Supplymentary Material of ART-ASyn

Overview

This supplementary material complements the main paper by providing detail explanation of the synthetic anomaly composition steps of ART-ASyn and quantitative analysis of several hyperparameters used in the paper.

§1. Clarification of Synthesis Method

In this section, the implementation details of synthesis method are provided along with the ablation study of the effect of each synthesis step.

§1.1. Implementation Details

As introduced in Section 3.2.1 of the main paper, the synthesis pipeline is built upon three primary transformations of crystallization, blurring, and intensity scaling. The following subsections provide detailed explanations of each transformation and their composition. For the full implementation, please refer to `preprocessing/ART_ASyn.py` in the code.

§1.1.1. Crystallization (T_{cryst})

In computer graphics, crystallization is a non-photorealistic rendering (NPR) technique that mimics the appearance of crystals or stained glass. Instead of preserving all fine details, the image is decomposed into cells with uniform colors, producing a stylized effect [1]. It follows the implementation of:

1. A fixed number of random "seed" points are sampled across the image grid. These act as nuclei around which regions will grow.
2. Each pixel in the image is assigned to the nearest seed using a spatial nearest-neighbor search utilizing `cKDTree` algorithm from Python module `scipy.spatial`. This yields a Voronoi diagram where the plane is divided into non-overlapping cells, each containing all pixels closest to its seed.
3. For each Voronoi cell, the algorithm computes the mean RGB color of all original pixels within the region. Every pixel in that cell is then replaced by this mean color.

This operation corresponds to the crystallization transformation mentioned in Section 3.2.1 of the main paper.

§1.1.2. Blurring (T_{blur})

Gaussian blurring is applied to both the raw anomaly layer and to crystallized regions utilizing `ImageFilter.GaussianBlur` from Python module of `PIL` with radii of 8 and 3, respectively. This simulates consolidation-like smoothness or the gradual density transitions observed in real lung pathologies.

§1.1.3. Intensity Scaling (T_{rib})

The alpha channel of anomalies is modulated by a rib-intensity–weighted mask computed from the original chest X-ray. This ensures that synthetic anomalies preserve radiographic consistency with surrounding rib shadows and do not unrealistically occlude high-density bony structures.

§1.1.4. Composition of Transformations

The three transformations are applied sequentially and composited together:

1. A blured version of brush painted anomaly is generated.
2. A crystallized version of brush painted anomaly is generated.
3. These layers are alpha-scaled and overlaid to create a composite with both sharp-edged and smoothed structures.
4. The anomaly alpha is refined with lung and rib-intensity masks, ensuring anatomical plausibility.

This layered application follows the conceptual description in Fig. 1 of the main paper, an initial painted anomaly is progressively refined by blurring, crystallization, and rib-aware scaling before being overlaid onto the normal chest X-ray to composite the synthetic abnormal image.

§1.2. Ablation Study

We conducted stepwise experiments to analyze the contribution of different synthesis transformations. The ablation starts with the baseline synthesis (A_{base}), then incorporates blurring (T_{blur}) and crystallization (T_{cryst}), and finally integrates all transformations into the complete pipeline (A_{final}). The transformations of blur and crystallization are introduced jointly, as they are designed to operate in a composite manner: blurring smooths structural boundaries while crystallization introduces irregular granular patterns, together forming a realistic density base.

The base of our synthetic abnormality (A_{base}) produces texture-based, cocoon-like patterns with irregular edges (see Figure S1-1). Quantitatively, A_{base} achieves 0.8315/0.6603 in AP/AUC on CheXpert and 0.9112/0.8797 on ZhangLab, compared to 0.7394/0.5929 and 0.7938/0.6842 under the no-synthesis setting, and improves Dice on QaTa-ZeroShot from 0.0707 to 0.3067. Compared to established methods, A_{base} already exceeds most baselines, although its AUC values of 0.6603 on CheXpert and 0.8797 on ZhangLab remain below the AnatPaste baseline of 0.7825 and 0.9103, respectively. This suggests that the texture painted A_{base} alone creates anomaly like structures.

Synthesis Step	CheXpert		ZhangLab		QaTa-ZeroShot
	AP	AUC	AP	AUC	Dice
None	0.7394	0.5929	0.7938	0.6842	0.0707
A_{base}	0.8315	0.6603	0.9112	0.8797	0.3067
$A_{\text{base}} + T_{\text{blur}} + T_{\text{cryst}}$	0.7840	0.5953	0.8536	0.7574	0.0954
A_{final}	**0.8639**	**0.7125**	**0.9621**	**0.9350**	**0.3275**

Table S1. Quantitative analysis of the effect of different synthesis steps on anomaly detection performance across CheXpert, ZhangLab, and QaTa-ZeroShot.

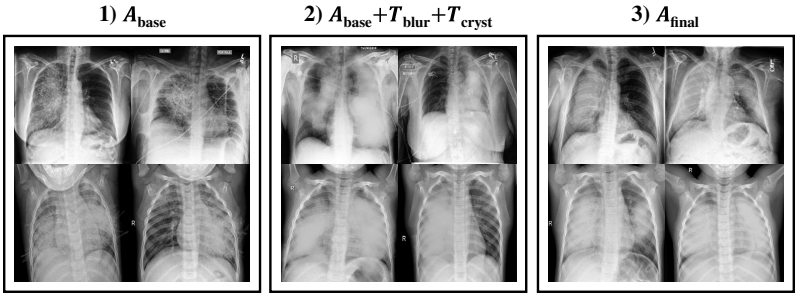

Figure S1. Visual comparison of different synthesis steps.

Adding blur and crystallization ($A_{\text{base}} + T_{\text{blur}} + T_{\text{cryst}}$) produces denser structures but tends to obscure anatomical boundaries (see Figure S1-2). This results in significant performance degradation compared to A_{base}. The variant lowers AP/AUC to 0.7840/0.5953 on CheXpert and 0.8536/0.7574 on ZhangLab, with Dice dropping to 0.0954 on QaTa-ZeroShot, which is only slightly better than the no-synthesis setting. adding an intensity-scaling step (T_{rib}) to restore anatomical visibility restores our ART-ASyn method, producing the A_{final} synthesis (see Figure S1-3). Collectively, blur and crystallization provide a necessary density base, and when combined with intensity scaling, they complement the texture-focused A_{base} to achieve superior anomaly realism as reported in Table 1 of the main paper and also Table S1.

In summary, the ablation demonstrates that texture synthesis alone (A_{base}) already enhances anomaly detection, but combining blur and crystallization with intensity scaling in the full A_{final} pipeline is essential to balance texture fidelity and anatomical consistency, resulting in the strongest and most consistent performance across datasets.

§2. Sensitivity of the Binarization Threshold for Anomaly Mask Computation

For ease of experimentation in the main paper, we fix the binarization threshold at $\tau = 0.2$ when deriving the anomaly mask $\hat{M}_{\text{anomaly}}$ from the pixel-wise reconstruction error between the synthetic input I_{syn} and its reconstruction $\hat{I}$. However, as a dataset-dependent thresholding hyperparameter, it is important to evaluate the model's sensitivity to its setting. We therefore explore a wide range of τ from 0.2 to 0.8 with a interval of 0.2 and report their respective anomaly detection performance on CheXpert and ZhangLab, and QaTa-ZeroShot as reported in Table S2.

On CheXpert, the best scores occur at $\tau = 0.6$ with AP/AUC $0.8694/0.7325$, while the lowest are at $\tau = 0.4$ with $0.8523/0.6981$, giving sensitivity ranges of 0.0171 (AP) and 0.0344 (AUC). On ZhangLab, the highest AP/AUC is at $\tau = 0.2$ with $0.9621/0.9350$ and the lowest at $\tau = 0.6$ with $0.9348/0.8930$, corresponding to ranges of 0.0273 (AP) and 0.0420 (AUC). For QaTa-ZeroShot, Dice peaks at $\tau = 0.6$ with 0.3319 and is lowest at $\tau = 0.4$ with 0.2963 (range 0.0356). Overall, the metrics are relatively insensitive to τ, with mild improvements near $\tau = 0.6$ for CheXpert and QaTa-ZeroShot, and near $\tau = 0.2$ for ZhangLab.

It is evident from the results that although the model's scores vary with the choice of τ, the magnitude of these changes is minor, below most of the performance margins by which our method exceeds competing baselines. Therefore, the performance of the model is relatively insensitive to the exact binarization threshold τ.

§3. Quantitative Analysis of the Effect of Loss Components

In Section 3.2.2 of the main paper, the loss is defined as:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{local}} + \mathcal{L}_{\text{global}} + \mathcal{L}_{\text{feat}} + \mathcal{L}_{\text{dice}}. \quad (1)$$

Threshold (τ)	CheXpert		ZhangLab		QaTa-ZeroShot
	AP	AUC	AP	AUC	Dice
0.2	0.8639	0.7125	**0.9621**	**0.9350**	0.3275
0.4	0.8523	0.6981	0.9417	0.9019	0.2963
0.6	**0.8694**	**0.7325**	0.9348	0.8930	**0.3319**
0.8	0.8531	0.7270	0.9459	0.9204	0.3133

Table S2. Sensitivity analysis of the effect of different binarization thresholds τ on anomaly detection result of CheXpert, ZhangLab and QaTa-ZeroShot.

Loss Configuration	CheXpert		ZhangLab		QaTa-ZeroShot
	AP	AUC	AP	AUC	Dice
$\mathcal{L}_{\text{global}}$	0.8250	0.6671	0.8124	0.7457	0.2771
$\mathcal{L}_{\text{global}} + \mathcal{L}_{\text{local}}$	0.8319	0.6675	0.8788	0.7575	0.3125
$\mathcal{L}_{\text{global}} + \mathcal{L}_{\text{feat}}$	0.8347	0.6742	0.9382	0.9053	0.2527
$\mathcal{L}_{\text{global}} + \mathcal{L}_{\text{dice}}$	0.8362	0.6860	0.9032	0.8297	0.2804
$\mathcal{L}_{\text{global}} + \mathcal{L}_{\text{local}} + \mathcal{L}_{\text{feat}}$	0.8573	0.7043	0.9409	0.9018	0.3227
$\mathcal{L}_{\text{global}} + \mathcal{L}_{\text{local}} + \mathcal{L}_{\text{dice}}$	0.8445	0.6801	0.9058	0.8550	0.3266
$\mathcal{L}_{\text{global}} + \mathcal{L}_{\text{feat}} + \mathcal{L}_{\text{dice}}$	0.8499	0.6762	0.9388	0.9003	0.3240
$\mathcal{L}_{\text{total}}$	**0.8639**	**0.7125**	**0.9621**	**0.9350**	**0.3275**

Table S3. Quantitative analysis of the effect of training under different loss component permutations on anomaly detection performance across CheXpert, ZhangLab, and QaTa-ZeroShot.

In this section, the effect of each permutation of different loss components on anomaly detecton performance for datasets CheXpert, ZhangLab, and QaTa-ZeroShot is shown in Table S3.

With the standard global reconstruction loss $\mathcal{L}_{\text{global}}$, the model reaches baseline anomaly detection performance of 0.8250/0.6671 on CheXpert and 0.8124/0.7457 on ZhangLab, with a Dice score of 0.2771 on QaTa-ZeroShot. This serves as the foundation upon which the effects of additional loss terms can be assessed.

Incorporating an extra loss component of the local reconstruction loss $\mathcal{L}_{\text{local}}$, the feature-alignment loss $\mathcal{L}_{\text{feat}}$, or the Dice loss $\mathcal{L}_{\text{dice}}$ consistently improves performance. Among these, the feature-alignment loss $\mathcal{L}_{\text{feat}}$ particularly effective on the one-class classification task. It elevates the AP/AUC of CheXpert to 0.8347/0.6742 and ZhangLab to 0.9382/0.9053, representing the largest gains relative to the baseline.

§4. Efficiency Analysis of PBTSeg

In the main paper, we propose a progressive binary thresholding lung segmentation method (PBTSeg). Table S4 reports the number of images processed per second on the same computer with different number of workers. It demonstrates the efficiency limitation of PBTSeg, as the simpler lung segmentation method in AnatPaste is approximately 10 times faster. This sacrifice in efficiency is offset by improved segmentation quality which is critical for ensuring that anomalies are synthesized within accurate lung boundaries, which directly impacts the anatomical plausibility and effectiveness of the generated samples. Additionally, efficiency limitation of PBTSeg represents a one-time preprocessing cost rather than an ongoing bottleneck during training as it only need to be computed once for the dataset and can then be reused throughout all subsequent experiments.

No. of Workers	AnatPaste	PBTSeg
	Image/Second	Image/Second
1	99.47	8.95
2	160.22	17.51
4	286.52	26.83
8	423.99	40.52

Table S4. Efficiency analysis of lung segmentation utilizing AnatPaste and PBTSeg.

References

[1] Narendra Ahuja, Byong An, and Bruce Schachter. Image representation using voronoi tessellation. *Computer Vision, Graphics, and Image Processing*, 29(3):286–295, 1985. 1