

Kaputt: A Large-Scale Dataset for Visual Defect Detection

Supplementary Material

I. VLM Prompts

We followed prompt engineering best practices for the zero-shot models and selected the one that performed best on a small validation set. The following prompt was used for the VLMs `Pixtral` and `Claude` to evaluate performance our dataset:

You are a highly skilled subject matter expert for inventory quality assurance and control. The presented image shows an item inside a tray. You have to determine whether the item is in pristine condition and can be sold as new and shipped to the customer as is, or whether it is damaged in any way and needs further attention before it can be shipped. Consider the following damage categories: crushed, tear, hole, deformed, ripped, deconstructed. Typical defects also include open boxes, or damaged and ripped packaging. Sometimes if the packaging is damaged, the item itself may become deconstructed and parts of the content may fall out. The container itself may be dirty which should not count as damage. However, if there is spillage that originated from a liquid item, then it must be called out as damage. Pay close attention to books and especially to corners of front or back pages. Moreover, items that are deconstructed, i.e. where the original packaging is damaged or fell off, should be flagged as damaged. In addition to the final decision specified by **DAMAGED** or **UNDAMAGED**, please also provide the severity score on a scale from 0 (pristine condition) to 10 (completely destroyed). Think step-by-step and provide the final response as json with keys "condition" and "severity".

II. CLIP Prompts

For the `CLIP` model 0-shot baseline, we tested the following prompts on the validation set:

- Image of an item with some damage. and Image of an item with no damage.
- Item without damage inside of a tray. and Item with damage inside of a tray.
- Image of an undamaged item. and Image of a damaged item.
- Image of an item without problems. and Image of an item with problems.

The last prompt performed best on the validation set and was used for the test set evaluation.

III. Dataset Details

We provide additional details about our defect taxonomy underlying our dataset (Figure 5) and experimental setup of our data collection station (Figure 6).

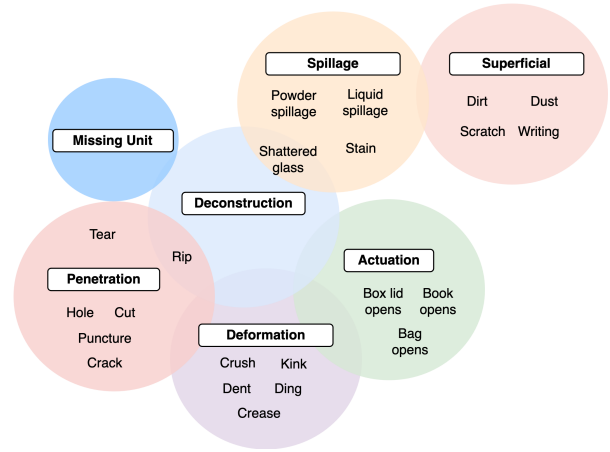


Figure 5. Overview of the defect types used to annotate defective samples (bold font) and related and colloquial characterization of the defect types (in bubbles). The proximity of the bubbles and their overlap indicates which defect types are similar/related.

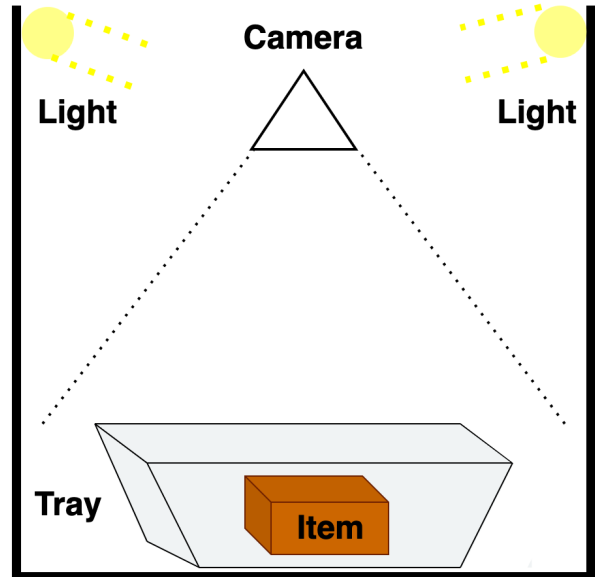


Figure 6. Schematic illustration of the data collection station. An overhead camera captures images from the top. Items reside in “trays”, common logistics containers.

IV. Error Analysis

The following tables systematically evaluate failure modes of all baselines. Table 4 provides quantitative statistics, while Tables 5-7 showcase exemplary correct and wrong predictions. Tables 8-35 further analyze model performance per item material and per defect type. Finally, Figures 7 and 8 show Precision-Recall and ROC curves.

Baseline	False Positives							False Negatives					
	Adversarial	Pack. Type	Query Distr.	Reference Var.	Minuscule	Unclear	Total	Subtle	Conf. Wrong	Reference Var.	Missing Unit	Unclear	Total
No training, no references (zero-shot, few-shot)													
CLIP	1	0	0	-	5	44	50	4	41	-	0	5	50
POMP	3	0	0	-	4	43	50	17	29	-	0	4	50
WinCLIP-zero	0	32	8	-	0	10	50	37	10	-	2	1	50
Claude-icl	8	18	13	-	0	11	50	42	8	-	0	0	50
Pixtral-zero	4	20	11	-	0	15	50	39	11	-	0	0	50
Pixtral-icl	7	11	5	-	0	27	50	38	12	-	0	0	50
No training, with references (few-shot, non-parametric, in-context learning)													
PatchCore50	1	0	0	42	1	6	50	40	1	0	0	9	50
WinCLIP-few	2	20	9	9	0	10	50	44	5	0	0	1	50
With training, no references (supervised/instruction fine-tuning)													
ResNet50	8	16	11	-	13	2	50	33	17	-	0	0	50
ViT-S	5	6	2	-	15	22	50	27	19	-	2	2	50
Pixtral-ft	1	11	7	-	3	28	50	40	10	-	0	0	50
AutoGluonMM	10	7	6	-	20	7	50	31	19	-	0	0	50
With training, with references (supervised with references, non-parametric with fine-tuning)													
PatchCore50-ft	1	3	2	36	1	7	50	38	5	0	7	0	50
AutoGluonMM-ref	3	9	3	16	4	15	50	21	20	3	0	6	50

Table 4. Error Analysis of Different Baseline Models. For each evaluated baseline model, we analyzed the top 50 (by classification score) false positives and false negatives. Upon visual examination, we classified them into 6 groups for the false positives and 5 groups for the false negatives. False positives were either *adversarial examples* (e.g. books with printed creases, oddly shaped), difficult *packaging types* (e.g. paper or plastic that wrinkles), *query distractions* (e.g. trash or tray scribbles), *reference variations* (e.g. confused or defective references), *minuscule defects*, or *unclear* reasons. False negatives were categorized into *confidently wrong* decisions by the baseline model, *missing units* (e.g. when multi-packs are captured with missing units that can only be identified by a model using references), or *minuscule*, *reference variations*, and *unclear* reasons, analogue to false positive cases.

Table 5. Failure and success modes of AutoGluonMM - With reference images. The images are arranged as squares in groups of four. The top-left example image shows the query image and the other three are the corresponding reference gallery images.

Model	TP		FP		TN		FN	
AutoGluonMM								
								
								
								

Table 6. Failure and success modes of PatchCore50-ft and WinCLIP - The images are arranged as squares in groups of four. The top-left images shows the query image, the top-right image the query with overlaid activation map, and the two bottom images are two of the reference images.

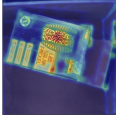
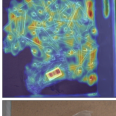
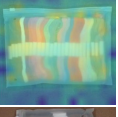
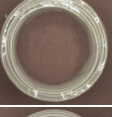
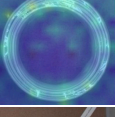
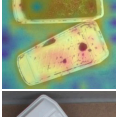
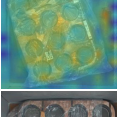
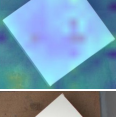
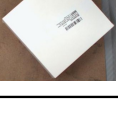
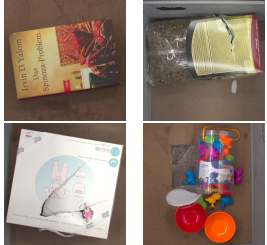
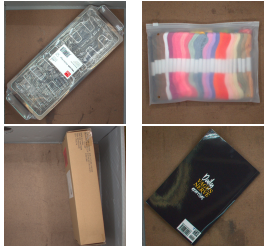
Model	TP		FP		TN		FN	
PatchCore50-ft								
								
								
								
WinCLIP								
								
								
								

Table 7. Failure and success modes of ResNet50, CLIP, and ViT-S - Those models were trained and evaluated without reference images. Each image shows a different example.

	TP	FP	TN	FN
ResNet50				
CLIP				
ViT-S				

V. Benchmark: Model Training Details

For the supervised methods we used validation set hyperparameter tuning. We employed standard augmentations (color, rotation, flip, zoom) for supervised baselines, and others already inherently apply similar concepts (WinCLIP’s windowed feature extraction). Unless stated otherwise, we used the default values for all methods. Finally, we addressed key issues, such as AD methods (PatchCore, WinCLIP) being distracted by the image background, by applying them to item crops rather than full tray crops.

CLIP and POMP For both the CLIP and POMP baselines we use the open-source implementation of POMP [21]. Thereby, we use the ViT-B/16 model, and 224×224 image input resolution. The evaluated prompts are listed in Section II. Example model predictions are shown in Table 7.

WinCLIP We use the open-source implementation of WinCLIP [13], based on a pre-trained openCLIP backbone (ViT-B/16 with 240×240 input resolution pretrained on LAION400M dataset). We use the original positive and negative text prompts (22 templates combined with 11 normal/anomaly states), using the generic item type `object`, e.g.

a photo of an object with damage
a photo of the object with damage
for visual inspection

Example model predictions are shown in Table 6.

Claude For our experiments with Claude we use the AWS Bedrock API. The model version is Claude Sonnet v3.5 and we use the default sampling parameters: temperature $t = 1$, nucleus sampling $top_p = 0.999$ and no top_k sampling.

Pixtral We use the HuggingFace library to fine-tune the Pixtral-12B model for our purpose. We use a constant learning rate of 3^{-5} and an effective batch size of 32. We run distributed training for one epoch on the entire training set across 8 A100 GPUs. The training data is constructed from the available labels as follows:

```
This item is {condition}.  
```json  
{
 "is_damaged_ge1": {true/false},
 "is_damaged_ge2": {true/false},
 "damage_intensity_median": {0, 1, 2},
 "condition": "{condition}",
 "severity": {severity}
}
```
```

where *condition* is set to DAMAGED or UNDAMAGED, according to the field *is_damaged_ge1*, and *severity* is the damage intensity median multiplied by 5 to bring it to the range between 0 and 10. We then use the same prompt as before (see Section I) and extract the value for *severity* as our damage confidence prediction.

PatchCore We used the PatchCore implementation from the anomalib package [3]. For each test image, a memory bank was built from up to 3 reference images (batch size = 1) with a coreset sampling ratio of 1.0 (no subsampling). Patch-level anomaly scores were computed using Euclidean distance to the $\max(1, \text{reference images})$ nearest neighbors. We used ResNet50 with 1024×1024 inputs, ImageNet normalization, and features extracted from layers 2 and 3. The image-level anomaly threshold was computed using anomalib’s Adaptive F1 method on a validation split containing both normal and anomalous samples. For this step, the memory bank was constructed from 10 normal validation images with a 0.01 coreset ratio. The resulting thresholds (43.76 for PatchCore50, 3.69 for PatchCore50-ft) were fixed during testing via anomalib’s Manual Threshold setting. Example predictions are shown in Table 6.

ResNet50 We fine-tune different ResNet backbones pre-training on ImageNet on a single V100 GPU using the `timm` library. After hyperparameter tuning, we arrived at the following settings (final choices in **bold**):

- Backbones: ResNet34, **ResNet50**, ResNet152
- Pre-training weights: ImageNet
- Input size: 1024×1024 , RGB
- Batch size 24 (ResNet152), **48 (ResNet50)**, 64 (ResNet34); Epochs: 20
- Optimizer: SGD with initial learning rate 0.005, momentum 0.8, ReduceLROnPlateau learning rate scheduler, reduction factor (0.15), patience epochs (1), minimum learning rate ($1e-5$)
- Dropout (dense layer for classification only, 0.5)
- Data augmentation: flip (horizontal/vertical), rotation ($0 \dots 180$), shift 10%, zoom 10%

Example model predictions are shown in Table 7.

ViT-S We fine-tuned a Vision Transformer with different pretraining weights on 8 V100 GPUs. After hyperparameter tuning, we arrived at the following settings:

- Backbones: ViT-small (22.1M parameters)[9].
- Pre-training weights: **DINOv2** [19]
- Input size: 1024×1024 , RGB, Patch size: 14×14 , RGB
- Batch size 8, Epochs: 30
- Optimizer: Adam with initial learning rate 5×10^{-6} (not scaled by Batch size), momentum 0.9, weight decay (0.05), Cosine scheduler
- Data augmentation: flip (horizontal/vertical) and AutoAugment set to `rand-m9-mstd0.5-inc1`

Example model predictions are shown in Table 7.

AutoGluon To train a classifier with reference images we employ AutoGluonMM, a freely available AutoML framework (<https://auto.gluon.ai>). Training is performed on a single V100 GPU using the default parameter settings. 20% of the training set are used for validation. The resulting multi-modal fusion MLP model uses a ViT

Table 8. Item Material KPIs for CLIP

| | Material | Total | Defects | AP _{any} [%] | AP _{major} [%] | AUROC | R@50%P [%] | R@1%FPR [%] |
|------|---------------------|-------|---------|-----------------------|-------------------------|--------------|---------------|-------------|
| book | other | 36 | 10 | 38.36 | 50.00 | 63.27 | 0.00 | - |
| | paper | 1664 | 671 | 46.85 | 18.81 | 57.02 | 11.92 | 1.79 |
| | plastic tight wrap | 127 | 38 | 30.80 | 23.08 | 52.69 | 0.00 | - |
| | cardboard | 3212 | 1261 | 41.84 | 12.77 | 53.43 | 0.79 | 1.11 |
| | other | 409 | 90 | 28.71 | 16.62 | 55.10 | 11.11 | 3.33 |
| | paper | 131 | 73 | 62.01 | 48.82 | 60.37 | 100.00 | 2.74 |
| | plastic bubble wrap | 187 | 36 | 35.59 | 20.39 | 70.37 | 36.11 | - |
| | plastic hard | 1032 | 148 | 17.27 | 14.35 | 53.98 | 0.00 | 1.35 |
| | plastic loose bag | 2419 | 684 | 29.94 | 19.52 | 52.66 | 0.15 | 1.32 |
| | plastic tight wrap | 850 | 195 | 29.05 | 21.55 | 59.73 | 0.51 | 1.03 |

Table 9. Damage Type KPIs for CLIP

| | Damage Type | Total | Defects | AP _{any} [%] | AP _{major} [%] | AUROC | R@50%P [%] | R@1%FPR [%] |
|--|--------------------|-------|---------|-----------------------|-------------------------|--------------|-------------|-------------|
| | penetration | 7827 | 966 | 15.22 | 9.28 | 57.96 | 0.21 | 1.55 |
| | deformation | 8646 | 1785 | 24.52 | 7.30 | 56.55 | 0.22 | 1.74 |
| | actuation | 8403 | 1542 | 21.43 | 11.83 | 55.72 | 0.00 | 1.75 |
| | superficial | 7585 | 724 | 11.25 | 2.29 | 56.08 | 0.00 | 1.38 |
| | spillage | 6943 | 82 | 1.33 | 0.92 | 52.15 | 0.00 | 2.44 |
| | deconstruction | 7888 | 1027 | 16.15 | 11.69 | 56.97 | 0.00 | 1.85 |
| | textttmissing unit | 6886 | 25 | 0.35 | 0.35 | 49.43 | 0.00 | 0.00 |

Table 10. Item Material KPIs for POMP

| | Material | Total | Defects | AP _{any} [%] | AP _{major} [%] | AUROC | R@50%P [%] | R@1%FPR [%] |
|------|---------------------|-------|---------|-----------------------|-------------------------|--------------|---------------|-------------|
| book | other | 36 | 10 | 29.22 | 16.67 | 52.31 | 0.00 | - |
| | paper | 1664 | 671 | 42.50 | 15.11 | 54.02 | 0.00 | 0.30 |
| | plastic tight wrap | 127 | 38 | 40.90 | 25.23 | 64.87 | 2.63 | 2.63 |
| | cardboard | 3212 | 1261 | 38.55 | 16.30 | 48.40 | 0.00 | 0.87 |
| | other | 409 | 90 | 24.89 | 17.41 | 52.19 | 0.00 | 2.22 |
| | paper | 131 | 73 | 56.09 | 46.61 | 49.71 | 100.00 | - |
| | plastic bubble wrap | 187 | 36 | 30.86 | 18.04 | 65.34 | 5.56 | 5.56 |
| | plastic hard | 1032 | 148 | 22.96 | 21.72 | 59.09 | 2.03 | 3.38 |
| | plastic loose bag | 2419 | 684 | 30.97 | 22.56 | 52.76 | 0.00 | 1.32 |
| | plastic tight wrap | 850 | 195 | 22.91 | 16.45 | 51.02 | 0.00 | 0.51 |

Table 11. Damage Type KPIs for POMP

| | Damage Type | Total | Defects | AP _{any} [%] | AP _{major} [%] | AUROC | R@50%P [%] | R@1%FPR [%] |
|--|----------------|-------|---------|-----------------------|-------------------------|--------------|------------|-------------|
| | penetration | 7827 | 966 | 13.10 | 8.72 | 51.11 | 0.00 | 1.14 |
| | deformation | 8646 | 1785 | 20.50 | 6.84 | 47.39 | 0.00 | 1.29 |
| | actuation | 8403 | 1542 | 19.89 | 11.79 | 52.37 | 0.00 | 1.10 |
| | superficial | 7585 | 724 | 8.26 | 1.47 | 42.67 | 0.00 | 1.11 |
| | spillage | 6943 | 82 | 3.04 | 2.85 | 65.33 | 0.00 | 3.66 |
| | deconstruction | 7888 | 1027 | 15.82 | 12.32 | 57.18 | 0.00 | 1.46 |
| | missing unit | 6886 | 25 | 0.54 | 0.54 | 57.76 | 0.00 | 0.00 |

Table 12. Item Material KPIs for WinCLIP-zero

| | Material | Total | Defects | AP _{any} [%] | AP _{major} [%] | AUROC | R@50%P [%] | R@1%FPR [%] |
|------|---------------------|-------|---------|-----------------------|-------------------------|--------------|---------------|--------------|
| book | other | 36 | 10 | 51.57 | 33.33 | 65.77 | 60.00 | - |
| | paper | 1664 | 671 | 42.26 | 14.91 | 53.46 | 0.00 | 0.89 |
| | plastic tight wrap | 127 | 38 | 36.84 | 23.23 | 57.61 | 5.26 | 2.63 |
| | cardboard | 3212 | 1261 | 42.02 | 15.74 | 51.22 | 5.79 | 2.22 |
| | other | 409 | 90 | 28.96 | 22.61 | 58.06 | 0.00 | 1.11 |
| | paper | 131 | 73 | 75.00 | 72.04 | 66.36 | 100.00 | 13.70 |
| | plastic bubble wrap | 187 | 36 | 39.00 | 25.27 | 72.46 | 16.67 | 8.33 |
| | plastic hard | 1032 | 148 | 23.12 | 19.58 | 60.02 | 0.00 | 2.70 |
| | plastic loose bag | 2419 | 684 | 31.13 | 20.46 | 54.61 | 1.32 | 1.61 |
| | plastic tight wrap | 850 | 195 | 23.87 | 17.21 | 47.32 | 2.05 | 2.05 |

Table 13. Damage Type KPIs for WinCLIP-zero

| | Damage Type | Total | Defects | AP _{any} [%] | AP _{major} [%] | AUROC | R@50%P [%] | R@1%FPR [%] |
|--|----------------|-------|---------|-----------------------|-------------------------|--------------|-------------|-------------|
| | penetration | 7827 | 966 | 15.87 | 11.44 | 55.90 | 0.10 | 2.17 |
| | deformation | 8646 | 1785 | 18.29 | 5.30 | 45.77 | 0.00 | 0.39 |
| | actuation | 8403 | 1542 | 22.36 | 13.46 | 56.72 | 0.07 | 1.88 |
| | superficial | 7585 | 724 | 8.49 | 1.21 | 45.85 | 0.00 | 0.14 |
| | spillage | 6943 | 82 | 2.12 | 1.76 | 64.97 | 0.00 | 1.22 |
| | deconstruction | 7888 | 1027 | 19.40 | 14.90 | 62.12 | 0.00 | 2.82 |
| | missing unit | 6886 | 25 | 0.85 | 0.85 | 49.95 | 0.00 | 4.00 |

Table 14. Item Material KPIs for Claude-icl

| | Material | Total | Defects | AP _{any} [%] | AP _{major} [%] | AUROC | R@50%P [%] | R@1%FPR [%] |
|------|---------------------|-------|---------|-----------------------|-------------------------|--------------|---------------|--------------|
| book | other | 36 | 10 | 42.22 | 100.00 | 58.46 | 20.00 | 20.00 |
| | paper | 1664 | 671 | 43.09 | 20.03 | 53.08 | 11.03 | 2.24 |
| | plastic tight wrap | 127 | 38 | 42.42 | 31.41 | 61.75 | 26.32 | 2.63 |
| | cardboard | 3212 | 1261 | 46.29 | 32.53 | 58.07 | 30.29 | 1.43 |
| | other | 409 | 90 | 27.92 | 21.71 | 56.08 | 5.56 | 3.33 |
| | paper | 131 | 73 | 69.51 | 65.91 | 66.37 | 100.00 | - |
| | plastic bubble wrap | 187 | 36 | 29.05 | 24.13 | 62.99 | 0.00 | - |
| | plastic hard | 1032 | 148 | 24.32 | 22.95 | 67.30 | 0.00 | 0.68 |
| | plastic loose bag | 2419 | 684 | 34.07 | 25.43 | 59.39 | 0.00 | 0.44 |
| | plastic tight wrap | 850 | 195 | 26.19 | 19.90 | 55.89 | 0.00 | 0.00 |

Table 15. Damage Type KPIs for Claude-icl

| | Damage Type | Total | Defects | AP _{any} [%] | AP _{major} [%] | AUROC | R@50%P [%] | R@1%FPR [%] |
|--|----------------|-------|---------|-----------------------|-------------------------|--------------|------------|-------------|
| | penetration | 7827 | 966 | 19.41 | 16.41 | 62.79 | 0.00 | 0.52 |
| | deformation | 8646 | 1785 | 21.85 | 9.70 | 53.40 | 0.00 | 0.17 |
| | actuation | 8403 | 1542 | 23.95 | 18.46 | 59.63 | 0.00 | 0.39 |
| | superficial | 7585 | 724 | 10.68 | 3.21 | 55.26 | 0.00 | 0.28 |
| | spillage | 6943 | 82 | 6.57 | 6.90 | 77.89 | 0.00 | 3.66 |
| | deconstruction | 7888 | 1027 | 22.20 | 19.83 | 65.18 | 0.00 | 0.58 |
| | missing unit | 6886 | 25 | 0.50 | 0.50 | 57.94 | 0.00 | 0.00 |

backbone and has 97.4M parameters. Example model predictions are shown in Table 5.

Table 16. Item Material KPIs for Pixtral-zero

| | Material | Total | Defects | AP _{any} [%] | AP _{major} [%] | AUROC | R@50%P [%] | R@1%FPR [%] |
|------|---------------------|-------|---------|-----------------------|-------------------------|--------------|---------------|--------------|
| book | other | 36 | 10 | 46.11 | 3.70 | 67.31 | 50.00 | 10.00 |
| | paper | 1664 | 671 | 40.97 | 15.13 | 49.75 | 3.28 | 0.30 |
| | plastic tight wrap | 127 | 38 | 31.42 | 17.24 | 53.02 | 0.00 | - |
| | cardboard | 3212 | 1261 | 40.57 | 14.07 | 51.28 | 3.81 | 1.03 |
| | other | 409 | 90 | 22.89 | 16.38 | 50.36 | 0.00 | 1.11 |
| | paper | 131 | 73 | 60.88 | 52.79 | 52.83 | 100.00 | 8.22 |
| | plastic bubble wrap | 187 | 36 | 22.84 | 11.32 | 51.38 | 0.00 | 2.78 |
| | plastic hard | 1032 | 148 | 15.98 | 14.21 | 50.18 | 0.00 | 0.00 |
| | plastic loose bag | 2419 | 684 | 30.49 | 20.64 | 53.87 | 0.44 | 1.02 |
| | plastic tight wrap | 850 | 195 | 25.05 | 17.21 | 49.87 | 2.56 | 2.56 |

Table 17. Damage Type KPIs for Pixtral-zero

| | Damage Type | Total | Defects | AP _{any} [%] | AP _{major} [%] | AUROC | R@50%P [%] | R@1%FPR [%] |
|--|----------------|-------|---------|-----------------------|-------------------------|--------------|------------|-------------|
| | penetration | 7827 | 966 | 13.77 | 8.51 | 52.43 | 0.00 | 1.55 |
| | deformation | 8646 | 1785 | 20.21 | 5.30 | 49.05 | 0.00 | 0.39 |
| | actuation | 8403 | 1542 | 19.46 | 10.80 | 52.12 | 0.00 | 0.78 |
| | superficial | 7585 | 724 | 9.49 | 1.39 | 50.03 | 0.00 | 0.14 |
| | spillage | 6943 | 82 | 4.01 | 3.38 | 66.58 | 0.00 | 3.66 |
| | deconstruction | 7888 | 1027 | 14.88 | 11.80 | 53.09 | 0.00 | 1.27 |
| | missing unit | 6886 | 25 | 0.35 | 0.35 | 42.24 | 0.00 | 0.00 |

Table 18. Item Material KPIs for Pixtral-icl

| | Material | Total | Defects | AP _{any} [%] | AP _{major} [%] | AUROC | R@50%P [%] | R@1%FPR [%] |
|------|---------------------|-------|---------|-----------------------|-------------------------|--------------|---------------|-------------|
| book | other | 36 | 10 | 44.88 | 8.33 | 70.77 | 20.00 | - |
| | paper | 1664 | 671 | 39.86 | 12.68 | 49.42 | 0.00 | 0.45 |
| | plastic tight wrap | 127 | 38 | 29.85 | 17.74 | 51.27 | 0.00 | - |
| | cardboard | 3212 | 1261 | 39.14 | 10.71 | 49.87 | 0.00 | 0.16 |
| | other | 409 | 90 | 19.31 | 13.33 | 43.85 | 0.00 | - |
| | paper | 131 | 73 | 58.52 | 50.67 | 54.97 | 100.00 | - |
| | plastic bubble wrap | 187 | 36 | 25.43 | 23.28 | 54.76 | 2.78 | 2.78 |
| | plastic hard | 1032 | 148 | 18.13 | 15.42 | 58.25 | 0.00 | 0.68 |
| | plastic loose bag | 2419 | 684 | 33.16 | 22.33 | 58.50 | 0.00 | 0.88 |
| | plastic tight wrap | 850 | 195 | 23.20 | 15.73 | 50.64 | 0.00 | 0.00 |

Table 19. Damage Type KPIs for Pixtral-icl

| | Damage Type | Total | Defects | AP _{any} [%] | AP _{major} [%] | AUROC | R@50%P [%] | R@1%FPR [%] |
|--|----------------|-------|---------|-----------------------|-------------------------|--------------|------------|-------------|
| | penetration | 7827 | 966 | 13.60 | 7.96 | 54.87 | 0.00 | 0.52 |
| | deformation | 8646 | 1785 | 19.00 | 4.58 | 45.86 | 0.00 | 0.56 |
| | actuation | 8403 | 1542 | 20.05 | 11.06 | 54.62 | 0.00 | 0.91 |
| | superficial | 7585 | 724 | 9.06 | 1.02 | 47.62 | 0.00 | 0.55 |
| | spillage | 6943 | 82 | 1.59 | 1.39 | 59.30 | 0.00 | 1.22 |
| | deconstruction | 7888 | 1027 | 15.56 | 12.10 | 58.35 | 0.00 | 1.17 |
| | missing unit | 6886 | 25 | 0.47 | 0.47 | 56.67 | 0.00 | 0.00 |

Table 20. Item Material KPIs for PatchCore50

| | Material | Total | Defects | AP _{any} [%] | AP _{major} [%] | AUROC | R@50%P [%] | R@1%FPR [%] |
|------|---------------------|-------|---------|-----------------------|-------------------------|--------------|---------------|-------------|
| | book other | 36 | 10 | 32.12 | 20.00 | 56.92 | 0.00 | - |
| | book paper | 1664 | 671 | 45.20 | 17.57 | 56.44 | 5.22 | 2.09 |
| book | plastic tight wrap | 127 | 38 | 41.24 | 39.53 | 58.87 | 10.53 | 5.26 |
| | cardboard | 3212 | 1261 | 42.16 | 12.17 | 52.75 | 5.31 | 2.46 |
| | other | 409 | 90 | 30.50 | 22.62 | 60.91 | 3.33 | 3.33 |
| | paper | 131 | 73 | 53.02 | 43.31 | 46.95 | 100.00 | - |
| | plastic bubble wrap | 187 | 36 | 32.21 | 24.43 | 71.12 | 0.00 | 2.78 |
| | plastic hard | 1032 | 148 | 20.79 | 18.75 | 63.25 | 0.00 | 3.38 |
| | plastic loose bag | 2419 | 684 | 31.42 | 21.86 | 54.53 | 0.59 | 1.02 |
| | plastic tight wrap | 850 | 195 | 26.97 | 21.36 | 54.77 | 0.00 | 2.56 |

Table 21. Damage Type KPIsfor PatchCore50

| | Damage Type | Total | Defects | AP _{any} [%] | AP _{major} [%] | AUROC | R@50%P [%] | R@1%FPR [%] |
|--|----------------|-------|---------|-----------------------|-------------------------|--------------|-------------|-------------|
| | penetration | 7827 | 966 | 14.45 | 8.55 | 56.60 | 0.00 | 1.66 |
| | deformation | 8646 | 1785 | 24.37 | 6.83 | 53.99 | 1.18 | 2.58 |
| | actuation | 8403 | 1542 | 20.70 | 12.00 | 55.57 | 0.00 | 1.17 |
| | superficial | 7585 | 724 | 12.80 | 1.90 | 52.06 | 1.52 | 4.28 |
| | spillage | 6943 | 82 | 1.75 | 1.31 | 60.44 | 0.00 | 2.44 |
| | deconstruction | 7888 | 1027 | 16.05 | 12.78 | 59.36 | 0.00 | 1.46 |
| | missing unit | 6886 | 25 | 0.50 | 0.50 | 59.88 | 0.00 | 0.00 |

Table 22. Item Material KPIs for WinCLIP-few

| | Material | Total | Defects | AP _{any} [%] | AP _{major} [%] | AUROC | R@50%P [%] | R@1%FPR [%] |
|------|---------------------|-------|---------|-----------------------|-------------------------|--------------|---------------|-------------|
| | book other | 36 | 10 | 49.95 | 33.33 | 64.62 | 60.00 | - |
| | book paper | 1664 | 671 | 42.73 | 15.50 | 53.90 | 0.89 | 1.19 |
| book | plastic tight wrap | 127 | 38 | 37.84 | 26.60 | 57.78 | 5.26 | 5.26 |
| | cardboard | 3212 | 1261 | 42.12 | 16.58 | 50.97 | 6.74 | 2.30 |
| | other | 409 | 90 | 29.01 | 22.66 | 58.62 | 0.00 | 2.22 |
| | paper | 131 | 73 | 73.74 | 70.17 | 66.01 | 100.00 | 5.48 |
| | plastic bubble wrap | 187 | 36 | 39.74 | 25.79 | 73.11 | 11.11 | 8.33 |
| | plastic hard | 1032 | 148 | 23.21 | 19.81 | 60.19 | 0.00 | 1.35 |
| | plastic loose bag | 2419 | 684 | 31.24 | 20.39 | 54.63 | 1.75 | 1.75 |
| | plastic tight wrap | 850 | 195 | 24.32 | 17.81 | 47.60 | 1.54 | 2.05 |

Table 23. Damage Type KPIsfor WinCLIP-few

| | Damage Type | Total | Defects | AP _{any} [%] | AP _{major} [%] | AUROC | R@50%P [%] | R@1%FPR [%] |
|--|----------------|-------|---------|-----------------------|-------------------------|--------------|------------|-------------|
| | penetration | 7827 | 966 | 15.96 | 11.35 | 55.94 | 0.00 | 2.69 |
| | deformation | 8646 | 1785 | 18.47 | 5.51 | 46.01 | 0.00 | 0.50 |
| | actuation | 8403 | 1542 | 22.48 | 13.56 | 56.87 | 0.00 | 2.08 |
| | superficial | 7585 | 724 | 8.55 | 1.25 | 45.81 | 0.00 | 0.69 |
| | spillage | 6943 | 82 | 2.18 | 1.80 | 65.47 | 0.00 | 1.22 |
| | deconstruction | 7888 | 1027 | 19.49 | 14.98 | 62.20 | 0.00 | 3.12 |
| | missing unit | 6886 | 25 | 0.80 | 0.80 | 49.99 | 0.00 | 4.00 |

Table 24. Item Material KPIs for ResNet50

| | Material | Total | Defects | AP _{any} [%] | AP _{major} [%] | AUROC | R@50%P [%] | R@1%FPR [%] |
|------|---------------------|-------|---------|-----------------------|-------------------------|--------------|---------------|--------------|
| | book other | 36 | 10 | 73.32 | 25.00 | 79.62 | 60.00 | 50.00 |
| | book paper | 1664 | 671 | 82.36 | 72.92 | 86.16 | 95.53 | 24.59 |
| book | plastic tight wrap | 127 | 38 | 82.54 | 82.69 | 85.99 | 81.58 | 44.74 |
| | cardboard | 3212 | 1261 | 81.11 | 73.70 | 84.66 | 92.94 | 27.44 |
| | other | 409 | 90 | 60.83 | 60.46 | 80.24 | 66.67 | 16.67 |
| | paper | 131 | 73 | 88.14 | 86.22 | 85.00 | 100.00 | 28.77 |
| | plastic bubble wrap | 187 | 36 | 92.21 | 86.03 | 97.15 | 97.22 | 66.67 |
| | plastic hard | 1032 | 148 | 76.58 | 79.37 | 90.78 | 79.05 | 47.97 |
| | plastic loose bag | 2419 | 684 | 83.28 | 79.69 | 91.55 | 93.42 | 33.63 |
| | plastic tight wrap | 850 | 195 | 76.01 | 78.50 | 87.23 | 81.54 | 37.44 |

Table 25. Damage Type KPIsfor ResNet50

| | Damage Type | Total | Defects | AP _{any} [%] | AP _{major} [%] | AUROC | R@50%P [%] | R@1%FPR [%] |
|--|----------------|-------|---------|-----------------------|-------------------------|--------------|--------------|--------------|
| | penetration | 7827 | 966 | 64.99 | 65.24 | 88.54 | 68.84 | 32.82 |
| | deformation | 8646 | 1785 | 72.46 | 60.20 | 89.19 | 83.31 | 27.90 |
| | actuation | 8403 | 1542 | 75.88 | 73.04 | 90.74 | 83.27 | 35.80 |
| | superficial | 7585 | 724 | 52.97 | 36.00 | 87.88 | 54.14 | 24.59 |
| | spillage | 6943 | 82 | 18.79 | 16.49 | 82.02 | 6.10 | 29.27 |
| | deconstruction | 7888 | 1027 | 73.02 | 73.43 | 91.10 | 80.33 | 39.44 |
| | missing unit | 6886 | 25 | 15.40 | 15.40 | 87.07 | 8.00 | 32.00 |

Table 26. Item Material KPIs for ViT-S

| | Material | Total | Defects | AP _{any} [%] | AP _{major} [%] | AUROC | R@50%P [%] | R@1%FPR [%] |
|------|---------------------|-------|---------|-----------------------|-------------------------|--------------|---------------|--------------|
| | book other | 36 | 10 | 88.52 | 100.00 | 94.23 | 100.00 | 50.00 |
| | book paper | 1664 | 671 | 90.68 | 92.63 | 92.35 | 99.40 | 49.03 |
| book | plastic tight wrap | 127 | 38 | 92.54 | 94.40 | 94.94 | 94.74 | 65.79 |
| | cardboard | 3212 | 1261 | 90.22 | 94.33 | 91.87 | 98.57 | 55.91 |
| | other | 409 | 90 | 77.69 | 79.73 | 88.74 | 84.44 | 48.89 |
| | paper | 131 | 73 | 96.72 | 97.14 | 95.77 | 100.00 | 69.86 |
| | plastic bubble wrap | 187 | 36 | 97.22 | 98.44 | 99.23 | 100.00 | 80.56 |
| | plastic hard | 1032 | 148 | 92.88 | 94.31 | 97.62 | 95.95 | 79.05 |
| | plastic loose bag | 2419 | 684 | 92.51 | 92.41 | 96.29 | 97.52 | 65.64 |
| | plastic tight wrap | 850 | 195 | 86.01 | 90.05 | 93.16 | 91.80 | 57.44 |

Table 27. Damage Type KPIsfor ViT-S

| | Damage Type | Total | Defects | AP _{any} [%] | AP _{major} [%] | AUROC | R@50%P [%] | R@1%FPR [%] |
|--|----------------|-------|---------|-----------------------|-------------------------|--------------|--------------|--------------|
| | penetration | 7827 | 966 | 83.17 | 88.80 | 95.11 | 86.65 | 64.49 |
| | deformation | 8646 | 1785 | 85.14 | 89.08 | 94.08 | 92.89 | 55.35 |
| | actuation | 8403 | 1542 | 89.18 | 92.78 | 96.01 | 93.52 | 68.81 |
| | superficial | 7585 | 724 | 70.61 | 75.96 | 93.63 | 73.07 | 46.55 |
| | spillage | 6943 | 82 | 44.74 | 45.77 | 89.86 | 48.78 | 52.44 |
| | deconstruction | 7888 | 1027 | 89.53 | 91.98 | 96.88 | 92.50 | 75.56 |
| | missing unit | 6886 | 25 | 59.23 | 59.23 | 89.78 | 60.00 | 68.00 |

Table 28. Item Material KPIs for Pixtral-ft

| | Material | Total | Defects | AP _{any} [%] | AP _{major} [%] | AUROC | R@50%P [%] | R@1%FPR [%] |
|------|---------------------|-------|---------|-----------------------|-------------------------|--------------|---------------|-------------|
| book | other | 36 | 10 | 27.78 | 3.70 | 50.00 | 0.00 | - |
| | paper | 1664 | 671 | 40.59 | 12.59 | 50.22 | 0.45 | 0.45 |
| | plastic tight wrap | 127 | 38 | 29.92 | 17.59 | 50.00 | 0.00 | - |
| | cardboard | 3212 | 1261 | 41.24 | 16.88 | 52.06 | 5.95 | 1.75 |
| | other | 409 | 90 | 22.01 | 15.61 | 49.37 | 0.00 | - |
| | paper | 131 | 73 | 55.73 | 41.41 | 49.14 | 100.00 | - |
| | plastic bubble wrap | 187 | 36 | 19.25 | 11.18 | 50.00 | 0.00 | - |
| | plastic hard | 1032 | 148 | 17.50 | 16.05 | 52.32 | 5.41 | 5.41 |
| | plastic loose bag | 2419 | 684 | 30.84 | 22.39 | 51.89 | 3.95 | 3.95 |
| | plastic tight wrap | 850 | 195 | 24.13 | 17.07 | 50.77 | 1.54 | 1.54 |

Table 29. Damage Type KPIs for Pixtral-ft

| | Damage Type | Total | Defects | AP _{any} [%] | AP _{major} [%] | AUROC | R@50%P [%] | R@1%FPR [%] |
|--|----------------|-------|---------|-----------------------|-------------------------|--------------|-------------|-------------|
| | penetration | 7827 | 966 | 14.35 | 9.84 | 51.86 | 2.69 | 4.45 |
| | deformation | 8646 | 1785 | 22.18 | 7.78 | 51.65 | 4.03 | 4.03 |
| | actuation | 8403 | 1542 | 20.71 | 13.79 | 51.78 | 4.28 | 4.28 |
| | superficial | 7585 | 724 | 10.18 | 1.52 | 51.22 | 0.00 | 3.18 |
| | spillage | 6943 | 82 | 1.36 | 1.04 | 51.46 | 0.00 | 3.66 |
| | deconstruction | 7888 | 1027 | 16.71 | 14.52 | 52.56 | 5.84 | 5.84 |
| | missing unit | 6886 | 25 | 0.36 | 0.36 | 49.63 | 0.00 | 0.00 |

Table 30. Item Material KPIs for AutoGloUnMM

| | Material | Total | Defects | AP _{any} [%] | AP _{major} [%] | AUROC | R@50%P [%] | R@1%FPR [%] |
|------|---------------------|-------|---------|-----------------------|-------------------------|--------------|---------------|--------------|
| book | other | 36 | 10 | 91.18 | 100.00 | 95.77 | 100.00 | 60.00 |
| | paper | 1664 | 671 | 87.99 | 89.72 | 90.00 | 98.21 | 37.71 |
| | plastic tight wrap | 127 | 38 | 88.07 | 89.84 | 92.82 | 94.74 | 44.74 |
| | cardboard | 3212 | 1261 | 87.11 | 89.92 | 89.69 | 97.70 | 41.08 |
| | other | 409 | 90 | 73.62 | 73.75 | 87.85 | 84.44 | 31.11 |
| | paper | 131 | 73 | 94.08 | 96.07 | 91.73 | 100.00 | 45.21 |
| | plastic bubble wrap | 187 | 36 | 93.55 | 89.45 | 98.73 | 100.00 | 77.78 |
| | plastic hard | 1032 | 148 | 88.37 | 88.35 | 95.21 | 89.87 | 74.32 |
| | plastic loose bag | 2419 | 684 | 90.27 | 89.32 | 94.59 | 95.91 | 55.85 |
| | plastic tight wrap | 850 | 195 | 82.71 | 86.68 | 92.10 | 91.80 | 48.72 |

Table 31. Damage Type KPIs for AutoGloUnMM

| | Damage Type | Total | Defects | AP _{any} [%] | AP _{major} [%] | AUROC | R@50%P [%] | R@1%FPR [%] |
|--|----------------|-------|---------|-----------------------|-------------------------|--------------|--------------|--------------|
| | penetration | 7827 | 966 | 76.85 | 81.41 | 92.75 | 79.81 | 50.83 |
| | deformation | 8646 | 1785 | 82.15 | 83.23 | 92.99 | 90.20 | 43.59 |
| | actuation | 8403 | 1542 | 85.41 | 87.45 | 94.55 | 91.38 | 55.97 |
| | superficial | 7585 | 724 | 65.02 | 64.72 | 92.38 | 69.89 | 32.46 |
| | spillage | 6943 | 82 | 41.72 | 43.75 | 86.44 | 39.02 | 46.34 |
| | deconstruction | 7888 | 1027 | 82.94 | 84.86 | 94.05 | 86.47 | 60.37 |
| | missing unit | 6886 | 25 | 37.16 | 37.16 | 84.94 | 36.00 | 48.00 |

Table 32. Item Material KPIs for PatchCore50-ft

| | Material | Total | Defects | AP _{any} [%] | AP _{major} [%] | AUROC | R@50%P [%] | R@1%FPR [%] |
|------|---------------------|-------|---------|-----------------------|-------------------------|--------------|---------------|-------------|
| | book other | 36 | 10 | 38.53 | 50.00 | 63.46 | 0.00 | - |
| | book paper | 1664 | 671 | 52.97 | 21.87 | 63.45 | 51.12 | 3.58 |
| book | plastic tight wrap | 127 | 38 | 52.01 | 42.58 | 72.30 | 42.11 | 7.90 |
| | cardboard | 3212 | 1261 | 45.20 | 15.65 | 56.09 | 10.15 | 3.01 |
| | other | 409 | 90 | 31.54 | 23.47 | 60.43 | 6.67 | 4.44 |
| | paper | 131 | 73 | 59.69 | 51.09 | 50.87 | 100.00 | 2.74 |
| | plastic bubble wrap | 187 | 36 | 36.15 | 32.39 | 76.29 | 11.11 | 0.00 |
| | plastic hard | 1032 | 148 | 23.11 | 21.63 | 62.94 | 6.76 | 6.76 |
| | plastic loose bag | 2419 | 684 | 36.65 | 25.60 | 63.65 | 0.00 | 1.61 |
| | plastic tight wrap | 850 | 195 | 30.72 | 21.10 | 62.62 | 0.00 | 0.51 |

Table 33. Damage Type KPIsfor PatchCore50-ft

| | Damage Type | Total | Defects | AP _{any} [%] | AP _{major} [%] | AUROC | R@50%P [%] | R@1%FPR [%] |
|--|----------------|-------|---------|-----------------------|-------------------------|--------------|-------------|-------------|
| | penetration | 7827 | 966 | 17.30 | 10.33 | 61.46 | 0.00 | 2.07 |
| | deformation | 8646 | 1785 | 27.28 | 8.27 | 59.16 | 0.56 | 2.30 |
| | actuation | 8403 | 1542 | 25.02 | 14.93 | 62.63 | 0.00 | 1.88 |
| | superficial | 7585 | 724 | 11.61 | 1.58 | 53.25 | 0.00 | 2.21 |
| | spillage | 6943 | 82 | 1.83 | 1.45 | 62.05 | 0.00 | 2.44 |
| | deconstruction | 7888 | 1027 | 19.16 | 15.30 | 66.08 | 0.00 | 1.36 |
| | missing unit | 6886 | 25 | 0.55 | 0.55 | 61.83 | 0.00 | 0.00 |

Table 34. Item Material KPIs for AutoGluonMM-gal

| | Material | Total | Defects | AP _{any} [%] | AP _{major} [%] | AUROC | R@50%P [%] | R@1%FPR [%] |
|------|---------------------|-------|---------|-----------------------|-------------------------|--------------|---------------|--------------|
| | book other | 36 | 10 | 50.70 | 25.00 | 79.62 | 70.00 | - |
| | book paper | 1664 | 671 | 75.12 | 66.10 | 81.07 | 93.29 | 13.86 |
| book | plastic tight wrap | 127 | 38 | 85.85 | 81.44 | 92.58 | 97.37 | 26.32 |
| | cardboard | 3212 | 1261 | 70.14 | 58.16 | 79.10 | 92.07 | 13.64 |
| | other | 409 | 90 | 56.90 | 56.83 | 76.70 | 53.33 | 22.22 |
| | paper | 131 | 73 | 79.82 | 74.69 | 80.99 | 100.00 | 5.48 |
| | plastic bubble wrap | 187 | 36 | 49.04 | 36.65 | 85.67 | 75.00 | 2.78 |
| | plastic hard | 1032 | 148 | 64.01 | 64.40 | 86.34 | 66.22 | 29.73 |
| | plastic loose bag | 2419 | 684 | 73.77 | 68.66 | 88.06 | 91.96 | 14.77 |
| | plastic tight wrap | 850 | 195 | 65.69 | 63.66 | 84.37 | 71.28 | 16.41 |

Table 35. Damage Type KPIsfor AutoGluonMM-gal

| | Damage Type | Total | Defects | AP _{any} [%] | AP _{major} [%] | AUROC | R@50%P [%] | R@1%FPR [%] |
|--|----------------|-------|---------|-----------------------|-------------------------|--------------|--------------|--------------|
| | penetration | 7827 | 966 | 48.35 | 47.07 | 84.40 | 46.27 | 16.87 |
| | deformation | 8646 | 1785 | 58.47 | 43.24 | 85.04 | 71.99 | 12.21 |
| | actuation | 8403 | 1542 | 62.11 | 57.57 | 86.93 | 75.36 | 16.73 |
| | superficial | 7585 | 724 | 32.67 | 19.27 | 83.59 | 2.62 | 7.18 |
| | spillage | 6943 | 82 | 6.34 | 6.65 | 69.51 | 0.00 | 18.29 |
| | deconstruction | 7888 | 1027 | 57.13 | 56.43 | 86.50 | 65.92 | 19.96 |
| | missing unit | 6886 | 25 | 16.27 | 16.27 | 84.46 | 8.00 | 36.00 |

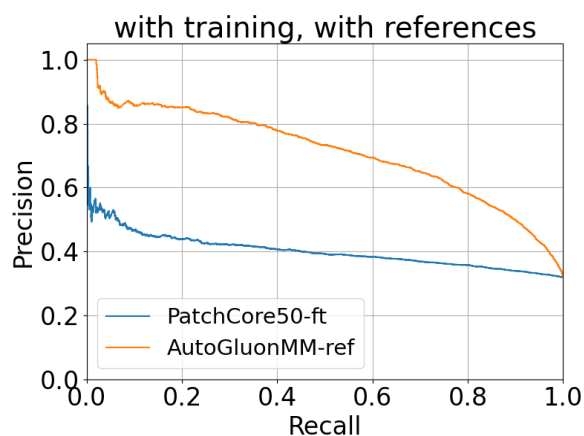
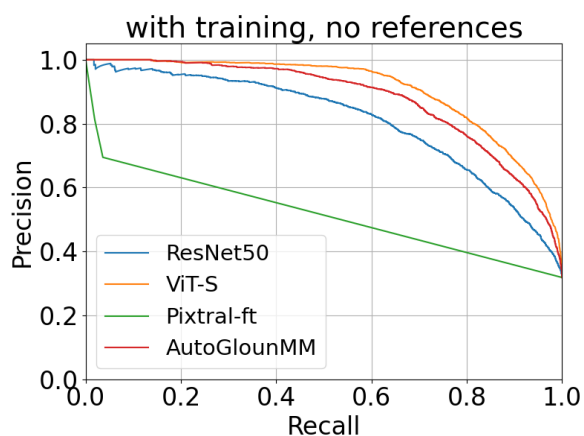
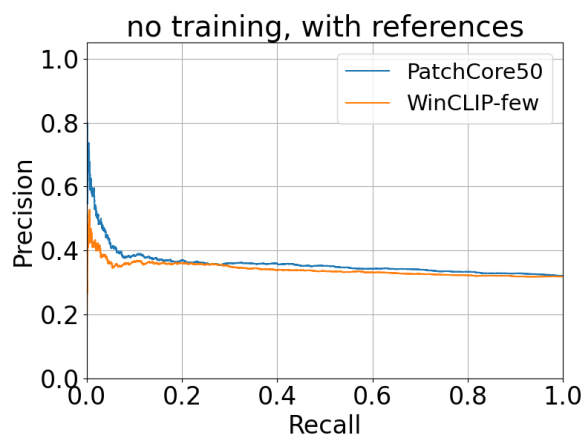
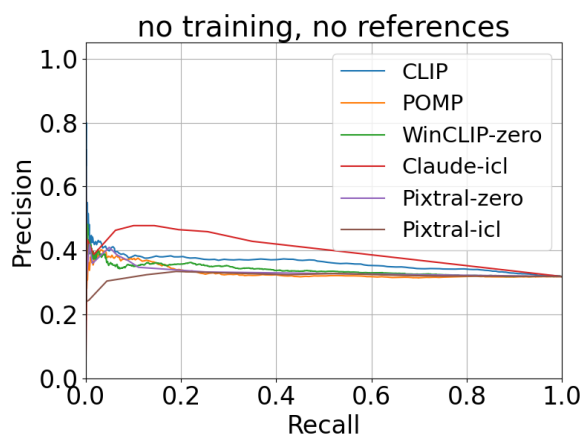


Figure 7. Precision-Recall curves for all four experiments. These curves illustrate the trade-off between precision and recall across different operating thresholds for each experimental configuration.

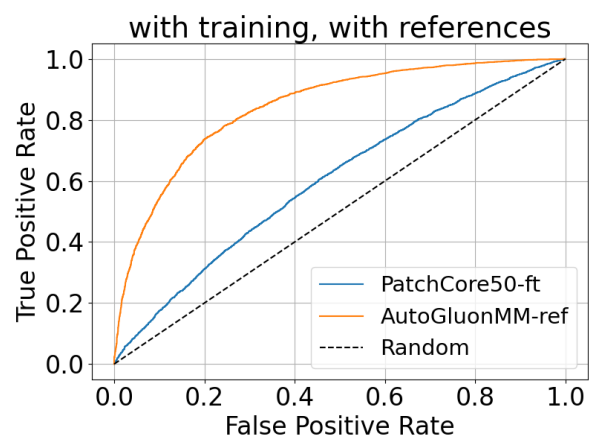
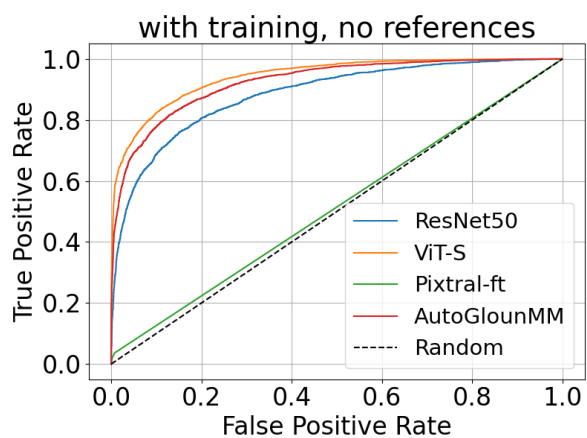
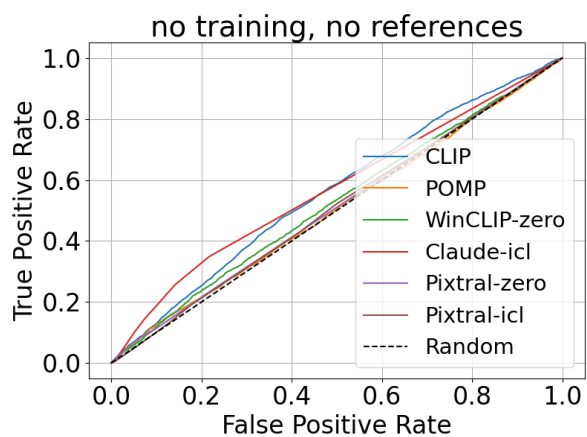


Figure 8. Receiver Operating Characteristic (ROC) curves for all four experiments. These curves show the performance of the classification models by plotting the true positive rate against the false positive rate at various threshold settings.