2021 3rd International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI) | 978-1-6654-1790-7/21/$31.00 ©2021 IEEE | DOI: 10.1109/MLBDBI54094.2021.00042

Image Classification and Generation Based on GAN Model

Han Meng†,*
School of Electrical and Electronic Engineering
Huazhong University of Science and Technology
Wuhan, China
* u201912134@hust.edu.cn

Fangru Guo†
Faculty of Science
University of Calgary
Calgary, Canada
fangru.guo@ucalgary.ca
†These authors contributed equally.

***Abstract*—The topic of image processing is becoming more and more popular in the field of artificial intelligence, and it can be applied to fields of biology, medicine, video games, art, and etc. In order to have a deeper understanding of how to optimize the image processing, this paper mainly proposed the Generative Adversarial Network (GAN), which is an emerging deep learning model with the ability to continuously improve modeling under the game, and there are already many applications related to image processing, such as video prediction, 3-dimensional object generation, image super-resolution and etc. In this paper, we mainly implement image generation and image classification based on GAN model. In order to indicate the performance of GAN model in image generation in detail, GAN models with linear layers and with convolution layers are trained and compared based on MNIST datasets. Furthermore, we train GAN model with linear layers, and GAN model with convolution layers, Multi-Layer Perceptron (MLP), Convolutional Neural Network (CNN), and Residual Network (ResNet) models in image classification based on the Mixed National Institute of Standards and Technology database (MNIST), and receive the training loss and testing accuracy of these models for different epochs in image classification. The experimental results demonstrated that GAN model with convolution layers performs best in both image generation and image classification.**

Keywords: Image generation; GAN model; Image classification; Experimental analysis

I. INTRODUCTION

As the most essential part of computer vision, image generation, and classification have always been a hot topic in Artificial Intelligence since it has a wide application to many different fields such as computed tomography in biochemistry, face recognition in security field, and transportation scene recognition in transportation field [1-3]. To satisfy the increasing need in this important task, a great number of excellent neural networks have been proposed and proved effective in a wide range of missions. Szegedy et al. [4], for instance, proposed the idea inception which successfully increased the depth of Convolutional Neural Network (CNN) without the problem of overfitting, and kept the computational budget by utilizing 1×1 convolution kernel. He et al. introduced ResNet with residual block to deal with the problems of network degradation, gradient disappearance, and explosion by employing the idea of identity mapping [5]. All of those previous works have made a huge contribution to image classification and achieved outstanding performance in various tasks. In this paper, we introduce an unsupervised model called Generative Adversarial Network, also known as Generative Adversarial Network (GAN), which is a deep learning model that can continuously improve modeling under the game [6]. It has already been widely used in the field of image generation, and has spun off plenty of relative works which have successful application in different tasks. For instance, in the field of game and Virtual Reality (VR) [7], image generation based on GAN model subverts the working mode of traditional human artists and generates fully real scenes [8]. Moreover, in the field of artistic creation, image colorization, super resolution, and image-to-image translation are all implemented by GAN model [9, 10]. As GAN has done a great job in the field of image generation, we make comparisons between GAN with linear layers and with convolution layers to find out their performances for image generation. We also apply GAN in image classification based on a large database of handwritten digits - Mixed National Institute of Standards and Technology database (MNIST) which contains a training set of 60,000 examples and a test set of 10,000 examples. Then, we compare the performance of GAN model to that of some classic networks such as Multi-Layer Perceptron (MLP), CNN, and Residual Network (ResNet) in image classification based on MNIST datasets.

The main aspects of this work can be listed as below:

- We introduce and analyze the image generation performance of GAN model with linear layers and with convolution layers.
- The image classification performances of MLP, CNN, and ResNet models are compared.

The rest of this work can be summarized as follows. In Section II, we introduce and compare the image generation performance of GAN model with linear layers and with convolution layers in detail, and compare other certain models such as MLP, CNN, and ResNet in terms of image classification. In Section III, the experiments on MNIST dataset are conducted and analyzed. Finally, Section IV concludes this paper.

978-1-6654-1790-7/21/$31.00 ©2021 IEEE

DOI 10.1109/MLBDBI54094.2021.00042

II. Methodology

The work can be mainly divided into three sections. First, we introduce the main structure and mechanism in image generation of GAN model in detail. Furthermore, in order to further evaluate the performance of GAN in image classification, this paper also compares and analyzes the classification performance of other models such as CNN and ResNet-18. Finally, with a little modification on the general GAN model in the first part, we attempt to implement it in a classification task and observe how it works.

A. Image Generation Based on GAN

GAN includes a generative model and a discriminative model. Through training them in an adversarial way, we can get a well-developed generator which can generate a seemingly real image given an input of random noise, and a discriminator which achieves a superior performance in discriminating real pictures from fake ones.

While the purpose of the generator is to fool the discriminator with the fake images it generates, the discriminator is trained to find the real ones. When optimizing the parameters in discriminator, its distribution starts to approach that of the real data. When optimizing the parameters in generator with parameters in the discriminator fixed, the distribution of generated images starts to approach that of the discriminator and it will be harder for the discriminator to tell the difference. After enough iterations, the distribution of the images generated by generator is really close to that of the real images. As a result, two models reach a state of Nash equilibrium, which is a desirable result.

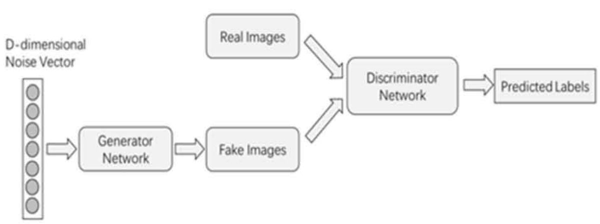

Figure 1. The main structure of GAN model

B. Other Image Classification Model

1) Multi-Layer Perceptron

MLP, also known as Artificial Neural Network (ANN), consists of an input layer, an output layer, and several hidden layers within. Every two layers are fully connected, which means the value of an individual neuron in the latter layer is decided by all the values of neurons in the previous layer. The procedure of forward propagation of each layer is to perform linear operation on each neuron, combine them altogether, and input the result in an activation function to obtain the value of one neuron in the next layer [11].

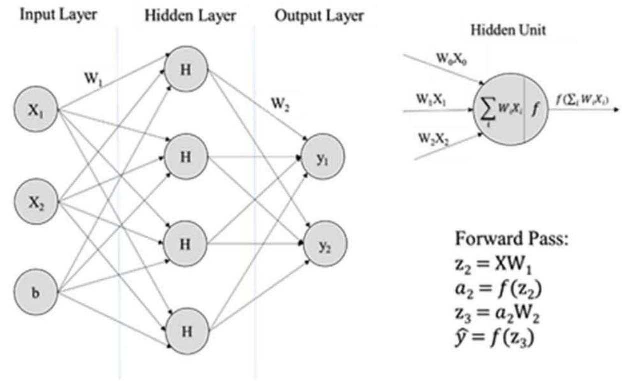

Figure 2. Illustration of MLP

2) Convolutional Neural Network

To better extract the local features in an image, and avoid the problem of over parameterization at the same time, CNN was proposed, characterized by its ability in representation learning and shift-invariant representation classification. A basic CNN consists of three structures, namely convolution, activation, and pooling. By passing the images through several basic blocks constructed of these three fundamental layers, we can extract a set of features lying in them. While dealing with the image classification task, we take this output from CNN as the input in a fully connected layer, to realize the mapping between image features domain and labels domain [12].

3) Residual Network

To deal with the challenging problems of gradient disappearance and network degradation, ResNet employs residual block by adding identity mapping to the convolution layer. By using identity mapping, the derivative of any input is converted from the multiplication of weights to addition, and an extra is added to keep the gradient from reducing to almost zero with the increase of layers. As to network degradation, which is resulted from non-identity mappings in extra layers after the network has already reached an optimal depth, is resolved by residual block since the identity mapping within it can make the extra layers regress faster to a near-identical mapping [13].

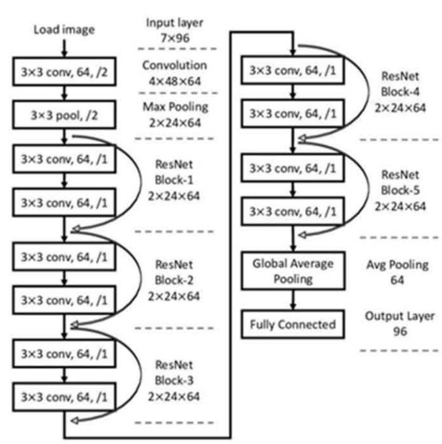

Figure 3. Illustration of ResNet-18

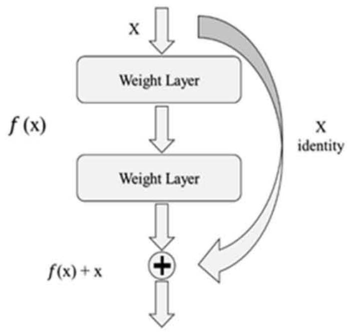

Figure 4. Identity mapping in residual block of ResNet-18

C. Image Classification with GAN

With a well-trained GAN model, we can leverage the discriminator to deal with image classification tasks. Since it has already been trained with the ability to discriminate the images in the training data, its latent space can totally match that of the datasets. With little tuning in the last fully connected layer in the last layer and do some extra training, it can be applied to the field of image classification.

III. EXPERIMENTS

A. Datasets

MNIST is a prevailed database of handwritten digits (28 by 28 pixels) provided by the National Institute of Standards and Technology (NIST), and is widely used in various computer vision tasks. The label for each image represents its number, ranging from 0 to 9.

B. Evaluation Metrics

For image generation, since current evaluation metrics aren't mature enough and cannot be applied to every task, we choose to evaluate the performance by presenting the images directly. Furthermore, considering the simplicity of handwritten digits, the quality of the picture is obvious to bare eyes.

As to image classification, we choose accuracy (*ACC*), the most classic and popular evaluation metrics in classification tasks:

$$ACC = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

where TP, TN, FP, and FN indicate true positive, true negative, false positive, and false negative in the prediction task, respectively.

We also provide cross-entropy loss:

$$H(p,q) = -\sum_i p(c_i)\log q(c_i) \quad (2)$$

where $p(c_i)$ represents target distribution and $q(c_i)$ represents predicted distribution. Since we use one-hot vectors to encode labels, we get

$$Cross\_Entropy(p,q) = -\log q(c_i) \quad (3)$$

where c_i is the class of the training example.

C. Experimental Results and Analysis

In the image generation task, we feed the data into GAN models with linear layers and convolution layers to train the generator and discriminator. We present the images generated by both two in Fig. 5, and find that GAN with convolution layers far outperforms basic GAN in image generation. However, both two GANs suffer from the inability to select the target labels for the generated images. Because we can only input random vectors, they will generate the random numbers as a result. In addition, if the parameters of discriminator and generator in GAN with convolution layers are updated same times in one epoch, it would probably generate the same digits as shown in Fig. 5.

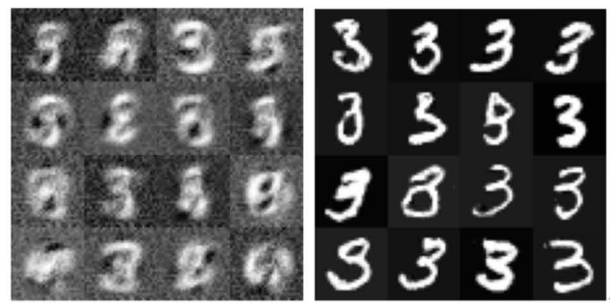

Figure 5. Comparison between images generated by GAN and GAN with convolution layers (Trained for 20 Epochs)

After the adversarial training with discriminator and generator in GAN, we apply the discriminator to the image classification task. Since the original output layer, a linear fully-connected layer, output vectors whose last dimension's size is 1 (discriminate the facticity of the image). We have to do a little modification on its output layer, as extending its output dimension to 10 (number of classes of the digits). Then, we will lock the parameters of the discriminator derived from generation task and train the output layer individually with the same epochs. In Tabs. I-II, we compare the performance of these two GANs with other models mentioned in Section II.

TABLE I. TRAINING LOSS OF DIFFERENT GAN, RESNET-18, MLP, AND CNN MODELS IN DIFFERENT EPOCHS.

Epochs / Models	1	2	5	10	15	20
GAN	326.351	133.076	63.205	26.822	12.044	6.941
GAN (conv)	104.747	28.873	11.822	4.104	1.556	1.398
ResNet-18	84.859	34.705	14.578	7.590	3.329	3.217
MLP	343.109	135.588	68.690	30.006	14.450	7.230
CNN	118.537	33.440	13.019	5.261	3.608	1.675

TABLE II. TESTING ACCURACY OF DIFFERENT GAN, RESNET-18, MLP, AND CNN MODELS IN DIFFERENT EPOCHS..

Epochs / Models	1	2	5	10	15	20

GAN	92.62%	94.33%	96.61%	97.42%	97.75%	97.73%
GAN (conv)	97.98%	98.64%	98.81%	99.07%	99.19%	99.26%
ResNet-18	98.10%	98.25%	98.56%	98.44%	98.79%	99.04%
MLP	92.36%	94.47%	96.50%	97.31%	97.66%	97.58%
CNN	97.75%	98.66%	98.89%	98.88%	98.95%	99.11%

When choosing the same learning rate, optimizer, and other hyper parameters, we find that the performance of GAN with convolution layers has the most desirable result among all other models we select according to both training loss and test accuracy. Noted that we choose the MLP and CNN model with the same layers and depth that of the discriminators of GAN and GAN with convolution except the output layers, the results evidently prove that discriminators that have been trained in image generation task contain more information which can enhance its ability in image classification. Although ResNet-18 is proved as a really strong network in image classification, its performance is not really desirable in MNIST due to the simplicity of the data, especially as the epochs grow. Beyond that, we also notice that when GAN has already been trained adversarially in image generation for many epochs, with only a few extra training epochs for the modified discriminator, its performance in classification may even be better than the normal network that is trained for a lot of epochs.

IV. Conclusion

This paper implements image generation and classification by using the GAN model and compared its performance to some other models such as MLP, CNN, and ResNet-18. In the image generation part, GAN with linear layers and with convolution layers are compared based on MNIST datasets. The output images indicate that the GAN model with convolution layers surpasses the GAN model with linear layers. At the same time, GAN with linear layers and with convolution layers, MLP, CNN, and ResNet-18 models are trained in order to get the training loss and testing accuracy in image classification. The experimental result shows that the performance of the ResNet-18 model contains the most efficient and accurate data when the epochs are low. However, with the growth of epochs, GAN with convolution layers performs better than the ResNet-18 model. It is concluded that the GAN model with convolution layers can efficiently and accurately realize image generation and image classification.

Image processing is a subject that full of challenges and values, and there are still a lot of spaces for research. In the future, we will continue to study some practical applications of the GAN model in various fields in reality. For example, unsupervised anomaly detection on complex data sets can play an important role in the medical field. Furthermore, we will also study the application of GAN in music fields: music generation, which can predict the next chord trend based on the previous melody.

References

[1] Q. Song, L. Zhao, X. Luo, and X. Dou, "Using deep learning for classification of lung nodules on computed tomography images," Journal of Healthcare Engineering, vol. 2017, 2017.

[2] K. W. Bowyer, "Face recognition technology: security versus privacy," IEEE Technology and Society Magazine, vol. 23, no. 1, pp. 9-19, 2004.

[3] S. Wang, S. Yao, K. Niu, C. Dong, C. Qin, and H. Zhuang, "Intelligent scene recognition based on deep learning," IEEE Access, vol. 9, pp. 24984-24993, 2021.

[4] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1-9, 2015.

[5] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770-778, 2016.

[6] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," Association for Computing Machinery, vol.63, no. 11, pp. 139-144. 2020.

[7] C. Yao and J. Xiao, "VR image generation based on aesthetical rules," 2008 Congress on Image and Signal Processing, pp. 806-810, 2008.

[8] D. Lee and H. Choi, "Text to game characterization: a starting point for generative adversarial video composition," 2018 IEEE International Conference on Big Data and Smart Computing (BigComp), pp. 717-720, 2018.

[9] F. C. Silva, P. André Lima de Castro, H. R. Júnior, and E. Cordeiro Marujo, "Mangan: assisting colorization of manga characters concept art using conditional GAN," 2019 IEEE International Conference on Image Processing (ICIP), pp. 3257-3261, 2019.

[10] B. Chen, T. Liu, K. Liu, H. Liu, and S. Pei, "Image super-resolution using complex dense block on generative adversarial networks," 2019 IEEE International Conference on Image Processing (ICIP), pp. 2866-2870, 2019.

[11] P. Zhang, E. Patuwo, and, M. Hu. "Forecasting with artificial neural networks: the state of the art," International Journal of Forecasting, vol. 14, pp. 35-62, 1998.

[12] S. Lawrence, C. L. Giles, A.C. Tsoi, and A. D. Back, "Face recognition: a convolutional neural-network approach," IEEE Transactions on Neural Networks, vol. 8, no. 1, pp. 98-113, 1997.

[13] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770-778, 2016.