National Tsing Hua University

1130IEEM 513600

Deep Learning and Industrial Applications Homework 3

Name: 黃品翰 Student ID: 112034511

1. I choose wood dataset

```
Number of defect classes: 6
Types of defect classes: ['color', 'combined', 'good', 'hole', 'liquid', 'scratch']
Number of images used in the dataset: 386

Distribution of data:
train: 247 images
test: 79 images
ground_truth: 60 images
Image dimensions (height, width, channels): (1024, 1024, 3)
```

2.

First :baseline(Optimizer =Adam，lr=1e-3、CosineAnnealingLR 、epochs=50)

```
Train loss: 0.7586, Train acc: 78.8961%, Val loss: 0.9717, Val acc: 74.3590%, Best Val loss: 0.8613, Best Val acc: 75.64%
```

Test accuracy :75.64%

Second: Unfreeze layer 4 and fc layer

```
Train loss: 0.3750, Train acc: 87.9870%, Val loss: 1.1077, Val acc: 74.3590%, Best Val loss: 0.8034, Best Val acc: 76.92%
```

Test accuracy :76.92%

Third: Apply rotation and color jitter in train transform block

```
Train loss: 0.9261, Train acc: 73.3766%, Val loss: 1.0943, Val acc: 70.5128%, Best Val loss: 0.9648, Best Val acc: 75.64%
```

Test accuracy :75.64%

Fourth: Use SGD instead of Adam (lr=0.01, momentum=0.9, weight_decay=1e-4)

```
Train loss: 0.7423, Train acc: 77.2727%, Val loss: 0.9222, Val acc: 75.6410%, Best Val loss: 0.8190, Best Val acc: 76.92%
```

Test accuracy :76.92%

Summary: Among the four different approaches, unfreezing certain layers for fine-tuning deeper features combined with the use of SGD resulted in better performance improvement. This may be due to allowing the model to adjust higher-level features or benefiting from the regularization effect of SGD, enabling the model to better adapt to task-specific characteristics.

3.

(i) **Long-tail distribution:**

long-tail distribution refers to a scenario where only a few classes contain a very large number of samples while most other classes have only a few examples. Kind of like the concept of data imbalance.In the MVTec AD dataset, the "Good" class contains significantly more images than any defect category. This imbalance can lead the model to overfit on the majority class and struggle to learn the features of rarely occurring defects.

(ii) **A proposed solution to data imbalance:**

A work published after 2021 called **Balanced MSE for Imbalanced Visual Classification**. The authors propose modifying the Mean Squared Error (MSE) loss to address class imbalance by introducing a carefully designed balance term. This Balanced MSE (BMSE) framework reweights samples from minority classes so that they contribute more effectively during training, mitigating overfitting to the dominant classes. By penalizing prediction errors more heavily for underrepresented categories, BMSE improves decision boundaries across all classes.

4. Most training images are good with few or no defect examples.So the key is to train a model to learn what normal looks like and then flag deviations as anomalies. One strategy is to use unsupervised classification methods. For example, autoencoder can be trained solely on normal images so that it learns to reconstruct them accurately. Then, images with high reconstruction errors can be flagged as anomalies. Other approaches include using generative adversarial networks or one-class SVMs that learn the distribution of normal images. Additionally, synthetic defect generation can sometimes help.

5.

(i) To fine-tune YOLO-World or SAM for anomaly detection, distinct dataset formats are required. For object detection, each image must be accompanied by bounding box annotations that tightly encapsulate the regions where anomalies appear. These labeled boxes specify both the location and class of the abnormality. For Segmentation, demands like pixel-level masks outlining the exact shape of each defect within the image. Accurate mask generation ensures that the model learns to differentiate defective areas from normal regions at a finer resolution.

(ii) Because they come pretrained on extensive image repositories, granting them robust feature extraction abilities. When refined with domain-specific samples, they adapt their learned representations to detect anomalies more effectively. This transfer learning approach typically requires fewer labeled images than training from scratch, helping quickly achieve high-performance detection or segmentation results while minimizing the effort involved in data annotation.