Evaluation of Deep Convolutional Generative Adversarial Networks for data augmentation of chest X-ray images

Contact Information:
Sagar Kora Venu
Department of Analytics and Data Science
Email: Skora@my.harrisburgu.edu

HARRISBURG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Abstract

Medical image datasets are generally highly imbalanced, due to over-representation of common medical problems and lack of data on unusual conditions. Training deep neural network model on such datasets to accurately classify the medical condition does not yield desired results and often over-fits the data on majority class samples. In order to address this issue, data augmentation is often performed on training data by position augmentation techniques such as scaling, cropping, flipping, padding, rotation, translation, affine transformation, and color augmentation techniques such as brightness , contrast, saturation, and hue to increase the dataset sizes. These augmentation techniques are not guaranteed to be advantageous in domains with limited data, especially medical image data, and could lead to further overfitting. In this work, we performed data augmentation on Chest X-rays dataset through generative modeling (deep convolutional generative adversarial network) which creates artificial instances retaining similar characteristics to the original data and evaluation of the model resulted in Fréchet Distance of Inception (FID) score of 1.289..

Introduction

Data sets for medical imaging are limited in size due to privacy issues and getting annotation of medical images is expensive and time-consuming, which often leads to having only small amounts of labeled medical imaging data to use for image classification tasks. Deep learning techniques need a huge volume of data to train effective models for tasks such as image recognition/ classification. Data augmentation is a technique commonly used in deep learning to expand data and prevent over-fitting in such data-limited situations. In this work, we investigate the use of Deep Convolutional Generative Adversarial Networks for generating chest X-ray images to augment the original dataset.Generative Adversarial Networks (GAN's) were introduced by Ian Goodfellow and his colleagues in 2014 [1]. GAN's utilize two neural networks, a generator which takes random noise as input to create samples (data) as realistic as possible to the original dataset and a discriminator to distinguish between data that is real (original data) vs fake (generated data) as shown in Figure 1. In this work, we used deep convolutional neural networks for both generator and discriminator as proposed by [3], hence the name Deep Convolutional Generative Adversarial Network.

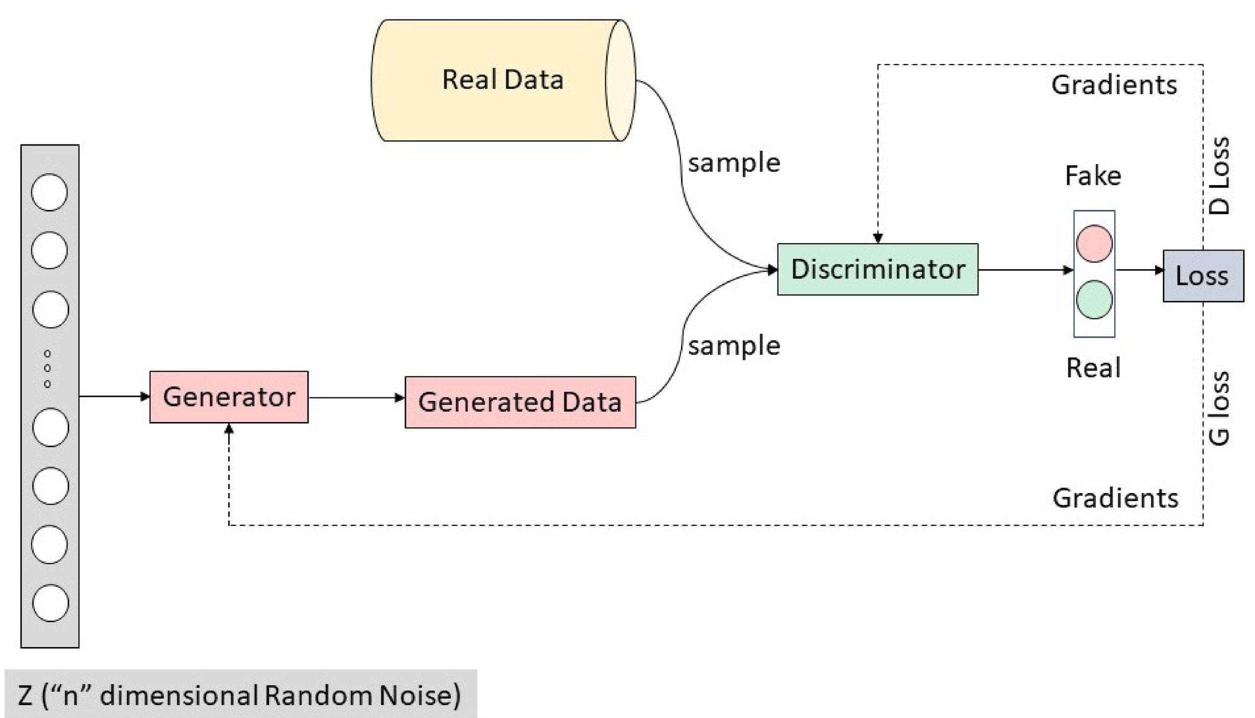

Figure 1: Generative Adversarial Network Architecture

Materials and Methods

We used X-ray images data obtained by [2] in the experiment. The dataset was already organized into three folders (train, test, val) and each folder contained sub-folders for each image category (Normal/ Pneumonia) with 5216 X-ray images in the train folder (1341 images are labeled with Normal and 3875 images labeled with Pneumonia), 16 X-ray images in val folder and 624 X-ray images in test folder. Its obvious that the data in the train folder is imbalanced and training a neural network to classify the data among two categories will over-fits the data . So, in this experiment, we augment the Normal X-ray images by Deep Convolutional Generative Adversarial Networks with the architecture as shown in Figure 2

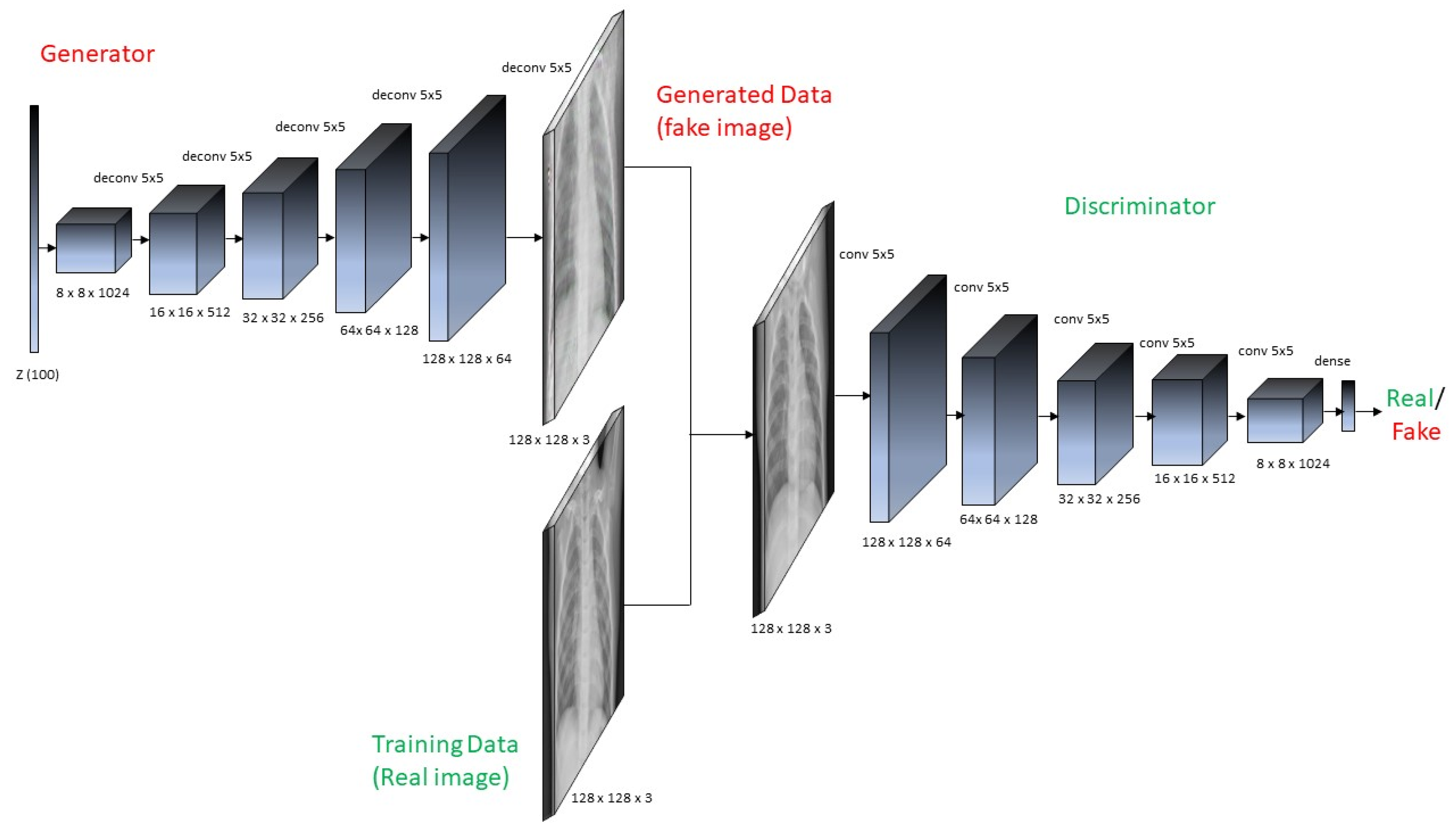

Figure 2: Deep Convolutional Generative Adversarial Network Architecture

The images were resized to 128x128 pixels due to GPU memory constraints and then the images are scaled to a [-1,1] pixel value range. In this architecture, a 100x1 noise vector is fed as an input to the generator. There are then four Convolutional layers with 2D-upsampling layers applied with Leaky ReLu activation function interlaced in between to scale to the appropriate 128x128 image size. The discriminator network is a similar network with four convolutional layers and a stride of 2 with leaky ReLU as an activation function except for the final node which is a sigmoid activation function to output if the image is real (original data) or fake (generated data).

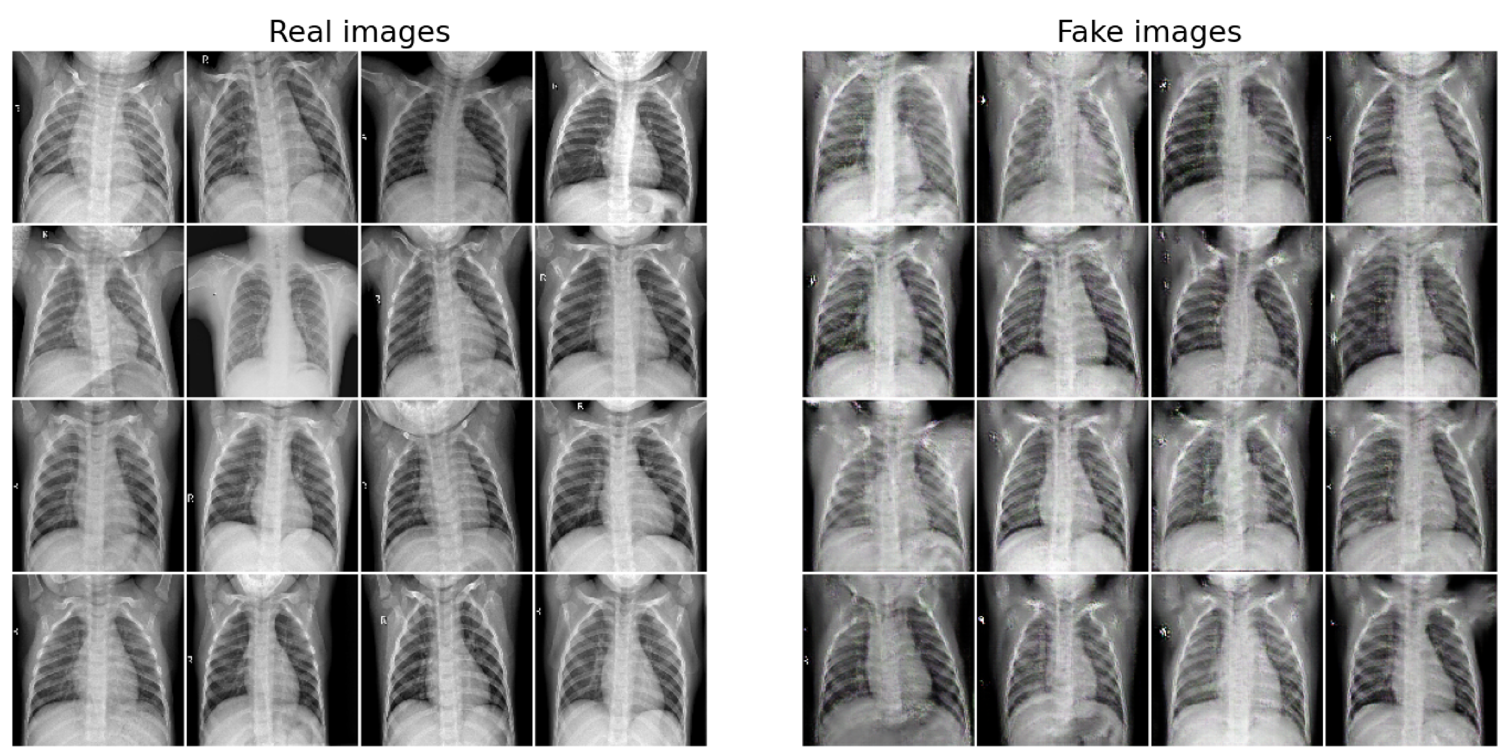

Figure 3: Images from Original dataset (Real) and Images generated by the generator of DCGAN (fake)

GAN Objective Function

The learning process of the GANs is to train a discriminator and a generator simultaneously, which is otherwise a min-max game between discriminator and generator where, the discriminator tries to maximize the loss function and the generator tries to minimize the loss function as shown in the equation (1).

$$\min_G \max_D V_{\mathrm{GAN}}(D,\ G) = \mathbb{E}_{\boldsymbol{x}\sim p_{\mathrm{data}}(\boldsymbol{x})}[\log D(\boldsymbol{x})] + \mathbb{E}_{\boldsymbol{z}\sim p_{\boldsymbol{z}}(\boldsymbol{z})}[\log(1 - D(G(\boldsymbol{z})))] \tag{1}$$

Results

The DCGAN was trained for 500 epochs and in just around 50 epochs, the DCGAN was able to generate images that resembled the chest X-ray images. As shown in figure 3, the generator of the DCGAN was able to generate realistic chest X-ray images (fake images). For comparison, we show a grid of Real images (original data) and the fake images (generated images).

Evaluation

In this work, we evaluate the DCGAN model using Fréchet Distance of Inception (FID) measure as shown in the equation (2).

$$d_{FID}(x, g) = \left\|\mu_x - \mu_g\right\|^2 + Tr\left[\Sigma_x + \Sigma_g - 2\left(\Sigma_x\Sigma_g\right)^{\frac{1}{2}}\right] \tag{2}$$

To calculate the Gaussian statistics (mean and covariance), the number of samples (real images and generated images respectively) should be greater than the dimension of the coding layer i.e., the samples should be greater than 2048 for the Inception-V3 pool 3 layer, otherwise the covariance is not full rank resulting in complex numbers and NAN's. Since, we had very limited samples (less than 2048) in our training dataset, we could not take advantage of the Inception-V3 pool3 layer, so we used the previous layer which is a Pre-aux classifier that is a 768-dimensional feature. We then calculated the Fréchet Distance of Inception (FID) score on the model which achieved a FID score of 1.289 (lower scores correspond to better GAN performance).

Conclusions

- The contributions of Generative Adversarial Networks to the field of Medical imaging are highly appreciated, especially where there is limited access to the medical imaging data and and the high costs of obtaining the labeled data.
- In this study, we are able to generate realistic chest X-ray images that resemble the chest X-ray images from the original dataset and evaluated the model using Fréchet Distance of Inception (FID) achieving a score of 1.289..

Forthcoming Research

- To test the visual quality of the generated X-ray images, we intend to supply the generated images to a clinician to label the images as either real or fake (generated).
- Develop a deep convolutional neural network to improve the accuracy in classifying the medical condition by utilizing the generated images alongside real training images.

References

[1] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

[2] Daniel S Kermany, Michael Goldbaum, Wenjia Cai, Carolina CS Valentim, Huiying Liang, Sally L Baxter, Alex McKeown, Ge Yang, Xiaokang Wu, Fangbing Yan, et al. Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell*, 172(5):1122–1131, 2018.

[3] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.

Acknowledgements

We thank [2] for making the datasets publicly accessible and we also thank Harrisburg University of Science and Technology for their support.