

GL-LCM: Global-Local Latent Consistency Models for Fast High-Resolution Bone Suppression in Chest X-Ray Images

Yifei Sun¹, Zhanghao Chen¹, Hao Zheng¹, Yuqing Lu², Lixin Duan³, Fenglei Fan⁴, Ahmed Elazab⁵, Xiang Wan⁶, Changmiao Wang^{6,✉}, and Ruiquan Ge^{1,✉}

¹ Hangzhou Dianzi University, Hangzhou, China
gespring@hdu.edu.cn

² The Chinese University of Hong Kong, Hong Kong, China

³ University of Electronic Science and Technology of China, Shenzhen, China

⁴ City University of Hong Kong, Hong Kong, China

⁵ Shenzhen University, Shenzhen, China

⁶ Shenzhen Research Institute of Big Data, Shenzhen, China
cmwangalbert@gmail.com

Abstract. Chest X-Ray (CXR) imaging for pulmonary diagnosis raises significant challenges, primarily because bone structures can obscure critical details necessary for accurate diagnosis. Recent advances in deep learning, particularly with diffusion models, offer significant promise for effectively minimizing the visibility of bone structures in CXR images, thereby improving clarity and diagnostic accuracy. Nevertheless, existing diffusion-based methods for bone suppression in CXR imaging struggle to balance the complete suppression of bones with preserving local texture details. Additionally, their high computational demand and extended processing time hinder their practical use in clinical settings. To address these limitations, we introduce a Global-Local Latent Consistency Model (GL-LCM) architecture. This model combines lung segmentation, dual-path sampling, and global-local fusion, enabling fast high-resolution bone suppression in CXR images. To tackle potential boundary artifacts and detail blurring in local-path sampling, we further propose Local-Enhanced Guidance, which addresses these issues without additional training. Comprehensive experiments on a self-collected dataset SZCH-X-Rays, and the public dataset JSRT, reveal that our GL-LCM delivers superior bone suppression and remarkable computational efficiency, significantly outperforming several competitive methods. Our code is available at <https://github.com/diaoquesang/GL-LCM>.

Keywords: Chest X-Ray · Bone Suppression · Global-Local · Latent Consistency Model · Dual-Energy Subtraction.

1 Introduction

Lung disease remains a major global health challenge, contributing to high rates of morbidity and mortality [15, 30]. Among available diagnostic tools, Chest X-

Ray (CXR) is widely regarded as the primary imaging modality for evaluating pulmonary conditions such as inflammation, tuberculosis, and lung masses. This preference is due to its accessibility, affordability, and low radiation exposure. However, interpreting CXR images poses notable difficulties. Even experienced radiologists may overlook lesions because superimposed bone structures can obscure critical diagnostic information [13]. In light of these challenges, the development of bone suppression techniques in CXR imaging offers great promise for enhancing the diagnostic capabilities of radiologists [12] and improving the performance of computer-aided lung lesion detection systems [23]. Currently, Dual-Energy Subtraction (DES) imaging is the most commonly employed bone suppression technique in clinical practice. By utilizing two X-ray exposures at different energy levels, DES effectively reduces the visual clutter caused by overlapping bones. However, DES imaging requires specialized equipment and increases radiation exposure, making it less accessible and impractical in resource-limited settings.

Conventional statistical-based methods for bone suppression in CXR imaging [10, 19] offer an alternative to specialized DES equipment. However, these methods require precise segmentation and boundary annotation, hampering their practical applicability. In contrast, recent advancements in deep learning have enabled models to automatically learn complex image features, providing more accurate bone suppression. This has led to the development of advanced techniques for bone suppression. For instance, Gusarev *et al.* [6] utilized an Auto-Encoder (AE) to generate soft tissue images from CXR images. Similarly, Zhou *et al.* [31] proposed a Multi-scale Conditional Adversarial Network (MCA-Net) drawing inspiration from Generative Adversarial Networks (GANs) [5]. Additionally, Wang *et al.* [23] introduced an improved U-Net with self-attention. Despite these efforts, performance limitations persist, limiting their applications in real-world clinical settings.

Recently, diffusion models have demonstrated strong capabilities in various generation tasks [4, 25, 28]. Although some studies have successfully used diffusion models for bone suppression in CXR imaging [1, 21], applying these models to perform fast high-resolution bone suppression still encounters challenges given the crucial requirement in clinical settings [24]. Particularly, these methods, despite being end-to-end, fail to balance global suppression of bone structures with local detail retention, primarily due to the difficulty of processing global bone residuals and local detail features through a single network. Additionally, existing diffusion models for bone suppression are computationally demanding, requiring significant processing time, which is impractical for clinical applications.

To overcome these challenges, we propose the Global-Local Latent Consistency Model (GL-LCM), a novel framework for fast high-resolution bone suppression in CXR images based on Latent Consistency Models (LCMs) [14]. Our key contributions are summarized as follows: First, the GL-LCM architecture facilitates effective bone suppression while retaining texture details. This is achieved through the design of dual-path sampling in the latent space combined with global-local fusion in the pixel space. Second, GL-LCM significantly enhances

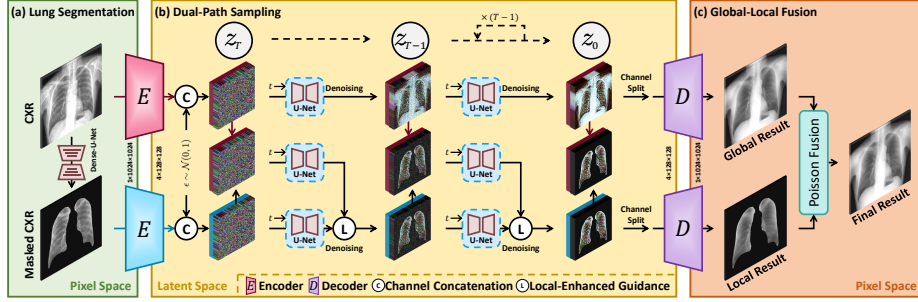


Fig. 1. Overview of GL-LCM framework. (a) Lung segmentation in the pixel space, (b) Dual-path sampling in the latent space, and (c) Global-local fusion in the pixel space.

inference efficiency, which requires only approximately 10% of the inference time of current diffusion-based methods, making it more suitable for clinical applications. Third, we introduce Local-Enhanced Guidance (LEG) to mitigate potential boundary artifacts and detail blurring issues in local-path sampling without additional training. Fourth, extensive experiments on both the self-collected dataset SZCH-X-Rays and the public dataset JSRT demonstrate exceptional performance and efficiency of our GL-LCM.

2 Proposed Method

The proposed GL-LCM framework, illustrated in Fig. 1, operates in three key stages. Initially, lung regions are segmented from CXR images in the pixel space using a pre-trained Dense-U-Net [26]. Next, using LEG, dual-path sampling of conditional LCMs is employed to produce both global and local soft tissue data. Finally, these global and local outputs are decoded back into the pixel space and integrated through Poisson Fusion to yield the final soft tissue image.

2.1 Conditional Latent Consistency Models

The LCMs [14] represent a new class of generative models that efficiently produce high-resolution images while preserving quality. Unlike traditional consistency models operating in the pixel space [20], LCMs reduce computational demands by transforming data from the pixel space X to a latent space Z . Informed by the principles of LCMs, our GL-LCM model incorporates both a forward and a reverse process within the latent space.

The forward process of our GL-LCM can be defined as:

$$z_t = \sqrt{\bar{\alpha}_t} z_0 + \sqrt{1 - \bar{\alpha}_t} \epsilon, \quad \epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \quad (1)$$

where ϵ is a noise component sampled from a Gaussian distribution, and $\bar{\alpha}_t := \prod_{s=0}^t \alpha_s$. Here, $\alpha_t = 1 - \beta_t$ is a differentiable function of timestep t determined

by the denoising LCM sampler. The training loss of GL-LCM is expressed as:

$$\mathcal{L}(\epsilon_\theta) = \sum_{t=1}^T \mathbb{E}_{z_0, \epsilon} \left[\|\epsilon_\theta(z_t, t, \tilde{z}) - \epsilon\|_2^2 \right], \quad (2)$$

where ϵ_θ represents the predicted noise at timestep t by the noise estimator network with parameters θ , and $\tilde{z}$ is a given image condition.

During the reverse process of GL-LCM, starting from random noise $z_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, the final result z_0 is predicted through a multi-step denoising process:

$$z_{t-1} = \begin{cases} \sqrt{\alpha_{t-1}} \left(c_{\text{out}}(t) \frac{z_t - \frac{1-\alpha_t}{\sqrt{1-\alpha_t}} \epsilon_\theta(z_t, t, \tilde{z})}{\sqrt{\alpha_t}} + c_{\text{skip}}(t) z_t \right) + \frac{1-\alpha_{t-1}}{\sqrt{1-\alpha_{t-1}}} \epsilon, & 1 < t \leq T \\ c_{\text{out}}(t) \frac{z_t - \frac{1-\alpha_t}{\sqrt{1-\alpha_t}} \epsilon_\theta(z_t, t, \tilde{z})}{\sqrt{\alpha_t}} + c_{\text{skip}}(t) z_t, & t = 1, \end{cases} \quad (3)$$

where T is the total sampling timesteps, and $c_{\text{out}}(t)$ and $c_{\text{skip}}(t)$ are differentiable functions such that $c_{\text{out}}(0) = 0$, and $c_{\text{skip}}(0) = 1$.

We adopt a U-Net architecture enhanced with multi-resolution attention [2] to serve as the noise estimator network with parameters θ . Conditional guidance is achieved through channel concatenation of z_t and $\tilde{z}$. Additionally, we utilize a VQGAN [3] as the encoder E and decoder D to convert data between the pixel and latent spaces.

2.2 Dual-Path Sampling

We introduce a dual-path architecture for LCM sampling to balance global bone suppression and the retention of local details. As illustrated in Fig. 1 (b), this architecture comprises a global path (top) implementing global bone suppression and a local path (bottom) focusing on preserving the pulmonary texture details.

Local-Enhanced Guidance. To implement dual-path sampling, we apply different conditions to the conditional LCM. Initially, CXR and masked CXR data were intended to guide the global and local paths, respectively. However, we observed that solely using a local condition in the local path could result in potential boundary artifacts and detail blurring, significantly diminishing the local path’s output quality. This issue may arise from insufficient local guidance, and unnecessary control outside the lung region. Therefore, we recommend enhancing local guidance while reducing global control to address these challenges effectively.

Classifier-Free Guidance (CFG) [7] is a widely used technique in diffusion models to enhance the guidance strength and the quality of generated outputs. This is achieved by incorporating both conditional and unconditional scenarios during the training phase and subsequently adjusting the weights of their predicted outcomes during the inference phase. Drawing inspiration from CFG, we propose LEG in our local-path sampling methodology, which is defined as:

$$\nabla_{z_{l,t}} \log p(z_{l,t} | \tilde{z}_l, \tilde{z}_g) = \alpha_l \cdot \nabla_{z_{l,t}} \log p(z_{l,t} | \tilde{z}_l) + (1 - \alpha_l) \cdot \nabla_{z_{l,t}} \log p(z_{l,t} | \tilde{z}_g), \quad (4)$$

where $z_{l,t}$ represents the sample at a given timestep t . The terms $\tilde{z}_l$ and $\tilde{z}_g$ refer to the local and global conditions, respectively, while α_l denotes the weight associated with the local condition. In our GL-LCM, LEG can be applied without necessitating additional training. This is because the dual-path sampling mechanism inherently encompasses training for both global and local conditions, thereby streamlining the integration of LEG into the model.

2.3 Global-Local Fusion

To seamlessly combine the global and local outcomes derived from dual-path sampling, we employ the Poisson Fusion method for global-local integration. This approach effectively preserves the gradient information within the local result, ensuring the retention of fine textures and edges. By utilizing image gradients, Poisson Fusion facilitates a smooth transition between the global low-frequency information and the local high-frequency details. This results in a final output with exceptional global bone suppression performance and remarkable local detail retention. Poisson Fusion is implemented by solving the following equations:

$$\begin{cases} \nabla \cdot (\nabla R) = \nabla \cdot (\nabla S_l), & \text{in } M(I_g) \\ R = S_g, & \text{outside of } M(I_g), \end{cases} \quad (5)$$

where R , S_l , S_g , and $M(I_g)$ represent the fusion result, the local-path result, the global-path result, and the lung region mask calculated by the Dense-U-Net from the CXR image, respectively. The boundary condition is defined as $R = S_g$.

3 Experiments

3.1 Data and Implementation

Data Preparation. Acquiring paired CXR and DES soft tissue images is challenging due to the significant scarcity of bone suppression datasets. To address this gap, we utilized the SZCH-X-Rays dataset, a large and high-quality collection from our partner hospital, along with JSRT [18], the largest publicly available bone suppression dataset to date, for comprehensive evaluation.

The SZCH-X-Rays Dataset comprises 741 pairs of posterior-anterior CXR and DES soft tissue images, obtained using a dual-exposure DES digital radiography system (Discovery XR656, GE Healthcare) from our partner hospital. Images affected by operational errors, significant motion artifacts, pleural effusion, and pneumothorax were excluded. This dataset is partitioned into 592 training images, 74 validation images, and 75 test images. All images were resized to a resolution of 1024×1024 pixels for consistency.

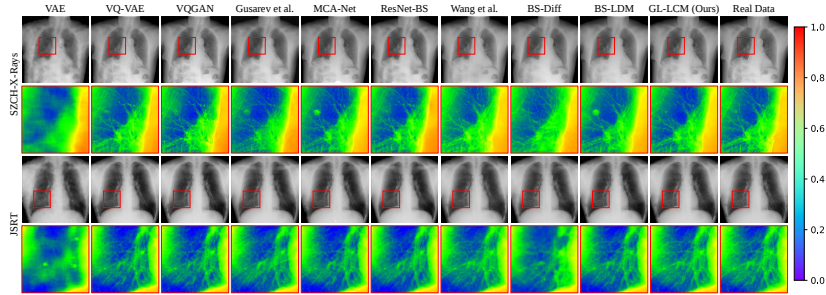


Fig. 2. Bone suppression and pseudo-color zoomed-in views of GL-LCM compared with state-of-the-art approaches on SZCH-X-Rays (top) and JSRT (bottom) datasets.

The JSRT Dataset includes 241 pairs of CXR and soft tissue images sourced from 14 medical centers. These images are categorized into 192 training images, 24 validation images, and 25 test images. Each image was resized to a resolution of 1024×1024 pixels and then transformed into a negative format.

Implementation Details. All experiments were carried out using PyTorch 2.0.1 on a single Nvidia A100 80G GPU within Ubuntu 20.04. The noise estimator network for GL-LCM was trained from scratch over 600 epochs (batch size of 4 and AdamW optimizer). Similarly, the VQGAN model integrated within GL-LCM was trained from scratch for 300 epochs (batch size of 4 and Adam optimizer). A dynamic learning rate schedule was applied in both cases, initializing at 0.0001. The scheduling parameter β_t varied from 0.00085 to 0.012, distributed across $T = 50$ total sampling timesteps using a scaled linear noise schedule. Finally, the weight α_l in LEG was set to 3.

3.2 Experimental Results

Comparisons. We benchmark our method against AE-based methods such as VAE [11], VQ-VAE [22] and Gusarev *et al.* [6], GAN-based methods like VQGAN [3] and MCA-Net [31], ResNet-based ResNet-BS [17], and U-Net-based Wang *et al.* [23]. We also compare with diffusion-based BS-Diff [1], and BS-LDM [21]. For evaluation, we use Bone Suppression Ratio (BSR) [8] as task-specific metric, along with three image quality assessment metrics: Mean Square Error (MSE), Peak Signal-to-Noise Ratio (PSNR) [9], and Learned Perceptual Image Patch Similarity (LPIPS) [29].

Table 1 presents a quantitative comparison of GL-LCM’s performance against these existing methods. Prior AE-based methods, such as VQ-VAE [22] and Gusarev *et al.* [6], show limited success due to their simplistic structures, which struggle with high-frequency details. Although diffusion-based techniques, such as BS-Diff [1] and BS-LDM [21], enhance image quality through iterative refinements, they fail to balance bone suppression and detail retention effectively.

GL-LCM stands out among current state-of-the-art methods. It achieves a 16.20%, 0.66 dB, and 20.00% improvement in MSE, PSNR, and LPIPS, re-

Table 1. Comparison of our method with other universal and task-specific approaches in the bone suppression task on the SZCH-X-Rays dataset. The best and second-best results are shown in **red** and **blue**, respectively.

Method	BSR (%) $\uparrow$	MSE (10^{-3}) $\downarrow$	PSNR $\uparrow$	LPIPS $\downarrow$
Universal Method				
VAE [11]	91.281 $\pm$ 3.088	1.169 $\pm$ 1.059	30.018 $\pm$ 2.007	0.237 $\pm$ 0.047
VQ-VAE [22]	94.485 $\pm$ 2.407	0.645 $\pm$ 0.596	32.600 $\pm$ 2.071	0.137 $\pm$ 0.029
VQGAN [3]	94.330 $\pm$ 3.402	0.923 $\pm$ 2.478	32.096 $\pm$ 2.420	0.083 $\pm$ 0.020
Task-Specific Method				
Gusarev <i>et al.</i> [6]	94.142 $\pm$ 2.666	1.028 $\pm$ 2.201	31.369 $\pm$ 2.385	0.156 $\pm$ 0.031
MCA-Net [31]	95.442 $\pm$ 2.095	0.611 $\pm$ 0.435	32.689 $\pm$ 1.939	0.079 $\pm$ 0.018
ResNet-BS [17]	94.508 $\pm$ 1.733	0.646 $\pm$ 0.339	32.265 $\pm$ 1.635	0.107 $\pm$ 0.022
Wang <i>et al.</i> [23]	89.767 $\pm$ 6.079	1.080 $\pm$ 0.610	29.963 $\pm$ 1.378	0.072 $\pm$ 0.016
BS-Diff [1]	92.428 $\pm$ 3.258	0.947 $\pm$ 0.510	30.627 $\pm$ 1.690	0.212 $\pm$ 0.041
BS-LDM [21]	94.159 $\pm$ 2.751	0.701 $\pm$ 0.293	31.953 $\pm$ 1.969	0.070 $\pm$ 0.018
GL-LCM (Ours)	95.611 $\pm$ 1.529	0.512 $\pm$ 0.293	33.347 $\pm$ 1.829	0.056 $\pm$ 0.015

Table 2. Comparison of our method with other universal and task-specific approaches in the bone suppression task on the JSRT dataset. The best and second-best results are shown in **red** and **blue**, respectively.

Method	BSR (%) $\uparrow$	MSE (10^{-3}) $\downarrow$	PSNR $\uparrow$	LPIPS $\downarrow$
Universal Method				
VAE [11]	85.646 $\pm$ 9.327	1.224 $\pm$ 0.749	29.814 $\pm$ 2.364	0.155 $\pm$ 0.032
VQ-VAE [22]	86.445 $\pm$ 8.881	0.986 $\pm$ 0.596	30.712 $\pm$ 2.273	0.062 $\pm$ 0.017
VQGAN [3]	86.594 $\pm$ 8.916	1.002 $\pm$ 0.606	30.635 $\pm$ 2.255	0.061 $\pm$ 0.017
Task-Specific Method				
Gusarev <i>et al.</i> [6]	89.283 $\pm$ 8.288	0.821 $\pm$ 0.570	31.700 $\pm$ 2.594	0.100 $\pm$ 0.024
MCA-Net [31]	86.887 $\pm$ 9.825	0.876 $\pm$ 0.625	31.577 $\pm$ 2.905	0.057 $\pm$ 0.017
ResNet-BS [17]	88.782 $\pm$ 8.905	0.960 $\pm$ 0.661	31.021 $\pm$ 2.576	0.060 $\pm$ 0.016
Wang <i>et al.</i> [23]	89.679 $\pm$ 9.477	1.013 $\pm$ 0.655	30.681 $\pm$ 2.431	0.075 $\pm$ 0.015
BS-Diff [1]	88.707 $\pm$ 8.859	1.003 $\pm$ 0.655	30.765 $\pm$ 2.504	0.154 $\pm$ 0.037
BS-LDM [21]	89.322 $\pm$ 9.562	0.783 $\pm$ 0.632	32.307 $\pm$ 3.231	0.058 $\pm$ 0.017
GL-LCM (Ours)	90.056 $\pm$ 10.635	0.746 $\pm$ 0.680	32.951 $\pm$ 3.799	0.052 $\pm$ 0.015

Table 3. Inference efficiency comparison of our method and other diffusion-based methods on the SZCH-X-Rays dataset. The best and second-best results are shown in **red** and **blue**, respectively.

Method	Sampler	Sampling Steps	Parameters	Inference Time (s)
BS-Diff [1]	DDPM	1000	254.7M	108.86
BS-LDM [21]	DDPM	1000	421.3M	84.62
GL-LCM (Ours)	LCM	50	436.9M	8.54

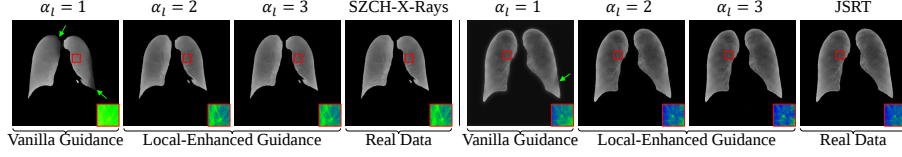


Fig. 3. The qualitative results of our ablation study of conditional guidance methods of GL-LCM on the SZCH-X-Rays and JSRT datasets. A pseudo-color zoomed-in view is shown in the bottom right corner, and the green arrows mark the boundary artifacts.

Table 4. Effect of various conditional guidance methods for local-path sampling on the SZCH-X-Rays and JSRT datasets. The best and second-best results are shown in red and blue, respectively.

Guidance Method	SZCH-X-Rays		JSRT	
	PSNR $\uparrow$	LPIPS $\downarrow$	PSNR $\uparrow$	LPIPS $\downarrow$
Vanilla Guidance	32.777 $\pm$ 2.091	0.058 $\pm$ 0.016	32.296 $\pm$ 3.454	0.073 $\pm$ 0.020
CFG [7]	32.315 $\pm$ 1.717	0.068 $\pm$ 0.013	32.613 $\pm$ 3.604	0.070 $\pm$ 0.015
LEG (Ours)	33.347 $\pm$ 1.829	0.056 $\pm$ 0.015	32.951 $\pm$ 3.799	0.052 $\pm$ 0.015

Table 5. Effect of various fusion strategies on the SZCH-X-Rays and JSRT datasets. The best and second-best results are shown in red and blue, respectively.

Fusion Strategy	SZCH-X-Rays		JSRT	
	PSNR $\uparrow$	LPIPS $\downarrow$	PSNR $\uparrow$	LPIPS $\downarrow$
$\times$	31.360 $\pm$ 2.079	0.091 $\pm$ 0.020	31.638 $\pm$ 3.078	0.074 $\pm$ 0.021
α -Fusion [16]	29.781 $\pm$ 1.522	0.181 $\pm$ 0.021	31.784 $\pm$ 3.043	0.092 $\pm$ 0.013
AE Fusion [27]	30.850 $\pm$ 1.806	0.141 $\pm$ 0.028	31.835 $\pm$ 3.075	0.061 $\pm$ 0.017
Poisson Fusion (Ours)	33.347 $\pm$ 1.829	0.056 $\pm$ 0.015	32.951 $\pm$ 3.799	0.052 $\pm$ 0.015

spectively, on the SZCH-X-Rays dataset. On the JSRT dataset, it demonstrates improvements of 5.49%, 0.64 dB, and 8.77% over other methods. In terms of BSR, GL-LCM achieves 95.61% on SZCH-X-Rays and 90.06% on JSRT. Fig. 2 illustrates the qualitative results for these datasets. Furthermore, the inference time of GL-LCM is about 10% of what is required by other diffusion-based methods, as outlined in Table 3.

Ablation Study. Our ablation study on SZCH-X-Rays and JSRT demonstrates the significant improvements brought by our approach. Table 4 and Table 5 present the quantitative results. The incorporation of LEG improves PSNR by 0.57 dB and 0.34 dB and reduces LPIPS by 3.45% and 25.71% on SZCH-X-Rays and JSRT, respectively. Unlike Vanilla Guidance solely using a local condition, LEG substantially enhances detail clarity and removes boundary artifacts, as shown in Fig. 3. Moreover, our global-local fusion strategy improves PSNR by 1.99 dB and 1.12 dB and reduces LPIPS by 38.46% and 14.75% on SZCH-X-Rays and JSRT, respectively. For comparison, the absence of a fusion strategy (LCM

baseline) is evaluated using only the global path. By combining the global-local design with LEG, our GL-LCM achieves superior performance overall.

4 Conclusion

In this study, we introduce a GL-LCM architecture, which is specifically designed to achieve fast and high-resolution bone suppression while preserving texture details. We also present LEG for effective guidance of local-path sampling, which mitigates potential boundary artifacts and detail blurring. Extensive experiments on self-collected and public datasets demonstrate that GL-LCM delivers impressive performance and efficiency, significantly surpassing many competitive methods. Looking ahead, we plan to further optimize the model by light-weighting its backbone and explore downstream applications, such as computer-aided diagnosis and clinical validation.

Acknowledgments. This work was funded by the National Undergraduate Innovation and Entrepreneurship Training Program of China (No. 202410336081), National Natural Science Foundation of China (No. 61702146, 62076084), and Guangdong Basic and Applied Basic Research Foundation (No. 2025A1515011617, 2022A1515110570).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Chen, Z., Sun, Y., Ge, R., Qin, W., Pan, C., Deng, W., Liu, Z., Min, W., Elazab, A., Wan, X., et al.: Bs-diff: Effective bone suppression using conditional diffusion models from chest x-ray images. In: 2024 IEEE International Symposium on Biomedical Imaging (ISBI). pp. 1–5. IEEE (2024)
2. Dhariwal, P., Nichol, A.: Diffusion models beat gans on image synthesis. *Advances in Neural Information Processing Systems* **34**, 8780–8794 (2021)
3. Esser, P., Rombach, R., Ommer, B.: Taming transformers for high-resolution image synthesis. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 12873–12883 (2021)
4. Fu, X., Yin, W., Hu, M., Wang, K., Ma, Y., Tan, P., Shen, S., Lin, D., Long, X.: Geowizard: Unleashing the diffusion priors for 3d geometry estimation from a single image. In: *European Conference on Computer Vision*. pp. 241–258. Springer (2024)
5. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. *Advances in Neural Information Processing Systems* **27** (2014)
6. Gusarev, M., Kuleev, R., Khan, A., Rivera, A.R., Khattak, A.M.: Deep learning models for bone suppression in chest radiographs. In: 2017 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB). pp. 1–7. IEEE (2017)
7. Ho, J., Salimans, T.: Classifier-free diffusion guidance. In: *NeurIPS 2021 Workshop on Deep Generative Models and Downstream Applications* (2021)

8. Hogeweg, L., Sánchez, C.I., van Ginneken, B.: Suppression of translucent elongated structures: applications in chest radiography. *IEEE Transactions on Medical Imaging* **32**(11), 2099–2113 (2013)
9. Huynh-Thu, Q., Ghanbari, M.: Scope of validity of psnr in image/video quality assessment. *Electronics Letters* **44**(13), 800–801 (2008)
10. Juhász, S., Horváth, Á., Nikhazy, L., Horváth, G., Horváth, Á.: Segmentation of anatomical structures on chest radiographs. In: *XII Mediterranean Conference on Medical and Biological Engineering and Computing*. pp. 359–362. Springer (2010)
11. Kingma, D.P., Welling, M., et al.: An introduction to variational autoencoders. *Foundations and Trends in Machine Learning* **12**(4), 307–392 (2019)
12. Li, H., Han, H., Li, Z., Wang, L., Wu, Z., Lu, J., Zhou, S.K.: High-resolution chest x-ray bone suppression using unpaired ct structural priors. *IEEE Transactions on Medical Imaging* **39**(10), 3053–3063 (2020)
13. Liu, Y., Zeng, F., Ma, M., Zheng, B., Yun, Z., Qin, G., Yang, W., Feng, Q.: Bone suppression of lateral chest x-rays with imperfect and limited dual-energy subtraction images. *Computerized Medical Imaging and Graphics* **105**, 102186 (2023)
14. Luo, S., Tan, Y., Huang, L., Li, J., Zhao, H.: Latent consistency models: Synthesizing high-resolution images with few-step inference. *arXiv preprint arXiv:2310.04378* (2023)
15. Pham, L., Phan, H., Palaniappan, R., Mertins, A., McLoughlin, I.: Cnn-moe based framework for classification of respiratory anomalies and lung disease detection. *IEEE Journal of Biomedical and Health Informatics* **25**(8), 2938–2947 (2021)
16. Porter, T., Duff, T.: Compositing digital images. In: *Proceedings of the 11th Annual Conference on Computer Graphics and Interactive Techniques*. pp. 253–259 (1984)
17. Rajaraman, S., Zamzmi, G., Folio, L., Alderson, P., Antani, S.: Chest x-ray bone suppression for improving classification of tuberculosis-consistent findings. *Diagnostics* **11**(5), 840 (2021)
18. Shiraishi, J., Katsuragawa, S., Ikezoe, J., Matsumoto, T., Kobayashi, T., Komatsu, K.i., Matsui, M., Fujita, H., Kodera, Y., Doi, K.: Development of a digital image database for chest radiographs with and without a lung nodule: receiver operating characteristic analysis of radiologists’ detection of pulmonary nodules. *American Journal of Roentgenology* **174**(1), 71–74 (2000)
19. Simkó, G., Orbán, G., Máday, P., Horváth, G.: Elimination of clavicle shadows to help automatic lung nodule detection on chest radiographs. In: *4th European Conference of the International Federation for Medical and Biological Engineering: ECIFMBE*. pp. 488–491. Springer (2009)
20. Song, Y., Dhariwal, P., Chen, M., Sutskever, I.: Consistency models. In: *International Conference on Machine Learning*. pp. 32211–32252. PMLR (2023)
21. Sun, Y., Chen, Z., Zheng, H., Ge, R., Liu, J., Min, W., Elazab, A., Wan, X., Wang, C.: Bs-ldm: Effective bone suppression in high-resolution chest x-ray images with conditional latent diffusion models. *arXiv preprint arXiv:2412.15670* (2024)
22. Van Den Oord, A., Vinyals, O., et al.: Neural discrete representation learning. *Advances in Neural Information Processing Systems* **30** (2017)
23. Wang, Y., Qu, Z., Dai, Z., Li, Y., Liu, Y., Wang, W., Xiao, L., Zhang, Y.: Rib suppression-based radiomics for diagnosis of neonatal respiratory distress syndrome in chest x-rays. *Scientific Reports* **15**(1), 4416 (2025)
24. Weber, T., Ingrisch, M., Bischl, B., Rügamer, D.: Cascaded latent diffusion models for high-resolution chest x-ray synthesis. In: *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. pp. 180–191. Springer (2023)

25. Wolleb, J., Bieder, F., Friedrich, P., Zhang, P., Durrer, A., Cattin, P.C.: Binary noise for binary tasks: Masked bernoulli diffusion for unsupervised anomaly detection. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 135–145. Springer (2024)
26. Yahyatabar, M., Jouvet, P., Cheriet, F.: Dense-unet: a light model for lung fields segmentation in chest x-ray images. In: 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). pp. 1242–1245. IEEE (2020)
27. Yao, R., Liu, C., Zhang, L., Peng, P.: Unsupervised anomaly detection using variational auto-encoder based feature extraction. In: 2019 IEEE International Conference on Prognostics and Health Management (ICPHM). pp. 1–7. IEEE (2019)
28. Ye, H., Lin, H., Han, J., Xu, M., Liu, S., Liang, Y., Ma, J., Zou, J.Y., Ermon, S.: Tfg: Unified training-free guidance for diffusion models. *Advances in Neural Information Processing Systems* **37**, 22370–22417 (2025)
29. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 586–595 (2018)
30. Zhen, K., Yang, P., Zhai, Z.: Epidemiology and management of paediatric pulmonary embolism: a much-needed focus of research. *The Lancet Respiratory Medicine* (2025)
31. Zhou, B., Lin, X., Eck, B., Hou, J., Wilson, D.: Generation of virtual dual energy images from standard single-shot radiographs using multi-scale and conditional adversarial network. In: Asian Conference on Computer Vision. pp. 298–313. Springer (2019)