Synthesis of Realistic Medical Images with Pathologies Using Diffusion Models with Application to Lung CT and Mammography

Arjun Krishna[a], Eric Papenhausen[b], Darius Coelho[b], and Klaus Mueller[ab]

[a]Computer Science Department, Stony Brook University, Stony Brook, NY, USA 11794
[b]12bit.ai, 1500 Stony Brook Road, Stony Brook, NY USA 11794

ABSTRACT

The scarcity of annotated medical images poses a significant challenge to developing and training accurate AI models for the detection of tumors and other pathologies. To address this, we introduce a novel method for synthesizing realistic medical images using Multi-Conditioned Denoising Diffusion Probabilistic Models, capable of generating images with and without tumors. Our approach leverages existing datasets to train the diffusion models, ensuring the synthetic images closely mimic real-world medical images. This controlled synthesis process allows for the creation of diverse datasets, enhancing the variability and richness of training data. By generating high-quality synthetic medical images, we aim to overcome the limitations of data scarcity and improve the performance and generalizability of AI models. While our primary goal is to increase the volume and diversity of training data, this method also holds potential benefits for underrepresented population groups by facilitating the inclusion of more varied demographic and pathological characteristics. Our results indicate that the diffusion model-generated medical images are indistinguishable from real images by radiologists, demonstrating their potential for effective use in AI model training. We also found that enriching sparse training data with our synthetic images can improve the accuracy of pathology detection AI classifiers. This innovative approach promises to significantly advance the field of AI-driven medical imaging, leading to more accurate and reliable diagnostic tools. Our paper presents first results on two specific important applications, lung CT and mammography.

Keywords: Synthetic Lung-CT, Synthetic Mammograms, Diffusion Models, Pathologies, Medical Imaging, Diagnostic Tools, Data Augmentation

1. INTRODUCTION

Significant progress has been made in deep learning-based applications for medical imaging; such as denoising,[1–3] cross-modality translations[4–9] and cancer detection.[10, 11] However, the potential of these applications is limited by the scarcity of highly specialized, accurate, and high-resolution annotated images necessary for robust training. Researchers have turned to synthetic medical image synthesis using generative AI, yet generating high-resolution phantom images with flawless anatomy and accurate annotations[12–14] remains a challenge. Current methods rely on unconditional generative models[15] but these often produce non-annotated images, failing to meet specific medical application needs.

Denoising Diffusion Probabilistic Models (DDPMs)[16] have emerged as an alternative, along with innovative techniques guiding the process towards specific distributions[17–20] during sampling or image generation. This paper explores a conditional generation approach via a condition-free trained DDPM,[19] expanding it to include multiple annotations/conditions for dynamic guidance during sampling. By synthesizing nodules, tumors, and lesions, both benign and malignant within anatomy-constrained lung-CT and mammograms, we demonstrate the model's versatility and ability to handle novel scenarios and rare cases broadening its applicability. To our knowledge, this is the first work to achieve such flexibility in generating full-resolution images with annotations, ensuring anatomical accuracy across multiple clinically relevant domains. We validate our approach using a Visual Turing Test[21] and other quantitative methods.

Further author information: Arjun Krishna: E-mail: arjkrishna@cs.stonybrook.edu

2. MULTI-CONDITIONED DENOISING DIFFUSION PROBABILISTIC MODEL

The DDPM iteratively transforms a Gaussian distribution into a lung-CT/mammogram image. The Markov Chain model learns the reverse of the forward diffusion process which is $q(x_t|x_{t-1}) \coloneqq N(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t I)$ with x_t as the latents with added noise and β_t as a fixed variance schedule.

Via the reparameterization trick; $x_t = \sqrt{\overline{\alpha}_t}x_0 + \sqrt{1-\overline{\alpha}_t}\epsilon$ with $\alpha_t \coloneqq 1$ - β_t and $\overline{\alpha}_t \coloneqq \prod_{i=1}^{t}\alpha_i$. The added noise $\epsilon \sim N(0, I)$ has the same dimensionality as the image and the sampled latents during training. The reverse diffusion process $p_\theta(x_{t-1}|x_t) = N(x_{t-1}; \mu_\theta(x_t, t), \sigma_t^2 I)$ is learned via a neural network p_θ via the below loss:

$$Loss = \|\epsilon - \epsilon_\theta(x_t, t)\|^2 = \|\epsilon - \epsilon_\theta(\sqrt{\overline{\alpha}_t}x_0 + \sqrt{1-\overline{\alpha}_t}\epsilon, t)\|^2 \tag{1}$$

Eq. 1 is used to train our DDPM, incorporating refinements from Nichol et al.[22] We train 2 DDPMs; one for the lung-CT on a dataset of lung-CT scans provided by the National Lung Screening Trial[23] and another for the mammograms on a dataset of mammography scans provided by RSNA.[24] We extracted around 50k images for training from scans contained in both datasets. For sampling post-training; x_{t-1} (via reparameterization):

$$x_{t-1} = \frac{1}{\sqrt{\alpha_t}}(x_t - \frac{\beta_t}{\sqrt{1-\overline{\alpha}_t}}\epsilon_\theta(x_t, t)) + \sqrt{\beta_t}\epsilon \tag{2}$$

Next, we discuss our modifications to the sampling strategy to facilitate multi-annotations guidance.

2.1 Multi-Condition Guidance

We explore the guidance techniques introduced by Choi et al.[19] and utilize them to exert precise control over the generation of lung CT images within all HU windows. The sampling process is guided towards a subset of images similar to a reference image y, by enforcing equivalence between the downsampled reference image and the intermediate latents x_t produced during sampling steps. Specifically, Choi et al. continuously refine the downsampled latent variable x_t within steps (T, a) to ensure that both x_t and y_t share low frequency contents. As such in each sampling step; y_t is computed from reference image y and x_t is refined as:

$$x_t = x_t + \phi_N(y_t) - \phi_N(x_t) \tag{3}$$

where $\phi_N(\dots)$ is a low-pass linear filter with N as the downsampling factor. The term is approximated by ensuring the latent x_t captures the missing low-frequency contents of y_t after sampling from the DDPM.

We argue that since the extent of low pass filtering of a linear filter ϕ can be controlled by its factor N, for any given set of conditional or guidance images $\{y_1, y_2 \dots y_m\}$ we should be able to fine-tune our algorithm with a set of integers $\{n_1, n_2 \dots n_m\}$, with each integer denoting the extent of downsampling for a linear filter corresponding to every conditional image, such that we could have a valid image generation via a trained DDPM sharing low level features (or similarity) with each of the conditional images. We modify Eq. 3 as:

$$x_{t-1} = x_{t-1} + \sum_{s=1}^{M} \begin{cases} (\phi_{n_s}(y_{s_{t-1}}) - \phi_{n_s}(x_{t-1})), & \text{if } t \geq a_s \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

Fig. 1a and Fig. 2 show some of the images generated with our model. For mammograms, we needed to perform texture/content transfer for tumor patches during the multi-conditioned sampling since at times the texture/style of the two conditioning images would vary greatly because of the density/fat content of the surrounding tissue or the entire breast; which is discussed in next section.

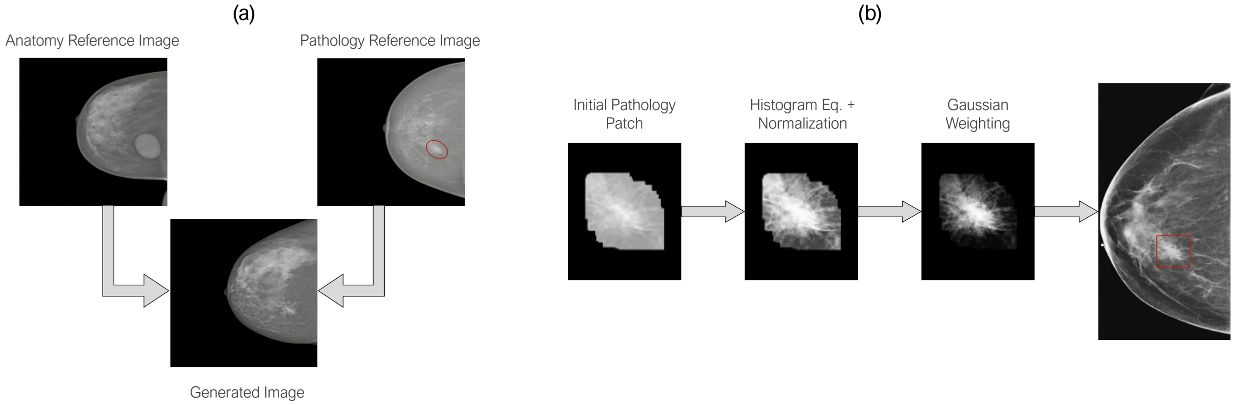

Figure 1: **(a)** Examples of a generated mammogram with pathology using multi-condition sampling, following the anatomy/texture of the anatomy reference image (top left) and the pathology reference image (top right). For mammograms, we pre-processed cancer tumors with texture and content transfer. **(b)** The transformations a pathology patch undergoes before multi-conditioned guidance. Histogram equalization improves contrast, followed by normalization to match pixel values with the reference image. Gaussian weighting reduces pixel intensity at the edges, eliminating artifacts from discontinuities between the patch and the reference image.

2.2 Mammography Diffusion Model

For the annotated mammography images the sampling process is conditioned on two images – a reference image from the RSNA dataset[24] and a pathology image from the much smaller annotated INbreast dataset[25] containing 8 cancer images. The reference image guides the diffusion model to generate a specific orientation (i.e. MLO or CC) as well as scanner and breast properties. The pathology image is taken from a second, annotated scan. It is transformed prior to conditioning to make its appearance closer to the reference image.

Fig. 1b illustrates the transformation applied to the pathology image to make it closer in appearance to the reference image. First, we perform histogram equalization followed by normalizing the values based on the reference image. Specifically, we normalize between the minimum reference image value and a quantile value between 0.975 and 0.995 chosen uniformly at random. Next, we apply a Gaussian weighting to reduce the intensity of toward the edge of the patch. We also apply a weighting during conditioning on the pathology image so that Eq. 3 becomes:

$$x_t = x_t + (\phi_N(y_t) - \phi_N(x_t)) * W \tag{5}$$

W is a weight matrix with the same dimension x_t and is zero for pixels outside the pathology patch (similar to the case for lung-CT tumors). For pixels inside the patch, W is proportional to a spherical bi-variate Gaussian function with a mean determined by the center of the patch and a variance based on the size of the patch.

3. EVALUATION

Lung-CT. We reran our Visual Turing Test[14] with the assistance of three radiologists to evaluate the realism of our generated lung CT images. The radiologists were asked to label a randomly selected lung CT image as “Real” or “Fake.” from a set of 30 real and generated images, one at a time. The images were randomly chosen from bone, lung, and soft-tissue windows.

Their responses had an accuracy of about 46% indicating our generative framework passed the Visual Turing Test, as expert radiologists could not distinguish between synthesized and real lung CT images.

Mammography. We evaluate the quality of the synthetic mammography images by comparing the classifier performance of a ConvNext[26] convolutional neural network trained for 20 epochs on real data augmented with an increasing amount of synthetic data. Specifically, we train on 7500 mammography images from the RSNA dataset[24] and evaluate classifier performance as we add synthetic data in increments of 2500. In all cases, 10% of the training dataset contained cancerous tumors while 90% contained no cancer.

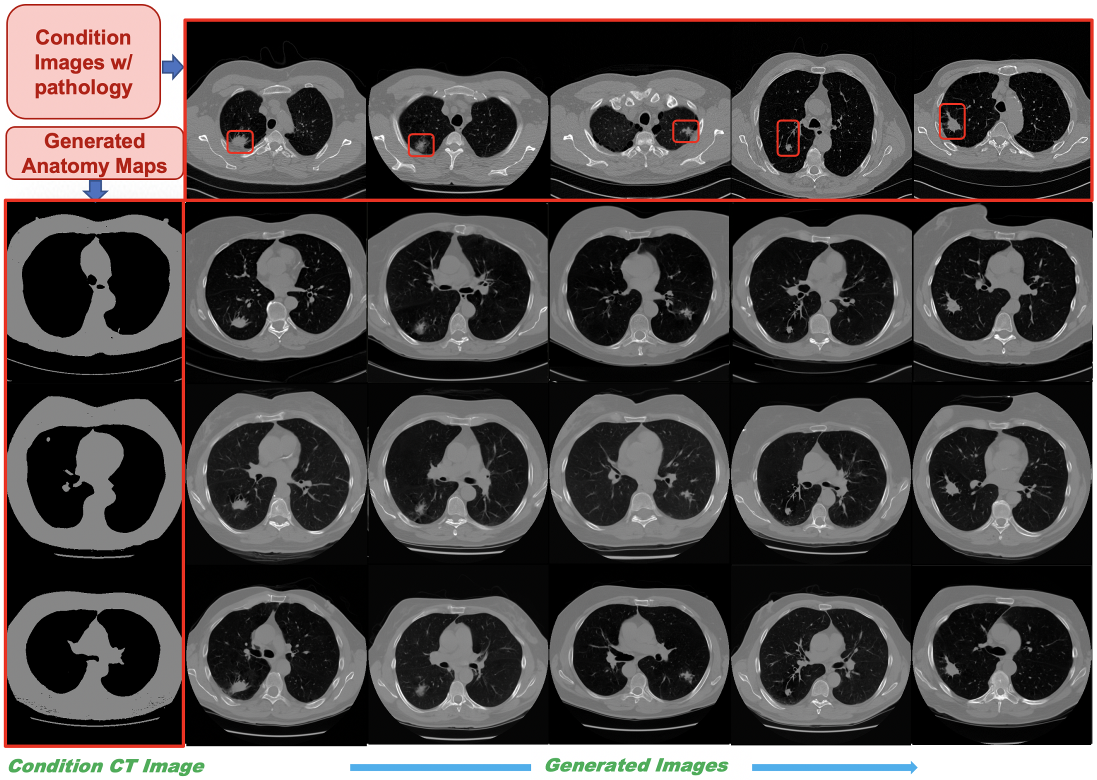

Figure 2: Examples of generated lung-CT images with pathology using multi-condition sampling. These images follow the anatomy of the conditional images in the left column (generated with B-Spline curves[14]) and the pathology of the real CT images in the top row. In many cases, the generated lesions realistically attach to the lung walls or surrounding tissues when in close proximity, this mimics the behavior of tumors in these regions.

To synthesize images containing cancer, a total of six pathology patches were taken from the INbreast[25] dataset. To increase tumor diversity, We applied random augmentations to these patches during image generation (i.e. rotation, translation, and flipping). The classifiers were evaluated on their ability to detect malignant tumors in mammography images taken from the mini-DDSM dataset.[27]

Table 1 shows that adding synthetic data improved the performance over real data alone. Interestingly, adding only 2500 synthetic images led to the best results, with performance degrading as more synthetic data was added. This is likely due to the small number of tumor patches used in synthesizing mammograms with cancer. Adding more synthetic images likely results in overfitting which is harming performance.

Training Size	7500 Real Data	+2500 Synth	+5000 Synth	+7500 Synth	+10000 Synth
AUC	0.588	0.671	0.639	0.603	0.599
F1 Score	0.324	0.391	0.363	0.326	0.319

Table 1: Performance metrics datasets with varying amounts of synthetic data.

4. CONCLUSION

Our paper presents a novel generative AI model for synthesizing realistic medical images using Multi-Conditioned Denoising Diffusion Probabilistic Models. This versatile model can generate images for unique scenarios and rare cases across various domains / modalities. We demonstrated that the synthetic images produced by our model can pass a visual Turing test and enhance classifier performance. In the future, we aim to broaden the range of domains and pathologies our model can address. Additionally, we will explore the potential for our synthetic images to entirely replace real data in the training of medical AI models.

REFERENCES

[1] Wolterink, J. M., Leiner, T., Viergever, M. A., and Išgum, I., "Generative adversarial networks for noise reduction in low-dose ct," *IEEE Transactions on Medical Imaging* **36**(12), 2536–2545 (2017).

[2] Yang, Q. et al., "Low-dose ct image denoising using a generative adversarial network with wasserstein distance and perceptual loss," *IEEE Transactions on Medical Imaging* **37**(6), 1348–1357 (2018).

[3] Kang, E., Koo, H. J., Yang, D. H., Seo, J. B., and Ye, J. C., "Cycle-consistent adversarial denoising network for multiphase coronary ct angiography," *Medical Physics* **46**(2), 550–562 (2019).

[4] Hiasa, Y. et al., "Cross-modality image synthesis from unpaired data using cyclegan," in [*International Workshop on Simulation and Synthesis in Medical Imaging*], 31–41 (2018).

[5] Yang, H. et al., "Unpaired brain mr-to-ct synthesis using a structure-constrained cyclegan," in [*Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*], 174–182 (2018).

[6] Jin, C.-B. et al., "Deep ct to mr synthesis using paired and unpaired data," *Sensors* **19**(10), 2361 (2019).

[7] Russ, T. et al., "Synthesis of ct images from digital body phantoms using cyclegan," *International Journal of Computer Assisted Radiology and Surgery* **14**, 1741 – 1750 (2019).

[8] Sandfort, V., Yan, K., Pickhardt, P. J., and Summers, R. M., "Data augmentation using generative adversarial networks (cyclegan) to improve generalizability in ct segmentation task," *Scientific Reports* **9** (2019).

[9] Zijlstra, F. et al., "Ct synthesis from mr images for orthopedic applications in the lower arm using a conditional generative adversarial network," in [*SPIE Medical Imaging: Image Processing*], **10949**, 387–393 (2019).

[10] Donnelly, J. e. a., "Asymmirai: Interpretable mammography-based deep learning model for 1-5-year breast cancer risk prediction.," *Radiology vol. 310,3* (2024).

[11] Yala, A. e. a., "Toward robust mammography-based models for breast cancer risk.," *Science translational medicine vol. 13,578* (2021).

[12] Han, K., Xiong, Y., You, C., Khosravi, P., Sun, S., Yan, X., Duncan, J., and Xie, X., "Medgen3d: A deep generative framework for paired 3d image and mask generation," (2023).

[13] Toda, R. et al., "Lung cancer ct image generation from a free-form sketch using style-based pix2pix for data augmentation," *Scientific Reports* **12** (2022).

[14] Krishna, A. e. a., "Image factory: A method for synthesizing novel ct images with anatomical guidance.," *Medical physics vol. 51,5* (2024).

[15] Park, H. Y. et al., "Realistic high-resolution body computed tomography image synthesis by using progressive growing generative adversarial network: Visual turing test," *JMIR Medical Informatics* **9** (2021).

[16] Ho, J., Jain, A., and Abbeel, P., "Denoising diffusion probabilistic models," *Advances in neural information processing systems* **33**, 6840–6851 (2020).

[17] Zhang, L., Rao, A., and Agrawala, M., "Adding conditional control to text-to-image diffusion models," (2023).

[18] Chung, H., Ryu, D., McCann, M. T., Klasky, M. L., and Ye, J. C., "Solving 3d inverse problems using pre-trained 2d diffusion models," (2022).

[19] Choi, J., Kim, S., Jeong, Y., Gwon, Y., and Yoon, S., "Ilvr: Conditioning method for denoising diffusion probabilistic models," *International Conference on Computer Vision* (2021).

[20] Ho, J. and Salimans, T., "Classifier-free diffusion guidance," (2022).

[21] Geman, D., Geman, S., Hallonquist, N., and Younes, L., "Visual turing test for computer vision systems," *PNAS* (2015).

[22] Nichol, A. Q. and Dhariwal, P., "Improved denoising diffusion probabilistic models," *openreview.net/forum?id=-NEXDKk8gZ* (2021).

[23] NIH(NCI), "National lung screening trial." https://www.cancer.gov/types/lung/research/nlst (2014). Accessed: January, 2022.

[24] Chris, C. Felipe, K. et al., "Rsna screening mammography breast cancer detection." https://kaggle.com/competitions/rsna-breast-cancer-detection (2022). Accessed: August, 2024.

[25] Ngx, T., "Inbreast 2012 [data set]." https://www.kaggle.com/datasets/tommyngx/inbreast2012 (2012). Accessed: August, 2024.

[26] Liu, Z., Mao, H., et al., "A convnet for the 2020s," in [*Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*], 11976–11986 (2022).

[27] Cheddad, A., "The complete mini-ddsm [data set]." https://www.kaggle.com/datasets/cheddad/miniddsm2 (2021). Accessed: August, 2024.