National Tsing Hua University

11220IEEM 513600

Deep Learning and Industrial Applications

Homework 3

Name:方品心 Student ID:312704005

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:
   metal_nut
   - Number of defect classes.：4
   - Types of defect classes.：Classes: ['bent', 'color', 'flip', 'scratch']
   - Number of images used in your dataset.：一個類別 20 張，總共 80 張
   - Distribution of training and test data.：training / test 劃分比例：80/20
   - Image dimensions.：3x700x700

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.)
   - Epoch : 200 (原:50，還未收斂)
   - Data Augmentation :
     - (改)

```
train_transforms = transforms.Compose(
    [
        transforms.Resize((224, 224)),
        transforms.RandomRotation(10),
        transforms.ColorJitter(brightness=0.2, contrast=0.2),
        transforms.AutoAugment(),
        transforms.RandomHorizontalFlip(),
        transforms.ToTensor(),
        # transforms.Normalize(mean=[0.485, 0.456, 0.406], std=[0.229, 0.224, 0.225]),
    ]
)
```

(原)

```
train_transforms = transforms.Compose(
    [
        transforms.Resize((32, 32)),
        transforms.AutoAugment(),
        transforms.RandomHorizontalFlip(),
        transforms.ToTensor(),
    ]
)
```

   - Resize：224x224 (原:32x32) 因為預訓練模型預期輸入圖片的大小是224x224或

256x256，且這樣的大小可以較好的保留圖片特徵。

- Pre-trained Model : Resnet50, weights='IMAGENET1K_V2'

  (原: Resnet18, weights='IMAGENET1K_V1')

```
Available weights for model 'resnet50':
ResNet50_Weights.IMAGENET1K_V1
ResNet50_Weights.IMAGENET1K_V2
```

將模型的預訓練模型從ResNet18更改為ResNet50，並使用來自IMAGENET1K_V2的權重，以提高模型的性能和特徵提取能力，使用ResNet50預訓練模型和數據增強技術的模型表現較佳。這是因為ResNet50具有更深的網絡結構，可以更好地捕捉數據中的特徵，並選擇了更新版本的預訓練權重。

- 其他參數皆無更動

Performance : Test accuracy is 81.25% (原:12%左右)

```
Epoch 198/200, Train loss: 0.3762, Train acc: 92.1875%, Val loss: 0.6051, Val acc: 75.0000%, Best Val loss: 0.5823 Best Val acc: 81.25%
Epoch 199/200, Train loss: 0.4154, Train acc: 93.7500%, Val loss: 0.6155, Val acc: 75.0000%, Best Val loss: 0.5823 Best Val acc: 81.25%
Epoch 200/200, Train loss: 0.4240, Train acc: 84.3750%, Val loss: 0.6189, Val acc: 75.0000%, Best Val loss: 0.5823 Best Val acc: 81.25%
```

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.'

   又稱為數據不平衡，長尾分布是指一種統計分佈，大多數觀察值或事件的發生頻率集中在少數的類別或值上，而剩下的觀察值或事件則散佈在較多的類別或值上，形成一條長長的尾部導致數據分佈不均衡。

   (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.

   由X. Li等人在2021年發布的 "Meta-Learning with Adaptive Class Hierarchy for Long-Tailed Visual Recognition"。論文提出了一種元學習方法，該方法適用數據集的類層次結構，在訓練期間動態地將更多資源分配給較少見的類別。他們的方法可以應用於這次的MVTec AD數據集，通過動態調整訓練過程中的採樣策略，將更多資源分配給較不常見的缺陷類別。這種資源的自適應分配有助於模型更好的從不均衡的數據分佈中學習，提高其有效檢測異常的能力。

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.)

A. **Unsupervised Learning**: 非監督式學習，使用聚類或密度估計，將正常圖像與潛

在的異常圖像區分開來。

B. **One-class Classification**: 將問題轉化為單類別分類問題，僅使用正常圖像訓練模型，從而將異常視為外部或未知類別。

C. **Semi-supervised Learning**: 利用有限的瑕疵樣本與大量正常樣本進行訓練，可以考慮使用半監督式學習來提高模型的性能。

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.

   Object Detection：需要準備帶有標註框的圖像，標註框用於指示圖像中的物體位置和邊界框。每個標註框要包含物體的類別標籤和邊界框的坐標信息。

   Segmentation：對於分割任務，需要準備圖像及其對應的像素級標籤，標籤為每個像素指定所屬的類別，才能實現對圖像中每個像素的精確分割。

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

   這些模型適合微調的原因有以下幾點：首先，這些模型在大規模數據上進行了預訓練，具有強大的特徵提取能力，能夠捕獲圖像中的豐富信息；其次，這些模型在訓練過程中已經學習到了最基本的圖像特徵，可以很好的適應不同的瑕疵檢測任務；此外，這些模型通常具有較高的準確性和泛化能力，可以在有限的訓練資料上進行有效的fine-tune。通過使用這些預訓練模型在自己的數據集上進行微調，可以快速且有效地針對特定瑕疵檢測任務進行模型訓練。