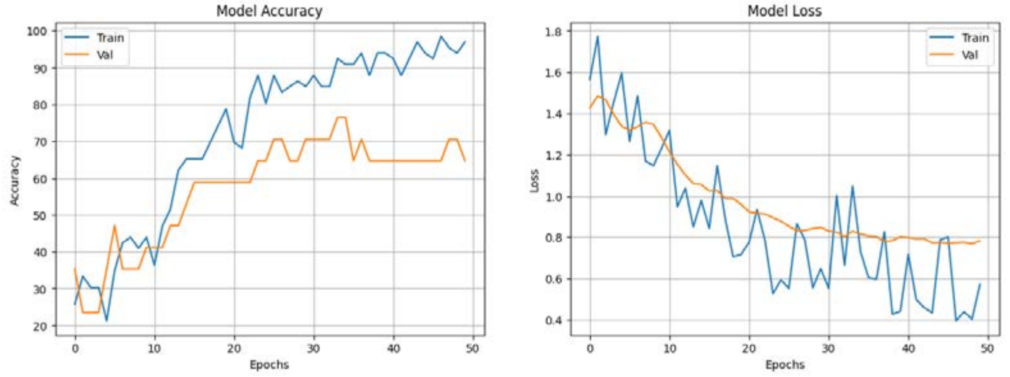
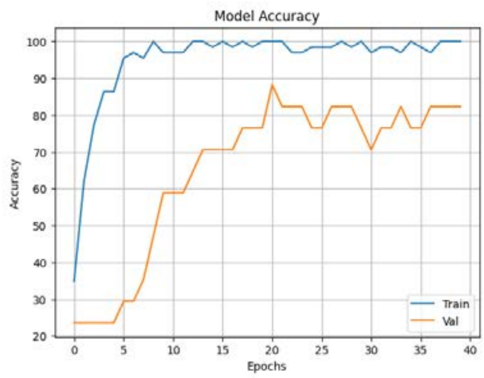
1. I choose the bottle dataset to carry out the experiments.

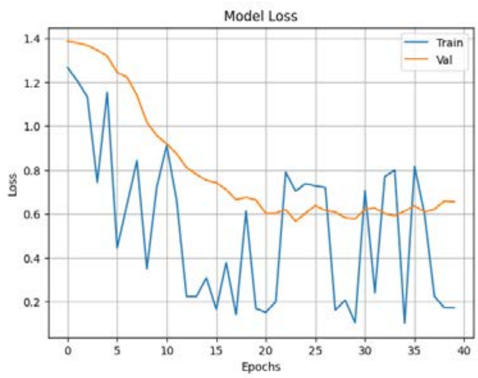
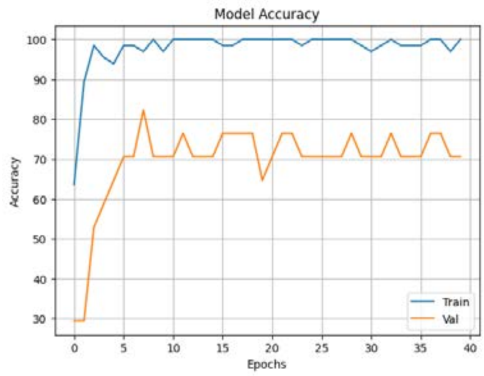
- Number of defect classes: 3
- Types of defect classes: ['broken_large', 'contamination', 'broken_small']
- Number of training images: 209
- Number of test images: 83
- Total images: 292
- Test distribution per class:
    - broken_large: 20
    - contamination: 21
    - broken_small: 22
    - good: 20
- Train distribution: good = 209
- Image size (width x height): (900, 900)

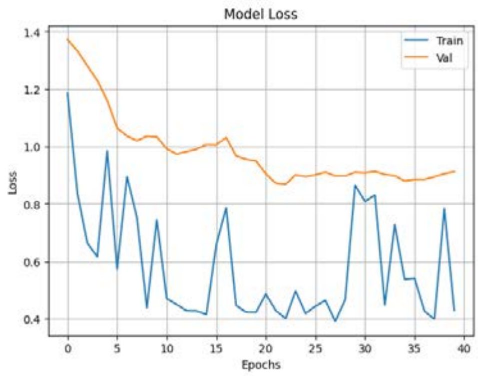
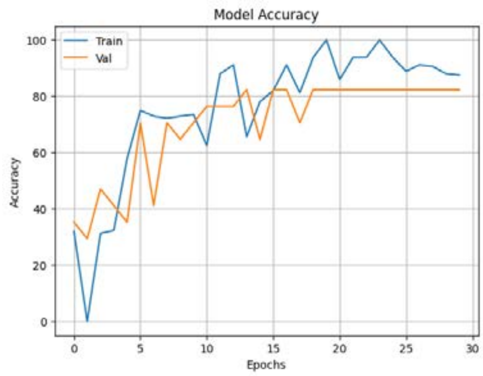
2.

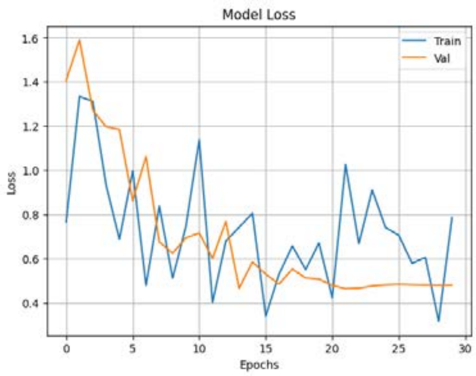
Four distinct approaches were evaluated for defect classification on the bottle dataset. Attempt 2, using ResNet50 with a custom multi-layer head, achieved the best performance with 88.24% accuracy, benefiting from a deeper backbone, batch normalization, dropout, and adaptive learning rate scheduling. Attempt 1 (ResNet18 fine-tuning) employed a two-stage training strategy and achieved 76.47% accuracy, showing improvement over the baseline through gradual unfreezing and CosineAnnealingLR. Attempt 3 leveraged an ensemble of ResNet34 and EfficientNet-B0, reaching 82.35% accuracy by combining complementary features and using label smoothing with CosineAnnealingWarmRestarts. Attempt 4, based on Vision Transformer (ViT-B/16), also achieved 82.35%, integrating strong augmentations, MixUp, Focal Loss, and class-balanced sampling. All models were trained using AdamW, with batch sizes ranging from 16 to 32, and learning rates tailored for backbone and head layers. Key contributors to performance were architecture depth, augmentation diversity, and effective regularization.

Approaches and its correspond Accuracy and Loss	
1	

<table>
<tr><td>2</td><td>


</td></tr>
<tr><td>3</td><td>


</td></tr>
<tr><td>4</td><td>


</td></tr>
</table>

3.

Long-tail distribution, or data imbalance, occurs when a few classes have many instances (head), while most others have few (tail), leading to the poor performance of machine learning models on minority classes. A paper published in 2022, "Balanced Meta-Softmax for Long-Tailed Visual Recognition" by Xiangyu Yue et al., proposes Balanced Meta-Softmax (BMS) to address this. BMS re-weights loss contributions using the effective number of samples per class, reducing overfitting to dominant classes and enhancing learning for rare ones. A meta-learning strategy further adjusts these weights during training to better match the dataset's characteristics.

As to apply to the MVTec AD dataset, which may have many “good” images and fewer defect samples, BMS can reduce bias towards the “good” class and boost recognition of underrepresented defects. This adaptive balancing could improve overall anomaly detection by ensuring better representation of all classes during training, especially in long-tailed setups typical of industrial anomaly datasets.

4.

When training an anomaly detection model mostly on “good” images, unsupervised or self-supervised methods are essential. One approach is using autoencoders to reconstruct normal images so that anomalies show higher reconstruction errors. Another method is feature embedding, where features extracted from “good” images form a distribution, and anomalies lie outside it. Generative models like GANs can also be used to generate normal samples, while anomalies are poorly reproduced. Last but not least, self-supervised learning tasks can learn meaningful representations from “good” data, aiding in identifying deviations. These techniques allow effective detection without requiring labeled anomaly samples.

5.

To apply object detection models like YOLO-World to the MVTec AD dataset, defective images must be annotated with bounding boxes around anomaly regions, labeled by defect type, while “good” images have no boxes. For segmentation models like SAM, pixel-level masks are required, where pixels within defect regions are labeled accordingly, and the rest are marked as background or “good”. Pre-trained models like YOLO-World and SAM are well-suited for fine-tuning due to their strong feature extraction capabilities learned from large datasets. Fine-tuning enables these models to adapt general visual patterns to detect specific anomalies in industrial images. This approach requires less data and time than training from scratch and offers better performance. Their built-in object localization and segmentation abilities also align well with the need to accurately detect and isolate defects in visual inspection tasks.