Synthetic Medical Image Augmentation for Classification of Pigmented Lesions

Nicolas Essipova, Peter Lakatos, Marios Chatiras
6th December 2019

1 Project Description

Generative adversarial networks, or GANs, are a class of machine learning systems that are used to generate new data given a data set, with the same statistical properties. Our work aims to display the effectiveness of new data generation as a form of image augmentation technique to improve the predictive performance of a convolutional neural network. Our work will be demonstrated with and without the synthetic image augmentation, for comparative purposes.

2 Implementation

We will be approaching this problem with our own variation of implementation of the technique and methodology first introduced in a paper by Frid-Adar et al. 2018 “GAN-based Synthetic medical Image Augmentation for Increased CNN Performance in Liver Lesion Classification” [1].

We will first describe and prepare our data to fit with the scope of our project. Thereafter, we will experiment with training our network before moving onto exploring the utilization of other pre-trained CNN models – all without the use of synthetic data. We will note the results before we proceed with training our Deep Convolutional Generative Adversarial Network, DCGAN, [2] and then training the same CNN models with the use of generated data from the DCGAN. We will then compare these results with the previously noted results and discuss our results.

Due to the distributional bias in the dataset, we will also need to specify the sensitivity and specificity metrics when evaluating our classifiers.

3 Tools

We will utilize a computational graph library such as TensorFlow 2.x or PyTorch 1.x for our implementation of CNN and DCGAN. There is a possibility we will have to use open source libraries for pre-processing. With consideration to the size of the dataset, we may also have to explore the option of conducting our work with distributed computing or cloud computing, such as on Paperspace or Google Cloud Platform.

4 Data

The dataset is from the International Skin Imaging Collaboration effort and it is for the diagnosis of pigmented skin lesions to spot signs of melanoma, which is the deadliest form of skin cancer. The dataset contains 10.000 dermatoscopic images and consists of all types of pigmented lesion categories [3].

References

[1] Frid-Adar et al. “GAN-based Synthetic Medical Image Augmentation for increased CNN Performance in Liver Lesion Classification”, 2018. Available at: https://arxiv.org/abs/1803.01229 [Accessed 061219]

[2] Alec Radford, Luke Metz, Soumith Chintala. “Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks”, 2015. Available at: https://arxiv.org/abs/1511.06434 [Accessed 061219].

[3] Skin Cancer MNIST: HAM10000. Available at, along with detailed data source and licensing: https://www.kaggle.com/kmader/skin-cancer-mnist-ham10000 [Accessed 061219].