National Tsing Hua University

11220IEEM 513600

Deep Learning and Industrial Applications

Homework 3

Name: **姚國慶** Student ID: **112003802**

1. (10 points) Download the MVTec Anomaly Detection Dataset from Kaggle. Select one type of product from the dataset. Document the following details about your dataset:
   - Number of defect classes.
     Ans. 5 classes come from metal_nut
   - Types of defect classes.
     Ans. good, bent, color, flip, scratch
   - Number of images used in your dataset.
     Ans.115 sheets
   - Distribution of training and test data.
     Ans.設計三種分布：
     (1) training 40pcs, test 10pcs
     (2) training 80pcs, test 20pcs
     (3) training 90pcs, test 10pcs
   - Image dimensions.
     Ans. Pixels 700x700

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.)

Ans. 測試4種方式, (1) Pre-trained model. (2)train/val.分配數量. (3) batch size. (4)epoch，共測試18個試驗，結果如下表。

Test#	(1) pre-train model	(2) the numbers of dataset		(3) batch size	(4) epochs	acc. (3 times avg.)	備註
		train	val				
Test1	resnet18(weights='IMAGENET1K_V1')	40	10	8	50	63%	
Test2				16	50	70%	
Test3				32	50	28%	
Test4		80	20	16	50	54%	
Test5				32	50	53%	
Test6		90	10	16	50	76%	
Test7				16	100	80%	epoch 50之前就已收斂
Test8				32	100	63%	epoch 50之前就已收斂
Test9	resnet50(weights='IMAGENET1K_V1')	40	10	16	50	60%	
Test10		80	20	16	50	58%	
Test11				32	50	65%	
Test12	resnet50(weights='IMAGENET1K_V2')	40	10	16	50	40%	
Test13		80	20	16	50	40%	
Test14		90	10	16	50	40%	
Test15				32	50	45%	
Test16	resnet101(weights='IMAGENET1K_V1')	90	10	16	100	60%	epoch 50之前就已收斂
Test17				16	50	70%	
Test18				32	50	60%	

(1) Pre-trained model: 以resnet18表現最好搭配train90/val.10,batch size 16,epoch50的acc. 76%~80%最佳，更多層數的CNN模型如resrnet50, resnet101的表現反而不好, acc.在40~70%。

(2) Train/val.分配，以90/10的表現最好,有較多的data可供training時模型表現較好

(3) Batch size以16表現最好。

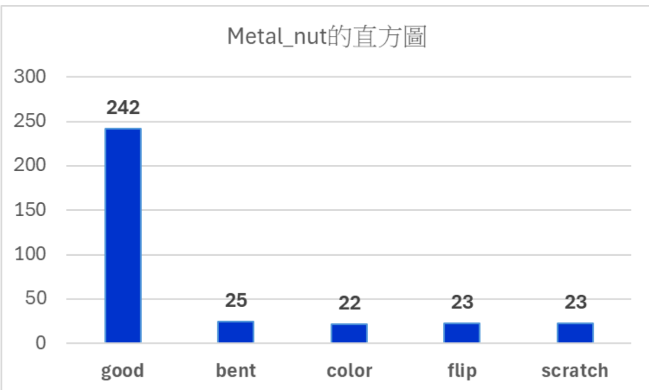
(4) Epoch幾乎在50之前便已收斂，更多的epoch沒有幫助。

以上測試的leaning rate皆為0.001, optimizer為Adam。

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.

Ans.(i)

長尾分佈，是指一種統計分佈，其中大部分的數據點集中在分佈的“頭部”，代表常見或頻繁發生的事件，而其餘的數據點則延伸到“尾部”，代表較不常見或罕見的事件。"長尾"一詞來自於這種分佈的圖形表示，其中尾部在x軸的右側延伸得很長。如本次程式作業metal_nut的類別分布如下圖，



Ans. (i i)

"Learning Imbalanced Datasets with Label-Distribution-Aware Margin Loss" by Huang et al., published in the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) in 2020.此論文提出了一種創新的損失函數，Label-Distribution-Aware Margin Loss（LDAM），用於解決瑕疵分類中的數據不平衡問題。傳統的損失函數，如交叉熵損失，無差別對待所有類別，可能導致模型偏向多數(頭部)類別的學習。LDAM將類別分布納入損失計算，稱為Label-Distribution-Aware Margin Loss，此損失函數能夠根據類別數據集中分佈情形進行調整，懲罰對少數(尾部)類別的預測損失，相對鼓勵模型更專注於從少數類別中學習，從而減輕數據不平衡的影響。在MVTec AD數據集，其中'Good'類的圖像比缺陷類別更多，可以應用LDAM來解決這個不平衡問題。使用LDAM作為損失函數，模型可以更關注缺陷類別，確保它們不會被多數'Good'類圖像過度訓練。

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.)

Ans.舉例2種策略改善MVTec AD資料集資料不平衡的問題：

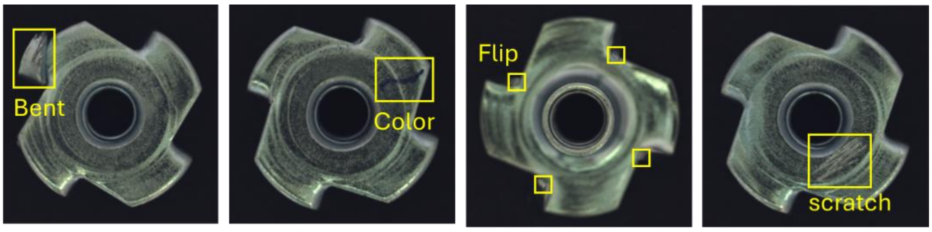
(1) Transfer Learning: 利用在open source如ImageNet-pretrained models的pre-trained模型來得到初始化模型的權重，再以MVTec AD不平衡數據集fine tuning模型權重，注意要修改pretrained model的輸出類別數量以吻合MVTec AD資料集的類別數量。

(2) 類似face recognition，利用siamese network，預先訓練好大量的'good'影像(捨去訓練較少量的anomaly影像)並encoding，將待測影像encoding後與資料集中的’good’影像code比對相似度，設定一個相似度參數，小於此相似度則為anomaly。

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 and SAM. (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?

Ans. (i)

使用Object detection方法，須事先將datasets中的要辨識的目標物件以bounding box標示出來，例如本作業MVTec AD metal nut資料集的defect:bent、color、flip、scratch等，如下圖示：



Ans. (ii)以 object detection or segmentation主有以下兩個好處：

(1) Transfer Learning: 這些model通常有很多pre-train model可以在open source上找到，例如ImageNet，參數已經訓練完成，也被實際測試有非常好的模型準確率，只要輸入使用者的任務資料集依需求做指定層參數的fine tune即可，另外這些pre-train model也有彈性讓使用者更改任務識別的class數。

(2) Efficiency: Yolo模型結構以sliding window detection的方式相對傳統CNN模型快速，適合做real-time或運算資源受限的應用。