National Tsing Hua University

11220IEEM 513600

Deep Learning and Industrial Applications

Homework 3

Name: Li-Chiao Wang (王麗喬)   Student ID: 112034804

Completed on April 10, 2024

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:

Select dataset: **tile**

• Number of classes: 6 defect classes.

• Types of classes: good, crack, glue_strip, gray_stroke, oil and rough.

• Number of images used in your dataset: 15*5 (defect images) + 15 (good images) = 90 (total).

• Distribution of training and test data: 80% for training (72 images), 20% for testing (18 images).

• Image dimensions: The dataset consists of 90 images, each measuring 840x840 pixels with RGB color channels.

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.)

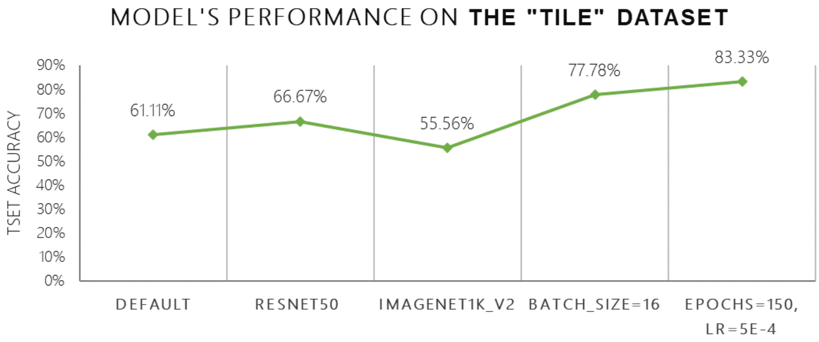
- Default setting: model=ResNet18, weights=IMAGENET1K, batch size=32, epochs=100, optimizer=Adam, learning rate=1e-3.
- Test accuracy (default setting) is 61.11%.

(1) Switching the model from ResNet18 to **ResNet50** resulted in a test accuracy of 66.67%, indicating an improvement over the default setting.

(2) Adjusting the model weights to **IMAGENET1K_V2** led to a decrease in performance, with a test accuracy of 55.56%.

(3) Changing the model weights to IMAGENET1K_V1 and adjusting the **batch size** from 32 to **16** resulted in a notable improvement, with a test accuracy of 77.78%.

(4) Increasing the number of **epochs** from 100 to **150** and reducing the **learning rate** from 1e-3 to **5e-4** yielded the highest test accuracy of 83.33%.



Overall, the performance of the model significantly improved through various modifications. Switching to a deeper architecture (ResNet50) and adjusting hyperparameters such as batch size, model weights, epochs, and learning rate played pivotal roles. The best performing model achieved an 83.33% test accuracy by fine-tuning these hyperparameters.

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.'

A long-tail distribution, or data imbalance, arises when certain categories occur much more frequently than others. This leads to a skewed distribution in datasets, with fewer instances in some classes (the tail) compared to others (the head).

(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.

The paper "Self-supervised Learning for Medical Image Classification Using Imbalanced Training Data" by Chen and Li (2022) proposes a method combining self-supervised pre-training with supervised fine-tuning, followed by Balanced-MixUp, to address data imbalance in medical image classification. This approach can be adapted for the MVTec AD dataset by initially pre-training the model using self-supervised learning on the abundant 'good' class images. Then, Balanced-MixUp can be employed during fine-tuning to handle the imbalance between the 'good' class and defect classes, resulting in improved defect detection performance.

- Chen, W., Li, K. (2022). **Self-supervised Learning for Medical Image Classification Using Imbalanced Training Data**. In: Li, K., Liu, Y., Wang, W. (eds) Exploration of Novel Intelligent Optimization Algorithms. ISICA 2021. Communications in Computer and Information Science, vol 1590. Springer, Singapore. https://doi.org/10.1007/978-981-19-4109-2_23

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.)

To address the lack of defect examples, several strategies can be employed:

(1) Augmentation Techniques: By synthetically introducing defects into 'good' images, the dataset can be enriched with examples of anomalies, providing the model with a more diverse training set.
(2) Transfer Learning: Start training on a dataset with defect images, then fine-tune the model using 'good' images from MVTec AD to enhance its anomaly detection capabilities.
(3) Unsupervised Learning Methods: Use autoencoders or GANs to understand the normal data distribution, enabling detection of anomalies by identifying deviations from this learned distribution.
(4) Semi-Supervised Learning: By incorporating a limited number of labeled defect images from external sources, the model can learn from both 'good' and 'defective' examples, improving its anomaly detection capabilities.

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.

For both object detection and segmentation, having labeled data is crucial. Object detection involves annotating images with bounding boxes that enclose objects, with each box labeled according to its class. On the other hand, segmentation tasks require pixel-level annotations, where each pixel is assigned a class label, enabling precise delineation of objects. Both tasks demand meticulous annotation to effectively train models.

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

Models like YOLO-World and SAM, pre-trained on vast datasets, are prime for fine-tuning. This initial training endows them with broad visual comprehension. Fine-tuning, accomplished with a smaller, task-specific dataset, adapts the model to the target domain's nuances. Leveraging pre-trained models accelerates learning, conserves computational resources, and capitalizes on robust representations, ultimately enhancing anomaly detection performance.