

National Tsing Hua University
11320IEEM 513600
Deep Learning and Industrial Applications
Homework 3

Name: 巫宛芸

Student ID: 113034601

Due on 2025/04/10

1. 我選用資料集中的 zipper 進行檢測。Zipper 資料集細節如下：

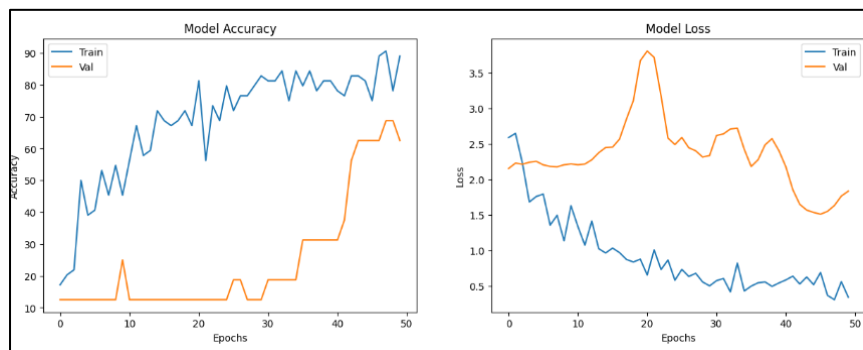
- 缺陷類別數量(Number of defect classes)：7 種
- 七種缺陷類型(Types of defect classes)：broken_teeth / combined / fabric_border / fabric_interior / rough / split_teeth / squeezed_teeth
- 圖片總數 (Number of images used)：共使用 391 張圖像
訓練資料集 (train/good)：240 張 (皆為正常樣本)
測試資料集 (test)：151 張 (包含 32 個正常樣本、七種異常樣本(各有 15~18 個樣本))

Train: 240, Test: 151, Total: 391
- 訓練和測試資料集分布(Distribution of training and test data)：訓練集和測試集的比例約為 6:4。訓練集符為單一正常樣本，符合非監督式的異常檢測設定；而測試集的七種異常分別有 15~18 個樣本，但正常樣本有 32 個，分布稍微不平均。
- 圖片尺寸(Image dimensions)：RGB 彩色，尺寸為 1024*1024

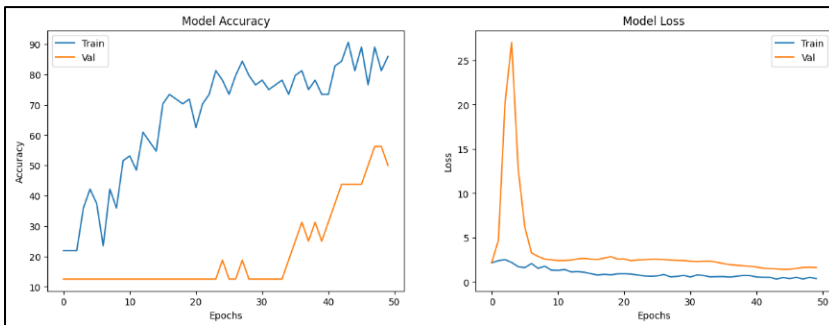
2. 下表整理四種方法，詳細說明在下方。

實驗	Pre-trained model	方法	驗證準確率
1	ResNet18	解凍 layer4、Resize 128	68.75%
2	ResNet50 (modifying the pre-trained model)	解凍 layer4、Resize 128	56.25%
3	ResNet18	(延續實驗 1) 解凍 layer4、Resize 128 調低 batch size、調高 epoch	75.0%
4	ResNet18	(延續實驗 3) 解凍 layer4、Resize 128 調低 batch size、調高 epoch 調低 learning rate	81.25%

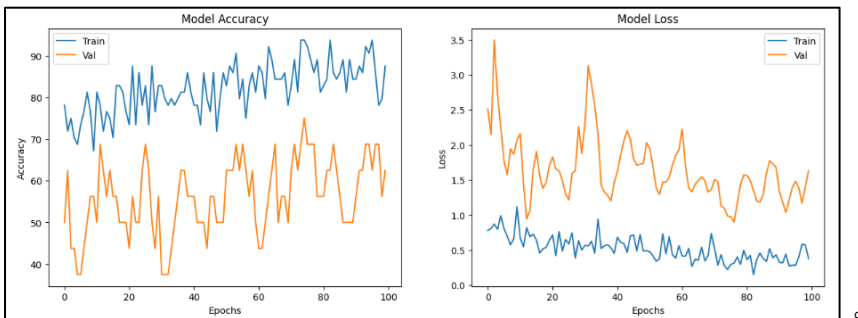
實驗 1：使用 ResNet18 作為預訓練模型，解凍最後一層 (layer4) 並且把圖片尺寸調整成 128×128，保留更多細節以提升模型辨識能力。最終驗證準確率達 68.75%。



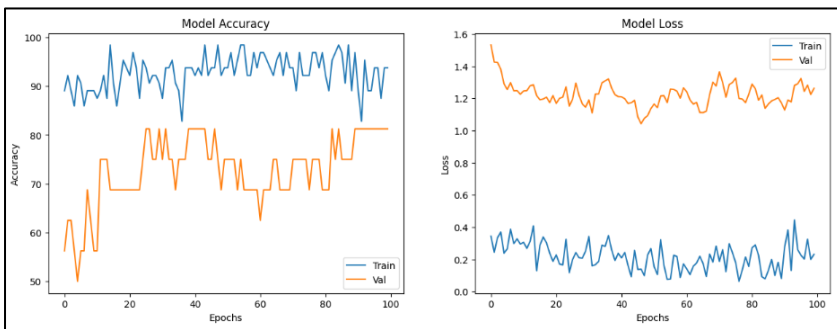
實驗 2：把 Pre-trained model 從 ResNet18 升級到 ResNet50，使用跟實驗 1 相同的輸入尺寸與解凍設定，驗證準確率最高為 56.25%。(可能因為資料量有限，模型出現過擬合，效能反而低於 ResNet18)



實驗 3：延續實驗 1 的模型與資料設定，進一步調整 hyperparameters，將 batch size 減少為 16(提升特徵學習的敏感度)，並增加 epoch 數為 50。在這組實驗，模型的驗證準確率最高達 75%。batch size 縮小同時導致準確率曲線震盪幅度變大



實驗 4：延續實驗 3 的設定，並調整學習率為 $1e-4$ ，讓模型參數更新更穩定，避免在小樣本情境下過度震盪。結果顯示驗證準確率提升至 81.25%，是四組實驗中最高。總結四組實驗，實驗 4 表現最好，關鍵在於保留 ResNet18 與解凍 layer4，並進一步調整超參數，讓模型在小樣本下更穩定收斂並學到有效特徵。



3.

(i) 定義 long-tail distribution：指的是資料中大多數樣本集中在少數主要類別 (head classes)，而其餘大多數類別 (tail classes) 則僅包含少量樣本。這種分布會讓模型在訓練時，更容易學會常見類別，忽略那些樣本比較少的類別，導致分類不平均，讓模型在處理樣本少的類別時表現較差。

(ii) [Long et al. \(2022\) 《Retrieval Augmented Classification for Long-Tail Visual Recognition》](#)

這篇論文提出一種 RAC (Retrieval Augmented Classification) 方法解決類別不平衡問題。RAC 在傳統分類模型中加入一個檢索模組，可從外部資料庫中找出與輸入圖像語意相近的範例與文字描述，並與主模型整合，提升對樣本較少的 tail classes 的辨識能力。應用我們的資料集中，透過建立包含缺陷樣本的檢索資料庫，即使各類缺陷的樣本數不多，仍能藉由語意補強方式提升分類表現。

4. 可以用 unsupervised 或 self-supervised 的異常檢測方法。譬如用 autoencoder 或 Variational Autoencoder(VAE)學習正常樣本的重建特徵，再透過 reconstruction error 判斷是否異常；另外也可以用 PatchCore 這類方法建立正常樣本的 feature memory，測試時比較特徵距離就能找出異常區域。這些方法不需要標記缺陷，適合 MVTec 這種缺陷樣本少的情況。

5.

(i)

- YOLO-World(object detection)：要準備每張圖片的 bounding box (x, y, width, height)，標記出每個缺陷區域的「座標」跟對應的「類別標籤」。格式可以用 YOLO 格式(.txt)或 COCO 格式(.json)。
- SAM(segmentation)：要準備每張圖片對應的 pixel-wise mask，用來標示缺陷區域的精確形狀。pixel-wise mask 格式可以用 PNG 或 COCO segmentation 格式。

(ii) 這些模型有良好的「特徵擷取能力」和「空間感知能力」，可以針對圖片中的異常區域進行定位與分割。使用少量有標註的缺陷樣本來進行 fine-tuning，就能快速適應像 zipper 這種資料集中的缺陷特徵，提升準確性與可解釋性，因此適合應用在我們資料集的缺陷辨識。