LACToS: Local Anomaly detection via Conditional Typicality of Superpixels

Anonymous authors
Paper under double-blind review

Abstract

Local anomaly detection is the task of identifying spatially confined anomalous or out-of-distribution (OOD) regions within an otherwise in-distribution (ID) image. Most existing methods are designed for a specific context, with a strong focus on industrial defect detection, where they achieve excellent performance on well-established benchmarks. However, we observe that these methods can exhibit significant performance drops in other contexts, especially when the inlier distribution is multimodal. To account for this inlier complexity, we propose LACToS, a principled local OOD detection method based on statistical typicality and made tractable through two innovations: (i) grouping pixels into superpixels and (ii) using a computationally efficient proxy to estimate typicality. We evaluate LACToS across three diverse applications spanning medical imaging, human faces, and X-ray security screening. LACToS performs more robustly than prior methods across these benchmarks.

1 Introduction

Local anomaly detection is the task of identifying out-of-distribution (OOD) regions that are spatially confined within an otherwise in-distribution (ID) image. It is central to applications such as industrial defect inspection, medical image analysis, and security screening, where an anomalous region may occupy only a small fraction of the image yet indicate a defect, tumor, or weapon. Unlike image-level OOD detection, which assigns a single anomaly score to an entire image, localized pixel-level anomaly detection involves generating a dense heatmap that indicates which pixels or regions of each outlier image are inconsistent with the inlier distribution. This makes the problem especially challenging: a local OOD detector must be sensitive to subtle local deviations while remaining robust to natural variation in normal image content.

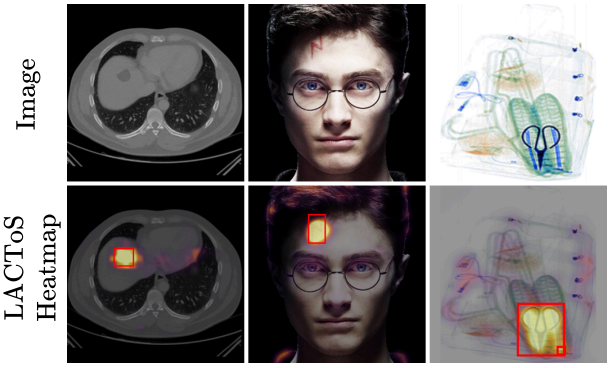


Figure 1: Using a diffusion model pretrained on ID images only, LACToS is able to identify locally OOD regions in images across various tasks, including tumors in CT scans, anomalies such as scars on faces, and harmful objects such as scissors in X-ray baggage scans.

Prior work on localized anomaly detection has largely followed representation-based and reconstruction-based approaches. Representation-based methods extract local features from normal (inlier) training images, and score test patches using their distance from the learned distribution or their likelihood under it (Cohen and Hoshen, 2020; Defard et al., 2021; Roth et al., 2022). Reconstruction-based methods instead reconstruct, inpaint, or denoise the target image under a prior for inlier images, and use the discrepancy between the input and its reconstruction as an anomaly signal (Mousakhan et al., 2024; Wu et al., 2024). Although both approaches have achieved strong empirical performance, especially for industrial defect detection (Bergmann et al., 2019), their anomaly scores are often heuristic: representation-based methods depend on the choice of feature and distance metric, while

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

reconstruction-based methods typically collapse the distribution of inlier completions into a point estimate. Conformal and risk-control methods can provide statistical guarantees for calibrated segmentation outputs (Shen and Liu, 2025), but they generally rely on large amount of OOD calibration data. Moreover, we empirically observe that these methods do not always generalize reliably beyond industrial defect detection.

We present LACToS, a statistical method for local OOD detection that is designed to address these limitations of existing methods. Our approach is fundamentally reconstruction-based, but uses the key insight that the distribution of the reconstructions contains valuable information that is lost when reconstruction error is computed using a single sample or an aggregated point estimate. To leverage this richer distributional information towards improved local anomaly detection, we propose an anomaly metric based on statistical typicality, a measure of how representative a sequence of observations is for a given distribution. Our use of typicality is inspired by previous typicality-based OOD detection methods (Nalisnick et al., 2019; Osada et al., 2023; Barkley et al., 2026). However, these works are focused on image-level OOD detection where an entire image is classified as OOD, and to the best of our knowledge, using typicality for local anomaly detection has not yet been explored.

We begin with the typicality score of Nalisnick et al. (2019), and adapt it to identify local rather than whole-image anomalies (Section 4.1). Instead of asking whether an entire image is typical with respect to the unconditional prior distribution, we ask whether each pixel is typical with respect to the *inpainting posterior* conditioned on all other pixels in the image. However, this substitution induces three major challenges: (i) failure to detect larger local anomalies that span multiple contiguous pixels, (ii) loss of statistical validity, because typicality testing requires sampling a sequence of i.i.d. observations, and (iii) prohibitive computational cost of performing posterior sampling separately for each pixel.

LACToS tackles these challenges through two innovations: we partition images into superpixels and replace pixel-level typicality with superpixel-level typicality (Section 4.2), and we use a single *denoising posterior* as a common proxy for all inpainting posteriors. Our use of superpixels addresses challenges (i) and (ii), because a local OOD feature is more likely to be captured by a superpixel than a single pixel, and the individual pixels in each superpixel can be treated as independent observations to evaluate superpixel-level typicality. Our use of denoising rather than inpainting addresses challenge (iii), by dramatically reducing the number of posterior samples needed to estimate typicality. Through these innovations, we show that LACToS achieves strong quantitative and qualitative performance across a wide range of local OOD detection tasks including detecting tumors in CT scans, detecting unusual facial accessories in photographs, and detecting weapons in X-ray baggage scans.

To summarize, we propose LACToS, a statistical approach to local OOD detection in images using an inlier diffusion model and typicality testing. Our method (1) is grounded in statistical typicality, and (3) is competitive with, and often more precise than, existing methods across diverse local OOD detection tasks with complex inlier distributions.

2 PRELIMINARIES

Appendix section A provides a brief introduction to diffusion models and their use in posterior sampling, which we use as a sub-component of LACToS. Here, we provide context on local OOD detection methods proposed in prior work, some of which inspired our use of statistical typicality in LACToS.

OOD detection is studied at both the *image* and *pixel* levels. Image-level OOD detection assigns a single anomaly score to an entire image, with the goal of identifying images that are not drawn from the inlier distribution seen during training. In contrast, pixel-level OOD detection produces a dense anomaly heatmap to localize anomalous regions within each image. Formally, given an image $x^\star$, the goal is to estimate the ground truth anomaly mask $m^\star$. Previous works (Mousakhan et al., 2024; Wyatt et al., 2022; Defard et al., 2021) define an anomaly score $f(x^\star, j)$ that captures how likely each pixel $j \in \mathcal{J}$ is to be OOD, where $\mathcal{J}$ is an index set of size $W \times H$ over all pixels. Then, they compute an anomaly heatmap A with elements $A_j = f(x^\star, j)$ for each pixel, fix some threshold τ, and set $\hat{m} = \mathbb{1}\{A \geq \tau\}$. Many of these works, especially those using diffusion models, are reconstruction-based: they generate an in-distribution reconstruction of the input using posterior

sampling and use discrepancies such as ℓ_1, ℓ_2, SSIM (Wang et al., 2004), LPIPS (Zhang et al., 2018), or feature-space distances as anomaly scores (Zavrtanik et al., 2021; Wu et al., 2024; Mousakhan et al., 2024; Pirnay and Chai, 2022). Although effective, these methods often collapse the distribution of plausible reconstructions into a first-order statistic, such as the mean reconstruction error, which does not capture the full reconstruction and can reduce robustness to complex inlier distributions.

Meanwhile, distribution-aware statistics have been explored across several image-level OOD detection methods. For example, EigenScore (Shoushtari et al., 2025) uses the eigenvalue spectrum of the posterior covariance, and DDPM-OOD (Graham et al., 2023) aggregates standardized reconstruction statistics. There are also typicality-based methods that compare the negative log-likelihood of the samples with the entropy of the inlier distribution (Nalisnick et al., 2019; Osada et al., 2023; Barkley et al., 2026). Specifically, let $(x^{(1)}, \ldots, x^{(N)})$ be independent draws from the inlier distribution p. The (ϵ, N)-typical set is defined as

$$\mathcal{T}^{p}_{\epsilon,N} = \left\{ (x^{(1)}, \ldots, x^{(N)}) \;\middle|\; \left| -\frac{1}{N} \log p(x^{(1)}, \ldots, x^{(N)}) - \mathbb{H}(p) \right| \leq \epsilon \right\}, \tag{1}$$

where $\mathbb{H}(p) = \mathbb{E}_{x\sim p}[-\log p(x)]$ is the entropy of p. By the Asymptotic Equipartition Property (Cover et al., 1991),

$$-\frac{1}{N} \log p(X^{(1)}, \ldots, X^{(N)}) \xrightarrow[N\to\infty]{} \mathbb{H}(p). \tag{2}$$

For image-level OOD detection, Nalisnick et al. (2019) estimate the deviation from typicality as

$$\hat{\epsilon}(x^{(1)}, \ldots, x^{(N)}) := \left| -\frac{1}{N} \sum_{n=1}^{N} \log p(x^{(n)}) - \mathbb{H}(p) \right|, \tag{3}$$

and classify a set of samples as OOD when $\hat{\epsilon}$ exceeds a threshold. Unlike likelihood alone, typicality accounts for whether the sampled distribution is consistent with the inlier distribution, and has been shown to be a more effective image-level OOD detection metric (Nalisnick et al., 2019; Osada et al., 2023).

3 Motivation

The local OOD detection problem has been extensively studied in the context of industrial defect detection. Prior works such as PatchCore (Roth et al., 2022), EfficientAD (Batzner et al., 2024), GLASS (Chen et al., 2024), and diffusion-based DDAD (Mousakhan et al., 2024) achieve excellent performance on widely used benchmarks such as MVTec AD (Bergmann et al., 2019) and VisA (Zou et al., 2022). However, we observe that these methods do not extend robustly to other OOD detection tasks, such as tumor detection, facial anomaly detection, and security screening using X-ray images, as shown in Table 1. One potential cause of this brittleness is the complexity of the inlier distribution. Industrial defect-detection datasets often have very structured inlier distributions, as products without defects are visually near-homogeneous, whereas CT scans, images of human faces, and X-ray baggage scans are far more heterogeneous even in the absence of anomalies. Moreover, defects in products are context-agnostic: a defective feature, such as a crack or fracture, is anomalous regardless of where in the image or product it occurs. In contrast, a dark region in a CT scan may be innocuous normal anatomy in one location, or indicative of a tumor in another.

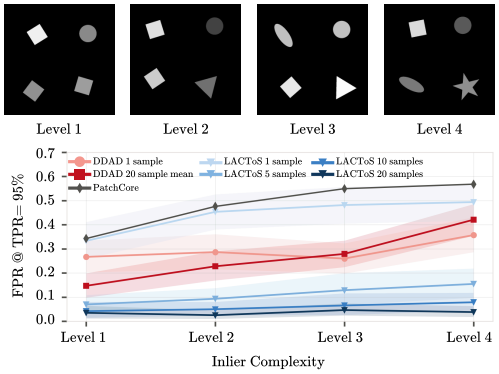


Figure 2: **Multimodality Toy Experiment.** (Top) Images from the inlier distributions across different complexity levels. The top-right corner always contains a circle, while other corners can contain different shapes, with a wider variety of shapes at higher levels. (Bottom) False positive rates at true positive rate of 95%, evaluated on images that contain squares in the top-right corner.

Specifically, when the conditional distribution over a local region given the rest of the image is multimodal, methods that rely on a first-order statistic may erroneously identify samples from a different inlier mode as anomalous. To illustrate this phenomenon through a toy experiment, we construct four inlier distributions with increasing complexity levels, and train a diffusion model on each inlier distribution. Each image contains four shapes, one per quadrant. In the upper-right quadrant, every inlier image contains a circle, which allows us to define all other shapes placed in that quadrant as local anomalies. The other three quadrants can contain a wider variety of shapes, as shown in the top row of Figure 2, with the candidate shapes at each level occurring with equal probability. From level 1 to level 4, the set of shapes that can be placed in these three quadrants progressively grows, which increases the number of modes of the inlier distribution. We then introduce local anomalies in test images by replacing the top-right circle with a square, and evaluate two baseline methods, DDAD (Mousakhan et al., 2024) and PatchCore (Roth et al., 2022), on how well they can correctly identify only the square as an anomaly. The bottom row of Figure 2 demonstrates that at a per-image true positive rate (TPR) of 95%, the false positive rate (FPR) increases with inlier multimodality for both baseline methods. This trend holds across the TPR values shown in Figure 7. For more information on the experimental setup, please refer to Section B.

4 METHOD

To tackle this problem, we propose LACToS, a typicality-based pixel-level OOD detection method using diffusion models, which is much more robust to inlier multimodality. In Section 4.1, we first describe an idealized extension of an existing typicality-based metric for image-level OOD detection to pixel-level anomaly localization, and several issues that limit its direct application. Then, in Section 4.2 and Section 4.3 we introduce two key techniques, superpixels and a denoising proxy, that help overcome these issues and allow for an effective and tractable approximation to pixel-level typicality. In Section 4.4, we additionally incorporate extracted pixel-level features which can provide more informative statistical context than color alone.

Throughout, we denote $x^\star$ to denote the test image and subscripts to identify specific pixels or superpixels. Similarly, we use x (again with appropriate subscripts) to denote the corresponding random variable under the prior or posterior, with Monte Carlo realizations indexed by superscripts.

4.1 TYPICALITY-BASED ANOMALY SCORE

A natural extension of the existing typicality-based OOD detection metric in Equation (3) from images to pixels is conditional typicality. It is defined as

$$f^{\text{inp}}(x^\star, j) := \left| -\log p(x_j^\star | x_{-j}^\star) - \mathbb{H}(p(x_j | x_{-j}^\star)) \right|, \qquad j \in \mathcal{J}, \tag{4}$$

where the superscript inp stands for inpainting, $x_j^\star$ and $x_{-j}^\star$ are the values of pixel j and all remaining pixels in the test image and all other pixels, respectively, and $p(x_j | x_{-j}^\star)$ is the distribution of prior-consistent values for pixel j when all other pixel values are fixed at their observed values $x_{-j}^\star$. Intuitively, we define a pixel in $x^\star$ to be OOD if it does not belong to the typical set with respect to this *conditional* data distribution $p(x_j | x_{-j}^\star)$, given all other pixel values.

Unfortunately, this direct extension of typicality to conditional distributions in Equation (4) is not a practical score for anomaly detection for the following reasons. **(P1)** Due to the spatial locality of pixel intensities and the smoothness of natural images, the value of a pixel given all others is often strongly constrained by nearby pixels. With conditional distributions that are concentrated on a tiny set determined by the neighboring pixels, the typicality score is unlikely to capture OOD features that span multiple adjacent pixels. **(P2)** In addition, a single pixel value is a sequence of length 1, which is insufficient for the theoretical guarantees of typical sets (Equation (2)) to hold. Indeed, the typicality test of Nalisnick et al. (2019) has been shown to exhibit poor empirical performance on single samples (Chauhan et al., 2022). **(P3)** It is also computationally demanding to compute the anomaly heatmap A by approximating Equation (4), as $f^{\text{inp}}(x^\star, j)$ has to be evaluated for every pixel $j \in \mathcal{J}$. Using N posterior samples for a Monte Carlo estimate $\hat{p}(x_j | x_{-j}^\star)$ of $p(x_j | x_{-j}^\star)$ for each pixel would require a total of $N \times |\mathcal{J}|$ posterior samples, which is computationally prohibitive at practical image resolution H, W. **(P4)** Finally, raw RGB values often do not capture sufficient information about local image contents, including semantic information, which can lead to subpar performance in anomaly detection tasks.

4.2 SUPERPIXEL-LEVEL TYPICALITY



Figure 3: A fine-grained superpixel decomposition of an example image.

Our first key insight is that superpixels (Ren and Malik, 2003; Achanta et al., 2012; Stutz et al., 2018), groups of spatially contiguous pixels that share similar color, can help us address **(P1)** and **(P2)**, by grouping OOD pixels together into small groups that are easier to detect and can be treated as a sequence to estimate typicality. The latter requires Assumption 4.1, which captures the intuition that, within each superpixel, individual pixel values are independent and identically distributed. An example of superpixels is visualized in Figure 3.

Concretely, we define $\mathcal{S}(x^\star) = \{S_k \subset \mathcal{J} \mid k \in \mathcal{K}\}$ to be a fine-grained superpixel partition of the pixels in $\mathcal{J}$, which can be obtained using off-the-shelf superpixel algorithms (Achanta et al., 2012; Van den Bergh et al., 2012; Liu et al., 2011; Li and Chen, 2015). Here k indexes superpixels and S_k is the set of pixels in superpixel k. Let x_{S_k} denote image pixels in superpixel k, and x_{-S_k} denote all other pixels.

Assumption 4.1. [Conditionally i.i.d. pixel values within each superpixel] Given $x^\star$ and $\mathcal{S}(x^\star)$, there exists a single-pixel distribution q_k^{inp} associated with each superpixel $k \in \mathcal{K}$ such that

$$x_j \mid x^\star_{-S_k} \overset{iid}{\sim} q_k^{\text{inp}} \quad \forall j \in S_k.$$

Under Assumption 4.1, we define the *superpixel-level typicality score* g^{inp} as

$$g^{\text{inp}}(x^\star, k) = \frac{1}{|S_k|}\left|-\log p(x^\star_{S_k}|x^\star_{-S_k}) - \mathbb{H}(p(x_{S_k}|x^\star_{-S_k}))\right| = \left|-\frac{1}{|S_k|}\sum_{j\in S_k}\log q_k^{\text{inp}}(x_j^\star) - \mathbb{H}(q_k^{\text{inp}})\right|. \tag{5}$$

where $p(x_{S_k}|x^\star_{-S_k})$ is the conditional distribution of the superpixel given all other pixel values, and the last equality holds due to the conditional independence assumption (Assumption 4.1) on the pixels in S_k. We then compute the pixel-level anomaly heatmap A by estimating the typicality score in Equation (5) using Monte Carlo sampling. For each superpixel k, this involves drawing M samples $x^{(1)}, \ldots, x^{(M)}$ from the inpainting distribution $p_k^{\text{inp}} := p(x \mid x^\star_{-S_k})$, extracting pixels $\{x_j^{(m)}\}_{j\in S_k, m\in[M]}$, fitting a histogram $\hat{q}_k^{\text{inp}}$ to estimate q_k^{inp}, and computing the estimate $\hat{g}^{\text{inp}}(x^\star, k)$. Finally, for pixel-level scores it is natural to duplicate the same score over all pixels $j \in S_k$, i.e. $\hat{f}^{\text{inp}}(x^\star, j) = \hat{g}^{\text{inp}}(x^\star, k)$ for all $j \in S_k$, and apply Gaussian smoothing to smooth out the hard boundaries between superpixels.

Estimating the superpixel-level typicality score g^{inp} rather than the pixel-level typicality score f^{inp} helps address these problems. **(P1)** By inpainting a group of similar pixels instead of a single pixel, g^{inp} can better identify local OOD features that span multiple pixels than f^{inp}. **(P2)** Under Assumption 4.1, we are given a sequence of $|S_k|$ independent observations $\{x_j^\star\}_{j\in S_k}$ for q_k^{inp}, which facilitates computing a typicality score for each superpixel. **(P3)** In terms of computational cost, the number of anomaly scores to evaluate is reduced from the number of pixels $|\mathcal{J}|$ to the number of superpixels $|\mathcal{K}|$. We can also use a much smaller number of samples $M \ll N$ to compute the histogram $\hat{q}_k$ for q_k^{inp}, since each posterior sample from p_k^{inp} translates to $|S_k|$ samples from q_k^{inp}.

4.3 DENOISING POSTERIOR

Although superpixel-level typicality reduces the required number of posterior samples from $|\mathcal{J}|N$ to $|\mathcal{K}|M$, the cost remains prohitive for fine superpixel partitions. For instance, Figure 3 shows a 256×256 image partitioned into 2,600 superpixels (approximately 25 pixels per superpixel on average), which would require a total of at least 2600 posterior samples using a diffusion model. This cost could be reduced by partitioning the image into fewer, larger superpixels, but these may struggle to localize small anomalies and fail to satisfy Assumption 4.1.

We overcome this problem **(P3)** by estimating q_k^{inp} for all $k \in \mathcal{K}$ simultaneously using a small number of posterior samples. Instead of using different samples from each of the *inpainting conditional*

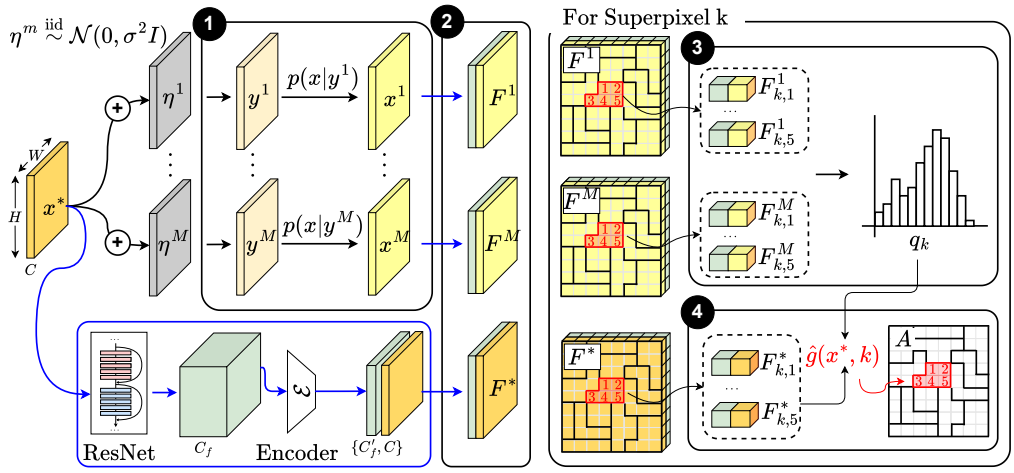


Figure 4: **LACToS Overview. (1) Posterior Sampling.** We sample noise independently from an isotropic Gaussian distribution and add it to the test image $x^\star$ to create noised versions of $x^\star$ denoted by y^m. Then, using diffusion-based posterior sampling methods, we sample reconstructions $x^m \sim p(x|y^m)$. **(2) Feature Extraction.** For $x^\star$ and each sample x^m, we pass the image through a pretrained ResNet model, extract activations, and compress them using an encoder to create a feature vector of dimension $C'_f + C$ for each pixel, where C is the original channel dimension. **(3) Histogram Fitting**. For a given superpixel k, we gather the features of the reconstructed pixels in the superpixel, and construct $\hat{q}_k^{\text{den}}$, a $C'_f + C$-dimensional histogram. **(4) Typicality Computation.** Finally, using the fitted histogram, we compute the typicality of the pixel features in $x^\star$ and report the anomaly heatmap A.

distributions $\{p_k^{\text{inp}}\}_{k\in\mathcal{K}}$, we use the same M samples from the *denoising conditional distribution*, defined as

$$p^{\text{den}}(x) := \mathbb{E}_{\eta\sim\mathcal{N}(0,\sigma^2 I)}\big[p(x \mid x^\star + \eta)\big]. \tag{6}$$

For each S_k, we pool the pixel values across these samples to form $\{x_j^m\}_{j\in S_k, m\in[M]}$, which we use to fit a histogram for q_k^{inp}, and compute the anomaly score.

Intuitively, p^{den} serves as a shared relaxation of the superpixel-specific inpainting posteriors $\{p_k^{\text{inp}}\}_{k\in\mathcal{K}}$. In each p_k^{inp}, the complement $x^\star_{-S_k}$ is fully observed while $x^\star_{S_k}$ is left unobserved. For the denoising posterior p^{den}, there is a tension between information on $x^\star_{-S_k}$ and lack of information on $x^\star_{S_k}$, controlled by the noise level σ. With increasing σ, we lose more information about $x^\star_{S_k}$, which is desirable, but we also lose information about the context $x^\star_{-S_k}$. The following theorem informally bounds the error introduced by replacing the inpainting posterior with the denoising posterior. Appendix section D states and proves the formal result and gives bounds without Assumption 4.1.

Theorem 4.1. (Informal). Let $g^{\text{inp}}(x^\star, k)$ and $g_\sigma^{\text{den}}(x^\star, k)$ be the anomaly scores obtained using the ideal inpainting distribution and the proxy denoising distribution with a noise distribution of $\mathcal{N}(0, \sigma^2 I)$, respectively. Under appropriate assumptions, the error can be bounded as

$$\left|g^{\text{inp}}(x^\star, k) - g_\sigma^{\text{den}}(x^\star, k)\right| \le \alpha_k \sigma + \beta_k \sigma^{-\gamma_k}$$

for some positive constants α_k, β_k, and γ_k.

4.4 Feature Augmentation

One final problem remains **(P4)**: RGB values alone may not distinguish local ID and OOD features. To address this issue, **(P4)** we follow the convention of previous works (Mousakhan et al., 2024; Roth et al., 2022; Defard et al., 2021) and augment RGB values with additional features. Specifically, we extract the activations of the last 3 layers of a pretrained ResNet (He et al., 2016) ϕ, as later layers of convolutional neural networks are known to encode higher-level visual structure and have been widely used as transferable representations for downstream vision tasks (Razavian et al., 2014; Zeiler and Fergus, 2014). Since the dimension of these activations, denoted as C_f, is large, which

inflates histogram error in $\hat{q}_k$ under limited samples, we perform dimensionality reduction using an autoencoder (trained on each ID dataset) to reach a compressed feature vector of length $C'_f (\ll C_f)$; in our experiments we use $C'_f = 3$.

Algorithm Summary. Our complete algorithm, illustrated in Figure 4, can be summarized as follows. First, we generate M samples from the denoising conditional distribution p^{den} through posterior sampling using a diffusion model. Second, each of the samples is passed through ϕ to extract features, which are then compressed using the trained encoder $\mathcal{E}$. Third, the pixel values and the compressed features are concatenated, resulting in

$$u_j^m := (x_j^m, \mathcal{E}(\phi(x^m)_j)) \in \mathbb{R}^{C+C'_f}$$

for each pixel j. Fourth, the augmented vectors u_j^m that belong to each superpixel S_k are collected to form $\{u_j^m\}_{j \in S_k, m \in [M]}$ and used to construct a histogram $\hat{q}_k$, an estimation of the $C+C'_f$-dimensional distribution q_k^{inp}. The fourth step is repeated over all superpixels, after which we can compute the anomaly map by computing the typicality of $x^\star$ for each superpixel (Equation (5)), and finally apply spatial Gaussian smoothing to refine pixel-level anomaly scores across superpixel boundaries.

5 Evaluation

Datasets. We evaluate LACToS on three local anomaly detection datasets spanning medical imaging, human faces, and X-ray security screening (Table 4). In each case, the diffusion model in LACToS is trained only on inlier distribution (ID) images, and all methods are evaluated on a held-out test set. Each test set includes an equal number of ID images and images containing local OOD features.

CHAOS (Kavur et al., 2019) provides abdominal CT scans of healthy patients. We train on 3,070 inlier slices and construct the test set containing 100 healthy test slices and 100 locally anomalous slices acquired by adding synthetic tumor to each of the healthy test slices. Following common practice in CT image analysis (Wang et al., 2021; Thummerer et al., 2023; Su et al., 2019), we apply the provided body mask at evaluation to exclude background pixels. **FFHQ** (Karras et al., 2019) is a dataset of high-quality face images, which we treat as the inlier distribution. We train on a random subset of 5,000 images, and manually curate an OOD test set of 50 photographs of movie and television characters with prosthetics or atypical facial features (e.g., scars, metal implants), verified not to overlap with FFHQ. The ID test set contains 50 randomly selected FFHQ test set images. **DVXray** (Ma et al., 2024) is a dataset of X-ray luggage scans, with an ID set of weapon-free bags and an OOD set containing weapons. We train on 5,000 inlier images and restrict the OOD test set to the *scissors* category, where targets are small and spatially confined. From randomly sampled scissor images, we used SAM 3 (Carion et al., 2026) to convert the dataset's bounding-box annotations into segmentation masks, manually discarding images with visibly inaccurate masks for a final OOD test set of 72 images. We construct an ID test set that also contains 72 images, none of which contains a weapon.

Baselines & Models. We compare LACToS against six strong baselines developed for industrial defect detection, spanning representation- and reconstruction-based methods. **PatchCore** (Roth et al., 2022) builds a coreset-subsampled memory bank of ID patch features from a frozen ImageNet-pretrained backbone and scores test patches by nearest-neighbor distance. **SimpleNet** (Liu et al., 2023) learns a feature adapter and a discriminator that separates adapted ID features from synthetic anomalies generated by injecting Gaussian noise in feature space. **SuperSimpleNet** (Rolih et al., 2024) extends SimpleNet with architectural and training improvements for fast, reliable surface defect detection. **EfficientAD** (Batzner et al., 2024) combines a lightweight student–teacher model with an autoencoder to detect local structural anomalies and violations of global logical constraints through feature discrepancies. **GLASS** (Chen et al., 2024) trains a discriminator to separate ID features from synthetic anomalies generated using gradient ascent with truncated projection in feature space and local texture overlays in image space. **DDAD** (Mousakhan et al., 2024) conditions a denoising diffusion model on the test image to produce an ID reconstruction and scores anomalies by feature-space discrepancy.

LACToS requires an unconditional diffusion model and a posterior sampler. For all benchmarks, we use the guided-diffusion U-Net of Dhariwal and Nichol (2021) with diffusion posterior sampling

Table 1: Pooled quantitative performance in local anomaly detection. AUROC and AP are computed over all pixels in each dataset jointly. AUPRO is computed over the dataset using per-region overlap for $\mathrm{FPR} \leq 0.3$. Higher is better for all metrics. The best result is shown in **bold** and the second-best is underlined.

Dataset	Metric	PatchCore	SimpleNet	SuperSimpleNet	EfficientAD	GLASS	DDAD	LACToS (Ours)
CHAOS	AUROC	0.9297	0.7464	0.3786	0.6664	0.8318	**0.9707**	0.9620
	AP	0.0590	0.0193	0.0051	0.0105	0.0410	0.3159	**0.4318**
	AUPRO	0.6811	0.4191	0.0332	0.2513	0.5530	0.8676	**0.8764**
FFHQ	AUROC	0.8491	0.7162	0.7647	0.6200	0.6707	0.7469	**0.8515**
	AP	0.0798	0.0537	0.0772	0.0405	0.0345	0.0563	**0.1177**
	AUPRO	0.4094	0.3174	0.4047	0.2365	0.2440	0.3083	**0.5320**
DVXray	AUROC	**0.9640**	0.9234	0.8688	0.9184	0.8944	0.8567	0.9607
	AP	0.0545	0.0254	0.0165	0.0235	0.0155	0.0118	**0.0568**
	AUPRO	0.8141	0.6577	0.5351	0.6592	0.6062	0.5183	**0.8179**

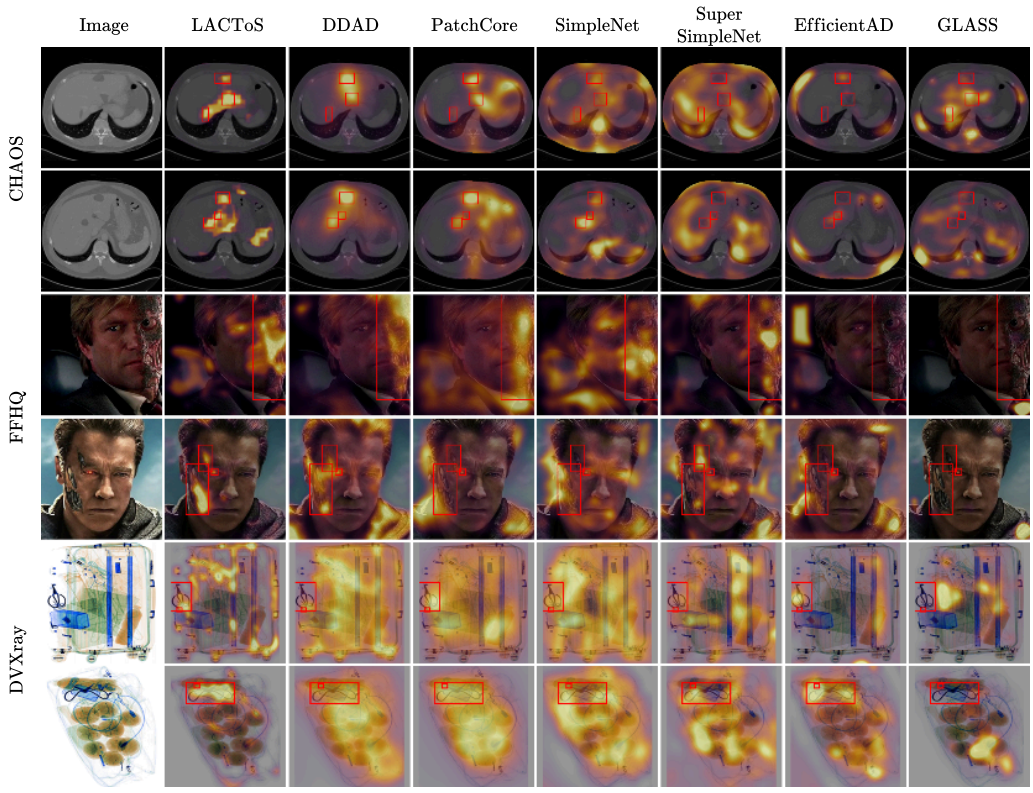


Figure 5: Anomaly heatmaps generated by each method on example images from each test dataset. Rows 1, 2 (**CHAOS**): detect liver tumors in CT scans. Rows 3, 4 (**FFHQ**): detect anomalous facial features. Rows 5, 6 (**DVXray**): detect weapons (scissors) in X-ray baggage scans.

(DPS) (Chung et al., 2022) at inference time. Full experimental details are provided in Section E and in our code, which will be released upon publication.

Results. Each method produces a heatmap of anomaly scores per test image. Following standard practice, we pool the anomaly score for each pixel across all test images and evaluate three complementary pixel-level detection metrics. **AUROC** captures the relative ordering of heatmap values against the ground-truth anomaly mask, but can reward low-contrast heatmaps that flag many ID pixels as nearly as anomalous as the true OOD pixels. **AP** is the area under the precision-recall curve obtained by sweeping a threshold across the pooled heatmaps. **AUPRO** is the normalized area under the per-region overlap curve up to a false positive rate of 0.3, where per-region overlap is the fraction

of correctly detected pixels within each connected ground-truth anomalous region, averaged equally across regions. When different anomalies have different sizes, AUPRO weights all anomalies equally rather than over-weighting large anomalies.

Table 1 reports these metrics, which indicate that LACToS is competitive with baselines across all tasks, though the best method differs across metrics. In addition, we can observe from the example heatmap visualizations in Figure 5 that LACToS tends to be more precise: it correctly marks OOD pixels but rarely flags ID pixels, at the cost of occasionally missing part or all of the OOD region. The baselines instead produce blurrier heatmaps that recover the OOD region but also flag many ID pixels.

Ablation Studies. To verify the robustness of LACToS, we perform ablation studies over various hyper-parameters, which we describe in detail in Section G. These include the number of posterior samples M and the average size of each superpixel. We observe that the performance of LACToS is mostly stable across different settings, demonstrating its robustness to hyper-parameter choices. In addition, we perform an ablation on the use of autoencoder-based features, which shows that the features can provide substantial improvement compared to typicality based solely on RGB color, though this improvement is dataset-dependent.

6 Discussion

We present LACToS, a framework for general-purpose local anomaly detection. We leverage a pretrained diffusion model to represent the inlier distribution, and combine it with a statistical approach that scores image regions by their typicality (or rather, atypicality) under the conditional distribution induced by their surrounding context. By combining superpixel partitioning with a shared denoising posterior as a tractable proxy for per-superpixel inpainting, LACToS evaluates a theoretically grounded typicality-based anomaly score at a cost of M posterior sampling runs per image, regardless of the number of superpixels. Empirically, LACToS shows strong performance across diverse local OOD detection tasks spanning medical imaging to security scanning.

We note that since LACToS is more robust to inlier multimodality, it is also less likely to detect features from minority inlier modes as anomalous, especially when using a larger M. We demonstrate this behavior through a toy experiment, and refer interested readers to Section C. It is possible that some features annotated as anomalies in the benchmarks we evaluate may be similar to features present with low frequency in the inlier (training) datasets, causing LACToS to consider them as inliers and thus generate false negative predictions according to the dataset anomaly labels. We therefore advise practitioners that, in order to maximize the accuracy of LACToS, it is important to make sure that the training set does not contain the anomalies one desires to detect.

Limitations and future work. Since LACToS targets general-purpose local anomaly detection, specialized approaches can outperform LACToS for specific highly structured datasets, such as those common in industrial defect detection. Please refer to table 7 for evaluation on **MVTec AD**. LACToS inherits the computational cost of diffusion posterior sampling, which is significantly higher per-image than feedforward representation-based methods such as PatchCore. The method also relies on Assumption 4.1, which may be violated when superpixels are large or when the anomalous region spans multiple superpixels with heterogeneous pixel statistics. LACToS also requires several hyperparameters, such as the noise variance in the denoising posterior sampling, the quantization granularity and smoothing bandwidth in the histogram estimation, the number of posterior samples, the granularity of superpixels, and the spatial smoothing variance of the final heatmap. While we provide ablation studies showing LACToS is fairly robust to these hyperparameters, future work may find mechanisms to remove or automatically set optimal values for these hyperparameters across tasks, potentially improving OOD detection performance. Future work may also replace our histogram estimation with a learned or parametric density model, or extend our LACToS framework to video and 3D data types where spatio-temporal or volumetric superpixels could provide longer sequences for improved typicality estimation.

AI USE STATEMENT

We used generative AI tools to assist with manuscript editing and formatting, literature search and summarization, and feedback on technical exposition, experimental methodology, and mathematical arguments. We did not use generative AI to generate experimental data or experimental results. All AI-assisted suggestions were manually reviewed by the authors and, where applicable, checked against primary sources, mathematical derivations, implementation code, and experimental outputs. The authors made all final decisions regarding the method, experiments, analysis, and claims, and take responsibility for the final content of this work, including text, claims, and artifacts produced with the aid of generative AI.

ETHICS STATEMENT

This work studies anomaly detection using existing datasets and does not involve new data collection or interaction with human subjects. Our experiments include medical CT images, human face images, and X-ray security-screening images. The face-image experiments are used only to evaluate local anomaly-detection methodology; the labels are not intended to characterize people or physical traits as intrinsically abnormal, and we do not perform identity recognition or infer sensitive personal attributes. Likewise, the medical and security-screening experiments are research benchmarks and are not intended for direct clinical or operational deployment without application-specific validation. We use the datasets only for the research purposes described in this paper and follow their respective terms of use.

REPRODUCIBILITY STATEMENT

We provide the complete method specification in Section 4. Dataset construction, preprocessing, diffusion-model and reconstruction settings, dataset-specific hyperparameters, baseline configurations, and feature-extraction details are provided in Section E. Ablation studies examining the principal hyperparameters are reported in Section G, and complete theoretical assumptions and proofs are given in Section D. We will release the implementation code upon publication.

REFERENCES

Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, and Sabine Süsstrunk. Slic superpixels compared to state-of-the-art superpixel methods. *IEEE transactions on pattern analysis and machine intelligence*, 34(11):2274–2282, 2012.

Brian DO Anderson. Reverse-time diffusion equation models. *Stochastic Processes and their Applications*, 12 (3):313–326, 1982.

Brett Barkley, Preston Culbertson, and David Fridovich-Keil. SCOPED: Score-curvature out-of-distribution proximity evaluator for diffusion. *International Conference on Learning Representations*, 2026.

Kilian Batzner, Lars Heckler, and Rebecca König. Efficientad: Accurate visual anomaly detection at millisecond-level latencies. In *2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 127–137. IEEE, 2024.

Paul Bergmann, Michael Fauser, David Sattlegger, and Carsten Steger. Mvtec ad–a comprehensive real-world dataset for unsupervised anomaly detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9592–9600, 2019.

Nicolas Carion, Laura Gustafson, Yuan-Ting Hu, Shoubhik Debnath, Ronghang Hu, Didac Suris, Chaitanya Ryali, Kalyan Vasudev Alwala, Haitham Khedr, Andrew Huang, Jie Lei, Tengyu Ma, Baishan Guo, Arpit Kalla, Markus Marks, Joseph Greer, Meng Wang, Peize Sun, Roman Rädle, Triantafyllos Afouras, Effrosyni Mavroudi, Katherine Xu, Tsung-Han Wu, Yu Zhou, Liliane Momeni, Rishi Hazra, Shuangrui Ding, Sagar Vaze, Francois Porcher, Feng Li, Siyuan Li, Aishwarya Kamath, Ho Kei Cheng, Piotr Dollár, Nikhila Ravi, Kate Saenko, Pengchuan Zhang, and Christoph Feichtenhofer. Sam 3: Segment anything with concepts, 2026. URL https://arxiv.org/abs/2511.16719.

Kushal Chauhan, Pradeep Shenoy, Manish Gupta, Devarajan Sridharan, et al. Robust outlier detection by de-biasing vae likelihoods. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9881–9890, 2022.

Qiyu Chen, Huiyuan Luo, Chengkan Lv, and Zhengtao Zhang. A unified anomaly synthesis strategy with gradient ascent for industrial anomaly detection and localization. In *European Conference on Computer Vision*, pages 37–54. Springer, 2024.

Hyungjin Chung, Jeongsol Kim, Michael T Mccann, Marc L Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. *arXiv preprint arXiv:2209.14687*, 2022.

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

Thomas M Cover, Joy A Thomas, and John Kieffer. *Elements of information theory*, volume 2. wiley New York, 1991.

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.

Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. *Advances in neural information processing systems*, 34:8780–8794, 2021.

Mark S Graham, Walter HL Pinaya, Petru-Daniel Tudosiu, Parashkev Nachev, Sebastien Ourselin, and Jorge Cardoso. Denoising diffusion models for out-of-distribution detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2948–2957, 2023.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

Aapo Hyvärinen and Peter Dayan. Estimation of non-normalized statistical models by score matching. *Journal of Machine Learning Research*, 6(4), 2005.

Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4401–4410, 2019.

Ali Emre Kavur, M. Alper Selver, Oğuz Dicle, Mustafa Barış, and N. Sinem Gezer. CHAOS - Combined (CT-MR) Healthy Abdominal Organ Segmentation Challenge Data, April 2019. URL `https://doi.org/10.5281/zenodo.3362844`.

Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising diffusion restoration models. *Advances in neural information processing systems*, 35:23593–23606, 2022.

Zhengqin Li and Jiansheng Chen. Superpixel segmentation using linear spectral clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1356–1363, 2015.

Yaron Lipman, Ricky TQ Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. *arXiv preprint arXiv:2210.02747*, 2022.

Ming-Yu Liu, Oncel Tuzel, Srikumar Ramalingam, and Rama Chellappa. Entropy rate superpixel segmentation. In *CVPR 2011*, pages 2097–2104. IEEE, 2011.

Zhikang Liu, Yiming Zhou, Yuansheng Xu, and Zilei Wang. Simplenet: A simple network for image anomaly detection and localization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 20402–20411, 2023.

Bowen Ma, Tong Jia, Mingyuan Li, Songsheng Wu, Hao Wang, and Dongyue Chen. Toward dual-view x-ray baggage inspection: A large-scale benchmark and adaptive hierarchical cross refinement for prohibited item discovery. *IEEE Transactions on Information Forensics and Security*, 19:3866–3878, 2024. doi: 10.1109/TIFS.2024.3372797.

Fernand Meyer and Serge Beucher. Morphological segmentation. *Journal of visual communication and image representation*, 1(1):21–46, 1990.

Arian Mousakhan, Thomas Brox, and Jawad Tayyub. Anomaly detection with conditioned denoising diffusion models. In *DAGM German Conference on Pattern Recognition*, pages 181–195. Springer, 2024.

Eric Nalisnick, Akihiro Matsukawa, Yee Whye Teh, and Balaji Lakshminarayanan. Detecting out-of-distribution inputs to deep generative models using typicality. *arXiv preprint arXiv:1906.02994*, 2019.

Genki Osada, Tsubasa Takahashi, Budrul Ahsan, and Takashi Nishide. Out-of-distribution detection with reconstruction error and typicality-based penalty. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 5551–5563, 2023.

Jonathan Pirnay and Keng Chai. Inpainting transformer for anomaly detection. In *International conference on image analysis and processing*, pages 394–406. Springer, 2022.

Ali Sharif Razavian, Hossein Azizpour, Josephine Sullivan, and Stefan Carlsson. CNN features off-the-shelf: An astounding baseline for recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 806–813, 2014.

Ren and Malik. Learning a classification model for segmentation. In *Proceedings ninth IEEE international conference on computer vision*, pages 10–17. IEEE, 2003.

Blaž Rolih, Matic Fučka, and Danijel Skočaj. Supersimplenet: Unifying unsupervised and supervised learning for fast and reliable surface defect detection. In *International Conference on Pattern Recognition*, pages 47–65. Springer, 2024.

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.

Cheng Shen and Yuewei Liu. Conformal segmentation in industrial surface defect detection with statistical guarantees. *Mathematics*, 13(15):2430, 2025.

Shirin Shoushtari, Yi Wang, Xiao Shi, M Salman Asif, and Ulugbek S Kamilov. Eigenscore: Ood detection using covariance in diffusion models. *arXiv preprint arXiv:2510.07206*, 2025.

Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*, 2020.

Yang Song, Liyue Shen, Lei Xing, and Stefano Ermon. Solving inverse problems in medical imaging with score-based generative models. *arXiv preprint arXiv:2111.08005*, 2021.

David Stutz, Alexander Hermans, and Bastian Leibe. Superpixels: An evaluation of the state-of-the-art. *Computer Vision and Image Understanding*, 166:1–27, 2018.

Kuan-Hao Su, Harry T Friel, Jung-Wen Kuo, Rose Al Helo, Atallah Baydoun, Christian Stehning, Adina N Crisan, Melanie S Traughber, Ajit Devaraj, David W Jordan, et al. Ute-mdixon-based thorax synthetic ct generation. *Medical physics*, 46(8):3520–3531, 2019.

Adrian Thummerer, Erik Van der Bijl, Arthur Galapon Jr, Joost JC Verhoeff, Johannes A Langendijk, Stefan Both, Cornelis (Nico) AT van den Berg, and Matteo Maspero. Synthrad2023 grand challenge dataset: Generating synthetic ct for radiotherapy. *Medical physics*, 50(7):4664–4674, 2023.

Michael Van den Bergh, Xavier Boix, Gemma Roig, Benjamin De Capitani, and Luc Van Gool. Seeds: Superpixels extracted via energy-driven sampling. In *European conference on computer vision*, pages 13–26. Springer, 2012.

Tonghe Wang, Yang Lei, Yabo Fu, Jacob F Wynne, Walter J Curran, Tian Liu, and Xiaofeng Yang. A review on medical imaging synthesis using deep learning and its clinical applications. *Journal of applied clinical medical physics*, 22(1):11–36, 2021.

Yinhuai Wang, Jiwen Yu, and Jian Zhang. Zero-shot image restoration using denoising diffusion null-space model. *arXiv preprint arXiv:2212.00490*, 2022.

Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.

Di Wu, Shicai Fan, Xue Zhou, Li Yu, Yuzhong Deng, Jianxiao Zou, and Baihong Lin. Unsupervised anomaly detection via masked diffusion posterior sampling. *arXiv preprint arXiv:2404.17900*, 2024.

Julian Wyatt, Adam Leach, Sebastian M Schmon, and Chris G Willcocks. Anoddpm: Anomaly detection with denoising diffusion probabilistic models using simplex noise. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 650–656, 2022.

Vitjan Zavrtanik, Matej Kristan, and Danijel Skočaj. Reconstruction by inpainting for visual anomaly detection. *Pattern Recognition*, 112:107706, 2021.

Matthew D Zeiler and Rob Fergus. Visualizing and understanding convolutional networks. In *European conference on computer vision*, pages 818–833. Springer, 2014.

Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–595, 2018.

Yang Zou, Jongheon Jeong, Latha Pemula, Dongqing Zhang, and Onkar Dabeer. Spot-the-difference self-supervised pre-training for anomaly detection and segmentation. In *European conference on computer vision*, pages 392–408. Springer, 2022.

A BACKGROUND

A.1 DIFFUSION MODELS

Diffusion models are generative models that have gained enormous popularity for their high sample quality Ho et al. (2020); Dhariwal and Nichol (2021). To sample from a desired data distribution p, these models first define a noising process (also called the forward process) as a stochastic differential equation (SDE) (Song et al., 2020) of the form

$$dx = f(x,t)dt + g(x,t)dw, \qquad t = 0 \to 1, \tag{7}$$

where w is a standard Wiener process (Brownian motion) and f, g are the drift and diffusion coefficients, respectively, which are determined by a chosen noise schedule. If we denote the marginal distribution of x at time t as p_t, this process transforms clean data samples $x \sim p_0 := p$ across time until the distribution becomes pure noise at $t = 1$, i.e. $p_1 = \mathcal{N}(0, I)$. Once the forward process is defined, there exists, under regularity conditions, a corresponding backward SDE with drift $\tilde{f}$ and diffusion $\tilde{g}$ (Anderson, 1982; Song et al., 2020):

$$dx = \tilde{f}(x,t)dt + \tilde{g}(x,t)d\bar{w}, \qquad t = 1 \to 0, \tag{8}$$

where $\bar{w}$ is also a Wiener process but with t indexed backwards. This backward SDE shares the same marginal distributions p_t as the marginal distributions induced by the forward process for all $t \in [0, 1]$. Since the drift coefficient $\tilde{f}$ of the backward SDE has the form

$$\tilde{f}(x,t) = f(x,t) - g(x,t)^2 \nabla_x \log p_t(x), \tag{9}$$

a diffusion model is optimized on either a score-matching loss to predict $\nabla_x \log p_t(x)$ Hyvärinen and Dayan (2005); Song et al. (2020) or on a flow-matching loss to predict $\tilde{f}(x,t)$ directly Lipman et al. (2022), and used to iteratively sample from p_t through discretization of the backward SDE.

A.2 POSTERIOR SAMPLING USING DIFFUSION MODELS

To condition generation on inference-time measurements using a pretrained diffusion model, prior works modify the reverse sampling process to approximately sample from the posterior distribution capturing simultaneous consistency with the prior encoded by the diffusion model and the likelihood induced by the measurement model Song et al. (2021); Kawar et al. (2022); Chung et al. (2022); Wang et al. (2022).

These methods decompose the conditional score using Bayes' rule as

$$\nabla_x \log p_t(x|y) = \nabla_x \log p_t(x) + \nabla_x \log p_t(y|x), \tag{10}$$

where $\nabla_x \log p_t(x)$ is the unconditional score, which is provided by the pretrained diffusion model, and $\nabla_x \log p_t(y|x)$ is the likelihood score, which is typically approximated to promote consistency with y while maintaining computational tractability. These works have shown strong empirical performance on various imaging inverse problems such as inpainting, denoising, deblurring, super-resolution, and tomography; in all of these tasks the condition is given in the form of measurements $y = \mathcal{A}(x) + \eta$ for linear or non-linear observation model $\mathcal{A}$ and noise η.

B TOY EXPERIMENT 1

Here we provide further details and results for the toy experiment conducted in Section 3. For each level of inlier complexity, we train a dedicated diffusion model on a set of inlier images that always contain a circle at the upper-right corner. The other 3 corners of the inlier images contain a shape randomly selected from the set of candidate shapes, with probabilities defined in Table 2. As the level increases, more shapes are equally probable at these three corners, thereby isolating the effect of increasing local multimodality in the inlier distribution. Some example images in each training set are shown in Figure 6.

Table 2: **Toy Experiment 1 Inlier Distribution Specification.**

Level	Set of Candidate Shapes	Probability
Level 1	{Circle, Square}	$[0.5, 0.5]$
Level 2	{Circle, Square, Triangle}	$[0.\overline{3}, 0.\overline{3}, 0.\overline{3}]$
Level 3	{Circle, Square, Triangle, Ellipse}	$[0.25, 0.25, 0.25, 0.25]$
Level 4	{Circle, Square, Triangle, Ellipse, Star}	$[0.2, 0.2, 0.2, 0.2, 0.2]$

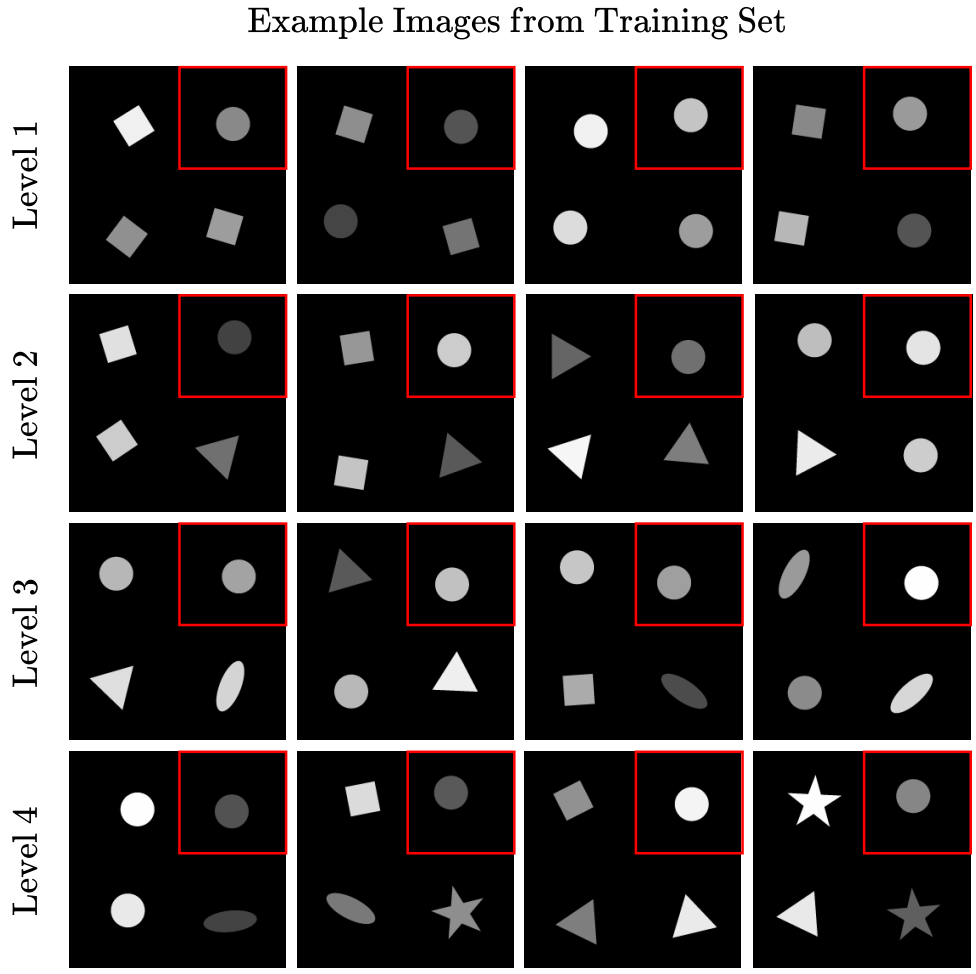


Figure 6: **Example Images from Training Sets.** The shape orientations are random, with positions sampled with jitter around fixed anchors in each quadrant. The upper-right corner, marked with a red box, always contains a circle. With increasing level, each of the other three corners has increasing modality, as more shapes are equally probable under the inlier distribution.

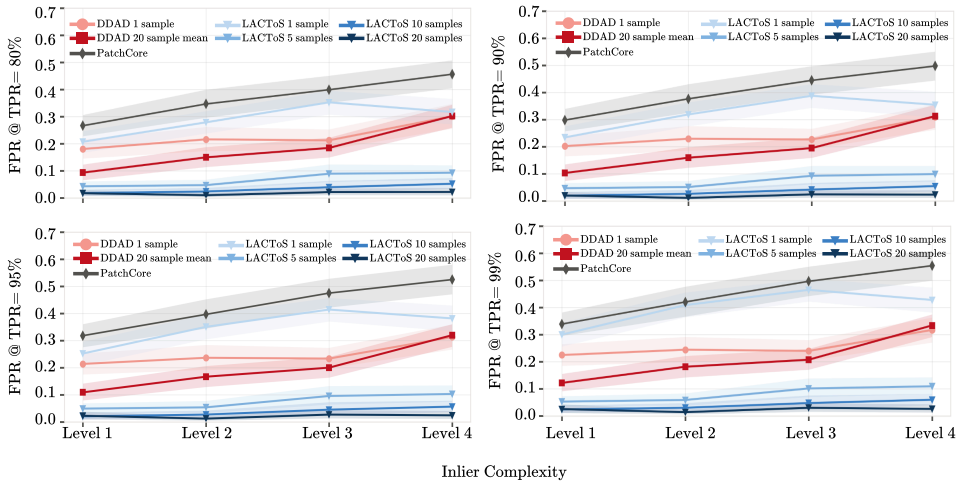


Figure 7: **LACToS Toy Experiment 1: Robustness Under Multimodality.** Average per-image False Positive Rates at True Positive Rate $\in \{80\%, 90\%, 95\%, 99\%\}$, evaluated on images that contain squares on the upper-right corner.

Using these diffusion models, we evaluate how the multimodality of the inlier distribution affects the performance of different local anomaly detection methods. We generate 50 test images that contain a square on the upper-right corner, and place other inlier shapes at the other corners. We label the pixels in the square as positive (anomalous), and all others as negative (non-anomalous). For DDAD Mousakhan et al. (2024), each test image is noised then denoised, which is used in conjunction with ResNet embeddings of the image to construct the anomaly heatmap. We also evaluate a variant of DDAD that uses more samples by repeating the noising-then-denoising process 20 times with different random seeds, averaging the denoised samples, and using the resulting average to construct the anomaly heatmap. For PatchCore Roth et al. (2022), a feed-forward model is trained for each inlier distribution, and a heatmap is generated by evaluating the test image with the feed-forward model. Finally, for LACToS, we follow the process described in Section 4, while varying the sample count from 1 to 20. Since the images are grayscale, we use the same set of parameters as for CHAOS, the dataset containing abdominal CT scans which are also in grayscale, for all methods.

To effectively isolate how likely different methods are to erroneously predict inlier shapes as anomalous, we report per-image averaged false- positive rate (FPR) at fixed values of per-image true-positive rate (TPR) $\in \{80\%, 90\%, 95\%, 99\%\}$. Figure 7 shows that the FPR of DDAD (Mousakhan et al., 2024) and PatchCore (Roth et al., 2022) increases with inlier multimodality, whereas the FPR of LACToS decreases rapidly with sample size, and stays negligible compared to the baseline methods when 20 samples are used, even when the inlier distribution is increasingly multimodal. This demonstrates that LACToS is more robust to inlier multimodality.

C TOY EXPERIMENT 2

We conduct a second toy experiment to better understand how LACToS and baseline methods treat ID but under-represented features, compared to dominant ID features and strictly OOD features. We hypothesize that, due to the distribution-aware nature of statistical typicality, LACToS considers these minority ID features to be closer to majority ID features, whereas baseline methods that use a first-order statistic may consider them to be closer to OOD. To demonstrate this concretely, we train 5 diffusion models on training distributions whose specifications are provided in Table 3.

Table 3: **Toy Experiment 2 Inlier Distribution Specification.**

Level	Set of Candidate Shapes	Probability
Level 1	{Circle, Square}	$\{0.5, 0.5\}$
Level 2	{Circle, Square}	$\{0.6, 0.4\}$
Level 3	{Circle, Square}	$\{0.7, 0.3\}$
Level 4	{Circle, Square}	$\{0.8, 0.2\}$
Level 5	{Circle, Square}	$\{0.9, 0.1\}$

Similar to the toy experiment in Section B, the upper-right corner still only admits a circle under the inlier distribution, so a square there is a strict OOD feature. On the other hand, the inlier probability of a square being placed in the other 3 corners is progressively lowered from 0.5 to 0.1. Example images in each training set are shown in Figure 8.

To verify the hypothesis, we measure two quantities. First, we measure the AUROC over pixels that belong to ID squares and circles, where squares are labeled positive and circles are labeled negative. A higher AUROC here means the method treats ID, but potentially under-represented, squares differently from ID circles, and vice versa. Second, we measure the AUROC over pixels that belong to squares, where OOD squares (those at the upper-right corner) are labeled positive and ID squares (those at other corners) are labeled negative. Higher AUROC here means that the method treats the under-represented ID squares differently from truly OOD squares, and vice versa. In Figure 9, we observe that LACToS is less likely to distinguish ID squares from ID circles and more likely to distinguish under-represented ID squares from truly OOD squares compared to DDAD. Both trends become more apparent with increasing number of sampled reconstructions. Together, these results verify our hypothesis, and inform practitioners that LACToS is a more reliable choice when one wishes to detect features that are not included in the training set without falsely flagging uncommon inlier features.

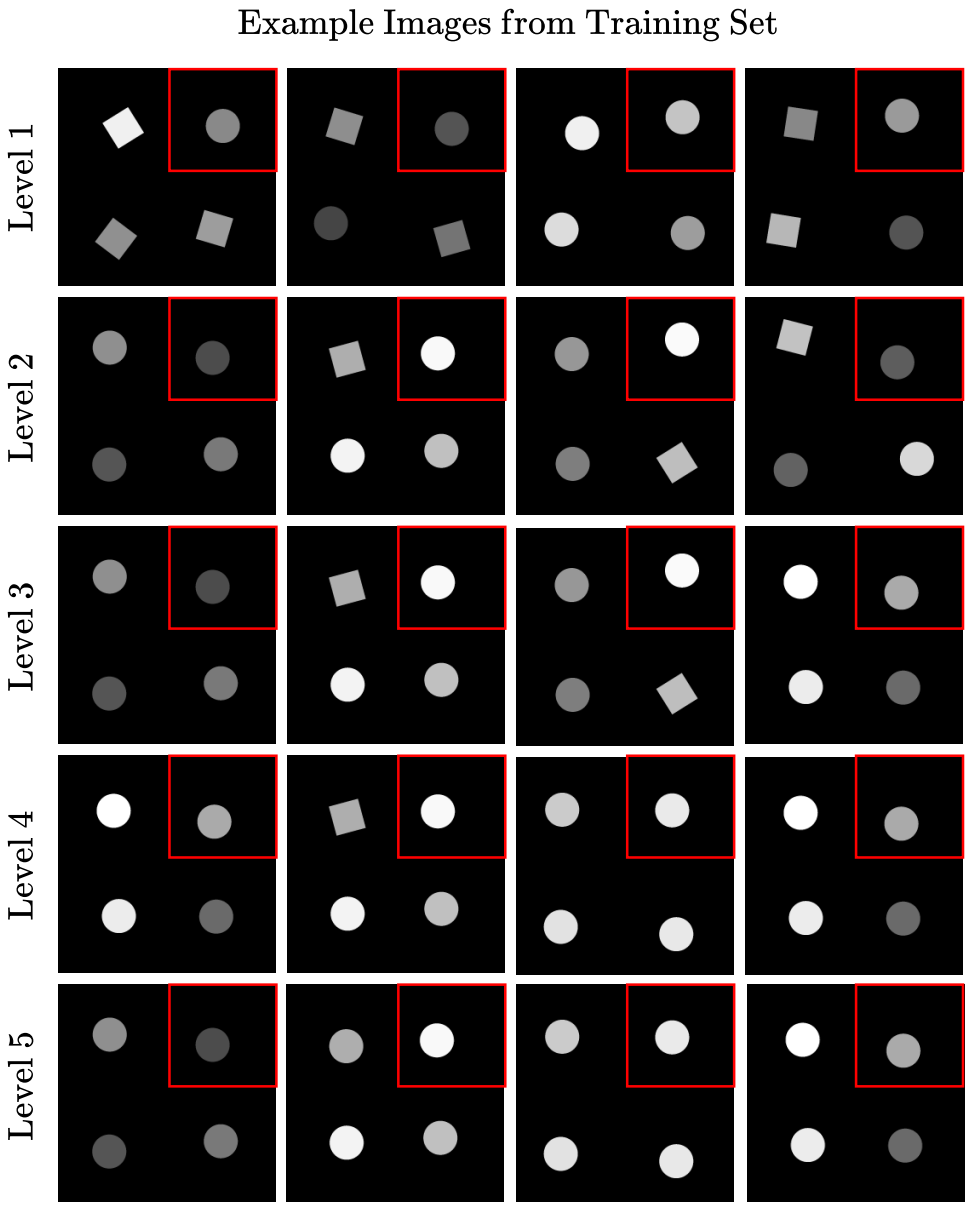


Figure 8: **Example Images from Training Sets.** Positions of the shapes are sampled with jitter around fixed anchors, with random orientations. The upper-right corner, marked with a red box, always contains a circle. With increasing level, the probability of a square at the other three corners decreases, making it more difficult to distinguish squares there from the square in the upper-right corner.

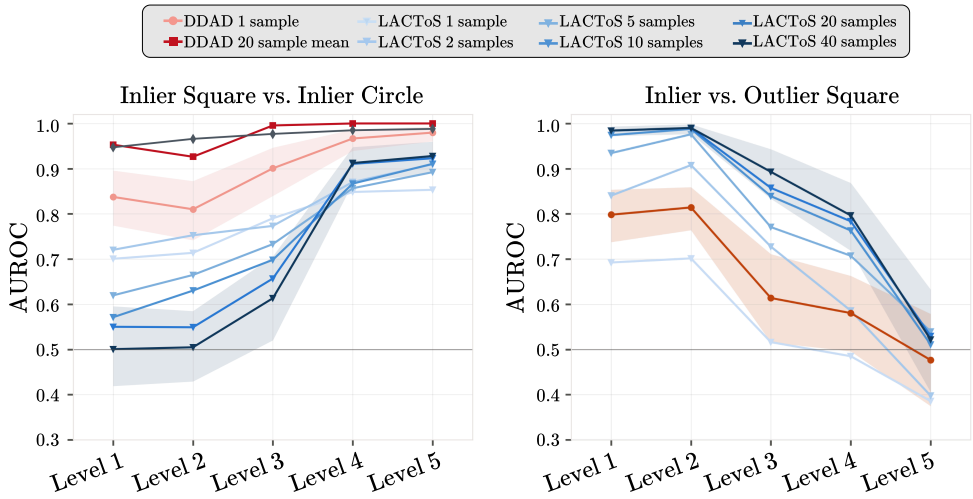


Figure 9: **LACToS Toy Experiment 2.** AUROCs across different inlier square probabilities, over methods and different sample sizes. (Left) AUROC is computed over pixels that belong to inlier squares and circles, with those pixels from squares labeled positive, and negative otherwise. Higher AUROC means the method is more likely to distinguish an inlier square (minority) from inlier circles (majority). (Right) AUROC is computed over pixels that belong to squares, with those pixels from the square in the upper-right corner (truly OOD) labeled positive, and negative otherwise (minority inlier). Higher AUROC means the method better distinguishes an inlier but uncommon square from a strictly out-of-distribution square.

D THEORY

Here we provide a formal version of Theorem 4.1. Fix a test image $x^\star$ and a superpixel $S_k \in \mathcal{S}(x^\star)$. Let $\mathcal{V} := \mathbb{Z}_{256}^C$, where $\mathbb{Z}_{256} = \{0, \dots, 255\}$ denotes the set of possible 8-bit values for one color channel. Thus, $\mathcal{V}$ is the single-pixel state space, under the finite quantized histogram setting. We consider features as an empirical extension, and do not include them in the discussion here. Under Assumption 4.1, the ideal inpainting posterior $p_k^{\text{inp}} = p(x \mid x^\star_{-S_k})$ induces the single-pixel distribution

$$q_k^{\text{inp}} := p(x_j \mid x^\star_{-S_k}), \qquad q_k^{\text{inp}} : \mathcal{V} \to [0,1], \quad j \in S_k. \tag{11}$$

For $\eta \sim \mathcal{N}(0, \sigma^2 I)$, write $y := x^\star + \eta$. For a realization y^m of y, let $y^m := x^\star + \eta^m$ denote the noised version of $x^\star$ used to sample x^m, and define the pooled single-pixel denoising marginal for superpixel S_k as

$$q_{k,y^m}^{\text{den}} := \frac{1}{|S_k|} \sum_{j \in S_k} p(x_j \mid y^m), \qquad q_{k,y^m}^{\text{den}} : \mathcal{V} \to [0,1]. \tag{12}$$

The denoising surrogate distribution used by the score is then

$$q_k^{\text{den},\sigma} := \mathbb{E}_\eta \left[q_{k,y}^{\text{den}} \right]. \tag{13}$$

The corresponding ideal and denoising scores are

$$g^{\text{inp}}(x^\star, k) := \left| -\frac{1}{|S_k|} \sum_{j \in S_k} \log q_k^{\text{inp}}(x_j^\star) - |S_k| \mathbb{H}(q_k^{\text{inp}}) \right|, \tag{14}$$

$$g_\sigma^{\text{den}}(x^\star, k) := \left| -\frac{1}{|S_k|} \sum_{j \in S_k} \log q_k^{\text{den},\sigma}(x_j^\star) - |S_k| \mathbb{H}(q_k^{\text{den},\sigma}) \right|. \tag{15}$$

For a realization y^m of y, define the auxiliary distribution

$$q_{k,y^m}^{\text{aux}} := \frac{1}{|S_k|} \sum_{j \in S_k} p(x_j \mid x^\star_{-S_k}, y^m_{S_k}), \qquad q_{k,y^m}^{\text{aux}} : \mathcal{V} \to [0,1]. \tag{16}$$

This is the ideal inpainting conditional after additionally revealing the noisy superpixel values, pooled over pixels in S_k. It is used only to separate the two sources of error: loss of context from observing the noised version of $x^\star_{-S_k}$ instead of $x^\star_{-S_k}$, and the leakage of information about $x^\star_{S_k}$ from observing a noisy instead of completely masked superpixel. We will refer to the former as *context gap*, and the latter as *self-leakage*.

The following assumption formalizes the bias–variance-like tradeoff induced by using denoising instead of inpainting. A denoising conditional observes a noisy version of the entire image, whereas the ideal inpainting conditional observes the clean complement $x^\star_{-S_k}$ and does not observe $x^\star_{S_k}$. As σ increases, the noisy observation outside S_k becomes a worse proxy for the clean outside context and widens that context gap, which we capture as a linear function of σ in the form of $\alpha_k \sigma$. In the opposite direction, when σ is too small, the noisy values inside S_k reveal too much about $x^\star_{S_k}$, allowing the denoising posterior to copy the held-out region rather than predict it from context; this self-leakage is represented by the $\beta_k \sigma^{-\gamma_k}$ term.

Assumption D.1 (Denoising–inpainting tradeoff)**.** There exist constants $\alpha_k, \beta_k, \gamma_k > 0$ and $0 < \underline{\sigma}_k \le \overline{\sigma}_k$ such that, for every $\sigma \in [\underline{\sigma}_k, \overline{\sigma}_k]$,

$$\mathbb{E}_\eta \left[\left\| q_{k,y}^{\text{den}} - q_{k,y}^{\text{aux}} \right\|_1 \right] \le \alpha_k \sigma, \tag{17}$$

$$\mathbb{E}_\eta \left[\left\| q_{k,y}^{\text{aux}} - q_k^{\text{inp}} \right\|_1 \right] \le \beta_k \sigma^{-\gamma_k}. \tag{18}$$

We add another assumption, which is a mild regularity condition ensuring that small distributional errors translate into small score errors. Since the score contains both a log-likelihood term and an entropy term, probabilities near zero can make the functional unstable. The lower bound by δ_k rules out this degenerate case on the finite single-pixel state space. In practice, the computed histogram-based scores can enforce such a floor by smoothing or clipping the estimated probabilities.

Assumption D.2 (Score regularity)**.** There exists $\delta_k \in (0,1)$ such that, for every $\sigma \in [\underline{\sigma}_k, \overline{\sigma}_k]$,

$$q_k^{\mathrm{inp}}(v) \geq \delta_k \quad \text{and} \quad q_k^{\mathrm{den},\sigma}(v) \geq \delta_k \qquad \text{for all } v \in \mathcal{V}. \tag{19}$$

We now state the formal version of Theorem 4.1.

Theorem D.1 (Formal denoising score approximation)**.** Under Assumptions A.1 and A.2, for every $\sigma \in [\underline{\sigma}_k, \overline{\sigma}_k]$,

$$\left| g^{\mathrm{inp}}(x^\star, k) - g_\sigma^{\mathrm{den}}(x^\star, k) \right| \leq L_{\delta_k} \left(\alpha_k \sigma + \beta_k \sigma^{-\gamma_k} \right), \tag{20}$$

where

$$L_{\delta_k} := \left(\delta_k^{-1} + 1 + \log(\delta_k^{-1}) \right). \tag{21}$$

Equivalently, after absorbing L_{δ_k} into the constants, there exist positive constants $\widetilde{\alpha}_k$ and $\widetilde{\beta}_k$ such that

$$\left| g^{\mathrm{inp}}(x^\star, k) - g_\sigma^{\mathrm{den}}(x^\star, k) \right| \leq \widetilde{\alpha}_k \sigma + \widetilde{\beta}_k \sigma^{-\gamma_k}. \tag{22}$$

Proof. First, we begin by inspecting the error of the anomaly scores. For any two distributions $q, r : \mathcal{V} \to [0,1]$ satisfying $q(v), r(v) \geq \delta_k$ for every $v \in \mathcal{V}$,

$$\begin{aligned}
&\left| \left| -\frac{1}{|S_k|} \sum_{j \in S_k} \log q(x_j^\star) - \mathbb{H}(q) \right| - \left| -\frac{1}{|S_k|} \sum_{j \in S_k} \log r(x_j^\star) - \mathbb{H}(r) \right| \right| \\
&\leq \left| \frac{1}{|S_k|} \sum_{j \in S_k} \big(\log r(x_j^\star) - \log q(x_j^\star) \big) + \big(\mathbb{H}(r) - \mathbb{H}(q) \big) \right| \\
&\leq \frac{1}{|S_k|} \sum_{j \in S_k} \left| \log r(x_j^\star) - \log q(x_j^\star) \right| + \left| \mathbb{H}(q) - \mathbb{H}(r) \right|,
\end{aligned} \tag{23}$$

by applying reverse triangle inequality, and then triangle inequality. Since $t \mapsto -\log t$ has derivative bounded in absolute value by δ_k^{-1} on $[\delta_k, 1]$,

$$\frac{1}{|S_k|} \sum_{j \in S_k} \left| \log r(x_j^\star) - \log q(x_j^\star) \right| \leq \delta_k^{-1} \|q - r\|_1. \tag{24}$$

Similarly, for $h(t) = -t \log t$, the derivative $h'(t) = -1 - \log t$ satisfies

$$|h'(t)| \leq 1 + \log(\delta_k^{-1}) \qquad \text{for } t \in [\delta_k, 1]. \tag{25}$$

Applying the mean-value theorem coordinate-wise gives

$$|\mathbb{H}(q) - \mathbb{H}(r)| \leq \left(1 + \log(\delta_k^{-1}) \right) \|q - r\|_1. \tag{26}$$

Combining (23)–(26),

$$\left| \left| -\frac{1}{|S_k|} \sum_{j \in S_k} \log q(x_j^\star) - \mathbb{H}(q) \right| - \left| -\frac{1}{|S_k|} \sum_{j \in S_k} \log r(x_j^\star) - \mathbb{H}(r) \right| \right| \leq L_{\delta_k} \|q - r\|_1. \tag{27}$$

It remains to control $\|q_k^{\mathrm{den},\sigma} - q_k^{\mathrm{inp}}\|_1$. By the definition of $q_k^{\mathrm{den},\sigma}$ and convexity of the ℓ_1 norm,

$$\begin{aligned}
\left\| q_k^{\mathrm{den},\sigma} - q_k^{\mathrm{inp}} \right\|_1 &= \left\| \mathbb{E}_\eta \left[q_{k,y}^{\mathrm{den}} - q_k^{\mathrm{inp}} \right] \right\|_1 \\
&\leq \mathbb{E}_\eta \left[\left\| q_{k,y}^{\mathrm{den}} - q_k^{\mathrm{inp}} \right\|_1 \right].
\end{aligned} \tag{28}$$

For each realization y^m of y, insert q_{k,y^m}^{aux} and apply the triangle inequality:

$$\left\| q_{k,y^m}^{\mathrm{den}} - q_k^{\mathrm{inp}} \right\|_1 \leq \left\| q_{k,y^m}^{\mathrm{den}} - q_{k,y^m}^{\mathrm{aux}} \right\|_1 + \left\| q_{k,y^m}^{\mathrm{aux}} - q_k^{\mathrm{inp}} \right\|_1. \tag{29}$$

Taking expectations and using Assumption A.1 yields

$$\left\| q_k^{\mathrm{den},\sigma} - q_k^{\mathrm{inp}} \right\|_1 \leq \alpha_k \sigma + \beta_k \sigma^{-\gamma_k}. \tag{30}$$

Finally, apply (27) with $q = q_k^{\mathrm{den},\sigma}$ and $r = q_k^{\mathrm{inp}}$. This proves (20). The absorbed-constant form (22) follows by setting $\widetilde{\alpha}_k = L_{\delta_k} \alpha_k$ and $\widetilde{\beta}_k = L_{\delta_k} \beta_k$. □

Corollary D.1 (Bound-optimal denoising noise)**.** Under the conditions of Theorem D.1, the upper bound in (20) is minimized over $\sigma \in [\underline{\sigma}_k, \overline{\sigma}_k]$ by

$$\sigma_k^{\mathrm{opt}} := \min\left\{\overline{\sigma}_k, \max\left\{\underline{\sigma}_k, \left(\frac{\gamma_k \beta_k}{\alpha_k}\right)^{1/(\gamma_k+1)}\right\}\right\}. \tag{31}$$

In particular, if

$$\left(\frac{\gamma_k \beta_k}{\alpha_k}\right)^{1/(\gamma_k+1)} \in [\underline{\sigma}_k, \overline{\sigma}_k], \tag{32}$$

then the bound is minimized at the interior value

$$\sigma_k^{\mathrm{opt}} = \left(\frac{\gamma_k \beta_k}{\alpha_k}\right)^{1/(\gamma_k+1)}, \tag{33}$$

and the minimum value of the right-hand side of (20) is

$$L_{\delta_k}(\gamma_k + 1)\gamma_k^{-\gamma_k/(\gamma_k+1)} \alpha_k^{\gamma_k/(\gamma_k+1)} \beta_k^{1/(\gamma_k+1)}. \tag{34}$$

Proof. The multiplicative factor L_{δ_k} in (20) is positive and therefore does not affect the choice of σ. It is enough to minimize

$$\alpha_k \sigma + \beta_k \sigma^{-\gamma_k}. \tag{35}$$

The derivative is

$$\alpha_k - \gamma_k \beta_k \sigma^{-\gamma_k - 1}, \tag{36}$$

which vanishes exactly when

$$\sigma = \left(\frac{\gamma_k \beta_k}{\alpha_k}\right)^{1/(\gamma_k+1)}. \tag{37}$$

The second derivative is

$$\gamma_k(\gamma_k + 1)\beta_k \sigma^{-\gamma_k - 2} > 0, \tag{38}$$

so this critical point is the unique unconstrained minimizer. Restricting σ to $[\underline{\sigma}_k, \overline{\sigma}_k]$ gives (31). When the unconstrained minimizer lies inside the interval, substituting (33) into $L_{\delta_k}(\alpha_k \sigma + \beta_k \sigma^{-\gamma_k})$ gives (34). □

This corollary should be interpreted as selecting the *bound-optimal* noise level rather than the necessarily optimal empirical denoising noise. The expression balances the two errors in Assumption D.1: increasing σ worsens the context gap, while decreasing σ worsens self-leakage. At an interior optimum, these two effects satisfy $\alpha_k \sigma_k^{\mathrm{opt}} = \gamma_k \beta_k (\sigma_k^{\mathrm{opt}})^{-\gamma_k}$.

We now analyze the typicality score without Assumption 4.1. The signed statistic remains centered under the ID model, while dependence affects its concentration through the covariance of the per-pixel information values.

Fix S_k and $x^\star_{-S_k}$, and write $n_k := |S_k|$. Since the single-pixel marginals may now differ, replace q_k^{inp} by their pooled marginal

$$\bar{q}_k^{\mathrm{inp}}(v) := \frac{1}{n_k} \sum_{j \in S_k} p(x_j = v \mid x^\star_{-S_k}), \qquad v \in \mathcal{V}. \tag{39}$$

This reduces to q_k^{inp} under Assumption 4.1. Retaining the probability floor of Assumption D.2 for $\bar{q}_k^{\mathrm{inp}}$, define

$$\begin{aligned} Z_{k,j} &:= -\log \bar{q}_k^{\mathrm{inp}}(x_j) - \mathbb{H}(\bar{q}_k^{\mathrm{inp}}), \\ T_k^{\mathrm{inp}} &:= \frac{1}{n_k} \sum_{j \in S_k} Z_{k,j}, \qquad G_k^{\mathrm{inp}} := |T_k^{\mathrm{inp}}|. \end{aligned} \tag{40}$$

Here G_k^{inp} is the per-pixel counterpart of the score in (14), using the pooled marginal.

By the definition of $\bar{q}_k^{\mathrm{inp}}$, the expected average information is $\mathbb{H}(\bar{q}_k^{\mathrm{inp}})$, and hence $\mathbb{E}_{p_k^{\mathrm{inp}}}[T_k^{\mathrm{inp}}] = 0$. Dependence enters through the total variance

$$\begin{aligned}\Gamma_k^{\mathrm{ID}} &:= \mathrm{Var}_{p_k^{\mathrm{inp}}}\left(\sum_{j\in S_k} Z_{k,j}\right) \\ &= \sum_{j\in S_k} \mathrm{Var}_{p_k^{\mathrm{inp}}}(Z_{k,j}) + 2\sum_{\substack{i,j\in S_k \\ i<j}} \mathrm{Cov}_{p_k^{\mathrm{inp}}}(Z_{k,i}, Z_{k,j}).\end{aligned} \tag{41}$$

Since $\mathrm{Var}_{p_k^{\mathrm{inp}}}(T_k^{\mathrm{inp}}) = \Gamma_k^{\mathrm{ID}}/n_k^2$, Chebyshev's inequality gives

$$\mathbb{P}_{p_k^{\mathrm{inp}}}(G_k^{\mathrm{inp}} \geq \tau) \leq \frac{\Gamma_k^{\mathrm{ID}}}{n_k^2\tau^2}, \qquad \tau > 0. \tag{42}$$

To express this in terms of effective sample size, let

$$v_{k,\mathrm{ID}}^2 := \frac{1}{n_k}\sum_{j\in S_k} \mathrm{Var}_{p_k^{\mathrm{inp}}}(Z_{k,j}), \qquad n_{k,\mathrm{eff}}^{\mathrm{ID}} := \frac{n_k^2 v_{k,\mathrm{ID}}^2}{\Gamma_k^{\mathrm{ID}}} \quad (\Gamma_k^{\mathrm{ID}} > 0). \tag{43}$$

The bound becomes $v_{k,\mathrm{ID}}^2/(n_{k,\mathrm{eff}}^{\mathrm{ID}}\tau^2)$. Under Assumption 4.1, the cross-covariances vanish and $n_{k,\mathrm{eff}}^{\mathrm{ID}} = n_k$; a positive sum of cross-covariances reduces the effective sample size and weakens concentration.

The same variance dependence appears in OOD detection. For an OOD conditional distribution R_k, let $\Delta_k := |\mathbb{E}_{R_k}[T_k^{\mathrm{inp}}]|$ and $\Gamma_k^{\mathrm{OOD}} := \mathrm{Var}_{R_k}(\sum_{j\in S_k} Z_{k,j})$. If $G_k^{\mathrm{inp}} < \tau < \Delta_k$, then $|T_k^{\mathrm{inp}} - \mathbb{E}_{R_k}[T_k^{\mathrm{inp}}]| \geq \Delta_k - \tau$. Chebyshev's inequality therefore yields

$$\mathbb{P}_{R_k}(G_k^{\mathrm{inp}} < \tau) \leq \frac{\Gamma_k^{\mathrm{OOD}}}{n_k^2(\Delta_k - \tau)^2}, \qquad 0 < \tau < \Delta_k. \tag{44}$$

Finally, let $G_{k,\sigma}^{\mathrm{den}}$ be the score in (15) divided by n_k, with its denoising marginal evaluated at the image being scored. If Assumptions D.1 and D.2 hold uniformly over the considered superpixel realizations, with common constants and $\bar{q}_k^{\mathrm{inp}}$ replacing q_k^{inp}, the proof of Theorem D.1 gives

$$|G_k^{\mathrm{inp}} - G_{k,\sigma}^{\mathrm{den}}| \leq e_k(\sigma), \qquad e_k(\sigma) := L_{\delta_k}\big(\alpha_k\sigma + \beta_k\sigma^{-\gamma_k}\big). \tag{45}$$

Applying (42) and (44) at the resulting shifted thresholds gives

$$\mathbb{P}_{p_k^{\mathrm{inp}}}(G_{k,\sigma}^{\mathrm{den}} \geq \tau) \leq \frac{\Gamma_k^{\mathrm{ID}}}{n_k^2(\tau - e_k(\sigma))^2}, \qquad \tau > e_k(\sigma), \tag{46}$$

$$\mathbb{P}_{R_k}(G_{k,\sigma}^{\mathrm{den}} < \tau) \leq \frac{\Gamma_k^{\mathrm{OOD}}}{n_k^2(\Delta_k - \tau - e_k(\sigma))^2}, \qquad 0 < \tau < \Delta_k - e_k(\sigma). \tag{47}$$

Under Assumption 4.1, the effective sample size equals the number of pixels in the superpixel. Without this assumption, the signed statistic remains centered under the ID model, but positive within-superpixel covariance reduces the effective sample size and weakens the concentration and detection bounds. The denoising approximation retains the same form, with the effect of dependence captured by the variance terms.

E Experimental Details

E.1 Dataset summary

Table 4 summarizes the train/test composition, masking, and image resolution for each dataset. **MVTec AD** and **CHAOS** are evaluated at 256×256 following standard practice; **FFHQ** is 256×256; **DVXray** is 600×800 in its native form, and we report per-pixel mean metrics so resolution does not bias cross-dataset comparison.

Table 4: Dataset composition. Training images contain only inlier (anomaly-free) examples. Test images contain local anomalies and are paired with pixel-level ground-truth masks.

Dataset	# train (inlier)	# test (ID + OOD)	Eval mask	Resolution
CHAOS	3,070	100 + 100	body silhouette	256×256
FFHQ	5,000	50 + 50	—	256×256
DVXray (scissors)	5,000	72 + 72	—	600×800
MVTec AD (15 categories)	3,629	1,258	—	256×256

E.2 Diffusion models and reconstruction

For **MVTec AD**, we use DDAD's checkpoints, which are ϵ-prediction UNet diffusion models, trained separately for each category. For all other benchmarks, we use the guided-diffusion UNet of Dhariwal and Nichol (Dhariwal and Nichol, 2021) trained from scratch on each inlier set, and produce reconstructions via DPS (Chung et al., 2022).

E.3 LACToS hyperparameters

For sample reconstruction using DPS (Chung et al., 2022), we utilize the identity forward model and zero-mean Gaussian noise with variance σ_{noise}. DPS also has a scale parameter that controls the posterior guidance strength.

Once M reconstructions are computed and the corresponding features are acquired, we rescale RGB values from $[0, 255]$ and autoencoder outputs from $[0, 1]$ to a common $[0, 255]$ range. Then, we partition the reconstructions and feature maps into superpixels containing an average of S_{avg} pixels each, using marker-controlled watershed segmentation (Meyer and Beucher, 1990). For each superpixel, we quantize the rescaled values into B_{RGB} and B_{feat} bins per dimension, for RGB values and autoencoder outputs respectively. While it is ideal to construct $C + C'_f$-dimensional histograms, the number of histogram bins grows exponentially with the dimension. Therefore, we construct two empirical histograms per superpixel, one over RGB values and one over feature values. We smooth each histogram with a Gaussian kernel of σ_{hist} to mitigate empty bins under limited samples, and use the histograms to compute the RGB and feature typicality scores for each superpixel. The two scores are summed with a weight parameter λ multiplied to the feature typicality to balance the trade-off between the two scores. Finally, we smooth the anomaly map that contains the same value over all pixels in each superpixel to acquire the pixel-level anomaly map. All parameters are selected by performing a grid search over a small subset containing 20 test images, and the configuration with the highest AUROC is used to evaluate the whole test set. The specific parameters used for each benchmark can be found in Table 5.

Table 5: Dataset-specific hyperparameters used for LACToS.

Parameter	**CHAOS**	**FFHQ**	**DVXray**
DPS σ_{noise}	0.1	0.01	2.0
DPS scale	0.4	0.1	1.0
S_{avg}	100	20	10
M	5	20	20
σ_{hist}	0.5	1.0	0
$(B_{\text{RGB}}, B_{\text{feat}})$	$(16, 64)$	$(16, 32)$	$(16, 16)$
λ	1.0	0.25	0

E.4 Baselines

For baselines, we use the official implementations of PatchCore, SimpleNet, SuperSimpleNet, EfficientAD, GLASS, and DDAD, with each method's default backbone: WideResNet-50 for PatchCore; ImageNet-pretrained ResNet/WideResNet for the SimpleNet variants.

For fair comparison, we perform parameter tuning for each baseline method using the same set of 20 test images. Note that we do not tune the parameters of SimpleNet, SuperSimpleNet, and GLASS as retraining models are required; we use the default parameters in the respective manuscripts or github repositories. For PatchCore, k, the number of nearest neighbors in the coreset to compare each test patch, is swept across $k \in \{1, 2, 3, 5, 7, 9, 15\}$. For EfficientAD, α, which balances the strengths of the student-teacher detector and the autoencoder detector, is swept across $\alpha \in \{0, 0.1, 0.2, \cdots, 1.0\}$. For DDAD, a 2d grid of $(w, v) \in \{1, 2, 4, 6, 8\} \times \{0.5, 1, 2, 4\}$ is searched, where w controls the guidance strength during denoising, and v balances the strengths of pixels and extracted features. The selected parameter values are described in Table 6.

Table 6: Dataset-specific hyperparameters used for baseline methods.

Parameter	**CHAOS**	**FFHQ**	**DVXray**
PatchCore	$k = 1$	$k = 1$	$k = 1$
EfficientAD	$\alpha = 0$	$\alpha = 0.7$	$\alpha = 0.2$
DDAD	$(w, v) = (2, 1)$	$(w, v) = (2, 4)$	$(w, v) = (4, 4)$

E.5 Feature extractor and autoencoder

The feature extractor ϕ is an ImageNet-pretrained ResNet-101 He et al. (2016), frozen at all stages. We extract activations from the last three residual stages, upsample them to the input resolution, and concatenate them. To reduce histogram error under limited samples, these features are compressed to $C'_f = 3$ dimensions by a per-pixel autoencoder trained on each inlier dataset.

F ADDITIONAL RESULTS

We provide additional visualizations of each benchmark in Figure 10, Figure 11, and Figure 12.

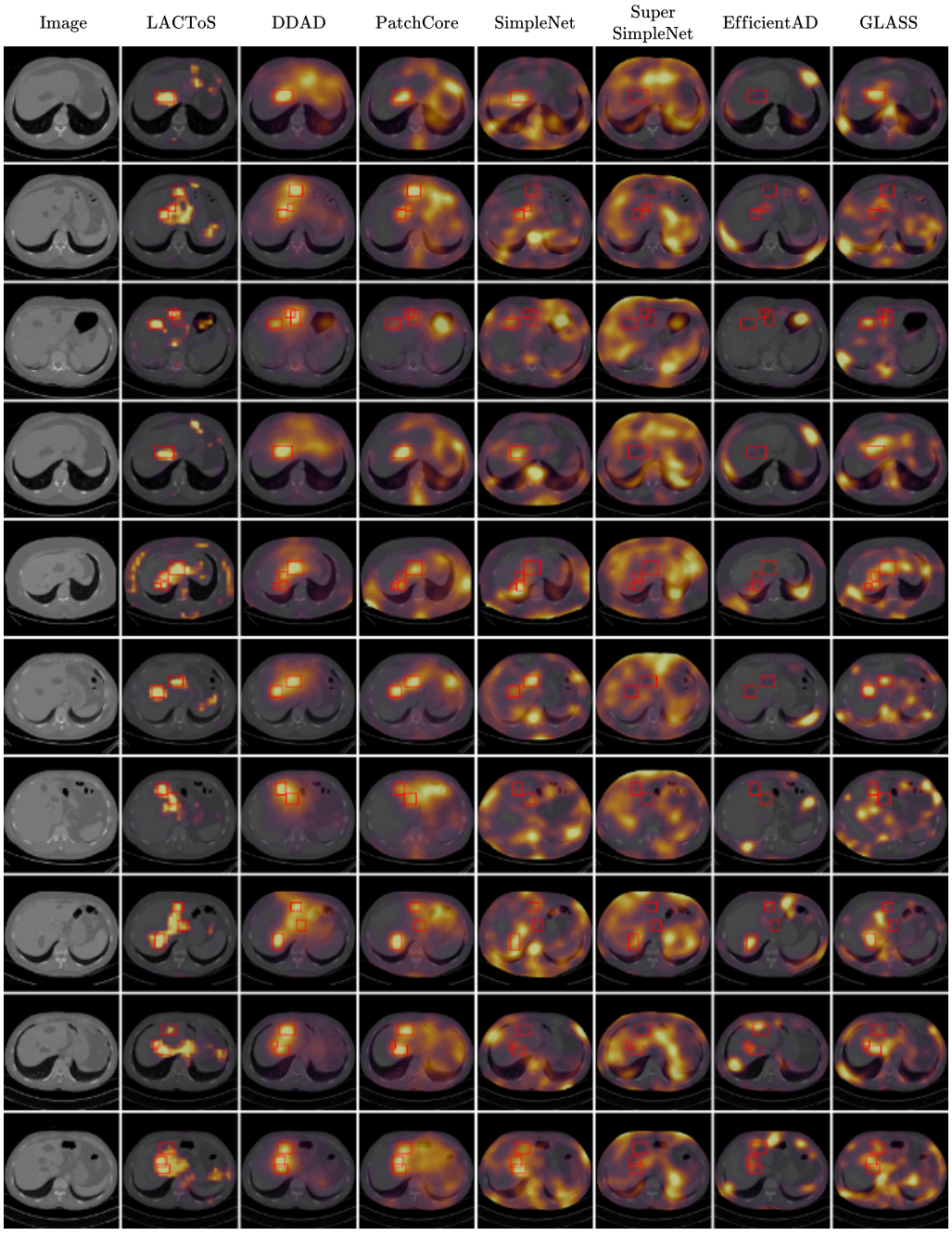


Figure 10: **CHAOS Additional Visualizations.** 10 randomly selected test images and the heatmaps for each method.

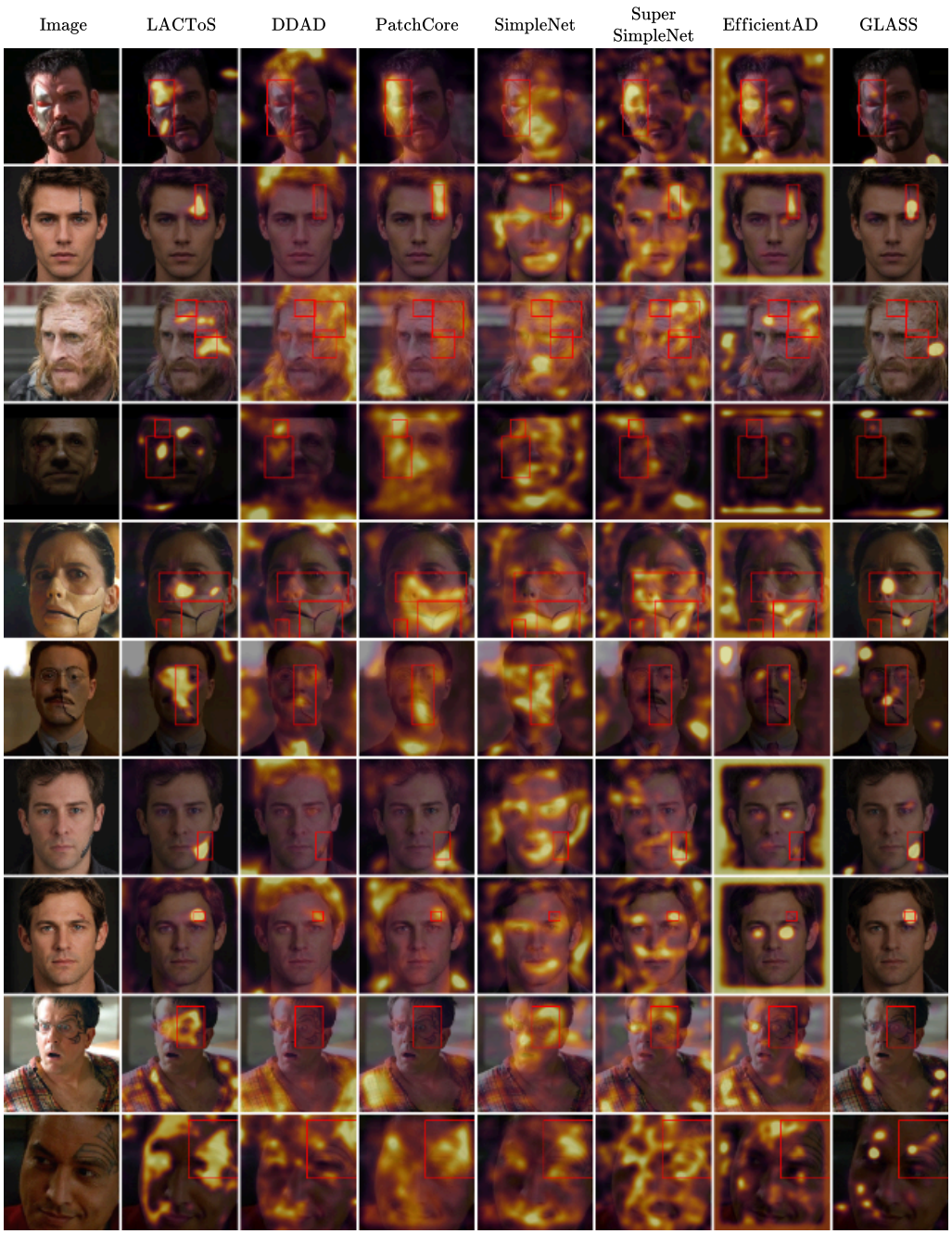


Figure 11: **FFHQ Additional Visualizations.** 10 randomly selected test images and the heatmaps for each method.

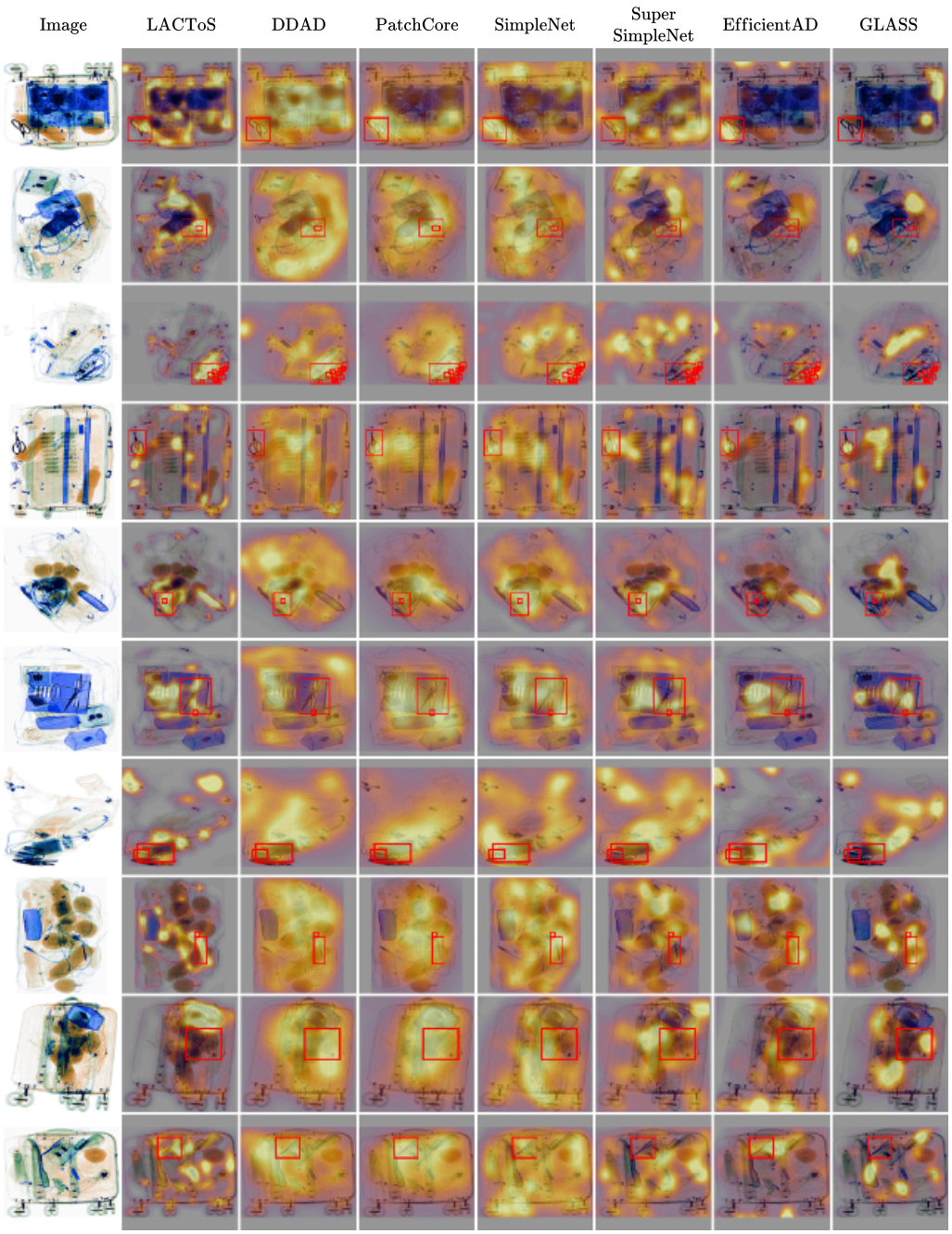


Figure 12: **DVXray Additional Visualizations.** 10 randomly selected test images and the heatmaps for each method.

In Table 7, we report the AUROC, AP, and AUPRO of all methods for each of the object category in the MVTec dataset.

Table 7: **Results on MVTec AD.** We report pixel-level AUROC, AP, and AUPRO for each object or texture category. AUROC and AP pool every pixel of all test images of the category (anomalous and anomaly-free); AUPRO is the region-based per-region overlap integrated up to an FPR of 0.3. Each method occupies three rows, one per metric. Higher is better for all metrics.

Method	Metric	Carpet	Grid	Leather	Tile	Wood	Bottle	Cable	Capsule
SimpleNet	AUROC	0.9765	0.9656	0.9898	0.9326	0.9177	0.9775	0.9688	0.9862
	AP	0.3650	0.2440	0.3257	0.5470	0.3673	0.6315	0.6151	0.3729
	AUPRO	0.9047	0.8757	0.9630	0.7781	0.7866	0.8945	0.8741	0.9093
SuperSimpleNet	AUROC	0.9832	0.9771	0.9878	0.9383	0.9208	0.9779	0.9728	0.9805
	AP	0.4548	0.2383	0.3492	0.5109	0.3867	0.6202	0.6151	0.3973
	AUPRO	0.9222	0.9236	0.9686	0.8107	0.8551	0.8982	0.8746	0.8753
PatchCore	AUROC	0.9877	0.9729	0.9903	0.9496	0.9301	0.9813	0.9831	0.9873
	AP	0.5770	0.2326	0.3533	0.4947	0.4016	0.6842	0.6258	0.3935
	AUPRO	0.9295	0.8932	0.9574	0.8212	0.8312	0.9253	0.9153	0.9227
DDAD	AUROC	0.9822	0.9795	0.9878	0.9790	0.9374	0.9860	0.9784	0.8693
	AP	0.6422	0.2710	0.4124	0.7853	0.5285	0.8047	0.6641	0.2313
	AUPRO	0.8994	0.9291	0.9811	0.9117	0.8604	0.9003	0.8835	0.6845
EfficientAD	AUROC	0.9831	0.9833	0.9916	0.9816	0.9680	0.9884	0.9863	0.9915
	AP	0.7511	0.4328	0.5421	0.7505	0.6273	0.7852	0.7593	0.4603
	AUPRO	0.9489	0.9144	0.9840	0.9249	0.9207	0.9550	0.9009	0.9445
GLASS	AUROC	0.9952	0.9908	0.9952	0.9963	0.9761	0.9922	0.9569	0.9898
	AP	0.8108	0.4528	0.5518	0.9640	0.7196	0.8966	0.4997	0.4798
	AUPRO	0.9832	0.9606	0.9815	0.9770	0.9467	0.9652	0.8470	0.9460
LACToS (Ours)	AUROC	0.9392	0.9402	0.9922	0.9377	0.8774	0.9782	0.9524	0.7731
	AP	0.2634	0.2504	0.5493	0.6480	0.4104	0.7594	0.5130	0.0420
	AUPRO	0.8128	0.7731	0.9465	0.8629	0.7300	0.8728	0.8265	0.4932

Method	Metric	Hazelnut	Metal Nut	Pill	Screw	Toothbrush	Transistor	Zipper	Average
SimpleNet	AUROC	0.9741	0.9747	0.9839	0.9887	0.9823	0.9746	0.9790	0.9715
	AP	0.3978	0.8049	0.7402	0.2325	0.3377	0.6181	0.4712	0.4714
	AUPRO	0.9297	0.8082	0.9009	0.9464	0.8404	0.9354	0.9091	0.8837
SuperSimpleNet	AUROC	0.9792	0.9855	0.9841	0.9873	0.9851	0.9618	0.9813	0.9735
	AP	0.5013	0.8630	0.7763	0.2242	0.4923	0.6081	0.4395	0.4985
	AUPRO	0.9447	0.9004	0.9326	0.9434	0.8473	0.9130	0.9304	0.9027
PatchCore	AUROC	0.9851	0.9659	0.9895	0.9908	0.9827	0.9742	0.9813	0.9768
	AP	0.5019	0.7060	0.7824	0.2585	0.3272	0.6312	0.4444	0.4943
	AUPRO	0.9559	0.8557	0.9517	0.9572	0.8613	0.9431	0.9244	0.9097
DDAD	AUROC	0.9832	0.9840	0.9461	0.8490	0.9729	0.9436	0.9751	0.9569
	AP	0.5405	0.8580	0.6220	0.0202	0.5126	0.6341	0.5388	0.5377
	AUPRO	0.9279	0.9072	0.5272	0.5060	0.6514	0.9233	0.9056	0.8266
EfficientAD	AUROC	0.9856	0.9847	0.9885	0.9945	0.9882	0.9775	0.9892	0.9855
	AP	0.5492	0.8220	0.7474	0.4125	0.5202	0.8170	0.5594	0.6358
	AUPRO	0.9457	0.9337	0.9606	0.9692	0.9315	0.9455	0.9615	0.9427
GLASS	AUROC	0.9863	0.8987	0.9727	0.9315	0.9629	0.5712	0.9948	0.9474
	AP	0.6267	0.5778	0.8330	0.2391	0.2321	0.1719	0.7725	0.5886
	AUPRO	0.9733	0.9004	0.8555	0.7053	0.7514	0.3643	0.9822	0.8760
LACToS (Ours)	AUROC	0.9852	0.9656	0.9419	0.8559	0.9512	0.9684	0.8602	0.9279
	AP	0.7546	0.8206	0.5889	0.0106	0.5025	0.7168	0.1435	0.4649
	AUPRO	0.8898	0.7961	0.7456	0.5294	0.5672	0.9245	0.6963	0.7644

G Ablation Studies

In this section, we ablate our method on the different choices of hyper-parameters.

Sample efficiency. Table 8 reports the AUROC, AP, and AUPRO of diffusion-based methods DDAD (Mousakhan et al., 2024) and LACToS as the number of denoising samples M varies from 1 to 20. LACToS exhibits improved performance with increasing M on **CHAOS** and **DVXray** compared to DDAD, which relies on point-estimates of reconstruction errors. This is consistent with the statistical properties of typicality: more samples yield better-estimations of $\hat{q}_k^{\text{inp}}$, which improves the reliability of the entropy and log likelihood estimates. At $M = 1$, LACToS degenerates toward a reconstruction-error-like score.

Superpixel size. Table 9 shows the effect of the number of superpixels' average size in the performance of LACToS, which is mostly stable but with different trends across benchmarks.

AutoEncoder features. The sensitivity of the metrics under varying λ is illustrated in Table 10. $\lambda = 0$ corresponds to using RGB typicality only. In most datasets, including feature typicality significantly improves performance.

Table 8: Ablation on sample size M in diffusion-based methods. AUROC, AP, and AUPRO are reported for each dataset.

Dataset	Metric	$M = 1$		$M = 5$		$M = 10$		$M = 20$	
		DDAD	LACToS	DDAD	LACToS	DDAD	LACToS	DDAD	LACToS
CHAOS	AUROC	0.9707	0.9552	0.9748	0.9620	0.9749	0.9615	0.9745	0.9620
	AP	0.3159	0.2536	0.3496	0.4318	0.3468	0.4236	0.3489	0.4338
	AUPRO	0.8676	0.8524	0.8827	0.8764	0.8825	0.8713	0.8814	0.8729
FFHQ	AUROC	0.7469	0.8446	0.7512	0.8521	0.7559	0.8503	0.7583	0.8515
	AP	0.0563	0.0937	0.0543	0.1197	0.0552	0.1119	0.0564	0.1177
	AUPRO	0.3083	0.4638	0.3065	0.5157	0.3202	0.5239	0.3151	0.5320
DVXray	AUROC	0.8567	0.9347	0.8518	0.9541	0.8523	0.9585	0.8495	0.9607
	AP	0.0118	0.0248	0.0119	0.0465	0.0119	0.0530	0.0118	0.0568
	AUPRO	0.5183	0.7385	0.4896	0.7964	0.4880	0.8055	0.4890	0.8179

Table 9: Ablation on the average size of superpixels. The larger the average size, the smaller the total number of superpixels. AUROC, AP, and AUPRO on **CHAOS**, **FFHQ**, and **DVXray**.

Dataset	Metric	Average size of superpixels						
		10	20	30	50	70	100	200
CHAOS	AUROC	0.9418	0.9504	0.9535	0.9578	0.9595	0.9620	0.9593
	AP	0.1271	0.1599	0.1857	0.2434	0.2643	0.4318	0.3677
	AUPRO	0.7992	0.8274	0.8385	0.8499	0.8571	0.8764	0.8600
FFHQ	AUROC	0.7930	0.8515	0.8502	0.8480	0.8462	0.8438	0.8383
	AP	0.0774	0.1177	0.1145	0.1107	0.1095	0.1065	0.0998
	AUPRO	0.3890	0.5320	0.5263	0.5180	0.5101	0.5068	0.4863
DVXray	AUROC	0.9607	0.9611	0.9593	0.9501	0.9468	0.9463	0.9330
	AP	0.0568	0.0570	0.0537	0.0495	0.0458	0.0418	0.0342
	AUPRO	0.8179	0.8122	0.8125	0.8040	0.7872	0.7817	0.7466

Table 10: Ablation on the feature-weighting parameter λ. $\lambda = 0$ corresponds to using RGB features only. AUROC, AP, and AUPRO are reported on **CHAOS**, **FFHQ**, and **DVXray**.

Dataset	Metric	$\lambda = 0$	$\lambda = 0.25$	$\lambda = 0.5$	$\lambda = 0.75$	$\lambda = 1.0$	$\lambda = 1.5$	$\lambda = 2$	$\lambda = 4$	Feature only
CHAOS	AUROC	0.8418	0.9638	0.9632	0.9625	0.9620	0.9613	0.9609	0.9601	0.9592
	AP	0.0729	0.3893	0.4318	0.4356	0.4318	0.4192	0.4056	0.3645	0.2733
	AUPRO	0.5689	0.8764	0.8778	0.8772	0.8764	0.8755	0.8748	0.8735	0.8717
FFHQ	AUROC	0.7605	0.8515	0.8488	0.8459	0.8437	0.8406	0.8386	0.8347	0.8293
	AP	0.0677	0.1177	0.1115	0.1075	0.1041	0.0997	0.0975	0.0944	0.0916
	AUPRO	0.3688	0.5320	0.5353	0.5345	0.5331	0.5305	0.5287	0.5252	0.5211
DVXray	AUROC	0.9607	0.9589	0.9522	0.9454	0.9397	0.9305	0.9235	0.9057	0.8088
	AP	0.0568	0.0535	0.0425	0.0336	0.0285	0.0232	0.0205	0.0165	0.0115
	AUPRO	0.8179	0.7980	0.7686	0.7432	0.7223	0.6905	0.6675	0.6122	0.4791