Synthetic Brain Image Generation for ADHD prediction based on Progressive Growing Generative Adversarial Network

Saadia Binte Alam*, Moazzem Hossain* and Syoji Kobashi*

* *Graduate School of Engineering, University of Hyogo, Japan.*
saadiabinte@ieee.org , munnam77@gmail.com , kobashi@amec-hyogo.org

Abstract: In this paper, we proposed a method to generate synthetic brain images using generative adversarial networks (GANs). In medical image analysis, it remains a difficult and important task to produce realistic medical images that are entirely different from the original ones and also the exchange of clinical image data is a crucial issue for the implementation of diagnostic support systems. Nonetheless, it is difficult for researchers to obtain medical image data because the images contain individual information. Recently proposed GAN models could learn how to distribute training images without seeing actual image data, and generated images can completely anonymized personal information. The produced images can be used as training images for the classification of medical image, promoting medical image analysis viable. Instead of collecting a large amount of MRI data, an approach to image generation has been implemented in our paper. We exploit a progressive growing GAN (PGGAN), a neonatal brain image generation method that can be used for brain MRI classification and ADHD prediction tasks. The PGGAN slowly discovers the features of ADHD in MRI images by adding new layers during the training phase. Our image generation approach shows that it can produce brain MR images avoiding artificial artifacts and have clinical characteristics of the target symptom.

Keywords: *Generative Adversarial Network, Synthetic Medical Image generation, PGGAN*

1. INTRODUCTION

With the advancement of technologies for image recognition, aspirations have been expressed for their application to clinical devices in the field of medicine [1]. While recent machine-based image recognition techniques have shown good prospects as diagnostic support systems [2], the use-fullness of these techniques is still limited. There have been several studies where machine learning techniques were used for world-renowned modalities (e.g. CT, MRI, and mammography) [3],[4],[5]. Well-equipped study environments such as large-scale annotated medical image databases have already been built in major inspections [6], [7] and we can easily access these large-scale public clinical datasets containing data from all over the world [8]. However, few studies on regional inspections have been carried out in a limited area (hereafter referred to as minor inspections). In the case of minor examinations, it is difficult to obtain high-quality annotated clinical data, as there are different types of imaging equipment affecting the quality of the images, in medical facilities and a thorough annotation requires specialized expertise. In addition, new imaging techniques have often been established for minor inspections, and even if an annotated clinical dataset can be built from such data, the updating of training data becomes a fundamental task, unlike that for major inspections. A new data collection, adapted to the latest imaging techniques, must be planned. Hence, minor inspections suffer from a lack of high-quality accessible data, and data-driven approaches are needed to apply machine learning techniques.

Data sharing is one of the most powerful information-driven solutions that can solve the problems of lack of high-quality data and the need to update training data [9]. Nonetheless, the protection of the privacy of patients should be a top priority in the exchange of clinical information process [10]. In the field of medicine, data privacy and convenience have been known to be inversely proportional, and this is one of the most daunting problems. Clinical image data includes not only information on privacy, but also identifiers such as social security number, gender, age and occupation, and these data should be carefully handled. The probability of re-identification increases when the number of data is small. While anonymization has been carried out for the removal of identifiers, no attention has been paid to the anonymization of image data. Simplified approaches to privacy prevention issues are needed in order to accelerate the use of machine-based learning support for minor

inspections.

Medical image generation methods that allow the generation of meaningful synthetic information have attracted a great deal of attention in recent years. The wide availability of synthetic data can enable researchers to develop and validate more sophisticated techniques for the recognition of images. Namely, since the methods of image generation know how to distribute training data without referring to actual images, anonymization of individual knowledge can be realized. The use of synthetic data would lead to image recognition tasks for minor inspections that include sharing and updating of clinical data. In this study, we targeted neonatal brain MRI images for the diagnosis of ADHD.

In this paper, we exploit PGGAN, a progressively growing Generative Adversarial Network. This proposed method learns how to distribute the target data and how to produce the image after distribution from a latent space. Therefore, produced synthetic images are not linked to individual patient image information and can easily be used by researchers to build support systems.

We deal with two research questions, which GAN architecture is well adapted for practical medical image generation and how can we treat MR images with unique intra sequence variability. So our contribution in this work is to exploit PGGAN and show that it can produce realistic brain MR images, potentially leading to effective clinical applications for data augmentation for machine learning and medical imaging tasks. This work examines how to use medical images with an underlying intra-sequence variation to maximize GAN-based synthetic data generation for medical imaging.

2. RELATED WORKS

2.1 Generated Adversarial Network

The basic GAN first proposed by [11] consisting of two neural network models, a generative model G that learns about the unknown data distribution of training and a bias model D that learns to decide whether samples come from the data distribution of training. When given the previous distribution of the latent variable z after the latent distribution $p_z(z)$, the generator G takes z as the input vector and outputs the sample G(z). On the other hand, the discriminator D takes the sample x as input and outputs D(x), which represents the probability that the data is real. Both models are trained at the same time as a stochastic gradient descent

(SGD) algorithm, and their training procedures can be seen as a two-player mini-max game with the following objective function:

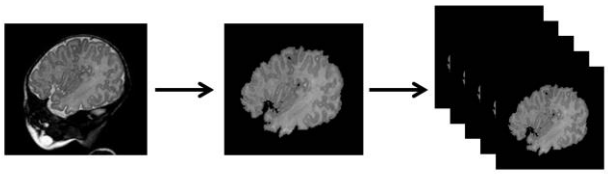
$$\min_X \max_Y V(X,Y) = \mathrm{E}_{a\sim p_{\mathrm{datal}}(a)}[\log X(a)] + \mathbb{E}_{\mathrm{z\text{-}p.}(s)}[\log(1 - X(Y(z)))] \quad (1)$$

In this equation (1), When the discriminator X attempts to maximize V (X, Y) while the generator Y attempts to minimize it. In other words, the discriminator X separates the images in a p_{data} from those of Y (z), while the generator Y produces samples to fool the discriminator X. The principle of GANs has been applied to various tasks. Supervised and unsupervised image domain transformations have been addressed in the sense of conditional image creation. For example, pix2pix learns the task of image-to-image translation using paired data samples. Nonetheless, despite the difficulty of collecting annotated paired samples, this method requires a large number of paired samples. Several researchers like UNIT[12], CoGAN[13], CycleGAN[14], and DiscoGAN[15] have suggested unpaired image-to-image translation frameworks to address this issue. In addition, recent domain classification based GANs that monitor the characteristics of the generated images by running a latent distribution have shown promising results. With a broad, easy-to-access training data set, conventional methods have already generated high-quality images.

3. MATERIALS AND METHOD

3.1 Proposed image generation method

Data preprocessing: We used 12 neonatal data which are segmented from RAW MRI and each of which has approximately 200 slices. The images are of 320x320 pixels. We pick the slices from #30 to #120 from all the slices to omit the initial / final slices, because they relay a marginal amount of useful information and negatively impact the training of PGGAN. Figure 1 provides data processing flow and examples of MR brain images used in this research.



RAW MRI Segmented Slices

Figure 1: Data Preprocessing

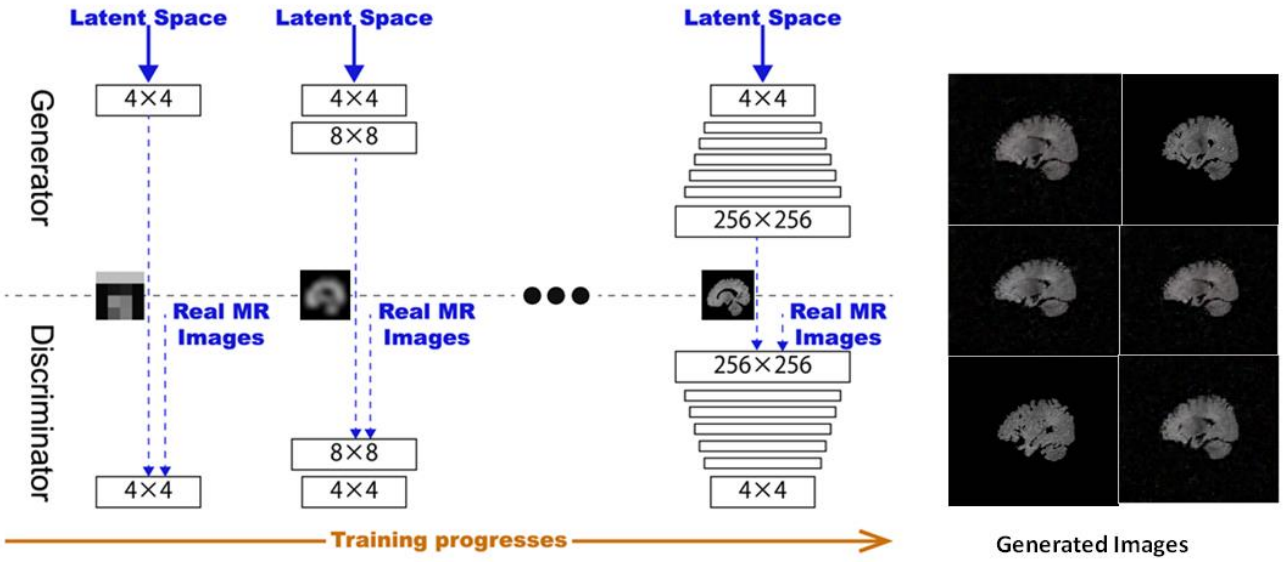


Figure 2: Network architecture to generate synthetic brain MR images.

PGGAN implementation details:

We use the Wasserstein loss PGGAN architecture with gradient penalty.

$$\mathop{\mathbb{B}}_{\hat{y}\sim\mathbb{P}_g}[X(\tilde{y})] - \mathop{\mathbb{B}}_{y\sim\mathbb{P}_r}[X(y)] + \lambda \mathop{\mathbb{B}}_{\hat{y}\sim\mathbb{P}_{\hat{y}}}\left[\left(\|\nabla_{\hat{y}}Y(\hat{y})\|_2 - 1\right)^2\right] \quad (2)$$

In the equation(2), where the discriminator X belongs to the set of 1-Lipschitz functions, P_r is the distribution of the data by the true data sample y, and P_g is the distribution of the model by the synthetic sample produced by the conditioning of the image noise samples y using a uniform distribution in[−1, 1]. The last word of the term is the gradient penalty for the random sample $\hat{y} \sim P_{\hat{y}}$. The training lasts 20,000 steps with a batch size of 4 and 2.0×10−4 learning rates for the Adam Optimizer. Once in three times, we flip the real / synthetic labels of the discriminator for robustness.

It is important to identify subtle distinctions between abnormal and normal characteristics when trying to generate synthetic images for the task of classification. Nonetheless, abnormal brain MR images frequently vary only marginally from normal brain MR images and are difficult to understand. In order to detect the subtle differences between abnormal and normal images, we use progressive growing adversarial network (PGGAN) architecture. PGGAN training starts with low resolution images and then gradually increases resolution by adding new layers to the generator and the discriminator. We train our networks with a low resolution of 4x4 pixels. Network architecture can be found in figure 2. PGGAN learns the broad outlines of the training images at the low resolution level. As training continues, layers are slowly added to the generator and the discriminator to obtain high resolution images. In a high-resolution phase, our PGGAN learns the detailed regions of the training images through following these progressive training methods, the generator will learn the features of ADHD from training images. Conditional information is also added to generate synthetic images in a high-resolution phase.

4. EXPERIMENT

4.1 Experimental settings

As for the experiment, 12 neonate MRI images were used and the images were grayscale and 320x320 pixels, separated into several patches of 256x256 pixels. The scale of the patches has been experimentally determined. In the image generation process, these patches have been resized for training. We randomly selected images from our original data and constructed our image generation training data. Several evaluation metrics, such as Inception Score[16], Fréchet Inception Distance (FID)[17] and Sliced Wasserstein Distance (SWD)][18], have been proposed for assessment of the quality of the images produced. Nevertheless, these criteria are not suitable for evaluating images for problems with classification and prediction. GAN-train measures the classification performance of a classifier trained on generated synthetic images and tests the output of a set of

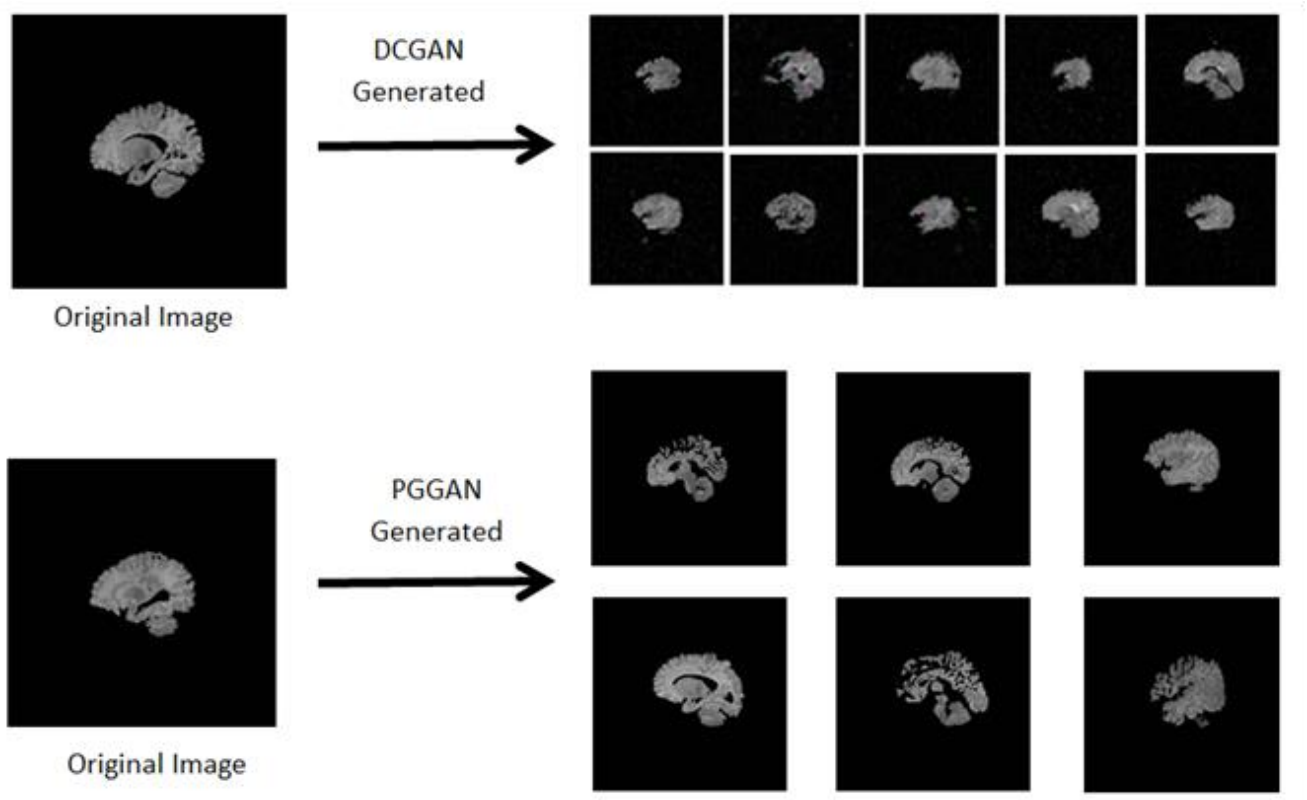


Figure 3: Our synthetic brain generation results shows that DCGAN can generate brain images with low image resolution and quality while our proposed image generation method PGGAN can generate high resolution synthetic images from original MRI images.

real images. If a set of images can be generated by an optimized GAN model that captures the target distribution perfectly, they are indistinguishable from the original training set. As anonymized generated images were used for classification in our analysis, we used GAN-train as our evaluation index.

In the experiment, the support vector machine (SVM) was used as an estimator for the MRI classification in the GAN-train. A deep learning-based estimator is the first choice in terms of MRI classification accuracy. Nevertheless, such an estimator has many parameters, and the efficiency of the classification relies heavily on setting the parameters. As a result, we used the simplest SVM as our estimator to correctly evaluate the effectiveness of the images generated. Types of features also affect the performance of the classification. Handcrafted features are an old-fashioned approach. In the experiment, we extracted high-level semantic features from the pre-trained deep model, namely the pre-trained VGG-16 models. Sensitivity (Sen), specificity (Spe) and harmonic mean Sen and Spe (HM) were used for the assessment. These parameters may be defined as follows;

Sensitivity (Sen) = TP / (TP + FN)

Specificity (Spe) = TN / (TN + FP)

Harmonic Mean (HM) = (2 x (Sen x Spe) / Sen + Spe)

Where, TP is the numbers of true positive samples, TN is true negative samples, FP is positive samples and FN is false negative samples, respectively.

4.2 Results and Discussion

The goal of our method is to generate realistic synthetic images that is shown in figure 3 for easy sharing and updating of clinical data. It is assumed that the synthesis anonymized data will be as successful as the real classification problems data. We contrasted the efficiency of the MR image classification using synthetic data as a quantitative assessment. The results of the GAN-train classification are shown in Tables 1. From the results, we can see that PGGAN outperformed the comparative methods in MR image classification efficiency, despite the fact that performance does not outperform when actual images were used as training data. On the other hand, we can see that the model trained at DCGAN generated images cannot correctly identify the actual data. Overall, we have verified that the progressively growing network architecture is successful in detecting real data distribution because DCGAN does not have such network architecture.

There are some limitations to this study. The classification performance of MR images in this study is not sufficient for clinical applications. In the experiment, instead of using deep neural networks that involve complicated parameter tuning processes, we used the simplest SVM models as our estimator since we focused on assessing the quality of the images produced.

Table1: Classification performance of PGGAN using generated images.

Training data	Sen	Spe	HM
PGGAN	**0.762**	**0.712**	**0.736**
DCGAN	0.584	0.613	0.598
Real Data	0.883	0.885	0.883

5. CONCLUSION

We have provided a synthetic brain MR image generation approach with gradually increasing adversarial learning PGGAN, which is a new, high-quality image generation system for the easier understanding, sharing and upgrading of clinical data for deep learning techniques. Besides the fact that our anonymized generated images were useful for the classification of MR images. Overall, our novel Data Augmentation approach based on PGGAN sheds light on diagnostic and prognostic medical applications, not limited to brain image generation; future studies are needed to expand our encouraging results.

REFFERENCES

[1] A. Canziani, A.Paszke, and E. Culurciello, ‘‘An analysis of deep neural network models for practical applications,’’ 2016.

[2] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi,M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez, ‘‘A survey on deep learning in medical image analysis,’’Med. Image Anal., 2017.

[3] H. R. Roth, L. Lu, J. Liu, J. Yao, A. Seff, K. Cherry, L. Kim, and R. M. Summers, ‘‘Improving computer-aided detection using convolutional neural networks and random view aggregation,’’ IEEE Trans. Med. Imag.,2016.

[4] X. Zhen, Z. Wang, A. Islam, M. Bhaduri, I. Chan, and S. Li, ‘‘Multi- scale deep networks and regression forests for direct bi-ventricular volume estimation,’’ Med. Image Anal., 2016.

[5] D. Zikic, Y. Ioannou, M. Brown, and A. Criminisi, ‘‘Segmentation of brain tumor tissues with convolutional neural networks,’’ in Proc. MICCAI- BRATS, 2014,

[6] G. Wang, W. Li, M. A. Zuluaga, R. Pratt, P. A. Patel, M. Aertsen, T. Doel, A. L. David, J. Deprest, S. Ourselin, and T. Vercauteren, ‘‘Interactive medical image segmentation using deep learning with image-specific fine tuning,’’ IEEE Trans. Med. Imag., 2018.

[7] H. Shan, Y. Zhang, Q. Yang, U. Kruger, M. K. Kalra, L. Sun, W. Cong, and G. Wang, ‘‘3D convolutional encoder-decoder network for low-dose CT via transfer learning from a 2D trained network,’’ IEEE Trans. Med. Image., 2018.

[8] H. C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, ‘‘Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning,’’ IEEE Trans. Med. Imag., 2016.

[9] J. Luo, M. Wu, D. Gopukumar, and Y. Zhao, ‘‘Big data application in biomedical research and health care: A literature review,’’ Biomed. Inform. Insights, 2016.

[10] B. Fabian, T. Ermakova, and P. Junghanns, ‘‘Collaborative and secure sharing of healthcare data in multi-clouds,’’ Inf. Syst., 2015.

[11] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, ‘‘Generative Adversarial Nets,’’ in Proc. Int. Conf. Adv. Neural Inf. Process. Syst. (NeurIPS),2014.

[12] M.-Y. Liu, T. Breuel, and J. Kautz, ‘‘Unsupervised image-to-image trans- lation networks,’’ in Proc. Int. Conf. Neural Inf. Process. Syst. (NeurIPS), 2017.

[13] M.-Y. Liu and O. Tuzel, ‘‘Coupled generative adversarial networks,’’ in Proc. Int. Conf. Neural Inf. Process. Syst. (NeurIPS), 2016.

[14] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, ‘‘Unpaired image-to-image translation using cycle-consistent adversarial networks,’’ in Proc. IEEE Int. Conf. Comput. Vis. (ICCV) 2017.

[15] T. Kim, M. Cha, H. Kim, J. K. Lee, and J. Kim, ‘‘Learning to discover cross-domain relations with generative adversarial networks,’’ in Proc. Int. Conf. Mach. Learn.(ICML), 2017.

[16] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, ‘‘Improved techniques for training GANs,’’ in Proc. Int. Conf. Adv. Neural Inf. Process. Syst. (NeurIPS), 2016.

[17] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, ‘‘GANs trained by a two time-scale update rule converge to a local nash equilibrium,’’ in Proc. Int. Conf. Adv. Neural Inf. Process. Syst. (NeurIPS), 2017.

[18] T. Karras, T. Aila, S. Laine, and J. Lehtinen, ‘‘Progressive growing of GANs for improved quality, stability, and variation,’’ (ICLR), 2017.