Lung Ultrasound Image Generation Using Generative Adversarial Networks and Its Evaluation Based on Quality Metrics.

Lisman, Iván A[1,2], Veiga, Ricardo A[2,3], Acquaticci, Fabián[1,3]
1. Instituto Nacional de Tecnología Industrial, San martín, Buenos Aires, Argentina.
2. Departamento de electrónica, Facultad de Ingeniería, Universidad de Buenos Aires, Buenos Aires, Argentina.
3. Instituto de Ingeniería Biomédica, Universidad de Buenos Aires, Argentina.

Abstract

One of the main problems in machine learning is the lack of data, especially in medicine. Thus, this research presents the development and evaluation of a generative adversarial network (GAN), trained with lung ultrasound images of actual patients, with the purpose of obtaining artificially generated samples from the same distribution. In particular, a Deep Convolutional-GAN (DCGAN) architecture was used. In order to train this neural network, a new lung ultrasound database was created by combining several publicly available datasets. They are lung ultrasound snapshots extracted from B-mode videos using a convex transducer. The dataset includes 10 healthy patients and 14 patients with acute respiratory distress syndrome (ARDS), predominantly due to COVID-19. Each video was segmented into 20 frames, resulting in a total of 200 images for healthy patients and 280 for sick patients. All images have a resolution of 256x256 pixels and a greyscale of 256 values. After training this neural network, successful generated images were achieved similar to those from the two types of patients. Two objective metrics were used to evaluate the similarity between the original and generated images. The Fréchet Inception and Kernel Inception Distances (FID and KID, respectively) were used to measure the distance between the most relevant feature representations of the generated images and those of the real ones. Both metrics were used not only to evaluate the performance of the network but to perform an optimization of the latent space dimension of the GAN. Several latent space dimensions were tested {4, 10, 16, 16, 50, 64, 100, 128, 256, 512} and, for each type of patient, a minimum was found. This optimal dimension was 4 for healthy patients. On the other hand, for ARDS patients, the optimal dimension was 128, possibly because these images are more complex in structure. Additionally, the values of FID and KID corresponding to the set of images generated with these optimal latent space dimensions were compared with those corresponding to degraded versions of the original set of images (with different types and levels of noise). It was found that the generated images have a similar quality to those with low levels of degradation. For example, for the "Salt&Pepper" noise, modifying only 60 random pixels from each original image would lead to the same FID obtained with the images generated by the optimized models. Additionally, the training of the GAN was analyzed to detect signs of memorization, mode collapse or other potential problems during the learning process. Finally, a subjective evaluation of the images was conducted with medical doctors to determine the diagnostic value of the generated images. Two different tests were presented to nine experts, assessing their preference on real and GAN-generated images and then asking them to rate their quality. The results showed that the participants did not have a preference for one type of image over the other, and the average scores for both categories of images were similar. This research adds value to the field of lung ultrasound image synthesis by using optimization based on a quality metric, offering a quantitative approach to enhance the quality of generated images. These results may be valuable for improving the generation of medical images, which could potentially be used for data augmentation to enhance the performance of classifier models trained on such images.