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（稱為頭部類別），而多數類別只有少量樣本（稱為尾部類別）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」，其中大多數類別都是樣本稀少的少數類。

   (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(CVPR 2024)採用兩階段架構解決數據不平衡問題：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

Ref: *Ho, C.-H., Peng, K.-C., & Vasconcelos, N. (2024). Long-Tailed Anomaly Detection with Learnable Class Names.*

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supervised的異常偵測方法。常見策略為訓練 AutoEncoder 或變分自編碼器（VAE），使其學會重建正常樣本，推論時以重建誤差判斷異常。另一方法為contrastive Learning，藉此學習正常樣本的特徵分佈。近年如 PatchCore 與 PaDiM 等技術也僅需 good 資料，透過建模正常分佈，將偏離樣本視為異常，且無須標註異常資料。

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.

   若要進行物件偵測，需準備每張圖片的瑕疵位置與類別的 bounding box，格式如 COCO 或 Pascal VOC。若採用 SAM 進行語意分割，則需更精細的像素級遮罩（mask），可用二值圖或多邊形方式標示異常區域，以便模型學習缺陷的邊界與形狀。

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

   YOLO-World 與 SAM 均在大規模多樣性資料上預訓練，具備強大泛化能力。微調能讓模型針對特定產品與瑕疵類型學習特化特徵。這些模型開源、模組化，適合遷移學習與客製化訓練，且可精準偵測與標註區域，特別適用於需高空間辨識度的瑕疵偵測任務。