工工系碩一 112034702 黃銘渝

1. I choose cable dataset
   - Number of defect classes: 8 class
   - Types of defect classes:
     bent_wire、cable_swap、combined、cut_inner_insulation、
     cut_outer_insulation、good、missing_cable、poke_insulation
   - Number of images used in your dataset: 466
   - Distribution of training and test data: 224, 242
   - Image dimensions: (3, 1024, 1024)
2. 4 different implementations
   - **Result**

Method	Learning Rate	Epochs	Accuracy
Original	1e-3	50	38.666%
Decoder	1e-4	50	40.666%
Decoder with Deeper Network	1e-4	50	47.333%
Decoder with Deeper Network + Loss function	1e-4	5	56%
Deep SVDD	1e-4	50	61.333%

   Optimizers are all Adam

   Loss functions

   Original: Cross Entropy Loss

   Others: MSE Loss

   - **Method Description**
     1. **Original**
        Predict 2 classes, anomaly and good with resnet18
     2. **Decoder**
        Connect resnet18 without fully connected layer with decoder
     3. **Decoder with deeper network**
        Replace resnet18 with the decoder part to resnet50
     4. **Decoder with Deeper Network + Loss function**
        Change loss function with structural similarity loss (SSIM) and combine with the weight of 0.5 with MSE Loss for faster convergence and better performance
     5. **Deep SVDD**
        Utilize Resnet50 as encoder to implement the Deep SVDD
3.

(i) The head of sorted numbers of occurrence is far from the center of the distribution. To be specific, small number of head categories contain a large

number of samples, while the majority of tail categories have a relatively limited number of samples.

(ii) *Yang Lu, Yiliang Zhang, Bo Han, Yiu-ming Cheung, & Hanzi Wang. (2023). Label-Noise Learning with Intrinsically Long-Tailed Data.*
In the article, they introduced Two-Stage Bi-Dimensional Sample Selection (TABASCO) method consisted of Bi-Dimensional Sample Separation stage and Adaptive Centroid Distance. In first stage, they utilized Weighted Jensen-Shannon Divergence (WJSD) and Adaptive Centroid Distance (ACD). WJSD is used to calculate inconsistency between model prediction and given label and adds a weight based on how much the model's highest prediction confidence differs from the observed class. As ACD, it computes the cosine similarity between a sample's feature and its class centroid in feature space. The centroid is also adaptively updated using only high-confidence samples to increase purity. In the second stage they used a Gaussian Mixture Model (GMM) to cluster samples in each class based on each metric (WJSD and ACD). Then, they selected the best metric per class based on separability statistics and chose the cleaner cluster by comparing average metric values and centroid similarities. Eventually, clean samples from all classes are used as labeled data, and others as unlabeled data, and the model is then trained via semi-supervised learning, refining predictions in the next round.

4. As I utilized in question 2, the anomaly model could be constructed with decoder. Since there are only good images in trains, the encoder-decoder based model could construct only good images. Consequently, we could calculate the similarities of reconstructed image and input image and define the anomaly data with lower similarities. For further improvement, we could utilize the structure of GAN to improve the performance of autoencoder for more accurate results.
5.
   (i) To fine-tune models for object detection, the dataset should consist of images paired with bounding box annotations. Each bounding box should include coordinates (x, y, width, height) and the corresponding object label (e.g., "defect"). For segmentation, the dataset requires pixel-level annotations, typically in the form of masks where each pixel is labeled to indicate the class it belongs to (e.g., background vs. defect).
   (ii) Models are well-suited for fine-tuning because they are pretrained on large, diverse datasets and have demonstrated strong generalization. YOLO-World supports open-vocabulary detection, allowing adaptation to new object classes with minimal data. SAM (Segment Anything Model) provides strong zero-shot segmentation capabilities, making it easier to adapt to unseen

categories. Fine-tuning these models with domain-specific data enhances their performance on custom tasks like defect detection.