National Tsing Hua University

Deep Learning and Industrial Applications

Homework3

Name: 謝伊婷 Student ID: 112034533

1.

Document the following details about your dataset: **carpet**

- Number of defect classes：**6**
- Types of defect classes：**color, cut, good, hole, metal_contamination, thread**
- Number of images used in your dataset：**6 * 17**
- Distribution of training and test data：(**0.8 , 0.2**)
- Image dimensions：**3 * 1024 * 1024**

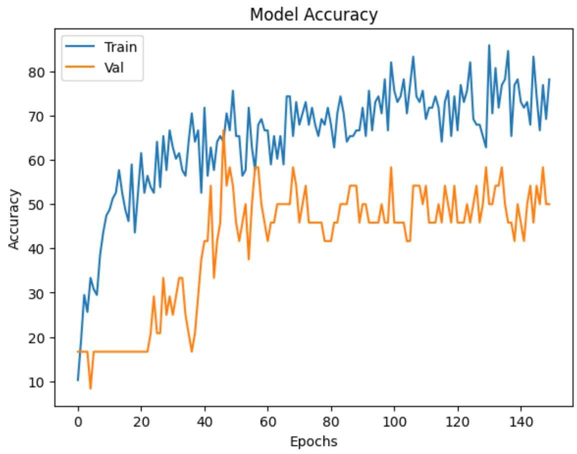
2.

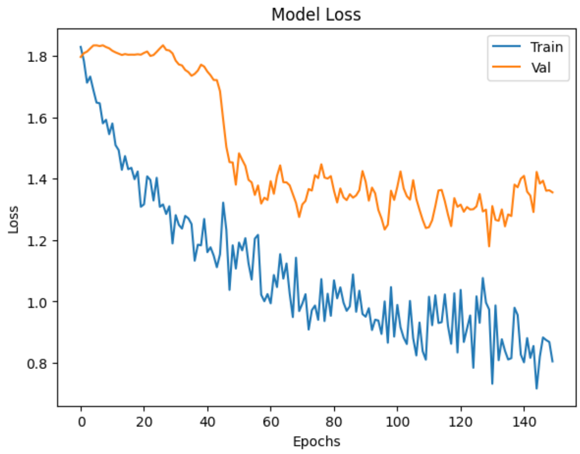
resize	epoch	learning rate	Train Loss	Train Accuracy(%)	Val Loss	Val Accuracy(%)	Test Accuracy(%)	pretrain model	weight
32	50	0.001	1.53	34.62	1.86	12.50	29.17	resnet18	IMAGENET1K_V1
64	50	0.001	1.52	33.33	1.80	25.00	29.17	resnet18	IMAGENET1K_V1
64	100	0.001	1.20	55.13	1.76	29.17	33.33	resnet18	IMAGENET1K_V1
64	100	0.001	1.08	65.38	1.65	37.50	50.00	resnet50	IMAGENET1K_V1
128	100	0.001	0.95	64.10	1.60	41.67	50.00	resnet50	IMAGENET1K_V1
128	100	0.001	1.04	66.67	1.26	54.17	62.50	resnet50	IMAGENET1K_V2
256	100	0.001	1.04	66.67	1.47	45.83	58.33	resnet50	IMAGENET1K_V2
128	150	0.001	0.80	78.21	1.36	50.00	66.67	resnet50	IMAGENET1K_V2
64	150	0.001	0.82	82.05	1.26	45.83	58.33	resnet50	IMAGENET1K_V2
256	150	0.001	0.84	82.05	1.56	45.83	54.17	resnet50	IMAGENET1K_V2
128	300	0.001	0.67	78.21	1.42	45.83	62.50	resnet50	IMAGENET1K_V2
32	300	0.001	1.48	47.44	1.96	29.17	45.83	resnet50	IMAGENET1K_V2
256	300	0.001	0.65	83.33	1.61	45.83	58.33	resnet50	IMAGENET1K_V2

實驗結果顯示，torchvision 的 models 把原本的 ResNet18 換成 ResNet50，並將 model 的 weight 調整成 IMAGENET1K_V2，model 的表現有明顯提升。

而把 epoch 和 resize 提高，model 的表現也會提高，特別是在上圖中最後一組 256x256、300 epoch 並搭配 ResNet50 model 的設定下，train accuracy 達到 83.33%，但是這組在 val set 的表現卻沒有提升很多，可能有過擬合的情況。

綜合來看，我認為這些設定中，表現最好的是 128x128、150 epoch 的這組，test accuracy 達到 66.67% (下圖是這組參數下 model performance 的 Visualizing)。





3.

(i) A long-tail distribution is a probability distribution that exhibits a large number of occurrences in the tail of the distribution, which represents a low frequency of occurrence for a wide variety of events, items, or conditions. This is in contrast to a small number of occurrences that represent the "head" and have a high frequency.

(ii)

The paper "Use Your Head: Improving Long-Tail Video Recognition" examines long-tail distribution in video recognition and proposes a method called Long-Tail Mixed Reconstruction (LMR). This method combats the overfitting to few-shot classes by reconstructing them with weighted combinations of head class samples. Furthermore, it uses label mixing to learn robust decision boundaries, which has shown state-of-the-art performance on various datasets.

https://arxiv.org/abs/2304.01143v1

4.

To develop an anomaly detection model with the MVTec AD dataset, which has a predominance of 'good' images, consider the following strategies:

- **Unsupervised Learning**: Employ autoencoders that learn to reconstruct 'good' images and detect anomalies as reconstruction errors.
- **Data Augmentation**: Generate synthetic defect examples using image manipulation or generative adversarial networks (GANs).
- **Transfer Learning**: Apply knowledge from models pre-trained on similar tasks with more diverse defect examples.
- **Few-Shot Learning**: Utilize methods that can learn to detect anomalies with very few examples of defects.

5.

(i) For object detection, you should prepare a dataset with bounding box annotations that highlight where each object is located in the image. For segmentation, the dataset must include pixel-wise masks that delineate the exact shape of each object, allowing the model to learn precise contours and edges of the anomalies.

(ii) These models are suitable for fine-tuning on a custom dataset because they've been pre-trained on vast amounts of data, capturing a wide range of features that can generalize well to new tasks with further training. Fine-tuning allows these robust models to apply their learned features to the specific task of anomaly detection, adjusting to the unique patterns of defects within the given dataset.