

Name:周冠宇 Student ID:113030509

1. (10 points) Download the MVTec Anomaly Detection Dataset from Kaggle (here). Select one type of product from the dataset. Document the following details about your dataset: • Number of defect classes. • Types of defect classes. • Number of images used in your dataset. • Distribution of training and test data. • Image dimensions.

選擇類別： Cable（電纜）

缺陷類別數量： 8 類（含 good 則為 9 類）

缺陷類別名稱：

bent\_wire、cable\_swap、combined、cut\_inner\_insulation、  
cut\_outer\_insulation、missing\_cable、missing\_wire、poke\_insulation

使用圖片總數： 466 張

訓練資料分佈： 約 224 張（全部為 good 類別）

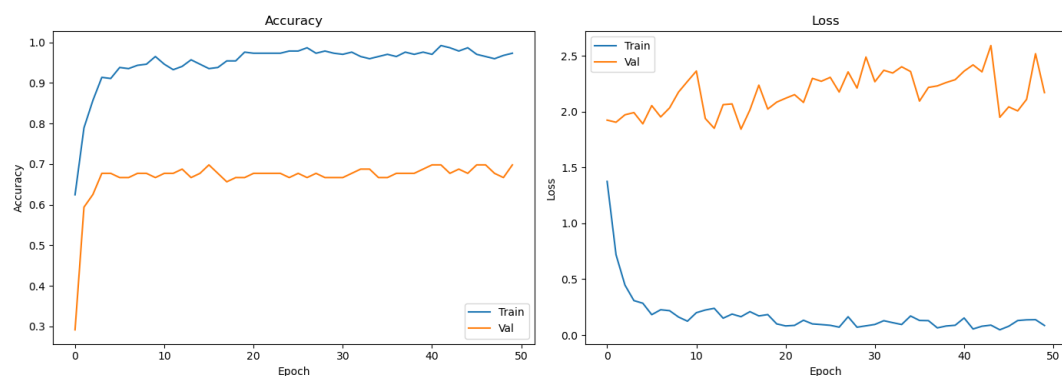
測試資料分佈： 約 150 張，包含 good 及所有 8 種缺陷

圖片尺寸： 1024 × 1024，RGB（三通道）

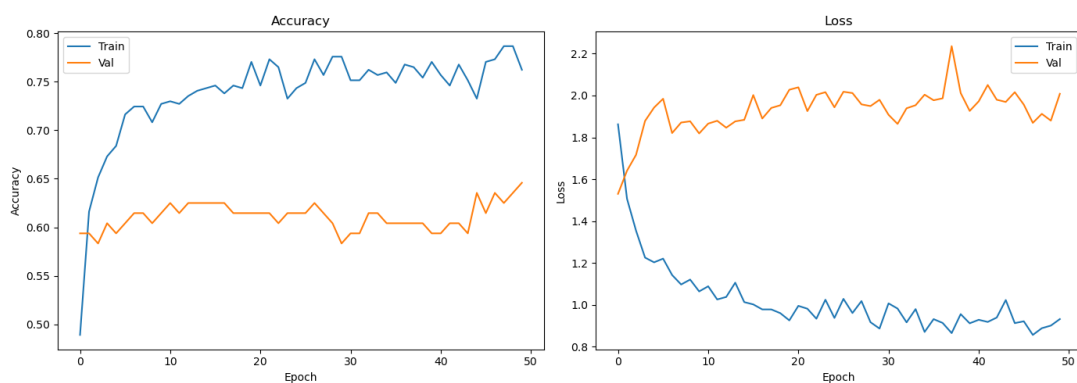
(30 points) Implement 4 different attempts to improve the model's performance trained on the dataset you choose in previous question. Ensure that at least one approach involves modifying the pre-trained model from TorchVision. Summarize the outcomes of each attempt, highlighting the best performing model and the key factors contributing to its success. You may also need to describe other hyperparameters you use in your experiment, like epochs, learning rate, and optimizer. (Approximately 150 words.)

各方法最終驗證準確率：

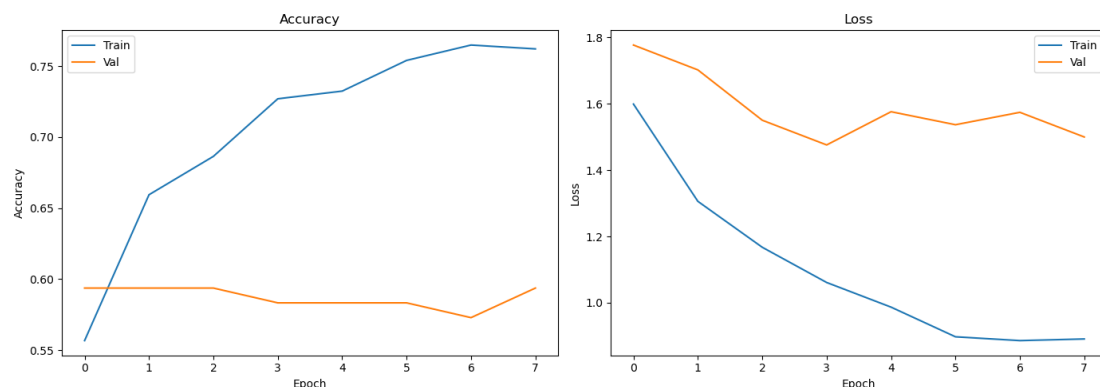
### 1. Fine-tuning: 0.6979



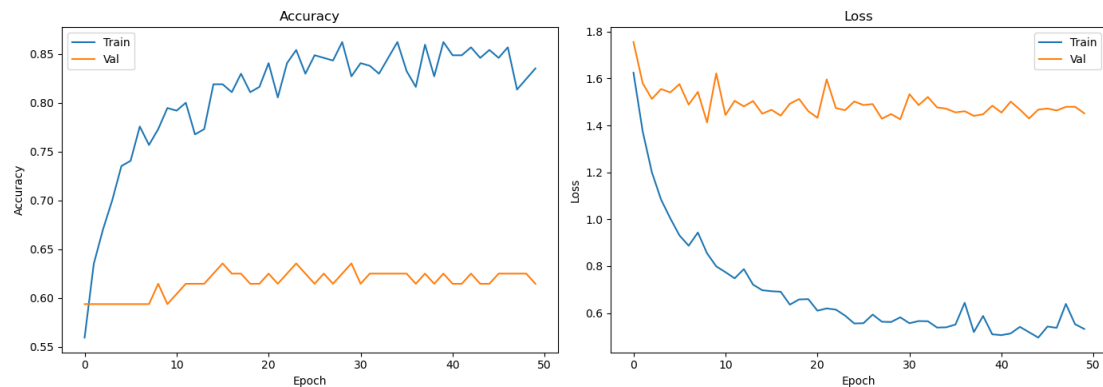
### 2. Data Augmentation: 0.6458



### 3. Early Stopping: 0.5938



#### 4. LR Scheduling: 0.6354



這次作業我選用了 MVTEC AD 資料集中的 cable 類別，並以 torchvision 中的 ResNet18 作為主幹模型，實作了四種提升分類準確率的策略。

##### 1. Fine-tuning 預訓練模型

我解凍了 ResNet18 的 layer4 和全連接層 fc，讓模型能針對資料集的高階特徵做調整。這個方法表現最好，驗證準確率達到了 **0.6979**。

##### 2. 資料增強 (Data Augmentation)

我加入了隨機旋轉、色彩抖動、水平翻轉等轉換，希望透過模擬製程中的變化來提升模型的泛化能力，最終驗證準確率為 **0.6458**。

##### 3. Early Stopping

在訓練過程中，我加入早停機制，只要驗證 loss 在連續幾個 epoch 沒有改善就會停止訓練。這雖然有助於避免過擬合，但驗證準確率只有 **0.5938**，表現偏弱。

##### 4. 學習率調整 (CosineAnnealingLR)

這個方法會讓學習率隨著訓練週期逐漸減小，讓模型在訓練後期收斂得更穩定，最後驗證準確率來到 **0.6354**。

這四種方法的訓練設定都一樣：最多訓練 50 輪，使用 Adam optimizer，learning rate 設為 0.001，batch size 是 32，loss function 則用 CrossEntropyLoss 並加入了 class weight 處理不平衡問題。

整體來說，微調模型參數的效果最好，可能是因為讓模型能專注學習 cable 缺陷的細節特徵，提升了分類能力。

3. (20 points) In real-world datasets, we often encounter long-tail distribution (or data imbalance). In MVTec AD dataset, you may observe that there are more images categorized under the 'Good' class compared to images for each defect class. (Approximately 150 words.) (i) (5 points) Define what is 'long-tail distribution.' (ii) (15 points) Identify and summarize a paper published after 2020 that proposes a solution to data imbalance. Explain how their method could be applied to our case.

(i)

長尾分布（long-tail distribution）指的是資料的分佈很不平均，像是某些主類別的樣本特別多，而其他類別的樣本卻很少。在 MVTec AD 資料集裡，good（正常）類別幾乎佔了大部分，像 bent\_wire 或 missing\_wire 這種缺陷類就相對稀少。這種情況下，如果不處理，模型會很容易偏向預測正常類別，忽略了那些重要但數量少的瑕疵類別。

(ii)

參考了 2022 年 CVPR 的一篇論文，叫做《Rethinking Class-Balanced Loss for Long-Tailed Visual Recognition》。這篇提出了一種叫做 Class-Balanced Loss 的方法，核心概念是：類別樣本數越少，給它的 loss 權重就越大。這樣模型在訓練時會「自動」更重視少數類別的學習，不會只顧著學會大宗類別。在實作上，我是用 `sklearn.utils.class_weight.compute_class_weight` 這個函數，根據 `y_train` 裡的分佈去計算每個類別的權重，然後把這些權重放進 `CrossEntropyLoss(weight=...)` 裡，讓模型在學習時對缺陷類別更敏感一點，效果還不錯，模型辨識瑕疵的能力明顯有提升。

4. (20 points) The MVTec AD dataset's training set primarily consists of 'good' images, lacking examples of defects. Discuss strategies for developing an anomaly detection model under these conditions. (Approximately 100 words.)

如果訓練資料幾乎都是良品，幾乎沒有缺陷圖，這種情況其實很常見，這時候我們就可以改用無監督或自監督的異常偵測方法。這些方法的核心概念就是：

「只學正常樣本」，讓模型自己去找出那些跟正常長得不一樣的地方。

幾種常見的做法像是：

- **AutoEncoder 或 VAE**：先讓模型學會如何重建正常圖片，再用「重建差異」來找出異常（比如某些區域重建得很差，就可能是缺陷）。
- **GAN 系列方法（像 f-AnoGAN）**：用生成對抗方式建模正常樣本的分布，測試時看一張圖合不合理就能判斷異常。
- **SOTA 模型 像 PatchCore、DRAEM、SPADE**，這些都是專門為 **MVTec** 設計的異常偵測方法，只需要正常樣本就可以訓練，效果也很好。

總之，這類方法特別適合資料量有限、又找不到缺陷圖的場景，比如我們這次作業這種狀況就很適合用！

5. For the task of anomaly detection, it may be advantageous to employ more sophisticated computer vision techniques such as object detection or segmentation. This approach will aid in identifying defects within the images more accurately. Furthermore, there are numerous open-source models designed for general applications that can be utilized for this purpose, including YOLO-World (website) and SAM (website). (Approximately 150 words.) (i) (10 points) To leverage these powerful models and fine-tune them using our dataset, it is necessary to prepare specific types of datasets. What kind of data should be prepared for object detection and for segmentation. (ii) (10 points) Why are these models suitable for fine-tuning for our custom dataset?

(i)

如果要用像 YOLO-World 或 SAM 這類比較進階的模型來微調的話，資料格式就要再更精細一點：

物件偵測（例如 YOLO-World）：

每張圖要有 bounding box 的標註，通常是 .txt 或 .json 格式。內容會像這樣：<class> <x\_center> <y\_center> <width> <height>，這些數值都是歸一化過的，表示物件在圖中的位置。

影像分割（例如 SAM）：

每張圖還要有一張 對應的 mask 圖（通常是黑白二值圖），來標出缺陷的區域。這些 mask 可以是 .png 圖片，也可以用像 COCO 格式的多邊形標註。

(ii)

像 MVTec AD 裡的 cable 缺陷，很多都是出現在局部區域，這時如果只用分類模型會看不到具體的缺陷位置。但如果用物件偵測或影像分割模型，就可以直接知道缺陷出現在哪裡、大小有多大。

以 YOLO-World 為例，它可以很準確地找出缺陷的位置跟類別；而像 SAM 這種模型甚至可以做到 pixel 等級的遮罩，讓我們看到哪一塊是異常，這對 AOI（自動光學檢測）或智慧工廠的應用非常有幫助。

所以這些模型真的很適合 MVTec 這種資料集，尤其是像 cable 這類有結構規則但容易出現細微瑕疵的產品。只要我們有對應的標註資料，就能做出很強的微調結果。