HW3

110034009蕭禮英

Q1

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.

📄 Datasetset Description

- **Product Type:** **Bottle**
- **Number of Defect Classes:** 3
- **Types of Defect Classes:**
  - broken_large
  - broken_small
  - contamination
- **Number of Images:**
  - ground truth(遮罩圖片):
    - broken_large: 19張
    - broken_small: 21張
    - contamination: 20張
  - train(訓練集): 全部都是正常樣本，共 208 張
  - test(測試集):
    - good: 19
    - broken_large: 19
    - broken_small: 21
    - contamination: 20

- **Distribution of Training and Test Data:**
  - train: good only
  - test: good + defects(broken_large, broken_small, contamination)
  - ground_truth: broken_large, broken_small, contamination
- **Image Dimensions:**
  - 900 x 900 pixels（圖片大小統一）

Q2

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.)

訓練方式

因為訓練集資料中只有正樣本(good)，資料極度不平衡，根據Q3閱讀的論文，
我們可以用Algorithm-Level techniques和Hybrid Approachs來進行訓練

Algorithm-Level techniques	**方式一: Autoencoder**
	方式二: VAE
Hybrid Approachs	**方式三: ResNet18(Frozen) + One-Class SVM**
	方式四: ResNet18(Finetuning) + One-Class SVM
	方式五: ResNet50(Frozen) + One-Class SVM

訓練方式	Hyperparameter			Outcome			
	Epochs	Learning Rate	Optimizer	Train Loss	Train Acc	Test Loss	Test Acc
Autoencoder	123	1e-3	Adam	0.000301	0.9759	0.000360	0.9732
VAE	10	1e-3	Adam	146047.51 1	0.9271	76390.67	0.9077
ResNet18 (Frozen) + One-Class SVM	無	無	無	無	無	無	0.9277
ResNet18(Fine-Tuning) + One-Class SVM	21	1e-3	Adam	14.4045	0.6259	無	0.61
ResNet50(Frozen) + One-Class SVM	無	無	無	無	無	無	0.8675

Summarize

我們實作了五種不同的異常偵測方法，其中**Autoencoder 表現最佳**，推測可能原因有二：

1. **Autoencoder 是在「像素級別」進行學習與比對，比只用向量分類（像 One-Class SVM）更細緻。**
2. ResNet 是依賴 ImageNet 進行預訓練， 但 ImageNet 的貓貓狗狗圖，跟工廠瓶子圖片差很多

Autoencoder 是直接針對本次case進行學習： ✓ 避免 domain gap、 ✓ 對小規模、單一類別的資料集更友善

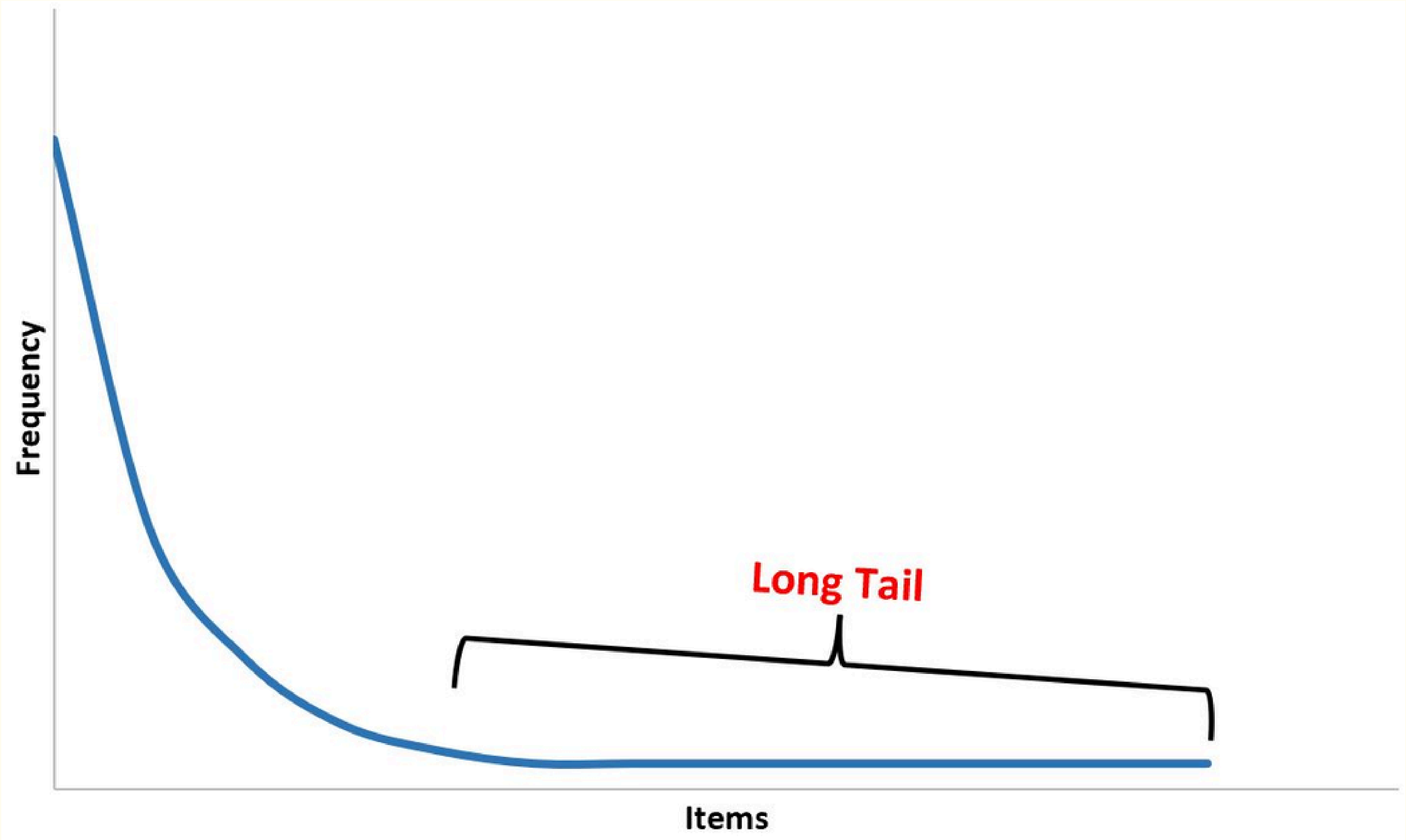
第二種方法 VAE 的準確率亦達 0.9077，但由於引入 KL 散度造成 loss 較高、模型較不穩定。
第三種方法，使用凍結參數的 ResNet18 搭配 One-Class SVM 得到 0.9277 的準確率，顯示預訓練模型萃取的特徵對異常分類有效。
第四種方法，微調 ResNet18 出現過擬合，測試準確率僅為 0.61。
第五種方法，使用凍結參數的 ResNet50 搭配 One-Class SVM 得到 0.8675 的準確率，可能因為神經網路架構較大導致過擬合，學習效果不如ResNet18。

Q3

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.

1. **Define what is longtail distribution**

In "long-tailed" distributions a high-frequency or high-amplitude population is followed by a low-frequency or low-amplitude population which gradually "tails off" asymptotically. The events at the far end of the tail have a very low probability of occurrence.



2. A Systematic Literature Review on AI Approaches To Address Data Imbalance In Machine Learning

Kutub Uddin Apu,+1 more - 06 Jan 2025

The main takeaways:

- **Prevalence of Data Imbalance:** Data imbalance is a common challenge in machine learning, leading to biased models and poor performance .
- **Diverse Strategies Identified:** The review categorizes various approaches to tackle data imbalance, including:
  - **Data-Level Methods:** Techniques like SMOTE (Synthetic Minority Over-sampling Technique) are widely used but can introduce noise and redundancy into the dataset.
  - **Algorithm-Level Techniques:** Cost-sensitive learning and ensemble methods show strong performance but require careful tuning of parameters and can be resource-intensive.

- **Hybrid Approaches:** Combining data-level and algorithm-level methods can enhance accuracy and adaptability, making them suitable for complex imbalance scenarios [1].
- **Advanced AI Techniques:** Methods such as Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), and deep learning architectures are powerful for high-dimensional datasets but face challenges like computational demands and risks of overfitting .

- **Foundation for Future Research:** By synthesizing insights from 92 peer-reviewed studies, the paper lays a groundwork for future research aimed at addressing the recurring challenges of data imbalance. It emphasizes the need for continued exploration of effective methodologies and the development of standardized metrics to facilitate better comparisons and advancements in the field.

3. How their method could be applied to our case?

- Data Level methods
  - ❌ 不適用，因為根本沒「瑕疵瓶子」資料，無法幫少數類別生成樣本。
- Algorithm-Level techniques
  - Autoencoder
  - 專門用來解「只給正常樣本」的分類問題，讓模型自己找出什麼是「異常」。
- Hybrid Approachs
  - ResNet18 + One-Class SVM
    - 先用ResNet18 把圖片壓縮成低維特徵，再用 One-Class SVM 判斷哪個落在異常區域。
  - GAN（生成對抗網路）變形：AnoGAN、f-AnoGAN
    - 用 GAN 學正常樣本，當生成器沒辦法重建某個圖像 → 那就代表「可能是異常圖像」。

Q4

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.)

在僅有正常樣本的情境下，本任務可視為一種「非監督式學習（Unsupervised Learning)」或「單類別學習（One-Class Learning)」問題。

此類問題適合採用**異常偵測模型（Anomaly Detection Model)**--透過學習「什麼是正常」，來識別與正常模式偏離的樣本作為潛在異常。

其中，**Autoencoder 或 VAE（Variational Autoencoder）**是常見且有效的非監督方法。這類模型僅需使用正常圖像進行訓練，**學習如何重建輸入圖像**。於測試階段，若模型面對異常樣本，將難以正確還原，進而產生較大的重建誤差，可作為判斷異常的依據。

此方法的最大優勢為不需提供異常樣本，即可有效發現潛在的異常區域。

又或者我們可以混合兩種以上方法(hybrid approach)來解決異常偵測問題。例如：

- ResNet18 + One-Class SVM：先用ResNet18 把圖片壓縮成低維特徵，再用 One-Class SVM 判斷哪個落在異常區域。
  - GAN（生成對抗網路）變形：AnoGAN、f-AnoGAN，用 GAN 學正常樣本，當生成器沒辦法重建某個圖像 → 那就代表「可能是異常圖像」。

Q5

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?

1. What kind of data should be prepared for object detection and segmentation?

項目	Object Detection（物件偵測）	Segmentation（圖像分割）
目標	找出「物件的位置」（用方框框起來）	精準標出「物件的形狀與區域」
標記方式	用「矩形框（Bounding Box）」標記物件位置	用「像素等級（Pixel-wise）」標記每一個像素屬於哪一類
標記資料內容	- 物件類別 - 方框座標（x, y, w, h）	- 每張圖對應一張 mask 圖 - 每個像素有類別標籤
標註格式	常見有 YOLO（.txt）、Pascal VOC（.xml）、COCO（.json）	mask 圖片（灰階圖或彩色圖），或對應的 label map（.png/.npy）

2．Why are these models suitable for fine-tuning for our custom dataset?

YOLO-World（物件偵測）與 SAM（Segment Anything Model，圖像分割）是基於大型通用圖像資料庫所預訓練的大型視覺模型，具備良好的泛化能力與視覺理解能力。

這些模型已具備辨識圖像中不同物件、區域與邊界的能力，只需搭配少量的自訂資料進行微調（fine-tuning），就能快速適應特定應用場景，例如瓶身瑕疵分類。

此外，這些模型支援開源架構、訓練流程成熟，並可與主流訓練框架（如 PyTorch、Detectron2）整合，大幅降低資料需求與開發門檻。因此，非常適合用於資料量有限、需快速上手的實際工業檢測應用。

THANK YOU