

Homework 3

學號：113034510 姓名：蘇謹

- 選擇的 product 為 capsule，以下是 dataset 的細節：
 - Number of defect classes：6
 - Types of defect classes：'crack', 'faulty_imprint', 'good', 'poke', 'scratch', 'squeeze'
 - Number of images：120 (20 images * 6 classes)
 - Distribution of training and test data：96 training (80%), 24 test (20%)
 - Image dimensions：3 * 1000 * 1000
- 由於原來的模型都使用 Epochs=50，我認為不能很好的觀察訓練得收斂程度，所以統一設定以下實驗的固定參數為：Epochs = 200、Learning rate = 0.001、Optimizer= Adam、LR_scheduler = CosineAnnealingLR。接下來針對 Model、Weight、Batch Size、Transformer Resize、Transformer Normalize 做超參數實驗如下：

Original Model Result

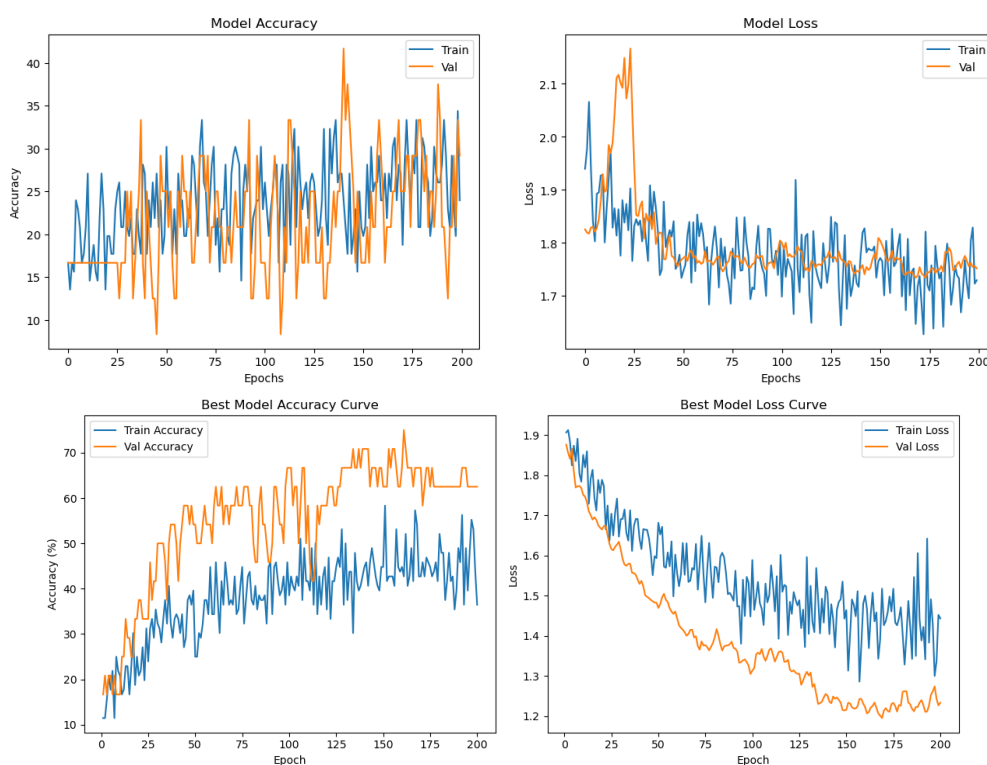
Model	Weight	Batch Size	Resize	Normalize	Train Loss	Train Acc(%)	Val Loss	Val Acc(%)	Test Acc(%)
ResNet18	IMAGENET1K_V1	32	32	Img/255	1.730	23.96	1.752	29.17	41.67

Experiment Result

Model	Weight	Batch Size	Resize	Normalize	Train Loss	Train Acc(%)	Val Loss	Val Acc(%)	Test Acc(%)
EfficientNet_v2_l	IMAGENET1K_V1	4	64	No	2.120	22.92	2.785	12.5	37.5
EfficientNet_v2_l	IMAGENET1K_V1	16	64	No	2.139	18.75	2.042	8.33	37.5
EfficientNet_v2_l	IMAGENET1K_V1	32	64	No	1.901	17.71	1.905	8.33	45.83
EfficientNet_v2_l	IMAGENET1K_V1	64	64	No	1.903	18.75	1.842	16.67	25
EfficientNet_v2_l	IMAGENET1K_V1	32	128	No	1.785	28.13	1.807	25	25
EfficientNet_v2_l	IMAGENET1K_V1	32	128	Yes	1.779	29.17	1.736	37.5	37.5
EfficientNet_v2_l	IMAGENET1K_V1	32	256	Yes	1.569	35.42	1.665	29.17	41.67
EfficientNet_v2_l	IMAGENET1K_V1	32	512	Yes	1.443	36.46	1.233	62.5	75
ResNet18	IMAGENET1K_V1	32	512	Yes	1.655	32.292	1.682	37.5	45.83

首先，Original Model Result 代表原始模型設定的結果，其中 `Img/255` 是訓練過程中使用的標準化手法，可以發現在這組參數設定下效果不彰 (Test Acc = 41.67%)。

所以接下來的 Experiment Result 中，第一步改動是增加 Transformer 的手法 (Random Rotation、Color Jitter) 並且換 Pretrained Model = EfficientNet_v2_l 以及做 Batch Size 調整，整體 Test Acc 有效提升，另外可以觀察到適合的 **Batch Size 選擇為 32**，有些微提升正確率。第二步是探討 Resize 的大小選擇，發現 `Resize = 512` 是最佳選擇，而且可以注意到 **Resize 對 Test Acc 影響十分關鍵**，畢竟照片解析度高一點才能更好的學習特徵。此外，我將標準化改用 **Transformer Normalize**，發現效果有小幅提升。最後，比較 ResNet18、EfficientNet_v2_l，使用 EfficientNet_v2_l 有更好的 Test Acc，代表更換 **Pretrained Model = EfficientNet_v2_l** 是很重要的因素。以下為原始模型 (上圖) 與最佳模型 (下圖) 的 Acc、Loss 比較：



3.

- (i). Long-tail distribution 也就是資料不平衡，是指資料中大多數樣本集中於少數或單一的 classes 中，而其他 classes 的樣本數量相較來說少很多，分佈於尾部，造成類別間極度不平衡。

- (ii). 在《Class-Imbalanced Semi-Supervised Learning with Adaptive Thresholding》論文中，提出了一種簡單有效的方法來處理資料不平衡的半監督學習問題，主要是根據每個類別的樣本數量，自適應調整未標記資料的 **pseudo-labels** 信心閾值。也就是對於樣本較少的類別，降低信心閾值，以增加該類別的 **pseudo-labels** 數量；反之，提高信心閾值。由此有助於平衡並提升模型在不平衡資料集上的表現。若使用這個方法於 MVtec AD，可以根據那些需要注意的缺陷類別之樣本稀少程度，降低 **pseudo-labels** 的信心閾值，增加缺陷類別的 **pseudo-labels** 數量，平衡資料分佈，這有助於模型更有效地學習缺陷類別的特徵，提升異常檢測的準確性。

(Reference : Lan-Zhe Guo, Yu-Feng Li Proceedings of the 39th International Conference on Machine Learning, PMLR 162:8082-8094, 2022. <https://proceedings.mlr.press/v162/guo22e>)

- 4. 對於訓練集大多為 'Good' 圖像，缺乏缺陷樣本，因此我覺得可以改為採用無監督或自監督的異常檢測方法。例如利用 GAN 這種的生成對抗網路來學習正常樣本的分佈，並在測試時比較輸入與重建結果的差異來偵測異常。此外，利用自監督學習生成正樣本的增強版本作為訓練目標，也能提升模型的異常偵測能力。

5.

- (i). 對於 **object detection**，需要準備每張圖像中所有缺陷區域的 **bounding boxes** 和與其對應的類別標籤。資料格式如 COCO 或 VOC 格式。對於 **segmentation**，需提供每張圖像中每個 **pixel** 的分類標籤，也就是缺陷區域的 **mask**，如同 MVtec AD 中 **ground truth** 資料中的圖樣，可使用 PNG 或 JSON 等格式。
- (ii). YOLO-World 和 SAM 是具高度泛化能力的預訓練模型，已在大量的資料上學習到足夠的特徵，所以使這些預訓練模型能在樣本數量有限的情況下透過 **fine tune** 快速地去適應我們想要的特定目標。運用 **fine tune** 的技術，能夠達成保留其強大特徵提取能力的同時，更使其專注微調於我們想要在資料集中關注的特定缺陷類型或目標，從而達到更精準且高效率開發的異常偵測或定位效果。