Deep Learning and Industrial Applications

Homework 3

113030511 簡若伃

1. **Select one type of product from the dataset: transistor**
   - Number of defect classes: **4**
   - Types of defect classes: **Bent lead, cut lead, damaged case, misplaced.**
   - Number of images used in your dataset: **313**
   - Distribution of training and test data: **250, 63**
   - Image dimensions: **256 × 256 × 3**
2. **Implement 4 different attempts to improve the model's performance trained on the transistor.**

Attempt	Model	Data Augmentation	epoch	optimizer	Train acc	Val acc	Test acc
Base	ResNet18		50	Adam	87.6	87.3	87.3
1	ResNet18	RandomRotation(15)	50	Adam	88.4	87.3	87.3
2	ResNet34	Same as base	50	Adam	92.4	88.9	88.9
3	ResNet50	Same as base	50	AdamW	91.2	88.9	90.5
4	ResNet50	RandomRotation(15)	100	AdamW	90.4	85.7	90.5

For the baseline, the data augmentation includes auto augment and random horizontal flip, using ResNet18, the optimizer is Adam, with learning rate using Cosine Annealing LR. Baseline with moderate accuracy. Attempt 1 plus data augmentation random rotation 15. Attempt 2 use ResNet34, and unfrozen least 2 layers, layer 4 and fc. Attempt 3 use ResNet50 and change optimizer to AdamW with weight decay le-4. Attempt 4 use data augmentation random rotation 15 plus ResNet50 plus optimizer AdamW with weight decay and epoch 100. Training for 100 epochs leads to slight overfitting despite higher train accuracy. We can see fine-tuning deeper layers brings significant improvement.

Among the five tested configurations, attempt 3 achieved the best performance with a validation accuracy of 88.9% and a test accuracy of 90.5%. Deeper feature extraction using ResNet50, AdamW optimizer improves generalization through decoupled weight decay, showing simpler and stable data augmentation pipeline, and cosine learning rate helps smooth convergence.

3. **(i) Define 'long-tail distribution.'**

   Long-tail distribution in statistics refers to a portion of the distribution contain many samples, we called “head”, while the other classes contain very few samples. It's a type of class imbalance where data is heavily skewed, the minority classes dominate the dataset, and most classes are underrepresented.

   (ii) Identify and summarize a paper proposes a solution to data imbalance. Explain how their method could be applied to our case.

   Supercharging Imbalanced Data Learning With Energy-based Contrastive Representation Transfer (2021)

   Authors: Junya Chen, Zidi Xiu, Benjamin Goldstein, Ricardo Henao, Lawrence Carin, Chenyang Tao

   The paper introduces a meta-distributional learning approach to address data imbalance, based on a causal assumption, the data-generating mechanism is invariant across different classes, even if their feature distributions differ. This enables knowledge transfer from majority to minority classes. The paper proposes ECRT (Energy-based Contrastive Representation Transfer) that learns causal, disentangled representations of features using contrastive learning. And applies data augmentation in the source space via shuffling or sampling, training the final classifier using both real and augmented minority-class representations.

   Apply to MVTec AD:

   We can use only the 'Good' class to pretrain a feature extractor, typically implemented with a ResNet-based encoder. Now this encoder is used to extract latent feature representations, Z, and then transformed into a source space, S, through Generalized Contrastive Learning, which utilizes an invertible function to ensure that the transformation preserves meaningful structure. Next, perform data augmentation to enrich the representation of underrepresented defect classes, it includes a non-parametric method that involves shuffling the feature dimensions within the same class, and a parametric approach that samples new feature vectors from a Gaussian distribution fitted to the minority class features. Finally, we have classifier in the source space using both real and synthetic data for defect classes.

4. **Strategies for developing an anomaly detection model.**
   - **Feature Embedding + Distance-Based Scoring**

     Use a pretrained CNN, like ResNet, to extract feature vectors for normal images. At inference, compute distance with cosine or Mahalanobis, between test image features and the "normal" feature set, larger the distance, the more abnormal it is. Methods which are common used like PaDiM, a multivariate Gaussian distribution to model feature distributions. And PatchCore, leverages nearest neighbor search on patch-level features.
   - **Self-Supervised Learning**

     Learn general representations by solving pretext tasks, like rotation prediction, contrastive learning, cut paste using only normal data. At test time, images that produce unexpected behavior, like low confidence on rotation are flagged as anomalies. Methods like CutPaste, self-supervised anomaly detection via copy-paste image patch task works well. Or SimCLR/ MoCo + kNN anomaly scoring.
5. **(i) What kind of data should be prepared for object detection and for segmentation?**

   For object detection models, we need defect images with class labels for each bounding box, like bent lead or cut lead. For segmentation models, we need images with pixel-level defects, corresponding segmentation masks for each defect, each pixel labeled as defect or background.

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

   YOLO-World and SAM are pretrained on large-scale datasets, enables effective transfer learning even when our dataset is small or highly specific. Besides, YOLO-World excels in real-time object detection, with precise bounding box prediction and category classification. For pixel-level segmentation, SAM is capable of delineating very fine structures, perfect for highlighting subtle surface defects. These models support class customization, which can detect or segment new categories with just a small amount of labeled data using fine-tuning or prompt-based learning.

Reference:

- https://en.wikipedia.org/wiki/Long_tail
- https://arxiv.org/abs/2011.12454