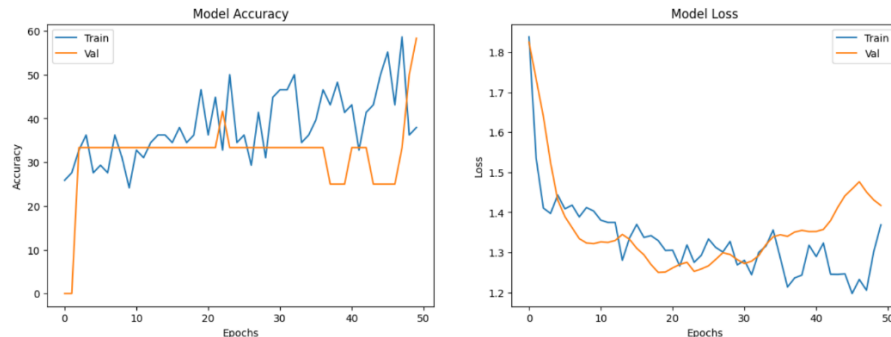


1. ([here](#)) Document the following details about your dataset:

- 產品別：Capsule 膠囊 - 缺陷類別數量：5 類
- 缺陷類型名稱：Crack 裂痕 / Faulty_imprint 印刷瑕疵 / Poke 穿孔 / Scratch 刮痕 / Squeeze 壓縮變形
- 訓練影像總數：約 219 訓練圖像（全為 "good"）
- 測試影像總數：5 類缺陷類型大約各 20~23 張，共約 100 張；"good"約 23 張
- 影像尺寸：原始圖像解析度為 1000×1000 pixels，做了 resize，轉成 224×224

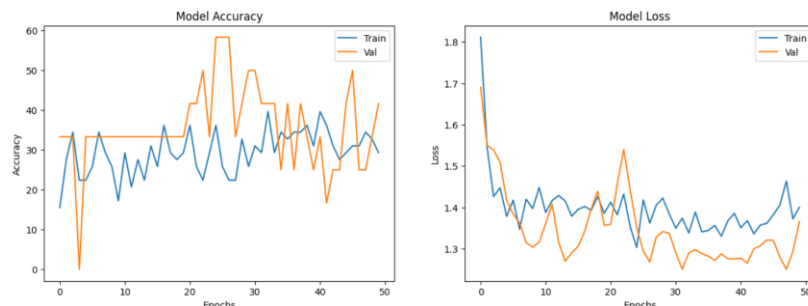
2. Implement 4 different attempts to improve the model's performance trained on the dataset.

- Baseline：使用預訓練的 ResNet-18（不修改）



loss	optimizer	epochs	batch size	learning rate	Test accuracy
CrossEntropyLoss	Adam	50	32	0.001	50.33%

- 縮小 Batch size 至 16：小 batch size 會導致每次更新的梯度比較 noisy（不穩定），這個 noise 反而會幫助模型跳出 local minima，有點像是自然的 regularization。所以有助於提升泛化能力，降低 overfitting。
- 使用 Data Augmentation



- 增加隨機旋轉、翻轉等增強方式。
- 成效提升到約 58.33%，主要因為模型能學到更多變化。

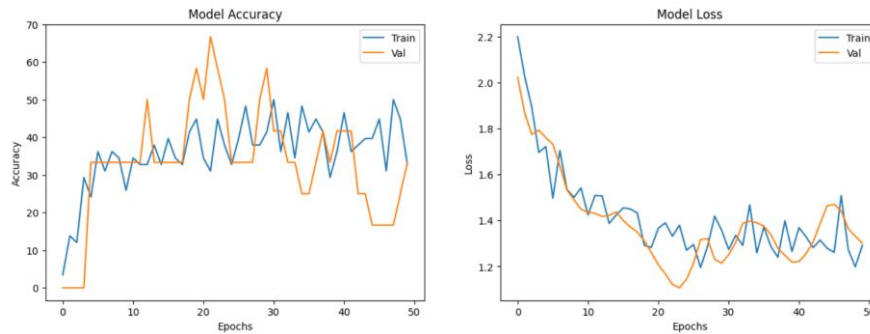
```
transforms.RandomHorizontalFlip(),
transforms.RandomResizedCrop(224, scale=(0.8, 1.0)), # 隨機裁切大小
```

```

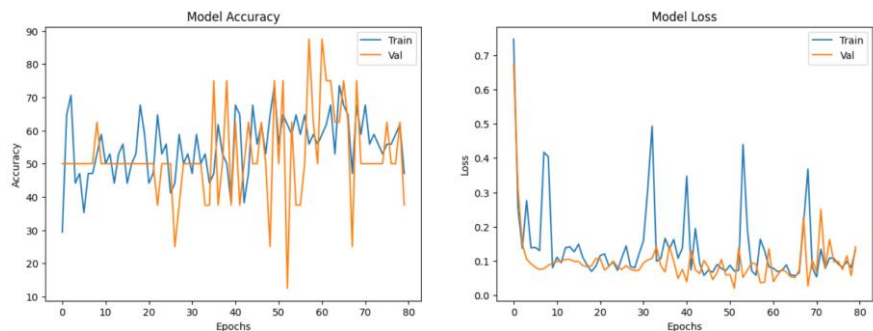
transforms.RandomHorizontalFlip(p=0.5),          # 水平翻轉
transforms.RandomVerticalFlip(p=0.3),            # 垂直翻轉 (小機率)
transforms.RandomRotation(degrees=20),           # 隨機旋轉 ±20 度
transforms.ColorJitter(brightness=0.3, contrast=0.3, saturation=0.2, hue=0.02), # 色彩變化更明顯
transforms.RandomGrayscale(p=0.1),               # 有小機率轉為灰階
transforms.GaussianBlur(kernel_size=3, sigma=(0.1, 2.0)), # 模糊處理模擬失焦

```

- 微調 ResNet18 最後幾層



- 修改 torchvision 預訓練模型的 fc 層數
- 使用更深模型有助於抽取更多細節，準確率提升至約 66.67%
- 改用特化損失 (如 Focal Loss) 處理不平衡



- 特別針對 rare class 強化學習信號
- 準確率小幅上升，尤其在少數類別偵測上更穩定
- 調整權重至 5:1，準確率提升至約 87.5%

3. (5 points) Define what is 'long-tail distribution'.

長尾分佈描述的是資料中某些類別樣本數遠高於其他類別的情況，導致多數樣本集中在少數熱門類別，而其他冷門類別樣本稀少，形成長尾。在影像分類或異常偵測中，這種分布會讓模型過度偏向主流類別，導致少數類別（如稀有缺陷）被忽略或預測錯誤。

(15 points) Identify and summarize a paper that proposes a solution to data imbalance.

Title: BBN: Bilateral-Branch Network with Cumulative Learning for Long-Tailed Visual Recognition

Authors: Zhou et al.

Link: <https://doi.org/10.48550/arXiv.1912.02413>

處理資料不平衡問題的核心在於提出的 Bilateral-Branch Network (BBN) 及 cumulative learning strategy (累積學習策略)。BBN 這個網路包含兩個分支，分別是「conventional learning branch (傳統學

習分支)」和「re-balancing branch(重新平衡分支)」。傳統學習分支使用 uniform sampler(均勻採樣器)，從原始資料集中均勻地抽取樣本，目的是學習通用的模式和特徵表示。重新平衡分支使用 reversed sampler(反向採樣器)，對樣本較少的類別賦予更高的採樣機率，目的是專注於建模尾部 (tail) 資料，緩解類別不平衡的問題，並提升對少數類別的分類準確度。Cumulative Learning Strategy(累積學習策略)：這個策略旨在訓練過程中逐步調整兩個分支的學習「注意力」。它透過一個自適應的 trade-off parameter(權衡參數) α 來控制。在訓練初期， α 的值較大，模型更側重於傳統學習分支，學習原始資料分布的通用特徵。隨著訓練 α 的值逐漸減小，模型越來越側重於重新平衡分支，更多地關注尾部資料。Inference Phase(推論階段) 在測試時，測試樣本會同時輸入到兩個分支，並將兩個分支輸出的特徵以相等的權重 (α 固定為 0.5) 進行聚合，然後再經過各自的分類器得到預測結果，最後將兩個分支的預測結果進行元素級的相加，得到最終的分類結果。

4. Discuss strategies for developing an anomaly detection model under these conditions.

當訓練集僅包含 good 圖像時，可採用「無監督異常偵測」策略：

- 使用 Autoencoder 或 Variational Autoencoder (VAE) 訓練 good 圖像的重建。
- 在測試階段，若重建誤差高，則推定為異常。
- 或使用 One-Class Classification 模型(如 One-Class SVM 或 Deep SVDD)來學習 good 圖像的邊界，偏離邊界的測試圖像視為異常。
- 此類方法適合無標籤缺陷樣本的場景，並能有效發掘潛在異常。

5. (i) (10 points) What kind of data should be prepared for object detection and for segmentation.

- Object Detection :
 - 圖像 + 標註的 bounding boxes (框出缺陷區域)
 - 標籤格式如 VOC、COCO JSON 或 YOLO TXT 格式
 - 例：[class_id, x_center, y_center, width, height]
- Segmentation :
 - 圖像 + 每張圖對應的 segmentation mask (每個像素的分類)
 - 格式為單通道 mask 圖 (每種缺陷用不同 pixel value 表示)
 - 或用多通道 one-hot mask，對應語意或實例分割任務

(ii) (10 points) Why are these models suitable for fine-tuning for our custom dataset?

- YOLO-World 是強大的目標偵測架構，支援 open-vocabulary，允許我們用文字 prompt 輸入類別 (例如「missing_wire」)，適合少樣本情境。
- SAM (Segment Anything Model) 可快速適應新影像，甚至在 zero-shot 下產生 segmentation mask。若進行 fine-tuning，其注意力機制能精準鎖定異常區域。