

Presentació Proves de Concepte en Dades i IA 27 maig 2026

Phase IV-AI LAND: Síntesi de TC toràctica 3D amb guiatge anatòmic mitjançant models de difusió de pulmons i nòduls



Anna Oliveras Tous

*investigadora de la Unitat Tecnològica Digital
Health Eurecat*

01

Problem & motivation

Why synthetic chest CT, and why it is still hard.

02

Related work

State of the art in 3D medical diffusion and where it breaks.

03

Vision & objectives

What we set out to build — and the constraints we accepted.

04

The LAND pipeline

Latent diffusion with anatomical mask conditioning, end to end.

05

Experiments & results

FID, MS-SSIM, memory budget, and texture control.

06

Conclusions & future work

What we shipped, what we learned, where we go next.

The data bottleneck.

Deep learning thrives on data. Medical imaging doesn't have it.

Acquisition cost & privacy regulations shrink the pool of usable volumes.

Expert annotation is slow and expensive — radiologist time is the rate-limiter.

Class imbalance: rare pathologies are exactly the cases AI most needs to see.

LUNG CANCER WORLDWIDE

#1 cause of cancer death

Globally, the leading cause of cancer-related mortality — and the modality most affected by data scarcity is also the most diagnostically critical.

LUNG NODULES

High variability in **size**, **shape**, **texture** and **location** means small per-class datasets and brittle downstream models.

State of the art & its ceiling.

Diffusion models lead image synthesis today. 3D medical diffusion is the new frontier — and the hardware bill grows with every cubic voxel.

MEDICAL DIFFUSION · 2023

3D MRI & CT synthesis. Limited to **3×**
lower resolution than LAND.

PATCHDDM · 2024

Patches conditioned on spatial coords. No
autoencoder. **~17.5 training days**.

WDM · 2024

Wavelet decomposition reduces spatial
dim. Still uses **~2×** the GPU memory of
LAND.

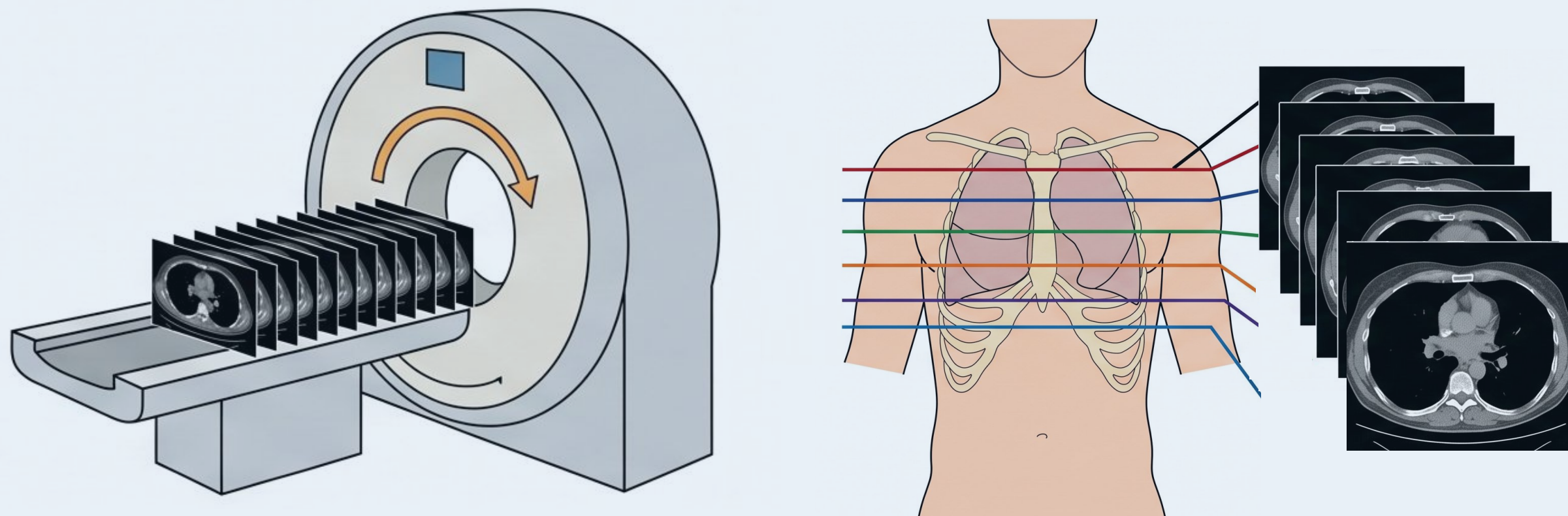
MAISI (NVIDIA) · 2025

512×512×768 CT — best-in-class
resolution. Requires **58 GB** of GPU
memory.

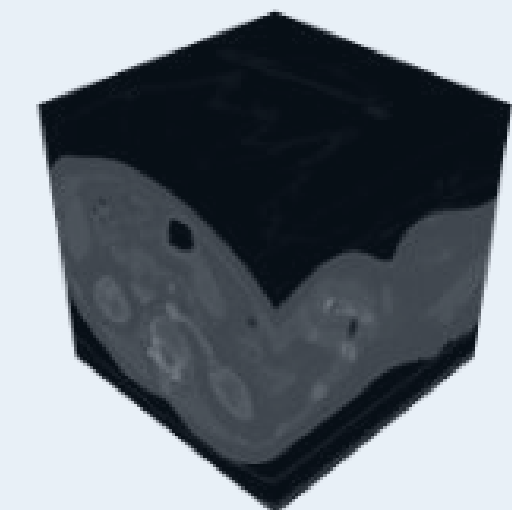
State of the art & its ceiling.

What is 3D medical imaging?

CT SLICES



CT VOLUME



256 x 256 x 256

State of the art & its ceiling.

3D medical image synthesis is a key tool for addressing **data scarcity**, **privacy concerns**, and **class imbalance** in medical AI — but the state of the art has three structural problems.

HARDWARE

Compute-heavy.

Existing 3D GANs and diffusion models require **multiple high-end GPUs** and long training times — limiting accessibility and scalability.

CONDITIONING

Limited inputs.

Conditioning is often based on lesion or nodule masks alone — leading to **anatomically implausible** or globally inconsistent chest structures.

CLINICAL IMPACT

Low realism.

These limitations reduce the clinical realism and downstream utility of synthesized CT scans for training, validation, and educational use.

Vision & objectives.

VISION

Scalable, controllable, resource-efficient synthesis of realistic 3D chest CT scans.

Lower the hardware barrier. Improve anatomical control. Make data augmentation, rare-case simulation, and educational use possible in computationally constrained research and clinical settings.

OBJECTIVES

Latent diffusion framework for conditional generation of 3D chest CT from anatomical masks.

Anatomical fidelity: measured by FID and MS-SSIM at full resolution.

Hardware floor: training and inference on a single mid-range GPU — no compromise on resolution.

Fine-grained control: location, size, and texture of nodules through the input mask.

The LAND pipeline.

A **latent diffusion model** for 3D chest CT synthesis that generates full-resolution **256 × 256 × 256** volumes at **1 mm isotropic spacing**, on a single mid-range GPU.

RESOLUTION

256³

at 1 mm isotropic — no compromise on spatial fidelity.

MEMORY BUDGET

20 GB

Single mid-range GPU — vs. 58 GB for MAISI & ~40 GB for WDM.

CONDITIONING

2-in-1

Global **lung anatomy** + local **nodule regions** in a single mask.

CONTROL

5 texture levels

Non-solid → solid nodules, dialed per-voxel via the mask value.

Experimental results highlight the importance of incorporating **global lung structure** to achieve anatomically coherent synthesis, and show that LAND can generate diverse CT volumes with and without pulmonary nodules of varying sizes, locations, and characteristics.

The LAND pipeline.

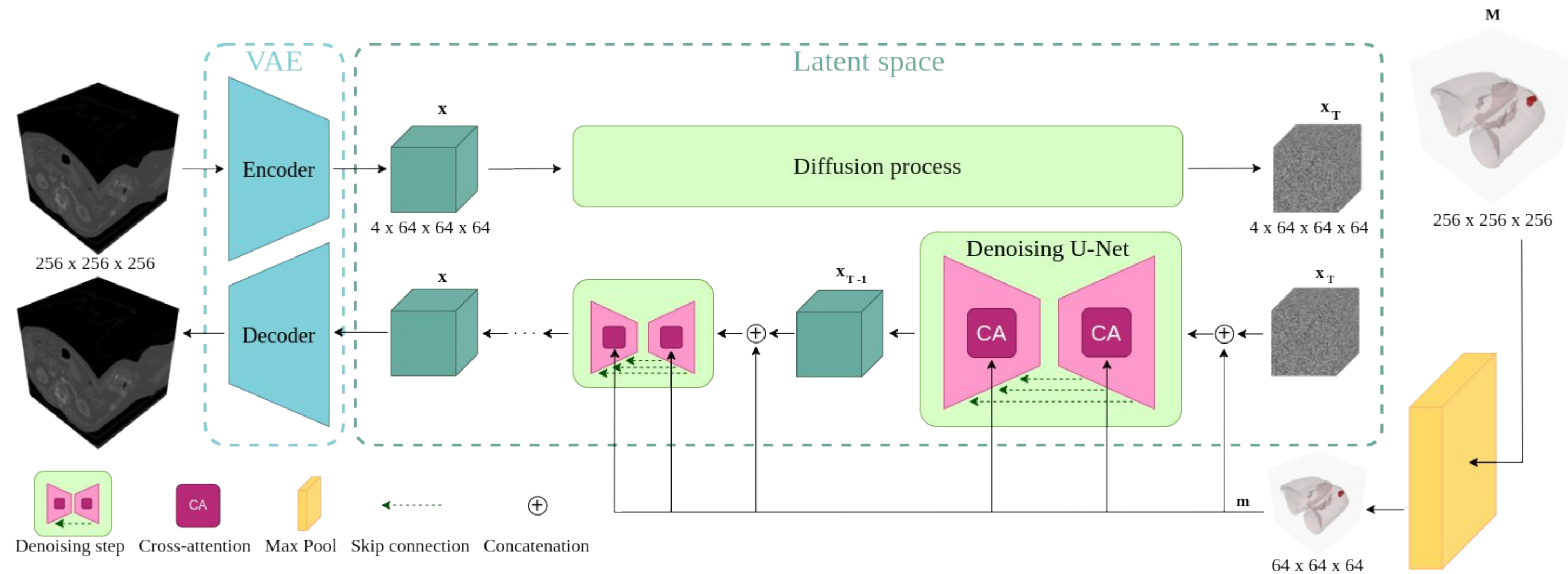


Fig. 1 · CT volumes are encoded into a latent space by a 3D VAE. A 3D U-Net learns the denoising process there, optionally conditioned on a downsampled anatomical mask injected through concatenation and cross-attention.

Architecture at a glance.

3D VAE

Compress 256^3 volumes into a $64^3 \times 4$ latent — $\times 4$ spatial reduction, $\times 4$ feature expansion.

Lightweight **MAISI variant**: 3 resolution levels, one residual block per level.

Symmetric encoder/decoder channel counts.

Loss = **L1 (MAE) + LPIPS + adversarial + KL**.

Trained **once** and then frozen — diffusion runs on top.

3D U-NET DENOISER

5 resolution levels, 2 residual blocks per level, additive skip connections.

Cross-attention modules re-inject conditioning masks at multiple resolutions.

Velocity prediction — estimates a linear combination of clean latent + noise; better stability at high resolution.

Linear noise schedule, $T = 1000$ timesteps.

Min-SNR-y weighting balances timestep contributions to the loss.

Lung + nodule conditioning.

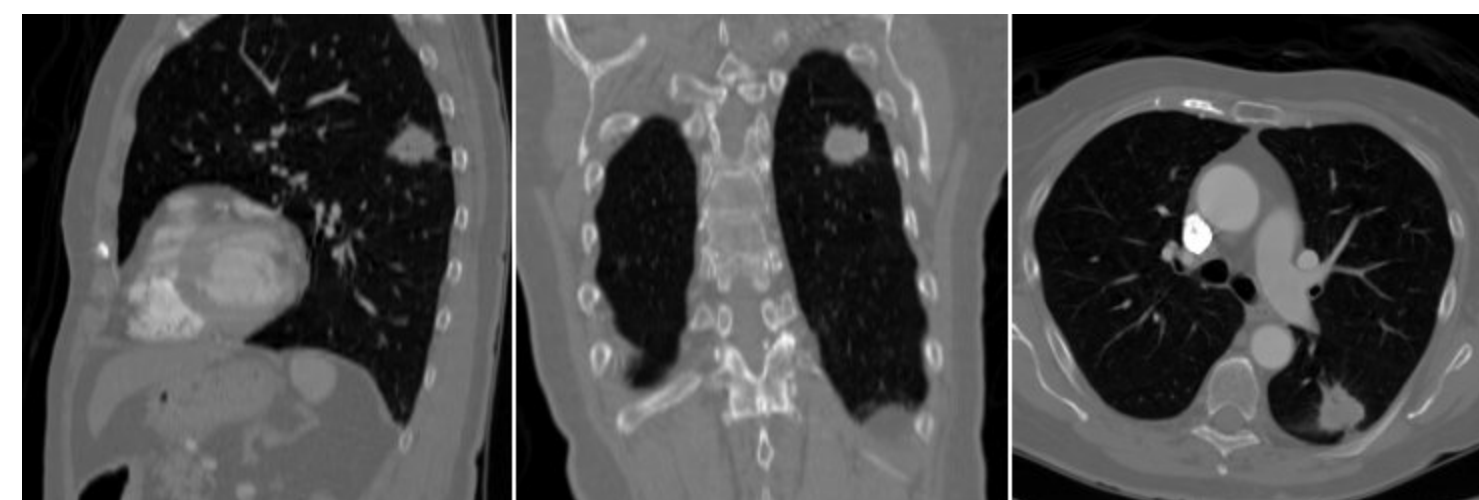
Conditioning on *nodule masks alone* often produces anatomically implausible nodule placements — outside the lung entirely.

LAND conditions on a single mask that encodes **both** global anatomy and local lesions:

0.5 Lung region — global anatomical context.

1 2 3 4 5 Nodule texture score — non-solid (1) → solid (5).

The normalized mask is **concatenated** with the noisy latent feeding the U-Net, and injected again through **cross-attention** at multiple resolutions. 3D max-pooling ×4 brings it to the latent grid — preserving all nodules through the spatial collapse.



Setup &
metrics.

DATASETS

LIDC-IDRI

1 010 CT volumes, four-radiologist nodule masks and attribute ratings. *Training*.

NLST

881 volumes selected (≥ 1 nodule). Nodule masks generated with an ad-hoc U-Net. *Inference / evaluation only*.

Lung areas segmented with a pretrained open-source U-Net (Hofmanninger et al.). Pre-processing: intensity clip, 1 mm isotropic resample, center-crop 256^3 , normalize to [0,1].

METRICS

FID

↓ better

Fréchet Inception Distance over 881 real + synthetic samples. ResNet-50 features pretrained on 23 medical datasets.

MS-SSIM

↓ = more diverse

Multi-scale structural similarity across 10 000 pairs of synthetic samples — measures within-set diversity.

Radiomics

Intensity + GLCM texture features on 32^3 nodule crops, by texture score.

TRAINING

GPU	A100-20C / 20 GB
Batch size	1
Optimizer	AdamW
LR (VAE)	1×10^{-4}
LR (U-Net)	1×10^{-5}
Timesteps T	1 000
β schedule	$1e-4 \rightarrow 0.02$
Total steps	500 000

All experiments trained to the same step budget — fair comparison against baselines run from scratch.

Quantitative

results.

TABLE 1 · LAND VS. BASELINES AND CONDITIONING CONFIGURATIONS

MODEL · CONDITIONING	FID · LIDC ↓	FID · NLST ↓	MS-SSIM ↓	MEM · GB
PatchDDM	317.534	376.397	0.390	19.61
WDM (wavelet, 3D)	15.236	32.662	0.270	7.27
LAND · unconditional	5.062	4.759	0.294	7.38
LAND · nodules only	4.518	5.821	0.297	7.52
LAND · lungs + nodules	4.475	3.373	0.290	7.52
LAND · lungs + textured nodules	4.603	3.865	0.289	7.52

FID multiplied by 10³ in the paper. Arrows summarise relative ranking — see paper for exact values. Crucially, only LAND trains within the 20 GB budget.

HEADLINE 1

Lower FID than WDM & PatchDDM

VAE-based latent compression captures richer semantic features more aligned with the generative task.

HEADLINE 2

Adding lung context improves FID

Nodules alone give only local information. Adding lung anatomy provides the global structural context the model needs.

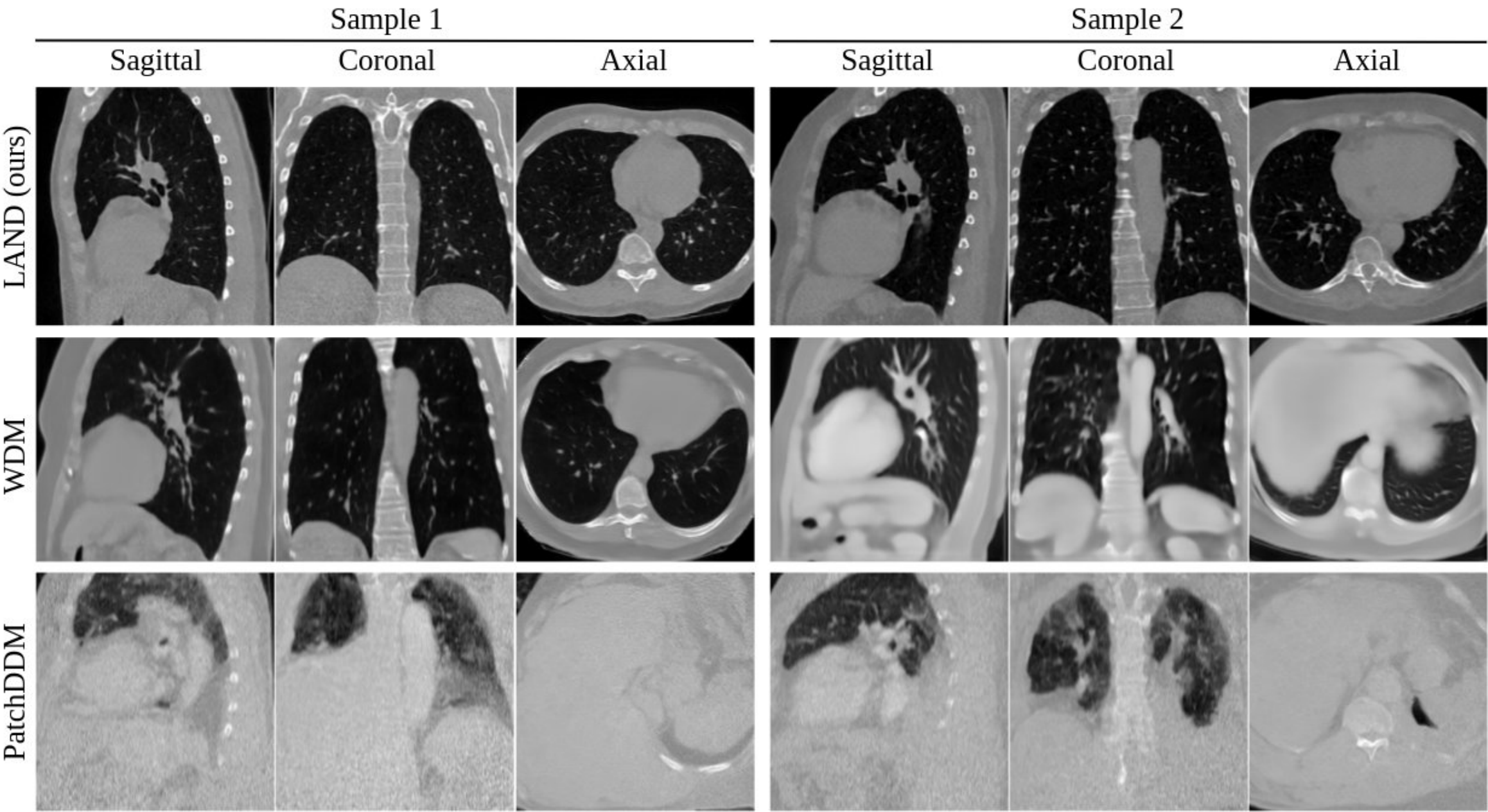
HEADLINE 3

No memorisation of training masks

FID improves on NLST (unseen masks) too — LAND adapts to novel lung contours.

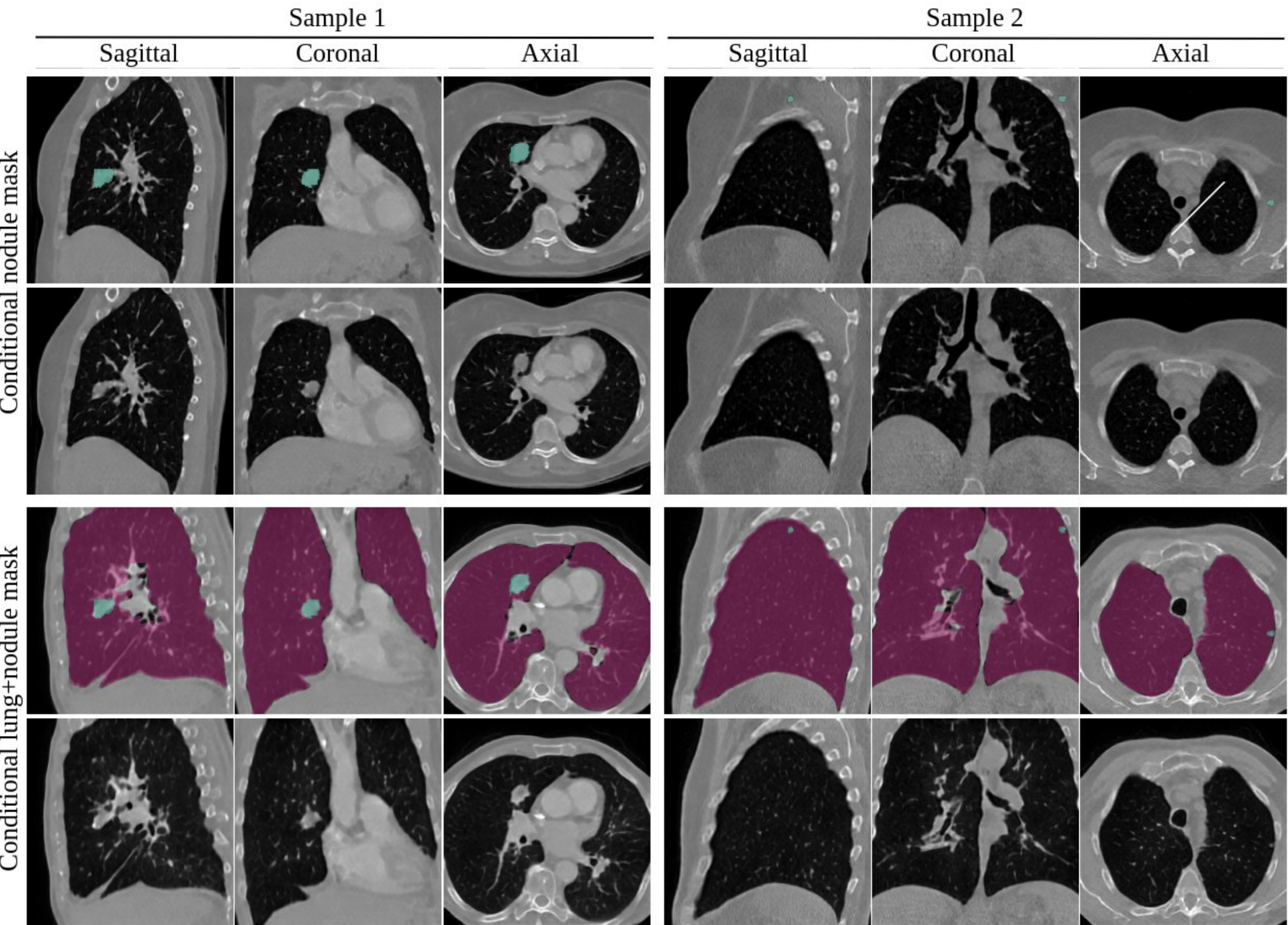
Qualitative results.

LAND VS. BASELINES UNCONDITIONAL CASE



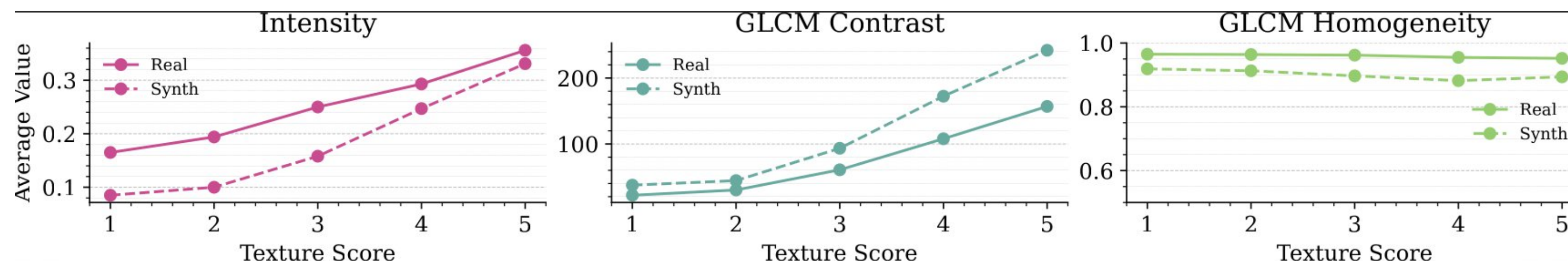
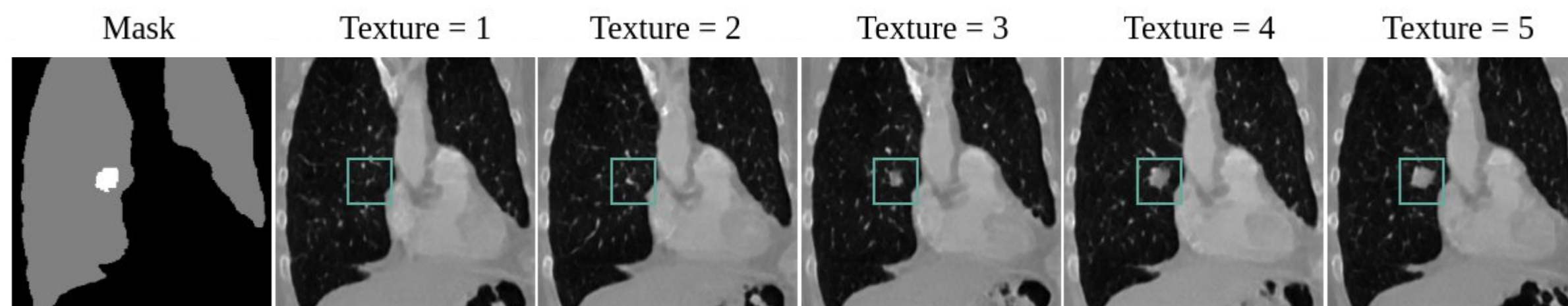
Qualitative results.

LAND CONDITIONAL CASE



Controllable **nodule texture.**

With the same anatomical mask, sweeping the texture score from 1 to 5 produces progressively more **solid** nodules — and the radiomic signatures track real data.



What we shipped, **what's next.**

CONCLUSIONS

LAND — a latent diffusion model for high-quality 3D chest CT synthesis conditioned on anatomical masks.

Fine-grained control over **lung shape**, **nodule location** and **nodule texture**.

Runs end-to-end on **a single mid-range GPU**, dramatically lowering the training cost of concurrent methods.

Promising path forward for AI-driven research and clinical applications where real data is scarce.

LIMITATIONS & FUTURE WORK

Downstream evaluation: measure how synthetic CTs help classification and segmentation.

Richer nodule attributes beyond texture — spiculation, sphericity, margin.

Mask generation module to free generation from predefined anatomical masks.

Faster sampling — currently 1 000 timesteps puts inference in the minutes range.

Extension to **multi-modal / multi-organ** synthesis beyond chest CT.

CIDAI · PHASE IV-AI LAND

Thank you.■

Questions, comments, and collaboration ideas — very welcome.

AUTHORS

Anna Oliveras Tous
Oriol Guardiola Olivella

EURECAT · CIDAI

C/ Bilbao, 72 Edif. A
08005 — Barcelona