# CYCLE GENERATIVE ADVERSARIAL NETWORK
# FOR UNPAIRED SKETCH-TO-CHARACTER TRANSLATION

LEENA ZEINI J ALSAATI

A dissertation submitted in partial fulfilment of the
requirements for the award of the degree of
Master of Science (Computer Science)

School of Computing
Faculty of Engineering
Universiti Teknologi Malaysia

APRIL 2019

# ACKNOWLEDGEMENT

I wish to express my sincere appreciation to my supervisor, Associate Professor Dr. Siti Zaiton Mohd Hashim, for her guidance, critics, and support. I am grateful to my parents for their guidance, advices, motivation, and support. My sincere appreciation also extends to my siblings, friends and colleagues who have provided assistance and emotional support. I am grateful to all. Without their support this thesis would not have been the same as presented here.

## ABSTRACT

Cartoon characters are currently being used in various applications such as comic and cartoon production. The ability to generate a variety of poses and facial expression of cartoon characters from simple sketches of stickfigures can ease the drawing process in production. Previous studies show only few research focused on the task of sketch to character translation. With low performance of detecting rare pose features and improving rare feature detection has not been significantly studied. The aim of our research is to investigate the capabilities of generative adversarial networks (GANs) in the application of Sketch to Character translation. A wide range of extended GAN versions has been reviewed and in this research, a new dataset collection has been proposed which consists of images of sketches and cartoon characters that are manually drawn. A Cycle GAN has been implemented and its performance against Conditional GAN is compared. Cycle GAN's cycle consistent loss is the main reason for learning a mapping between the domain of source images and the domain of target images without the need of paired training samples. Cycle GAN has been proven successful in handling a verity of applications in unpaired translation setting. The Conditional GAN has been also proven successful in a wide range of applications, however, it requires paired training samples. Results show that Conditional outperforms the Cycle GAN in accurately mapping the cartoon characters to the stickfigure, which is due to the nature of the paired training sample. However, the Cycle GAN still managed to produce sharper images that compete with the results of a Conditional GAN.

# ABSTRAK

Watak-watak kartun kini sedang digunakan dalam pelbagai aplikasi seperti pengeluaran komik dan kartun. Keupayaan untuk menghasilkan pelbagai pose dan ekspresi muka watak kartun dari lakaran mudah melekat dapat memudahkan proses lukisan dalam pengeluaran. Kajian terdahulu menunjukkan hanya sedikit penyelidikan yang difokuskan pada tugas lakaran untuk terjemahan aksara. Dengan prestasi rendah mengesan ciri pose yang jarang berlaku dan meningkatkan pengesanan ciri langka belum banyak dikaji. Tujuan penyelidikan kami adalah untuk menyiasat keupayaan rangkaian adversarial generatif (GANs) dalam penerapan terjemahan Sketch to Character. Pelbagai versi GAN yang diperluaskan telah dikaji dan dalam kajian ini, koleksi dataset baru telah dicadangkan yang terdiri daripada imej lakaran dan watak kartun yang ditarik secara manual. Kitaran GAN telah dilaksanakan dan prestasinya berbanding GAN Bersyarat dibandingkan. Siklus kitaran GAN yang konsisten adalah sebab utama pembelajaran pemetaan di antara domain imej sumber dan domain imej sasaran tanpa memerlukan sampel latihan berpasangan. Kitaran GAN telah terbukti berjaya dalam mengendalikan kebenaran aplikasi dalam tetapan terjemahan yang berpasangan. GAN Conditional juga telah terbukti berjaya dalam pelbagai aplikasi, bagaimanapun, ia memerlukan sampel latihan berpasangan. Keputusan menunjukkan bahawa Conditional mengungguli GAN Kitaran dengan tepat memetakan watak-watak kartun dengan tepat, yang disebabkan oleh sifat sampel latihan berpasangan. Walau bagaimanapun, Siklus GAN masih dapat menghasilkan imej yang lebih tajam yang bersaing dengan hasil GAN Bersyarat.

## TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

## LIST OF ABBREVIATIONS

| | | |
|---|---|---|
| ANN | - | Artificial Neural Network |
| CNN | - | Convolutional Neural Network |
| GAN | - | Generative Adversarial Network |
| CGAN | - | Conditional Generative Adversarial Network |
| DCGAN | - | Deep Convolutional Generative Adversarial Network |
| VAE | - | Variational Autoencoder |
| BN | - | Batch Normalization |

# LIST OF SYMBOLS

| | | |
|---|---|---|
| G | - | Generator |
| D | - | Discriminator |
| xi | - | Input sample |
| yi | - | Output sample |
| X | - | Input domain |
| Y | - | Output domain |
| z | - | Noise |
| N | - | Number of samples |

# CHAPTER 1

# INTRODUCTION

## 1.1 Introduction

Cartoon characters have been a part of many of us from an early age. We have been exposed to them as early as when we started perceiving and opening our eyes to the world. Drawing is used as a form of visual art in which an artist expresses their thoughts and imagination though simple lines of sketches to detailed illustrations to present a picture and further, using colours and shading to enhance its reality. Drawing is used in many commercial productions such commercial illustrations, comic and animation production, architectural designing, and many more. The traditional technique for illustration productions was to manually draw them on traditional equipment such as pencil, paper, paint, light tables, etc. Moreover, the process of the actual production involved for each image to be drawn repeatedly with slight changes to indicate movement. Now a days, the majority of the process is digitized, and the rise of animation software has become more popular. Animation and storyboarding software such as Toon Boom, have been used in the production of numerous famous cartoon series; Family Guy, The Simpsons, Rick and Morty, Bob’s Burgers, and many more. However, the reliance on the repetitive task of redrawing the character between time lapse to represent an action such as walking or running still exists. A different tool that was provided by Toom Boom software is the use of joint configuration to assemble movement points that allow the character to move on those points. Also, another software designed for computer animation and modelling is Maya. The process of production of 3D computer graphics creation involves 3D modelling and forming the object's shape, animation and placement of movement of objects, and 3D rendering based on light, surface types, and other qualities to generate the image.

Having the Imagination to create but no skills of drawing can make the process more difficult. However, the ability to sketch is easy. A sketch is defined as fast, freehand drawing or a form of doodling, and most likely is not intended to be complete or claimed as finished work. It is mostly used a first draft which acts as a base for something more detailed. Generative Adversarial Networks (GANs), a model that generates new images, are applicable to many applications. An eye-catching implementation by MIT is *Nightmare Machine*, where they used GANs for image translation, specifically style transfer. Moreover, training GANs to auto generate illustrations of characters by drawing simple sketches with defining the special features, can ease the process tremendously by lowering the production time consumed in drawing each character as more characters are generated in fewer time, thus, avoiding repetitive tasks.

## 1.2 Problem Background

Existing methods of cartoon and comic production is mainly revolved on having drawing talent and skills in using software tools. We take a step into automating the process of generating cartoon characters.

Generative models are models that learn the natural characteristics of a dataset and generate new samples that cannot be differentiated from existing data. The generative algorithm learns about the distribution of joint probability over data observations. GANs have been proven successful in understanding a variety of data types such as images, audios, and videos. Many studies have come up to address the problem of image generation and have proposed many modifications of the original GAN. These architectures have been applied on a wide variety of applications, which will be discussed in chapter two.

Despite the success of GANs, it can suffer from several serious problems including Mode Collapse, and Non-convergence (Goodfellow, Pouget-abadie, Mirza, Xu, & Warde-farley, 2014). Mode Collapse occurs when the model learns only a single class from the set of target images that achieves a low error value and uses it repeatedly until all of the generated images are of that only class. Non-convergence is the inability for the model to converge to a final state, in more detail, the gradient becomes unsure which direction to place the next step. (Goodfellow et al., 2014).

Another problem that can affect the learning process is the data availability. Having enough training samples is important for it to perform well. In most cases, collecting a paired dataset by having a corresponding input image for every target image, can be expensive. This creates a problem as there are cases where paired training samples are not available or does not exist. As a result, the performance of the model may degrade when it is given unseen training samples. State of the arts models now exist to handle rare or unseen training samples using unpaired training. As a result, it alleviates the big burden in obtaining label-intensive pixel-to-pixel supervision. Existing research attempted to address unpaired datasets by developing extended versions of GAN. The most popular and domain diverse model is the Cycle GAN, where the authors have adopted it for many applications (Jun-yan Zhu, Park, & Efros, 2018). Furthermore, other state of the art models such as BIGAN, CoGAN, and DiscoGAN share in common an objective function that focuses on unpaired dataset. However, not many researches have addresses the problem of sketch-to-character translation

## 1.3 Problem Statement

Despite the successes recorded in the area of image-to-image translation, several challenges continue to persist particularly in the areas of robustness of the models being able to identify and generate rare features. Low performance in detecting rare features which leads to the model generating blurred images for unseen data samples. Furthermore, previous studies examined the problem of sketch to cartoon character translation but involved only training the model using paired data samples. Moreover, most research focused on various applications, such as sketch-to-face translation and a very few focusing on sketch-to-character generation. The main purpose of this study was to adopt and examine Cycle GAN model that can effectively identify and generate unseen features using an unpaired training dataset. Having a corresponding input and its label in a dataset can be difficult and the process of creation can be expensive, therefore, the need to explore unpaired based model.

## 1.4 Aim of the Study

The study adopted a Cycle GAN model to translate simple sketches, with defined facial and body features, to detailed cartoon characters. Specifically, it trained the model in an unpaired setting, to produce synthesised cartoon images that represents that features given by the stick figure.

## 1.5 Objectives of the Study

The basis of comparison in this study is to make use of two approaches of feeding data to models. As follow, the following objectives achieved in this study:

I. To generate images of cartoon characters, from unpaired images of stick figure sketches and target cartoon characters, using Cycle GAN.

II. To compare its performance with synthesized images, from paired images of stick figures and cartoon characters, generated by Conditional GAN.

## 1.6 Research Questions of the Study

The study answers the following research questions:

I. Can a Cycle GAN translate unpaired images of cartoon characters from stick figure sketches and produce good results?

II. How does the cartoon characters generated by Cycle GAN compare to Conditional GAN which translates based on paired images of cartoon characters from stick figure sketches?

## 1.7 Significance of the Study

The outcome of this study will be of benefit to both individuals and industries who will need to produce good quality cartoon characters, with defined poses and facial expressions, from simple sketches without much talent needed. This will allow individuals and artists to produce many amounts of characters with little time and effort, and thus improving the production rate. Moreover, it eliminates the expensive process of collecting a dataset to train a model, and thus automatically generate sketch based images of cartoon characters. The significance of the study therefore encompasses the following:

I. The created dataset yields a better incite on the effect of paired and unpaired training samples on the model's performance.
II. The adopted Cycle GAN model eliminates the need for paired training examples; Sketches matched with its corresponding cartoon character.
III. The study produces an adopted model that generates images of cartoon characters from sketches with defined facial and body pose features.
IV. The study shows a comparison between the performances of paired and paired based models, where the unpaired based model gives competing results with the paired based model.

## 1.8 Scope of the Study

The scope of this study has been limited to the image-to-image translation task of generating cartoon characters from stick figure sketches, with the use of paired and unpaired datasets. Moreover, the study focused on:

I. Difficulty rating of *Easy* and *Moderate*; images will have unified sizing, framing, and include only a single, and with varied pose orientation. It will not cover the *Hard* rating as our scope is to only generate single subjects (does not include other objects such as furniture, pets, accessories).

II. Defined facial and body pose features.

III. Used Python programming language, and TensorFlow and Pytorch Libraries to code the models.

## 1.9 Organization of the Dissertation

Chapter one of the dissertation introduces the true essence of the research proposal. It begins with an introduction to the research topic, brief commentary of the background to the problem, and the problem statement. The chapter then highlights the objectives, aim, and scope of the research.

Chapter Two presents the literatures review. It is conducted in order to gain an in-depth understanding of the current research area and find the research gaps in the domain of image-to-image translation. Firstly, the definition of Generative Adversarial Network, problems, and optimization techniques are introduced. Secondly, the problem of image-to-image translation and its applications are explained. Thirdly, previous researches related to the tasks of sketch-to-character translation are studied. Finally, the concept of paired and unpaired datasets is explained.

## REFERENCES


Almahairi, A., Rajeswar, S., Sordoni, A., Bachman, P., & Courville, A. (2018). Augmented CycleGAN : Learning Many-to-Many Mappings from Unpaired Data. *International Conference on Machine Learning, Stockholm, Sweden, PMLR 80, 2018*.

Authors, A. (2019). Harmonic Unpaired Image to Image Translation. *Under Review at ICLR*, 1–17.

Baldi, P. (2012). Autoencoders, Unsupervised Learning, and Deep Architectures. *JMLR: Workshop and Conference Proceedings*, 37–50. https://doi.org/10.1561/2200000006

Bin, H., Weihai, C., Xingming, W., & Chun-Liang, L. (2017). *High-Quality Face Image SR Using Conditional Generative Adversarial Networks*. https://doi.org/10.4310/MRL.1997.v4.n4.a1

Chen, N., & Xu, L. (2015). Predicting Depth, Surface Normals and Semantic Labels with a Common Multi-Scale Convolutional Architecture. *Environmental Science and Pollution Research*, *24*(3), 2928–2935. https://doi.org/10.1007/s11356-016-8065-3

Chuang, Y., Chen, Y., Wang, Y., & Kao, M. (2018). *Deep Photo Enhancer : Unpaired Learning for Image Enhancement from Photographs with GANs*.

Doersch, C. (2016). *Tutorial on Variational Autoencoders*. 1–23. https://doi.org/10.3389/fphys.2016.00108

Eitz, M., Hays, J., & Alexa, M. (2012). How do humans sketch objects? *ACM Transactions on Graphics*, *31*(4), 1–10. https://doi.org/10.1145/2185520.2185540

Fukumoto, Y., Shimizu, D., & Shibata, C. (2018). Generation of Character Illustrations from Stick Figures using a Modification of Generative Adversarial Network. *2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC)*, *01*, 183–186. https://doi.org/10.1109/COMPSAC.2018.10225

Goodfellow, I. J., Pouget-abadie, J., Mirza, M., Xu, B., & Warde-farley, D. (2014). *Generative Adversarial Nets*. 1–9.

Ha, D., & Eck, D. (2017). *A Neural Representation of Sketch Drawings*. Retrieved from http://arxiv.org/abs/1704.03477

Hertzmann, A., Jacobs, C. E., Oliver, N., Curless, B., & Salesin, D. H. (2001). Image analogies. *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques - SIGGRAPH '01*, 327–340. https://doi.org/10.1145/383259.383295

Hitawala, S. (2018). *Comparative Study on Generative Adversarial Networks*.

Huang, Y., & Khan, S. M. (2017). DyadGAN : Generating Facial Expressions in Dyadic Interactions Yuchi Huang. *IEEE Conference on Computer Vision and Pattern Recognition Workshops DyadGAN:*, 2259–2266. https://doi.org/10.1109/CVPRW.2017.280

Isola, P., Zhu, J. Y., Zhou, T., & Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. *30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 5967–5976. https://doi.org/10.1109/CVPR.2017.632

Jakub Langr, V. B. (2018). GANs in ActioN. *MEAP Edition, Copyright 2018 Manning Publications For, 4*.

Jin, Y., Zhang, J., Li, M., Tian, Y., & Zhu, H. (2017). Towards the High-quality Anime Characters Generation with Generative Adversarial Networks. *Machine Learning for Creativity and Design, NIPS Workshop, Long Beach, California, USA*, 1–13.

Johnson, J., Alahi, A., & Fei-Fei, L. (2016). *Perceptual Losses for Real-Time Style Transfer and Super-Resolution*. 1–18. https://doi.org/10.1023/A:1015856924451

Kataoka, Y., Matsubara, T., & Uehara, K. (2017). Automatic Manga Colorization with Color Style by Generative Adversarial Nets. *IEEE SNPD, Kanazawa, Japan 495*, 495–499.

Kazemi, H., Iranmanesh, M., Dabouei, A., Soleymani, S., & Nasrabadi, N. M. (2018). Facial Attributes Guided Deep Sketch-to-Photo Synthesis. *IEEE Winter Conference on Applications of Computer Vision Workshops*, 1–8. https://doi.org/10.1109/WACVW.2018.00006

Kingma, D. P., & Welling, M. (2014). *Auto-Encoding Variational Bayes*. 1–14.

Kumar, S. (2018). "*Basic Generative Adversarial Networks take random noise as input and output realistic data, what kind of generative networks takes a drawing as input and outputs a photo",* [Accessed 18 Mar. 2019].

Laffont, P.-Y., Ren, Z., Tao, X., Qian, C., & Hays, J. (2014). Transient attributes for high-level understanding and editing of outdoor scenes. *ACM Transactions on Graphics*, *33*(4), 1–11. https://doi.org/10.1145/2601097.2601101

Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., … Shi, W. (2017). Photo-realistic single image super-resolution using a generative adversarial network. *30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 105–114. https://doi.org/10.1109/CVPR.2017.19

Liang, X., Lee, L., Dai, W., & Xing, E. P. (2017). Dual Motion GAN for Future-Flow Embedded Video Prediction. *Proceedings of the IEEE International Conference on Computer Vision*, 1762–1770. https://doi.org/10.1109/ICCV.2017.194

Luo, V., Straka, M., & Li, L. (2018). *Historical and Modern Image-to-Image Translation with Generative Adversarial Networks*.

Mirza, M., & Osindero, S. (2014). *Conditional Generative Adversarial Nets*. 1–7.

P.Kingma, D., & Ba, J. L. (2017). ADAM: A Method for Stocastic Optimization. *ICLR 2015*, 1–15. https://doi.org/10.1016/j.patrec.2017.03.023

Petrov, T. (2019). "*Why doesn't gradient descent work to optimize generative adversarial networks?".* [online][Accessed 18 Mar. 2019].

Philip, C., & Jong, L. H. (2017). Face Sketch Synthesis using Conditional Adversarial Networks. *ICTC*, 373–378.

Quan, X., Shunli, H., Chunyan, J., Lei, S., & Zhikai, L. (2015). Ultra-low permeability rocks: Laboratory measurement and low pressure gradient analysis. *Petroleum Science and Technology*, *31*(9), 941–948. https://doi.org/10.1080/10916466.2010.535080

Radford, A., Metz, L., & Chintala, S. (2016). Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. *ICLR*, 1–16. https://doi.org/10.1051/0004-6361/201527329

Reed, S., Akata, Z., Yan, X., Logeswaran, L., Schiele, B., & Lee, H. (2016). Generative Adversarial Text to Image Synthesis. *Proceedings of the 33rd International Conference on Machine Learning, New York, NY, USA, 2016*. https://doi.org/10.1017/CBO9781107415324.004

Royer, A., Bousmalis, K., Gouws, S., Bertsch, F., Mosseri, I., Cole, F., & Murphy, K. (2017). *XGAN: Unsupervised Image-to-Image Translation for Many-to-Many Mappings*. 1–19. Retrieved from http://arxiv.org/abs/1711.05139

Selseng, S. (2017). *Guiding the training of Generative Adversarial Networks*.
Shih, Y., Paris, S., Durand, F., & Freeman, W. T. (2013). Data-driven hallucination of different times of day from a single outdoor photo. *ACM Transactions on Graphics*, 1–11. https://doi.org/10.1145/2508363.2508419

Wang, X., & Gupta, A. (2016). *Generative Image Modeling using Style and Structure Adversarial Networks*. https://doi.org/10.1007/978-3-319-46493-0
Xie, S., & Tu, Z. (2015). *Holistically-Nested Edge Detection*. 1–10. https://doi.org/10.1002/pssb.200301569

Xiong, W., Luo, W., Ma, L., Liu, W., & Luo, J. (2018). *Learning to Generate Time-Lapse Videos Using Multi-Stage Dynamic Generative Adversarial Networks*. https://doi.org/10.1109/CVPR.2018.00251

Yeh, R. A., Chen, C., Lim, T. Y., Schwing, A. G., Hasegawa-johnson, M., & Do, M. N. (2017). *Semantic Image Inpainting with Deep Generative Models*.

Zhang, H., Xu, T., Li, H., Zhang, S., Huang, X., Wang, X., & Metaxas, D. (2017). Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks. *IEEE Int. Conf. Comput. Vision (ICCV)*, 5907–5915. Retrieved from internal-pdf://4.110.82.218/Zhang_StackGAN_Text_to_ICCV_2017_paper.pdf

Zhang, M., Member, S., Ma, K. T., & Lim, J. H. (2017). Anticipating Where People Will Look Using Adversarial Networks. *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. X, NO. X, JUNE 2017 Anticipating*. https://doi.org/10.1109/TPAMI.2018.2871688

Zhang, R., Isola, P., & Efros, A. A. (2016). *Colorful Image Colorization*. 1–29. https://doi.org/10.1109/ICCV.2015.55

Zhang, S., Ji, R., Hu, J., Lu, X., & Li, X. (2018). Face Sketch Synthesis by Multidomain Adversarial Learning. *IEEE Transactions on Neural Networks and Learning Systems*, 1–10. https://doi.org/10.1109/TNNLS.2018.2869574

Zhu, J., Park, T., & Efros, A. A. (2018). *Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks*.

Zhu, J., Xie, J., & Fang, Y. (2018). Learning Adversarial 3D Model Generation with 2D Image Enhancer. *The Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18)*, 7615–7622.

Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2018). Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. *Proceedings of the IEEE International Conference on Computer Vision*, 1–18. https://doi.org/10.1109/ICCV.2017.244