

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

Name: 楊佳蓁

Student ID: 109034048

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:

我選擇的是hazelnut，以下是該資料集的資訊

- Number of classes. : 5
- Types of classes. : hole 、 print 、 cut 、 crack 、 good
- Number of images used in your dataset. : 50
- Distribution of training and test data. : training set 為每項類別資料的80%(共40筆) ， test set為剩下的20%(共10筆)
- Image dimensions. : (1024,1024,3)

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.)

原本的validation accuracy為20% 、 validation loss為1.6444，我調整了epoch、pre-trained model、batch size以及training set和test set resize的數值，以下為每次變動一項參數時performance的改變。

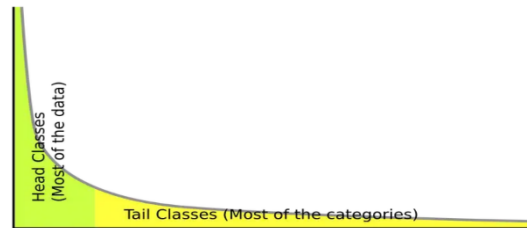
| Hyperparameter | 調整前/後 | validation accuracy | validation loss |
|-------------------|---------------------------|---------------------|-----------------|
| epoch | 50 / 70 | 40% | 1.5540 |
| pre-trained model | renet18 / mobilenet_v2 | 60% | 1.6101 |
| batch size | 32 / 16 | 60% | 1.2897 |
| resize | (32,32) / (128,128) | 70% | 1.1782 |

這些改變當中，調整resize能產生最好的效益，我認為應該是resize的數值太小會讓model失去一些重要資訊，放大後才能捕捉更多圖片資訊。不過其實四項參數同時調整才能讓model有最好的performance，其validation accuracy為80%、validation loss為0.6703，更重要的是這樣模型每次訓練的表現會更穩健，因此作業繳交的ipynb檔為四項參數同時調整的結果。

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.'

long-tail distribution代表在資料當中，少數類別佔大多數，而其他多數類別的資料較少，如下圖。



(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.

Paper: <https://arxiv.org/ftp/arxiv/papers/2108/2108.00071.pdf> (2021)

這篇paper提出解決data imbalance的問題，主要方法包刮「增加少數類別的樣本數」及「減少多數類別的樣本數」，前者例如使用SMOTE(oversampling的一種)基於少數資料合成相似的新樣本，或是利用GAN生成更擬真的資料；後者例如各種undersampling，最簡單的就是random undersampling隨機刪除多數類別的資料。如果以本次作業為例，將各項類別的資料都縮減為10筆以處理data imbalance就是一種undersampling，但如果不想要浪費辛苦蒐集資料，可以花點時間利用前面提到的方法合成少數類別的資料。

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.)

要處理dataset資料不平衡的問題，除了第三題提到改變樣本的方式(「增加少數類別的樣本數」或「減少多數類別的樣本數」)之外，也可以將F1 score作為評估指標，F1 score將precision和recall合併為單一指標，即使資料不平衡仍然可以提供全面的評估，較不會受到資料中某項類別數量過多或過少的影響。

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.

對於object detection，資料需要標註每個圖像中物體的位置(包含邊界框)和類別。

對於segmentation，資料需標記mask，以指定每個像素的類別，使其能夠在像素級別上劃分物體邊界。

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

這些model已經在大型數據集上訓練過，具有較好的performance與generalization的能力。另外，訓練的資料及權重等相關資訊也都公開透明，可以較容易選擇適合的model。