SAC-Diff: A Scan-Aware Consistency-Enhanced Diffusion Framework for Unsupervised Chest CT Anomaly Detection

Xinyuan Zheng[1,2] XINYUAN.ZHENG@YALE.EDU
Yoshihisa Shinagawa[1] YOSHIHISA.SHINAGAWA@SIEMENS-HEALTHINEERS.COM
Sepehr Farhand[1] SEPEHR.FARHAND@SIEMENS-HEALTHINEERS.COM
Chi Liu[2] CHI.LIU@YALE.EDU
Gerardo Hermosillo[1] GERARDO.HERMOSILLOVALADEZ@SIEMENS-HEALTHINEERS.COM
Xueqi Guo[1] XUEQI.GUO@SIEMENS-HEALTHINEERS.COM
[1] *Siemens Medical Solutions USA, Inc., Malvern, PA, USA*
[2] *Department of Biomedical Engineering, Yale University, New Haven, CT, USA*

Editors: Under Review for MIDL 2026

Abstract

Anomaly detection in medical imaging is important but challenging due to diverse and imbalanced pathologies. Supervised methods rely on large annotated datasets and generalize poorly to unseen conditions. Unsupervised generative methods, especially diffusion models, can learn normal anatomy and detect outliers, but often hallucinate because of the Gaussian noise design and insufficient anatomical guidance. To address these challenges, we propose **SAC-Diff**, a **S**can-**A**ware **C**onsistency-Enhanced **Diff**usion framework for unsupervised anomaly detection in automated lung disease screening using chest CT. SAC-Diff adopts simplex noise for detail-preserving diffusion perturbation, integrates scan awareness via (A) subject-aware anatomical priors into conditional diffusion and (B) background-aware masking for scan-specific variations and heterogeneous lung anomalies, and enhances robustness by enforcing consistency and quantifying uncertainty through multi-sample ensembling. We evaluate SAC-Diff on two diseased datasets with various anomalies, COVID-19 and interstitial lung disease (ILD), and observe substantial improvements over prior methods. On COVID-19, SAC-Diff achieves an IoU of 0.39 (+3.75% improvement compared to existing methods) and Dice of 0.53 (+2.99%); on ILD, it improves IoU to 0.31 (+74.45%) and Dice to 0.44 (+60.40%). Our results demonstrate promise toward robust and annotation-free CT anomaly detection in hospital deployment.

Keywords: computer-aided diagnosis, unsupervised anomaly detection, CT

1. Introduction

Anomaly detection (Pang et al., 2021; Samariya and Thakkar, 2023; Liu et al., 2024) is a fundamental challenge due to the inherent scarcity of well-annotated anomalous data, particularly within the domain of medical imaging (Wolleb et al., 2022). This scarcity originates from the low prevalence of certain pathological conditions, the high cost of annotation that requires expert radiological input, and strict privacy restrictions that limit data availability. Consequently, training robust and generalizable models in this context remains a nontrivial task.

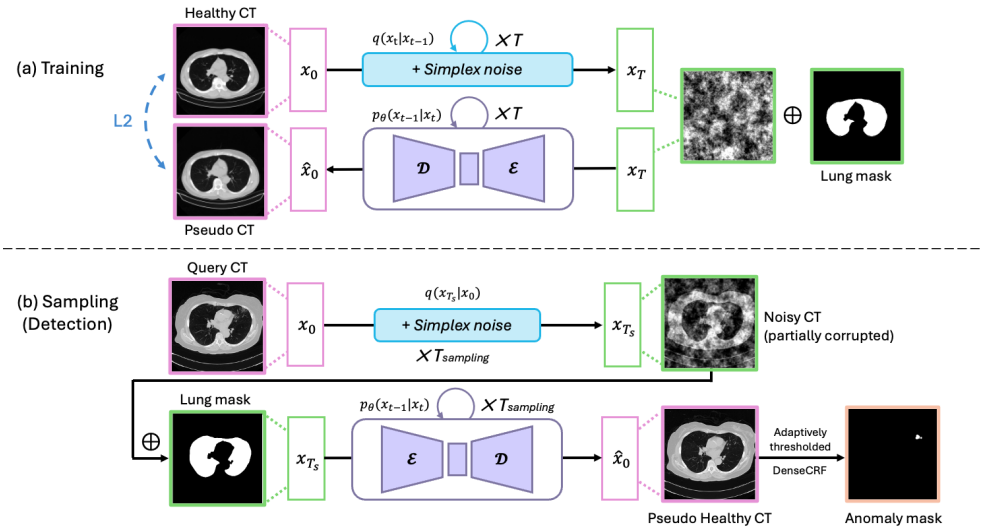

Figure 1: SAC-Diff training and detection workflow. (a) During training, healthy CT scans are diffused for T steps by adding simplex noise until a completely corrupted image is obtained. A pre-segmented lung mask is incorporated as anatomical guidance, and the base model (residual UNet) is then trained to denoise these corrupted scans and reconstruct pseudo-healthy CTs. (b) At inference time, query CT scans are diffused for T_s steps to obtain a partially corrupted representation. This noisy CT, together with the lung mask as anatomical guidance, is subsequently denoised by the trained network to generate a pseudo-healthy reconstruction. The anomaly mask is derived from the difference MSE map through a scan-aware background-adaptive masking strategy.

To address the scarcity of abnormal medical data, previous approaches involve artificially injecting lesion-like regions into normal images (Pezeshk et al., 2017; Salem et al., 2019; Huang et al., 2022) or synthesizing pathological data to increase the scale of training set (Mok and Chung, 2019; Abdelhalim et al., 2021a; Basaran et al., 2024; Abdelhalim et al., 2021b; Li et al., 2020). However, these approaches often fail to capture the complex anatomical variability present in real pathological cases. The heterogeneity of abnormal patterns and the imbalance between common and rare anomalies further complicate the task. In the context of chest CT, this challenge is amplified by the breadth of abnormalities that may appear, including but not limited to: asbestosis, bronchiectasis, pneumonia, fibrosis, abscess, ground-glass opacities, honeycombing, and various types of nodules (Brixey et al., 2024; Akira et al., 2003; Ma et al., 2022; Baratella et al., 2021; Chaganti et al., 2020a; Pu et al., 2021; Gaillandre et al., 2023). The diversity and subtlety of these patterns make it difficult to simulate or synthesize anomalies or collect training data containing various abnormal presentations. These limitations highlight the need for unsupervised frameworks that can learn robust representations of normal anatomy and identify deviations without relying on human annotations or synthetic anomalies.

Therefore, rather than synthesizing abnormal images which often result in unrealistic or oversimplified pathologies, some works (Wolleb et al., 2022; Wyatt et al., 2022; Cai et al., 2025; Bercea et al., 2025) have turned to using generative models to learn the distribution

of normal data. In this paradigm, models are trained to capture healthy anatomy, enabling the detection of any out-of-distribution anomalies at inference time as deviations from the learned manifold. This idea has been applied for autoencoders (Zimmerer et al., 2019; Zhou and Paffenroth, 2017) and GAN-based methods (Akcay et al., 2019; Schlegl et al., 2019) showing initial success on natural and medical images. However, these approaches are constrained by training instability, mode collapse, and difficulties in capturing high-fidelity structural details. The emergence of diffusion models has introduced a more stable and expressive generative framework. Following the seminal works of denoising diffusion probabilistic models (DDPMs) (Ho et al., 2020; Dhariwal and Nichol, 2021; Song et al., 2020), diffusion-based methods have increasingly been adopted for anomaly detection (He et al., 2024; Beizaee et al., 2025; Wyatt et al., 2022; Zhang et al., 2023; Yao et al., 2025; Yu et al., 2023) for high-fidelity normal image reconstruction. Notably, Wolleb et al. (2022) extended diffusion models to unsupervised anomaly detection in medical contexts. AnoDDPM (Wyatt et al., 2022) introduces simplex noise during forward diffusion to improve sensitivity in detecting low-frequency anomalies in brain MRI. THOR (Bercea et al., 2024) refines the reverse diffusion process by incorporating implicit guidance via intermediate anomaly maps. These methods (Wolleb et al., 2022; Wyatt et al., 2022; Bercea et al., 2024; Pinaya et al., 2022; Beizaee et al., 2025; Yu et al., 2023) focus on enhancing the quality of generated pseudo-healthy outputs.

Despite aforementioned benchmarks, most anomaly detection methods remain limited in scope, typically focusing on lesion or tumor detection, and fail to exploit anatomical and structural regularities inherent in medical images. Detection accuracy and sensitivity remain areas for improvement. In this work, we propose a diffusion-based framework with scan awareness and consistency enhancement for unsupervised anomaly detection to address these limitations. We summarize our main contributions as follows:

- Awareness of subject anatomy: We integrate a conditioning mechanism to preserve anatomical fidelity in the reconstruction of pseudo-healthy scan, reducing false positives related to incorrect organ shape or boundary.
- Awareness of foreground distinction: We introduce an adaptive masking strategy based on background statistics to binarize anomalies within the organ of interest (lung), accounting for scan-specific variations and heterogeneity of pathological patterns.
- Consistency enhancement via ensembling: We enhance detection performance and support uncertainty quantification by exploiting the inherent consistency of ensemble inferences from generative models, overcoming the limitation of single-sample predictions.
- Evaluation on heterogeneous anomalies: We validate on two clinically relevant datasets. The model generalizes across focal and diffuse patterns (nodules, abscesses, fibrosis, ground-glass opacities), demonstrating superior performance compared to existing methods.

2. Methodology

2.1. Proposed SAC-Diff

The overall workflow of proposed SAC-Diff is illustrated in Fig. 1 and the backbone model is detailed in Fig. 2. Our SAC-Diff is built upon DDPM (Ho et al., 2020), incorporating the simplex noise modifications proposed in Wyatt et al. (2022). We further introduce conditioning and ensembling strategies for scan-aware pseudo-healthy reconstruction and

consistency-enhanced detection. These adaptations enable the model to exploit anatomical context and improve robustness.

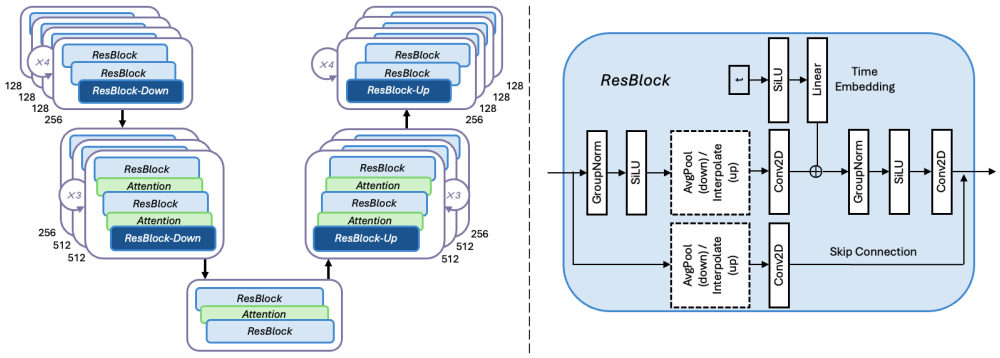

Figure 2: Backbone model of SAC-Diff. Numbers denote feature channels. The network follows a symmetric design with 7 downsampling and 7 upsampling modules (each with 3 residual blocks), connected by a bottleneck module with 2 residual blocks.

2.2. Simplex Noise

The Gaussian white noise used in standard diffusion models has a flat spectral density; however, natural images have been shown to have a power law distribution of frequencies (Ruderman, 1997). Assuming that normal and anomalous medical images follow similar power-law characteristics, using Gaussian noise can lead to disproportionate corruption. Low-frequency regions, such as large pathological structures, tend to remain relatively uncorrupted during the forward process and are therefore reconstructed in the reverse pass, reducing anomaly detection sensitivity. To address this, multi-octave simplex noise was introduced for spatially coherent perturbations with stronger low-frequency corruption (Wyatt et al., 2022). Following the approach of Wyatt et al. (2022), we apply multi-octave simplex noise during the forward diffusion process, using a starting frequency of $\nu = 2^{-6}$, an octave count of $N = 6$, and a frequency decay factor of $\gamma = 0.8$.

2.3. Subject-Anatomy-Aware Conditioning

Standard DDPMs are trained to model the unconditional distribution $p(x_0)$ of the data. In medical imaging applications, anatomical structures such as organs or tissue boundaries are known a priori and can serve as useful conditioning signals to guide the generative process. To incorporate such knowledge, we extend the formulation to a conditional generative model $p_\theta(x_{t-1} \mid x_t, c)$ where c denotes the auxiliary information, in our setting, a binary segmentation mask of the lung field acquired from a lightweight, pre-trained segmentation model (Chaganti et al., 2020b). Since only lung slices need to be processed, the segmentation is a necessary step in extracting the lung region and does not add overhead or introduce additional computational burden. We adopt an early fusion strategy, where the conditioning signal c is concatenated channel-wise with the noised image x_t as the input

to the denoising network at each step, formulated as $\hat{\epsilon}_\theta([x_t \oplus c], t)$. Since the conditioning mechanism is agnostic to disease labels, it generalizes naturally to any unseen pathologies.

Algorithm 1: Sampling (Detection)

Input: Query image x, condition c
$\epsilon \sim \text{Simplex}(\nu = 2^{-6}, N = 6, \gamma = 0.8)$
Construct noisy input $x_T = \sqrt{\bar{\alpha}_T}\, x + \sqrt{1 - \bar{\alpha}_T}\, \epsilon$
for $t = T, T-1, \ldots, 1$ **do**
  Predict noise: $\hat{\epsilon}_\theta([x_t \oplus c], t)$
  Compute mean $\mu_\theta(x_t, t, c) = \frac{1}{\sqrt{\alpha_t}} \left(x_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\, \hat{\epsilon}_\theta([x_t \oplus c], t) \right)$
  $z \sim \text{Simplex}(\nu = 2^{-6}, N = 6, \gamma = 0.8)$ if $t > 1$, else $z = 0$
  $x_{t-1} = \mu_\theta(x_t, t, c) + \sigma_t z$
end
return $\mathcal{E}(x) = \|x - x_0\|^2$

2.4. Background-Aware Adaptive Masking

During inference (see Alg. 1), we apply a partial forward diffusion ($T_{\text{sampling}} < T_{\text{training}}$) to each query sample from the abnormal datasets. This ensures that the corrupted image preserves anatomical information, while still introducing enough perturbation to enable effective reconstruction and detection. Following reconstruction, we compute the voxel-wise squared error between the original input x and the reconstructed sample x_0. The resulting error map $\mathcal{E}(x) = \|x - x_0\|^2$ serves as an initial estimation of the anomaly map, where larger values indicate deviations from the learned distribution of normal anatomy.

As the intensity range of the generated class activation maps varies significantly across heterogeneous anomalies and scan-specific characteristics (Guo et al., 2023), fixing a threshold for anomaly mask binarization is suboptimal. To address this, we propose an adjustable background-adaptive thresholding strategy to determine the cutoff used for binarizing anomaly masks. Specifically, we compute a volume-specific threshold α_a based on the statistics of the background region (i.e., voxels outside the lung field): $\alpha_a = \text{mean}\big(\mathcal{E}[c < 1]\big) + \lambda \cdot \text{std}\big(\mathcal{E}[c < 1]\big)$, where $\mathcal{E}$ denotes the predicted MSE map, c is the binary lung mask for the corresponding CT volume, $\mathcal{E}[c < 1]$ denotes $\mathcal{E}$ outside the lung region ($c < 1$), and λ is an adjustable parameter.

By changing λ, we can control how strict or lenient the threshold is for detecting anomalies based on how far a pixel's MSE deviates from the background distribution. The threshold α_a captures both the background bias and the noise level specific to each scan, which yields more robust and consistent segmentation of abnormal regions across subjects with varying intensity distributions. We further apply a fully connected conditional random field (Krähenbühl and Koltun, 2012) to enforce spatial consistency and obtain the final anomaly mask. Morphological operations (erosion followed by dilation) are used to remove noise and close small gaps in the detected regions.

2.5. Consistency-Enhanced Ensembling

Figure 3 illustrates our proposed consistency-enhanced ensembling strategy for inference-time detection. Specifically, multiple (K) pseudo-healthy reconstructions are averaged to

produce a stable anomaly estimate $\bar{x}_0 = \frac{1}{K}\sum_{k=1}^{K} x_0$, while their voxel-wise standard deviation $\sigma(x) = \left(\frac{1}{K}\sum_{k=1}^{K}(x_0^{(k)} - \bar{x}_0)^2\right)^{1/2}$ provides an estimate of epistemic uncertainty, offering insight into model confidence. Compared with the conventional single-sampling inference procedure, we exploit the inherent consistency of ensemble inferences from the generative model by aggregating multiple posterior samples. This approach improves robustness, reduces spurious predictions or hallucinations, and supports interpretability via voxel-wise uncertainty maps.

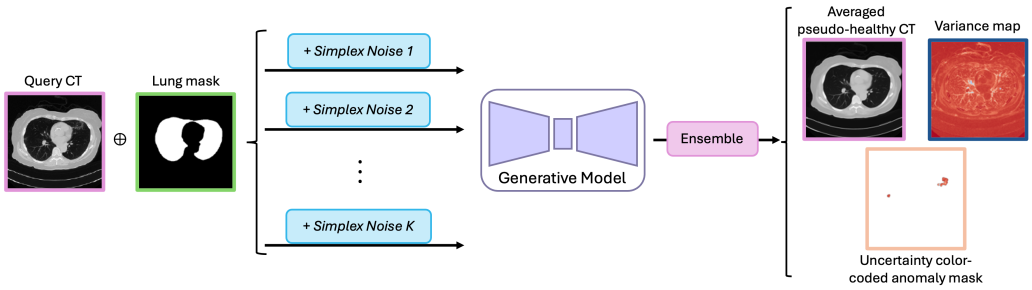

Figure 3: Consistency-enhanced ensembling inference strategy. In inference, each query image is perturbed with K noise realizations. The mean and variance of the resulting K reconstructions are used for robust anomaly detection and uncertainty estimation.

3. Experiments

3.1. Dataset

Our dataset includes a cohort of chest CT volumes including (a) a total of 253 healthy subjects with no abnormal or actionable findings in the lung, (b) 23 subjects diagnosed with COVID-19, and (c) 23 subjects with ILD. All CT scans were reconstructed with sharp reconstruction kernels and calibrated with a slice thickness of 5 mm and an in-plane resolution of 512×512 pixels. Each volume contains approximately 80 axial slices. Preprocessing included standard CT chest windowing to normalize intensities into [-1, 1]. For all chest CT scans, the lung masks were automatically extracted using a pre-developed lung segmentation model (Chaganti et al., 2020b). In addition to conditioning, we also used lung masks to select CT slices that are within the lung region for diffusion model training and anomaly detection inference, yielding a total of 22,734 normal slices from all 253 healthy subjects. Voxel-wise dense annotations of abnormal regions were provided by thoracic radiologists for all abnormal cases. Both COVID-19 and ILD datasets were randomly split for validation (3 subjects) and testing (20 subjects).

3.2. Implementation

The SAC-Diff model was trained on a single NVIDIA H100 GPU with a batch size of 4 for 125 epochs. We used the AdamW optimizer (Loshchilov and Hutter, 2018) with an initial learning rate of 1×10^{-4} and cosine weight decay. The training objective was the L2 loss. Based on empirical experimental results, we set $T_{\text{training}} = 1000$, $T_{\text{sampling}} = 550$, $\lambda = 1$ for

COVID dataset and $\lambda = -0.25$ for ILD dataset, and number of samples $K = 7$ in ensemble. Input data consisted of 2D axial slices, which were stacked to obtain 3D outputs. To ensure slice consistency within the same subject, we used a fixed random seed for simplex noise generation across slices within the same 3D scan at inference time.

The trained model was evaluated using dense annotations on both COVID-19 and ILD datasets, each containing 20 subjects. Qualitatively, we compare the reconstructed pseudo-healthy images against their original input CT scans, and the anomaly maps along with segmentation results against the radiologist-annotated ground truth. Quantitatively, we report mean Intersection over Union (IoU), Dice Similarity Coefficient (Dice), Precision, Recall, and False Positive Rate (FPR) on both test sets.

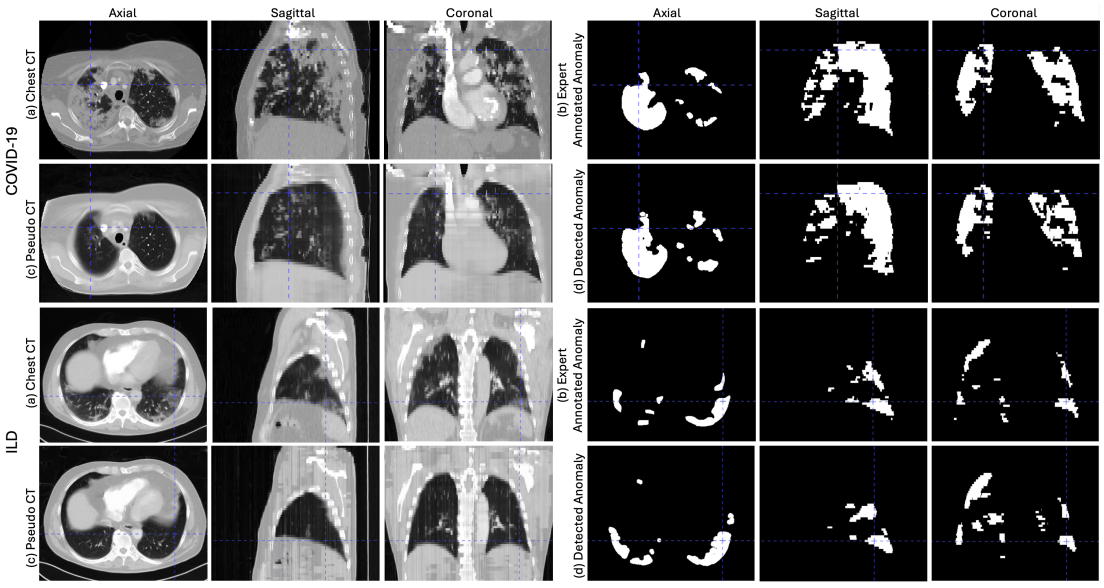

Figure 4: Examples of automatic anomaly detection using the proposed method. (a) 3D chest CT scan from a COVID-19 patient and an ILD patient; (b) Manually annotated anomaly map by radiologist; (c) Synthetic CT scan generated by the model, with anomalies suppressed; (d) Anomaly detected by our model.

3.3. Detection Result Visualization

Fig. 4 presents two 3D examples of anomaly detection using our proposed method, one from a COVID-19 case and one from an ILD case. The reconstructed images suppress heterogeneous anomalies while preserving lung structures in both cases. As illustrated across the coronal, sagittal, and axial views, the model effectively captures the size, structure, and location of various anomalies for different diseases.

3.4. Baseline Comparison and Ablation Study

We conducted sensitivity tests and ablation studies by comparing SAC-Diff with other baselines and configurations, including a supervised method dedicated to COVID lesions (Biondi

et al., 2021), a state-of-the-art medical foundation model MedSAM2 (Ma et al., 2025), and generative model baselines DDPM (Ho et al., 2020) and AnoDDPM (Wyatt et al., 2022). We followed the same training pipeline on the same dataset and optimized the optimal hyperparameter settings used in subsequent experiments.

Qualitative comparisons between the baselines and our model are shown in Fig. 5. In pseudo-normal CT reconstruction, our proposed architecture effectively removes both large and subtle anomalies while preserving lung boundaries. From visual comparisons, prior methods either fail to preserve structural details or fail to suppress significant anomalies. As highlighted by the pink and blue arrows, SAC-Diff successfully detects abnormal regions missed by previous approaches.

Tables 1 and 2 compare SAC-Diff with prior methods on COVID-19 and ILD datasets, respectively. Compared with other baselines and variants (a–e), the proposed SAC-Diff significantly outperforms all methods on both datasets in IoU and Dice. The supervised COVIDSeg achieved the highest recall, but generalized poorly on our in-house COVID set and a more diverse ILD set. MedSAM2 has the lowest FPR on both sets. Despite its on-par performance on the COVID set, both IoU and Dice scores of MedSAM2 dropped drastically on the ILD dataset with more diverse anomalies. These are the limitations of the current supervised methods and foundation models.

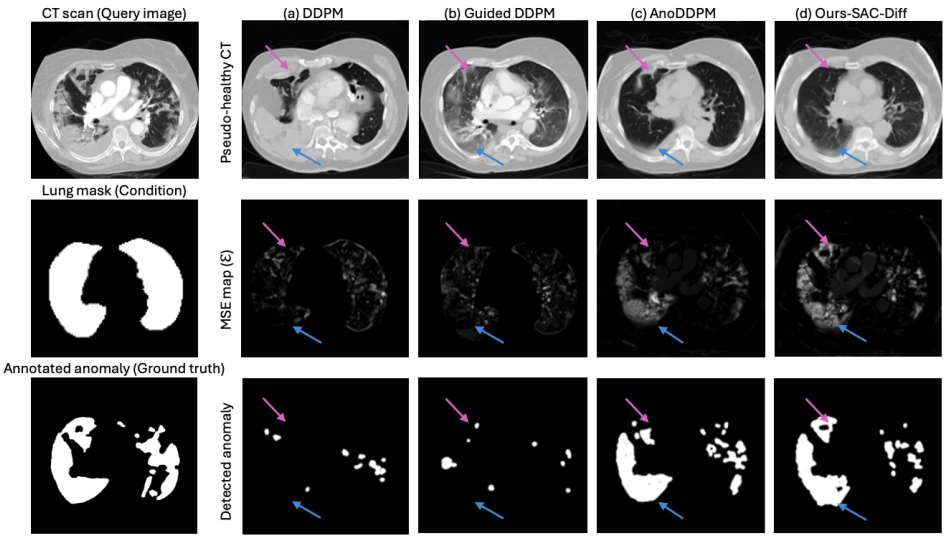

Figure 5: Comparison of anomaly detection between previous models and our proposed model. (a) DDPM; (b) guided DDPM; (c) AnoDDPM; (d) Ours: SAC-Diff. The arrows indicate abnormal regions that were missed by previous methods but successfully detected by the proposed approach.

Comparing DDPM with variant (a), and AnoDDPM with variant (c), we observe anatomy guidance alone could improve performance for the COVID-19 dataset but not sufficient for ILD. Simplex noise yields consistent improvements across settings, as seen in (a) vs. (c) and (b) vs. (d). When background-adaptive thresholding is combined with simplex noise, it achieves strong performance, as in variant (d). In addition, uncertainty-aware ensem-

Table 1: Baseline comparison and ablation study on COVID dataset. Checkmarks indicate which components are enabled in each DDPM variant. The best results are marked **in bold**. An asterisk (*) denotes statistically significant improvement over prior methods.

Config	Simplex Noise	Anatomy-Guidance	Background-Adaptive	Ensembling	IoU ↑	Dice ↑	Precision ↑	Recall ↑	FPR ↓
DDPM (Ho et al., 2020)					0.1260 (0.1025)	0.2099 (0.1560)	0.1462 (0.1238)	0.5330 (0.2256)	0.0228 (0.0199)
COVIDSeg (Supervised) (Biondi et al., 2021)					0.2088 (0.1672)	0.3169 (0.2088)	0.2440 (0.2100)	**0.7033** (0.3230)	0.0166 (0.0186)
AnoDDPM (Wyatt et al., 2022)	✓				0.1596 (0.1535)	0.2487 (0.2046)	0.2269 (0.1990)	0.4660 (0.2870)	0.0215 (0.0181)
MedSAM2 (Ma et al., 2025)					0.3731 (0.1800)	0.5186 (0.2014)	**0.5306** (0.2398)	0.6215 (0.2724)	**0.0076** (0.0089)
Variant (a)		✓			0.1411 (0.1036)	0.2335 (0.1548)	0.1715 (0.1322)	0.4980 (0.2088)	0.0227 (0.0191)
Variant (b)			✓		0.0230 (0.0251)	0.0438 (0.0468)	0.0240 (0.0266)	0.3551 (0.2776)	0.0195 (0.0190)
Variant (c)	✓	✓			0.2147 (0.1943)	0.3161 (0.2421)	0.2735 (0.2391)	0.5092 (0.2705)	0.0134 (0.0128)
Variant (d)	✓		✓		0.2635 (0.1749)	0.3869 (0.2197)	0.3250 (0.2291)	0.6204 (0.2300)	0.0129 (0.0125)
Variant (e)		✓	✓		0.0189 (0.0258)	0.0358 (0.0476)	0.0193 (0.0266)	0.5025 (0.3653)	0.0197 (0.0194)
SAC-Diff w/o Ensemble	✓	✓	✓		0.3140 (0.1939)	0.4439 (0.2323)	0.3737 (0.2434)	0.6849 (0.2404)	0.0117 (0.0121)
SAC-Diff	✓	✓	✓	✓	**0.3871** (0.1727)*	**0.5341** (0.1957)*	0.5008 (0.2483)	0.6662 (0.2262)	0.0094 (0.0098)

Table 2: Baseline comparison and ablation study on ILD dataset.

Config	Simplex Noise	Anatomy-Guidance	Background-Adaptive	Ensembling	IoU ↑	Dice ↑	Precision ↑	Recall ↑	FPR ↓
DDPM (Ho et al., 2020)					0.1193 (0.0991)	0.2000 (0.1509)	0.1379 (0.1201)	0.5384 (0.2274)	0.0232 (0.0198)
COVIDSeg (Supervised) (Biondi et al., 2021)					0.1785(0.1537)	0.2765(0.2137)	0.2005(0.1792)	**0.7981**(0.2781)	0.0282(0.0214)
AnoDDPM (Wyatt et al., 2022)	✓				0.1535 (0.1467)	0.2417 (0.1968)	0.2195 (0.1998)	0.4613 (0.2769)	0.0207 (0.0176)
MedSAM2 (Ma et al., 2025)					0.0270 (0.0350)	0.0540 (0.0637)	0.3369 (0.4977)	0.3324 (0.4626)	**0.0142** (0.0120)
Variant (a)		✓			0.1399 (0.1029)	0.2319 (0.1529)	0.1704 (0.1321)	0.5035 (0.2045)	0.0230 (0.0189)
Variant (b)			✓		0.0147 (0.0125)	0.0287 (0.0239)	0.0151 (0.0129)	0.4687 (0.2054)	0.0331 (0.0213)
Variant (c)	✓	✓			0.1282 (0.1317)	0.2064 (0.1883)	0.1516 (0.1624)	0.5183 (0.2933)	0.0245 (0.0181)
Variant (d)	✓		✓		0.1285 (0.1210)	0.2094 (0.1722)	0.1499 (0.1580)	0.6020 (0.2601)	0.0285 (0.0185)
Variant (e)		✓	✓		0.0116 (0.0172)	0.0224 (0.0325)	0.0117 (0.0173)	0.5709 (0.3360)	0.0332 (0.0215)
SAC-Diff w/o Ensemble	✓	✓	✓		0.2733 (0.1929)	0.3974 (0.2233)	0.3954 (0.2392)	0.5488 (0.2708)	0.0207 (0.0165)
SAC-Diff	✓	✓	✓	✓	**0.3114** (0.2000)*	**0.4435** (0.2184)*	**0.4921** (0.2405)*	0.5229 (0.2648)	0.0179 (0.0156)

bling further improves the model's anomaly detection performance through the enhanced inter-sample consistency. With all the components, the proposed SCA-Diff significantly outperforms other baselines in IoU and Dice, achieving an average IoU of 0.3871 (+3.75%) and an average Dice score of 0.5341 (+2.99%) on COVID-19, and average IoU of 0.3114 (+74.45%) and an average Dice score of 0.4435 (+60.40%) on ILD.

In addition to improved accuracy, the ensembling strategy enables uncertainty quantification by computing the voxel-wise standard deviation across reconstructed results. The resulting uncertainty map provides a per-voxel measure of confidence, offering insights into the model's reliability and facilitating a more robust clinical interpretation. We illustrate this effect with three representative examples in Fig. 6. The predicted anomaly maps are visualized using a color scale: warmer colors indicate higher model confidence related to high cross-inference consistency, while cooler colors reflect greater uncertainty. The arrows highlight regions where ensembling leads to improved detection.

4. Conclusion and Future Directions

In this work, we presented SAC-Diff for unsupervised anomaly detection in chest CT. SAC-Diff uses simplex noise perturbation, subject-aware anatomical conditioning, background-aware masking, and ensemble inference with uncertainty quantification to deliver reliable anomaly detection. Experiments show that SAC-Diff consistently outperforms previous supervised, diffusion-based, and foundation model-based methods, highlighting its ability to localize heterogeneous anomalies. Potential future directions are discussed below:

Domain Shift. Our training dataset contains primarily non-contrast CT images, whereas the test sets include contrast-enhanced scans. This domain shift may slightly

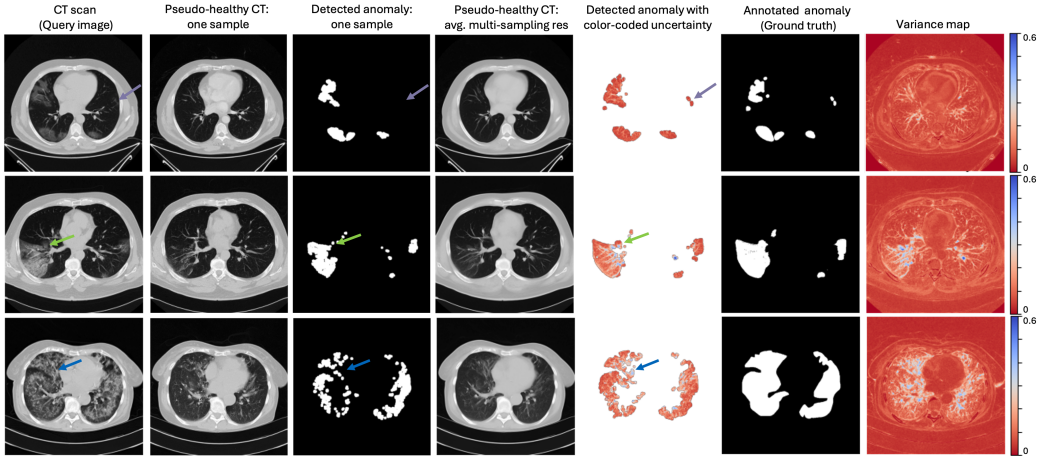

Figure 6: Examples of automatic anomaly detection results with uncertainty-aware ensembling enabled. The predicted anomaly maps and variance maps are represented using the same color scale, where warmer colors correspond to higher model confidence. The arrows highlight regions where inherent consistency in ensembling leads to improved detection.

affect performance, as contrast agents alter intensity distributions, though mostly outside the lung (see Fig. 4 COVID-19 (a) and (c)). While the background awareness of SAC-Diff shows strong robustness under this shift, future work should address contrast variability through domain adaptation techniques or include contrast-enhanced scans into training.

Model Interpretability. The ensembling improves reliability and interpretability through uncertainty quantification. On the uncertainty maps, low variance indicates high model confidence achieved under inherent consistency, and high variance indicates detection with potential ambiguity. This reduces the risk of over-reliance on spurious predictions, making it more suitable for integration into real-world clinical workflows.

Dependence on Segmentation Accuracy. Since the model is guided by a lung mask, the segmentation accuracy directly impacts anomaly detection. Anomalies along the chest wall or airway borders, such as wall thickening, are therefore relatively difficult to capture. This limitation underscores the need for task-specific lung field segmentation methods tailored for anomaly detection. In this work, we used a pre-trained network for lung segmentation (Chaganti et al., 2020b); however, developing an end-to-end pipeline that jointly optimizes organ segmentation and anomaly detection can be a promising direction.

Conditioning Strategies. To provide subject-aware anatomical guidance, we condition the model on a lung mask, which informs the denoising process at every step. This simple yet effective conditioning improves anomaly localization using anatomical context. Beyond lung masks, conditioning on positional encoding or adjacent CT slices could further preserve lung boundaries and enforce spatial coherence across volumes. While early fusion is computationally efficient, more expressive conditioning mechanisms, such as cross-attention (Rombach et al., 2022) or FiLM-based feature modulation (Perez et al., 2018), can enable richer interactions between x_t and c. We leave these extensions for future work.

References

Ibrahim Saad Aly Abdelhalim, Mamdouh Farouk Mohamed, and Yousef Bassyouni Mahdy. Data augmentation for skin lesion using self-attention based progressive generative adversarial network. *Expert Systems with Applications*, 165:113922, March 2021a. ISSN 0957-4174. doi: 10.1016/j.eswa.2020.113922. URL https://www.sciencedirect.com/science/article/pii/S0957417420307132.

Ibrahim Saad Aly Abdelhalim, Mamdouh Farouk Mohamed, and Yousef Bassyouni Mahdy. Data augmentation for skin lesion using self-attention based progressive generative adversarial network. *Expert Systems with Applications*, 165:113922, March 2021b. ISSN 0957-4174. doi: 10.1016/j.eswa.2020.113922. URL https://www.sciencedirect.com/science/article/pii/S0957417420307132.

Samet Akcay, Amir Atapour-Abarghouei, and Toby P. Breckon. GANomaly: Semi-supervised Anomaly Detection via Adversarial Training. In C. V. Jawahar, Hongdong Li, Greg Mori, and Konrad Schindler, editors, *Computer Vision – ACCV 2018*, pages 622–637, Cham, 2019. Springer International Publishing. ISBN 978-3-030-20893-6. doi: 10.1007/978-3-030-20893-6_39.

Masanori Akira, Satoru Yamamoto, Yoshikazu Inoue, and Mitsunori Sakatani. High-resolution CT of asbestosis and idiopathic pulmonary fibrosis. *AJR. American journal of roentgenology*, 181(1):163–169, July 2003. ISSN 0361-803X. doi: 10.2214/ajr.181.1.1810163.

Elisa Baratella, Barbara Ruaro, Cristina Marrocchio, Natalia Starvaggi, Francesco Salton, Fabiola Giudici, Emilio Quaia, Marco Confalonieri, and Maria Assunta Cova. Interstitial Lung Disease at High Resolution CT after SARS-CoV-2-Related Acute Respiratory Distress Syndrome According to Pulmonary Segmental Anatomy. *Journal of Clinical Medicine*, 10(17):3985, January 2021. ISSN 2077-0383. doi: 10.3390/jcm10173985. URL https://www.mdpi.com/2077-0383/10/17/3985. Publisher: Multidisciplinary Digital Publishing Institute.

Berke Doga Basaran, Weitong Zhang, Mengyun Qiao, Bernhard Kainz, Paul M. Matthews, and Wenjia Bai. LesionMix: A Lesion-Level Data Augmentation Method for Medical Image Segmentation. In Yuan Xue, Chen Chen, Chao Chen, Lianrui Zuo, and Yihao Liu, editors, *Data Augmentation, Labelling, and Imperfections*, pages 73–83, Cham, 2024. Springer Nature Switzerland. ISBN 978-3-031-58171-7. doi: 10.1007/978-3-031-58171-7_8.

Farzad Beizaee, Gregory Lodygensky, Christian Desrosiers, and Jose Dolz. MAD-AD: Masked Diffusion for Unsupervised Brain Anomaly Detection, July 2025. URL http://arxiv.org/abs/2502.16943. arXiv:2502.16943 [cs].

Cosmin I. Bercea, Benedikt Wiestler, Daniel Rueckert, and Julia A. Schnabel. Diffusion Models with Implicit Guidance for Medical Anomaly Detection . In *proceedings of Medical Image Computing and Computer Assisted Intervention – MICCAI 2024*, volume LNCS 15011. Springer Nature Switzerland, October 2024.

Cosmin I. Bercea, Benedikt Wiestler, Daniel Rueckert, and Julia A. Schnabel. Evaluating normative representation learning in generative AI for robust anomaly detection in brain imaging. *Nature Communications*, 16(1):1624, February 2025. ISSN 2041-1723. doi: 10.1038/s41467-025-56321-y.

Riccardo Biondi, Nico Curti, Francesca Coppola, Enrico Giampieri, Giulio Vara, Michele Bartoletti, Arrigo Cattabriga, Maria Adriana Cocozza, Federica Ciccarese, Caterina De Benedittis, Laura Cercenelli, Barbara Bortolani, Emanuela Marcelli, Luisa Pierotti, Lidia Strigari, Pierluigi Viale, Rita Golfieri, and Gastone Castellani. Classification Performance for COVID Patient Prognosis from Automatic AI Segmentation—A Single-Center Study. *Applied Sciences*, 11(12):5438, January 2021. ISSN 2076-3417. doi: 10.3390/app11125438. URL https://www.mdpi.com/2076-3417/11/12/5438. Publisher: Multidisciplinary Digital Publishing Institute.

Anupama Gupta Brixey, Andrea S. Oh, Aseel Alsamarraie, and Jonathan H. Chung. Pictorial Review of Fibrotic Interstitial Lung Disease on High-Resolution CT Scan and Updated Classification. *Chest*, 165(4):908–923, April 2024. ISSN 1931-3543. doi: 10.1016/j.chest.2023.11.037.

Yu Cai, Weiwen Zhang, Hao Chen, and Kwang-Ting Cheng. MedIAnomaly: A comparative study of anomaly detection in medical images. *Medical Image Analysis*, 102:103500, May 2025. ISSN 1361-8415. doi: 10.1016/j.media.2025.103500. URL https://www.sciencedirect.com/science/article/pii/S1361841525000489.

Shikha Chaganti, Philippe Grenier, Abishek Balachandran, Guillaume Chabin, Stuart Cohen, Thomas Flohr, Bogdan Georgescu, Sasa Grbic, Siqi Liu, François Mellot, Nicolas Murray, Savvas Nicolaou, William Parker, Thomas Re, Pina Sanelli, Alexander W. Sauter, Zhoubing Xu, Youngjin Yoo, Valentin Ziebandt, and Dorin Comaniciu. Automated Quantification of CT Patterns Associated with COVID-19 from Chest CT. *Radiology. Artificial Intelligence*, 2(4):e200048, July 2020a. ISSN 2638-6100. doi: 10.1148/ryai.2020200048.

Shikha Chaganti, Philippe Grenier, Abishek Balachandran, Guillaume Chabin, Stuart Cohen, Thomas Flohr, Bogdan Georgescu, Sasa Grbic, Siqi Liu, François Mellot, Nicolas Murray, Savvas Nicolaou, William Parker, Thomas Re, Pina Sanelli, Alexander W. Sauter, Zhoubing Xu, Youngjin Yoo, Valentin Ziebandt, and Dorin Comaniciu. Automated Quantification of CT Patterns Associated with COVID-19 from Chest CT. *Radiology: Artificial Intelligence*, 2(4):e200048, July 2020b. doi: 10.1148/ryai.2020200048. URL https://pubs.rsna.org/doi/full/10.1148/ryai.2020200048. Publisher: Radiological Society of North America.

Prafulla Dhariwal and Alexander Quinn Nichol. Diffusion Models Beat GANs on Image Synthesis. November 2021. URL https://openreview.net/forum?id=AAWuCvzaVt.

Yann Gaillandre, Alain Duhamel, Thomas Flohr, Jean-Baptiste Faivre, Suonita Khung, Antoine Hutt, Paul Felloni, Jacques Remy, and Martine Remy-Jardin. Ultra-high resolution CT imaging of interstitial lung disease: impact of photon-counting CT in 112

patients. *European Radiology*, 33(8):5528–5539, August 2023. ISSN 1432-1084. doi: 10.1007/s00330-023-09616-x.

Florin-Cristian Ghesu, Bogdan Georgescu, Yefeng Zheng, Sasa Grbic, Andreas Maier, Joachim Hornegger, and Dorin Comaniciu. Multi-Scale Deep Reinforcement Learning for Real-Time 3D-Landmark Detection in CT Scans. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(1):176–189, January 2019. ISSN 1939-3539. doi: 10.1109/TPAMI.2017.2782687. URL https://ieeexplore.ieee.org/document/8187667.

Xueqi Guo, Mohamad Abdi, Yoshihisa Shinagawa, Anna Jerebko, and Sepehr Farhand. SEAM-STRESS: A Weakly Supervised Framework for Interstitial Lung Disease Segmentation in Chest CT. In *2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI)*, pages 1–4, April 2023. doi: 10.1109/ISBI53787.2023.10230747. URL https://ieeexplore.ieee.org/abstract/document/10230747. ISSN: 1945-8452.

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 Thirty-Eighth AAAI Conference on Artificial Intelligence and Thirty-Sixth Conference on Innovative Applications of Artificial Intelligence and Fourteenth Symposium on Educational Advances in Artificial Intelligence*, volume 38 of *AAAI'24/IAAI'24/EAAI'24*, pages 8472–8480. AAAI Press, February 2024. ISBN 978-1-57735-887-9. doi: 10.1609/aaai.v38i8.28690. URL https://doi.org/10.1609/aaai.v38i8.28690.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising Diffusion Probabilistic Models. In *Advances in Neural Information Processing Systems*, volume 33, pages 6840–6851. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper/2020/hash/4c5bcfec8584af0d967f1ab10179ca4b-Abstract.html.

Weikai Huang, Yijin Huang, and Xiaoying Tang. AugPaste: One-Shot Anomaly Detection for Medical Images. In *Ophthalmic Medical Image Analysis: 9th International Workshop, OMIA 2022, Held in Conjunction with MICCAI 2022, Singapore, Singapore, September 22, 2022, Proceedings*, pages 1–11, Berlin, Heidelberg, September 2022. Springer-Verlag. ISBN 978-3-031-16524-5. doi: 10.1007/978-3-031-16525-2_1. URL https://doi.org/10.1007/978-3-031-16525-2_1.

Philipp Krähenbühl and Vladlen Koltun. Efficient Inference in Fully Connected CRFs with Gaussian Edge Potentials, October 2012. URL http://arxiv.org/abs/1210.5644. arXiv:1210.5644 [cs].

Qingyun Li, Zhibin Yu, Yubo Wang, and Haiyong Zheng. TumorGAN: A Multi-Modal Data Augmentation Framework for Brain Tumor Segmentation. *Sensors (Basel, Switzerland)*, 20(15):4203, July 2020. ISSN 1424-8220. doi: 10.3390/s20154203.

Jiaqi Liu, Guoyang Xie, Jinbao Wang, Shangnian Li, Chengjie Wang, Feng Zheng, and Yaochu Jin. Deep Industrial Image Anomaly Detection: A Survey. *Machine Intelligence Research*, 21(1):104–135, February 2024. ISSN 2731-5398. doi: 10.1007/s11633-023-1459-z. URL https://doi.org/10.1007/s11633-023-1459-z.

Ilya Loshchilov and Frank Hutter. Decoupled Weight Decay Regularization. September 2018. URL https://openreview.net/forum?id=Bkg6RiCqY7.

Jun Ma, Zongxin Yang, Sumin Kim, Bihui Chen, Mohammed Baharoon, Adibvafa Fallahpour, Reza Asakereh, Hongwei Lyu, and Bo Wang. MedSAM2: Segment Anything in 3D Medical Images and Videos, April 2025. URL http://arxiv.org/abs/2504.03600. arXiv:2504.03600 [eess].

Ruimin Ma, Shuang Li, Yuanying Wang, Shuqiao Yang, Na Bao, and Qiao Ye. High-resolution computed tomography features of asbestosis versus fibrotic hypersensitivity pneumonitis: an observational study. *BMC Pulmonary Medicine*, 22(1):207, May 2022. ISSN 1471-2466. doi: 10.1186/s12890-022-01967-3. URL https://doi.org/10.1186/s12890-022-01967-3.

Tony C. W. Mok and Albert C. S. Chung. Learning Data Augmentation for Brain Tumor Segmentation with Coarse-to-Fine Generative Adversarial Networks. In Alessandro Crimi, Spyridon Bakas, Hugo Kuijf, Farahani Keyvan, Mauricio Reyes, and Theo van Walsum, editors, *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, pages 70–80, Cham, 2019. Springer International Publishing. ISBN 978-3-030-11723-8. doi: 10.1007/978-3-030-11723-8_7.

Alexander Quinn Nichol and Prafulla Dhariwal. Improved Denoising Diffusion Probabilistic Models. October 2020. URL https://openreview.net/forum?id=-NEXDKk8gZ.

Guansong Pang, Chunhua Shen, Longbing Cao, and Anton Van Den Hengel. Deep Learning for Anomaly Detection: A Review. *ACM Comput. Surv.*, 54(2):38:1–38:38, March 2021. ISSN 0360-0300. doi: 10.1145/3439950. URL https://dl.acm.org/doi/10.1145/3439950.

Ethan Perez, Florian Strub, Harm de Vries, Vincent Dumoulin, and Aaron Courville. FiLM: visual reasoning with a general conditioning layer. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence and Thirtieth Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Educational Advances in Artificial Intelligence*, AAAI'18/IAAI'18/EAAI'18, pages 3942–3951, New Orleans, Louisiana, USA, February 2018. AAAI Press. ISBN 978-1-57735-800-8.

Aria Pezeshk, Nicholas Petrick, Weijie Chen, and Berkman Sahiner. Seamless lesion insertion for data augmentation in CAD training. *IEEE transactions on medical imaging*, 36(4):1005–1015, April 2017. ISSN 0278-0062. doi: 10.1109/TMI.2016.2640180. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC5509514/.

Walter H. L. Pinaya, Mark S. Graham, Robert Gray, Pedro F. da Costa, Petru-Daniel Tudosiu, Paul Wright, Yee H. Mah, Andrew D. MacKinnon, James T. Teo, Rolf Jager, David Werring, Geraint Rees, Parashkev Nachev, Sebastien Ourselin, and M. Jorge Cardoso. Fast Unsupervised Brain Anomaly Detection and Segmentation with Diffusion Models. In Linwei Wang, Qi Dou, P. Thomas Fletcher, Stefanie Speidel, and Shuo Li, editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2022*, pages 705–714, Cham, 2022. Springer Nature Switzerland. ISBN 978-3-031-16452-1. doi: 10.1007/978-3-031-16452-1_67.

Jiantao Pu, Joseph K. Leader, Andriy Bandos, Shi Ke, Jing Wang, Junli Shi, Pang Du, Youmin Guo, Sally E. Wenzel, Carl R. Fuhrman, David O. Wilson, Frank C. Sciurba, and Chenwang Jin. Automated quantification of COVID-19 severity and progression using chest CT images. *European Radiology*, 31(1):436–446, January 2021. ISSN 1432-1084. doi: 10.1007/s00330-020-07156-2.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-Resolution Image Synthesis with Latent Diffusion Models, April 2022. URL http://arxiv.org/abs/2112.10752. arXiv:2112.10752 [cs].

Daniel L. Ruderman. Origins of scaling in natural images. *Vision Research*, 37(23):3385–3398, December 1997. ISSN 0042-6989. doi: 10.1016/S0042-6989(97)00008-4. URL https://www.sciencedirect.com/science/article/pii/S0042698997000084.

Mostafa Salem, Sergi Valverde, Mariano Cabezas, Deborah Pareto, Arnau Oliver, Joaquim Salvi, Àlex Rovira, and Xavier Lladó. Multiple Sclerosis Lesion Synthesis in MRI Using an Encoder-Decoder U-NET. *IEEE Access*, 7:25171–25184, 2019. ISSN 2169-3536. doi: 10.1109/ACCESS.2019.2900198. URL https://ieeexplore.ieee.org/document/8645628.

Durgesh Samariya and Amit Thakkar. A Comprehensive Survey of Anomaly Detection Algorithms. *Annals of Data Science*, 10(3):829–850, June 2023. ISSN 2198-5812. doi: 10.1007/s40745-021-00362-9. URL https://doi.org/10.1007/s40745-021-00362-9.

Thomas Schlegl, Philipp Seeböck, Sebastian M. Waldstein, Georg Langs, and Ursula Schmidt-Erfurth. f-AnoGAN: Fast unsupervised anomaly detection with generative adversarial networks. *Medical Image Analysis*, 54:30–44, May 2019. ISSN 1361-8415. doi: 10.1016/j.media.2019.01.010. URL https://www.sciencedirect.com/science/article/pii/S1361841518302640.

Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-Based Generative Modeling through Stochastic Differential Equations. October 2020. URL https://openreview.net/forum?id=PxTIG12RRHS.

Julia Wolleb, Florentin Bieder, Robin Sandkühler, and Philippe C. Cattin. Diffusion Models for Medical Anomaly Detection. In Linwei Wang, Qi Dou, P. Thomas Fletcher, Stefanie Speidel, and Shuo Li, editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2022*, pages 35–45, Cham, 2022. Springer Nature Switzerland. ISBN 978-3-031-16452-1. doi: 10.1007/978-3-031-16452-1_4.

Julian Wyatt, Adam Leach, Sebastian M. Schmon, and Chris G. Willcocks. AnoDDPM: Anomaly Detection with Denoising Diffusion Probabilistic Models using Simplex Noise. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 649–655, June 2022. doi: 10.1109/CVPRW56347.2022.00080. URL https://ieeexplore.ieee.org/document/9857019. ISSN: 2160-7516.

Dong Yang, Daguang Xu, S. Kevin Zhou, Bogdan Georgescu, Mingqing Chen, Sasa Grbic, Dimitris Metaxas, and Dorin Comaniciu. Automatic Liver Segmentation Using

an Adversarial Image-to-Image Network. In Maxime Descoteaux, Lena Maier-Hein, Alfred Franz, Pierre Jannin, D. Louis Collins, and Simon Duchesne, editors, *Medical Image Computing and Computer Assisted Intervention MICCAI 2017*, pages 507–515, Cham, 2017. Springer International Publishing. ISBN 978-3-319-66179-7. doi: 10.1007/978-3-319-66179-7_58.

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 Aleš Leonardis, Elisa Ricci, Stefan Roth, Olga Russakovsky, Torsten Sattler, and Gül Varol, editors, *Computer Vision – ECCV 2024*, pages 1–17, Cham, 2025. Springer Nature Switzerland. ISBN 978-3-031-73209-6. doi: 10.1007/978-3-031-73209-6_1.

Jongmin Yu, Hyeontaek Oh, and Jinhong Yang. Adversarial Denoising Diffusion Model for Unsupervised Anomaly Detection. October 2023. URL https://openreview.net/forum?id=PvHhhn1iX9.

Xinyi Zhang, Naiqi Li, Jiawei Li, Tao Dai, Yong Jiang, and Shu-Tao Xia. Unsupervised Surface Anomaly Detection with Diffusion Probabilistic Model. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 6759–6768, October 2023. doi: 10.1109/ICCV51070.2023.00624. URL https://ieeexplore.ieee.org/document/10377534. ISSN: 2380-7504.

Chong Zhou and Randy C. Paffenroth. Anomaly Detection with Robust Deep Autoencoders. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '17, pages 665–674, New York, NY, USA, August 2017. Association for Computing Machinery. ISBN 978-1-4503-4887-4. doi: 10.1145/3097983.3098052. URL https://doi.org/10.1145/3097983.3098052.

David Zimmerer, Simon Kohl, Jens Petersen, Fabian Isensee, and Klaus Maier-Hein. Context-encoding Variational Autoencoder for Unsupervised Anomaly Detection. April 2019. URL https://openreview.net/forum?id=BylLiVXptV.

Appendix A. Lung Mask Segmentation

For lung mask segmentation, we first use multiscale deep reinforcement learning to identify anatomical landmarks (Ghesu et al., 2019). The carina bifurcation serves as the primary reference point for locating the lung region of interest (ROI); when this landmark cannot be detected, the sternum tip is used as an alternative. The dimensions and spatial placement of the lung ROI relative to the detected landmark are defined based on annotated lung datasets.

The extracted lung ROI is then resampled to a 2-mm isotropic grid and passed through an adversarial Image-to-Image Network (DI2IN) (Yang et al., 2017) to produce the lung segmentation. The main structure of DI2IN is a symmetric convolutional encoder-decoder and the discriminator is a CNN. Afterward, the predicted ROI segmentation is mapped back to the original image space to restore the native resolution and dimensions. This pipeline is pre-trained from heterogeneous patient population to ensure performance across diverse pathologies and applied to our datasets.

Appendix B. Denoising Diffusion Probabilistic Models

DDPMs (Ho et al., 2020; Nichol and Dhariwal, 2020) have emerged as a state-of-the-art approach in generative modeling, achieving high sample fidelity and superior mode coverage. It consists of two core components:

Forward diffusion process: The forward process $q(x_t|x_{t-1})$ gradually corrupts a clean sample $x_0 \sim q(x_0)$ into Gaussian noise over a sequence of T time steps. At each step, the x_t is sampled from a Gaussian distribution centered around the previous state:

$$q(x_t|x_{t-1}) = \mathcal{N}(x_t; \sqrt{\alpha_t}x_{t-1}, (1-\alpha_t)\mathbf{I}),$$

where $\{\alpha_t\}_{t=1}^T$ is a predefined noise schedule. By recursively applying this process, a closed-form expression for the noised input at any timestep t is

$$x_t = \sqrt{\bar{\alpha}_t}x_0 + \sqrt{1-\bar{\alpha}_t}\,\epsilon,$$

where $\bar{\alpha}_t = \prod_{s=1}^t \alpha_s$ and $\epsilon \sim \mathcal{N}(0, \mathbf{I})$.

Learned reverse denoising process: The generative process in DDPMs is defined by a learned reverse Markov chain:

$$p_\theta(x_{0:T}) = p(x_T) \prod_{t=1}^T p_\theta(x_{t-1}|x_t),$$

where the prior is defined as $p(x_T) = \mathcal{N}(0, \mathbf{I})$. Each reverse step is modeled as a Gaussian distribution with a mean $\mu_\theta(x_t, t)$ and fixed variance. The mean can be reparameterized using a neural network $\hat{\epsilon}_\theta(x_t, t)$ that predicts the noise used in the forward process

$$\mu_\theta(x_t, t) = \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\,\hat{\epsilon}_\theta(x_t, t)\right).$$

This parameterization allows us to train the model to predict the noise $\epsilon \sim \mathcal{N}(0, \mathbf{I})$ that perturbed x_0 into x_t. The model is optimized using a denoising score-matching objective:

$$\mathcal{L}(\theta) = \sum_{t=1}^{T} \mathbb{E}_{x_0 \sim q(x_0), \epsilon \sim \mathcal{N}(0,\mathbf{I})} \left[\|\epsilon - \hat{\epsilon}_\theta(x_t, t)\|_2^2 \right].$$

We adapted the DDPM framework described above, replacing the Gaussian noise with Simplex noise.

Appendix C. Training Algorithm

Algorithm 2: Training

$x_0 \sim q(x_0)$
$t \sim \mathcal{U}(\{1, \ldots, T = 1000\})$
$\epsilon \sim \text{Simplex}(\nu = 2^{-6}, N = 6, \gamma = 0.8)$
repeat
  Compute noisy input $x_t = \sqrt{\bar{\alpha}_t} \cdot x_0 + \sqrt{1 - \bar{\alpha}_t} \cdot \epsilon$
  Take a gradient descent step on $\nabla_\theta \left\| \epsilon - \hat{\epsilon}_\theta([x_t \,\|\, c], t) \right\|^2$
until *converged*;