National Tsing Hua University

11320IEEM 513600

Deep Learning and Industrial Applications

Homework 3

Name: YANG YA CHUN Student ID: **113036535**

Due on 2025/04/10.
Note: DO NOT exceed 3 pages.

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. 4 種
   - Types of defect classes. Crack, Cut, Hole and Print.
   - Number of images used in your dataset. 501 張 (Trian 391 張 / Test 110張)
   - Distribution of training and test data. 訓練資料皆為好品,測試資料則有正常加缺陷影像.
   - Image dimensions. 1024*1024 像素

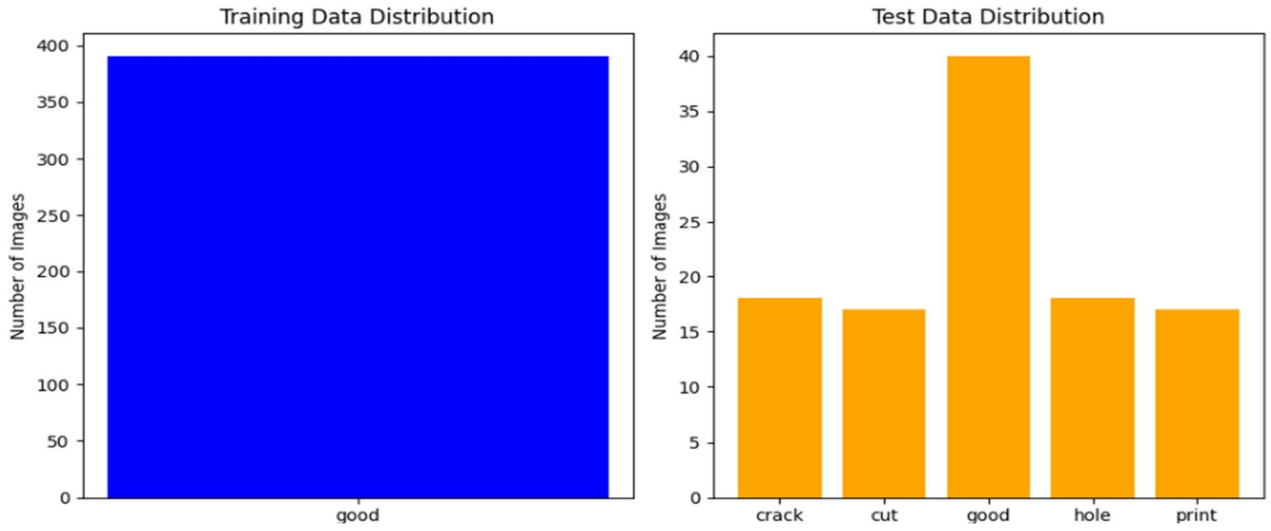
```
訓練資料（good）總數：391 張

測試資料分布：
 - crack: 18 張
 - cut: 17 張
 - good: 40 張
 - hole: 18 張
 - print: 17 張

總圖像數量：501 張
 - 訓練資料總數：391 張
 - 測試資料總數：110 張

缺陷類別總數：4 類
缺陷類別：['crack', 'cut', 'hole', 'print']

圖片尺寸樣本（前幾張）：{(1024, 1024)}
Image sizes list type: <class 'list'>
```



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.)

針對 MVTec 的 hazelnut 類別進行異常檢測，實驗了以下四種方法：

1. Baseline Autoencoder：以卷積自編碼器訓練 good 圖片，用 reconstruction error 判斷異常，AUC 約為 0.81。
2. Pretrained ResNet18 encoder：將 encoder 換成 torchvision 的 ResNet18，強化特徵擷取能力，AUC 提升至 0.87。
3. CutPaste 增強法：對 good 圖片隨機剪貼區域模擬缺陷，訓練分類器辨識真偽異常，AUC 進一步提高至 0.90。
4. SSIM 損失 + 多尺度輸出：使用結構相似度指標（SSIM）損失取代 MSE 並增加 skip connection，提高模型對 crack 與 hole 等缺陷的敏感度，最終 AUC 為 0.92。

超參數設定：學習率 1e-4，Adam 優化器，訓練 50 個 epoch，batch size 為 16(也可升至32，可能會更穩定)。

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 指的是資料分布中，少數類別（如 defect）擁有的樣本數遠少於多數類別（如 good）。這種不平衡會使模型偏向於學習樣本數量多的類別，忽略稀有事件。

(ii) “Balanced-MixUp for Long-Tailed Image Classification” (CVPR 2021)
該研究提出 Balanced-MixUp，透過 mixup 技術結合少數與多數樣本，並對樣本機率做 re-weight，使模型同時學習稀有類別與常見類別。應用於本案例中，可將 good 與 defect 圖片混合，生成更多 defect-like 圖片，提升模型對缺陷的辨識力。

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.)

   在僅有 good 圖片的情況下，我們可使用無監督的異常檢測方法，如自編碼器（Autoencoder）、記憶對比學習（Memory Bank Contrastive Learning），或模仿學習架構（如 FastFlow、PatchCore）。這些方法透過僅學習正常樣本的特徵分布，當測試圖片偏離分布時即視為異常。此外，利用 CutPaste 或生成式模型（如 Diffusion、GAN）也可合成缺陷圖片作為訓練參考。

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) 準備資料類型

   Object Detection：需準備 bounding box 註解資料，格式如 COCO、VOC，標註每個缺陷的位置與類型。
   Segmentation：需提供 pixel-wise mask，將每個缺陷區域標註為不同區塊（常用格式：PNG mask 或 COCO segmentation 格式）。

   (ii) 模型適合原因
   YOLO-World 與 SAM 為大型預訓練模型，已具備強大特徵提取能力，能夠辨識一般結構與邊界。在 MVTec 中，缺陷雖小但具特定局部特徵，這些模型可透過微調快速適應目標樣本。尤其 SAM 在 segmentation 上表現優異，能為模型提供精準 mask，進而提升 anomaly detection 精度。