National Tsing Hua University

11320IEEM 513600

Deep Learning and Industrial Applications

Homework 3

Name : 魏詠新　　　　　　　　Student ID: 113034542

Due on 2025/04/10.

Note: DO NOT exceed 3 pages.

1. (10 points) Download the MVTec Anomaly Detection Dataset from Kaggle (here). Select one type of product from the dataset. Document the following details about your dataset:
   Choose "pill"
   - Number of defect classes : 7
   - Types of defect classes : scratch, pill_type, faulty_imprint, crack, contamination,combined, color
   - Number of images used in your dataset : 434
   - Distribution of training and test data : 0.62:0.38
   - Image dimensions : 800 × 800 × 3

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.)

   【作法一】分配原訓練集good資料夾的圖片跟原測試集的圖片，分別以比例70:30合併成新的train set跟 val set。使用的epochs=15, learning rate=0.001, optimizer是optim.Adam, batch size=32

Attempts	Train Accuracy	Val accuracy
1. 改變學習率0.0005、資料增強(隨機剪裁到指定大小、隨機翻轉、旋轉、改變顏色屬性、標準化)	69.36%	67.05%
2. 原先使用BCEWithLogitsLoss改嘗試 Focal Loss 來改善學習效果	69.64%	67.94%

3. 模型修改 fc 層為多層（Sequential + Dropout + ReLU）、凍結模型的前幾層，並微調模型的後幾層、加入 regularization	83.17%	85.50%
4. 再修改模型從 resnet18 到 resnet50	94.39%	90.08%

*黃色標註為表現最佳，延續前面四個改變再修改模型從 resnet18 到 resnet50，逐步達到較佳的表現

【方法二】 因訓練集中只有good資料夾，故嘗試以自編碼器（Autoencoder）開始，batch size設定32、lr=0.001、epochs=25、optimizer為 Adam 優化器

Attempts	測試集整體準確率
1. 資料增強(PIL影像、隨機翻轉、隨機旋轉、亮度、隨機裁剪)、ROC分析	62.28%
2. 將原先autoencoder模型加入Skip Connection（U-Net 風格），在 Encoder 和 Decoder 之間加上 Skip Connection	46.11%
3. 移除Skip Connection，使用預訓練的 ResNet50 作為 Encoder	88.02%
4. 嘗試修改其他lr、batch size，發現並未改善準確率	85.03%

*黃色標註為表現最佳，延續前面再使用預訓練的 ResNet50 作為 Encoder，再加上先前作的資料增強、ROC分析等，使模型達到較佳的表現。

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.'
   長尾分佈是一種統計分佈，其中大多數觀察值集中在少數類別（稱為「頭部」），而其他類別則包含極少數的觀察值（稱為「長尾」），導致資料集中的類別分佈極度不均衡。
   (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.

《A Balanced Meta-Softmax Loss for Long-Tailed Visual Recognition》（2020）提出了 Balanced Meta-Softmax (BMSoftmax) 的方法，該方法BMSoftmax 結合了 Balanced Softmax 和 Meta Sampler 。Meta Sampler 透過學習來調整每個類別的採樣率，確保在訓練過程中，即使是樣本數量極少的類別也能被模型充分學習。我們可以使用 BMSoftmax Loss 來訓練分類模型，讓模型即使面對極度不平衡的情況（如大量 good 圖片與少量 defect 圖片），也能更準確辨識缺陷類別。

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.)

適合採用非監督式或自監督式的異常偵測方法，例如自編碼器（Autoencoder）、變分自編碼器（VAE）或使用預訓練模型提取特徵後，以距離或重建誤差判斷異常，也可使用 One-Class SVM 或基於特徵分布的方法。這些策略能透過學習正常樣本的模式，偵測與之差異較大的異常情況。

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 (website) and SAM (website). (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.
   物體檢測需要準備包含有標註的圖像數據集，每張圖像應該包含邊框，並且每個邊框應該標註對應的物體類別和位置（坐標）。對於圖像分割，數據集應該包含像素級別的標註，每個像素都有對應的標註，指示它屬於哪個物體或背景，從而將圖像分割成不同的區域。
   (ii) (10 points) Why are these models suitable for fine-tuning for our custom dataset?
   這些模型適合進行微調以適應我們的自定義數據集，因為它們已經在大規模數據集上進行了預訓練，捕捉了各種特徵和模式。這樣的預訓練使得模型能夠很好地進行遷移學習，並且通過微調，可以使模型的知識適應我們數據集的特定特徵，提高對異常檢測或物體分割任務的準確性和效率。