

# An improvement on common optimization methods based on SuperstarGAN

**Songang Liu**

Internet of Things Engineering Department, Beijing University of Posts and Telecommunications, Peking, 100102, China

liusongang@bupt.edu.cn

**Abstract.** Image processing has long been a focal point of research, offering avenues to enhance image clarity and transfer image features. Over the past decade, Generative Adversarial Networks (GANs) have played a pivotal role in the field of image conversion. This study delves into the world of GANs, focusing on the SuperstarGAN model and its optimization techniques. SuperstarGAN, an evolution of the well-known StarGAN, excels in multi-domain image-to-image conversion, overcoming limitations and offering versatility. To better understand its optimization, this study explored the effects of different optimizers, such as Adam, SGD, and Nadam, on SuperstarGAN's performance. Using the CelebA Face Dataset with 200 million images and 40 features, this study conducted experiments to compare these optimizers. The results revealed that while SGD and Nadam can achieve comparable results to Adam, they require more iterations and careful tuning, with SGD showing slower convergence. Nadam, with its oscillatory nature, shows promise but requires proper learning rate adjustments. This research sheds light on the critical role of optimizer choice in training SuperstarGAN. Adam emerges as the most efficient and stable option, but further exploration of Nadam's potential is warranted. This study contributes to advancing the understanding of optimization techniques for generative adversarial networks, with implications for high-quality facial image generation and beyond.

**Keywords:** StarGAN, CelebA, Image to Image, Optimizer.

## 1. Introduction

As the most direct information carrier, image processing has always been a popular research object for many researchers, such as making the image clear or transferring some features of the image. Generative adversarial networks (GANs) have been an important technique in image conversion area in recent ten years.

There are many derivative techniques of Gans, and their improvements have led them to focus their work on different areas of image processing. The Conditional Generative Adversarial Network (cGAN) creatively uses three discriminators to render realistic facial expressions back to the original facial expressions for expression recognition purposes, and it can also accomplish the work of reducing specks in optical coherence tomography (OCT) [1].

CycleGAN is an improvement on traditional Gans that uses two generators and discriminators for image-to-image conversion, which allows it to have each generator convert an image to display features. However, in the face of multiple domains of image to image conversion tasks. For example, using the

CelebA dataset, it required training 780 models to satisfy the 40 domains of the dataset. StarGAN is the dominant model in this field, which requires only one model to complete the image-to-image conversion of multiple domains. While StarGAN also has the limitation of not being able to learn secondary features, after the introduction of ControlGAN's framework, SuperstarGAN was born, which is undoubtedly a top student of multi-domain image-to-image conversion.

The emergence of StarGAN corresponds to people's needs for image conversion. It not only breaks through the barrier of multi-domain conversion but also allows images to be converted from one domain to another, which is very useful for domain adaptive and transfer learning tasks. StarGAN addressed domain adaptation and style transfer challenges but had limitations. These limitations led to exploration for improvements. SuperstarGAN, with advanced capabilities in facial image generation and attribute manipulation, represents progress in the field. However, ongoing evaluation is needed to refine its strengths and meet complex task demands. There is still a lot of room to explore the application side of SuperstarGAN.

The present study delves deep into the intricacies of SuperstarGAN. Two paramount aspects under scrutiny encompass the intricate interplay between initial learning rate configurations and the optimization methodologies that underpin the SuperstarGAN framework. Given the substantial influence of StarGAN's architectural lineage and configuration on SuperstarGAN's foundation, an exhaustive exploration becomes essential to understand the nuances of integration and adaptation. At the same time, considering the complex premise of learning rate setting, this paper only explores the profound impact of model optimization methods on the model.

## 2. Method

### 2.1. Dataset description and preprocessing

In the selection of data sets, the widely known CelebA face dataset was used for experiments in this study. It is a public face dataset containing more than 200,000 photos of human faces from more than 10,000 celebrities. The mode of images in the CelebA dataset are based on the RGB, almost all of which have a resolution of  $178 \times 218$  pixels or higher. Figure 1 presents the sample images of the collected dataset. Simultaneously, this dataset possesses a characteristic highly conducive to image processing, with each image being accompanied by an extensive set of binary labels representing prevalent facial attributes. Due to these attribute tags, it is easy to identify key information such as gender and age in this dataset. Another consideration for the use of the CelebA dataset in this article is to maintain consistency with the datasets used in the StarGAN and SuperstarGAN experiments for the purpose of controlling variables.



**Figure 1.** The sample images on the collected dataset.

### 2.2. GAN model

#### 2.2.1. StarGAN

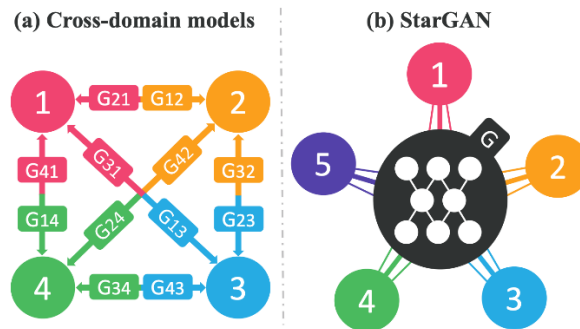
The basic design idea of StarGAN is to design a single GAN and make it complete the multi-domain image-to-image conversion task by itself. While traditional approaches might require training a separate

model for each domain, StarGAN attempts to combine these tasks into a single model, resulting in greater efficiency and versatility. The model generates an image of the domain by input conditional information, such as the label of the target domain [2].

In the context of StarGAN, three essential components are present: a generator, a discriminator, and an attribute classifier. The generator's primary role involves transforming the input image into an image representative of the target domain. It relies on conditional information as input, facilitating the determination of style and domain attributes within the generated image.

The discriminator's function, on the other hand, centers on the distinction between generated images and genuine images. This process serves a dual purpose. First, it enhances the generator's ability to produce high-quality, lifelike images through adversarial training. Secondly, in the context of StarGAN, the discriminator takes on an additional role in multi-domain transformation. Specifically, it is responsible for identifying the domain associated with the generated image, thereby ensuring alignment with the intended target domain [3].

StarGAN shown in Figure 2, conceived as a solution to address the limitations of CycleGAN, draws inspiration from the concept of ACGAN (Auxiliary Classifier GAN). It effectively incorporates conditional input within the generator and auxiliary classifier within the discriminator. Consequently, StarGAN demonstrates remarkable efficiency in executing image-to-image conversions across multiple domains while utilizing a single generator-discriminator pair.



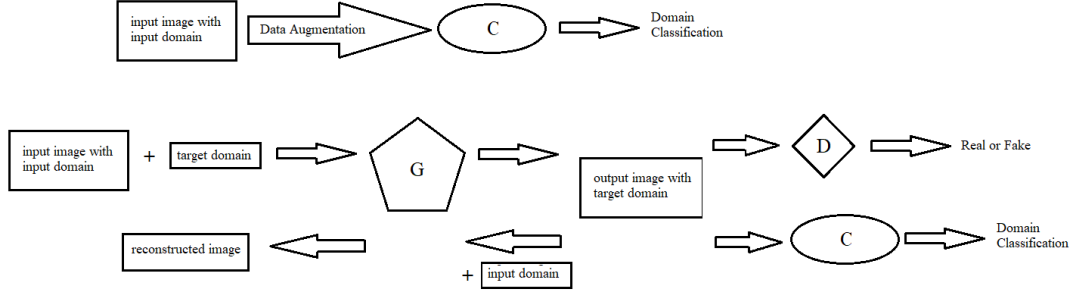
**Figure 2.** StarGAN is designed differently from ordinary Gans [3].

### 2.2.2. SuperstarGAN

SuperstarGAN shown in Figure 3 represents a significant advancement over StarGAN, effectively addressing some of its limitations. Notably, SuperstarGAN excels in learning complex mappings between large-scale domains, a challenge that StarGAN struggled with. Moreover, it overcomes the constraint of expressing subtle variations in features that was beyond StarGAN's capabilities.

One of the key limitations of traditional StarGAN lies in its use of ACGAN discriminators. These discriminators suffer from a tendency to overfit the training data, making them less effective at conveying essential information for training generators. This issue arises from the inherent integration of the classifier and discriminator in ACGAN's structure.

SuperstarGAN adopts an innovative approach inspired by ControlGAN. It introduces data augmentation techniques to train an independent classifier specifically designed to process StarGAN. This strategy empowers SuperstarGAN to effectively capture fine-grained features belonging to the target domain, enabling image-to-image conversion at a large-scale domain level. Despite these improvements, SuperstarGAN retains the core modules present in StarGAN, including the generator, discriminator, and domain discriminator [4].



**Figure 3.** SuperstarGAN model structure diagram

### 2.3. Optimizer

In the previous SuperstarGAN experiment, the Adam optimizer was used by the model. In this paper, it is believed that Adam was selected in the previous experiments for more consideration of convenient comparison with stargan to control the test performance of variables, and it does not mean that Adam is the optimizer with the best performance for SuperstarGAN. Nadam and SGD, for example, are better than Adam in some use cases and are worth considering.

## 3. Results and discussion

The experiment used the CelebA Face Dataset, which contains 200 thousand facial images, each with 40 features to train on. The data specifications used for training are shown in Table 1 and Table 2 [5, 6].

The research method adopted in this paper is to change the technology of the optimizer and train the model with common parameters in the case of fixed variables, and observe the change rate of generator loss data generated by training in iteration and the picture example of model output to judge the advantages and disadvantages of the optimization method.

When only SGD is used instead of Adam as the optimizer. By observing the training data, it can be found that the initial loss of the model is high, the convergence speed of the model is significantly reduced, and the local minimum will appear [7]. When using SGD, the model essentially took four times as many iterations and times to reach the same level as Adam. Considering the typical shortcomings of SGD, it is clear that SGD lacks a self-learning mechanism and requires more iterations to achieve a comparable level of performance compared to Adam algorithms. In general, optimization with SGD requires a more nuanced selection of learning rate and rate decay strategies compared to Adam, all with the aim of improving stability. When testing the SGD algorithm, momentum is set to 0.9, which is a relatively general parameter, but it is probably not optimal.

When using only Nadam without adam as the optimizer, the training process and results become more optimistic compared to SGD, but the overall fitting speed is still not as good as Adam. Given the NADAM's oscillatory nature, it is impossible to obtain optimal test results without adjusting the learning rate to accommodate the NADAM. Nadam has the potential to outperform Adam as the best optimization technique for SuperstarGAN. Nadam also requires less hyperparameter tuning. Considering that SuperstarGAN is a deep learning model, Nadam has its own unique advantages.

Figure 4 is the sample image generated by Adam and SGD during 2000 iterations. By looking at the image, it can be found that the sample image generated by the two optimization methods has a big gap in the clarity of the face image. It can be said that Adam has obvious advantages compared with SGD optimization method in SuperstarGAN. It can already produce highly realistic images in a few iterations [8]. In addition, attention module may be considered for further improving the performance in different situations due to their excellent performance in various tasks [9, 10].

**Table 1.** Generator total loss

C/loss	Adam	SGD	Nadam
500 iterations	15.1630	20.0141	16.4725
1000 iterations	12.9426	18.1839	13.8967
2000 iterations	12.7871	16.2811	14.0875

**Table 2.** Datasets used in the experiments [4].

Name	Num.of samples	Num.of labels	Average Num.of samples per label	Label category	Training iteration
CelebA	200,599	40	5015	Multi-lable	2,000



**Figure 4.** Sample images generated by SGD (left) and Adam (right) in 2000 iterations

#### 4. Conclusion

In conclusion, the experiments conducted using the CelebA Face Dataset with 200 thousand facial images and 40 features for training have provided valuable insights into the optimization methods for SuperstarGAN. The focus was on comparing the performance of the SGD and Nadam optimizers against the commonly used Adam optimizer while keeping other parameters fixed. The findings highlight that while SGD can achieve comparable results to Adam, it requires significantly more iterations and careful tuning of learning rates and decay strategies to achieve stability. It tends to suffer from slower convergence and a higher risk of getting stuck in local minima. In contrast, Nadam shows promise as an alternative optimizer, offering improved results compared to SGD and requiring less hyperparameter tuning. Its unique advantages make it a strong contender for optimizing SuperstarGAN, particularly given the model's depth and complexity. The choice of optimizer plays a crucial role in the training of SuperstarGAN, with Adam proving to be the most efficient and stable option in these experiments. However, further exploration of Nadam's potential is encouraged, as it may offer a promising avenue for improving the optimization process. Overall, this research contributes to a better understanding of optimization techniques for generative adversarial networks and their application in generating high-quality facial images.

## References

- [1] Deng J et al 2019 cGAN based facial expression recognition for human-robot interaction. IEEE Access 7 9848-9859
- [2] Choi Y et al 2020 Stargan v2: Diverse image synthesis for multiple domains. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition pp. 8188-8197
- [3] Choi Y et al 2018 Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 8789-8797).
- [4] Ko K et al 2023 Superstargan: Generative adversarial networks for image-to-image translation in large-scale domains. Neural Networks, 162, 330-339.
- [5] Kameoka H et al 2018 Stargan-vc: Non-parallel many-to-many voice conversion using star generative adversarial networks. In 2018 IEEE Spoken Language Technology Workshop (SLT) (pp. 266-273). IEEE.
- [6] Anoosheh A et al 2018 Combogan: Unrestrained scalability for image domain translation. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops (pp. 783-790).
- [7] Arjovsky M et al 2017 Wasserstein generative adversarial networks. In International conference on machine learning (pp. 214-223). PMLR.
- [8] Goodfellow I et al 2014 Generative adversarial nets. Advances in neural information processing systems, 27.
- [9] Qiu Y et al 2022 Pose-guided matching based on deep learning for assessing quality of action on rehabilitation training. Biomedical Signal Processing and Control, 72, 103323.
- [10] Zhou Q et al 2023 Medical image classification using light-weight CNN with spiking cortical model based attention module. IEEE Journal of Biomedical and Health Informatics, 27(4), 1991-2002.