

A Unified Latent Schrödinger Bridge Diffusion Model for Unsupervised Anomaly Detection and Localization

Supplementary Material

Shilhora Akshay Patel^{1,*}

Niveditha Lakshmi Narasimhan²

Jacob George²

Vineeth N Balasubramanian¹

¹Indian Institute of Technology, Hyderabad ²KLA Corporation

*shilhora.akshay333@gmail.com

Our code is available at <https://github.com/ShilhoraAkshayPatel/LASB> for further research and reproducibility. This supplementary material contains additional details that we could not include in the main paper due to space constraints, including the following information:

Contents

S1 Generalization Ability	1
S2 Extended Results	2
S2.1MvTec Results	2
S2.2VisA Results	2
S3 Sampling Stability	2
S4 Hyperparameters for Our LASB Model	3
S5 LASB Robustness to Transformations	4
S6 Visualizations	6
S7 Limitations	6
S8 Broader Impact	6

S1. Generalization Ability

Hold-out test set generalization is essential for assessing how well the proposed Latent Space Anomaly Schrödinger Bridge (LASB) can handle new, unseen data, which is pivotal for real-world applications. This type of generalization ensures the model can identify anomalies not just in controlled settings but in practical environments where conditions may vary widely from the training data. In the context of anomaly detection, the ability to generalize effectively means the model can accurately detect deviations without

	Category	2x	4x	8x
Objects	Bottle	99.7/98.7/96.3	99.7/98.7/96.3	99.62/98.62/96.3
	Cable	96.35/99.25/97.3	96.35/99.25/97.3	96.33/99.23/97.3
	Capsule	98.2/99.3/97.3	98.2/99.3/97.3	98.2/99.3/97.35
	Hazelnut	99.65/99.25/99.45	99.65/99.25/99.45	99.63/99.26/99.43
	Metal Nut	99.5/98.4/97.2	99.5/98.4/97.2	99.5/98.42/97.2
	Pill	99.15/98.7/96.15	99.15/98.7/96.15	99.06/98.67/96.18
	Screw	96.6/97.7/97.5	96.6/97.7/97.5	96.52/97.62/97.45
	Toothbrush	99.6/99.65/99.3	99.6/99.65/99.3	99.57/99.56/99.25
	Transistor	99.45/99.4/99.15	99.45/99.4/99.15	99.43/99.4/99.08
	Zipper	99.75/99.3/99.2	99.75/99.3/99.2	99.68/99.25/99.2
Textures	Carpet	99.25/99.15/99.4	99.25/99.15/99.4	99.26/99.21/99.37
	Grid	99.55/99.5/98.95	99.55/99.5/98.95	99.53/99.45/98.93
	Leather	99.65/99.55/99.2	99.65/99.55/99.2	99.58/99.48/99.2
	Tile	99.25/99.65/99.35	99.25/99.65/99.35	99.23/99.58/99.33
	Wood	99.55/99.6/99.25	99.55/99.6/99.25	99.48/99.55/99.23
	Mean	99.01/99.14/98.33	99.01/99.14/99.83	98.98/99.11/98.32

Table S1. Detection results on MVTEC-AD dataset for multi-class anomaly detection with AUROC_{cls}/AP_{cls}/F1max_{cls} metrics using multiple samplings (2x, 4x, 8x denote the number of samplings considered)

being misled by noise or irrelevant data variations, thereby avoiding costly false positives or missed defects.

When applying the LASB model to a diverse dataset like VisA, which is known for its varied anomaly presentations and complex visual settings, the hold-out test set generalization becomes a critical measure of the model’s utility across different scenarios. The performance metrics across various models in the Table. S4 reveal that the LASB model consistently outperforms most other models in different categories, achieving impressive results in detecting anomalies. Particularly notable is LASB’s superior performance in categories like *pcb4*, *pipe fryum*, and *chewinggum*, where it scores remarkably high precision and recall rates. LASB’s outstanding mean scores of 94.2% precision and 94.5% recall across all categories underscore its capability to generalize well across different types of data and anomaly conditions. The second-highest performing model, DiAD [6], has mean precision and recall scores of 86.8% and 85.1%

Category		2x	4x	8x
Objects	Bottle	99.4/89.2/62.5	99.42/89.17/62.52	99.4/89.2/62.5
	Cable	96.5/65.3/58.2	96.52/65.32/58.22	96.5/65.3/58.2
	Capsule	98.2/61.3/66.3	98.22/61.27/66.32	98.2/61.3/66.3
	Hazelnut	99.25/88.45/86.4	99.25/88.47/86.42	99.23/88.43/86.4
	Metal Nut	99.25/99.15/98.5	99.22/99.17/98.52	99.23/99.16/98.5
	Pill	98.15/75.6/64.45	98.12/75.62/64.4	98.13/75.6/64.43
	Screw	98.75/59.7/62.7	98.77/59.67/62.75	98.73/59.65/62.7
	Toothbrush	99.2/79.65/80.25	99.22/79.67/80.22	99.2/79.63/80.21
	Transistor	97.25/69.5/57.25	97.22/69.47/57.22	97.21/69.45/57.21
	Zipper	99.25/78.55/79.7	99.22/78.52/79.72	99.21/78.51/79.7
Textures	Carpet	98.65/78.55/68.25	98.67/78.57/68.3	98.63/78.53/68.26
	Grid	99.05/76.25/66.25	99.17/76.27/66.25	99.06/76.23/66.23
	Leather	99.25/75.3/66.95	99.27/75.32/67.05	99.26/75.3/67.01
	Tile	99.25/97/73.7	99.27/97.07/73.67	99.23/97.05/73.65
	Wood	97.3/78.6/67.45	97.32/78.57/67.47	97.3/78.55/67.43
Mean		98.58/78.14/70.59	98.59/78.14/70.61	98.57/78.12/70.58

Table S2. Localization results on MVTEC-AD dataset for multi-class anomaly localization with $AUROC_{seg}/AP_{seg}/F1max_{seg}$ metrics using multiple samplings (2x, 4x, 8x denote the number of samplings considered)

respectively, which are significantly lower than LASB’s 94.2% and 94.5%. This highlights LASB’s superior detection capabilities and robustness by a margin of over 7%, underscoring its enhanced effectiveness in anomaly detection.

Now in localization, the performance of LASB from Table S3 distinctly highlight the LASB model’s superior capability in anomaly localization on the VisA dataset, outperforming other models significantly. With mean scores of 98.2% for $AUROC_{seg}$, 46.4% for AP_{seg} , and 52.6% for $F1max_{seg}$, LASB demonstrates exceptional accuracy and reliability in identifying and localizing various types of anomalies across different categories. The second-highest performing model, DiAD [6], shows mean scores of 96.0% for $AUROC_{seg}$, 26.1% for AP_{seg} , and 33.0% for $F1max_{seg}$. The substantial performance gap, especially in AP_{seg} and $F1max_{seg}$ where LASB leads by over 20%, illustrates LASB’s enhanced precision and effectiveness. This comparison underlines LASB’s robustness and its advanced ability to manage complex anomaly detection and localization tasks with greater efficiency.

S2. Extended Results

S2.1. MvTec Results

Table S5 presents an exhaustive comparison of state-of-the-art methods for anomaly detection and localization on the widely-used MVTEC-AD dataset. The methods are categorized into class-based and unified (multi-class) approaches, and their performances are reported in terms of two key metrics: Image-level AUROC (I-AUROC) and Pixel-level AUROC (P-AUROC). Each column represents a specific object class from the dataset, while the rows outline methods published in top-tier conferences, highlighting the venue, model type, and performance across the

object classes. The methods vary significantly in design, ranging from embedding-based models (e.g., SimpleNet, PatchCore) to advanced diffusion-based frameworks (e.g., DiffAD, LASB). Notably, LASB, the proposed method, achieves the highest average scores in both I-AUROC (99.66) and P-AUROC (99.15), demonstrating state-of-the-art performance.

The table S5 underscores the evolving nature of anomaly detection models, transitioning from traditional embedding and GAN-based architectures to more powerful diffusion models. Diffusion-based methods like LASB consistently outperform earlier approaches, particularly in challenging classes like "Screw" and "Transistor," where subtle anomalies often go undetected. This highlights the robustness of diffusion frameworks in modeling complex patterns and their generalizability across diverse object types. The results validate LASB’s capability to outperform other diffusion methods, such as TransFusion [3] and D3AD [16], solidifying its position as a state-of-the-art solution in both class-based and unified setups.

S2.2. VisA Results

Table S6 details the performance of various methods on the VisA dataset, another prominent benchmark for anomaly detection and localization. Similar to Table S5, methods are divided into class-based and unified multi-class approaches, with I-AUROC and P-AUROC metrics used for evaluation. The dataset includes unique classes like PCB variants, macaroni shapes, and food items, providing a challenging testbed for models. LASB, the proposed method, achieves competitive results, outperforming many state-of-the-art techniques in both class-based (I-AUROC: 99.4, P-AUROC: 99.3) and unified scenarios (I-AUROC: 94.2, P-AUROC: 98.18).

The table S6 highlights LASB’s strong generalization across diverse classes, particularly in the class-based setting. Notably, LASB achieves near-perfect scores in most PCB and macaroni categories, reflecting its robustness in detecting subtle, class-specific anomalies. Compared to other advanced methods like OmniaL [22] and TransFusion [3], LASB demonstrates superior localization accuracy (P-AUROC), especially in complex multi-class setups. This reinforces the efficacy of diffusion-based approaches in modeling intricate object features, making LASB a leading choice for anomaly detection tasks in industrial and real-world applications.

S3. Sampling Stability

In the field of anomaly detection using diffusion models, achieving consistent outcomes during the sampling or inference stage is notably challenging due to the inherent stochastic nature of generative processes. This inconsistency can be particularly problematic in real-world applica-

Category	DRAEM [19]	UniAD [18]	DDPM [7]	LDM [12]	DiAD [6]	DiffAD [21]	LASB (ours)
pcb1	94.6/31.8/37.2	93.3/3.9/8.3	75.7/1.1/2.8	84.5/2.1/4.9	98.7/49.6/52.8	72.5	99.4/52.1/56.1
pcb2	92.3/10.0/18.6	93.9/4.2/9.2	76.2/0.7/1.6	89.5/2.5/6.7	95.2/7.5/16.7	69.9	98.4/33.4/48.3
pcb3	90.8/14.8/24.4	97.3/13.8/21.9	83.3/1.4/3.2	94.4/9.4/17.4	96.7/8.0/18.4	72.4	98.6/35.2/47.1
pcb4	94.4/31.0/37.5	99.4/14.7/22.9	70.1/4.3/5.8	80.4/2.2/4.2	97.0/17.6/27.2	69.2	99.2/44.3/45.2
macaroni1	95.0/19.1/24.1	97.4/3.7/9.7	87.0/4.1/10.4	81.6/0.3/1.3	94.1/10.2/16.7	73.2	97.5/39.2/45.4
macaroni2	94.6/3.9/12.4	95.2/0.9/4.3	84.7/3.0/7.1	87.2/0.3/3.9	96.8/3.1/9.3	68.8	98.7/41.7/42.3
capsules	97.1/27.3/37.7	88.7/3.3/7.4	77.1/1.1/2.1	75.5/1.1/2.7	97.3/10.2/21.0	70.2	99.7/36.2/46.1
candle	82.2/10.1/19.0	98.5/17.6/27.9	76.4/0.7/1.4	85.3/0.9/1.9	97.3/12.3/22.7	71.0	98.5/32.4/46.4
cashew	80.7/9.9/17.5	98.6/15.7/27.9	74.5/1.1/2.5	90.5/1.6/3.1	99.0/53.1/60.9	69.3	98.4/66.3/67.1
chewinggum	91.0/62.3/63.3	98.5/5.4/9.0	74.7/1.4/3.0	90.8/5.5/11.4	99.5/4.7/15.6	69.6	98.7/39.5/48.2
fryum	92.4/38.8/38.5	95.9/34.0/40.0	77.4/9.1/14.8	89.9/14.8/24.8	97.5/9.0/17.0	73.5	98.5/66.4/67.1
pipe_fryum	91.1/38.1/39.6	98.9/50.2/57.7	87.0/6.9/12.9	96.4/31.0/37.2	99.4/72.7/69.9	71.2	98.8/78.6/71.8
Mean	91.3/23.5/29.5	95.9/21.0/27.0	79.7/2.2/4.5	86.6/6.0/9.9	96.0/26.1/33.0	71.2/-	98.2/46.4/52.6

Table S3. Results on VisA dataset for multi-class anomaly localization with AUROC_{seg}/AP_{seg}/F1max_{seg} metrics

Category	DRAEM [19]	UniAD [18]	DDPM [7]	LDM [12]	DiAD [6]	DiffAD [21]	LASB (ours)
pcb1	71.9/72.2/70.0	92.8/92.7/87.8	54.1/47.7/67.1	51.2/46.9/66.8	88.1/88.7/80.7	75/-	91.2/90.4/92.3
pcb2	78.4/78.2/76.2	87.8/87.7/83.1	50.8/48.5/66.7	57.0/63.4/67.5	91.4/91.4/84.7	94.6/-	95.2/92.1/94.5
pcb3	76.6/77.4/74.7	78.6/76.8/76.1	53.4/51.2/60.8	62.7/69.6/72.0	86.2/87.6/77.6	94.7/-	93.6/89.6/93.7
pcb4	97.3/97.4/95.9	98.8/98.8/97.5	56.0/49.4/66.6	54.4/47.1/66.0	99.6/99.5/97.6	97.5/-	99.1/98.8/98.2
macaroni1	69.8/68.5/70.9	79.9/79.8/72.7	50.9/55.1/68.0	56.2/49.6/64.8	85.7/85.2/78.8	87.6/-	91.5/88.6/93.3
macaroni2	59.4/60.7/68.0	71.6/71.6/69.9	54.4/51.8/67.1	56.8/52.7/66.6	62.5/57.4/69.6	88.6/-	91.8/79.7/74.6
capsules	83.4/91.1/82.1	55.6/55.6/76.9	58.9/62.7/78.2	57.7/71.4/77.3	83.9/82.9/87.6	94.4/-	89.7/82.5/97.4
candle	69.3/73.9/68.0	94.1/94.0/86.1	52.1/48.3/66.6	50.4/52.2/68.2	96.2/90.2/87.6	94.4/-	94.8/93.3/94.6
cashew	81.7/89.7/87.3	92.8/92.6/91.4	63.5/78.9/86.0	61.1/71.0/87.1	95.1/95.7/98.6	81.4/-	93.6/96.6/93.8
chewinggum	93.7/97.1/91.0	96.3/96.1/90.4	50.9/65.6/80.0	53.9/65.5/81.4	99.4/99.5/95.9	94.1/-	98.8/98.9/94.4
fryum	89.1/95.0/86.6	83.0/83.4/85.0	51.0/62.0/80.4	63.7/71.5/81.3	95.9/95.8/87.2	87.1/-	94.3/96.5/94.5
pipe_fryum	82.8/91.2/83.9	94.7/94.7/93.9	56.9/74.9/80.8	56.1/75.5/80.3	92.8/91.9/93.7	92.7/-	97.7/99.4/94.4
Mean	79.1/81.9/78.9	85.5/85.5/84.4	54.5/57.9/72.3	56.7/61.4/73.1	86.8/88.3/85.1	89.79/-	94.2/92.2/94.5

Table S4. Results on VisA dataset for multi-class anomaly detection on the AUROC_{cls}/AP_{cls}/F1max_{cls} metrics

Method	Venue	Model Method	Bottle	Cable	Capsule	Hazelnut	Metal Nut	Pill	Screw	Toothbrush	Transistor	Zipper	Carpet	Grid	Leather	Tile	Wood	Avg
Class-based																		
SimpleNet [9]	CVPR2023	Embedding-based	100/98.0	99.9/97.6	97.7/98.9	100/97.9	100/98.8	99.0/98.6	98.2/99.3	99.7/98.5	100/97.6	99.9/98.9	99.7/98.2	99.7/98.8	100/99.2	99.8/97.0	100/94.5	99.6/98.1
PatchCore [13]	CVPR2022	Embedding-based	100/98.6	99.5/98.4	98.1/98.8	100/98.7	100/98.4	99.8/98.9	98.1/99.4	100/98.7	100/96.3	99.4/98.8	98.7/98.9	99.7/98.3	100/99.3	100/99.3	99.2/95.0	99.1/98.1
DSR [20]	ECCV2022	-	100/91.5	98.1/53.3	100/68	100/62.5	97.5/65.7	100/93.9	97.8/41.1	100/78.5	93.8/70.4	100/78.2	95.6/87.3	98.5/67.5	96.2/52.5	99.7/74.2	96.3/68.4	98.2/70.2
PaDiM [1]	ICPR 2021	Memory Bank	98.3/94.8	96.7/88.8	98.5/93.5	98.2/92.6	97.2/85.6	95.7/92.7	98.5/94.4	98.8/93.1	97.5/84.5	98.5/95.9	99.1/96.2	97.3/94.6	99.2/97.8	94.1/86.0	94.9/91.1	97.5/92.1
CS-Flow [14]	WACV2022	Normalization Flow	100/85.3	97.0/86.3	90.7/88.2	99.5/89.5	97.2/93.2	95.3/92.0	89.6/77.9	96.9/89.2	98.7/96.0	98.7/80.3	99.5/84.7	98.2/74.6	100/85.1	97.2/77.5	97.2/84.5	97.2/84.5
CFLow-AD [4]	WACV 2022	Normalization Flow	100/98.1	89.3/95.5	94.5/98.8	100/99.0	99.5/97.9	92.4/98.3	90.8/97.9	89.7/98.5	94.3/89.7	98.4/98.0	98.6/98.6	96.2/96.8	100/99.3	99.9/96.8	99.3/92.4	96.2/97.1
OCR-GAN [8]	TIP 2023	GAN	99.6/-	99.1/-	96.2/-	98.5/-	99.5/-	98.3/-	100/-	98.7/-	98.3/-	99.0/-	99.4/-	99.6/-	97.1/-	95.5/-	95.7/-	98.3/-
RD4AD [2]	CVPR2022	Embedding-based	99.6/97.8	84.1/85.1	94.1/98.8	60.8/97.9	100/93.8	97.5/97.5	97.7/99.4	97.2/99	94.2/85.9	99.5/98.5	98.5/99	98/99.2	100/99.3	98.3/95.3	99.2/95.3	98.5/97.8
DRAEM [19]	ICCV 2021	AutoEncoder	97.5/87.6	57.8/71.3	65.3/50.5	93.7/96.9	72.8/62.2	82.2/94.4	92/95.5	90.6/97.7	74.8/64.5	98.8/98.3	98/98.6	99.3/98.7	98.7/97.3	99.8/98	98.9/86	88.1/87.2
ADSPR [15]	ICCV 2023	Diffusion	99.9/95.9	94.4/96.9	97.3/96.6	98.3/98.7	96.7/96.6	95.4/98.2	98.6/98.5	98.6/98.7	91.5/96.4	100/98.9	99.9/98.8	99.1/99.9	99.8/99.3	99.8/96.8	96.1/95.4	97.6/97.36
DiffAD [21]	ICCV 2023	Diffusion	100/98.8	94.6/96.8	97.5/98.2	100/99.4	99.5/99.1	97.7/97.7	97.2/99	100/99.2	96.1/93.7	100/99	98.3/98.1	100/99.7	100/99.1	100/99.4	100/96.7	98.7/98.26
D3AD [16]	CVPRW 2023	Diffusion	100/98.6	97.8/93.3	96.6/97.9	98/98.8	98.9/96.1	99.2/98.2	83.9/99	100/99	96.8/95.6	98.2/98.3	94.2/97.6	100/99.2	98.5/94.4	95.5/94.7	99.7/95.9	97.15/97.44
DDAD [11]	Arxiv 2023	Diffusion	100/98.2	100/98.2	99.4/95.7	100/98.4	100/99	100/99.1	99/99.3	100/98.7	100/95.3	100/98.2	99.3/98.7	100/99.4	100/99.4	100/98.2	100/95	99.84/98.05
TransFusion [3]	ECCV 2024	Diffusion	100/97.3	97.9/85.5	98.5/92.1	100/97.7	100/94.1	98.3/96.2	97.2/97	100/94.1	98.3/83.9	100/97.2	99.2/95.9	100/98	100/96.2	99.8/95	99.4/94.8	99.24/94.33
LASB (Ours)	-	Diffusion	100/99.5	99.2/98.5	99.4/99.1	100/99.3	99.8/99.2	99.1/99.2	98.7/99.2	99.8/99.2	99.8/98.7	100/99.4	99.7/99.4	99.9/99.3	99.9/99.6	99.8/99.5	99.9/98.2	99.66/99.15
Unified (multi-class)																		
HVQ-Trans [10]	NeurIPS 2023	Non-Diffusion	100/98.3	99.0/98.1	95.4/98.8	100/98.8	99.9/96.3	95.8/97.1	95.6/98.9	93.6/98.6	99.7/97.9	97.9/97.5	99.9/98.7	99.0/97.0	100/98.8	99.2/92.2	97.2/92.4	98.0/97.3
MambaAD [5]	NeurIPS 2024	Non-Diffusion	100/98.8	98.8/95.8	94.4/98.4	100/99.0	99.9/96.7	97.0/97.4	94.7/99.5	98.3/99.0	99.3/98.4	99.8/99.2	100/99.2	100/99.4	98.2/93.8	98.8/94.4	98.6/97.7	98.6/97.7
OmniAL [22]	CVPR2023	Non-Diffusion	100/99.2	98.2/97.3	95.2/96.9	95.6/98.4	99.2/99.1	97.2/98.8	88.0/98.0	100/99.4	93.8/93.3	100/99.5	98.7/99.4	99.0/99.3	99.6/99.0	93.2/97.4	97.2/98.3	97.2/98.3
LAD [17]	ECCV 2024	Diffusion	98.1/98.4	97.5/97.7	98.4/98.2	97.2/96.7	98.1/97.6	97.4/96.8	98.4/97.2	98.4/97.2	97.6/95.3	95.2/96.3	98.3/97.5	97.6/98.3	98.2/98.1	97.7/97.9	98.6/97.6	97.5/97.4
UniAD [18]	NeurIPS 2024	-	99.7/98.1	95.5/97.3	88.1/98.5	99.9/98.1	97/94.8	94/95	88.8/98.3	95.8/98.4	99.7/97.9	94.9/96.8	99.7/98.5	97/96.5	100/98.8	99.3/91.8	98.5/93.2	96.5/96.8
DIAD [6]	AAAI 2022	Diffusion	99.7/98.4	94.8/96.8	89/97.1	99.5/98.3	99.1/97.3	95.7/95.7	99.7/99.9	99.7/99.9	98.8/95.1	95.1/96.2	99.4/98.6	98.5/96.6	96.8/92.4	99.7/93.3	97.15/96.76	97.15/96.76
LASB (Ours)	-	Diffusion	99.8/99.5	96.5/96.6	98.3/98.3	99.8/99.3	99.7/99.2	99.1/98.1	96.7/98.9	99.8/99.3	99.6/97.3	99.9/99.3	99.3/98.8	99.7/99.2	99.8/99.3	99.4/99.4	99.8/97.4	99.14/98.66

Table S5. MVTec-AD dataset results for class-based and unified Anomaly Detection/Localization (I-AUROC / P-AUROC).

tions where reliable and stable detection is critical. Therefore, assessing the stability of model outputs across multiple inferences is essential. Our approach involves training the model once and then conducting multiple sampling or inference tests to evaluate if the outcomes remain consistent over time.

Our findings, as detailed in Table S1, S2, reveal that the LASB model exhibits strong stability across all evaluation metrics, showing negligible variance across numerous inferences. This consistency is attributed to the model’s capability to maintain structural integrity and effectively transition from anomalous to normal latent spaces. The LASB model is designed to reconstruct a normal image regard-

less of the underlying anomalies, compelling it to disregard anomalous features during reconstruction. This process not only ensures that anomalies of various patterns, sizes, and orientations are effectively handled but also enhances the overall reliability of the model. Consequently, LASB’s performance surpasses existing SOTA methods, establishing it as a robust solution for practical deployment in anomaly detection scenarios.

S4. Hyperparameters for Our LASB Model

The proposed LASB model employs a set of hyperparameters specifically configured to optimize its performance in anomaly detection tasks. One of the foundational settings

Method	Venue	Model Method	pcb1	pcb2	pcb3	pcb4	macaroni1	macaroni2	capsules	candle	cashew	chewinggum	fryum	pipe.fryum	Avg
Class-based															
SimpleNet [9]	CVPR2023	Embedding	99.2/99.8	99.2/98.8	98.6/99.2	98.9/98.6	97.6/99.6	83.4/96.4	89.5/99.2	96.9/98.6	94.8/99.0	100/98.5	96.6/94.5	99.2/99.3	96.2/98.5
PatchCore [13]	CVPR2022	Embedding	96.0/99.8	95.1/98.4	93.0/98.9	99.5/98.3	90.1/98.5	63.4/93.5	68.8/96.5	98.7/99.2	97.7/99.2	99.1/98.9	91.6/95.9	99.0/99.3	91.0/98.1
DSR [20]	ECCV2022	-	-	-	-	-	-	-	-	-	-	-	-	-	-
PaDiM [11]	ICPR2021	Memory Bank	94.7/91.3	88.5/88.7	91.0/84.9	97.5/81.6	87.0/92.1	70.5/75.4	70.7/76.9	91.6/95.7	93.0/87.9	98.8/83.5	88.6/80.2	97.0/92.5	89.1/85.9
CS-Flow [14]	WACV2022	Normalization Flow	-	-	-	-	-	-	-	-	-	-	-	-	-
CFLOW-AD[4]	WACV2022	Normalization Flow	-	-	-	-	-	-	-	-	-	-	-	-	-
OCR-GAN [8]	TIP2023	GAN	96.1/-	98.3/-	98.1/-	99.7/-	97.2/-	95.1/-	98.8/-	98.9/-	97.4/-	99.4/-	96.3/-	99.7/-	97.9/-
RD4AD [2]	CVPR2022	Embedding-based	97.2/98.3	96.5/99.3	99.4/98.2	98.7/99.3	91.4/99.1	91.8/99.4	96.2/98.9	98.7/94.4	99.3/97.6	96.9/96.4	99.6/99.1	96.9/98.3	96.7/99.6
DRAEM [19]	ICCV2021	Augmentation	71.9/94.6	78.4/92.3	76.6/90.8	97.3/94.4	69.8/95.0	59.4/94.6	83.4/97.1	69.3/82.2	81.7/80.7	93.7/91.0	89.1/92.4	82.8/91.1	79.1/91.3
ADSPR [15]	ICCV2023	Diffusion	-	-	-	-	-	-	-	-	-	-	-	-	-
DiffAD [21]	ICCV2023	Diffusion	75.0/72.5	94.6/69.9	94.7/72.4	97.7/69.2	87.6/73.2	90.7/68.8	87.6/70.2	94.4/71.3	81.4/72.4	94.0/69.6	87.1/73.5	92.7/71.2	89.7/91.18
D3AD [16]	CVPR2024	Diffusion	92.5/98.3	98.3/94.0	97.4/94.2	99.8/86.4	94.3/99.3	92.5/98.3	88.5/95.7	95.6/92.7	94.2/89.4	99.7/94.1	96.5/91.7	96.9/97.2	95.51/94.27
DDAD [11]	ArXiv2023	Diffusion	100/93.4	99.7/97.4	97.2/96.3	100/98.5	99.2/98.7	99.2/98.2	100/99.5	99.9/98.7	94.5/97.4	98.1/96.5	99.0/96.9	100/99.5	98.9/97.58
TransFusion [3]	ECCV2024	Diffusion	98.9/92.4	99.7/85.1	99.2/92.0	99.6/89.4	99.4/94.0	96.5/95.6	99.6/97.3	98.3/88.6	93.7/82.8	99.6/83.2	98.3/77.8	99.6/87.9	98.53/88.84
GLASS	ECCV2024	Diffusion	98.7/99.4	98.4/98.3	99.5/99.3	99.1/98.6	100/99.8	96.2/99.7	96.8/99.3	99.3/99.2	98.9/98.9	100/99.4	99.0/94.8	100/99.1	98.82/98.81
LASB (Ours)	-	Diffusion	98.4/99.3	98.2/98.8	98.6/98.9	99.4/99.3	96.8/98.4	97.2/99.4	98.7/99.1	99.2/99.3	98.4/99.2	99.8/98.9	98.2/98.9	99.4/99.3	98.52/99.06
Unified (multi-class)															
HVQ-Trans [10]	NeurIPS 2023	Non-Diffusion	96.7/99.4	93.4/98.0	92.0/98.3	99.5/97.7	93.1/99.4	86.2/98.5	77.1/99.0	96.8/99.2	94.9/99.2	99.4/98.8	90.4/97.7	98.5/99.4	93.2/98.7
MambaAD [5]	NeurIPS 2024	Non-Diffusion	95.4/99.8	94.2/98.9	93.7/99.1	99.9/98.6	91.6/99.5	81.6/99.5	91.8/99.1	96.8/99.0	94.5/94.3	97.7/98.1	95.2/96.9	98.7/99.1	94.3/98.5
OmniAL [22]	CVPR2023	Non-Diffusion	96.6/98.7	99.4/83.2	96.9/98.4	97.4/98.5	96.9/98.9	89.9/99.1	87.9/98.6	85.1/90.5	97.1/98.9	94.9/98.7	97.0/89.3	91.4/99.1	94.2/96.0
GLAD [17]	ECCV 2024	Diffusion	92.6/98.3	91.8/97.8	92.3/98.9	91.7/98.8	92.5/98.6	92.2/97.4	89.8/98.6	90.6/93.6	92.8/96.8	91.6/98.4	92.3/97.6	91.4/98.8	91.8/97.8
UniAD [18]	NeurIPS 2022	Transformer	92.8/93.3	87.8/93.9	78.6/97.3	98.8/94.9	79.9/97.4	71.6/95.2	55.6/88.7	94.1/98.5	92.8/98.6	96.3/98.5	83/95.9	94.7/98.9	85.5/95.925
DIAD [6]	AAAI 2024	Diffusion	88.1/98.7	91.4/95.2	86.2/96.7	99.6/97	85.7/94.1	62.5/93.6	58.2/97.3	92.8/97.3	91.5/90.9	99.1/94.7	89.8/97.6	96.2/99.4	86.75/96.04
LASB (Ours)	-	Diffusion	91.2/99.4	95.2/98.7	93.6/98.6	99.1/99.2	91.5/97.5	91.8/98.7	89.7/97.9	94.8/98.5	93.6/94.6	98.8/97.8	93.4/98.5	97.7/98.8	94.2/98.18

Table S6. VisA dataset dataset for class-based and multi-class Anomaly Detection / Localization (I-AUROC / P-AUROC).

is the shape of the latent space vector, z , which is defined as $64 \times 64 \times 3$. This configuration supports the model’s ability to encode detailed features and textures at a manageable resolution, suitable for capturing nuances in data. The model operates over a diffusion process with 1000 steps, denoted as T , which is indicative of a gradual and detailed transformation process, ensuring that the model effectively captures the data distribution.

Another crucial aspect of the LASB model is its sampling strategy during the denoising diffusion probabilistic modeling phase. It uses 100 DDPM sampling steps [20], a considerably lower number than the diffusion steps, which balances computational efficiency with the quality of sample generation. The variance schedule for the diffusion process is tightly controlled within a range of $[0.1, 0.3]$ for β , which regulates the noise level added during the diffusion steps, optimizing the balance between detail preservation and noise introduction.

Parameters Name	Model Name	
	Latent Denoising	Autoencoder
z shape	$64 \times 64 \times 3$	-
Diffusion steps T	1000	-
DDPM sampling steps T	100	-
β [min, max]	[0.1, 0.3]	-
Model input shape	$64 \times 64 \times 3$	$256 \times 256 \times 3$
Embed dim	-	3
Channels	224	128
Num res blocks	2	2
Channel Multiplier	1,4,8	1,2,4
Attention resolutions	32,16,8	-
Num Heads	8	-
Batch Size	8	4
Iterations	500000	50000
Learning Rate	$5e-5$	$4.5e-06$

Table S7. Hyperparameters for our LASB model

In terms of architecture, the model inputs are images of size $64 \times 64 \times 3$, aligning with the latent vector’s dimensions to maintain consistency in processing scales. The architecture details include 320 channels across the network with 2 residual blocks, indicating a relatively deep network capable of complex feature extraction. The channel multiplier is set at progressive stages of 1, 4, and 8, which increases the network’s capacity as it goes deeper, allowing for a more refined feature hierarchy. Attention mechanisms are employed at resolutions 32, 16, and 8, which focus the model’s capacity on important features at different scales, essential for effective anomaly detection in varied contexts. The model is equipped with 8 heads in its multi-head attention layers, facilitating parallel attention across different representation subspaces, enhancing its ability to capture diverse patterns and anomalies.

S5. LASB Robustness to Transformations

As proposed approach translates anomalies into clean images, allowing LASB to unlearn anomalies while preserving structural details unlike diffusion models that rely on Gaussian noise and often degrade fine features. To demonstrate LASB robustness, we applied transformation based corruptions, including lighting variations (random brightness/contrast), pose variations (random flip/rotation), and unaltered conditions.

Class-Based			
Dataset	Lighting Variation	Pose Variation	W/o Variations
MVTec-AD	97.36 / 97.89	97.15 / 97.55	99.66 / 99.15
VisA	96.84 / 97.21	96.42 / 97.18	98.52 / 99.06
Multi-Class Based			
Dataset	Lighting Variation	Pose Variation	W/o Variations
MVTec-AD	97.22 / 97.42	96.92 / 97.14	99.14 / 98.66
VisA	92.82 / 96.52	92.42 / 96.24	94.2 / 98.18

Table S8. Robustness to various transformations

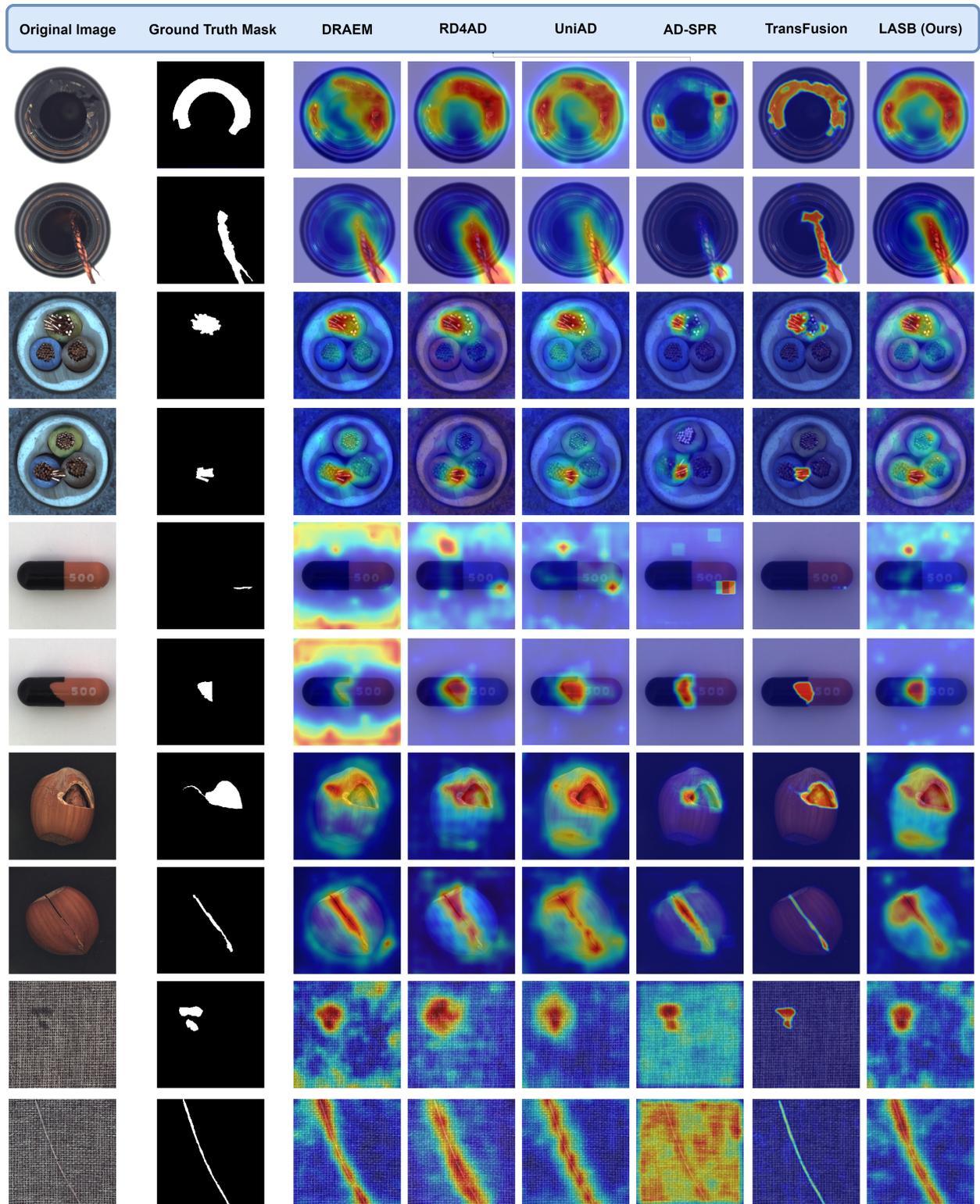


Figure S1. Visual representation of test samples from the MVTEC dataset for anomaly localization, depicted through heatmaps for various models for different categories.

As evidenced by the results in Table S8, LASB maintains high detection accuracy even under significant image transformations. The minimal performance degradation across different corruption types highlights LASB model exceptional generalization capabilities and practical applicability in real-world scenarios where lighting and pose variations are common. This resilience to transformations is particularly valuable in industrial inspection settings where environmental conditions cannot always be perfectly controlled.

S6. Visualizations

The LASB model distinguishes itself through its unified approach to anomaly detection, capable of reconstructing normal images across various classes without discrimination. This characteristic is pivotal as it ensures that the model is not biased towards any specific type of normalcy but is robust in identifying deviations from a completely unseen samples across diverse datasets. Such a unified nature is crucial in practical scenarios where anomalies can vastly differ in appearance and context, requiring a versatile system that maintains high performance without the need for class-specific tuning. To further justify the performance we illustrate the effectiveness of the LASB model in highlighting anomalous regions by employing heat-map visualizations in Figure S1. These visualizations are instrumental in providing intuitive, visual feedback about the regions the model identifies as anomalous. By overlaying heat-maps on the original images, one can visually assess the precision with which the model detects and localizes irregularities. This method not only confirms the model’s capability to discern subtle and overt anomalies but also helps in verifying the performance of its localization in a transparent manner. Furthermore, the use of heat-map visualizations aids in refining the model by providing direct insights into its operational dynamics to understand the model’s sensitivity and specificity across different conditions and settings.

S7. Limitations

While the LASB model demonstrates state-of-the-art performance in anomaly detection and localization tasks, it persists some limitations which we leave to future work. Firstly, the method relies on a fine-tuned VQ-VAE model for latent space compression, which may introduce domain-specific biases if the encoder-decoder architecture is not adequately trained for diverse datasets and samples. Secondly, the LASB’s design is optimized for high-resolution datasets like MVTec-AD and VisA, and its performance on low-resolution, noisy, or highly imbalanced datasets remains less explored. Additionally, while the model significantly reduces computational overhead compared to pixel-space counterparts, the inference process still demands substantial resources when scaled to ultra-large datasets or real-time

streaming scenarios. Finally, although LASB is designed to operate in a unified framework, handling edge cases involving extreme variations in anomaly size, shape, or appearance may require further refinement in the structural reconstruction capabilities of the latent space diffusion process and need much challenging datasets to solve.

S8. Broader Impact

The LASB model has the potential to make a significant impact across a range of critical domains, from industrial quality control to medical diagnostics and beyond. In industrial settings, LASB can improve production efficiency by reliably detecting minute anomalies in high-resolution images, ensuring stringent quality standards while reducing human inspection costs. In healthcare, LASB’s ability to generalize across diverse anomaly types could lead to breakthroughs in detecting rare or subtle pathological conditions, thereby assisting medical professionals in timely and accurate diagnoses. Furthermore, the model’s inherent scalability and computational efficiency align with global efforts toward more sustainable AI, enabling its application in resource-constrained environments. However, as with any AI model, ethical considerations such as unintended biases, the potential for misuse in surveillance applications, and the need for transparency in critical decision-making systems must be actively addressed to ensure equitable and responsible deployment.

References

- [1] Thomas Defard, Aleksandr Setkov, Angélique 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. 3, 4
- [2] Hanqiu Deng and Xingyu Li. Anomaly detection via reverse distillation from one-class embedding. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9737–9746, 2022. 3, 4
- [3] Matic Fučka, Vitjan Zavrtanik, and Danijel Skočaj. Transfusion—a transparency-based diffusion model for anomaly detection. In *European conference on computer vision*, pages 91–108. Springer, 2024. 2, 3, 4
- [4] Denis Gudovskiy, Shun Ishizaka, and Kazuki Kozuka. Cflow-ad: Real-time unsupervised anomaly detection with localization via conditional normalizing flows. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 98–107, 2022. 3, 4
- [5] Haoyang He, Yuhu Bai, Jiangning Zhang, Qingdong He, Hongxu Chen, Zhenye Gan, Chengjie Wang, Xiangtai Li, Guanzhong Tian, and Lei Xie. MambaAD: Exploring state space models for multi-class unsupervised anomaly detection. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024. 3, 4

- [6] Haoyang He, Jiangning Zhang, Hongxu Chen, Xuhai Chen, Zhishan Li, Xu Chen, Yabiao Wang, Chengjie Wang, and Lei Xie. A diffusion-based framework for multi-class anomaly detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 8472–8480, 2024. 1, 2, 3, 4
- [7] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020. 3
- [8] Yufei Liang et al. Omni-frequency channel-selection representations for unsupervised anomaly detection. *IEEE Transactions on Image Processing*, 2023. 3, 4
- [9] 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. 3, 4
- [10] Ruiying Lu, YuJie Wu, Long Tian, Dongsheng Wang, Bo Chen, Xiyang Liu, and Ruimin Hu. Hierarchical vector quantized transformer for multi-class unsupervised anomaly detection. *Advances in Neural Information Processing Systems*, 36:8487–8500, 2023. 3, 4
- [11] Arian Mousakhan, Thomas Brox, and Jawad Tayyub. Anomaly detection with conditioned denoising diffusion models. *arXiv preprint arXiv:2305.15956*, 2023. 3, 4
- [12] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10684–10695, 2022. 3
- [13] 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. 3, 4
- [14] Marco Rudolph, Tom Wehrbein, Bodo Rosenhahn, and Bastian Wandt. Fully convolutional cross-scale-flows for image-based defect detection. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1088–1097, 2022. 3, 4
- [15] Woosang Shin, Jonghyeon Lee, Taehan Lee, Sangmoon Lee, and Jong Pil Yun. Anomaly detection using score-based perturbation resilience. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 23372–23382, 2023. 3, 4
- [16] Justin Tebbe and Jawad Tayyub. Dynamic addition of noise in a diffusion model for anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 3940–3949, 2024. 2, 3, 4
- [17] Hang Yao, Ming Liu, Zhicun Yin, Zifei Yan, Xiaopeng Hong, and Wangmeng Zuo. Glad: towards better reconstruction with global and local adaptive diffusion models for unsupervised anomaly detection. In *European Conference on Computer Vision*, pages 1–17. Springer, 2024. 3, 4
- [18] Zhiyuan You, Lei Cui, Yujun Shen, Kai Yang, Xin Lu, Yu Zheng, and Xinyi Le. A unified model for multi-class anomaly detection. *Advances in Neural Information Processing Systems*, 35:4571–4584, 2022. 3, 4
- [19] Vitjan Zavrtanik, Matej Kristan, and Danijel Skočaj. Draem—a discriminatively trained reconstruction embedding for surface anomaly detection. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 8330–8339, 2021. 3, 4
- [20] Vitjan Zavrtanik, Matej Kristan, and Danijel Skočaj. Dsr—a dual subspace re-projection network for surface anomaly detection. In *European conference on computer vision*. Springer Nature Switzerland, 2022. 3, 4
- [21] Xinyi Zhang, Naiqi Li, Jiawei Li, Tao Dai, Yong Jiang, and Shu-Tao Xia. Unsupervised surface anomaly detection with diffusion probabilistic model. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6782–6791, 2023. 3, 4
- [22] Ying Zhao. Omnia: A unified cnn framework for unsupervised anomaly localization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3924–3933, 2023. 2, 3, 4