Innovation/Impact: Our work shows that using transfer learning and data augmentation dramatically improves the quality of synthetically generated medical images and could be applied to any research in which generative adversarial networks (GANs) are used. It also evaluates the StyleGAN2 network on a medical imaging dataset. The StyleGAN2 network is widely regarded as the state-of-the-art (SOTA) generative model but is infrequently applied to medical tasks due to its computational complexity.

Motivation: In our broader work of GAN-based anomaly detection, we use GANs to model the training distribution of clinically deployed deep learning models. Contemporary work has shown that training GANs on limited data leads to poor quality images. The training distributions of deployed models are often comprised of limited data, necessitating methods to improve the GAN's capabilities within these constraints.

Key Results: In Table 1, we present the quantitative and qualitative results of our experiments. We see that while performing both pretraining and data augmentation improves image quality individually, the best results are achieved when the methods are used in tandem. Permutation tests on the Fréchet Inception Distance (FID) scores showed the improvements of Experiments 2, 3, and 4 upon Experiment 1 to be statistically significant with 99% confidence. In addition, permutation tests showed the FID improvement of Experiment 4 upon Experiments 2 and 3 to be statistically significant with 99% confidence. The false positive (FPR) and false negative (FNR) rates of the visual Turing test demonstrate that transfer learning and data augmentation improved the visual quality of the images. When transfer learning and data augmentation were used together, almost all participants achieved results consistent with random guessing.

Experiments	FID (↓)	FPR (↑)	FNR (↑)
1. Vanilla	10.70 (±0.72)	24.17 (±26.91)%	29.17 (±20.35)%
2. Pretraining	7.62 (±0.35)	28.33 (±22.51)%	32.67 (±16.33)%
3. Augmentations	7.51 (±0.89)	46.67 (±11.69)%	33.33 (±16.63)%
4. Pretraining and Augmentations	**5.22 (±0.17)**	**53.00 (±9.06)%**	**42.50** (±10.37)%

Table 1: The results of an ablation study evaluating the effects of transfer learning and data augmentation on synthetic image quality.

Figure 1 shows the average losses for each experiment across the five training runs. We see that data augmentation significantly reduces the generator's loss (and thereby increases the discriminator's loss). Figure 2 shows randomly chosen synthetic images from Experiments 1 and 4. There is a noise artifact in all images generated by Experiment 1 that is not present in any of the images from Experiment 4.

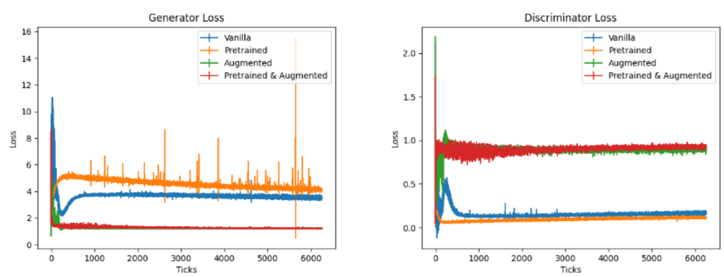

Figure 1: The averaged loss of the generator (left) and discriminator (right) across training time.

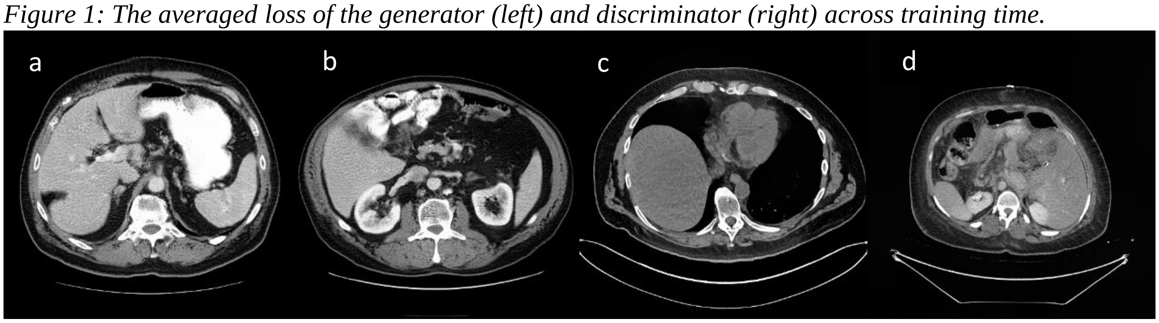

Figure 2: Randomly chosen generated examples from Experiments 1 (a, b) and 4 (c, d).

Discussion: GANs underperform when presented with limited datasets, which is often the case in medical tasks. Pretraining using large amounts of data vastly enhances the GANs' performance, even when the data is not directly related to the task at hand. For instance, in our experiments we pretrained on facial images. Data augmentation also improves the GANs' performance. However, the most substantial gains are achieved when both transfer learning and data augmentation are used simultaneously.