VISVESVARAYA TECHNOLOGICAL UNIVERSITY
BELGAUM-590014

Synopsis Entitled

"Simulation of Musical Instruments using Neural Networks"

Submitted for

BACHELOR OF ENGINEERING
In
COMPUTER SCIENCE AND ENGINEERING

For the Academic year **2017-2018**

Submitted by

Chandra Mahendra Vikram Singh	**1MV14CS027**
Himanshu Kumar	**1MV14CS040**
Jubin George Mathew	**1MV14CS046**
K Vishnudev	**1MV14CS049**

Project carried out at
Sir M. Visvesvaraya Institute of Technology
Bangalore-562157

Under the Guidance of
Mrs. Sheela S. Kathavate
Associate Professor, Department of CSE
Sir M Visvesvaraya Institute of Technology, Bangalore

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
SIR M. VISVESVARAYA INSTITUTE OF TECHNOLOGY
HUNASAMARANAHALLI, BANGALORE-562157

ABSTRACT

Music is an important entity in everyone's life. Similarly, musical instruments have a paramount place in Music. The art of playing musical instruments have been traditionally carried on from generation to generation, from primitive stage to this modern age. As Music is a performing art, which is being creative and cannot be static, hence the gradual developments and experiments have always given new ideas to the modern generation.

This project highlights experimentation in Instrumental Music which leads to digitalised instruments in this sector. Electronic Instruments are instruments which help to electrically produce a symphony of its own may it be an electric guitar in a rock concert or a keyboard in orchestra. This project introduces an end-to-end neural network model for playing the sound of a musical instrument based on a silent video of it being played. At a high level, the model consists of a convolutional neural network (CNN) to extract features from the raw video frames and neural autoregressive models to encode the spatiotemporal features of the video.

TABLE OF CONTENTS

INTRODUCTION

Nowadays machine learning is in every domain of the industry like online advertising, recommendation system, search engine, machine locomotive, pattern recognition etc. Also due to its pronounced effectiveness its applicability keeps growing bigger. Machine learning is an application of artificial intelligence that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it to learn for themselves.

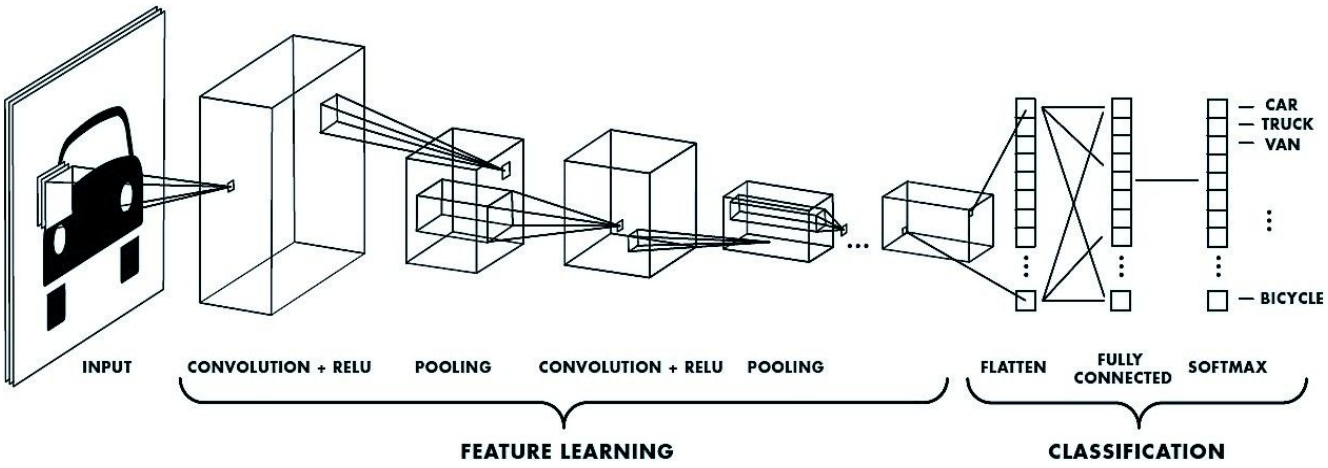
In machine learning, a concept called convolutional neural network (CNN, or ConvNet) is used extensively in image and video recognition. It is a class of deep, feed-forward artificial neural networks that has successfully been applied to analyzing visual imagery. CNNs use a variation of multilayer perceptrons designed to require minimal preprocessing. They are also known as shift invariant or space invariant artificial neural networks (SIANN), based on their shared-weights architecture and translation invariance characteristics. CNNs use relatively little pre-processing compared to other image classification algorithms. This means that the network learns the filters that in traditional algorithms were hand-engineered. This independence from prior knowledge and human effort in feature design is a major advantage. They have applications in image and video recognition, recommender systems and natural language processing.

This project redefines the way one uses music instruments to generate sound to compose music. A keyboard instrument is a musical instrument played using a keyboard, a row of levers which are pressed by the fingers. The most common of these are the piano, organ, and various electronic keyboards, including synthesizers and digital pianos. Today, the term keyboard often refers to keyboard-style synthesizers. Under the fingers of a sensitive performer, the keyboard may also be used to control dynamics, phrasing, shading, articulation, and other elements of expression—depending on the design and inherent capabilities of the instrument.

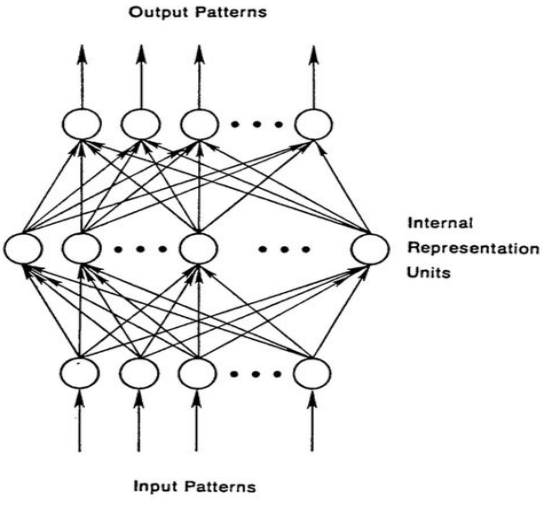
LITERATURE SURVEY

There are wide range of instruments for music generation, they are either expensive or colossal to carry and sometimes both. So, this model tries to develop a system to mimic multiple instruments or input devices using the same environment in a more economic manner. Different Input Output devices such as keyboard, mouse or gesture could be used as input for electronic gadgets. They also suffer the same set of limitations as musical instruments. Thus to overcome such limitations, we have an elegant solution which involves building a generic learning model which can be trained for any specific device and the same can be achieved by passing the input received from these devices to a Convolutional neural network.

In automated learning, convolutional neural networks (CNN, or ConvNet) [1] are a class of deep-forward artificial neural networks that have been used for image analysis. These networks uses convolution layers in its core. The images are broken down into smaller units, which are a set of lernable filters called kernels. As a result, the network learns filters that activate when it detects some specific type of feature at some spatial position in the input.



Recurrent neural networks (RNN) [2] were created in the 1980s but have recently gained popularity with advances to network designs and increased computational power of graphics processing units. They are particularly useful with sequential data because each neuron or unit can use internal memory to maintain information about previous inputs.



OBJECTIVE

The project aims to provide a digitalized experience to the end user by simulating musical instruments.

Key objectives are:

1. To research feature extraction from live video feed.
2. To research and develop a convolutional neural network to classify and extract features from the feed.
3. To develop a model which can effectively predict the user activity.
4. To research and build a prototype of a system to present the working model of end product. The prototype will provide means to input live video via a camera module and an audio output device for generating sound.
5. To research and develop a model for playing/generating music in correspondence with key pressed.

SCOPE

Scope of the project is to develop and implement a working model to detect the movement of hands and play music accordingly. The focus is to develop a cost effective and a portable alternative to existing musical instruments that are expensive and bulky.

The setup includes an optical device to capture video for detection of motion. This project is limited only to keyboard instruments and other instruments are out of scope of this project. Accuracy of the model, quality of capturing device and nondeterministic variation in processing time may widely affect the performance of this model.

The project is expected to be completed within a period 4 - 6 months. This time span will involve building a custom hardware package from existing embedded system to give an independent kit to practice music.

METHODOLOGY

We will design a layout of the instrument and a camera module for capturing the motion of the hand while interacting with the instrument. The simulated instrument will have a printed canvas to showcase the physical layout of the instrument.

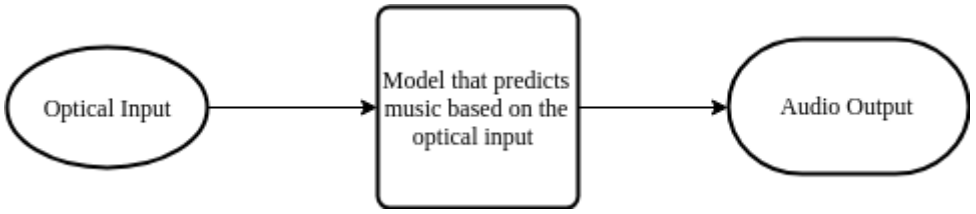


Figure 5.1

The live video feed from the camera module is used as an input for processing. Video frames from input are used to train the model. After the frames are analysed in a sequential manner, sound is played by predicting keys pressed. These musical notes are determined based on the motion and orientation of the hand.

The captured video is preprocessed to ensure feasible image processing. The images are again processed so as to make the process of pattern recognition easier, thus leading to faster and accurate prediction.

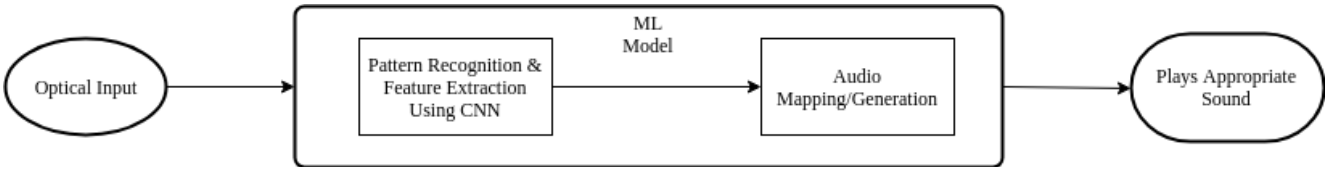


Figure 5.2

The captured images are analysed are passed through the ML system which uses the neural network (NN). The features extracted by the Neural Networks is used for searching patterns in the image so that the appropriate prediction can be predicted. The music is thus played based on the predicted action which are taken by the user.

TECHNOLOGY

Tools and Libraries

- **Gradient Descent** : Gradient descent is a first-order iterative optimization algorithm for finding the minimum of a function.
- **TensorFlow** : An open source software library for numerical computation using data flow graphs. It was originally developed by the Google Brain Team within Google's Machine Intelligence research organization for machine learning and deep neural networks research.
- **Numpy** : a python library used for numerical computations and multi dimensional matrix operations.
- **Python** : 2.7 or above

Hardware requirements

- **CPU** : 2.4 GHz quad core, x86-64 architecture
- **GPU** : NVIDIA or any other TensorFlow supported GPU, CUDA or cuDNN
- **RAM** : 4 GB memory

CONCLUSION

This project uses Convolutional Neural Networks model for detecting hand motion of the end user and hence playing the audio in correspondence with the musical instrument from the video feed. This model uses convolutional neural network which parses each video frame and extracts respective feature set to feed the model. The model doesn't require explicit manual labeling, which helps to avoid the overhead . In future, we will be working on improving its latency in order to make it more realistic and expeditious. This would lead to an increase in its real time applications and gradually an alternative for bulky and costly instruments.

BIBLIOGRAPHY

[1] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems* 25, pages 1097–1105. Curran Associates, Inc., 2012.

[2] https://deeplearning4j.org/lstm.html

[3] Yu, Fisher, and Vladlen Koltun. "Multi-scale context aggregation by dilated convolutions" arXiv preprint arXiv:1511.07122 (2015). https://arxiv.org/abs/1511.07122

Head of Department

Project Guide

Project Coordinator