

姓名：孫秉新

學號 113034511

1.

我選擇的產品是 bottle。

Number of defect classes : 3

Types of defect classes : broken_large、broken_small、contamination

Number of images used in your dataset :

訓練資料集總共有 209 筆資料，測試資料集在每種分類都有大約 20 筆資料，詳細資料集資訊如圖 1 所示。

```
training data:
{'good': 209}
test data:
{'broken_large': 20, 'good': 20, 'broken_small': 22, 'contamination': 21}
ground_truth data:
{'broken_large': 20, 'broken_small': 22, 'contamination': 21}
```

圖 1

Distribution of training and test data :

根據原始資料集的資料數量，我的訓練與驗證資料的比例為 4:1，由於每個缺陷類別的資料大約有 20 筆，所以就從每個類別各取 20 筆資料劃分為訓練和驗證資料，如圖 2 所示。

```
Shape of x_train: (48, 3, 900, 900)
Shape of x_val: (12, 3, 900, 900)
Shape of y_train: (48,)
Shape of y_val: (12,)
```

圖 2

Image dimensions :

圖片的維度為 900 x 900，如圖 2 所示。

2.

我的策略是先調整好訓練過程，在套不同的模型看結果，首先我將 batch size 從

32 改成 16，因為訓練資料只有 48 筆，所以我覺得 32 太大了，而 8 又太小會導致訓練曲線非常動蕩。然後我將 epoch 數調整到 300，比較看得出訓練曲線的收斂狀況，跑出來的結果也會比較穩定。然後將 learning rate 調整為 $5e-4$ ，模型才能比較穩定的學習。我有嘗試將學習率調整的部分改成使用 ReduceLROnPlateau，但是訓練出來的效果不好，所以還是使用原來的方法。

我嘗試了四種預訓練模型，分別是 resnet18、resnet50、efficientnetb0 和 mobilenetV3，在四種模型都凍結權重，只加最後一層輸出層的情況下，resnet18 和 resnet50 表現得比較好，因此我再去調整 resnet18 和 resnet50 的輸出層數量以及神經元數量，並且開放預訓練模型訓練第四層的權重，這樣調整後，表現最好的模型是 resnet18，測試準確度可達 100%，而 resnet50 也不錯，測試準確度可達 91.67%。

3.

(1)

長尾分佈指的是數據集中少數類別佔據大部分的樣本，而多數類別則擁有較少的樣本，形成長尾形狀的統計分佈。這會導致模型在少數樣本類別上的表現不佳，因為它主要學習來自較多樣本的類別。在 MVTec AD 數據集中，我們可以觀察到「Good」類別的樣本數量遠多於各種缺陷類別，我認為這會使模型更傾向於預測為「Good」，而無法有效識別缺陷類別。

(2)

我有查到一篇研究數據不平衡問題的論文，題目為 "Balanced Contrastive Learning for Long-Tailed Visual Recognition"。作者提出了一種基於對比學習的方法來處理長尾分佈問題。該方法透過以下技術來平衡數據：

1. Contrastive Learning：利用特徵空間中的相似性學習，確保不同類別的樣本能夠獲得區別性較高的表示。
2. Memory Bank：使用存儲機制來平衡長尾類別，讓稀少類別的特徵不會因為數據量不足而被模型忽略。
3. 類別平衡策略：在訓練時動態調整不同類別的樣本權重，確保長尾類別在訓練中得到足夠的關注。

應用於 MVTec AD 數據集的可能性

在 MVTec AD 中，由於「Good」類別的樣本遠多於缺陷類別，我們可以透過這種 Contrastive Learning 來增強模型對缺陷類別的學習能力。此外也可以透過 Memory Bank 技術，可以有效保存並增強少數類別的特徵學習，使模型能夠更準確地辨識不同種類的缺陷。

4.

根據網路上的資料，主要有以下三種策略。並且這些方法可結集成技術，提高異常檢測的準確度與穩健性。

1. Autoencoder

自編碼器透過學習「良品」圖像的重建，當遇到缺陷圖像時，重建誤差會顯著增加，進而判斷異常。

2. Generative Models

生成對抗網絡（GAN）或擴散模型可以學習「良品」圖像的分佈，並在測試時判斷圖像是否偏離該分佈，從而識別異常。

3. Self-Supervised Learning

Contrastive Learning 讓模型學習區分人工擾動的良品圖像與原始圖像，從而提升異常檢測能力。

5.

(1)

如果要用 YOLO-World 來做物件偵測，需要準備標註好的圖片，標註方式是用方框（Bounding Box）把缺陷圈起來，並記錄這些框的位置和對應的缺陷類別。

如果要用 SAM 來做影像分割（準確標出缺陷的形狀），則需要更細緻的標註，不只是方框，而是用遮罩（Mask）標出缺陷的範圍，就像在圖片上「塗顏色」來標示出異常區域。

(2)

由於 YOLO-World 擅長即時物件偵測，能夠快速識別圖像中的異常區域，並可透過轉移學習（Transfer Learning）適應 MVTec AD 的缺陷檢測任務。SAM 具備強大的泛化能力，能夠根據提供的提示進行精細分割，有助於準確區分缺陷區域，提高異常檢測的細緻度。此外，這些模型均為開源，且經過大規模數據訓練，能夠透過少量標註數據微調，使其適應特定產業的缺陷檢測需求。因此具備開源與可遷移性。