113030507\_HW3

#1.

本次作業選擇 transistor 資料集進行分析，有關資料集的各項資訊如下圖所示 :

* 缺陷類別的數量 : 4
* 缺陷類別 : bent\_lead、cut\_lead、damaged\_case、misplaced
* 所使用的圖片總數 : 313
* 訓練與測試資料的分配情況
  + Train distribution: {'good': 213}
  + Test distribution: {'misplaced': 10, 'good': 60, 'damaged\_case': 10, 'bent\_lead': 10, 'cut\_lead': 10}
* 圖片的尺寸 : 1024 x 1024

#2.

本次實驗主要針對下列參數與方法進行調整，包括 Batch\_size、epoch、NN Model、正規化、資料不平衡處理等，其餘未提及的參數如learning rate 與 optimizer 皆與 sample code 相同。而表格以下將分別針對各組實驗序號進行結果分析，包含訓練過程的 loss/accuracy 之趨勢變化。

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 實驗  序號 | Batch\_size | epoch | NN Model | 正規化  (拿掉/255) | 處理資料不平衡(Criterion) | Test Accuracy |
| 0 (原本) | 32 | 50 | ResNet18 | 無 | nn.CrossEntropyLoss() | 0.4 |
| 1 | 32 | 200 | ResNet18 | 無 | nn.CrossEntropyLoss() | 0.6 |
| 2 | 16 | 50 | ResNet18 | 無 | nn.CrossEntropyLoss() | 0.6 |
| 3 | 16 | 200 | ResNet50 | 無 | nn.CrossEntropyLoss() | 0.7 |
| 4 | 16 | 200 | ResNet18 | 有 | nn.CrossEntropyLoss() | 0.8 |
| 5 | 16 | 200 | ResNet18 | 有 | FocalLoss(gamma=2.0, alpha=0.25) | 0.8 |

(1)

實驗 1 中，將 epoch 從原本的 50 提升至 200，測試準確率從 40% 明顯提升至 60%。從訓練與驗證的 loss/accuracy 曲線可觀察到，模型在 50 個 epoch 時尚未收斂，驗證準確率波動且學習效果不穩定。延長訓練時間後，loss 持續下降且趨勢更平滑，顯示模型逐步學會了資料中的特徵模式。由此可見，適當增加訓練次數能夠有效提升模型的表現，尤其在資料量不大但模型較深的情況下特別重要。

一張含有 螢幕擷取畫面, 繪圖, 圖表, 行 的圖片

AI 產生的內容可能不正確。

(2)

實驗 2 採用 ResNet18 搭配較小的 batch size(16) 與 50 個 epoch。從圖表可觀察到，模型雖能穩定降低訓練損失，但驗證損失震盪劇烈、準確率長時間維持低點，直到後段才開始上升。這顯示模型在前期容易發生過擬合與泛化困難，可能原因包括 batch 波動性增加與訓練輪數不足。最終模型測試準確率為 60%，雖較原始模型提升，但整體表現仍受限於學習時長與前處理配置。延長訓練時間或改善輸入正規化方式可望進一步提升效能。

一張含有 文字, 行, 繪圖, 圖表 的圖片

AI 產生的內容可能不正確。

(3)

對於像 MVTec 這種缺陷檢測資料集，許多細節 (如裂痕、變形、微小的材質變化) 需要更深層的感受野 (Receptive Field) 與非線性處理能力來捕捉，這正是 ResNet50 的優勢。實驗 3 將模型從 ResNet18 替換為 ResNet50，使其擁有更深層的結構與更強大的特徵擷取能力。ResNet50 採用了 bottleneck 區塊與更多的卷積層，使其能夠捕捉到缺陷圖像中細微的紋理與形狀變化。這項結構升級帶來了顯著的效能提升，驗證準確率從原本的 60% 上升至超過 70%，顯示出在精細化異常分類任務中，深層網路結構具有明顯優勢。

一張含有 螢幕擷取畫面, 繪圖, 行, 文字 的圖片

AI 產生的內容可能不正確。

(4)

在前處理階段，我們原先在應用了 transforms.ToTensor() 之後，又手動將影像除以 /255.0。然而，ToTensor() 本身就已經將像素值從 [0, 255] 正規化為 [0.0, 1.0]，這使得額外的除法操作，意外地將像素強度壓縮到 [0.0, 0.0039] 的極小範圍。這種過度正規化大幅降低了影像特徵的清晰度，進而影響模型的學習與判斷能力。因此，實驗 4 在我們移除多餘的除法後，模型的驗證準確率從約 60% 明顯提升至 80%，證明了**正確的影像正規化流程對模型效能的關鍵影響**。此外，在這裡沒有沿用方法三的 restnet50 是因為測試結果僅達到 test\_acc=0.7，不及使用 restnet18。

一張含有 螢幕擷取畫面, 繪圖, 文字, 行 的圖片

AI 產生的內容可能不正確。

(5)

實驗 5 為了解決資料集中「正常類」與「缺陷類」樣本分布不均的問題，我們實驗性地將損失函數由 CrossEntropyLoss 改為 Focal Loss。Focal Loss 能夠降低對簡單樣本的過度學習，進而讓模型更聚焦在少數且難以辨識的異常類別上。雖然整體驗證準確率與原本相近，皆維持在 80%，但這個改動有助於模型在處理不平衡資料時更穩健學習，也提升了不同類別間的預測平衡性。這項調整在精細分類任務中具有潛在價值，特別適用於異常檢測或醫療影像等情境。

一張含有 螢幕擷取畫面, 繪圖, 文字, 行 的圖片

AI 產生的內容可能不正確。

#3.

(1)

長尾分布指的是資料集中，大多數樣本集中在少數幾個「主類別」，而剩下大量「稀有類別」的樣本數極少，呈現類似長尾的統計分布曲線。這會造成訓練模型時偏向主類別，忽略了尾端類別的學習，導致嚴重的類別不平衡問題。

(2)

* 論文名稱：《Balanced Contrastive Learning for Long-Tailed Visual Recognition》(Zhu et al., CVPR 2022)
* 該研究針對類別不平衡問題，提出了改良的對比學習方法 Balanced Contrastive Learning (BCL)，透過「類別平均」與「類別補全」策略，讓每個類別在特徵學習中獲得相對均衡的學習權重。此方法能有效修正長尾資料造成的表示偏差，並與交叉熵分類器結合，強化尾端類別的辨識效果。在本任務中，我們可將 BCL 應用於 MVTec 資料集中，使模型在訓練時同時學會「正常」與「缺陷」類別的穩定表示，有效提升異常類別的辨識準確率。

#4.

由於 MVTec AD 的訓練集多為「正常品」，可採用無監督或自監督學習方法建立異常檢測模型。常見策略如：訓練自編碼器 (Autoencoder) 或變分自編碼器(VAE) 重建正常樣本，異常品將呈現較高重建誤差；或使用對比學習 (如 CutPaste、DRAEM)，透過人工生成擾動模擬缺陷樣本，讓模型學會區分正常與異常特徵。此類方法不依賴標記的瑕疵樣本，特別適合資料不平衡的情境。

#5.

(1)

針對物件偵測模型 (如 YOLO-World)，需要準備包含 「邊界框 (bounding boxes)」與對應類別標籤的標註資料，格式如 COCO、VOC 等。例如：標記每張圖片中缺陷的位置與類型。而對於 語意分割模型 (如 SAM)，則需為每張圖片提供「像素級的遮罩 (mask)」標註，即將缺陷區域以不同顏色或數字標註出來。這類資料可使用工具如 Labelme 或 CVAT 製作，常見格式為 PNG mask 或 JSON 格式。

(2)

YOLO-World 與 SAM 都是基於大型通用資料訓練的預訓練模型，具備強大的視覺表徵能力，能在少量資料下快速遷移學習。YOLO-World 支援多語意目標偵測，適合用來找出缺陷的區域與類別；而 SAM (Segment Anything Model) 專精於圖像分割，能準確描出異常區域的輪廓。這些模型透過微調後可提升缺陷定位與解釋性，特別適合應用於如 MVTec AD 這種需要高精度異常檢測的任務