Advances in Engineering Research : Possibilities and Challenges
ISSN Print: 3079-5192
ISSN Online: 3079-5206
https://doi.org/10.63313/AERpc.2015

Implementation and Optimization of Generative Adversarial Networks in Handwriting Image Modeling

Youyu Ting[1]. Wenan Tan[1*] and Jinlong Lv[2]
[1]School of Computer and Information Engineering, Shanghai Polytechnic University, China
[2]School of Intelligent Manufacturing and Control Engineering, Shanghai Polytechnic University, China
Email: watan@sspu.edu.cn

How to cite this paper: Ting, Y. Y., Tan, W. A., & Lv, J. L. (2025). Implementation and Optimization of Generative Adversarial Networks in Handwriting Image Modeling. Advances in Engineering Research : Possibilities and Challenges, 1(3), 33-40. ISSN Print: 3079-5192; ISSN Online: 3079-5206. https://doi.org/10.63313/AERpc.2015
Published: 2025-06-12

Abstract

Generative Adversarial Networks (GAN), as an important research direction in the field of deep learning in recent years, have been widely used in many fields such as image and speech due to their excellent data generation ability. In this paper, an adversarial model for MNIST handwritten digital image generation is designed and implemented based on the classical GAN framework. The model adopts fully connected neural networks to construct the generator and discriminator and com-bines optimization techniques such as Label Smoothing and Batch Normalization to improve the training stability and image generation quality effectively. Through experiments on the MNIST dataset, this paper systematically analyzes the training process of the model and its performance. The results demonstrate that the model can gradually learn the distribution characteristics of real images and generate handwritten digital images with clear morphology and a reasonable structure, ver-ifying the effectiveness and feasibility of the designed method. The research in this paper not only deepens the understanding of the working mechanism of GAN but also provides a foundation and reference for subsequent research on image generation in more complex scenes.

Keywords

Generative Adversarial Networks; MNIST; Fully Connected Neural Networks; La-bel Smoothing; Batch Normalization

1. Introduction

With the rapid development of artificial intelligence technology, deep learning has achieved remarkable results in computer vision [1], natural language processing [2], and other fields. Among them, the Generative Adversarial Network (GAN), as a novel generative model, has received widespread attention since it was proposed by Goodfellow[3] and others in 2014. By introducing an adversarial mechanism between Generator and Discriminator, GAN enables the model to learn the distributional features of the data to generate realistic samples, and performs well in tasks such as image generation, image restoration, and style migration.

Traditional image generation methods rely on hand-designed features or probabilistic graph-based models, which suffer from complex modeling and limited generation quality [4]. GAN, on the other hand, adopts unsupervised learning and requires only raw data for training, which effectively solves the problem of sample modeling in high-dimensional space. Especially when dealing with the task of image generation, GAN shows strong data representation ability and has become one of the most representative generative models at present.

MNIST handwritten digits dataset is a classic test dataset in the field of deep learning, which consists of 60,000 training images and 10,000 test images, each of which is a grayscale map of size 28×28, with the content of handwritten digits from 0 to 9. Based on the simplicity of its data structure and its clear annotations, MNIST has been widely used for verifying the validity of image recognition and generation models [5].

The aim of this paper is to realize the task of generating MNIST handwritten digit images based on a generative adversarial network with a fully connected structure. By constructing shallow Generator and Discriminator models, an adversarial training strategy is used to continuously optimize the generator to improve the quality of the generated images. At the same time, Label Smoothing and Batch Normalization are introduced in the training process to enhance the stability and convergence of the model. Finally, the feasibility and effectiveness of the constructed model in the image generation task are verified through image visualization and loss function curve analysis. This study not only helps to understand the basic principles and training mechanisms of GAN models but also lays the foundation for subsequent more complex image generation tasks. Through the experimental exploration of the MNIST dataset, it can be further extended to more challenging practical application scenarios such as face image generation and art style migration.

2. Related work

In recent years, with the continuous development of deep learning technology, generative models have become a research hotspot. Among them, Generative Adversarial Network (GAN), as a kind of generative model based on the idea of the game, is widely used in the fields of image, speech, text, etc. because of its powerful data distribution modeling ability.

2.1. Mechanisms of GAN model generation

Generative Adversarial Network was first proposed by Goodfellow et al. in 2014, and its core idea is to train the model through the game relationship between two neural networks, Generator and Discriminator[6]. The Generator is responsible for generating samples approximating real data from random noise, while the Discriminator is used to discriminate whether the input samples are real data or generated data. The whole training process can be represented as a min-max game with the following loss function:

$$\min_G \max_D V(D,G) = E_{x\sim p_{data}(x)}[\log D(x)] + E_{z\sim p_z(z)}[\log(1-D(G(z)))] \quad (1)$$

In the original GAN structure, the generator and discriminator are usually constructed using fully connected neural networks, which are suitable for the generation task of low-resolution images, such as MNIST handwritten digital images. Although the structure is simple and intuitive, problems such as unstable training and poor sample quality often occur in high-dimensional data modeling, and most of the subsequent research has been improved and optimized on this basis.

2.2. Improvement and Extension of GAN Models

After the basic structure of GAN was proposed, many scholars have improved and extended it. Since the adversarial training mechanism of GAN is more complex and susceptible to problems such as gradient vanishing[7]and mode collapse, researchers have proposed various optimization methods to improve the training stability and generation effect. For example, Radford et al. proposed a Deep Convolutional Generative Adversarial Network (DCGAN), by introducing the structure of the convolutional neural network, the model is more suitable for processing image data and significantly improves the clarity and stability of the generated samples. In addition, in order to improve the training stability, the researchers also proposed a variety of optimization methods such as Label Smoothing, Gradient Penalty, etc., to alleviate the common problems of Mode Collapse and Gradient Disappearance during the GAN training process[8].

2.3. Application of the GAN model

GAN has been widely used in several fields since it was proposed, especially in image processing.

In image generation tasks, GAN is used to generate high-quality images such as handwritten digits, face images, and natural scenes. Taking MNIST handwritten digits as an example, researchers often realize high-fidelity reproduction of 28×28 pixel digital images by constructing generators and discriminators with lightweight or convolutional structures. Besides, GAN has been applied in the following directions:

1.image style migration: applying the style of one image to another (e.g. CycleGAN);
2.image restoration: self-filling of damaged image regions;
3.image super-resolution: improve the clarity and resolution of images (e.g. SRGAN); medical image synthesis, anime avatar generation, face dressing, etc.

3.3. GAN implementation in handwriting image modeling

3.1. GAN Principles

Generative adversarial networks contain two main sub-networks:

1. Generator: Receives a random noise vector with dimension 100 (sampled from standard normal distribution), maps it into an image of size 28×28 through multi-layer nonlinear transformation, to generate a "dummy image" that is as close as possible to the distribution of the real image. 2.

2. Discriminator (Discriminator): Receive an input image (maybe a real image or generated image), and input a probability value, indicating the possibility that the image is a real image. The goal of the discriminator is to maximize the distinction between authentic images and generated images.

The two form a zero-sum game: the generator continuously optimizes to "trick" the discriminator. In contrast, the discriminator tries to recognize the difference between the generated image and the real image. Eventually, the two networks reach a Nash equilibrium, and the generator outputs an image that approximates the real image distribution.

3.2. Network Architecture Design

3.2.1.Generator Architecture Design

The generator consists of a four-layer fully connected neural network with the network structure shown in the following table.

Layers	Operation Type	Output Dimension	activation function	Description
Input Layer	100-dimensional noise vector	-	-	Random initialization
Hidden Layer 1	Fully Connected (Linear)	256	LeakyReLU+BN	BN is not used
Hidden Layer 2	Fully connected	512	LeakyReLU+BN	BatchNorm speeds up convergence
Hidden Layer 3	Fully connected	1024	LeakyReLU+BN	-BatchNorm
Output Layer	Fully connected	784(28×28)	Tanh	Output image in the range [-1,1]

Table 1. Generator network structure

The final output of the generator is activated by the Tanh() activation function, which compresses the pixel values into the interval [-1, 1], which corresponds to the normalized range of the input data.

3.2.2.Discriminator structure design

The discriminator consists of a three-layer fully connected neural network with the structure shown in the following table:

Layers	Operation type	Output Dimension	activation function
Input Layer	Flatten to 784-dimensional vector	-	-
Hidden Layer 1	Fully Connected	512	LeakyReLU
Hidden Layer 2	Fully connected	256	LeakyReLU
Output Layer	Fully connected	1	Sigmoid

Layers	Operation type	Output Dimension	activation function

Table 2. Discriminator network structure

The discriminator finally outputs a probability value for the likelihood that the image is a real sample.

3.3. Loss function design

To train the generator and the discriminator, this paper uses the classical Binary Cross Entropy Loss (BCEL) function to calculate the objective function of the discriminator and the generator respectively.

3.3.1.Discriminant loss function

The loss of the discriminator consists of two components, the loss of judgment on the real image and the loss of judgment on the generated image:

$$L_D = -E_{x\sim p_{data}(x)}[\log D(x)] - E_{z\sim p_z(z)}[\log(1-D(G(z)))] \tag{2}$$

In the implementation, to improve training stability, the labels of the real images were set to 0.9 (label smoothing technique) instead of the traditional 1.0 for the generated images.

3.3.2.Generator Loss Function

The goal of the generator is to "trick" the discriminator, i.e., it wants the discriminator to recognize the generated image as the real one, so its loss function is:

$$L_G = -E_{z\sim p_z}[\log D(G(z))] \tag{3}$$

3.4. Training strategies and optimization methods

3.4.1.Optimizer Configuration

- both networks use Adam optimizer for parameter updating.
- the learning rate is 0.0002,β1=0.5, and β2=0.999, this setting has been verified to have good training stability in GAN.

3.4.2.Training process

Each training cycle (Epoch) consists of the following steps:

- acquiring a batch of images from real data.
- generating an image by sampling a noise vector from a normal distribution.
- using the discriminator to judge the real image and the generated image separately, calculating the discriminator loss, and updating the discriminator.
- sample the noise vector again and update the generator with the generated image so that the generated image is more likely to be judged as real by the discriminator.
- record the loss changes for each epoch and save the generated image for visualization.

4. Experimental results

4.1. Authors and Affiliations

The dataset used in this experiment is the MNIST (Mixed National Institute of Standards and Technology) handwritten digital image dataset, provided by the National Institute of Standards and Technology, which is one of the most commonly used image datasets in the field of deep learning. The dataset contains a total of 70,000 grayscale images, of which 60,000 are in the training set and 10,000 are in the test set. The size of each image is 28×28 pixels, and the image content is handwritten numbers 0 to 9, with a total of 10 categories, and the images are stored in standard black and white grayscale with pixel values ranging from 0 to 255.In this experiment, the test set (10,000 images) is selected as the real image inputs for the training of the discriminator, which provides an adversarial target for the generator.The following figure shows the dataset of MINST trained by GAN.

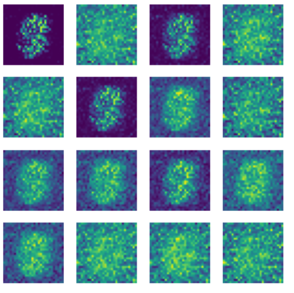

Figure 1. data set

4.2. Experimental setup

The experimental environment is a workstation configured with NVIDIA RTX 4060 GPU, an Intel Core i9-10900K CPU, and 64GB RAM, and the deep learning framework is PyTorch 2.0. The generator consists of a four-layer fully connected network with an input of a random noise vector of dimension 100, and the output is a 28×28 grayscale image. The discriminator consists of a three-layer fully connected network with a 784-dimensional vector (28×28 expansion) as input and a probability value between 0 and 1 as output.LeakyReLU is used for the intermediate layer, Tanh is used for the generator output, and Sigmoid is used for the discriminator output. The stability of the training is enhanced by the introduction of Batch Normalization in the generator. The discriminator, on the other hand, sets the real image label to 0.9 to prevent overfitting operations.

4.3. Training process and results

For each training round (Epoch), the discriminator is first trained to accurately distinguish the real image from the generated image, and then the generator is trained to improve its ability to “deceive” the discriminator, and at the end of each Epoch, the loss of the generator and discriminator is recorded once, and an image is generated using 64 random noise vectors. After each Epoch, the loss of the generator and the discriminator is recorded once, and the image is generated using 64 random noise vectors for visualization, and then the generator is trained for 100 iterations, and the G_loss and D_loss curves of the final training results are shown below.

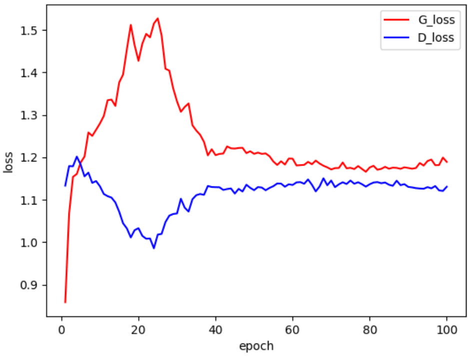

Figure 2. Training results

5. Conclusion

In this paper, a neural network model for handwritten digit image generation is designed and implemented based on the basic theoretical framework of generative adversarial networks. In terms of model structure, both the generator and discriminator are constructed using a multilayer fully connected network. The generator takes random noise as input and outputs images with high similarity to real handwritten digits; the discriminator discriminates the authenticity of the input images. In order to improve the training stability and the quality of the generated images, this paper introduces techniques such as Label Smoothing, Batch Normalization, and uses the Adam optimizer to update the model parameters.

Experiments on the MNIST handwritten digital image dataset show that the constructed GAN model is able to learn the real data distribution more accurately and gradually improve the quality of the generated images. With the increase in the number of training rounds, the image output by the generator gradually tends to be clear and structurally complete from the initial fuzzy and unrecognizable random image, and the final generated result has a certain digital form and realism, which verifies the effectiveness of the model. At the same time, the loss trend of the generator and discriminator during the training process also reflects that the model training has entered a stable state, and there is no obvious abnormal situation, such as pattern collapse.

Although the structure of the model used in this paper is relatively simple, through a reasonable training strategy and optimization method, it still achieves a more satisfactory generation effect, demonstrating the powerful ability of GAN in the image generation task. In future work, we can further introduce convolutional neural networks (CNN) to construct deeper generator and discriminator structures, such as DCGAN, to improve image quality and diversity; we can also explore variants of Conditional GAN (CGAN), Wasserstein GAN (WGAN), and other models to enhance generation control and training stability. In addition, applying GAN to more complex data scenes (e.g., face images, natural scene images, etc.) will also be an important direction for further research.

References

[1] PAN Miaozhong,WANG Yingyi. Dual encoder image forgery detection based on Con-vNeXt-Mamba[J/OL]. Computer Engineering and Applications,1-11[2025-03-26].Hu, T. and Desai, J.P. (2004) Soft-Tissue Material Properties under Large Deformation: Strain Rate Effect. Proceedings of the 26th Annual International Conference of the IEEE EMBS, San Francisco, 1-5 September 2004, 2758-2761.

[2] LI Zhongwei,LI Keyi,LIU Xin. Sequential recommendation algorithm based on data aug-mentation and fused self-attention[J/OL]. Computer Engineering,1-14[2025-03-26].

[3] Z.W. Li, K .Y. Li, X. Liu. Sequential recommendation algorithm based on data augmentation and fused self-attention[J/OL]. Computer Engineering,1-14[2025-03-26]..

[4] Fan Li. Research and application of image data enhancement technology based on gene-ra-tive adversarial network [D]. Zhejiang University,2022.

[5] W.C. Liu, K.W. Guo. A review of research on attack and defense methods for deep neural network visual model adversarial robustness[J]. Network Security Technology and Ap-pli-cation,2025,(01):42-48.

[6] Ren X.W., Wang R., Zhao S., et al. Text Intent Recognition Method Based on VAE-GAN and Improved RoPE[J/OL]. Journal of Shaanxi University of Science and Techno-lo-gy,1-8[2025-03-26].

[7] Xindong Zhao, Chunyu Liu, Chen Wang, et al. A computational imaging method for simple optical system of remote sensing camera based on MsRN-Deblur TransGAN generative adversarial network[J/OL]. Advances in Lasers and Optoelectron-ics,1-29[2025-03-26].

[8] Yinlong Kan, Xuebin Chen, Zhongrui Jing, et al. A data enhancement scheme in federated learning based on conditional generative adversarial networks[J/OL]. Computer Applica-tions,1-16[2025-03-26].