MUSIC GENERATION FOR SCENE EMOTION USING GENERATIVE AND CNN MODEL

D.I.D.D. Jayawardena

168285U

Degree of Master of Science in Artificial Intelligence

Department of Computational Mathematics

University of Moratuwa

Sri Lanka

March 2019

MUSIC GENERATION FOR SCENE EMOTION USING GENERATIVE AND CNN MODEL

D.I.D.D. Jayawardena

168285U

Thesis submitted in partial fulfilment of the requirements for the degree of Master of Science in Artificial Intelligence

Department of Computational Mathematics

University of Moratuwa

Sri Lanka

March 2019

Declaration

I declare that this is my own work and this thesis does not incorporate without acknowledgement any material previously submitted for a Degree or Diploma in any other University or institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Also, I hereby grant to University of Moratuwa the non-exclusive right to reproduce and distribute my thesis, in whole or in part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as articles or books).

Name of Student:
D.I.D.D.Jayawardena.

Signature of Student: Date:

The above candidate has carried out research for the Master's Dissertation under my supervision.

Name of Supervisor:
Dr. Subha Fernando.

Signature of Supervisor: Date:

Abstract

Generate music using emotional semantics of an image is quiet challenging task due to the complexity of extracting emotional features of an image and generate music according to the emotion.This paper proposes an enhanced deep neural network backed by Generative adversarial network for scene emotion categorization and LSTM based music generator for music generation. In developed system system functions in three parts.Initially we have generated fake images which more looks like real images using the generator of generative adversarial network which will help to enrich the dataset and increase the size of the dataset.Our dataset contains mainly three emotion categories (Happy,Angry,Sad). Second part of the system is image classifier developed using convolutional neural network which is trained using enhanced image dataset of scene emotions.Image classifier helps to identify the probabilities of the input scene which fed in to music generator for creation of training music dataset for each uploaded scene.Third and the last part of the system is the music generator which is developed using convolutional neural network with Long short term memory model.With the use of LSTM model developed deep neural network model got the capability of remember and predict next step.MIDI dataset from raw music files of songs created for each category to train the music generator. Since music composing is more human centric task,best way to evaluate the system is using musicians.So we have tested the system with two musicians and single listener.And also we have compare the image classifier using dataset which contains GAN generated images and without GAN generated images. After improving the dataset using generated images by GAN,we were able to achieve 80% of categorical accuracy and 85% of validation accuracy in image classification.Based on the evaluation done by musicians on generated sounds more than 50% of the sounds were in good quality and they have confirmed the musics were appealing to hear.

Acknowledgements

It is quiet a long journey to learn and successfully complete a research.So lot of people from academic side and family end helped to succeed in this work.First I would like to thank my supervisor for giving me full freedom to do the work as my way and guiding me when I need help. Then I have thank my parents and sister helping me and encouraging me to go for success.My friends were always with me to help when I really need motivation to work on research with the hectic work life in this country.

Contents

Declaration i

Abstract ii

Acknowledgements iii

1 Introduction 1
- 1.1 Prolegomena . . . 1
- 1.2 Objectives . . . 1
- 1.3 Background and Motivation . . . 2
- 1.4 Problem in brief . . . 3
- 1.5 Proposed solution . . . 3
- 1.6 Resource requirements . . . 4
- 1.7 Structure of the thesis . . . 4
- 1.8 Summary . . . 4

2 Issues and Challenges 5
- 2.1 Introduction . . . 5
- 2.2 Gestation of music generation for scene emotion . . . 5
- 2.3 Major developments in music generation . . . 6
- 2.4 Major development in image sentiment identification . . . 9
  - 2.4.1 Problems In Image Classification . . . 10
- 2.5 Problem definition . . . 13
- 2.6 Summary . . . 13

3 Technology For Scene Emotion Detection And Music Generation 14
- 3.1 Introduction . . . 14
- 3.2 Convolution Neural Networks(CNN) . . . 14
  - 3.2.1 Convolution Neural Networks performance Improvement 15
- 3.3 Generative Adverserial Networks(GANs) . . . 16
  - 3.3.1 Wasserstein Generative Adverserial Networks(WGANs) 16
  - 3.3.2 How Generative Adverserial Networks used in image classification . . . 17
- 3.4 Long Short Term Memory(LSTM) based music generation . . . 17
- 3.5 Other libraries and technologies . . . 18
- 3.6 Summary . . . 18

4 Approach 19
- 4.1 Introduction . . . 19
- 4.2 Hypothesis . . . 19
- 4.3 Process . . . 19
- 4.4 Input . . . 20
- 4.5 Output . . . 20
- 4.6 Features . . . 21
- 4.7 Users . . . 21

4.8 summary . . . 21

5 Design **22**
5.1 Introduction . . . 22
5.2 Architecture of the project . . . 22
5.3 Image classifier for emotion detection . . . 23
5.3.1 GAN fake image generator . . . 24
5.4 Music generation . . . 24
5.5 Summary . . . 25

6 Implementation **26**
6.1 Introduction . . . 26
6.2 Dataset preparation . . . 26
6.3 GAN implementation . . . 26
6.4 Image classifier implementation . . . 29
6.5 Music generator implementation . . . 30
6.6 Web Api implementation . . . 30
6.7 Summary . . . 32

7 Evaluation **33**
7.1 Introduction . . . 33
7.2 Evaluation Procedure . . . 33
7.2.1 Evaluation Procedure of music files . . . 33
7.2.2 Evaluation Procedure of Image classifier . . . 33
7.3 Image Classification Results . . . 33
7.4 Sound Generation Results . . . 35
7.5 summary . . . 36

8 Conclusion and Future Work **37**
8.1 Introduction . . . 37
8.2 Conclusion . . . 37
8.3 Future work . . . 38
8.4 summary . . . 38

A Code Implementations **39**
A.1 Image Classifier Implementation . . . 39
A.2 Music Generator . . . 40
A.3 GAN Implementation . . . 46