Memory-Efficient Deep Learning for High-Resolution Synthetic Chest X-ray Generation

Shane Woloszyn[1,*]

[1]Kennebunk High School, Kennebunk, 04043, United States
[*]woloszyn.shane@gmail.com

ABSTRACT

In the field of medical imaging, the generation of high-quality images has the potential to have profound implications for diagnostic algorithms, training datasets, and educational tools. This project introduces the use Conditional Generative Adversarial Network (CGAN) as well as a Super Resolution Generative Adversarial Network (SRGAN) to create highly realistic chest X-ray images without concern for memory limitations, making it highly viable for large-scale applications. By using established deep-learning methods, this study expands on previous works and puts the theoretical applications into a practical, deployable model. By leveraging conditional labels, this approach creates new opportunities for enhancing medical diagnostics and facilitating the training of medical students. While this model only analyzes its use for chest X-rays, with simple modifications it has the potential to been generalized across a variety of fields.

Introduction

In recent years, the use of Generative Adversarial Networks (GANs) has emerged as a promising approach in the field of artificial intelligence, particularly for the creation of synthetic, but realistic, images. These networks have shown great potential ranging from artistic content creation to medical diagnostic applications. This paper explores the development and application of a CGAN in connection with an SRGAN, specifically for the generation of chest X-ray images with different pathologies. Leveraging the power of deep learning, the project aims to create a tool capable of producing highly realistic synthetic medical images based on conditional inputs, to assist in medical training and diagnostics. Through this tool, we seek to demonstrate how GANs could be fine-tuned for specific medical uses. The outcomes of this work have the potential to make a significant impact on healthcare training, especially in environments with limited access to diverse imagery.

Related Work

GANs have emerged as a powerful tool for their ability to create high-quality images based on a dataset. However, their use for creating actual images hasn't been utilized heavily in the medical field, with the majority of subject matters being art, fashion, or entertainment. Earlier work in the field, such as **Radford et. al. (2015)** with DCGANs, has demonstrated the power of GANs for creating realistic images from random noise, showing the potential for more advanced applications[1].

The application of medical image synthesis was greatly advanced with the introduction of CGANs, which allowed for the generation of images based on specific class labels. The work of **Odena et. al. (2016)** specifically with AC-GANs (Auxillary Classifier GAN) was a significant contribution to this domain, paving the way for more targeted and specific image generation tasks which were especially valuable for use in the medical field[2]. **Skandarani et. al. (2020)** conducted an empirical study testing GAN models for medical image synthesis. In this study, they tested a variety of GANs, but just analyzed the successes and failures of each model, rather than using the models for a specific purpose[3]. While this work was influential, it was more exploratory, and at the end, they called for research to be done regarding GANs that can account for different subtleties.

Building on these advancements, **Bluthgen et al. (2024)** introduced an approach for generating high-resolution synthetic medical images using a diffusion model with adaptive noise sampling[4] This technique provided a significant improvement in generating photorealistic medical images with fine-grained details. Their findings demonstrated that adaptive noise control could better capture nuanced features in medical imaging datasets, offering an alternative pathway for improving image synthesis techniques.

Baur et. al. (2018) focused on the generation of realistic images of skin lesions using GANs[5]. This study demonstrated that GANs were capable of creating highly realistic skin lesion images, that can be used for training models or enhancing datasets. While this work was significant, it focused on a relatively simple domain and didn't involve any use of conditional labels.

In contrast, our work advances the application of GANs in the medical field by addressing complex challenges unexplored by other studies. Our approach introduces the use of conditional labels to generate more complex and diverse medical images.

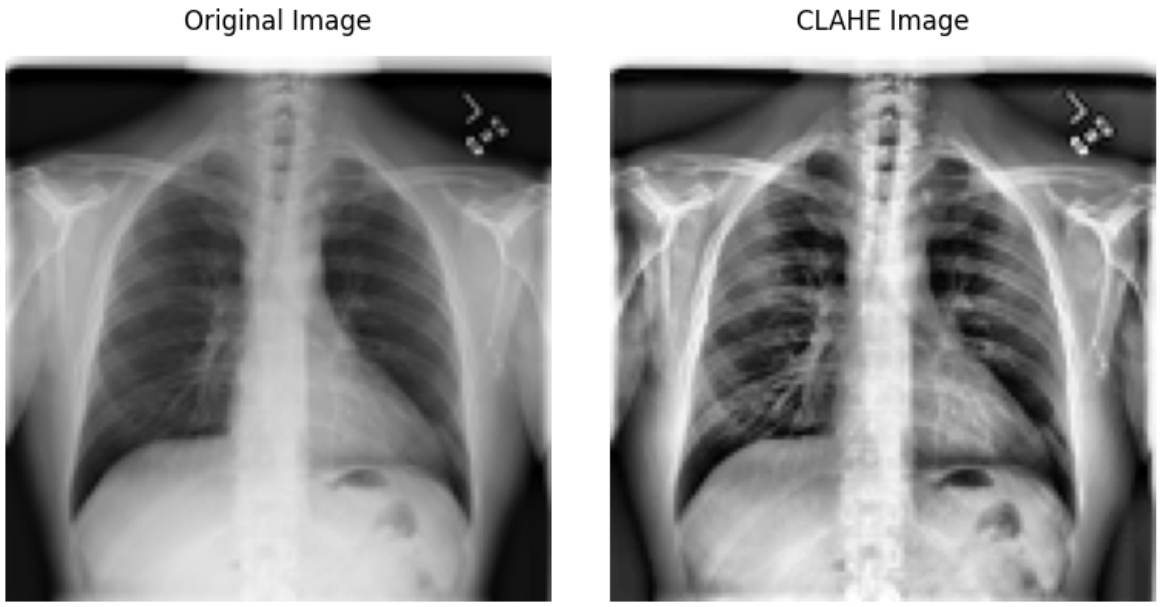

Figure 1. Comparison of the original and preprocessed image. The image on the left represents the original input from the dataset, while image on the right shows the improved contrast and feature clarity as an effect of CLAHE preprocessing.

This enables the synthesis of medical images that aren't only realistic but also highly relevant and applicable to a much wider range of tasks. Notably, our methodology achieves this without requiring massive computational resources, making it one of the few approaches capable of producing high-quality images efficiently.

Data

Our data was comprised of grayscale chest X-rays from the ChestX-ray8 dataset[6]. The images in this dataset were sourced from the National Institute of Health, and labeled for one of 14 pathologies (or "No Findings"). Although the images were diagnosed using Computer Aided Detection, the estimated accuracy for the diagnosis is >90%. Of the full dataset, we used 5,000 images diagnosed "No Findings", 5,000 diagnosed "Mass", and 5,000 diagnosed "Pneumonia". Each pathology class was then assigned a numerical label, with 0 being "No Finding", 1 being "Mass", and 2 being "Pneumonia". These labels were used as conditional labels to be interpreted by the CGAN.

The images were then resized from 1024x1024 resolution to a set of 512x512 images, and a set of 128x128 due to computing limitations, while maintaining clarity needed for model training. To further enhance the quality and detail within each image, the images were pre-processed using Contrast Limited Adaptive Histogram Equalization (CLAHE). This process boosts local contrast, making subtle variations more prominent and aiding the models in detecting characteristics for each pathology. The improvements in clarity after applying CLAHE can be seen in *Figure 1*.

Methods

This study implemented a Conditional Generative Adversarial Network (CGAN) using PyTorch, to generate synthetic 128x128 gray scale chest X-ray images based on three labeled pathologies (No Finding, Mass, Pneumonia). The CGAN architecture expands on the basic GAN framework by introducing a label input, allowing the model to control the pathology types for images generated. The GAN structure contains two components: a generator and a discriminator, engaged in an adversarial game.

The generator receives noise tensors, sampled from a Gaussian distribution, with assigned conditional labels to represent the different pathologies. These tensors are then put through multiple convolutional networks to reach the target resolution, through layers of up-sampling then activation. For our activation we chose LeakyReLu within both the generator and discriminator, to prevent "dead neurons." Dead neurons arrive when neurons consistently output zero across multiple training iterations, which is usually caused by standard ReLu functions due to their output of 0 for negative inputs. To prevent this, LeakyReLu applies a

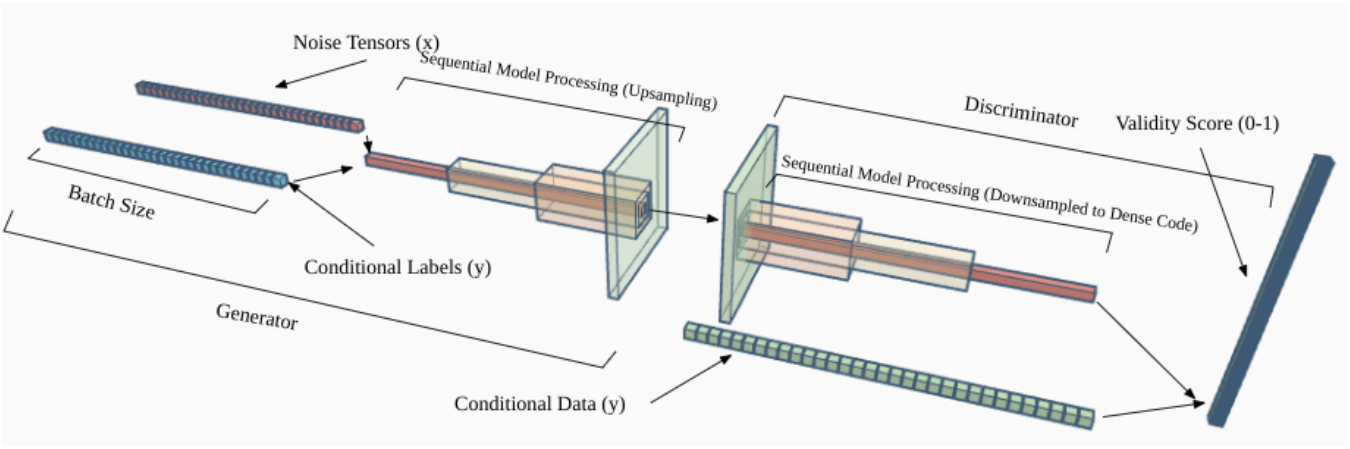

Figure 2. Basic structure of the Conditional GAN model. This diagram shows the generator and discriminator with their respective input sources and processing layers.

small slope α to negative values, allowing gradients even with negative input. This can be represented mathematically below:

$$f(x) = \begin{cases} x & x > 0 \\ \alpha x & x \leq 0 \end{cases} \tag{1}$$

Here $x > 0$ represents the positive input region, where the gradients are unchanged, where $x \leq 0$ represents the negative input region, where the gradients are modified slightly by α rather than the typical 0 by normal ReLu.

The discriminator, in contrast, serves as a binary classifier with conditional input, distinguishing between real images as 1, and fake images as 0, given the conditional label. The discriminator goes through layers of down-sampling, activation (LeakyReLu), and dropout (to prevent mode collapse). Mode collapse occurs when the discriminator becomes too strong, and the generator begins to focus on a certain set of patterns, preventing diversity and ultimately hurting the learning. Dropout prevents this by dropping units with a certain probability, weakening the discriminator. The overall structure of the Conditional GAN model is depicted in *Figure 2*. This diagram outlines the flow of conditional inputs and noise tensors, and how they are processed through each layer in the generator and discriminator.

The theoretical foundation of GANs is rooted in game theory, with the conditional additional only slightly modifying the process. This process can be modeled as a minimax operation, where the generator and discriminator aim to minimize their own objective functions in an adversarial game. The generator's objective function (or generator loss) can be defined as:

$$\mathscr{L}_g = -\mathbb{E}_{z\sim p_z, y\sim p_y} \log[D(G(z \mid y))] \tag{2}$$

where $D(G(z \mid y))$ represents the discriminator's confidence that the generated image $G(z \mid y)$ being judged as real, given the conditional label y. Using this loss, the generator can adjust its parameters to maximize the probability that the images are judged as real, making the generated images begin to appear more real and like the real images.

The discriminator's objective function (or discriminator loss) can be defined as:

$$\mathscr{L}_d = -\mathbb{E}_{x\sim p_x(x)} \log D(x \mid y) - \mathbb{E}_{z\sim p_z, y\sim p_y}[\log(1 - D(G(z \mid y)))] \tag{3}$$

where $D(x \mid y)$ represents the discriminators confidence that real image x is real given conditional label y, and $1 - D(G(z \mid y))$ represents the discriminators confidence that the generated image $G(z \mid y)$, conditioned on label y. Using this loss the discriminator can adjust its parameters to better judge real images as real, and fake images as fake, making it harder for the generator to fool the discriminator effectively making the generator need to produce more "real looking" images.

Having defined the conditional loss functions for both our generator and our discriminator, we can formalize this objective into one adversarial function. The discriminator D aims to maximize its accuracy in distinguishing real images as real and fake images as fake, maximizing $-\mathscr{L}_d$. The generator aims to fool the discriminator, producing synthetic images that are distinguished as real, effectively maximizing $\mathscr{L}_d$, while minimizing $\mathscr{L}_g$. This adversarial objective of G and D allows us to derive a conditional adversarial objective function, which can be written as a minimax game:

$$\min_G \max_D \mathbb{E}_{x,y\sim p_{\text{data}}} [\log D(x \mid y)] + \mathbb{E}_{z\sim p_z, y\sim p_y} [\log(1 - D(G(z \mid y)))] \tag{4}$$

This function correctly shows the competition between G and D, with D wanting to maximize the likelihood of correctly classifying images and G minimizing the likelihood of producing distinguishable fake images.

Training was carried out with an NVIDIA GeForce GTX 1050Ti with it taking around 12 hours to finish training ($\sim$500 epochs). After completing training, the discriminator and generator can be saved as individual models, without the need for any other structure. While the generator can create images alone, by passing generated images through the discriminator again with a "real threshold", we are able to further ensure the images are accurate, as they have to be marked real by the discriminator again.

With our CGAN trained and able to create 128x128 images, we created an SRGAN to up-sample these images to a higher resolution. This approach minimizes the computing capabilities necessary for training, as the models are able to be trained separately, splitting the memory needed.

Similar to the CGAN, the SRGAN is trained with a generator and discriminator engaged in an adversarial game. To train the SRGAN, we curated a dataset of paired low- and high-resolution images (still using the ChestX-ray8 dataset). The low-resolution images were generated by down-sampling the high-resolution images, ensuring that the model learn to reconstruct missing details.

The generator was then given a low-quality image and tasked with up-sampling it to a higher resolution image. The discriminator then distinguishes between the generated high-resolution images and the real high-resolution images, returning a probability score that guides the generator in improving the realism of its outputs.

In summary, this study implemented a CGAN to generate pathology-specific synthetic chest X-ray images and an SRGAN to enhance their resolution. By training these models separately, we were able to optimize computational efficiency while maintaining high-quality outputs. The trained models provide a framework for generating high resolution pathology-labeled chest X-rays, which can be utilized for a variety of applications.

Results

The combined CGAN and SRGAN model successfully generated synthetic chest X-ray images conditioned on 3 pathology labels: *No Findings, Mass, and Pneumonia*. To visually estimate the performance of the model, we generated one synthetic image per pathology and arranged them in a 2x3 grid (see *Figure 3* below). Each image was generated based on the discriminator-assessed "real threshold" of 0.9, indicating high confidence in the generated images' authenticity. This generation process took less than a second to complete. As a comparison, randomly selected real images for each pathology were placed below the generated in *Figure 3*

Discussion

The CGAN-SRGAN model developed in this study presents a novel approach to generating high-resolution, condition-specific synthetic chest X-ray images. By initially training a CGAN to create 128×128 images and subsequently upscaling them with an SRGAN, this method significantly reduces computational demands while maintaining image realism. This approach enables the creation of diverse, labeled datasets, which can be instrumental in training and validating diagnostic algorithms. The ability to generate pathology-specific images enhances the diversity of training data, potentially improving the robustness of deep learning models in real-world medical applications. Furthermore, since these images are entirely synthetic, they alleviate concerns regarding patient privacy and data-sharing restrictions.

Beyond diagnostic model development, this synthetic imaging framework holds promise for medical education. By generating an unlimited number of high-resolution images tailored to specific pathologies, medical students and radiologists-in-training can gain exposure to a wider array of cases, including rare conditions that may be underrepresented in traditional datasets. This ability to curate customized learning materials enhances the accessibility and effectiveness of medical training.

Despite these advantages, certain limitations must be acknowledged. The CGAN-generated images, while visually convincing, may lack subtle anatomical details crucial for clinical interpretation. The reliance on the initial dataset also introduces the risk of propagating existing biases or artifacts present in real X-ray collections. Additionally, while the SRGAN enhances resolution, it does not inherently add new diagnostic information—it simply refines features inferred from the lower-resolution input. Therefore, while these synthetic images may be useful for training and research purposes, caution should be exercised before considering them for direct clinical application.

Future work should explore strategies to further enhance anatomical accuracy and diagnostic utility. This could include training on 3D volumetric data, incorporating additional patient data such as age and sex to improve contextual accuracy, or integrating domain-specific constraints to enforce anatomical consistency. Furthermore, expanding this methodology to other imaging modalities, such as MRI or CT scans, could further demonstrate the general value of this approach in medical imaging.

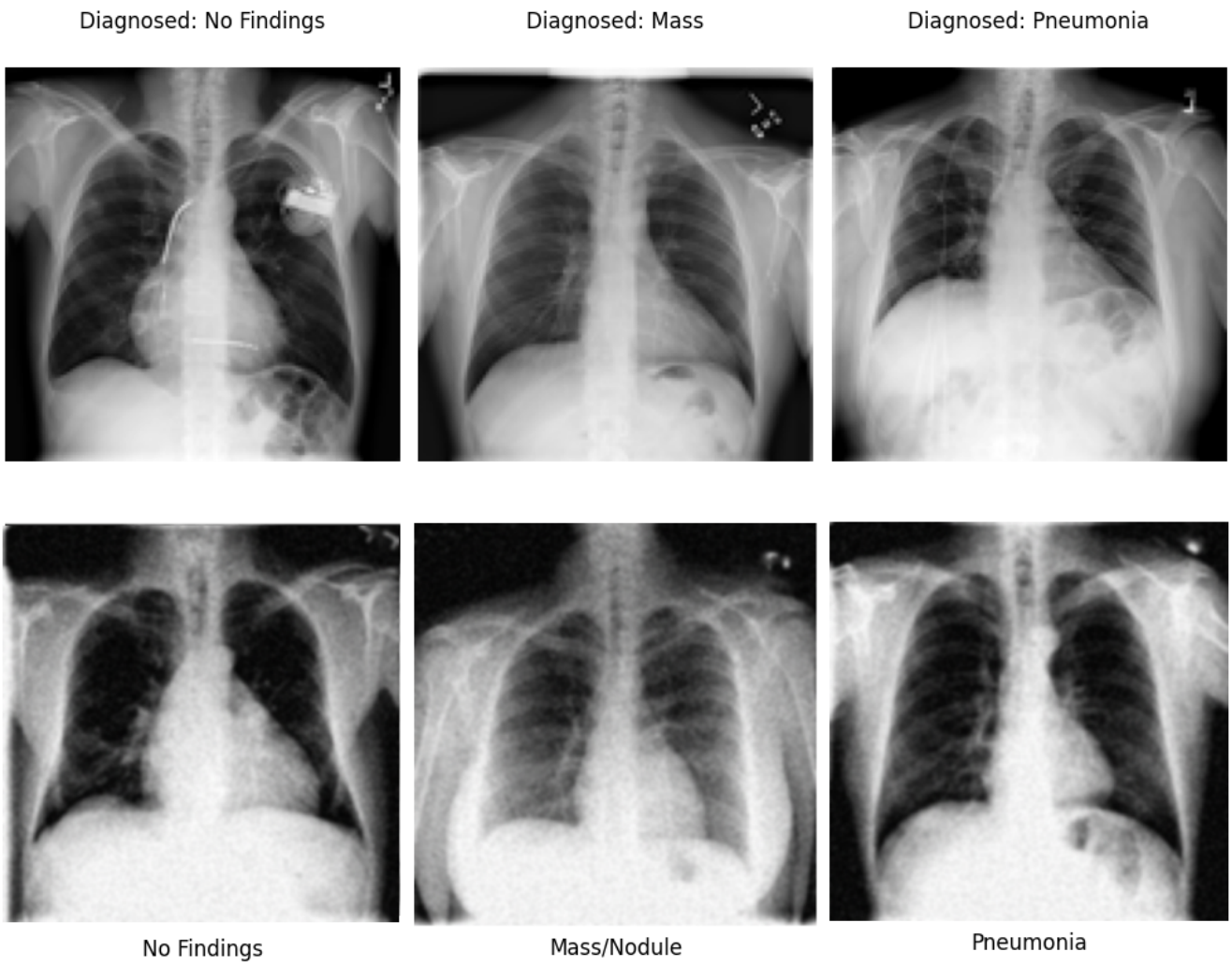

Figure 3. Comparison of real X-rays (top) with synthetic X-rays (bottom).

Conclusion

This study presents a CGAN/SRGAN model that generates realistic, condition-specific synthetic chest X-rays, underscoring its potential to assist in diagnostic and educational imaging in the medical field. While challenges remain in optimizing image quality, and avoiding artifacts, this work provides a foundation for future innovations in medical synthetic imaging.

Acknowledgements

I would like to thank Dr. Stuart Damon of University of New England for his inital discussions that helped inspire this work, and Dr. Sylvain Jaume of University of New England for the mentorship and valuable discussions throughout the research process.

References

1. Radford, A., Metz, L. & Chintala, S. Unsupervised representation learning with deep convolutional generative adversarial networks. (2016). 1511.06434.
2. Odena, A., Olah, C. & Shlens, J. Conditional image synthesis with auxiliary classifier gans. (2017). 1610.09585.
3. Skandarani, Y., Jodoin, P.-M. & Lalande, A. Gans for medical image synthesis: An empirical study. (2021). 2105.05318.
4. Bluethgen, C. *et al.* A vision–language foundation model for the generation of realistic chest X-ray images. *Nat. Biomed. Eng.* 1–13, DOI: 10.1038/s41551-024-01246-y (2024).
5. Baur, C., Albarqouni, S. & Nava, N. Generating highly realistic images of skin lesions with gans. *Int. J. Comput. Vis.* (2018).
6. Wang, X. *et al.* ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases (2017).