National Tsing Hua University

1130IEEM 513600

Deep Learning and Industrial Applications

Homework 3

Name: Ting, Kuan-Fu Student ID: 113003803

Due on 2025.04.10

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:
   - Number of defect classes.
   - Types of defect classes.
   - Number of images used in your dataset.
   - Distribution of training and test data.
   - Image dimensions.

Answer:

Category	Total Images	Train Images	Test Images	Ground Truth Images	# Defect Classes	Defect Types	Image Dimensions
bottle	229	209	20	0	0	None	900x900x3
cable	55	0	0	55	5	bent_wire, combined, cut_outer_insulation, missing_cable, missing_wire	1024x1024x3
capsule	459	219	132	108	5	crack, faulty_imprint, poke, scratch, squeeze	1000x1000x3
carpet	98	0	45	53	1	metal_contamination	1024x1024x3
grid	56	0	34	22	2	metal_contamination, thread	1024x1024x3
hazelnut	409	391	18	0	1	hole	1024x1024x3
leather	135	15	69	51	2	glue, poke	1024x1024x3
metal_nut	340	199	48	93	2	bent, scratch	700x700x3
pill	78	0	40	38	2	contamination, faulty_imprint	800x800x3
screw	374	119	160	95	5	manipulated_front, scratch_head, scratch_neck, thread_side, thread_top	1024x1024x3
tile	200	0	117	83	5	crack, glue_strip, gray_stroke, oil, rough	840x840x3
toothbrush	72	0	42	30	1	defective	1024x1024x3
transistor	323	213	80	30	2	bent_lead, misplaced	1024x1024x3
wood	39	0	29	10	1	hole	1024x1024x3
zipper	177	0	118	59	5	broken_teeth, fabric_interior, rough, split_teeth, squeezed_teeth	1024x1024x3

2. (30 points) Implement 4 different attempts to improve the model's performance trained on the dataset you choose in the 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.)

Answer:

1. Try data augmentation. The defect data is few. I try to do data augmentation to enhance training data variety. After data augmentation, the model performance can be improved and more stable if repeated training is used. I split the dataset into training/val/test datasets. I can really know model performance after splitting the dataset into training/val/test datasets.
2. Reference to the article (Li, H., Wu, J., Wu, L. Y., Chen, H., Liu, D., Wang, M., & Wang, P. (2023). Industrial Anomaly Detection and Localization Using Weakly-Supervised Residual Transformers. arXiv preprint arXiv:2306.03492.) I try to add BatchNorm and LeakyReLU, enlarge model depth, and Use additional Linear + ReLU + Dropout before the final classification to enhance expressiveness. The model performance is improved after enlarging the model size but requires a drop to decrease the overfitting.
3. Using Focal Loss, AdamW optimizer, and modifying learning rate and scheduler. After changing the loss function, learning rate, and scheduler, the training is faster and with higher accuracy.
4. Increase training epoch. After increasing the training epoch, I noticed that the val accuracy is stable and test accuracy is good, not overfitting. Increase training epoch. After increasing the training epoch, I noticed that the value accuracy is stable and test accuracy is good, not overfitting.

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.'

(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.

Answer:

(i) Long-tail distribution: A small number of classes have many samples. Many classes have few samples.

(ii) Reference to the article (Li, H., Wu, J., Wu, L. Y., Chen, H., Liu, D., Wang, M., & Wang, P. (2023). I try the following methods.

Defining Neural Networks

- Add BatchNorm to improve stability and convergence speed.
- Use LeakyReLU instead of ReLU to deal with the problem of neuron death when the sample size is small.
- Increase the number of channels layer by layer, for example, 32 → 64 → 128 → 256.
- Finally, AdaptiveAvgPool2d is still used to keep the input adaptability unchanged.
- Use additional Linear + ReLU + Dropout before final classification to enhance expressiveness.

Training the Neural Network

- Change Loss to Focal Loss to help deal with class imbalance.
- Use AdamW optimizer.
- Adjust the learning rate and scheduler.
- Added early stopping mechanism.
- Increase the number of epochs to 100 and use the warmup strategy.

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.)

Answer:

According to the paper (Chen, W., Yang, K., Yu, Z., Shi, Y., & Chen, C. P. (2024). A survey on imbalanced learning: latest research, applications, and future directions. Artificial Intelligence Review, 57(6), 137.) Suggest the following methods.

Data-level methods: Resample the dataset using techniques like SMOTE, Borderline-SMOTE, or undersampling.

Algorithm-level methods: Modify learning algorithms via cost-sensitive learning, adaptive thresholds, or weighted loss.

Hybrid methods: Combine data and algorithm-level strategies for better performance.

Ensemble learning: Use methods like Bagging, Boosting, or cost-sensitive ensembles to improve classification.

Deep long-tail learning: Address extreme class imbalance using class-balanced loss, information augmentation, and representation learning.

Applications: Applied in fault detection, fraud detection, medical diagnosis, and more.

Future directions: Focus on noisy data, data streams, dynamic imbalance, and unsupervised settings.

5. For the task of anomaly detection, it may be advantageous to employ more sophisticated computer visrion 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.
   (ii) (10 points) Why are these models suitable for fine-tuning for our custom dataset?

Answer:
(i)

Feature	**Object Detection**	**Segmentation**
Output	Bounding boxes + labels	Pixel-wise masks (class or instance)
Annotation	Coordinates + class labels	Per-pixel labels or polygons

Label complexity	Lower	Higher
Common format	VOC, COCO, YOLO	PNG masks, COCO polygons

(ii)

These models are pretrained on large-scale datasets, learning general features like edges, textures, and shapes. Fine-tuning them on MVTec helps them adapt to specific industrial textures and subtle variations.