

第 1 題：Capsule 類別分析

資料集：Capsule (MVTec AD)

1. 缺陷類別數量

‘capsule’ 類別共包含 5 種缺陷類別，不包括「good」類別。

2. 缺陷類別種類

根據資料集中的資料夾結構，五種缺陷類別如下：

- faulty imprint
- crack
- scratch
- poke
- squeeze

3. 使用的圖片總數

訓練集圖片數量：219 張（全部為 good 圖片）

測試集圖片數量：132 張（包含 good 與缺陷圖片）

Ground truth 缺陷標註圖數量：109 張（僅對缺陷測試圖提供標註）

4. 訓練與測試資料分佈

訓練集僅包含 good 圖片，用於學習正常樣本特徵。

測試集包含正常與缺陷樣本，用於評估分類與異常檢測能力。

Ground truth 標註圖對應於缺陷樣本，可用於像素層級的分割評估。

5. 影像尺寸

所有圖片皆為 RGB 彩色圖像，解析度為 1000×1000 像素，在訓練集、測試集與標註圖中皆一致。

第 2 題：模型改進

	學習率	epoch	優化器	改進點	準確率(%)
baseline	1e-3	50	adam	X	28.57
dropout	1e-4	50	adam	增加 dropout(0.7)	32.14
unfreeze resize224	1e-4	50	adam	預訓練模型解凍 更改成 resize(224, 224)	53.57
data_argument resnet50	1e-4	50	adam	更多數據增強 更改成 resnet50	57.14
L2	1e-4	50	adam	加入 L2 正則化	64.29

為了提升模型對 MVtec AD 中 capsule 產品的分類表現，我實作了四種改進策略。首先，我在 baseline 架構中加入 Dropout(0.7)，提升模型的正則化能力，使準確率從 28.57% 提升至 32.14%。接著，解凍預訓練模型（ResNet）並將輸入尺寸調整為 (224, 224)，顯著提升表現至 53.57%。第三，我採用多種資料增強技術並改用 resnet50，有效提升至 57.14%。最後，在上述設定基礎上加入 L2 正則化（weight_decay=1e-4），最終達到 64.29% 的準確率。整體而言，使用更深的模型結構、適當的資料增強與正則化策略，是促使模型泛化能力提升的關鍵。所有實驗皆訓練 50 epochs，使用 Adam 優化器與學習率 1e-4。

第 3 題：long-tail distribution 和解決方法

Long-Tailed Distribution 是一種類別數量遵循重尾規律的分布，其中少數類別佔據大多數樣本，而大多數類別僅擁有極少量樣本。

《Distribution Alignment : A Unified Framework for Long-tail Visual Recognition》

DisAlign 透過兩階段訓練架構解決長尾分布問題：第一階段學習特徵表徵，第二階段固定特徵並校正分類器輸出分數。其核心在於針對每類別加入可學的縮放與偏移參數，搭配動態信心函數調整 logit，使模型更公平地對待尾部類別。同時提出廣義重加權機制，根據類別樣本數調整學習權重，有效改善資料不平衡造成的分類偏差。

第 4 題：缺少異常資料的解決方法

在僅有正常樣本的情況下，可利用預訓練模型提取影像特徵，建立正常樣本的特徵分布（如多變量高斯分布），再透過 Mahalanobis 距離或機率密度函數評估測試樣本是否偏離該分布。若距離過遠或密度極低，則判定為異常，無需依賴缺陷樣本即可實現高效的異常偵測。

第 5 題：

1. 若進行物件偵測模型（如 YOLO-World）的微調，需準備每張影像的缺陷類別與對應的邊界框（bounding box）標註。若使用語意分割模型（如 SAM），則需提供每張影像的像素遮罩（segmentation mask），可由 MVTec AD 中的 `ground_truth` 資料夾取得。這些資料必須與原圖精確對齊，才能進行有效的模型訓練。
2. YOLO-World 與 SAM 皆為預訓練大型模型，具備少量樣本下的遷移能力。MVTec AD 資料集具有缺陷樣本稀少、不平衡等特性，因此適合使用 few-shot 或 prompt-based 模型進行微調。此外，SAM 的即時遮罩生成能力，有助於缺陷位置與形狀的精準標示，提升異常偵測解析度。