National Tsing Hua University
11220IEEM 513600
Deep Learning and Industrial Applications
Homework 3

Name:　　陳建宏　　　　　　　　Student ID: 111003807

Due on 2024/04/11.
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:

    Bottle dataset
    • Number of　classes. : 4 classes
    • Types of classes. (broken_large, broken_small, contamination,good)
    • Number of images used in your dataset. 4
    • Distribution of training and test data. Training80% test20%
    • Image dimensions. 900*900
    https://github.com/CCHung1981/nthu_2024_dlia_hw.git

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.)

	epochs	learning rate	Batch_size	Test accuracy
原始參數	50	1e-3	32	50%
方案 1	100	1e-3	32	50%
方案 2	100	0.5e-3	32	37.5%
方案 3	100	0.5e-3	16	62.5%
方案 4 ResNet-50	100	0.5e-3	16	87.5%

　　最初的模型在使用 50 個訓練週期和 1e-3 學習率以及 32 的批量大小的基本設置下，達到了 50%的測試準確率。第一次改進嘗試將訓練週期增加到 100，測試準確率維持 50%。第二次嘗試將學習率降低到 0.5e-3，並保持

相同的週期和批量大小，準確率下降到了 37.5%。第三次嘗試保持改進後的學習率不變，同時將批量大小減少到 16，準確率提升至在 62.5%。最後一次也是最成功的嘗試涉及了修改來自 TorchVision 的預訓練 ResNet-50 模型，同時保持 0.5e-3 的學習率和 16 的批量大小。這一變化顯著提高了模型的性能，達到了 87.5%的測試準確率。

最佳表現模型成功的關鍵因素是結合了預訓練的 ResNet-50 架構，它由於其深度和複雜性可能引入了更精細的特徵檢測器，以及較低學習率和較小批量大小的最佳組合，這可能允許在訓練期間進行更精緻的權重更新。

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.'

這通常指的是少數類別（"頭部"）擁有大量觀察，而大多數類別（"長尾"）只有相對少數的觀察。例如，在一個產品評價的數據集中，少數產品可能會收到大量評價，而大多數只有幾個。在機器學習中，長尾分佈可能會構成挑戰，因為在這樣的數據上訓練的模型可能會偏向於更頻繁出現的類別，導致對不常見類別的性能較差。

(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.

Simplifying Neural Network Training Under Class Imbalance,2023

這篇文章介紹了針對不平衡數據集的一系列優化方法，包括聯合自監督學習（Joint-SSL）、梯度累積（SAM）和標籤平滑，這些原本設計於平衡數據集的方法經改良後更適合不平衡數據的訓練。研究顯示，在高度不平衡的環境中，較小的批次大小有益於減少對多數類別的過適配，並且數據增強對於少數類的改善尤為顯著。這項研究強調了選擇適當的數據增強策略的重要性，並發現在不同類別不平衡水平上，最佳的增強策略可能有所不同。將這些發現應用到 MVTec AD 數據集可能包括採用不同的批量大小和數據增強方法，以及選擇適當的模型預訓練策略，以平衡數據並提高對少見缺陷類別的識別能力。

3. (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.)

在以"良好"圖像為主、缺陷範例不足的 MVTec AD 數據集中開發異常檢測模型時，可以考慮以下策略：使用自監督學習，讓模型學習區分"良好"圖像中的正常特徵。再採用生成模型，如 GAN，來合成少量的缺陷圖像，增加訓練數據的多樣性。此外，實施異常檢測特定的模型架構或損失函數，如使用自編碼器檢測圖像中的異常。最後，可以考慮一些數據增強技術，以提高模型對罕見或未見過的缺陷的識別能力。

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.

為了利用這些強大的模型並用我們的數據集進行微調，需要準備特定類型的數據。針對物體檢測，應該準備包含標記物體位置（如邊界框）的標註數據；對於分割任務，則需要每個像素的類別標註，通常以圖像掩碼的形式提供，用於區分不同物體或區域。

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

這些模型適合用於自定義數據集的微調，因為它們已經在廣泛和多樣化的數據上預訓練過，學習到了強大的特徵表示。這意味著它們能夠較好地識別和理解各種圖像特徵，即使是在新的應用場景中。通過微調，這些模型可以進一步調整以專門針對特定的數據集特點和任務要求，從而提高在異常檢測等特定任務上的性能。