National Tsing Hua University

11320IEEM 513600

Deep Learning and Industrial Applications

Homework 3

Name: 李宜庭 Student ID: 112030751

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:
   - Number of defect classes.
   - Types of defect classes.
   - Number of images used in your dataset.
   - Distribution of training and test data.
   - Image dimensions.

- **Number of defect classes**:
  - **3** defect classes
- **Types of defect classes**:
  - broken_large, broken_small, contamination
- **Number & of images used in your dataset**:
  - Training set (only good samples): **209** images
  - Test set:
    - good: 20 images
    - broken_large: 20 images
    - broken_small: 22 images
    - contamination: 21 images
- **Number & Distribution of training and test data**:
  - **Training**: Only contains **209 good images**
  - **Testing**:
    - good: 20
    - broken_large: 21
    - broken_small: 16
    - contamination: 26
    - **Total test**: 83 (20 good + 63 defective)
- **Image dimensions**:
  - All original images are of size **900 × 900 pixels**, RGB format.

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.)

Hyperparameters: 模型1(epochs=20)、模型2-4(epochs=10)、lr=0.001、optimizer=Adam

模型組合	**模型類型**	**Accuracy**	**Good Recall**	**Defect Recall**	**F1 (Defect)**	**備註**
1. AutoEncoder	Unsupervised	0.43	0.95	0.27	0.42	僅用 good 資料訓練
2. ResNet + CrossEntropy	Supervised	0.79	1.00	0.75	0.86	有 overfitting risk
3. ResNet + Weighted CE	Supervised	0.90	1.00	0.85	0.92	改善Imbalance 問題
4. ResNet + Weighted CE + Dropout + Weight Decay	Supervised	**0.91**	**1.00**	**0.88**	**0.93**	表現最佳

模型1為AutoEncoder-based unsupervised anomaly detection，完全基於Good類別之Train Data訓練而成，而模型2-4於資料處理階段時，有針對Test Data每個Defect類別中隨機抽取5張資料加入Train Data。整體而言，AutoEncoder 在完全未見過 defect 類別的情況下，對defect的辨識能力較弱，導致整體準確率與 F1 較低，屬於 baseline。而ResNet with CrossEntropyLoss 能有Classification的功能，但面對Imbalance問題時效果有限。加入Weighted CrossEntropy後，有效改善 defect 類別被忽略的問題，使模型對Defect更為敏感。最後結合 Dropout + weight decay 的Regularization版本更進一步提升模型的泛化能力，為目前表現最佳的組合。

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.'

   當資料集中有少數類別擁有大量樣本（head classes），而多數類別只有少量樣本（tale classes）時，就會出現「long-tail distribution」。這種不平衡會在類別頻率分布中形成一條「long-tail」，而樣本主要集中在少數的 head classes 上，導致模型在訓練時容易偏向預測head classes。

   (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.

   LTAD(2024)採用兩階段架構解決Data Imbalance的問題：Phase 1利用pre-train model（如 CLIP）生成缺陷的偽類別描述，並透過 VAE 在特徵空間合成樣本以平衡數據分佈；Phase 2 訓練 Transformer 重建模組與語義二元分類器，同時處理多類缺陷檢測。此方法無需預定義缺陷名稱，對於 MVTec AD 的極端不平衡場景（β=200），顯著提升小樣本缺陷的Recall rate（+26%）與 AUROC（+8.2%），適用於稀有或未見過的缺陷模式檢測。

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.)

   當訓練資料僅包含正常（good）圖片時，可採用unsupervised或self-supervised的異常偵測方法。常見策略為訓練 AutoEncoder 或 Variational AutoEncoder（VAE），透過重建誤差偵測異常。另一方法為Contrastive Learning，透過學習「正常樣本」的特徵分佈，來找出那些「明顯不同」或「偏離這個分佈」的輸入，進而判定它是異常（defect）。

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.

   若使用 YOLO-World 進行物件偵測，需為每張影像標註瑕疵的 bounding box 位置及對應類別，標註格式可採 COCO 或 Pascal VOC。若使用 SAM 進行語意分割，則需提供 pixel-level 的遮罩（mask），以binary mask image或polygon明確標示瑕疵的輪廓與區域。這些標註將協助模型學習缺陷的位置、邊界與形狀，提升偵測與分割的準確性。

   (ii) (10 points) Why are these models suitable for fine-tuning for our custom dataset?

   YOLO-World 與 SAM 均在大規模、多樣化資料集上進行pre-train，具備優異的泛化能力。透過微調，模型能快速適應產品領域中的瑕疵特徵，進而強化檢測效能。兩者皆為開源且模組化架構，可進行Transfer Learning與自訂化的訓練流程。其出色的空間辨識能力，特別適合用於對缺陷區域需高度準確標註的 anomaly detection 任務。