Title: Improving the Generation of Synthetic Medical Images using Data Augmentation and Transfer Learning

Authors:

Institutions: University of Texas MD Anderson Cancer Center, Rice University, Baylor College of Medicine, UC San Diego

Purpose: Generative adversarial networks (GANs) have recently been used in medical imaging for anomaly detection, data augmentation, image translation, image segmentation, treatment planning, and image reconstruction. Past works have demonstrated that limited data leads to poor quality images. We aim to show that both data augmentation and transfer learning improve synthetic medical image generation.

Methods: Our dataset contained 97 non-contrast and 108 contrast-enhanced abdominal computed tomography scans windowed with level 50 and width 350. All slices containing the liver were mapped to PNG images. The final training set contained 10,600 512x512 images. We trained a StyleGAN2 network with four experimental setups: (1) no data augmentation or pretraining, (2) pretraining only, (3) data augmentation only, and (4) both pretraining and augmentations. For the pretraining experiments, we used weights from a StyleGAN2 network trained on the Flickr-Faces-HQ dataset. For data augmentation, we used horizontal flipping and ADA augmentations. We repeated each experiment five times. We evaluated the quality of the synthetic images with the Fréchet Inception Distance (FID) and a visual Turing test (VTT). The VTT was administered to six participants with medical physics backgrounds. It consisted of four subtests, one per experiment. Each test was presented in a Google Form and contained 20 real and 20 fake images.

Results: The average FID scores for Experiments 1, 2, 3, and 4 were 10.70, 7.62, 7.51, and 5.22 respectively (lower is better). One-sided permutation tests revealed that the perceived quantitative improvements of transfer learning and data augmentations were statistically significant (99% confidence). The average false positive and false negative rates calculated from the VTT results increased across each experiment, showing that transfer learning and data augmentations improved the perceptual generation quality.

Conclusion: We find that both transfer learning and data augmentations resulted in substantial qualitative and quantitative improvements in synthetic image quality.