Application of Generative Adversarial Networks in augmentation of medical image datasets

Feng Cheng, Yu Zhao Ye
College of Physical Science and Technology XMU
1075068497@qq.com

Abstract

How to cope with the questions (e.g. overfitting, owe fitting) that by small datasets and limited amount of annotated samples in the CNN-based Image Identification is a significant problem faced in the medical imaging domain. In this paper, we present a data augmentation method that generates synthetic medical images using Generative Adversarial Networks (GANs). Our work proposing a training scheme that first uses classical data augmentation to enlarge the training set and then further enlarges the data size and its diversity by applying GAN techniques for synthetic data augmentation. We compare the classification performance before and after using GAN to expand the data set. The classification performance using GAN to expand the data set batter than using only classic data augmentation.

Introduction

On the one hand, how to train a model by small datasets and limited amount of annotated samples when employing supervised machine learning algorithms that require labeled data and larger training examples. Especially in the field of medical imaging, most annotations that made by radiologists with expert knowledge on the data are time consuming. Although public medical datasets are available online, and grand challenges have been publicized, most datasets are still limited in size and only applicable to specific medical problems. [4] Moreover, collecting medical image-data is a complex and expensive procedure that requires assisting work lots of professional doctors and diagnostic specialists.

Most of researchers use data augmentation schemes commonly including simple modifications of dataset images such as rotation translation flip and scale to overcome this problem. [6] Using such data augmentation to improve the training process of networks has become a standard procedure in computer vision tasks.

But relying on traditional methods to augment datasets image alone is not enough. The datasets image is hard to get close to accuracy of the real image data. Therefore, we propose a way to improve the case. This promising approach for training a model that synthesizes images is known as Generative Adversarial Networks (GANs). [3] GANs are a framework for the estimation of generative models via an adversarial process in which two models, a discriminator D and a generator G, are trained simultaneously. The generator G aim is to capture the data distribution, while the discriminator D estimates the probability that a sample came from the training data rather than G. To learn a generative distribution pg over the data x the generator builds a mapping from a prior noise distribution P_z to a data space as $G(z; \theta_G)$, where θ_G are the generator parameters. The discriminator outputs a single scalar representing the probability that x came from real data rather than from pg. The generator function is denoted with $D(x; \theta_D)$, where θ_D are discriminator parameters[8]. GANs have gained great popularity in the computer vision community and different variations of GANs were recently proposed for generating high quality realistic natural images. [2]

Recently, the driving focus of the medical image analysis (MIA)community has been on the supervised learning of decision boundaries, while generative tasks have been on the back seat. This changed dramatically with the advent of generative adversarial networks (GANs) , which lead to a new age of generative modeling and distribution learning. [7] With their abilities to mimic data distributions and to synthesize images at yet unprecedented levels of realism, GANs have carved open new ways to bridge the gap between supervised learning and image generation. GANs can discover the high dimensional latent distribution of data, which has led to significant performance gains in the extraction of visual features. [5]

Most studies have employed the image-to-image GAN technique to create label-to-segmentation translation, segmentation to-image translation or medical cross modality translations [4]. We focus on study the GAN framework to synthesize high quality medical images for data augmentation. We compare the classification performance before and after using GAN to expand the data set. The classification performance using GAN to expand the data set batter than using only classic data augmentation.

RELATED WORK PROPOSED SOLUTION

Even a small CNN has thousands of parameters that need to be trained. When using deep networks with multiple layers or dealing with limited numbers of training images, there is a danger of overfitting which especially the sample size is small. The Classic solution to reduce overfitting is data augmentation that artificially enlarges the dataset. Classical augmentation techniques on images include mostly affine transformations. To enrich the training data we apply here an image synthesis technique based on the GAN network. Examples of real and synthetic retina are shown in Figuere 1.

Classic Data Augmentation

Classic augmentation techniques on the images include mostly affine transformations such as translation, rotation, scaling, flipping and shearing. In order to preserve the retina characteristics, we avoided transformations that cause shape deformation (like shearing and elastic deformations). We took N = 1080 images as the training set. The training set images are vertical flipped N-vf times with probability 1 ($p = 1$), and horizontal flipped N-hf times with probability 0.1 ($p = 0.1$). In addition, we generated the retina images N-col times that has a difference in brightness (between 0.7 and 0.9), contrast (between 0.7 and 0.9) and saturation (between 0.7 and 0.9). As a result of the augmentation process, the total number of augmentations was $N = N(1 + N - vf + N - hf + N - col)$. Bicubic interpolation was used to resize the images to a uniform size of 64 $\times$64.

GAN Networks for Retina Synthesis

GANs are a specific framework of a generative model. It aims to implicitly learn the data distribution p data from a set of samples (e.g. Images) to further generate new samples drawn from the learned distribution. We employed the Wasserstein GAN [1]for synthesizing retina images. The model consists of two deep CNNs that are trained simultaneously. A sample x is input to the discriminator (D), which outputs D(x), its probability of being a real sample. The generator (G) gets input samples z from a known simple distribution pz, and maps G(z) to the image space of distribution pg. During training, the generator improves its ability to synthesize more realistic images while the discriminator improves its ability to distinguish the real from the synthesized images. Hence the moniker of adversarial training.

We put a vector of 100 random numbers drawn from a uniform distribution as input and outputs a retina image of size $64 \times 64 \times 3$. The generator network architecture consists of a fully connected layer reshaped to size $4 \times 4 \times 1024$ and four fractionally-strided convolutional layers to up-sample the image with a 5×5 kernel size. The discriminator network has a typical CNN architecture that takes the input image of size $64 \times 64 \times 3$, and outputs a decision - if the retina is real or fake. In this network, four convolution layers are used, with a kernel size of 5×5 and a fully connected layer. Strided convolutions are applied to each convolution layer to reduce spatial dimensionality instead of using pooling layers.

EXPERIMENTS AND RESULTS

Data and Implementation

The dataset was mass up of 1480 retina images, and splited 1080 train dataset and 400 test dataset. The train dataset

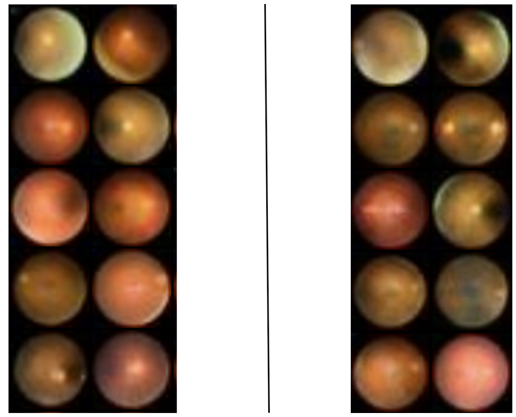

Figure 1: Left side: Real retina; Right side: Synthetic retina

contained 5 class images which were established by Fundus Camera and were reshaped size of 64×64. We used the train dataset to fit the original distribution by WGAN, and we ended up with about 2000 fake images. We selected 443 high quality images from the fake ones, and splited the high quality images into 5 class (label: number of images, [0]: 101, [1]: 96, [2]: 84, [3]: 80, [4]: 82)

We use Resnet101 and Densenet121 to verify data augmentation results. For training, we used a batch size of 64 with a learning rate of 0.0001 for 40 epochs. For the implementation of the retina classification CNN we used pytorch framework. For the implementation of the WGAN architectures, we used pytorch framework. In addition, all training processes were performed using an NVIDIA GeForce GTX 1080 Ti GPU in Colab.

Figure 2: Classic Augmentation Only

Evaluation of the Synthetic Data Augmentation

We started by examining the effects of using resnet101 and densenet121 of none data augmentation for the retina classification task, but due to small number of samples, overfitting is inevitable. We then using only classic data augmentation to cope with overhitting, but with little success. Then we synthesized the retina images and using data augmentation with it and examined the classification

Figure 3: Classic and Synthesized Images Augmentation

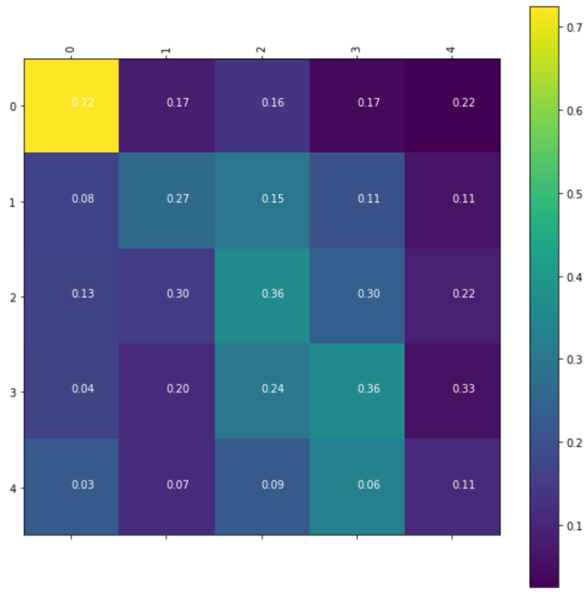

Figure 4: Confusion Matrix for the Classical Data Augmentation

results after adding the synthesized retina images to the training set. As is shown in Figure 2 and Figure 3, the way of data augmentation by synthesize images improved the overfitting effectively.

As is shown in Table1, we used none augmentation as our baseline. We recorded the classification results for the retina classification CNN for increasing amounts of data augmentation over the original training set. In order to examine the effect of adding increasing numbers of examples, we formed the data group to increase the images to 4320. But Resnet and Densenet went up from 45.96% and 46.44% to 47.08% and 48.18% respectively. The confusion matrix for the test results of Resnet appears in Figure 4.

The second step of the experiment consisted of generating synthetic the retina images for data augmentation using GAN. We employed the WGAN architecture to train the retina images, and in all the steps of the learning procedure

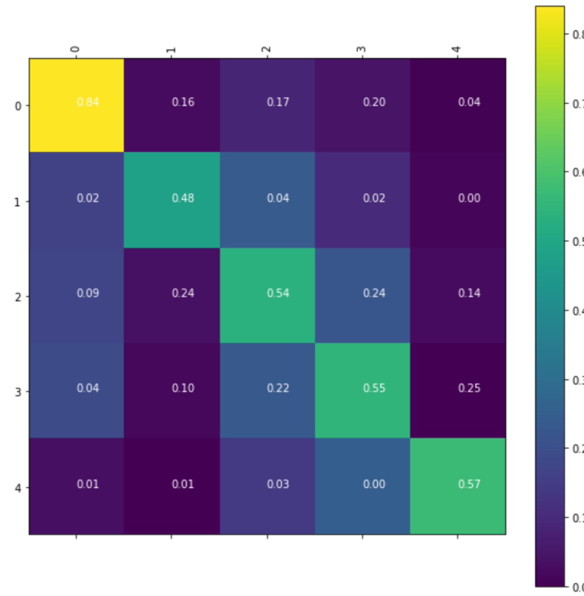

Figure 5: Confusion Matrix for the Synthetic Data Augmentation

	Resnet	**Densenet**
None Aug	45.96%	46.44%
Classic Aug	47.08%	48.18%
Classic and GAN Aug	62.86%	71.27%

Table 1: Evaluation of the Synthetic Augmentation

we maintained a complete separation between train and test subsets. In Generator training, we used a batch size of 8 with a learning rate of 0.0001 for the frist 800 epochs and with a learning rate of 0.00005 for the last 600 epochs, in total 1400 epochs. After the generator had learned the retina distribution, it was able to synthesize new examples by using an input 100dim vector of normal distributed samples.

Since our dataset was too small for effective training, we incorporated classic augmentation for the training process to increase the images of training dataset to 5649. The classification results significantly improved from 47.08% and 48.18%(Resnet and Densenet) to 62.86% and 71.27%. The confusion matrix for the best classification results using synthetic data augmentation is presented in Figure 5.

CONCLUSION

To conclude, in this work we presented a method that uses the generation of synthetic medical images for data augmentation to improve classification performance on a medical problem with limited data. We demonstrated this technique on a retina classification task and achieved a significant improvement of 15.78% and 23.09% using synthetic augmentation over the classic augmentation.

References

[1] Arjovsky, M.; Chintala, S.; and Bottou, L. 2017. Wasserstein GAN.

[2] Frid-Adar, M.; Klang, E.; Amitai, M.; Goldberger, J.; and Greenspan, H. 2018. Synthetic data augmentation using GAN for improved liver lesion classification. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, 289–293. doi:10.1109/ISBI.2018.8363576.

[3] Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; and Bengio, Y. 2014. Generative Adversarial Nets. In Ghahramani, Z.; Welling, M.; Cortes, C.; Lawrence, N.; and Weinberger, K. Q., eds., *Advances in Neural Information Processing Systems*, volume 27, 2672–2680. Curran Associates, Inc. URL https://proceedings.neurips.cc/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf.

[4] Greenspan, H.; van Ginneken, B.; and Summers, R. M. 2016. Guest Editorial Deep Learning in Medical Imaging: Overview and Future Promise of an Exciting New Technique. *IEEE Transactions on Medical Imaging* 35(5): 1153–1159. ISSN 1558-254X. doi:10.1109/TMI.2016.2553401.

[5] Kazeminia, S.; Baur, C.; Kuijper, A.; van Ginneken, B.; Navab, N.; Albarqouni, S.; and Mukhopadhyay, A. 2020. GANs for medical image analysis. *Artificial Intelligence in Medicine* 109: 101938. ISSN 0933-3657. doi:https://doi.org/10.1016/j.artmed.2020.101938. URL http://www.sciencedirect.com/science/article/pii/S0933365719311510.

[6] Krizhevsky, A.; Sutskever, I.; and Hinton, G. E. 2017. ImageNet Classification with Deep Convolutional Neural Networks. *Commun. ACM* 60(6): 84–90. ISSN 0001-0782. doi:10.1145/3065386. URL https://doi.org/10.1145/3065386.

[7] Litjens, G.; Kooi, T.; Bejnordi, B. E.; Setio, A. A. A.; Ciompi, F.; Ghafoorian, M.; van der Laak, J. A.; van Ginneken, B.; and Sánchez, C. I. 2017. A survey on deep learning in medical image analysis. *Medical Image Analysis* 42: 60 – 88. ISSN 1361-8415. doi:https://doi.org/10.1016/j.media.2017.07.005. URL http://www.sciencedirect.com/science/article/pii/S1361841517301135.

[8] Mattia, F. D.; Galeone, P.; Simoni, M. D.; and Ghelfi, E. 2019. A Survey on GANs for Anomaly Detection. *CoRR* abs/1906.11632. URL http://arxiv.org/abs/1906.11632.