

CONFLUX: A Latent Diffusion Model for 3D Chest-CT Synthesis with RL Post-Training

Max Van Puyvelde^{1,2*}, H. Ibrahim Gulluk³, Wim Van Criekinge², and Olivier Gevaert¹

¹ Department of Biomedical Data Science, Stanford University School of Medicine

² Department of Mathematical Modelling, Statistics & Bioinformatics, Ghent University

³ Department of Electrical Engineering, Stanford University
maxvpuyv@stanford.edu, gulluk@stanford.edu, wim.vancriekinge@ugent.be, ogevaert@stanford.edu

Abstract. Controllable generative models of 3D medical images can synthesize volumes with specified clinical attributes, but this demands samples that are simultaneously high-fidelity, natively 3D, and faithful to the requested conditioning. We present CONFLUX, a latent diffusion model for chest computed tomography (CT): a 3D variational autoencoder compresses each volume, and a rectified-flow transformer generates in the latent space. Generation is conditioned on structured radiological metadata (18 abnormality findings, sex, age, and reconstruction kernel) through adaptive layer normalization. The model leads strong volumetric baselines on tri-planar Fréchet distance (FID 32.3 vs. 74.6 for MAISI) while exposing direct control over clinical attributes. To strengthen that control we add an online reinforcement-learning post-training stage (group-relative policy optimization) that rewards how reliably a classifier recovers the requested findings from each generated volume. Judged by a separate, independent classifier, post-training removes 47% of the shortfall relative to real-scan reliability. We release the model and a $\sim 200k$ synthetic chest-CT dataset with conditioning metadata spanning a wide variety of clinical findings.

Keywords: Medical image synthesis · Latent diffusion · Rectified flow · Controllable generation · Reinforcement learning.

1 Introduction

Generative models of 3D medical images are used to augment under-represented cohorts, share data under privacy constraints, and synthesize counterfactual volumes. Each use requires the model to be controllable: a sample requested to exhibit a set of clinical attributes must realize them. Chest CT, which is volumetric, high-resolution, and described by structured radiological metadata, is

* M. Van Puyvelde and H. I. Gulluk are joint first authors; W. Van Criekinge and O. Gevaert are joint senior authors.

an especially hard case; a model that is both high-fidelity and controllable on it is directly useful for cohort augmentation and controlled study design.

High-resolution synthesis has converged on a latent rectified-flow design: an autoencoder compresses the signal and a single-stream transformer trained under a flow-matching objective generates in the learned latent space [15,13,3,2]. We present CONFLUX, a natively 3D instance of this design for chest CT, conditioned on structured radiological metadata (abnormality findings, sex, age, and reconstruction kernel) through adaptive layer normalization. We show it is competitive with a strong volumetric baseline on distribution-level quality while exposing direct control over clinical attributes.

Flow matching trains the model so its outputs reproduce the data distribution for each conditioning vector, an aggregate property that holds over many samples but is not guaranteed for any single one. The objective rewards overall realism and never checks whether an individual generated volume exhibits its requested findings, so a sample can look realistic yet under-express or omit one. The gap is observable post hoc: a classifier trained on real volumes reads requested attributes from generated samples less reliably than from real ones, but the likelihood objective cannot target it. Reinforcement learning (RL) optimizes it directly, maximizing an explicit reward on the model’s own samples. Group-relative policy optimization has recently been adapted from language models to flow-matching image generators [10,20]; we adapt it to 3D structured-attribute medical synthesis as a post-training stage that improves conditioning faithfulness, verified by an independent judge model.

Contributions.

- **A controllable 3D latent rectified-flow model for chest CT.** A 3D convolutional VAE and a single-stream rectified-flow transformer conditioned directly on structured radiological metadata through adaLN modulation, with synthesis quality competitive with a strong 3D CT diffusion baseline (Sec. 3.1–3.2, Sec. 4.1).
- **RL post-training that improves conditioning faithfulness.** A group-relative policy optimization stage that fine-tunes the flow model against a faithfulness reward, improving how reliably requested findings appear in generated volumes. The gain is verified by an independent held-out judge; to our knowledge it is the first GRPO post-training of a 3D medical flow model (Sec. 3.3, Sec. 4.3).
- **A released model and synthetic dataset.** The trained model together with a $\sim 200,000$ -volume controllable, faithfulness-optimized synthetic chest-CT dataset whose conditioning metadata spans a wide variety of findings (Sec. 4.4).

2 Related Work

Latent diffusion compresses volumes with an autoencoder and learns a diffusion or flow model over the latent representation [15]. The image-generation architecture has converged on the diffusion transformer [13], trained under rectified-flow

/ flow-matching objectives [11,9,3] with adaptive-normalization conditioning, as in recent open systems [2]. In 3D medical imaging, conditional latent diffusion has been applied to brain MRI [14,17] and to whole-body and chest CT [5,18,6]; MAISI, a 3D CT latent diffusion model with mask- and metadata-based conditioning [5], is our reference competitor for synthesis quality (Sec. 4.1).

Treating denoising as a multi-step decision process enables policy-gradient fine-tuning of diffusion models against non-differentiable rewards [1,4]. Group-relative policy optimization (GRPO) [16], which replaces a learned value function with a group-relative advantage, was adapted to flow-matching generators by Flow-GRPO [10] via an ODE-to-SDE conversion, and to several visual generation settings by DanceGRPO [20]; subsequent work studies the stability of the importance ratio in this regime [19]. Faithfulness, the agreement between requested and realized attributes, is commonly measured by scoring generated samples with a classifier trained on real data [14].

3 Method

CONFLUX is a three-stage latent diffusion model (Fig. 1): a VAE compresses a CT volume to a low-resolution latent representation (Sec. 3.1); a single-stream rectified-flow transformer generates in that latent space (Sec. 3.2); and a third stage post-trains this model to improve its faithfulness to the requested conditioning (Sec. 3.3).

3.1 Latent autoencoder

Generating directly in voxel space is costly, so the first stage learns a compact latent representation in which the flow model operates. A 3D convolutional variational autoencoder encodes each preprocessed volume $\mathbf{x} \in \mathbb{R}^{1 \times D \times H \times W}$ into a diagonal-Gaussian latent $\mathbf{z} \in \mathbb{R}^{C \times D/f \times H/f \times W/f}$ (mean and per-element variance from an encoder E), downsampling by $f=8$ per spatial axis to $C=16$ channels; a deeper decoder G reconstructs $\hat{\mathbf{x}} = G(\mathbf{z})$. Both use group-normalized residual blocks with a 3D self-attention block at the lowest resolution, the encoder-decoder asymmetry following latent-diffusion autoencoders [15]. Training minimizes an ℓ_1 reconstruction loss, a Kullback-Leibler regularizer toward $\mathcal{N}(\mathbf{0}, \mathbf{I})$ (weight 10^{-6}), and a tri-planar LPIPS perceptual loss [21] averaged over slices along the three canonical planes (LPIPS is 2D). After training E, G are frozen. Rectified-flow training assumes near-unit-variance targets, so the flow model operates on the normalized latent $\tilde{\mathbf{z}} = (\mathbf{z} - m)s$ with a single global scale s and shift m estimated from the training latents, inverted as $\mathbf{z} = \tilde{\mathbf{z}}/s + m$ before decoding or scoring (values in Table 4).

3.2 Conditional rectified-flow transformer

The second stage is a single-stream transformer that generates a latent embedding from a Gaussian noise sample. Under the rectified-flow formulation,

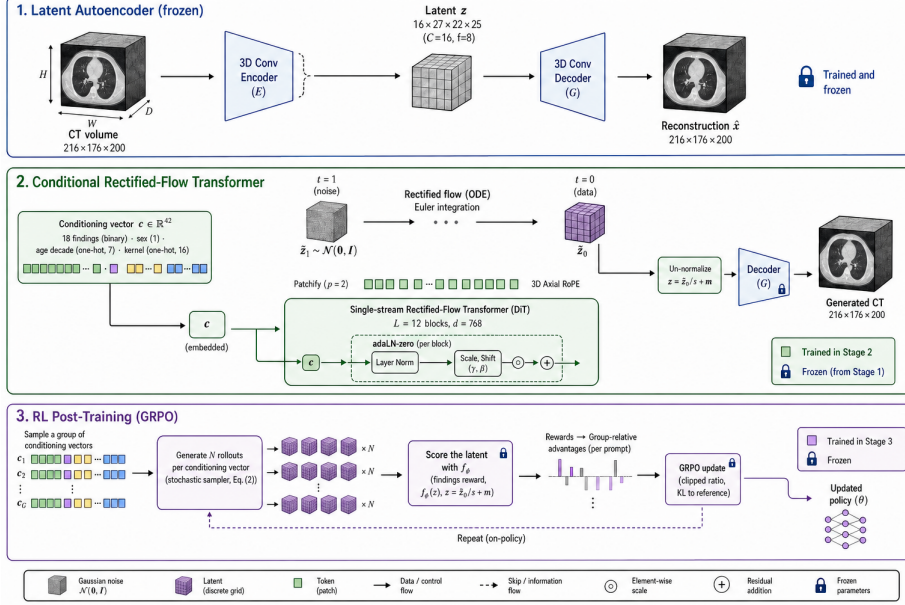


Fig. 1: CONFLUX architecture. **Stage 1:** a 3D convolutional VAE ($f=8$, $C=16$; $216 \times 176 \times 200 \rightarrow 16 \times 27 \times 22 \times 25$) encodes a CT volume to a latent representation and is frozen thereafter. **Stage 2:** a single-stream rectified-flow transformer ($L=12$, $d=768$, patch size $p=2$) is trained from scratch over the normalized latent space by flow matching; patchified tokens carry 3D axial RoPE and are modulated block-wise by the structured metadata vector c (Eq. (1)) through adaLN-zero, and a sample is drawn by integrating the Euler ODE from $t=1$ to $t=0$ and decoding with the frozen Stage-1 decoder. **Stage 3:** GRPO post-training draws a group of rollouts per prompt with the stochastic sampler (Eq. (2)), scores each by the frozen findings-reward classifier (Eq. (3)), and turns the group-relative advantages into a clipped, KL-anchored policy update against the frozen pre-RL reference.

a data latent $\tilde{z}_0$ and noise $\epsilon \sim \mathcal{N}(0, I)$ are joined by the straight-line path $\tilde{z}_t = (1 - t)\tilde{z}_0 + t\epsilon$ ($t=0$ data, $t=1$ noise), travelled at the constant velocity $\epsilon - \tilde{z}_0$. The network $v_\theta(\tilde{z}_t, t, c)$ is trained to predict this velocity by minimizing its expected squared error against the target $\epsilon - \tilde{z}_0$, and at sampling time the predicted velocity field is integrated from noise back to a data latent. The latent volume is patchified into tokens with 3D axial rotary position embeddings; conditioning enters through adaptive layer-normalization (adaLN-zero). The conditioning vector concatenates the available structured metadata,

$$c = [c_{\text{find}} \in \{0, 1\}^{18}, c_{\text{sex}} \in \{0, 1\}, c_{\text{age}} \in \Delta^6, c_{\text{ker}} \in \Delta^{15}] \in \mathbb{R}^{42}, \quad (1)$$

i.e. 18 binary findings, sex, a one-hot age decade (Δ^k the k -simplex vertices), and a one-hot reconstruction kernel. For classifier-free guidance [8] c is dropped to a null embedding with probability 0.1 during training; sampling integrates the

probability-flow ODE $d\tilde{z} = \mathbf{v}_\theta dt$ from $t=1$ to $t=0$ on a time-shifted Euler grid. Training uses precomputed latent moments, so E is never evaluated in stage two.

3.3 Reinforcement-learning post-training

Flow matching trains the generator to match the data distribution, but supplies no signal that a particular requested attribute is realized in a given sample. The post-training stage adds that signal: it samples volumes from the model, scores how well each matches its requested conditioning, and updates the model to increase that agreement, adapting Flow-GRPO [10] to this setting. Policy-gradient updates require a probability for each sampling step, which the deterministic ODE sampler does not provide, so we replace it with a stochastic sampler that preserves the model’s marginals while injecting noise at each step,

$$\tilde{z}_{t+\Delta t} = \tilde{z}_t + \Delta t \mathbf{d}(\tilde{z}_t, t) + \sigma(t) \sqrt{|\Delta t|} \boldsymbol{\xi}, \quad \mathbf{d} = \mathbf{v}_\theta + \frac{\sigma(t)^2}{2t} (\tilde{z}_t + (1-t)\mathbf{v}_\theta), \quad (2)$$

($\Delta t < 0$, action noise $\boldsymbol{\xi} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$). Each step is now a Gaussian draw, giving a T -step rollout a tractable log-probability. Its per-step term is averaged over the $d \approx 2.4 \times 10^5$ latent dimensions rather than summed: summation over so many dimensions inflates the ratio between the updated and sampling policies and destabilizes training, a known failure mode of flow-model RL. The reward measures how faithfully a generated volume realizes its requested findings, read by a frozen classifier f_ϕ ,

$$r(\tilde{z}_0, \mathbf{c}) = - \sum_{g \in \{\text{find}, \text{sex}, \text{age}, \text{ker}\}} \omega_g \ell_g(f_\phi(\mathbf{z}), \mathbf{c}_g), \quad \mathbf{z} = \tilde{z}_0/s + m, \quad (3)$$

the negative weighted cross-entropy between predicted and requested conditioning, with the latent un-normalized first because f_ϕ is trained on raw autoencoder latents. This classifier is a compact 3D convolutional network (group-normalized convolutional blocks, global average pooling, and a linear head over the 18 findings) trained on real latent representations; the independent judge used in evaluation (Sec. 4.3) shares this architecture but scores decoded volumes at the voxel level. We optimize the findings group only ($\omega_{\text{find}}=1$, the others 0); the unweighted groups serve as a specificity check (Sec. 4). For each step we draw G conditioning vectors and N rollouts per vector and form group-relative advantages in the standard way, subtracting each prompt’s mean reward and dividing by the batch-wide standard deviation (more stable than a per-group estimate at the small N affordable here). The policy is updated with the usual clipped PPO surrogate on the per-step importance ratio $\pi_\theta/\pi_\theta^{\text{old}}$, penalized by a Kullback–Leibler divergence to the frozen pre-RL reference π_{ref} . With one on-policy epoch the clip is inactive and this penalty is the sole brake on drift; the configuration is given in Table 4.

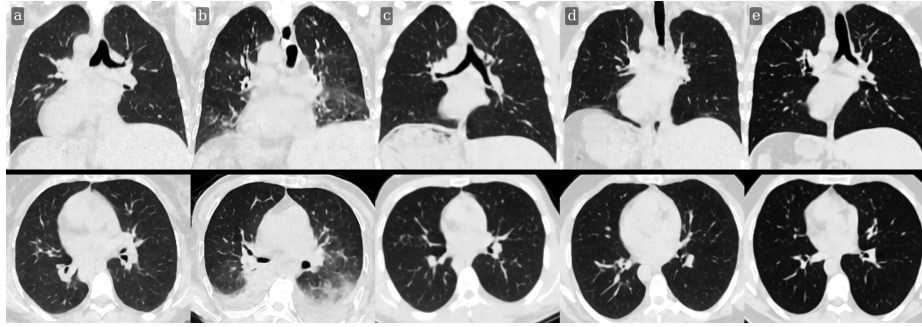


Fig. 2: CONFLUX samples. Generated volumes conditioned on real CT-RATE metadata; coronal (top) and axial (bottom) mid-slices, lung window. **(a)** F, 40–50: cardiomegaly, lung nodule, mosaic attenuation; **(b)** M, 70+: pleural effusion, opacity, consolidation; **(c)** M, 20–30 and **(d)** M, 30–40: no abnormality; **(e)** M, 30–40: lung nodule, bronchiectasis. Not all listed findings appear in the shown mid-slice. More samples with per-volume conditioning are in Appendix C.

4 Experiments

We use CT-RATE [7], a chest-CT corpus with paired structured metadata and 18 NLP-extracted abnormality labels. Volumes are cropped to a lung bounding box and resized to $216 \times 176 \times 200$ with intensities in $\sim [-1, 1]$. A quality-control pass removes scans failing thresholds on lung fraction, left–right balance, and voxel spacing, and removes feet-first and bone-kernel acquisitions; the autoencoder trains on the QC-passing set ($\sim 40,800$ volumes) and the flow model on a one-scan-per-patient subset of 18,417, with 3,039 patient-disjoint volumes held out for validation. Because the 18 findings are NLP-extracted rather than expert-annotated, faithfulness is measured against these predicted labels.

4.1 Synthesis quality

We measure distribution-level quality with a tri-planar 2D-FID (Inception-v3 pool3 features on evenly spaced axial/coronal/sagittal slices, averaged), complemented by density and coverage [12] on the same tri-planar features ($k=5$; fidelity vs. manifold coverage) and diversity (mean pairwise MS-SSIM among generated volumes) against the real floor. Each method is scored on $N=501$ generated volumes against the held-out validation reference, with identical lung windowing and seed. Competitors are MAISI [5], a 3D CT latent diffusion model, and GenerateCT [6], a text-conditional 3D chest-CT generator; VAE reconstruction upper-bounds any latent model on our autoencoder. Fig. 2 shows generated volumes across a range of conditioning profiles, with matching coronal and axial views.

Our model leads both baselines (Table 1): tri-planar FID 32.3 against 74.6 (MAISI) and 145.4 (GenerateCT), approaching the 22.6 VAE-reconstruction

Table 1: Synthesis quality vs. 3D CT baselines ($N=501$; $\pm 95\%$ bootstrap CI). Tri-planar and axial 2D-FID ($\downarrow$), density and coverage [12] ($\uparrow$), and diversity (pairwise MS-SSIM, real floor 0.41). VAE reconstruction is the autoencoder upper bound; best generative result in bold.

Method	FID _{tri} $\downarrow$	FID _{ax} $\downarrow$	Density $\uparrow$	Coverage $\uparrow$	Diversity
GenerateCT [6]	145.4 $\pm$ 2.3	70.2 $\pm$ 3.1	0.008 $\pm$.002	0.009 $\pm$.002	0.486
MAISI [5]	74.6 $\pm$ 3.3	55.6 $\pm$ 2.9	0.016 $\pm$.003	0.017 $\pm$.003	0.471
CONFLUX (ours)	32.3$\pm$1.9	24.7$\pm$1.3	0.056$\pm$.006	0.105$\pm$.008	0.507
VAE recon (ceiling)	22.6 $\pm$ 1.8	18.9 $\pm$ 1.3	0.161 $\pm$.010	0.466 $\pm$.015	0.453

ceiling. Every FID and density/coverage gap to ours is significant (unpaired bootstrap, 95% CIs exclude zero); our samples are $3.5\times$ denser and cover $6\times$ more of the real manifold than MAISI. Diversity sits near the real floor for all methods (0.51 ours vs. 0.41 real; collapse would approach 1). Competitors are re-sampled into our lung space, inflating their cross-plane FID, so the resize-robust axial plane is the cleanest comparison and still favors ours (24.7 vs. 55.6 and 70.2).

4.2 Reward classifier

The GRPO reward is a frozen classifier that reads the 18 findings from the latent (Sec. 3.3), so its accuracy bounds the quality of the reward signal. On the real validation split it reaches macro AUROC 0.793 over the 18 labels, above every CT-CLIP variant reported on the same CT-RATE benchmark [7] (Table 2), outperforming CT-CLIP zero-shot on 17/18 findings. It does so while scoring the $8\times$ -compressed latent rather than full-resolution CT, indicating the tokenizer preserves the diagnostic signal. The comparison is not fully controlled (we score a one-scan-per-patient subset from the latent; the baselines, the full validation split at full resolution), but the dataset, labels, findings, and metric are identical.

4.3 Conditioning faithfulness

We measure faithfulness as agreement between requested and realized conditioning, read by a classifier on the generated samples. The *reward* classifier operates in the latent space and is the model optimized by GRPO; because using it for evaluation is circular, we judge instead with an *independent image-space* classifier trained on decoded real volumes and never used as the reward. We generate from 200 fixed validation prompts with 4 samples each at 50 ODE steps and report macro average precision (AP) and macro AUROC over the 18 findings. Post-training raises both under the independent judge (Table 3): AP 0.330 $\rightarrow$ 0.344 and AUROC 0.684 $\rightarrow$ 0.699. The absolute gains are small, but so is the headroom: the judge reads *real* CT at only AP 0.360, so the gain already recovers 47% of the base-to-real gap. A paired-by-prompt bootstrap ($B=2000$),

Table 2: Per-finding AUROC on CT-RATE (real validation): our latent classifier vs. CT-CLIP variants and CT-Net [7]. Top 8 findings by our AUROC; macro over all 18 (full table in Table 5). Best per row in bold.

Finding	Ours	CT-Net	CT-CLIP _{zs}	CT-CLIP _{ft}
Cardiomegaly	0.92	0.79	0.86	0.92
Pleural effusion	0.91	0.77	0.90	0.93
Arterial wall calcif.	0.91	0.63	0.85	0.87
Coronary artery calcif.	0.91	0.63	0.85	0.86
Interlobular septal thick.	0.88	0.73	0.77	0.80
Pericardial effusion	0.83	0.65	0.77	0.78
Mosaic attenuation	0.81	0.74	0.77	0.78
Bronchiectasis	0.78	0.57	0.65	0.61
Macro (all 18)	0.793	0.629	0.731	0.756

Table 3: Conditioning faithfulness under the *independent* image-space judge, findings macro over 18 labels ($n=800$ generated). The real-data ceiling is the judge scored on real volumes; Δ and its ± 1 s.e. come from a paired-by-prompt bootstrap, both gains significant. Headroom recovered is Δ as a fraction of the base-to-ceiling gap.

	AP $\uparrow$	AUROC $\uparrow$
Base (pre-RL)	0.330	0.684
CONFLUX	0.344	0.699
Real-data ceiling	0.360	0.746
Δ (CONFLUX – base)	$+0.014 \pm 0.008$	$+0.015 \pm 0.006$
Headroom recovered	47%	24%

which cancels prompt-difficulty noise, confirms both gains significant ($p=0.042$ and 0.014). The unweighted groups (sex, age, kernel) stay unchanged within noise, so the gain is specific to the optimized findings.

4.4 Model and dataset release

We release the trained model and, conditioning it on the real CT-RATE meta-data proportions, a $\sim 200,000$ -volume synthetic chest-CT dataset. Each volume is paired with its conditioning vector, spanning a wide variety of findings, for cohort augmentation and conditional study design at a scale unavailable in real corpora. The dataset is available at <https://huggingface.co/datasets/gevaertlab/conflux-chest-ct> and the model checkpoints at <https://huggingface.co/gevaertlab/conflux>.

5 Conclusion

We presented CONFLUX, a controllable, natively 3D latent rectified-flow model for chest CT, conditioned on structured radiological metadata, that leads strong 3D baselines on quality while giving direct control over clinical attributes. An RL post-training stage raises conditioning faithfulness, the first GRPO post-training of a 3D medical flow model to our knowledge. We release the model and a $\sim 200,000$ -volume synthetic chest-CT dataset.

References

1. Black, K., Janner, M., Du, Y., Kostrikov, I., Levine, S.: Training diffusion models with reinforcement learning. In: ICLR (2024)
2. Black Forest Labs: FLUX. <https://blackforestlabs.ai> (2024), technical report
3. Esser, P., Kulal, S., Blattmann, A., et al.: Scaling rectified flow transformers for high-resolution image synthesis. In: ICML (2024)
4. Fan, Y., et al.: DPOK: Reinforcement learning for fine-tuning text-to-image diffusion models. In: NeurIPS (2023)
5. Guo, P., Zhao, C., Yang, D., et al.: MAISI: Medical AI for synthetic imaging. arXiv preprint arXiv:2409.11169 (2024)
6. Hamamci, I.E., Er, S., et al.: GenerateCT: Text-conditional generation of 3d chest CT volumes. In: ECCV (2024)
7. Hamamci, I.E., et al.: Developing generalist foundation models from a multimodal dataset for 3D computed tomography. arXiv preprint arXiv:2403.17834 (2024)
8. Ho, J., Salimans, T.: Classifier-free diffusion guidance. In: NeurIPS Workshop (2021)
9. Lipman, Y., Chen, R.T.Q., Ben-Hamu, H., Nickel, M., Le, M.: Flow matching for generative modeling. In: ICLR (2023)
10. Liu, J., et al.: Flow-GRPO: Training flow matching models via online RL. arXiv preprint arXiv:2505.05470 (2025)
11. Liu, X., Gong, C., Liu, Q.: Flow straight and fast: Learning to generate and transfer data with rectified flow. In: ICLR (2023)
12. Naeem, M.F., Oh, S.J., Uh, Y., Choi, Y., Yoo, J.: Reliable fidelity and diversity metrics for generative models. In: ICML (2020)
13. Peebles, W., Xie, S.: Scalable diffusion models with transformers. In: ICCV (2023)
14. Pinaya, W.H.L., et al.: Brain imaging generation with latent diffusion models. In: MICCAI Workshop DGM4MICCAI (2022)
15. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: CVPR (2022)
16. Shao, Z., et al.: Deepseekmath: Pushing the limits of mathematical reasoning in open language models. arXiv preprint arXiv:2402.03300 (2024)
17. Van Puyvelde, M., Gulluk, I., Van Criekinge, W., Gevaert, O.: BrainG3N: A dual-purpose tokenizer for controllable 3D brain MRI generation. arXiv preprint arXiv:2606.19651 (2026)
18. Wang, H., et al.: 3D MedDiffusion: A 3D medical latent diffusion model for controllable and high-quality medical image generation. arXiv preprint arXiv:2412.13059 (2024)

19. Wang, J., Liang, J., Liu, J., et al.: GRPO-guard: Mitigating implicit over-optimization in flow matching via regulated clipping. arXiv preprint arXiv:2510.22319 (2025)
20. Xue, Z., et al.: DanceGRPO: Unleashing GRPO on visual generation. arXiv preprint arXiv:2505.07818 (2025)
21. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. In: CVPR (2018)

A Training Parameters

Table 4: Training configuration for the three stages. Latent moments are cached; the post-training SDE uses the FLUX time-shifted Euler grid; the reward classifier and reference policy are frozen. Per-channel latent std is 1.27–2.48.

Stage	Hyperparameter	Value
Autoencoder	downsampling factor f	8
	latent channels C	16
Flow transformer	hidden width d	768
	depth L	12
	attention heads	12
	patch size p	2
	latent norm. scale s	0.5509
	latent norm. shift m	0.1869
GRPO post-training	prompts per step G	2
	rollouts per prompt N	16
	SDE steps T	10
	action noise $\sigma(t)$	$0.5\sqrt{t}$
	PPO clip ϵ	0.2
	KL weight β_{KL}	0.03
	inner epochs	1
	learning rate	3×10^{-6}

B Per-finding reward-classifier accuracy

Table 5: Per-finding AUROC on CT-RATE (real validation), all 18 findings: our latent classifier vs. CT-CLIP variants and CT-Net [7]. Sorted by our AUROC; best per row in bold.

Finding	Ours	CT-Net	CT-CLIP _{zs}	CT-CLIP _{ft}
Cardiomegaly	0.92	0.79	0.86	0.92
Pleural effusion	0.91	0.77	0.90	0.93
Arterial wall calcif.	0.91	0.63	0.85	0.87
Coronary artery calcif.	0.91	0.63	0.85	0.86
Interlobular septal thick.	0.88	0.73	0.77	0.80
Pericardial effusion	0.83	0.65	0.77	0.78
Mosaic attenuation	0.81	0.74	0.77	0.78
Bronchiectasis	0.78	0.57	0.65	0.61
Atelectasis	0.76	0.61	0.68	0.69
Consolidation	0.76	0.66	0.72	0.75
Medical material	0.76	0.66	0.72	0.75
Lung opacity	0.76	0.60	0.63	0.69
Hiatal hernia	0.75	0.55	0.70	0.72
Emphysema	0.73	0.54	0.74	0.75
Lymphadenopathy	0.72	0.61	0.70	0.70
Peribronchial thick.	0.71	0.53	0.69	0.71
Pulmonary fibrotic seq.	0.70	0.52	0.57	0.65
Lung nodule	0.67	0.55	0.57	0.65
Macro (all 18)	0.793	0.629	0.731	0.756

C Additional samples and conditioning



Fig. 3: CONFLUX samples. Forty synthetic chest CT volumes drawn at random from the model, each as a coronal (top) and axial (bottom) mid-slice in a lung window, numbered 0–39; per-patient conditioning is given in Table 6. The range of body habitus, anatomy, and abnormality reflects the diversity and realism of generated volumes; matching coronal/axial views show cross-plane 3D coherence.

Table 6: Conditioning metadata for each synthetic volume in Fig. 3.

#	Sex	Age	Kernel	Conditioned abnormalities
0	F	0–20	Br40f	none
1	F	70+	YA	Arterial wall calcification, Cardiomegaly, Coronary artery wall calcification, Lymphadenopathy, Atelectasis, Lung nodule, Lung opacity
2	M	60–70	Bl56f	Arterial wall calcification, Coronary artery wall calcification, Hiatal hernia, Pulmonary fibrotic sequela, Bronchiectasis
3	M	30–40	YA	Lung opacity, Consolidation, Interlobular septal thickening
4	M	30–40	EA	Hiatal hernia, Atelectasis
5	F	20–30	Br40f	none
6	M	30–40	Br40f	Lymphadenopathy, Peribronchial thickening, Consolidation, Bronchiectasis
7	F	20–30	A	Atelectasis, Lung nodule
8	M	60–70	YA	Arterial wall calcification, Lung opacity
9	F	40–50	YA	Arterial wall calcification, Hiatal hernia, Lymphadenopathy, Atelectasis, Lung nodule, Pulmonary fibrotic sequela, Consolidation
10	F	30–40	EA	Lymphadenopathy, Lung nodule
11	M	40–50	YA	Lymphadenopathy, Lung nodule, Lung opacity
12	M	50–60	YB	Lymphadenopathy, Lung nodule
13	F	50–60	L	Coronary artery wall calcification, Lung nodule, Lung opacity, Pulmonary fibrotic sequela
14	F	70+	Bl56f	Arterial wall calcification, Coronary artery wall calcification, Emphysema, Atelectasis, Lung nodule, Pulmonary fibrotic sequela, Mosaic attenuation pattern
15	F	70+	YA	Arterial wall calcification, Coronary artery wall calcification, Atelectasis
16	M	50–60	Bl57d	Lymphadenopathy, Lung opacity
17	M	20–30	Br36d	none
18	M	20–30	EA	none
19	F	40–50	YA	Arterial wall calcification, Lymphadenopathy, Emphysema, Lung nodule, Lung opacity, Pulmonary fibrotic sequela
20	M	50–60	YA	Lung opacity
21	F	40–50	YA	none
22	M	40–50	EA	Lymphadenopathy, Lung nodule
23	F	50–60	Br40f	Cardiomegaly, Lymphadenopathy, Consolidation

#	Sex	Age	Kernel	Conditioned abnormalities
24	F	60–70	YA	Arterial wall calcification, Lymphadenopathy, Lung nodule, Pulmonary fibrotic sequela, Mosaic attenuation pattern
25	F	70+	YA	Coronary artery wall calcification
26	M	50–60	L	Lung opacity, Consolidation
27	M	70+	Bl56f	Arterial wall calcification, Coronary artery wall calcification, Atelectasis, Lung opacity, Pulmonary fibrotic sequela
28	F	20–30	YA	none
29	M	30–40	EA	Lung nodule, Pulmonary fibrotic sequela
30	F	50–60	Br40f	Arterial wall calcification, Coronary artery wall calcification, Hiatal hernia, Lung nodule
31	F	20–30	B	none
32	F	50–60	YA	Lymphadenopathy, Atelectasis, Lung opacity, Pulmonary fibrotic sequela
33	F	60–70	Br36d	Lung opacity, Pulmonary fibrotic sequela, Peribronchial thickening, Consolidation, Bronchiectasis, Interlobular septal thickening
34	M	20–30	EA	none
35	M	30–40	B	Lung nodule
36	F	70+	YA	Arterial wall calcification, Coronary artery wall calcification, Emphysema, Consolidation
37	F	40–50	Br40f	none
38	M	20–30	Br40f	none
39	F	30–40	YA	none



Fig. 4: CONFLUX samples (continued). A further 40 synthetic chest CT volumes drawn at random, each a coronal (top) and axial (bottom) mid-slice in a lung window, numbered 0–39; per-patient conditioning is given in Table 7.

Table 7: Conditioning metadata for each synthetic volume in Fig. 4.

#	Sex	Age	Kernel	Conditioned abnormalities
0	M	40–50	B	Medical material, Atelectasis, Lung nodule, Lung opacity
1	F	20–30	YB	Lung nodule
2	M	50–60	L	Arterial wall calcification, Atelectasis
3	M	20–30	YA	Lung opacity, Interlobular septal thickening
4	M	50–60	EA	Hiatal hernia, Atelectasis, Lung nodule, Pulmonary fibrotic sequela
5	M	50–60	EA	Arterial wall calcification, Lung opacity
6	F	20–30	Bl56f	Arterial wall calcification, Cardiomegaly, Pericardial effusion, Coronary artery wall calcification, Hiatal hernia, Lymphadenopathy, Lung nodule
7	M	30–40	Bl57d	Emphysema, Atelectasis
8	F	50–60	Br40f	none
9	F	30–40	Br40f	none
10	M	40–50	Bl56f	Hiatal hernia, Lymphadenopathy, Atelectasis, Lung nodule, Bronchiectasis
11	F	30–40	Bl57d	none
12	M	50–60	L	Lymphadenopathy, Lung opacity
13	M	50–60	EA	Lung nodule
14	F	30–40	Br60f	Medical material, Emphysema, Pulmonary fibrotic sequela
15	F	30–40	Br40f	none
16	M	40–50	B	Emphysema, Lung opacity, Consolidation
17	F	70+	YA	Emphysema, Atelectasis, Lung nodule, Peribronchial thickening
18	F	60–70	YA	Arterial wall calcification, Hiatal hernia, Atelectasis, Lung opacity
19	F	40–50	YA	Emphysema
20	F	70+	Br40f	Arterial wall calcification, Coronary artery wall calcification, Lymphadenopathy, Lung nodule, Pulmonary fibrotic sequela
21	M	20–30	YA	Lung opacity, Peribronchial thickening, Bronchiectasis
22	F	40–50	Br40f	Coronary artery wall calcification, Lung nodule
23	M	50–60	EA	Medical material, Coronary artery wall calcification, Lymphadenopathy
24	F	50–60	YA	Arterial wall calcification, Lymphadenopathy, Emphysema, Atelectasis
25	M	20–30	Br40f	Lymphadenopathy, Lung nodule, Lung opacity

#	Sex	Age	Kernel	Conditioned abnormalities
26	M	60–70	B	Arterial wall calcification, Cardiomegaly, Pericardial effusion, Lymphadenopathy, Emphysema, Atelectasis, Lung opacity, Pulmonary fibrotic sequela, Pleural effusion, Peribronchial thickening, Bronchiectasis
27	F	30–40	Br40f	none
28	M	60–70	B	Arterial wall calcification, Coronary artery wall calcification, Hiatal hernia, Lung opacity
29	M	30–40	Bl56f	Lymphadenopathy, Lung opacity, Bronchiectasis
30	M	40–50	Br40f	Medical material, Emphysema, Lung nodule, Pulmonary fibrotic sequela
31	M	20–30	YB	none
32	M	20–30	YA	Lung nodule
33	M	50–60	YA	Hiatal hernia, Lung nodule, Pulmonary fibrotic sequela, Mosaic attenuation pattern, Peribronchial thickening
34	M	30–40	A	Lung opacity, Interlobular septal thickening
35	M	40–50	YA	Hiatal hernia, Lung opacity
36	M	30–40	Br60f	Bronchiectasis
37	F	30–40	Br40f	Medical material, Pericardial effusion, Atelectasis, Lung nodule, Pulmonary fibrotic sequela
38	M	30–40	Bl56f	none
39	M	50–60	B	Lung opacity



Fig. 5: CONFLUX samples (continued). A further 40 synthetic chest CT volumes drawn at random, each a coronal (top) and axial (bottom) mid-slice in a lung window, numbered 0–39; per-patient conditioning is given in Table 8.

Table 8: Conditioning metadata for each synthetic volume in Fig. 5.

#	Sex	Age	Kernel	Conditioned abnormalities
0	M	30–40	YA	Lung opacity
1	F	70+	YA	Arterial wall calcification, Coronary artery wall calcification, Atelectasis, Lung nodule, Pulmonary fibrotic sequela
2	M	60–70	YA	Arterial wall calcification, Coronary artery wall calcification, Emphysema, Pulmonary fibrotic sequela
3	F	50–60	YA	Lung nodule, Lung opacity, Pulmonary fibrotic sequela
4	F	40–50	YA	Medical material, Lung nodule
5	F	20–30	YA	Pericardial effusion, Pleural effusion, Peribronchial thickening, Consolidation
6	F	70+	Bl56f	Arterial wall calcification, Coronary artery wall calcification, Hiatal hernia, Lymphadenopathy, Lung nodule, Lung opacity, Pulmonary fibrotic sequela
7	M	50–60	YA	Arterial wall calcification, Coronary artery wall calcification, Lymphadenopathy, Lung nodule, Lung opacity
8	M	40–50	EA	Lung nodule, Pulmonary fibrotic sequela
9	M	30–40	EA	Emphysema, Lung nodule, Pulmonary fibrotic sequela
10	M	40–50	Br40f	Hiatal hernia, Atelectasis
11	F	60–70	YA	Lymphadenopathy
12	M	60–70	YA	Medical material, Arterial wall calcification, Cardiomegaly, Coronary artery wall calcification, Lymphadenopathy, Emphysema, Lung nodule, Pleural effusion, Mosaic attenuation pattern
13	M	70+	Br40f	Arterial wall calcification, Cardiomegaly, Coronary artery wall calcification, Lymphadenopathy, Emphysema, Lung opacity, Pleural effusion, Peribronchial thickening, Consolidation
14	F	60–70	Br60f	Cardiomegaly, Lymphadenopathy, Atelectasis
15	M	30–40	EA	Lung nodule
16	F	40–50	YA	Medical material, Atelectasis, Lung opacity, Consolidation
17	M	40–50	B	Coronary artery wall calcification, Pulmonary fibrotic sequela
18	F	60–70	Bl56f	Atelectasis, Lung opacity
19	M	30–40	Br40f	Lymphadenopathy, Emphysema

#	Sex	Age	Kernel	Conditioned abnormalities
20	M	20–30	other	none
21	F	70+	Br60f	Arterial wall calcification, Coronary artery wall calcification, Hiatal hernia, Lymphadenopathy, Emphysema, Lung nodule, Lung opacity, Consolidation
22	F	50–60	Bl56f	Lung nodule
23	F	70+	YA	Coronary artery wall calcification, Hiatal hernia, Atelectasis, Pleural effusion
24	F	70+	Br60f	Medical material, Arterial wall calcification, Coronary artery wall calcification, Lung nodule, Lung opacity
25	M	40–50	Bl56f	Lung nodule, Lung opacity
26	M	50–60	B	Arterial wall calcification, Coronary artery wall calcification, Lymphadenopathy, Emphysema, Lung nodule, Pulmonary fibrotic sequela
27	F	30–40	Bl56f	Lymphadenopathy, Lung nodule, Pulmonary fibrotic sequela
28	M	70+	YA	Arterial wall calcification, Coronary artery wall calcification, Hiatal hernia, Lymphadenopathy, Lung nodule, Lung opacity, Pulmonary fibrotic sequela
29	M	50–60	Bl56f	Medical material, Arterial wall calcification, Coronary artery wall calcification, Hiatal hernia, Emphysema, Lung nodule, Pulmonary fibrotic sequela, Bronchiectasis
30	M	50–60	B	Medical material, Consolidation, Bronchiectasis
31	F	40–50	YA	Lung nodule, Lung opacity
32	M	20–30	Br40f	Lung opacity, Consolidation
33	M	30–40	B	Lymphadenopathy, Consolidation
34	M	50–60	EA	Lung nodule
35	M	30–40	Bl56f	Lymphadenopathy, Emphysema, Atelectasis, Lung nodule
36	M	60–70	B	Arterial wall calcification, Coronary artery wall calcification, Atelectasis, Lung nodule, Mosaic attenuation pattern
37	M	30–40	Br40f	none
38	M	60–70	B	Lung nodule, Pulmonary fibrotic sequela
39	F	40–50	YA	Emphysema, Lung nodule, Lung opacity, Pulmonary fibrotic sequela