DOI: 10.54254/2755-2721/49/20241085

Investigation the influence related to parameters configuration of Generative Adversarial Networks in face image generation

Yuebo He

Department of Computer Science, Faculty of Science, Western University, London, ON, N6A 3K7, Canada

Yhe534@uwo.ca

Abstract. Due to the excellent performance of Generative Adversarial Networks (GAN) for age regression on face images, it is particularly important to explore the effect of different parameters on model training. In this study, the origin and development of Artificial Intelligence (AI) is first discussed, from which the concept and principles of GAN are derived. This is followed by a brief introduction of the UTKface dataset used in this research, and the Conditional Adversarial Autoencoder (CAAE) framework based on the GAN technique. The division of labor and roles of the encoder, generator, and the two discriminators in the model are described. The various learning rates as well as batch size combinations attempted in this study are then illustrated, and the training results of the model are shown in the form of graphs and plots of the loss value function. A situation where the model stops learning is highlighted in the results, which is similar to pattern descent in GAN, and is shown to be characterized by the inability of the discriminator to successfully recognize it. Ultimately, drawing from the acquired outcomes, it can be deduced that employing a larger batch size serves to enhance the pace of model training. It is advisable to concurrently elevate the learning rate by an equivalent factor when augmenting the batch size, thereby ensuring a consistent trajectory for model convergence.

Keywords: Artificial Intelligence (AI), Generative Adversarial Network (GAN), Learning Rate, Batch Size.

1. Introduction

With the advancement of computer hardware, an increasing number of statistical theories can be realized, and a new branch of computer science Machine Learning (ML) has been subdivided. ML is the process of finding associations and building models in data by a computer through certain learning algorithms [1]. This process is like humans learning through experience, except that computers learn faster and rely on the "experience" provided to them. Along with the introduction of Graphical Processing Units (GPU) in ML training, humans were able to implement Artificial Neural Networks (ANN) on computers using biological neural networks as templates. The computer autonomously generates a large number of neurons during training, and these neurons are linked in a certain structure to form a neural network that can mimic human beings in making decisions and judgments. Utilizing artificial neural network technology, deep learning has emerged as a crucial component within the

DOI: 10.54254/2755-2721/49/20241085

diverse landscape of machine learning techniques. The application of this method is widespread in many different fields, including but not limited to picture restoration, natural language processing, and image recognition [2, 3]. The concept of Generative Adversarial Networks (GAN) as a specific method in deep learning, was first proposed by Juergen Schmidhuber in 1991, and this model is characterized by the simultaneous training of two neural networks, one of which is a generator and the other a discriminator. The discriminator must accurately discriminate between genuine images and created images, whereas the generator tries to produce images to "trick" the discriminator [4]. GAN has been heavily utilized for picture generation in recent years, in 2017 a GAN-based implementation of Age Progression / Regression on face images was proposed by Zhang et al. In their proposed method, there are two adversarial networks called encoder and generator. The two discriminators are trained simultaneously and the performance of encoder and generator is analyzed separately [5]. By using UTKFace as a dataset, the face images are processed into individual Z-vectors with an encoder and an attempt is made to have the generator reduce the Z-vectors into images. Based on their findings, it can be observed advantages of this model for face image generation. However, the effect of different parameters on model training and results is missing in the study. Furthermore, in recent years there has been no effective answer to the setting and adjustment of parameters. Especially for the learning rate and batch size, small changes in these two parameters can have a huge impact on the training effect [6]. Therefore, the settings of different learning rates and batch sizes need to be further explored and confirmed.

This study focuses on analyzing the effect of different learning rates and batch size on training in the implementation of face image generation using GAN techniques. The number of training epochs, the model's loss value, its convergence speed, and an examination of the generated images will be used to gauge the model's efficacy and react to the influence of the parameters on the outcomes. Additionally, this study seeks to identify the patterns that influence model training under various parameter settings and investigates the ideal learning rate and batch size settings for the application's model.

2. Method

2.1. Dataset and preparation

The dataset that has been used in this study is UTKFace, a dataset that consists of more than 20,000 different face images collected from the Internet. It contains different age spans (ranging from 0 to 116 years old), different poses, expressions, light effects illumination, occlusion, facial decorations, resolution, and other huge differences. These images are labeled and categorized by age, gender, and ethnicity through the DEX algorithm [5, 7]. This face image dataset that includes a variety of features is well suited for training machine learning models for face detection, age estimation, etc. Some sample images can be found in Figure 1.

Figure 1. Selected examples of photos from UTKface, each with its own unique elements [5].

DOI: 10.54254/2755-2721/49/20241085

These face images are cropped to 128×128×3 size and are based on the RGB color gamut. Before training, a random portion of the images are separated to form a validation set and to demonstrate the training effect of the model, to avoid overfitting of the model to the validation set images.

2.2. CAAE in GAN

The model structure used in this study is the Conditional Adversarial Autoencoder (CAAE) framework shown in Figure 2 implemented based on GAN technology. It contains an encoder, a generator (decoder), and two discriminators, and this structure introduces Variational Autoencoder (VAE) a machine learning based autoencoding function based on GAN [5]. That is to say in this study two adversarial networks were trained, one for adversarial training of the encoder and the other for adversarial training of the generator. The general concept is that the convolutional encoder will turn the facial picture into a vector, and the anti-convolutional generator will turn the vector back into the original image. During this procedure, the model will learn how to effectively encode, maintain, and restore the image's features in order to trick the discriminator [8].

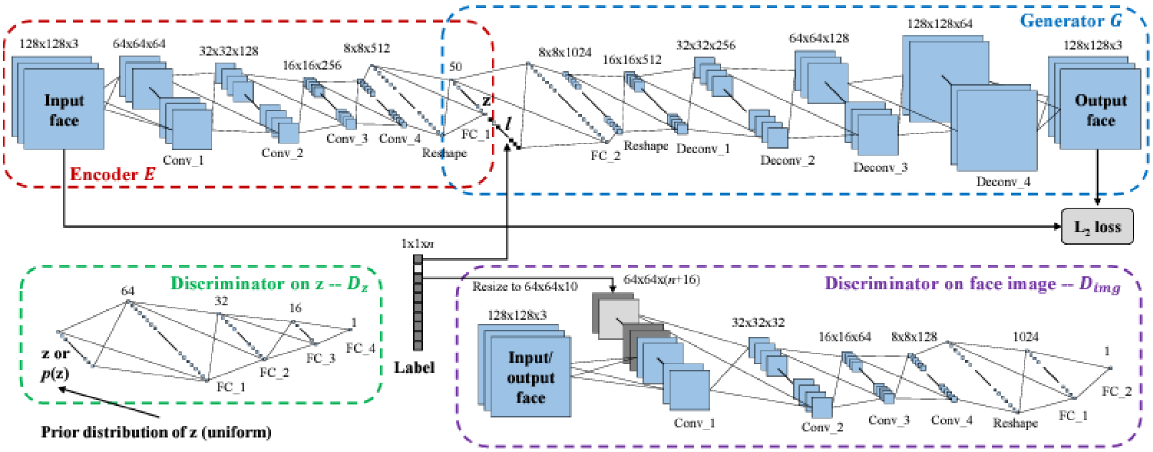

Figure 2. The overall logic of the CAAE structure [5].

2.2.1. Encoder

Four convolutional layers and one fully connected layer make up the encoder, which consecutively applies mapping to compress the 128×128×3 input face picture data. Through the processing of the convolutional and link layers, the encoder will output a Z-vector of size 1×1×50, which contains the feature information of the original image. Compared to existing GAN which uses the introduction of random noise to disrupt the original image, CAAE's encoding process can better preserve the specific features of the image and the original data [9].

2.2.2. Generator

The generator is responsible for taking the Z-vector output from the encoder and re-passing it through the inverse convolutional layer to output a new RGB face image of size 128×128×3. It works more similar to the decoder in VAE, decoding the Z vector and the product is the image.

2.2.3. Discriminator – Z

It is mainly responsible for accepting the Z vector generated by the encoder and supervising the encoder achieving uniform distribution. The encoder is set to generate some vectors that confuse the discriminator Z. Without the introduction of the discriminator Z, the encoder produces vector data that is not homogeneous enough to encode the features of the original image well enough to be recorded. This situation will lead to the generator not getting effective data for training, which will affect the model performance.

2.2.4. Discriminator – IMG
The discriminator-img works similarly to the discriminator in a traditional GAN. It is responsible for distinguishing whether the image output by the generator is "machine generated" or not. A strong discriminator will help motivate the generator to try to produce results that are finer and fit the original image.

2.3. Implementation details
This study experimented with different parameters in the above mentioned CAAE model architecture and observed the impact. Based on the available research data, this study set the baseline parameters as: learning rate = $2 \times e^{-4}$; batch size = 64; epoch = 100. Base of that, the learning rate changed to $1 \times e^{-4}$, $4 \times e^{-4}$, $1 \times e^{-5}$ and $1.5 \times e^{-4}$. The batch changed to 128 and 32. Epoch's performance based on the training of different parameters is set to 10, 20 and 50. Eventually, this study recorded the change of the loss value and the validation set of the model after the model finishes training.

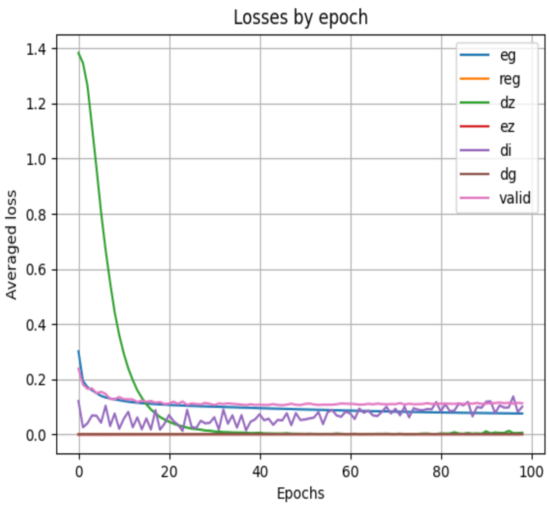

Figure 3. Plot of each loss function with epoch.

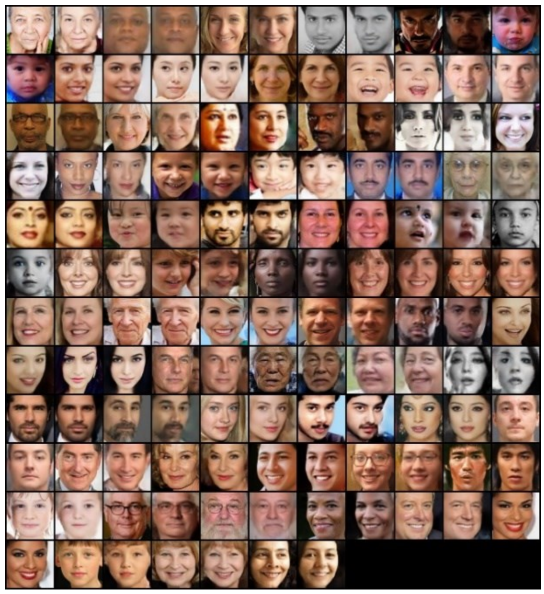

Figure 4. Validation set output of the model.

These two figures (Figure 3 and Figure 4) show the results produced by the model under the basic parameters. Where the blue line "eg" in the left graph indicates the loss value between the original and generated images, and the pink "valid" indicates the loss value for the validation set. The green "dz" and purple "di" indicate the performance of the discriminator Z and the discriminator img, respectively. The validation set's original and generated images are shown side by side on the right, with the original image on the left and the model-derived image on the right, for comparison.

3. Results and discussion
As shown in Table 1, result ID 3 is the outcome achieved using the standards established in the body of existing literature. The studies with various parameters set for a limited number of epochs produced ID 10-15. The studies with various values for longer epochs produced ID 16-21. The results of this time with ID 14 appeared to be dramatically different from the other tests, resulting in values that were not in the same range as the other tests, and this test will be discussed below as a special case. The final data for outcome ID 19, 21 produced no change compared to the benchmark results, but the epoch used was half of the original parameters. With the same batch size of 128 and the learning rate successively decreased, the end data for the three tests with result IDs 16, 17, and 18 did not significantly differ, but the overall loss value in the test with the lowest learning rate was somewhat higher than the other two at 50 epochs.

DOI: 10.54254/2755-2721/49/20241085

From the linear plots in Figure 5, it can be exhibited that the linear variations of the resultant ID 16, 17 and 21 are similar to the linear plots of the baseline parameters 3, which helps in analysing the search for the optimal parameters. Whereas the line as a function of the loss value decreases very slowly in the test with result ID 20, this is the plot of the results using the smallest learning rate of all the tests.

Table 1. Data results for different parameters.

Result ID	epoch	learning rate	batch size	loss value	eg	dz	di	valid
3	100	$2 \times e^{-4}$	64	0.0817	0.0758	0.0063	0.1271	0.1165
10	10	$1 \times e^{-4}$	64	0.1263	0.1245	1.0049	0.0241	0.1208
11	10	$2 \times e^{-4}$	128	0.1371	0.1381	0.7722	0.0179	0.1556
12	10	$4 \times e^{-4}$	128	0.1441	0.1488	0.1976	0.0351	0.1531
13	10	$1.5 \times e^{-4}$	128	0.1387	0.1399	0.9363	0.0191	0.1402
14	13	$4 \times e^{-4}$	128	0.3454	0.3455	0.4429	0.0008	0.3484
15	20	$4 \times e^{-4}$	128	0.1196	0.1169	0.0396	0.2041	0.1314
16	50	$4 \times e^{-4}$	128	0.0991	0.0998	0.0011	0.0872	0.1113
17	50	$2 \times e^{-4}$	128	0.0998	0.0997	0.0068	0.0102	0.1046
18	50	$1 \times e^{-4}$	128	0.1021	0.1039	0.0981	0.0107	0.1071
19	50	$1 \times e^{-4}$	32	0.0874	0.0855	0.0137	0.0616	0.0994
20	31	$1 \times e^{-5}$	32	0.1211	0.1268	1.3472	0.0119	0.1304
21	50	$2 \times e^{-4}$	32	0.0845	0.0884	0.0144	0.0668	0.1014

According to the above results indicated that different learning rates and epochs have relatively more impact on the results of model training than batch size. For result ID 20 alone, too small a learning rate severely reduces the learning speed of the model. It may be confirmed that the learning rate exhibits a positive correlation with the variety of batch size since the data and linear variation charts of results ID 16, 19 are very consistent with the benchmark data plots. This means that, given a benchmark batch size and learning rate, and the learning rate is varied by a factor of several, along with the batch size, and the resulting curves will resemble the benchmark curves as a result. The model will therefore achieve comparable performance.

Based on the data results of result ID 14, it can be observed that the discriminator in the model is not able to correctly discriminate the generating graph, which causes the generator to stop learning. This behaviour is similar to the well-known Mode drop in GAN, which occurs very occasionally and has a probability of occurring when experiments are performed with the same parameters [10].

4. Conclusion

Based on the CAAE model with a GAN-centered method, an attempt has been made in this study to investigate how parameters affect results and the best parameter settings for age regression of face photos. In the tests, several combinations of learning rate and batch size were tested, and the models were trained using various epoch settings. The experimental findings show that a bigger batch size enables the model to converge more quickly, greatly lowering the model's training time. Despite the loss in final accuracy, the convergence curve of the model will remain largely unchanged as long as the learning rate is guaranteed to keep increasing by the same multiple as the batch size. This helps to speed up model training when hardware is limited. Although this study found some of the effects of each parameter on model training, more research is still needed to confirm the specific parameter settings. Meanwhile, with the improvement of hardware performance, small batch size training may be able to give the model higher accuracy and generalization ability, and these aspects need to be investigated more deeply in the future.

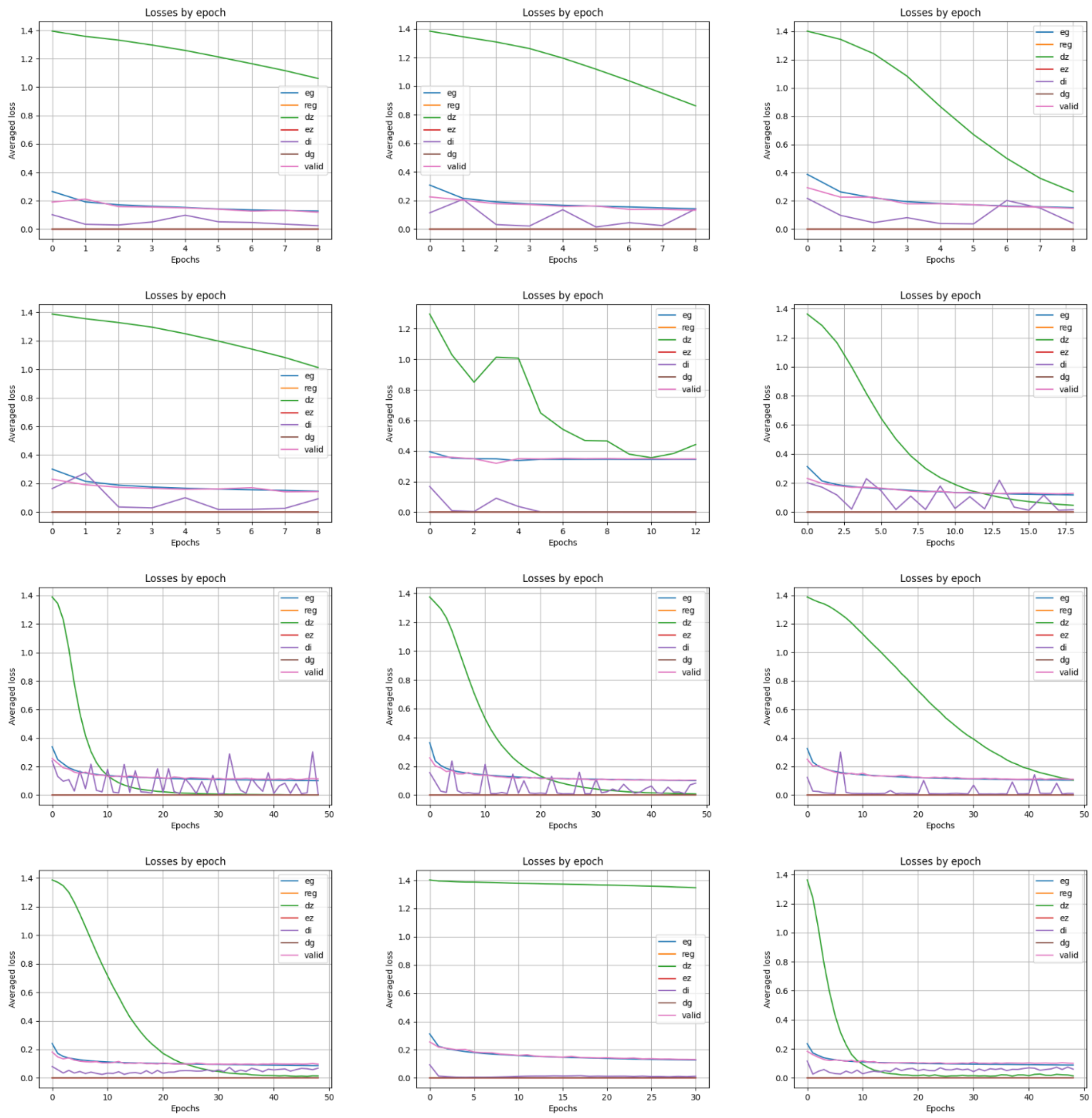

Figure 5. Variations in loss values for 12 distinct parameters. summarized and organized according to the result ID from top to bottom and from left to right.

References

[1] Zhou Z H 2021 Machine learning (Springer Nature)
[2] Salehi A W et al 2023 A Study of CNN and Transfer Learning in Medical Imaging: Advantages, Challenges, Future Scope. Sustainability 15(7) 5930
[3] Qiu Y et al 2020 Improved denoising autoencoder for maritime image denoising and semantic segmentation of USV China Communications 17(3) 46-57
[4] Schmidhuber J 2020 Generative adversarial networks are special cases of artificial curiosity 1990 and also closely related to predictability minimization 1991 (Neural Networks) chapter 127 pp 58-66
[5] Zhang Z Song Y and Qi H 2017 Age progression/regression by conditional adversarial autoencoder (IEEE conference on computer vision and pattern recognition) pp 5810-5818
[6] He F Liu T and Tao D 2019 Control batch size and learning rate to generalize well: Theoretical and empirical evidence (Advances in neural information processing systems) chapter 32

DOI: 10.54254/2755-2721/49/20241085

[7] Rothe R Timofte R and Gool L V 2015 IMDB-WIKI-500k+ face images with age and gender labels

[8] Goodfellow I Pouget-Abadie J Mirza M Xu B Warde-Farley D Ozair S and Bengio Y 2014 Generative adversarial nets (Advances in neural information processing systems) chapter 27

[9] Yu X and Porikli F 2016 September Ultra-resolving face images by discriminative generative networks (European conference on computer vision) pp 318-333 Cham: Springer International Publishing

[10] Yazici Y Foo C S Winkler S Yap K H and Chandrasekhar V 2020 October Empirical analysis of overfitting and mode drop in gan training (2020 IEEE International Conference on Image Processing) pp 1651-1655 IEEE