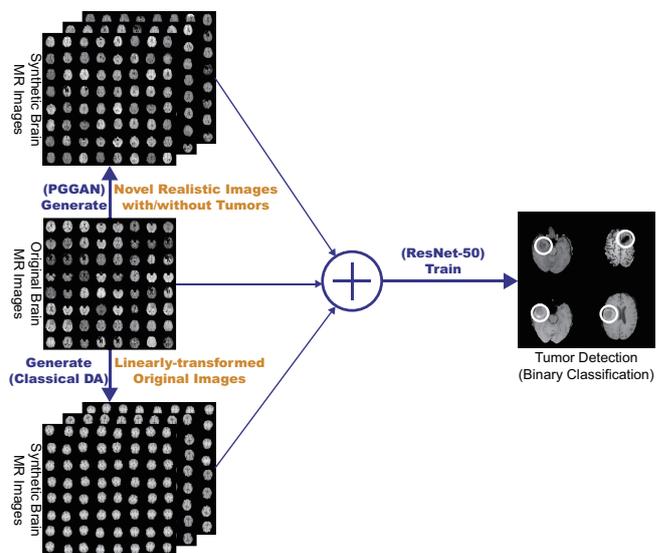# Infinite Brain Tumor Images:
Can GAN-based Data Augmentation
Improve Tumor Detection on MR Images?

Changhee HAN[1,a)]   Leonardo RUNDO[2,b)]   Ryosuke ARAKI[3,c)]   Yujiro FURUKAWA[4,d)]
Giancarlo MAURI[2,e)]   Hideki NAKAYAMA[1,f)]   Hideaki HAYASHI[5,g)]

Abstract

Due to the lack of available annotated medical images, accurate computer-assisted diagnosis requires intensive Data Augmentation (DA) techniques; however, those transformed images intrinsically have a similar distribution to the original ones, leading to limited performance improvement. 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). This study exploits Progressive Growing of GANs (PGGANs), a multi-stage generative training method, to generate original-sized 256×256 MR images for convolutional neural network-based brain tumor detection, which is challenging *via* conventional GANs; difficulties arise due to unstable GAN training with high resolution and a variety of tumors in size, location, shape, and contrast. Our preliminary results show that this novel PGGAN-based DA method can achieve promising performance improvement, when combined with classical DA, in tumor detection and also in other medical imaging tasks.

1   Graduate School of Information Science and Technology, The University of Tokyo, Tokyo, Japan
2   Department of Informatics, Systems and Communication, University of Milano-Bicocca, Milan, Italy
3   Graduate School of Engineering, Chubu University, Aichi, Japan
4   Kanto Rosai Hospital, Kanagawa, Japan
5   Department of Advanced Information Technology, Kyushu University, Fukuoka, Japan
a)   han@nlab.ci.i.u-tokyo.ac.jp
b)   leonardo.rundo@disco.unimib.it
c)   ryorsk@mprg.cs.chubu.ac.jp
d)   juliusdamirdomaplus@yahoo.co.jp
e)   mauri@disco.unimib.it
f)   nakayama@ci.i.u-tokyo.ac.jp
g)   hayashi@ait.kyushu-u.ac.jp

Fig. 1   PGGAN-based DA for better tumor detection: the PGGANs method generates a number of realistic brain tumor/non-tumor MR images and the binary classifier uses them as additional training data.

1.   Introduction

Convolutional Neural Networks (CNNs) have dramatically improved medical image analysis, such as brain Magnetic Resonance Imaging (MRI) segmentation [1], primarily thanks to large-scale annotated training data. Unfortunately, obtaining such massive medical data is challenging; consequently, better training requires intensive Data Augmentation (DA) techniques, such as geometry/intensity transformations of original images [2].

However, those transformed images intrinsically have a similar distribution with the original ones, leading to limited performance improvement; thus, generating realistic (i.e., similar to the real image distribution) but completely new samples is essential to fill the real image distribution uncovered by the original dataset. In this context, Gener-

ative Adversarial Network (GAN)-based DA is promising, as it has shown excellent performances in computer vision, revealing good generalization ability, such as drastic improvement in eye-gaze estimation using SimGAN [3].

Also in medical imaging, realistic retinal image and Computed Tomography (CT) image generation have been tackled using adversarial learning [4], [5]; a very recent research reported performance improvement with synthetic training data in CNN-based liver lesion classification, using a small number of 64×64 CT images for GAN training [6]. However, GAN-based image generation using MRI, the most effective modality for soft-tissue acquisition, has not yet been reported due to the difficulties from low-contrast MR images, strong anatomical consistency, and intra-sequence variability; in our previous work [7], we generated $64 \times 64/128 \times 128$ MR images using conventional GANs and even an expert physician failed to accurately distinguish between the real/synthetic images.

So, how can we generate highly-realistic and original-sized 256×256 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—the most commonly used sequence in tumor detection thanks to its high contrast—for CNN-based tumor detection. This 256×256 image generation is extremely challenging: (*i*) GAN training is unstable with high-resolution inputs and severe artifacts appear due to strong consistency in brain anatomy; (*ii*) brain tumors vary in size, location, shape, and contrast. However, it is beneficial, because most CNN architectures adopt around 256×256 input sizes (e.g., InceptionResNetV2: 299×299, ResNet-50: 224×224 [10]) and we can obtain better performance with original-sized image augmentation—towards this, we use Progressive Growing of GANs (PGGANs) [8], a multi-stage generative training method. Using the synthetic images, our novel PGGAN-based medical DA method achieves better performance in CNN-based tumor detection, when combined with classical DA (Fig. 1).

Contributions. Our main contributions are:

- **MR Image Generation:** This research explains how to exploit MR Images to generate realistic and original-sized 256×256 MR images using PGGANs, while maintaining clear tumor/non-tumor features.
- **MR Image Augmentation:** The proposed study shows encouraging results on PGGAN-based DA, when combined with classical DA, for better brain tumor detection and other medical imaging tasks.
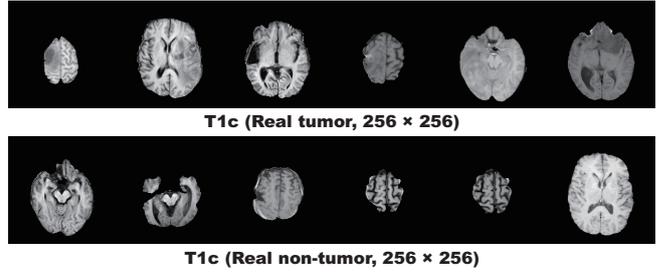

T1c (Real tumor, 256 × 256)

T1c (Real non-tumor, 256 × 256)

Fig. 2 Example real MR images used for PGGAN training.

2. Materials and Methods

2.1 The BRATS 2016 Training Dataset

This paper exploits a dataset of 240×240 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 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, these labels are often incorrect/ambiguous on 2D slices; so, we discard (*i*) tumor images tagged as non-tumor, (*ii*) non-tumor images tagged as tumor, (*iii*) unclear boundary images, and (*iv*) too small/big 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 training set's images are zero-padded to reach a power of 2, 256×256 from 240×240 pixels for better PGGAN training. Fig. 2 shows example real MR images. **PGGANs** is a novel training method for GANs with a progressively growing generator and discriminator [8]: starting from low resolution, newly added layers model fine-grained details as training progresses. As Fig. 3 shows, we adopt PGGANs to generate highly-realistic and original-sized 256×256 brain MR images; tumor/non-tumor images are separately trained and generated. **PGGAN Implementation Details.** We use the PGGAN architecture with the Wasserstein loss using gradient penalty [9]. Training lasts for 100 epochs with a batch size of 16 and 1.0×10^{-3} learning rate for Adam optimizer.
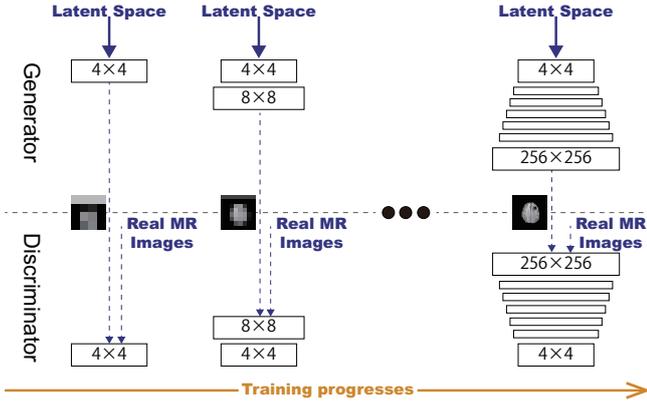
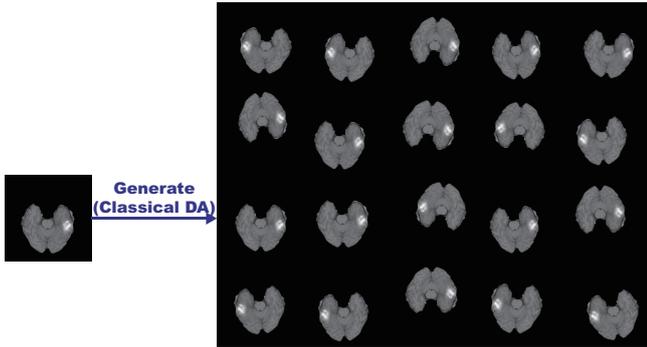**Fig. 3** PGGAN architecture for MR image generation.

Fig. 4 Example real MR image and its geometrically-transformed synthetic images.

2.3 Tumor Detection Using ResNet-50

Pre-processing. To fit ResNet-50's input size, we center-crop all images from 240×240 to 224×224 pixels.

ResNet-50 is a residual learning-based CNN with 50 layers [10]: unlike conventional learning unreferenced functions, it reformulates the layers as learning residual functions for sustainable and easy training. We adopt ResNet-50 to detect tumors in brain MR images, i.e., the binary classification of images with/without tumors.

To confirm the effect of PGGAN-based DA, the following 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; 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 (Fig. 4). For better DA, highly-unrealistic PGGAN-generated images are manually discarded.

ResNet-50 Implementation Details. We use the ResNet-50 architecture with a dropout of 0.5 before the final softmax layer, along with a batch size of 192, 1.0×10^{-3} learning rate for Adam, and early stopping of 10 epochs.
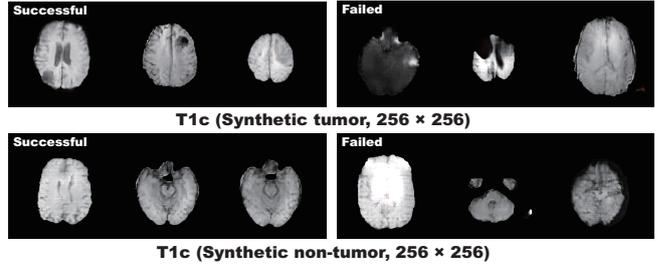

Fig. 5 Example synthetic MR images yielded by PGGANs: (a) successful cases; (b) failed cases.

3. Results

This section shows how PGGANs generates synthetic brain MR images. The results include instances of synthetic images and their influence on tumor detection.

3.1 MR Images Generated by PGGANs

Fig. 5 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 tumor appearance while maintaining the realism of the original brain MR images; however, for about 25% of cases, the generated images lack clear tumor/non-tumor features or contain unrealistic features, such as hyper-intensity, gray contours, and odd artifacts.

3.2 Tumor Detection Results

Table 1 shows the classification results for detecting brain tumors with/without DA techniques. As expected, the test accuracy improves by 0.64% with the additional $200,000$ geometrically-transformed images for training. When only the PGGAN-based DA is applied, the test accuracy decreases drastically with almost 100% of sensitivity and 6.84% of specificity, because the classifier recognizes the synthetic images' prevailed unrealistic features as tumors, similarly to anomaly detection.

However, surprisingly, when it is combined with the classical DA, the accuracy increases by 1.02% with higher sensitivity and specificity; probably, this occurs because the PGGAN-based DA fills the real image distribution uncovered by the original dataset, while the classical DA provides the robustness on training for most cases.

4. Conclusion

Our preliminary results show that PGGANs can generate original-sized 256×256 realistic brain MR images and achieve higher performance in tumor detection, when

Table. 1 Binary classification results for detecting brain tumors with/without DA.

	Accuracy (%)	Sensitivity	Specificity
ResNet-50 (w/o DA)	90.06	85.27	97.04
ResNet-50 (w/ 200k classical DA)	90.70	88.70	93.62
ResNet-50 (w/ 200k PGGAN-based DA)	62.02	**99.94**	6.84
ResNet-50 (w/ 200k classical DA + 200k PGGAN-based DA)	**91.08**	86.60	**97.60**

combined with classical DA. This occurs because PG-GANs' multi-stage image generation achieves good generalization and synthesizes images with the real image distribution unfilled by the original dataset. However, considering the Visual Turing Test results, yet unsatisfactory realism with high resolution strongly limit DA performance, so we plan to (*i*) discard unrealistic images or (*ii*) generate only realistic images, and then (*iii*) refine synthetic images more similar to the real image distribution.

For (*i*), classifier two-sample tests [11], assessing whether two samples are drawn from the same distribution, can help discard images not from the real image distribution, as manual removal is demanding. Regarding (*ii*), we can map an input random vector onto each training image [12] and generate images with suitable vectors, to control the divergence of generated images; virtual adversarial training [13] can be also integrated to control the output distribution. Lastly, (*iii*) can be achieved by GAN-based image-to-image translation, such as CycleGAN [14], considering SimGAN's remarkable performance improvement after refinement [3]. Moreover, we should further avoid real images with ambiguous/inaccurate annotation for better tumor detection.

To evaluate the realism of the generated images, t-Distributed Stochastic Neighbor Embedding (t-SNE) [15] can visualize the distribution of tumor/non-tumor images by extracting the features from the last layer of a trained CNN. Overall, our novel PGGAN-based DA approach sheds light on diagnostic and prognostic medical applications, not limited to tumor detection; future studies are needed to extend our encouraging results.

References

[1] M. Havaei, A. Davy, D. Warde-Farley, et al., Brain tumor segmentation with deep neural networks, Med. Image Anal., vol. 35, pp. 18-31, Jan.2017.

[2] O. Ronneberger, P. Fischer, and T. Brox, U-Net: Convolutional networks for biomedical image segmentaion, Proc. Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp. 234-241, Munich, Germany, Oct.2015.

[3] A. Shrivastava, T. Pfister, O. Tuzel, et al., Learning from simulated and unsupervised images through adversarial training, Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), pp. 2107-2116, Hawaii, The US, July.2017.

[4] P. Costa, A. Galdran, M. I. Meyer, et al., End-to-End Adversarial Retinal Image Synthesis, IEEE Transactions on Medical Imaging, vol. 37, no. 3, pp. 781–790, March 2018.

[5] M. J. M. Chuquicusma, S. Hussein, J. Burt, et al., How to Fool Radiologists with Generative Adversarial Networks? A Visual Turing Test for Lung Cancer Diagnosis, Proc. IEEE International Symposium on Biomedical Imaging (ISBI), Washington, D.C., The US, Apr.2017.

[6] M. Frid-Adar, I. Diamant, E. Klang, et al., GAN-based Synthetic Medical Image Augmentation for increased CNN Performance in Liver Lesion Classification, arXiv preprint arXiv:1803.01229, 2018.

[7] C. Han, H. Hayashi, L. Rundo, et al., GAN-based Synthetic Brain MR Image Generation, Proc. IEEE International Symposium on Biomedical Imaging (ISBI), Washington, D.C., The US, Apr.2017.

[8] 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.

[9] I. Gulrajani, F. Ahmed, M. Arjovsky, et al., Improved Training of Wasserstein GANs, Proc. Conf. on Neural Information Processing Systems (NIPS), pp. 5767–5777, California, The US, Dec.2017.

[10] K. He, X. Zhang, S. Ren, et al., Deep Residual Learning for Image Recognition, Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), pp. 770–778, Nevada, The US, June.2016.

[11] D. Lopez-Paz and M. Oquab, Revisiting classifier two-sample tests, Proc. International Conference on Learning Representations (ICLR), arXiv preprint arXiv:1610.06545, 2017.

[12] T. Schlegl, P. Seebock, S. M. Waldste, et al., Unsupervised Anomaly Detection with Generative Adversarial Networks to Guide Marker Discovery, Proc. Int. Conf. on Information Processing in Medical Imaging (IPMI), pp. 146–157, North Carolina, The US, June.2017.

[13] T. Miyato, S. Maeda, M. Koyama, et al., Virtual Adversarial Training: a Regularization Method for Supervised and Semi-supervised Learning, Proc. International Conference on Learning Representations (ICLR), arXiv preprint arXiv:1704.03976, 2016.

[14] J. Zhu, T. Park, P. Isola, et al., Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks, Proc. Int. Conf. on Computer Vision (ICCV), pp. 2223–2232, Venice, Italy, Oct.2017.

[15] L. Maaten and G. Hinton, Visualizing Data using t-SNE, Journal of Machine Learning Research (JMLR), vol. 9, pp. 2579–2605, November 2018.