Transfer-PGGAN: High-Resolution Block Fine-Tuning of Pretrained PGGANs for Domain-Adapted Chest X-Ray Synthesis

Will Li **Finn Guo** **Hengfan Zhang**
Department of Electrical and Computer Engineering, Duke University
{will.li, xiaofeng.guo, hengfan.zhang}@duke.edu

Abstract

Medical image synthesis using generative adversarial networks (GANs) has the potential to alleviate scarcity of labeled data to promote supervised learning tasks. However, GAN training is computationally expensive and prone to image domain shift bottlenecks. In this project, we proposed a GAN transfer learning pipeline to finetune pretrained Progressive Growing GAN (PGGAN) to synthesize domain-adapted, realistic chest X-ray images. Leveraging the pretrained generator and DenseNet-based discriminator based on the ChestX-ray14 dataset, we finetune the higher-resolution ScaleBlocks of PGGAN to adapt to the novel CheXpert dataset. Through visual inspection and quantitative assessment with Structural Similarity Index Measure (SSIM), we showed that our finetuned generator outputs images more closely aligned with real-world radiographs from target dataset compared to those produced by the pretrained generator. The proposed pipeline provides a reproducible, lightweight setup to transfer existing GAN models across domain-shifted datasets, bridging the gap for more effective, resource-efficient augmentation strategies to supply high-fidelity medical image datasets.

1 Introduction

Deep learning models have demonstrated exceptional performance across various medical imaging tasks, including pathology detection, segmentation, and classification. However, the success of deep learning models is constrained by the limited availability of high-quality annotated datasets, particularly in the medical domain where data is expensive to collect and sensitive to share. In contrast to natural image datasets, traditional data augmentation techniques such as random flipping or color jittering may not be applicable in medical settings, where small geometric or color distortions can lead to loss of clinically significant features for diagnosis.

To address this challenge, medical image synthesis using generative models, especially generative adversarial networks (GANs), has emerged as a promising approach to populate synthetic image samples with high-fidelity. Yet, training full-scale GAN from scratch remains computationally expensive and vulnerable to domain-shift bottlenecks when the training and target image distributions differ due to variances in imaging equipment , acquisition protocol, or patient population. In this project, we propose a transfer learning pipeline that employs a pretrained Progressive Growing GAN (PGGAN) models[1] and finetune its higher-resolution scale blocks to achieve adaptations to generate images similar to our target chest X-ray image dataset. Specifically, we finetuned a PGGAN generator pretrained on the ChestX-ray14[2] dataset using a curated subset from CheXpert dataset[3] by freezing the early and intermediate resolution layers and fine-tuning the last two ScaleBlocks via an adversarial training process. With the hierarchical image upsampling architecture of PGGAN models, we hypothesize that we can preserving the model's ability to encode radiographic and anatomical features in the lower level layers while allowing the model to capture fine-grained color and textural details

36th Conference on Neural Information Processing Systems (NeurIPS 2022).

inherent in the new distribution through fine-tuning the high-resolution layers only. For adversarial supervision, we used a pretrained DenseNet-121 model[4] as discriminator backbone and added a learnable binary classification head to provide reliable contrastive feedback under the WGAN-GP loss[5]. Quality of our image synthesis and adaptation are measured through visual inspections and qualitative metrics such as Structural Similarity Index Measure (SSIM). We found that our finetuned generator produces images that align more closely in both visual appearance and SSIM scores with the real image data in the CheXpert dataset than those generated by the original ChestX-ray14-trained PGGAN. The code for our work are available at https://github.com/liwilliam127/transfer_PGGAN.

2 Related Work

Recent advances in generative models have dramatically changed the landscape of medical image analysis. In particular, methods using generative adversarial networks (GANs) have emerged as a powerful tool for synthesizing high-fidelity radiological images for data augmentation. However, challenges such as training cost and preservation of domain-specific structures still hinder the effectiveness of these models in clinical settings. To illustrate our work, we review related research in GAN-based image synthesis methods for medical imaging and explore practical strategies of transfer learning and fine-tuning to address these challenges.

2.1 Generative Adversarial Networks for Medical Imaging

Generative Adversarial Network is a framework for training implicit generative model to reproduce images following a latent training distribution through adversarial optimization[6]. GANs have been widely used for image synthesis in medical imaging studies[7]. For example, previous studies have applied GANs to generate chest X-rays images for downstream disease classification tasks[8]. Many of these GAN-based methods fails to address training instability issues in high-resolution medical image generation. Therefore, another work by Segal et al. investigated using Progressive Growing GANs (PGGANs) [1] to stably generate images with hierarchical resolutions upsacling, making them particularly suitable for medical image synthesis. However, PGGAN generators remain computationally expensive to train due to its multi-resolution image rendering, especially as given limited data and computational power in conventional medical settings.

2.2 Transfer Learning and Partial Finetune in Medical Domains

Transfer learning has emerged as a effective technique in deep learning to enable pretrained models on large-scale source domains to be adapted to new tasks and domains with limited data and computational resource[9]. The method is critical in medical imaging due to the scarcity of annotations especially as for medical images, differences between institutions, imaging hardware, or patient population can introduce domain shift and reduce the generalization ability of the model[10]. The work of Raghu et al. [11] highlights that features from early layers in pre-trained models generalize well across a variety of medical tasks. They show that in low-data settings, fine-tuning only task-specific layers can match or even exceed the performance of the full model. Therefore, our approach focuses on adjusting the pretrained PGGAN generator through higher-level layer fine-tuning to account for such distribution shift.

3 Dataset and Preprocessing

We conduct all experiments on a curated subset of the Stanford CheXbert dataset [3], a widely-used collection of frontal and lateral-view chest radiographs with radiologist-labeled pathologies for supervised learning tasks (available with license at https://github.com/jfhealthcare/Chexpert). The original dataset contains 224k high-resolution (around 3000×3000 pixels) over 14 pathology labels, in we selected a subset of 73,752 frontal-view images of three representative chest X-ray diagnosable pathologies, including cardiomegaly, pneumothorax, and pleural effusion. Given the limited computational budget and grayscale nature of the chest X-ray radiographs, we design a task-specific preprocessing pipeline to address modality mismatch and dimensional alignment.

3.1 Image Resize and Grayscale Transformation

The original CheXpert images are visually grayscale radiographs, but the images are inherently provided in RGB format with three channels, which can be computationally expensive to train given our GPU constraints. Additionally, since we adapt the pretrained DenseNet121 model from TorchXRayVision library[4] as the discriminator for our adversarial training process, which requires single-channel grayscale image input, we also convert the training set images to single-channel grayscale using the Luminance coefficients. Since the generator internally operates on RGB outputs via 3-channel to_RGB layers, we retain the RGB pipeline for image generation and apply the same grayscale transformation on the fly prior to feeding into the discriminator. The conversion allows us to preserve most of model's capability to generate high-fidelity, radiographic-style images under our resource constraints.

3.2 Spatial Preprocessing and Resolution Matching

The raw chest X-ray images in CheXpert are high-resolution and vary in aspect ratio. For standardization and resource efficiency during the training, we resized input images to 224×224 using bilinear interpolation-based resize method from TorchVision library[12] as a standard resolution widely adopted in previous GAN training [13]. The downsampling also allows for alignment with the expected discriminator input resolution. During the training, while the generator dimension are inherently 1024×1024, we applied the same transformation to downsample images to 224×224 prior to feeding them into the adversarial loop. This step ensures consistency and compatibility between modules while allowing for training on limited GPU resources.

4 Methodology

Our method builds upon fine-tuning a pretrained 1024×1024 Progressive Growing GAN (PGGAN) model in a generative adversarial training pipeline illustrated in Figure 1. The generator was initialized with the publicly available PGGAN weights released by Segal et al [14], which was trained based on 800k images from the NIH ChestX-ray14 dataset[2] with model and license are available at: https://github.com/BradSegal/CXR_PGGAN. A DenseNet-121 from torchxrayvision pretrained on the CheXpert dataset serves as the backbone for the discriminator in our GAN framework. All convolutional layers of the discriminator remain frozen to preserve the model's ability to extract radiographic image features. A pooling layer is applied to the features extracted, and then we adapt the generator with a two-layered binary MLP head for binary classification, which will serve to output real/fake logit in the adversarial process.

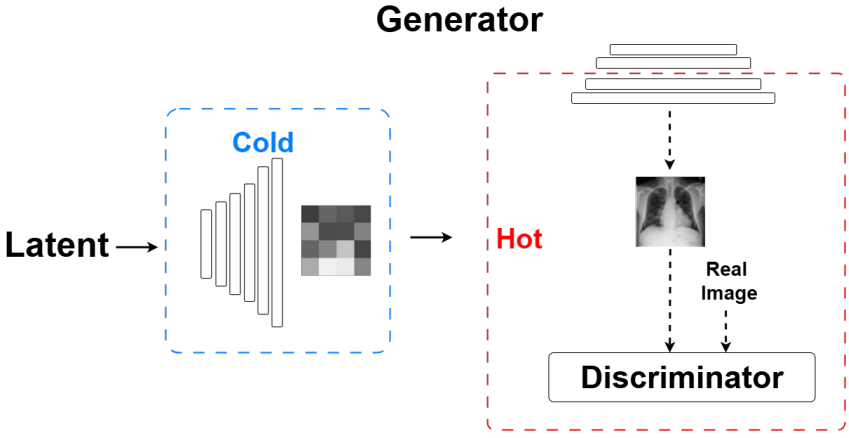

Figure 1: Training Pipeline Overview.

4.1 Domain Adaptation via Partial Generator Tuning

Progresive Growing GAN, as shown in Supplementary Material Figure 3, is a hierarchical GAN model architecture that generates images through its sequence of "ScaleBlocks" that upscale the

image resolution in each stage ($4 \times 4 \rightarrow 8 \times 8 \rightarrow \ldots \rightarrow 1024 \times 1024$). A 1×1 to_RGB layer is applied after each Scaling layer for channel reduction, with the last few to_RGB layers responsible for the final rendering of the images to our visually perceivable RGB images. Therefore, we propose a partial fine-tuning of the model by freezing the earlier ScaleBlocks (1-5) and their To_RGB layers while training the final ScaleBlock 6-7 (which upsamples from $256 \times 256 \rightarrow 1024 \times 1024$) and corresponding to_RGB layers. The method allows for preserving model's early layers that have learned to encode broad radiographic and anatomical features. By fine-tuning the highest resolution components, we can train these ScaleBlocks and RGB_rendering layers to capture domain-specific grayscale tonality and structural details of the target image distribution. By reducing the parameter update through freezing early-layers, we also expect lower memory use and training wall-time, which reinforces our idea to propose a lightweight transfer learning pipeline. In related studies, researchers also demonstrated that similar layer-selective finetune strategies can help to effectively stabilize GAN transfer[15] and prevent model collapses.

4.2 Training Overview and Objectives

We will train the model using a standard adversarial loss objective, where the DenseNet-based discriminator aims to distinguish between real and synthetic images and the generator learns to produce samples to fool the discriminator. At every training iteration, a latent embedding vector $z \in$ (batch_size, 512) is sampled from $\mathcal{N}(0, I)$ and forwarded to the PGGAN generator to generate three-channel, 1024×1024 images. These images will undergo the same resize transformations and grayscale conversion as discussed in Section 3.2 to match the resolution and modality of preprocessed real image. Before feeding to the discriminator, we normalize both real and synthetic image with tanh to ensure that both grayscale inputs are mapped to the identical range of [-1, 1]. Then, the results are forwarded to the discriminator, which will produce a real/fake logit for loss training. The pipeline ensures that real and synthetic images will share resolution, modality, and normalization range, which will allow the discriminator's gradients to target image semantics rather than trivial cues for differentiation.

We employ Wasserstein GAN with gradient penalty (WGAN-GP) loss as our training objective and follow Gulrajani et al.[5] to perform five critic (discriminator) updates per every generator update (n_critic = 5) in order to ensure stable Wasserstein distance approximation. However, after certain number of epochs, we noticed a tendency of discriminator quickly overpowering the generator as we observe continuously descending d_loss versus diverging g_loss overtime. To address the imbalance, we implemented different forms of 'regularization' to constrain the learning of the discriminator, including reducing the n_critic to 2, lowering discriminator learning rate, and increasing the lambda for gradient penalty. With these adjustments, we were able to retain the downward and stabilizing generator loss trend and observe steady improvement in output image quality over time.

5 Experiments

5.1 Training Setup

To train the partially finetuned generator and the adapted discriminator, we used stochastic gradient descent with Adam optimizer with $\beta_1 = 0.5$ and $\beta_2 = 0.9$. We used the Wasserstein GAN loss with gradient penalty (WGAN-GP) as the objective function with $\lambda = 20$. The learning rate for discriminator is set to 2e-5, and that of generator is set to 1e-4 to mitigate the overpower of discriminator over the generator. For the same reason, we used n_critic = 2 after tunings. Experiments were conducted on one an NVIDIA A100 (40GB) GPU, and 30 epochs of training were completed with batch size of 32.

5.2 Qualitative Assessment

For qualitative assessment of the generated images of our finetuned PGGAN generator, we randomly sampled 50 synthetic images and manually inspected their quality and visual features, and two images from each group are randomly selected for demo Figure 2. Comparing the synthetic image of our finetuned generator with real CheXpert X-ray images and those generated by the pretrained PGGAN, we were able to identify multiple consistent differences. First, the generated images with

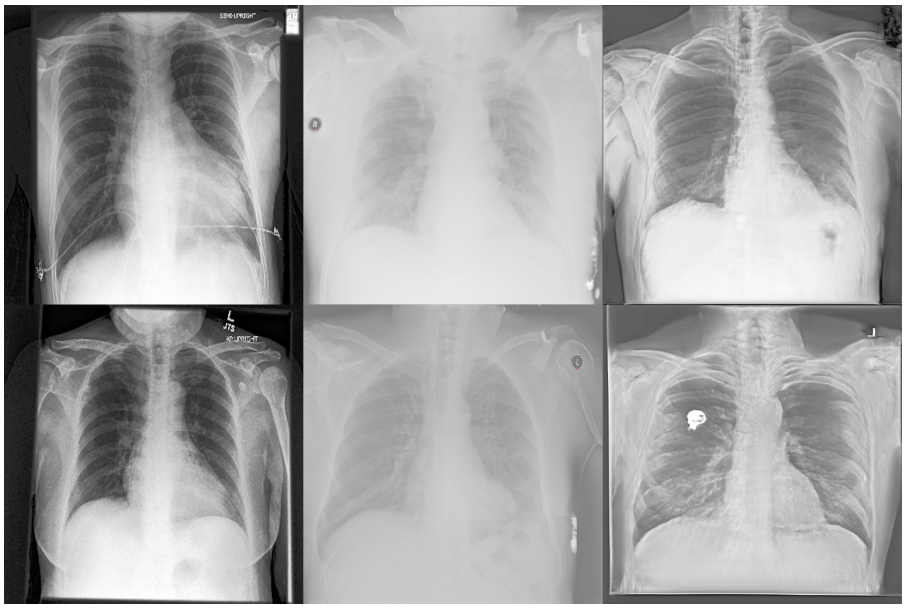

Figure 2: Visual comparison between real CheXpert X-rays (left), pretrained generator output (middle) and finetuned generator (right).

fine-tuning are more consistent with the CheXpert images than synthetic images without finetune do in terms of contrast and grayscale distribution. In the pretrained generator images, soft tissue regions, such as the cardiac silhouette, mediastinum, and diaphragm, tend to exhibit more overlapped grayscale intensity as air regions (e.g. lung areas). After fine-tuning, the generator produces images with consistently darker air regions and less intense soft tissues, more closely resembling the tonal characteristics shown in the real CheXpert images. Overall, the images exhibit better alignment of their luminance distribution and dynamic range of grayscale intensity to target dataset images as a sign of successful adaptation through fine-tuning.

Second, the finetuned generator outputs exhibited sharper anatomical and structural separation. For example, by visual inspections, we found that on average, the finetuned generator produce images with more distinct lung region contours from surrounding area and crisper edges especially around the ribs and clavicles. Comparing to the pretrained generator outputs, the units of anatomical structures are more salient and approximate to those in real CheXpert dataset. This observation aligns with our expectation regarding the different dataset provenance and preprocessing pipelines of CheXpert[3] and Chest X-ray14[2]dataset, including the original resolution differences (3000×3000 vs. 1024×1024) and preprocess compression (near-lossless vs. contrast-normalized).

Lastly, despite the improvements from the original generator, we observed that the synthetic images still exhibited mildly rtanheduced high-frequency textures compared to the the real CheXpert images, such as blurred small vascular branches and nosie artifact at pixel level. It indicates a residual gap in our generator's ability to fully replicate fine-grained radiographic detail in real world chest X-rays. However, we also noted that, comparing the generated images from 20 to 30 epochs, the richness of these high-frequency features shows gradually improvements over time, consistent with the training dynamic discussed in related GAN literature.[1]. Therefore, we expect to reduce this gap in our future work given extended training time and computational resources.

5.3 Quantitative Evaluation

Furthermore, we evaluate our finetuned generator by comparing the quantitative metric performance of our finetuned generator image versus the pretrained generator images. Due to the grayscale conversion as part of our image preprocessing, we rendered our synthetic images using single-channel grayscale for best visual alignment with real X-ray images. Therefore, to achieve stable and less

biased comparisons, we avoided using Fréchet Inception Distance, which was based on Inception-v3 pretrained on RGB ImageNet. Alternatively, we selected samples from CheXpert dataset, original generator images, and our finetuned generator images and compared using pair-wise Structural Similarity Index (SSIM), a standard metric for similarity between images. SSIM, as shown in Equation 1, is derived as the product of their luminance, contrast, and structural ratios based on mean, variance, and covariances of each image.

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{1}$$

As the result, the original generator images achieves an average SSIM of 0.3739 to the CheXpert images, while the finetuned generator improves this score to 0.5925, indicating a significant increase in structural similarity between our synthetic images and real X-rays. This numerical improvement is consistent with our visual observation and confirms the effectiveness of the fine-tuning process. While the current SSIM score still leaves room for improvement, we expect that in future work, with access to more GPU resources to handle higher-resolution images and longer training cycles, the generator could potentially produce more realistic results to fully demonstrate its capability for domain adaptation.

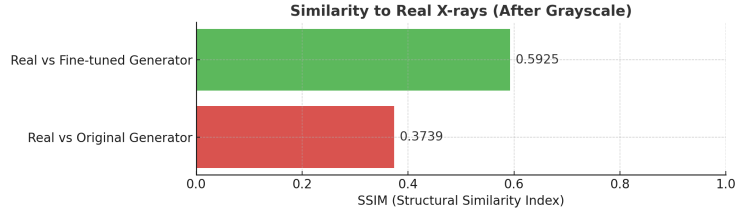

Figure 3: Visual comparison between real chest X-rays and generated outputs.

6 Conclusion

In this project, we proposed a lightweight pipeline to achieve domain-adaptation of pretrained generative models (PGGAN) for medical image synthesis. By freezing early ScaleBlock layers of the generator and fine-tuning only the high-resolution and color rendering blocks, our approach exhibited effective domain adaptation with minimal computation without large-scale retraining. We demonstrate that with our partial fine-tuning strategy, we were able preserve the pretrain model's learning of radiographic image generation and also improves the grayscale and structural similarities between the new synthetic images and real X-ray images of target domains. For evaluations, we measured the improvement in synthetic image similarity to real chest radiographs through visual inspections and quantitative SSIM. Our results highlight the potential of this efficient, lightweight pipeline to achieve generative model adaptation in clinical settings where image data augmentation are in demand and computational power are limited.

6.1 Limitations and Future Work

Our approach demonstrates effective domain adaptation through partial fine-tuning of the pretrained generators. Due to the time and computational resource limitation of the project, we only conducted experiments with two final ScaleBlock layers unfreezed, which may limit the extent of domain adaptation and quality of image generation. Similarly, more pretrained generative models and medical image datasets should be evaluated to more comprehensively demonstrate the robustness of our method in addition to the feasibility. Lastly, our evaluation focused primarily on visual inspections and SSIM-based similarity. Furthermore, additional downstream tasks with clinical-relevant objectives, such as classification or detection tasks using datasets augmented with our generator should be assessed to demonstrate the practical utility of the method. In the next step, we will generalize the method using more diverse generative models (e.g. StyleGAN or diffusion models) and medical image modalities (e.g. CT scans). Also, we will aim to conduct downstream task evaluations to further assess the performance of our synthesized chest X-ray images in supervised learning tasks such as lung segmentation and pathology subtype classification.

References

[1] T. Karras, T. Aila, S. Laine, and J. Lehtinen. Progressive growing of gans for improved quality, stability, and variation. In *International Conference on Learning Representations (ICLR)*, 2018.

[2] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers. Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3462–3471, 2017.

[3] J. Irvin, P. Rajpurkar, M. Ko, et al. Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, 2019.

[4] J. P. Cohen, M. Luck, and S. Honari. Torchxrayvision: A library of chest x-ray datasets and models, 2020.

[5] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville. Improved training of wasserstein gans, 2017.

[6] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial networks, 2014.

[7] N. Park, M. Mohammadi, K. Gorde, S. Jajodia, H. Park, and Y. Kim. Data synthesis based on generative adversarial networks. *Proceedings of the VLDB Endowment*, 11(10):1071–1083, 2018.

[8] H. Salehinejad, E. Colak, T. Dowdell, J. Barfett, and S. Valaee. Generalization of deep neural networks for chest pathology classification in x-rays using generative adversarial networks. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018.

[9] S. J. Pan and Q. Yang. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10):1345–1359, 2010.

[10] Y. Wang, Z. Li, Z. Wu, H. Wang, J. Song, and P. Xie. Analyzing cross-population domain shift in chest x-ray image classification and mitigating the gap with deep supervised domain adaptation. In *Proceedings of the 27th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2024.

[11] M. Raghu, C. Zhang, J. Kleinberg, and S. Bengio. Transfusion: Understanding transfer learning for medical imaging. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2019.

[12] S. Marcel and Y. Rodriguez. Torchvision: Computer vision library for pytorch. `https://github.com/pytorch/vision`, 2010.

[13] H. Talebi and P. Milanfar. Learning to resize images for computer vision tasks, 2021.

[14] B. Segal, D. M. Rubin, G. Rubin, and A. Pantanowitz. Evaluating the clinical realism of synthetic chest x-rays generated using progressively growing gans. *SN Computer Science*, 2(4), 2021.

[15] Y. Wang, C. Wu, L. Herranz, J. van de Weijer, A. Gonzalez-Garcia, and B. Raducanu. Transferring gans: Generating images from limited data, 2018.

7 Supplementary Material

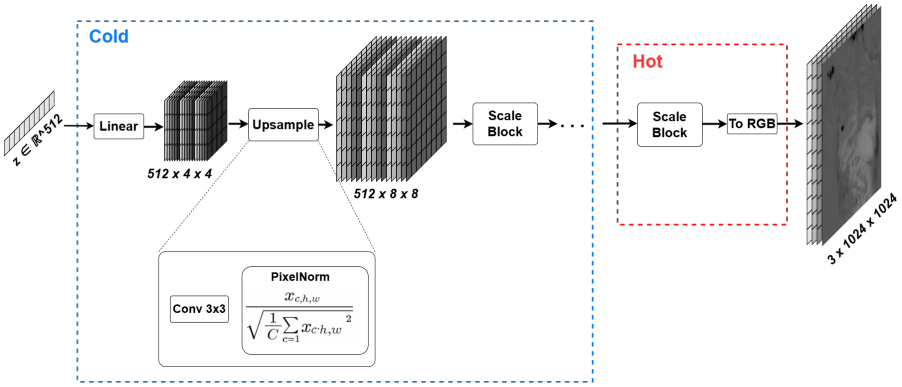

Figure 4: Progressive Growing GAN Model Architecture