

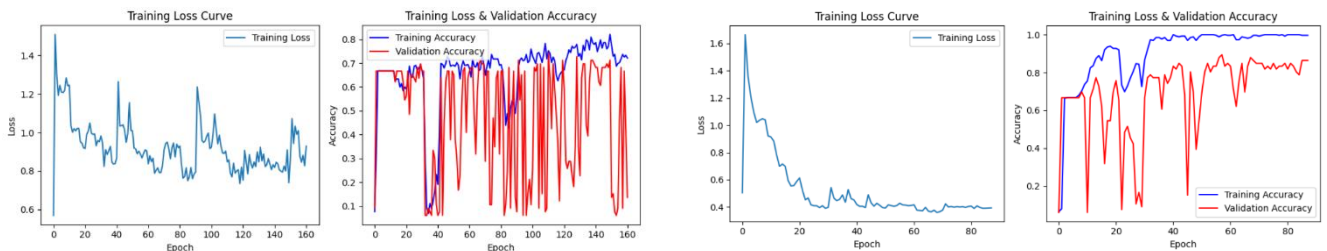
深度學習 Homework 2

1. 請分別使用「自建的 ResNet」與「有預訓練過的 ResNet」對資料集進行分類，並比較兩者的結果。接著用相同流程比較其他的預訓練模型與有預訓練過的 ResNet。

(3) (10pt) 比較自己建立的 ResNet 與預訓練 ResNet 的表現：

Pre-train ResNet 的 loss 下降速度較快，曲線也較平滑，顯示模型更穩定，且僅用 66 個 epochs 就達到最佳表現，而自建模型則要訓練到 131 個 epochs 才穩定，驗證表現也更穩定；自建模型的 loss 波動較大，也更容易 overfitting。

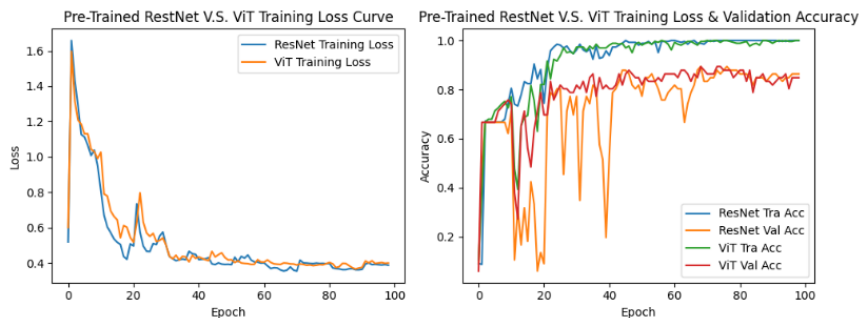
	自建 ResNet(左藍、右藍橘)	預訓練 ResNet(左橘、右綠紅)
訓練速度	24m 8.8s (161 epochs)	16m 19.2s (96 epochs)
訓練損失	0.8243	0.4083
收斂時機	131 epoch	66 epoch
準確率	0.7424	0.8636
可正確分類之異常類別(不分到 good)	faulty_imprint、squeeze	皆有學習到



(4) (25pt) 比較不同預訓練模型的表現

	預訓練 ResNet	預訓練 Vision Transformer
訓練速度	16m 19.2s (96 epochs)	17m 1.8s (99 epochs)
訓練損失	0.4083	0.3954
收斂時機	66 epoch	69 epoch
準確率	0.8636	0.8939

ViT 表現較佳，不僅準確率較高且 loss 也更低、更穩定。ViT val 準確率在約 2~30 epoch 之後變化趨緩並穩定上升，顯示泛化能力較好；ResNet 則有較大的震盪，表示模型對資料較敏感。而 Training loss 顯示兩者收斂速度相近，ViT 初期較高但後期下降更為穩定，較能避免 overfitting。



2. 在實際的製造環境中，異常樣本通常非常稀少且難以蒐集，對樣本進行標註也需要仰賴專業知識，並伴隨高昂的人力與時間成本。因此，要取得完整且多樣的異常資料相當困難。

(1) (5pt) 在上述情境中無法像一般分類任務那樣收集足夠的異常樣本，請問此時應採用何種分類設定來訓練模型？（提示：老師上課有講過）

可採用 Anomaly Detection 或 One-Class Classification 的概念來設定。這類方式使用大量的 typical (normal) data 進行訓練，先讓模型學習正常樣本的分布，日後遇到偏離正常分布 (high reconstruction error) 的樣本即可判定為異常。

(2) (10pt) 承上題，請說明在資料集中僅包含大量正常樣本而幾乎沒有異常樣本的情況下，瑕疵檢測模型通常會採取什麼樣的訓練策略？

1. 只使用正常樣本進行訓練的策略；讓 Autoencoder 會壓縮與重建正常樣本的特徵。其訓練目標為：讓模型學會「正常樣本的 reconstruction error」，如果未來的 reconstruction error 較大，表示該樣本偏離正常樣本，即代表可能是異常。
2. 使用 Self-Supervised Learning 方式進行特徵表示學習。

(3) (10pt) 請舉出一種常見的瑕疵檢測訓練方法，是能夠只使用正常樣本或少量異常樣本就有效地訓練模型，並簡單講解該方法的原理。

Autoencoder 為基礎的異常檢測方法，Autoencoder 是一種 Unsupervised Learning Model，由 Encoder 和 Decoder 組成，其訓練目標是學會如何將輸入資料 encode 壓縮成 latent representation 後再用 decoder 來 reconstruct 還原回原始模樣，越接近原始模樣則 reconstruction error 愈低。

使用正常樣本訓練，讓模型學習正常樣本的分布，若測試樣本與正常分配的差距大 (high reconstruction error) 則代表可能為異常樣本。

3. (10pt) 在瑕疵檢測任務中，除了要對整張圖片進行正常／異常分類外，若圖片中存在異常，我們也希望模型能準確地指出瑕疵所在的位置或範圍。請問這種需求還可以透過哪些類型的電腦視覺任務來達成？寫出一種並簡短說明該任務的基本概念，比方說如何標示異常的位置。

可以用 Semantic Segmentation，目標是對影像中的每個 pixel 分類，輸出一張與輸入同大小的 mask，用來標示每個 pixel 所屬的類別。在瑕疵檢測中，可預測哪些 pixel 屬於異常區域，進而標示瑕疵位置或範圍。

以基於 Encoder-Decoder 的 U-Net 為例，由 encoder (contracting path) 提取特徵、Decoder (expansive path) 負責 upsampling 逐步還原解析度，並透過 skip connection 結合 encoder 的細節資訊，輸出與輸入同樣大小的「像素分類圖」。在訓練階段，輸入為影像，對應的標註為「異常區域的 mask」（黑白遮罩圖）：白色像素（1）代表異常的位置、黑色像素（0）代表正常位置。訓練完成後，U-Net 可以針對任意輸入影像，預測出異常可能出現的區域，以 pixel 為單位輸出異常區域圖。