SOUND SYNTHESIS FROM SCRIPT

A Thesis

Submitted by

Nivetha S (RCAS2021MDB021)

in partial fulfillment for the award of the degree of

MASTER OF SCIENCE
SPECIALIZATION IN
DATA SCIENCE AND BUSINESS ANALYSIS

DEPARTMENT OF COMPUTER SCIENCE

RATHINAM COLLEGE OF ARTS AND SCIENCE

(AUTONOMOUS)

COIMBATORE - 641021 (INDIA)

MAY-2023

RATHINAM COLLEGE OF ARTS AND SCIENCE
(AUTONOMOUS)
COIMBATORE - 641021

BONAFIDE CERTIFICATE

This is to certify that the Thesis entitled **"SOUND SYNTHESIS FROM SCRIPT"** submitted by **Nivetha S,** for the award of the Degree of Master in Computer Science specialization in **"DATA SCIENCE AND BUSINESS ANALYSIS"** is a bonafide record of the work carried out by her under my guidance and supervision at Rathinam College of Arts and Science, Coimbatore.

Ms. Sujina S, M.Tech., (Ph.D).,
Supervisor

Dr.P.Sivaprakash, M.Tech., Ph.D.,
Mentor

Submitted for the University Examination held on 09.05.2023

INTERNAL EXAMINER

EXTERNAL EXAMINER

RATHINAM COLLEGE OF ARTS AND SCIENCE

(AUTONOMOUS)

COIMBATORE - 641021

DECLARATION

I, **S.Nivetha**, hereby declare that the Thesis entitled **" SOUND SYNTHESIS FROM SCRIPT",** is the record of the original work done by me under the guidance of **Ms. Sujina S, M.Tech., (Ph.D).,**, Faculty Rathinam college of arts and science, Coimbatore. To the best of my knowledge this work has not formed the basis for the award of any degree or a similar award to any candidate in any University.

Signature of the Student

Nivetha S

Place: Coimbatore

Date: 09.05.2023

COUNTER SIGNED

Ms. Sujina S, M.Tech., (Ph.D).,

Supervisor

Contents

Acknowledgement

On successful completion for project look back to thank who made in possible. First and foremost, thank **"THE ALMIGHTY"** for this blessing on me without which I could have not successfully our project.I am extremely grateful to **Dr.Madan.A. Sendhil, M.S., Ph.D.,** Chairman, Rathinam Group of Institutions, Coimbatore and **Dr. R.Manickam MCA., M.Phil., Ph.D.,** Secretary, Rathinam Group of Institutions, Coimbatore for giving me opportunity to study in this college.I am extremely grateful to **Dr.S Balasubramanian,M.Sc.,PhD,(Swiss),PDF(SwissuSA)** Principal Rathinam College of Arts and Science(Autonomous), Coimbatore.Extend deep sense of valuation to **Mr.A.Uthiramoorthy, M.C.A., M.Phil., (Ph.D),** Rathinam College of Arts and Science (Autonomous) who has permitted to undergo the project.

Unequally I thank **Dr.P.Sivaprakash, M.Tech., Ph.D.,** Mentor and **Dr.Mohamed Mallick, M.E.,** Project Coordinator, and all the Faculty members of the Department - iNurture Education Solution Pvt Ltd for their constructive suggestions, advice during the course of study.I convey special thanks, to the supervisor **Ms.Sujina S, M.Tech., (Ph.D).,** who offered their inestimable support, guidance, valuable suggestion, motivations, helps given for the completion of the project.

I've dedicated sincere respect to my parents for their moral motivation in completing the project.

List of Figures

List of Abbreviations

NN	Neural Network
CNN	Convolutional Neural Network
RNN	Recurrent Neural Network
CTC	Connectionist Temporal Classification
OCR	Optical Character Recognition
TTS	Text-To-Speech

Abstract

The growing volume of handwritten data presents challenges in terms of accurate and efficient processing. To address this issue, this research proposes a cost-effective solution that enables users to listen to the content of text rather than reading it. The technology used combines text-to-speech (TTS) and neural networks (NNs) principles. Specifically, convolutional neural networks (CNNs) are used to recognize the structure of handwritten characters and words, aiding in the automated extraction of distinguishing properties. With this approach, texts in various formats can be identified and converted into speech. The proposed system makes it easier for visually impaired individuals to use computers and other technologies by converting text into speech. This technology has the potential to benefit many people who are unable to read and write letters, making it possible for them to access educational content and become more educated individuals. Overall, this research presents an innovative solution for converting handwritten text into audio, enabling greater accessibility and inclusivity in education.

Chapter 1

Introduction

There are numerous jobs that machine cannot complete on their own can be performed by humans. Despite the fact that text recognition in handwritten documents has been one of the most popular research areas over the past few decades. Handwritten Text Recognition is one of the challenging tasks because of huge variation and inconsistency in handwriting styles which varies from person to person. Converting handwritten text into machine-readable text formats is difficult due to the wide variation in handwriting styles across individuals and the inferior quality of handwritten text compared to machine-printed text. However, it's a critical issue that needs to be resolved for a number of sectors, including banking, healthcare, and insurance.

As computer technology advanced, the format of handwritten text quickly shifted to computer-generated digital text, necessitating the creation of a method to convert handwritten text to digital text since it speeds up and simplifies the processing of such data. To convert these simplified handwritten texts to audio, numerous automatic handwritten systems have been developed by various academics in the past which can convert the words to text and audio. In this paper, a process which extracts handwritten

text from scanned or printed text images into editable text and further transforming the text speech synthesis is provided.

Neural Networks are designed based on the structure of the human brains and are particularly effective at tackling issues like pattern recognition, classification of distinct classes, data mining, and series prediction that cannot be solved in a sequence of straightforward stages. These Neural Networks Model consists of Convolution Neural networks (CNN) layers, Recurrent Neural Network (RNN) layers and Connectionist Temporal Classification (CTC) layers which is used to convert handwritten words to text.

The Designing computer system which deals with the voice/speech in the field of computer science is used to synthesis the written text. The basic idea of Text-to-Speech (TTS) technology, is to transform the written output to the spoken output. Thus TTS, makes the computer to speak for the better understanding of handwritten words.

1.1 Objective of the project

A Language is a tool to share our ideas and thus making a communication between us. It is possible to make it done through written signs (text), hand gestures, and spoken words. It is a distinctive feature of human beings to be the only creatures to use such these systems. Text is the most common and oldest form of human communication. The artificial fabrication of converting text to speech is referred to as "Speech Synthesis", sometimes known as "Text to Speech Synthesis". A voice synthesizer is a type of computer system that serves this purpose and can be included in the software. Speech

synthesis has evolved over the years and has been incorporated into many different systems and applications. The method of creating speech synthesis is challenging. TTS has numerous objectives, but its primary objective is to give consumers access to high-quality speeches from handwritten text.

The general objective of this project is to convert handwritten text to speech synthesizer for physically disabled and vocally troubled people. The specific goals of the project are,

- To implement an isolated whole-word speech synthesizer that is capable of converting text and responding with speech
- To make it possible for people who are deaf or dumb to communicate and contribute to the expansion of an organization.
- To provide an accessible computer interface for the aged and blind.
- To create a user-friendly interface for people who are unable to read words and letters.
- To create modern technology and awareness of paving the way of making educated individuals.

.

1.2 Scope of the Project

The study is focused on an ideal combination of a human-like behavior with computer application to build a one-way interactive medium between the computer and the user. The scope of the project is to build an application that aid people with disabilities especially on reading and also helps them to get easier information without any stress.This could also help children learning and improving their communication skills,pronouncing words and making them the subjects easily.

Any researcher who wishes to explore the Impact of employing a Computer Speech Program for brain augmentation and assimilation process in Humans" will find the speech synthesizer to be of great assistance. This text-to-speech synthesizing system will help close the digital divide, by enabling the semi-illiterates to evaluate and read through electronic materials.

The most evident disadvantage of text as a medium for knowledge construction and it lacks the inherent expressiveness of speech. When a text is transcribed to audio, it loses many of its unique qualities like tone, rhythm, repetition of the words and so on. Correctness of all these loses leads to increase in educated individuals, modern technologies and making blind to learn the beautiful words around them, etc.,

1.3 Module Description

- **Neural Network:**

  A neural network is an artificial intelligence technique that instructs computers to

analyze data in a manner modelled after the human brain. It is a kind of artificial intelligence technique known as deep learning that makes use of interconnected neurons or nodes in a layered structure to mimic the human brain. Computers can use this to build an adaptive system that helps them continuously get better by learning from their failures. Neural networks adapt the changing input so, the network generates the best possible output without needing to redesign again.

- **CUDA Toolkit:**

  CUDA (Compute Unified Device Architecture), is a parallel computing platform and programming model that dramatically boosts computing performance by utilizing the capabilities of the graphics processor unit (GPU).

- **MXNet:**

  Thus the library named MXNet is mainly used for the conversion of handwritten words to text. To define, train, and deploy deep neural networks on a variety of devices, from cloud infrastructure to mobile devices, using the open-source deep learning framework MXNet. It enables a flexible programming model, different languages, and great scalability, enabling quick model training.

- **CV2:**

  The module import name for OpenCV-python, "Unofficial pre-built CPU-only OpenCV packages for Python" is cv2. Traditional OpenCV requires extra and lengthy steps that involve creating the module from scratch which is unnecessary.

- **TTS**

  TS or text-to-speech, is a software function that reads text and produces synthesized speech from it.TTS is very precise, yet it tends to lack the entire range of emotional expression that humans naturally produce. Almost any text-based message can be transformed into an understandable spoken message using TTS. TTS has a wide range of practical commercial uses, particularly in assisting teams to efficiently and affordably provide bulk notifications across text and spoken communication channels

  The Benefits of Text-To-Speech conversion are listed as:

  1..Customizable phrase tokenizer for a voice that maintains appropriate intonation, abbreviations, and decimals while allowing for infinite text lengths to be read.

  2.Text pre-processors with custom settings that, for instance, can fix pronunciation.

  3.Retrieval of supported languages automatically.

1.4 Existing System

Handwritten to Text

Now a days the handwritten text are becoming less as there is a increase in the technologies. Thus, the words are digitalized to make everyone understand and easy.The most typical application of neural networks may be pattern recognition. A different

class of target vectors and their corresponding input vectors are supplied to the neural network.This consist of convolutional neural network(CNN) layers, Recurrent neural network(RNN) and Connectionist Temporal Classification(CTC) layer.

Handwritten to Text using Tensor flow: The system is built with the Nerual Network(NN) which is trained on word images as a input layers(and therefore also all the other layers) can be kept small for word-images, NN-training is feasible on the CPU which is used to convert the handwritten words to the text. Thus this procedure is barely minimum in HTR using TF.

Handwritten Text Recognition using Neural Networks:

The existing system proposed the study of converting the handwritten words to the text. A System that can process handwritten English characters as input, extract the best features, train a neural network by identifying the type of input text, and finally produce the computerized version of the input text.The task is performed by using NN model.

Chapter 2

Literature Survey

The realisation of handwritten characters is precisely sizzling due to a wide range of applications, such as banking, postal service, and digital libraries. Application form processing, digitising historical documents, postal code processing, bank transaction processing, and many other applications are examples of the maturation comedian in the field of handwritten character processing. The handwritten recognition feature of the device translates the user's handwritten characters or words into a language that the computer can understand. Numerous machine learning and deep learning techniques have been planned for effective recognition. In this research, we present a comprehensive analysis of the phases of handwritten character recognition as well as numerous approaches and techniques in CNN, RNN in neural Networks.

[1]In this essay, recent character recognition methods are compared and analysed. Our goal is to evaluate the effects of machine learning on character recognition. In order to speed up various processes, character recognition has many applications in the sectors of finance, healthcare, and other industries. These applications include searchability, storability, readability, editability, accessibility, etc. Thus classification

methods, traditional machine learning techniques including neural networks, support vector machines, random forests, etc. have been employed. Deep learning algorithms have now come to be thanks to improvements in computer hardware and productive research in the area of artificial intelligence. Deep learning is being used in recent papers to identify characters. They also show how various functions enhance the field's performance.

[2]A discipline known as optical character recognition makes it possible to convert many kinds of texts or photos into editable, searchable, and analyzable data. In the past ten years, academics have developed systems that automatically evaluate printed and handwritten documents in order to convert them to electronic format. This review paper's goals are to present research directions and a summary of previous studies on character recognition in handwritten texts.An OCR system is primarily dependent on the extraction of features and the classification/discrimination of these features (based on patterns).As a subspecialty of OCR, handwritten OCR has drawn more and more attention. Depending on the input data, it is further divided into offline and online systems. While the nature of input in online systems is more dynamic and is based on the movement of a pen tip with a specific velocity, projection angle, position, and locus point, the offline system is a static system in which input data is in the form of scanned photographs. Because it eliminates the issue of input data overlap that exists in the offline system, an online system is seen as more complex and advanced.

[3]How to identify and classify images is one of the computer-related issues that is being looked for and investigated by computers that can detect pictures just like

people do. Handwriting is one that can be identified from an image, and it can be used to support human tasks like check analysis and the processing of handwritten forms. In image recognition, the view angle, lighting, and clarity of the captured image all have an impact on how well the image is recognised. The method that can be used to recognise handwriting is one of the many methods that will be explored in this paper.

[4]In order to improve the real accuracy attained in correctly recognising written language, further research is required given the growing amount of handwritten data and the new advances in computing power. In line with the fact that convolutional neural networks (CONVNETs) algorithms are incredibly effective at extracting structure from images, they are also incredibly capable of recognising handwritten textual characters/words in ways that facilitate automatic recognition of distinctive characteristics. The best approach for tackling problems with handwritten text recognition is therefore to use CONVNETs algorithms. Text in a variety of forms will be recognised using this method. A vast range of handwritten textual characters, including numbers, characters, scripts, and other symbols, demonstrate how handwriting has evolved into a more complex art form.

[5]Computer vision systems still struggle to recognise unrestricted handwritten text. Traditionally, two models—the first for line segmentation and the second for text line recognition—are used to recognise text in paragraphs. To do this, we put out a comprehensive end-to-end model that makes use of hybrid attention. This model is made to iteratively process an image of a paragraph line by line.It converts the handwritten words to text.

[6]Due to a variety of backgrounds, noises, writing styles, and several touches between characters, offline handwritten text detection remains a difficult task. We suggest a model of the 2D Self-Attention Convolutional Recurrent Network (2D-SACRN) in this paper for identifying handwritten text lines.

[7]In the near future, the digitization and processing of the current paper documents could play a significant role in the creation of a paperless environment. Deep learning techniques for handwritten recognition have been extensively studied by various researchers. Deep neural networks can be trained quickly thanks to a lot of data and other algorithmic advancements. In the literature, various methods for extracting text from handwritten manuscripts have been proposed. The suggested model is tested using the IAM and RIMES handwritten databases, yielding results that are competitive in terms of word and letter accuracy as well as text conversion.

Chapter 3

Methodology

3.1 Neural Network

Deep learning techniques are based on neural networks, sometimes referred to as artificial neural networks (ANNs) or simulated neural networks (SNNs), which are a subset of machine learning. Their structure and nomenclature are modelled after the human brain, mirroring the communication between organic neurons.

A node layer of an artificial neural network (ANN) consists of an input layer, one or more hidden layers, and an output layer. Each node, or artificial neuron, is connected to others and has a associated weight and threshold that go along with it.Any node whose output exceeds the defined threshold value is activated and begins providing data to the network's uppermost layer. Otherwise, no data is transmitted to the network's next tier.

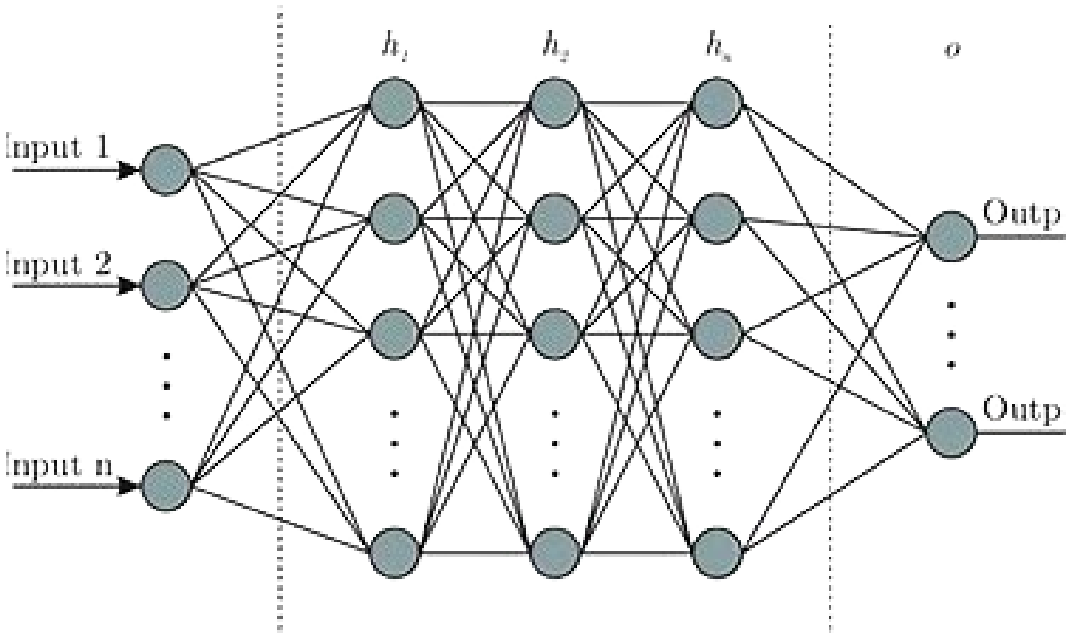

Figure 3.1: Neural Network

3.2 Convolutional Neural Network

Artificial neural networks are employed in a variety of categorization tasks including words, audio, and visual input. Different types of neural networks are employed for various tasks. For example, predicting the order of words requires the use of recurrent neural networks, specifically an LSTM, while classifying images requires the use of convolutional neural networks. There are three categories of layers in a standard neural network:

- Input Layers: It is the layer where we input data into our model. The overall number of characteristics in our data is equal to the number of neurons in this layer (number of pixels in the case of an image).

- Hidden Layer: The hidden layer then gets the data from the input layer. Depending on our model and the size of the data, there may be numerous hidden levels. The number of neurons in each hidden layer might vary, but they are typically more than the number of features. Each layer's output is calculated by matrix multiplication of the 14 output from the layer before it with learnable weights from that layer, adding learnable biases after that, and then computing the activation function, which makes the network nonlinear.

- Output Layer: The output of each class is then converted into the probability score for each class using a logistic function, such as sigmoid or softmax, using the data from the hidden layer as input

Convolutional Neural Network Design A convolutional neural network's architecture is a multi-layered feed-forward neural network created by sequentially stacking numerous hidden layers on top of one another. Convolutional neural networks can learn hierarchical features because of their sequential construction.

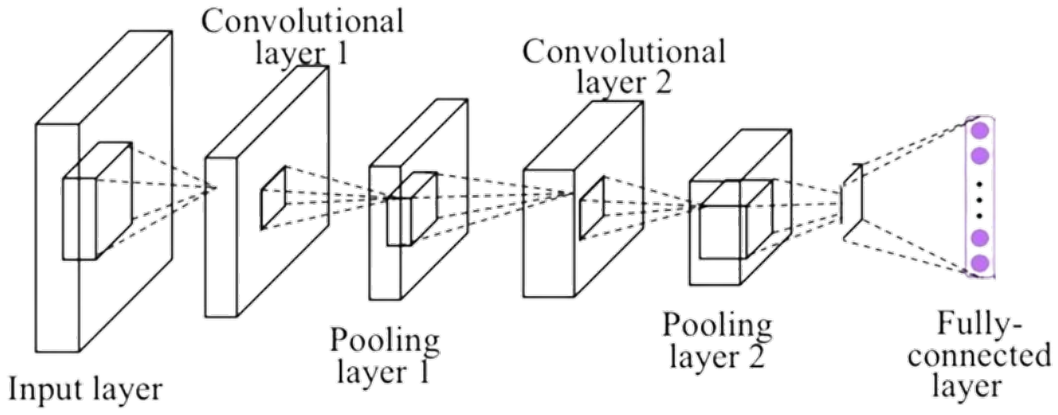

Figure 3.2: Convolutional Neural Network

Convolutional layers are frequently followed by activation layers, some of which are then followed by pooling layers, as the hidden layers.

For the implementation of CNN:

To start with importing the libraries, data preprocessing followed by building a CNN, training the CNN, and lastly, we will make a single prediction. All the steps will be carried out in the same way as we did in ANN,the only difference is that now we are not pre-processing the classic dataset, but some images, which is why the data preprocessing is different and will consist of doing two steps, i.e., in the first, we will pre-process the training set and then will pre-process the test set.

3.3 Recurrent Neural Network

An artificial neural network that employs sequential data or time series data is known as a recurrent neural network (RNN). These deep learning algorithms are included into well-known programmes like Siri, voice search, and Google Translate. They are frequently employed for ordinal or temporal issues, such as language translation, natural language processing (nlp), speech recognition, and image captioning. Recurrent neural networks (RNNs) use training data to learn, just like feedforward and convolutional neural networks (CNNs) do. They stand out due to their "memory," which allows them to affect the current input and output by using data from previous inputs.

Recurrent Neural Network vs. Feedforward Neural Network

Recurrent networks are distinguished by the fact that each layer of the network uses the same parameters. Recurrent neural networks share the same weight parameter

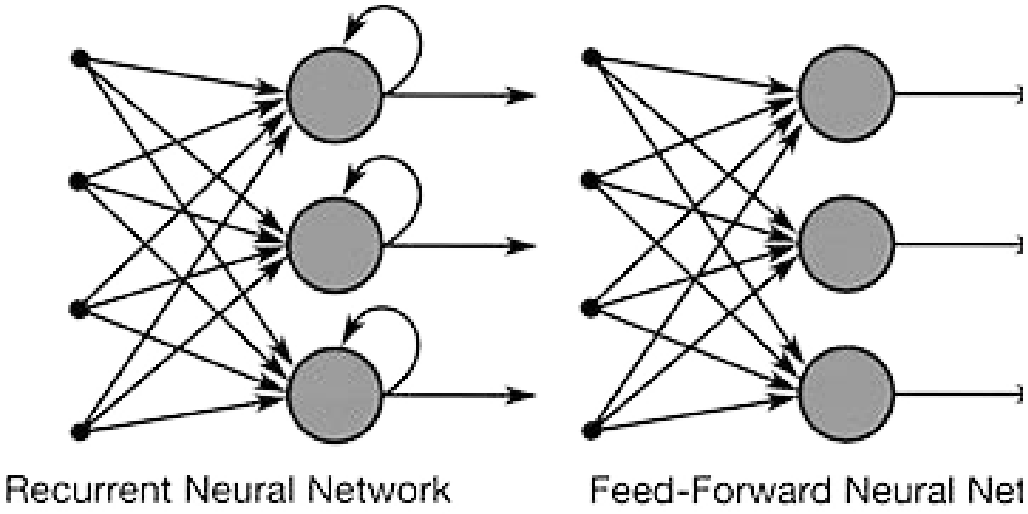

Figure 3.3: Recurrent Neural Network

inside each layer of the network, in contrast to feedforward networks, which have distinct weights across each node. However, to support reinforcement learning, these weights are still modified using the techniques of backpropagation and gradient descent.

3.4 Long Short-Term Memory - LSTM

LSTM basically poss ed with memory blocks, simply it creates an input memory block which reduces the smaller gradient effect, and previous inputs which are fed into the input layer is controlled by forget gate and make LSTM ahead from the vanishing gradient issue, forget gate used to determine the state of input whether its remembered or forgot-ted. Totally LSTM has three gates input gate, output gate and forgot gate.

The entire flow of LSTM is been shown in fig, x_t as input, h_{t-1} as a output cell. In

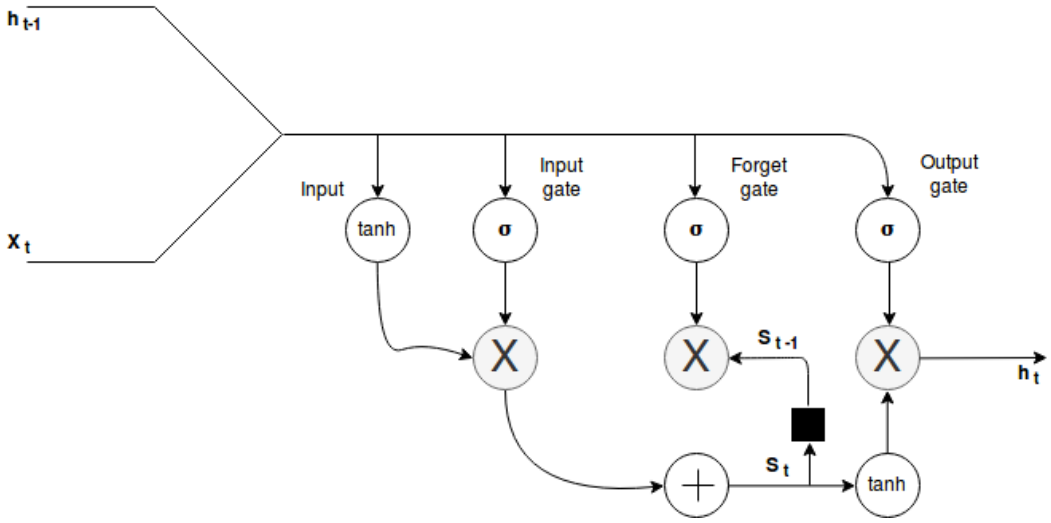

Figure 3.4: Long short term memory

this activation function *tanh* is take and distinguishing values are from -1 to 1. The following theroy is expressed as

$$g = tanh(b^g + x_t R^g + h_{t-1} J^g) \tag{3.1}$$

Where weight from the previous cell input and output is denoted by R^g and J^g. Bias is denoted by b^g, g is the input weight of the neuron.

The output of the input gate undergoes element wise multiplication, with sigmoid nodes with x_t of weights and input value ofh_{t-1} . The input gate of LSTM is expressed as

$$k = \sigma(b^k + X_t U^k + h_{t-1} V^k \tag{3.2}$$

The output of the LSTM cell is expressed as where $\circ$ is multiplication operator element wise.

$$g \circ k \tag{3.3}$$

In the forget gate a new state is introduce S_t . This inner state S_t provides a internal loop with time step adding to the $g \circ k$ input state. Forget gate usually set a node for activation function which is S_{t-1}. It Remembers the information from the previous input state, this helps LSTM to learn the exact context fed into the network. The Forget gate is expressed as

$$l = \sigma(b^1 + X_t U^1 + h_{t-1} V^1) \tag{3.4}$$

S_{t-1} is output of the forget gate and each time the inputs added to the this gate, all inputs are filtered without multiplication it is mixed along with sigmod function and weights. This helps in eliminating vanishing gradient problem. Finally the output gate posses with gating function from each cell and produce the output.

3.5 TTS

The goal of text-to-speech (TTS) is to produce natural-sounding voice from text input.

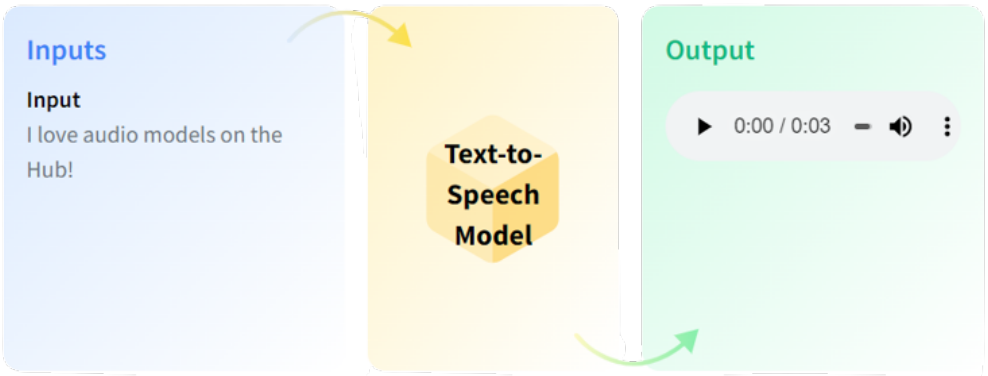

Figure 3.5: Text-To-Speech

TTS models can be improved such that a single model can provide speech for a variety of speakers and languages.

Use Cases: Any voice-enabled application that requires text to speech conversion can employ text-to-speech (TTS) models.Text-to-speech is used by a small firm to configure its auto-attendant. The software makes it simple for the business to add consistent and understandable voice messages to its phone menu system as new items or advancements are introduced. Depending on the preferences of the students, an educational institution will use text-to-speech to deliver mass notice notifications about future events to both text and recorded phone messages.

Voice Assistants:

On smart devices, voice assistants are made using TTS models. Since the outputs of TTS models include components of natural speech, such as emphasis, they are a preferable alternative to concatenative approaches, where the assistant is constructed by recording sounds and mapping them.

Announcement Systems: TTS models are frequently used in airport announcement systems and public transportation announcement systems to turn announced text into speech.

3.6 Working of TTS:

A frontend and a backend make up the engine that drives text-to-speech technology.Raw text with symbols is transformed into written-out words at the front end. The process of text normalisation is followed by a phonetic transcription match on the backend.

The real sound you hear is produced by the backend as well. For this reason, experts call the backend the synthesiser.

Depending on the service you're using, you may or may not hear a specific voice. Many people have not previously employed text to speech technologies because these synthetic voices have historically sounded unnatural.Even synthetic voices with the ability to add intonation and emotion are now present in many of the more advanced TTS programmes.Advanced TTS solutions currently provide a number of commercial advantages that make this assistive technology accessible to a broad audience.

3.7 Pyttsx3:

A Python library called pyttsx3 offers a cross-platform API for producing synthesised speech. It generates voice using the operating system's built-in text-to-speech engines. It provides a cross-platform API for generating synthesized speech.Using the operating system's built-in text-to-speech algorithms, you can turn written text into spoken words.The library is easy to use and provides a simple API for generating speech in Python scripts and applications. It can be used to create speech-enabled applications, generate audio books, and assist people with visual impairments.

3.8 Advantages

By transforming the scribbled words to text and audio, the initiative offers a big advantage for many industries, including the medical, IT, banking, healthcare, and insurance sectors. As a result, the proposed system approach opens the door for deaf and dumb

people to work in numerous organisations. Many impaired individuals who are unable to talk benefit from the transformed audio. The majority of people are better listeners than readers. These individuals can now fully understand the information because to this project.Thus the suggested system helps the people to become educated in every sectors.

3.9 System Design

The input image is fed into the CNN, RNN and CTC layers These layers are trained to extract relevant features from the image. Thus the extracted input is further converted to the audio by using TTS technology. The text-to-speech/audio system technology makes it possible for a computer to speak. When text is supplied into the TTS system, a computer algorithm known as the TTS engine analyses it, pre-processes it, and then uses some mathematical models to synthesis voice.

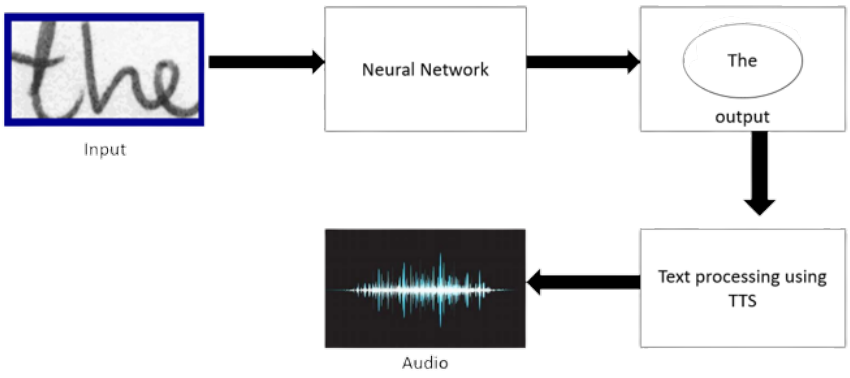

Figure 3.6: Architectural diagram

Chapter 4

Experimental Setup

4.1 Data Set

The make everyone speak to the world, the data is collected from IAM which contains number of handwritten images.Thus these are fed into the layers for text conversion and further converted to audio by using TTS technologies.

An putpose is the intention of the user interacting with everyone or the intention behind each message should be correctly received to everyone.

4.2 Data Preperation

Load the Zip file and extract the required data. The given variable holds all the tokenized data (which are sample text messages in images) for training. The variable holds all target labels/images correspond to each training data and these data are further trained by the by the provided model. After this define letters, variable, and characters from input data for training to give the desired text result.

This is the basic overview of data pre-processing model. In data classification, there will be two sets of data as Training data and Testing data. Once the model is created.

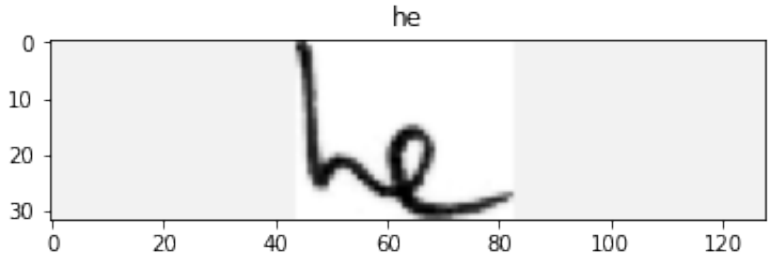

Figure 4.1: Dataset flow work

4.3 Model Training

Define Neural Network architecture for proposed model. Variables are defined to store the values which are needed to design and train the model. Further the content is computed and featurized by the layers of CNN.

The featurized data is encoded by Long Short Term Memory(LSTM). The LSTM will take the data in temporal order (left to right), SEQ LEN slices and will encoded them. Then the custom layer is defined where the data is reshaped from (N, W, H) to (SEQ LEN, N, CHANNELS) , run the biDirectional LSTM, and again reshape the data from (SEQ LEN, N, HIDDEN UNITS) to (N, SEQ LEN, HIDDEN UNITS). The resulting encoded slices will be fed to a fully connected layer of size ALPHABET SIZE and the result will be fed to a CTC loss.

After the detection of the loss, the values are assembled and the parameters are initialized for training. After training, evaluation begins and the loop continues until the layers of the data ends. Then the evaluated data is loaded and saved in a variable for the further process.

Figure 4.2: Training of the dataset

The above figure is the training of the handwritten dataset. Thus the image is fed into the model in the form of png(where the backgorund is excluded) and stored in the variables. Further, the model is trained with the different types of alphabet and character which are not invloved in the before training.Now, the letters and characters are totally trained which are needed to convert the handwritten word to text.In the following, the length, width and weight of the words are defined, so that the model can give the accurate text result.

Fig 4.3 is the sample of the input data which contains the handwritten words. Thus these words are analyzed by the model and trained by CNN, RNN layers. The loss is calculated by CTC layer. By continuous training of the model with the handwritten words, the CTC loss can be reduced to give the high efficiency text.

Figure 4.3: Training of the dataset

Thus the converted handwritten text is converted into the speech/audio by the technolgy called Text-To-Speech(TTS). The main work of the TTS is to convert the text to audio(to make the computers to speak).Communication is the better way for making one understand rather than the text.The improvement in the technology paves for these ideas. The developing technology of TTS, helps the world to communicate with each other by speaking. Thus conversion of words to audio makes the people more knowledgeable and educated, which finds the way for modern world and technology.

The pyttsx3 is a python library which is used to convert the text to an audio file with the help of the TTs library. Here, the handwritten dataset is first converted to the text, Further, processing of the data leads to the audio convertion. For creating synthesised speech, it offers a cross-platform API.You can convert written text into spoken words by using the text-to-speech algorithms that are already built into the operating system.The

library is simple to use and offers a straightforward API for creating voice in Python programmes and scripts.

```python
import pyttsx3


# If you receive errors such as No module named win32com.client,
# No module named win32, or No module named win32api, you will need to additionally

class TextToSpeech:
    engine: pyttsx3.Engine

    def __init__(self, voice, rate: int, volume: float):
        self.engine = pyttsx3.init()
        if voice:
            self.engine.setProperty('voice', voice)
        self.engine.setProperty('rate', rate)
        self.engine.setProperty('volume', volume)  # Between 0 and 1

    def text_to_speech(self, text: str, save: bool = False, file_name='output.mp3'):
        self.engine.say(text)
        print('I\'m speaking...')

        if save:
            # On linux make sure that 'espeak' and 'ffmpeg' are installed
            self.engine.save_to_file(text, file_name)

        self.engine.runAndWait()

    def list_available_voices(self):
        voices: list = [self.engine.getProperty('voices')]

        for i, voice in enumerate(voices[0]):
```

Figure 4.4: manual testing of audio

Chapter 5

Results and Discussions

The classification of this work includes the procedure of Convolutional Neural Network (CNN), and Recurrent Neural Network (RNN) layers. To check the loss of the values, the Connectionist Temporal Classification layer is used. The observed outcome of this project is that the text is extracted from handwriting and thus converted to audio. A variety of scanned handwritten images with various writing styles were used to test the Handwritten Character Recognition system. The outcomes were incredibly positive. The suggested system pre-processes the image to get rid of the noise. The proposed system has an advantage as it uses a small number of images for training and thus gives an effective result.

For the images with handwritten text in various writing styles, sizes, and alignments against a variety of backgrounds, the proposed technique produced good results. Even when there is noise in either the background or the characters of the image, it accurately categorizes the majority of handwritten characters.

Fig 5.1 displays the result of the handwritten words to text. After training of the model, the words are manually tested to check the desired output.

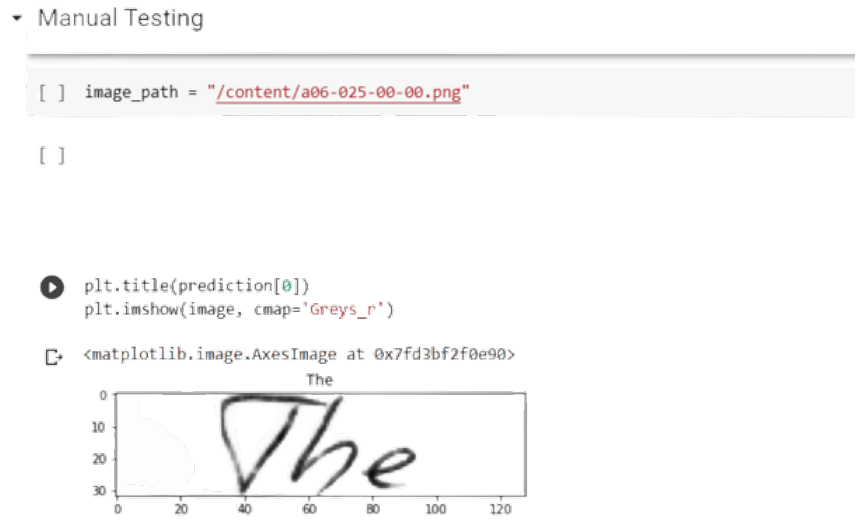

Figure 5.1: Handwritten words to Text

The proposed system undergoes a validation process to ensure accuracy, including manual checking of handwritten words to text and audio conversion verification using the Text-To-Speech (TTS) and pyttsx3 libraries. TTS is a general term for any technology that converts written text to spoken words, while pyttsx3 provides a Python API for generating synthesized speech across multiple platforms. By utilizing these libraries, the system can be manually tested to ensure the audio output accurately reflects the original text. This step is crucial for reliability, especially for individuals with visual impairments who rely on audio output for accessing information.

The developed system was tested using the pyttsx3 and TTs Python libraries, resulting in successful manual testing of text-to-audio conversion. With a variety of voice recognition options available in multiple languages through the pyttsx3 library, the system can effectively convert user input to speech. This inclusive solution pro-

```python
class TextToSpeech:
    engine: pyttsx3.Engine

    def __init__(self, voice, rate: int, volume: float):
        self.engine = pyttsx3.init()
        if voice:
            self.engine.setProperty('voice', voice)
        self.engine.setProperty('rate', rate)
        self.engine.setProperty('volume', volume)  # Between 0 and 1

    def text_to_speech(self, text: str,
                       save: bool = False, file_name='output.mp3'):
        self.engine.say(text)
        print('I\'m speaking...')

        if save:
            # On linux make sure that 'espeak' and 'ffmpeg' are installed
            self.engine.save_to_file(text, file_name)

        self.engine.runAndWait()
```

Figure 5.2: manual testing audio

vides valuable support for those with visual impairments, and its flexibility in language support makes it suitable for a broad range of users.

```python
if __name__ == '__main__':
    tts = TextToSpeech
    ('com.apple.speech.synthesis.voice.daniel', 200, 1.0)
    # tts.list_available_voices()
    tts.text_to_speech
    ("This is Nivetha from palani.", save=True, file_name='output.mp3')
```

Figure 5.3: manual testing audio

```
Python 3.9.12 (main, Apr 4 2022, 05:22:27) [MSC v.1916 64 bit (AMD64)]
Type 'copyright', 'credits' or 'license' for more information
IPython 8.2.0 -- An enhanced Interactive Python. Type '?' for help.

✓ import pyttsx3 ···

I'm speaking...
```

Figure 5.4: manual testing audio

Chapter 6

Conclusion

There are tons of new technologies growing in the Metamorphism of the world nowadays. Similarly, the requirements for these technologies are also getting increased. This research meets the demand for assistive technology. In this research, a well-developed solution to convert handwritten text to speech is provided. With this technology, blind people could be benefited mainly. While looking deeper into the application, we come to know that this can be very useful not only for the physically challenged (blind) but also for others who are unable to read or struggles to read. This technology could be more useful when they are integrated with other recent technologies. This technology is useful in educational systems also by allowing students to focus on the content rather than on the act of reading, resulting in a better understanding of the material.

6.1 Limitations

Every coin has two sides. Similarly, this Text to Speech technology also has its own drawbacks. The most evident disadvantage is that it lacks speech's natural expressiveness during the transmission of information. when Speech is translated into text, it

loses a lot of distinctive characteristics including tone, rhythm, tempo, and repetition, which lowers the cognitive load and supports comprehension. Errors can happen in the text if an error happens in recognizing the speech. Although a transcript may faithfully reproduce the spoken words, the strategic, emotive, and impact aspects of speech are lost on the written page.

6.2 Future Works

The present project has converted only the handwritten words to text. Further, it will be continued to develop the system which is used to deliver the text information by audio.Voice synthesis-based applications will continue to advance as there is a peak increase in the new technologies. The handwriting of the individuals differs in the styles, sizes and so on. Thus, this handwritten words can be highly trained by newly invented technologies for betterment of clearance in the audio quality .In future work, hybrid feature extraction methods will be developed in order to enhance the accuracy. Although the audio quality is developed, it is far from perfect yet. Thus it may be improved by tuning.Also better classification methods will be investigated in order to minimize the miss classified image.

References

1. A. Sing, A. Bist A wide scale survey on handwritten digit recognition using machine learning. Int. J. Comput Sci. Eng. 124–134 (2019)

2. J. Memon, M. Sami, R. Ahmed Khan, Mueenuddin, Handwritten optical character recognition (OCR): a comprehensive systematic literature review (SLR). IEEE Access (2020), 142642–142668

3. P. Sharma, R.K. Pamula, Handwritten text recognition using machine learning techniques in applications of NLP. Int. J. Innov. Technol. Exploring Eng. (IJITEE),1394–1397 (2019)

4. D. Prabha Devi, R. Ramya, P.S. Dinesh, C. Palanisamy, G. Sathish Kumar, Design and simulation of handwritten recognition system. Mater Today Proc Elsevier (2019)

5. Deep Learning for Handwritten Text Recognition (ConvNet RNN) Deep Learning for Handwritten Text Recognition (ConvNet RNN), July 2021

6. Am Tuan Ly, Hung Tuan Nguyen Masaki Nakagawa, 2D Self-attention Convolutional Recurrent Network for Offline Handwritten Text Recognition, sep 2021

7. P.Thangamariappan, J.C.Miraclin, J. Pamila, Handwritten recognition by using machine learning approach. Int. J. Eng. Appl. Sci. Technol. 564–567 (2020)

8. R. Geetha, T. Thilagam T. Padmavathy, Effective offline handwritten text recognition model based on a sequence-to-sequence approach with CNN–RNN networks, Jan 2021

9. S.S. Rosyda, T.W. Purboyo, A review of various handwriting recognition methods. Int. J. Appl. Eng. Res. (IJAERV) 1155–1164 (2018)

10. L. Abhishek,Optical character recognition using ensemble of SVM, MLP and extra trees classifier, in 2020 International Conference for Emerging Technology (INCET) (INCET) (IEEE 2020),1–4

11. Rohini G. Khalkar, Adarsh Singh Dikhit, Anirudh Goel, Manisha Gupta, "Handwritten Text Recognition using Deep Learning (CNN RNN)", DOI: 10.17148/IARJSET.2021.86148, IARJSET, 2021, pp. 870-881

12. R. Sharma, B. Kaushik, N. Gondhi, Character recognition using machine learning and deep learning a survey, in 2020 International Conference on Emerging Smart Computing and Informatics (ESCI) (IEEE, 2020), 341–345

13. P. Bojja, N.S.S.T. Velpuri, G.K. Pandala, S.D. Lalitha Rao Sharma Polavarapu, P.R. Kumari, Handwritten text recognition usingmachine learning techniques in applications of NLP. Int. J. Innov. Technol. Exploring Eng. (IJITEE) 1394–1397 (2019)

14. B.M.Vinijit, M.K. Bhojak, S. Kumar, G. Chalak, A review on hand-written character recognition methods and techniques, In International Conference on Communication and Signal Processing (2020), 1224–1228

15. A. Tahir, A. Pervaiz, Hand written character recognitoin using SVM. Pacific Int. J. 39–43 (2020)

16. Pooja A.Gundle, R.K. Chavan, Survey on Text to Speech Synthesis Models and Methods, nternational Journal of Scientific Engineering Research Volume 10, Issue 7, July-2019 235

17. Kurhade, J. Naveenkumar, and A. K. Kadam, "An experimental on top-k high utility itemset mining by efficient algorithm Tkowithtku," Int. J. Innov. Technol. Explor. Eng., vol. 8, no. 8 Special Issue 3, pp. 519–522, 2019.

18. Prabhat Pathak, Text to speech conversion with Python, May 31, 2020

19. Mohini Agarwal, Handwriting Text Recognition, Dec 11, 2019

20. Text to Speech Conversion October 2016, Indian Journal of Science and Technology

21. T. S. Gunawan, A. F. R. M. Noor, and M. Kartiwi, "Development of english handwritten recognition using deep neural network," 2018.

22. G. Cohen, S. Afshar, J. Tapson, and A. van Schaik, "Emnist: an extension of mnist to handwritten letters," axis preprint arXiv:1702.05373, 2017