GAN-based Data Augmentation for Improved Liver Lesion Classification

Maayan Frid-Adar
Department of Biomedical Engineering
Tel Aviv University
Tel Aviv, Israel
maayan.frid@gmail.com

Eyal Klang, Michal Amitai
Department of Diagnostic Imaging
The Chaim Sheba Medical Center
Tel-Hashomer, Israel
eyalkla@hotmail.com
michal.amitai@sheba.health.gov.il

Jacob Goldberger
Faculty of Engineering
Bar-Ilan University
Ramat-Gan, Israel
jacob.goldberger@biu.ac.il

Hayit Greenspan
Department of Biomedical Engineering
Tel Aviv University
Tel Aviv, Israel
hayit@eng.tau.ac.il

Abstract

We propose a method for generating synthetic medical images using deep learning Generative Adversarial Networks (GANs). We show that the generated medical images can be used for synthetic data augmentation, and improve the performance of CNN for medical image classification. Our novel method is demonstrated on a limited dataset of computed tomography (CT) images of 182 liver lesions (Cysts, Metastases and Hemangiomas). Using synthetic data augmentation we achieved an improvement of $\sim 7\%$ in accuracy for the liver lesion classification task. This work was recently submitted for journal publication [1].

1 Introduction

Deep learning methods, and in particular convolutional neural networks (CNNs), have led to an enormous breakthrough in a wide range of computer vision tasks, primarily by using large-scale annotated datasets. However, obtaining such datasets in the medical domain remains a challenge, mainly because expert annotation is expensive the diseases are scarce [2].

In our paper [1], we present methods for synthesis of high quality focal liver lesions from CT images using generative adversarial networks (GANs), design of a CNN-based solution for the liver lesion classification task and augmentation of the CNN training set using the generated synthetic data - for improved classification results.

2 Generating Synthetic liver lesions

Even a small CNN has thousands of parameters that need to be trained. The standard solution to reduce overfitting of a small dataset is data augmentation that artificially enlarges the dataset [3]. Classical augmentation techniques on gray-scale images include mostly affine transformations. To enrich the training data we apply here an image synthesis technique based on the GAN network. The approach we propose involves several steps: in the first step, standard data augmentation is used to create a larger dataset which is then used to train a GAN. The synthetic examples created by the GAN are next used as an additional resource for data augmentation. The combined standard and synthetic...
augmentation is finally used to train a lesion classifier. Examples of real and synthetic lesions are shown in Figure 1.

The GAN network [4] is a specific framework of a generative model. It aims to implicitly learn the data distribution $p_{\text{data}}$ from a set of samples (e.g. images) to further generate new samples drawn from the learned distribution. We employed the Deep Convolutional GAN (DCGAN) [5] for synthesizing labeled lesions for each lesion class separately. The model consists of two deep CNNs that are trained simultaneously, as depicted in Figure 2.

Figure 1: Left side: Real lesions; Right side: Synthetic lesions. Lesion ROI examples of Cysts (top row), Metastases (middle row) and Hemangiomas (bottom row).

The GAN network [4] is a specific framework of a generative model. It aims to implicitly learn the data distribution $p_{\text{data}}$ from a set of samples (e.g. images) to further generate new samples drawn from the learned distribution. We employed the Deep Convolutional GAN (DCGAN) [5] for synthesizing labeled lesions for each lesion class separately. The model consists of two deep CNNs that are trained simultaneously, as depicted in Figure 2.

Figure 2: Deep Convolutional GAN Architecture (generator+discriminator).

3 Experiments and Results

The dataset contains 182 2D CT scans: 53 cysts, 64 metastases, 65 hemangiomas. An expert radiologist marked the margin of each lesion and determined its corresponding diagnosis which was established by biopsy or a clinical follow-up. We use a liver lesion classification CNN of 5 layers to determine the network predictions into the three lesion classes. For training we used a batch size of 64 with a learning rate of 0.001 for 150 epochs. The input to our classification system are ROIs of $64 \times 64$ cropped from CT scans. In all experiments and evaluations we used 3-fold cross validation with case separation at the patient level and each fold contained a balanced number of cyst, metastasis and hemangioma lesion ROIs.

3.1 Evaluation of the Synthetic Data Augmentation

We started by examining the effects of using only classic data augmentation for the liver lesion classification task (our baseline). We recorded the classification results for the liver lesion classification CNN for increasing amounts of data augmentation over the original training set. The second step of the experiment consisted of generating synthetic liver lesion ROIs for data augmentation using GAN. We employed the DCGAN architecture to train each lesion class separately, using the same 3-fold cross validation process and the same data partition. After the generator had learned each lesion class data distribution separately, it was able to synthesize new examples. We examined the classification results after adding the synthesized lesion ROIs to the training set.

Results of the GAN-based synthetic augmentation experiment are shown in Figure 3. We see the total accuracy results for the lesion classification task, for each group of data. The red line shows the effect of adding classic data augmentation. The results improved as the number of training examples increased, up to saturation around 78.6% where adding more augmented data examples failed to improve the classification results. We note that the saturation starts with $\sim 5000$ samples per fold. The blue line shows the effect of adding synthetic data augmentation. The classification results...
significantly improved from 78.6% with no synthesized lesions to 85.7% for ∼ 8000 samples per fold. Overall, the classification performance using only classic data augmentation yielded 78.6% sensitivity and 88.4% specificity. By adding the synthetic data augmentation the results increased to 85.7% sensitivity and 92.4% specificity.

In the paper [1], we present additional evaluations of the generated lesions. We use t-SNE visualization of the lesion classification CNN features to show the improvement of the features separation using generated examples. Furthermore, we asked expert radiologists to evaluate the quality of the synthesized lesions. We found that the experts had similar classification accuracy results for the real set, as well as the synthesized lesions set, indicating to us the validity of the lesion generation process. As a final experiment, we compared the performance of the CNN-based system, to state-of-the-art methods for liver lesion classification [6] and demonstrated improved performance.

4 Conclusion

In conclusion, we presented a method that uses the generation of synthetic medical images for data augmentation to improve performance on a medical problem with limited data. We demonstrated this technique on a liver lesion classification task and achieved an improvement of ∼ 7% using synthetic augmentation over the classic augmentation. We believe that other medical problems can benefit from using synthetic augmentation, and that the presented approach can lead to stronger and more robust radiology support systems.

References


