

Optimizing GAN parameters for efficient and accurate image generation: A study of WGAN-GP in brain tumor dataset

Yuanjun Feng

School of Information Science and Technology, Beijing Forestry University, Beijing, 100083, China

fyj0531@bjfu.edu.cn

Abstract. In this era of unprecedented rapid development of artificial intelligence, researchers are moving forward to develop new neural networks. Nevertheless, few have considered optimizing the extant generative adversarial network (GAN) to achieve the most effective and accurate solutions to provide optimal results for the challenges. This study aims to investigate the critical role of parameter optimization in GAN neural networks, with a particular focus on the Wasserstein Generative Adversarial Network with Gradient Penalty (WGAN-GP) architecture applied to the generation of medical images, especially images depicting brain tumors. Therefore, the project used Kaggle's brain tumor data set as a canvas to conduct an in-depth study of the impact of these parameters on the training model and generated results. For the starter, to investigate how alterations in the learning rate impact the model, this article selects a series of values for meticulous analysis to determine the most effective configuration. Then, evaluate and find a better and more suitable choice between Adam and SGD optimizers through comparison, focusing on their impact on training dynamics. As the last one, this study examines how the Tanh activation function constrains pixel values and shapes image realism through comparative results. By dissecting and understanding the interaction of these parameters in detail, we lay the foundation for optimizing GAN neural networks, increasing their efficiency, and producing accurate solutions for accurate diagnostics and healthcare applications. The journey through the labyrinth of GAN parameter tuning ultimately provided valuable insights into seamlessly synthesizing medical images.

Keywords: Generative Adversarial Networks (GANs), Parameter Optimization, Medical Image Generation, WGAN-GP.

1. Introduction

Artificial Intelligence (AI) has had profound impacts across various domains, and Generative Adversarial Networks (GANs), a significant technology within this realm, have shown immense potential in fields e.g. computer vision and natural language processing. GANs consist of two opposing neural networks, a generator and a discriminator, which engage in competitive and cooperative iterations to optimize information. This procedure produces simulation data that exhibits a greater resemblance to authentic and realistic information. From the early stages of AI development to the emergence of GANs, this technology has transcended traditional methods based solely on rules and templates, achieving more realistic data generation. When it comes to the field of healthcare, GANs are extensively used to form synthetic medical images with biometric features to imitate various disease scenarios. GANs have made

notable advancements in image and data augmentation, not to mention their potential for image registration, segmentation, and other tasks, leading to a significant boost in the quality and diversity of medical images. Consequently, they are indispensable in vital fields such as diagnosing cases and identifying tumors.

As technology continues to evolve, diverse forms of GANs have surfaced. For instance, Conditional GANs (CGANs) permit controlled generation processes, as seen in models like Pix2Pix, which transform sketches into lifelike images, finding broad applications in art creation and design [1]. Style Transfer GANs (StyleGANs), on the other hand, can form images with discrete styles, fostering innovation in artistic creation and the fashion sector [2]. However, this is just the tip of the iceberg. Different types of GANs can achieve similar goals within the same domain, and additionally, a single GAN type can be employed to address various problems. Specifically, Cycle-Consistent GANs (CycleGANs) also have the potential for image style transfer, much like the work done by StyleGAN. This technology finds applications in urban planning, such as transforming real-world street scenes into landscapes of different seasons, aiding decision-makers in improved planning. In addition, CycleGANs can be utilized for image transformation, such as turning horses into zebras, handling the problem that Pix2Pix's training data must be in pairs whereas getting these pairs is usually problematic [3]. In summary, CycleGANs can be deployed in various fields, including but not restricted to, image translation and augmented reality.

Furthermore, in the domain of Natural Language Processing (NLP), GANs are employed to generate text resembling human writing for applications such as creating advertising copy. Also, as mentioned above, GANs play a vital role in the medical imaging field by generating medical images to supplement insufficient training data, which enables the analysis of a far more comprehensive range of possible pathological impacts and eventually enhances the accuracy of medical image diagnostics.

To sum up, as a pivotal technology in the realm of artificial intelligence, Generative Adversarial Networks have demonstrated considerable creativity and application potential across multiple domains. New generations of GANs are iterated upon and updated, resolving an increasing number of novel problems. Nevertheless, few have considered the pursuit of optimizing existing GAN neural networks to achieve the most efficient and accurate solutions, providing an optimal outcome for these challenges. Thus, this paper aims to profoundly investigate the relationship between the parameters of GAN neural networks and their corresponding performance. Starting with the WGAN-GP network[4], using the brain tumor dataset obtained on Kaggle, it delves into aspects such as the structure of the generator and discriminator, learning rates, and optimizers, investigating their influence on the accuracy of results provided by the network by comparison.

2. Methodology

2.1. Dataset preparation

In this study, the Brain Tumor dataset provided by Kaggle was employed [5]. The dataset comprises features about brain tumors, which include five first-order characteristics, eight second-order characteristics that can also be named texture characteristics, and the target category. The first-order features include mean, variance, standard deviation, skewness, and kurtosis. The texture features are contrast, energy, Angular Second Moment (ASM), entropy, homogeneity, dissimilarity, correlation, and roughness. Additionally, the Class column represents whether the image depicts a tumor (1 = tumor, 0 = no tumor). The dataset contains 3,762 sample data, with each measuring 240×240. Upon classification analysis, the dataset comprises 2079 tumor-negative samples and 1683 tumor-positive ones. Several sample images are shown in Figure 1.

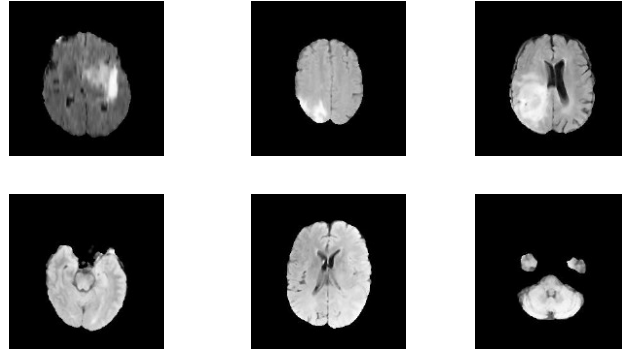


Figure 1. Tumor Positive (Top) and Negative Images (Bottom).

2.2. WGAN-GP

2.2.1. WGAN. The overall structure of the WGAN-GP of this project is as follows: The generator begins with a random noise vector input of size 128. Then, through a linear layer, a fully connected layer, it maps the noise vector to a higher-dimensional tensor to be processed by the following set of residual blocks. The second layer is residual blocks. The generator consists of four residual blocks for gradually increasing the resolution of the generated images. Each residual block contains two convolutional layers and a batch normalization layer, with different channel numbers and convolution kernel sizes, and gradually builds higher-resolution feature maps through residual connections. The structure of the residual block helps prevent the vanishing gradient problem and generates higher-quality images. The resulting image features are then normalized using a batch normalization layer. Then comes a convolutional layer that maps the final feature map to 3 channels (corresponding to RGB images), with a convolution kernel size of 3. Nonlinear transformation is implemented using the ReLU activation function and the output is limited to the range of $[-1, 1]$ via the Tanh activation function. At the end of the generator, the output image is transformed into a 3-channel image using the contiguous function and reshaped, to match the shape of the training data.

The discriminator starts with a convolutional layer. First, the input image (3-channel RGB) goes through a convolutional layer for feature extraction. The discriminator also has four residual blocks to do the opposite of the generator, that is, to gradually reduce the resolution of the image and increase the abstraction of the features. The number and parameters of these residual blocks correspond to those of the generator. Finally, the image features are mapped to a scalar via a linear layer, utilized to represent the realism of the input image.

2.2.2. WGAN-GP. WGAN-GP refers to Wasserstein Generative Adversarial Network with Gradient Penalty, which is an advanced variant of the GAN model. First, instead of using traditional JS or KL divergence to measure the difference between the distributions of real and generated data, WGAN-GP uses Wasserstein distance, also known as Earth Mover distance (EMD) that provides a more meaningful measure of distance between distributions, which is more stable than the cross-entropy loss function used in GAN, and can avoid problems such as training instability and mode collapse that occur in GAN. Secondly, WGAN-GP introduces a gradient penalty term in the loss function of the discriminator, which can constrain the gradient norm of the discriminator, that is, ensure that the Lipschitz constraint is satisfied. This encourages a more balanced training dynamic between the generator and the discriminator, which avoids the problem of exploding and vanishing gradients. To sum up, WGAN-GP has been improved and improved compared to GAN in terms of training stability, loss function, gradient constraints, and avoidance of mode collapse, so it is more commonly used and effective in practical applications.

2.3. Implementation details

This article mainly conducts a comparative study of projects based on changes in three parameters, including learning rate, optimizer, and activation function.

To begin with, the learning rate is discussed. This hyperparameter holds significant importance in the training of deep neural networks. It plays a pivotal role in dictating the scope of model parameter adjustments in each iteration, thereby exerting a direct influence on both the training pace and the eventual model's performance. Therefore, this project aims to determine the optimal learning rate by repeatedly testing various values, including 0.001, 0.0001, 0.00001, and so on, to achieve fast convergence and high-quality models.

Additionally, selecting the suitable optimizer is essential when training WGAN-GP, with options including the commonly used SGD and Adam optimizer. Both optimizers may have superior performance in different scenarios, so determining the most suitable optimizer for the task requires comparing their convergence speed and generated image quality. Although it is generally recognized that Adam performs well in many situations, specific situations may benefit from SGD, so comprehending their characteristics and applicable scenarios is essential.

Finally, the activation function is also instrumental in deep neural networks. They are responsible for introducing nonlinear properties that enable networks to learn and represent complex functional relationships. Being the predecessor of modern activation functions, the tanh function is characterized by its bounded output, making it an ideal candidate to serve as the input for the next layer. This inherent property naturally positions it as the primary choice for fulfilling the role of a nonlinear function. Therefore, it has become the primary research object of this project. This study aims to undertake a comprehensive exploration of the tanh function's influence on model performance, with the intention of offering valuable insights and guidance for the design and application of generative models.

3. Results and discussion

3.1. Learning Rate

First, the following Figure 2 represents the results of each group that underwent changes in learning rate parameters (**lr** represents the learning rate):

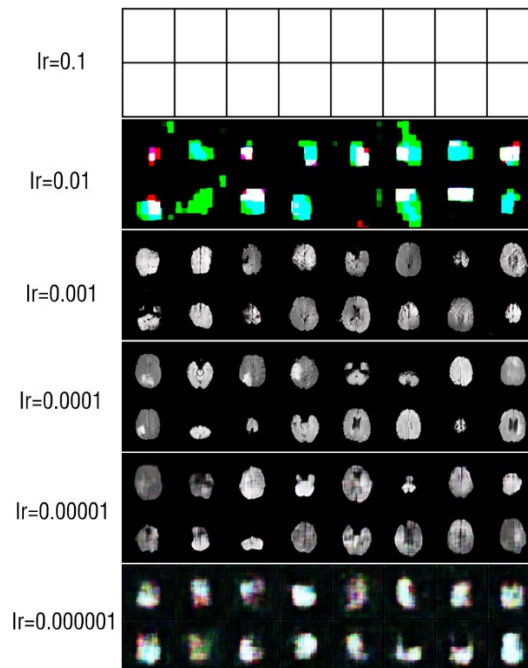


Figure 2. The generated images based on different learning rates.

At such a high learning rate of 0.1, instability becomes apparent, leading to a training imbalance between the generator and the discriminator. Typically, the discriminator can be so powerful that it is difficult for the generator to generate realistic images. The Wasserstein distance may fluctuate and not decrease consistently, causing the generator to be incapable of producing images.

Reducing the learning rate to 0.01 usually improves training stability but may still result in a somewhat unstable training process. The performance may improve for both the discriminator and the generator, but the quality of generated images may still be relatively poor. The Wasserstein distance may exhibit more stability but could converge to higher values.

A learning rate of 0.001 is often considered a good starting point in WGAN. It performs relatively well in improving training stability and image quality generation. The Wasserstein distance is likely to stabilize, and the quality of generated images may improve.

A learning rate of 0.0001 is a commonly used learning rate in WGAN. It typically exhibits good smoothness during the training process, and the quality of generated images may significantly improve. The Wasserstein distance usually stabilizes and results in more realistic generated images.

Further reducing the learning rate to 0.00001 may result in a longer training process, which means a slower training speed, but generally results in higher quality generated images after long training periods. Wasserstein distance may be more stable but requires more iterations to get the best results.

A learning rate of 0.000001 is typically considered extremely small and may demand a significant amount of time to converge. Nevertheless, it can generate highly realistic images with prolonged training. The Wasserstein distance may remain relatively stable, but it may take much longer to achieve peak performance.

3.2. Optimizer

The author compares the impact of two mainstream optimizers shown in Figure 3: Adam and SGD, on the generated results. The evaluation of the performance of the two optimizers will be carried out from three aspects, which are: Training Stability, Generated Images' Quality and Stability of Wasserstein Distance.

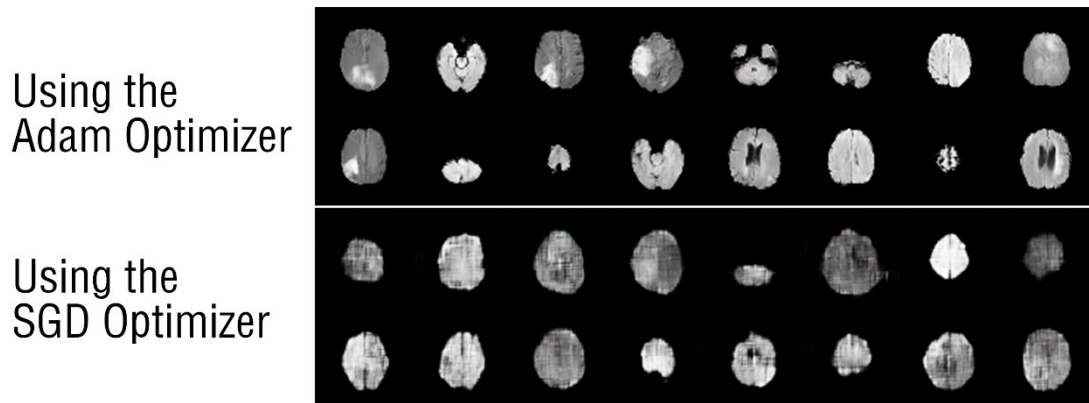


Figure 3. The generated images based on different optimizers.

1) Adam [6]: Training Stability: Adam usually exhibits good training stability. It can converge quickly in the early stages of training, reducing the risk of the model getting stuck in local minima, which is beneficial for WGAN training.

Generated Images' Quality: Adam typically generates higher-quality images. Due to its adaptive learning rate mechanism, it helps the generator learn to produce realistic image distributions more rapidly.

Stability of Wasserstein Distance: Adam usually brings down the Wasserstein distance to a stable state more quickly, aiding in obtaining high-quality generated images faster.

2) SGD [7]: Training Stability: Compared to Adam, SGD generally has poorer training stability. It may require more tuning and manual adjustments to achieve a stable state.

Generated Image Quality: SGD typically requires more time for training to produce high-quality generated images. Due to the lack of an adaptive learning rate mechanism, the model may get stuck in local minima, resulting in lower image quality.

Stability of Wasserstein Distance: SGD usually takes more time to lower the Wasserstein distance to a stable state, which means that the model may require more iterations to achieve satisfactory results.

3.3. Activation Function

Obviously, without the Tanh function constraining the pixel values of the generated images, they become more contrasty shown in Figure 4, which is reflected in the result that their colors become richer.

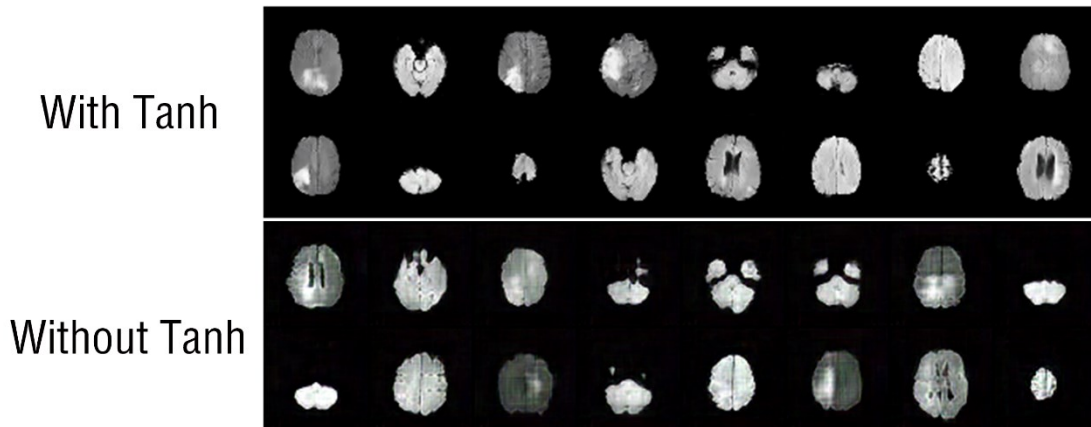


Figure 4. The generated images based on different activation functions.

3.4. Discussion

In terms of the learning rate, from the above comparison results, it is evident that the selection of the learning rate holds significant importance for the success of the training process. An excessively high learning rate can introduce instability, whereas an overly low one can result in a sluggish training process or cause the model to become trapped in a local minimum during training. Experimental results show that smaller learning rates (such as 0.0001 in this study) generally perform better in WGANs.

The second is the optimizer. Different optimizers show different performances in WGAN training. The previous article evaluated the results of the optimizer comparison from three perspectives: training stability, quality of generated images, and stability of the Wasserstein distance. Compared with SGD, the Adam optimizer has faster convergence speed, higher generated image quality, and reaches a stable state of the loss function in a shorter time. These advantages are all due to its adaptive learning rate mechanism, which enables Adam to target Different parameters and assign different learning rates to control the direction and size of the gradient.

Lastly, when discussing the activation function Tanh function, experimental results show that using the Tanh activation function in WGAN usually produces smooth transitions and realistic details in the generated images. Simultaneously the pixel value restrictions brought by the Tanh function also put constraints on the result images, making them more realistic. The Tanh function does help the generator to forge visually pleasing authentic images, although naturally, sometimes the generated images can be too smooth and lack some sharp edges and details. In the future, to further improve the performance of the generation, some advanced technologies e.g. transfer learning and attention mechanisms may be considered [8-11].

4. Conclusion

In this paper, the investigation revolves around the relationship between key parameters of GANs and their consequential impact on performance, with specific attention directed towards the WGAN-GP

network's role in generating brain tumor images. A comprehensive exploration encompasses learning rates, optimizer options (Adam vs. SGD), and activation functions (Tanh vs. no Tanh), ultimately yielding insightful revelations pertinent to GAN neural network optimization. The overarching goal of this study lies in augmenting the efficiency and precision of GAN-driven data generation, with a pronounced emphasis on the realm of medical imaging. The findings illuminate the pivotal influence these parameters wield in GAN training. Notably, smaller learning rates, exemplified by 0.0001, surface as the preferred choice for WGANs, while the adaptive learning rate mechanism intrinsic to the Adam optimizer significantly bolsters training stability, image quality, and convergence speed. Moreover, the deployment of the Tanh activation function is unveiled as a catalyst for the creation of realistic images, albeit with a potential trade-off involving some loss of fine details. While this research underscores the influence of parameter optimization on outcomes and identifies superior options through multiple comparative analyses, the quest for the definitive optimal choice remains ongoing. Future research endeavors aim to establish methodologies capable of pinpointing, for instance, the ideal learning rate.

References

- [1] Isola P Zhu J Y Zhou T and Efros A A 2017 Image-to-image translation with conditional adversarial networks In Proceedings of the IEEE conference on computer vision and pattern recognition pp 1125-1134
- [2] Karras T Laine S and Aila T 2019 A style-based generator architecture for generative adversarial networks In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition pp 4401-4410
- [3] Zhu J Y Park T Isola P and Efros A A 2017 Unpaired image-to-image translation using cycle-consistent adversarial networks In Proceedings of the IEEE international conference on computer vision pp 2223-2232
- [4] Gulrajani I Ahmed F Arjovsky M Dumoulin V and Courville A C 2017 Improved training of wasserstein gans Advances in neural information processing systems 30
- [5] Kaggle 2020 Brain tumor <https://www.kaggle.com/datasets/jakeshbohaju/brain-tumor>
- [6] Reddi S J Kale S and Kumar S 2019 On the convergence of adam and beyond arXiv preprint arXiv:1904.09237
- [7] Woodworth B et al 2020 Is local SGD better than minibatch SGD?. In International Conference on Machine Learning (pp. 10334-10343) PMLR
- [8] Qiu Y et al 2019 Semantic segmentation of intracranial hemorrhages in head CT scans In 2019 IEEE 10th International Conference on Software Engineering and Service Science (ICSESS) pp. 112-115) IEEE
- [9] Niu Z Zhong G and Yu H 2021 A review on the attention mechanism of deep learning. Neurocomputing 452 48-62
- [10] Qiu Y Wang J Jin Z Chen H Zhang M and Guo L 2022 Pose-guided matching based on deep learning for assessing quality of action on rehabilitation training Biomedical Signal Processing and Control 72 103323
- [11] Bengio Y 2012 Deep learning of representations for unsupervised and transfer learning. In Proceedings of ICML workshop on unsupervised and transfer learning pp 17-36 JMLR Workshop and Conference Proceedings