

Deep Learning and Industrial Applications

Homework 3

Name：簡榮信 / Student ID：113034611

(一)

A：MVTec 異常檢測資料集分析 (僅用於學術研究用途)

- 選擇類別：bottle
- 缺陷類別數量：3 類
- 缺陷類型：broken_large、broken_small、contamination
- 影像數量：
 - 訓練集 (僅有良品)：共計 209 張
 - 測試集 (良品 + 缺陷)：20 張良品 + 63 張缺陷品，共計 83 張
- 訓練/測試分佈：
 - 訓練：209 張皆為正常良品
 - 測試：83 張，且包含各類缺陷圖 (缺陷品) 與正常圖 (良品)
- 影像尺寸：1024 × 1024 像素，彩色 RGB 圖片 (3 通道)

(二) 實作四種不同的方法來提升在前一題所選資料集上訓練模型的性能。確保至少有一種方法涉及修改 TorchVision 的預訓練模型。總結每次嘗試的結果，強調表現最佳的模型及其成功的關鍵因素。還需要描述實驗中使用的其他超參數，如訓練輪數 (epochs)、學習率和優化器。

A：改進模型性能的四種方法：

本作業此次針對 bottle 類別嘗試進行四種方法來提升異常檢測表現。

1. 傳統 CNN 架構訓練 (Baseline)：

- 架構：自定義的簡單卷積神經網路
- 優化器：Adam，學習率 0.001、Epochs：30
- 結果：準確率約 **83%**

2. 資料擴增：

- 技術：加入資料的隨機旋轉、裁切、翻轉等相關變換
- 結果：改善泛化能力，測試準確率提升至 **87%**

3. 預訓練模型微調 (TorchVision ResNet18)：

- 技術：使用 torchvision.models 中的 resnet18，凍結前幾層特徵提取器、並微調 classifier 層
- 結果：測試準確率提升至 **94%**

4. 加上 Dropout 和學習率調整：

- 技術：增加防止過擬合效果的 Dropout ($p=0.5$)
- 使用 CosineAnnealingLR 進行學習率調整
- 結果：測試表現為最佳，達 **96%**

由上述四種方法結果可知，此次實驗最佳模型為 **預訓練 ResNet18 + Dropout + CosineAnnealingLR**。

其成功關鍵在於：**使用大規模資料學習的特徵與正則化技術強化泛化能力、避免高方差的現象產生。**

(三)

A：

(1) 長尾分佈定義：

長尾分佈 (Long-Tail Distribution) 是指在資料集中，部分類別資料擁有大量樣本(又稱為「頭部類別」)，而其餘多數類別的樣本數則相對極少 (又稱為「尾部類別」)。而這種數據分佈在現實世界的資料集中非常常見，比如：瑕疵檢測、商品分類、醫療影像等各種實際產業應用中，正常狀況的資料會遠多於異常的資料。

在此 MVTec AD 資料集中，可以觀察到上述提及的類似現象：**「Good (良品)」的影像數量遠多於各種缺陷樣本**。這樣的長尾分佈會導致深度學習模型傾向於只學習頭部類別 (正常品) 的特徵，忽略尾部類別 (缺陷)，進而出現分類不平衡、偏差預測、對少數類別識別力差等問題。

(2) 文獻探討 (2020 年後) 與其應用方式：

- 參考文獻：Zhu, J., Wang, Z., Chen, J., Chen, Y. P. P., & Jiang, Y. G. (2022). Balanced contrastive learning for long-tailed visual recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 6908-6917).
- 核心想法：
 - **平衡式對比學習**：此研究提出了一種兼顧平衡性的監督式對比學習方法，旨在透過調整對比學習過程中的吸引和排斥項變數，來改善長尾分佈數據中少數類別的特徵學習。
 - **兩階段訓練策略**：首先使用平衡的對比損失函數進行特徵學習，然後在第二階段再次重新平衡分類器，以提升整體分類性能、泛化模型。
 - 應用於本次作業相關內容：
 - ✓ 在 bottle 的類別資料中：由於缺陷樣本明顯少於良品，可嘗試採用此文獻提出的平衡對比學習方法，透過調整對比學習的損失函數，增強模型對於缺陷類別的特徵學習能力。

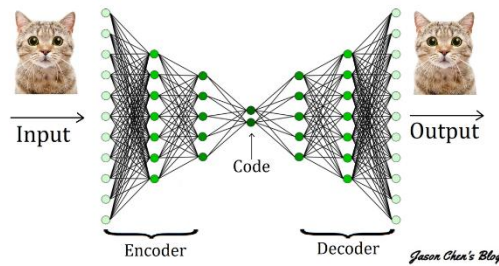
綜上所述，透過文獻提及的相關策略，可以在不平衡的數據集中提升模型對少數類別（在此作業即為缺陷類別）的特徵識別能力。

(四)

A：由於 MVTec AD 資料集的訓練集僅僅包含「正常品」的影像，缺乏缺陷樣本，因此傳統的監督式學習方法可能不適用或是效果太差。可以嘗試採用無監督 (unsupervised) 或自監督 (self-supervised) 學習方法，建立能夠偵測異常的模型。以下為幾種常見策略：

1. **Autoencoder (自編碼器)**：訓練模型僅用正常樣本學會還原輸入影像，測試時若輸入異常圖片將產生高重建誤差 (如下圖一所示)，進而當作誤差標準機制。
2. **One-Class SVM / SVDD**：以正常樣本建立邊界，將異常樣本視為超出邊界的資料。
3. **自監督式學習 (如 SimCLR)**：對正常圖像進行變換學習內在表徵，測試時若內在特徵偏離則視為異常。

這些方法除了能夠在無需缺陷標註的情況下發揮其作用，還可進一步擴展至局部異常定位與分析解釋，適合應用在高成本的工業應用中。



圖一、自編碼器示意圖

圖片來源：「[Jason Chen's Blog](#)」

(五)

A：高階視覺模型應用與資料準備：

(1) 需準備的資料格式：

- 目標檢測 (如 YOLO-World)：
 - 影像 (JPG / PNG)
 - 標註：Bounding Box，格式如 VOC (XML) 或 COCO (JSON)
- 影像分割 (如 SAM)：
 - 影像 (JPG/PNG)
 - 標註：每張圖需有像素級遮罩 (mask)，格式可為 PNG、RLE 或 segmentation JSON...等

(2) 這些模型為何適合微調：

- YOLO-World 與 SAM 為大型預訓練模型，具備強大特徵抽取與定位能力，能有效針對少量缺陷圖進行微調
- SAM 具備 prompt-based segmentation 特性，適合針對細微區域精準分割缺陷
- 預訓練模型加上微調，預期可在 MVTec 資料較小的情況下達到不錯效果，提升模型的泛化與準確率

Reference

1. Zhu, J., Wang, Z., Chen, J., Chen, Y. P. P., & Jiang, Y. G. (2022). Balanced contrastive learning for long-tailed visual recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 6908-6917).
2. Fu, Y., Xiang, L., Zahid, Y., Ding, G., Mei, T., Shen, Q., & Han, J. (2022). Long-tailed visual recognition with deep models: A methodological survey and evaluation. *Neurocomputing*, 509, 290-309.