# Unseen Visual Anomaly Generation

## Supplementary Material

## Overview

The supplementary material presents the following sections to strengthen the main manuscript:

- **Sec. A** shows more implementation details.
- **Sec. B** presents a user study on anomaly generation quality.
- **Sec. C** presents more anomaly generation results.
- **Sec. D** presents results on mask controllability.
- **Sec. E** shows ablations on hyperparameters.
- **Sec. F** shows more ablations on attention-guided and prompt-guided optimization.
- **Sec. G** shows more anomaly detection results.

## A. More Implementation Details

**Stable Diffusion** For our proposed AnomalyAny, we set the inference steps to 100 and $\gamma = 0.25$ with stable-diffusion-v1-5. For optimization, we set $\lambda = 10$ and $\Delta t$ to $1.0/T$. As implemented in [6], the maximum value threshold to stop the iterative optimization at one time step $t$ is set to $0.05, 0.5, 0.8$, increasing with the denoising diffusion process. All experiments are run on a single NVIDIA A100-SXM4-80GB GPU.

**Anomaly detection framework** We adapt the anomaly detection framework proposed in AnomalyGPT [13], which deploys CLIP [25] to compute the vision-language and vision-vision similarities and aggregate these similarities for anomaly detection. The similarity between visual tokens and text embeddings for normal/anomalous states can indicate the abnormal level of visual tokens [18]. Specifically, for a given image, we first extract its patch tokens $\mathbf{F}^i_{patch} \in \mathbb{R}^{H_i \times W_i \times C_i}$ and image token $\mathbf{F}_{image} \in \mathbb{R}^{1 \times C}$ using the CLIP visual encoder, where $i$ indicates the tokens that are extracted from the $i$-th stage of the image encoder. Then, text embeddings $\mathbf{F}_{text} \in \mathbb{R}^{2 \times C}$ representing normal/abnormal states are extracted via the CLIP text encoder. Since the extracted patch tokens have not undergone the final image-text alignment and cannot be directly compared with text features, we use a lightweight feature adapter comprising only a linear layer to project patch tokens for both fine-tuning and dimension alignment between visual and text embeddings, producing $\hat{\mathbf{F}}^i_{patch} \in \mathbb{R}^{H_i \times W_i \times C}$. The detection and localization results based on vision-language similarity can then be obtained as follows:

$$S_{VL} = \text{softmax}(\mathbf{F}_{image} \cdot \mathbf{F}^T_{text}), \tag{14}$$

$$M_{VL} = \text{Upsample}\left(\sum_{i \in \mathcal{H}} \text{softmax}(\hat{\mathbf{F}}^i_{patch} \cdot \mathbf{F}^T_{text})\right), \tag{15}$$

where $\mathcal{H}$ is the list of selected stages, and $S_{VL}$ and $M_{VL}$ denote the image- and pixel-level anomaly scores, respectively. When some normal samples are available, we utilize the same visual encoder and feature adapter to extract multi-hierarchy normal patch tokens and store them in memory banks $\mathbf{B}^i \in \mathbb{R}^{N_i \times C}$. Then, for the testing patch tokens, we compute the distance between each token and its most similar counterpart in the memory bank, and the localization result $M_{VV}$ based on vision-vision similarities is yielded as follows:

$$M_{VV} = \text{Upsample}\left(\sum_{i \in \mathcal{H}} (1 - \max(\hat{\mathbf{F}}^i_{patch} \cdot (\mathbf{B}^i)^T))\right) \tag{16}$$

The maximum value of $M_{VV}$ is taken as the image-level anomaly scores $S_{VV}$. The predictions from vision-language and vision-vision similarities are summed up as final predictions. We train the anomaly detection model on the available normal sample and the synthetic samples for 200 epochs with batch size 16. We use Adam optimizer with a learning rate of 1e4 and the CosineAnnealingLR scheduler.

## B. User Study on Anomaly Generation Quality

To better assess the quality of our generated anomalous samples, we conducted a user study with 20 participants. The participants were shown exemplar normal samples of the five tested categories and asked to choose the most realistic anomalous images. We provided the participants with two groups of samples, as shown in Figure 11. In group 1, we randomly sampled three images each from 100 anomalous samples generated by Cut&Paste, DRAEM, NSA, AnomalyDiffusion, and our proposed AnomalyAny. Participants were asked to choose the three most realistic images for each category. For this group, we get a total vote of 300: $20(participants) \cdot 5(categories) \cdot 3(selected\ samples\ per\ category)$. In group 2, we randomly sampled two images each from real anomalies in the test set and from the 100 images generated by our method. Participants were then asked to choose the two most realistic images for each category. For this group, we get

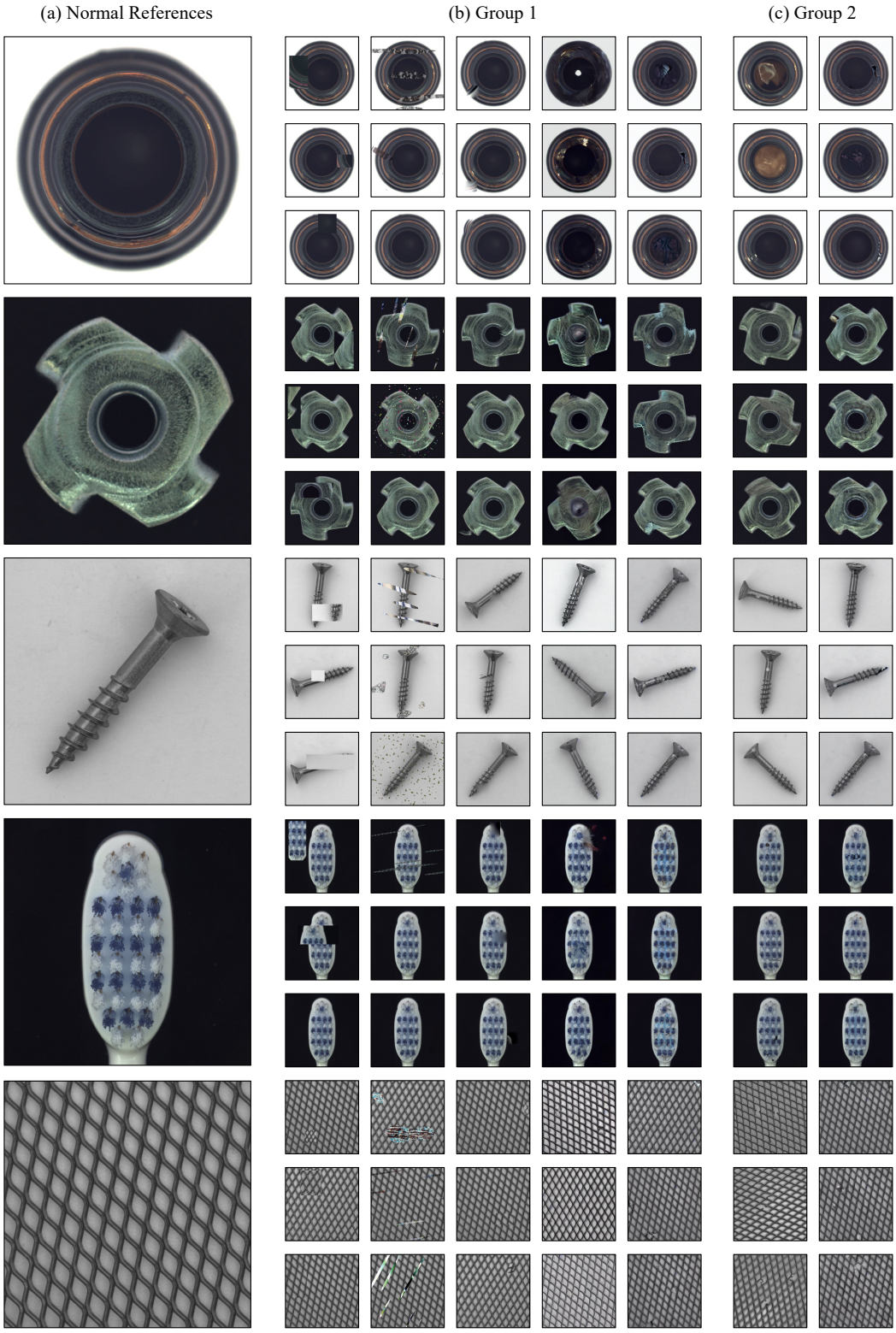


Figure 11. **Samples for user study.** (a) Normal references. (b) Group1: Samples of each column from left to right are generated by: Cut&Paste; DRAEM; NSA; AnomalyDiffusion; the proposed AnomalyAny; (c) Group2: Samples of each column from left to right are real anomalous data samples from the dataset and anomalous images generated by our method AnomalyAny.

a total vote of 200: $20(participants) \cdot 5(categories) \cdot 2(selected\ samples\ per\ category)$. The survey results are reported in Table 4 where we show the total votes from the 20 participants for each group. It is evident that our method surpasses other anomaly generation methods in terms of authenticity, even compared to AnomalyDiffusion, which leverages real test samples for training. Additionally, when mixed with real normal samples, our generated anomalous images are realistic enough to be misclassified.

| Group 1 | | | | | Group 2 | |
|---|---|---|---|---|---|---|
| Cut&Paste | DRAEM | NSA | AnomalyDiffusion | **Ours** | Real | **Ours** |
| 6 | 33 | 61 | 64 | 136 | 99 | 101 |

Table 4. Results of **user study** for anomaly generation quality assessment across two groups.

## C. More Anomaly Generation Results

More comparisons of anomaly generation results between other anomaly generation methods and AnomalyAny are provided in Figure 12. Additionally, to show the generalization ability and controllability of AnomalyAny, Figure 13 show more generation results with different object types and anomaly types with descriptions provided in Table 5.

## D. Ablations on Mask Control

In Figure 14, we show results using manually labeled regional masks for more precise anomaly locations, demonstrating our finer controllability over anomaly regions when needed.

## E. Ablations on Hyperparameters

In Figure 15, we present the visual results ablations on the hyperparameters $\lambda$ and $\Delta t$ in Equation (9). It shows that a large $\lambda$ value causes artifacts due to overly fast updates, while a small $\lambda$ value results in insufficient optimization. A large $\Delta t$ introduces artifacts from fast updates early in the process, while a small $\Delta t$ leads to artifacts from inadequate detail refinement in the final steps.

## F. More Ablations on attention-guided & prompt-guided optimization

In Figure 16, we provide more results from ablations on different optimization strategies, as shown in (b)(c)(d), and on prompt-guided optimization objectives, as shown in (d)(e)(f)(g). The results from ablations on different optimization strategies further validate that our attention guidance module effectively enforces the generation of specified

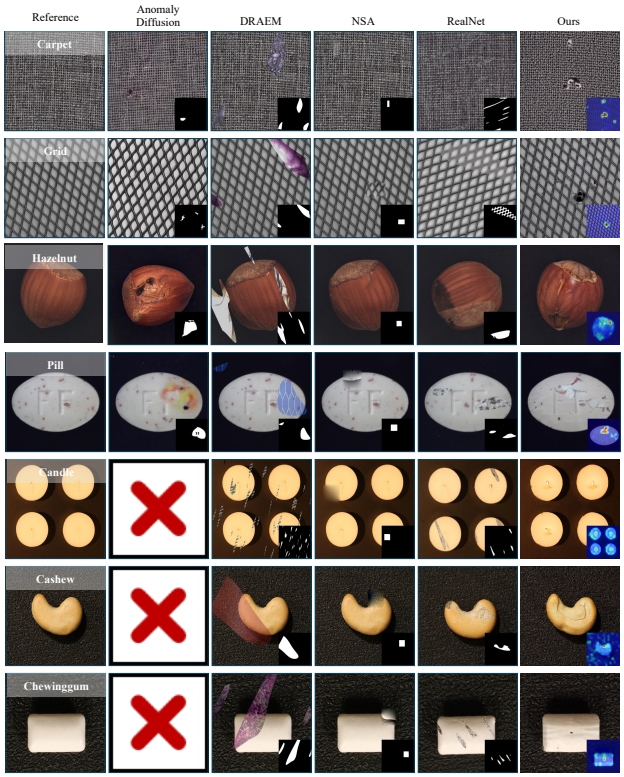


Figure 12. **Qualitative comparisons between existing anomaly generation methods.** Since AnomalyDiffusion does not provide results on VisA, its corresponding generation results are replaced by a blank.

anomalies, while the detailed prompt guidance module enhances semantic richness and improves generation quality.

For the prompt-guided optimization objectives, the best overall results are achieved with our proposed method, which incorporates all optimization objectives. We observe that in some cases (as shown in the first two rows for capsule and leather examples), the image and prompt optimization objectives significantly improve generation results for challenging concepts. In other cases, the difference is less pronounced. Across all scenarios, $L_{img}$ consistently helps generate more salient anomaly patterns in the image with more concentrated anomaly attention maps. Meanwhile, in cases where the anomaly description is less ambiguous (e.g., "rust" or "hole"), the impact of $L_{prompt}$ on optimization is relatively minor.

## G. More Anomaly Detection Results

In this section, we present additional results in Table 6 for 2-shot and 4-shot anomaly detection. Specifically, we condition on 2 and 4 normal images, generating 200 and 400 synthetic anomalous images respectively for training. Additionally, we provide per-category few-shot anomaly de-

| [CLS] | [Anomaly State] | GPT-Generated Detailed Descriptions |
|---|---|---|
| Screw | Damaged | The head of the screw is damaged or worn-out. |
| | Scratched | A screw with several deep, thin scratches. |
| | Broken on the top | The screw is partially broken on the tip. |
| | Bent | The screw is bent along its shaft, making it difficult to use. |
| | Rust | The screw is covered in rust, weakening its structure and making it difficult to use. |
| Leather | Damaged | Leather that has dried out or aged can develop damaged cracks. |
| | Scratched | Leather that has been scratched with deep, gouging lines. |
| | Cut | The leather has a visible cut, creating a split in its surface. |
| | Stained | Leather that is discolorations caused by spills (oil, ink, or dyes). |
| | Wet | Leather that appears darker, unevenly discolored, and may develop a tacky texture or water spots as it dries. |
| Bowl | Broken | A broken bowl that has visible cracks with jagged or uneven edges. |
| | Dirty | A dirty bowl that has visible stains on its surface. |
| | Colored | The bowl has uneven patches of color, giving it a stained or discolored appearance. |
| | Cracked | The bowl has a visible crack running through its surface, compromising its strength and usability. |
| | Stained | The bowl has visible stains, with patches of discoloration or residue marring its surface. |
| Vase | Broken | The vase has fragments missing, making it unusable. |
| | Cracked | The vase has a visible crack running along its surface, threatening its structural integrity. |
| | Colored | The vase has patches of uneven hues that alter its original appearance. |
| | Hole | The vase has a hole piercing its surface, preventing it from holding liquids properly. |
| | Deformed | The vase is misshapen, with uneven curves. |
| Phone Screen | Scratched | The phone screen has visible scratches, with fine lines marring its smooth surface. |
| | Broken | The phone screen is shattered, with pieces of glass splintered or missing. |
| | Cracked | The phone screen has a crack running across it |
| | Colored | The phone screen displays abnormal patches of color, such as rainbow streaks. |
| | Fingerprint | The phone screen has a smudged fingerprint, leaving an oily mark on its surface. |

Table 5. **Corresponding descriptions for anomaly generation** in Figure 13

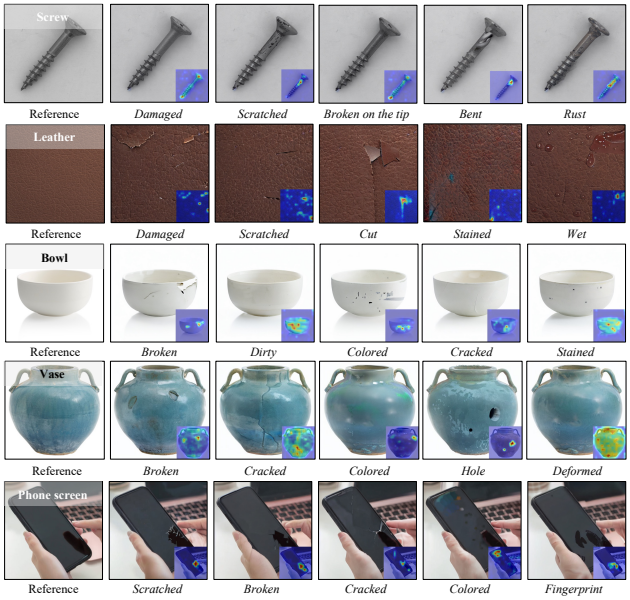


Figure 13. **Anomaly generation results** for arbitrary objects and anomaly descriptions.

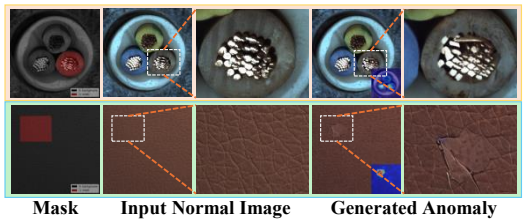


Figure 14. **Ablation on mask control.**


Figure 15. **Ablation on hyperparameters.**

tection results on MVTec and VisA in Table 8 to Table 13. Visual results of few-shot anomaly detection are provided in Figure 17. Full-shot detection results are listed in Table 7 with per-category results in Table 14 and Table 15. Specifically, we condition on all normal images, generating 3-5 anomalous images with each normal image for training. We provide t-test results of few-shot anomaly detection in Table 16 for statistical completeness.

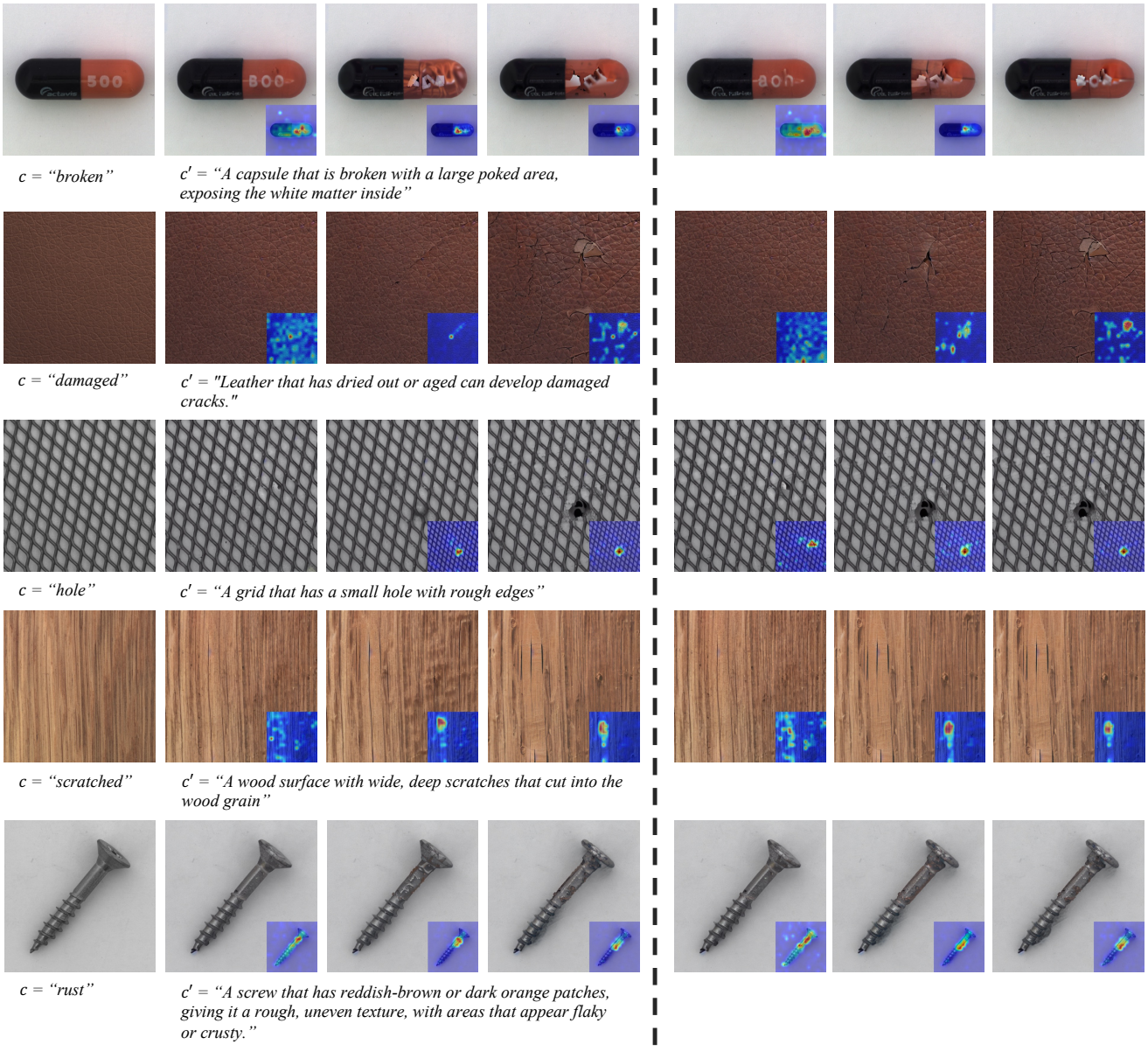


Figure 16. **Ablation on optimization strategies and objectives.**

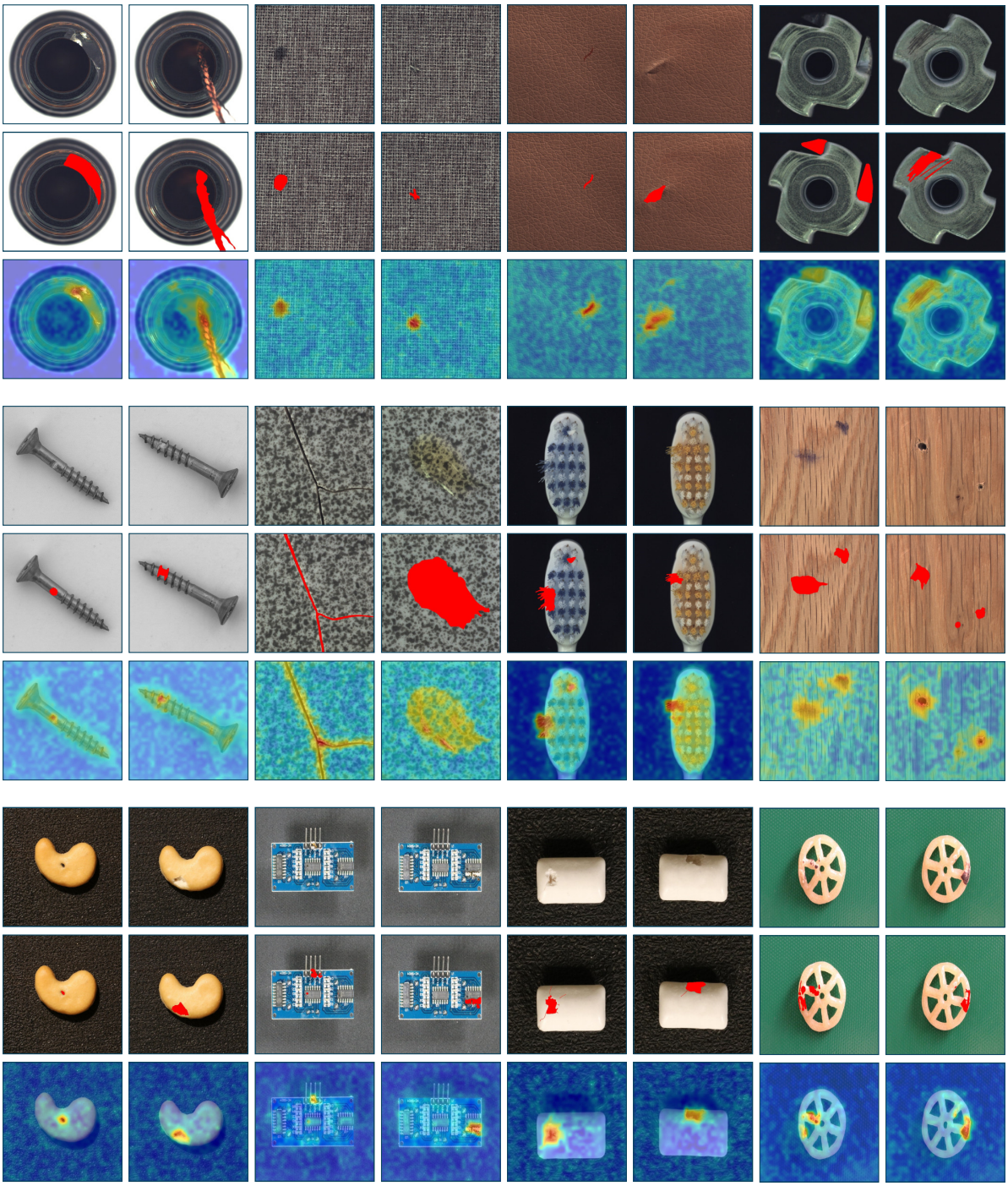

Figure 17. **Anomaly detection results in the 4-shot setup.** For each pair, the original image, ground truth, and detection results are listed from the top to the bottom.

| Setup | Methods | MVTec AD | | | | | VisA | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | I-AUC | I-F1 | P-AUC | P-F1 | PRO | I-AUC | I-F1 | P-AUC | P-F1 | PRO |
| 1-shot | PaDiM | 76.6±3.1 | 88.2±1.1 | 89.3±0.9 | 40.2±2.1 | 73.3±2.0 | 62.8±5.4 | 75.3±1.2 | 89.9±0.8 | 17.4±1.7 | 64.3±2.4 |
| | PatchCore | 83.4±3.0 | 90.5±1.5 | 92.0±1.0 | 50.4±2.1 | 79.7±2.0 | 79.9±2.9 | 81.7±1.6 | 95.4±0.6 | 38.0±1.9 | 80.5±2.5 |
| | WinCLIP+ | 93.1±2.0 | 93.7±1.1 | 95.2±0.5 | 55.9±2.7 | 87.1±1.2 | 83.8±4.0 | 83.1±1.7 | 96.4±0.4 | 41.3±2.3 | 85.1±2.1 |
| | AnomalyGPT | 94.1±1.1 | - | 95.3±0.1 | - | - | 87.4±0.8 | - | 96.2±0.1 | - | - |
| | PromptAD | 94.6±1.7 | - | **95.9±0.5** | - | - | 86.9±2.3 | - | 96.7±0.4 | - | - |
| | **Ours** | **94.9±0.4** | **94.7±0.4** | 95.4±0.2 | **57.3±0.0** | **91.9±0.0** | **89.7±0.8** | **85.8±0.5** | **97.7±0.4** | **43.2±0.4** | **92.5±0.1** |
| 2-shot | PaDiM | 78.9±3.1 | 89.2±1.1 | 91.3±0.7 | 43.7±1.5 | 78.2±1.8 | 67.4±5.1 | 75.7±1.8 | 92.0±0.7 | 21.1±2.4 | 70.1±2.6 |
| | PatchCore | 86.3±3.3 | 92.0±1.5 | 93.3±0.6 | 53.0±1.7 | 82.3±1.3 | 81.6±4.0 | 82.5±1.8 | 96.1±0.5 | 41.0±3.9 | 82.6±2.3 |
| | WinCLIP+ | 94.4±1.3 | 94.4±0.8 | 96.0±0.3 | 58.4±1.7 | 88.4±0.9 | 84.6±2.4 | 83.0±1.4 | 96.8±0.3 | 43.5±3.3 | 86.2±1.4 |
| | AnomalyGPT | 95.5±0.8 | - | 95.6±0.2 | - | - | 88.6±0.7 | - | 96.4±0.1 | - | - |
| | PromptAD | 95.7±1.5 | - | **96.2±0.3** | - | - | 88.3±2.0 | - | 97.1±0.3 | - | - |
| | **Ours** | **95.8±0.2** | **95.2±0.2** | 96.0±0.2 | **58.8±0.2** | **92.6±0.1** | **91.3±0.4** | **87.2±0.6** | **97.9±0.4** | **44.9±0.3** | **92.7±0.1** |
| 4-shot | PaDiM | 80.4±2.5 | 90.2±1.2 | 92.6±0.7 | 46.1±1.8 | 81.3±1.9 | 72.8±2.9 | 78.0±1.2 | 93.2±0.5 | 24.6±1.8 | 72.6±1.9 |
| | PatchCore | 88.8±2.6 | 92.6±1.6 | 94.3±0.5 | 55.0±1.9 | 84.3±1.4 | 85.3±2.1 | 84.3±1.3 | 96.8±0.3 | 43.9±3.1 | 84.9±1.4 |
| | WinCLIP+ | 95.2±1.3 | 94.7±0.8 | 96.2±0.3 | 59.5±1.8 | 89.0±0.8 | 87.3±1.8 | 84.2±1.6 | 97.2±0.2 | 47.0±3.0 | 87.6±0.9 |
| | AnomalyGPT | 96.3±0.3 | - | 96.2±0.1 | - | - | 90.6±0.7 | - | 96.7±0.1 | - | - |
| | PromptAD | **96.6±0.9** | - | **96.5±0.2** | - | - | 89.1±1.7 | - | 97.4±0.4 | - | - |
| | **Ours** | 96.4±0.1 | **95.1±0.1** | 96.2±0.1 | **59.8±0.1** | **93.0±0.0** | **91.7±1.0** | **87.1±0.1** | **97.8±0.4** | **47.9±0.2** | **93.4±0.1** |

Table 6. **Comparison of few-shot anomaly detection on MVTec AD and VisA.** Results are reported over 5 runs. The best results are in **bold**, and the second-best results are underlined.

| Methods | MVTec AD | | | | | VisA | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | I-AUC | I-F1 | P-AUC | P-F1 | PRO | I-AUC | I-F1 | P-AUC | P-F1 | PRO |
| UniAD | 96.5 | 98.8 | 96.8 | 43.4 | 90.7 | 88.8 | 90.8 | 98.3 | 33.7 | 85.5 |
| SimpleNet | 95.3 | 98.4 | 96.9 | 45.9 | 86.5 | 87.2 | 87.0 | 96.8 | 34.7 | 81.4 |
| DiAD | 97.2 | **99.0** | 96.8 | 52.6 | 90.7 | 86.8 | 88.3 | 96.0 | 26.1 | 75.2 |
| AnomalyGPT | 97.4 | - | 93.1 | - | - | - | - | - | - | - |
| **Ours** | **98.4** | 96.9 | **97.4** | **65.1** | **94.7** | **95.8** | **91.9** | **98.7** | **58.7** | **97.7** |

Table 7. **Comparison of full-shot anomaly detection on MVTec AD and VisA.** The best results are in **bold**, and the second-best results are underlined.

| Category | I-AUC | I-F1 | P-AUC | P-F1 | PRO |
|---|---|---|---|---|---|
| bottle | 98.9±0.1 | 97.9±0.5 | 96.3±0.0 | 71.0±0.2 | 92.7±0.0 |
| cable | 89.1±7.6 | 87.8±4.5 | 91.7±0.0 | 36.4±2.2 | 82.8±0.2 |
| capsule | 93.8±1.6 | 94.4±0.5 | 97.3±0.0 | 44.8±0.0 | 96.2±0.0 |
| carpet | 100.0±0.0 | 100.0±0.0 | 99.0±0.0 | 75.1±0.1 | 97.5±0.0 |
| grid | 97.3±1.3 | 96.1±2.2 | 97.2±0.0 | 49.8±0.1 | 92.0±0.1 |
| hazelnut | 99.9±0.0 | 99.3±0.2 | 98.4±0.0 | 63.5±0.1 | 97.3±0.0 |
| leather | 100.0±0.0 | 100.0±0.0 | 99.7±0.0 | 62.2±0.2 | 99.5±0.0 |
| metal_nut | 95.6±9.3 | 95.3±4.3 | 91.6±0.1 | 58.8±0.5 | 90.7±0.0 |
| pill | 93.7±0.4 | 96.0±0.1 | 94.3±0.0 | 56.6±0.2 | 97.0±0.0 |
| screw | 74.9±3.1 | 87.2±1.3 | 97.8±0.0 | 42.5±0.3 | 92.3±0.0 |
| tile | 99.6±0.0 | 98.8±0.2 | 95.6±0.0 | 73.2±0.0 | 93.3±0.0 |
| toothbrush | 94.1±2.7 | 93.1±5.6 | 98.5±0.0 | 56.6±1.3 | 94.7±0.0 |
| transistor | 91.4±2.1 | 80.7±4.4 | 79.6±0.0 | 37.1±0.2 | 64.0±0.2 |
| wood | 99.6±0.0 | 98.5±0.1 | 96.6±0.0 | 70.0±0.0 | 96.8±0.0 |
| zipper | 96.4±0.8 | 96.1±0.4 | 97.4±0.0 | 61.4±0.4 | 91.4±0.0 |
| Average | 94.9±0.4 | 94.7±0.4 | 95.4±0.2 | 57.3±0.0 | 91.9±0.0 |

Table 8. **Per-category anomaly detection performance on MVTec AD in the 1-shot setup.** We report the mean and standard deviation over 5 random seeds for each measurement.

| Category | I-AUC | I-F1 | P-AUC | P-F1 | PRO |
|---|---|---|---|---|---|
| bottle | 99.5±0.1 | 98.6±0.1 | 93.2±0.1 | 62.3±1.8 | 88.7±0.4 |
| cable | 89.3±1.4 | 86.5±2.0 | 94.9±0.0 | 41.1±0.8 | 86.0±0.3 |
| capsule | 95.1±1.1 | 94.6±0.9 | 96.6±0.2 | 44.9±0.0 | 96.1±0.1 |
| carpet | 100.0±0.0 | 100.0±0.0 | 99.3±0.0 | 76.5±0.1 | 98.1±0.0 |
| grid | 95.4±1.2 | 93.5±1.9 | 98.2±0.0 | 52.2±0.2 | 93.9±0.0 |
| hazelnut | 99.8±0.0 | 98.9±0.1 | 98.0±0.0 | 57.4±0.3 | 96.9±0.0 |
| leather | 100.0±0.0 | 100.0±0.0 | 99.7±0.0 | 66.1±0.3 | 99.5±0.0 |
| metal_nut | 99.8±0.0 | 99.1±0.4 | 94.6±0.1 | 66.7±0.5 | 93.2±0.1 |
| pill | 96.4±0.8 | 96.8±0.3 | 94.8±0.0 | 59.0±0.3 | 96.9±0.0 |
| screw | 78.8±3.2 | 88.4±0.1 | 98.2±0.0 | 46.6±0.2 | 93.5±0.0 |
| tile | 100.0±0.0 | 99.9±0.1 | 97.1±0.0 | 74.5±0.0 | 94.8±0.0 |
| toothbrush | 93.4±0.9 | 93.1±0.1 | 98.7±0.0 | 59.5±1.0 | 94.3±0.0 |
| transistor | 88.9±2.0 | 82.5±4.1 | 85.1±0.2 | 41.2±0.0 | 67.6±0.7 |
| wood | 99.4±0.0 | 97.7±0.1 | 96.9±0.0 | 70.9±0.1 | 97.0±0.0 |
| zipper | 99.4±0.0 | 98.3±0.4 | 97.7±0.0 | 63.6±1.5 | 92.8±0.4 |
| Average | 95.8±0.1 | 95.2±0.2 | 96.0±0.2 | 58.8±0.2 | 92.6±0.1 |

Table 9. **Per-category anomaly detection performance on MVTec AD in the 2-shot setup.** We report the mean and standard deviation over 5 random seeds for each measurement.

| Category | I-AUC | I-F1 | P-AUC | P-F1 | PRO |
|---|---|---|---|---|---|
| bottle | 99.2±0.0 | 98.0±0.5 | 96.9±0.0 | 73.0±0.1 | 93.7±0.0 |
| cable | 90.6±0.8 | 87.1±3.1 | 93.7±0.0 | 43.4±0.9 | 86.6±0.1 |
| capsule | 96.4±0.4 | 95.8±0.2 | 98.1±0.0 | 46.9±0.1 | 97.3±0.0 |
| carpet | 100.0±0.0 | 100.0±0.0 | 99.2±0.0 | 75.7±0.0 | 97.8±0.0 |
| grid | 98.4±0.8 | 96.9±2.6 | 98.1±0.0 | 52.4±0.0 | 93.6±0.0 |
| hazelnut | 98.9±0.2 | 97.6±0.4 | 98.6±0.0 | 63.1±0.3 | 97.3±0.0 |
| leather | 100.0±0.0 | 100.0±0.0 | 99.7±0.0 | 63.2±0.2 | 99.4±0.0 |
| metal_nut | 99.3±0.3 | 98.4±0.4 | 93.2±0.0 | 64.7±0.1 | 92.8±0.0 |
| pill | 97.3±0.1 | 97.4±0.1 | 94.9±0.0 | 59.4±0.1 | 97.3±0.0 |
| screw | 85.5±0.1 | 88.6±0.1 | 98.3±0.0 | 49.0±0.3 | 93.6±0.0 |
| tile | 99.8±0.0 | 98.7±0.1 | 95.9±0.0 | 73.3±0.0 | 93.6±0.0 |
| toothbrush | 94.3±1.3 | 94.1±0.8 | 98.6±0.0 | 60.2±0.9 | 94.2±0.1 |
| transistor | 87.5±1.9 | 75.9±4.1 | 82.8±0.0 | 39.5±0.2 | 67.4±0.1 |
| wood | 99.5±0.0 | 97.9±0.2 | 96.5±0.0 | 69.6±0.0 | 96.7±0.0 |
| zipper | 98.8±0.1 | 98.5±0.2 | 97.9±0.0 | 64.5±0.1 | 93.2±0.0 |
| Average | 96.4±0.1 | 95.1±0.1 | 96.2±0.1 | 59.8±0.1 | 93.0±0.0 |

Table 10. **Per-category anomaly detection performance on MVTec AD in the 4-shot setup.** We report the mean and standard deviation over 5 random seeds for each measurement.

| Category | I-AUC | I-F1 | P-AUC | P-F1 | PRO |
|---|---|---|---|---|---|
| candle | 90.8±0.3 | 85.3±0.7 | 98.9±0.0 | 37.6±0.1 | 98.0±0.0 |
| capsules | 91.1±2.6 | 88.5±0.4 | 98.2±0.0 | 47.9±2.0 | 94.2±0.1 |
| cashew | 88.9±16.6 | 88.9±5.1 | 96.6±0.1 | 59.3±0.9 | 95.6±0.0 |
| chewinggum | 97.4±0.1 | 95.2±0.3 | 99.6±0.0 | 77.1±0.2 | 92.6±0.1 |
| fryum | 96.2±0.9 | 94.7±0.8 | 95.4±0.0 | 40.8±0.1 | 92.2±0.0 |
| macaroni1 | 86.6±5.2 | 79.6±5.9 | 99.7±0.0 | 30.1±0.8 | 96.2±0.1 |
| macaroni2 | 79.2±2.1 | 73.0±1.2 | 98.4±0.0 | 28.3±1.0 | 90.5±0.7 |
| pcb1 | 90.8±2.0 | 86.1±1.5 | 98.5±0.0 | 49.5±4.6 | 93.6±0.0 |
| pcb2 | 84.3±4.1 | 78.1±4.7 | 96.2±0.0 | 30.8±0.4 | 83.7±0.0 |
| pcb3 | 78.0±5.2 | 74.3±1.7 | 94.9±0.0 | 32.7±11.4 | 86.0±0.0 |
| pcb4 | 96.8±1.0 | 92.4±3.7 | 96.9±0.0 | 36.1±0.3 | 90.2±0.3 |
| pipe_fryum | 96.4±0.6 | 93.1±2.3 | 98.5±0.0 | 47.7±0.1 | 97.6±0.0 |
| Average | 89.7±0.8 | 85.8±0.5 | 97.7±0.4 | 43.2±0.4 | 92.5±0.1 |

Table 11. **Per-category anomaly detection performance on VisA in the 1-shot setup.** We report the mean and standard deviation over 5 random seeds for each measurement.

| Category | I-AUC | I-F1 | P-AUC | P-F1 | PRO |
|---|---|---|---|---|---|
| candle | 90.5±0.5 | 84.4±0.9 | 98.9±0.0 | 39.0±0.1 | 98.0±0.0 |
| capsules | 94.5±1.1 | 91.0±1.1 | 98.3±0.0 | 49.3±0.5 | 94.5±0.2 |
| cashew | 90.7±0.7 | 89.5±0.2 | 97.0±0.0 | 59.6±0.1 | 95.9±0.0 |
| chewinggum | 97.6±0.2 | 95.2±0.5 | 99.5±0.0 | 76.5±0.3 | 92.4±0.0 |
| fryum | 97.0±0.1 | 95.0±0.2 | 96.1±0.0 | 44.9±0.1 | 92.6±0.1 |
| macaroni1 | 89.6±0.1 | 82.4±0.6 | 99.7±0.0 | 29.3±0.3 | 96.3±0.0 |
| macaroni2 | 77.4±19.9 | 72.7±6.9 | 98.4±0.0 | 28.3±0.4 | 88.8±0.0 |
| pcb1 | 92.0±0.5 | 87.7±0.5 | 98.5±0.0 | 49.3±3.0 | 93.9±0.0 |
| pcb2 | 86.7±2.0 | 80.2±2.0 | 96.8±0.0 | 34.5±1.9 | 84.5±0.1 |
| pcb3 | 84.6±0.4 | 78.6±0.2 | 95.4±0.0 | 40.5±4.5 | 86.1±0.1 |
| pcb4 | 97.0±0.2 | 93.2±0.2 | 97.1±0.0 | 35.9±0.2 | 91.1±0.2 |
| pipe_fryum | 97.9±0.0 | 95.2±0.9 | 98.7±0.0 | 51.2±0.1 | 97.6±0.0 |
| Average | 91.3±0.4 | 87.2±0.6 | 97.9±0.4 | 44.9±0.3 | 92.7±0.1 |

Table 12. **Per-category anomaly detection performance on VisA in the 2-shot setup.** We report the mean and standard deviation over 5 random seeds for each measurement.

| Category | I-AUC | I-F1 | P-AUC | P-F1 | PRO |
|---|---|---|---|---|---|
| candle | 92.8±0.3 | 86.1±0.9 | 99.0±0.0 | 39.8±0.1 | 98.3±0.0 |
| capsules | 94.7±0.9 | 91.6±2.3 | 98.4±0.0 | 50.5±0.3 | 94.6±0.1 |
| cashew | 89.9±13.4 | 89.5±5.2 | 97.1±0.0 | 62.6±0.3 | 95.8±0.0 |
| chewinggum | 97.2±0.3 | 95.1±0.1 | 99.5±0.0 | 75.8±0.4 | 92.5±0.3 |
| fryum | 96.4±1.1 | 93.7±1.2 | 96.3±0.0 | 46.2±0.1 | 93.1±0.0 |
| macaroni1 | 89.3±2.0 | 82.2±2.8 | 99.7±0.0 | 29.0±2.2 | 96.8±0.0 |
| macaroni2 | 79.8±16.4 | 74.7±8.8 | 98.7±0.0 | 29.0±0.5 | 90.9±0.1 |
| pcb1 | 88.3±11.4 | 81.3±17.4 | 99.2±0.0 | 70.1±1.1 | 93.9±0.0 |
| pcb2 | 87.8±1.0 | 80.5±1.9 | 97.1±0.0 | 35.7±0.6 | 85.7±0.1 |
| pcb3 | 89.6±1.2 | 82.3±3.6 | 96.2±0.0 | 47.2±1.3 | 89.0±0.1 |
| pcb4 | 97.1±0.6 | 93.0±2.3 | 97.5±0.0 | 36.9±0.0 | 92.6±0.1 |
| pipe_fryum | 97.3±1.7 | 94.9±1.2 | 98.8±0.0 | 52.3±0.6 | 97.6±0.0 |
| Average | 91.7±1.0 | 87.1±0.1 | 97.8±0.4 | 47.9±0.2 | 93.4±0.1 |

Table 13. **Per-category anomaly detection performance on VisA in the 4-shot setup.** We report the mean and standard deviation over 5 random seeds for each measurement.

| Category | I-AUC | I-F1 | P-AUC | P-F1 | PRO |
|---|---|---|---|---|---|
| bottle | 99.7 | 98.4 | 97.5 | 74.9 | 94.7 |
| cable | 94.9 | 92.6 | 95.9 | 53.0 | 90.6 |
| capsule | 98.3 | 97.3 | 98.5 | 51.7 | 98.0 |
| carpet | 100.0 | 99.4 | 99.3 | 75.3 | 98.2 |
| grid | 99.8 | 99.1 | 99.1 | 52.7 | 96.6 |
| hazelnut | 99.7 | 98.6 | 99.4 | 76.3 | 98.5 |
| leather | 100.0 | 100.0 | 99.7 | 60.4 | 99.5 |
| metal_nut | 100.0 | 100.0 | 95.8 | 73.9 | 95.2 |
| pill | 98.5 | 98.3 | 96.3 | 63.6 | 97.9 |
| screw | 96.6 | 95.1 | 99.0 | 60.6 | 96.0 |
| tile | 100.0 | 99.4 | 97.2 | 74.6 | 94.8 |
| toothbrush | 96.9 | 95.2 | 99.5 | 71.5 | 97.8 |
| transistor | 94.3 | 84.2 | 88.0 | 45.2 | 70.0 |
| wood | 99.7 | 98.4 | 97.1 | 70.9 | 97.0 |
| zipper | 98.0 | 96.7 | 98.8 | 71.5 | 96.3 |
| Average | 98.4 | 96.9 | 97.4 | 65.1 | 94.7 |

Table 14. Per-category anomaly detection performance on **MVTec AD** in the **full-shot** setup.

| Category | I-AUC | I-F1 | P-AUC | P-F1 | PRO |
|---|---|---|---|---|---|
| candle | 95.6 | 90.0 | 99.3 | 40.1 | 98.2 |
| capsules | 96.2 | 93.8 | 99.1 | 60.1 | 96.6 |
| cashew | 97.4 | 94.5 | 98.2 | 70.4 | 94.6 |
| chewinggum | 98.7 | 97.0 | 99.5 | 75.3 | 91.8 |
| fryum | 98.4 | 97.5 | 97.4 | 53.6 | 93.9 |
| macaroni1 | 95.3 | 88.5 | 99.9 | 36.4 | 98.5 |
| macaroni2 | 84.7 | 79.3 | 99.6 | 28.9 | 96.7 |
| pcb1 | 95.9 | 91.8 | 98.8 | 41.8 | 95.4 |
| pcb2 | 94.1 | 88.2 | 98.2 | 40.3 | 91.5 |
| pcb3 | 95.9 | 90.4 | 97.5 | 52.9 | 93.4 |
| pcb4 | 99.4 | 96.6 | 98.4 | 46.5 | 94.9 |
| pipe_fryum | 98.4 | 95.9 | 99.1 | 58.7 | 97.9 |
| Average | 95.8 | 91.9 | 98.7 | 50.4 | 95.3 |

Table 15. Per-category anomaly detection performance on **VisA** in the **full-shot** setup.

| Method | PaDiM | PatchCore | WinCLIP+ | AnomalyGPT | PromptAD |
|---|---|---|---|---|---|
| p-value | $5e-7$ | $8e-5$ | $1e-7$ | 0.005 | 0.7 |

Table 16. t-test on anomaly detection results.