Study on Deep Generative Models
Generative Adversarial Networks

A forward study of anime face generation

Vivian Sedov

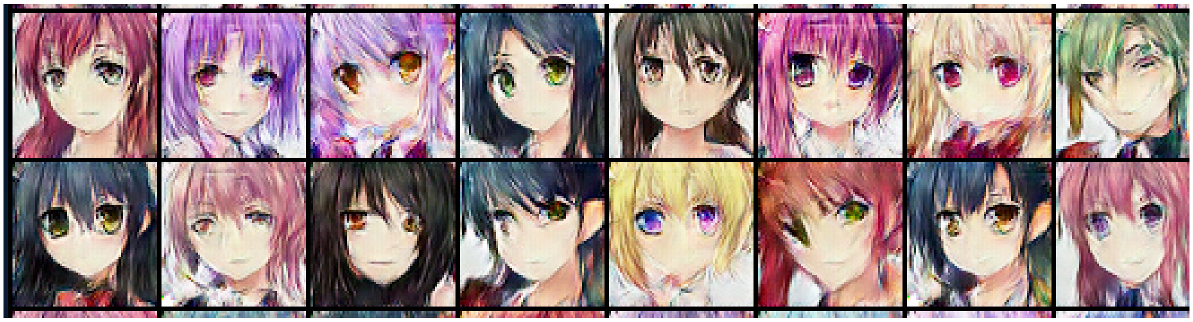
Final Report
BSc (Hons) in Computer Science and Artificial Intelligence
Supervised By: Dr Yunkuen Cheung

Department of Computer Science
Royal Holloway, University of London

March 2023

Contents

List of Figures



Abstract

Since the introduction of generative networks, image generation has made remarkable progress, particularly in the field of facial construction. Ian Goodfellow's introduction of Generative Adversarial Networks (GANs) [2] in 2014 opened up new possibilities for generating realistic images. In recent years, many attempts have been made to use GANs for generating anime faces. However, generating anime faces poses unique challenges due to their stylized and cartoon-like features. Unlike realistic human faces, anime faces often have exaggerated features, such as large eyes and bright-colored hair, that can be difficult to represent accurately using traditional image generation techniques. Additionally, anime artists may use a wide range of styles and techniques, which can make it challenging to develop a generalizable model for anime face generation.

To address these challenges, we propose using active label synthesis and feature labeling tools such as Illustration2Vec to identify and tag features of anime characters. This approach allows us to compile a more accurate and well-suited dataset for training GAN models. We will also use unsupervised learning techniques to apply GANs and evaluate the effectiveness of different generative models. By combining data and model perspectives, we aim to quantitatively analyze facial production and generate high-quality anime faces that are faithful to their original style and characteristics [5].

The proposed study aims to generate anime faces using unsupervised learning with GANs and to compare the effectiveness of different generative models. We will discuss the operation and architecture of several generative models. Such that in this paper we will discuss the integration of ilustrationtovec with respect to Active label Synthesis, and a applied auxiliary classifier generative adverserial network (AC-GAN) [9]. Such that the whole process is automatic and creditable with its art style.[1] [15] [13].

Generative models, Unsupervised learning, Generative Adversarial Network, Machine Learning

March 17, 2023

1 Introduction

The fast growth of Japanese animation industry has resulted in the production of many animated shows that have piqued the attention of diverse audiences. Each anime has a unique trait that caters to a distinct users taste. Such that the idea of creating our custom ones come to mind. However, it takes tremendous efforts to master this skill, to draw and generate concise facial data. To bridge this gap, we have the principle of automatic generation of characters based on prior datasets. Due to the powerful ability of generative models, GANs have achieved great success in manipulating and generating images, with up-scaling and image generation being prime examples.

There are three essential requirements that must be met for the development of anime faces to ensure the success of this endeavour. I) Produced faces must be those of female and male faces, with a discernible difference between them; II) Each face must have unique traits that are sufficiently distinct from the initial training batch; III) The quality of the generated faces must be high[5]. To verify that these standards are adhered to, I will quantify the variety of the produced and input data using the following techniques: By having a suitable way to evaluate the performance of the produced faces, would imply the comparison between the generated photos towards the present data set. Various approaches, such as mean squared error or a similarity index based on the entropy of the produced pictures, may be used to quantify this. To confirm the image's viability, we may assess the quality of created pictures using the Inception Score *The Frechet Inception Distance (FID) : It measures the distance between two distributions, in this case the distribution of generated images from a GAN and the distribution of real images.*

Parametric models have shown great promise in the field of facial generation due to their ability to create new and unique faces by manipulating a set of parameters. However, there are inherent challenges in designing parametric models that can capture the complexities and diverse range of facial structures, especially in the context of anime faces. One significant challenge is the non-linear nature of facial structures, which can lead to inconsistencies and artefacts when generating new faces. Additionally, the high dimensionality of the feature space can make it difficult to control the generated faces' specific attributes, such as hair and eye color [6]. However, despite the availability of large datasets such as Danbooru2021, the inherent challenges of dataset dependability, quality, and biases still persist. Ensuring that the dataset is accurately labeled and free from erroneous or inconsistent information is critical, as GANs can be notoriously difficult to train. Another challenge that remains is the accurate generation of both male and female anime faces, as gender distinction is a crucial aspect of generating distinct and unique characters.

Various efforts have been made in the literature to generate anime facial data, with early studies by Alec Radford [11], Mattya [7], and Rezoolab [12] focusing on the key characteristics of anime face data. Subsequent projects such as IllustrationGAN [15] and AnimeGAN [3] have attempted to generate high-quality anime faces, but often report issues with fuzzy and deformed images, it is still difficult to develop a high standard anime faces for anime characters.

In this project, the proposal is a novel approach to overcome these limitations and generate high-quality anime faces by leveraging Illustration2Vec for active label synthesis and AC-GANs for image generation [13]. The use of Illustration2Vec allows us to automatically and accurately label the dataset with tags corresponding to style, gender, hair color, and eye color. These labeled features are essential for training AC-GANs, which in turn facilitate the generation of more diverse and high-quality anime faces while maintaining control over desired attributes.

By incorporating Illustration2Vec and AC-GANs in our approach, we aim to address the challenges faced by previous projects and generate visually appealing anime faces that respect the unique style and philosophy of anime characters. This combination of techniques not only helps to improve the overall quality of the generated images but also allows for more flexibility and control over the output, resulting in a more practical and versatile solution for anime face generation.

Keywords *Generative models, Unsupervised hyperref colorearning, Generative Adversarial Network, Machine Learning.*

2 Scope

2.1 Aim and Scope

The primary aim of this project is to develop a generative model capable of producing high-quality and unique anime character faces while addressing the inherent challenges and building upon the limitations of existing methods. This study will focus on exploring the potential of ACGAN and DCGAN architectures in generating distinct male and female faces with diverse traits and maintaining high image quality. Moreover, the application of Illustration2Vec will be investigated for accurately labeling the dataset and extracting semantic information from anime-style illustrations, which will facilitate generating images based on specific user requirements.

The objectives of this project are as follows:

- Examine the fundamental problems associated with GANs, such as mode collapse, convergence issues, and vanishing gradient, within the context of anime face generation, and explore potential solutions.
- Conduct a comparative analysis of different GAN architectures, ranging from basic GANs to more advanced variants like DCGANs and ACGANs, to assess their suitability for generating diverse and high-quality anime faces.
- Investigate the utility of Illustration2Vec for accurately labeling and extracting semantic information from the dataset, enabling the generation of anime faces based on specific user-defined attributes.
- Evaluate the performance of the developed model by quantifying the variety and quality of the generated faces, ensuring that they meet the essential requirements outlined in the introduction [5].

2.2 Data objective

Ideally in this study I want to seek to bridge the gap between the need for developing personalised virtual pictures through the principle of generating images based on personalised labels. Through leberaging the capabilities of Illustration2Vec [13] and AC-GANs / DCGANS [10]. In essence, a weakly supervised learning process using the aggregation of 2 core distinct learners, when being parsed into the generative network.

Such that the core objectives for the data it self are the following:

- Obtaining the ground truth for style labels: Eye colour, hair colour, and having a modular approach.
- Interpolating the image and parser to provide a more accurate representation of the image.

2.2.0.0.1 Pre-Processing As some datasets require customized images or pre-processed images based on their categorical information, it is essential to have more information about the images themselves. The core dataset used in this study has some labels, but it is necessary to consider additional labels and pre-trained models to expand on this information.

Experiments will be conducted on Illustration2Vec [15], a CNN-based technique for predicting tags for anime illustrations. Given an anime image, the network should predict the probability that it has generic anime features, referred to as tags (e.g., a smile, blue hair).

Although the core dataset I will be using has some labels, it is vital to have more information about the images it self. While the idea of manually annotating a ptoion of a gender label and then utilising it in conjuction with an incomplete labels sounds fruitful, it is also important to consider additional labels, and some pretrained models to expand on.

During the data pre-processing phase, the objectives are to:

- Determine the optimal way to represent and communicate the data to the generative network and the model
- Validate image quality
- Validate labels
- Parse forward to the generative adversarial network

Consequently, within the project's scope, I will strive to bridge the aforementioned disparities in a principled and organised way, enabling the user to easily design and produce pictures depending on particular image qualities. Such that this project aims to address the existing challenges with image labels and generate high-quality personalized virtual images.

2.3 Objectives - MileStones

2.3.1 Completed Objectives

- Gained an understanding of the mathematical principles behind GANs and the fundamental reasons why they are challenging to train.
- Investigated different GAN architectures by hyperparameter tuning and testing multiple network models, including DCGANs and MPL GAN (Multi Perceptron Layer).
- Created the initial anime dataset and applied a style network for feature extraction.

Upon the successful completion of these bjectives, this project will contribute to the development of a robus and effective Gan based solution for generating high quality, unique anime characters. The feats from this project will have future implications, particularly in the context of CGANs and a core future focus on ContraGans, which aim to enhance the controllability and interpretability of generated images. The insights gained from this project will pave the way for further exploration and refinement of GAN-based models, ultimately enabling more effective and efficient generation of diverse, high-quality, and user-specific content in various domains.

2.4 Related Literature

2.4.0.0.1 Generative Adversarial Networks (GANs) , introduced by Goodfellow et al [2]., demonstrate remarkable results in various generation tasks, such as image generation, image transfer, super-resolution, and more. The core concept of GANs involves training a generator and a discriminator model simultaneously. The discriminator's goal is to distinguish real examples, sampled from ground-truth images, from the samples produced by the generator. Conversely, the generator aims to create realistic samples that the discriminator cannot differentiate from ground-truth samples. This idea manifests as an adversarial loss applied to both the generator and discriminator during training, effectively encouraging the generator's outputs to resemble the original data distribution.

2.4.0.0.2 Conditional GANs (cGANs) , played a pivotal role in the early stages of this project, serving as a building block for generating images conditioned on specific input information. Mirza Osindero introduced cGANs, which incorporate conditional information, such as class labels, into both the generator and discriminator networks [8]. By conditioning the networks on additional input data, cGANs enable more controlled and targeted image generation, leading to better and more diverse results. The use of cGANs in this project allowed for a more structured and meaningful approach to generating anime-style images, as it permitted the model to generate images based on specific input conditions. This technique improved the overall quality and relevance of the generated images while maintaining a strong foundation for further research and development.

2.4.0.0.3 Conditional Auxiliary Classifier GANs (AC-GANs) , Odena et al. introduced the concept of AC-GANs in their paper titled "Conditional Image Synthesis with Auxiliary Classifier GANs" [9]. AC-GANs integrate the ideas of cGANs [8] and auxiliary classifiers to generate images conditioned on class labels, which results in higher-quality image generation. The discriminator acts as an auxiliary classifier, predicting not only the realness of an image but also its class label. This dual role of the discriminator enables the generation of more accurate and diverse images, as the generator is encouraged to produce images that not only look realistic but also conform to the specified class label.

2.4.0.0.4 AnimeGAN , **AnimeGan & AnimeGANv2** are two GAN-based models specifically designed for generating anime-style images [4, 14]. These models inspired the development of this project's anime image generation framework. By analyzing the strengths and weaknesses of both AnimeGAN and AnimeGANv2, this research aims to improve upon existing techniques for generating anime-style images and address the limitations of these models.

2.4.0.0.5 Illustration2Vec , Illustration2Vec, proposed by Saito Matsumoto [13], is a model designed to extract semantic information from illustrations. It utilizes a deep convolutional neural network (CNN) to learn feature representations of illustrations, enabling various applications such as annotation, retrieval, and generation. By incorporating Illustration2Vec into the image generation process, it is possible to leverage the semantic information to create more accurate and diverse images. For example, combining Illustration2Vec with a GAN model allows the generator to produce images that not only look visually appealing but also possess meaningful semantic content. This can lead to more realistic and contextually appropriate image generation results.

Numerous GAN variants have been proposed for image generation. Radford et al. applied convolutional neural networks (CNNs) in GANs to generate images from latent vector inputs[11]. Instead of generating images from latent vectors, several methods utilize the same adversarial idea for generating images with more meaningful input. Mirza Osindero introduced Conditional Generative Adversarial Nets, which use image class labels as conditional input to generate specific MNIST numbers [8]. Reed et al. further employed encoded text as input to generate images that match textual descriptions. Odena et al. proposed ACGAN, which trains the discriminator as an auxiliary classifier to predict the input condition, instead of only feeding conditional information as input [9].

3 Proposed Method

3.1 Preliminaries and Analysiss

3.1.1 Background: Generative Adversarial Networks

Simply said, generative adversarial networks (GANs) are two neural network players competing to defeat one another. During this process, both neural networks gain a significant lot of information and produce more accurate findings, which we may exploit to create fresh synthetic data. This new information closely matches the original information exploited by the networks. GANs are being employed in the generation of pictures, sounds, and movies.

Generative Modeling is an unsupervised learning issue in machine learning that creates fresh content.The favourite example of Generative Adversarial Networks (which appears to be engrained at this point, comparable to Bob, Alice, and Eve for cryptography) is the situation of money counterfeiting. The Generator is meant to produce counterfeit currency, whilst the Discriminator is intended to discriminate between authentic and counterfeit dollars. As the Generator advances, the Discriminator must also progress. This implies a dual training strategy in which one model strives to outperform the other (i.e. via more learning) (i.e. through additional learning

Particular a random vector/matrix, the generator offers a "fake" sample, and the discriminator attempts to detect whether a given sample is "genuine" (chosen from the training set) or "fake" (produced by the generator) (generated by the generator). Training proceeds concurrently: the discriminator is taught for a few epochs, followed by the generator, etc. Thus, both the generator and the discriminator grow more adept at their respective vocations. GANs are extremely sensitive to hyperparameters, activation functions, and regularisation. They are notoriously tough to train.

3.1.2 Core Mathematical Understanding of Vanilla GAN

High level overviews are only half the understanding, and there is a deeper principle for how generative networks are presented. In this section, I will take a core deep level proof and understanding for the principles used within the project and the core mathematical understanding. To prove anything, the purpose must be explicitly (mathematically) stated. This not only guides the proof technique but also provides a target to "keep an eye out for" while reviewing the evidence. The objective is for the Generator to provide instances that are distinct from actual data. This is the mathematical idea of random variables having equal distribution (or in law). Also, their probability density functions (i.e., the probability measure produced by the random variable on its range) are identical:

$$p_G(x) = P_{data}(x)$$

. This is the precise approach of the evidence: describe an optimization issue where the ideal solution is G satisfies $p_G(x) = P_{data}(x)$.

3.2 Methods Used

3.2.1 Deep Convolutional GANs

3.2.2 Auxiliary Classifier GANs

3.2.2.1 Feature Extraction

3.2.2.2 Limitation and Discussions

3.2.2.3 Overview of this structure

3.2.3 Reason for label synthesis

Active label synthesis is a method for computer-generated images that generates anime faces from specific labels. This technology is helpful because it can generate realistic human faces without requiring vast datasets or intricate neural networks. This is advantageous since it makes it possible to build anime faces fast and effectively while keeping the required degree of realism.

Moreover, active label synthesis may build anime faces with a specific set of attributes. This is achieved by the use of labels or tags that indicate certain facial characteristics, such as eyes, nose, mouth, and hair colour. Without individually creating each face, it is feasible to make anime faces with the appropriate appearance and feel using this approach.

Additionally, active label synthesis is advantageous because of its capacity to produce anime faces from many sources. For instance, it may be used to produce anime faces from photographs already stored in a database or from images uploaded by the user. This is helpful since it allows users more control over the process of making anime faces, enabling them to precisely tailor the picture to their desired appearance.

Active label synthesis is an effective method for creating anime faces. It can build realistic anime faces fast and effectively from a certain set of labels.

4 Experiments

4.1 Dataset and Evaluation Methods

4.2 Implementation Details

4.3 Quantitative Results

4.4 Output Structure

5 Software Engineering

5.1 ML Ops

6 Assessment

6.1 Professional Considerations for this project

6.1.1 Ethical Considerations for this project

7 Timeline

The optimal method would be to focus on core implementation during the first term and to have establish a solid foundation and working program for the second year. During the second term, emphasis will be placed on fine turning, and further tests will be conducted. Assuming all goes as planned, I will integrate more ideas to expand this project.

7.1 Term 1

TABLE 1 Term One

Week 1	Study more into GANs and VAE
Week 2	Formulate a full understanding of these representations
Week 3-4	Implement the code in both ML and none ML Context, and produce first Early deliverable
Week 4-5	Create a comparative view of the generated data
Week 7	Look into different variations of GANs and VAEs and compare the generated data.
Week 8-9	Fine tune and optimise my model and code
Week 10	Prepare for interim report and presentation
Week 11	Look into further detail about what I could add towards the report.

- Week 1 through week 4 Within these weeks, it was a base line for me to understand How generative networks work, more over, a high level overview, for what they are.
- Week 5 6 7 A modification against the original plan listed above, where i spent a large degree of time to review the mathematical proofs towards generative networks, instead of actively coding it . So in this instance, I was testing fully connected models and dcgans, in which i had concluded that DCGans provide better results.
- week 8 Was the week in which i started creating the model, and the code to evaluate the maths and start the data pre process on two test datasets. Cats and dogs and Human faces datasets.
- Week 9 Week 9 was taken up by allot of coursework , and had limited time to work on this project. Such that in this instance, I had decided to train and test my currentmodel, for 1000 epochs.
- week 10 and 11 Was purely based on refactoring the code, and cleaning up an impurities within the dataset using torch libs, to manipulate the dataset.

8 Diary

Please be advised that this diary is not a complete record of the work done on this project, but rather a summary of the most important events and decisions made during the project. All information relating to this literature can be found on the gitlab repo.

Abbreviations

FID The Frechet Inception Distance 3

GANs Generative Adversarial Networks 3

Bibliography

[1] Jaydeep T. Chauhan. *Gans vs VAE*. 2018. URL: https://www.ijcaonline.org/archives/volume182/number22/chauhan-2018-ijca-918039.pdf.

[2] Ian J. Goodfellow et al. *Generative Adversarial Networks*. 2014. arXiv: 1406.2661 [stat.ML].

[3] Jie Lei. *Animegan*. 2020. URL: https://github.com/jayleicn/animeGAN.

[4] Bing Li et al. "AniGAN: Style-Guided Generative Adversarial Networks for Unsupervised Anime Face Generation". In: *CoRR* abs/2102.12593 (2021). arXiv: 2102.12593. URL: https://arxiv.org/abs/2102.12593.

[5] Hongyu Li and Tianqi Han. *Towards Diverse Anime Face Generation: Active Label Completion and Style Feature Network*. Ed. by Paolo Cignoni and Eder Miguel. 2019. DOI: 10.2312/egs.20191016.

[6] Ziqiang Li et al. *A Comprehensive Survey on Data-Efficient GANs in Image Generation*. 2022. DOI: 10.48550/ARXIV.2204.08329. URL: https://arxiv.org/abs/2204.08329.

[7] Mattya. *Chainer-dcgan*. 2018. URL: https://github.com/mattya/chainer-DCGAN.

[8] Mehdi Mirza and Simon Osindero. "Conditional Generative Adversarial Nets". In: *CoRR* abs/1411.1784 (2014). arXiv: 1411.1784. URL: http://arxiv.org/abs/1411.1784.

[9] Augustus Odena, Christopher Olah, and Jonathon Shlens. *Conditional Image Synthesis With Auxiliary Classifier GANs*. 2016. DOI: 10.48550/ARXIV.1610.09585. URL: https://arxiv.org/abs/1610.09585.

[10] Alec Radford, Luke Metz, and Soumith Chintala. *Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks*. 2015. DOI: 10.48550/ARXIV.1511.06434. URL: https://arxiv.org/abs/1511.06434.

[11] Alec Radford, Luke Metz, and Soumith Chintala. *Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks*. 2016. arXiv: 1511.06434 [cs.LG].

[12] Rezoolab. *Make illustration on computer with chainer*. 2015. URL: http://qiita.com/rezoolab/items/5cc96b6d31153e0c86bc.

[13] Masaki Saito and Yusuke Matsui. "Illustration2Vec: a semantic vector representation of illustrations". In: *SIGGRAPH Asia 2015 Technical Briefs* (2015).

[14] TachibanaYoshino. *AnimeGANv2*. 2020. URL: https://github.com/TachibanaYoshino/AnimeGANv2.

[15] tdrussell. *IllustrationGAN*. 2016. URL: https://github.com/tdrussell/IllustrationGAN.

Articles only

[4] Bing Li et al. "AniGAN: Style-Guided Generative Adversarial Networks for Unsupervised Anime Face Generation". In: *CoRR* abs/2102.12593 (2021). arXiv: 2102.12593. URL: https://arxiv.org/abs/2102.12593.

[8] Mehdi Mirza and Simon Osindero. "Conditional Generative Adversarial Nets". In: *CoRR* abs/1411.1784 (2014). arXiv: 1411.1784. URL: http://arxiv.org/abs/1411.1784.

[13] Masaki Saito and Yusuke Matsui. "Illustration2Vec: a semantic vector representation of illustrations". In: *SIGGRAPH Asia 2015 Technical Briefs* (2015).

Gans

[2] Ian J. Goodfellow et al. *Generative Adversarial Networks*. 2014. arXiv: 1406.2661 [stat.ML].

[6] Ziqiang Li et al. *A Comprehensive Survey on Data-Efficient GANs in Image Generation*. 2022. DOI: 10.48550/ARXIV.2204.08329. URL: https://arxiv.org/abs/2204.08329.

[10] Alec Radford, Luke Metz, and Soumith Chintala. *Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks*. 2015. DOI: 10.48550/ARXIV.1511.06434. URL: https://arxiv.org/abs/1511.06434.

Misc

[9] Augustus Odena, Christopher Olah, and Jonathon Shlens. *Conditional Image Synthesis With Auxiliary Classifier GANs*. 2016. DOI: 10.48550/ARXIV.1610.09585. URL: https://arxiv.org/abs/1610.09585.

[11] Alec Radford, Luke Metz, and Soumith Chintala. *Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks*. 2016. arXiv: 1511.06434 [cs.LG].

[12] Rezoolab. *Make illustration on computer with chainer*. 2015. URL: http://qiita.com/rezoolab/items/5cc96b6d31153e0c86bc.

Github

[3] Jie Lei. *Animegan*. 2020. URL: https://github.com/jayleicn/animeGAN.

[7] Mattya. *Chainer-dcgan*. 2018. URL: https://github.com/mattya/chainer-DCGAN.

[14] TachibanaYoshino. *AnimeGANv2*. 2020. URL: https://github.com/TachibanaYoshino/AnimeGANv2.

[15] tdrussell. *IllustrationGAN*. 2016. URL: https://github.com/tdrussell/IllustrationGAN.

Index