Supplementary Material

WinCLIP: Zero-/Few-Shot Anomaly Classification and Segmentation

A. Experimental details

Compositional prompt ensemble. Figure 6 provides a detailed list of prompts we adopt to perform compositional prompt ensemble proposed in Section 4.1 of the main text. Recall that we consider two levels of prompts: *i.e.*, (a) state-level, and (b) template level. A complete prompt can be composed by replacing the token `[c]` in a template-level prompt with one of state-level prompt, either from the normal or anomaly states. Each of the state-level prompt takes an object-level label `[o]`. In our experiments, we use the object name words available for both MVTec-AD and VisA per dataset to replace `[o]`.

Data pre-processing. For CLIP-based models, including our proposed WinCLIP and WinCLIP+, we apply the data pre-processing pipeline given in OpenCLIP [4] for both MVTec-AD and VisA datasets to minimize potential train-test discrepancy. Specifically, it performs a channel-wise standardization with the pre-computed mean `[0.48145466, 0.4578275, 0.40821073]` and standard deviation `[0.26862954, 0.26130258, 0.27577711]` after normalizing each RGB image into $[0, 1]$, followed by a bicubic re-sizing based on the `Pillow` implementation. By default, we make the input resolution to be 240 for the shorter edge from the re-sizing, to be compatible with ViT-B/16+ in our experiments. This re-sizing policy also applies to other baseline models for fairer comparisons, although we keep the remaining parts of their original data pre-processing pipelines. In addition, similar policy can also be used in other backbones with input of different resolutions.

Evaluation metrics. Although the AUROC is a good metric for balanced dataset, it provides an inflated view of model performance in imbalanced dataset, especially in anomaly segmentation where the normal pixels dominate anomalies. This is also discussed by Zou et al. [8]. F_1-max is computed from the precision and recall for the anomalous samples at the optimal threshold, which is a more straightforward metric to measure the upper bound of anomaly prediction performance across thresholds. Thus we acknowledge that the low-shot anomaly segmentation is still not solved since our best model only achieves $< 60\%$ F_1-max for both MVTec-AD and VisA, even though WinCLIP+ achieves $> 95\%$ pixel-AUROC. In addition, our WinCLIP and WinCLIP+ outperform all the compared methods in terms of all these metrics on the setups, demonstrating the effectiveness of the proposed methods.

Other implementation details. (*i*) The ViT-B/16+ architecture [4], that we mainly adopt in our experiments, is a modification of ViT-B/16 [6] with (a) an increased dimension in both image ($768 \rightarrow 896$) and text ($512 \rightarrow 640$) embeddings, as well

(a) *State*-level (normal)

- `c := "[o]"`
- `c := "flawless [o]"`
- `c := "perfect [o]"`
- `c := "unblemished [o]"`
- `c := "[o] without flaw"`
- `c := "[o] without defect"`
- `c := "[o] without damage"`

(b) *State*-level (anomaly)

- `c := "damaged [o]"`
- `c := "[o] with flaw"`
- `c := "[o] with defect"`
- `c := "[o] with damage"`

(c) *Template*-level

- `"a cropped photo of the [c]."`
- `"a cropped photo of a [c]."`
- `"a close-up photo of a [c]."`
- `"a close-up photo of the [c]."`
- `"a bright photo of a [c]."`
- `"a bright photo of the [c]."`
- `"a dark photo of the [c]."`
- `"a dark photo of a [c]."`
- `"a jpeg corrupted photo of a [c]."`
- `"a jpeg corrupted photo of the [c]."`
- (cont'd) `"a blurry photo of the [c]."`
- `"a blurry photo of a [c]."`
- `"a photo of a [c]."`
- `"a photo of the [c]."`
- `"a photo of a small [c]."`
- `"a photo of the small [c]."`
- `"a photo of a large [c]."`
- `"a photo of the large [c]."`
- `"a photo of the [c] for visual inspection."`
- `"a photo of a [c] for visual inspection."`
- `"a photo of the [c] for anomaly detection."`
- `"a photo of a [c] for anomaly detection."`

Figure 6. Lists of multi-level prompts considered in this paper to construct compositional prompt ensemble.

MVTec-AD (few-shot)			Anomaly classification			Anomaly segmentation		
Setup	Method	Backbone	pAUROC	PRO	F_1-max	AUROC	AUPR	F_1-max
1-shot	PatchCore [5]	WRN-50-2	83.4±3.0	92.2±1.5	90.5±1.5	92.0±1.0	79.7±2.0	50.4±2.1
	PatchCore (hidden)	ViT-B/16+	79.9±4.8	88.6±2.7	88.7±1.1	91.8±1.0	74.1±2.3	47.6±2.6
	PatchCore (last)		83.3±3.8	90.7±2.1	89.8±1.4	92.3±0.9	74.5±2.2	47.7±2.9
	WinCLIP+ (ours)		**92.7**±1.9	**96.7**±0.7	**93.5**±1.0	**95.2**±0.5	**87.1**±1.2	**55.9**±2.7
2-shot	PatchCore [5]	WRN-50-2	86.3±3.3	93.8±1.7	92.0±1.5	93.3±0.6	82.3±1.3	53.0±1.7
	PatchCore (hidden)	ViT-B/16+	84.1±2.9	90.7±1.9	90.2±1.2	93.5±0.7	77.9±1.8	51.4±2.1
	PatchCore (last)		86.5±2.5	92.3±1.4	91.1±1.6	93.1±0.9	76.8±2.0	49.8±2.2
	WinCLIP+ (ours)		**94.0**±1.7	**97.0**±0.7	**94.0**±1.0	**96.0**±0.3	**88.4**±0.9	**58.4**±1.7
4-shot	PatchCore [5]	WRN-50-2	88.8±2.6	94.5±1.5	92.6±1.6	94.3±0.5	84.3±1.6	55.0±1.9
	PatchCore (hidden)	ViT-B/16+	87.5±3.1	92.5±1.8	91.7±1.4	94.7±0.6	81.2±1.6	54.4±1.9
	PatchCore (last)		89.9±2.2	94.2±1.4	92.7±1.1	94.0±0.6	78.9±1.6	52.2±1.5
	WinCLIP+ (ours)		**94.8**±1.5	**97.5**±0.7	**94.2**±0.9	**96.2**±0.3	**89.0**±0.8	**59.5**±1.8

Table 8. Comparison of few-shot performances on MVTec-AD. We report the mean and standard deviation over 5 random seeds for each measurement. Bold indicates the best performance.

MVTec-AD (zero-shot)			
Model	Size	AC	AS
RN50	224^2	79.8	65.6
RN101	224^2	79.2	62.4
RN50x4	288^2	81.9	71.3
RN50x16	384^2	82.3	65.3
ViT-B/16	224^2	86.1	71.1
ViT-B/16+	240^2	90.8	85.1
ViT-L/14	224^2	86.1	64.4

Table 9. Comparison of WinCLIP performance in AUROC (for AC) and pAUROC (for AS) on zero-shot MVTec-AD, across different CLIP backbone architectures.

as in (b) the input resolution ($224^2 \rightarrow 240^2$; $196 \rightarrow 225$ tokens); (*ii*) We note that CLIP models require the square-shaped resolution, *e.g.*, 240^2 for ViT-B/16+, to be compatible with the attention layers inside. Although the MVTec-AD benchmark already consists of square images, most of images in the VisA benchmark are non-squared (*e.g.*, 1500×1000) and simply taking a crop can affect the anomaly status of the given images. In this respect, to enable CLIP-based models properly handle non-squared images in our experiments, we perform a simple "image tiling" scheme. Specifically, for such non-squared images, we first extract multiple overlapping (squared) "tiles" of size the shorter edge L_s, by taking a sliding window across the longer edge. Then we average the predictions from the tiles to get the final (either in image- and pixel-level) prediction. The stride for the sliding is set to $0.8 \cdot L_s$ at most, *i.e.*, the tiles have overlaps with its neighbors at least in $0.2 \cdot L_s$; (*iii*) In addition, for the baseline results, we use our re-implementation of SPADE [2] and PaDiM [3], and adopt the official implementation of PatchCore[1] in our experiments.

B. Additional results on ablation study

Comparison with CLIP-based PatchCore: PatchCore [5], a current state-of-the-art considered in our experiments, is originally based on the internal features of convolutional network: *e.g.*, WideResNet-50-2 (WRN-50-2) [7] pre-trained on ImageNet. In Table 8, we test whether PatchCore can further benefit from the CLIP-based backbone that our WinCLIP+ is based on. Specifically, we additionally consider two variants of PatchCore that take the patch-token features of CLIP-based ViT-B/16+ backbone, one from (a) the 6^{th}- and 9^{th}-layer of ViT (which corresponds to `block2` and `block3` in ResNet-like models as considered by [5]; "*hidden*"), and the other one from (b) the last layer of ViT ("*last*"). Overall, we have the following observations. First, in case of the ViT-B/16+ backbone, PatchCore performs better with the last layer, which is in contrast to the cases of convolutional backbones. Second, compared to the original PatchCore, the CLIP-based variants achieve no better performances. Third, WinCLIP+ significantly outperforms "PatchCore (last)" where our WinCLIP+ also utilizes the last patch-token features, namely as referred as $\mathbf{F}^{\text{P}}$ (Section 4.3 of the main text). The results confirm the effectiveness of (a) our simple association-based module over a more sophisticated PatchCore[2] in ViT, and (b) the WinCLIP features $\mathbf{F}^{\text{W}}$.

Effect of different CLIP backbones: Table 9, on the other hand, explores the effect of different CLIP architectures to the WinCLIP zero-shot performance. Specifically, on zero-shot setups, we compare AUROC (and pixel-AUROC) from WinCLIP in AC (and AS) testing over the CLIP pre-trained models available at OpenCLIP,[3] including our default choice of ViT-B/16+. To apply WinCLIP for ResNet-based backbones, we notice that the CLIP implementation of ResNet architectures incorporates an attention layer to perform the feature pooling, namely as *attention pooling*, similar to ViT-based architectures. In this respect, for the CLIP-ResNet models, we apply our window-based inference to perform zero-shot AS from the convolutional feature map before the attention pooling, in the same way of applying WinCLIP for ViTs. Here, we remark that the effective patch size of each pixel on the last feature map (before the pooling) of ResNet-based models is designed to be 32 (the downsampling

[1] https://github.com/amazon-science/patchcore-inspection

[2] Technically, PatchCore incorporates several techniques upon a patch-level memory scheme, *e.g.*, local patch aggregation, clustering and score re-weighting.

[3] https://github.com/mlfoundations/open_clip

rate), which is larger than those of ViTs we test, *e.g.*, of 16. Overall, we observe that ViT-based models generally show better performance compared to ResNets, in both AC and AS. The particular gap in AS is possibly due to the bigger patch sizes in ResNets, which can result in more blurry outputs. Still, we observe the performance benefits from larger models or resolutions in both types of architecture.

C. Additional qualitative results

In Figure 7-10, we provide further qualitative results obtained from our (zero-shot) WinCLIP and (few-shot) WinCLIP+ for anomaly segmentation, both in MVTec-AD and VisA considered in our experiments. Specifically, we report MVTec-AD results in Figure 7 and 8, and VisA results in Figure 9 and 10.

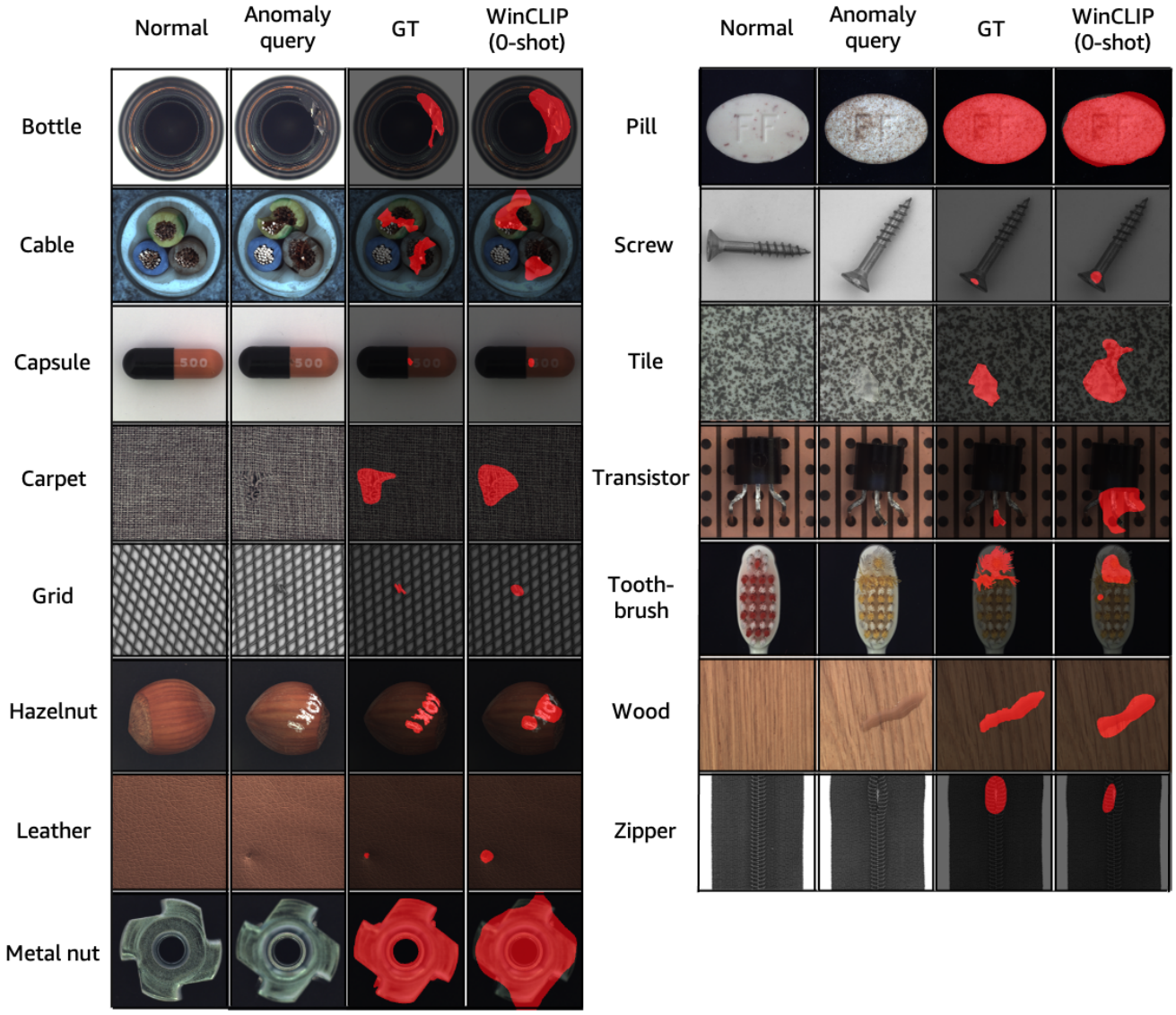

Figure 7. Additional qualitative results from WinCLIP (0-shot), tested on MVTec-AD.

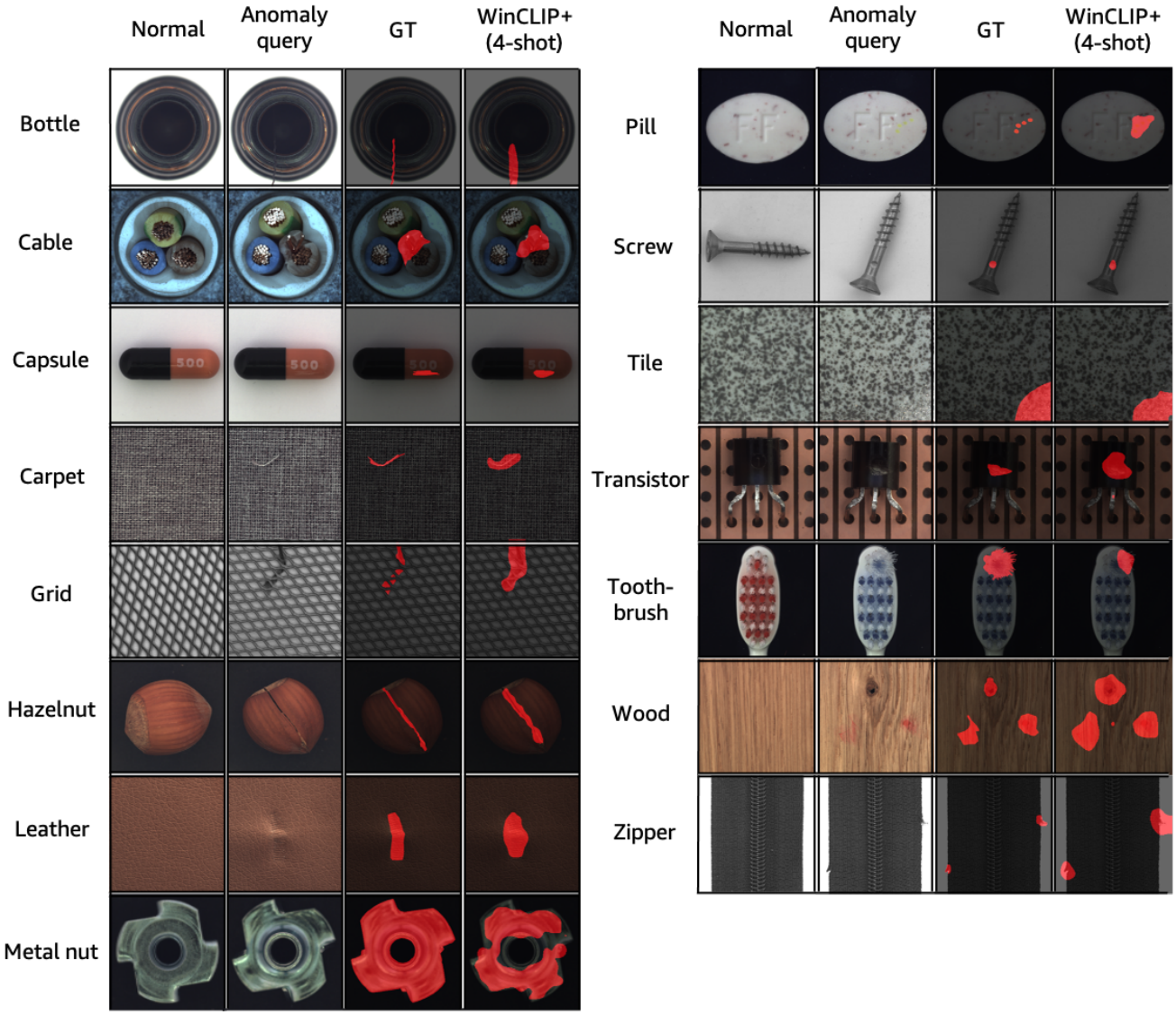

Figure 8. Additional qualitative results from few-shot WinCLIP+ (4-shot), tested on MVTec-AD.

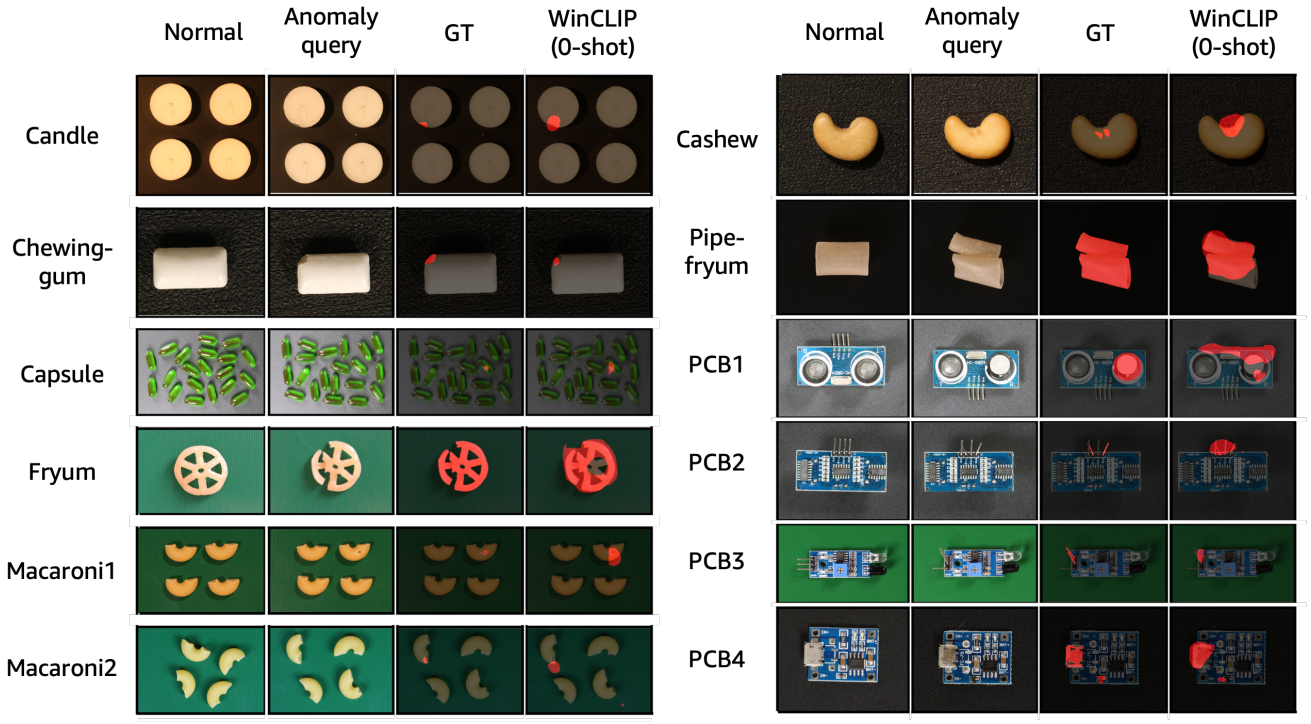

Figure 9. Additional qualitative results from WinCLIP (0-shot), tested on VisA.

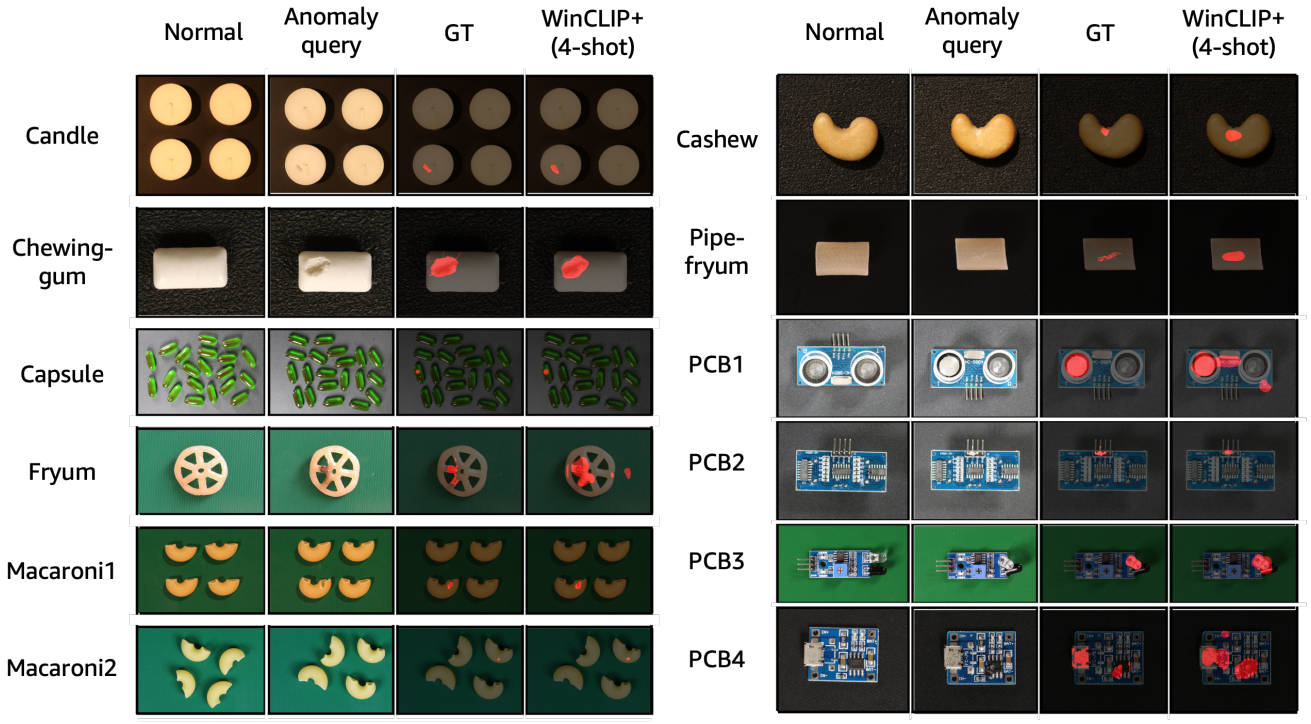

Figure 10. Additional qualitative results from few-shot WinCLIP+ (4-shot), tested on VisA.

Failure cases. We present some failure examples from both MVTec-AD and VisA for language driven zero-shot WinCLIP in Figure 11. Note that the normal images are shown just for better illustration and are not used in model prediction. The first major factor causing the failure is the logical anomaly [1] illustrated in Figure 11(a), *e.g.*, misplaced axis in cable, missing text on capsule, missing capacitor in PCB1 and bent component in PCB3. Such type of anomalies need to be clarified by normal reference images while language might be not sufficient. The issues are alleviated by our few-normal-shot WinCLIP+. The second major factor refers to tiny defect illustrated in Figure 11(b), such as the ones in carpet, wood, capsule, macaroni1. We conjecture that spatial features with more local details might improve these cases, which is left for future exploration. The third major factor is the irrelevant deviation from normality that are not defects of interests illustrated in Figure 11(c), *e.g.*, the tiny red/white dots in pill/hazelnut, extra ingredient on cashew, designed holes and acceptable scratches in PCB2. We hypothesize that more clarification on these deviation and a pre-trained model with better understanding on these states might alleviate the problem. Lastly, although WinCLIP can roughly localize anomalies such as the cases in bottle, tile, PCB4 and fryum, it makes some errors around the true positives, illustrated in Figure 11(d). However, we argue this is minor as the rough anomaly localization is sufficient to explain where the defects are for visual inspection.

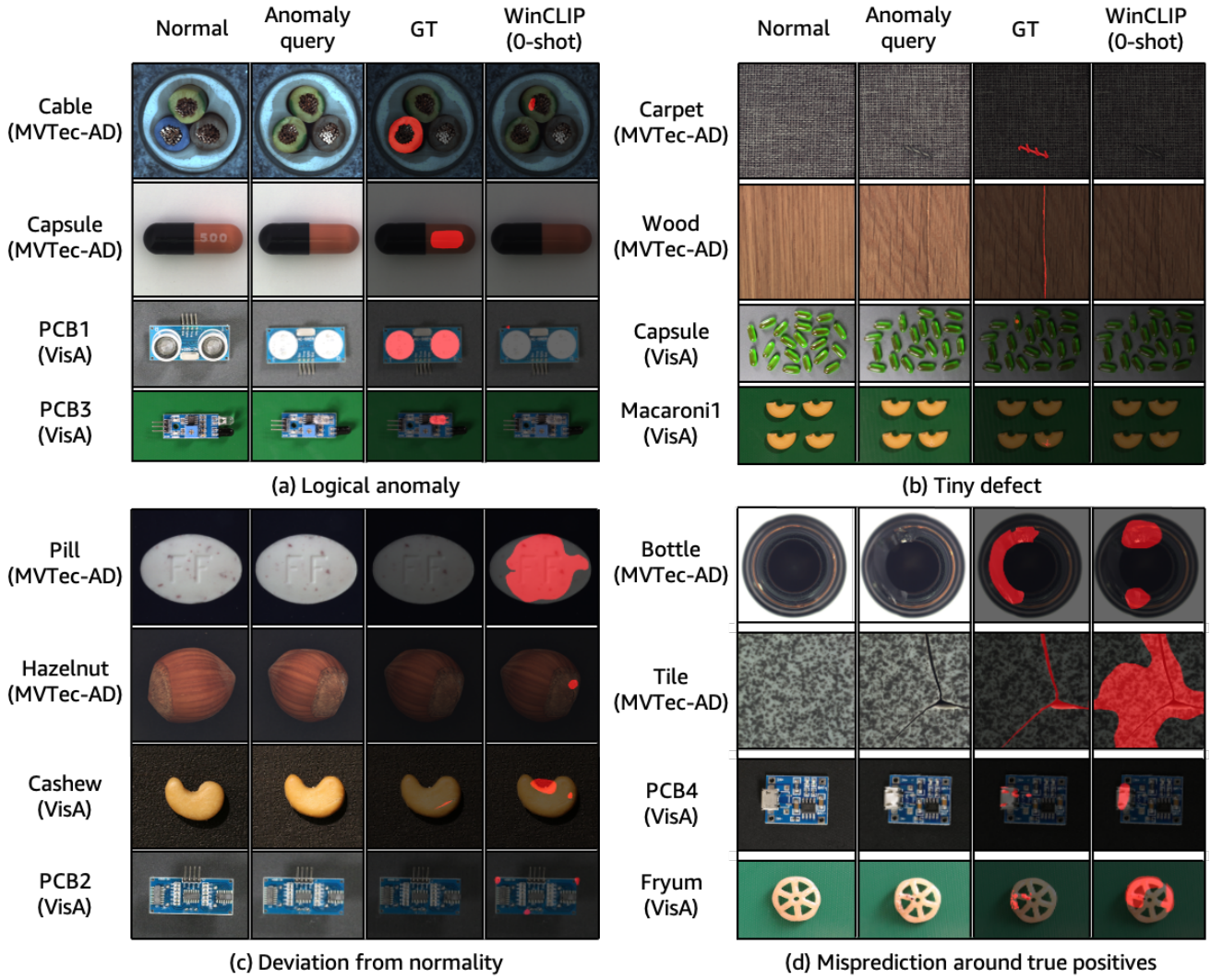

Figure 11. Curated illustrations of failure cases from zero-shot WinCLIP.

D. Detailed quantitative results

In this section, we report the detailed, subset-level performance values for the evaluation metrics provided in Table 1 and 4 of the main text. Specifically, we report MVTec-AD results in Table 10-15 and VisA results in Table 16-21.

MVTec-AD (AC)	$K=0$	$K=1$				$K=2$				$K=4$			
AUROC	**WinCLIP**	SPADE	PaDiM	PatchCore	**WinCLIP+**	SPADE	PaDiM	PatchCore	**WinCLIP+**	SPADE	PaDiM	PatchCore	**WinCLIP+**
Bottle	99.2±0.0	98.7±0.6	97.4±0.7	99.4±0.4	98.2±0.9	99.5±0.1	98.5±1.0	99.2±0.3	99.3±0.3	99.5±0.2	98.8±0.2	99.2±0.3	99.3±0.4
Cable	86.5±0.0	71.2±3.3	57.7±4.6	88.8±4.2	88.9±1.9	76.2±5.2	62.3±5.9	91.0±2.7	88.4±0.7	83.4±3.1	70.0±6.1	91.0±2.7	90.9±0.9
Capsule	72.9±0.0	70.2±3.0	57.7±7.3	67.8±2.9	72.3±6.8	70.9±6.1	64.3±3.0	72.8±7.0	77.3±8.8	78.9±5.5	65.2±2.5	72.8±7.0	82.3±8.9
Carpet	100.0±0.0	98.1±0.2	96.6±1.0	95.3±0.8	99.8±0.3	98.3±0.4	97.8±0.5	96.6±0.5	99.8±0.3	98.6±0.2	97.9±0.4	96.6±0.5	100.0±0.0
Grid	98.8±0.0	40.0±6.8	54.2±6.7	63.6±10.3	99.5±0.3	41.3±3.6	67.2±4.2	67.7±8.3	99.4±0.2	44.6±6.6	68.1±3.8	67.7±8.3	99.6±0.1
Hazelnut	93.9±0.0	95.8±1.3	88.3±2.6	88.3±2.7	97.5±1.4	96.2±2.1	90.8±0.8	93.2±3.8	98.3±0.7	98.4±1.3	91.9±1.2	93.2±3.8	98.4±0.4
Leather	100.0±0.0	100.0±0.0	97.5±0.7	97.3±0.7	99.9±0.0	100.0±0.0	97.5±0.9	97.9±0.7	99.9±0.0	100.0±0.0	98.5±0.2	97.9±0.7	100.0±0.0
Metal nut	97.1±0.0	71.0±2.2	53.0±3.8	73.4±2.9	98.7±0.8	77.0±7.9	54.8±3.8	77.7±8.5	99.4±0.2	77.8±5.7	60.7±5.2	77.7±8.5	99.5±0.2
Pill	79.1±0.0	86.5±3.1	61.3±3.8	81.9±2.8	91.2±2.1	84.8±0.9	59.1±6.4	82.9±2.9	92.3±0.7	86.7±0.3	54.9±2.7	82.9±2.9	92.8±1.0
Screw	83.3±0.0	46.7±2.5	55.0±2.5	44.4±4.6	86.4±0.9	46.6±2.2	54.0±4.4	49.0±3.8	86.0±2.1	50.5±5.4	50.0±4.1	49.0±3.8	87.9±1.2
Tile	100.0±0.0	99.9±0.1	92.2±2.2	99.0±0.9	99.9±0.0	99.9±0.1	93.3±1.1	98.5±1.0	99.9±0.2	100.0±0.0	93.1±0.6	98.5±1.0	99.9±0.1
Toothbrush	87.5±0.0	71.7±2.6	82.5±1.2	83.3±3.8	92.2±4.9	78.6±3.2	87.6±4.2	85.9±3.5	97.5±1.6	78.8±5.2	89.2±2.5	85.9±3.5	96.7±2.6
Transistor	88.0±0.0	77.2±2.0	73.3±6.0	78.1±6.9	83.4±3.8	81.3±3.7	72.8±6.3	90.0±4.3	85.3±1.7	81.4±2.1	82.4±6.5	90.0±4.3	85.7±2.5
Wood	99.4±0.0	98.8±0.3	96.1±1.2	97.8±0.3	99.9±0.1	99.2±0.4	96.9±0.5	98.3±0.6	99.9±0.1	98.9±0.6	97.0±0.2	98.3±0.6	99.8±0.3
Zipper	91.5±0.0	89.3±1.9	85.8±2.7	92.3±0.5	88.8±5.9	93.3±2.9	86.3±2.6	94.0±2.1	94.0±1.4	95.1±1.3	88.3±2.0	94.0±2.1	94.5±0.5
Mean	**91.8±0.0**	81.0±2.0	76.6±3.1	83.4±3.0	**93.1±2.0**	82.9±2.6	78.9±3.1	86.3±3.3	**94.4±1.3**	84.8±2.5	80.4±2.5	88.8±2.6	**95.2±1.3**

Table 10. Comparison of anomaly classification (AC) performance in terms of class-wise AUROC on MVTec-AD. We report the mean and standard deviation over 5 random seeds for each measurement.

MVTec-AD (AC)	$K=0$	$K=1$				$K=2$				$K=4$			
AUPR	**WinCLIP**	SPADE	PaDiM	PatchCore	**WinCLIP+**	SPADE	PaDiM	PatchCore	**WinCLIP+**	SPADE	PaDiM	PatchCore	**WinCLIP+**
Bottle	99.8±0.0	99.6±0.1	99.2±0.2	99.8±0.1	99.4±0.3	99.8±0.0	99.6±0.3	99.8±0.1	99.8±0.1	99.9±0.0	99.7±0.0	99.8±0.1	99.8±0.1
Cable	91.2±0.0	79.6±2.3	64.9±3.8	93.8±2.2	93.2±1.1	84.5±3.1	69.6±6.6	95.1±1.3	92.9±0.6	88.8±1.9	76.1±5.6	97.1±0.7	94.4±0.3
Capsule	91.5±0.0	91.2±0.9	86.9±2.2	89.4±2.0	91.6±2.7	91.6±2.1	88.4±0.8	91.0±2.9	93.3±3.6	94.4±1.9	87.8±0.8	94.9±1.1	95.1±3.3
Carpet	100.0±0.0	99.4±0.0	99.0±0.2	98.7±0.2	99.9±0.1	99.5±0.1	99.4±0.1	99.0±0.1	99.9±0.1	99.6±0.1	99.4±0.1	98.8±0.2	100.0±0.0
Grid	99.6±0.0	66.9±2.1	75.0±3.3	81.1±4.9	99.9±0.1	68.3±2.1	82.5±2.3	84.1±4.0	99.8±0.1	68.8±4.2	83.0±1.8	86.4±4.0	99.9±0.0
Hazelnut	96.9±0.0	97.9±0.6	93.3±1.7	92.9±2.2	98.6±0.7	98.0±1.1	94.1±0.5	96.0±2.0	99.1±0.4	99.1±0.7	94.8±0.6	97.0±1.2	99.1±0.2
Leather	100.0±0.0	100.0±0.0	99.2±0.2	99.1±0.2	100.0±0.0	100.0±0.0	99.2±0.3	99.3±0.2	100.0±0.0	100.0±0.0	99.6±0.1	99.6±0.1	100.0±0.0
Metal nut	99.3±0.0	91.7±0.8	82.0±2.7	91.0±1.1	99.7±0.2	93.7±2.4	82.2±1.4	92.3±4.0	99.9±0.0	94.1±1.8	85.5±1.7	97.0±2.6	99.9±0.1
Pill	95.7±0.0	97.0±0.8	88.3±1.3	96.5±0.6	98.3±0.5	96.5±0.4	87.9±2.6	96.6±0.7	98.6±0.1	97.0±0.2	87.0±1.2	96.9±0.4	98.6±0.2
Screw	93.1±0.0	71.3±1.8	78.1±1.0	71.4±2.3	94.2±0.6	71.0±1.4	77.3±1.3	72.9±3.4	94.1±1.5	73.7±2.4	75.7±2.8	71.8±1.9	94.9±0.8
Tile	100.0±0.0	100.0±0.0	97.2±0.7	99.6±0.3	100.0±0.0	100.0±0.0	97.6±0.4	99.4±0.4	100.0±0.1	100.0±0.0	97.6±0.2	99.6±0.1	100.0±0.0
Toothbrush	95.6±0.0	88.3±0.6	93.7±0.5	93.5±1.4	96.7±2.0	90.8±1.3	95.2±1.6	94.1±1.4	99.0±0.6	91.3±2.6	95.8±0.7	94.8±0.7	98.7±1.1
Transistor	87.1±0.0	76.2±1.7	66.2±7.5	77.7±5.5	79.0±4.0	81.6±3.4	69.0±6.5	89.3±3.9	80.7±2.3	80.3±2.6	77.6±8.4	84.5±9.0	80.7±3.2
Wood	99.8±0.0	99.6±0.1	98.8±0.3	99.3±0.1	100.0±0.0	99.7±0.1	99.0±0.1	99.5±0.2	100.0±0.0	99.7±0.2	99.1±0.0	99.5±0.2	99.9±0.1
Zipper	97.5±0.0	96.9±0.5	95.5±0.9	97.2±0.3	96.8±1.8	98.2±0.8	95.4±1.0	97.8±1.0	98.3±0.4	98.6±0.4	96.2±0.8	99.1±0.7	98.5±0.2
Mean	**96.5±0.0**	90.6±0.8	88.1±1.7	92.2±1.5	**96.5±0.9**	91.7±1.2	89.3±1.7	93.8±1.7	**97.0±0.7**	92.5±1.2	90.5±1.6	94.5±1.5	**97.3±0.6**

Table 11. Comparison of anomaly classification (AC) performance in terms of class-wise AUPR on MVTec-AD. We report the mean and standard deviation over 5 random seeds for each measurement.

MVTec-AD (AC)	$K=0$	$K=1$				$K=2$				$K=4$			
F_1-max	**WinCLIP**	SPADE	PaDiM	PatchCore	**WinCLIP+**	SPADE	PaDiM	PatchCore	**WinCLIP+**	SPADE	PaDiM	PatchCore	**WinCLIP+**
Bottle	97.6±0.0	97.8±0.8	96.3±1.2	98.3±0.6	96.5±1.3	98.7±0.4	97.1±1.1	97.5±0.6	97.7±0.7	98.6±0.3	97.9±0.4	97.9±0.8	97.8±0.6
Cable	84.5±0.0	79.6±2.3	77.2±1.1	85.2±3.6	86.1±1.3	80.4±1.7	78.7±1.2	86.1±2.4	85.2±0.7	83.8±2.5	81.1±1.1	91.3±1.0	87.2±0.6
Capsule	91.4±0.0	92.0±0.6	91.0±0.2	92.0±1.0	91.6±0.7	92.1±0.4	92.1±0.9	93.6±0.6	92.1±0.7	92.7±0.3	92.8±0.9	94.3±0.3	92.5±0.5
Carpet	99.4±0.0	96.5±0.2	95.1±0.5	94.9±0.5	99.2±0.8	96.6±0.3	96.5±0.4	95.3±0.5	99.3±0.7	96.9±0.3	96.6±0.3	94.3±0.8	99.9±0.2
Grid	98.2±0.0	84.5±0.3	84.5±0.3	86.2±1.1	98.9±0.4	84.8±0.3	85.3±0.9	86.9±2.3	99.1±0.0	84.8±0.5	85.0±0.5	87.5±2.0	99.1±0.0
Hazelnut	89.7±0.0	92.4±1.3	87.4±1.6	87.0±1.4	94.7±2.3	93.2±2.8	89.3±1.1	91.0±3.7	95.6±1.6	95.9±2.0	90.0±1.7	92.8±1.2	96.2±1.0
Leather	100.0±0.0	99.9±0.2	96.2±0.9	95.9±0.7	99.5±0.0	100.0±0.0	96.6±1.3	95.7±0.9	99.7±0.2	100.0±0.0	97.9±0.2	97.5±0.7	99.8±0.2
Metal nut	96.3±0.0	90.1±0.6	90.1±0.3	91.4±1.2	97.7±1.0	90.5±1.1	90.0±0.3	91.9±0.9	98.4±0.5	90.6±0.9	90.3±0.4	93.6±1.4	98.5±0.6
Pill	91.6±0.0	93.5±0.4	91.6±0.0	91.9±0.3	93.8±0.7	93.4±0.3	91.7±0.1	92.0±0.3	94.3±0.4	93.6±0.5	91.7±0.1	92.1±0.3	94.1±0.4
Screw	87.4±0.0	85.3±0.0	85.6±0.3	85.8±0.2	88.5±0.3	85.6±0.2	85.5±0.1	85.7±0.2	89.0±0.6	85.8±0.7	85.7±0.3	86.8±0.6	89.6±0.7
Tile	99.4±0.0	99.2±0.5	90.7±2.0	97.5±1.2	98.9±0.2	99.2±0.3	91.5±1.3	96.9±1.5	99.2±0.3	99.4±0.0	91.0±0.9	97.5±0.4	99.2±0.3
Toothbrush	87.9±0.0	85.6±1.3	85.4±1.3	88.9±2.2	94.1±1.9	87.1±1.4	90.1±3.1	90.8±1.6	96.7±1.8	86.5±1.8	90.6±2.1	92.6±2.2	96.8±2.3
Transistor	79.5±0.0	70.3±1.7	66.4±4.7	70.6±7.2	75.1±3.1	73.4±3.4	65.9±3.7	85.3±6.1	75.9±2.4	72.3±2.7	74.8±7.7	78.3±11.5	76.6±2.8
Wood	98.3±0.0	97.0±0.8	94.3±1.1	96.1±0.3	99.4±0.3	97.7±1.0	94.9±0.3	96.6±0.9	99.5±0.4	97.6±0.7	94.8±0.5	96.5±0.9	99.2±0.9
Zipper	92.9±0.0	91.0±0.9	91.2±0.8	95.3±0.5	92.1±2.5	93.5±1.8	92.2±0.6	95.4±0.5	94.4±0.3	94.7±0.8	92.5±0.8	96.5±0.3	94.7±0.4
Mean	**92.9±0.0**	90.3±0.8	88.2±1.1	90.5±1.5	**93.7±1.1**	91.1±1.0	89.2±1.1	92.0±1.5	**94.4±0.8**	91.5±0.9	90.2±1.2	92.6±1.6	**94.7±0.8**

Table 12. Comparison of anomaly classification (AC) performance in terms of class-wise F_1-max on MVTec-AD. We report the mean and standard deviation over 5 random seeds for each measurement.

MVTec-AD (AS)	$K=0$	$K=1$				$K=2$				$K=4$			
pAUROC	**WinCLIP**	SPADE	PaDiM	PatchCore	**WinCLIP+**	SPADE	PaDiM	PatchCore	**WinCLIP+**	SPADE	PaDiM	PatchCore	**WinCLIP+**
Bottle	89.5±0.0	95.3±0.2	96.1±0.5	97.9±0.1	97.5±0.2	95.7±0.2	96.9±0.1	98.1±0.0	97.7±0.1	96.1±0.0	97.1±0.1	98.2±0.0	97.8±0.0
Cable	77.0±0.0	86.4±0.2	88.4±1.2	95.5±0.8	93.8±0.6	87.4±0.3	90.0±0.8	96.4±0.3	94.3±0.4	88.2±0.2	92.1±0.4	97.5±0.3	94.9±0.1
Capsule	86.9±0.0	96.3±0.2	94.5±0.6	95.6±0.4	94.6±0.8	96.7±0.1	95.2±0.5	96.5±0.4	96.4±0.3	97.0±0.2	96.2±0.4	96.8±0.6	96.2±0.5
Carpet	95.4±0.0	98.2±0.0	97.8±0.2	98.4±0.1	99.4±0.0	98.3±0.0	98.2±0.0	98.5±0.1	99.3±0.0	98.4±0.0	98.4±0.0	98.6±0.1	99.3±0.0
Grid	82.2±0.0	80.7±1.3	70.2±2.8	58.8±4.9	96.8±1.0	83.5±1.0	70.8±2.0	62.6±3.2	97.7±0.8	87.2±1.1	77.0±1.8	69.4±1.3	98.0±0.2
Hazelnut	94.3±0.0	97.2±0.1	95.4±0.7	95.8±0.6	98.5±0.2	97.6±0.1	96.8±0.3	96.3±0.6	98.7±0.1	97.7±0.1	97.2±0.2	97.6±0.1	98.8±0.0
Leather	96.7±0.0	99.1±0.0	98.5±0.1	98.8±0.2	99.3±0.0	99.1±0.0	98.7±0.1	99.0±0.1	99.3±0.0	99.1±0.0	98.8±0.0	99.1±0.0	99.3±0.0
Metal nut	61.0±0.0	83.8±0.7	74.6±1.1	89.3±1.4	90.0±0.6	85.8±1.1	80.3±2.1	94.6±1.4	91.4±0.4	87.1±0.7	82.7±3.9	95.9±1.8	92.9±0.4
Pill	80.0±0.0	89.4±0.4	84.8±1.0	93.1±1.1	96.4±0.3	89.9±0.2	87.3±0.7	94.2±0.3	97.0±0.2	90.7±0.2	88.9±0.5	94.8±0.4	97.1±0.0
Screw	89.6±0.0	94.8±0.2	83.3±0.7	89.6±0.5	94.5±0.4	95.6±0.4	89.8±0.8	90.0±0.7	95.2±0.3	96.4±0.4	90.8±0.2	91.3±1.0	96.0±0.5
Tile	77.6±0.0	91.7±0.3	84.1±1.1	94.1±0.5	96.3±0.2	92.0±0.1	87.7±0.2	94.4±0.2	96.5±0.1	92.2±0.1	88.9±0.3	94.6±0.1	96.6±0.1
Toothbrush	86.9±0.0	94.6±0.6	97.3±0.3	97.3±0.4	97.8±0.1	96.2±0.3	97.7±0.3	97.5±0.2	98.1±0.1	97.0±0.6	98.4±0.2	98.4±0.4	98.4±0.5
Transistor	74.7±0.0	71.4±1.3	90.2±2.8	84.9±2.7	85.0±1.8	72.8±0.9	92.3±2.1	89.6±0.9	88.3±1.0	73.4±0.7	94.0±2.7	90.7±1.4	88.5±1.2
Wood	93.4±0.0	93.4±0.1	90.7±0.4	92.7±0.9	94.6±1.0	93.8±0.1	91.9±0.1	93.2±0.7	95.3±0.4	93.9±0.1	92.2±0.1	93.5±0.3	95.4±0.2
Zipper	91.6±0.0	94.9±0.3	93.9±0.8	97.4±0.4	93.9±0.8	95.8±0.2	95.4±0.3	98.0±0.1	94.1±0.7	96.2±0.1	96.1±0.2	98.1±0.1	94.2±0.4
Mean	**85.1**±**0.0**	91.2±0.4	89.3±0.9	92.0±1.0	**95.2**±**0.5**	92.0±0.3	91.3±0.7	93.3±0.6	**96.0**±**0.3**	92.7±0.3	92.6±0.7	94.3±0.5	**96.2**±**0.3**

Table 13. Comparison of anomaly segmentation (AS) performance in terms of class-wise pixel-AUROC on MVTec-AD. We report the mean and standard deviation over 5 random seeds for each measurement.

MVTec-AD (AS)	$K=0$	$K=1$				$K=2$				$K=4$			
PRO	**WinCLIP**	SPADE	PaDiM	PatchCore	**WinCLIP+**	SPADE	PaDiM	PatchCore	**WinCLIP+**	SPADE	PaDiM	PatchCore	**WinCLIP+**
Bottle	76.4±0.0	91.1±0.4	89.8±0.8	93.5±0.3	91.2±0.4	91.8±0.5	91.7±0.2	93.9±0.3	91.8±0.3	92.5±0.1	92.2±0.2	94.0±0.2	91.6±0.2
Cable	42.9±0.0	63.5±0.7	59.1±3.2	84.7±1.0	72.5±2.3	66.7±0.9	66.5±2.8	88.5±0.9	74.7±2.3	69.5±0.4	74.2±1.8	91.7±0.6	77.0±1.1
Capsule	62.1±0.0	92.7±0.4	80.0±2.0	83.9±0.9	85.6±2.7	93.4±0.3	82.3±2.1	86.6±1.0	90.6±0.6	94.1±0.6	85.7±1.3	87.8±1.9	90.1±1.5
Carpet	84.1±0.0	96.1±0.0	92.9±0.3	93.3±0.3	97.4±0.4	96.2±0.0	93.9±0.2	93.7±0.4	97.3±0.3	96.3±0.0	94.4±0.2	93.9±0.4	97.0±0.2
Grid	57.0±0.0	67.7±1.9	41.2±4.6	21.7±9.5	90.5±2.7	72.1±1.5	45.1±3.6	23.7±3.8	92.8±2.5	78.0±1.5	55.5±3.4	30.4±4.6	93.6±0.6
Hazelnut	81.6±0.0	94.9±0.3	85.7±1.9	88.3±1.3	93.7±0.9	95.6±0.2	89.4±0.9	89.8±1.3	94.2±0.3	95.6±0.1	90.4±0.7	92.0±0.3	94.2±0.3
Leather	91.1±0.0	98.7±0.0	95.6±0.2	95.2±1.0	98.6±0.0	98.8±0.0	96.2±0.2	95.9±0.3	98.3±0.4	98.8±0.0	96.3±0.1	96.4±0.1	98.0±0.4
Metal nut	31.8±0.0	73.4±1.1	38.1±1.6	66.7±2.9	84.7±1.1	78.1±1.8	48.2±5.0	79.6±4.2	86.7±0.8	81.2±1.4	54.0±8.8	83.8±5.5	89.4±0.1
Pill	65.0±0.0	92.8±0.3	78.9±0.6	89.5±1.6	93.5±0.2	93.3±0.2	84.3±0.4	91.6±0.5	94.5±0.2	93.9±0.2	86.6±0.4	92.5±0.4	94.6±0.3
Screw	68.5±0.0	85.0±0.8	51.6±1.7	68.1±1.3	82.3±1.1	87.2±1.2	69.5±2.1	69.0±2.1	84.1±0.5	89.5±1.3	72.3±0.8	72.4±3.1	86.3±1.8
Tile	51.2±0.0	84.2±0.4	66.7±1.5	82.5±1.1	89.4±0.4	84.6±0.2	71.9±0.5	82.5±0.5	89.6±0.4	84.9±0.1	73.6±0.9	83.0±0.1	89.9±0.3
Toothbrush	67.7±0.0	83.5±1.3	82.1±1.5	79.0±2.4	85.3±1.0	87.4±1.1	83.3±2.6	81.0±0.7	84.7±1.4	89.0±1.1	87.1±1.7	85.5±3.0	86.0±3.3
Transistor	43.4±0.0	55.3±2.0	70.3±7.0	70.9±4.6	65.0±1.8	57.6±1.4	76.5±5.5	78.8±1.5	68.6±1.1	58.5±0.7	82.2±7.4	79.5±2.8	69.0±1.1
Wood	74.1±0.0	92.9±0.1	86.5±0.6	87.1±1.0	91.0±0.6	93.1±0.1	88.0±0.2	86.8±1.4	91.8±0.6	93.2±0.1	88.4±0.2	87.7±0.4	91.7±0.3
Zipper	71.7±0.0	86.8±0.6	81.7±2.0	91.2±1.1	86.0±1.7	89.0±0.4	85.6±0.7	92.8±0.4	86.4±1.6	90.1±0.2	87.2±0.8	93.4±0.2	86.9±0.7
Mean	**64.6**±**0.0**	83.9±0.7	73.3±2.0	79.7±2.0	**87.1**±**1.2**	85.7±0.7	78.2±1.8	82.3±1.3	**88.4**±**0.9**	87.0±0.5	81.3±1.9	84.3±1.6	**89.0**±**0.8**

Table 14. Comparison of anomaly segmentation (AS) performance in terms of class-wise PRO on MVTec-AD. We report the mean and standard deviation over 5 random seeds for each measurement.

MVTec-AD (AS)	$K=0$	$K=1$				$K=2$				$K=4$			
F_1-max	**WinCLIP**	SPADE	PaDiM	PatchCore	**WinCLIP+**	SPADE	PaDiM	PatchCore	**WinCLIP+**	SPADE	PaDiM	PatchCore	**WinCLIP+**
Bottle	58.1±0.0	61.5±0.3	68.2±1.9	74.8±0.4	72.8±0.8	62.7±0.4	70.7±0.4	75.1±0.1	73.2±0.9	64.3±0.3	71.4±0.4	75.0±0.2	73.3±0.6
Cable	19.7±0.0	25.9±1.2	27.4±1.8	59.8±1.4	49.4±3.3	28.5±0.8	29.5±1.6	62.2±1.0	51.2±1.3	30.2±0.4	34.5±1.1	65.5±1.1	54.7±1.1
Capsule	21.7±0.0	37.5±3.5	27.1±2.8	32.3±2.1	29.7±7.8	39.6±3.0	33.1±2.6	37.9±4.5	43.5±1.4	40.8±3.4	37.0±2.0	39.0±6.3	40.7±4.9
Carpet	49.7±0.0	67.1±0.2	62.4±0.5	67.3±0.4	73.3±1.5	67.6±0.2	62.6±0.2	67.0±0.7	72.9±1.3	68.1±0.2	62.9±0.2	67.4±0.3	72.0±0.7
Grid	18.6±0.0	17.0±1.0	9.4±2.1	5.5±2.2	50.7±4.5	18.9±1.1	13.1±1.5	5.2±1.2	53.4±3.8	23.1±1.6	18.0±1.9	10.0±5.3	52.7±1.5
Hazelnut	37.6±0.0	62.8±0.8	47.7±3.3	50.1±3.9	68.9±2.6	65.1±0.3	57.1±0.7	53.6±3.7	70.5±1.7	65.2±0.6	58.0±1.3	60.8±1.5	71.0±0.3
Leather	39.7±0.0	55.6±0.1	52.3±0.8	58.6±0.4	58.0±0.7	55.8±0.6	52.8±0.2	58.8±0.3	57.5±0.6	55.5±0.1	52.5±0.2	58.8±0.3	56.3±1.0
Metal nut	32.4±0.0	46.4±1.1	38.2±0.9	55.1±2.6	59.4±1.7	48.7±1.4	44.5±2.0	70.4±4.8	62.7±1.5	50.4±0.9	47.5±4.4	74.8±6.7	67.4±1.6
Pill	17.6±0.0	29.6±0.8	25.3±0.6	54.5±4.0	64.7±1.8	31.2±0.4	28.8±1.2	59.4±1.7	67.8±0.5	33.0±0.5	32.7±1.1	61.7±1.6	67.9±0.4
Screw	13.5±0.0	11.6±1.0	3.5±0.1	6.4±0.4	22.2±2.8	14.7±2.3	5.9±0.3	6.5±0.4	22.4±2.8	20.1±5.2	6.4±0.2	7.4±0.5	30.1±4.3
Tile	32.6±0.0	57.3±0.5	42.2±1.4	64.2±1.4	71.2±0.4	58.0±0.2	47.3±0.4	64.4±0.8	71.9±0.6	58.4±0.2	48.8±0.6	65.0±0.1	72.2±0.6
Toothbrush	17.1±0.0	40.0±2.5	59.5±3.8	63.4±3.2	62.7±3.6	46.9±1.8	62.4±2.7	61.5±2.4	65.8±2.2	51.0±3.7	65.0±1.4	64.9±0.5	69.4±4.6
Transistor	30.5±0.0	21.4±1.9	41.6±8.1	48.2±5.6	39.1±3.5	23.2±1.4	47.5±7.7	54.6±1.5	45.6±2.3	23.8±0.8	54.0±10.5	55.7±2.6	46.6±2.2
Wood	51.5±0.0	56.4±0.3	46.6±0.7	52.9±1.1	65.2±1.4	57.0±0.1	47.2±0.2	52.9±1.8	65.8±0.6	57.1±0.4	47.7±0.3	53.3±0.7	65.1±0.5
Zipper	34.4±0.0	45.1±0.3	51.3±2.5	62.5±2.4	50.6±3.9	48.7±0.5	53.1±1.1	65.3±0.7	50.9±4.5	51.2±0.6	55.2±1.7	65.1±0.5	52.8±2.7
Mean	**31.7**±**0.0**	42.4±1.0	40.2±2.1	50.4±2.1	**55.9**±**2.7**	44.5±1.0	43.7±1.5	53.0±1.7	**58.4**±**1.7**	46.2±1.3	46.1±1.8	55.0±1.9	**59.5**±**1.8**

Table 15. Comparison of anomaly segmentation (AS) performance in terms of class-wise F_1-max on MVTec-AD. We report the mean and standard deviation over 5 random seeds for each measurement.

VisA (AC)	$K=0$	$K=1$				$K=2$				$K=4$			
AUROC	**WinCLIP**	SPADE	PaDiM	PatchCore	**WinCLIP+**	SPADE	PaDiM	PatchCore	**WinCLIP+**	SPADE	PaDiM	PatchCore	**WinCLIP+**
Candle	95.4±0.0	86.1±5.6	70.8±4.1	85.1±1.4	93.4±1.4	91.3±3.3	75.8±2.1	85.3±1.5	94.8±1.0	92.8±2.1	77.5±1.6	87.8±0.8	95.1±0.3
Capsules	85.0±0.0	73.3±7.5	51.0±7.8	60.0±7.6	85.0±3.1	71.7±11.2	51.7±4.6	57.8±5.4	84.9±0.8	73.4±7.1	52.7±3.4	63.4±5.4	86.8±1.7
Cashew	92.1±0.0	95.9±1.1	62.3±9.9	89.5±4.4	94.0±0.4	97.3±1.4	74.6±3.6	93.6±0.6	94.3±0.5	96.4±1.3	77.7±3.2	93.0±1.5	95.2±0.8
Chewinggum	96.5±0.0	92.1±2.0	69.9±4.9	97.3±0.3	97.6±0.8	93.4±1.0	82.7±2.1	97.8±0.6	97.3±0.8	93.5±1.4	83.5±3.7	98.3±0.3	97.7±0.3
Fryum	80.3±0.0	81.1±4.0	58.3±5.9	75.0±4.8	88.5±1.9	90.5±3.9	69.2±9.0	83.4±2.4	90.5±0.4	92.9±1.6	71.2±5.9	88.6±1.3	90.8±0.5
Macaroni1	76.2±0.0	66.0±10.5	62.1±4.6	68.0±3.4	82.9±1.5	69.1±8.2	62.2±5.0	75.6±4.6	83.3±1.9	65.8±1.2	65.9±3.9	82.9±2.7	85.2±0.9
Macaroni2	63.7±0.0	55.8±6.1	47.5±5.9	55.6±4.6	70.2±0.9	58.3±4.4	50.8±2.9	57.3±5.6	71.8±2.0	56.7±3.2	55.0±2.9	61.7±1.8	70.9±2.2
PCB1	73.6±0.0	87.2±2.3	76.2±1.2	78.9±1.1	75.6±23.0	86.7±1.1	62.4±10.8	71.5±20.0	76.7±5.2	83.4±8.5	82.6±1.5	84.7±6.7	88.3±1.7
PCB2	51.2±0.0	73.5±3.7	61.2±2.0	81.5±0.8	62.2±3.9	70.3±8.1	66.8±2.0	84.3±1.7	62.6±3.7	71.7±7.0	73.5±2.4	84.3±1.0	67.5±2.6
PCB3	73.4±0.0	72.2±1.0	51.4±12.2	82.7±2.3	74.1±1.1	75.8±5.7	67.3±3.8	84.8±1.2	78.8±1.9	79.0±4.1	65.9±1.9	87.0±1.1	83.3±1.7
PCB4	79.6±0.0	93.4±1.3	76.1±3.6	93.9±2.8	85.2±8.9	86.1±8.2	69.3±13.7	94.3±3.2	82.3±9.9	95.4±2.3	85.4±2.0	95.6±1.6	87.6±8.0
Pipe fryum	69.7±0.0	77.9±3.2	66.7±2.2	90.7±1.7	97.2±1.1	78.1±3.0	75.3±1.8	93.5±1.3	98.0±0.6	79.3±0.9	82.9±2.2	96.4±0.7	98.5±0.4
Mean	**78.1**±0.0	79.5±4.0	62.8±5.4	79.9±2.9	**83.8**±4.0	80.7±5.0	67.4±5.1	81.6±4.0	**84.6**±2.4	81.7±3.4	72.8±2.9	85.3±2.1	**87.3**±1.8

Table 16. Comparison of anomaly classification (AC) performance in terms of class-wise AUROC on VisA. We report the mean and standard deviation over 5 random seeds for each measurement.

VisA (AC)	$K=0$	$K=1$				$K=2$				$K=4$			
AUPR	**WinCLIP**	SPADE	PaDiM	PatchCore	**WinCLIP+**	SPADE	PaDiM	PatchCore	**WinCLIP+**	SPADE	PaDiM	PatchCore	**WinCLIP+**
Candle	95.8±0.0	86.5±4.3	69.2±3.9	86.6±2.3	93.6±1.5	90.7±3.2	72.8±1.0	86.8±1.7	95.1±1.1	92.6±1.9	72.5±1.1	88.9±1.1	95.3±0.4
Capsules	90.9±0.0	79.4±4.9	63.4±5.7	72.3±5.3	89.9±2.5	79.9±5.8	63.4±2.0	73.6±4.7	88.9±0.7	81.1±4.5	63.0±2.3	78.4±3.1	91.5±1.4
Cashew	96.4±0.0	97.9±0.4	78.2±5.7	94.6±2.0	97.2±0.2	98.6±0.6	86.1±2.2	96.9±0.3	97.3±0.2	98.3±0.6	88.4±2.0	96.5±0.7	97.7±0.4
Chewinggum	98.6±0.0	96.4±0.9	79.8±3.6	98.9±0.1	99.0±0.3	97.1±0.4	89.5±1.9	99.1±0.2	98.9±0.3	97.1±0.6	88.5±3.2	99.3±0.1	99.0±0.1
Fryum	90.1±0.0	89.8±1.8	74.5±2.9	87.6±2.4	94.7±1.0	94.5±2.3	81.0±5.4	92.1±1.3	95.8±0.2	95.8±1.0	81.5±3.0	95.0±0.6	96.0±0.3
Macaroni1	75.8±0.0	61.9±11.2	60.4±2.9	67.8±3.4	84.9±1.2	64.5±9.5	63.1±4.3	74.9±5.2	84.7±1.5	60.2±2.7	64.9±2.1	82.1±3.5	86.5±0.6
Macaroni2	60.3±0.0	52.7±4.2	51.7±5.0	54.9±3.2	68.4±1.8	55.9±3.1	52.7±1.5	57.2±2.6	70.4±1.8	51.9±2.3	54.9±2.5	60.2±3.0	69.6±2.8
PCB1	78.4±0.0	84.9±3.7	68.6±2.4	72.1±2.5	76.5±19.0	83.8±2.1	60.4±7.7	72.6±16.4	78.3±4.3	83.2±7.2	77.4±2.9	81.0±9.2	87.7±1.7
PCB2	49.2±0.0	74.9±2.9	63.3±1.2	84.4±0.4	64.9±3.3	71.7±6.6	68.9±2.6	86.6±1.1	65.8±4.0	74.2±5.0	75.0±1.7	86.2±1.0	71.3±3.4
PCB3	76.5±0.0	75.5±2.1	52.3±10.8	84.6±1.5	73.5±1.6	78.3±5.2	65.2±3.8	86.1±0.5	80.9±1.6	81.0±3.6	64.5±2.4	88.3±1.1	84.8±1.8
PCB4	77.7±0.0	92.9±1.6	74.7±2.6	92.8±3.1	78.5±15.5	81.9±11.2	67.6±11.9	93.2±3.4	72.5±16.2	94.8±2.9	84.0±2.0	94.9±1.2	85.6±8.9
Pipe fryum	82.3±0.0	88.3±2.0	79.2±1.5	95.4±0.6	98.6±0.5	88.1±1.7	84.5±1.7	96.8±0.7	99.0±0.3	88.8±1.0	89.8±1.7	98.3±0.3	99.2±0.2
Mean	**81.2**±0.0	82.0±3.3	68.3±4.0	82.8±2.3	**85.1**±4.0	82.3±4.3	71.6±3.8	84.8±3.2	**85.8**±2.7	83.4±2.7	75.6±2.2	87.5±2.1	**88.8**±1.8

Table 17. Comparison of anomaly classification (AC) performance in terms of class-wise AUPR on VisA. We report the mean and standard deviation over 5 random seeds for each measurement.

VisA (AC)	$K=0$	$K=1$				$K=2$				$K=4$			
F_1-max	**WinCLIP**	SPADE	PaDiM	PatchCore	**WinCLIP+**	SPADE	PaDiM	PatchCore	**WinCLIP+**	SPADE	PaDiM	PatchCore	**WinCLIP+**
Candle	89.4±0.0	80.4±6.1	72.0±2.0	79.5±0.6	87.8±1.2	85.5±2.9	74.8±2.9	78.7±0.8	89.1±1.3	86.9±2.4	76.7±1.3	80.5±0.9	88.9±1.0
Capsules	83.9±0.0	81.0±2.4	77.5±0.8	77.9±1.5	84.9±2.0	80.4±3.9	77.2±0.4	77.2±0.3	85.4±0.6	79.5±1.8	77.2±0.3	77.3±0.6	86.0±0.9
Cashew	88.4±0.0	94.8±1.8	80.8±0.7	89.6±3.6	90.7±0.7	95.5±2.2	82.4±1.2	92.3±0.5	90.9±0.7	95.7±0.9	82.5±1.2	91.1±2.1	91.6±1.3
Chewinggum	94.8±0.0	89.7±2.2	83.8±2.0	95.9±0.8	95.6±0.9	90.5±1.2	86.7±0.8	97.0±0.5	95.4±0.6	91.3±1.5	87.9±0.8	97.4±0.6	95.7±0.5
Fryum	82.7±0.0	85.3±2.1	80.6±0.8	82.9±1.7	87.2±1.4	90.9±1.8	82.7±2.1	84.7±1.4	88.4±0.6	91.9±1.7	82.9±1.8	86.6±0.6	88.9±0.8
Macaroni1	74.2±0.0	71.9±2.0	69.2±2.3	70.4±1.9	76.2±1.4	72.8±3.1	68.8±1.8	74.3±2.1	76.7±2.0	70.8±1.2	70.0±1.6	78.9±1.4	78.2±1.2
Macaroni2	69.8±0.0	68.1±0.8	67.1±0.2	67.6±0.7	72.3±1.1	68.2±1.2	67.1±0.4	67.6±1.2	73.9±0.9	67.9±0.6	68.4±1.0	68.8±0.8	73.1±1.6
PCB1	71.0±0.0	85.5±0.2	80.3±0.8	84.5±0.4	81.3±6.6	85.8±0.2	71.4±5.4	78.3±6.7	73.2±3.7	81.2±6.4	83.1±0.6	85.6±1.8	83.1±2.2
PCB2	67.1±0.0	70.9±1.9	68.6±1.4	75.9±0.8	67.2±0.3	71.5±2.7	69.2±0.6	78.1±2.1	67.3±0.3	71.1±3.2	72.0±2.3	79.2±1.9	67.7±0.6
PCB3	71.0±0.0	70.2±1.4	67.7±0.9	76.7±2.5	73.5±1.5	73.3±3.6	69.9±1.1	78.9±1.0	73.9±1.3	75.5±3.3	69.0±0.7	80.7±0.5	77.0±1.4
PCB4	74.9±0.0	87.5±1.7	74.5±2.2	90.4±2.7	86.1±2.1	83.1±5.9	74.6±3.9	91.3±4.1	86.8±3.8	90.6±2.1	81.0±1.6	92.2±3.4	84.6±7.0
Pipe fryum	80.7±0.0	82.7±0.7	81.2±0.5	89.2±2.4	94.4±0.7	82.7±1.4	83.1±0.6	91.4±1.1	95.4±0.8	82.8±0.6	85.4±0.8	93.9±1.2	95.6±0.7
Mean	**79.0**±0.0	80.7±1.9	75.3±1.2	81.7±1.6	**83.1**±1.7	81.7±2.5	75.7±1.8	82.5±1.8	**83.0**±1.4	82.1±2.1	78.0±1.2	84.3±1.3	**84.2**±1.6

Table 18. Comparison of anomaly classification (AC) performance in terms of class-wise F_1-max on VisA. We report the mean and standard deviation over 5 random seeds for each measurement.

VisA (AS)	$K=0$	$K=1$				$K=2$				$K=4$			
pAUROC	**WinCLIP**	SPADE	PaDiM	PatchCore	**WinCLIP+**	SPADE	PaDiM	PatchCore	**WinCLIP+**	SPADE	PaDiM	PatchCore	**WinCLIP+**
Candle	88.9±0.0	97.9±0.3	91.7±2.2	97.2±0.2	97.4±0.2	98.1±0.2	94.9±0.8	97.7±0.3	97.7±0.1	98.2±0.1	95.4±0.2	97.9±0.1	97.8±0.2
Capsules	81.6±0.0	95.5±0.5	70.9±1.1	93.2±0.9	96.4±0.6	96.5±0.9	75.7±1.7	94.0±0.2	96.8±0.3	97.7±0.1	79.1±0.7	94.8±0.5	97.1±0.2
Cashew	84.7±0.0	95.9±0.5	95.5±0.6	98.1±0.1	98.5±0.2	95.9±0.4	96.4±0.4	98.2±0.2	98.5±0.1	95.9±0.3	97.2±0.3	98.3±0.2	98.7±0.0
Chewinggum	93.3±0.0	96.0±0.4	90.1±0.4	96.9±0.3	98.6±0.1	96.0±0.3	93.1±0.7	96.6±0.1	98.6±0.1	95.7±0.3	94.4±0.5	96.8±0.1	98.5±0.1
Fryum	88.5±0.0	93.5±0.3	93.3±0.6	93.3±0.5	96.4±0.3	93.9±0.2	94.1±0.6	94.0±0.3	97.0±0.2	94.4±0.1	95.0±0.4	94.2±0.2	97.1±0.1
Macaroni1	70.9±0.0	97.9±0.2	89.4±0.9	95.2±0.4	96.4±0.6	98.5±0.2	91.7±0.3	96.0±1.3	96.5±0.7	98.8±0.1	93.5±0.5	97.0±0.3	97.0±0.2
Macaroni2	59.3±0.0	94.1±1.0	86.4±1.1	89.1±1.6	96.8±0.4	95.2±0.4	90.1±0.8	90.2±1.9	96.8±0.6	96.4±0.2	90.2±0.3	93.9±0.3	97.3±0.3
PCB1	61.2±0.0	94.7±0.4	89.9±0.3	96.1±1.5	96.6±0.6	96.5±1.5	90.6±0.6	97.6±0.9	97.0±0.9	96.8±1.5	93.2±1.5	98.1±1.0	98.1±0.9
PCB2	71.6±0.0	95.1±0.2	90.9±1.4	95.4±0.2	93.0±0.4	95.7±0.1	93.9±0.9	96.0±0.3	93.9±0.2	96.3±0.0	93.7±1.0	96.6±0.2	94.6±0.4
PCB3	85.3±0.0	96.0±0.1	93.9±0.3	96.2±0.3	94.3±0.3	96.6±0.1	95.1±0.5	97.1±0.1	95.1±0.2	96.9±0.0	95.7±0.1	97.4±0.2	95.8±0.1
PCB4	94.4±0.0	92.0±0.6	89.6±0.6	95.6±0.6	94.0±0.9	92.8±0.3	90.7±0.9	96.2±0.4	95.6±0.3	94.1±0.2	92.1±0.5	97.0±0.2	96.1±0.3
Pipe fryum	75.4±0.0	98.4±0.2	97.2±0.6	98.8±0.2	98.3±0.2	98.7±0.1	98.1±0.4	99.1±0.1	98.5±0.2	98.8±0.0	98.5±0.1	99.1±0.0	98.7±0.1
Mean	**79.6±0.0**	95.6±0.4	89.9±0.8	95.4±0.6	**96.4±0.4**	96.2±0.4	92.0±0.7	96.1±0.5	**96.8±0.3**	96.6±0.3	93.2±0.5	96.8±0.3	**97.2±0.2**

Table 19. Comparison of anomaly segmentation (AS) performance in terms of class-wise pixel-AUROC on VisA. We report the mean and standard deviation over 5 random seeds for each measurement.

VisA (AS)	$K=0$	$K=1$				$K=2$				$K=4$			
PRO	**WinCLIP**	SPADE	PaDiM	PatchCore	**WinCLIP+**	SPADE	PaDiM	PatchCore	**WinCLIP+**	SPADE	PaDiM	PatchCore	**WinCLIP+**
Candle	83.5±0.0	95.6±0.5	81.5±5.3	92.6±0.4	94.0±0.4	95.6±0.4	87.3±1.2	93.4±0.6	94.2±0.2	95.7±0.1	88.3±0.7	94.1±0.4	94.4±0.2
Capsules	35.3±0.0	83.1±1.1	30.6±1.1	66.6±4.5	73.6±3.5	85.4±3.1	38.4±3.7	67.9±2.3	75.9±1.9	89.0±1.2	43.3±2.0	69.0±3.2	77.0±1.4
Cashew	76.4±0.0	89.8±1.1	73.4±2.1	90.8±0.2	91.1±0.8	90.4±0.5	78.4±2.7	91.4±1.0	90.4±0.6	90.4±0.6	81.2±2.8	92.1±0.3	91.3±0.9
Chewinggum	70.4±0.0	73.9±1.2	58.1±0.6	78.2±1.3	91.0±0.5	73.8±1.1	63.7±2.4	78.0±0.4	90.9±0.7	72.7±0.9	67.2±1.8	79.3±0.8	91.0±0.4
Fryum	77.4±0.0	83.7±1.2	71.1±1.6	78.7±2.3	89.1±1.0	84.5±0.9	71.2±0.8	81.4±2.8	89.3±0.2	86.2±0.9	73.2±1.3	81.0±1.2	89.7±0.5
Macaroni1	34.3±0.0	92.0±0.6	62.2±4.4	83.4±1.3	84.6±2.3	93.9±0.8	71.8±2.4	86.2±4.6	85.2±1.4	95.1±0.4	76.6±2.1	89.6±0.7	86.8±0.8
Macaroni2	21.4±0.0	80.0±3.3	54.9±3.6	66.0±3.0	89.3±2.4	81.7±1.5	65.6±3.4	67.2±6.5	88.6±1.7	86.0±0.8	65.9±1.5	78.3±0.9	90.5±1.3
PCB1	26.3±0.0	81.3±5.7	63.9±1.8	79.0±10.7	82.5±6.0	87.2±2.3	68.4±4.1	86.1±1.7	83.8±5.0	88.0±2.7	70.2±3.3	88.1±2.6	87.9±2.1
PCB2	37.2±0.0	83.7±0.6	64.4±3.8	80.9±0.5	73.6±1.5	85.5±1.0	72.9±3.4	82.9±1.8	76.2±0.9	87.0±0.5	71.9±2.6	83.7±1.0	78.0±1.3
PCB3	56.1±0.0	84.3±1.0	69.0±1.2	78.1±2.0	79.5±2.5	86.1±0.6	74.0±2.3	82.2±1.1	82.3±1.8	87.7±0.6	77.2±0.8	84.4±1.9	84.2±1.0
PCB4	80.4±0.0	66.9±2.0	59.1±1.8	77.9±3.1	76.6±4.1	69.3±1.1	62.6±3.6	79.5±4.8	81.7±1.2	74.7±1.0	67.9±2.6	83.5±2.5	84.2±0.7
Pipe fryum	82.3±0.0	94.3±0.5	83.9±0.8	93.6±0.5	96.1±0.6	95.0±0.2	86.9±0.9	94.5±0.4	96.2±0.6	95.0±0.3	88.7±1.3	95.0±0.5	96.6±0.2
Mean	**56.8±0.0**	84.1±1.6	64.3±2.4	80.5±2.5	**85.1±2.1**	85.7±1.1	70.1±2.6	82.6±2.3	**86.2±1.4**	87.3±0.8	72.6±1.9	84.9±1.4	**87.6±0.9**

Table 20. Comparison of anomaly segmentation (AS) performance in terms of class-wise PRO on VisA. We report the mean and standard deviation over 5 random seeds for each measurement.

VisA (AS)	$K=0$	$K=1$				$K=2$				$K=4$			
F_1-max	**WinCLIP**	SPADE	PaDiM	PatchCore	**WinCLIP+**	SPADE	PaDiM	PatchCore	**WinCLIP+**	SPADE	PaDiM	PatchCore	**WinCLIP+**
Candle	22.5±0.0	37.1±0.8	19.6±2.6	40.9±1.0	42.7±1.7	37.6±0.5	21.6±1.4	40.4±1.0	42.2±0.8	38.3±1.0	21.6±0.7	41.0±1.0	43.0±0.9
Capsules	9.2±0.0	28.5±7.8	2.3±0.2	35.7±6.8	58.2±1.3	37.4±8.4	3.0±0.4	37.8±5.7	57.0±3.7	48.0±2.0	3.9±0.4	47.0±3.0	59.8±1.8
Cashew	13.2±0.0	51.1±1.4	35.1±4.9	60.4±1.1	59.5±2.1	52.8±1.2	40.8±4.0	60.3±0.4	60.5±2.4	54.1±0.5	47.2±2.9	60.7±0.4	62.3±1.1
Chewinggum	41.1±0.0	58.7±1.1	19.7±2.0	64.5±1.0	65.3±0.5	59.9±0.5	29.5±6.4	63.9±0.4	64.8±0.9	59.5±0.6	37.8±4.4	64.4±0.6	65.2±0.2
Fryum	22.1±0.0	34.0±1.5	32.3±1.5	37.2±1.4	50.8±1.8	36.6±2.1	36.5±3.6	41.1±2.9	54.8±1.7	40.3±1.8	44.5±1.3	44.6±2.9	56.5±0.6
Macaroni1	7.0±0.0	28.5±3.4	4.8±0.7	16.5±2.6	34.1±1.7	39.2±3.5	5.5±1.2	20.0±8.2	33.2±1.9	37.6±5.8	7.0±0.7	21.6±1.9	33.8±0.9
Macaroni2	1.0±0.0	6.7±2.8	1.9±0.5	2.7±0.7	34.4±3.0	8.6±1.8	2.4±0.1	5.1±2.3	29.9±3.4	18.3±3.0	2.4±0.3	10.9±1.7	35.1±2.5
PCB1	2.4±0.0	16.6±1.1	8.6±0.7	30.7±3.2	25.9±2.6	34.8±19.9	8.9±0.5	47.7±19.7	34.6±16.2	37.1±20.8	13.9±3.6	55.9±20.3	50.9±20.4
PCB2	4.7±0.0	35.0±1.9	9.8±2.2	33.3±1.0	18.7±1.5	39.2±1.1	16.6±2.5	33.3±0.6	24.0±1.1	42.6±1.0	15.8±3.0	33.6±0.3	27.8±1.9
PCB3	10.3±0.0	43.9±1.6	18.1±0.3	36.6±0.6	31.2±6.7	44.9±0.7	20.6±1.6	37.1±0.2	37.1±2.8	47.5±0.7	22.4±1.3	37.3±0.1	42.5±1.1
PCB4	32.0±0.0	30.7±1.0	13.3±1.5	34.8±1.4	22.8±2.9	35.0±3.4	15.5±1.6	40.9±4.3	30.6±2.4	39.2±5.2	19.8±2.5	44.1±4.4	31.9±3.0
Pipe fryum	12.3±0.0	55.5±1.9	43.3±3.6	62.6±2.6	51.8±2.0	59.3±1.4	52.9±5.5	64.5±0.6	53.6±2.6	61.1±0.7	58.2±0.8	65.0±0.5	55.1±1.1
Mean	**14.8±0.0**	35.5±2.2	17.4±1.7	38.0±1.9	**41.3±2.3**	40.5±3.7	21.1±2.4	41.0±3.9	**43.5±3.3**	43.6±3.6	24.6±1.8	43.9±3.1	**47.0±3.0**

Table 21. Comparison of anomaly segmentation (AS) performance in terms of class-wise F_1-max on VisA. We report the mean and standard deviation over 5 random seeds for each measurement.

References

[1] Paul Bergmann, Kilian Batzner, Michael Fauser, David Sattlegger, and Carsten Steger. Beyond dents and scratches: Logical constraints in unsupervised anomaly detection and localization. *International Journal of Computer Vision*, 130(4):947–969, 2022. 6

[2] Niv Cohen and Yedid Hoshen. Sub-image anomaly detection with deep pyramid correspondences. *arXiv preprint arXiv:2005.02357*, 2020. 2

[3] Thomas Defard, Aleksandr Setkov, Angelique Loesch, and Romaric Audigier. PaDiM: a patch distribution modeling framework for anomaly detection and localization. In *International Conference on Pattern Recognition*, pages 475–489. Springer, 2021. 2

[4] Gabriel Ilharco, Mitchell Wortsman, Ross Wightman, Cade Gordon, Nicholas Carlini, Rohan Taori, Achal Dave, Vaishaal Shankar, Hongseok Namkoong, John Miller, Hannaneh Hajishirzi, Ali Farhadi, and Ludwig Schmidt. *OpenCLIP*. Zenodo, July 2021. 1

[5] Karsten Roth, Latha Pemula, Joaquin Zepeda, Bernhard Schölkopf, Thomas Brox, and Peter Gehler. Towards total recall in industrial anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14318–14328, 2022. 2

[6] Hugo Touvron, Matthieu Cord, Matthijs Douze, Francisco Massa, Alexandre Sablayrolles, and Hervé Jégou. Training data-efficient image transformers & distillation through attention. In *International Conference on Machine Learning*, pages 10347–10357. PMLR, 2021. 1

[7] Sergey Zagoruyko and Nikos Komodakis. Wide residual networks. In Edwin R. Hancock Richard C. Wilson and William A. P. Smith, editors, *Proceedings of the British Machine Vision Conference (BMVC)*, pages 87.1–87.12. BMVA Press, September 2016. 2

[8] Yang Zou, Jongheon Jeong, Latha Pemula, Dongqing Zhang, and Onkar Dabeer. SPot-the-Difference self-supervised pre-training for anomaly detection and segmentation. In *Proceedings of the European Conference on Computer Vision*, 2022. 1