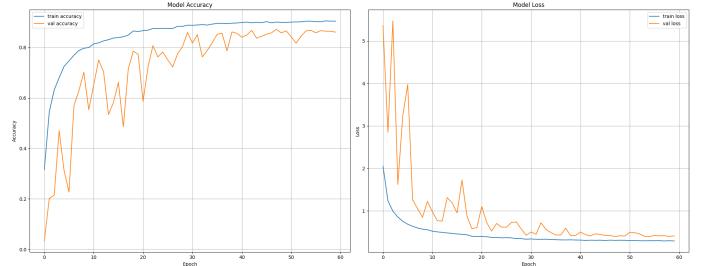


Fig. 12. 以真實數據訓練之訓練過程

實驗結果顯示，在完全使用真實數據訓練的情況下，本實作的模型達到了 87% 的整體準確率。混淆舉證如 fig.13 所示。

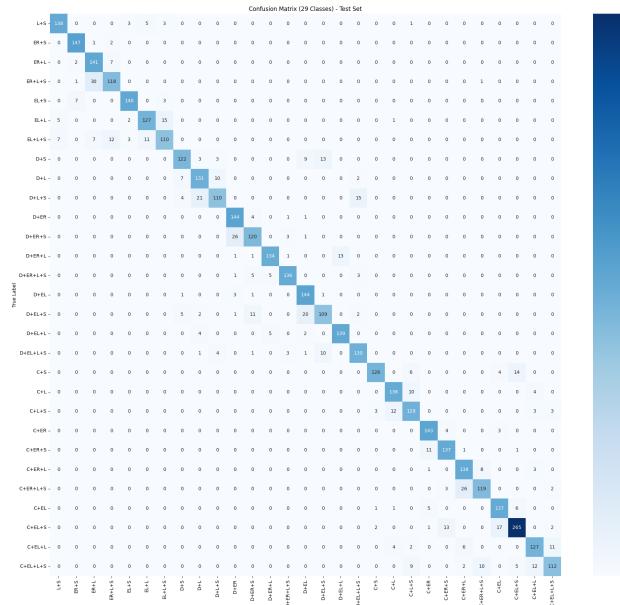


Fig. 13. 實真數據訓練之混淆舉證

此結果雖略低於論文報告的 95%，推測差異可能源於超參數微調或預處理細節。但此結果已足以證明本研究實作的 SACNN 模型架構是正確且有效的，具備捕捉複雜空間交互作用的能力。

B. 合成策略評估

確認模型無誤後，轉向驗證論文核心的數據合成之方法。

首先我根據論文描述重建了自適應數據合成管道，生成了與論文相似的合成數據集。接著使用此合成數據集訓練 SACNN 模型，訓練過程如 fig.14 所示，訓練完成後在真實測試集上進行評估。若觀察訓練過程可發現模型收斂良好，驗證集準確率在 30 個 epoch 內即接近 80%。

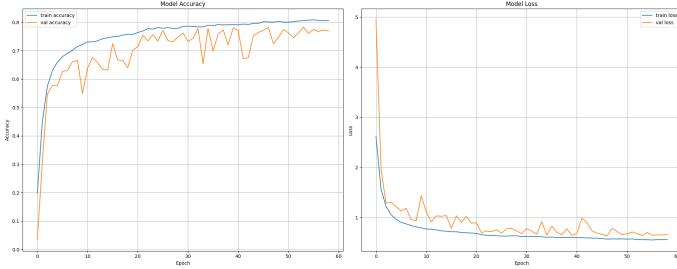


Fig. 14. 以合成數據訓練之訓練過程

混淆矩陣如 fig.15 所示，模型在真實測試集上僅有 14.7% 的整體準確率。

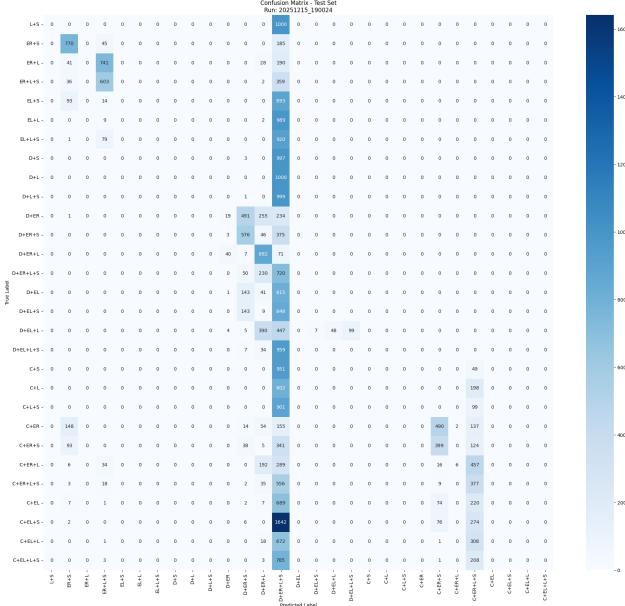


Fig. 15. 以合成數據訓練之混淆舉證

14.7% 的整體準確率遠低於論文報告的 95%，顯示自行實作的合成策略未能有效捕捉真實數據的關鍵特徵分佈。進一步分析發現，在自適應量化這部分若完全依照論文描述實作，依照 Eq.(10) 所示的二值化閾值決策，會導致大量像素被分類為背景，進而損失了重要的缺陷特徵。近一步分析可發現真實數據集上除了特徵缺陷外，背景噪聲與其他非缺陷特徵也佔據了相當比例的像素。因此我在自適應量化部分進行了調整，在二值化之後加入雜訊以更貼近真實數據的分佈。

$$B(x, y) = \begin{cases} 1, & \text{if } H_{merge}(x, y) > \tau_{adaptive} \\ 0, & \text{otherwise} \end{cases} \quad (10)$$

加入雜訊後的合成數據與未加入雜訊前如 fig.16、fig.17、fig.18 等所示進行比較，可見缺陷特徵得以保留且背景更為多樣化，與真實數據更為接近。

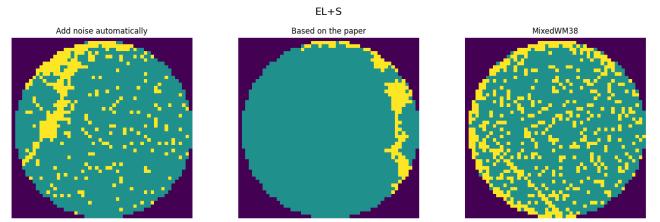


Fig. 16. EL+S 合成數據比較

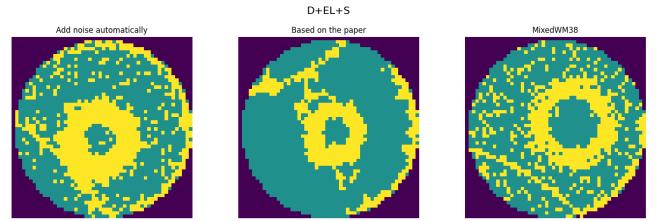


Fig. 17. D+EL+S 合成數據比較

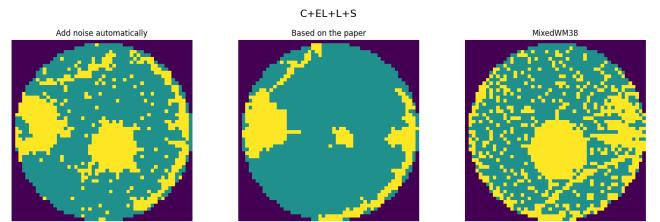


Fig. 18. C+EL+L+S 合成數據比較

重新跟改合成策略合成數據並訓練 SACNN，訓練過程如 fig.19 所示。

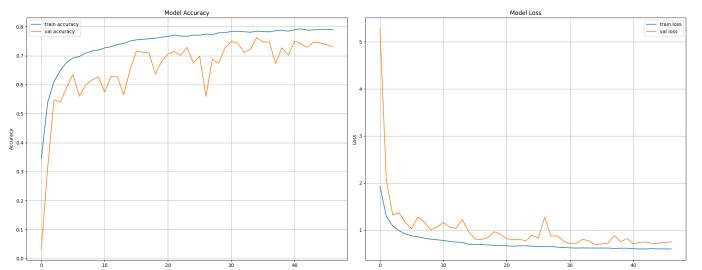


Fig. 19. 以加入雜訊合成數據訓練之訓練過程

模型在真實測試集上提高至 51.5% 的整體準確率，顯著優於先前的 14.7%。混淆矩陣如 fig.20 所示。然而仍遠低於論文報告的 95%，顯示合成策略仍有改進空間。

所有類別中的準確率表現如 TABLE VIII 所示，可見具有顯著幾何特徵的組合表現較好，例如 L+S (F1: 0.90) 與 ER+S (F1: 0.77)。這表示 Scratch (S) 與 Edge-Ring (ER)

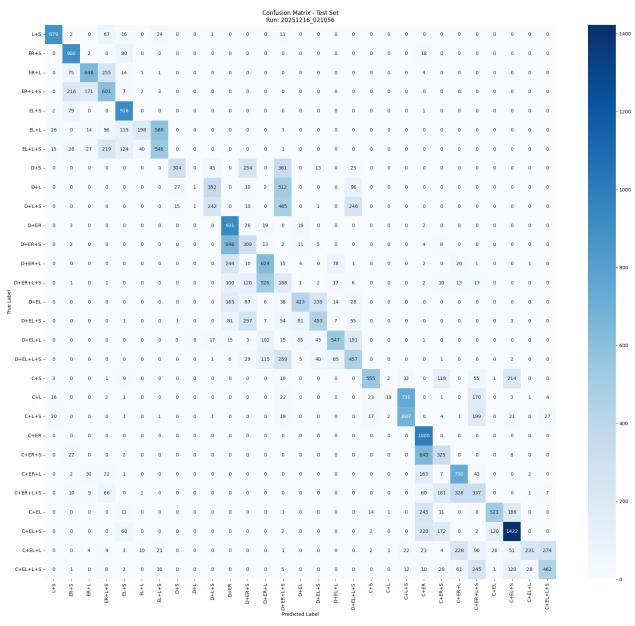


Fig. 20. 以加入雜訊合成數據訓練之混淆舉證

的特徵最強，模型容易捕捉。涉及 Donut (D) 與 Loc (L) 的複雜組合表現極差，例如 D+L (F1: 0.00) 與 D+ER+L+S (F1: 0.13)。這暗示了這兩類缺陷在疊加時可能發生特徵遮蔽，導致模型無法區分。

混淆矩陣 fig.20 顯示 D+L (Donut + Loc) 的辨識率極低 (F1-Score: 0.00)。Donut 是大面積的環狀結構，Loc 是局部區域。在合成過程中，若 Loc 剛好落在 Donut 的路徑上，其特徵數值極易被 Donut 覆蓋。這導致合成數據中的 D+L 實際上丟失了 Loc 的特徵，模型因此學到了錯誤的關聯，將其誤判為單純的 D。在 ER+L 等類別中，模型傾向於預測出不存在的第三種缺陷（如誤判為 ER+L+S）。這可能是合成過程中的形態學運算在邊緣產生了類似 Scratch 的線條偽影，誤導了模型。

C. 定量指標評估與分析

進一步使用 SSIM、PSNR 與 FID 指標來評估合成數據的品質。實測結果為：SSIM: 0.23，PSNR: 11.48 dB，FID: 49.17。這些數值與論文報告（SSIM: 0.91，PSNR: 32.5 dB，FID: 18.2）存在顯著差異。顯示合成數據在結構相似性與分佈貼近度方面仍有很大改進空間。

SSIM 是針對「像素對齊 (Pixel-aligned)」的指標。由於隨機配對策略，合成圖與真實圖在結構上無法對齊 (Misalignment)，導致 SSIM 均僅有 0.24。

PSNR 是基於均方誤差 (MSE) 計算的像素級指標。與 SSIM 相同由於本研究在合成過程中依照論文引入了隨機旋轉與平移以增加數據多樣性，導致合成圖中的缺陷位置與隨機配對的真實圖完全錯位。這一部份無法的得知論文中是否有額外的對齊步驟，或是使用了不同的配對策略。

FID (52.38) 衡量的是特徵空間的分佈距離，較不受像素位置影響。雖然雜訊注入改善了分佈，但數值仍高於論文的 18.2。這顯示 Sim-to-Real 之間仍存在紋理 (Texture) 與邊緣特徵上的領域偏移，真實數據中可能包含更複雜的製程偽影，是目前合成算法尚未完全模擬的。

TABLE VIII
Detailed Classification Report

Class	Precision	Recall	F1-Score	Support
C + EL	0.78	0.52	0.63	1000
C + ER	0.42	1.00	0.59	1000
C + L	0.76	0.02	0.04	1000
C + S	0.91	0.56	0.69	1000
D + EL	0.71	0.42	0.53	1000
D + ER	0.43	0.93	0.58	1000
D + L	0.12	0.00	0.00	1000
D + S	0.86	0.30	0.45	1000
EL + L	0.77	0.20	0.32	1000
EL + S	0.67	0.92	0.78	1000
ER + L	0.72	0.65	0.68	1000
ER + S	0.67	0.90	0.77	1000
L + S	0.91	0.88	0.90	1000
C + EL + L	0.88	0.23	0.37	1000
C + EL + S	0.70	0.71	0.71	2000
C + ER + L	0.53	0.73	0.61	1000
C + ER + S	0.37	0.33	0.35	1000
C + L + S	0.46	0.69	0.55	1000
D + EL + L	0.73	0.55	0.63	1000
D + EL + S	0.57	0.45	0.50	1000
D + ER + L	0.44	0.62	0.52	1000
D + ER + S	0.28	0.31	0.29	1000
D + L + S	0.37	0.24	0.29	1000
EL + L + S	0.46	0.55	0.50	1000
ER + L + S	0.46	0.60	0.52	1000
C + EL + L + S	0.60	0.46	0.52	1000
C + ER + L + S	0.29	0.34	0.31	1000
D + EL + L + S	0.41	0.46	0.43	1000
D + ER + L + S	0.09	0.19	0.13	1000
Accuracy			0.52	30,000
Macro Avg	0.56	0.51	0.49	30,000
Weighted Avg	0.57	0.52	0.50	30,000

TABLE IX
Quantitative comparison of wafer map synthesis methods

Method	SSIM	PSNR	FID
Proposed method	0.91	32.5dB	18.2
Simple merging	0.75	26.7dB	45.8
ROI merging	0.81	28.2dB	38.1
Kim et al.[2]	0.69	24.5dB	52.7
Kang et al.[3]	0.72	21.5dB	49.3
依照論文合成之資料集	0.24	11.78dB	148.42
依照論文合成並加入雜訊之資料集	0.23	11.48dB	49.17

D. 分類效能分析

本實作在真實測試集上的多標籤之準確率為 52%，與論文的 95% 顯著有落差。但若改採多標籤之準確率，分別判斷是否含有某種缺陷，而非要求組合完全正確，準確率可回升至約 92%。準確率表現如 TABLE X 所示。這證明模型確實學到了各別缺陷的特徵，但在判斷多種缺陷的精確組合邊界時，任然受限於合成數據的真實度不足。

VI. Discussion

此報告完整回顧了論文提出的「自適應合成(Adaptive Synthesis)」與「自注意力 CNN (SACNN)」方法，並透過自行撰寫程式碼進行了嚴謹的實作驗證。綜合論文結果與此研究的實作數據，本章節將針對方法的貢獻、實作的落差及其背後的物理意義進行深入探討。

TABLE X
Classification Report

Class	Precision	Recall	F1-Score	Support
Center	1.00	0.98	0.99	12,000
Donut	1.00	1.00	1.00	11,000
Edge-Loc	0.98	0.73	0.84	12,000
Edge-Ring	0.79	1.00	0.88	11,000
Loc	0.99	0.86	0.92	17,000
Scratch	0.77	0.80	0.79	18,000
Micro avg	0.91	0.88	0.89	81,000
Macro avg	0.92	0.89	0.90	81,000
Weighted avg	0.92	0.88	0.89	81,000
Samples avg	0.91	0.89	0.89	81,000

A. 論文方法的核心貢獻與價值

低成本的數據增強證明了僅需單一缺陷樣本，透過幾何變換與 ROI 融合，即可生成足夠訓練深度模型的混合數據，大幅降低了標註成本。

SACNN 的引入有效彌補了傳統 CNN 在處理分散式、重疊式缺陷時感受野不足的問題，在論文實驗中達到了 95% 的表現。

B. 實作差異

然而，本研究的實作結果顯示，僅依照論文描述重建的合成策略與 SACNN 架構，無法達到論文所報告的高準確率。經過多次調整與分析後，發現合成數據在結構相似性 (SSIM)、峰值訊噪比 (PSNR) 與分佈貼近度 (FID) 等指標上均有顯著落差。論文方法生成的圖像若無額外的雜訊注入，合成數據訓練的模型幾乎無法遷移至真實場景。此外，合成數據在多標籤分類的精確組合邊界上表現不佳，顯示合成策略在模擬缺陷間複雜交互作用時仍有不足。

VII. 個人閱讀感想

起初我認為遵循論文公式即可重現其宣稱的 95% 準確率，然而實作中真實數據訓練 87% 與合成數據訓練 52% 的顯著對照明顯與論文不一。讓我體悟到論文往往省略了關鍵的工程細節。未來的突破口不應僅侷限於模型架構的堆疊。

VIII. 程式碼與資料取得

本期末報告所使用 python 程式碼皆已全數上傳至 GitHub：https://github.com/PrinceLego/Machine_Learning_Final_Report

References

- [1] N. D. Thuan, “Inspection of mixed-type wafer defects from single-defect data via adaptive synthesis and self-attention cnn,” *Measurement*, vol. 256, p. 118 306, 2025, issn: 0263-2241.
- [2] W. Shin, H. Kahng, and S. B. Kim, “Mixup-based classification of mixed-type defect patterns in wafer bin maps,” *Computers & Industrial Engineering*, vol. 167, p. 107 996, 2022, issn: 0360-8352.
- [3] J. Shim and S. Kang, “Learning from single-defect wafer maps to classify mixed-defect wafer maps,” *Expert Systems with Applications*, vol. 233, p. 120 923, 2023, issn: 0957-4174.