Synthetic Chest X-ray Image Generator

Whitepaper by

Faustina Selvadeepa, Assistant Manager - Mphasis NEXT Labs | Solution Development
Dr. Revendranath T, Project Manager - Mphasis NEXT Labs | Advisor
Dr. Udayaadithya Avadhanam, Principal & Vice President - Mphasis NEXT Labs | Advisor
Sai Barath Sundar, Senior Manager - Mphasis NEXT Labs | Advisor

Mphasis
The Next Applied

Contents

1. Introduction

The medical industry covers a wide array of fields focused on studying, diagnosing, treating and preventing diseases and conditions. Collaborations among healthcare professionals, doctors, researchers and organizations are required to advance medical knowledge and improve patient outcomes.

Synthetic medical images play a vital role in addressing challenges related to data scarcity, privacy and diversity. The demand for synthetic medical images arises from the limitations in the availability of diverse, well-annotated datasets, especially for training and validating Machine Learning models. Generating synthetic images enables researchers and practitioners to augment existing datasets, ensuring the robustness and generalization of artificial intelligence algorithms used in medical imaging, diagnostics and treatment planning, all while avoiding privacy concerns.

This whitepaper examines the transformative technology of synthetic chest x-ray generation, leveraging artificial intelligence to create realistic medical images for various applications. Recent breakthroughs in synthetic image generation using GANs (Generative Adversarial Networks) have been made, but challenges like mode-collapse still exist. Mode-collapse occurs when the generator model produces a narrow set of images, failing to capture the diversity of input data. Diffusion models, however, prove to be more robust, capable of producing high-quality synthetic images, particularly in the medical domain. The DiffuseVAE architecture, which combines Variational Autoencoders (VAEs) and Denoising Diffusion Probabilistic Models (DDPMs), demonstrates improved synthesis quality on standard image synthesis benchmarks like CIFAR-10 and CelebA-64.

To enhance the model's understanding of the intricate details of chest x-ray images, we propose to extend the DiffuseVAE architecture by incorporating a representation learning model to influence the synthetic image generation process.

2. Problem Statement

Despite the recent advancements in Deep Learning and medical imaging, the availability of large-scale, annotated chest x-ray datasets remains a bottleneck in developing accurate and generalizable Computer Aided Diagnosis (CADx) systems due to two reasons. First, existing datasets are often small, biased towards certain pathologies and limited in diversity, hindering the robustness and scalability of Machine Learning models trained on such datasets. Second, concerns regarding patient privacy and data sharing further exacerbated this challenge, making it difficult to obtain annotated chest x-ray images at scale.

Therefore, a critical need arises for advanced synthetic chest x-ray generation methods that can accurately replicate the intricacies of real patient images while addressing concerns related to data privacy and scarcity. Such techniques must generate high-resolution, diverse and realistic chest x-ray images across a spectrum of pathologies and anatomical variations to facilitate the development of robust CADx systems (capable of detecting and classifying abnormalities, performing lesion segmentation).

The medical images differ from non-medical images in two aspects. First, noise represents critical structures of a disease or condition in medical images, whereas in non-medical images, noise is ignored. Second, pixel intensity values in medical images are directly related to tissue density and other physical properties of the imaged object, whereas in non-medical images, these values may not carry specific quantitative significance. Furthermore, different medical conditions, such as pneumonia and pleural effusion, can sometimes cause similar changes in the appearance of an x-ray.

3. Solution

The proposed solution to address the technical challenges in generating synthetic chest x-ray images is based on the DiffuseVAE architecture. The solution has two modules as shown in Figure 1. First, the image generation module targets at synthesizing high-fidelity images based on the input data. Second, the representation learning module aims to learn the correct representation in the latent space based on input data and tasks at hand.

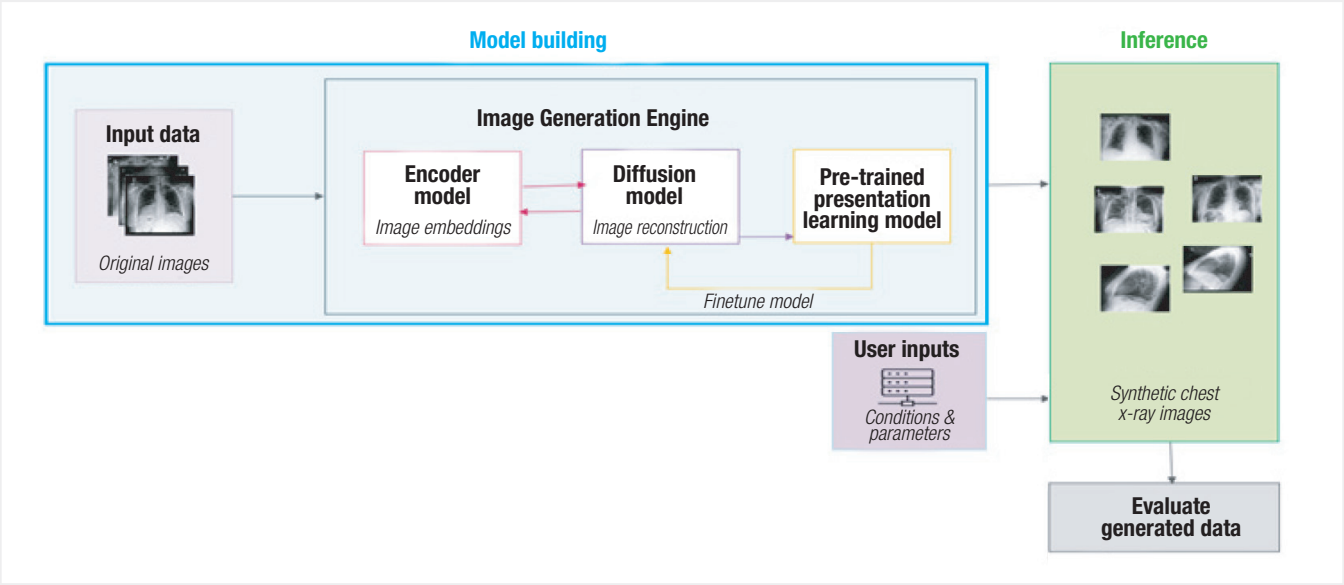


Figure 1: Solution Architecture

3.1. Image Generation Module

The image generation module encompasses a 3-stage process as discussed below:

3.1.1 Encoder Model Training

A VAE is a type of artificial neural network used in unsupervised learning. VAE learns a lower-dimensional, continuous latent space representation of input data, such as images, without requiring labeled data. VAEs consist of two main components: (1) an encoder, maps the input data into a compact latent space and (2) a decoder, reconstructs the original data from the latent space representation. VAEs are trained using variational inference techniques to approximate the posterior distribution of the latent space and therefore learn the ability to generate new data samples by sampling from the learned latent space distribution.

3.1.2. Diffusion Model Training

DDPM is a probabilistic generative model used for image generation and denoising. DDPM is trained using maximum likelihood estimation, where the goal is to maximize the likelihood of observing the clean images given the noisy inputs. DDPM works by iteratively denoising an image with progressively increasing levels of noise and leveraging the concept of Langevin dynamics to model the diffusion process. At each iteration, DDPM performs two steps: (1) estimates the conditional distribution of the clean image given the noisy input and (2) samples the distribution in step (1) to obtain a denoised version of the image.

3.1.3. DiffuseVAE: Combining the VAEs and Diffusion Models

The overall objective of DiffuseVAE is to learn a low-dimensional latent code that captures the underlying structure of the data distribution, while also preserving the high-frequency information in the generated samples. DiffuseVAE combines the VAE and DDPM frameworks by conditioning the DDPM on the latent code inferred by the VAE. This is done in two stages:

Stage 1: VAE Encoding

In the first stage, the VAE encoder takes the original image as input and generates a latent code. This latent code serves as a compact, low-dimensional representation of the input data, preserving essential structural information while filtering out irrelevant or high-frequency noise.

Stage 2: DDPM Reverse Diffusion

The second stage involves using the DDPM to model the training data. The DDPM employs a reverse diffusion process that is conditioned on the VAE-generated latent code. This conditioning enables the DDPM to generate high-quality samples while leveraging the structural information captured by the VAE, ensuring the final outputs retain intricate details.

The design choices and parameterized choices for VAE and DDPM, as detailed in the paper, are carefully selected to optimize the performance of DiffuseVAE.

3.2. Representation Learning Module

The image generation module produces images resembling chest x-rays. However, it is uncertain whether these reconstructed images accurately reflect the data on which the model was trained due to two reasons. First, the model does not incorporate any metadata, such as specific health conditions, which could guide the generation of images for particular types of conditions. Second, the generated images need validation to ensure their alignment with the requirements of healthcare professionals.

To overcome image reconstruction challenges, a self-supervised image representation learning framework based on Momentum Contrastive Learning (MoCo) is used to fine-tune the DiffuseVAE model. MoCo learns data representations by maximizing the similarity between augmented positive samples while minimizing the similarity between positive and negative samples. MoCo utilizes a momentum encoder, an exponentially moving average of the original encoder, to compute representations of negative examples. The momentum encoder significantly contributes to stabilizing the optimization process and enhancing the quality of the learned representation. MoCo uses a contrastive loss function where, given an anchor image, the objective is to maximize the similarity between the anchor and augmented versions (positive examples) of the anchor, while minimizing the similarity between the anchor and other images (negative examples) in the dataset. Negative examples are sampled from a queue of previously seen examples, updated using a First In, First Out (FIFO) strategy with the embeddings of the current batch.

Refer to Figure 2 for an architecture diagram illustrating the key components of the MoCo framework and working functionality in the context of self-supervised representation learning. MoCo is trained using Stochastic Gradient Descent (SGD) with a momentum optimizer. During training, the encoder is updated to minimize the contrastive loss and learns to distinguish between positive and negative examples in a high-dimensional space. Eventually, MoCo learns representation capturing the underlying structure of the data, which can serve as features for downstream tasks.

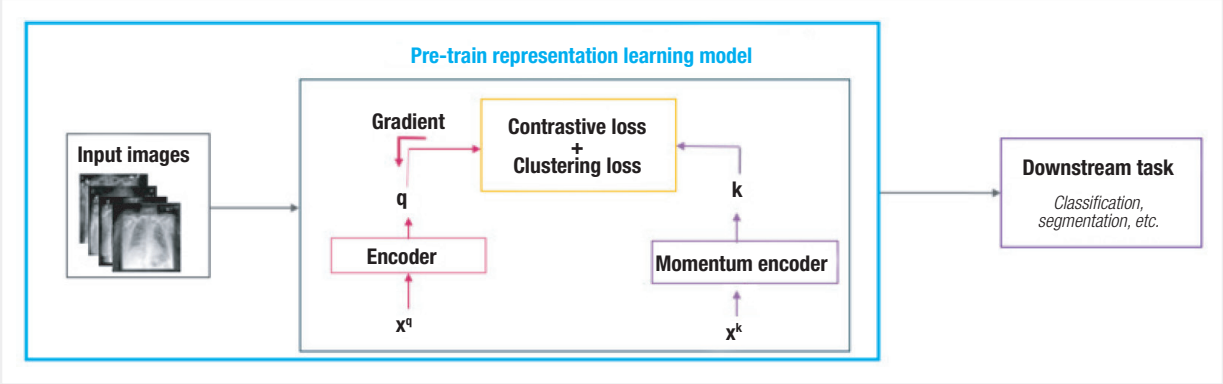


Figure 2: Representation Learning Module

3.3. Representation Learning Model with Auxiliary Loss Function

MoCo learns effective image representations through image augmentation. However, incorporating meta-information about the images to further guide the representations is challenging. To incorporate meta-information about the images, an auxiliary loss function called clustering loss is introduced, which acts as a regularization term to encourage the model to learn representations using meta-information. Specifically, the training dataset is divided into K distinct categories, with each category representing a different distribution of visual patterns captured during image generation (e.g., different diseases or normal vs. abnormal images).

Combining the contrastive loss with the clustering loss encourages MoCo to do the following: (1) learn to distinguish between positive and negative pairs of images (contrastive loss) and (2) group similar images together (clustering loss).

The total loss function is defined as:

$$L = L_{moco} + \lambda L_{KMeansLoss}$$

Where lambda is a hyperparameter that controls the importance of the auxiliary clustering loss.

3.4. Fine-tune Diffusion Model with Representation Learning Model

The modified loss function in the representation learning model encourages learning better representations in the embedding space. During forward diffusion, a noisy image is generated by adding noise to the original image according to the noise schedule. In the reverse process, the reconstructed image is computed by passing the noisy image through the diffusion model. The encoder of the representation learning model computes the embeddings of both the original and reconstructed images. A reconstructed representation loss is computed between the original and reconstructed embeddings. The fine-tuned model loss is defined as:

$$Finetuned\ model\ Loss\ (L) = L_{Diffusion\ loss} + \lambda L_{Representation\ loss}$$

Where lambda is the hyperparameter controlling the weight of the representation learning module loss.

During the image generation process, DDPM utilizes the learned representations to enhance the underlying data distribution, and therefore results in higher-quality generated images. The clustering loss encourages grouping similar images together in the embedding space. Images with similar features (similar patterns indicative of the same disease) are placed closer together, fostering the creation of more coherent clusters. Consequently, the image generation process better captures nuances and variations within each cluster. Leveraging improved representations and coherent image clusters, DDPM's sampling process yields more realistic and high-quality samples sensitive to meta information. DDPM gains a deeper understanding of the underlying data distribution, enabling the generation of images more closely resembling the original data.

4. Evaluation of the Solution

DDPM evaluation uses FID (Fréchet Inception Distance), a metric that quantitatively measures the quality of generated images from generative models. A pre-trained neural network, typically trained for image classification, extracts features from both real and generated images.

An ideal FID score approaches 0, with a lower FID indicating generated images more closely resembling real images. Conversely, a higher FID score suggests that generated images differ more from real images in terms of feature distribution. For medical images, FID scores typically range from 100 to 300, considered too high, which happens because FID struggles to detect subtle differences in medical images. The Inception-V3 model used in FID calculation is trained on the ImageNet dataset, consisting of natural images, such as scenes and animals. When applied to medical images without additional training or fine-tuning, the network may not be well-suited to the unique characteristics and features found in medical images.

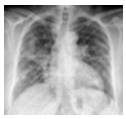
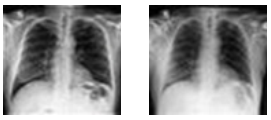
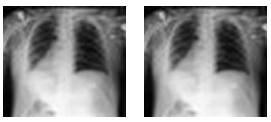
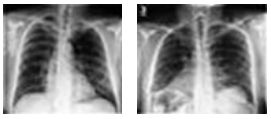
Due to limitations in the FID score for medical images, a custom evaluation metric using MoCo called the **reconstructed representation score** is used to assess the similarity between the original and synthesized images. Similar to the FID score, a lower reconstructed representation score indicates that the generated images are closer to the real images, while a higher score means the generated images diverge more from the real images. The reconstructed representation score typically ranges from 0.03 to 0.045.

5. Dataset

The CheXpert dataset (Irvin et al., 2019), developed by the Stanford ML Group, is utilized for the experiments and model training. CheXpert is a large dataset of chest radiographs that is designed to advance the field of medical image analysis. CheXpert consists of 224,316 chest radiographs from 65,240 patients, annotated for the presence of 14 common chest conditions.

6. Results

The table below summarizes the results of the experiment conducted to evaluate the performance of the model in generating synthetically-generated chest x-ray images for three different health conditions. The table presents the FID scores and reconstructed representation scores for both the model-generated images and the original images.

Health condition	Original image	Generated images	** FID score	*Reconstructed representation score
Pneumonia			418.71	0.03, 0.04
Pleural Effusion			315.16	0.045, 0.038
Healthy Lungs			251.12	0.039, 0.040

&reconstructed representation score is computed for each generated images whereas FID is computed for both the generated images under each health condition.*

7. Impact of Solution

Generating synthetic chest x-ray images offers several advantages:

1) **Cost and Time Efficiency:** Eliminates costly and time-consuming data collection processes, such as obtaining patient consent, conducting imaging studies and managing sensitive medical records.
2) **Accelerated Research:** Provides researchers quick and affordable access to a diverse range of synthetic images, accelerating research activities.
3) **Patient Privacy Preservation:** Preserves patient privacy by generating anonymized images closely resembling real-world x-ray scans, allowing studies without compromising confidentiality or encountering regulatory challenges.
4) **Simulation of Rare Conditions:** Enables simulation of rare or complex conditions difficult to encounter in clinical practice.
5) **Machine Learning Dataset Creation:** Serves as a valuable tool in medical imaging research by creating large datasets for training and testing Machine Learning models without manual annotation.
6) **Customized Image Generation:** Facilitates generating images with specific characteristics, such as different diseases or anomalies, crucial for studying their impact on diagnostic accuracy.

The approach for synthetic image generation described in this white paper is available for consumption on the AWS Marketplace: (AWS Marketplace: Synthetic Chest X-Ray Image Generator (amazon.com)).

8. Conclusion and Future work

Potential opportunities for future research and development to enhance the capabilities of the synthetic chest x-ray image generator are given below:

- **Multi-Disease Image Generation**
  The model now generates images depicting a single condition, i.e., pneumonia, pleural effusion and healthy lungs. Significant advancement would involve expanding the model to generate images with multiple co-occurring diseases or anomalies. Doing so would require training on a more diverse dataset that includes images with various conditions and developing innovative techniques for encoding and decoding such complex scenarios.
- **Prompt-Based Image Generation**
  Integrating the image generator into a larger system that allows users to specify prompts or descriptions for the desired images could prove highly beneficial. For instance, a user could input a textual description, such as "bilateral pneumonia with pleural effusion," and the system could generate a synthetic image that aligns with the description.

In conclusion, the synthetic chest x-ray image generator represents a significant step forward in the field of medical imaging. However, numerous opportunities exist for further enhancement and development of the approaches for synthetic chest x-ray image generation for healthcare.

9. References

- Pandey, K., Mukherjee, A., Rai, P., & Kumar, A. (2022). DiffuseVAE: Efficient, Controllable and High-Fidelity Generation from Low-Dimensional Latents. arXiv preprint arXiv: 2201.00308
- He, K., Fan, H., Wu, Y., Xie, S., & Girshick, R. (2020). Momentum Contrast for Unsupervised Visual Representation Learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 9729-9738).
- Abdusalomov, A., Nasimov,R., Nasimova,N., Muminov, B., & Whangbo, T. (2023). Evaluating Synthetic Medical Images Using Artificial Intelligence with the GAN Algorithm. PMID: 37050503 PMCID: PMC10098960 DOI: 10.3390/s23073440
- (Irvin et al., 2019) arXiv:1901.07031 Jeremy Irvin *, Pranav Rajpurkar *, Michael Ko, Yifan Yu, Silviana Ciurea-Ilcus, Chris Chute, Henrik Marklund, Behzad Haghgoo, Robyn Ball, Katie Shpanskaya, Jayne Seekins, David A. Mong, Safwan S. Halabi, Jesse K. Sandberg, Ricky Jones, David B. Larson, Curtis P. Langlotz, Bhavik N. Patel, Matthew P. Lungren, Andrew Y. Ng

For more information, contact: marketinginfo.m@mphasis.com

USA
Mphasis Corporation
41 Madison Avenue
35th Floor, New York
New York 10010, USA
Tel: +1 (212) 686 6655

UK
Mphasis UK Limited
1 Ropemaker Street, London
EC2Y 9HT, United Kingdom
T : +44 020 7153 1327

INDIA
Mphasis Limited
Bagmane World Technology Center
Marathahalli Ring Road
Doddanakundhi Village, Mahadevapura
Bangalore 560 048, India
Tel.: +91 80 3352 5000

WAS 17/04/25.05 LETTER BASIL 9511

www.mphasis.com