HW3

113034547
郭又瑜

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:

- Chosen dataset: bottle
- Number of defect classes: 3
- Types of defect classes: broken_small, broken_large, contamination
- Number of images used in your dataset: train: 48, val: 12
- Distribution of training and test data: train: 0.8, val: 0.2
- Image dimensions: (900, 900, 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.)

Hyperparameters used in each attempt: batch_size = 32, epochs = 50, lr = 0.001, optimizer = Adam, lose function = CrossEntropyLoss

- Attempt 1: (ResNet18 with unfrozen layer 4 and fully connected layer)
  Epoch 50/50, Train loss: 0.1190, Train acc: 95.8333%, Val loss: 0.5009, Val acc: 83.3333%, Best Val loss: 0.3365 Best Val acc: 83.33%
- Attempt 2: (EffficientNet B0)
  Epoch 50/50, Train loss: 0.7493, Train acc: 66.6667%, Val loss: 0.7660, Val acc: 58.3333%, Best Val loss: 0.7660 Best Val acc: 66.67%
- **Attempt 3: (Resize input images + ResNet50 with unfrozen layer 3, 4 and fully connected layer)**
  Epoch 50/50, Train loss: 0.0444, Train acc: 100.0000%, Val loss: 1.1309, Val acc: 66.6667%, Best Val loss: 0.5236 **Best Val acc: 91.67%**
- Attempt 4 : (DenseNet121)
  Epoch 50/50, Train loss: 0.5719, Train acc: 79.1667%, Val loss: 0.6634, Val acc: 75.0000%, Best Val loss: 0.6634 Best Val acc: 75.00%

Attempt 3 使用 ResNet50 並解凍 layer 3、layer 4 與 fc layer，搭配較高解析度的輸入圖片（128×128），使模型能有效學習缺陷位置的細微特徵。相較於完全凍結的模型，部分解凍允許深層特徵提取器根據資料分布進行微調，提升辨識力。同時採用 Cosine Annealing 學習率排程器，有助於模型穩定收斂。這些設定大幅改善驗證集表現，成功將準確率提升至 91.67%，顯示深層微調與輸入解析度是關鍵因素。

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.'

長尾分布指的是資料集中存在少數幾個類別樣本數非常多，而其他多數類別的樣本數極少的情況。在影像資料中，可能出現主流類別（如正常圖）佔據大多數資料，而稀有類別（如特定缺陷）僅有少量樣本。由於樣本數極不平均，傳統模型容易偏向預測出現頻率高的類別，導致少數類別的辨識準確率偏低，進而影響整體效能。

(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.

論文《Distributional Robustness Loss for Long-Tail Learning》，提出了一種新的損失函數，旨在提升深度學習模型在長尾分佈資料上的表現。 他們的研究指出，傳統的模型在處理樣本數較少的尾部類別時，往往學習到較差的特徵表示，導致分類性能下降。為了解決這個問題，作者引入了 Distributionally Robust Optimization (DRO) 的概念，設計了一種新的損失函數，能夠在考慮最壞情況的分佈下，學習更穩健的特徵表示。

參考論文

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.)

為了在缺乏缺陷樣本的情況下進行異常偵測，常見策略是使用重建式模型，訓練模型只會學習重建正常樣本的影像分布。當輸入異常影像時，由於重建效果差，可透過計算重建誤差判斷是否為異常。另一種常見方法是使用特徵對比模型，例如 PaDiM 或 FastFlow，這類模型會先從正常樣本中提取深度特徵並建立分布模型，測試時比較輸入圖像的特徵與正常樣本的特徵分佈差異，以此判斷異常程度。

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.

Object Detection：需要為每張影像建立標記檔，內容包含一個或多個 bounding boxes。每個 box 需標示缺陷區域的位置（x, y, width, height）及其對應類別（如：contamination、broken）。

Segmentation：需要提供每張影像的 pixel-level mask image，標示出異常的準確邊界。每個 mask 圖像可為單通道二值圖（單類缺陷）或彩色圖（多類別），與原始圖像同尺寸，像素值代表每個像素所屬的類別。

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

YOLO-World 和 SAM 都是針對實際應用設計的開源模型，具有良好的 transfer 能力與速度。YOLO 系列模型具備高速且準確的物件偵測能力，能快速標定缺陷區域，而 SAM 則具備強大的 zero-shot 分割能力，即使在訓練樣本有限的情況下，也能達到精細的區域分割效果。透過微調這些預訓練模型，即可使其適應 MVTec 資料集中不同類型的缺陷樣貌，提升異常定位與解釋能力。