National Tsing Hua University

11220IEEM 513600

Deep Learning and Industrial Applications

Homework 3

Name : 龔友謙 Student ID : 112030512

1. Document the following details about your dataset (I choose bottle for my dataset)
   (1) Number of classes : 4
   (2) Types of classes : good, broken_large, broken_small, contamination
   (3) Number of images used in your dataset：80 (20 each)
   (4) Distribution of training and test data : 64 training data, 16 testing data
   (5) Image dimensions : (900, 900, 3) represent 900x900 pixels. With 3 channels

2. Implement **4** different attempts to improve the model's performance

Test accuracy	original	1st attempt	2nd attempt	3rd attempt	4th attempt
	31.25%	37.5%	37.5%	65.0%	56.25%
Epochs	50	50	50	200	500
Learning rate	0.001	0.0001 (for the unfrozen layers)	0.001	0.001	0.00001
Optimizer	Adam	Adam	Adam	Adam	Adam
Regularization			Dropout with a rate of 0.2	Dropout with a rate of 0.2	Dropout with a rate of 0.2

1st attempt：Fine-tuning the Pre-trained ResNet-18 Model，having the model to learn more specific features for the dataset.

2nd attempt ：applying various data augmentation techniques and regularization to prevent overfitting.

3rd attempt : change the epochs to 200, trying to get better result by having more epochs.

4th attempt : change the epochs to 500, change the learning rate to 0.00001.

Summarize : It boost the test accuracy from 37.5% to 65.0%. From the data visualization it shows that about 150 epochs is enough, it is hard to improve after this.

3.

(i) Long-tail distribution : also known as data imbalance, meaning that where a few classes or categories have a significantly larger number of instances compared to the rest of the classes. This results in an uneven distribution of data, with a small number of classes dominating the dataset while the remaining classes are represented by a comparatively smaller number of instances.

(ii) A paper that proposes a solution to data imbalance is "Adaptive Focal Loss for Imbalanced Data Classification" by Kai Chen et al., published in 2020. The authors introduce an Adaptive Focal Loss (AFL) method that addresses the problem of data imbalance by dynamically adjusting the focal loss based on the distribution of the data.

In the case of the MVTec AD dataset, where the 'Good' class has significantly more instances compared to the defect classes, the AFL method could be applied to mitigate the data imbalance issue. By dynamically adjusting the focal loss function during training, the model would be encouraged to pay more attention to the defect classes, leading to improved overall performance on the dataset, even with the imbalanced distribution.

4.

(1) We can apply the method mentioned in the last question, having the ability to dynamically adjust the focal loss based on the class imbalance ratio

(2) Transfer Learning, like the code TA provide in the lab, The pre-trained model can be fine-tuned on the 'good' samples from the MVTec AD dataset, and the resulting features can be used for anomaly detection.

(3) Generative Adversarial Networks (GANs), GANs can be leveraged to generate synthetic defect samples, which can then be used to balance the training data.

5.

(i)

For Object Detection:

- The dataset should consist of images with bounding box annotations around the defective regions. This allows the model to learn to detect the presence and locations of defects within the images.
- Each image should be labeled with the corresponding class of the defect, and the bounding box coordinates should be provided for each instance of the defect.

For Segmentation:

- The dataset should include images with pixel-level segmentation masks, where the defective regions are delineated at the pixel level.
- The segmentation masks should be provided as separate annotations, aligned with the corresponding input images.

(ii)

(1) Transfer Learning: These models have already been trained on large, general-purpose datasets. This means they have learned useful visual features that can be adapted to the specific task of anomaly detection on the MVTec AD dataset.

(2) Efficiency: It's more efficient to fine-tune these pre-trained models than to train a new model from scratch, especially when the custom dataset is relatively small.

(3) Specialized Architectures: Models like YOLO-World and SAM are designed specifically for object detection and segmentation, which are key capabilities for identifying and locating defects in the images.