

Synthetic Generation of Brain Tumor MRIs

Cameron Giarrusso¹ and Logan Vegna²

Cornell University, Ithaca, NY, U.S.

Corresponding Author Email ID:

fcg26@cornell.edu¹

lnv8@cornell.edu²

Abstract. MRI scans have allowed for great strides in the medical field. Particularly, MRI scans of patients with brain tumors have allowed for the continued research, diagnosis and advanced treatment of benign and cancerous brain tumors. For only neurologists have been able to identify and classify tumors in MRI scans, typically requiring patients to make visits to multiple medical facilities and specialists for a diagnosis. However, recent machine vision research has introduced methods for the automatic detection, and classification of brain tumors in various medical imaging formats. This could potentially reduce the time it takes for a brain tumor diagnosis. Unfortunately due to its private medical nature, large datasets of brain tumor MRIs are scarce. In an attempt to alleviate this issue this paper explores synthetic generation of brain tumor MRI images using Generative Adversarial Networks to supplement the scarce data.

Keywords: MRI: Magnetic Resonance Imaging · GAN: Generative Adversarial Network

1 Introduction

With the strides in computer vision capabilities since 2012, deep learning techniques have been adopted by the medical imaging community with blinding speed. A large reason for this widespread adoption is its potential in image classification and to augment image representation. In this paper, we focus on the application of generative adversarial networks (GANs) as a data augmentation tool, specifically when applied to brain MR images.

1.1 Brain MR Images

MR Images consistently play a vital role as the most common technique used by radiologists when analyzing for brain tumors. [3] With the abundance of medical data, response time for diagnoses became more important than ever. [15] Historically, interpreting and properly segmenting the brain tumor images is a time-consuming and tedious procedure that can take anywhere from 3 - 5 hours. [1, 12] However, it also heavily relies on the expertise of individual radiologists, which is why human raters consistently score a 74% - 85% Dice

score. [13] When compared to automated methods that take anywhere from a few seconds to 20 minutes, it is no wonder that manual segmentation methods have fallen by the wayside as fully automated segmentation methods are far more practical for day-to-day use. [2]

1.2 GANs

Simple GAN A simple GAN is trained to generate samples from a learned distribution without actually modeling the underlying probability density function. It accomplishes this by pitting two neural networks against each other: a generator G and a discriminator D .

Intuitively, the generator G can be thought of as a blacksmith trying to forge counterfeit products and the discriminator D is a quality control officer trying to catch any counterfeits. From this perspective, any counterfeits that make it past the discriminator results in a reward signal to G . This highlights how the information is back propagated from D to G , so after every training iteration G learns how to more effectively fool D . [19]

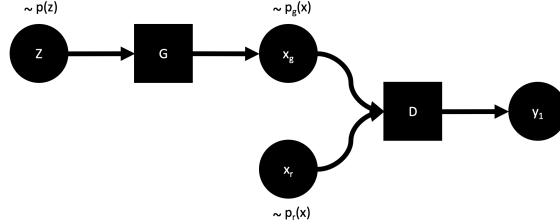


Fig. 1. Simple GAN architecture [19]

The generator takes noise z sampled from some prior distribution $p(z)$ and outputs a fake image X_g . Here θ_g parametrizes the non-linear mapping function learned by G as $X_g = G(z; \theta_g)$. Here it is typically expected that X_g and the real images X_r share some visual similarities. When combined together, these real and generated images are fed into the discriminator one at a time. The input to D is the real or generated image. The output of D is y which is a binary output stating that the image is either real or generated. $y = D(x; \theta_d)$ represents the mapping learned by D as parametrized by θ_d . Thus the generated samples form a distribution $p_g(x)$ which is meant to approximate $p_r(x)$ when trained successfully. Below, Equations 1 and 2 are the loss functions for the simple GAN. [19]

$$\mathcal{L}_D^{GAN} = \max_D E_{x_r \sim p_r(x)} [\log D(x_r)] + E_{x_g \sim p_g(z)} [1 - \log D(x_g)] \quad (1)$$

$$\mathcal{L}_G^{GAN} = \min_G E_{x_g \sim p_g(z)} [1 - \log D(x_g)] \quad (2)$$

While D and G are trained together, there is no guarantee of balance between the two. Thus it's possible for one network to become much more powerful than its counterpart, typically this is D . This imbalance leads a vanishing gradient as X_g is too easy to discern from X_r so the back propagated gradient approaches 0 and G stops being trained. Another common problem is mode collapse, this occurs when the generator G focuses on too narrow of a distribution $p_g(x)$ and instead of generating a wide, diverse array of images it selects from a limited set of images. [19]

StyleGAN2 For this project we elected to the state-of-the-art network StyleGAN2 for our image synthesis. Historically, Jensen-Shannon divergence was used as the distance metric to measure the distance between $p_g(x)$ and $p_r(x)$. [10] Jensen-Shannon divergence measures the vertical difference in heights of the output distribution between $p_g(x)$ and $p_r(x)$. More recent GANs, including StyleGAN2, use Wasserstein divergence which measures the horizontal distance between image output distributions as the metric for updating generator weights. [18]

What elevates StyleGAN2 to the task of brain MR image generation above other GANs is its signature use of intermediate latent space manipulation. [11] Where as our traditional GAN just plugs and chugs our initial input z , StyleGAN2 first has a mapping network f that first transforms it to some intermediate w . Affine transformations of w are then fed into the the generator g via adaptive instance normalization (AdaIN). [9, 6, 8] Here affine transformations to w within the latent domain results in shifts in direction, shape, color, and structure of faces in pretrained models. [11] StyleGAN2 also uses path length regularization which amounts to a fixed-sized step in the latent space W that contains w results in a fixed-magnitude change to the image.

2 Previous Work

GANs are typically used in one of two ways in the medical imaging field: image discrimination and image synthesis. GANs that focus on the discriminative aspect are useful as a regularizer for detecting abnormal images. While GANs that focus on the generative aspect are great for addressing the data scarcity problem plaguing the field due to patient privacy restrictions. [19]

2.1 Medical Image Synthesis

A recurring problem in the intersection the medical fields and computer vision is the problem of patient privacy. Where as more modern deep learning techniques require large annotated data sets, many of the necessary diagnostic images have not been released into the public domain. [5]

GANs have been employed as a novel data augmentation technique as many neuroradiologists claim generated MR data is of comparable quality to real images, save some discrepancies in anatomic accuracy. [4] Synthetically generated

medical data has been proven to help with classification tasks. Specifically in the case of liver lesions where Frid-Adar et al. [7] used three DCGANs (a type of GAN used for unconditional synthesis) to generate three classes of samples for liver lesions: cysts, metastases, and hemangiomas. When combined with real training data, the generated images were proven beneficial in increasing both sensitivity and specificity.

GANs have been a huge boon in dealing with the insufficient amount of data for many rarer diseases. Typically in these large data sets, image augmentation techniques include elastic deformation, flipping, rotation, scaling, and translation. [16] However, this doesn't take into account the appearance, location, shape, or size of the tumor relative to the brain. This paper hopes to address this shortcoming by using StyleGAN2 to encode aspects of images like appearance, location, shape, and size into the latent space W ; the idea being, taking a fixed-size step in any direction in W will result in a uniform change to various aspects the tumor and the image unreachable by traditional augmentation methods.

3 Methods

For this project we elected to use the 256x256 resolution pretrained model used in [11] for our image synthesis. Using a pretrained model greatly reduced training time and system requirements.

3.1 Training

Initially, we elected to combine data from three different tumor classifications (meningioma, glioma, and pituitary tumor) into one network. These images were originally 512x512 resolution, of which the outer 128 pixels on each side was cropped. The new 256x256 images were then down sampled to 32x32 to quickly train the network. While this initial training showed promising results with regards to brain structure and tumor placement, because of the 32x32 resolution down sampling many of the important details that would help classifiers and segmenters with specificity were lost.

Now that we knew the generator showed promise, for the next stage we opted to no longer down sample the images and we separated them into the three categories depending on tumor type. From here we trained three separate 256x256 models to generate each of the above tumor MR images. Each model was trained for a few thousand epochs. Figure 2 demonstrates an example of the MR images generation sampled from each category.

3.2 Evaluation

We used two main methods of evaluation in this paper: Inception Score, and subjective analysis. For subjective analysis we simply created a large sample of synthetic data and judged it through observation. Much more complicated than

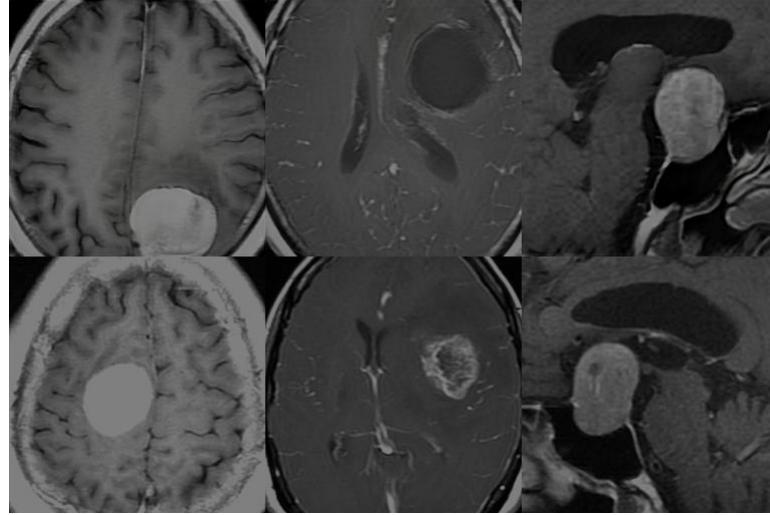


Fig. 2. Generated and real brain MR images, from left to right images represent a meningioma, glioma, and pituitary tumor; images on the top layer are generated, images on the bottom layer are from the real training set.

subjective analysis, Inception score provides an empirical insight into the quality of the generated images, and is closely correlated to how a human would rate the images. [14] Inception score relies on an additional neural network to classify the generated images. For this classifier we used the InceptionV3 model [17]. We trained the model for thirty epochs with a batch size of 64 and used Adam as the optimizer. This ultimately resulted in a validation accuracy of approximately 90 percent. Unlike the approach which we took in generating images, the inception model combines all three classes of brain tumor. This will allow us to understand the efficacy of utilizing the generated images in real world tumor identification models which aim to identify the type of brain tumor rather than just their existence. The metric which Inception Score produces is somewhat ambiguous thus it is necessary to compare it to a baseline. To do this we simply calculated the inception score of three hundred real examples and 300 generated examples.

4 Results

Our models each trained for about 1 week straight on a Tesla P100 while the transfer learning model took many GPU years.

4.1 Emperical

Calculating the Inception Scores for real and generated examples, as detailed above, manifested the result that our generator performed extremely well. The

final inception scores were as follows: Baseline 2.6, Generator 2.4. The score of the generator is approximately 90 percent of the baseline score indicating that the generator produces examples which are on average 90 percent as realistic as the real data. This is an extremely strong result and is reflected by the subjective quality of the images. The excellent inception score also indicates that our generator will be effective in providing supplemental training data for tumor identification classifiers.

4.2 Examples

Above in Figure 2 we selected images from each class with comparable counterparts in the training set. Below in Figure 3 we generated nine randomly sampled examples from each generator to illustrate the general quality and detail of our images. Further below that, Figure 4 is 300 generated images of each class (meningioma, glioma, and pituitary tumor) to illustrate the diversity of the generator.

5 Conclusion

We have illustrated the quality and diversity of MR images state-of-the-art generators like StyleGAN2 can produce. We also took advantage of transfer learning to vastly improve the production speed of the models. [11]

References

1. Abd-Ellah, M.K., Awad, A.I., Khalaf, A.A., Hamed, H.F.: Design and implementation of a computer-aided diagnosis system for brain tumor classification. In: 2016 28th International Conference on Microelectronics (ICM). pp. 73–76. IEEE (2016)
2. Akil, M., Saouli, R., Kachouri, R., et al.: Fully automatic brain tumor segmentation with deep learning-based selective attention using overlapping patches and multi-class weighted cross-entropy. Medical image analysis **63**, 101692 (2020)
3. Akram, M.U., Usman, A.: Computer aided system for brain tumor detection and segmentation. In: International conference on Computer networks and information technology. pp. 299–302. IEEE (2011)
4. Bermudez, C., Plassard, A.J., Davis, L.T., Newton, A.T., Resnick, S.M., Landman, B.A.: Learning implicit brain mri manifolds with deep learning. In: Medical Imaging 2018: Image Processing. vol. 10574, p. 105741L. International Society for Optics and Photonics (2018)
5. of the Clinical Practice Committee, D.S., et al.: Informed consent for medical photographs. Genetics in Medicine **2**(6), 353 (2000)
6. Dumoulin, V., Shlens, J., Kudlur, M.: A learned representation for artistic style. arXiv preprint arXiv:1610.07629 (2016)
7. Frid-Adar, M., Diamant, I., Klang, E., Amitai, M., Goldberger, J., Greenspan, H.: Gan-based synthetic medical image augmentation for increased cnn performance in liver lesion classification. Neurocomputing **321**, 321–331 (2018)

8. Ghiasi, G., Lee, H., Kudlur, M., Dumoulin, V., Shlens, J.: Exploring the structure of a real-time, arbitrary neural artistic stylization network. arXiv preprint arXiv:1705.06830 (2017)
9. Huang, X., Belongie, S.: Arbitrary style transfer in real-time with adaptive instance normalization. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 1501–1510 (2017)
10. Karras, T., Aila, T., Laine, S., Lehtinen, J.: Progressive growing of gans for improved quality, stability, and variation. arXiv preprint arXiv:1710.10196 (2017)
11. Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., Aila, T.: Analyzing and improving the image quality of stylegan. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8110–8119 (2020)
12. Kaus, M.R., Warfield, S.K., Nabavi, A., Black, P.M., Jolesz, F.A., Kikinis, R.: Automated segmentation of mr images of brain tumors. Radiology **218**(2), 586–591 (2001)
13. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., et al.: The multimodal brain tumor image segmentation benchmark (brats). IEEE transactions on medical imaging **34**(10), 1993–2024 (2014)
14. Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., Chen, X.: Improved techniques for training gans (2016)
15. Schneider, T., Mawrin, C., Scherlach, C., Skalej, M., Firsching, R.: Gliomas in adults. Deutsches Ärzteblatt International **107**(45), 799 (2010)
16. Simard, P.Y., Steinkraus, D., Platt, J.C., et al.: Best practices for convolutional neural networks applied to visual document analysis. In: Icdar. vol. 3 (2003)
17. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2818–2826 (2016)
18. Wu, J., Huang, Z., Thoma, J., Acharya, D., Van Gool, L.: Wasserstein divergence for gans. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 653–668 (2018)
19. Yi, X., Walia, E., Babyn, P.: Generative adversarial network in medical imaging: A review. Medical image analysis **58**, 101552 (2019)

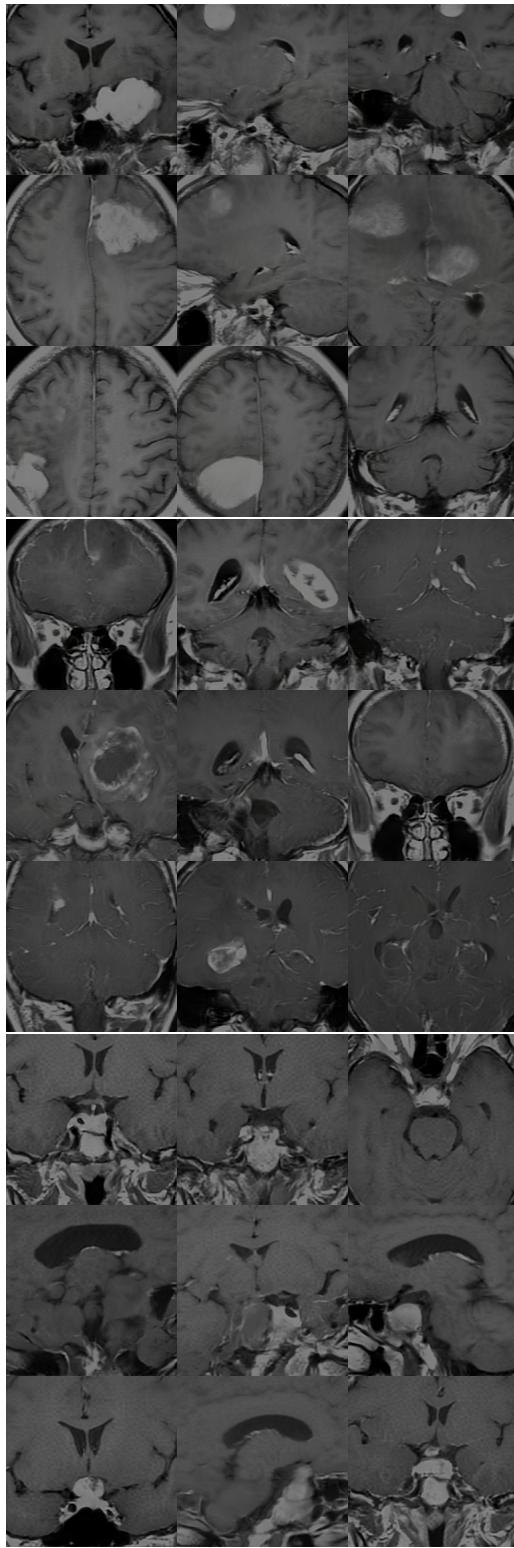


Fig. 3. Generated brain MR images, from top to bottom images represent a meningioma, glioma, and pituitary tumor.

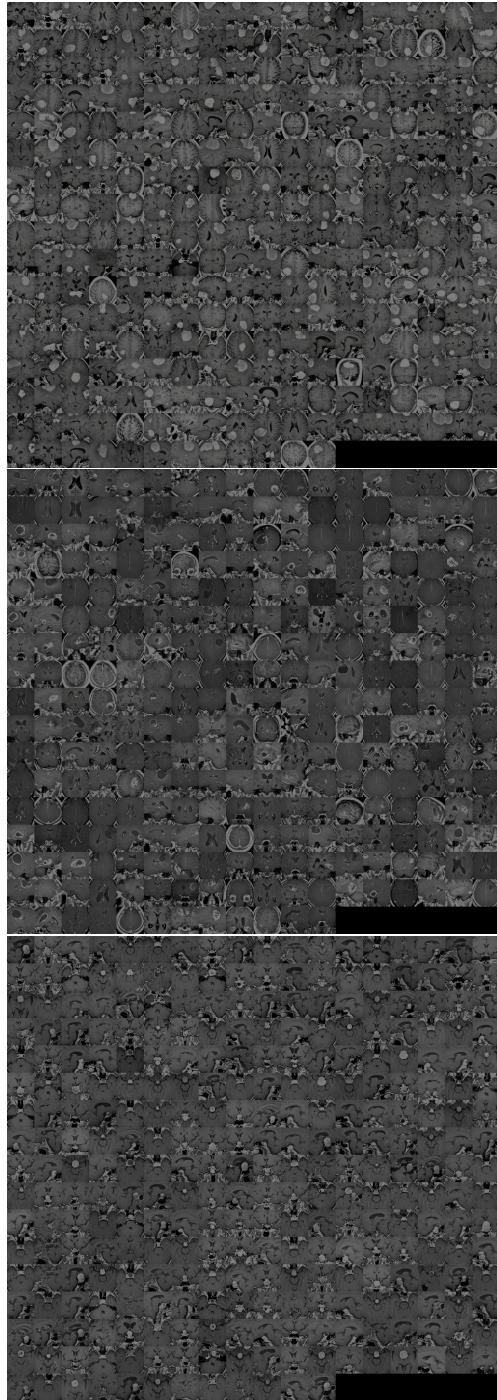


Fig. 4. Generated brain MR images, from top to bottom images represent a meningioma, glioma, and pituitary tumor.