

**Purpose:** To compare the performance improvements between transfer learning, data augmentation, and additional training data when applied to the state-of-the-art generative model, StyleGAN2, on a high-resolution, limited-size medical dataset.

**Methods:** Our dataset contained 3,234 abdominal computed tomography scans from 456 patients windowed with level 50 and width 350. All axial slices containing liver anatomy were mapped to 512x512 PNG images. Five datasets of various sizes were created such that every dataset was a subset of all larger datasets. We trained a StyleGAN2 network with eight experimental setups: 10,600 images (1) with no pretraining or augmentation, (2) with pretraining only, (3) with augmentation only, (4) with pretraining and augmentation, and (5) 20,579 images, (6) 44,578 images, (7) 84,799 images, and (8) 153,945 images with no pretraining or augmentation. For pretraining, we used weights from StyleGAN2 trained on the Flickr-Faces-HQ dataset. For data augmentation, we used horizontal flipping and adaptive discriminator augmentations. We evaluated the quality of the synthetic images with the Fréchet Inception Distance (FID) and visual Turing tests.

**Results:** Using transfer learning and data augmentation together resulted in a greater improvement of the FID (63%) than increasing the size of the dataset about fifteen-fold (44%). All methods reduced noise artifacts, enhanced detail, and provided superior anatomical accuracy in the synthetic images. When transfer learning and data augmentation were used, it became harder for participants to distinguish between real and fake images (90% increase in the false positive rate). For our best performing generative model, radiologists and radiation oncologists labeled fake images as real 42% of the time. Finally, data augmentation stabilized the generative adversarial network training dynamics.

**Conclusion:** Using data augmentation and transfer learning resulted in greater performance gains than including additional training data. Data augmentation proved to be an effective tool in mitigating training divergence on medical images.