Learning More with Less: GAN-based Medical Image Augmentation

Changhee HAN*1,2, Kohei MURAO*2, Shin’ichi SATOH*2, Hideki NAKAYAMA*1

Abstract

Accurate computer-assisted diagnosis using Convolutional Neural Networks (CNNs) requires large-scale annotated training data, associated with expert physicians’ time-consuming labor; thus, Data Augmentation (DA) using Generative Adversarial Networks (GANs) is essential in Medical Imaging, since they can synthesize additional annotated training data to handle small and fragmented medical images from various scanners; those images are realistic but completely different from the original ones, filling the data lack in the real image distribution. As a tutorial, this paper introduces background on GAN-based Medical Image Augmentation, along with tricks to achieve high classification/object detection/segmentation performance using them, based on our empirical experience and related work. Moreover, we show our first GAN-based DA work using automatic bounding box annotation, for robust CNN-based brain metastases detection on 256 × 256 MR images; GAN-based DA can boost 10% sensitivity in diagnosis with a clinically acceptable amount of additional False Positives, even with highly-rough and inconsistent bounding boxes.

Keywords: Generative Adversarial Networks, Data Augmentation, Medical Image Augmentation

1. Introduction

Convolutional Neural Networks (CNNs) have revolutionized medical image analysis, including brain Magnetic Resonance Imaging (MRI) segmentation, primarily thanks to large-scale annotated training data. However, obtaining such a huge amount of annotated medical data is demanding; thus, accurate diagnosis requires intensive Data Augmentation (DA) techniques, such as geometric/intensity transformations of original images. Yet, such augmented images intrinsically have a similar distribution to the original ones, leading to limited performance improvement; in this context, Generative Adversarial Network (GAN) [1]-based DA can boost the performance by filling the real image distribution uncovered by the original dataset, since it synthesizes realistic but completely novel samples showing good generalization ability; GANs showed outstanding performance in Computer Vision, including 21% performance improvement in eye-gaze estimation [2]. This GAN-based DA trend especially applies to Medical Imaging for handling various types of small/fragmented medical datasets from multiple scanners: some researchers used it for classification on brain tumor MR [3] and skin lesion images [4]; the others used GAN-based Region of Interest (ROI) DA for segmentation on 3D lung nodule CT images [5].

2. Tricks to Learn More with Less

1) Pre-Processing
Since variety in size, location, shape, and visual appearance heavily influences GAN performance especially when training data are small and fragmented, as usually in Medical Imaging, we should remove irrelevant information on medical images (e.g., denoising/skull-stripping) and then crop/resize the remaining parts to a power of 2 (e.g., \(256 \times 256/32 \times 32 \times 32\)) for better GAN training; the Spatial Pyramid Pooling (SPP) layer in the discriminator can be used to avoid the effect of resizing [6]. Classical DA techniques (i.e., geometric/intensity transformations) may help GAN training (at least horizontal and vertical flipping), along with Webbly Supervised Learning [7] for external data.

2) **GAN-based Image Generation**

We can either generate whole images including pathological parts (e.g., whole brain MR images with tumors) [1] or the Regions of Interest (ROIs) alone (e.g., skin lesion images) [4] for 2D image generation, but only ROIs alone (e.g., 3D lung CT nodules) [5] for 3D image generation due to heavy computation power—–if needed, we can paste the generated ROIs naturally to the whole images for DA. Moreover, Progressive Growing of GANs (PGGANs) [8] with the Wasserstein loss using gradient penalty [9] can generate a variety of realistic high-resolution (i.e., \(\geq 256 \times 256\)) medical images. From the latent space, whether to sample points using a normal or uniform distribution does not empirically make significant difference to DA.

3) **Post-processing**

Removing GAN-generated images with weird artifacts does not boost DA, since realism confirmed by humans, including via Visual Turing Test [10] by expert physicians [11], is not strongly associated with DA performance.

4) **Image-quality Evaluation on GAN-generated Images**

\(t\)-Distributed Stochastic Neighbor Embedding (\(t\)-SNE) algorithm [12] can visualize the distribution of real and synthetic images by directly embedding those high-dimensional data into a 2D/3D space. Visual Turing Test by physicians can also quantitatively evaluate how realistic or certain disease-like the GAN-based synthetic images are by supplying, in a random order, a random selection of the same number of real/synthetic images [11] per each class.

5) **Application to Classification**

We can use GAN-generated images for classification, generating normal vs. pathological images (e.g., non-tumor vs. tumor) [3] or several types of pathological images [4]. Some studies report better DA performance when generating labeled examples for each pathological class separately, rather than incorporating class conditioning to generate labeled examples all at once [4]. Balancing between real and synthetic images is essential during classifier training, since adding over-sufficient synthetic images leads to worse performance [3]; we can either classically augment only real images or both real/synthetic images (empirically, the best balance between real and synthetic images is 1-to-1 or 1-to-2). Pre-training on ImageNet might not achieve better sensitivity than training from scratch. We can also use \(t\)-SNE to differentiate classification results with/without DA by visualizing features extracted from the last layer of the trained classifier.

6) **Application to Object Detection**

For minimizing time-expensive expert physicians’ annotation tasks, we can also exploit GAN-generated images for Object Detection by incorporating bounding box conditions into GANs during training and generating images based on the annotation slightly different from training images using a random combination of classical DA during testing.
Adding additional training images can achieve higher sensitivity with an acceptable amount of False Positives (FPs). The next section describes our novel GAN-based DA work using bounding boxes [13].

7) Application to Segmentation

Similarly to Object Detection, we can condition rigorous segmentation into GANs to achieve higher Dice score.

3. GAN-based MR Image Augmentation for Brain Metastases Detection

1) Brain Metastases Dataset

This paper exploits a small/fragmented dataset of T1c brain axial MR images collected by the authors, containing 180 brain metastases cases from multiple MRI scanners. For tumor detection, our whole brain metastases dataset (180 patients) is divided into: (i) a training set (126 patients); (ii) a validation set (18 patients); (iii) a test set (36 patients); only the training set is exploited for our GAN training to be fair. Our experimental dataset consists of:

- Training set (2,813 images/5,963 bounding boxes);
- Validation set (337 images/616 bounding boxes);
- Test set (947 images/3,094 bounding boxes).

After skull-stripping on all images with diverse resolution, remaining brain parts are cropped and resized to 256 × 256 pixels. In our first GAN-based medical DA for object detection, we lazily annotate tumors with highly-rough and inconsistent bounding boxes to minimize expert physicians' labor.

2) Proposed GAN-based Image Generation

We are considering an innovative training method for GANs, incorporating bounding box conditions into PGGANs [8]. The original PGGANs exploits a progressively growing generator and discriminator: starting from low resolution, newly-added layers model fine-grained details as training progresses. Fig. 1 shows example real and GAN-generated images, including tumor bounding boxes.

3) Brain Metastases Detection Using YOLOv3

YOLOv3 [14] is a fast and accurate CNN-based object detector, dividing the image into regions and predicts bounding boxes/probabilities for each region. We detect brain metastases on MR images using YOLOv3.

To confirm the influence of GAN-based DA, the following detection results are compared: (i) 2,813 real images without DA, (ii) with 4,000 GAN-based DA. Due to the risk of overlooking the diagnosis via computer-assisted diagnosis, higher sensitivity matters more than a lower amount of FPs; thus, we aim to achieve higher sensitivity with a clinically acceptable amount of FPs, adding the additional synthetic training images. Since our annotation is highly-rough, we calculate sensitivity/FPs per slice with both IoU threshold 0.5 and 0.25. For better DA, synthetic images with unclear tumor appearance are manually discarded.

4) Brain Metastases Detection Results

Table 1 indicates the tumor detection results with/without DA. As expected, the mAP/sensitivity remarkably increase with the additional training images, while FPs per slice also increase. Adding synthetic data leads to more FPs, also detecting blood vessels which are small/hyper-intense on T1c MR images, very similarly to the enhanced tumor regions (i.e., the contrast agent is perfused throughout the blood vessels). Adding 4,000 GAN-generated images
Fig. 1 Example real/synthetic $256 \times 256$ MR images and resized $32 \times 32$ tumor bounding boxes.

shows a sensitivity improvement by 0.10 with IoU threshold 0.5 and by 0.08 with IoU threshold 0.25—the improved robustness during training produces sensitivity 0.91 with moderate IoU threshold 0.25 despite our highly-rough bounding box annotation. Fig. 2 also visually shows that it can alleviate the risk of overlooking the tumor diagnosis, with clinically acceptable FPs (i.e., the highly-overlapping bounding boxes only require a physician's single check). It should be noted that we cannot increase FPs to achieve such high sensitivity without GAN-based DA.

4. Conclusion

We introduce background on GAN-based Medical Image Augmentation, along with the tricks to achieve high classification/object detection/segmentation performance using them, based on our experience and related work. As an example, we also show that our novel GAN can generate $256 \times 256$ MR images with brain metastases of random shape, naturally at desired position/size and achieve high sensitivity in tumor detection—even with small/fragmented training data from multiple MRI scanners and lazy annotation using highly-rough bounding boxes.

Acknowledgement

This research was supported by AMED under Grant Number JP18lk1010028. The authors appreciate Dr. Tomoyuki Noguchi for preparing the dataset at the Department of Radiology, National Center for Global Health and Medicine.
Fig. 2 Example detection results obtained by the DA setups on four different images, compared against the ground truth: (a) ground truth; (b) without GAN-based DA; (c) with 4k GAN-based DA. Red V symbols indicate the brain metastases undetected without GAN-based DA, but detected with 4k GAN-based DA.

Table 1 YOLOv3 brain metastases detection results with/without GAN-based DA using bounding boxes with 0.1% detection threshold.

|                   | IoU ≥ 0.5 | IoU ≥ 0.25 |
|-------------------|-----------|------------|
|                   | mAP       | Sensitivity| FPs per slice | Sensitivity| FPs per slice |
| 2,813 real images | 0.51      | 0.67       | 4.11         | 0.83       | 3.59         |
| + 4,000 GAN-based DA | **0.54** | **0.77**   | 7.64         | **0.91**   | 7.18         |

References

[1] Goodfellow I, Pouget-Abadie J, Mirza M et al: Generative adversarial nets. In Advances in Neural Information Processing Systems (NIPS): 2672-2680, 2014

[2] Shrivastava A, Pfister T, Tuzel O et al: Learning from simulated and unsupervised images through adversarial training. In Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR): 2107-2116, 2017

[3] Han C, Rundo L, Araki R et al: Infinite brain tumor images: Can GAN-based data augmentation improve tumor detection on MR images? In Proc. Meeting on Image Recognition and Understanding (MIRU), 2018

[4] Frid-Adar M, Diamant I, Klang E et al: GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. Neurocomputing 321: 321-331, 2018

[5] Jin D, Xu Z, Tang Y et al: CT-realistic lung nodule simulation from 3D conditional generative adversarial networks for robust lung segmentation. In Proc. International Conference on Medical Image Computing and Computer-Assisted
Biography

Changhee Han
Ph.D. Student at Graduate School of Information Science and Technology, The University of Tokyo. He received a Master's Degree in Computer Science from The University of Tokyo. His research interests include Deep Learning for Medical Imaging.

Kohei Murao
Research Associate Professor at Research Center for Medical Bigdata, National Institute of Informatics. He received a Ph.D. Degree in Physics from Tohoku University. His research interests include Medical Image Analysis and Informatics.

Shin’ichi Satoh
Professor at Digital Content and Media Sciences Research Division, National Institute of Informatics. He received a Ph.D. Degree in Computer Science from The University of Tokyo. His research interests include Computer Vision and Pattern Media.

Hideki Nakayama
Associate Professor at Graduate School of Information Science and Technology, The University of Tokyo. He received a Ph.D. Degree in Computer Science from The University of Tokyo. His research interests include Computer Vision and Deep Learning.