My Work at Qure.AI

As an AI Scientist Intern, from May 2024 to Dec 2025

Aryan Goyal

Qure.ai
Indian Institute of Technology Bombay

March 21, 2026

Outline

Internship Overview

DiffusionXRay

Controllable Lung Nodule Synthesis

Overview

During my internship, I worked on diffusion models for **synthetic chest X-ray data**, with a focus on keeping fine-grained details intact.

I worked on:

- Image enhancement / super-resolution of DRRs
- Generating diverse synthetic lung nodules

Also explored:

- evals for pulmonary nodule
- flow models

Primary Research Outputs

- *DiffusionXRay: A Diffusion and GAN-Based Approach for Enhancing Digitally Reconstructed Chest Radiographs*
  Aryan Goyal[†], Ashish Mittal[†], Pranav Rao, Manoj Tadepalli, Preetham Putha
  DEMI Workshop, MICCAI 2025 (Accepted)
  Paper — Code — Poster

- *A Diffusion-Driven Fine-Grained Nodule Synthesis Framework for Enhanced Lung Nodule Detection from Chest Radiographs*
  Shreshtha Singh[†], Aryan Goyal[†], Ashish Mittal[†], Manoj Tadepalli, Piyush Kumar, Preetham Putha
  MIDL 2026 (Submitted)
  OpenReview

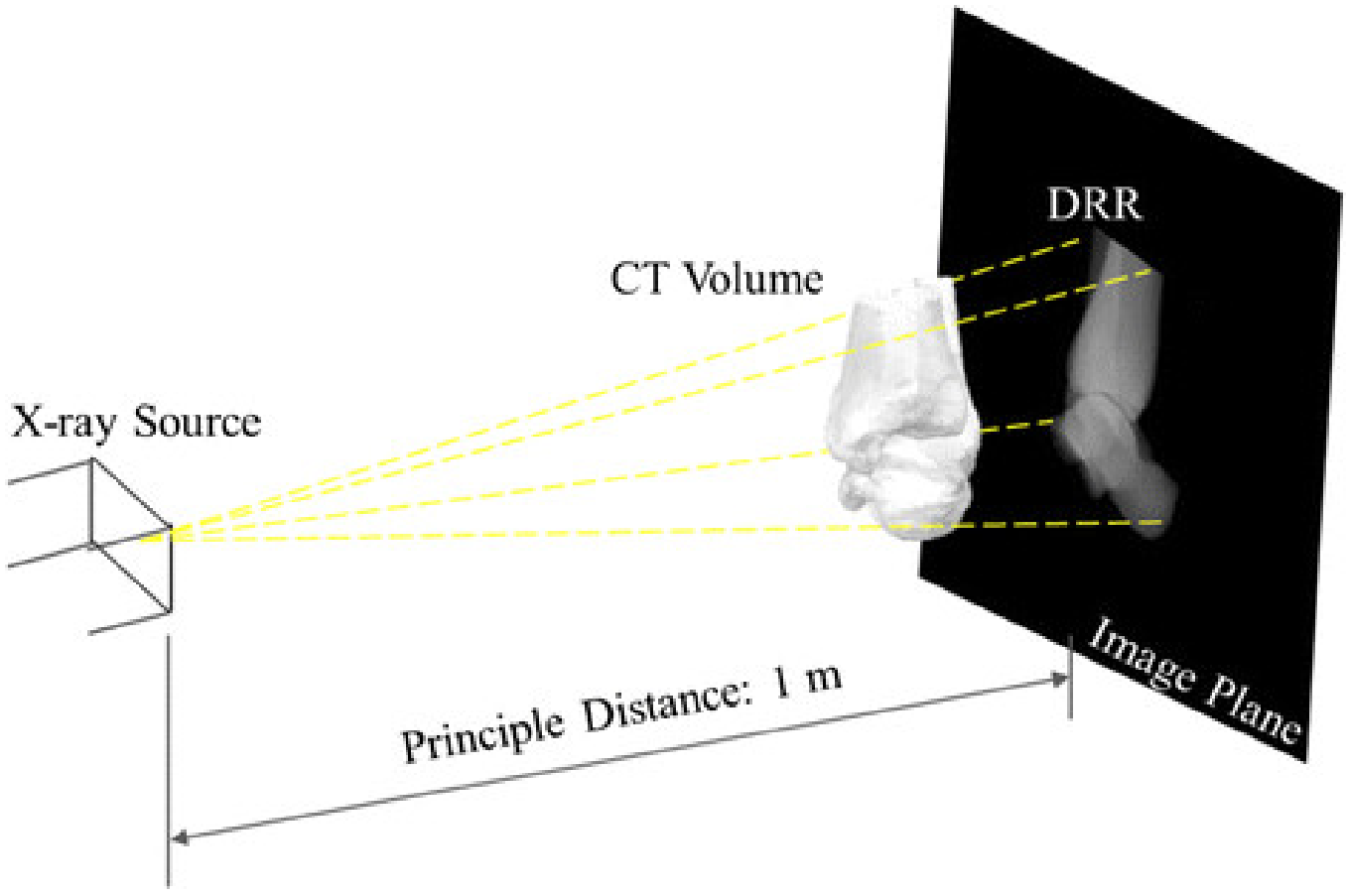
- *A System and Method for Projecting Synthetic Nodules in Medical Imaging*
  Indian Patent Application No. 202521024259 **(Filed)**

[†] Equal contribution

Motivation

- Early lung-cancer cues (subtle nodules) are hard to see on CXRs; labeled datasets are scarce.

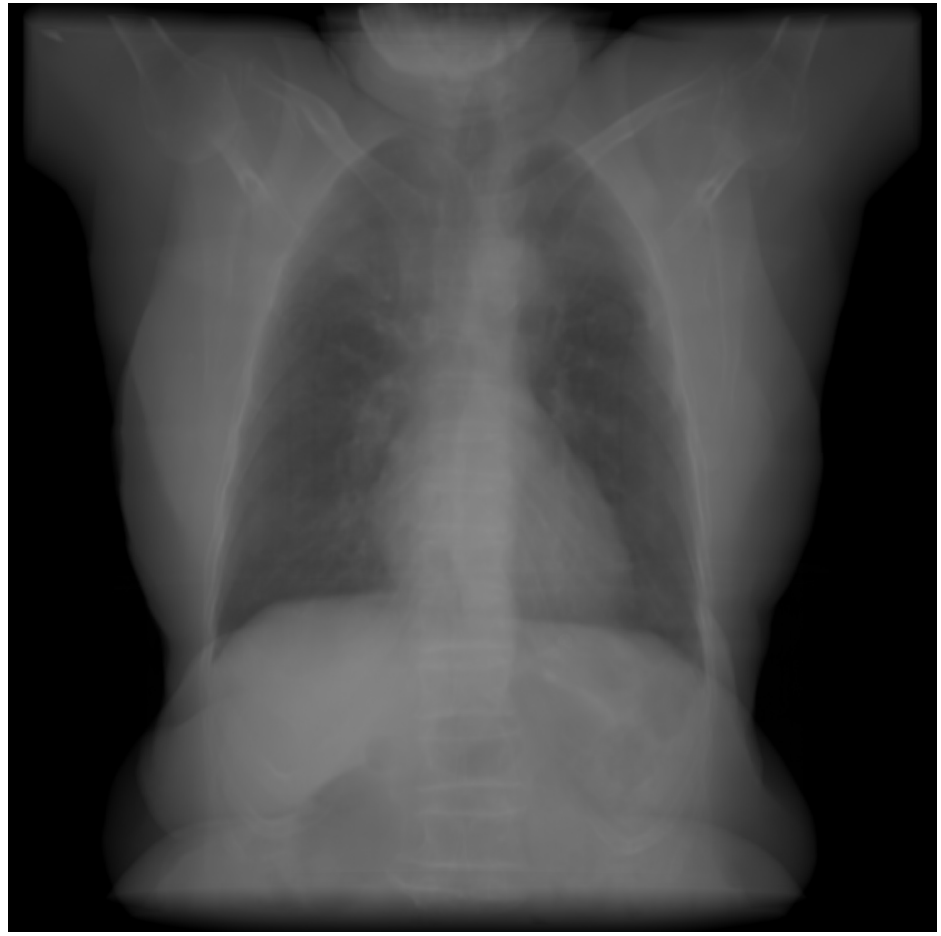
DRRs offer a solution



DRRs (CT-projected X-rays) are scalable but suffer blur, loss of fine structures, artifacts.

DRR = Digitally Reconstructed Radiographs

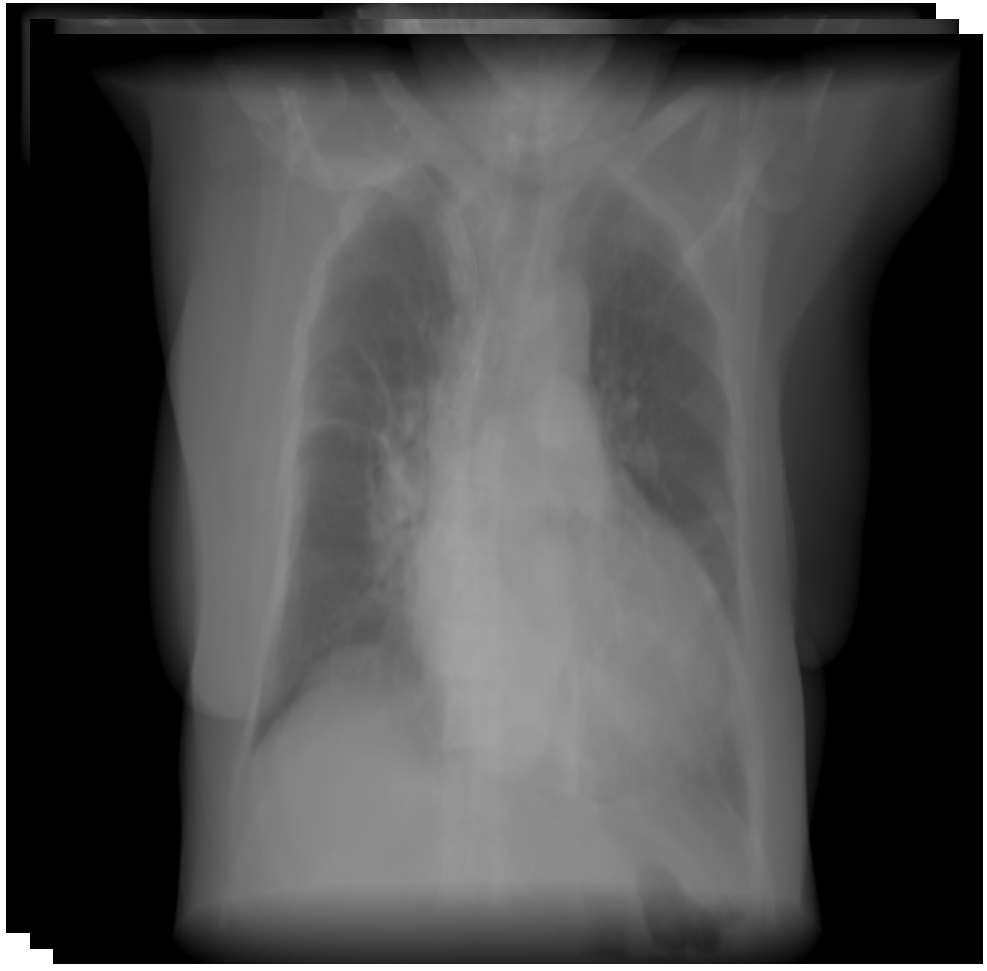
DRRs Obtained via Projection



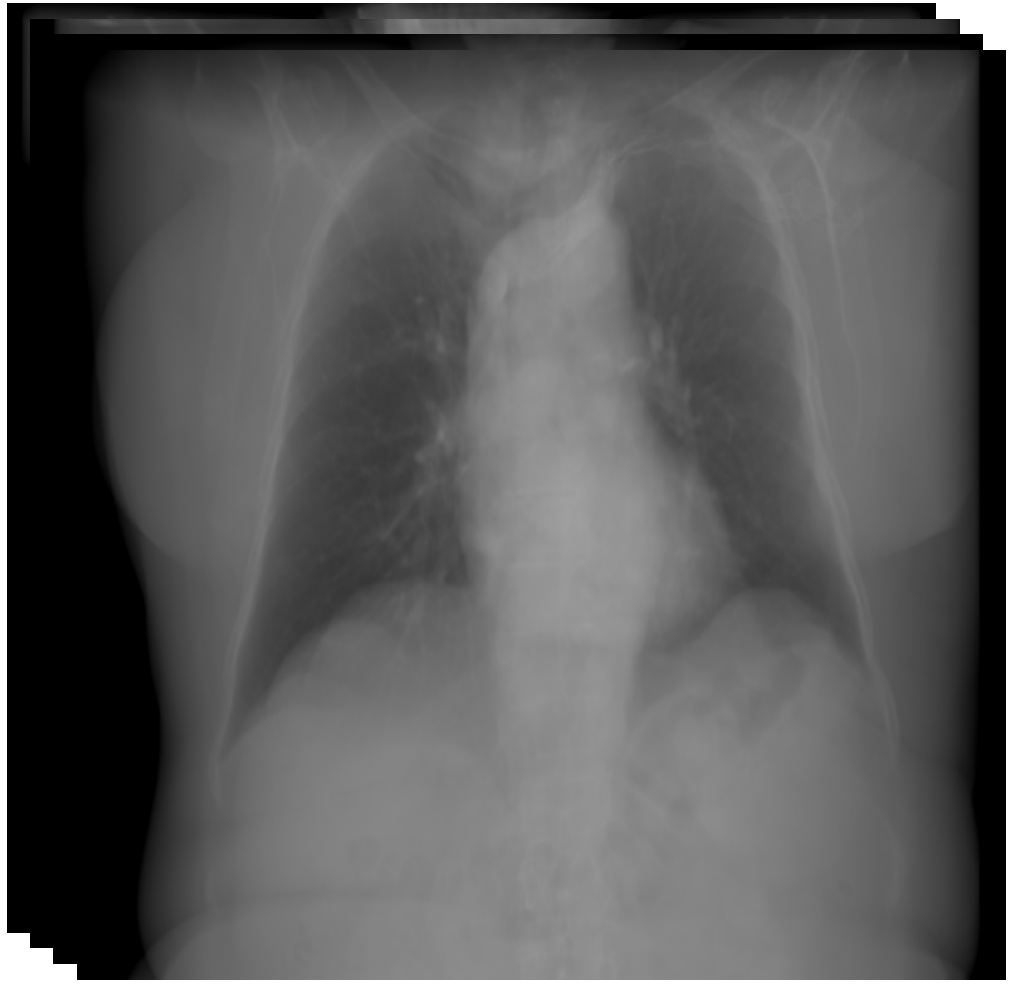
DRRs Obtained via Projection



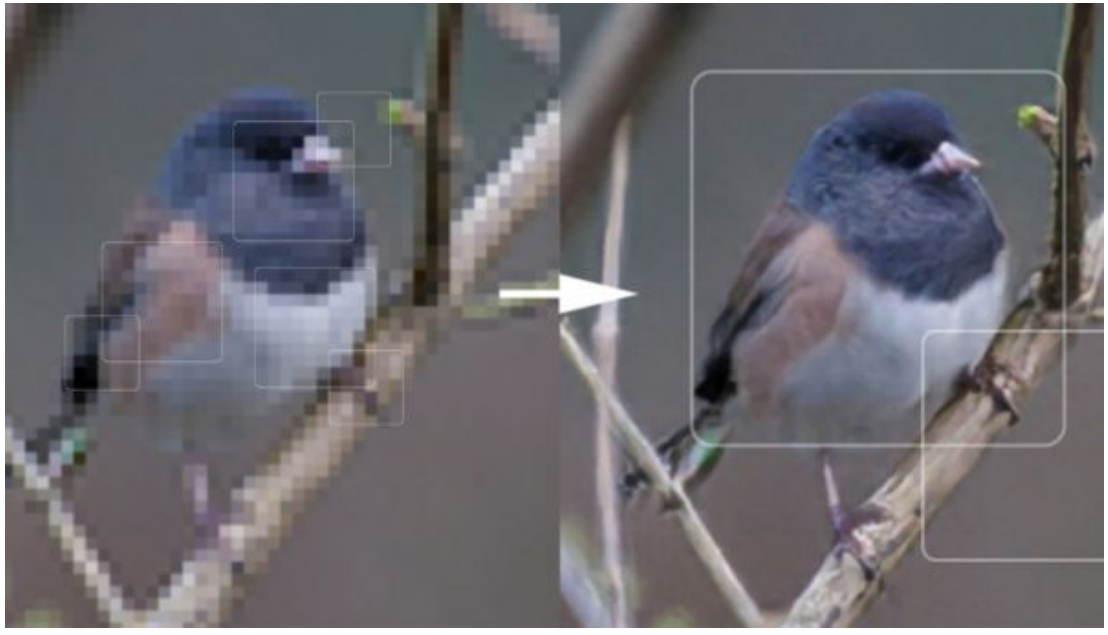
DRRs Obtained via Projection



DRRs Obtained via Projection



Super-Resolution/enhancement Models



Let's generate paired data using HQ X-rays and train a super-resolution/enhancement model, simple enough?

Results from Transformer-Based Methods



SwinFIR (Zhang et al., 2023), SwinIR (Liang et al., 2021)

Results from Transformer-Based Methods



SwinFIR (Zhang et al., 2023), SwinIR (Liang et al., 2021)

Results from Transformer-Based Methods



SwinFIR (Zhang et al., 2023), SwinIR (Liang et al., 2021)

Results from Bicubic Interpolation



Input

Results from Bicubic Interpolation



Input



Bicubic
Interpolation

Results from Bicubic Interpolation



Input

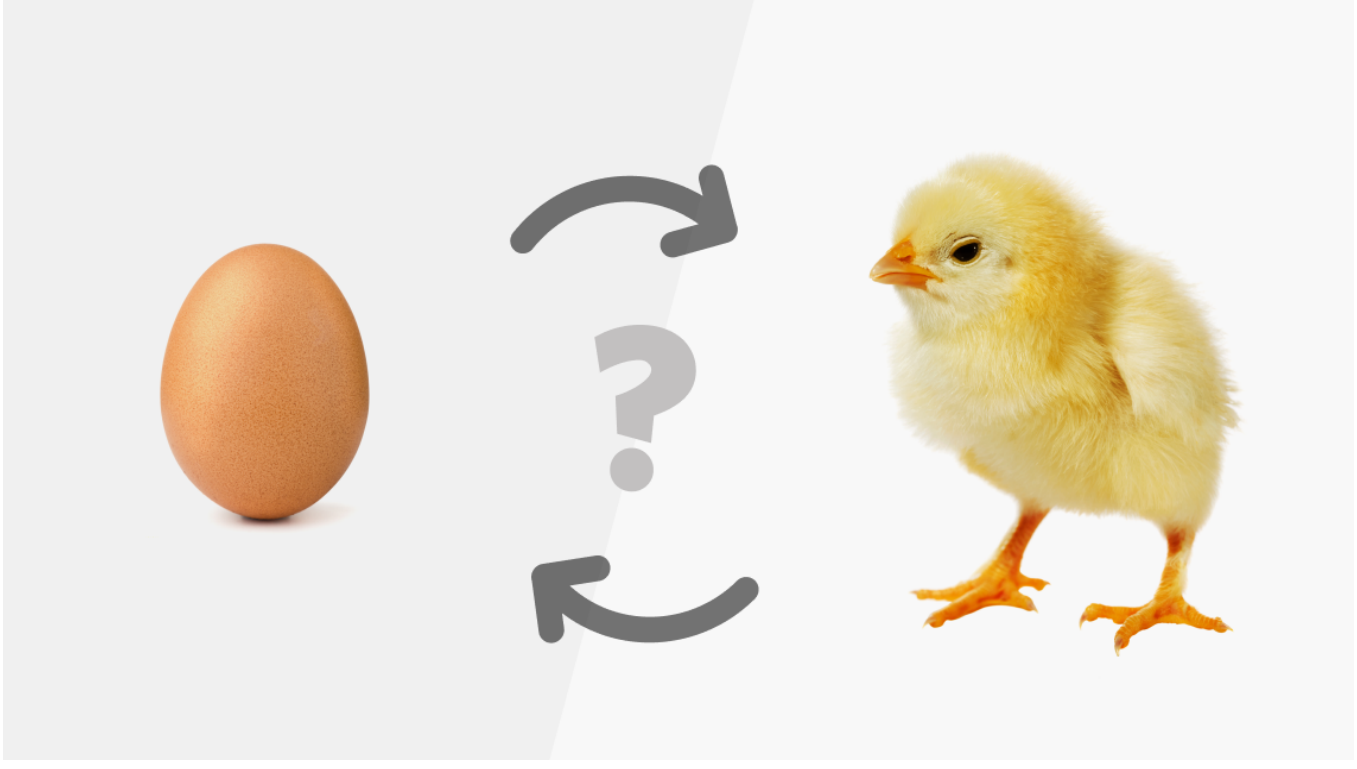


Bicubic
Interpolation



Reference

Stuck!



Helpful Papers on Domain Adaptation

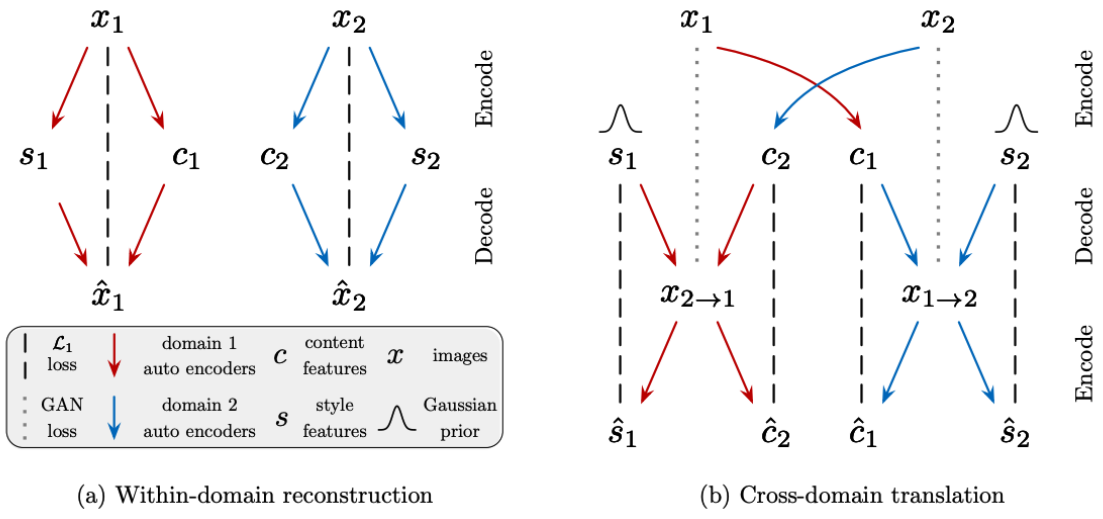
- **Synthesizing Realistic Image Restoration Training Pairs** Tao Yang et al., arXiv 2023 arXiv:2303.06994 — Diffusion-based generation of realistic LR–HR training pairs.
- **Towards Realistic Data Generation for Real-World Super-Resolution** Long Peng et al., arXiv 2025 arXiv:2406.07255 — Addresses domain gap in real-world SR via realistic data synthesis.
- **Multimodal Unsupervised Image-to-Image Translation (MUNIT)** Xun Huang et al., ECCV 2018 arXiv:1804.04732 — Multimodal I2I framework widely used for domain adaptation.

DiffusionXRay: A 2-Stage Pipeline

1. Generate realistic LQ CXRs from HQ:
   - MUNIT (Multimodal Unsupervised Image-to-Image Translation)-LQ
   - DDPM-LQ (Denoising Diffusion Probabilistic Models).
2. Train DDPM (Denoising Diffusion Probabilistic Models)-HQ enhancer on resulting paired HQ–LQ data.

Methodology

MUNIT-LQ (unpaired style transfer)



(a) Within-domain reconstruction

(b) Cross-domain translation

- Disentangles content vs style; combines HQ content with LQ style.

Methodology

DDPM-LQ (diffusion degradation model)

- Conditional DDPM learns realistic degradations.
- Stage 1: train on real LQ DRRs (unconditional).
- Stage 2: fine-tune conditioned on HQ to generate paired LQ.

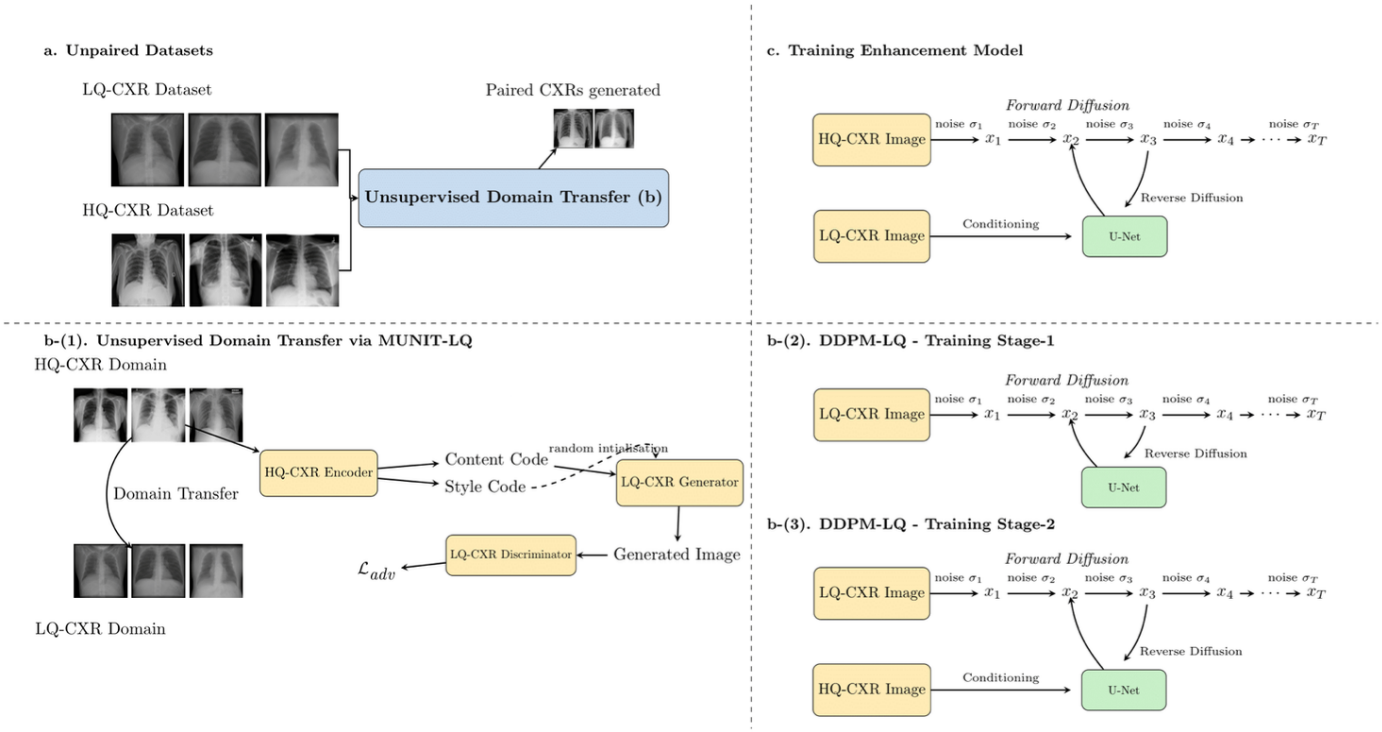


Fig. 1: Overview of our proposed DiffusionXRay framework. (a) Unpaired high and low-quality (CXRs) serve as training data for our unsupervised domain transfer models. (b) Implementation of domain transfer using: (b-1) MUNIT-LQ for style-guided translation or(b-2) DDPM-LQ for diffusion-based degradation modeling using 2 stage training. (c) The paired data generated from these models is subsequently used to train our final enhancement network.

Qualitative Enhancement Results

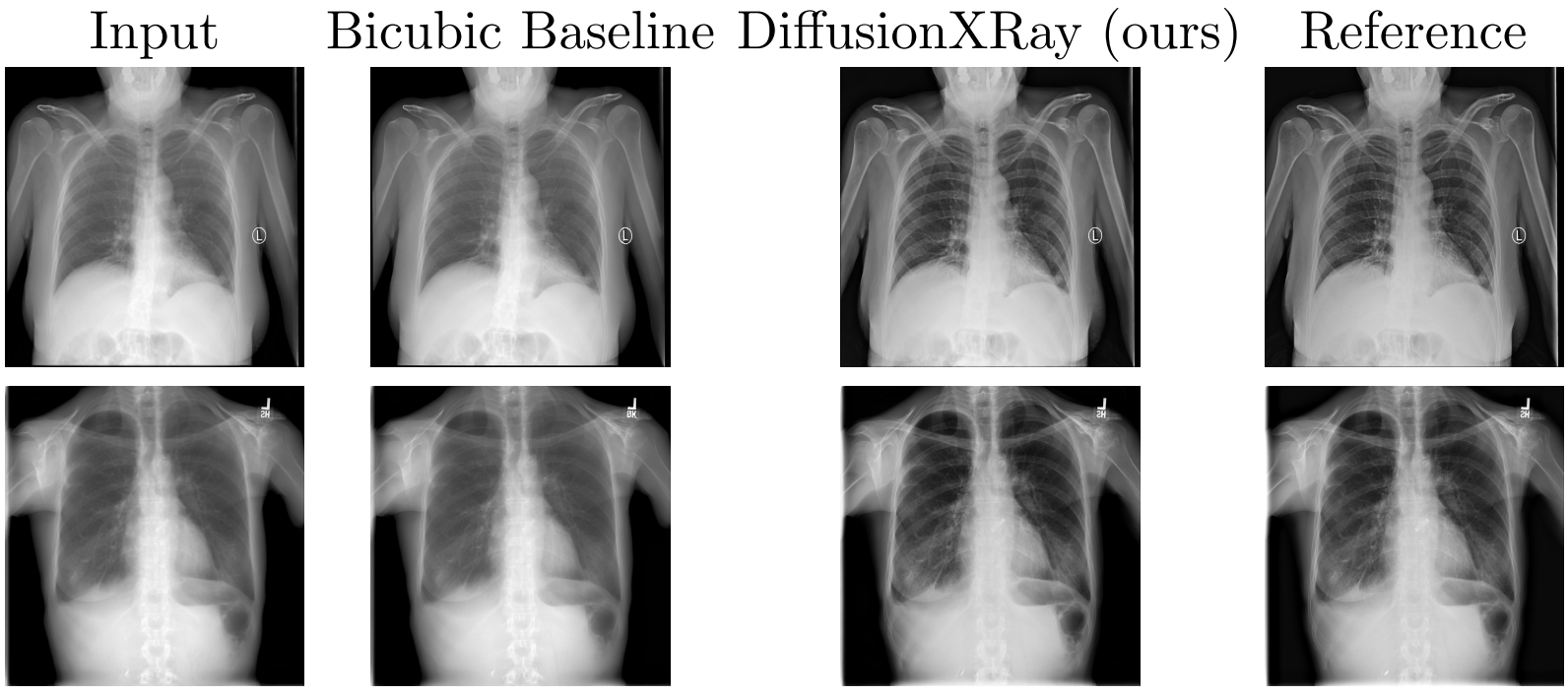


Figure 1: Qualitative comparison of image enhancement and super-resolution results (512×512 → 1024×1024) on synthetically degraded CXRs.

Qualitative LQ-CXR synthesis CXR Results

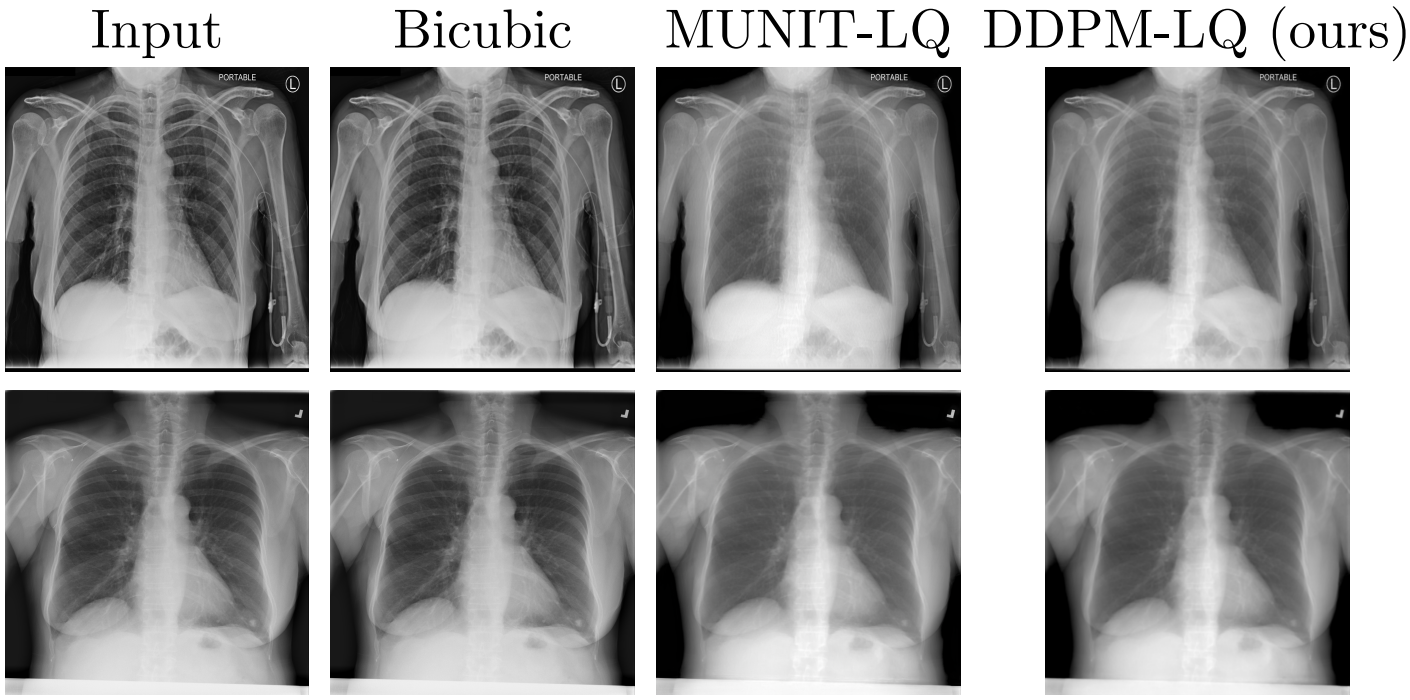


Figure 2: Qualitative comparison of low-quality CXR synthesis using Bicubic, MUNIT-LQ, and DDPM-LQ.

Quantitative (SR + Enhancement)

Setting	PSNR ↑	SSIM ↑
Results with MUNIT-LQ generated data		
Bicubic-DDPM	20.08	0.83
Diffusion-Xray (Ours)	**27.50**	**0.92**
Results with DDPM-LQ generated data		
DDPM-LQ (baseline)	19.85	0.78
DDPM-LQ (Ours)	**22.21**	**0.78**

Table 1: Comparison of different low-quality simulation strategies for chest X-rays. Higher PSNR and SSIM indicate better reconstruction quality.

SR = Super Resolution

Results: Radiologist Evaluation

Task	Question	DiffusionXRay (Ours)	Bicubic-DDPM
Task 1: Nodule preservationTask 1: Nodule preservationTask 1: Nodule preservationpt¡ -Task 1: Nodule preservationpt¿	Q1: Nodule easily visible? ↑	**100%**	6.6%
	Q2: Confusion with other structures? ↓	**0%**	30%
Task 2: Quality assessmentTask 2: Quality assessmentTask 2: Quality assessmentpt¡ -Task 2: Quality assessmentpt¿	Q1: Lung field clarity improved? ↑	**100%**	66.7%
	Q2: Significant noise increase? ↓	72.9%	25%

- 100% nodule visibility with 0% confusion ⇒ clinically meaningful preservation of subtle findings.
- Lung-field clarity improved in all cases vs. 66.7% baseline, consistent with higher SSIM under MUNIT-LQ training.
- Visibility gains come with 72.9% perceived noise, reflecting a purposeful trade-off favoring diagnostic conspicuity over smoothing.

Limitations & Future Work

Limitations

- Computational cost of training and inference is high.

Future Directions

- Lighter backbones or distillation.
- Extend to non-DRR scenarios such as portable X-rays.

Controllable Lung Nodule Synthesis

Motivation

- Nodules in CXRs come in various shapes, sizes and textures
- Assembling such representative datasets is hard

Related Work: Lung Nodule Synthesis (II)

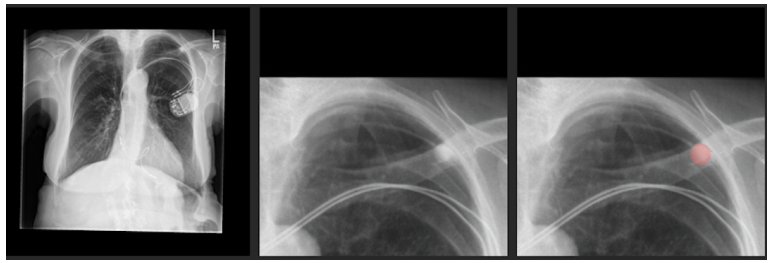
- Zhenrong Shen, Xi Ouyang, Bin Xiao, Jie-Zhi Cheng, Dinggang Shen, and Qian Wang. *Image Synthesis with Disentangled Attributes for Chest X-ray Nodule Augmentation and Detection.* Medical Image Analysis, 84:102708, 2022.

- Sebastian Gündel, Arnaud A. A. Setio, Saša Grbić, Andreas Maier, and Dorin Comaniciu. *Extracting and Leveraging Nodule Features with Lung Inpainting for Local Feature Augmentation.* In Proceedings of MICCAI Workshops, 2021.

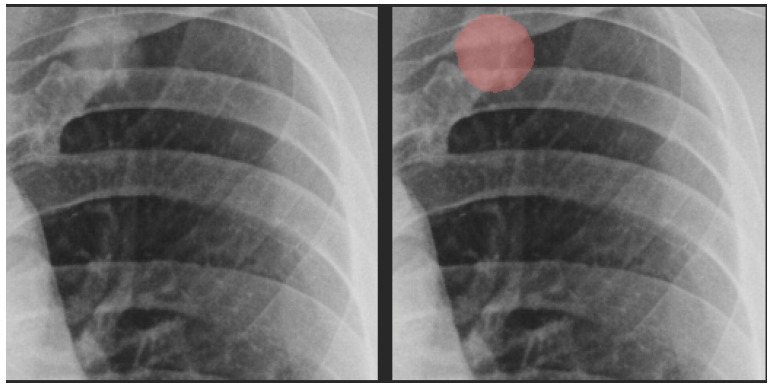
- Ecem Soganciogl u, Shi Hu, Davide Belli, and Bram van Ginneken. *Chest X-ray Inpainting with Deep Generative Models.* arXiv preprint arXiv:1809.01471, 2018.

Characteristic Definitions (I): Texture

Homogeneous Uniform density throughout the nodule.

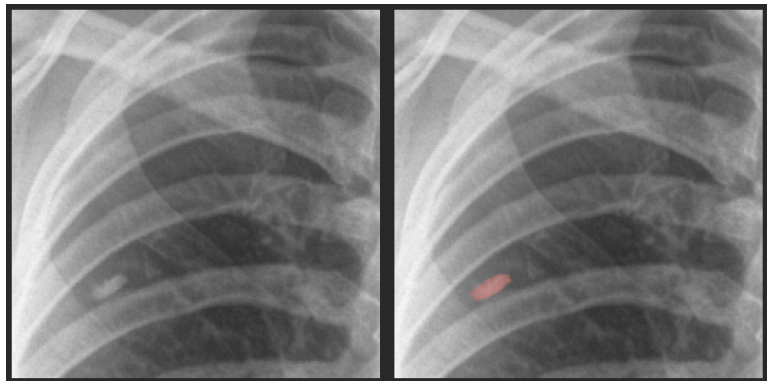


Inhomogeneous Uneven density with mixed bright and dark regions.

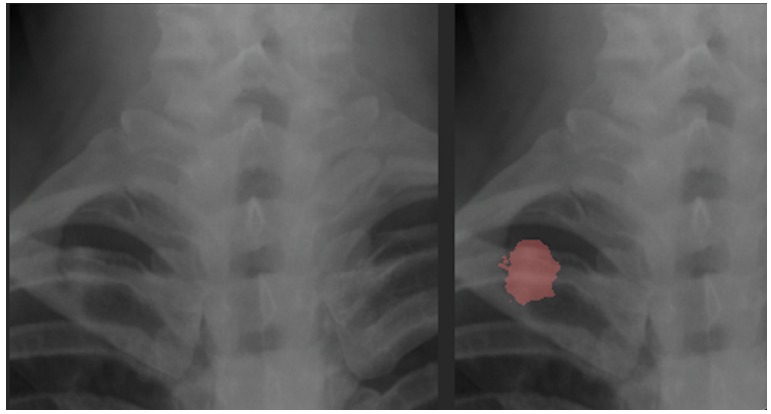


Characteristic Definitions (II): Boundary Morphology

Regular Smooth, well-defined, and clearly separated from lung tissue.



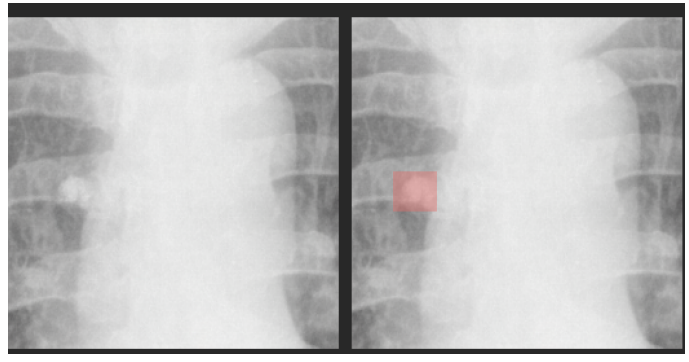
Irregular Spiculated, lobulated, or poorly defined margins.



Characteristic Definitions (II)

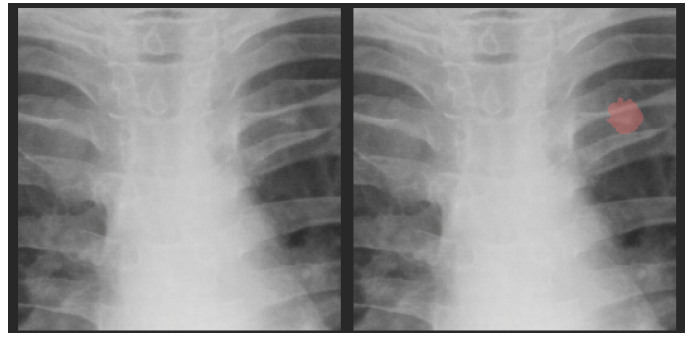
Calcification

Bright, high-intensity regions caused by calcium deposits, commonly seen in **benign nodules**.



Subtlety

Low-contrast nodules that blend with surrounding lung tissue, making them difficult to detect.



Method Overview

- Mask-conditioned Diffusion Transformer backbone
- Characteristic-specific LoRA adapters
- Orthogonality-constrained adapter merging

Training proceeds in stages:

1. Large-scale CXR pretraining
2. Nodule patch finetuning with spatial masks
3. Independent LoRA training per attribute
4. Joint training with orthogonality regularization

(a) Original patch (b) Nodule mask (c) Generated Patch

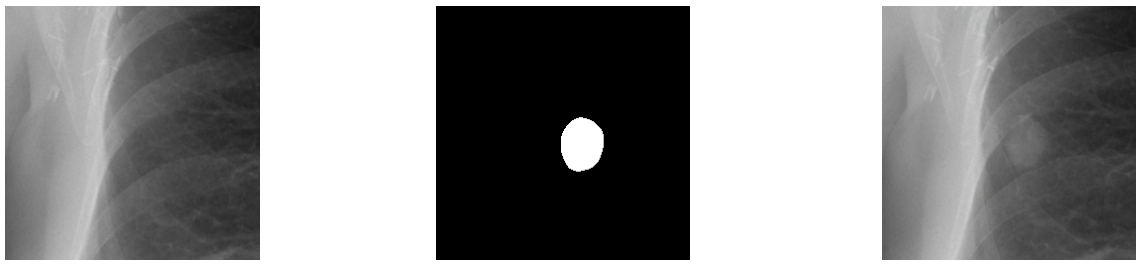


Figure 3: Given an original CXR patch and a binary nodule mask, the model generates a nodule within the masked region

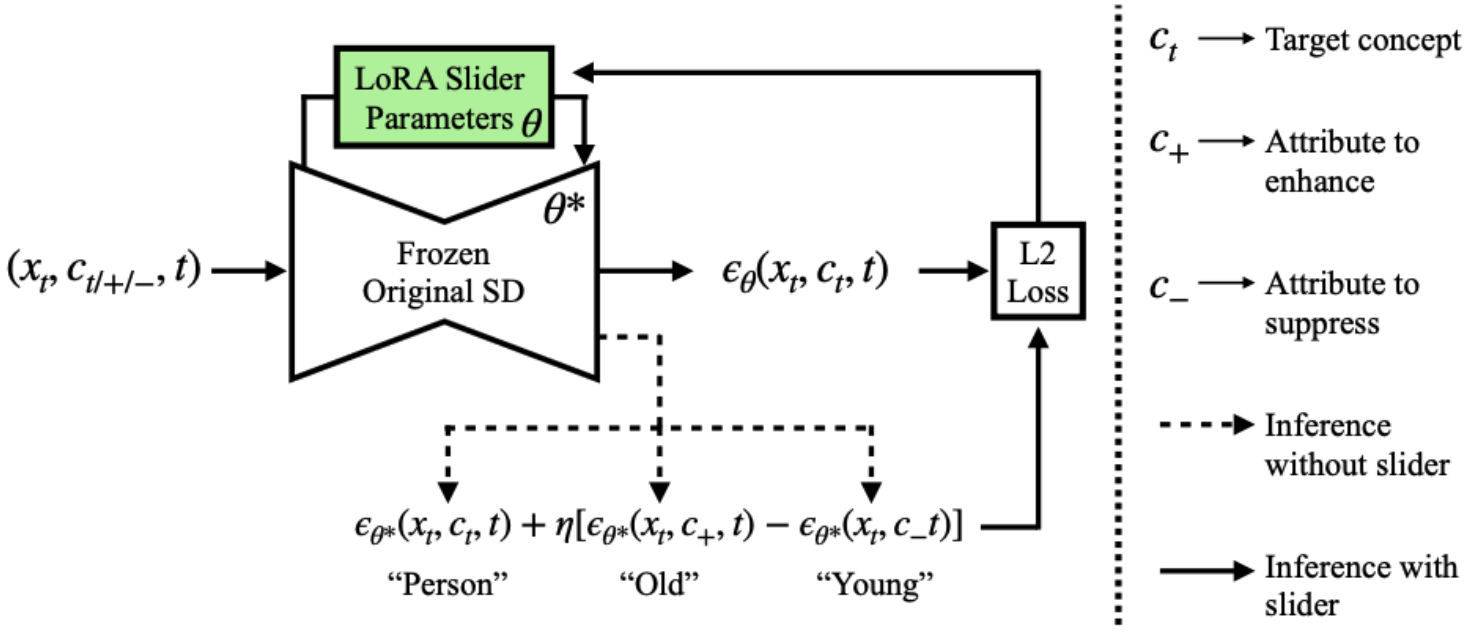
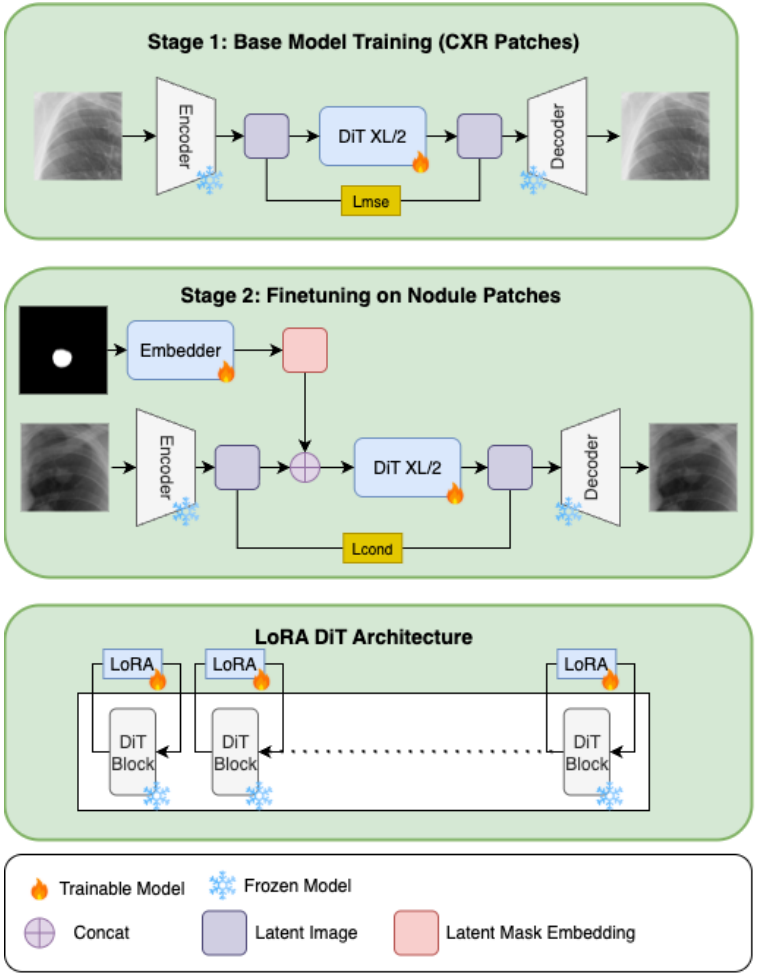
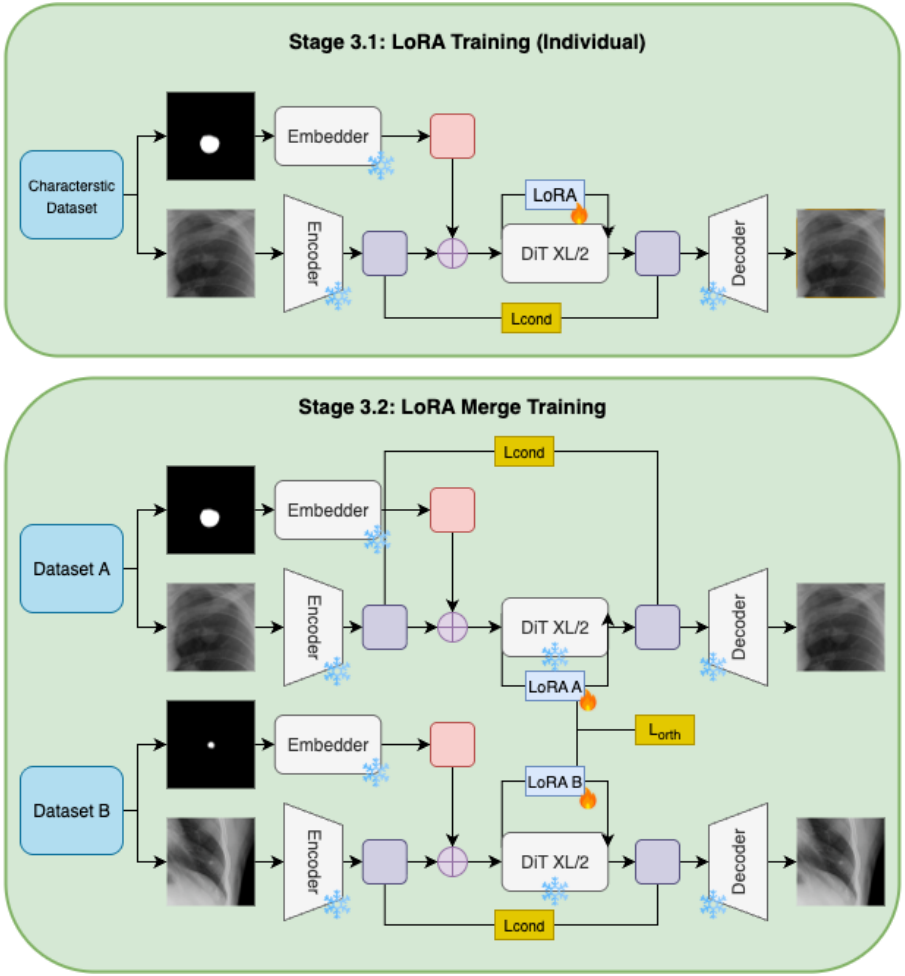


Figure 4: The slider model generates samples x_t by partially denoising Gaussian noise over timesteps 1 to t, conditioned on the target concept c_t.

Source: R. Gandikota et al., *Concept Sliders: LoRA Adaptors for Precise Control in Diffusion Models*, arXiv:2311.12092, 2023.

At a glance


Stage 1: Base Model Training (CXR Patches)
Encoder
DiT XL/2
Decoder
Lmse
Stage 2: Finetuning on Nodule Patches
Embedder
Encoder
DiT XL/2
Decoder
Lcond
LoRA DiT Architecture
LoRA
LoRA
LoRA
DiT Block
DiT Block
DiT Block
Trainable Model
Frozen Model
Concat
Latent Image
Latent Mask Embedding


Stage 3.1: LoRA Training (Individual)
Characterstic Dataset
Embedder
Encoder
LoRA
DiT XL/2
Decoder
Lcond
Stage 3.2: LoRA Merge Training
Dataset A
Dataset B
Embedder
Embedder
Encoder
Encoder
Lcond
DiT XL/2
DiT XL/2
Decoder
Decoder
LoRA A
LoRA B
Lorth
Lcond

Diffusion Backbone Formulation

Forward diffusion process:

$$q(x_t \mid x_{t-1}) = \mathcal{N}(\sqrt{1-\beta_t}x_{t-1}, \beta_t I)$$

Training objective:

$$\mathcal{L}_{diff} = \mathbb{E}_{x_0,t,\epsilon}\left[\|\epsilon - \epsilon_\theta(x_t, t, c)\|_2^2\right]$$

Conditioning c includes binary nodule masks to enforce spatial control.

Characteristic-Specific LoRA Adapters

For a frozen backbone weight W_0:

$$W = W_0 + \alpha AB$$

where:

- $A \in \mathbb{R}^{d \times r}, B \in \mathbb{R}^{r \times k}$
- $r \ll \min(d, k)$
- α controls attribute strength

Separate adapters were trained for:

- Calcification
- Border regularity
- Texture homogeneity
- Subtlety

Subtlety Control via Continuous LoRA Scaling

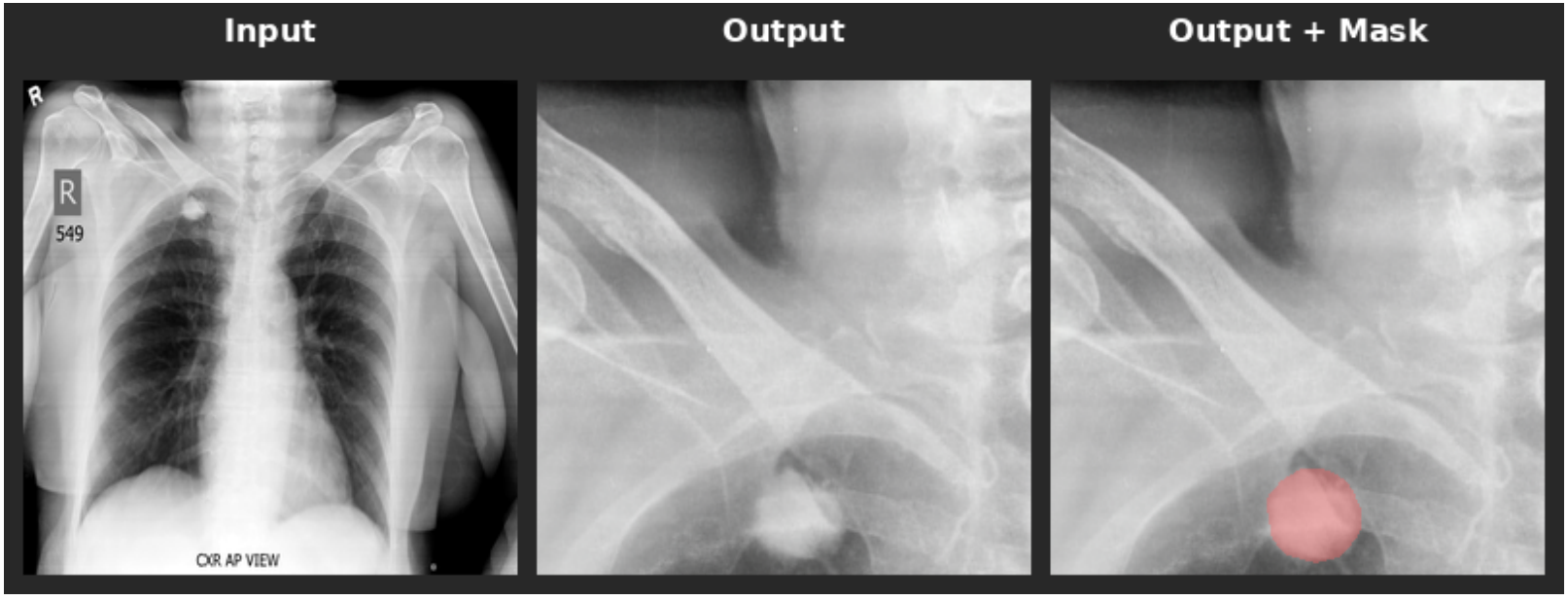
Radiologist-annotated subtlety scores $s \in \{1, \ldots, 5\}$ were mapped as:

$$\alpha(s) = 22 + s$$

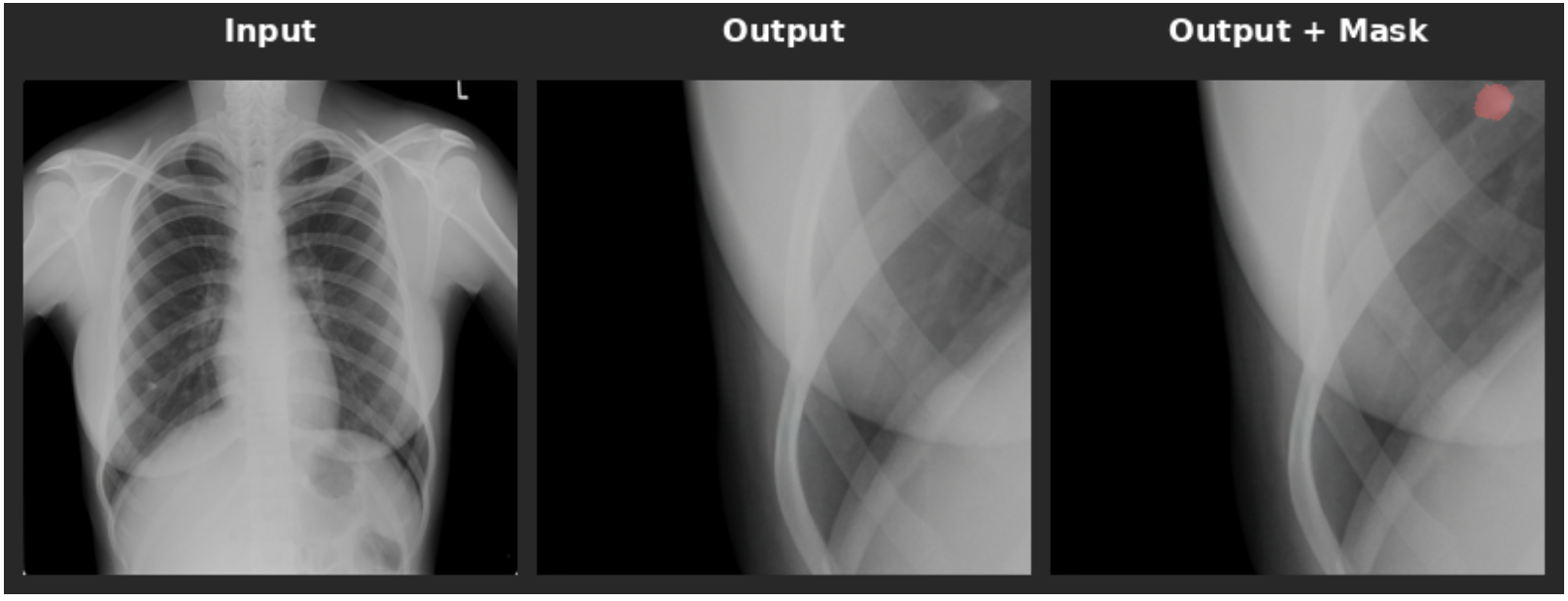
Lower α produces fainter nodules, higher α increases visibility. This enables:

- Ordered perceptual progression
- Clinically meaningful difficulty control

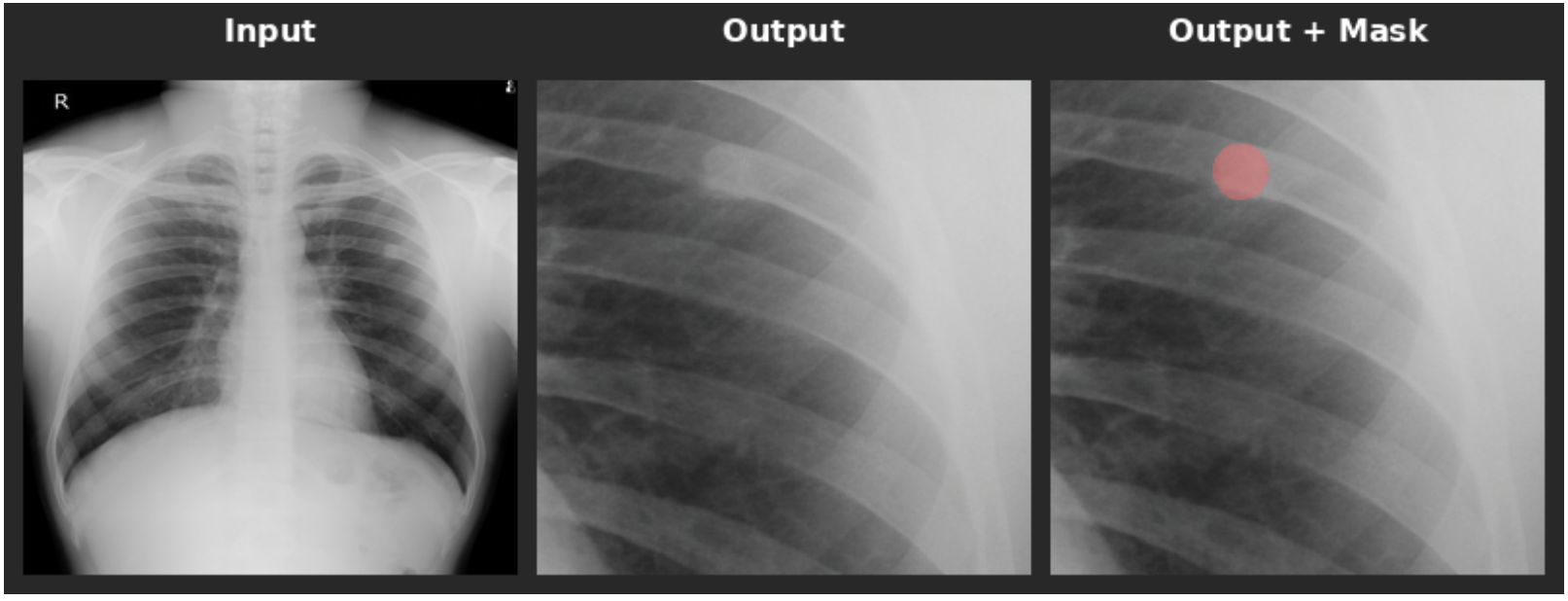
Texture — Homogeneous



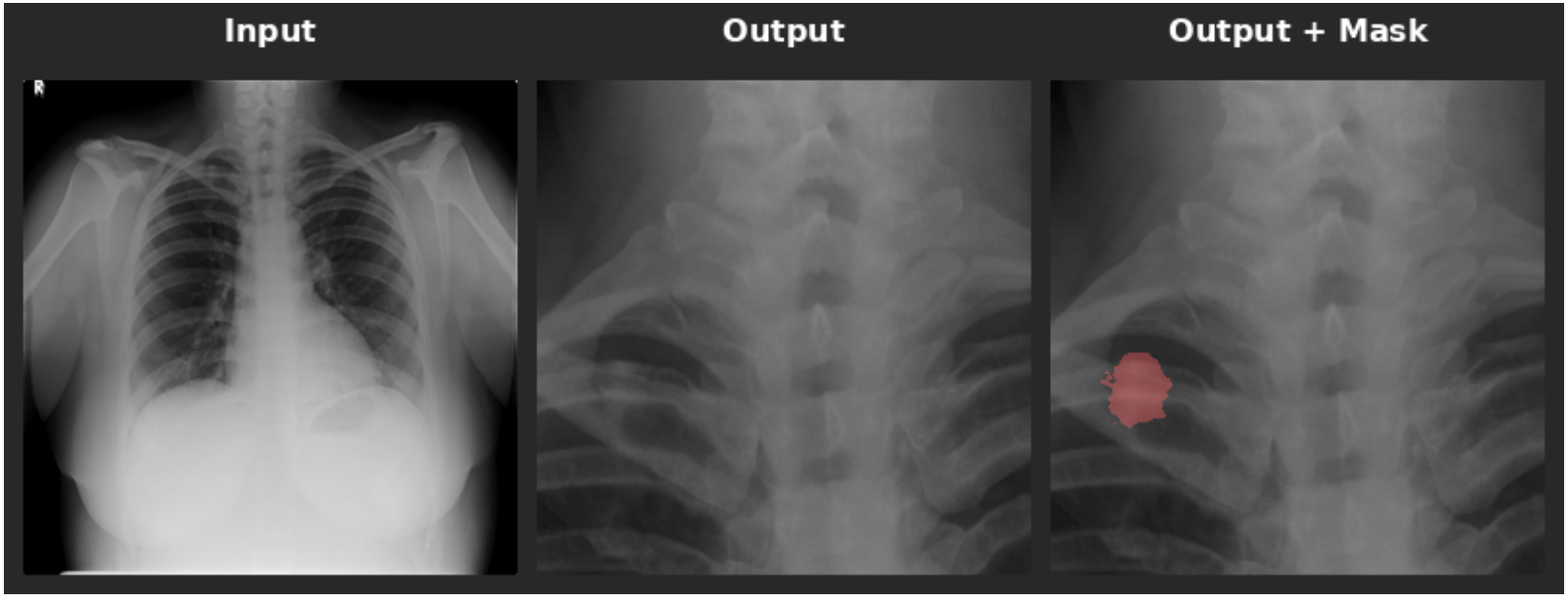
Texture — Inhomogeneous



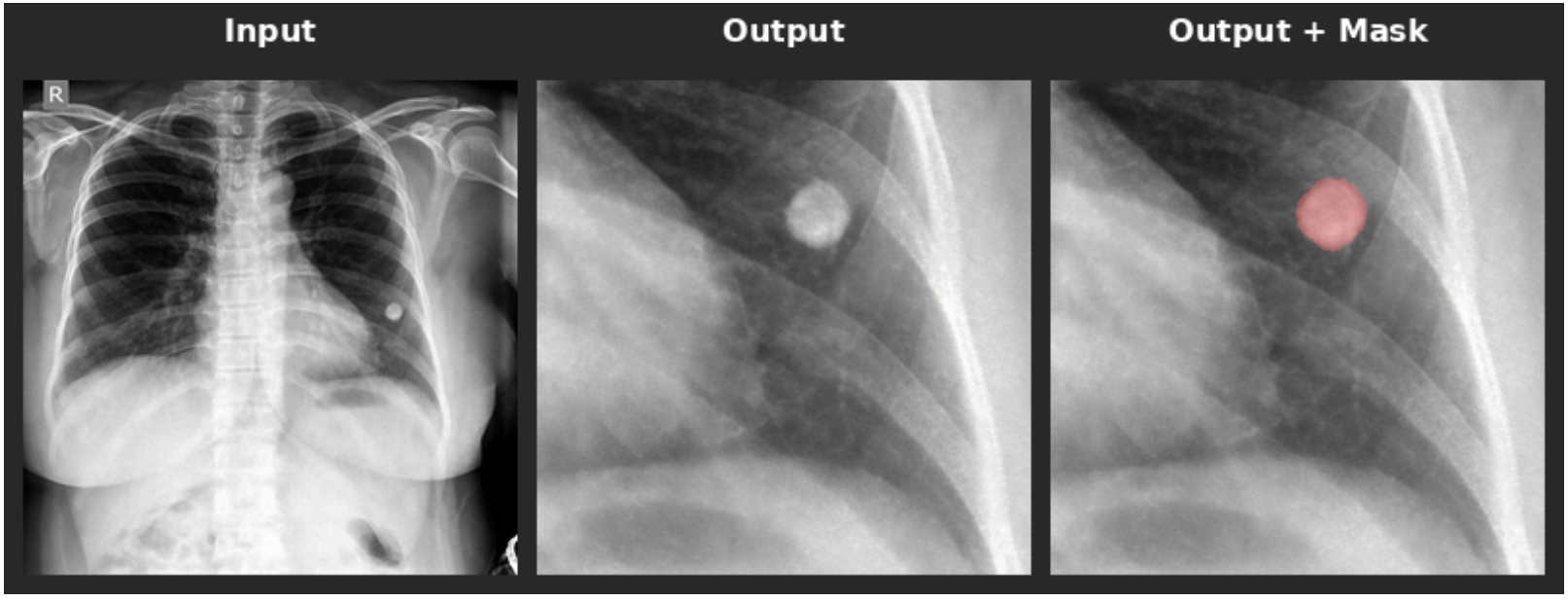
Boundary Morphology — Regular



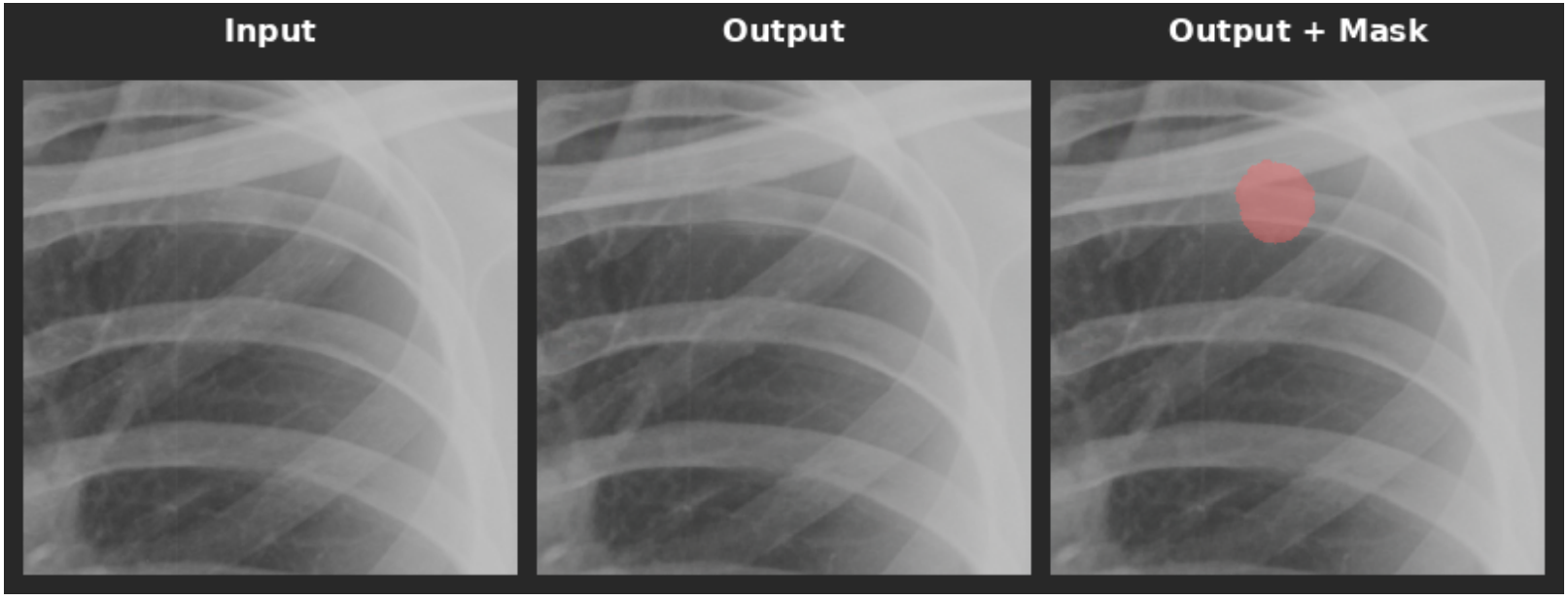
Boundary Morphology — Irregular



Calcified



Subtle Nodules



Radiologist Evaluation — Task 2: Controllability

For each characteristic, 10 nodules were generated using characteristic-specific LoRA adapters. Three radiologists verified whether the intended trait was present.

Agreement Rates:

Nodule Characteristic	**Agreement (%)**
Calcification	80
Regular border	90
Irregular border	100
Homogeneous texture	90
Inhomogeneous texture	100

Did you notice something about the characteristics?

Nodule characteristics are *not mutually exclusive*.

A regular nodule can be homogeneous or inhomogeneous, while being subtle or calcified.

What if we could control these combinations?

ex - I want to generate irregular-shaped nodules,
hence I will use irregular masks, but I also want to control
the texture as well

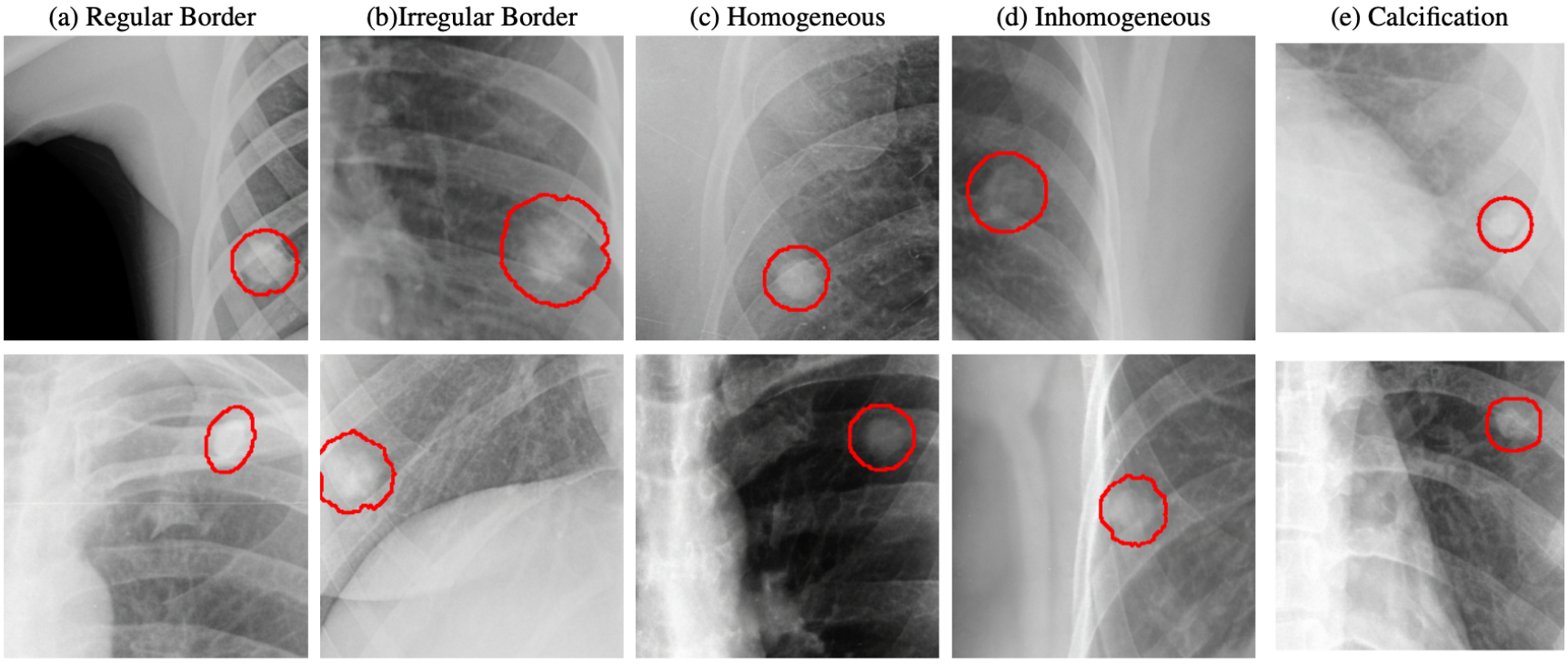


Figure 5: Results of characteristic-specific LoRA training, more than one characteristic!

This is where the design choice of LoRA pays off!!

- Characteristic specific LoRA adapters are used to steer the generation process
- Already explored with a diffusion backbone with multiple adapters in prior work[1].

[1] R. Gandikota et al., *Concept Sliders: LoRA Adaptors for Precise Control in Diffusion Models*, arXiv:2311.12092, 2023.

bad outputs?

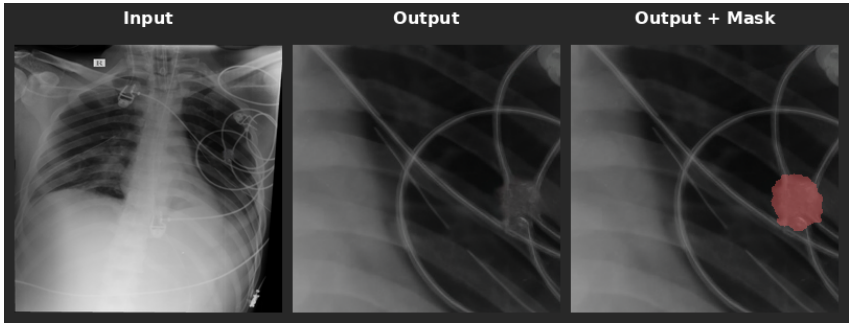


Figure 6: looks very synthetic, not a good sample!

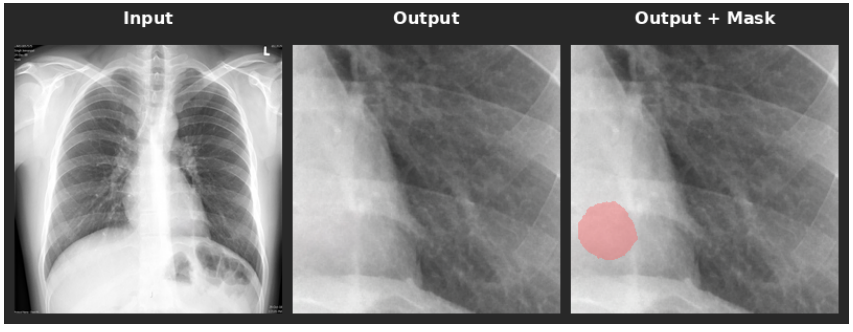
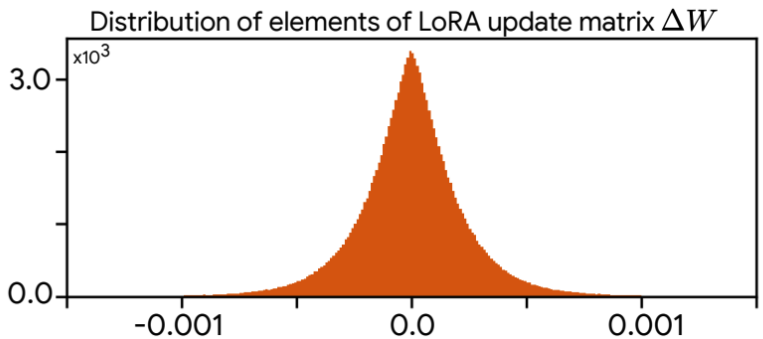


Figure 7: Calcified and Homogeneous, but not calcified?

ain't that simple

- Conventional merging strategies such as linear merge and switching assume that LoRA updates combine uniformly across layers.
- This can lead to:
  - one characteristic being suppressed by another
  - weird or nonsensical outputs
- A key reason is that LoRA weights are highly sparse (60–70% near zero, $|w| < 0.01$), meaning that a small subset of parameters drives most perceptual changes[2].



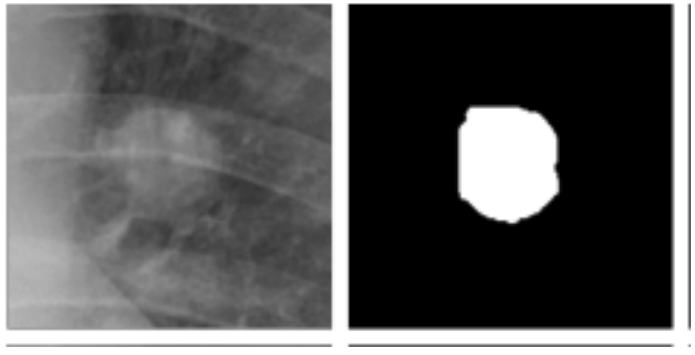
[2] Z. Ouyang, Z. Li, and Q. Hou, *K-LoRA: Unlocking Training-Free Fusion of Any Subject and Style LoRAs*, arXiv:2502.18461, 2025.

Why does this happen

- overlapping attention region



vs

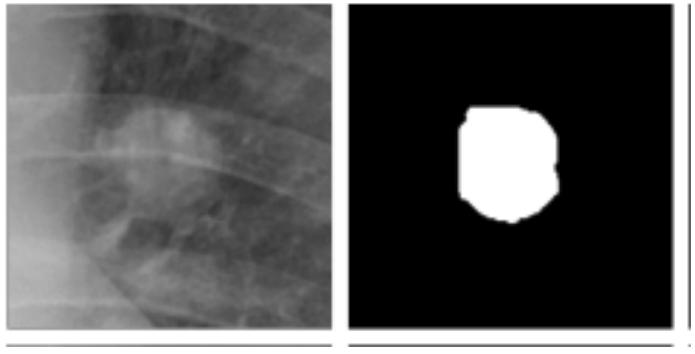


Why does this happen

- overlapping attention region



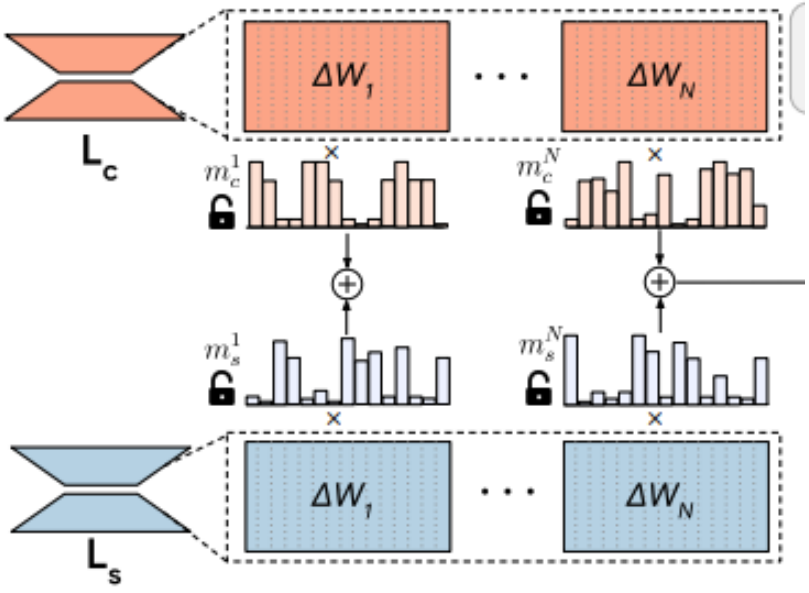
vs



can't be fixed in our case unfortunately :(

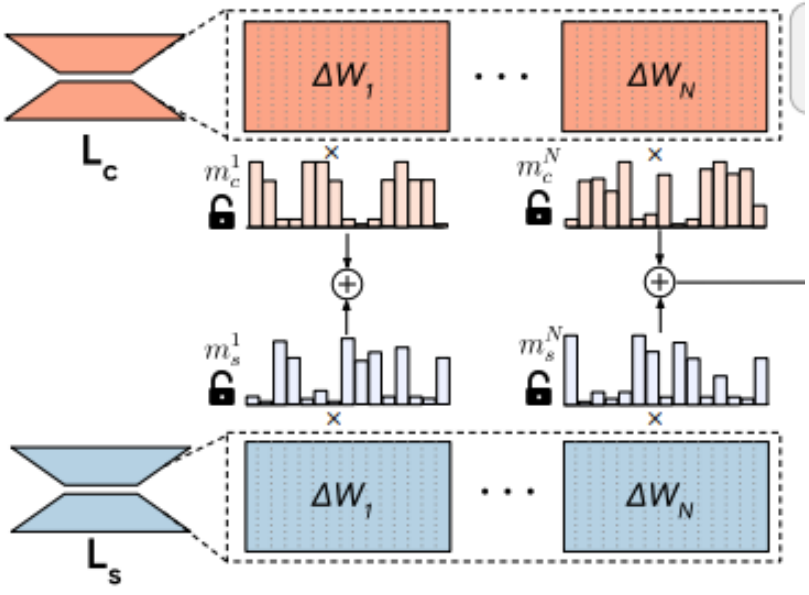
Why does this happen

- non-orthogonal weights



Why does this happen

- non-orthogonal weights



This can be fixed :)

Related Work (LoRA Merging)

- Viraj Shah et al. *ZipLoRA: Any Subject in Any Style by Effectively Merging LoRAs.* arXiv preprint arXiv:2311.13600, 2023. Merges multiple LoRAs by learning layer-wise scaling coefficients that selectively combine subject and style adapters.
- Ziheng Ouyang, Zhen Li, and Qibin Hou. *K-LoRA: Unlocking Training-Free Fusion of Any Subject and Style LoRAs.* arXiv preprint arXiv:2502.18461, 2025. Performs training-free LoRA fusion by reweighting sparse LoRA parameters based on their magnitude and layer sensitivity.
- Shenghe Zheng et al. *Decouple and Orthogonalize: A Data-Free Framework for LoRA Merging.* arXiv preprint arXiv:2505.15875, 2025. Applies orthogonalization to independently trained LoRA updates at merge time to reduce interference between adapters.

All methods perform LoRA conflict resolution at merge time, whereas

Orthogonality-Constrained LoRA Merging

Naive LoRA merging causes interference due to correlated updates.
We introduce:

$$\mathcal{L}_{orth} = \|W_1 W_2^\top\|_F^2$$

This encourages:

- Decoupled parameter subspaces
- Stable multi-attribute composition
- Reduced spatial competition

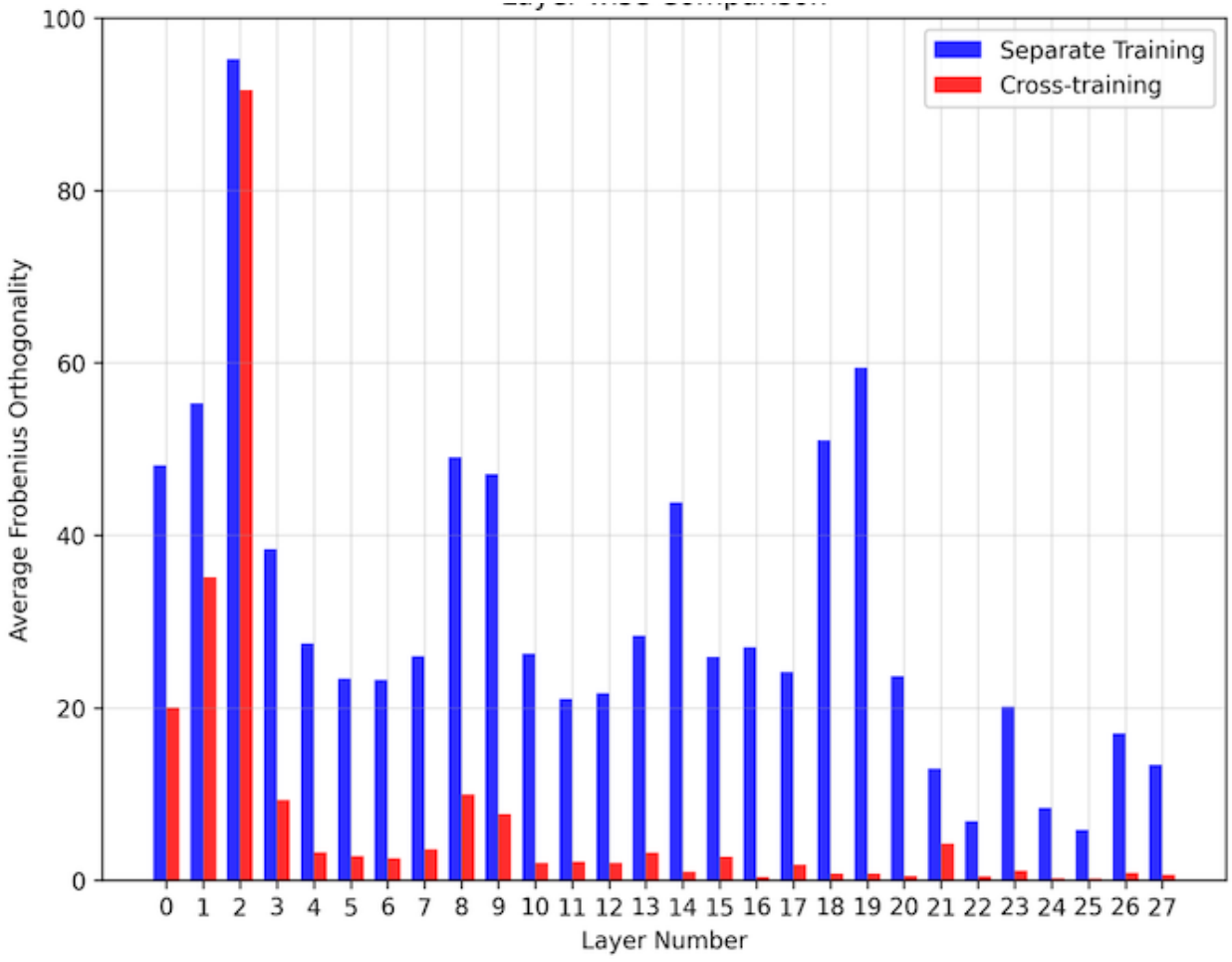


Figure 8: Layer-wise Frobenius norm comparison of cross-trained vs. separately trained adapters across two characteristic pairs.

Radiologist Evaluation — Task 1: Realism

50 pulmonary nodules were shown to 3 expert radiologists (25 real + 25 synthetic from our base diffusion model).

Results:

- 90% of real nodules were correctly identified as real
- 80% of the synthetic nodules were also labeled as real

Qualitative

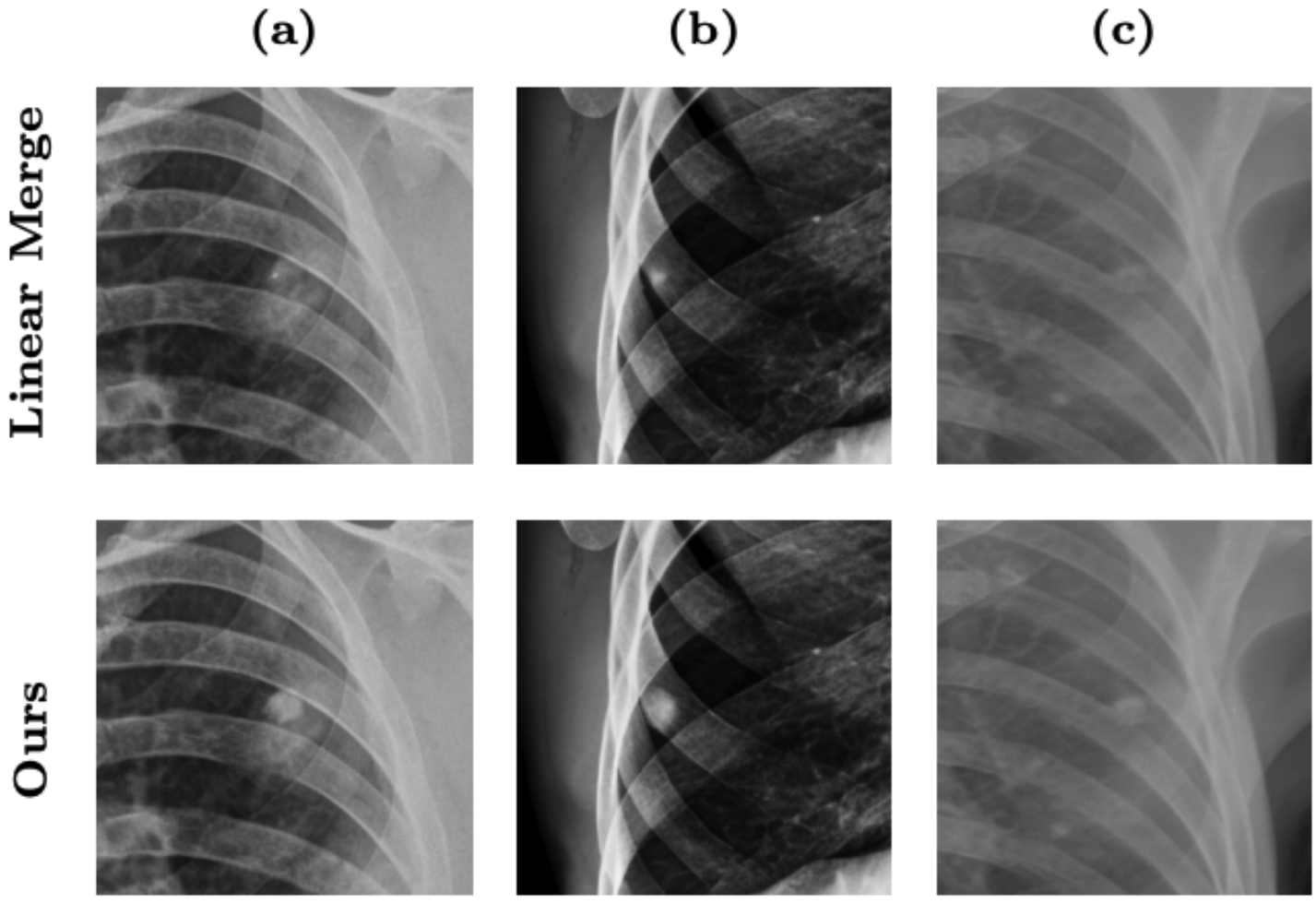


Figure 9: (a) calcified + irregular, (b) calcified + homogeneous, (c) regular + homogeneous

Qualitative

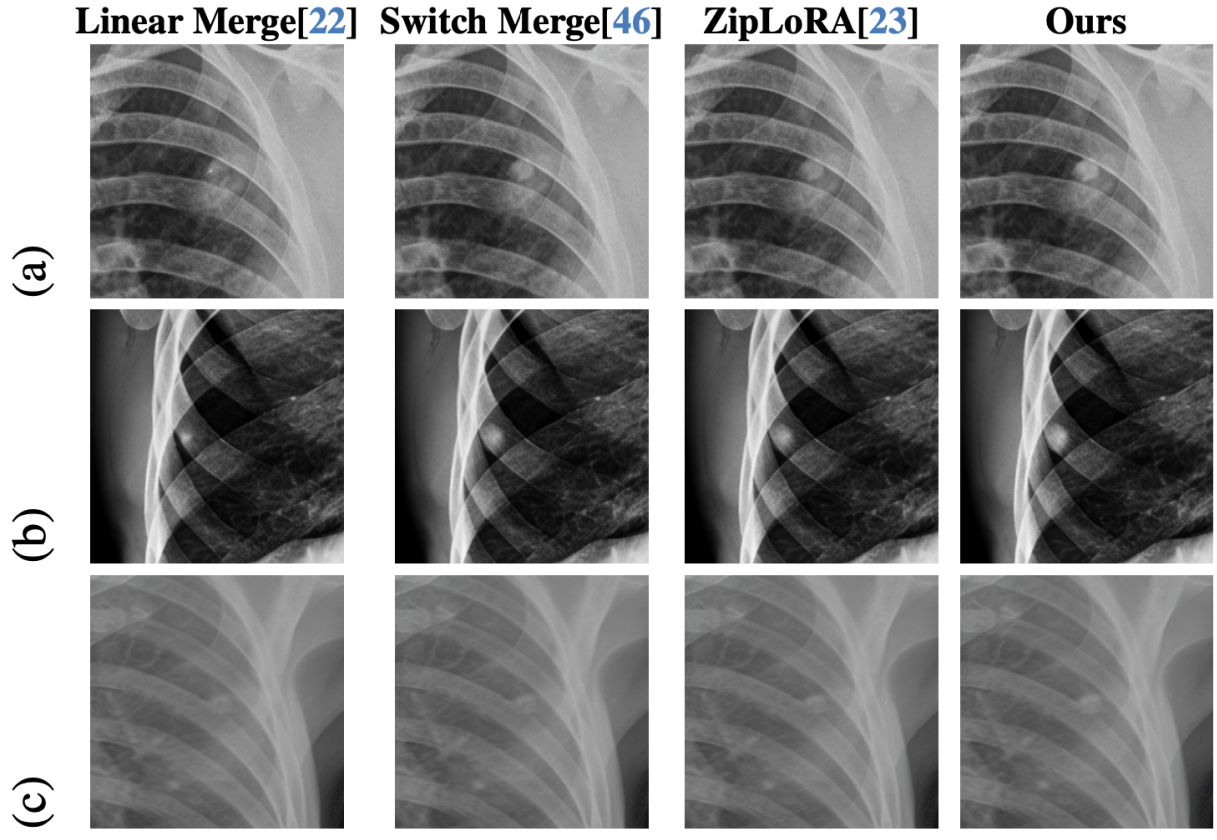


Figure 10: (a) calcified+irregular, (b)calcified+homogeneous, (c)irregular+homogeneous. Rows in- dicate configurations; columns show methods

Downstream Task Performance - nodule detection and segmentation

Train Data	**In-house**		**JSRT**		**CheX-ray14**	
	AUC	IoU	AUC	IoU	AUC	IoU
10k real	0.9705	0.3090	0.8560	0.2475	0.9008	0.5285
10k real + 2k syn	0.9788	0.3222	0.8639	0.2743	0.9168	0.5293
10k real + 4k syn	0.9780	0.3197	0.8780	0.2589	0.9245	0.5500
10k real + 6k syn	**0.9802**	0.3247	0.8940	0.2894	0.9315	0.5750
10k real + 8k syn	0.9796	**0.3274**	0.8864	0.2923	**0.9341**	**0.5954**
10k real + 10k syn	0.9801	0.3056	**0.9023**	**0.3091**	0.9318	0.5613

- Adding synthetic data consistently improves performance across datasets, especially on **JSRT** and **CheX-ray14**.
- Best performance is achieved with **6k–8k synthetic samples**, beyond which gains saturate or slightly regress.

Metric	Characteristic	5k Real	5k Real + 2k fake
AUC	Nodule	0.9531	0.9647
	Calcification	0.8338	0.8454
	Regular Border	0.8955	0.8994
	Irregular Border	0.6033	0.6522
	Homogeneous	0.7533	0.7649
	Inhomogeneous	0.7936	0.8099
IoU	Nodule	0.2696	0.3002
	Calcification	0.2879	0.3199
	Regular Border	0.2941	0.3301
	Irregular Border	0.2733	0.3080
	homogeneous	0.2695	0.3050
	Inhomogeneous	0.2706	0.2963

Figure 11: Characteristic-specific LoRA adapter evaluation

Swin-Tiny with a multi-head classifier was trained using ∼400 synthetic nodules per characteristic. Data augmentation improves both AUC and IoU across all radiological characteristics.

Augmentation	JSRT[45]	ChestX-ray14[44]
10k real	0.8560	0.9008
10k real + 10k ACGAN [1]	0.8780	0.9281
10k real + 10k ReACGAN [2]	0.8808	0.9259
10k real + 10k CRFILL [3]	0.8786	0.9296
10k real + 10k DiT-XL/2(Ours)	**0.9023**	**0.9318**

Figure 12: Comparison of the effect of synthetic-data augmentation on nodule AUC scores across ChestX-ray14 and JSRT

Fin.

Reach out to me:

X.com - @arygoy
garyan18@gmail.com