Learning More with Less:
Conditional PGGAN-based MRI Augmentation with Highly-Rough Annotation for Brain Metastases Detection

CHANGHEE HAN\textsuperscript{1,2,3,a)} Kohei MURA\textsuperscript{3} Tomoyuki NOGUCHI\textsuperscript{3} Yusuke KAWATA\textsuperscript{3} Fumiya UCHIYAMA\textsuperscript{3} Leonardo RUNDO\textsuperscript{4} Hideki NAKAYAMA\textsuperscript{2} Shin’ichi SATOH\textsuperscript{1}

Abstract

Accurate computer-assisted diagnosis can alleviate the risk of overlooking the diagnosis in a clinical environment. Towards this, as a Data Augmentation (DA) technique, Generative Adversarial Networks (GANs) can synthesize additional training data to handle the small/fragmented medical imaging datasets collected from various scanners; those images are realistic but completely different from the original ones, filling the data lack in the real image distribution. However, we cannot easily use them to locate disease areas, considering expert physicians’ expensive annotation cost. Therefore, we propose Conditional Progressive Growing of GANs (CPGGANs), incorporating highly-rough bounding box conditions incrementally into PGGANs to place brain metastases at desired positions/sizes on $256 \times 256$ Magnetic Resonance (MR) images, for Convolutional Neural Network-based tumor detection; this first GAN-based medical DA using automatic bounding box annotation improves the training robustness. The results show that CPGGAN-based DA can boost 10% sensitivity in diagnosis with clinically acceptable additional false positives. Surprisingly, further tumor realism, achieved with additional normal brain MR images for CPGGAN training, does not contribute to detection performance, while even three physicians cannot accurately distinguish them from the real ones in Visual Turing Test.

1. Introduction

Accurate computer-assisted diagnosis with high sensitivity can alleviate the risk of overlooking the diagnosis in a clinical environment. Specifically, Convolutional Neural Networks (CNNs) have revolutionized medical imaging, such as diabetic eye disease diagnosis [1], mainly thanks to large-scale annotated training data. However, obtaining such annotated medical big data is demanding; thus, better diagnosis requires intensive Data Augmentation (DA) techniques, such as geometric/intensity transformations of original images [2]. Yet, those augmented images intrinsically have a similar distribution to the original ones, leading to limited performance improvement; in this context, Generative Adversarial Network (GAN) [3]-based DA can boost the performance by filling the real image distribution uncovered by the original dataset, since it generates realistic but completely new samples showing good generalization ability.

In medical imaging, where the primary problem lies in small and fragmented imaging datasets from various scanners, GAN-based DA performs effectively: researchers improved classification by augmentation with noise-to-image GANs (e.g., random noise samples to diverse pathological images) [4] and segmentation with image-to-image GANs (e.g., a benign image with a pathology-conditioning image to a malignant one) [5]. Nevertheless, unlike bounding box-based object detection, simple classification cannot locate disease areas and rigorous segmentation requires physicians’ expensive annotation.

So, how can we achieve high sensitivity in diagnosis using GANs with minimum annotation cost, based on highly-rough and inconsistent bounding boxes? We aim to generate GAN-based realistic/diverse $256 \times 256$ brain MR images with brain metastases at desired positions/sizes based on bounding boxes; the object detector uses them as additional training data.

Fig. 1: CPGGAN-based DA for better tumor detection: our CPGGANs generates a number of realistic brain MR images with tumors at desired positions/sizes based on bounding boxes; the object detector uses them as additional training data.

\textsuperscript{1} Research Center for Medical Big Data, National Institute of Informatics
\textsuperscript{2} Graduate School of Information Science and Technology, The University of Tokyo
\textsuperscript{3} Department of Radiology, National Center for Global Health and Medicine
\textsuperscript{4} Department of Radiology, University of Cambridge
\textsuperscript{a)} han@slab.ci.i.u-tokyo.ac.jp

\textsuperscript{CONFIDENTIAL EXTENDED ABSTRACT. DO NOT DISTRIBUTE ANYWHERE.}
Regions of Interest (ROIs) alone, however, could facilitate detection because it provides more image details and most CNN architectures adopt around $256 \times 256$ input pixels. Therefore, as a conditional noise-to-image GAN not relying on an input benign image, we propose Conditional Progressive Growing of GANs (CPGGANs), incorporating highly-rough bounding box conditions incrementally into PGGANs [6] to naturally place tumors of random shape at desired positions/sizes on MR images. Moreover, we evaluate the generated images’ realism via Visual Turing Test [7] by three expert physicians, and visualize the data distribution via t-Distributed Stochastic Neighbor Embedding (t-SNE) algorithm [8]. Using the synthetic images, our novel CPGGAN-based DA boosts 10% sensitivity in diagnosis with clinically acceptable additional False Positives (FPs). Surprisingly, we confirm that further realistic tumor appearance, judged by the physicians, does not contribute to detection performance. 

**Contributions.** Our main contributions are as follows:

- **Conditional Image Generation:** This first bounding box-based $256 \times 256$ whole pathological image generation shows that CPGGANs can generate realistic/diverse images with objects naturally at desired positions/sizes for clinical applications in oncological imaging, such as DA, data anonymization, and physician training.

- **Misdiagnosis Prevention:** This study allows to achieve high sensitivity in automatic computer-assisted diagnosis using small/fragmented medical imaging datasets from various scanners with minimum annotation labor based on highly-rough/inconsistent bounding boxes.

- **Brain Metastases Detection:** This first bounding box-based brain metastases detection method successfully detects tumors exploiting CPGGAN-based DA.

2. **Materials and Methods**

2.1 **Brain Metastases Dataset**

For robust clinical applications, we exploit a dataset of T1c brain axial MR images collected by the authors, containing 180 brain metastases cases from multiple MRI scanners—those images differ in contrast, magnetic field strength (i.e., 1.5 T, 3.0 T), and matrix size (i.e., $190 \times 224, 216 \times 256, 256 \times 256, 460 \times 460$ pixels). We also use additional brain images from 193 normal subjects only for CPGGAN training, not in tumor detection.

2.2 **Proposed CPGGAN-based Image Generation**

**Data Preparation** For tumor detection, our brain tumor dataset (180 patients) is divided into: (i) a training set (126 patients); (ii) a validation set (18 patients); (iii) a test set (30 patients):

- **Training set:** (2, 813 images/5, 963 bounding boxes);
- **Validation set:** (337 images/616 bounding boxes);
- **Test set:** (947 images/934 bounding boxes).

Only the training set is used for CPGGAN training to be fair. The training set is relatively small/fragmented for CNN-based detection, considering the very similar appearances of the same patient’s tumor slices. To confirm the effects of realism/diversity—provided by combining PGGANs and bounding box conditioning—on tumor detection, we compare the following GANs: (i) CPGGANs trained only with the brain metastases images; (ii) CPGGANs trained also with additional 16, 962 brain images from 193 normal subjects; (iii) Image-to-image GAN trained only with the brain metastases images. After skull-stripping on all images with various resolution, remaining brain parts are cropped and resized to $256 \times 256$ pixels (i.e., a power of 2 for better CPGGAN training). We lazily annotate tumors with rough bounding boxes to minimize physicians’ labor (Fig. 2).

CPGGANs is a conditional noise-to-image GAN training method, incorporating rough bounding box conditions incrementally into PGGANs [6]. The original PGGANs progressively grow a generator/discriminator: starting from low resolution, new layers model details as training progresses. As Fig. 3 shows, to generate realistic/diverse $256 \times 256$ MR images with tumors at desired positions/sizes using only bounding boxes without an input benign image under limited training data/highly-rough annotation, we further condition the generator/discriminator:

- **Conditioning input:** prepare a $256 \times 256$ black image (i.e., pixel value: 0) with white bounding boxes (i.e., 255) describing tumor positions/sizes for attention;

- **Generator input:** resize the conditioning image to the previous generator’s output resolution/channel size and concatenate them (the first $4 \times 4$ images are generated from noises);

- **Discriminator input:** concatenate the conditioning image with a real or synthetic image.
CPGGAN Implementation Details We use the CPGGAN architecture with the Wasserstein loss using gradient penalty. Training lasts for 3,000,000 steps with a batch size of 4 and $2.0 \times 10^{-4}$ learning rate for Adam optimizer. During testing, as tumor attention images, we use training image annotation with a random combination of horizontal/vertical flipping, width/height shift up to 10%, and zooming up to 10%; these generated images are used as additional training images for tumor detection.

Image-to-image GAN is a conventional conditional GAN with a U-Net-like generator with 4 convolutional/deconvolutional layers in encoders/decoders, along with a discriminator with 3 decoders.

2.3 Brain Metastases Detection Using YOLOv3

YOLOv3 [9] is a fast/accurate CNN-based object detector. We adopt YOLOv3 to detect brain metastases on MR images since its high efficiency can play a clinical role for real-time tumor alert.

To confirm the effect of GAN-based DA, the following detection results are compared: (i) 2,813 real images without DA, (ii), (iii), (iv) with 4,000,000/8,000/12,000 CPGGAN-based DA, (v), (vi), (vii) with 4,000,000/8,000/12,000 CPGGAN-based DA, trained with additional normal brain images, (viii), (ix), (x) with 4,000,000/8,000/12,000 image-to-image GAN-based DA. Due to the risk of overdiagnosing the diagnosis, higher sensitivity matters more than less FPs; thus, we aim to maximize sensitivity with clinically acceptable FPs, adding the additional synthetic training images. Since our annotation is highly-rough, we calculate sensitivity/FPs per slice with both Intersection over Union (IoU) threshold 0.5 and 0.25. For better DA, GAN-generated images with unclear tumor appearance are manually discarded.

YOLOv3 Implementation Details We use the YOLOv3 with Darknet-53. During training, we use a batch size of 64, $1.0 \times 10^{-3}$ learning rate for Adam optimizer, and classic DA of geometric/intensity transformations. The network resolution is set to 416 $\times$ 416 pixels during training and 608 $\times$ 608 pixels during validation/testing to detect small tumors better. For testing, we pick the model with the best sensitivity on validation with detection threshold 0.1%/IoU threshold 0.5 between 96,000-240,000 steps to avoid severe FPs while achieving high sensitivity.

2.4 Clinical Validation Using Visual Turing Test

To quantitatively evaluate how realistic the CPGGAN-based synthetic images are, we supply, in a random order, to three expert physicians a random selection of 50 real/50 synthetic tumor images. They take four tests in ascending order: (i), (ii) test1, 2: real vs. CPGGAN-generated resized 32 $\times$ 32 tumor bounding boxes, trained without/with additional normal brain images; (iii), (iv) test3, 4: real vs. CPGGAN-generated 256 $\times$ 256 MR images, trained without/with additional normal brain images. Then, the physicians are asked to constantly classify them as real/synthetic.

2.5 Visualization Using t-SNE

To visually analyze the distribution of real/synthetic 32 $\times$ 32 resized tumor bounding box and 256 $\times$ 256 whole MR images, we use t-SNE [8] on a random selection of:

- 500 real tumor images;
- 500 CPGGAN-generated tumor images;
- 500 CPGGAN-generated tumor images, trained with additional normal brain images.

Table 1: YOLOv3 brain metastases detection results with/without DA, using bounding boxes with 0.1% detection threshold.

<table>
<thead>
<tr>
<th>Detection Method</th>
<th>Sensitivity</th>
<th>FPs per slice</th>
<th>Sensitivity</th>
<th>FPs per slice</th>
</tr>
</thead>
<tbody>
<tr>
<td>2,813 real images</td>
<td>0.67</td>
<td>4.11</td>
<td>0.63</td>
<td>3.59</td>
</tr>
<tr>
<td>+ 4,000 CPGGAN-based DA</td>
<td>0.77</td>
<td>7.64</td>
<td>0.91</td>
<td>7.18</td>
</tr>
<tr>
<td>+ 8,000 CPGGAN-based DA</td>
<td>0.71</td>
<td>6.36</td>
<td>0.87</td>
<td>5.85</td>
</tr>
<tr>
<td>+ 12,000 CPGGAN-based DA</td>
<td>0.78</td>
<td>11.77</td>
<td>0.91</td>
<td>11.29</td>
</tr>
<tr>
<td>+ 4,000 Image-to-Image GAN-based DA</td>
<td>0.69</td>
<td>7.16</td>
<td>0.86</td>
<td>5.61</td>
</tr>
<tr>
<td>+ 8,000 Image-to-Image GAN-based DA</td>
<td>0.73</td>
<td>8.10</td>
<td>0.89</td>
<td>7.59</td>
</tr>
<tr>
<td>+ 12,000 Image-to-Image GAN-based DA</td>
<td>0.74</td>
<td>9.42</td>
<td>0.89</td>
<td>8.95</td>
</tr>
<tr>
<td>+ 4,000 Image-to-Image GAN-based DA (+ normal)</td>
<td>0.72</td>
<td>6.21</td>
<td>0.87</td>
<td>5.76</td>
</tr>
<tr>
<td>+ 8,000 Image-to-Image GAN-based DA (+ normal)</td>
<td>0.68</td>
<td>3.80</td>
<td>0.84</td>
<td>2.99</td>
</tr>
<tr>
<td>+ 12,000 Image-to-Image GAN-based DA (+ normal)</td>
<td>0.74</td>
<td>7.20</td>
<td>0.89</td>
<td>6.72</td>
</tr>
</tbody>
</table>

3. Results

3.1 MR Images Generated by CPGGANs

Fig. 4 shows example GAN-generated MR images. CPGGANs successfully captures the T1c-specific texture/tumor appearance at desired positions/sizes. Due to rough bounding boxes, the generated tumors’ shapes largely vary within the bounding boxes. When trained with additional normal brain images, it clearly maintains the realism with less odd artifacts, including tumor bounding boxes, which the additional images do not include. But as expected, image-to-image GAN, without progressive growing, generates completely unrealistic images without an input benign image due to the limited training data/highly-rough annotation.

3.2 Brain Metastases Detection Results

Table 1 shows the tumor detection results with/without GAN-based DA. As expected, the sensitivity remarkably increases with the additional synthetic training data while FPs per slice also in-
Fig. 5: Example detection results by the seven DA setups on four images, compared against the ground truth: (a) ground truth; (b) without GAN-based DA; (c), (d), (e) with 4k/8k/12k CPGGAN-based DA; (f), (g), (h) with 4k/8k/12k CPGGAN-based DA, trained with additional normal brain images. Red V symbols indicate the tumors undetected without GAN-based DA, but detected with 4k CPGGAN-based DA.

Table 2: Visual Turing Test results by three physicians for classifying real vs synthetic images: (a), (b) test1, 2: real vs CPGGAN-generated resized 32 × 32 tumor bounding boxes, trained without/with additional normal brain images; (c), (d) test3, 4: real vs CPGGAN-generated 256 × 256 MR images, trained without/with additional normal brain images. Proximity to 50% of accuracy indicates superior performance.

Table 2: Visual Turing Test results by three physicians for classifying real vs synthetic images: (a), (b) test1, 2: real vs CPGGAN-generated resized 32 × 32 tumor bounding boxes, trained without/with additional normal brain images; (c), (d) test3, 4: real vs CPGGAN-generated 256 × 256 MR images, trained without/with additional normal brain images. Proximity to 50% of accuracy indicates superior performance.

### Acknowledgments

This research was supported by AMED Grant Number JP18ik1010028.

### References


