This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

AMFG: Anatomy-grounded Multi-Finding Guidance for Training-Free Chest X-Ray Generation

Yeon Gyu Han[1⋆], Junah Jung[2⋆], Chang Min Park[3†], and Dongheon Lee[3,4†]

[1] Department of Biomedical Engineering, Chungnam National University College of Medicine, Daejeon, Republic of Korea
[2] Interdisciplinary Program in Medical Informatics, Seoul National University College of Medicine, Seoul, Republic of Korea
[3] Department of Radiology, Seoul National University College of Medicine, Seoul National University Hospital, Seoul, Republic of Korea
[4] Interdisciplinary Program in Artificial Intelligence, Seoul National University, Seoul, Republic of Korea
dusrb37@gmail.com; goldfish0907@gmail.com; morphius@snu.ac.kr; dhlee.jubilee@gmail.com

Abstract. Text-conditioned diffusion models have enabled chest X-ray (CXR) synthesis, yet when prompted with multi-finding radiology reports they frequently omit findings or place them at anatomically incorrect locations. Existing approaches improve this through reinforcement learning or adapter training, but their reliance on additional training limits portability across base models. We propose AMFG (Anatomy-grounded Multi-Finding Guidance), a training-free framework applied at inference time to a frozen CXR diffusion model. AMFG comprises three complementary guidance signals: (1) finding presence guidance using a pretrained pathology classifier, (2) anatomical grounding guidance that constrains cross-attention maps to clinically correct regions via segmentation masks, and (3) finding interaction guidance that enforces spatial consistency among co-occurring findings. All components use only existing public models with no additional training. Experiments on MIMIC-CXR demonstrate that AMFG achieves the highest anatomical correctness among all compared methods and that this accuracy translates into superior downstream classifier performance, demonstrating that explicit anatomical grounding—rather than additional training alone—is key to clinically useful synthetic CXR data.

Keywords: Chest X-ray synthesis · Training-free guidance · Multi-finding generation.

1 Introduction

Text-conditioned diffusion models have enabled chest X-ray (CXR) synthesis from free-text radiology reports. RoentGen [1] first demonstrated that visually

⋆ Equal contribution. † Corresponding authors.

plausible CXRs can be generated from report descriptions, offering a promising avenue for alleviating data scarcity and privacy constraints in medical imaging [2,3,4]. This capability is especially valuable for uncommon findings such as pneumothorax or lung masses, where limited labeled examples hinder classifier training and synthetic augmentation offers a practical remedy [3].

However, real-world radiology reports typically describe multiple concurrent findings with implicit anatomical context (*e.g.*, "bilateral pleural effusion with cardiomegaly"). When prompted with such multi-finding descriptions, existing CXR models frequently omit individual findings or place them at anatomically incorrect locations (Fig 1). For instance, a model prompted with "left pleural effusion" may diffuse opacity into the right costophrenic angle, producing an image that is visually plausible yet clinically misleading. Such misalignment is not merely a quality issue: synthetic images with incorrect finding–anatomy associations can propagate spurious correlations into downstream classifiers, undermining the very purpose of synthetic data augmentation [5,6].

Recent methods tackle this misalignment through additional training, such as reinforcement learning [7] or learned adapter modules [8], but coupling guidance to a specific base model limits portability. A natural alternative is training-free inference-time guidance, which steers generation without touching model weights and has shown strong results for compositional control in natural images [9,10,11]. Yet no existing training-free method addresses multi-finding medical image generation, because spatial control in CXR demands *anatomical domain knowledge* that text-based cues alone cannot provide.

We propose AMFG (**A**natomy-grounded **M**ulti-**F**inding **G**uidance), a training-free framework that steers a frozen CXR diffusion model at inference time without modifying its weights (Fig. 1). Our key insight is that faithful multi-finding CXR generation requires not only ensuring that each finding is *present*, but also constraining *where* it appears and how multiple findings *interact* spatially. Accordingly, AMFG comprises three complementary signals: finding presence guidance via a pathology classifier, anatomical grounding guidance that constrains cross-attention to clinically correct regions using segmentation masks, and finding interaction guidance that enforces spatial consistency among co-occurring findings.

Our contributions are as follows: (1) We present AMFG, a training-free guidance framework that steers a frozen CXR diffusion model to generate anatomically faithful multi-finding images at inference time, using only publicly available pretrained models and requiring no additional training. (2) Through systematic ablation, we show that the bottleneck in spatial supervision: a simple grounding signal that constrains cross-attention to clinically correct regions yields a larger gain than all other guidance components combined. (3) We empirically demonstrate that anatomically correct synthesis directly improves downstream classification, with the largest gains on rare findings—establishing that spatial accuracy, not just visual realism, determines the clinical utility of synthetic CXR data.

2 Related Work

Text-Driven CXR Generation. RoentGen [1], its demographic extension [12], and LLM-CXR [13] generate CXRs from text but provide no spatial control over finding placement. CXRL [7] adds RL rewards for diagnostic accuracy, yet its image-level reward cannot penalize per-finding misplacement. XReal [8] trains anatomy adapters for mask-guided generation, achieving precise spatial control but requiring retraining for each new base model. SegGuidedDiff [14] and InstructX2X [15] similarly require training—the former conditions on segmentation masks from scratch, the latter targets single-finding edits rather than multi-finding synthesis. Vilouras *et al.* [16] showed that cross-attention maps in CXR diffusion models are poorly calibrated to anatomical structures—the very signal our method exploits. None of these methods provides weight-preserving guidance applicable to a frozen CXR diffusion model without retraining.

Training-Free Diffusion Guidance. In natural images, attention-based methods (A&E [9], STORM [10]) improve multi-concept presence but do not constrain *where* each concept appears, while gradient-based methods (TFG [11]) lack finding–region mappings for CXR. PAG [17] and FreeU [18] operate globally without finding-level targets. In medical imaging, training-free guidance has addressed reconstruction [19] and enhancement [20], but not multi-finding generation. The missing ingredient is *anatomical domain knowledge*: AMFG fills this gap by injecting a pretrained segmentor and a clinical finding–anatomy mapping φ into inference-time guidance.

3 Method

Given a frozen CXR diffusion model and a multi-finding report, AMFG steers the denoising process through three complementary signals (Fig. 1): finding presence guidance (§3.2), anatomical grounding guidance (§3.3), and finding interaction guidance (§3.4)—addressing *what* to generate, *where* to place it, and *how* findings should interact.

3.1 Preliminaries

We assume a frozen CXR latent diffusion model ϵ_θ with text encoder τ. At step t, the predicted clean latent is $\hat{z}_0 = (z_t - \sqrt{1-\bar{\alpha}_t}\,\epsilon_\theta(z_t, t, \tau(y)))/\sqrt{\bar{\alpha}_t}$, and the decoded image is $\hat{x}_0 = \mathrm{dec}(\hat{z}_0)$, where $\mathrm{dec}(\cdot)$ is the VAE decoder. The cross-attention map $A_i \in \mathbb{R}^{h \times w}$ for finding d_i indicates which spatial locations attend to the corresponding text tokens. From the report y, we extract a finding set $\mathcal{D} = \{d_1, \ldots, d_N\}$ using the CheXpert labeler [21], retaining only positively mentioned findings and discarding negated or uncertain mentions.

3.2 Finding Presence Guidance

We leverage a frozen pathology classifier f_{cls} (DenseNet-121 trained on NIH ChestX-ray14 [22], provided by TorchXRayVision [23]) to ensure each prompted

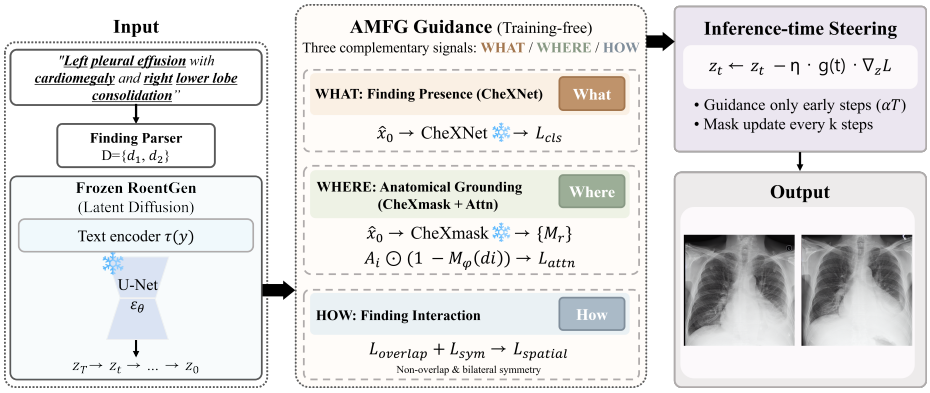

Fig. 1. Overview of AMFG. When prompted with a multi-finding report, a frozen CXR diffusion model frequently omits findings or places them at incorrect locations. AMFG steers the denoising process via three complementary signals: finding presence guidance ($\mathcal{L}_{\text{cls}}$), anatomical grounding guidance ($\mathcal{L}_{\text{attn}}$), and finding interaction guidance ($\mathcal{L}_{\text{spatial}}$), using only pretrained public models with no additional training.

finding appears in the generated image. At each step, we pass $\hat{x}_0$ through f_{cls} and maximize the predicted probability for every finding:

$$\mathcal{L}_{\text{cls}} = -\sum_{i=1}^{N} \log\, \sigma(f_{\text{cls}}(\hat{x}_0)_{d_i})\,, \tag{1}$$

However, $\mathcal{L}_{\text{cls}}$ ensures presence but not placement; the next section addresses this.

3.3 Anatomical Grounding Guidance

Each finding must appear at an anatomically correct location—*e.g.*, cardiomegaly in the cardiac silhouette, pleural effusion in the lower lung zones. We translate this domain knowledge into a cross-attention constraint.

Anatomy mask extraction. Using a frozen segmentation model f_{seg} (CheXmask [24]), we extract anatomical region masks from the predicted clean image: $\{M_r\} = f_{\text{seg}}(\hat{x}_0)$, where $r \in \{\texttt{heart}, \texttt{left\_lung}, \texttt{right\_lung}\}$.

Finding–anatomy mapping. A fixed mapping φ associates each CheXpert finding with its expected region: cardiac findings (*e.g.*, cardiomegaly) → `heart`; pulmonary findings (*e.g.*, pleural effusion, pneumothorax) → `ipsilateral lung` (laterality parsed from report; bilateral when unspecified); findings without a single target region (*e.g.*, fracture) are guided by $\mathcal{L}_{\text{cls}}$ only. The mapping φ is intentionally conservative and grounds a finding only when a single, anatomically reliable anchor exists; AMFG does not claim universal anatomical grounding for all CXR labels.

Grounding loss. For findings whose anatomical target is defined by φ, we constrain the corresponding attention map A_i to concentrate within the target region:

$$\mathcal{L}_{\text{attn}} = \sum_{i:\,\varphi(d_i)\ \text{defined}} \left\| A_i \odot \left(1 - M_{\varphi(d_i)}\right) \right\|_F^2, \tag{2}$$

where $\odot$ denotes element-wise multiplication. This penalizes attention mass for finding d_i that falls outside its anatomical region, with the gradient $\nabla_{z_t}\mathcal{L}_{\text{attn}}$ shifting attention toward the correct locations. Unlike text-based spatial cues in prior work [10], our mapping φ directly encodes clinical domain knowledge via a pretrained segmentation model.

3.4 Finding Interaction Guidance

When multiple findings co-occur, their spatial relationships must remain consistent. We introduce two interaction constraints.

Non-overlap. The overlap term is a soft attention-space regularizer that discourages distinct finding tokens from attending to the same region, encouraging each finding to occupy its own anatomical area:

$$\mathcal{L}_{\text{overlap}} = \sum_{i \neq j} \| A_i \odot A_j \|_F^2 . \tag{3}$$

Bilateral symmetry. For bilateral findings (*e.g.*, "bilateral pleural effusion"), attention distributions over the left and right lungs should be symmetric:

$$\mathcal{L}_{\text{sym}} = \sum_{i \in \mathcal{B}} \left\| A_i^L - \text{flip}(A_i^R) \right\|_F^2, \tag{4}$$

where $\mathcal{B}$ is the set of bilateral findings and A_i^L, A_i^R are attention maps masked to the left and right lungs, with $\text{flip}(\cdot)$ denoting horizontal reflection. The combined interaction loss is $\mathcal{L}_{\text{spatial}} = \mathcal{L}_{\text{overlap}} + \lambda_{\text{sym}}\,\mathcal{L}_{\text{sym}}$.

3.5 Inference

The total guidance loss is:

$$\mathcal{L} = w_1\,\mathcal{L}_{\text{cls}} + w_2\,\mathcal{L}_{\text{attn}} + w_3\,\mathcal{L}_{\text{spatial}}. \tag{5}$$

At each step, we update the latent with a noise-level-aware schedule:

$$z_t \leftarrow z_t - \eta \cdot g(t) \cdot \nabla_{z_t}\mathcal{L}, \tag{6}$$

where η is the step size and $g(t) = \sqrt{1 - \bar{\alpha}_t}$ scales guidance strength proportionally to the noise level, applying stronger correction at noisier (earlier) steps. Following prior work [10,9], we apply guidance only during the first $\lfloor \alpha T \rfloor$ steps, allowing later steps to refine details without interference. Segmentation masks are recomputed every k denoising steps and cached in between to limit overhead to approximately 1.3× the base model's inference time (measured with k=5 on a single A100).

4 Experiments

4.1 Experimental Setup

Dataset and Baselines. We evaluate on 1,000 PA-view test pairs from MIMIC-CXR-JPG [25], including a multi-finding subset of 500 pairs with two or more findings. All methods are evaluated on the same test set using identical metric pipelines with publicly available code and weights. *Training-free* (applied to frozen RoentGen [1] with identical DDIM schedule, random seeds, and test prompts): FreeU [18], PAG [17], A&E [9], STORM [10], TFG [11].

Training-required: XReal [8] (adapter training), reproduced with its publicly released weights.

Implementation details. 50-step DDIM, α=0.6, k=5, w_1=1, w_2=10, w_3=5; single NVIDIA A100 GPU. All results are averaged over 5 random seeds.

4.2 Evaluation

We employ four complementary metrics: (1) **FID**: image quality in domain-specific DenseNet-121 feature space [1]. (2) **mAUROC**: finding presence assessed by a CheXpert-pretrained DenseNet-121 classifier [23] that is *independent* of the guidance classifier f_{cls} (trained on NIH ChestX-ray14). (3) **Dice**: for each finding d_i, we compute the GradCAM activation map from the evaluation classifier, binarize it at τ=0.5, and measure its overlap with the anatomical mask $M_{\varphi(d_i)}$ from f_{seg} (CheXmask [24]). Although f_{seg} is shared between guidance and evaluation, cross-validation with an independent segmentor (PSPNet [23]) yields r=0.94 (p<0.001), supporting robustness to the anatomical mask source; we note this is a correlation-based check rather than a fully independent evaluation. (4) **MF-Acc**: the fraction of generated images for which *all* prompted findings satisfy Dice(d_i) > δ (δ=0.2), *i.e.*, every finding is correctly localized to its expected anatomical region.

4.3 Comparison with Baselines

Table 1 presents quantitative results. Among training-free methods, AMFG achieves the best performance across all four metrics, with particularly pronounced margins on anatomical metrics: Dice and MF-Acc show substantial gains over the strongest training-free baselines.

Notably, AMFG outperforms XReal—which requires adapter training—in both Dice and MF-Acc, while remaining competitive in FID and mAUROC. This suggests that, on localization metrics, training-free anatomical grounding can outperform trained spatial control, even without adapter training.

Figure 2 shows representative AMFG generations for challenging multi-finding prompts, including bilateral, laterality-specific, and three-finding cases.

Table 1. Quantitative comparison on MIMIC-CXR test set (mean ± std over 5 seeds). All methods evaluated with identical metric pipelines. Best training-free in **bold**; best overall underlined. ‡: requires adapter training.

Method	FID↓	mAUROC↑	Dice↑	MF-Acc(%)↑
RoentGen [1]	42.3±1.2	.681±.012	.218±.015	11.4±1.8
+ FreeU [18]	39.8±1.1	.689±.010	.225±.013	12.6±1.5
+ PAG [17]	38.5±1.0	.695±.011	.231±.014	13.8±1.7
+ A&E [9]	37.1±1.3	.724±.009	.267±.016	18.2±2.1
+ STORM [10]	35.6±1.1	.738±.010	.289±.018	21.6±2.3
+ TFG [11]	34.2±0.9	.751±.008	.304±.015	24.8±1.9
+ AMFG (Ours)	**33.5**±1.0	**.763**±.009	**.458**±.017	**42.6**±2.5
XReal‡ [8]	31.8±0.8	.772±.007	.421±.019	38.4±2.2

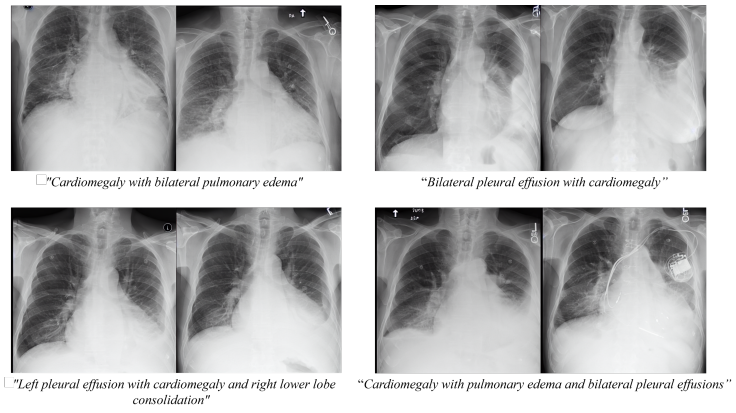

Fig. 2. Representative AMFG generations for multi-finding prompts. Clockwise from top-left: (1) cardiomegaly + bilateral pulmonary edema; (2) bilateral pleural effusion + cardiomegaly; (3) cardiomegaly + pulmonary edema + bilateral pleural effusions; (4) left pleural effusion + cardiomegaly + right lower-lobe consolidation.

4.4 Ablation Study and Downstream Task

Ablation Study. Table 2 presents the contribution of each component. Adding $\mathcal{L}_{\text{cls}}$ alone improves mAUROC substantially but raises Dice only marginally, confirming that presence alone does not ensure anatomical correctness. Adding $\mathcal{L}_{\text{attn}}$ yields the largest single Dice gain, the core contribution: anatomical grounding fundamentally improves spatial placement. Notably, $\mathcal{L}_{\text{spatial}}$ *without* $\mathcal{L}_{\text{attn}}$ yields only a modest Dice improvement, confirming $\mathcal{L}_{\text{attn}}$ as the primary driver. The full combination further improves both Dice and MF-Acc, while the adaptive schedule outperforms uniform guidance, validating the noise-level-aware design.

Downstream Task. The practical value of synthetic CXR data lies in its ability to improve downstream classifiers, particularly for rare findings where labeled examples are scarce. We evaluate whether anatomically correct gener-

Table 2. Ablation study of AMFG guidance components (mean over 5 random seeds).

$\mathcal{L}_{\text{cls}}$	$\mathcal{L}_{\text{attn}}$	$\mathcal{L}_{\text{spatial}}$	FID↓	mAUROC↑	Dice↑	MF-Acc(%)↑
			42.3	.681	.218	11.4
✓			36.8	.742	.236	14.2
✓	✓		34.1	.755	.431	38.6
✓		✓	35.5	.748	.271	19.4
✓	✓	✓	**33.5**	**.763**	**.458**	**42.6**
Guidance schedule:						
Uniform $g(t)$=1			34.8	.756	.438	39.2
Adaptive (default)			**33.5**	**.763**	**.458**	**42.6**

ation translates into downstream utility. We train a DenseNet-121 classifier on the MIMIC-CXR training split and augment it with synthetic images generated by each method (1,000 images per method, matched to the finding distribution of the training set). We report mAUROC on the held-out test split, separately for all 14 findings and for the rare-finding subset (pneumothorax, lung lesion, consolidation, pneumonia; each with prevalence <5%).

Table 3. Downstream classification performance (mAUROC) when augmenting a DenseNet-121 classifier with synthetic CXRs from each generation method (mean over 3 seeds).

Augmentation source	All findings	Rare findings
None (real only)	.798	.672
+ RoentGen (unguided)	.804	.683
+ TFG	.811	.698
+ XReal[‡]	.817	.714
+ AMFG (Ours)	**.824**	**.738**

Table 3 shows that AMFG-augmented training yields the highest downstream mAUROC, with the gap most pronounced on rare findings. Unguided RoentGen augmentation provides only marginal gains, and TFG—despite improving finding presence—does not substantially benefit rare-finding detection. Here, anatomically correct placement prevents the classifier from learning spurious spatial correlations, confirming that anatomical correctness in synthetic data directly translates into downstream utility.

5 Discussion

The dominant role of $\mathcal{L}_{\text{attn}}$ is consistent with Vilouras *et al.* [16]: the cross-attention miscalibration they report is precisely what our grounding loss corrects, and this correction propagates to downstream gains—most notably on

rare findings—suggesting that Dice and MF-Acc should complement FID and mAUROC as standard CXR generation metrics. This points to a broader conclusion: the bottleneck is *spatial supervision* rather than model capacity. Without any adapter training, inference-time anatomical constraints remain competitive in FID/mAUROC and surpass trained adapters on localization, while operating through standard latent-diffusion interfaces that make AMFG applicable to other frozen CXR backbones. Validating this portability across additional backbones and datasets, and automating φ via knowledge graphs [26], are natural next steps.

Acknowledgments. This work was supported in part by the Institute of Information & Communications Technology Planning & Evaluation (IITP) grant No. RS-2021-II211343 (Artificial Intelligence Graduate School Program, Seoul National University) and the National Research Foundation of Korea (NRF) grant No. RS-2024-00354666, both funded by the Korea government (MSIT); and by the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare (Grant No. RS-2025-02307233).

Disclosure of Interests. The authors have no competing interests.

References

1. Bluethgen, C., et al.: A vision–language foundation model for the generation of realistic chest X-ray images. Nat. Biomed. Eng. **9**, 494–506 (2025)
2. Kazerouni, A., et al.: Diffusion models in medical imaging: A comprehensive survey. Med. Image Anal. **88**, 102846 (2023)
3. Ktena, I., et al.: Generative models improve fairness of medical classifiers under distribution shifts. Nat. Med. **30**, 1166–1173 (2024)
4. Zhou, Z., Guo, Y., Tang, R., et al.: Privacy enhancing and generalizable deep learning with synthetic data for mediastinal neoplasm diagnosis. npj Digit. Med. **7**, 293 (2024)
5. Cooke, L.H., Jung, M., Brendel, J.M., Kerkovits, N.M., Foldyna, B., Lu, M.T., Raghu, V.K.: RoentMod: a synthetic chest X-ray modification model to identify and correct image interpretation model shortcuts. npj Digit. Med. **9**, 324 (2026).
6. Pérez-García, F., Bond-Taylor, S., Sanchez, P.P., van Breugel, B., Castro, D.C., Sharma, H., Salvatelli, V., Wetscherek, M.T.A., Richardson, H., Lungren, M.P., Nori, A., Alvarez-Valle, J., Oktay, O., Ilse, M.: RadEdit: stress-testing biomedical vision models via diffusion image editing. In: Computer Vision – ECCV 2024, pp. 358–376. Springer Nature, Cham (2025).
7. Han, W., Kim, C., Ju, D., Shim, Y., Hwang, S.J.: Advancing text-driven chest X-ray generation with policy-based reinforcement learning. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2024, LNCS, vol. 15003, pp. 56–66. Springer, Cham (2024)
8. Hashmi, A.U.R., et al.: XReal: Realistic anatomy and pathology-aware X-ray generation via controllable diffusion model. arXiv:2403.09240 (2024)
9. Chefer, H., et al.: Attend-and-excite: Attention-based semantic guidance for text-to-image diffusion models. ACM Trans. Graph. **42**(4) (2023)

10. Han, W., et al.: Spatial transport optimization by repositioning attention map for training-free text-to-image synthesis. In: CVPR 2025 (2025)
11. Ye, H., Lin, H., Han, J., Xu, M., Liu, S., Liang, Y., Ma, J., Zou, J., Ermon, S.: TFG: Unified Training-Free Guidance for Diffusion Models. In: Advances in Neural Information Processing Systems 37 (NeurIPS 2024) (2024).
12. Moroianu, S.L., et al.: Improving performance, robustness, and fairness of radiographic AI models with finely-controllable synthetic data. arXiv:2508.16783 (2025)
13. Lee, S., et al.: LLM-CXR: Instruction-finetuned LLM for CXR image understanding and generation. In: ICLR 2024 (2024)
14. Konz, N., et al.: Anatomically-controllable medical image generation with segmentation-guided diffusion models. In: MICCAI 2024 (2024)
15. Min, H., et al.: InstructX2X: An interpretable local editing model for counterfactual medical image generation. In: MICCAI 2025 (2025)
16. Vilouras, K., et al.: Anatomy-Grounded weakly supervised prompt tuning for chest X-ray latent diffusion models. arXiv:2506.10633 (2025)
17. Ahn, D., Cho, H., Min, J., et al.: Self-rectifying diffusion sampling with perturbed-attention guidance. In: Computer Vision – ECCV 2024, LNCS, vol. 15131, pp. 1–17. Springer, Cham (2024)
18. Si, C., et al.: FreeU: Free lunch in diffusion U-Net. In: CVPR 2024 (2024)
19. Askari, H., et al.: Training-free medical image inverses via bi-level guided diffusion models. In: WACV 2025 (2025)
20. Fei, B., Li, Y., Yang, W., Gao, H., Xu, J., Ma, L., Yang, Y., Zhou, P.: A diffusion model for universal medical image enhancement. Commun. Med. **5**, 294 (2025).
21. Irvin, J., et al.: CheXpert: A large chest radiograph dataset with uncertainty labels and expert comparison. In: AAAI 2019, pp. 590–597 (2019)
22. Rajpurkar, P., et al.: CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning. arXiv:1711.05225 (2017)
23. Cohen, J.P., et al.: TorchXRayVision: A library of chest X-ray datasets and models. In: Medical Imaging with Deep Learning (2022)
24. Gaggion, N., Mosquera, C., Mansilla, L., et al.: CheXmask: a large-scale dataset of anatomical segmentation masks for multi-center chest X-ray images. Sci. Data **11**, 511 (2024)
25. Johnson, A.E.W., Pollard, T.J., Greenbaum, N.R., et al.: MIMIC-CXR-JPG, a large publicly available database of labeled chest radiographs. arXiv:1901.07042 (2019)
26. Jain, S., et al.: RadGraph: Extracting clinical entities and relations from radiology reports. In: NeurIPS 2021 Datasets and Benchmarks (2021)