PURBANCHAL UNIVERSITY

DEPARTMENT OF COMPUTER ENGINEERING

KHWOPA ENGINEERING COLLEGE
LIBALI, BHAKTAPUR

A PROJECT REPORT

ON

TEXT TO PHOTO REALISTIC IMAGE USING GAN

Project work submitted in partial fulfillment of requirements for the award of the degree of Bachelor of Engineering in Computer Engineering

SUBMITTED BY

Biraj Khatri (710308)

Milan Adhikari (710315)

Nabin Bikram Sah (710316)

Rajesh Bdr. Shrestha (710325)

Sahas Dangol (710335)

UNDER THE GUIDANCE OF

Er. Shree Ram Khaitu

18th November, 2018

ACKNOWLEDGEMENT

Firstly, we would like to express our deepest gratitude to Department of Computer Engineering for accepting the topic for us to carry out the project. It gives us immense pleasure to express our deepest sense of gratitude and sincere thanks to our highly respected and esteemed guide Er. Shree Ram Khaitu, our project supervisor, for his valuable guidance, encouragement and help for completing this work. His useful suggestions for this whole work and co-operative behavior are sincerely acknowledged.

We would like to express our sincere thanks to Er. Reena Manandhar, Head of Department of Computer Engineering, for giving us this opportunity to undertake this project. We are also grateful to our lecturers for their constant support and guidance.

We would also like to express our sincere thanks to all our friends and others who helped us directly or indirectly during this project work. Finally, we expect constructive criticisms and valuable suggestions from our teachers for further improvement and usability of this project.

With Kind Regards

Biraj Khatri (710308)

Milan Adhikari (710315)

Nabin Bikram Sah (710316)

Rajesh Bdr. Shrestha (710325)

Sahas Dangol (710335)

ABSTRACT

Human beings are quickly able to conjure and imagine images related to natural language descriptions. For example, when we read a story about a sunny field full of flowers, an image of a beautiful field with blossoming flowers might pop into head. Artificial synthesis of images using text descriptions or human cues could have profound applications in visual editing, animation, and digital design.

Generating high-quality image from the given text description is a challenging problem in computer vision and has different advantages in many applications. The sample image generated by existing text-to-image approach could be roughly reflect the meaning of the given description, but they fail to contain necessary details and vivid object parts which is necessary to build the image properly.

In this report, Stacked Generative Adversarial Networks (StackGAN) is used to generate 256x256 photo-realistic image conditioned on the text description. Using decomposition technique, the hard problems is sub divided into sub-problems through a sketch-refinement process. The stage-I GAN sketches the primitive shape and color of the object based on the given text description, and in this stage-I we get low-resolution image. The Stage-II GAN takes Stage-I results and text description as input, and generates high-resolution images with photo-realistic details. Some of the defects that arrives in stage-I will be rectified and add compelling details with the refinement process. After refinement process is done than the significant improvements on generating photo-realistic image is generated.

Keywords: artificial, synthesis, computer vision, Stack GAN, Stage-I GAN, Stage-II GAN, Vivid, high-resolution, Photo-realistic

TABLE OF CONTENTS

Chapters	Page no.
Cover Page	
Certificate	
Acknowledgement	i
Abstract	ii
Table of Content	iii
List of Figures	v
List of Abbreviation	vi
1. Introduction	
1.1 Background	1
1.2 Statement of Problem	2
1.3 Objectives	2
1.4 Scope	2
1.5 Limitation	2
2. Literature Review	3
3. Methodology	
3.1 Background	5
3.2 Tools and Platforms	6
3.3 Stacked Generative Adversarial Network	
3.3.1 Conditioning Augmentation	7
3.3.2 Stage-I GAN	7
3.3.3 Stage-II GAN	8
3.3 Algorithm	10
3.4 Architecture	11
3.5 Implementation and Training	12
3.6 Datasets and Features	12

4. Result and Discussion

4.1 Testing

4.1.1 Time for execution 13

4.1.2 Average time for training 15

4.2 Output 16

5. Conclusion and Future Enhancement 18

References 19

Appendix 20

LIST OF FIGURES

Figures	Page no.
3.5.1 Architecture of StackGAN for text to photorealistic image generation	10
4.2.1 Output for the text “a colorful bird with a bright yellow body, a black crown and throat, orange bill, and black primaries and secondaries.”	15
4.2.2 Output for the text “a black bird with a bright yellow body, a black crown and throat, orange bill, and black primaries and secondaries.”	15
4.2.3 Output for the text “small, roundish bird with off white breast and belly, light brown crown, brown and black colored wings.”	15
	15
4.2.4 Output for the text “This small blue bird has a short pointy beak and brown on its wings.”	16
4.2.5 Output for the text “This bird is completely red with black wings and pointy beak.”	16
4.2.6 Output for the text “A small sized bird that has a cream belly and short pointed bill.”	16
4.2.7 Output for the text “A small bird with a black head and wings and features grey wings.”	16
4.2.8 Output for the text “This bird is black with green and has a very short beak.”	16

LIST OF ABBREVIATION

CNN	Convolutional Neural Network
CUB	Caltech-UCSD Bird
DCGAN	Deep Convolutional Generative Adversarial Networks
GAN	Generative Adversarial Network
MSCOCO	Microsoft Common Object in Context
PPGN	Plug & Play Generative Networks