

**2D HUMAN ANIMATION SYNTHESIS FROM VIDEOS USING
GENERATIVE ADVERSARIAL NEURAL NETWORKS.**

Pasan Nadeera Udawatta

199489J

Degree of Master of Science in Artificial Intelligence

Department of Computational Mathematics

Faculty of Information Technology

University of Moratuwa Sri Lanka

March 2022

2D HUMAN ANIMATION SYNTHESIS FROM VIDEOS USING GENERATIVE ADVERSARIAL NEURAL NETWORKS.

Pasan Nadeera Udawatta

199489J

Thesis submitted in partial fulfillment of the requirements for the Degree of
Masters of Science in Artificial Intelligence

Degree of Master of Science in Artificial Intelligence

Department of Computational Mathematics

Faculty of Information Technology

University of Moratuwa Sri Lanka

March 2022

Declaration

I declare that this dissertation does not incorporate, without acknowledgment, any material previously submitted for a Degree or a Diploma in any University and to the best of my knowledge and belief, it does not contain any material previously published or written by another person or myself except where due reference is made in the text. I also hereby give consent for my dissertation, if accepted, to be made available for photocopying and for interlibrary loans, and for the title and summary to be made available to outside organization.

Name of the Student

P. N Udawatta

Signature of Student

Date:

Supervised by

Dr.Subha Fernando

Signature of the Supervisor

Date:

Dedication

I dedicate this thesis to my parents and my wife who were there for me for my successes and failures.

Acknowledgment

The success of this thesis would not have been possible without the guidance and support of many people around me.

Many thanks to my adviser Dr. Subha Fernando for her effort, patience and wisdom, and guidance provided throughout the project. Her expertise was the guiding light that enabled me to successfully formulate the research question and the methodology.

I would also like to thank Prof. Asoka Karunananda for the valuable guidance that directed me towards the correct path of conducting research.

I would also like to gratefully recognize the efforts of Dr.Sagara Sumathipala for his valuable support to conduct this research promptly. I would also like to offer my gratitude to all my lecturers who provide me with support, guidance, and knowledge that enable me to complete this thesis. I would also like to offer my sincere gratitude to non-academic staff members who were there to help throughout my postgraduate process.

I would also like to recognize my dear colleagues for all the support and guidance provided.

Most importantly, I would like to thank my parents and my wife for their unconditional, unequivocal, and loving support.

Abstract

Synthesizing 2D human animation has many industrial applications yet is currently done manually by animators utilizing time and resources. Therefore, many types of research have been conducted to synthesize human animation using artificial intelligence techniques. However, these approaches lack the quality as well as capability to generalize to various visual styles. Thus, synthesizing high-quality human animations across different visual styles remains a research challenge

We hypothesize that given video references for motion and appearance, synthesizing high-quality human animations across a variety of visual styles can be achieved via generative adversarial networks.

Here we have come up with the solution known as HumAS-GAN, an acronym for **H**uman **A**nimation **S**ynthesis **G**enerative **A**dversarial **N**etworks. HumAS-GAN accepts video references for motion and appearance and synthesis 2d Human animations. HumAS-GAN has three main modules, motion extraction, motion synthesis, and appearance synthesis. In motion extraction, the motion information is extracted via pre-trained human pose extraction [21], The motion synthesis module synthesizes a motion representation matching the target human's body structure which is then combined with the human pose coordinates to be used by the appearance synthesis module to generate the Human animation. HumAS-GAN is focused on improving the quality of the animation as well as the ability to use cross-domain/visual-style references to generate animation. This solution will be beneficial for many multimedia-based industries as it is capable of generating high human animations and quickly switching to any visual style they prefer.

HumAS-GAN is evaluated against other methods using a custom dataset and a set of 3 experiments designed to evaluate the capability of generating human animations across various visual styles. Evaluations results prove the superiority of HumAS-GAN over other methods in synthesizing high-quality 2d human animations across a variety of visual styles.

Contents

Declaration	i
Dedication	ii
Acknowledgment	iii
Abstract	iv
Chapter 1 Introduction	1
1.1 Prolegomena	1
1.2 Background and motivation	1
1.3 Aim and objectives	2
1.4 Problem definition	2
1.5 HumAS-GAN	2
1.6 Resource requirements	3
1.7 Structure of the thesis	3
1.8 Summary	3
Chapter 2 2D Human animation synthesis – Past, Present & Future	4
2.1 Introduction	4
2.2 2D Human animations	4
2.3 Quality of Human animation	5
2.4 Introduction Human animation synthesis	5
2.5 Early developments that influenced Human animation synthesis	5
2.5.1 Image to Image Translation	5
2.5.2 Pose Guided Human Image Synthesis	6
2.6 Approaches of Human animation synthesis	6
2.6.1 The generalized model approach for human animation synthesis	6
2.6.2 Specialized models approach for Human motion synthesis	8
2.6.3 Similarities in specialized and generalized approaches	9
2.6.4 Human Motion representation methods of Specialized and General approaches ...	10
2.7 Problem definition	10
2.8 Summary	11
Chapter 3 Technologies used for HumAS-GAN	12
3.1 Introduction	12

3.2	Artificial Neural networks.....	12
3.1	Generative algorithms	12
3.2	Generative adversarial network (GAN)	13
3.3	GAN: The Generator.....	13
3.4	GAN: The Discriminator.....	13
3.5	Adversarial training.....	14
3.6	Conditional generative adversarial networks (cGAN).....	14
3.7	Conditional generative adversarial networks for human motion synthesis.....	15
3.7.1	Pix2Pix GAN architecture	15
3.7.2	Pix2Pix HD GAN architecture.....	16
3.8	Human segmentation extraction.....	17
3.9	Human pose extraction.....	17
3.9.1	Summary	18
Chapter 4	Approach for HumAS-GAN.....	19
4.1	Introduction	19
4.2	Hypothesis.....	19
4.3	Input	19
4.3.1	Input of the HumAS-GAN.....	19
4.3.2	Tanning Label input for HumAS-GAN	19
4.4	Output.....	20
4.5	Process.....	20
4.6	Users.....	21
4.7	Feature.....	21
4.8	Summary	21
Chapter 5	Design of HumAS-GAN.....	22
5.1	Introduction	22
5.2	Motion extraction module	22
5.3	Motion Synthesis module.....	23
5.4	Appearance Synthesis module	23
5.5	Summary	24
Chapter 6	Implementation of HumAS-GAN.....	25
6.1	Introduction	25

6.2	Motion extraction module	25
6.3	Motion Synthesis module.....	25
6.4	Appearance Synthesis module	26
6.5	Training HumAS-GAN model.....	26
6.5.1	Training data	26
6.5.2	The training process of the HumAS-GAN Motion Synthesis module	27
6.5.3	The training process of the HumAS-GAN Appearance Synthesis module	28
6.6	Summary	28
Chapter 7	Evaluation of HumAS-GAN.....	29
7.1	Introduction	29
7.2	Quantitative evaluation	29
7.2.1	Data preparation.....	29
7.2.2	Indexes of qualitative evaluation	29
7.2.3	Experiment process.....	30
7.2.4	Quantitative comparison	31
7.3	Qualitative evaluation	32
7.4	Experiment 1 Qualitative evaluation.....	33
7.4.1	Experiment 2 Qualitative evaluation	34
7.4.2	Experiment 3 Qualitative evaluation	35
7.5	Summary	37
Chapter 8	Conclusion and future work.....	38
8.1	Introduction	38
8.2	Conclusion.....	38
8.2.1	Achievement of Project Objectives	38
8.3	Overall Conclusion.....	38
8.4	Limitations and further work	39
8.5	Summary	39
References	40
APPENDIX	43
Chapter 1	43
1.1	Appendix: I: Open CV Library	43
1.2	Appendix II: Normalization algorithm.....	43

1.3	Appendix III: Everybody dance now[8] implementation git repository	45
1.4	Appendix IV: Pix2PixHD[1] implementation	45
1.5	Appendix V: IOU loss equation	46
1.6	Appendix VI: Evaluator	46
1.7	Appendix VIII: Inference script	47
1.8	Appendix IX: Train model	49
1.9	Appendix X: Custom human motion dataset	52

List of tables

Table 7.1 : Evaluation experiments	31
Table 7.2 – mean SSIM and LPIPS value comparison from experiment 1	31
Table 7.3 – mean SSIM and LPIPs values comparison for experiment 2	31
Table 7.4 – mean SSIM and LPIP value comparisons from experiment 3.....	32

List of figures

Figure 2.1: Failure to separate body structure from motion reference	7
Figure 3-1 – Example of a Human segmentation image generated by Human Parsing[20] model	17
Figure 3-2 – Example of a Pose image generated by OpenPose, which is an implementation of [21] approach.....	17
Figure 5-1 Top-level architecture of HumAS-GAN.....	22
Figure 6.1, The training process of the G1	27
Figure 6.2, The training process of the G2.	28
Figure 7.1: The experiment for evaluating HumAS-GAN against other human synthesizing methods.....	30
Figure 7.2: animation of CGI character	33
Figure 7.3: Animation of a photorealistic character	34
Figure 7.4: Animation of character with a significantly different body different to the motion reference character	35
Figure 7.5: Comparison of synthesized animations against appearance reference.....	36
Figure 7.6: Influence of target motion image for the final animation	37

Abbreviations

GAN	Generative Adversarial Network
CGI	Computer Generated imagery
HumAS-GAN	Human Animation Synthesis Generative Adversarial Network
cGAN	Conditional Generative Adversarial Network
GPU	Graphics Processing Unit
HD	High Definition
SSIM	Structural Similarity Index Measure
LPIPS	Learned Perceptual Image Patch Similarity