National Tsing Hua University

11220IEEM 513600

Deep Learning and Industrial Applications

Homework 3

Name: 劉奕　　Student ID: 111003817

1. Select one type of product from the dataset. Document the following details about your dataset:
   - Product type: ‘wood’
   - Number of defect classes: 5
   - Types of defect classes: 'liquid', 'hole', 'color', 'good', 'combined'
   - Number of images used in your datasets: 79.
   - Distribution of training and testing data: training set: 53, testing set: 26.
   - Image dimensions: (1024, 1024, 3)
2. Implement 4 different attempts to improve the model's performance trained on the dataset you chose in the previous question.

Exp	Arch.	Num epochs	Optimizer	Learning rate	Training Strategy	Performance (train/test)
Baseline	Simple CNN	30	SGD	1e-3	Train from scratch	24.5%/23.0%
Attempt 1	Simple CNN	30	Adam	1e-3	Train from scratch	30.1%/26.9%
Attempt 2	ResNet18	30	Adam	1e-3	Train from scratch	35.8%/34.6%
Attempt 3	ResNet18	30	Adam	1e-3	Transfer learning	60.3%/61.5%
Attempt 4	ResNet 18	30/10	Adam	1e-3/1e-4	Transfer learning/Fine-tuning	83.0%/69.2%

The above table outlines the different attempts to improve the model's performance. This experiment employed a straightforward Convolutional Neural Network (CNN) as the baseline model, the specifics of which can be found in the notebook titled hw3.ipynb. It utilized an SGD optimizer with a learning rate of 1e-3 across 30 epochs for training. The findings indicate that the key factor leading to improvement was the incorporation of a pre-trained model, particularly when combined with a suitable training strategy: initially, only the final fully connected layer of the model was trained for 30 epochs, followed by training the entire model's weights with a reduced learning rate for an additional 10 epochs.

3. Questions about the long-tail distribution (data imbalance):
   A. Define what is 'long-tail distribution.': A long-tail distribution is a probability distribution with many occurrences far from the "head" or central part of the distribution. It characterizes phenomena where a low frequency of high-amplitude events coexists with a high frequency of low-amplitude events. It is often visualized as a long tail extending to the right on a chart.

   B. 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.
   Paper: Deep Representation Learning on Long-tailed Data: A Learnable Embedding Augmentation Perspective (https://arxiv.org/abs/2002.10826)
   The paper presents a method for learning deep representations from long-tailed data distributions, focusing on the challenge that the lack of intra-class diversity in tail classes leads to distorted feature spaces. It introduces a novel "feature cloud" approach that augments tail class instances in the deep feature space to enhance their distribution. This augmentation mimics the intra-class diversity observed in head classes, thereby improving the overall discriminative capability of the learned features. The methodology involves calculating the intra-class angular distribution from head classes and transferring this distribution to tail classes, effectively expanding their feature space during training. Since the proposed approach is model architecture-agnostic, it can be easily applied to other classification models by integrating additional modules to learn the appropriate embeddings, such as the intra-class angular distribution module and the feature cloud construction for the tail data.

4. Discuss strategies for developing an anomaly detection model under these conditions.
   Developing an anomaly detection model with the MVTec AD dataset, which primarily consists of 'good' images without examples of defects, necessitates strategies that focus on learning the normal variation of data to identify anomalies. One approach is to use unsupervised learning techniques such as autoencoders, which learn to compress and reconstruct the input images; high reconstruction errors then identify anomalies, as the model is less adept at reconstructing unseen, defective examples. Another strategy is one-class classification, such as One-Class SVM or isolation forests, which models normal data distribution and predicts outliers as anomalies. Additionally, augmenting the dataset with synthetic anomalies generated through techniques like GANs or by applying transformations to good images can provide the model with examples of defects to improve its detection capabilities.

5. Questions about the usage of open-source models:
   A. What kind of data should be prepared for object detection and for segmentation?
   Both object detection and segmentation tasks need images and correlated annotations. The annotation must include the bounding box (the object's location) and the object's class name for object detection. For segmentation, the annotations include masks that categorize each pixel in the image. Additionally, preparing the corresponding text data might be required to fine-tune YOLO-world or SAM.

   B. Why are these models suitable for fine-tuning for our custom dataset?
   One reason is that these models are trained with very large and diverse datasets to understand the fundamental features of images (we call these models vision foundation models). Because of their strong ability, these models are suitable for fine-tuning the custom dataset, especially since the images of the custom dataset are not out of the distribution of the original pre-training data. To fine-tune the model with the custom datasets more efficiently, adapter-like [1] techniques are suitable to decrease the training time and avoid the loss of model generalization ability.
   [1] https://arxiv.org/pdf/1902.00751.pdf