National Tsing Hua University

11320IEEM 513600

Deep Learning and Industrial Applications

Homework 3

Name: 林郁希 Student ID:312350045(交大)

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:

- Number of defect classes.: 8 classes
- Types of defect classes.: Bent wire, cable swap, combined(multiple defects), cut inner insulation, cut outer insulation, missing cable, missing wire, poke insulation
- Number of images used in your dataset.: 9-13 per class
- Distribution of training and test data: 298/75
- Image dimensions: 1024X1024

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.)

針對A. Restnet:18 B. resize:32 C. learning rate D. epoch number，這些模型內組件進行調整，learning rate數值為0.001比0.0001的表現好，推測為0.0001的optimize效率太低，會在local minimum卡住，reshape數值為224比32表現好，推測為能抓取的圖片細節比較多，利於模型捕捉到缺陷，epoch number 50、100和Restnet 34、18各有不同的對應表現，在50 epochs的情況下34表現會比18好，100 epochs的情況下34表現會比18好，推測為34層數較深，前期會因此學習速度較快，而後期會對於細節過度敏感而在小地方收斂。最後為A: 18 B: 224 C: 0.001 D: 150 表現最好，可以到60% or 63% test accuracy。

hyperparameters and modifying data:

A. Restnet:18 B. resize:32 C. learning rate D. epoch number

```
train_transforms = transforms.Compose([
    transforms.Resize((32, 32)),
    transforms.RandomHorizontalFlip(),
    transforms.AutoAugment(),
    transforms.ToTensor(),
    transforms.Normalize(mean=[0.485, 0.456, 0.406],
                         std=[0.229, 0.224, 0.225])
])
val_transforms = transforms.Compose([
    transforms.Resize((32, 32)),
    transforms.ToTensor(),
    transforms.Normalize(mean=[0.485, 0.456, 0.406],
                         std=[0.229, 0.224, 0.225])
])
```

A: 34 B: 32 C: 0.0001 D: 50

```
Test accuracy is 40.0%
```

```
Epoch 50/50, Train loss: 2.1518, Train acc: 30.00%, Val loss: 2.0184, Val acc: 40.00%, Best Val acc: 43.33%
```

A: 18 B: 32 C: 0.001 D: 50

```
Test accuracy is 50.0%
```

```
Epoch 50/50, Train loss: 1.7185, Train acc: 42.50%, Val loss: 1.6474, Val acc: 50.00%, Best Val acc: 53.33%
```

A: 34 B: 224 C: 0.001 D: 50

```
Test accuracy is 56.7%
```

```
Epoch 50/50, Train loss: 1.0107, Train acc: 70.00%, Val loss: 1.3266, Val acc: 56.67%, Best Val acc: 60.00%
```

A: 18 B: 224 C: 0.001 D: 50

```
Test accuracy is 53.3%
```

```
Epoch 50/50, Train loss: 1.0131, Train acc: 70.00%, Val loss: 1.1800, Val acc: 53.33%, Best Val acc: 56.67%
```

A: 34 B: 224 C: 0.001 D: 100

```
Test accuracy is 40.0%
```

```
Epoch 100/100, Train loss: 0.6801, Train acc: 83.33%, Val loss: 1.2665, Val acc: 40.00%, Best Val acc: 56.67%
```

A: 18 B: 224 C: 0.001 D: 100

```
Test accuracy is 60.0%
```

```
Epoch 100/100, Train loss: 0.7794, Train acc: 78.33%, Val loss: 1.0000, Val acc: 60.00%, Best Val acc: 66.67%
```

A: 18 B: 224 C: 0.001 D: 150

```
Test accuracy is 63.3%
```

```
Epoch 150/150, Train loss: 0.5732, Train acc: 85.00%, Val loss: 0.9337, Val acc: 63.33%, Best Val acc: 70.00%
```

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.'
(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.

'long-tail distribution.'是指一個類別內的物件(資料)，其中大部分的物件對於總體的產出(利潤、精準度、貢獻)，對於總體影響較小，反而小部分的物件對於對於總體影響較較大。在我們的模型上就會指向"good"是大部分的情況產線的情況，然而我們期望模型的目的是偵測到電線異常的狀況，但這方面的資料又較少。解決方法:來自 Deep Long-Tailed Learning: A Survey，1. Class Re-sampling: 解決資料不平衡，2. Information Augmentation: 一方面可以用 transfer learning 的方式達成，另一方面是產出更多新資料，3. Module Improvement: 調整 module

的使用找出適合的最優解

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.)

跟上題 long-tail distribution 的狀況相似，解決的策略也會相近。顯然 defect 的資料會比 good 的少，所以可以用 re-sampling 重複採計，或是 augment 製造更多的資料來讓資料平衡一些，更細微的可以用 re-weighting 或 re-margining 的方式調整資料採計的權重。另外假如其他人有提供同樣的 long-tail distribution 訓練後模型，我們或許可以用 transfer learning 的方式來沿用，提高自己訓練的效率。最後對於模型參數，可以調整不同的 hyperparameter 找到最佳的策略。

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.

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

對於 YOLO-World 類模型，我們需要準備 bounding box 的圖像資料，內容為富有目標缺陷的 label，常見的格式為 COCO JSON 或是 YOLO TXT。而對於 SAM，我們需要準備 pixel-level masks，內容為可以精確描邊出缺陷的形狀，常見格式為 COCO 或是 Pascal VOC。至於為甚麼，這兩類模型適合我們的自訂資料集進行微調，是因為他們已經有充分的預訓練，有良好辨別形狀、輪廓的能力，讓 transfer-learning 的 fine-tuning 成本極低，另外因為資料集要明確框出缺陷位置，讓不同資料只要能做到標出缺陷位置和提供類別，就能使用，泛用性很高。