# TOWARDS ANNOTATING LESS MEDICAL IMAGES: PGGAN-BASED MR IMAGE AUGMENTATION FOR BRAIN TUMOR DETECTION

*Changhee Han*[1], *Hideaki Hayashi*[2], *Leonardo Rundo*[3], *Ryosuke Araki*[4], *Yudai Nagano*[1], *Yujiro Furukawa*[5], *Giancarlo Mauri*[6], *Hideki Nakayama*[1]

[1]Grad. School of Information Science and Technology, The University of Tokyo, Tokyo, Japan
[2]Dept. of Advanced Information Technology, Kyushu University, Fukuoka, Japan
[3]Dept. of Radiology, University of Cambridge, Cambridge, The UK
[4]Grad. School of Engineering, Chubu University, Aichi, Japan
[5]Kanto Rosai Hospital, Kanagawa, Japan
[6]Dept. of Informatics, Systems and Communication, University of Milano-Bicocca, Milan, Italy

## ABSTRACT

How can we tackle the lack of available annotated medical image data through Data Augmentation (DA) techniques for accurate computer-assisted diagnosis? To fill the data lack in the real image distribution, we synthesize brain contrast-enhanced Magnetic Resonance (MR) images—realistic but completely different from the original ones—using Generative Adversarial Networks (GANs). Especially, we use Progressive Growing of GANs (PGGANs) to generate original-sized $256 \times 256$ brain MR images. Our novel PGGAN-based medical DA method achieve better performance, when combined with classical DA and GAN-based refinement, in convolutional neural network-based tumor detection.

## 1. INTRODUCTION

How can we generate highly-realistic and original-sized $256 \times 256$ images, while maintaining clear tumor/non-tumor features using GANs? Our aim is to generate GAN-based synthetic contrast enhanced T1-weighted (T1c) brain MR images [1, 2]—the most commonly used sequence on tumor detection due to its high contrast—for CNN-based tumor detection. This $256 \times 256$ image generation is challenging, as GAN training is unstable with a high-resolution; towards this, we use Progressive Growing of GANs (PGGANs) [3]. Using the synthetic images, our novel PGGAN-based medical DA method achieves better performance in CNN-based tumor detection, when combined with classical DA. We also compare PGGAN-generated images with their refined images using GAN-based image-to-image translation techniques—SimGAN and UNIT.

Email:han@nlab.ci.i.u-tokyo.ac.jp

## 2. MATERIALS AND METHODS

### 2.1. The BRATS 2016 Training Dataset

This paper exploits a dataset of $240 \times 240$ T1c brain axial MR images containing 220 High-Grade Glioma cases to train PGGANs with sufficient data and resolution.

### 2.2. Proposed PGGAN-based Image Generation

**Pre-processing.** We select the slices from $\#30$ to $\#130$ among the whole $155$ slices to omit initial/final slices, since they convey a negligible amount of useful information and could affect the training of both the PGGANs and ResNet-50. For tumor detection, our whole dataset (220 patients) is divided into: $(i)$ a training set (154 patients); $(ii)$ a validation set (44 patients); $(iii)$ a test set (22 patients). Only the training set is used for the PGGAN training to be fair. Since tumor/non-tumor annotations are based on 3D volumes and highly incorrect on 2D slices, we discard $(i)$ tumor images tagged as non-tumor, $(ii)$ non-tumor images tagged as tumor, $(iii)$ unclear boundary images, and $(iv)$ too small images; after all, our datasets are composed of:

- Training set (5,036 tumor/3,853 non-tumor images);
- Validation set (793 tumor/640 non-tumor images);
- Test set (1,575 tumor/1,082 non-tumor images).

The images from the training set are zero-padded to reach a power of 2, $256 \times 256$ from $240 \times 240$ for better PGGAN training. Fig. 1 shows some real MR images.

**PGGANs.** We adopt PGGANs to generate highly-realistic and original-sized $256 \times 256$ brain MR images; tumor/non-tumor images are separately trained and generated.

**Table 1**. Tumor detection (binary classification) results with various DA, with/without ImageNet Pre-training.

| | **Accuracy** (%) | **Sensitivity** (%) | **Specificity** (%) |
|---|---|---|---|
| ResNet-50 (w/o DA) | 93.26 (86.38) | 90.95 (88.94) | 95.87 (83.62) |
| ResNet-50 (w/ 200k classical DA) | 95.02 (92.21) | 93.63 (90.21) | 96.57 (95.11) |
| ResNet-50 (w/ 400k classical DA) | 94.93 (93.24) | 91.90 (90.91) | 98.39 (95.97) |
| ResNet-50 (w/ 200k PGGAN-based DA) | 93.95 (86.25) | 92.48 (87.25) | 95.56 (84.78) |
| ResNet-50 (w/ 200k PGGAN-based DA w/o clustering and discard) | 94.80 (80.54) | 91.82 (80.02) | 98.39 (81.25) |
| ResNet-50 (w/ 200k classical DA + 200k PGGAN-based DA) | 96.18 (95.63) | 94.12 (94.24) | **98.79** (97.28) |
| ResNet-50 (w/ 200k UNIT-refined DA) | 94.31 (83.68) | 93.26 (87.75) | 96.02 (78.48) |
| ResNet-50 (w/ 200k classical DA + 200k UNIT-refined DA) | **96.70** (96.34) | 95.48 (**97.53**) | 98.29 (94.96) |
| ResNet-50 (w/ 200k SimGAN-refined DA) | 94.49 (77.66) | 92.39 (82.03) | 97.18 (71.98) |
| ResNet-50 (w/ 200k classical DA + 200k SimGAN-refined DA) | 96.36 (95.04) | 95.11 (95.07) | 97.88 (94.96) |

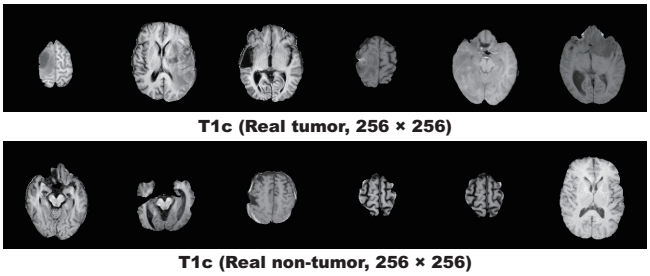


**Fig. 1**. Example real MR images used for PGGAN training.

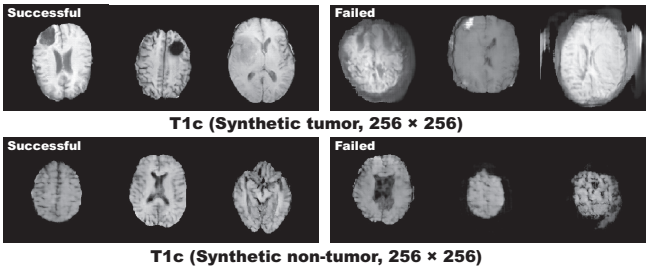


**Fig. 2**. Example synthetic MR images yielded by PGGANs: (a) successful cases; (b) failed cases.

### 2.3. Proposed ResNet-50-based Tumor Detection

**Pre-processing.** As ResNet-50's input size is $224 \times 224$, we center-crop the whole images from $240 \times 240$.

**ResNet-50.** We adopt ResNet-50 to detect tumors in brain MR images—the binary classification of images with/without tumors.

The classification results are compared ($i$) without DA, ($ii$) with 200,000 classical DA (100,000 for each class), ($iii$) with 200,000 PGGAN-based DA, and ($iv$) with both 200,000 classical DA and 200,000 PGGAN-based DA to confirm the effect of PGGAN-based DA; the classical DA adopts a random combination of horizontal/vertical flipping, rotation up to 10 degrees, width/height shift up to 8%, shearing up to 8%, zooming up to 8%, and constant filling of points outside the input boundaries.

## 3. RESULTS

### 3.1. MR Images Generated by PGGANs

Fig. 2 illustrates examples of synthetic tumor/non-tumor images by PGGANs. In our visual confirmation, for about 75% of cases, PGGANs successfully captures the T1c-specific texture and the appearance of the tumors.

### 3.2. Tumor Detection Results

Table 1 shows the classification results using ResNet-50 for detecting brain tumors. Excpet the 5th row, we remove failure cases and only use success cases. When the PGGAN-based DA, especially after refinement, is combined with classical DA, the accuracy increases significantly.

## 4. CONCLUSION

PGGANs can generate $256 \times 256$ realistic brain MR images and achieve better performance in tumor detection.

## 5. REFERENCES

[1] C. Han, H. Hayashi, L. Rundo, et al., GAN-based synthetic brain MR image generation, Proc. IEEE International Symposium on Biomedical Imaging (ISBI 2018), pp. 734–738, Washington D.C., The US, 2018.

[2] C. Han, L. Rundo, R. Araki, et al., Infinite Brain Tumor Images: Can GAN-based Data Augmentation Improve Tumor Detection on MR Images?, Proc. Meeting on Image Recognition and Understanding (MIRU 2018), Sapporo, Japan, 2018.

[3] T. Karras, T. Aila, S. Laine, et al., Progressive Growing of GANs for Improved Quality, Stability, and Variation, Proc. International Conference on Learning Representations (ICLR), arXiv preprint arXiv:1710.10196, 2018.