National Tsing Hua University

11220IEEM 513600

Deep Learning and Industrial Applications

Homework 3

Name: 林芮伃　　　　Student ID: 112030509

Due on 2024/04/11.

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: Choose zipper from the dataset.
   - Number of classes. 8
   - Types of classes.
     Color, Combined, Contamination, Crack, faulty_imprint, Good, pill_type, scratch
   - Number of images used in your dataset.
     Number of images used in your dataset: 80
   - Distribution of training and test data.
     Number of images in training data: 64
     Number of images in testing data: 16
   - Image dimensions.
     Image dimensions (height, width, channels): 1024 1024 3

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.)
   1. Modify batch size: The choice of batch size affects both the training speed and performance of the model. A larger batch size may speed up training but could also make the model converge more slowly.
   2. Adjust learning rate: Adjustments are made by monitoring the model's performance on the validation set.
   3. Adjust training epochs: The training epochs need to be long enough but also avoided overfitting.

4. Modify pre-trained model: The pre-trained model was modified by specifically altering its fully connected layers and adding activation functions (ReLU), dropout layers, and a LogSoftmax layer.

epochs	learning rate	optimizer	Batch size	Model	accuracy
50	lr=1e-3	Adam	32	ResNet-18	12.5%
50	lr=1e-3	Adam	128	ResNet-18	31.75%
100	lr=1e-3	Adam	128	ResNet-18	25%
100	lr=2e-3	Adam	128	ResNet-18	37.5%
100	lr=2e-3	Adam	128	ResNet-18 model.fc= nn.Sequential	31.25%

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.'
   Long-tail distribution refers to a situation in a dataset where some categories have very few samples, while other categories have a significantly larger number of samples. This leads to insufficient predictive ability of the model for the minority classes during training, impacting the overall performance of the model.

   (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.

   One paper published after 2020 that addresses the data imbalance issue is "Meta-Learning for Few-Shot Image Classification." The paper proposes a meta-learning-based approach to tackle data imbalance. Their method involves two main steps: first, during meta-training, they train a meta-model using samples from the minority classes; then, during meta-testing, they use the meta-model to guide the model's predictions. This approach can be applied to our case by emphasizing training on the minority classes during meta-training to enhance the model's predictive ability for these minority classes.

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.)

- Generate Defect Data: Use synthetic data techniques like GANs to introduce random noise, blur, or disturbances into normal images, creating defect images.
- Unsupervised Learning: Utilize autoencoders to learn feature representations of normal images and then detect different anomaly images.
- Transfer Learning: Transfer knowledge from models trained on other tasks to improve training effectiveness.
- Feature Engineering: Employ convolutional automatic feature extraction techniques in deep learning to capture differences between normal and anomal images.

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 object detection, a dataset with object bounding box annotations is required. This means annotating each image to indicate the objects present and their location information, such as rectangular bounding boxes. Such annotations enable the model to learn to identify objects and their positions within the image. For segmentation, a dataset with pixel-level labels is needed. This involves labeling each pixel to indicate whether it belongs to a certain object or the background, creating mask information corresponding to each pixel. Such labeling enables the model to achieve precise image segmentation.
   (ii) (10 points) Why are these models suitable for fine-tuning for our custom dataset?
   These models are suitable for fine-tuning because they have been validated and perform well on general computer vision tasks. By fine-tuning these pre-trained models, we can leverage the generic features they have learned on large datasets to achieve more accurate anomaly detection. Additionally, these models typically use deep neural network architectures, allowing them to effectively handle complex image features, making them suitable for application to our custom dataset.