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.

我選用資料集中的 bottle 進行檢測，Bottle 資料及細節如下：

- 缺陷類別數量（Number of defect classes）： 3 種
- 缺陷類型（Types of defect classes）： broken_large、contamination、broken_small
- 總圖像數（Total images used）： 292 張
- 訓練與測試資料分佈（Distribution of training and test data）：
  - 訓練資料集（good）：209 張
  - 測試資料集（good + defective）：83 張
- 圖片尺寸（Image dimensions）： 900 × 900 × 3（RGB 彩色圖像）

```
Dataset Summary for 'bottle':
Number of defect classes: 3
Types of defect classes: ['broken_large', 'contamination', 'broken_small']
Total images used: 292
  - Training images (only 'good'): 209
  - Test images (good + defective): 83
Image dimensions: 900 x 900 x 3
```

2. (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.)

實驗編號	模型 (backbone)	批次大小	學習率	訓練回合數 (Epochs)	方法說明	驗證準確度
1	ResNet18	32	1e-3	30	基礎設定	84.51%
2	ResNet50	32	1e-3	30	更換模型為 ResNet50	90.14%
3	ResNet18	16	1e-3	50	減少批次大小並增加 epochs	85.92%
4	ResNet18	16	1e-4	50	減低學習率	83.10%

我針對 MVTec AD 資料集中 bottle 類別的影像，進行了四組實驗，以提升模型對於正常與瑕疵樣本的分類效能。第一組為基礎設定，使用 ResNet18、batch size 32、學習率 1e-3，驗證準確率為 84.51%。與第一組相比，第二組更換為預訓練的 ResNet50，其他設定相同，準確率顯著提升至 90.14%，顯示更深層的模型與轉移學習有助於學習更複雜的特徵。第三組在 ResNet18 架構下，將 batch size 減少至 16 並將訓練回合數提高至 50，準確率提升至 85.92%，說明增加訓練時間與較小 batch 有助模型泛化。第四組延續第三組設定，僅將學習率降至 1e-4，準確率反而降至 83.10%，顯示學習率過低可能導致收斂緩慢。整體而言，與第一組相比，選擇預訓練的 ResNet50 是提升效能最有效的方法，超參數調整則提供次要的微幅優化。

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) Long-tail distribution（長尾分布）指的是資料中的樣本數主要集中於少數幾個「主要類別」(head classes)，而其他大量類別（tail classes）則僅有極少數樣本。這會導致模型在訓練時偏向常見類別，進而影響對稀有類別的辨識效果與整體分類性能。

(ii) 近年來，CVPR 2022 發表的論文《Balanced Contrastive Learning for Long-Tailed Visual Recognition》，由 Zhou 等人提出了一種針對長尾分布資料問題的改進方法，稱為平衡對比式學習（Balanced Contrastive Learning, BCL）。這個方法的核心理念是將資料中的類別分為常見的「head 類別」與較少出現的「tail 類別」，並為這兩類別分別建立記憶機制，使用不同的對比學習策略，以避免在訓練過程中模型過度偏向 head 類別。

這個方法若應用在 MVTec AD 資料集上，可以將大量的正常樣本（good）視為 head 類別，而將三種缺陷類型視為 tail 類別。透過 BCL，可以幫助模型在學習過程中更好地保留並辨識少數缺陷類別的特徵，進一步提升模型對異常樣本的辨識能力，減少資料不平衡所帶來的影響。

Paper 網址：

https://openaccess.thecvf.com/content/CVPR2022/papers/Zhu_Balanced_Contrastive_Learning_for_Long-Tailed_Visual_Recognition_CVPR_2022_paper.pdf

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.)

MVTec AD 的訓練資料幾乎都是正常（good）圖像，缺陷樣本非常少，所以很適合用無監督或自監督的異常偵測方法。像 Autoencoder 或變分自編碼器（VAE）這類方法，可以先學習正常圖像的特徵分佈，之後再透過「重建誤差（reconstruction error）」來判斷新圖是不是異常。另一種方式是使用 PatchCore 這類記憶型方法，先記住正常圖的特徵，測試時就比較輸入圖的特徵跟記憶庫之間的距離，來判斷是否異常。這些方法都不需要用到缺陷的標註資料，對像 MVTec 這種幾乎只有正常樣本的情況來說非常實用。而且，相比起傳統的分類模型，這類方法通常有更好的泛化能力，能夠抓到那些模型從沒見過的異常狀況。

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)

- YOLO-World（物件偵測）：需為每張圖片標註缺陷的 bounding box（x, y, width, height）及對應類別。可儲存為 YOLO 格式（.txt）或 COCO 格式（.json）。
- SAM（語意分割）：每張圖需對應 pixel-wise mask，標註異常區域的精確形狀。可使用 PNG 遮罩圖，或轉為 COCO segmentation 格式。

(ii)這些模型具備優秀的特徵提取和空間感知能力，能有效針對圖像中的異常區域進行準確的定位與分割。即使只有少量標註的缺陷圖，也能透過 fine-tune 快速適應像 bottle 這類缺陷樣式多變、資料分布不均的情況。相比單純的分類模型，這些方法不僅能提升偵測的準確率，還能清楚指出異常位置，讓結果更具可解釋性，因此特別適合用在需要具體標出異常區域的任務上。